Published in final edited form as:

J Am Stat Assoc. 2013 July 1; 108(502): 456–468. doi:10.1080/01621459.2013.779829.

Straight to the Source: Detecting Aggregate Objects in Astronomical Images with Proper Error Control

David A. Friedenberga [Statistician III] and

Battelle Memorial Institute, 505 King Ave., Columbus, OH 43214

Christopher R. Genovese [Professor]

Department of Statistics, Carnegie Mellon University, Pittsburgh PA 15213

David A. Friedenberga: friedenbergd@battelle.org; Christopher R. Genovese: genovese@stat.cmu.edu

Abstract

The next generation of telescopes, coming on-line in the next decade, will acquire terabytes of image data each night. Collectively, these large images will contain billions of interesting objects, which astronomers call *sources*. One critical task for astronomers is to construct from the image data a detailed *source catalog* that gives the sky coordinates and other properties of all detected sources. The source catalog is the primary data product produced by most telescopes and serves as an important input for studies that build and test new astrophysical theories. To construct an accurate catalog, the sources must first be detected in the image. A variety of effective source detection algorithms exist in the astronomical literature, but few if any provide rigorous statistical control of error rates. A variety of multiple testing procedures exist in the statistical literature that can provide rigorous error control over pixelwise errors, but these do not provide control over errors at the level of sources, which is what astronomers need. In this paper, we propose a technique that is effective at source detection while providing rigorous control on source-wise error rates. We demonstrate our approach with data from the Chandra X-ray Observatory Satellite. Our method is competitive with existing astronomical methods, even finding two new sources that were missed by previous studies, while providing stronger performance guarantees and without requiring costly follow up studies that are commonly required with current techniques.

Keywords

Blind Source Detection; Multiple Testing; Astrostatistics

1 Introduction

The typical astronomical image records the intensity of light over some range of frequencies across a section of sky that contains many celestial objects (called *sources*) of various size, shape, and luminosity. Each pixel essentially counts how many photons have struck the corresponding detector during the exposure, but the counts give a distorted image of the astronomical scene. Thermal noise and background emissions (collectively called *background*) corrupt the data and obscure the sources' signals. Diffraction and atmospheric effects blur the image, reducing resolution and washing out the fainter sources. And instrument artifacts can produce a variety of systematic errors and missing data. The goal of the astronomical source detection problem is to identify and characterize the sources within such an image and to produce a *source catalog* mapping each source to its coordinates and other identifiable properties.

A source catalog is the main data product of most astronomical surveys and the basic input to downstream scientific studies. This has been true for some time. Early catalogs – from the data of ancient astronomers Shi Shen and Hipparchus, each listing about 1000 stars, to the compendium of deep-sky objects produced by William and Caroline Herschel in the 1700s (Herschel, 1786) – were based on direct visual observations. Later work, especially in the 20th century, used photographic plates, both improving resolution and allowing the detection of much fainter objects. But either way, compiling a source catalog has been a slow and painstaking affair, often requiring years to collect data on only a handful of objects. Until recently, catalogs comprising a few hundred objects were considered large, a few thousand epic.

All this changed with the advent of new technologies – digital imaging, advanced designs for telescope mirrors, and computer automation – and with increases in available computing power and storage. With relative suddenness, astronomers found that they could observe wider, deeper, and faster than ever before. For instance, the Sloan Digital Sky Survey (York et al., 2000) has measured hundreds of millions of objects. The upcoming Large Synoptic Survey Telescope (LSST, Tyson and the LSST Collaboration, 2002) will scan the entire sky every few days, collecting several terabytes of data per night into a catalog comprising *billions* of objects. Over the past two decades, astronomy has gone from data poor to data rich.

The sheer size of these source catalogs raises both practical and statistical challenges. A practical challenge is creating a catalog whose *impurity*, the proportion of detections that are not real objects, is sufficiently small. Every catalog, even if derived from manual observations, is likely to contain errors: sources that are missed, located incorrectly, misclassified, or created spuriously. An impure catalog complicates and distorts downstream scientific analyses through the presence of false or misclassified objects. To mitigate such errors, it is common for follow-up observations to verify the sources, reducing errors to a workable level. Yet in the coming generation of astronomical surveys, the massive size of the catalog will preclude comprehensive, manual follow-up observations. Downstream scientific analyses that use the catalog will likely be based on samples or selections made by automatic, statistical criteria. Without rigorous bounds on the catalog's error rate, the validity of scientific inferences produced by these analyses will be in question.

A statistical challenge is controlling the error rate for the detection of unknown, aggregate *objects* (e.g., sources) given measurements at the level of *pixels*. The statistics of individual pixels are typically well-understood, and common multiple testing methods (e.g., Benjamini and Hochberg (1995)) can be used to control pixelwise error rates. But the pixels are arbitrary constructs of the measurement process, and interest usually centers instead on detecting sources. The relationship between pixelwise and sourcewise error rates is complicated and obscure, so control of pixelwise error rates does not provide adequate guarantees of sourcewise control.

In this paper, we develop an accurate source detection method that provides rigorous control on the rate of falsely detected *sources* in (astronomical) images. Our method extends the False Cluster Proportion (FCP) controlling procedure of Perone Pacifico et al. (2004), which is reviewed in Section 5. Because of the Poisson statistics common in astronomical images, the original FCP procedure, which we call PP-FCP, underperforms. We improve the FCP algorithm in two ways. In Section 6, we introduce a simulation-based technique for p-value calculation that makes FCP suitable for a wider range of noise models. In Section 7, we use a multi-scale transform to enhance sources and significantly improve power. Taken together, these extensions lead to a new procedure that we call the Multi-scale False Cluster Proportion (MS-FCP) procedure. MS-FCP is a powerful source detection technique that

provides rigorous, tunable control over the rate of false *sources*, whereas techniques in current use provide control over the rate of false *pixels*, if they provide any control at all. In Section 8, we present a simulation study to demonstrate validity and then apply the method to data from the Chandra X-Ray Observatory to demonstrate value-added. (The data are described in Section 3.) We show that MS-FCP has detection power that is competitive with currently used procedures, even detecting sources that were overlooked in previous studies, while providing rigorous control over the rate of falsely detected sources. For the massive-data astronomical surveys of the near future, MS-FCP can meet astronomers' needs for accurate source detection in terms of power, computation, and error control. More generally, MS-FCP provides a useful method for identifying peak-like structures in a random field when one needs to control the error rate on detected objects.

2 The Source Detection Problem

We assume that the input image is an $n \times m$ array of pixels, with the value recorded at pixel (i, j) denoted by Y_{ij} . The photons that contribute to Y_{ij} arise from two components: *sources*, the emission produced by the celestial objects of interest, and *background*, which includes thermal noise, the emission of unresolved objects, interfering radiation sources, atmospheric emission, and all other anomalies or artifacts. Astronomical images essentially measure photon counts for which a Poisson model is appropriate (Cash, 1979). So for our base model, we assume that the Y_{ij} 's are independent with

$$Y_{ij} \sim \text{Poisson}(\lambda_{1,ij} + \lambda_{0,ij}), \quad (1)$$

where $\lambda_{1,ij} \geq 0$, $\lambda_{0,ij} \geq 0$ denote the mean intensity of sources and background, respectively and $\lambda_{1,ij} + \lambda_{0,ij} > 0$. The idea here is that the pixels are measuring the counts in disjoint cells of a Poisson random field across the sky.

The source detection problem is to identify which pixels contain sources and to separate the sources from the background. If $\lambda_{1,ij} > 0$, then we take pixel (i, j) to be a source pixel; otherwise, it is a background pixel. So it is natural to consider this as a multiple testing problem with the null hypothesis at each pixel being that $\lambda_{1,ij} = 0$. At a coarse level, we want to characterize the set $\mathcal{S} = \{(i, j) : \lambda_{1,ij} > 0\}$ of source pixels, but our more specific goal is to identify and locate the underlying sources, so that an accurate catalog can be constructed. This requires a more stringent criterion for success because the objects are coherent, localized aggregates. As Figure 1 shows, with the same number of pixel-wise type I and type II errors, it is possible to get widely varying sources in the resulting catalog. Put another way, our loss function operates on the catalog, not on the pixels themselves.

3 Data: The Chandra Deep Field

We demonstrate our techniques on data from the Chandra X-ray observatory (Weisskopf et al., 2000), one of the most powerful X-ray telescopes available. X-rays are emitted when matter is heated to millions of degrees, such as in the hot gas surrounding large galaxy clusters, during supernovae, or when matter circles a black hole. X-ray data informs astronomers about the nature of celestial objects (such as stars, quasars, galaxies, and clusters), and the distribution of mass and energy in the universe, and can in turn help us understand the universe's history and evolution.

The data we use comprise part of the Chandra Deep Field South (CDFS), a composite image of 11 Chandra observations of a small patch of sky with nearly one million seconds of observing time. This is a long exposure, so we should be able to resolve very distant and faint sources. It is also in a location that can be observed by several complementary ground-

based telescopes. The part of the CDFS that we will analyze and discuss in this section can be seen in Figure 2. The first step in converting the raw image data into an actionable scientific dataset is cataloging all the sources in the image. The sources in the CDFS are thought to be active galaxies and quasars with massive black holes in their centers.

The original analysis of the CDFS was published in Giacconi et al. (2002) [GI]. Their main catalog was created by combining the catalogs from two different source detection algorithms: a modified version of the SExtractor algorithm (Bertin and Arnouts, 1996), and the WAVDETECT algorithm (Freeman et al., 2002). The SExtractor algorithm estimates the background and catalogs as a source any region that is 2.4σ above the estimated background that is also at least 5 pixels in area. These parameters are determined using images simulated to look like the real telescope images. They are selected to create a large catalog which will likely include several false sources. Independently, the WAVDETECT procedure was run using a wavelet transformation specifically tuned for this type of image to detect sources. Again, parameters are calibrated using simulated images. The two catalogs are merged and then refined using follow-up observations in an effort to remove sources that may be false. This last step is an attempt to replicate the detection with independent observations, thus if an object is in the catalog we are fairly certain it is real since it has been observed multiple times, often with a completely different instrument. It is important to note that this detection strategy is designed for a small patch of sky with the resources to follow-up all potential detections. In this scenario, it makes sense to use a detection scheme that casts a wide net. As long as it is possible to go back and verify or reject each potential detection, it is worth tolerating a larger number of spurious detections initially in exchange for the ability to detect some fainter sources. But in future surveys such as the LSST, where the number of objects precludes comprehensive followup,¹ this strategy will fail. Our alternative is to provide statistical control of source-wise error rate in one “pass”.

4 Previous Work in Astronomical Source Detection

Due to its prevalence and difficulty, the source detection problem has generated a wide variety of methods in both the statistical and astronomical literatures. The natural connection in the statistical literature is to multiple-testing methods. Here, one tests the null hypothesis at each pixel in the image that the pixel belongs to the background (i.e., $\mu_{i,j} = 0$ in the language of Section 2) and adjusts the p-values to account for multiplicity. Any standard multiple testing method can be used, but recently developed techniques based on (or motivated by) controlling False Discovery rates are particularly appropriate. See, for instance, Benjamini and Hochberg (1995)[BH], Benjamini et al. (2006), Storey (2002), Sun and Cai (2007), Meinshausen and Rice (2006), Perone Pacifico et al. (2004), and Hopkins et al. (2002).

Astronomical images from deep-sky surveys tend to be sparse – with a strong majority of pixels belonging to the background – and sparsity tends to lead to higher relative power for many of these multiple testing methods (e.g. Benjamini et al. (2006) and Genovese and Wasserman (2004)). However, in some types of astronomical images, there are several components of variation besides the sources themselves. These nuisance terms in the background often have complicated spatial structure, can exhibit long-range dependence, and can produce signal variations comparable to (or greater than) the sources themselves (e.g. Sehgal et al. (2007)). This complicates the analysis for multiple-testing methods because simple comparisons of pixel intensity no longer distinguish source from background. One effective approach is to filter the data to suppress the nuisance terms while

¹Note that repeated observations are planned for all these surveys. While this reduces the error rate, it does not eliminate the problem, especially as interest tends to focus on the deepest objects, which are the least reliably detected.

enhancing sources. This can often be accomplished by using filters which can take advantage of well-defined spectral structures for certain types of nuisance phenomena and sources. In many instances a well chosen filter can sharply improve performance of multiple-testing procedures for source detection.

A more fundamental problem with existing multiple-testing methods for astronomical source detection is that they control error rates only at the pixel level. Unfortunately, there is not a direct calibration between pixel-level and source-level error rates, which limits the direct usefulness of these techniques. Multiple testing approaches that are less pixel-centered have been developed in the related problem of analyzing functional magnetic resonance imaging (fMRI) data. In this problem, the sources are regions of neural activity that reveal themselves through a measurable change in blood flow response. Worsley et al. (1996) and Worsley et al. (2002) use level sets of a random field formed from test statistics to identify regions containing sources. Heller et al. (2006) construct a test based on clusters instead of pixels in the fMRI setting, but this technique takes advantage of a temporal dimension that may not be available in general for astronomical problems.

Source detection understandably garners much attention in the astronomical literature. The source detection methods used by astronomers fall into three general classes: simple thresholding, peak-finding algorithms, and Bayesian algorithms. Simple thresholding consists of choosing an intensity cutoff for pixel-wise statistics and classifying any pixel above threshold as a source. It is popular because it is simple and fast, easily computed by the popular SExtractor software (Bertin and Arnouts, 1996). Before thresholding, filters are often applied to the raw data to suppress confounding background signals as in Vikhlinin et al. (1995) and Melin et al. (2006). Hopkins et al. (2002) also use simple thresholding, choosing the threshold to control the False Discovery Rate via the method of Benjamini and Hochberg (1995), using simulated telescope images to calibrate between rate of false pixels and the rate of false sources.

Peak-finding algorithms search for local maxima in the denoised image and catalog them as sources. Vale and White (2006) and Sehgal et al. (2007) use peak-finding algorithms to look for large galaxy clusters in simulated radio telescope images. Wavelet-based techniques are commonly used to detect peaks in X-ray images as in Freeman et al. (2002), Valtchanov et al. (2001) and Giacconi et al. (2002). Damiani et al. (1997) and Gonzalez-Nuevo et al. (2006) provide good overviews of the popular Mexican Hat wavelet as a tool for source detection.

Bayesian techniques have become popular with astronomers and several have applied Bayesian methods to source detection as in Savage and Oliver (2007), Strong (2003), and Hobson and McLachlan (2003). Bayesian detection algorithms typically define models for sources, often two-dimensional Gaussians, and attempt to distinguish them from backgrounds via inference on a posterior distribution. Guglielmetti et al. (2009) use Bayesian mixture models to separate sources from background, while Carvalho et al. (2009) propose ways to speed up computations in Bayesian source detection which can be slower than other methods. Bayesian algorithms typically require stronger a priori assumptions and can be more computationally expensive than thresholding or peak-finding algorithms.

An important difference between the approaches in the statistical and astronomical literature lies in how performance is assessed. The statistical approaches are assessed theoretically, with a focus on formal bounds on ensemble criteria such as error rates and power. The astronomical approaches are often assessed empirically, using realistic, high-end simulations of the telescope images to test and calibrate the methods. Formal guarantees provide reassurance that the methods will perform well despite variations in the data. While such

guarantees require assumptions, the assumptions can in practice be quite weak, such as the minimal assumptions required for many multiple testing methods. Moreover, as we will argue below, the guarantees provide essential information for downstream studies, particularly in the context of massive astronomical surveys. For formal guarantees to be useful, however, the criteria that are bounded must match the scientific goals, and the procedures that achieve the bounds must be computable within the allowable time. Empirical assessments, on the other hand, are less concerned with possible but unrealistic data sets but rather test performance on data that look as realistic as possible a priori. Astronomers expend great ingenuity and effort constructing simulations that account accurately for all known effects; these simulations can require months of computing for a single run. But the success of empirical assessments relies on the simulated and real data being both quantitatively and qualitatively similar. Unexpected effects or even an unexpected balance of effects can move the real data outside the domain of the assessments, leading to undesirable performance. Because multi-year astronomical surveys typically must decide upon data-analysis methods far in advance of data collection and build these methods into the system, the unexpected can have serious consequences.

In this paper, our goal is to provide a fast source-detection method that provides formal performance guarantees while providing good performance for realistic source configurations. Our focus is on controlling error rates for the sources themselves rather than pixels in the image. (Specifically, we develop a procedure for controlling a source-wise false discovery rate.) Statisticians have developed numerous methods that give bounds on pixelwise error rates but have not been focused on detecting aggregate objects, while astronomers have been thinking about detecting objects but without a rigorous approach to controlling error rates. The family of techniques we develop do both: control the error rates while making inferences about the sources themselves.

5 The False Cluster Proportion Algorithm

An alternative to traditional pixelwise testing is the False Cluster Proportion (FCP) procedure introduced by Perone Pacifico et al. (2004), which is designed to use the sources themselves as the unit of inference.

The FCP procedure treats the image as a realization of an underlying random field, where each location (pixel) belongs to either the *background* or *signal*. It identifies sources from the data so that the proportion of false sources among all sources detected is below a pre-specified bound with high probability. A false source here refers to a cluster whose overlap with the background pixels is above a pre-specified tolerance.

Let S_0 denote the (unknown) set of background pixels. The procedure begins by constructing a random set U such that

$$P(U \supseteq S_0) \geq 1 - \alpha. \quad (2)$$

We call U a $1 - \alpha$ *confidence superset* for S_0 . For any cluster of pixels C , we call that cluster false (at threshold ε) if the proportion of background pixels in C is at least ε ; that is,

$$\frac{\#(C \cap S_0)}{\#(C)} \geq \varepsilon. \quad (3)$$

Notice that with probability at least $1 - \alpha$,

$$\frac{\#(C \cap U)}{\#(C)} \geq \frac{\#(C \cap S_0)}{\#(C)} \geq \varepsilon. \quad (4)$$

For any t , let L_t denote the set of all pixels that have value $> t$. Each L_t can be decomposed into k_t connected components $C_{1,t}, C_{2,t}, \dots, C_{k_t,t}$ where any two pixels in a cluster can be joined by a path of contiguous pixels². We refer to each component as a cluster, and our goal is to decide which of these clusters correspond to the sources we want to detect and which are false in the sense defined above.

In the typical Astronomical source detection problem, clusters range in size from one pixel to a few hundred, and are roughly circular with the intensity usually greatest at the center³. The goal is to choose a threshold statistic T that ensures the proportion of detected objects that are false is sufficiently low.

Define the true *false cluster proportion* as a function of threshold, (t)

$$\Xi(t) = \frac{1}{k_t} \# \{1 \leq i \leq k_t : \#(C_{it} \cap S_0) \geq \varepsilon \#(C_{it})\}, \quad (5)$$

and an analogous envelope function built on the confidence superset

$$\bar{\Xi}(t) = \frac{1}{k_t} \# \{1 \leq i \leq k_t : \#(C_{it} \cap U) \geq \varepsilon \#(C_{it})\}. \quad (6)$$

By the definition of the confidence superset, it follows that

$$P(\Xi(t) \leq \bar{\Xi}(t) \forall t) \geq 1 - \alpha. \quad (7)$$

Moreover, if we choose T such that $\bar{\Xi}(T) = \alpha$, then $P(\Xi(T) > \alpha) = 1 - \alpha$, so we have controlled the source-wise false discovery proportion with a specified probability. To find T , we must search over possible values of t , and then we construct a catalog using the k_T clusters derived from the components of L_T . For example, suppose we want to bound the rate of false clusters to be less than 10%. Equation (7) tells us that if we can find the value $T_{.1}$ such that $\bar{\Xi}(T_{.1}) = .1$ then the catalog containing the clusters $C_{1,T_{.1}}, \dots, C_{k_{T_{.1}},T_{.1}}$ as sources is likely to have less than 10% false sources. This procedure gives valid control, but its power will be reduced to the degree that the envelope function overcovers. The key to good power is finding as small a confidence superset as possible.

FCP produces a source catalog that satisfies a useful, overall probabilistic guarantee. But it can also be helpful in practice, at least as a diagnostic, to assess the strength of detection for each individual cluster. One statistic for this purpose is analogous to a p-value: just as a p-value is the smallest level at which we can detect a significant difference from the null, we defined p_C for a given cluster C to be the smallest FCP at which the cluster would be considered a detected source. To find p_C , begin by finding the value of its brightest pixel in C ,

²For a given (interior) pixel, we call the eight pixels surrounding it contiguous pixels to the original pixel.

³Note that in our definition two overlapping sources will be treated as one “theoretical” source. This is a common simplifying assumption in many source detection algorithms. Since sources are sparse on the sky, this is unlikely; however there are post-processing methods that can be employed to disambiguate overlapping sources. SExtractor has one such disambiguation algorithm built in to the software

$$t_c = \max\{Y_{ij} : (i, j) \in C\}. \quad (8)$$

Now for any threshold $t > t_c$, the cluster C will *not* be detected, and hence

$$p_c = \bar{\Xi}(t_c), \quad (9)$$

which can be computed directly from the envelope function. The possible values of p_c lie in $[0, 1]$ with smaller values corresponding to “stronger” detections.

5.1 The PP Method for Calculating U

To control the FCP, it is necessary to construct a confidence superset U that satisfies (2) for a specified $0 < \alpha < 1$. Perone Pacifico et al. calculate U by looking for the smallest set A for which the null of the following test cannot be rejected.

$$H_0: A \subset S_0 \text{ versus } H_1: A \not\subset S_0 \quad (10)$$

In practice, we use the maximum pixel value in the set A as the test statistic, yielding a p-value

$$p(x, A) = P_0(\max(Y_{ij} \in A) \geq x). \quad (11)$$

U is calculated by starting with the entire image and then progressively removing the pixels with the highest intensity (those least likely to be part of the background) until the p-value of the remaining set passes α . In general, the p-value in Equation (11) is difficult to calculate exactly, but there are several approximations that can be used (Adler and Taylor, 2007). Perone Pacifico et al. (2004) use the Piterbarg approximation (Piterbarg, 1996) which assumes that the field is a locally stationary Gaussian random field whose covariance function is sufficiently smooth. This method yields valid confidence supersets and works well in practice with smooth Gaussian data. For more details, see Perone Pacifico et al. (2004). Henceforth, we will refer to this construction of U as the PP method and the FCP algorithm using this U as the PP-FCP method.

5.2 PP-FCP Applied to CDFS Data

To illustrate how FCP is used and to provide a standard for comparison, we now apply the PP-FCP method to the CDFS data. In the CDFS data, the potential sources are active galaxies and quasars, which emit X-rays at a high rate. We would like to detect as many of these objects as possible without making too high a portion of spurious detections. Each pixel records the number of photons that strike the detector, and the pixel values are modeled as Poisson random variables. The mean background in these images is small, approximately 0.3 photons per pixel,⁴ which violates the Gaussian background assumption of the PP procedure. Our initial approach to dealing with the non-Gaussianity is to transform the data so that the Gaussianity assumption is more appropriate. Specifically, we first smooth the data with a Gaussian filter and then calculate the square root of each smoothed pixel. This operation may wash out very faint sources and effectively reduces the resolution, but we have found that with only a small amount of smoothing we can satisfy the Gaussianity assumption without losing too much information. Smoothing the data with a

⁴The background is estimated using the median of the data after trimming pixels that are more than five IQR from the center of the data. Since the vast majority of pixels are background, the estimate is robust to the amount of trimming

Gaussian filter effectively aggregates the counts while preserving most of the structure of the Poisson data. The former (aggregation) demands a larger bandwidth; the latter (preserving structure) demands a smaller bandwidth. So in practice, we choose the smallest bandwidth that gives a reasonable number of terms in the average. We use a plug-in estimator based on image size and (estimated) noise level giving the bandwidth for which the distribution of smoothed Poisson pixels is close to Normal for the given noise level as determined by simulation. The square root transformation then makes the background approximately Normal, similar to Anscombe (1948). Using simulations, we have tested the effectiveness of our smooth-then-square-root procedure for low rate Poisson data and found that for rates as low as 0.2, it yields approximately normal data. While this depends on the nature of the sources, the distribution is driven almost entirely by the low count pixels. Hence, we compute the plug-in estimator, via simulation, in the worst case and use the resulting bandwidth for the data.

After the normalizing transformation, we calculate the 95% confidence superset for the background using the PP method. We then run PP-FCP on the transformed data as described in Section 5, bounding the rate of false sources to be less than 10%. Because our U is very conservative (it contains every pixel from the null with probability at least 0.95), we select less than but close to 1 (here 0.99). Specifically, we consider any cluster that lies completely within U as false. The intent of this choice is to maximize our detection power: given that U covers S_0 , any cluster intersecting in U^c corresponds to a source. In this sense, the choice of α and β are dependent, and our choice puts all the tuning into the latter. Both on real and simulated data we have found this value to work well – all of the examples presented in the Results section use this choice of α . In addition, we do not have to make strong assumptions about the number or distribution of sources. The main assumption is that our normalizing transformation does indeed give us a background that is approximately Gaussian, an assumption which can easily be verified. FCP, using the PP method to calculate U , detects 17 sources, all of which are in the previously published GI catalog as seen in Figure 3. Since this catalog has been verified with follow up observations we are confident that all the detections are real. However, this analysis is clearly missing many clusters that should be detectable – nine of the 26 clusters in the Main catalog of GI were missed using FCP with the PP method.

While PP-FCP gives detections with valid control, its power can be improved. Comparing the detections in Figure 3 to Figure 2, it is clear that the clusters that FCP missed are small and faint. Because FCP compares the intensity of pixels to a global threshold and does not utilize local information, it struggles to detect these faint clusters. One way to boost the power is to filter the data in a way that enhances sources to help separate them from the background. In this case, we will show that a multi-scale filter that enhances peaks will bring out the peak-like sources, making them easier to detect from the background. However, the filter will change the distribution of the pixels from the Poisson assumption. To detect sources after filtering while maintaining the desired control over the rate of false sources, we need to adapt FCP so that it can handle more varied noise conditions. We will develop these ideas in the next two sections. First, we extend the FCP method by using a more general method for calculating the confidence superset U . Then, we construct a multi-scale filter for enhancing astronomical sources.

6 Extending the FCP to Non-Gaussian Noise Families

The PP method for constructing a confidence superset is limited to cases where pixels are (approximately) or can be transformed to look like a realization of a smooth Gaussian random field. But in many situations, no suitable approximation or transformation is

available. Here we develop a simulation-based approach to constructing p-values for building a $1 - \alpha$ confidence superset from a broader family of noise distributions.

Assume that we have n total pixels with background noise distribution F . (For concreteness, we take these as arranged in a rectangular image, but for these purposes, other domains are acceptable.) We assume that we can sample numerically from F . When F is known (or assumed), this is straightforward. More commonly, F is specified in terms of parameters that must be determined empirically, either from the image itself or from other data. For instance in astronomical images, the background photon intensity and the seeing, which are essentially constant across images with small fields of view, can be often estimated, respectively, from background patches in the data and the neighborhoods of identified stars. We then set F from the noise distribution in the family with that estimated parameter (possibly overdispersed slightly), with only slight loss of variation. More generally, we can set F implicitly by resampling the values of pixels classified as background by a pixelwise multiple-testing procedure.

Our procedure begins by simulating images from the noise distribution F and calculating the distribution of maximum values over subsets of the images. Because the p-values $p(x, A) = P_0(\max(Y_{ij} : (i, j) \in A) \geq x)$ (like the Piterbarg approximation to it) are dependent primarily on $a = \text{area}(A)$ rather than the shape of A , it is in practice sufficient to consider one subset for each possible value of a . For each subset we build up empirical distributions for the maxima, and calculate empirical p-values accordingly

$$p(x, A) \approx \frac{1}{B} \# \{1 \leq i \leq B : Y_{(a)}^i \geq x\}, \quad (12)$$

for selected subsets A . We then use these p-values (in place of the Piterbarg approximate p-values) in the FCP method (see Section 5), yielding a set of detected sources.

Algorithm: Simulating p-values:

For each $b \in \{1, \dots, B\}$

- 1 Simulate an image, Y^b , with the same dimensions as the data, with noise distribution F and no sources
 - 2 Record the maximum value of the image, $Y_{(n)}^b$
 - 3 For $k = 1, 2, \dots, s$ randomly remove k pixels and record the maximum remaining pixel value $Y_{(n-k)}^b$. $n - s$ is the smallest number of pixels for which you want to calculate the p-value, thus $n - s$ only needs to bound the number source pixels.
-

In fact, the dependence on the source pixels is so weak that we can get a conservative approximation to the procedure that still performs well by thresholding method. Specifically, we note that $P(x, A) > P(x, A - k)$ $k > 0$. Since we are dealing with large images, where the source signal is sparse relative to the background, the maximum over n pixels is not going to be substantially different than the maximum over $n - k$ pixels when $n \gg k$. Thus if we use the following algorithm we will get a comparable confidence superset directly and with less computation.

Algorithm: Simulating a Confidence Superset

- 1 For each $b \in \{1, \dots, B\}$, simulate an image, Y^b , with the same dimensions as the data, with noise distribution F and no sources, then record its maximum
- 2 Calculate $r_{1-\alpha}$, the $1 - \alpha$ percentile of the maxima
- 3 Using the original data image, calculate the level set

$$U = \{ Y_{ij} : Y_{ij} < r_{1-\alpha} \}$$

Using the second algorithm is faster because we avoid the computation of the sB maxima at the cost of having a bigger confidence superset U . However as long as $n \gg k$, which will be the case for telescope images where sources are relatively sparse on the sky, the size difference is negligible. The second algorithm also allows us to calculate U more directly without having to search over sets or approximate any p-values.

It is worth highlighting the difference between the simulations we are using for computing p-values and those astronomers use for empirical evaluation of source detection methods. Our simulations depend on the nature of the noise distribution, which is rather robustly determined by the nature of the instrument, and such assumptions can be readily checked and varied. On the other hand, the simulations used by astronomers for evaluation, while tremendously detailed and sophisticated, depend strongly on the specific nature of the physical processes determining the *sources*, which are much less constrained by prior knowledge. The results of the evaluation are more sensitive to these assumptions and are difficult to validate.

In addition to handling a wide range of noise distributions – any distribution from which we can draw numerically – our simulated confidence superset U can easily incorporate additional pre-processing steps, such as filters used to enhance sources, remove bad pixels, correct for background contamination (such as the effects of seeing, as represented through the point-spread function), or transform to an alternate domain (e.g., wavelet transforms). The basic filtered algorithm is shown below.

Algorithm: Simulating a Confidence Superset for Filtered Data

- 1 For each $b \in \{1, \dots, B\}$, simulate an image, Y^b , with the same dimensions as the data, with noise distribution F and no sources.
- 2 Apply a filter G to Y^b to get $G(Y^b)$
- 3 Record the maximum pixel value from $G(Y^b)$
- 4 Calculate $r_{1-\alpha}$, the $1 - \alpha$ percentile of the maxima
- 5 Using the filtered data image, $G(Y)$, calculate the level set

$$U = \{ G(Y)_{ij} : G(Y)_{ij} < r_{1-\alpha} \}$$

7 Improving FCP by Enhancing Detection Power

We have seen with the Chandra data that FCP, using the intensity at each pixel and the PP method, does not achieve power comparable to the methods in Giacconi et al. (2002)[GI]. In

Section 6, we outlined a simulation-based approach which replaces the PP method, allowing for more general test statistics that do not necessarily have to be Gaussian. In this section, we describe a new technique which directly address a problem contributing to the lack of power seen in Section 5.2. The problem is that rules based on the magnitude of a pixel, are not necessarily the best at discriminating sources aggregated over pixels from background. We expect, due to random fluctuations, to get some individual background pixels with relatively large counts. Compared to rules based on the magnitude of pixels, rules that take into account local information can better differentiate between these occasional pixels with large counts and a clump of pixels that all have large counts. This is in part the reason why wavelets are commonly used in source detection. We introduce an alternative, the Multi-scale Derivative transform. The Multi-scale Derivative transform uses local information to identify regions that stand out from the background as sources, instead of looking at the magnitude of each pixel individually.

The idea behind the Multi-scale Derivatives is to apply a filter to the data simultaneously over a range of bandwidths and to examine the derivative of the filtered image with respect to the bandwidth. Figure 4 shows how this works in a stylized example. Start at a location with no source. Then at small bandwidths, our smooth estimate of the null location remains low, and this will stay roughly the same until we reach a bandwidth that is big enough to hit a source, at which point our smoothed estimate will jump up as the source gets smoothed into the null space. Alternatively, at a source location we will have a high value for our smooth estimate if our bandwidth is small and as we increase the bandwidth, our estimate will get lower and lower as null areas are smoothed into the source.

More formally, the multi-scale derivative procedure can be expressed in two steps: i) apply a convolutional filter with a specially designed kernel over a set of bandwidths $\mathcal{H}$, and ii) adaptive local selection of bandwidths. The result is an image M that will enhance fine scale sources in the image without significant reduction in detection power for larger sources.

Let ϕ be the two-dimensional standard Normal density, and let ϕ_h be the scaled version, $\phi_h(z) = h^{-2} \phi(z/h)$. Define $K_h = -\frac{\partial}{\partial h} \phi_h$ and note that

$$K_h(x) = -h \operatorname{trace} \phi_h''(x) = \frac{1}{h} \left(d - \frac{\|x\|^2}{h^2} \right) \phi_h(x), \quad (13)$$

where ϕ_h'' is the Hessian of ϕ_h with respect to x .

For a field Y , the multi-scale derivative for a given bandwidth h is a field of the form

$$\widehat{M}_h = K_h \star Y. \quad (14)$$

Note that we have reversed the sign (relative to Figure 4) so that large positive values correspond to sources.

We construct the filtered image M_h over a pre-specified grid of bandwidths $\mathcal{H}$, selected to cover the well-known range of source sizes. The smoothing operation can be calculated quickly by convolving the data with a Gaussian kernel in Fourier space, and we can use the same trick to calculate the required derivative of the smoother. The multi-scale derivative image, M is created by taking the maximum value at each pixel over all scales:

$$\widehat{M}(x) = \max_{h \in \mathcal{H}} \widehat{M}_h(x) \quad (15)$$

We then use M as the input to the FCP method, looking for large positive values to identify sources.

In the appendix, we show that when $Y = f + Z$, where f is a function representing sources and Z is a sufficiently smooth, mean zero, Gaussian random field representing background noise, then M enhances the power for detecting sources that are sharper than the typical covariance scale of Z , without significantly reducing power for larger (and typically easier to detect) sources. Because deep (and most often small) are usually of greatest interest in astronomical surveys, the multi-scale derivative enhances the marginal detection power for the most desirable targets.

As a demonstration of the technique, Figure 5 shows a small dataset due to Sehgal et al. (2007) simulated to look like data from the Atacama Cosmology Telescope (Kosowsky, 2003), a ground-based radio telescope used to detect large galaxy clusters. The clusters we wish to detect are visible as bright spots in the simulated image on the left. In practice, we cannot observe these clusters directly because they are obscured by several other confounding signals. To isolate the clusters, Sehgal et al. use a Wiener filter, which combines data from three different radio frequencies, as seen in the middle image. Our goal is to detect the galaxy clusters from the filtered image. The brightest sources are visible in the Wiener filtered image, but the background is spatially varying and is much brighter in certain areas than in others. Any detection algorithm based on the absolute magnitude of pixels will not perform well on this image. Instead we calculate the Multi-scale Derivative image from the filtered image as seen in the right panel of Figure 5. The Multi-scale Derivative image suppresses the background and highlights the sources, which will make source detection much easier. In this example, the Multi-scale Derivative is used as a filter which enhances sources in the image while suppressing background. Alternatively, it can be viewed as a test-statistic which incorporates local information to test for peaks in the data. We can calculate a confidence superset⁵ for M which can be used to create a catalog from M with False Cluster Proportion control; we call this the MS-FCP procedure. The goal is to exploit the enhanced power to produce a purer catalog.

8 Results

8.1 Simulation Results

To evaluate both PP-FCP and MS-FCP, we ran the procedures on images simulated to look like the Chandra data. One hundred 512×512 pixel images were made using code developed by Dr. Peter Freeman. The number and intensities of the sources is simulated according to Virani et al. (2006), using exposure times and a point spread function consistent with the CDFS data. For each image we do the following:

Algorithm: Simulation Procedure

- 1 Simulate an image as described above and add Poisson noise to each pixel
- 2 Smooth the image with a Gaussian filter using a bandwidth of 1.5 pixels, then take the square root at each pixel and standardize the pixels. This makes the background pixels look approximately standard Normal

⁵Algorithm: Simulating a Confidence Superset for Filtered Data in Section 6

- 3 Generate two catalogs by running the PP-FCP and MS-FCP algorithms. Both algorithm's parameters are set so that we control the proportion of false clusters to be less than 10% with probability .95, and using $\alpha = 0.99$.

Each image was subjected to four noise conditions. In the first, no additional noise was added while in the other three, Poisson noise was added to each pixel with means one, two, and five. PP-FCP and MS-FCP were then run on each of the four hundred images such that at most 10% of the detections will be false positives with probability .95. For each run, we checked each detection made by our algorithms against the true catalog used to simulate the image. If there was true source whose center was within 2 pixels of a detected source we declare the detected source to be real, otherwise it is spurious. From this information, we calculate the *purity* of the created catalog as the proportion of detections that correspond to real objects. Table 1 shows the mean purity for the two algorithms across the noise conditions.

The purities indicate that we are not making many false detections; on average less than 0.5% of PP-FCP's detections are false while MS-FCP has slightly more false detections, between 1.5–2%, which is below the target threshold and reflects somewhat improved power. Table 2 show the proportion of trials in which the percentage of false clusters exceed the 10% target for the two methods.

In the simulations we rarely exceed the pre-specified 10% falsely detected clusters bound – significantly less than the 5% we are willing to tolerate. This also indicates there may be room for improvement. Future work will examine ways to get closer to the bound – by getting closer to the bound we should be able to get more power while maintaining an acceptable false detection rate. One potential method to get closer to the bound, is using the Higher Criticism method proposed by Donoho and Jin (2008) as a starting point to generate a confidence superset. These simulations also indicate that our choice of $\alpha = 0.99$ is reasonable in the sense that it is not generating a large number of spurious sources.

In these simulations, our methods control the rate of false positive detections as advertised. But for them to be useful they also need to be able to detect the sources we want to detect. Figure 6 shows boxplots of the percentage of sources recovered by noise type and by detection procedure. As expected MS-FCP has better detection power across all noise conditions. Detection power is directly related to the source intensity – brighter sources are easier to detect than faint ones. Thus we can model detection power as a function of source intensity. Figure 7 shows logistic regressions, modeling the probability of a detection as a function of source intensity, for FCP and MS-FCP from the highest noise condition data. We see that MS-FCP is much better at detecting fainter sources than FCP, for instance we predict that 95% of sources with source intensity 67.3 will be detected by MS-FCP while PP-FCP would only detect about 27% of these sources.

8.2 MS-FCP Applied to CDFS Data

Using the Multi-scale Derivative filter on the CDFS image is intended to enhance the clusters, making them stand out more from the background. After performing the smooth-then-root transformation as described previously, we compute the Multi-scale Derivative image. We then ran MS-FCP with the same parameters as before. With MS-FCP, we detect 24 of the 26 sources detected in the main catalog. We do not make any detections outside of the catalog. Since the detections in the catalog have been replicated in follow-up observations this suggests that we are not making any false detections while still detecting almost all of the real objects.

Without having to simulate images made to look like the telescope images, making minimal assumptions about the underlying astronomy, and without conducting follow-up observations, we can still make the statement that with probability .95 less than 10% of our detections will be false. The other algorithms cannot make such an explicit guarantee of a pure sample.

In Giacconi et al. (2002), which we refer to as GI, their algorithm is designed for a small patch of sky with the ability to follow-up all potential detections. In this scenario, it makes sense to tune the detection scheme to cast a wide net. Because we can go back and verify or reject each potential detection, we will tolerate a larger number of spurious sources initially in exchange for the ability to detect some fainter sources. The wide net approach can be incorporated into FCP methods by relaxing the False Cluster Proportion threshold. For instance, if we allow for the proportion of false sources to grow to 20% we detect 26 sources, two of which are new detections outside of the main catalog. Both these new sources have been subsequently detected using other datasets (Luo et al., 2008) so we are confident that they are both real. Using the wide-net approach we have not only captured almost all of the detections from the previous analysis (but with rigorous error control), we have also found two new sources that were missed by the astronomers using both thresholding and wavelet-based methods.

We expect the proportion of false sources one is willing to tolerate to be a parameter that astronomers will have to determine to suit the problem at hand. For instance, the “wide-net” approach works adequately for small data sets, but newer telescopes, like LSST, will collect such a large amount of data that following up every one of billions of potential detections will be expensive if it is even possible. In this scenario having a detection procedure that controls for the proportion of false detections automatically in one pass will be vitally important.

9 Discussion

Source detection from image data is a long-standing problem in numerous fields, notably in astronomy, where it is an essential part of constructing the basic data sets used in astronomical research. In astronomy, source detection algorithms are one step in a multi-step process, where sources detected by the algorithm are verified by follow-up observation, and then used for downstream analyses and studies. This has been an effective approach because the process can achieve high power while tolerating a large proportion of false positives in the initial step. Indeed, as long as follow-up is cheap, the standard trade-off between specificity and sensitivity breaks down. Power is paramount, and type I error rates need only be coarsely controlled. In other words, astronomers have had little reason to focus on formal statistical performance criteria, such as error rates.

In the coming generation of astronomical surveys, we believe this will change, that having formal error rates for source detection procedures will become critical. First, the sheer number of sources will be so great that a small to moderate false positive rate will require a large number of follow up observations. That is, follow up costs will not be ignorable. While new technologies may make it faster and more efficient to conduct limited follow-up, by directly controlling the false detection rate in the detection algorithm, we can scale back the amount of follow up needed (freeing up the telescopes for other studies) while at the same time provide a measure of security that the catalog is not overly polluted with false detections. Second, with the enormous source catalogs to come, downstream studies will rely more and more on large samples from the main catalog. The validity of such studies will rely on having a suitably pure sample, requiring a bound on the false positive rate.

Our approach in this paper is to construct a fast and accurate source detection method that can provide rigorous bounds on error rates. Unlike most approaches in the statistical literature, we obtain error rates at the source level – the error rates of interest for astronomers. We extend the FCP method of Perone Pacifico et al. (2004) to handle a wider variety of noise models, and we introduce a multi-scale statistic that substantially improves power for detection of focal regions such as astronomical sources. We show that the resulting procedure, which we call MS-FCP, is competitive *without followup* with existing astronomical source detection techniques that use followup while providing performance bounds that existing techniques do not.

There are several directions for improvement. First, our confidence superset construction tends to be conservative, giving fewer false positives than would be acceptable and thus giving lower power than the technique can achieve. Incorporating more problem specific assumptions into our p-value calculations is one possible improvement to consider. Another promising approach would be to exploit the sparsity of objects in the astronomical images by adapting the Higher-Criticism method (Donoho and Jin, 2008) as a starting point. Second, with regard to surveys like the LSST, we can account for the temporal aspect of the data, with each patch of sky observed every few nights. Using this temporal dimension we will be able to detect small changes in the sky that occur over time. Future work will examine using methods like MS-FCP to detect these transient objects by incorporating the objects that are not changing over short time scales into the background. Finally, while the multi-scale derivative statistic performs well for astronomical sources, it is possible to tailor the approach to different source configurations and different classes of objects. The general approach behind the MS-FCP generalizes. We combine two steps: apply FCP with a suitable confidence superset to make inferences at the level of *objects* and use a filter to increase power for a relevant class of objects. As the Appendix demonstrates, the multi-scale derivative improves contrast for sharp, localized features like those in modern astronomical images. For different classes, alternative filters will provide better enhancement, such as simple smoothers for broad-and-low regions of signal. We have demonstrated another example in the analysis of neuroimaging data (Friedenberg, 2010).

The MS-FCP method is applicable for source detection problems in astronomy but applies as well to a wide variety of problems in other fields. It exhibits good power, fast computation, and gives a performance bound that allows inferences on aggregate objects in an image.

Acknowledgments

This work is supported by National Science Foundation (NSF) grants 486560 and DMS0806009, National Institute of Health (NIH) grant 1R01NS047493, National Aeronautics and Space Administration (NASA) grant NNX07AH61G and the Bruce and Astrid McWilliams Center for Cosmology. The authors would like to thank Arthur Kosowsky and Neelima Seghal for the Atacama Cosmology Telescope data, Peter Freeman for his help with the Chandra X-ray telescope data and simulations, and the referees and AE for their insightful comments.

References

- Adler, R.; Taylor, J. Random fields and geometry. Vol. 115. Springer Verlag; 2007.
- Anscombe F. The transformation of Poisson, binomial and negative-binomial data. *Biometrika*. 1948; 35:246–254.
- Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistic Society*. 1995; B:289–300.
- Benjamini Y, Krieger AM, Yekutieli D. Adaptive Linear Step-Up Procedures that control the False Discovery Rate. *Biometrika*. 2006; 93:491–507.
- Bertin E, Arnouts S. SExtractor: Software for source extraction. *A&AS*. 1996; 117:393–404.

- Carvalho P, Rocha G, Hobson MP. A fast Bayesian approach to discrete object detection in astronomical data sets - PowellSnakes I. *MNRAS*. 2009; 393:681–702.
- Cash W. Parameter estimation in astronomy through application of the likelihood ratio. *The Astrophysical Journal*. 1979; 228:939–947.
- Damiani F, Maggio A, Micela G, Sciortino S. A Method Based on Wavelet Transforms for Source Detection in Photon-counting Detector Images. I. Theory and General Properties. *The Astrophysical Journal*. 1997; 483:350–369.
- Donoho D, Jin J. Higher criticism thresholding: Optimal feature selection when useful features are rare and weak. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105:14790–14795. [PubMed: 18815365]
- Freeman PE, Kashyap V, Rosner R, Lamb DQ. A Wavelet-Based Algorithm for the Spatial Analysis of Poisson Data. *The Astrophysical Journal Supplement Series*. 2002; 138:185–218.
- Friedenberg, DA. PhD thesis. Carnegie Mellon University; 2010. Adaptive Cluster Detection.
- Genovese C, Wasserman L. A stochastic process approach to false discovery control. *The Annals of Statistics*. 2004; 32:1035–1061.
- Giacconi R, Zirm A, Wang J, Rosati P, Nonino M, Tozzi P, Gilli R, Mainieri V, Hasinger G, Kewley L, Bergeron J, Borgani S, Gilmozzi R, Grogin N, Koekemoer A, Schreier E, Zheng W, Norman C. Chandra Deep Field South: The 1 Ms Catalog. *The Astrophysical Journal Supplement Series*. 2002; 139:369–410.
- Gonzalez-Nuevo J, Argueso F, Lopez-Caniego M, Toffolatti L, Sanz JL, Vielva P, Herranz D. The Mexican Hat Wavelet Family. Application to point source detection in CMB maps. *MNRAS*. 2006; 369:1603.
- Guglielmetti F, Fischer R, Dose V. Background-source separation in astronomical images with Bayesian probability theory - I. The method. *MNRAS*. 2009; 396:165–190.
- Heller R, Stanley D, Yekutieli D, Rubin N, Benjamini Y. Cluster-based analysis of fMRI data. *Neuroimage*. 2006; 33:599–608. [PubMed: 16952467]
- Herschel W. Catalogue of One Thousand New Nebulae and Clusters of Stars. By William Herschel, LL.D. F. R. S. *Royal Society of London Philosophical Transactions Series I*. 1786; 76:457–499.
- Hobson MP, McLachlan C. A Bayesian approach to discrete object detection in astronomical data sets. *MNRAS*. 2003; 338:765–784.
- Hopkins AM, Miller CJ, Connolly AJ, Genovese C, Nichol RC, Wasserman L. A New Source Detection Algorithm Using the False-Discovery Rate. *AJ*. 2002; 123:1086–1094.
- Kosowsky A. The Atacama Cosmology Telescope. *New Astronomy Reviews*. 2003; 47:939.
- Luo B, Bauer FE, Brandt WN, Alexander DM, Lehmer BD, Schneider DP, Brusa M, Comastri A, Fabian AC, Finoguenov A, Gilli R, Hasinger G, Hornschemeier AE, Koekemoer A, Mainieri V, Paolillo M, Rosati P, Shemmer O, Silverman JD, Smail I, Steffen AT, Vignali C. The Chandra Deep Field-South Survey: 2 Ms Source Catalogs. *The Astrophysical Journal Supplement Series*. 2008; 179:19–36.
- Meinshausen N, Rice J. Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Annals of Statistics*. 2006; 34:373–393.
- Melin J-B, Bartlett JG, Delabrouille J. Catalog extraction in SZ cluster surveys: a matched filter approach. *A&A*. 2006; 459:341–352.
- Perone Pacifico M, Genovese C, Verdinelli I, Wasserman L. False Discovery Control for Random Fields. *Journal of the American Statistical Association*. 2004; 99:1002–1014.
- Piterbarg, V. Asymptotic Methods in the theory of Gaussian processes and fields. *American Mathematical Society*; 1996.
- Savage RS, Oliver S. Bayesian Methods of Astronomical Source Extraction. *The Astrophysical Journal*. 2007; 661:1339–1346.
- Sehgal N, Trac H, Huffenberger K, Bode P. Microwave Sky Simulations and Projections for Galaxy Cluster Detection with the Atacama Cosmology Telescope. *The Astrophysical Journal*. 2007; 664:149–161.
- Storey JD. A Direct Approach to False Discovery Rates. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*. 2002; 64:479–498.

- Strong AW. Maximum Entropy imaging with INTEGRAL/SPI data. *A&A*. 2003; 411:L127–L129.
- Sun W, Cai TT. Oracle and Adaptive Compound Decision Rules for False Discovery Rate Control. *Journal of the American Statistical Association*. 2007; 102:901–912.
- Tyson JA, the LSST Collaboration. Large Synoptic Survey Telescope: Overview. *PROCSPIE INTSOCOPE*. 2002; 4836:10.
- Vale C, White M. Finding Clusters in SZ Surveys. *New Astron*. 2006; 11:207.
- Valtchanov I, Pierre M, Gastaud R. Comparison of source detection procedures for XMM-Newton images. *A&A*. 2001; 370:689–706.
- Vikhlinin A, Forman W, Jones C, Murray S. Matched Filter Source Detection Applied to the ROSAT PSPC and the Determination of the Number-Flux Relation. *The Astrophysical Journal*. 1995; 451:542–.
- Virani SN, Treister E, Urry CM, Gawiser E. The Extended Chandra Deep Field-South Survey: X-Ray Point-Source Catalog. *The Astronomical Journal*. 2006; 131:2373.
- Weisskopf M, Tananbaum H, van Speybroeck L, O'dell S. Chandra X-Ray Observatory: Overview, X-Ray Optics, Instruments, and Missions. *Proc SPIE*. 2000:4012.
- Worsley K, Liao J, Aston C, Petre V, Duncan G, Morales F, Evans A. A General Statistical Analysis for fMRI data. *NeuroImage*. 2002; 15:1–15. [PubMed: 11771969]
- Worsley K, Marrett S, Vandal A, Friston K, Evans A. A Unified Statistical Approach for Determining Significant Statistical Signals in Images of Cerebral Activation. *Human Brain Mapping*. 1996; 4:58–73. [PubMed: 20408186]
- York DG, Adelman J, Anderson JE Jr, Anderson SF, Annis J, Bahcall NA, Bakken JA, Barkhouser R, Bastian S, Berman E, Boroski WN, Bracker S, Briegel C, Briggs JW, Brinkmann J, Brunner R, Burles S, Carey L, Carr MA, Castander FJ, Chen B, Colestock PL, Connolly AJ, Crocker JH, Csabai I, Czarapata PC, Davis JE, Doi M, Dombeck T, Eisenstein D, Elnan N, Elms BR, Evans ML, Fan X, Federwitz GR, Fiscelli L, Friedman S, Frieman JA, Fukugita M, Gillespie B, Gunn JE, Gurbani VK, de Haas E, Haldeman M, Harris FH, Hayes J, Heckman TM, Hennessy GS, Hindsley RB, Holm S, Holmgren DJ, hao Huang C, Hull C, Husby D, Ichikawa S-I, Ichikawa T, Ivezić Z, Kent S, Kim RSJ, Kinney E, Klaene M, Kleinman AN, Kleinman S, Knapp GR, Korienek J, Kron RG, Kunszt PZ, Lamb DQ, Lee B, Leger RF, Limmongkol S, Lindenmeyer C, Long DC, Loomis C, Loveday J, Lucinio R, Lupton RH, MacKinnon B, Mannery EJ, Mantsch PM, Margon B, McGehee P, McKay TA, Meiksin A, Merelli A, Monet DG, Munn JA, Narayanan VK, Nash T, Neilsen E, Neswold R, Newberg HJ, Nichol RC, Nicinski T, Nonino M, Okada N, Okamura S, Ostriker JP, Owen R, Pauls AG, Peoples J, Peterson RL, Petravick D, Pier JR, Pope A, Pordes R, Prosapio A, Rechenmacher R, Quinn TR, Richards GT, Richmond MW, Rivetta CH, Rockosi CM, Ruthmansdorfer K, Sandford D, Schlegel DJ, Schneider DP, Sekiguchi M, Sergey G, Shimasaku K, Siegmund WA, Smee S, Smith JA, Snedden S, Stone R, Stoughton C, Strauss MA, Stubbs C, SubbaRao M, Szalay AS, Szapudi I, Szokoly GP, Thakar AR, Tremonti C, Tucker DL, Uomoto A, Berk DV, Vogeley MS, Waddell Pi, Wang S, Watanabe M, Weinberg DH, Yanny B, Yasuda N. The Sloan Digital Sky Survey: Technical Summary. *The Astronomical Journal*. 2000; 120:1579–1587.

Appendix

The purpose of this appendix is to clarify the value added by the multi-scale derivative (MSD) filter discussed in the paper. We give two results related to the continuous version of the MSD. The first shows that the MSD brings out the Laplacian of the underlying function, suggesting that the MSD will be well suited to enhancing sharp localized features like the peaks produced by galaxies in typical astronomical surveys. The second result gives conditions under which the MSD enhances the contrast of a sharp, localized feature relative to background noise.

As in the text of the paper, let ϕ_h be the d -dimensional standard Normal density, and let ϕ_h be

the scaled version, $\phi_h(z) = h^{-d} \phi(z/h)$. Define $K_h = -\frac{\partial}{\partial h} \phi_h$ and note that

$$K_h(x) = -h \operatorname{trace} \varphi_h''(x) = \frac{1}{h} \left(d - \frac{\|x\|^2}{h^2} \right) \varphi_h(x), \quad (16)$$

where φ_h'' is the Hessian of φ_h with respect to x .

For the first result, we apply this filter to a function g :

$$(K_h \star g)(x) = \frac{1}{h} \int \left(d - \frac{\|u\|^2}{h^2} \right) \varphi_h(u) g(x-u) du \quad (17)$$

$$= \frac{1}{h} \int (d - \|v\|^2) \varphi(v) g(x-hv) dv \quad (18)$$

$$\equiv \frac{1}{h} \int \xi(v) g(x-hv) dv, \quad (19)$$

where for convenience, we have defined $\xi(v) = (d - \|v\|^2) \varphi(v)$ on $\mathbb{R}^d$.

For small h , we can expand

$$g(x-hv) = g(x) - hg'(x)^T v + \frac{h^2}{2} v^T g''(x) v + \frac{h^3}{6} g_x'''(v^T, v, v) + O(h^4), \quad (20)$$

where g_x''' is the third derivative tensor at x . We have that $\int v_i^2 \xi(v) dv = -2$, and by symmetry, $\int v_i^{2a+1} v_j^b v_k^c \xi(v) dv = 0$ for non-negative integers a, b , and c .⁶ Hence,

$$(K_h \star g)(x) = -h \operatorname{trace} g''(x) + O(h^3), \quad (21)$$

where $\operatorname{trace} g''$ is the Laplacian of g . This is the first result.

Now suppose that g has a maximum peak (or other feature) at some fixed x . Define

$$\operatorname{Contrast}(g, h) = \frac{1}{hg(x)} \int (d - \|v\|^2) \varphi(v) g(x-hv) dv, \quad (22)$$

where we suppress dependence on the fixed x for simplicity. For example, if $x = 0$ and $g = \frac{1}{a} e^{-\|x\|^2/a}$ for some $a > 0$,

$$\operatorname{Contrast}(g, h) = \frac{hd}{a^2} (1 + (h/a)^2)^{-\frac{d+2}{2}}. \quad (23)$$

We will use the Contrast function for such peaks to quantify the filter's performance.

⁶When the region is compact and symmetric, the third integral equal continues to hold while the first two hold to good approximation for small h .

Assume the data is of the form $Y = f + Z$, where the function f on $\mathbb{R}^d$ represents the signal (i.e., source profile) and Z be a mean-zero, Gaussian noise process with covariance function R , such that Z 's sample paths are almost surely smooth. Then, $K_h Y = K_h f + K_h Z$.

Power will be enhanced when the Contrast of f is larger than that of Z . Define

$$\tilde{Z}(x) = (K_h \star Z)(x) = \frac{1}{h} \int \xi(v) Z(x - hv) dv. \quad (24)$$

We have that

$$\text{Cov}(\tilde{Z}(x), \tilde{Z}(y)) = h^{-4} \int \int \xi(u) \xi(v) R(x - hu, y - hv) du dv. \quad (25)$$

When $R(a, b)$ takes the form $R(a, b) = \sum_k \lambda_k^2 \psi_k(a) \psi_k(b)$, then

$$\frac{\text{Var } \tilde{Z}(x)}{\text{Var } Z(x)} = \frac{\sum_k \lambda_k^2 \psi_k^2(x) \left(\frac{1}{h \psi_k(x)} \int \xi(u) \psi_k(x - hu) du \right)}{\sum_k \lambda_k^2 \psi_k^2(x)} \quad (26)$$

$$\leq \sum_k \frac{\lambda_k^2 \psi_k^2(x)}{\sum_j \lambda_j^2 \psi_j^2(x)} \text{Contrast}^2(\psi_k, h). \quad (27)$$

$$\leq \max_k \text{Contrast}^2(\psi_k, h). \quad (28)$$

For concreteness, assume that f has an isolated and approximately Gaussian peak of width s at 0. Assume also that Z has a squared-exponential covariance function with width ℓ . Then, $\tilde{Z}$ is approximately a shifted Gaussian with width $\kappa \sqrt{2}$, so it follows that

$$\frac{\text{Var } \tilde{Z}(x)}{\text{Var } Z(x)} \leq \text{Contrast}^2(\varphi_{\kappa \sqrt{2}}, h). \quad (29)$$

If $\ell = s/\kappa$, then $\text{Contrast}(f, h) \approx \text{Contrast}(Z, h)$ whenever $\ell \leq \sqrt{2}$, with $d = 2$. That is, sharp features in f are enhanced relative to the noise by the multi-scale filter. When $\ell > \sqrt{2}$, however, we can choose h so that the difference in contrasts from signal to noise is arbitrarily small. (Interestingly, for $d > 2$, the direction of this relation is reversed.)

By adaptive choice of h , as described in the text, we can improve power for the peaks (or other features) in f that are sharper than the correlation scale of the noise process, without significantly reducing power for other peaks. Because the faintest, smallest sources are typically of greatest interest in astronomical surveys, the multi-scale derivative should enhance sources at the margin of detection, as desired.

Figure 1.

The 45 pixels that have any overlap with the red circle are considered sources and black pixels indicate sources detected via some detection algorithm. Each of the three images have the same number of Type I and Type II error pixels. While all these pictures have the same number of pixel-wise Type I and Type II errors they lead to very different conclusions about the number, shape and location of the sources in the image, thus a testing criteria based on the aggregate sources, instead of pixels, is necessary.

Figure 2.

A 512 pixel by 512 pixel patch of the Chandra Deep Field South. The image was smoothed and then all pixels with count greater than 15 were set to 15. This image is the log of that image plus one. The transformations were performed so that the brightest sources do not wash out the fainter ones for display purposes. The transformed image was not used in any of the analysis. The complete data set is publicly available from <http://cxc.cfa.harvard.edu/cda/>

Figure 3.

CDFS data after applying the False Cluster Proportion procedure using the PP method to calculate U . Black indicates sources detected by FCP and the circles are detections from both the Main Catalog and Secondary Catalogs from GI. There are 17 detections from FCP all of which are in the Catalog and 9 missed sources.

Figure 4.

Left: A toy source with a background and a source point highlighted. A: The intensity of the background point after smoothing with different values of the smoothing parameter h . The intensity stays constant around zero until the source begins to get smoothed into the background, at which point we see a small increase in the intensity. B: The intensity of a point in the source after smoothing with different values of the smoothing parameter h . The intensity drops sharply for small amounts of smoothing and continues to decline for larger values of h . Thus estimates of the derivative of the source point with respect to h should give large negative values whereas the background point will not.

Figure 5.

Left: The galaxy clusters we want to detect. In practice these will be obscured by confounding radio signals Middle: The Wiener filtered image which combines information from three different frequencies in an attempt to pick out the signal from the galaxy clusters while eliminating the confounding signals. This filter has introduced a non-constant background to the image. Right: The Multi-scale Derivative image calculated from the Wiener filtered image recovers most of the sources we are trying to detect while getting rid of most of the background variation. Data courtesy of Sehgal et al. (2007)

Figure 6. Boxplots comparing the percent of real clusters detected for PP- FCP and MS-FCP across different noise conditions.

Figure 7. Logistic regressions for the high noise conditions show that MS-FCP is better at detecting faint sources than PP-FCP

Figure 8. CDFS data after applying MS-FCP procedure. Black indicates sources detected by FCP and the circles are detections in both the Main Catalog and the Secondary Catalogs from GI

Table 1

Purities averaged over 100 runs for each of PP-FCP and MS-FCP.

Mean Purity for Simulated Data		
Poisson Noise Added	Purity PP-FCP	Purity MS-FCP
None	0.995	0.984
= 1	0.996	0.984
= 2	0.997	0.982
= 5	0.997	0.977

Table 2

The number of times the percentage of falsely detected clusters exceeded our 10% bound over 100 runs for each of PP-FCP and MS-FCP.

Poisson Noise Added	PP-FCP	MS-FCP
None	0	0
= 1	0	0.01
= 2	0	0.01
= 5	0.01	0.01

Table 3

Summary of Detection Routines on CDFS

Comparison of Detection Procedures on CDFS			
Method	False Cluster Proportion Tolerance (c)	% of Main Catalog in GI recovered	# of New Verified Detections outside of GI
PP-FCP	10%	65.3%	0
MS-FCP	10%	92.3%	0
MS-FCP	20%	92.3%	2