MADX – a simple technique for source detection and measurement using multiband imaging from the Herschel–ATLAS survey

S. J. Maddox$^{1,2}$ and L. Dunne$^{1,2}$

$^1$School of Physics and Astronomy, Cardiff University, Queen Buildings, The Parade, Cardiff CF24 3AA, UK
$^2$Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

Accepted 2020 February 12. Received 2020 February 11; in original form 2019 June 18

ABSTRACT

We describe the method used to detect sources for the Herschel–ATLAS (Herschel Astrophysical Terahertz Large Area Survey) survey. The method is to filter the individual bands using a matched filter, based on the point spread function (PSF) and confusion noise, and then form the inverse variance-weighted sum of the individual sources, including weights determined by a chosen spectral energy distribution. Peaks in this combined image are used to estimate the source positions. The fluxes for each source are estimated from the filtered single-band images, interpolated to the exact subpixel position. We test the method by creating simulated maps in three bands with PSFs, pixel sizes, and Gaussian instrumental noise that match the 250, 350, and 500 μm bands of Herschel–ATLAS. We use our method to detect sources and compare the measured positions and fluxes to the input sources. The multiband approach allows reliable source detection a factor 1.2–3 lower in flux compared to single-band source detection, depending on the source colours. The false detection rate is reduced by a factor between 4 and 10, and the variance of the source position errors is reduced by about a factor 1.5. We also consider the effect of confusion noise and find that the appropriate matched filter gives a further improvement in completeness and noise over the standard PSF filter approach. Overall the two modifications give a factor of 1.5–3 improvement in the depth of the recovered catalogues compared to a single-band PSF filter approach.

Key words: methods: data analysis – techniques: image processing; software: data analysis.

1 INTRODUCTION

There are many well-known algorithms to detect sources in imaging data, from simple identification of connected pixels above a threshold (Irwin 1985; Bertin 1996), through matched filtering (Stetson 1987; Tegmark & de Oliveira-Costa 1998; Herranz et al. 2002) to wavelet techniques (Vielva et al. 2003; González-Nuevo et al. 2006, Grumitt et al. 2020). These have generally been developed with single passbands in mind, but recently the increase in availability of multiwavelength data has spurred the development of techniques that make optimal use of several passbands. This has been particularly useful for submillimetre data (Naselsky, Novikov & Silk 2002; Herranz et al. 2005, Lanz et al. 2010; Planck Collaboration Int. LIV. 2018).

This paper describes the method that was used to detect sources for the Herschel Astrophysical Terahertz Large Area Survey, hereafter H-ATLAS (Eales et al. 2010; Rigby et al. 2011; Valiante et al. 2016, Maddox et al. 2018). The H-ATLAS is based on observations in the 100, 160, 250, 350, and 500 μm bands of the Herschel Space Observatory$^1$ (Pilbratt et al. 2010), which provide maps, covering ∼600 sq deg of sky in the five bands. The unprecedented depth of the Herschel data and the desire to have a blind far-infrared selected survey meant that we could not rely on data from other surveys to identify sources; the source detection had to be based on the Herschel maps alone. Also, the depth of the Herschel data mean that the source density is high, and the maps are significantly affected by source blending and confusion, so standard methods do not perform well.

It is fairly straightforward to show that the optimal way to detect an isolated point source in a map with a uniform background and simple Gaussian noise, is to filter the data with the point spread function (PSF) and find the peak in the filtered map (e.g. North 1943, Pratt 1978; Stetson 1987; Kay 1998). The value of the peak is equivalent to a least-squares fit of the PSF to the data at the position of the peak, and provides the minimum variance flux estimate of the source. Our method is based on this matched filter approach, but

---

* E-mail: maddoxs@cardiff.ac.uk

© 2020 The Author(s)
Published by Oxford University Press on behalf of the Royal Astronomical Society

---

$^1$Herschel is an ESA space observatory with science instruments provided by European-led Principal Investigator consortia and with important participation from NASA.
includes significant improvements: namely that the matched filter includes the effect of confusion noise; the application of the filter includes a locally defined noise weighting; that several bands can be combined in an optimal way to maximize the efficiency of detecting sources; and that the fluxes are estimated sequentially to reduce the effects of source blending. Simply detecting images in each band individually and merging the catalogues is not the optimal way to construct a combined catalogue, and combining multiwavelength data in an optimal way enhances the source detection reliability and automatically produces a band-matched catalogue. There has been extensive research in this area, considering correlated noise between bands, variable source sizes, and different spectral behaviour, as reviewed by Herranz, Argüeso & Carvalho (2012). We developed our method to find sources in the H-ATLAS survey, where data are available in five bands, with different angular resolution, and each with spatially varying noise. Note that the spatially non-uniform noise distributions mean that it is not a good approximation to assume simple Gaussian noise with known power spectra and cross-correlation functions. This means that methods such as the matched multi-filter approach of Lanz et al. (2010) are not directly applicable to the H-ATLAS data.

The next four sections in this paper describe the steps in the detection and extraction process: in Section 2, we estimate and subtract a non-uniform background; in Section 3, we filter the map for each waveband; in Section 4, we combine the wavebands and detect sources; and in Section 5, we estimate the source positions and fluxes. Then, in Section 6, we describe simulations which show the improvements of our method compared to single-band PSF filtered catalogues.

## 2 Background Estimation

The first step in detecting sources is to estimate the background, which may be spatially varying. A background may be instrumental or a real astronomical signal which is a contamination to the point sources we wish to extract. In the H-ATLAS, the background was largely local ‘cirrus’ from dust emission in our Galaxy. In general, it is impossible to differentiate between multiple confused sources and a smoothly varying background (and/or foreground) component, but in either case, it is necessary to remove the contribution of the background flux from each individual source. So, we need to determine the local background at all relevant positions in the map. We have done this by splitting the map into blocks of pixels corresponding to $10 \times 10$ pixels. For an isolated source on a uniform background with uniform noise distributions mean that it is not a good approximation to assume simple Gaussian noise with known power spectra and cross-correlation functions. This means that methods such as the matched multi-filter approach of Lanz et al. (2010) are not directly applicable to the H-ATLAS data.

The background at each pixel is then estimated using a bi-cubic interpolation between the coarse grid of backgrounds, and subtracted from the data. This approach to background estimation is similar to the \texttt{nebuliser} algorithm,² developed by the Cambridge Astronomical Survey Unit. After the initial analysis of the science-verification data for the H-ATLAS survey (Rigby et al. 2011), we used \texttt{nebuliser} to perform the background subtraction rather than the inbuilt background subtraction (Valiante et al. 2016; Maddox et al. 2018). This choice was largely based on the much faster run-time of \texttt{nebuliser} compared to the MADX version. The catalogues described in his paper use the built-in MADX background subtraction.

## 3 Filtering

Typically image data are sampled finely enough that point sources have 2 or 3 pixels within the FWHM in each direction. This means that the flux from a source is spread over many pixels, and the optimal estimate of the source flux is given by a weighted sum over pixels. For an isolated source on a uniform background with uniform Gaussian errors, the minimum variance estimate of the source flux is given by the sum of the data weighted by the PSF at the true position of the source. Cross-correlating the data by the PSF gives the PSF-weighted sum at the position of each pixel, and choosing the peak in the filtered map gives the minimum variance estimate of the source position and source flux (see e.g. Stetson 1987).

²http://casu.ast.cam.ac.uk/surveys-projects/software-release/background-filtering

![Figure 1. Simulated histogram of pixel values in a background block. The model has a true background of 10 mJy with 6 mJy Gaussian noise and a single 1 Jy source in the centre. The red line is the best-fitting Gaussian. The fitted peak is 10.3 mJy (red dashed line), the median is 11.3 mJy (blue dashed line), and the mean is 42 mJy (green dashed line). The dotted blue lines are the $\pm 2\sigma$ from the median.](https://academic.oup.com/mnras/article-abstract/493/2/2363/5739926)
If the pixel uncertainties vary spatially, the optimal weighting must also include the inverse of the estimated variance, as derived by Serjeant et al. (2003). If the power spectrum of the noise is not flat, the optimal filter is different from the PSF. In particular, when the source density is high, confusion noise is important, and the optimal filter is narrower than the PSF. The optimal filter, \( Q \), can be estimated using a matched filter approach that includes confusion noise (Chapin et al. 2011). In this case, the noise-weighted filtered map, \( F \), is given by

\[
F = \frac{(DW) \otimes Q}{W \otimes PQ},
\]

where \( D \) is the background subtracted data in each pixel, the weight \( W = 1/\text{var}(D) \) is the inverse of the variance of each pixel, \( P \) is the PSF, and \( \otimes \) represents the cross-correlation operator. Assuming that the instrumental noise on each pixel in the unfiltered map is uncorrelated, the variance of each pixel in the filtered map is given by

\[
V = \frac{W \otimes Q^2}{(W \otimes PQ)^2}.
\]

This is the generalization of the PSF-filtering approach derived by Serjeant et al. (2003); setting \( Q = P \) yields exactly the Serjeant et al. results. The noise weighting in this step is particularly important for the H-ATLAS data, since the noise varies dramatically on small angular scales depending on the number of detector passes a particular sky position has (Maddox et al. 2018).

The filtered map gives the minimum variance estimate of the flux that a source would have at any given pixel in the map. Standard Fourier transform routines allow easy calculation of the filtered maps at integer pixel positions, but in practice, sources are not centred on pixels. In order to find the best flux estimates, we need to allow for subpixel positioning. Without this, the fluxes will be significantly underestimated, particularly when a source lies at the edge of a pixel. Our approach to solving this problem is discussed in Section 5.2.

## 4 COMBINING WAVEBANDS

The filtered map in a single band provides an estimate of source flux and uncertainty at any position, and this approach can be extended to include any other wavebands that are available. If we know the observed spectral energy distribution (SED), of a source, \( S(\lambda) \), then the flux in a band with response \( R(\lambda) \), is given by

\[
F = \int \frac{S(\lambda)R(\lambda)d\lambda}{\int R(\lambda)d\lambda}.
\]

where we assume the detector measures total energy, as in a bolometer, not the total count of photons as in a CCD. We define the normalized SED as \( S_0(\lambda) \) where

\[
S_0(\lambda) = \frac{S(\lambda)}{\int S(\lambda)d\lambda},
\]

so the observed SED of the source is \( A S_0(\lambda) \), where \( A = \int S(\lambda)d\lambda \). Given a set of filter passbands, \( R_k \) and the source SED, the true broad-band flux in each band is \( F_{k0} = AF_{k0} \), where

\[
F_{k0} = \int \frac{S_0(\lambda)R_k(\lambda)d\lambda}{\int R_k(\lambda)d\lambda}.
\]

Since the value of \( A \) does not depend on wavelength, the filtered map in each band gives an independent estimate of \( A \). In order to combine the maps, we need to have the estimates at exactly the same position. As discussed in Section 5.2, it is reasonable to use a bi-cubic interpolation to estimate the source flux at non-integer pixel positions. If we interpolate the lower resolution maps to the pixel centres of the highest resolution map, then we can take the inverse variance-weighted sum to obtain the minimum variance estimate of \( A \) at the pixel positions of the highest resolution map. For waveband \( k \) the estimate of \( A \) at position \( x \) is \( A_k = F_k(x)/F_{k0} \), and the variance is \( \sigma^2_{A_k} = V_k(x)/F_{k0}^2 \). Hence, the overall minimum variance estimate of \( A_{tot} \) is given by

\[
A_{tot} = \sum_k \frac{F_k}{\sigma^2_k},
\]

and the uncertainty on \( A_{tot} \) is given by

\[
\sigma^2_A = \frac{1}{\sum_k \frac{1}{\sigma^2_k}}.
\]

So, the significance of a source detection at any position is \( A_{tot}/\sigma_A \). This is a very simple derivation of the standard result first presented by Naselsky et al. (2002). As for the case of a filtered map in a single band, we estimate the most likely position of the source as the position of the peak in the combined significance map.

Note that these formulae include a factor \( F_{k0}^2 \) as part of the weight given to the waveband \( k \), so that the true SED acts as a weighting term for each band as well as the inverse variance factor. This makes intuitive sense: if a source’s flux is expected to peak in a particular band, we should give that band the most weight in determining the position of the source; if a source has a flat spectrum, so that the flux is equal in all bands, then all bands are given equal weight. In general, we do not actually know the true SED of any particular source, and clearly do not know the SED of sources that we have not yet detected. However, we can maximize the detection rate of a particular type of source by choosing an SED prior to match.

This derivation considers an isolated source, but we can filter and combine the full area of available data to produce a global significance map, and find all the peaks to consider as potential sources. For H-ATLAS, we kept those that are more than 2.5\( \sigma \) as potential sources. In principle we could retain all peaks, but rejecting the low-significance peaks gives a large saving in computing time, while not losing any significant detections.

## 5 SOURCE PARAMETERS

### 5.1 Estimating positions and fluxes

To estimate the position of each source, we perform a variance weighted least-squares fit of a Gaussian to the 5 \( \times \) 5 pixels around each peak. The position of the peak is allowed to vary freely, and is not constrained to be at integer pixel positions. We fit only to pixels near the peak to minimize the effects of confusion from other nearby sources. Since the individual maps have been filtered, the peak pixels already include data from the surrounding raw pixels, combined in an optimal way; the peak fitting is solely to find the position at the sub-pixel level.

In order to estimate the flux in each band, we use the individual filtered maps, and interpolate to find the value of each map at the position of the peak in the combined map. For an isolated source, this will provide the optimal flux estimates. However, if there are sources that are close together, so that the PSFs significantly overlap, this simple approach will ‘double-count’ flux, because the wings of each source add on to the peak of its neighbour. A simple way to avoid this problem is to sort the sources in order of decreasing...
flux based on their initial peak-pixel value, and then estimate the optimal fluxes in sequence. After getting the optimal fluxes for a source, we subtract the scaled filtered PSF from the maps before estimating the fluxes for the next source. This is similar in concept to the clean algorithm (Högbom 1974) but with just one pass. This process is done separately for each band, so the ‘clean’ works from the brightest sources in each band. In principle, the procedure could be iterated to a stable solution, but in practice the difficult cases are blends of sources that require a more sophisticated deblending technique to improve the flux estimates. So iterating this simple clean algorithm would provide a very small gain in reliability at a large computational cost.

To provide the uncertainty for each flux measurement, we create a map of the filtered noise, using equation (2) and perform the same interpolation to estimate the flux variance in each band at each source position.

5.2 Subpixel subtleties

The above analysis ignores the complication that our data typically sample the sky with only 2 or 3 pixels across the FWHM of the PSF. When the PSF is sampled into coarse pixels, the value of the peak pixel is averaged over the whole area of the central pixel, and so is suppressed relative to the peak of the true PSF. For a PSF that is close to a Gaussian with 3 pixels across the FWHM, this suppression is typically ~5 percent. Since we use the pixelated PSF (the Point Response Function – or PRF) when filtering the data, the filtered data are boosted by the suppression factor, and so the estimated flux for source that is centred in a pixel is unbiased. For a source that is not centred in a pixel, the observed peak value is suppressed compared to a pixel centred source.

In order to obtain the most accurate estimates of source flux for an arbitrary position, we need to consider the true flux distribution within the footprint of each pixel. An obvious way to improve on the flux estimated from individual pixel values is to interpolate between them. However, a bilinear interpolation does not remove the bias, as can be seen by considering a source that is exactly half-way between two pixels; each pixel will have the same value that is less than the true peak value, so the interpolated value will also be biased low. A bicubic interpolation allows the interpolated value to be higher than either individual pixel, and so gives a better flux estimate.

Since we estimate the source positions by fitting a Gaussian to the central 5 × 5 pixels of each source, the positions are not directly affected by the pixelization; for signal-to-noise ratio greater than 20 we find that the source positions are accurate to better than 1/12 of a pixel (see Fig. 7). So, to estimate the flux of the source, we use the filtered maps interpolated to the best-fitting subpixel position. If the source position lies at the centre of a pixel, the interpolation returns that pixel value, and this will be an unbiased flux estimate. If the source position lies at the boundary between two pixels, the pixel values are suppressed relative to the peak of the pixel-centred PSF, but the bicubic interpolation means that the estimated flux will be higher than the pixel values, thus reducing the suppression. We tested this effect by creating simulated data at higher resolution, averaging over the small pixels to produce a low-resolution data, and then measuring the recovered flux from the interpolated, filtered low-resolution data. We find the interpolation reduces the suppression due to pixelization, but does leave a slight underestimate of the actual peak. The fractional flux error as a function of position is shown in Fig. 2. The error is zero at the centre of a pixel, and smoothly increases towards the pixel edges. It is largest for a source at the corner of 4 pixels when the flux

is underestimated by ~1 percent. Although the simulations are specific to a simple Gaussian PSF, this is a good approximation to the H-ATLAS data. In fact, the PSF in most astronomical data can be approximated by a Gaussian near the peak, and the pixel scales are typically chosen to sample the FWHM at a similar spacing, and so similar improvements are likely for other data.

6 TESTS OF THE METHODS

6.1 Simulations

As a simple test of the source detection algorithm we generated maps covering 3.4 × 13.6°, in three bands: 250, 350, and 500 μm. This is equivalent to a single one of the three H-ATLAS equatorial fields (Valiante et al. 2016). Sources are placed on a grid of positions separated by 3 arcmin on the sky with a small blank area around the edges of the maps, leading to 17 750 sources in the maps. Each source is assigned a small random offsets from the exact grid centre, so that the pixels sample the PSF with random offsets from the pixel centres. The PSF for each band is chosen to be a Gaussian with FWHM of 18, 24, and 36 arcsec, chosen to match the H-ATLAS maps. Each source is given a 250 μm flux between 1 mJy and 1 Jy, uniformly spaced in log flux. This is clearly not a good match to the flux distribution of real sources, but is a simple way to provide good statistics over the full flux range. The 350 and 500 μm fluxes for each source are then assigned so that the SED matches a modified blackbody with β chosen from a uniform random distribution between 1 and 2, and temperature, T, randomly chosen from a lognormal distribution centred on T = 25 K, and ranging from 20 to 35 K, as shown in Fig. 3. This distribution roughly matches the SEDs of low-redshift galaxies seen in the H-ATLAS survey (Smith et al. 2012).

Sources are included following a two-component redshift distribution with a low-redshift population peaking at z = 0.3, and a high-redshift population extending to z ~ 2 with a peak at z = 1.2. This reproduces the F250/F350 colour distribution observed in the H-ATLAS survey, as shown in Fig. 4. Although these simulations provide a reasonable match to the F250/F350 colour of real data, they

Figure 2. Mean fractional flux errors as a function of the precise source position within a pixel. The contours are linearly spaced from 0 to ~0.9 per cent. The central thicker contour corresponds to mean error of 0, and the outer thicker contour is ~−0.5 per cent. The maximum error is ~−1 per cent for a source in the corner of 4 pixels.

Downloaded from https://academic.oup.com/mnras/article-abstract/493/2/2363/5739926 by Cardiff University user on 10 March 2020
Finally, we add Gaussian noise to the maps. The standard deviation is set to roughly match the instrumental noise in the H-ATLAS survey (Valiante et al. 2016). The values we use here are 9.3, 9.8, and 13.5 mJy pixel$^{-1}$ in the 250, 350, and 500 $\mu$m bands respectively. Note that the sources are positioned on a grid, and so do not suffer from confusion noise, meaning that the appropriate matched filter is the PSF. We consider confusion noise and the modified matched filter in Section 6.3.

We then run MADX on the simulated maps to detect the sources and measure their position and fluxes. We used several different priors: first using only the single bands for the detection: weights 1,0,0 for the 250 $\mu$m band; weights 0,1,0 for the 350 $\mu$m band; weights 0,0,1 for the 500 $\mu$m band. Second, we used equal weighting for each band (weights 1,1,1), corresponding to a flat spectrum source. By design, the sources are on a grid, and cannot overlap, so a simple positional match allows us to associate the detected sources to the corresponding input sources, and calculate the errors in position and fluxes. We identify a recovered source with an input source if the recovered position is within one pixel of the input position. For very low signal-to-noise detections, the large standard deviation of the positional errors means that some detections are not matched within the one pixel radius. This has a small effect on the catalogue completeness, but the dominant source of incompleteness is simply noise on the flux estimates.

### 6.2 Results

For each simulation, we measure the completeness by simply counting the fraction of input sources that are detected as a function of flux. Fig. 5 shows the completeness as a function of flux in each band for both the colour-matched and red-population simulations, and for catalogues using the single-band and the flat priors. The blue lines are for colour-matched simulations with catalogues using the single-band prior and the black lines use the flat prior. It is clear that including information from all three bands significantly improves the completeness of the resulting catalogue. The flux limit at 50 per cent completeness in the 250 and 350 $\mu$m band samples is a factor $\sim 1.4$ deeper using the multiband approach compared to the single-band method. The gain for the 500 $\mu$m band is about a factor of 3, reflecting the very large gain in signal-to-noise ratio by using the information from the other bands to identify sources. The improvements are very similar for the simulations with extra red sources.

The noise in the maps leads to peaks that are detected as a source, but do not correspond to an input source. The number of false detections per beam area is shown as a function of signal-to-noise ratio for each band in Fig. 6. Using the flat-prior detection reduces the number by a factor between 4 and 10 in all bands compared to the single-band detection. Including extra red sources in the simulations makes no significant difference to the rate of false detections. For the single-band detections, the expected number of noise peaks for Gaussian noise with a 2.5$\sigma$ cut is shown by the dashed line. The observed number of false detections follows this very well above the 2.5$\sigma$ cut. There are a small number of sources with fluxes below the 2.5$\sigma$ limit, and this is because cut is applied to the initial peak pixel value, while the final flux plotted on x-axis uses the flux measured at the interpolated peak position. The noise between the two flux estimates scatters some sources below the initial cut. In practice, for the real H-ATLAS data we apply a much higher limit (between 4$\sigma$ and 5$\sigma$) in the final catalogue, so this effect is not visible.

Next we compare the measured positions to the input positions. While including data from lower resolution, or lower signal-to-
Figure 5. Completeness of recovered source catalogues as a function of flux in each band. The blue and black lines are for simulations which match the H-ATLAS colour distribution; the red and magenta include a high-redshift red population. The dotted lines are for source detection using only the relevant single band in each panel, and the solid lines are for detection using the equal weighting of bands. The dashed curves show the expected completeness from Gaussian errors and the vertical dashed lines show the 2.5σ detection threshold. In the 250μm band, using the flat prior pushes the 50% completeness limit about a factor 1.25 deeper. In the 350 and 500μm bands, the gains using the flat prior are factors of 1.28 and 3 for the colour-matched simulations. The effect of the red population is to slightly reduce the completeness in the 350 and 500μm bands. This is a small effect as a percentage of the full catalogue, but represents a significant improvement for the red population itself.

Figure 6. The number of false detections per beam as a function of signal-to-noise ratio. The blue and black lines are for simulations which match the H-ATLAS colour distribution; and the red and magenta include a high-redshift red population. The dotted lines are for source detection using only the relevant single band and the solid lines show the equal weighting of bands. The dashed lines show the expected rate of false detections from Gaussian errors and the 2.5σ threshold applied during the MADX detection. For the 250 and 350μm bands, the flat prior reduces the false detection rate by about a factor 4 at the 4σ limit, and a factor 6 at 3σ. In the 500μm band, the gain is roughly a factor 10 between 2σ and 4σ. The inclusion of the extra red population does not significantly change the false detection rates.
noise bands could increase the positional errors, extra information is provided by summing data with the correct weighting, and the positional accuracy is improved. This can be seen in Fig. 7, which shows the rms position error in RA and Dec. as a function of signal-to-noise ratio, defined as \( \sigma = \sigma_A \) from equations (6) and (7). Using only the 250 \( \mu m \) band to detect sources leads to positional errors in good agreement with the theoretical expectations from Ivison et al. (2007). Including all bands significantly reduces the positional errors, even though the other bands have poorer resolution. At low signal-to-noise ratio (\( S/N \lesssim 4 \)), the apparent positional error starts to flatten because only the sources within one pixel are counted as matches, and the apparent standard deviation is biased too low.

Finally, we compare the measured and input fluxes for the sources, in terms of both random and systematic errors. To assess the random errors, we simply calculate the standard deviation of the flux for real data. Valiante et al. (2007) created realistic simulations and used them to estimate the boosting as a function of signal-to-noise ratio. At high signal-to-noise ratio, there is a small (\( \sim 0.5 \) per cent) underestimate of flux due the peak pixelization issues discussed in Section 5.2. For sources with fluxes near the detection limit, there is a systematic bias to higher fluxes when using the single bands to detect sources (flux boosting). This is related to Eddington/Malmquist bias when selecting sources to be above a signal-to-noise threshold: faint sources with negative errors are not retained in the catalogue, whereas those with positive errors are detected to a fainter level. The precise form of the boosting depends on the detection of true source fluxes, as well as the measurement errors. For our current simulations, we chose to distribute sources uniformly in log flux, and so they do not match real source flux distributions, even though the colours are realistic. Hence, we cannot use them to estimate the boosting as a function of flux for real data. Valiante et al. (2016) created realistic simulations and used them to estimate both the completeness and boosting correction factors that apply to the H-ATLAS data.

Using the flat-prior combination to detect sources includes information from all three bands, and so reduces the bias from noise peaks in any single band, leading to significantly reduced flux boosting. For the 500 \( \mu m \) band, the angular size of the PSF is larger, and the signal-to-noise ratio is significantly smaller than the other two bands. This mean that it contributes only a small amount to the detection signal and positional measurement. The positional errors mean that the local peak is missed and the flux estimate is systematically underestimated. This bias can be corrected using the average values measured from simulations (Valiante et al. 2016).

The flux boosting and biases are not significantly affected by the inclusion of extra red sources in the simulations.

6.3 Confusion noise

The simulations described so far contain no confusion noise, and so the appropriate optimal filter is simply the PSF. To test the performance of the modified matched filter approach, we have added confusion noise to each pixel in the simulations as an extra term consisting of PSF-filtered Gaussian noise, with the variance as seen in the H-ATLAS data (Valiante et al. 2016). The corresponding standard deviation is 7 mJy pixel\(^{-1}\) in all three bands. We use these values of confusion noise to calculate the matched filters as described in appendix A of Chapin et al. (2011). We then re-run the image detection based on the single-band priors, and used both the PSF and the matched filters. We also used the flat-prior detection with the matched filters. The resulting completeness comparisons are shown in Fig. 9. The matched filter selection improves the completeness at a given signal-to-noise cut, providing a catalogue \( \sim 20 \) per cent deeper. As before, using the flat-prior detection also provides an improvement of a factor \( \sim 1.3 \) in flux at the same completeness level for the 250 and 350 \( \mu m \) bands, and a factor 2 for the 500 \( \mu m \) band. Overall the two modifications give a catalogue that is a factor of 1.5–3 deeper in flux at 50 per cent completeness.
Figure 8. The ratio of mean measured flux compared to the mean input flux as a function of input signal-to-noise ratio for different source detection priors. The blue lines show the single-band priors, and the black lines use a flat prior. Panels (a)–(c) show the results for the 250, 350, and 500 µm fluxes respectively. Using the single-band source detection leads to significant flux boosting in the measured fluxes. For the 250 and 350 µm bands, the flat prior reduces the boosting effect at fainter fluxes. For the 500 µm band, the flat-prior fluxes are systematically underestimated at fainter fluxes. The red and magenta lines show the results from simulations with extra red sources. The red sources do not introduce any significant changes.

The added confusion noise means that the flux errors are larger than for the simulations with only Gaussian noise. Using the matched filter reduces the errors by about 10 per cent compared to using the PSF. As an aside, we note that using the matched filter for the Gaussian noise simulations leads to a 14 per cent increase in flux errors compared to the PSF filter. This is exactly as would be expected since the optimal filter for data with Gaussian noise is the PSF, and the optimal filter for data with confusion noise is the appropriate matched filter.

As shown in Fig. 10, using the matched filter reduces the number of false detections by a factor 8 at the 4σ limit for the 250 µm band. Using the flat prior as well reduces the rate by a further factor 2. In the 350 and 500 µm bands, the matched filter provides similar reductions in the false detection rate.

Figure 9. Completeness of recovered source catalogue as a function of flux in each band for simulations with confusion noise included. The blue lines are for source detection using the appropriate single-band prior and PSF filtering; the magenta lines are for source detection using the single-band priors, but with the matched filters; and the black lines use the matched filters and equal weighting of bands.
Figure 10. The number of false detections per beam as a function of signal-to-noise ratio for simulations with confusion noise. The blue lines are for source detection using the appropriate single-band priors and PSF filtering; the magenta lines are for source detection using the single-band priors, but with the matched filters; and the black lines use the matched filters and the equal weighting of bands. In the 250 µm band, using the matched filter reduces the false detection rate by a factor 8 at the 4σ limit. Using the flat-prior source detection reduces it by a further factor of 3. In the 350 and 500 µm bands, the gain is roughly a factor 10 between 2σ and 4σ.

Fig. 11 shows the ratio of measured to input flux when confusion noise is included. The behaviour is very similar to that shown in Fig. 8 for simple Gaussian noise. At high signal-to-noise ratio, there is a small (∼0.5 per cent) underestimate of flux due to the peak pixelization issues discussed in Section 5.2. Near the detection limit, the single-band detection leads to significant flux boosting. Using the flat prior to detect sources reduces this effect. For the 500 µm band, the positional errors again mean that the fluxes are underestimated, but the effect is slightly smaller than for Gaussian noise because the confusion noise has less power on the small scales that directly affect the position estimates.
7 SUMMARY

We have presented a simple approach to detecting sources in data which consists of multiple broad-band images. For high signal-to-noise sources, fitting a Gaussian to estimate the source position and using bicubic interpolation to estimate fluxes significantly improves the accuracy over single pixels estimates. Combining multiple bands in an optimal way before detecting images leads to a significant improvement in sensitivity to faint sources, a reduction in the number of false detections, and an improvement in positional accuracy. Using a matched filter which accounts for confusion noise improves the signal-to-noise ratio of individual flux measurements and so further improves the source detection reliability. Together the two modifications provide catalogues a factor 2–3 deeper than possible with a standard single-band PSF filter approach.

ACKNOWLEDGEMENTS

LD and SJM acknowledge support from the European Research Council (ERC) in the form of Consolidator Grant COSMIDUST (proposal ERC-2014-CoG-647939, PI: H. L. Gomez), and support from the ERC in the form of the Advanced Investigator Program, COSMISM (proposal ERC-2012-ADG-321302, PI: R.J. Ivison).

REFERENCES

Bertin E., 1996, A&AS, 117, 393
Chapin E. L. et al., 2011, MNRAS, 411, 505
Eales S. et al., 2010, PASP, 122, 499
Gonzalez-Nuevo J., Argüeso F., Lopez-Caniego M., Toffolatti L., Sanz J. L., Vielva P., Herranz D., 2006, MNRAS, 369, 1603
Grumitt R. D. P. et al., 2020, preprint (arXiv:1910.08583v1)
Herranz D., Sanz J. L., Barreiro R. B., Lopez-Caniego M., 2005, MNRAS, 356, 944
Herranz D., Sanz J. L., Hobson M. P., Barreiro R. B., Diego J. M., Martínez-González E., 2002, MNRAS, 336, 1057
Herranz D., Argüeso F., Carvalho P., 2012, Adv. Astron., 2012, 410965
Högbohm J. A., 1974, ApJS, 15, 417
Irwin M. J., 1985, MNRAS, 214, 575
Ivison R. J. et al., 2007, MNRAS, 380, 199
Ivison R. J. et al., 2016, ApJ, 832, 78
Kay S. M., 1998, Fundamentals of Statistical Signal Processing: Detection Theory. Prentice-Hall, London
Lanz L. F., Herranz D., Sanz J. L., Gonzalez-Nuevo J., Lopez-Caniego M., 2010, MNRAS, 403, 2120
Maddox S. J. et al., 2018, ApJS, 236, 30
Naselsky P., Novikov D., Silk J., 2002, MNRAS, 335, 550
North D. O., 1943, Proc IRE, 51, 1016 (rep. July 1963)
Pilbratt G. et al., 2010, A&A, 518, 1
Planck Collaboration Int. LIV., 2018, A&A, 619, A94
Pratt W. K., 1978, Digital Image Processing. Wiley, New York
Rigby E. et al., 2011, MNRAS, 415, 2336
Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525
Serjeant S. et al., 2003, MNRAS, 344, 887
Smith D. J. B. et al., 2012, MNRAS, 427, 703
Stetson P. B., 1987, PASP, 99, 191
Tegmark M., de Oliveira-Costa A., 1998, ApJ, 500, 83
Valiante E. et al., 2016, MNRAS, 462, 3146
Vielva P., Martínez-Gonzalez E., Gallegos J. E., Toffolatti L., Sanz J. L., 2003, MNRAS, 344, 89

This paper has been typeset from a \LaTeX file prepared by the author.