A NEW SOURCE DETECTION ALGORITHM USING THE FALSE-DISCOVERY RATE

A. M. HOPKINS,[1,2] C. J. MILLER,[3] A. J. CONNOLLY,[1] C. GENOVESE,[4]
R. C. NICHOL,[3] AND L. WASSERMAN[4]

ABSTRACT

The false-discovery rate (FDR) method has recently been described by Miller et al., along with several examples of astrophysical applications. FDR is a new statistical procedure due to Benjamini & Hochberg for controlling the fraction of false positives when performing multiple hypothesis testing. The importance of this method to source detection algorithms is immediately clear. To explore the possibilities offered, we have developed a new task for performing source detection in radio telescope images, SFIND 2.0, which implements FDR. We compare SFIND 2.0 with two other source detection and measurement tasks, IMSAD and SExtractor, and comment on several issues arising from the nature of the correlation between nearby pixels and the necessary assumption of the null hypothesis. The strong suggestion is made that implementing FDR as a threshold-defining method in other existing source detection tasks is easy and worthwhile. We show that the constraint on the fraction of false detections as specified by FDR holds true even for highly correlated and realistic images. For the detection of true sources, which are complex combinations of source pixels, this constraint appears to be somewhat less strict. It is still reliable enough, however, for a priori estimates of the fraction of false source detections to be robust and realistic.

Key words: methods: data analysis — methods: statistical — techniques: image processing

1. INTRODUCTION

Detecting and measuring the properties of objects in astronomical images in an automated fashion is a fundamental step underlying a growing proportion of astrophysical research. There are many existing tasks, some quite sophisticated, for performing such analyses. Regardless of the wavelength at which an image has been made, however, each of these tasks has one thing in common: a threshold needs to be defined above which pixels will be believed to belong to real sources.

Defining an appropriate threshold is a complex issue and, owing to the unavoidable presence of noise, *any* chosen threshold will result in some true sources being overlooked and some false sources measured as real. Varying the chosen threshold to one extreme or the other will minimize one of these types of errors at the expense of maximizing the other. Clearly, choosing a threshold to jointly minimize both types of errors is not trivial, but even more problematic is that it is not even clear that one can, a priori, make a well-defined estimate of the magnitude of each type of error. Typically, this is done by comparing measured source counts with existing estimates for the expected number of sources. This is an unsatisfactory solution for surveys reaching to new sensitivity limits or at previously uninvestigated wavelengths (where there can be no estimate of the expected number of sources from existing studies) and also for small-area imaging, where the properties of large populations are affected by clustering or small number statistics.

Throughout this paper, we shall use the term "source pixel" to mean a pixel in an image that is above some threshold and, thus, assumed to be part of a true source. The term "source" shall be used to mean a contiguous collection of source pixels that corresponds to an actual astronomical object, a star or galaxy, for example, whose properties we are interested in measuring. In this work, we concentrate on radio images, specifically images produced by radio interferometers. We emphasize, though, that all the conclusions presented here are valid for any pixelized map where the "null hypothesis" is known at each pixel. Throughout this paper, the null hypothesis is taken to be the "sky background" at each pixel (after the image is normalized).

Now consider the following possible criteria applied to some chosen threshold: (1) that there be no falsely discovered source pixels; (2) that the proportion of falsely discovered source pixels be some small fraction of the total number of pixels (background plus source); (3) that the fraction of false positives (i.e., the number of falsely discovered sources over the total number of discovered sources) be small. The first of these can be achieved by using a very high threshold called the Bonferoni threshold (Miller et al. 2001). This threshold is rarely used, since although guaranteeing no false detections, it detects so few real sources. The second criterion is most often applied in astronomy and can be achieved by simply choosing the appropriate significance threshold. A threshold of 3 σ, for example, ensures that 0.1% of the total number of *pixels* are falsely discovered. Unfortunately, this is not the same as a constraint on the fraction of false detections compared with total number of *detections*. This quantity is a more meaningful measure to use in defining a threshold for the following reason: Consider a 3 σ threshold in an image composed of 10^6 pixels (1000 × 1000) and containing only Gaussian noise. This would yield, on average, 1000 pixels above the threshold. If real sources are also present, these 1000 pixels appear as false source pixels, and if it happens that only 2000 pixels

[1] Department of Physics and Astronomy, University of Pittsburgh, 3941 O'Hara Street, Pittsburgh, PA 15260; ahopkins@phyast.pitt.edu.
[2] Hubble Fellow.
[3] Department of Physics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213.
[4] Department of Statistics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213.

are measured as source pixels, then half the detections are spurious. If many more true source pixels are present, of course, this threshold may be quite adequate.

The third criterion defines a more ideal threshold. Such a threshold allows one to specify a priori the maximum number of false discoveries, on average, as a fraction of the total number of discoveries. Such a method should be independent of the source distribution (i.e., it will adapt the threshold depending on the number and brightness of the sources). The false-discovery rate (FDR) method (Benjamini & Hochberg 1995; Benjamini & Yekutieli 2001; Miller et al. 2001) does precisely this, selecting a threshold that controls the fraction of false detections. We have implemented the FDR technique in a task for detecting and measuring sources in images made with radio telescopes.

Radio images were chosen for the current analysis for several reasons, including previous experience at coding of radio source detection tasks, but also because the conservative nature of constructing many radio source catalogs allows the value of this method to be emphasized. Traditionally, radio source catalogs are constructed in a fashion aimed at minimizing spurious sources, accomplished by selecting a very conservative threshold, which is usually 5 σ or even 7 σ. This is partly driven by the difficulty of completely removing residual image artifacts such as sidelobes of bright sources, even after applying very sophisticated image reconstruction methods, and the desire to avoid classifying these as sources. Many of the issues surrounding radio source detection are described by White et al. (1997). While such a conservative approach may minimize false detections, it has the drawback of not detecting large numbers of real, fainter sources important in many studies of the submillijansky and microjansky populations, for example. An FDR-defined threshold may allow many more sources to be included in a catalog while providing a quantitative constraint to be placed on the fraction of false detections.

Table 1 briefly describes the algorithms used in some common radio source detection and measurement tasks. There is much similarity in the selection of tasks available within the two primary image analysis packages, AIPS[5] and MIRIAD,[6] as alluded to in the summaries given in this table, although the specifics of each such task certainly differ to a greater or lesser extent. In addition to these packages, AIPS++ also has an image measurement task, IMAGE-FITTER, similar in concept to the IMFIT tasks in AIPS and MIRIAD. A stand-alone task, SExtractor (Bertin &

Arnouts 1996), is used extensively in image analysis for detecting and measuring objects, primarily in images made at optical wavelengths. Given the flexibility of the SExtractor code provided by the many user-definable parameters, it is possible to use this task also for detecting objects in radio images.

In the analysis below, we compare the effectiveness of SFIND 2.0, which runs under the MIRIAD package, with that of IMSAD (also a MIRIAD task and which employs essentially the same algorithm as SAD and VSAD in AIPS) and SExtractor. Section 2 describes the operation of the SFIND 2.0 task and the implementation of the FDR method. Section 3 describes the Monte Carlo construction of artificial images on which to compare SFIND 2.0, IMSAD, and SExtractor, the tests used in the comparison, and their results. Subsequent implementation using a portion of a real radio image is also presented. Section 4 discusses the relative merits of each task and the effectiveness of the FDR method and raises some issues regarding the validity of the null hypothesis (i.e., the background level) and the correlation between neighboring pixels. Section 5 presents our conclusions, along with the strong suggestion that implementing FDR as a threshold-defining method in other existing source detection tasks is worthwhile.

2. SFIND 2.0

Here we describe the algorithm used by SFIND 2.0, which implements FDR for identifying a detection threshold. We include the version number of the task simply to differentiate it from the earlier version of SFIND, also implemented under MIRIAD, as the source detection algorithm is significantly different. Subsequent revisions of SFIND will continue to use the FDR thresholding method. The elliptical Gaussian fitting routine used to measure identified sources has not changed, however, and is the same as that used in IMFIT and IMSAD. An example of the use of the earlier version of SFIND can be found in the source detection discussion of Hopkins et al. (1999).

The first step performed is to normalize the image. A Gaussian is fitted to the pixel histogram in regions of a user-specified size to establish the mean and standard deviation, σ, for each region. Then for each region of the image, the mean is subtracted and the result is divided by σ. Ideally, this leaves an image with uniform noise characteristics, defined by a Gaussian with zero mean and unit standard deviation. In practice, the finite size of the regions used may result in some nonuniformity, although a judicious choice of size for these regions should minimize any such effect. We note that radio interferometer images often contain image artifacts such as residual sidelobes arising from the image

---

TABLE 1
Existing Radio Source Detection and Measurement Tasks

AIPS	MIRIAD	Short Description
IMFIT/JMFIT	IMFIT	Fits multiple Gaussians to all pixels in a defined area
SAD/VSAD/HAPPY	IMSAD	Defines " islands " encompassing pixels above a user-defined threshold, and fits multiple Gaussians within these areas
...	SFIND 2.0	Defines threshold using FDR to determine pixels belonging to sources, and fits those by a Gaussian

processing, sampling effects, and so on. With adequate sampling, these effects should be statistically random with zero mean and simply add to the overall image noise.

Next, the FDR threshold is calculated for the whole image. The null hypothesis is taken to be that each pixel is drawn from a Gaussian distribution with zero mean and unit standard deviation. This corresponds to the "background pixels." In the absence of real sources, each pixel has a probability p (which varies with its normalized intensity) of being drawn from such a distribution. In images known to contain real sources, a low p-value for a pixel (calculated under the assumption that no sources are present) is often used as an indicator that it is a "source pixel." The p-values for all N pixels in the image are calculated and ordered. The threshold is then defined by plotting the ordered p-values as a function of i/N (where N is the total number of pixels and i is the index, from 1 to N) and finding the p-value, p_{cut} say, corresponding to the last point of intersection between these and a line of slope α/c_N. Here α is the maximum fraction of falsely detected source pixels to allow, on average (over multiple possible instances of the noise), and $c_N = 1$ if the statistical tests are fully independent (the pixels are uncorrelated). If the tests are dependent (the pixels are correlated), then

$$c_N = \sum_{i=1}^{N} i^{-1} \ . \tag{1}$$

Since most radio images (and, indeed, astronomical images in general) show some degree of correlation between pixels but tend not to be *fully* correlated, i.e., the intensity of a given pixel is not influenced by that of *every* other pixel, we have chosen to take an intermediate estimate for c_N reflecting the level of correlation present in the image. This is related to the synthesized beam size, or point-spread function (PSF). If n is the (integer) number of pixels representing the PSF, we define $c_N = \sum_{i=1}^{n} i^{-1}$. This will be discussed further in § 3.1. A diagrammatic example of the threshold calculation is shown in Figure 1. It becomes obvious from this

figure that increasing or decreasing the value chosen for α corresponds to increasing or decreasing the resulting p-value threshold, p_{cut}, and the number of pixels thus retained as source pixels. The FDR formalism ensures that the average fraction of false source pixels will never exceed α. As described in Miller et al. (2001), this explanation of implementing FDR does not explain or justify the validity of FDR. The reader is referred to Miller et al. (2001) for a heuristic justification and to Benjamini & Hochberg (1995) and Benjamini & Yekutieli (2001) for the detailed statistical proof.

Finally, once the FDR threshold is defined, the pixels with $p < p_{cut}$, corresponding to source pixels, can be analyzed. Each of the source pixels are investigated in turn as follows: A hill-climbing procedure starts from the source pixel and finds the nearest local maximum from among the contiguous source pixels. From this local peak, a collection of contiguous, monotonically decreasing pixels are selected to represent the source. At this point, it is possible to either (1) use all the pixels around the peak that satisfy these criteria or (2) use only those that are themselves above the FDR threshold. The latter is the default operation of SFIND 2.0, but the user can specify an option for choosing the former method as well. The former method, which allows pixels below the FDR threshold to be included in a source measurement, may be desirable for obtaining more reasonable source parameters for sources close to the threshold. In either case, the resulting collection of source pixels is fitted in a least-squares fashion by a two-dimensional elliptical Gaussian. If the fitting procedure does not converge, the source is rejected. This is typically the case when the potential source consists of too few pixels to well constrain the fit. It is likely that most such rejections will be due to noise spikes, which typically contain a small number of pixels, although some may be due to real sources lying just below the threshold such that only a few true source pixels appear above it. If the fit is successful, the source is characterized by the fitted Gaussian parameters, and the pixels used in this process are flagged as already "belonging" to a source to prevent them from being investigated again in later iterations of this step. On completion, a source catalog is written by the task, and images showing (1) the pixels above the FDR threshold and (2) the normalized image may optionally be produced.

3. TASK COMPARISONS

To compare the effectiveness of SFIND 2.0, IMSAD, and SExtractor, 100 artificial images, 361×391 pixels in size, were generated. Each of these contained a different instance of random Gaussian noise and the same catalog of 72 point sources with known properties (position and intensity). The intensity distribution of the sources spans a little more than 2 orders of magnitude, ranging from somewhat fainter than the noise level to well above it. Many fewer bright sources were assigned than faint sources, in order to produce a realistic distribution of source intensities. The artificial images were convolved with a Gaussian to mimic the effects of a radio telescope PSF. The test images have $2''$ pixels, and the Gaussian chosen to represent the PSF has FWHMs of $11''73 \times 5''52$ with a position angle of $16°3$. The (convolved) artificial sources in the absence of noise can be seen in Figure 2, along with one of the test images.
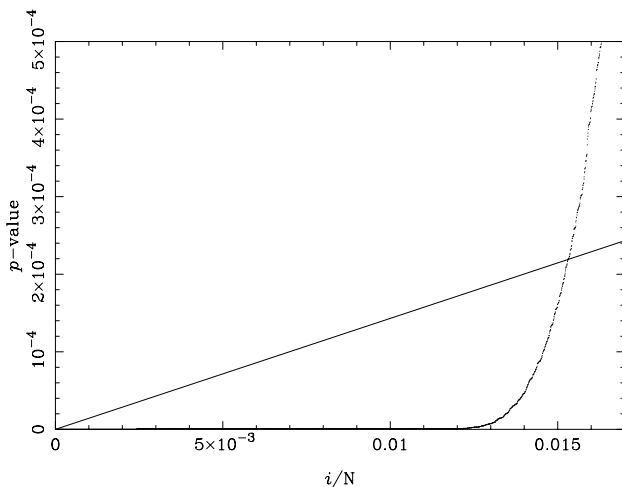

FIG. 1.—Graphical example of how the FDR threshold is calculated. This diagram shows only the relevant portion of the full graph (which has abscissa spanning 0–1). The points show p-values of the pixels from one of the artificial images. The line has a slope of α/c_N with $\alpha = 0.05$ and $c_N \approx 3.495$, corresponding to a PSF area covering 18 pixels. The last intersection point gives $p_{cut} = 2.19 \times 10^{-4}$. Hence, all pixels with $p < 2.19 \times 10^{-4}$ are considered "source pixels."
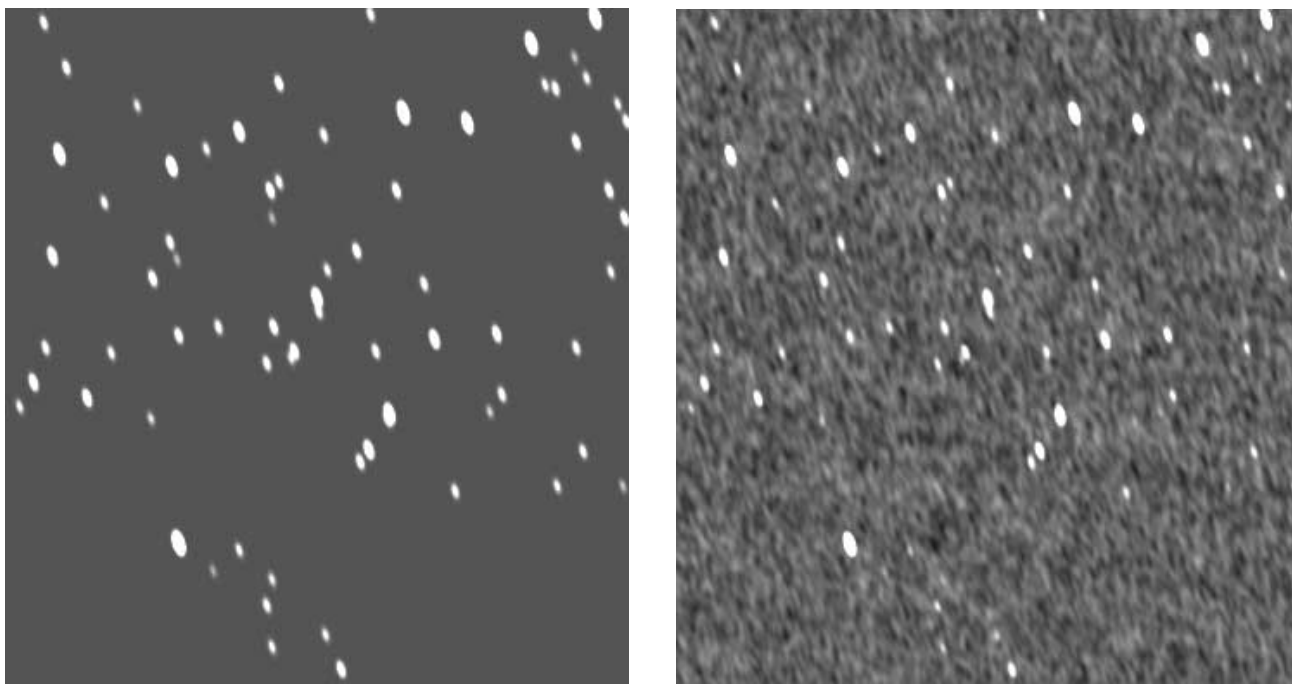
Fig. 2.—Example artificial images. *Left*, artificial sources only; *right*, artificial sources in the presence of noise as used in the simulations, emphasizing that real sources close to or below the noise level become difficult or impossible to detect, even visually.

On each image, SFIND 2.0 was run with $\alpha = 0.01$, 0.05, and 0.1, and the resulting lists of detected sources compared with the input catalog. By way of an example, the sources detected in a single test of SFIND 2.0 are shown in Figure 3, which indicates by an ellipse the location, size, and position angle of each detected source. This example demonstrates the difficulty of detecting faint sources and the ability of noise to mimic the characteristics of faint sources.

3.1. *Source Pixel Detection*

Miller et al. (2001) examined the simplest possible scenario, source pixels placed on a regular grid in the presence of uncorrelated Gaussian noise. Here we investigate a much more realistic situation. The "source pixels" now lie in contiguous groups comprising "real" sources, in the sense that the whole image (background and sources) has been convolved with a PSF. The number of pixels in each source above a certain threshold will vary depending on the intensity of the source. We now confirm that the FDR method works consistently on these realistic images.

To verify the reliability of the FDR defined threshold, the number of pixels detected above the FDR threshold in each test were recorded along with the number that were unassociated with any true source. The distribution of this fraction of false FDR pixels should never exceed the value specified for α, and this can be seen in the histogram in Figure 4. This figure also shows how the distribution of falsely detected pixels changes with the form chosen for c_N, emphasizing that the assumption of complete independence of the pixels is not justified (as expected), but neither is the image fully correlated, evidenced by the conservative level of false detections seen under this assumption. Our choice for the form of c_N, which is not in fact a result of the rigorous statistical proof, appears to be a feasible and reliable intermediate for such "partially correlated" images.

3.2. *Source Detection*

The FDR formalism ensures that the average fraction of falsely detected *pixels* will be less than α. The connection between numbers of pixels and numbers of sources is complex, however, and the same criterion cannot be said to be true for the fraction of falsely detected *sources*. The number of source pixels per source will vary according to both instrumental effects, such as the sampling and the PSF, and intrinsic source sizes compared to the instrumental resolution and the source brightnesses compared to the noise level in an image. Even if all sources are pointlike, and hence should appear in the image as a PSF, the number of source pixels above a given threshold for a given source depends on its brightness, and the number of source pixels per source would not be expected to be constant. To investigate the effect of this complex relation, we explore empirically the results of applying FDR thresholding to our simulated images. The fraction of falsely detected *sources* in each image, as well as the fraction of true sources not detected, are shown in Figure 5 as distributions for each tested value of α. By construction, a number of the artificial sources have intensities comparable to or lower than the noise level in the images, so not all sources will be able to be recovered in every image. This is reflected in the fact that somewhat more than 5% of sources are missed (by all tasks tested) even with very liberal thresholds.

The histograms in Figure 5a show that for $\alpha = 0.1$, where up to 10% of pixels could be expected to be false, the fraction of false sources is not much more. The result for $\alpha = 0.05$ is also quite reasonable, although for $\alpha = 0.01$ the outliers are further still, relatively speaking, from the expected fraction. While the strict constraint applicable to the fraction of false source pixels no longer holds for false sources, it still seems to be quite a good estimator. For the case in which only the peak pixel is required to be above the
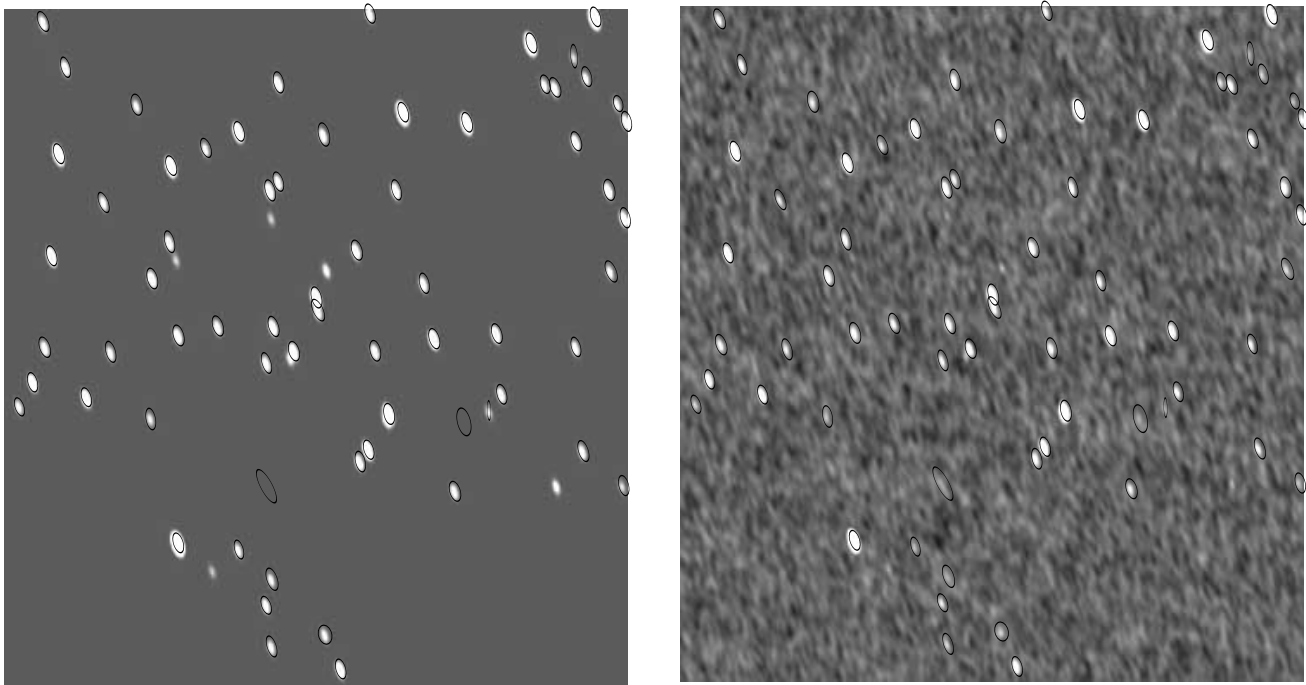
Fig. 3.—Example artificial images, showing objects detected as sources by SFIND 2.0. The noiseless image (*left*) is shown for reference to make it clear which objects have been correctly detected or missed in this instance.

FDR threshold, Figure 5c, the fractions of falsely detected sources are not so strongly constrained. For $\alpha = 0.1$, the fraction may be almost twice that expected. In both cases, although with greater reliability in the former, this allows the FDR method to provide an estimate of the fraction of false sources to expect. Even though the constraint may not be rigorous and clearly the estimate will be much more reliable in the case where all source pixels are required to be
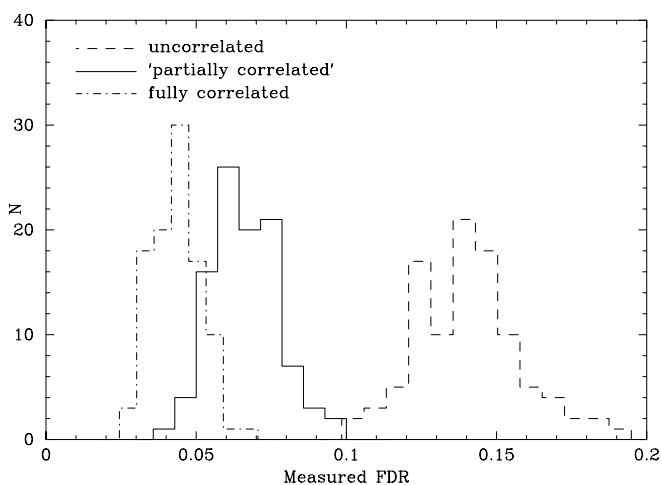

Fig. 4.—Fraction of falsely detected pixels for $\alpha = 0.1$, assuming different forms for c_N. For the uncorrelated assumption, $c_N = 1$. For fully correlated, $c_N = \sum_{i=1}^{N} i^{-1}$. For "partially correlated," $c_N = \sum_{i=1}^{n} i^{-1}$, where n represents the number of pixels covering the "correlation size" of the image, corresponding in our tests to the PSF size. For both the fully correlated and partially correlated cases, it can be seen empirically that $\langle \text{FDR} \rangle \leq \alpha$ (where the angle brackets represent an ensemble average over replications of the data).

above the FDR threshold, the FDR method allows a realistic a priori estimate of the fraction of false detections to be made. This feature is not possible with the simple assumption of, say, a 5 σ threshold.

To test IMSAD and SExtractor in the same fashion as SFIND 2.0, a choice of threshold value as a multiple of the image noise level (σ) was required. Simply selecting a canonical value of 3, 5, or 7 σ, for example, would complicate the comparison between these tasks and SFIND 2.0, since this would be testing not only different source measurement routines, but also potentially different thresholds. The values of α selected for testing SFIND 2.0 result in threshold levels that correspond approximately (since the noise level varies minimally from image to image) to 4.1, 3.6, and 3.3 σ. A 5 σ threshold in these simulations would correspond to a value of $\alpha \approx 0.0001$.

As a brief aside, it should be emphasized that this particular correspondence between a choice of α and a particular σ-threshold is only valid for the noise and source characteristics of the images used in the present simulations. For images with different noise levels or different source intensity distributions, any particular value of α will correspond to some different multiple of the local noise level. The primary advantage of specifying an FDR α value over choosing a 5 σ threshold, say, is that the FDR threshold is adaptive. The FDR threshold will assume a different value depending on the source intensity distribution relative to the background. This point is made very strongly in Miller et al. (2001) in the diagrams of their Figure 4. As a specific example in the context of the current simulations, we investigated additional simulated images containing the same noise as in the current simulations but containing sources having very different intensity distributions. We chose one intensity distribution such that every source was 10 times
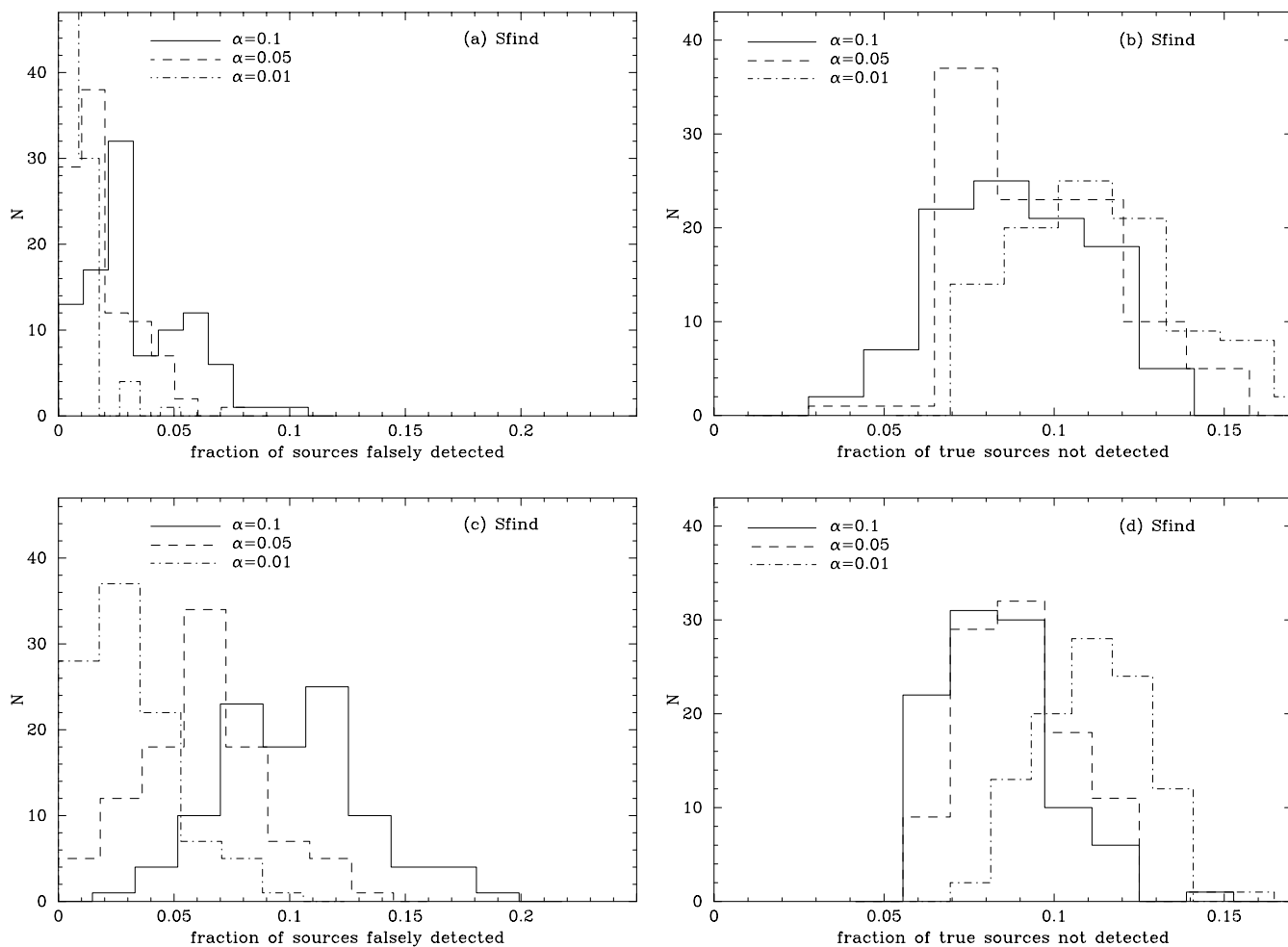
FIG. 5.—Comparison of different source detection codes. The fractions of sources falsely detected (*left*) or missed (*right*) are shown for each of the tested tasks. (*a, b*) Results for SFIND 2.0 in its default mode, where all source pixels are required to be above the FDR threshold; (*c, d*) results for SFIND 2.0 when only the peak pixel is required to lie above the FDR threshold; (*e, f*) fractions of falsely detected and missed sources for IMSAD; (*g, h*) those for SExtractor. The SExtractor results are based on setting a minimum requirement of 7 pixels for a source to be detected.

brighter than in the current simulations and one such that every source was 10 times fainter. In the brighter case, the FDR threshold for $\alpha = 0.01$, ensuring that on average no more than 1% of source pixels would be falsely detected, corresponded not to 4.1 σ, but to about 3.8 σ. The reason here is that as the source distribution becomes brighter, many more pixels will have low p-values. To retain the constant fraction of false pixels more background pixels must be included, so the threshold becomes lower. In the fainter case, where the artificial sources are very close in intensity to the noise level, the same FDR threshold corresponds to about 4.4 σ. Of course, in this case, very few sources are detected, for obvious reasons, but the same constraint on the fraction of falsely detected pixels applies. Here fewer pixels will have low p-values, and thus fewer background pixels are allowed, maintaining a constant fraction of false pixels, and the threshold increases. In the brighter case, the simple assumption of, say, a 4 σ threshold would give a lower rate of false pixels, while in the fainter case it would give a higher fraction. The importance of these examples is to emphasize that FDR provides a consistent constraint on the fraction of false detections in an adaptive way, governed by the source intensity distribution relative to the background, which cannot be reproduced by the simple assumption of, for example, a 4 σ threshold. While the source

distributions in most astronomical images typically lie between the two extremes presented for this illustration, the adaptability of the FDR thresholding method still presents itself as an important tool.

Returning now to the comparison between SFIND 2.0, IMSAD, and SExtractor, the IMSAD and SExtractor thresholds were set to correspond to those derived from the α-values used in SFIND 2.0. The distributions of falsely detected and missed sources were similarly calculated. These are shown in Figures 5e–5h. One of the features of SExtractor is the ability to set a minimum number of contiguous pixels required before a source is considered to be real, and obviously the number of detected sources varies strongly with this parameter. After some experimentation, we set this parameter to 7 pixels, since this resulted in a distribution of false detections most similar to that seen with SFIND 2.0, for the case where only pixels above the FDR threshold are used in source measurements. Values larger than seven reduced the number of false detections at the expense of missing more true sources and vice versa. From this comparison, SFIND 2.0 appears to miss somewhat fewer of the true sources than SExtractor when SExtractor is constrained to the same level of false detections. This is also true if SExtractor is constrained to a similar distribution of false detections as obtained by SFIND 2.0 for the case where
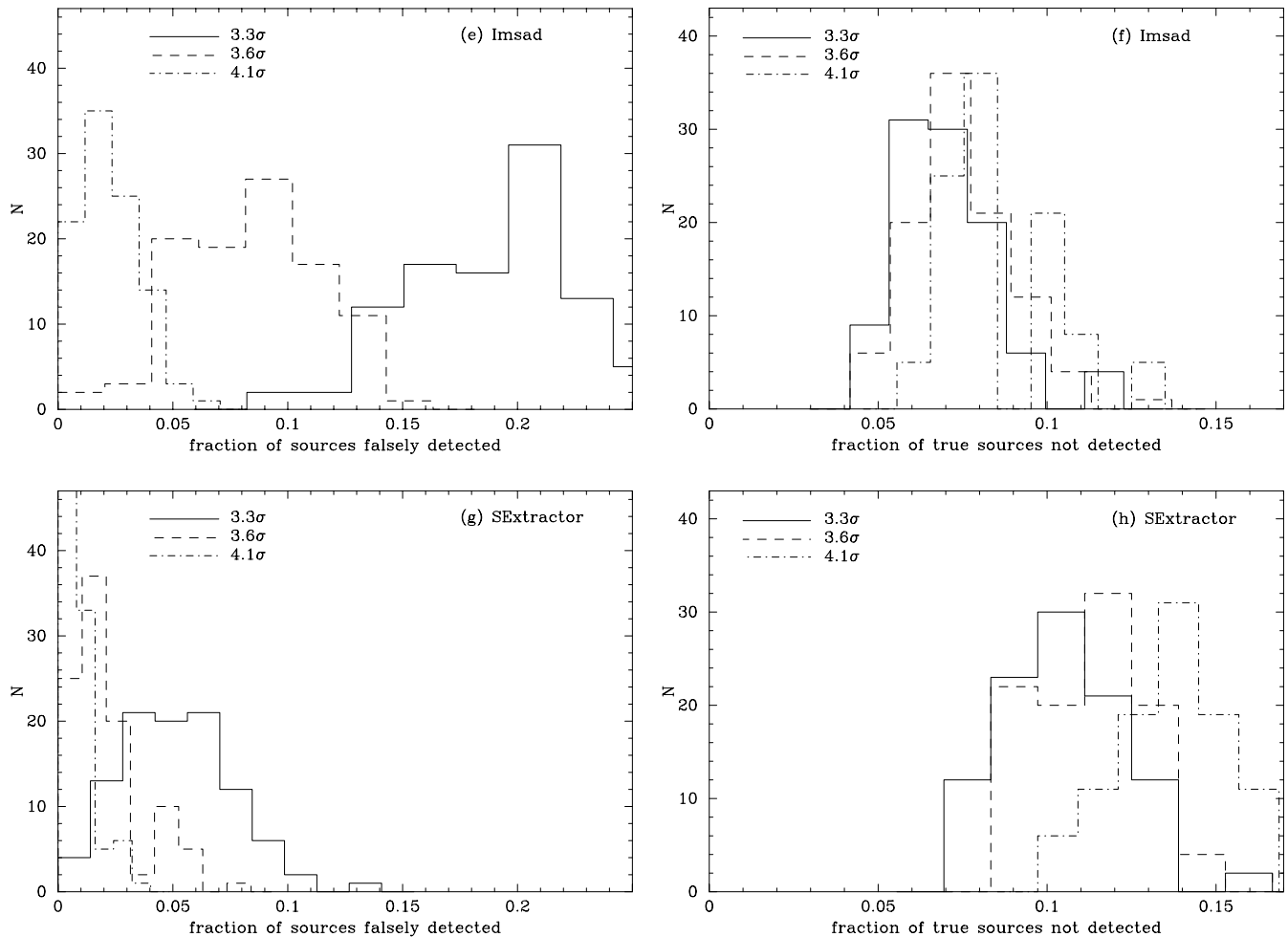
Fig. 5.—*Continued*

only the peak pixel is required to be above the FDR threshold (corresponding to four contiguous pixels). In both cases, allowing SExtractor to detect more true sources by lowering the minimum pixel criterion introduces larger numbers of false detections. SFIND 2.0 also performs favorably compared with IMSAD. While SFIND 2.0 seems to miss a few percent more real sources than IMSAD, IMSAD seems to detect many more false sources than SFIND 2.0 in either of its source measurement modes.

Of primary importance in source measurement is the reliability of the source parameters measured. Figure 6 shows one example from the 100 tests comparing the true intensities and positions of the artificial sources with those measured by SFIND 2.0. Similar results are obtained with IMSAD, which uses the same Gaussian fitting routine. As expected, the measured values of intensity and position become less reliable as the source intensity becomes closer to the noise, although they are still not unreasonable. A comprehensive analysis of Gaussian fitting in astronomical applications has been presented by Condon (1997), and the results of the Gaussian fitting performed by SFIND 2.0 (and IMSAD) are consistent with the errors expected. In addition, the assumption that a source is pointlike, or only slightly extended, and thus well fitted by a two-dimensional elliptical Gaussian is clearly not always true. Complex sources in radio images, as in any astronomical image, present difficulties for simple source detection algorithms such as the ones investigated here. It is not the aim of the

current analysis to address these problems, except to mention that the parameters of such sources measured under the point-source assumption will suffer from much larger errors than indicated by the results of the Gaussian fitting.

As a final test, SFIND 2.0 was used to identify sources in a real radio image, a small portion of the Phoenix Deep Survey (Hopkins et al. 1999). This image contains sources with extended and complex morphologies, as well as point sources. Figure 7 shows the results, with SFIND 2.0 reliably identifying point-source and extended objects, as well as the components of various blended sources and complex objects.

4. DISCUSSION

The main aims of this analysis have been to (1) investigate the implementation of FDR thresholding to an astronomical source detection task and (2) compare the rates of missed and falsely detected sources between this task and others commonly used. Implementation of the FDR thresholding method is very straightforward, (evidenced by the seven-step IDL example of Miller et al. 2001, in their Appendix B). The FDR method performs as expected in providing a statistically reliable estimate of the fraction of falsely detected *pixels*. Performing source detection on a set of pixels introduces the transformation of pixel groups into sources. This ultimately results in the strong constraint on the false fraction of FDR-selected pixels becoming a less rig-
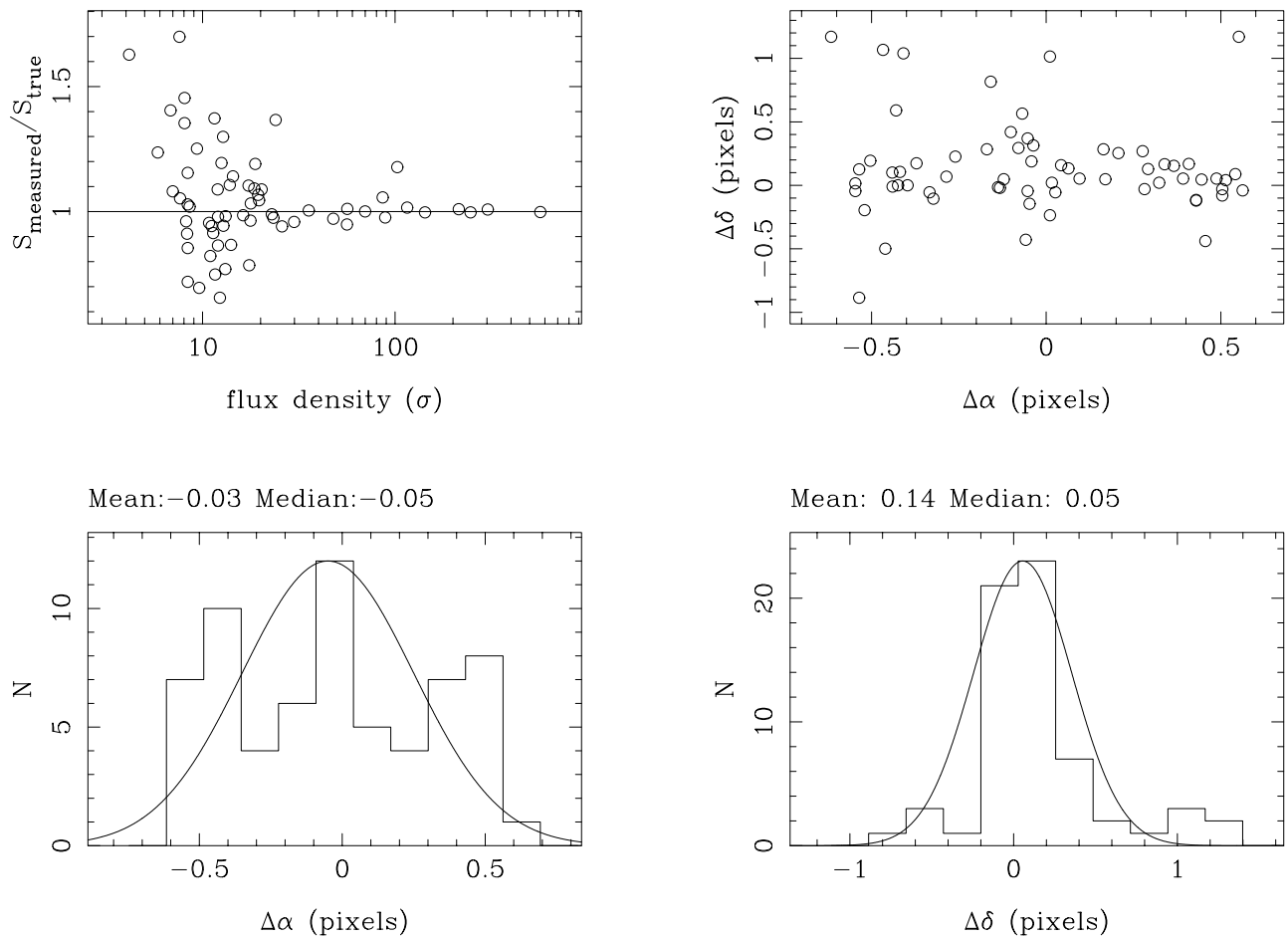
FIG. 6.—Comparison of measured source parameters with the true values. *Top left*, ratio of measured to true flux density as a function of the flux density; *top right*, scatter plot of position errors; *bottom left*, distribution of errors in right ascension; *bottom right*, distribution of errors in declination. The Gaussians over the two histograms have $\sigma = 0.35$ pixels and are indicative rather than fitted to the distributions.

orous, but still useful and reliable, estimate of the fraction of false sources. As already mentioned, this is still a more quantitative statement than can be made of the rate of false sources in the absence of the FDR method. It is possible that rigorous quantitative constraints on the fraction of false source detections may be obtained empirically for individual images or surveys. By performing Monte Carlo source detection simulations with artificial images having noise properties similar to the ones under investigation, the trend of falsely detected sources with α may be able to be reliably characterized. This has not been examined in the current analysis, but will be included in subsequent work with SFIND 2.0. While constraints on the fraction of falsely detected sources may be possible, neither FDR nor any other thresholding method provides constraints on the numbers of true sources remaining undetected.

A study is also ongoing into whether a more sophisticated FDR thresholding method for defining a source may be feasible. This would involve examining the combined size and brightness properties of groups of contiguous pixels to define a new *p*-value. This would represent the likelihood that such a collection of pixels comes from a " background distribution " or null hypothesis, corresponding to the properties exhibited by noise in various types of astronomical images. Using this new *p*-value, an FDR threshold, now in the size-brightness parameter space, could be applied for

defining a source catalog. Clearly, much care will need to be taken to avoid discriminating against true sources that may lie in certain regions of the size-brightness plane, such as low surface brightness galaxies. The assumption in many existing source detection algorithms that sources are point-like already suffers from such discrimination though, so even if such bias is unavoidable, some progress may still be achievable.

The form assumed for c_N in this analysis is not in fact a rigorous result of the formal FDR proof. Instead, it is a " compromise " estimate that seems mathematically reasonable and gives reliable results in practice. To be strictly conservative, the form of c_N given by equation (1) should be adopted to ensure that the fraction of falsely detected " source pixels " is strictly less than α. This rigorous treatment, however, is dependent on the number of pixels present in the image. Now consider analysis of a subregion within an image. As the size of this subregion is changed, the number of pixels being considered similarly changes, and this will have the effect of changing the threshold level and the resulting source catalog. This perhaps nonintuitive aspect of the FDR formalism is the adaptive mechanism that allows it to be rigorous in constraining the fraction of false detections.

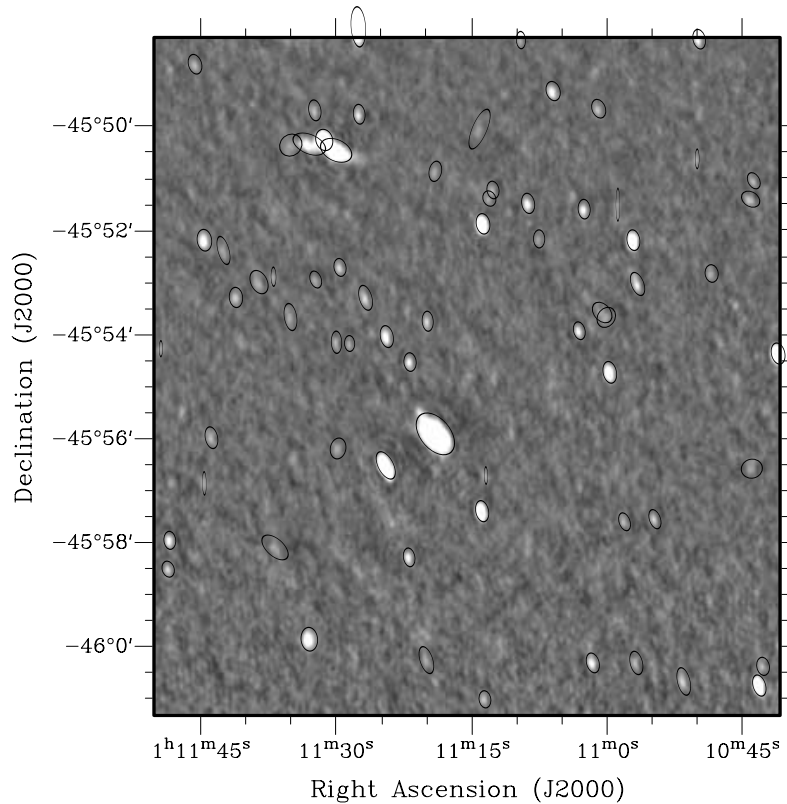There are additional complicating factors that must be taken into account when performing source detection. The

Fig. 7.—Example of a real radio image (a portion of the Phoenix Deep Survey). The marked sources have been detected using SFIND 2.0 and an FDR threshold corresponding to $\alpha = 0.01$. There are 69 detected sources, and the FDR method suggests that less than one of these are likely to be falsely detected.

null hypothesis assumed for the FDR method (and indeed for all the source detection algorithms) is that the background pixels have intensities drawn from a Gaussian distribution (or other well-characterized statistical distribution such as a Poissonian). This is not strictly true for radio images, where residual image processing artifacts may affect the noise properties, albeit at a low level. In all cases, such deviations will result in a larger fraction of false pixels than expected, some of which may be clumped in a fashion sufficient to mimic and be measured as sources, thus increasing the fraction of falsely detected sources. This comment is simply to serve as reminder to use caution when analyzing images with complex noise properties.

5. CONCLUSIONS

We have implemented the FDR method in an astronomical source detection task, SFIND 2.0, and compared it with two other tasks for detecting and measuring sources in radio telescope images. SFIND 2.0 compares favorably with both in the fractions of falsely detected sources and undetected true sources. The FDR method reliably selects a threshold that constrains the fraction of false pixels with respect to the total number of "source pixels" in realistic images. The fraction of falsely detected sources is not so strongly constrained, although quantitative estimates of this fraction are still reasonable. More investigation of the relationship between "source pixels" and "sources" is warranted to determine whether a more rigorous constraint can be established. With the ability to quantify the fraction of false detections provided by the FDR method, we strongly recommend that it is worthwhile implementing as a threshold-defining method in existing source detection tasks.

REFERENCES

Benjamini, Y., & Hochberg, Y. 1995, J. R. Stat. Soc. B, 57, 289
Benjamini, Y., & Yekutieli, D. 2001, Ann. Stat., in press
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Condon, J. J. 1997, PASP, 109, 166

Hopkins, A., Afonso, J., Cram, L., & Mobasher, B. 1999, ApJ, 519, L59
Miller, C. J., et al. 2001, AJ, 122, 3492
White, R. L., Becker, R. H., Helfand, D. J., & Gregg, M. D. 1997, ApJ, 475, 479