

GALAPAGOS-C: analysis of galaxy morphologies using high-performance computing methods

Andreas Hiemer,¹★ Marco Barden,¹ Lee S. Kelvin,¹ Boris Häußler^{2,3}
and Sabine Schindler¹

¹*Institute for Astro and Particle Physics, A-6020 Innsbruck, Austria*

²*Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK*

³*University of Hertfordshire, Hatfield, Hertfordshire AL10 9AB, UK*

Accepted 2014 August 11. Received 2014 August 11; in original form 2014 March 21

ABSTRACT

We present GALAPAGOS-C, a code designed to process a complete set of survey images through automation of source detection (via SExtractor), postage stamp cutting, object mask preparation, sky background estimation and complex two-dimensional light profile Sérsic modelling (via GALFIT). GALAPAGOS-C is designed around the concept of MPI-parallelization, allowing the processing of large data sets in a quick and efficient manner. Further, GALAPAGOS-C is capable of fitting multiple-Sérsic profiles to each galaxy, each representing distinct galaxy components (e.g. bulge, disc, bar), in addition to the option to fit asymmetric Fourier mode distortions. The modelling reliability of our core single-Sérsic fitting capability is tested thoroughly using image simulations. We apply GALAPAGOS-C to the Space Telescope A901/902 Galaxy Evolution Survey to investigate the evolution of galaxy structure with cosmic time and the dependence on environment. We measure the distribution of Sérsic indices as a function of local object density in the A901/902 cluster sample to provide one of the first measures of the Sérsic index–density relation. We find that the fraction of galaxies with a high Sérsic index ($2.5 < n < 7.0$) is higher in denser environments (~ 35 per cent), halving towards sparsely populated regions (~ 15 per cent). The population of low Sérsic index galaxies ($0.4 < n < 1.6$) is lower in denser environments (~ 35 per cent), increasing towards sparsely populated regions (~ 60 per cent). The population of intermediate Sérsic index galaxies ($1.6 < n < 2.5$) approximately follows the trend of the high Sérsic index types.

Key words: surveys – galaxies: evolution – galaxies: statistics – galaxies: structure.

1 INTRODUCTION

The modern era of extensive, wide-area photometric surveys has allowed for in-depth, robust studies of galaxy populations and their morphological properties on a hitherto unforeseen scale. Imaging surveys are usually composed of several image regions in one or more optical wavelength ranges to a variable depth and resolution and based upon ground or space data sets. Recent surveys include for example the Space Telescope A901/902 Galaxy Evolution Survey (STAGES; Gray et al. 2009), the Galaxy Evolution From Morphology And SED survey (GEMS; Rix et al. 2004), the Galaxy And Mass Assembly survey (Driver et al. 2009), the Sloan Digital Sky Survey (Abazajian et al. 2003; York et al. 2012) and the CANDELS survey (Grogin et al. 2011).

Galaxies come in a wide array of shapes and sizes, principally elliptical, lenticular, barred and non-barred spiral galaxies, grouping galaxies together according to the most common features, aligning these galaxies along the Hubble Tuning Fork (Hubble 1927). Massive elliptical galaxies are mostly characterized by a smooth, one-component profile, lenticular types contain a featureless disc and a spheroid component in the centre, the bulge, and spiral types consist of a bulge and a mostly blueish disc component. Additional components like bars, irregularities and profile truncations are common features, and transition states like unusual dusty red spirals (e.g. Wolf et al. 2003; Bösch et al. 2013), dusty ellipticals (Rowlands et al. 2012) or blue ellipticals (Schawinski et al. 2009) may also be observed.

It has long been known that this variety of galaxy morphologies is closely related to both the internal and external evolutionary processes a galaxy encounters during its lifetime. For example, observations show that the average properties of galaxy populations are strongly affected by their environment. One such example is

★E-mail: andreas.hiemer@uibk.ac.at

quantified by the morphology–density relation (Dressler 1980), which shows that the relative fractions of elliptical, lenticular and spiral galaxies is strongly correlated with the local environmental density they reside in. Different cluster-specific interaction processes such as ram pressure stripping, high-speed fly-bies, harassment, strangulation, starvation, amongst others can influence the morphological properties and the relative fractions of these populations. Galaxy clusters are locations where the hierarchical mass accumulation with time, as predicted by Λ cold dark matter, is most pronounced. Observations show that infalling galaxies are accelerated by the potential well of a galaxy cluster, often towards a massive (brightest) cluster galaxy (BCG). This cluster environment drives a variety of interaction processes affecting the morphology of the involved galaxies.

A prominent phenomenon affecting morphological change is the triggering and suppressing of star formation via ram pressure stripping (e.g. Steinhauser et al. 2012). Rapid motions of galaxies in a cluster environment through the intracluster medium (ICM) create ram pressure ($\sim \rho v^2$) on the gas in a galaxy (e.g. Quilis, Moore & Bower 2000). If this ram pressure is strong enough to overcome the galactic potential well of the galaxy itself and its potentially existing dark matter halo, the cold gas can be compressed and star formation can be induced on short time-scales (Steinhauser et al. 2012). On longer time-scales, the gas will be removed from the galaxy and star formation gets quenched, which has obvious implications on the emergent Hubble type. If the intensity of the interaction with the ICM is lower, galaxy strangulation or starvation occurs (e.g. Bekki et al. 2002) and only the thin gaseous halo around the galaxy is affected as it is gravitationally less strongly bound. This will usually lead to tidal effects that allow the gas contained in the galaxy to escape and will hence suppress further star formation.

Morphological distortions among neighbouring galaxies are often indicators for interaction processes due to spatially close encounters. Typical examples of such indicators are tidal tails, warps or asymmetric and lopsided galaxy features in general. Additionally, the emergence of galactic bars, usually created through disc instabilities or external triggering mechanisms (Athanasoulas, Machado & Rodionov 2013) amongst other mechanisms, can be used as an indirect indicator of soft environmental interactions. At high relative velocities in galaxy clusters, galaxy harassment (e.g. Gunn & Gott 1972; Moore et al. 1996) is the most meaningful process leading to short bursts of star formation and strong distortions in their morphology, e.g. asymmetries, tidal tails or even changes in the Hubble types of the involved galaxies. Especially in the denser core regions where the relative velocities for galaxy mergers are too high in order for galaxy merging to occur, the galaxies are exposed to successive fly-bies from other cluster members causing tidal forces in the galaxy potential well. During this process, many morphological properties of a galaxy can be radically disturbed and changed as it affects the stellar and gas distribution alongside its overall shape. In the case of spiral galaxies with relatively low mass, the faint, outer regions are more sensitive to soft interaction processes. In environments in which galaxies show lower relative velocities, e.g. at the outskirts of clusters, in galaxy groups or in simple galaxy pairs, galaxy mergers become more important as a transformation process. Galaxy–galaxy mergers are an efficient mechanism to produce spheroids from late-type galaxies (e.g. Toomre & Toomre 1972).

For example, Maltby et al. (2012) find that ~ 50 percent of the spiral galaxies in A901/902 show morphological distortions (truncations or antitruncations) in their outer disc regions with no environmental dependence, e.g. the effect seems to be driven by galaxy

internal processes. Due to the close relation to physical processes, the evolution of these galaxies is reflected in their morphological properties. In this context, galaxy light – and the morphological parameters derived from this – can be seen as a result of the interplay of different processes with different intensities that affect a galaxy over its lifetime.

Disentangling the morphological properties of a galaxy and its components is a non-trivial task and requires a statistical, but detailed, study of galaxy morphologies for a large number of objects with a range of histories, even when neglecting the distortions discussed above and only considering the basic properties like shape and size. A contemporary and increasingly popular means by which to tackle this subject is via the analysis of large, automatically produced data sets containing robust morphological parameters. Recently, several galaxy fitting algorithms and data pipelines have been presented, e.g. GIM2D (Simard et al. 2002), IMFIT (Erwin, private communication), SIGMA (Kelvin et al. 2012), BUDDA (Gadotti, de Souza & Dos Anjos 2003) or SExtractor 2.0 (Bertin 2011), GALFIT (Peng et al. 2010a) and GALAPAGOS (Barden et al. 2012).

GALAPAGOS, the software also discussed in this paper, allows a complete set of survey images to be analysed without user interaction, using several code-internal procedures. It utilizes SExtractor (Bertin & Arnouts 1996) for automated source detection, while ensuring that repeatedly detected objects from overlapping tiles are removed automatically. It then independently measures the local sky background and constructs both postage stamp images and object masks for every source. For detailed profile estimation via Sérsic (Sérsic 1963) profile fitting, GALFIT (Peng et al. 2010a) is applied to each object. Finally, an output catalogue is compiled, which contains the light profile information for all objects. We adopt GALFIT as our preferred 2D-model-fitting software as it has been shown to compare well in comparison to other fitting algorithms both in accuracy and speed (Häussler et al. 2007), allows for a wide range of light profile models, for example Sérsic (Sérsic 1963), Gaussian or exponential models amongst others and is generally easy to use, making it ideal for an automation routine like GALAPAGOS-C. The original version of GALAPAGOS was written in the Interactive Data Language (IDL). It includes most of the features presented in this paper and is able to process typically about 1000 sources per 24 h when performing single-Sérsic fits in case of typical *Hubble Space Telescope* (HST) surveys like GEMS/STAGES/COSMOS or similar on a single 2.2 GHz CPU (GALAPAGOS; Barden et al. 2012).

Even when using several processors simultaneously, the code has to abide a strict order in which to fit objects (from bright to faint), while avoiding simultaneous fitting of galaxies close together. This can occasionally lead to the code being ‘blocked’, waiting for the results of the brighter object to be returned, while allocated CPUs are able to fit a different, independent object in the meantime, hence increasing the overall time needed to fit a survey.

In this paper, we present a new version of GALAPAGOS, written/coded in C, which we call GALAPAGOS-C. This new version incorporates a complex sorting mechanism to handle dependences for neighbouring objects (similar to the one applied in Häußler et al. 2013). This new scheme enables parallel scheduling on a multitude of processors and ensures optimal usage of all available CPUs/cores at all times. The additional computational resources lead to an overall substantially increased computational speed, even when using the same number of processors, and allow for a more detailed insight into the sub-structures of Sérsic profiles, e.g. the decomposition of galaxy components (Simard et al. 2011), which uses additional CPU power. Additionally, GALFIT allows for Fourier mode fitting to quantify morphological distortions and asymmetries

of profiles that deviate from axisymmetric galaxy profiles that are assumed by most of the codes mentioned above. These deviations can carry valuable information on the galaxies morphologies and – more crucially – assembly history and should hence be exploited. With GALAPAGOS-C, this exploitation is possible in an automated fashion on all or most galaxies in a survey, offering a vast amount of additional information that was previously inaccessible.

In this paper, we highlight the new adaptations incorporated in this new code. In order to clearly separate the two versions, we use the term GALAPAGOS wherever we make reference to the original, IDL-based version of the code, while we refer to ‘GALAPAGOS in C’ (or GALAPAGOS-C) whenever we discuss this new, C-based version. Additionally, the focus of this paper is to present various tests of the GALFIT single-Sérsic fitting on large data sets, especially in comparison to the previous version of GALAPAGOS. We focus on *HST* space-based data provided by GEMS (in the *Chandra Deep Field South*) and STAGES (targeting the Abell 901/902 cluster group). This allows a clean and complete comparison as the tests of GALAPAGOS in IDL base on the same data set (Barden et al. 2012).

This paper is structured as follows. We start with an overview of the new, parallelized code in C and refer the interested reader to Barden et al. (2012) for further details of the philosophy of the code (Section 2). We describe the code input control via a series of set-up scripts (Section 3), details on the code components and its new features (Section 4) and describe the simulated data set (Section 5) which was used extensively during our extensive tests of the reliability of the fitting results. These are presented in Section 6 using artificial image simulations together with the speed-up performance of the code, carrying out single-Sérsic fits. Section 6 also presents a detailed comparison of the results of both GALAPAGOS and GALAPAGOS-C. A scientific application of the code is presented in Section 7 where the dependence of the average Sérsic indices of galaxies as a function of their environmental density is measured. We present an outlook and a summary in Section 8. The appendix provides further information on tests on the sky estimation (A), the stability of the point spread function (PSF, B) and the typical sources of error (C).

As this version of GALAPAGOS in C is based upon ‘GALAPAGOS: from pixels to parameters’ (Barden et al. 2012) and incorporates SExtractor (Bertin & Arnouts 1996) and GALFIT (Peng et al. 2010a), an established basic working knowledge of all these codes is assumed. Within this paper, we present the essential of the code framework, but focus on a detailed description of the new features implemented in GALAPAGOS in C. We refer the reader to the publications above for further details. GALAPAGOS-C is freely available for download from the astrophysical code library at <http://ascl.net/1408.011>.

2 OVERVIEW OF GALAPAGOS-C STRUCTURE

GALAPAGOS-C aims to keep the general structure of both the set-up file and the code as close to the original GALAPAGOS version as possible wherever this seems desirable, e.g. both set-up file and code structure are still arranged in several main blocks that are familiar to the experienced GALAPAGOS user. These blocks are arranged as follows.

- (i) Source detection using SExtractor, see Section 4.1 ‘Source detection’.
- (ii) Postage stamp cutting, see Section 4.2 ‘Postage stamps’.
- (iii) Background estimation, see Section 4.3 ‘Background estimation’.

- (iv) Preparation and execution of GALFIT, see Section 4.4 ‘GALFIT’.
- (v) Catalogue compilation for all objects.

The basis of GALAPAGOS-C is the detection of every astronomical light source in every survey image using SExtractor. Modern survey images usually feature a wide range of object sizes, shapes and luminosities and require SExtractor to be run several times with amplified set-up parameters correspondingly (high dynamic range mode). GALAPAGOS-C applies a high dynamic range (HDR) mode for SExtractor, which is well suited for handling both space-based (e.g. *HST*) and wide area data (see Section 4.1 ‘Source detection’). Spurious detections or overly deblended sources may be removed from the SExtractor catalogue using the bad detection list, if desired. After the SExtractor procedure, GALAPAGOS-C combines the catalogues from the individual images into one complete survey catalogue. It also sorts the contained objects according to the SExtractor magnitudes. Duplicate sources – resulting from overlapping images – are removed automatically, picking out the ‘best’ and most complete of all detections by choosing the detection with the greatest distance to the next image border.

For further analysis with GALFIT, a separate image region for every detected source, containing only the primary object including its direct potential neighbours is cut out from the image mosaic. This ‘postage stamp’ can be passed over to GALFIT in order to reduce the overall fitting time. GALAPAGOS-C automatically calculates the necessary size for every postage stamp image from the SExtractor parameters and some user input, e.g. keywords. This is done for every detected object in the output table (see Section 4.2 ‘Postage stamps’).

For detailed object profiling and sky estimation, GALAPAGOS-C computes a ‘skymap’ for every input image tile. This skymap provides information on the nature of every pixel, categorizing each into ‘bad pixel’, ‘source’ and ‘sky’, respectively. This map is used in the sky estimation step to identify sky pixels and for masking out spurious detections for the further procedure, e.g. in GALFIT.

Even though GALFIT has the ability to fit the sky simultaneously to the galaxy profile itself using the postage stamp image, GALAPAGOS-C does not make use of this option to avoid situations when the sky value by GALFIT might be biased by neighbouring contamination or by fit degeneracies with the wing structure of spiral arms. For this calculation, GALAPAGOS-C measures a sophisticated flux growth curve excluding forbidden pixels from the skymap (see Section 4.3 ‘Background estimation’).

After this preparatory work, the essential working step is to fit every astronomical object in the source catalogue individually using GALFIT (see Section 4.4 ‘GALFIT’). Precise fitting analysis by GALFIT deserves good initial parameter guesses to start the fit, the determination of a correct sky background and a robust identification of neighbours. GALAPAGOS-C executes the fitting jobs in a rank-ordered sequence beginning with the brightest objects in the survey and proceeding to the fainter sources. It often happens that neighbouring sources with a significant light influence overlap and influence each other. When a fit for the bright source already exists, GALAPAGOS can hold the fit parameters fixed at the best values and include the previously determined profile of the bright source for the fit of the fainter source in order to improve the fit.

With GALAPAGOS-C, this fitting procedure can be executed on modern supercomputers using a large number of CPU cores in parallel. For an infinite number of CPU cores, the user could theoretically distribute every individual fitting job to a separate CPU core and thus reduce the overall fitting time of the whole survey to the fitting time required by the most time-consuming single source. But for

CODE STRUCTURE OVERVIEW

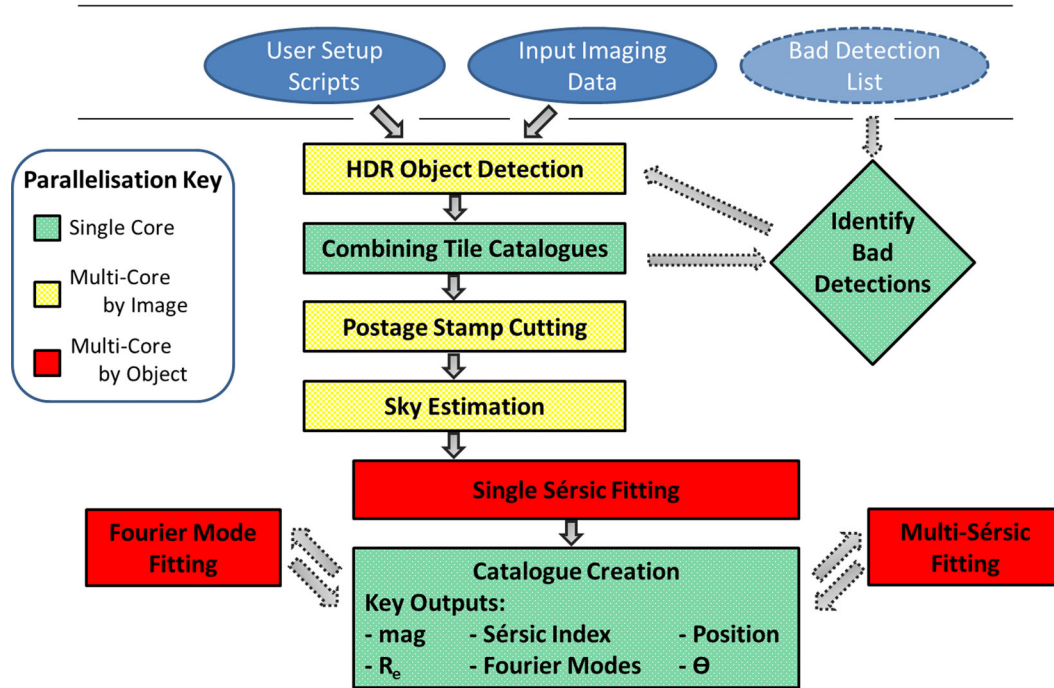


Figure 1. Block diagram of the code structure. The top shows the user input files (blue ellipses). Dashed boxes and arrows are optional features. Emphasis is put on parallelization: the yellow blocks are parallelized for every image, so can be carried out quickly on a multicore machine, each processor working on a specific image tile. Using more processors than image tiles would not speed up these steps any further. In the red blocks, GALAPAGOS-C is parallelized for every object. Here, a master processor distributes GALFIT jobs to slave processes working on a specific object (see Section 4.5 ‘Parallelization’). Combining catalogues and catalogue creation (green) are sequential blocks and must be done after the SExtractor and every GALFIT block.

practical implementation with a limited number of CPUs, it must be considered that overlapping sources are influencing each other and that fitting jobs are not independent. GALAPAGOS-C overcomes this problem by constructing a time-saving internal schedule: the fitting jobs are still distributed to the available CPU cores in a rank-ordered procedure according to their magnitudes. If a certain object is influenced by a brighter, neighbouring source and this brighter object is currently been fit, the fainter object is blocked and must wait until the fit for the brighter neighbouring source is finished. After that, the source will be unlocked and the existing fit parameters for the brighter, neighbouring source can be included as fixed parameters. In order to use computation resources economically, the CPU node which was allocated to fit the blocked source is provided with the next brightest independent source in the meantime. The blocked source might be processed on the first free CPU core, at the moment when it becomes unlocked. Due to the limitations of this procedure, it would in practice not make sense to use as many CPU cores as astronomical sources are present in the survey. For large, modern surveys, this is in any case not possible due to the vast numbers of sources and the user always has to find the most suitable compromise for the number of CPU cores. Thus, the speed-up will always depend on the density of the sources in the survey and the degree of mutual influence. The speed-up is shown to behave nearly linear up to a certain number of CPU cores. Tests on the speed-up behaviour of GALAPAGOS-C will help the user in finding the appropriate number of CPU cores (see Section 4.5 ‘Parallelization’).

Using modern supercomputers with a large number of CPU cores allows for integrating more sophisticated analytical galaxy profiles. The new version of GALAPAGOS-C integrates GALFIT features for fitting asymmetrical distortions by applying a Fourier mode expansion

on elliptical isophotes as well as fitting several galaxy components. These new additional fitting steps can be executed after single-Sérsic fitting (see Sections 4.4.4 ‘Fourier modes’ and 4.4.5 ‘Multicomponent fitting’). The tests, the applications and the results will be presented in different papers.

The last working step in GALAPAGOS-C is the compilation of a catalogue containing the fitting information for every astronomical source that was detected. GALAPAGOS-C automatically reads the headers from the GALFIT output image files. The fitting information for every object will be added to the source catalogue and stored in a FITS table.

The flexibility of GALAPAGOS-C for running certain segments of the code independently without having to repeat the entire procedure (see Fig. 2) is maintained and then carried out following the hierarchy of the separate code blocks indicated in Fig. 1.

Fig. 1 describes how the GALAPAGOS-C code blocks are executed and depend on each other according to the internal order, e.g. the sky estimation can be executed without existing postage stamps but not without the data from SExtractor. The advantage of this flexible code block construction is that the user does not have to rerun the whole GALAPAGOS-C procedure in the event of an error, or the user can add or change the code output after every block as desired. Notably, the user can run every code block with the desired number of necessary CPU cores (for details, see Section 4.5 ‘Parallelization’).

2.1 New features of GALAPAGOS-C

GALAPAGOS-C includes several new features. We upgraded the source detection with HDR SExtractor in order to refine object deblending

and to improve the application on data sets with a wide range of object sizes. Instead of using a two-step detection scheme, GALAPAGOS-C allows for making use of an arbitrary number of SExtractor set-up files (for details, see Section 4.1 ‘Source detection’).

Most importantly, GALFIT allows for fitting more sophisticated light profiles to galaxies, e.g. the measuring of morphological distortions with Fourier mode fitting or the decomposing galaxies into several galaxy components like bulges, discs or bars. GALAPAGOS-C provides several approaches to multicomponent fitting. The simplest approach is to fit a fixed number of Sérsic profiles to every source in the survey. This approach would leave it up to the user to decide whether a Sérsic component for a galaxy is meaningful or not. In addition, GALAPAGOS-C includes routines in order to automatically quantify the residual images and the meaning of additional Sérsic components to a fit. One routine is embedded in logical filters to quantify the meaning of additional Sérsic components automatically and thus automate the multicomponent fitting procedure. However, we have not carefully tested and quantified its output and thus do not provide this routine in the official version of GALAPAGOS-C. GALAPAGOS-C will include this feature in a future version, but as it has not been tested, it is not included in the current version. However, the actual multicomponent fitting does work, so a user could use this feature as long as he/she will analyse the results themselves.

Making use of the above-mentioned light profiles and applying GALAPAGOS on large survey data sets requires additional computational resources. Due to the large time consumption of GALAPAGOS in IDL when applying it to large survey data sets, this version of GALAPAGOS has been pushed to its limits. For this reason, GALAPAGOS has been re-written in C and is now MPI-parallelized in order to be executed on modern supercomputers. For the parallelization, GALAPAGOS-C uses a `master--slave` concept (for details, see Section 4.5 ‘Parallelization’). An additional new aspect thereby is that the workload of the `master` process is constructed to be as low as possible and the communication amongst the processors is highly optimized. As a result, GALAPAGOS-C can be applied for supercomputers with a large number of CPU cores without risking computational overhead.

An overview of the new features of GALAPAGOS-C is shown in the list below:

- (i) amplification of the HDR source detection mode to an arbitrary number of SExtractor set-ups,
- (ii) computational optimization of several routines, e.g. background estimation and sub-routines,
- (iii) parallelization and preparation for the application on large data sets,
- (iv) GALFIT Fourier mode fitting,
- (v) GALFIT multicomponent fitting,
- (vi) automated residual quantification and component analysis in combination with logical filters (in future version, currently being tested).

3 GALAPAGOS-C SET-UP AND CONTROL

3.1 GALAPAGOS-C set-up script

GALAPAGOS-C is controlled by a set of scripts similar to the original version of GALAPAGOS. An example start-up script for GALAPAGOS-C is shown in Fig. 2. This start-up script controls the execution of GALAPAGOS-C and consists of six parts (I–VI), similar to the code

components described in Section 2. Each set-up parameter is described by a keyword and explained using a brief comment.

The first set of parameters is the block I ‘file locations’. This block is always mandatory for the execution of GALAPAGOS-C and can hence not be disabled. It controls the input and output file locations. The other five set-up blocks reflect the program blocks as described in Section 2. These blocks are dynamic in a similar fashion to the former original version of GALAPAGOS. They can be activated or skipped according to the user’s particular setup of GALAPAGOS-C code blocks as indicated in Fig. 2. Skipping a certain code block can only work if a code block that it depends on has been executed already and the output data exists, e.g. skipping block II when executing block III is only possible if the SExtractor output data exist.

With the set-up block II ‘HDR SExtractor Setup’, HDR source extraction can be activated. Depending on the linked number of ‘.sex’ setup files in the GALAPAGOS-C start-up script, the HDR mode of SExtractor will be automatically activated if more than one link is provided. The number of SExtractor catalogue files and segmentation image files must be consistent with the given number of SExtractor set-up files. The user must note that every SExtractor run is controlled by a set-up script as well, which depends on additional files. SExtractor requires a neural network file ‘default.nnw’ and an optional convolution filter, e.g. ‘block3x3.conv’, which have to be specified by the user in the SExtractor set-up scripts. For a detailed description on SExtractor and the SExtractor set-up, see Bertin & Arnouts (1996).

The set-up block III ‘Postage Stamp Cutting’ defines postage stamps; the block IV ‘Sky Preparation Block’ prepares a skymap and starts the background estimation procedure using the parameter configuration. The set-up block V ‘GALFIT Setup’ defines the fitting procedure of the astronomical sources. With this version of GALAPAGOS in C, the user can activate GALFIT single-Sérsic fitting, GALFIT Fourier mode fitting and GALFIT multi-Sérsic fitting. GALFIT Fourier mode fitting and GALFIT multicomponent fitting can work in combination with GALFIT single-Sérsic fitting or as stand-alone routines only if the GALFIT single-Sérsic fitting data exist. The last set-up block VI ‘Output Catalogue Setup’ defines the coordination number of the survey geometry to clean the catalogue from multiple detections caused by overlapping tiles and the filename of the output catalogue. Duplicate detections are cleaned after the detection when the HDR source detection block has been executed.

Usually, an astronomical survey is split into several image tiles and GALAPAGOS-C is constructed to include exposure time maps (weight images) which can be associated with science images. GALAPAGOS-C requires a list containing the file location of the science images, the weight images, a link to an output directory and a file preposition for every image tile. This information must be contained in an ASCII file as shown in the example in Fig. 3. In this figure, we give an example for the framework of a hypothetical survey which consists of three image tiles.

3.2 Running GALAPAGOS-C on a supercomputer

Modern supercomputers usually request a submission script including information on the execution queue, the amount of memory which is required by every node and the number of CPU cores. This information is usually provided in the user manual of the supercomputer. GALAPAGOS-C is started by submitting the script to the supercomputer queue. The execution command consists of the ‘mpirun’ command, the specified number of CPU cores (`-np N_{CPU}`), the executable name of GALAPAGOS-C and the path to the GALAPAGOS-C

```

#===== I FILE LOCATIONS =====
galafile:      /home/Gala/Setup/gala.files  # file containing image and directory path
outdir:       /home/Gala/cat              # directory for combined SExtractor catalogue
#===== II HDR SExtractor SETUP =====
dosex:        y                          # execute the SExtractor block?
sexexe:       /usr/local/bin/sex          # SExtractor executable path
setup:        /home/Gala/Setup/cold.sex   # setup filename(s) for cold SExtractor
\             /home/Gala/Setup/hot1.sex  # hot1 SExtractor setup
\            /home/Gala/Setup/hot2.sex   # hot2 SExtractor setup
cat:          /cold.cat                  # SExtractor catalogue file(s)
\            /hot1.cat                   # hot1 SExtractor catalogue file
\            /hot2.cat                   # hot2 SExtractor catalogue file
seg:          /coldseg.fits              # segmentation image(s)
\            /hotseg1.fits               # hot1 SExtractor segmentation image
\            /hotseg2.fits               # hot2 SExtractor segmentation image
sex_par:      /home/Gala/Setup/default.param # SExtractor parameters in .param format
zerop:        26.486                     # zeropoint
enlarge:      1.1                        # enlarging factor for ellipses to exclude hot objects
out_par:      /home/Gala/Setup/out.param  # SExtractor output parameters
out_seg:      /out_seg.fits              # output segmentation image
out_tab:      outtab.fits                # SExtractor output table name(per tile/
chkim:        yes                        # check images ("yes"/"no")?
check:        check.fits                 # check image
chktype:      APERTURES                  # check image modulus
#===== III STAMP SETUP =====
dostamps:     y                          # do stamp block?
stampfile:    stamps                    # descriptor file for postage stamps
stampname:    s                          # preposition for postage stamps
stampfac:     2.5                        # sizefactor for postage stamps
#===== IV SKY PREPARATION BLOCK =====
doskymap:     y                          # execute the skymap block
size:         3                          # enlarging factor for KRON ellipses
offset:       20                         # !efinition of sky isophotes: additional offset scale factor
skymapname:   map.fits                  # skymap name
dosky:        y                          # execute sky preparation block AND GALFIT
outsky:       outsky                    # filename for sky parameters
nslope:       15                         # number of measuring points for slope calculation
bandsize:     60                         # definition of sky isophotes: individual width
isodist:      30                         # definition of sky isophotes: !istance between isophotes
startrrad:    1.                         # enlarging factor for SExtractor KRON ellipses to start procedure
expand:       3                          # enlarging factor for KRON ellipses
#===== V GALFIT SETUP =====
dogalfit:     y                          # execute the GALFIT block?
dofourier:    y                          # execute GALFIT fourier modes?
ncomponents:  1                          # number Sersic components
path:         galfit                     # GALFIT executable including path
object:       gf.obj                     # filename preposition for GALFIT objects
galfitset:    GALFIT-Setup               # filename prefix for GALFIT setup files
psf:          /home/Gala/psf.fits         # PSF filename including path
psffactor:    1                          # PSF fine sampling factor relative to data
badpix:       mask                       # bad pixel mask (FITS image or ASCII coordinates list)
constraints:  constraint                 # file with parameter constraints (ASCII file)
re_max:       750                        # constraint max Re (half-light radius)
mag_min:      -5                         # constraint min magnitude deviation (minus)
mag_max:      5                          # constraint max magnitude deviation (plus)
convbox:      257                        # size of PSF convolution box
platescale:   0.03                       # platescale of the images [arcsec/pixel]
display:      regular                    # display type (regular, curses, both)
secenlarge:   1.5                        # scale factor by which isophote is enlarged (for neighbours)
magexcl:      5.                         # magnitude faint end limit for secondaries when fitting galaxies
fluxextend:   1.4                        # power for FLUX_RADIUS to be converted to a half light radius
distmax:      30.                        # maximum radius to search for contrib targets (IN ARCSEC!)
magcrit:      3.                         # magnitude cut below which objects are considered as contributing
contribmax:   2                          # maximum number of contributing targets
#===== VI OUTPUT CATALOGUE SETUP =====
docat:        y                          # execute catalogue combination block
neighbortiles: 8                        # maximum number of neighboring tiles
sexcomb:      cat.fits                  # filename for output catalogue in A

```

Figure 2. Example for a GALAPAGOS-C start-up script.

```

/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.

```

Figure 3. Example for the file location list required by the GALAPAGOS-C start-up script. This list defines the file names for the survey science images, the corresponding weight images (exposure time maps), the output directories for every GALAPAGOS-C file (e.g. sky value, GALFIT set-up script) for every object and the output file preposition, a total of four columns. This example would define a survey with three tiles.

set-up file as arguments in quotation marks. Usually, GALAPAGOS-C is started on a supercomputer using a command similar to

```
mpirun -np <NCPU> ./GALAPAGOS -s './.../gala.setup'.
```

Note that the specification command for the number of CPU cores (N_{CPU}) depends on the supercomputer and on the survey/data set.

4 SOFTWARE COMPONENTS OF GALAPAGOS-C

In the following section, we describe the new code components of GALAPAGOS-C in detail. For the traditional code components, which we simply translated from IDL into C, we repeat the most important aspects and refer the reader to Barden et al. (2012) for a detailed description. First, we focus on the adaptations of HDR SExtractor (Section 4.1) and postage stamp cutting (Section 4.2). In Section 4.3, we briefly repeat the concept of background estimation and describe the meaning of the necessary set-up parameters and the code adaptations in GALAPAGOS-C. Next, we describe the new features of GALAPAGOS-C, namely two additional code components: we upgraded the application of GALFIT with an optional Fourier mode expansion to quantify the intensity of the morphological distortions on the elliptical isophotes of the Sérsic profiles (see ‘Fourier modes, Section 4.4.4). Additionally, the new version of GALAPAGOS-C is much faster, as the fitting procedure with GALFIT is entirely parallelized and can thus be applied on modern supercomputers. This will be described in Section 4.5 (‘Parallelization’).

4.1 Source detection

GALAPAGOS-C uses the same approach as GALAPAGOS in detecting astronomical objects on the survey images, namely it employs SExtractor for this task. The embedding of SExtractor in GALAPAGOS is described in detail in Barden et al. (2012, section 3.1).

For the detection of astronomical sources, SExtractor identifies flux peaks above the sky background. It measures the background values within a grid locally or globally around an astronomical source. With a separate set-up script for SExtractor, the user defines a minimum number of connected pixels. If the flux in such a connected image region exceeds a certain limit above the sky background, the object is detected as a light source and enters a source catalogue. SExtractor includes an internal algorithm to add the flux of a detected object in the surrounding area either to the detected source or to declare it as a stand-alone object. The user can influence the deblending procedure, with a number of set-up options. A useful description of SExtractor is given by the manual in Bertin & Arnouts (1996).

SExtractor has previously been successfully applied to both ground- and space-based data sets (e.g. Kelvin et al. 2012; Häussler et al. 2007). Yet, modern large-area surveys with a tremendously high resolution highlight a limitation of SExtractor. In these surveys, galaxies in the near and distant Universe, highly resolved sources and very faint objects are observed at the same time, testing the limitations of any object detection algorithm. Fine-tuning the code to detect even the faintest and smallest objects can only be done at the cost of splitting up the brightest and most structurally complex sources.

For large surveys with a wide range of object luminosities and sizes being observed at the same time, GALAPAGOS-C allows for a sophisticated approach, called ‘HDR SExtraction’ mode. This mode is based on the original two-step SExtractor implementation in GALAPAGOS (see also fig. 2 in Barden et al. 2012). The concept of us-

ing a normal single-stage SExtractor configuration or combining a catalogue within the two-stage HDR SExtraction mode as was previously done (Barden et al. 2012) has been extended to a user-defined number of intermediate SExtractor set-up configurations.

Our detection procedure remains analogous to the concept of ‘cold’ and ‘hot’ mode (see section 3.1.1 in Barden et al. 2012, for details). In a first step – the so-called coldest mode – only the brightest sources will be detected at the cost of missing the faint sources. With each ‘hotter’ configuration of SExtractor, more and more emphasis must be put on depth and object deblending. The combination of these deeper SExtractor catalogues into the final output catalogue is done iteratively after each hotter SExtractor run according to the same concept as described in Barden et al. (2012), section 3.1.1, each time combining the already-existing detection catalogue (as ‘cold’) with the newer one (as ‘hot’). At each stage, GALAPAGOS-C automatically adds new object detections found in the ‘hotter’ SExtractor runs, creating the ‘cold’ catalogue for the next iteration. Spurious detections are erased automatically if they are located inside the Kron ellipse of the ‘colder’ detection. The user has the possibility to artificially enlarge every Kron ellipse by a scaling factor using the parameter ‘enlarge’ in the master set-up file.

If the user activates the catalogue compilation procedure using the keyword ‘docat’ in the GALAPAGOS-C set-up control file, GALAPAGOS-C will automatically compile a combined source catalogue containing all sources in the whole image survey. If image tiles might overlap and objects might be detected multiply, GALAPAGOS-C will automatically remove multiple entries from the combined catalogue. In the final step of GALAPAGOS-C, the resulting source catalogue will be completed with the GALFIT output parameters from single-Sérsic model and (optionally) the Fourier mode fitting procedure and the multicomponent fitting results.

4.2 Postage stamps

For each detected source, GALAPAGOS-C cuts a smaller section, centred on the object, from the science image. This reduces the number of pixels and thus the degrees of freedom for the fitting procedure with GALFIT. Excluding a large fraction of free sky in the science image and using only the relevant areas of the image leads to a significant optimization of the overall fitting time. Additionally, the creation of masks for excluding neighbouring sources without any significant light influence (tertiaries) is accelerated. To obtain an appropriate fit result, the directly influencing neighbouring objects (secondaries) must be visible on the postage stamp image to be included into the fitting procedure. If an influencing ‘contributor’ (for a definition see Section 4.3 ‘Background estimation’) is not entirely visible on the postage stamp, GALAPAGOS-C solves this problem by including a fixed Sérsic profile of the contributor which was measured before (for details on the scheduling of the fitting procedure, see Section 4.5 ‘Parallelization’). If a secondary is only partially visible on a postage stamp image, GALFIT is able to fit truncated profiles. GALAPAGOS-C uses the size of the SExtractor Kron ellipses to calculate the size of the postage stamps as a rectangular area around the Kron ellipse. The set-up parameter `stampfac` allows us to enlarge the rectangular area by the scalefactor. This scaling factor is a compromise between using a big area for the fit that includes the central primary source and all relevant neighbours and a small area to exclude as many unessential pixels – including irrelevant neighbours – as possible in order to shorten the computation time of GALFIT. Even if relevant neighbouring source are not entirely present on the postage stamp image, GALAPAGOS-C will

automatically include the necessary Sérsic profiles for these sources (see Sections 4.4.1 ‘Simultaneous fits’ and 4.5 ‘Parallelization’) by placing the centre of these objects outside the fitting image. A scalefactor of 2.5 was found to work well for STAGES and GEMS one-orbit *HST* surveys (see Barden et al. 2012).

4.3 Background estimation

Measuring a precise sky level is fundamental for accurate object profile fitting with GALFIT (see e.g. de Jong 1996; Häussler et al. 2007). An internal sky estimation in GALAPAGOS-C is carried out equivalently to the background estimation procedure presented by Barden et al. (2012, section 3.4). It was translated into C, computationally optimized and several new concepts of GALAPAGOS-C were implemented additionally (e.g. ‘potential contributors’, see below).

Even though GALFIT is able to fit the background sky on a postage stamp as a free parameter, GALAPAGOS-C incorporates its own routine to estimate a robust background level for every object on the whole image tile. This is necessary to overcome the risk of biased fit results and fit degeneracies between the wings of the Sérsic profile and the sky itself. Contaminations can often occur in crowded fields, especially in the neighbourhood of objects with a large and slowly decreasing light profile (e.g. big ellipticals). Additionally, a situation might arise in which not enough sky pixels are left on the postage stamp image, making it impossible for GALFIT to find a good estimate for the sky. In order for GALFIT to be able to get the sky level right, it would require a huge postage stamp, severely slowing down the speed of the fits. GALAPAGOS-C, however, can compute the sky in the surrounding regions on the original image tile separately, hence using a much larger area, if necessary. Holding the sky fixed at this derived values will significantly reduce the overall fitting time in GALFIT. This also avoids a huge possible degeneracy between the wings of the Sérsic profile and the sky itself, as it will likely be introduced when fitting the sky with GALFIT on a small postage stamp. This effect is possibly easiest to see in the case of magnitudes.

A wrong sky value results in the assignment of a wrong amount of flux to the object, hence biasing the magnitude outcome of the fit, but similar effect exists for other parameters, especially for the half-light radius r_e and the Sérsic index n . Not fitting the sky removes that degeneracy.

GALAPAGOS-C incorporates its own sub-routine to calculate a robust sky level for each source before the fitting, keeping it always fixed to the best values. The GALAPAGOS values are shown to improve significantly over the SExtractor values (Barden et al. 2012). GALAPAGOS-C includes a modified copy of the working principle from the IDL version of GALAPAGOS. We adopt a flux growth method to calculate a precise local background sky level for every object (see Fig. 4). GALAPAGOS-C measures the average flux in elliptical azimuthally averaged rings centred on the position of the primary object to calculate the background flux as a function of distance to the source, while excluding both image defects and other sources. GALAPAGOS-C calculates the distance to the object where the flux contribution levels off. It then uses the last few annuli at this radius to determine the background sky level.

To mask neighbouring sources and image defects, GALAPAGOS-C uses the sky map calculated in step III. This map defines the nature of every pixel, where a value of 0 indicates a blank background sky pixel, a value of 1 or higher indicates a masked object and a value of -1 is an invalid pixel with zero exposure time (for further details, see Barden et al. 2012). The masking of objects in the sky map is based on the SExtractor Kron ellipses. The user can enlarge the size of every Kron ellipse using the scalefactor *size* (for STAGES and GEMS, *size* = 3) in the GALAPAGOS-C set-up file in order to use a more conservative sky estimation at larger radii. However, a user should be careful in densely populated fields to not force the radii to be too large e.g. to measure a sky value so far out that its values become unrelated to the primary object. Additionally, the value *offset* (for STAGES and GEMS, *offset* = 20) allows for adding a fixed number of pixels to enlarge the Kron ellipse.

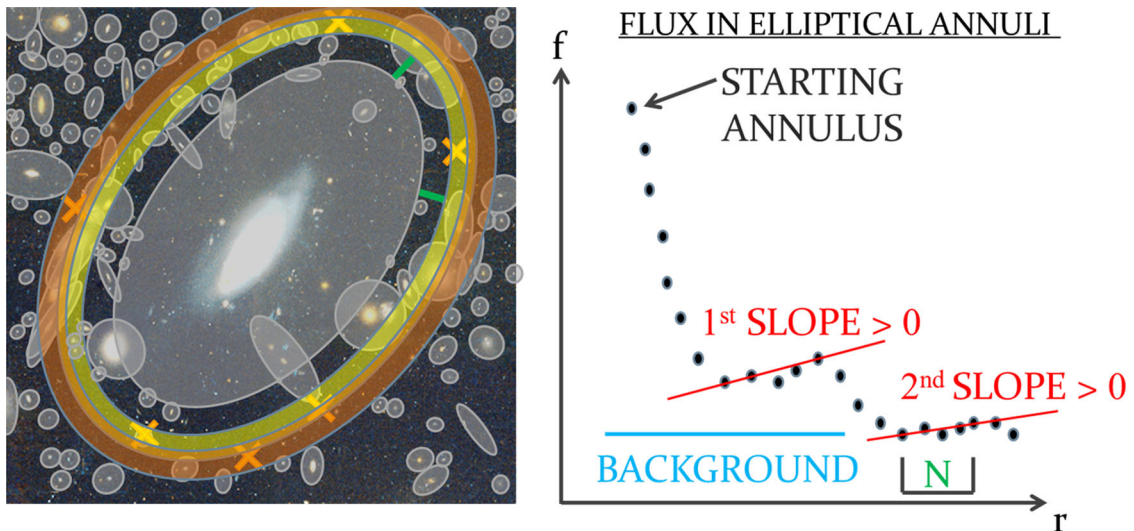


Figure 4. Sky estimation. Left: the average sky background values are measured in elliptical rings centred around an object (yellow and orange). The grey regions indicate regions which are excluded for the calculation for every source according to the sky map. The yellow crosses indicate outlier pixels, which were excluded from the calculation by clipping 3σ outliers from a Gaussian distribution. The distance indicated by the green bar shows the constant offset distance to the enlarged Kron ellipse. Right: the azimuthally averaged flux in every elliptical ring is plotted as a function of radius r . The first measuring point was obtained by the innermost circle at the starting radius. With successively larger radii, the flux contamination by the galaxies light profile decreases and once the rings are sufficiently large, the average slope randomly changes its sign. The iteration is stopped when the slope gets zero or positive for a second time. The measurements are computed using the last N sky estimates to obtain the BACKGROUND level.

For the calculation of the flux growth curve, the user must configure the geometry of ellipses within which the background flux is measured. The width of the rings can be specified using the parameter `bandsize` (for STAGES and GEMS, `bandsize` = 60 pixels). The distance between the rings is defined with the parameter `isodist` (for STAGES and GEMS, `isodist` = 30 pixels). The Kron ellipse can be expanded using the factor `expand` (for STAGES and GEMS, `expand` = 3, see Fig. 4 for illustration). It is thus possible – and recommended – to use overlapping annuli to refine the process without losing a significant amount of computation time and to increase the signal-to-noise ratio in each ring, while avoiding making the rings too big. GALAPAGOS-C internally calculates the average flux mean in elliptical annuli as a function of radius. A regression line is calculated according to the mean sky values covered by the last `nslope` annuli. GALAPAGOS-C measures the slope of the regression line (see Fig. 4, right). The parameter `nslope` (for STAGES and GEMS, `nslope` = 15) defines the number of measuring points for the calculation of the regression line. If the slope is 0 or positive for a second time, the background value for the source is calculated as the mean value from the average background sky values in the measuring points. The procedure allows for one positive slope to overcome possible internal fluctuations caused by e.g. a possible ring structure or a missed neighbour. Using a slope change higher than the second one (e.g. three slope changes) however has not been found to be necessary in practice.

The background estimation procedure can be affected by bright neighbouring sources around the current primary object. To that end, GALAPAGOS-C calculates the light flux influence of every source around a certain radius at the central position of the object. This radius must be specified using the parameter `distmax` (in arcsec) in the GALAPAGOS-C set-up script. From examining the most influential source in the whole survey, we find a value of 30 arcsec for the *HST* surveys STAGES and GEMS. Those numbers will strongly vary for other surveys e.g. through the presence or absence of bright, big foreground galaxies.

For single-Sérsic fitting, GALAPAGOS-C adopts the shape information given by SExtractor. For later fitting procedures (e.g. multi-component fitting), GALAPAGOS-C makes use of the already existing single-Sérsic fit profiles wherever possible, and will thus be able to work more accurately. At this point of the code procedure, however, this information is incomplete due to missing Sérsic fits for all sources, so GALAPAGOS-C must assume a conservative scenario with a high-light contribution. The light flux influence is calculated under the worst-case assumption that all sources in the neighbourhood have a high Sérsic index of 4 and a half-light radius calculated from the SExtractor flux radius using a conversion exponent (see Fig. 7). The parameter `fluxextend` allows the user to define an exponent α to convert the SExtractor flux radius to an empirically ‘true’ half-light r_e radius according to $r_e = (\text{flux_radius})^\alpha$. Barden et al. (2012) found $\alpha = 1.4$ as a good approximation for typical *HST* surveys. To identify a neighbouring source as a ‘potential contributor’, the calculated flux has to exceed user-specified values, given by the parameter `magcrit`, at the central position of the current object. In this case, the fit for the current object will be blocked until a fit profile exists for the contributing source and the calculation will be redone with the existing fit parameters. Only if the flux influence still exceeds the value given by the user parameter `magcrit`, the neighbouring source is identified as a ‘(confirmed) contributor’. In this case, the calculation of the background sky value will be redone, subtracting the single-Sérsic profile of the contributing source from the original image temporarily for the calculation and the remaining pixel values are used for the sky calculation. Even though the

contributor is still masked out from the image using the sky map, GALAPAGOS-C this way avoids situations when a contributor might not be masked completely from the image. In the worst case, after a certain number of iterations, the area of the growing annuli can overlap with the remaining non-Sérsic sub-structures of the contributor (e.g. its spiral structure). These sub-structures in the residual image can cause fluctuations in the slope of the flux growth curve and thus end the procedure at the outer regions of the contributor. However, these fluctuations – except potential tidal features – are most likely in the centre of the objects and are generally found to be masked out.

In this procedure, a fainter object can never be identified as a potential contributor to a brighter source and thus block the brighter source. With the set-up parameter `contribmax`, the user can define the maximum number of contributing sources in order to exclude the theoretical possibility of too many contributors. This is as a feature to have full control over the maximum number of secondaries for all objects to avoid situations where CPU cores might be overloaded with exceedingly time-consuming fits. If more than the specified number of contributing sources are identified, only the most influential objects will be included in the calculation. For STAGES and GEMS, we found only 3 objects out of $\sim 80\,000$ with more than 2 contributors using the given set-up parameters (for STAGES and GEMS, we use `contribmax` = 3).

Additionally, the calculation of the sky values in GALAPAGOS-C is significantly accelerated. For extended galaxies, the number of rings can get large with a large number of pixels and thus the calculation can consume a large amount of time. To that end, if the computation time for a ring gets longer than necessary due to exceeding a maximum number of pixels, GALAPAGOS-C will only look at a fraction of the pixels in the ring for the calculation. This fraction is calculated by a code internal procedure in order to provide enough pixels for a robust mean value of the sky. It will randomly select enough pixels in a ring without counting pixels twice and thus still provide robust values. GALAPAGOS-C provides GALFIT with the first three numbers after the decimal point. Testing the sky estimation routine in GALAPAGOS-C, we found that this hard-coded sky value in the GALFIT set-up file has not been varying significantly anymore using more than $\sim 10^5$ pixels in a ring. GALAPAGOS-C uses internal routines to take care for robust number statistics for the number of pixels in the rings. For this test, we ran GALAPAGOS-C 10 times of a data set provided by the STAGES image in tile 2 which contains 1062 detected objects. Wherever the rings are large enough, GALAPAGOS-C will self-check to provide robust number statistics. If the ring sizes can get insufficiently large due to inappropriate parameters in the user set-up file, GALAPAGOS-C provides a warning message. Additionally, as done in GALAPAGOS, point-like sources, e.g. cosmic rays, are excluded from the sky estimation by using a robust 3σ clipping function when determining the average values within each ring.

In contrast to GALAPAGOS, the sky estimation procedure is executed as a stand-alone block, independently of the GALFIT block. The sky estimation is only redone for objects with a confirmed contributor in the GALFIT block and the former background value is overwritten in this case. This is necessary in order to overcome biased sky values due to the large extent of the contributor’s light influence. When the fit profile of the contributor exists, GALAPAGOS-C can make use of this profile in order to subtract it from the original image for the sky calculation of the fainter neighbour. Including this procedure is shown to provide more accurate sky values, which are fundamental in providing accurate fit results. For details on the performance of the sky estimation routine, see Appendix A. In GALAPAGOS-C, the sky

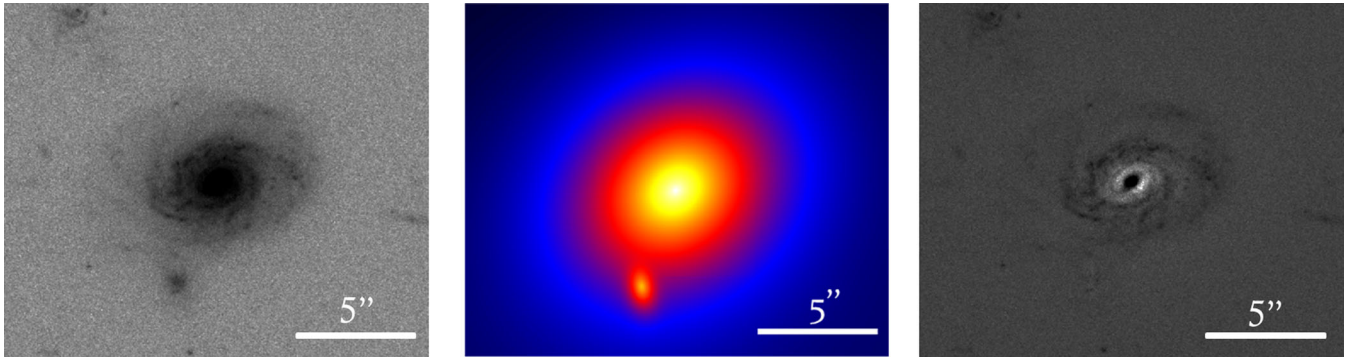


Figure 5. Single-Sérsic modelling with GALFIT. The left-hand panel shows an example of a galaxy image (primary source), which is fitted with a single-Sérsic profile (centre). One neighbouring source with a significant influence (secondary) was detected and thus included in the fitting procedure. The right-hand image shows a residual image for the whole fitted region. Details on the definition of primary, secondary and tertiary objects are given in Section 4.4.1 (‘Simultaneous fits’).

estimation procedure is executed in parallel for every survey image. This means that every available CPU node is provided with a survey image in order to estimate local sky values for every detected source (see also Section 4.5 ‘Parallelization’). Using more CPU nodes, that survey images would not make sense for this code block.

4.4 GALFIT

The essential and computationally most time-consuming step in GALAPAGOS is fitting every astronomical source with GALFIT. All the previous work within GALAPAGOS-C is exclusively done to start GALFIT with the best possible set-up file for every detected source. GALFIT incorporates a series of mathematical functions to describe light profiles of galaxies and galaxy components. The Sérsic model (Sérsic 1968) is the most general profile and is used by GALAPAGOS-C to quantify every detected source:

$$\Sigma(r) = \Sigma_e \exp\{\kappa(n)[(r/r_e)^{(1/n)} - 1]\}. \quad (1)$$

The Sérsic function generally describes galaxy profiles with Σ_e being the effective surface brightness, r_e the effective (or half-light) radius to describe the surface brightness $\Sigma(r)$ as a function of radius r . The most important parameter however, is the Sérsic index n with $\kappa(n)$ as a normalization constant. With this parameter, the Sérsic profile is a generalized/transitional profile that can describe a de Vaucouleurs profile (with $n = 4$) or an exponential profile (with $n = 1$) for different types of galaxies.

Single-Sérsic fitting is a core component of GALAPAGOS-C to quantify galaxies in the same way as before, but only as an initial approach. Multiplied computational resources (see ‘Parallelization’, Section 4.5) allow us to make use of several light profiles simultaneously in order to perform e.g. bulge–disc decomposition. Using additional light profiles relative to the single-Sérsic model permits a deeper insight into the morphological sub-structure of galaxies and asymmetric features (see ‘Fourier modes’, Section 4.4.4). A superposition of these models is used to identify the morphological components and their geometries – e.g. for bulge–disc decomposition.

The GALFIT block within GALAPAGOS-C adopts the same concept of primary, secondary and tertiary sources as the original version of GALAPAGOS; neighbouring sources are included or excluded according to the same concepts, bad pixel masks and parameter constraints are used in the same way. In its concept, this part of the code is a one-to-one translation of GALAPAGOS into a faster computing environment, but realization in a parallel computation environment on

a supercomputer requires additional internal scheduling (see ‘Simultaneous fits and ‘Parallelization’, Sections 4.4.1 and 4.5). Fig. 5 provides an illustration for the single Sérsic fitting with GALFIT. The spiral galaxy in the centre of the left panel is the primary source and the smaller, neighbouring galaxy below is included simultaneously into the fitting as a secondary. The middle panel shows the single Sérsic profiles for both galaxies. In the residual image on the left, both images are subtracted in order to highlight the elements of the galaxy, which have not been approximated adequately via single Sérsic fitting.

4.4.1 Simultaneous fits

It regularly occurs that sources are influenced by secondary light profiles, because they are close together. To obtain the best values for the fit parameters for both sources, it is necessary to include neighbouring sources into the fit of either source simultaneously (Barden et al. 2012, section 3.5.1). We define the target object for the current fit as the **primary** source. Objects with a direct overlap of the expanded Kron ellipses with the primary are **secondary** sources and always included in the fit procedure of the **primary** source simultaneously. Objects with no influence and no overlap with the **primary** source are **tertiary** sources. They are not regarded to be important for the fit of the primary source and thus excluded from the fit using a GALFIT pixel mask (for details on the pixel mask, see Section 4.4.2 ‘Object masking’ and Barden et al. 2012, section 3.5.2).

Additionally, there exist objects with a significant influence on the primary source without any overlap of the expanded SExtractor Kron ellipses due to a light contribution. For the identification of these sources, the code follows the same criteria as described in Section 4.3 ‘Background estimation’ and Section 4.5 ‘Parallelization’. If the value of the flux contribution of a neighbouring object at the central position of the primary object exceeds the user-given value defined with the parameter `magcrit` in the GALAPAGOS-C set-up file, the source will be identified as a so-called contributor. GALAPAGOS-C will search for contributors in a circle around the primary source. The user has to define the radius of this circle using the set-up parameter `distmax`. GALAPAGOS-C will handle such a contributor as a secondary and thus include it into the fit of a primary source. Barden et al. (2012) show that the best-fitting values can be obtained by fitting the sources in a strict order according to their magnitudes, beginning with the brightest source in the entire survey and proceeding towards the fainter sources. The advantage of this approach

is that the fit parameters of brighter neighbours can be included into the fits of fainter sources.

Using the fit profile of a brighter neighbouring source, GALAPAGOS-C can hold the fit parameters fixed to the best values and include it into the fit of the fainter source. The advantage of this procedure is that the time-consuming fits for the bright sources are not repeated multiple times for every fainter source. In a sequential approach, where one object is fitted after the other proceeding through the source catalogue according to the object magnitudes, the fitting process is easy to implement, as the fit profile for every brighter source is already in existence. In case a fit to a brighter secondary might not have been successful, GALAPAGOS-C uses SExtractor parameters and includes the given values with free starting parameters every influenced source. Distributing the fitting jobs on a large number of CPU cores requires a sophisticated scheduling of the fitting procedure (see Section 4.5 ‘Parallelization’).

It can happen that overlapping objects do not have a significant influence on a bright source if they are comparatively faint. For this purpose, the user defines a determining magnitude difference between the bright and the faint neighbour beyond which a faint source is always excluded from the fit using the set-up parameter `magexclude`. If the magnitude difference between a primary and a secondary source is larger than the magnitudes specified by `magexclude`, the secondary source is downgraded by GALAPAGOS-C to be a tertiary source and thus not included in the fit for the primary source. The advantage of using this set-up parameter is that GALFIT cannot use the additional Sérsic component for the faint neighbour wrongly in order to start a multicomponent fit instead of a single-Sérsic fit for the current primary source. Barden et al. (2012) found that 5 mag is a good value for the set-up parameter `magexclude`. Using a magnitude-ordered catalogue makes it easy and quick for GALAPAGOS-C to exclude faint neighbours.

4.4.2 Object masking

GALFIT allows us to make use of object masks for every postage stamp image in order to exclude tertiary neighbouring sources, image defects and region with zero exposure time. GALAPAGOS-C automatically masks tertiary sources but excludes the regions of secondary sources from the mask. In the case of an overlapping tertiary source with the expanded Kron ellipse of the secondary source, the overlapping region is not masked completely. GALAPAGOS-C uses the SExtractor segmentation image to mask only the core region of the tertiary source in the overlapping region with the secondary source in order to obtain the best-fitting results for the secondary source (please see Barden et al. 2012, section 4.5 for details).

4.4.3 Fit constraining

GALFIT can be provided with a constraint file. The GALAPAGOS-C user can optionally constrain the fit parameters for the object magnitudes, the half-light radii and the Sérsic indices. These constraints must always be stronger than the internal constraints of GALFIT, otherwise the values are not included. GALAPAGOS-C will automatically provide GALFIT with a constraint file for every object. Fit constraints should be used with caution. The constraints in the parameter space might not only affect the resulting values but also impact the minimization algorithm itself. When the parameter space for every step in the algorithm is confined, the fitting results can be biased even if the final results are significantly different from any constraint. For details on fit constraining, see Barden et al. (2012, section 4.5). In the case of multicomponent fitting, GALAPAGOS-C automatically includes the

constraints which have been specified in the GALAPAGOS-C set-up file for every component.

4.4.4 Fourier modes

Fourier modes in GALFIT are a suitable tool for studying lopsidedness, asymmetries and morphological distortions of galaxies in general (Peng et al. 2010a, section 4). In general, the isophotal shapes of galaxies differ from perfect ellipsoids, showing perturbations at different intensities. In Fig. 6, we show hypothetical examples of Sérsic profiles with a Fourier mode expansion on the axis-symmetric elliptical isophotes in different intensities. To quantify the intensity and the direction of these perturbations, the Fourier perturbation on a perfect radially symmetric ellipsoid shape is defined as follows:

$$r(x, y) = r_0(x, y) \left(1 + \sum_{m=1}^N a_m \cos(m(\theta + \phi_m)) \right) \quad (2)$$

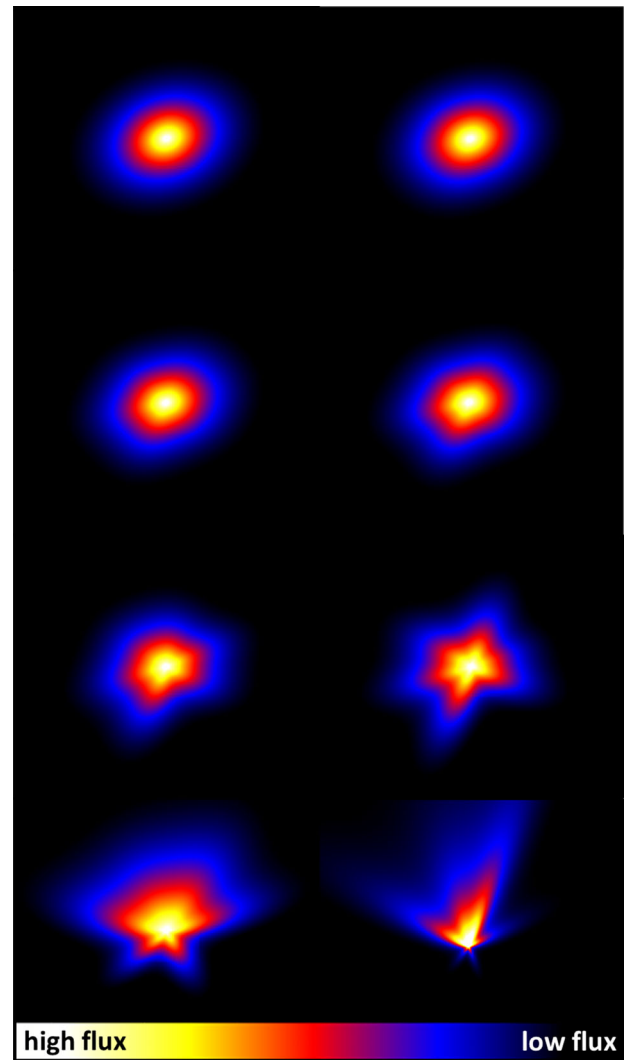


Figure 6. Hypothetical examples for Sérsic profiles with Fourier mode distortions in different intensities in arbitrary directions. Fourier modes F1, F3, F4, F5 and F6 (see text for details) are applied in simultaneously growing intensities from top left to bottom right. The first object is a pure Sérsic profile without distortions. The total Fourier mode intensities are 0, 0.025, 0.05, 0.1, 0.25, 0.5, 1.0 and 2.0. The flux intensities of the objects are indicated by the colour bar below.

with r_0 being the axisymmetric single-Sérsic profile and N being the number of Fourier modes. A significant parameter in this context is the Fourier mode amplitude a_m for every mode m . The Fourier mode amplitude a_m describes the fractional and radial deviation from a perfect Sérsic isophote and hence the importance of the respective mode. The phase angle ϕ_m is the relative orientation of the Fourier mode m towards the position angle θ of the whole object, making it possible to measure these modes at other orientations than along the semimajor axis, e.g. asymmetries perpendicular to the main axis of a galaxy.

These two free parameters a_m and ϕ_m can be used with an unrestricted number of Fourier modes m . The physically most meaningful and easiest to understand of these modes is the Fourier mode $m = 1$, describing the intensity and direction of lopsidedness for a galaxy. Higher Fourier modes $m > 2$ only have a mathematical meaning, but provide a good basis to quantify the overall symmetrical deviation from a perfect ellipsoidal shape of a galaxy. In that sense, it is meaningful to use the overall sum of the absolute values for every Fourier mode amplitude:

$$A_E = \sum_m^N |a_m|. \quad (3)$$

We follow the suggestion in the GALFIT user manual (Peng et al. 2010a) combining only the most useful low-ordered Fourier modes ($m = 1, 3, 4, 5, 6$) and desist from the use of Fourier mode $m = 2$, in order to avoid situations where it could be partially or even entirely degenerate with the classical axial ratio parameter, q , for an ellipse.

4.4.5 Multicomponent fitting

The Sérsic profile can be used to approximate several galaxy components, e.g. bulges, discs or bars. For example, de Jong (1996), Laurikainen, Salo & Buta (2005) and Gadotti (2008) used multicomponent decomposition in order to treat bars as an important component in 2D galaxy fitting. In general, bulges are shown to have a higher Sérsic index in comparison to discs, while bars usually can be approximated with an even lower Sérsic index than discs. GALFIT enables fitting of several Sérsic components to one object at the same time.

GALAPAGOS-C makes use of this option with the following approach: using the set-up parameter `ncomponents` allows GALAPAGOS-C to fit up to five Sérsic components to fit any astronomical source in the survey. Multicomponent fitting is always based on the information from the single-Sérsic fit. If a single-Sérsic fit was successful for an astronomical source, GALAPAGOS-C constructs a two-component GALFIT set-up file for the source in the next step. GALAPAGOS-C automatically fills the set-up file with starting guess values using the information from the single-Sérsic fit. Table 1 shows an example for a two-Sérsic-component fit. The starting values for central positions of the sub-components are always the same as the central position of the single-Sérsic profile. For the component magnitudes, it splits the overall single-Sérsic magnitude equally between all components. For the disc profile, it uses the same half-light radius as provided by the single-Sérsic fit and for the bulge component, the single-Sérsic half-light radius is divided by 2. The starting values for the Sérsic indices are $n = 1$ for the disc component and $n = 4$ for the bulge component. For the axial ratios, it uses the same values as found in the single-Sérsic fit for the disc components, but starts with 1 for the bulge components. It keeps the position angles of the several components apart by adding and subtracting 1° , respectively, in order to disentangle the components

Table 1. Calculation of starting parameters for two-Sérsic-component fits in GALAPAGOS-C. The index ‘Sérsic’ always indicates the result from the previous GALAPAGOS-C single-Sérsic fit.

Parameter	First component	Second component
Position	$\text{pos}_{\text{Sérsic}}$	$\text{pos}_{\text{Sérsic}}$
Magnitude	$\text{mag}_{\text{Sérsic}} + 0.7526$	$\text{mag}_{\text{Sérsic}} + 0.7526$
Half-light radius	$r_{\text{e,Sérsic}}/2$	$r_{\text{e,Sérsic}}$
Sérsic index	4	1
Axial ratio	1	$a_{\text{r,Sérsic}}$
Position angle	$+1^\circ$	-1°

in the fitting procedure. If more than two components are desired by the GALAPAGOS-C user, it will automatically construct multicomponent set-up profiles by always dispersing the single-Sérsic values in graduated steps based on the single-Sérsic results.

However, many sources in galaxy surveys might consist of less components than specified in the set-up file. GALAPAGOS-C provides two approaches to evaluate if the fit of an additional component makes sense or not. It analyses and compares the residual images of the single-Sérsic fits and the multicomponent fits at the object positions within the SExtractor Kron ellipses using two independent criteria: the first possible approach is to sum up the overall residual flux within the SExtractor Kron ellipse by summing up absolute values and calculating the mean rest flux per pixel within the SExtractor Kron ellipse for every object in a relative sense. GALFIT gives estimates of the fit quality by providing a reduced χ^2 value for the fit. This value is calculated across the complete postage stamp region. However, the reduced χ^2 provided by GALFIT will take the residual flux from the secondaries into account or might be biased by unmasked flux resulting from bad neighbouring detections (e.g. image artefacts which cannot be well approximated with a Sérsic profile). Focusing only on the pixels inside the primary Kron ellipse can provide more informative values, e.g., in the presence of a bright (either secondary or tertiary) source on the postage stamp. This superior approach helps to improve the focus on the object of interest. The second approach is to calculate the Bayesian information criterion (BIC; Schwarz 1978) for the same region:

$$\text{BIC} = \chi^2 + k \cdot \ln(n), \quad (4)$$

where χ^2 is a measure for the goodness of the fit in an absolute sense (not the reduced χ^2). Hence, lower BIC implies a better fit. Here, k is the number of degrees of freedom and n is the number of pixels in the original SExtractor Kron ellipse.

These values can be utilized in combination with logical filters to decide whether the additional Sérsic component is meaningful or not and thus to determine the number of object components. In our testing scenarios, the logical filter is provided with threshold values in order to decide whether a component is meaningful or not. GALAPAGOS-C can automatically compare the object parameter results from the single-Sérsic fits and the multicomponent fits. The threshold values for the compared fit parameters can be specified in order to compare the magnitudes and the half-light radii of the fits, respectively. The logical filter can also be provided with threshold values for the Sérsic indices and the axial ratios in the subsequent component. The logical filter in GALAPAGOS-C is able to analyse the parameters in comparison in order to state if a fit component is meaningful or not.

The computational framework for the multicomponent fitting procedure exists and it was successfully applied to the STAGES and the GEMS survey data sets. However, we have not yet tested its per-

formance and data reliability for large data sets. For this reason, the official version of GALAPAGOS-C implements multicomponent fitting as described above without using the automated residual quantification yet.

The current version of GALAPAGOS-C, includes secondaries with single-Sérsic profiles according to the approach as described in Section 4.4.1 ('Simultaneous fits'). Including multicomponent profiles for secondaries needs further investigation and is planned to be included in a future version of GALAPAGOS-C.

4.5 Parallelization

On a supercomputer, GALAPAGOS-C can distribute the actual GALFIT fitting jobs on a large number of CPU cores according to a so-called `master--slave` concept. One CPU processor is selected to be the master processor (master). The master distributes the GALFIT jobs to the remaining number of slave processors (slaves).

The master will proceed through the source catalogue according to the object magnitudes as given by SExtractor, starting with the brightest source in the survey and proceeding towards fainter sources. As explained in Section 4.4.1 ('Simultaneous fits'), object fits can depend on the influence of neighbouring sources, due to direct overlap of the expanded SExtractor ellipses (direct secondaries) or due to direct influence by a significant amount of light (contributors). In both cases, the neighbouring objects have to be included into the fit of the current primary as secondary objects. If the secondary is brighter, it will be included with its parameters held fixed to the best-fitting values. If the secondary is fainter than the primary source or the fit for a brighter neighbouring source was not successful, GALAPAGOS-C uses the SExtractor parameters of the secondary source to provide the GALFIT set-up file with free starting parameters.

However, when the sources are fitted with GALFIT in parallel, it often happens that the fit profile for a brighter secondary is still being developed on another CPU node and thus the fit parameters for a brighter neighbouring source do not exist at that moment. In

this case, the master process blocks the GALFIT job with brighter secondaries until the fit for the brighter secondaries are done. The CPU core which was allocated for the blocked source will work on an unblocked source in the meantime and the blocked primary will be fitted by the first available CPU node as soon as the fit profile of the influencing secondary source becomes available.

In order to know, which sources are influencing a primary source, GALAPAGOS-C must identify the directly overlapping secondaries and the contributors first. As the fit profiles for contributors might not exist yet, GALAPAGOS-C has to estimate the influence of contributors in the neighbourhood. To overcome this problem, GALAPAGOS-C assumes a worst-case scenario for the light contribution of every object in the neighbourhood. Fig. 7 shows an overview for the procedure. GALAPAGOS-C assumes every source to have a high Sérsic index of 4 and calculates a virtual Sérsic profile for every object in the neighbourhood with the information from SExtractor (green profile). If the influence of a virtual source exceeds a critical magnitude which is defined by the set-up parameter `magcrit` (bold red line) at the central position of the current primary (orange), the neighbouring object is identified as a 'potential contributor'. The flux contribution in the worst-case scenario of the neighbouring source has to exceed the dashed red line at the centroid position of the current primary object. The fitting job of the primary will be blocked in this case until the fit profile of the potential contributor exists. With the true information about the profile of the potential contributor (blue profile) available, GALAPAGOS-C calculates again whether the true light contribution is indeed high enough. Only if the light contribution is still exceeding the critical value at the position of the current primary, it is identified as a true contributor and is thus included in the fit for the current primary with fixed parameters. Searching for contributors for every astronomical source might take a long time, if the survey is large. For this reason, the user must specify a radius around any source to define the neighbourhood area using the set-up parameter `distmax`. GALAPAGOS-C only analyses every source within radius of `distmax` around the primary source. A good estimate for this value can be found orientating at the most

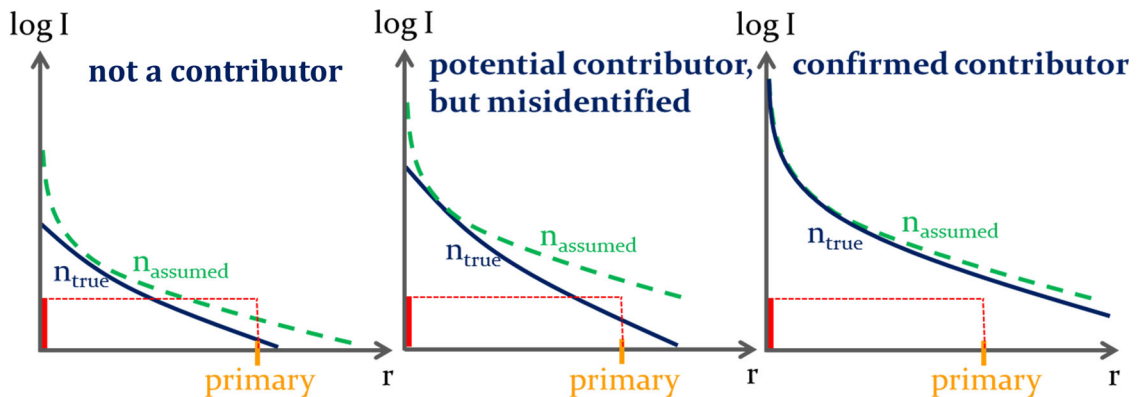


Figure 7. Identification of contributors in GALAPAGOS-C (for a definition, see Section 4.4.1 'Simultaneous fits'). The diagrams show the logarithmic intensity of a neighbouring light source around a primary source at every distance r to the centre of the neighbouring source. In order to identify sources with a significant light influence on the current primary source in the surrounding image areas, GALAPAGOS-C measures the mean influence of every source in a user-defined radius around a primary source at the centroid position of the primary (orange). The user has to define the distance using the set-up parameter `distmax` and the distance from the primary source to the neighbouring source must be smaller than `distmax` for the investigation. As the Sérsic profiles for the neighbouring sources are not available, GALAPAGOS-C has to assume a worst-case approach for every source in the neighbourhood using a high Sérsic index of 4 (green profiles). If even the green worst-case profile does not exceed the user-given luminosity threshold value `magcrit` (indicated by the bold red line and the vertical red line) at the centroid position of the current primary source, the source is not a contributor (left). In the case that the worst-case profile exceeds the threshold value, the neighbouring source is identified as a 'potential contributor' and GALAPAGOS-C has to block the fitting procedure for the primary until the true fit (blue profile) for the potential contributor exists. If the true light contribution of the potential contributor at the centroid position of the current primary source is indeed below the threshold value (centre), GALAPAGOS-C will start the fitting procedure without including the neighbouring source. If the true light contribution still exceeds the threshold value, GALAPAGOS-C will include the confirmed contributor (right) in the fitting procedure with its parameters fixed to the best values.

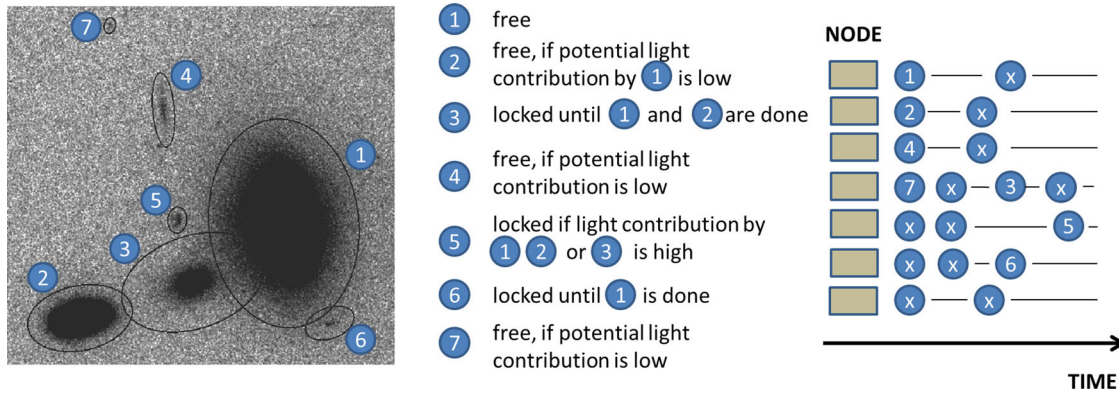


Figure 8. Scheduling of overlapping object profiles. Left: the brightest object at the right-hand side will be treated first. Objects with a direct overlap of Kron ellipses (objects 3 and 6) will be blocked until the fit for object 1 is done. The objects with no direct overlap (objects 2, 4, 5, and 7) might be blocked as well, depending on the user-defined threshold value for significant light contribution. The same concept is applied for the other objects. In order to calculate the ‘worst-case’ light contribution by neighbouring Sérsic profiles, GALAPAGOS-C assumes every object to have a high Sérsic index of 4 to construct a robust schedule for the fitting procedure. For details, see Fig. 7. Right: a possible example for a schedule for the treatment of the objects. Other objects in the survey are marked with ‘x’. Note that object 3 is free, when objects 1 and 2 are done. Object 6 is free when object 1 is done and object 5 must wait for objects 1 and 3 (GALAPAGOS-C usually uses enlarged Kron ellipses for secondary identification). The exact schedule depends on set-up parameters and the ratio of CPU nodes to the number of sources in the whole survey.

influencing source in the whole survey. For this estimate, it makes sense to use the largest elliptical galaxy and appraise its influencing region.

Fig. 8 shows a possible scenario for the distribution of fitting jobs in a parallel environment: If the brightest source (1) is the brightest source in the survey or not influenced by any ‘potential contributor’, it is free to be sent to a free CPU node for fitting with GALFIT. For the second brightest source (2), GALAPAGOS-C has to calculate the light contribution by source (1). If the fit profile does not exist at the moment of the procedure, GALAPAGOS-C has to assume a worst-case scenario with a high Sérsic index of 4 for source (1). If the light contribution exceeds the critical value, which was defined in the set-up file, the source (2) will be blocked until the fit profile for source (1) exists. Once it does, GALAPAGOS-C will recalculate the true light contribution with the existing parameters to decide whether the source must be included in the fit or not. Source (3) must be locked until the brighter sources (1) and (2) are done in any case, as they are both identified as secondaries due to direct overlap with the SExtractor Kron ellipses. Source (4) and (5) might be blocked if the threshold value for a critical magnitude contribution by any neighbouring source is exceeded and source (6) is blocked until source (1) is done. In the case of source (7), GALAPAGOS-C will calculate a worst-case light influence by the neighbouring sources, if the distance from the object centres is lower than the given value by the set-up parameter *distmax*. In any case, the light contribution for source (7) will probably be too low and the source will be free to be fitted on the first CPU core, which is available.

Fig. 8 (right) shows a possible schedule for the fitting procedure. While influenced sources are blocked until the fits for the brighter secondaries and contributors are done, GALAPAGOS-C will work on independent and free sources (x) in the meantime. In order to occupy every CPU and not waste computation time, the number of CPU cores must be significantly lower than the number of astronomical sources. GALAPAGOS-C blocks the allocated CPU nodes and they cannot be employed by other users of the supercomputer in the meantime. A meaningful ratio of astronomical source to CPU nodes strongly depends on the object density and the reciprocal light influence of the astronomical sources in the survey. We intensively tested the speed-up behaviour of this procedure using the GEMS

and the STAGES survey (see Section 6.2 ‘Speedup performance’). We found that if the ratio of source to cores is at lower values than 100, a linear speed-up behaviour can be achieved.

5 DATA SAMPLE

5.1 HST surveys

This highly parallelized version of GALAPAGOS-C is designed to be applied to any data sets for surveys with a stable PSF (for a stability test of the PSF, see Appendix B). Here, we use STAGES (Gray et al. 2009) survey science images to test its performance using supercomputers. This survey of the Abell 901/2 multicluster system at $z \sim 0.167$ is ideally suited for this purpose as it includes a number of regions with both low and high object density. It is the same data set that GALAPAGOS was tested on by Barden et al. (2012), making the comparison both easy and clean. Especially the high-density regions provide many overlapping sources and hence provide an ideal test-environment to test the quality and performance of the parallelization methods used by GALAPAGOS-C.

In detail, we use the 80 individual one-orbit-depth *F606W* HST/Advanced Camera for Surveys (ACS) images, that make up a mosaic with a combined number of $\sim 80\,000$ detected sources within a region that covers 0.5×0.5 ($\sim 5 \times 5$ Mpc²) area of the multicluster system. The field was observed in the framework of the 17-band COMBO-17 survey (Wolf et al. 2003), which complements the ACS imaging with photometric redshift data and observed/rest-frame spectral energy distributions (SEDs) between 350 and 930 nm (*UBVR* broad-band filters). Additionally, *Spitzer*, *Galaxy Evolution Explorer*, the Giant Metrewave Radio Telescope and *XMM-Newton* provide useful information for scientific purposes, but are obviously not yet used by GALAPAGOS-C.

Detailed optical SEDs, structural and morphological parameters are collected for all cluster galaxies with $M_R \leq 24$ (equivalent to $M_V \sim -16$ at $z \sim 0.165$) as well as information about fore- and background galaxies to observe a wide range of environments. The 80 individual images/tile observed by STAGES mosaic in the *F606W* filter and generally show overlaps of the order of ~ 100 pixels (see Fig. 9), making this an ideal survey set-up for

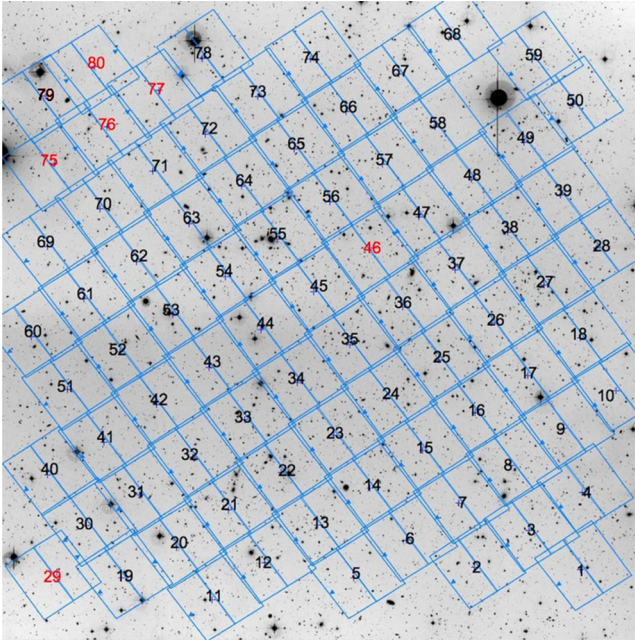


Figure 9. Arrangement of the STAGES image tiles.

analysis within GALAPAGOS-C. For further details, see Gray et al. (2009).

The parameter space of absolute magnitude, surface brightness and object size amongst other parameters is limited in the framework of survey observation and data analysis. In order to illustrate our survey information limits, we may analyse the bivariate brightness distribution (BBD) which relates surface brightness to absolute magnitude (Driver 1999). Fig. 10 shows a BBD for our data set of the confirmed cluster members in A901/902. The absolute average effective surface brightness (i.e. the average surface brightness within the half-light radius) is shown against the absolute $F606W$ GALFIT magnitude. The absolute magnitudes are calculated from conversions of the apparent GALFIT magnitudes into absolute magnitudes. The absolute average surface brightness ($\langle \mu_e \rangle$) is given by

$$\langle \mu_e \rangle = \langle \mu_e \rangle - 10 \log_{10}(1+z) - K(z), \quad (5)$$

where $\langle \mu_e \rangle$ is the average surface brightness within the half-light radius r_e and $K(z)$ considers the rest-frame K -correction. For the calculation of the data points within the BBD, we neglect K -corrections and assume $K(z) = 0$. In order to compare all galaxies at the given redshift $z \approx 0.165$, we neglect an evolutionary correction $E(z)$. The average surface brightness within the half-light radius ($\langle \mu_e \rangle$) is given by

$$\langle \mu_e \rangle = m + 2.5 \log_{10}(2) + 2.5 \log_{10}(\pi r_e^2). \quad (6)$$

The Sérsic index populations are indicated by contours in three bins: the red contours show the distribution of galaxies with a high Sérsic index ($2.5 < n < 7.0$), the green contours show the populations of galaxies with an average Sérsic index ($1.6 < n < 2.5$) and the population with a low Sérsic index ($0.4 < n < 1.6$) is shown in blue.

The grey lines show the parameter limits within which robust survey analysis can be performed. The solid lines indicate absolute physical limits. These are the detection limit, the object size limits and the surface brightness limit for faint objects and a brightness

limit e.g. due to CCD saturation for bright sources. The dashed grey lines provide information about the limit down to which GALFIT is shown to provide robust values.

Fig. 11 shows that GALFIT on average determines the parameters for Sérsic profiles within a 3σ scatter below ~ 10 per cent if the apparent magnitude is more than 2 mag brighter than the detection limit. For this reason, we add 2 mag to the average detection limit (at ~ -13 mag) and consider every object which has a brighter absolute magnitude as indicated by the dashed, vertical line at ~ -15 mag to be bright enough for our data analysis. The lower limits for the apparent object size of a galaxy have been provided by tests on GALFIT image simulations and by Häussler et al. (2007). It is shown that the object parameter errors slightly increase if the half-light radius is smaller than the half width at half-maximum (HWHM) of the PSF. The dashed grey line on the upper right indicates objects with a half-light radius, which have the size of the HWHM of the PSF. This corresponds to a value of ~ 3 pixels for the STAGES survey. Objects with a larger half-light radius are considered to be large enough for scientific data analysis. The solid grey line indicates a physical constraint on the object size for objects with a half-light radius $r_e < 1$ pixel. Häussler et al. (2007) found that GALFIT produces robust fitting parameters for STAGES survey data even at this object size. The surface brightness limit is indicated by the grey, horizontal lines at the bottom of the plot. The stability of fitting parameters for single Sérsic profile fits in the STAGES survey has been quantified as a function of surface brightness by Barden et al. (2012). Using image simulations of more than 1.5 million objects, it is shown that the recovery of input single-Sérsic profile parameters is stable if the simulated mean surface brightness μ_{sim} within r_e is brighter than ~ 24.2 mag arcsec $^{-2}$. The average deviations within a 1σ scatter are lower than 10 per cent for magnitudes and half-light radii and lower than 20 per cent for the Sérsic indices in a sample with a low Sérsic index ($0 < n < 2.5$, ~ 1.1 million objects). In the same manner, the uncertainties for the single-Sérsic profile parameters are still lower than 20 per cent if the analysis is limited to the sample with high Sérsic indices ($2.5 < n < 8.0$, ~ 470 000 objects). In both cases, the profile parameters do not show any significant trends. We adopt this result in Fig. 11 and assume that objects with an average effective surface brightness brighter than ~ 24.2 mag arcsec $^{-2}$ are well suited for our data analysis. In addition, we include the physical detection limit for objects which are fainter than ~ 26.2 mag arcsec $^{-2}$ with a solid dashed line at 26.2 mag arcsec $^{-2}$, as the fraction of faint, undetected objects is significant below an apparent magnitude of $\mu_e \sim 26.2$ mag arcsec $^{-2}$.

The majority of our data sample is located well within the boundaries inside the dashed lines, implying that the sample is robust. Only a small fraction of the areas in the contour line which contain at least 90 per cent of the complete data sample overlap with regions where the robustness of the fitting results may perhaps suffer. To the right of the figure, a small fraction of objects tend to be too dim. The result for these possible low-surface-brightness galaxies is that the Sérsic index may be underestimated. This fraction however only contains a number of galaxies, which is too low to bias our statistical parameter results significantly. At the bottom of the figure, a small fraction of all Sérsic types touch the faint magnitude limit. As a result, the fit uncertainties for this fraction can increase. By visual inspection, we note that a significant fraction of these galaxies consist of dwarf galaxies and objects with a similar structure to e.g. the so-called Green Pea galaxies (e.g. Cardamone et al. 2009; Amorín et al. 2013). Such low-mass, compact and oxygen-rich emission-line galaxies with a high star formation rate are typically observed at redshifts between $z \sim 0.1$ and ~ 0.35 (Cardamone et al. 2009).

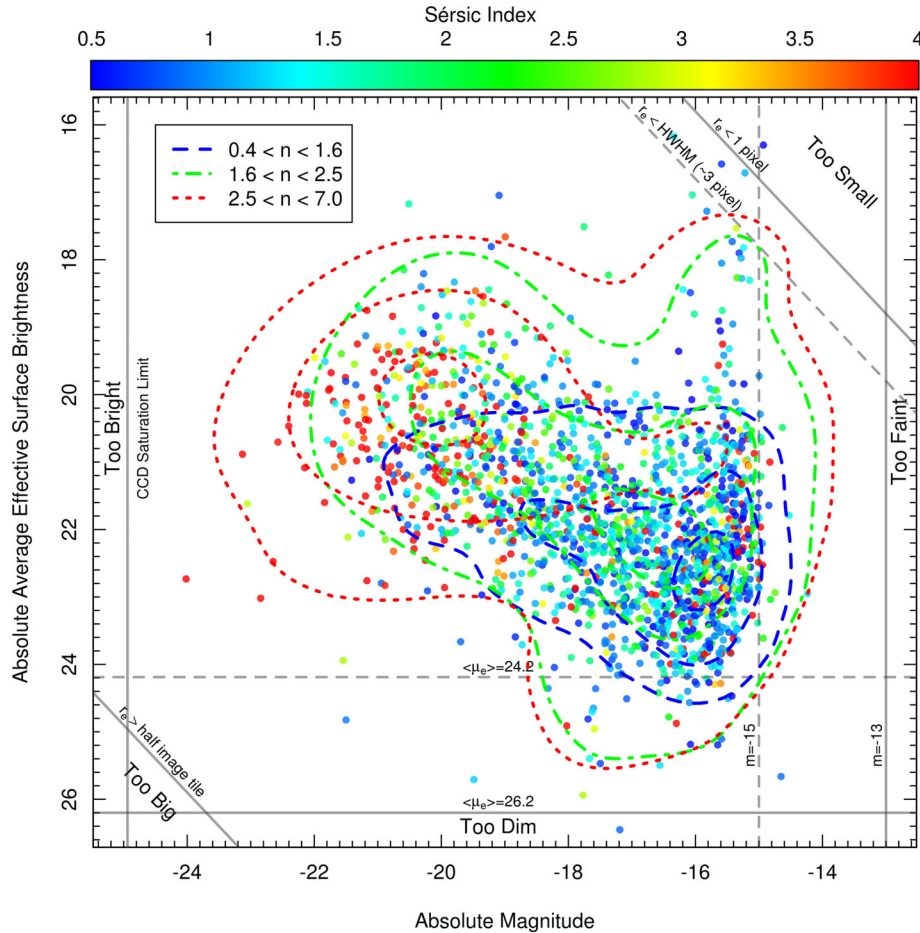


Figure 10. The BBD in the $F606W$ band for all galaxies in our data sample. The absolute average effective Sérsic surface brightness (i.e. the average surface brightness within the half-light radius) is plotted against the absolute Sérsic magnitude. Colours indicate the Sérsic indices from low (blue) to high (red). Contour lines indicate the density of every population, i.e. the outermost contour of every colour contains 90 per cent of the data points, the middle contour contains 50 per cent of the data points and the innermost contour contains 10 per cent of the data points and indicates the region with the highest density and indicate the object density in every region. The borderlines indicate the magnitude, surface brightness and size limits within which the data analysis can be performed per se (see text for details). The solid lines represent physical limits (e.g. detection limits). The dashed lines give an estimate if enough information is provided for GALFIT to produce robust fitting results. As all galaxies are located at approximately the same redshift, the borderlines should be considered valid for the whole data sample.

A large majority of our data lies well inside the constraint boundaries. The BBD shows that objects which are too bright or too big do not occur. The number of objects which are too small in our data sample is negligible and objects which are too dim should not influence our results in practice. Close to the faint magnitude limit, a significant fraction of objects are approaching the border. As a consequence, the fitting uncertainties could increase as we approach the faint magnitude limit. However, the vast majority of our data lies safely within the boundaries. Based on the BBD for our galaxy population, we conclude that the derived luminosities and surface brightness are sufficient for the statistical analysis of galaxy morphologies in our data sample.

5.2 Simulations

Image simulations are the most convenient tool for testing the performance and the data reliability of quantitative fitting pipelines and they have been used for similar purposes by many other groups (e.g. Häussler et al. 2011; Simard et al. 2011). Especially, it has been shown previously that GALFIT models generally correspond well to

more accurately sampled (but much more CPU-intensive) profiles (e.g. by analytic integration of the model profile over a pixel area; Häussler et al. 2007). Hence, for the purpose of creating model images, GALFIT cannot only be used to extract the profile parameters from galaxy images, but can also be utilized to create the above mentioned images in the first place, creating perfect mathematical galaxy profiles from user-given input parameters.

For this purpose, GALFIT must be provided with a complete set of galaxy parameters for a whole survey containing a large number of objects. Following the idea of Häussler et al. (2007), Gray et al. (2009) and others, we use GALAPAGOS to create an object catalogue in the first place and use the resulting parameters to re-simulate the survey itself, creating a realistically looking – but entirely artificial – data set for which the true fitting values are known for an analysis of the fitting code in detail, a large number of simulated galaxy profiles is required in a fashion that resembles the real data as realistically as possible. The performance of GALAPAGOS-C is then measured by applying it to this simulated version of the STAGES survey and comparing fitting results to input values. While this

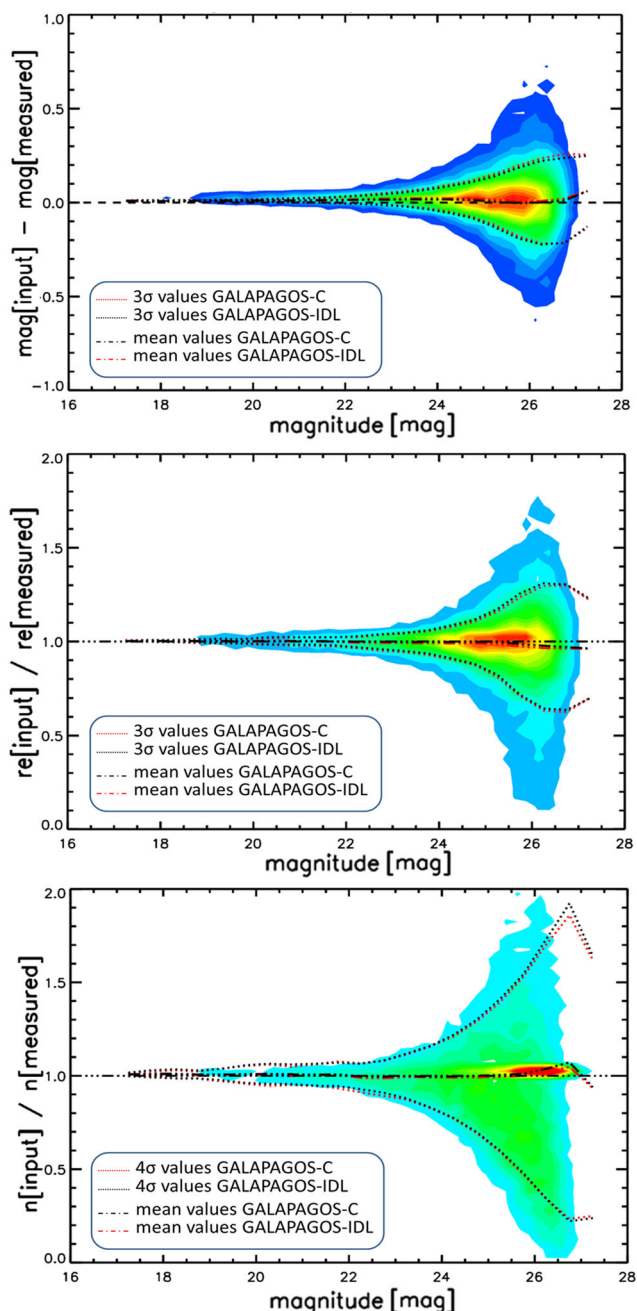


Figure 11. Parameter recovery for testing GALAPAGOS and GALAPAGOS-C with simulated images. Every detected object with an existing GALFIT profile in the STAGES survey was re-simulated and re-analysed. Important parameters are absolute magnitude deviation (top), ratio of measured to true half-light radius (middle) and the ratio of measured to true Sérsic index (bottom). A comparison between GALAPAGOS in IDL (black) and GALAPAGOS-C (red) is shown by the dashed lines which indicate mean values, mean+ 3σ and mean+ 4σ deviations, respectively. As one can see, both versions return virtually identical accuracies.

scheme is potentially biasing against objects that are hard to fit, as it starts from GALAPAGOS results, this effect has been found to be negligible (Gray et al. 2009; Haeussler et al. 2011).

As the original GALAPAGOS version was also used and tested on the same data set, this allows a convenient comparison to the performance of the IDL code (see Barden et al. 2012, section 5).

5.2.1 Simulating image tiles with single-Sérsic profiles

For testing purposes, we use the same approach with simulated data as described in detail in Häussler et al. (2007). GALAPAGOS provides a well-suited and tested data set for (re)simulating Sérsic profiles of astronomical sources at their original positions with their morphological parameters.

These Sérsic profiles (81 365 objects) were placed at their original positions into a mosaic of empty images (no sky background) and convolved with the same PSF used for real STAGES data as described in Gray et al. (2009). Next, an artificial sky background was added, assuming a Gaussian distribution of pixel values and Poisson noise was added to the image. The sky mean values and the variation were measured through an image in regions of pure sky pixels in the original STAGES image tiles and applied to the simulations. During this process, remaining object contribution to the sky measurement was avoided by using 3σ clipping when calculating the mean sky value.

The GALFIT data for saturated stars was included as well as Sérsic profiles using their fit results. However, image defects and other artefacts (for example cosmic rays) which cannot be sensibly approximated by a Sérsic profile were not included in the simulations, creating a cleaner, somewhat easier-to-use set of images, while the galaxy distributions in our simulated images were created to reflect real data as close as possible. We measure the performance of GALAPAGOS-C using the described simulated data and test the data quality in comparison to the former IDL version of GALAPAGOS.

6 DATA QUALITY AND PERFORMANCE

One primary aim of this new version of GALAPAGOS, but recoded in C, was to increase the efficiency of code, so it can be run on large sets of data on sensible time-scales, but without degrading the quality of the values returned. In this section, we test the data reliability using artificial image simulations (see Section 5) and its speed-up performance on real astronomical data.

6.1 Single-Sérsic fitting

We used the output parameters of the IDL version of GALAPAGOS to re-simulate the STAGES survey. In order to compare the fitting results of both the GALAPAGOS in IDL and the GALAPAGOS-C code, we apply the same GALAPAGOS set-up for fitting the simulated STAGES data set with both existing codes.

Of the 81 365 objects that were simulated, 70 132 objects were re-detected by SExtractor. Especially, spheroid-dominated galaxies can be detected easily as they mostly consist of a centrally concentrated light profile and low-surface-brightness galaxies with a wide light distribution are proved to be more difficult in the detection process (Rix et al. 2004; Haeussler et al. 2011). Among the lost objects, a large fraction consists of very faint sources close to the detection limit. They were not re-detected mainly due to the different nature of the Gaussian sky background or because of relatively too much noise in the original measurements. Another fraction of the re-simulated objects consist of image defects and artefacts, e.g. one spike of a star is re-simulated by an accumulation of many faint sources. These artefacts cannot usually be simulated easily with a Sérsic profile and thus are not well re-simulated. Therefore, we find that stand-alone artefacts within an accumulation of overlapping, artificial objects at the position

of an image defect are not well re-detected and thus sources are missed.

As the same SExtractor set-up was used by both runs, including a two-step detection procedure with the identical ‘cold’ and ‘hot’ modes, the data knowledge is the same to both codes at this stage making the result comparison ultimately ‘fair’.

Starting the fitting procedure with GALFIT using the same constraints for the Sérsic index n ($0.2 < n < 8$), the half-light radius r_e ($0.3 < r_e < 750$) and the magnitude deviation from SExtractor magnitude ($|m_{\text{Source Extractor}} - m_{\text{GALFIT}}| < 5$) that we used in previous versions of the code (see Barden et al. 2012, section 4), only 12 fits on simulated and detected objects failed in both codes. In total 70 120 sources were fitted successfully by both GALAPAGOS-C and GALAPAGOS. The GALFIT crashes were not caused by the same objects in both cases. Small deviations in the GALFIT set-up parameters can lead to different results, especially if a small deviation in the background sky values occurred (see Appendix A, ‘Background estimation test’).

Consistent to the way we presented similar results in previous papers, Fig. 11 shows the results for the three most important profile parameters as a function of the original profile magnitude. While we see statistical errors (manifested in the widening 3σ lines), we see no significant systematic trends and we conclude that our latest version of GALAPAGOS-C shows an accurate data reproduction. As expected, one can see that with fainter magnitudes – and lower signal-to-noise ratios – the uncertainties increase drastically, the decreasing values at very faint objects are dominated by low number statistics.

Fig. 11 also shows a direct comparison of both codes. Both the mean values of the profile parameters and the 3σ scatter are in excellent agreement with each other. Small deviations between the two codes can be explained by small differences in the sky calculation routine as GALAPAGOS-C uses a more optimized approach (see Section 4.5, ‘Parallelization’) and by small differences caused by the different handling of contributing sources within both codes (see ‘Simultaneous fits’).

Over all, we find the data quality of the new version of GALAPAGOS-C to be as robust as the IDL version presented in Barden et al. (2012), section 4.

6.2 Speed-up performance

Again, we refer to the single-orbit *HST* survey STAGES for testing GALAPAGOS-C performance. This data set is ideally suited for testing the deblending processes during the object detection on the one hand while complex objects like mergers and especially overlapping objects and high object densities will demand much computing time when modelling with GALFIT.

‘The fitting process with GALFIT is the main limitation for GALAPAGOS performance’ (Barden et al. 2012, section 5). Increasing the total performance of GALAPAGOS was the main aim to be achieved with a new, truly parallel code written in C. As explained in Barden et al. (2012, section 5), doubling the number of CPU cores does not necessarily imply to reduce the required computation time by a factor of 2. Finding the most appropriate number of CPU cores for a data set is highly non-trivial as ostensibly two factors compete against each other. A deep survey with complicated geometry will lead to many bright objects to be fitted with a large number of free parameters due to overlapping objects. This requires a multiple amount of time (see Barden et al. 2012, fig. 17). To guarantee the correct handling for the fainter neighbours – and subsequently again their fainter neighbours – in a truly parallel procedure, these fainter objects must be blocked

until the fit for the brighter source is done. Looking at an entire survey, this results in a tree-like scheduling of object treatment (see Section 4.5 ‘Parallelization’). The overall process time can then not be shorter than the time it takes to fit the most CPU/time intensive branch in this ‘fitting tree’. Under these circumstances, it becomes less sensible to use a larger number of processors as additional processors would not lead to an overall improvement on processing time.

The most important factor in this context is the overall number of sources in the survey, or better – the number of branches in the fitting tree – the more sources it contains, the more ‘branches’ will be in the ‘fitting tree’. In a survey with many branches, more processors should have a positive effect on the CPU time as more branches can be worked on simultaneously while one single processor is working its way through the most time intensive branch, from the brightest object to its fainter neighbours.

However, if too many CPUs are used, most of those will stand still and wait for the slowest branch to finish. For a sensible number of CPU processors – compared to the number of branches in the fitting tree – enough time should be left to execute the most work intensive part, while the other processors will be busy preparing and fitting the other objects. These other processors should ideally end their work at the same time, and no CPU power is wasted (see Fig. 12) e.g. one processor would work through several ‘branches’, with a combined computing time equivalent to the one of the slowest ‘branch’. However, for a high number of processors, the procedure will always be dominated by the slowest branch. Fig. 12 shows an example: if up to 64 cores are applied to process the STAGES survey data set, the CPU nodes finish their work at approximately the same time. If 128 processors or even more are applied, the procedure is too fast for the most time consuming source (and all its depending neighbours). This leads to the effect that CPU nodes cannot be provided with fitting jobs and they become inactive towards the end of the fitting procedure whilst others are still working. The relative amount of overall CPU waiting time (red area) therefore increases. It is therefore necessary to find the optimum number of processors for every survey, between using a sufficient number of CPUs in order to go through all the fitting quickly, while not having CPUs stand still, waiting for a different processor to finish it fits.

Our suggestion for the number of CPU cores is for them to be geared to the number of astronomical sources. It makes sense to test the obviously most complicated and time demanding source including all its secondary objects for its fitting time to estimate a meaningful number of CPU cores. However, the exact number will depend on many factors within a survey, so this approach serves for giving an estimate for the number of CPU cores. For the scientific application of this code, it is not necessary to spend too much time on this optimization of the number of cores used, as the output parameters do not depend on the number of CPU cores in practice. This is a result of the fact that different ‘branches’ of the fitting tree are independent of each other.

The connection of speed-up to overall number of sources in a survey leads to the advantage that the speed-up will adapt automatically to the size of a survey. Fig. 13 shows the speed-up performance for our test example using the STAGES data set (~81 000 galaxies). In this example, an appropriate speed-up performance can be accomplished using up to ~100 cores. We find it to be largely impossible to fit the STAGES survey in less than 8 h, no matter how many cores are used. We measured the continuous curve by fitting the STAGES survey with 4, 8, 16, 32, 64, 128, 256 and 512 CPU cores. If even more cores are used, the only advantage is that the most time-intensive GALFIT operation will start slightly earlier. However,

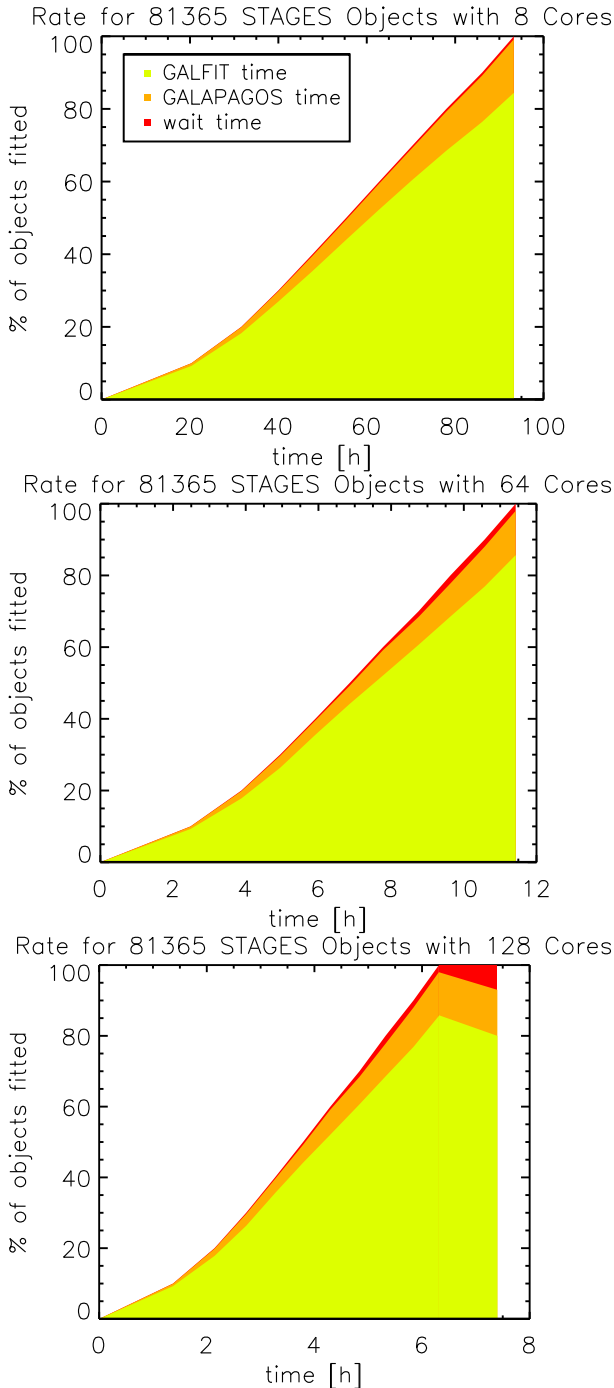


Figure 12. Fractional distribution of computation time for the slave processes for the STAGES sample with a standard set-up for GALAPAGOS-C. Most of the time is consumed by fitting galaxy profiles with GALFIT (yellow). The orange area shows the fraction of time needed by GALAPAGOS-C to prepare the GALFIT process for every source, mainly I/O time for mask creation. The red area shows the fraction of average waiting time for every slave process. In the upper two cases with 8 and 64 nodes, all processes finish nearly at the same time. In the example at the bottom, some processes are still working at time-intensive GALFIT jobs, while the others have already finished their work resulting in an increased red fraction of waiting time towards the end of the procedure where almost all object fits are done and GALAPAGOS-C has to wait for the most time-consuming fit(s). Hence, we would argue that using between 64 and 128 cores for this survey is the most meaningful approach.

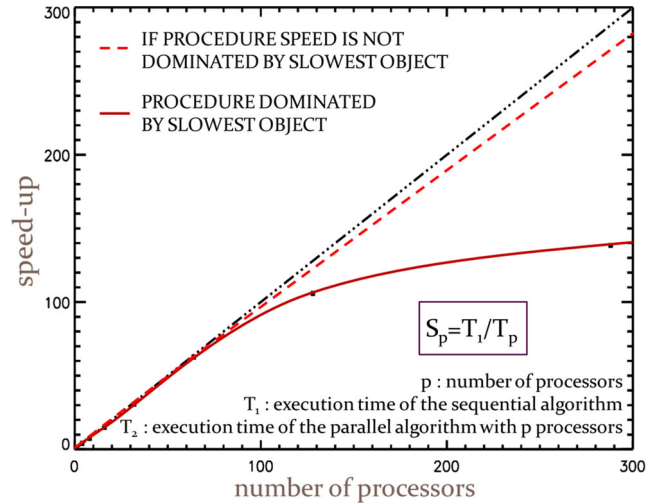


Figure 13. Speed-up performance of GALAPAGOS-C with a standard set-up for the STAGES survey. The black triple-dot-dashed line shows a linear speed-up behaviour which would represent an ideal case. The red line shows the speed-up for $\approx 81\,000$ objects for the STAGES survey. Using up to ≈ 100 nodes, a nearly linear speed-up is observed. Using more nodes leads to no further significant speed-up as the procedure is dominated by the slowest object in this case (see Fig. 12, for an example using 128 cores). Testing with multiple number of STAGES survey objects shows the speed-up behaviour in the red dashed line.

the slowest branch starts to dominate the overall finishing time and the speed-up performance will be truncated.

Dealing with even larger surveys with more objects, the speed-up will perform according to the dashed red line. The dashed line was measured by fitting the STAGES survey with every galaxy multiplied four times after the HDR detection run. In order to provide a realistic order of different magnitudes in the combined catalogue, the magnitudes of every source were slightly altered at the fifth position after the decimal point. Thus, we constructed a large data set which contained every detected object in the STAGES survey four times and provided GALAPAGOS-C with $\sim 323\,000$ objects. For the measurement of the red dashed line in Fig. 13, we used 32, 64, 128, 256 and 512 cores.

7 SERSIC INDEX RELATION IN A901/902

Since the pioneering work of Dressler (1980), progressively more evidence for a correlation between the morphological properties of galaxies and their population density has been noted in the literature e.g. Postman & Geller (1984), Dressler et al. (1997), Springel et al. (2001), Kauffmann et al. (2004), Baldry et al. (2006), Bamford et al. (2009), Vulcani et al. (2012), Calvi et al. (2012), van der Wel (2008), Peng et al. (2010b), Blanton et al. (2005) or Pimbblet & Jensen (2012).

Number statistics show that less dense field environments are usually dominated by late-type galaxies, whereas early-type galaxies represent a proportionally large fraction of the total galaxy population in dense regions in the field (morphology–density relation).

Disc-dominated systems usually can be approximated by a pure exponential profile with a Sérsic index of $n \approx 1$, when the contribution of a bulge is taken into account. Early-type galaxies are typically well described by a Sérsic profile (for details see Section 4.4 ‘GALFIT’) with a comparatively higher Sérsic index. Ellipticals and bulges cover a wide but on average higher range of Sérsic indices.

The Sérsic index for a complete galaxy generally increases towards more massive bulges (Kelvin et al. 2012).

Furthermore, the visually classified Hubble type of a galaxy has been shown to correlate closely with the Sérsic index of its measured light profile (Kelvin et al. 2012). To expand on this study, we now investigate the correlation of the Sérsic index with environmental density, thus providing a complementary measure of the morphology–density relation, the Sérsic–density relation (SDR) measured in the Abell 901/902 system.

7.1 Data

We adopt the cluster environment in A901/902 as our sample data set. GALAPAGOS-C provides information on the Sérsic profiles of 2281 cluster members. Our cluster sample is a combination of the confirmed cluster members as previously defined in Wolf et al. (2003) and Gray et al. (2009). The Wolf et al. (2003) cluster galaxies are selected using the COMBO-17 survey data set, which confirms 775 objects. The Gray et al. (2009) cluster galaxies are selected using the *HST* STAGES survey data set, which provides additional 1504 objects; a combined of total 2279 galaxies. Two-degree field spectroscopy has been used to verify cluster membership and provides redshift information for the individual galaxies and the mean redshifts of the galaxy sub-clusters. Matched X-ray point sources have likewise been used to verify cluster membership according to Gilmour et al. (2007). The sample was constructed with 90 per cent (Gray et al. 2009) completeness using the following approach: identified galaxies must fulfil the criterion

$$\text{abs}(z - 0.17) < dz, \quad (7)$$

where

$$dz = \sqrt{0.015^2 + (1.65\sigma(R))^2} \quad (8)$$

$$\sigma(R) = 1.17 \times 0.005 \sqrt{1 + 10^{0.6+R_{\text{mag}}-20.5}} \quad (9)$$

with z being the measured redshift and R_{mag} the measured R -band magnitude (Gray et al. 2009). For the purpose of defining cluster memberships, we rely on the most conservative selection according to the STAGES master catalogue (COMBO_FLAG ≥ 4). This flag is the strictest on the difference in redshift which is allowed for a certain magnitude. We matched the detected galaxies within GALAPAGOS-C using a position matching algorithm. Due to a mismatch in the object position or a seemingly poor galaxy model fit (as indicated by Sérsic indices $n < 0.4$ or $n > 7$), 141 objects were discarded from the data set. In addition, 71 objects are excluded from the data analysis according to the given constraints in the BBD (see Fig. 10). Therefore, our final data set consists of 2063 galaxies spanning a range of local density environments in the cluster environment A901/902.

7.2 Object classification

Fig. 14 shows the distribution of Sérsic indices for all galaxies in the sample. The solid blue and red lines show exponential (disc-like) and de Vaucouleurs (spheroid-like) Sérsic indices, respectively. The increased fraction of objects at the leftmost and rightmost bin result from bad model fits (<6.5 per cent) as the fit ended up on or very close to a fitting constraint, effectively preventing GALFIT to carry out a free fit. Thus, we discard possibly biased fits for galaxies with a Sérsic index below $n = 0.4$ or above $n = 7$ (grey shaded areas). We now divide our sample into three Sérsic index bins: low ($0.4 <$

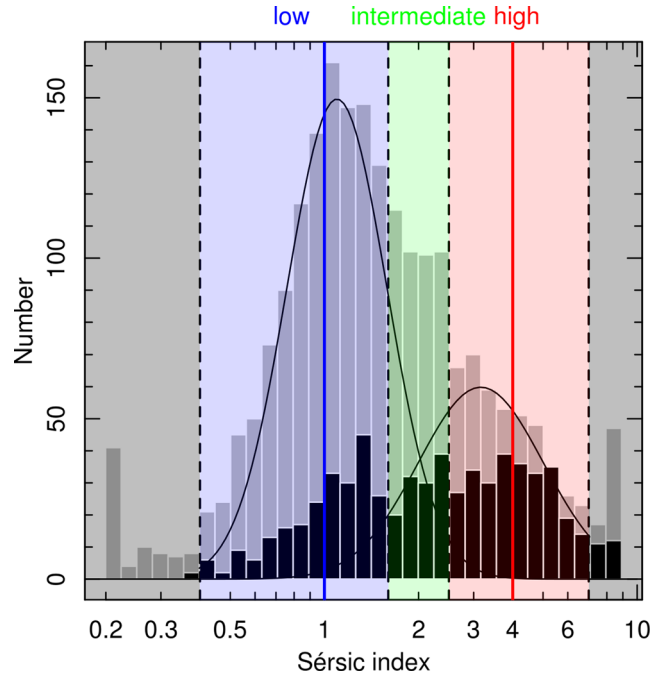


Figure 14. A histogram showing the distribution of Sérsic indices for galaxies in our sample (2138 objects). The blue line indicates Sérsic profiles of typical disc-dominated systems at a Sérsic index of $n \approx 1$ and the red line indicates spheroid-dominated systems at $n \approx 4$. The black dashed lines show the selection cuts for our samples at $n = 0.4, 1.6, 2.5$ and 7.0 distinguishing between low-Sérsic galaxies (blue shaded area), intermediate-Sérsic galaxies (green shaded area) and high-Sérsic galaxies (red shaded area), respectively. Galaxies outside this range are not considered for further scientific analysis (grey shaded areas). A vaguely indicated bimodality of the object distribution is quantified by a simultaneous fit of two Gauss curves to the main histogram (black curves). The black histogram shows the object distribution for a sub-sample of the galaxies in our sample with a brighter apparent magnitude than 21.2 mag (~ 30 per cent of the overall sample).

$n < 1.6$), intermediate ($1.6 < n < 2.5$) and high ($2.5 < n < 7.0$). Our internal bin limits are chosen to be equidistant in log-space in the regime $1 < n < 4$ in order to split the total population into at least two bimodal groupings focusing on the disc-like $n = 1$ population whilst the second should be focused on the spheroid-like $n = 4$ population. Because of the well-known overlap between these two Gaussian-like populations, we also add a transition ‘green-valley’ population, which would ensure that our $n = 1$ and 4 populations remained relatively free of contamination.

These bins are presented in Fig. 14 between the dashed lines. The probability density function for the distribution of Sérsic indices (see e.g. Kelvin et al. 2012, fig. 15) indicates a bimodal distribution of Sérsic types in the nearby Universe ($z < 0.25$). In their g -band sample, which corresponds roughly to our sample when the rest-frame wavelength is corrected for redshift, a bimodal distribution of Sérsic types occurs. A pronounced peak for galaxies with a low Sérsic index at $n \approx 1$ exists and a less pronounced peak is visible for galaxies with a high Sérsic index at $n \approx 3.5$. However, in our sample, only a hint for bimodal distribution of Sérsic profiles at $n \approx 1$ and ≈ 3.5 is indicated. Our statistic is quantified by a simultaneous fit of two Gauss functions to the number distribution (black curves). The first and most pronounced peak at $n \approx 1.09$ coincides well with the expected distribution. Interestingly, the position of the second Gauss distribution at $n \approx 3.16$ is slightly shifted towards lower Sérsic indices and much less pronounced. The reason for the shift of

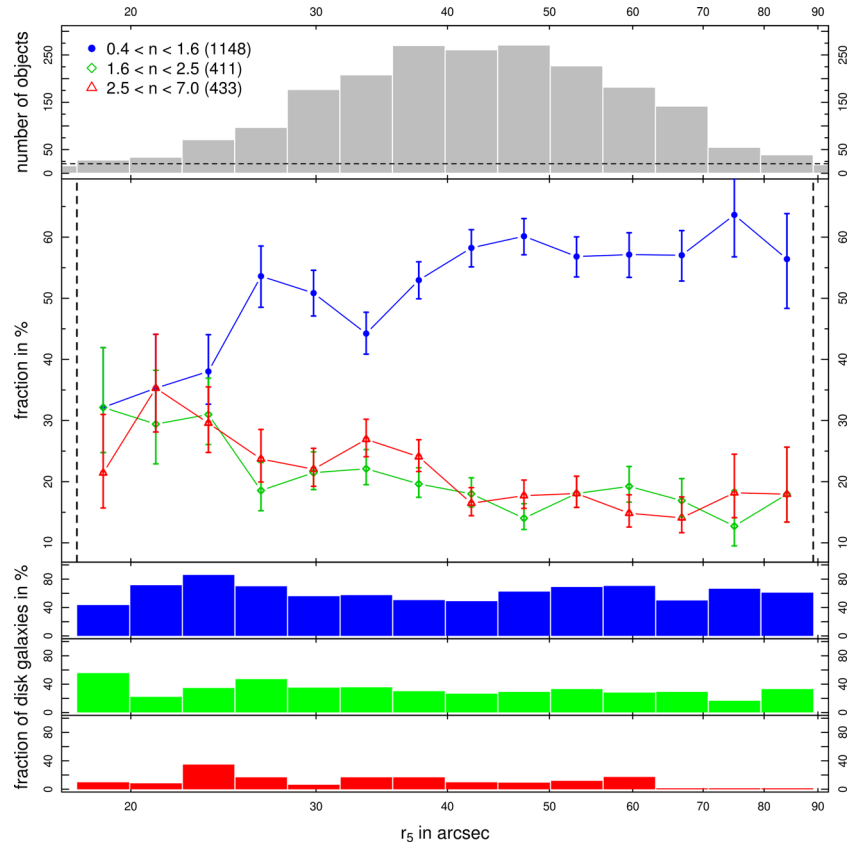


Figure 15. The relative fractions of Sérsic types and the total number of objects as a function of local object density. Objects with a high Sérsic index ($2.5 < n < 7.0$) are plotted in red, objects with an intermediate Sérsic index ($1.6 < n < 2.5$) in green and objects with a low Sérsic index ($0.4 < n < 1.6$) are blue. The local object density is estimated using an r_5 -criterion, representing the distance on the sky to the fifth next neighbour. Each population is binned using equal bins in \log_{10} -space with an exponent of 0.05. The upper plot shows the number of objects in each density bin. Bins with a number of objects below the black horizontal dashed line are discarded. Due to low number statistics ($N < 10$ objects) in the bins lower than 15 and higher than 90 arcsec, we discard these bins from the analysis and the lower plot. Note that also the relative fraction of Sérsic types in the leftmost bin might be biased by low number statistics. The lower three histograms show the fractions of galaxies with a disc in every Sérsic index bin.

the population with high Sérsic indices towards lower values may be the presence of dust within a significant fraction of the galaxies with a high Sérsic index in our sample. A well-studied effect of dust is the blocking of photons at short wavelengths from the core regions of galaxies. As a consequence, dust could bias the measured Sérsic indices towards lower values (Gadotti, Baes & Falony 2010; Pastrav et al. 2012). In addition, the reason for the observed differences to the distribution in Kelvin et al. (2012) is a mixture of several possible effects: first, Kelvin et al. (2012) use a probability density function in order to quantify the distribution of their sample in a smoothed fashion. Thus, the influence of statistical fluctuations might become more distinct in our distribution. Secondly, the morphologies of the galaxies and the distribution of the Sérsic index populations in our sample should be more affected by cluster-specific effects in comparison to field-dominated surveys. Thirdly, and the most important, the *HST* data set is very deep in comparison to ground-based surveys. The black histogram illustrates the same distribution of Sérsic types, but only for the brightest galaxies with an apparent magnitude which is brighter than 21.2 mag (~ 30 per cent of the sample). In this histogram, the distribution of galaxies with a low Sérsic index is much more even in comparison to the galaxies with a high Sérsic index. A comparison between the grey and the black histogram highlights that a significantly higher fraction of galaxies with a low Sérsic index is missing in comparison to the galaxies with a high Sérsic index. The result is that galaxies with Sérsic indices in

the range of $n = 1$ are on average fainter, dimmer and thus harder to detect than comparable objects with a high Sérsic index. Hence, the visible relative fraction of galaxies with low Sérsic indices in the range of $n = 1$ increases in deep *HST* surveys. As a consequence, there exists a broad, prominent peak, which dominates the overall distribution at $n = 1$ in the histogram in Fig. 14.

For each of the 2063 cluster members in our data set, we calculate the r_5 -criterion, the apparent distance to the fifth next neighbour amongst the identified galaxy-cluster members, to statistically estimate the density of the environment. Due to low number statistics in the bins with very high object density ($r_5 < \approx 17.78$ arcsec) and in the bins very low object density ($r_5 > \approx 89.13$ arcsec), further 71 objects are discarded from the analysis. In total, 1992 galaxies are included in Figs 15 and 17 (~ 87.4 per cent of the total sample of cluster galaxies). Objects are binned according to the calculated r_5 -distance using equal bins in \log_{10} -space with an exponent of 0.05. As a result, the calculated r_5 -distances are binned into 20 bins between 10 and 100 arcsec.

7.3 Results

7.3.1 The Sérsic index–density relation

Fig. 15 shows the relative fraction of Sérsic types as a function of local object density. The top panel shows a histogram for the

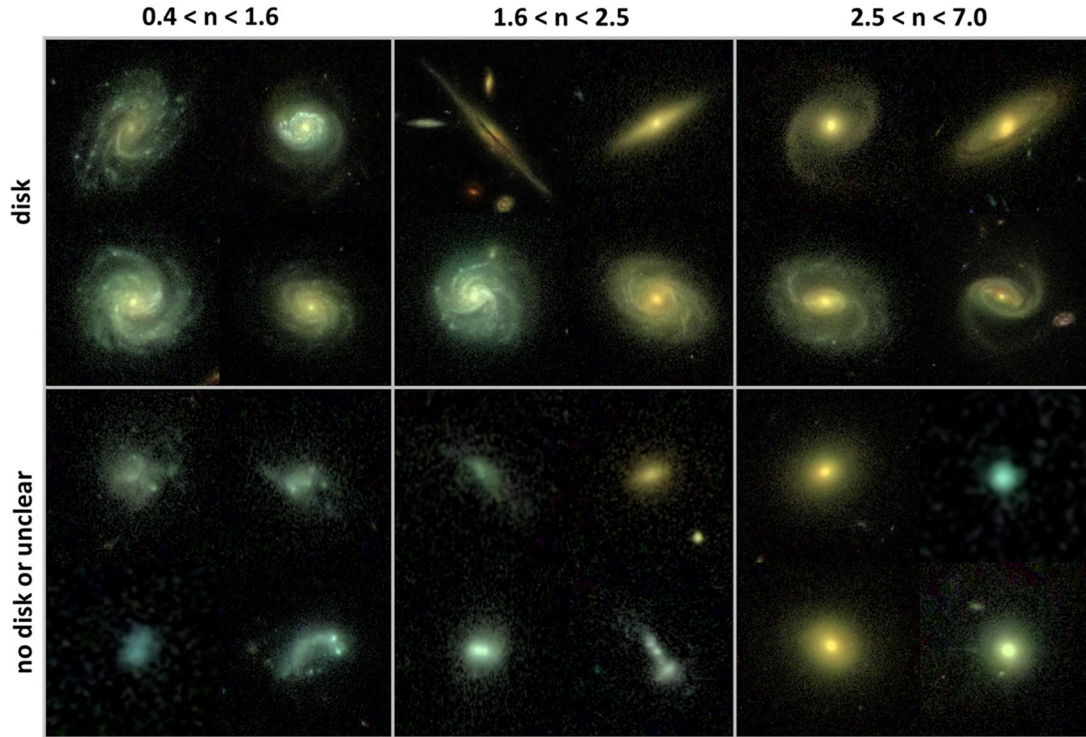


Figure 16. Colour postage stamp examples for 24 galaxies in the STAGES data set. Galaxies with a low Sérsic index ($0.4 < n < 1.6$) are shown on the left, galaxies with an intermediate Sérsic index ($1.6 < n < 2.5$) are shown in the middle and the right-hand panels show galaxies with a high Sérsic index ($2.5 < n < 7.0$). The upper row specifies objects which we visually classified as galaxies which host a disc and galaxies without a disc or unclear cases are shown below. Each postage stamp varies with angular size. The angular size is chosen to be approximately of the size of the Kron ellipse for each galaxy accordingly in order to zoom in on each object for the visual classification. If a galaxy is too small to see a disc (e.g. the galaxy in lowest left image), it is classified as a galaxy without a disc.

objects in every r_5 -bin using equal bins in \log_{10} -space with an exponent of 0.05. The bottom panel shows the fraction of Sérsic types in every bin. The distribution of low-Sérsic types is plotted in blue, intermediate-Sérsic types in green and high-Sérsic types in red. Error bars indicate $\pm 1\sigma$ binomial confidence intervals (Cameron 2011). The population in denser cluster environments close to the cluster centre (left-hand side of the figure) shows a relatively higher fraction of high (red) and intermediate Sérsic index objects (green) than in the less dense environments (right-hand of the figure). These two populations halve in fraction until an average r_5 -value of ~ 45 arcsec, beyond which their number fractions remain roughly constant at a relatively low value of slightly below 20 per cent. Interestingly, the population of intermediate Sérsic types (green data points) follows the trend of the high Sérsic population at all densities. This intermediate Sérsic index population, however, should not be considered to be analogous to a lenticular S0 population in the sense of the morphology–density relation as presented by Dressler (1980) as it will also contain other morphological types e.g. bulge–disc composite galaxies. The fraction of galaxies with low Sérsic indices (blue data points) increases going towards sparse environments before remaining at a constantly high level of nearly 60 per cent. Hence, low-density environments at large cluster radii seem to be statistically dominated by galaxies with a low Sérsic index. Overall, our result coincides well with the morphology–density relation.

7.3.2 The Sérsic index–density relation for disc galaxies

We extracted colour postage stamps from the original STAGES image tiles in order to classify every object visually into galaxies

with and without a disc. For the construction of the postage stamps, the object luminosities in the STAGES survey are combined with additional colour information from the COMBO-17 survey in the same field of view. In order to be classified as a galaxy with a disc, the object must fulfil *at least* one of the following criteria:

- (i) a flattened, extended disc is clearly visible to the observer,
- (ii) at least one spiral arm is visible to the observer,
- (iii) a dust lane exists.

In any other case, the galaxy is classified as a galaxy without a disc.

Fig. 16 shows example postage stamps for 24 galaxies with disc galaxies in the upper panels and galaxies without a disc in the lower panels. The columns indicate the Sérsic index bins from low (left) to high (right). We find that the sample of disc galaxies with a low Sérsic index is dominated by late-type systems. The sample of disc galaxies with a high Sérsic index is dominated by early-type disc galaxies. The sample of objects without a disc and a low Sérsic index mainly consists of irregular galaxies. Conversely, in the sample of galaxies without a disc and a high Sérsic index, we mainly observe elliptical galaxies as well as dwarf ellipticals. Both in the sample of galaxies with a disc as well as in the sample of galaxies without a disc, the observed objects with an intermediate Sérsic index can be seen as a transition state and a mixture of both populations.

The fraction of galaxies with a disc is presented in the lower three histograms in Fig. 15 for every Sérsic index bin, respectively. The histograms show the fraction of disc galaxies with a low Sérsic index in blue, the fraction of disc galaxies with an intermediate Sérsic index are shown in green and the red histograms represent the fraction of galaxies with a disc in the high Sérsic index bin. We find that the majority of galaxies with a low Sérsic index (~ 62 per cent) have

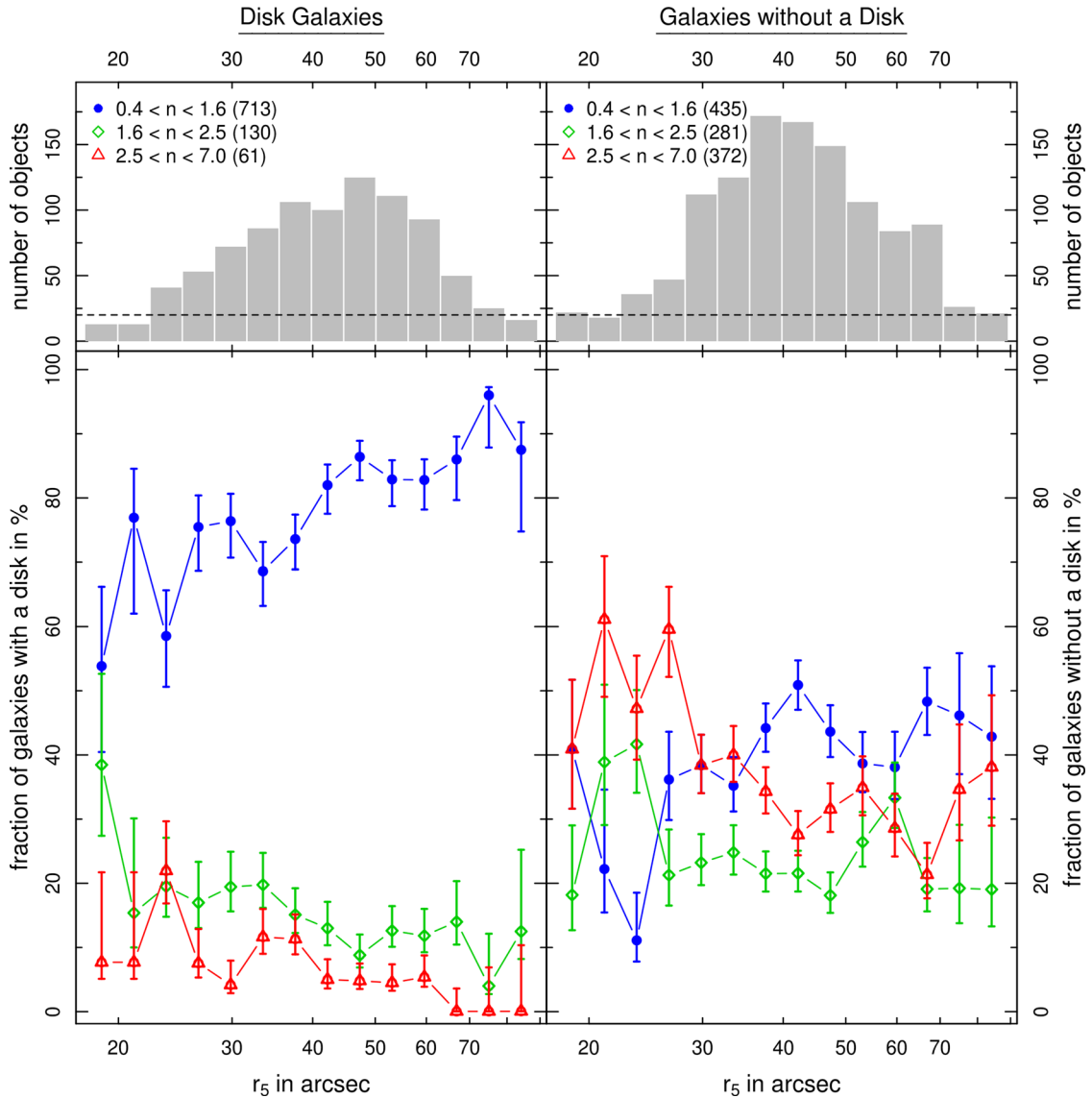


Figure 17. The relative fractions Sérsic index types as a function of local object density for the galaxies with a disc (left) and for galaxies without a disc (right). The same criteria as in Fig. 15 are applied for a sub-sample of 904 galaxies with a disc (~ 45.4 per cent) and 1088 galaxies without a disc (~ 55.6 per cent). Note that the leftmost two bins and the rightmost bin might be biased by low number statistics.

a disc. The fraction of disc galaxies decrease strongly for galaxies with higher Sérsic indices. Only ~ 32 per cent of the galaxies with an intermediate Sérsic index and ~ 14 per cent of the galaxies with a high Sérsic index are observed to have a disc. In addition, we find that the overall average fraction of disc galaxies increases from regions with a high local object density towards sparse regions (~ 35 per cent in dense regions and ~ 60 per cent in sparse regions). At the same time, interestingly, the fractions of disc galaxies in each of the Sérsic index bins do not show significant trends. During our visual classifications, we observe that the majority of disc galaxies with a low Sérsic index consists of late-type systems. On the other side, the majority of disc galaxies with an intermediate or a high Sérsic index are early-type disc systems and are dominated by a bright bulge and/or a bar in the centre. In addition, every sample contains a significant fraction of unusually small objects and dwarf galaxies without a disc. Together with irregular galaxies, these objects contribute in lowering the overall disc fractions in every Sérsic index bin.

Fig. 17 shows the relative fractions of Sérsic index types as a function of local object density in the same way as in Fig. 15, but the plot distinguishes between a sub-sample of galaxies with a disc and a sub-sample of galaxies without a disc. The histograms in the top panels represent the distribution of the galaxies in equivalent r_5 -bins. In the lower panels, the fraction of Sérsic types in every bin is shown again in the same way. This is done for galaxies with a disc (left) and for galaxies without a disc (right).

The population of disc galaxies (left) shows similar trends in comparison to the overall sample, but the relative fractions of Sérsic index types are more bimodal. We find a higher fraction of galaxies with a low Sérsic index and a lower fraction of galaxies with intermediate or high Sérsic indices. In dense regions, ~ 60 – 70 per cent amongst the galaxies with a disc have a low Sérsic index, ~ 25 per cent are galaxies with an intermediate Sérsic index and ~ 10 per cent have a high Sérsic index. The fraction of galaxies with a lower Sérsic index constantly increases until at an average r_5 -value of ~ 45 arcsec and stays roughly constant at ~ 80 – 90 per cent

in sparser regions. The other way round, the population of galaxies with intermediate and high Sérsic indices constantly decrease from dense regions towards sparse regions. Beyond an average r_5 -value of ~ 45 arcsec, the fraction of disc galaxies with an intermediate Sérsic index is lower than 15 per cent and the fraction of disc galaxies with a high Sérsic index nearly vanishes. The population of galaxies without a disc is shown in the right-hand panel. The trends are still in agreement with our previous findings. The fraction of systems with a low Sérsic index still increases towards sparse regions and the fraction of systems with higher Sérsic indices decreases. One might expect that the population of galaxies without a disc should be dominated by objects with a high Sérsic index. However, this is only the case in regions with a high density. As the total sample is dominated by galaxies with a low Sérsic index (1148 objects), the remaining number of objects with low Sérsic index in the sample without a disc (435 objects) is still high in comparison to the number of galaxies with higher Sérsic indices. The fraction of galaxies without a disc and a low Sérsic index mainly consists of irregular galaxies. The result is that the cumulative distribution of Sérsic index types in the sub-sample of galaxies without a disc is relatively even and flat.

We find a distinct correlation between the relative fractions of galaxies with a disc and the Sérsic indices of a galaxy population. With a higher fraction of disc-dominated systems, the fraction of galaxies with a low Sérsic index increases. In addition, our results support the well-known findings that late-type disc-dominated systems are usually well approximated with a low Sérsic index $0.4 < n < 1.6$, if the light contribution of the bulge is low. For early-type galaxies with a disc, a bright bulge and/or when a bar component becomes pronounced, the Sérsic index increases on average. Elliptical and bulge-dominated systems are usually well approximated with a higher Sérsic indices. Irregular galaxies and unusual objects (e.g. Green Peas) occur in a wide range of Sérsic indices and lower the relative fraction of disc galaxies in each sample.

8 SUMMARY

We present a new version of the survey image processing pipeline GALAPAGOS (Barden et al. 2012), which we term GALAPAGOS-C. GALAPAGOS-C is designed for processing large imaging data sets automatically based on a uniform single set-up and without further user interaction. While it is based on GALAPAGOS, it incorporates many more features and is heavily optimized for computing speed and efficiency as well as fitting accuracy. After applying SExtractor (Bertin & Arnouts 1996) to detect astronomical sources, GALAPAGOS-C automatically removes multiple entries in overlapping tiles, cuts postage stamps, estimates a local sky background and prepares an object mask for every object. After this preparation and fitting set-up, single-Sérsic fitting is performed using GALFIT (Peng et al. 2010a). The resulting fit parameters are stored in a FITS table.

GALAPAGOS-C is written in C and uses MPI-parallelization to be able to process large data sets on modern supercomputers. GALAPAGOS-C can distribute the actual GALFIT fitting jobs on a large number of CPU cores according to a so-called master--slave concept. The code is optimized for both speed and stability. A new aspect thereby is that the workload of the master process is constructed to be as low as possible and the communication amongst the processors is highly optimized. As a result, GALAPAGOS-C can be applied for supercomputers with a large number of CPU cores without risking computational overhead. In addition, the user must determine an

appropriate number of CPU cores for applying GALFIT to every object on a supercomputer in terms of efficiency. This number strongly depends on the number of detected objects, the object density and the quality of the survey data relating to image defects and saturated stars, but in practice does not influence the output quality of the data.

Additionally, this new version integrates GALFIT features for fitting asymmetrical distortions by applying a Fourier mode expansion on elliptical isophotes, allowing studies of morphological distortions of galaxies and their intensities due to various intrinsic and environmental effects. With this new MPI-parallelized version of GALAPAGOS-C, the feature of multicomponent Sérsic fitting has now been implemented but will be presented in a future paper. This allows the fitting of bulges, discs, bars and further components and will allow the wider scientific community to gain a deeper insight into galaxy formation and evolution across a diverse range of galaxy morphological types.

GALAPAGOS-C is controlled by a single, simple set-up file with a similar structure to the prior version of GALAPAGOS in order to control all features of the code in a cleanly arranged manner. Based on very similar concepts as the IDL version of GALAPAGOS, its wide area of applications on extragalactic surveys is still given independently on ground- or space-based data sets with a stable PSF. The user must provide a good PSF for the entire survey region before running GALAPAGOS-C.

We have tested the performance of single-Sérsic fitting for GALAPAGOS-C using artificial image simulations with Sérsic light profiles covering a wide and realistic parameter range. We also compare the outputs of GALAPAGOS-C to comparable results from GALAPAGOS finding very good agreement across an extensive parameter space.

However, the versatility of GALAPAGOS-C could be further increased by implementing further recent development, e.g. multiband fitting (MegaMorph, etc.; Häußler et al. 2013) and by implementing usage of a variable PSF for every object. In this new version, GALFIT is still provided with just one image for each objects and a single PSF for the entire data set. Software for this procedure is now given by PSFEX (Bertin 2011), which extracts model PSFs from large FITS images in a sophisticated manner. Here, GALAPAGOS-C could benefit from a combination of data from PSFEX with GALAPAGOS-C in order to implement PSFs ‘on the fly’, similar to the procedure in other codes (Kelvin et al. 2010).

In this paper, we apply GALAPAGOS-C to a data set consisting of $\approx 81\,000$ galaxies including 2279 galaxies in the A901/902 cluster environment. Through the analysis of 1992 single-Sérsic fits to each confirmed cluster member, we provide one of the first estimates of the Sérsic index–density relation. We find that the fraction of galaxies with low Sérsic indices ($0.4 < n < 1.6$) increases going towards sparse environments before remaining at a constantly high level (from ~ 35 to ~ 60 per cent). The population in denser cluster environments closer to the cluster centre shows a relatively higher fraction of high Sérsic index ($2.5 < n < 7.0$) and intermediate Sérsic index ($1.6 < n < 2.5$) galaxies (~ 35 per cent) than in sparse environments (~ 15 per cent). This result confirms and completes the findings of Dressler (1980) and many others.

In addition, this data set is visually classified into galaxies with and without a disc. We find that at least 60 per cent of galaxies with a low Sérsic index have a disc, ~ 30 per cent of the galaxies with an intermediate Sérsic index and only ~ 15 per cent of the galaxies with a high Sérsic index have a disc. We find that these fractions are relatively constant across all environmental object densities.

GALAPAGOS-C is freely available for download from the astrophysical code library at <http://ascl.net/1408.011>.

ACKNOWLEDGEMENTS

We thank the referee for the constructive comments which helped a lot to improve the manuscript. AH, LK and SS are grateful for funding by the Austrian Funding Organization FWF through grant P23946–N16. This work was supported by the Austrian Ministry of Science BMWF as part of the UniInfrastrukturprogramm of the Focal Point Scientific Computing at the University of Innsbruck. We thank Asmus Böhm for providing a sample of local PSF from the STAGES survey for the PSF stability test in Appendix B. We thank the supervisor of the local Astro–Cluster supercomputer at the University of Innsbruck, Josef Schafer, for the competent support.

REFERENCES

- Abazajian K. et al., 2003, *AJ*, 126, 2081
- Amorín R. O., Vílchez J. M., Pérez-Montero E., Papaderos P., 2013, *Rev. Mex. Astron. Astrofis. Ser. Conf.*, 27, 49
- Athanassoula E., Machado R. E. G., Rodionov S. A., 2013, *MNRAS*, 429, 1949
- Baldry I. K., Balogh M. L., Bower R. G., Glazebrook K., Nichol R. C., Bamford S. P., Budavari T., 2006, *MNRAS*, 373, 469
- Bamford S. P. et al., 2009, *MNRAS*, 393, 1324
- Barden M., Häußler B., Peng C. Y., McIntosh D. H., Guo Y., 2012, *Astrophysics Source Code Library*, record ascl:1203.002
- Bekki K., Forbes D. A., Beasley M. A., Couch W. J., 2002, *MNRAS*, 335, 1176
- Bertin E., 2011, in Evans I. N., Accomazzi A., Mink D. J., Rots A. H., eds, *ASP Conf. Ser. Vol. 442, Astronomical Data Analysis Software and Systems XX*. Astron. Soc. Pac., San Francisco, p. 435
- Bertin E., Arnouts S., 1996, *A&AS*, 117, 393
- Blanton M. R., Eisenstein D., Hogg D. W., Schlegel D. J., Brinkmann J., 2005, *ApJ*, 629, 143
- Böhm A. et al., 2013, *A&A*, 549, A46
- Bösch B. et al., 2013, *A&A*, 549, A142
- Calvi R., Poggianti B. M., Fasano G., Vulcani B., 2012, *MNRAS*, 419, L14
- Cameron E., 2011, *PASA*, 28, 128
- Cardamone C. et al., 2009, *MNRAS*, 399, 1191
- Davari R., Ho L. C., Peng C. Y., Huang S., 2014, *ApJ*, 787, 69
- de Jong R. S., 1996, *A&AS*, 118, 557
- Dressler A., 1980, *ApJ*, 236, 351
- Dressler A. et al., 1997, *ApJ*, 490, 577
- Driver S. P., 1999, *ApJ*, 526, L69
- Driver S. P. et al., 2009, in Andersen J., Nordström B., Bland-Hawthorn J., eds, *Proc. IAU Symp. 254, The Galaxy Disk in Cosmological Context*. Cambridge Univ. Press, Cambridge, p. 469
- Gadotti D. A., 2008, *MNRAS*, 384, 420
- Gadotti D. A., de Souza R. E., Dos Anjos S., 2003, *Bull. Astron. Soc. Brazil*, 23, 18
- Gadotti D. A., Baes M., Falony S., 2010, *MNRAS*, 403, 2053
- Gilmour R., Gray M. E., Almaini O., Best P., Wolf C., Meisenheimer K., Papovich C., Bell E., 2007, *MNRAS*, 380, 1467
- Gray M. E. et al., 2009, *MNRAS*, 393, 1275
- Grogin N. A. et al., 2011, *ApJS*, 197, 35
- Gunn J. E., Gott J. R., III, 1972, *ApJ*, 176, 1
- Haeussler B., Bamford S., Vika M., Rojas A., 2011, in Evans I. N., Accomazzi A., Mink D. J., Rots A. H., eds, *ASP Conf. Ser. Vol. 442, Galapagos: A Semi-Automated Tool for Galaxy Profile Fitting*. Astron. Soc. Pac., San Francisco, p. P73
- Häussler B. et al., 2007, *ApJS*, 172, 615
- Häußler B. et al., 2013, *MNRAS*, 430, 330
- Hubble E. P., 1927, *The Observatory*, 50, 276
- Jahnke K. et al., 2004, *ApJ*, 614, 568
- Kauffmann G., White S. D. M., Heckman T. M., Ménard B., Brinchmann J., Charlot S., Tremonti C., Brinkmann J., 2004, *MNRAS*, 353, 713
- Kelvin L., Driver S., Robotham A., Hill D., Cameron E., 2010, in Debattista V. P., Popescu C. C., eds, *AIP Conf. Proc. Vol. 1240, Hunting for the Dark: The Hidden Side of Galaxy Formation*. Am. Inst. Phys., New York, p. 247
- Kelvin L. S. et al., 2012, *MNRAS*, 421, 1007
- Laurikainen E., Salo H., Buta R., 2005, *MNRAS*, 362, 1319
- Maltby D. T., Hoyos C., Gray M. E., Aragón-Salamanca A., Wolf C., 2012, *MNRAS*, 420, 2475
- Moore B., Katz N., Lake G., Dressler A., Oemler A., 1996, *Nature*, 379, 613
- Pastrav B. A., Popescu C. C., Tuffs R. J., Sansom A. E., 2012, in Tuffs R. J., Popescu C. C., eds, *Proc. IAU Symp. 284, The Spectral Energy Distribution of Galaxies*. Cambridge Univ. Press, Cambridge, p. 306
- Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2010a, *AJ*, 139, 2097
- Peng Y.-j. et al., 2010b, *ApJ*, 721, 193
- Pimbblet K. A., Jensen P. C., 2012, *MNRAS*, 426, 1632
- Postman M., Geller M. J., 1984, *ApJ*, 281, 95
- Quilis V., Moore B., Bower R., 2000, *Science*, 288, 1617
- Rhodes J. D. et al., 2007, *ApJS*, 172, 203
- Rix H.-W. et al., 2004, *ApJS*, 152, 163
- Rowlands K. et al., 2012, *MNRAS*, 419, 2545
- Schawinski K. et al., 2009, *BAAS*, 41, 277
- Schwarz G., 1978, *Ann. Stat.*, 6, 461
- Sérsic J. L., 1963, *Bol. Asociacion Argentina Astron. La Plata Argentina*, 6, 99
- Sérsic J. L., 1968, *Bull. Astron. Inst. Czech.*, 19, 105
- Simard L. et al., 2002, *ApJS*, 142, 1
- Simard L., Mendel J. T., Patton D. R., Ellison S. L., McConnachie A. W., 2011, *ApJS*, 196, 11
- Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, *MNRAS*, 328, 726
- Steinhauser D., Haider M., Kapferer W., Schindler S., 2012, *A&A*, 544, A54
- Toomre A., Toomre J., 1972, *ApJ*, 178, 623
- van der Wel A., 2008, *ApJ*, 675, L13
- Vulcani B. et al., 2012, *MNRAS*, 420, 1481
- Wolf C., Meisenheimer K., Rix H.-W., Borch A., Dye S., Kleinheinrich M., 2003, *A&A*, 401, 73
- York D. G. et al., 2012, *MNRAS*, 423, 3692

APPENDIX A: BACKGROUND ESTIMATION TEST

A robust value for the local sky level at the position of a source is the most important basis for correct fitting results. Therefore, we intensively tested the sky estimation routine which is incorporated

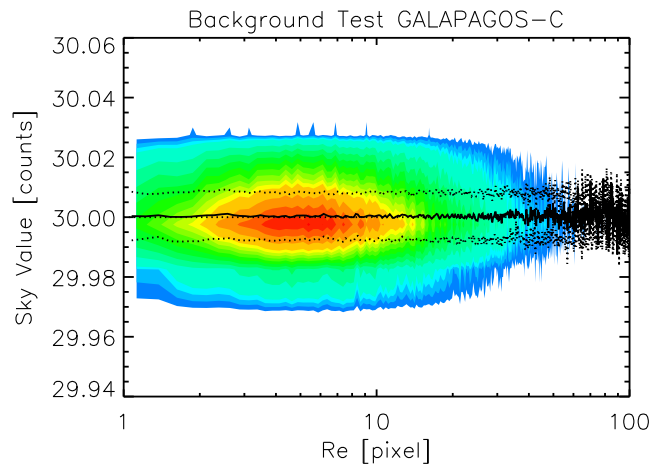


Figure A1. Test of GALAPAGOS-C sky estimation routine on an artificial image simulation. The true value is 30.00 counts. The detected values are plotted as a function of object size measured in half-light radii (r_e). The dashed line indicates the average value in bins of 0.25 pixel. The dotted lines show the 1σ deviations.

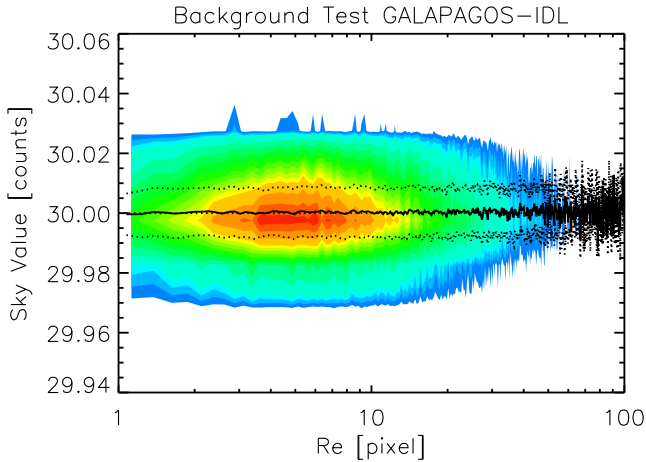


Figure A2. Test of GALAPAGOS IDL sky estimation routine on an artificial image simulation. The true value is 30.00 counts. The detected values are plotted as a function of object size measured in half-light radii (r_e). The dashed line indicates the average value in bins of 0.25 pixel. The dotted lines show the 1σ deviations.

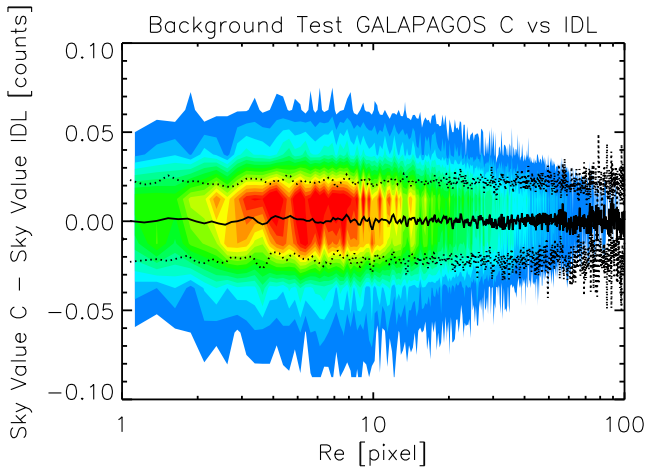


Figure A3. Comparison of the GALAPAGOS-C sky estimation routine with the GALAPAGOS in IDL routine using 80 image tiles from the STAGES survey images. The detected values by GALAPAGOS in IDL are subtracted from the values measured by GALAPAGOS-C and are plotted as a function of object size measured in half-light radii (r_e) by GALAPAGOS-C. The dashed line indicates the average value in bins of 0.25 pixel. The dotted lines show the 1σ deviations. The plot shows that both codes perform equally well, independent of the primary objects size.

in GALAPAGOS-C and compared our results to the values given by the well-tested routine from former version of GALAPAGOS in IDL. We applied both routines to a completely artificial data set. We used the data from image tile 55 in the STAGES survey which is the tile with the highest density of galaxies in the whole survey and thus the tile with the most contributors (for a definition see Section 4.4.1 ‘Simultaneous fits’) and measured the average global sky level. For this purpose, we used the GALAPAGOS-C skymap, where every detected source was masked out. The mask was constructed using expanded SExtractor Kron ellipses by expanding their semiminor and -major axis by a factor of 3. The average sky value was then measured using 3σ clipping for the remaining pixels. A 1σ value for this distribution was obtained as well. The mean values for sky in the image tile was measured and rounded to 30.00 counts and the average value for 1σ 1.90.

We placed a single Sérsic profile at the position of every detected source with the parameters measured by GALAPAGOS and added Poisson noise. We added a sky background that shows the same Gauss distribution for the same values as above. Varying the single-Sérsic parameters for every detected source within its measurement errors, we constructed 80 different tiles with the same sky behaviour, but slightly different galaxy populations. We applied both GALAPAGOS-C and GALAPAGOS to this data to measure the local sky background for every source. Figs A1 and A2 show the distributions of the measured sky values for 81 365 objects. We see good agreement with the simulated sky values, no systematic trends and no significant difference in the values measured by both codes.

In order to compare the data output of both GALAPAGOS-C and GALAPAGOS, we subtracted the measured sky values from real STAGES images measured by both codes. This was done for 81 340 detected objects in 80 image tiles. Fig. A3 shows again a density plot for the subtracted values for every object. We see no systematic trends and no significant difference in the values measured by both codes.

APPENDIX B: PSF STABILITY TEST

The PSF of *HST* ACS is very stable in time in comparison to ground-based telescopes. Significant changes in the shape of the PSF arise from astigmatism, coma and/or defocus in the field of view, which can be caused by height variations of the two CCDs or changes in the focal length of the *HST* (Jahnke et al. 2004). Rhodes et al. (2007) amongst others investigated the temporal and the spatial stability of the *HST* Advanced Camera for Surveys Wide Field Camera PSF based on COSMOS survey data. It was found that the PSF undergoes low temporal and spatial instabilities, leading to a slow periodic focus change in the *HST* image data set.

However, for the purpose of including one single overall PSF for every source in the STAGES data set, we measure the relative impact of the PSF instabilities on the fitting results. For example, Davari et al. (2014) show that that size determination of galaxies in *HST* surveys is in practical not affected from PSF instabilities within a realistic range of PSFs. It is shown that *HST* PSFs are not even limited to specific survey applications. In order to quantify the variations, we use six different PSFs from the STAGES survey. PSF 1 (standard PSF, see Fig. B1, left) was used for every application within this paper. This PSF is averaged from a large number of global point sources over the entire ACS CCD. The remaining five PSFs are averaged from 35 stars which are located within 40 arcsec at a given CCD position. Thus, they represent five different places at the ACS CCD. One PSF is located in the centre of the CCD and the remaining four PSFs sample the four corners of the CCD at a distance of ~ 1500 pixels from the CCD border, respectively. The PSFs have been constructed according to the description by Böhm et al. (2013). This means that the standard PSF, in theory, should be more robust globally, whilst the latter five PSFs are more appropriate for detailed local studies using *HST* imaging. We used 2D Gauss fits in order to quantify the given PSFs. The measured full width at half-maxima (FWHM) are shown in Table B1 for the six PSFs.

We apply this set of PSFs to 10 arbitrarily selected STAGES image tiles (02, 09, 16, 24, 35, 41, 49, 55, 60, 72) which contain 10 970 objects in total. The single-Sérsic fitting results with the five local PSFs are compared to the results from the fitting with standard PSF.

The result is presented in Fig. B2. We find that the used, averaged PSF shows no significant trends in comparison to the other

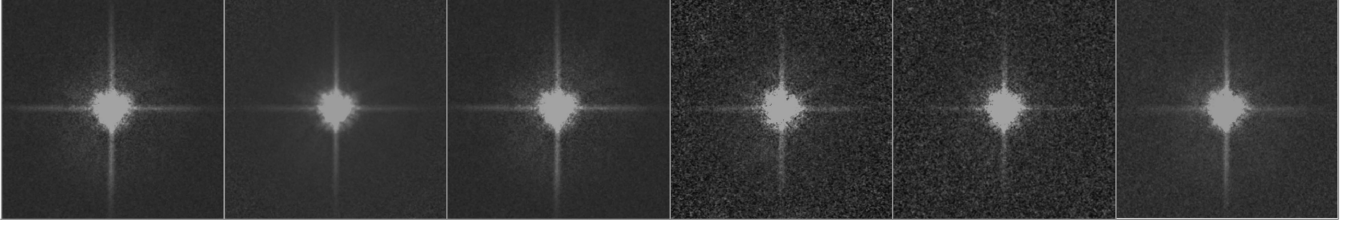


Figure B1. Six different PSF from the *HST* STAGES survey. The left PSF is an averaged superposition from a large number of point sources across the whole ACS CCD. The remaining five PSFs are averaged superpositions from 35 point sources around specific positions at the ACS CCD. The left PSF (standard PSF) is used by GALAPAGOS-C for every application in this paper. The remaining five PSFs provide GALFIT with different probes in order to test the overall PSF stability of the survey.

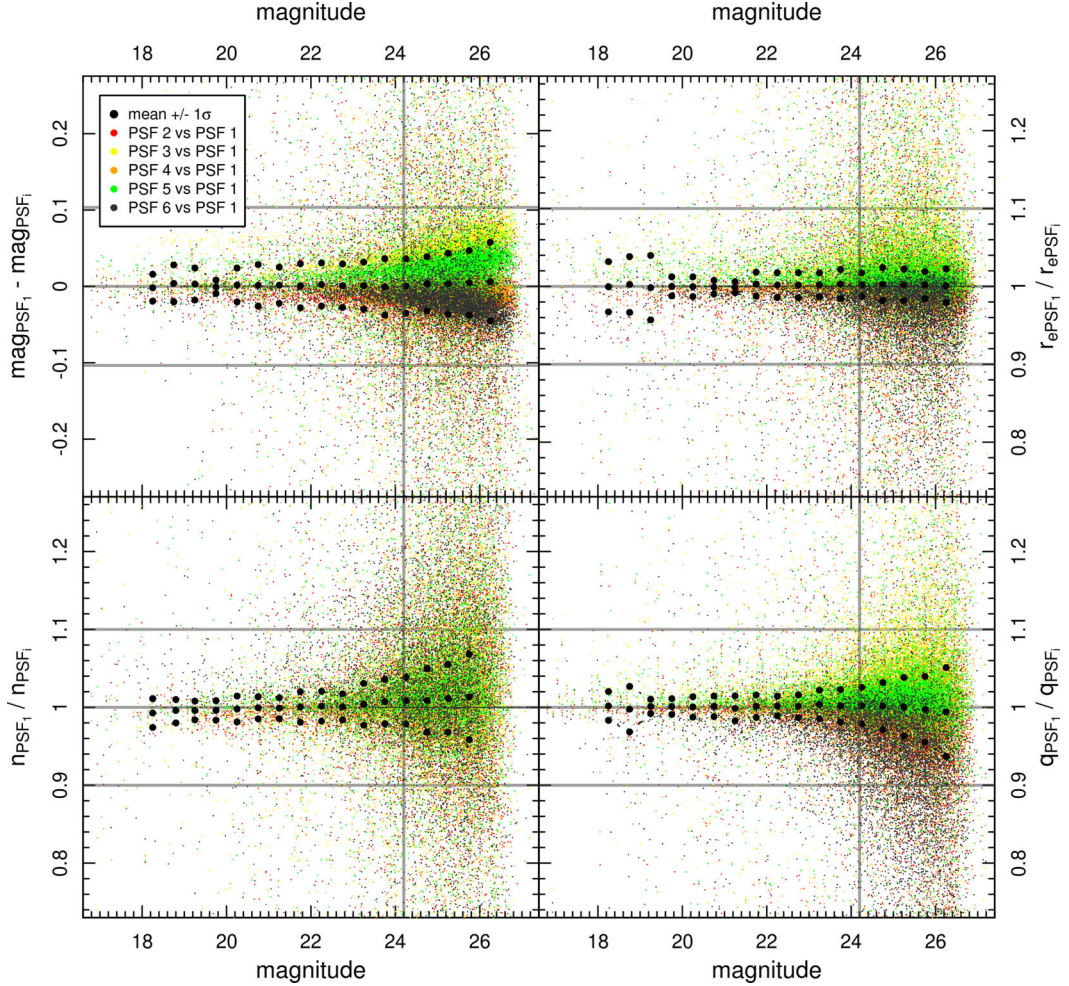


Figure B2. Fitting results for 10 970 objects in the STAGES survey. First every object has been fitted by GALFIT using the standard, global PSF (see Fig. B1, left). Next, every object has been fitted with the remaining five (local) PSFs in Fig. B1. Every single-Sérsic profile parameter (magnitude, half-light radius, Sérsic index, axial ratio) which is measured with the standard PSF is compared with the profile parameter which is measured using the remaining five PSFs. The black, dotted lines show the mean value and the $\pm 1\sigma$ deviations of the point distribution for the deviation for every parameter. The colours indicate the fitting results for every PSF in comparison with the standard PSF. The horizontal, grey lines indicate a deviation of 10 per cent from the parameter which was measured using the standard PSF. The vertical, grey lines margin the scientifically relevant magnitude space down to an apparent brightness limit of ~ 24.2 mag.

PSFs. However, the Sérsic profile parameters of other PSFs show slight trends as expected due to the different FWHM of the PSFs (see Table B1). For example, PSF 5 tends to overestimate magnitudes, half-light radii and axial ratio in comparison to the standard PSF, whereas PSF 6 tends to systematically underestimate the said parameters. However, Fig. B2 shows that the variations in the

scientifically relevant parameters space for objects brighter than 24.2 mag are clearly lower than 5 per cent for all Sérsic profile parameters. We conclude that Sérsic profile parameter variations, which are induced by temporal and spatial PSF variations, are low enough that one single PSF can be applied for the complete STAGES data set.

Table B1. FWHM (in pixel) of the tested PSF.

PSF	<i>x</i> FWHM	<i>y</i> FWHM
PSF1	3.89	3.91
PSF2	3.92	3.93
PSF3	3.78	3.83
PSF4	3.80	3.85
PSF5	3.79	3.86
PSF6	4.12	4.07

APPENDIX C: IDENTIFY ERROR SOURCES AND REMEDY IN GALAPAGOS-C

Due to the eternally wide range of object morphologies, no contemporary fitting algorithm is able to provide perfect profile parameters for every source in the data set. Throughout this paper is shown that GALAPAGOS-C provides good profile approximations for a large fraction of objects. The following enumeration (in the order of the code steps) might help the user to understand the circumstances of error sources and failed steps in GALAPAGOS-C. It provides clues on remedy and helps the user to make sensible choices for some set-up parameters.

C1 Detection failures

The fundamental operation step in GALAPAGOS-C is the object detection with SExtractor. The SExtractor manual provides useful information on setting up the code and well-tested starting parameters for *HST* survey data can be found in the example set-up files of GALAPAGOS-C.

It has been shown that a two-step set-up configuration (cold and hot) provides robust object detections for many applications. Irrespective of the number of set-up configurations, the user is provided with a combined check-image from the HDR runs in every image tile in order to control the output of the HDR detection procedure. Still, it can happen that nearby but separate objects are not detected as stand-alone sources. In addition, it is possible that the two nearby objects are detected as stand-alone source, e.g. the brighter one is detected in the cold run and the fainter one is detected in the hot run. But as the fainter source is located too close to the brighter source, the fainter source might be erased mistakenly and not enter the catalogue. As a result, the parameters of the missed source are

missing and secondly, the fitting procedure for the brighter source can be biased due to the missing neighbouring profile. Both have impact on further object profiling. Fig. C1 (left-hand image) shows an example: a bright elliptical galaxy is closely located next to two much smaller galaxies below. These two galaxies are visible on the residual images (right-hand three images). The two small neighbours have only been detected in the hot SExtractor run. As the central positions of the two neighbouring companions are located inside the Kron ellipse of the bright elliptical and the enlarging factor ('enlarge') for erasing hot mode detections was >1 , they have been removed from the detection catalogue mistakenly.

In order to overcome detections failures, the user has several options. First, the user should check whether the SExtractor set-up file(s) are as accurate as possible for the present survey data set. The SExtractor manual and the additional starting guide 'SExtractor for Dummies' (Dr Benne W. Holwerda) present useful information on providing well-suited set-up parameters. Secondly, it is possible to change the set-up parameter 'enlarge' to a slightly lower value in order to erase a lower fraction of very close, neighbouring sources. In this case, the user must take special care and check if the wide majority of spurious hot detections is still erased correctly. In addition, it could make sense to add another SExtractor set-up between the 'cold' and the 'hot' configuration in order to refine the detection procedure. The more SExtractor set-ups are used the more it makes sense to slightly lower the value of the set-up parameter 'enlarge'.

C2 Intracuster light and sky bias

Several circumstances can lead to situations where the local sky background is not flat around the object of interest. GALAPAGOS-C is shown to provide robust sky values in most of the cases, but e.g. the presence of intracuster light can significantly bias the local sky estimation procedure towards too high values. In addition, GALAPAGOS-C does not include any internal routine in order to overcome fitting biases due to present intracuster light on a postage stamp. The result is that this light can be included into the fitting profiles and thus magnitudes, object sizes, Sérsic indices or Fourier mode amplitudes can be overestimated and other fitting results are significantly influenced.

In the case that intracuster light is present for a significantly large fraction of objects in the survey, GALAPAGOS-C must be used with caution and sky values, profile parameters and residual images must

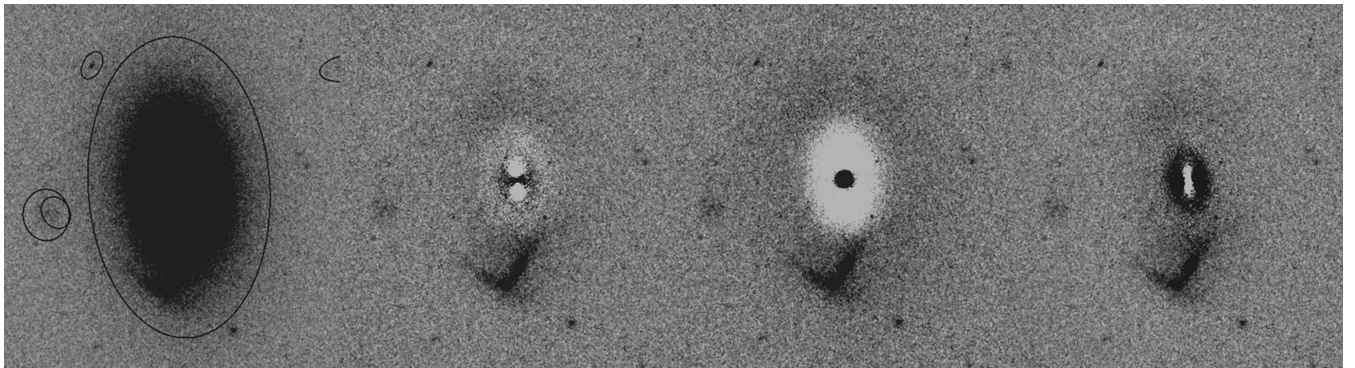


Figure C1. An elliptical galaxy with a Sérsic index of $n \approx 4$ (left). The galaxy profile is approximated with a single-Sérsic profile. Every detected source in the neighbourhood with a possible overlap is downgraded from secondary to tertiary due to its comparatively low magnitude. The leftmost residual image shows the fitting result for the true Sérsic profile with a Sérsic index of 4. In the middle residual image, the Sérsic index is artificially underestimated ($n = 2$). The right image shows a residual for an overestimated Sérsic index ($n = 6$). Note the distinct ring structures of over- and underestimated flux in the left two residual images.

be controlled carefully for the biased sources. In order to minimize the influence of intracluster light on the fitting parameters, the user can re-run GALAPAGOS-C with smaller postage stamps (change the set-up parameter ‘stampfac’ to a value between 1 and 1.5). The fraction of biased sources must be replaced manually in the output catalogue.

C3 Stars and artefacts

Certain objects in a survey cannot be well approximated with a Sérsic profile (non-Sérsic objects). Examples are point-like sources and very bright objects, which lead to the saturation of pixels in the CCD array and thus cause artefacts due to the architecture of the telescope.

GALAPAGOS-C overcomes a large fraction of this light by object masking. Saturated pixels usually have an exposure time of zero in the exposure time map (weight image) and hence these pixels are masked automatically. However, if a non-Sérsic source is located too close in the neighbourhood of a galaxy, GALAPAGOS-C automatically includes it into the profile fit as a secondary source with a Sérsic profile. The result is a bad fit for the neighbouring source. The remaining flux, which was not properly approximated in the fitting procedure, can falsify the fitting parameters for the primary source.

The user must be aware of these impacts and thus should carefully control the fits of sources with nearby non-Sérsic objects. If a significant fraction of objects is biased, the user should include these sources into the bad detection list and re-run GALAPAGOS-C. It also makes sense to take a corrective action before running the GALFIT block and erase every source within a certain radius around non-Sérsic objects from the combined catalogue.

C4 Insufficient fits of bright neighbouring galaxies

The light contribution of bright objects with a high Sérsic index can influence a significant number of neighbouring sources in the survey due to the nature of these profiles showing bright and extended wings. To this end, GALAPAGOS-C takes special care in order to include brighter light profiles correctly for fainter, neighbouring sources. However, if the fit for the brighter source is insufficient, the fit error can propagate and bias the fainter, neighbouring sources. This error becomes significant in impacting the subsequent fits when the Sérsic index of the contributing source is strongly overestimated and the included profile pretends light where no light contribution exists. In the same manner, when the Sérsic index is underestimated, an influencing light profile is not sufficiently taken into account and the fitting procedure of neighbouring sources can be biased.

To this end, the user must take special care for the fits of large objects with high Sérsic indices and a large-area light contribution. Fig. C1 shows an example for an elliptical galaxy with a Sérsic index of $n \approx 4$. The small and faint secondary profiles on the left are automatically downgraded to tertiary sources due to their comparatively low magnitudes and are masked out from the fits. Despite the detection failures of the relatively faint neighbours, the left residual shows a relatively good approximation with the correct Sérsic index. In the middle residual image, the Sérsic index is (artificially) underestimated and in the right residual image, the Sérsic profile is insufficient due to an (artificially) overestimated Sérsic index. The result in the latter cases is that the residual shows distinct, concentric rings around the centre of the source. If the Sérsic index is overestimated, too much flux is assigned to the central region of the object and the outermost regions. In these regions, the flux is significantly overestimated. Otherwise, too less flux is assigned to the object in regions of intermediate radial distance. If a fainter, neighbouring source (e.g. the small source on the top left) is located in one of these regions, its fitting parameters can be strongly falsified. In both examples with wrong Sérsic indices, the falsification effects lead to similar biases for the fitting parameters of neighbouring sources. To this end, it makes sense to control the residuals of the most prominent and influencing high-Sérsic sources in the survey e.g. bright ellipticals and BCG.

High-Sérsic sources are sensitive to a correct sky value due to relatively low signal-to-noise ratios in the outer regions of the light profile where the wings in the light profiles become important. If a systematic trend of over- or underestimated Sérsic indices (as indicated in Fig. C1) for the mentioned sources is observed, the user should refine the sky estimation routine parameters. If the sky is generally overestimated, it makes sense to increase the enlarging factor ‘size’ for the Kron ellipses when preparing the sky maps or the user could increase the additional offset factor for the sky isophotes ‘offset’ in order to start the sky estimation at larger distances from the object and thus minimize the light contribution. In addition, it makes sense to force the sky estimation routine to converge later at larger distances from the objects e.g. by increasing the parameter ‘nslope’ (number of measuring points for slope calculation). Further, it can make sense to use a bigger width of sky isophotes ‘bandsize’ with a wider distance interval by increasing the set-up parameter ‘isodist’ (distance between the sky isophotes). The GALAPAGOS-C example set-up file provides well-tested example starting parameters for *HST* data sets.

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