Scientific report on existing software

INAF – OAR for the Astrodeep project

ASTRODEEP

"Unveiling the power of the deepest images of the Universe"

THEME [SPA.2012.2.1-01]

[Exploitation of space science and exploration data]

Grant agreement for: Collaborative project

Grant agreement no: 312725

Deliverable 3.1

ABSTRACT

Report on existing software tools dedicated to source extraction and definition of optimal technique for catalog production in multi-wavelength high-density fields

Prepared by: E. Merlin, F. Buitrago, N. Bourne, N. Cappelluti, M. Castellano, M. Comastri, J. Dunlop, D. Elbaz, H. Ferguson, A. Fontana, R. McLure, P. Santini, X. Shu, T. Wang
Approved by: AEC
Date: 30/12/2013

Goal

The ASTRODEEP consortium is committed to process the data resulting from the deepest surveys of the Universe and to deliver the resulting data products to the worldwide community.
This document describes the tools that already exist and are used to obtain photometric catalogs in deep extragalactic fields. In most cases, the procedures described here are available within the members of the ASTRODEEP consortium, some of them being public tools (e.g. TFIT or CONVPHOT) and others as recipes developed internally. In the document we also describe other procedures, not currently used in the ASTRODEEP team, that however are sufficiently well described and used in the literature and are therefore useful reference for future developments.

1 Preparation software

Before starting with the real work to obtain photometric catalogs from images, it is necessary to pre-process them to obtain an initial catalog of the sources to be analysed and to to subtract background light.

1.1 Image segmentation

One of the preliminary steps to any photometric analysis is the detection of the astrophysical objects through the so-called "segmentation" process, which consists in separating the pixels of the image where objects lie from the surrounding "empty" regions where only the sky background is present. Since the value of each pixel in an astronomical image is determined by the sum of the background signal and the photons coming from the sources in that region of the sky, the process of object detection is intimately connected with the determination/subtraction of the background, which will be described in details in Sect. 1.2.

The public, open-source software SExtractor (Bertin & Arnouts 1996) has become a standard for the segmentation of images aimed at the analysis of extragalactic sources. Image segmentation in SExtractor can be schematically described as a sequence of three steps which are controlled by the user through input parameters:

1. *Background estimation* (Sect. 1.2).
2. *Thresholding*. The software individuates groups of pixels touching each other at their sides or angles ("8-connectivity") and whose value exceeds a given threshold. The threshold value is specified by the user in units of the background's standard deviation (parameter DETECT_THRESH). Only groups exceeding a minimum number of pixels (defined by the parameter DETECT_MINAREA) are considered as "reliable" detections. Similarly, the maximum dimension of an object tobe kept in the final segmentation is specified as DETECT_MAXAREA.
3. *Deblending*. Connected set of pixels are analysed following a multi-thresholding approach (Beard et al. 1990). The set is re-thresholded at N levels (parameter DEBLEND_NTHRESH), exponentially spaced (in the case of CCD images) between the

extraction threshold and the peak value. Subsets are individuated at any of the N thresholds and they are considered as separate components if two or more have an integrated pixel intensity greater than a given fraction (parameter DEBLEND_MINCONT) of the total intensity of the original object.

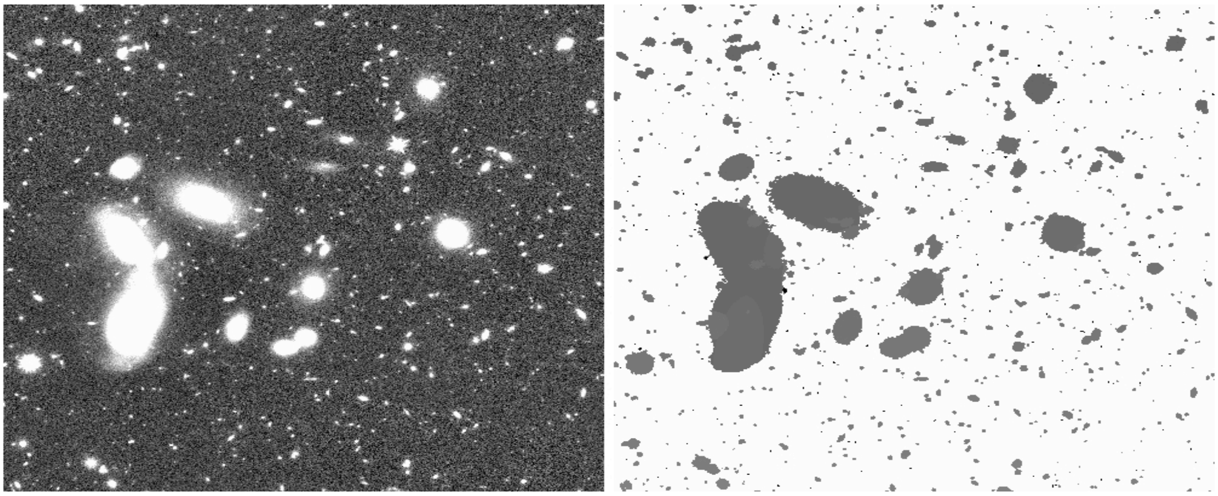

Fig. 1. Science image (left, F160W band from WFC3-HST) and relevant "segmentation map" (right) obtained with SExtractor.

A "segmention map" of the same size of the input image can be saved as output of SExtractor (Fig. 1). In the segmentation map the original pixels where objects have been individuated according to the above process are assigned an integer value corresponding to the object identification number. In turn, object-free regions are set to zero.

1.1.1 Segmentation dilation

Since objects in a scientific image are defined as connected pixels above a given threshold over the background noise, in the external regions of the sources the noise can be large enough that the outer regions of the objects are not correctly identified. The area assigned to a given object in the segmentation map is therefore smaller than the typical extent of the real objects.

A specific software, DILATE, has been developed at INAF-OAR to enlarge the object area and recover most of the actual size of the objects. This code was originally designed within the CONVPHOT package and has recently been revised to analyse the CANDELS data, as described in Galametz et al. 2013.

In the original approach, sources with an area above a minimum threshold m_{AREA} were dilated by a constant factor of 4 (i.e., doubling the radius). Sources smaller than m_{AREA} were dilated to reach this minimum threshold.

In the current version a gradual increase of the dilation factor has been introduced:

• If ISOAREAF IMAGE > 1000: Dilated area = area

• If 60 < ISOAREAF IMAGE < 1000:

Dilated area = −0.0004×area2 + 1.25×area + 166.

• If ISOAREAF IMAGE< 60: Dilated area=4×area.

1.2 Background subtraction

The proper evaluation of the background is a critical step in all kind of photometric measurements. In principle, the background could be measured by evaluating the median intensity of the image in the regions devoid of sources. Unfortunately, the very existence of areas "devoid" of sources is poorly defined. First of all, areas devoid of detected sources do contain fainter sources that are not individually detected but that can contribute to an average background. In addition, as the density of sources increases (i.e. as the depth of the image increases) and especially when the PSF becomes comparable or even larger than the typical separation between sources, uncontaminated portions of the image can be very hardly identified.

For this reason a proper background subtraction is a critical step in any procedure, and at some level a still unsolved problem in deep photometric analysis. It is important to remark that the background can be dealt with in a different way in two different steps of the photometric process: the object detection (since objects are defined above a given threshold over the background) and in the final estimate of the magnitude.
We describe here two procedures that have been adopted: the first is the standard method adopted in the SExtractor package, and the second is a specialized technique that has been adopted to process Spitzer low resolution crowded images.

1.2.1 The SExtractor approach.

This approach is based on the assumption that most discrete sources do not overlap too severely, which is generally the case for high galactic latitude fields in good seeing conditions.

During the detection process, SExtractor constructs the background map by making a first pass through the pixel data, computing an estimator for the local background in each mesh of a grid that covers the whole frame. The background estimator is a combination of $\kappa\sigma$ clipping and mode estimation, similar to the one employed in DAOPHOT (Stetson 1987). Briefly, the local background histogram is clipped iteratively until convergence at $\pm 3\sigma$ around its median; if σ is changed by less than 20% during that process, the field is considered not crowded, and the mean of the clipped histogram is taken as a value for the background. Otherwise, the mode is estimated with:
$Mode = 2.5 \times Median - 1.5 \times Mean$
(note that his expression is different from the usual approximation
$Mode = 3 \times Median - 2 \times Mean$,
e.g. Kendall and Stuart 1977, but was found to be more accurate with simulations).

SExtractor also allows the user to compute a local background, obtained through an average (with the above formula) in a rectangular annulus around the object. This additional procedure makes it possible to compensate for local deviations of the background (including those due to small-scale background fluctuations or nearby bright objects). This option is also coded in the CONVPHOT code (see Section 3.1.2).

1.2.2 Background estimation in crowded fields

Background subtraction in very crowded fields is more challenging. It is difficult to mask out sources and decide how much of the extended wings of their light to treat as background for neighbouring sources.

An iterative approach to deal with background subtraction in images for which we have prior information on source positions in high-resolution images has been developed at STScI to analyse CANDELS IRAC data.

First pass

The approach begins by applying a standard background-subtraction technique. This background estimate will be discarded and replaced by a better estimate in the second pass, so the details are not particularly important. The main goal is to remove large-scale gradients. The procedure is as follows:

- Identifying and masking bright sources
- Smoothing and subtracting the residual on large scales
- Identifying and subtracting fainter sources
- Smoothing and subtracting the residual on slightly smaller scales.

Second pass

Once this background is subtracted, the sources are fitted using TFIT or similar codes (see Section 3 below). The sources are subtracted and the pixels associated with the bright parts of sources are masked. This therefore presents an image that has had the lower-surface-brightness wings of the sources subtracted. The background in the source-free pixels is then measured by laying down square apertures of some number of entirely unmasked pixels (9x9 pixels works well for IRAC). A sigma-clipped mean of the pixel fluxes in these apertures is then estimated and the result is a sparsely-sampled irregular grid of "clean" background estimates scattered across the image. The background estimate for each pixel is then computed from this grid using an "N-nearest-neighbor" interpolation scheme (optionally weighted by a power of the distance to the N-th neighbor). For IRAC, we have found that inverse-distance weighting using the 19 nearest apertures works well. This approach is attractive because it is naturally adaptive. Where the density of background estimates is high, the background estimate is quite local (but still using scales much larger than typical sources), while it works acceptably well even near the image edges or where the density of background estimates is low.

2 SExtractor-based Aperture Photometry

Basic photometry of images with different resolution can be performed through an appropriate use of SExtractor as well as of other codes for aperture photometry (e.g. DAOPHOT, IRAF *digiphot* tasks.).

These techniques suffer from major shortcomings which will be discussed below in some details, but provide a useful reference to the principles of multi-band photometry.

To obtain consistent photometry from two images at different resolution it is necessary to perform measurements in the same physical region of each source. To this aim, as a first step, the high resolution image (HRI) must be smoothed with an appropriate kernel to match the PSF of the low resolution image (LRI). Publicly available PSF-matching codes are e.g. “PSFex” (Bertin 2011) and IRAF *psfmatch* task.

Normally, the process to obtain the kernel is based on a back-and-forward Fourier transformation technique. Since by definition

$$PSF_{LRI} = PSF_{HRI} * K$$

where * is the convolution operator, the Convolution Theorem gives

$$K = FT^{-1}\{FT\{PSF_{LRI}\}/FT\{PSF_{HRI}\}\},$$

where *FT* is the Fourier transform operator. A low passband filter should be applied in the Fourier domain to suppress high frequency fluctuations and remove the effect of noise, and the kernel should finally be normalized to unity, in order to preserve the flux of the objects.

Different choices of the measurement area can then be adopted. The choice is driven by several factors, in particular by image properties (crowding, depth, etc.) and by the unavoidable compromise between the quest for an high signal to noise (S/N) ratio (achieved in the central region of each object) and the need of avoiding systematic biases (which typically require using large apertures). Indeed, while the aperture diameter yielding the highest S/N can be analytically determined to be ~1.35 Full Width Half Maximum (FWHM) for point-like Gaussian sources, as far as extragalactic sources are concerned larger apertures can be preferred, in order to minimize systematic effects due both to uncertainties in astrometry and PSF-matching and to color-gradients and wavelength-dependent morphologies.

As a consequence, the measurement area is not generally expected to include all the flux in the LRI and the smoothed-HRI images, and the flux ratio measured from them is simply used as a "scaling factor" between the unknown total flux in the LRI and the total flux in the HRI, the latter being previously measured in large circular apertures or in scalable elliptical apertures (following e.g. Petrosian 1976 or Kron 1980):

$$F_{tot,LRI} = \frac{F_{LRI}}{F_{LRI,smoothed}} \times F_{tot,HRI} \; .$$

While aperture photometry has been often employed in the literature, on PSF-matched images it suffers from major drawbacks: PSF homogenization degrade photometric quality of the higher-resolution images, and it yields to pixel-to-pixel noise correlation and to flux contamination in crowded fields, as extragalactic deep fields usually are. Small size apertures can give acceptable and stable photometric quality for small/faint isolated objects but yield to significant biases in the photometry of extended sources having wavelength-dependent morphology. On the other hand, large apertures are more subject to flux contamination from neighbouring sources and give sub-optimal measurements in terms of S/N ratio.

2.1 Isophotal magnitudes

As already discussed, the segmentation process (Sect. 1.1) individuates the isophotal area of each object, i.e. the connected region of the image where pixel values exceeds a defined threshold with respect to the background.

As a result, a self-consistent flux measurement can in principle be obtained by the integral of pixel values within the isophotal area in the smoothed-HRI and in the LRI, with detection of the object performed in either of the two.

This approach has been previously adopted in the analysis of HST image datasets where small differences exist between the (nearly) diffraction-limited PSFs at different wavelengths (e.g. Galametz et al. 2013). As a matter of fact, the isophotal area obtained through image segmentation has no physical meaning when objects are unresolved, i.e. their size is lower or comparable to the PSF FWHM. For this reason, this approach is not effective to the analysis of most datasets where images have widely different PSFs, e.g. when the LRI is a ground-based optical image severely affected by atmospheric seeing, or mid/far-infrared images from space. In addition, it is well known that isophotal fluxes suffer from strong flux-dependent biases being the isophotal area of faint and low surface brightness objects a very poor indicator of their physical size.

2.2 PSF-matched aperture magnitudes

A simpler but more versatile approach exploits pre-determined circular or elliptical apertures (Fig. 2).

Fluxes are usually measured in circular apertures chosen to include >70% of the flux from point-sources (e.g. Castellano et al. 2010, McLure et al. 2013). Depending on image crowding and on astrometric and PSF-matching accuracy flux ratios can also be measured in small scalable elliptical apertures (e.g. Bouwens et al. 2011).

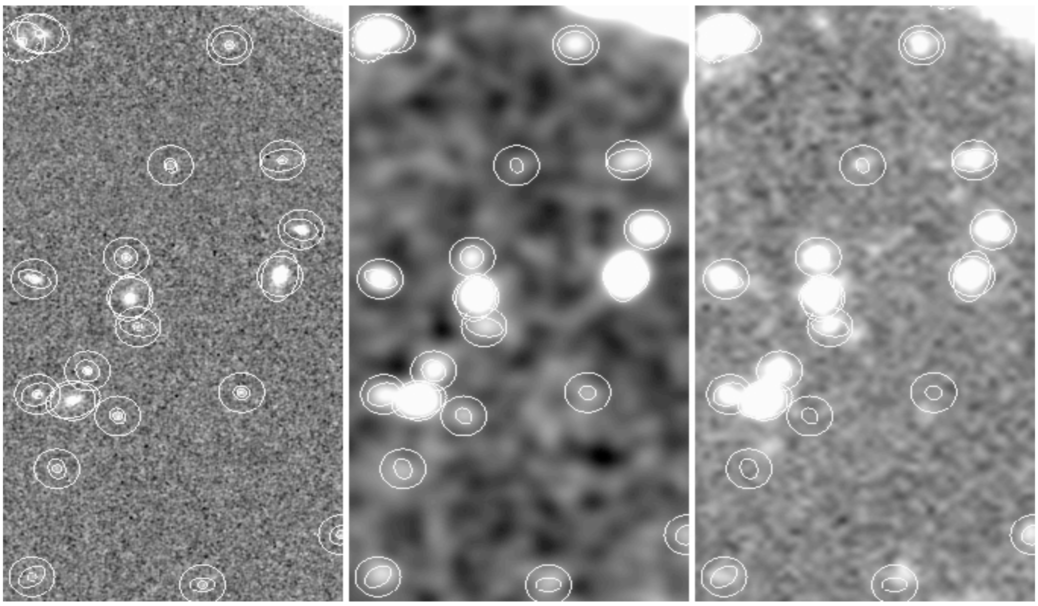

Fig. 2. Example of photometry on PSF-matched images, from left to right: a) original high resolution image (F160W band, WFC3-HST), b) ground-based low resolution image (B band, SUBARU-Suprime Cam), c) HRI smoothed to the LRI PSF. Flux can be consistently measured in b) and c) within the circular and elliptical regions superimposed on the images.

3 PSF-matched photometry using priors

This method is essential to compare and combine observations from different wavelength domains, obtained using both space-based and ground-based imaging instruments with very different sensitivities, pixel scales, angular resolutions, and survey depths (e.g., Agüeros et al. 2005; Obrić et al. 2006; Wilson et al. 2007). As pointed out in the previous Section, aperture photometry in a low-resolution image at positions measured in a corresponding higher resolution image is frequently biased by neighbour contamination for reasonably sized apertures. On the other hand, performing source extraction on both images and matching the resulting catalogs is compromised by the inability to deblend neighbouring objects. PSF-matching techniques that degrade high-resolution data to match the lowest resolution data discard much of the valuable information obtained with observatories like HST, reducing all images to the “lowest common denominator” of angular resolution. Finally, crowded-field, PSF-fitting photometry packages such as DAOPHOT perform well if the sources in the low-resolution images are unresolved but are unsuitable for analysis of even marginally resolved images of extragalactic sources.
Therefore, it is of great advantage to use the *a priori* knowledge of the existence, positions and morphologies of sources in a deep HRI (“*priors*”)to improve photometric measurements of objects in a corresponding LRI of the same field. Given the HRI and LRI of the same galaxy, the spatial distribution of the light from the HRI can be used to constrain the photometric measurement of the LRI. This can be achieved either using real object profiles or using idealized models of the objects fitted to the sources; each of the two approaches has pros and cons (see below for further description and discussion on this issue).

Like in the aperture photometry approach (Sect. 2), a convolution kernel selected to match the PSF of the HRI with the PSF of the LRI is required, but this time to singularly convolve each prior at once.

Then, using a catalog of the HRI sources and their WCS information to locate them with respect to the LRI, a linear system can be created:

$$I = B + F_1 P_1 + \ldots + F_{N_{obj}} P_{N_{obj}} \ ,$$

where I contains the pixel values of the flux in the LRI, P_i are the convolved models (“templates”) for the N_{obj} objects in the (region of the) LRI being fitted (their fluxes having been normalized to 1), F_i are the multiplicative scaling factors for each object, and B is an additive constant. In physical terms, F_i represent the fluxes of each object in the LRI, and B allows for a possible background component in the image.

Once the templates for each object in the (region of the) LRI have been generated, the best fit to their fluxes can therefore be derived by minimizing a χ^2 statistic,

$$\chi^2 = \left[\frac{\sum_{m,n} I(m,n) - B_i - M(m,n)}{\sigma(m,n)}\right]^2$$

where m and n are the pixel indexes,

$$M(m,n) = \sum_{i=1}^{N_{obj}} F_i P_i(m,n) \ ,$$

and s is the RMS value in the pixel.
The output quantities are the best-fit solutions of the minimization procedure, i.e. the F_i parameters and their relative errors. They can be obtained resolving the linear system

$$\frac{\partial \chi^2}{\partial F_i} = 0 \left(i = 1, \ldots, N_{obj} \right).$$

Since the P_i model for each object is normalized to unit flux, the resulting total magnitude of the *i*-th object in the LRI is simply

$$-2.5 \log \left(F_i \right) + ZP_{LRI}$$

where ZP_{LRI} is the zero-point of the measure image. It has been shown that this total magnitude is a reliable measure of the actual total flux of the objects, somewhat less prone to systematic effects than e.g. the Kron magnitudes computed by SExtractor (Laidler 2007).

Although in practical cases the number of free parameters (i.e., objects) can be quite large, the linear system to be resolved is very sparse, since non null off-diagonal terms represent only the rare overlapping/blended sources, and the minimization can be performed in a quite efficient way by using standard numerical techniques.
Note that this method may assign non-physical negative fluxes in some particular cases, in order to optimize the fit (this occurs for low-confidence detections in the LRI, or for objects in the vicinity of a bright, poorly modeled object; this most likely occurs when the PSF transfer kernel is not accurately determined, or possibly when there are real, intrinsic morphological differences in the galaxy between the wavelengths of the HRI and LRI). Such negative fluxes might be removed in a post-processing phase.
Formal errors are usually assigned to the computed fluxes by taking the square root of the variance of the fitted parameters *Fi*, which serves as the uncertainties in the flux.
As already pointed out in Sect. 2, the magnitudes obtained this way can hardly be compared with the SExtractor magnitudes of the HRI image, so that reliable colors cannot be directly obtained. A robust color estimation should be carried out on a *same* area of each object and possibly be extended on a large region of the sources, in order not to be biased by red nuclei or by strong color gradients. At this purpose, as discussed above, the fluxes *Fi* should be compared with the fluxes determined by the same method on the HRI.

The whole method relies on the following assumptions:

1. The HRI, used to detect objects, isolate their area and to obtain models of the sources, should be deep enough to allow all the sources in the LRI to be detected. Ideally, such image should be well sampled and allow a proper resolution for most sources.

2. The PSF must be accurately estimated in both images, and a convolution kernel must be obtained to smooth the HRI to the PSF of the LRI. This is a key issue which deserves particular attention, since even small inaccuracies in the definition of a PSF or in the derivation of the kernel may end up in large errors in photometry.

3. The objects are considered to have no measurable proper motion.

4. Morphology and positions of the objects should not change significantly between the two bandwidths (HRI and LRI). This is obviously not always the case, and may lead to

significant errors. This is particularly true for far-IR images obtained with Herschel.

5. Finally, the objects should be well separated in the HRI, although they may of course be blended in the LRI.

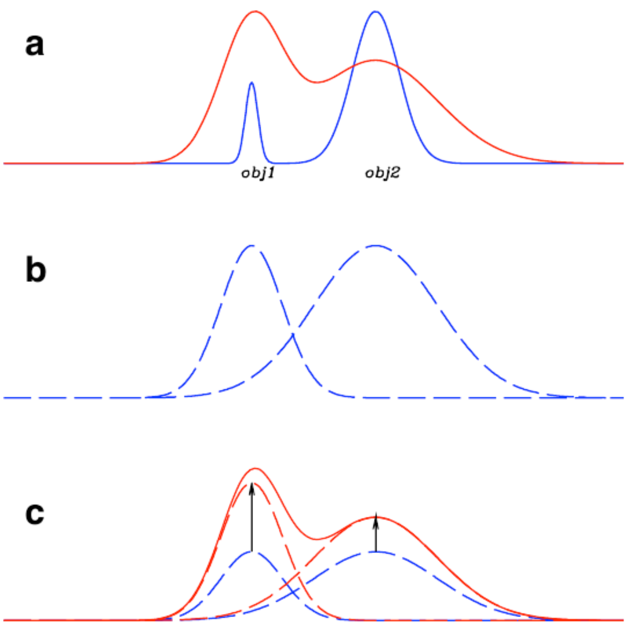

Fig. 3. A schematic representation of the PSF-matching algorithm. (a) Two objects are clearly detected and separated in the HRI (blue, solid-thin line). The same two objects are blended in the LRI (red, solid-thick line) and have quite different colours. (b) The two objects are isolated in the HRI and are individually smoothed to the PSF of the LRI, to obtain the "model" images. (c) The intensity of each object is scaled to match the global profile of the measure image. (from De Santis et al., 2006)

There are two alternative methods to obtain priors for the positions and morphologies of the objects in the HRI. The first possibility is to use the real images of sources, directly cut from the HRI, and straightforwardly smooth them down to the resolution of the LRI. The second one is to obtain analytical models of sources.

3.1 Tools based on real profiles

This approach has the advantage of not relying on any a priori assumption on the features of the sources, but requires a detailed preparation of the images (e.g., they should be reduced to the same pixel scale before processing them).

Beginning with the catalog of *(x,y)* positions, local backgrounds, and isophotal fluxes, along with a segmentation map that have been obtained by processing the HRI, a library of "*cutouts*" is constructed, each a subset of the HRI containing a single object. Pixels outside the segmentation map for each object are set to zero flux. The background can be subtracted from each of these cutouts, and, for convenience, each cutout can be also normalized to unit flux.

These "real" priors are then convolved to the spatial resolution of the LRI, using the previously obtained kernel. It is assumed that these model template images are "truth" images for the objects in the low resolution frame.

There are three available software that follow these guidelines, namely TFIT, CONVPHOT and the McLure et al. code.

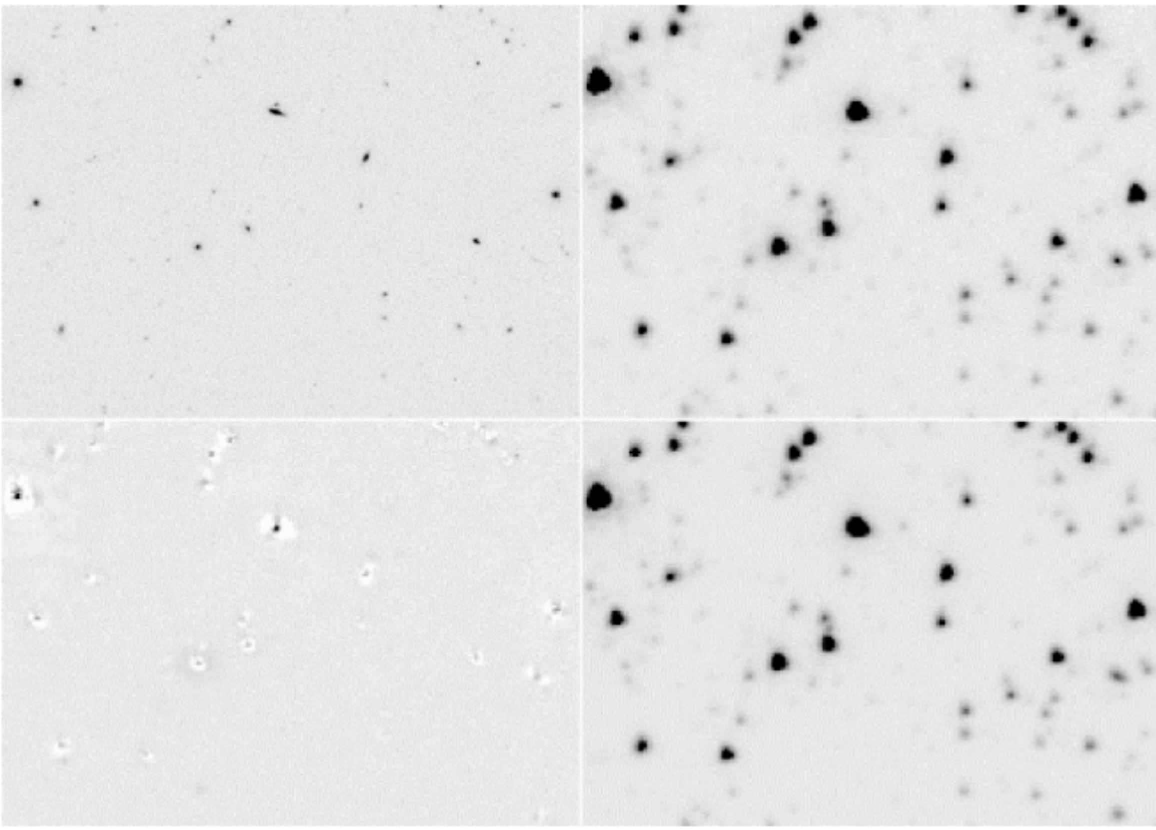

Fig. 4. This figure shows (*clockwise from top left*) *HST* ACS *z*850 data (the HRI), SSC IRAC 3.6 μm data (the LRI), a model image (a reconstruction of the LRI constructed by creating a collage of the object templates, scaled by the flux measurement for each object) and the residual image (constructed by subtracting the model image from the actual LRI). Inspection of the residual image can reveal objects that had no counterpart in the high-resolution image, as well as any problems with misregistration or the transfer kernel. All images are displayed at the same scale. Imperfections in the residual image are due to inaccuracies in the shape of the *Spitzer* IRAC PSF. (from Laidler et al. 2007)

3.1.1. TFIT

TFIT (Laidler et al. 2007) was initially conceived and written by C. Papovich in 1999 as a C++ code.

In its latest version, TFIT consists of a Python envelop performing the entirety of tasks, except for the fitting core routine, which is based on the original C++ routines. It requires the pre-installation of the following software: Python (and some of its standard modules/interfaces, such as *numpy, scipy, matplotlib, anfft, pyraf*); CFITSIO libraries; STSCI; IRAF; STSDAS; FFTW3.

The HRI and the LRI must be aligned and their pixel scales must be the same or have integer ratio (these requirements can be accomplished for example using SWARP, Bertin et al. 2002, and/or the IRAF tasks CCMAP and SREGISTER).
TFIT can subtract a constant background (given in the input catalog), but it is instead recommended to perform a detailed background subtraction *before* running TFIT, thus putting the term *B* equal to zero in the linear system to be solved.
The preliminary steps (cutting out of priors and their convolution with the kernel) are performed in "stages", in which all sources are processed before passing at the following step. The cutout stage is performed invoking the IRAF task *imcopy*, via *pyraf* call. The convolution stage is performed using a Fast Fourier Transform routine (FFTW3), invoked via the *anfft* Python module (in TFIT original release, the convolution was instead performed via straight pixel-by-pixel summation). Cutouts and templates are stored as FITS files, generally requiring quite a large amount of memory.

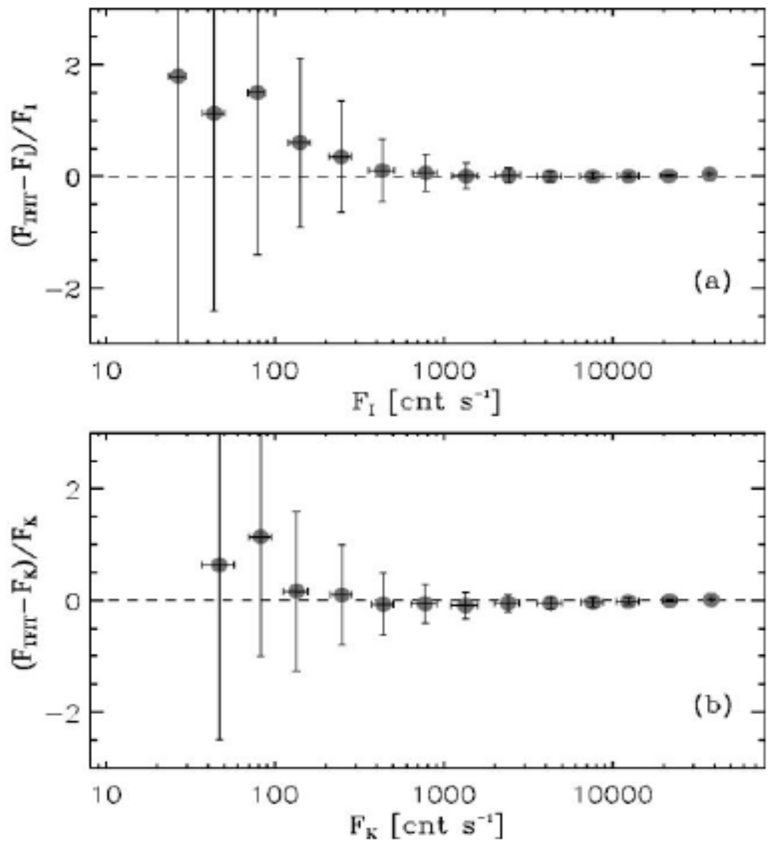

Fig. 5. Comparison between the SExtractor isophotal flux (FI; top) and the SExtractor Kronlike flux (FK; bottom) from the high-resolution image vs. the flux derived by TFIT from the low-resolution images for ≈1300 sources in a simulated images. (a) Flux difference, (FLRI ⌅ FI)/FI , vs. isophotal flux from the high-resolution image. The sources have been binned in intervals of quarter magnitudes, and the error bars indicate the standard deviation for the sources in that bin. For this simulation, the standard error in the mean is one-tenth the standard deviation. (b) Same as (a), except TFIT fluxes are compared to the Kronlike fluxes of SExtractor.

In TFIT the linear system is not built on the LRI as a whole at once. Rather, the LRI is divided into an arbitrary grid of “cells”, and a linear system is built and solved for each of these cells. The typical dimension of a cell should be more than say 30 times the LRI FWHM (a wrong choice of the cell dimension may lead to catastrophic errors); the grid is constructed in such a way that cells overlap, and each cell is then expanded once to completely contain the sources which partly fell into its original dimensions. Moreover, a second fitting run is performed using a shifted (dithered) grid. In this way, each source is fitted more than once (usually at least 4 times for a standard choice of the dimensions of cells). At the end of the fitting procedure, the “best choice” for each source is selected picking the fit obtained within the cell in which the object is at the minimal distance from the geometrical center.
This method strongly reduces the computational time, because the solution of a very large sparse linear system is usually computationally far more expensive than the solution of a large number of small ones. On the other hand, it introduces some degrees of arbitrariness in the procedure, which may lead to systematic inaccuracies in the determination of the fluxes. Further analysis is being performed on the issue.

The linear system solution is performed by the C++ core code, via LU decomposition of the matrix (in the TFIT original version, the Singular Value Decomposition method was adopted instead). No constraint is imposed on the fitting constants, so negative fluxes can be found and kept as "right" solutions.

Diagnostics and error estimates are also computed during this stage, and outputted both numerically and graphically (i.e., covariance matrix - whose diagonal consists of the squares of the errors to be assigned to each source - and residual images).

After the first fitting procedure is completed, a "collage model image" is constructed using the templates, each multiplied by its fitting constant. Then, each region of the LRI is cross-correlated with the corresponding region of this collage image, and the resulting "best fitting shifts" in *x* and *y* are found. Finally, a set of shifted convolution kernels which maximize the correlation for each region are produced. All of these procedures are performed invoking IRAF tasks, namely *xregister* and *imlintran*.

The stages of convolution and fitting are then repeated from scratch, this time degrading each HRI cutout using the new "shifted" convolution kernel found for the region of the LRI to which it belongs. In this way, a higher degree of precision is obtained in the astrometry registration of each source.

A complete (double) run on a standard astronomical field (say, one Goods-S Hawk-I field) requires ca. 48 hours on a standard machine, this estimate strongly varying depending on the number of sources in the HRI catalog falling outside the LRI (these are cut and convolved anyway, and later excluded from the minimization procedure). This long executing time is largely due to the slowness of the Python procedures.

TFIT is currently being reviewed and improved, and a new version is expected to be released in the next months.

3.1.2. CONVPHOT

CONVPHOT (De Santis et al. 2006) is a C code developed at INAF-OAR. Its only pre-requirement is the CFTISIO library.

CONVPHOT can compute the background for each source, and subtract it on-the-fly. The pixel scale of the HRI and the LRI must be the same; the pixels offsets between HRI and LRI are needed as an input.

The preliminary steps (cutout and convolution) are organized as a unique flow for each source, which is located, extracted, cut and convolved on-the-fly. The convolution is performed via straight pixel-by-pixel summation. Cutouts, templates, and also segmentations for each object can be stored as FITS files (requiring a large amount of memory), but the code may be run without storing images.

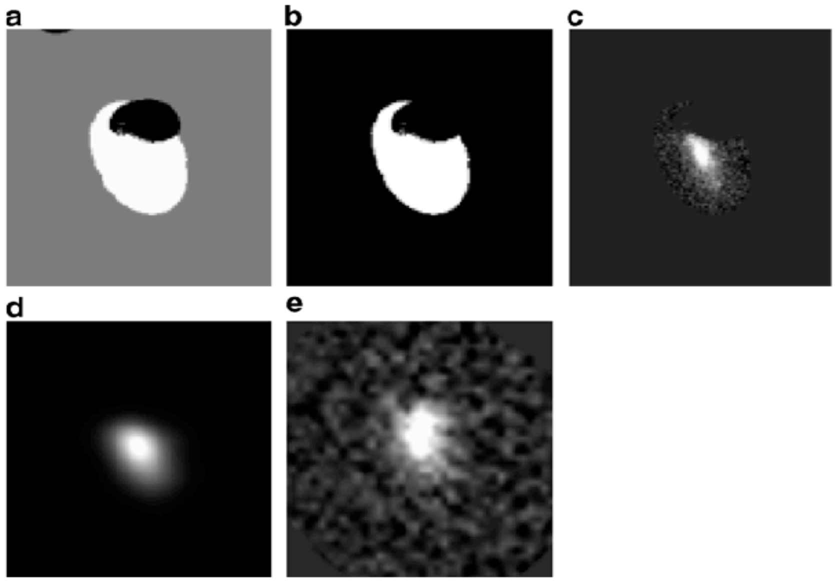

Fig. 6. Example of the processing steps to create the thumbnail of the model profile of a given object. From left to right: (a) the segmentation of the object (white) is extracted, (b) other objects are masked, (c) the object profile is extracted from the detection image and the local background is subtracted, (d) the object profile is smoothed to the measure PSF and normalised to obtain the model profile and (e) the same object is extracted from the measure image subtracting its local background. (from De Santis et al. 2006)

The minimization procedure is performed on the LRI as a whole at once. This has the obvious advantage of being fully self-consistent, excluding the introduction of possible arbitrary biases. On the other hand, if the LRI is large the method is highly demanding both in terms of RAM memory and of computational time. Indeed, the total computational time is comparable to the TFIT one, but in CONVPHOT most of the time is spent in the fitting routine, another source of slowness being the pixel-by-pixel convolution instead of the FFT one.

Outputs include fluxes, errors and diagnostic images, as for TFIT.

CONVPHOT does not include a double-pass procedure with registered kernels. On the other hand, it allows for the possibility of fitting each source using only the pixels with fluxes above a relative threshold $t_i = P_{i,\mathrm{MAX}} * t_f$, where $P_{i,\mathrm{MAX}}$ is the normalized model profile maximum of the i-th object, and t_f is an input threshold parameter. This should avoid spurious fits caused by the wings of the PSF.

Moreover, CONVPHOT allows for the iterative exclusion of sources with resulting strong (say, more than 3σ) negative flux, ruling them out after a first fitting pass and repeating the whole procedure without taking them into consideration.

3.1.3. McLure et al. code

McLure et al. (2011) developed a code to analyse the ultra-deep IRAC imaging available in the Hubble Ultra Deep Field (HUDF), based on using HST H-160 band imaging as prior information. The basic algorithm is very similar to that employed by TFIT and CONVPHOT. It is written in Fortran, it allows for FFT convolution of high resolution cutouts, and it uses a standard Gauss-Jordan elimination routine to solve the linear system. The images are fitted as a whole without cells subdivision.

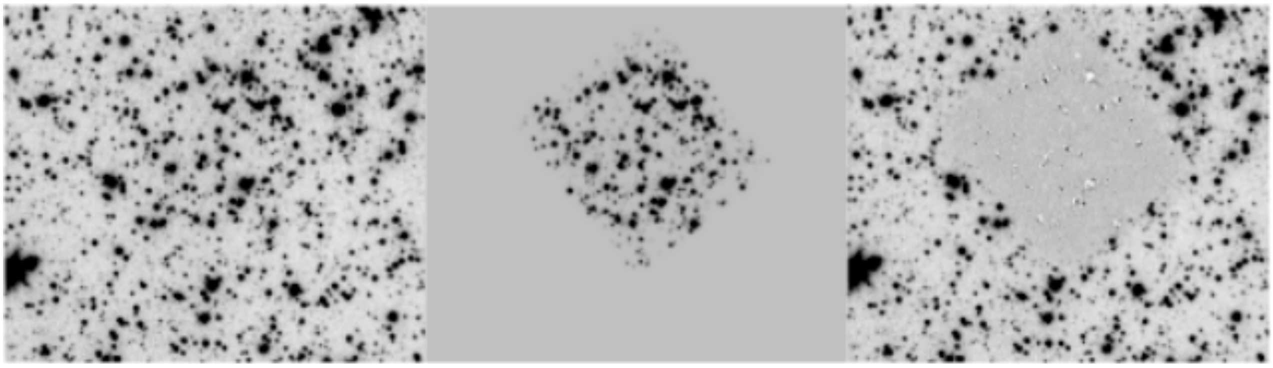

Fig. 7 - Illustration of the IRAC deconfusion algorithm developed by McLure et al. (2011). The left-hand panel shows the inverse-variance weighted stack of the epoch1+epoch2 4.5 μm imagingcovering the HUDF. The middle panel shows the best-fitting model of the IRAC data, based on using the H160 WFC3/IR imaging to provide model templates, and a matrix inversion procedure to determine the best-fitting template amplitude. The right-hand panel shows the model subtracted image (note that the WFC3/IR imaging does not cover the full area of the HUDF).

3.2 Tools based on parametric fit

The other possible approach to the problem of obtaining reliable priors from the HRI involves a parametric fit of the detected sources before the convolution to the low resolution. It has been a common exercise, especially after the advent of CCD cameras, to fit the observed surface brightness profiles of galaxies by mathematical functions in order to shed light into the structure of the different galaxy components and their connections. To this end, a mathematical form

$$I = I(P_1, \ldots, P_n, x, y)$$

is usually adopted, where I is the surface brightness distribution of the galaxy, ($P_1, \ldots, P_n$) are the parameters which describe it and x,y are the spatial coordinates of the image. So, in general, the light distribution can be written as

$$I(P_1, \ldots, P_n, x, y) = \sum_{i=0}^{N} I_i(P_{i,1}, \ldots, P_{i,n}, x, y)$$

where the indices i account for the N components that pertain to the galaxy. There are a number of empirically motivated light distribution laws in the literature to be used for the various galaxy constituents, e.g. the $R^{1/4}$ law (De Vaucouleurs 1948) for ellipticals and spiral bulges, the exponential law (Freeman 1970) for disks, etc. However, the preferred option nowadays is usually the Sérsic (1968) function

$$I = I_e \exp\left\{-b_n\left[\left(\frac{R}{a_e}\right)^{1/n} - 1\right]\right\},$$

where I_e is the intensity at the effective radius, R is the radial coordinate and a_e is this effective radius along the semimajor axis enclosing half of the flux from the model light profile. The quantity b_n is a function of the radial shape parameter n (the so-called the Sérsic index), which defines the global curvature in the luminosity profile, and is obtained by solving the expression $\Gamma(2n) = 2\gamma(2n,b_n)$, where $\Gamma(a)$ and $\gamma(a,x)$ are, respectively, the gamma function and the incomplete gamma function. The Sérsic law is a generalization of the previously

mentioned De Vaucouleurs (case *n=4*) and Freeman (case *n=1*) profiles. It is also noteworthy that, by fitting several Sérsic profiles at the same time, it is possible to obtain bulge-disk decompositions or more elaborated luminosity profiles. To define unambiguously a Sérsic function in two dimensions, we need also to supplement the axis ratio – ratio of the semi-minor over the semi-major axis – and position angle of the galaxy.
Using parametric fits as priors has the following advantages:

- in principle, the only limitations of this procedure are the ones given by the galaxy flux and the sky noise. In fact, low S/N sources are better modeled using these methods as the importance of noise in their light surface brightness profiles is inherently reduced;
- it is possible to take into account the light contamination by galactic neighbours, which minimizes the impact of crowding;
- the knowledge of the galaxy profiles can be extrapolated to distances that are hidden underneath the sky noise;
- by having the description of the galaxy light, one can always "play" with these infinite resolution models, convolving them with her favourite choice of PSFs, pixel scales and/or any other observational conditions.

On the other hand, it must be kept in mind that the method relies on the assumption that the galaxy light can be described by a given analytical function (or some combination of functions), which may not be the case, as these functions might be inaccurate or depart from the assumed profile at galactocentric radii not proven by the images. Moreover, the analytic functions are 2D symmetric, and it is well known that real galaxies have asymmetries (fans of stars, shells, tidal features, …) which cannot be properly model using this approach.

There exists a plethora of software packages to fit mathematical laws to the galaxies' 2D luminosity profiles (e.g., GIM2D, Simard 1998, Simard et al. 2002; 2DPHOT, La Barbera et al. 2008; GASPHOT, Pignatelli, Fasano & Cassata 2006; etc.), being GALFIT (Peng et al. 2002, 2010) the most widely utilized. GALFIT convolves Sérsic $r^{1/n}$ 2D models with the PSF of the images and determines the best fit by comparing the convolved model with the observed galaxy surface brightness distribution using a Levenberg-Marquardt algorithm to minimise the χ^2 of the fit. It has a series of auxiliary functionalities such as image masking, customizable initial conditions and functions, etc. which makes it very attractive to the final user. The executable program is distributed for free but the source code is not publicly available.

3.2.1 MEGAD (Multi-cpu Edinburgh Galfit-based Algorithm for Deconfusion)

MEGAD (Buitrago et al. 2014, in preparation) is a Multi-CPU IDL-coded tool based on some of the previously mentioned software packages (IRAF, SExtractor and GALFIT), and it is easily configurable according to the final user's requirements (it can be fed with any kind of image or PSF, and it is also possible to change any of the fit parameters, i.e.: functions to be fit, structural parameters' initial guesses and ancillary data to improve the results). It does not consist of a single wrap-up program but a combination of numerical routines. However, it comes with a series of added-value programs to help the user creating the necessary inputs at each step of the execution and also for building catalogs storing the output information.
As for the previously described codes, the LRI and the HRI must be aligned and have the same pixel scales. MEGAD automatically divides the galaxies' analysis into as many directories as the number of used CPUs. Independent directory trees are also created within these folders in order to keep track of each individual analysis, in case the user would like to refit/pay more attention to any interesting object. SExtractor outputs are used as in the previously described codes to obtain cutouts of all target galaxies. Such cutouts are then fitted using GALFIT, with

the following conditions:

1. SExtractor output parameters are used as initial guesses;
2. the value of the sky flux is let free to vary;
3. masks are generated for distant neighbours, but any galaxy object close to the target galaxy will be analysed at the same time to take into account its light contribution. Usually an extra mask is built, which only covers the very central object, for improving the analysis in very crowded fields;
4. several PSFs could be used in order to take into account the spatial variation of this parameter.

From the various output files per object at study (one per given mask and PSF combination) the best fits are selected as those with the best reduced χ^2 values, and their structural parameters are the ones which will move forward in the analysis. As stated, every time GALFIT is used all its intermediate files are kept in case any object fit is worth a closer revision.

At this point, catalogs with the best outcomes must be created (in a fixed format) to feed the next step of the algorithm (software packages to make things easy for the user are available). In addition, the user is supplied with a program to create a total model image where every galaxy model is convolved with its best fit PSF. By subtracting it to the original HRI, the user can check visually how good the fits were.

Finally, MEGAD takes the final master catalog from the previous step and fits the counterparts of the galaxy stamps in the LRI. This is done freezing all the structural parameters of the 5σ detection galaxies in the HRI except their magnitudes. No masks are used now, and a zero sky is assumed. Again, all these possible configurations might vary, being a typical change at this stage to allow the galaxy centroids to move ± 1 pixel (at the cost of computational time). As it happened for the HRI, several PSFs are usually taken into account to reflect the variation of this parameter within our imaging.
As in the previous step, software to produce a final master catalog and a total model and the associated residual image are provided.

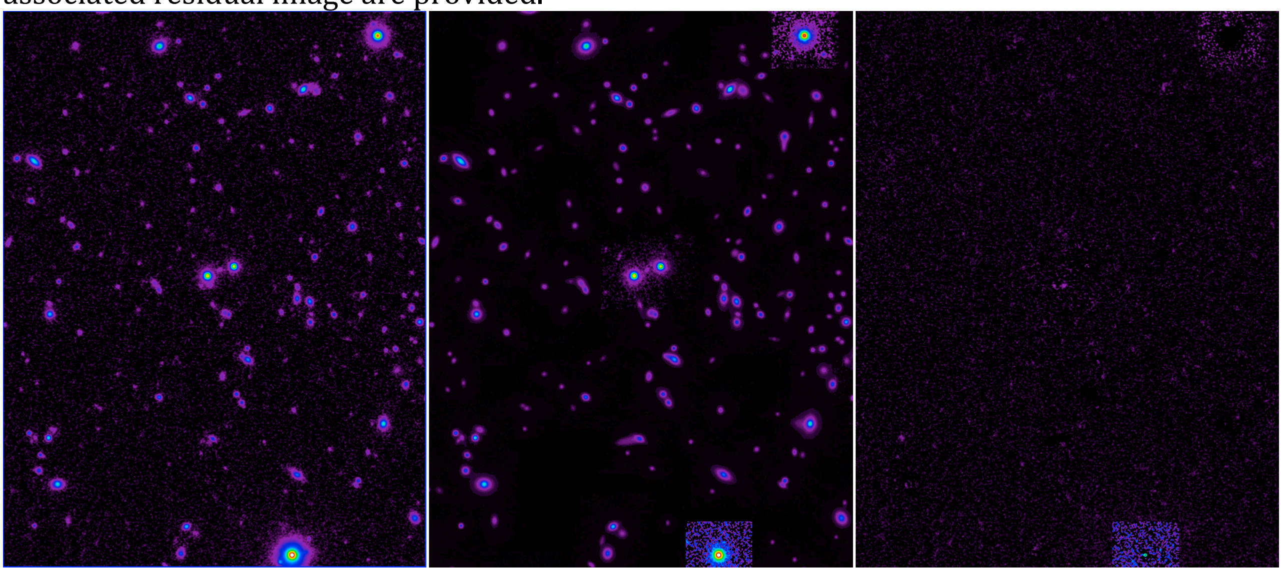

Fig. 8 - Example of how MEGAD works. From left to right, original central part of the ULTRAVISTA K-band image, model and residuals (all to the same color scale). Small remaining defects are the presence of objects barely above the sky noise or the couple of minor galaxies outshined by the central star which SExtractor was not able to identify as individual objects.

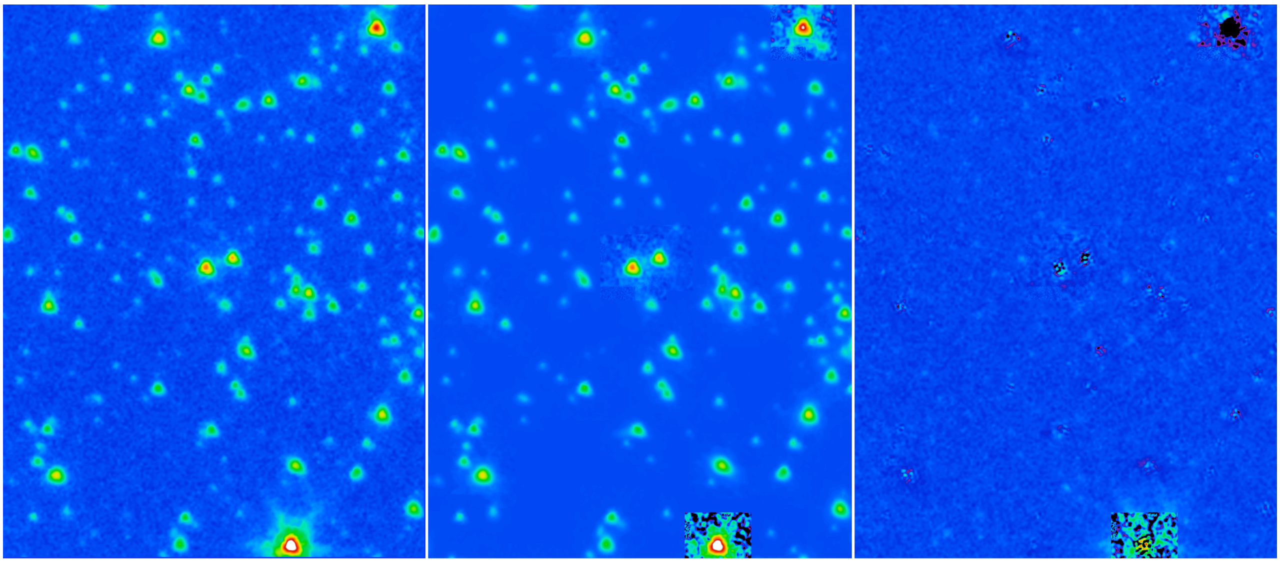

Fig. 9 - Same are as in Fig. 4, but this time the MEGAD analysis was done using the SCOSMOS survey 3.6 microns image. From left to right: original image, model and residuals (all to the same color scale). K-band fits (the previous figure's models) were the input for these IRAC analyses. Of course, objects not modeled previously do not appear in the final models.

3.2.2 PyGFIT

PyGFIT (Mancone et al. 2013) differs in a number of steps and in the way they are implemented from MEGAD, although the underlying philosophy is very similar. To begin with, it is written in Python and, although it also internally uses SExtractor, it has its own Levenberg-Marquardt algorithm to conduct the galaxy fits instead of relying on GALFIT for that.

As an input, PyGFIT needs to receive single Sérsic fits to the HRI, for example using the aforementioned GALFIT. Then, SExtractor is run over the LRI with a two-folded purpose: first, to create a segmentation map to determine the blended objects; second, to calculate a background map to be subtracted to the LRI in order to account for the sky noise. The next step consists of the alignment between the HRI and LRI, to correct by any offset between this two. To this end, it finds isolated objects and calculates the offset via a least squares minimization.

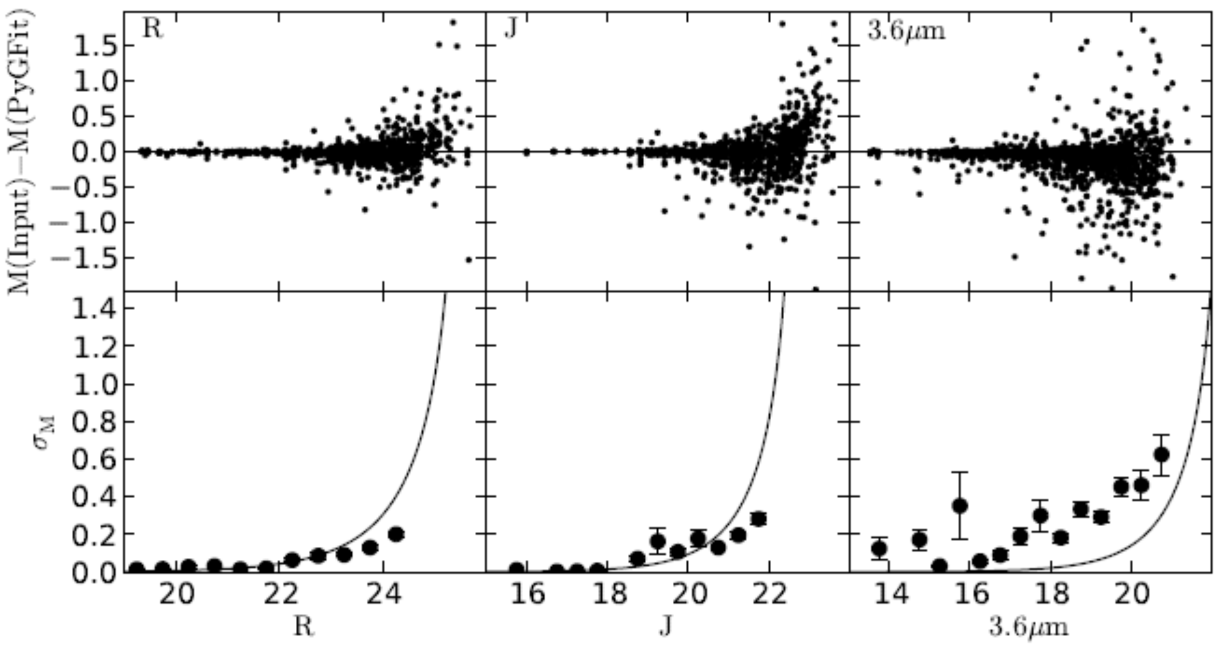

Fig. 10 - These plots are extracted from the Figure 5 in Mancone et al. (2013). On the top row, they are shown the differences between the magnitudes of simulated galaxies and the PyGFIT outputs for three photometric bands. On the bottom row, the magnitude errors are displayed according to the total magnitude. The error bars depict the standard deviations as a function of aperture magnitude. The solid line corresponds to the limit of Poisson sky noise in the absence of crowding for 4 arcsec apertures.

Once this is done, it proceeds to cut square stamps large enough to contain the full segmentation region of the low resolution source, plus the allowed position shift during the fit and also adding two extra pixels. Then, it automatically finds the overlapping objects in the HRI and starts the fit. It is noticeably to state that it starts from the brightest low resolution source to the faintest. During the fit, all the structural parameters remain fixed, except the fluxes and the centroids, allowing a ± 1 pixel change in its position. After each object is fit, its model is subtracted from the LRI to remove any contribution to nearby objects.

The final output consist of a master catalog and a residual image. PyGFIT has a built-in capability to quantify errors by conducting simulations on artificial galaxies. The authors claim it produces good results, although the uncertainties increase as a function of neighbour distance and neighbour brightness.

For a more detailed information, please consult the webpage http://www.baryons.org/pygfit .

3.2.3 MegaMorph

MegaMorph (Häussler et al. 2013, and Vika et al. 2013) is a multi-purpose galaxy fitting software that, even though not primarily developed for attempting galaxy PSF-matched photometry, can be used for this task. The program is written in IDL and uses SExtractor and GALFIT, much as MEGAD does. Nevertheless, GALFIT is modified for specific purposes (referring as GALFITM to this new version in MegaMorph documentation) and it is utilized by a wrap-up package called GALAPAGOS (Barden et al. 2012) which manages object deblending, nearby object contamination, which object should be processed next and other technicalities.

The concept that distinguish MegaMorph from the other codes based on parametric fits is that, using the various photometrical bands to obtain information for fitting the lower signal-to-noise images, it provides the possibility to change the structural parameters with wavelength, allowing for morphological k-corrections in a controlled manner. For ASTRODEEP, this approach may be useful for exploring the LRI permitting a change in the galaxies' structural parameters.

The core of the MegaMorph algorithm consist of replacing every galaxy model with a wavelength-dependent function (following the same annotation convention as in the beginning of this Section 3.2)

$$I(P_1,\dots,P_n)=\sum_{i=0}^{N} I_i\left(\tilde{P}_1(\lambda;q_{1,1},\dots,q_{1,m_1}),\dots,\tilde{P}_n(\lambda;q_{n,1},\dots,q_{n,m_1}),x,y\right)$$

where λ is the wavelength, and P_j-*tilde* are the functions, with m_j variables called $q_{j,k}$ that describe the variation of the model parameters. MegaMorph assumes that these parameters change according to Chebysev polinomials T_k:

$$\tilde{P}_j(\lambda;\{q_{j,k}\})=\sum_{j=0}^{n} q_{j,k}T_k[z(\lambda)],$$

being z the variable which changes with the wavelength. The reason behind k ranges from 0 to m_k resides in the fact that the each parameter may change in a different way: for example, in the case of the Sérsic function the P_j-*tilde* parameters would be magnitude, effective radius, Sérsic index, axis ratio and position angle (*j=5*, although in reality we need to add the position of the galaxy's centroid). One can assume 0-th order of change for the first parameter (so the magnitude must remain constant among the different bands in this case), a quadratic change for the effective radius and so forth. The user is allowed to choose which kind of variation is permitted per parameter, but the authors recommend to use an order one less than the number of photometric bands utilized.
All the free parameters $q_{j,k}$ of the model are then fitted to the multi-band data simultaneously by minimizing the usual χ^2 equation

$$\chi^2=\sum_{u,v,w}\frac{\left[Data_{u,v,w}-Model\left(P_j\left(\lambda_w;\{q_{j,k}\}\right),x_u,y_v\right)\right]^2}{\sigma^2}.$$

For a more detailed information, please consult the webpage http://stevenbamford.com/research/projects/megamorph/.

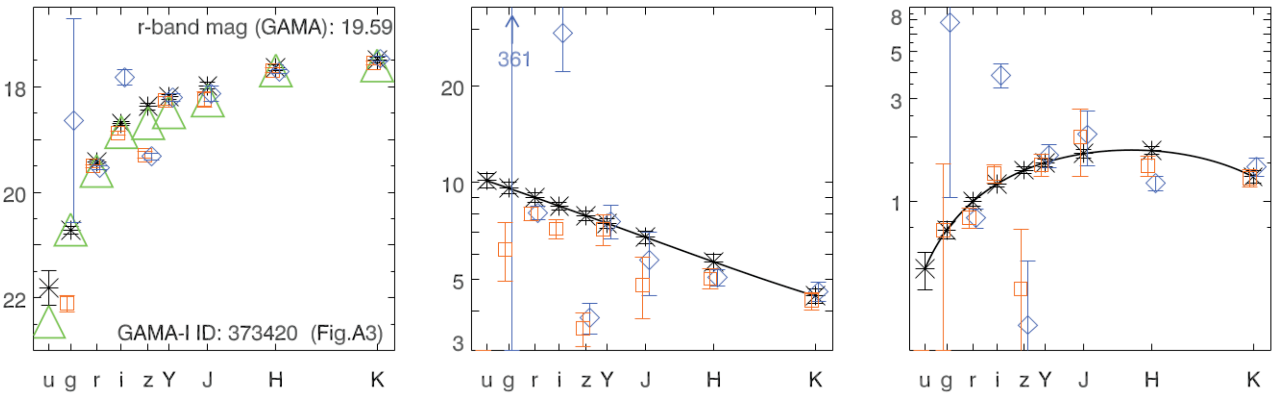

Fig. 11 - Plots taken from Häussler et al. (2013) Figure 1. From left to right: magnitudes, effective radii (in pixels) and Sérsic indices for the galaxy GAMA-I373420 (cf. GAMA survey, Driver et al. 2011). MegaMorph outputs are the black stars, while

green triangules are SExtractor MAG_AUTO results, orange squares are single band fits in Kelvin et al. (2012) and the blue rhombuses are the authors' single fits. The solid lines represent the polynomial variation of the structural parameters and the errors bars come from GALFIT results. It is conspicuous that MegaMorph reduces the multiband scatter on the outputs.

4 PSF-fitting photometry on mid to far infrared images

4.1 "Blind" techniques

STARFINDER

STARFINDER (Diolaiti et al. 2000a,b) is an IDL program for stellar fields analysis. The code starts detecting a catalog of presumed stars above a pre-fixed threshold in the background-removed image. Then, the images of the presumed stars are analysed in order of decreasing luminosity, and a synthetic model image of the observed field is constructed by placing an intensity- scaled PSF template in the position of each identified star. The PSF template can be extracted directly from the field, or constructed accurately by means of other methods. The analysis of a given object includes the following operations (in iteration):

1. re-identification: subtraction of already known stars, to reject spurious detections associated to PSF features of brighter sources;
2. correlation check, to measure the similarity of the object with the PSF; a correlation threshold is fixed to distinguish between stellar-like objects and spurious detections (e.g. noise spikes);
3. fitting: to determine position and flux, taking into account the contribution of neighbouring brighter stars and of the local background;
4. updating of the stellar field model, which contains a replica of the PSF for each detected star; it is basically used to keep track of the contribution of bright sources when analysing fainter ones.

If the extension of a given detected star results to be significantly larger than the PSF area, the object is assumed to be a blend and de-blending is attempted by iteratively searching for statistically significant residuals and subsequent fitting. Note that the method relies on the assumptions of accurate knowledge of the PSF, and the flux errors are artificially small. STARFINDER has been applied to produce the blind catalogs for the *Herschel* PEP survey, and its performance in comparison to the PSF fitting using 24mm priors is presented in PEP-GOODSH data release. For a more detailed information, please consult the webpages http://www.mpe.mpg.de/resources/PEP/DR1_tarballs/readme_PEP_global.pdf; http://www.mpe.mpg.de/resources/PEP/DR1_tarballs/readme_PEP_SPIRE.pdf.

GETSOURCES

GETSOURCES (Men'shchikov et al. 2012) is a multi-resolution blind source detection algorithm developed primarily for large far-infrared surveys of galactic star-forming regions with Herschel, but which can also be applied to deep extragalactic surveys. GETSOURCES analyses filtered single-scale decompositions of detection images over a wide range of spatial scales (separated by only 5% in dimensions) and across all wavelengths, thus linking the information over a large dataset at various levels of resolution. This is particularly important for detecting and measuring sources in Herschel images, which have angular resolutions differing by a factor ~7, the coarsest beam being FWHM~36″.

The procedure is as follows. All images (e.g. at 100, 160, 250, 350 and 500µm) are first processed to have the same pixel size and the same reference pixel (e.g. using SWARP). Then each image (at each wavelength) is decomposed in single scale images over a wide range of spatial scales (of incremental spatial frequency) starting from FWHM/3 of that wavelength. This decomposition is obtained convolving the original images with circular Gaussians and subtracting them from one another, starting from the highest resolution (see Fig. 12 for a schematic illustration):

$$I_j(\lambda)=G_{(j-1)}*I(\lambda)-G_j*I(\lambda)(j=1,\ldots,N)$$

where:

- I(l) is the original image
- I_j(l) are the "single-scale" decomposition images
- G_j are the smoothing Gaussians with a $FWHM_j$ = *fs* x $FWHM_{j-1}$, where *fs* = 1.05, $FWHM_0$ = 2 pixels (e.g. 2x1.2") and $FWHM_{max}$ =2 x largest FWHM beam size (e.g. 2x36" for Herschel).

Each one of the obtained "single scale" images are then cleaned of noise and background by iterating searches for the appropriate cut-off levels.
Combined single scale images are then produced by summing up the single scale images at each individual wavelength that have the same resolution.

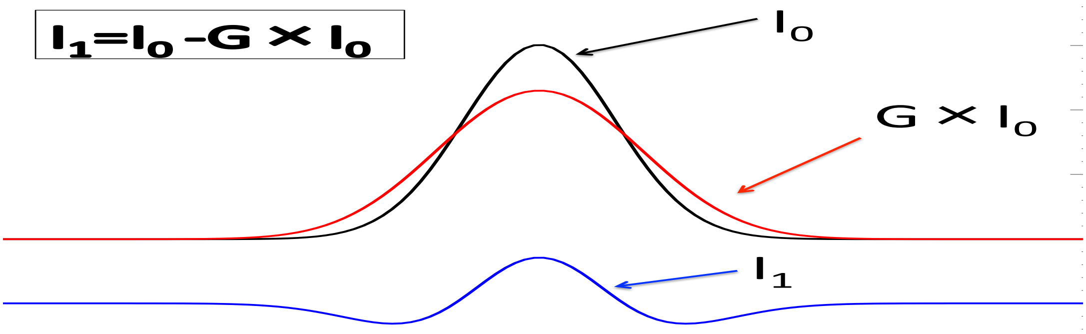

Figure 12: A schematic representation of the spatial decomposition in GETSOURCES.

Since shorter wavelengths will have different S/N ratios than larger wavelengths due to the different beam sizes and flux density of the source as a result of its SED and redshift, the single scale images are weighted as a function of resolution, in order to give the strongest weight to the wavelength which FWHM is closest to the scale resolution (see Eq.8 of Men'shchikov et al. 2012). Another normalization is applied to avoid too strong gradients in peak intensities from one scale to another in the combination of wavelengths (see Eq.6 of Men'shchikov et al. 2012). This "renormalization" process ensures that signals only detected in high resolution images are not diluted when combined with much smaller resolution images. Yet, when two wavelengths are relatively close (such as e.g. the 100 and 160µm ones), then the resolutions are close enough to allow the detection of sources to be enhanced by the combination of images; e.g. 2.5s sources in both images may become 3s detections, which is a clear improvement as compared to blind detection techniques that are applied to single wavelengths (e.g. STARFINDER).

Objects are then identified in the combined detection image at each scale by finding segmentation maps of the objects, and by tracking their appearance and "evolution" from small to large scales. In the combined clean images, the sources begin to appear at some relatively small scales; their brightest peak is systematically found at the scale that roughly corresponds to the source size (FWHM); and they eventually vanish at significantly larger scales.

After individual objects are detected in the combined single scale images as described above, their properties (sizes and flux densities at each wavelength) are measured on the original background-subtracted image, where they are deblended with an algorithm that iteratively divides the intensity of a pixel between surrounding objects according to the two-dimensional Moffat function.

The end result of the process is a final catalog containing coordinates of all detections and estimates of the objects' S/N ratios, peaks, total fluxes (with uncertainties), and sizes and orientations at each wavelength.

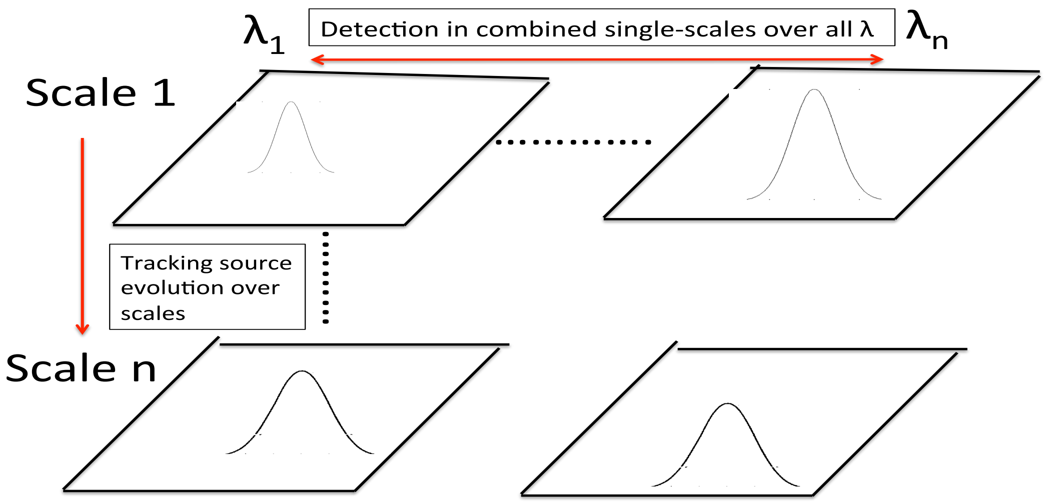

Figure 12: schematic illustration of the source detection in GETSOURCES.

Typical computational times are of the order of 2-3 hours from decomposition to measurement. The algorithm may require considerable storage space, depending on the numbers of pixels, spatial scales, wavelengths, iterations, and potential sources detected.

4.2 "Prior" techniques

4.2.1 Standard GOODS prior PSF-fitting technique

The basic conceptual approach of this method is similar to the one described in Sect. 3. Given a prior information about the expected positions of sources, this method aims to solve the linear system $B=AX$, where:

- B is a N_{pixel} vector containing the flux density in each pixel of the scientific image;
- A is a N_{pixel} x $(N_{prior}+1)$ matrix;
- X is a $N_{prior}+1$ vector, containing the $N_{prior}+1$ free parameters of the system, i.e., flux density of each prior and background level.

The N_{prior} first columns of A contain the expected flux density of the i-th prior in each pixel of the scientific image, given its position and PSF. The last column of A contains the flux density of an unknown but constant background in each pixel of the scientific image.

Because $N_{pixel} > N_{prior}+1$, this system is overdetermined and can be solved using a single value decomposition (SVD) method, implemented in IDL (namely SVDC and SVSOL modules). This method is equivalent to a χ^2 minimization method of the $||Ax-B||^2$ system, but assuming a constant noise level across the scientific image. This assumption of a constant noise level is suitable for mid- and far-infrared observations, because their coverage maps are very uniform.
The measure image is divided into cells and a system is constructed and solved for each cell.

For each prior, the method proceeds as follows:

- a cutout stamp of (10xFHWM) x (10xFWHM) centered at the prior position is extracted from the scientific image; this particular size is defined to be large enough to contain all priors influencing the flux density of the central prior;
- the corresponding $B=AX$ system is solved using the SVD method;
- because the method is not bounded to positive solutions, some priors might result having unphysical negative flux densities; therefore, they are removed from the system of equations, and a new system is built and solved again; the procedure is iterated until all remaining priors have positive flux densities.

Flux densities of all priors within 3xFWHM from the center of the stamp are saved. Outside this area flux densities are not saved because they might be significantly affected by priors not situated in the current stamp. To evaluate the quality of our fits, we convolve the residual map with the PSF. The quality of the fit is then given by the dispersion of pixel values in the convolved residual map. This estimate is recorded along with the prior flux density estimates.

Following this procedure, each prior can have several flux density estimates, i.e. being the central prior of a stamp and being a close neighbour (<3xFWHM) in another prior stamp. In such cases, the flux density of the prior is defined as being the one with the highest quality of fit.

4.2.2 FASTPHOT

FASTPHOT (Béthermin et al., 2010) is a quick and simple PSF fitting routine optimized for prior extraction. It fits all the detected sources at the same time and is consequently efficient for deblending.

Assuming that the noise is Gaussian and the position of sources is known, FASTPHOT determines the flux of each source by maximizing the likelihood:

$$L(m|s)=\prod_{pixels} C\times\exp\left[-\frac{\left(m-\sum_{i-1}^{Nsources} PSF_{x_i,y_i}\times s_i-\mu\right)^2}{2\sigma^2}\right],$$

where m and s are the map and the noise map, $PSF_{xi,yi}$ is a unit-flux PSF centered at the position (x_i, y_i), m is a constant background, C is a normalization constant which depends only of the value of the noise map, and s is a vector containing the flux of the sources.

The linear equation

$$\frac{\partial \log(L(m|s))}{\partial s_i}=0$$

is equivalent to the standard *($N_{sources}$ + 1)x($N_{sources}$ + 1)* matrix equation *AX=B,*

described in Sect. 3.

FASTPHOT allows fitting all the sources in the field simultaneously, but also allows to first cut the field into several cells and then fit each area separately. It is fast (it can perform simultaneous PSF fitting photometry of 1000 sources in SPIRE maps in less than 1 second); it takes into account the noise map which can aid in obtaining more accurate flux densities, and calculates the covariance matrix, which can be used to measure the quality of fitting.
For a more detailed information, please consult the webpage http://www.ias.u-psud.fr/irgalaxies .

4.2.3 DESPHOT: DEblended Spire PHOTometry

The issues with deblending sources in heavily confused far-IR/sub-mm images such as those from BLAST, Herschel and SCUBA-2 are distinct from those encountered at shorter wavelengths for a number of reasons:

- steep source counts lead to extreme crowding in low-resolution images which are *δ*-confusion-limited;
- as a result the background is dominated by unresolved faint sources rather than empty sky;
- the flux of a given source at these wavelengths is generally uncorrelated with that at most other wavelengths;
- sources are unresolved at all but the most local redshifts.

However, the last point, although contributing to increase confusion, actually becomes an advantage when it comes to resolve blends, because it allows to mathematically decompose a severely confused image into a linear combination of point sources. Given their positions as a prior, the fluxes of those sources can be computed since their profiles can be assumed to be identical and defined by the Point Response Function (PRF). In algebraic form, the image is a vector of pixels given by

$$d=\sum_{i=1}^{n} P_i f_i+\delta$$

where P_i is the PRF at the position of source *i*, f_i the total flux of source *i*, and *δ* the noise. This approach has been used to measure stacked properties of undetected populations (e.g. Marsden et al., 2009; Kurczynski & Gawiser, 2010; Bourne et al., 2012; Viero et al., 2013a), but these methods are limited by the assumption that a complete catalogue of source positions is provided as an input. The DESPHOT method, developed by Roseboom et al. (2010, 2012, 2013) (see also Viero et al., 2013b; Wang et al., 2013), is more flexible and implements a lossminimisation procedure to iterate towards the optimal solution. A prior catalogue is required, based on positions from (e.g.) MIPS and radio imaging, but this catalogue is not static. The algorithm only includes sources from the prior which are necessary to optimise the solution, but can add sources that are absent from the prior if they appear in the residual. Optimisation is achieved by solving the matrix form of the system equation, i.e.

$$d = Af + \delta,$$

in which A contains the contributions from each source to each pixel in the map. The solution is found using the non-negative, weighted 'least absolute shrinkage and selection operator' (LASSO; Tibshirani, 1996; ter Braak et al., 2010), an algorithm which finds the smallest 'active set' of sources which are necessary for a non-negative (i.e. no negative sources) least-squares fit, discarding unnecessary ('inactive') sources altogether. The algorithm begins with all sources flagged as inactive, and in each iteration chooses to either activate a new source or increment the flux of an existing one, in order to achieve the maximum possible reduction in χ^2. The iteration process is continued until the gradient of the χ^2 reaches zero, indicating that the minimum has been found.

The main steps of the photometry algorithm can be summarized as follows:

- the image is divided into cells defined by islands of blended sources surrounded by a ring of low signal-to-noise pixels, and each segment is processed in the following steps individually (this step is optional and is included purely to reduce computing resources);
- a solution is obtained for the positions in the prior catalogue using LASSO;
- the background level of the map is estimated from the residual of the data image and the solution model;
- additional sources are identified in the residual (using STARFINDER, see Sect. 4.1) and are appended to the prior catalogue. In the case of SPIRE, since the algorithm is used on all three bands simultaneously the sources identified in any one band can be included in the updated prior for all bands;
- a second pass through LASSO is performed to obtain the optimal solution with the measured background and additional sources;
- finally, errors are estimated by inversion of the Fisher information matrix,
  $$\left(A^T N_d^{-1} A\right)^{-1}$$

(N_d being the covariance matrix), which contains the instrumental noise plus confusion noise between active sources. The additional confusion noise from background sources that are not extracted is estimated from the variance in the residual image after the final solution model is subtracted.

DESPHOT has previously been applied for Herschel-SPIRE photometry using priors from 24-μm and near-IR bands (Roseboom et al., 2010, 2013; Viero et al., 2013b; Wang et al., 2013). In ASTRODEEP the algorithm can be tested and developed to provide the optimal multi-wavelength deblending of the far-IR/submm images including possible weighting of the lower-resolution blends using information from the solutions in higher-resolution images.

5 Photometry in X-ray deep Chandra images

The detection of significant features in 2-D X-ray images poses some special challenges that mandate specially developed methods for source detection. Chandra data are usually reduced with standard software (CIAO) provided by the Chandra X-ray Center. The CIAO Detect package includes three different source detection programs which statistically identify significant brightness enhancements, deriving from both unresolved (point) and resolved (extended) X-ray sources. The detection of faint sources in deep and ultra deep Chandra fields (see Xue et al. 2012) is performed with "blind" search techniques on the images, which are described in the following sections.

5.1 Blind "one step" techniques: WAVDETECT

Wavelet source detection is performed using the CIAO software WAVDETECT (Xue et al. 2012). WAVDETECT correlates the image with wavelets of different scales (selected by the user) and then searches the results for significant correlations. It is a wrapper for the tools WTRANSFORM and WRECON, and works in two steps:

- Step 1: WTRANSFORM detects probable source pixels within a dataset by repeatedly correlating it with "Mexican Hat" wavelet functions with different scale sizes;
- Step 2: WRECON generates a source list with information from each wavelet scale. For each source, a cell that contains the majority of the source flux is constructed, and source properties are computed within it. The WRECON step can incorporate PSF information through a PSF map image, congruent with the input image, whose pixel values represent the size of the PSF at that image location.

5.2 Blind "two step" techniques with priors: CLMDETECT

For an input list of source positions in Chandra images, simultaneous maximum likelihood PSF fits to the source count distribution are performed in all energy bands of each ACIS-CCD (a description of the main properties of the detection algorithm can be found in Cruddace et al. 1988). This is achieved using the code CLMDETECT.

The most important fit parameters are the sources location, sources extent (Gaussian sigma or beta model core radius), and sources count rates. The PSF fitting may either be performed in single-source or in multi-source mode, in the latter neighbouring sources with overlapping PSFs being fitted simultaneously. Detection likelihoods are optimized for all the overlapping sources simultaneously, and detection likelihoods per source are calculated.
Selection of sources for simultaneous fitting is controlled by a distance parameter and by a parameter *nmaxfit* that gives the maximum number of sources to be fit simultaneously ($1 \leq$ *nmaxfit* ≤ 10).
It is also possible to fit several PSFs for each input source position, by setting the parameter *nmulsou* to the corresponding value ($1 \leq$ *nmulsou* ≤ 3, *nmaxfit*$*$*nmulsou* ≤ 10).
Two parameters determine the image region on which a source fit is performed: (that is, both the size of the subimage around each source used for fitting, and the radius around each source in which other input sources are considered for multi-PSF fitting, if *nmulsou* > 1). Both these parameters are given as encircled energy fractions of the calibration PSF, so the actual radii in pixel units slightly change with energy band and source position; alternatively, they can be given as a fixed value in units of image pixels.
All detection likelihoods are transformed to equivalent likelihoods L_2, corresponding to the

case of two free parameters, to allow comparison between detection runs with different numbers of free parameters (i.e., when different numbers of input images are used):
where P is the incomplete Gamma function, n is the number of energy bands involved, ν is the number of degrees of freedom of the fit, and $L_i = C_i/2$ with C as defined by Cash (1979). The

$$L_2 = -\ln(1 - P(\frac{\nu}{2}, L')) \quad \text{with} \quad L' = \sum_{i=1}^{n} L_i$$

equivalent detection likelihoods obey the simple relationship $L_2 = -\ln(p)$, where p is the probability for a random Poissonian fluctuation to have caused the observed source counts. Note that for very small numbers of source counts (less than ≈ 9 counts, Cash 1979), this relation has to be treated with caution. Therefore, it will only give a rough estimate of the number of expected spurious sources.
Errors in right ascension and declination are then calculated as error ellipses at $C = C_{min} + 1$ and extended to 68% confidence levels (chi-square) by multiplying them by a factor of $\sqrt{2}$. The final value for the positional error is calculated as the square root of the quadratic sum of the errors in R.A. and Dec.

CMLDETECT has been initially developed for ROSAT and XMM-Newton (with the original name EMLDETECT), and was later adapted for use with Chandra with the customized version which makes use of a Chandra PSF-Library. However, unlike the XMM-Newton PSF, the Chandra-PSF does not depend exclusively on the off-axis angle but also on the azimuthal position. In order the adapt the software for Chandra, an ad-hoc PSF library has been created by averaging over all the azimuthal angles the PSF templates in energy and off-axis angle bins. This approximation has been tested with several Montecarlo simulation and real data in the Chandra COSMOS survey (Puccetti et al. 2009) and by comparing them with other software outputs. This method ensures the best control on the thresholds and the best result on fluxing precision.

The software needs as input a background map, which is previously computed with a spline fit of the source removed image; however, this method left spurious features in the background maps, and it was therefore replaced with the use of the background templates provided by the Chandra X-ray data Center (CXC). Basically, these are long exposure images taken with the ACIS detector in stow mode, so that the only recorded events have instrumental origin ora are particle background events. The remaining part of the background (purely cosmic diffuse background), which unlike the particle background suffers from the off-axis dependent vignetting, has been modeled according to the exposure map. A full description of the method can be found in Cappelluti et al. (2013).

The input source list can be produced by either a blind one step detection or by simply inputting any catalog of known sources. For most of the X-ray surveys using a two-step blind source detection, the list of source candidates (priors) is computed with EBOXDETECT, a software tool developed for XMM-Newton.
In EBOXDETECT, source counts are accumulated from a 3x3 or 5x5 pixels window with the option to use a position and energy dependent PSF weighted filter. Detection of extended objects is again achieved by doubling the pixel sizes in up to three consecutive detection runs. Background-subtracted source counts are obtained by applying correction factors to account for the fractions of source counts falling in each source and background area, respectively. The respective off-axis angle dependent correction factors are calculated using the current

calibration PSF (medium accuracy model).
The PSF correction of source and background counts as implemented in the code proceeds as follows. The enboxed energy fraction in the source box is

$$\alpha = \sum_{n\times n} PSF\ ,$$

while the fraction of source counts in the background counting area is

$$\beta = \sum_{(n+4)\times(n+4)} PSF - \sum_{n\times n} PSF\ .$$

The raw box counts are given by

$$BOX_{CTS} = \sum_{n\times n} image\ ,$$

while raw background map is

$$BG_{RAW} = \left(\sum_{(n+4)\times(n+4)} image - \sum_{n\times n} image\right) / \left((n+4)^2 - n^2\right)\ .$$

The PSF corrected and background subtracted source counts are given by:

$$STCS = \frac{BOX_{CTS} - BG_{RAW} \times n^2}{\alpha - \beta \times n^2 / \left((n+4)^2 - n^2\right)}$$

and the PSF corrected background map is

$$BG_{MAP} = BG_{RAW} - SCTS \times \frac{\beta}{(n+4)^2 - n^2}\ .$$

Finally, the source counts in map detection mode are given by

$$SCTS = \frac{BOX_{CTS} - BG_{MAP} \times n^2}{\alpha}\ .$$

Count rate errors are calculated by assuming Poissonian statistics in both the source and background cells. Positional errors are assumed to be equal to the standard deviation of the distribution of the counts in the detection cell. The errors of the derived parameters, such as count rates, fluxes, and source positions in celestial coordinates are derived from the count and image pixel positional errors, respectively.

Following the definition which was e.g. used by the ROSAT mission, detection likelihoods (per energy band and total) are given for each source in the form $L=-ln(p)$ where p is the probability of Poissonian random fluctuation of the counts in the detection cell which would have resulted in at least the observed number of source counts. The value of p is calculated using the incomplete Gamma function $P(a,x)$ as a function of raw source counts and raw background counts in the detection box (see Press et al., Numerical Recipes, chapter 6.2 for the details on the calculation of *P)*, thus:

$$LIKE = -\ln p\left(BOX_{CTS}, BG_{RAW} \times n^2\right)$$

The source list produced with a low threshold in *LIKE* is the passed to CMLDETECT.

6 References

Agüeros M.A. et al., 2005, AJ, 130, 1022
Barden M. et al., 2012, MNRAS, 422, 449
Beard S.M. et al., 1990, MNRAS 247, 311
Bertin E., & Arnouts S., 1996, A&AS, 117, 393
Bertin E., 2002, ASPC, 281, 228B
Bertin E., 2011, ASPC 442, 435
Béthermin M., Dole H., Cousin M., & Bavouzet N., 2010, A&A, 516A,b43B
Bourne N. et al., 2012, MNRAS, 421, 3027
Bouwens R.J. et al., 2011, ApJ, 737, 90B
Cappelluti N. et al., 2013, ApJ, 769, 68C
Cash W. et al., 1979, ApJ, 228, 939C
Castellano M. et al., 2010, A&A, 524, 28C
Cruddace R.G, Hasinger G.R. & Schmitt J.H., 1988, ESOC, 28, 177C
De Santis C., Grazian A., Fontana A. & Santini P., 2007, NewA, 12, 271
De Vaucouleurs G., 1948, Ann. Astrophys., 11, 247
Diolaiti E., Bendinelli O., Bonaccini D. et al., 2000, A&AS, 147, 335
Diolaiti E. et al., 2000, SPIE, Conference Series, Vol. 4007, 879–888
Driver S.P. et al., 2011, 413, 971
Freeman K.C., 1970, ApJ, 160, 811
Galametz A., et al., 2013, ApJs, 206, 10G
Häussler B. et al., 2013, MNRAS, 430, 330
Kelvin L.S., 2012, MNRAS, 421, 1007
Kendall M. & Stuart A., 1977, “The advanced theory of staistics. Vol. 1”
Kron G.E., 1980, R.A., 10, 3K
Kurczynski P. & Gawiser E., 2010, AJ, 139, 1592
La Barbera F. et al. 2008, PASP, 120, 681
Laidler V.G. et al., 2007, PASP, 119, 1325
Mancone C.L. et al., 2013, arXiv:1310.6046
Marsden G. et al., 2009, ApJ, 707, 1729
McLure R.J. et al., 2011, MNRAS, 418, 2074M
McLure R.J. et al., 2013, MNRAS, 432, 2696M
Men’shchikov A. et al., 2012, A&A 542, 81
Obrić M. et al. 2006, MNRAS, 370, 1677
Peng Y. et al., 2002, AJ, 124, 266
Peng Y. et al., 2010, AJ, 139, 2097
Petrosian V., 1976, ApJ, 210L,53P
Pignatelli E., Fasano G. & Cassata P., 2006, A&A, 446, 373
Press W.H. et al., 1992, “Numerical Recipes in FORTRAN”
Puccetti S. et al., 2009, ApJS, 185, 586P
Roseboom I. G. et al., 2010, MNRAS, 409, 48

Roseboom I. G. et al., 2012, MNRAS, 419, 2758
Roseboom I. G. et al., 2013, MNRAS, 436, 430
ter Braak C.J.F., Boer M.P., Radu Totir L., Winkler C., Smith O. S., Bink M.C.A.M., 2010, Genetics, 185, 1045
Sersic J.L., 1968, Atlas de Galaxias Australes, Observatorio Astronómico, Córdoba, Argentina
Simard L., 1998, Astronomical Data Analysis Software & Systems, Astron. Soc. Pac., p. 108
Simard et al., 2002, ApJS, 196, 11
Stetson P.B., 1987, PASP, 99, 191
Tibshirani R., 1996, J. R. Statist. Soc. B, 58, 267
Viero M. P. et al., 2013a, ApJ, 779, 32
Viero M. P. et al., 2013b, ApJS submitted (ArXiv:1308.4399)
Vika M. et al., 2013, MNRAS, 435, 623
Wang L. et al., 2013, MNRAS submitted (ArXiv:1312.0552)
Wilson G. et al. 2007, ApJ, 660, L59
Xue Y.Q. et al., 2012,ApJ, 758, 129