

GALAXY AND MASS ASSEMBLY (GAMA):
BULGE-DISK DECOMPOSITION OF KIDS AND
VIKING DATA IN THE NEARBY UNIVERSE

DISSERTATION

ZUR ERLANGUNG DES DOKTORGRADES
AN DER FAKULTÄT FÜR MATHEMATIK, INFORMATIK UND NATURWISSENSCHAFTEN
FACHBEREICH PHYSIK
DER UNIVERSITÄT HAMBURG

VORGELEGT VON

SARAH CASURA

HAMBURG
2022

Gutachter der Dissertation:	Prof. Dr. Jochen Liske A/Prof. Dr. Aaron S. G. Robotham
Zusammensetzung der Prüfungskommission:	Prof. Dr. Jochen Liske Prof. Dr. Robi Banerjee Prof. Dr. Peter Hauschildt Prof. Dr. Dieter Horns Prof. Dr. Johannes Haller
Vorsitzender der Prüfungskommission:	Prof. Dr. Peter Hauschildt
Datum der Disputation:	17.06.2022
Vorsitzender des Fach-Promotionsausschusses PHYSIK:	Prof. Dr. Wolfgang J. Parak
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Günter H. W. Sigl
Dekan der Fakultät MIN:	Prof. Dr. Heinrich Graener

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben.

Hamburg, den 27.04.2022



(Sarah Casura)

Zusammenfassung

Quantitative Strukturanalysen von Galaxien sind unerlässlich um theoretische Modelle und Simulationen einzugrenzen, welche die Entstehung und Entwicklung von Galaxien sowie des gesamten Universums beschreiben. Entsprechende Daten werden von Himmelsdurchmusterungen in hoher Qualität geliefert. Um diese vollumfänglich nutzen zu können, sind automatisierte Methoden notwendig. Diese Promotionsarbeit präsentiert Strukturparameter für die Komponenten einer großen Anzahl von GAMA Galaxien im nahen Universum und verbessert die Bildanalyse, die Modellierung von Galaxien sowie die Nachbearbeitung solcher Modelle zur Qualitätssicherung, im Kontext von automatisierten Galaxiezerlegungsstudien. Die Probe umfasst 13096 Galaxien in den GAMA II Equatorialregionen bei Rotverschiebungen $z < 0.08$. Die Bilddaten stammen von den KiDS und VIKING Himmelsdurchmusterungen, welche den optischen und nah-infraroten Wellenlängenbereich abdecken (*ugriZYJHK_s*-Filter). Wir fitten die Oberflächenhelligkeitsverteilung jeder Galaxie in jedem Filter mit drei Modellen: einer Sérsicfunktion, einer Sérsic- plus Exponentialfunktion sowie einer Punktquelle plus Exponentialfunktion. Die Modellanpassung erfolgt mit einer vollautomatisierten Markov-Chain-Monte-Carlo-Analyse mit dem bayesschen zweidimensionalen Profilanpassungs-Code **ProFit**. Die Vorarbeit, einschließlich der Bildsegmentierung, Hintergrundsubtraktion, Modellierung der Punktspreizfunktion und Schätzung von Anfangswerten, wird mit dem Bildanalysen-Code **ProFound** ausgeführt. Nach der Modellierung wird eine Modellauswahl getroffen und Galaxien für welche keines der Modelle angemessen ist - z.B. asymmetrische und verschmelzende Systeme - markiert. Die Modellierungsqualität wird durch visuelle Inspektion, Vergleiche mit bestehenden Veröffentlichungen sowie zwischen unabhängigen Ergebnissen für mehrfach-Aufnahmen derselben Galaxie und mit speziell auf diese Arbeit zugeschnittenen Simulationen überprüft. Diese werden auch für eine Untersuchung von systematischen Fehlerquellen verwendet. Die Modellierungsergebnisse erweisen sich als verlässlich für eine große Spannbreite an verschiedenen Galaxietypen und Bildqualitäten. Es liegen nur minimale systematische Messabweichungen vor. Die systematischen Fehler sind einen Faktor von 2-3 größer als zufällige Fehler. Die automatisierte Modellauswahl stimmt mit einer Genauigkeit von $> 90\%$ mit der visuellen Inspektion einer Teilprobe von Galaxien überein. Eine Untersuchung der $g - r$ -Farben von Galaxiekomponenten und des zugehörigen Farbe-Helligkeits-Diagramms ergibt, dass diese im Einklang mit bestehenden Veröffentlichungen sind, trotz der größeren Flexibilität der Modelle dieser Arbeit. Alle Resultate sind auf der GAMA-Datenbank integriert.

Abstract

Quantitative measurements of the structural components of the galaxy population are crucial to constrain theory and simulations, which - in turn - constrain the formation and evolution of galaxies and the universe as a whole. This requires analysing large and diverse samples of galaxies at multiple wavelengths. Corresponding high-quality data is delivered by current and future large galaxy imaging surveys. To fully exploit these, automated analysis methods need to be developed further. In this thesis, we derive a catalogue of robust structural parameters for the components of a large sample of nearby GAMA galaxies while at the same time contributing to the advancement of image analysis, surface brightness fitting and post-processing routines for quality assurance in the context of automated large-scale bulge-disk decomposition studies. The sample consists of 13096 galaxies at redshifts $z < 0.08$ in the GAMA II equatorial survey regions with imaging data from the Kilo-Degree Survey (KiDS) and the VISTA Kilo-Degree INfrared Galaxy (VIKING) survey spanning the optical and near-infrared (u, g, r, i, Z, Y, J, H and K_s bands). We fit three models to the surface brightness distribution of each galaxy in each band individually: a single Sérsic model, a Sérsic plus exponential and a point source plus exponential. The fitting is performed with a fully automated Markov-chain Monte Carlo (MCMC) analysis using the Bayesian two-dimensional profile fitting code `ProFit`. All preparatory work, including image segmentation, background subtraction, point spread function estimation, and obtaining initial guesses, is carried out using the complementary image analysis package `ProFound`. After fitting the galaxies, we perform model selection and flag galaxies for which none of our models are appropriate, mainly mergers and irregular galaxies. The fit quality is assessed by visual inspections, comparison to previous works, comparison of independent fits of galaxies in the overlap regions between KiDS tiles and bespoke simulations. The latter two are also used for a detailed investigation of systematic error sources. We find that our fit results are robust across various galaxy types and image qualities with minimal biases. Errors given by the MCMC underestimate the true errors typically by factors 2-3. Automated model selection criteria are accurate to $> 90\%$ as calibrated by visual inspection of a subsample of galaxies. We also present $g - r$ component colours and the corresponding colour-magnitude diagram, consistent with previous works despite our increased fit flexibility. All results are integrated into the GAMA database.

Previous publications

The following publications make use of results obtained in this thesis:

- Casura S., Liske J., Robotham A. S. G., Brough S., Driver S. P., Graham A. W., Häußler B., et al., “Galaxy And Mass Assembly (GAMA): Bulge-disk decomposition of KiDS data in the nearby universe”, MNRAS (submitted)
- Oh S., Colless M., D’Eugenio F., Croom S. M., Cortese L., Groves B., Kewley L. J., et al., “The SAMI Galaxy Survey: the difference between ionized gas and stellar velocity dispersions”, 2022, MNRAS, 512, 1765. doi:10.1093/mnras/stac509
- Häußler B., Vika M., Bamford S. P., Johnston E. J., Brough S., Casura S., Holwerda B. W., et al., “Galapagos-2/Galfitm/GAMA – multi-wavelength measurement of galaxy structure: separating the properties of spheroid and disk components in modern surveys”, 2022, arXiv, arXiv:2204.05907
- Oh S., Colless M., Barsanti S., Casura S., Cortese L., van de Sande J., Owers M. S., et al., “The SAMI Galaxy Survey: decomposed stellar kinematics of galaxy bulges and disks”, 2020, MNRAS, 495, 4638. doi:10.1093/mnras/staa1330
- Cluver M. E., Jarrett T. H., Taylor E. N., Hopkins A. M., Brough S., Casura S., Holwerda B. W., et al., “Galaxy and Mass Assembly (GAMA): Demonstrating the Power of WISE in the Study of Galaxy Groups to $z < 0.1$ ”, 2020, ApJ, 898, 20. doi:10.3847/1538-4357/ab9cb8
- Robotham A. S. G., Davies L. J. M., Driver S. P., Koushan S., Taranu D. S., Casura S., Liske J., “ProFound: Source Extraction and Application to Modern Survey Data”, 2018, MNRAS, 476, 3137. doi:10.1093/mnras/sty440

Contents

1	Introduction	1
1.1	Background and aims	3
1.1.1	The structure of galaxies	4
1.1.2	Galaxy surface brightness modelling	6
1.1.3	Science aims	8
1.2	Data and sample	11
1.2.1	GAMA	11
1.2.2	KiDS	12
1.2.3	VIKING	12
1.2.4	Sample selection	13
1.3	Methods and code	14
1.3.1	Bayesian analysis	14
1.3.2	Sérsic models	16
1.3.3	ProFit	18
1.3.4	ProFound	19
2	Bulge-disk decomposition pipeline	21
2.1	Pipeline overview	21
2.1.1	Preparatory steps	22
2.1.2	Galaxy fitting	26
2.1.3	Post-processing	33

2.2	Pipeline development	41
2.2.1	Data inputs and setup	42
2.2.2	Background subtraction choices	45
2.2.3	PSF estimation details	57
2.2.4	Segment sizes	78
2.2.5	Modelling decisions	87
2.2.6	Fitting specifics	96
2.2.7	Manual calibrations	109
2.3	Pipeline updates	122
2.3.1	VIKING data products	122
2.3.2	Segmentation changes	124
2.3.3	Joint fitting	128
2.3.4	Manual re-calibrations	134
3	Results	141
3.1	BDDcomp DMU	141
3.1.1	BDDcomp v01	142
3.1.2	BDDcomp v02	142
3.1.3	BDDcomp v03	142
3.1.4	BDDcomp v04	144
3.1.5	BDDcomp v05	145
3.1.6	BDDcomp v06	146
3.2	Catalogue statistics	147
3.2.1	Statistics overview	147
3.2.2	Outlier statistics	155
3.2.3	Model selection differences	160
3.3	Parameter distributions	167
3.3.1	v04 parameter distributions	167
3.3.2	v05 parameter distributions	171

3.3.3	Galaxy and component colours	179
3.4	Catalogue limitations	182
3.4.1	Model limitations	182
3.4.2	Model selection caveats	182
3.4.3	Drawbacks of tight fitting segments	184
3.4.4	Sources of systematic uncertainties	184
3.4.5	GAMA-KiDS RA/Dec offset	184
3.4.6	Completeness limits	184
4	Quality control	187
4.1	Comparison to previous works	187
4.1.1	Comparison of catalogue statistics	187
4.1.2	Comparison of component colours to literature	189
4.1.3	Comparison to size-stellar mass relations of Lange et al. (2015)	193
4.1.4	Comparison to single Sérsic fits of Kelvin et al. (2012)	195
4.2	Systematic uncertainties and biases	197
4.2.1	Overlap sample comparison	197
4.2.2	Simulations: parameter recovery	199
4.2.3	Simulations: model selection accuracy	200
4.2.4	Systematic uncertainties	203
5	Summary, conclusion and outlook	211
	Bibliography	i

1 Introduction

Most of the information we obtain about the universe is in the form of electromagnetic radiation. The majority of electromagnetic radiation is emitted by galaxies, which are one of the most important constituents of the universe. Studying the galaxy population, their structure and evolution is therefore one of the main pathways to understand our universe (e.g. Zeilik & Gregory, 1998; Hammer et al., 2017; Kembhavi & Khare, 2020; Häußler et al., 2022).

Galaxies are complex objects consisting mainly of stars, gas, dust, dark matter and a supermassive black hole, sometimes including an active galactic nucleus (AGN). They differ greatly in their physical properties including their size, mass and its distribution in different components, their age, composition (relative amounts of gas and dust, types of stars contained, abundances and distribution of different elements, dark matter component), star formation activity, assembly and evolutionary history. Most of those properties can only be inferred indirectly and suffer from many uncertainties (e.g. Zeilik & Gregory, 1998; Vika et al., 2013; Hammer et al., 2017, see also Section 1.1.1 for an overview over different galaxy structures).

The picture becomes even more complex when considering the interactions between the different constituents of galaxies as well as their evolution over time. For example, stars can form from gas via collapse and then feed back into the environment from which they formed via radiation and mass-loss processes (e.g. Kelvin et al., 2012; Driver et al., 2016, and references therein). A significant fraction of the radiation emitted by stars is absorbed by dust and re-emitted at longer wavelengths (Driver et al., 2007a; Popescu et al., 2011). The visual appearance of a galaxy at any given point in time is therefore not only determined by the relative amounts of stars, gas and dust it contains, but also their relative distributions and the galaxy geometry and orientation with respect to the line of sight. Further, most galaxies are not isolated systems but exhibit a range of interactions with their environments. Exactly how all of these aspects are related and lead to the formation and evolution of a variety of galaxies is still not fully understood, with much debate on the relative importance of different processes such as mergers, feedback or secular processes (Driver et al., 2009; Lange et al., 2016; Hammer et al., 2017; Nedkova et al., 2021).

To constrain the formation and evolution of galaxies, observations are key. The two main branches of observations of galaxies are photometry and spectroscopy. While the latter has the advantage of providing the intensity as a function of wavelength, the former offers larger

samples at higher spatial resolution and can also include at least some colour information via multi-wavelength observations in different broad-band filters (Zeilik & Gregory, 1998; Vika et al., 2014; Häußler et al., 2022). For both methods, the optical part of the electromagnetic spectrum is of special importance. In part, this is because optical telescopes (and the analysis methods of corresponding data) have reached greater maturity than their counterparts at other wavelengths due to their much longer history. However, there is also physical motivation for studying galaxies at (near-)visible wavelengths since this is the range in which most stars emit most of their radiation (Zeilik & Gregory, 1998; Driver et al., 2012; Vika et al., 2013).

Along with technological advances, observations of galaxies have become more sophisticated over time (Driver et al., 2009; Häußler et al., 2022). To fully exploit the increased data quality and quantity of modern galaxy surveys and maximise their science return, corresponding advances in analysis methods are needed (Hammer et al., 2017; Robotham et al., 2017, 2022; Häußler et al., 2022). Quantitative measurements of the sizes and structures of the galaxy population are particularly important to constrain theory and simulations, which - in turn - constrain the formation and evolution of galaxies and the universe as a whole (e.g. Driver et al., 2009; Simard et al., 2011; Kelvin et al., 2012; Nedkova et al., 2021, and see also Section 1.1.2 for a more comprehensive overview of galaxy surveys and corresponding analysis methods).

This thesis presents the analysis of the photometric images of 13096 galaxies from the Galaxy and Mass Assembly (GAMA; Driver et al., 2009) survey in nine broad-band optical and near-infrared (NIR) filters. We fit several models to the surface brightness distribution of each galaxy in each band, in order to obtain information about their internal stellar structure. The resulting catalogue can be used to analyse the properties of galaxies as a function of wavelength, morphology or other parameters and serves as a basis for the comparison to theory and simulations. At the same time, we have contributed to the advancement of software, methodologies and analysis techniques in the field of large automated bulge-disk decomposition studies of galaxies.

We provide more details on the aims and achievements of this thesis at the end of Section 1.1, after introducing the more general context. The remainder of this chapter then presents the data (Section 1.2) as well as the analysis methods and code packages used (Section 1.3), including a discussion of the distinguishing features of this study compared to previous work. Chapter 2 details the pipeline we developed for the bulge-disk decomposition of our sample of galaxies, including preparatory work and post-processing; and the evolution of all steps over time. The main results of this pipeline are then shown in Chapter 3. Chapter 4 focuses on the quality control of the fits by comparison to previous work and a detailed investigation into systematic uncertainties and biases. We conclude with a summary and outlook in Chapter 5. We assume a standard cosmology of $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.3$ and $\Omega_\lambda = 0.7$ throughout.

Parts of this work have already been submitted for publication in Casura et al. (submitted). In particular, Section 1.1.2, parts of Section 1.1.3, most of Sections 1.2.1, 1.2.2, 1.2.4, 1.3.3 and 1.3.4, Section 2.1, parts of Section 3.2.1, Sections 3.3.1, 3.3.3 and 3.4, Chapter 4 and the majority of Chapter 5 are either heavily based on or directly taken from Casura et al. (submitted) with only minor modifications. This includes Figures 2.1, 2.2, 2.3, 2.11 and 2.39, the top panel of Figure 3.1, Figures 3.8, 3.9, 3.10, 3.18, 3.19, 4.1, 4.2 and 4.3, the right panels of Figures 4.4 and 4.5, Figures 4.6, 4.7, 4.8 and 4.9 as well as Tables 2.1, 2.2, 2.3, 2.5, 3.1, 4.1 and 4.2. The remaining work of this thesis has not been presented elsewhere to date. Where we have taken large amounts of content from Casura et al. (submitted), we also indicate this at the beginning of the corresponding chapter or section.

The written work presented here is accompanied by several data products which are bundled in the `BDDcomp` data management unit (DMU) on the GAMA team database¹. It includes several catalogues containing the results of the preparatory work, the galaxy fitting and the post-processing as well as a detailed description of all processing steps and the columns contained in each catalogue. Various diagnostic plots for each galaxy fit and all inputs used for the fitting (i.e. the outputs of the preparatory work) are stored on the GAMA file server. To date, there have been four releases of the `BDDcomp` DMU (v01 to v04), with the fifth (v05) to be published alongside this thesis. More information on the `BDDcomp` DMU and its different releases is given in Section 3.1. In addition, all of the results stored on the GAMA database as well as many test runs and the corresponding code can be found on the local machines at Hamburg observatory, specifically on `/hs/fs11/data/gama/profit/` and the corresponding directories on the `fs12`, `fs13` and `fs14` machines. The “readme”-files in those directories provide orientation. The (publicly available) KiDS and (preprocessed) VIKING data are also on the `fs11` machine, in the `/hs/fs11/data/gama/data/imaging/` directory.

1.1 Background and aims

In this section, we explain the context and motivation for the current work. We begin with a more detailed description of the different types of galaxies and their structure (Section 1.1.1), followed by a brief overview on galaxy surface brightness modelling including available tools and common practices in the field (Section 1.1.2). This provides the background for the aims of this thesis, which are described in Section 1.1.3. Most of Section 1.1.2 and parts of Section 1.1.3 are taken from the introduction of Casura et al. (submitted).

¹<http://www.gama-survey.org/db/>



Figure 1.1: The Hubble sequence. Elliptical galaxies are labelled E and shown in the left part of the diagram, with increasing numbers referring to increasing ellipticities. Spiral galaxies are shown on the right, split into “ordinary” S (top) and barred SB (bottom) spirals. Letters a to c represent the transition from “early-types” to “late-types”, characterised by an increased dominance of the disk relative to the bulge combined with less tightly wound spiral arms. Image credit: Ville Koistinen, Wikimedia Commons [https://commons.wikimedia.org/wiki/File:Hubble_sequence_photo.png].

1.1.1 The structure of galaxies

Here, we give a brief overview over the different types of present-day (low z) galaxies and their main properties, mostly following the discussion in Zeilik & Gregory (1998). We note that the picture drawn here is somewhat simplified and galaxies are more complex in detail, with many exceptions to the rule.

Traditionally, three main types of galaxies are distinguished according to their visual appearance: elliptical galaxies, spiral galaxies and irregular galaxies. These are represented in the Hubble sequence introduced by Hubble (1926) with a schematic shown in Figure 1.1.

Elliptical galaxies have a smooth appearance with near to no substructure, gas or dust. Their stellar population is relatively old with little ongoing star formation such that they appear reddish in colour. They are pressure-supported with random stellar motions and typically follow a de Vaucouleurs (1948) radial profile, i.e. the intensity I drops approximately as $10^{r^{-1/4}}$ with r being the radius from the centre of the galaxy.

Spiral galaxies show more structure than elliptical galaxies. They classically have two main stellar components: a bulge and a disk. The bulge sits at the centre and is similar to an elliptical galaxy in its properties, due to which bulges and ellipticals are often classed together and named spheroids. The disk is flat and rotationally supported, contains gas and dust and has a younger stellar population with ongoing star formation and bluer colours. It typically

follows an exponential decrease in intensity from the centre, although with local variations due to substructure like spiral arms.

As can be seen in Figure 1.1, spiral galaxies are subdivided into two branches according to whether or not they have a central bar. The bar is disk-like in its stellar composition, but, as the name suggests, has an elongated shape. There are also spiral galaxies that host a pseudo-bulge instead of the classical bulge described above. Pseudo-bulges, like classical bulges, appear approximately spherical and are located at the centre of the galaxy. However, pseudo-bulges are younger and bluer in colour than classical bulges and show a less pronounced increase in intensity towards the centre (e.g. Kennedy et al., 2016; Lange et al., 2016; Häußler et al., 2022).

The distinct differences between spheroids and disks can most easily be explained by different formation channels (e.g. Driver et al., 2013; Lange et al., 2016). In the two-phase model, spheroids form in a so-called “hot” mode, whereby they rapidly assemble their mass through collapse and mergers at early times. This leads to (now) old stellar populations with a high concentration at the nucleus and a generally spheroidal, pressure-supported shape. Disks are formed later and more slowly in the “cold” mode via the accretion of gas regulated by feedback processes. Therefore, disk stellar populations are younger than spheroids, flattened due to the conservation of angular momentum and rotationally supported. Bars and pseudo-bulges can form from disk instabilities.

In between spiral and elliptical galaxies lie S0 or lenticular galaxies. They have both a dominant bulge and disk, but no spiral arms or other prominent substructure. Their colour and composition is intermediate in between that of spiral galaxies and elliptical galaxies; their formation channel is still being debated (e.g. Barsanti et al., 2021).

The last of the three main categories of galaxies are irregulars, which show no clear symmetry. They tend to be bluer in colour than spiral galaxies and can also contain substructure, gas and dust. They are thought to be the result of interactions and mergers between galaxies, which can disturb the regular morphology and trigger star formation.

In addition to these categories of giant galaxies, there are dwarf galaxies. Dwarfs can be elliptical or irregular in shape. They are the most numerous type, but contribute only a small fraction to the total stellar mass and light in the universe. They are also difficult to detect due to their faintness.

The relative fractions of galaxies that fall into each category depend strongly on the observational limits. For magnitude-limited surveys, spirals are usually dominant in number, followed by ellipticals since these two classes of galaxies are brightest. In the local volume, however, most galaxies are small irregular galaxies with spirals only contributing around 33% and ellipticals 13% (Zeilik & Gregory, 1998).

For the GAMA survey, Driver et al. (2022) have derived morphological classifications for all galaxies at a redshift of $z < 0.08$. They find that approximately 10% of their sample are elliptical galaxies, 45% bulge-disk systems (i.e. spirals and S0) and another 45% late-type systems with only a single component, including asymmetric systems (irregulars and disk-only systems with no discernible bulge). Since this sample of galaxies closely matches our sample selection (see Section 1.2.4), we might expect a very similar distribution of galaxy types for our catalogue.

1.1.2 Galaxy surface brightness modelling

To obtain physical quantities from the observations of galaxies provided by surveys, they need to be modelled. Due to its importance in understanding the history of the universe, the quantitative modelling of galaxy surface brightness distributions has a long history dating back to de Vaucouleurs (1948), Sérsic (1963) and even earlier works; see Graham (2013) for a review of the development of light profile models. While the early works focused on azimuthally averaged galaxy profiling with a single functional form (e.g. Kormendy, 1977), modern codes allow users to decompose galaxies into several distinct components (e.g. bulges, disks) and to take into account the full two-dimensional information. To this end, there are many different techniques, methods and code packages, all of which have become increasingly sophisticated as the quality and quantity of available astronomical data have grown.

Broadly, they can be divided into parametric and non-parametric modelling, as well as one-dimensional and two-dimensional methods. Which of these is most appropriate to use depends on the science case and the available data. This work falls into the regime of large-scale automated analyses of galaxies with often barely resolved components, for which we want to obtain structural parameters that are easily comparable between galaxies. Hence, two-dimensional parametric analysis is most appropriate (see also the discussion in Robotham et al. 2017 and references therein).

Examples of such two-dimensional, parametric fitting tools used for large-scale automated analyses include GIM2D (Simard et al., 2002), BUDDA (de Souza et al., 2004), GALFIT3 (Peng et al., 2010), GALFITM (Vika et al., 2013), IMFIT (Erwin, 2015), ProFit (Robotham et al., 2017) and PHI (Argyle et al., 2018). Each of these tools comes with its own advantages and disadvantages, which goes to show how difficult the problem of galaxy modelling is, especially when automated for large samples of the very diverse galaxy population. Usually, some form of post-processing is needed to assess the influence of systematic uncertainties, judge the convergence, exclude bad fits and identify the most appropriate model to use for each galaxy. This can be achieved via visual inspection (for small enough samples), logical filters, frequentist statistics such as the F -test, Bayesian inference, or similar methods (see, e.g., Allen et al., 2006; Gadotti, 2009; Simard et al., 2011; Vika et al., 2014; Meert et al., 2015;

Lange et al., 2016; Méndez-Abreu et al., 2017).

Despite the associated difficulties (e.g. convergence and quality of fit metrics), many authors have performed two-dimensional surface brightness profile fitting for large numbers of galaxies, modelling the radial light profile as a simple functional form, most often a Sérsic function (Blanton et al., 2003, 2005; Barden et al., 2005; Trujillo et al., 2006; Hyde & Bernardi, 2009; La Barbera et al., 2010; Kelvin et al., 2012; van der Wel et al., 2012; Häußler et al., 2013; Shibuya et al., 2015; Sánchez-Janssen et al., 2016, to name just a few). The results of such analyses have been used to derive a number of key relations between different galaxy properties, their formation and evolutionary history, and interactions with the environment. For example, many works have studied the distribution of, and relation between, size and mass or luminosity for different galaxy types (split by e.g. Sérsic index or colour), sometimes including morphology, surface brightness, internal velocity, environment, wavelength, colour, or redshift effects (e.g. Shen et al., 2003; Barden et al., 2005; Blanton et al., 2005; Trujillo et al., 2006; Hyde & Bernardi, 2009; La Barbera et al., 2010; Kelvin et al., 2014; van der Wel et al., 2014; Lange et al., 2015; Shibuya et al., 2015; Nedkova et al., 2021).

With improving data quality of surveys, the fitting of more than one component - i.e. decomposing galaxies - has become more common. While some authors, such as Gadotti (2009), Salo et al. (2015) or Gao & Ho (2017) also account for bars, central point sources, spiral arms or other additional morphological features, most works focus on the bulge and disk. The focus on only two components is especially true when running automated analyses of large samples, since in many cases the data quality is not sufficient to meaningfully constrain more than one or two components, or it would require extensive manual tuning based on visual inspection. From a more physical point of view, the majority of the stellar mass in the local universe resides in ellipticals, disks and classical bulges, with pseudo-bulges and bars only contributing a few percent (Gadotti 2009; and see also Section 1.1.1). Hence, for automated analyses it is common practice to fit only two components, where the term “bulge” is used to describe the central component, irrespective of whether it is a classical bulge, pseudo-bulge, bar, lens, AGN, or a mixture thereof, while “disk” refers to a more extended component with typically lower surface brightness and potential additional structure such as spiral arms, breaks, flares or rings.

Examples of large bulge-disk decomposition studies include Simard et al. (2002, 2011); Allen et al. (2006); Benson et al. (2007); Gadotti (2009); Lackner & Gunn (2012); Fernández Lorenzo et al. (2014); Head et al. (2014); Mendel et al. (2014); Vika et al. (2014); Meert et al. (2015, 2016); Kennedy et al. (2016); Kim et al. (2016); Lange et al. (2016); Dimauro et al. (2018); Bottrell et al. (2019); Cook et al. (2019); Barsanti et al. (2021); Domínguez Sánchez et al. (2022); Häußler et al. (2022); Hashemizadeh et al. (2022); and Robotham et al. (2022). Such catalogues can then be used to determine the relative numbers of different galaxy components

as well as their luminosity or stellar mass functions, size-mass or size-luminosity relations, including their redshift evolution and dependence on other properties of the galaxy and its environment (similar to the studies of entire galaxies mentioned earlier). For example, this has been done by Driver et al. (2007b); Dutton et al. (2011); Tasca et al. (2014); Kennedy et al. (2016); Lange et al. (2016); Moffett et al. (2016) and Dimauro et al. (2019).

In addition, quantitative measures for the components of galaxies aid the comparison of observational data to theory and simulations. Bulges and disks are often decisively different not only in their visual appearance but also in their structure, dynamics, stellar populations, gas and dust content and are thought to have different formation pathways (Section 1.1.1 and Cole et al. 2000; Cook et al. 2009; Driver et al. 2013; Lange et al. 2016; Dimauro et al. 2018; Lagos et al. 2018; Oh et al. 2020). Consequently, bulge-disk decomposition studies provide stringent constraints on the formation and evolutionary histories of galaxies and their physical properties that are not easily measured directly such as the dark matter halo, the build-up of stellar mass (in different components) over time, or merger histories (examples include Driver et al., 2013; Bottrell et al., 2017; Bluck et al., 2019; Rodriguez-Gomez et al., 2019; de Graaff et al., 2022). Hence, consistently measuring the structure of the stellar components is essential to make full use of current and future large-scale observational surveys such as the Kilo-Degree Survey (KiDS; de Jong et al., 2013) and the VISTA Kilo-Degree INfrared Galaxy (VIKING; Edge et al., 2013) Survey or the Legacy Survey of Space and Time (LSST; Ivezić et al., 2019), and of cosmological hydrodynamical simulations such as Illustris (Vogelsberger et al., 2014) and IllustrisTNG (The Next Generation; Pillepich et al., 2018) or Evolution and Assembly of GaLaxies and their Environments (EAGLE; Schaye et al., 2015).

1.1.3 Science aims

This work is the first in a series of planned contributions to the field of galaxy structure and evolution, with the final aim of the series being to constrain the nature and distribution of dust in galaxy disks. Driver et al. (2007a) have shown that in the B -band of the Millennium Galaxy Catalogue (MGC), an average of 37% of photons produced in the disk and 71% of photons produced in the bulge are absorbed before even leaving their galaxy of origin. Accounting for this effect of internal dust attenuation is therefore vital to avoid biases in structure formation and galaxy evolution studies, especially since the dust properties evolve as well. Due to the geometry of the system (see also Section 1.1.1), bulges and disks are affected differently and the severity of the attenuation depends strongly on the inclination angle of the galaxy relative to our line of sight (see figure 11 in Driver et al. 2007a). Bulges and disks therefore need to be studied separately.

A sufficiently large sample of robust structural parameters for these two components can be used to constrain the nature and distribution of dust. This can be achieved by comparing

the distribution of bulges and disks in the luminosity-size plane as a function of inclination to dust radiative transfer models such as those presented in Popescu et al. (2011) and preceding papers of this series. The analysis is particularly powerful if observations in multiple filters are available since dust attenuation also varies strongly as a function of wavelength (Popescu et al., 2011).

This thesis presents the first step towards achieving this final goal: we obtain single Sérsic fits and bulge-disk decompositions for 13096 GAMA galaxies in the KiDS u , g , r and i bands and the VIKING Z , Y , J , H and K_s bands. We choose `ProFit` (see Section 1.3.3) as our modelling software due to its Bayesian nature (allowing full MCMC treatment including more realistic error estimates), its suitability to large-scale automated analyses and its ability, in combination with `ProFound` (Section 1.3.4), to serve as a fully self-contained package covering all steps of the analysis from image segmentation through to model fitting. We supplement this functionality with our own routines for the rejection of unsuitable fits, model selection, and a characterisation of systematic uncertainties. The aim of our future work is to use those structural parameters for the bulges and disks to significantly expand the analysis of Driver et al. (2007a), exploiting our more and better data at several wavelengths.

With these science aims in mind, we are most interested in obtaining structural parameters that are directly comparable amongst each other, i.e. consistent within the dataset; and correctly represent the statistical properties of the entire sample, with less emphasis placed on capturing all aspects of the detailed structure of individual galaxies. Consequently, we choose to model a maximum of two components for each galaxy and use the terms “bulge” and “disk” in their widest senses, in line with previous automated decompositions of large samples. In particular the “bulges” we obtain are often mixtures of classical or pseudo-bulges, bars, lenses and AGN. Similarly, we place more emphasis on the central, high surface brightness regions of galaxies by modelling only a relatively tight region around each galaxy of interest. While most of the fits we obtain are not perfect (because galaxies are more complex than two simple components), they do achieve the aims specified above and are comparable to similar studies.

While much of our procedure was focused towards obtaining suitable fits for our final goal of studying dust properties, there are many other analyses that can be built on our results. Some of the most obvious of these include deriving the stellar mass functions of bulges and disks, studying component colours and investigating the trends of structural parameters such as bulge or disk size with wavelength, all of which belong to our plans for future work. Additionally, (an earlier version of) the resulting catalogue has been used to aid the kinematic bulge-disk decomposition of a sample of galaxies in the Sydney-AAO Multi-object Integral-field spectroscopy (SAMI) Galaxy Survey (Oh et al., 2020), to examine the properties of galaxy groups (Cluver et al., 2020), to investigate the difference between ionised gas and stellar velocity dispersions (Oh et al., 2022), to cross-check the results of other multi-wavelength

bulge-disk decomposition studies (Häußler et al., 2022; Robotham et al., 2022), to study the alignment of galaxy spin axes with filaments of the cosmic web as a function of different galaxy (component) properties (Barsanti et al., in prep.) and to validate a cosmological galaxy simulation against observations using an unsupervised machine learning technique (Turner et al., in prep.). Several student projects, including Bachelor’s and Master’s theses, have also made use of our results. For example, Roschlaub (2022) tested the usage of a new deconvolution algorithm presented in Nammour et al. (2021) in the context of galaxy fitting, Targaczewski (in prep.) is working on estimating the supermassive black hole mass from the bulge Sérsic index for a sample of AGN in our catalogue, Ehbrecht (in prep.) is investigating under which circumstances (seeing and noise) a given Sérsic profile can be reasonably constrained, Porter-Temple (in prep.) is going to look at the difference the number of spiral arms make on the bulge-to-disk flux ratio and Porter (in prep.) will compare the bulge-to-disk flux ratios of void galaxies to those of the remaining GAMA sample.

Apart from these scientific insights enabled by the final product of our pipeline (the catalogue), advancements on the technical side should not be neglected. As briefly mentioned before, improvements in methodology are crucial to make full use of current and future datasets of large-scale galaxy surveys; and we have contributed to this aspect in several ways. First of all, `ProFit`, and even more so `ProFound`, were under active development during the time of our own pipeline development, with this thesis project being one of the first large-scale automated applications of both packages. Consequently, our work has contributed to improving both packages by discovering bugs, suggesting additional features and serving as inspiration for the implementation of various automated procedures. In addition, we have added routines for the swapping of bulge and disk components (see Section 2.1.2) and post-processing of fits (mainly model selection and flagging of bad fits, Section 2.1.3); and have performed numerous tests with varying combinations of `ProFound` and `ProFit` routines and their tuning parameters for the preparatory work (image segmentation, background subtraction, point spread function estimation, obtaining initial guesses) as well as the galaxy fitting. The resulting procedures, as well as the numerous alternatives that we found to be less optimal are described in detail in Chapter 2 and can serve as guides for similar bulge-disk decomposition works on current and future large-scale galaxy surveys. Our detailed analysis of systematic uncertainties based on the overlap sample and bespoke simulations (Section 4.2.4) can also give orientation towards constraining and quantifying the dominant sources of error in other analyses.

The goal of this thesis therefore is not only to present our final catalogue of robust structural parameters for the components of galaxies, but also the large amounts of technical work surrounding the pipeline development.

1.2 Data and sample

After the general introduction of the context and aims of this thesis, we now present the data products that we use from GAMA (Section 1.2.1), KiDS (Section 1.2.2) and VIKING (Section 1.2.3) followed by the selection of our sample of galaxies (Section 1.2.4). Sections 1.2.1, 1.2.2 and 1.2.4 are largely based on Casura et al. (submitted, their section 2).

1.2.1 GAMA

The Galaxy and Mass Assembly (GAMA)² survey is a large low-redshift spectroscopic survey covering $\sim 238\,000$ galaxies in 286 deg^2 of sky (split into 5 survey regions) out to a redshift of approximately 0.6 and a depth of $r < 19.8$ mag. The observations were taken using the AAOmega spectrograph on the Anglo-Australian Telescope and were completed in 2014. The survey strategy and spectroscopic data reduction are described in detail in Driver et al. (2009); Baldry et al. (2010); Robotham et al. (2010); Driver et al. (2011); Hopkins et al. (2013); Baldry et al. (2014) and Liske et al. (2015).

In addition to the spectroscopic data, the GAMA team collected imaging data on the same galaxies from a number of independent surveys in more than 20 bands with wavelengths between 150 nm and $500\ \mu\text{m}$. Details of the imaging surveys and the photometric data reduction are given in Liske et al. (2015); Driver et al. (2016, 2022); and relevant publications of the corresponding independent surveys. The combined spectroscopic and multiwavelength photometric data at this depth, resolution and completeness provide a unique opportunity to study a variety of properties of the low-redshift galaxy population.

In this work, we focus on the KiDS and VIKING imaging data in the nine optical and near-infrared filters $u, g, r, i, Z, Y, J, H, K_s$ (see Sections 1.2.2 and 1.2.3) in the GAMA II equatorial survey regions, which are 3 regions of size $12^\circ \times 5^\circ$ located along the equator at 9, 12 and 14.5 hours in right ascension (the G09, G12 and G15 regions). For our sample selection, we make use of the equatorial input catalogue³ (`EqInputCat:TilingCatv46`, Baldry et al., 2010) and the most recent version of the redshifts originally described by Baldry et al. (2012) (`LocalFlowCorrection:DistancesFramesv14`), see details in Section 1.2.4. For the stellar mass-size relation (Section 4.1.3), we also use the Data Release (DR) 3 version of the stellar mass catalogue first presented in Taylor et al. (2011) (`StellarMasses:StellarMassesv19`); for the comparison to previous work (Section 4.1.4) we use the single Sérsic fits of Kelvin et al. (2012) (`SersicPhotometry:SersicCatSDSSv09`); and in order to correct galaxy colours for Galactic extinction, we use the corresponding table provided along with the equatorial input

²<http://www.gama-survey.org>

³For the sake of reproducibility, we always give the exact designation of a catalogue on the GAMA database in parentheses: the data management unit (DMU) that produced the catalogue (e.g. `EqInputCat`) followed by the catalogue name (e.g. `TilingCat`) and the version used (e.g. `v46`).

catalogue (EqInputCat:GalacticExtinctionv03). All of these catalogues can be obtained from the GAMA database.²

1.2.2 KiDS

The Kilo-Degree Survey (KiDS, de Jong et al., 2013) is a wide-field imaging survey in the Southern sky using the VLT Survey Telescope (VST) at the ESO Paranal Observatory. 1350 deg² are mapped in the optical broad-band filters u, g, r, i ; while the VIKING Survey (Edge et al. 2013, Section 1.2.3) provides the corresponding near-infrared data in the Z, Y, J, H, K_s bands. The GAMA II equatorial survey regions have been covered as of DR3.0.

KiDS provides $\sim 1^\circ \times 1^\circ$ science tiles calibrated to absolute values of flux with associated weight maps (inverse variance) and binary masks. The science tiles are composed of 5 dithers (4 in u) totalling 1000, 900, 1800 and 1200 s exposure time in u, g, r, i , with all dithers aligned in the right ascension and declination axes (i.e. no rotational dithers) and taken in immediate succession. The r -band observations were performed during the best seeing conditions in dark time; while g, u and i have progressively worse seeing and i was additionally taken during grey time or bright moon. During co-addition, the dithers across all four bands were re-gridded onto a common pixel scale of $0''.2$. The magnitude zeropoint of the science tiles is close to zero with small corrections given in the image headers for DR4.0. The r -band point spread function (PSF) size is typically $0''.7$ and the limiting magnitudes in u, g, r, i are $\sim 24.2, 25.1, 25.0, 23.7$ mag respectively (5σ in a $2''$ aperture). This high image quality, depth, survey size and wide wavelength coverage in combination with VIKING make KiDS data unique. For details, see Kuijken et al. (2019).

For this work, we use the the u, g, r and i band science tiles, weight maps and masks from KiDS DR4.0 (Kuijken et al., 2019), which are publicly available⁴ for our selected sample of galaxies (Section 1.2.4). Our primary band, used for most analyses during pipeline development, is the r -band since it is the deepest and was taken during the best seeing conditions. Observations in g and i are of comparable quality, such that these three bands together form our core bands where we obtain the best and most comparable fits (these are also what Casura et al. submitted is based on). The u -band is of considerable worse data quality especially in terms of depth and therefore we place less emphasis on its analysis. More details are given in Section 2.2, also explaining our choice of which of the KiDS data products to use (Section 2.2.1).

1.2.3 VIKING

The VISTA Kilo-degree INfrared Galaxy (VIKING) survey is a wide-field, intermediate-depth near-infrared imaging survey using the Visible and Infrared Survey Telescope for Astronomy

⁴<http://kids.strw.leidenuniv.nl/DR4/index.php>

1.2. DATA AND SAMPLE

(VISTA) at the ESO Paranal Observatory. 1350 deg² were mapped in the broad-band filters Z, Y, J, H, K_s over two areas of sky matched to the KiDS footprint. The observations are complete and the fourth and final public data release is described in Edge & Sutherland (2020).

VIKING provides astrometrically and photometrically calibrated tiles of size $\sim 1.5^\circ$ in right ascension (RA) and $\sim 1^\circ$ in declination (Dec) composed of 6 pawprints each, arranged in a fashion to cover the gaps between detector chips. Each tile is observed twice, with observations sometimes years apart: first in the Z, Y, J filters with total exposure times of 480, 400 and 200 s; and then again in the J, H, K_s filters with exposure times of 200, 300 and 480 s. Combining the two J -band observations results in median magnitude limits of 21.4, 20.6, 20.1, 19.0 and 18.6 mag in the Z, Y, J, H and K_s bands respectively, although some fields can be shallower by up to 0.3 mag which is the quality threshold applied by the VIKING team. In addition to the stacked tiles, the pawprints are publicly available. Both pawprints and stacks are approximately aligned in RA and Dec, have a pixel size close to $0''.34$ and various Vega magnitude zeropoints around 30, with the exact values given in the image headers. They also both have associated confidence maps which give the per-pixel exposure time and exclude the “bad patches” of two detectors that have flat-fielding issues. More details are given in Edge & Sutherland (2020).

For this work, we use the Z, Y, J, H and K_s individual detector images from the pawprints. In particular, we use the preprocessed versions from Wright et al. (2019) that were kindly provided to us by Angus Wright and the KiDS team. These have the advantage that they are specifically processed in order to allow a consistent analysis in combination with KiDS data: they have been rotated slightly to be exactly aligned in RA and Dec, corrected for atmospheric extinction, re-calibrated to remove the exposure time from the image units and re-scaled onto a common AB magnitude zeropoint of 30. In addition, Wright et al. (2019) also perform background subtraction, produce weight maps and perform a quality control of each chip. More details and the reasons for our choice to use these preprocessed individual chips instead of the VIKING stacked tiles are given in Section 2.3.1.

1.2.4 Sample selection

Our main sample consists of all GAMA II equatorial region main survey targets with a reliable redshift in the range $0.005 < z < 0.08$, which are a total of 12958 objects.⁵ In addition, we include all 2404 targets of the “GAMA sample” of the SAMI Galaxy Survey⁶ (Bryant et al., 2015), the majority of which are already in our main sample. The combination of both results

⁵In detail, we select all targets with $NQ \geq 3$, $SURVEY_CLASS \geq 4$ and $0.005 < Z_CMB < 0.08$ from `EqInput-Cat:TilingCatv46` joined to `LocalFlowCorrection:DistancesFramesv14` on CATAID.

⁶Taking the CATIDs listed in `sami_sel20140413_v1.9_publiclist` from https://sami-survey.org/data/target_catalogue

in the full sample of 13096 unique physical objects, which were imaged a total of 14966 times in each of the KiDS g , r and i bands due to small overlap regions between the tiles. 11301, 1742, 31 and 22 objects were imaged once, twice, three and four times respectively. For versions of our bulge-disk decomposition pipeline up to `BDDcomp v04` (Casura et al., submitted), we keep these multiple data matches to the same physical object separate during all processing steps to serve as an internal consistency check (Chapter 4). For `v05`, where we add the KiDS u and the five VIKING bands to the analysis, we then changed this to instead fit all data matches in the same band jointly. This is necessary due to our decision to work at the individual detector chip level for VIKING (see Section 1.2.3), resulting in many data matches to the same physical objects (more than 20 for some objects). See Section 2.3 for details.

1.3 Methods and code

With the galaxy sample and input data defined, we now turn towards the methods used in the analysis of those. After a brief introduction into the theoretical background of Bayesian analysis and Sérsic modelling in Sections 1.3.1 and 1.3.2, we present the two main code packages used for the galaxy modelling and preparatory work respectively, `ProFit` and `ProFound` (Sections 1.3.3 and 1.3.4; taken from section 2 of Casura et al. (submitted)).

1.3.1 Bayesian analysis

Bayesian probability theory, originally introduced by Reverend Thomas Bayes in the 18th century and further developed by Laplace in the early 19th century, has recently experienced a “re-discovery” with rapidly increasing popularity. In the alternative (for a long time more popular) frequentist approach, probability is defined as the “long-run relative frequency” of an outcome and hence requires many repetitions of an experiment. In contrast to this, Bayesian probability is defined as a “degree of belief”, which has proven to be a powerful approach especially in fields such as astronomy, where the repeatability of experiments is often very limited (Sivia & Skilling, 2006; Gregory, 2005). We briefly review its most basic concepts here following Sivia & Skilling (2006). We refer the reader to this and similar works for a more detailed treatment.

Bayesian analysis can be used both for parameter estimation and model selection. Its most basic building block is Bayes’ theorem, which can be expressed as

$$p(B|A, C) = \frac{p(A|B, C)p(B|C)}{p(A|C)}. \quad (1.1)$$

Here, $p(B|A, C)$ denotes the probability p of statement B given information A and C .

1.3. METHODS AND CODE

In the context of parameter estimation, this becomes

$$p(\theta|data, M) = \frac{p(data|\theta, M)p(\theta|M)}{p(data|M)}, \quad (1.2)$$

where θ denotes one or several parameters of model M . It gives the posterior probability of the parameter(s) [$p(\theta|data, M)$, the quantity that is desired in parameter estimation], as the product of the probability of the data given the parameter values and the model [$p(data|\theta, M)$, which is easily computed] with the prior of the parameter [$p(\theta|M)$, its probability distribution before taking any data] and a normalisation constant that depends only on the data and the model, not the parameter [$p(data|M)$, called evidence]. Note that if the model contains several parameters, the posterior probability is a joint probability for all of those parameters and one needs to marginalise (integrate) over all other parameters to obtain the correct one-dimensional posterior probability of any given parameter of choice. In practice, the most efficient way to achieve this usually is sampling the likelihood space via Markov Chain Monte Carlo (MCMC) or similar methods. Provided convergence is achieved, the MCMC chain points are representative samples from the joint posterior, allowing easy projection of the distribution onto any axis (parameter) of interest.

For model selection, Bayes' theorem (Equation 1.1) takes the following form:

$$p(M_k|data, I) = \frac{p(data|M_k, I)p(M_k|I)}{p(data|I)}. \quad (1.3)$$

This now expresses the posterior probability of Model k as the product of the model likelihood with the model prior, normalised by the probability of the data given any relevant background information I . The ratio of two model posteriors gives the odds ratio between two models:

$$O_{1:2} = \frac{p(M_1|data, I)}{p(M_2|data, I)} = \frac{p(data|M_1, I) p(M_1|I)}{p(data|M_2, I) p(M_2|I)}, \quad (1.4)$$

where the normalisation constant cancelled out since it is the same for both models. This gives the relative probability of the two models (for any parameter values) given the same set of data and background information. The last term on the right hand side is the prior odds ratio, i.e. the relative probability of the two models before taking any data. Unless there is a strong reason to prefer one model over the other (e.g. previous data contained in I), this ratio should be set to 1 to allow a fair comparison of models.

The most relevant term is hence the first term on the right hand side, called the Bayes factor. It consists of the likelihoods for each model, which is the probability of the data given the

model, marginalised (integrated) over all possible parameter values:

$$\begin{aligned} p(\text{data}|M_k, I) &= \int p(\text{data}, \theta|M_k, I)d\theta \\ &= \int p(\text{data}|\theta, M_k, I)p(\theta|M_k, I)d\theta. \end{aligned} \tag{1.5}$$

Due to this integral over all parameters, the marginalised likelihood will generally become smaller when adding parameters that do not improve the fit significantly (since the likelihood $p(\text{data}|\theta, M_k, I)$ is then approximately constant while the integral over the prior probabilities for the parameters $p(\theta|M_k, I)$ decreases due to the larger prior range). Equation 1.4 hence naturally implements Ockham's factor (also called Ockham's razor) which states that if two models can explain the data equally well, then the simpler one should be preferred (i.e. to avoid overfitting). Note that the model likelihood in Equation 1.5 is also the same as the evidence in parameter estimation (the denominator in Equation 1.2) and for this reason the Bayes factor can also be called the ratio of evidences or the ratio of marginalised likelihoods. Computing the model likelihood is often non-trivial especially for models with many parameters, since it requires high-dimensional integrals. We return to this in Section 2.1.3.

1.3.2 Sérsic models

The most important component for both parameter estimation and model selection - apart from the data - is the model M with its parameters θ . If the model is wrong in the sense that it cannot represent the data adequately, then any estimated parameter values will have limited validity. Comparing two inadequate models consequently also produces an odds ratio with little meaning. We hence carefully explain our choice of models (and parameters) here, including its implications.

The most common radial profile to fit the surface brightness distribution of galaxies with is the Sérsic (1963) function. It gives the intensity I as a function of radius r :

$$I(r) = I_e \exp \left[-b_n \left(\left(\frac{r}{R_e} \right)^{1/n} - 1 \right) \right]. \tag{1.6}$$

Here, n is the Sérsic index (the main shape parameter), R_e is the effective radius where half of the total flux is included (the size), I_e is the intensity at R_e (the overall normalisation) and b_n is a normalisation constant which can be calculated from n . The Sérsic function becomes a Gaussian for $n=0.5$, exponential for $n=1$ and a de Vaucouleurs (1948) profile for $n=4$ and can be adjusted to account for ellipticity and/or boxyness in two dimensions (Robotham et al., 2017). Due to this flexibility, the Sérsic function can fit a wide variety of galaxy (component) shapes and types such as exponential disks or classical de Vaucouleurs bulges (Section 1.1.1) and correspondingly is extremely popular in the galaxy fitting community (Graham & Driver,

2005, and references therein). It is, however, a purely empirically derived function with no profound physical meaning.

Adding several Sérsic profiles together allows to fit multiple galaxy components. Common examples include a de Vaucouleurs bulge plus an exponential disk (i.e. Sérsic functions with n fixed to 4 and 1 respectively), a (free n) Sérsic bulge plus exponential disk or a double Sérsic profile. In all cases, the bulge may or may not be forced to be round (i.e. circular instead of elliptical in two dimensions) and there can additionally be further constraints on the parameters such that for example the centres of the bulge and disk components must align. Further morphological features, such as bars, can also be accounted for by adding more (Sérsic or other) profiles.

However, there are also problems associated with increasing numbers of components and fitting parameters: parameter degeneracies become more common, convergence is more difficult to achieve and when the fitting algorithm has converged it needs to be ensured that the solution is also physical, i.e. that the different model components do indeed represent different physical components of the galaxy. As briefly mentioned in Section 1.1.2, this problem is particularly pronounced for automated analyses of large samples of galaxies with varying properties, where manual intervention and control of the fits is limited. Depending on the image quality, physical properties and redshift of the galaxy, it is possible that only one component can reasonably be constrained by the data even though the galaxy physically consists of several. Others may need three or more components to adequately represent all morphological features. Conversely, there are also a number of galaxies (e.g. ellipticals) that physically contain just a single component, leaving any additional model components unconstrained. While in theory these should be easily recognized during model selection, it is often non-trivial in practice due to overlapping point sources, neighbouring objects, imperfect PSF estimates or sky subtraction, image artifacts and similar issues.

For these reasons, we decided to follow most other works in the field of large automated analyses of galaxy surface brightness fitting (Section 1.1.2) and fit a maximum of two components to each galaxy. For the disk component, we use an exponential profile that is elliptical in two dimensions. This ensures that we can correctly capture disks with differing brightnesses, sizes and inclination angles. We do not allow the Sérsic index to vary (i.e. we fit exponential disks instead of Sérsic disks) as most disks do follow an exponential profile on average (Zeilik & Gregory, 1998), with spiral arms only contributing local perturbations. Other deviations from the exponential profile, such as disk breaks and flares are most common in the disk outskirts, where we place less emphasis by considering only a relatively tight region around each galaxy for fitting. For the bulge profile, however, we leave the Sérsic index free in order to not only capture classical de Vaucouleurs bulges (with Sérsic indices around 4) but also pseudo-bulges (with Sérsic indices usually below 2) and bars (also low Sérsic indices and additionally

elongated shapes; see Section 1.1.1). For the same reason we also do not constrain the bulge - or, more precisely, the central component - to be round. However, we do constrain the bulge and disk to lie precisely on top of each other to avoid one of the components wandering off to fit overlapping point sources.

In addition to this two-component model, we fit two simpler ones: a single Sérsic model to represent those galaxies that either physically have just one component or where the data quality is not sufficient to constrain more than one component; and a “1.5-component” point source bulge plus exponential disk model. While the former is also routinely done in the literature since it is a wise decision in general to start with a simple model before adding complexity (also for Bayesian model selection), the latter is less common. The reason for us adding this model is that after fitting, we identified a population of bulges that were clearly present, but had ill-constrained parameters due to their faintness and small sizes. To obtain reliable magnitudes for these bulges, we considered it best to limit the freedom of the fitting parameters to a minimum and fit the point source model (consisting of only a magnitude and a position) instead of the Sérsic bulge model.

In summary, we fit three models to each galaxy: a single Sérsic, a Sérsic plus exponential and a point source plus exponential. More details are given in Sections 2.1.2 and 2.2.5. We note that according to the morphological classification performed by Driver et al. (2022) (see Section 1.1.1), the majority of galaxies in our sample can be approximately represented by one of these three models. However, in a statistical sense, our models are not an appropriate representation of the detailed structure of most galaxies. We therefore expect highly correlated residuals caused by - for example - spiral arms, rings, nuclear lenses, bars, AGN and bulges (especially if several of these features are present and the “bulge” Sérsic function is forced to compromise between fitting them) as well as asymmetries and irregular features of all kinds. These model shortcomings need to be considered when assessing the fit quality and for model selection, see Sections 2.1.3 and 2.2.7.

1.3.3 ProFit

Once the data are obtained and the model and its parameters are defined, the decision comes to the software to use for fitting. We opt to use `ProFit`⁷ (v1.3.2) which is a free and open-source, fully Bayesian two-dimensional profile fitting code specifically developed to fit the surface brightness distributions of galaxies (Robotham et al., 2017). `ProFit` offers great flexibility: there are several built-in profiles to choose from, it is easy to add several components of the same or different profiles, there is a choice of likelihood calculations and optimisation algorithms that can be used (various downhill gradient options, genetic algorithms, over 60 variants of MCMC methods), parameters can be fitted in linear or logarithmic space, it

⁷<https://github.com/ICRAR/ProFit>

is possible to add complex priors for each, as well as constraints relating several parameters; and much more. The pixel integrations are performed using a standalone C++ library (`libprofit`), making it both faster and more accurate than other commonly used algorithms such as `GALFIT` (Peng et al. 2010; see detailed comparison in Robotham et al. 2017). This allows us to fit galaxies with the computationally more expensive MCMC algorithms, overcoming the main problems of downhill gradient based optimisers: their susceptibility to initial guesses and their inability to easily derive realistic error estimates (e.g. Lange et al., 2016). This makes `ProFit` highly suitable for the decomposition of large sets of galaxies with little user intervention.

1.3.4 ProFound

`ProFit` (Section 1.3.3) requires a number of inputs apart from the (sky-subtracted) science image and the chosen model to fit, most importantly initial parameter guesses, a segmentation map specifying which pixels to fit, a sigma (error) map and a PSF image. To provide these inputs in a robust and consistent manner, the sister package `ProFound`⁸ (Robotham et al., 2018) was developed, which also serves as a stand-alone source finding and image analysis tool. The main novelties of `ProFound` compared to other commonly used free and open-source packages such as `Source Extractor` (Bertin & Arnouts, 1996) are that, rather than elliptical apertures, `ProFound` uses dilated “segments” (collections of pixels of arbitrary shape) with watershed de-blending across saddle-points in flux. This means that the flux from each pixel is attributed to exactly one source (or the background) and apertures are never overlapping or nested. It also allows for extracting more complex object shapes than ellipses while still capturing the total flux due to the segment dilation (expansion) process. This makes it less prone to catastrophic segmentation failures (such as fragmentation of bright sources or blending of several sources into one aperture), reducing the need for manual intervention or multiple runs with “hot” and “cold” deblending settings, making `ProFound` particularly suitable for large-scale automated analysis of deep extragalactic surveys (Robotham et al., 2018; Davies et al., 2018; Bellstedt et al., 2020).

Apart from the segmentation map, the main function of the package, `profoundProFound`, also returns estimated sky and sky-RMS maps (if not given as inputs) and a wealth of ancillary data including a list of segments and their properties such as their size, ellipticity and the flux contained. The latter is particularly useful to obtain reasonable initial parameter guesses for galaxy fitting; or for identifying certain types of sources (e.g. stars for PSF estimation). The package also contains many additional functions for further image analysis and processing, all within the same framework. In addition, combining `ProFound` with `ProFit` allows the user to estimate a PSF (see Section 2.1.1), entirely removing any dependence on external

⁸<https://github.com/asgr/ProFound/>

tools. Finally, both packages come with comprehensive documentation and many extended examples and vignettes which serve as great resources for newcomers to the fields of source extraction and galaxy fitting.

We use `ProFound` (v1.9.2) along with `ProFit` (v1.3.2) for all preparatory steps (image segmentation/source identification, sky subtraction, initial parameter estimates and PSF determination; see Section 2.1.1 for details) producing the inputs needed for the galaxy fitting with `ProFit`.

2 Bulge-disk decomposition pipeline

In this chapter, we present the bulge-disk decomposition pipeline we built around the `ProFound` and `ProFit` software packages. It consists of three main steps: preparatory work, galaxy surface brightness fitting and post-processing. The main inputs to the pipeline are a list of galaxies (GAMA CATAIDs) and the path to a directory containing (KiDS or VIKING) imaging files with associated weight maps and masks. The fully automated pipeline then performs all necessary steps to proceed from these basic inputs to a final catalogue of properties of the galaxies and their components, including numerous quality assessment parameters.

We start with an overview of all steps of the pipeline in Section 2.1. This is based on section 3 in Casura et al. (submitted) and describes the status quo at the time of submission, with `v04` being the corresponding `BDDcomp` DMU version on the GAMA database. It is limited to the processing of the KiDS g , r and i bands. Section 2.2 details how we have arrived at all of the procedures and the evolution of the pipeline from `v01` through to `v04`. Most of these studies focused on the KiDS r -band only. Finally, in Section 2.3 we summarise the changes made to the pipeline after the submission of Casura et al. (submitted). These updates were mainly performed in order to process VIKING data and resulted in `v05` of the `BDDcomp` DMU, which encompasses all 9 KiDS and VIKING bands ($u, g, r, i, Z, Y, J, H, K_s$).

We use the free and open-source programming language R (R Core Team, 2020) for all scripting.

2.1 Pipeline overview

This section is directly taken from section 3 in Casura et al. (submitted) with very minor adjustments and gives an overview of the entire bulge-disk decomposition pipeline, including all preparatory steps, the actual galaxy fitting and the post-processing. It reflects the status of the data processing at the time of submission, meaning it refers to the processing of the KiDS g , r and i bands and `v04` of the corresponding `BDDcomp` DMU on the GAMA database. The adjustments relative to Casura et al. (submitted) were mainly to fit the style of this thesis and to point to relevant further tests, examples and diagnostics in Section 2.2 that are not contained in Casura et al. (submitted).

2.1.1 Preparatory steps

The first part of our bulge-disk decomposition pipeline consists of several image processing steps that we group under the term “preparatory work”.

Cutouts and masking

KiDS imaging tiles are registered to the same pixel grid across all four bands (with matching weight maps and masks), such that a joint analysis of the bands is straightforward. They are also aligned such that the x -axis corresponds to right ascension (RA) and the y -axis to declination (Dec). Hence, we obtain a $400'' \times 400''$ cutout of the KiDS tile, associated weight map and mask for each object in our sample, and for each of the three KiDS bands we used (g, r, i). The masks of all three bands are then combined and all pixels which have a value greater than 0 in any of the masks are excluded from analysis. This results in approximately 20% of all pixels being masked out. This large fraction of masking is primarily due to the reflection halos of bright stars that are also clearly visible in the data (see de Jong et al. 2015 for details). We combine the masks in this way to ensure that the pixels used for analysis are exactly the same in all bands and so the results are most directly comparable. Objects for which the central pixel is masked ($\sim 20\%$ of all galaxies) are skipped in the galaxy fitting. More details on the choice of KiDS data products and the cutout size to use are given in Section 2.2.1.

Image segmentation

We perform image segmentation in order to determine which pixels to fit for each of our objects, identify other nearby sources, improve the background subtraction and obtain reasonable initial guesses for the galaxy parameters. This is performed on the joint g, r, i cutouts with `ProFound` in several steps.

First, we add the cutouts in the $g, r,$ and i bands using inverse variance weighting and compute the joint weight map. We then estimate the (joint gri) sky by running the stacked image through `profoundProFound` passing in the correct magnitude zeropoint, mask and weight, but leaving `skycut` on its default of 1. This means that all pixels with a flux at least 1σ above the median are progressively assigned to segments (collections of pixels belonging to an object) using an iterative process: starting with the brightest pixel in the image, segments are grown by adding neighbouring pixels with lower flux; new segments are started when a pixel shows more flux than its neighbours (within some tolerance) or when all neighbouring pixels above the `skycut` value have been assigned. Once all pixels above `skycut` have been assigned, the resulting segments are additionally expanded until flux convergence is reached. For more details, see Robotham et al. (2018).

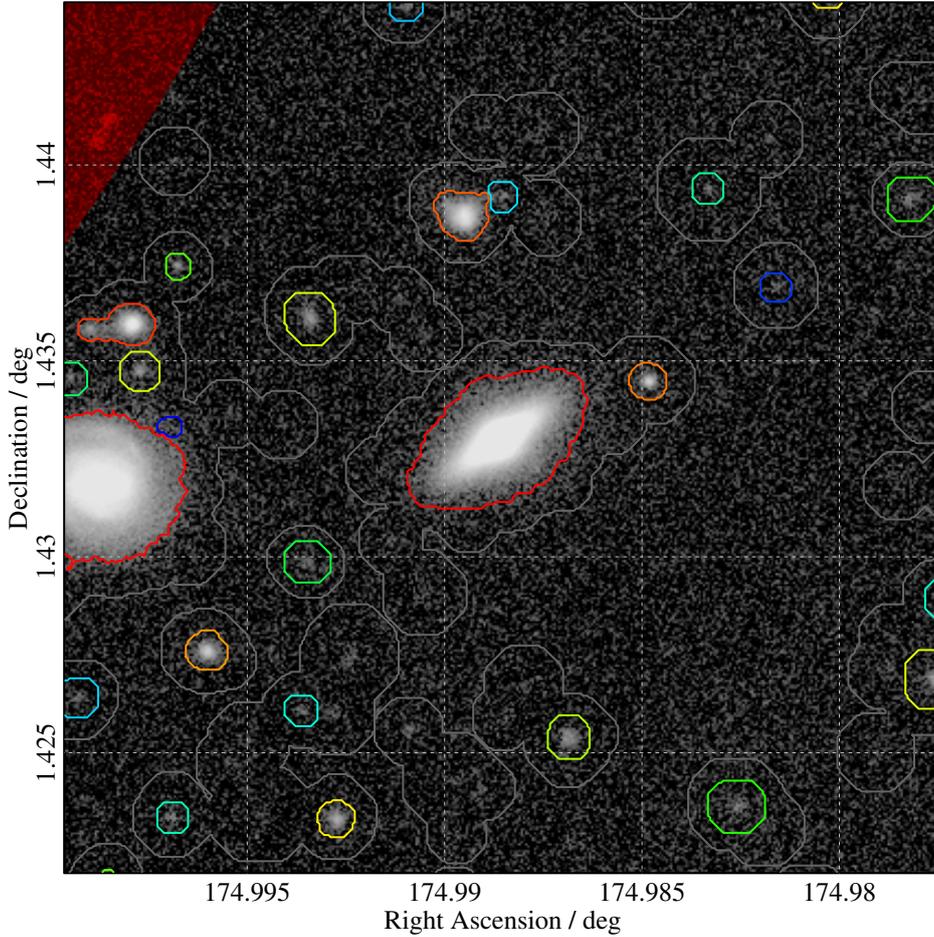


Figure 2.1: The ProFound segmentation map obtained for the GAMA galaxy 396740 overlaid on the KiDS r -band image. Note this is only a cutout of the full segmentation map showing the central $100'' \times 100''$. Identified objects (segments) are shown with contours, coloured from red to blue according to the flux contained. Grey contours indicate the more dilated segmentation map used for the background subtraction. Masked areas are shaded red.

Along the way, ProFound estimates the sky background several times since object detection relies on accurate background subtraction and vice versa.¹ For the final sky estimate, the already-dilated segments are expanded even further to ensure that no object flux will bias the background determination. This very aggressive object mask is indicated with grey contours in Figure 2.1. We use it for the joint- gri sky estimate here and also for the band-specific background determination detailed below (performed in the same way).

For the galaxy fitting, however, we decided to use tighter segments that do not push that deeply into the sky. Besides speeding up the fit, this results in the best possible fit to the inner, high signal-to-noise regions of the galaxy that we are most interested in and reduces

¹The sky variance can also be estimated, but in our case this is already provided as the KiDS weight maps and given to ProFound as an input.

the sensitivity to background subtraction problems, flux from the wings of other objects and features that cannot be captured by our models such as disk breaks and flares, and edge-on disks requiring the inclined disk model (van der Kruit & Searle, 1981). Note, however, that this choice comes with some trade-offs, most notably that the fit frequently overpredicts the flux outside the segment boundary. We address this in more detail in Section 2.1.3.

To obtain these tighter segmentation maps, we run `profoundProFound` again with the sky now fixed and a higher `skycut` value of 2. This means that only pixels with a flux at least 2σ above the background level are considered in the segmentation, which ensures that fewer noise fluctuations are “detected” and segment borders are smooth. In order to capture all flux of the galaxy wings, the segment for the object of interest (only) is then expanded further (using `profoundMakeSegimDilate`) such that its area increases by typically around 30%. This last step also ensures unbiased smooth borders of the segment since it is entirely independent of noise fluctuations. The resulting segmentation map is indicated with coloured contours in Figure 2.1 and is used for galaxy fitting in all bands, so that exactly the same pixels are fitted in each band (the segmentation statistics are re-calculated in each band).

Section 2.2.4 contains more details on the procedures for defining the segmentation map, reasons behind our decision to use different segmentation maps for background subtraction and galaxy fitting and the effect that different segment sizes have on galaxy fitting.

Background subtraction

KiDS tiles are background-subtracted already, however we opt to use the sky estimated by `ProFound` to even out inhomogeneities on smaller scales. For this, we split our $400'' \times 400''$ cutout into 16 square boxes and mask out all objects using the aggressively dilated segmentation map indicated with grey contours in Figure 2.1. The sky is then estimated as the median of the remaining (background) pixels in each box independently; and the solutions between the boxes interpolated with a bicubic spline.² This is done for each band independently, however the segmentation maps used to mask out objects are the same in all bands.

This procedure for the background subtraction was chosen after extensive testing during pipeline development which is summarised in Section 2.2.2. In short, we found that the `ProFound` sky adopted here does not subtract object wings while still homogenising the background well enough to avoid having to fit it along with the object of interest (introducing possible parameter degeneracies). It also decreases the sensitivity of the fit to the chosen segment size.

²This is done by `profoundProFound` internally; with the box size and the order of the interpolation spline being some of the variables we set.

Sigma maps

Once the image segmentation and background subtraction is completed, we also calculate the sigma (error) map for each cutout (independently in each band). This is a combination of the KiDS weight map (where $\sigma = 1 / \sqrt{weight}$) and the object shot noise. The latter is estimated as \sqrt{N} , where N is the number of photons per pixel (using positive-valued pixels only). This, in turn, is obtained by converting the image into counts using the gain provided in the meta-data associated with each KiDS tile.

PSF estimation

PSF fitting is performed on the background-subtracted $400'' \times 400''$ cutouts with corresponding masks and sigma maps in each band. The segmentation statistics returned by `ProFound` are used to identify isolated stars (round, bright, small and highly concentrated objects with few nearby segments). More details on the star candidate selection are given in Section 2.2.3. These objects are then fit with a Moffat (1969) function using `ProFit`; fitting all parameters except boxyness, i.e. the position, magnitude, full width at half maximum (FWHM), concentration index, axial ratio and position angle. Scale parameters are fitted in logarithmic space, a Normal likelihood function is used, initial guesses are taken from the segmentation statistics and we use the BFGS algorithm from `optim` (R Core Team, 2020), which is a fast downhill gradient optimisation using a quasi-Newton method published simultaneously by Broyden (1970); Fletcher (1970); Goldfarb (1970); Shanno (1970).

Some of the objects fitted above may not actually be suitable for PSF estimation as they can be too faint or bright (close to saturation), have irregular features, bad pixels or additional small objects included in the fitting segment. Unsuitable objects are excluded by a combination of hard cuts in reduced chi-squared (χ^2_ν), position and magnitude relative to the `ProFound` estimates and an iterative 2σ -clipping in FWHM, concentration index, angle and axial ratio. Again, more details can be found in Section 2.2.3. Finally, we take the median of the Moffat parameters of a maximum of 8 suitable stars (the closest 2 from each quadrant where possible to ensure an even distribution around the position of interest) and use these Moffat parameters to create a model PSF image. The size of the PSF image is adjusted to include at least 99% of the total flux; or to a maximum of the median segment size within which the stars were fitted, with pixels in the corners of the image set to zero to avoid having a rectangular PSF.

More details of the PSF fitting procedure and example diagnostic plots are given in Section 2.2.3, which also includes a description of other options we explored for obtaining PSFs, a summary of the PSF quality control and an overview of the PSF effects on galaxy fitting.

Outputs

For the fitting, we are only interested in the central galaxy and the closest neighbouring sources (for potential simultaneous fitting and to gain a better overview during visual inspection). Hence, we do not save the entire $2000 \times 2000 \text{ pix}^2$ cutouts used in the preparatory work as that would unnecessarily waste storage space and computational time used on reading and writing files. Instead, the image, corresponding mask, segmentation map, sigma map and sky image are cut down to the smallest possible size that includes the object of interest (centred) and its neighbouring (touching) segments before saving. These 5 files, the model PSF image and some ancillary information such as the segment statistics are the main outputs of the preparatory work pipeline and serve as inputs for the galaxy fitting, which we describe in Section 2.1.2.

2.1.2 Galaxy fitting

The second major step in the pipeline is the model fitting, based on the outputs of the preparatory work.

Inputs and models

We use the Bayesian code `ProFit` (Robotham et al., 2017) to perform 2-dimensional multi-component surface brightness modelling in each band independently, assuming elliptical geometry and a (combination of) Sérsic function(s) as the radial profile. The Sérsic (1963) function is given in Equation 1.6 and described by three main parameters: the Sérsic index n giving the overall shape (with special cases $n = 0.5$: Gaussian; $n = 1$: exponential and $n = 4$: de Vaucouleurs (1948) profile), the effective radius R_e including half of the total flux and the overall normalisation which we specify as total magnitude m . In addition, in 2 dimensions the axial ratio b/a gives the ratio of the minor to the major axis of the elliptical model and the position angle PA its orientation, while x and y are used to define the position in RA and Dec. Throughout this thesis, R_e refers to the effective radius along the major axis of the elliptical model. The Sérsic model is detailed in Graham & Driver (2005); and see also Section 1.3.2.

The data inputs for `ProFit` are a background-subtracted image, corresponding mask, segmentation map, sigma map and PSF. All of these are obtained during the preparatory steps (Section 2.1.1). On the modelling side, the main choices are the profile(s) to fit with initial parameter guesses and priors, the likelihood function to use, the fitting algorithm and convergence criteria; which are detailed below. In short, we choose to fit each object with 3 different models in a four-step procedure:

- (i) Single component Sérsic fits with initial guesses from segmentation statistics.

2.1. PIPELINE OVERVIEW

- (ii) Double component Sérsic bulge plus exponential disk fits with initial guesses from single component fits.
- (iii) Double component re-fits for a subset of galaxies which seemed to have the bulge and disk components swapped in step (ii), see below.
- (iv) “1.5-component” point source bulge plus exponential disk fits with initial guesses from double component fits.

Further details on this choice of models are given in Sections 1.3.2 and 2.2.5. Note that, for brevity, we will call the central component “bulge” throughout this work, even if it may not be a classical bulge. In particular, we do not distinguish classical bulges from pseudo-bulges, bars, AGNs, nuclear disks, combinations thereof or anything else that may emit light near the centre of a galaxy. Hence, we also use the term “bulge” for 1.5-component fits where the central component is unresolved and for double component central components with low Sérsic index and/or low axial ratios.

Table 2.1: The fitting parameters for each of our three models.

parameter	single	double		1.5-comp.	
		bulge	disk	bulge	disk
x -centre	free	free		free	
y -centre	free	free		free	
m	free	free	free	free	free
$\log_{10}(R_e)$	free	free	free	N/A	free
$\log_{10}(n)$	free	free	1	N/A	1
$\log_{10}(b/a)$	free	free	free	N/A	free
PA	free	free	free	N/A	free
boxyness	0	0	0	N/A	0

To implement our three models, we make use of two of the many models built into `ProFit`, namely the Sérsic and point source models. We fit all parameters except `boxyness` (i.e. we do not allow deviations of components from an elliptical shape) and, for the double and 1.5-component models, tying the positions of the two components together. Exponential disks are implemented using a Sérsic profile with the Sérsic index fixed to 1. This leaves 7 free parameters for our single Sérsic and 1.5-component models and 11 free parameters of the double component fits, which are summarised in Table 2.1; and see also Section 2.2.5. Scale parameters (Sérsic index, effective radius and axial ratio) are treated in logarithmic space throughout, i.e. the actual fitting parameters are $\log_{10}(X)$ for scale parameters X (Section 2.2.6).

The 1.5-component model is needed for around 15-30% of our double component systems where the bulge is too small relative to the image resolution to meaningfully constrain its Sérsic parameters (the exact number depends on the band due to the different PSF sizes).

With the point source profile, at least we can determine the existence of a second component and constrain its magnitude and hence the bulge-to-total (or AGN-to-total, bar-to-total, etc.) flux ratio. An example is given in Figure 2.21 (Section 2.2.5).

If the centre of an object is masked or the PSF estimation failed (which happens if large fractions of the surrounding area are masked), then the object is skipped and no fits are obtained. This affects approximately 20% of the galaxies. All other objects are fitted with all three models; and the best model is selected subsequently (see Section 2.1.3 on details of the model selection and Section 3.2 for the corresponding statistics).

Initial guesses

Since we use MCMC algorithms, our fits do not strongly depend on the initial guesses. However, reasonable starting parameters are still required for convergence within finite computing times.

The initial guesses for the single component Sérsic parameters are obtained directly from the segmentation statistics output by `profoundProFound` (Section 1.3.4) where we use the position, magnitude, effective radius (R_{50}), axial ratio and angle as given; and the inverse of the concentration ($1/\text{con}$) for the Sérsic index.

For the double component fits, we convert the single component fits into initial guesses as follows: the position is taken unchanged, the magnitude of the single component fit is split equally between the two components, the bulge and disk effective radii are taken as $1/2$ and 1 times the single component effective radius respectively, the Sérsic index of the bulge is set to 4 and its axial ratio to 1 (round), the disk axial ratio is set to the axial ratio of the single component fit and the position angles of both components are taken as that of the single component fit. See also Section 2.2.5.

Initial guesses for the 1.5-component fits are taken from the double component fits (after making sure the components are not swapped, see below), where the bulge magnitude is used as the point source magnitude and the disk parameters are taken unchanged.

Priors, intervals, constraints

All parameters are limited to fixed intervals. In addition, there can be constraints between parameters (such that, e.g., the bulge and disk positions can be tied together). If a (trial) parameter is outside the bounds of its interval or constraint during any step of the fitting process, `ProFit` moves it back onto the limit before the likelihood is evaluated.

The limits for single-component fits are given in Table 2.2. In addition, the position angle is constrained such that if it leaves its interval, it is not just moved back onto the limit but

2.1. PIPELINE OVERVIEW

Table 2.2: The fitting limits for single-component fits.

parameter(s)	lower limit	upper limit
x - and y -centre	0	cutout side length
magnitude	10	25
effective radius	0.5 pixels	$\sqrt{2}$ cutout side length
Sérsic index	0.1	20
axial ratio	0.05	1
position angle	-90°	270°

jumps back 180° (which is the same angle, just more in the centre of the fitting interval).

There are no additional priors or constraints for single component fits. This means that in effect, we use unnormalised uniform priors which are 1 everywhere in the respective interval and zero otherwise. For scale parameters (which are fitted in logarithmic space) the priors are uniform in logarithmic space, corresponding to Jeffreys (1946), i.e. uninformative, priors.

The limits and constraints for double and 1.5-component fits are the same as for the single component fits (for both bulges and disks), except for the magnitude where the individual component magnitudes have infinity as their upper limit and instead the *total* magnitude is constrained to be within the magnitude limits. This is most consistent and also allows the fitting procedure to discard one of the two components for systems which can equally well be fitted with a single Sérsic function (we then take this into account in the model selection). Further considerations on additional constraints between parameters (that we did not enforce in the end) are summarised in Section 2.2.5.

Note that the above procedure results in unnormalised posteriors. The lack of normalization does not impede our analysis because the only time when we compare posteriors is during model selection, where we effectively fold the normalisation into the calibration during visual inspection (Section 2.1.3).

Likelihood function

We use a Normal likelihood function for all fits. We have tested a t-distribution likelihood function which is less sensitive to outliers/unfittable regions; but found that the Normal likelihood function is better suited to our needs for several reasons.

First of all, the t-distribution fits often preferred to use the freedom of the bulge parameters to fit disk features instead (e.g. rings, bumps, flares, etc. that cannot be captured by the exponential model), treating the bulge as an outlier since the t-distribution prefers a few strong outliers (the bulge pixels) over many weak ones.

Second, the t-distribution fits fail for galaxies which are perfectly fitted by the model since

then the errors truly are distributed Normally. This is a relatively common occurrence.

Hence some galaxies ($\sim 20\%$) need to be fitted with a Normal distribution anyway, which, third, makes model selection much harder since the likelihood values obtained with different likelihood functions cannot easily be compared to each other.

See Section 2.2.6 for more details and examples.

Fit and convergence

All fits are performed on the sky-subtracted image within the galaxy segment only using the `convergeFit` function from the `AllStarFit` package (Taranu, 2022). This function uses a combination of different downhill gradient algorithms available in the `nloptr` package (Johnson, 2017) followed by several MCMC fits with `LaplacesDemon` (Statisticat & LLC., 2018) until convergence is reached. The exact procedure is described in Section 2.2.6.

The downhill gradient algorithms are used first to improve the initial guesses. The MCMC chain is not very sensitive to the initial guesses, but converges much faster if starting closer to the peak of the likelihood. Once the MCMC chains have converged, 2000 further likelihood points are collected to ensure a stationary sample for the subsequent analysis of the galaxy. We test this in Section 2.2.6.

We only fit the primary object of interest. While simultaneously fitting neighbouring sources is possible in `ProFit` and might have improved the fit on a few objects, the effects are generally small since the galaxies we study are not in highly crowded fields and the segmentation process usually excludes the vast majority of the flux from other sources. This is especially true since we use tight fitting segments within which the galaxy flux is dominant (cf. Section 2.1.1); and considering that the watershed algorithm of `ProFound` cleanly separates even overlapping sources, so neighbours are automatically masked (Section 1.3.4). Hence we opted for the simpler and computationally cheaper option of only fitting the main objects. We confirm that this does not lead to major biases in Section 4.2.2).

An example fit for an object which is well-represented by our 2-component model is shown in Figure 2.2.

Component swapping

Approximately 20-30% of the double component fits have their bulge and disk components swapped, i.e. the exponential component fitting the central region and the Sérsic component fitting the wings (this is a common problem in galaxy fitting, first pointed out by Allen et al. 2006). In particular, the freedom of the Sérsic component is often used to fit disks that do not follow pure exponential profiles while at the same time being the dominant component

2.1. PIPELINE OVERVIEW

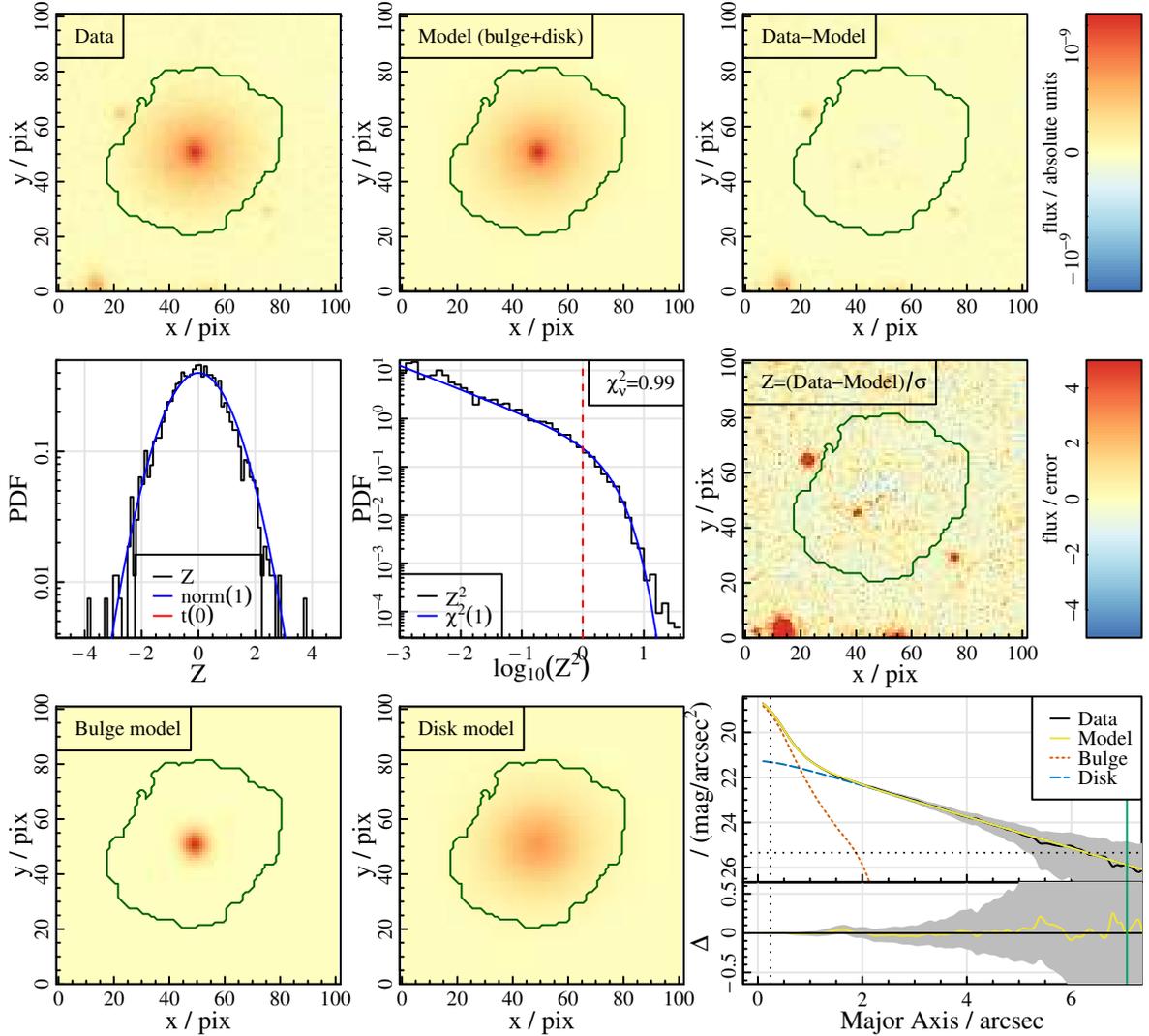


Figure 2.2: The result of the 2-component (Sérsic bulge plus exponential disk) fit for the galaxy 611298 in the KiDS r -band. **Top row:** the data, 2-component model and residual between them shown in absolute values of flux given by the colour bar on the right. The green contour indicates the segment used for fitting. Note that the flux scaling here is non-linear and optimised to increase visibility of galaxy features, but is the same in all 3 panels. **Middle row:** goodness of fit statistics. The right panel is the normalised residual Z (colour bar on the right) capped at $\pm 5\sigma$. The left panel is the one-dimensional distribution (measured probability density function, PDF) of Z within the segment; with blue and red curves showing a Normal distribution with a standard deviation of 1 and a Student's-t distribution with the relevant degrees of freedom for comparison. The middle panel shows the measured PDF of Z^2 compared to a χ^2 -distribution of 1 degree of freedom (blue). The reduced chi-squared, χ^2_v (sum over Z^2 divided by the degrees of freedom of the fit) is given in the top right corner. **Bottom row:** The bulge and disk models in 2 dimensions on the same flux scale as the top row; and the bulge, disk, and total model compared against the data in one-dimensional form (azimuthally averaged over elliptical annuli). The FWHM of the PSF and the approximate 1σ surface brightness limits are indicated by vertical and horizontal dotted lines for orientation. The vertical solid green line indicates the segment radius beyond which our model is extrapolated. The pixel scale is $0''.2$ for KiDS data, i.e. $1''$ corresponds to 5 pix.

in terms of flux (which is the case for most galaxies). To solve this problem, we devised an empirical swapping procedure guided by the visual inspection of a subsample of our galaxies.

First, we select the galaxies that are most likely to have swapped components based on a cut in the plane of the ratio of Sérsic indices and the ratio of effective radii for the single component fits and the bulge of the double component fits. The reasoning for choosing this parameter space to calibrate the cut was that we would generally expect bulges alone to be more concentrated (i.e. smaller effective radius and higher Sérsic index) than when mixed with their respective disks in the single Sérsic fits. This results in approximately 30% of our sample being flagged as possibly swapped, which we then re-fit in a second step.

The re-fit is performed in exactly the same way as the original fit, except that we now use the results of the previous double-component fit as initial guesses, swapping around the bulge and disk components (except for the bulge Sérsic index for which we use a value of 4). While the MCMC chain is less sensitive to initial guesses than a downhill gradient algorithm, it will still show some dependency for finite run-times. In particular, in our double component model the two components are *nearly* interchangeable with the only difference being the Sérsic index (fixed to 1 for the disk, free for the bulge). Hence there will always be 2 high maxima in likelihood space, which are far apart in the 11-dimensional parameter space. Moving from one to the other would require changing 9 parameters (all except position) at once in the right direction and hence is statistically unlikely. Therefore, we assist the code in finding the other maximum by manually swapping the initial guesses.

In approximately 5% of all re-fits, the code still converges on the same fit as before the swapping, but in most cases we find another likelihood maximum which corresponds to the bulge and disk components being reversed. As a third step we then select between the old and the new fit to obtain the physically more appropriate one. For this we first check whether either of the fits has a bulge Sérsic index smaller than 2 *and* a bulge effective radius at least 10% larger than the disk effective radius *and* a bulge-to-total ratio above 0.7 (i.e. the “bulge” component is close to exponential, larger than the disk and contains the majority of the flux). If this is the case for only one of the fits, we choose the other one. If it is true for both or neither of the fits, then we apply our main criterion, which is that we choose the fit with the higher absolute value of bulge flux in the central pixel. These selection criteria are again based on visual inspection guided by the notion that we expect the bulge to be smaller and steeper than the disk and have proven to work very well. Note that the fit we select in this way is the one that is physically better motivated (i.e. with the bulge at the centre), and not necessarily the one which is statistically better.

After this procedure, the number of galaxies which still have the bulge and disk components swapped (and are classified as double component fits in model selection) is reduced to ~ 1 -2%. The corresponding diagnostic plot based on the visual inspection as well as an example are

shown in Section 2.2.7 (Figures 2.35, 2.36 and 2.37).

2.1.3 Post-processing

To assess and improve the quality of the fits, we perform a number of post-processing steps, namely the flagging of bad fits, model selection, and truncating fits to segment radii.

Flagging of bad fits

After all three models have been fitted to all objects, we run them through our outlier flagging process (separately in each band). Each model is treated separately first; they are then combined during the model selection (see below).

The criteria for flagging bad fits (outliers) are: a very irregular fitting segment, an extreme bulge-to-total flux ratio, numerical integration problems, a parameter hitting its fit limits, poor χ^2 statistics, a large distance between the input and fitted positions and a small fraction of model flux within the fitting segment. Additionally, there are some cautionary flags that identify fits which should be treated with extra care. All criteria are derived from and calibrated against visual inspection and described in more detail below and in Section 2.2.7 (including diagnostic plots). For orientation, we give the percentage of affected r -band fits in parentheses for each criterion below. The corresponding percentages for the g and i band fits are given in Table 3.3a in Section 3.2.2. Note, however, that bad fits tend to fall into multiple of these categories, so the total number of bad fits is smaller than the sum of flagged objects in each category. Overall, approximately 9%, 11% and 10% of all non-skipped fits are flagged in the g , r and i bands, respectively (after model selection).

Very irregular segment (5.4%): we calculate the difference between the magnitude of the model contained within the segment and the magnitude contained within the “segment radius”, which is defined as the maximum distance between the centre of the fit and the edge of segment. Objects where this magnitude difference is larger than 0.3 are flagged, as this is an indication for irregular segments (shredded, partly masked or cut off by another object for example). Note this criterion often shows overlap with the criterion on the fraction of model flux contained within the segment (see description below).

Extreme bulge-to-total ratio (0.1%): we flag double component and 1.5-component fits with a bulge-to-total ratio smaller than 0.001 or larger than 0.999 because in these cases the second component has negligible flux and a single component fit is better suited.

Numerical integration problems (0.2%): `ProFit` includes an oversampling scheme for accurate pixel flux integration where pixels containing steep flux gradients are recursively oversampled up to an oversampling factor of 4096; in the central pixel even up to $\sim 10^9$

(for more details see Robotham et al., 2017). However, for very extreme model parameters, even this procedure may not be accurate enough anymore, leading to significant errors in the pixel flux calculations. This could be improved by changing the default oversampling values to achieve higher accuracy (at the cost of increased computational time), however we opted for simply excluding those cases since usually this only happens for unresolved bulges which are better represented by the 1.5-component fits anyway.

Parameter hitting limit (5.8 %): we flag objects where the magnitude, effective radius or Sérsic index hit either of their limits (cf. Section 2.1.2); or the axial ratio hit its lower limit (for double component fits this applies to both components individually). The axial ratio upper limit is not flagged because fits are allowed to be exactly round, but there is a cautionary flag for all objects which hit any of its parameter limits (6.5 %). We also add a cautionary flag for suspiciously small or large errors on any parameter, where “suspicious” is defined as being an outlier in the respective distribution of errors (2.1 %).

Poor χ^2 statistics (0.1 %): we flag fits with a χ^2_{ν} larger than 80; or where the χ^2 in the central pixel is more than 1000 times larger than the average χ^2 per pixel since that is an indication that the bulge was not fitted.

Large distance between input and output position (0.3 %): we flag fits with a distance between the input and output position of more than 2'' (10 pix), which are usually highly asymmetrical objects, mergers, objects with very nearby other objects (especially small objects embedded in the wings of much larger objects), or objects in regions of the image with unmasked instrumental effects. Often the fitted object then is not the one that we intended to fit. There is also a cautionary flag for offsets above 1'' (1.3 %).

Small fraction of model flux within fitting segment (1.4 %): we flag fits where the amount of model flux (of any component) that falls within the fitting segment is less than 20 %. With so little flux to work on `ProFit` cannot constrain the parameters well anymore and these are often objects which are cut off by a masked region (e.g. a bright star) or other nearby objects. There is a cautionary flag for objects where the fraction of model flux (of any component) that falls within the segment is less than 50 % (9.3 %).

Model selection

As detailed in Section 1.3.1, model selection in Bayesian analysis is performed by computing the posterior odds ratio, which in turn depends on the marginalised likelihoods for the two models in question. Since this is often difficult to compute in practice, many information criteria tests have been developed which are based on the (non-marginalised) likelihood (or χ^2) combined with some penalty term depending on the number of model parameters.

2.1. PIPELINE OVERVIEW

This penalty term serves to judge whether a more complicated model is justified and takes the role of Ockham’s factor. Commonly used tests include the Akaike information criterion (AIC, Akaike, 1974), the Bayesian information criterion (BIC, Schwarz, 1978), or the deviance information criterion (DIC, Spiegelhalter et al., 2002). We choose to use the deviance information criterion, which is usually recommended over the AIC or BIC in Bayesian analysis (Hilbe et al., 2017) and straightforward to compute from an MCMC output. Brief tests using the BIC or the estimated log marginal likelihood output by `LaplacesDemon` showed similar results.

The DIC is a direct output of the `LaplacesDemon` function (see Section 2.1.2) and is defined as:

$$\text{DIC} = \text{Dev} + pD = \text{Dev} + \text{var}(\text{Dev})/2, \quad (2.1)$$

where pD is a measure of the number of free parameters in the model and $\text{Dev} = -2 \times \log\text{-likelihood}$ is the deviance. In theory, then, if the DIC difference ΔDIC between two models is negative, the first model is preferred and if it is positive, then the second model is preferred; with differences larger than approximately 4 being considered meaningful (Hilbe et al., 2017). However, for the case of galaxy fitting where many features are present that cannot be captured by the model (bars, spiral arms, disk breaks or flares, tidal tails, mergers, foreground objects, etc.), we want to choose the model that we consider physically more appropriate rather than better in a strictly statistical sense. This requires visual classification, logical filters, detailed simulations or a manual calibration of the ΔDIC cut (or whichever other chosen diagnostic) by visual inspection of a representative sub-sample (e.g. Allen et al., 2006; Simard et al., 2011; Vika et al., 2014; Argyle et al., 2018; Kruk et al., 2018; Cook et al., 2019; Robotham et al., 2022). We choose the latter approach, which has the added advantages that we do not need to worry about normalising our likelihoods (cf. Section 2.1.2), hence circumventing dependencies of the results on prior widths; nor the fact that our pixel values are correlated (due to the PSF) – these effects are simply folded into the visual calibration.

We use a random sample of ~ 700 non-skipped objects per band (i.e. ~ 2000 objects in total) for the calibration; and a further 1000 r -band objects that were previously inspected for cross-checking the results. In addition, our model selection procedure takes into account some of the outlier flagging. For each of the ~ 700 objects in each band, we visually inspected the fits of all three models and classified the object into one of the categories: “single component”, “1.5-component”, “double component”, “not sure if 1.5- or double component”, “not sure at all”, “unfittable” (outlier). We then calculate the DIC differences between all three models (i.e. $\Delta\text{DIC}_{1-1.5}$, ΔDIC_{1-2} and $\Delta\text{DIC}_{1.5-2}$) and calibrate them for model selection in two steps: first, we select between single component fit or not; of the ones that are not single component fits we then select between double component or 1.5-component fits.

For the first step of model selection calibration, the $\Delta\text{DIC}_{1-1.5}$ and ΔDIC_{1-2} cuts are op-

timised such that the minimum number of fits is classified wrongly. “Wrong” in this case means a fit was manually classified as “single” but is now a double/1.5; or a fit was manually classified as “1.5”, “double”, or “not sure if 1.5 or double” but is now a single. “Unfittable” and “not sure at all” cases are ignored. For the second step of model selection calibration, the $\Delta\text{DIC}_{1.5-2}$ cut is optimised in the same way; where “wrong” now means that the fit was manually classified as “1.5” but is now a double or vice versa, with all other categories being ignored. For the two steps of the calibration, we bootstrap the manual sample 1000 and 500 times respectively and repeat the optimisation to get an estimate of the error on the chosen DIC cuts. These errors are chosen as the 1σ quantiles (i.e. they contain the central 68% of DIC cut distributions). Hence, all our calibrated DIC cuts have a median, a lower limit and an upper limit. Any object within these limits is flagged as unsure in the model selection, i.e. the DIC differences are not conclusive for this object.

To perform the actual model selection, the calibrated DIC cuts in each band are then applied to the entire sample, again in a two-step procedure: the single component fit is selected if neither of the 1.5- or double component fits are significantly better (as indicated by the DIC differences). Double component fits need to be significantly better than 1.5-component fits, too. In all cases, if the DIC difference is very clear, we do the model selection first; then flag objects as outliers if needed.³ In the unsure region of the DIC difference, we choose the model that is not flagged as outlier; if neither is flagged, the DIC cut is applied.

Compared against visual inspection (keeping in mind that visual classification is not free of errors either), roughly 7%, 9% and 6% of the galaxies end up in the wrong category in total in the g , r and i bands respectively (in both steps of model selection combined, ignoring cases which were visually classified as “unsure”). Table 2.3 gives the detailed confusion matrix for the r -band. Note that we do not consider the success of the outlier flagging here, so for outliers we show what the galaxy would have been classified as if it were not flagged (absolute value of the `NCOMP` column in our catalogue). We highlight those galaxies that are correctly classified in bold and show those that were ignored during the model selection calibration process in grey font. The remaining (black) numbers add up to the 9% quoted above. Corresponding confusion matrices for the g and i bands are given in Section 2.2.7 (Tables 2.5a and 2.5b). Note that since we minimise the *total* number of fits classed wrongly, there is a slight bias against the rarer categories in the automated model selection. For example, the relative fraction of true 1.5-component objects (as per the visual inspection) that is classified wrongly by the automated selection is higher simply because 1.5-component objects are much rarer than single or double component objects.

In addition to this band-specific model selection, we perform a joint model selection for all

³This means that it is possible (and not uncommon) that a galaxy which is classified as an outlier has a non-flagged fit in another model (but the fit that was chosen was significantly better than the other one, despite it being an outlier).

2.1. PIPELINE OVERVIEW

Table 2.3: The confusion matrix for our model selection based on a DIC difference cut compared against visual inspection for the r -band. All values are in percent of the total number of visually inspected r -band galaxies. Bold font highlights galaxies classified correctly, while grey shows those that were ignored during the calibration.

visual classification	number of components		
	1	1.5	2
“single”	41.6	0	2.7
“1.5”	2.2	2.4	0.9
“double”	3.1	0.1	9.2
“1.5 or double”	0.3	0.6	3.0
“unsure”	16.1	0.4	13.1
“unfittable”	0.9	0.6	2.7

three bands. For this, we sum the DIC values of all three bands for each model before computing the DIC differences. Then we perform the same optimisation procedure as for the single bands (using all ~ 2000 visually classified objects across the three bands) to obtain cuts in DIC difference which we subsequently apply for the model selection. Note that the model selected in this way is by necessity a compromise between the different bands, which have different depth and seeing. In this procedure, approximately 9% of fits are classified wrongly across all bands compared to visual classification. The corresponding confusion matrix is shown in Table 2.5c in Section 2.2.7.

The accuracy of the model selection is also confirmed using simulations, to the extent to which our simulations allow us to do so (see Section 4.2.3 for details).

Truncating to segment radii

As detailed in Section 2.1.1, we produce segmentation maps that define the fitting region, meaning that only pixels within the fitting segment are considered during the evaluation of the likelihood of the model (equivalent to giving all pixels outside the segment zero weight in the fit). We choose tight fitting segments (cf. Section 2.1.1) in order to obtain the best possible fit in the inner, high signal-to-noise ratio regions of the galaxies and be less sensitive to disk breaks, flares, nearby other objects, sky subtraction problems and similar. The disadvantage of this approach is that profiles are not necessarily forced to zero for large radii, i.e. our Sérsic fits often show unphysically large effective radii combined with high Sérsic indices.

To mitigate this effect, we define a “segment radius” for each galaxy segment, which is simply the maximum distance between the fitted galaxy centre and the edge of the segment and can be understood as the upper limit to within which our model is valid. We then calculate the “segment magnitude”, m_{seg} , which is the magnitude of the (intrinsic, not PSF-convolved) profile integrated to the segment radius (rather than infinity); and the “segment effective

radius”, $R_{e,seg}$, which is the radius containing half of the flux defined by the segment magnitude. These values (and quantities derived from them, such as segment bulge-to-total flux ratios) are provided in the catalogue (labelled `*_SEGRAD`) and we strongly recommend using these instead of the Sérsic values integrated to infinity whenever they are available. For a direct parameter comparison to other works, the values in those catalogues should also be appropriately truncated.

In the following, we explain this recommendation in more detail; with further points to note in Sections 2.2.4, 4.1.4 and 4.2.4.

Figure 2.3 illustrates the effects produced by our tight fitting segments, how to mitigate those by truncating the magnitude and effective radius appropriately; and the circumstances under which this correction is necessary. For two example galaxies - 214264 and 3896188 - we show a detailed comparison of our single Sérsic fit to both a fit using a larger segment (from v05 of our pipeline, see Section 2.3) and to the fit obtained in Kelvin et al. (2012). We present a more general (statistical) comparison of our fit results to those of Kelvin et al. (2012) in Section 4.1.4, where we also give more details on how they derived their fits. For the purposes of the analysis in this section it suffices to say that Kelvin et al. (2012) used much larger fitting regions than we do, while the remaining analysis is in many ways analogous to ours (although they use different data, code and procedures in detail).

Focusing on the left half of Figure 2.3, the top two rows show the KiDS r -band data, our single Sérsic model, the residual and various goodness of fit statistics as described in the caption of Figure 2.2. Rows three and four show the same for a larger fitting segment as indicated by the green contour. Rows five and six again show the same for the Kelvin et al. (2012) fit, where we note that this was originally performed on r -band Sloan Digital Sky Survey (SDSS, York et al. 2000) data but is now evaluated on the r -band KiDS data. The Kelvin et al. (2012) fits were performed on cutouts larger than the size shown here, i.e. they include all visible pixels (and more) in the fit. Note that the reduced chi-squared value quoted in the bottom middle panel of each set of plots always is evaluated within the smallest segment so that they can be directly compared.

Finally, the bottom panels show a direct comparison of the one-dimensional profiles of all three fits, which we will now study in detail. We show the surface brightness (azimuthally averaged over elliptical annuli) against the projected major axis for the data (solid black line with grey uncertainty region), our model fit for the fiducial segment (dashed blue line) and the larger segment (dash-dotted pink line) and the Kelvin et al. (2012) model fit (dotted orange line). The vertical green solid and dashed lines indicate the segment radii (for the two segment sizes respectively) beyond which our model is an extrapolation. The vertical dotted line shows the half width at half maximum (HWHM) of the PSF and the horizontal dotted line is the 1σ surface brightness limit of the data. The inset in the bottom left of this plot

2.1. PIPELINE OVERVIEW

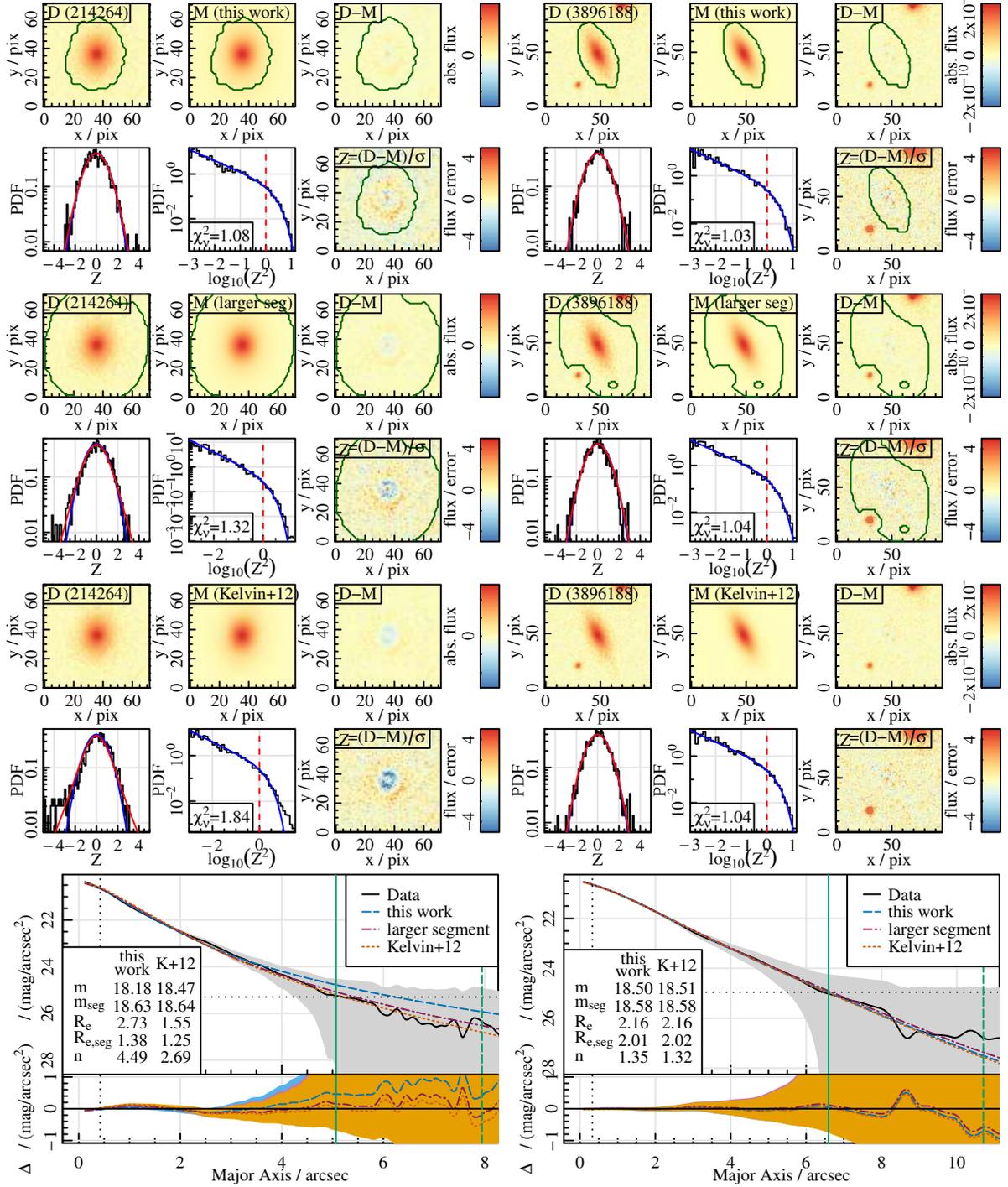


Figure 2.3: **Left:** detailed comparison of our single-Sérsic fit, our fit using a larger segment, and the Kelvin et al. (2012) fit to the galaxy 214264, which in reality is a 1.5-component system. **Right:** the same for galaxy 3896188, which is well-described by a single Sérsic component. **Top two rows:** our fit to the KiDS r -band data with panels the same as those in Figure 2.2. **Rows three and four:** the fit we obtained by using a larger fitting segment as indicated. **Rows five and six:** the Kelvin et al. (2012) fits (originally performed on SDSS r -band data) evaluated on the KiDS r -band data, which use a fitting region larger than the cutout shown. **Bottom panels:** direct comparison of the one-dimensional profiles, see text for details.

shows the fitted magnitude m , effective radius R_e in arcseconds and Sérsic index n values for our and the Kelvin et al. (2012) fits; and the corresponding segment-radius-truncated values for m and R_e . Below that, we show the difference between all three models and the data (with errors): our fiducial fit in blue with a dashed line, the fit in the larger segment in pink with a dash-dotted line and the Kelvin et al. (2012) fit in orange with a dotted line.

Our model is a better fit to the inner regions of the galaxy than the Kelvin et al. (2012) fit (out to about $2''$, also evident from the two-dimensional plots and from the reduced χ^2 -value within the segment decreasing from 1.84 to 1.08), owing to the higher Sérsic index which better represents the steep bulge at the centre. However, it has a large effective radius and considerable amounts of model flux at large radii which are not observed in the data. In particular in the region beyond the segment radius, where our model is merely extrapolated, it is clearly oversubtracting the data (also visible in the 2-dimensional plots). Correspondingly, the truncated segment quantities differ substantially from the fitted Sérsic values. The Kelvin et al. (2012) fits, instead, use a larger fitting region and hence follow the data out to larger radii, which results in a worse fit of the central regions but does not contain such large amounts of excess flux beyond the surface brightness limit. Hence, truncating to segment radii has a smaller effect on the parameter values. The truncated values for both models are then in reasonable agreement with each other, except for the Sérsic index, for which no truncated version exists as it would be unclear how to define such a value. Our fit in the larger segment is in between the two others in all respects, since it has a fitting region intermediate to the other two.

Note that the differences only come about when the model is not (in a formal statistical sense) a good representation of the data, i.e. when there is a need to compromise between fitting different regions. In the case of the left side of Figure 2.3, the galaxy shown is better described by a 1.5-component model ($m^B = 20.47$, $m^D = 18.79$, $R_e^D = 1.89''$), although in general there are many objects in our sample for which even a two-component model cannot capture all aspects of the data. For comparison, in the right half of Figure 2.3, we show a galaxy that is well-described by a single Sérsic model: here, both our and the Kelvin et al. (2012) fits arrive at virtually the same solution despite the different fitting regions. In fact, all three models and the data are nearly indistinguishable all the way down to the 1σ surface brightness limit.

In short, there is no perfect way to fit a Sérsic function to an object which intrinsically does *not* have a pure Sérsic profile. For such objects, which unfortunately comprise the majority of our sample, the fitted parameters will always depend on the exact fitting region used as well as the quality of the data (its depth in particular). Most previous works, including Kelvin et al. (2012), opted to use large fitting regions in order to include enough sky pixels to ensure that the profiles are constrained to approach zero flux at large radii (although a Sérsic function technically never reaches zero exactly). Here, we choose a different approach by

using smaller fitting segments. This means that the profiles are not constrained to approach zero flux at large radii. Instead more emphasis is placed on adequately representing the inner regions of the galaxies. We choose this approach since it is most appropriate for our science case, where we are primarily interested in comparing the high signal-to-noise regions of galaxies from the same data set amongst each other. In addition, it decreases the sensitivity of our fits to deviations from a Sérsic profile in the low surface brightness wings of objects (arguably no galaxy truly follows a Sérsic profile to infinity) as well as nearby other objects and inaccuracies in the sky subtraction. We stress that this means that our parameters are not directly comparable to other works using larger fitting segments. In particular, our Sérsic indices tend to be systematically higher (see Section 4.1.4) since high Sérsic indices result in high amounts of flux at large radii and are hence suppressed when constraining the models to zero flux at large radii. Magnitudes and effective radii can be compared to those of other studies by truncating to segment radii.

2.2 Pipeline development

This section contains technical details on the development of the bulge-disk decomposition pipeline and the numerous tests performed at various stages during that process with a particular focus on the preparatory work. Building on the brief overview over the final v04 pipeline in Section 2.1 (reproduced from Casura et al., submitted, their section 3; a summary of which can also be found in the description accompanying the `BDDcomp` DMU on the GAMA database), we provide additional information on each step. This includes supplementary tests and diagnostic plots as well as explanations of how we have arrived at each of the decisions listed in Section 2.1 and what other options have been explored over the course of the pipeline development. Further, we detail the pipeline evolution from v01 through to v04; while the changes made for v05 are given in Section 2.3. A summary of the key changes between all five different versions is provided in Section 3.1.

Note, however, that many of the procedures presented here have been developed in an iterative way since most choices are interconnected and influence or depend on each other (both within the preparatory work pipeline and also for the galaxy fitting). `ProFit` and even more so `ProFound` were also actively developed during the time of pipeline development with frequent changes in the procedures and default values, necessitating corresponding adaptations to our routines. Consequently, the decisions presented here were not done in any strict chronological order and instead many of the test runs varying all of those parameters were performed and repeated numerous times at different points during pipeline development. We do not describe all of those iterations here, but instead try to summarise the most important findings and give evidence of when each decision was made (be it originally or finally). Unless stated otherwise, the tests in this section have been carried out on the KiDS r -band data.

2.2.1 Data inputs and setup

We begin with a more detailed description on the general setup of the preparatory work pipeline and the data products it uses as inputs.

KiDS data products

Before starting the data processing, we needed to decide which of the KiDS data products to use. The basic data unit for KiDS photometric data is a science tile. Each tile is approximately 1 deg^2 or $18500 \times 19500 \text{ pix}^2$ in size, is astrometrically and photometrically calibrated with a uniform pixel size of $0''.2$ and a magnitude zeropoint of zero. They are stacked images (coadds) composed of 5 (4 in u) slightly offset (dithered) frames taken in direct succession and arranged to close the gaps between the charge-coupled device (CCD) chips (see details in de Jong et al. 2015, 2017; Kuijken et al. 2019 and cf. also Section 1.2.2).

Since both re-gridding and stacking of individual exposures can result in problems for accurate photometry (e.g. correlations between pixels, abrupt changes in the background levels or PSF distortions), we briefly considered working with individual dithers instead of the science tiles. However, the dithers are not publicly available which means that they did not benefit from the same analysis and data reduction procedures as the full tiles. In particular, parts of the processing and quality control were performed on the full tiles and not at the individual dither level, see details in de Jong et al. (2015, 2017); Kuijken et al. (2019), leaving the dithers at a lower data quality. The science tiles, on the other hand, show a very high data quality despite the three main problems of re-gridded and stacked images listed above: pixel correlations are present, but limited to very small scales (see Section 2.2.2). Abrupt changes in the background levels at the edges of individual CCDs are very rare due to the careful background subtraction performed by the KiDS team before stacking. PSF distortions are minimal since all dithers were taken in direct temporal succession, so the seeing did not vary much between individual exposures.

We concluded that it is more advisable to use the publicly available science tiles for KiDS, which the team themselves also use to create their photometric catalogues. This also significantly facilitated the data processing and galaxy fitting, especially since `ProFit` did not yet support multi-frame fitting at that stage. Note that for VIKING data, our decision was different and we decided to use the individual frames instead of the coadds (see Section 2.3.1).

Each science tile has an associated weight map and flag image (mask). The weight images give the inverse variance for each pixel in the same flux units as the science tiles. They include information from the dithering pattern, flat fields, dark frames and the first step of the masking: defects that affect individual frames, such as cosmic rays, hot and cold pixels, saturated pixels or satellite tracks are masked in individual frames when adding them up to

2.2. PIPELINE DEVELOPMENT

tiles, with the weight of that pixel (in the final co-add) being reduced accordingly (de Jong et al., 2015).

We noted at various points during our analysis (see, e.g., Section 2.2.2), that the KiDS weight maps are conservatively estimated, i.e. the errors resulting from $1/\sqrt{weight}$ are slightly larger than the typical standard deviation of sky pixels (after our background subtraction and object masking routines). However, in view of the wealth of information included in the weight maps and in particular the strongly and abruptly varying weights across the tiling pattern, we decided it is still better to use those conservative weights than to not use them at all or try to estimate them using `ProFound`. Due to this, perfect fits (both for PSF estimation and galaxy fitting) usually have χ^2_ν values around 0.7 to 0.9 instead of the nominal expectation value of unity.

In addition to the weight maps, each KiDS tile has an associated mask image. These are mostly produced by an automated procedure developed specifically for the purpose of masking critical areas related to bright stars: saturated pixels, readout spikes due to saturated pixels, spikes caused by diffraction by the mirror supports and up to three reflection halos produced by the optics components (de Jong et al., 2015). In addition, for DR1.0 and DR2.0, defects not related to bright stars were manually masked. In DR3.0 the manual masks were not included (de Jong et al., 2017). For DR4.0 a semi-automatic procedure was then developed which includes the majority of these remaining issues with minimal manual intervention (Kuijken et al., 2019).

The masks are binary flags where each of the affected areas has a different value. It is therefore possible to exclude only certain types of areas, such as saturated pixels, while still using those in others, e.g. weak tertiary reflection halos. However, we have decided against such an approach and instead mask all pixels with a flag value greater than zero. This results in approximately 20% of pixels being masked out in *gri* and correspondingly, $\sim 20\%$ of galaxies in our sample are skipped since their centre is masked. This is a large fraction, but it is not a limiting factor to our analysis. Our emphasis is on obtaining the most directly comparable fits to a statistically large sample of galaxies, for which it is optimal to use this highest data quality and treat all bands consistently. The sample size could instead be increased by including more galaxies at higher redshift if desired.

If there is a particular interest in individual masked galaxies, it is still possible to re-run those ignoring some or all of the mask values. We have done this in the context of a Master's project focusing on AGN, of which there are only few in our sample (Targaczewski in prep.). However, this in turn required careful visual control of the resulting fits. For our own large scale automated analysis, it is hence preferable to use all flag values as masks. Importantly, the masks will randomly affect all types of galaxies and therefore not introduce any bias in our statistical analysis.

In summary, we opt to use the KiDS science tiles with associated weight maps and masks, including all binary flag values. For v01 and v02 of the BDDecomp DMU we used KiDS DR1.0, DR2.0 and DR3.0, which were incremental data releases and together cover all of the GAMA equatorial regions. From v03 onwards, we moved to KiDS DR4.0, which became newly available then. This did not add any tiles in our region of interest which was complete already in DR3.0. However, in contrast to the previous incremental data releases, DR4.0 was a complete re-release of all tiles with a number of processing changes. Most notably, the re-processing included a photometric homogenisation across all tiles (see Kuijken et al. 2019). DR4.0 therefore supersedes all previous KiDS data releases.

Pipeline setup

Given the list of galaxies in our sample and the list of KiDS tiles with associated weight maps and masks, the first decision to make is whether one should work on a “per-galaxy-basis”, i.e. taking cutouts of the corresponding tile(s) for each galaxy or a “per-tile-basis”, i.e. treating all galaxies within a tile at the same time. We opted to work on the galaxy level, mostly due to computational memory limits since the tiles are rather large (~ 1.5 GB each). Hence for each galaxy in our sample, we find the corresponding RA and Dec position from the GAMA database, identify the tile(s) that this position falls into and take a cutout around the galaxy position, with corresponding cutouts taken from the weight and mask images.

This approach also has the advantage that the galaxies, which are the more physically meaningful quantity, are taken as the basis instead of the tiles, which can vary in different datasets. This allowed a relatively easy expansion from KiDS to VIKING data; and - in that context - a switch from treating several data matches to the same galaxy individually to treating them jointly (see Section 2.3). It also facilitates multi-band fitting, which we will attempt in future work.

The relatively large cutout size of $400'' \times 400''$ was chosen to allow for a reasonable number of stars for PSF estimation (see Section 2.2.3); but is small enough to be handled computationally without problems. All preparatory work (image segmentation, background subtraction, PSF estimation) is carried out on these cutouts. We then further reduce the cutout size to only the region of interest around the galaxy itself before storage and subsequent galaxy fitting (see Section 2.1.1).

Note that we have decided to split the pipeline into three main steps: the preparatory work, the galaxy fitting and the post-processing (Section 2.1). After each of these steps, the results are stored to disk. This has the advantages that the galaxies can be fitted with different models or procedures using the same preparatory work; and that the galaxy fitting is better reproducible since all inputs are stored in the preparatory work directory. Also, the post-

2.2. PIPELINE DEVELOPMENT

processing is more flexible in the sense that e.g. model selection and outlier flagging can be re-calibrated without having to re-fit any galaxies. This is particularly important since the galaxy fitting takes the vast majority of computational time, with the preparatory work contributing around 2-3% of the total run-time (6-7% for VIKING) and the post-processing being negligible. Since each galaxy is treated independently of all other galaxies, the pipeline runs in parallel on many cores.

2.2.2 Background subtraction choices

After the trivial step of taking cutouts, the first main task of the preparatory work pipeline is to perform image segmentation to identify the pixels that belong to the object of interest and mask out neighbouring sources. This, however, is intimately linked to the background subtraction, since a reliable source identification needs a well-known background estimate and vice versa. Hence, the main function of `ProFound`, `profoundProFound`, iteratively performs image segmentation and background subtraction simultaneously by default. However, since KiDS image tiles are already background-subtracted, the question arose as to whether we should use this `ProFound`-estimated sky (e.g. to even out smaller-scale background inhomogeneities and to make the treatment of KiDS data more comparable to the treatment of VIKING data) or set it to zero a priori.

In addition, `ProFit` has the option to fit a constant background along with the source(s) of interest, which could also be used for local sky estimates instead of or in addition to the `ProFound`-estimated sky. We will refer to the former as (`ProFit`) background fitting and the latter as (`ProFound`) sky subtraction. Note that the two methods are not equivalent and describe slightly different “types” of sky: `ProFound` explicitly attempts to exclude faint undetected background sources and extended low-surface brightness wings of objects, while `ProFit` includes all undetected light in the background estimate (see details below).

Hence, the main choices to make with regards to background subtraction are whether or not to use the `ProFound`-estimated sky and/or fit a constant background with `ProFit` along with the source (in addition to the background subtraction already performed by the KiDS team); and if so, which algorithms and options to use for the background estimation. Interconnected choices are how deep to go in the object detection (most strongly influenced by the value of `skycut` in `profoundProFound`) and how large to make the segments for fitting objects.

The decisions made with regards to these choices are summarised in Table 2.4. In the following, we give more details on the metrics that we used to inform these choices. Note that these studies were only performed on KiDS r -band data. The final pipeline uses a joint treatment of the g , r and i bands as described in Section 2.1.1 and has been updated in many other respects, too. However, the choices made at this stage remained. We also note that

Table 2.4: Sky subtraction/background fitting choices

Choice to make	Decision made
Perform ProFound sky subtraction	Yes
Sky box size	1/4 of large cutout side length, i.e. 500 pix or 100''
Sky grid size	Equal to sky box size (default)
Interpolation type	Bilinear (default)
Clipping	Yes (default)
Type of sky estimate	Median (default)
Improve sky with FFT	Yes up to BDDcomp v02 , then no
Object masking (for FFT)	Aggressive
Perform ProFit background fitting	No
Size of cutout to fit	Star segment only
Skycut value (see Section 2.2.4)	1 for sky subtraction, 2 for object fitting

while we focus on the background subtraction in this section, we inevitably touch upon image segmentation, too. In particular, we discuss the choices made for object masking during background estimation, while the details of the segmentation maps used for the galaxy fitting are de-coupled (one of the choices made during the background studies) and explained in more detail in Section 2.2.4.

ProFound sky subtraction

As a first metric to judge the effects of different sky subtraction options, we use the distribution of **ProFit** backgrounds fitted to stars used during PSF estimation around a test sample of ~ 200 galaxies (one in each KiDS tile, approximately uniformly distributed across the galaxy magnitude range). The stars have the advantage that they show much less variation than galaxies and can usually be (near-)perfectly fitted with Moffat (1969) functions. They also do not suffer from PSF uncertainties since they do not need PSFs; they are more numerous than our galaxies and fitting them is much faster since simple downhill gradient algorithms suffice for such simple systems. In the following, we give a brief overview of the main results of these detailed investigations.

For Figure 2.4, we fitted a constant background in addition to the Moffat parameters to a few thousand stars around a random test sample of galaxies using **ProFit**. We used all pixels within a square cutout around the star for fitting, except those that are identified as belonging to other objects during the segmentation procedure. The cutout side length ranges from 6 to 30 times the FWHM of the PSF given in the header of the corresponding KiDS tile, which is typically around 0.6-0.8''. The top panel shows the mean, median and standard deviation of the distribution of fitted backgrounds (in arbitrary units of flux since the cutouts were normalised before fitting) for different cutout sizes and for the two cases of also performing a **ProFound** sky subtraction before fitting (blue symbols) or not (red symbols). For the largest

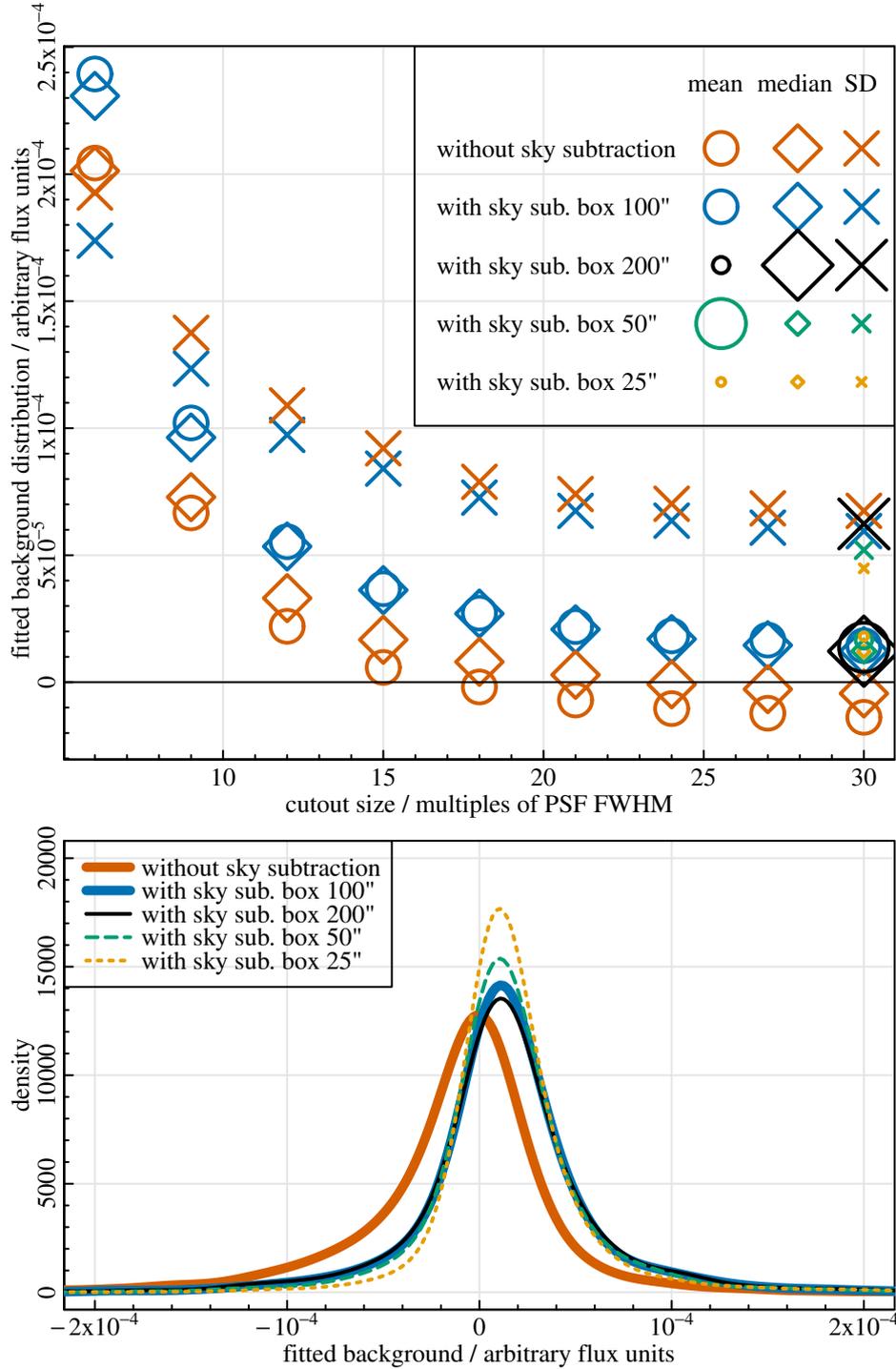


Figure 2.4: Top panel: the mean (circles), median (diamonds) and standard deviation (crosses) of the distribution of fitted ProFit backgrounds as a function of the fitting cutout size and with (blue) or without (red) previous ProFound sky subtraction. For the largest cutout size, we additionally show the effect of the chosen box size for sky subtraction (black, green, yellow). **Bottom panel:** the full distribution of backgrounds for the largest cutout size (corresponding to the rightmost set of symbols in the top panel).

cutout size, we also show results for different sky box sizes, which is set to the fiducial value of $100''$ for all other runs: black symbols denote an increase in the box size by a factor of 2, while green and yellow symbols represent a decrease by a factor of 2 and 4 respectively. The bottom panel shows the full distribution of fitted backgrounds for the largest cutout size, corresponding to the rightmost set of symbols in the top panel.

For small cutout sizes, the distributions are clearly biased positive (top panel of Figure 2.4), i.e. the background fit is dominated by the wings of the star. For larger cutout sizes, the distributions then converge onto a slightly positive or negative value (with and without **ProFound** sky subtraction respectively). In general, the distributions with **ProFound** sky subtraction are slightly narrower and more symmetric than those without (i.e. the standard deviation is smaller and the mean and median are closer together). Due to these reasons (width, asymmetry and convergence onto negative values for non-**ProFound**-sky-subtracted images), we concluded that the **ProFound** sky subtraction is useful to even out smaller-scale inhomogeneities even though KiDS images are already background subtracted. In addition, it will make the comparison to VIKING data (which have potentially different backgrounds to KiDS) more consistent.

The fact that the mean and median of the fitted distributions with previous sky subtraction converge onto a positive value instead of zero is expected due to the different “types” of sky that are estimated by the two options: **ProFit** will estimate a slightly higher sky than **ProFound** because the former simply uses all non-object pixels while the latter explicitly tries to avoid being biased by undetected (wings of) objects by clipping outliers and using the part of the flux distribution smaller than the mode only to estimate the median and standard deviation of the background in each box.

The box size should be chosen as small as possible to be able to detect small-scale inhomogeneities in the background, yet larger than the largest (complexes of) objects so as to not subtract their wings. For our purposes, we found a box size of $100''$ to be suitable. Increasing the box size by a factor of two only increases the standard deviation of the distribution, but does not change the mean or median. Decreasing the box size by a factor of two decreases the standard deviation but also shifts the mean away from the median, making the distribution more skewed. Decreasing the box size by a factor of four (i.e. to $25''$) makes the distribution significantly more asymmetric and - upon visual inspection of a random sample of fitted backgrounds - also reveals clear correlations between the fitted background and the location of objects in the images. Hence the box size of $100''$ seems to be optimal in that it is independent of the object distribution yet still detects some small-scale background variations.

We also varied other parameters to test their influence on the sky estimate (grid size on which to place the boxes, bilinear or bicubic interpolation between boxes, clipping outliers or not, using the mean or median for the sky estimate, dilating the object masks more or less) but

2.2. PIPELINE DEVELOPMENT

have found either little influence or adverse effects on the sky estimate. Hence we decided to use the `profoundProFound`-defaults for these parameters, which were optimised for typical optical and NIR survey data (such as KiDS) during the development of `ProFound` (Robotham et al., 2018).

Finally, we should mention that we have also carried out tests where we divide each of the `ProFit` background fits by its respective error and investigate the resulting distributions. In theory, this should result in a Gaussian with mean zero and standard deviation one. A small positive shift away from the mean is to be expected according to the previous analysis, although much smaller than one standard deviation (Figure 2.4). A standard deviation larger than one would indicate that the background offset (i.e. the positive shift) is not constant for different cutouts. This analysis, however, relies on realistically estimating the uncertainty of the fitted background value. We have tried this in two different ways: either during the downhill gradient fitting with `optim` (see Section 2.1.1) or directly from the KiDS weight map. Both methods have unresolved problems, so we did not use these results further and only add a short description here for completeness.

In the first case, we use `optim`'s option to return the Hessian along with the best fit value of all parameters. The diagonal elements of the inverse of the Hessian matrix are the variances of the parameters and hence the entry corresponding to the background fit can be used to estimate its uncertainty (ignoring covariances and systematic uncertainties). For unknown reasons, however, the variance of the background was often negative, resulting in an imaginary standard deviation. Taking the absolute value before taking the square root gives errors of the correct order of magnitude, such that the normalised background distributions then have a standard deviation of approximately one and a slight positive offset.

For the second approach, we use the fact that each background pixel is drawn from a normal distribution with a standard deviation given by $\sigma = 1 / \sqrt{weight}$. `ProFit` essentially finds the mean of this background in each cutout weighted according to the errors (σ). The error on the mean (uncertainty coming from randomness that cannot be avoided) is $\sigma_{bg} = \bar{\sigma} / \sqrt{N}$ where $\bar{\sigma}$ is the mean of σ and N the number of pixels used to fit the background. This ignores any possible correlations between pixels or fitting parameters and assumes that the weight map is an accurate representation of the true errors on pixels. The problem with this estimate is that it results in background uncertainties that are approximately a factor of five smaller than those estimated from the Hessian; and the background fit distributions are correspondingly broader.

It is unclear why the two methods provide such different results for the background uncertainty. There are correlations between pixels on small scales (see below) that will additionally contribute to the uncertainty estimated from the weight map, but this is unlikely to explain a factor of five. Correlations between parameters are ignored in both cases and hence cannot

explain the difference. The weight maps tend to be on the conservative side compared to the true pixel errors (see Section 2.2.1), but this also does not explain a factor of five and moreover would tend to make the difference even larger if corrected for. Given these open issues, we decided to focus on the distribution of fitted backgrounds in absolute values as presented above; and did not consider the normalised distributions further.

ProFit background fitting

Since the above analysis relies on the distribution of fitted **ProFit** backgrounds as a metric, it does not allow to decide whether or not the fitted background should be used in addition to the **ProFound** estimated sky. Therefore, as a further test, Figure 2.5 shows the systematic differences that occur in the PSF estimates as a function of the different background fitting and cutout size options. This now uses the model PSFs estimated for the 223 test galaxies (one in each KiDS tile within our region) rather than the fits to individual stars. For any two runs, the differences between the 223 pairs of model PSFs are calculated and then added together (stacked) to show systematic trends. Ideally, PSF estimates should be robust against changes in the processing details (e.g. the cutout sizes used for fitting them), such that the differences should be small.

The first row shows the difference between fitting a **ProFit** background or not for four different combinations of cutout size and **ProFound** sky subtraction as labelled in the panels. The larger cutouts (first and third panel) show less of a difference than the smaller cutouts. Note, for comparison to Figure 2.4, that “small” cutouts always mean 6 times the PSF FWHM given in the header, while “large” cutouts refer to 18 times the PSF FWHM (we unfortunately did not perform these test runs with larger cutout sizes than this due to the order in which the decisions were originally taken). For both cutout sizes, there is also less of a difference in the fitted PSFs without **ProFound** sky subtraction (third and fourth panel) than with. This means that performing a **ProFit** background fit on the already-**ProFound**-sky-subtracted image systematically changes the PSF estimate - likely because there is no background left to fit and so the additional degrees of freedom are used to change (improve) the PSF fit instead. Note all residuals here are positive at the centre, indicating that fitting a **ProFit** background generally makes the PSFs more peaked at the centre.

The second row shows the difference between subtracting a **ProFound** sky or not for four different combinations of cutout size and **ProFit** background fitting. The second and fourth panels which show differences for small cutouts show negative residuals while the first and third panels (large cutouts) show positive residuals. In general, all residuals are small, which confirms that the PSF estimate is generally robust against performing a **ProFound** sky subtraction or not (i.e. the sky subtraction does not subtract the wings of the PSF or similar). The best results (i.e. smallest residuals) are obtained using the large cutouts and no **ProFit**

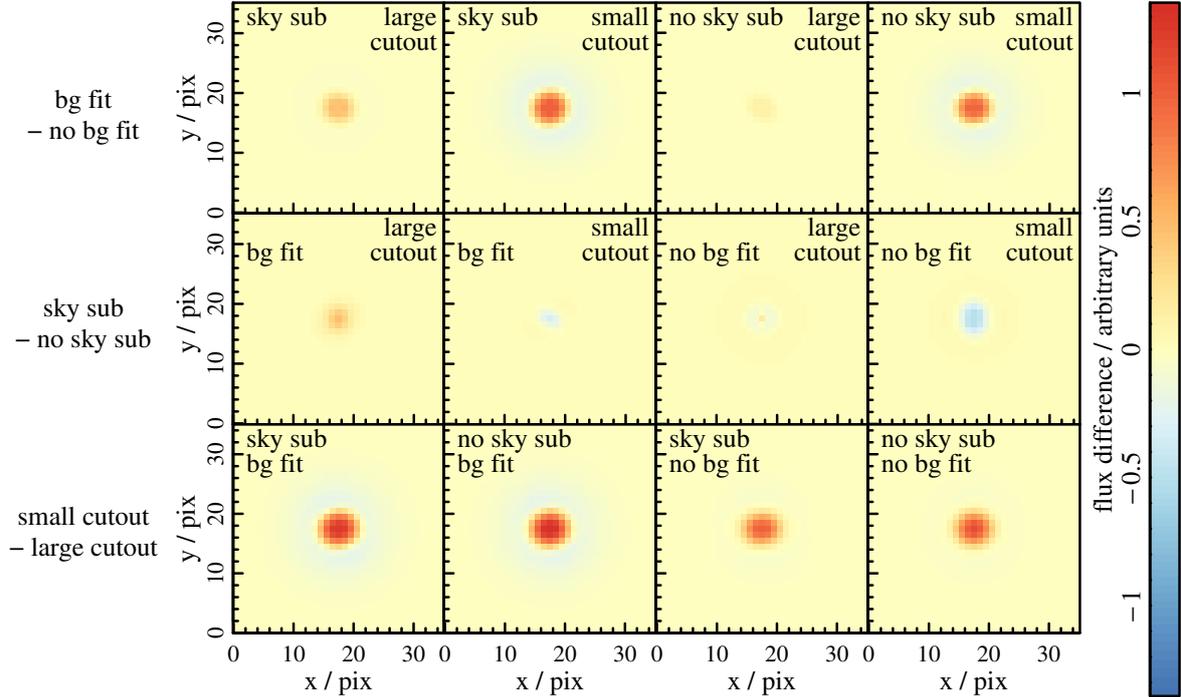


Figure 2.5: Stacked differences between model PSFs obtained from test runs with different combinations of ProFound sky subtraction (sky sub), ProFit background fitting (bg fit) and cutout sizes as labelled. **Top row:** the difference between model PSFs obtained by fitting or not fitting a ProFit background for four different combinations of ProFound sky subtraction and cutout size as labelled in the panels. “Small” cutouts refer to a side length of 6 times the PSF FWHM, “large” means 18 times the FWHM. **Middle row:** the difference between model PSFs obtained with ProFound sky subtraction vs. those obtained without, for four different combinations of ProFit background fitting and cutout sizes, as indicated. **Bottom row:** the difference in model PSFs obtained from fitting large or small cutouts for four different combinations of ProFound sky subtraction and ProFit background fitting, as labelled.

background fitting (third panel). When using smaller cutouts there seem to be less residuals with ProFit background fitting (second panel) than without (fourth panel). This can be interpreted as follows: for small cutouts, the star dominates the cutout region, so the ProFound sky subtraction does not influence the fit of the star. This is even clearer when we additionally fit a background within the small cutout region. For larger cutouts, the PSF estimate is also robust against performing a ProFound sky subtraction or not, provided no additional ProFit background is fitted. An additional background fit makes the PSF estimate dependent on the sky subtraction (first panel), most likely because now the fit has too many degrees of freedom.

The third row of Figure 2.5 shows the difference between a large and a small cutout for four different combinations of ProFound sky subtraction and ProFit background fitting. As observed before, the most robust results are obtained with ProFound sky subtraction only (third panel). Both results using ProFit background fitting (first two panels) are more

dependent on the cutout size used.

As a result of these studies, we saw our decision to use the `ProFound` sky subtraction reinforced and decided to not use additional `ProFit` background fitting: after `ProFound` sky subtraction, the background is homogeneous enough that an additional `ProFit` background fitting changes the PSF rather than fitting the actual background. Also, as evident from Figure 2.4, we need very large cutouts for robust `ProFit` background fits. This has problems on its own as it will necessarily lead to more contamination from the wings of other objects or undetected faint objects (masking is never perfect), increases the chance of hitting an image edge or masked area, and increases computational time. In addition, there is a degeneracy between the fitted background and the concentration index of the Moffat parameter in that it can never be unambiguously determined whether the background is used to fit the wings of the star or the wings of a low-concentration Moffat function are used to fit the background (or wings of other objects).⁴ For very large cutout sizes, the star may even become insignificant relative to the background so that the fitting does not focus on the main object of interest anymore (and instead the Moffat function is used primarily to even out background inhomogeneities). Fitting galaxies introduces a wealth of additional problems if their outskirts do not follow Sérsic functions anymore, e.g. due to disk breaks, flares, rings and similar (cf. Section 2.1.3).

In summary, we opted to use the `ProFound` estimated sky with default parameters except for the box size of $100''$; but not to perform an additional `ProFit` background fit. This in turn allows to use tight segmentation maps for fitting. These choices were made before the first DMU release and hence did not change between v01 and v04.

Pixel correlations

To detect problems in the sky subtraction and object masking, an analysis of correlations between pixels can be useful. `ProFound` offers a set of functions to achieve this via a Fast Fourier Transform (FFT) of the normalised image (each pixel divided by its error) after sky subtraction and object masking. This FFT image gives information on the pixel correlations on different spatial scales. In the ideal case, this should result in pure uncorrelated white Gaussian noise. The idea then is that power on small scales represents undetected sources, while on larger scales it captures complex features of the sky background.

For KiDS images, we expect some residual correlation on small spatial scales. This is mainly because before stacking, all KiDS frames have been re-gridded onto the same pixel scale of $0''.2$ from their native scales of $\sim 0''.21$. This ensures that the tiles are pixel-matched across all four bands with the axes aligned in RA and Dec (see Sections 1.2.2, 2.1.1 and Kuijken et al. 2019). However, it will also introduce correlations on very small scales. In addition, there

⁴Similar arguments apply to the Sérsic index of a galaxy.

2.2. PIPELINE DEVELOPMENT

can be intrinsic instrumental correlations, typically on the scale of a few pixels. Investigating these in detail goes beyond the scope of this work. The PSF, although it is several pixels across, is not expected to play a significant role for seeing-limited data such as KiDS since most noise is “added” at a later stage. For perfect sky subtraction and object masking, we hence expect correlations to reach zero beyond the scale of a few pixels. In practice, we might expect a small positive excess resulting from very faint, individually undetectable objects.

Figure 2.6 shows the main diagnostic outputs produced by the `profoundPixelCorrelation` function for six different input images (A-F). In the left panel of each set of plots, we show a so-called correlogram: pixel correlation as a function of x (RA) and y (Dec) pixel lag (solid blue and red lines respectively). In addition, the dashed lines show the difference between the correlations of positive pixels and those of negative pixels (also in x and y); while the dotted lines indicate the pixel lag implied relative standard deviation. All correlograms are shown once with a y -axis range of -1 to 1 (top of the left panels) and once as a zoom into the range -0.1 to 0.1 (bottom). The x -axis is the same for all correlograms, ranging from a lag of 1 pixel to 2000 (the image size) on a logarithmic scale. Generally, a positive excess in this plot indicates that object detection or sky subtraction are not aggressive enough, while systematically negative correlations suggest that the image contains large pools of negative pixels, hinting towards problems in the sky subtraction. To the right of each set of correlograms, we show the corresponding two-dimensional FFT image, where white colour means a relative excess power (phase ignored) and small scales are centrally located. Note the linear axes scales compared to the logarithmic x -axis of the correlogram; while conversely the colour scale now uses an asinh stretching instead of the linear y -axis of the correlogram. While the quantitative interpretation of these images is less straight-forward, they can be used to detect correlations at different scales and in different directions at least qualitatively. In both cases, masked pixels (including those identified as belonging to objects) are excluded from the analysis by `ProFound`.

Starting from the top left, the first set of panels A), show the result obtained for a generated image of pure Gaussian noise with a mean of zero and standard deviation of one. As expected, the pixel correlations are zero on all scales, with the standard deviation being one. The exception is the last data point (at 2000 pixels), where the lines diverge due to unreliable low number statistics once we reach the image size. We hence only consider values up to 1000 pix on the x -axis for all correlograms. Correspondingly, the FFT image contains no structure (showing that the random number generator is working well). This represents the ideal case which we would like to achieve for our real images (after sky subtraction and object masking), but are likely to never fully reach given the limitations explained above.

Panels B) show the first step of complications that is known to exist in real (KiDS) images: pixel correlations due to the resampling of the raw images before stacking. For this analysis,

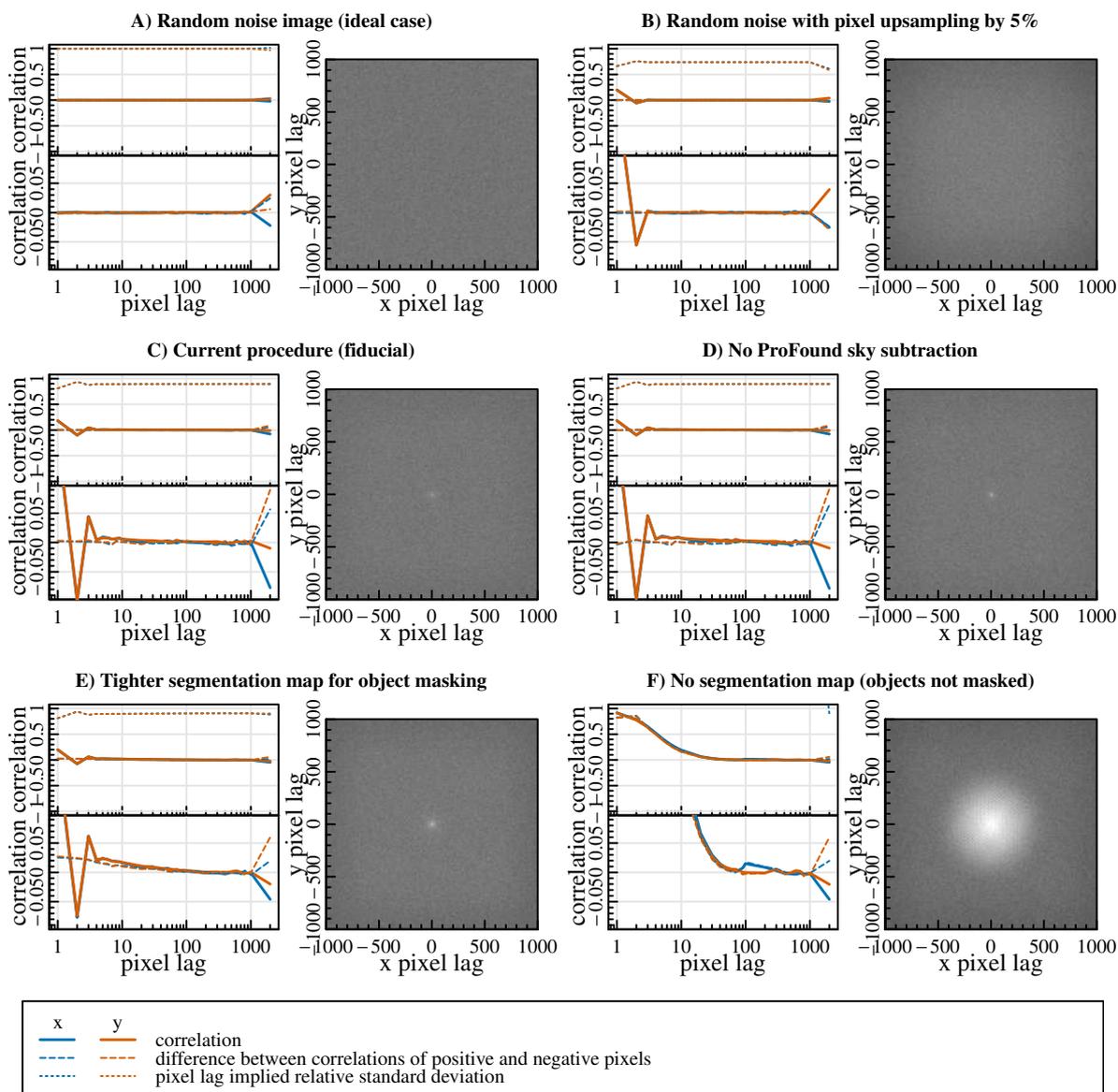


Figure 2.6: Correlations between pixels for **A)** a random noise image, **B)** a random noise image with pixel upsampling by 5 %, **C)** a real image around the example galaxy 396740 (the same as in Figure 2.1 for comparison) treated as in the final pipeline, **D)** the same real image as C), but leaving out the ProFound sky subtraction during treatment, **E)** the same real image as C) but using tighter segmentation maps for object masking (coloured instead of grey contours in Figure 2.1) and **F)** the same real image as C) without any object masking. The left panels always show the correlogram, i.e. pixel correlation versus pixel lag in x and y , with the meaning of the different lines indicated by the legend at the bottom of the plot. The bottom panel is a zoom (on the y -axis) into the top panel. The images to the right of the correlograms show the corresponding FFT image, scaled such that excess power is white (phase ignored, asinh scaling) and small scales are at the centre of the image.

2.2. PIPELINE DEVELOPMENT

we have created a random noise image of slightly smaller dimension ($1905 \times 1905 \text{ pix}^2$), which we have then upsampled onto the usual $2001 \times 2001 \text{ pix}^2$ grid and subsequently passed through the FFT analysis. This roughly corresponds to the upsampling performed by the KiDS team from a native instrumental pixel scale of $\sim 0''.21$ to $0''.2$ in the tiles. This introduces correlations on small scales (up to $\sim 3 \text{ pix}$), which are clearly visible in the correlograms. Since positive and negative pixels are equally affected by the upsampling, the difference between positive and negative correlations (dashed lines) stay zero. Note, however, that since we have not corrected the weight map (which is uniformly one for a random image of standard deviation one), the pixel lag implied standard deviation is now below one (i.e. the true standard deviation of the upsampled pixels is smaller than one).

Panels C) show the results obtained for one of our real images after sky subtraction and object masking with our final pipeline. We choose the same example galaxy as that shown in Figure 2.1, to allow a direct comparison with the segmentation maps. The most striking feature here is the characteristic pattern of correlations at very small scales, which - judging from panels B) - we mostly attribute to the upsampling (and possibly intrinsic instrumental effects on small scales, which we did not investigate). These features consistently persist for all images (although we only show one example here). Beyond a scale of $\sim 3 \text{ pix}$, the correlation drops to zero rapidly, although a small excess can be seen at scales up to 10 pix , maybe even up to 100 pix . These correlations are also visible as a faint white dot at the centre of the FFT image. However, there seem to be no asymmetries in our analysis (no structure in the FFT, x and y correlations always the same, negative and positive pixels affected equally). Note that the standard deviation is also slightly below one here, implying that the true standard deviation of background pixels (after sky subtraction and object masking) is slightly smaller than that implied by the weight maps provided by KiDS. We discuss this further in Section 2.2.1.

Panels D) show the same as panels C), only that in this case, we do not perform the **ProFound** sky subtraction. The plots are nearly indistinguishable from those in panels C), demonstrating that pixel correlations are neither introduced nor removed by our sky subtraction procedure and instead correlations are primarily caused by remaining undetected objects. Since the box size of the sky subtraction is $100''$ or 500 pix (Section 2.1.1), we only expect effects on large scales here.

We test the effect of undetected (wings of) objects in panels E), where we use a tighter segmentation map for object masking. Instead of the segmentation map generated for the background estimation, we use that produced for fitting the galaxies (cf. Section 2.1.1 and Figure 2.1: here, we use the coloured contours instead of the grey ones for object masking). This goes less deep in both object detection and segment dilation. In other words, we mask less of the faintest objects; and the regions masked around each (bright or faint) object are

smaller. This results in a clear positive excess at scales up to ~ 100 pix. The dashed lines also deviate from zero at these scales, indicating that positive pixels are more correlated on average than negative pixels (since the undetected sources of light are all positive). The FFT shows a corresponding white dot at its centre, but no additional two-dimensional structure.

Finally, panels F) show the pathological case of ignoring the segmentation map altogether, i.e. no objects are masked (except for those that are saturated and hence included in the KiDS masks already at an earlier stage). The correlation of pixels reaches values near unity for small scales and is substantial up to scales of 100 pix or even beyond. The dashed lines closely follow the solid lines, indicating that this is entirely caused by correlations between positive pixels (i.e. there are no negative flux objects in the image). The asymmetry between x and y correlations at scales of 100 – 200 pix is unique to this particular image and not generally visible for other example galaxies. The corresponding FFT image also shows a large power excess at scales up to several hundred pixels.

In conclusion, our procedure for background subtraction and object masking seems to remove the vast majority of correlations between pixels, with only the characteristic features of up-sampled pixels remaining at very small scales; and a slight excess caused by faint, individually undetectable objects at slightly larger scales. It is, however, important to use the very aggressive, deep and highly dilated object mask that we generate for accurate background subtraction; rather than the tighter fitting segmentation map (for details on the latter, see Section 2.2.4). For this reason, we eventually decided to use two separate segmentation maps for these two processes, to better accommodate the different aims that we try to achieve with them (cf. Section 2.1.1). This change came into effect with v03 of the `BDDcomp` DMU.

In addition to the above, it should be mentioned that `ProFound` offers a function to improve an already-estimated sky on the basis of an FFT analysis (`profoundSkySplitFFT`). The idea is that, on large scales, excess power captures complex features of the sky background that cannot be described by the bilinear or bicubic interpolation between boxes that is performed by `profoundProFound`. Hence, this part of the FFT image can be used to refine the sky estimate; the “new” (refined) sky is returned by the function. In early versions of the pipeline (up to the release of `BDDcomp v02`), we have used this refined sky. However, starting from v03 we reverted to using the “original” sky estimated by `profoundProFound` since it turned out to be more robust especially in crowded fields and/or in the presence of nearby masked objects, image edges and artifacts. On the other hand, for “well-behaved” fields there is very little structure to be removed on large scales (see Figure 2.6). Hence for most images, this additional step of sky refinement had very little effect, while for some fields it had adverse effects (also confirmed by simulations). For our large automated analysis with varying image properties and intrinsically relatively flat sky, we therefore found `profoundSkySplitFFT` to be unsuitable.

2.2.3 PSF estimation details

The last major step in the preparatory work pipeline is the PSF estimation. An accurate and precise estimate of the PSF is crucial to the success of the galaxy fitting. Any deviations of the model PSF from the truth will introduce systematic uncertainties in the galaxy parameters that are not easily accounted for. Unfortunately, however, the PSF estimation is not straightforward as there are many uncertainties involved and choices to make: how to identify and select stars suitable for PSF estimation, which function and algorithm to fit them with (or whether to fit them in the first place), how many stars to use and how to combine the different estimates into a final model PSF for each galaxy of interest.

In view of these difficulties, we first contacted the KiDS team to ask whether they would be able to provide PSFs even though they are not publicly released. The team were willing to provide PSFs in the form of shapelet functions for each tile. However, these are not the PSFs that the team use for their own weak lensing analyses, since those are performed on individual pawprints rather than stacked tiles. They did therefore not undergo a detailed quality control. A brief comparison with a selection of stars in a number of test tiles showed that the shapelets did not fully capture the extended wings of some PSFs. In addition, we already anticipated the addition of VIKING data, for which there are no such shapelets available. We therefore decided that it would be best to devise our own method of PSF estimation to obtain consistent estimates in all bands and have full control over their quality.

Since PSF estimation using `ProFound` and `ProFit` has not previously been performed, there were no established procedures for us to follow. Consequently, our treatment was developed as a mixture between theoretical considerations, procedures that have proven to work well with other software and a considerable fraction of trial and error combined with visual inspection of the results. A very brief overview of the resulting process is given in Section 2.1.1. Here, we describe the treatment in much more detail, explaining some of the reasons for our choices and other possibilities that we explored. We finish with an overview of the PSF quality control and its effects on the galaxy fitting.

Many of the details changed during the pipeline evolution. For example, the star candidate selection process depends on certain cuts that need to be adjusted when making significant changes to the way segments are defined (i.e. between `v02` and `v03` of the `BDDcomp` DMU). Since these cuts were mostly tuned by visual inspection and also changed frequently between test runs, we do not give detailed justifications for each pipeline version and instead focus on the status quo in `v04` (which is also what Casura et al. submitted is based on). We note that, while the detailed numbers may have changed several times, the overall procedure remained the same.

Star candidate selection

The first decision to make is whether to estimate PSFs separately for each individual object of interest, or fit all stars in a large region (e.g. an entire KiDS tile) at once and then interpolate the solution for each position of interest. Both approaches have their advantages and disadvantages. We took the first approach, whereby for each position of interest we independently estimate a PSF. This has the benefits that local PSF variations are captured without worrying about interpolation uncertainties. It also allows to flexibly change or improve individual PSFs without having to re-fit all stars in the field; and it allows to use the independent PSF estimates of neighbouring galaxies for quality control (see below). In addition, it is straightforward to handle computationally since it can just be done along with the local background subtraction and other preparatory work that is run for each galaxy.

Disadvantages of this approach are that each individual PSF estimate is based on a relatively low number of stars (to keep it local), potentially introducing more noise; and that there will be a number of galaxies with missing PSFs if no suitable nearby stars are available (for example because large areas in the vicinity of the galaxy are masked). However, the latter only affects less than 1% of galaxies (see Table 3.1 in Section 3.2) the majority of which tend to be in regions of decreased data quality (near bright, saturated stars, image edges or artifacts). To decrease the noise on our PSF estimates, we introduce a fitting step with `ProFit` instead of - as is sometimes done in other software - just averaging a large number of star cutouts. We find the fitted Moffat parameters for stars near each other to be very consistent (see below).

In summary, we perform PSF estimation on the single-band $400'' \times 400''$ cutouts with associated sigma (error) maps and masks, after background subtraction and image segmentation. For v04 of the pipeline, we use the segments from the stacked *gri* images, but re-calculate the segment statistics for each individual band. For the test runs which this section is based on, we only used *r*-band images.

From the segmentation statistics returned by `profoundProFound`, we select relatively round and isolated objects as follows:

- objects that do not touch other segments, masked regions or image edges (edge fraction = 1)
- objects with a regular boundary geometry (edge excess < 1)
- objects with an axial ratio (minor/major axis) larger than 0.5
- objects which were not flagged as possibly spurious

These are trivial selections based on theoretical considerations to remove the segments with least reliable statistics; with the exception of the axial ratio. For the latter, we originally used

2.2. PIPELINE DEVELOPMENT

an axial ratio cut based on the PSF ellipticity given in the header of KiDS science tiles. The header value refers to the average PSF ellipticity across the entire tile, which is typically very low (corresponding to axial ratios $\gtrsim 0.9$). However, we found that the PSF can vary strongly locally across the tile and is often elongated significantly especially towards the tile corners (see Figure 2.14), with axial ratios as low as ~ 0.6 . Due to these local variations, a cut based on the average PSF ellipticity proved to be unsuitable and we instead decided to simply use a fixed cut of 0.5 on the axial ratio to remove very elongated objects from further analysis.

Of these relatively round and isolated objects, a given fraction (depending on the depth of the image and source extraction, typically 4-8 %) are identified as star candidates via a joint cut in R50 (semi-major axis containing half the flux) and the concentration (R50/R90, where R90 is the semi-major axis containing 90 % of the total flux). A diagnostic plot of this step is returned with an example shown in Figure 2.7, again for galaxy 396740. We chose this cut based on the notion that we would expect stars to be small (i.e. low R50) and highly concentrated (i.e. low R50/R90). The plot of these two quantities coloured by axial ratio (example in Figure 2.7) typically shows a rather distinct population of objects that are constant in (small) R50 and (high) axial ratio, with low R50/R90. To capture most of this population of objects without including too many others, we then empirically devised the joint cut in R50 and R50/R90 as indicated in Figure 2.7.

We found the best results using a fixed percentage of objects to accommodate both crowded fields and those with only few objects and/or large masked areas. The chosen percentage depends on the depth of the source extraction (i.e. the `skycut` value in `profoundProFound`) and the depth of the image. For the stacked *gri* KiDS images and a `skycut` value of 2, we found that classifying 4 % of all segments as star candidates delivers reasonable results for the vast majority of cutouts. While this does not always capture all stars in the frame, it is usually enough to allow for a robust PSF estimate. Conversely, for some frames it may include a few objects that are not actually stars, but as long as they are not the majority, they will be screened out at a later stage in the PSF estimation when we reject outliers.

Note that since the number of segments and the fraction of how many of those are stars are determined by the segmentation map alone, this value can be used for images of varying depth as long as they use the same segmentation map. In other words, even though for example the *i*-band images are shallower than the *gri* stacks; and the segment properties (including R50 and R90) are re-calculated in the *i*-band, the 4 % selection cut is appropriate because we originally defined the segments on the *gri* stacked image. Hence from v03 of the pipeline onwards, when we started using the *gri* stacks to define common segments for all bands and fixed `skycut` to two, we always use the 4 % selection cut ($0.02 \times \text{skycut}$). Earlier (single-band) versions of the pipeline use $0.03 \times \text{skycut}$ for *g* and *r*, $0.04 \times \text{skycut}$ for *i* and $0.06 \times \text{skycut}$ for *u*.

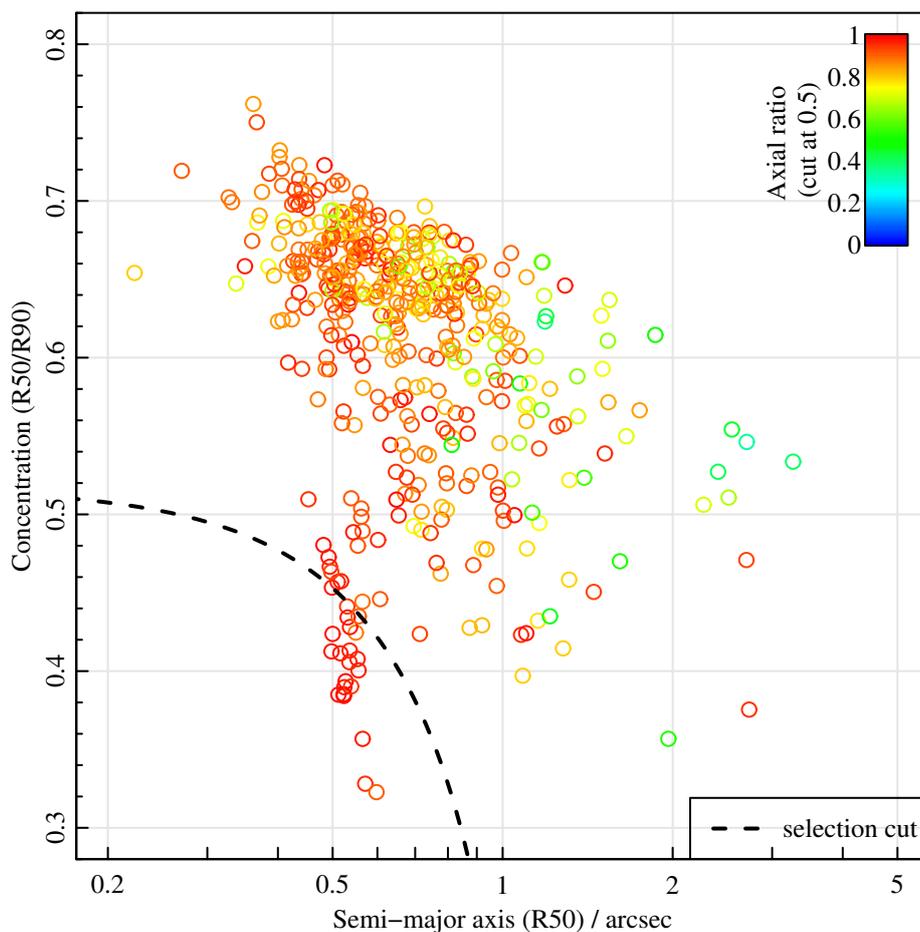


Figure 2.7: A diagnostic plot of the selection of star candidates for example galaxy 396740. Coloured points show the concentration $R50/R90$ against $R50$ (semi-major axis containing half of the total flux) for all segments in the $400'' \times 400''$ cutout around the galaxy; coloured by axial ratio. The joint selection cut in $R50$ and $R50/R90$ is indicated with a dashed line.

Around each of these star candidates, a smaller cutout is taken (side length equal to $2 \times R100$; $6 \times R100$ if a background is to be fit as well) and a subsample selected:

- objects brighter than the 5-sigma point source detection limit and fainter than the saturation limit (both taken from the headers of the corresponding KiDS tile)
- objects where less than 10% of pixels in the cutout belong to other segments (rising to 30% for the larger cutouts when fitting a background)
- objects where the star cutout does not overlap with the edge of the large cutout
- objects with a positive sum of the cutout (excluding poorly background-subtracted and/or purely noise-dominated objects)

The cutout size is large enough to include the entire star segment, which are the pixels that

2.2. PIPELINE DEVELOPMENT

we use for fitting the star. When additionally fitting a background, we need larger cutouts to include a sufficient number of sky pixels (see Section 2.2.2). This only affects a number of test runs related to Section 2.2.2; none of the full runs used additional background fitting. The remaining selection criteria then ensure that the stars are neither saturated nor too faint to produce a reliable fit; and that they are isolated and fully contained within the image. We also experimented with constraining the magnitude range further (e.g. fitting only stars that are at least 2 mag brighter than the 5-sigma point source detection limit), but found that this does not significantly improve the fitting quality and instead can limit the number of stars too much for fields in which only few star candidates are available.

The order in which we apply these selection cuts was essentially defined by trial and error: the cut in size and concentration (Figure 2.7) is best applied after a pre-selection of isolated relatively round objects, since otherwise the number of stars in crowded fields is reduced too much. This is because in these fields, the fraction of stars is usually higher, which we can at least partially account for by the pre-selection of round objects. However, the magnitude cuts should be applied after the concentration and size cuts since otherwise the assumed fraction of stars in the frame (4%) needs to be carefully tuned to the chosen magnitude cuts. The reason here is that stars tend to belong to the brightest objects in the frames, so for example cutting 2 mag off the faint end will increase the fraction of stars substantially.

If no stars pass the above criteria (usually this only happens when the vast majority of the large cutout is masked), then a default perfectly round PSF is created with the FWHM taken from the KiDS header and a Moffat concentration index of 2.1 (the median of all measured *r*-band concentrations). If this is the case, a warning is given and an appropriate flag value added to the PSF quality flag, such that galaxies for which only this default (dummy) PSF exists are skipped during the fitting procedure. If more than 32 star candidates pass the above criteria, we limit the fitting to the 32 objects closest to the galaxy in order to save computational time.

Ideally, we would use only those stars that are on the same detector chip as the galaxy of interest; or at least stars that have the same number of dithers as the galaxy. However, the KiDS instrumental setup and dithering pattern result in frequent changes of the number of dithers across a tile (cf. the darker stripes in the weight map of Figure 2.1). Trying to identify the corresponding stars is hence non-trivial and usually results in only very few star candidates being left. Moreover, different parts of the galaxy itself are often covered by different numbers of dithers, which would - to be consistent - then require the galaxy to be fit with different PSFs for different regions. These difficulties could be avoided by working at the pawprint level (like we do for VIKING, Section 2.3) instead of using the stacked science tiles. For reasons outlined in Section 2.2.1, we however decided to use the science tiles for KiDS. Since all dithers of a tile were taken in close temporal succession, they should have very

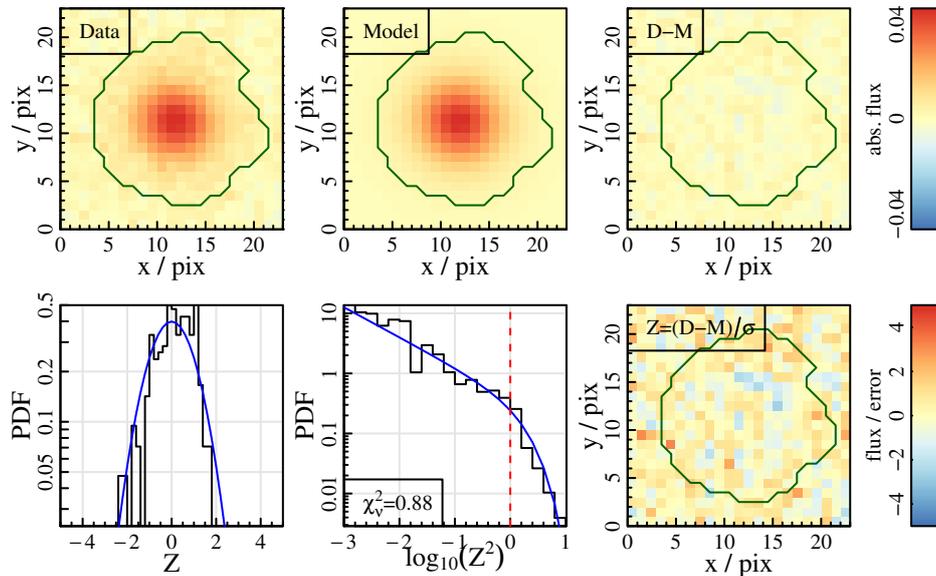


Figure 2.8: The Moffat function fit to an example star near galaxy 396740. Panels are the same as the top two rows in Figure 2.2.

similar PSFs in general. Therefore the stacking and the changing number of dithers across the field will not have a large effect on the PSF. Indeed we do find the PSF to vary only slowly across tiles, with no abrupt changes (cf. Figure 2.14).

Star fitting

In the next step, the star candidate cutouts are normalised to a magnitude of 0, masked appropriately and fitted with a Moffat function using `ProFit` (see fitting details in Section 2.1.1). We choose a Normal likelihood function since we need near-perfect fits for adequate PSF representation and so the residuals are expected to be distributed Normally. The normalisation of the PSF to a magnitude of 0 brings the sum of the cutout close to 1 (for the KiDS magnitude zeropoint of 0) and so individual pixels mostly have values around a few $\times 10^{-4}$. We found this to be close to the optimum in terms of fit accuracy when using the default tolerance levels for convergence in `optim` (it is also where the fitting time is longest). If no background is fitted, only the segment around the star is used for fitting; with a background fit the entire cutout is used but with all objects (except the segment of interest) masked. The diagnostic plot of the fit returned by `ProFit` is saved for each star and an example shown in Figure 2.8.

As briefly mentioned above, we introduce this fitting step (instead of a purely empirical estimate based on, e.g., simply averaging appropriately scaled cutouts around suitable stars) since it reduces the noise in our PSF estimate. This is particularly important given that we

2.2. PIPELINE DEVELOPMENT

only use a relatively small number of stars to obtain the model PSF. Additionally, it gives an opportunity to easily remove outliers, so the star candidate selection process need not be perfect. It does, however, require finding a modelling function that can adequately represent the PSF, since otherwise one will introduce systematic biases. We chose to fit a Moffat (1969) function since that is known to parameterise telescope PSFs well and, in fact, was devised for that exact purpose.

As one can see, a Moffat function is a perfect fit for the star shown in Figure 2.8. The normalised residual Z (bottom right panel) shows pure noise and its distribution follows a Gaussian function (bottom left panel), while Z^2 follows a chi-squared function as it should (bottom middle panel). The fact that the reduced chi-squared (χ^2_ν given in the bottom middle panel) is below the ideal value of one is because the KiDS weight maps tend to slightly overestimate the true pixel error, see Section 2.2.1.

Most isolated stars are represented very well by Moffat functions (like the example shown), with two exceptions. First, very bright stars sometimes show residuals at their cores, which can be systematic in the sense that all bright stars in a field show similar residuals. This indicates that a Moffat function is not able to capture the true shape of the PSF at the cores of bright stars. The residual pattern usually shows a red core surrounded by a blue ring or arc, often not entirely symmetric, see the example in the top panels of Figure 2.9. Nonetheless, the parameters fitted to these bright stars are indistinguishable from the parameters fitted to fainter stars of the same field. In addition, these bright stars reach surface brightnesses close to the saturation limit, where CCD detectors are known to have a non-linear response function. It is therefore unclear whether the residuals are due to actual deviations of the seeing PSF from a Moffat function or are simply due to the CCD response function.

The former case would mean that we introduce a systematic bias into our galaxy parameters from systematically wrong PSFs, while the latter would not affect the galaxy fitting since the galaxies in our sample have much lower surface brightnesses, typically far from the saturation regime. While we cannot be sure, we assumed that the latter is the case and we need not worry about the brightest stars since they lie in a different regime of the dynamic range of the CCD than our galaxies. This belief was formed mainly by the observation that stars of intermediate brightness typically show no such systematic residuals (i.e. the residuals are not just weaker but disappear completely), while at the same time we obtain very consistent results for the fitting parameters across the stars of a field. We therefore focused on stars of intermediate brightness for which we can obtain perfect fits with a Moffat function.

The second case where we see clear systematic residuals in the star fits is for very elongated objects near the corners of KiDS tiles. An example is shown in the top two rows of Figure 2.10. Here, there are clear asymmetries in the residuals for stars of all brightnesses, which show that the true PSF is not perfectly elliptical but instead egg-shaped, with the wider part

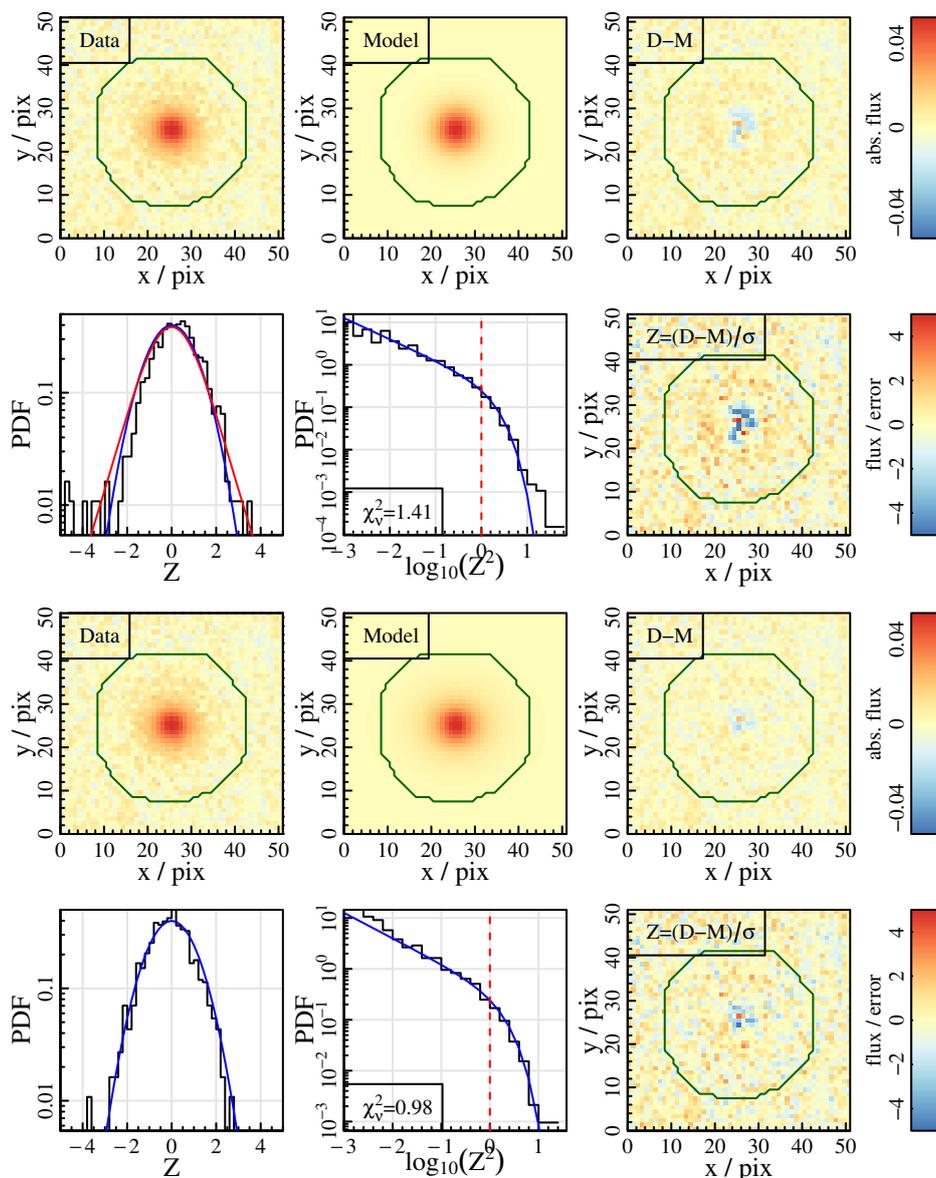


Figure 2.9: **Top two rows:** the Moffat function fit to a relatively bright example star near galaxy 7623. Panels are the same as the top two rows in Figure 2.2. **Bottom two rows:** the double Moffat function fit to the same star.

pointing towards the tile corner. Using the perfectly elliptical Moffat PSF will introduce a systematic uncertainty for the parameters of the corresponding galaxy. In particular we might bias the position angle and axial ratio, which are most severely influenced by incorrect PSFs (see Section 4.2.4). Due to the symmetry of the problem, however, only galaxies with a major axis oriented at approximately 45° with respect to the PSF elongation will be significantly affected. In addition, not all tile corners show elongated PSFs and for those that do the region is usually limited to the very outer edges of a tile. This also means that there are

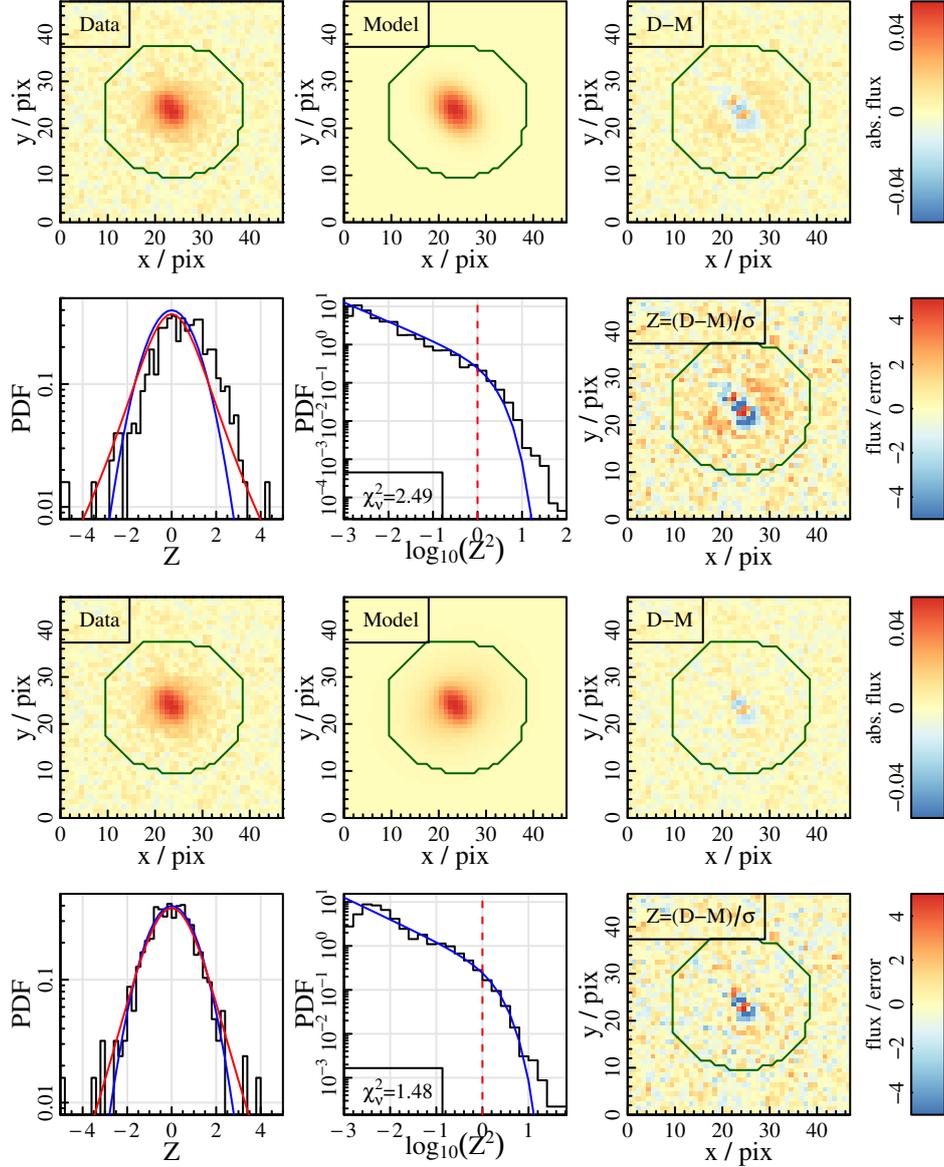


Figure 2.10: **Top two rows:** the Moffat function fit to an elongated, egg-shaped example star near galaxy 105600. Panels are the same as the top two rows in Figure 2.2. **Bottom two rows:** the double Moffat function fit to the same star.

often multiple matches (i.e. the galaxy belongs to the overlap sample), not all of which are necessarily compromised. Therefore, the overall number of galaxies affected is low and limited to those objects suffering from the worst data quality in many other respects, too.

In an attempt to improve the fits to very bright and/or elongated stars, we experimented with fitting double Moffat functions instead of just a single Moffat function to each star (following a suggestion by Dan Taranu, who uses this in the `AllStarFit` package). However, the success was limited: although the fit does improve owing to the increased number of parameters, in

both cases residuals remained (see the bottom panels in Figures 2.9 and 2.10). In particular, the systematic residual due to the egg-shape of the elongated PSFs is still present and also the bright star still shows hints of the residual pattern visible in the single Moffat fit. Conversely, the double Moffat fits (apart from increasing the computational time for PSF estimation by a factor of ~ 20) often produced unreliable results for those stars that are perfectly represented by a single Moffat function, leaving the second Moffat function unconstrained. Fitting some stars with a double Moffat and some with a single Moffat function bears its own problems, starting with having to decide which star to fit with which function and ending with having to devise a method to combine the different estimates into a single model PSF for the galaxy. Additionally, there are always parameter degeneracies when fitting two identical functions.

For these reasons, we concluded that fitting double Moffat functions to KiDS stars is only viable when jointly fitting all stars of a field. This increases the signal-to-noise ratio with respect to individual fits, so all parameters can be reasonably constrained; and eliminates the problem of combining the resulting PSFs, since there will only be one model. While this is the approach that e.g. Taranu et al. (2017) have taken (and was the original reason for the development of the `AllStarFit` package which we have used in other contexts), we decided against it for reasons outlined in Section 2.2.3. Therefore we decided to proceed with single Moffat PSFs despite its shortcomings for a small minority of galaxies; in particular since we believe that the imperfect fits to bright stars do not affect our analysis and so only a few galaxies in tile corners with very elongated PSFs are compromised. The test runs with double Moffat PSFs did also not show any significant or systematic improvement to the galaxy fits compared to galaxy fits with single Moffat PSFs.

Model PSF generation

After fitting, suitable stars for PSF estimation are determined as follows:

- The fitted centre in x and y must be within ± 1 pixel of the centre of the cutout
- The fitted magnitude must be within ± 0.1 mag of 0
- The reduced chi-squared (χ^2_{ν}) of the fit must be smaller than 3 (where χ^2_{ν} is evaluated within the star segment only even if a larger region was fit)
- FWHM, concentration index, angle, axial ratio and background (if fit) must not be equal to the fit limits (except for the axial ratio, which is allowed to be exactly 1 although this is the upper limit of the fit).
- Outliers in any of FWHM, concentration index, angle, axial ratio or background are rejected via an iterative 2σ clip (in logarithmic space where appropriate).

The first two criteria of these are fixed hard cuts since the input image was centred on the star

2.2. PIPELINE DEVELOPMENT

and normalised to a magnitude of 0 according to the segmentation statistics. Any deviation from these values indicates a difference in the position or magnitude estimated by `ProFound` and `ProFit` which likely points to bad segmentation, additional objects in the segment or a bad model fit. The cut in χ^2_{ν} is also a hard cut, chosen to remove objects that visually appear as bad fits (mostly the very bright stars discussed earlier). Note that the example star shown in Figure 2.9 is well below this limit. The reason is that we show an example star from a test run here, in order to directly compare it to the double Moffat fit. In this test run, however, we also used an additional cut on the magnitude of star candidates, excluding the brightest objects (see above). Consequently, there are only stars of intermediate brightness in this test run, one of which we show in Figure 2.9. During subsequent pipeline development, we have then changed the strategy and stopped using the magnitude cut, except for objects at the saturation limit (see discussion above). Instead, we now exclude these brightest stars with the cut in reduced chi-squared listed here.

We also exclude stars that hit their fit limit since these parameter estimates are not reliable and all fitting intervals were chosen very generously (except for the upper bound on axial ratio, which has a physical limit at 1, meaning perfectly round). Finally, we remove outliers in the respective distribution of the main fitting parameters to exclude any remaining objects that are not true point sources.

The stars fulfilling these criteria are classified as suitable, from which the selection is made:

- The closest two from each quadrant (8 in total) are selected to make sure they are roughly evenly distributed around the galaxy.
- If one or more quadrant contains less than two stars, the closest stars from any other quadrant (which are not already used) are taken instead to give 8 stars in total. A warning is given and the quality flag adjusted.
- If there are less than 8 stars in total, all of them are used with a warning (and a quality flag adjustment)
- If there are no stars classified as suitable, the default PSF mentioned above (round, FWHM from KiDS header, default Moffat concentration index) is used, with a warning and the corresponding flag.

There are two diagnostic plots summarising these selection steps (three if the background was fit as well), examples are shown in Figures 2.11 and 2.12. A list of stars with the relevant segmentation parameters and the fit parameters is saved for later reference.

From Figure 2.11 it is obvious that the Moffat models fitted to all star candidates in the field are very consistent in all parameters (shape, orientation, size and concentration). This is true even for stars which were not classified as suitable for PSF estimation (symbols with black

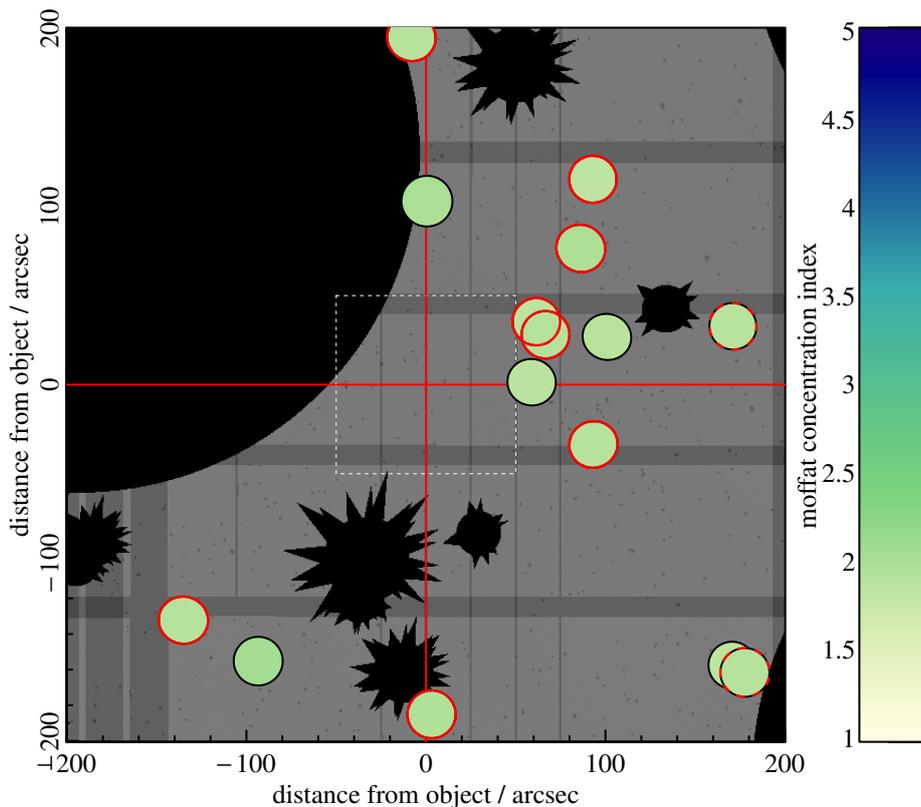


Figure 2.11: The result of the PSF fitting for the galaxy 396740 in the KiDS r -band with the dashed white square indicating the cutout shown in Figure 2.1. The greyscale image shows the r -band weight map with lighter colours meaning higher weight. Masked areas from the stacked gri -masks are shown in black (zero weight). The vertical and horizontal red line indicate the position of the object of interest (galaxy 396740) and split the image into its 4 quadrants. All fitted PSFs are shown as coloured ellipses with the size (FWHM multiplied by 20), axial ratio, orientation angle and concentration index (colour) taken from the fitted Moffat parameters. Stars selected for estimating the final model PSF have red borders; dashed red borders mean a fit was classified as suitable, but not selected because the maximum of 8 stars was already reached.

borders). We make the same observation in most cutouts, even though the PSFs themselves can vary greatly between frames (see Figure 2.14), with only occasional outliers appearing (which are subsequently excluded from the model PSF estimation).

Figure 2.12 shows that all star candidates of this frame are well-represented by a Moffat function. This is also generally the case except for very bright objects (close to the bright magnitude limit, of which there are none in this field), very elongated ones and those that have secondary objects included in their segmentation map. The latter of these are excluded either by the χ^2_ν cut or by the magnitude cuts since they will generally show a disagreement between `ProFound` and `ProFit` magnitudes. Note there is a slight systematic offset for all stars in this field, too, in that all stars have fitted magnitudes smaller than zero. As a reminder,

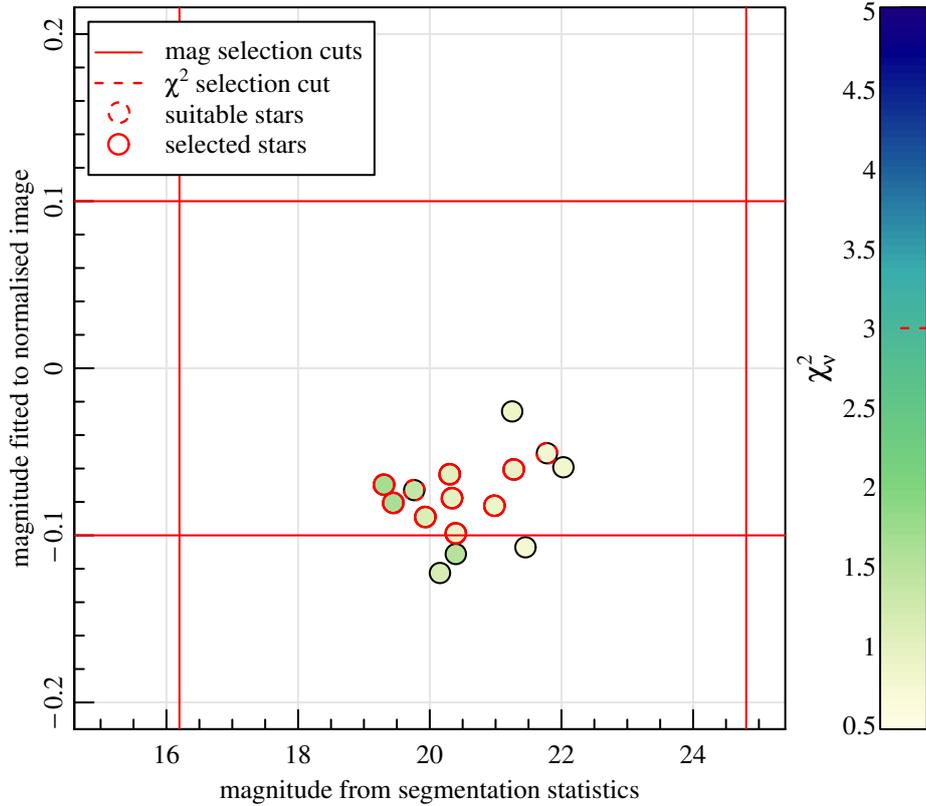


Figure 2.12: The second main diagnostic plot for the PSF fitting result, again for galaxy 396740 in the KiDS r -band. Here, we show the magnitude fitted to the normalised image of each star on the y -axis against the magnitude from the ProFound segmentation statistics on the x -axis, with symbols coloured by the reduced chi-squared of the fit. The magnitude and χ^2_v selection cuts are shown as solid and dashed red lines respectively (the latter on the colour bar); stars classified as suitable and selected are indicated with dashed and solid red borders.

we use the ProFound segment magnitudes to normalise the star images to a magnitude of 0 before fitting, so any deviation from zero in the fitted magnitude indicates that ProFit and ProFound estimate different magnitudes for the same object. However, small deviations below ~ 0.1 mag are no reason for concern since there are inherent differences in the two estimates: the ProFound estimate is simply the sum of all flux within the segment, while the ProFit estimate is the total magnitude of the Moffat function integrated to infinity, which can easily be somewhat brighter. However, larger deviations are usually indications of a bad model fit, bad segmentation, intruding (wings of) nearby objects or image artifacts and hence are excluded from further consideration.

Finally, the model PSF is created as an image of a two-dimensional Moffat function with parameters that are the medians of the selected stars. Exceptions are the centre of the star (in x and y), which is forced to the centre of the model image; the magnitude, which is forced

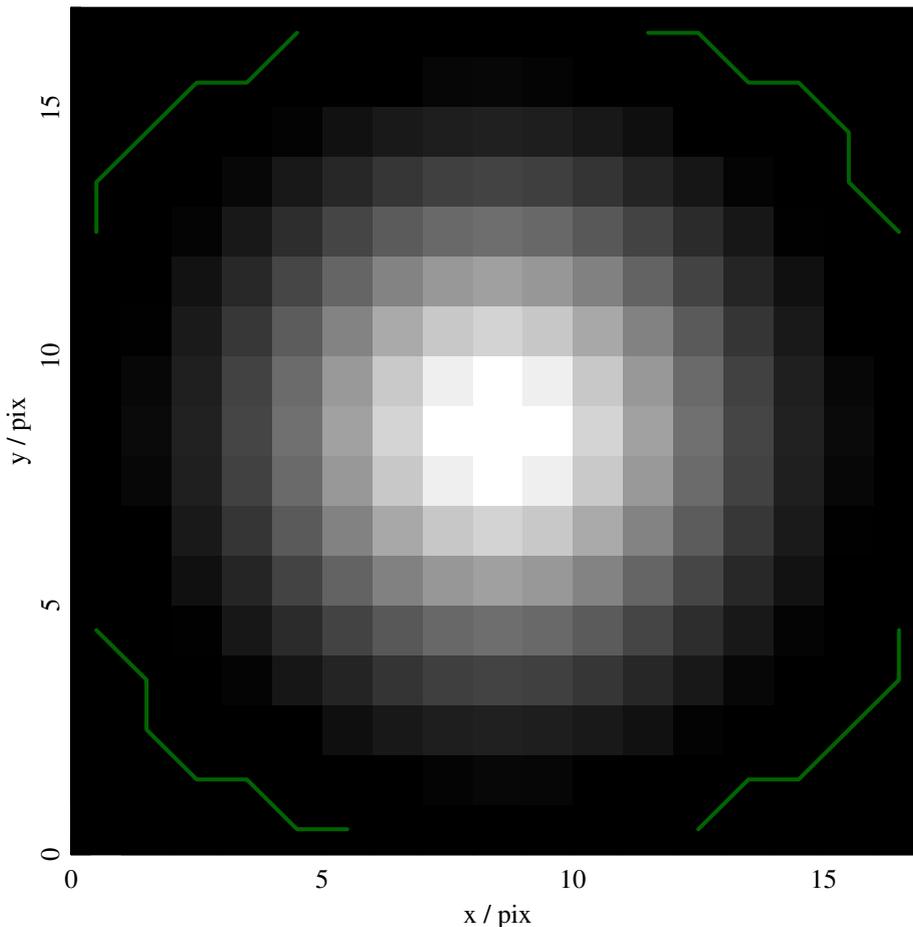


Figure 2.13: The model PSF obtained for galaxy 396740 in the KiDS r -band with asinh stretching of the flux scale. The green contours indicate the region beyond which the PSF is set to zero exactly.

to 0 and the background, which is ignored (i.e. set to 0 even if fit). The reason for forcing these values in the model PSF is to avoid systematic offsets when convolving the galaxy to be fit with its model PSF. The size of the image is adjusted to include at least 99% of the flux; or to a maximum of the median segment size within which the stars were fitted (to avoid extrapolation). Pixels in the corners of the image are set to 0 to avoid having a rectangular PSF and the image is subsequently re-normalised to sum to 1. The model PSF is stored as a FITS file and an image of it is returned as well, with an example shown in Figure 2.13.

There are many other ways in which we could have chosen to combine the individual PSF estimates into a final model PSF. We chose the approach of simply medianing the Moffat parameters mostly since it is practical, fast and robust. Taking the median ensures that we are not susceptible to outliers in the distribution (should there be any left after all of the above treatment), while working at the parameter level ensures that the resulting PSF is a smooth Moffat function with no abrupt changes. We avoid extrapolation of the model beyond

2.2. PIPELINE DEVELOPMENT

the region in which it was fitted by limiting the size of the final PSF accordingly; and in turn ensure an approximately round PSF by setting the very small values in the corners to zero exactly. We sample the PSF onto the same pixel scale as the KiDS images, so it can be used for direct convolution.

A first potential improvement to this procedure would be to finesample the PSF (and correspondingly the galaxy image) to allow a more accurate convolution. However, even the smallest possible oversampling factor of 3 (meaning a factor 9 increase in the number of pixels to convolve) leads to a prohibitive increase in the computational time for galaxy parameter estimation since the PSF convolution needs to be repeated for each model evaluation and constitutes a large fraction of the total computational time. Also, finesampling is only effective in increasing the accuracy of the convolution if the PSF is known to the level of accuracy implied by the finesampling. This can be achieved with PSF estimates based on large numbers of stars (best combined empirically or fit jointly) and/or theoretical knowledge about instrumental effects. In our case, where we fit only a relatively low number of stars individually, we cannot ensure or verify the PSF accuracy for significant finesampling factors.

Another potential improvement in creating the model PSF would be to consider parameter correlations. Currently, these are ignored since we take the medians of all parameters individually. One shortcoming of this is that the median of the axial ratio is taken independently of its orientation angle; which could lead to artificially elongated PSFs. For example, suppose the true PSF is perfectly round. Due to noise, the fitted axial ratios could be less than 1 (at least for some of the stars), with associated orientation angles that are randomly distributed. Taking the median of these two parameters independently of each other could therefore lead to a slightly elongated model PSF profile oriented at a random angle. In theory, this effect could lead to our PSFs being systematically too elongated, especially since it does not work the other way around (i.e. we will not have estimates which are systematically too round for truly elongated PSFs, since that would require the noise to be distributed in the same way for all stars of a field, which is highly unlikely).

After noticing this shortcoming, we have double-checked many of our model PSFs by studying the residual left after subtracting it from all stars that it was based on; and found no evidence of the PSF being systematically elongated. We also compared the average axial ratio of all model PSFs in a tile to the average PSF ellipticity given in the KiDS header of the same tile and found our estimates to be consistent with the KiDS value for individual tiles; and slightly rounder than the KiDS value on average across all tiles. Since the above effect cannot produce PSFs that are too round, we conclude that there are no systematic biases in our PSF shape introduced by our method. The difference between the KiDS values and our averages are likely because our average is based on the model PSFs which are the medians of a highly selected set of stars; while the KiDS value is computed as the average ellipticity of all individual stars

in the field which is likely to be noisier (and therefore less round).

While we conclude that the PSF estimates are not compromised by this effect, it would still be preferable to consider parameter correlations (between angle and axial ratio, but also other parameters) when computing the final model PSF. One way to achieve this is to perform a joint fit of several stars. This has the added advantages that it increases the signal-to-noise ratio relative to fitting individual stars, allowing to improve the PSF accuracy and possibly fitting more detailed models (e.g. a double Moffat function as mentioned above). We have briefly explored this route, but unfortunately it takes much longer than fitting the stars individually since the computational time scales non-linearly with the number of parameters and there are many parameters in a joint fit since each star needs its own position (x and y) and magnitude to be fitted, even if the PSF shape is the same for all stars. Furthermore, implementing a joint fit requires significant improvements to the star candidate selection process, which is not critical in the current approach since outliers can easily be excluded after fitting. For a joint fit, this step would have to be replaced by an alternative - or else we would have to first fit stars individually, select suitable ones and then perform a joint fit of those (instead of medianing the parameters), additionally increasing computational time. For these reasons, we did not implement such an approach in the current work. Nonetheless, it remains an interesting avenue to explore in future work since it would solve several shortcomings of the PSF estimation at once.

As a last point of how the PSF pipeline could potentially be improved, we would like to discuss the number of stars we chose for fitting. Generally, fitting a higher number of stars will result in a more robust PSF estimate. A competing effect is that stars closer to the galaxy will reflect the PSF at the galaxy position better than those further away. A robust local PSF estimate can therefore only include stars within a field small enough such the PSF does not vary significantly across it. Larger fields would require fitting multiple PSFs and interpolating between them, introducing additional uncertainties. The size of the large cutout to identify stars in ($400'' \times 400''$) is chosen to balance these two effects. It is small enough such that the PSF does not usually change significantly across it (see Figure 2.11), since the PSF in KiDS tiles does vary, but slowly (Figure 2.14). At the same time, the cutout is large enough to contain a reasonable number of stars for PSF fitting (~ 10 - 20 on average) unless large fractions of the cutout are masked. In order to make the PSF estimates comparable for crowded fields (sometimes containing more than 30 stars) and sparse ones or those with masked areas (where there are often only a few stars available), we take a maximum of 8 stars for each PSF estimate even if more would be available.

PSF quality control

As mentioned above, we produce a number of PSF quality control plots for each galaxy (Figures 2.7, 2.8, 2.11, 2.12, 2.13) which are stored on the GAMA file server for reference. In addition, we investigated the robustness of our estimates by comparing model PSFs for different galaxies in the same KiDS tile against each other. The main diagnostic plot of this comparison is shown in Figure 2.14 for four different KiDS tiles. Similar to Figure 2.11, we show the weight map as greyscale in the background with masked areas in black; and the PSFs as coloured contours with the size ($\text{FWHM} \times 100$), axial ratio, orientation angle and concentration index (colour) taken from the model PSF estimated for the galaxy at this position. Note that in contrast to Figure 2.11, the PSFs shown here are not fits to individual stars, but Moffat models created from the combination of several suitably selected stars. Therefore, the meaning of the red contours changed: solid red indicates model PSFs based on 8 stars, while dotted red means that less than 8 suitable stars were available. Black contours are drawn around default (dummy) model PSFs that had no suitable stars available and so should be ignored. Also, we now multiply the FWHM by 100 instead of 20 to enhance visibility in the much larger area ($\sim 1 \text{ deg}^2$ instead of $400'' \times 400''$).

The top left panel shows the tile `KiDS_DR4.0_175.0_1.5_r_sci.fits` in which galaxy 396740 (our usual example galaxy) resides near the centre. For orientation, we indicate the $400'' \times 400''$ cutout that is shown in Figure 2.11 as a white rectangle. The top right panel shows a tile with generally much smaller PSFs, `KiDS_DR4.0_216.0_-0.5_r_sci.fits`. On the bottom left we see `KiDS_DR4.0_130.0_1.5_r_sci.fits`, which shows larger and more elongated PSFs and finally, the bottom right image shows `KiDS_DR4.0_213.0_1.5_r_sci.fits` which has more pronounced PSF variations and is also the tile in which the example elongated star of Figure 2.10 resides (near the bottom right corner).

Comparing all tiles we can see that the PSF differences between tiles are much larger than the variations within a tile. The latter are generally relatively small and smooth with no abrupt changes. This is what one would expect for KiDS data, since it is seeing-limited and the seeing will vary greatly with time (i.e. between tiles); while all dithers composing a tile were taken in close temporal succession. The fact that we can observe this in the overview plots is therefore reassuring. Remember that each PSF in Figure 2.14 is a model PSF that was estimated entirely independently from all other model PSFs in the tile (except for galaxies very close to each other which could have some stars overlapping). The generally similar appearance of all model PSFs in a tile (with no clear outliers) combined with the slow variation across it is a sign of our PSF estimates being robust.

Investigating the variation across the individual tiles further, it can be seen that generally, the PSFs tend to become more elongated towards tile edges and corners, with the concentration and FWHM also increasing. This effect is more pronounced in some tiles (e.g. bottom right)

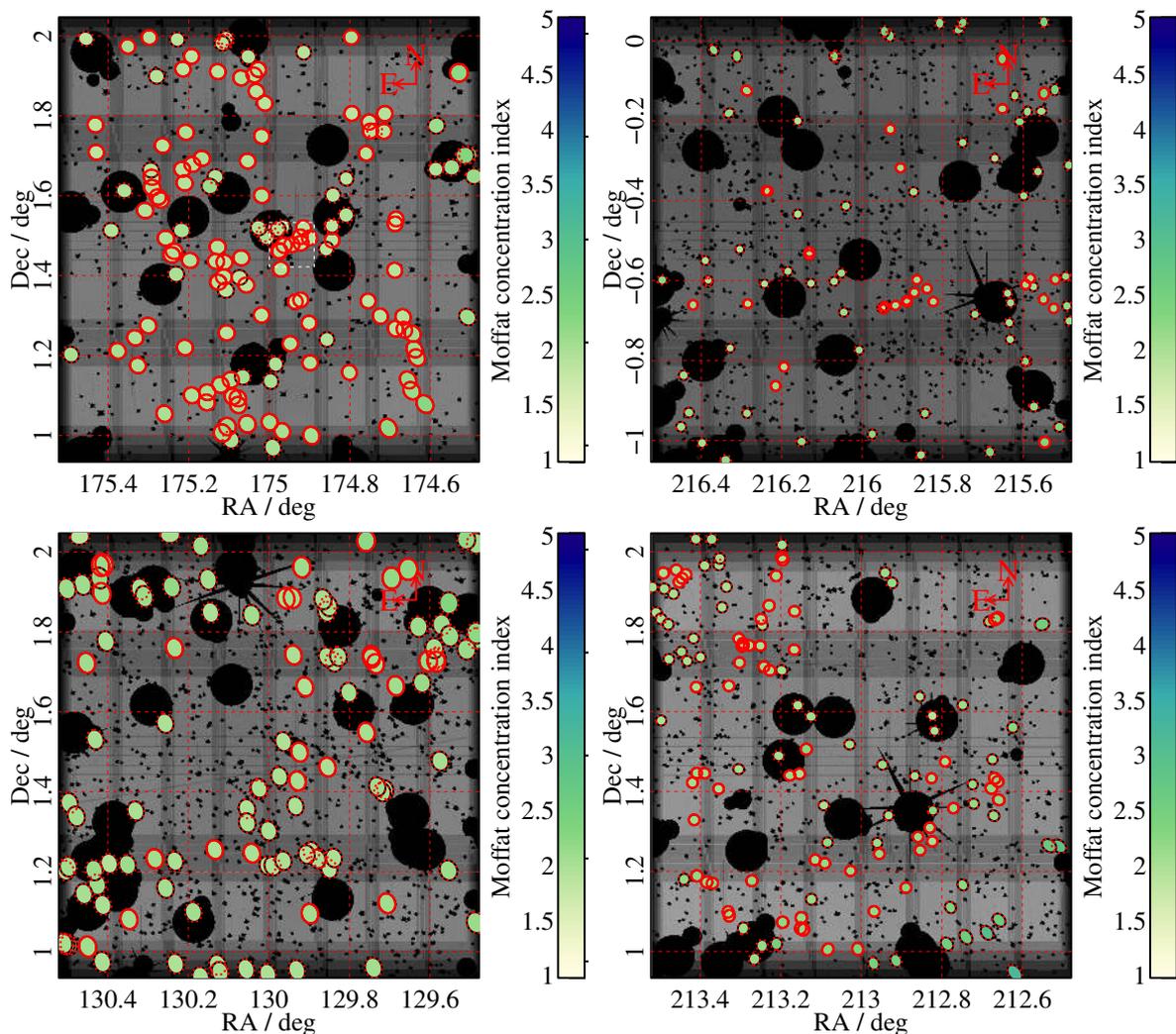


Figure 2.14: The model PSFs obtained for all galaxies in four different KiDS r -band tiles (top left: `KiDS_DR4.0_175.0_1.5_r_sci.fits`, top right: `KiDS_DR4.0_216.0_0.5_r_sci.fits`, bottom left: `KiDS_DR4.0_130.0_1.5_r_sci.fits`, bottom right: `KiDS_DR4.0_213.0_1.5_r_sci.fits`). Similar to Figure 2.11, the greyscale image shows the weight map with masked areas shown in black; while the coloured ellipses represent the PSFs ($\text{FWHM} \times 100$, axial ratio, orientation angle and concentration index). The major difference to Figure 2.11 is that we now do not show fits to individual stars, but instead the model PSFs derived for the galaxy positions. Solid red contours now mean that the model PSF is based on 8 stars, while dotted red contours are drawn around PSF estimates based on less than 8 stars. Black contours mean this is a default (dummy) PSF and should not be considered further. The white rectangle in the top left tile indicates the region shown in Figure 2.11.

than in others. Also, even within a tile, not all corners are equally affected, our impression was that the bottom right and top right corners show elongated PSFs more often than the top left and bottom left corners. The orientation angle seems to rotate such that the major axis always points towards the corners. For tiles that have intrinsically elongated PSFs, this

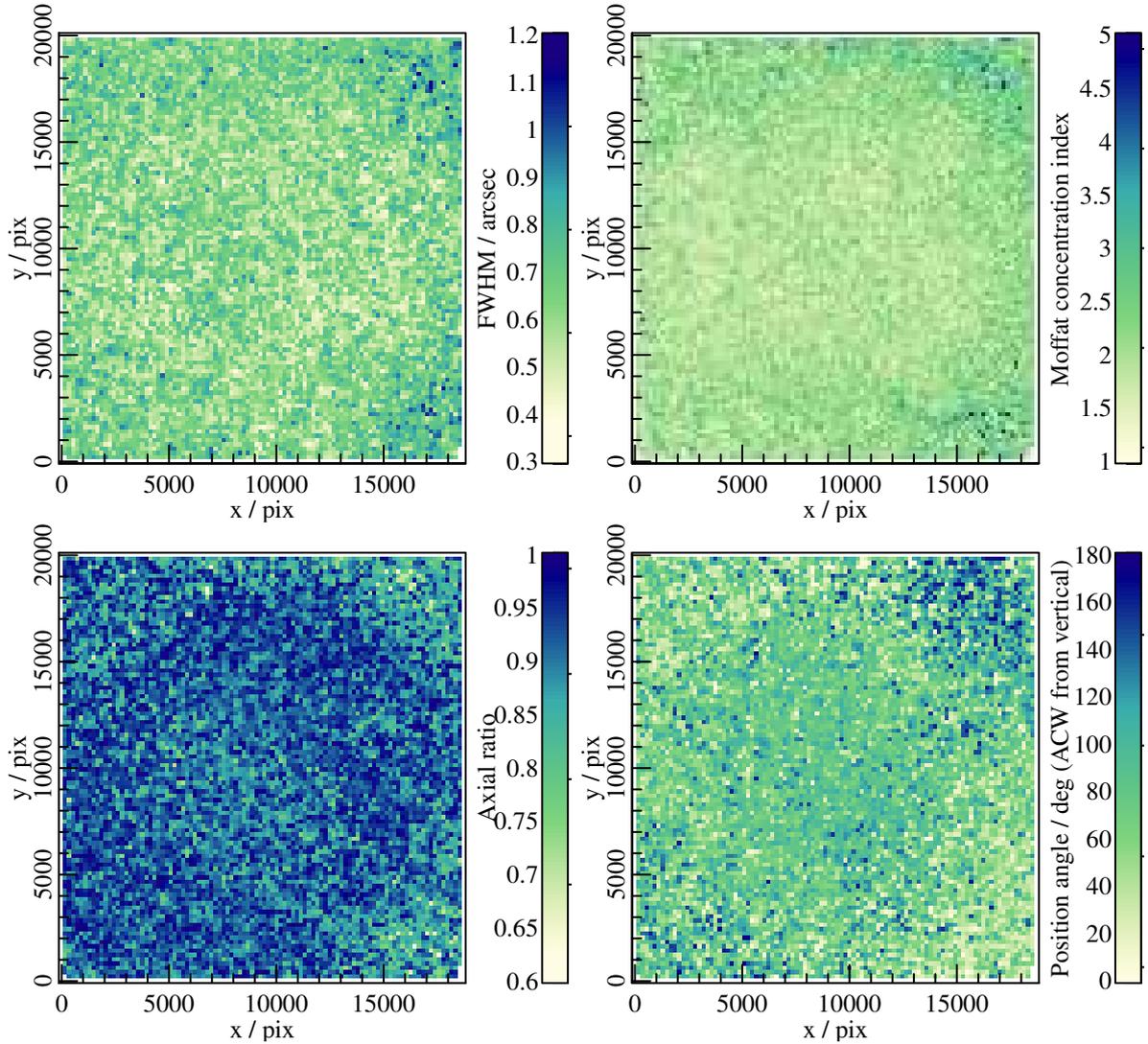


Figure 2.15: The variation of the FWHM (top left), Moffat concentration index (top right), axial ratio (bottom left) and position angle (bottom right) for model PSFs averaged across all KiDS r -band tiles. The position angle is measured anticlockwise (ACW) from the vertical axis and limited to the range between 0 and 180° due to symmetry.

can result in PSFs actually becoming rounder towards corners (e.g. in the top right corner of the bottom left panel).

To gain a better understanding of these systematic trends, we stacked the results for all KiDS r -band tiles, the result of which can be seen in Figure 2.15. As expected from the examples shown in Figure 2.14 above, the stacked plots show that the FWHM and concentration index generally increase towards the corners of tiles, while the axial ratio decreases and the angle rotates such that the major axis always points towards the nearest corner. Our visual impression that the right side of the tiles seems to be more affected than the left side is also confirmed. Interestingly, there is also a region near the centre of the tiles where the FWHM

and concentration index increase while the axial ratio decreases slightly, but with no clear corresponding trend in the position angle. The latter instead seems to be randomly distributed in the central tile region, such that averaging over all tiles results in a value of approximately 90° , i.e. the centre of the possible range of angles.

We have also visually compared these tile overview plots to the PSF diagnostic plots available from the KiDS database and found the general trends to agree well.

PSF effects on galaxy fitting

Besides the internal PSF consistency checks presented above, we investigated the effects that using different PSFs has on the fitted galaxy parameters. To this end, we fitted a test sample of galaxies with various different PSFs, sampled from the stars classified as suitable. An example can be seen in Figure 2.16.

We show the result of the fitted parameters for four different “versions” of the same galaxy: yellow shows the result for an image of the galaxy using 8 different PSFs sampled from the suitable stars (2000 MCMC samples each, i.e. 16000 points in total). Orange is for the same image of the galaxy, just using the usual (medianed) model PSF (2000 points). Light blue and dark blue show the same for a different image of the same galaxy, since the galaxy sits in the overlap region between KiDS tiles. The top left half of the plot then shows the MCMC samples for each combination of the seven single Sérsic parameters: position in RA and Dec, magnitude m , effective radius R_e , Sérsic index n , the position angle PA and the axial ratio b/a . The bottom right half shows the corresponding contours, with dashed, solid and dotted lines including 50 %, 68 % and 95 % of the data. On the diagonal we show the one-dimensional distributions for each parameter. Crosses indicate the mean of each distribution.

There is a clear discrepancy in the fitted RA and Dec centres for the two images, which is caused by the accuracy of the KiDS astrometric solution, see Section 4.2.4. Apart from that, all distributions overlap, with varying levels of consistency for the different parameters. The distributions sampling from the PSFs are broader than those using just the medianed model PSF. This shows that the PSF introduces systematic uncertainties which are not accounted for in the MCMC errors. The difference is larger for the second of the two images (blue), where the sample using the model PSF is also not centred on the same values as the sample obtained by varying the PSF. Nonetheless, the parameters fitted to the same image with different PSFs are in closer agreement than those fitted to different images. This is because for different images, there are additional sources of systematic uncertainties caused by, e.g., the background subtraction and different segment sizes. We investigate all of these systematic uncertainties in more detail in Section 4.2.4, where we derive the average systematic errors for each parameter using the overlap sample and point out parameter biases using simulations.

2.2. PIPELINE DEVELOPMENT

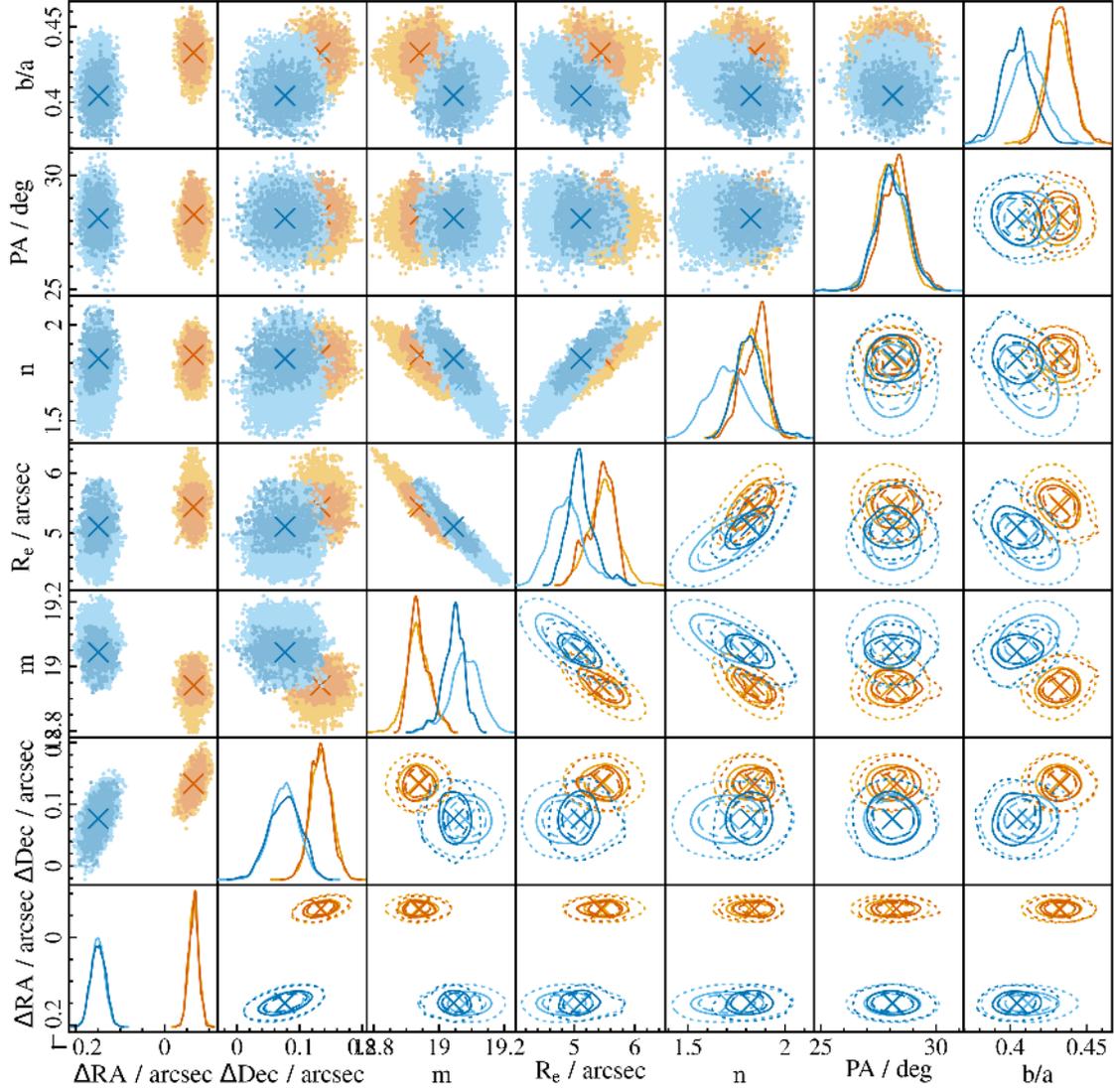


Figure 2.16: The effect of different PSFs on the fitted galaxy parameters for the galaxy 105723, which is part of the overlap sample. For each combination of the seven single Sérsic parameters we show the distribution of the MCMC chain points in the upper left half of the plot; while the lower right half shows the corresponding contours including 50 %, 68 % and 95 % of the data for dashed, solid and dotted lines respectively. The diagonal shows the one-dimensional distribution for each parameter. Yellow and orange lines and points show the parameters fitted for the first match of this galaxy, where the former indicate the full distribution obtained by sampling from the PSFs and re-fitting the galaxy 8 times (16000 MCMC samples in total), while the latter shows the result using just the model PSF (i.e. the medianed version, 2000 MCMC samples). Light and dark blue lines and points show the same for the second match, i.e. for the same physical object but a different image since the galaxy sits in the overlap region between KiDS tiles. Crosses show the mean of each distribution. For RA and Dec, instead of the absolute values, we show the difference to the GAMA input positions in arcseconds (for readability).

2.2.4 Segment sizes

We have briefly discussed image segmentation in Sections 2.1.1 and 2.2.2. One of the conclusions drawn there was that it is best to use two different segmentation maps for the background subtraction and the galaxy fitting due to the different aims that we would like to achieve with them. We make use of the different settings in `profoundProFound` to produce these segmentation maps. In the following, we detail the definition of the segmentation maps and how it has evolved over the time of pipeline development. We focus on the fitting segmentation map rather than the background subtraction segmentation map since that has already been discussed in Section 2.2.2. The latter part of the section then describes our own routine for segmentation map fixing, which also makes use of the fitted galaxy parameters. Hence this section is somewhat at the intersection between the preparatory work and the galaxy fitting choices.

skycut and dilation

The most important parameter that influences the segmentation in `profoundProFound` is `skycut`, which determines how deep the source extraction pushes into the noise (see Section 2.1.1 and Robotham et al. 2018 for details). A lower value of `skycut` means that more (fainter) sources are detected and segments for individual sources are larger. In addition, already-defined segments can be dilated (expanded) with the `profoundMakeSegimDilate` function, which is also used internally by `profoundProFound` to achieve flux convergence during the segment growing phase and to obtain a more aggressive object mask by further dilating all segments after convergence. There are additional parameters that influence the segmentation, most notably `pixcut`, `tolerance` and `sigma` (minimum number of pixels needed for segment detection, the tolerance to use for merging segments and the radius within which to search for neighbouring objects). We experimented with varying all of those parameters, but their effect is secondary to `skycut`. In the end we reverted to using the `profoundProFound` defaults for these parameters and will not discuss their effects further.

For the background subtraction, described in detail in Section 2.2.2, we use the rather low `skycut` value of one and the aggressive object mask returned by `profoundProFound` after additionally dilating the converged segments. This choice was mainly driven by theoretical considerations and ensures that we obtain a clean sample of sky pixels without (wings of) undetected objects for background subtraction. It may lead to an increased fraction of spurious detections and a significant number of sky pixels included in the segmentation maps, but this does not affect the background estimation as long as there are enough sky pixels remaining to obtain robust statistics.

Up to v02 of the pipeline, we used the same segmentation map for fitting objects, just without

2.2. PIPELINE DEVELOPMENT

the additional dilation for aggressive object masking. However, this often resulted in very jagged segment borders which include noisy, slightly positive pixels but exclude slightly negative pixels, thus artificially biasing the flux in galaxy outskirts. This is particularly true for large, bright objects where the number of curve of growth iterations performed to reach flux convergence is often zero or one. For the background subtraction this problem is alleviated by the additional dilation performed to obtain the aggressive object mask, but it sometimes caused issues for the galaxy fitting where no such additional dilation is performed.

From v03 onwards, we therefore decided to run `profoundProFound` a second time to obtain the fitting segments. Here we use a higher `skycut` value of two resulting in smaller segments with smoother borders. We then perform an additional dilation of the galaxy segment, increasing its area by typically 30%, which ensures that the edges are smooth and unbiased since the dilation is independent of the pixel values. These dilated segments starting with a `skycut` value of two are then approximately the same size as the previous undilated segments with a `skycut` value of one. An example of the difference can be seen in the last two panels of Figure 2.20, where we discuss the segmentation map fixing. At the same time we started defining all segments on the stacked *gri* images instead of in individual bands, to ensure that the fitted pixels are exactly the same in all bands to make the measurements most directly comparable.

We double-check the chosen `skycut` value based on the fraction of objects returned. Assuming a Normal distribution, the limit at which approximately 50% of the object pixels actually belong to objects rather than the sky background is given by the Normal quantile of one minus the fraction of pixels identified as objects. `skycut` should be larger than this limit to ensure that the majority of “object” pixels belong to real objects rather than spurious detections or overly extended segments pushing well beyond the object wings. But `skycut` should also not be much larger than this limit if one desires to detect faint sources and low surface brightness wings of extended objects. We therefore add a warning and raise a quality flag if the chosen value of `skycut` is below the “50-50-limit”, or if it is more than 1σ above it.

The computed “50-50-limit” is around unity in most cases, confirming that our assumption of Normal statistics is valid. For the early version of the pipeline, using a `skycut` value of one, the flag was therefore raised frequently due to small random fluctuations. Raising the `skycut` value to two decreased the percentage of objects with this flag raised from approximately 65% to 0.3%. Visual inspection of those galaxies revealed they are in regions where large fractions of the background are masked, such that the object statistics are not reliable anymore.

We conclude that our approach for defining fitting segmentation maps is suitable for our purposes. The initially higher `skycut` value of two prevents noise fluctuations from being detected, while the additional segment dilation ensures that the majority of the flux from the galaxy wings is included. Note that the segments defined in this way are still relatively tight

in that they do not include many sky pixels. This is a deliberate choice we made to obtain the best possible fit in the high signal to noise regions of the galaxy, at the expense of not necessarily normalising the fitted profile to zero at large radii. We discuss the benefits and drawbacks of this approach in detail in Section 2.1.3, also in comparison to previous works.

Systematics from segment sizes

Complementary to the statistical discussions in Section 2.1.3 and Section 4.2.4, we present a case-study on the systematic effect of different segment sizes here.

Figure 2.17 shows a corner plot similar in nature to the one presented in Figure 2.16, also for the same galaxy 105723. The distributions of the fitted parameters are shown for all combinations of the seven single Sérsic parameters, with MCMC samples in the top left half of the plot, corresponding contours in the bottom right half and one-dimensional distributions on the diagonal. Lines and points in colours ranging from red to orange to yellow show the parameters fitted to the first image of that galaxy for continuously increasing segment sizes. Blue through to green colours show the same for the second image of the galaxy. In both cases, the results are shown for five different segment sizes: no additional dilation after the `profoundProFound` run (red and blue) and dilations which increase the segment diameter by approximately factors of 1.5, 2, 2.5 and 3 respectively.

Note that for our final `v04` pipeline, as described above, we settled on an additional dilation based on increasing the area of the segment by approximately 30%. This corresponds to a diameter increase of less than 1.5, i.e. it is not shown in the plot. The reason we do not show the final version here is that the segment size studies have been carried out on a relatively early test run, with many changes in the subsequent pipeline development both for the preparatory work (e.g. background subtraction) and the fitting routines which were still under development at that stage. We therefore only use this case study to gain an intuitive understanding of the effect of different segment sizes and return to a more quantitative analysis of the final segment choice in Figure 2.19 using our simulations.

The most obvious feature in Figure 2.17 is that the RA and Dec positions fitted to the two different images of the galaxies disagree, as already seen in Figure 2.16. We believe this to be due to slightly different astrometric solutions for the two different KiDS tiles based on our investigation of the same issue in Section 4.2.4. For a given image, the fitted positions are in perfect agreement for all segment sizes. This is different for most other parameters, in particular for the highly correlated magnitude, effective radius and Sérsic index: the fits using the smallest, undiluted segments (shown in red and blue) are inconsistent with those using larger segments. As a general trend, also observed in other galaxies, fits to these small undiluted segments tend to result in brighter and larger models with higher Sérsic indices.

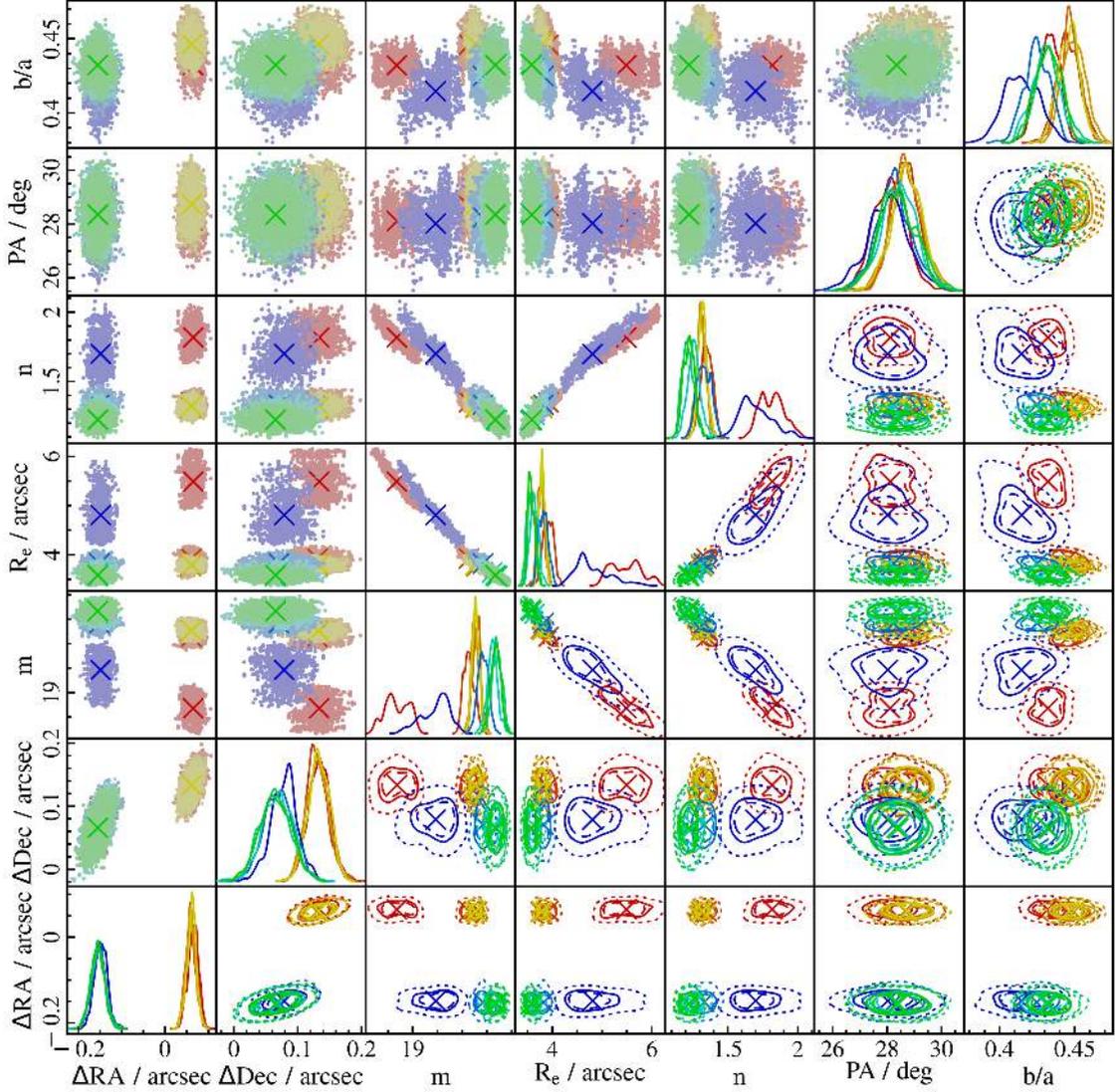


Figure 2.17: The effect of different segment sizes on the fitted galaxy parameters for the galaxy 105723, which is well-represented by a single Sérsic fit. The general layout of the plot is the same as for Figure 2.16. Red to orange to yellow colours represent the parameters fitted for the first match of the galaxy for continuously increasing segment sizes. Blue through to green colours show the same for the second match of the galaxy.

This re-inforces our belief that for undiluted segments, the galaxy flux in the outskirts is biased positive, increasing the fitted Sérsic index. Extrapolating this to infinity to obtain the total Sérsic magnitude then results in a model that is too bright since it includes significant amounts of flux beyond the segment borders where it is unconstrained, with a corresponding increase in effective radius. The position angle and axial ratio are uncorrelated to these three primary parameters and less affected by the segment size. For increasing segment sizes (with

additional unbiased dilations) the models rapidly converge onto common values, with small differences between the two images due to other sources of systematic uncertainties.

This rapid convergence is not achieved for all galaxies. Instead, it is only the case when the model is an adequate representation of the data, i.e. the galaxy (in this case) truly follows a single Sérsic profile. For objects that have intrinsically different shapes than the profile(s) we try to fit to them, there is a continuous evolution of the fitted parameters with the segment size, as the focus of the fit is shifted more and more from the inner to the outer regions. The convergence regime is potentially never reached in these cases; or at a much later stage when the segment size is so large that the background normalisation dominates the fit rather than the galaxy itself.

Figure 2.18 shows an example of such a fit. The layout of the plot and the meaning of the colours are exactly the same as for Figure 2.17, only that now we show a different galaxy, namely 107016. This galaxy is not well-represented by a single Sérsic fit and instead needs (at least) two components.⁵ There is therefore no unambiguous, correct solution for a single Sérsic fit to that galaxy and the fitted parameters will always depend on the exact segment size used.

This is a point we already made in Section 2.1.3, but it is re-inforced here. Choosing the optimal segment size is therefore non-trivial and depends on the science aim. Since our focus - as explained before - is on obtaining the best possible fits of the high signal to noise regions of galaxies, we do not want to make the segments very large. Nonetheless, they should be large enough to reach the convergence regime at least for cases where there is an unambiguous solution, such as that shown in Figure 2.17.

The best way to determine where the convergence regime begins is using simulations. These have two major advantages: we know that the fake galaxy is well-represented by a single Sérsic fit if it was created as such; and we know the true parameter values used to create the galaxy. In addition, we know the true PSF used to convolve the generated galaxy with, eliminating PSF uncertainties from the analysis. We have created such simulations for our studies of systematic uncertainties in Section 4.2. More details on their creation can be found there, for the analysis in this section it suffices to say that they are all perfect single Sérsic objects, but realistic in all other properties. After creation, we run them through our pipeline as if they were real galaxies.

Figure 2.19 shows an example similar to Figures 2.17 and 2.18 for a simulated galaxy. Since the simulations were run at a later stage during pipeline development than the segment size studies shown above, the colours have changed in meaning slightly: red to yellow still shows the fits to the first match of the simulated galaxy (which we also chose to lie in the overlap

⁵The double component fit of the object is acceptable, although there are still slight residuals visible at the centre that may point to the existence of a bar.

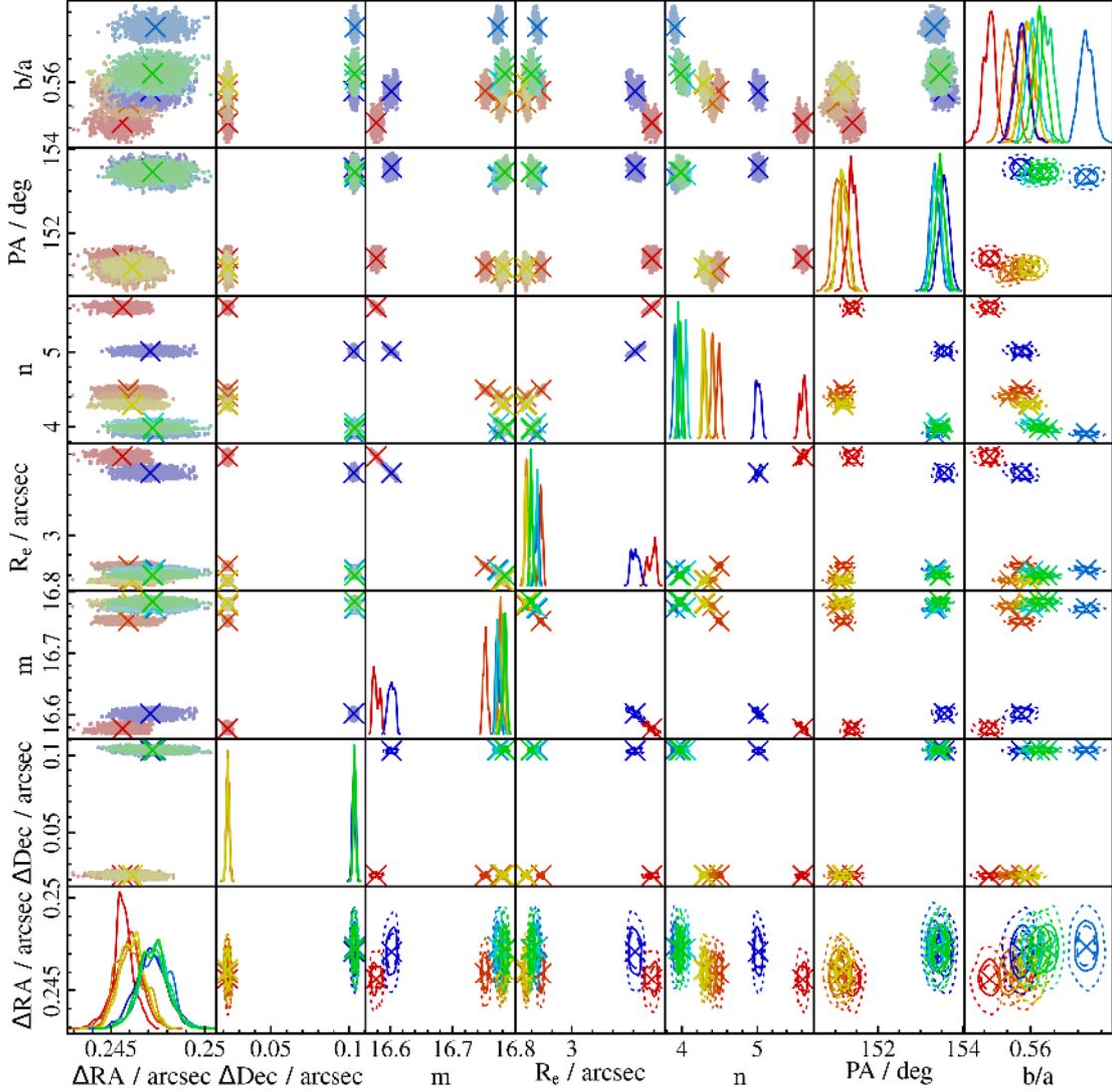


Figure 2.18: The same as Figure 2.17 for galaxy 107016, which is not well-represented by a single Sérsic model.

region) with increasing segment size; while blue to turquoise shows the same for the second match. There are now only three segment sizes shown: no additional dilation (red and blue), dilation such that the area increases by approximately 30% (i.e. the version also used for the final v04 pipeline, orange and light blue) and dilation with a kernel five times larger than the value used for v04, corresponding to an area increase by about a factor of 2.5 (yellow and turquoise). Green lines and crosses show the true parameter values.

In contrast to Figures 2.16 to 2.18, the RA and Dec values for both matches in Figure 2.19 are in perfect agreement since the fake galaxy was inserted after the astrometric solution

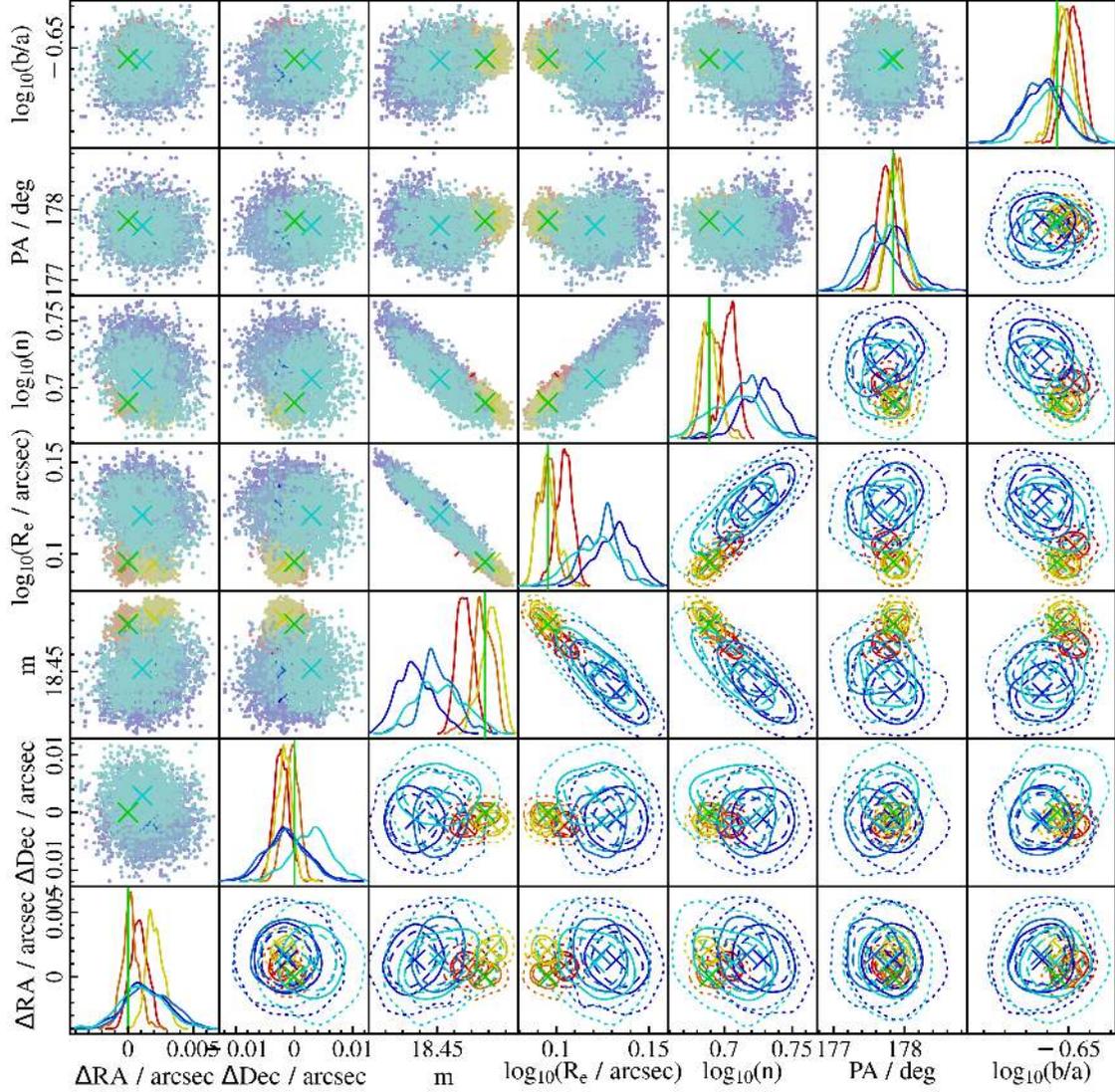


Figure 2.19: Similar to Figures 2.17 and 2.18, but now showing a random single Sérsic simulated galaxy. Again, red through to yellow and blue through to turquoise colours show the fits obtained for the first and second match respectively for increasing segment sizes. The exact segment sizes shown differ from those in Figures 2.17 and 2.18, see text for details. Green crosses and lines indicate the true values used for generating the galaxy.

had already been fixed. The agreement between the two matches and the true value for all other parameters is reasonable: the first match (red, orange, yellow) is generally closer to the truth than the second. It is also the deeper one of the two images, evidenced by its higher constraining power (narrower distributions). The second, shallower, image seems to suffer from some systematic uncertainty, possibly from the background subtraction (it cannot be due to the PSF since we use the true PSF used for generating the galaxy also for fitting). We

2.2. PIPELINE DEVELOPMENT

do not investigate this in detail here but refer to Section 4.2.4, where we explore systematic uncertainties in more detail.

The differences between the different segment sizes are generally small. In particular, both versions of the dilated segments show a high degree of overlap despite their very different sizes (30 % area increase vs. a factor of 2.5). The undilated segment again tends to converge onto a solution that is brighter, larger and with higher Sérsic index and differs more from the other two. This is despite it being relatively close in size to the first one of the dilated versions and again emphasises the importance of a dilation step that is independent of the pixel values to obtain an unbiased segment solution.

In conclusion, an unbiased dilation step after the `profoundProFound` curve of growth convergence is necessary. A relatively small dilation, increasing the area only by approximately 30 % is sufficient to reach the parameter convergence regime for cases where there is an unambiguous solution, i.e. where the model represents the data well. At the same time it minimises influence from, e.g., background subtraction uncertainties or neighbouring objects that can become more dominant for larger segment sizes. It also ensures that when working with a model that is not a perfect fit to the data, the focus is on correctly representing the regions dominated by the galaxy flux rather than the low signal to noise outskirts or the sky background.

Segmentation map fixing

Early versions of `ProFound` had a tendency to split up large, well-resolved galaxies with prominent substructure into several smaller segments. To combat this unwanted shredding effect, we devised a segmentation map fixing procedure using `ProFit`. This routine was used up to v02 of the `BDDcomp` DMU. It then became redundant since we upgraded `ProFound` to its newest version which saw major changes in how the segmentation was performed. The updated version of `ProFound` was much less likely to shred galaxies, even for large and well-resolved cases with a lot of substructure. We therefore turned off our own segmentation map fixing from v03 onwards since it was consuming unnecessary computational time and had the unwanted side effect of sometimes including faint and small sources in the segmentation map for the galaxy. Nonetheless, we briefly describe the segmentation map fixing routine here.

After the `profoundProFound` run during the preparatory work, the galaxy is fitted with a single component Sérsic model using the same fast downhill gradient algorithm used for the star fitting in PSF estimation (Section 2.1.1). Like for the “real” galaxy fits (Section 2.1.2), we fit all parameters except boxyness, use only the pixels within the galaxy segment for fitting and take initial guesses from the segmentation statistics.

For each segment in the cutout around the galaxy, the flux it contains is then compared

against the flux it should contain according to the fitted model: if the model flux accounts for more than 50% of the total flux in the segment and the segment contains more than 1% of the total model flux, then the segment is added to the galaxy segment. Point sources within the segment of interest are also identified by comparing to a smoothed version of the image and masked out. The galaxy is then re-fitted and the process repeated until no more segments are added. A diagnostic plot is returned and the fitted parameters are written to the header of the segmentation map to serve as better initial guesses in the galaxy fitting process.

Since the success of this segmentation map fixing crucially depends on the model fit, we repeat the process several times during the actual galaxy fitting. Therefore, after the first round of single Sérsic MCMC fits, the new (better) model image is used to check again whether any segments need to be added to the galaxy segment. If this is the case, the object is flagged, the segmentation map is updated and the fit is repeated. The updated segmentation map and initial guesses (in the header) are saved along with a diagnostic plot and the final fits. This process is repeated a maximum of five times, if the galaxy segment still changes in the last round the object is flagged.

The segmentation maps are then checked and fixed again on the basis of the model images resulting from the double component fits (using the same procedure). Galaxies where the segmentation map changed during this step are then re-fitted with single Sérsic models to ensure that the single Sérsic and double component fits are directly comparable (i.e. using the same segmentation map). The segmentation maps are not fixed again during this re-fitting step nor during any subsequent (1.5-component) fits of the galaxy.

Figure 2.20 shows an example for a galaxy with a nearby companion and two small point sources that led to it being split into several segments in `v02` of the pipeline. This can be seen in the first (leftmost) panel, where the galaxy segment is shown as a green contour and all other segments in blue. The second panel then shows the model fitted to that image during the segmentation fixing process, using the original galaxy segment only. In the third panel, we see the fixed segmentation map obtained after two iterations. Three segments have been added to the galaxy segment and two point sources masked out, such that the segmentation is now acceptable. It includes the majority of the galaxy flux, but excludes the neighbouring source. These first three panels are what is also shown in the segmentation map fixing diagnostic plot mentioned above.

The fourth panel shows the result of the segmentation directly after upgrading `ProFound` with our own segmentation fixing routine turned off. While the segmentation is not perfect, it is much better than that shown in the first panel. The potential improvement obtained by running an additional segmentation map fixing would therefore not warrant the associated effort. Finally, the last panel shows the segmentation map obtained after all other updates and represents the one that was used for `v03` and `v04`. The main change here was the switch

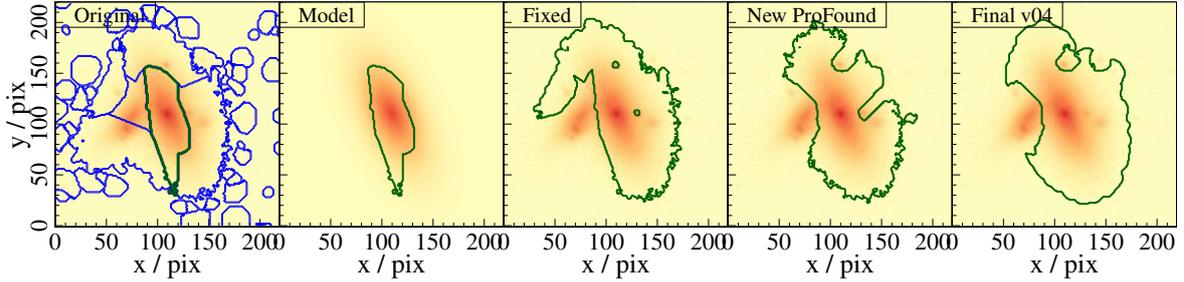


Figure 2.20: An example of the segmentation map fixing for galaxy 544891. **First panel:** the galaxy segment (green) and all neighbouring segments (blue) obtained with the version of ProFound used up to v02 of our pipeline overlaid on the galaxy image. **Second panel:** the model image fitted to the original galaxy segment during the segmentation map fixing procedure. **Third panel:** the galaxy segment obtained after two iterations of our segmentation map fixing procedure again overlaid on the galaxy image. **Fourth panel:** the segmentation obtained with the version of ProFound used from v03 of the pipeline onwards, with all other parameters unchanged with respect to the third panel. **Fifth panel:** the final v03 and v04 segmentation after further updates to the procedure (see text for details). Note that for clarity we do not show neighbouring segments in panels two to five.

from a `skycut` value of one to two and the addition of a further segment dilation after flux convergence is reached (see above). The segment now has a smoother, less biased border than that in panel four, while still being of comparable size.

2.2.5 Modelling decisions

Now that we have all the inputs defined, we move on to galaxy modelling. This is the second major step in the bulge-disk decomposition pipeline and described in some detail in Section 2.1.2 and Casura et al. (submitted). An introduction to Bayesian analysis, the Sérsic function and the most important reasoning for which models we choose is given in Sections 1.3.1 and 1.3.2. Here, we expand on these sections by pointing out details that we did not previously cover. We assume the reader to be familiar with the contents of Sections 1.3.1, 1.3.2 and 2.1.2 before reading this section.

Free fitting parameters

From a model selection perspective, it is advisable to have nested models, where the single Sérsic model is entirely contained within the double component model (i.e. the double component model is an extension of the single Sérsic fit). If this is not the case, then for some galaxies, the single Sérsic fit may be more appropriate in some regions, while the double component fit is better in others, leaving the question of which region of the galaxy is “more important” to fit. For nested models, the double component fit will always be at least as good as the single Sérsic model and the only decision to make remains whether it is significantly

better (i.e. so much better that the increased number of parameters is justified) or not. This can be determined by the Bayes factor (Section 1.3.1) for models that accurately represent all features of the galaxy; or by simulations which capture all of said features; or by visual inspection as we did since we have neither perfect models nor perfect simulations for our diverse sample of galaxies. Having said that, we also considered some versions of non-nested models during pipeline development, as we outline below.

The single Sérsic function in two dimensions has a total of eight parameters. Six of those, namely position in x and y , magnitude, effective radius, position angle and axial ratio definitely need to be fitted for all galaxies in the single Sérsic fit, leaving only the Sérsic index and the **boxyness** for consideration. Since our galaxy sample is diverse, containing classical elliptical galaxies as well as (pure) disk galaxies, double component galaxies and irregulars, we decided to leave the Sérsic index of the single component fit as a free parameter for fitting. This allows to capture all of those different galaxy morphologies in just one fitting round and still leads to rapid convergence of the model parameters.

We briefly experimented with fitting the **boxyness** as well, i.e. allowing deviations from an elliptical shape of the profile in two dimensions, at least for the single Sérsic profile. However, this increased the time needed to achieve convergence by a factor of a few without major improvements in the fit quality determined from visual inspection. In addition, we would then need to allow for **boxyness** in the double component fits as well to keep the models nested, further increasing the fitting time and the problem of parameter degeneracies. Fixing the **boxyness** to zero leaves seven free parameters for the single Sérsic fits, as mentioned previously.

The next complicated one of our three models is the 1.5-component model, which has a total of 11 parameters (eight for the Sérsic function plus three for the point source). For the same reason as above, we did not consider **boxyness** and - as explained below for the double component model - we fixed the bulge and disk positions to lie on top of each other. Leaving the Sérsic index free then results in an extension of the single Sérsic model by just one parameter, namely the point source magnitude. However, the reason why we introduced the 1.5-component model was to fit unresolved bulges for double component galaxies, i.e. it was meant as a simplification of the double component fits and not an extension of the single Sérsic fit. In fact, we originally added the 1.5-component fits after the single and double component fits for v01 of the BDDcomp DMU were already processed, after noticing the $\sim 20\%$ of bulges with unconstrained parameters (very small or unrealistically large effective radii combined with excessively high or low Sérsic indices and sometimes also low axial ratios, often hitting the fit limits in at least one of those parameters; see also Figures 2.21 and 2.22).

To make the 1.5-component fits most directly comparable to the double component fits, we therefore fixed the disk Sérsic index to 1. While it is still not strictly nested in the double

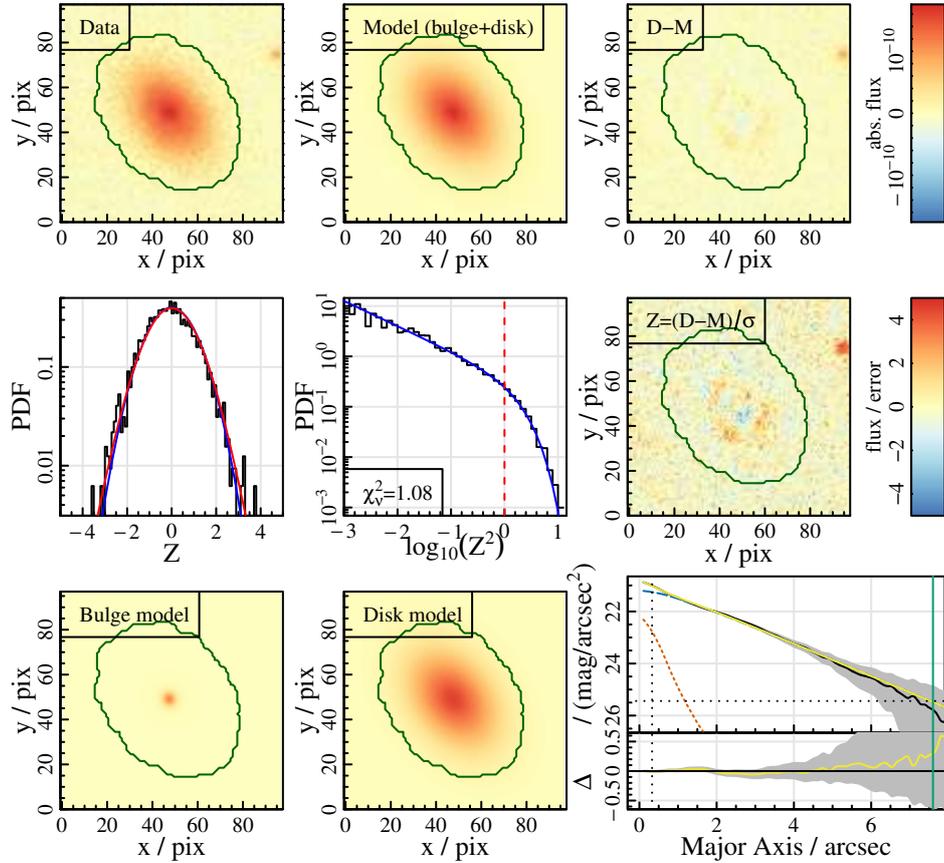


Figure 2.21: An example of a 1.5-component fit for galaxy 124052 in the KiDS r -band. Panels are the same as those in Figure 2.2.

component model, the 1.5-component model in this way simply replaces the Sérsic bulge by a point source bulge (with just one parameter), thus serving its purpose of “saving” the double component fits for which the bulge is ill-constrained. For the same reason, we also left the 1.5-component fits last in the order of processing even though in general it is advisable to start with simpler models and then increase complexity. Processing the 1.5-component fits last has the additional advantage that we need not worry about swapping of the bulge and disk since that has already been taken care of for the doubles (see below) and we can directly use the double component initial guesses. In summary, our 1.5-component model consists of a point source plus an exponential disk, tied to the same position. Note that this means the 1.5-component fit also has just 7 parameters and often results in a worse fit than the single Sérsic model (e.g. for elliptical galaxies that cannot be represented well with an exponential disk).

An example of such a 1.5-component fit is shown in Figure 2.21, with panels the same as in Figure 2.2. For direct comparison, we show the double component fit in Figure 2.22. The two fits are visually indistinguishable and their disk parameters are very similar. However, the

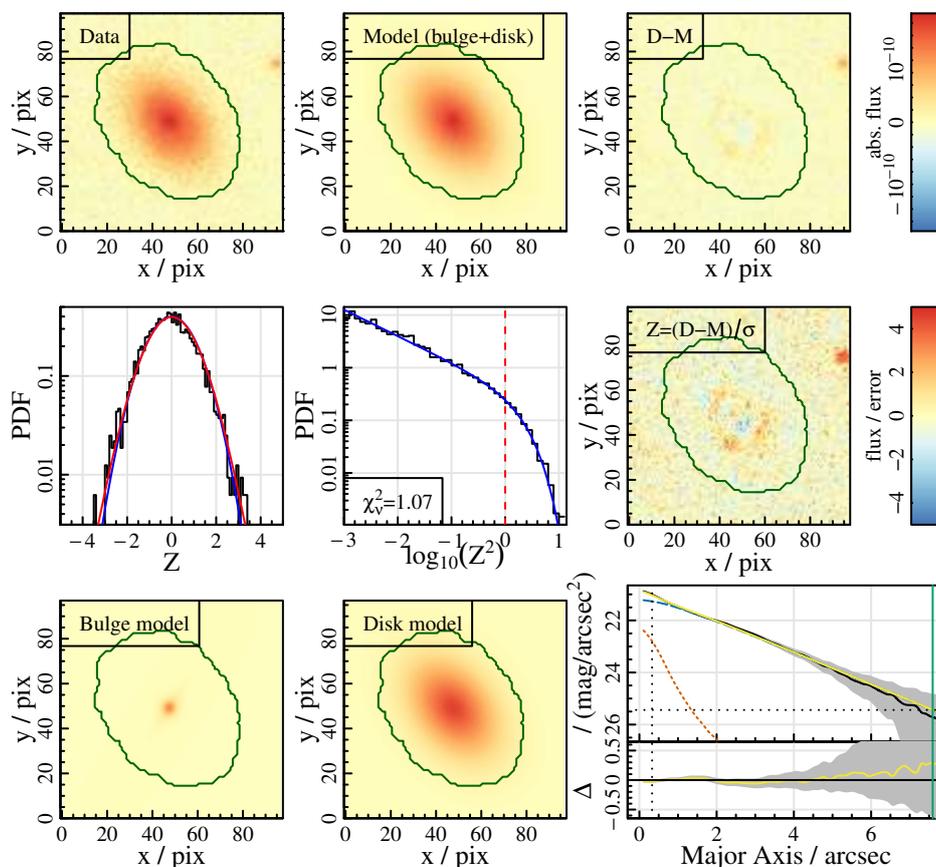


Figure 2.22: The double component fit for galaxy 124052 (classified as 1.5-component object) in the KiDS r -band, for direct comparison to the 1.5-component fit in Figure 2.21.

double component fit has a bulge Sérsic index of 20 and an effective radius of $0.1''$ (0.5 pix), combined with an axial ratio of 0.05, thereby hitting all three of those fit limits. These values (as well as the position angle) are meaningless, but they do influence the fitted magnitude, which is 21.5 mag for the double component model. The point source model does not fit any of the meaningless parameters and instead produces a more robust magnitude estimate of 22.4 mag.

Both the single Sérsic and 1.5-component fits tend to converge rapidly and without problems as there are only few parameter correlations. As explained above, the decisions for which parameters to fit were mostly done on a theoretical basis and straightforward. Most of the remaining discussion focuses on the double component models, where there are many more parameter correlations and degeneracies and therefore also more options to consider in how to limit those.

The double component model has a total of 15 parameters (8 for the Sérsic function and 7 for the exponential). Based on our experience with the single Sérsic fits, we did not consider

boxyness for either of the two components. Another natural choice to make is to tie the bulge and disk positions together due to the symmetry of “typical” galaxies. Nonetheless, we also considered leaving the bulge and disk positions independent from each other to better capture disturbed morphologies. Unfortunately, this frequently led to one of the components wandering off to fit overlapping point sources or wings of other objects, specifically when the main part of the galaxy can be adequately represented by a single component. Conversely, it only significantly increased the fit quality for objects with an irregular morphology that are difficult to represent with our symmetrical models even when allowing offset bulges. Hence, we decided to tie the bulge and disk positions together to lie exactly on top of each other. This leaves a total of 11 free parameters for the double component model, which we also used in the end despite numerous considerations to limit the degrees of freedom further (see below).

Constraints on bulges

Some of the considerations to reduce the degrees of freedom of the double component fit involved tying the position angle of the bulge to that of the disk or forcing the bulge to be round (i.e. setting the axial ratio to 1, or > 0.8 or similar); and/or setting the Sérsic index of the bulge to 4 or limiting it to be larger than, e.g., 2. All of these measures can be used to ensure that the bulge component does indeed fit a classical bulge and not other features such as pseudo-bulges or bars. However, as explained in Section 1.3.2, it is not our aim to fit classical bulges only and instead we use the Sérsic component of the model to explicitly fit all morphological features near the centre of the galaxy. Constraining the bulge to be round or have a high Sérsic index would therefore be counterproductive.

Nonetheless, for versions of the pipeline up to `BDDcomp v02`, we used a lower limit on the bulge Sérsic index of 1. The idea was that the bulge profile is then always steeper than the disk (or of the same steepness), so it is more likely that it will dominate the inner parts of the galaxy and not the outskirts (i.e. to avoid swapped fits). However, there were several problems with that approach. First of all, a significant fraction of bulges converged onto the lower limit ($\sim 18\%$, with a further 33% being consistent with the lower limit within errors, i.e. 51% in total), thereby producing unreliable fits with unrealistic error estimates. Second, a number of pseudo-bulges, bars and broad cores (often galaxies with slightly disturbed morphologies, for which, however, acceptable fits are still achievable) could not be appropriately represented as they showed flattened profiles towards the centre. Third, we imposed the same lower limit on the single Sérsic fits to keep the models nested, which led to $\sim 11\%$ of these converging onto the lower limit as well (with another 13% consistent with the lower limit within errors, so 24% in total). An example of a single Sérsic object with a flattened core (Sérsic index of 0.6) is shown in Figure 2.23. For comparison, Figure 2.24 shows the fit to the same galaxy in

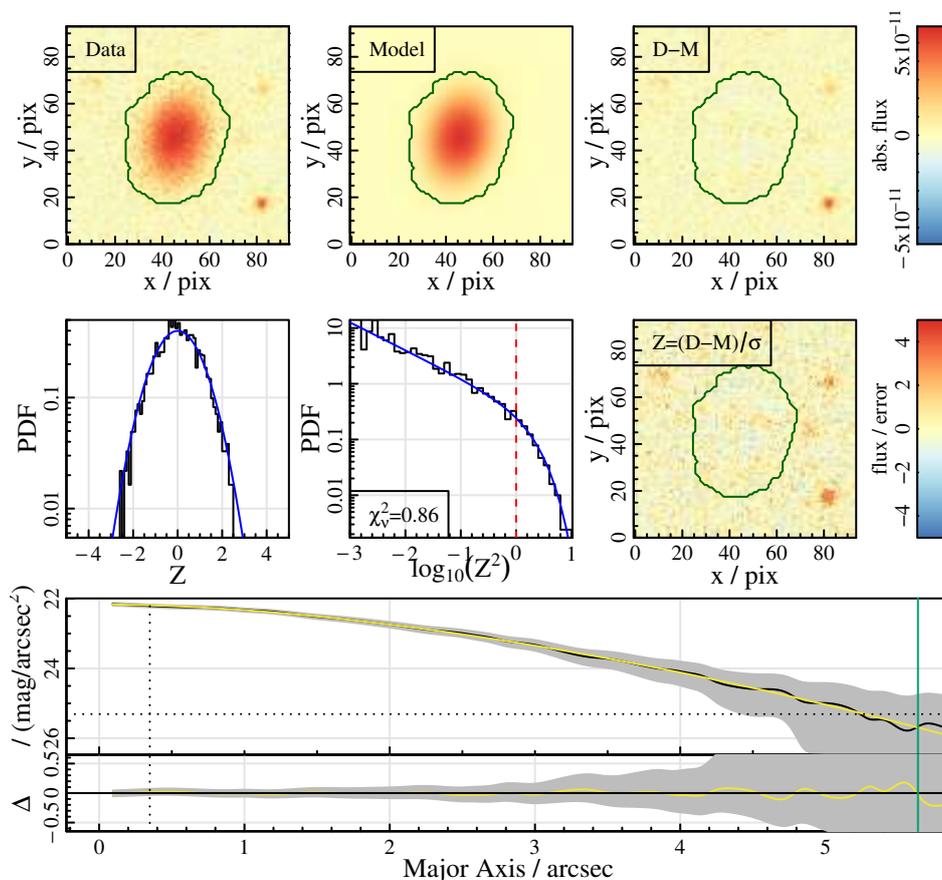


Figure 2.23: An example of a single Sérsic fit with a low Sérsic index (0.6) for galaxy 220084 in the KiDS r -band. Panels in the top two rows are the same as those in Figure 2.2, while the bottom row shows the one-dimensional fit only, corresponding to the rightmost panel of the bottom row in Figure 2.2.

$v01$ of the catalogue, where the Sérsic index converged onto its lower limit of 1 and is clearly not able to capture the profile of the galaxy accurately.

Finally, limiting the bulge Sérsic index to be larger than one only helped in part against swapped fits since the flux ratio of the components is not solely determined by their respective Sérsic indices but also by their relative effective radii, magnitudes and axial ratios; and high Sérsic index bulges in particular can also dominate both the inner and the outer parts of the galaxy at the same time. From $v03$ of the pipeline onwards, we therefore adjusted the Sérsic index lower limit to 0.1 for all models. At the same time, we increased the upper limit on all Sérsic indices from 12 to 20 to avoid fits hitting the limits (even though the difference in the actual profile is small at such high Sérsic indices). This reduced the number of fits hitting either of the Sérsic index limits to 0.3% for the single Sérsic models (including those consistent with the limit within errors), of which approximately two thirds were classified as outliers and the last third as 1.5-component fits. For the double Sérsic models, 15% still

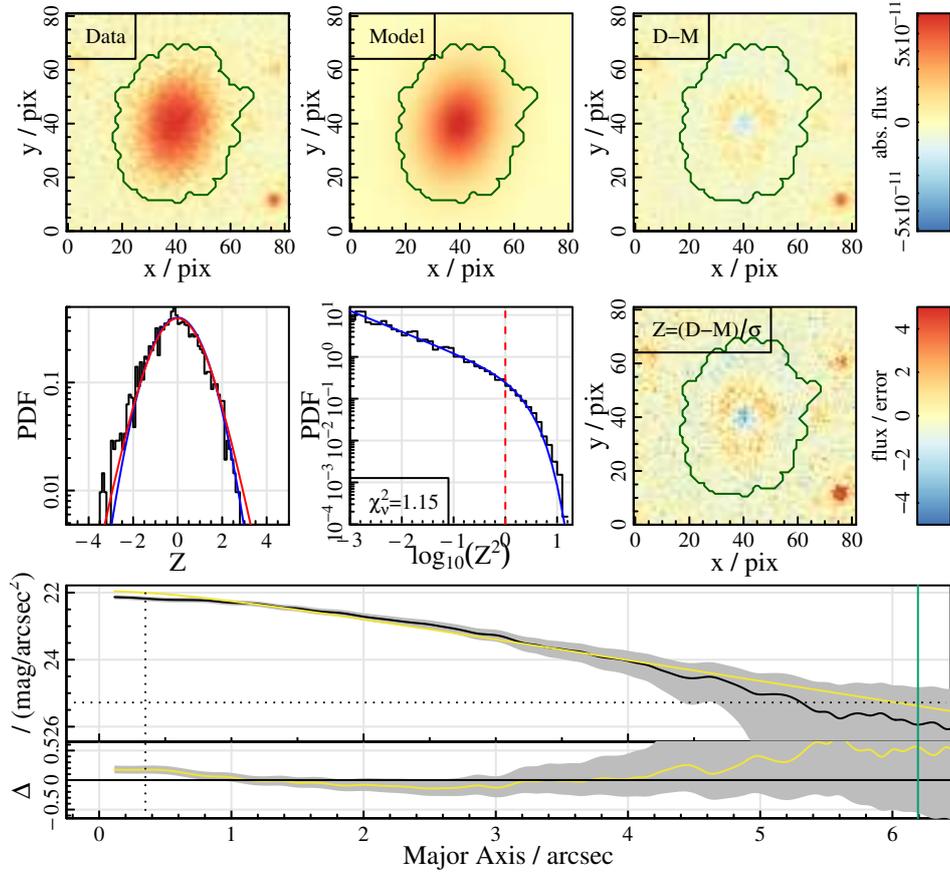


Figure 2.24: The v01 single Sérsic fit to galaxy 220084, where the Sérsic index converged onto its (then) lower limit of unity; for direct comparison to Figure 2.23.

hit their bulge Sérsic index limits, of which 60% were classified as single Sérsic fits, 14% as 1.5-component fits and the remaining 26% as outliers.

In another attempt to stop the bulge and disk components from swapping, we also experimented with limiting the effective radius of the bulge to be smaller than that of the disk. Unfortunately, this did also not have the desired effect, as the precise meaning of the effective radius strongly depends on the Sérsic index and the two parameters (as well as the magnitude) are highly correlated. For example, a bulge with a high Sérsic index can easily dominate the flux in the central parts of the galaxy even if its effective radius is comparable to or larger than that of the disk (with a Sérsic index of 1).

An example is shown in Figure 2.25. The bulge and disk clearly dominate the inner and outer parts of the galaxy respectively. Nonetheless, the bulge effective radius is a factor of more than six larger than that of the disk due to its high Sérsic index of 15.8. This is frequently the case especially since we fit only the pixels in relatively small segments, so the extended

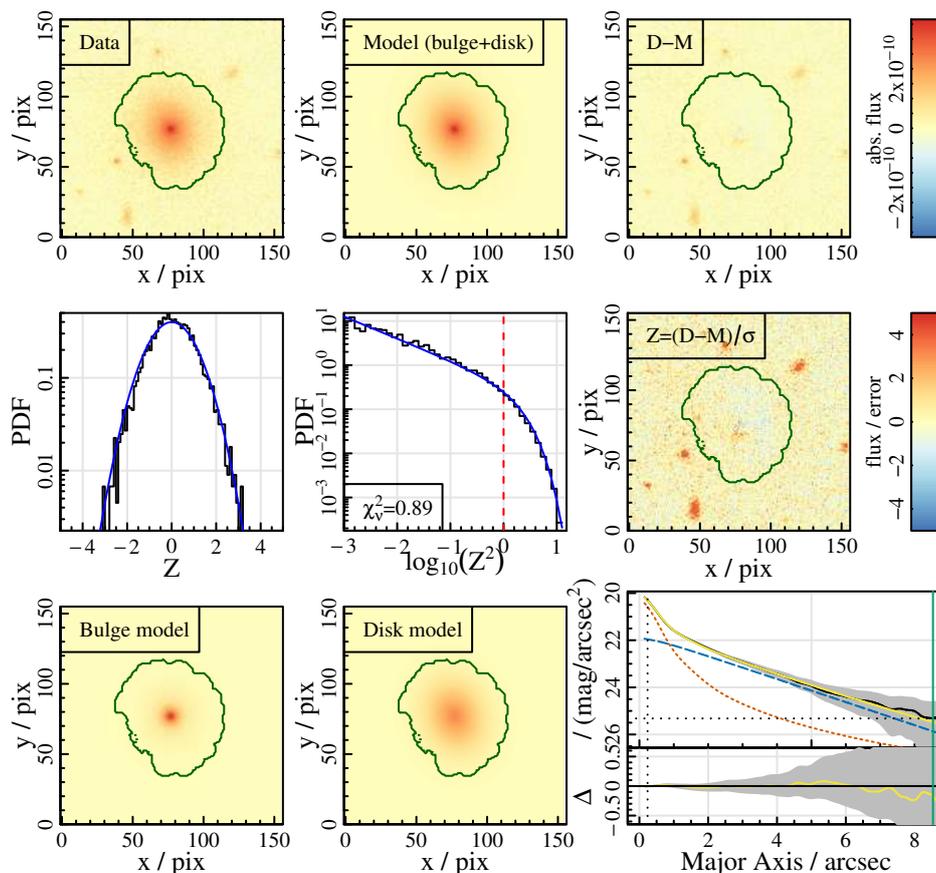


Figure 2.25: The double component fit to galaxy 136866, where the bulge effective radius is much larger than that of the disk. Panels are the same as in Figure 2.2

wings of a high n Sérsic function fall beyond the fitting region and are effectively ignored.⁶

Given the limited success of our attempts to avoid swapped components by constraining the bulge parameters, we have instead devised a separate routine to swap the components of galaxies where necessary, see Sections 2.1.2 and 2.2.7.

Priors and initial guesses

One other route towards influencing the fit without imposing hard cuts is by using differing initial guesses, as we do in our swapping routine (Section 2.1.2). However, one of the main strengths of an MCMC analysis is that it is less dependent on initial guesses than e.g. downhill gradient fitting. When starting very close to a local maximum (as we do for the swapped fits), the fit is more likely to converge onto that. Similarly, when starting very far from any

⁶On a side note, when limiting the effective radii to the segment values, the disk R_e is a factor of about 2 larger than that of the bulge. However, since we truncate to segment radii in a post-processing step, we cannot use this during the fit itself.

2.2. PIPELINE DEVELOPMENT

maximum (i.e. in the flat part of the likelihood space), convergence takes much longer and fails more frequently. For any “reasonable” initial guesses, however, the fit result will not depend strongly on them. We obtain such reasonable initial guesses from the segmentation output of `profoundProFound` and further improve them via a fast downhill gradient algorithm before starting the MCMC (see Section 2.2.6). The details of, e.g., how to convert the `profoundProFound` concentration to a Sérsic index initial guess, how to split the total flux between the bulge and disk initially, what ratio of effective radii to use and the starting value of the bulge Sérsic index were the subject of several investigations but were found to have little effect on the fit results.

Arguably the most “Bayesian” way to constrain parameters is via priors. These can be hard cuts - like the fixed intervals that we use for fitting parameters - or soft, like for example Gaussian functions of a certain width centred on the most likely value. However, when imposing such informative priors, they should be based on previous knowledge (e.g. from previous data) rather than general trends or notions since they will directly affect the posterior and can even dominate over the likelihood if the prior is strong and/or the likelihood is weak due to a limited quality or quantity of data. To get the most unbiased estimate of the parameters based on the data alone it is therefore wise to choose uninformative, broad priors.

We do this by imposing fixed limits, but using flat priors within these limits (in logarithmic space where appropriate). The limits are listed in Table 2.2 and are deliberately very broad. Most of them are based on physical considerations (e.g. the axial ratio cannot be larger than 1) or limitations of the data (e.g. if the effective radius is less than 0.5 pix, more than half of the total light of the profile is contained within a single pixel, at which point the accuracy of the PSF is very likely not sufficient anymore). The only exception to this is the Sérsic index, for which we discuss the limits in some detail above.

Note that as we mention in Section 2.1.2, we do not normalise our priors, leading to unnormalised posteriors. In theory, normalised posteriors are important to obtain meaningful Bayes’ factors for model selection. In practice, we found that in our case even with normalised posteriors the model selection needs manual calibration due to the model inadequacy. Leaving the priors unnormalised has the advantage that the prior ranges do not influence the posterior and so we do not need to re-calibrate the model selection every time for different test runs with different prior intervals. It also allows to easily put infinity as the upper limit on component magnitudes and instead constrain the total magnitude of the double component fit, which otherwise would require a somewhat non-trivial joint prior to be normalisable.

2.2.6 Fitting specifics

After the models are defined, there are a number of decisions to be made with regards to the actual fitting. These are summarised in Section 2.1.2. In the following, we point out additional details regarding the logarithmic fitting of scale parameters, the choice of likelihood function, the fitting algorithms and routine and the assessment of convergence.

Logarithmic fitting

From a theoretical point of view, scale parameters (in our case those are the effective radius, Sérsic index and axial ratio) are best fitted in logarithmic space, while location parameters (position, magnitude and angle) should be fitted in linear space. This is because a “step size” for location parameters remains constant across the entire parameter range, so, e.g., the difference between a position angle of 5° and 10° is the same as that between 170° and 175° . For scale parameters, the step size changes across the parameter range (in linear space), for example the difference in profile shape between a Sérsic index of 0.5 or 1 (Gaussian or exponential) is much larger than that between a Sérsic index of 18.5 or 19 (nearly indistinguishable). Converting scale parameters into logarithmic space, i.e. fitting $\log_{10}(X)$ instead of X for scale parameters X , equalises the step size across the (now logarithmic) parameter range again.⁷ Note that intensity or brightness are also scale parameters, but since magnitude is already a logarithmic measure of flux, it becomes a location parameter.⁸

Given this theoretical background and the functionality of `ProFit` to easily specify which parameters should be fitted in logarithmic space (with conversions to and from logarithmic space performed internally), it seemed natural to fit scale parameters in logarithmic space. Nevertheless, we tested the effect of fitting in linear space for one or several of the scale parameters during our test runs. Fitting all parameters in linear space approximately doubled the computational time needed for single Sérsic fits compared to logarithmic fitting of all scale parameters, while no systematic differences in the estimated parameters could be observed.

The difference in computational time highlights the importance of choosing similar step sizes in all parameters for rapid convergence. With the logarithmic fitting of scale parameters and our chosen fitting intervals, we arrive at comparable step sizes for all parameters with one exception: the position angle. To ensure convergence of the position angle we therefore (linearly) re-scaled this parameter into units of 30° . All fitting parameters then have similar ranges in absolute terms (in their respective, potentially logarithmic, fitting units), as also evident from Figure 4.9 (the absolute values of all parameters vary in a range of about ± 0.05 around the true value; and the absolute errors are on comparable scales for all parameters).

⁷For the same reason, relative errors are more meaningful than absolute errors for scale parameters; and vice versa for location parameters.

⁸Hence why colours are calculated as magnitude differences rather than ratios.

Likelihood function

In Bayesian modelling, the probability of the observed data given an (assumed) model is called the likelihood (Section 1.3.1). Calculating it requires knowing or assuming a likelihood function that takes the data and model as input and returns the corresponding probability. This can in theory be any function, but is most often a “standard” function appropriate to the statistical problem at hand, such as a Gaussian, chi-squared, Poisson or Student’s t-distribution, all of which are supported by `ProFit`.

KiDS data is expected to have random Normal errors with pixel-specific standard deviations expressed in the sigma map (Section 2.1.1), which in turn is a combination of the KiDS weight maps and the object shot noise. For a perfect model, the log-likelihood can therefore simply be calculated as the sum of the Gaussian probabilities to obtain each measured flux value given the model flux value of the corresponding pixel and its error from the sigma map (`ProFit` does this internally). The correct likelihood function to use is therefore the Normal one, provided the model is a perfect fit to the data.

To accommodate models that are not perfect fits to the data at hand (i.e. most of our models for the majority of our galaxies), `ProFit` supports the use of a Student’s t-likelihood instead. This evaluates the probability of the data given the model and errors under the assumption of a Student’s t-distribution rather than a Gaussian one, optimising the degrees of freedom of the Student’s t-distribution at the same time. Due to the broader wings of the Student’s t-distribution, this effectively downweights the regions of the galaxy that cannot be captured by the model during the fit. This makes it more robust for the purpose of galaxy fitting.

As an example, we show the single Sérsic fit to galaxy 177705 using a Student’s t-likelihood in Figure 2.26 and the fit to the same galaxy with a Normal likelihood in Figure 2.27. This galaxy needs two components to be accurately represented. Fitting it with just a single Sérsic function gives different results for the two likelihood functions: the Student’s t-likelihood is more likely to “tolerate” a few pixels that are in high tension with the model (i.e. pixels belonging to the bulge) if this allows a better fit on average to the remainder of the galaxy. A normal likelihood instead “tries harder” to avoid strong outliers and compromises more between fitting both components, even if that results in a generally slightly worse fit over larger areas such as the disk. While in the example here it is debatable which option is to be preferred, a Student’s t-distribution is definitely preferable if the pixels in high tension are - e.g. - an overlapping foreground point source or a clumpy component of a star-forming spiral arm.

Since our models are generally a simplification of the true complexity of galaxies, we would expect a Student’s t-distribution to be more suitable to our analysis. This was therefore the default likelihood we used initially. Unfortunately, though, it turned out to have several

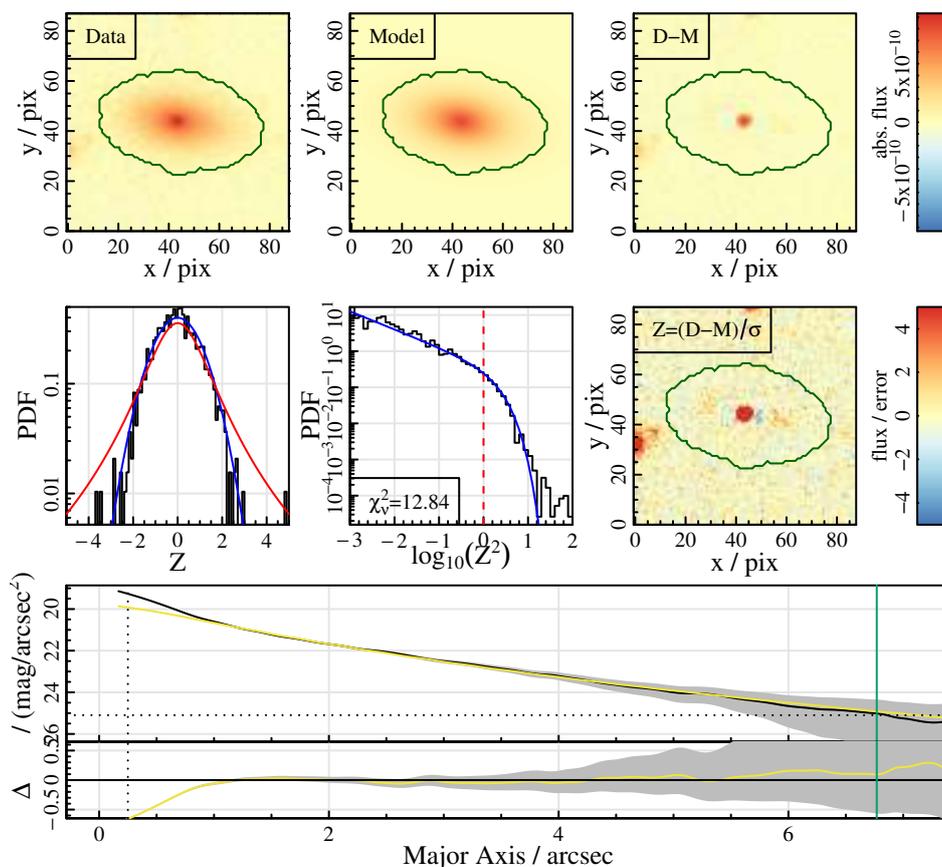


Figure 2.26: The Student’s t -likelihood function single Sérsic fit to galaxy 177705 (a double component object) in the KiDS r -band. Panels in the top two rows are the same as those in Figure 2.2, while the bottom row shows the one-dimensional fit only, corresponding to the rightmost panel of the bottom row in Figure 2.2.

problems as already outlined in Section 2.1.2. For the final pipeline, we therefore use a Normal likelihood function for all galaxies. Since this decision was made entirely during the test runs, the BDDcomp DMU has only Normal likelihood function fits from v01 onwards already. However, we still like to point out the development over the test runs, since this was a major point of concern and one aspect in which the final decision does not meet the theoretical expectations.

The first problem of using Student’s t -distribution likelihood functions became apparent for galaxies that are perfectly represented by their respective model. In this case, as explained above, the errors truly are distributed Normally. The Student’s t -distribution, however, “expects” a certain fraction of pixels that are in tension with the model and - if they are not present - will artificially produce them by making the fit unnecessarily worse. An example for this is shown in Figures 2.28 and 2.29 for galaxy 534802. This object can be perfectly represented by a single Sérsic component, as evident from the fit with a Normal likelihood

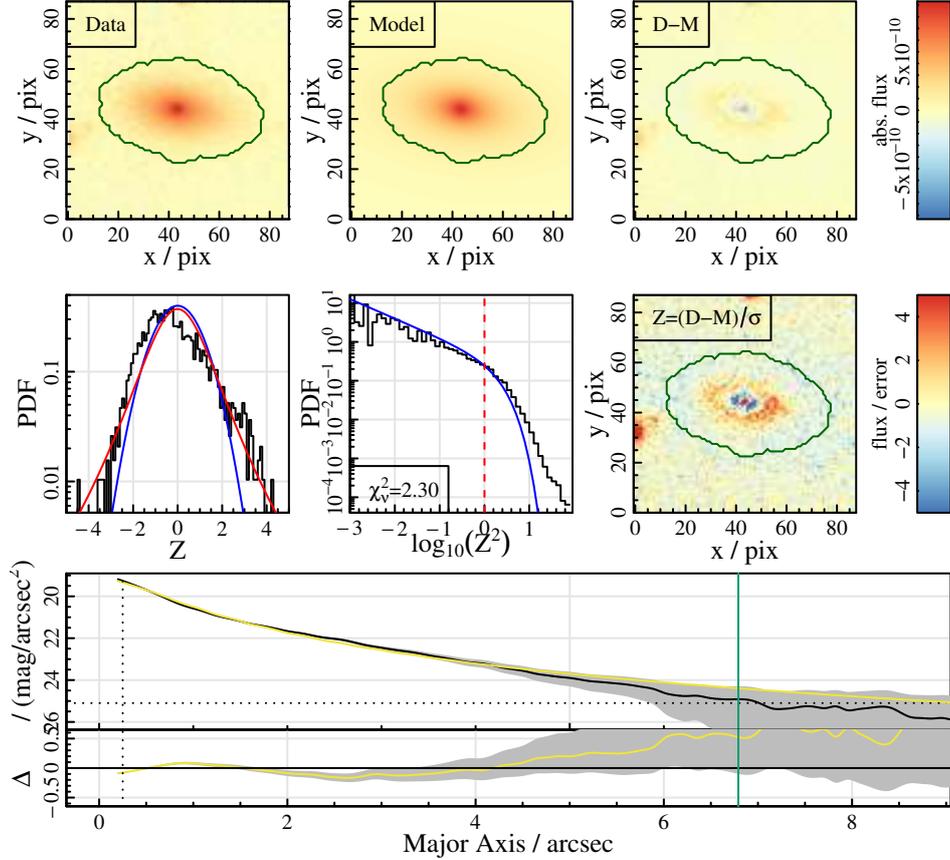


Figure 2.27: The Normal likelihood function single Sérsic fit to galaxy 177705, for direct comparison to Figure 2.26.

function. The fit with the Student’s t-likelihood, though, shows considerable residuals.

We concluded that fitting all galaxies with a Student’s t-likelihood is unsuitable and instead a certain fraction, namely those approximately 20% of galaxies that are perfectly captured by one of our models, need a Normal likelihood function. The correct way to decide which likelihood function to use would be to fit each galaxy twice (once with each likelihood function) and then choose the one which achieves the better fit, i.e. the higher likelihood. However, we noticed that the degrees of freedom estimated for the Student’s t-likelihood during the fit can provide a shortcut to save computational time.

As the degrees of freedom of the Student’s t-function approach infinity, its distribution approaches that of a Gaussian. It is therefore not surprising that the degrees of freedom converge onto a large value for those galaxies that would be better fitted by a Normal likelihood. First fitting all galaxies with a Student’s t-likelihood therefore allows to select only a subset of galaxies for re-fitting with a Normal likelihood based on a cut in the degrees of freedom of the Student’s t-function.

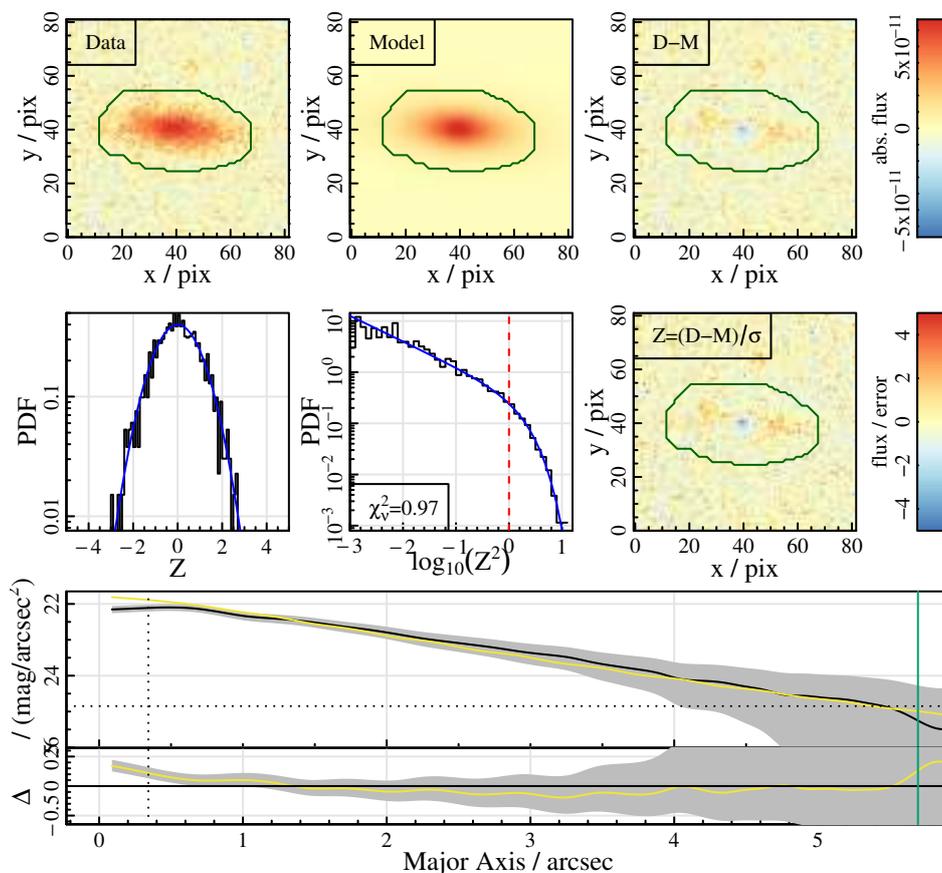


Figure 2.28: The Student’s t-likelihood function single Sérsic fit to galaxy 534802, which is well-represented by a single Sérsic profile, in the KiDS r -band. Panels in the top two rows are the same as those in Figure 2.2, while the bottom row shows the one-dimensional fit only, corresponding to the rightmost panel of the bottom row in Figure 2.2.

We demonstrate this in Figure 2.30 for a test sample of KiDS r -band galaxies that we fitted with both likelihood functions. The solid black histogram shows the degrees of freedom of the Student’s t-distribution for all galaxies in the test sample. There is a clear peak encompassing the majority of galaxies at low values of degrees of freedom and a long tail to very large values (note the logarithmic scaling of both axes). With a dashed orange line, we overplot the subset of galaxies that are better fitted with a Student’s t-likelihood, i.e. those that achieve a higher likelihood with a Student’s t-likelihood than with a Normal one. This encompasses essentially all of the galaxies with low degrees of freedom and none of those with high values (for which a Normal likelihood function achieves a better fit).

The dotted blue line indicates our selection cut for deciding which galaxies to re-fit with a Normal likelihood function, namely those for which the degrees of freedom of the Student’s t-likelihood function fit converged to values larger than 100. Note that this cut is deliberately more to the left, since galaxies to the left of this line will only be fitted with a Student’s t-

2.2. PIPELINE DEVELOPMENT

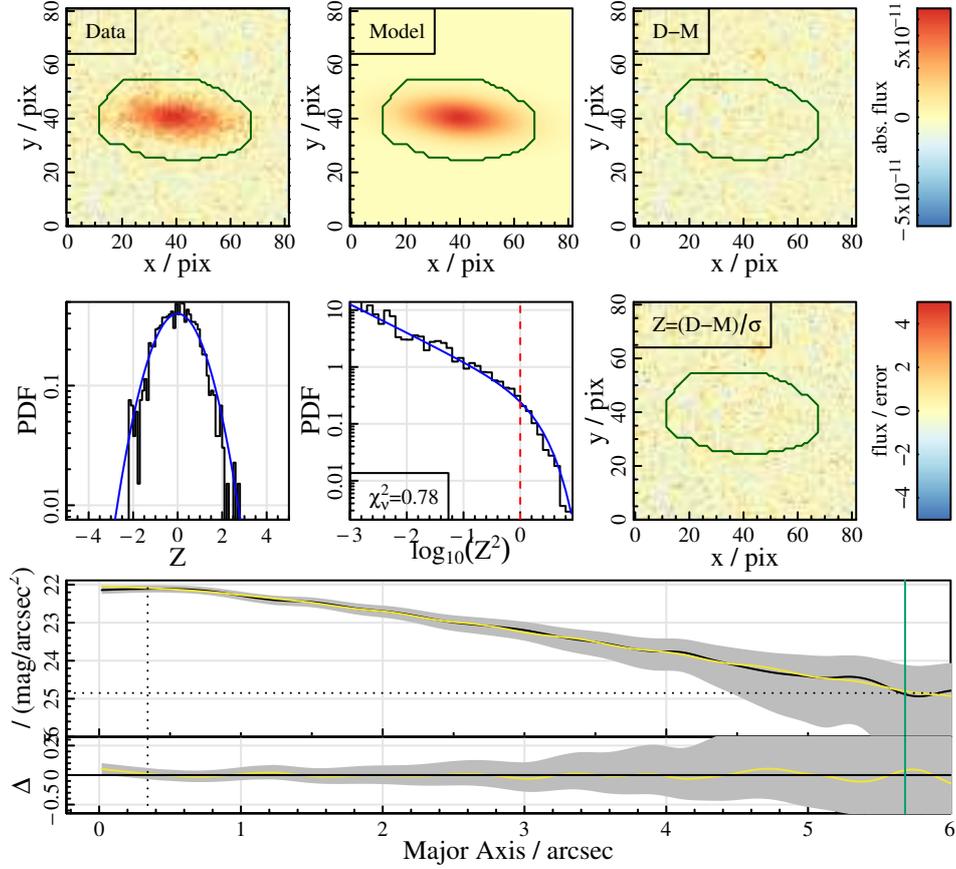


Figure 2.29: The Normal likelihood function single Sérsic fit to galaxy 534802, for direct comparison to Figure 2.28.

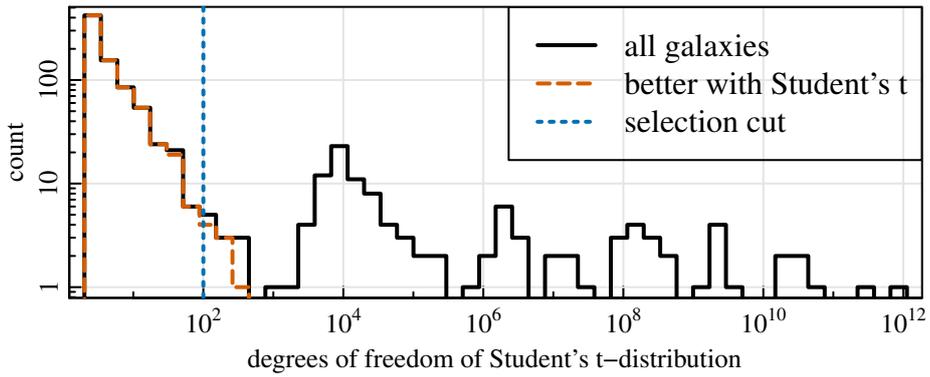


Figure 2.30: The distribution of the Student's t -distribution degrees of freedom for a set of test galaxies (solid black line) and for the subset of those that achieves a better fit with a Student's t -likelihood compared to a Normal likelihood (dashed orange line). The dotted blue line shows the selection cut for deciding which likelihood function to use.

distribution (therefore missing the best fit for any galaxies in that sample for which a Normal likelihood would be better suited), while those to the right of the line will be fitted with both likelihoods such that the optimal likelihood can still be chosen based upon which achieves the better fit (therefore only wasting a small amount of computational time on the galaxies in that sample for which a Student's t-likelihood is better suited).

These distributions look similar for double component fits and in other bands, although the relative numbers of galaxies which are better fitted with a normal or a t-distribution likelihood function changes due to the different depths and seeing. In summary, at early stages during pipeline development, we first fitted all galaxies using a Student's t-likelihood function. Then, galaxies for which the degrees of freedom of the Student's t-distribution converged onto values larger than 100 were fitted again with a Normal likelihood function (for all models independently). For all galaxies fitted twice with the same model, we selected the better fit according to which achieved the higher likelihood.

The exact order of processing (including the segmentation map fixing steps used in earlier versions of the pipeline, cf. Section 2.2.4) was as follows:⁹

- (i) Run the galaxies through the preparatory work pipeline, including downhill-gradient single Sérsic fits for the segmentation map fixing.
- (ii) Do single component Student's t-likelihood function fits on all galaxies using the output of the preparatory work as initial guesses (more precisely the fits from the segmentation fixing) and the corresponding segmentation maps. Check whether segmentation maps need fixing after the fit and iterate if needed.
- (iii) Do double component Student's t-likelihood function fits on all galaxies now using the outputs of the previous step as initial guesses and also using the updated segmentation map. Again check whether the segmentation maps need updating and iterate if needed.
- (iv) Re-run the single component Student's t-likelihood function fits for the galaxies for which the segmentation map changed again during the double component fits. This uses the output of the previous single component fit as initial guess, but takes the updated segmentation map (after the double component fit) to make all segmentation maps used consistent for model selection. Do not change the segmentation maps anymore in this run.
- (v) Run single component Normal likelihood function fits for the galaxies for which the degrees of freedom of the single component Student's t-likelihood fit were greater than 100 or the fit failed (which happened in a number of cases when the degrees of freedom

⁹We subsequently extended this procedure to include 1.5-component fits; and then simplified it again by first dropping the Student's t-likelihood fits and later also the segmentation map fixing; see Section 2.1.2 for the final v04 procedure.

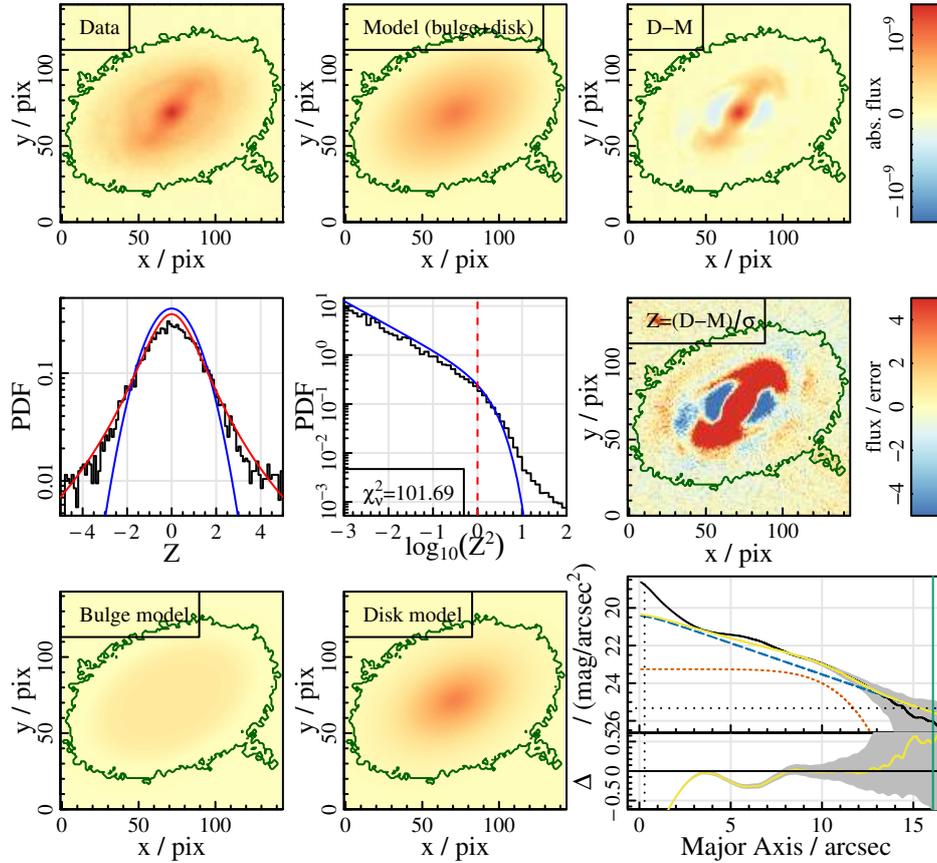


Figure 2.31: The Student’s t-likelihood function double component fit to galaxy 595088 in the KiDS r -band. Panels are the same as those in Figure 2.2.

reached infinity). Use the previous single component fits as initial guesses and do not change the segmentation maps.

- (vi) Run double component Normal likelihood function fits for the galaxies for which the degrees of freedom of the double component Student’s t-likelihood fit were greater than 100 or the double component fit failed. Use the previous double component fits as initial guesses and do not change the segmentation maps.

- (vii) Repeat these steps for all desired bands.

This procedure solved the first problem that we encountered using Student’s t-likelihood functions (the unnecessary residuals for perfect fits). There was, however, one other major problem which eventually prompted us to use Normal likelihood functions for all fits: the Student’s t-likelihood function fits frequently missed the bulges in double component fits.

An example of this is shown in Figures 2.31 and 2.32 for galaxy 595088. This is a bright and well-resolved object with clear evidence of a bulge, bar, disk, ring and spiral arms. Fitting

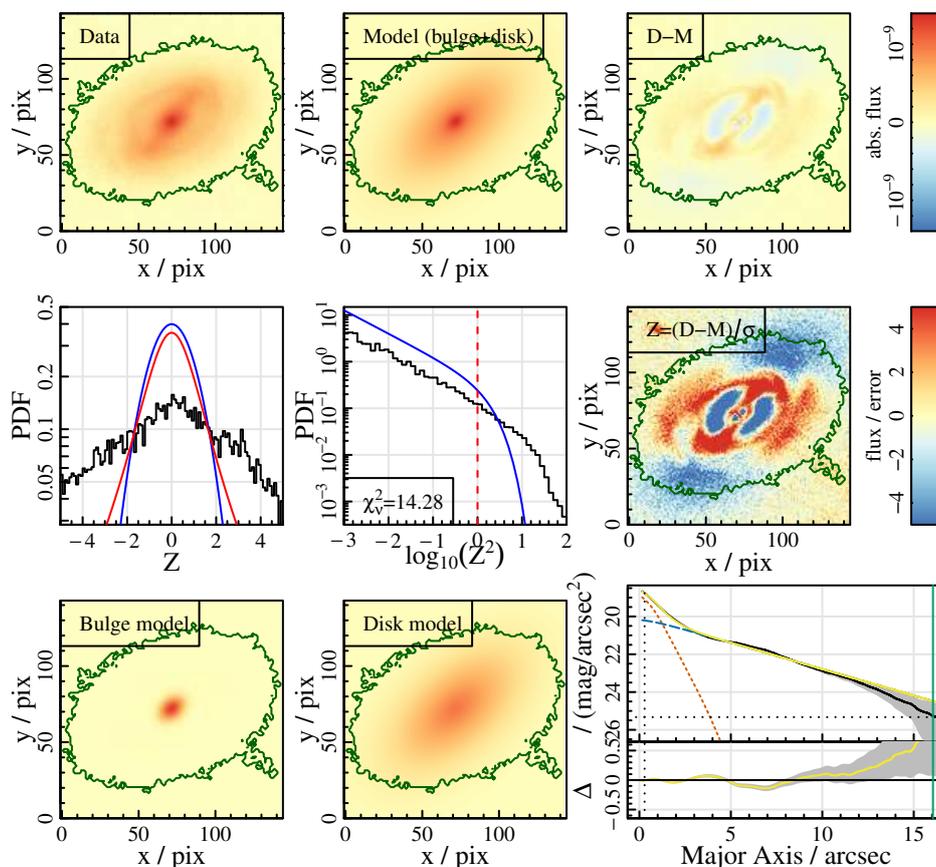


Figure 2.32: The Normal likelihood function double component fit to galaxy 595088, for direct comparison to Figure 2.31.

only two components will not allow to capture all of these details, and the results obtained with the two different likelihood functions differ greatly. Similar to the single Sérsic example in Figures 2.26 and 2.27, the Student’s t -likelihood fit ignores the bulge and instead uses the freedom of the Sérsic component to fit disk features that cannot be captured by the exponential model. This is because the Student’s t -function “prefers” a few strong outliers over many weak ones, which can be an advantage as explained above. In this case, though, this behaviour is highly undesirable since we explicitly intend the Sérsic component of the double component fit to represent the central regions of the galaxy and not to capture deviations of the disk from an exponential profile. The Normal likelihood function fit instead fits both components, as desired.

Many of our experiments with initial guesses, priors and constraints on bulge parameters (Section 2.2.5) were attempts to recover those “missing” bulges. As explained above, they were not successful and in the end, the solution was to use a Normal likelihood function for all fits. This is therefore what we do in the current pipeline (and earlier versions including v01 of the BDDecom DMU). It has the added advantage that each model only needs to be fitted

with one likelihood function, thereby saving computational time. In addition, it facilitates the manual calibration of the model selection, which is non-trivial if different galaxies and potentially also different models for the same galaxy use different likelihood functions.

Optimisation algorithm

One of the main assets of `ProFit` (cf. Section 1.3.3) is that its pixel integration is faster and more accurate than that of other commonly used algorithms, which allows us to use a more robust but also more computationally expensive MCMC algorithm instead of simple downhill gradient optimisers even for our relatively large sample of galaxies. As briefly mentioned before, `ProFit` allows great flexibility and can be combined with a variety of optimisation algorithms, including more than 60 variants of MCMC from the `LaplacesDemon` package as well as various downhill gradient optimisers and genetic algorithms. While we definitely want to make use of the advantages of the MCMC options, the questions remained whether it would be best to combine this with some of the other options, which MCMC variant(s) to use and how to ensure and assess convergence.

Instead of devising our own routine, we opted to use the `convergeFit` function from the `All-StarFit` package (Taranu, 2022), which has already been optimised to achieve convergence within reasonable computational times by Taranu et al. (2017). As briefly mentioned in Section 2.1.2, this function uses a combination of different downhill gradient optimisers followed by several MCMC fits until its internal convergence criterion is met.

The downhill gradient optimisers are taken from the `nloptr` package (Johnson, 2017) and are used first to improve the initial guesses with little computational effort. The MCMC chain is not very sensitive to the initial guesses, but converges much faster if starting close to the peak of the likelihood. In most cases, the downhill gradient algorithm will improve the initial guesses by moving closer to the peak of the likelihood, saving computational time. There can be cases where the downhill gradient moves in the “wrong” direction and/or becomes stuck in a local maximum, in which case it may then take the MCMC chain slightly longer to leave the local maximum and find the globally best fit. However, since this only happens in a minority of cases and the time loss is small compared to the time savings in successful cases, we still save large amounts of computational time overall. Apart from that, even an MCMC algorithm would tend to first explore the local maximum if starting close to it, so skipping the downhill gradient step may not actually shorten fitting times for galaxies with initial guesses close to local maxima.

After the downhill gradient algorithms have converged, `convergeFit` starts an MCMC chain with 500 iterations using the Hit-and-Run-Metropolis (HARM)¹⁰ algorithm in the `Laplaces-`

¹⁰Both HARM and CHARM are variants of the Hit-And-Run algorithm introduced by Turchin (1971).

`Demon` package (Statisticat & LLC., 2018). This process is repeated until convergence is reached, where convergence is defined as the fractional change in the log-likelihood between two runs being less than ϵ . We compare this criterion to other measures of convergence and stationarity below.

As a last step, the Componentwise Hit-and-Run-Metropolis (CHARM), also from `Laplaces-Demon`, is run with 2000 iterations and again repeated until convergence (usually only once since the chains have converged already using HARM). This algorithm is more robust but also a lot slower than HARM, hence it is only called in the end to collect likelihood samples around the peak which has already been found in previous steps. For further analysis of the galaxy it is then assumed that all samples returned by CHARM are stationary, see below for a test of this assumption.

On average, for the r -band fits of v04 of the `BDDecomp` DMU, the entire fitting procedure takes 7.6 minutes per galaxy for the single Sérsic fits, 28.9 minutes for the initial double component fits, 27.5 minutes for the swapped double component fits (only performed on approximately one third of all galaxies) and 6.0 minutes for the 1.5-component fits. In total, the computational time to fit all models and including the preparatory work and post-processing is just under one hour per galaxy, although with strong variations for individual objects depending on their size and complexity (i.e. how rapidly convergence can be reached) and whether or not they enter the second round of double component fits for swapping.

We also tested just using 10000 CHARM iterations, without any of the downhill gradient or HARM steps and simply assuming convergence is reached after such a large number of iterations. This is the approach taken in some other works based on `ProFit` fitting of galaxies, e.g. Cook et al. (2019) and Hashemizadeh et al. (2022). Compared to the `convergeFit` function, this procedure drastically increased the computational time for the vast majority of our galaxies (by a factor of about four on average) with no systematic differences in the fit results and no improvements to the convergence or stationarity criteria.

MCMC convergence

Judging the convergence (or equivalently, stationarity) of a chain is one of the main challenges in MCMC analysis with no universal solution.¹¹ Hence, all algorithms require user input regarding the stopping criterion; in the HARM and CHARM algorithms this is simply the number of iterations to perform. However, depending on the galaxy morphology, the number of iterations required to achieve chain convergence can be vastly different, so to achieve stationary results for all galaxies of a large sample while still retaining some computational efficiency, a fixed number of iterations is not suitable and a more flexible stopping criterion

¹¹We use the terms “convergence” and “stationarity” interchangeably here, although in detail they are not exactly the same.

2.2. PIPELINE DEVELOPMENT

is needed. In the `convergeFit` function, this is implemented via repeated chains of 500 (or 2000 for CHARM) iterations each, which are judged for convergence comparing to the previous batch of 500 (2000) iterations: a chain is considered stationary if the fractional change in the log-likelihood obtained from the current batch is less than e compared to the previous batch. The threshold (e) is somewhat arbitrary and has been chosen based on visual inspection (Dan Taranu, private communication). According to this criterion, all of our fits have converged.

There are many other ways to assess stationarity, such as the BMK diagnostic (Boone et al., 2014), the Geweke Diagnostic (Geweke et al., 1992), the Heidelberger Diagnostic (Heidelberger & Welch, 1981, 1983), or the Kolmogorov-Smirnov Convergence Diagnostic (Brooks et al., 2003) to name just a few of the tools available in the `LaplacesDemon` package. All of them have in common that they are ultimately based on an arbitrary threshold of some form; and that there can be difficulties with multivariate chains, especially if correlations between parameters exist (which they do in our case, in particular the magnitude, Sérsic index and effective radius are generally (anti-)correlated). Nonetheless, we investigated the first two of those in more detail.

The BMK diagnostic is based on calculating the Hellinger distances (Hellinger, 1909) between consecutive batches of the chain and is automatically computed when running `LaplacesDemon`. According to this criterion, 29% of our r -band single component fits are not converged; and the majority of the double component fits. The Geweke diagnostic looks for trends or changes in moments in the given samples to assess stationarity and can easily be computed from the MCMC chain. According to this, 87% of the r -band single component fits (and almost all double component fits) are non-stationary. Cross-comparing the two criteria shows that 30% of the Geweke-non-stationary single component fits are not converged according to the BMK diagnostic; and vice versa 89% of the BMK-non-converged single component fits are not stationary according to the Geweke diagnostic. These fractions are very similar to the fractions in the full sample (30% vs. 29% and 89% vs. 87%), so the two criteria seem only marginally (if at all) correlated. Also, the percentages are very similar when increasing the number of iterations; when re-fitting the galaxies using the previous fits as initial guesses and when using simulated galaxies for which we know that parameters are well-recovered (cf. Section 4.2).

In addition, we visually inspected some of the chains that were flagged as converged/stationary by both criteria as well as some that are non-converged/non-stationary according to both criteria and could not see an obvious difference between the two samples. As a demonstration, we show the chains for the single component fits for two example galaxies in Figures 2.33 and 2.34, where the chain that is not converged according to either criterion visually appears to be more stable than the one that is converged according to both criteria.

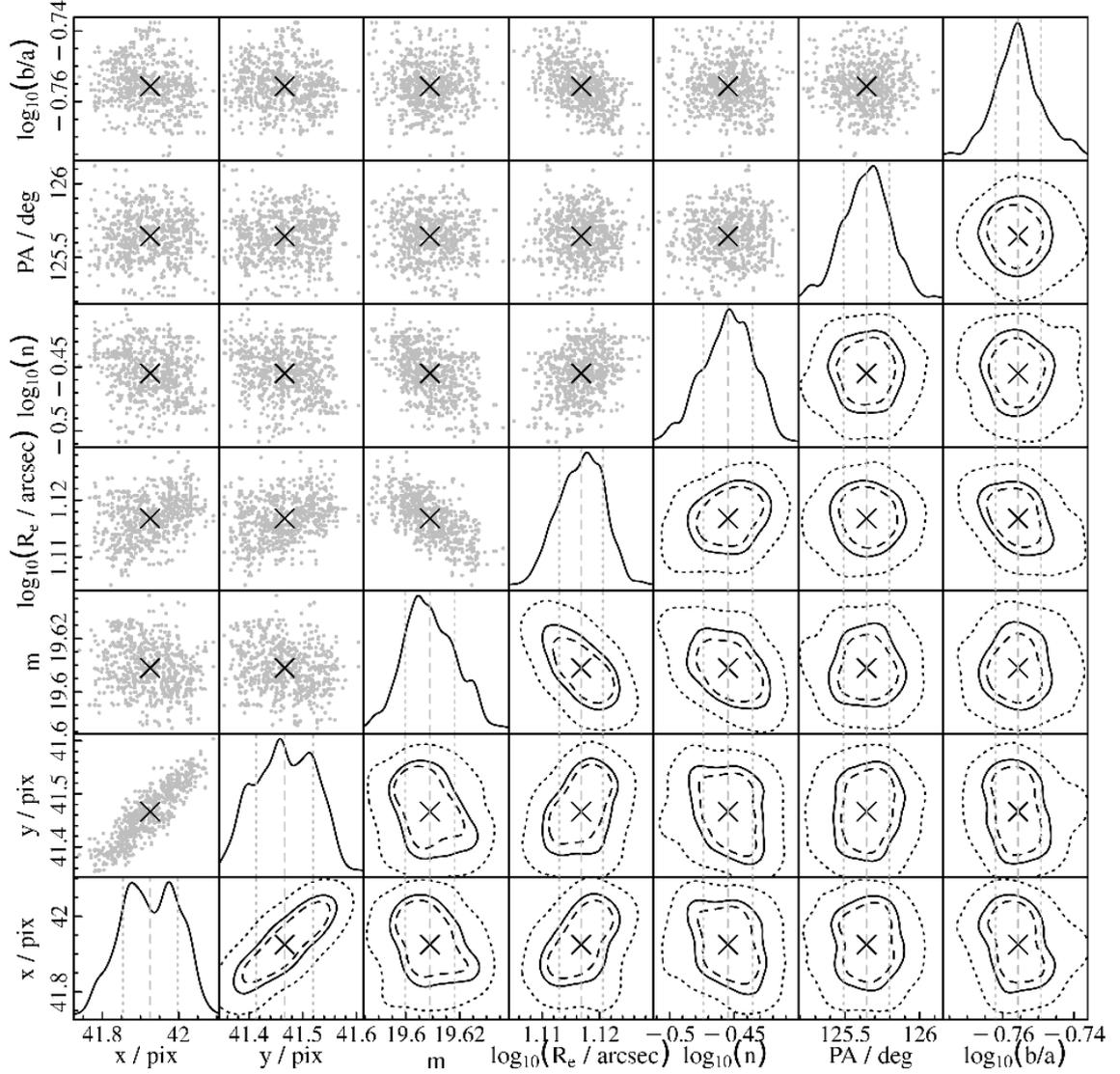


Figure 2.33: The MCMC chain for the r -band single component fit of galaxy 136524 that is neither converged according to the BMK diagnostic nor stationary according to the Geweke diagnostic. Panels are the same as in Figure 2.16.

For all of these reasons, we decided to keep the default convergence criterion used in the `convergeFit` function rather than implementing a more complicated diagnostic with no obvious advantages. We assume the last 2000 CHARM samples are sufficiently stationary for our needs based on the visual inspection of a number of example chains and the simulations (Section 4.2).

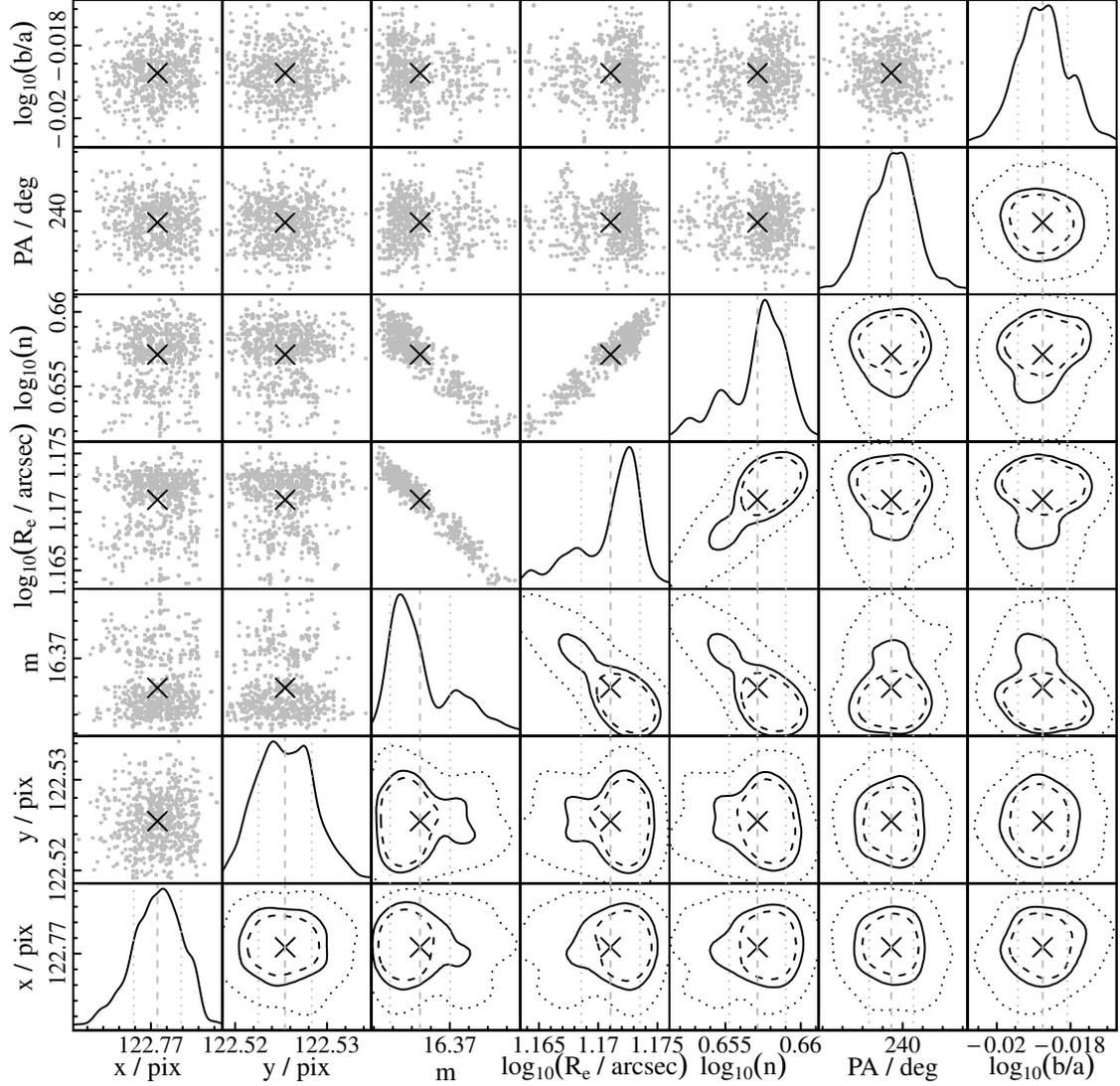


Figure 2.34: The MCMC chain for the r -band single component fit of galaxy 136851 that is converged/stationary according to both the BMK diagnostic and the Geweke diagnostic, for direct comparison to Figure 2.33.

2.2.7 Manual calibrations

While the $v04$ pipeline is fully automated, it depends on several tuning parameters that required manual calibrations based upon visual inspection during pipeline development. We point these out here as they may not be directly transferable to other bands or types of data with different depths and resolutions and therefore require special attention. We consequently also consider them separately in Section 2.3, where we introduce the VIKING data and hence need to re-calibrate some of the visual inspections (Section 2.3.4).

The first notable manual calibration is the selection of star candidates for the PSF estimation, which we describe in Section 2.2.3. It depends on the depth of the data used for the segmentation map and the exact segmentation procedure. We hence re-calibrated this for each of the individual bands as well as for the segmentation maps obtained from the stacked *gri* images and double-checked the calibrations (adjusting when needed) every time when there were major changes in the segmentation procedure. See details in Section 2.2.3.

The second manual calibration of the preparatory work pipeline is the χ^2_{ν} -cut to exclude stars which visually appear as a bad fit from the model PSF creation, see Section 2.2.3. Which cut is appropriate depends mainly on the segment size, since the reduced chi-squared within the segment will generally decrease for larger segments due to the larger number of background pixels with little deviations from the model. This effect is amplified by the fact that the KiDS weight maps are generally conservative, such that the true pixel errors tend to be smaller than the standard deviations quoted in the sigma map (cf. Section 2.2.1). We therefore re-calibrated this cut whenever there were major changes in the segmentation procedure, such as between v02 and v03 of the pipeline.

Since we have described both of these manual calibrations in the preparatory work pipeline in detail before, we do not elaborate on them further here. However, there are three more procedures based upon visual inspection during the fitting and post-processing, namely the criteria for swapping components, those for flagging bad fits and the calibration of the DIC cuts for the model selection. These are listed in Sections 2.1.2 and 2.1.3, with more details and diagnostic plots given below.

In addition to these manual calibrations, our estimation of systematic errors (Section 4.2.4) is based on simulations that were tuned to reproduce KiDS *r*-band single Sérsic fits for our sample of galaxies with our pipeline. They may not be transferable to other bands, other models, a different pipeline or different samples of galaxies.

Swapping of components

The two components of the double component model are nearly identical, with the only difference being the Sérsic index, which is a free fitting parameter for the bulge and fixed to 1 (exponential) for the disk. This similarity of the profiles leads to frequent swapping of the two components, such that around 20-30% of the double component fits have the “disk” dominating the flux in the inner regions of the galaxy and the “bulge” the outskirts. This is a common problem in galaxy fitting and usually requires some form of post-processing, see e.g. the logical filter in Allen et al. (2006). We attempted to reduce the fraction of swapped fits by constraining the bulge parameters in various ways, see Section 2.2.5. These methods, which would solve the problem at the fitting stage, had very limited success. We therefore developed

2.2. PIPELINE DEVELOPMENT

a separate routine for reducing the number of swapped components that is somewhat at the interface between fitting and post-processing: the double component models are fitted, then the swapping routine is applied, including an additional fitting step for a subsample of galaxies (about one third of the full sample). This routine is outlined in Section 2.1.2 and results in the number of swapped fits being reduced to $\sim 1\text{-}2\%$. We supplement Section 2.1.2 here with more information on how the routine was developed and some diagnostic plots.

Note that the 1.5-component fits do not suffer from swapped components for two reasons. First, they are last in the order of processing, such that the double component models have already been fitted and swapped where necessary. The 1.5-component fits benefit from these results as initial guesses and are therefore very unlikely to swap the components again. Second, and maybe even more importantly, the 1.5-component fits have much fewer degrees of freedom than the double component fits. The point source bulge is defined by just a single parameter - the magnitude - and offers very little flexibility to fit deviations of the disk from an exponential profile (which is the most common reason why the Sérsic component in the double component model fits the disk rather than the bulge). We therefore only apply the swapping routine to the double component fits.

As a reminder, the basic idea that solved the problem of swapping components is very simple: our two components are very similar and nearly interchangeable. In general, therefore, there will be two high maxima in likelihood space that are far apart. Moving from one to the other is statistically unlikely given limited run times as it requires changing 9 of the 11 double component parameters at once. The code - even the MCMC - is hence more likely to converge onto the maximum that is closer to the initial guesses. By swapping the initial guesses and re-fitting a galaxy, we can assist the code in finding the “correct” maximum (not necessarily the statistically better one but the one we find physically more appropriate, i.e. with the bulge at the centre).

Manual calibrations enter at two stages in this process, namely for selecting galaxies to enter the re-fit (to avoid having to fit all galaxies twice) and for selecting the better of the two fits after the re-fit. Both selection criteria were determined by visual inspections of r -band galaxies during pipeline development; and double-checked for their suitability to g - and i -band fits.

Figure 2.35 shows a diagnostic plot of the swapping results. We show the logarithm of the ratio of the single Sérsic effective radius to that of the bulge in the double component fit on the y -axis against the logarithm of the ratio of the bulge Sérsic index to that of the single Sérsic fit on the x -axis. This is the plane in which we select galaxies to enter the swapping procedure (i.e. to be re-fitted). For un-swapped fits, we would expect both of these quantities to be larger than zero. To allow for some scatter, we do not cut exactly at zero, but instead re-fit all galaxies that have $2x + y < 0.1$ and $2y + x < 0.1$. This cut is indicated with the yellow

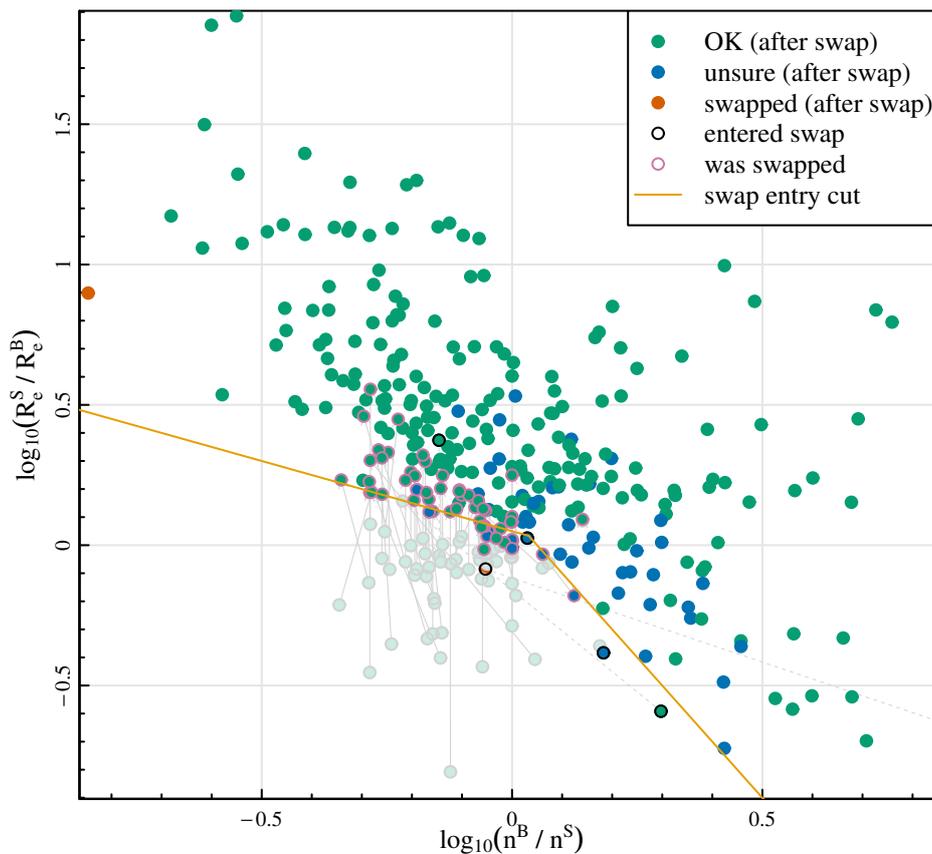


Figure 2.35: A diagnostic plot of the swapping procedure for a test sample of visually inspected r -band galaxies classified as double component systems. The x -axis shows the logarithm of the ratio of the bulge Sérsic index to that of the single Sérsic fit; while the y -axis shows the logarithm of the ratio of the single Sérsic effective radius to that of the bulge in the double component fit. The yellow line shows our selection cut for determining which galaxies to re-fit (those below the line). The colours of the points indicate the result of the visual inspection (performed after the swapping procedure) as detailed in the legend. Points encircled in pink represent galaxies that were swapped (i.e. they entered the swapping procedure and the new fit was considered better). Lighter points show the corresponding original fits (that were replaced during the swapping procedure), connected by grey arrows. Points encircled in black show galaxies that entered the swapping procedure but were not swapped (i.e. the original fit was considered better). They are connected to the points where they would have ended up if they were swapped by dashed grey arrows.

line, any galaxies for which the original fit falls below the cut enter the re-fitting stage.

We have experimented with many other parameter combinations also including the bulge-to-total flux ratio, the relative bulge and disk effective radii, the bulge and disk axial ratios and the chi-squared value of the central pixel, but have found the current procedure to produce the best results. Note that the exact cuts were refined several times during pipeline development, we only show the final v04 versions here. Cuts for earlier version of the pipeline were made in the same plane and generally similar. We only adjusted them slightly based on new visual

2.2. PIPELINE DEVELOPMENT

inspections (also making use of visual inspections done in the context of the other manual calibrations) to minimise the number of galaxies entering the swapping stage while at the same time minimising the number of galaxies with swapped components in the final catalogue.

The coloured points show where the fits fall in this plane after the swapping procedure, with the colours indicating the result of the visual inspection (which was performed on the final fits, i.e. after the swapping procedure): green indicates unswapped fits, orange shows swapped fits and blue indicates those that were unclear, e.g. if both profiles are very similar, or the galaxy may better be represented by a single Sérsic fit (i.e. a model selection fail) or because the galaxy is very irregular.

The points encircled in pink are the ones that entered the swapping procedure and were swapped, i.e. these coloured points represent the new (better) fits after the swapping. The corresponding original fits for the same galaxies are shown as lighter points (of the same colour) connected by light grey arrows. All of the original fits lie below the selection cuts, while most of the re-fits lie above it; i.e. generally galaxies are moved towards the “good” areas of the plot during the swapping procedure, even though this criterion is not applied in the selection of the better fit (see below). After the swapping, all fits are also considered un-swapped or at least unsure. Note that we did not do a visual calibration on the fits before swapping (so the colours of the lighter points just represent the final result for the galaxy after the swapping), but given that they changed their parameters significantly during the procedure it is very likely that they indeed had swapped components before.

The points encircled in black are those that entered the swapping procedure but were not swapped, i.e. the original fit was considered better. Light grey dotted arrows connect them to their corresponding re-fits in lighter colours (that were then discarded). There are only five such objects in the entire sample of 300 visually inspected galaxies, i.e. the number of re-fits done in vain is very small. Note one of these is actually above the selection cuts and only entered the re-fits because it is the second match of a galaxy in the overlap sample for which the first match required swapping (for ease of processing, we enter the entire galaxy for re-fitting if any of its matches were identified as being potentially swapped). Three more of these objects are close to the selection cut and either did not change their parameters much during the re-fit (for two of them) or were fine before the re-fit (the third).

Finally, the last object that entered the swapping procedure and did not get swapped is one of the two objects (of a total of 300) that remain swapped after our procedure, i.e. a “failed swap”. For this object (partly covered by another object so the orange colour is difficult to see), the re-fit seems to have failed as indicated by the dotted grey arrow pointing far beyond the plotting region. The second of the failed swaps is the leftmost point in the plot and did not enter the swapping procedure. Reducing the number of swapped components from around one third to only two out of 300 can be seen as a success though. Further visual inspections

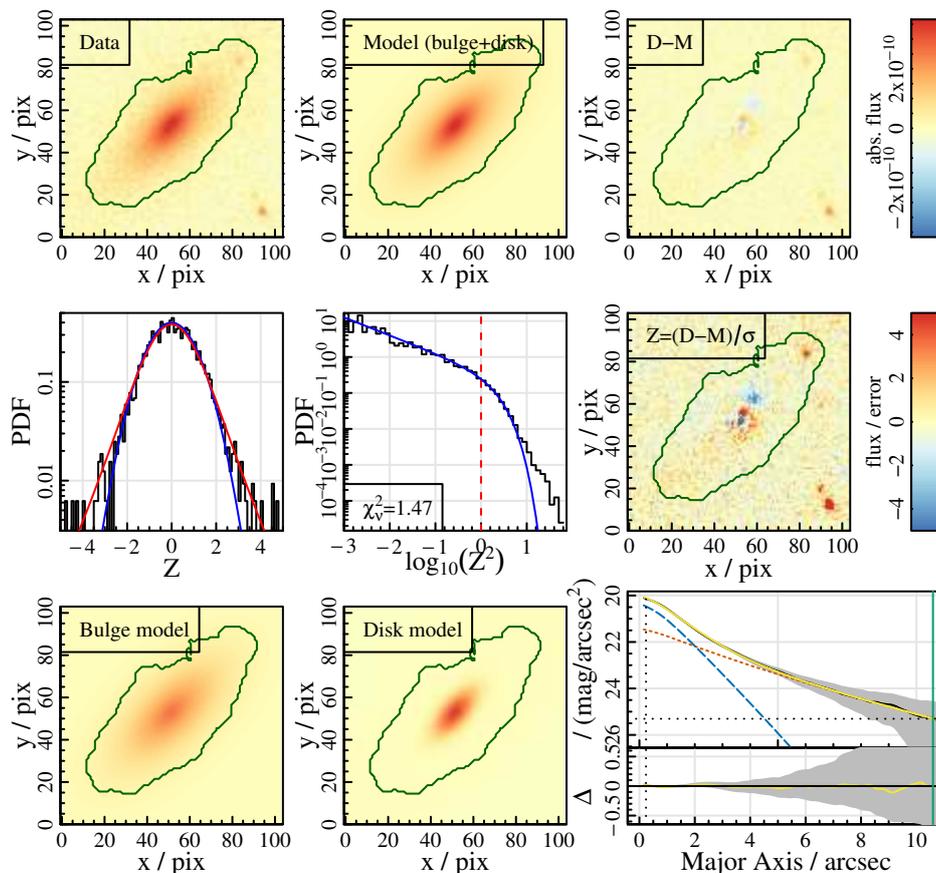


Figure 2.36: The double component fit to galaxy 593119 in the KiDS r -band before it entered the swapping procedure. Panels are the same as those in Figure 2.2.

indicate that the number of failed swaps is generally around 1 – 2% in all bands.

The last point that remains to be explained is how we select the better one of the re-fits. This criterion is stated in Section 2.1.2 and is a combination of physical considerations and visual inspections. To recap, if the “bulge” component is close to exponential, larger than the disk and contains the majority of the flux for only one of the two fits, then we choose the other one. For all galaxies for which this criterion is met by both or neither of the fits (which is the majority), we select the one with the higher absolute value of bulge flux in the central pixel. We devised and calibrated these criteria during visual inspection of samples of galaxies that were fitted twice (once with initial guesses swapped), comparing the two fits against each other and flagging the physically most appropriate one.

We also experimented with many different criteria for this selection, e.g. the higher absolute value of flux in central regions of differing sizes (instead of only one pixel), higher values of bulge flux relative to the disk flux (rather than absolute), chi-squared values in the central pixel(s), combinations of other parameters such as Sérsic index, bulge-to-total ratio and ef-

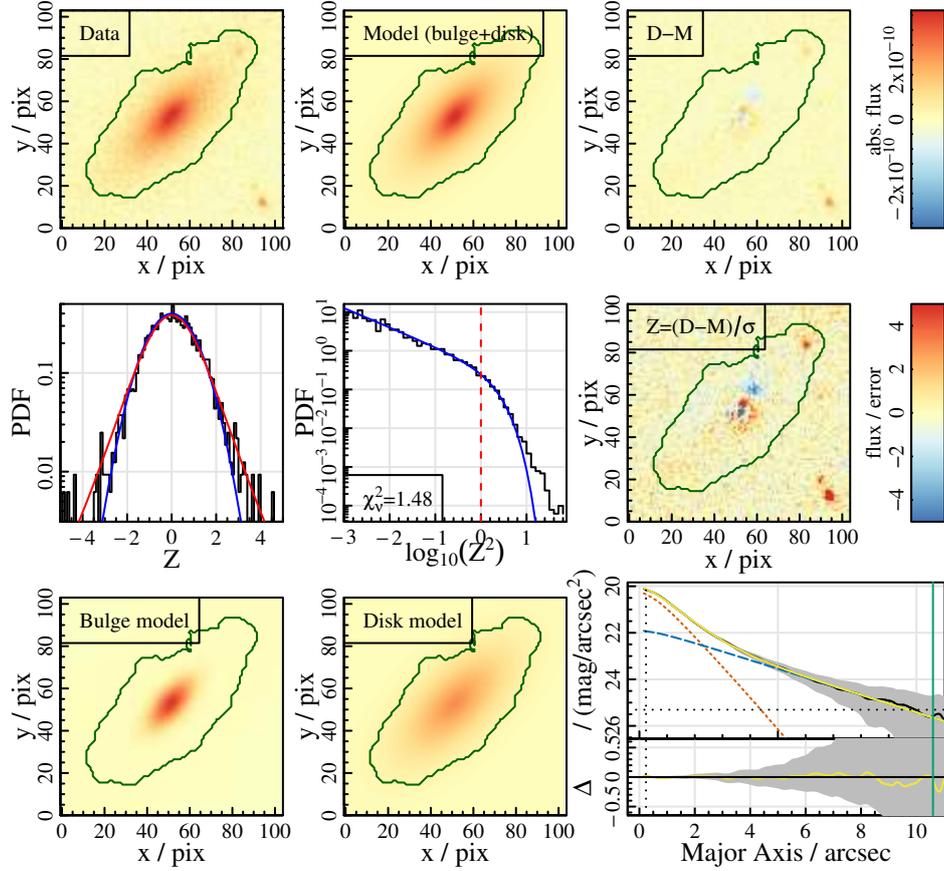


Figure 2.37: The double component fit to galaxy 593119 after the swapping procedure, for direct comparison to Figure 2.36.

fective radii or selections that are statistically better motivated, for example choosing the fit with the higher likelihood. In the end, the rather simple selection of the fit with the higher absolute value of bulge flux in the central pixel proved to be most suitable to our needs.

Figures 2.36 and 2.37 show an example galaxy pre- and post-swapping. The resulting fit after the swapping procedure is not perfect (galaxies that can be perfectly represented by our models rarely suffer from swapped components) and in fact has a slightly higher reduced chi-squared value than the original fit, i.e. it is statistically worse. Still, the fit post-swapping is physically more appropriate than the original one as the bulge now dominates the flux in the central region and the disk that in the outskirts.

Outlier flagging

Similar to the swapping procedure, the flagging of bad fits (outliers) proceeded in a very iterative way, experiencing multiple refinements during pipeline development. The basic procedure was to visually inspect a random sample of galaxies and flag them as outlier or not;

then visualise the distributions of outliers and good fits as a function of many (combinations of) metrics that we would expect to capture some aspects of “bad fits” either from a theoretical point of view or from the visual impression gained while inspecting our sample of galaxies. We then proceeded to visually inspect more (non-random) samples of galaxies in the regions where we identified possible metrics, to better define the exact cuts.

This process or parts thereof were repeated several times when we changed the preparatory work procedure (in particular the segmentation) or revised some modelling decisions such as the fitting limits or parameter constraints. The resulting metrics that we used for outlier flagging in v04 of the BDDcomp DMU are listed in Section 2.1.3, including an indication of how many fits they affect. They were all calibrated on the r -band fits, but double-checked and found to be suitable for the g and i band fits as well.

Initially, the outlier flagging was heavily linked to the swapping and model selection calibration, since all three processes are interconnected. For these reasons, many of the calibrations are in fact based on the same visual inspections, where we flagged numerous categories such as “the double component fit is swapped but could probably be a good fit if it were not”, “borderline between single/double component fit and outlier”, “this galaxy would need at least three components” or “failed segmentation map (could maybe be a good fit otherwise)” along with more straight-forward categories for galaxies which can be appropriately represented by one of our models¹².

Once we had finalised the swapping procedure, the number of swapped components was very low such that this effect was de-coupled from the outlier flagging and model selection. The latter two remain connected though, since the decision of whether or not a given galaxy can be appropriately represented by one of the models depends on the available model fits. The visual inspection for the model selection therefore always included an “outlier” category, which we in turn could also use to define the outlier criteria (with further refinements based on non-random samples as stated above; which did not enter the model selection calibration). Further details on the final order of processing of the model selection and outlier rejection routines are given in Section 2.1.3.

Note that in contrast to the swapping procedure, the outlier flagging and model selection are entirely post-processing routines, i.e. they do not require any re-fits and also did not require specific test runs to be calibrated during pipeline development. Nonetheless, some of the visual inspections performed in the context of model selection and outlier flagging also influenced our modelling decisions such as the definition of the fitting limits for the Sérsic index (cf. Section 2.2.5) and not least the addition of the 1.5-component models.

¹²The 1.5-component fits were not considered during most of this iterative process since they were added at a later stage during pipeline development. The outlier flagging was then only adapted slightly to accommodate the new model fits.

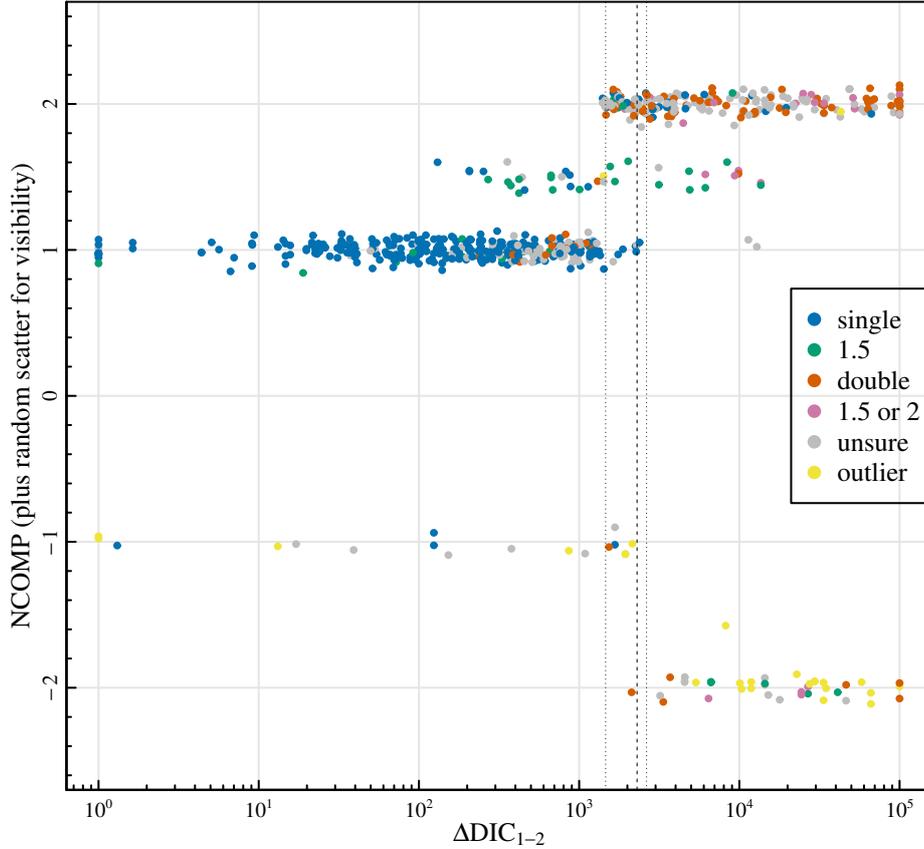


Figure 2.38: A diagnostic plot of the outlier rejection and model selection routines. We show the number of components assigned by the automated routine (NCOMP) on the y -axis against the DIC difference between the single and double component fits on the x -axis. Each point represents one of the ~ 700 galaxies used during model selection calibration, coloured by the result of the visual inspection as indicated in the legend. The vertical black dashed and dotted lines indicate the calibrated DIC cut with its bootstrap uncertainties.

Figure 2.38 shows a diagnostic plot of the combined outlier flagging and model selection. It shows the number of components assigned by the automated routine against the DIC difference between the single and double component models, clipped to the plot limits (for simplicity, we do not show the other two DIC differences). As a reminder, NCOMP values of 1, 1.5 and 2 mean that the object was classified as single, 1.5- and double component fit respectively, while objects with negative NCOMP are outliers, where the absolute value indicates the category that the object would have been assigned to if it were not an outlier. Each coloured point is one of the ~ 700 galaxies used for v04 model selection calibration and sorted into one of the six categories indicated by the legend during visual inspection (cf. Section 2.1.3). If all routines were perfect, we would expect all yellow points to be in the lower part of the plot; while all blue, green and red points would be on the $\text{NCOMP} = 1, 1.5$ and 2 lines respectively (except for the random scatter that we added for visibility). Pink points can either be on the 1.5 or 2 lines and grey points (unsure; excluded from all analyses) are allowed to be anywhere,

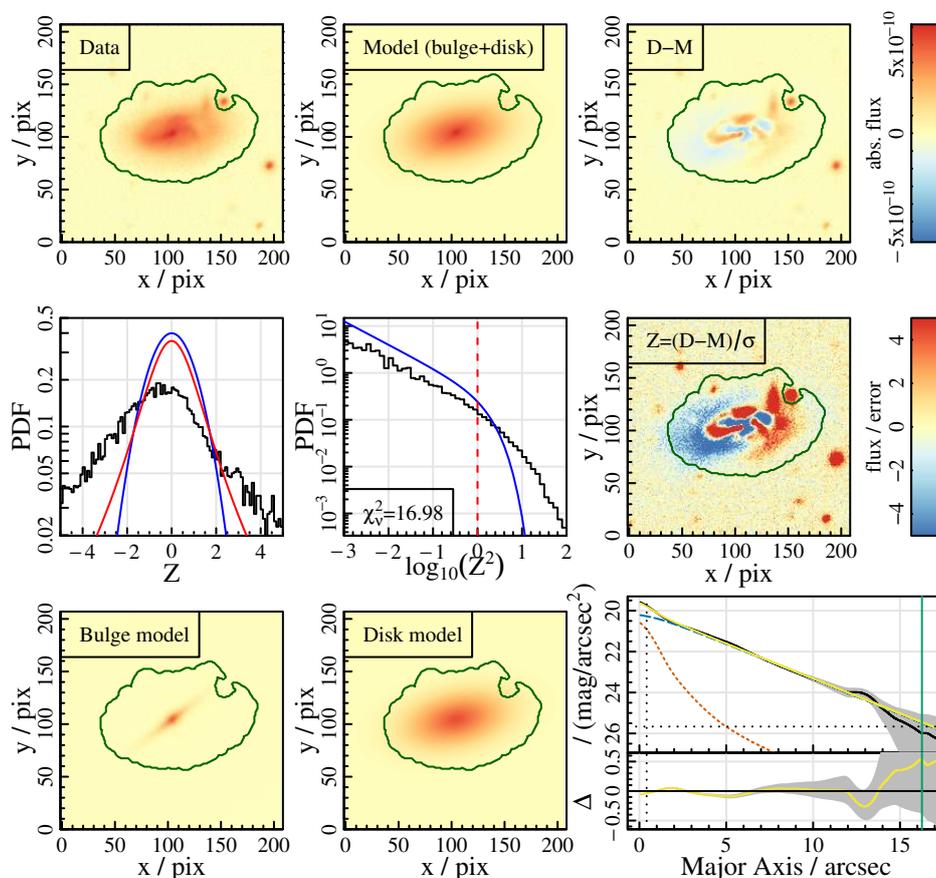


Figure 2.39: The double component fit to galaxy 278760, classified as outlier ($\text{NCOMP} = -2$) in the KiDS r -band. Panels are the same as those in Figure 2.2.

though preferentially in the positive region.

We further investigate the model selection accuracy below. Focusing on just the outlier flagging, there are a total of three objects that were visually classified as outliers but are considered good fits in the automated selection. Vice versa, there are 20 objects classified as acceptable fits which are flagged as outliers by the automated routine. This amounts to a total of approximately 3% of all objects being classified wrongly by the outlier flagging routine. A further 14 objects (2%) are visually classified as “unsure” and ended up being flagged as outlier. The “unsure” category includes galaxies for which it was unclear which model is the best (since several models delivered equally good fits) as well as those for which it was unclear whether or not they are an outlier (borderline cases and/or several models delivered equally bad fits). The majority of the grey points should hence lie in the positive regions of the plot, but we cannot judge how many exactly are erroneously flagged as outliers.

Depending on whether the “unsure” category is considered or not, the outlier flagging routine has an overall accuracy of 95-97% compared against visual inspection for the KiDS r -band.

The corresponding fractions in the g and i bands are 94-97% and 95-97% respectively. An example of a galaxy classified as outlier is shown in Figure 2.39.

Model selection calibration

The third and arguably most important calibration based on visual inspection is that of the DIC difference cuts for model selection. As mentioned above, it is intrinsically linked to the flagging of outliers and to a lesser extent also to the swapping routine. The theoretical background for Bayesian model selection is given in Section 1.3.1. Our model selection routine (including its visual calibration) is then explained on that basis in Section 2.1.3.

We emphasize again that purely statistical measures cannot achieve the aim of our model selection, namely to select the physically most appropriate fit, especially in the context of our models being statistically inadequate for the majority of galaxies in our sample. A purely visually-based classification is possible (see, e.g., Hashemizadeh et al. 2022 or Driver et al. 2022), but time-consuming for large samples of galaxies such as ours. This is particularly true because the visual classification would need to be repeated for each band individually as long as bands are also fitted individually. Classifying galaxies from e.g. colour images to determine the number of physically distinct components they contain, sometimes even before (or without subsequent) model fitting, is common in the galaxy fitting community (see e.g. Hashemizadeh et al. (2022)) and certainly desirable from a scientific perspective. From a statistical perspective though, if the single-band image used for the fit is too shallow to constrain two components, then a double component fit will not produce reliable parameters even for galaxies which physically do consist of two components. For this reason, we calibrate the model selection on single-band diagnostic plots comparing the fits achieved by our three models and selecting the best (or - if there are several equally good fits - simplest) one.

We then use the result of the visual inspection of a subsample of galaxies to calibrate a cut based on statistical measures (the DIC) that we subsequently apply to the full sample. This limits the number of galaxies that need visual inspection while overcoming the unsuitability of a statistical measure alone. Nonetheless, the model selection is limited by the statistical measure to some extent, as the statistical measure will never be able to distinguish between different “kinds” of bad model fits. For example, the automated procedure based on a DIC cut will always select the double component fit if it is significantly better than the single component fit (where “significantly better” is the part that we calibrate during visual inspection), irrespective of whether the bulge and disk components are swapped or for example the bulge dominates both the centre and the outskirts of the galaxy.

To further limit these shortcomings of a DIC difference cut, we experimented with combining it with more physical measures, e.g. a Δ DIC cut as a function of bulge Sérsic index or bulge

to total flux ratio (and many other versions). None of these more complicated joint cuts improved the model selection significantly, so we chose the simplest version of just a (one-dimensional) cut in ΔDIC . The accuracy that we achieve compared against visual inspection (see confusion matrices in Section 2.1.3 and below) is sufficient for our needs despite the intrinsic limitations of the method. The only viable technique to overcome such limitations in the future is machine learning, as investigated by e.g. Dimauro et al. (2018). However, the majority of recent efforts to apply machine learning to the problem of galaxy classifications is focused on reproducing visual morphological classifications based on images alone (e.g. Nolte et al. 2019; Turner et al. 2021; Tarsitano et al. 2022), which are to be distinguished from model selection.

Figure 2.38, which was already introduced above, shows the result of the model selection compared against visual calibration. Most single component galaxies (blue points) are correctly classified as such, especially for low ΔDIC ; while those with high ΔDIC are correctly identified as being the double component fits (orange dots). The ΔDIC cut with its uncertainty region (from bootstrapping) is indicated with vertical dashed and dotted black lines; and can also be inferred from the distribution of points in the $\text{NCOMP} = 1$ and $\text{NCOMP} = 2$ classes. Most confusion between the single and double component models is near those cuts, where coincidentally (but not entirely surprisingly) we can also find the highest number of galaxies classified as “unsure”. Interestingly, the 1.5-component models have intermediate values of ΔDIC_{1-2} (which is the quantity we show on the x -axis; we do not show $\Delta\text{DIC}_{1-1.5}$ or $\Delta\text{DIC}_{1.5-2}$ which are the relevant quantities for 1.5-component model selection).

The last point to note about Figure 2.38 is that most galaxies that are wrongly classified as outliers (i.e. blue, green, orange and pink points in the lower half of the plot) are at least classified in their correct model category, i.e. the majority of blue points in the lower half of the plot lies on the $\text{NCOMP} = -1$ line, while the majority of orange points is at $\text{NCOMP} = -2$. For this reason, and in an attempt to characterise the accuracy of the model selection alone rather than in combination with the reliability of the outlier rejection routine, we use absolute values of NCOMP for all model selection statistics and ignore the “outlier” category (essentially adding the yellow points to the grey population).

The full confusion matrix for the r -band model selection is given in Table 2.3 in Section 2.1.3. For completeness, we add the corresponding confusion matrices for the g and i bands and the joint model selection in Table 2.5, although the statistics are generally similar between bands. For reference, we also list the calibrated DIC difference cuts between all three models for all three bands and the joint model selection in Table 2.6. As one would expect (due to Ockham’s factor), the cuts are generally higher for models with larger differences in the numbers of parameters and for bands with better data quality.

2.2. PIPELINE DEVELOPMENT

Table 2.5: The confusion matrices for our model selection based on a Δ DIC cut compared against visual inspection for the g and i bands and the joint model selection. All values are in percent of the total number of visually inspected galaxies in the respective band(s). Bold font highlights galaxies classified correctly, while grey shows those that were ignored during the calibration.

(a) The g -band model selection confusion matrix.				(b) The i -band model selection confusion matrix.			
visual class.	number of components			visual class.	number of components		
	1	1.5	2		1	1.5	2
“single”	48.9	0.4	1.0	“single”	51.8	0.4	1.5
“1.5”	2.1	3.7	0.6	“1.5”	1.9	3.0	1.5
“double”	1.8	0.1	6.0	“double”	0.7	0	8.9
“1.5 or double”	0.4	1.5	1.9	“1.5 or double”	0	1.0	1.9
“unsure”	19.5	0.9	7.3	“unsure”	10.0	0.7	12.4
“unfittable”	1.3	0.3	2.1	“unfittable”	1.2	0.3	2.2

(c) The joint model selection confusion matrix.			
visual class.	number of components		
	1	1.5	2
“single”	47.4	0.5	1.5
“1.5”	2.2	2.4	1.4
“double”	2.2	0.4	7.3
“1.5 or double”	0.3	1.1	2.1
“unsure”	15.0	0.7	11.1
“unfittable”	1.0	0.4	2.4

Table 2.6: The calibrated DIC difference cuts used for v04 of the BDDcomp DMU. For each band, we show the lower bound (LB), the actual cut and the upper bound (UB) for each of the three DIC differences between our models. For the joint model selection, the cuts refer to the differences between the summed DICs of all bands (cf. Section 2.1.3).

band	Δ DIC _{1-1.5}			Δ DIC ₁₋₂			Δ DIC _{1.5-2}		
	LB	cut	UB	LB	cut	UB	LB	cut	UB
g	43	68	93	1760	2273	4856	551	817	940
r	57	172	248	1463	2298	2636	690	853	947
i	33	57	127	396	473	520	308	336	438
joint	183	255	341	5707	5751	5797	1650	1824	3104

2.3 Pipeline updates

This section describes the updates to the pipeline performed when adding the processing of VIKING data, i.e. in between v04 and v05 of the BDDcomp DMU, the latter of which is to be released on the GAMA database alongside the publication of this thesis. At the same time we included the KiDS u -band, such that now our catalogue covers all 9 optical and near-infrared bands from KiDS and VIKING ($u, g, r, i, Z, Y, J, H, K_s$). This work has not been presented to date.

2.3.1 VIKING data products

As for KiDS (see Section 2.2.1), we needed to decide which of the VIKING data products to use. VIKING provides data products at two different levels: pawprints and tiles, the latter of which are made of six pawprints stacked together to close the gaps between detector chips. In addition, for the J -band, there are “deep stacks” composed of two tiles each, since the J -band tiles have been observed twice in the VIKING observing strategy (see Section 1.2.3). All of these data products are astrometrically and photometrically calibrated. In contrast to KiDS, however, they are not re-gridded, so they have various pixel sizes (close to $0''.34$), are only approximately aligned in RA and Dec and are calibrated to differing Vega zeropoints given in the image headers. The frames are also not corrected for atmospheric extinction and the flux units still contain the exposure time. The sky background variations are subtracted, but only on large scales (Edge & Sutherland, 2020).

The most important difference to KiDS images for our analysis, however, is that the pawprints making up a tile are not taken in direct succession but can be taken hours or even months apart. This leads to seeing variations and hence considerable PSF differences between the pawprints making up a tile (Edge & Sutherland, 2020). For accurate photometry, it is therefore necessary to apply an aperture correction that is a function of position in the tile, termed “grouting”. This is done by the VIKING team to obtain their photometric catalogue, but with various different pipeline versions as their data reduction procedure evolved during the time of the survey. The tiles themselves cannot easily be corrected for this and are hence left with potentially strongly and abruptly varying PSFs since different pairs of pawprints contribute to different areas of the tile given the VIKING dithering pattern, see Edge & Sutherland (2020).

This drove our decision to work at the pawprint level instead of using stacked image tiles, since a reliable PSF estimation is vital for an accurate galaxy fit. Very similar considerations led Wright et al. (2019) and Driver et al. (2016) to the same conclusion. Instead of downloading the pawprints and re-calibrating them ourselves, we decided to directly use the preprocessed detector chips of Wright et al. (2019) (who were following Driver et al. 2016 in their analysis).

2.3. PIPELINE UPDATES

They are not publicly available, but the KiDS team kindly provided them to us for this work. Given that the KiDS team use these re-processed chips for their own photometric analysis combining KiDS and VIKING data in Wright et al. (2019), they seemed highly appropriate to use for our work as well.

Wright (private communication) provided individual detector chips (i.e. each pawprint split into its 16 chips) with a size of $\sim 2200 \times 2200 \text{ pix}^2$ or approximately $750'' \times 750''$. Apart from the data reduction already performed by VIKING, they have been re-calibrated with a multiplicative correction factor to account for atmospheric extinction, remove the exposure time from the image units and convert the images onto a common AB magnitude zeropoint of 30 (see Wright et al. 2019 for details). In addition, the chips are rotated slightly such that their x and y axes align exactly in RA and Dec and at the same time the background is subtracted and a weight map created. This is performed using the SWarp software (Bertin, 2010) after truncating the chip edges slightly to exclude unreliable pixels (Wright et al., 2019, and Wright, private communication). It results in a background-subtracted chip image with a corresponding weight map that is zero around the edges of the chips and has a uniform value elsewhere, giving the average inverse variance over the chip. Chips with unreliable photometry are then identified via a cut in the recalibration factor and discarded.

Note that Wright et al. (2019) do not use the VIKING confidence maps in their analysis. They exist for pawprints and tiles and give the pixel-by-pixel variation in exposure time, with a value of 100 referring to the median exposure time in the given band. They also include masking for detector artifacts such as bad pixels and the two “bad patches” that do not flat-field well (Edge & Sutherland, 2020). Apart from these bad patches, the confidence maps are relatively uniform for individual chips, with variations for the tiles mostly caused by the dithering pattern. Since Wright et al. (2019) exclusively work at the detector level and use a weighted combination of the individual detector fluxes with outlier rejection, there was no need to consider the confidence maps in their analysis.

Following the same arguments, we decided to simply use the uniform-valued weight maps of Wright et al. (2019) for our analysis as well. While we do not perform a weighted combination of fluxes with outlier rejection, each of our galaxy is covered by typically 3-4 VIKING exposures and up to 40 in the J -band overlap regions. Should any of these (partly) fall into a “bad patch”, it will likely not dominate the galaxy fit. The same applies to satellite tracks and similar artefacts that only affect individual chips. Nonetheless, a potential improvement of our analysis of the VIKING data would be to re-scale the Wright et al. (2019) weight maps with the corresponding confidence maps, after rotating the latter onto the same pixel grid.

The last difference to the KiDS data that we would like to mention is that VIKING do not provide masks for bright stars and their possible reflection halos. These could be created following, e.g., the procedure in Barnett et al. (2021), which is another potential improvement

to our analysis of VIKING data. For the current work, we decided against such a procedure and instead rely on the rather conservative KiDS masks that were considered during image segmentation (see Section 2.3.2).

In summary, we opt to use the individual VIKING chips recalibrated by Wright et al. (2019) with associated uniform-valued weight maps and no bright star masks. This results in many data matches for each galaxy (a median of 3 in Z, Y, H, K_s and 6 in J ; although it can be more than 20 in overlap regions) with significantly different PSFs, making a joint fit necessary to avoid stacking. This became possible with the multi-frame fitting functionality of ProFit v2.0.0, released in February 2021.

2.3.2 Segmentation changes

Most of the preparatory work remained unchanged between v04 and v05 of the pipeline, except for the technical details to enable the analysis of the VIKING data (e.g. differing naming conventions). We still take cutouts around each object for each data match, estimate the sigma maps and do the background subtraction and PSF estimation separately for each image. The only major difference is in the treatment of the segmentation maps and masks.

For the KiDS g, r and i bands, we used a stacked image to define a joint segmentation map in all three bands, including a joint mask (see Section 2.1.1). This was straight-forward since all tiles are registered to the same pixel grid across all KiDS bands, so there is always an exact correspondence of a tile in the r -band to one in the g - and one in the i -band; and they can simply be added (apart from very slight variations in the tile sizes that are easily taken account of). The VIKING chips generally also cover similar areas of sky in the different bands (except for the “missing chips”, see Section 2.3.3), but they are not registered to the same grid and have differing pixel sizes.

A joint analysis of all bands is still possible, but requires more care to correctly account for these effects. In addition, due to the large number of matches and the much smaller sizes of VIKING chips compared to KiDS tiles ($750''$ vs. 1° per side), galaxies are more frequently covered only partly by individual exposures. Considering only the area covered (and not masked) by all exposures in all bands would unnecessarily decrease the amount of available data for fitting and result in no data being left in many cases (see also Figure 2.40). Nonetheless, we want to have the same fitting regions in all bands, to avoid systematic effects due to different segment sizes (see Sections 2.1.3 and 2.2.4).

For these reasons, we decided to define the KiDS g, r and i bands as our “core” bands (cf. Section 1.2.2) and perform the segmentation on the gri stacks as for v04. For the u, Z, Y, J, H and K_s bands, we then simply transfer these gri segmentation maps (one for sky subtraction and one for galaxy fitting) onto the corresponding world coordinates and pixel

2.3. PIPELINE UPDATES

grid of the target image using `profoundSegimWarp`. In case there are several segmentation maps available for the same galaxy (i.e. for the overlap sample in *gri*), we use the first match for transferring to other bands.¹³ Should the segmentation map extend beyond the edge of the target image, it is truncated accordingly.

For the *u*-band, we then additionally apply the *u*-band mask, although the vast majority of areas masked in the *u*-band are already included in the stacked *gri* masks that were considered during the segmentation. As mentioned before, we do not use any additional masks for VIKING data (apart from those included in the weight maps, see Section 2.3.1). Stars that are saturated in KiDS data, as well as reflection haloes, are conservatively masked during our analysis (Section 2.2.1) and will automatically be excluded from the segmentation. We consider it very unlikely that a star would be saturated and/or produce significant reflection haloes in one of the VIKING bands while being unmasked in all of the KiDS *g*, *r* and *i* bands. This is particularly true since we use individual detector chip images for VIKING and stacked (hence much deeper) tiles for KiDS.

This procedure ensures that the fitting regions across all bands are as similar as possible (and exactly identical in our three core bands), while maximising the available data for fitting. The result for an example galaxy can be seen in Figure 2.40, where we give a qualitative impression of the matches available for each band and the corresponding segmentation maps. All images are shown on their native pixel scales and for this reason, the cutout sizes and flux scalings are also different. Since the figure is only meant to give a visual impression, we omit axis labels and colour bars.

The galaxy shown is part of the overlap sample in the KiDS *g, r, i* bands and also has many matches in the VIKING bands (we show the first eight for each band). The two segmentation maps for the two matches in *gri* are similar, but not identical. For all other bands, the first of those was transferred to all matches. Matches for which the galaxy centre falls into a white region (missing data) are skipped, i.e. not used during the joint fit. Some further matches are skipped due to PSF failures, see Section 2.3.3. For the *r*-band, we additionally show the *v04* segmentation maps on the same axis scale as their corresponding *v05* versions for direct comparison (see below for reasons why the segment size increased in *v05*).

Potential improvements to this segmentation procedure would be to jointly treat multiple matches in *gri* (i.e. for the overlap sample, where we still produce separate segmentation maps for each match) and/or use the correct corresponding match in the *u*-band, which is also pixel-matched to the other KiDS bands (instead of simply transferring the larger segmentation map to all matches). Ideally, we would in the future consider a stack¹⁴ of all bands used for analysis

¹³We actually intended to use the largest segmentation map instead of the first match, but this did not work due to a small bug in our code. The bug is fixed for any future runs, but we did not consider this to be a problem that would justify re-processing the entire *v05* of the `BDDcomp` DMU.

¹⁴Stacking is not a problem in this context, since PSF variations do not affect the segmentation heavily.

to define the segmentation maps, with regions missing from individual frames (because they are masked or beyond the edge of the corresponding image) downweighted appropriately without excluding them entirely. This is not possible with the current version of `ProFound` and would be non-trivial to implement.

One last change in the v05 segmentation maps came from two default changes in `profound-ProFound`: the `skycut` default changed from 1 to 1.5 and the `SBdilate` default changed from `NULL` to 2.

For the meaning and effect of `skycut`, see Section 2.2.4. We use two different versions of segmentation maps for the background subtraction and galaxy fitting with explicitly different aims (see also Sections 2.1.1, 2.2.2 and 2.2.4) that are produced with `skycut` values of 1 and 2 respectively. This was the optimal procedure we found after numerous experiments with different `skycut` values, including 1.5. We therefore considered it best to discard the new default and instead set `skycut` to 1 explicitly in all places where it was previously left on its default.

The `SBdilate` option specifies how many magnitudes to push beyond the surface brightness limit in segment dilation if the number of pixels to be added is larger than `SBN100` (with a default of 100 that we left unchanged). After defining initial segments, `profoundProFound` expands those until its convergence criterion is met. By default, the convergence criterion is that the total flux within the segment increases by less than 5% during the dilation (Robotham et al., 2018). For bright and extended galaxies, this means the number of dilations is often zero or one (cf. Section 2.2.4) since the flux already contained in the segment is large, so the fractional change during dilation is small. For these objects, the `SBdilate` option can be used to dilate the segments further and capture the low surface brightness flux in the extended wings. In detail, if the dilated annulus contains at least `SBN100` pixels and has a mean surface brightness brighter than the sky surface brightness limit plus `SBdilate`, then the segment dilation continues (i.e. the annulus will be added to the current segment even if it contributes less than 5% additional flux). For the new default values this means that we assume that 100 pix are sufficient to safely push the detection 2 mag beyond the sky surface brightness limit of individual pixels.

We had not previously considered the `SBdilate` option since it was not available in the early versions of `ProFound` that we used for most of the preparatory work pipeline development; and we did not re-visit this after `SBdilate` was introduced. After the most recent `ProFound` update, we considered simply setting it back to its previous default of `NULL`, meaning that the `SBdilate` routine is not triggered. However, we decided to adopt the new defaults for several reasons. First, the logic behind `SBdilate` seemed reasonable and we could indeed observe that it typically does not change the star segments (for PSF estimation) but does dilate most galaxy segments, especially those of bright and extended objects. For these objects we had

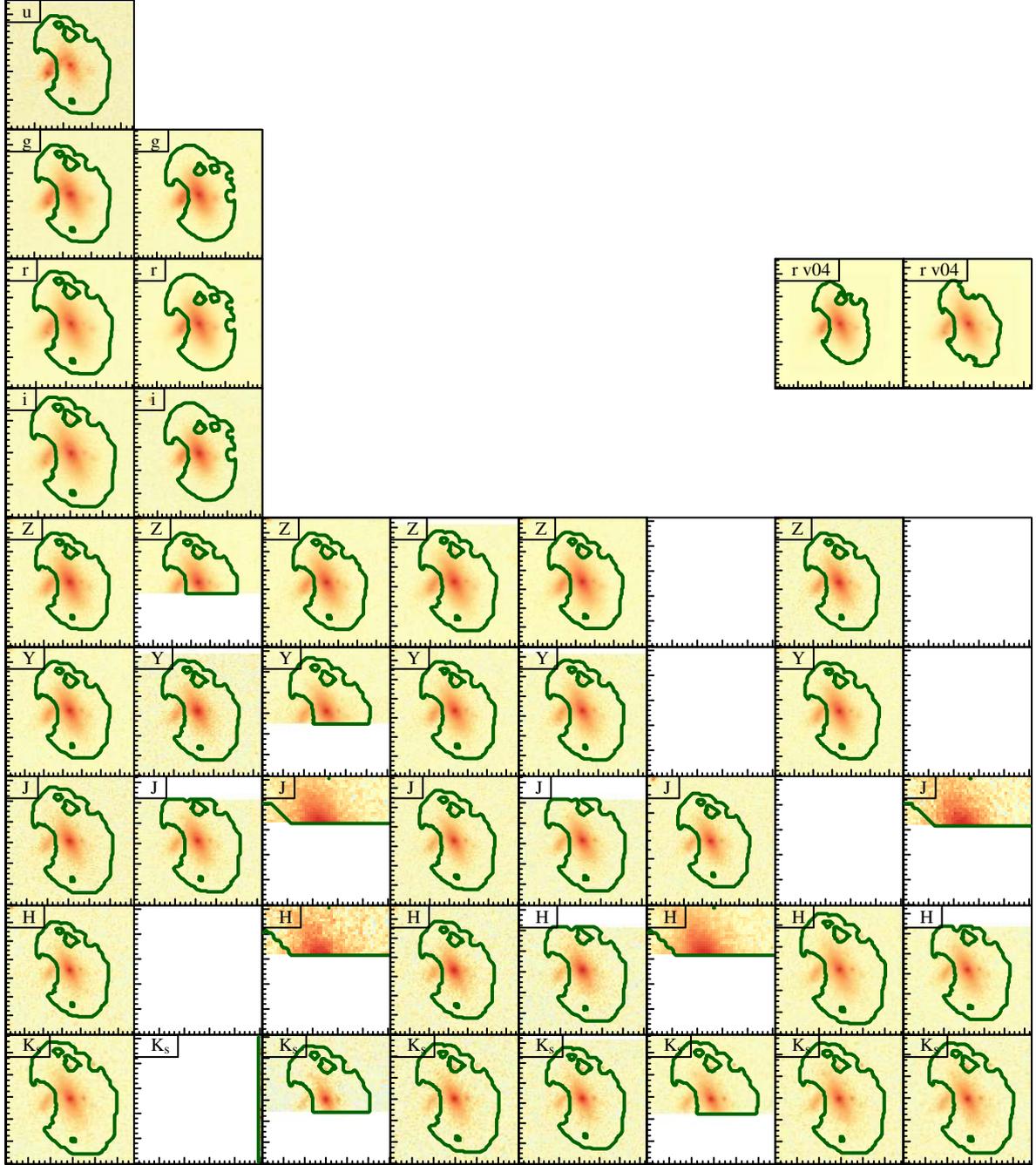


Figure 2.40: An example of the images and v05 segmentation maps (green contours) used for fitting galaxy 544891 in all nine bands (the same object as that shown in Figure 2.20). This galaxy returned one data match in the u -band, two in g, r, i (of which none were skipped), 14 in Z (5 of which were skipped), 13 in Y (5 skipped), 21 in J (9 skipped), 8 in H (5 skipped) and also 8 in K_s (1 skipped). For the VIKING bands, we only show the first 8 matches each. White regions indicate missing data beyond the image edge or in the region of the chip that has zero weight assigned (cf. Section 2.3.1). For the r -band, we also show the v04 segmentation maps for direct comparison (the second of which corresponds to the rightmost panel in Figure 2.20). These are shown in the same cutout size as their corresponding v05 versions for easier visual comparison. All other cutouts are shown in their original size (i.e. with differing numbers of pixels) and with individual flux scaling for best visibility (due to the different pixel scales). We omit axis labels and colour bars due to the qualitative nature of this plot.

also noticed (during the detailed analysis of the v04 results) that the azimuthally averaged one-dimensional flux profiles often continue to be positive and well-defined beyond the v04 segment region, see e.g. Figure 2.2. This is an indication that while individual pixels are below the surface brightness limit, their sum is still above it (and would be sufficient to constrain the model profile to larger radii), i.e. we have not reached the actual sky background yet. While our aim is to have relatively tight segments that do not include many background pixels, we did not intend to discard the extended wings of galaxies.

In addition, since we now use the segments defined on the *gri* images also for the other bands, we need to allow for potential variations of object sizes as a function of wavelength. There can also be variations in depth between bands and even within the same band (in particular in the overlap regions) due to the fact that we now jointly fit all matches but still define segments on individual matches of the *gri* stacks. For all of these reasons, it seemed safer to use somewhat larger, more dilated segments for v05 compared to v04 and we decided to adopt the new default of `SBdilate`.

An example of the resulting v05 segmentation map can be seen in Figure 2.40 for galaxy 544891. This is the same object that we also showed in Figure 2.20 to allow for a visual impression of the evolution of segmentation maps during pipeline development. For a direct comparison to the v04 segmentation maps, we also show these in Figure 2.40 in the rightmost columns as indicated. They are shown in the same cutout size as their corresponding v05 version. Clearly, the segment size increased due to the `SBdilate` procedure.

2.3.3 Joint fitting

Since we work at the individual exposure level for VIKING data, galaxies frequently have multiple data matches as can be seen in the top panel of Figure 2.41. Due to our joint processing of the *g*, *r* and *i* bands (cf. Section 2.3.2), these have the exact same number of matches (with a minimum and median of 1 and a maximum of 4). The *u*-band follows the *g*, *r* and *i* bands closely (also with a minimum and median of 1 and a maximum of 4 matches), while the VIKING bands have many more matches on average but also some missing matches. The *Z*, *Y*, *H* and *K_s* bands show similar distributions ranging from 0 to around 20 with medians of 3. The *J* band has typically twice as many matches (a median of 6 and a maximum of 39 matches) due to the VIKING observing strategy, see Section 1.2.3.

The number of missing (zero) matches is very similar in all VIKING bands. They are due to small gaps in the VIKING data caused by a combination of incomplete sky coverage (see Edge & Sutherland 2020 and also figure 1 in Wright et al. 2019) and the quality control cuts applied by the VIKING team as well as by Wright et al. (2019), see Section 2.3.1. This affects approximately 5% of the galaxies in our sample which are largely the same in all VIKING

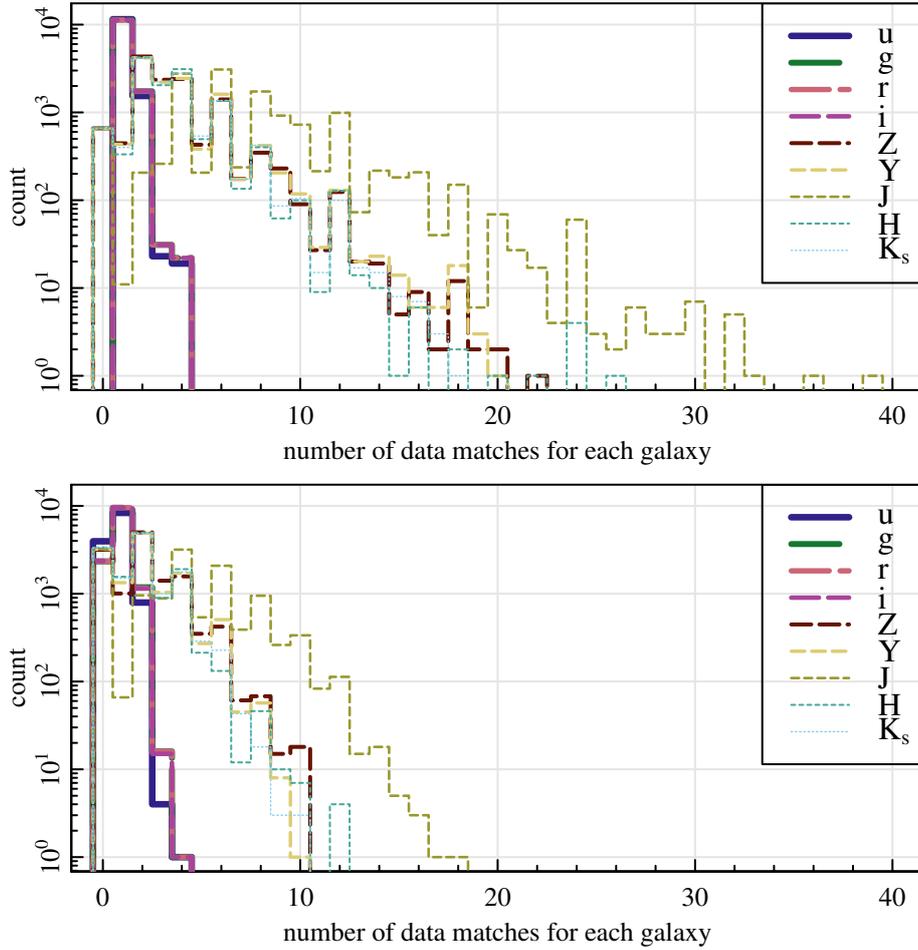


Figure 2.41: Top: the number of data matches to each of our 13096 galaxies for all nine KiDS and VIKING bands as indicated in the legend. Note the logarithmic scaling on the y -axis. **Bottom:** the same but showing only non-skipped matches (i.e. those that were actually used for fitting).

bands (cf. Table 3.2) and preferentially distributed around the edges of the three equatorial GAMA regions.

Except for these “missing objects”, all matches pass through the preparatory work pipeline. However, some may be skipped at the fitting stage if the galaxy centre falls within a masked region or the PSF estimation failed. The number of images actually used for fitting each galaxy is shown in the bottom panel of Figure 2.41. For the g, r, i bands, around 20% of individual matches are skipped due to the KiDS masking and a further $\sim 1\%$ due to PSF fails (cf. Table 3.1). Approximately 18% of galaxies have no matches available for fitting and are skipped entirely (Table 3.2).

For the u -band, the number of galaxies with no matches available is higher due to the additional masking, the decreased data quality and the slightly smaller footprint of the u -band

tiles compared to the g , r and i bands (since the u -band tiles consist of only four dithers, while for the other bands it is five, cf. Section 1.2.2). The smaller footprint by itself does not exclude any objects (i.e. there is no missing data in the u -band like in the VIKING bands), but it decreases the overlap regions between tiles. Hence galaxies in these regions are more often covered by only one exposure and/or less dithers (i.e. shallower data). These effects combined - plus the additional u -band masking - lead to more galaxies being skipped due to masking ($\sim 2\%$) and PSF failures ($\sim 11\%$), see also Table 3.2 in Chapter 3. The latter effect in particular is very significant in the u -band due to the shallower data (not only in the overlap regions) in combination with our usage of the gri segmentation maps: many candidate stars are too faint in u to reliably estimate a PSF. Those that would be bright enough are masked, since they are saturated in the other bands. The median and maximum numbers of matches remain one and four respectively for all KiDS bands.

Due to our use of the gri segmentation maps, all galaxies that are skipped due to masking in the core bands (i.e. around 18%) also have zero images available for fitting in the VIKING bands. Additionally, there are the $\sim 5\%$ of galaxies for which no matches were returned in the first place, such that the number of objects skipped entirely during the fit in the VIKING bands is slightly higher than that in the KiDS g , r and i bands, although still lower than in the u -band. The number of images available for fitting is, however, reduced significantly in all bands, especially for those objects with many data matches. This is mainly due to the truncation of chip edges by Wright et al. (2019), whereby unreliable pixels around the edges of chips are assigned zero weight (Section 2.3.1). A galaxy falling within this region of the chip is hence counted as a match, but may be covered only partly or not at all by the actual data. We count these objects as being “masked” in all subsequent analysis.

Since this affects the chip edges only (there are no additional masks for the VIKING data), it exclusively affects the VIKING overlap regions. In addition, some matches may be skipped due to PSF failures. This also preferentially affects the VIKING overlap regions where there are less stars available for fitting (both due to the missing data around the edge of each chip as well as the large cutouts used for PSF estimation extending beyond the actual edge of the chip). As we typically have many data matches available in the overlap regions, it is relatively rare that a galaxy ends up with zero matches due to either of these reasons (both combined contribute around $1\text{-}2\%$ in addition to the KiDS masking, cf. Table 3.2). The number of matches available for fitting therefore remains higher than for KiDS data, with values typically around two to four in the Z, Y, H, K_s bands and a few more in the J -band (with medians of two and four respectively). The maximum number of matches is reduced to 18 in the J -band and around 10 in all other bands. An example of a galaxy with many matches is shown in Figure 2.40.

This high number of data matches made treating each match individually (as we did for

2.3. PIPELINE UPDATES

v04) impractical from a computational point of view and - more importantly - unwise from a scientific point of view as we would lose too much depth of the data. Since we want to avoid stacking due to the different PSFs in VIKING frames, we perform a multi-frame fit whereby we jointly fit all images of the same galaxy in each band. Note the bands are still treated individually, though, i.e. we perform multi-frame rather than multi-band fits. For consistency, we also re-process the KiDS g , r and i bands.

This means that we pass all (individually sky-subtracted) images for the same physical object in the same band - with their corresponding masks, segmentation maps, sigma maps and PSFs - to `ProFit` at once. The model is then fitted to all images jointly, with appropriate weighting and taking account of the different PSFs. Since we treat each band individually, we do not allow any model parameters to vary between images, i.e. there is only one model applied to all images. The only additional user input needed is an additive offset between the individual images in x and y (in pixels) and in rotation angle (in degrees); as well as a multiplicative offset in pixel scale.

The rotation angle offset is zero in our case since all KiDS tiles as well as all VIKING chips re-processed by Wright et al. (2019) are exactly aligned in RA and Dec. For KiDS data, the pixel scale offset is also zero. The VIKING chips have slightly varying pixel scales ranging between $0''.337$ and $0''.341$. We do not consider this since the early versions of the `ProFit` multi-frame fitting mode did not offer support for varying pixel scales (it does now, so we can improve this in the future). This means that the estimates of the effective radius (the only parameter that depends on the pixel scale) are inconsistent at the 1% level. This is well within its uncertainty for the vast majority of galaxies, with the median relative error on single Sérsic effective radius ranging from 4% in the Z -band to 10% in K_s .

The only non-zero offsets in our analysis are the differences between the x and y positions of the galaxy centre in the different frames. These can be calculated easily from the world coordinate system headers of the corresponding images, assuming the astrometric solutions to be exact. The KiDS DR4.0 astrometric solutions show a scatter of approximately $0''.04$ each in RA and Dec (Kuijken et al., 2019). As we discuss in Section 4.2.4, this - although it is well within the pixel scale of $0''.2$ - is a factor of approximately four larger than the median MCMC error on position when fitting individual images (see also Figure 4.9; and examples of the same effect can be seen in Figures 2.16 and 2.17). For the joint fit, it is therefore possible to see visual offsets of the galaxy centre in the different matches for extreme cases.

An example for such an extreme case in the r -band is shown in Figures 2.42 and 2.43. There is a discrepancy in the astrometric solution between the two images: for the individual fits in v04, the difference in the fitted RA and Dec positions for the two images are $0''.1$ and $0''.4$ respectively. For the joint fit in v05, this leads to the shallower of the two fits in particular being offset with respect to the actual centre of the galaxy, showing up as red and blue regions

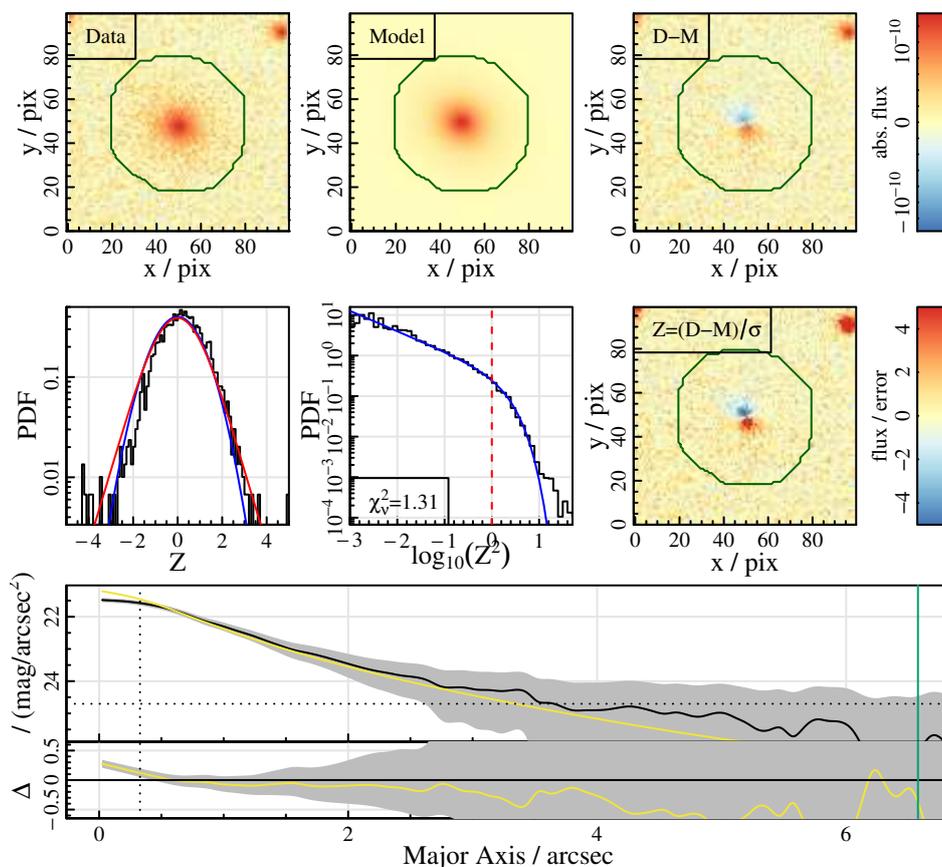


Figure 2.42: The (joint v05) single Sérsic fit to the first image of galaxy 345472 in the KiDS r -band. Panels in the top two rows are the same as those in Figure 2.2, while the bottom row shows the one-dimensional fit only, corresponding to the rightmost panel of the bottom row in Figure 2.2.

in Figure 2.42. The fit to the other image (Figure 2.43) shows a corresponding offset in the other direction, although not as clearly visible since it is the deeper image and dominates the fit. This adds onto the systematic uncertainties. The vast majority of offsets are much smaller than the extreme case shown here and cannot be visually seen in the model fits. It also only affects the overlap sample in KiDS where there is more than one match; and at the same time the astrometric solutions as well as the overall data quality are worst.

For VIKING data, the scatter of the astrometric solution is $0''.09$ (Edge & Sutherland, 2020). This is a factor of about two larger than for KiDS. However, the pixel size is also a factor of approximately two larger and the median MCMC uncertainty on position ranges from $0''.03$ in Z to $0''.05$ in K_s in both x and y . The systematic uncertainty from the astrometric solution is therefore only a factor of two to three larger than the random uncertainty, compared to a factor of four for KiDS. It is therefore likely that the astrometric uncertainty affects VIKING bands less than KiDS bands despite the increased number of matches for each galaxy. Our

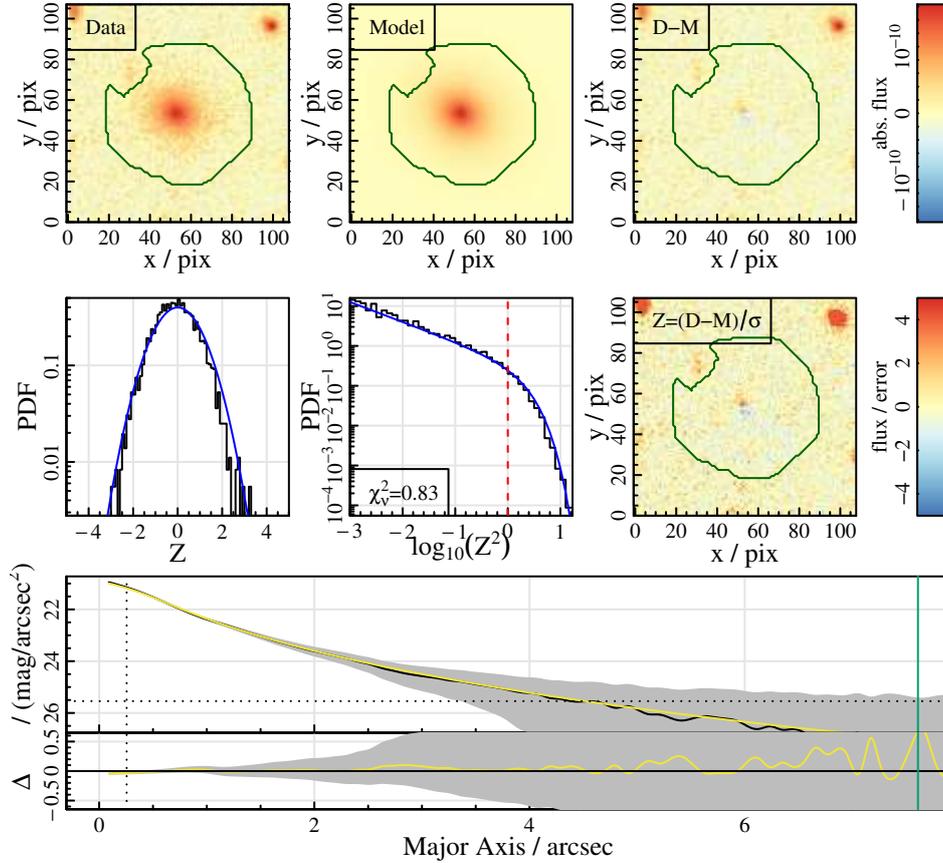


Figure 2.43: The (joint v05) single Sérsic fit to the second image of galaxy 345472 in the KiDS r -band, for direct comparison to Figure 2.42.

visual impression was also that offsets visible by eye (such as those shown in Figures 2.42 and 2.43) are less frequent in VIKING bands than in the KiDS g , r and i bands.

To account for the uncertainty in the astrometric solution, we would need to allow for varying x and y centres to be fitted to each image of the same galaxy. This was not supported by early versions of the ProFit multi-frame fitting mode and can have several caveats as it would need to be ensured that the fitted offset is indeed due to the astrometric solution and not due to differing noise, image artefacts or similar issues. We therefore did not implement this for v05 of the pipeline. However, it could be added in future work, minimising the caveats for example by imposing a (Gaussian) prior with a width taken from the known scatter of the astrometric solution.

There are no changes with respect to v04 on the modelling side, except for some technical details, where, e.g., we needed to modify the `convergeFit` function to make it compatible with the ProFit multi-frame fitting mode. Other than that, we use the same three models, the same initial parameter guesses and priors, the Normal likelihood function and the same

fitting algorithm and convergence criteria. The only difference is that we now obtain just a single fit for each physical object, referring to the first (non-skipped) data match where necessary (i.e. for x and y positions in our case; and potentially also for the position angle and effective radius for non-zero offsets in rotation angle and pixel scale).

The advantage of this approach is that it makes full use of the data available for each physical object; and there is only one result for each object so there is no need to combine or choose between different measurements of the same galaxy. The disadvantage is that we lose the overlap sample for internal consistency checks and estimation of systematic uncertainties, as we have done in Section 4.2.4 for the v04 results. In addition, the method relies heavily on the accuracy of the astrometric solutions (see above).

The natural next step would be to move to multi-band fits, i.e. fitting all matches from all bands for the same physical object jointly. This is the approach taken by the MEGAMORPH project team (Häußler et al., 2013; Vika et al., 2013; Häußler et al., 2022) and has many advantages since it ensures smooth wavelength trends while preserving physical variation and additionally allows more robust fits to fainter magnitudes. We did not consider this approach initially due to the lack of support for multi-frame fitting in early versions of ProFit. With ProFit v2.0.0 this is now possible, and certainly an interesting option for future work. We did not implement multi-band fitting for v05 yet as it requires a major effort to be implemented computationally and above all requires the definition of how parameters can vary across wavelength.

The newly developed package ProFuse (Robotham et al., 2022), combining ProFit with the spectral energy distribution fitting package ProSpect (Robotham et al., 2020) is a first step towards fitting physically meaningful variations as a function of wavelength. However, like other works using simultaneous multi-band fits (e.g. Häußler et al., 2022), it (for now) fixes the structural parameters and only leaves the magnitudes free to vary. This is not an option for our work since variations in size as a function of wavelength are crucial for our science aim (Section 1.1.3). Hence the upgrade from v04 to v05 of our catalogue focused on expanding our previous work to all nine bands and to multi-frame fitting. It is a first step towards multi-band fitting and at the same time complementary to current works in that area; and can even serve to inform choices of “allowed” variations in other parameters.

2.3.4 Manual re-calibrations

We point out the tuning parameters of the v04 pipeline that are based on manual calibration steps in Section 2.2.7. These may not all be directly transferable to VIKING data and hence we re-calibrated some of them for v05. There are five such calibrations in total.

The first one is the cut chosen for the candidate star selection during PSF estimation (the

2.3. PIPELINE UPDATES

percentage of all segments that are likely to be stars, see Section 2.2.3). It depends on the depth of the data used for the segmentation map and the total number of segments returned by the segmentation procedure. Since we use the *gri* segmentation maps for the VIKING bands (and the `SBdilate` option only changes the size and not the number of segments), we do not need to re-calibrate this cut. We visually inspected a random sample of the diagnostic plots of this step (example shown in Figure 2.7) to ensure that the 4% selection cut is indeed appropriate for the KiDS *u* and all five VIKING bands provided that the *gri* segmentations are used.

The second visually tuned parameter is the cut in reduced chi-squared to exclude bad star fits from the model PSF creation (Section 2.2.3), which is influenced mainly by the segment size. While the galaxy segments for v05 generally increased in size due to the `SBdilate` option (cf. Section 2.3.2), the star segments did not change since stars are compact and not extended objects. We therefore also do not need to adjust this cut for v05.

The third manually tuned process is the swapping of bulge and disk components where necessary. This procedure was developed with many visual inspections and test runs of *r*-band fits. However, the resulting procedure, described in detail in Sections 2.1.2 and 2.2.7, does not require any tuning parameters in absolute terms and instead relies only on relative quantities such as the ratio of the double component bulge effective radius to that of the single Sérsic fit. It is therefore expected to work well for other bands, too, as long as the galaxies in that band show comparable properties as those in the *r*-band. We confirmed this in detail using the *g* and *i* band fits in v04 by repeating all visual inspection steps. For the bands added in v05, we checked the most important statistics: the fraction of objects entering the swapping procedure and the fraction of those actually being swapped are very similar to the *r* band fractions. The fraction of objects that remain swapped after the procedure is also very low in all bands based on the visual inspection of galaxies in the context of the model selection re-calibration (see below). We conclude that the swapping procedure is appropriate for all nine bands and does not require re-calibration.

The fourth set of criteria that were developed with a large number of visual inspections are those for flagging bad fits (listed in Section 2.1.3 with further details in Section 2.2.7). Similar to the swapping procedure, we calibrated this on the *r*-band but found it to be appropriate for the *g* and *i* bands as well. We therefore assumed that it would work reasonably well for the *u* and VIKING bands, too, and did not perform any re-calibration for v05. However, in contrast to the swapping procedure, there are a number of hard cuts in the outlier flagging that depend on the pixel scale and segment size or can be influenced by the depth of the data. We investigate this in detail in Section 3.2.2. In summary, we find that the outlier flagging is adequate for v05 and only the flag value “very irregular segment” could potentially have benefitted from a re-calibration. This can easily be added since the outlier flagging is a

pure post-processing procedure and we provide the parameters for all model fits irrespective of their outlier status. We also provide full details of the criteria used for flagging outliers in the corresponding flag of the `BDDcomp` DMU. Users are therefore free to ignore certain flag values or discard our flagging entirely and perform their own identification of outliers.

The fifth and last manual calibration is that of the DIC difference cut for model selection, which is described in Sections 2.1.3 and 2.2.7. These are cuts in absolute values that are strong functions of data quality (depth; and to a lesser extent also seeing) and need to be re-calibrated for each band individually. We therefore visually inspected 200 randomly selected galaxies for each band (1800 objects in total) and re-calibrated all Δ DIC cuts. The procedure is exactly the same as that used for `v04`, so we refer the reader to Sections 2.1.3 and 2.2.7 for a description. Due to the differences in processing for `v04` and `v05` and for maximal consistency of the manual calibration between all nine bands, we also re-calibrated the three core bands. The differences between the g , r and i calibrations in `v04` and `v05` are investigated further in Section 3.2.3. Here, we focus on `v05` only.

Since we perform joint simultaneous fits on all images of the same object in a given band, the most consistent way to calibrate the model selection is on combined images composed of a stack of all individual frames. Otherwise it may be difficult to evaluate the model fit for objects that have many data matches (apart from having to look at 10 diagnostic plots for a single object), since the individual frames may be too shallow to visually identify two components, but the combination of all allows to constrain them during the fit. Note that since the stacking is performed after model fitting, we can avoid the problems with the different PSFs that led us to use joint fits in the first place. To this end, we first create model images for each individual frame using its native pixel scale and PSF. We then use the `magwarp` function to re-map these PSF-convolved model images, corresponding data images and sigma maps onto the world coordinate system (pixel grid) of the first data match. Subsequently, we stack all individual data frames and model frames, both weighted appropriately by the sigma maps. Like this, the model images and data images have undergone the exact same procedure and can be directly compared. An example of such a stacked diagnostic plot composed of 12 J -band images is shown in Figure 2.44 for galaxy 544891 (which we also show in Figure 2.40).

We use the visual inspections to calibrate individual model selections for each band and two versions of joint model selections: one for the core bands only (gri , using the summed DICs from these three bands) and one for all bands ($ugriZYJHK_s$, summing all nine DICs). The resulting confusion matrices for all bands are listed in Table 2.7 and the calibrated DIC cuts are given in Table 2.8. The overall percentages of fits classified wrongly compared to visual inspection amount to $\sim 3, 15, 17, 9, 9, 3, 3, 5, 5, 17$ and 18% for the $u, g, r, i, Z, Y, J, H, K_s$ individual, gri joint and $ugriZYJHK_s$ joint model selections respectively. The high fractions in the g, r and joint selections are driven by large numbers of galaxies visually classified as

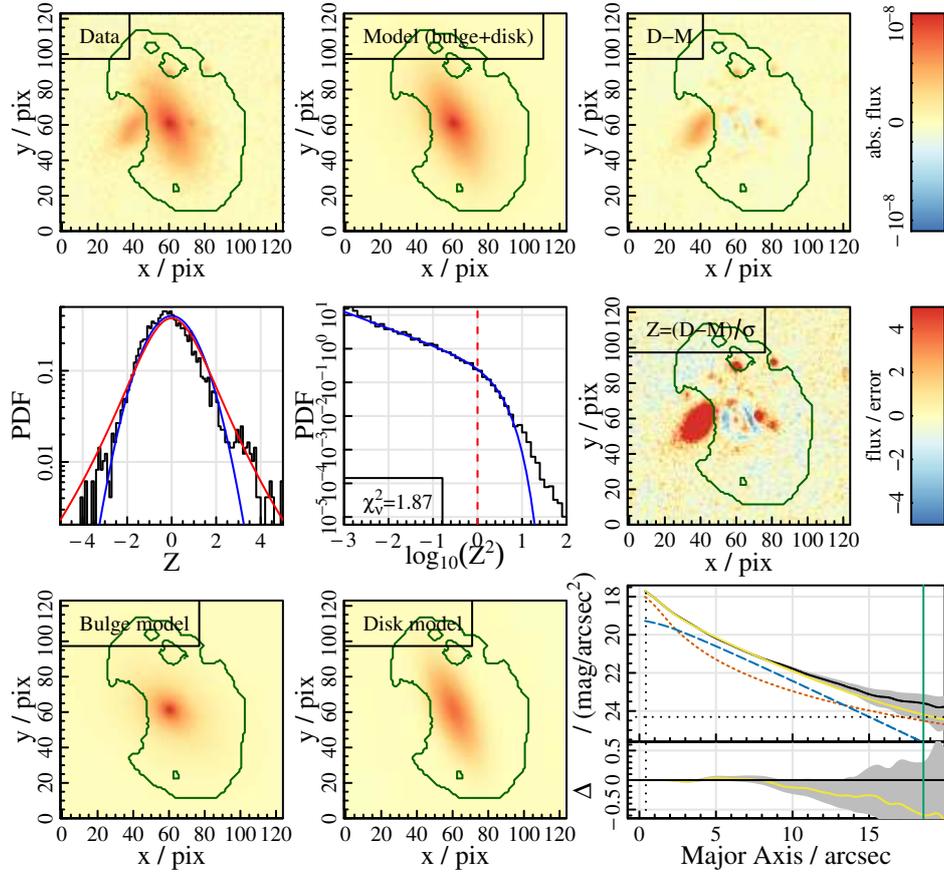


Figure 2.44: The (joint $v05$) double component fit to the galaxy 544891 in the VIKING J -band. The diagnostic plot is composed of 12 individual frames for both the data and the model that are stacked in a consistent way, taking PSF differences into account. This is the same galaxy as that shown in Figure 2.40, so the stacked version here can be compared to the individual frames shown in that figure. Panels are the same as those in Figure 2.2.

double component fit that ended up in the $\text{NCOMP} = 1$ category. We comment on this further in Section 3.2.3. For all other bands, the confusion rate is acceptable.

For the shallower bands (especially KiDS u), the data quality is not sufficient to constrain more than one component in most cases, so the fraction of single component fits increases drastically. In part, this is the reason for the low confusion rates in these bands: essentially all fits are (correctly) classified as single Sérsic objects. This is enhanced even more by the fact that we minimise the absolute number of fits classified wrongly during the model selection calibration, resulting in a bias against the rarer categories (cf. Section 2.1.3). Additionally, the corresponding DIC difference cuts are often poorly constrained due to the low number of objects available for calibration in those rare categories. This is also apparent from Table 2.8, where some of the cuts show large uncertainty regions or take on unreasonably large or small

Table 2.7: The confusion matrices for the v05 model selection based on a DIC difference cut compared against visual inspection for all nine KiDS and VIKING bands and the joint model selections in the core bands only as well as in all bands. All values are in percent of the total number of visually inspected galaxies in the respective band(s). Bold font highlights galaxies classified correctly, while grey shows those that were ignored during the calibration.

	<i>u</i> -band			<i>g</i> -band			<i>r</i> -band		
	number of comps.			number of comps.			number of comps.		
visual class.	1	1.5	2	1	1.5	2	1	1.5	2
“single”	91.5	0	0	58.5	0	2.5	47.5	0.5	2.5
“1.5”	0	1.0	0	1.5	0	1.5	0.5	0.5	2.5
“double”	2.5	0	0.5	7.5	0.5	12.5	10.5	0	23.0
“1.5 or double”	0	0.5	0	1.5	0.5	2.0	0.5	0.5	2.0
“unsure”	2.5	0	0	6.5	0	3.0	4.0	0	3.5
“unfittable”	1.5	0	0	1.5	0	0.5	1.0	0	1.0

	<i>i</i> -band			<i>Z</i> -band			<i>Y</i> -band		
	number of comps.			number of comps.			number of comps.		
visual class.	1	1.5	2	1	1.5	2	1	1.5	2
“single”	69.0	0.5	2.5	63.5	2.0	1.5	73.0	0	0
“1.5”	1.0	1.0	1.0	1.5	4.0	1.0	0.5	4.0	0
“double”	4.0	0	11.5	2.5	0	12.0	1.5	1.0	12.5
“1.5 or double”	0	0	1.0	0	1.5	1.0	0	0	1.5
“unsure”	3.0	0	3.0	4.0	1.0	2.0	1.0	0.5	2.5
“unfittable”	2.0	0	0.5	1.5	0	1.0	1.0	0	1.0

	<i>J</i> -band			<i>H</i> -band			<i>K_s</i> -band		
	number of comps.			number of comps.			number of comps.		
visual class.	1	1.5	2	1	1.5	2	1	1.5	2
“single”	76.5	0	0.5	79.0	0.5	1.0	75.0	0	1.0
“1.5”	0.5	1.0	1.0	0.5	5.5	1.0	2.5	2.0	0.5
“double”	0.5	0.5	14.0	0.5	0.5	7.0	0	0.5	13.5
“1.5 or double”	0	0	1.0	0.5	0	0.5	0	0.5	0
“unsure”	1.5	0.5	1.0	1.0	0	2.0	2.0	0	2.0
“unfittable”	0.5	0.5	0.5	0.5	0	0	0.5	0	0

	joint <i>gri</i>			joint <i>ugriZYJHK_s</i>		
	number of comps.			number of comps.		
visual class.	1	1.5	2	1	1.5	2
“single”	59.4	0.3	1.5	71.3	0.7	0.5
“1.5”	0.8	1.2	1.2	3.6	0	0.1
“double”	10.9	1.7	10.7	11.6	0.1	4.1
“1.5 or double”	0.7	0.5	1.5	1.4	0	0.1
“unsure”	5.5	0	2.2	4.4	0.1	0.4
“unfittable”	1.5	0	0.5	1.4	0	0.1

2.3. PIPELINE UPDATES

Table 2.8: The calibrated DIC difference cuts used for v05 of the `BDDcomp` DMU. For each band, we show the lower bound (LB), the actual cut and the upper bound (UB) for each of the three DIC differences between our models. For the joint model selection, the cuts refer to the differences between the summed DICs of all bands.

band	$\Delta\text{DIC}_{1-1.5}$			ΔDIC_{1-2}			$\Delta\text{DIC}_{1.5-2}$		
	LB	cut	UB	LB	cut	UB	LB	cut	UB
<i>u</i>	66	77	370	265	10^{18}	10^{40}	54	73	101
<i>g</i>	32	63	106	2096	2612	3867	10^{-11}	66	260
<i>r</i>	29	39	64	645	1881	3146	10^{-13}	116	228
<i>i</i>	137	180	10^{17}	1051	1247	1616	481	512	576
<i>Z</i>	28	31	281	464	650	819	171	306	944
<i>Y</i>	21	52	65	256	316	382	53	119	175
<i>J</i>	106	116	223	524	743	944	61	77	442
<i>H</i>	47	52	60	389	523	642	165	200	734
<i>K_s</i>	44	10^{12}	10^{35}	190	264	447	329	395	520
<i>gri</i>	175	330	27391	15145	17827	18885	10^{-54}	10^{-24}	10^{-4}
<i>ugriZYJHK_s</i>	10^4	10^8	10^{19}	59994	79722	82856	10^{-53}	10^{-23}	10^{-4}

values (essentially tending to zero or infinity).¹⁵ The latter problem could be alleviated by visually inspecting a larger subsample of galaxies in each band.

The joint model selection encompassing all nine bands benefits from the high number of 1800 visually inspected galaxies for calibration. However, this does not solve the underlying problem of the bands being too shallow to constrain two components. Also, joint model selection is always a compromise between all bands by necessity. Since most bands have relatively high numbers of single Sérsic fits, the joint model selection does, too. This means that for the deeper bands with the best seeing, many fits are classified as single Sérsic objects in the joint model selection even though they do have two well-defined components. Vice versa, an object classified as double component fit in the joint model selection may well show unconstrained components in the shallowest bands. For this reason, the joint model selection should be used with care. The compromise becomes worse for greater differences in data quality between bands.

For this reason, we also added the second version of the joint model selection, concentrating on the core bands only (like in v04). These are most directly comparable not only in their treatment throughout our pipeline, but also in their data quality, so there are less caveats in their joint model selection (although it is still a compromise, see also Section 2.1.3). To avoid such inconsistencies, some works (e.g. Lackner & Gunn 2012; Kim et al. 2016) take the structural parameters from one band (e.g. the *r*-band) and simply impose them onto the other fitting bands without further adjustments; only fitting the component magnitudes. Others

¹⁵The lower cuts tend to zero rather than negative infinity because we calibrate the ΔDIC cuts in logarithmic space.

(e.g. Simard et al. 2011; Kennedy et al. 2016) use simultaneous multi-band fits but again only allow the magnitudes to vary between bands. While this may result in robust colours, it does not capture physical variations of structural parameters with wavelength, which are a key quantity we want to obtain. The only way to avoid these issues and successfully use shallower bands for multi-component fits (other than deeper observations) are simultaneous multi-band fits with varying parameters as a function of wavelength. Constraints on the “allowed” structural variation between bands can be motivated by theory, simulations or previous work fitting bands individually, such as this one.

3 Results

In this chapter, we present the results of our bulge-disk decomposition pipeline. This is mainly a catalogue of parameters for the (components of) galaxies in our sample along with a wealth of ancillary data and quality control metrics. We investigate some of these metrics here. A more detailed quality control of the catalogue is then presented in Chapter 4, where we compare our work to previous works in the field and perform a detailed study of systematic uncertainties. Since the quality control focuses on v04 of the `BDDcomp` DMU, we concentrate on this version here, too (following Casura et al. submitted). However, we expand on v04 with the new results from v05 that have not been presented previously, highlighting and discussing differences where relevant. This anchors the v05 results to those of v04, thereby benefitting from the detailed quality control of that version presented in Chapter 4. Results from DMU versions prior to v04 are not presented since they were somewhat preliminary with limited quality control and were superseded entirely by v04.

3.1 BDDcomp DMU

Our main result is the `BDDcomp` DMU on the GAMA database. It contains two catalogues per band: `BDInputs` with the most important outputs of the preparatory work pipeline (segmentation, PSF estimation, initial guesses) and `BDModelsAll` with the output from the galaxy fitting and post-processing (model selection, flagging of bad fits and truncating to segment radii). From v03 onwards, where we include several bands in the analysis, the catalogue names also specify the band to which they refer (e.g. `BDModelsAllR`). Up to v04, all images of the same object in the same band (data matches) are listed separately since they were also treated individually. For v05, they were fitted jointly so each galaxy only has one entry in the `BDModelsAll` table. The matches remain separate in the `BDInputs` tables since the preparatory work was still carried out on each image individually.

In addition to these band-specific tables, `BDModels` gives the most important columns of the `BDModelsAll` table(s), combining results from all bands from v03 onwards. It also has a few additional joint columns (mainly joint model selection). Finally, the table `BDModelsAlt` (added from v03 onwards) presents the same information as `BDModels` just with the different bands arranged in rows instead of columns. This results in a total of three tables up to v02 (*r*-

band only), eight tables for v03 and v04 (g, r, i) and 20 tables for v05 ($u, g, r, i, Z, Y, J, H, K_s$).

Each table is accompanied by comprehensive documentation including descriptions of all columns, details on the processing steps and practical tips for using the catalogue. The DMU also provides all input data used for the fitting (i.e. image cutouts, masks, error maps, segmentation maps, sky estimates, PSFs) as well as various diagnostic plots of the fit results on the GAMA file server, where detailed descriptions of these files can be found.

Sections 3.1.1 to 3.1.6 summarise the key changes between the different DMU versions (also listed on the GAMA database) and give the basis for each DMU version, as labelled in the directory tree of our local machines. Furthermore, they provide the release date and the versions of `ProFit` and `ProFound` used. Individual aspects of this pipeline evolution are pointed out throughout Chapter 2 and we refer the reader to this chapter for details.

In the remainder of this chapter we then present an overview over the contents of the main catalogue, `BDModels`, in v04 and v05.

3.1.1 BDDecomp v01

v01 is the first release of our bulge-disk decomposition catalogue, published on 2019-02-08. It is limited to the r -band only and is based on the preparatory work run labelled "run3" and the galaxy fitting "run5". All previous runs as well as numerous test runs were not published on the GAMA database. `ProFit` and `ProFound` versions used were from 2018-04-24.

3.1.2 BDDecomp v02

v02 is only a minor update of v01 (published 2019-02-27) with two small mistakes fixed in the post-processing of the `BDModels` and `BDModelsAll` tables:

- The `*_OUTLIER_FLAGS` were fixed for 297 single component fits, 2211 double component fits and 3437 1.5-component fits.
- The `P_BDQUAL_FLAGS` were fixed for all 1.5-component fits.

The preparatory work and galaxy fitting results did not change; and both are based on the same runs as v01.

3.1.3 BDDecomp v03

v03 is a major update which was released on 2019-12-03 and is based on preparatory work "run4" and fitting "run6". We included the g and i bands in addition to the r -band, increasing the number of catalogues from three to eight (two per band plus two joint ones). This also resulted in a slight re-labelling of the catalogues themselves and their column names to

distinguish all values between the three bands. At the same time, we upgraded to KiDS DR4.0 (which had become available in the meantime), including the photometric homogenisation given in the DMAG header keyword. We also updated `ProFit` and `ProFound` to the versions from 2019-08-19, which saw major changes since the v01/v02 DMU release. All of the above, as well as our own investigations and feedback from both collaborators and users of the first catalogue versions resulted in numerous changes to the pipeline.

In the preparatory work pipeline (see the first half of Section 2.2 for details):

- We turned off our own routine for segmentation map fixing since the new version of `ProFound` was much less prone to “shredding” galaxies and our routine tended to include faint secondary objects in the segments.
- We increased the `skycut` value in `profoundProFound` from 1 to 2 resulting in smaller segments with smoother borders.
- We added one additional dilation of the galaxy segment after the `profoundProFound` run, which ensures that the edges are smooth and unbiased (without this, the segment border can be very jagged including noisy, slightly positive pixels but excluding slightly negative pixels; especially for large bright objects where the number of curve of growth iterations is often zero or one). The dilated segments are then approximately the same size as they were in the previous run due to the higher `skycut` value.
- We defined the segmentation maps on stacked images of the g, r, i bands (with corresponding stacked masks) and used these segmentation maps in all bands.
- We defined and used two different segmentation maps for the sky estimation and object fitting; the one used for sky estimation is more dilated to exclude faint sources and extended wings from the sky (aggressive object masking and lower `skycut`).
- The star fraction cut to identify candidate stars and the chi-squared cut to exclude bad fits from PSF modelling were adjusted slightly to account for the new segments.
- We stopped using the `profoundSkySplitFFT` routine and instead take the sky estimate from `profoundProFound` directly, as we found that to be more robust.

In the galaxy fitting and post-processing (see the second half of Section 2.2 for details):

- We increased the limits for the Sérsic index fitting to 0.1 to 20 (previously 1 to 12) allowing for flatter bulges and causing fewer fits hitting the limits.
- We improved the swapping criteria slightly (both for the selection of fits to enter the swapping procedure and for selecting the better of two swapped fits) based on new visual inspections.

- We updated the outlier rejection criteria to reflect the new Sérsic index limit; and we included a new criterion based on the difference between the magnitude within the segment and within the segment radius (also based on new visual inspections).
- We re-calibrated the model selection for each band (new visual inspections) and additionally provided a joint model selection that is based on the summed DICs for all bands.
- We added several alternative measurements of the magnitudes, effective radii and bulge-to-total (B/T) ratios for each fit, resulting in the catalogues having more columns:
 - The Sérsic magnitude and effective radius.
 - The magnitude contained within the segment and the corresponding effective radius (containing half of the segment flux).
 - The magnitude and effective radius within the “segment radius”, which is defined as the maximum distance between the centre of the fit and the edge of the segment. This is the upper limit to which our model fits are valid, everything beyond that is extrapolated and unconstrained. We recommend using these values for magnitude, effective radius and bulge-to-total flux ratios because many of our fits have high Sérsic indices with unphysically large wings beyond the segment borders. This is because we opted for relatively tight segments focusing on fitting the central regions well rather than forcing the models to zero at large radii (Section 2.1.3). For reproducibility and ease of comparison to other catalogues, the segment radius (`RAD_SEG`) is provided in a separate column as well.
 - The magnitude and effective radius within 10 Sérsic effective radii. Note that these values may still include unphysical results (see above) and were provided for completeness only.

3.1.4 BDDecomp v04

v04 is again a relatively minor update with no changes to the statistical properties of the galaxy parameters. It is based on preparatory work "run5" and fitting "run7", used ProFit and ProFound versions from 2019-08-19 and was released on 2021-06-25. Its major effort was in better characterising the systematic uncertainties. This is the version that Casura et al. (submitted) is based on and is the currently newest version on the GAMA database.

In detail:

- Parameter errors labelled `*_ERR` now include our best estimate of systematic uncertainties. The purely random MCMC errors have been re-labelled to `*_ERR_MCMC`. The

`*_ERR` values are obtained from the `*_ERR_MCMC` values by multiplying with the “error underestimate” factors for each parameter listed in table 4 of Casura et al. (submitted) and reproduced here in Table 4.2. We do not apply the bias corrections listed in the same table since they are only applicable to the averages of large statistical samples, not individual galaxies. Since the error underestimates are slightly different for `*_SEGRAD` values, these values now also have associated uncertainties (`*_SEGRAD_ERR`). Note the error underestimates are derived from single component r -band fits. We apply them to the g and i bands and bulge and disk parameters for 1.5 and double component fits as well, since that should result in more realistic errors than the random ones alone. However, it is likely that the true errors are even larger for component parameters (in particular for the non-dominant component) and/or the g and i bands (also for single Sérsic fits, since the r -band is the deepest and best-resolved).

- B/T values now also have errors (`*_BT_ERR`) given for convenience. They are derived from the bulge and disk magnitude errors assuming independent variables (which is not actually true) and hence should be taken as lower limits.
- We fixed a small bug in the preparatory work pipeline where in `v03` we accidentally performed the KiDS zeropoint homogenisation correction (with the `DMAG` header keyword) twice, resulting in initial guesses for magnitudes being biased by about 0.03 mag on average. Since the MCMC fits are not strongly dependent on the initial guesses, this did not affect the fits (the zeropoint was correct during the actual fitting). Nonetheless, we re-ran the entire preparatory work and fitting pipelines.
- Position angles are now all in the interval 0 to 180° (which they were already supposed to be in `v03`, but were not in all cases).
- Galaxies are now also flagged as outliers if the single Sérsic fit failed (even if the 1.5- or double component fits were successful) because then the segment radius (and thus all `*_SEGRAD` columns) are missing; and selecting good single Sérsic fits with `NCOMP > 0` fails. This affects only one fit in the g -band and none in the other bands.
- For consistency we renamed the columns `*_FLUX_ERR_SEG` to `*_FLUX_SEG_ERR` and `*_RAD_SEG` to `*_SEGRAD`.
- We added a quick-start guide to the DMU description on the GAMA database to cover the most important aspects in catalogue usage.

3.1.5 BDDecomp v05

`v05` is another major update, to be released on the GAMA database alongside this thesis. It uses preparatory work "run6" and fitting "run8" with `ProFit` and `ProFound` versions

last updated on 2022-02-11. In addition, we moved from Ubuntu 16.04 to Ubuntu 20.04 on all of our local machines, which necessitated updating `R` along with all of its packages and replacing the `astro` package by the (relatively new) `Rfits` package for writing FITS files. On the scientific side, the major upgrades are that we now include the KiDS u and VIKING Z, Y, J, H and K_s bands and do multi-frame fitting. The corresponding changes are described in detail in Section 2.3.

In summary:

- We now process all nine KiDS and VIKING bands, using the science tiles for KiDS but individual detector chips preprocessed by Wright et al. (2019) for VIKING. This increases the number of catalogues from eight to 20 (two per band plus two joint ones).
- We now do multi-frame fitting, i.e. when there is more than one data match to the same galaxy in the same band, they are fitted jointly rather than separately (assuming the astrometric solutions to be the ground truth). For the preparatory work, the matches are still treated individually and also listed individually in the corresponding `BDInputs` tables; but for the `BDModels` tables, there is now only one entry per galaxy. This means that `CATAID` is a unique identifier again (along with the `OBJECTIDS` which we keep for compatibility with earlier versions).
- To reflect these changes, the `BDQUAL_FLAGS` and `CUTOOUT_FLAGS` have changed in meaning slightly, the `BEST_IMG` column (identifying the best image to use in the case of multiple matches) has been replaced with the `N_MATCH` and `N_FIT` columns (giving the number of data matches and the number of images actually used for fitting). The pixel scale (which varies for VIKING) has been added in the `PIXSC` column to give meaning to all columns with units of pixels. `JOINT_*` columns have been renamed into `GRI_*` (or whichever bands exactly they refer to, to differentiate between the different versions of joint model selection).
- We define the KiDS g, r and i bands as our core bands and transfer the segmentation maps defined on a stack of those core bands to all other bands.
- The galaxy segment sizes are generally larger than in `v04` due to a default change in `profoundProFound`. We use the new default of `SBdilate = 2`, but discarded the `skycut` default change.
- We re-calibrated the model selection with new visual inspections in all bands.

3.1.6 BDDecomp v06

We envision future versions of the DMU to provide simultaneous multi-band fits for all KiDS and VIKING observations of our sample.

3.2 Catalogue statistics

We begin the presentation of the contents of the `BDMODELS` catalogue with an overview of the numbers of galaxies classified in each model selection category, including outliers and skipped fits. For this, Section 3.2.1 presents two large tables (Table 3.1 for `v04` and 3.2 for `v05`) detailing the evolution of the numbers of galaxies at each step in our pipeline, with the most important information condensed in Figure 3.1. The `v04` versions of this are heavily based on Casura et al. (submitted). In Sections 3.2.2 and 3.2.3 we then investigate the outlier statistics and model selection statistics in more detail, with a particular focus on comparing `v04` and `v05`.

3.2.1 Statistics overview

Table 3.1 gives an overview of the fit and post-processing results for `v04`. Table 3.2 presents the analogous table for `v05`. Starting with our full sample (13096 galaxies from the combination of our main and SAMI samples, see Section 1.2.4), we show how the number of galaxies evolves through all steps of the pipeline in both versions. The results are split per-band and per-model where necessary. At some steps, we also include percentages of galaxies lost or remaining (grey font). In short, for `v04` we lose nearly 20% of our sample to masking and a further almost 10% to the flagging of bad fits; where the former is a random subset while the latter preferentially affects certain types of galaxies (e.g. mergers and irregulars). These fractions are similar for `v05` in the core bands, but increase for the KiDS u and the VIKING bands - especially the longer wavelength ones - due to missing data (VIKING only), additional masking, a higher fraction of PSF failures (u -band only) and a larger fraction of outliers mainly caused by the shallower data. The former three effects are explained in Section 2.3.3, while the latter one is investigated in more detail in Section 3.2.2. For `v05`, the numbers of fit failures also increases significantly in all bands (including the core bands), but this does not affect the final statistics since preferentially those fits failed that would not have been selected in the model selection anyway (Section 3.2.3).

Note that for both versions, we used stacked gri images for segmentation and masking (plus additional masking in the non-core bands), but then treated the galaxies independently in all bands except for the model selection, where we performed both a per-band and a joint version. Therefore, the column “joint gri ” in Table 3.1 always gives the number of galaxies that were “good” in all three bands (hence why numbers are generally lower), except for the model selection, where it shows the results of the joint model selection (cf. Section 2.1.3). The same is true for the corresponding “ gri ” and “ $ugriZYJHK_s$ ” columns in Table 3.2, except that for step 2a) we now - somewhat arbitrarily - give the sum of all data matches in all respective bands since there is no one-to-one correspondence between data matches in the KiDS and VIKING bands, so it is impossible to define the number of joint data matches in all

Table 3.1: Fit results for v04 of the BDDecom DMU: numbers (black) and percentages (grey) of galaxies remaining or lost at each step in our pipeline, split per-band and per-model where necessary.¹ Individual steps are numbered through, see text for details.

band	<i>g</i>			<i>r</i>			<i>i</i>			joint <i>gri</i>		
model (components)	1	1.5	2	1	1.5	2	1	1.5	2	1	1.5	2
number of:												
1) unique objects (galaxies)							13096					
2) images (independent fits)							14966					
3) images not masked							11989					
lost due to masking (%)							20					
4) successful PSFs	11838			11872			11946			11683		
lost due to PSF fails (%)	1			0.8			0.3			2		
5) successful fits	11837	11837	11831	11872	11870	11861	11946	11943	11945	11682	11678	11665
lost due to fit fails (%)	< 0.01	< 0.01	0.05	0	0.01	0.07	0	0.02	< 0.01	< 0.01	0.03	0.12
6) fits not flagged	10951	7122	8022	11025	8164	8759	11086	7620	7775	10680	6446	5870
not flagged/successful (%)	93	60	68	93	69	74	93	64	65	91	55	50
7) selected fits	8294	740	1743	7061	585	2935	7411	662	2663	7308	621	2009
selected/successful (%)	70	6	15	59	5	25	62	6	22	63	5	17
total number (per band) of:												
8a) good flagged skipped fits	10777 1061 3128			10581 1291 3094			10736 1210 3020			9938 1745 3283		
good f. s./all images (%)	72 7 21			71 9 21			72 8 20			66 12 22		
8b) good flagged skipped gal.	9722 935 2439			9545 1145 2406			9687 1059 2350			8998 1559 2539		
good f. s./unique objects (%)	74 7 19			73 9 18			74 8 18			69 12 19		

¹Based on information given in the *_BDQUAL_FLAG, *_OUTLIER_FLAG and *_NCOMP columns of the BDModels catalogue.

Table 3.2: Fit results for v05 of the BDDecom DMU: numbers (black) and percentages (grey) of galaxies remaining or lost at each step in our pipeline, split per-band and per-model where necessary.¹ Individual steps are numbered through, see text for details.

band	<i>u</i>			<i>g</i>			<i>r</i>			<i>i</i>		
model (components)	1	1.5	2	1	1.5	2	1	1.5	2	1	1.5	2
number of:												
1) unique objects (galaxies)	13096			13096			13096			13096		
2a) images (data matches)	14738			14966			14966			14966		
2b) objects with > 0 images	13096			13096			13096			13096		
lost due to missing data (%)	0			0			0			0		
3) objects not masked	10540			10768			10768			10768		
lost due to masking (%)	20			18			18			18		
4) successful PSFs	9141			10768			10752			10731		
lost due to PSF fails (%)	11			0			0.1			0.3		
5) successful fits	9087	8902	8504	10715	10640	9805	10706	10609	9845	10674	10566	9875
lost due to fit fails (%)	0.4	2	5	0.4	1	7	0.4	1	7	0.4	1	7
6) fits not flagged	6165	2408	880	9762	6096	5913	9803	7079	7000	9568	6527	4865
not flagged/successful (%)	68	27	10	91	57	60	92	67	71	90	62	49
7) selected fits	5888	254	57	7213	607	1691	5522	682	3201	7207	537	1686
selected/successful (%)	65	3	0.7	67	6	17	52	6	33	68	5	17
total number (per band) of:												
8) good flagged skipped fits	6199 2936 3961			9511 1249 2336			9405 1345 2346			9430 1298 2368		
good f. s./all objects (%)	47 22 30			73 10 18			72 10 18			72 10 18		

band	<i>Z</i>			<i>Y</i>			<i>J</i>			<i>H</i>		
model (components)	1	1.5	2	1	1.5	2	1	1.5	2	1	1.5	2
number of:												
1) unique objects (galaxies)	13096			13096			13096			13096		
2a) images (data matches)	46225			47678			95396			45918		
2b) objects with > 0 images	12437			12437			12436			12429		
lost due to missing data (%)	5			5			5			5		
3) objects not masked	9989			9969			9961			9945		
lost due to masking (%)	19			19			19			19		
4) successful PSFs	9942			9894			9898			9825		
lost due to PSF fails (%)	0.4			0.6			0.5			0.9		
5) successful fits	9899	9825	8943	9861	9721	9148	9848	9768	8929	9687	9534	8770
lost due to fit fails (%)	0.3	0.9	8	0.3	1	6	0.4	1	7	1	2	8
6) fits not flagged	8796	5305	3868	8261	4414	2538	8120	5081	2938	7368	4529	2299
not flagged/successful (%)	89	54	43	84	45	28	82	52	33	76	48	26
7) selected fits	6363	949	1395	6247	610	1239	6336	475	1257	5597	712	1024
selected/successful (%)	64	10	16	63	6	14	64	5	14	58	7	12
total number (per band) of:												
8) good flagged skipped fits	8707 1221 3168			8096 1786 3214			8068 1820 3208			7333 2405 3358		
good f. s./all objects (%)	66 9 24			62 14 25			62 14 24			56 18 26		

band	K_s			gri			$ugriZYJHK_s$		
model (components)	1	1.5	2	1	1.5	2	1	1.5	2
number of:									
1) unique objects (galaxies)	13096			13096			13096		
2a) images (data matches)	45993			14966 × 3			340846		
2b) objects with > 0 images	12438			13096			12426		
lost due to missing data (%)	5			0			5		
3) objects not masked	9955			10768			9738		
lost due to masking (%)	19			18			21		
4) successful PSFs	9800			10715			8082		
lost due to PSF fails (%)	1			0.4			13		
5) successful fits	9660	9516	8694	10597	10315	8635	7964	7149	4457
lost due to fit fails (%)	1	2	8	0.9	3	16	0.9	7	28
6) fits not flagged	6687	4136	1910	9371	5378	3464	4312	1443	357
not flagged/successful (%)	69	43	22	88	52	40	54	20	8
7) selected fits	5366	343	929	7202	604	1097	4051	37	141
selected/successful (%)	56	4	11	68	6	13	51	0.5	3
total number (per band) of:									
8) good flagged skipped fits	6638	3093	3365	8903	1788	2405	4229	3912	4955
good f. s./all objects (%)	51	24	26	68	14	18	32	30	38

¹Based on the *_N_MATCH, *_BDQUAL_FLAG, *_OUTLIER_FLAG and *_NCOMP columns.

bands. For the remaining rows, however, we then list the numbers of objects that were “good” in all bands again, resulting in very low numbers for the 9-band joint selection. Also note that for v05, all numbers except those in step 2a) refer to the numbers of galaxies (unique objects) since all images (data matches) of a galaxy were fitted jointly (per band). For v04, individual images were fitted separately, so the numbers in steps 2) to 8a) refer to individual data matches, not unique objects.

Explanations of each step as numbered through in Tables 3.1 and 3.2:

1. The full sample results from the combination of our main and SAMI samples (Section 1.2.4).
2. For KiDS bands, some galaxies have been imaged more than once due to overlap regions between the tiles. For VIKING bands, most galaxies have been covered by more than one exposure since we work at the individual chip level. For v04, we treat these duplicate observations of the same physical object independently throughout our pipeline, so numbers in all subsequent rows of Table 3.1 refer to individual images, not unique objects. For v05, all images of the same object in the same band are fitted jointly, so all numbers in Table 3.2 refer to unique objects. For the VIKING bands, about 5% of objects are not covered by any chip due to small gaps in the data coverage (Section 2.3.3).
3. For the core bands, we use the associated KiDS masks, combining the three bands. *Images* for which the central galaxy pixel is masked ($\sim 20\%$) are skipped during the fitting (Section 2.1.1). This results in $\sim 18\%$ of *unique objects* being skipped in the core bands. Since we use the *gri* segmentation maps for all other bands, too, these objects are automatically skipped in all bands. 1-2% of objects are additionally skipped in the KiDS *u* and VIKING bands due to the masking in the respective bands (Section 2.3.3).
4. For each image in each band, a PSF is then estimated by fitting nearby stars. If the PSF estimation fails, the image is skipped during the fit (Section 2.1.1). Note that technically, we estimate PSFs also for galaxies that are masked in step 3, but we do not list those here. For v05, an object is only skipped if all of its data matches are either masked or have a failed PSF. We count those for which all matches are masked in step 3 (irrespective of whether they also have failed PSFs or not); and those for which all matches failed the PSF as well as those for for which there was a mixture between matches that were masked (but have a PSF) and failed the PSF (but were not masked) here in step 4.
5. For each non-masked image with a successful PSF estimate, we attempt 3 fits: a single Sérsic (1), a pointsource + exponential (1.5) and a Sérsic + exponential (2). In v05, we attempt a joint fit of all images for each object that has at least one non-masked

3.2. CATALOGUE STATISTICS

image with a successful PSF estimate. Sometimes, the fit attempts fail with an error (Section 2.1.2).

6. Each fit (for each model independently) is passed through our outlier flagging process, identifying bad fits (Section 2.1.3). We further assess the differences between v04 and v05 in Section 3.2.2.
7. Of the non-flagged (i.e. good) fits, we then select the most appropriate one during model selection (Section 2.1.3). v04 and v05 model selection results are compared in detail in Section 3.2.3.
8. Summing up the selected fits for each model (step 7) gives the total number of good fits. The difference between the good and successful fits (step 5) stems from the outlier flagging. Skipped fits are due to missing data, masking, PSF or fit fails (steps 2b, 3, 4, 5). For v04, the sum of good, flagged and skipped fits in step 8a gives the total number of independent fits (step 2). Removing duplicate fits for the same physical objects gives the number of good, flagged and skipped galaxies in step 8b, which sum to the number of unique objects (step 1). For this, we always use the best available result for each galaxy, i.e. it is counted as “good” if at least one of the multiple fits was “good”. For v05, there is only one fit per galaxy, so the good, flagged and skipped fits from step 8 sum directly to the number of unique objects (step 1).

Figure 3.1 visualises the most important information given in Tables 3.1 and 3.2, namely the final number of objects classified in each category. For v04 (top panel), lighter bars in the background refer to individual fits (total 14966) with the number of unique galaxies (total 13096) overplotted. When several fits to the same galaxy were classified in different categories, we allocate it to the highest of those,¹ which is consistent with Table 3.1. For v05 (bottom panel), we only show the number of unique objects, since all images were fitted jointly.

$\text{NCOMP} = -999$ means the object was skipped (not fitted) because it is masked or the PSF estimation failed (usually because of large masked areas in the immediate vicinity of the object). For VIKING bands, objects can also be skipped due to small gaps in the sky coverage. $\text{NCOMP} = 1, 1.5$ or 2 indicates that this is a good fit classified as single, 1.5- or double component fit. $\text{NCOMP} = -1, -1.5$ or -2 indicates that this is a bad fit (outlier) which would have been classified as single, 1.5- or double component fit if it were not an outlier (most often these are mergers/irregular galaxies for which our models are not appropriate; or galaxies that are partly masked). We keep these three classes separate since automated outlier identification can never be perfect; and what should be considered a bad fit will depend on the use case. The

¹This means that a galaxy is classified as “outlier” if *all* fits to it are outliers and it is “skipped” only if *all* fits are skipped. Galaxies with good fits are allocated to the most complex model of the available fits (assuming that one of the images was deeper and allowed to constrain more components than the other(s)), while within the outlier categories we allocate it to the simplest model. Note that in Table 3.1 we only show the total number of flagged fits and do not split them into the different outlier categories.

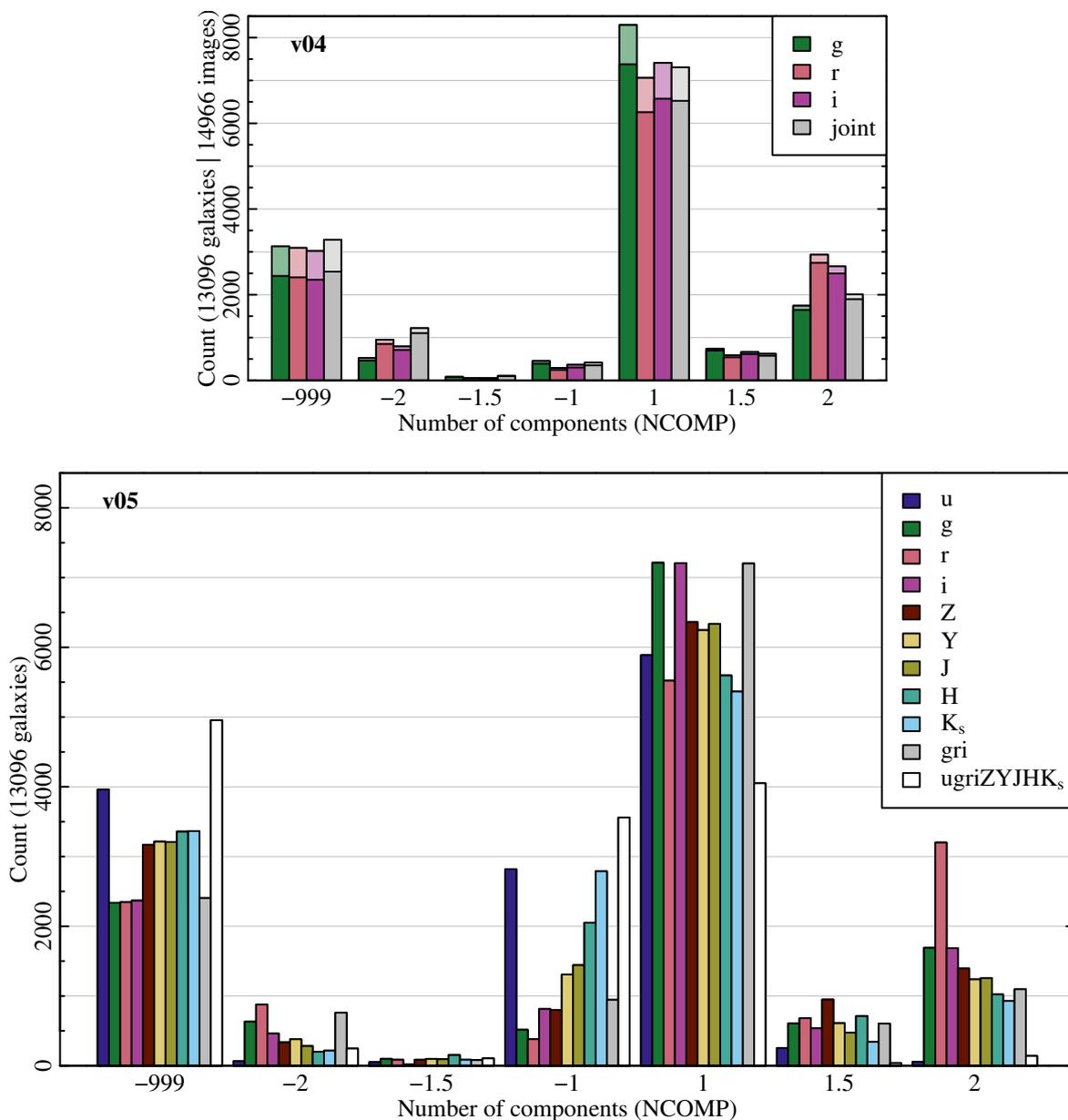


Figure 3.1: The number of components assigned in our model selection procedure for individual bands and the joint analysis for v04 (**top**) and v05 (**bottom**) of the pipeline. 1, 1.5 and 2 mean single Sérsic, point source bulge + exponential disk and Sérsic bulge + exponential disk models respectively. Negative values indicate that the chosen (best) fit was flagged as unreliable (mostly irregular or partly masked galaxies). -999 is assigned to skipped fits, either because the galaxy centre is masked (most cases), the galaxy falls into a gap in the data (VIKING bands only) or because the PSF estimation failed. For v04, the lighter (higher) bars show the number of images, whereas saturated bars indicate the number of unique objects in both figures; see text and Tables 3.1 and 3.2 for details. Horizontal grey lines in the background are placed at equal intervals in both figures to ease the direct comparison.

flagging of fits is hence only intended as a guide and all available information in the catalogue is retained for all fitted objects. We analyse Figure 3.1 in more detail in Sections 3.2.2 and 3.2.3, where we compare the outlier flagging and model selection statistics for v04 and v05 of the BDDcomp DMU.

Examples of fits classified in each category can be found throughout Chapter 2, see e.g. Figure 2.2 for a double component fit, Figure 2.21 for a 1.5-component object, Figure 2.29 for a single Sérsic galaxy and Figure 2.39 for an outlier.

3.2.2 Outlier statistics

In this section, we analyse the numbers of outliers in Figure 3.1 and the corresponding Tables 3.1 and 3.2 in more detail. Since the outlier flagging was calibrated on v04 (cf. Section 2.3.4), we put particular emphasis on comparing the v05 results to those of v04 for the core bands (g , r , i) and compare all other bands against those three within v05. The differences in the numbers of skipped fits ($\text{NCOMP} = -999$) are explained in Section 2.3.3. In short, the g , r and i bands show the same number of skipped fits in v04 and v05 since that is mainly due to masking, which did not change. The u -band has a higher number of skipped fits due to its smaller footprint, shallower data and the additional masking. VIKING bands have higher numbers of skipped fits mainly due to small gaps in the data coverage. All objects that are skipped in at least one of the bands are counted as skipped for the joint categories, which is the reason for the higher number of skipped objects for the 9-band joint analysis. The differences in the model selection are investigated in Section 3.2.3. In this section, we focus on differences in the outlier categories, i.e. $\text{NCOMP} = -1$, -1.5 and -2 .

The core bands in general have similar numbers of flagged fits in v05 with respect to v04. The VIKING Z -band is also comparable to the core bands. For the longer wavelength bands and the KiDS u -band, the number of outliers steadily increases. To investigate the origin of this further, Table 3.3 gives the percentages of fits flagged according to each criterion for all bands in v04 and v05. To decouple the analysis from differences in the model selection, we show the total number of outliers, i.e. categories $\text{NCOMP} = -2$, -1.5 and -1 combined. A detailed description of all criteria is given in Section 2.1.3, where the corresponding r -band v04 percentages are also indicated. All values are given in percent of the total number of non-skipped fits; counting only those in the respective model selection category.² Note that the total number of outliers is smaller than the sum of the individual flag values since bad

²E.g. the 7.21% of u -band fits with an irregular segment are composed of the fraction of objects that had this flag raised for their single Sérsic fit *and* were classified as single Sérsic fits or single Sérsic outlier (absolute value of NCOMP equal to one), plus the corresponding fractions of 1.5- and double component fits. For fractions of flagged fits for each model independently, see step 6 of Tables 3.1 and 3.2. Note also that the total number of flagged fits in step 8 of those tables is given as a fraction of the total number of objects, whereas Table 3.3 gives all values as percentages of non-skipped objects.

Table 3.3: The percentage of fits flagged as outlier according to each criterion (see text for details) in all bands in v04 and v05 of the BDDecomp DMU. All values are in per cent of the total number of non-skipped fits, except for the “extreme B/T ratio” criterion, which is in per cent of the total number of non-skipped 1.5- and double component fits (since it does not apply to single Sérsic fits). Grey font indicates cautionary flags that are not considered during the final outlier flagging. Bold font in the v05 table highlights major changes in the fractions relative to v04 (for g , r and i bands) and relative to the (v05) r -band for all other bands.

(a) Outlier statistics in v04.

flag	g	r	i
irregular segment	5.48	5.37	5.27
extreme B/T ratio	0.00	0.05	0.04
numerical problems	0.08	0.20	0.17
rel. param hit limit	3.67	5.82	4.94
any param hit limit	4.60	6.46	8.09
small or large error	2.14	2.06	4.13
poor χ^2 statistics	0.04	0.14	0.05
position offset $> 2''$	0.42	0.35	0.39
position offset $> 1''$	1.50	1.29	1.41
flux in seg $< 20\%$	1.42	1.42	1.29
flux in seg $< 50\%$	8.04	9.32	9.33
total outliers	8.95	10.87	10.13

(b) Outlier statistics in v05.

flag	u	g	r	i	Z	Y	J	H	K_s
irregular segment	7.21	6.85	6.83	6.68	7.06	6.81	7.88	7.65	7.71
extreme B/T ratio	0.36	0.07	0.04	0.10	0.13	0.18	0.19	0.50	0.98
numerical problems	1.38	0.31	0.62	0.34	0.36	0.83	0.64	1.56	2.90
rel. param hit limit	25.40	4.91	5.60	5.67	5.25	11.18	10.99	17.11	23.60
any param hit limit	28.66	5.61	6.23	6.71	6.42	13.47	13.01	20.48	28.78
small or large error	28.86	5.73	6.26	6.73	6.51	13.37	12.91	20.45	28.94
poor χ^2 statistics	0.10	0.09	0.17	0.11	0.08	0.13	0.10	0.15	0.15
position offset $> 2''$	1.02	0.39	0.32	0.34	0.48	0.54	0.59	1.12	1.93
position offset $> 1''$	3.65	1.49	1.25	1.40	1.84	2.08	2.17	3.20	4.59
flux in seg $< 20\%$	1.54	1.09	1.27	1.14	0.86	0.84	1.13	1.51	1.71
flux in seg $< 50\%$	6.23	5.25	6.45	5.77	3.88	3.59	4.88	5.51	6.39
total outliers	32.14	11.61	12.50	12.10	12.30	18.06	18.41	24.70	31.79

fits frequently fall into multiple outlier categories. Grey font indicates cautionary flags that are not taken as criteria for flagging outliers. We only list these for completeness and do not comment on them further. In Table 3.3b, bold font highlights values that show large differences with respect to the v04 percentages for the g , r and i bands; and with respect to the v05 r -band for all other bands.

The fractions of outliers in the g , r and i bands in v05 are approximately 12%, 13% and 12%

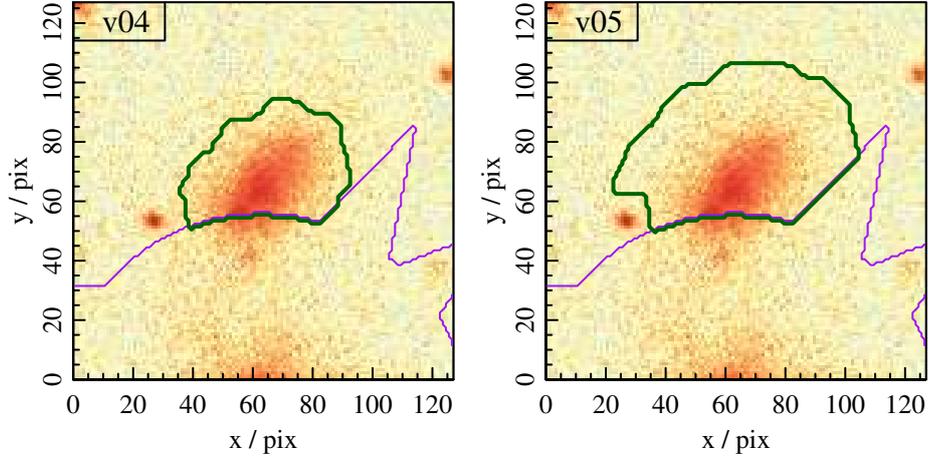


Figure 3.2: The r -band segmentation map (dark green contour) for galaxy 14912, which was flagged as having an irregular fitting segment in v05 (**right**), but not in v04 (**left**). Purple contours indicate masked regions.

respectively. These are 2% to 3% more outliers than in v04. The main driver of this is the “very irregular fitting segment (irregular segment)” criterion (+1.5%). The g and i bands also seem to have a higher number of (reliable) parameters hitting their fit limits (+1%), which is not observed in the r -band. In addition, there are increases in the numerical problems and to a lesser extent in the extreme B/T ratio objects, but the overall numbers of fits classified as outliers according to these criteria are so low that they do not affect the final outlier fraction much.

There are no differences in the data used between v04 and v05 and the numbers of skipped fits are nearly identical for the core bands. The only possible reasons for the increased fraction of outliers are differences in the processing, namely the increased segment sizes and the joint fitting of all images (Section 2.3). Considering only galaxies with a single data match (in the core bands) does not change any flag values significantly. This suggests that the joint fitting is not the cause of the increased outlier fractions and instead it is the slightly different (larger) segments. Since the segments used are the same in all bands, the fraction of fits flagged according to the “irregular segment” criterion are very similar across all bands (with slight differences between bands arising from the definition of this criterion, which also depends on the fit itself, cf. Section 2.1.3).

Visual inspection of a number of fits, which were flagged for having an irregular segment in v05 but not in v04, showed that they were all borderline cases, mostly with large fractions of the segment cut off by a masked region or neighbouring object. An example is shown in Figure 3.2 for the single Sérsic object 14912: both segments (dark green contours) are cut off by the bright star mask (purple contour). The segment for v05 (right panel) is not “worse” than that for v04 (left panel), suggesting that the “irregular segment” flag could benefit from

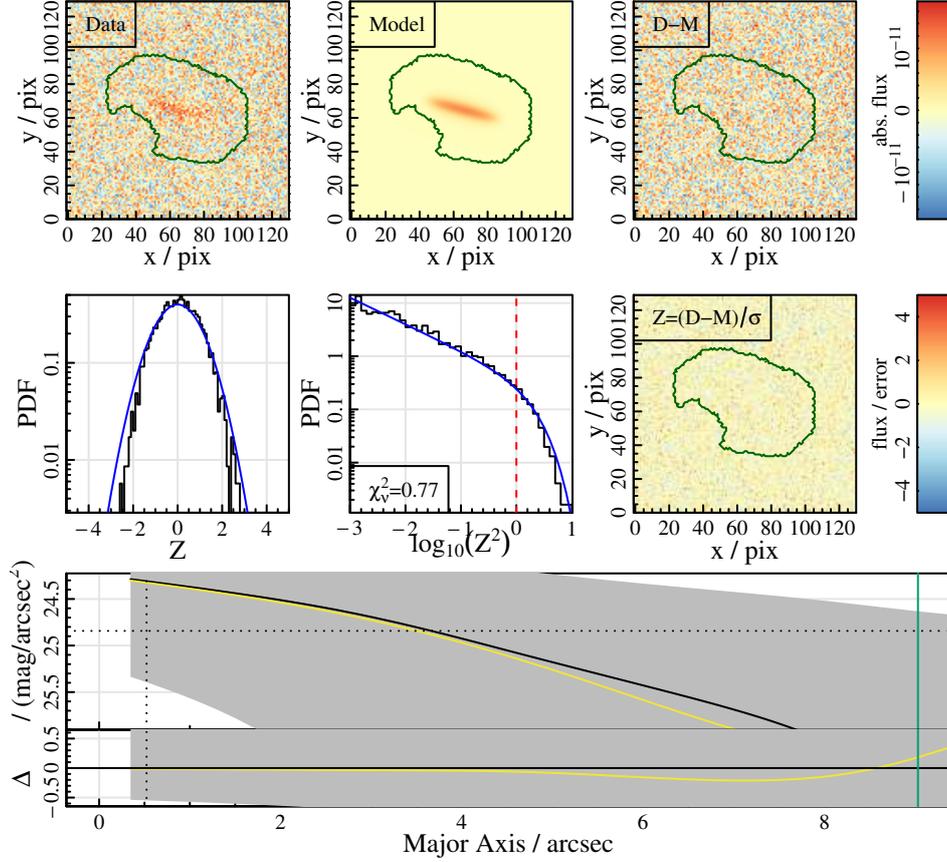


Figure 3.3: The u -band single Sérsic fit to galaxy 16537, which is classified as an outlier since it hit its fitting limits in several parameters. Panels in the top two rows are the same as those in Figure 2.2, while the bottom row shows the one-dimensional fit only, corresponding to the rightmost panel of the bottom row in Figure 2.2.

a re-calibration in `v05`. However, since it only affects a relatively small number of objects, all of which are borderline cases, we decided against a re-calibration at this stage. Since we provide all modelling results and flag values for all objects in the catalogue, this can still be revised in the future.

The fraction of outliers in the VIKING Z -band (12%) is comparable to that in the core bands, while the KiDS u and the VIKING Y, J, H, K_s bands have significantly higher fractions, ranging between 18% and 32%. Part of the reason for these higher fractions are the lower numbers of non-skipped fits (i.e. lower normalisation), but as evident from Figure 3.1, the absolute numbers of outliers increase, too. In all cases, the dominant reason for this are (reliable) parameters hitting their fit limits, with maybe a slight secondary contribution from the position offset flag (Table 3.3b). All other flag values show only small increases that can be attributed to the lower normalisation. For the VIKING bands, the fractions of fits flagged generally increases from Z through to K_s for all flag values (in part again due to the higher

3.2. CATALOGUE STATISTICS

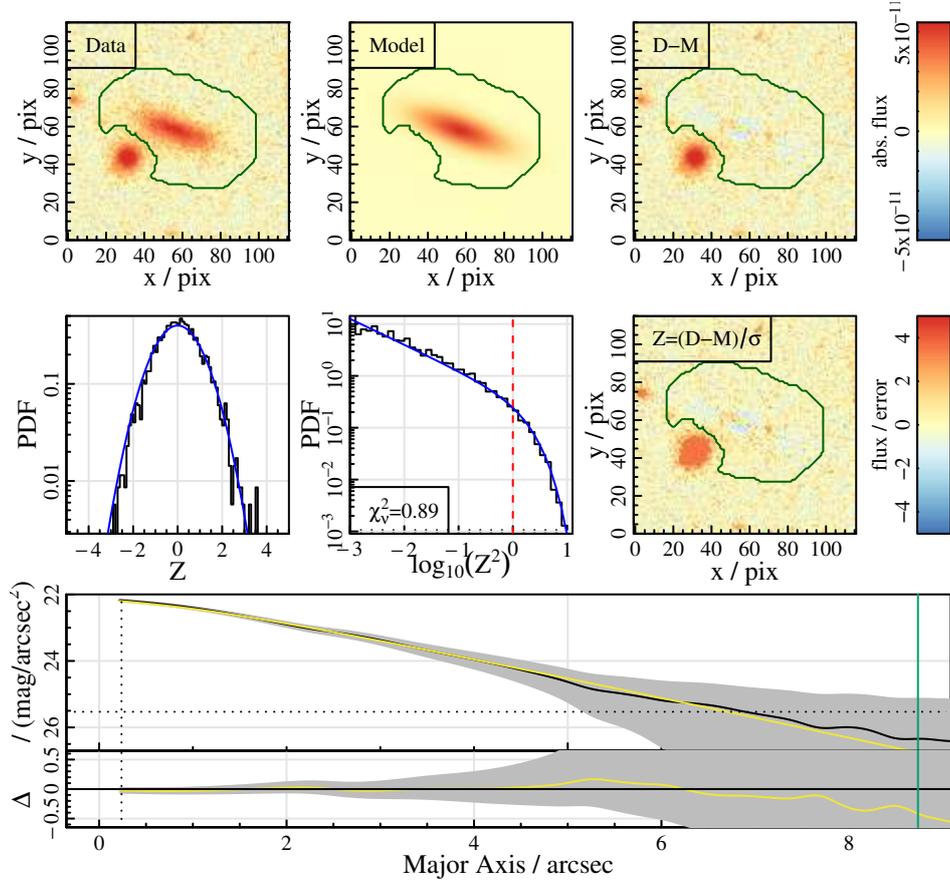


Figure 3.4: The r -band single Sérsic fit to galaxy 16537, classified as a single Sérsic object, for direct comparison to Figure 3.3.

number of skipped fits from Z through to K_s).

The parameter that hits its fit limit most frequently for single Sérsic fits is the axial ratio, followed by the Sérsic index (with a large overlap between the two) and to a lesser extent the effective radius. Almost all of the parameters hitting their fit limits also have a suspiciously large error (meaning they are an outlier in their respective error distribution), which is a cautionary flag. Both of these are indications that the parameters are ill-constrained and can be explained by the decreasing data quality (in particular depth) as a function of wavelength for VIKING (Section 1.2.3) and the KiDS u -band (Section 1.2.2; which was also the reason why the u -band was not considered during v04 and is not used as a “core” band in v05). The shallowest bands (u and K_s) are hence not only too shallow to constrain two components in most cases (cf. Section 2.3.4), but also too shallow to constrain a single Sérsic fit for approximately one quarter of our sample and therefore become flagged as outliers. As we discuss in Section 2.3.4, this is a fundamental problem that can only be solved with deeper data or simultaneous multi-band fits.

For the same reason, the fractions of non-flagged fits provided in Table 3.2 decrease drastically as a function of data quality and model complexity. For example, in the u -band, 90% of the attempted double component fits are flagged as outliers. The vast majority of these, however, were not classified as double component objects since the data can equally well be represented by a single Sérsic fit. What drives the higher fraction of outliers in these shallow bands are therefore mostly the objects for which even the single Sérsic fits cannot be constrained anymore. Figure 3.3 shows an example of such a fit where the u -band data is of insufficient quality to constrain a single Sérsic model. The fit was flagged as an outlier since it hit its parameter limits for the axial ratio and Sérsic index. For comparison, we show the r -band version of the same galaxy in Figure 3.4. This fit was classified as a (good) single Sérsic fit.

3.2.3 Model selection differences

We now turn to the model selection differences between v04 and v05, as shown in Figure 3.1 and Tables 3.1 and 3.2. The model selection differences between individual bands in v05 have already been discussed in Section 2.3.4 where we also give full confusion matrices against visual inspection for the model selection in all bands and the joint model selection in v05. Corresponding v04 confusion matrices are provided in Sections 2.1.3 and 2.2.7. Here, we supplement these by a new type of confusion matrix, namely that between the v04 and v05 model selections for the core bands. For the most direct comparison, we consider only those galaxies that had a single data match and were not skipped in any bands in either catalogue (leaving 8900 objects), and to decouple the analysis from outlier flagging differences (Section 3.2.2), we take absolute values of N_{COMP} .

Table 3.4 shows the resulting confusion matrices for all bands and the joint gri model selection, with bold colours highlighting the galaxies that were classified in the same category in both catalogue versions. For better overview, we also show the sums of each row and column (total numbers of objects classified in each model selection category for each version independently) and the sum of the diagonal, i.e. the total number of objects for which the model selection agrees between the two versions. All values are in percent of the total number of objects.

The overall agreement between the model selections in v04 and v05 ranges between $\sim 81\%$ for the joint model selection to 90% for the g -band. The largest rate of confusion is observed between single and double component objects. For the r -band, there are more double and 1.5-component fits in v05 than in v04 (also evident from Figure 3.1). This is reversed in the i -band and the joint model selection, where the number of single component fits is higher in v05. For the g -band, the relative numbers of 1.5-/double component fits and single Sérsic objects is approximately the same in both versions.

Comparing the ΔDIC_{1-2} cuts shown in Tables 2.6 and 2.8, we can see that for the g -band, the

3.2. CATALOGUE STATISTICS

Table 3.4: The confusion matrices between the v04 and v05 model selections in percent of the total number of objects that have a single data match and were not skipped in any of the core bands in either catalogue. Bold font highlights those galaxies classified in the same category for both versions. Total fractions of objects classified into each of the three categories (i.e. the sum of each row/column) are also given; as well as the total percentage of objects for which the model selection is in agreement between the two versions (sum of the diagonal).

<i>g</i> -band					<i>r</i> -band			
v04 NCOMP	v05 NCOMP			total	v05 NCOMP			
	1	1.5	2		1	1.5	2	total
1	69.2	0.9	2.3	72.4	53.1	2.1	5.0	60.2
1.5	0.8	4.2	2.4	7.4	0.2	3.0	2.2	5.5
2	1.9	1.5	16.8	20.2	1.6	2.0	30.8	34.3
total	71.8	6.7	21.5	90.2	54.8	7.1	38.1	86.9

<i>i</i> -band					<i>gri</i> joint			
v04 NCOMP	v05 NCOMP			total	v05 NCOMP			
	1	1.5	2		1	1.5	2	total
1	63.3	0.1	0.3	63.8	63.4	0.4	0.8	64.6
1.5	2.5	3.5	0.2	6.2	1.6	2.8	2.0	6.3
2	8.9	1.8	19.4	30.1	11.1	3.3	14.7	29.1
total	74.7	5.3	19.9	86.2	76.0	6.5	17.5	80.9

Table 3.5: The same as Table 3.4 but using the v04-calibrated Δ DIC cuts for the v05 model selection.

<i>g</i> -band					<i>r</i> -band			
v04 NCOMP	v05 NCOMP (v04 cal.)			total	v05 NCOMP (v04 cal.)			
	1	1.5	2		1	1.5	2	total
1	69.3	1.5	1.7	72.4	57.7	0.7	1.8	60.2
1.5	0.8	6.1	0.5	7.4	0.3	4.6	0.5	5.5
2	1.6	1.7	16.9	20.2	1.9	2.1	30.3	34.3
total	71.6	9.3	19.1	92.3	60.0	7.4	32.6	92.6

<i>i</i> -band					<i>gri</i> joint			
v04 NCOMP	v05 NCOMP (v04 cal.)			total	v05 NCOMP (v04 cal.)			
	1	1.5	2		1	1.5	2	total
1	61.4	0.9	1.5	63.8	62.1	1.1	1.4	64.6
1.5	0.5	5.3	0.4	6.2	0.6	5.4	0.3	6.3
2	1.7	1.7	26.6	30.1	2.9	3.9	22.2	29.1
total	63.6	7.9	28.5	93.2	65.6	10.4	24.0	89.7

Δ DIC cuts are similar in v04 and v05.³ For the r -band, the v05 cut is slightly lower than in v04, although still within the uncertainty region. The i -band and the joint gri model selection both have significantly higher Δ DIC₁₋₂ cuts in v05 than in v04, with no overlap between the “unsure” regions. This explains the low confusion rate between v04 and v05 in the g -band, the slightly higher number of double component fits in v05 in r , and the significantly lower number of double component fits in the i -band and joint model selection.

Since the procedure for the model selection calibration is identical in v04 and v05, these differences must stem from differences in the fits themselves or in the manual (re-)calibrations. To judge the effect of differences in the fits (either from the random nature of the MCMC sampling or from systematic differences e.g. due to the larger segments), we additionally performed the model selection on v05 using the Δ DIC cuts from v04. The resulting confusion matrices between the model selection categories for v04 and v05 (with v04 calibration) are given in Table 3.5. The total fractions of fits classified in the same model selection category (sum of the diagonals) rise to 92.3%, 92.6%, 93.2% and 89.7% respectively for the g, r, i individual and gri joint model selection, compared to 90.2%, 86.9%, 86.2% and 80.9% in Table 3.4. The relative fractions of single Sérsic and 1.5-/double component fits in the two versions are now within 0.2% of each other for all bands ($\sim 1\%$ for the joint selection). The fractions of 1.5-component fits are about 2% higher in v05 than in v04 for all bands (4% for the joint gri selection), with the double component fractions correspondingly lower, despite using the same Δ DIC cuts.

This indicates that confusion between models on the level of approximately 7-8% (10% for the joint selection) stems from differences between the fits themselves (to the same galaxies), either of random nature or due to the different fitting procedures in v04 and v05, most likely the larger segments (see Section 2.3). This introduces no systematic differences in the classification between the single Sérsic and 1.5-/double component models, but generally increases the numbers of 1.5-component fits relative to the double component fits. The reasons for the latter remain to be investigated.

The differences between Tables 3.4 and 3.5 are caused by differences in the Δ DIC cuts for v04 and v05, which in turn must be caused by differences in the visual calibrations. Reasons could be statistical fluctuations due to the randomly selected calibration samples of galaxies (that differ in v04 and v05) as well as human error, both amplified by the relatively small sample size of 200 objects per band in v05. Since the visual classifications were performed several years apart, there could also be systematic differences due to a deepened understanding of the fit results, modelling limitations, systematic uncertainties and model selection caveats based on the detailed investigations of the v04 results in the meantime. In particular, the v05

³Note that the absolute DIC values in v05 are generally a factor of about two larger than in v04 due to the larger segments. However, the DIC *differences* are comparable since all models are equally affected by the larger segments. We therefore would expect the Δ DIC cuts to be similar in both catalogue versions.

3.2. CATALOGUE STATISTICS

calibration put special emphasis on classifying as few fits as possible in the “unsure” or “outlier” categories since that further reduces the already relatively low number of objects available for model selection calibration. We also explicitly de-coupled model selection from outlier rejection, the identification of swapped fits and segmentation failures. If, for example, the double component fit (within the segment) is significantly better than the other two models, in the v05 calibration we labelled this object as a double component object regardless of whether it is swapped, would be better classified as an outlier or the segmentation map failed (e.g. containing secondary objects). This means that the two categories “unsure” and “unfittable/outlier” became equivalent (as they always were in the ΔDIC cut calibration, where both are ignored). Conversely, if the object clearly has a bulge but the 1.5- and double component models failed to fit it and are not better than the single Sérsic fit, then the object was labelled as “single”. While the general notions were the same during the v04 visual inspections - driven by the limitations of a model selection based on a statistical measure - the criteria were not as sharply defined.

To investigate such manual calibration differences, Figure 3.5 shows the visual classifications in v04 (blue) and v05 (orange) as a function of the DIC difference between the single and double component models, where we see most confusion in Table 3.4. Note that the galaxies shown are not the same and the number of objects is higher for v04 (~ 700 per band; 2000 in the joint selection) than for v05 (200 per band, 600 in the joint selection). The corresponding calibrated ΔDIC_{1-2} cuts with their “unsure” regions (solid and dashed lines) are also indicated. The visual inspection categories that are relevant for calibrating the ΔDIC_{1-2} cut are “single”, “double” and “1.5 or double” (cf. Section 2.1.3); we show the other categories for completeness only.

Overall, the distribution of blue and orange points in each panel is similar, indicating that there are no fundamental differences in the visual calibrations for v04 and v05. For the g and r bands, the DIC difference cuts are consistent between the two versions, as already mentioned above when comparing Tables 2.6 and 2.8. Hence, the confusion rates in these bands is low with especially the g -band coming close to the confusion from fit differences alone (Tables 3.4 and 3.5). The cuts in the i -band and joint selection are not consistent, although they are still relatively close to each other compared to the vast range the DIC differences observed. Therefore, the confusion rates between v04 and v05 in these bands are still moderate despite the numerical differences in the ΔDIC cuts (Tables 3.4 as well as 2.6 and 2.8).

The higher cut in v05 in the i -band appears to be caused by a cluster of orange points around a ΔDIC_{1-2} value of 10^3 , that has no blue counterpart (remember that there are a factor of 3.5 more blue points than orange ones) and is also not observed in the g and r bands. It is possible that this is simply caused by an unfortunate selection of galaxies or random human error and would be downweighted by a larger sample of objects used for calibration. Alternatively, it

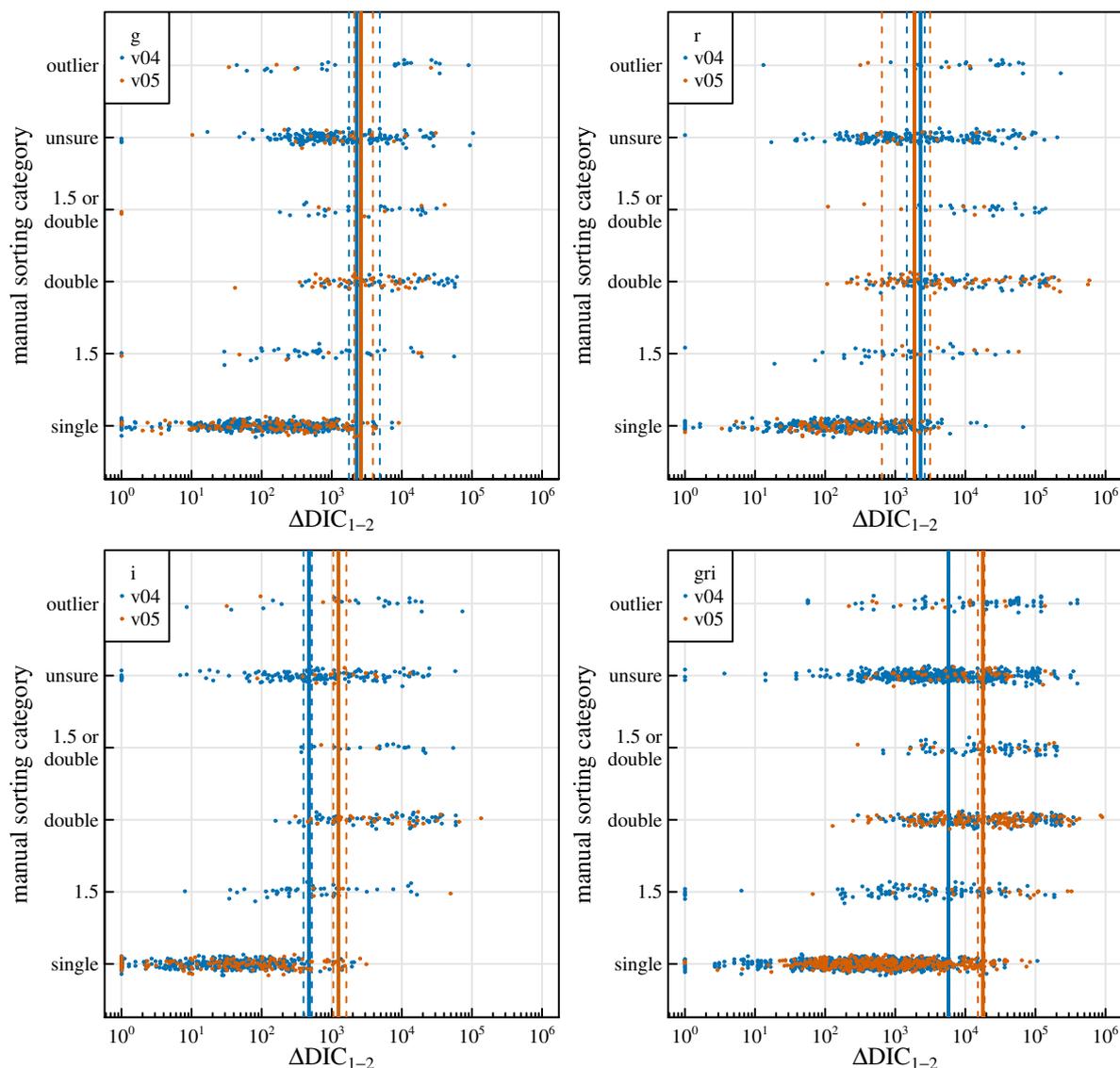


Figure 3.5: The visual classification during manual sorting plotted against ΔDIC_{1-2} for v04 (blue points) and v05 (orange points) in the g (**top left**), r (**top right**) and i (**bottom left**) bands as well as for the gri joint model selection (**bottom right**; with DICs summed for all bands). Vertical solid and dashed lines indicate the corresponding ΔDIC_{1-2} cuts for v04 (blue) and v05 (orange).

could be systematic due to the sharpened criteria for the v05 calibration. Visual inspection of this cluster of datapoints reveals them all to be borderline cases. Figures 3.6 and 3.7 show an example, where the double component fit is better than the single Sérsic fit, but it is debatable whether it is “significantly” better in the sense that it justifies the increased number of parameters. In the v05 calibration we decided that this was not the case, but maybe the decision would have been different during the v04 calibration; or even when inspecting the same galaxy at a later or earlier time in the v05 calibration, as there is always a certain

3.2. CATALOGUE STATISTICS

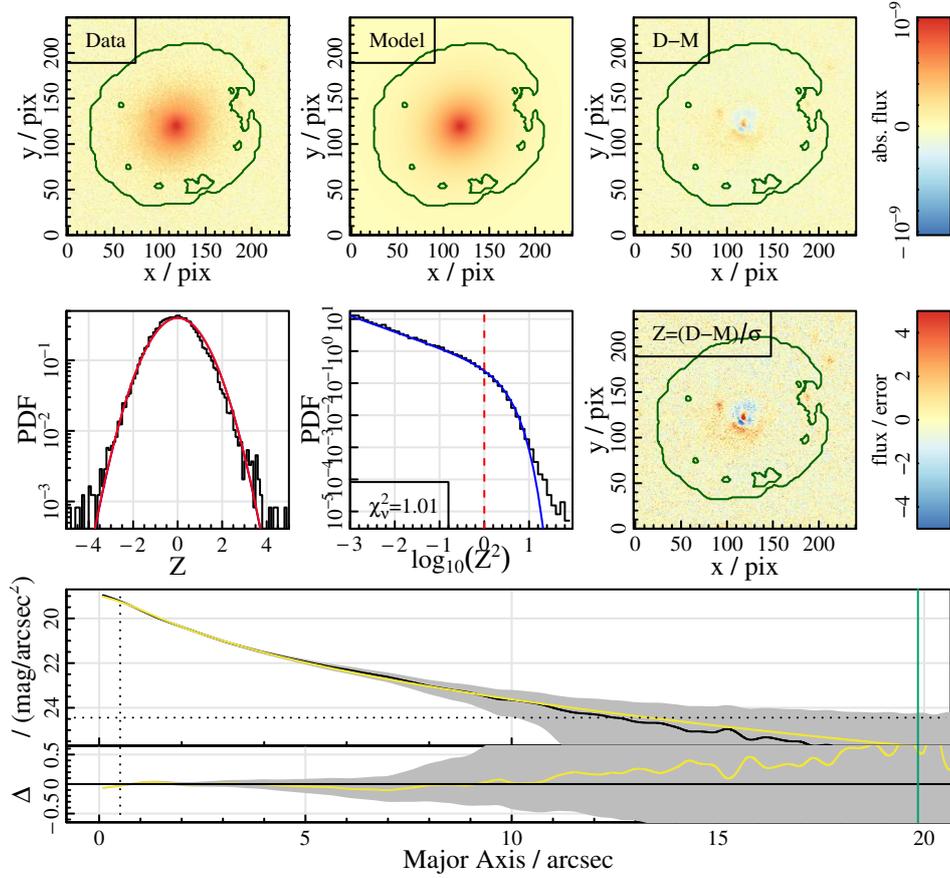


Figure 3.6: The single Sérsic fit to galaxy 585561 in the i -band, which was visually classified as a single Sérsic object despite its relatively high ΔDIC_{1-2} value of ~ 1500 . Panels in the top two rows are the same as those in Figure 2.2, while the bottom row shows the one-dimensional fit only, corresponding to the rightmost panel of the bottom row in Figure 2.2.

amount of scatter associated with visual inspection.

The last point of the model selection that we would like to return to are the high confusion rates against visual inspection for the g and r bands in v05 that were apparent from Table 2.7. As a reminder, the v05 g and r bands and the two versions of the joint model selections have approximately 15%, 17%, 17% and 18% of fits classified wrongly by the automated procedure compared to visual inspection, mostly due to $\text{NCOMP} = 1$ objects that were visually classified as doubles. For all other bands and also for the v04 versions of the same bands, the confusion rate is 9% at most. Based on the above analysis of the v04 and v05 model selection differences, we believe that this is mainly due to the lower number of fits classified as “unsure” in the v05 model selection.

Going back to Figure 3.5, the objects responsible for the high confusion are orange points in the “double” manual sorting category but to the left of the orange ΔDIC cut. There is a much higher number of those in the g and r bands than in i , partly due to the generally

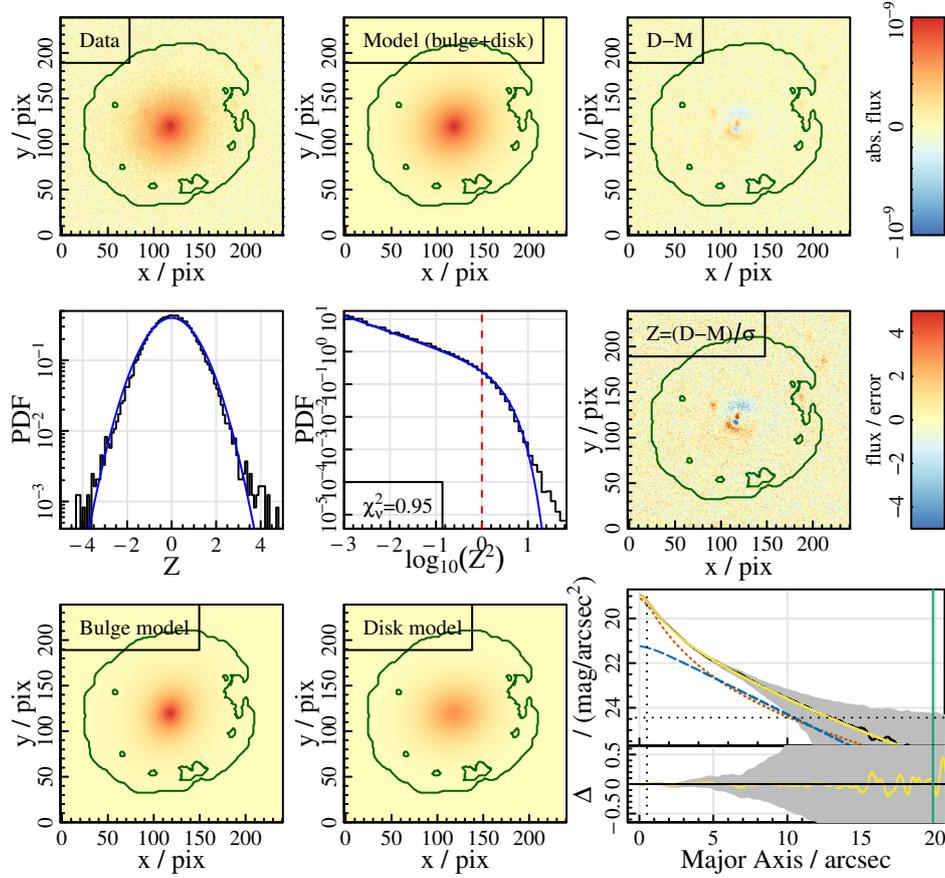


Figure 3.7: The double component fit to galaxy 585561 in the i -band, for direct comparison to Figure 3.6. Panels are the same as those in Figure 2.2.

lower number of double component objects in i . Most blue points ($v04$ model selection) in that region of ΔDIC_{1-2} instead populate the “unsure” category, where there is generally a very low number of orange points due to our explicit attempt to classify as few galaxies as possible as “unsure” in the $v05$ calibration. From this - and the consistency of the $v04$ and $v05$ ΔDIC_{1-2} cuts in the g and r bands - we conclude that the automated model selection itself is not “worse” in $v05$ than it was in $v04$. Instead, the higher confusion statistics is simply caused by a large number of ambiguous objects that were more or less randomly assigned to one of the two categories in the $v05$ visual inspection, whereas they were classified as “unsure” and ignored during $v04$ model selection.

Considering all of the above differences and uncertainties, we conclude that the confusion rates between the model selections in $v04$ and $v05$ as well as those between the $v05$ automated and manual classifications are acceptable. The quality of the $v05$ model selection is comparable to that of $v04$ despite the lower number of visual classifications per band.

3.3 Parameter distributions

After the analysis of the outlier rejection and model selection statistics, we take a closer look at the fitted structural parameters, considering only good fits within the respective model selection category for each band individually. We do not use the joint model selection or take a combined sample across all bands due to the differences in data quality between the bands that result in vastly different relative numbers of galaxies in each category. Section 3.3.1 shows the parameter distributions for v04 of the catalogue, taken directly from Casura et al. (submitted). Section 3.3.2 adds the new v05 versions (not included in Casura et al. submitted or elsewhere) and compares them to those of v04. Finally, Section 3.3.3 focuses on galaxy and component colours, again based on Casura et al. (submitted).

3.3.1 v04 parameter distributions

Figure 3.8 shows the distribution of the main parameters - magnitude, effective radius and Sérsic index - for v04 of the BDDcomp DMU in all three bands (g , r and i) for single Sérsic fits, bulges and disks. The single Sérsic fit distributions are shown for all galaxies with $\text{NCOMP} > 0$ (i.e. all non-outliers) in black and for those galaxies which were actually classified as single component systems ($\text{NCOMP} = 1$) in yellow. Red dotted and blue dashed lines show bulges and disks, respectively. For disks we show the 1.5-component fits and double component fits combined (i.e. the 1.5-component parameters for objects with $\text{NCOMP} = 1.5$ and double component parameters for those with $\text{NCOMP} = 2$ added into one histogram); the Sérsic index is not shown since it was fixed to 1. Bulge magnitudes are also shown for 1.5- and double component fits combined; effective radii and Sérsic indices are only shown for the double component fits since they do not exist in the point source model. The legend indicates the numbers of objects in each histogram, which can also be inferred from Table 3.1. Magnitudes and effective radii are truncated at the segment radii which we found to give more robust results than using the Sérsic values extrapolated to infinity (see Sections 2.1.3, 4.1.4 and 4.2.4).

The first thing apparent from Figure 3.8 is that the distributions in the three bands are generally very similar, which is reassuring given that the fits were performed independently. Looking at the first column, the single Sérsic number counts increase up to a sharp drop just before 20 mag in all bands, which is not surprising given the GAMA survey limit of 19.8 mag. The faintest of these objects are all classified as single component galaxies (the yellow lines are on top of the black lines), while some of the brighter objects are successfully decomposed into bulges and disks. Disks are generally slightly brighter than bulges. The bulges show a second, smaller peak at very faint magnitudes which we found to be the ones from the 1.5-component fits (unresolved, faint bulges). There is a slight trend for magnitudes to become brighter moving from g to r to i for all components, as expected from typical galaxy colours.

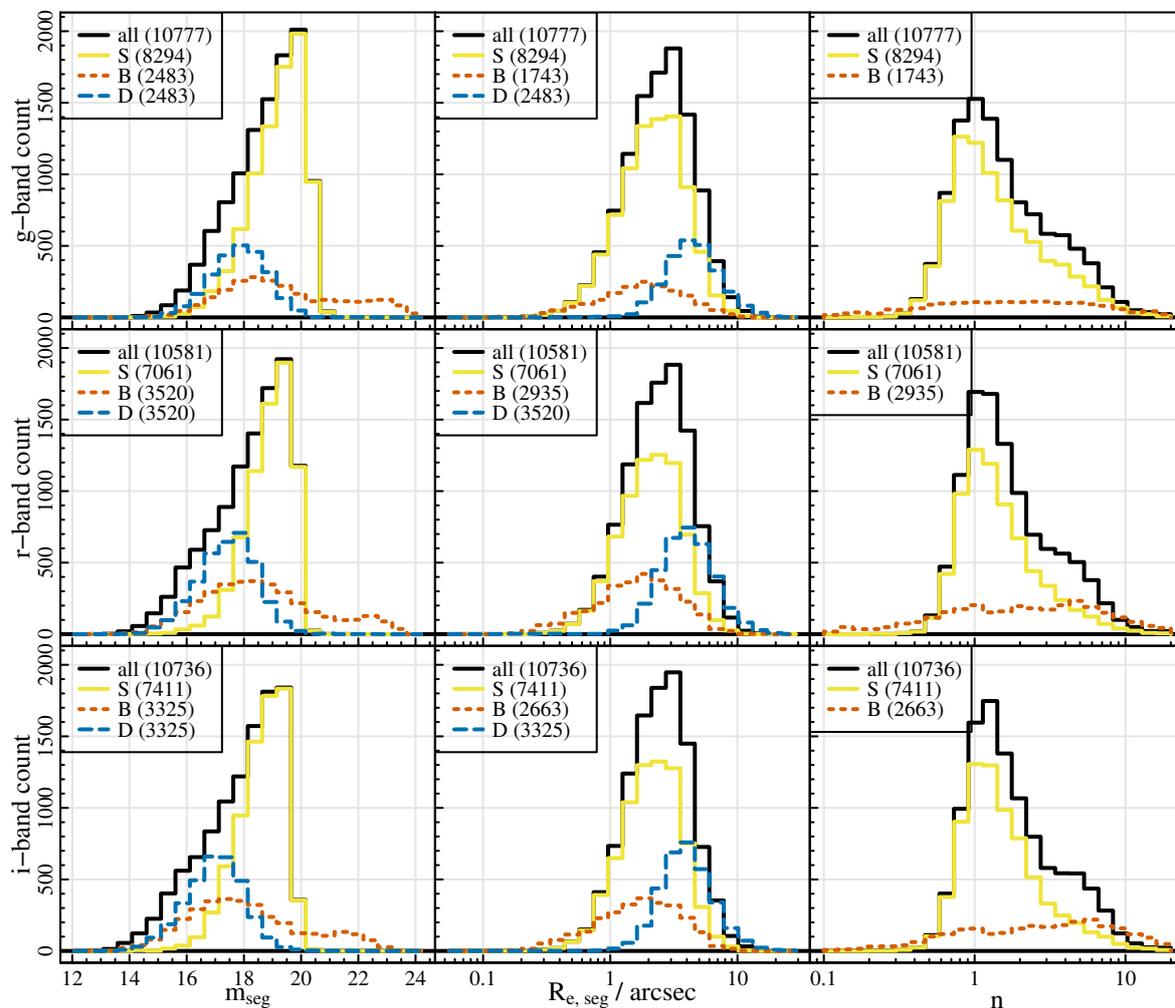


Figure 3.8: The distribution of the main parameters (limited to segment radii) for all bands and models in v04. Left, middle and right columns show magnitude, effective radius and Sérsic index while top, middle and bottom rows show the g -band, r -band and i -band respectively. The solid yellow lines are the single Sérsic (S) values for those galaxies which were classified as single component systems, dotted red and dashed blue lines show bulges (B) and disks (D), respectively, for those objects classified as 1.5- or double component systems. For reference the solid black line shows the single Sérsic fits for all galaxies with $\text{NCOMP} > 0$ (i.e. including those classified as 1.5- or double component systems). The number of objects in each histogram is given in the legends, where the number of bulges and disks differs for effective radii and Sérsic indices because these parameters do not exist for 1.5-component fits (point source bulge). We do not show disk Sérsic indices since they were fixed to 1 (exponential).

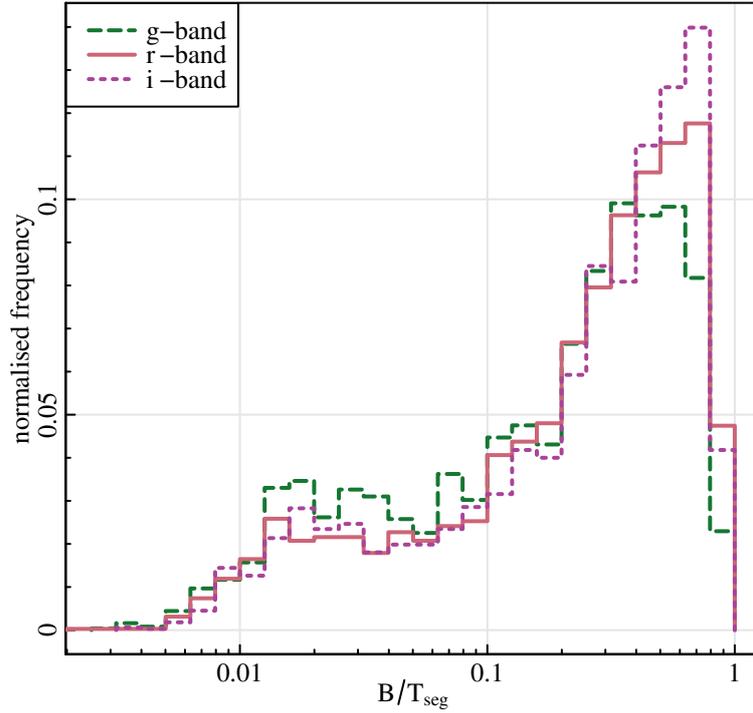


Figure 3.9: The distribution of the bulge to total flux ratio (limited to segment radii) for the 1.5- and double component fits in all bands in v04. Dashed green, solid light red and dotted dark pink lines refer to the g , r and i bands, respectively. The histograms have been normalised by their respective total number of fits (cf. Figure 3.8) to make the bands directly comparable.

We investigate the colours further in Section 3.3.3.

From the middle column it becomes obvious that bulges tend to be smaller than disks by a factor ~ 2 , while single Sérsic fits span a wide range of sizes. Similar to the trend observed in the magnitudes, the smallest objects are classified as single component systems, while some of the larger galaxies can be successfully decomposed.

The Sérsic indices of single component systems show a clear peak around a value of 1 (exponential), a sharp drop-off at lower values and a longer tail towards higher values. Interestingly, the single Sérsic distributions showing all systems (black lines) have a secondary “bump” around a value of 4 or 5 (classical de Vaucouleurs bulge), which is not apparent in those galaxies classified as single component systems (yellow line). Hence most of those high Sérsic index objects were found to contain bulges and were classified as double component systems. The bulges themselves show a wide range of Sérsic indices with (at least in r and i bands) a slightly double-peaked nature around values of 1 and 4-6. At this point, we would like to remind the reader that we use the term “bulge” to refer to all kinds of central components of galaxies, including classical bulges, pseudo-bulges, bars and AGN (cf. Section 2.1.2). Hence the “bulge” distribution will include a variety of physical components and their combinations,

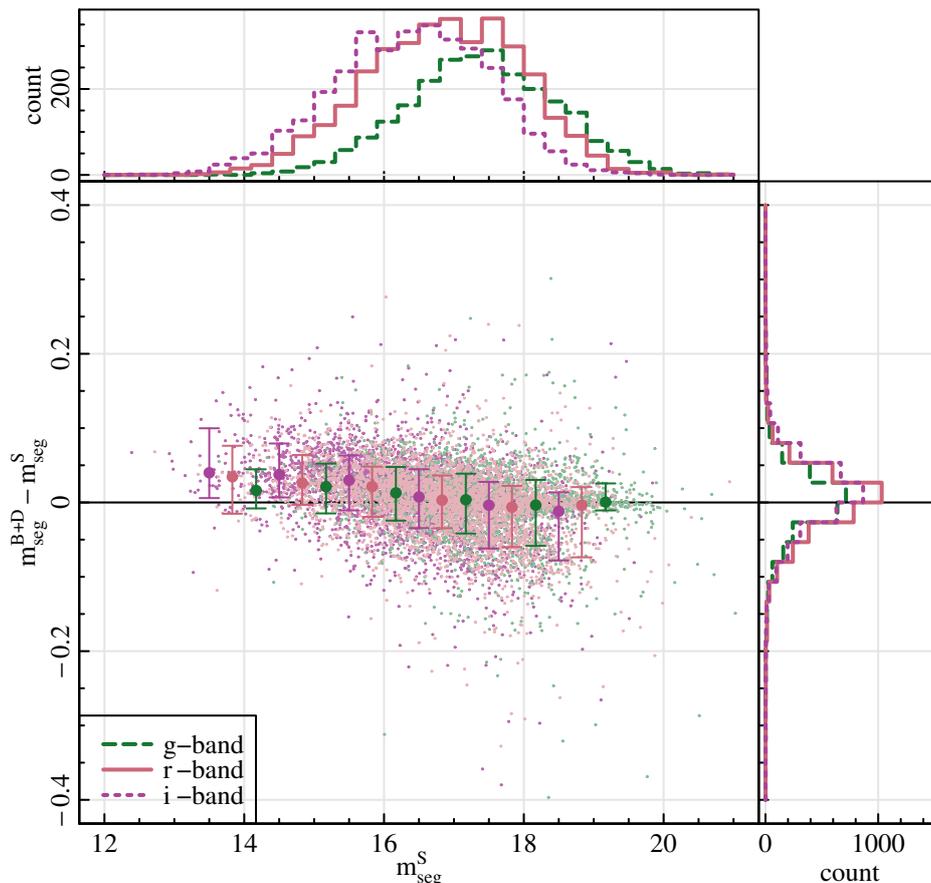


Figure 3.10: The difference between the single Sérsic magnitude and the total magnitude derived from the double or 1.5-component fits for those galaxies that were classified as such, for all bands in v04 (all magnitudes limited to segment radii). The scatter plot shows the difference between the two magnitudes against the single Sérsic magnitude for all three bands with the running medians and 1σ -percentiles overplotted. The top and right panels show the respective marginal distributions. Dashed green, solid light red and dotted dark pink lines refer to the g , r and i bands respectively.

leading to the wide spread of values. In addition, the Sérsic index tends to be the parameter with the largest uncertainty, with typical galaxies showing relative errors on their bulge Sérsic index of 1-10%, adding further scatter to the distribution.

Since the bulge to total flux ratio is a derived parameter that is frequently of interest, we additionally show it in Figure 3.9 for all three bands; for those galaxies that were classified as a 1.5- or double component fit in the respective band. The majority of systems have intermediate values of B/T with only a few percent at the extreme end above 0.8. The secondary peak at very low B/T values around 0.02 stems from the 1.5-component fits. The B/T ratio generally increases from g to r to i , as expected (see Section 4.1.2).

Finally, as a first consistency check, we show the difference between the single Sérsic magni-

3.3. PARAMETER DISTRIBUTIONS

tude and the total magnitude derived from the double or 1.5-component fits, all limited to segment radii, in Figure 3.10. The distributions for all three bands are highly peaked around zero, with the vast majority of objects having total magnitudes consistent with the single Sérsic magnitudes within 0.1 mag (over the entire magnitude range). We only show galaxies that were classified as 1.5- or double component fits here to ensure reliable bulge and disk magnitudes, but note that the distribution is very similar when including objects classified as single Sérsic fits. This indicates that the total magnitude is well-constrained even in the case when the individual component magnitudes are not (see also Section 3.3.3).

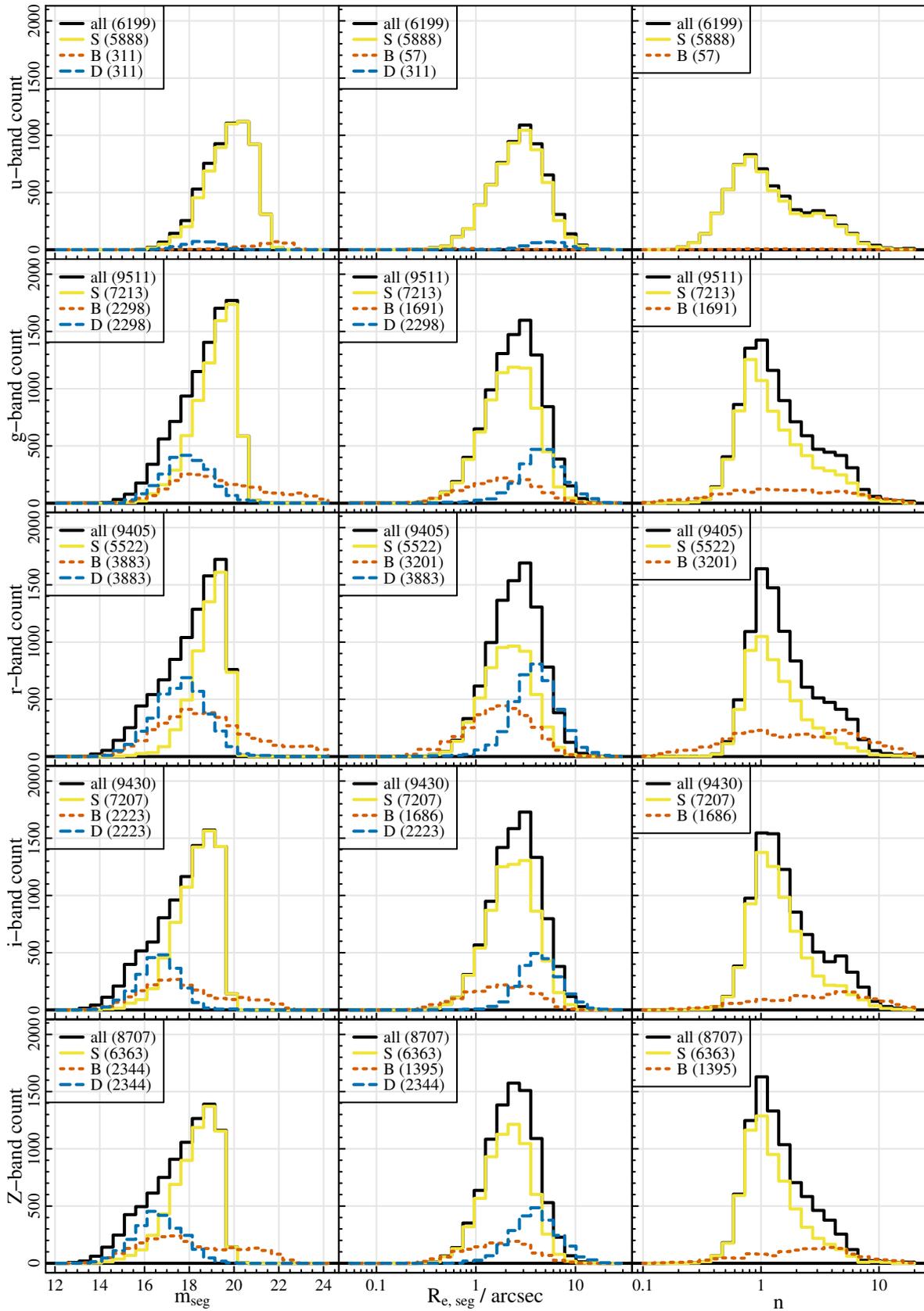
3.3.2 v05 parameter distributions

Figure 3.11 shows the v05 equivalent of Figure 3.8: the distribution of the three main Sérsic parameters in all nine bands for single Sérsic fits, bulges and disks. The layout of the plot, the colour coding of the lines and the scales of the axes are identical to those in Figure 3.8. A more detailed, direct comparison between the results in v04 and v05 is shown in Figures 3.14 to 3.17 at the end of this section.

Generally, the distributions of the parameters in Figure 3.11 are similar in all bands and similar to those in Figure 3.8. The total number of fits in the core bands is lower in v05 than in v04 due to the joint fitting: in v04 we show all fits obtained to individual images of the same galaxy, while in v05 there is only one fit per galaxy. For all other bands, the total numbers are lower due to the increased fractions of skipped fits and outliers, cf. Sections 2.3.3 and 3.2.2. The number of objects shown for individual components varies greatly due to the differences in the model selection between bands (Sections 2.3.4 and 3.2.3), where - as for v04 - the brightest and largest objects, often with high single Sérsic indices, are most frequently decomposed successfully.

Magnitudes become systematically brighter as a function of wavelength due to typical galaxy colours. Effective radii and Sérsic indices do not show obvious trends as a function of wavelength, with no indications of inconsistencies in the analysis between different datasets or bands. For all bands, disks tend to be brighter than bulges, with the bulge magnitudes showing a double-peaked nature due to the 1.5- and double component fits. Disks are also typically larger than bulges. The bulge Sérsic indices again show a large range of values, although with a slight trend towards higher values for longer wavelength bands.

The v05 equivalent to Figure 3.9 is shown in Figure 3.12. It shows the distribution of the bulge to total flux ratio for all nine bands as a density plot on a logarithmic x -axis scale. The general trend observed in Figure 3.9 persists: the B/T ratio increases as a function of wavelength. The double-peaked nature is again due to the 1.5- and double component fits, where the former populate the region of very low B/T values and the latter the intermediate and high



3.3. PARAMETER DISTRIBUTIONS

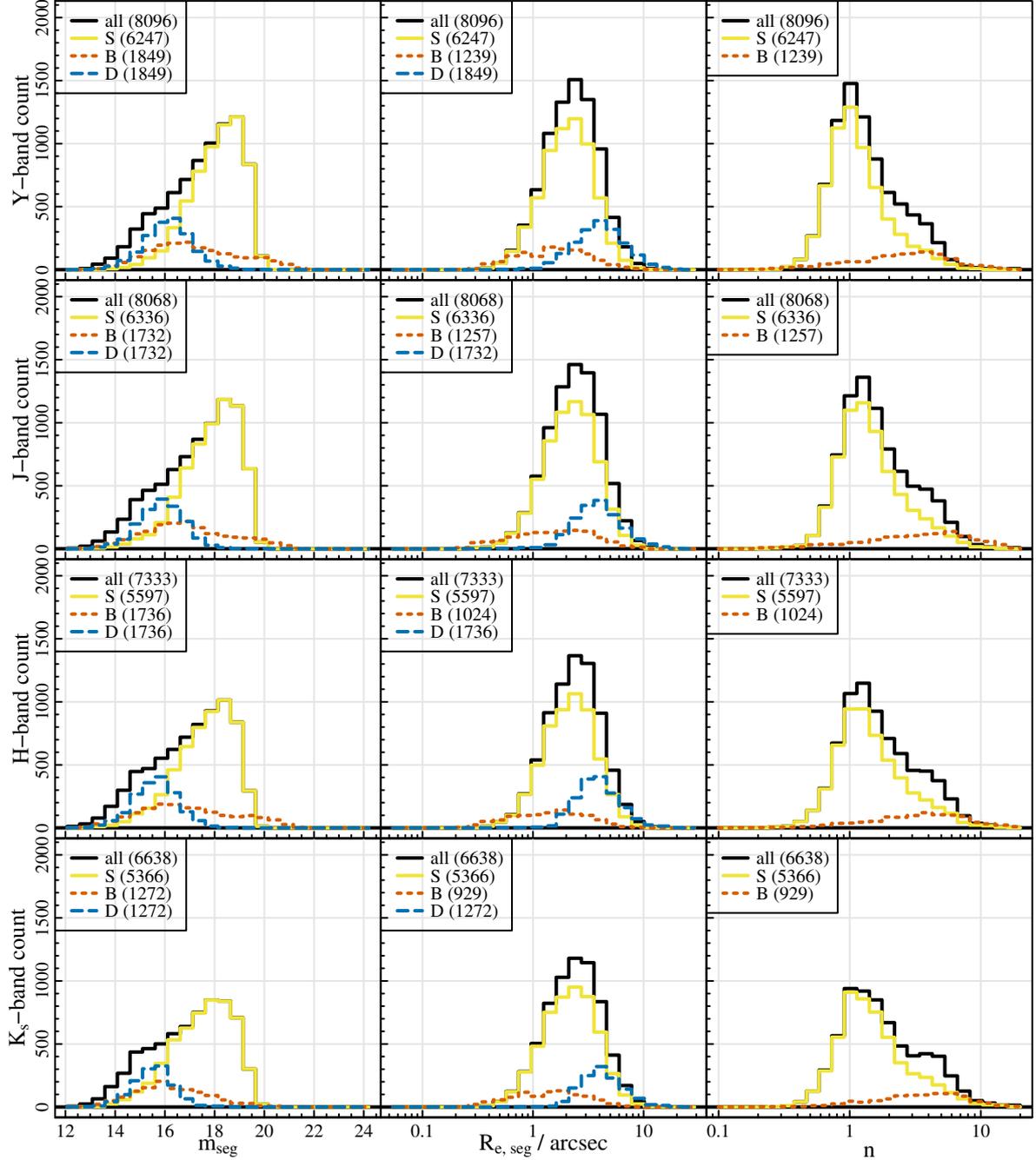


Figure 3.11: The distribution of the main parameters (limited to segment radii) for all bands and models in v05. Panels are the same as in Figure 3.8, except that we now show all nine bands ($u, g, r, i, Z, Y, J, H, K_s$ from top to bottom). The axis scales and bin sizes are the same for all bands and also with respect to Figure 3.8.

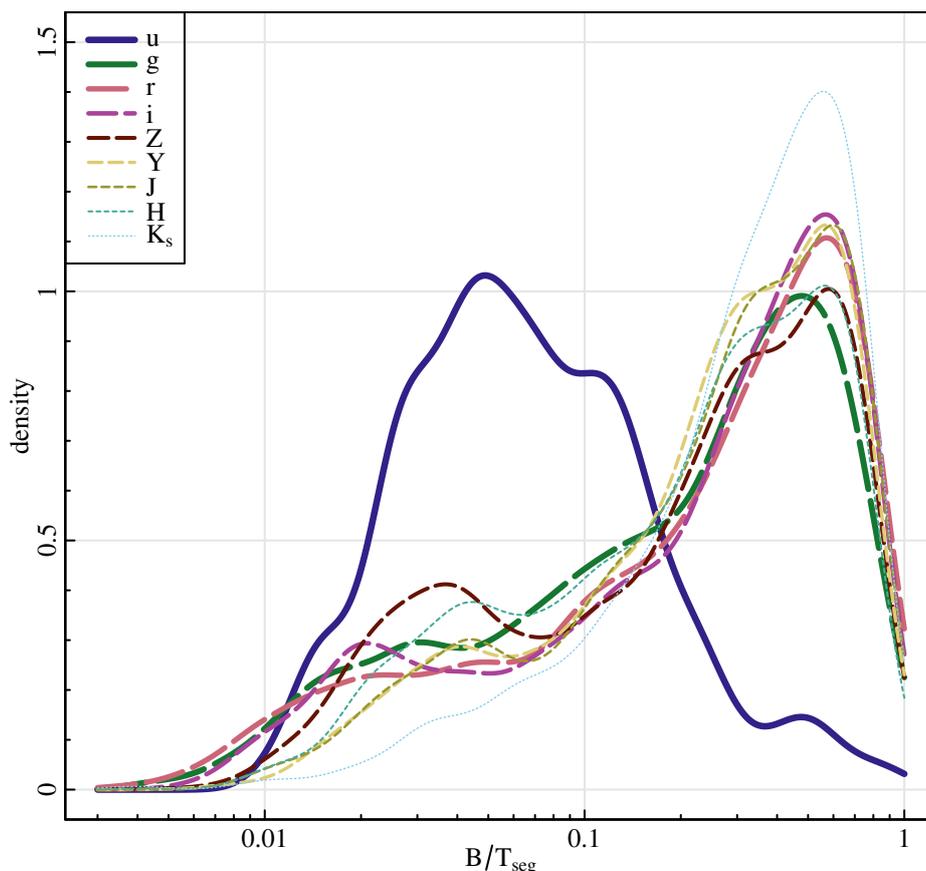


Figure 3.12: The distribution of the bulge to total flux ratio (limited to segment radii) for the 1.5- and double component fits in all bands in v05. This figure is equivalent to Figure 3.9, except that we now show density plots instead of normalised histograms for better visibility due to the higher number of bands.

values. The u -band is very different from the others since it has a factor of ~ 10 less B/T ratio measurements available than the core bands; and almost all of those are 1.5-component fits (see first row in Figure 3.11). Notable exceptions from the general trend of increased B/T with wavelength are observed for the Z and H bands, both of which have exceptionally high numbers of 1.5-component fits relative to the double component fits (see relative numbers in rows 5 and 8 in Figure 3.11; also visible as larger “bumps” at low magnitudes). This results in lower B/T values on average for these bands.

Figure 3.13 shows the last of the v05 equivalents to Section 3.3.1, namely the difference between the single Sérsic magnitudes and the total magnitudes derived from the sum of the bulge and disk fluxes (corresponding v04 version in Figure 3.10). For all bands, these two versions of the total magnitude are consistent within 0.1 mag for the vast majority of galaxies over the entire magnitude range. There is, however, a slight positive bias especially for the u, i, J, H and K_s bands (see the histogram on the right of Figure 3.13), i.e. the single Sérsic

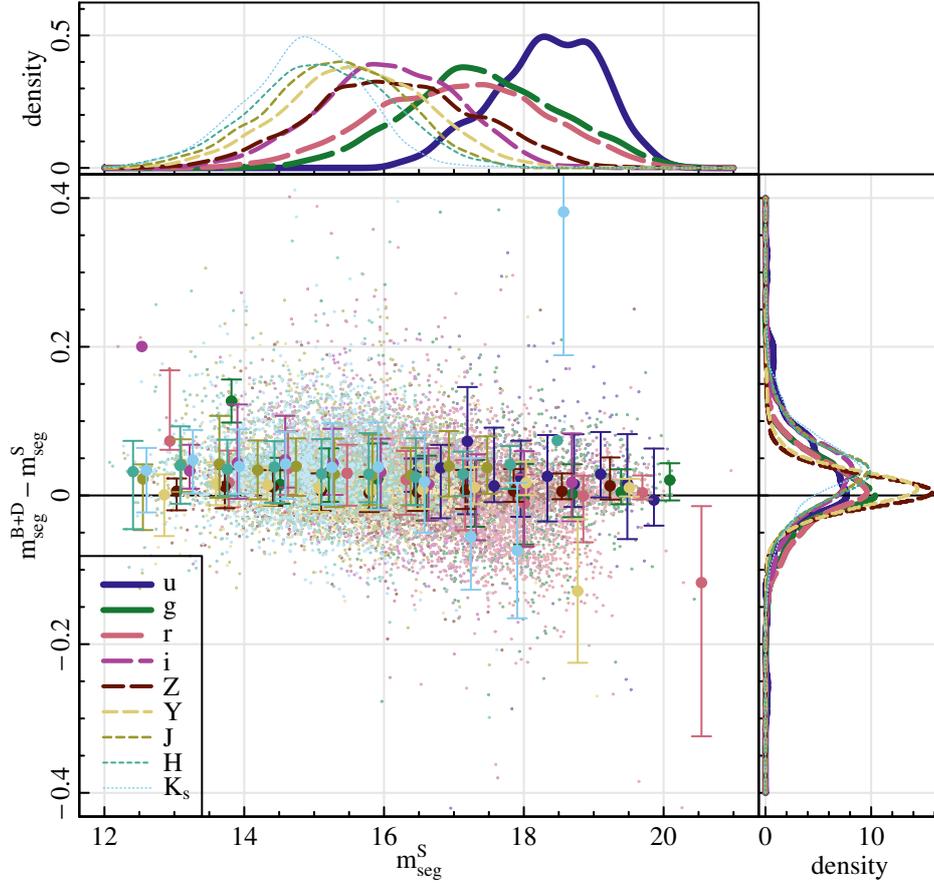


Figure 3.13: The difference between the single Sérsic magnitude and the total magnitude derived from the double or 1.5-component fits for those galaxies that were classified as such, for all bands in v05 (all magnitudes limited to segment radii). The scatter plot shows the difference between the two magnitudes against the single Sérsic magnitude for all bands with the running medians and 1σ -percentiles overplotted. The top and right panels show the respective marginal distributions. This figure is equivalent to Figure 3.10 for v04.

magnitudes tend to be slightly brighter than the total magnitudes obtained from adding the bulge and disk fluxes. A hint of this is also visible in Figure 3.10.

To finish this Section, Figures 3.14 to 3.17 provide a direct comparison of the fitted parameters in v04 and v05 in the r -band. For all single Sérsic parameters, we show the difference between the two fits (v05 - v04 for position RA and Dec, magnitude m , and position angle PA; and v05 / v04 for the scale parameters effective radius R_e , Sérsic index n , and axial ratio b/a) against the v04 fits (in logarithmic units for scale parameters) in Figure 3.14. Running medians and 1σ quantiles are shown as red points with error bars with numerical values indicated in the top left corners of each row. The sample is limited to those objects that had a single data match in the core bands and were neither skipped nor flagged as outlier in either catalogue. The corresponding g and i band plots are nearly indistinguishable with only very slightly

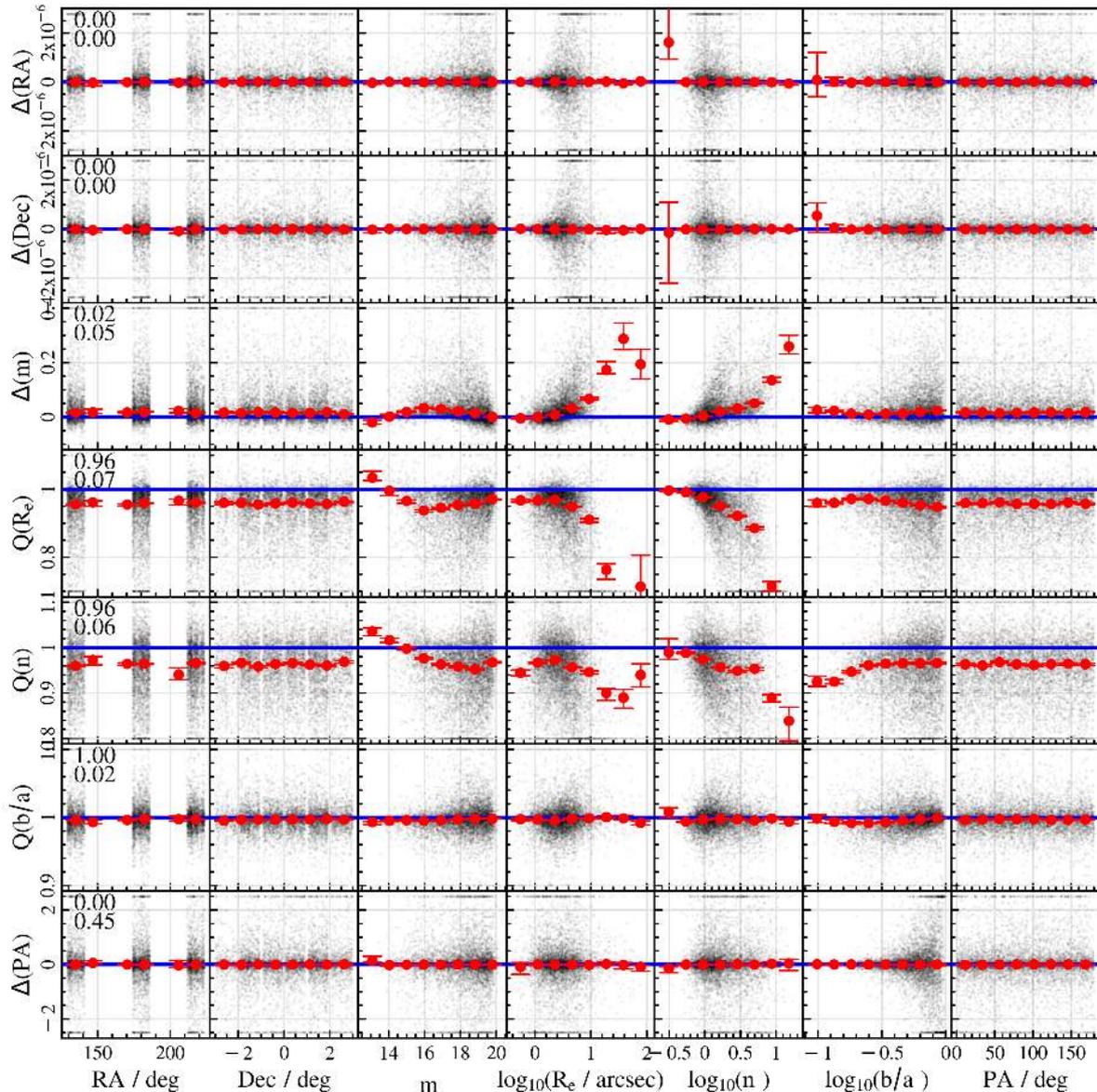


Figure 3.14: The difference Δ or quotient Q (for scale parameters) between the v05 and v04 fits plotted against the v04 fits for all single Sérsic parameters in the r -band. Black dots show all fits that had a single data match and were neither skipped nor flagged as outliers in either catalogue, red dots with error bars show the running median and its error in evenly spaced bins and horizontal blue lines indicate no difference between the fits. The numbers in the top left corners of the first row of panels show the median and 1σ -quantile of the respective distribution in the y -direction (which is identical for all panels of a row). Results for the g and i bands are not shown since they are almost indistinguishable from those in the r -band.

3.3. PARAMETER DISTRIBUTIONS

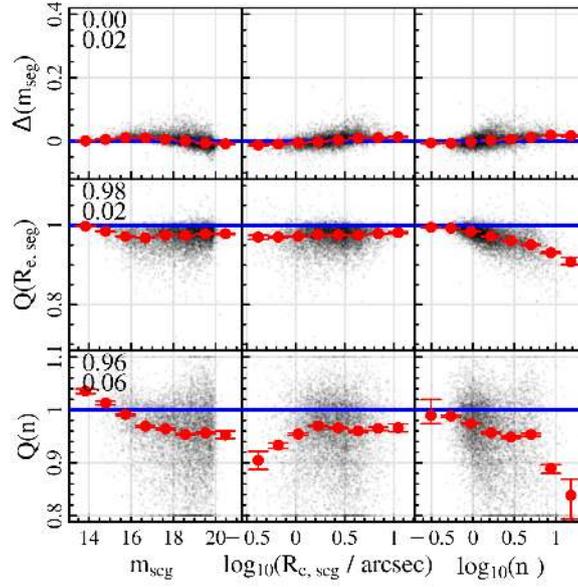


Figure 3.15: The same as Figure 3.14, but for the three main parameters magnitude, effective radius and Sérsic index only, truncating the former two to the v04 segment radii for both catalogues. The axis scales are the same as in Figure 3.14 (central three by three panels).

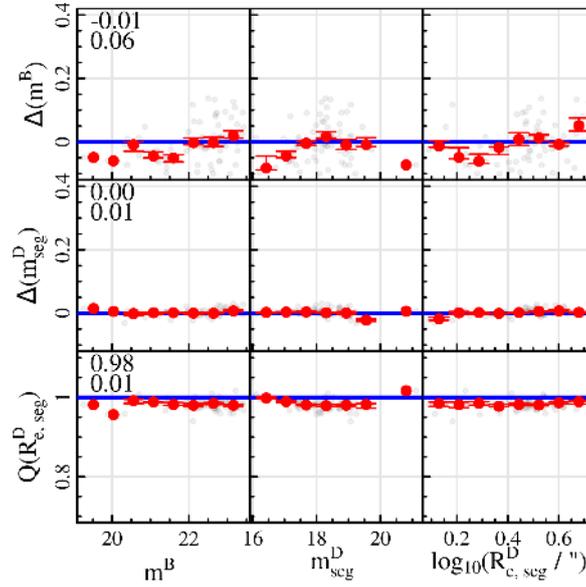


Figure 3.16: The same as Figure 3.14, but for the main parameters of the 1.5-component fits (bulge and disk magnitudes and disk effective radius); all truncated to the v04 segment radii for both catalogues (except for the point source magnitude since the point source flux is entirely contained within the segment by definition). The axis scales are the same as those for the corresponding single Sérsic parameters in Figures 3.14 (central three by three panels) and 3.15. The sample is limited to objects classified as 1.5-component fits in both catalogues in the r -band.

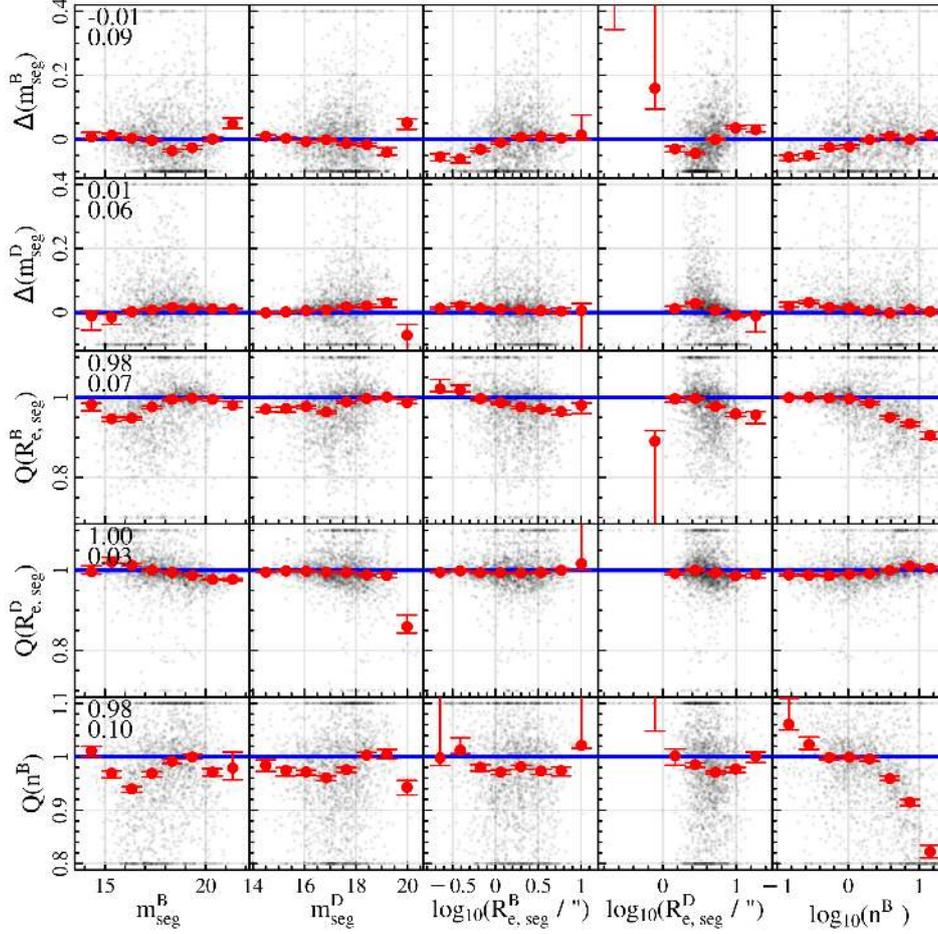


Figure 3.17: The same as Figure 3.14, but for the main parameters of the double component fits (bulge and disk magnitudes, bulge and disk effective radii and bulge Sérsic index); with magnitudes and effective radii truncated to the $v04$ segment radii for both catalogues. The axis scales are the same as those for the corresponding single Sérsic parameters in Figures 3.14 (central three by three panels) and 3.15. The sample is limited to objects classified as double component fits in both catalogues in the r -band.

increased scatter in the Sérsic index (1σ -quantiles of 0.08 for both bands) and position angle (1σ -quantiles of 0.53 in g and 0.55 in i), so we do not show them.

Galaxy positions in RA and Dec (top two rows of Figure 3.14) are recovered near-perfectly (the y -axis range corresponds to $0''.01$; or 5% of a pixel). The only prominent feature are the three GAMA II equatorial survey regions that are clearly visible in RA. The small gaps in the distribution in Dec at intervals of 1° (the KiDS tile size) are due to the limitation to objects with a single match, excluding the overlap sample. There are no deviations of RA or Dec as a function of any other parameter (first two rows); nor does any other parameter show systematic trends as a function of RA and Dec (first two columns). Similarly consistent results between $v04$ and $v05$ are obtained for the axial ratio and position angle (bottom two

3.3. PARAMETER DISTRIBUTIONS

rows; and last two columns), which agree to within $\sim 2\%$ and half a degree respectively.

The three main parameters (central three by three panels in Figure 3.14) show more deviations and also systematic trends: magnitudes are on average 0.01 mag fainter in v05 than in v04; while effective radii and Sérsic indices are 4% and 3% smaller respectively. Bright and large objects with high Sérsic indices are particularly strongly affected. These differences are caused by the larger segments in v05 as explained in Sections 2.1.3 and 2.2.4. We also investigate this effect further in Section 4.1.4, where we compare the v04 results to previous work.

Truncating the effective radius and magnitude to the v04 segment radii (the smaller segments) for both catalogue versions removes the systematic trends in those parameters as shown in Figure 3.15. The Sérsic index remains different between the catalogue versions since it cannot be corrected for different segment sizes. For reasons outlined in Section 2.3.2, we still opted for slightly larger segments in v05.

For completeness, Figures 3.16 and 3.17 show the analogous plots for the 1.5- and double component fits, for objects that were classified as such in both catalogues. We show only the three main Sérsic parameters (for bulges and disks where relevant) and limit magnitudes and effective radii to segment radii. The sample sizes are much smaller (especially in the 1.5-component plot), and the scatter generally increases, in particular for the bulge parameters. The overall agreement between the v04 and v05 parameters remains high, with systematic trends mostly limited to Sérsic indices. The average offsets in Sérsic index decreases to 2% (from 4% for single Sérsic fits) and the offset in the disk effective radius disappears entirely. This is most likely due to the generally better model fit of the double component model, since the differences only arise when the model cannot represent the data adequately (cf. Section 2.1.3).

3.3.3 Galaxy and component colours

For the study of colours, we focus on the core bands since those are the deepest exposures and most directly comparable in our analysis. Additionally, they benefit from the joint *gri* model selection, resulting in a relatively high number of objects with reliable bulge and disk magnitudes. To maximise similarity in terms of depth and seeing, we use $g-r$ colours and r -band absolute magnitudes, M_r . Results for $g-i$ and/or M_i are qualitatively similar, albeit a bit more noisy. Using other bands reduces the number of available components considerably and further increases the scatter (which is already substantial for $g-r$ as we will see below). Following Casura et al. (submitted), we use v04 of the BDDcomp DMU. Using v05 instead does not change the results.

The left panel of Figure 3.18 shows the distribution of $g-r$ colours for galaxies and their components. The colours are corrected for Galactic extinction, but not for dust attenuation in the

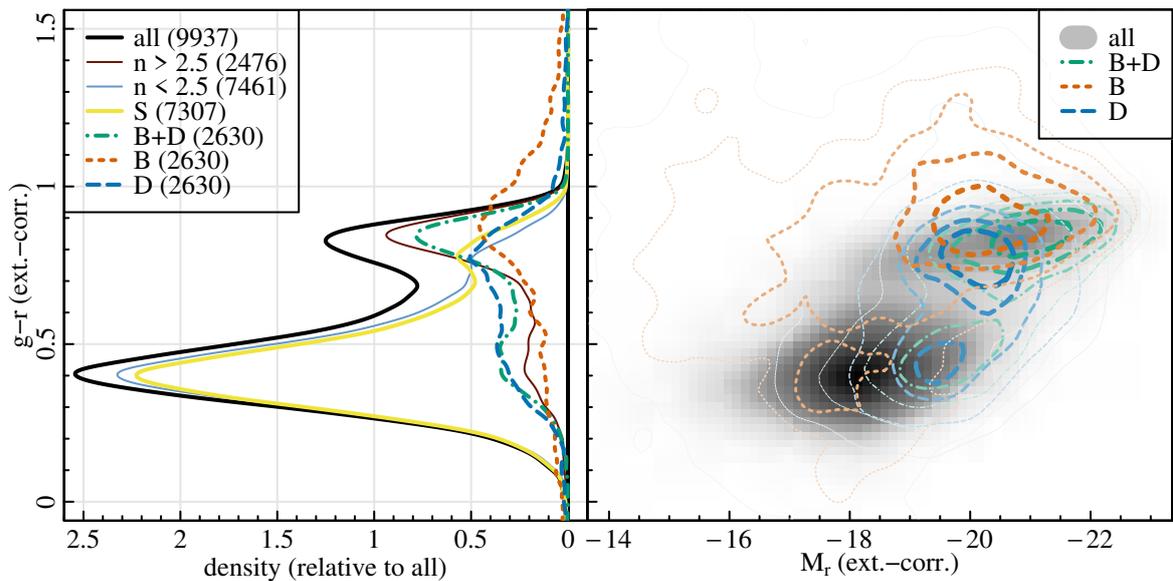


Figure 3.18: Left panel: The Galactic extinction-corrected $g-r$ colour distributions (limited to segment radii) for galaxies and their components. The colour coding of the lines is the same as for Figures 3.8 and 3.11, although with a few additions: The solid black line shows single Sérsic fits for all galaxies with $\text{NCOMP} > 0$ in the joint model selection; the thinner dark red and light blue solid lines split this sample into those with $n > 2.5$ and $n < 2.5$ in the r -band. The solid yellow line gives the single Sérsic values for those galaxies which were classified as single component systems, dotted red and dashed blue lines show bulges and disks, respectively, for those objects classified as 1.5 or double component systems (always in the joint model selection). The dot-dashed green histogram gives the total galaxy colour (derived from the addition of bulge and disk flux) for 1.5 and double component systems. The number of objects in each histogram is given in the legend. **Right panel:** The Galactic extinction-corrected $g-r$ vs. M_r colour-magnitude diagram (limited to segment radii) for galaxies and their components. The colour coding of the lines is the same as for the left panel, although note we do not show the single Sérsic contours here for clarity. Contours include 10, 25, 50, 75 and 90% of the sample.

emitting galaxy. The Galactic extinction was obtained from v03 of the `GalacticExtinction` catalogue accompanying the equatorial input catalogue on the GAMA database.

The solid black line shows the colour distribution for all single Sérsic fits that were not classified as outliers in the joint model selection ($\text{NCOMP} > 0$). It is clearly bimodal, with redder colours typically belonging to higher Sérsic index objects as indicated by the thinner dark red and light blue lines splitting the distribution at $n = 2.5$ (in the r -band). Not entirely surprisingly (given the distribution of Sérsic indices in Figure 3.8), the distribution of single Sérsic objects actually classified as such ($\text{NCOMP} = 1$, solid yellow line) mostly follows the distribution of low Sérsic index objects; while the high Sérsic index objects tend to be classified as double component systems. For the latter, we show total colours with a dash-dotted green line, bulge colours with a dotted red and disk colours with a dashed blue line. As expected, bulges tend to be redder than disks, although the scatter is large.

3.3. PARAMETER DISTRIBUTIONS

The right panel of Figure 3.18 shows the corresponding colour-magnitude diagram. Colours and absolute magnitudes are both corrected for Galactic extinction but not for dust attenuation in the emitting galaxy. The absolute magnitude was calculated using the distance modulus provided in v14 of the `DistancesFrames` catalogue from the GAMA database which we also used to obtain redshifts for the sample selection.

The grey density plot in the background shows the single Sérsic fits for all non-outlier ($\text{NCOMP} > 0$) galaxies, corresponding to the black line in the left panel of Figure 3.18. The bimodality of the distribution is even clearer here, with the red sequence and blue cloud being well-separated. The green contours indicate the part of the sample that was classified as 1.5 or double component object⁴: as expected, this is concentrated towards the bright end of the galaxy distribution and hence encompasses mostly galaxies located in the red sequence. Correspondingly, bulges and disks are both relatively red, with bulges on average slightly redder than the total galaxies and disks slightly bluer (while both components - obviously - are fainter than the total galaxy). However, both components show a large scatter and overlap with each other: both faint blue bulges exist as well as bright red disks.

A detailed study of component colours and the different populations in the right panel of Figure 3.18 is beyond the scope of this thesis. However, we note that the total galaxy colours show much less scatter; indicating that the scatter results from a different splitting of the light into bulge and disk components in the g and r bands, while the total amount of light is well-constrained (cf. also Figure 3.10). A further brief investigation into extreme systems (blue bulges with red disks and also excessively red bulges with very blue disks) suggests that they are caused by a variety of remaining uncertainties in our analysis, e.g. swapped components in one of the two bands (Section 2.1.2), small faint bulges that are barely detected in the r -band and missed in g , the “bulge” component dominating both small and large radii in one of the two bands (cf. Section 3.4.2) or failures in the flagging of bad fits, all combined with model selection uncertainties and the necessity of joint model selection to compromise between the bands. While each of these processes by itself only affects a small number of galaxies, in sum across both bands they do reach the 10-20% level. Still, on average our colours do follow the expected trends, as we show in Section 4.1.2 with an overview of similar studies in the literature. We will study the colours of galaxies and their components in more detail in forthcoming work, also including the other bands ($uZYJHK_s$) and taking full account of inclination effects due to dust in the emitting galaxies (see, e.g. Driver et al., 2008). We will then also assess trends in other parameters, such as the component effective radii, with wavelength; and use these to constrain the nature and distribution of dust in galaxy disks.

⁴To be precise, the green contours were derived by adding the respective bulge and disk fluxes of the 1.5 or double component objects (for consistency with the bulge and disk contours), while the grey density plot is based on the single Sérsic fits (for robustness at low magnitudes). These two versions of the total galaxy magnitude are generally very similar as evidenced by Figure 3.10.

3.4 Catalogue limitations

We finish this chapter by pointing out a few limitations of our results that users of the catalogue should be aware of. Most of these have been discussed before and are only summarised here, with references to the relevant sections.

3.4.1 Model limitations

All of our models are axially symmetric and monotonically decreasing in intensity from the centre. We are unable to capture asymmetries such as spiral arms, offset bulges, tidal tails, mergers, star-forming regions etc.; or disk features such as rings, bumps, truncations or flares. If such features are present in the data, they may bias or skew the model parameters. We also remind the reader that when we talk about “bulges”, what we really mean are the central components. This could be a classical bulge, a pseudo-bulge, an AGN, a bar, or any combination (sometimes resulting in the model trying to fit a mixture between e.g. a bar and a bulge). We make no attempt to distinguish between these cases.

3.4.2 Model selection caveats

Model selection is accurate to $> 90\%$ compared to what could be achieved by visual classification (Section 2.1.3). However, it is important to note that our aim in the model selection is to determine which one of our three models is most appropriate to use for the given data; and not how many physically distinct components an object consists of. The reason for this is that for a given galaxy, the data quality will strongly influence how many fitting parameters can be meaningfully constrained and using more model parameters will inevitably overfit the data and lead to unphysical results. Hence, even in the joint model selection, we base our visual classification on the fit and residuals in individual bands (which is what we fit to), rather than e.g. colour images. Due to the different depths and resolutions of the bands, it is common for the same galaxy to be classified as double component in one band, but single component in another (cf. also discussions in Sections 2.3.4 and 3.2.3).

In an attempt to make fitting parameters more directly comparable across bands, we introduced the joint *gri* model selection (Section 2.1.3), yet this is necessarily a compromise between the different bands. For example, we lose bulges that are resolved in the *r*-band but not in *g* and *i* due to the larger PSFs; or there may be some ill-constrained *i*-band fitting parameters for an extended low-surface brightness disk that is visible in *r* and *g* but not in the shallower *i*-band image. There are also more skipped fits and outliers in the joint model selection than in the band-specific ones because all objects that are skipped or flagged in at least one of the three bands are skipped or flagged in the joint model selection. These

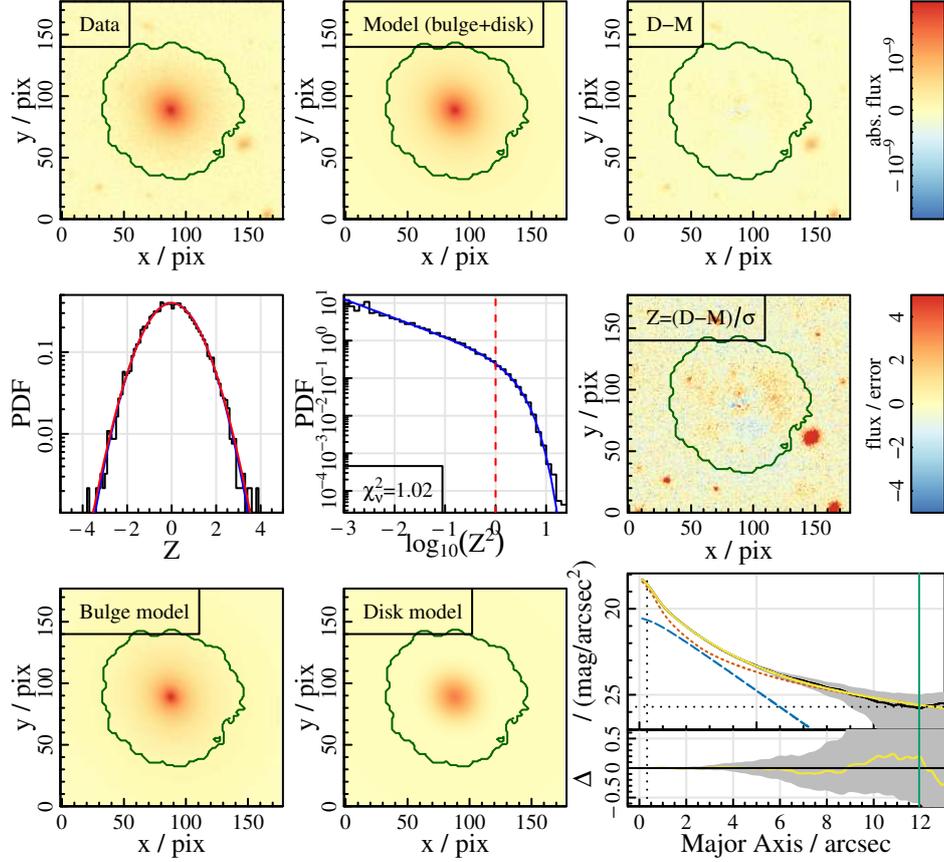


Figure 3.19: The double component fit to galaxy 549706, classified as double component object but with a very high B/T ratio of 0.71 in the KiDS r -band. Panels are the same as those in Figure 2.2.

problems become significantly worse for the nine-band joint model selection encompassing all of $ugriZYJHK_s$ due to the larger range in depth and seeing for individual bands.

Irrespective of the result of the model selection, we provide all fitted parameters of all models in the catalogue (along with the postage stamps of all fits and a flag indicating the preferred model). This allows users to perform their own selection if desired; but also requires care as not all provided parameters will be meaningful. While single Sérsic fits to double component objects are mostly reasonable; double component fits to true single component galaxies will have unconstrained and potentially unphysical parameters for at least one of the components.

We are also aware of a population of objects that are classified as double component fits but have the bulge component dominating both the centre and the outskirts, with the disk only dominating at intermediate radii or even staying “below” the bulge at all radii. We believe these are essentially single component systems that do not follow a Sérsic law (e.g. Sérsic index would be higher at centre than outskirts); and so the freedom of the disk is used to

offset this. This population is easily identifiable by the high bulge-to-total ratio ($B/T \gtrsim 0.6$ or 0.7). The single Sérsic fits may be more appropriate to use in these cases (see also the discussion of this issue in Allen et al., 2006). An example is shown in Figure 3.19.

3.4.3 Drawbacks of tight fitting segments

As detailed in Section 2.1.3, we use relatively tight segments around the galaxies for fitting, which results in the best possible fit of the inner regions of the galaxy but can lead to large, unphysical wings. Hence we recommend using only integrated properties, i.e. the summed flux/magnitude within the region that was fitted and the corresponding effective radii and bulge-to-total ratios as given by the corresponding *_SEGRAD properties in the catalogue. For comparisons to other catalogues using larger fitting segments, their profiles should also be appropriately truncated (see details in Sections 2.1.3, 2.2.4 and 4.1.4).

3.4.4 Sources of systematic uncertainties

We provide errors for each fitted parameter in the catalogue including our best estimate of systematic uncertainties taken from Table 4.2. However, we do not apply the (small) bias corrections given in the same table, since they are only applicable to large random samples of our galaxies and not to individual objects. In addition, we would like to point out that the systematic errors were estimated from single Sérsic r -band fits. We expect that individual components as well as other bands are affected by similar systematics, but we did not test for this. Also, there are some systematic uncertainties that we do not account for in our simulations, most obviously galaxy features that cannot be captured by our models. For these reasons, the given errors should still be considered as lower limits of the true errors.

3.4.5 GAMA-KiDS RA/Dec offset

We observed an average offset between the input and output (fitted) positions of galaxies in both RA and Dec of approx. 0.4 pix ($0''.08$). This is due to an offset between the GAMA (SDSS) and KiDS positions; the same offset can be seen when comparing the KiDS source catalogue with the Gaia catalogue; see also figure 15 in Kuijken et al. (2019). We correct for this during the outlier rejection, but give the original (uncorrected) fitted values for position otherwise.

3.4.6 Completeness limits

Due to our sample selection (Section 1.2.4), our spectroscopic completeness is 100% and even the faintest objects in our sample are well-resolved and bright enough to allow for robust single

3.4. CATALOGUE LIMITATIONS

Sérsic fits in the core bands. However, this is not the case for the KiDS u and the longest wavelength VIKING bands, as discussed in Section 3.2.2: in the shallowest bands (K_s and u), we lose approximately 25% of our sample entirely; and robust double component fits can only be obtained for a minority of objects. Also, there is a systemic limit to the component magnitude in that the samples of bulges and disks with magnitudes fainter than the GAMA limit ($r < 19.8$ mag) are incomplete. For example, a bulge with a magnitude of 22 mag in the r -band will only be contained in our sample if the corresponding disk is bright enough such that the total magnitude is below 19.8 mag. Hence, the sample of bulges with 22 mag is incomplete. This applies almost exclusively to the faint bulges from the 1.5-component fits as can be seen in the first column of Figures 3.8 and 3.11.

4 Quality control

After presenting some of the contents of our catalogue of bulge-disk decompositions, we now turn towards demonstrating its robustness. For this, we first compare to previous works in Section 4.1, including work on the same galaxy sample. Section 4.2 then describes additional internal consistency checks and a detailed study of biases and systematic errors with bespoke simulations. Both of these sections are taken from Casura et al. (submitted, their sections 5 and 6) and are focused on the fitting results of v04 of the `BDDecom` catalogue. This has the great advantage that duplicate images of the same galaxy were fitted independently and therefore this overlap sample can serve to identify systematic uncertainties. The v05 results have been anchored to those of v04 in Chapter 3, such that the quality control presented here is applicable to that version as well and we do not repeat the analysis for v05. Quality control metrics and diagnostic plots of individual steps of our pipeline, in particular for the preparatory work, have been presented in Section 2.2.

4.1 Comparison to previous works

This section begins the quality control by comparing to previous works. We start with an overview of the catalogue statistics and colours to corresponding literature values; then turn towards a more detailed comparison with the results from Lange et al. (2015) and Kelvin et al. (2012), who also use GAMA galaxies. In addition, comparisons to v04 of our catalogue are provided in Häußler et al. (2022, their figure B2) for the single Sérsic and double component models and in Robotham et al. (2022, their figures 19 and 20) for single Sérsic fits, both finding reasonable agreement.

4.1.1 Comparison of catalogue statistics

As a first check, we compare our model selection statistics to those of other bulge-disk decomposition works, although care must be taken in judging these results since they will depend on the sample selection, data quality and observational band.

Table 4.1 summarises the corresponding percentages including a few notes on the most important differences of the quoted works to ours (more details on the majority of these studies are

Table 4.1: Comparison of our v04 catalogue statistics to previous works (in the r -band unless stated otherwise). All values are given in percent.

reference	single	double	unsuitable	notes
this work	47	23	30	double including 1.5-comp.; unsuitable including masking (20%)
Domínguez et al. (2022)	Sánchez 42	55	3	unsuitable corresponding to fit failure, single including galaxies where both single and double fits were acceptable
Hashemizadeh et al. (2022)	48	45	7	I -band; classification by prior visual inspection; difficult objects excluded in sample selection; double Sérsic fits
Robotham et al. (2022)	68	31	1	stellar mass instead of light; simultaneous nine-band plus SED fit; unsuitable meaning fit failure; lower redshift limit
Barsanti et al. (2021)	47	28	25	cluster S0 galaxies
Dimauro et al. (2018)	27	63	10	mostly NIR filters; bright and massive galaxies only ($\log_{10}(M_*/M_\odot) > 10.3$)
Lange et al. (2016)	66	16	18	selection based on visual morphology; unsuitable counting all flagged galaxies
Meert et al. (2015)	44	39	17	larger sample up to higher redshift but smaller magnitude range
Head et al. (2014)	19	35	46	g -band; early-type sample; more stringent criteria for “good” fits
Lackner & Gunn (2012)	35	29	36	single corresponding to pure exponential or de Vaucouleurs; unsuitable corresponding to their “Sérsic” category
Simard et al. (2011)	73	26	1	unsuitable corresponding to failure rate of fitting routine; no selection of “good” fits given
Allen et al. (2006)	43	34	23	B -band; unsuitable galaxies excluded through cuts in redshift, galaxy size and surface brightness

given in Section 4.1.2). In short, for the automated decomposition of large samples of galaxies in the r -band, most authors - including ourselves - class roughly half of all galaxies as being well-represented by a single Sérsic model, with the other half split approximately evenly into double component fits and objects unsuitable for fitting with such simple models. This is also in broad agreement with the morphological classifications obtained by Driver et al. (2022).

4.1.2 Comparison of component colours to literature

$g-r$ colours of galaxy components, such as those we present in Figure 3.18 and Section 3.3.3, are not found frequently in the literature, although a number of authors have presented bulge-disk decompositions in several bands. For example, Simard et al. (2011) perform bulge-disk decompositions for a large sample of galaxies in the SDSS g and r bands but only present colour-magnitude diagrams for the total galaxies (their figures 9 and 10). These are visually comparable to our total galaxy colours as indicated by the dot-dashed green contours in the right panel of Figure 3.18. Mendel et al. (2014) add the SDSS u , i and z bands to the analysis of Simard et al. (2011) and present component masses in $ugriz$ but also do not study component colours.

Similarly, Meert et al. (2015) present a large r -band catalogue which is extended to include the g and i bands in Meert et al. (2016). Colour-magnitude diagrams, however, are again only presented for total galaxies, with the authors noting that component colours can be calculated from their catalogue but should be used with care since they are subject to large uncertainties.

More recently, Dimauro et al. (2018) provide (UVJ) component colours in their catalogue but defer their study to future work; while Bottrell et al. (2019) present $ugriz$ colour-magnitude diagrams for total galaxies (colour-coded by B/T); but again not for individual components. Domínguez Sánchez et al. (2022) decompose a sample of $\sim 10\,000$ galaxies in the g , r and i bands, but do not study the colours.

Among the first to show component colours for a large sample of galaxies were Lackner & Gunn (2012) in their study of $\sim 70\,000$ $z < 0.05$ SDSS galaxies in the g , r and i bands. However, in contrast to our fits, their g and i band fits are not independent. Instead, in order to decrease the noisiness of the colours, the structural parameters are taken from the r -band and only the magnitude is adjusted. Additionally, Lackner & Gunn (2012) (along with e.g. Mendel et al. 2014) fix the Sérsic index of the bulge to either 1 or 4 for their double component fits to limit the number of free parameters since the data is insufficient to constrain the bulge light profile. Keeping these differences in mind, their figure 32 showing the $g-r$ vs. M_r colour-magnitude diagram for bulges and disks as contours superimposed on the greyscale background for all galaxies can be compared to the right panel of our Figure 3.18 (for a more detailed description of Figure 3.18, see Section 3.3.3). In general, both plots are very similar¹: the grey background shows a large blue cloud and a well-separated red sequence. The double component fits populate the red sequence, green valley and the brighter part of the blue cloud. The bulges tend to be slightly redder than the red sequence but with a large scatter especially

¹For reference, the Lackner & Gunn (2012) cyan contours represent 6684 galaxies with a bulge Sérsic index of 1 and the magenta contours show 14042 objects with a bulge Sérsic index of 4. Also note that their x -axis is reversed with respect to ours.

at the faint end. Disks spread from the red sequence towards the green valley with a smaller population also in the blue cloud. Lackner & Gunn (2012) also note the large scatter in colour for bulges in particular, despite their fitting constraints and lower reshift limit. Hence it is not surprising that even with the higher quality KiDS data and our new fitting routines, we get a large scatter in component colours, especially since we leave the bulge Sérsic index free and perform independent fits in both bands. The latter can lead to very extreme colours since it is not guaranteed that the “bulge” and “disk” models actually fit the same features in both images (in particular when there are additional features present that are not fully captured by the models; see also Section 3.3.3).

Kim et al. (2016) found similar difficulties when performing g -band and r -band decompositions on $\sim 10\,000$ large bright and approximately face-on SDSS galaxies. While they leave the Sérsic index free as we do, the g -band structural parameters are again taken from the r -band fits with only the magnitudes adjusted. Despite this, they find it necessary to remove almost 40% of their sample after fitting because they show excessively red bulge colours (combined with low B/T values in the r -band). After this cut, the $g - r$ vs. M_r colour-magnitude diagram for bulges shown in their figure 7 is slightly less noisy than ours (Figure 3.18), although still comparable. Kim et al. (2016) did not study the properties of the disks in their sample.

One of the most direct comparisons to make is with Kennedy et al. (2016) who study GAMA galaxies in the G09 region (a subset of our sample) in the $ugrizYJHK$ bands from the SDSS (York et al., 2000) and the United Kingdom Infra-red Telescope Infrared Deep Sky Survey (UKIDSS, Lawrence et al. 2007). They use the MEGAMORPH multi-band fitting method with GALAPAGOS-2 and GALFITM (Häußler et al., 2013; Vika et al., 2013) to perform simultaneous Sérsic plus exponential fits across all 9 bands. The structural parameters are constrained to be the same in all bands, with only the component magnitudes allowed to vary freely, therefore providing robust colours. While the paper focuses on studying $u - r$ colours, the corresponding catalogue on the GAMA database (`MegaMorph:MegaMorphCatv01`) contains the information for the fits in all 9 bands such that $g - r$ colours can easily be derived.

This comparison is shown in Figure 4.1 for those galaxies that were present in both catalogues and classified as double component fits (`NCOMP=2`) in the joint model selection of our fits (Kennedy et al. 2016 perform neither model selection nor outlier rejection). In addition to the scatter plot directly comparing the component colours, we show the corresponding distributions in the left (this work) and top (Kennedy et al., 2016) panels of Figure 4.1. As always, bulges are shown in red (points and dotted lines) and disks in blue (points and dashed lines). To aid the direct comparison of the distributions, we additionally show the Kennedy et al. (2016) bulge colour distribution in the left panel as a solid orange line and the disk distribution from this work in the top panel as a light blue solid line. Component colours are all corrected for Galactic extinction and limited to segment radii for our fits.

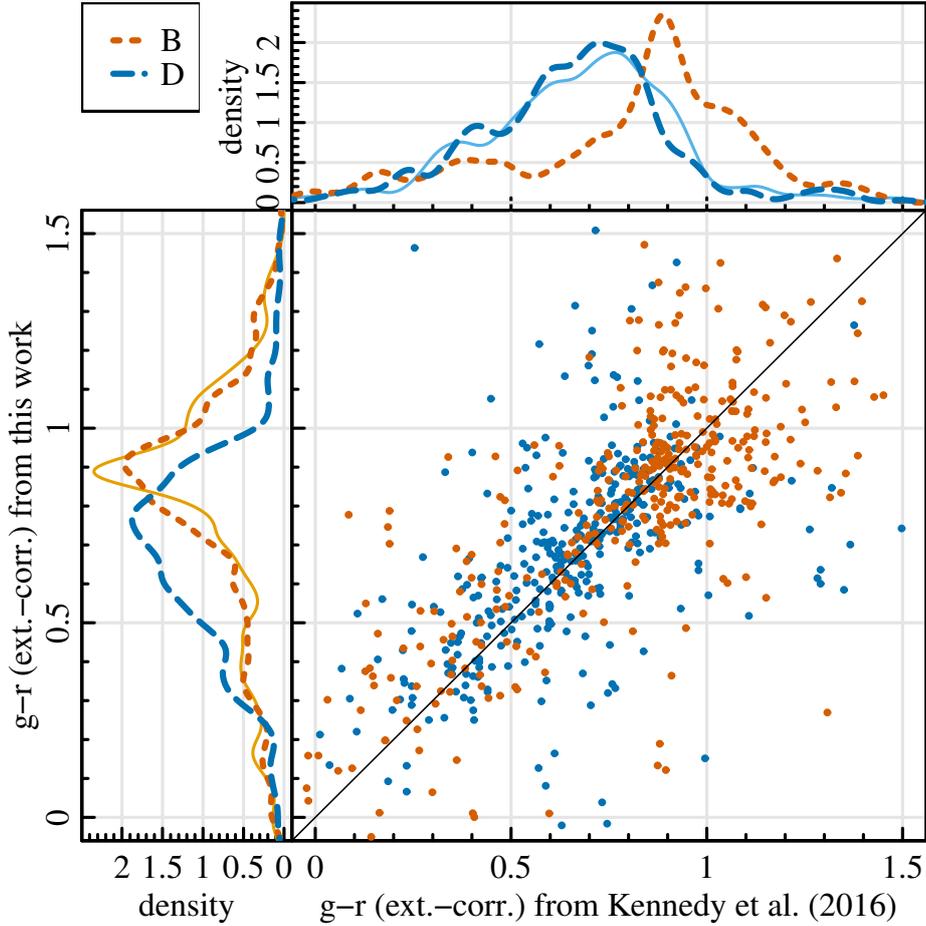


Figure 4.1: Our Galactic extinction-corrected $g-r$ component colours (limited to segment radii) compared against those from Kennedy et al. (2016) for a subsample of 390 objects that appear in both catalogues and were classified as double component fits in the joint v04 model selection. The scatter plot shows the direct comparison, while the density plots show the respective distributions in both catalogues (ours on the left, Kennedy et al. 2016 on the top). Bulges are again shown in red with dotted lines and disks in blue with dashed lines. To aid the direct comparison of the distributions, the lighter solid lines also show the Kennedy et al. (2016) bulge distribution on the left and our disk distribution on the top.

Despite the large scatter, it can be seen that our component colours are generally in agreement with those from Kennedy et al. (2016) with no systematic differences. The scatter in both catalogues is also comparable, although Kennedy et al. (2016) perform multi-band simultaneous fits with fixed structural parameters that should lead to more robust component colours. This advantage of their work seems to be balanced by advantages of our work, such as the improved data quality of KiDS, the robustness of the fitting procedure with `ProFound` and `ProFit` and our post-processing steps (in particular outlier rejection and model selection).

In addition to the large $g-r$ component colour studies discussed above, there are a number of publications focusing on the $g-i$ colours of bulges and disks for samples ranging between

~ 100 and ~ 1000 objects (i.e. roughly a factor of 10 smaller than ours), namely Gadotti (2009); Head et al. (2014); Vika et al. (2014); Fernández Lorenzo et al. (2014); Cook et al. (2019); Barsanti et al. (2021). We briefly compare our work to their results here, noting that all above authors have more stringent constraints on their fits than we do and also report problems in deriving bulge colours. For example, Fernández Lorenzo et al. (2014), although they fit the galaxies in both bands, use fixed aperture photometry to derive more stable bulge colours. Vika et al. (2014), while performing *simultaneous* multi-band fits, do not allow for any variation of structural parameters (except magnitudes) with wavelength. Head et al. (2014), in addition to varying magnitudes, allow for a trend in disk sizes with wavelength in approximately 30% of their sample, noting that this leads to increased scatter. Cook et al. (2019), who use `ProFit` like this work, allow disks to deviate slightly from the exponential profile but fix all bulges to be exactly round (axial ratio of 1), again only allow magnitudes and disk sizes to vary between bands and employ a sophisticated, visually-guided re-fitting procedure to obtain physically meaningful fits for “difficult” objects. Barsanti et al. (2021), also employing `ProFit`, additionally allow for differing bulge sizes and Sérsic indices in the different bands (but fixing bulge and disk axial ratios and position angles and performing model selection in the r -band only), but class approximately half of their double component fits as “unreliable”. Gadotti (2009), fitting bulges, bars and disks to a sample of face-on, visually-selected “well-behaved” galaxies refrain from automated fitting and instead treat each galaxy individually.

After these notes on the inherent difficulties associated with deriving component colours, we can now turn to the corresponding results: Head et al. (2014), in their study of early-type red-sequence galaxies in the Coma cluster, measure an average $g - i$ difference between bulges and disks of 0.09 ± 0.01 mag. Similarly, Barsanti et al. (2021) find a bulge-disk $g - i$ difference of 0.11 ± 0.02 mag for their sample of S0 cluster galaxies. Fernández Lorenzo et al. (2014), on the other hand, have a sample of mostly late-type spirals (with $B/T < 0.1$ for $\sim 66\%$ of their objects) and find a difference of 0.29 mag in the median $g - i$ bulge and disk colours, i.e. a factor of ~ 3 larger. In line with this, Vika et al. (2014) report that the bulge and disk colours are similar for early-type galaxies but differ significantly for late-types. The $g - i$ differences for the different morphological classes given in their table 2 range from 0.03 ± 0.04 mag for ellipticals to 0.28 ± 0.06 mag for late-type spirals; with the overall average (comprising approximately two thirds late-types) being 0.19 ± 0.04 mag. Similarly, the average $g - i$ colour difference of the Gadotti (2009) sample of varying galaxy types amounts to 0.18 ± 0.04 mag (from the online-version of their table 2).

Our results are perfectly in line with this: the median bulge-disk $g - i$ colour difference for our 1.5- or double component fits is 0.17 ± 0.01 mag, consistent with the Vika et al. (2014) and Gadotti (2009) results. Limiting to objects with a total $g - i > 1$ (red-sequence galaxies) reduces the value to 0.14 ± 0.01 mag; while focusing on 2-component fits only (excluding 1.5-

4.1. COMPARISON TO PREVIOUS WORKS

component fits) yields 0.10 ± 0.02 mag, in agreement with Head et al. (2014) and Barsanti et al. (2021). This is because our double component galaxies lie predominantly on the red sequence, as can be seen in Figure 3.18. 1.5-component fits on the other hand, have very small (namely unresolved) bulges by definition and hence belong to the class of late-type spirals. In fact, 87% of the 1.5-component objects have a (*g*-band) B/T ratio less than 0.1, with the median value as low as 0.02. Computing the bulge-disk *g* – *i* colour difference for this sample of objects yields a value of 0.46 ± 0.02 mag, suggesting that the trend described in Vika et al. (2014) continues at very low B/T.

From all of these comparisons we conclude that our component colours - although noisy - are in line with previous work. In order to increase the colour robustness while preserving the ability to capture physical trends with wavelength (i.e. not fixing the structural parameters to be the same in all bands), a simultaneous fit in all bands is needed. This has many advantages as shown by the MEGAMORPH project team using GALAPAGOS and GALFITM (Häußler et al., 2013; Vika et al., 2013; Häußler et al., 2022), especially for automated analyses, since it naturally ensures smooth wavelength trends while preserving physical variation and additionally allows more robust fits to fainter magnitudes. With ProFit v2.0.0, released in February 2021, now supporting a multi-band fitting mode - and the newly developed package ProFuse (Robotham et al., 2022) even combining this with a spectral analysis - this is certainly an interesting avenue to explore in future work and could provide a valuable alternative. It would also solve some of the other challenges we faced during the individual fits, as we discuss in Sections 2.3.3, 2.3.4 and 3.2.2.

4.1.3 Comparison to size-stellar mass relations of Lange et al. (2015)

Figure 4.2 shows the size-stellar mass relation obtained from our *r*-band single Sérsic fits in combination with the redshifts and distance moduli of v14 of the DistancesFrames catalogue (Baldry et al., 2012) and v19 of the StellarMasses catalogue (Taylor et al., 2011); both from the GAMA database. The aperture-derived stellar masses have been scaled to match the Sérsic total flux using the fluxscale keyword provided in the StellarMasses catalogue. The *g*-band and *i*-band results are very similar to those from the *r*-band so we do not show them.

The sample is limited to objects which were not flagged during our outlier rejection (Section 2.1.3) and split into early- and late-type galaxies according to our fitted Sérsic index ($n \lesssim 2.5$; analogous to Lange et al. 2015). We also limit the redshift range to $0.0001 < z < 0.06$ and the stellar mass range to $M_* > 10^9 M_\odot$, thus avoiding the need for volume corrections. For comparison, we show the $M_* - R_e$ relations obtained by Lange et al. (2015) by fitting a single power law to the single component *r*-band fits of Kelvin et al. (2012) (pre-release of SersicPhotometry:SersicCatSDSSv09) combined with an earlier version of the stellar

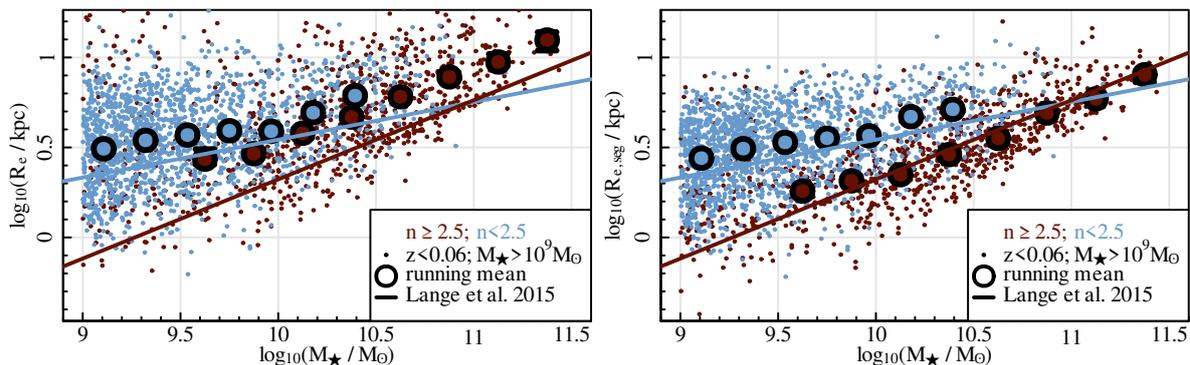


Figure 4.2: The size-stellar mass relation for our r -band fits (dots) compared to the Lange et al. (2015) fits (lines). The sizes are obtained from our single Sérsic effective radii (**left panel:** extrapolated to infinity; **right panel:** limited to segment radii) and the distance moduli provided in the `DistancesFrames` catalogue originally described by Baldry et al. (2012). The stellar masses are taken from the most recent version of the `StellarMasses` catalogue initially presented in Taylor et al. (2011). The sample is limited to the redshift range $0.0001 < z < 0.06$ (redshifts also from `DistancesFrames`) and the stellar mass range $M_* > 10^9 M_\odot$. Large circles with error bars indicate the running median with its error (usually smaller than the data point). Solid lines show the single exponential $M_* - R_e$ relation fits obtained by Lange et al. (2015) for their single component r -band sample, split by a Sérsic index cut at $n = 2.5$ (taken from their tables 2 and 3).

masses catalogue of Taylor et al. (2011) (`StellarMasses:StellarMassesv16`). We note that the stellar masses did not change much between v16 and v19: the mean and standard deviation of $\Delta \log_{10}(M_*/M_\odot)$ are 0.006 and 0.07 respectively for our sample. The two panels show the results obtained with effective radii taken directly from the Sérsic fits (left; extrapolated to infinity by definition) or limited to the segment radius within which they were fitted (right; see Section 2.1.3). To guide the eye, we also show the running median and its error for our data; where the error is taken as the 1σ -quantile divided by the square-root of data points within that bin (usually smaller than the size of the data points).

In both cases, the slope of the mass-size relation obtained from our data agrees well with the Lange et al. (2015) fit results.² There is an offset in the absolute sizes, but those will inherently depend on the exact definition of the size measurement at hand as well as the (depth of the) data used. Already calculating effective radii within the segments within which we fitted for them (right panel) brings our results much closer to those of Lange et al. (2015); although the measurements are then not directly comparable to their fits anymore since they use Sérsic values extrapolated to infinity (which will, in turn, depend on the segment size used for fitting). We now discuss these issues further by directly comparing our fits to those of Kelvin et al. (2012), which the Lange et al. (2015) results were based on.

²For reference, the slopes of the plotted lines are 0.21 and 0.44 for the late- and early-types respectively; taken from tables 2 and 3 of Lange et al. (2015)).

4.1. COMPARISON TO PREVIOUS WORKS

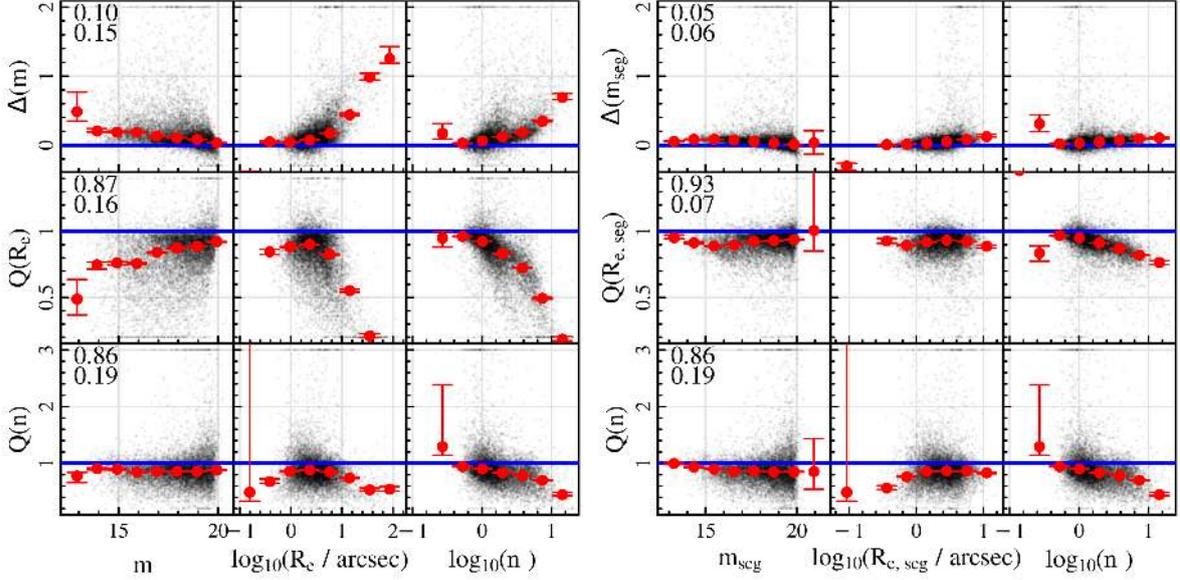


Figure 4.3: The difference Δ or quotient Q (for scale parameters) between the Kelvin et al. (2012) fits and our fits plotted against our fits for the three most important single-Sérsic parameters magnitude, effective radius and Sérsic index in the r -band. The top panels show the Sérsic parameters extrapolated to infinity, while for the bottom panels we calculated the magnitude and radius within the segment radius for both our and the Kelvin et al. (2012) fits. Outliers are clipped to the plotting interval; which is the same in both cases. Black dots show all fits, red dots with error bars show the running median and its error in evenly spaced bins and horizontal blue lines indicate no difference between the fits. The numbers in the top left corners of the first row of panels show the median and 1σ -quantile of the respective distribution in the y -direction (which is identical for all panels of a row). The sample is limited to fits that are available in the Kelvin et al. (2012) catalogue and were classified as single component in our model selection.

4.1.4 Comparison to single Sérsic fits of Kelvin et al. (2012)

To further investigate the size offset observed in Section 4.1.3, we directly compared our fits to those of Kelvin et al. (2012) (`SersicPhotometry:SersicCatSDSSv09` on the GAMA database). Since Kelvin et al. (2012) do not provide double component fits, the analysis is limited to single Sérsic fits. We again use the r -band as an example for the discussion, but note that results are very similar for the g and i bands.

The Kelvin et al. (2012) fits are based on the Structural Investigation of Galaxies via Model Analysis (`SIGMA`) code applied to data from SDSS DR7. `SIGMA` is a wrapper around `Source Extractor` (Bertin & Arnouts, 1996), `PSF Extractor` (Bertin, 2011) and `GALFIT 3` (Peng et al., 2010) performing similar steps to what we do in our pipeline (Section 2.1), i.e. source identification, background subtraction, PSF estimation and 2D model fits to the surface brightness profile of the galaxies. The differences lie in the data and code used, where we upgrade SDSS to KiDS, `Source Extractor` to `ProFound`, `PSF Extractor` to a combination of `ProFound` and `ProFit` and `GALFIT` to `ProFit`; with all the advantages described in Sec-

tions 1.2.2, 1.3.3 and 1.3.4. In addition, we also perform multi-component fits and model selection. For the comparison to the Kelvin et al. (2012) results, we focus on the three most important single Sérsic fit parameters: magnitude m , Sérsic index n and effective radius R_e , which tend to be the least “well-behaved” (position, axial ratio and angle are generally more easily constrained and uncorrelated, see e.g. Figure 3.14 in Section 3.3.2).

Figure 4.3 shows the difference between our fits and the Kelvin et al. (2012) fits for these three parameters. In line with Figures 3.14 to 3.17, we show the difference in magnitudes (Kelvin et al. (2012) fits - our fits), while for the effective radius and Sérsic index (scale parameters) we show the quotient (Kelvin et al. (2012) fits / our fits) on the y -axis; always plotted against our fitted values on the x -axis (in logarithmic space for scale parameters). Again, we show the results for Sérsic parameters extrapolated to infinity (left panels) and for the magnitude and effective radii calculated within the segment radius (right panels), where we limit both our fits and those of Kelvin et al. (2012) to our fitting segment (which are generally smaller than the fitting regions used in Kelvin et al. (2012)) to obtain directly comparable results.

For the Sérsic parameters extrapolated to infinity (left panels), large differences can be seen in all fitted parameters, including systematic trends across the parameter space. Note the larger axis range relative to Figures 3.14 to 3.17. This shows once again that fitted Sérsic parameters are not directly comparable given the differences in the data, code and processing steps with a wealth of potentially different systematic uncertainties (Section 2.1.3). However, when we limit the analysis to our segment sizes (bottom panels), the fits become much more comparable. On average, now, our fits are ~ 0.03 mag brighter and approximately 7% larger than the Kelvin et al. (2012) fits to the same galaxies, which is not surprising given the increased depth and resolution of KiDS compared to SDSS and the numerous sources of different systematic uncertainties (e.g., differing sky subtraction and PSF estimation). Also, there are fewer trends across the parameter space, indicating that systematic differences arise mainly from the extrapolation to infinity.

The exception to this is the Sérsic index, which still shows some trends. The reason is that the Sérsic index, unlike the magnitude and effective radius, cannot be corrected for different fitting regions. The Kelvin et al. (2012) fits, which were performed within larger fitting regions than our fits, will inevitably have to compromise more between the inner and outer regions of the galaxy to be fitted (unless the light profile truly follows a single Sérsic profile with no deviations out to very large radii, which is rarely the case). Our tight fitting segments, on the other hand, will result in better fits to the inner regions of the galaxy at the expense of producing unphysical wings when extrapolated beyond the fitting segment (Section 2.1.3). Thus, fitted Sérsic indices are always a weighted average (or compromise) across a range of radii and their absolute values will never be directly comparable between catalogues unless the fitting regions are exactly the same (or the galaxies studied follow perfect Sérsic profiles).

4.2 Systematic uncertainties and biases

The MCMC chain errors returned by the fitting procedure do not include systematic uncertainties which arise due to galaxy features not accounted for in the models, nearby other objects, imperfect PSF estimation, background subtraction inaccuracies and similar effects. For an individual galaxy, the presence of such “features” will systematically shift the fitted parameters away from the true values, thus introducing a bias. For a statistically large enough sample of galaxies, however, most of these effects are expected to cancel out on average since they are random from one galaxy to the next (e.g. nearby other sources shifting the fitted positions). These “random systematics” can - for statistical samples - be accounted for by simply increasing the given parameter errors such that in most cases, the true values are included in the credible intervals again. Such systematics can be studied using overlap sample galaxies, i.e. those that appeared in more than one KiDS tile (cf. Section 1.2.4). In addition, there can be “one-sided effects” that lead to an overall bias across the sample, e.g. due to excess flux from nearby objects. These can only be detected using simulations. In the following, we study both of these effects using our bespoke simulations, the overlap sample of real galaxies and the overlap sample of simulated galaxies; where we refer to the random systematics as “error underestimates” and to the one-sided effects as “biases”.

The final corrections for both of these effects are listed in Table 4.2. In short, biases are very small ($\lesssim 1\%$), while systematic errors are a factor of 2-3 larger than the random MCMC errors alone. The error underestimate corrections are also applied to the released catalogues (for v04 and v05), while the bias corrections are not since they are only valid for a large random subset of our galaxies and not for individual objects. Note that the systematic error studies were carried out on results from v03 of the BDDcomp DMU, while the remainder of this chapter refers to v04. However, since v04 is statistically identical to v03 (see Section 3.1.4), the results can directly be transferred. We would also like to point out that we focus on single Sérsic r -band fits in this section. We expect individual components in the 1.5- and double component fits as well as the other bands to be affected by similar systematics. Effects are likely to become worse for fainter and/or less well-resolved objects (i.e. bulges in particular; and objects in the KiDS u and the longer wavelength VIKING bands).

4.2.1 Overlap sample comparison

As an internal consistency check, we compared the fit results obtained from multiple observations of the same physical object (Section 1.2.4) in Figure 4.4. The plots are very similar to the ones in Figure 4.3 (see description in Section 4.1.4; but note the different y -axis scale), except that we now show the differences between two of our own fits to different KiDS images of the same galaxy (in the overlap region between the KiDS tiles). Hence all fits shown in Figure 4.4 are based on KiDS data and use the exact same pipeline for analysis, though the

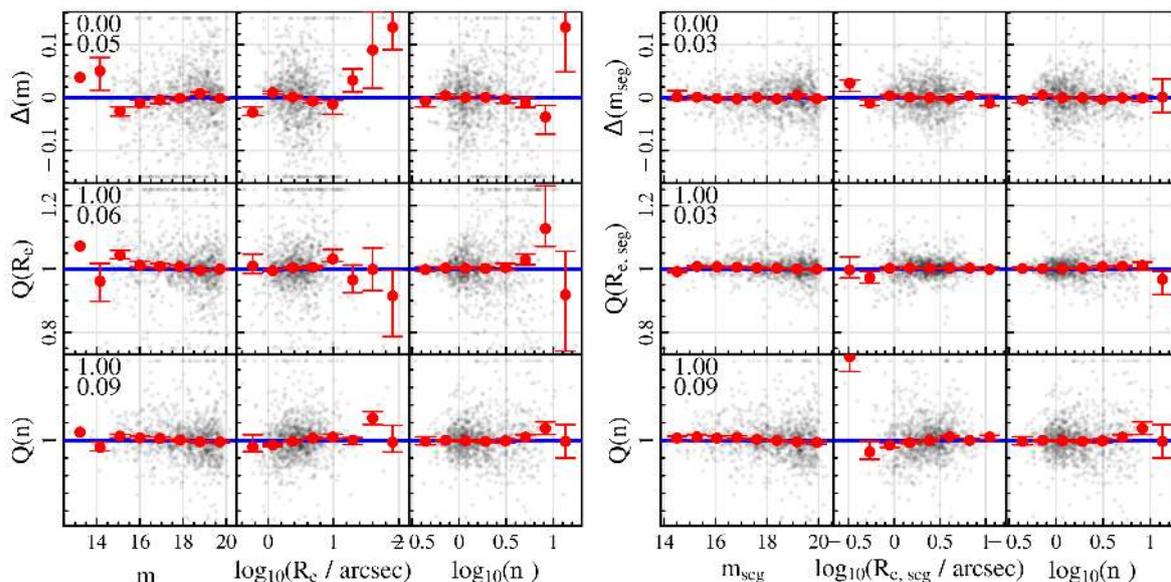


Figure 4.4: Similar to Figure 4.3 but now showing an internal consistency check of our catalogue using galaxies that were imaged (and successfully fitted) at least twice in the KiDS r -band. We pass these duplicate observations of the same physical objects through our pipeline independently and then compare the fit to the shallower image with the fit to the deeper (higher signal-to-noise ratio) image. Note the different plotting ranges relative to Figure 4.3 (especially on the y -axis).

different observations are treated entirely independently. We always use the deeper image as the reference image (the image depth at the edge of KiDS tiles can vary greatly depending on the number of dithers - between 1 and 5 - that cover the area). For the right set of panels, we evaluate the magnitude and effective radius within whichever of the two segment radii is smaller to avoid extrapolation and obtain consistent results.

For both versions (Sérsic parameters extrapolated to infinity on the left and truncated to segment radii on the right), there are very little differences between the two fits to the same galaxy; and there are no systematic trends across the parameter space. The running median is consistent with 0 (or 1, for scale parameters) in almost all bins, which shows that there are no inherent systematic differences in our fits related to image depth. This holds true despite the segments being systematically larger for the deeper images. The difference in segment size is too small on average to detect systematic trends across the entire sample: the median difference between the two segment radii is only 6%. However, there are a number of outliers visible as rows of points at the top and bottom of the panels since they were clipped to the plotting intervals (for plotting only). These correspond to the fits where the segment sizes differ substantially. Evaluating both fits within the same region before comparison (right panels) removes those outliers and reduces the overall scatter.

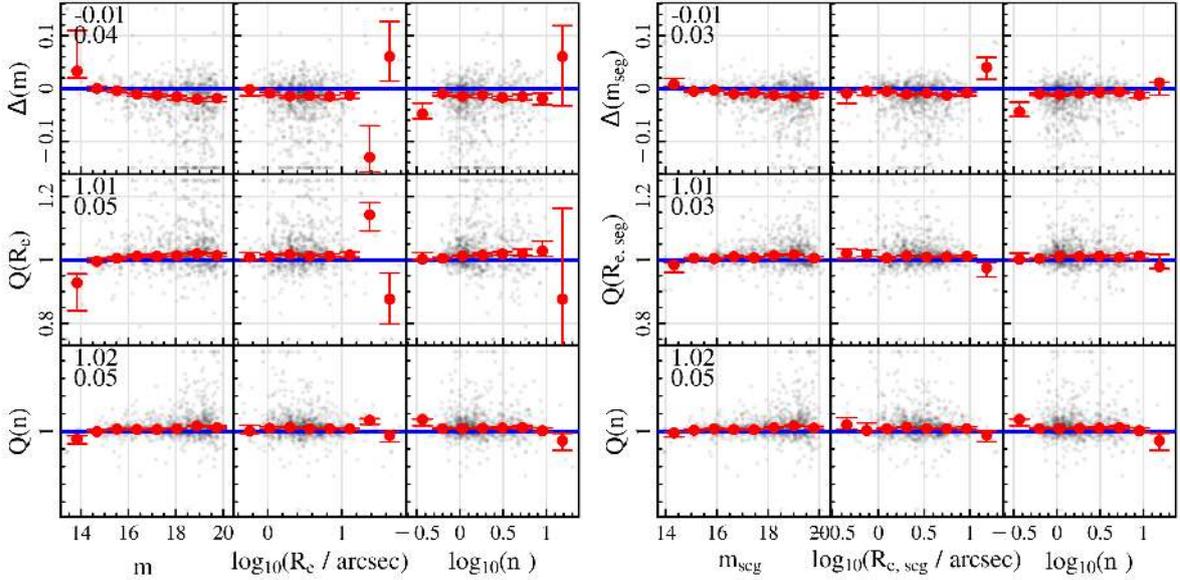


Figure 4.5: Similar to Figure 4.4 but comparing the fitted parameters of simulated images to the true (input) parameters; both limited to segment radii for the right set of panels.

4.2.2 Simulations: parameter recovery

As a final test of our pipeline, we ran simulations where we insert single Sérsic model galaxies convolved with an appropriate PSF at random locations in the KiDS data. To obtain a realistic distribution of parameters for the model galaxies (including correlations), we use the fitted single Sérsic parameters of a random sample of 1000 r -band galaxies that were not classified as outliers. The PSF to convolve with is taken as the model PSF that was fitted to the nearest real galaxy (at the position where the model galaxy is inserted), which is close to the real PSF at that image location. We then simply add the PSF-convolved galaxy to the KiDS data and run the resulting image through our entire pipeline (segmentation, sky subtraction, PSF estimation, galaxy fitting, outlier flagging, model selection).

In this way we are able to check for intrinsic biases in our entire pipeline, with 3 exceptions: issues due to galaxy features not represented by our models (bars, spiral arms, disk breaks, mergers, etc.), problems in the data processing performed by the KiDS team (if any), and deviations of the true PSF from a Moffat function.

In Figure 4.5 we show the corresponding plots to Figures 4.3 and 4.4; where on the x -axis we now have the true (input) parameters of our simulated galaxies and on the y -axis the difference between the fitted and the true values; both limited to segment radii for the right set of panels.

Generally, all parameters are recovered well, although for both versions of the plot the magnitudes show a slight offset of ~ 0.01 mag (with corresponding trends in effective radius and

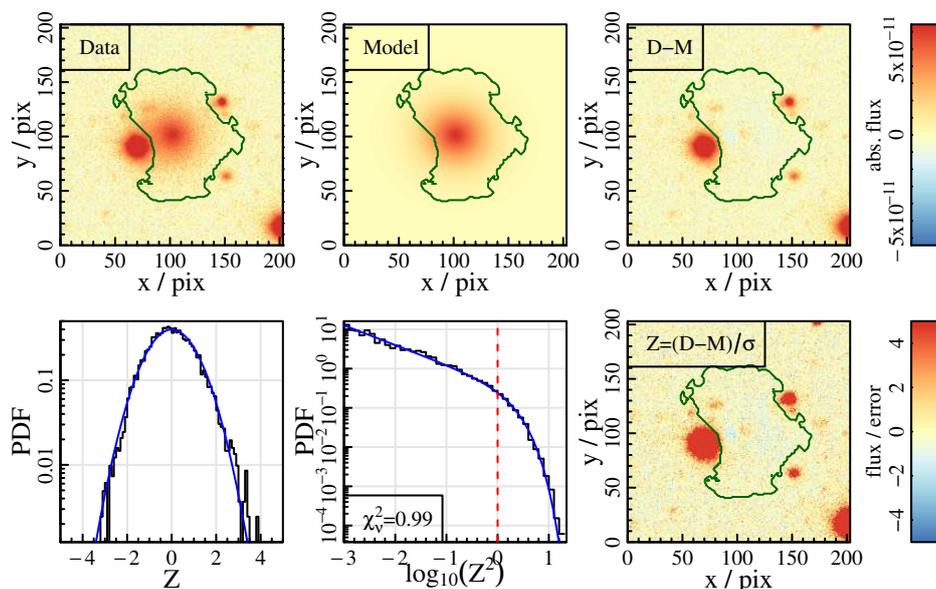


Figure 4.6: An example fit to a simulated galaxy where the difference between the true and the fitted magnitude is large due to the wings of a nearby bright object and a small faint object included in the segmentation map. Panels are the same as the top two rows in Figure 2.2.

Sérsic index since these parameters are correlated); worsening for faint objects. This offset is driven by a number of galaxies scattering to very low values, i.e. where the fit attributes significantly more flux to the galaxy than what we put into the simulation. Visual inspection of these simulated objects revealed that all of them have additional flux from other objects included in the segmentation maps. Figure 4.6 shows an example, where the difference between the fitted and the true magnitude is -0.17 for the values extrapolated to infinity and -0.14 for the segment truncated values. Truncating to segment radii only leads to limited improvement of the agreement since the cause of the offsets observed in Figure 4.5 is flux from nearby objects; and all simulated galaxies are intrinsically represented well by a single Sérsic model as they were created as such.

Nearby objects affect approximately 5-10% of our simulated fits in this way. Since it is a one-sided effect (there are no sources with negative flux), it results in a slight overall bias across the sample. This is expected to occur at a similar level also in the fits to real galaxies and could only be improved by simultaneously fitting nearby sources (see also the discussion of this issue in Häussler et al. 2007). However, for this work we decided against this option as explained in Section 2.1.2. We may revisit this decision in future work.

4.2.3 Simulations: model selection accuracy

Since we know that all of our input galaxies were perfect single Sérsic systems, the model selection and outlier rejection statistics can be used to judge the failure rate of these routines

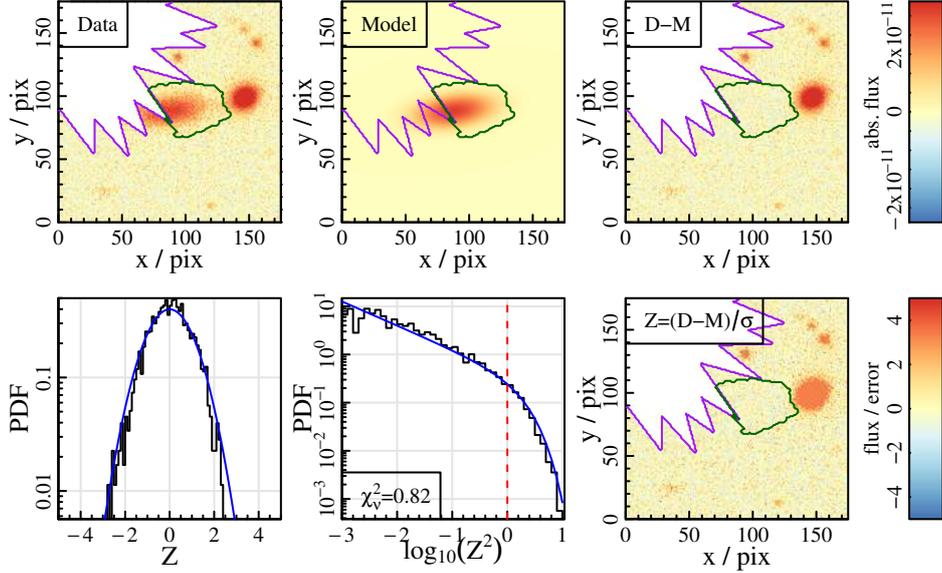


Figure 4.7: An example fit to a simulated galaxy which was flagged as a bad fit due to a nearby masked area from a bright star chopping up the segmentation map. Panels are the same as the top two rows in Figure 2.2.

(cf. also Sections 2.1.3 and 3.2). We simulated 1000 galaxies at random locations; which resulted in 1126 objects to be fit (due to the overlap regions between KiDS tiles). Of these, 262 (23%) were skipped; which is similar to the fraction of skipped fits for real galaxies, as expected since the main reason for this are the KiDS masks. Of the remaining objects, 94% are classified as single component fits, 3% are 1.5- or double component fits and 3% are flagged as outliers.

The number of outliers is significantly less than the 11% of real r -band galaxies flagged (Section 2.1.3) because in the simulations all galaxies are intrinsically “well-behaved”. Figure 4.7 shows an example of the most commonly occurring reason for being flagged as an outlier (in the simulations), namely the mask of a nearby bright star chopping up the segmentation map.

The fact that 97% of non-outlier simulated galaxies are correctly classified as single component fits confirms that model selection is accurate provided the galaxy can be unambiguously assigned to the single Sérsic model (cf. Section 2.1.3). We visually inspected the 3% 1.5- and double component fits and found nearby interfering objects in all of them. Figure 4.8 shows an example, where the fit attempts to capture the additional “features” with the freedom of a second component. Note that since we only simulated single Sérsic objects, we cannot comment on the accuracy of the model selection procedure for double component objects here. However, our model selection procedure was optimised on all types of real galaxies (not just single Sérsic objects); and our comparison to visual inspection in Section 2.1.3 also indicates a high accuracy for double component systems.

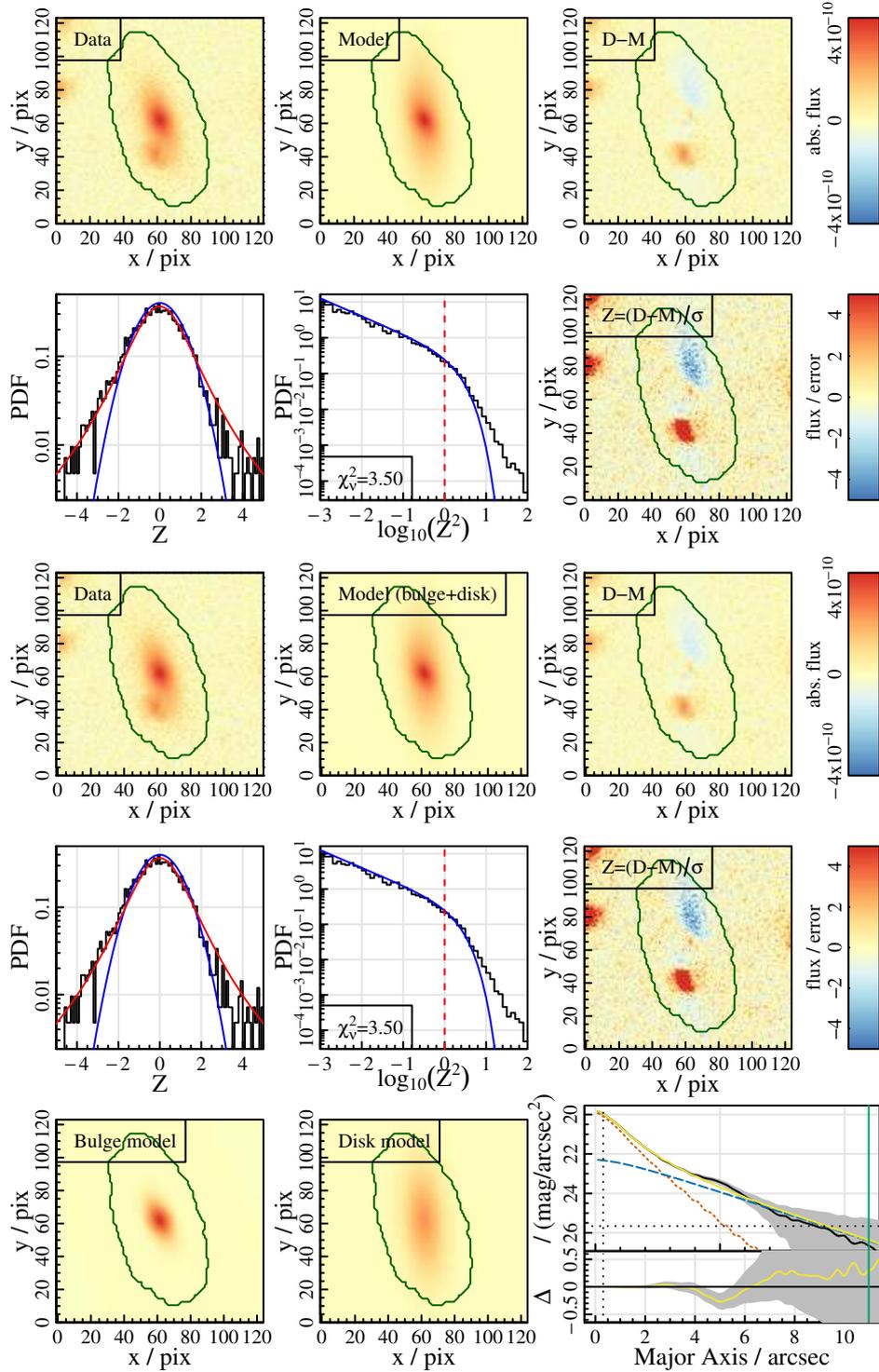


Figure 4.8: An example fit to a simulated galaxy which was classified as a double component fit. **First two rows:** the single Sérsic fit. **Last three rows:** the double component fit. Panels are the same as in Figure 2.2.

4.2.4 Systematic uncertainties

Figure 4.9 shows the results of our systematic error study, which we will now discuss in detail. Going from left to right we show three different plot types as labelled on the x -axis and described in more detail in the caption and below. Going from top to bottom, each plot type is shown for each single Sérsic parameter as labelled in the left panels; and the colours of the lines in the plot indicate which sample was used according to the legend at the bottom of the figure. The left panels show the distribution of absolute differences between the values fitted to both versions of a galaxy (in the indicated units); the middle panels show the corresponding error distribution (errors added in quadrature for the two fits; in the same units as the fit); and the right panels show the distribution of the absolute difference divided by its error (unitless). The solid black lines labelled “real data overlap” are using the overlap sample. The dashed dark blue lines labelled “simulation overlap” show the same for the overlap sample in simulated galaxies, where we have run more simulations specifically to boost the number of simulated overlap galaxies to a similar value as we have for the real overlap sample (~ 700). The dashed light blue line labelled “sim. overlap (t. PSF)” is the same as the dashed dark blue line, just that when fitting the galaxy, instead of the estimated PSF we passed in the true PSF (i.e. the one we used to convolve the model galaxy with originally). The dotted dark orange line labelled “simulation fit vs. true” shows the difference between the fitted value and the true value (instead of between the two fitted values in the overlap region) for the same sample of galaxies. Note the errors here now are just the errors of the fit since the true values do not have errors. The dotted light orange line (“sim. fit vs. true (t. PSF)”) is the same for the run that used the true PSFs.

All values are clipped to the plotting intervals (for plotting only). For scale parameters, all distributions are shown in logarithmic space (which the parameters were also fitted in). To make the scales comparable to the other parameters, the angle is shown in units of 30° (which it was also fitted in to make the MCMC step size comparable to the other parameters). For the magnitude and effective radius, we show both the fitted Sérsic values and the segment truncated values. Comparing rows 3 and 4 (Sérsic and segment magnitudes) or rows 5 and 6 (Sérsic and segment effective radii) against each other, it becomes clear immediately that the distributions for the segment values are narrower, i.e. limiting to segment sizes increases the stability and reduces the scatter in those parameters as already observed in many previous sections. Note, though, that our simulated galaxies do follow perfect single Sérsic profiles, so the differences between the segment and Sérsic values are generally expected to be much smaller in the simulations than in real data (see Section 2.1.3 for details).

For each distribution, we also give the median of the absolute difference and the errors; and the 1σ -quantile (half of the range containing the central 68% of data) of the normalised difference. These values (with uncertainties) are also given in Table 4.2.

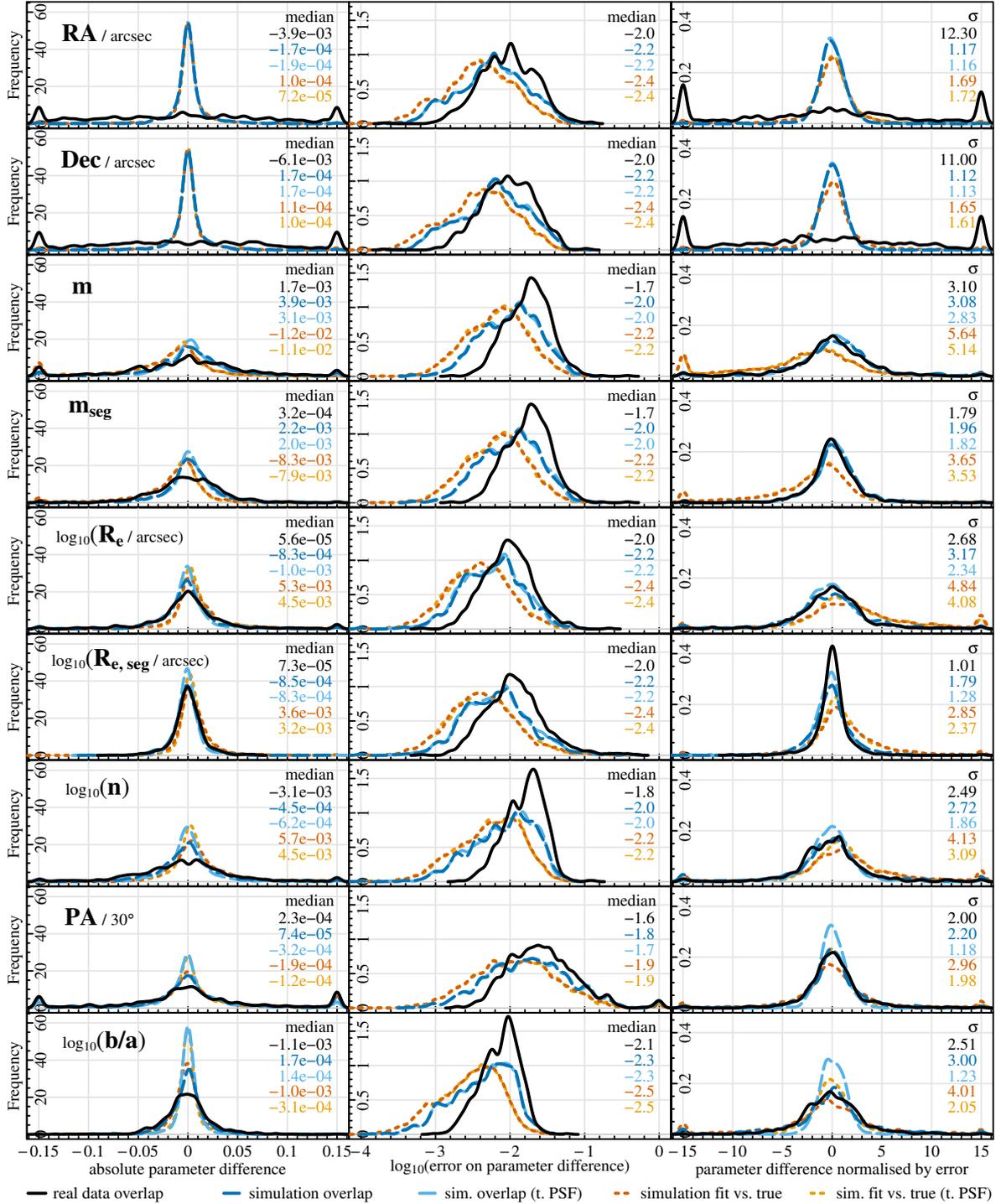


Figure 4.9: For all single Sérsic parameters as labelled top to bottom: **Left column:** The distribution of the absolute difference between the fitted and true values for simulated galaxies; or between the fitted values to two versions of the same (simulated or real) galaxy in the overlap sample. The legend at the bottom indicates which difference is shown; scale parameters are treated in logarithmic space throughout. See text for details. **Middle column:** The error on the parameter difference shown in the left column. **Right column:** The parameter difference normalised by its error (i.e. left column divided by middle column).

Overlap sample: real vs. simulated

Focusing on the real data and simulation overlap samples (solid black and dashed dark blue lines), which are most directly comparable, it can be seen that the distributions of the parameter differences (left column) are broader for the real galaxies than for the simulated galaxies for all parameters; i.e. for two versions of the same galaxy in the overlap sample, the fitted values are on average closer to each other in the simulation than in the real data. This could be due to two reasons: either irregular galaxy features in combination with noise (i.e. perfect single Sérsic objects are just more easily constrained/less easily influenced by noise fluctuations); or differences in the KiDS data processing between tiles (e.g. inaccuracies in their background subtraction procedure) that affect the real galaxies but not the simulated ones since those were added later. In reality, it is probably a combination of these two effects (with the first one presumably dominating). All steps of our own analysis affect both the simulated and the real galaxies and on average will have the same effect on both.

The errors (second column) do reflect this additional uncertainty in real galaxies in that they are larger by 0.2-0.3 dex for all parameters. In fact, the errors on the simulated galaxies seem to be more severely underestimated than those on the real galaxies, which becomes clear when looking at the parameter differences normalised by the respective errors (third column). In an ideal world, these would all be Gaussians centred on zero with a standard deviation of 1. As there will always be a few outliers due to interfering objects or image artifacts, instead of the mean and standard deviation we will consider their more robust equivalents, the median and 1σ -quantile (shown in plots). All overlap sample distributions (simulated and real, i.e. black, dark blue and light blue lines) are centred on zero, as already expected from the results shown in Figure 4.4. However, it can be seen that for all parameters except segment effective radius, the 1σ -quantile is larger than 1 both for the simulated and the real sample: values generally range between 2 and 3 with simulated galaxies performing slightly worse due to the underestimated errors (which are most likely caused by the PSFs, see discussion below).

The exception to this is position (RA and Dec, top two rows), for which the normalised distribution is a factor of approximately 10 broader for real galaxies than for simulated ones. In fact, a considerable fraction of these distributions fall outside of the plotting range, such that the clipping to these intervals results in prominent peaks at the plot edges (top right two panels of Figure 4.9). We believe this to be mainly due to the accuracy of the astrometric solution of the KiDS data, which shows a scatter of approximately $0''.04$ in DR4.0 in both RA and Dec (Kuijken et al. 2019; and we also confirmed this using the KiDS r -band source catalogues). This is a factor of 4 larger than the median MCMC error on position (top two panels in the middle row of Figure 4.9). Accounting for this additional source of scatter between the tiles (which only affects real objects but not the simulations since those were inserted after the astrometric calibration) would bring the normalised error distribution for

the real data overlap sample into much closer agreement with the simulated version. The remaining factor of ~ 2 -3 difference could also be due to the astrometry, considering that the overlap sample by definition sits at tile edges, where the astrometric solution is the most uncertain; or - as for all other parameters - due to irregular galaxy features in combination with noise (see discussion above). As a last point we would like to note that the absolute differences in position are usually still within 1 pixel ($0''.2$), i.e. although it stands out from the plot, this is a sub-pixel effect.

Simulated overlap: imperfect vs. true PSFs

These distributions of the simulated overlap sample can be compared to their equivalent distributions using the true PSFs (dashed light blue lines in Figure 4.9). This allows us to determine which parameters are affected by imperfect PSF estimates. However, we note that we can only make qualitative and relative statements here since we do not know how close to the truth our estimated PSFs for the real galaxies are. When simulating our galaxies, we convolve it with the model PSF fitted to the nearest real galaxy, i.e. this is the true PSF (cf. Section 4.2.2). When processing the simulated galaxy through our pipeline, the estimated PSF is then obtained by fitting nearby stars in the usual way. The nearest galaxy (which the true PSF is based on) is typically around $200''$ away, with the distribution ranging between ~ 0 and $\sim 500''$. This is close enough to provide a realistic PSF for the position of interest since KiDS tiles are much larger than this ($\sim 1 \text{ deg}^2$) and the PSF varies only slowly across the tiles. However, it is further away than the stars used to obtain the estimated PSF, which are typically within ~ 100 - $200''$ and can at the very maximum be $\sqrt{2} \times 200''$ away since the large cutouts used for PSF fitting are $400''$ on each side. This is expected since the density of stars is much higher than that of GAMA galaxies in KiDS data. However, it implies that the deviations of the estimated to the true PSFs in the simulations will on average be larger than for real data, leading to the errors on simulated galaxies being more severely underestimated as noted above.

Comparing the simulations with true PSFs to the real data, it can be seen that the simulations now perform better than the real data for all parameters except segment effective radius (1σ -quantiles between 1 and 2.8). In addition, comparing the simulations with the true and the false PSFs against each other, we can assess which parameters are most affected (relatively speaking): position angle and axial ratio are most severely influenced; followed by Sérsic index, effective radius and magnitude, while the position is nearly unaffected. This makes sense: the axial ratio and position angle are very sensitive to mistakes in the ellipticity and orientation of the PSF; while the fitted Sérsic index, effective radius and magnitude depend on the concentration and FWHM of the PSF. The position is only very weakly affected since the PSF is always centred and symmetric.

Note that all 1σ -quantiles are still larger than 1 (ranging from 1.1 to 2.8) even for the simulations with true PSFs. This indicates that the error underestimates of these parameters are not exclusively caused by the effects studied so far (galaxy and/or image processing features not accounted for in the simulations and PSF uncertainties) but there is an additional contribution from features that are also present in the simulations such as nearby objects, noise fluctuations, background subtraction inaccuracies or image artifacts.

Simulated sample: fitted vs. true values

Finally, for the simulated samples we can compare the fitted values to the true values (instead of the overlap sample comparison), which is shown as dotted orange lines in Figure 4.9. This allows us to detect biases, but is less directly comparable to the sample of real galaxies where the true value is unknown. Note the errors are generally slightly smaller compared to the overlap studies since for those, the errors for both fits were added in quadrature while the true values now do not have errors. Correspondingly, the normalised distributions are slightly broader even though the absolute differences between parameters are comparable. Most notably, however, the median of the distribution is now shifted away from zero for magnitude, effective radius and Sérsic index (see also Table 4.2). This is due to the bias caused by nearby objects already described in Section 4.2.2: magnitudes are too bright by ~ 0.01 mag; effective radii and Sérsic indices too large by approximately 1% (always a bit better for segment values and/or simulations using the true PSFs). All other parameters still have their distributions centred on zero, i.e. do not show any bias - at least not one that we can test with our simulations. This makes sense since position, axial ratio and angle will be influenced by nearby objects (and other effects) as well, but without any preferred direction and so on average this leads to an error underestimate rather than an overall bias. Using the true PSFs (dotted light orange lines) narrows all distributions slightly as expected; but again there is only an error underestimate rather than an overall bias introduced by the wrong PSFs since they are “randomly wrong”.

One source of potential bias that we cannot test with the simulations are galaxy features not accounted for in the models. If, for example, there is a large population of galaxies that have bars; and these bars lead to the bulge axial ratios being systematically underestimated, this is again a one-sided effect that could lead to an overall bias. For this reason, we explicitly use the term “bulge” in its broadest sense, including all kinds and possible combinations of central galaxy components. In addition, such features could further increase the error underestimate because they will tend to influence both fits to a galaxy in the overlap sample in the same way and hence are difficult to detect in the above analysis. If there are systematic one-sided deviations of the true PSFs from Moffat functions, these could lead to one-sided systematically wrong PSF estimates which could in turn also introduce an additional bias that cannot be

Table 4.2: Biases and error underestimates for all single Sérsic parameters derived from our systematic error studies (Section 4.2.4). The bias is additive (indicated with \pm) for those parameters that were treated in linear space and multiplicative (indicated with $*$) for those treated in logarithmic space. Error underestimates are always multiplicative. The column “bias/ σ ” gives the significance of each bias.

param.	bias (using true PSFs)	bias/ σ	error underest.
RA	$\pm (7 \pm 18) \times 10^{-5}$ arcsec	0.39	12.27 ± 0.71
Dec	$\pm (10 \pm 19) \times 10^{-5}$ arcsec	0.56	10.98 ± 0.53
m	$\pm (-11.4 \pm 0.8) \times 10^{-3}$	-13.83	3.10 ± 0.16
m_{seg}	$\pm (-8 \pm 0.5) \times 10^{-3}$	-14.92	1.79 ± 0.09
R_e	$*$ (1.0105 \pm 0.0009)	11.61	2.68 ± 0.12
$R_{e,seg}$	$*$ (1.0074 \pm 0.0005)	14.39	1.01 ± 0.05
n	$*$ (1.010 \pm 0.001)	10.93	2.49 ± 0.10
PA	$\pm (-4 \pm 19) \times 10^{-3}$ deg	-0.19	2.00 ± 0.10
b/a	$*$ (0.9993 \pm 0.0004)	-2.07	2.51 ± 0.09

tested by the simulations which use Moffat model PSFs. However, based on the PSF quality control, we do not believe that our PSF estimates are systematically wrong, see discussion in Section 2.2.3.

Corrections for systematics and their validity

Table 4.2 summarises the results of the systematic error studies: for all single Sérsic parameters (plus the segment magnitude and segment effective radius), we give the average bias and error underestimates with uncertainties. The bias is estimated from the median of the offset between fitted and true values in the simulation using the true PSFs (light orange numbers in the first column of Figure 4.9). The errors on the median are taken as the 1σ -quantiles of these distributions divided by the square-root of the number of data points (~ 2000). Logarithmic parameters are converted back into linear space to simplify bias correction in the catalogue. Nonetheless, scale parameters have multiplicative correction factors while location parameters have additive corrections (in given units). In other words: to correct for the bias, subtract the values in Table 4.2 from the catalogue values for position, magnitude and position angle; and divide by the given values for effective radius, Sérsic index and axial ratio. Users should note, however, that due to the way these biases were estimated, they do not include all sources of potential bias (e.g. galaxy features such as bars and spiral arms; or systematically wrong PSFs). Also, we recommend to apply the bias correction only to statistically large and random samples; they are average values not applicable to individual galaxies as evident from Figure 4.5.

The next column in Table 4.2 gives the significance of each bias, which is the deviation of the median from 0 (or 1 for scale parameters) divided by its error. It can be seen that

4.2. SYSTEMATIC UNCERTAINTIES AND BIASES

position, position angle and axial ratio are not biased (consistent with 0/1 within 2σ), while magnitude, effective radius and Sérsic index are biased (deviation from 0/1 of $> 5\sigma$); as found and discussed before.

Finally, the last column in Table 4.2 gives the error underestimates estimated from the width of the distribution of the normalised difference between fits to two versions of the same galaxy in the overlap sample (black numbers in the last column of Figure 4.9). The uncertainties in this case were estimated by bootstrapping the distributions 1000 times each to get an estimate of the variation of the distribution width. Since these distributions were normalised (by the respective errors), there is no need to convert between linear and logarithmic space and the given values can directly be used as correction factors for the MCMC errors. We have applied the relevant correction to all quoted errors in the catalogue, but also give the original (purely random) errors for completeness. Also note that since these values are now based on real data, they do include PSF uncertainties (in contrast to the biases). Also, since the overlap sample is in many ways the worst in terms of data quality (sitting at tile edges), these are likely upper limits. However, they still do not include error underestimates caused by (galaxy) features not accounted for in the models such as bars, rings, spiral arms or similar, as well as nearby objects. Since these are physical (rather than related to the data taking or image processing), they will be present in both versions of the overlap sample galaxy and influence both fits in similar ways - leading to a (random) bias on individual galaxy fits; and hence an error underestimate for a large enough sample. These issues should be kept in mind when using the catalogue.

5 Summary, conclusion and outlook

In this thesis we presented our pipeline for the single Sérsic fits and bulge-disk decompositions of 13096 galaxies at redshifts $z < 0.08$ in the GAMA II equatorial survey regions in the KiDS u, g, r, i and the VIKING Z, Y, J, H, K_s bands. The galaxy modelling is done using `ProFit`, the Bayesian two-dimensional surface profile fitting code of Robotham et al. (2017), fitting three models to each galaxy:

1. a single Sérsic component,
2. a two-component model consisting of a Sérsic bulge plus exponential disk and
3. a two-component model consisting of a point source bulge plus exponential disk (for unresolved bulges).

The preparatory work (image segmentation, background subtraction and obtaining initial parameter guesses) is carried out using the sister package `ProFound` (Robotham et al., 2018); with the PSF estimated by fitting nearby stars using a combination of `ProFound` and `ProFit`. Segmentation maps are defined on joint gri -images, while the remaining analysis is performed individually in each band except for the model selection, for which we offer both a per-band and a joint version. The analysis is fully automated and self-contained with no dependency on additional tools.

In addition to the galaxy fitting, we performed a number of post-processing steps including the flagging of bad fits and model selection. An overview of the number of galaxies successfully fitted in each band as well as the number classified in each category is given in Tables 3.1 and 3.2 as well as Figure 3.1. For our planned applications of the catalogue, which involves the statistical study of dust attenuation effects, we need fits that are most directly comparable to each other. Hence, we choose to model a maximum of two components for each galaxy even if more features may be present; and focus on achieving good fits in the high signal-to-noise regions of the galaxies by using relatively small segments for fitting. Consequently, we recommend using truncated magnitudes and effective radii for all analyses instead of the Sérsic values which are extrapolated to infinity. The quality of the fits was ensured by visual inspection, comparing to previous works (Kelvin et al., 2012; Lange et al., 2015), studying independent fits of galaxies in the overlap regions of KiDS tiles and bespoke simulations. The latter two were also used for a detailed analysis of how systematic uncertainties affect our fits.

We found that the combination of `ProFound` and `ProFit` is well-suited to the automated analysis of large datasets. The fully Bayesian MCMC treatment enabled by `ProFit` is able to overcome the main shortcomings of traditionally used downhill-gradient based optimisers, namely their susceptibility to initial guesses and their inability to easily derive realistic error estimates. The watershed deblending algorithm used by `ProFound` is less prone to catastrophic segmentation failures and allows us to extract more complex object shapes than other commonly used algorithms based on elliptical apertures; while still preserving the total flux well. With its wealth of utility functions, it not only facilitates the robust segmentation of large sets of images but also provides sky background estimates and reasonable initial guesses for the MCMC fitting.

These characteristics, in combination with our own routines for quality assurance, led to results that are robust across a variety of galaxy types and image qualities and in reasonable agreement with previous studies given the different data, code and focus of the analysis. The outlier rejection routine efficiently identifies objects for which none of our models is appropriate such as irregular galaxies or those compromised by masked areas. Model selection is based on a ΔDIC cut and accurate to $> 90\%$ compared to what could be achieved by visual inspection. There is a minimal bias in the fitted magnitude, effective radius and Sérsic index of approximately 0.01 mag, 1% and 1% respectively (on average across the full sample) caused by excess flux from nearby other objects. The errors obtained from the MCMC chains are underestimated with respect to the true errors by factors of typically between 2 and 3 (see Table 4.2) and can easily be corrected for statistically large samples of galaxies.

All results are integrated into the GAMA database as part of the `BDDcomp` DMU. The DMU consists of a number of catalogues giving the results of the preparatory work, the 2D surface brightness distribution fits and the post-processing of all 13096 galaxies in our full sample ($z < 0.08$ in the GAMA II equatorial survey regions) in all bands; with additional diagnostic plots and all fit inputs available on the GAMA file server (see Section 3.1 for details). So far, four DMU versions have been released, while `v05` will be made available alongside the publication of this thesis. The full DMU is currently available to GAMA team members with a version restricted to SAMI galaxies available to the SAMI team. It will be made publicly available in one of the forthcoming GAMA data releases. Readers interested in using (parts of) the catalogue before it is publicly released are encouraged to contact the authors to explore the possibilities for a collaboration¹.

Both the GAMA and the SAMI teams have actively made use of the catalogue versions already released, with many more studies in progress and planned (see Chapter 1). We therefore detailed not only the final version, but also different stages during pipeline development, including quality control steps of the preparatory work, in Chapter 2. Since `v05` of the

¹<http://www.gama-survey.org/collaborate/>

catalogue has not been previously published, a particular focus is placed on ensuring the reliability of those results by comparing it to v04 (Chapter 3). The latter has benefitted from an extensive quality control detailed in Chapter 4.

In addition to the projects of collaborators mentioned above, we have many own plans and ideas for scientific analyses on the basis of our bulge-disk decomposition results. These include studying bulge and disk colours and trends of structural parameters with wavelength as well as deriving component stellar masses with the aims of investigating the stellar mass functions, stellar mass-to-light ratios and scaling relations such as the size-mass relation for individual components as a function of wavelength. By comparison with dust radiative transfer models, we will then also be able to constrain the nature and distribution of dust in galaxy disks, thereby contributing to the understanding of systematic uncertainties affecting studies of galaxy structure, formation and evolution, which in turn are vital for our understanding of the universe as a whole (see Section 1.1.3).

On the technical side, further development of the pipeline foresees a number of minor improvements in the near future, for example considering the uncertainty of the astrometric solution for the joint fit, simultaneously fitting nearby sources and improving our treatment of the VIKING data by accounting for the slight differences in pixel size between frames, using the provided confidence maps and generating bright star masks. In the longer term, we will exploit the relatively new multi-band fitting functionality of `ProFit` - or better yet the new package `ProFuse` - to achieve simultaneous fits across all wavelength bands with physically motivated variations of structural parameters. Improvements to the preparatory work pipeline foresee the creation of a joint segmentation map including all bands, eliminating all manually calibrated tuning parameters and simultaneously fitting all suitable stars for more robust PSF estimates. In the post-processing, the main focus of development is on minimising manual intervention - especially the re-calibration of model selection - to allow easy scaling of the code to larger samples of galaxies from potentially different bands and datasets.

To summarise, we obtained a catalogue of robust structural parameters for the components of a sample of 13096 nearby GAMA galaxies while at the same time contributing to the advancement of image analysis, surface brightness fitting and post-processing routines for quality assurance in the context of automated large-scale bulge-disk decomposition studies. The further development of such methods and new approaches is vital to fully exploit the data of future sky surveys that will provide multi-wavelength imaging for millions of galaxies at unprecedented depth and resolution. The resulting measured parameters in turn are crucial to test and improve theoretical models and simulations and ultimately understand the formation, structure, composition and evolution of our universe.

Acknowledgements

GAMA is a joint European-Australasian project based around a spectroscopic campaign using the Anglo-Australian Telescope. The GAMA input catalogue is based on data taken from the Sloan Digital Sky Survey and the UKIRT Infrared Deep Sky Survey. Complementary imaging of the GAMA regions is being obtained by a number of independent survey programmes including GALEX MIS, VST KiDS, VISTA VIKING, WISE, Herschel-ATLAS, GMRT and ASKAP providing UV to radio coverage. GAMA is funded by the STFC (UK), the ARC (Australia), the AAO, and the participating institutions. The GAMA website is <http://www.gama-survey.org/>.

Based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs 177.A-3016, 177.A-3017, 177.A-3018 and 179.A-2004, and on data products produced by the KiDS consortium. The KiDS production team acknowledges support from: Deutsche Forschungsgemeinschaft, ERC, NOVA and NWO-M grants; Target; the University of Padova, and the University Federico II (Naples).

This publication has made use of data from the VIKING survey from VISTA at the ESO Paranal Observatory, programme ID 179.A-2004. Data processing has been contributed by the VISTA Data Flow System at CASU, Cambridge and WFAU, Edinburgh.

The total computing time for all runs and test runs performed in the context of this thesis sums to approximately 676 700 hours on CPU. With an electricity usage of 12 W per CPU, our total energy consumption for the project is 8120 kWh. Assuming an average CO₂ equivalent of 500 g/kWh (corresponding to the average of the German energy mix during the period of the project²), results in 4 tons of CO₂ emitted by this PhD thesis. This is comparable to the worldwide average of CO₂ emissions per person per year.³ It is still a lower limit since it does not count any non-automated calculations nor failed runs; and completely ignores other activities related to the project, such as travelling.

²<https://www.umweltbundesamt.de/presse/pressemitteilungen/bilanz-2019-co2-emissionen-pro-kilowattstunde-strom>

³<https://de.statista.com/statistik/daten/studie/167877/umfrage/co-emissionen-nach-laendern-je-einwohner/>

Thanks

First of all, I would like to thank my supervisor Jochen Liske for going through this project with me start to end. Jochen is a very skilled and enthusiastic scientist who is never short of ideas and was a great source of inspiration for me. On the more technical side, much support came from Aaron Robotham. I am very grateful to him for always being available and helping me out countless times during the course of my PhD. Furthermore, let me thank Robi Banerjee, Peter Hauschildt and Dieter Horns for being part of my defense committee.

It was great to be welcomed into the GAMA team, a very friendly and supportive group of scientists. Working with the SAMI and KiDS teams was also a pleasure, with members of both collaborations always happy to help out. My stays at ICRAR will remain in good memory due to the hospitality and friendliness of everybody I met there. While it is impossible to name everyone who contributed to the success of this project, a few people who come to my mind - apart from Jochen and Aaron of course - are Dan Taranu, Angus Wright, Jarkko Laine, Robin Cook, Hosein Hashemizadeh, Stefania Barsanti, Ben Henderson, Sree Oh, Simon Driver, Boris Häußler, Benne Holwerda, Ned Taylor, Amanda Moffett, Konrad Kuijken, Luke Davies, Michelle Cluver and Christos Georgiou.

A special thanks goes to Denis Wittor and Volker Heesen for proof-reading this thesis and suggesting many improvements. Janis Kummer was so kind to provide his thesis to use as a template and Jörg Knoche offered much needed emergency IT-support on several occasions. Together with all other current and former members of the Hamburger Sternwarte - many of whom I count to my friends - they made my PhD time a lot of fun and one of the best periods of my life.

Last but not least, my thanks goes to my family and friends - especially my parents, but also my sister, grandparents, some of the wider family and many close friends - all of whom always gave me the feeling that I could achieve anything if only I wanted to. I am extremely lucky to have Lars, the best husband that I could have wished for and an amazing dad to our son Tim. Without his unlimited support - and the happy nature of Tim - I could have never finished this project.

Bibliography

- Akaike H., 1974, *IEEE Transactions on Automatic Control*, 19, 716
- Allen P. D., Driver S. P., Graham A. W., Cameron E., Liske J., de Propris R., 2006, *MNRAS*, 371, 2
- Argyle J. J., Méndez-Abreu J., Wild V., Mortlock D. J., 2018, *MNRAS*, 479, 3076
- Baldry I. K., et al., 2010, *MNRAS*, 404, 86
- Baldry I. K., et al., 2012, *MNRAS*, 421, 621
- Baldry I. K., et al., 2014, *MNRAS*, 441, 2440
- Barden M., et al., 2005, *ApJ*, 635, 959
- Barnett R., Warren S. J., Cross N. J. G., Mortlock D. J., Fan X., Wang F., Hewett P. C., 2021, *MNRAS*, 501, 1663
- Barsanti S., et al., 2021, *ApJ*, 911, 21
- Bellstedt S., et al., 2020, *MNRAS*, 496, 3235
- Benson A. J., Džanović D., Frenk C. S., Sharples R., 2007, *MNRAS*, 379, 841
- Bertin E., 2010, *SWarp: Resampling and Co-adding FITS Images Together* (ascl:1010.068)
- Bertin E., 2011, in Evans I. N., Accomazzi A., Mink D. J., Rots A. H., eds, *Astronomical Society of the Pacific Conference Series Vol. 442, Astronomical Data Analysis Software and Systems XX*. p. 435
- Bertin E., Arnouts S., 1996, *A&AS*, 117, 393
- Blanton M. R., et al., 2003, *ApJ*, 594, 186
- Blanton M. R., Eisenstein D., Hogg D. W., Schlegel D. J., Brinkmann J., 2005, *ApJ*, 629, 143
- Bluck A. F. L., et al., 2019, *MNRAS*, 485, 666
- Boone E. L., Merrick J. R., Krachey M. J., 2014, *Journal of Statistical Computation and Simulation*, 84, 833

- Bottrell C., Torrey P., Simard L., Ellison S. L., 2017, MNRAS, 467, 2879
- Bottrell C., Simard L., Mendel J. T., Ellison S. L., 2019, MNRAS, 486, 390
- Brooks S. P., Giudici P., Philippe A., 2003, Journal of Computational and Graphical Statistics, 12, 1
- Broyden C. G., 1970, Journal of the Institute of Mathematics and Its Applications, 6, 76
- Bryant J. J., et al., 2015, MNRAS, 447, 2857
- Casura S., et al., submitted, MNRAS
- Cluver M. E., et al., 2020, ApJ, 898, 20
- Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
- Cook M., Lapi A., Granato G. L., 2009, MNRAS, 397, 534
- Cook R. H. W., Cortese L., Catinella B., Robotham A., 2019, MNRAS, 490, 4060
- Davies L. J. M., et al., 2018, MNRAS, 480, 768
- Dimauro P., et al., 2018, MNRAS, 478, 5410
- Dimauro P., et al., 2019, MNRAS, 489, 4135
- Domínguez Sánchez H., Margalef B., Bernardi M., Huertas-Company M., 2022, MNRAS, 509, 4024
- Driver S. P., Popescu C. C., Tuffs R. J., Liske J., Graham A. W., Allen P. D., de Propris R., 2007a, MNRAS, 379, 1022
- Driver S. P., Allen P. D., Liske J., Graham A. W., 2007b, ApJ, 657, L85
- Driver S. P., Popescu C. C., Tuffs R. J., Graham A. W., Liske J., Baldry I., 2008, ApJ, 678, L101
- Driver S. P., et al., 2009, Astronomy and Geophysics, 50, 5.12
- Driver S. P., et al., 2011, MNRAS, 413, 971
- Driver S. P., et al., 2012, MNRAS, 427, 3244
- Driver S. P., Robotham A. S. G., Bland-Hawthorn J., Brown M., Hopkins A., Liske J., Phillipps S., Wilkins S., 2013, MNRAS, 430, 2622
- Driver S. P., et al., 2016, MNRAS, 455, 3911
- Driver S. P., et al., 2022, MNRAS,
- Dutton A. A., et al., 2011, MNRAS, 410, 1660

- Edge A., Sutherland W., 2020, VIKING (VISTA Kilo-degree Infrared Galaxy Survey): Imaging and Catalogue Data Release 4, ESO Science Archive Facility, <https://www.eso.org/rm/api/v1/public/releaseDescriptions/135>
- Edge A., Sutherland W., Kuijken K., Driver S., McMahon R., Eales S., Emerson J. P., 2013, *The Messenger*, 154, 32
- Erwin P., 2015, *ApJ*, 799, 226
- Fernández Lorenzo M., et al., 2014, *ApJ*, 788, L39
- Fletcher R., 1970, *Computer Journal*, 13, 317
- Gadotti D. A., 2009, *MNRAS*, 393, 1531
- Gao H., Ho L. C., 2017, *ApJ*, 845, 114
- Geweke J., Bernardo J. M., Berger J. O., Dawid A. P., Smith A., 1992, *Bayesian statistics*
- Goldfarb D., 1970, *Mathematics of Computation*, 24, 23
- de Graaff A., Trayford J., Franx M., Schaller M., Schaye J., van der Wel A., 2022, *MNRAS*, 511, 2544
- Graham A. W., 2013, *Elliptical and Disk Galaxy Structure and Modern Scaling Laws*. pp 91–140, doi:10.1007/978-94-007-5609-0_2
- Graham A. W., Driver S. P., 2005, *Publ. Astron. Soc. Australia*, 22, 118
- Gregory P., 2005, *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica® Support*. Cambridge University Press, doi:10.1017/CBO9780511791277
- Hammer François Puech M., Flores H., Rodrigues M., 2017, *Studying Distant Galaxies: a Handbook of Methods and Analyses*. WORLD SCIENTIFIC (EUROPE) (<https://www.worldscientific.com/doi/pdf/10.1142/q0016>), doi:10.1142/q0016
- Hashemizadeh A., et al., 2022, arXiv e-prints, p. arXiv:2203.00185
- Häussler B., et al., 2007, *ApJS*, 172, 615
- Häußler B., et al., 2013, *MNRAS*, 430, 330
- Häußler B., et al., 2022, arXiv e-prints, p. arXiv:2204.05907
- Head J. T. C. G., Lucey J. R., Hudson M. J., Smith R. J., 2014, *MNRAS*, 440, 1690
- Heidelberger P., Welch P. D., 1981, *Commun. ACM*, 24, 233
- Heidelberger P., Welch P. D., 1983, *Operations Research*, 31, 1109

- Hellinger E., 1909, *Journal für die reine und angewandte Mathematik*, 136, 210
- Hilbe J. M., de Souza R. S., Ishida E. E. O., 2017, *Bayesian Models for Astrophysical Data: Using R, JAGS, Python, and Stan*. Cambridge University Press, Cambridge
- Hopkins A. M., et al., 2013, *MNRAS*, 430, 2047
- Hubble E. P., 1926, *ApJ*, 64, 321
- Hyde J. B., Bernardi M., 2009, *MNRAS*, 394, 1978
- Ivezić Ž., et al., 2019, *ApJ*, 873, 111
- Jeffreys H., 1946, *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 186, 453
- Johnson S. G., 2017, The NLOpt nonlinear-optimization package. <http://ab-initio.mit.edu/nlopt>
- de Jong J. T. A., Verdoes Kleijn G. A., Kuijken K. H., Valentijn E. A., 2013, *Experimental Astronomy*, 35, 25
- de Jong J. T. A., et al., 2015, *A&A*, 582, A62
- de Jong J. T. A., et al., 2017, *A&A*, 604, A134
- Kelvin L. S., et al., 2012, *MNRAS*, 421, 1007
- Kelvin L. S., et al., 2014, *MNRAS*, 444, 1647
- Kembhavi A., Khare P., 2020, *Electromagnetic Radiation: The Key to Understanding the Universe*. Springer Singapore, Singapore, pp 5–32, doi:10.1007/978-981-15-5709-5_2, https://doi.org/10.1007/978-981-15-5709-5_2
- Kennedy R., et al., 2016, *MNRAS*, 460, 3458
- Kim K., Oh S., Jeong H., Aragón-Salamanca A., Smith R., Yi S. K., 2016, *ApJS*, 225, 6
- Kormendy J., 1977, *ApJ*, 218, 333
- van der Kruit P. C., Searle L., 1981, *A&A*, 95, 105
- Kruk S. J., et al., 2018, *MNRAS*, 473, 4731
- Kuijken K., et al., 2019, *A&A*, 625, A2
- La Barbera F., de Carvalho R. R., de La Rosa I. G., Lopes P. A. A., Kohl-Moreira J. L., Capelato H. V., 2010, *MNRAS*, 408, 1313
- Lackner C. N., Gunn J. E., 2012, *MNRAS*, 421, 2277

Bibliography

- Lagos C. d. P., et al., 2018, MNRAS, 473, 4956
- Lange R., et al., 2015, MNRAS, 447, 2603
- Lange R., et al., 2016, MNRAS, 462, 1470
- Lawrence A., et al., 2007, MNRAS, 379, 1599
- Liske J., et al., 2015, MNRAS, 452, 2087
- Meert A., Vikram V., Bernardi M., 2015, MNRAS, 446, 3943
- Meert A., Vikram V., Bernardi M., 2016, MNRAS, 455, 2440
- Mendel J. T., Simard L., Palmer M., Ellison S. L., Patton D. R., 2014, ApJS, 210, 3
- Méndez-Abreu J., et al., 2017, A&A, 598, A32
- Moffat A. F. J., 1969, A&A, 3, 455
- Moffett A. J., et al., 2016, MNRAS, 462, 4336
- Nammour F., Schmitz M. A., Ngolè Mboula F. M., Starck J.-L., Girard J. N., 2021, J Fourier Anal Appl, 27, 88
- Nedkova K. V., et al., 2021, MNRAS, 506, 928
- Nolte A., Wang L., Bilicki M., Holwerda B., Biehl M., 2019, Neurocomputing, 342, 172
- Oh S., et al., 2020, MNRAS, 495, 4638
- Oh S., et al., 2022, MNRAS, 512, 1765
- Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2010, AJ, 139, 2097
- Pillepich A., et al., 2018, MNRAS, 473, 4077
- Popescu C. C., Tuffs R. J., Dopita M. A., Fischera J., Kylafis N. D., Madore B. F., 2011, A&A, 527, A109
- R Core Team 2020, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>
- Robotham A., et al., 2010, Publ. Astron. Soc. Australia, 27, 76
- Robotham A. S. G., Taranu D. S., Tobar R., Moffett A., Driver S. P., 2017, MNRAS, 466, 1513
- Robotham A. S. G., Davies L. J. M., Driver S. P., Koushan S., Taranu D. S., Casura S., Liske J., 2018, MNRAS, 476, 3137

-
- Robotham A. S. G., Bellstedt S., Lagos C. d. P., Thorne J. E., Davies L. J., Driver S. P., Bravo M., 2020, MNRAS, 495, 905
- Robotham A. S. G., Bellstedt S., Driver S. P., 2022, MNRAS,
- Rodriguez-Gomez V., et al., 2019, MNRAS, 483, 4140
- Roschlaub J.-L., 2022, B.Sc. thesis, Universität Hamburg
- Salo H., et al., 2015, ApJS, 219, 4
- Sánchez-Janssen R., et al., 2016, ApJ, 820, 69
- Schaye J., et al., 2015, MNRAS, 446, 521
- Schwarz G., 1978, Annals of Statistics, 6, 461
- Sérsic J. L., 1963, Boletín de la Asociación Argentina de Astronomía La Plata Argentina, 6, 41
- Shanno D. F., 1970, Mathematics of Computation, 24, 647
- Shen S., Mo H. J., White S. D. M., Blanton M. R., Kauffmann G., Voges W., Brinkmann J., Csabai I., 2003, MNRAS, 343, 978
- Shibuya T., Ouchi M., Harikane Y., 2015, ApJS, 219, 15
- 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
- Sivia D. S., Skilling J., 2006, Data Analysis: A Bayesian Tutorial, 2nd edn. Oxford Science Publications, Oxford University Press
- de Souza R. E., Gadotti D. A., dos Anjos S., 2004, ApJS, 153, 411
- Spiegelhalter D., Best N., Carlin B., Van Der Linde A., 2002, Journal of the Royal Statistical Society. Series B: Statistical Methodology, 64, 583
- Statisticat LLC. 2018, LaplacesDemon: Complete Environment for Bayesian Inference. <https://web.archive.org/web/20150206004624/http://www.bayesian-inference.com/software>
- Taranu D. S., 2022, AllStarFit: R package for source detection, PSF and multi-component galaxy fitting, Astrophysics Source Code Library (ascl:2201.005)
- Taranu D. S., et al., 2017, ApJ, 850, 70
- Tarsitano F., Bruderer C., Schawinski K., Hartley W. G., 2022, MNRAS, 511, 3330

Bibliography

- Tasca L. A. M., et al., 2014, *A&A*, 564, L12
- Taylor E. N., et al., 2011, *MNRAS*, 418, 1587
- Trujillo I., et al., 2006, *ApJ*, 650, 18
- Turchin V., 1971, *Theory of Probability and its Applications*, 16, 720
- Turner S., et al., 2021, *MNRAS*,
- de Vaucouleurs G., 1948, *Annales d'Astrophysique*, 11, 247
- Vika M., Bamford S. P., Häußler B., Rojas A. L., Borch A., Nichol R. C., 2013, *MNRAS*, 435, 623
- Vika M., Bamford S. P., Häußler B., Rojas A. L., 2014, *Monthly Notices of the Royal Astronomical Society*, 444, 3603
- Vogelsberger M., et al., 2014, *MNRAS*, 444, 1518
- van der Wel A., et al., 2012, *ApJS*, 203, 24
- van der Wel A., et al., 2014, *ApJ*, 788, 28
- Wright A. H., et al., 2019, *A&A*, 632, A34
- York D. G., et al., 2000, *AJ*, 120, 1579
- Zeilik M., Gregory S. A., 1998, *Introductory Astronomy and Astrophysics*, 4th edn. Saunders golden sunburst series, Fort Worth : Saunders College Pub.