

ProFound: Source Extraction and Application to Modern Survey Data

A. S. G. Robotham,^{1*} L. J. M. Davies,¹ S. P. Driver,¹ S. Koushan,¹
D. S. Taranu,^{1,2} S. Casura,³ J. Liske³

¹ICRAR, M468, University of Western Australia, Crawley, WA 6009, Australia

²ARC Centre of Excellence for All-sky Astrophysics (CAASTRO)

³Hamburger Sternwarte, Universität Hamburg, Gojenbergsweg 112, 21029 Hamburg, Germany

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

We introduce PROFOUND, a source finding and image analysis package. PROFOUND provides methods to detect sources in noisy images, generate segmentation maps identifying the pixels belonging to each source, and measure statistics like flux, size and ellipticity. These inputs are key requirements of PROFIT, our recently released galaxy profiling package, where the design aim is that these two software packages will be used in unison to semi-automatically profile large samples of galaxies. The key novel feature introduced in PROFOUND is that all photometry is executed on dilated segmentation maps that fully contain the identifiable flux, rather than using more traditional circular or ellipse based photometry. Also, to be less sensitive to pathological segmentation issues, the de-blending is made across saddle points in flux. We apply PROFOUND in a number of simulated and real world cases, and demonstrate that it behaves reasonably given its stated design goals. In particular, it offers good initial parameter estimation for PROFIT, and also segmentation maps that follow the sometimes complex geometry of resolved sources, whilst capturing nearly all of the flux. A number of bulge-disc decomposition projects are already making use of the PROFOUND and PROFIT pipeline, and adoption is being encouraged by publicly releasing the software for the open source R data analysis platform under an LGPL-3 license on GitHub (github.com/asgr/ProFound).

Key words: methods: data analysis – techniques: image processing – techniques: photometric

1 INTRODUCTION

Consistent and reliable source detection and photometric extraction has been a rich vein of research in astronomy. Clearly it is preferable to have a quantitative and reproducible means to analyse images, and over the years a number of fully automatic tools have been developed to achieve such outcomes (e.g. Bertin & Arnouts 1996).

Our group recently developed the PROFIT 2D galaxy profiling tool (Robotham et al. 2017), which requires a number of reasonable inputs that require tools outside of the package. Critically important for achieving a good fit are: a pixel matched sigma map (reflecting the local uncertainty in the image provided); a segmentation map that flags the

pixels to use when computing the fit likelihoods; a careful sky subtraction; and reasonable initial guesses for the profile parameters.

A mixture of tools written in a number of languages cover most of the input requirements for PROFIT, however in practice how these tools are combined when scripting PROFIT for a large automatic analysis of galaxy profiles has a critical impact on how successful the fitting procedure is. With a particular focus on sky subtraction, object segmentation and initial parameter estimates (the three most difficult aspects of galaxy profiling outside of the optimization problem itself) we developed the PROFOUND photometry package using the R data language (R Development Core Team 2016). Ostensibly this package is used to create good quality automatic inputs for further 2D decompositions with PROFIT, however it also serves as an extensively featured

* E-mail: aaron.robtham@uwa.edu.au

source detection and photometric extraction package in its own right. PROFOUND is designed to work well with relatively deep large-area images where at least a significant minority (25+%) of the pixels belong to the sky, i.e. of the type that you might use for galaxy profiling. To combat image artefacts it supports the use of per pixel masks, but in general it works best of smoothly varying well calibrated images, i.e. images without serious pedestal mosaicking discontinuities.

Blind source finding, as it is often known, has a long history in astronomy (see the recent detailed review in [Masias et al. 2012, 2013](#)). In the earliest days it was a necessarily visual and heuristic process, where astronomers would identify sources in photographic images essentially by eye. As technology moved towards the era of digital detectors and large arrays of imaging pixels, computer techniques advanced to automate these results in a more deterministic manner. Early techniques included simple sigma thresholding of the data in reference to the root mean square (RMS) fluctuations measured in the sky. This approach works well when the sources of interest are well above the sky noise. When sources move closer the surface brightness limit of the data, this technique can become increasingly ineffective, and lead to a higher than ideal false-positive rate. That is the number of new real sources can become subdominant compared to statistical fluctuations in the sky ([Davies et al. 2005](#)).

To combat this effect many improvements have been identified in the literature, e.g. simple schemes that smooth the data with an appropriate kernel and require a certain number of pixels to be above the RMS threshold and within a certain spatial separation on sky (see [Sabatini et al. 2003](#), for a discussion on such techniques for uncovering marginally detected sources). In practice a matched filter is often the optimal smoothing kernel, where for convolution this is the transpose of the point spread function of the image (for the one dimensional matched filtering argument see [Van Vleck & Middleton 1946](#)). Whilst these approaches are often applied in an ad-hoc manner (e.g. the exact matched filter is often not chosen as the convolution kernel), they work well to qualitatively reduce the false-positive rate by essentially requiring a spatial correlation in image fluctuations, which reduces the chance sky noise fluctuations far below that implied by the threshold applied.

Another area that is heuristic in nature but has been seen to work quite well in practice is source de-blending. This is a complex problem that is only satisfactorily resolved using a full generative model, e.g. the 2D galaxy profiling code PROFIT offers a mechanism for doing such an extraction. However, in many applications this approach is prohibitively computationally expensive. A pragmatic option has been to process the image pixels with a source de-blending algorithm. These usually work on a variant of the so-called ‘watershed’ de-blending. How these operate can differ in detail, but a generic feature is they separate the image into regions of distinct flux by approximating the image flux as belonging to different topographic structures, i.e. if the image was inverted these would approximately be seen as valleys (the positive flux sources) and flat noisy regions (the sky and the sky noise). If this topographic structure was steadily filled with water it is easy to see that structures that begin as distinct bodies of water will start to merge together as the

image becomes entirely flooded. There are various methods to use this insight to define genuinely distinct sources, but the basic approach is the same.

Mixed in with the above issues, there are a large number of subtle effects that must be handled carefully. These include the sky estimation, the sky RMS estimation, and the growth of apertures to fully contain the flux. Each of these have long histories in astronomy literature, but they largely all share a heuristic approach. This is usually for pragmatic reasons of computational complexity rather than aiming to be the ideal solution in a demonstrative sense. The calculation of the sky and sky RMS are simpler problems to tackle for the most part (exceptions include very crowded fields and confusion limited data), and a large part of the Methods (Section 2) discusses the main approaches in detail. Choosing an appropriate method to fully capture the flux present is a more difficult problem to solve, and many approaches have been advocated in the literature.

The earliest attempt at systematically identifying the flux for extended sources can be found in [Petrosian \(1976\)](#). The basic idea behind the Petrosian magnitude is to determine the radius to be scaled based on surface brightness properties of the galaxy, namely the ratio of the integrated surface brightness within some radius compared to the instantaneous surface brightness at the same radius. By incorporating the surface brightness in the numerator and the denominator when calculating the Petrosian radius the results of many observational effects are naturally removed, e.g. cosmological surface brightness dimming, variable imaging depth of the data under consideration and different observing conditions which can produce variable seeing amongst other effects.

In theory the Petrosian magnitude is an elegant route to extract flux measurements, since in principle extracted extended source fluxes are not highly sensitive to the observing conditions. In practice things are not as simple as we would like, and galaxies are not well represented as having a shared fundamental profile, which whilst not immediately clear is implicitly assumed in the Petrosian magnitude system ([Graham & Driver 2005](#)). Since galaxies can have a very broad range of profiles the magnitude extracted is in fact highly sensitive to the profile Sérsic index ([Sérsic 1963](#); [Graham & Driver 2005](#)). Since galaxies are also convolved with the atmospheric seeing in ground based data there is an additional dependence beyond the intrinsic profile, namely that the same galaxy shifted to higher redshift will return a different fraction of the true flux because the profile will have evolved away from its intrinsic value towards the atmospheric value, which for a mixture of reasons is usually very close to a canonical Sérsic index of 0.5, i.e. a Normal distribution.

A popular and computationally simpler alternative was presented in [Kron \(1980\)](#). In this approach the inner bright moments of light are used to estimate the Kron radius. It is then up to the user to identify a sensible Kron multiplier to scale this radius by in order to capture a certain quantity of the flux. For stars this factor is not especially important since the majority of the object flux is captured within the inner part of the profile. For more extended objects the number chosen for this multiplier can have a significant impact since much of the flux of extended objects comes from the outer parts of the profile. The ‘AUTO’ magnitude returned

by `SEXTRACTOR` (Bertin & Arnouts 1996) is closely related to this version of the Kron magnitude, and can be run in such a mode that this multiplying factor is chosen intelligently for each source rather than using a single fixed value.

A key feature of both of the above methods for extracting photometry is that they use either circular or elliptical apertures. The earliest applications predominantly used circular apertures, with the elliptical variants being more modern and widely popular today. This suggests an obvious limitation of either approach: galaxies are not simple ellipses, and decisions have to be made regarding how flux is distributed between potentially overlapping apertures.

The above motivated the most novel aspect of the `PROFOUND` code presented in this paper: a move away from simple elliptical apertures and towards apertures that properly identify the parts of galaxies containing the significant proportion of the flux. A few methods were looked at when first addressing this issue, with the result that dilated segments that follow the surface brightness distribution of the galaxies act as a better method to identify the true flux belonging to a given galaxy.

In the regime of bright compact elliptical sources, which are fairly isolated from other sources, the dilated segments follow the extent of traditional apertures fairly closely. However, for very extended sources the differences can be quite pronounced, with complex source geometry not being accurately captured by simple circular or even elliptical apertures. The dilation approach also offers a few other advantages when it comes to source de-blending, namely that segments are never allowed to overlap on the sky in the way the expanded apertures can. In fact it is non-trivial to determine how best to split flux between adjacent and overlapping apertures, and a number of different ad-hoc and heuristic schemes are usually applied to account for these effects.

The hard flux boundaries created by using segmented apertures also have a number of advantageous side-effects when it comes to determining fluxes in regions that have pathological issues such as very bright halos around bright stars that have saturated the detector. These often create biased photometry over an extended region, where the Kron or Petrosian aperture is often compromised and the expanded aperture overlaps with multiple fainter sources. We highlight some specific examples of this later in the paper, but it is a common feature of survey data (Wright et al. 2016). The method of iterative dilation used in `PROFOUND` naturally prevents extreme expansion artefacts since segments are not allowed to grow into each other. This is not to say the photometry extracted will not be compromised at all, but the segmentation map and the approximate sources properties are much nearer to the intrinsic values and serve as better inputs for `PROFIT`, which was the initial design goal of the new software.

In Section 2 we describe the methodology behind the most critical aspects of the package design. In Section 4 we look at the application of `PROFOUND` to fully simulated wide-field images, with a focus on the completeness and purity of the detection, and the accuracy of the photometric properties. In Section 5 we apply `PROFOUND` to the UltraVISTA multi-band imaging data, with a detailed comparison of some of the output properties compared to the public catalogues.

2 METHODS

In the following description of the main methods behind `PROFOUND` source extraction we use the same test Z-band data shown in Figure 1. This was taken from the public VISTA (Visible and Infrared Survey Telescope for Astronomy) Kilo-degree INfrared Galaxy (VIKING; Edge et al. 2013) survey that used the VIRCAM instrument on ESO’s 4m VISTA facility. The galaxy at the centre of the image was a main survey target (G5458748) for the Galaxy And Mass Assembly survey (GAMA; Driver et al. 2011; Liske et al. 2015). This image has a number of properties that make it ideal as a small case study: it contains a mixture of bright and faint galaxies and stars; it contains a mixture of compact and extended galaxies, the central region contains a number of reasonably confused sources; the background root mean square (RMS) in the sky varies distinctly in the frame due to its stacked origin; and it has objects contained entirely within the image and near to the edge.

The example data is included with the `PROFOUND` package so it is easy for a new user to recreate the plots in this paper using the many worked examples and vignettes¹. The thorough package documentation (the embedded PDF manual is 61 pages, with every function, variable and output described) and long-form vignettes have been influenced by the clear utility of the ‘SEXTRACTOR for Dummies’ guide which has been hugely beneficial to the community who regularly use `SEXTRACTOR` (Holwerda 2005).

This paper is not intended as a user manual, so we will not discuss the technicalities of the detailed settings and parameters here. Except where mentioned explicitly the code has been run in close to default mode, with the notable difference being the setting of the magnitude zero point (which has to be set explicitly since there is no standard format to specify this in FITS headers). Otherwise meta-data is largely extracted from the FITS header and extraction properties are estimated dynamically using the data itself.

Functions in the `PROFOUND` package are all named with a leading lowercase ‘profound’. This is to remove the potential for clashing function names since R does not trivially support package aliasing in the way that some high level languages do (e.g. PYTHON). The `PROFOUND` package includes a few different hierarchies of functions. The expectation is that some of these will be used routinely (e.g. the highest level `PROFOUND` object extraction and photometric measurement function, also called `PROFOUND`), and some will rarely be used by a typical end user (e.g. the linear interpolation function `INTERP2D`).

Between these two extremes there are a large number of mid-level functions that more advanced users might want to use directly in order to manipulate the data in a specific manner. The highest level `PROFOUND` function effectively links a large number of these mid-level functions together in a manner that achieves good quality source extraction and photometric analysis for a range of typical two dimensional astronomy data (particularly imaging and radio continuum data, but not limited to such applications). During development the focus has been on optical and NIR survey data, but it has also been used successfully on ultra-violet (UV) data and far-infrared (FIR) data.

¹ <http://rpubs.com/asgr/>

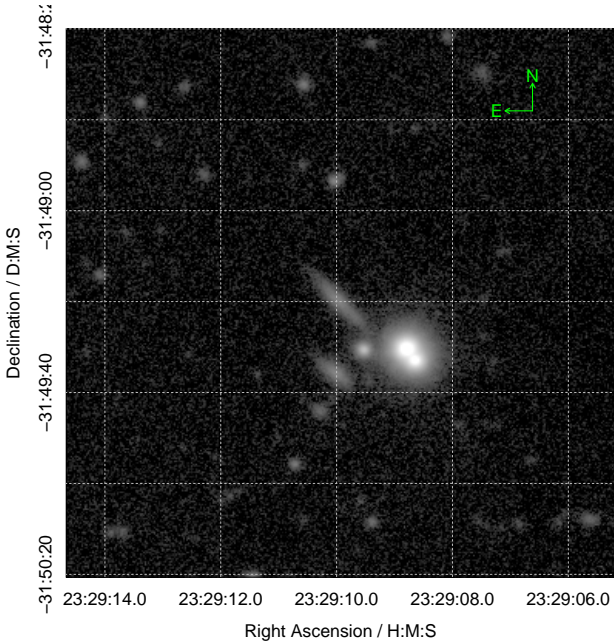


Figure 1. Example VISTA Z-band data taken from the VIKING survey included with the PROFOUND package and used in various parts of this paper (GAMA galaxy ID G5458748). In this Figure we stretch the z-scale to make the double star nature of the two bright central sources clear. In latter Figures we use a different mapping that enhances the contrast of fainter sources and visually merges these two stars together.

2.1 PROFOUND Source Extraction

The highest level PROFOUND function is ultimately a structured calling of a range of mid-level functions. The full code diagram is presented in Figure 2 with the various parts we discuss in more detail labelled by Section. The simplified version of Figure 2 is provided below, where the mid-level function executed is named at each relevant stage.

- (i) Make a rough sky map (see Section 2.2) — MAKESKYGRID
- (ii) Using this rough sky map, make an initial segmentation map (see Section 2.3) — MAKESEGIM
- (iii) Using this segmentation map, make a better sky map — MAKESKYGRID
- (iv) Using this better sky map, extract basic photometric properties (see Section 2.5) — SEGIMSTATS
- (v) Using the current segmentation map, dilate segments and re-measure photometric properties for the new image segments, by default it iterates six times (see Section 2.4) — MAKESEGIMDILATE / SEGIMSTATS
- (vi) Using the iterative dilation statistics, every object is checked for convergence, by default convergence of flux is used (see Section 2.4) — SELECTCOG
- (vii) Make final segmentation map by combining the segments when each source has converged in flux
- (viii) Make a conservative object mask by aggressively dilating the final segmentation map — MAKESEGIMDILATE
- (ix) Using this conservative objects mask make a final sky map — MAKESKYGRID
- (x) Using the final segmentation map and the final sky

map compute the final comprehensive photometric properties — SEGIMSTATS

(xi) Return a list containing the input image pixel-matched final segmentation map (called **segim**), the pre-dilation segmentation map (**segim_orig**), the binary object/sky mask (**objects**), the conservatively dilated binary object/sky mask (**objects_redo**), the sky image (**sky**), the sky-RMS image (**skyRMS**) and the effective surface brightness limit image (**SBlim**)

(xii) Return the data-frame of photometric properties for every detected source (**segstats**)

Some other simple properties are passed through and included in the output of ProFound: the original image, the image header (**header**) if it is attached to the input image, the magnitude zero point specified (**magzero**), the gain in electrons per astronomical data unit (**gain**), the pixel scale in arc-seconds per pixel (**pixscale**) and finally the full function call (**call**).

The above detection and extraction sequence was developed on a range of test data, where the desire was that good quality results should be achievable when running on default parameters. The latter was deemed important since parameter tuning can be non-obvious and complicated for novice users. Even in examples where qualitatively better extraction could be achieved by changing the parameters away from defaults, the changes were usually small and the impact marginal (i.e. pathological failure is very rare).

The emphasis during development was on robustness rather than speed. E.g. as written the sky determination is a relatively expensive operation, and by default this is done three times with increasingly aggressive object masks in order to be robust against biases due to extended objects and crowded fields. Even with the iterative object dilation and sky subtraction routines turned off, PROFOUND is still notably slower and more memory intensive than SExtractor (Bertin & Arnouts 1996), typically a factor of a few slower for the same data when achieving a similar number of source extractions. Since it is largely written in R in a highly functional manner, there is a lot of data copying between different levels of functions, although efforts are made to minimise this whilst still preserving the safe and functional nature of R code.

Typically the processing time scales with the number of pixels, i.e. a 2kx2k image will take a similar time to process as four 1kx1k images. However, this scaling only continues until the point the various objects produced within PROFOUND during processing can still be held in random access memory (RAM) without having to use disk based virtual memory. Since various internal objects are necessarily created that have the same dimensions as the input image, the maximum practical image size without subsetting the image is of order a tenth the available RAM (before potentially compromising the processing speed). On a modern computer with ~8GB of RAM this is in the region of a 10kx10k image. On a MacBook Pro with 16GB of RAM we were able to process 20kx20k images without resorting to disk based virtual memory and the associated slow down. There is also a low memory mode which radically cuts down the RAM requirements to be nearer to double the image size, but removes many of the useful outputs to save RAM (e.g. the sky and sky-RMS maps).

Code Flow of PROFOUND

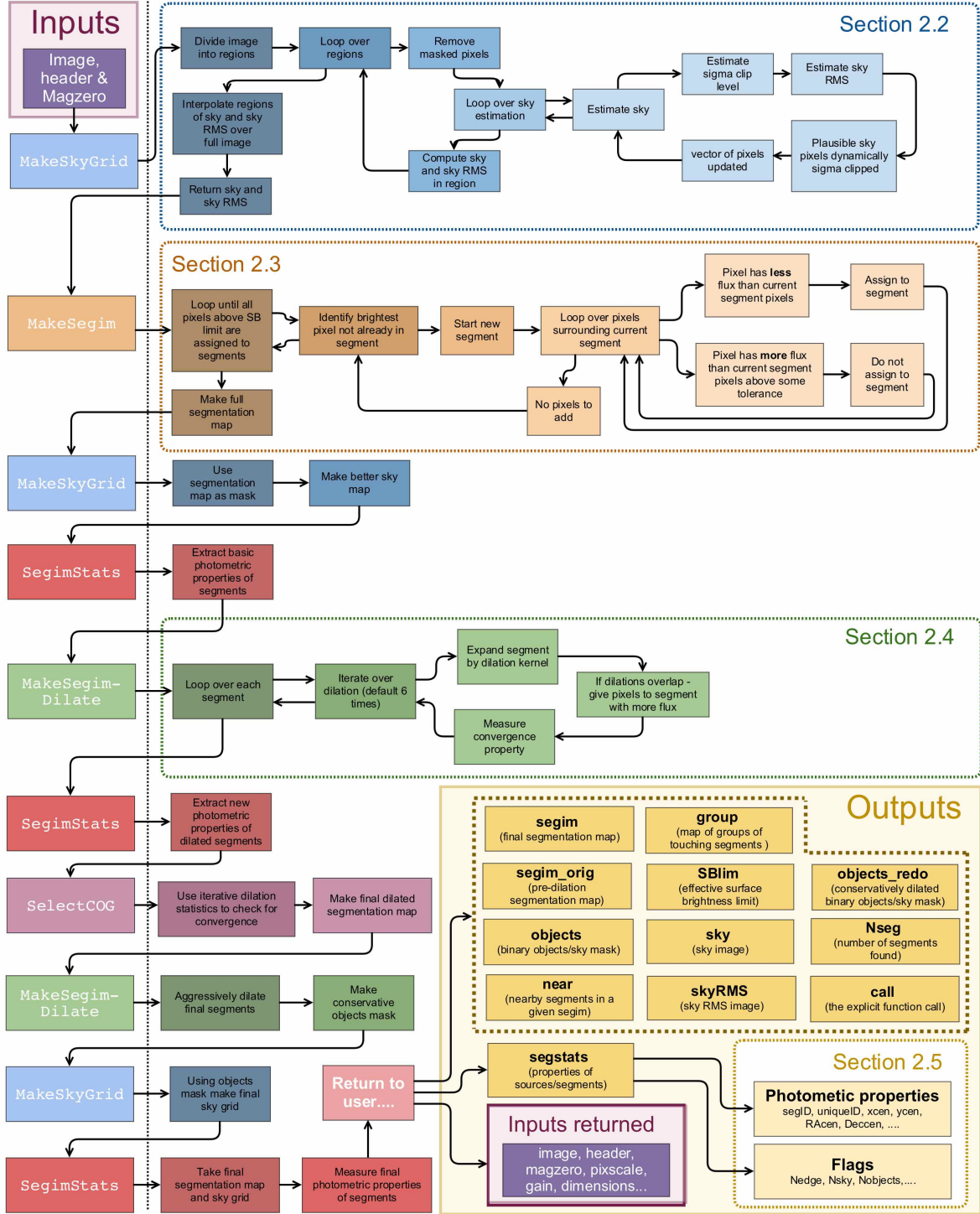


Figure 2. Detailed code diagram for the highest level PROFOUND function. This calls a number of mid-level functions that we discuss in detail in the relevant labelled Section.

A limitation compared to SEXTRACTOR is that the main PROFOUND function cannot natively handle matrix-like images much larger than 46k×46k pixels (strictly a hard limit of $2^{31} - 1$ pixels in total), even on machines with much more memory available. A higher-level function (PROFOUNDLARGE) is included to process very large FITS files, which extracts overlapping subregions directly from a target FITS file, and recombines them in an unambiguous manner. The UltraVISTA survey data discussed later in this paper was processed using the PROFOUNDLARGE function since it is marginally too large to be extracted in a single pass and served as useful test data, although the sub-region of interest for the Deep Extragalactic Visible Legacy Survey (DEVILS; Davies et al. in prep) is ultimately smaller than the $2^{31} - 1$ pixel limit. A reason to take this approach is that PROFOUNDLARGE can compute the data in an embarrassingly parallel manner, whereas most of the routines within PROFOUND are inherently single-threaded in nature. Routes exist to compile R with support for multi-threading, but this requires reasonably advanced knowledge, and our assumption is most users of PROFOUND will be using the standard single-threaded version of R.

For a typical survey image run with default parameters the various stages take fairly predictable proportions of the total computing time: calculating the full watershed deblend for the segmentation map dominates the total time (~50%), followed by calculating the sky map (done three times by default, ~20%), iteratively dilating the image segments (done six times by default, ~20%) and calculating the photometric properties of the segments (done once per dilation step and again at the end, so seven times in total, ~10%). Since the watershed stage dominates the time and uses an efficient external function written in C, the potential optimization gains are quite moderate. Reductions in processing time are possible if the number of sky calculations and/or the number of dilations are reduced from the defaults. When run in matched segment mode the processing time is significantly reduced since the watershed deblend stage is no longer required. If the sky subtraction and dilation steps are also turned off then the total processing time can be reduced by a factor ~10, i.e. the minimum requirement is that photometric properties of the provided segments are computed.

2.2 Sky Subtraction

An important step in almost any approach to object extraction from astronomical imaging data is the sky subtraction. Depending on the origin of the data the image might arrive to the user fairly flat and featureless in the background (e.g. optical drift scan data with the Sloan telescope; Ahn et al. 2014) or complex and variable at a number of different scales (e.g. near-infrared data with variable fringing artefacts; Andrews et al. 2014).

Given the potential complexity of the sky background it is rarely sensible to attempt to construct a formal statistical model of the sky since it is very difficult to meaningfully parameterise the range of behaviour observed (see Bijaoui 1980; Irwin 1985). Instead most popular astronomy applications have taken the route of a heuristic but visually appealing and pragmatically achievable scheme: coarsely sampled sky measurements combined with a polynomial interpolation

(Bertin & Arnouts 1996). Key parts of this process are that *true* sky pixels need to be identifiable in the target image, and some manner of estimating their variance and absolute level is possible.

The above is achieved in a practical manner by clipping likely objects out of the data and using a sliding box car filter on a grid to measure image properties. With a meaningful sampling of the sky and sky-variance a traditional scheme to interpolate between grid points (e.g. bilinear or bicubic) can then be used to construct a per-pixel estimate of the sky. The caveats to this process are that objects need to be well masked (else they will systematically bias the estimators) and the box car scale has to be well chosen so as to remove the real sky variations and not structure that belongs to objects in the image. The sky is extrapolated at the edges, which can cause artefacts if it is changing rapidly and bicubic interpolation is used. In this scenario bilinear interpolation is the safer option since it does not use higher order polynomial terms that cause this effect.

The main high-level sky estimation routine in PROFOUND (MAKESKYGRID) allows users to pass an image, masks (both for flagged pixels and for identified objects), the box car size, the grid sampling and the type of sky interpolation to use (bilinear or bicubic). This fairly simple functional interface allows for a large amount of flexibility in usage. The default code flow is as follows (names of the R functions used are displayed in small caps):

- (i) Divide the image into regions based on the requested box car and grid sampling. By default the grid sampling inherits the box car size, meaning pixels are evaluated once
- (ii) Each sub region is analysed separately in a large loop:

- (a) The masked pixels are removed from analysis, leaving a vector of fiducial sky pixels (*skyfid-pix*)

- (b) Then the following is computed iteratively (either until the clipping is converged, or after 5 iterations):

- (1) The sky value is estimated as $sky = \text{MEDIAN}(skyfid-pix)$

- (2) The dynamic sigma clip level is estimated to be $\sigma_{clip} = \text{QNORM}(1 - 2/N_{sky})$

- (3) The standard deviation of the sky pixels is estimated as $sky_{RMS} = \text{QUANTILE}(skyfid-pix, 0.5) - \text{QUANTILE}(skyfid-pix, 0.159)$

- (4) The plausible sky pixels are dynamically sigma clipped such that pixels satisfying $skyfid-pix > sky + sky_{RMS}\sigma_{clip} \vee skyfid-pix < sky - sky_{RMS}\sigma_{clip}$ are removed

- (5) The vector of *skyfid-pix* is updated and these new fiducial sky pixels are used for the next iteration

- (c) Once convergence has been achieved the final computed *sky* and *sky_{RMS}* values are returned for the region under consideration

- (iii) With all regions having a unique estimate of the *sky* and *sky_{RMS}* a bilinear or bicubic interpolation scheme is used to calculate plausible *sky* and *sky_{RMS}* values for all pixels

- (iv) The *sky* and *sky_{RMS}* images are returned to the user or higher level function in a list

As discussed above, the sky is measured a number of times during PROFOUND source extraction. The basic design philosophy is that it is possible to measure more accurate values for the sky and sky RMS as the objects are extracted, leaving behind increasingly certain sky pixels, more accurate identification of true sources, and better photometric measurements of the segments.

2.2.1 Different Sky Estimators

In principle the clipping process works on positive and negative pixels, but unless there are serious artefacts in the data it will work predominantly on the positively valued pixels by removing undetected and un-extracted sources. The clipping process can be changed so as to not clip out possibly biased pixels, and the type of estimator for the sky can also be changed from median to mean or mode. Whether or not these options are used depends on the type of image being analysed, and what the aim of the source extraction is. An implicit assumption in PROFOUND (and indeed SEXTRACTOR) is that the sky fluctuations are symmetrically distributed around some intrinsic value. This will only be true in detail when the number of sky photons is fairly large (many dozens or more) and we can use the Normal \sim Poisson approximation.

The difference between the sky level estimators (most commonly the mean, median or mode) and sky RMS level estimator (e.g. quantile versus standard-deviation) is an interesting point to consider. In the toy situation described it is preferable to use the mode (or possibly median if the mode is noisy and/or poorly sampled) for the sky and the quantile for the sky RMS, since these are both systematically nearer to the specified ‘intrinsic’ sky for these simple estimators (Irwin 1985). However, what really matters is the origin of the positively biased flux that skews the distribution. There are a number of processes that generate the ‘sky’ and contribute to the extended area signal in a typical digital detector image, in roughly descending order of importance:

- (i) The actual night sky caused by Earth’s atmosphere glowing. Can be quite spatially and temporally variable (e.g. NIR images, distant artificial lights turning on and off)
- (ii) Flux scattered around the image due to telescope optics or instruments producing scattered light and/or very broad \sim Lorentzian wings
- (iii) Scattering of astronomical light by the Earth’s atmosphere, usually at very small scales (close to Gaussian usually)
- (iv) Intrinsically broad features caused by the Milky-Way’s foreground cirrus
- (v) Intrinsically broad wings caused by extra-galactic sources (e.g. the low surface brightness wings of galaxies or intra halo/cluster light)
- (vi) Undetectable faint compact sources, can be structured (Milky-Way stars) or effectively uniform in distribution (high redshift galaxies)

The question is then: which of these do we wish to remove? The answer clearly depends on what sources we are trying to extract photometry for. If we are measuring stellar photometry of a bright star then we almost certainly want to remove 1/4/5/6, i.e. we want to model the light from the star that has been scattered by the atmosphere and the

telescope (assuming it is the dominant contribution close to the star being modelled). If we want to profile a faint galaxy then you probably need to remove 1/2 (assuming it is mostly caused by other sources)/3 (assuming it is mostly caused by other sources)/4/6, i.e. we want to keep the faint wings of the target galaxy intact. There is not a trivially right answer, but PROFOUND does offer a few routes to compute these different types of sky which are described at length with examples in some of the available vignettes².

In summary, a PROFOUND user might reasonably prefer a median, mode or mean type sky, depending on the use case. If the source sits on top of the sky (whatever makes it) then the user probably wants to use the more biased mean estimator. If the source is the dominant part of the observable background (e.g. when profiling the faint wings of galaxies) then the user might prefer the less biased median or mode estimators and more aggressive source clipping.

A final issue is whether you can extract better galaxy profile models by also using PROFIT to model the background for a given source, where PROFIT has the capacity to model a flat local floor for the sky. The three obvious options are:

- (i) Use PROFOUND sky subtraction on the image and do not fit a sky background in PROFIT
- (ii) Use PROFOUND sky subtraction on the image and also fit a sky background in PROFIT
- (iii) Do not use the PROFOUND sky subtraction on the image and only fit a sky background in PROFIT

For pragmatic reasons of needing to remove complex sky that cannot be fully generatively modelled with PROFIT, options (i) or (ii) are likely to be the best strategy for the majority of use cases.

2.3 Segmentation Map

Due to the emphasis on making segmentation maps that are useful inputs for PROFIT galaxy profiling, the method of making segmentation maps differs in important ways to the method used in Bertin & Arnouts (1996) (i.e. SEXTRACTOR, which in turn was inspired by Beard et al. (1990)). The most significant practical difference is that pixels that are flagged as being above a requested surface brightness threshold (be that stated in terms of sky RMS fluctuations or absolute surface brightness) are de-blended using a non-discretised watershed algorithm that creates flux de-blends through two-dimensional saddle-point cuts in image space, rather than one-dimensional cuts in flux space. The watershed approach outlined here is nearest in spirit to the ‘FOCUS’ de-blend method of Jarvis & Tyson (1981) and variants of the popular Meyer (1994) ‘priority flood’ algorithm (see Zhang et al. 2015; Zheng et al. 2015, for recent astronomical applications). PROFOUND uses the iterative process outlined below:

- (i) Identify the brightest pixel in the image above the specified surface brightness level which is not already assigned to a segment

² <http://rpubs.com/asgr/>

(ii) Progressively search unassigned image pixels surrounding the current segment, for each pixel searched:

- (a) If a searched pixel has less flux than any neighbouring pixels already assigned to the segment, then assign to the current segment if no unassigned pixels neighbouring the pixel under consideration have more flux
- (b) If a searched pixel has more flux than its neighbours already assigned to the segment above some tolerance level then do not assign it to the current segment
- (c) If no more pixels can be assigned to the current segment then terminate the growing process

(iii) Select the next brightest unassigned pixel remaining in the image and assign it to a new segment, then repeat the above segment growing process

(iv) Once all pixels above the specified surface brightness level have been assigned to a segment terminate the watershed process

There are a small number of parameters that have a significant effect on the watershed process, and in practice these need to be slightly altered to best segment the data under consideration. The most important is the ‘tolerance’, which specifies to what degree pixel growth is allowed to traverse uphill within a segment. In practice this determines the level of de-blending between closely separated flux peaks, where a higher tolerance means less splitting up of extended regions of flux. This is always specified in terms of the RMS of the sky and is 4 by default, i.e. a fluctuation would need to be more than 4 deviations of the sky RMS above the neighbouring segment pixel in order to not be included as part of the current segment being grown. The next most important parameter is the smoothing applied to the image (sigma), where by default a Gaussian kernel with a standard deviation of 1 pixel is used to blur the image. The smoothing can be turned off entirely, but this is rarely a good idea since there is always a large degree of pixel-to-pixel noise in all but most correlated images (typically flux values only appear smooth on the scale of a few pixels). Instead it is sometimes justifiable to increase the smoothing size, with an upper limit of 3 or 4 sometimes more suitable for images of large well-resolved galaxies. Related to this is the ‘ext’ parameter which is passed directly to the EBIMAGE WATERSHED function used in PROFOUND (Gregoire et al. 2010). This determines the allowed search radius around each pixel (rather than restricting the search to immediately adjacent pixels), where the default is 2 pixels. The smoothing ‘sigma’ and the search radius ‘ext’ have a similar impact and small changes in at least one of them (over the range 1–4) are common when first applying PROFOUND to a new dataset and tuning for optimal segmentation. Increasing the smoothing is an effective strategy for extracting extended low surface brightness sources.

Figure 3 demonstrates how the de-blender used in PROFOUND works in practice for different watershed tolerance levels. The main consequence of using the above approach is that islands of segments within larger segments can never be created. Instead the de-blended segments reflect the segment water would flow into if the flux map was turned upside down (ignoring dynamical effects like momentum, and just specifying the pixels that water would next flow into if the velocity was set to zero). There are other definitions of ‘wa-

tershed’, but this is the most standardised in the field of geology and accurately reflects a true gravitationally influenced watershed map. An internal Sloan Digital Sky Survey memo discusses the relative merits of different de-blending approaches³, with the general remark that all approaches make compromises and assumptions. This is true even for full generative modelling, given the requirement to define the parameterization of the model.

For our stated aims of producing good segmentation maps for passing into PROFIT, the type of saddle point segmentation discussed above works well since it errs on the side of ignoring pixels compromised by nearby sources when modelling the profile of a target galaxy. Qualitatively we also see relatively few occasions of segments being grouped together in a common aperture erroneously, which we sometimes see to occur when running SExtractor (see Wright et al. 2016, and the UltraVISTA example below for examples of such situations).

The full PROFIT model of the central complex shown in Figure 3 yields photometric properties (most notably flux) not far removed from the PROFOUND estimates (less than 0.5 mag differences). However, it is in general more important to get the correct number of segments and mode locations than having very good initial estimates for the fluxes and sizes. This is because estimating the number of components and mixtures is more difficult when galaxy modelling than optimising the parameters of a given model. Whilst PROFOUND does include routines to improve object measurements based on symmetry and flux sharing, these are turned off by default since they are computationally costly and generally unnecessary (i.e. the raw measurements are good enough).

The underlying code that computes the watershed de-blend comes from the image processing package EBIMAGE that is already available in R and widely used for low-level image processing (Gregoire et al. 2010). Its design and focus was for cellular biology (the EB in the name standing for European Bioinformatics) where the main task tackled was how to correctly segment images of cells taken by microscopes. As often noted anecdotally, there is much similarity between an image of biological systems and astronomy images, the former probably being the more complex to organise and segment in a systematic manner. For this reason it is not surprising that a tool developed for such an application works well for astronomical images. An important feature of the routine used is that it does not discretise the flux levels in the image (as SExtractor does) and uses the full flux resolution available. This fact means that it is quite slow (despite the underlying code being written in C), and the watershed step usually dominates the computation time for larger images since it scales as $O(n \log n)$ whereas nearly every other subroutine scales as n or better (where n is the number of pixels in the target image). The effect of this is that it can be faster to run PROFOUND on sub-regions rather than one very large image. It is only beyond the size of 10k×10k images (more than 10^8 pixels) that the difference becomes worth considering.

For convenience when using the outputs of PROFOUND the identities of neighbouring segments with respect to all other segments and also the friends-of-friends groups of de-

³ <http://www.astro.princeton.edu/~rhl/photomisc/deblender.pdf>

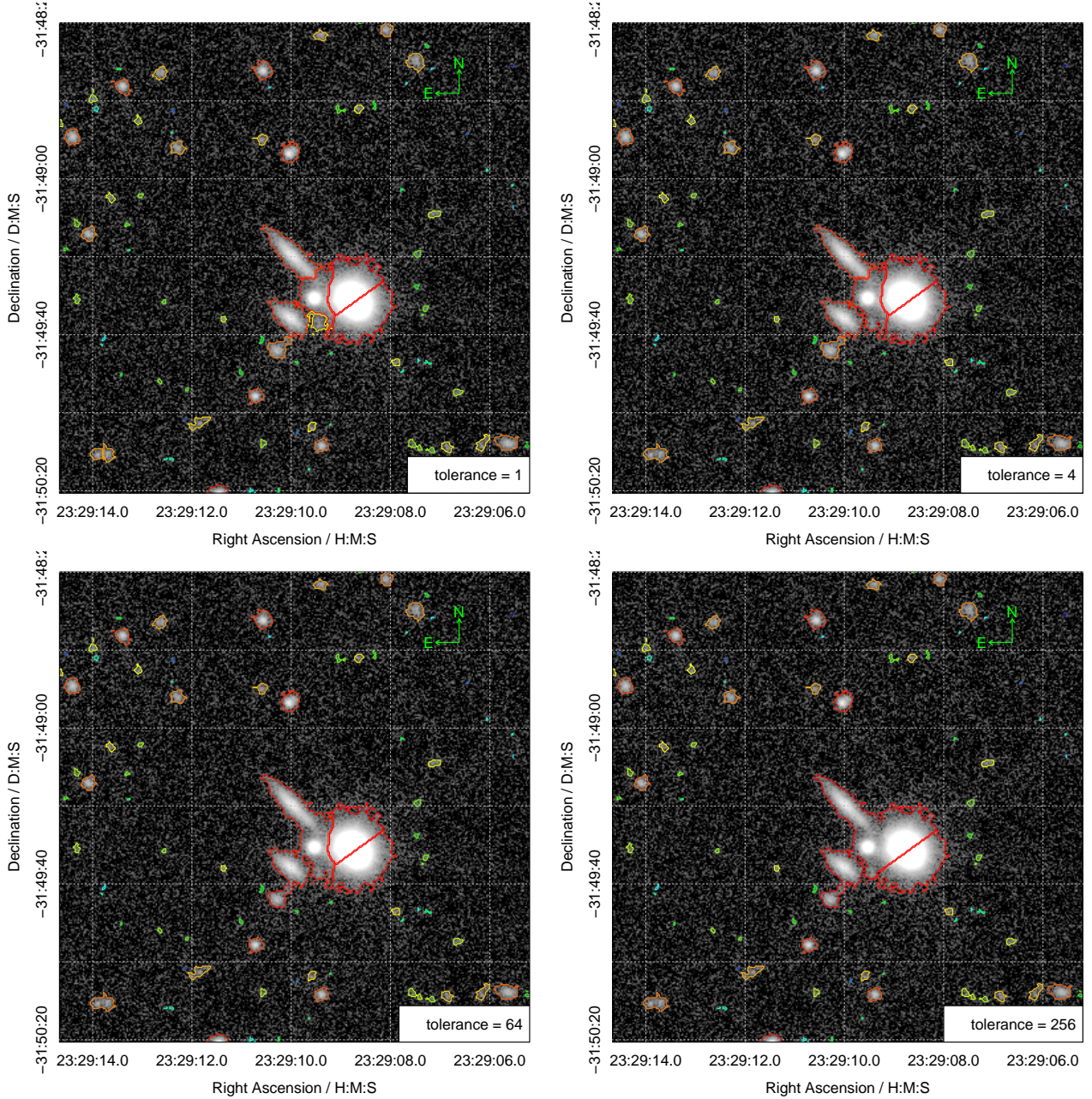


Figure 3. Examples of using different watershed de-blend tolerance levels on the initial segmentation map, as per labelled on each panel. There is no truly objective approach to say which solution is preferred, but most professional astronomers would probably suggest the best answer lies in the regime of tolerance 1–4, where the central confused complex has been broken up into its plausible sub-components.

blended regions can also be returned. The latter is particularly important when using PROFOUND as an input to PROFIT, since you should minimally try to fit all the segments in a grouped friends-of-friends region when trying to profile blended objects. These two types of grouped structures are not trivially returned by SExtractor, so for creating profiling inputs this is a clear advantage of using PROFOUND.

2.4 Segment Dilation

The next phase of the source extraction routine grows the segments using circular top-hat (by default) dilation opera-

tions until convergence has been achieved. As highlighted in the introduction, the segment dilation method is the most novel aspect of how PROFOUND operates. At no stage are fluxes or object properties calculated using apertures (be they circular or elliptical). Instead all integrated properties related to the source are estimated using the dilated segments alone. The procedure is iterative and therefore relatively expensive compared to simply expanding the inner Kron or Petrosian radius by some factor that approximately contains a large fraction of the flux.

By default the code looks for convergence in flux at some tolerance level, but in theory any property of the catalogue generated can be used to ascertain whether the seg-

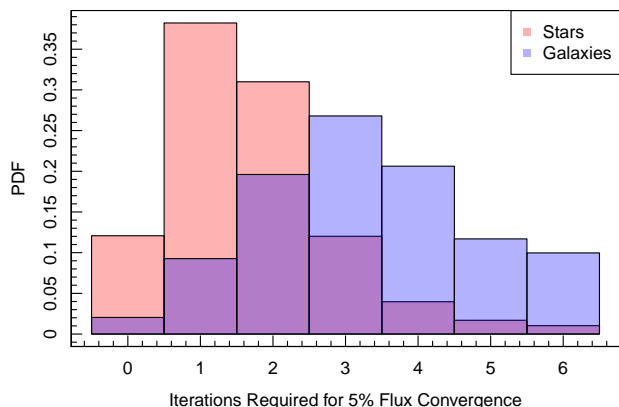


Figure 4. Histograms of the number of iterations required to reach flux convergence for both stars and galaxies. The data is taken from the large suite of simulations that we discuss in Section 4.

ment being grown has converged and the dilation should stop. Figure 4 shows the required number of dilation iterations before flux convergence is achieved for a large suite of simulated data containing 40k stars and galaxies that we discuss in detail later. It is notable that typically stars require fewer iterations than galaxies in order to achieve the default level of convergence (5%). The minimum number of iterations is zero, but only a very small fraction of stars require so few dilations. More typically stars require two dilations and galaxies require four.

The dilation distribution has not fallen to zero for stars, and more clearly galaxies, even by iteration six. This suggests that galaxies in particular have such extended flux envelopes that they need even more dilations. It is possible to increase the maximum allowed number of dilations to greater than six, but this was considered to be a sensible compromise default value. The sources requiring six dilations are notable for having the lowest integrated surface brightness levels of all sources (i.e. these are the most marginal detections), so the danger of pushing to much more aggressive dilation levels is that a significant amount of noise is incorporated into the aperture, compromising the photometric properties of the object being measured.

With the main design considerations for the segment dilation process now justified, the basic flow is as follows:

- (i) Execute a number (the default is six) of dilation operations on all segments, for each dilation operation:
 - (a) Expand every segment with a dilation kernel (by default this is a circular top-hat with a diameter of nine pixels)
 - (b) If segment dilations overlap then give all pixels to the segment containing more flux in the current iteration
 - (c) Measure the convergence property of interest for the new dilated segmentation map (by default this is the flux)
- (ii) After all dilations have been made look through all segments and determine when each segment has converged within some tolerance (this is within a factor 1.05 in flux by default)

- (iii) Put these converged segments together to make a final dilated segmentation map

The dilation operation is executed by the DILATE function in the EBIMAGE package. This is optimised in design for detecting the full extent of pixels belonging to an already labelled biological cell. A useful feature of PROFOUND is that it can accept any segmentation map as long as it has the generic feature that the segments are non-zero integers and the sky is labelled as 0. This means it is perfectly possible to pass in a segmentation from another software package (e.g. SExtractor) and then use PROFOUND to execute the dilation and flux convergence algorithms.

There is trade off to be made between setting the initial detection threshold of the image higher and allowing a larger number of more aggressive dilation operations, but the key point is that at least three (by default) pixels must be identifiable as a segment above the detection threshold in order to even be diluted. The default parameters work well on a range of common survey imaging data (e.g. SDSS (Ahn et al. 2014), KiDS (Kuijken et al. 2015), VIKING), so the need for large deviations from the defaults should be rare. In fact the settings related to the dilation operations are almost never altered in normal usage.

Figure 5 shows the segments that define the bright initial components of the segmentation maps and the fully dilated segments for the same objects. Comparing the undiluted segments (left panel) and the dilated segments (right panel) it is clear the bright stars do not dilate much before their flux converges, but fainter extended galaxies grow much larger in order to capture their converged share of flux. It is also notable that the geometries of the segments are kept largely intact during dilation.

Internally the dilated segments are used to compute a number of traditional geometric parameters such as effective size and ellipticity of the segment were it forced to be an ellipse. The approximate semi-major axis is output as ‘R100’ in PROFOUND, where the ratio between this and the standard Kron aperture (which is computed using the first order moments of the pixels in each segment) can be thought of as the expansion factor, which tends to be set to values between 2 and 4 when using aperture based photometric tools. Rather than being specified this is measured in PROFOUND. Figure 6 shows the typical factors for a suite of simulated VIKING depth data (introduced and discussed in more detail later in this paper). It is clear there is an absolute lower limit near to two, and only the very brightest stars and galaxies require expansion factors beyond four. The discreteness seen for the stars is a consequence of the iteration process, where the thickest branch contains stars that require two dilations for convergence, as seen in Figure 4.

Figure 7 gives an idea of how similar the dilated segments and the approximated elliptical apertures are in practice. The thick dashed white ellipses would perfectly follow the multi-coloured segments if the relationship was perfect. In practice they trace fairly similar shapes and largely contain the same pixels, hence Figure 6 should give a broadly accurate impression of the true expansion factors required. Where this relationship breaks down is in the highly clustered (and therefore de-blended) regions, in particular in the example data the central cluster of objects that PROFOUND

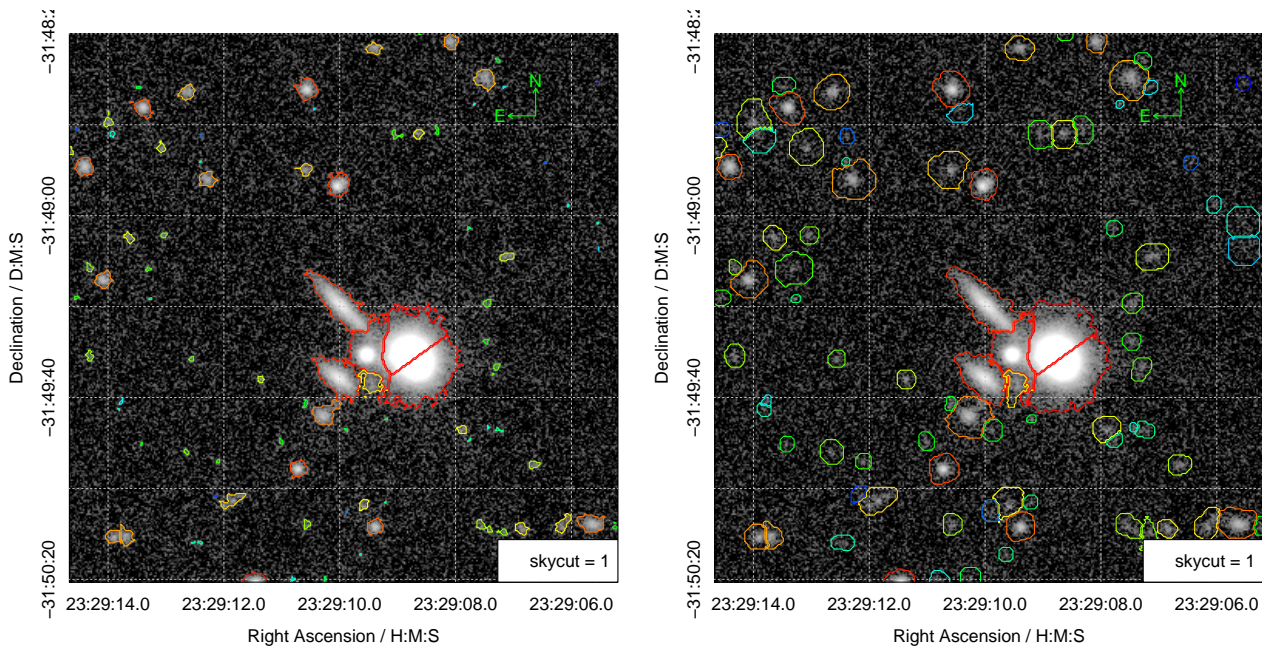


Figure 5. An example of the default diagnostic output from the main PROFOUND source extraction routine using the example VISTA Z-band data taken from the VIKING survey (as shown in Figure 1). The left panel shows the identified flux (non-dilated) segments via multi-coloured contours (red showing the brightest sources, via green, through to blue showing the faintest). The right panel is similar, but the contours now represent the fully dilated and flux converged segments. The left panel is actually identical to the top-right panel in Figure 3, but it is repeated here to aid direct comparison of pre and post dilation.

has de-blended through multiple saddle points. Since these segments will have necessarily non-elliptical geometry there is no good elliptical approximation for the segment regions. Since the elliptical apertures are not used to compute photometry within PROFOUND this is not a concern to people using the software in an isolated fashion, but it does mean that care has to be taken when attempting to apply these apertures using other programs (e.g. LAMBDA [Wright et al. 2016](#)).

2.5 Photometric Properties

Once a segmentation map has been constructed a separate internal function is used to calculate a large suite of photometric properties and segment flags. Internally this is achieved by associating all pixels with their respective segments, looping through each segment, extracting the relevant pixels for the current segment, and computing photometric properties using just the pixels flagged as belonging to a particular segment. To achieve this process rapidly PROFOUND uses the highly efficient `DATA.TABLE` package⁴ which is optimised for subsetting and processing on large datasets.

The main photometric properties returned are listed in Table 1 (ignoring the various types of flags, alternative definitions of some quantities, and error columns). These outputs are sufficient to provide reasonable initial guesses for a single Sérsic profile fit of a galaxy using PROFIT, which was the main initial design focus for PROFOUND. The assumption is that a multi-component profile will be built in complexity iteratively, i.e. in order to achieve good inputs

for a two component fit you would first start with a simple single component fit.

PROFOUND does not execute sophisticated algorithms to de-blend flux between neighbouring sources/segments. In comparison, methods presented in [Irwin \(1985\)](#) (simultaneous maximum likelihood of sources) and [Bertin & Arnouts \(1996\)](#) (heuristic flux division via symmetry expectation) do. In order to extract truly optimal photometry the identified blended sources should be further run through generative modelling software such as PROFIT (e.g. [Kelvin et al. 2012](#), see Section 3 for an example). However, PROFOUND includes a number of schemes to flag and improve photometry in complex and confused regions. Improved flux reconstruction is possible using the ‘rotated’ flux output for sources (‘flux_reflect’ and ‘mag_reflect’ in Table 1), which assumes flux symmetry of sources. This option determines a plausible amount of missing flux by rotating each segment about the central pixel and determining how much flux does not fall onto a mirrored segmented pixel. Running on the VIKING example data, the median difference between the raw segment flux and the rotated version is ~ 0.1 mag (i.e. the sources get brighter), so for most sources the difference is fairly small. The scale of these differences is also in line with the kind of flux differences seen when attempting full profile modelling with PROFIT (see Section 3).

PROFOUND also returns a large number of flags that can help decide whether there are issues with the segmentation, or other potential issues with the photometry (like lying very close to a frame edge). Table 2 is a summary of the major flags that are generated when the photometric properties of an image are computed. The `edge_frac` flag is particularly useful for flagging well isolated objects, and when this drops much below 1 it is a sign that future galaxy profiling might

⁴ <https://CRAN.R-project.org/package=data.table>

Table 1. A selection of photometric properties computed in PROFOUND

Name	Description
segID	Segmentation ID
uniqueID	Unique ID
xcen	Flux weighted x centre
ycen	Flux weighted y centre
RAcen	Flux weighted Right Ascension centre
Deccen	Flux weighted Declination centre
flux	Total flux in the segment in ADUs
mag	The flux in the segment scaled to a magnitude
flux_reflect	Total flux in the segment in ADUs scaled by flux missing under a segment rotation
mag_reflect	The flux_reflect in the segment scaled to a magnitude
N50/90/100	The number of pixels containing 50% / 90% / 100% of the flux
R50/90/100	Approximate elliptical semi-major axis containing 50% / 90% / 100% of the flux
SB_N50/90/100	Mean surface brightness containing 50% / 90% / 100% of the flux
con	The concentration, defined here as R50/R90.
axrat	Axial ratio of ellipse
ang	Orientation angle of the ellipse

Table 2. Flags and diagnostics computed in PROFOUND

Name	Description
Nedge	Number of edge segment pixels that make up the outer edge of the segment
Nsky	Number of edge segment pixels that are touching sky
Nobject	Number of edge segment pixels that are touching another object segment
Nborder	Number of edge segment pixels that are touching the image border
Nmask	Number of edge segment pixels that are touching a masked pixel
edge_frac	Fraction of edge segment pixels that are touching the sky
edge_excess	Ratio of the number of edge pixels to the expected number given elliptical geometry
flag_border	A binary flag telling the user which image borders the segment touches
flag_keep	A Boolean flag suggesting whether the object should be kept based on the flux growth and iterations

be compromised by nearby sources unless effort is made to execute a model that also accounts for these sources.

2.6 Colour Photometry

Colour photometry is a catch-all term that usually refers to measuring fluxes in multiple bands using common apertures, where the differences in fluxes can be mapped onto traditional optical colours for visualization purposes. The higher level PROFOUND function that provides the main interface to both source extraction and photometric analysis offers a few routes to extracting colour photometry. The top level interface can take a number of inputs that bypass internal routines to calculate them, e.g. segmentation maps, sky maps and object masks. Since the image provided does not have to be the same as the one used to create the segmentation map provided, it is easy to extract forced photometry by passing into PROFOUND a pixel matched image that was observed using a different filter to the detection band, and turning off the option to dilate the segments.

A more advanced method, useful in the case where the point spread function (PSF) varies significantly between bands, is to allow the segments to dilate to best extract converged flux in the target band. This is referred to as soft colour photometry in PROFOUND, and is a sensible method to extract total photometry across multiple bands with different depths and seeing conditions. In general the highest

image quality and/or deepest band should be used as the detection image. Additionally, PROFOUND includes routines to optimally stack images based on S/N properties, in which case a stacked image can be used as the detection image (this was used for the UltraVISTA data analysis presented in Section 5).

One issue is that the above methods require the images to be pixel matched. The PROFOUND package includes routines to remap images onto a common target Tan-Gnomonic world coordinate system (WCS), should the images not have a common projection. To do this PROFOUND uses the image warping routines available in the CIMG image analysis library, and accessible in R via the IMAGER package.

An example of this being applied to mismatching VISTA Ks-band (pixel scale 0.339 asec/pix [Edge et al. 2013](#)), Visual Survey Telescope r -band (VST; 0.2 asec/pix [Kuijken et al. 2015](#)) and Galaxy Evolution Explorer NUV (GALEX; 1.5 asec/pix [Martin et al. 2005](#)) can be seen in Figure 8. The remapping allows us to make a coordinate matched RGB colour image shown in Figure 9, where the VISTA Ks-band and GALEX NUV-band data are remapped onto the WCS of the VST r -band data. The upsampling conserves flux, and by default uses bilinear interpolation (bicubic or nearest pixel, and forward or backward mapping, are also options).

Using these remapped images it is simple to extract matched segment photometry by applying segments extracted from a detection band (in this case the VST r -band)

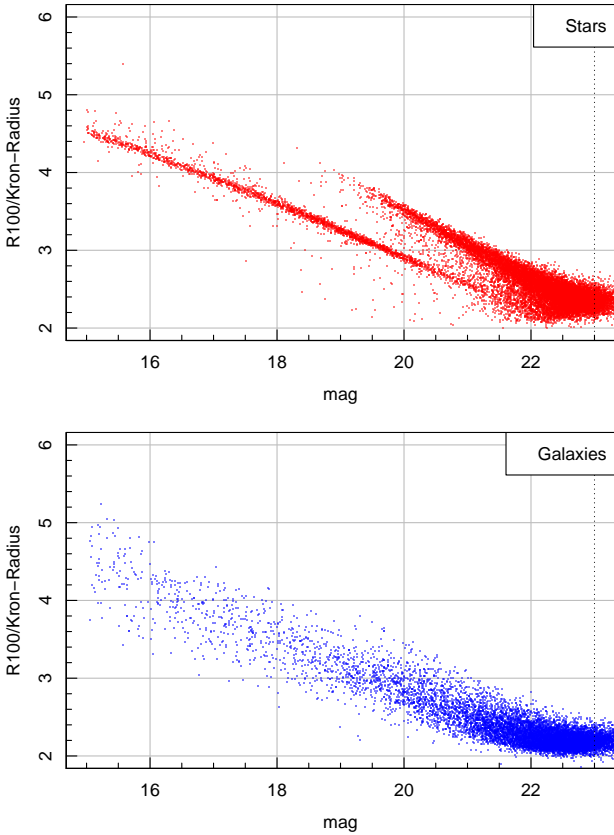


Figure 6. The Kron to fully diluted PROFOUND semi-major axis (R100) expansion factor for simulated stars and galaxies.

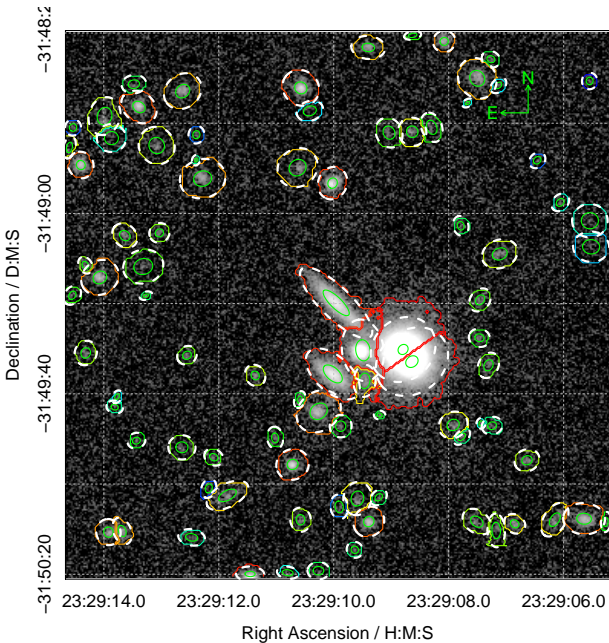


Figure 7. Overlay of the intrinsic dilated segments (multi-coloured solid thin lines) and the inner Kron aperture (green ellipses) and approximated elliptical apertures that best describe the dilated segment geometries (thick white dashed ellipses).

on the remapped target bands. Figure 10 shows what this extraction might look like internally, where it is clear the majority of the GALEX NUV flux associated with the central spiral galaxy is enclosed by the VST r -band derived segments. Some of the fainter NUV features would be very hard to extract blindly, but should produce a reasonable signal when extracted in such a forced manner.

The above approaches are the best solution to extracting matched aperture total photometry. However, more accurate spectral energy distributions (SEDs) are often derived from using only the brighter inner parts of sources, since the outskirts usually include lower S/N pixels which will act to increase the scatter between colours. Popular approaches to such colour photometry include fixed apertures (e.g. 2 arcsec circles placed on each source) or computing colours with a certain surface brightness level. The latter is achieved trivially in PROFOUND, since one of the objects returned is the segmentation map before dilation, i.e. the segmentation that only includes pixels that are independently above some surface brightness threshold. An example of such a bright segmentation map is shown in Figure 5, where the pixels identified are much brighter than the fully diluted segmentation map shown in the right panel of Figure 5. By using this map and turning off the option to dilate the segments high surface brightness colours can be extracted, leading to less scatter in the colour photometry measured (as we see in detail later using UltraVISTA data).

Finally, a hybrid colour is possible, where the bright segment map is used as the starting point in each target image, but the segments are allowed to dilate independently in each target band in order to achieve converged flux. This is often similar in output to just applying the diluted segmentation map to each target band without allowing for independent dilation, however it can be a sensible option if the detection image has a much larger PSF than one or more of the target bands, where the diluted aperture might be much larger than is actually necessary, and includes a large quantity of sky pixels which will lower the fidelity of the photometry extracted.

3 COMBINING PROFOUND WITH PROFIT

As discussed above, much of the design philosophy behind PROFOUND was to provide good quality inputs for PROFIT galaxy profiling. This includes: careful sky subtraction; sigma map construction via estimating the local sky-RMS map; good quality segmentation maps for extended sources and accurate initial conditions for the profile parameters to use when fitting with the PROFIT engine.

Here we use these various elements of PROFOUND to prepare the example VIKING data for a large multi-component fit⁵. Running PROFOUND in default mode, but with the ‘boundstats’ option turned on, creates all the outputs we need. In this example we aim to fit the central group of objects that have touching segments.

Figure 11 shows the main results, where the top panels show the initial parameter estimates taken from PROFOUND,

⁵ for full details see the ‘Complex Fit’ vignette at <http://rpubs.com/asgr/>.

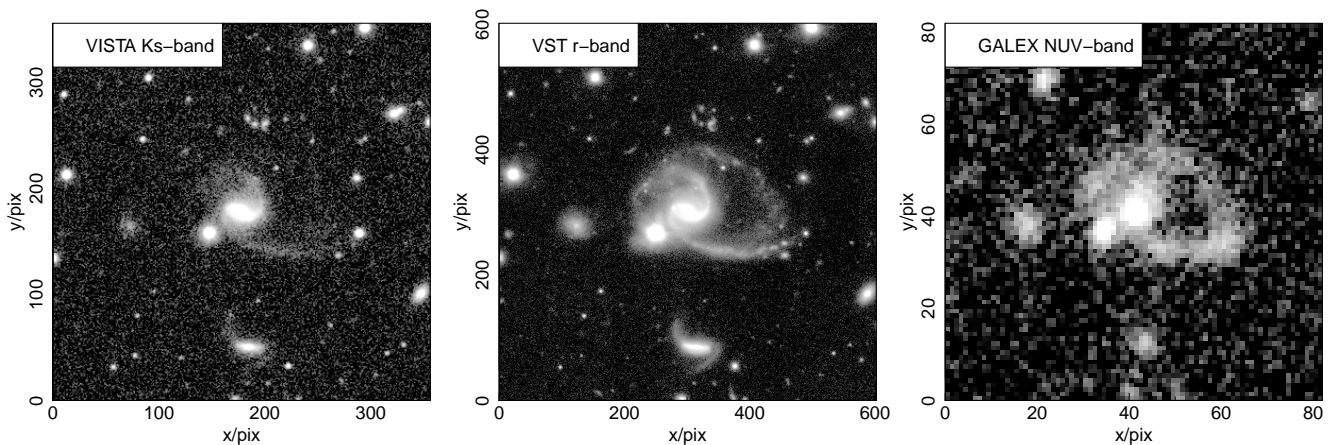


Figure 8. The PROFOUND package comes with some highly WCS mismatched images of the same galaxy (GAMA galaxy ID G202627). The left panel shows VISTA Ks-band (pixel scale 0.339 asec/pix). The middle panel shows VST r -band (0.2 asec/pix). The right panel shows GALEX NUV-band (1.5 asec/pix). Each image has dimensions of $2' \times 2'$.

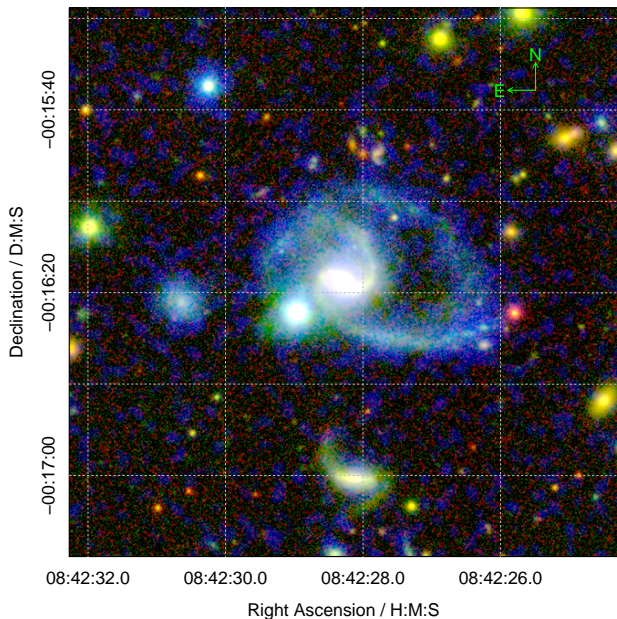


Figure 9. RGB composite image of mismatching projections combining VISTA Ks-band, VST r -band and GALEX NUV-band (as shown in Figure 8) for the red / green / blue channels respectively.

and the bottom panels show the PROFIT BFGS optimised solution. This particular fit uses Sérsic profiles for the two visually extended elliptical sources, and Moffat profiles for the remaining three objects which have PSF like characteristics. The Moffat PSF required for convolving the image and modelling the point-sources was estimated from fitting a number of isolated bright PSFs.

Overall we can achieve an excellent and rapidly converged simultaneous fit using this approach. The two extremely bright stars in the bottom-right of the fit region have some residual structure, but the relative flux residuals are generally small (a fraction of a percent). The other three sources are very well modelled, in particular the two

extended elliptical sources that have profiles very close to pure exponential discs.

The input PROFOUND and output PROFIT source fluxes all agree within 0.4 mag, and the differences are typically less than 0.1 mag. Even the two close bright stars are well estimated by PROFOUND, with the differences being 0.07 mag fainter and 0.16 mag brighter. The total modelled flux in the fit region agrees to better than 1% with the extracted PROFOUND flux. The R_e for the Sérsic index is also well estimated, the PROFOUND input increasing by $\sim 5\%$ for both of the clearly extended sources.

To achieve a rapidly converged fit for such complexes of objects the key requirements are that the objects are well segmented along flux saddle-points, and the initial estimates for the fluxes and sizes are within a factor of ~ 2 of the correct solution. PROFOUND can easily achieve these requirements if it is run in a sensible (usually near to default) manner. This suggests that using PROFOUND combined with PROFIT as part of an automated pipeline is a reasonable goal for future large scale decomposition tasks.

4 SIMULATIONS

4.1 PROFIT Simulations

To check the performance of PROFOUND we ran a number of tests using simulated data that was designed to approximately mimic the sky variations, sky RMS, PSF, magnitudes, sizes and profiles of a mixture of stars and galaxies in a typical VIKING survey frame using the image generation capabilities of PROFIT to make the simulated frames. Table 3 details the various parameters and sampling ranges used when generating the simulated images. Code to replicate very similar types of simulations are also available online for user experimentation⁶. In order to ensure we have sources going much deeper than the noise threshold of the VIKING data, the selection of magnitudes, source sizes and

⁶ see Simulated Images vignette at <http://rpubs.com/asgr/>

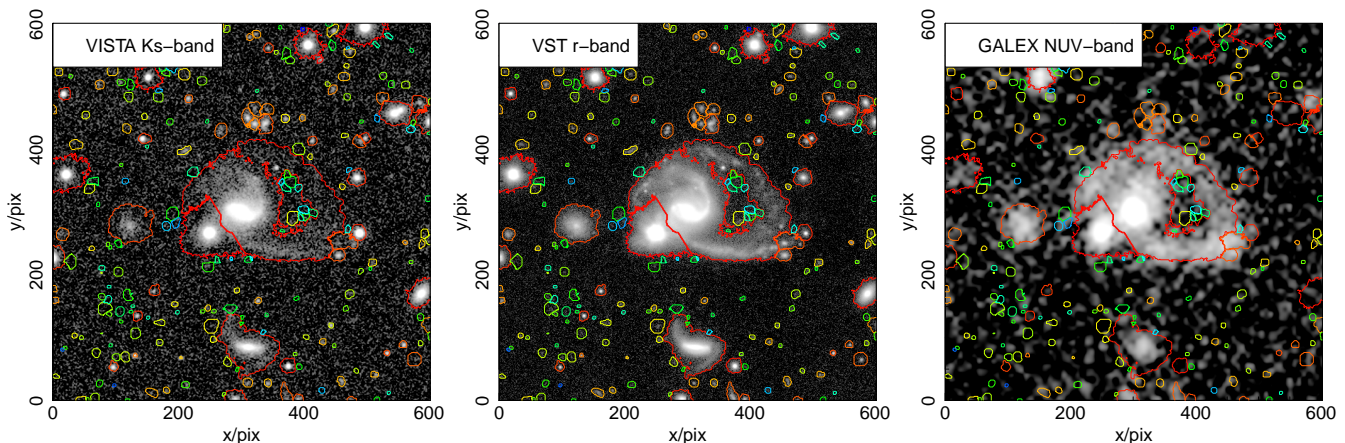


Figure 10. Schematic view of the PROFOUND segments defined using a common WCS system. In this case we use the WCS scheme from the VST r -band image as seen in Figure 8. The left panel shows the segments overlaid on the VISTA Ks-band. The middle panel shows the segments overlaid on the VST r -band they were defined with. The right panel shows the segments overlaid on the GALEX NUV-band. Each image has dimensions of $2' \times 2'$. It is notable that some of the GALEX NUV flux extends outside of the segment. PROFOUND can capture this additional flux if it is allowed to further dilate the provided segment (this is the default mode).

Table 3. The simulation setup parameters

Simulation Parameter	Value
N simulations	100
Sky	0
Sky RMS	10
magnitude zero point	30
x image pixels	1000
y image pixels	1000
PSF FWHM	5 pixels
N stars	200
Magnitude range	15–23
Magnitude Power-law slope	1.5
N galaxies	200
Magnitude range	15–23
Magnitude Power-law slope	2.0
R_e Poisson λ	5
Sérsic index range	1–4
Axial-ratio range	0.3–1
Boxiness range	-0.3–0.3

shapes was approximated using the deeper UltraVISTA survey recovered using PROFOUND (see Section 5). The PSF full width half max (FWHM) was chosen to be at the poorer extreme of values observed for VIKING: ~ 1.7 asec, or 5 pixels at the 0.339 asec/pixel scale of VIKING images (Edge et al. 2013).

Figure 12 is an example of the noise free model generated by PROFIT for the typical distribution of sources used for our simulations, the addition of noise and sky (estimated from PROFOUND on real VIKING data) to create more realistic looking data, and the extraction of sources using PROFOUND. In this example almost all of the stars were successfully recovered, and around half of the galaxies are detected, the rest being below the 1σ surface brightness threshold of the VIKING survey.

PROFOUND was run with close to default settings, with the difference being the de-blend tolerance (how many sky RMS deviations to use to de-blend sources during the wa-

tershed stage) was set to 1. These settings mimic the qualitatively optimal settings we have determined for processing VIKING and UltraVISTA data with PROFOUND (which in turn informed the majority of the default PROFOUND settings), but the broad results are quite robust to small changes in these settings.

One hundred $1k \times 1k$ frames were generated randomly with PROFIT model stars and galaxies and extracted with PROFOUND, with 200 stars and 200 extended galaxies generated per frame. This produces a final catalogue of 40k stars and galaxies generated. This is used to compute the false-positive and true-positive rates for stars and galaxies, and also quantify the measurement biases in PROFOUND compared to the intrinsic sources. The latter is important since one of the main design aims of PROFOUND (along with better sky subtraction and flux converged segmentation maps) is to create reasonable initial conditions for PROFIT fits. These do not need to be perfect, but it helps to be reasonably close to the global maximum likelihood in order to speed up the fitting time. Based on our experience with the simulations presented in Robotham et al. (2017), a factor of two in flux and/or size is a reasonable starting point for efficient convergence (in detail this statement is clearly algorithm dependent, and PROFIT offers no particular restriction on the optimization routine, giving out-the-box access to over 100).

Figure 13 shows the main completeness and spurious source extraction results for PROFOUND and SExtractor both run with default extraction parameters. The main trends with surface brightness and magnitude are as should be expected, where fainter objects are harder to accurately extract. As source brightness gets close to the 1σ surface brightness limit of the data the completeness drops off very sharply and the false-positive rate increases. Note that in the default mode used (skycut=1, which uses the 1σ sky RMS level as the extraction limit) the false-positive rate is always far below the true-positive completeness. This is a consistent finding when running simulations at a large range of depths, and suggests PROFOUND can be safely operated in this mode. The false-positive rate tends to get close to the

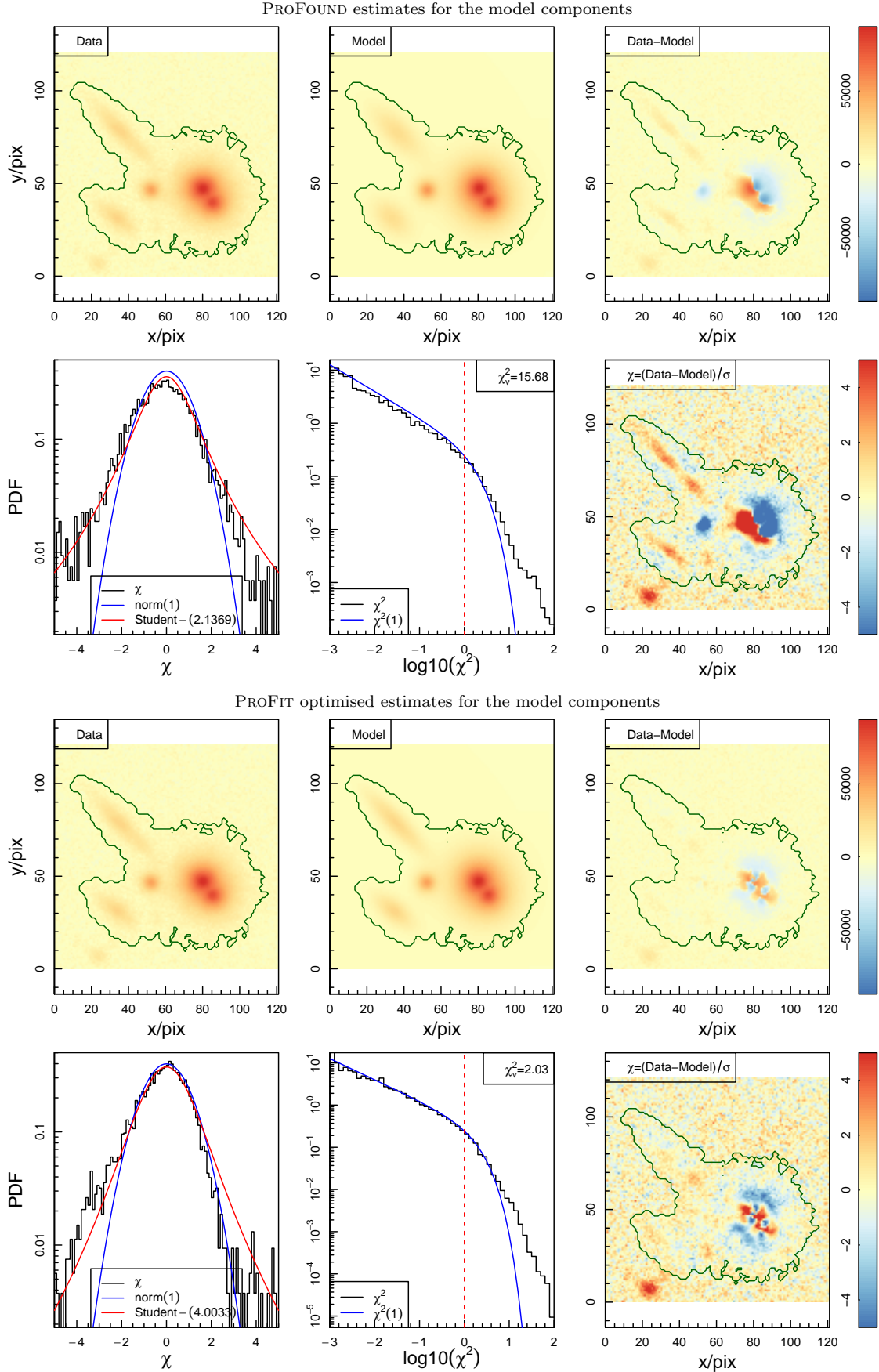


Figure 11. Example of fitting a group of confused sources by combining the initial sky subtraction, segmentation map, sigma map and parameter estimates from PROFOUND with the profiling and fitting engine of PROFIT.

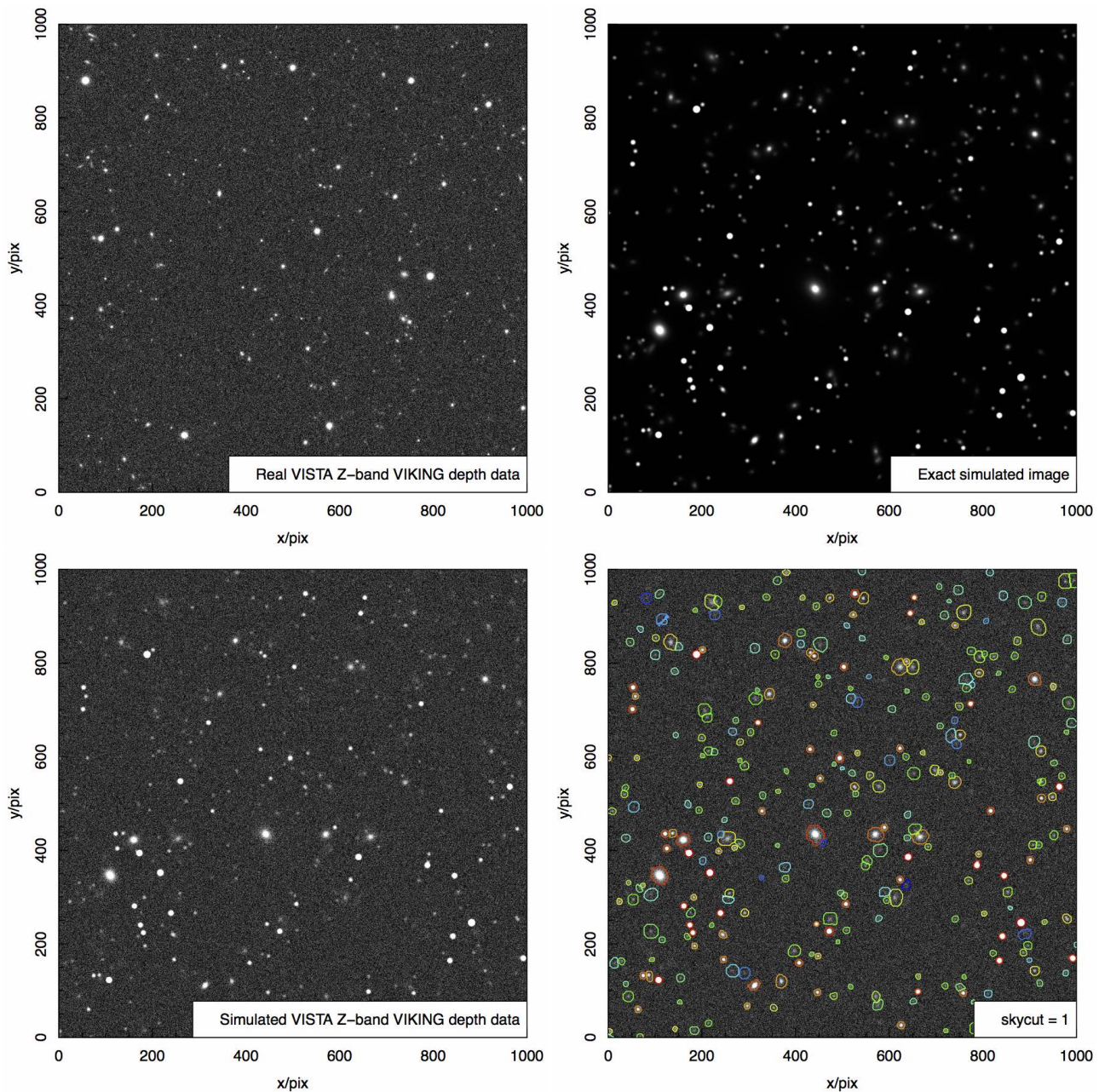


Figure 12. Top-left is a real VISTA Z-band VIKING depth frame. Top-right panel shows the pure model created by PROFIT with no noise added. Bottom-left panel adds realistic VIKING-like noise along with a variable sky background. The bottom-right panel shows the extracted PROFOUND segments.

true-positive rate for the faintest sources when skycut \sim 0.8, after which point you almost certainly would not want to push the extraction much further since most of the additional sources will be spurious. It should be noted that this limit is not trivially an extraction limit, since internal pixel smoothing and local clustering information is used to flag pixels that are likely to belong to real sources. Depending on the quality of the data and the pixel-to-pixel covariance the reasonable limit of skycut extraction might need to be higher than the value used here, however it is unlikely you could push much lower.

It is notable that the default extraction parameters of SExtractor are a bit more conservative, extracting fewer

sources (and suffering the associated incompleteness) but with a lower false-positive rate. This comparison should not be considered exhaustive, since very different curves are possible by changing the parameter setup of both PROFOUND and SExtractor, however it does give an indication of how things compare with little or no tuning for the data. Pushing deep into the sky noise is a complex topic that we do not aim to explore fully here, however efforts are under way to explore how best to operate PROFOUND in order to extract extremely faint and extended low surface brightness sources.

Figure 14 shows the accuracy in the extraction of magnitudes and sizes for stars and galaxies in the simulated images. As seen in the left panels, in general stars can be ex-

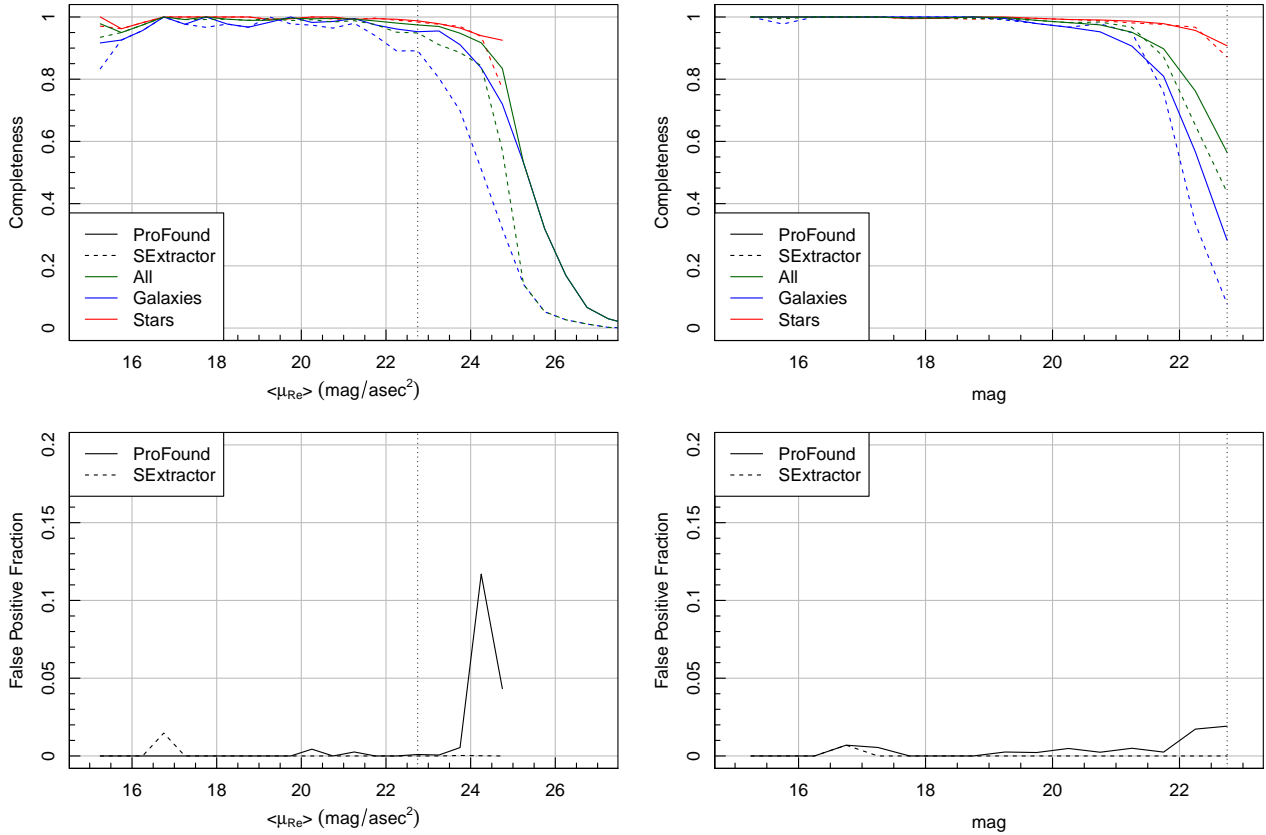


Figure 13. Top-panels show the detection completeness of stars and galaxies. Bottom-panels show the false-positive rate of spurious detections in PROFOUND. Left-panels are as a function of mean surface brightness within R_e . Right-panels are as a function of total magnitude. The approximate surface brightness and magnitude limits used for simulating the data is shown as a vertical dashed line.

tracted with better accuracy in terms of flux and size for a given magnitude. This is because the average surface brightness is typically much higher than for extended sources (i.e. galaxies). The extracted flux of stars show no systematic behaviour in the running median, even at the faint magnitude limit. The extracted sizes tend to show a small negative bias for brighter stars, which is due to pixel discreteness. At the faint magnitude limit, the sizes of stars becomes under-estimated because the dilation is halted by pixel flux noise.

The galaxy extractions show more scatter and also some systematic effects as a function of magnitude. Galaxies tend to have under-estimated fluxes at all magnitudes, but not surprisingly this effect is stronger for fainter objects. PROFOUND extracts fluxes for extended sources that are closer to the intrinsic values than SExtractor for all magnitudes, reflecting the utility of our dilated flux converged aperture approach. As we saw in Figure 4, extended sources tend to require more iterations before the flux is considered to be converged. Fainter sources tend to require more dilations to converge, which also means a greater fraction of their flux is introduced during dilation (in these simulations nearly 60% of flux is introduced during dilation for sources that require six iterations). In all these cases it is noise in the image that halts the dilation (since any drop in flux immediately terminates the dilation process), hence it is increasingly difficult to extract the same fraction of intrinsic flux for the faintest sources. Instead the sky RMS starts to dominate over the

object flux and the dilation halts. The behaviour of the extracted size is a bit more complicated, where very bright galaxies tend to have their sizes slightly overestimated. This is because large and brighter galaxies are more likely to be merged with nearby sources, further inflating their size. Towards the faint magnitude limit of the survey, the galaxies start to become under-estimated in terms of size, but the running median always stays within a factor of two of the input simulations.

In summary, PROFOUND is able to extract objects at close to the noise limit without introducing a large number of false-positive detections, and the sources extracted have photometric properties that have relatively little bias with source flux and within the tolerance we would want for further profiling using PROFIT.

4.2 LSST Simulations

In order to more directly compare PROFOUND against other source finding codes recently discussed in the literature, we ran it in default mode on some publicly available LSST simulation data (Connolly et al. 2010; Zheng et al. 2015). This data was used as a testing data set in Zheng et al. (2015), which presented a source finding approach that used a mixture of techniques to optimally recover sources blindly, with the focus on future large scale imaging surveys such as LSST. To allow a simple comparison we used default settings

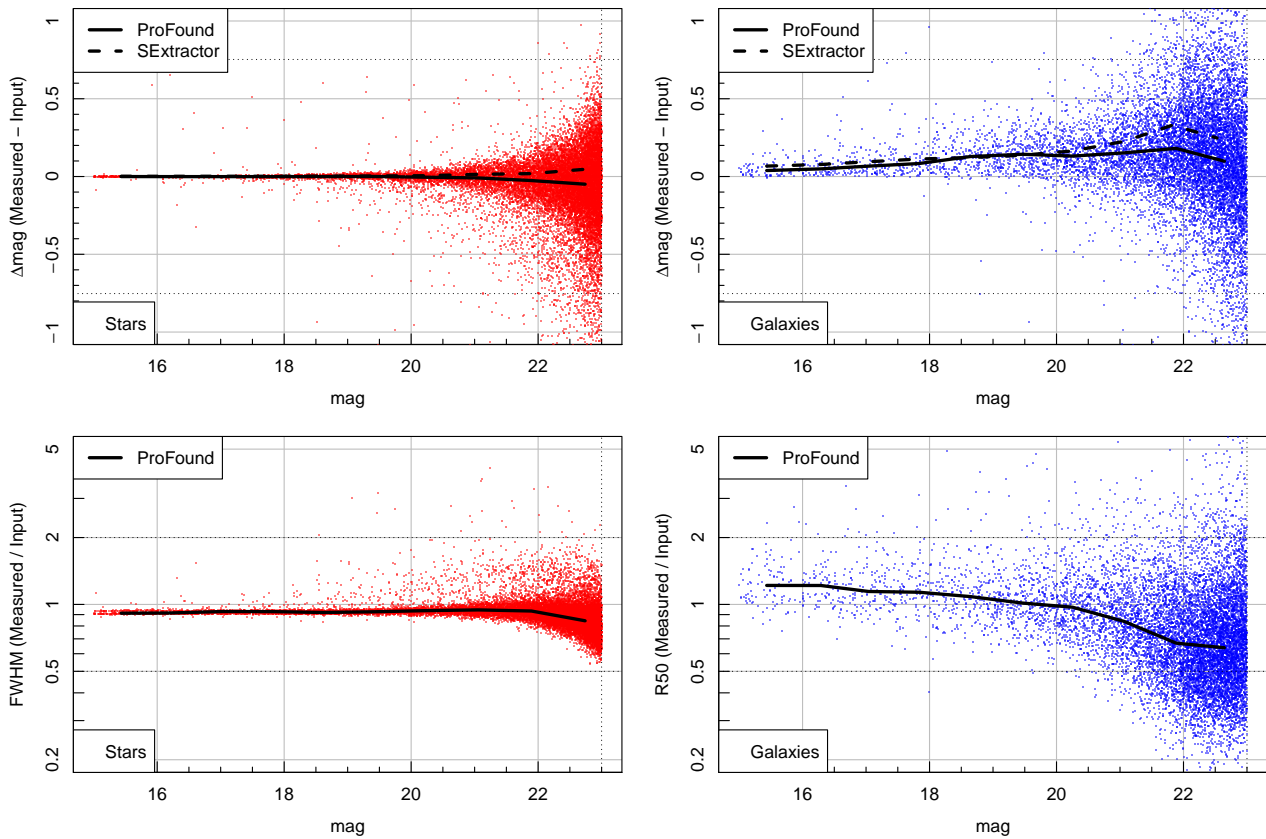


Figure 14. Main simulation results. The left panels show the star results in red. The right panels show the galaxy results in blue. The top panels show the difference between input and extracted magnitudes for both PROFOUND (points and running median line) and SExtractor AUTO (running median line). The bottom panels show differences in the estimated objects sizes (full width half max [FWHM] for stars, R_{50} for galaxies) but only for PROFOUND (SExtractor does not directly compute these sizes). The horizontal dotted lines show a factor of two error in flux or size, where 99% of stars and 70% of galaxies are within this tolerance.

throughout except for the sky background filter size, which was set to 64×64 in order to match the mode SExtractor had been run on using the same test data.

Running exactly the same images as analysed in Zheng et al. (2015) (Deep-32 and Deep-36, D32 and D36 from here) results in more detections than found in Zheng et al. (2015) (both the new extraction software presented and SExtractor). To ensure consistency we re-ran SExtractor with the parameters suggested, which resulted in an identical number of recovered sources to those presented in Zheng et al. (2015). Since the method of defining a true-positive match was not clearly described, we utilised our own matching criterion based on proximity in spatial position and flux, these becoming the three dimensions we use to determine true matches. Measuring the sky match distance in arcseconds and the flux in magnitudes, we define a match to be successful if the 3D match distance is within a radius of 2, i.e. if the intrinsic and recovered sources lie on top of each other, they are only matched in their fluxes are within two magnitudes.

Using this matching scheme, we find that of the 1548/1571 good sources (i.e. `flag_good=TRUE`) recovered by PROFOUND from the D32/D36 images, 1387/1413 were considered to be true matches (91.3% and 88.3% respectively). In direct comparison SExtractor recovered 1441/1386, of

which 1249/1182 were considered to be true matches (86.5% and 85.2% respectively). The total number of sources recovered using SExtractor is identical to the analysis in Zheng et al. (2015), however we find more true matches with our matching criteria (in that work 1189/1138 of the SExtractor sources are considered to be true matches). This suggests that our matching criterion is a bit more generous, but it at least allows for a fairly direct comparison between extraction methods. Just running with default parameters, PROFOUND recovers more sources at a higher true-positive rate than SExtractor, and more sources at a similar true-positive rate to the code presented in Zhang et al. (2015). A direct comparison with this code is harder given the differences in the matching criteria, but we certainly recover a lot more sources: they extract 1433/1375 sources with a claimed true-positive rate $\sim 91\%$.

It is possible to increase the true-positive rate of PROFOUND by changing a number of the available parameters. Equally, limiting the catalogue to the brightest 1433/1375 sources (to better match the extraction depth in Zheng et al. (2015)) increases the true-positive rate to 93.8%/91.3%. The appropriate balance of extraction depth and fidelity will be science dependent, but it is possible to investigate this thoroughly through the use of receiver operating characteristic (ROC) curves (true-positive versus false-positive plots). This

process can only be tackled through comparison to an absolute truth baseline, hence simulations should always be used when selecting appropriate parameters for a new data set.

Hence the more important finding is that PROFOUND performs in a highly competitive manner, even when run with default parameters.

5 APPLICATION TO ULTRA-VISTA DATA

The upcoming Deep Extragalactic Visible Legacy Survey (DEVILS; Davies et al. in prep.) on the Anglo Australian Telescope (AAT) will be conducting single-fibre redshift focussed spectroscopy for nearly 100% of galaxies down to a total Y magnitude limit of 21.2. DEVILS will be targeting three main regions: two VISTA Deep Extragalactic Observations (VIDEO; Jarvis et al. 2013) survey fields and the UltraVISTA field in the Cosmic Evolution Survey (COSMOS; Scoville et al. 2007). The source catalogues for UltraVISTA data release three (DR3) use a tailored run of SExtractor for total photometry and a mixture of forced aperture sizes for forced colour photometry (McCracken et al. 2012). For the purposes of DEVILS we need well converged total magnitudes and source apertures for likely galaxies. As discussed in Wright et al. (2016) and Andrews et al. (2017), the apertures returned by SExtractor often fail in crowded regions in a manner that creates spurious apertures that loop around nearby bright sources. When visually inspecting the SExtractor total photometry catalogues made available for the UltraVISTA regions it became clear a large number of bright sources (above our Y 21.2 mag limit) have unusually large apertures and erroneously bright flux measurements. Rather than manually intervene and run the fixed apertures through LAMBDA (mimicking the work flow discussed in Wright et al. 2016) we ran PROFOUND with close to default settings, creating a useable extra-galactic source catalogue in the process (Davies et al. in prep will discuss the construction of the DEVILS input catalogues in detail).

Figure 15 shows a sub region of the UltraVISTA survey. The left panel shows an all-band (Y, J, H, Ks) stacked image overlaid with the AUTO (Kron like) apertures from the full SExtractor catalogue that had been run on the UltraVISTA Y-band data. The right panel shows the PROFOUND pixel segments overlaid with the same apertures. It is clear that in isolated regions the AUTO apertures from SExtractor and the dilated flux converged segments from PROFOUND agree rather well. However, in some of the more crowded regions it is clear the segmentation and de-blending is producing quite different extractions, e.g. the top-right region near the green compass shows an example of multiple sources being grouped together in the SExtractor de-blend. Also, due to the mode PROFOUND was run in, the extraction depth is typically deeper with PROFOUND, i.e. there are effectively no SExtractor objects missed by PROFOUND, but there are some fainter (but visually real) sources extracted by PROFOUND missing in the public UltraVISTA SExtractor based catalogue.

Whilst there is much similarity between the SExtractor apertures and the PROFOUND segments, in general the PROFOUND segments appear to be larger. This seems to be particularly true for very bright stars (where the SExtractor aperture does not encompass all of the halo light in

general) and very faint objects. Figure 16 shows the difference quantitatively, where we compare the SExtractor AUTO aperture values for the semi-major and semi-minor axes against the ellipse implied sizes for the same quantities in PROFOUND. Whilst these apertures are produced by PROFOUND, they are not directly used in the photometry (only the pixel segments are used when computing photometric properties). Instead they are typically used for diagnostic tests and to aid star galaxy separation (which we describe later in this paper).

In general the estimated ellipses agree quite well, but it is clear there is a lot of correlation between the major and minor axes comparison, i.e. SExtractor and PROFOUND tend to agree on the shape of the ellipse, but there can be systematic differences in the total size of the objects. There is a small bias towards PROFOUND finding slightly larger apertures in general (the cross is above the 1–1 lines in both x and y), but the majority of aperture sizes agree within a factor of two.

Figure 17 shows how the extracted PROFOUND magnitudes compare to the Kron-like AUTO magnitudes returned by SExtractor. For brighter objects with unambiguous segmentation solutions the PROFOUND and SExtractor photometry agrees very well for all bands (well within 0.1 mag for the majority of sources). SExtractor does tend to have a number of brighter sources with close to double the flux of the PROFOUND source. These cases tend to be where PROFOUND has de-blended two very close by stars into two sources, whereas they are seen as being a single source in the published UltraVISTA catalogue. In these cases the de-blended solution in PROFOUND appears to be preferable.

Mimicking the analysis made in Wright et al. (2016) and Andrews et al. (2017), we use the narrowness of the colour distributions to judge the quality of the matched aperture photometry. The logic is that any colour aperture error will generate an additional random component of noise, and hence an increase in the spread of the intrinsic colour distributions. I.e. there are few scenarios where any additional random error could decrease the colour distribution intrinsic scatter (see Robotham & Obreschkow 2015, for a general discussion on intrinsic scatter and how best to model it). As discussed in Andrews et al. (2017), the 2nd aperture colours in UltraVISTA create very tight NIR colours, so these act as an excellent reference distribution for the matched segment PROFOUND colours. Matched segment colours can be computed in a number of ways in PROFOUND, but for this comparison we use the segmented pixels above a 2 σ sky-RMS threshold.

Figure 18 shows the comparison of PROFOUND and 2nd apertures taken from the published SExtractor derived UltraVISTA catalogue. Only two-way matching objects are shown for both distributions (i.e. the same ones as plotted in Figure 17). Since there are the same number of PROFOUND and UltraVISTA objects in each colour histogram, higher peaks will generally highlight tighter colour distributions. The overall result is that PROFOUND colours, run in matched segment mode, produce even tighter distributions than we find for the tightest SExtractor UltraVISTA colours. This is encouraging since the photometry of the main UltraVISTA survey was optimised for good quality colours for photo-z estimation.

A key issue for extra-galactic photometric surveys is

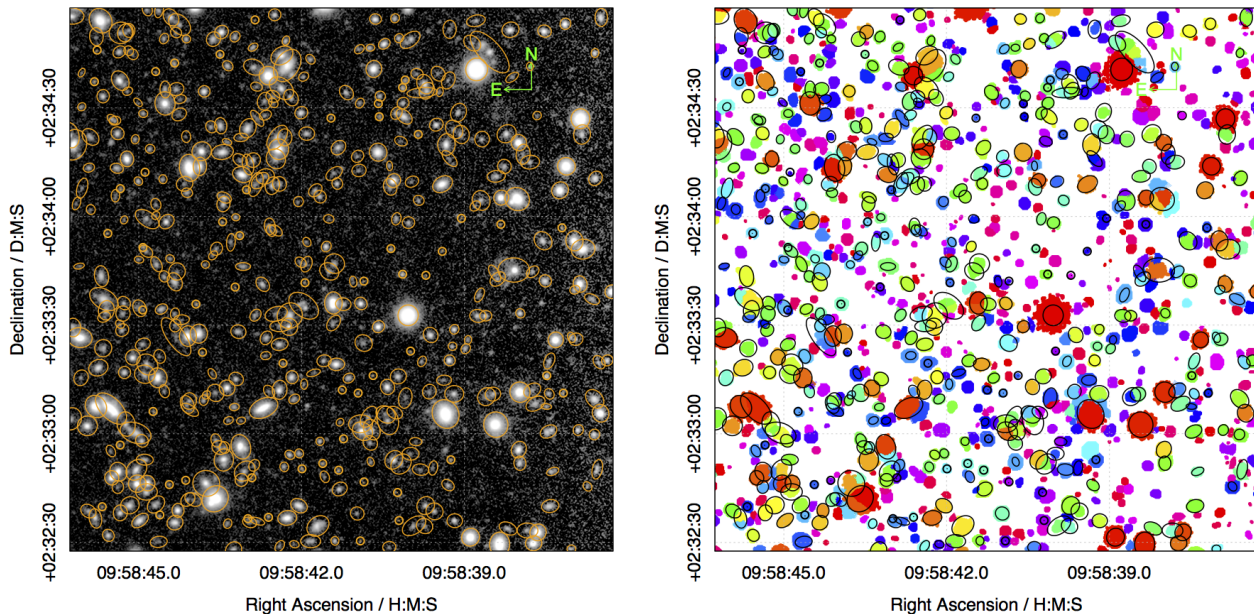


Figure 15. The left panel shows an example stacked region taken from UltraVISTA, where the publicly released SETRACTOR catalogue AUTO apertures are overlaid as orange ellipses. The right panel shows the corresponding PROFOUND segments in colour (redder means brighter sources and bluer means fainter sources), with the same SETRACTOR AUTO apertures overlaid in black. In isolated regions there is excellent qualitative agreement between the PROFOUND segments and the SETRACTOR AUTO apertures.

correctly identifying stars and galaxies. This is especially important when creating input catalogues for follow-up surveys (e.g. DEVILS) since we do not want to waste time observing stars or miss real galaxy targets. PROFOUND does not return a single parameter that flags stars and galaxies (like ‘Class-star’ in SETRACTOR), but using a combination of size, surface brightness, ellipticity and concentration outputs it is possible to accurately identify likely stars in optical and NIR photometry. In the case of UltraVISTA, the NIR colours add highly constraining discriminative information.

Figure 19 shows a particularly successful projection of surface brightness and NIR colour. The NIR (H-Ks) - (Y-J) multi-colour creates an almost perfectly horizontal separation between stars and galaxies, but using the PROFOUND surface brightness allows extremely bright stars with potentially erroneous colours (due to ghosting and saturation effects) do be easily selected. This selection is applied in a number of other geometric projections, where in each case the photometry has been extracted from the Y-band UltraVISTA data. It is clear that the NIR colours produce good separation agreement with the natural surface brightness versus size relationship, and also the surface brightness versus ellipticity relationship. These other two projections are not nearly as clear in isolation however, so in the case of UltraVISTA data the best possible star galaxy separation criteria must make use of accurate NIR colours. The details of the upcoming DEVILS survey star-galaxy separation using VISTA NIR colours and other geometric properties recovered with PROFOUND will be discussed in detail in Davies et al. in prep. In parallel to the processing of UltraVISTA data in order to create input catalogues for DEVILS, we are also using PROFOUND to process VIKING survey data that covers the upcoming Wide Area Vista Extragalactic

Survey (WAVES; Driver et al. 2016, circa 2021 start date)⁷. This future spectroscopic survey will cover ~ 1300 square degrees and is a major project for the upgraded 4-metre Multi-Object Spectroscopic Telescope (4MOST; de Jong 1996). The preparation of the NIR VIKING data using PROFOUND will be described in detail in an upcoming paper (Koushan et al., in prep).

6 CONCLUSIONS

In this paper we have presented the new PROFOUND source finding and extraction package. Its primary design goal is to provide all of the various inputs needed by the PROFIT galaxy modelling package, but it also serves as an effective stand-alone blind source finder. In these concluding remarks we will focus on the most novel aspects of the software compared to other open source astronomy alternatives.

- It uses a saddle point based source segmentation strategy that minimises the occurrence of source islands being created. A common artefact that essentially cannot be formed in PROFOUND is an aperture for an extended source looping around a bright star. This has removed the necessity of manual aperture fixing for extended sources (see Wright et al. 2016; Andrews et al. 2017).
- It defines photometric properties using dilated apertures rather than ellipses. This maintains the major geometric features of the sources, whilst also guaranteeing good flux convergence (by default, though other properties can be used to define the convergence criterion).

⁷ <https://wavesurvey.org>

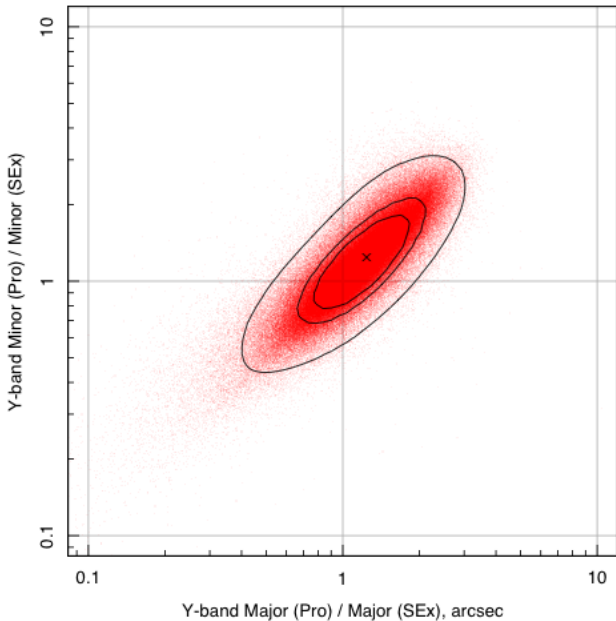


Figure 16. Comparison between the publicly released SEXTRACTOR AUTO aperture major/minor axes and the PROFIT extracted major/minor axes. PROFIT does not directly use the ellipses it estimates for any photometry, but they can be used by other code, e.g. LAMBDA. The three black contours show the high-density region containing 50% / 68% and 90% of the data, and the cross shows the median point.

- It creates all of the basic parameter inputs for PROFIT, in particular the effective radius is computed for all sources and the ellipticity and geometric rotation are defined in the same sense as for PROFIT (removing simple conversion mistakes that are otherwise common).
- It can operate in a number of modes for matched aperture (colour) photometry: e.g. simple per pixel segment matching, or with the capacity to let sources naturally dilate in each target band.
- It returns a comprehensive set of meta data by default, in particular the segmentation, sky and sky RMS maps. It can also compute the grouped complexes of touching sources, making it easy to fit multiple potentially overlapping sources with PROFIT.
- It offers a number of tools to aid the analysis of the source extraction and further prepare for running PROFIT, in particular it can create and sensibly track groups of touching/overlapping sources, and can extract surface brightness ellipses belonging to certain segments.
- It is fully open source and LGPL-3 compliant and written using a mixture of available R packages (available on the Comprehensive R Archive Network; CRAN) and R language, removing the need to compile the software for the majority of users.
- As well as the main package being hosted and maintained on GitHub (github.com/asgr/ProFound) it offers a large number of long form tutorial vignettes to aid inexperienced users tackling complicated extraction and fitting problems (<http://rpubs.com/asgr/>).

As with any piece of software, PROFIT has some

limitations that we will be keen to address in the longer term. As mentioned previously, it is predominantly single-threaded, and does suffer memory limitation issues for very large images. The immediate aim is to improve the memory limitations, since this appears to present the biggest limitation in typical use cases (e.g. processing large surveys). Also, in comparison to other photometry packages, its approach to assign flux to blended sources is fairly rudimentary. However, it is now possible to provide PROFIT segmentation maps as inputs to LAMBDA, which provides access to more sophisticated flux de-blending schemes. A final comment is that by design the saddle-point segmentation approach, whilst avoiding the possibility of some classes of serious de-blend errors, lacks the flexibility to create segments deeply embedded within segments. It is possible to combine different segmentation maps to get round this limitation, but it does mean running the segmentation function more than once for highly extended sources if the initial map poses a problem for further profiling with PROFIT⁸. That said, there is often no easy way for even humans to distinguish between very compact clumpy substructure that belongs to a galaxy and foreground stars that do not. Even colour information (unless very extensive) does not always remove this inherent ambiguity.

In summary, PROFIT offers a number of novel approaches to tackling some common issues with image source detection and extraction. It is particularly well suited for preparing inputs for further processing with PROFIT and similar galaxy profiling software (e.g. Barsanti et al. in prep, Casura et al. in prep, Cook et al. in prep, Hashemizadeh et al. in prep). A number of ongoing projects are already using PROFIT for generic survey source extraction (Koushan et al. in prep, Davies et al. in prep). We are also actively investigating how best to use PROFIT for the low surface brightness extraction of extremely faint and extended sources in KiDS data (Turner et al. in prep).

In the future there is a clear opportunity to create a more tightly coupled user experience for the automated fitting of galaxy profiles (in the general spirit of Barden et al. (2012) and Kelvin et al. (2012)). However, there is clear utility in using PROFIT simply as a source detection and extraction tool, so this explicit coupling has not been enforced at this stage. Prototypes of fully integrated packages that are inspired by some of the ideas and issues discussed in this work already exist, and will likely be released along with the corresponding scientific analysis over the next few years (e.g. Barsanti et al. in prep, Casura et al. in prep, Cook et al. in prep, Hashemizadeh et al. in prep).

ACKNOWLEDGEMENTS

We would like to thank the referee, Emmanuel Bertin, for his helpful suggestions on how to present aspects of this work. Much of the work presented here was made possible by the free and open R software environment (R Development Core Team 2016). All figures in this paper were made using the RMAGICAXIS package (Robotham 2016). This publication has made use of data from the VIKING survey from VISTA at

⁸ see Segmentation Map vignette at <http://rpubs.com/asgr/>

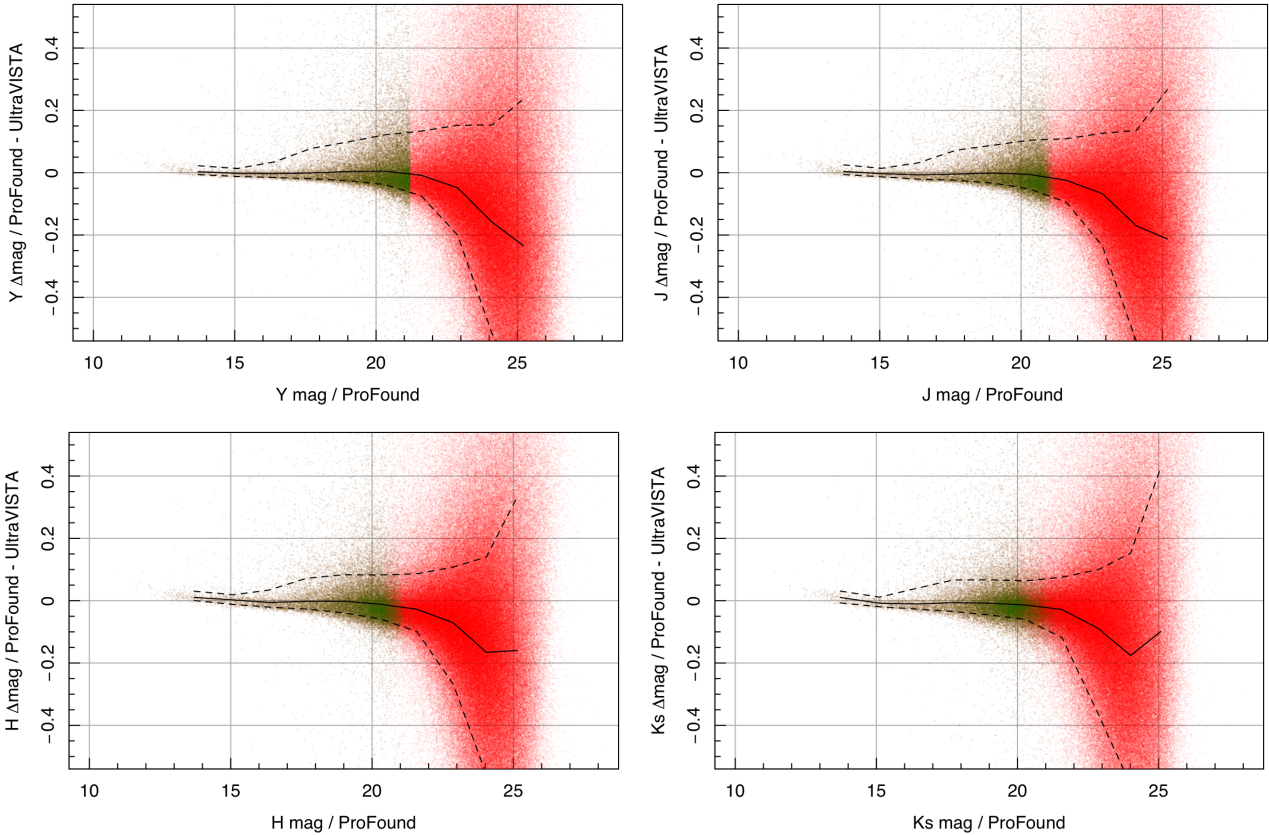


Figure 17. Comparison of PROFOUND diluted segment magnitudes versus SETRACTOR AUTO magnitudes for Y/J/H/Ks bands for 1⁴ matched sources. The solid black line shows the running median of the data, and the dashed lines show the range of 1 σ scatter around the median. The limit of the DEVILS sample ($Y < 21.2$) is shown in dark green in each panel. For all bands we see that on average PROFOUND returns more flux (i.e. the magnitude difference drops below 0 in these plots). The difference is typically less than 0.2 mag until ~ 23 mag, after which the amount of object scatter increases significantly. The typically brighter source extraction is the reason the faint limit has some diagonal structure.

the ESO Paranal Observatory, programme ID 179.A-2004 (Edge et al. 2013). Data processing has been contributed by the VISTA Data Flow System at CASU, Cambridge and WFAU, Edinburgh. Based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under ESO programme ID 179.A-2005 and on data products produced by TERAPIX and the Cambridge Astronomy Survey Unit on behalf of the UltraVISTA consortium. (McCracken et al. 2012). Parts of this research were conducted by the Australian Research Council Centre of Excellence for All-sky Astrophysics (CAASTRO), through project number CE110001020.

REFERENCES

Ahn C. P., et al., 2014, ApJS, 211, 17
 Andrews S. K., Kelvin L. S., Driver S. P., Robotham A. S. G., 2014, PASA, 31, 4
 Andrews S. K., Driver S. P., Davies L. J. M., Kafle P. R., Robotham A. S. G., Wright A. H., 2017, MNRAS, 464, 1569
 Barden M., Häußler B., Peng C. Y., McIntosh D. H., Guo Y., 2012, MNRAS, 422, 449
 Beard S. M., McGillivray H., Thanisch P. F., 1990, MNRAS, 247, 311
 Bertin E., Arnouts S., 1996, A&AS, 117, 393

Bijaoui A., 1980, AA, 84, 81
 Connolly A. et al., 2010, The Large Synoptic Survey Telescope Image Simulation Group database
 Davies J. I., Davies L. J. M., Keenan O. C., 2016, MNRAS, 456, 1607
 de Jong R. S., et al., 2014, Proc. SPIE, 9147, 91470M
 Driver S. P., et al., 2011, MNRAS, 413, 971
 Driver S. P., Davies L. J., Meyer M., Power C., Robotham A. S. G., Baldry I. K., Liske J., Norberg P., 2016, ASSP, 42, 205
 Edge A., Sutherland W., Kuijken K., Driver S., McMahon R., Eales S., Emerson J. P., 2013, The Messenger, 154, 32
 Ferrarese L. et al., 2012, ApJS, 200, 4
 Jarvis J. F., Tyson J. A., 1981, AJ, 86, 476
 Graham A. W., Driver S. P., 2005, PASA, 22, 118
 Holwerda B. W., 2005, ArXiv Astrophysics e-prints, 12139
<https://arxiv.org/abs/astro-ph/0512139>
 Irwin M. J., 1985, MNRAS, 214, 575
 Jarvis M. J., 2013, MNRAS, 428, 1281
 Kelvin L. S., et al., 2012, MNRAS, 421, 1007
 Kron R. G., 1980, ApJS, 43, 305
 Kuijken K., et al., 2015, MNRAS, 454, 3500
 Liske J., et al., 2015, MNRAS, 452, 2087
 Martin D. C., et al., 2005, ApJ, 619, L1
 Masias M., Freixenet J., Lladó X., Peracaula M., 2012, MNRAS, 422, 1674
 Masias M., Peracaula M., Freixenet J., Lladó X., 2013, ExA, 36,

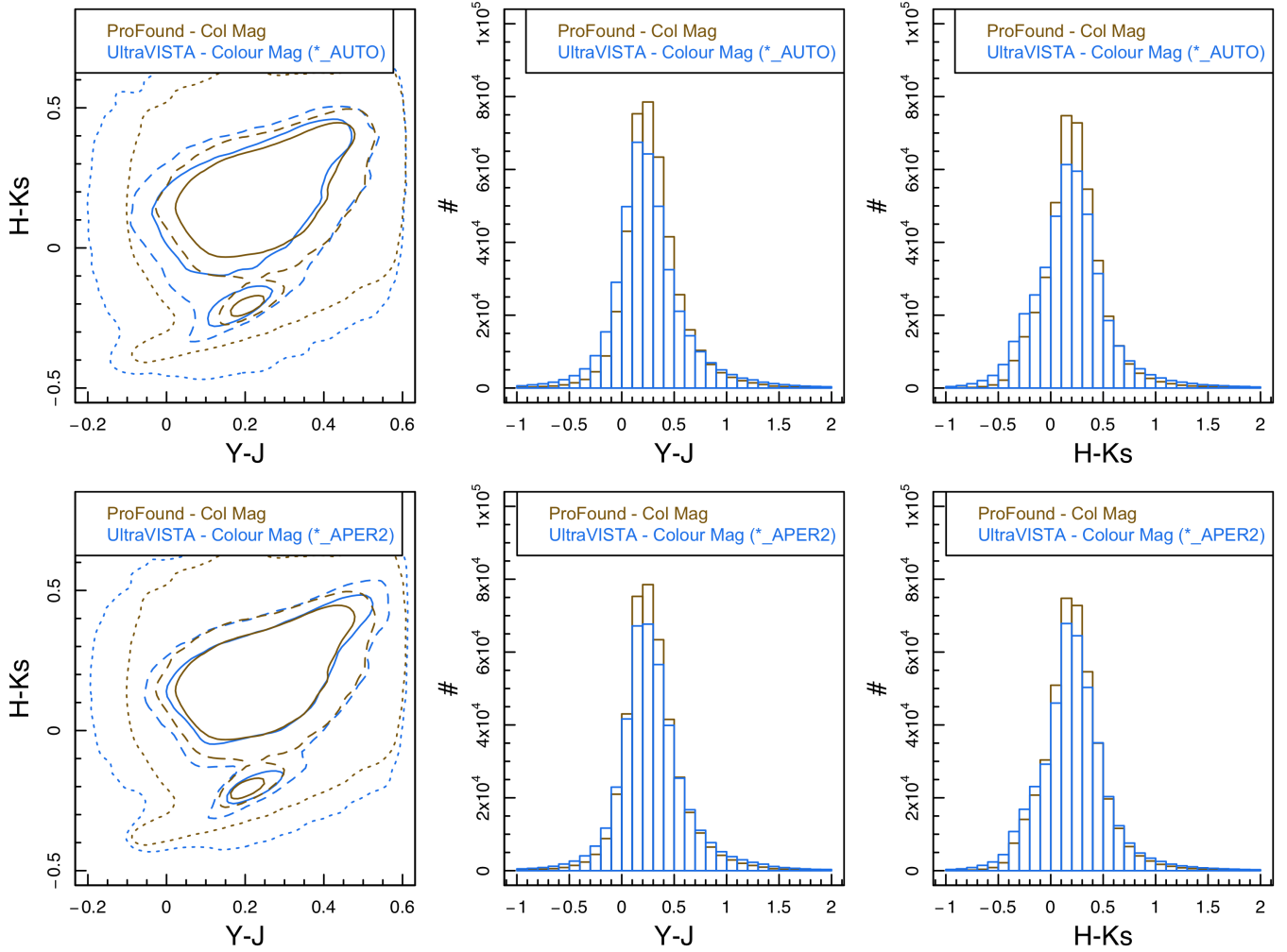


Figure 18. Comparison of the NIR colour distributions between the released SEXTRACTOR UltraVISTA photometric catalogues and those generated by PROFOUND. In these plots the sources have been explicitly matched against each other to ensure a fair comparison, though the same trends are evident if all detections at a comparable total magnitude cut are compared instead. The left panels show NIR colour-colour contour plots, where the contours contain the highest density 50%/68%/95% of the data. The other panels show the bluest (Y-J, middle column) and reddest (H-Ks, right column) colour histograms for the comparison region. In the top and bottom rows the PROFOUND photometry shown is the same, however the SEXTRACTOR UltraVISTA photometry is varied between the band optimised AUTO photometry (top row) and the fixed 2" circular aperture photometry (bottom row).

591

McCracken H., et al. 2012, *A&A*, 544, A156Meyer F., 1994, *Signal Processing*, 38, 113Gregoire P., Florian F., Oleg S., Michael B., Wolfgang H., 2010, *Bioinformatics*, 26, 979Petrosian V., 1976, *ApJ*, 209, 1R Development Core Team, 2016, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, ViennaRobotham A. S. G., & Obreschkow D., 2015, *PASA*, 32, 33Robotham A. S. G., 2016, *Astrophysics Source Code Library*, record ascl:604.004Robotham A. S. G., Taranu D. S., Tobar R., Moffett A., Driver S. P., 2017, *MNRAS*, 466, 1513Sabatini S., Davies J., Scaramella R., Smith R., Baes M., Linder S. M., Roberts S., Testa V., 2003, *MNRAS*, 341, 981Scoville N., et al., 2007, *ApJS*, 172, 1Sérsic J. L., 1963, *Bol. de la Asociacion Argentina Astron. La*

Plata Argentina, 6, 41

Van Vleck J. H. & Middleton D., 1946, *JAP*, 17, 940Wright A. H., 2016, *MNRAS*, 460, 765Zhang Y., McKay T. A., Bertin E., Jeltema T., Miller C. J., Rykoff E., Song J., 2015, *PASP*, 127, 1183Zheng C., Pulido J., Thorman P., Hamann B., 2015, *MNRAS*, 451, 4445

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.

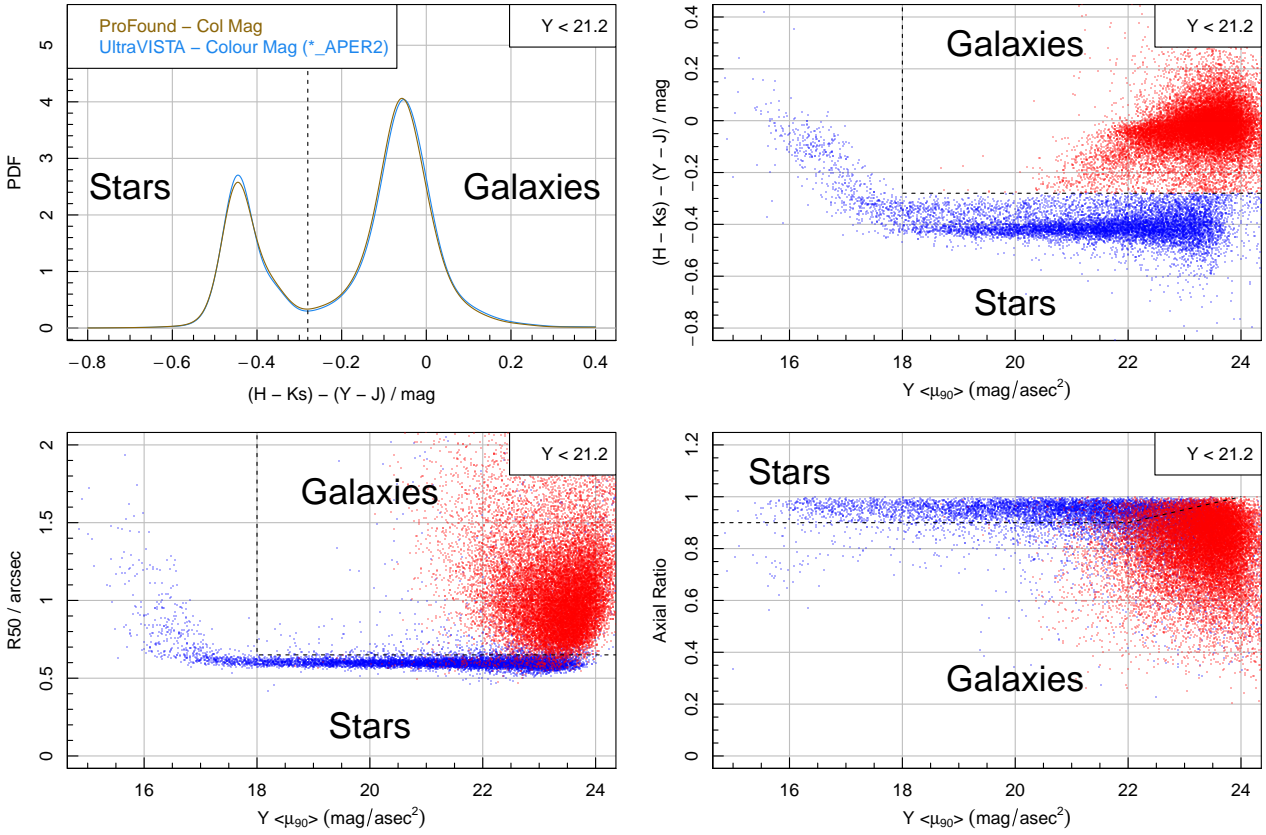


Figure 19. Demonstration of the star galaxy separation possible using outputs from PROFOUND. The top-left panel shows a simple multi-colour histogram ($[H-Ks]-[Y-J]$) where the use of the double colour term removes the main curvature seen in the left panels of Figure 18. This is evident in the top-right panel where the same PROFOUND multi-colour is shown on the y-axis against the mean surface brightness within the radius containing 90% of the flux (the separation shown by the dashed line is applied to the subsequent panels by colouring nominal stars blue and galaxies red). The bottom-left panel shows the same surface brightness against R_{50} , and the bottom-right panel has axial ratio on the y-axis. The multi-colour clearly acts as a very effective star-galaxy separator on its own, with the additional projections acting mostly as consistency checks rather than improving the fidelity of the separation. Sources which might have been incorrectly assigned in the top-right panel are those that have a colour differing from the dominant population inside the dashed regions (where the dashed regions are just indicative, and estimated by eye to approximately maximise the star galaxy discrimination).