Confusion limit resulting from galaxies: using the Infrared Array Camera on board SIRTF

Petri Väisänen,1,2* Eric V. Tollestrup2 and Giovanni G. Fazio2

1Observatory, PO Box 14, 00014 University of Helsinki, Finland
2Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA 02138, USA

Accepted 2001 March 23. Received 2001 March 23; in original form 2000 November 6

ABSTRACT
Recent ISO data have allowed, for the first time, observationally based estimates for source confusion in mid-infrared surveys. We use the extragalactic source counts from ISOCAM in conjunction with K-band counts to predict the confusion resulting from galaxies in deep mid-infrared observations. We specifically concentrate on the near-future Space Infrared Telescope Facility (SIRTF) mission, and calculate expected confusion for the Infrared Array Camera (IRAC) on board SIRTF. A defining scientific goal of the IRAC instrument will be the study of high-redshift galaxies using a deep, confusion-limited wide-field survey at 3–10 μm. A deep survey can reach 3-μJy sources with reasonable confidence in the shorter wavelength IRAC bands. Truly confusion-limited images with the 8 μm will be difficult to obtain because of practical time constraints, unless infrared galaxies exhibit very strong evolution beyond the deepest current observations. We find L* galaxies to be detectable to z = 3–3.5 at 8 μm, which is slightly more pessimistic than found in 1999 by Simpson & Eisenhardt.

Key words: methods: observational – techniques: photometric – galaxies: evolution – galaxies: photometry – cosmology: observations – infrared: galaxies.

1 INTRODUCTION
One of the successes of the Infrared Space Observatory (ISO) mission has been the new extragalactic source counts in the mid-infrared (mid-IR), extracted from both deep (Oliver et al. 1997; Elbaz et al. 1999) and wide-field (Serjeant et al. 2000) surveys. These counts are important in planning effective observing strategies for the next generation of space-borne IR missions, the SIRTF (Fanson et al. 1998) and ASTRO-F (Pearson et al. 2001), particularly with respect to expected source densities and confusion.

Confusion is usually defined as the fluctuations of the background sky brightness below which sources cannot be detected individually – these fluctuations are caused by intrinsically discrete extragalactic sources (we do not consider confusion due to Galactic cirrus in this work). Thus the fluctuations, or noise, come from the same sources that one wishes to study.

Confusion sets an important and fundamental limit to astronomical observations. For a given wavelength and resolution, it will not be productive to extend the exposure time indefinitely since a flux density limit will be reached beyond which no additional discrete sources can be extracted. This is the limit usually referred to as the confusion limit. Confusion cannot be avoided even with arbitrarily high resolution because of the finite sizes of sources on the sky.

Just as new sub-millimetre data allowed Blain, Ivison & Smail (1998) to set observational limits on source confusion at those wavelengths, we are now able to estimate more accurately the confusion at mid-IR wavelengths using deep galaxy counts from ISO data. Confusion in mid-IR has been discussed in connection with the individual deep surveys (e.g. Oliver et al. 1997 for the ISO HDF survey).

A useful theoretical definition of confusion was developed by Scheuer (1957) and followed up in numerous other works. More recently, essentially the same formalism has been used to predict confusion in observations done at radio (e.g. Condon 1974; Wall et al. 1982; Franceschini 1989), sub-millimetre (Toffolatti et al. 1998) and X-ray (Scheuer 1974; Barcons 1992) wavelengths. It turns out that sources at the flux density level where the surface density is approximately one source per beam produce the bulk of the confusion noise. This is easy to understand: at higher surface densities the multitude of sources per beam suppresses the fluctuations (contributing more to the absolute background level), while at lower surface density levels the many ‘empty’ beams also decrease the fluctuations.

However, confusion is not a very straightforward limit since it is a function of wavelength, resolution and other instrument response-dependent factors. Moreover, calculations of confusion limits for a given instrument are highly dependent on source-count
model. ‘Confusion’ is also sometimes confused with source detection probabilities, i.e. completeness level estimation when performing source extraction and photometry. Bright sources are often neglected from confusion calculations, but these bright objects can be a major factor in lowering the completeness of detections in a given survey. The aspect of completeness versus confusion is perhaps the least studied in the otherwise extensive literature on confusion.

The goal of this paper is two-fold: (i) to study the relation of confusion noise to actual detection of sources, and thus arrive at a practical definition for confusion; and (ii) to predict this confusion limit in mid-IR specifically for the forthcoming Space Infrared Telescope Facility (SIRTF) mission.

We start from analytical and theoretical confusion determination and proceed to a more observer-oriented completeness analysis. In particular, we will define confusion ‘noise’ and confusion ‘limit’ and investigate their relation to completeness levels in source extraction. We specify when the different methods of deriving confusion are most appropriate, and present both methods for the Infrared Array Camera (IRAC). The relevant quantitative numbers for confusion with IRAC/SIRTF are presented in Section 6 and Table 1.

2 ANALYTICAL SOLUTION FOR CONFUSION NOISE

Given the source-count distribution and the response pattern of the detector, we can estimate the confusion analytically. This has been done by many authors (see above) – here we mainly follow the notation of Franceschini et al. (1989, 1991). Let \( N(S) \) be the differential distribution of sources of flux \( S \) \( [N(S) \) is used for differential counts throughout this paper instead of \( dN/dS] \). We assume that the sources are distributed randomly on the sky. Now let us include the information of the observing instrument: let \( f(\theta, \varphi) \) be the point spread profile (normalized to unity) of the detector. Then the amplitude of the response at location \((\theta, \varphi)\) from the beam axis is

\[
x = Sf(\theta, \varphi).
\]

The mean number of responses, \( R(x) \), with amplitudes between \( x \) and \( x + dx \) in a solid angle \( d\Omega \) is

\[
R(x) \, dx = \int_0^{\infty} N \left( \frac{x}{f(\theta, \varphi)} \right) \frac{d\Omega}{f(\theta, \varphi)} \, dx,
\]

where we integrate over the whole beam \( \Omega_0 \).

The intrinsic fluctuations of \( R(x) \) can be written simply as the second moment \( \sigma_{\text{conf}}^2 \) of the \( R(x) \) distribution:

\[
\sigma_{\text{conf}}^2 = \int_0^{\infty} x^2 R(x) \, dx.
\]

The lower integration limit is 0 because \( S \) is always positive. The upper limit is not so straightforward; the variance would diverge if high-flux truncations are not introduced in either the source counts \( N(S) \) or the response distribution \( R(x) \). The limit is often set to \( x_c = Q\sigma \), with \( Q \) in the range of 3–5. This kind of upper limit makes the calculation an iterative process. The logic of having a response cut-off rather than a cut-off at specific flux \( S \) is that the latter would eliminate the contribution by bright sources altogether whereas in reality their presence – even relatively far away from the beam centre – could significantly affect the measurement because of the wings of their point spread functions (PSFs).

Notice that if \( Q = 3 \), for example, it does not follow that the objects detected will be ‘3\(\sigma\)' detections in the conventional sense. The \( Q \) value merely tells how large responses are included in the calculation of the confusion: this corresponds to setting a limit below which a response would not be detected with sufficient confidence as such. Thus for Poisson and Gaussian distributions \( Q = 3 \) corresponds to a probability \( \leq 0.3 \) per cent that a response \( x \) results from a large fluctuation of the background. The \( \sigma_{\text{conf}} \) value from equation (3) must be multiplied by 3 to get the ‘3\(\sigma\) confusion limit’ for the detection of objects regardless of the \( Q \) used.

For simplicity we will assume a Gaussian beam pattern

\[
f(\Psi) = \exp(4\Psi \ln 2),
\]

where \( \Psi = (\theta/\theta_0)^2 \). Performing the integration over the solid angle \( \Omega_0 \), equation (3) can be written as

\[
\sigma_{\text{conf}}^2(x_c) = \pi \theta_0^2 I(x_c),
\]

where

\[
I(x_c) = \int_0^{x_c} x^2 \, dx \int_0^{\infty} N \left( \frac{x}{f(\Psi)} \right) \exp(4\Psi \ln 2) \, d\Psi.
\]

Because of the integration limit \( (x_c = Q\sigma) \) it is easier to compute \( \theta_0 \) [the full width at half-maximum (FWHM) of the beam] as a function of \( \sigma \), i.e.

\[
\theta_0 = \sigma/[\pi I(x_c)]^{1/2},
\]

and then determine the confusion for a given \( \theta_0 \).

If the source counts have a power-law form, \( N(S) = kS^\gamma \), equation (5) can be integrated to give

\[
\sigma_{\text{conf}}^2 = \left( \frac{k\Omega_0}{3 - \gamma} \right)^{1/2} x_c^{3-\gamma/2},
\]

where

\[
\Omega_0 = \int f(\Psi)^{\gamma-1} \, d\Omega
\]

is the effective beam size (e.g. Condon 1974; Hacking & Houck 1987). If a Gaussian beam is assumed, as above, we can find the relation

\[
\sigma_{\text{conf}} \propto \theta_0^{(\gamma-1)}
\]

in the absence of cut-offs \( S_c \) in source fluxes.

In summary, it is evident that confusion will be a function of the beam size, the source counts, cut-offs related to the source and response distribution, and the \( Q \) term:

\[
\sigma_{\text{conf}} = \sigma_{\text{conf}}(\theta_0, N(S), x_c, S_c, Q).
\]

3 THE IRAC INSTRUMENT

The Infrared Array Camera (IRAC; Fazio et al. 1998) is one of the three focal-plane instruments on board SIRTF. IRAC is a four-channel camera that provides simultaneous images at 3.6, 4.5, 5.8 and 8 \( \mu \)m. Two adjacent 5.12 \( \times \) 5.12 arcmin\(^2\) fields of view are seen by these four channels in pairs. All four detectors are 256 \( \times \) 256 pixels, with 1.2 arcsec pixel\(^{-1}\) scale. The two short-wavelength channels use InSb detectors, while the two longer wavelengths have Si:As IBC detectors.

In the following, we concentrate on the shortest and longest wavebands, the 3.6- and 8-\(\mu\)m channels. The sensitivities in these
two bands and characteristics of galaxy models, as well as confusion estimates, bracket those expected from the two intermediate channels. The expected $5\sigma$ sensitivities in 200 s for the 3.6- and 8-$\mu$m bands are 3.1 and 23.2 $\mu$Jy, respectively, for point sources, and 0.9 and 5.1 $\mu$Jy arcsec$^{-2}$ for extended sources. The bandwidths are 21 and 38 per cent, respectively.

The major scientific objectives for IRAC and SIRTF are: (i) to study the early Universe; (ii) to study ultraluminous galaxies and active galactic nuclei; (iii) to search for and study brown dwarfs and superplanets; and (iv) to discover and study protoplanetary and planetary debris discs. These objectives are also mirrored in the six large ‘SIRTF Legacy Science Programmes’, which will utilize nearly 3200 h of SIRTF observing time primarily during the first year of the mission. The projects are: ‘The SIRTF Galactic Plane Survey’; ‘GOODS: Great Observatories Origins Deep Survey’; ‘From Molecular Clouds to Planets’; ‘SINGS: The SIRTF Nearby Galaxies Survey – Physics of the Star-Forming ISM and Galaxy Evolution’; ‘SWIRE: The SIRTF Wide-area Infrared Extragalactic Survey’; and ‘The Formation and Evolution of Planetary Systems: Placing Our Solar System in Context’. All of these are expected to result in substantial data bases that will be invaluable for archival research, and in planning subsequent programmes on SIRTF and on other space-borne, airborne and ground-based observatories.

While the motivation of this paper is connected to deep surveys of the high-redshift Universe (e.g. GOODS), confusion limit due to galaxies is a factor also in the deeper brown dwarf searches. Additionally, confusion due to stars becomes important in these surveys when operating close to the Galactic plane, at $b < 20^\circ$ (e.g. the Galactic Plane Survey project).

4 SOURCE-COUNT MODELS

4.1 Baseline model

We explore the effects of confusion and completeness with simple models of mid-infrared galaxy counts. As a first step, observed near-infrared $K$-band source counts are extrapolated to mid-IR.

To predict the counts at another wavelength, the specifics of the observing filter are needed, as well as the shape of the spectral energy distribution (SED) of the contributing sources at different redshifts and the numbers of these sources. The idea is to construct a model that fits reasonably well the observed counts, and then calculate the counts of the target band with the same model. Thus, a consistent extrapolation requires knowledge of the SEDs of different galaxy types, the local galaxy luminosity functions (LF) of the types, possible evolution of the sources, and a cosmological model. The observed counts themselves can then be shifted to the other band using information of the predicted colours as a function of magnitude.

In the following the baseline model is an extrapolation, such as described above, constructed to fit observed $K$-band counts. To compute the galaxy counts we used the general formalism presented earlier in Väisänen (1996). For the SEDs and luminosity evolution of galaxies we use the ‘bc96’ models (Bruzual & Charlot 1993) as presented in Gardner (1998). We also made use of ncmod, a useful general-purpose galaxy number count model presented in Gardner (1998). We adopt a prescription for internal absorption by dust in galaxies from therein as well as the included six different galaxy types (passively evolving E/S0, Sab, Sbc, Scd, Irr, and a small population of constantly star-forming dwarfs). The galaxy mix is from Yoshii & Takahara (1988) and the $K$-band local luminosity function (same for all galaxy types) from Gardner et al. (1997). For cosmology we used $H_0 = 50\,\text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1}$ and $q_0 = 0.02$.

Using the model above, we shift the observed $K$-band counts, a compilation from Väisänen et al. (2000), to 3.6-$\mu$m – these are shown as crosses in Fig. 1. The predicted source count of the same model is overplotted as the thick solid line. This same baseline model is calculated also for 6.7- and 8-$\mu$m bands. The baseline model corresponds to the one in Gardner’s (1998) fig. 11. There are other models overplotted in Fig. 1, corresponding to different cosmologies. The dash-dot curve uses the local near-IR luminosity function of galaxies from Szokoly et al. (1998), which we have found to predict near-IR counts fitting well our own wide-field galaxy counts (Väisänen et al. 2000).

4.2 ISO data-based model

For the 5.8- and 8-$\mu$m predictions, the mid-IR spectral properties of galaxies become important, and extrapolations such as used for the baseline model are not trustworthy. This is because the ‘bc96’ SEDs used above do not include the broad emission features in the spectra of galaxies between 3 and 12 $\mu$m, commonly thought to be the signature of polycyclic aromatic hydrocarbons (PAH). While global properties of galaxy populations at mid-IR still remain uncertain, recent ISO data have certainly shown that PAH features are an important factor that cannot be ignored (e.g. Mattila, Lehtinen & Lemke 1999; Genzel & Cesarsky 2000). This is particularly true for any dusty and infrared-bright objects, which might well dominate counts at mid-IR. Modelling of the relevant
dust and reradiation processes is complicated (see e.g. Guiderdoni et al. 1998; Silva et al. 1998; Xu et al. 1998); therefore, to be as observationally based as possible, our chosen recipe for predicting counts for the longer IRAC bands is to take the newly available mid-IR extragalactic counts at 6.7 \(\mu\)m and shift these to the neighbouring IRAC bands.

We thus fitted an experimental source-count slope to the 6.7-\(\mu\)m extragalactic counts found by various ISOCAM surveys: ELAIS (Serjeant et al. 2000), ISO-HDF project (Oliver et al. 1997), a survey at the Lockman Hole (Taniguchi et al. 1997) and the CFRS (Flores et al. 1999). The faint slope beyond the deepest observations is adjusted to the typical shape of existing (evolutionary) models (Pearson & Rowan-Robinson 1996; Franceschini et al. 1997; Roche & Eales 1999). Preliminary ISO counts from the HDF-S field (Oliver et al. 2001) are also in excellent agreement with this experimental source-count model. Though various independent counts are in very good agreement, it should be noted that there are still uncertainties in the calibration of the ISOCAM data (see e.g. Aussel et al. 1999; Serjeant et al. 2000). Note also that the ELAIS counts are not corrected for a possible Eddington/Malmquist type bias due to a non-Gaussian flux-error distribution (see fig. 1 in Serjeant et al. 2000).

Fig. 2 shows the recent mid-IR counts with our fitted model overlapped as the thick solid curve and the equivalent of the 3.6-\(\mu\)m baseline model as the lower thin solid line. We then translated both the 6.7-\(\mu\)m counts and the fitted model to the 8-\(\mu\)m band (Fig. 3).

Naturally the translation to the 8-\(\mu\)m band is model-dependent, but since the shift is not large (the ISOCAM filter partially overlaps the IRAC 8-\(\mu\)m filter), the uncertainty is low, even given the uncertain mid-IR spectral features. We used the baseline model in the transformation. We computed numerous model counts at 6.7 and 8 \(\mu\)m, with different LFs, galaxy mixes, cosmologies and SEDs (with and without PAH features), and found that all these changes in the transformation of counts are well within the error bars of the observed counts.

4.3 Reliability of models

As is seen from Fig. 2, the observations do not constrain the models fainter than about 20 \(\mu\)Jy, so the faint counts therefore remain model-based. However, the baseline model provides a realistic slope prediction since it is derived from near-IR counts, which extend two decades fainter. Furthermore, the faint counts result mostly from high-z objects seen at their redshifted near-IR rest frames, which are well modelled by the adopted SEDs.

We briefly compare some 6.7-\(\mu\)m model counts with the models used here. We have plotted the model of Franceschini et al. (1997) in Fig. 2 for reference. Models of Franceschini et al. (1991) are similar to the baseline curve here. Pearson & Rowan-Robinson (1996) models match our thick solid curve very well at and beyond the break point. The evolving model of Roche & Eales (1999) is very close to our experimental ISO-fitted curve. Overall, the faint mid-IR counts at 6.7 \(\mu\)m (and especially at 15 \(\mu\)m) are surprisingly high and have a steep slope.

At brighter flux levels, the 6.7-\(\mu\)m counts fit well the models of Pearson & Rowan-Robinson (1996), as well as those of Roche & Eales (1999), while they are in clear excess of models of e.g. Franceschini et al. (1997) and the baseline prediction of our near-IR extrapolated model.

Where does this large difference at the mJy level number count regime come from? One explanation is the effect of PAH features.
As a test, we calculated the baseline model in an alternative way, substituting all the spiral galaxy SEDs beyond 1 \( \mu m \) with mid-IR spectra taken from GRASIL results (Silva et al. 1998), which include the broad mid-IR spectral features. The effect at 1 mJy is substantial: in the 6.7-\( \mu m \) band there is almost a factor of 10 increase in counts, raising the baseline model just to the level of our experimental counts. At 8 \( \mu m \) the increase is a factor of 30, which takes the baseline curve slightly higher than the ISO-fitted experimental curve. By 10 \( \mu m \) the counts with and without PAH features are within a factor of 2 again. Whether it is appropriate to use such SEDs for the whole population of disc galaxies is uncertain. Nevertheless, it is clear that the experimental ISO-fitted counts used here are by no means unrealistic, and, moreover, the PAH features in mid-IR spectra are a crucial factor in explaining mid-IR counts.

5 DERIVING CONFUSION

5.1 Confusion versus beam size

Equation (7) is used to compute confusion as a function of the FWHM of the beam size (\( \theta_b \)). This is equivalent to confusion-limit curves in the papers of Franceschini et al. (1989, 1991). Several different cases are plotted in Figs 4 and 5, corresponding to source counts in Figs 1 and 3.

Fig. 4 shows effects in the 3.6-\( \mu m \) IRAC filter confusion due to typical changes in cosmology and of parameters in the \( \sigma_{\text{conf}} \) calculation. The confusion noise \( \sigma_{\text{conf}} \) for the baseline model is calculated using \( Q = 3 \) (see equation 3 and discussion following it) and plotted as the solid line. Confusion from the same model is calculated also using \( Q = 2 \) and 5; these are plotted as dotted lines, below and above the first curve, respectively. The first model is then modified to flat cosmology \( (q_0 = 0.5 \text{, dashed line}) \); it is unrealistic, but is included to compare with some earlier determinations) and to flat cosmology with a cosmological constant \( (\lambda_0 = 0.8 \text{, long-dash line}) \). Finally, the dash-dot curve corresponds to the same alternative local LF as used in Fig. 1.

Depending on which model one adopts, IRAC/SIRTF’s 1\( \sigma \) confusion noise from the realistic models at 3.6 \( \mu m \) seems to be 0.2–0.5 \( \mu m \) Jy for the IRAC resolution of \( = 2.3 \text{arcsec} \). It is noteworthy that changing the \( Q \) value changes the confusion as much as extreme changes in cosmology. It is thus important to check which \( Q \) value was used when directly adopting some confusion calculation for a given mission [e.g. Toffolatti et al. (1998) use \( Q = 5 \) when predicting confusion in far-IR and sub-millimetre].

Fig. 5 shows various models for the IRAC 8-\( \mu m \) filter. Next we investigate changes in confusion due to different galaxy population models. The solid curve shows the adopted baseline model, equivalent to the baseline 3.6-\( \mu m \) model. The corresponding source count can be found in Fig. 3. The baseline model is close to the confusion prediction of Franceschini et al. (1991), which has been used during the SIRTF planning phases. The resulting ISO-fitted confusion curve is plotted as the thick solid line. The expected 1\( \sigma \) confusion noise is \( \sigma_{\text{conf}} = 0.6 \mu m \) Jy.

It turns out that the flux levels contributing to the confusion of the longer-wavelength IRAC filter are just beyond the sensitivity limit of deepest ISOCAM observations. Thus the exact amplitude and location of the ‘bump’ seen in the mid-IR counts essentially determines the expected IRAC confusion limit. We return to this issue in Section 5.3.

© 2001 RAS, MNRAS 325, 1241–1252
Although both of the above confusion-limit analyses are often used, they are at opposite ends with respect to their approach. The analytical calculation presented first determines the fluctuations caused by unresolved sources. The second method is based on the realization that the bulk of fluctuations resulting from these unresolved sources are caused by sources that have a source density of about one per beam. Using any given (cumulative) source-count law, one can calculate where that level is, and assign it as the confusion limit. This latter method thus ignores the unresolved sources, and instead starts out by calculating numbers of bright sources. It is thus more related to completeness analysis in counting of individual galaxies – hence the term ‘confusion limit’ used here. Additionally, it is difficult to associate a meaningful noise figure, one that could be added to other noise sources, to this confusion limit; it does not add in quadrature like random noise.

### 5.3 Sensitivity of confusion to models

As seen earlier, the 6.7-µm model is not constrained by observations below \( \approx 20 \mu\)Jy. While the faint extrapolation is based on near-IR data and realistic models, it is worth while to investigate how sensitive the confusion noise prediction is to the model-dependent deep counts in the 8-µm band. Consider four ad hoc number-count slopes beyond the ISO data regime. The dashed curves in Fig. 3 show models where the amplitude of the bump at the break point of the counts (at \( \sim 10 \mu\)Jy) is either higher or lower than expected from models fitted to ISOCAM data. The dotted curves assume that the break is determined correctly, but the faint slope is shallower or steeper than the one adopted previously.

Fig. 5 shows the calculated \( \sigma_{\text{conf}} \) for each of these source counts. It is obvious that the faint-end slope (beyond \( \approx 6 \mu\)Jy) does not significantly affect the confusion level. The amplitude of the bump, on the other hand, is the most important factor contributing to confusion noise.

The dependence of confusion on the ‘bump’ is expected. The 8-µm sources start to approach beam-per-source surface densities with IRAC resolution at about the same region where the bump, or break, in the counts occurs. In addition, with steep counts, the sources at the cut-off (or roll-over) level are the population contributing most to confusion (e.g. Toffolatti et al. 1998).

Basically, depending on whether one normalizes the faintest 6.7-µm counts according to the HDF-N counts (high; Oliver et al. 1997), or to the Lockman Hole counts (low; Taniguchi et al. 1997), one gets almost a factor of 10 difference in \( \sigma_{\text{conf}} \) at IRAC resolution. The deep ISOCAM observations of Aliteri et al. (1999), through a cluster lens, support the high normalization.

Nevertheless, a word of warning concerning the mid-IR ISOCAM counts is warranted. Some of the observations have been pushed quite deep, for example, beyond a 20 beams-per-source level. Our photometric simulations show that the accuracy of photometry is very low at these levels, and the trend is systematically towards higher fluxes. In addition to unstable photometry, astrometric errors become significant (Hogg 2001). Thus, until the counts are confirmed by better resolution (and hopefully larger sky coverage), it remains a possibility that the deepest mid-IR counts (at both 6.7 and 15 µm) are significantly overestimated. If that is the case, the 8-µm baseline model would be more appropriate for estimating confusion levels than the ISO-fitted one.

© 2001 RAS, MNRAS 325, 1241–1252
6 DISCUSSION

6.1 Completeness of counts in IRAC bands

Both completeness and reliability must be considered when detecting faint sources. Certain detection techniques might give high completeness rates while suffering from poor reliability. On the other hand, requiring high reliability means lower completeness at a given flux level. We study both aspects, along with the effect of source confusion, using simulations of images that are realizations of model counts introduced earlier.

An analogue of the \( \sigma_{\text{conf}}(b) \) calculation in Section 5.1 is to test how the extraction of sources from a simulated image changes with the convolving PSF. We constructed simulated images from the baseline 3.6- and 8-\( \mu \)m models – Figs 8(a) and 9(a) show the 3.6- and 8-\( \mu \)m intrinsic cumulative model counts as dashed curves. Sources were drawn from \( N(S) \) and distributed randomly on an image with 1.2 arcsec pixels (fractions of pixels were allowed as the centre of a source). The image size was 768 pixel\(^2\), corresponding to a 3 \( \times \) 3 mosaic of IRAC frames. The image was then convolved with different Gaussian PSFs and sources were extracted using SExtractor (Bertin & Arnouts 1996), and aperture magnitudes with approximately the size of the FWHM were calculated. The cumulative ‘observed’ counts are overplotted as solid lines.

A comparison to confusion can be made by defining 70 per cent, for example, as a satisfactory level of completeness and determining the flux density level where this completeness is reached. These were determined from each simulation and plotted in Figs 6 and 7 as squares. The required 70 per cent completeness corresponds to 3–6\( \sigma_{\text{conf}} \) confusion noise levels at FWHMs of several arcseconds and up to 10\( \sigma_{\text{conf}} \) at smaller FWHMs.

The dashed lines in Figs 6 and 7 show the limits where there are 20 beams per source. The completeness limit is close to this level. Confusion limits are often quoted as being at 30–40 beams per source, which is realistic when higher completeness is required [or the definition of a beam is smaller, as e.g. in Hogg (2001)]. Also, the example above showed a very idealized case regarding object detection – perfect Gaussian point sources, no additional noise or source clustering. We will return to the issue of detection below.

Another way to derive confusion limits from simulations is to test how the extraction of sources changes as noise is added to an intrinsically noiseless image. The same simulated images as above were used, except that now the real IRAC PSFs for the two bands were used (FWHMs for both bands are \( \approx 2.3 \) arcsec). Fig. 8(b) shows the same intrinsic cumulative counts as earlier and the extracted counts corresponding to different amounts of additional noise. Starting from the bottom, the noise levels correspond approximately to the following integration times (IT, in seconds): 12, 24, 48, 60, 90, 100, 200, 300, 1200, 3100, 19 800 and 49 600 s. We have used the low background environment and the full-array readout mode, which has only ITs of 12, 30, 100 and 200 s available, each of which has its own sensitivity (an IT of 60 s results in a different noise level depending on whether it is acquired with five co-adds of 12 s ITs or two readouts of 30 s ITs). The topmost observed curve is the truly confusion-limited case, where no noise is added. It is somewhat lower than the case with a Gaussian PSF of approximately the same size as used in Fig. 8(a) (thick line), showing the effect of the extended wings in the real PSF.

The corresponding case for the 8-\( \mu \)m baseline model is shown in Fig. 9(b). The corresponding integration times (in kiloseconds) for

\[
N(S) / deg^2 = \frac{S}{\sigma_S^2}
\]

6.2 Reliability of detections

Naturally the extraction algorithm plays a role in construction of
the observed source counts and the exact determination of completeness levels. For example, it is seen that the extracted counts show an excess compared to input sources in Figs 8 and 9. Examination of input and output sources in our simulations showed that this bump is mostly caused by overestimated fluxes of sources close to the detection limit. The overestimation is due to a ‘pedestal effect’: unresolved sources are entering the aperture where the source’s magnitude is measured, and normal sky-subtraction routines are biased because of the non-Gaussian noise coming from the positive-only signal of faint sources. We found this to be typical for all photometric systems based on aperture or isophotal photometry in very crowded fields. PSF-fitting techniques, or extremely small apertures with subsequent aperture corrections, give more accurate results in point-source cases, such as the current test; however, these methods in turn have complications with real data where other shapes of objects are expected as well. The common Malmquist/Eddington-like bias is also partially responsible for the bump in the counts. It affects directly the numbers of sources counted: as there are more objects per flux (or magnitude) interval below the flux limit than above it, photometric errors preferentially scatter fainter sources to a brighter flux bin than vice versa.

The depth reached can be optimized by tuning appropriate detection parameters in the selected extraction method. However, a
detailed study of detection and photometric techniques is beyond the scope of this paper. To isolate the completeness versus confusion effects, we merely took care to be consistent from one test to another using small aperture magnitudes within SExtractor.

It is worth pointing out, however, that regardless of the extraction algorithm used, the source-count slope and the bright sources have a significant and large effect on confusion and completeness. In the traditional parlance of Section 2, the high-response tail of the $R(x)$ distribution strongly affects the detection of sources. Also it is the slope of the counts that determines $R(x)$. To illustrate this point, we performed simulations with pure power-law source counts.

Consider two source counts, shown in Fig. 10(a). The differential counts have a power-law form $N(S) \propto S^{-\gamma}$, where the shallow curve has $\gamma = 1.5$ while the steep one obeys Euclidean counts with $\gamma = 2.5$. Both are adjusted to give counts similar to the ISO-fitted 8-μm counts (dotted line) at $\sim 10\, \mu$Jy. These counts are unrealistic, of course, but do illustrate clearly the dangers of adopting confusion limits without checking how they are derived. The squares show the observed counts from images made using the two galaxy counts. The characteristic bump of the observed counts is seen with the steep slope. A steady decrease of completeness is evident with the shallow counts: the 70 per cent completeness limit is at 10$\mu$Jy while for the Euclidean counts the limit is closer to 5$\mu$Jy. The reason for incompleteness is seen in panel (b) where the cumulative surface density of sources is plotted. There are many bright objects in the shallow count; 50 beams per source is reached at 100$\mu$Jy where the completeness starts to drop. The 70 per cent completeness is at 17 beams per source. For the steep counts the drop also begins at $\approx 50$ beams per source, but is much faster, and by $\approx 20$ beams per source a completeness correction of a factor of 5 would be needed. This is because we are approaching confusion noise instead of confusion limit.

Panel (c) shows the pixel histograms of the images; these are equivalent to the total $R(x)$ distribution (equation 2). The distribution approaches a Gaussian as the slope steepens; with shallow slopes the histogram approaches the $N(S)$ distribution (e.g. Condon 1974). The width of the distribution is related to the calculated confusion noise $\sigma_{\text{conf}}$. This is shown in panel (d) as a function of FWHM. The $\sigma_{\text{conf}}(\theta)$ dependence of equation (9) is also verified. It is seen that the $\sigma_{\text{conf}}(\theta = 2.3\theta')$ from the Euclidean counts is a factor of 2 higher than from the shallow counts, in direct contrast to the expectation from source extraction simulation and surface density analysis above. For the steeper counts one can assign, for example, 5$\mu$Jy as the confusion limit, which would be equivalent to 4.5$\sigma_{\text{conf}}$ or 32 beams per source. The corresponding numbers at 70 per cent completeness (10$\mu$Jy) for the shallow curve are 19$\sigma_{\text{conf}}$ and 17 beams per source.

It thus turns out that the slope of the counts determines which sources dominate the confusion. If the source-count slope is shallow, it is the bright sources that dominate, reducing the completeness of detection of an object at a given flux level. In this case the ‘rule of thumb’ taken at $\approx 20$ beams per source (lower for shallow counts and higher for steeper ones) gives a good idea of the practical confusion limit.

If the slope is steep ($\gamma > 2$) the confusion noise due to fainter sources starts reducing the completeness faster than the effect of bright neighbouring sources (see also the discussion in Helou & Beichman 1990). In this case the traditional way of calculating $\sigma_{\text{conf}}$ from equation (5), and using $\approx 10\sigma_{\text{conf}}$ for the confusion limit, is more appropriate.

### 6.3 Integration times for IRAC

Figs 11, 12 and 13 show the expected completeness levels using different integration times for observations with IRAC’s 3.6- and 8-μm filters with realistic PSFs. Here we used a comprehensive Monte Carlo simulation: rather than extracting the source count once from a simulated image (as was done for Figs 8 and 9), test sources of a given flux were placed one-by-one on the simulated frames and then extracted. In each flux bin for every simulation with different noise floor, the test was performed 2000 times. The integration times corresponding to the different 1σ noise floors are printed in kiloseconds. We show the baseline models and also the ISO-fitted model for the 8-μm band.

A ‘detection’ can be defined in various ways. Here it was defined as finding a source (the closest one) within a FWHM radius and within a 4σ flux interval determined from the observed spread of

![Figure 11](image1.png)

**Figure 11.** Fractional completeness level as a function of the flux of the input source. The image in this Monte Carlo simulation is constructed from the baseline model at 3.6 μm. The curves are labelled with the integration time in thousands of seconds. The thick solid line is the truly confusion-limited case, with no noise added to the image, i.e. infinite integration time.

![Figure 12](image2.png)

**Figure 12.** Same as Fig. 11, but for the 8-μm band baseline model.
fluxes at the given flux density. Making the selection in this way we, first of all, avoid a serious a priori knowledge bias, which would be introduced if exactly the same input source were searched for from the simulated image. Secondly, we chose the mentioned flux cut to be dependent as little as possible on the used photometric technique. If, for instance, we had required a detection to be within ±30% of the input flux (this is used in Hogg 2001), there would have been more non-detections due to the typical brightening of sources when approaching the confusion limit. In this case the resulting no-noise completeness curves in Figs 11–13 would be lower by a factor of ~1.5 at the few μJy level.

Figs 11 to 13 can be used to estimate the longest useful integration times for the IRAC instrument in deep extragalactic surveys. To be totally confusion-limited at 3.6 μm, i.e. that the incompleteness would be due only to confusion, ITs of approximately 50 ks (14 h) would be needed. This is within the scope of a Legacy Science type project, where a ‘noise map’ of unresolved background sources from a reasonable size of area could be obtained for fluctuation studies.

On the other hand, as far as direct source counts are concerned, much shorter ITs are adequate. Consider a 500-s exposure. At 6 μJy, sources would still be detected to the confusion limit; that is, sources would be lost at 20 per cent rate due to confusion only. Half of all 4-μJy objects would be undetected; ~50 per cent of these are lost due to confusion and the other half due to sky noise. This 4-μJy level corresponds to approximately 15σconf as calculated from equation (7). Moreover, it is easy to see that the confusion limit is approached; if the IT is increased by a factor of 10, sources only 1.5 times fainter would be detected with the same confidence level.

If only the 3.6-μm band were considered, we would conclude that it does not make much sense to extend the integration time beyond 1 h. With this IT, 3-μJy sources are well detected. However, one is pushing confusion already since only ~60 per cent of all sources at that flux level could be extracted (it becomes unreliable to interpret and model the count slopes if completeness corrections exceed a factor of 2). This limit corresponds to approximately 25 beams per source.

In the longest IRAC waveband, with the baseline model, integration times exceeding 10 h would still uncover new sources at 3–4 μJy level with only modest completeness corrections of ~1.2. However, truly confusion-limited images at 8 μm are much harder to reach than at 3.6 μm – they would require nearly 100 h of integration time.

A practical 8-μm confusion limit is reached at a brighter flux level if galaxy counts are better modelled with the ISO-fitted counts. Requiring 70 per cent completeness (with ≈10σ ideal photometric detections), sources at 4 μJy would be detected with the ISO-fitted model, while 2-μJy objects could be reached with the baseline model. However, in the latter case the IT needed would exceed the scope of any realistic project.

For the 8-μm band we conclude that an integration time of about 15 h would be both sufficient and useful for performing a deep survey (at the same time this would allow truly confusion-limited images from the two shortest-wavelength filters). Sources at 4 μJy would be at 8σconf or 11σconf depending on whether galaxy counts are closer to the ISO-fitted or baseline model, respectively. Respective completeness levels would be at ~60 and 80 per cent. The ISO and baseline models give 20 and 40 beams per source at 4 μJy, respectively.

Table 1 summarizes the integration times along with the expected values of confusion for the main models used in this work. It gives the confusion limit and noise in the several different methods presented in this work. Specifically, it can be seen that the one-beam-per-source limit is nearly two orders of magnitude too low. Because the faint mid-IR counts are flatter in the 3.6 μm band, the overall confusion limit is best described by approximately 15σconf. In the longer IRAC waveband the limit is ≲10. The steepest counts are the baseline 8 μm – the resulting adopted confusion is approximately 25 beams per source, while for the other two cases in the table confusion could be described as ≲20 beams per source.

There are indications that the IRAC resolution might prove to be better than the instrument specifications used here (2.3 arcsec). If this is the case, confusion estimates change somewhat, especially for the two shortest-wavelength bands. If the 3.6-μm band were to achieve an in-flight PSF of ≈1 arcsec, the σconf value would decrease by a factor of 7 (taking into account the 1.2 arcsec pixel size). A limit of 25 beams per source would be at 0.5 μJy. It would therefore become difficult to reach confusion-limited images in

Table 1. Confusion limits in μJy for two IRAC bands using several models. Different values of confusion are given according to the method by which the limit is derived. In all the models in this table a value at ≈20 beams per source, or ≈10σconf gives a good estimate (‘adopted confusion limit’) of the flux density level that could ultimately be reached with sufficient accuracy (10σ with ideal photometry). However, for a practical limit, the required integration time (IT) would need to be considered as well.

<table>
<thead>
<tr>
<th>Method</th>
<th>3.6-μm baseline</th>
<th>8-μm baseline</th>
<th>8-μm ISO-fitted</th>
</tr>
</thead>
<tbody>
<tr>
<td>one source per beam</td>
<td>0.02</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td>20 sources per beam</td>
<td>2.0</td>
<td>2.3</td>
<td>4.2</td>
</tr>
<tr>
<td>confusion noise σconf</td>
<td>0.25</td>
<td>0.40</td>
<td>0.51</td>
</tr>
<tr>
<td>flux at 70 per cent compl.</td>
<td>2.2</td>
<td>2.3</td>
<td>4.1</td>
</tr>
<tr>
<td>adopted confusion limit</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>corresponding IT (ks)</td>
<td>3</td>
<td>150</td>
<td>50</td>
</tr>
</tbody>
</table>
this band. The 8-µm band estimates would not change significantly from the values presented above. The possibly improving resolution will be counteracted by real images, which are not as ideal as the point-source simulations performed here but will be spatially resolved galaxies.

Finally, note that if 90 per cent completeness were desired, confusion limits are more appropriately at \( \sim 10 \mu Jy \), or at around 100 beams per source. If 80 per cent completeness is enough, IRAC confusion limits are at \( \sim 6 \mu Jy \).

6.4 Detection of high-\( z \) objects

The study of the high-redshift Universe with SIRTF is highly dependent on obtaining accurate enough photometric redshifts of distant galaxies. This can be performed by tracking the signature of the H\(^\alpha\) opacity feature in galaxy spectra as it passes through the IRAC filters (Wright, Eisenhardt & Fazio 1994). As far as IRAC sensitivities are concerned, Simpson & Eisenhardt (1999) show that a realistic goal for the deep galaxy surveys is to determine these redshifts accurately for \( L^* \) galaxies up to \( z \sim 4 \). However, they did not consider confusion in their calculations, though they refer to the \( 1\sigma_{\text{conf}} \approx 0.5 \mu Jy \) expected from Franceschini et al. (1991). The results thus seem optimistic compared to work here. Flux densities of 1–2 \( \mu Jy \) seem impossible to reach. Certainly many individual sources at this level will be extracted, but they come from deep within the confused regime, so that obtaining global properties would be highly uncertain as a result of large completeness and photometric corrections.

According to our confusion estimates, and using model spectra of an evolved \( L^* \) galaxy (Fazio et al. 1998), these galaxies can be detected out to \( z \approx 3.5 \) using the baseline model case. This is obtained by setting 3 \( \mu Jy \) as the lowest flux density level where object photometry can be determined with enough accuracy (we assume 10\( \sigma \) in the ideal case, which still leaves room for source extraction in practice, when objects are not point sources). In the ISO-fitted case \( z \sim 3 \) could still be reached.

The situation changes if the highest-amplitude ISOCAM galaxy counts (HDF-N counts of Oliver et al. 1997) prove to be the most accurate counts. If the bright bump model (higher dashed line in Figs 3 and 5) were used, the 8-µm band would already become practically confusion-limited at 10 \( \mu Jy \). This would limit high-\( z \) observation to \( z \sim 2 \).

It is just these faint IR galaxies with unknown population characteristics that SIRTF projects will be studying. If the deep surveys detect high counts and surface densities, as would be expected from ISOCAM results, it would immediately imply strong evolution (e.g. Elbaz et al. 1999), and possibly hitherto unknown evolutionary stages of infrared-bright objects. On the other hand, that would also necessarily mean that the IRAC instrument is becoming confusion-limited sooner than expected, and the highest-\( z \) studies of galaxies may suffer. Surveys reaching galaxies at \( z \geq 3 \) are possible only if the mid-IR counts are at a moderate level and/or the in-flight IRAC PSF performance will be at the most optimistic values in all IRAC filters.

7 SUMMARY

We have presented empirically based estimates for confusion due to galaxies in mid-infrared wavelengths. For SIRTF’s IRAC bands the confusion limits are at 3–4 \( \mu Jy \), if completeness corrections by a factor of 1.3 to intrinsic counts are allowed. The 1\( \sigma \) confusion noise ranges from 0.2 to 0.5 \( \mu Jy \) on moving from the 3.6-µm waveband to the 8-µm. If the highest-amplitude ISO-based mid-IR galaxy counts will be confirmed, the confusion limit is higher by a factor of 2.

We have shown the differences between commonly used confusion definitions. Specifically we separated confusion noise from confusion limit, so that the first relates to the fluctuations from unresolved sources and the latter is defined by source surface density arguments or derived by completeness analysis of source detections. The slope of counts determines which confusion determination is more appropriate to use. With steep counts it is the faint source fluctuations that dominate, and when the count slope is shallow bright sources decrease completeness efficiently and thus determine the confusion limit. The combination of IRAC resolution and typical mid-IR source counts is an intermediate case, where the overall confusion limit (\( \sim 4 \mu Jy \)) can be found at 10–20 times the confusion noise \( \sigma_{\text{conf}} \).

We conclude that, to make a useful high-redshift (\( z \sim 3 \)) deep survey with SIRTF /IRAC, integration times of at least 15 h should be used. Such a survey would be confusion-limited at 3.6 and 4.5 µm, if the original IRAC design requirement resolution is used. If sharper images are obtained, 3.6-µm images will become confusion-limited at similar integration times to the 8-µm channel, i.e. only after 100 h.

ACKNOWLEDGMENTS

PV thanks the Finnish Academy and the Smithsonian Institution for support during this research. We acknowledge useful comments from Kalevi Mattila and Andrew Blain.

REFERENCES


© 2001 RAS, MNRAS 325, 1241–1252


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