Submitted to Bernoulli
arXiv: arXiv:1602.08296

Multiple testing of local maxima for detection of peaks on the (celestial) sphere

DAN CHENG[1] VALENTINA CAMMAROTA[2] YABEBAL FANTAYE[3] DOMENICO MARINUCCI[4] and ARMIN SCHWARTZMAN[5]

[1]*Department of Mathematics and Statistics, Texas Tech University, USA.*
E-mail: cheng.stats@gmail.com

[2]*Department of Statistics, Sapienza University of Rome, Italy.*
E-mail: valentina.cammarota@uniroma1.it

[3]*African Institute for Mathematical Sciences and Department of Mathematics, University of Stellenbosch, South Africa. E-mail:* yabi@aims.ac.za

[4]*Department of Mathematics, University of Rome Tor Vergata, Italy.*
E-mail: marinucc@mat.uniroma2.it

[5]*Division of Biostatistics, University of California, San Diego, USA.*
E-mail: armins@ucsd.edu

We present a topological multiple testing scheme for detecting peaks on the sphere under isotropic Gaussian noise, where tests are performed at local maxima of the observed field filtered by the spherical needlet transform. Our setting is different from the standard Euclidean large domain asymptotic framework, yet highly relevant to realistic experimental circumstances for some important areas of application in astronomy, namely point-source detection in cosmic Microwave Background radiation (CMB) data. Motivated by this application, we shall focus on cases where a single realization of a smooth isotropic Gaussian random field on the sphere is observed, and a number of well-localized signals are superimposed on such background field. The proposed algorithms, combined with the Benjamini-Hochberg procedure for thresholding p-values, provide asymptotic control of the False Discovery Rate (FDR) and power consistency as the signal strength and the frequency of the needlet transform get large.

Keywords: Gaussian random fields, Sphere, CMB, Height distribution, Overshoot distribution, Needlet transform, False discovery rate, Power.

1. Introduction

A classical problem of modern high-dimensional statistics is multiple testing in the presence of background noise. Applications are common in the areas of neuroimaging, genomic arrays and astronomy. These issues become particularly challenging when the background noise is allowed to exhibit more realistic properties than the simple *i.i.d.* framework, in particular when noise is modeled as a stochastic process or a random field. In this setting, important progresses have been recently obtained combining ideas from

two different streams of research, namely techniques from the multiple testing literature, such as False Discovery Rate (FDR) algorithms, and techniques to investigate excursion probabilities and local maxima for random fields [29, 7, 8]. These works have covered applications in a univariate and multivariate Euclidean setting; analytic properties have been derived under a large sample asymptotic framework, i.e., under the assumption that the domain of observations is growing steadily, together with the signals to be detected.

In this paper we introduce a related multiple testing procedure in a setting that is different from the standard Euclidean large domain asymptotic framework, yet highly relevant to realistic experimental circumstances for some important areas of application in astronomy. More precisely, we focus on cases where a single realization of a smooth isotropic Gaussian random field on the sphere is observed, and a number of well-localized signals are superimposed on such background field. This is exactly the setting for the so-called point-source detection issue in cosmic Microwave Background radiation (CMB) data experiments [24, 25, 26]. As discussed now in any modern textbook in cosmology [12], CMB data can be viewed as a single realization of an isotropic Gaussian random field, which represents a "snapshot" of the Universe at the *last scattering surface,* i.e. the time (approximately 4×10^5 years after the Big Bang, or 1.38×10^{10} years ago) when photons decoupled from electrons and protons and started to travel nearly without interactions in space. As such, CMB has been repeatedly defined as a goldmine of information on cosmology - two very successful satellite experiments, WMAP from NASA (`http://map.gsfc.nasa.gov/`) and Planck from ESA (`http://www.esa.int/Our_Activities/Space_Science/Planck`) have now produced full-sky maps of CMB radiations, and these data have been used in several thousand papers over the last few years to address a number of fundamental questions on the dynamics of the Big Bang, the matter-energy content of the Universe, the mechanisms of structure formation, and several others.

From the experimental point of view, it is very important to recall that, superimposed to CMB radiation, a number of foreground contaminants are present; as a first approximation, we can view them as point-like objects (galaxies or clusters of galaxies, typically). A major statistical challenge in the analysis of CMB data is the proper identification of such sources; on one hand this is important for the proper construction of filtered CMB maps, on the other hand these sources are of great interest on their own as proper astrophysical objects (in some cases they can be matched with existing catalogues, while in other cases they lead to new discoveries). A number of algorithms have been proposed for these tasks and these solutions have all been shown to perform well in practice, see i.e., [28] and the references therein. However, these procedures have all avoided the specific challenges of multiple testing and none of them has been shown to control in any proper statistical way any aggregate statistics such as the classical Family-Wise Error Rate (FWER) (the probability of observing any false detections at all), False Discovery Proportion (FDP) (the fraction of false detections among the detections in a given sample) or False Discovery Rate (FDR) (the expected FDP in a population).

Our purpose in this paper is to develop in such a spherical framework a rigorous statistical procedure to control error rates in a multiple testing framework. Our starting idea is to extend to these circumstances the STEM procedure advocated in [29, 8], and investigate rigorously its statistical properties. While our construction follows in several

ways what was earlier done by these authors, we wish to stress that the new spherical framework poses some major technical and foundational new challenges.

The first of these new challenges is the proper definition of filters and point-like signals in a spherical framework. Here, natural solutions can be found by exploiting recent developments in the analysis of spherical random fields an spherical wavelets. In particular, we can define bell-shaped signals by adopting a natural definition of a Gaussian distribution on the sphere, motivated in terms of diffusion processes; likewise, filtering can be implemented by wavelet techniques - we find particularly convenient the Mexican needlet construction introduced by [13, 14], see also [27] for some earlier applications to CMB data.

A second, more delicate, issue is the rigorous investigation of asymptotic statistical properties. A crucial staple of the STEM algorithm is the possibility to control the FDR, assuming convergence of the empirical distribution of local maxima to its theoretical counterpart. In standard settings this can be done by resorting to ergodicity properties, assuming that the domain of the observations grows larger and larger. This form of ergodic properties cannot be exploited here because our spherical domain is compact. We shall hence require a convergence result on the empirical distribution of local maxima in a high-frequency/fixed domain setting, i.e., as the observed frequencies diverge: this result extends to the case of the needlet transform some related computations which were recently performed in [6] for the case of random spherical harmonics. In this sense, our setting is related to the increasingly popular fixed-domain asymptotics approach for the analysis of random fields [18].

The plan of this paper is as follows. Our basic setting and model is introduced in Section 2, with Section 3 devoted to a careful discussion on the nature and effects of filtering. Section 5 provides a description of the multiple testing scheme and discusses the error and power definitions, including the derivation of asymptotic p-values and our adoption of Benjamini and Hochberg's (1995) pioneering approach [4]. The proofs of FDR control and tests consistency are collected in Section 6, while in Section 7 we provide numerical results on the empirical performance of the proposed procedures in simulated CMB fields.

Because of the technical issues involved in processing of real astrophysical data, we limit ourselves in this paper to the algorithmic and theoretical issues described above and defer the application to real CMB data to follow-up work.

2. The model

As motivated above, our purpose here is to represent a situation where a large number of "point sources", defined here as "signal", are superimposed on some isotropic background "noise". Of course, the notions of "noise" and "signal", here as in any other motivating field, are very much conventional. For CMB-oriented applications the background Gaussian field is the primary object of physical interest, while the super-imposed point sources are contaminants to be removed; in other astrophysical areas, the identification of the sources may be by itself a major scientific goal (a very recent catalogue of detected

point sources/astrophysical objects is given for instance by [26]). Moreover, "follow-up" observations of these sources by means of ground telescopes are expensive, hence we need to minimize false detections, in a suitable sense to be discussed below.

2.1. Spherical Gaussian

To introduce our model for the signal, we first need to justify the notion of a "Gaussian-shaped" density on the sphere. Our "point source" signal will be built from a set of such bell-shaped distributions. Let $\mathbb{S}^2$ denote the unit sphere in $\mathbb{R}^3$. The diffusion equation on the sphere is given by

$$\frac{\partial}{\partial t}h(x;t,x_0) = -\Delta_{\mathbb{S}^2}\, h(x;t,x_0), \qquad h(x;0,x_0) = \delta_{x_0}(x), \qquad x \in \mathbb{S}^2,$$

where $\Delta_{\mathbb{S}^2}$ is the Laplacian operator of $\mathbb{S}^2$ and $\delta_{x_0}(x)$ is the Dirac's delta measure centered at $x_0 \in \mathbb{S}^2$, and the second line is to be understood in the weak sense. It is standard to write the solution in terms of diffusion operators as

$$h(x;t,x_0) = \exp(-t\Delta_{\mathbb{S}^2}\,)h(x;0,x_0) = \sum_{\ell} \frac{2\ell+1}{4\pi}\exp(-t\lambda_\ell)P_\ell(\langle x_0, x\rangle), \tag{2.1}$$

where again the first equality is to be understood in the weak sense, $-\lambda_\ell := -\ell(\ell+1)$ denotes the set of eigenvalues of the spherical Laplacian, $\ell = 1, 2, ...$, while $\{P_\ell(\cdot)\}$ represents the family of Legendre polynomials

$$P_\ell(u) = \frac{(-1)^\ell}{2^\ell}\frac{d^\ell}{du^\ell}(1-u^2)^\ell, \qquad \ell = 1, 2, \ldots, \tag{2.2}$$

i.e., $P_1(u) = u$, $P_2(u) = (3u^2-1)/2$, $P_3(u) = (5u^3-3u)/2$, etc., and $\langle\cdot,\cdot\rangle$ denotes inner product on $\mathbb{R}^3$. By a straightforward analogy with the Euclidean case, it is natural/customary to view $h(\cdot;t,x_0)$ as the density on $\mathbb{S}^2$ of a spherical Gaussian centered on x_0 and having variance t.

2.2. Signal plus noise model

We can hence introduce the following sequence of signal-plus-noise models, for $N = 1, 2, \ldots$,

$$y_N(x) = \mu_N(x) + z(x), \qquad x \in \mathbb{S}^2, \tag{2.3}$$

where $\mu_N(x)$ denotes a sequence of deterministic functions on the sphere defined by

$$\mu_N(x) = \sum_{k=1}^{N} a_k h(x; t_{k,N}, \xi_k), \qquad a_k > 0, \tag{2.4}$$

and $h(x; t_{k,N}, \xi_k)$ is the family of "spherical Gaussian distributions" on $\mathbb{S}^2$ (centered on ξ_k and with variance $t_{k,N}$) which we introduced in (2.1). As mentioned earlier, we will set $t_{k,N} \to 0$ as $N \to \infty$, so that each kernel $h(x; t_{k,N}, \xi_k)$ will become in the limit more and more concentrated around its center ξ_k.

Let us now focus on the "noise" component $z(x)$. Assume that $\{z(x),\ x \in \mathbb{S}^2\}$ is Gaussian, zero-mean and isotropic. For such fields, it is well-known that the following representation holds in the mean square sense [19]:

$$z(x) = \sum_{\ell=1}^{\infty} z_\ell(x), \qquad z_\ell(x) = \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(x), \tag{2.5}$$

where $\{Y_{\ell m}(.)\}$ denotes the family of spherical harmonics, which form an orthonormal basis for the $L^2(\mathbb{S}^2)$ space of square integrable functions on the sphere (see for instance [19], Chapters 3 and 5); on the other hand, $\{a_{\ell m}\}$ denotes the uncorrelated array of random spherical harmonic coefficients, which can be obtained by means of the spherical Fourier transform

$$a_{\ell m} = \int_{\mathbb{S}^2} z(x) \overline{Y}_{\ell m}(x) dx; \tag{2.6}$$

here, $\mathbb{E}[a_{\ell m} \overline{a}_{\ell' m'}] = C_\ell \delta_\ell^{\ell'} \delta_m^{m'}$, where δ_a^b is the Kronecker delta function, and the sequence $\{C_\ell\}$ represents the so-called angular power spectrum of the field; (2.5) is hence the Karhunen-Loeve expansion on the sphere. As pointed out in [20], under isotropy and finite-variance the sequence $\{C_\ell\}$ necessarily satisfies $\sum_\ell C_\ell(2\ell+1)/(4\pi) = \mathbb{E}[z^2(x)] < \infty$ and the random field $z(x)$ is mean square continuous, meaning that $\lim_{x \to x_0} \mathbb{E}(z(x) - z(x_0))^2 = 0$ for all $x_0 \in \mathbb{S}^2$. Its covariance function is given by

$$\Gamma(x_1, x_2) = \mathbb{E}\left[z(x_1) z(x_2)\right] = \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} C_\ell P_\ell(\langle x_1, x_2 \rangle), \qquad x_1, x_2 \in \mathbb{S}^2.$$

The Fourier components $\{z_\ell(x)\}$, can be viewed as random eigenfunctions of the spherical Laplacian:

$$\Delta_{\mathbb{S}^2} z_\ell = -\ell(\ell+1) z_\ell, \qquad \ell = 1, 2, \ldots;$$

the asymptotic behaviour of critical points for $z_\ell(x)$ has been studied for instance by [6].

3. Filtering and smoothing

An important step in the implementation of the STEM algorithm is smoothing of the observed data. Given the very delicate nature of the asymptotic results in our setting, the definition of the kernel function requires here special care. We shall propose here to adopt a kernel which is based upon the so-called Mexican needlet construction introduced by [13]; see also [17, 22, 27] for the investigation of stochastic properties and statistical applications of these techniques.

Mexican needlets can be viewed as a natural development of the standard needlet frame which was introduced by [23]. Loosely speaking, Mexican needlets differ from the standard needlet construction inasmuch as they allow for providing a kernel which is unboundedly supported in the spectral domain; they can hence be shown to enjoy better localization properties in the spatial domain, i.e., faster (Gaussian rather than nearly exponential) decay of their tails. For our purposes, these better localization properties in the spatial domain turn out to be very important, as they allow a tight control of leakage in the signals.

The Mexican needlet transform of order $p \in \mathbb{N}$ can be defined by

$$\Psi_j(\langle x_1, x_2\rangle) = \Psi_j(\langle x_1, x_2\rangle; B, s) := \sum_{\ell=0}^{\infty} b\left(\frac{\ell}{B^j}; s\right) \frac{2\ell+1}{4\pi} P_\ell(\langle x_1, x_2\rangle); \tag{3.1}$$

here the function $b(\cdot; s)$ is defined by $b(u; s) = u^{2s}e^{-u^2}$, with $u \in \mathbb{R}_+$; it is easily seen to belong to the Schwartz class (i.e., all its derivatives decay faster than any polynomial). The user-chosen integer parameter s is usually labelled p in the literature, but we avoid this choice here for possible notational conflicts with $p-$values; it has been shown that higher values of s entail better properties in the frequency domain, but worse spatial localization, in particular, a higher number of sidelobes. Finally, $B > 1$ plays the role of a bandwidth parameter: for any given value j, a higher value of B implies better localization properties in the spatial domain, and worse in the spectral domain; values in the range 1.1-1.3 have been shown to perform well in practice [27].

4. High frequency asymptotics

Our asymptotic theory will be developed in the so-called "high-frequency" framework as the filtering frequency j increases. To achieve non-trivial limits, we shall allow the number of point sources N to grow with j ($N = N_j \to \infty$ as $j \to \infty$), and their shape to become more and more localized ($t_{k,N} = t_{k,N_j} \to 0$ as $N_j \to \infty$). Under this regime, in what follows we shall only use the index j for asymptotics.

The above conditions can be understood by an analogy with the (now standard) high-dimensional asympotics framework where the number of parameters is allowed to grow to infinity in the presence of growing number of observations. From the point of view of astronomy, this reflects a growing number of sharper and sharper sources as the resolution of our experiments grow better and better, or equivalently as the scales that we are able to probe become smaller and smaller.

It should be noted that, in the absence of these conditions, all our procedures to follow have properties that can be trivially established: in particular, the power of our detection procedures is very easily seen to converge to unity. More explicitly, we believe that our setting is meaningful and relevant as a guidance for applied scientists; indeed, in many circumstances the cardinality of possible signals (i.e., the number of potential galactic sources in an astrophysical environment) is in the order of several thousands, so it seems

more useful to consider this quantity as diverging to infinity together with the number of observations.

To formally achieve the right balance, we shall introduce the following assumption.

Condition 1. *Let* $t_{N_j} = \max_{k=1,\dots,N} t_{k,N_j}$. *As* $j \to \infty$, $t_{N_j} B^{2j} \to 0$.

In words, we are assuming that both the signal and the filter become more and more localized, the former more rapidly to make identification meaningful.

4.1. The filtered signal

Applied to (2.3), the kernel transform (3.1) produces the sequence of observable smoothed fields

$$y_j(x) := \mu_j(x) + \beta_j(x), \tag{4.1}$$

where $\mu_j(x)$ denotes the smoothed signal and $\beta_j(x)$ denotes the smoothed noise (the change in notation in Eq. (4.1) with respect to Eq. (2.3) reflects the dependence of all the quantities, including N, on j). We describe both components in detail below.

For the filtered signal we obtain

$$\begin{aligned}\mu_j(x) &:= \sum_{k=1}^{N} a_k \langle \Psi_j(\langle x, y\rangle), h(y; t_{k,N}, \xi_k)\rangle_{L^2(\mathbb{S}^2)} \\ &= \sum_{k=1}^{N}\sum_{\ell=0}^{\infty} a_k b\left(\frac{\ell}{B^j}; s\right) \exp(-\ell(\ell+1)t_{k,N}) \frac{2\ell+1}{4\pi} P_\ell(\langle \xi_k, x\rangle).\end{aligned}$$

In words, both the filtered noise and signals are averaged versions, in the spectral domain, of (random and deterministic, respectively) Fourier components. We neglect the numerical approximations that can arise when computing these integrals on real data (these numerical approximations are certainly of lower order with current packages such as Healpix [15]).

For further analysis, it is convenient to introduce the simple approximation

$$\begin{aligned}\mu_j(x) &= \sum_{k=1}^{N}\sum_{\ell=0}^{\infty} a_k \left(\frac{\ell}{B^j}\right)^{2s} \exp\left(-\ell(\ell+1)t_{k,N} - B^{-2j}\ell^2\right) \frac{2\ell+1}{4\pi} P_\ell(\langle \xi_k, x\rangle) \\ &= \sum_{k=1}^{N}\sum_{\ell=0}^{\infty} a_k \left(\frac{\ell}{B^j}\right)^{2s} \exp\left(-B^{-2j}\ell^2\right) \frac{2\ell+1}{4\pi} P_\ell(\langle \xi_k, x\rangle) + o_j(1) \\ &= \sum_{k=1}^{N} a_k \Psi_j(\langle x, \xi_k\rangle) + o_j(1),\end{aligned}$$

where the second line can be easily justified resorting to Condition 1 above. It is also known that the smoothing filter $\Psi_j(\langle \cdot, \xi_k\rangle)$, for any $\xi_k \in \mathbb{S}^2$, has Gaussian tails (up to a

polynomial factor), and hence decays faster than exponentially; more precisely one has that there exists a constant C_s such that [13, 14]

$$|\Psi_j(\langle x, \xi_k\rangle)| \leq C_s B^{2j} e^{-\frac{B^{2j} d^2(x,\xi_k)}{4}} \left(1 + \left|H_{2s}\left(B^j d(x, \xi_k)\right)\right|\right), \tag{4.2}$$

where $d(x, y) = \arccos(\langle x, y\rangle)$ is the standard geodesic distance on the sphere and $H_q(\cdot)$ denotes the Hermite polynomial of degree q, which is defined by

$$H_q(x) = (-1)^q e^{x^2/2} \frac{d^q}{dx^q}\left(e^{-x^2/2}\right).$$

It is also possible to provide a useful analytic approximation for the functional form of the needlet filter at high frequencies j; indeed Geller and Mayeli [13] proved the following.

Lemma 4.1. *Let $s = 1$ and let $\xi_k \in \mathbb{S}^2$ be fixed. Then as $j \to \infty$,*

$$\Psi_j(\langle x, \xi_k\rangle) = g(d(x, \xi_k))(1 + O(B^{-2j})),$$

where

$$g(\theta) = \frac{1}{4\pi} B^{2j} e^{-\frac{B^{2j}\theta^2}{4}} \left(1 - \frac{B^{2j}\theta^2}{4}\right), \quad \theta \in [0, \pi],$$

and d is the standard geodesic distance on the sphere.

As a consequence, we have the following analytic expression for our signal when $s = 1$, as $j \to \infty$:

$$\mu_j(x) = \sum_{k=1}^{N} a_k \frac{1}{4\pi} B^{2j} e^{-\frac{B^{2j} d^2(x,\xi_k)}{4}} \left(1 - \frac{B^{2j} d^2(x, \xi_k)}{4}\right)(1 + O(B^{-2j})) + o_j(1). \tag{4.3}$$

It is readily verified that the function $g(\cdot)$ has the global maximum $g(0) = \frac{1}{4\pi} B^{2j}$ and a local minimum $g(2\sqrt{2}B^{-j}) = -\frac{1}{4\pi} e^{-2} B^{2j}$.

4.2. The filtered noise

Our next step is to focus on the sequence of filtered noise fields.

Recalling the standard addition theorem for spherical harmonics (see [19], Eq. (3.42))

$$\sum_{m=-\ell}^{\ell} Y_{\ell m}(x_1)\overline{Y}_{\ell m}(x_2) = \frac{2\ell + 1}{4\pi} P_\ell(\langle x_1, x_2\rangle),$$

it is then easy to see that the "filtered noise" is given by

$$\begin{aligned}\beta_j(x) &:= \langle \Psi_j(\langle x, y\rangle), z(y)\rangle_{L^2(\mathbb{S}^2)} \\ &= \sum_{\ell=0}^{\infty} b\left(\frac{\ell}{B^j}; s\right) \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(x) = \sum_{\ell=0}^{\infty} b\left(\frac{\ell}{B^j}; s\right) z_\ell(x),\end{aligned}$$

where the last line is due to (2.5).

As derived in (4.4), they can be expressed as averaged forms of random spherical eigenfunctions,

$$\beta_j(x) = \sum_{\ell=1}^{\infty} b\Big(\frac{\ell}{B^j}; s\Big) z_\ell(x), \qquad j = 1, 2, \dots \tag{4.4}$$

It is convenient to normalize these fields to have unit variance, as follows:

$$\tilde{\beta}_j(x) = \frac{\beta_j(x)}{\sqrt{\mathbb{E}[\beta_j^2(x)]}}, \qquad j = 1, 2, \dots \tag{4.5}$$

Let us define also

$$\tilde{y}_j = \frac{y_j}{\sqrt{\mathbb{E}[\beta_j^2(x)]}} = \tilde{\beta}_j + \frac{\mu_j}{\sqrt{\mathbb{E}[\beta_j^2(x)]}}. \tag{4.6}$$

In this paper, we assume that the angular power spectrum of the spherical noise field is known, and from this it is immediate to compute the normalizing variance needed to construct $\widetilde{\beta}_j(x)$ and $\widetilde{y}_j$. In practical applications, it should be noted that there are usually both tight theoretical predictions on the angular power spectrum and very efficient empirical estimates which converge very rapidly to the true values, in the high-frequency sense, so that these statistics can be usually implemented without difficulties.

A rigorous investigation of the asymptotic properties of these smoothed fields requires some mild regularity assumptions on the power spectrum C_ℓ, which are customary in this branch of literature. More precisely (see for instance [19], page 257, or [2, 21, 17, 22]),

Condition 2. *There exist $M \in \mathbb{N}, \gamma > 2$ and a function $G(\cdot) \in C^\infty$ such that*

$$C_\ell = \ell^{-\gamma} G(\ell) \tag{4.7}$$

where $G(\ell) > 0$ for all ℓ, and for some $c_1, \dots, c_M > 0$ and $r = 1, \dots, M$, we have

$$\sup_u \left| \frac{d^r}{du^r} G(u) \right| \leq c_r u^{-r}.$$

Condition 2 entails a weak smoothness requirement on the behaviour of the angular power spectrum, which is satisfied by cosmologically relevant models; for instance, this condition is fulfilled by models of the form (4.7), where $G(\ell) = P(\ell)/Q(\ell)$ and $P(\ell), Q(\ell) > 0$ are two positive polynomials of the same order. In what follows we denote by G_0 the limit $G_0 := \lim_{\ell \to \infty} G(\ell)$.

Under Condition 2, it is possible to establish an upper bound on the correlation function of $\{\tilde{\beta}_j(\cdot)\}$, as follows (see [17, 22] for a proof).

Proposition 4.2. *Assume Condition 2 holds with $\gamma < 4s+2$ and $M \geq 4s+2-\gamma$; then there exists a constant $K_M > 0$, not depending on j, x, and y, such that the following inequality holds*

$$|\mathrm{Cor}(\beta_j(x), \beta_j(y))| \leq \frac{K_M}{(1+j^{-1}B^j d(x,y))^{4s+2-\gamma}}, \tag{4.8}$$

where $d(x,y) = \arccos(\langle x, y\rangle)$ is the standard geodesic distance on the sphere.

The inequality (4.8) is qualitatively similar to others which were earlier established in the case of standard needlets; see for instance [2]. A quick comparison with the results in [2] shows an important difference, namely that the rate of decay for the bound on the right-hand side depends on the shape of the kernel (in particular, on the parameter s) and on the rate of decay of the angular power spectrum (i.e., on the parameter γ); none of these values affect the rate of convergence in the standard needlet case. As a consequence, in the case of Mexican needlets, asymptotic uncorrelation only holds under the assumption that $\gamma < 4s+2$, so that higher values of s are needed to ensure asymptotic uncorrelation for larger values of γ. We believe this issue can be easily addressed by a plug-in procedure; for instance, for the CMB applications we mentioned earlier there are strong theoretical motivations and experimental constraints that allow to set $2 < \gamma < 3$, so that taking $s = 1$ is already enough to ensure the correlation function decays to zero: ample numerical evidence on the uncorrelation properties of Mexican needlets is collected in [27]. The term j^{-1} appearing in the denominator of (4.8) is a consequence of some standard technical difficulties when dealing with boundary cases such as $M = 4s+2-\gamma$.

Recall that a Gaussian random field $\{X(x),\ x \in \mathbb{S}^2\}$ is isotropic if and only if its covariance function is invariant to rotations, i.e.

$$\Gamma(x,y) := \mathrm{Cov}(X(x), X(y)) = \mathrm{Cov}(X(g\cdot x), X(g\cdot y))$$

for all elements $g \in SO(3)$ of the three-dimensional group of rotations. For such fields the covariance functions can be expressed by

$$\Gamma(x,y;\{C_\ell\}_{\ell=1,2,\dots}) = \sum_{\ell=1}^{\infty} \frac{2\ell+1}{4\pi} C_\ell P_\ell(\langle x, y\rangle);$$

it turns now to be convenient to make explicit the dependence of the covariance function from the angular power spectrum sequence $\{C_\ell\}_{\ell=1,2,\dots}$. The fields $\beta_j(x)$ are obviously isotropic and from Proposition 4.2 it is immediate to obtain a bound on the covariance (rather than correlation) function, indeed we have

$$\Gamma_{j,s}(x,y) := \mathbb{E}[\beta_j(x)\beta_j(y)] = \sum_{\ell=1}^{\infty} b^2\Big(\frac{\ell}{B^j}; s\Big) C_\ell \frac{2\ell+1}{4\pi} P_\ell(\langle x, y\rangle)$$

$$= \Gamma(x,y;\{b^2\Big(\frac{\ell}{B^j}; s\Big) C_\ell\}_{\ell=1,2,\dots}) \leq \frac{K_M}{(1+j^{-1}B^j d(x,y))^{4s+2-\gamma}} \sum_{\ell=1}^{\infty} b^2\Big(\frac{\ell}{B^j}; s\Big) C_\ell \frac{2\ell+1}{4\pi}.$$

For the implementation of our multiple testing procedures, we shall need to write down an analytic formula for the asympotic distribution of maxima of the noise components; to this aim, we need the exact limiting behaviour of higher-order derivatives of the covariance function, evaluated at the origin. Let us first introduce the functions

$$c_{s,2n}(\gamma) := 2^{\gamma/2-2-n-2s}\Gamma(1-\gamma/2+n+2s), \qquad \Gamma(t) := \int_0^\infty x^{t-1}\exp(-x)dx.$$

Let us also define

$$\Gamma'(C_\ell) := \sum_{\ell=1}^{\infty} \frac{2\ell+1}{4\pi} C_\ell P'_\ell(1), \qquad \Gamma''(C_\ell) := \sum_{\ell=1}^{\infty} \frac{2\ell+1}{4\pi} C_\ell P''_\ell(1),$$

where

$$P'_\ell(1) = \frac{\ell(\ell+1)}{2} \quad \text{and} \quad P''_\ell(1) = \frac{\ell(\ell-1)(\ell+1)(\ell+2)}{8}$$

represent the derivatives of the Legendre polynomials evaluated at 1. Moreover, let us write also

$$\eta_j = \frac{\sqrt{\Gamma'(C_\ell)}}{\sqrt{\Gamma''(C_\ell)}}, \qquad \kappa_j = \frac{\Gamma'(C_\ell)}{\sqrt{\Gamma''(C_\ell)}}. \tag{4.9}$$

In the sequel, it should be kept in mind that η_j and κ_j are for the standardized fields $\tilde{\beta}_j$ (and not for β_j).

Proposition 4.3. *As $j \to \infty$, we have*

$$\Gamma'(\tilde{\beta}_j) \sim \frac{c_{s,2}(\gamma)}{2c_{s,0}(\gamma)} B^{2j}, \quad \Gamma''(\tilde{\beta}_j) \sim \frac{c_{s,4}(\gamma)}{8c_{s,0}(\gamma)} B^{4j},$$

and therefore

$$\eta_j \sim \frac{2\sqrt{c_{s,2}(\gamma)}}{\sqrt{c_{s,4}(\gamma)}} B^{-j}, \qquad \kappa_j \sim \frac{\sqrt{2}c_{s,2}(\gamma)}{\sqrt{c_{s,0}(\gamma)c_{s,4}(\gamma)}};$$

where $a_j \sim b_j$ denotes $\lim_{j\to\infty} a_j/b_j = 1$.

The proof is provided in the Supplement. We note that the expressions for η_j and κ_j will be used in the applied sections below for the numerical evaluation of p-values. In the sequel, we let $s = 1$; all the results below can be trivially extended to choices of other forms of filtering, with different values of s (we stress that the case $s = 1$ is the choice that has been usually adopted for applications, see [28, 27]).

5. The multiple testing scheme

5.1. The signal and null regions

To properly approach the detection of point sources as a multiple testing problem, we first need to carefully define the spatial region occupied by the needlet-transformed point sources.

Define the *signal region* $\mathbb{D}_1^\rho = \cup_{k=1}^N D(\xi_k, \rho)$ and *null region* $\mathbb{D}_0^\rho = \mathbb{S}^2 \setminus \mathbb{D}_1^\rho$, where $\rho > 0$ is a pre-specified location tolerance parameter and $D(\xi_k, \rho)$ is the geodesic ball on $\mathbb{S}^2$ with center ξ_k and radius ρ. The presence of a tolerance parameter is not only required to settle properly the theoretical framework, but is also consistent with the common scientific practice, see again [28]. We introduce now a further condition.

Condition 3. *As $j \to \infty$, we have*

$$\rho = \rho_j \sim j^\nu B^{-j} \ , \ \min_{1 \le k \ne k' \le N} d(\xi_k, \xi_{k'}) > 2\rho, \ \textit{and} \ \min_{1 \le k \le N} a_k > a_0, \quad \forall N,$$

where $d(x, y)$ is the geodesic distance on the sphere, and ν and a_0 are positive constants.

Condition 3 is meant to ensure that the tolerance radius ρ decays to zero asymptotically faster than the distance between separate sources. This assumption is needed for our theoretical results below; in particular, for the power consistency argument to go through we require some form of convexity to hold (with high probability) in neighbourhoods around the maxima of radius equal to the tolerance parameter, and this would not necessarily be the case should two maxima be "too close" (see i.e., the proof of Lemma 6.10). Notice that the restriction $\min_{1 \le k \ne k' \le N} d(\xi_k, \xi_{k'}) > 2\rho$ in Condition 3 yields

$$\text{Area}(D(\xi_1, \rho_j))N_j < 4\pi = \text{Area}(\mathbb{S}^2),$$

for every j, implying that the area of null region is always positive and that N_j cannot grow too fast, specifically $N_j = O(\rho_j^{-2}) = O(j^{-2\nu} B^{2j})$. It is easy to check that

$$\text{Area}(D(\xi_1, \rho_j)) = 2\pi(1 - \cos \rho_j) \sim \pi \rho_j^2 \sim \pi j^{2\nu} B^{-2j}.$$

Here are some examples for Condition 3. If $N_j = \lfloor B^{2j(1-\delta)} \rfloor$ for some $0 < \delta < 1/2$ ($\lfloor . \rfloor$ denoting integer part), then the area of signal region tends to 0. If $N_j \sim c_0 \rho_j^{-2} \sim c_0 j^{-2\nu} B^{2j}$ for some $c_0 \in (0, 4)$, then the area of signal region tends to πc_0.

5.2. The STEM algorithm on the sphere

As some general notation, for a smooth Gaussian random field $\{X(x), x \in \mathbb{S}^2\}$, define the number of local maxima of X exceeding the level $u \in \mathbb{R}$ over a domain $D \subset \mathbb{S}^2$ as

$$M_u(X; D) = \#\left\{x \in D : X(x) > u, \ \nabla X(x) = 0, \ \nabla^2 X(x) \prec 0\right\}; \tag{5.1}$$

here $\nabla X(x)$ and $\nabla^2 X(x)$ denote the gradient and Hessian of the field X at x, and $\nabla^2 X(x) \prec 0$ means the Hessian $\nabla^2 X(x)$ is negative definite. The gradient and Hessian can be computed as $\nabla X = (E_1 X, E_2 X)$ and $\nabla^2 X = (E_i E_j X)_{1\le i,j\le 2}$, respectively, where E_1 and E_2 are orthonormal tangent vectors. In spherical coordinates, at any given point $x = (\theta, \varphi), 0 \le \theta \le \pi$, $0 \le \varphi < 2\pi$ these are given by

$$E_{1,x} = \frac{\partial}{\partial \theta}, \quad E_{2,x} = \frac{1}{\sin\theta}\frac{\partial}{\partial \varphi}.$$

For convenience, denote by $M(X; D) = M_{-\infty}(X; D)$ the total number of local maxima of X over D.

Suppose now we observe $y_N(t)$ on $\mathbb{S}^2$ defined by (2.3). In accordance to [29, 8], we call the following procedure STEM (Smoothing and TEsting of Maxima).

Algorithm 1 (STEM algorithm)**.**

1. Smoothing*: For a given j, apply the needlet transform to the observed field* (2.3) *to obtain the filtered field* (4.1)*. Normalize by the (known) noise variance to obtain the field* (4.6)*.*
2. Candidate peaks*: Find the set of local maxima of $\tilde{y}_j(x)$ on $\mathbb{S}^2$*
$$\tilde{T}_j = \left\{x \in \mathbb{S}^2 : \ \nabla \tilde{y}_j(x) = 0, \ \nabla^2 \tilde{y}_j(x) \prec 0\right\}. \tag{5.2}$$
3. p-values*: For each $x \in \tilde{T}_j$, compute the p-value $p_j(x)$ for testing the (conditional) hypothesis*
$$\begin{aligned} \mathcal{H}_0(x) &: \ \{\mu_j(x') = 0 \text{ for all } x' \in D(x, \rho_j)\} \quad \text{vs.} \\ \mathcal{H}_A(x) &: \ \{\mu_j(x') > 0 \text{ for some } x' \in D(x, \rho_j)\} \end{aligned} \tag{5.3}$$
where $B(x, \rho_j)$ is a geodesic ball centered at x on the sphere and of radius equal to the tolerance radius ρ_j.
4. Multiple testing*: Notice that $M(\tilde{y}_j; \mathbb{S}^2) = \#\{x \in \tilde{T}_j\}$ is the number of tested hypotheses. Perform a multiple testing procedure on the set of $M(\tilde{y}_j; \mathbb{S}^2)$ p-values $\{p_j(x), x \in \tilde{T}_j\}$, and declare significant all local maxima whose p-values are smaller than the significance threshold.*

Next, we carefully define detection errors and power for this testing scheme.

5.3. Error and power definitions

We shall consider two forms of inference, one with fixed significance threshold u and one where the significance threshold is determined by the data according to a multiple testing procedure (see Section 5.4). For fixed $u \in \mathbb{R}$, denote by $\tilde{T}_j(u)$ the set of local maxima

of $\tilde{y}_j$ exceeding u defined via (5.1). Define the total number of detected peaks and the number of falsely detected peaks as

$$R_j(u) = M_u(\tilde{y}_j; \mathbb{S}^2), \quad V_{\rho_j}(u) = M_u(\tilde{y}_j; \mathbb{D}_0^{\rho_j}), \tag{5.4}$$

respectively. Both are defined as zero if $\tilde{T}_j(u)$, is empty. As usual, the False Discovery Proportion (FDP) is proportion of falsely detected peaks, i.e.

$$\mathrm{FDP}_{\rho_j}(u) = \frac{V_{\rho_j}(u)}{R_j(u) \vee 1} = \frac{M_u(\tilde{y}_j; \mathbb{D}_0^{\rho_j})}{(M_u(\tilde{y}_j; \mathbb{D}_0^{\rho_j}) + M_u(\tilde{y}_j; \mathbb{D}_1^{\rho_j})) \vee 1}, \tag{5.5}$$

while the False Discovery Rate (FDR) is the expected FDP, i.e.

$$\mathrm{FDR}_{\rho_j}(u) = \mathrm{E}\left\{ \frac{V_{\rho_j}(u)}{R_j(u) \vee 1} \right\}. \tag{5.6}$$

We shall denote $W_{\rho_j}(u) = R_j(u) - V_{\rho_j}(u)$.

Finally, again following the same conventions as in [29, 8], we define the power of Algorithm 1 as the expected fraction of true discovered peaks

$$\mathrm{Power}_{\rho_j}(u) = \mathrm{E}\left(\frac{1}{N_j} \sum_{k=1}^{N_j} \mathbb{1}_{\{\tilde{T}_j(u) \cap D(\xi_k, \rho_j) \neq \emptyset\}} \right) = \frac{1}{N_j} \sum_{k=1}^{N_j} \mathrm{Power}_{\rho_j,k}(u), \tag{5.7}$$

where $\mathrm{Power}_{\rho_j,k}(u)$ is the probability of detecting peak k

$$\mathrm{Power}_{\rho_j,k}(u) = \mathbb{P}\left(\tilde{T}_j(u) \cap D(\xi_k, \rho_j) \neq \emptyset \right). \tag{5.8}$$

The indicator function in (5.7) ensures that only one significant local maximum is counted within the same peak support, so power is not inflated. In other words, no matter if the number of elements in T_j falling into D_k^j is one or larger the value of the indicator function will be always unity.

5.4. P-values and BH procedure

For each observed local maximum, its P-value is the probability, under the null hypothesis, of obtaining such height as the one observed or higher. Given the observed heights $\tilde{y}_j(x)$ at the local maxima $x \in \tilde{T}_j$, the p-values in step (3) of Algorithm 1 are computed as $p_j(t) = F_j\left(\tilde{y}_j(t)\right)$, $t \in \tilde{T}_j$, where

$$F_j(u) = \mathbb{P}\left(\tilde{\beta}_j(x) > u \;\middle|\; x \in \tilde{T}_j \right) \tag{5.9}$$

denotes the right tail probability of $\tilde{\beta}_j(x)$ at the local maximum $x \in \tilde{T}_j$, evaluated under the null hypothesis that there are no point masses at all, i.e., $\mu_N(x) = 0, \forall x$.

Note that, since only local maxima are evaluated as candidates for point sources, the probability (5.9) must be conditional on the location x being a local maximum. Such probabilities have been derived in [9]. Applying the technique there, we have that

$$F_j(u) = \frac{\mathbb{E}[M_u(\tilde{\beta}_j;\mathbb{S}^2)]}{\mathbb{E}[M(\tilde{\beta}_j;\mathbb{S}^2)]} = \int_u^\infty f_j(x)\,dx, \tag{5.10}$$

where

$$\begin{aligned} f_j(x) = \frac{2\sqrt{3+\eta_j^2}}{2+\eta_j^2\sqrt{3+\eta_j^2}} \Bigg\{ & \left[\eta_j^2 + \kappa_j^2(x^2-1)\right]\phi(x)\Phi\left(\frac{\kappa_j x}{\sqrt{2+\eta_j^2-\kappa_j^2}}\right) \\ & + \frac{\kappa_j\sqrt{(2+\eta_j^2-\kappa_j^2)}}{2\pi} x e^{-\frac{(2+\eta_j^2)x^2}{2(2+\eta_j^2-\kappa_j^2)}} \\ & + \frac{\sqrt{2}}{\sqrt{\pi(3+\eta_j^2-\kappa_j^2)}} e^{-\frac{(3+\eta_j^2)x^2}{2(3+\eta_j^2-\kappa_j^2)}} \Phi\left(\frac{\kappa_j x}{\sqrt{(2+\eta_j^2-\kappa_j^2)(3+\eta_j^2-\kappa_j^2)}}\right) \Bigg\}. \end{aligned}$$

Here η_j and κ_j are defined in (4.9) above, and $\phi(x)$ and $\Phi(x)$ denote the standard normal density and cumulative distribution functions, respectively.

As a multiple testing procedure in step (4) of Algorithm 1, we apply the Benjamini-Hochberg (BH) procedure [4]. In the context of this work, it is implemented directly as follows. Consider the N_j p-values in increasing order. For a fixed significance level $\alpha \in (0,1)$, let k be the largest index for which the ith smallest P-value is less than $i\alpha/M(\tilde{y}_j;\mathbb{S}^2)$. Then the null hypothesis $\mathcal{H}_0(x)$ at $x \in \tilde{T}_j$ is rejected if

$$p_j(x) < \frac{k\alpha}{M(\tilde{y}_j;\mathbb{S}^2)} \quad \Longleftrightarrow \quad \tilde{y}_j(x) > \tilde{u}_{\mathrm{BH},j} = F_j^{-1}\left(\frac{k\alpha}{M(\tilde{y}_j;\mathbb{S}^2)}\right), \tag{5.11}$$

where $k\alpha/M(\tilde{y}_j;\mathbb{S}^2)$ is defined as 1 if $M(\tilde{y}_j;\mathbb{S}^2) = 0$. Since $\tilde{u}_{\mathrm{BH},j}$ is random, definition (5.6) is hereby modified to

$$\mathrm{FDR}_{\mathrm{BH},\rho_j} = \mathbb{E}\left\{\frac{V_{\rho_j}(\tilde{u}_{\mathrm{BH},j})}{R_j(\tilde{u}_{\mathrm{BH},j}) \vee 1}\right\}, \tag{5.12}$$

where $R_j(\cdot)$ and $V_{\rho_j}(\cdot)$ are defined in (5.4) and the expectation is taken over all possible realizations of the random threshold $\tilde{u}_{\mathrm{BH},j}$.

Since $\tilde{u}_{\mathrm{BH},j}$ is random, similarly to the definition of $\mathrm{FDR}_{\mathrm{BH},\rho_j}$ (5.12), we define

$$\mathrm{Power}_{\mathrm{BH},\rho_j} = \mathbb{E}\left(\frac{1}{N_j}\sum_{k=1}^{N_j} \mathbb{1}_{\{\tilde{T}_j(\tilde{u}_{\mathrm{BH},j}) \cap D(\xi_k,\rho_j) \neq \emptyset\}}\right). \tag{5.13}$$

6. FDR Control and Power Consistency

6.1. FDR Control

The strategy to prove FDR control is to first quantify the expected number of local maxima above any level u over the null region (false discoveries) and signal region (true discoveries). This is given in Lemmas 6.1 and 6.3 below.

Lemma 6.1. *Let $u \in \mathbb{R}$ be fixed. Then under Condition 2 as $j \to \infty$, the expected number of local maxima of $\tilde{y}_j$ above u in the null region $\mathbb{D}_0^{\rho_j}$ is*

$$\mathbb{E}[M_u(\tilde{y}_j; \mathbb{D}_0^{\rho_j})] = [4\pi - 2\pi(1-\cos\rho_j)N_j]r_j(u) + o(e^{-j^\nu}), \tag{6.1}$$

where

$$r_j(u) = \mathbb{E}[M_u(\tilde{\beta}_j; \mathbb{S}^2)] = F_j(u)r_j, \quad u \in \mathbb{R}, \tag{6.2}$$

is the expected number of local maxima of $\tilde{\beta}_j$ exceeding u over a unit area on $\mathbb{S}^2$, $F_j(u)$ is the tail distribution function (5.10), and

$$r_j = r_j(-\infty) = \frac{1}{4\pi} + \frac{1}{2\pi\eta_j^2\sqrt{3+\eta_j^2}}. \tag{6.3}$$

Proof. Recall $\text{Area}(D(\xi_k, \rho_j)) = 2\pi(1-\cos\rho_j)$ for every k, therefore

$$\text{Area}(\mathbb{D}_0^{\rho_j}) = \text{Area}(\mathbb{S}^2) - \text{Area}(D(\xi_k,\rho_j))N_j = 4\pi - 2\pi(1-\cos\rho_j)N_j.$$

By the Kac-Rice metatheorem, Lemma 4.1 and Condition 3,

$$\begin{aligned}
&\mathbb{E}[M_u(\tilde{y}_j; \mathbb{D}_0^{\rho_j})] \\
&= \int_{\mathbb{D}_0^{\rho_j}} \frac{1}{2\pi\sqrt{\det\text{Cov}(\nabla\tilde{y}_j(x))}} \mathbb{E}[|\det(\nabla^2\tilde{y}_j(x))| \mathbb{1}_{\{\tilde{y}_j(x)>u,\, \nabla^2\tilde{y}_j(x)\prec 0\}} | \nabla\tilde{y}_j(x) = 0] dx \\
&= \int_{\mathbb{D}_0^{\rho_j}} \frac{1}{2\pi\sqrt{\det\text{Cov}(\nabla\tilde{\beta}_j(x))}} \mathbb{E}[|\det(\nabla^2\tilde{\beta}_j(x))| \mathbb{1}_{\{\tilde{\beta}_j(x)>u,\, \nabla^2\tilde{\beta}_j(x)\prec 0\}} | \nabla\tilde{\beta}_j(x) = 0] dx \\
&\quad + O(N_j B^{mj} j^{2\nu} e^{-j^{2\nu}}),
\end{aligned}$$

where m is some positive constant. Evaluating the integral yields (6.1), where

$$r_j(u) = \frac{1}{2\pi C'(\tilde{\beta}_j)} \mathbb{E}[|\det(\nabla^2\tilde{\beta}_j(x))| \mathbb{1}_{\{\tilde{\beta}_j(x)>u,\, \nabla^2\tilde{\beta}_j(x)\prec 0\}} | \nabla\tilde{\beta}_j(x) = 0]$$

is the expected number of local maxima of $\tilde{\beta}_j$ exceeding u over a unit area on $\mathbb{S}^2$. The exact expression (6.2) follows from (5.10), while (6.3) was proved in [9]. □

Remark 6.2 **[Asymptotics of $r_j(u)$.]** By Proposition 4.3, as $j \to \infty$,

$$r_j \sim \frac{c_{s,4}(\gamma)}{8\pi\sqrt{3}c_{s,2}(\gamma)} B^{2j} \quad \text{and} \quad r_j(u) = F_j(u) r_j \sim F_j(u) \frac{c_{s,4}(\gamma)}{8\pi\sqrt{3}c_{s,2}(\gamma)} B^{2j}.$$

Lemma 6.3. *Let $u \in \mathbb{R}$ be fixed. Then under Conditions 2,3 as $j \to \infty$, the number of local maxima of $\tilde{y}_j$ over the signal region $\mathbb{D}_1^{\rho_j}$ satisfies*

$$M_u(\tilde{y}_j; \mathbb{D}_1^{\rho_j}) \geq N_j + O_p(B^{-2j}) \ , \ \mathbb{E}[M_u(\tilde{y}_j; \mathbb{D}_1^{\rho_j})] \geq N_j + O(B^{-2j}).$$

Proof. Let $\tilde{\rho}_j = B^{-j} < \rho_j = j^\nu B^{-j}$. By Lemma 4.1, within the domain $B(\xi_k, \tilde{\rho}_j)$, the mean function of $\tilde{y}_j$ satisfies the assumptions of the unimodal signal model in [8] with the signal strength being $a = B^{2j}$. It then follows from similar arguments as in [8] that

$$M_u(\tilde{y}_j; \mathbb{D}_1^{\tilde{\rho}_j}) = N_j + O_p(B^{-2j})$$
$$\mathbb{E}[M_u(\tilde{y}_j; \mathbb{D}_1^{\tilde{\rho}_j})] = N_j + O(B^{-2j}).$$

The desired results then follow immediately from the observation

$$M_u(\tilde{y}_j; \mathbb{D}_1^{\rho_j}) \geq M_u(\tilde{y}_j; \mathbb{D}_1^{\tilde{\rho}_j}),$$

where the inequality admits the possibility of there being other local maxima in the flatter areas $D(\xi_k, \rho_j) \setminus D(\xi_k, \tilde{\rho}_j)$ of the needlet transform impulse response. The exact expected number of these is presumably small, but hard to estimate. □

Remark 6.4 **[The rate of ρ_j]** The proof of Lemma 6.3 explains why we make the assumption $\rho_j \sim j^\nu B^{-j}$ in Condition 3 above. This choice of rate for ρ_j, decaying slightly less slowly than B^{-j}, allows obtaining an asymptotic limit to the number of local maxima over the null region $\mathbb{E}[M_u(\tilde{y}_j; \mathbb{D}_0^{\rho_j})]$, while the number of local maxima over the signal region $\mathbb{E}[M_u(\tilde{y}_j; \mathbb{D}_1^{\rho_j})]$ can be bounded asymptotically. If we had chosen the rate of $\tilde{\rho}_j$ for ρ_j, decaying at a rate B^{-j}, then as shown in the proof of Lemma 6.3, we could obtain an exact limit for the number of local maxima over the signal region; however in that case, the number of local maxima over the null region would be difficult to quantify due to complicated behavior of the mean function $\tilde{\mu}_j$ (after the needlet tranform) immediately outside that radius.

The following ergodic result, which will be used in the proof of Theorem 6.6 below, shows that the variance of the number of local maxima goes to zero after normalization by the expected value. The result itself is theoretically important and the proof is given in the Supplement.

Theorem 6.5. *Under Conditions 1, 2, 3, as $j \to \infty$,*

$$\mathrm{Var}[M_u(\tilde{\beta}_j; \mathbb{S}^2)] \leq c(u) j^2 B^{2j} + o(j^2 B^{2j}),$$

where the constant $c(u)$ is uniformly bounded with respect to u and the $o(\cdot)$ term is universal.

The following is the first main result of this paper, showing control of FDP and FDR.

Theorem 6.6. *Let Conditions 1, 2, 3 hold.*
(i) Suppose that Algorithm 1 is applied with a fixed threshold u, then

$$\mathrm{FDP}_{\rho_j}(u) \le \frac{[4\pi - 2\pi(1-\cos\rho_j)N_j]r_j(u)}{[4\pi - 2\pi(1-\cos\rho_j)N_j]r_j(u) + N_j}(1+o_p(1)), \tag{6.4}$$

where $r_j(u)$ is defined in (6.2).
(ii) Suppose that Algorithm 1 is applied with the random threshold $\tilde{u}_{\mathrm{BH},j}$ (5.11), then

$$\mathrm{FDR}_{\mathrm{BH},\rho_j} \le \alpha\frac{[4\pi - 2\pi(1-\cos\rho_j)N_j]r_j}{[4\pi - 2\pi(1-\cos\rho_j)N_j]r_j + N_j} + o(1) \le \alpha + o(1), \tag{6.5}$$

where r_j is given by (6.3).

Proof. (i) By Theorem 6.5 and Chebyshev's inequality,

$$\begin{aligned}\mathrm{FDP}_{\rho_j}(u) &= \frac{M_u(\tilde{y}_j;\mathbb{D}_0^{\rho_j})/B^{2j}}{M_u(\tilde{y}_j;\mathbb{D}_0^{\rho_j})/B^{2j} + M_u(\tilde{y}_j;\mathbb{D}_1^{\rho_j})/B^{2j}} \\ &\le \frac{\mathbb{E}[M_u(\tilde{y}_j;\mathbb{D}_0^{\rho_j})]/B^{2j}}{\mathbb{E}[M_u(\tilde{y}_j;\mathbb{D}_0^{\rho_j})]/B^{2j} + M_u(\tilde{y}_j;\mathbb{D}_1^{\rho_j})/B^{2j}}(1+o_p(1)).\end{aligned}$$

It then follows from Lemmas 6.1 and 6.3 that

$$\mathrm{FDP}_{\rho_j}(u) = \frac{[4\pi - 2\pi(1-\cos\rho_j)N_j]r_j(u)}{[4\pi - 2\pi(1-\cos\rho_j)N_j]r_j(u) + N_j}(1+o_p(1)).$$

(ii) Following a similar argument to that in [8], we use the fact that $\tilde{u}_{\mathrm{BH},j}$ is the smallest u satisfying $\alpha\tilde{G}_j(u) \ge F_j(u)$, where

$$\tilde{G}_j(u) = \frac{M_u(\tilde{y}_j;\mathbb{D}_0^{\rho_j}) + M_u(\tilde{y}_j;\mathbb{D}_1^{\rho_j})}{M(\tilde{y}_j;\mathbb{D}_0^{\rho_j}) + M(\tilde{y}_j;\mathbb{D}_1^{\rho_j})}$$

and $F_j(u)$ is the height distribution (5.10) of $\tilde{\beta}_j$. Notice that

$$F_j(u) = \frac{\mathbb{E}[M_u(\tilde{y}_j;\mathbb{D}_0^{\rho_j})]}{\mathbb{E}[M(\tilde{y}_j;\mathbb{D}_0^{\rho_j})]}.$$

Similarly to the proof of part (i), we have

$$\begin{aligned}\tilde{G}_j(u) &\ge \frac{\mathbb{E}[M(\tilde{y}_j;\mathbb{D}_0^{\rho_j})]F_j(u) + N_j}{\mathbb{E}[M(\tilde{y}_j;\mathbb{D}_0^{\rho_j})] + N_j} + o_p(1) \\ &= \frac{[4\pi - 2\pi(1-\cos\rho_j)N_j]r_jF_j(u) + N_j}{[4\pi - 2\pi(1-\cos\rho_j)N_j]r_j + N_j} + o_p(1).\end{aligned}$$

Solving the equation

$$\alpha\frac{[4\pi-2\pi(1-\cos\rho_j)N_j]r_jF_j(u)+N_j}{[4\pi-2\pi(1-\cos\rho_j)N_j]r_j+N_j}+o_p(1)=F_j(u)$$

gives an asymptotic solution

$$\tilde{u}^*_{\mathrm{BH},j}=F_j^{-1}\left(\frac{\alpha N_j}{N_j+(1-\alpha)[4\pi-2\pi(1-\cos\rho_j)N_j]r_j}\right)+o_p(1). \tag{6.6}$$

Since $\tilde{u}_{\mathrm{BH},j}\le\tilde{u}^*_{\mathrm{BH},j}$ almost surely, we have

$$\begin{aligned}\mathrm{FDR}_{\mathrm{BH},j}&=\mathbb{E}\left[\frac{V_{\rho_j}(\tilde{u}_{\mathrm{BH},j})}{V_{\rho_j}(\tilde{u}_{\mathrm{BH},j})+W_{\rho_j}(\tilde{u}_{\mathrm{BH},j})}\right]\le\mathbb{E}\left[\frac{V_{\rho_j}(\tilde{u}^*_{\mathrm{BH},j})}{V_{\rho_j}(\tilde{u}^*_{\mathrm{BH},j})+W_{\rho_j}(\tilde{u}^*_{\mathrm{BH},j})}\right]\\&\le\frac{\mathbb{E}[V_{\rho_j}(\tilde{u}^*_{\mathrm{BH},j})]}{\mathbb{E}[V_{\rho_j}(\tilde{u}^*_{\mathrm{BH},j})]+\mathbb{E}[W_{\rho_j}(\tilde{u}^*_{\mathrm{BH},j})]}(1+o(1))\le\frac{\mathbb{E}[M(\tilde{y}_j;\mathbb{D}_0^{\rho_j})]F_j(\tilde{u}^*_{\mathrm{BH},j})}{\mathbb{E}[M(\tilde{y}_j;\mathbb{D}_0^{\rho_j})]F_j(\tilde{u}^*_{\mathrm{BH},j})+N_j}(1+o(1))\\&=\alpha\frac{[4\pi-2\pi(1-\cos\rho_j)N_j]r_j}{[4\pi-2\pi(1-\cos\rho_j)N_j]r_j+N_j}+o(1).\end{aligned}$$

□

Remark 6.7 [Threshold for FDP.] To make the bound for FDP asymptotically equal to a significance level α, the corresponding threshold u must satisfy the equation

$$\frac{[4\pi-2\pi(1-\cos\rho_j)N_j]r_j(u)}{[4\pi-2\pi(1-\cos\rho_j)N_j]r_j(u)+N_j}=\alpha,$$

implying

$$r_j(u)=\frac{\alpha N_j}{(1-\alpha)[4\pi-2\pi(1-\cos\rho_j)N_j]}. \tag{6.7}$$

By Remark 6.2,

$$r_j(u)\sim F_j(u)\frac{c_{p,4}(\gamma)}{8\pi\sqrt{3}c_{s,2}(\gamma)}B^{2j},$$

implying that as $j\to\infty$ and $u\to\infty$, $\log(r_j(u))\sim\log(B^{2j}e^{-u^2/2})$. Solving the equation

$$B^{2j}e^{-u^2/2}=\frac{\alpha N_j}{(1-\alpha)[4\pi-2\pi(1-\cos\rho_j)N_j]}$$

yields the approximate solution

$$u\sim\sqrt{2\log(B^{2j}/N_j)}. \tag{6.8}$$

According to Condition 3, $N_j\rho_j^2=O(1)$, implying $N_j=O(\rho_j^{-2})=O(j^{-2\nu}B^{2j})$. Therefore, $u\ge2\sqrt{\nu\log(j)}\to\infty$.

Remark 6.8 [Comparison between FDP and BH Procedure.] Dividing both sides of (6.7) by r_j yields

$$F_j(u) = \frac{\alpha N_j}{(1-\alpha)[4\pi - 2\pi(1-\cos\rho_j)N_j]r_j},$$

implying the following threshold by FDP for controlling significance level α:

$$u_\alpha = F_j^{-1}\left(\frac{\alpha N_j}{(1-\alpha)[4\pi - 2\pi(1-\cos\rho_j)N_j]r_j}\right).$$

In comparison, for controlling significance level α by the BH procedure, the asymptotic threshold is given by (6.6). If we replace α by

$$\tilde{\alpha} = \alpha\frac{[4\pi - 2\pi(1-\cos\rho_j)N_j]r_j}{[4\pi - 2\pi(1-\cos\rho_j)N_j]r_j + N_j},$$

then the FDP threshold at significance level $\tilde{\alpha}$ is given by

$$u_{\tilde{\alpha}} = F_j^{-1}\left(\frac{\alpha N_j}{N_j + (1-\alpha)[4\pi - 2\pi(1-\cos\rho_j)N_j]r_j}\right).$$

This coincides with the asymptotic threshold $\tilde{u}^*_{\mathrm{BH},j}$ (6.6) by the BH procedure. Since $r_j = O(B^{2j})$ and $N_j = O(\rho_j^{-2}) = O(j^{-2\nu}B^{2j})$, we see that the upper bound in (6.5) tends to α in the limit of high-frequency.

Remark 6.9 [Comparison with FWER control (the expected Euler characteristic thresholding).] For high values of the threshold u, the expected Euler characteristic exceeding u can be approximated by $r_j(u)$; hence, the threshold for controlling the FWER can be obtained by solving the equation $r_j(u) = \alpha$. By the discussion in the previous remark, this equation becomes $B^{2j}e^{-u^2/2} = \alpha$, which gives the solution $u \sim \sqrt{2\log(B^{2j})} = 2\sqrt{j\log(B)}$. In comparison, the FDR threshold increases at rate $2\sqrt{\nu\log(j)}$, which is much slower.

6.2. Power Consistency

To prove power consistency, we first show that, asymptotically, there will be at least one local maximum of $\tilde{y}_j$ within a small ball centered at every point source.

Lemma 6.10. *Under Conditions 1, 2, 3, for each fixed k, there exists $c > 0$ such that for sufficiently large j,*

$$\mathbb{P}\left(\#\{x \in \tilde{T}_j(u) \cap D(\xi_k, \tilde{\rho}_j)\} \geq 1\right) \geq 1 - \exp\left(-cB^{j(\gamma-2)/2}\right),$$

where $\tilde{\rho}_j = B^{-j}$ and $u = \sqrt{2\log(B^{2j}/N_j)}$ is the asymptotic BH threshold (6.8).

Proof. For each k, the probability $\tilde{y}_j(x)$ has at least one local maximum above u in $D(\xi_k, \tilde{\rho}_j)$ is the complement of the probability that: (1) $\tilde{y}_j(x)$ has no local maxima in $D(\xi_k, \tilde{\rho}_j)$, or (2) $\tilde{y}_j(x)$ is below u everywhere in $D(\xi_k, \tilde{\rho}_j)$.

For (1), this is less than the probability that there exists some $x \in D(\xi_k, \rho_1)$ such that $\langle \nabla \tilde{y}_j(x), \xi_k - x \rangle \leq 0$, since all $x \in \partial D(\xi_k, \tilde{\rho}_j)$ satisfying $\langle \nabla \tilde{y}_j(x), \xi_k - x \rangle > 0$ would imply the existence of at least one local maximum in $D(\xi_k, \tilde{\rho}_j)$. This probability is bounded above by

$$\begin{aligned}
&\mathbb{P}\left(\inf_{\partial D(\xi_k, \tilde{\rho}_j)} \langle \nabla \tilde{\beta}_j(x), \xi_k - x \rangle \leq - \inf_{\partial D(\xi_k, \tilde{\rho}_j)} \langle \nabla \tilde{\mu}_j(x), \xi_k - x \rangle \right) \\
&= \mathbb{P}\left(\sup_{\partial D(\xi_k, \tilde{\rho}_j)} -\left\langle \nabla \tilde{\beta}_j(x), \frac{\xi_k - x}{\|\xi_k - x\|} \right\rangle \geq \inf_{\partial D(\xi_k, \tilde{\rho}_j)} \left\langle \nabla \tilde{\mu}_j(x), \frac{\xi_k - x}{\|\xi_k - x\|} \right\rangle \right) \\
&\leq \mathbb{P}\left(\sup_{x \in D_k^{\tilde{\rho}_j}} \sup_{\|\tau\|=1} \langle \nabla \tilde{\beta}_j(x), \tau \rangle \geq c_1 B^{2j + j(\gamma-2)/2} \right),
\end{aligned}$$

where c_1 is a positive constant and the last inequality is due to the result proved in the Supplement that $\mathrm{Var}(\beta_j) \sim c_{s,0}(\gamma)/B^{j(\gamma-2)}$, together with (4.3), Lemma 4.1 and the fact that $\partial D(\xi_k, \tilde{\rho}_j)$ is contained in the closure of $D(\xi_k, \tilde{\rho}_j)$. By Proposition 4.3, there exists $c_2 > 0$ such that for sufficiently large j

$$\sup_{x \in D(\xi_k, \tilde{\rho}_j)} \sup_{\|\tau\|=1} \mathrm{Var}(\langle \nabla \tilde{\beta}_j(x), \tau \rangle) \leq c_2 B^{2j}.$$

Then by the Borell-TIS inequality, there exists $c_3 > 0$ such that for sufficiently large j,

$$\mathbb{P}\left(\#\{x \in \tilde{T}_j \cap D(\xi_k, \tilde{\rho}_j)\} = 0 \right) \leq \exp\left(-c_3 B^{j(\gamma-2)/2} \right).$$

On the other hand, for (2), the probability that $\tilde{y}_j(x)$ is below u everywhere in $D(\xi_k, \tilde{\rho}_j)$ is bounded above by $1 - \Phi(|u - B^{2j + j(\gamma-2)/2}|)$. The desired result then follows from the observation

$$\mathbb{P}\left(\#\{x \in \tilde{T}_j(u) \cap D(\xi_k, \tilde{\rho}_j)\} \geq 1 \right) \geq 1 - \exp\left(-c_3 B^{j(\gamma-2)/2} \right) - \left(1 - \Phi(|u - B^{2j + j(\gamma-2)/2}|) \right),$$

where the last term in parentheses is much smaller than the second when $u = \sqrt{2 \log(B^{2j}/N_j)}$. □

Following is the second main result of this paper, showing that the detection power tends to one asymptotically.

Theorem 6.11. *Let Conditions 1, 2, 3 hold.*

(i) Suppose that Algorithm 1 is applied with a fixed threshold u, then $\mathrm{Power}_{\rho_j}(u) \to 1$.

(ii) Suppose that Algorithm 1 is applied with the random threshold $\tilde{u}_{\mathrm{BH}}$ (5.11), then $\mathrm{Power}_{\mathrm{BH}, \rho_j} \to 1$.

Proof. The desired results follow directly from Lemma 6.10 and the definitions of power (5.7) and (5.13). □

7. Numerical Validation

In this section we present numerical evidence on the performance of the algorithm advocated in this work. One crucial step in the STEM algorithm (Algorithm 1) is the computation of p-values of detected peaks, based on the distribution of peak heights under the null hypothesis, i.e., no point masses at all. We therefore start our validation by comparing the analytical peak height distribution function given in (5.10) with the empirical result from filtered noise Monte Carlo simulations.

Once we establish the validity of the peak height distribution on the noise field, we add simulated point sources to form the full signal-plus-noise Monte Carlo simulations. These simulations are used to evaluate the numerical performance of the asymptotic FDP approximation and FDR control of Section 6. For all the results we report, the Mexican needlet parameters used are $B = 1.2$ and $s = 1$.

7.1. Simulation of the CMB noise field

All our maps and the corresponding spherical harmonic coefficients are generated using the *HEALpix* package, which is now the standard routine software for handling cosmological data: see [15] for a detailed discussion on this package and its main features. In HEALpix one can use the *create_alm* routine to generate random spherical harmonic coefficients, $a_{\ell m}$, with a given power spectrum. The code *alm2map* takes these coefficients and generate a pixelized Gaussian map; the inverse process is implemented using the *map2alm* code. To decompose a map into Mexican needlet components, we filter the $a_{\ell m}$ coefficients by the Mexican needlet window functions as given in (4.4).

A single HEALpix pixel has an area of $4\pi/N_{\text{pix}}$ where $N_{\text{pix}} = 12N_{\text{side}}^2$ is the total number of pixels on a given map. The resolution is specified by the N_{side} parameter, which is a multiple of 2.

As it is the case in typical experimental circumstances, we assume that known astrophysical contaminants (for instance, the Milky Way) have been preliminarily removed from the data. To simulate our noise field, we then generated 100 Gaussian realization maps of the CMB sky starting from the Planck CMB power spectrum. We refer to Figure 11 on page 21 of [24], which provides a plot showing the theoretical prediction on the C_ℓ, and their estimated values on Planck CMB data; the errors bars at high ℓ become so narrow that they do not show up in the figure, being actually thinner than the thickness of the theoretical line prediction. This plot shows both that very efficient estimates for the power spectrum exist, and that theoretical models in this area should be considered very reliable.

All maps are simulated with a pixel resolution of $N_{\text{side}} = 1024$. The standard deviation, also called root mean square (RMS), of the simulated noise field is given by

$$\sigma^2_{noise} = \sigma^2_{cmb} = \sum_{\ell} \frac{(2\ell + 1)C_\ell}{4\pi}, \tag{7.1}$$

where C_ℓ is the Planck CMB power spectrum [25].

To simulate the finite resolution of the measuring instrument, these maps are then smoothed by a Gaussian filter with full-width half max (FWHM) of 10 arcmin. In the literature, this is usually referred to as a 10 arcmin Gaussian beam. Its effect can be thought of as part of the noise autocovariance function, although it is essentially negligible at $\ell \sim 1000$ as in our needlet analysis.

7.2. Simulation of point sources

As mentioned above, a point source in the sky is observed by a detector which has a finite angular resolution. With some abuse of nomenclature, the opening angle of the smallest resolvable angular unit, t_N, is called the beam of the detector. The typical angular size of galaxies is of a few arcsecs (i.e., one degree divided by 60^2) while the detector beam sizes for typical CMB experiments (10 arcmin) are an order of magnitude larger. This means that galaxies and other objects with angular size smaller than the beam can be viewed as point sources. As argued in the previous sections, the convolution of the point sources by the detector beam yields a Gaussian bell-like profile in the final map with the peak of the Gaussian being at the location of the point sources, ξ_k. The signal part of our simulations is hence given by (2.4) above, where the coefficient a_k represent the brightness of the kth point source and N is their total number.

It would be possible to consider more realistic models for these point sources, for instance using the so-called Planck sky model (see for instance [10]). However, this would require a rather lengthy technical discussion on some specific astrophysical and experimental settings, which would not add anything substantial to the understanding of our current algorithm, nor would alter significantly our numerical results. We therefore delay a more complete analysis of these practical issues to a future, more applied paper.

To simulate our signal model with N point sources, we first generated N coordinate points randomly with a uniform probability density over the sphere. Second we found the pixels that correspond to these locations on the HEALpix map; third we set the values a_k of these pixels as draws from a uniform distribution in the range 0 and $A_{\max}$. These amplitudes are given as a multiple of the RMS of the noise (7.1). Finally, to simulate the instrumental resolution, we convolved the map obtained in the last step by means of a Gaussian beam of FWHM = 10 arcmin. This final map is now a pixelized version of (2.3). Notice that for clarity we have described the smoothing process as a separate operation in the noise and signal maps, but this is, of course, equivalent to doing a single smoothing operation on a signal plus noise map.

The Gaussian beam decreases the sources magnitude by a factor proportional to the ratio between the area of a pixel and the area covered by the detector beam. For our choice of the beam and the pixel resolution, this factor is an order of magnitude. Since we desired to generate point sources uniformly distributed between 0 and $3\sigma_{noise}$ after smoothing, we used $A_{max} = 30\sigma_{cmb}$. We considered different values for the total number of sources, i.e., N=1000, 3000, 5000.

The final signal-plus-noise Monte Carlo simulations are then obtained by adding the point sources map to the 100 noise simulations; an example is provided in Figure 1. Note

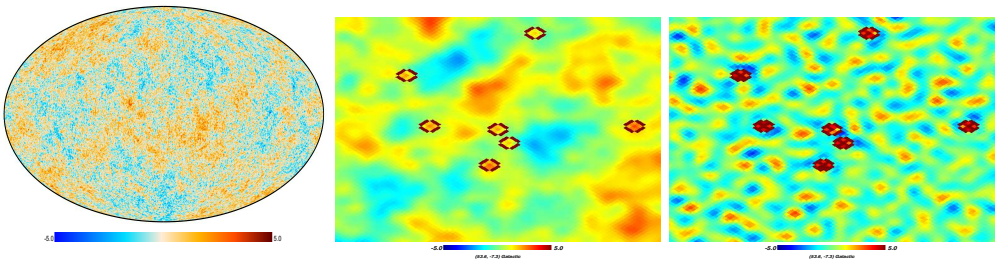

Figure 1. **Signal plus noise maps:** Upper panel is an equal area stereographic projection (the so called Mollweide projection) of the signal plus noise simulation before needlet filtering; the color map is given in standardized RMS units. Bottom left panel is a gnomonic projection of the unfiltered map around a point source with 5 degree diameter; bottom right panel is a similar gnomonic projection around the same point source but from a Mexican needlet filtered map. The red marks have been added only to visualize the location of the point sources but are not part of the simulation.

that the point sources are weak and hard to find without statistical analysis.

7.3. Distribution of peak heights

The theoretical distribution of local maxima (peaks) on a Mexican needlet filtered Gaussian map is given by (5.10). In Figure 2 we present the comparison of the theoretical density, $h_j(x)$, with what we obtained empirically using 100 Gaussian map simulations with no point sources. The upper panels from left to right respectively present the normalized Gaussian peak PDFs for needlet frequency $j = 31, 34$, with Mexican needlet parameters $B = 1.2$ and $p = 1$, while in Figure 1 we had taken $B = 1.2$ and $j = 38$. The choice of the parameters B and j needs some discussion. In short, it is known (see again [28, 27]) that the Mexican needlet filter at scale j is centered around the multipole $\ell = B^j$ and has most of its power concentrated in the multipoles $[B^{j-1}, B^{j+1}]$. In practical experimental circumstances, the multipoles to be covered will be fixed mainly by the technical characteristics of the device collecting the data. For instance, it is known that the antennae in the satellite Planck can reach a resolution of about 7 arcminutes, corresponding roughly to $\ell = 1800$; however, a detailed technical analysis (which is discussed in the Planck papers, see for instance [24]) has shown that the highest multipoles are affected by instrumental noise which can make the corresponding data unreliable. Focussing for instance on $B = 1.2, j = 38$ means that we will be concentrating the statistical analysis on multipoles in the order of 1021 ± 200; in this range, instrumental noise is known to be basically negligible. This form of knowledge on the technical features of the experiment should be known in practice to any applied scientist which is investigating a specific data set (not necessarily in astrophysics); the range of values that we reported are also meant to illustrate the convergence result as j grows larger and larger.

In the lower panel of Figure 2 we show the relative percentage difference between the

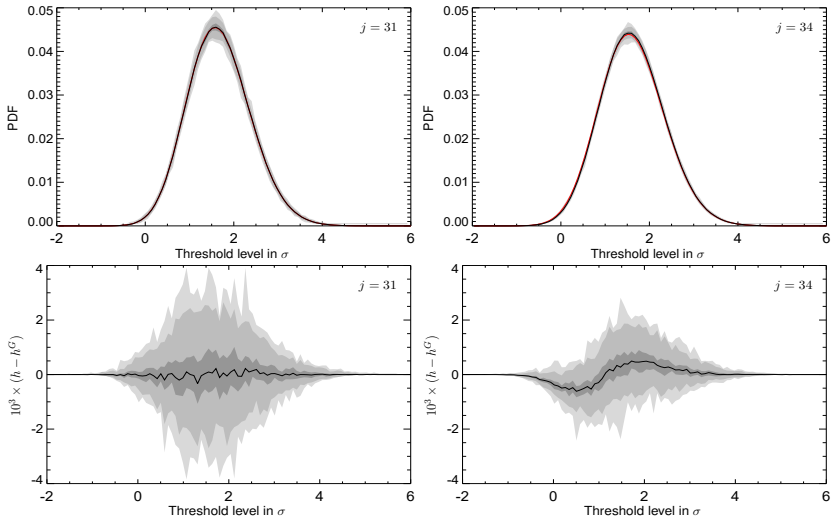

Figure 2. **Peaks PDF:** probability density of height of local maxima. In the upper panels, the red curves represent the analytical values, while the black curves and the gray contours are the mean and the 68,95,99 % percentiles from the simulations from 100 Monte Carlo simulations with no point sources. The lower panels show the difference between the analytical (h^G) and numerical (h) result. Here $j = 31, 34$, which corresponds to central multipoles of $\ell = [284, 492]$.

analytical and simulation results. It is easy to see from these figures that the theory fits the numerical results remarkably well. Moreover, the dispersion around the expected value of the PDF decreases as j increases, consistently with the ergodicity result of Theorem 6.5.

7.4. Application of the STEM algorithm

The first step in the STEM algorithm, after needlet filtering, is to normalize the map using its standard deviation, as defined in (4.5), to obtain (4.6).

To find local peaks on a map we compute the first and second derivatives using HEALpix 's routine *alm2map_der*. The pixels where the first derivative is close to zero (within a precision of 10^{-6}) are classified as the local extrema. We then partition these extrema into maxima, minima and saddle using the eigenvalue decomposition of the Hessian matrix - of course, maxima are those with all the eigenvalues negative.

It is instructive to look at how the brightness of point sources increase as we filter the signal plus noise map with Mexican needlets. In Figure 3, we plot the PDF of point source amplitudes before adding noise (grey curve), after adding noise but before needlet filtering (thick black curve), and after filtering with increasing j. For the high frequency Mexican needlet we considered, $j = 38$, filtering increases the brightness by a factor greater than

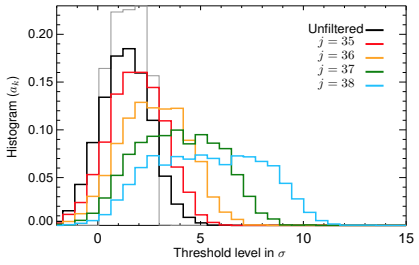

Figure 3. **Needlet filtering increases signal-to-noise ratio:** Histogram of signal amplitudes at the location of the point sources, before adding noise (grey curve), after adding noise but before filtering (thick black curve), and after filtering with the Mexican needlet at different j.

4. The negative values in the histogram are due to the added Gaussian noise; we do not expect to detect such weak sources based on their amplitude information only.

7.5. False Discovery Proportion (FDP)

In (6.5) of Section 6 we provided the analytical result on the upper bound of the FDP as a function of the power spectrum of the noise, the total number and the spatial profile of the sources. Here we compare this result with what is obtained from numerical simulations.

The empirical FDP is computed using the following steps: locate maxima on needlet filtered signal-plus-noise Monte Carlo simulations using our peak detection code; classify peaks as *True discovery* if the location of a maxima corresponds to a known (input) point source within ρ pixel radius or *False discovery* if there are no input sources within ρ pixels radius of the peak (ρ corresponding to the tolerance parameter); count the number of True and False discoveries as a function of ρ and the RMS of the noise.

The empirical FDP as a function of u, which is in units of the RMS of the noise, and the source detection tolerance parameter ρ is computed, according to (5.5), as

$$\widehat{\mathrm{FDP}}_{\rho_j}(u) = \frac{\#\text{ of False discoveries above } u}{\text{total }\#\text{ of peaks above } u} \tag{7.2}$$

In Figure 4 we illustrate the comparison of the FDP for thresholds in the filtered map above 3σ and $\rho = 3$ for different values of j. We found that setting $2 \leq \rho \leq 8$ does not alter significantly our results (note that the smallest practical radius is $\rho = 2$). The red curve in these plots corresponds to the first term on the right hand side of (6.5), while the black curve is from the mean of the simulations. The contours from dark to light gray corresponds to the $68, 95$ and 99% confidence intervals. We note that, as expected, the analytic results for the upper bound become larger than the numerical simulations as j increases.

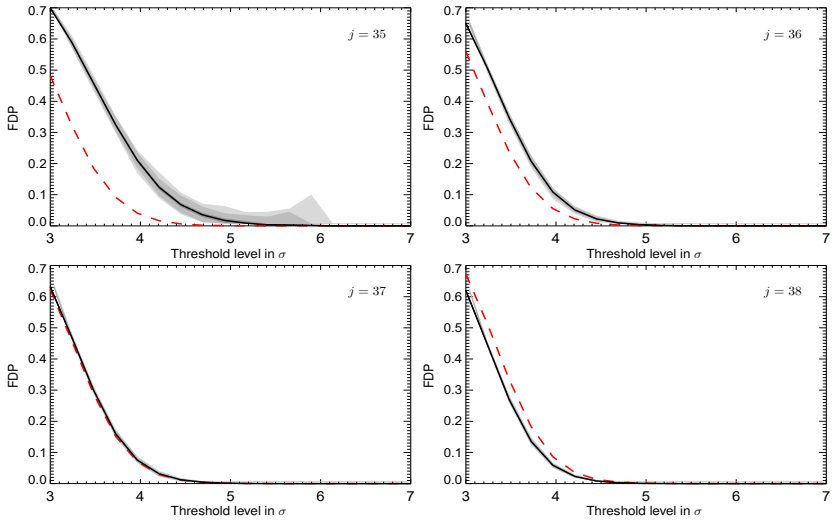

Figure 4. **False Discovery Proportion:** FDP as a function of threshold (in units of standard deviation) for different needlet scales. The dashed red curves are for the analytical upper bounds while the black curves are for the mean of the empirical FDPs from 100 Monte Carlo simulations. The gray shades are for percentiles 68, 95 and 99%.

7.6. False Discovery Rate (FDR)

We now proceed in validating the analytical formalism established in Section 6 to control the false discovery rate (FDR). This is done by comparing the analytical upper bound of the FDR, which is given by (6.5), with the empirical result from simulations. In Figure 5, it is shown that for a given error rate, the empirical FDR is always below the upper limit set by the theory; in Figure 6 we present the mean FDP and FDR curves together with the corresponding theoretical results for different number of input sources. Again, the FDP and FDR are bounded above by the theoretical bounds.

7.7. Detection power

To quantify how many of the input point sources we discovered in our analysis, in Figure 7 we show the number of peaks that matches the true sources, i.e. the numerator of (5.7), which measures the statistical power of the algorithm. These results show that the power of the STEM algorithm is almost 100% in detecting bright sources - indeed, we have detected all input sources whose brightness was above 1σ in the unfiltered simulated maps.

Overall, we believe that the results in this section provide a strong numerical support for the asymptotic findings that we described earlier in this paper. Applications to real

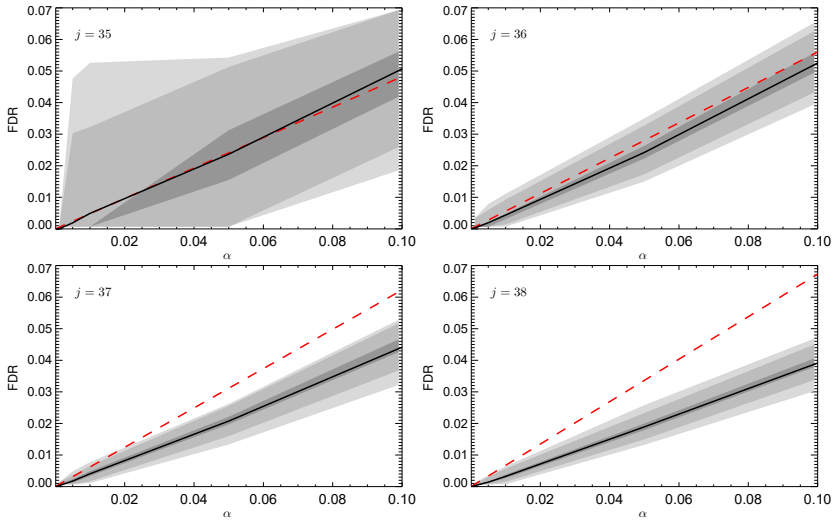

Figure 5. **False Discovery Rate:** FDR as a function of error rate α for different needlet scales. The red curve is a plot of $\alpha * FDP(u = 3)$, while the black curve is the mean from 100 Monte Carlo simulations. The gray shades are for percentiles $68, 95$ and 99%. The number of point sources is 5000.

CMB data from the Planck satellite are currently being developed and will appear in a follow-up paper.

Acknowledgements

Research partially supported by the ERC Grant no. 277742 *Pascal* (PI Domenico Marinucci), NIH grant R01-CA157528 (PI Armin Schwartzman), NSF grant DMS 1811659 (PI Armin Schwartzman), and NSF grant DMS 1811632 (PI Dan Cheng).

References

[1] Adler, R. J. and Taylor, J. E. (2007), *Random Fields and Geometry*, Springer.

[2] Baldi, P., Kerkyacharian, G., Marinucci, D. and Picard, D. (2009a), Asymptotics for Spherical Needlets, *Annals of Statistics*, **37**, 1150-1171.

[3] Baldi, P., Kerkyacharian, G., Marinucci, D. and Picard, D. (2009b), Subsampling needlet coefficients on the sphere, *Bernoulli*, **15**, 438–463.

[4] Benjamini, Y. and Hochberg, Y. (1995), Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society Series B*, **57**, 289–300.

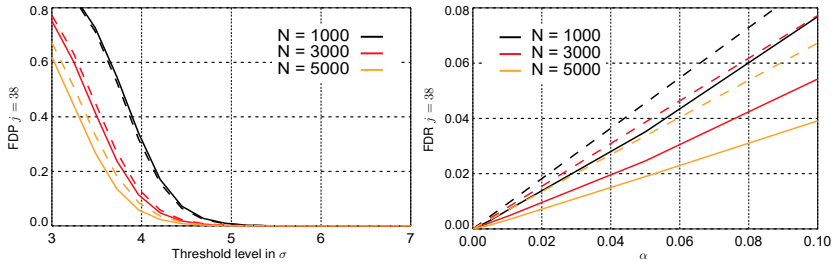

Figure 6. **Theoretical vs numerical results for FDP and FDR:** FDP as a function of thresholds and FDR as a function of global P-value, α, for needlet scale $j = 38$. The dashed curves are the analytical upper bounds while the solid curves are the corresponding empirical results from 100 Monte Carlo simulations.

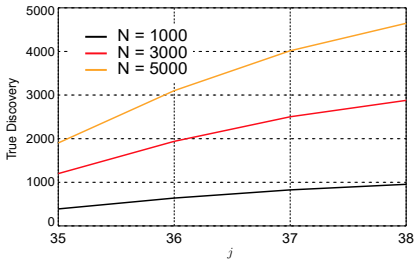

Figure 7. **Detection power:** number of true discoveries for threshold $u > 3$ as a function of needlet scales, j. The three curves are the mean of 100 Monte Carlo simulations for the corresponding cases. The legend shows the number of input point sources in simulations.

[5] Bobin, J., Sureau, F., Starck, J.-L., Rassat, A. and Paykari, P. (2014), Joint Planck and WMAP CMB Map Reconstruction, *Astronomy and Astrophysics*, **563**, A105, arXiv:1401.6016.

[6] Cammarota, V., Marinucci, D. and Wigman, I. (2016), On the distribution of the critical values of random spherical harmonics, *Journal of Geometric Analysis*, **26**, 3252–3324.

[7] Cheng, D. and Schwartzman, A. (2015), Distribution of the height of local maxima of Gaussian random fields. *Extremes*, **18**, 213–240.

[8] Cheng, D. and Schwartzman, A. (2017), Multiple testing of local maxima for detection of peaks in random fields. *Annals of Statistics*, **45**(2), 529–556.

[9] Cheng, D. and Schwartzman, A. (2018), On the Explicit Height Distribution and Expected Number of Local Maxima of Isotropic Gaussian Random Fields, *Bernoulli*, **4B**, 3422-3446.

[10] Delabrouille, J. et. al. (2013), The pre-launch Planck Sky Model: a model of sky emission at submillimetre to centimetre wavelengths, *Astronomy and Astrophysics*, **553**, A96.

[11] Durastanti, C. (2017), Tail Behaviour of Mexican Needlets, *Journal of Mathematical Analysis and its Applications*, **447**, no. 2, 716-735.

[12] Durrer, R. (2008), *The Cosmic Microwave Background*, Cambridge University Press.

[13] Geller, D. and Mayeli, A. (2009a), Continuous Wavelets on Compact Manifolds, *Mathematische Zeitschrift,* **262**, 895–927, arXiv: 0811.4440.

[14] Geller, D. and Mayeli, A. (2009b), Nearly Tight Frames and Space-Frequency Analysis on Compact Manifolds, *Mathematische Zeitschrift,* **263**, 235–264, arXiv: 0706.3642.

[15] Górski, K.M., Hivon, E., Banday, A.J., Wandelt, B.D., Hansen, F.K., Reinecke, M. and Bartelmann, M. (2005), HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere, *Astrophysical Journal*, **699**, 759–771.

[16] Kato, T. (1995), *Perturbation theory for linear operators*, Classics in Mathematics, Springer-Verlag.

[17] Lan, X. and Marinucci, D. (2009), On the dependence structure of wavelet coefficients for spherical random fields, *Stochastic Processes and their Applications* **119**, 3749–3766.

[18] Loh, W.-L. (2005), Fixed-Domain Asymptotics for a Subclass of Matérn-type Gaussian Random Fields, *Annals of Statistics*, **33**, 2344–2394.

[19] Marinucci, D. and Peccati, G. (2011), *Random Fields on the Sphere. Representation, Limit Theorem and Cosmological Applications*, Cambridge University Press

[20] Marinucci, D. and Peccati, G. (2013), Mean Square Continuity on Homogeneous Spaces of Compact Groups, *Electronic Communications in Probability*, **18**, no. 37, 1–10.

[21] Marinucci, D., Pietrobon, D., Balbi, A., Baldi, P., Cabella, P., Kerkyacharian, G., Natoli, P., Picard, D. and Vittorio, N., (2008), Spherical Needlets for CMB Data Analysis, *Monthly Notices of the Royal Astronomical Society*, **383**, 539–545.

[22] Mayeli, A. (2010), Asymptotic Uncorrelation for Mexican Needlets, *Journal of Mathematical Analysis and its Applications*, **363**, 336–344.

[23] Narcowich, F. J., Petrushev, P. and Ward, J.D. (2006a), Localized Tight Frames on Spheres, *SIAM Journal of Mathematical Analysis* , **38**, 574–594.

[24] Planck Collaboration (2016) Planck 2015 results. I. Overview of Products and Scientific Results, *Astronomy and Astrophysics*, **594**, A1, http://xxx.lanl.gov/abs/1507.02704.

[25] Planck Collaboration (2016) Planck 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters, *Astronomy and Astrophysics*, **594**, A11, http://xxx.lanl.gov/abs/1507.02704.

[26] Planck Collaboration (2016) Planck 2015. XXVI. The second Planck catalogue of compact sources, *Astronomy and Astrophysics*, **594**, A26.

[27] Scodeller, S., Rudjord, O. Hansen, F.K., Marinucci, D., Geller, D. and Mayeli, A. (2011), Introducing Mexican needlets for CMB analysis: Issues for practical applications and comparison with standard needlets, *Astrophysical Journal*, **733**, 121.

[28] Scodeller, S., Hansen, F.K. and Marinucci, D. (2012), Detection of new point sources in WMAP 7 year data using internal templates and needlets, *Astrophysical Journal*, **753**, 27, arXiv:1201.5852.

[29] Schwartzman, A., Gavrilov, Y. and Adler, R. J. (2011), Multiple testing of local maxima for detection of peaks in 1D, *Annals of Statistics*, **39**, 3290–3319.