Astronomy & Astrophysics manuscript no. galapagos
January 15, 2009

GALAPAGOS: From Pixels to Parameters

Marco Barden[1,2], Boris Häußler[2,3], Chien Y. Peng[4,5], Daniel H. McIntosh[6,7], and Yicheng Guo[7]

[1] **Current address:** Institute of Astro- and Particle Physics, University of Innsbruck, Technikerstrae 25, A-6020 Innsbruck, Austria
e-mail: marco.barden@uibk.ac.at
[2] Max-Planck-Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany
[3] **Current address:** Schools of Physics & Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK
[4] **Current address:** NRC Herzberg Institute of Astrophysics, 5071 West Saanich Road, Victoria, British Columbia V9E 2E7, Canada
[5] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
[6] **Current address:** Department of Physics, University of Missouri-Kansas City, 5100 Rockhill Road, Kansas City, MO 64110, USA
[7] Department of Astronomy, University of Massachusetts, 710 North Pleasant Street, Amherst, MA 01003, USA

submitted to A&A 16. Jan. 2009

ABSTRACT

Context. To automate source detection, two-dimensional light-profile Sérsic modelling and catalogue compilation in large survey applications, we introduce a new code GALAPAGOS, Galaxy Analysis over Large Areas: Parameter Assessment by GALFITting Objects from SExtractor.
Aims. Based on a single setup, GALAPAGOS can process a complete set of survey images. It detects sources in the data, estimates a local sky background, cuts postage stamp images for all sources, prepares object masks, performs Sérsic fitting including neighbours and compiles all objects in a final output catalogue.
Methods. For the initial source detection GALAPAGOS applies SEXTRACTOR, while GALFIT is incorporated for modelling Sérsic profiles. It measures the background sky involved in the Sérsic fitting by means of a flux growth curve. GALAPAGOS determines postage stamp sizes based on SEXTRACTOR shape parameters. In order to obtain precise model parameters GALAPAGOS incorporates a complex sorting mechanism and makes use of modern CPU's multiplexing capabilities. It combines SEXTRACTOR and GALFIT data in a single output table. When incorporating information from overlapping tiles, GALAPAGOS automatically removes multiple entries from identical sources. GALAPAGOS is programmed in the Interactive Data Language, IDL.
Results. We test the stability and the ability to properly recover structural parameters extensively with artificial image simulations. Moreover, we apply GALAPAGOS successfully to the STAGES data set. For one-orbit *HST* data, a single 2.2 GHz CPU processes about 1000 primary sources per 24 hours. Note that GALAPAGOS results depend critically on the user-defined parameter setup. This paper provides useful guidelines to help the user make sensible choices.

Key words. Methods: data analysis – Surveys – Galaxies: structure – Galaxies: statistics

1. Introduction

Imaging surveys provide a general tool to access the average properties of galaxy populations. A survey data set usually consists of an arrangement of primary images in one or several (optical) filters. These data are often accompanied by supplementary data in other wavelengths from x-ray to radio, including both imaging and/or spectroscopy data taken from spacecraft or ground-based telescopes. Examples for such surveys are the ground-based COMBO-17 (Wolf et al. 2004) or CFHT legacy surveys, or the Hubble Space Telescope (*HST*) based DEEP1/DEEP2 Groth Strip surveys (Vogt et al. 2005), the GOODS (Giavalisco et al. 2004) or the COSMOS surveys (Scoville et al. 2007). The smallest unit in this category are single exposure deep pencil beam surveys like e.g. the Hubble Ultra Deep Field (Beckwith et al. 2006).

Such surveys resulted in numerous scientific publications including a wide variety of topics. These include studies of star formation rates in galaxies (e.g. Bell et al. 2005; Shioya et al. 2008), the galaxy luminosity function (e.g. Faber et al. 2007; Liu et al. 2008), rates of galaxy mergers (e.g. Bell et al. 2006; Lotz et al. 2008) and the frequency of bars (e.g. Jogee et al. 2004; Sheth et al. 2008). Structural properties of galaxies (e.g. Barden et al. 2005; McIntosh et al. 2005), their morphologies (e.g. Bundy et al. 2005) and dependencies of these with environment (e.g. Balogh et al. 2004; Nuijten et al. 2005) were investigated. Attention was also focused on the hosts of QSOs (e.g. Sánchez et al. 2004) and other exotic types of objects (e.g. Moustakas et al. 2004; Smolčić et al. 2007). Finally, large surveys are ideal tools for galaxy lensing applications (e.g. Parker et al. 2007; Heymans et al. 2008) and the identification of supernovae (e.g. Astier et al. 2006).

Common to all imaging surveys are the specific reduction methods involved in the data analysis. After reducing the imaging data, which normally consists of a mosaic of many potentially (partly) overlapping tiles, scientific sources are detected on the individual tiles and compiled in a source catalogue. The source catalogue is then cleaned from duplicate source entries resulting from overlaps in the tiling scheme of the survey. Depending on the scientific goals, more sophisticated methods are then applied to analyse the morphology of the sources, i.e. quantify the structure of their light-profiles. Finally, the resulting additional structural parameters are added to the source catalogue.

For the main task, source detection and extraction, the code SEXTRACTOR by Bertin & Arnouts (1996) has become the standard tool in astronomy. Based on a simple setup script SEXTRACTOR detects sources, estimates a background sky level, measures primary shape information, like position, position an-

gle and axis ratio, and even performs aperture photometry. A key feature is the ability to properly deblend close companion sources, while at the same time avoid breaking single large sources up into several pieces. Other features include a neural network to separate stars and galaxies or the option to associate the detected objects with a given list of input positions. SExtractor is designed with minimum user interaction, support for large images and high execution speeds in mind.

In order to analyse galaxy light profiles quantitatively, many codes have been developed. The ones that are most widely used employ a two-dimensional fitting method to model ellipsoidal radial profiles, and include convolution with a point spread function (PSF).

One of these codes is Gim2d, which was first employed as part of an Iraf pipeline to analyse survey imaging data (Simard et al. 2002). Based on a Metropolis algorithm to find the minimum in χ-space, Gim2d mainly uses the Sérsic profile (Sersic 1968), which is a general expression that includes both the de Vaucouleurs and exponential forms (see Sec. 4.5 and eq. 2). The minimisation method performs a global parameter space search. As a result, Gim2d is robust, however it requires large amounts of CPU time compared to other codes (e.g. Häussler et al. 2007).

Another application for modelling light profiles is Budda (de Souza et al. 2004). Budda was initially developed to perform bulge/disc decomposition. However, it has recently been updated to include also bar and central point source modelling. Moreover, it now also features a double exponential profile for discs.

Finally, a rather versatile and effective method was presented by Peng et al. (2002): Galfit. Like the aforementioned programmes, it is a two-dimensional fitting code to extract structural components from galaxy images. It is designed to model galaxies in as flexible a manner as possible, by allowing the user to fit any number of components and functional forms. Galfit therefore allows for the possibility to not only fit simple situations, but also for bulge, disk, bar, halo, etc. This freedom has the major advantage that not only may the object of prime interest be fitted, but so may the neighbouring sources – at the same time, as some situation may demand. Various light profile models are built into the code, including the "Nuker" law (Lauer et al. 1995), the Sérsic profile (Sersic 1968), an exponential disc, Gaussian or Moffat functions and even a pure PSF for modelling stars. All model profiles, except for the PSF itself, are convolved by the PSF to simulate image smearing by the Earth's atmosphere and telescope optics.

Although a scientist has a multitude of options to choose from for fitting and detecting objects, analysing a complete survey to the end of obtaining a source catalogue with galaxy parameters, requires many intermediate steps. For example, duplicate sources from tile overlaps have to be differentiated; the detection and fitting codes have to be set up; a proper local background sky value has to be estimated; resulting source parameters have to be compiled in a catalogue. As these steps are fairly general we have built a code that simplifies all these steps and largely automates the entire process. Our code, Galapagos, performs all the required steps from a single setup and with minimal manual interaction provides a fitting catalogue. It runs SExtractor to detect sources and performs an automated Sérsic fit using Galfit. Amongst the various codes introduced above, we opted to use Galfit because it outperforms Gim2d both in speed and reliability (Häussler et al. 2007) and allows a much wider range of light-profile models than Budda. Upcoming versions will include additional features like automated multi-component fitting. The code is available freely for public download from our website at: `http://astro.uibk.ac.at/~barden/galapagos/index.html`.

The layout of the paper is as follows. We start by giving an overview of the structure of the code (Sec. 2) and explaining its setup and how to interact manually with it (Sec. 3). Then we elaborate on the methods involved in the individual components (Sec. 4). Next, we present some fitting results based on simulated data and provide details concerning the reliability of the code (Sec. 5). Subsequently, we give estimates on the performance of Galapagos (Sec. 6), followed by a summary (Sec. 7). In the course of the paper we assume a working knowledge of SExtractor and Galfit and refer the reader to the publications by Bertin & Arnouts (1996) and Peng et al. (2002).

2. Overview of Code Structure

To handle an observational setup as described in Sec. 1 and provide an object property catalogue, Galapagos is divided into four main blocks, each of which is executable independently from the others. This allows the flexibility of repeating or optimising certain segments of the analysis without re-running the entire pipeline. These blocks are:

1. Detect sources by running SExtractor (`B`)
2. Cut out postage stamps for all detected objects (`C`)
3. Estimate the sky background, prepare and run Galfit (`D`)
4. Compile a catalogue of all galaxies (`F`)

Note that letters in brackets correspond to the respective sections in the Galapagos setup file (see Sec. 3). We visualise this structure in Fig. 1. In the first block (`B`), SExtractor is run to detect sources on the individual survey images. Optionally, Galapagos features a high dynamic range (HDR) mode for source extraction (Sec. 4.1), which is ideally suited for wide area and/or space-based, e.g. HST, data. After a first pass, the user may refine this catalogue by identifying "bad" detections followed by re-running SExtractor. This may be required to fix sources which have been overly deblended, or to remove spurious source detections (for details see Sec. 4.6). Galapagos assembles the resulting source list for each tile automatically into a catalogue. Once all the tiles are analysed, Galapagos combines the individual output catalogues rejecting duplicate sources (see Sec. 4.1) and optionally bad detections like cosmic rays etc. (see Sec. 4.6).

For galaxy fitting, it is worthwhile to first extract each galaxy from the survey mosaic, thereby reducing the amount of time needed to ingest an image into Galfit, or in the event that one wishes to repeat an analysis. Therefore, in the second block (`C`), Galapagos estimates a size for each object based on its SExtractor parameters. With this information it computes the extent of a postage stamp. From the original survey images, Galapagos then creates such a cutout for every object. It performs the subsequent fitting with Galfit on these postage stamps (see Sec. 4.3). At this stage, Galapagos creates for every survey image a "sky-map" containing information about the nature of the pixel flux (either "no flux", "sky" or "source"). It uses this map later on to identify blank sky pixels (see Sec. 4.4).

The third block (`D`) performs the major fitting work. For every object in the source catalogue it prepares and runs Galfit (see Sec. 4.5). Accurate fitting analysis by Galfit requires some careful consideration, which includes identifying the proper sky background, identifying neighbours and providing initial parameter guesses to start the fitting.

Sky estimation in galaxy fitting may sometimes be tricky to obtain in crowded fields, in small images, or when an object has

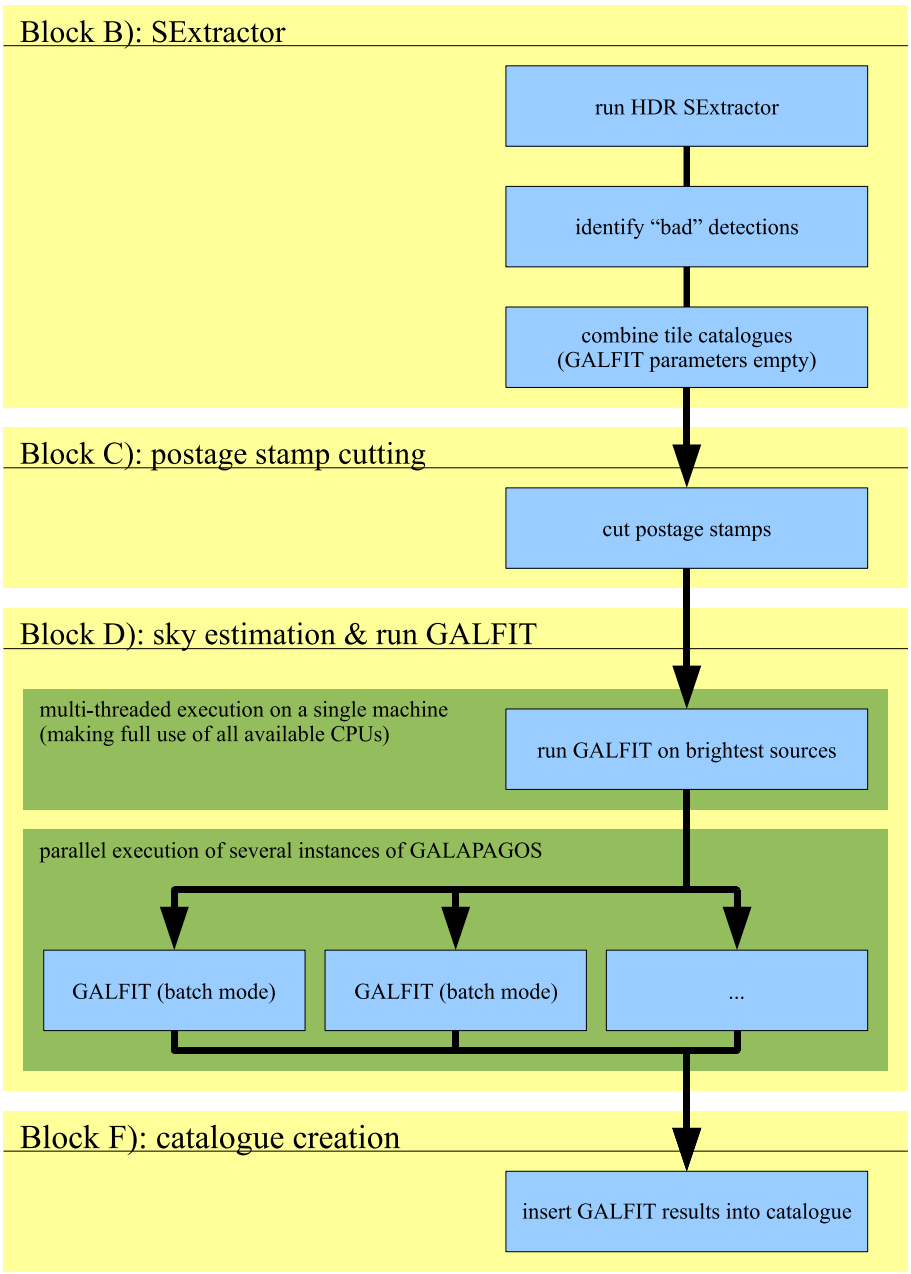

Fig. 1. Code structure. A yellow background indicates the four main blocks. Fitting objects with GALFIT is a two-stage process (see Sec. 4.5), which requires most of the total computation time (green background). We mark smaller tasks by blue boxes. For further details see Sec. 2.

a slowly tapering wing (e.g. elliptical galaxies). For instance, in crowded fields, the "sky background" may depend on whether neighbouring contamination is masked out or simultaneously fitted. Because of the interface with GALFIT, GALAPAGOS uses the full survey image and not the small postage stamp to compute the sky. It measures the sky using a flux growth curve including pixel rejection based on the "sky-map", which was calculated in the previous step (see Sec. 4.4). Note that even though GALFIT can fit the sky, GALAPAGOS does not use this option to avoid instances when neighbouring contamination makes accurate determination infeasible, and to reduce the degree of freedom in the fit.

To optimise the execution time, GALAPAGOS performs fitting in a rank-ordered sequence starting with the brightest source in the survey and progressing to the fainter ones. The advantage of this procedure is twofold:

a) Faint neighbours of bright sources do not have to be included in a simultaneous fit (as a second component), as they do not influence the resulting fit parameters of the bright object significantly. The magnitude difference between faint and bright neighbours is a free user parameter (`D16`, `D17`; for further details see Sec. 4.6).
b) When a faint source has a brighter neighbour, which has to be included in the fit as well, parameters for that object will already exist from a previous fit. Hence the variables for that component can safely be held fixed to the best values. This reduces the total number of degrees of freedom and increases the computation speed for a large number of sources tremendously (Sec. 4.5).

The weakness of the sequential approach is that it voids the speed benefits of parallel processing. To alleviate this problem, we devise two methods:
a) Consecutive sources in a rank-ordered list are usually suf-

ficiently far apart to not affect one another (the average object size is much smaller than their typical distance). Therefore, GALAPAGOS starts the next object in the sequence as a new process on another CPU (core), given that its distance from other sources in the queue is large enough.
b) If the objects are located on different tiles, generally it is possible to parallelise the analysis, unless they are sufficiently large or bright to contaminate neighbouring tiles. In latter cases, these objects are treated first in a sequential manner. When GALAPAGOS has analysed the last of these, it can then switch to a parallel analysis on different machines or CPU cores, working on one survey image at a time. A user-defined variable specifies after what fraction of brightest sources GALAPAGOS treats the tiles as decoupled, i.e. analysis of all sources affecting several tiles is assumed to be finished. At that point processing of individual tiles in parallel is available. Note that while in tile-by-tile parallel mode, GALAPAGOS still treats sources sequentially, i.e. on a single CPU (core), and execution time does not profit from point a) above.

Therefore, the total number of parallel threads for the two optimisations amounts to the total number of tiles at maximum in the second stage of fitting b) and the total number of available CPU (core)s in the first stage a) (for further details see Sec. 4.5). A requirement of the first stage a) is, that all CPUs must be able to see the whole dataset, i.e. they have to have access to the same harddiscs, because several threads are interacting with each other and working on the same data. In the second stage b), one might think that, as the tiles are decoupled, only a single tile is accessed at a time. However, in the presence of big, bright sources that affect neighbouring tiles, this is not the case any more. Therefore, even when fitting individual tiles in parallel the whole data set must be accessible. As now only the information for the current tile is changed, the fitting may be distributed to different harddiscs, though (by creating identical copies).

In the last block `(F)`, GALAPAGOS reads the results of the fitting from the headers of the GALFIT output images and puts them into the source catalogue (see Sec. 4.2). Here, it removes a second set of "bad" detections from the catalogue. Namely those that were required in the fitting process to allow optimal results for neighbouring objects. Usually, these are bright artefacts in close proximity to relatively faint real sources (see Sec. 4.6). The resulting catalogue, GALAPAGOS compiles in a FITS-table.

3. Code Setup and Control

A set of scripts controls GALAPAGOS, some of which are optional. We show an example for the main startup script in Fig. 2. It contains all references for file locations and controls the programme execution. The startup script is divided into six parts `(A)` through `(F)`, closely related to the four blocks described in Sec. 2. The first set of parameters `(A)` defines the input and output file locations. Section `(E)` contains options that help setting up GALFIT. These two parameter sets are "static"; the remaining four are "dynamic" in the sense that these can be activated or skipped when executing GALAPAGOS. They correspond to the four programme blocks that were previously defined in Sec. 2. Parameter set `(B)` starts SEXTRACTOR; `(C)` is responsible for defining and cutting the postage stamps; `(D)` performs the estimation of the sky background, prepares GALFIT and starts the fitting; finally parameter set `(F)` reads out the fit results and creates the output catalogue. The difference between the "static" and "dynamic" parameter sets is, that the latter control code execution while the prior only define file locations and setup parameters. Note that the setup files for SEXTRACTOR `B03` (and optionally `B06` in HDR-mode) require additional files to be accessible. These are the neural network file "default.nnw" and an optional convolution filter, e.g. "tophat_3.0_3x3.conv". For details on how to setup SEXTRACTOR see Bertin & Arnouts (1996).

A file location list, or short "file list" `A00`, provides the information required for defining the organisation of the survey. Just like the startup script, this is an ASCII file containing information about the individual tiles. For each survey pointing it provides the file location of the actual tile, the corresponding weight image (which is an exposure time map), a path for storing the fitting output of the individual tiles, and a prefix, which is attached to all output files. We give an example of such a file list in Fig. 3 for a hypothetical survey with 10 science tiles.

In order to refine the output of SEXTRACTOR, the user has the option to remove sources. After an initial run with the optimal SEXTRACTOR setup, the user may create a list that contains the file name of the respective tile together with an x/y-pixel position, e.g.:

```
/path/to/survey/tile01.fits 234 567
/path/to/survey/tile01.fits 765 432
/path/to/survey/tile02.fits 453 678
...
```

GALAPAGOS rejects any detection within a certain radius `B16` automatically from the SEXTRACTOR catalogue on a subsequent run of the code. Thus, if one wants to refine the catalogue, the SEXTRACTOR section of the code has to be run twice, i.e. GALAPAGOS needs to be started first with only the SEXTRACTOR section activated and then run a second time with the SEXTRACTOR section and optionally others enabled as well. The first execution is required for identifying bad detections; the second run then treats them. For details on what sources should be removed and how the code deals with them see Sec. 4.6.

In Sec. 2 we briefly introduced the mechanisms to optimise the total programme execution time. According to this scheme, after cutting the postage stamps and fitting the brightest objects, GALAPAGOS may be run in parallel on several computers each working on a section of the survey. In order to specify the region that the respective pipeline should work on, the user must provide a list `E01` that contains the file names of the individual tiles in question. If one were to fit the survey from Fig. 3 in parallel on three CPUs, one could set up the first computer with a batch list containing tiles 1 to 3, the second one with 4 to 6 and the third one with 7 to 10. As an example, the batch list for the first CPU would look like this:

```
/path/to/survey/tile01.fits
/path/to/survey/tile02.fits
/path/to/survey/tile03.fits
```

To summarise, to process to run GALAPAGOS on a complete survey requires the following steps:

1. setup a startup script and file list (including SEXTRACTOR setup).
2. run the first block `(B)` (optionally in HDR mode).
3. optionally identify "bad" detections and manually create the respective "bad detection lists".
4. if "bad detection lists" were created re-run the first block `(B)`.
5. run the second block `(C)` to prepare and cut postage stamps.
6. run the third block `(D)` on the brightest galaxies.
7. create batch lists for parallel processing and for each batch list a new corresponding startup script.

```
#================================FILE LOCATIONS================================
A00) /path/to/survey/setup/gala_files #example below
A01) /path/to/survey/cat              #output directory for catalogues
#==============================SEXTRACTOR SETUP================================
B00) execute                          #execute the SExtractor block
B01) /path/to/SExtractor-binary/sex   #SExtractor executable including path
B02) /path/to/survey/setup/gala.param #output parameters in .param-format
B03) /path/to/survey/setup/coldsex    #SExtractor setup file (cold)
B04) coldcat                          #output catalogue (cold)
B05) coldseg.fits                     #output segmentation map (cold)
B06) /path/to/survey/setup/hotsex     #SExtractor setup file (hot)
B07) hotcat                           #output catalogue (hot)
B08) hotseg.fits                      #output segmentation map (hot)
B09) 1.1      #enlarge the cold isophotes for catalogue combination by a factor
B10) outcat                           #output combined catalogue
B11) outseg.fits                      #output combined segmentation map
B12) outparam                         #output parameter file
B13) check.fits                       #check image filename
B14) apertures                        #check image type
B15) /path/to/survey/setup/gala_exclude #list of "critical" detections
B16) 1.5                              #radius in pix used to exclude objects
B17) all #if set 'outonly': hot/cold catalogues/segmaps are deleted
B18) /path/to/survey/setup/gala_bad   #list of "catalogue" detections
B19) sexcomb                          #combined sextractor catalogue
#================================STAMP SETUP=================================
C00) execute #execute the Stamps creation block
C01) stamps  #descriptor file for postage stamps
C02) v       #preposition for postage stamps
C03) 2.5     #scale factor by which the sextractor isophotes are enlarged
#============================SKY PREPARATION SETUP============================
D00) execute #execute the sky preparation block
D01) skymap  #output object/sky-mapfile
D02) outsky  #output filename for sky values
D03) 3       #scale factor by which SEx isophote is enlarged (for skymap)
D04) 1.5     #scale factor by which SEx isophote is enlarged (for neighbours)
D05) 20      #additional offset to scale factor
D06) 30      #distance between individual sky isophotes
D07) 60      #width of individual sky isophotes
D08) 30      #gap between sextractor isophote and inner sky isophote
D09) 3       #cut below which objects are considered as contributing
D10) 2       #nobj_max; max number of allowed contributing sources
D11) 1.4     #power by which the flux_radius is raised to convert to Re
D12) 5       #fraction of sources to be treated first (in %; 5=5%)
D13) 15      #calculate the slope of the sky from the x last determinations
D14) -0.3    #slope in fwhm_image vs. mag_best below which object is star
D15) 6.8     #zeropoint in fwhm_image vs. mag_best below which object is star
D16) 5       #magnitude faint end limit for secondaries when fitting galaxies
D17) 2       #magnitude faint end limit for secondaries when fitting stars
D18) 8       #number of neighbouring frames
D19) 4       #maximum number of parallel processes
D20) 360     #minimum distance (in arcseconds) for D19)
#================================GALFIT SETUP================================
E00) /path/to/galfit-binary/galfit  #Galfit executable including path
E01) /path/to/survey/setup/batchlist.XX #tile batch list
E02) obj                            #object file preposition
E03) v_gf                           #preposition for GALFIT output files
E04) /path/to/survey/setup/psf.fits #PSF filename including path
E05) mask                           #mask file preposition
E06) constr                         #constraint file preposition
E07) 257                            #convolution box size
E08) 26.486                         #zeropoint
E09) 0.03                           #plate scale of the images [arcsec/pixel]
E10) 706                            #exposure time
E11) 750                            #constraint max Re
E12) -5                             #constraint min magnitude deviation (minus)
E13) 5                              #constraint max magnitude deviation (plus)
E14) notnice                        #nice?
E15) 2.1c                           #GALFIT version string. E.g. 2.0.3c
#============================OUTPUT CATALOGUE SETUP============================
F00) execute      #execute catalogue combination block
F01) combcat.fits #filename for output catalogue in A01)
```

Fig. 2. Example of a startup script for GALAPAGOS.

```
/path/to/survey/tile01.fits /path/to/survey/wht01.fits /path/to/survey/t01 t01.
/path/to/survey/tile02.fits /path/to/survey/wht02.fits /path/to/survey/t02 t02.
/path/to/survey/tile03.fits /path/to/survey/wht03.fits /path/to/survey/t03 t03.
...
/path/to/survey/tile10.fits /path/to/survey/wht10.fits /path/to/survey/t10 t10.
```

Fig. 3. Example of a file location list for GALAPAGOS. The middle section is left out. In this example a survey with 10 tiles is defined. The four columns represent science image, corresponding weight image (exposure time map), output directory and output file preposition, respectively.

8. re-run block (D) in parallel on several machines.
9. when parallel processing is finished on all computers, run block (F).

Note that if the survey is small enough, steps 6 to 8 may be combined, either by setting the brightest galaxies fraction D12 to 100%, thus taking full advantage of the available CPUs on the machine or by simply providing only one batch file containing all tiles. The latter option does not provide advantages over the first one and is best used only for testing purposes.

4. Components

In the following sections we describe in detail the methods involved in the individual components of GALAPAGOS. These include SEXTRACTOR and high dynamic range (HDR) source extraction (Sec. 4.1), compiling a combined source catalogue (Sec. 4.2), the cutting of postage stamps (Sec. 4.3), estimating a background sky level robustly (Sec. 4.4), and fitting with GALFIT (Sec. 4.5). In the last part of this section we introduce some technical mechanisms to optimise the code for robustness and speed (Sec. 4.6).

4.1. SExtractor

Source identification in SEXTRACTOR is performed by detection of flux peaks above the sky background. SEXTRACTOR measures the sky background either globally or locally by interpolation over a grid of points. If a user-defined total number of adjoining pixels exceeds a specified flux limit above the sky background it meets the definition of being a legitimate source and is thus entered into the object catalogue. The flux surrounding this source SEXTRACTOR then attributes to the object or a neighbour by a tree algorithm. If the total flux in two different branches above a certain surface brightness level differs significantly it considers both branches to belong to the same source – if they are similar, it breaks up the source into two objects. The user has various options to influence this deblending including the filtering of spurious detections. Details on how to operate SEXTRACTOR can be found in Bertin & Arnouts (1996).

As a by-product of this tree-algorithm SEXTRACTOR can provide a segmentation map as output. This is a copy of the input image where individual pixel values reference the object number they were accounted to. Blank sky pixels get the value 0.

To judge the extent of a galaxy SEXTRACTOR provides the ideal tool. One of the output parameters is the "Kron radius". Kron (1980) and Infante (1987) found that for both stars and galaxies, when convolved with a Gaussian seeing, 90% of their flux lies within a radius $r_K = k \cdot r_1$, with $k = 2$ and

$$r_1 = \frac{\sum r\, I(r)}{\sum I(r)}, \tag{1}$$

where $I(r)$ is the source intensity at pixel location r from the centre. The quantity $K = r_K \times a$, with the semi-major half axis a, together with the axis ratio q describes an ellipse around a source encircling most of its flux. When using the term "Kron radius", in the following discussion we mean K. The internal SEXTRACTOR routines also use the Kron radius to estimate the total source flux.

SEXTRACTOR has been used successfully with both ground- and space-based data. Yet, recent large CCD arrays put the code to its limits due to the wide range of object sizes and luminosities that are being observed simultaneously. In classic pencil beam surveys, the objects of interest are mostly faint and small. SEXTRACTOR is then fine-tuned to pick up such sources properly at the cost of splitting up the occasional big bright spiral galaxy into many pieces. On the other hand, wide area surveys traditionally do not reach very deep. When fine-tuning SEXTRACTOR for these surveys, emphasis is put on correctly deblending the larger objects while losing some depth. In both applications one reaches the dynamic range limit and has to make a compromise of depth and proper deblending. Unfortunately, modern space surveys with the *HST* in particular are very demanding on dynamic range. They cover large areas and the science involves observation of galaxies in the near and distant universe including the biggest and brightest as well as the faintest and smallest sources.

Fortunately, there is a rather simple two-step approach using SEXTRACTOR to overcome this problem. Firstly, one runs SEXTRACTOR in a so-called "cold" mode in which only the brightest sources are picked up and properly deblended. As this will miss many faint sources, in a second setup emphasis is put on depth. The second run we term "hot" mode. Then one combines both output catalogues in such a way as to include all cold sources, but only hot sources at positions where no cold source was detected. We term this mode "High Dynamic Range (HDR) SExtraction".

To illustrate the process of including hot sources outside the Kron radius of cold sources into a combined catalogue, see Fig.4. In the upper left hand panel we show a "cold" run. The big central spiral galaxy is deblended correctly with the fainter galaxy below it. Also, the clumpy low surface brightness spiral in the upper left corner is detected as a single source. All three sources are taken over into the combined catalogue. Requiring a proper deblending of the bright objects results in missing the faintest sources, though. The "hot" run (upper right hand panel) picks those up. However, it brakes the brighter galaxies up into many sources. In the example, an off-centre knot of the upper left galaxy was detected as a separate object. Moreover, the outer regions of the central (and upper left) galaxy are assigned separate source IDs. These "spurious" detections change the effective size of the central galaxy (compare diameters of the Kron ellipses in the left and right figures). Interestingly, the relatively bright galaxy below the central object is not deblended properly in the hot run. Furthermore, the size and position angle of the upper left detection demonstrates the lower detection threshold of the hot setup. In the hot setup a larger fraction of the low surface brightness flux is included in the calculation of the position angle, thus providing a much better estimate than the cold setup, which is more heavily weighted towards the inner regions of the

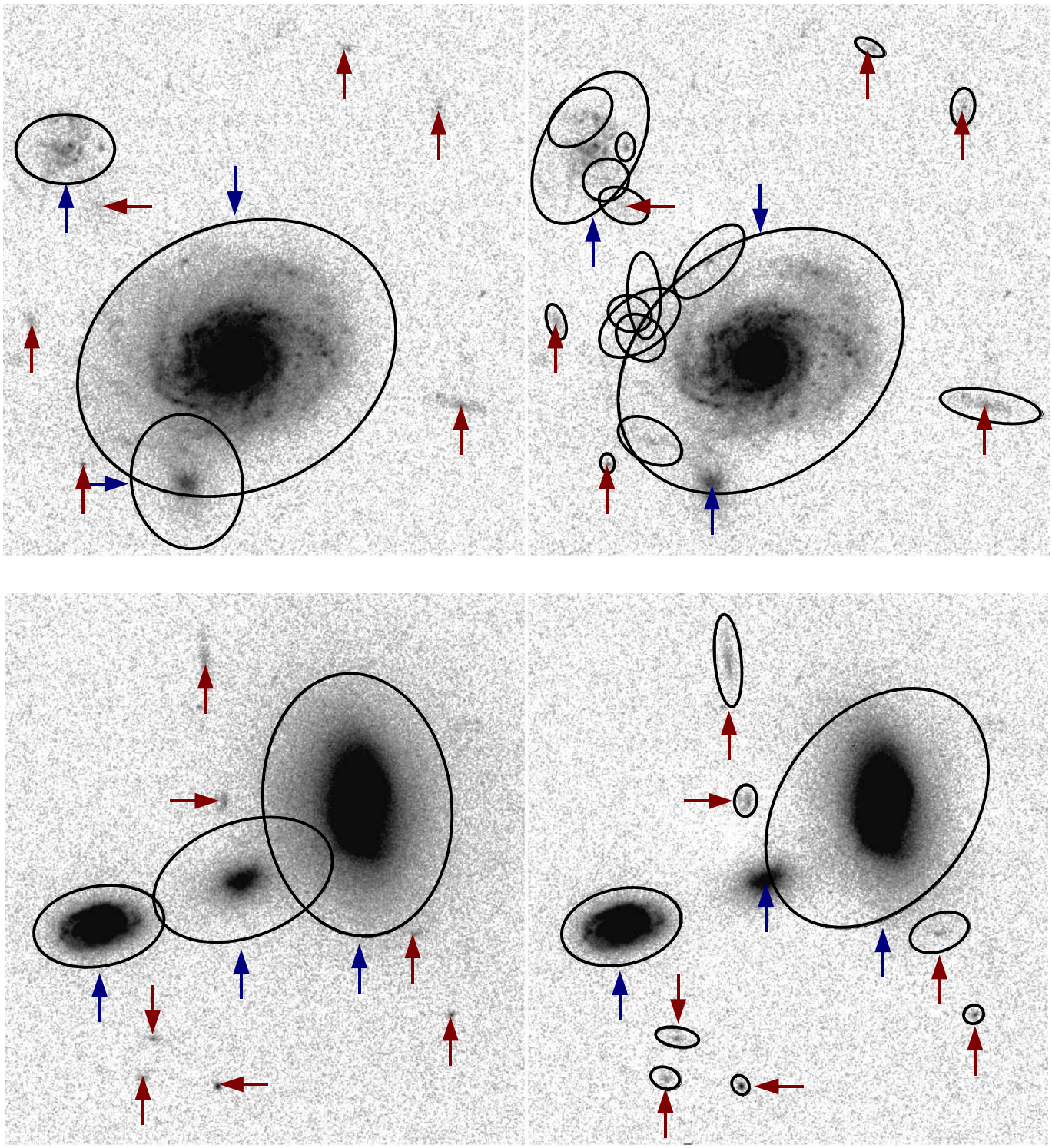

Fig. 4. Combining "hot" and "cold" SExtractor catalogues. The two panels (*upper* and *lower*) show examples of a cold (left side) and a hot (right side) SExtraction. Ellipses indicate the SExtractor Kron ellipses of the detected sources. Arrows mark objects from the hot (red) and the cold (blue) catalogue that were incorporated into the combined catalogue. Additional hot sources not marked with arrows (e.g. in the upper right panel) were excluded from the combined catalogue as described in detail in Sec. 4.1.

sources. Yet, the values from the cold run enter the combined catalogue as deblending is the more important source of error. Also, Galfit calculates structural parameters like the position angle much more reliably. The lower panels in Fig. 4 show another example. Again, the deblending in the hot run is bad, while in the cold run it is correct. The faintest sources are only detected in the hot run. Bad deblending in the hot run strongly affects the calculation of the position angle of the brightest source, while in the cold run it is acceptable.

We developed and tested this method for the GEMS survey (Rix et al. 2004; Caldwell et al. 2008; for tests see Häussler et al. 2007). Subsequently, other major surveys have adopted it as well, including COSMOS (Koekemoer et al. 2007; Leauthaud et al. 2007) and STAGES (Gray et al. 2008). Galapagos provides

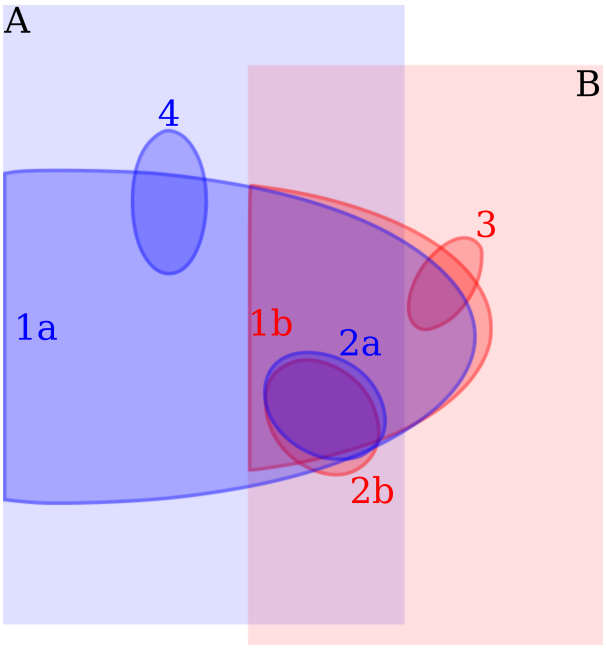

Fig. 5. Combining SExtractor catalogues from neighbouring tiles. Tile A contains sources 1a, 2a and 4, while objects 1b, 2b and 3 were detected on tile B. Ellipses show the corresponding sizes from SExtractor. For a description of what source ends up in the resulting table see Sec. 4.2.

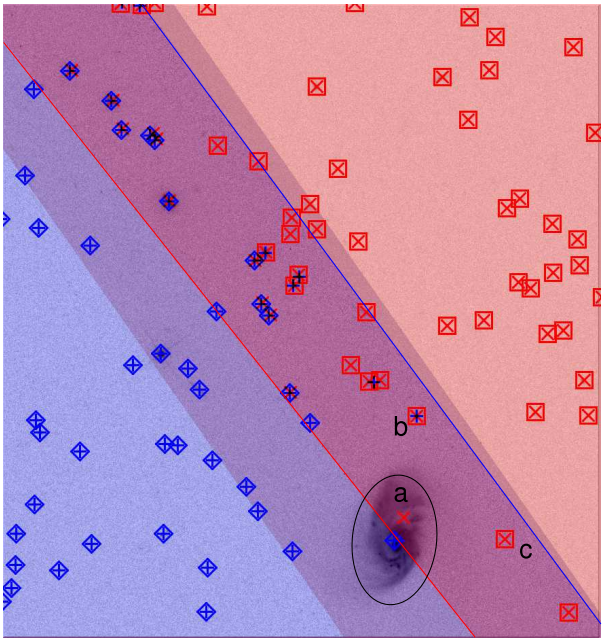

Fig. 6. Combining SExtractor catalogues from neighbouring tiles. The left image (blue area) extends out to the right (blue) diagonal line; the right image (red area) extends out to the left (red) line. Shaded areas outside of the lines corresponding to the respective image did not receive sky flux. Pluses (blue) indicate source detections (hot and cold already combined) from the left (blue) image; crosses (red) mark detections from the right (red) image. Diamonds (blue) highlight objects that are contained in the combined catalogue if they originated from the left (blue) image; boxes (red) highlight those that were taken from the right (red) image. Catalogue combination is based on the SExtractor ellipses (see Sec. 4.2). Such an ellipse is shown in case a). The source from the right image (red) is rejected as it lies inside an ellipse of a detection in the left image (blue), which is further away from the respective image boundary (blue and red lines). For the same reason in case b) the red source is kept. The detection in case c) does not even have a counterpart in the other catalogue.

the option for running SExtractor in two-stage HDR or normal single-stage SExtractor configuration.

4.2. Catalogue Compilation

Compiling the output source catalogue is a two-stage process. Galapagos creates a first combined catalogue from the SExtractor output tables. In the subsequent model fitting process, Galapagos fills this catalogue with the Galfit output parameters.

When putting together catalogues from potentially overlapping images, Galapagos has to take care of removing detections of the same source on multiple images. To this end, it uses the world coordinate system of the images to translate pixel coordinates from one to another image. Next, Galapagos calculates the distance to the image border for each source in both catalogues. The area containing flux (pixels with non-zero values) defines the image border. This is crucial in particular for non-rectangular images (e.g. from *HST*). Now, Galapagos sorts the two catalogues by border-distance. It starts with the source farthest from the edge, which we assume to be on image *A* (source 1a; see Fig. 5). Then it checks whether there are sources inside the Kron ellipse of the current object in the neighbouring image *B* (sources 1b, 2b and 3). If it finds any such targets, Galapagos removes them from the list. Note that Galapagos does not remove objects overlapping with the source from image *A* from the list (sources 2a and 4). Following this scheme it works through the complete list, from the farthest to the closest objects to the boundary, and constantly updates the list in the process. A problem arises for sources, say in image *A* (source 1a), extending over a radius larger than the size of the overlap area and having overlapping detections on image *B*, which are not covered by image *A* (source 3 in Fig. 5). In such a case, Galapagos includes the main source from image *A* (source 1a) in the catalogue and all overlapping sources from image *A* (sources 2a and 4). Overlapping sources from image *B* it removes, though (sources 2b and 3). Thus, unfortunately this method removes source 3 from the catalogue. Although this problem cannot be avoided, in practice it rarely occurs. It can be avoided completely if the largest source in the survey is smaller than the overlap between survey images.

Fig. 6 shows an example for this procedure to remove duplicate detections. The bright galaxy a) is just on the edge of the red image. As only half its flux is visible on that image, the calculated centre is far off from the real position. The blue image fully contains this galaxy. The two central positions (in red and blue) being so different, a normal nearest neighbour matching algorithm with a maximum matching radius would not have been able to identify the two detections as the same source. In the proposed scheme, though, the red detection is not put into the combined catalogue, for being inside the Kron ellipse of a source that is further from the image edge in the blue image. Similarly, the red source b) is further from the image edge than the blue source and thus, we reject the blue object. Objects without counterpart in the other image, as in c), we do keep in the combined catalogue.

As Galapagos performs duplicate removal before running Galfit, it does not fit sources twice. Note that for fitting sources at the edge of an image, Galapagos takes objects on neighbouring survey images into account as well (see Sec. 4.6).

4.3. Postage Stamps

To optimise galaxy fitting with GALFIT, GALAPAGOS cuts the science images into smaller sections centred on individual sources. The advantage of using such postage stamps is that the total fitting time can be reduced significantly. Even rather deep optical surveys contain large fractions of empty sky, which can mostly be excluded from the fit once the information from the sky pixels is effectively used to estimate the background (see Sec 4.4). In a typical one-orbit *HST* survey around a factor of 2 in the total number of pixels can be saved. Moreover, although GALFIT allows simultaneous fitting of multiple sources, modelling more than a handful of objects at the same time quickly becomes rather impractical. Thus, to optimise automated fitting of large numbers of sources, GALAPAGOS incorporates a postage stamp cutting facility.

We note that a disadvantage of using postage stamps for fitting *with the background sky as a free fit parameter* is a potential risk of biased fit results if the postage stamp does not contain enough empty sky pixels. In such a case the χ^2 of the fit might indicate a good fit, yet the result would be flawed by attributing too much or too little flux to the object. This could potentially also have a strong impact on other structural parameters. Therefore, GALAPAGOS does *not* allow a free fit of the sky background within GALFIT, but estimates a value before the fitting. We give details on the background estimation in the following section.

To determine the size of the postage stamps, GALAPAGOS uses the Kron radius. The user specifies a scale factor `C03` by which the Kron radius is enlarged. The decision for this scaling should be guided by trying to find a compromise between maximal area, to include as much flux of the central source (and maybe the closest neighbours) as possible, and minimal area, to speed up computation time of GALFIT. Finding a good compromise is important as elliptical galaxies require a larger area than spiral galaxies, owing to their extended and slowly dimming, low surface brightness wings. For the one-orbit *HST* surveys GEMS and STAGES, we found a factor of 2.5 to work well. GALAPAGOS does not enlarge the size of the postage stamps in the presence of close neighbouring galaxies. However, they are properly taken into account in the fitting process (see Sec. 4.5).

4.4. Sky Estimation

Obtaining a precise sky level is the most critical systematic in galaxy surface brightness profile fitting (see e.g. de Jong 1996; Häussler et al. 2007). To obtain a precise background measurement GALFIT is capable of including the sky as a free parameter when fitting a celestial source. However, using the sky as a free parameter requires an appropriate size of the input image, i.e. it has to contain *all* the flux of the primary source and most of the flux of neighbouring sources that are to be fitted simultaneously and ample sky. For estimating a proper sky background, the image should be as large as possible. However, as detailed above, large postage stamps become impractical once too many neighbouring sources are included. Only a manual setup may allow using the sky as a free model parameter. To enable automated processing of large numbers of objects, GALAPAGOS incorporates its own subroutine to obtain an optimal sky measurement before running GALFIT and hence uses a fixed value during fitting. With the proper setup, the resulting GALAPAGOS estimate improves significantly over values obtained from SEXTRACTOR.

We use a flux growth method to estimate the local sky around an object. Calculating the average flux in elliptical annuli centred on the object of interest while excluding other sources or image defects, we obtain the background flux as a function of radius. Once the slope over the last few measurements levels off, GALAPAGOS stops and determines the sky from those last few annuli (see Fig. 7).

For this procedure to work, we create a "sky-map", i.e. a copy of the input images where the pixel values indicate the nature of the contained flux. In the sky-map a pixel value of 0 stands for blank background sky, while positive numbers indicate the presence of a source. A value of -1 indicates no flux at all, as happens with *HST* images that are geometrically distorted (see Fig. 8). One might think that to make the decision between source or sky the aforementioned SEXTRACTOR segmentation map might suffice. Unfortunately, the level out to which SEXTRACTOR detects objects is rather limited. In particular with elliptical galaxies SEXTRACTOR underestimates the flux belonging to the object significantly. Changing the SEXTRACTOR setup parameters cannot totally remedy this. Therefore, a significant number of pixels still containing some source flux would be assigned as "sky". To circumvent this problem, we instead use Kron ellipses to determine the extent of an object. GALAPAGOS regards any pixel within a specified number of Kron radii `D03` (in the case of STAGES three) plus an additional fixed offset `D05` (for STAGES: 20 pix) as containing source flux. Note that this scaling factor `D03` does not have to coincide with the scale for the size of the postage stamps `C03`. Note also that GALAPAGOS records the total number of objects that might contribute to a certain pixel, i.e. when e.g. two sources overlap, the value in the intersection of the two Kron ellipses is also two. A weight map (exposure time map) specified by the user defines the off-chip pixels, which are given a value of -1. GALAPAGOS identifies these as pixels with zero exposure time.

GALAPAGOS takes special care to minimise the impact of large nearby sources on the background estimation process for the current object. To that end, GALAPAGOS relies on the SEXTRACTOR output catalogue to provide shape information. Under the assumption that all sources have a Sérsic index $n = 4$ and a half-light radius $r_e = (\texttt{flux\_radius})^{\alpha}$, with the SEXTRACTOR catalogue parameter `flux_radius` and a user specified power α (we chose $\alpha = 1.4$; `D11`) to convert the SEXTRACTOR `flux_radius` to a "true" half-light radius, GALAPAGOS calculates the flux of all catalogue objects at the position of the current source. Any source exceeding a user specified limit `D09`, GALAPAGOS regards as an important flux contributor for the current object. Subsequently, we will term the sources that are selected that way "contributors". Note that `D09` has the units of a magnitude, i.e. "exceeding" the given limit implies a number smaller than this value. As the SEXTRACTOR `flux_radius` is a rather poor proxy for the true half-light radius and without proper estimate for the Sérsic index, we opt for a rather conservative limit of this flux cut.

If a proper GALFIT fit exists for the contributors, GALAPAGOS subtracts their model profile from the input image temporarily, i.e. for the time of the current background estimation. Note that removal of a model profile includes convolution with the telescope PSF before subtraction. In order to optimise the profile subtraction, GALAPAGOS processes the SEXTRACTOR source catalogue in order of increasing magnitude. As the very few brightest sources have a significant impact on both the sky estimation and fitting of a large number of fainter sources, starting with the brightest galaxies is essential. We give further details about the sorting process in Sec. 4.6.

Normally, the Kron ellipse of the current object defines the starting radius for the iterative measurement of the sky background in increasing annuli. In case of the presence of potentially dominant flux contributors, for which no GALFIT model ex-

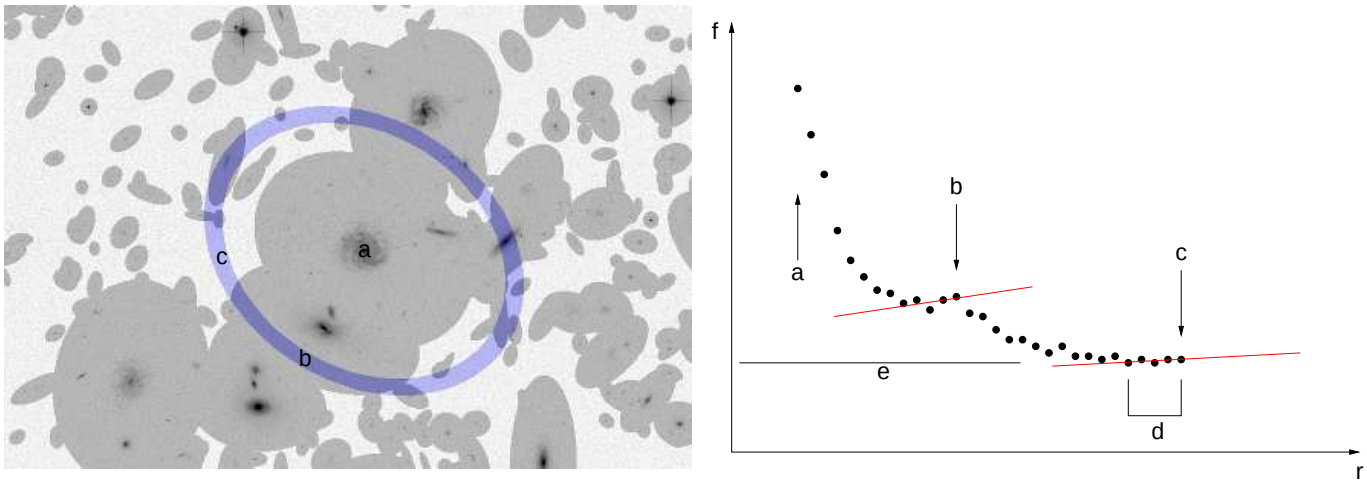

Fig. 7. Sky estimation. *Left:* The average flux f measured in elliptical annuli (blue) centred on an object (here a) determines the background level. In each annulus, we exclude regions surrounding other sources from the calculation (shaded area). For the indicated annulus, we exclude dark blue shaded regions b – only light blue regions c define the average background flux. *Right:* Flux f measured in an elliptical annulus as a function of radius r. *a:* Starting radius. *b:* Slope (indicated by the diagonal lines) turns positive for the first time, e.g. due to galactic structure at large radii. *c:* Slope turns positive for the second time. Here we stop the iteration. *d:* We compute slope measurements from the last n sky estimates (here: n=5; n is a user parameter). *e:* The adopted background sky level. See Sec. 4.4 for details.

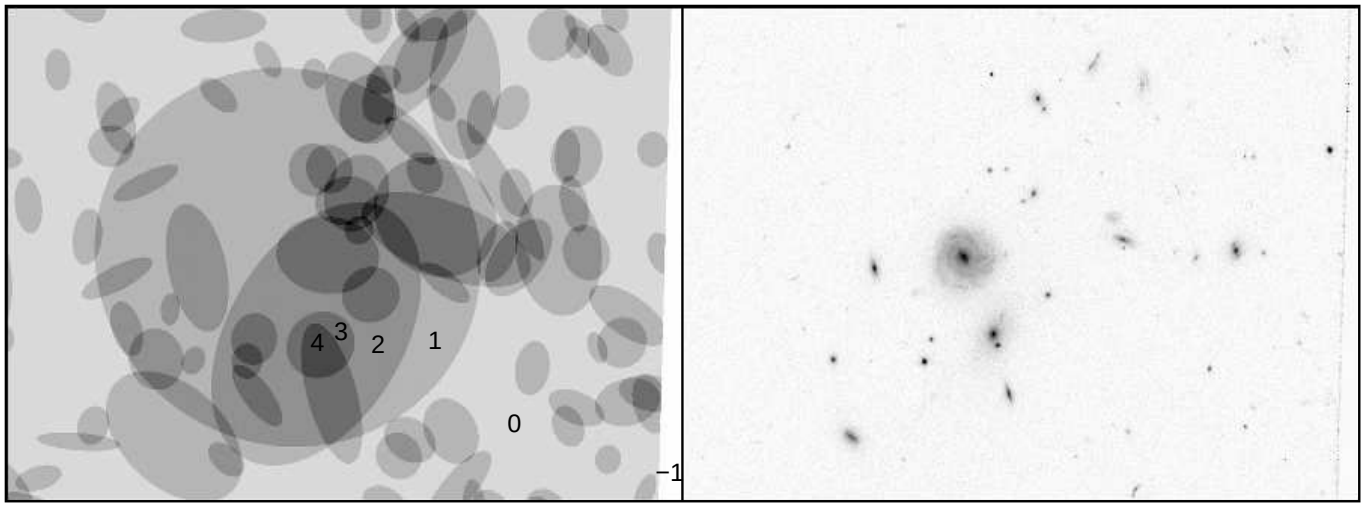

Fig. 8. The "skymap" (*left*): for each object that was detected in the image (*right*) we calculate the Kron ellipse and scale it up. Pixels inside a Kron ellipse get the value one. Pixel values stack, e.g. where two Kron ellipses overlap, the pixel value is two. Blank sky has a value of zero; pixels without astronomical flux, as occurs after removing image distortions e.g. in *HST* images, have a value of -1. Some pixel values are indicated.

ists yet, and hence were not subtracted from the input image, Galapagos increases the starting radius to the maximal distance of all such sources from the current. For each sky annulus, it estimates an average flux value excluding any pixels that were flagged as containing an object (or that were flagged as having a defect or no flux) in the skymap. Firstly, of the distribution of the remaining pixels, Galapagos clips all 3σ outliers. Then it fits a Gaussian function to the left over distribution, producing a mean value for the current annulus. After each new sky annulus measurement, Galapagos calculates a robust linear fit to the last few estimates (`D13`; in the case of STAGES 15 measurements). As long as source flux is still measurable, this slope is negative. Once this process reaches the true background, the estimated slope should start to randomly change its sign. When this happens for the second the time, Galapagos stops the loop and obtains the final background value from the last measurements. Stopping the process at the first positive slope measurement often results in suboptimal estimates as galactic inhomogeneities (like spiral arms) might produce dips sufficient to produce a slope sign change. However, using a much later slope change (than two) in practice is not necessary. Note that neighbouring sources are not a problem for the termination of this iteration as the method takes special care to take their influence into account (as shown above). This whole process is fully user-configurable, including options for the width of the sky annuli `D07`, their spacing `D06`, the initial starting radius `D08` and the magnitude cut `D09`.

4.5. GALFIT

Of the various light profiles built into Galfit, the most general one for galaxy fitting is the Sérsic model, which is used by Galapagos:

$$\Sigma(R) = \Sigma_e \cdot \exp\left(-\kappa\left[(R/R_e)^{1/n} - 1\right]\right), \quad (2)$$

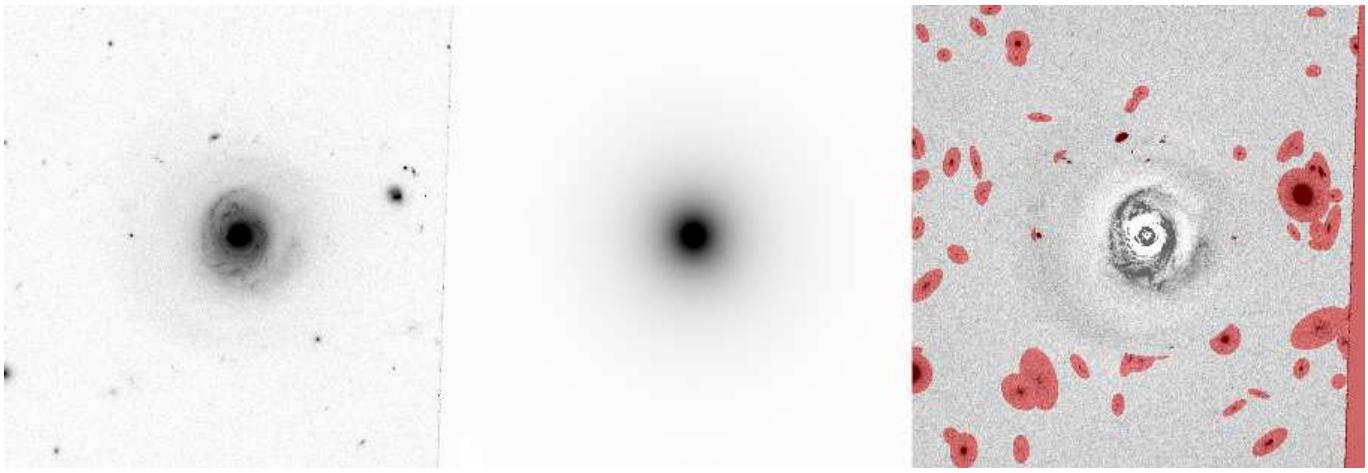

Fig. 9. Fitting Sérsic profiles with GALFIT. From left to right, the panels show the original galaxy image, the Sérsic model and the residual of image and model, respectively. GALAPAGOS excludes (masks) areas shaded in red from the fit. In this example no bright secondary sources were detected. The next brightest object after the primary source is too far away to become a secondary (for details on the definition of primary and secondary sources see Sec. 4.5). Note that the masked region at the right edge of the image results from the irregular shape of the HST images. This area has not received any flux and thus GALAPAGOS masks it as well.

where R_e is the effective or half-light radius, Σ_e is the effective surface brightness, $\Sigma(R)$ is the surface brightness as a function of radius R, n is the Sérsic index and $\kappa = \kappa(n)$ is a normalisation constant. The Sérsic profile is a generalisation of a de Vaucouleurs profile with variable Sérsic index n. An exponential profile has $n = 1$ while a de Vaucouleurs profile has $n = 4$.

A simple setup script controls profile modelling with GALFIT. It contains information about input and output file locations, PSF image, bad pixel mask, etc. A list of starting guesses defines what light-profiles are to be fitted. Although the downhill gradient method incorporated in GALFIT is often speculated to be prone to converging to a local instead of the global minimum, in practice we find it to be extremely robust, even in comparison to global parameter space search algorithms (Häussler et al. 2007). In application to high redshift survey data, the other two noteworthy features are the included bad pixel mask (i.e. pixels that are excluded from the fitting) and the number of simultaneously fitted objects.

We show an example of GALFIT output in Fig. 9. The left image presents the input postage stamp. In this case a single component (one object) was fitted. We show the resulting Sérsic model in the middle. Note that the brightness cuts and scaling in both left and middle panels are the same. The right panel displays the difference image of input minus model. Bright spiral features and dark dust lanes that strongly deviate from the smooth Sérsic profile are very prominent in this image. In order not to bias the fit by neighbouring sources and the image boundaries, GALAPAGOS excludes the shaded region from the fit.

In order to define which objects do not have a high importance for the current fit and hence may be masked instead of being fitted simultaneously, we define the following terminology: the target for the current fit is the *primary source*; any object whose Kron ellipse overlaps with that of the primary are *secondary sources*; objects without any overlap with the primary we term *tertiary sources*. We consider tertiary sources not to be important for the quality of the fit. As a result, we mask and exclude them from the modelling. Secondary sources might have an impact on the outcome of the parameters of the primary. Therefore, we do fit them simultaneously with the primary. We treat contributors (for a definition see Sec. 4.4) like secondaries. The difference between contributors and secondaries is, that we do not require an overlap of the Kron ellipses for contributors.

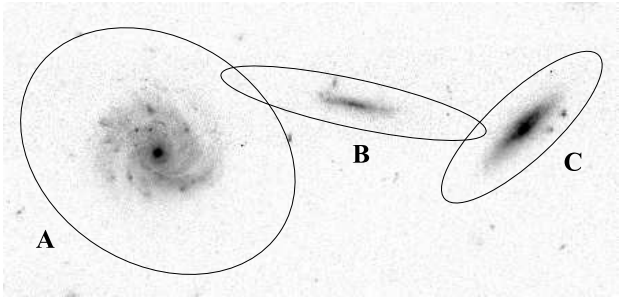

Fig. 10. Optimisation of the GALFIT setup. Circles indicate the Kron ellipses used for classifying the detected objects (as secondaries or tertiaries). This example was taken from real data. However, for clarity we do not mark the faintest detections in this image. For details see Sec. 4.5.

Simultaneous Fits

Often, sources are so close to each other, that they are best fitted simultaneously. One might argue that after a simultaneous fit of two sources, A and B, the best parameters are known for both objects. However, in general this is not true. Exemplary we construct a situation with three sources A, B and C, where C is on the opposite side of A with B being in the middle (see Fig. 10). Let A be the brightest source of the three, i.e. A is fitted firstly (see also Sec. 4.6). The Kron ellipses of A and B and those of B and C overlap, while the Kron ellipses of A and C do not. Fitting of A implies fitting of B simultaneously: B is a secondary to the primary A. C is not connected to A and therefore following our prescription we mask it. As a tertiary we exclude it from the fit. The resulting fit for B thus is not optimal, as it neglects the presence of C, which is important for fitting B, but not for fitting A. To obtain the optimal fit for B, we have to fit B as a primary. In this case A and C are secondaries as their Kron ellipses both overlap with B, and are fitted simultaneously. To speed up the fitting of B, we can now insert the known parameters for object A, thus effectively removing one component from the fit. This

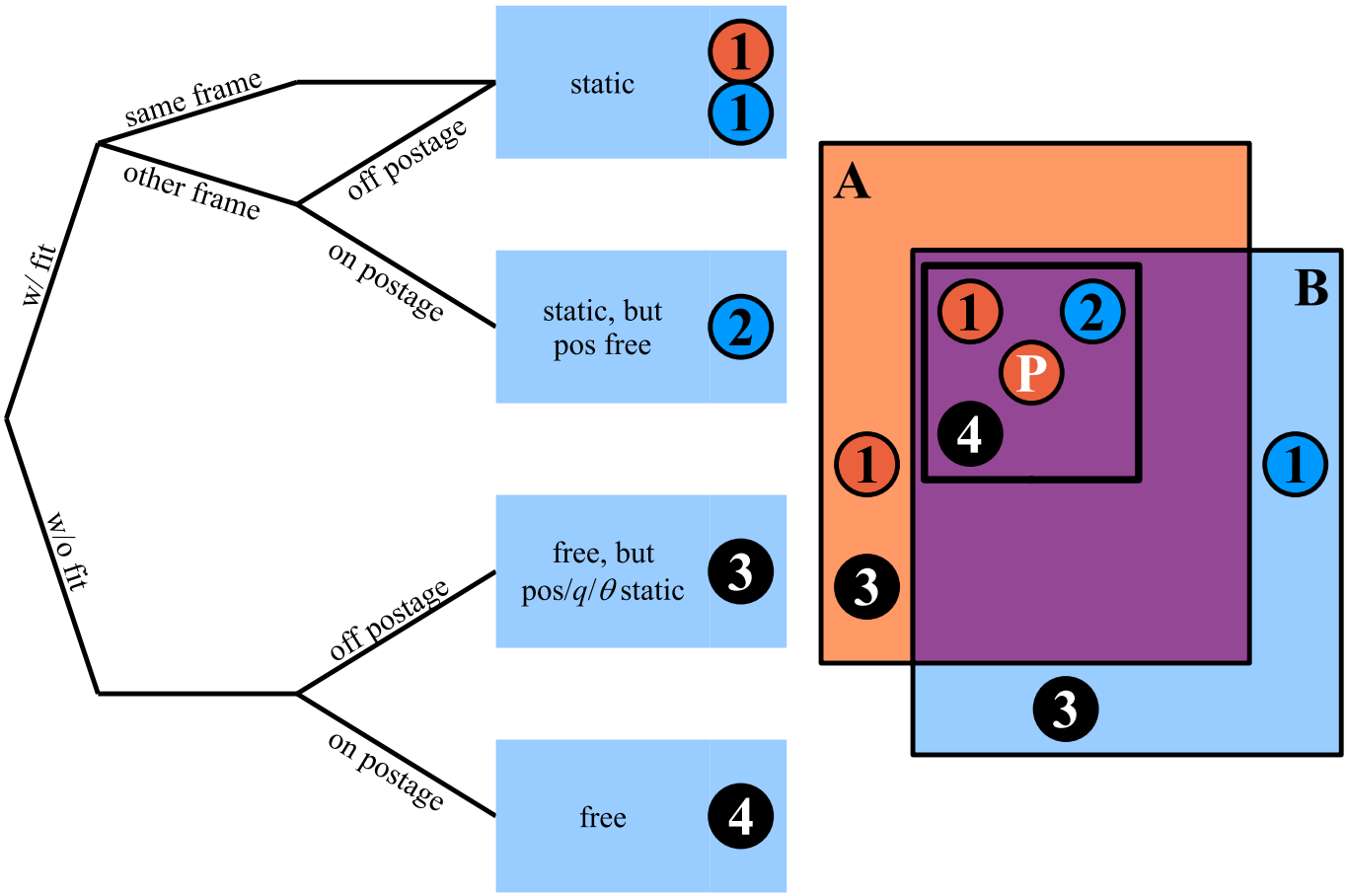

Fig. 11. GALFIT parameter setup scheme for secondaries and contributors. Depending on the relative position to the primary target, i.e. on the same postage stamp or not, with a pre-existing fit from the same or another survey image or without a fit, we show the setup for the GALFIT parameters (*left panel*): static implies all parameters are fixed to their initial guess (i.e. the SEXTRACTOR estimates), while free means that they are variable throughout the fit. In some cases the position pos, the axis ratio q or the position angle θ take a different state than the remaining fit parameters. We visualise the situation in the *right panel*: The current primary P is located on the red survey image A, which has some overlap (purple) with the blue image B. The solid black outline indicates the postage stamp corresponding to P. Potential secondaries or contributors are numbered. For sources with a black background (*3* & *4*) no prior fit exists (SEXTRACTOR values are used as static/fixed profile parameters), while for targets shown either in red (*1*) or blue colour (*1* & *2*) a fit from the respective survey image is available. For further details see Sec. 4.5.

example highlights the importance of fitting all objects once as primaries, while secondaries may be made static if a fit already exists.

Normally secondary sources are fitted simultaneously with the current primary object (see above). Using a pre-existing fit (as in the example) as static parameters for a secondary source, thus, is an exception to this rule. A further complication is that the existing fit for the secondary may have been obtained from a different survey image as the current primary. In that case, the central position of the secondary has to be converted via the world coordinate system information from the original pixel coordinates to the current system of the primary. Therefore, to allow optimal centring after such a conversion, GALAPAGOS fixes all parameters for the secondary, but its central coordinates. If in the previous example, when fitting object B with a pre-existing fit of source A, the fit for A was performed on a different survey image than B, then the pixel centre of A would not be static. However, a free pixel centre is only required if the centre of the secondary A is also inside the postage stamp of the current primary B. If the centre is off the postage stamp, sub-pixel accuracy is not required any more for an optimal fit, and all components of A are made static. We visualise this situation in Fig. 11 (case *1* and *2*).

Furthermore, if no fit exists for a secondary, a free fit for that source is not always the best solution. In the case that the centre of the secondary is not on the postage stamp, a free fit results in too many degrees of freedom. In GALAPAGOS we opt to then fix the position, axis ratio and position angle to the values provided by SEXTRACTOR (while leaving the remainder as free parameters). This is justified because, on one hand, more than half the flux of the secondary cannot be seen by GALFIT, thus making it increasingly difficult to come up with precise estimates for these parameters. On the other hand, the values given by SEXTRACTOR usually have high enough accuracy not to bias the fit of the primary significantly. Essentially, in such a case GALFIT will improve on the profile shape (size and Sérsic index) while leaving the overall geometry (position, axis ratio and position angle) at the SEXTRACTOR values (see Fig. 11; case *3* and *4*).

In addition to the "normal" sources (secondaries and tertiaries) in the immediate surroundings of the current object, GALAPAGOS has to take bright and large contributors as defined in Sec. 4.4 into account as well, although these sources may be off the current survey image. It treats them as secondaries without the requirement of their Kron ellipse to overlap with the Kron ellipse of the primary. In terms of the parameter setup, GALAPAGOS handles them exactly like other secondaries.

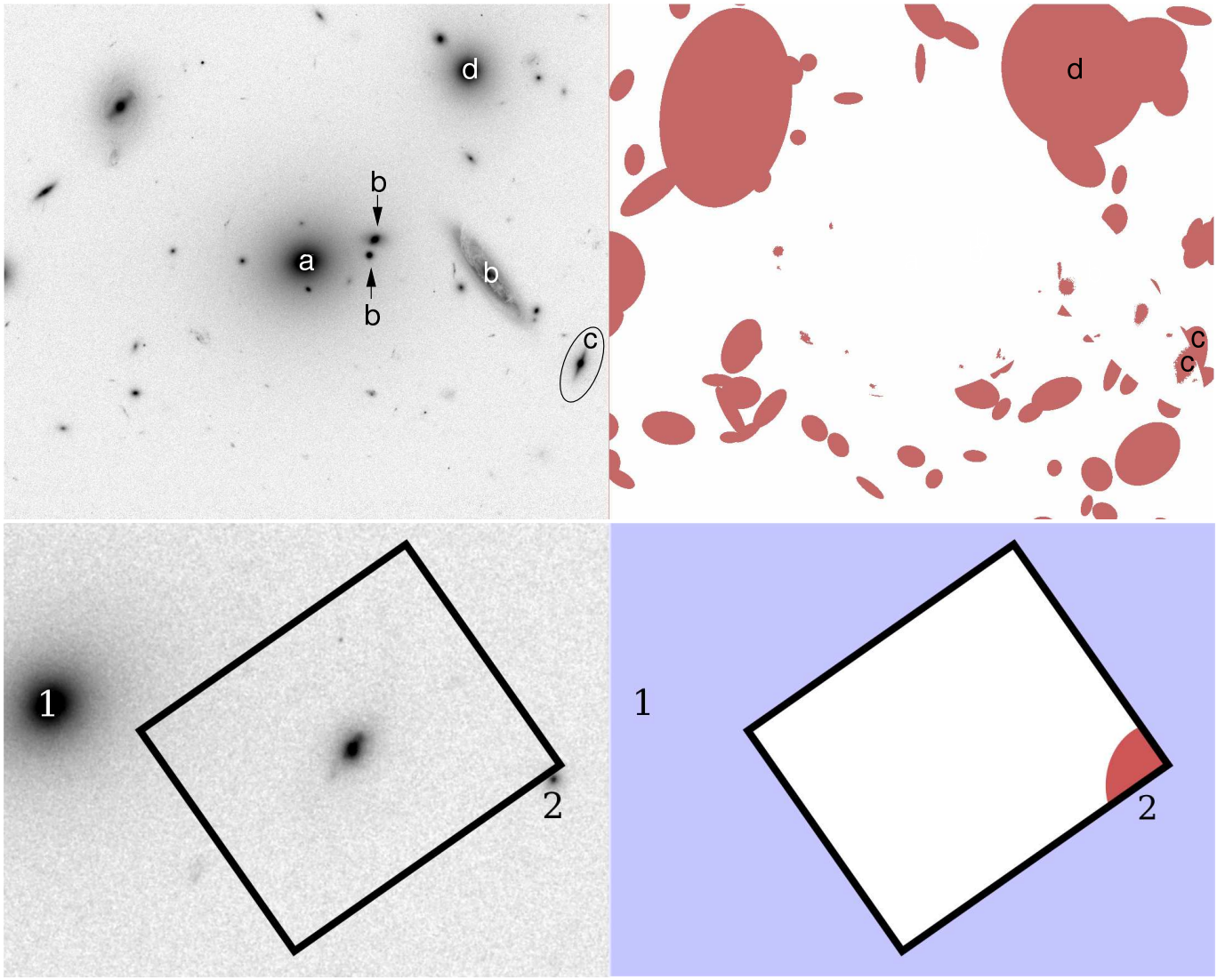

Fig. 12. Mask creation. The left panels show galaxy images; the right panels the corresponding bad pixel masks. In the bad pixel masks white and red represent good and bad pixels, respectively. *Upper panels:* a) and b) indicate primary and secondary sources, respectively. c) and d) mark examples of tertiary sources: c) is only partly masked, as it overlaps with a secondary source; d) is masked completely for not having any overlap with the primary or a secondary. *Lower panels:* The plotted area shows a postage stamp (indicated by the solid rectangles) and some of its surroundings. Note that the postage stamp is tilted (representation in world and not pixel coordinates) and that the blue area is actually not part of the postage stamp. 1) is a source that might potentially contribute to the fit of the primary, due to its brightness, and is included as a secondary source (with parameters fixed from a previous fit) although its centre is off the current postage stamp. 2) is a tertiary source without overlap with the primary and being too faint to contribute significantly, and is therefore masked completely (red pixels inside the postage stamp).

Bad Pixel Masks

GALFIT supports so-called bad pixel masks (see Peng et al. 2002) to exclude image regions from fitting and thus speed up the fitting process. As tertiary sources may overlap with secondaries, we take the following approach to define the area to be masked. In general, GALAPAGOS masks the full Kron ellipse, enlarged by a user-specified factor `D04` (which may have a different value as the one used for computing the skymap `D03`) and an additional offset `D05`, for the fitting. If the Kron ellipse of the tertiary overlaps with the Kron ellipse of the secondary, GALAPAGOS *includes* the intersection in the fit. However, as the included area might contain significant flux (maybe even the nucleus) of the tertiary, it *excludes* any pixel marked in the SEXTRACTOR segmentation map as belonging to the tertiary. Thus, the resulting shape of the mask may look complicated, yet this procedure ensures having the fit of the secondary targets only mildly affected. The primary source should not be significantly affected at all.

To speed up the fitting process by reducing the number of simultaneous fits, GALAPAGOS masks secondary objects based on a magnitude criterion `D16` (for extended and `D17` for point sources) in comparison to the primary source. In the case that they are too faint compared to the primary, GALAPAGOS "downgrades" them to tertiary status and treats them as such, i.e. it masks their Kron ellipses completely, but for parts which overlap with other secondaries or the primary that are not covered by the SEXTRACTOR segmentation map.

Pixels that have a value of zero in the weight map, i.e. an exposure time of zero, GALAPAGOS also masks. Obeying these rules results in masks as shown in Fig. 12.

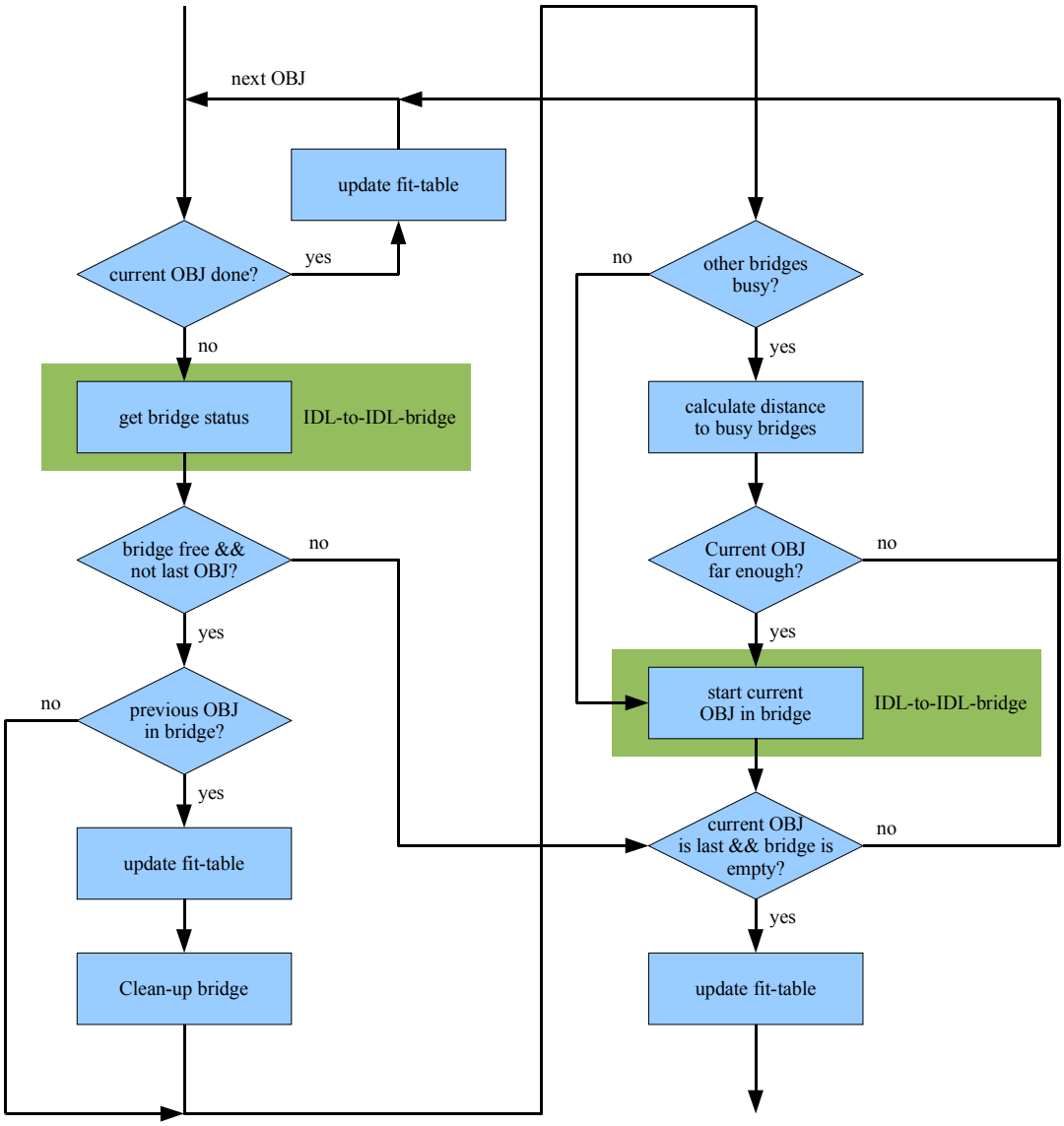

Fig. 13. Schematic of the main loop for fitting the brightest sources. Actions that make direct use of the IDL-to-IDL-bridge are shown with a green background.

4.6. Computational Optimisation

In the following section we will describe additional characteristics of GALAPAGOS that increase the efficiency and robustness of the code.

Sorting and Parallel Computation

After running SEXTRACTOR and cutting postage stamps, GALFIT fits the individual catalogue sources. Because accurate estimation of the sky background requires ordered processing of the data, which is extremely inefficient in terms of total CPU time, we have developed methods to speed up this sequential process. In the next paragraphs we will describe the mechanisms that are incorporated into GALAPAGOS to switch from sequential to parallel processing and to increase the overall efficiency and robustness of the code.

The first method to increase the computational efficiency is to optimise the specific setup of GALFIT in cases where several objects are to be fitted at the same time. Under certain circumstances the parameters of secondaries may be made static. This is shown in detail in Sec. 4.5. The fact that GALAPAGOS sorts all objects by SEXTRACTOR magnitude to increase the accuracy of the sky background estimation maximises the impact of this method. Beginning with the brightest sources assures that the sources that take the longest to fit are held fixed when fitting a larger number of the fainter objects. Another reason in favour of sorting objects by magnitude is that the efficiency with which bright contributors are included in the current fit is greatly enhanced.

In order to allow parallel processing while still maintaining the advantages of a sequential mode, we built different mechanisms into GALAPAGOS. One method is to encapsulate the sky fitting and GALFIT processing and running the code on one survey image only, at a time. This will then enable the user to run several instances of the code simultaneously on n different computers, thus reducing total computation time by a factor of n. This is realised in GALAPAGOS by specifying which tiles are to be processed in a so-called "batch file" (`E01`). The problem with this approach is that sources may extend from one survey image onto the next. Therefore, one might run into the situation where tile A is fitted before tile B, with the brightest source in the two tiles being on B and reaching into A. In this case, a fit

for the brightest object is not available for estimating the optimal sky background for a number of galaxies on tile *A*. Note also that this method assumes that the average object size is much smaller than the size of a survey image.

To alleviate the problem we divided Galapagos into two parts:
a) In the first part, Galapagos treats a fraction of all sources on all tiles in a sorted order. This assures that the brightest galaxy from tile *B* is fitted before Galapagos treats tile *A* or *B*. This part still requires sequential processing without the possibility to run other instances of Galapagos at the same time. Also, it produces a rather large computational overhead, as potentially with every new source a number of large images (science image, weight image, segmentation map, etc.) are to be loaded into memory and processed. A possible working definition for the fraction of sources that have to be fitted sequentially might encompass all sources that span an area larger than the size of the overlaps resulting from the survey's tiling scheme.
b) The second part is kept as detailed above: Galapagos processes all objects within a tile in order of increasing magnitude. Several instances of Galapagos may be run simultaneously on *different* tiles. With the sources that potentially reach into neighbouring tiles already processed in step a), now survey images may be treated as individual entities, which can be processed out of order and simultaneously.

We built a last mechanism into Galapagos to speed up the first sequential stage a) and make it "quasi-parallel". With IDL 6.3 the "IDL-to-IDL-bridge" was introduced, allowing spawning multiple IDL sessions within the one currently running. With this feature it is possible to start processing one object in "thread" a) and then continue with the next object in "thread" b) (see Fig. 13). Only the total number of available CPUs on the system limits the amount of parallelisation. On current dual processor mainboards with a quad-core CPU up to eight objects may be treated simultaneously. The primary limitation comes from the requirement to process objects sequentially, though. Before starting with a new source, Galapagos checks whether the distance between all currently running objects and the source that is to be started is larger than a user specified distance `D20`. The extent of the brightest object in the survey determines this distance and it should exceed the limit out to which this object might have an influence on the fitting of neighbours. If the next source in the queue is too close to objects currently being executed, the code waits for these objects to finish. The second stage (lower section in Block D; see Fig. 1) does not require this optimisation as starting multiple instances of the code allows a much more direct interaction with it and a simpler mode to partially interrupt and restart operation.

Neighbouring Tiles

During the sky background estimation Galapagos calculates the influence of all objects on the currently processed source. Depending on the size of the survey, this check for contributors takes up a significant fraction of the complete source loop computation time. However, the sources immediately required for processing the current object are only the ones that may have an impact on the fitting or background estimation. Therefore, by specifying the "reach" of the brightest sources allows to restrict the computation to a much smaller fraction of all sources. This is done by providing the total number of closest tiles n that are to be included in the calculation (`D18`). If the tiles are taken on a regular grid, $n = 8$ defines a ring surrounding the tile of the current object (see Fig. 14). Note that in case of a tile at the edge of the survey this "ring" is not cut in half, but all tiles are selected on just one side. Galapagos *always* selects the nearest n neighbours. It calculates the distance between tiles from the centres of the images.

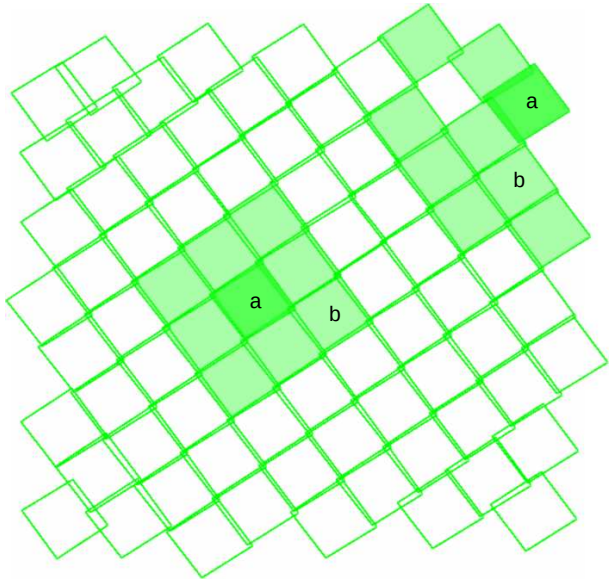

Fig. 14. Definition of neighbouring tiles. For each survey image the n closest neighbours define its immediate neighbourhood. In a checkerboard configuration $n = 8$ corresponds to a 3×3 pattern. At the edge of the survey more distant tiles are included. The light green shaded areas *b* indicate the neighbourhoods for the central tiles *a*.

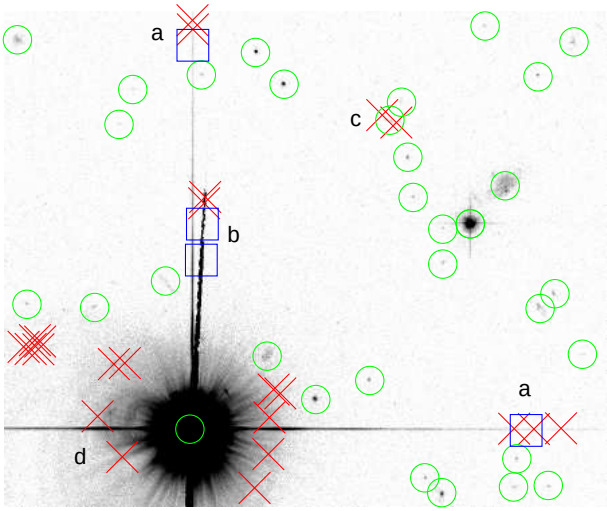

Fig. 15. Correction of detection errors. Red crosses indicate "critical" detections; blue boxes mark "catalogue" detections; green circles show "good" source detections. For a definition of these terms see Sec. 4.6. *a:* The diffraction spike of the star was picked up as multiple individual sources. Most of them are critical failures. Yet, one detection in each spike was kept as a catalogue source, in order to guarantee that the fitting of nearby objects is not biased. *b:* Pixel bleeding from the star. Again, some detections were flagged critical, others as catalogue sources. *c:* An over-deblended object. The excess detections are critical errors. *d:* Spurious detections in the vicinity of the bright star. All are critical failures. Note that the categorisation of sources is an optional, subjective process, which is not performed automatically by Galapagos.

Detection Flags

A *perfect* setup for SExtractor never exists. In a small fraction of all detections one or the other failure occurs, e.g. (over-) deblending, non-detection, spurious detection, etc. In particular in the surroundings of bright stars (or even galaxies) these errors accumulate. With respect to setting up Galfit properly, there are two classes of failures: the "critical" and the "catalogue" failures. Depending on their relative brightness compared to nearby "real" objects they either have to be removed before the fitting (faint sources; "critical") or after (bright sources; "catalogue").

A critical failure is a detection error that should be corrected *before* running Galfit. Critical detections *do not affect the fitting of neighbouring real sources*. Examples are over-deblends, cosmic rays or a bad detection at the image edge. Critical failures include any unwanted detection that might erroneously include additional unnecessary components in the fitting of real objects. We give an example for an over-deblended source in Fig. 15 *c* and indicate several spurious detections in Fig. 15 *d*.

In contrast, catalogue failures are detections that one has to remove *after* running Galfit. They are bright in relation to neighbouring sources and they *might affect the fitting of nearby objects* if not included as separate components. Typically, they are connected to cosmetic "defects" of the image. A common example for these are diffraction spikes of stars, which may not be included in the PSF model. Therefore, Galfit may not properly fit a galaxy close to such a spike, as too much flux is in the spike compared to the source. Common are also satellite trails or pixel column bleeding of saturated stars. We show some examples in Fig. 15 *a* and *b*.

Galapagos can optionally take care of both these failures. If the user provides a (manually created) list of positions for critical and/or catalogue failures (one file each), Galapagos will remove any source found within a specified radius `B16` around these positions from the catalogues at the proper stage in the process. Otherwise, Galapagos treats them simply as normal sources and provides Sérsic fits for them.

To classify unwanted detections (into one of the two categories), the user should decide whether an object is required for obtaining a proper fit with Galfit for neighbouring "real" sources, or not. In principle it is save to put any detection error into the catalogue failure list. This might lead to prolonged fitting times, though. In practice, most detection errors are faint enough to not influence neighbours and should therefore be put into the list of critical failures.

Note that the definition of whether an object is a critical or catalogue failure is subjective and depends on the user. However, the correction of these errors is an option. Galapagos will run perfectly well without any manual treatment. In that case, the user will have to live with the fact that some (small) fraction of sources might be affected by this.

Treatment of Stars

A problem related to fitting bright saturated stars is that they are often much brighter than the stars that one can use as a PSF model. Because there is a limited dynamic range, the PSF cannot adequately capture the tails seen around brighter stars, which may then contaminate neighbouring galaxies. To deal with this situation, we fit Sérsic models to stars instead of the usual PSF model, because a high Sérsic index produces a model with extended tails. However, in so doing, it may cause Galfit to not converge within a reasonable amount of time. As the focus of Galapagos is on modelling the properties of galaxies, no further attempt was made to apply a different, more elaborate model (instead of the Sérsic profile).

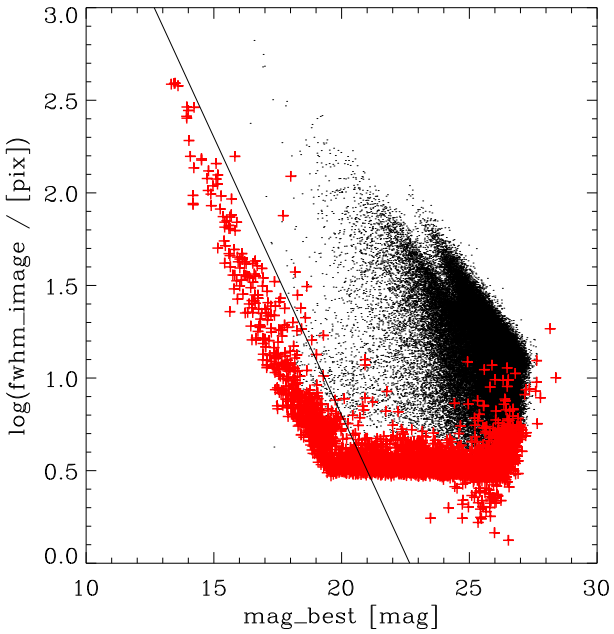

Fig. 16. Treatment of saturated stars. Here we show source detections from the STAGES survey in the log(`fwhm_image`) vs. `mag_best` plane. Red pluses mark objects identified as stars (see Gray et al. 2008). The line indicates the cut used to identify saturated stars (left of the line). Black dots show other (extragalactic) sources.

To resolve the resulting problem with the convergence of Galfit Galapagos identifies saturated stars in the magnitude-size plane, which is represented by the SExtractor parameters `mag_best` and log(`fwhm_image`) (see Fig. 16). The user specifies the zeropoint `D15` and slope `D14` of a line below which Galapagos treats objects as saturated stars (i.e. on the bright and compact/small side). The reason for many of the brighter stars to fail in the fitting is the detection of a large number of neighbouring secondary sources (including stellar diffraction spikes), which have to be modelled simultaneously. To reduce the number of these secondaries the user specifies a relative magnitude cut `D17`, below which secondaries are not fitted any more and treated as tertiaries. For the STAGES data, all objects more than two magnitudes fainter than the primary star ($m_{\rm star} - m_{\rm object} > 2$) were subject to this. Note that for galaxies the same limitation applies, but at a much weaker level. Again a magnitude limit `D16` may be specified (e.g. $m_{\rm galaxy} - m_{\rm object} > 5$). Restricting the number of secondaries to those objects bright enough to influence the fit and removing the fainter ones resolves the issue.

5. Data Quality

We have tested Galapagos thoroughly using simulated data as described in more detail in Häussler et al. (2007) and Gray et al. (2008). For the simulations applied here, we use the same setup as for fitting the STAGES survey. All in all, the simulated datasets contain around 7 million galaxies. Excluding the ones that are not recovered by Galapagos for being below the detection threshold (3 million) and the ones that ran into any given fitting constraint (~280 000; the following constraints for the Sérsic index n, the half-light radius $R_{\rm e}$ and the magnitude m were applied: $0.2 < n < 8$, $0.3 < R_{\rm e}$ [pix] < 750, $|m_{\rm GALFIT} - m_{\rm SExtractor}| < 5$) or where the fit crashed (293), leaves us with around 3.7 million successfully fitted galaxies.

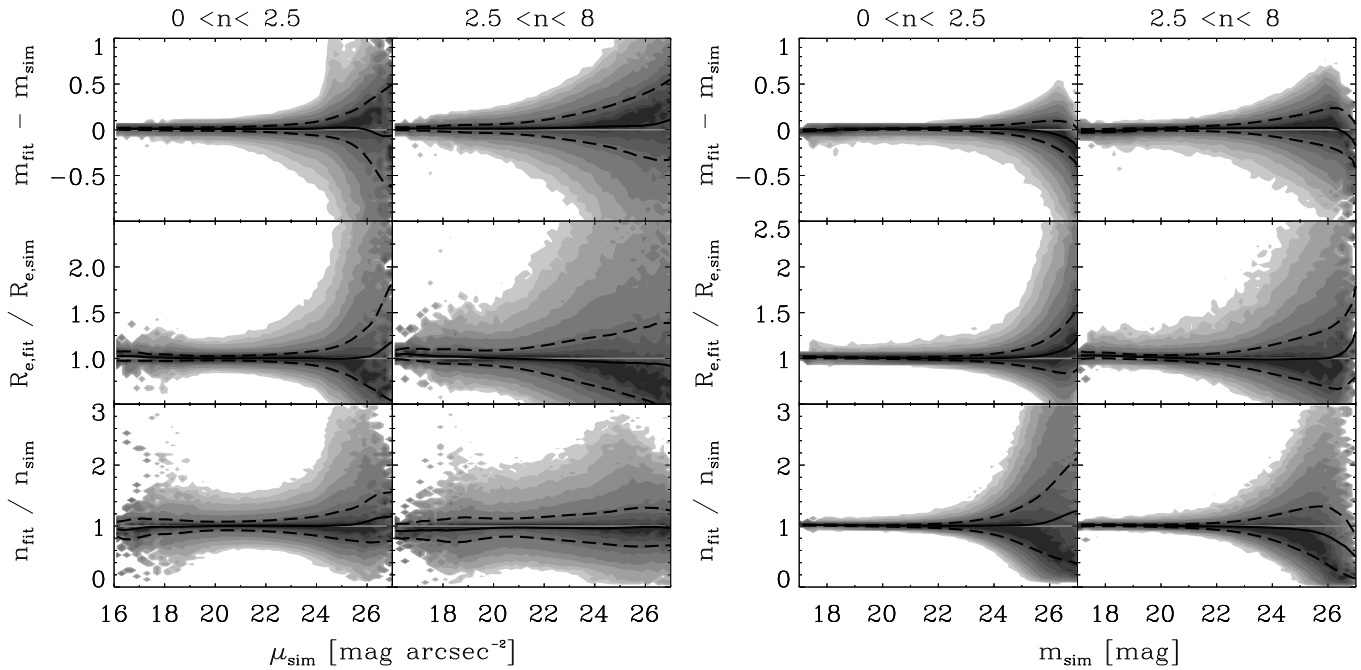

Fig. 17. Parameter recovery as a function of simulated mean surface brightness $\mu_{\rm sim}$ within $R_{\rm e}$ (*left panel*) and simulated magnitude $m_{\rm sim}$ (*right panel*) for two different Sérsic index ranges (disc-like galaxies with $n \approx 1$ on the left-hand side, early-type galaxies with $n \approx 4$ on the right-hand side). Contours show galaxy density, with each bin being normalised to its own peak value. As a result, contours roughly resemble mean value and standard deviation. Due to an asymmetric distribution and different binning, the mean value (black line) deviates slightly from the peak values for fainter galaxies. The 1-σ scatter of the distributions is shown as well (dashed lines). The light grey line indicates the ideal zero-level. Fainter galaxies (both as function of magnitude and surface brightness) and galaxies with higher n are fitted less accurately. Also, for the brightest galaxies in the sample, the deviation increases. Most likely their brightness (and size) makes them the most difficult objects to setup for fitting, because of a large number of simultaneously fitted neighbours and because of having the highest uncertainty in their background sky estimate.

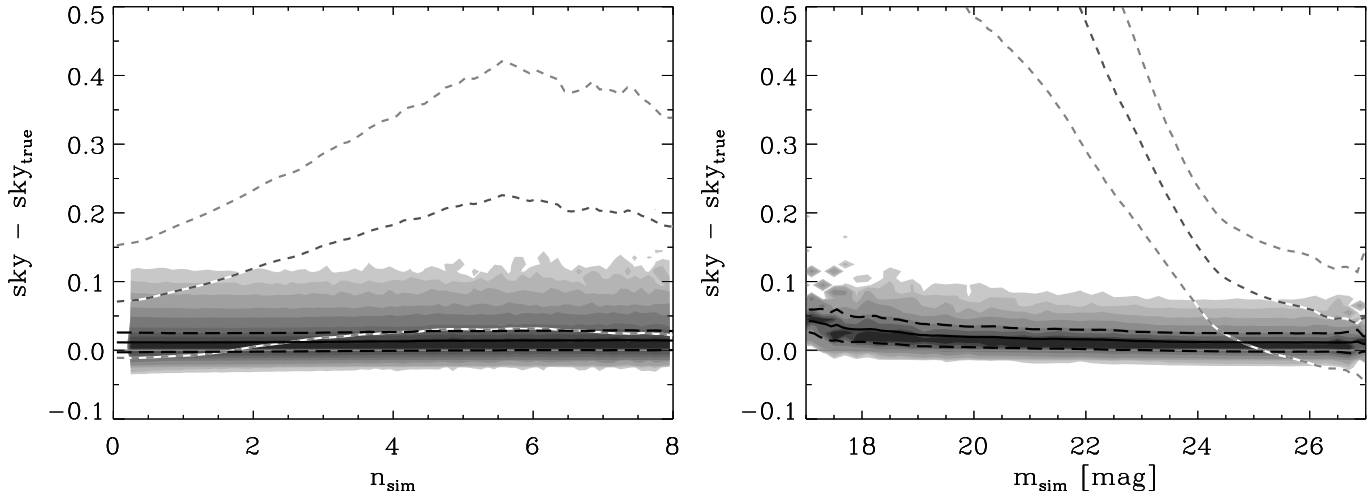

Fig. 18. Sky recovery (flux difference in counts) as a function of simulated galaxy Sérsic index (*left panel*) and simulated magnitude (*right panel*). Contours and black lines show the distribution, mean and σ of the estimated sky as recovered by Galapagos, white/grey dashed lines indicate mean and σ as provided by SExtractor. Galapagos recovers a very accurate sky value independent of galaxy structure, whereas SExtractor overall exhibits a much larger offset, scatter and dependence on galaxy morphology. At the brightest magnitudes a slight trend is seen in Galapagos while SExtractor performs worse by a factor of ~50. Note that the right panel shows only objects with $3 < n < 5$, and thus portrays a rather conservative scenario.

The left panel of Fig. 17 shows the deviations of the three most important fitting parameters magnitude m, effective radius $R_{\rm e}$ and Sérsic index n as a function of simulated mean surface brightness $\mu_{\rm sim}$ within $R_{\rm e}$ for two different regimes of Sérsic index. We choose the samples such that the completeness as a function of magnitude is roughly 90% for all galaxies. The low Sérsic index sample ($m_{\rm sim} < 24.5$) contains ~1.1 million galaxies, the high Sérsic sample ($m_{\rm sim} < 25.25$) contains ~470 000 galaxies. Obviously, Galapagos' performance decreases at faint magnitudes and high Sérsic indices. The right panel of Fig. 17 shows the same plot, but as a function of simulated magnitude $m_{\rm sim}$ rather than surface brightness to illustrate the same effects in another commonly used parameter space. Again, we choose a cut to select only galaxies with a surface brightness completeness exceeding 90%. The low Sérsic index sample ($\mu_{\rm sim} < 22.25$) contains ~780 000 galaxies, the high

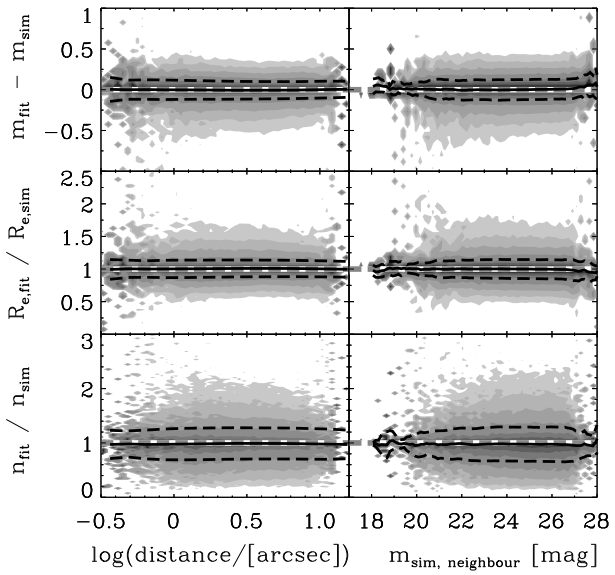

Fig. 19. Parameter deviations as a function of both distance (*left panel*) and magnitude (*right panel*) of the nearest neighbouring galaxy. The thick grey/white dashed lines indicate the zero-level; black solid and dashed lines show the mean and σ of the recovered parameter, respectively. Grey contours represent the normalised distribution of recovered parameter values.

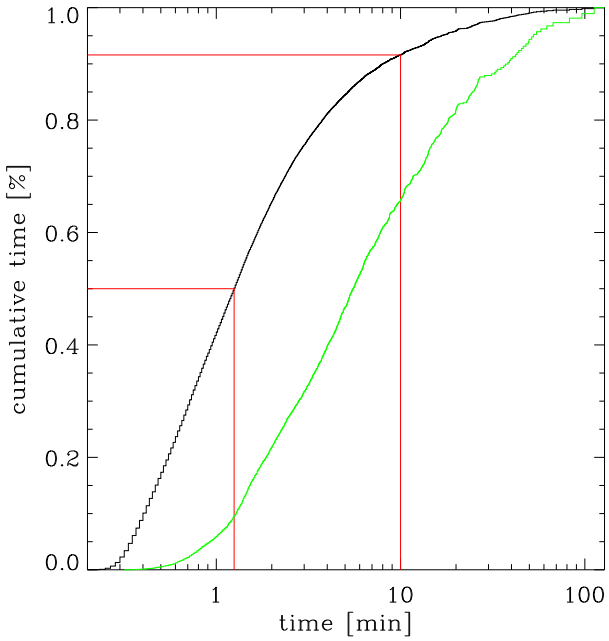

Fig. 20. Performance of the galaxy modelling with GALFIT. Cumulative histogram of the fitting time per object as a fraction of the total fitting time. The two histograms show all galaxies (black/left) and the brightest 5% (green/right).50% of all sources take less than 1.25 min to fit and more than 90% of all objects are done within 10 min (red lines). The brightest 5% take about a factor of 5 longer.

Sérsic sample ($\mu_{\rm sim} < 23$) contains ~295 000 galaxies. At the faint end, quite expectedly, the recoverability of parameters gets worse. In both panels of Fig. 17, we see no significant systematic trends down to very faint levels.

The left panel in Fig. 18 shows the deviation of the sky value (as recovered by GALAPAGOS) from the true sky value (as derived from the empty noise image used for the galaxy simulations) as a function of the simulated Sérsic index n of the primary object. Obviously, the recovery of the sky in GALAPAGOS is completely independent of n. Compared to the SEXTRACTOR value for the local sky, which shows both a much bigger offset and a larger standard deviation, the recovery is close to ideal with very small offset and scatter. We derive the true sky value for this plot from simple statistics on an empty noise image used for the simulations.

Furthermore, we investigate the magnitude dependence of the sky recoverability (see right panel of Fig. 18). Here we select only objects with a Sérsic index $3 < n < 5$ (~300 000 objects), which due to their extended low surface brightness wings are hardest to fit and estimate a background value. Thus, we portray a conservative worst case scenario. While for the large majority of all objects there is no trend to be seen at all, at the bright end the estimates provided by GALAPAGOS do diverge slightly: at $m = 17, 18$ the mean sky moves off by ~0.04/0.03 with a scatter of ~0.02, respectively. For comparison, the values recovered by SEXTRACTOR are ~2.3/1.3 with a scatter of ~0.4 at the same brightness.

To examine the influence of neighbouring galaxies in a similar fashion as was shown in Häussler et al. (2007), we plot parameter deviations over both magnitude of and distance from the next neighbour. The next neighbour we here define as the closest simulated galaxy that was found by SEXTRACTOR. This does not necessarily imply that this galaxy had to be properly deblended and simultaneously fitted when running GALFIT (i.e. assuming a rather conservative definition resulting in a worst case scenario). We show these deviations in Fig. 19. In contrast to the analysis in Häussler et al. (2007), we now have enough statistical significance to separate both effects. We only show neighbours with $21 < m_{\rm sim} < 23$ (~330 000 galaxies) in the left panel and neighbours with a distance $1 < d$ [arcsec] < 1.6 (~280 000 galaxies) in the right panel of Fig. 19 to not confuse the two distinct effects: contamination by bright neighbours and contamination by close neighbours. As one can see from these plots, GALAPAGOS results do not show any dependence on either of these parameters. From this plot we conclude that the deblending and fitting scheme applied in GALAPAGOS works well and successful deblending of clustered fields (as e.g. STAGES) is possible.

6. Performance

We measure the performance of GALAPAGOS when applying it to the single-orbit *HST* survey STAGES (see Gray et al. 2008). STAGES is a mosaic composed of 80 tiles in the F606W filter containing ~75 000 sources. The survey being centred on a nearby galaxy cluster system at redshift z ~0.16, it provides the ideal test case including a high fraction of large and also peculiar objects. Large objects serve as a test for the deblending process during the source extraction, while peculiar objects like mergers or saturated stars with diffraction spikes pose a challenge for the modelling with GALFIT.

The fitting process with GALFIT is the main limitation for GALAPAGOS ' performance. The largest amount of time was spent on fitting the fainter 95% of all sources in the parallel mode (second part of block (D) in Fig. 1). Using eight 2.2 GHz CPUs in parallel, this process (i.e. the slowest of the eight) takes ~260 hours. There is potential for further improvement by increasing the total number of CPUs. This would also reduce the overhead resulting from individual pipelines not finishing at the same time (i.e. pipelines with fewer sources finish sooner), resulting normally in much less than the total number of available CPUs running simultaneously at the end of the fitting.

For the first part of block (D), the fitting of the brightest 5% of all sources we used four 2.4 GHz CPUs. This part of the fitting takes ∼150 hours. Note that moving from four CPUs to eight does not necessarily imply halving the required computation time. The performance increase at this stage depends on the survey geometry. A wide area survey has a higher efficiency than a smaller survey of the same depth, because of the higher probability that the brightest objects in the survey are further apart from each other, thus allowing a higher multiplicity. Fig. 20 shows a cumulative histogram of the fitting time per object. Note that the brightest objects take considerably longer to fit than the rest thus explaining the necessity to find a good compromise between the time spent in the two stages.

The remaining blocks take up an almost negligible fraction of the total processing time. Block (B), the SExtractor stage, takes ∼13.5 hours, including HDR mode. Cutting the postage stamps in block (C) requires ∼2.5 hours and the last block (F), compilation of the output catalogue, finishes within ∼0.7 hours.

Note that overheads for adjusting the setup and preparing the parallel fitting is not taken into account in the numbers cited above. Also, for varying survey layouts/configurations relative fractions of the total processing times between the various stages might vary significantly.

7. Summary

We present Galapagos, a software for automating the process of detecting sources and modelling them with single Sérsic profiles. Galapagos incorporates SExtractor and Galfit to perform these two tasks. In addition, it provides HDR source extraction, a postage stamp cutting facility and a robust means of estimating a local sky background. It stores results in a combined FITS table, excluding duplicates resulting from detections in overlapping tiles. We optimised the code for speed and stability, making use of modern multi-core CPUs and allowing a high degree of multiplicity. Another aim was to present the user with a simple setup, yet enabling control over all features of the code. As a result, Galapagos can be used on a wide variety of survey applications, from single tile deep observations to wide area shallow surveys. Galfit's ability to work with any given PSF enables application of Galapagos to both space- and ground-based data. We tested Galapagos on an extensive set of simulations and find it to be extremely robust in terms of parameter recoverability. Note that the results of the fitting depend on the choice of the input parameters. For example, a bad SExtractor setup will have a significant impact on the fitting procedure and thus lower the quality of the output catalogue.

The main feature that will be implemented in Galapagos in the near future is the option for a consistent two-component bulge-disc fitting. This will also include an estimator providing information about whether the increased amount of data potentially allows further insight into the structural composition of the object or not. Based on this idea, we will also investigate the automated fitting of bars and the application of Fourier mode fitting, built into the upcoming version of Galfit.

Another potential aspect for increasing the versatility of Galapagos could be the implementation of a variable PSF. Currently, just one PSF is used for convolving the Galfit model profiles for the whole survey. Instead, one could allow using a different PSF depending on the position on the tile or even varying tile by tile.

Galapagos is freely available for download from our webpage at: `http://astro.uibk.ac.at/~barden/galapagos/index.html`

Acknowledgements. MB was supported by the *Austrian Science Foundation FWF* under grant P18416. BH is grateful for support from the *Science and Technology Facilities Council (STFC)*. DHM acknowledges support from the *National Aeronautics and Space Administration (NASA)* under LTSA Grant NAG5-13102 issued through the *Office of Space Science*. CYP acknowledges support from the Canadian NRC-HIA Plaskett Fellowship and the STScI Institute/Giacconi Fellowship programs.

References

Astier, P., Guy, J., Regnault, N., et al. 2006, A&A, 447, 31
Balogh, M., Eke, V., Miller, C., et al. 2004, MNRAS, 348, 1355
Barden, M., Rix, H.-W., Somerville, R. S., et al. 2005, ApJ, 635, 959
Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, AJ, 132, 1729
Bell, E. F., Papovich, C., Wolf, C., et al. 2005, ApJ, 625, 23
Bell, E. F., Phleps, S., Somerville, R. S., et al. 2006, ApJ, 652, 270
Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393
Bundy, K., Ellis, R. S., & Conselice, C. J. 2005, ApJ, 625, 621
Caldwell, J. A. R., McIntosh, D. H., Rix, H.-W., et al. 2008, ApJS, 174, 136
de Jong, R. S. 1996, A&AS, 118, 557
de Souza, R. E., Gadotti, D. A., & dos Anjos, S. 2004, ApJS, 153, 411
Faber, S. M., Willmer, C. N. A., Wolf, C., et al. 2007, ApJ, 665, 265
Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, ApJ, 600, L93
Gray, M. E., Wolf, C., Barden, M., et al. 2008, ArXiv e-prints
Häussler, B., McIntosh, D. H., Barden, M., et al. 2007, ApJS, 172, 615
Heymans, C., Gray, M. E., Peng, C. Y., et al. 2008, MNRAS, 385, 1431
Infante, L. 1987, A&A, 183, 177
Jogee, S., Barazza, F. D., Rix, H.-W., et al. 2004, ApJ, 615, L105
Koekemoer, A. M., Aussel, H., Calzetti, D., et al. 2007, ApJS, 172, 196
Kron, R. G. 1980, ApJS, 43, 305
Lauer, T. R., Ajhar, E. A., Byun, Y.-I., et al. 1995, AJ, 110, 2622
Leauthaud, A., Massey, R., Kneib, J.-P., et al. 2007, ApJS, 172, 219
Liu, C. T., Capak, P., Mobasher, B., et al. 2008, ApJ, 672, 198
Lotz, J. M., Davis, M., Faber, S. M., et al. 2008, ApJ, 672, 177
McIntosh, D. H., Bell, E. F., Rix, H.-W., et al. 2005, ApJ, 632, 191
Moustakas, L. A., Casertano, S., Conselice, C. J., et al. 2004, ApJ, 600, L131
Nuijten, M. J. H. M., Simard, L., Gwyn, S., & Röttgering, H. J. A. 2005, ApJ, 626, L77
Parker, L. C., Hoekstra, H., Hudson, M. J., van Waerbeke, L., & Mellier, Y. 2007, ApJ, 669, 21
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266
Rix, H.-W., Barden, M., Beckwith, S. V. W., et al. 2004, ApJS, 152, 163
Sánchez, S. F., Jahnke, K., Wisotzki, L., et al. 2004, ApJ, 614, 586
Scoville, N., Aussel, H., Brusa, M., et al. 2007, ApJS, 172, 1
Sersic, J. L. 1968, Atlas de galaxias australes (Cordoba, Argentina: Observatorio Astronomico, 1968)
Sheth, K., Elmegreen, D. M., Elmegreen, B. G., et al. 2008, ApJ, 675, 1141
Shioya, Y., Taniguchi, Y., Sasaki, S. S., et al. 2008, ApJS, 175, 128
Simard, L., Willmer, C. N. A., Vogt, N. P., et al. 2002, ApJS, 142, 1
Smolčić, V., Schinnerer, E., Finoguenov, A., et al. 2007, ApJS, 172, 295
Vogt, N. P., Koo, D. C., Phillips, A. C., et al. 2005, ApJS, 159, 41
Wolf, C., Meisenheimer, K., Kleinheinrich, M., et al. 2004, A&A, 421, 913

Appendix A: Starting Parameters

In this appendix we give a detailed description of the starting parameters in the Galapagos startup file. Lines starting with "#" are treated as comments and are ignored by the code. Examples for SExtractor related setup files (items `B02`, `B03`, `B06` in the table below) can be found in the corresponding documentation (Bertin & Arnouts 1996). In order not to execute the blocks `B00`, `C00`, `D00` or `F00` the user should either replace "execute" with something else or simply comment out the respective line, e.g. "#B00) execute". Files in the table below without a directory descriptor can be found in the output directory defined for each survey image in the file list unless otherwise noted.

		File Locations
A00)	/path/to/survey/setup/gala_files	setup of the survey tiling (path and filename). For an example see Fig. 3
A01)	/path/to/survey/cat	output directory for catalogues. In this directory, the combined SExtractor catalogue (item B19); in ASCII format) and the final output catalogue (item F01); in FITS format) are placed
		SExtractor Setup
B00)	execute	execute the SExtractor block
B01)	/path/to/SExtractor-binary/sex	path and filename of SExtractor executable
B02)	/path/to/survey/setup/gala.param	path and filename of SExtractor output parameters in .param-format
B03)	/path/to/survey/setup/coldsex	path and filename of SExtractor setup file (cold)
B04)	coldcat	filename of SExtractor output catalogue (cold)
B05)	coldseg.fits	filename of SExtractor output segmentation map (cold)
B06)	/path/to/survey/setup/hotsex	path and filename of SExtractor setup file (hot)
B07)	hotcat	filename of SExtractor output catalogue (hot)
B08)	hotseg.fits	filename of SExtractor output segmentation map (hot)
B09)	1.1	factor by which the cold isophotes are enlarged when combining hot/cold catalogues
B10)	outcat	filename of combined SExtractor output catalogue
B11)	outseg.fits	filename of combined SExtractor output segmentation map
B12)	outparam	filename of SExtractor output parameter file
B13)	check.fits	filename of SExtractor check image
B14)	apertures	type of SExtractor check image
B15)	/path/to/survey/setup/gala_exclude	path and filename of list of "critical" detections (removed *before* the fitting). Set to non-existing file if not required
B16)	1.5	radius in pix used to exclude "bad" detections
B17)	all	if set "outonly": hot/cold catalogues/segmaps are deleted, else all files are kept
B18)	/path/to/survey/setup/gala_bad	path and filename of list of "catalogue" detections (removed *after* the fitting). Set to non-existing file if not required
B19)	sexcomb	filename of combined SExtractor catalogue. Output directory is A01)
		Stamp Setup
C00)	execute	execute the postage stamps creation block
C01)	stamps	output descriptor file for postage stamps. Per line, this ASCII file contains: SExtractor number, x/y source centre, x-range, y-range
C02)	v	filename preposition for postage stamps. E.g. for C02) = "v", a global file preposition "im1." (from file list) and SExtractor detection number "234", the output filename would be: "im1.v234.fits"
C03)	2.5	scale factor by which the SExtractor isophotes (Kron ellipses) are enlarged to calculate postage stamp size
		Sky Preparation
D00)	execute	execute the sky preparation block
D01)	skymap	filename of output object/sky-mapfile
D02)	outsky	filename of output list with sky values
D03)	3	scale factor by which SExtractor isophote is enlarged (for calculating the skymap)
D04)	1.5	scale factor by which SExtractor isophote is enlarged (for neighbouring source treatment)
D05)	20	Definition of sky isophotes: additional offset to scale factor (in pix), for sky measurement
D06)	30	Definition of sky isophotes: distance between individual sky isophotes
D07)	60	Definition of sky isophotes: width of individual sky isophotes
D08)	30	Definition of sky isophotes: gap between SExtractor isophote and inner sky isophote
D09)	3	cut below which objects are considered as contributing to the actual primary source
D10)	2	max number of allowed contributing sources per primary source
D11)	1.4	power by which the flux_radius is raised to be converted to a half-light radius
D12)	5	fraction of sources to be treated first (in %; "5"=5%), using multiple CPUs

D13)	15	calculate the slope of the sky from the x last determinations
D14)	-0.3	slope in FWHM_IMAGE vs. MAG_BEST below which an object is considered a star. Used for treating secondary sources
D15)	6.8	zeropoint in FWHM_IMAGE vs. MAG_BEST below which an object is considered a star. Used for treating secondary sources
D16)	5	magnitude faint end limit for secondaries when fitting galaxies. Objects more than x mag fainter than the primary are not included as secondary sources but tertiaries
D17)	2	magnitude faint end limit for secondaries when fitting stars. See also D16)
D18)	8	number of neighbouring tiles. See Sec. 4.6
D19)	4	maximum number of parallel processes for fitting the brightest sources, defined by D12)
D20)	360	minimum distance (in arcseconds) between all simultaneously fitted sources for D19). If current source in fitting queue is closer, fitting is delayed until other sources are done and the criterium is fulfilled
		GALFIT Setup
E00)	/path/to/galfit-binary/galfit	path and filename of GALFIT executable
E01)	/path/to/survey/setup/batchlist.XX	path and filename of a batch list, used for parallel processing of several tiles. For an example see Sec. 3
E02)	obj	object file preposition. E.g. for E02) = "obj", a global file preposition "im1." (from file list) and SEXTRACTOR detection number "234", the output filename would be: "im1.obj234"
E03)	v_gf	preposition for GALFIT output files. E.g. for E03) = "v_gf", a global file preposition "im1." (from file list) and SEXTRACTOR detection number "234", the output filename would be: "im1.v_gf234.fits". Note, GALFIT output files contain 3 FITS extensions: the original image, the model (including fit parameters in the header) and the residual image
E04)	/path/to/survey/setup/psf.fits	PSF filename including path
E05)	mask	mask file preposition used in GALFIT. E.g. for E05) = "mask", a global file preposition "im1." (from file list) and SEXTRACTOR detection number "234", the output filename would be: "im1.mask234.fits".
E06)	constr	constraint file preposition. E.g. for E06) = "constr", a global file preposition "im1." (from file list) and SEXTRACTOR detection number "234", the output filename would be: "im1.constr234".
E07)	257	size of PSF convolution box
E08)	26.486	magnitude zeropoint
E09)	0.03	plate scale of the images [arcsec/pixel]
E10)	706	effective exposure time (after image reduction, multidrizzling, etc.)
E11)	750	constraint: maximum allowed half-light radius
E12)	-5	constraint: minimum magnitude deviation (minus) from SEXTRACTOR measurement, i.e. the fit magnitude is constrained to not more than E12) mag brighter than the SEXTRACTOR value
E13)	5	constraint: maximum magnitude deviation (plus) from SEXTRACTOR measurement. See also E12)
E14)	notnice	use the UNIX facility "nice" when starting the fitting with GALFIT. Set E14) to "nice" to activate "nicing"
E15)	2.1c	GALFIT version string. E.g. 2.0.3c
		Output Catalogue Setup
F00)	execute	execute catalogue combination block
F01)	combcat.fits	filename of combined FITS output catalogue. Output directory is A01)