

Dissertation  
submitted to the  
Combined Faculties of the Natural Sciences and Mathematics  
of the Ruperto-Carola University of Heidelberg, Germany  
for the degree of  
Doctor of Natural Sciences

Put forward by  
Ellen Andrae  
born in Berlin, Germany  
Oral examination: 27.06.2014



The GALEX-GAMA Survey and its  
Application to the Statistical Inference of  
the Attenuation of Starlight by Dust in  
Spiral Galaxies

Referees: Prof. Werner Hofmann  
Dr. Jochen Liske



## **Abstract**

This thesis describes a new widefield survey of the UV broadband emission from a complete flux-limited sample of galaxies in the local Universe conducted with the Galaxy Evolution Explorer (GALEX). I develop techniques to reliably measure UV photometry from galaxies catalogued in the Galaxy and Mass Assembly (GAMA) spectroscopic survey, accounting for the influence of varying UV background, neighbouring sources and measurement artefacts. I also develop the framework for a statistical technique for the inference of the intrinsic UV emission of stars in the population of spiral galaxies, correcting for the effects of attenuation of UV light by dust in flux-limited samples. This technique incorporates for the first time an explicit treatment of the dependency of the attenuation on orientation, size, and morphology of the galaxies. This is a prerequisite for the use of the UV data to infer the present and past evolution of local Universe galaxies from the UV-optical photometry.

## **Zusammenfassung**

Diese Arbeit stellt eine großflächige Himmelsdurchmusterung im Breitband-UV vor, die vom Galaxy Evolution Explorer (GALEX) durchgeführt wurde. Sie vermisst die UV-Emission einer flusslimitierten vollständigen Untergruppe von Galaxien im nahen Universum. Ich entwickle Methoden zur zuverlässigen Bestimmung der UV-Flüsse von Galaxien aus dem Katalog der Galaxy And Mass Assembly (GAMA) spektroskopischen Himmelsdurchmusterung. Diese Methoden berücksichtigen sowohl variierende UV-Hintergrundflüsse, als auch Konfusion durch benachbarte Quellen sowie Messartefakte. Des Weiteren entwickle ich eine statistische Methode zur Schätzung der intrinsischen UV-Emission der Spiralgalaxienpopulation, die der Abschwächung von UV-Strahlung durch Staub in der Scheibe der jeweiligen Galaxie Rechnung trägt und diese dafür korrigiert. Diese Methode ermöglicht es zum ersten Mal, die astrophysikalischen Zusammenhänge der Abschwächung von UV-Strahlung durch Staub in der Scheibe der jeweiligen Galaxie zu betrachten und dabei die Orientierung, Größe und Morphologie der Galaxie explizit in die Analyse mit einzubeziehen. Dies ist eine notwendige Voraussetzung, um aus den UV-Daten Rückschlüsse auf die gegenwärtige und vergangene Evolution von Galaxien im nahen Universum ziehen zu können.



## Acknowledgements

I would like to thank Mark Seibert for his assistance in compiling the GALEX data. GALEX (Galaxy Evolution Explorer) is a NASA Small Explorer, launched in April 2003. We gratefully acknowledge NASA's support for construction, operation, and science analysis for the GALEX mission, developed in cooperation with the Centre National d'Etudes Spatiales (CNES) of France and the Korean Ministry of Science and Technology.

I would then like to acknowledge the GAMA consortium. 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 programs 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>.

In particular, my thanks go to Meiert Grootes for the development of the cell-based selection technique for spiral galaxies. I would also like to thank Cristina Popescu for the development over many years of the radiation transfer model for the panchromatic analysis of the spectral energy distributions of spiral galaxies, which forms the basis of my technique for deattenuating the UV measurements. Without this model the statistical precision on which the scientific outcome relies would not have been possible.

Penultimately, I would like to thank my supervisor Richard J. Tuffs both for the countless fruitful discussions over the course of my thesis research, as well as for the comments and suggestions which have helped me improve this manuscript.

And finally, I would like to thank my family for their support, my little son for enforcing a work-life balance, and especially my husband for countless late-night fruitful discussions and for proof-reading my drafts.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>I</b>	<b>The GALEX-GAMA survey</b>	<b>9</b>
<b>2</b>	<b>GALEX-GAMA input data</b>	<b>11</b>
2.1	GAMA . . . . .	11
2.1.1	Input & Tiling catalogues . . . . .	12
2.1.2	Spectroscopic redshifts . . . . .	13
2.1.3	Sizes and inclinations via Sérsic fits . . . . .	14
2.2	GALEX . . . . .	14
2.2.1	GALEX optics, detector and data acquisition . . . . .	15
2.2.2	Pipeline products . . . . .	17
2.2.3	Imaging performance . . . . .	19
<b>3</b>	<b>The GALEX-GAMA data products</b>	<b>23</b>
3.1	Preprocessing . . . . .	24
3.1.1	GALEX-GAMA shallow, medium, and deep coverage . . . . .	25
3.1.2	Primary/secondary region . . . . .	26
3.1.3	UV Background . . . . .	28
3.1.4	Redefinition of measurement error . . . . .	30
3.2	Blind catalogue . . . . .	33
3.3	Simple matching catalogue . . . . .	34
3.4	Advanced matching catalogue . . . . .	35
3.4.1	Description of the technique . . . . .	35
3.4.2	Correlation with catalogue depth and galaxy quantities . . . . .	38
3.5	Curve-of-growth analysis . . . . .	41
3.6	Choosing the best method . . . . .	46
3.6.1	Comparison of all three flux estimates . . . . .	46
3.6.2	Matching completeness . . . . .	51
3.6.3	“Best-mag” catalogue . . . . .	52
3.7	Ancillary Information . . . . .	53
3.7.1	Observation information catalogue (“ObsInfoCat”) . . . . .	53
3.7.2	HEALPIX maps of the GAMA fields . . . . .	53
3.8	UV detection limit . . . . .	54
3.9	Unmatched GALEX sources . . . . .	56

<b>II</b>	<b>Attenuation of starlight by dust in spiral galaxies</b>	<b>59</b>
<b>4</b>	<b>Fundamental considerations</b>	<b>61</b>
<b>5</b>	<b>GAMA input data &amp; constraints</b>	<b>65</b>
5.1	Basic data . . . . .	65
5.1.1	Distances . . . . .	65
5.1.1.1	Redshift and local-flow correction . . . . .	65
5.1.1.2	Redshift completeness . . . . .	66
5.1.2	Optical photometry . . . . .	66
5.1.2.1	r-band photometry . . . . .	66
5.1.2.2	Foreground extinction – A-correction . . . . .	67
5.1.2.3	Rest frame magnitude – k-correction . . . . .	67
5.1.3	NUV photometry . . . . .	68
5.2	Sample selection . . . . .	68
5.2.1	Spiral selection technique . . . . .	69
5.2.2	GAMA spiral galaxy sample . . . . .	69
5.3	Derived physical quantities . . . . .	69
5.3.1	Stellar mass . . . . .	70
5.3.2	Star formation rate . . . . .	71
5.3.3	Structural parameters . . . . .	72
5.3.3.1	Axis ratios and sizes . . . . .	72
5.3.3.2	Bulge-to-total ratio . . . . .	72
5.4	Predicted dust attenuation . . . . .	74
5.4.1	Radiation transfer model . . . . .	75
5.4.2	Disk opacity models . . . . .	77
<b>6</b>	<b>Simulating Data</b>	<b>79</b>
6.1	Simulation of intrinsic properties of population . . . . .	79
6.1.1	r-band magnitude . . . . .	79
6.1.2	Redshift . . . . .	82
6.1.3	Stellar mass $M_s$ . . . . .	83
6.1.3.1	Simulation of g-i colour . . . . .	84
6.1.3.2	Simulation of i-band magnitude . . . . .	84
6.1.4	Star formation rate (SFR) . . . . .	85
6.1.5	Absolute intrinsic NUV magnitude $M_{UV,intr}$ . . . . .	85
6.1.6	Dust content . . . . .	86
6.1.7	Intrinsic morphology . . . . .	86
6.1.7.1	Intrinsic r-band radius . . . . .	87
6.1.7.2	Intrinsic inclination . . . . .	88
6.1.7.3	Intrinsic single Sérsic index $n$ . . . . .	89
6.1.7.4	Intrinsic bulge-to-disk-ratio . . . . .	90
6.1.7.5	Spiral galaxy selection . . . . .	90
6.2	Simulation of apparent (directly observable) properties of the GAMA galaxies . . . . .	91
6.2.1	Apparent morphology . . . . .	91
6.2.1.1	Observed B/D, single Sérsic index $n$ , and size . . . . .	91
6.2.1.2	Observed axis ratio $b/a$ . . . . .	91
6.2.2	Noiseless apparent magnitudes . . . . .	94

6.2.2.1	Dust attenuation . . . . .	94
6.2.2.2	Foreground extinction and k-correction . . . . .	95
6.2.3	Apparent NUV magnitude with noise . . . . .	95
6.2.3.1	UV background flux and exposure time . . . . .	95
6.2.3.2	NUV noise . . . . .	96
6.2.4	Apparent NUV-r colour with noise . . . . .	96
6.2.5	Motivation of log-normal distribution of intrinsic luminosity . . . . .	96
6.3	Simulation of flux-limited samples . . . . .	98
6.3.1	Detection limits . . . . .	98
6.3.1.1	r-band detection limit . . . . .	98
6.3.1.2	NUV detection limit . . . . .	99
6.3.2	Simulated data sets . . . . .	99
6.3.2.1	Data set A . . . . .	100
6.3.2.2	Data set B . . . . .	100
6.3.2.3	Data set C . . . . .	100
6.3.2.4	Data set D . . . . .	100
6.3.2.5	Data set E . . . . .	101
6.3.2.6	Data set F . . . . .	101
6.3.3	Sample biases . . . . .	101
6.3.3.1	Colour bias caused by the r-band . . . . .	101
6.3.3.2	Colour bias caused by NUV noise . . . . .	102
6.3.3.3	Inclination bias . . . . .	102
<b>7</b>	<b>Statistical inference of the attenuation of starlight by dust in spiral galaxies</b>	<b>105</b>
7.1	Method . . . . .	105
7.1.1	Utilisation of ellipticities . . . . .	106
7.1.2	Utilisation of scatter in the intrinsic colours . . . . .	106
7.1.3	Present limitations . . . . .	108
7.2	Inference of inclination from observed $b/a$ . . . . .	108
7.3	Likelihood of the intrinsic colour distribution . . . . .	110
7.3.1	Intrinsic colour distribution . . . . .	111
7.3.2	Maximum likelihood fit of the colour . . . . .	113
7.3.3	Tests on simulated data and uncertainty of $\mu$ and $\sigma$ . . . . .	113
7.4	Deriving the best-fitting parameters of a given dust model . . . . .	114
7.5	Tests on simulated data . . . . .	115
7.5.1	Parameter estimation for simulated data with scatter-free intrinsic colour . . . . .	116
7.5.2	The effect of intrinsic scatter on the parameter estimation . . . . .	120
7.5.3	The effect of UV noise on the parameter estimation . . . . .	123
7.5.4	Effect of equating $b/a$ with $\cos(i)$ . . . . .	128
7.6	Distinguishing between dust models . . . . .	128
7.7	Application to real data . . . . .	132
7.7.1	Best-fitting dust model (parameter) for the GAMA spiral galaxy sample . . . . .	132
7.7.2	Derived intrinsic colours . . . . .	136
7.7.3	Technical discussion of the results . . . . .	138
7.7.3.1	r-band detection limit . . . . .	138

7.7.3.2	Bulge-to-disk proxy . . . . .	138
7.7.3.3	Effects of dust attenuation on size and inclination estimates . . . . .	138
7.7.3.4	NUV noise . . . . .	139
7.7.3.5	Form of parametric distribution in colour . . . . .	139
7.7.4	Scientific discussion of the results . . . . .	140
<b>8</b>	<b>Summary and outlook</b>	<b>143</b>
8.1	Accounting for r-band detection limit and NUV noise . . . . .	144
8.2	Maximum likelihood with prior . . . . .	144
8.3	Probabilistic treatment of spiral selection . . . . .	144
8.4	Iterative approach to the derivation of the dust model and of the attenuation corrected r-band luminosity function . . . . .	145
8.5	Self-consistent calculation of stellar mass . . . . .	145
8.6	Accounting for the effects of dust attenuation on measured size, axis ratio, and Sérsic index . . . . .	146
8.7	Empirical correlations of dust content with physical quantities . . .	146
8.8	Comparing subsamples of galaxies . . . . .	147
8.9	Alternative attenuation models . . . . .	147
8.10	Multi-wavelength approach . . . . .	147
<b>III</b>	<b>Appendices</b>	<b>149</b>
<b>A</b>	<b>r-band Luminosity Function</b>	<b>151</b>
A.1	Functional form of the luminosity function . . . . .	151
A.1.1	Luminosity function of the intrinsic absolute magnitude . . .	151
A.1.2	Distance distribution . . . . .	152
A.1.3	Detection limit . . . . .	153
A.1.4	Dust attenuation . . . . .	155
A.1.5	Restframe magnitude . . . . .	158
A.1.6	Full r-band model . . . . .	159
A.2	Deriving the best-fitting luminosity function parameters . . . . .	159
A.2.1	Calculating the amplitude $\phi^*$ . . . . .	159
A.2.2	Deriving best-fitting $\alpha$ and $M^*$ using maximum likelihood method . . . . .	160
A.3	Tests on simulated data . . . . .	162
A.3.1	Best-fitting parameters for dust-free simulation . . . . .	162
A.3.2	Normalization of the probability distribution containing dust	162
A.3.3	Effect of starting point of the fit . . . . .	164
A.3.4	Effect of sample size . . . . .	165
A.4	fit parameter for GAMA galaxies . . . . .	166
<b>B</b>	<b>Theoretical framework for the treatment of NUV noise</b>	<b>169</b>
<b>C</b>	<b>Accounting for r-band detection limit</b>	<b>173</b>

<b>D GALEX-GAMA DMU product details</b>	<b>175</b>
D.1 Blind catalogue . . . . .	175
D.2 Simple match catalogue . . . . .	180
D.3 Advanced match catalogue . . . . .	181
D.4 Curve-of-Growth catalogue . . . . .	182
D.5 Best-mag catalogue . . . . .	183
D.6 ObsInfo catalogue . . . . .	187



# Chapter 1

## Introduction

The identification, characterisation, and quantification of the populations of young stars in galaxies constitute key observational constraints on the evolution of the visible Universe. Such populations probe the contemporaneous conversion of the interstellar medium (ISM) into stars. This, in turn, enables empirical investigations of the dependence of star formation on the physical and chemical properties of the ISM on various linear scales, as well as on macroscopic properties of the host galaxy. In principle, detailed studies of local Universe galaxies can be used to constrain physical theories of the process of star formation which can then be applied to distant galaxies to ultimately provide a complete physical description the conversion of baryons into stars over cosmic time.

Such studies must embrace both processes on relatively short time- and length-scales, which govern the collapse into stars of self-gravitating structures in the ISM, as well as processes operating on longer length- and time-scales which govern the cooling and energisation of the gas in the ISM, and which regulate the rate at which the self-gravitating structures form. Common physical quantities, for example metallicity, radiation field intensity, magnetic field, dust content and cosmic ray content, play a role in the physics of both these regimes, so it is of particular interest to investigate the empirical relations between these quantities on the one hand, and the star formation rates (SFR) of galaxies on the other. Indeed, one of the main motivations of large statistical surveys of the multiwavelength emission of galaxies from radio to X-rays is to establish such relations as a function of global properties of galaxies such as mass, morphology, and environment. Multi-wavelength statistical investigations of this nature are also essential to empirically constrain processes operating on Kpc to Mpc scales which govern the flow of gas into and out of galaxies, regulating the amount of gas in the ISM available to form stars. Recent empirically-orientated reviews on small and large-scale processes linked to star formation in the Milky Way and in external galaxies are by Molinari et al. (2014) and Kennicutt & Evans (2012).

In seeking empirical connections between the process of star formation and the global properties of galaxies it is essential to use an observational tracer that isolates and quantifies the SFR of a galaxy. Several techniques have been used to measure the SFR of galaxies (see Kennicutt & Evans, 2012). The most direct and robust method is to measure the bolometric output of young stars in the non-ionising ultraviolet continuum (at wavelengths between 912 and ca. 3000Å). The advantage of this technique is apparent from Fig. 1.1, which shows the predicted

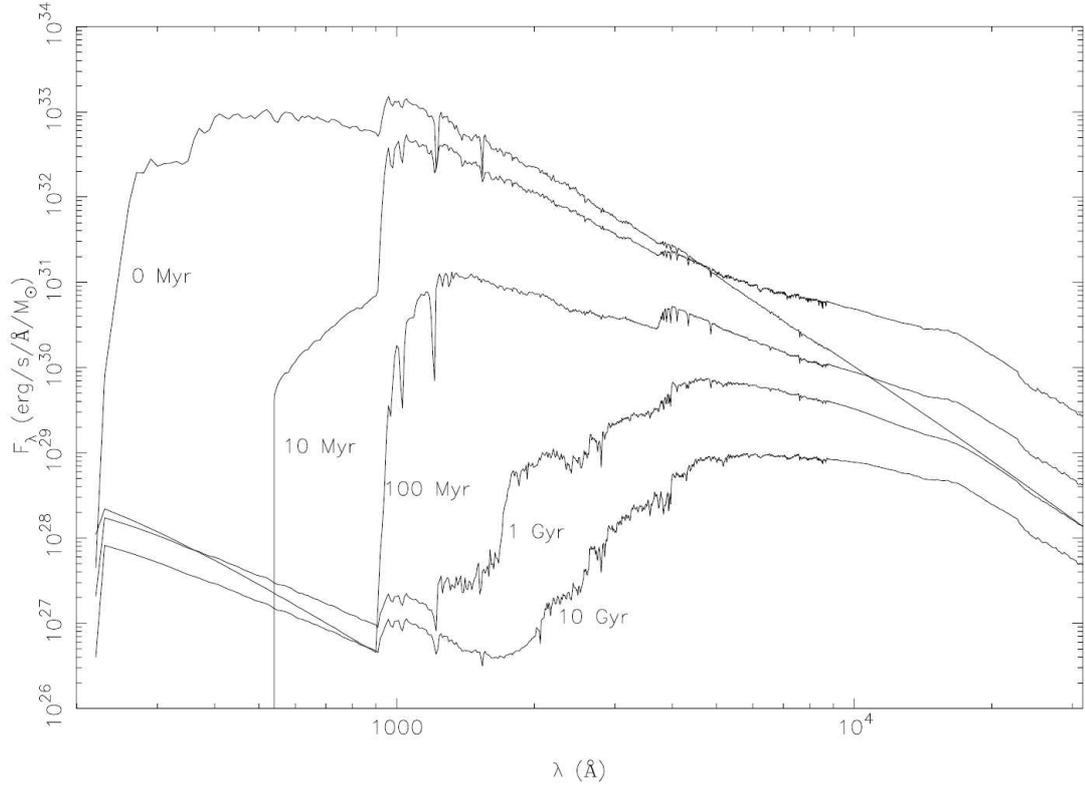


Figure 1.1: The spectral evolution of the starlight

The predicted dependence on time and wavelength of the UV and optical emission from a single burst of star formation, (reproduced from Fig.1 of Fioç & Rocca-Volmerange, 1997). The assumed initial mass function is that of Rana & Basu (1992), derived from the multiplicity of stars in the solar neighbourhood.

dependence on time and wavelength of the UV and optical emission from a single burst of star formation for an Initial Mass Function (IMF) of the stars constrained by observations of the youngest stars in the solar neighbourhood. One sees a rapid decrease in luminosity with time, such that at early times, within a few hundred Myr of the burst, most of the energy is radiated in the non-ionising UV (at wavelengths between 912 and ca. 3000Å). By contrast, at later times, virtually all light is emitted in the optical (longwards of 3000Å). This behaviour reflects the predicted relation between colour and luminosity of stars on the main sequence, combined with the expectation that all stars with masses greater than ca.  $6M_{\odot}$  (which are the stars that dominate the UV output of the burst), disappear through supernova explosions within a few hundred Myr. Even when considering systems with a constant star formation rate (SFR), there is a corresponding strong sensitivity of the luminosity-weighted mean age of the radiating stars to the observing wavelength, such that, even at the shortest wavelengths accessible to observation from the ground, only stellar populations with ages greater than 1Gyr can be investigated with broad-band photometry.

For targets sufficiently bright to be measured spectroscopically this difficulty can be readily circumvented by measuring optical spectral lines emitted in the HII regions around the massive stars - most notably the optical recombination lines of Hydrogen. Indeed, imaging the  $H\alpha$  recombination line at 6563Å is undoubtedly



Figure 1.2: M51 in the optical (left) and in the UV (right)

LH panel: A colour composite of images of optical emission from M51 (overlaid with  $H\alpha$  emission in red). RH panel: A colour composite of the 1600 and 2200Å emission from M51 measured by GALEX. Picture credit: NASA/ESA/S. Beckwith & Hubble Heritage Team (STScI/AURA); NASA/JPL-Caltech

the primary technique for resolved studies of star formation in very local galaxies. This is due to the arcsec angular resolution of optical telescopes combined with the fact that, as can also be seen from Fig. 1.2, stellar ionising radiation probes stars with lifetimes of ca. 10Myr, an order of magnitude shorter than the timescales probed by spaceborne measurements of the non-ionising UV. Despite this difference in stellar ages probed by the two methods, in practice measurements of the UV continuum shortwards of 3000Å probe almost the same emission structures in galaxies as  $H\alpha$ . This can be seen through the comparison of the images of M51 taken in the spaceborne UV and the optical (Fig. 1.2).

For the study of large statistical samples of faint distant galaxies, however, measuring SFR in the UV continuum has decisive advantages over  $H\alpha$  measurements<sup>1</sup>. Firstly, for statistical surveys of large numbers of galaxies, measurements

---

<sup>1</sup>Other SF indicators are also in common use in extragalactic surveys, notably measurements in the FIR continuum of starlight which has been absorbed by dust grains in the ISM (e.g. Eales et al., 2010) and in the radio continuum (e.g. Schinnerer et al., 2007). However with the exception of very optically thick starburst galaxies, surveys of local Universe galaxies made with FIR facilities are much less sensitive to star-forming galaxies than surveys made either in  $H\alpha$  or in the spaceborne UV continuum. Moreover substantial fractions of the FIR emission is powered by optical photons from old stellar populations, complicating the elucidation of SFR. Similar considerations apply to the measurement and interpretation of radio continuum emission, present radio surveys being even less sensitive to star-forming galaxies in the local Universe than FIR surveys. Moreover, whereas free free radio emission directly traces the ionising UV luminosity of

in the broad-band UV continuum emission reach deeper completeness limits on SFR than spectroscopic surveys, despite the need (for galaxies within redshifts of  $\sim 0.5$ ) to observe with spaceborne facilities to attain the coverage shortwards of  $3000\text{\AA}$  needed to ensure that the detected signal is dominated by recently formed massive stars. This is largely due to the advent of the NASA's Galaxy Evolution Explorer (GALEX; Martin et al., 2005; Bianchi et al., 2014). Secondly, an essential characteristic of statistical surveys of local Universe galaxies (within redshifts of  $\sim 0.1$ ) is the measurement of morphology and size using optical imaging. Spectroscopic measurements, although essential to measure redshifts then have the disadvantage that the width of the slit or fibre diameter covers only a fraction of the angular extent of the targets, so integrated fluxes are not measured. This is a particularly critical issue if, as may well be the case for rotationally supported systems, galaxies grow from the inside out, such that the UV and  $H\alpha$  emissivity is more extended than the optical emissivity from older stellar populations. Finally, measurements in the non-ionising UV continuum, in particular in the GALEX NUV band centred at  $2200\text{\AA}$ , are much less sensitive to the form of the IMF than  $H\alpha$ , which is an advantage if the main inference to be made is the rate at which the ISM is condensing into stars. This is because we do not yet have clear empirical constraints on possible variations of the IMF with global properties of galaxies.

All these considerations have motivated the GALEX-GAMA survey, which is a medium deep blind continuum survey in the FUV ( $1650\text{\AA}$ ) and NUV ( $2200\text{\AA}$ ) bands of almost the entire 290 sq. deg. footprint of the Galaxy And Mass Assembly spectroscopic redshift survey (Driver et al., 2011) of local Universe ( $z < 0.4$ ) galaxies made with the Anglo Australian Observatory (AAO) 3.9m telescope. A visual impression of the 3d coverage in solid angle and redshift of GAMA is shown in Fig. 1.3, comparing GAMA with the SDSS redshift survey and with deep pencil beam surveys. A principle characteristic of GAMA is its present and planned multiwavelength coverage from the radio to X-rays (Driver et al., 2009). Most particularly from the point of view of studying star formation in galaxies, the GAMA fields are prime targets for measurements in the HI 21cm line of neutral Hydrogen with the Australia Square Kilometre Array Pathfinder telescope. In addition, the entire GAMA footprint has been surveyed with Herschel at FIR/submm wavelengths (Eales et al., 2010).

The first part of this dissertation describes the analysis of the GALEX-GAMA data, which was taken in the latter part of the lifetime of the observatory, mainly in the 2008-2013 time frame, and embraces around 400 hours of observing time to cover the GAMA fields to a depth of 1500s. The fields to be observed in the ultra-deep survey mode of ASKAP were covered to 3000s depth. A particular aim of this analysis was to provide catalogues of GAMA galaxies extracted with a homogeneous processing and accurate integrated photometry with quantitative understanding of uncertainties. This is particularly important for the analysis of complete statistical populations of dwarf galaxies and quiescent spiral galaxies which are reached at the limit of the depth of the GALEX-GAMA survey at redshifts out to ca. 0.1.

---

galaxies, most of the emission shortwards of 10GHz is emitted by relativistic electrons radiating synchrotron emission in magnetic fields. The empirical and physical connection of both the relativistic electrons and magnetic fields on galaxy wide scales to the star formation process are completely open issues.

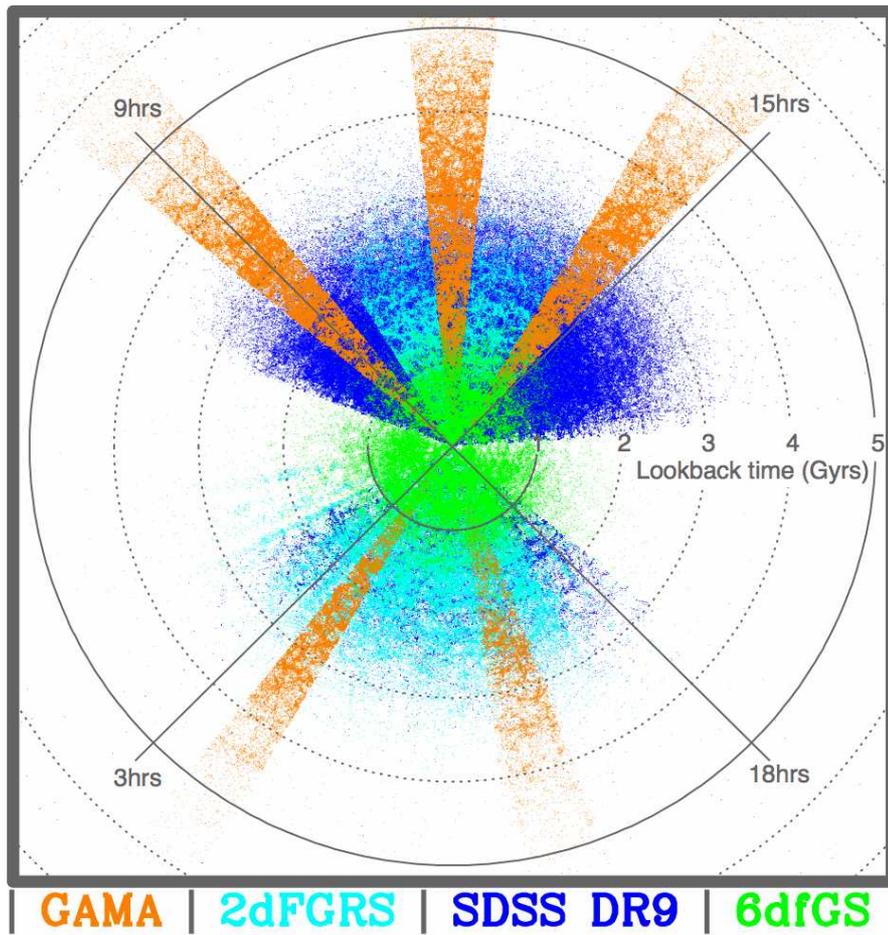


Figure 1.3: Spatial distribution of GAMA galaxies

RA -  $z$  distribution of redshifts collapsed in DEC for GAMA (orange), SDSS (blue), 2dFGRS (azure), and 6dFGS (green). Taken from the GAMA website <http://www.gama-survey.org>. Picture credit: A. Robotham/J. Liske

One principle obstacle in the scientific exploitation of the GALEX-GAMA data (while at the same time being of intrinsic scientific interest in its own right) is however the attenuation of the UV starlight due to the presence of interstellar dust in the ISM of star-forming galaxies. Attenuation of the UV light caused by dust is physically unavoidable since massive stars form gas in dense clouds, and the dust present in these clouds plays a major role in the necessary cooling processes prior to the star formation.

However, it is not only the dust in the birth clouds that attenuates the UV light. The attenuation by dust distributed through the whole ISM also impacts (and probably dominates) the attenuation of UV light, despite the close physical association of the youngest (and therefore very luminous) most massive stars with the opaque dust clouds hosting them. The presence of dust local to star formation regions and on large scales in regions between the spiral arms can clearly be seen in the HST image of M51 shown in Fig. 1.2.

Statistical studies of the luminosity and colour distribution of large samples of galaxies have shown the severe effect of dust attenuation in the optical bands (Driver et al., 2007; Shao et al., 2007; Driver et al., 2008). Dust extinction laws

(e.g. Calzetti, 2001) predict an even more severe effect of dust attenuation at UV wavelengths. Dust attenuation therefore effectively hinders us from making direct quantitative inferences on SFR from UV data. For instance, the determination of Star-Formation Histories (SFR) is biased due to the dimming of the emission at the considered wavelength, or the determination of SFH is biased due to the dust-reddening/age degeneracy.

Apart from the effect of attenuation, dust grains play a major role the physical processes inside galaxies. Some of these processes have a direct impact on the star formation process. For example grains mediate the heating of the ISM through photoelectric heating, are the main agents radiating energy out of very optically thick collapsing stellar cores, and are thought to be the main sites for formation of hydrogen molecules. In addition, dust grains may play a fundamental role in controlling the feedback of energy and matter between galaxies and the circumgalactic medium, including the mediation of AGN feedback (Fabian, 1994).

There are still many open questions about dust in galaxies and dust attenuation, and how this is related to the galaxy populations. The fundamental question of what are the sources and sinks of grains in the ISM of galaxies is still open. A host of mundane issues abound, such as to what extent the sophisticated models for grain optical properties developed for the Milky Way (see e.g. Weingartner & Draine, 2001) are valid for external galaxies.

For all these reasons it is crucial to have a quantitative understanding of the amounts and properties of grains present in the disks of different populations of galaxies (absorption, scattering as well as emission). There are several methods which are used to this end. The most direct technique is to directly measure the starlight absorbed by the dust in the FIR. However, in practice the depth of FIR surveys is sufficient only to reach the more massive and nearby galaxies in the GAMA spectroscopic sample. Another technique is to use measurements of the nebular line emission spectrum of the galaxies obtained as a byproduct of the spectroscopic survey to deduce the dust attenuation. This technique has for example been very successfully used for SDSS galaxies (e.g. Brinchmann et al., 2004). However, here again the disadvantages of spectroscopic observations in terms of sensitivity and poor coverage of the integrated emission from the targets come into play.

These considerations motivate the second part of this work, in which a method is developed to extract quantitative information about the dust content of disk galaxies and the attenuation of UV light by the dust, and the relation of these quantities to stellar mass. The method uses statistical measurements of the NUV emission of GAMA galaxies measured in the GALEX-GAMA survey, in combination with the ellipticities of the galaxies obtained from optical imaging available in GAMA, to statistically constrain the orientations of the disk and thereby break the colour-reddening degeneracy for a statistically well defined sample of spiral galaxies. A special focus is made on the qualitative investigation of biases caused by detection limits and measurement noise in this analysis. To this end we set up detailed simulations of galaxy properties, especially the optical/UV emission, colour, stellar masses, and morphology. We simulate how dust attenuation affects and biases the measurement of the photometry and thereby causes the observed colour distribution of the galaxy sample to differ from the intrinsic. Furthermore, we reveal the existence of severe biases in the colour distribution of flux limited sam-

ples caused by noise, biased inclination measurements, and the colour-luminosity relation combined with the detection limit. Finally, we provide the groundwork for an extension of the analysis allowing the investigation of the dependence of dust content and attenuation on other derived physical properties such as SFR.

Throughout this work, we assume a standard  $\Lambda$ CDM cosmology with parameter values  $\Omega_M = 0.3$ ,  $\Omega_\lambda = 0.7$ ,  $h \cdot H_0 = 70 \text{ kms}^{-1} \text{ Mpc}^{-1}$  ( $h = 0.7$ ) (Spergel et al., 2003).



# Part I

## The GALEX-GAMA survey



# Chapter 2

## GALEX-GAMA input data

### 2.1 GAMA

The Galaxy And Mass Assembly survey (GAMA<sup>1</sup>; Driver et al., 2009, 2011; Hopkins et al., 2013) is a joint European-Australian project based around a spectroscopic survey using the 2dF instrument and the AAOmega spectrograph on the Anglo-Australian Telescope at the Australian Astronomical Observatory. Complementary imaging of the GAMA regions has been, or is being obtained by a number of independent survey programs including GALEX, VST KiDS, VISTA VIKING, WISE, Herschel-ATLAS, GMRT and ASKAP providing UV to radio coverage (see Table 2.1). The survey aims to test the  $\Lambda$ CDM model of structure formation as well to study galaxy evolution.

GAMA aims to obtain spectra of all galaxies with SDSS Petrosian r-band magnitude (foreground attenuation corrected)  $r_{Petro,0} < 19.8\text{mag}$ , in a 290 sq.deg region of the sky divided into five fields. Targets are selected using the SDSS photometric survey (York et al., 2000; Stoughton et al., 2002, , see Sect. 2.1.1).

The GAMA survey consists of five fields: three equatorial fields centred on  $9^h$ (G09),  $12^h$ (G12),  $14.5^h$ (G15), as well as two southern fields centred on  $2.15^h$ (G02) and  $23^h$ (G23). Due to technicalities of its completion, the survey is divided into GAMA I (Driver et al., 2011, see Table 2.2) consisting of the three equatorial fields (G09, G12 and G15), and GAMA II (Liske et.al., in preparation) encompassing the full survey range (see Tables 2.2 and 2.3). The spectroscopic survey is complete for G09, G12, G15, and G02 to the survey depth of  $r_{petro,0} < 19.8\text{mag}$ , which is more than two magnitudes deeper than the spectroscopic depth of SDSS. The G23 field is currently being observed. Already today, GAMA provides 286 705 galaxy spectra of 237 822 unique targets (see Baldry et al., 2014).

As illustrated by Fig. 2.1, the unique combination of depth, area, and high redshift completeness places the GAMA survey between the deep pencil beam surveys such as VIMOS VLT Deep Survey deep sample (VVDS-deep; Le Fèvre et al., 2005) and the shallow wide-field surveys such as SDSS Main Galaxy Sample (SDSS-main; Strauss et al., 2002).

---

<sup>1</sup>The GAMA website is <http://www.gama-survey.org/>.

Survey	Facility	Wavelength / band	Type of data used	Reference
XXL	XMM-Newton <sup>a</sup>	0.5–2 keV	Proprietary data	Pierre et al. (2011)
GALEX-GAMA	GALEX <sup>b</sup>	0.15, 0.22 $\mu\text{m}$	Public (MIS <sup>c</sup> ) and own data	
SDSS	Sloan telescope	$u, g, r, i, z$	Public data	DR7: Abazajian et al. (2009)
KiDS <sup>d</sup>	VST <sup>e</sup>	$u, g, r, i$	Proprietary data	de Jong et al. (2013)
CFHTLenS <sup>f</sup>	CFHT <sup>g</sup>	$u, g, r, i, z$	Proprietary data	Heymans et al. (2012)
UKIDSS LAS <sup>h</sup>	UKIRT <sup>i</sup>	$Y, J, H, K$	Public data	Lawrence et al. (2007)
VIKING <sup>j</sup>	VISTA <sup>k</sup>	$Z, Y, J, H, K_s$	Proprietary data	
WISE All-Sky DR	WISE <sup>l</sup>	3.4, 4.6, 12, 22 $\mu\text{m}$	Public data	Wright et al. (2010)
H-ATLAS <sup>m</sup>	Herschel	100, 160, 250, 350, 500 $\mu\text{m}$	Proprietary data	Eales et al. (2010)
DINGO <sup>n</sup>	ASKAP <sup>o</sup>	21 cm	In planning	see Duffy et al. (2012)
GMRT-GAMA	GMRT <sup>p</sup>	92 cm	Own data	Mauch et al. (in prep.)

<sup>a</sup>X-ray Multi-Mirror Mission; <sup>b</sup>Galaxy Evolution Explorer; <sup>c</sup>Medium Imaging Survey; <sup>d</sup>Kilo Degree Survey; <sup>e</sup>VLT Survey Telescope; <sup>f</sup>CFHT Lensing Survey; <sup>g</sup>Canada France Hawaii Telescope; <sup>h</sup>UKIRT Infrared Deep Sky Survey – Large Area Survey; <sup>i</sup>United Kingdom Infrared Telescope; <sup>j</sup>VISTA Kilo-Degree Infrared Galaxy Survey; <sup>k</sup>Visible and Infrared Survey Telescope for Astronomy; <sup>l</sup>Wide-Field Infrared Survey Explorer; <sup>m</sup>Herschel Astrophysical Terahertz Large Area Survey; <sup>n</sup>Deep Investigation of Neutral Gas Origins; <sup>o</sup>Australian Square Kilometre Array Pathfinder; <sup>p</sup>Giant Metrewave Radio Telescope;

Table 2.1: Independent surveys from which GAMA is using imaging and/or photometric data, sorted by wavelength. *Taken from Liske et al. (2014, in preparation).*

## 2.1.1 Input & Tiling catalogues

There are two main GAMA catalogues, which represent the interface for all complementary GAMA data and catalogues. The Input catalogue InputCatAv06 (Baldry et al., 2010) is a list of all objects (both stars and galaxies) in the SDSS DR7 catalogue which are at least as bright (after a consideration of Milky Way dust attenuation, see Sects.5.1.2.1 and 5.1.2.2) as  $r_{Petro,0} < 20\text{mag}$  and lie within the GAMA footprint for the equatorial regions.

All galaxies with  $r_{Petro,0} < 19.8\text{mag}$  in this catalogue are listed again in the TilingCatv41<sup>2</sup>. While the InputCatAv06 represents a list of all sources in the GAMA equatorial fields, the TilingCatv41 contains all GAMA galaxies - the so-called GAMA sources. It lists the main photometric information extracted from

<sup>2</sup>This catalogue is thus named as it was used to plan the observations at the AAO. The tiling algorithm is presented in Robotham et al. (2010).

FIELD id	min RA	max RA	min DEC	max DEC	r-band limit
G09	129.0	141.0	-1.0	+3.0	19.4
G12	174.0	186.0	-2.0	+2.0	19.8
G15	211.5	223.5	-2.0	+2.0	19.4

Table 2.2: GAMA I regions

FIELD id	min RA	max RA	min DEC	max DEC	r-band limit
G02	30.2	38.8	-10.25	-3.72	19.8
G09	129.0	141.0	-3.0	3.0	19.8
G12	174.0	186.0	-3.0	3.0	19.8
G15	211.5	223.5	-3.0	3.0	19.8
G23	338.1	351.9	-35.0	-30.0	19.8

Table 2.3: GAMA II regions

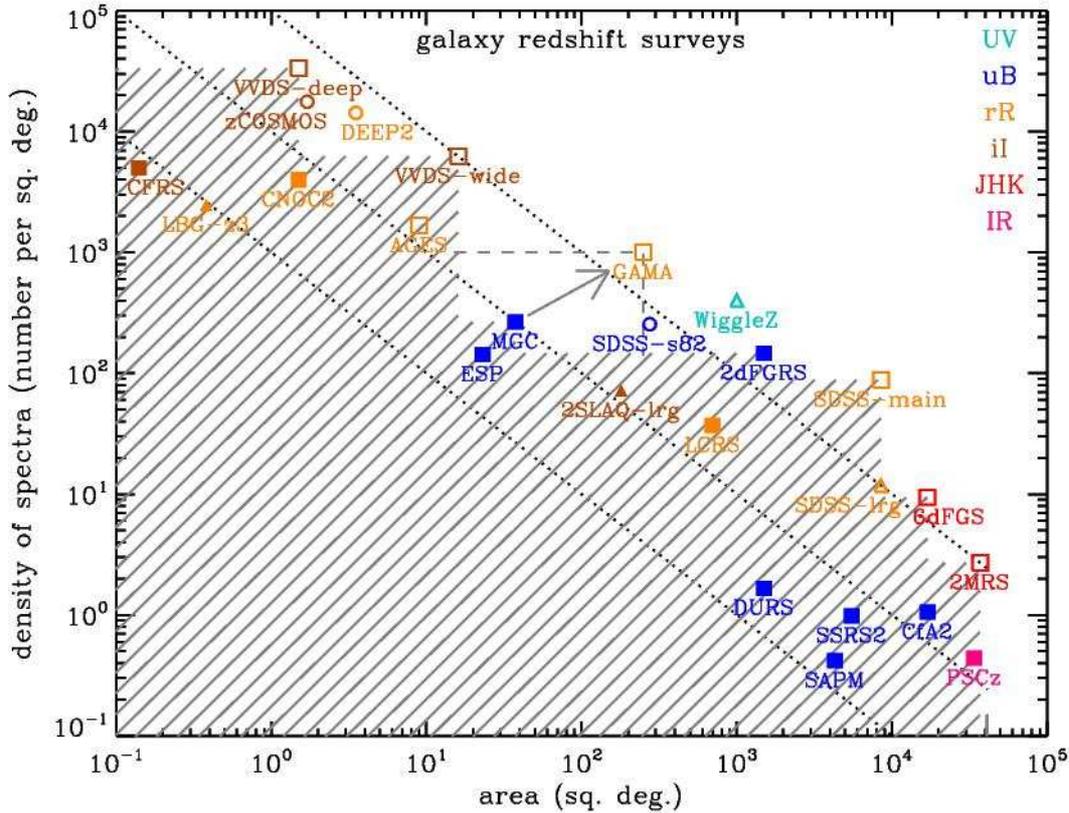


Figure 2.1: Field galaxy survey overview

Comparison between field galaxy surveys with spectroscopic redshifts: squares represent predominantly magnitude-limited surveys; circles represent surveys involving colour cuts for photometric redshift selection; triangles represent highly targeted surveys. The colours represent different principal wavelength selections as in the legend. Filled symbols represent completed surveys. *Taken from Baldry et al. (2010).*

SDSS along with the redshifts measured by GAMA (added a posteriori).

There exist (preliminary) versions of these catalogues for the non-equatorial GAMA fields. In this work, we focus on the description of the equatorial fields. The same methods were applied to generate the catalogues of the non-equatorial GAMA fields.

## 2.1.2 Spectroscopic redshifts

Deriving the redshift from a given galaxy spectrum is the key task in generating a spectroscopic redshift survey. Spectra of stars are very different from spectra of galaxies, and galaxy spectra again can be dominated by absorption lines or emission lines.

In December 2013, the `autoz` program was completed (Baldry et al., 2014), which is capable of fully automatically deriving redshifts from spectra. It uses a cross-correlation method for both absorption line and emission line spectra. 47 different spectroscopic templates, ranging from spectral templates for different young and old stars to different kinds of galaxy spectra, were combined to produce redshift measurements with a median error of 33km/s.

`autoz` was trained using GAMA spectroscopic redshifts, zCOSMOS spectro-

scopic redshifts and COSMOS photometric redshifts.

The overall spectroscopic completeness of GAMA is  $> 98\%$  to its limiting depth. GAMA is also spectroscopically complete at  $> 95\%$  for galaxies with up to 5 neighbours within  $40''$  (Driver et al., 2011; Robotham & Driver, 2011).

The median redshift of the GAMA galaxies is 0.2. 50% of the GAMA galaxy redshifts are in the range of 0.13 - 0.27, and 99% of the GAMA galaxies have a redshift of  $z < 0.5$  (Baldry et al., 2010).

### 2.1.3 Sizes and inclinations via Sérsic fits

Structural information of every galaxy is crucial for understanding galaxy evolution. Ellipticity and size measurements in particular are critical inputs for extracting UV photometry of galaxies from GALEX data.

Kelvin et al. (2012) have performed single component light profile fits on SDSS and UKIDSS-LAS imaging data for all GAMA galaxies using single Sérsic profiles (Sérsic, 1963). They employ a masking technique to disentangle close or even overlapping objects.

In the catalogue SersicCatv09, they provide (amongst other information) the Sérsic index, the ellipticity (i.e. axis ratio), orientation angle and effective radius along the semi-major axis (i.e. half-light radius)<sup>3</sup>. These quantities are used here as input for the matching procedures to define optical apertures.

## 2.2 GALEX

The Galaxy Evolution Explorer (GALEX; Martin et al., 2005; Morrissey et al., 2005, 2007, GALEX observer’s guide<sup>4</sup>, GALEX Technical Documentation<sup>5</sup>) was launched on April 28, 2003. Its mission was to survey the sky in two photometric bands in the non-ionizing UV. The bands are NUV (1350-1750 Å) and FUV (1750-2750 Å). The main scientific goal of GALEX was to study the present day star formation in galaxies in the local Universe, and, in combination with other data, to understand the past evolution of galaxies. It was shut down on June 28th, 2013 after 10 years of service<sup>6</sup>.

For the GALEX-GAMA survey, we included all GALEX NUV and FUV data overlapping the GAMA fields, ranging from data obtained from the All-Sky survey (AIS; average exposure time 100s, limiting NUV magnitude  $m_{AB}$ <sup>7</sup>  $\sim 20.5$ ), the Medium Imaging Survey (MIS; average exposure time  $\sim 1500$ s, limiting NUV magnitude  $m_{AB} \sim 23$ ), the Deep Imaging Survey (DIS; average exposure time 30000s, limiting NUV magnitude  $m_{AB} \sim 25$ ), the dedicated guest investigator program “GALEX-GAMA:UV/Optical/Near-IR/Far-IR/Radio Observations of  $\sim 100$ k Galaxies” (GI5-048; PIs: R. Tuffs, C. Popescu), and in the last year of

<sup>3</sup>Due to the shape of the light profile, galaxies with high Sérsic index (e.g. caused by the presence of a bright bulge) will have smaller effective radii than those with low Sérsic index (e.g. pure disk galaxies), even if their disks have the same physical size.

<sup>4</sup>[http://galexgi.gsfc.nasa.gov/docs/galex/Documents/ERO\\_data\\_description\\_2.htm](http://galexgi.gsfc.nasa.gov/docs/galex/Documents/ERO_data_description_2.htm)

<sup>5</sup><http://www.galex.caltech.edu/researcher/techdocs.html>

<sup>6</sup>see press release <http://www.galex.caltech.edu/newsroom/glx2013-03r.html>

<sup>7</sup>The GALEX products use AB magnitudes. The zero point of the AB apparent magnitude scale is defined to be 3631 Jy. The absolute magnitude (i.e. at standard distance of 10 pc) is thus  $M_{AB} = -2.5 \log S_\nu + 34.1$ , where  $S_\nu$  is the luminosity measured in  $\text{W Hz}^{-1}$ .

GALEX operations, as part of the “GAMA/H-Atlas/Dingo Consortium California Institute of Technology/GALEX Sky Completion Project”.

We list the coverage of the individual GAMA fields in Table 2.4 (see following sections for details).

### 2.2.1 GALEX optics, detector and data acquisition

The GALEX telescope uses a 50-centimeter diameter (19.7-inch) mirror with a circular field of view of 1.2 degree diameter. The NUV and FUV bands are obtained simultaneously using a dichroic beam splitter with a dielectric multilayer inside, which reflects the FUV and transmits the NUV (for more details see Martin et al., 2005). The light path is shown in the cross section of the GALEX satellite shown in Fig. 2.2.

GALEX can only observe when it is in the Earth’s shadow, or eclipse, because on the day side of the orbit atmospherically-scattered sunlight and airglow would swamp and might damage the detectors (especially the NUV detector). Even the small amount of residual atmosphere at the 700-km GALEX orbital altitude scatters significant flux into the telescope (see GALEX Technical Documentation<sup>8</sup> for more details). One GALEX orbit around the Earth takes  $\sim 100$  minutes, with about 1/3 of the orbit in the Earth’s shadow. The maximum observation time of an eclipse is slightly less. For each pointing (visit), the detector collects photons in a circular area (see below). Usually, only one visit is done during each eclipse. However, multiple visits can be coadded, generating exposure times longer than an eclipse. The coadds are then processed to produce the so-called tile-level data – data within the FOV of a (set of) pointing(s).

The detector itself is a photon counting device with 65 mm<sup>2</sup> active area and crossed delay-line anodes for photon event position readout. In order to prevent an over-exposure of the detector active area by bright objects, the spacecraft performs a dither pattern of  $\sim 1'$  diameter (see Fig. 2.3), which will cause a point source to have a doughnut-like shape in detector coordinates. The pipeline then transforms the detector coordinates to celestial coordinates (see Fig. 2.4), restoring the point-like shape of the source on the final data maps<sup>9</sup>. Note that the orientation of the spacecraft is not aligned with the RA-DEC coordinates. Therefore, the maps in detector coordinates are arbitrarily rotated compared with the final maps.

During the generation of the final maps, the data is converted from counts  $c$  to flux<sup>10</sup>. In a first step, the counts per second  $cps$  are calculated by dividing the counts per pixel by the effective exposure time  $t_e$  of the according pixel.

$$cps = c/t_e \tag{2.1}$$

Using the zero-point magnitude  $Z_{NUV}$  and  $Z_{FUV}$  in the NUV and FUV respec-

<sup>8</sup>or also GALEX Observer’s Guide <http://www.galex.caltech.edu/researcher/faq.html>

<sup>9</sup>A doughnut-like shape can still appear on the final data maps due to artefacts, and a blind pixel will manifest itself as a dark doughnut on the exposure time maps. Note that a blind pixel of the detector will not result in a blind pixel on the final map, as the dither pattern will propagate the coverage of a single detector pixel over several final map pixels. Therefore, each final map pixel has coverage from several detector pixels.

<sup>10</sup>Described in detail at [http://galexgi.gsfc.nasa.gov/docs/galex/FAQ/counts\\_background.html](http://galexgi.gsfc.nasa.gov/docs/galex/FAQ/counts_background.html)

Field ID	Field Area [deg <sup>2</sup> ]	GALEX depth	GALEX band	Overlap Area [deg <sup>2</sup> ]	Cover Percent [%]	Unflagged Area [deg <sup>2</sup> ]	Unflagged Area [%]
G02	55.48	GSC	FUV	54.21	97%	53.59	96%
G02	55.48	GSC	NUV	55.52	100%	54.93	99%
G09	60.00	GSC	FUV	48.27	80%	47.59	79%
G09	60.00	GSC	NUV	54.04	90%	53.21	88%
G12	60.00	GSC	FUV	44.11	73%	43.75	72%
G12	60.00	GSC	NUV	54.46	90%	54.15	90%
G15	60.00	GSC	FUV	49.78	82%	49.58	82%
G15	60.00	GSC	NUV	50.42	84%	50.16	83%
G23	42.15	GSC	FUV	30.29	71%	30.18	71%
G23	42.15	GSC	NUV	36.66	86%	36.18	85%
G02	55.48	GMC	FUV	29.76	53%	29.54	53%
G02	55.48	GMC	NUV	54.70	98%	54.05	97%
G09	60.00	GMC	FUV	50.01	83%	49.32	82%
G09	60.00	GMC	NUV	52.01	86%	51.31	85%
G12	60.00	GMC	FUV	46.30	77%	45.98	76%
G12	60.00	GMC	NUV	49.38	82%	48.87	81%
G15	60.00	GMC	FUV	56.27	93%	55.97	93%
G15	60.00	GMC	NUV	56.86	94%	56.50	94%
G23	42.15	GMC	FUV	8.83	20%	8.57	20%
G23	42.15	GMC	NUV	42.03	99%	41.40	98%
G02	55.48	GDC	FUV	9.41	16%	9.36	16%
G02	55.48	GDC	NUV	19.02	34%	18.91	34%
G09	60.00	GDC	FUV	0.03	0%	0.03	0%
G09	60.00	GDC	NUV	0.03	0%	0.03	0%
G12	60.00	GDC	FUV	0.96	1%	0.95	1%
G12	60.00	GDC	NUV	0.96	1%	0.95	1%
G15	60.00	GDC	FUV	0.00	0%	0.00	0%
G15	60.00	GDC	NUV	0.00	0%	0.00	0%
G23	42.15	GDC	FUV	0.00	0%	0.00	0%
G23	42.15	GDC	NUV	0.00	0%	0.00	0%

Table 2.4: GAMA II coverage by GALEX

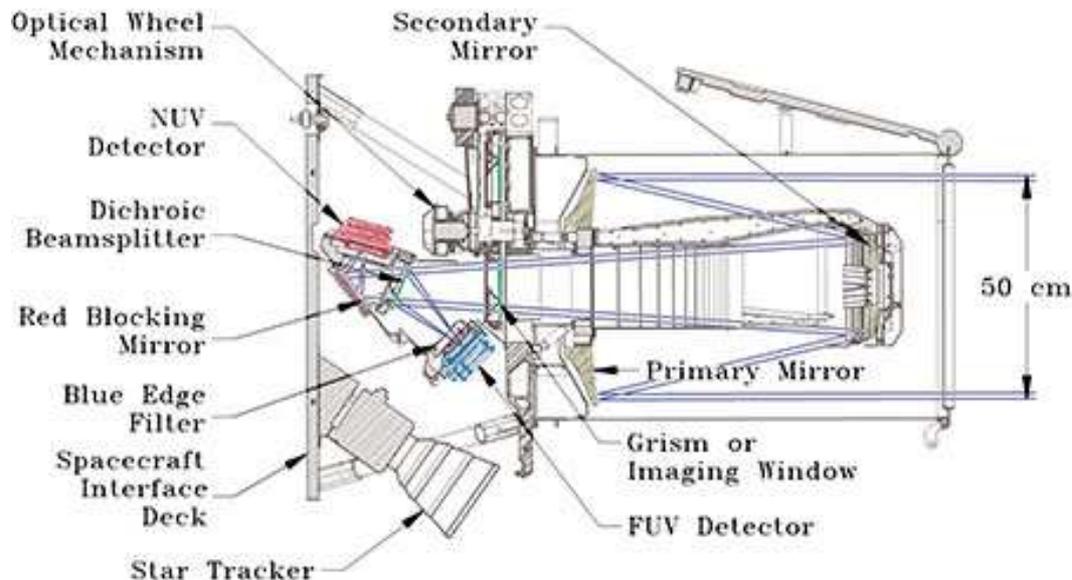


Figure 2.2: The GALEX satellite

A cross section of the GALEX instrument showing the light path (double blue lines). Taken from Morrissey et al. (2005).

tively, the conversion from *cps* to AB magnitude is then<sup>11</sup>

$$m_{NUV} = -2.5 \cdot \log_{10}(cps) + Z_{NUV} \quad \text{with} \quad Z_{NUV} = 20.08 \quad (2.2)$$

$$m_{FUV} = -2.5 \cdot \log_{10}(cps) + Z_{FUV} \quad \text{with} \quad Z_{FUV} = 18.82 \quad (2.3)$$

and to flux  $f$  (units  $\mu\text{Jy}$ ) is

$$f_{NUV} = cps_{NUV} \cdot N_{CUJ} \quad (2.4)$$

$$f_{FUV} = cps_{FUV} \cdot F_{CUJ} \quad (2.5)$$

with

$$N_{CUJ} = 10^{(Z_{NUV}-23.9)/-2.5} = 33.72869 \quad (2.6)$$

$$F_{CUJ} = 10^{(Z_{FUV}-23.9)/-2.5} = 107.6470 \quad (2.7)$$

The FUV detector stopped working in the latter part of the mission when most of the GALEX-GAMA data was taken. This led to areas with NUV coverage only. Nevertheless, there is still significant coverage in the FUV (see Table 2.4).

## 2.2.2 Pipeline products

The basic input data from GALEX used to create the GALEX-GAMA catalogues are the standard output data from the GALEX pipeline v7 (see GALEX Technical Documentation<sup>12</sup> for more details). The data is divided into a set of images and catalogues associated with each pointing (tile) in the mosaic covering the GAMA fields.

<sup>11</sup>see GALEX technical documentation

<sup>12</sup><http://www.galex.caltech.edu/researcher/techdocs.html>  
[http://www.galex.caltech.edu/researcher/gr6aocs/GI\\_Doc\\_Ops7.pdf](http://www.galex.caltech.edu/researcher/gr6aocs/GI_Doc_Ops7.pdf)

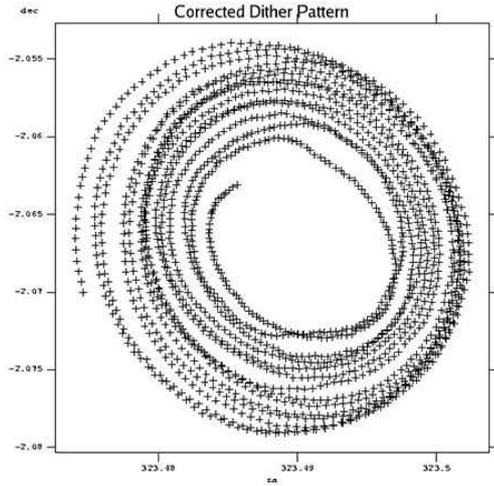


Figure 2.3: Dither pattern

An example of a typical dither pattern (for the example of eclipse 23456). Note that it represents a continuous spiral pattern, about 1 arcmin across, made over the course of the whole observation. The conversion from detector coordinates to an RA and DEC position on the sky is done by using position and orientation information of the spacecraft and by tracking sources in the FOV over time against the known locations of those same stars in catalogues. Units are in degree and the time resolution is 1 second. Figure taken from GALEX technical documentation.

The **count maps** are generated by converting the detected photon counts from detector coordinates into celestial coordinates (see Fig. 2.4) with a pixel scale of 1.5 arcsec. The **relative response maps** give the effective exposure time of each data map pixel (pixel scale of 1.5 arcsec). The edges of this map are not sharp, and blind detector pixels will show on this map as darker rings due to the dither pattern of the telescope. Besides the benefit of preventing over-exposure of the detector active area, another positive side effect of the dither pattern is the fact that even if the detector itself has a blind pixel, the data maps have a reduced, but non-zero exposure in that position.

Combining count maps and relative response maps, the **intensity maps** (units counts per second, pixel scale of 1.5 arcsec) are produced. These are the basis on which SExtractor (Bertin & Arnouts, 1996) then identifies the background (smoothed over an area of several pixels and provided as separate **background maps** with pixel scale of 1.5 arcsec) and extracts the sources. Background-removed maps are provided and, amongst other information<sup>13</sup>, position, size, axis ratio, orientation angle and Kron magnitudes (Kron, 1980) of the identified sources are listed in the **tile-level source catalogue**<sup>14</sup> for sources passing a signal-to-noise cut of 2.5. For GALEX-GAMA, we employ the NUV-based data. Thus both NUV and FUV quantities are derived inside an “NUV aperture”, containing the extent of the galaxy as determined by the pipeline shape fit algorithm on the NUV map. Using this NUV based shape is preferable because the S/N of the NUV data is significantly higher than that of the FUV data. It improves the reliability and completeness of the FUV detections and provides meaningful UV colours. It should, however, be noted that because source detection is carried out in the NUV, it is possible for negative flux measurements in the FUV to be listed in the pipeline catalogues. Only the NUV measurement must pass the detection limit of 2.5 signal-to-noise in order for the source to make it into the pipeline source catalogue.

<sup>13</sup>see [http://www.galex.caltech.edu/researcher/files/mcat\\_columns\\_long.txt](http://www.galex.caltech.edu/researcher/files/mcat_columns_long.txt)

<sup>14</sup>The tile-level source catalogues are combined to produce the blind catalogue, see Sect. 3.2

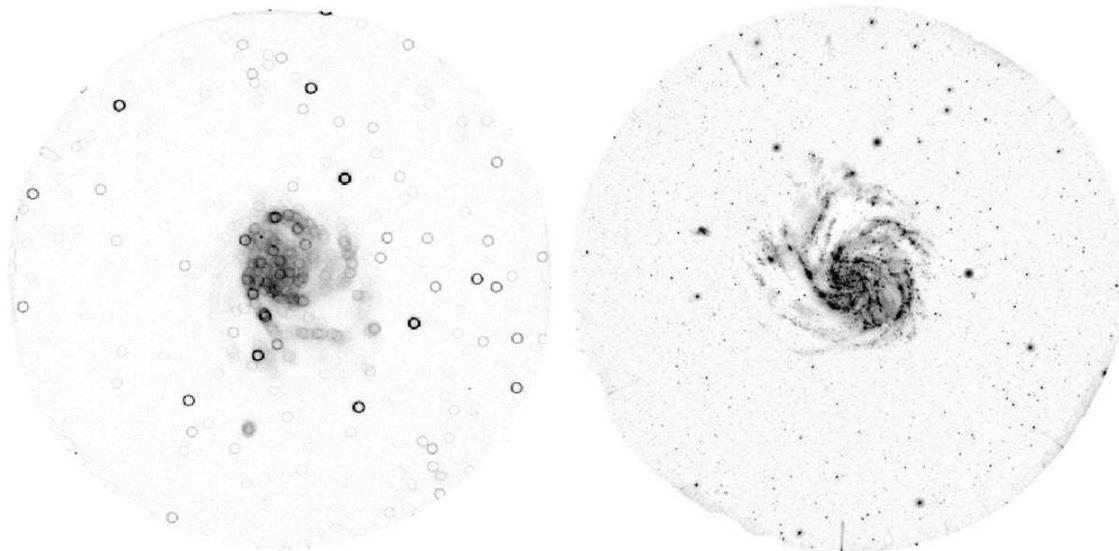


Figure 2.4: pipeline processing

Left: A GALEX NUV count map of M101 in detector coordinates clearly showing the effect of the 1' dither pattern, which is used for all observations to smooth out the effects of the detector flat field and avoid over-exposure of the detector. Stars and compact sources in the disk of the galaxy (i.e. point sources) appear as dark rings on the map in detector coordinates, due entirely to the dither pattern.

Right: The fully processed, background-subtracted NUV image of M101 in sky coordinates, which provides the basis for the pipeline source catalogue data.

Note that the detector coordinates are not aligned with the RA-DEC coordinates, resulting in a rotation of the count map in detector coordinates compared to the final map. Figure taken from Morrissey et al. (2005).

For quality control, the pipeline produces low resolution (15" pixel scale) **flag maps** which highlight regions with potentially compromised photometry due to artefacts (see Table 2.5 for a description of all types of artefacts). The most severe artefacts are dichroic reflections (reflections of bright stars on the dichroic beam splitter) and window reflections (NUV only). Both effects manifest themselves as rings on the count maps, causing SExtractor to misinterpret them either as separate sources themselves or as part of a nearby true source. Edge reflections produced by bright stars slightly outside of the field of view, are generally too conservatively flagged. However, there are extreme cases which need to be selected against (see Sect. 3.7.2).

### 2.2.3 Imaging performance

The point-spread function (PSF) and astrometric precision of GALEX are critical quantities in our matching to the GAMA survey (see chapter 3) and in general for the extraction of photometry. The image resolution (FWHM of the PSF) is 5.3" and 4.2" in the NUV and FUV, respectively, with small variations across

---

and for column description see App. D.1.

Flag name	Bit	Flag_ N value	Meaning
Edge	1	1	Detector bevel edge reflection (NUV only)
Window	2	2	Detector window reflection (NUV only)
Dichroic	3	4	Dichroic reflection
Varpix	4	8	Variable pixel based on time slices.
Brtedge	5	16	Bright star near edge of field (NUV only). Just like #1 but with a higher count rate requirement (about 10 cps) on the generating star. Cuts down on false positives at the cost of a higher false negative rate.
Rim	6	32	Detector rim (annulus) proximity (>0.6 deg > from FOV center).
Dimask	7	64	Dichroic reflection artefact mask flag. Used only when a coadd has enough visits at enough position angles that masking the dichroic reflection doesn't decrease flux by more than 1/3.
Varmask	8	128	Masked pixel determined by varpix.
Hotmask	9	256	Masked detector hotspots.

Table 2.5: GALEX pipeline artefact description

the detector (see Morrissey et al. (2007), especially their Figs. 9 and 10). The astrometric precision (i.e. systematic uncertainty) is of the order of 0.3" for the GALEX-GAMA survey.<sup>15</sup>

Fig. 2.5 shows a density histogram of the offsets between the GAMA positions of stars in the INPUTcatAv06 and GALEX positions after cross-matching the Blind GALEX (see Sect. 3.2) and GAMA catalogues. The offsets are not systematic and vary from tile to tile, which is why we did not correct for this position bias. The principle cause of the astrometric offsets is the limited precision of astrometric catalogues used to derive plate solutions within the GALEX pipeline.

However, this small bias will have little effect on the matching performance described in the following Chapter 3, as 50% of all simple matched sources (matching radius of 4", see Sect. 3.3) lie within 0.9" distance (difference between optical and NUV central coordinates), and 90% lie within 2.6" distance.

<sup>15</sup>For MIS depth, Morrissey et al. (2007) even quote an NUV position error of  $\sigma \sim 0.48''$ . This translates to  $1\sigma \approx 1/3$  pixel of the final map.

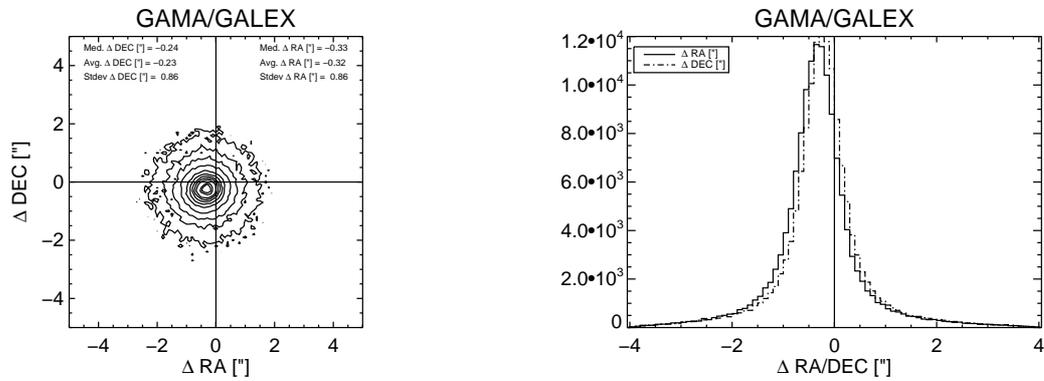


Figure 2.5: astrometric precision of GALEX

Left: The stacked UV flux distribution of stars in the INPUTcatAv6 GAMA catalogue (contour levels=[0.01, 0.025, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 0.9] · maximum value) revealing a small offset between the GALEX position and the GAMA position. Right: Same distribution collapsed onto the RA (solid line) and the DEC axis (dashed line). The overall astrometric precision is 0.3", and slightly varies from tile to tile. Figure provided by Mark Seibert.



# Chapter 3

## The GALEX-GAMA data products

The basic goal of this work is to provide a set of robust UV photometry for all galaxies in the GAMA optical spectroscopic survey. This can be done in two different ways. Firstly, one can extract a list of fluxes and positions of UV sources in the GAMA footprint, which can be identified on the UV maps without any reference to optical positions. This UV source list is then compared with the list of positions of GAMA sources and positional coincidences (within some well defined tolerance criteria) are considered to constitute an identification of the UV counterpart of the optical source. This process is called “matching”. The second approach is to examine the UV maps at the positions of the optical galaxy and extract UV fluxes within an aperture defined according to some clear presumptions. This is called “aperture photometry”. Matching and aperture photometry both have advantages and disadvantages. Both techniques were employed in the creation of the GALEX-GAMA catalogues.

Both matching and aperture photometry are non-trivial processes in real surveys. For example, differences in depth and PSF, or different physics in different wavebands can significantly affect the identification of sources, as can an object’s size, shape, and surface brightness. All this can systematically affect the content of the catalogue.

Specifically considering the matching methods, the nearest neighbour match (see e.g. Wyder et al., 2007, or Sect. 3.3 for this work) is a straightforward approach as it ignores any individual characteristics of the catalogues and instruments used. More elaborate methods use a Bayesian approach and assume a spherical normal distribution for the distance between the two matching partners (e.g. Budavári & Szalay, 2008; Sutherland & Saunders, 1992), different colour probabilities for different kinds of stars/galaxies (e.g. Naylor et al., 2013), probabilistic positions of the catalogued sources (e.g. Heinis et al., 2009; Brewer et al., 2013). A redistribution of the flux<sup>1</sup> of the catalogue derived from data with larger PSF among nearby objects of the catalogue derived from data with smaller PSF was presented by e.g. Robotham & Driver (2011).

Aperture photometry can also be done in a range of ways: The simplest is PSF-homogenisation followed by matched-aperture photometry (Hill et al., 2011, for GAMA aperture-matched photometry in the optical/NIR bands). This is a

---

<sup>1</sup>They are employing weights proportional to the inverse of the distance between the sources in order to imitate the source positioning of SExtractor, which is based on the first moment of the light profile.

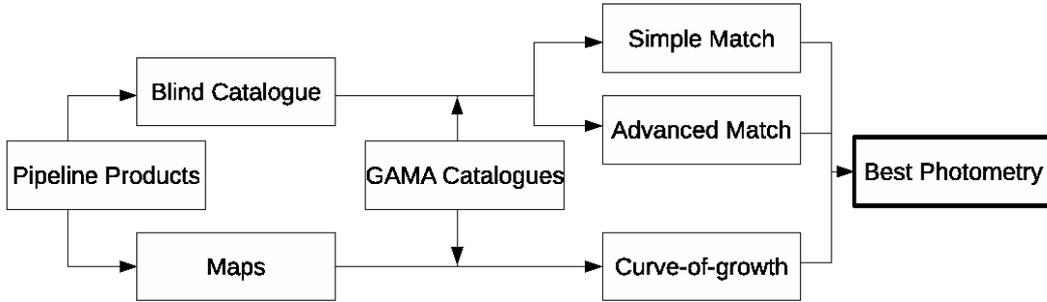


Figure 3.1: Flowchart of the matching procedure

re-measurement of the flux inside a common aperture defined in the optical.

In this chapter, we design three methods for measuring UV fluxes for objects in the GAMA InputCatAv06 (the GAMA II SDSS-based input catalogue): a simple (nearest neighbour) matching technique (Sect. 3.3; so-called “simple match”), an “advanced matching” technique employing flux redistribution between objects in the optical and the UV catalogues (Sect. 3.4), and a curve-of-growth technique (COG; Sect. 3.5). The latter is a form of aperture photometry in which the aperture is mainly defined from the UV data using prior information about the optical shape. A flow chart showing the relation of the basic products (simple matched, advanced matched, and COG catalogues) to the GALEX pipeline data is given in Fig. 3.1.

It is necessary to pre-process the tile-level GALEX pipeline products before we can use them as input to the process of matching and extraction of photometry. The pre-processing is described in Sect. 3.1.

The GALEX-GAMA data base will be available to the community through the data releases of GAMA. This means that the data products are organised within a so-called Data Management Unit (DMU), which can be linked together with a whole range of other GAMA products (e.g. optical spectroscopy or submm photometry), allowing multi-wavelength analysis. This chapter outlines the components of the GALEX-GAMA DMU. It consists of a blind catalogue (Sect. 3.2) combining the tile-level GALEX pipeline products into a single catalogue and the catalogues produced using the different matching techniques and aperture photometry (Sects. 3.3, 3.4, and 3.5). A catalogue combining the different techniques to provide an off-the-shelf best UV magnitude catalogue is presented in Sect. 3.6.3. Ancillary data products, i.e. a catalogue listing meta data of the GALEX observations at GAMA source positions and HEALPIX overview maps, are presented in Sect. 3.7.

All these catalogues are provided for each of the three GALEX depths. We list the column names, units, default values and a brief column description in Appendix D.

### 3.1 Preprocessing

The tile-level GALEX pipeline source catalogues and maps, described in Sect. 2.2.2, are the main input to the GALEX-GAMA products. One point to note is that no attempt was made to create mosaics combining the maps for each tile and extract photometry of sources from these mosaics. This would be advantageous, since regions where tiles overlap would have deeper photometry. Furthermore,

the task of collating source lists from the individual tiles would have been much simplified. However, this task is quantitatively difficult and error prone due to the variable background and particularly due to reflection artefacts on the edges of the individual fields (see background HEALPIX map of G09 in Fig. 3.21, middle panel). While low resolution HEALPIX mosaic maps are generated (see Sect. 3.7.2), the pipeline products (e.g. high resolution maps or pipeline source catalogues) are unfortunately not available for these mosaics. We therefore rely on the tile-level pipeline products as input to GALEX-GAMA.

In this section, we describe the handling of these tile-level pipeline catalogues before they are used to build the GALEX-GAMA catalogues. Since processing is dependent on the depth of the exposures, we define three survey depths (exposure time ranges) in Sect. 3.1.1. A key concept is the definition of regions associated with each tile from which source photometry is extracted. We explain these so-called “primary” and “secondary” regions and their definition in Sect. 3.1.2. The characteristics of the background in the GAMA fields are presented in Sect. 3.1.3. Lastly, it was necessary to redefine the flux measurement error (compared to the pipeline values) based on the local background and exposure time. This process is presented in Sect. 3.1.4.

### 3.1.1 GALEX-GAMA shallow, medium, and deep coverage

The (NUV) exposure time of the GALEX tiles overlapping the GAMA regions varies considerably from less than 100 seconds to 115,000 seconds. Because of this, the noise characteristics and the relative importance of systematic and random effects change markedly between low exposure and high exposure tiles. In addition, while in general more exposure time is beneficial, since it allows deeper data and higher S/N, there are also downsides to it. Long exposure times are achieved by co-adding several individual visits (see Sect. 2.2.1) of shorter exposure time. Since the spacecraft has a different orientation angle for each visit, not only are the photon counts co-added, so are the artefacts. At some point, the disadvantage of an increasing number of artefacts outweighs the advantage of deeper data. Furthermore, it may also be of interest to have similar exposure times, e.g. if a more or less homogeneous detection limit is desired.

Therefore, GALEX-GAMA is structured in three survey depths, i.e. nominal exposure time<sup>2</sup> ranges: GALEX-GAMA shallow coverage (GSC, 0-800 seconds), GALEX-GAMA medium coverage (GMC, 800-10,000 seconds), and GALEX-GAMA deep coverage (GDC, >10,000 seconds). The GSC data is taken from the GALEX All Sky Survey, while the GMC data is taken from the GALEX Medium Imaging Survey, the GALEX-GAMA guest investigator program, and the sky completion. The GDC data was taken from the Deep Imaging Survey. A histogram of the effective exposure time<sup>3</sup> of the three equatorial GAMA fields for each survey depth is shown in Fig. 3.2. As the relative response of the NUV detector is slightly less than that of the FUV detector and due to various corrections along the pipeline (e.g. correction for the dead time of the detector in high flux situations), the FUV has

---

<sup>2</sup>**Nominal exposure time** refers to the actual time the tile was observed.

<sup>3</sup>**Effective exposure time** refers to the exposure time at a given coordinate. Due to the dither pattern, blind pixels, dead time and the relative response of the detector, this number differs from the nominal exposure time and varies over the tile.

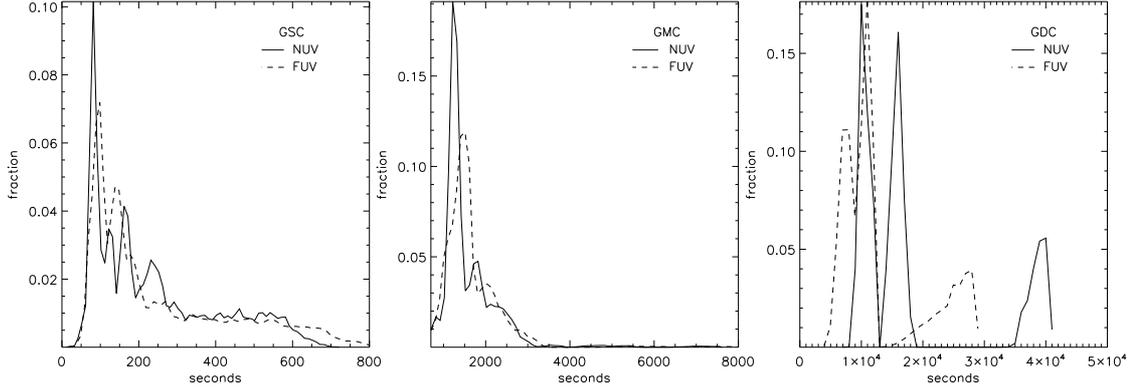


Figure 3.2: Histogram of the NUV and FUV effective exposure time.

The effective exposure time distributions (normalised histograms) for the equatorial fields are plotted for all three survey depths in the NUV and in the FUV. The distributions are similar in both bands, as the data are mostly taken simultaneously.

slightly more effective exposure time than the NUV if both bands are observed. It is also possible that tiles (or sub-visits of the considered co-adds) have NUV coverage only due to the shut-down of the FUV detector in the later part of the GALEX mission. The latter is the main reason for the spread between NUV and FUV effective exposure in the GDC. The three equatorial GAMA fields have only 5 tiles in GDC, of which only 4 have FUV exposure. Therefore, the effective exposure time distribution is less smooth than in the GMC and GSC. Furthermore, all these tiles are co-added from several tiles with lower exposure. Not all tiles used in the co-add had FUV exposure, leading to a lower FUV exposure than NUV exposure in the GDC.

All products are generated for the three survey depths separately. It should be noted that the artefact density in GDC is very high, demanding great caution when using GDC data.

As a rule of thumb, one can expect a given coordinate with coverage in a longer exposure category to have exposure in the lower ones as well. However, there are exceptions where single visits<sup>4</sup> were long enough to qualify for the GMC category without existing data in the GSC.

### 3.1.2 Primary/secondary region

Given the circular shape of the GALEX tiles, one must decide what to do with overlapping regions. Ideally, one should co-add all data onto a common grid in RA and DEC. However, difficulties arise, e.g., from varying backgrounds (e.g. caused by large edge reflections from bright nearby stars) which would lead to a discontinuous background in the co-adds.

As an alternative, the GALEX instrument team defines “primary” regions on

<sup>4</sup>These longer visits are in regions that were not covered by the GALEX all sky survey. This is generally due to the fact that in the beginning of the mission (when the data for the GALEX all sky survey was observed) the avoidance of bright foreground stars was executed more rigorously than in the later part of the GALEX mission, where the majority of the GALEX-GAMA GMC data was observed.

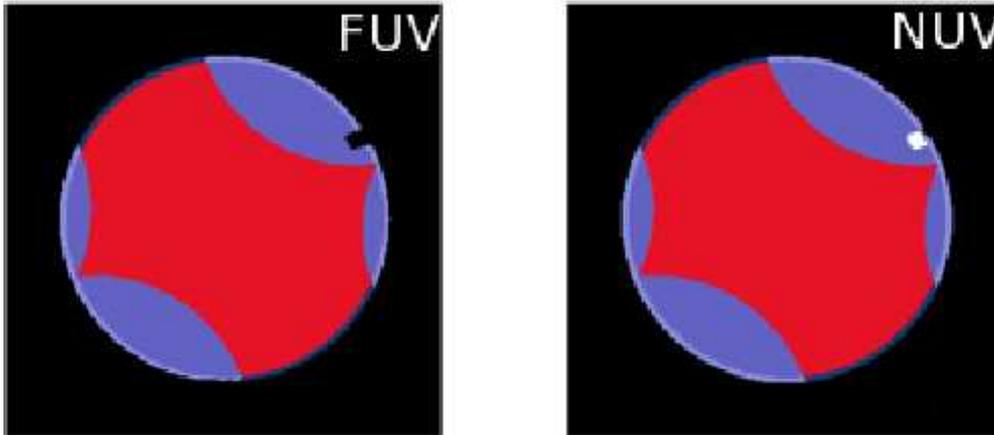


Figure 3.3: Illustration of primary and secondary regions.

For the tile MISDR1\_33675\_0589, the primary region (red) and secondary region (different shades of blue) are shown. The dysfunctional part on the FUV detector (black square) is also discarded from the NUV data (white square; see text for detail).

each tile, i.e. identifies the region of each tile containing data which is considered superior to the data of other tiles covering the same area (“secondary” regions). This way, each point on the sky has primary coverage on exactly one tile (or no coverage at all).

The definition of primary and secondary regions we employ for GALEX-GAMA is slightly different than the one used by the GALEX instrument team in the Imaging Surveys. The standard procedure of the GALEX instrument team is to adopt a hexagonal shape for the primary region, thus a RA/DEC coordinate always has primary coverage on that tile where it is close to the centre of the tile. However, since our data is less homogeneous in exposure time, this definition might lead to defining higher exposure time regions as secondary since they are further away from their tile centre.

We therefore define the tile with primary coverage as the tile with the higher nominal exposure time. Since only part of the data has FUV coverage (see Sect.2.2.1) and we are interested in having as much FUV coverage as possible, we assign regions with the greatest FUV nominal exposure time as primary. For regions with no FUV exposure, we adopt the rule with maximum NUV nominal exposure time in assigning primary regions.

A further aspect of declaring a region as secondary is the inferior data quality in the very outer area of a tile. There, the sensitivity is less-well characterised and reflection artefacts are common (see GALEX Technical Documentation<sup>5</sup>). We therefore define the primary region of a tile to have a distance to the tile centre of  $\leq 35$  arcmin rather than the 36 arcmin for the full FOV<sup>6</sup>.

Furthermore, a small region on the outer edge is declared secondary by defi-

<sup>5</sup>See also GALEX Observer’s Guide <http://www.galex.caltech.edu/researcher/faq.html> .

<sup>6</sup>In other words, the regions which have a distance of more than 35 arcmin to the centre are not used in the GALEX-GAMA survey. However, rather than discarding those regions completely, they are defined as secondary region. It should be noted that at the edges of the GALEX-GAMA coverage footprint, this might lead to GAMA sources having only secondary coverage, as they have a distance of more than 35 arcmin to the centre of the only tile that covers the source’s coordinates.

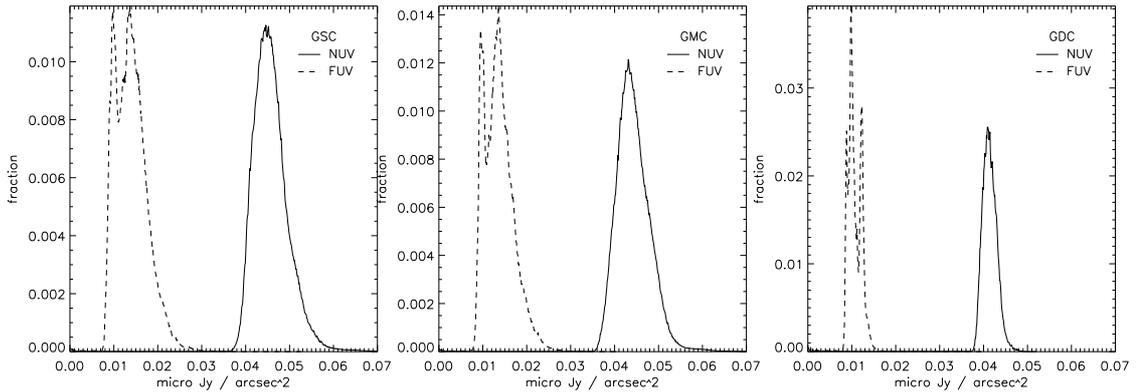


Figure 3.4: Histograms of the NUV and FUV background brightnesses

The distributions of the background emission in the background maps for the three survey depths for the equatorial GAMA fields in the NUV and in the FUV, all normalised histograms, are shown. They are very narrow and significantly higher than the background averaged over the full sky (see text).

dition (Fig. 3.3, white region in the NUV, black in the FUV). In this region, the FUV detector is malfunctional and by declaring the regions secondary, tiles with coverage of this region at the sky but lower nominal (FUV) exposure time are enabled to be declared primary.

Fig. 3.3 shows the primary/secondary map (15" pixel scale; which is the resolution of the pipeline flag maps) for the example tile MISDR1\_33675\_0589. The map encodes the primary/secondary classification bitwise, which is represented as colour code in the figure. The red region represents the primary region. Secondary regions due to overlap with other tiles are represented in blue, while the secondary regions beyond 35 arcmin are indicated as dark blue. Secondary regions where both criteria apply are indicated as light blue. Black indicates no coverage, and white indicates the small region in the NUV which is discarded due to the lack of FUV data.

### 3.1.3 UV Background

The background measured by GALEX has slightly different origins in the NUV and FUV band (see GALEX Technical Documentation<sup>5</sup> for more details). The diffuse Galactic light (starlight scattered by dust grains in our Galaxy) dominates the background in the FUV. It varies depending on the column density of the dust and therefore displays large-scale structure. The NUV band is dominated by the zodiacal light, although there is also a substantial contribution from diffuse Galactic light. The zodiacal background varies with the position of the Earth around the Sun, i.e. it changes with observation date.

Nightglow of the residual atmosphere produces a moderate background in both bands, increasing at the beginning and the end of each eclipse. Detector background is very low in comparison ( $<1\%$ ), except in local “hot spots” which are masked in the pipeline processing.

Altogether, the UV background is very low. Expectation values given by the GALEX instrument team for the average background of the sky<sup>7</sup> are of the order of

<sup>7</sup>See GALEX Technical Documentation.

Field ID	GALEX depth	GALEX band	Median Background $\mu\text{Jy}/\text{arcsec}^2$	Median Background counts/pixel
G02	GSC	FUV	0.0098	0.0339
G02	GSC	NUV	0.0388	0.415
G09	GSC	FUV	0.0156	0.0587
G09	GSC	NUV	0.0471	0.534
G12	GSC	FUV	0.0102	0.0376
G12	GSC	NUV	0.0439	0.502
G15	GSC	FUV	0.0146	0.0753
G15	GSC	NUV	0.0456	0.806
G23	GSC	FUV	0.0098	0.0342
G23	GSC	NUV	0.0358	0.435
G02	GMC	FUV	0.0103	0.304
G02	GMC	NUV	0.0385	3.42
G09	GMC	FUV	0.0152	0.612
G09	GMC	NUV	0.0451	5.36
G12	GMC	FUV	0.0101	0.298
G12	GMC	NUV	0.0419	3.72
G15	GMC	FUV	0.0142	0.450
G15	GMC	NUV	0.0447	4.10
G23	GMC	FUV	0.0089	0.255
G23	GMC	NUV	0.0365	3.28
G02	GDC	FUV	0.0099	4.86
G02	GDC	NUV	0.0374	49.2
G09	GDC	FUV	0.0124	2.82
G09	GDC	NUV	0.0409	42.8
G12	GDC	FUV	0.0010	2.20
G12	GDC	NUV	0.0414	48.1
G15	GDC	FUV	-	-
G15	GDC	NUV	0.0414	32.5
G23	GDC	FUV	-	-
G23	GDC	NUV	-	-

Table 3.1: NUV and FUV background level.

Listed are the median background values for the three survey depths for the five GAMA fields (area per pixel = 1.5" x 1.5"). The numbers illustrate the strong large-scale dependence of the UV background on wavelength and position. Note that since the median in NUV and FUV is not taken over the same area for a given field and depth (NUV only tiles exist), we do not expect a constant factor between the NUV and FUV flux or count rates.

$\sim 10^{-3}$  counts per second per pixel ( $\sim 0.015 \mu\text{Jy}/\text{arcsec}^2$ ) in the NUV and  $\sim 10^{-4}$  counts per second per pixel ( $\sim 0.005 \mu\text{Jy}/\text{arcsec}^2$ ) in the FUV.

Fig. 3.4 shows histograms of the NUV and FUV background flux for the three survey depths for the equatorial GAMA fields. The background distributions are quite narrow, but significantly higher than the all-sky average. This is due to the position of the equatorial fields,<sup>8</sup> which have moderate Galactic latitudes and therefore have more diffuse Galactic light than regions perpendicular to the Galactic plane.

Tab. 3.1 lists the median background flux for all five GAMA fields. G09 and G15 have a higher median background in both NUV and FUV than the southern fields G02 and G23. G12 has a median background comparable with southern fields, highlighting the variability in position of the background.

### 3.1.4 Redefinition of measurement error

As mentioned in Sect. 2.2.2, we need to make use of the measurement uncertainties to select sources from the pipeline catalogue to include in the GALEX-GAMA catalogues<sup>9</sup>.

The GALEX pipeline measures the UV flux of a source by counting individual photons over the time period of the exposure time (the detector of GALEX is a photon count device, see Sect. 2.2.1). Thus, the flux measurement is not affected by readout noise. However, the photon count is subject to Poisson noise. While total counts of individual background pixels can be quite low (NUV background rate  $\sim 10^{-3}$  per second per pixel,  $1.5''$  pixel scale; see Sect. 3.1.3), a point source (with FWHM of  $5''$ ) covers an area corresponding to 22 pixels. For the GMC exposure time, the total background counts are already high enough to be well approximated by a Gaussian error (see Fig. 3.5 for the background counts distribution in the equatorial fields for the three survey depths and Table 3.1 for the average background level).<sup>10</sup>

The GALEX observer's guide claims that the pipeline error follows a Gaussian error  $\sigma_c$  on the photon counts

$$\sigma_c = \sqrt{(s_c + bg_c)} \quad (3.1)$$

where  $s_c$  and  $bg_c$  are the total counts of the source and the background integrated over the source aperture. (Conversion to flux error is done using Eqs. 2.4 and 2.5.) A comparison with the pipeline error estimates, however, showed that the pipeline

---

<sup>8</sup>The GAMA fields were optimised for accessibility for ground-based facilities. The equatorial fields can be observed from the northern as well as the southern hemisphere. The coordinates of each field's centre (Galactic longitude; latitude) are: G02 (172; -61), G09 (229; +28), G12 (276; +60), G15 (348; +54), G23 (14; -65).

<sup>9</sup>To some extent this decision was already made by the GALEX instrument team when the tile-level pipeline catalogues were built. However, we will need to reproduce this decision in the matching process.

<sup>10</sup>For the shallow coverage, the NUV background count is of the order of 0.5 counts per pixel, barely justifying the Gaussian approximation. However, the total flux of a detected source must be at least 2.5 times the noise estimate, resulting in source + background counts high enough to justify the Gaussian error approximation, bearing in mind point sources cover 22 pixels. The FUV background flux is technically too low for a Gaussian error approximation. Nevertheless, since the FUV measurement does not need to pass a detection limit, we adopt the NUV error definition for the FUV as well.

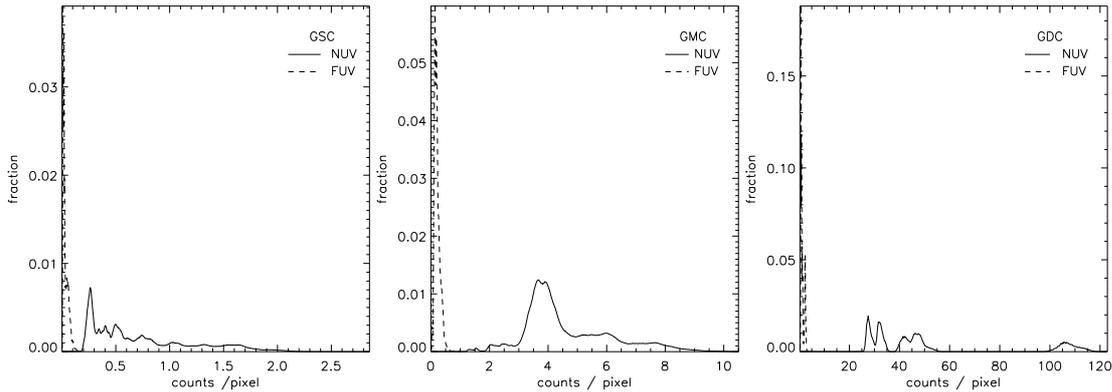


Figure 3.5: Histograms of the NUV and FUV background counts.

The distribution of the background counts per pixel in the equatorial GAMA fields in the NUV (solid line) and in the FUV (dashed line), both normalised histograms, are shown. They are dominated by the distribution of the exposure time. The background flux distribution is very narrow, as can be seen very well in the NUV distribution for GDC, where the individual distributions of each tile are visible.

error estimates seem to follow a modified Gaussian model with a slightly enhanced contribution of the background (which we will refer to as the “simple-model” in this section).

$$\sigma_c = \sqrt{(s_c + 1.3 \cdot bg_c)} \quad (3.2)$$

Fig. 3.6 (left) shows the NUV pipeline error estimate divided by the NUV simple-model error as function of the NUV magnitude for a narrow exposure time range for GMC sources. Fig. 3.6 (right) compares the actual error prediction of the pipeline (black) with that of the simple-model (red). For the majority of sources ( $\sim 75\%$ ), the simple-model reproduces the pipeline error estimate well<sup>11</sup>.

However, for  $\sim 25\%$  of the sources, the pipeline significantly underestimates the error. There appear to exist three groups with different error definitions than the group in agreement with the simple-model (the “correct” pipeline estimate). Investigations on our side and by the GALEX instrument team were not able to discover the reason for this discrepancy. All sources of a given tile either have correct pipeline errors or the same bias. There seems to be a tendency that single-visit tiles have correct pipeline errors and co-added tiles have too low pipeline errors, however there are exceptions to this.

We therefore decided to redefine the error estimate in order to remove the multi-modality of the pipeline error estimates. We also used the opportunity to include components into the error budget which were not considered in the pipeline error.

The new errors consist of two parts: A Gaussian error  $\sigma_f$  from the photon count statistic with (even further) increased background to account for a background measurement error. Another Gaussian component  $\sigma_{ff}$  (1% of the source flux)

<sup>11</sup>The scatter for faint sources is most likely due to small variations of the background over the source area, which is not represented in the single background flux number quoted in the tile-level catalogue.

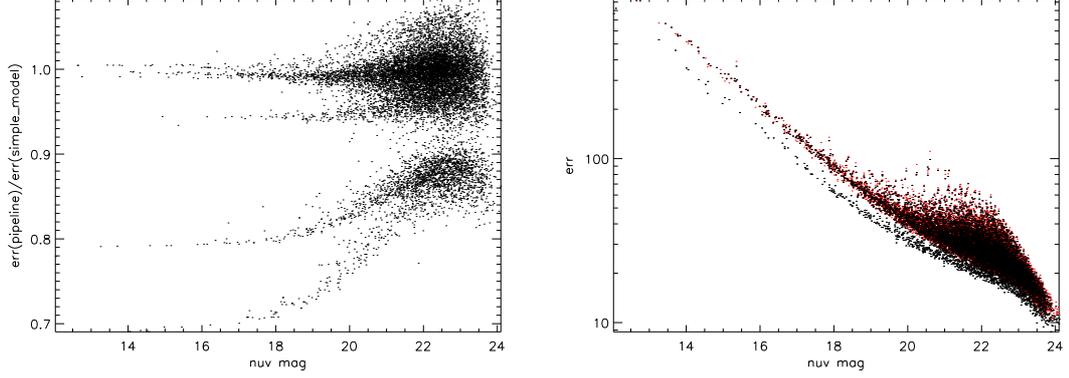


Figure 3.6: Redefinition of measurement error

Left: NUV pipeline error divided by NUV simple-model error as function of NUV magnitude (for the GMC). While for one group of sources the pipeline error estimate agrees well with the simple-model error (“correct” pipeline prediction), there exist three groups where the pipeline error estimate is significantly lower. For faint sources (fainter than  $\sim 20$ mag), random noise on the map seems to dominate the pipeline error prediction.

Right: NUV error estimate as function of NUV magnitude using the pipeline error (black) and the simple-model error (red). The simple-model error estimate reproduces the “correct” pipeline error prediction well. The further division of the groups visible in the plot on the left is not visible in this plot due to the distribution of background.

Sources were selected from the GAMA equatorial fields with a narrow range of nominal exposure time (1100-1400 seconds).

accounts for the flat field error. Both components are added in quadrature.

$$\sigma_c = \sqrt{(s_c + 1.5 \cdot bg_c)}$$

$$\sigma_{cps} = \sigma_c / t_e = \sqrt{\frac{(s_{cps} + 1.5 \cdot bg_{cps})}{t_e}}$$

$$\sigma_f = \sigma_{cps} \cdot cuj = \sqrt{\frac{(s_f + 1.5 \cdot bg_f) \cdot cuj}{t_e}} \quad (3.3)$$

$$\sigma_{ff} = s_f \cdot 0.01 \quad (3.4)$$

$$flux\_err = \sqrt{\sigma_f^2 + \sigma_{ff}^2} \quad (3.5)$$

where  $s_c$  and  $bg_c$  are total source and background counts,  $s_{cps}$  and  $bg_{cps}$  are total source and background counts per second,  $t_e$  is the effective exposure time in seconds,  $cuj$  the conversion factor defined in Eqs. 2.4 and 2.5 and  $s_f$  is the source flux in  $\mu Jy$ . Note that the noise is a function of both the source flux and the background flux, as well as of the exposure time.

The newly defined error calculated using Eq. 3.5 slightly overpredicts the error compared with the “correct” pipeline error, which is especially noticeable for bright sources, where the flat field error dominates the new error budget. Nevertheless, this redefined error takes into account effects that were neglected in the pipeline error budget. We therefore consider it to be more accurate than the “correct” pipeline estimate and use it in the construction of the GALEX-GAMA catalogues and throughout this work.

Note that the catalogue lists an error estimate for both flux and magnitude. The quoted error is equivalent to the Gaussian width of the error distribution only for the flux. The magnitude will have an asymmetric error distribution. The magnitude error is quoted for the convenience of users, who often use a Gaussian approximation of the magnitude error distribution. It is defined as

$$\text{mag\_err} = \frac{2.5}{\ln(10)} \cdot \frac{\text{flux\_err}}{\text{flux}} \quad (3.6)$$

which follows directly from the Gaussian error propagation of the flux-magnitude relation.

## 3.2 Blind catalogue

In order to obtain a catalogue of all the sources detected by GALEX in the GAMA footprint<sup>12</sup>, independent of whether or not there is a correlation with a GAMA galaxy, it is necessary to combine the tile-level pipeline source catalogues. However, this cannot simply be done by cocatenation of all the individual tile-level pipeline catalogues. Multiple exposures in regions with overlapping tiles would lead to multiple entries of sources in the resulting catalogue. The GALEX team therefore cocatenates the pipeline catalogue entries only for sources with primary coverage only (see Sect. 3.1.2 definition of primary coverage).

As mentioned in Sect. 2.2.2, the GALEX pipeline produces NUV-based catalogues. The specific shape of a source as fitted to the NUV maps is used in both NUV and FUV to extract the source flux. Thus, the FUV sources are not measured independently from the NUV. Since only the NUV measurement must pass the detection limit of signal-to-noise ratio (S/N) of 2.5, it is possible for a source to have very small FUV count rates, even below the background level, (in which case a negative FUV source flux results).

Appendix D.1 lists and describes all columns of this **blind catalogue**, produced by the GALEX instrument team specifically for this project. For practicality, the pipeline tile-level catalogues are reduced to the most useful columns. The full pipeline product is in principle accessible via the unique GALEX ID "GGOID" (see Sect. D.1 for column description).

Note that we deliberately did not apply a S/N cut to the blind catalogue. Due to the overestimation of the S/N by the pipeline (due to the underestimation of the pipeline error), faint sources were included in the tile-level catalogues which would not pass the S/N cut of 2.5 using the redefined errors (see Sect. 3.1.4). We include these faint sources in the blind catalogue in order to supply the user with as much data as possible. However, during the GALEX-GAMA matching described in the following sections, we will apply the cut in order to generate a homogeneous data set.

Even though the primary regions are sharply defined, the uncertainty of the source position can cause a single source, identified on two overlapping tiles, to have primary exposure in both tiles ( $\sim 0.1\%$  of sources). This scenario leads to multiple entries of the same source in the blind catalogues. In the same way, a

---

<sup>12</sup>For convenience, we generate three times five blind catalogues, one for each GAMA field and survey depth combination.

source can happen to have secondary exposure in both tiles and thus be excluded from the catalogue. The multiple entries can be identified via the distance to their nearest neighbour. A distance between two sources smaller than the FWHM of the PSF would have led to a blending of these sources during the pipeline source extraction. The majority of multiple entries are filtered out during the construction of the blind catalogue.

### 3.3 Simple matching catalogue

Having generated the blind catalogue, we now identify the GALEX counterpart for every GAMA source by matching the blind catalogue to the InputCatAv06 (the GAMA II SDSS-based input catalogue). The easiest way to match two catalogues is a simple nearest neighbour match. For each source in the InputCatAv06, we identify the nearest UV neighbour (i.e. source in the blind catalogue) within 4 arcsec<sup>13</sup>.

The simple match catalogue reports GAMA ID along with the nearest neighbour blind catalogue ID (see Sect. D.1). GAMA objects without a GALEX nearest neighbour within a distance of 4 arcsec are considered unmatched and are not included in this catalogue. To distinguish between GAMA objects that were not detected in the GALEX data and those that were not covered by GALEX, we refer to the ObsInfo catalogue (Sect. 3.7.1). In order to avoid unnecessary duplication of information, this catalogue does not include the properties of a given GAMA object's nearest neighbour GALEX object. To obtain them, the user can access the blind catalogue using the listed GALEX ID.

Since the resolution of the GALEX (NUV) imaging is  $\sim 5.3$  arcsec, a given GALEX detection in the blind catalogue may not represent the UV flux of a single GAMA object (defined in the optical) but the joint UV flux of multiple GAMA objects. In these cases it would be incorrect to assign all of the GALEX flux to any single GAMA object. Furthermore, doing so would introduce an environmental bias: GAMA objects with many other close-by objects would be more likely to be associated with a GALEX detection than isolated GAMA objects. Furthermore, as the nearest UV object is found individually and independently for each GAMA object, it is possible for a single UV object to be the nearest neighbour of more than one GAMA object.

Hence, as a quality indicator of the match, the number of GALEX objects within 4 arcsec distance to the GAMA objects is listed in column NMATCH4. The column MANY2ONE gives the number of GAMA objects this GALEX object has been matched to. If one of these two columns is larger than 1, this must be considered a multiple match and we recommend the use of the advanced match catalogue entry (Sect. 3.4.1) for this GAMA object.

As a further quality indicator, an ID list of all GALEX objects within 20 arcsec of a GAMA object and their actual distance to that object are also provided.

We refrain from applying a signal-to-noise cut<sup>14</sup> on this catalogue in order to

---

<sup>13</sup>This is the matching radius used by the GALEX team to match to optical surveys. For the GAMA survey, 50% of all simple matched sources lie within 0.9" distance (difference between optical and NUV central coordinates), and 90% lie within 2.6" distance.

<sup>14</sup>As discussed in Sect. 3.1.4, we use a redefinition of the noise estimate. The signal-to-noise cut based on the old pipeline estimate applied to the blind catalogue is carried through into this

preserve as many objects as possible and leave this decision to the user.

## 3.4 Advanced matching catalogue

In the light of the deficiencies of the simple nearest neighbour match discussed above, the necessity of a more advanced matching technique becomes apparent. Accordingly, we developed the advanced matching procedure which takes into account the possibility of multiple matches between GALEX and GAMA objects and seeks to reconstruct the true UV flux of a given GAMA object.

We describe the advanced matching technique in detail in Sect. 3.4.1, and discuss trends and biases as function of galaxy quantities and catalogue depth in Sect. 3.4.2. Further diagnostics and comparisons to the other matching techniques are discussed later in Sect. 3.6.1.

### 3.4.1 Description of the technique

The advanced matching procedure takes into account the possibility of multiple matches between GALEX sources from the blind catalogue and GAMA sources from the InputCatAv06 and seeks to reconstruct the original UV flux of a given GAMA source. We provide both flux measurements, the redistributed UV flux (FLUXSPLIT\_XXX; where XXX substitutes for NUV or FUV) for each GAMA source along with the total UV flux within the GAMA source's optically defined aperture, as explained below (FLUXTOT\_XXX; which might be a combination of the source's flux and the flux of neighbouring sources).

We illustrate the routine for an example GAMA object and refer to Fig. 3.8 for an illustrating cartoon and Fig. 3.7 for a flowchart of the routine. First, optical shape information for the object (r-band ellipticity, position angle, and effective radius) is extracted from the catalogue SersicCatv09, and used to define a target area within which GALEX counterparts will be deemed to be associated with the GAMA object. To take into account scatter between apparent positions and real positions in the UV and possibly unavailable optical shape information, the minimal search region, i.e. optical target area, is a circle of radius 4 arcsec (the matching radius used in the simple match). The question that is then to be answered is which GALEX objects in the blind catalogue have their central coordinates within this optically defined search region. The number, the IDs and the total flux of all such GALEX objects are recorded in columns NMATCHUV, UVIDXLIST and FLUXTOT\_XXX of the advanced match catalogue, respectively (see Sect. D.3 for the column description).

Since it is possible that a GALEX object may itself be extended, or that it has more than one potential close-lying optical counterpart in InputCatAv06, we also need to take into account the possibility that one or more of the GALEX objects in the search region around the GAMA object may themselves be related to further GAMA objects. Hence, in the second step of the advanced matching procedure, a UV area is constructed for each GALEX object within the optical search region, using the object's position angle, ellipticity and size along its semimajor axis from the blind catalogue. InputCatAv06 is then searched for GAMA objects within the

---

catalogue. However, the redefinition will now cause some objects to have a S/N of less than 2.5.

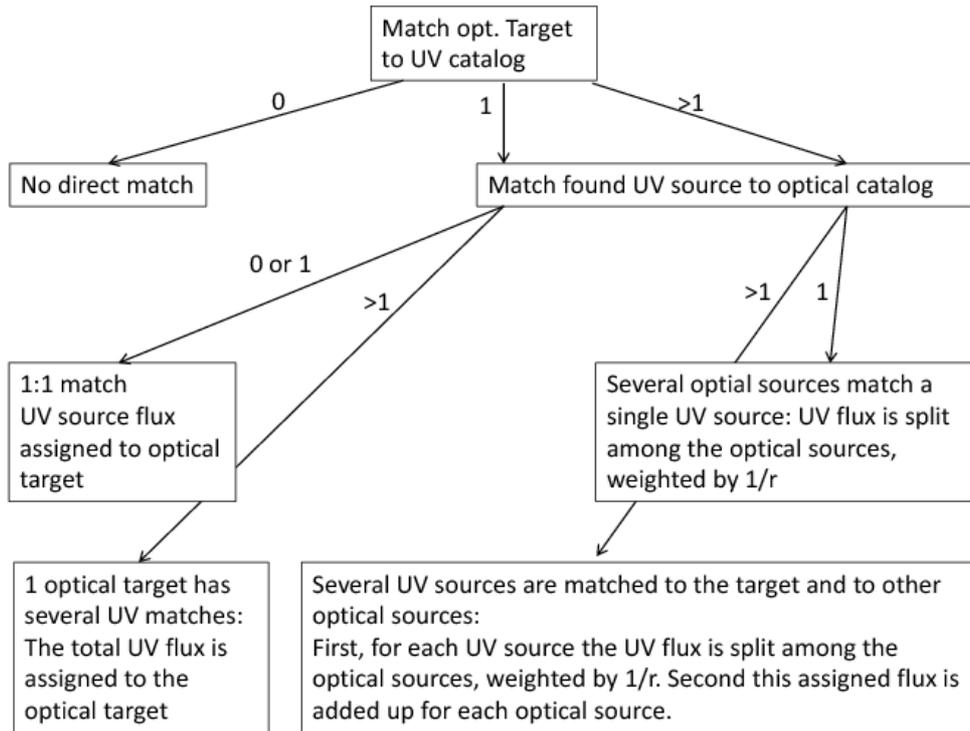


Figure 3.7: An overview of the decision tree of the advanced matching.

UV area of each of the GALEX objects. The number and IDs of the GAMA objects found in this step are recorded in the columns `NMATCHOPT` and `OPTIDXLIST`. The columns `NMATCHUV` and `UVIDXLIST` for the GAMA objects found in this step are updated to include this GALEX object.<sup>15</sup>

In cases where there is a 1-to-1 match all of the flux of the GALEX object is attributed to the matching GAMA object. This is the most common case (57%). If, however, there is more than one potential optical counterpart to one or more of the GALEX objects in the optical search region, the UV flux of each of the GALEX objects is split among all potential optical counterparts of that GALEX object, weighted inversely by angular distance (using a minimum distance of 0.3 arcsec to account for positional errors), thus mimicking the source positioning of SExtractor (Bertin & Arnouts, 1996). Finally, the UV flux contributions from all of the GALEX objects in the optical search radius to the GAMA object under consideration are summed up and recorded in the columns `FLUXSPLIT_XXX`. `FLUXSPLITERR_XXX` is the square-root of the weighted quadratic sum of the errors of all GALEX objects from which the object under consideration received a flux contribution, where the weights in the quadratic sum are the same weights as those used for the flux itself.

Note that it is possible for a GAMA object not to have a `FLUXTOT_XXX` value associated with it but to still have a `FLUXSPLIT_XXX` value associated with it. This will happen if there are no GALEX objects within the GAMA object's optical search region (leading to `FLUXTOT_XXX` being set to its default value

<sup>15</sup>Note that in case of multiple optical matches for a GALEX object, both GAMA objects will include the full flux of the same GALEX object in `FLUXTOT_XXX`.

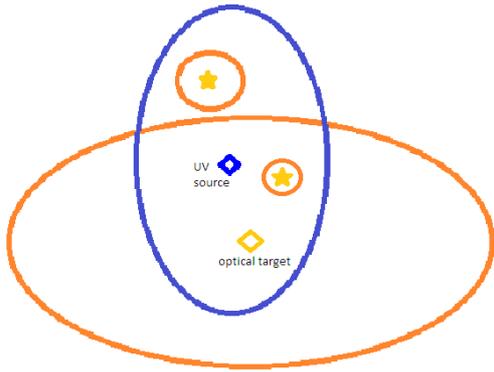


Figure 3.8: Cartoon illustrating the advanced matching

The yellow diamond represents the optical position of the galaxy of interest (the target), the yellow stars indicate positions of other optical sources, and the orange ellipses represent the size of these sources in the optical. The blue diamond indicates the position of the UV source with a size indicated by the blue ellipse. In this example, the UV source occupies a region of sky that is not only populated by the target, but also by two other stars. The UV flux reported for this source needs to be distributed among these three optical sources during the matching process (see text for details).

# opt. matches	1	2	3	> 3
# sources	301389	141381	47359	34463
# UV matches	1	2	3	4
# sources	515438	9060	92	2

Table 3.2: Statistics of multiplicity of matches for the INPUTCatAv06 (stars and galaxies) matched to the GMC advanced match catalogue.

of -99.0) and if the GAMA object lies inside the UV search region(s) of one or more GALEX objects, which in turn lie in the optical search region of at least one other GAMA object. In this case, the first GAMA object participates in the redistribution of the UV flux from these GALEX objects only. This catalogue only includes matched GAMA objects with  $S/N\_NUV \geq 2.5$  in the redistributed flux FLUXSPLIT\_NUV (i.e. a S/N cut is applied). Unmatched GAMA objects are not included in this catalogue. To distinguish between GAMA objects that were not detected in the GALEX data and those that were not covered by GALEX, and to obtain flux limits, we refer to the ObsInfo catalogue (Sect. 3.7.1).

We also point out that there are substantial differences between the objects included in this catalogue and those included in the simple match catalogue. An object may be included in this catalogue but not in the simple match catalogue and vice versa. The former happens when a GAMA object is more than 4 arcsec away from its nearest GALEX neighbour but still receives flux from one or more GALEX objects during the flux re-distribution (with  $S/N\_NUV \geq 2.5$ ). The latter happens when the flux redistribution results in the S/N of the NUV flux received by a GAMA object listed in the simple match catalogue dropping below the threshold of 2.5.

For multiple matches, the redistributed flux estimate is superior to the simple match catalogue entry, as it is less vulnerable to the biases discussed above. See Tab. 3.2 and Fig. 3.10 for statistics of the match multiplicity. For a statistical comparison of the flux measurements derived using the advanced match and the simple match, see Sect. 3.6.1.

We recognise that the advanced match flux redistribution scheme (i.e the weighting by  $1/r$ ) may not be optimal, as the true flux distribution of the optical counterparts is not known. Nevertheless, it significantly improves the reliability of the flux measurement over the population of galaxies compared with the simple match scheme (see Sect. 3.6.1), although is, necessarily, imprecise for individual targets.

### 3.4.2 Correlation with catalogue depth and galaxy quantities

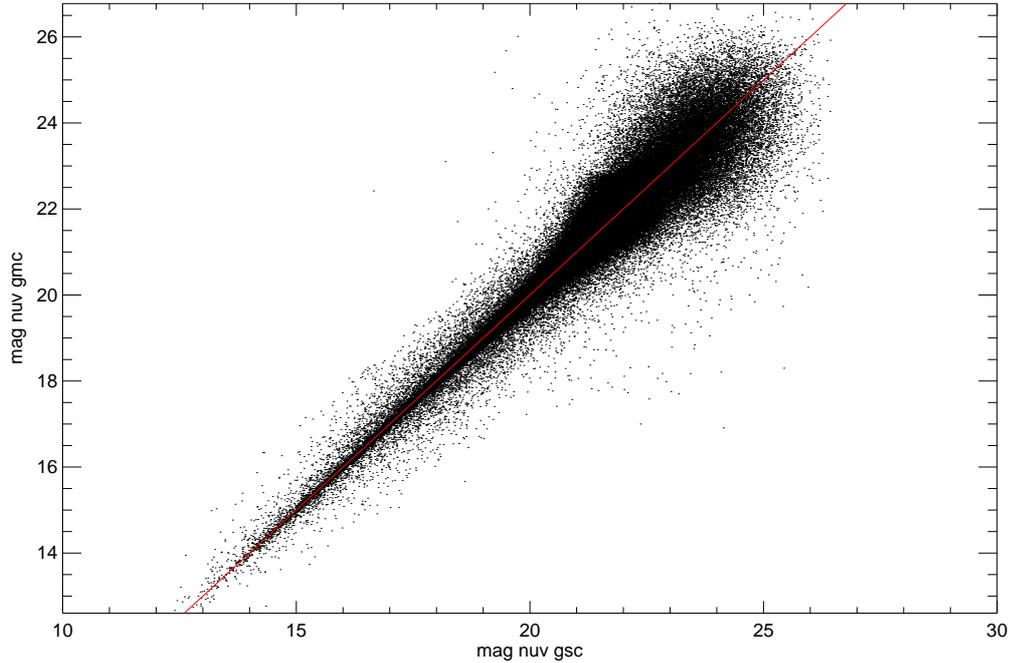


Figure 3.9: Comparison of the advanced match NUV magnitude listed in both the medium and shallow catalogues.

The comparison of the advanced match redistributed NUV magnitude between objects listed in both the medium and the shallow coverage catalogue shows the good agreement between the two depths and illustrates the noise of the measurements.

The multiplicity of matches using the advanced matching technique is highly dependent on the surface density of source in the optical and the UV catalogues. The probability of a random object to be within the search radius increases with the surface density. This is especially important for the optical catalogue, as objects blended together in the UV can be identified as individual sources in the optical. Nevertheless, it is important to include all sources which potentially contribute to the UV flux in the matching process. In theory, there is a sweet spot for the depth of the optical catalogue which counterbalances the advantages and disadvantages of the catalogue depth. One could even include further information about the optical sources, such as optical colours, in order to judge the probability of a source having UV emission. However, we deliberately decided against the use of the optical colours in order to not bias the matching against true UV-optical colours which are not consistent with the assumptions made.

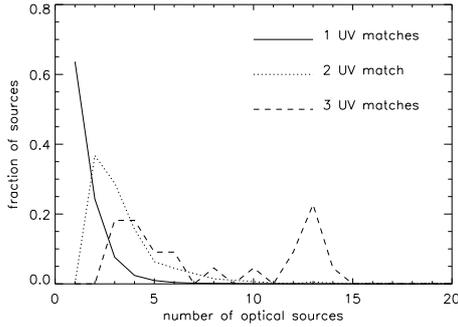


Figure 3.10: GMC advanced match multiplicity distribution.

For one, two and three UV matches, the distribution of number of optical matches is shown.

Fig. 3.9 shows a comparison of the redistributed NUV magnitude using the GSC and the GMC UV data (using optical sources that have a UV match in both catalogues). There is no systematic bias noticeable between the redistributed magnitudes listed for the two depths. The measurement noise in the UV can be seen in the scatter around the red line, symbolizing identical measurements.

The distribution of multiplicity of matches in the optical for different multiplicity of matches in the UV is shown in Fig. 3.10. The majority of sources has a single UV match (solid line) and no further optical source is involved in the matching (i.e. a unique match). A high multiplicity in UV matches correlates with a high multiplicity in optical matches, and vice versa.

The angular size of a source, both in the UV and in the optical, is another parameter which strongly influences the advanced match. Large objects are more likely to have neighbouring sources close enough to be within their matching radius. Fig. 3.11 (right figure is the zoomed-in continuation of the left) illustrates the distribution of unique matches (i.e. 1-to-1 matches: dotted line) and multiple matches (dashed line) as function of the r-band effective radius. As expected, the distribution for multiple matches is much more prominent for large radii than the one for unique matches. The same trend, even more prominent, is seen for the dependency on the NUV semimajor axis shown in Fig. 3.12.

An apparent dependence of the match multiplicity on redshift is shown in Fig. 3.13. While both multiple and unique matches roughly follow the same redshift distribution, multiple matches are slightly more frequent for low redshifts. However, since angular size depends on redshift, this was to be expected and does not imply a direct redshift dependence of the match multiplicity.

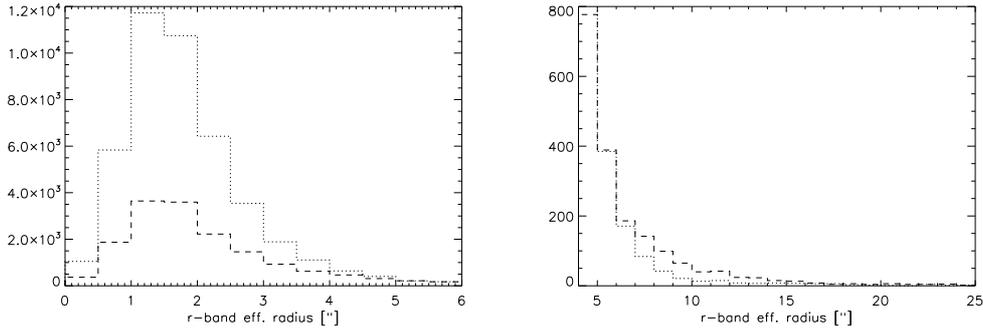


Figure 3.11: These figures show the distribution of the r-band effective radii for sources with unique matches (dotted line) and multiple matches (dashed line). For small effective radii (left), the majority of sources have unique matches, i.e. unresolved sources usually have 1-to-1 matches. For large radii (right), there are more multiple matches than unique ones. This is to be expected, as the probability to find another source (e.g. a foreground star or another optical source in general) within or very close to the area of a source increases for large radii. (Note: These “other sources” might be small, but get involved with multiple matches this way.)

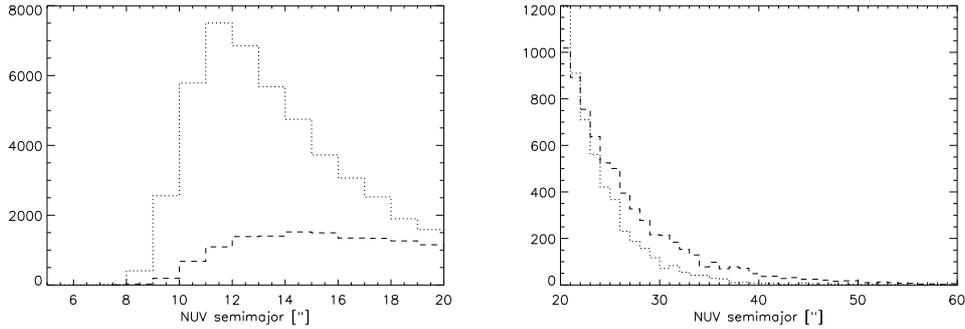


Figure 3.12: These figures show the distribution of the NUV semimajor axes for unique matches (dotted line) and multiple matches (dashed line). For small semimajor axis (left), the majority of sources have unique matches, i.e. unresolved sources usually have 1-to-1 matches. For large semimajor axes (right), there are more multiple matches than unique ones. This is to be expected, as the probability to find another source (e.g. foreground star, generally another optical source) within or very close to the area of a source increases for large radii.

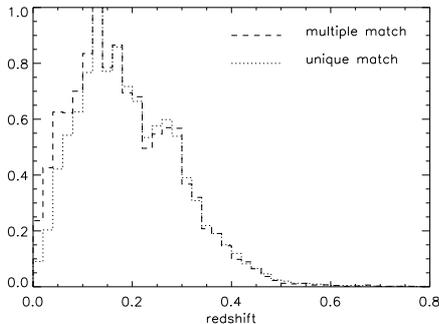


Figure 3.13: Redshift dependence of advanced match multiplicity

A histogram of the redshift (normalised to the highest bin) for unique matches (dotted line) and multiple matches (dashed line) is shown. The redshift has no direct influence on the multiplicity of the matching. However, it mirrors the size distribution with red shift, resulting in a higher multiplicity of the matches for low red shift, i.e. more large resolved objects.

### 3.5 Curve-of-growth analysis

The pipeline is optimised for extracting the photometry for small sources without structure. However, by fitting light profiles, an assumption about the intrinsic light profile of the object is made. In the case of large, resolved objects, especially large galaxies, or in the case of crowded fields, this standard assumption may not be valid, leading to biased flux estimates. It therefore becomes unreliable for sources with a semimajor axis of more than  $\sim 20$  arcsec (for a comparison of the COG flux and the redistributed flux from the advanced matching as function of NUV semimajor axis, see Fig. 3.17). Since both the simple match as well as the advanced match use the pipeline flux estimates as input, they inherit this drawback.

As a benchmark for the pipeline photometry, we use a curve-of-growth (COG) method<sup>16</sup> for galaxies in the TilingCatv41, making use of prior knowledge about the target and neighbouring sources, derived from optical photometry (i.e. position, ellipticity and position angle). Furthermore, this method is especially valuable as an alternative flux measurement for the following reasons: Firstly, it enables reliable flux measurement in the case of crowded fields, i.e. in presence of neighbouring or even overlapping sources, which the GALEX pipeline merges into a single source. Secondly, COG measures the flux for all optical galaxies, even for sources without a UV match (using any of the matching techniques mentioned above). Thus, we have a COG flux measurement where we would otherwise have to work with the 2.5 S/N detection limit as upper limit. Thirdly, the GALEX pipeline becomes unstable for large sources. They are likely to be shredded into several small sub-sources and thus heavily bias low the matched UV flux. The non-parametric curve-of-growth technique used to re-measure the UV flux directly on the UV maps relies as little as possible on assumptions about the UV shape of the target. The technique places only basic constraints on the position and axis ratio of a considered target imported from the optical shape fits. In the case of extremely UV faint galaxies, the optical effective radius is used to constrain the size of the target.

Having highlighted the benefits of the COG measurement, we now present the algorithm in detail.

For each galaxy in the TilingCatV41 (GAMA catalogue of galaxies only), we extract the size and ellipticity information from the SersicCatv09 (Kelvin et al., 2012, default value: circular point source). The GALEX tile in which the galaxy’s coordinates have primary exposure is identified and a large area around the target galaxy, a cutout of the tile-level intensity map, flag map, and exposure time map is made (cutout edge length = 10 times the r-band effective radius of the target). A further small map is generated which flags the area beyond 36 arcmin from the tile center, thus avoiding edge artefacts which would otherwise bias the measurement<sup>17</sup>.

A mask is generated which identifies “good” pixels to be used in the COG measurement and “bad” pixels which are likely to have flux biased by neighbouring sources or artefacts. To this end, we first mask all window and dichroic reflections

---

<sup>16</sup>The COG method was jointly developed with Meiert Grootes, who applied an adapted version to HERSCHEL data.

<sup>17</sup>This is effectively a mask selecting against secondary regions in case the source itself has primary exposure, but is close enough to the edge to have secondary region in its immediate vicinity. However, this is a slightly weaker selection than the primary/secondary selection, which declares everything beyond 36 arcmin away from the tile centre as secondary.

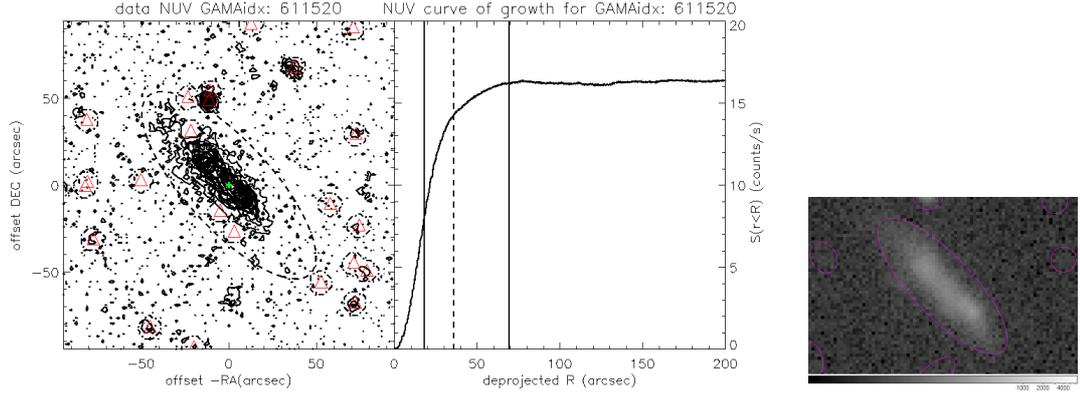


Figure 3.14: COG for the example of GAMA galaxy 611520.

The contour plot of GAMA galaxy 611520 (optical centre indicated by the green cross) on the left is overplotted with the galaxy’s aperture as found by COG (dashed lines). Also shown are neighbouring sources (red triangles) and the masked area around them (dashed-dotted lines). Inside the galaxy’s aperture, the interpolated flux is shown inside the masked areas. Outside the galaxy’s aperture, the actual observed flux is shown.

In the center, the actual curve-of-growth is overplotted with the COG half-power radius, the optical effective radius and the COG semimajor (solid line, dashed line, solid line, from left to right).

On the right, the NUV flux map is overplotted with the contour found by the GALEX pipeline. See text for details.

as listed in the flag map. Then, we import the position information of neighbouring sources from the InputCatAv06 and, for galaxies, their size and ellipticity information. A circular aperture with radius of  $1.1 \cdot \text{FWHM}$  of the  $5.3''$  NUV PSF (or, if larger, an ellipse with the observed ellipticity and a semimajor axis equal to three times the observed effective r-band radius) around the neighbouring source and flagged as potentially containing biased flux. While the majority of the listed stars are not detectable in the UV, some do need to be masked and individual, very bright stars can even have a significant flux outside of the above defined aperture. Therefore, and in order to mask neighbouring NUV sources that are not listed in the InputCatAv06, the size and ellipticity information of the blind UV sources around the target are used to flag<sup>18</sup> these known UV sources (with aperture defined by the ellipticity, semimajor and semiminor axis as listed in the blind catalogues).

We now calculate the mean flux of the “good” pixels within elliptical<sup>19</sup> annuli (width 3 arcsec, i.e. two pixels measured along the semimajor axis) around the optical centre of the target galaxy. Assuming symmetry, this average value for the “good” pixels is identical (within the sampling noise) with the average flux per pixel of a bias-free annulus.

We have chosen the cutout size such that we can safely measure 10 annuli beyond the source which are free of target flux and contain background only.

<sup>18</sup>Since the UV blind catalogues will, in the majority of cases, include the target itself, we use this UV based flag only for the background estimation, and NOT for masking sources overlapping with the target.

<sup>19</sup>We use the optical prior knowledge of the orientation angle and axis ratio convolved with the UV PSF (i.e. assumed Gauss profile, FWHM  $5.3''$ ).

Thus, the mean flux<sup>20</sup> of these 10 annuli is taken as first background estimate and the background level error  $\sigma_{bg}$  is calculated.

The semi-major axis of the target galaxy is then defined as the smaller of the following: the outer radius of the smallest annulus in which the average flux falls below the background level plus  $1.0 \cdot \sigma_{bg}$  or the outer radius + 3 arcsec of the smallest annulus in which the average flux falls below the background level plus  $1.5 \cdot \sigma_{bg}$ . This value was found to be a good compromise between including very faint target emission at large radii on the one hand and the need to avoid pure noise integration or blending with neighbouring objects on the other.

Having defined the aperture of the target galaxy, the background is now defined as the mean value of all (unmasked) pixels outside of the galaxy's aperture. The total flux of the target is obtained by integrating over the flux inside of the aperture after having removed the background. Masked areas inside the galaxy area are accounted for by renormalising the unmasked galaxy flux measured at the same elliptical radius as the mask.

In addition to the semimajor axis, the half light radius is determined, identifying the radius at which the curve-of-growth has reached half of the target's flux.

As a quality control, the column "CONFSOURCE\_XXX" reports the presence of masked pixels inside the target galaxy's COG-defined aperture. Targets with a masked area in the central radius bin are flagged as not measurable since the true flux distribution in the centre is unknown and cannot be reproduced. Targets with a semimajor axis smaller than two radius bins must be considered to be noise fluctuations since their size is smaller than the UV PSF.

Galaxy aperture determination as well as flux integration are done independently for the NUV and the FUV. If both bands have exposure time, a comparison between the radii found can serve as a quality control. Stars, in particular, tend to be much fainter in the FUV. Thus, even very bright stars are very unlikely to have excess FUV flux outside the masked area, even though they may have excess NUV flux outside the masked area. Hence, the excess NUV flux might lead to the false inclusion of the neighbouring source inside the galaxy's aperture and bias high the semimajor axis in the NUV, while the mask will be sufficient in the FUV, leading to an unbiased semimajor axis in the FUV. If the semimajor axis in the FUV is more than 30 arcsec smaller than the semimajor axis in the NUV, the NUV measurement is considered as biased by a neighbouring star and is set to the default value 999.

In addition to the flux within the COG-defined size of the galaxy, we also report the total flux measured within the optical aperture (defined by the observed ellipticity and a semimajor axis equal to 3 times the r-band effective radius) of the galaxy. Especially for galaxies with very low UV flux, it is beneficial to integrate over a known galaxy footprint, rather than to try to define it from noise-dominated data.

Note that the COG measurements (both the actual COG flux and the inte-

---

<sup>20</sup>As the UV background is very low ( $\sim 10^{-3}$  counts per second per pixel in the NUV and  $10^{-4}$  in the FUV), even for GMC exposure times of 1,500 seconds we expect a Poisson distribution with  $\mu = 1.5$  background photons per pixel (also see Fig. 3.5 for the distribution of the background values). A median pixel value as background estimate would therefore be very insensitive to the overall background level due to the quantised nature of the photon counts.

grated flux within the optical aperture) do not have a non-detection in the classical S/N cut definition. We report the integrated (noisy) background-subtracted flux within a galaxy’s aperture, thus it may be negative.

Fig. 3.14 shows the COG result for an example galaxy (GAMA CATA\_INDEX 611520). On the left, a contour plot of the NUV emission is shown, overplotted with the symbols for the known neighbouring sources. Masked regions are enclosed by dashed-dotted lines. Inside the target area (dashed line), the interpolated flux is displayed inside the masked area. Outside of the target area, the actual observed flux is displayed. While there is no flux from neighbouring sources left inside the target area (as intended), the bright star just above the target area is still visible. The middle panel displays the actual curve-of-growth, with the two vertical lines indicating the half-power radius and the full radius of the target as found by COG. On the right, a cutout of the NUV map is shown, overplotted with the NUV aperture found by the GALEX pipeline<sup>21</sup>.

Fig. 3.15 displays performance tests of COG on simulated images. For point sources on the left (green), extended Gaussians in the middle (blue), and extended Sérsic profiles (with Sérsic indices up to 8) on the right (red), the measured COG flux (“recovered flux”) is compared with the input flux of the simulated galaxies (overplotted in black are the integrated fluxes within the simulated “optical” aperture equal to an ellipse with semimajor axis of 3 times the “optical” effective radius<sup>22</sup>). In the top row, both fluxes are plotted against each other. The COG flux reproduces the input flux very well for point sources and extended Gaussians. For extended Sérsic profiles, the correspondence is still reasonably good, but COG slightly underestimates the input flux since the faint wings of the source fade into the background, thus underestimating the source size. The integrated flux within the optical aperture is even more affected by the extended wings as it assumes by definition that all of the source’s flux is included within 3 times the effective radius.

In the middle row, the ratio of the recovered flux and the input flux is plotted vs. the surface brightness (as measured by COG) / background. Again, one can see the input flux is well reproduced for point sources and extended Gaussians, and it is underestimated for the extended Sérsic profiles. For very faint sources (surface brightness / background < 0.7), COG has problems determining the radius of the source and tends to underestimate the input flux, while the integrated flux within the optical aperture still reproduces the input value on average. We therefore define surface brightness / background = 0.7 as the flux level below which we recommend to use the integrated flux within the optical aperture instead of the COG flux. In the bottom row, the ratio between the COG effective radius (i.e. half-power radius) and the input effective radius as function of surface brightness (as measured by COG) / background is shown. For point sources, the COG effective radius suffers from the discretisation of the annuli used for the size estimate. While the effective radii of extended Gaussian sources are well reproduced, they are underestimated for sources with extended Sérsic profiles, as discussed above. For all three profile

---

<sup>21</sup>Even though the GALEX pipeline correctly identified and measured the target, the different flux distributions in the optical and the NUV caused the optical and NUV central coordinates to be almost 7 arcsec apart. Thus, this galaxy does not have a match in the simple match catalogue, and we rely on the measurements in the advanced match catalogue or the COG catalogue.

<sup>22</sup>For the simulation, we used the same shape for the NUV and the optical flux distribution. For extended sources, in particular, this will not be the case for real galaxies.

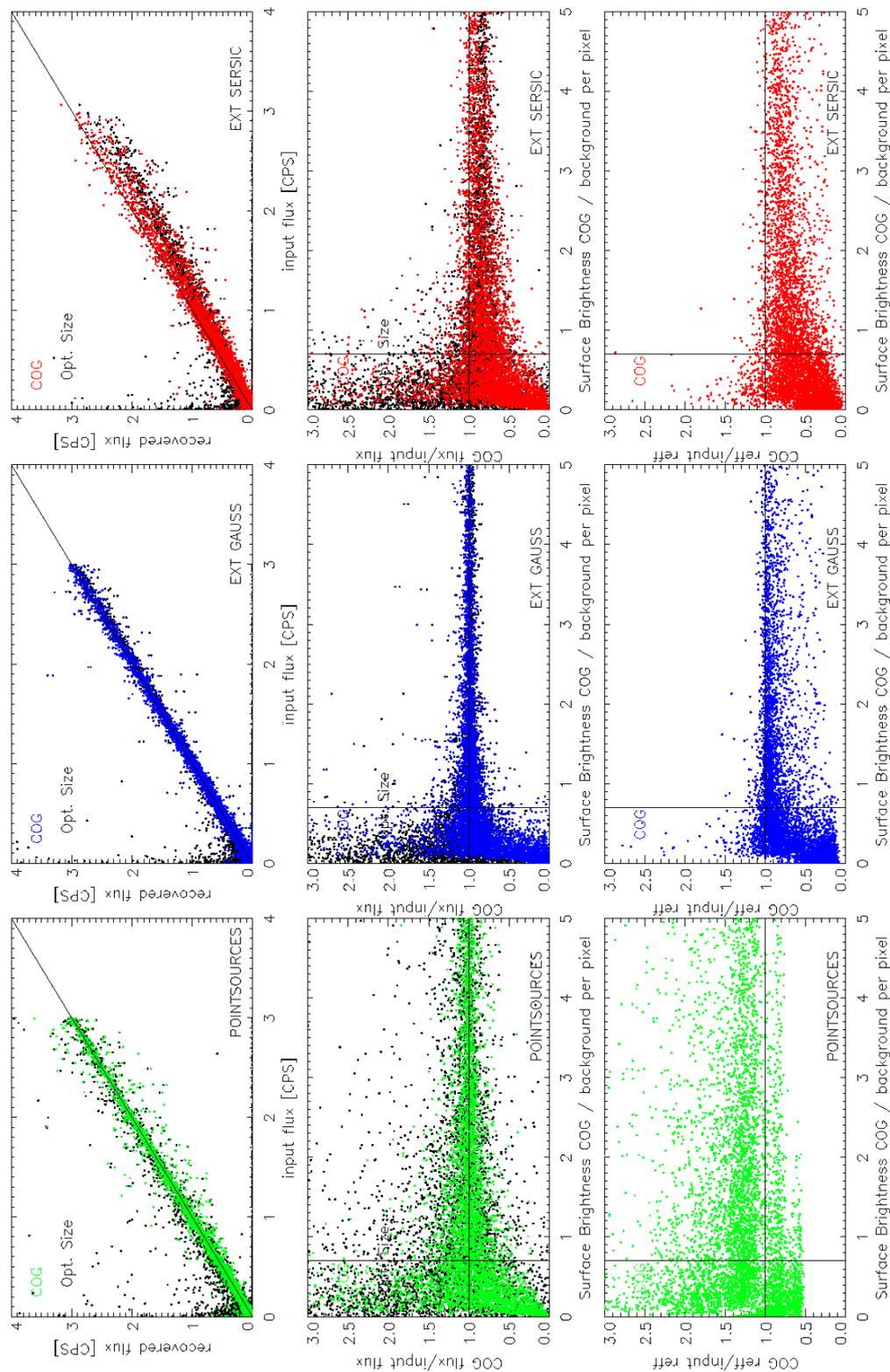


Figure 3.15: COG reliability

The images show the reliability of the COG measurements for simulated galaxies (point sources on the left in green, extended Gaussians in the middle in blue, and extended Sérsic profiles on the right in red.) and the comparison with the measurement using the optical aperture (black, same panels).

Top row shows COG flux vs. input flux, middle row shows the COG flux - input flux ratio vs. the COG surface brightness - background level ratio, and the bottom row shows the COG effective radius - input effective radius ratio vs. the COG surface brightness - background level ratio. The vertical back line at COG surface brightness - background level ratio = 0.7 indicates the suggested switch from COG flux (for bright sources) to the integrated flux within the optical aperture (for faint sources). See text for details.

types, the radius is underestimated if the surface brightness / background falls below 0.7.

## 3.6 Choosing the best method

### 3.6.1 Comparison of all three flux estimates

	GSC	GMC	GDC
blind	573520	650048	18439
simple match	356097	418060	6632
advanced match	418229	524592	7960
COG	-	167315	-

Table 3.3: Number of objects in each GALEX-GAMA DMU catalogue for the equatorial fields

The blind catalogue, being the input to the matching process, contains the most objects. The simple match catalogue contains less matched sources than the advanced match catalogues due to the smaller matching radius. The actual number of sources in each survey depth depends on both the survey depth and covered area.

COG re-measures the flux of galaxies (and not stars) and is currently only available for the GMC depth. It should be noted that the COG catalogue also contains galaxies with  $S/N < 2.5$ .

Having presented the individual catalogues above, we now compare the different methods of flux estimation. In Tab. 3.3, we list the number of GAMA objects with a GALEX match in each of the catalogues for all three depths. As is to be expected, the advanced match catalogue lists the most GAMA objects, as it is designed to find at least the simple match counterpart. The COG catalogue is by design the smallest, as it does not attempt to measure flux for all GAMA objects, but only for the galaxies (and currently only for GMC depth).

For 1-to-1 matches, by design the simple match catalogue reports the same flux value as the redistributed flux in the advanced match catalogue (in case the distance between the optical centre and the UV centre is less than 4"; see Fig. 3.17, top panel, black crosses). Therefore, the 1-to-1 match situation is used as the “gold standard” with which the multiple matches are compared.

Fig. 3.16 shows the color magnitude diagrams of GAMA galaxies (equatorial fields; GMC depth) for the gold standard 1-to-1 matches (left) and for multiple matches using the magnitude from the advanced match catalogue (top right) and from the simple match catalogue (bottom right). For comparison, both multiple match colour magnitude diagrams are overplotted in red with the outer contour of the gold standard. The lack of red galaxies in the 1-to-1 matches sample compared with the multiple-match sample is a selection bias; elliptical galaxies tend to be large and red, so their chance of having foreground sources is larger than for spirals, and not a feature of the matching techniques. While the colour-magnitude distribution using the advanced match magnitudes very much resembles the 1-to-1 match distribution, the colour-magnitude distribution using the simple match

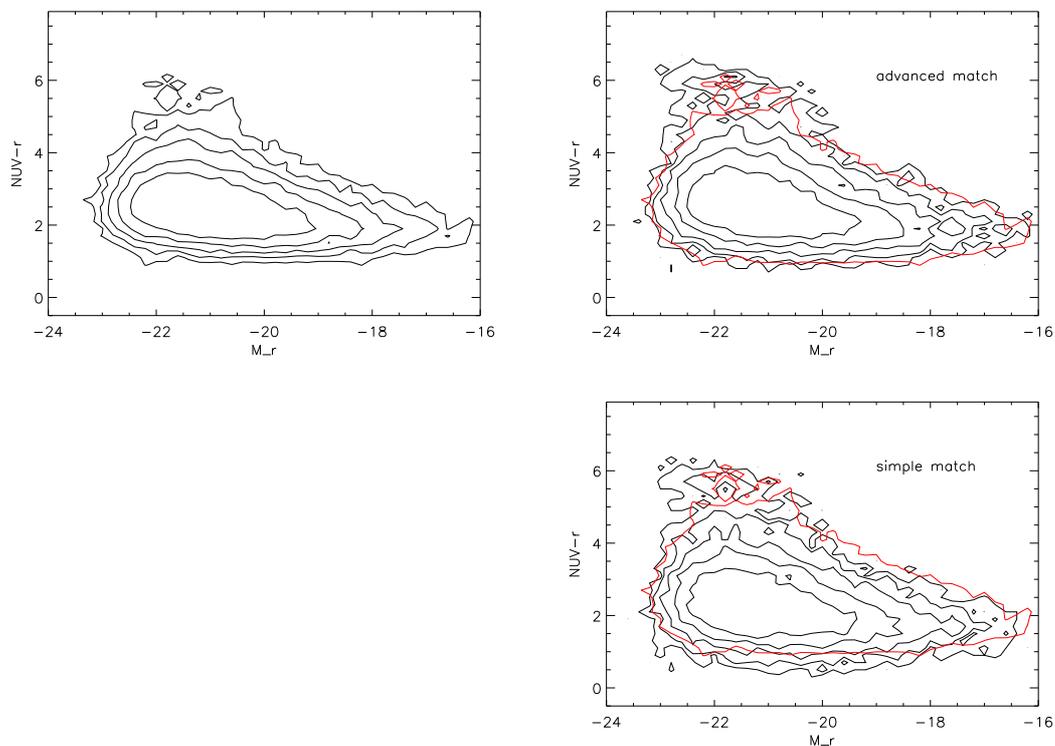


Figure 3.16: NUV-r vs. r colour magnitude diagram

The top right panel shows the advanced match NUV-r colour vs. r for TilingCatv41 galaxies, GMC depth, multiple matches only. The bottom right panel shows the simple match NUV-r colour vs. r for the same subsample. The top left panel shows the NUV-r colour vs. r for unique matches. Since these can be considered as intrinsic quantities, it is desired to reproduce this distribution. The outer contour of the NUV-r colour vs. r for unique matches is overlaid in red on both multiple match plots for easier comparison. Contour levels are [0.02,0.04,0.08,0.16,0.32] per cent of the according subsample. It is obvious that for multiple matches, the simple match is about  $\sim 0.2$  mag bluer than the unique matching, while the advanced matching reproduces the same colour distribution. The slight overpopulation of red, r-band bright sources for both multiple match methods is due to selection effects of the subsamples. Resolved sources, which tend to be brighter and/or nearer than the average, are more likely multiple matches than unresolved sources.

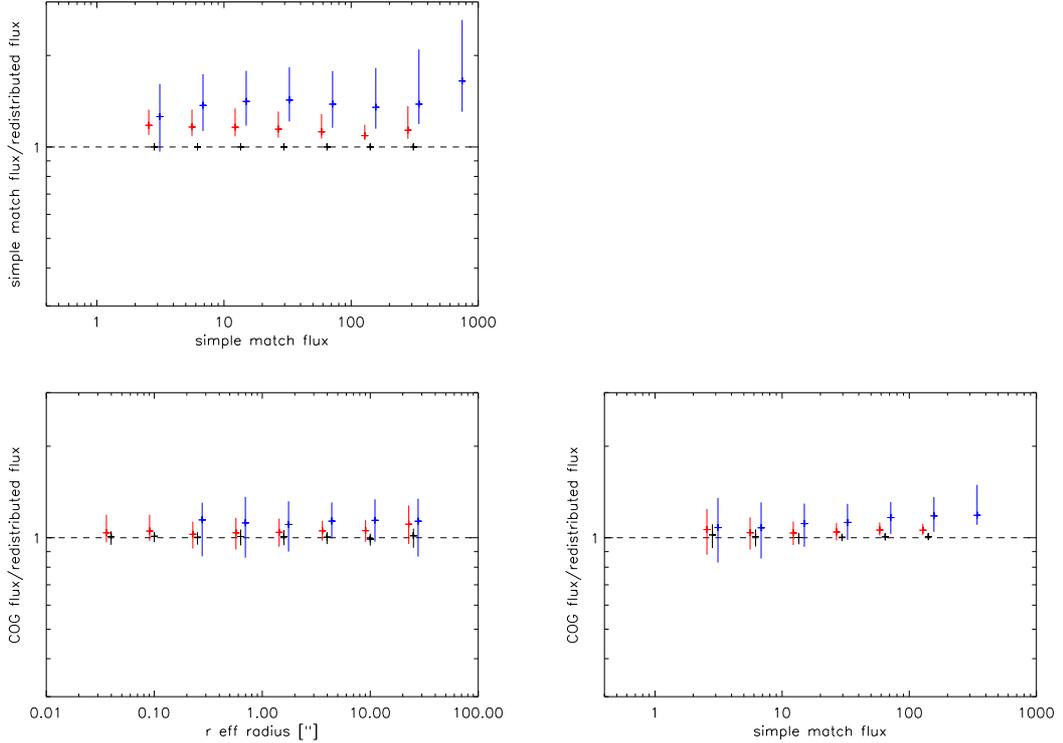


Figure 3.17: Comparison of the three matching techniques

For 1-to-1 matches (black), low multiplicity matches (2-to-1 or 1-to-2; red) and high multiplicity matches (blue), the top panel shows the ratio of simple match flux and redistributed advanced match flux, as function of simple match flux (median value indicated by the cross, inner quartile ranges by the error bar). Below, the ratio of the COG flux and the redistributed advanced match flux is shown as function of effective r-band radius (left), and simple match flux (right).

While the 1-to-1 matches have the same flux using all three measurement techniques, the flux of sources involved in multiple matches are generally strongly overpredicted by the simple match technique. Even the advanced match technique still tends to significantly overpredict the flux of objects involved in high multiple matches.

magnitudes is clearly biased towards the blue. Statistically speaking, the simple match technique is assigning too much flux to multiple match galaxies and the advanced match technique is able to restore the actual flux of the galaxy<sup>23</sup>. Fig. 3.17 top panel illustrates the same feature. The ratio of simple match flux and redistributed flux from the advanced match is unity for 1-to-1 matches (black). Low multiplicity matches (red) and especially high multiplicity matches (blue) show the overprediction of the simple match flux. As the advanced matching technique does not only distribute the UV flux of a single UV source among several optical sources, but may also assign (parts of) the UV flux of several UV sources to optical sources, the ratio of simple match and advanced match flux can become less than unity.

Comparing the COG flux and the redistributed flux in Fig. 3.17, bottom row,

<sup>23</sup>This comparison only claims a correct flux reconstruction for a statistical sample of galaxies. For individual galaxies, the accuracy of the redistributed flux depends mainly on the agreement of the assumed and real flux geometry.

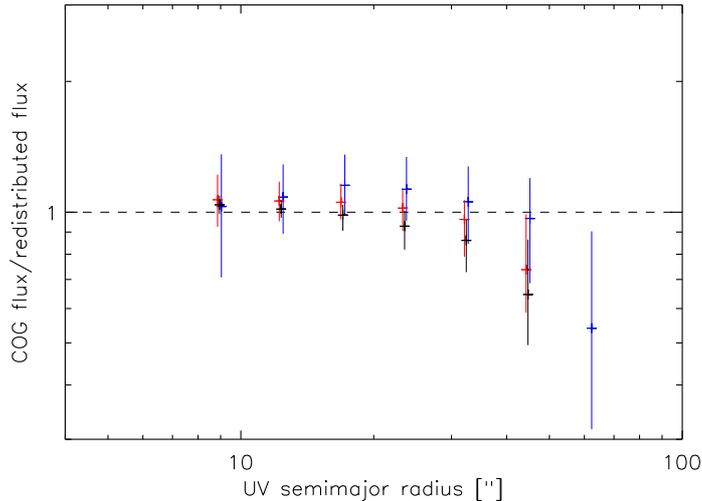


Figure 3.18: Ratio of COG and redistributed NUV flux vs. UV size

The ratio of COG and redistributed NUV flux is roughly constant for sources with an NUV semimajor axis smaller than 20 arcsec. For sources with larger radii, the GALEX pipeline measurement becomes unreliable. Same colour code as in Fig. 3.17.

we see a similar trend. For 1-to-1 matches, both fluxes are roughly identical. For multiple matches, the COG flux is statistically higher than the redistributed flux. In the example scenario of Fig. 3.8, the advanced match technique will redistribute the flux of the single UV source among three optical sources. As all three optical centres are roughly equally far away from the UV centre, the  $1/r$  weighting will result in assigning roughly  $1/3$  of the UV flux to each optical source, even though the majority of the flux most likely came from the galaxy and not the two stars. The COG technique, however, masks the area of the known neighbouring sources, and re-measures the flux of the galaxy alone. The COG flux is superior to the redistributed flux for multiple matches, as it does not rely on assumptions about how the flux of the target galaxy and of the neighbouring sources are geometrically distributed.

Fig. 3.18 shows the ratio of the COG flux and the redistributed flux as function of UV size. The GALEX pipeline has problems to properly measure (UV) large objects, which manifests itself in the overestimation of the redistributed flux for large UV pipeline semimajor axes. If UV sources with pipeline semimajor axis larger than 20 arcsec are involved in the matching, we consider the COG flux superior to the redistributed flux, irrespective of the multiplicity of the match.

Fig. 3.19 compares the size estimates from the optical with the different measures in the UV for spiral galaxies (blue) and a bright subsample (`petromag_r < 16`; red). The top left panel shows the very good agreement of the size estimates of the half-light r-band radius taken from the SersicCatv09 and the half-light radius from COG for bright spirals, and it shows a subset of galaxies which are significantly larger in the UV than in the optical. The top right panel shows the comparison of the COG critical radius (equivalent to the semimajor axis) and the pipeline nearest neighbour semimajor axis. The scatter between these two semimajor axis estimates is very large. A loose correlation can be seen for the bright

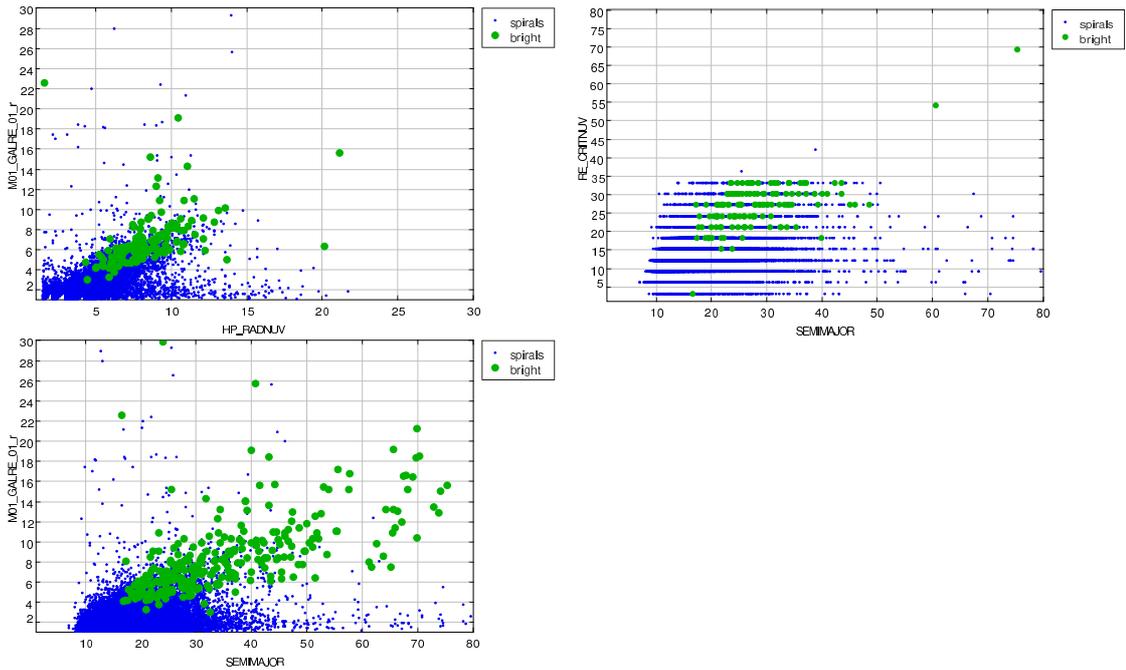


Figure 3.19: Comparison of optical with UV sizes

Top left: Comparison of the effective r-band radius (half-light radius) from the Sersic-Catv09 and the COG half-light radius for all spiral galaxies (blue) and for bright spirals only ( $\text{petromag\_r} < 16$ ; green). Bottom: Comparison of the effective r-band radius and the UV nearest neighbour semimajor axis. While the COG radius and the optical radius correlate very well, the UV semimajor axis (expected to be of the order of 3 times larger since it includes all flux) correlates less well and seems to be even larger than expected. Top right: Comparison of the COG critical radius (enclosing all flux) and the UV semimajor axis pipeline estimate. While the bright sample shows a reasonable correlation of the size estimates, the full sample shows large scatter.

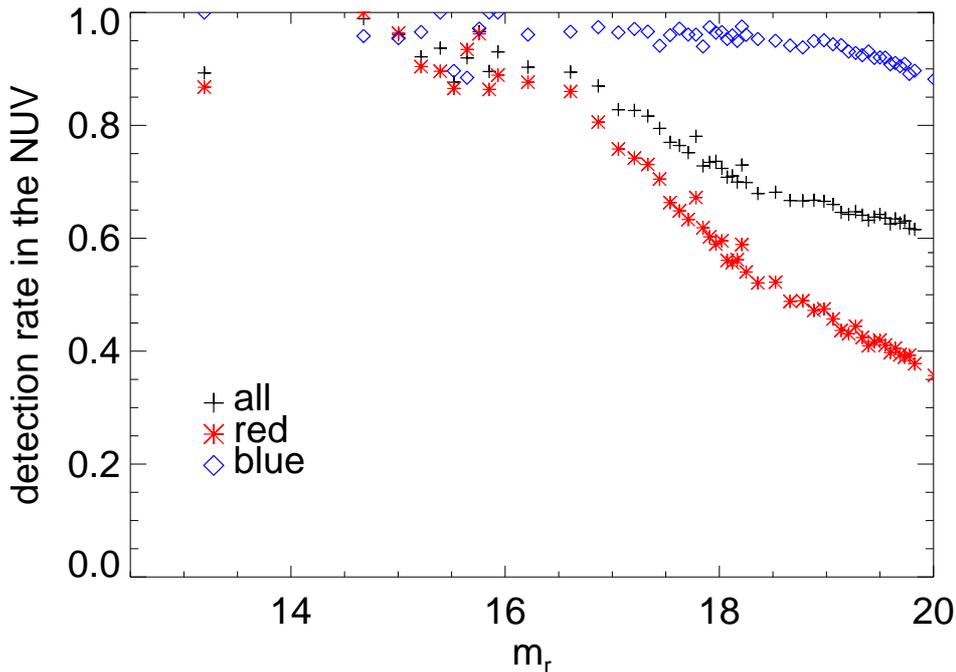


Figure 3.20: Advanced match detection rate at GMC depth

subsample, but in general there appears to be little correlation. Especially the pipeline semimajor axis values larger than 40 arcsec do not have an equally large COG counterpart. As this is also the case for 1-to-1 matches, the reason for the lacking correlation cannot be argued to be the pipeline UV merging together several small optical sources. It rather illustrates the inability of the GALEX pipeline to reliably measure large sources with UV semimajor  $> 20$  arcsec<sup>24</sup>.

### 3.6.2 Matching completeness

A real detection rate for the GAMA galaxies (i.e. the fraction of actual detections and all galaxies that should have been detected) can only be determined using a superior NUV catalogue containing a complete NUV source list down to the detection limit as “truth”. Here, we assume the r-band based TilingCatv41 to be complete, i.e. we assume that it contains all galaxies (brighter than the r-band limit of 19.8mag). However, due to the intrinsic colour of galaxies, it is possible that a galaxy is listed in the TilingCatv41, but has a UV magnitude fainter than the detection limit. This constellation will result in a non-detection in the UV. Moreover, it is correct that this galaxy is not detected in the UV and it is not a problem of completeness that it is missing (since the completeness is always defined as complete up to the detection limit).

Fig. 3.20 shows the detection rate for galaxies in the NUV as function of the apparent r-band magnitude (i.e. matching rate of the NUV-based GALEX-GAMA advanced match catalogue with the TilingCatv41 for galaxies with NUV exposure and without window or dichroic artefact flags). Blue and red galaxies (defined

<sup>24</sup>See also GALEX observer’s guide.

following Baldry et al., 2004) show very different behaviour due to their intrinsic colour. The blue galaxies (blue diamonds) are almost complete over the full considered r-band magnitude range, their completeness is only slightly dropping beyond 18 mag. However, even for bright galaxies, the completeness remains at  $\sim 98\%$ . The red galaxies (red asterisks), as is to be expected, already starts to become incomplete for bright r-band magnitudes, at  $\sim 15$  mag. The full galaxy sample (black crosses), being the combination of the red and blue galaxy sample, shows a completeness between the two samples.

The incompleteness at bright r-band magnitudes is a potential cause for biases in the analysis of the UV flux of GAMA galaxies as it falsely suggests 5% of the r-band bright galaxies to have a UV flux below the detection limit. The COG measurements of the GAMA galaxies, however, are by definition carried out for *all* galaxies in the TilingCatv41 and the COG catalogue reports the re-measured flux at the optical position disregarding of a UV detection limit. Missing COG measurements are not due to too faint fluxes, but only due to masking in the central regions (in the order of 10% of the cases), which is independent of galaxy properties and will thus not introduce a bias in the analysis of the UV flux of GAMA galaxies (even though it is more likely to happen in crowded fields like, for example, galaxy clusters).

### 3.6.3 “Best-mag” catalogue

The matched catalogues above described provide flux measurements derived with different techniques. Each technique has its own advantages and disadvantages, thus each is preferable in different situations. We therefore provide a ready-to-use best-mag catalogue, which combines the strengths of the individual matching techniques and employs the most suitable one for each object individually<sup>25</sup>.

Besides providing an excerpt of the most useful columns of each of the above mentioned catalogues, the best-mag catalogue provides the best flux (BEST\_FLUX\_XXX, BEST\_MAG\_XXX, ...) estimate for a GAMA object, decided individually for each object as outlined below.

As discussed in Sect. 3.6.1, the COG flux is superior to the redistributed flux for multiple matches and for large sources, as it masks neighbouring sources rather than relying on assumptions made for the flux redistribution. Furthermore, for 1-to-1 matches, both techniques return similar flux values. We therefore quote the COG flux as “best” if available. In cases where COG photometry is not available (e.g. due to close-to-center masking or non-galaxy sources), we resort to the advanced match fluxes. In the case of COG non-detections (due to small flux or COG galaxy area smaller than the PSF; not non-existing measurements), we substitute the integrated flux inside of the optical aperture for the “best” flux<sup>26</sup>.

For easy comparison, we also quote the flux measurements of all techniques in this catalogue, making this the UV catalogue of choice for general applications. See Sect. D.5 for individual column descriptions.

---

<sup>25</sup>The user can decide based on the task at hand whether a consistently derived UV photometry is crucial (then the photometry of a single technique should be used) or whether the photometry considered best on an object-to-object base is more beneficial.

<sup>26</sup>We emphasise that we refer to the total flux on the maps within the optical aperture, *not* the total flux of all UV blind catalogue sources within the optical aperture.

## 3.7 Ancillary Information

### 3.7.1 Observation information catalogue (“ObsInfoCat”)

When analysing galaxy properties, not only the flux of a detected galaxy is of interest. A certain galaxy may not be listed in either the blind or the matched catalogue. This can be because of three reasons. First, it truly has a flux so low that it did not meet the signal-to-noise cut. Then, the flux corresponding to the S/N cut can be used as upper limit. Second, the galaxy’s position on the sky was not observed. In this case, we cannot say anything about that galaxy and need to remove it from the UV sample. Third, the galaxy has flux above the S/N cut, but was not detected due to artefacts nearby. This case is the most error-prone one, as it might even lead to false flux reports in the catalogue for galaxies that were detected.

In order to distinguish between these three cases, we list exposure time, background and flag value extracted from the tile-level<sup>27</sup> exposure time, background level and artefact flag maps, respectively, alongside with the upper limits (calculated from the exposure time and background) at the position of each GAMA source for all three coverage depths. This enables the user to check whether a source was observed or not and whether the listed flux is potentially biased by artefacts. See appendix D.6 for individual column descriptions.

### 3.7.2 HEALPIX maps of the GAMA fields

The HEALPIX maps provide useful meta information for each GAMA field in all three GALEX depths. With their help, the user can obtain the desired information exactly at the position of interest, even if there is no catalogued GALEX source. The resolution of the HEALPIX maps is nested scheme  $n_{\text{side}} = 2^{14}$ , which translates roughly to  $12.9 \times 12.9$  arcsec<sup>2</sup> pixel size. Fig. 3.21 shows all HEALPIX maps for the example field G09 in the NUV.

#### Intensity maps

The HEALPIX intensity maps show the background subtracted emission for the GAMA fields. Visual inspection of these maps can bring useful insight for crowded fields or overlapping sources.

#### Flag maps

The HEALPIX flag maps merge the tile-level flag maps. They allow the user to extract the flag value of a given optical coordinate, thus enabling the user to identify whether an object of interest is potentially corrupted.

In addition, we provide HEALPIX manual flags maps, which flag regions which contain previously unflagged artefacts identified during visual inspection. The artefacts are predominantly large edge reflection from bright, nearby stars which are significantly larger than the pipeline flag.

---

<sup>27</sup>While it is slightly easier to extract this information from the HEALPIX maps for each field (which already incorporate the selection against secondary regions) the resolution of the tile-level maps are significantly better and therefore preferable.

### Background maps

The HEALPIX background maps show the background distribution in the GAMA fields. As the “observed” background consists of both, the “real” background (i.e. large-scale emission caused by UV reflection of Milky Way dust and zodiacal light) and large-scale edge reflection artefacts caused by bright nearby stars, the map looks somewhat patchy, but it still shows the filament structure of the Milky Way dust.

### Exposure time maps

The tile-level exposure time maps are produced with 6 arcsec resolution in detector coordinates and then remapped in 1.5 arcsec resolution to ra-dec coordinates (see GALEX Technical Documentation). These are then mosaiced together to produce a GAMA field overview of the exposure time. Together with the HEALPIX background map, this map enables the user to, e.g., calculate the upper limit for a non-detected object.

### Coverage fraction maps

This rather simplistic HEALPIX coverage fraction map provides information about the coverage of the GAMA field in the two bands NUV and FUV, and can have continuous values from 0 for no coverage to 1 for full coverage. To some extent redundant with the HEALPIX exposure time map, it allows the user to extract the information whether or not an object was covered in the NUV or FUV. This map is particularly useful for objects on the edge of the coverage area, which might be within a HEALPIX pixel that has only partial coverage. While the HEALPIX exposure time map will report a single exposure time for the full HEALPIX pixel, a coverage fraction value  $<1$  but  $>0$  will tell the user that the HEALPIX pixel only had partial coverage.

## 3.8 UV detection limit

Detecting bright sources is always easier than the detection of faint sources. The detection limit of the GALEX UV survey is defined such that the faintest detectable apparent magnitude  $m_{lim,UV}$  has a signal-to-noise ratio of 2.5.

As GALEX is observing the sky in small, circular tiles with different exposure times, there is no homogeneous exposure time to the GAMA dataset (see Sect. 6.2.3.1). This affects the detection limit such that regions with higher exposure time are able to detect fainter sources. Furthermore, there are variations in the UV background levels, such that regions with large background flux will be slightly less sensitive to faint objects. One easy way to deal with this bias is to limit the analysis to sources brighter than the brightest detection limit, ignoring fainter data. However, since we are interested in using as much data as possible, it is important to use the individual exposure times and background levels, resulting in individual detection limits for each considered position on the sky (i.e. for each individual galaxy).

The Gaussian error of a flux measurement  $s$  was defined in Eq. 3.5. Since in this section, we are only considering faint sources equal to or fainter than the detection limit, we can safely assume  $0.01 \cdot s < bg$ , such that we can neglect the flat field

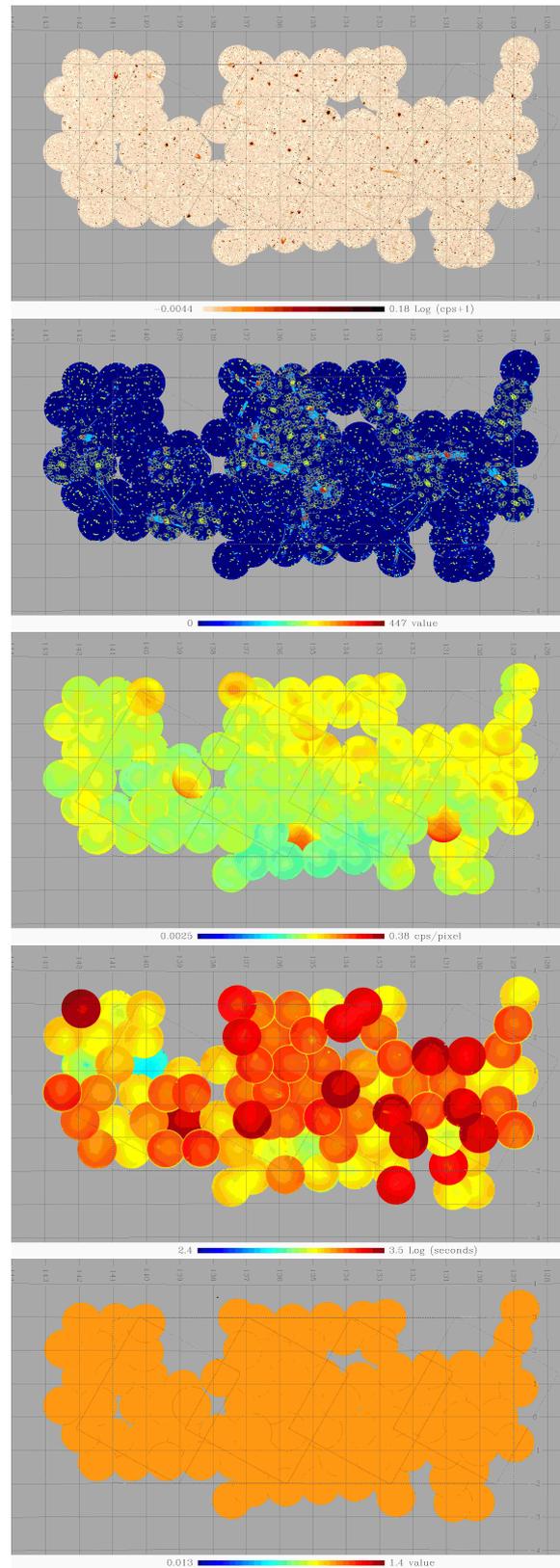


Figure 3.21: HEALPIX maps for the example of the G09 field in the NUV, GMC depth. Top to bottom: HEALPIX maps of background subtracted intensity, flags, background, exposure time, and coverage fraction of the HEALPIX pixel for the G09 field in the NUV. Especially in the exposure time map, one can clearly see the individual circular tiles which shape the GALEX-GAMA footprint.

error. Thus, the Gaussian error is defined as  $ds$

$$ds = \frac{\sqrt{(s + 1.5 \cdot bg \cdot a)t_e}}{t_e}$$

where  $t_e$  is the exposure time in seconds,  $a$  is the area of the source on the map in arcsec<sup>2</sup>, and  $bg$  is the local background level in counts per second per arcsec<sup>2</sup>.

The signal-to-noise ratio  $S/N$  is therefore

$$S/N = \frac{s \cdot t_e}{\sqrt{(s + 1.5 \cdot bg \cdot a)t_e}} = \frac{s \cdot \sqrt{t_e}}{\sqrt{s + 1.5 \cdot bg \cdot a}} \quad (3.7)$$

The flux  $s$  which has a  $S/N$  identical to the signal-to-noise cut chosen for the data set represents the detection limit  $s_{lim}$ . We can derive the detection limit for the GALEX pipeline  $s_{lim}$  by (setting  $S/N=2.5$  and) solving Eq. 3.7 for  $s_{lim}$ .

$$\begin{aligned} S/N^2 &= \frac{s_{lim}^2 \cdot t_e}{s_{lim} + 1.5 \cdot bg \cdot a} \\ s_{lim}^2 \cdot t_e &= s_{lim} \cdot S/N^2 + 1.5 \cdot bg \cdot a \cdot S/N^2 \\ t_e \cdot s_{lim}^2 - S/N^2 \cdot s_{lim} - 1.5 \cdot bg \cdot a \cdot S/N^2 &= 0 \end{aligned}$$

Bearing in mind that all variables in the equation above cannot be negative, we can obtain a unique solution of the equation above using the quadratic formula.

$$s_{lim} = \frac{S/N^2 + \sqrt{S/N^4 + 4 \cdot t_e \cdot 1.5 \cdot bg \cdot a \cdot (S/N)^2}}{2 \cdot t_e} \quad (3.8)$$

The GALEX-GAMA catalogues use a  $S/N$  cut of 2.5 and we derive detection limits for point sources (i.e. circular area  $a = \pi \cdot 3.4^2$  arcsec<sup>2</sup>, the area of the point-spread function, no redshift dependence of the size) at each position of GAMA sources.

Using the NUV zero point magnitude  $ZP = 20.08$ , the flux limit  $s_{lim}$  can be converted into the magnitude limit  $m_{lim}$  with

$$m_{lim} = ZP - 2.5 \cdot \log_{10}(s_{lim}) \quad (3.9)$$

For further improvement, the model can incorporate a physical size distribution of the galaxy population, which is then transformed to apparent size and can be used as area  $a$  in Eq. 3.8. It should be noted that this improvement would introduce a redshift dependency of the noise and the detection limit.

### 3.9 Unmatched GALEX sources

The simple match and the advanced match processes both start from an optical input catalogue and search for UV sources within the optical galaxy area. In theory, it is possible that there are UV sources which should have an optical match, but do not since the UV central coordinates do not lie within the area of an optical galaxy. This can easily happen for extended sources which are treated as point sources by the advanced matching due to missing size information.

As a sanity check, we start the matching from the UV side and search for extended, bright (i.e.  $S/N \geq 10$ ) UV sources that do not have an optical counterpart. Since the UV pipeline size is always known, a scenario equivalent to the one described above does not apply here. We then eyeball these unmatched UV sources in order to either identify true UV-only sources, which are worth further investigation, or to identify a potential of improvement of the advanced matching routine<sup>28</sup>.

We found that unmatched UV objects with high S/N are predominantly present in regions with long exposure times or low background. Therefore, these objects actually are not UV bright, they only have higher S/N values than the majority of the objects with comparable brightness. In other words, even for typical NUV-r colours, their optical counterpart did not meet the r-band magnitude cut for entering the InputCatAv06.

---

<sup>28</sup>As the GALEX blind catalogues deliberately extends slightly beyond the GAMA footprint, one needs to be aware that regions with GALEX coverage which lie outside of the region covered by the InputCatAv06 will of course lead to unmatched UV sources. Furthermore, there are altogether 4 small regions (order of 0.2 degree x 0.2 degree) inside of the GAMA equatorial fields, which have no objects listed in the InputCatAv06, i.e. there are small “holes” in the InputCatAv06. These regions will lead to unmatched UV sources as well.



## Part II

# Attenuation of starlight by dust in spiral galaxies



# Chapter 4

## Fundamental considerations

In the second part of this work, we are interested in the question: "What would the photometry of galaxies be if it had not been affected by dust attenuation<sup>1</sup>?" In order to answer this question, we need to answer a similar question first: "How much dust is in a galaxy, and how is it distributed with respect to stars?"

As elliptical galaxies are believed to be largely dust free, we restrict our analysis to spiral galaxies (see Sect. 5.2.1 for the derivation of the GAMA spiral galaxy subsample). Furthermore, we wish to avoid complications which might arise from galaxy evolution and therefore restrict the analysis to galaxies in the local universe (i.e. redshift  $\leq 0.13$ ). This is also a necessary practical constraint, as reliable measurements of galaxy morphology are not yet available for more distant GAMA galaxies<sup>2</sup>.

### Choice of NUV for analysis of dust attenuation

It has long been known that dust attenuation is a function of wavelength (e.g. O'Donnell, 1994). The shorter the wavelength, the more light is absorbed and/or scattered by dust grains and thus the galaxy's light is more dimmed. Consequently, if one wants to study the effect of dust attenuation, the expected signal is strongest for rather short wavelengths. 912 Å presents a fundamental limit as shortwards of this wavelength the photons are energetic enough to ionise hydrogen. Consequently, the attenuation shortwards 912 Å is dominated by neutral gas rather than by dust. The GALEX NUV band represents a good compromise in choice of band to study dust attenuation, as it is strongly subjected to dust attenuation while still directly tracing young, blue, luminous stars which are quite tightly correlated to readily observable global properties of galaxies at the present epoch such as stellar mass. In principle, the GALEX FUV band covers an even shorter wavelength range and thus is even more affected by dust attenuation. However, the combined FUV emission of all stars in a galaxy with a constant SF rate, while being comparable to that in the NUV, is more strongly affected by the high mass

---

<sup>1</sup>Throughout we use the term "attenuation" to denote the reduction in observed star light from galaxies due to dust inside the galaxies. This should not be confused with the term extinction, which refers to the reduction in flux of point sources due to dust between the source and the observer. Whereas attenuation depends both on the geometrical distribution of stars and dust in a galaxy, as well as on optical properties of the grains, extinction is solely dependent on the optical properties of the grains.

<sup>2</sup>This will however be rectified when the ongoing VST survey of the GAMA fields is complete

cut off of the initial mass function (IMF; e.g. Salpeter (1955); Kennicutt (1983); Kroupa et al. (1993) or Chabrier (2003) slightly differ in the high mass cut off). This would constitute an additional uncertainty in the analysis, since we do not know a priori whether the IMF systematically changes with galaxy properties such as mass or SFR. One may also anticipate a higher intrinsic scatter in luminosities of galaxies measured at the very shortest UV wavelengths due to fluctuations in SFR with time. We therefore favour the NUV over the FUV data to understand the effects of dust on the UV light.

Another fundamental consideration in extracting information of attenuation of dust from the GALEX-GAMA data is the fact that the GAMA survey is not flux-limited in the NUV, but in the optical r-band. This poses a complication, since the r-band fluxes are themselves affected by dust attenuation, very likely to a different degree as the UV fluxes. In addition, we need to consider that the attenuation may vary systematically with UV and/or r-band luminosity, so we need to account for the relative number of faint and bright galaxies in the flux-limited sample. To this end, we cannot simply employ an NUV luminosity function for our study, as the sample would be missing those galaxies that would be detectable in the NUV but did not pass the r-band detection limit (i.e. distant low mass blue galaxies). We therefore need to consider the effect of dust on the r-band in order to account for selection biases.

The method we adopt to address this issue is to analyse a conditional probability, i.e. calculate the probability of the NUV measurement, given the r-band detection for a given hypothetical attenuation model  $P(m_{NUV,obs}|m_{r,obs}, \phi)$ , where  $\phi$  denotes the attenuation model.

This is equivalent to the probability of an observed colour  $c_{obs} = m_{NUV,obs} - m_{r,obs}$ ,  $P(c_{obs}|m_{r,obs}, \phi)$ . It is then formally straight forward to express P in terms of the probability  $P(m_{r,intr}|m_{r,obs}, \phi)$  that a galaxy has an intrinsic<sup>3</sup> r-band magnitude  $m_{r,intr}$ , given the observed magnitude  $m_{r,obs}$  and the hypothesized dust model  $\phi$ :

$$P(c_{obs}|m_{r,obs}, \phi) \propto \int_{m_{r,intr}} P(c_{obs}|m_{r,intr}, \phi, \xi) \cdot P(m_{r,intr}|m_{r,obs}, \phi) dm_{r,intr} \quad (4.1)$$

Here  $P(c_{obs}|m_{r,intr}, \phi, \xi)$  is the probability that a galaxy has an observed colour  $c_{obs}$  given an intrinsic r-band magnitude  $m_{r,intr}$ , for the dust model  $\phi$  and a hypothesized model  $\xi$  for the distribution in intrinsic colour of a galaxy  $c_{intr} = m_{NUV,intr} - m_{r,intr}$ . Since the transformation between  $c_{intr}$  and  $c_{obs}$  is a function only of  $\phi$ ,  $P(c_{obs}|m_{r,intr}, \phi, \xi)$  can be expressed as:

$$P(c_{obs}|m_{r,intr}, \phi, \xi) \propto \int_{c_{intr}} P(c_{obs}|c_{intr}, \phi) \cdot P(c_{intr}|m_{r,intr}, \xi) dc_{intr} \quad (4.2)$$

Lastly, motivated both by physical and empirical considerations, one can seek to express both the model for the intrinsic colour  $\xi$  and the dust model  $\phi$  as a

---

<sup>3</sup>Here, the term "intrinsic" is used to denote a property of a galaxy that would be observed in the absence of dust.

function of the readily observable and robustly inferable stellar mass  $M_s$ <sup>4</sup>.

$$P(c_{obs}|m_{r,intr}, \phi, \xi, M_s) \propto \int_{c_{intr}} P(c_{obs}|c_{intr}, \phi(M_s)) \cdot P(c_{intr}|m_{r,intr}, \xi(M_s)) dc_{intr} \quad (4.3)$$

In Eq. 4.3, it can be seen that  $\xi$  and  $\phi$  appear in separate terms. Thus, the effect of dust attenuation on colour predicted by  $\phi(M_s)$  is completely degenerate with the form of the distribution of intrinsic colours predicted according to  $\xi(M_s)$  (since we have to assume we have a priori no theory for  $\xi(M_s)$  and  $\phi(M_s)$ ). This is the well known colour-reddening degeneracy for galaxies which prevents direct inference of intrinsic colours of dusty galaxies. This degeneracy can, however, be potentially broken if the galaxies are disk-like, and we can measure the inclination of the disk, since whereas the intrinsic colour is independent of inclination, the reddening by dust is affected by the viewing angle.

The second part of this thesis is devoted to developing an algorithm that can exploit information on inclination derived from the shape measurements of galaxies to break the colour-reddening degeneracy and test simple hypotheses  $\xi(M_s)$  for the form of the probability distribution in intrinsic colour of the population of disk galaxies  $P(c_{intr}|m_{r,intr}, \xi(M_s))$  for different hypotheses  $\phi(M_s)$  for the dust attenuation  $P(c_{obs}|c_{intr}, \phi(M_s))$ .

In addition, we will investigate whether it is possible to extract information about dust attenuation by making constraints on  $P(c_{intr}|m_{r,intr}, \xi(M_s))$ , specifically that the intrinsic distribution in colour should be very narrow for disk galaxies.

Optimizing the parameter combination  $\phi(M_s)$  and  $\xi(M_s)$  via a maximum likelihood fit, we will seek a description for the intrinsic colour as function of stellar mass as well as the dust distribution. This approach is not only capable of deriving the best-fitting parameters for a given dust model, but is also capable of intercomparing different dust models.

The GALEX-GAMA survey is an ideal basis for this work, because of the well understood and complete statistical content and depth of the GAMA survey, the availability of sizes and ellipticities for all GAMA galaxies, and the deep and almost complete coverage in NUV provided by the GALEX observations, as done in part I.

In this work we will focus on the term  $P(c_{obs}|m_{r,intr}, \phi, \xi)$  in Eq. 4.1. The second term  $P(m_{r,intr}|m_{r,obs}, \phi)$ , which describes the selection bias introduced by the r-band detection limit for a given dust model is not considered. Its evaluation in principle requires an analysis of the r-band luminosity function for each dust model and each parameter set of the dust model (see outlook and appendix C).

## Simulation of galaxy population

The GALEX-GAMA data set, as described in detail in Part II of this work, is part of the r-band based GAMA survey. Galaxies with foreground dust attenuation

---

<sup>4</sup>Another advantage of expressing hypotheses for the dust model and colour in terms of stellar mass is that, as explained in Sect. 5.3.1, stellar mass can to a good accuracy be extracted from long wavelength broadband optical photometry with relatively small sensibility to dust attenuation.

corrected observed magnitudes  $< 19.8\text{mag}$  are selected to be part of the survey. This selection criterion is unfortunately subject to biases.

The redshift of a galaxy to the observer does not only result in a dimming of the galaxy's light due to the distance, but also causes a frequency shift towards the red. This effect can be corrected for using so-called k-corrections (see Sect. 5.1.2.3), but unfortunately these corrections are not included in the r-band detection limit definition.

Another selection bias is the so-called inclination bias. Even though the r-band is less affected by dust attenuation than the NUV, galaxies with intrinsic r-band luminosity close to the r-band detection limit can have their light attenuated enough to drop below the detection limit. While individual dust attenuation models differ in their assumed dust distribution from single-sheet to complex geometries, in the assumed grain structure and heating mechanisms, the attenuation is generally a function of dust content and inclination. In the presence of a detection limit (or a lower bin boundary), galaxies with high dust content and/or high inclination will appear fainter and will be underrepresented in the detected data set (or in the considered bin). The inclination bias (discussed in detail in Sect. 6.3.3.3) itself is a function of the luminosity function in the r-band.

In order to investigate the selection biases of the r-band defined GAMA survey as well as to verify our analysis technique, we therefore simulate data sets of spiral galaxies, including a r-band luminosity function derived from the GAMA data itself (see Appendix A), dust attenuation, k-corrections and foreground dust attenuation, and the r-band detection limit. Furthermore, we simulate the UV properties of the galaxy population via an intrinsic NUV-r colour model, including NUV detection noise<sup>5</sup> (as laid out in detail in Sect. 6). We also simulate the bias on the inclination measurement caused by the intrinsic thickness of the galaxy's disk (see Sect. 7.2), which prevents galaxies to have infinite thinness even if viewed perfectly edge-on.

Having obtained an understanding of the selection biases, we can account for them in our analysis and proceed to the actual aim of this work, the derivation of dust attenuation parameter.

This part of this work is outlined as follows: Chapter 5 presents all data, constraints and the dust model needed for the analysis of the intrinsic colour. In Chapter 6 we present our detailed simulation of spiral galaxy data set and furthermore discuss the fundamental concepts and functional forms used for the simulation and later in the analysis. In Chapter 7, the actual analysis of the dust attenuation via the intrinsic colour is developed, tested and applied to the GAMA spiral galaxy data set. We give a summary of our work and an outlook in Chapter 8.

---

<sup>5</sup>The NUV noise is much larger than the r-band noise, which can be neglected.

# Chapter 5

## GAMA input data & constraints

In order to analyse the dust properties of local spiral GAMA galaxies as outlined before, we need more information than just the r-band and UV fluxes, which we presented in Sect. 2.1.1 and Chapter 3, respectively. In this chapter, we will present all additional data needed in Sect. 5.1 and define the final data set to be used in Sect. 5.2. We also present the concepts used to derive quantities from the available data which are needed for the analysis (Sect 5.3) and introduce the different hypotheses for the dust model (Sect 5.4).

### 5.1 Basic data

Along with the tiling catalogue presented in Sect. 2.1.1 and the `best_mag_NUV` from the GALEX-best catalogue (Sect. 3.6.3), we need more information about the GAMA galaxies. This section lists all the auxiliary information we need to extract from the GAMA catalogues in this work.

#### 5.1.1 Distances

##### 5.1.1.1 Redshift and local-flow correction

Distances are derived from local-flow corrected redshifts. The GAMA collaboration provides spectroscopic redshift measurements for all GAMA galaxies. These redshifts are derived from frequency shifts of spectral lines. Especially in the local universe, the infall velocity into a galaxy cluster contributes significantly to the frequency shift of a galaxy’s spectral lines. Thus the measured redshifts are biased and need to be corrected. Baldry et al. (2012) derive local-flow corrected redshifts for the GAMA galaxies using the flow model of Tonry et al. (2000), consisting of Hubble flow, peculiar motion, two attractors, one of them consistent with the Virgo Cluster, and a dipole and a quadrupole correction to the Hubble flow.

In terms of this analysis, we will treat these local-flow corrected redshifts as the “true” redshifts of the galaxies and use them to derive the distances of the galaxies according to Eq. 6.8.

In order to avoid the necessity of considering the effects of galaxy evolution, we restrict our analysis to the local universe. We choose a redshift limit of  $z \leq 0.13$ , which is identical with the redshift limit for which Grootes et al. (2014) derived their spiral selection method (see Sect. 5.2.1). We apply this limit to the local-flow corrected redshifts.

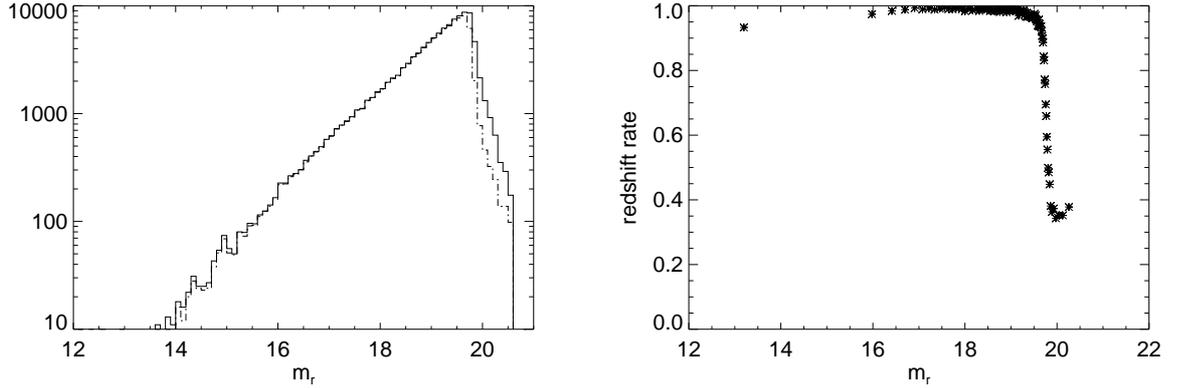


Figure 5.1: Redshift completeness

The left plot shows the distribution of the r-band magnitudes of all galaxies in GAMA in the equatorial fields (solid line) and those with redshift (dashed-dotted line) as function of the r-band. The redshift completeness is shown on the right. The main region of incompleteness is at r-band magnitudes  $> 19.6$  mag.

### 5.1.1.2 Redshift completeness

Figure 5.1, left shows the distribution of all galaxies in the equatorial fields (solid line) and those with redshift (dashed-dotted line) in the r-band (Petrosian magnitudes). The main region of incompleteness in the r-band is for magnitudes  $> 19.6$  mag. Figure 5.1, right shows the redshift completeness, i.e. the fraction of objects with redshift and all objects, as a function of the r-band magnitude. While the r-band redshift rate is almost unity ( $\sim 97\%$ ) out to a magnitude of  $\sim 19$ , it drops off rapidly for fainter sources.

## 5.1.2 Optical photometry

In this section, we specify the optical photometry used in this work and describe the necessary correction of the photometry for both the optical and the NUV.

### 5.1.2.1 r-band photometry

Objects (both stars and galaxies) in the SDSS DR7 catalogues which have a foreground extinction (see Sect. 5.1.2.2) corrected r-band Petrosian magnitude  $r_{Petro,0} < 20$  mag and lie within the GAMA equatorial fields are listed in the input catalogue InputCatAv06 (Baldry et al., 2010, see Sect. 2.1.1).

Apart from being a pure source list, the InputCatAv06 also provides the SDSS photometry for the GAMA sources. In this work, we use two definitions for the r-band magnitude, depending on context. In the context where sample selection plays a role, we use the SDSS "PETROMAG\_R" as r-band magnitude.

For the derivation of the stellar mass (see Sect. 5.3.1), we use r-band Sérsic magnitudes, derived by Kelvin et al. (2012). The r-band Sérsic magnitudes were derived from single Sérsic fits of SDSS images and are thought to provide a better representation of the contribution to integrated flux of faint emission from outskirts of galaxies.

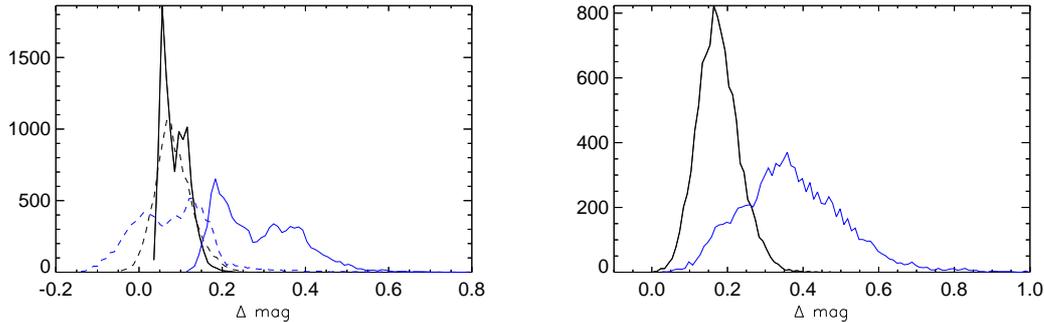


Figure 5.2: Histogram of the A- and k-corrections of the GAMA data set. The left-hand histogram shows the A-correction (solid line) and the k-correction (dashed line) for the r-band (black) and the NUV (blue) for the spiral galaxies of the GAMA data set (no redshift cut). The right-hand histogram shows the sum of both corrections (same colour code).

### 5.1.2.2 Foreground extinction – A-correction

The light of an extragalactic source must travel through our Galaxy before it reaches the telescope. On its way, it is subject to foreground extinction due to the dust present in our Galaxy. Since the amount of dust in the line of sight varies, the foreground extinction is not only a function of wavelength, but also a function of the celestial coordinates.

Baldry et al. (2010) provide in the catalogue “GalacticExtinctionV02” foreground extinction values (so-called A-correction) for the GAMA galaxies in all GAMA bands. They used the  $E(B-V)$  values taken from the dust maps of Schlegel et al. (1998) in combination with the extinction law of O’Donnell (1994) in the optical. In the UV, they follow Wyder et al. (2005), who provide a formula for the UV extinction values as function of  $E(B-V)$  values, using the extinction law of Cardelli et al. (1989).

The observed magnitudes  $m_{obs}$  of the galaxies can be corrected for the foreground extinction  $A_{corr}$  using these values, thus restoring the apparent magnitude attenuated only by the dust present in the considered galaxy itself.

Note that the r-band cut (at the “detection limit”  $m_{detlim}$ ) applied to the GAMA data of the input and tiling catalogue is based on the  $A_{corr}$ -corrected data, i.e.

$$m_{obs} - A_{corr} \leq m_{detlim} \quad (5.1)$$

The distribution of  $A_{corr}$  is shown in Fig. 5.2 for the r-band (black) and the NUV (blue) as solid line. Fig. 5.3 left shows the dependency of r-band  $A_{corr}$  with the RA coordinate. This RA dependence is interesting, but not important, since the detection limit was applied to foreground attenuation corrected data.

### 5.1.2.3 Rest frame magnitude – k-correction

Observing galaxies in a given wavelength range, one needs to be aware of and correct for the fact that the wavelength range in which the light is observed is not identical with the wavelength range in the rest frames of the galaxies in which the

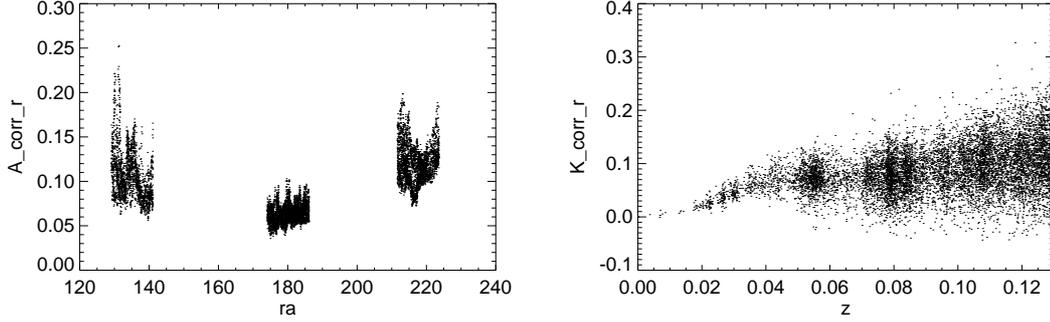


Figure 5.3: A-correction vs. RA and k-correction vs. redshift

light was emitted. It is possible to employ galaxy spectra templates to calculate the galaxies’ emission in the wavelength range of the filter.

The correction for this redshift dependent frequency shift in the observed band (so-called “k-correction”, Humason et al., 1956) for all GAMA galaxies is provided by Loveday et al. (2012) for the UV and the optical bands in the catalogue “kCorrectionsV03”. The k-corrections were calculated with `kcorrect v4_2` (Blanton & Roweis, 2007) using SDSS DR7 model magnitudes. Using these k-corrections, we can correct the data for the redshift dependent frequency shift in the observed band.

The distribution of  $k_{corr}$  is shown in Fig. 5.2 for the r-band (black) and the NUV (blue) as dashed line. Fig. 5.3 right shows the dependency of r-band  $k_{corr}$  with redshift. This redshift dependence is very small over the considered redshift range. Beyond  $z=0.04$ , the k-correction distribution is dominated by intrinsic scatter.

### 5.1.3 NUV photometry

The GALEX-GAMA data products (see Sect. 3) provide the NUV photometry for the GAMA data. We use the redistributed magnitudes from the “advanced match” catalogue (see Sect. 3.4) for the NUV photometry. In principle, COG photometry is superior (see Sect. 3.5). However, the COG photometry was finalised only late during the preparation of this work.

Furthermore, we import the background level and the exposure time at the optical position of each GAMA galaxy from the ObsInfoCat (see Sect. 3.7.1)

## 5.2 Sample selection

Besides the detailed morphological information, it is crucial to be able to distinguish between spiral and spheroidal galaxies. The latter are known to have much lower dust mass than the former (e.g. Smith et al., 2012). As this work addresses attenuation in spiral galaxies, we need to exclude dust-free galaxies from our data sample.

Identifying a GAMA galaxy as a spiral or as an elliptical galaxy is a difficult task, as galaxies are too numerous for selection by human inspection of images. Furthermore, automatic classifications can be ambiguous, with the results differing dependent on the criteria used. This is because there is a broad overlap in the

parameters recovered for spirals and ellipticals by automated morphological fitting routines (e.g. Baldry et al., 2004; Lintott et al., 2011, Taylor et al., 2014, in prep.).

As we want to probe the dust properties of local spiral galaxies, it is nonetheless crucial to have a reliable separation of spiral and elliptical galaxies.

### 5.2.1 Spiral selection technique

Grootes et al. (2014) developed a non-parametric cell-based proxy technique for selecting spiral galaxies using different parameter combinations which can select almost pure samples of disk galaxies. We chose to adopt the following parameter combination:

- single component Sérsic index  $n$
- effective radius  $r_{eff}$  (linear size)
- absolute (attenuated) i-band magnitude  $M_{i,d}$

as this selection was found by Grootes et al. (2014) to be one of the most reliable ones (i.e. best in classifying the training data). Furthermore, the selection is independent of UV flux, stellar mass estimates, or short wavelength optical colours, allowing a morphologically defined sample of galaxies almost unbiased with respect to stellar mass or star formation rate.

Grootes et al. (2014) trained the technique on the morphological classifications of Galaxy Zoo Data Release 1 (DR1; Lintott et al., 2011). For the parameter combination used here, Grootes et al. (2014) report a contamination rate by elliptical galaxies of  $\sim 2\%$  and a spiral completeness of  $\sim 78\%$ . Most of the excluded spirals have small effective radii, leading to a rather complete and pure sample of Milky-Way-type galaxies or larger. It is these more massive galaxies which are presumably most affected by dust attenuation.

### 5.2.2 GAMA spiral galaxy sample

We construct the GAMA spiral galaxy sample from the 221 373 galaxies of the equatorial fields as listed in the TilingCatv41. Only 41 962 of these have local flow-corrected redshifts of  $z \leq 0.13$ . Of these, 39 373 galaxies have valid Sérsic fits and stellar mass estimates.

11 977 of the 39 373 galaxies are identified as spiral galaxies using the spiral selection technique. Of these, 11 173 galaxies have NUV coverage in the GMC and make it into the final GAMA spiral galaxy sample.

The distribution in absolute r-band magnitude of the spiral galaxy sample is shown in Fig. 5.4.

## 5.3 Derived physical quantities

Some quantities needed as input for the analysis are not provided in the GAMA catalogue, or they are significantly biased and need to be corrected. This section presents the methods employed to derive the necessary quantities and presents the salient properties of the spiral sample defined in Sect. 5.1.

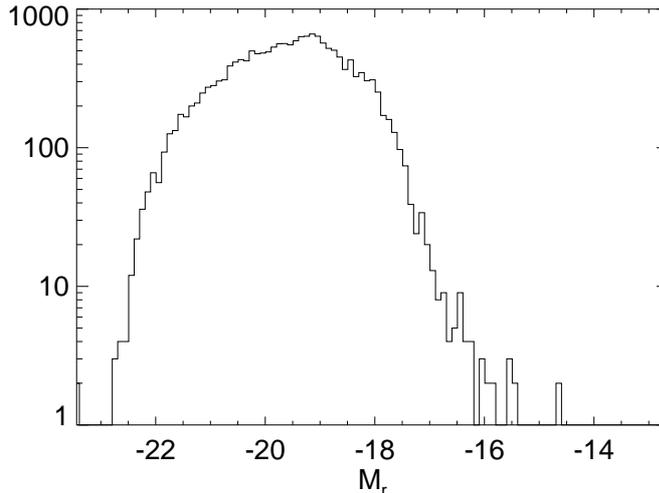


Figure 5.4: Histogram of the absolute r-band magnitude of the GAMA spiral galaxy sample

The distribution of the absolute r-band magnitude follows a Schechter function on the bright side. The kink at  $M_r = -19$  is caused by the incompleteness created by the detection limit. The second kink slightly fainter than  $M_r = -18$  is an artefact of the cell-based spiral selection.

### 5.3.1 Stellar mass

Stellar masses ( $M_s$ ) have been derived for all GAMA sources in the equatorial fields with redshifts  $0.002 \leq z \leq 0.65$  by Taylor et al. (2011). Making use of the broadband GAMA photometry and Bayesian parameter estimation, they performed population synthesis modelling based on stellar population models of Bruzual & Charlot (2003) and a Chabrier (2003) stellar initial mass function.

Since GAMA aperture photometry (Hill et al., 2011) was used to derive stellar masses (because better colour estimates are provided) which might miss some of the galaxy’s flux, the results were scaled by the ratio of the r-band Sérsic magnitude to the r-band aperture magnitude to determine total stellar masses.

It should be noted that Taylor et al. (2011) corrected the stellar mass estimates for the effect of dust attenuation using the Calzetti et al. (2000) model. This is principally inconsistent with the radiation transfer model used here. Furthermore the Calzetti et al. (2000) model does not account for or predict systematic dependencies of attenuation on inclination, individual disk opacity, or B/D ratio. However, Taylor et al. (2011) argue that the effects of dust on the estimated stellar mass are small compared to the individual effects on colour and luminosity, so the choice of dust attenuation model is not critical in this context<sup>1</sup>.

The formal random uncertainties on the derived stellar masses (derived in a Bayesian way) are quoted in Taylor et al. (2011) to be  $\sim 0.1 - 0.2$  dex on average.

It should be noted that the stellar mass estimation becomes incomplete for low stellar mass and red galaxies. This is predominantly due to the fact that those

<sup>1</sup>Taylor et al. (2011) argue that determination of stellar mass is not strongly affected by dust since the reddening vectors are parallel in a stellar mass/luminosity vs. colour plot.

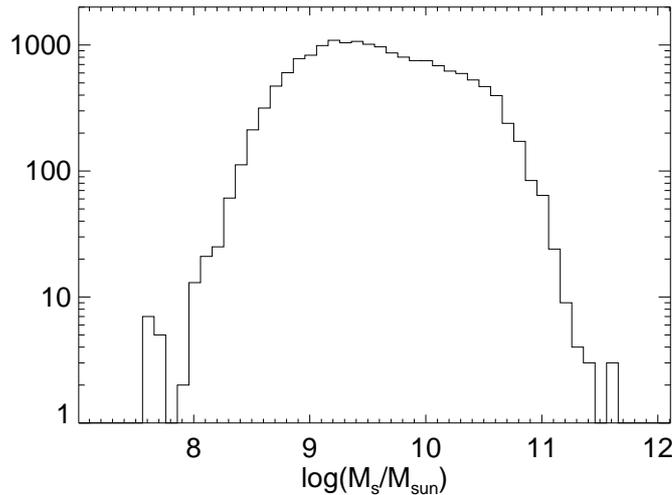


Figure 5.5: Histogram of the stellar mass of the GAMA spiral galaxy sample

galaxies do not pass the r-band detection limit and are thus not included in the GAMA survey. For the redshifts  $0 < z < 0.13$ , spiral galaxies are  $\sim 90\%$  complete and the full galaxy sample is  $\sim 75\%$  complete to  $\log(M_s/M_\odot) \lesssim 9.0$  (see Fig. 6 of Taylor et al., 2011). A histogram of  $M_s$  of the spiral sample defined in Sect. 5.2.1 is given in Fig. 5.5. One can clearly see the turnover of the  $M_s$  distribution for  $\log(M_s/M_\odot) \lesssim 9.0$  due to the incompleteness of the sample.

### 5.3.2 Star formation rate

The star formation rate (SFR) represents the rate at which gas is converted into stars in a galaxy. There are several recipes to derive this quantity from observables, ranging from using UV fluxes, FIR fluxes, nebulae emission lines, or population synthesis modelling of broadband colors (see e.g. Kennicutt, 1998). We estimate the SFR from UV fluxes.

Kennicutt (1998) defines the UV derived SFR using a Salpeter (1955) IMF and a Bruzual & Charlot (2003) stellar population model as

$$SFR[M_\odot \text{ yr}^{-1}] = 1.4 \cdot 10^{-28} L_{UV} [\text{erg s}^{-1} \text{ Hz}^{-1}] \quad (5.2)$$

Salim et al. (2007) and Treyer et al. (2005) derive a scale factor to convert this relation to a SFR using a Chabrier (2003) IMF. For the Chabrier IMF, which we will use in this work,

$$SFR[M_\odot \text{ yr}^{-1}] = 1.4 \cdot 10^{-28} / 1.58 L_{UV} [\text{erg s}^{-1} \text{ Hz}^{-1}] \quad (5.3)$$

Given the NUV flux of the galaxy  $f$  (i.e. the observed NUV flux  $f_{obs}$  corrected for dust attenuation, foreground extinction and k-corrections applied) in  $\mu Jy$  and the luminosity distance  $D_L$  in  $Mpc$ , we can calculate the NUV luminosity.

$$\begin{aligned} L &= f \cdot 4 \cdot \pi \cdot D_L^2 [\mu Jy \cdot Mpc^2] \\ &= f \cdot 4 \cdot \pi \cdot D_L^2 [10^{-25} \text{erg s}^{-1} \text{ Hz}^{-1} \text{m}^{-2} \cdot (3.0857 \cdot 10^{22})^2 \text{m}^2] \end{aligned} \quad (5.4)$$

$$= f \cdot 4 \cdot \pi \cdot D_L^2 [3.0857^2 10^{19} \text{ erg s}^{-1} \text{ Hz}^{-1}]$$

Since the foreground extinction  $A_{\text{corr},UV}$  and the k-corrections  $k_{\text{corr},UV}$  are provided by the GAMA collaboration in magnitudes, we can correct the observed flux  $f_{\text{obs}}$ , obtaining the dust attenuated flux  $f_d$ .

$$f_d = f_{\text{obs}} \cdot 10^{\frac{A_{\text{corr},UV} + k_{\text{corr},UV}}{2.5}} \quad (5.5)$$

To derive the SFR, this flux must then be de-attenuated assuming a given dust model (see Sect. 5.4).

### 5.3.3 Structural parameters

Dust attenuation is a strong function of the morphology of a galaxy (spiral or spheroidal), as well as morphological parameters such as linear size. In particular, for a fixed dust mass, the size of a galaxy will determine the surface density of dust, i.e. whether the galaxy is optically thick or thin. Furthermore, the attenuation of the galaxy's light is different for different inclination angles. Critical measurements used in this analysis are therefore the effective r-band radius and the axis ratio obtained via single Sérsic fits provided in the SersicCatv09 (see Sect. 2.1.3).

#### 5.3.3.1 Axis ratios and sizes

The morphological parameters ellipticity and size are extracted from the single Sérsic fit by Kelvin et al. (2012). The sizes are effective radii, therefore in principle a function of the Sérsic index  $n$  of the galaxy. However, since spiral galaxies are disk dominated, they have a rather small spread in Sérsic indices. Thus, we use the effective radius as radius of the galaxy independent of other parameters. Fig. 5.6 shows the angular size as a function of redshift for four stellar mass bins. One can see very well that the trend in the distribution of angular size with redshift is the one we expect for a constant distribution of physical size. Furthermore, one can also see very well the decreasing size of the galaxy with decreasing stellar mass. In the lowest stellar mass bin, galaxies with redshifts beyond  $z = 0.1$  have on average a higher effective radius than the predicted trend (indicated by the red line). This is an artefact caused by the resolution of the SDSS data used for the single Sérsic fit by Kelvin et al. (2012).

The lower envelope of the size distribution shows several steps, which are caused by the hard cut<sup>2</sup> in physical radii of the spiral selection.

The fitted axis ratio  $b/a$  is obtained from the single Sérsic fits as well. While for perfectly flat disks we can assume the inclination  $i$  - axis ratio relation to be  $b/a = \cos(i)$ , we need to consider the intrinsic height for real galaxies. We present the conversion from  $b/a$  to  $i$  in Sect. 7.2.

#### 5.3.3.2 Bulge-to-total ratio

Dust attenuation of light from bulges and disks has very different characteristics, due to the different geometrical distribution of the stars in these morphological

---

<sup>2</sup>Since the spiral selection is a cell-based method, a hard cut in physical radius is a hard cut at the last cell, i.e. the position of the cut can be different for different combinations of  $M_{i,d}$  and  $n$ , causing the size distribution to have two steps on the lower side.

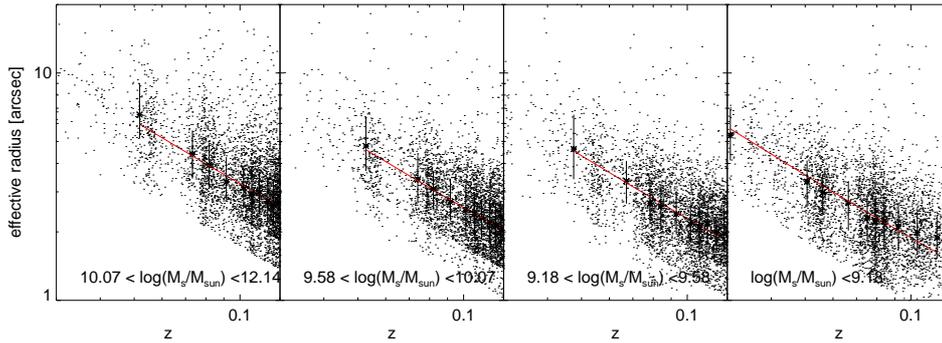


Figure 5.6: Angular size as function of redshift

The angular sizes of galaxies in the GAMA spiral galaxy sample is shown as function of redshift in four stellar mass bins. The angular size as function of redshift of an average physical size galaxy in this stellar mass bin is shown as the red line.

structures in relation to the dust. Therefore, it is necessary to have prior knowledge of the fraction of light in the bulge to understand dust attenuation characteristics of galaxies. This can in principle be achieved through measurement of the bulge-to-disk ratio ( $B/D$ ) in an optical band. However, since GAMA does not yet have available structural decompositions of bulges and disks, we must make use of a proxy for this quantity. The proxy we use is the Sérsic index from the single component Sérsic fits to the GAMA galaxies.

In order to relate the single Sérsic index  $n$  to the  $B/D$ , we make use of the work of Simard et al. (2011), who provide both single Sérsic fits with free Sérsic index  $n$  as well as fits of bulge + disk assuming  $n_{disk} = 1$  and  $n_{bulge} = 4$  for  $\sim 110,000$  galaxies from the SDSS spectroscopic sample and report the resulting bulge-to-total ratio ( $B/T$ ).

First, we select a subsample of spiral galaxies from SDSS with  $z < 0.13$  applying the same recipe of Grootes et al. (2014) we used for our spiral galaxy sample (see Sec. 5.2.1)<sup>3</sup>. We use this SDSS spiral galaxy sample to construct the proxy for the bulge-to-disk ratio.

Plotting a 2D histogram (see Fig. 5.7, left) of the single Sérsic  $n$  vs. the  $B/T$  of the exponential disk - de Vaucouleurs bulge fit, we can see a reasonable correlation between  $n$  and  $B/T$ . For bins of  $n$ , we derive the median  $B/T$  value (red line) and the quartile  $B/T$  values (white lines). We can now derive the  $B/T$  value for a given GAMA galaxy by assigning the median  $B/T$  value of the SDSS spiral sample for the corresponding  $n$  bin, where  $n$  for a GAMA galaxy is found from the single Sérsic fits of Kelvin et al. (2012).

In Fig. 5.7 one sees that there is an acceptably tight correlation up to  $n$  of 2.5. For  $n > 2.5$  the constant value of  $B/T = 0.3$  is used as there is not enough data in the SDSS spiral sample beyond that point to further constrain the correlation. Fig. 5.7 (right) shows a cumulative plot of the  $B/T$  for the real  $B/T$  value of the SDSS spiral sample (black) and the reassigned  $B/T$  using the single Sérsic index proxy (blue). The agreement is acceptable up to a the  $B/T$  value of 0.3, which we set as the highest value possible for the  $B/T$  obtained via the proxy.

<sup>3</sup>Applying the fit quality cuts as suggested by Simard et al. (2011), we select against those galaxies that prefer single Sérsic fits or require a flexible  $n_{bulge}$ . However, none of these galaxies would have passed the spiral selection in first place.

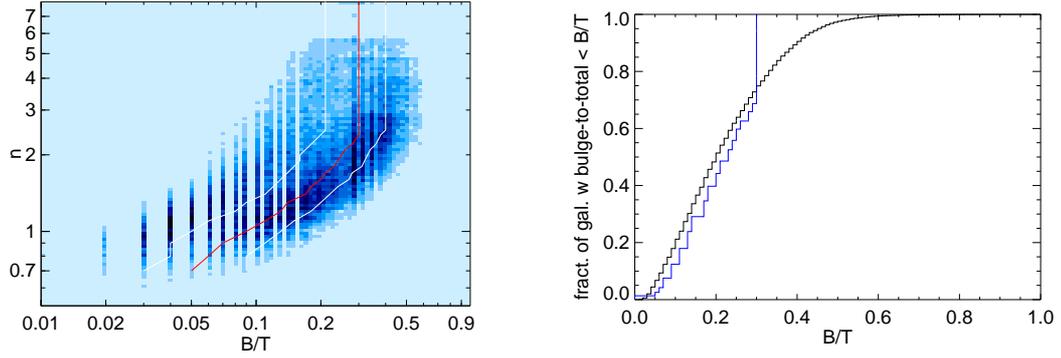


Figure 5.7: Single Sérsic index  $n$  as proxy for bulge-to-total  $B/T$  ratio

Left: 2D histogram for the SDSS spiral sample visualises the correlation between  $n$  and  $B/T$  (both quantities have discrete values). The red line represents the median  $B/T$  value in every  $n$  bin, the white lines the quartile range. The edges in the bin population as function of  $n$  are an artefact from the cell-based spiral selection.

Right: Cumulative plot of the real bulge-to-total ratio (black) and the reassigned bulge-to-total ratio using the single Sérsic index proxy (blue) for the SDSS spiral data set. For the full dynamical range of the lookup table, the reassigned bulge-to-total ratio traces the real one acceptably well. The main difference arises in the high  $B/T$  range beyond 0.3, where the reassigned  $B/T$  is fixed to 0.3.

From the apparent  $B/T$ , it is straightforward to calculate the apparent bulge-to-disk ratio ( $B/D$ ).

$$B/T = \frac{B}{B+D} \quad \longrightarrow \quad B/D = \frac{1}{\frac{1}{B/T} - 1} \quad (5.6)$$

$B/D$  is called the apparent bulge-to-disk ratio because it refers to the observed, i.e. dust attenuated quantities. By calculating the attenuation of each component individually, we can reconstruct the total de-attenuated magnitude of a galaxy (see Sect. 5.4).

## 5.4 Predicted dust attenuation

The goal of this work is to derive a model of the attenuation of starlight by dust in spiral galaxies. There are many models in use in the literature, of varying degrees of complexity, all of which can in principle be tested using the statistical technique developed in this thesis. However, since, as already mentioned, attenuation depends on the relative spatial distribution of dust and stars, the potentially most powerful methods are those based on models for the transfer of starlight through realistic dusty disks. Accordingly, and because direct measurements of structural parameters of GAMA galaxies are available which can be used to constrain the geometry and orientation of the galaxies, we make use in this work of the radiation transfer model of Popescu et al. (2011), which is applicable to a wide range of non-starburst, late-type galaxies. For a detailed technical description of the model, its parameters as well as tests of the performance of the model, the reader is referred to Popescu et al. (2011, 2000) and Tuffs et al. (2004).

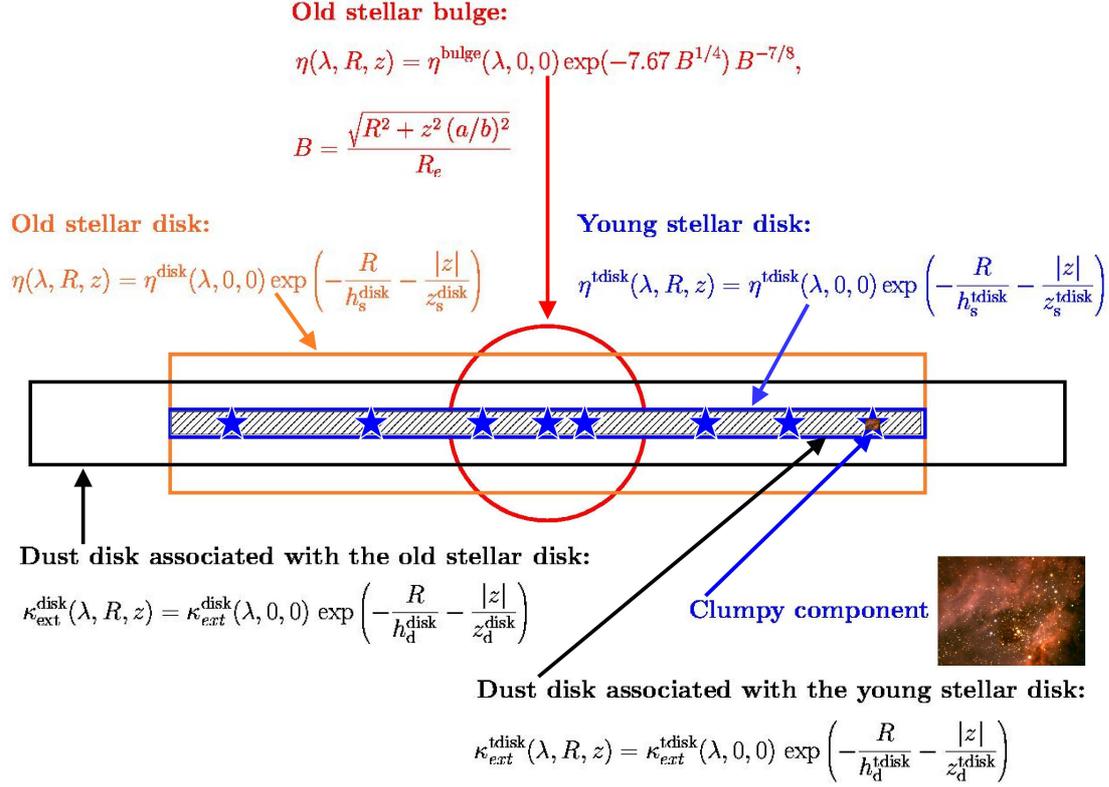


Figure 5.8: Geometric components of the radiation transfer model of Popescu et al. (2011).

Shown is a schematic representation of the geometrical distributions of stellar and dust emissivity together with a mathematical prescription of the stellar emissivities and dust opacities used in the model. See text for details.

### 5.4.1 Radiation transfer model

The distribution of the stellar light is modelled in the radiation transfer (RT) model by a superposition of a de Vaucouleurs bulge consisting of old stellar population, a thick exponential disk consisting of an old stellar population and a thin exponential disk consisting of a young stellar population (see Fig. 5.8). Each disk has its own scale length and scale height. The emission of the dustless bulge is parametrised via the bulge-to-disk ratio ( $B/D$ ), making the model applicable to a variety of geometries along the Hubble sequence.

The optical properties of the dust grains in the disks are modelled using the Weingartner & Draine (2001) dust grain model. The geometrical distribution of the dust is modelled with two components. i) The diffuse component is smoothly distributed in the young and old stellar disk, and is modelled by two exponential disks. It can be seen as representing the large-scale dusty structures (e.g. Cirrus clouds) and has a unity projected surface filling factor. ii) A clumpy component which is representing the dust surrounding star formation regions. This dust mainly attenuates the UV emission of the young stars in these SF regions. The fraction of the UV light from these young stars escaping the surrounding dust into the diffuse ISM is given by the factor  $1-F$  (with  $F$  being fixed to  $F=0.41$  for this work following Popescu et al., 2011).  $F$  is a fixed geometrical blocking factor, so this component of attenuation is independent of wavelength in any given SF

region. However, since stars migrate away from their birth clouds with time, and since the birth clouds become disrupted due to the effect of mechanical energy input from stellar winds and supernovae, more UV light escapes from older clouds than from younger clouds. Furthermore, the escaping light from older clouds is redder due to the colour evolution of the stars with time. This leads to a wavelength dependency of locally absorbed light in SF regions, despite the fundamentally geometrical nature of the blocking (which is independent of the grain properties since the clouds are very optically thick). Unlike the attenuation of light in the diffuse disk, however, the attenuation of light in the clouds is independent of inclination.

The relative scale lengths and scale heights of the stars and diffuse dust in the exponential disks have been calibrated, and fixed to the one found in local edge-on spiral galaxies analysed with the radiation transfer analysis by Xilouris et al. (1999). Doing so, the wavelength dependence of the scale lengths of the stellar emission of the disks (with blue disks being larger than red disks) is also fixed.

The attenuation of UV/optical emission by dust in the diffuse disks is independent of the strength of the emission, and, for a given wavelength, is only a function of the inclination and the disk opacity, which is parametrised via the B-band face-on central optical depth  $\tau_B^f$  in the diffuse dust disks. (For simplicity, the B-band face-on central optical depth  $\tau_B^f$  is hereafter referred to as  $\tau$  or “dust” and is the total opacity, i.e. sum of the opacities of the two disks.) Therefore, Popescu et al. (2011) tabulate the attenuation values  $\Delta m$  resulting from their radiation transfer calculations for different disk opacity and inclination values for the different UV/optical wavelengths. They tabulate the attenuation separately for bulges (attenuated by the dust present in the disk) and disks, enabling the user to calculate the combined attenuation of bulge and disk emission  $A(\tau, i)$  for any given B/D value.

Specifically, we use the equations 17 and 18 in Tuffs et al. (2004) for the UV and optical attenuation, respectively:

$$A_r(\tau, i) = 2.5 \log_{10} \left( (1 - B/T) 10^{\frac{\Delta m_r^{disk}(\tau, i)}{2.5}} + B/T \cdot 10^{\frac{\Delta m_r^{bulge}(\tau, i)}{2.5}} \right) \quad (5.7)$$

$$A_{NUV}(\tau, i) = \Delta m_{NUV}^{disk}(\tau, i) - 2.5 \log_{10} (1 - F \cdot f_\lambda) \quad (5.8)$$

where  $F \cdot f_\lambda$  is the fraction of the emitted UV flux density at wavelength  $\lambda$  which is locally absorbed in star-forming regions and for this work, it is fixed to  $F \cdot f_\lambda = 0.41 \cdot 0.883 = 0.362$  following Popescu et al. (2011). The bulge-to-total ratio  $B/T$  corresponds to the bulge-to-disk ratio derived via the Sérsic index (see Sect.5.3.3.2).

Figure 5.9 shows the model prediction of the NUV-r attenuation as a function of inclination for different  $\tau$  values (shown for  $B/D = 0$ ). It is interesting to note that though the NUV-r attenuation is strongly dependent on both inclination and  $\tau$ , the attenuation curves for different  $\tau$  values are almost parallel. Thus, the main difference in attenuation between model predictions with different  $\tau$  values is caused by the different face-on attenuation, rather than by differences in inclination.

This poses a challenge for the extraction of information of dust attenuation from the data, as the face-on colour is 100% degenerate with the (a priori unknown) intrinsic colour of the stellar emission. In Sect. 7.4 we describe how the degeneracy can be overcome.

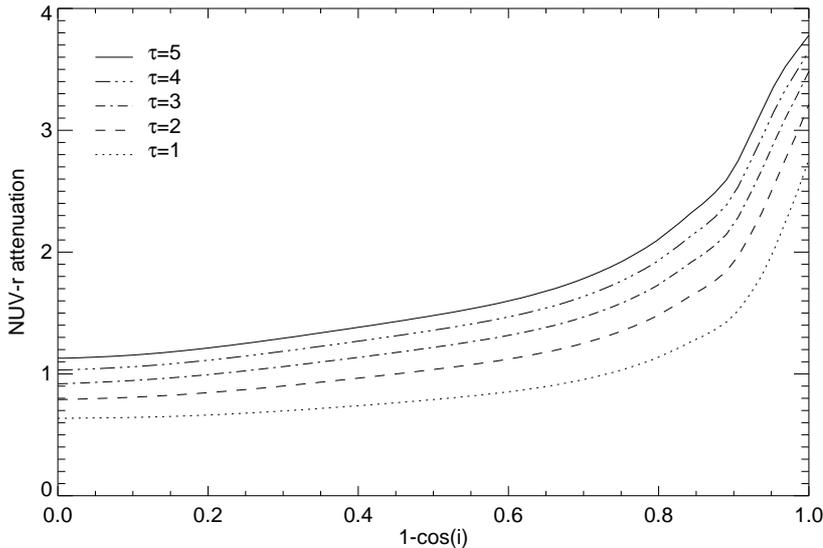


Figure 5.9: Attenuation of NUV-r colour

The radiation transfer model prediction of the attenuation of the NUV-r colour for spiral galaxies (bulge-to-disk ratio = 0) for different  $\tau$  values as function of inclination.

Therefore, in this work, we focus on the different changes of the attenuation with inclination for different  $\tau$  values, which is a comparably smaller signal (see also Fig. 7.5).

### 5.4.2 Disk opacity models

As described above, the fundamental parameters of the radiation transfer model affecting the attenuation of starlight in the diffuse ISM is the disk opacity, i.e. B-band face-on central optical depth  $\tau_B^f$ . The ultimate aim of this work is to optimise the model description of  $\tau$ , linking it to physical characteristics of the galaxy.

As a first step, we employ two models for the  $\tau$  value of a spiral galaxy sample. The first one, using for testing the algorithm, is a naive model of a Dirac delta function of  $\tau$ , i.e. we assume the same value for each galaxy in the data set, disregarding of physical quantities. We refer to this dust model as the “simple dust model”.

As the second model for  $\tau$ , we adopt the stellar mass - dust mass correlation of Grootes et al. (2013)<sup>4</sup>. They used a GAMA spiral galaxy subsample with infrared/submm data to calibrate the stellar mass - dust mass relation, resulting in a  $\tau$  model for spiral galaxies as a function of stellar mass surface density  $\Sigma_s$ .

$$\begin{aligned} \Sigma_s &= M_s / \Lambda \\ \log_{10}(\tau) &= 1.12 \cdot \log_{10}(\Sigma_s) - 8.6 \end{aligned} \tag{5.9}$$

where  $\Lambda$  is the area of the galaxy seen face-on defined by  $\Lambda = 2 \cdot \pi \cdot r_{eff}^2$  with  $r_{eff}$  being the linear size in units kpc. The dependence on the galaxy’s size introduces a further scatter in attenuation besides the inclination since galaxies of a given stellar mass are observed to have a range in sizes.

<sup>4</sup>The stellar mass - dust mass correlation translates to stellar mass surface density -  $\tau$  correlation.

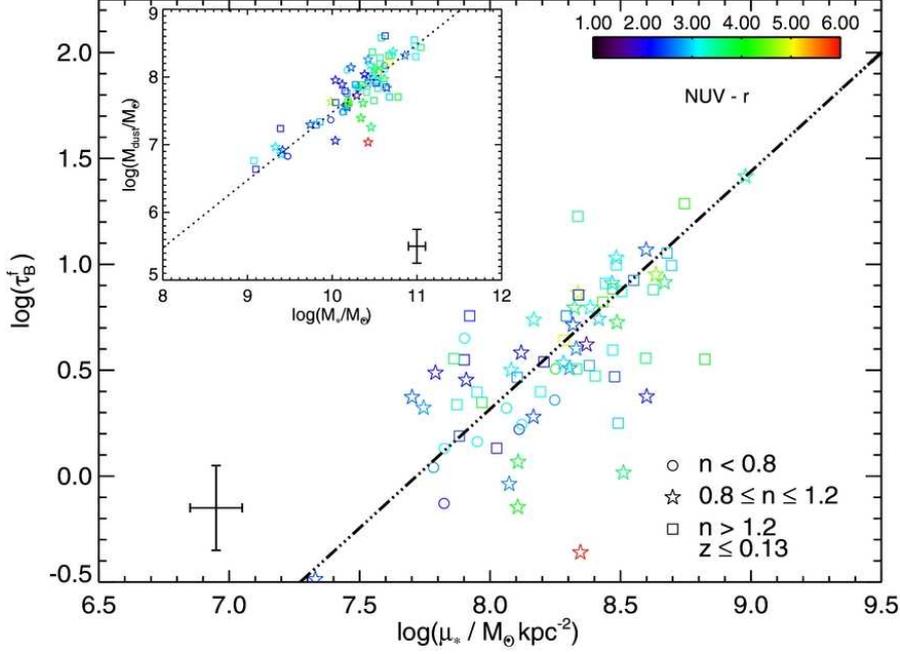


Figure 5.10: B-band face-on central optical depth  $\tau_B^f$  vs. stellar mass surface density. Symbols are coded according to Sérsic index  $n$  and NUV-r color (see figure). The dash-dotted line represents the best fit. The inset depicts the dust mass as a function of stellar mass. The dotted line represents a reference value with a slope of unity and an offset corresponding to  $M_{dust}/M_s = 0.003$ . Taken from Grootes et al. (2013).

We furthermore introduce a scaling factor  $\beta$  to linearly scale the  $\tau$  values of Grootes et al. (2013), and optimise for that scaling factor in this work. We refer to this dust model as the “stellar mass dependent dust model”.

$$\tau = \beta \cdot \frac{\Sigma_s^{1.12}}{10^{8.6}} \quad (5.10)$$

In order to avoid unreasonably high  $\tau$  values, the maximum value for  $\tau$  is set to be 10.

# Chapter 6

## Simulating Data

In developing a complex algorithm for data analysis, it is essential to test the response of the code to simulated data, where the supposed analysis outcome is known in advance. In addition, the simulations can enable data to be better understood, guiding the design of the algorithm and its implementation. For example, selection effects and other biases can be isolated and understood, thus the code can be designed to account for them. The test data is also used to verify the outcome of (intermediate steps of) the analysis.

We therefore simulate the data, including many features in the simulation to make it realistic. We seek a good, but not necessarily perfect imitation of the real data, as the benefit of perfectly fine-tuning of all the parameters in the simulation is little compared to the effort needed. In this chapter, we describe the data simulation in detail. We introduce and discuss the physical assumptions made in this simulation, which are the same as made in the data analysis.

We start with the simulation of the intrinsic quantities of the population in Sect. 6.1, namely intrinsic r-band magnitude, redshift, stellar mass, star formation rate, intrinsic NUV magnitude, size, dust content and intrinsic morphology. Next, we simulate the apparent (observed) properties of the GAMA galaxies in Sect. 6.2: apparent morphology, noiseless apparent magnitude and apparent NUV magnitude with noise. At the end of this chapter in Sect. 6.3.2, we define the flux-limited samples via the detection limit in the r-band and the NUV, present the different simulated data sets used in the following chapters to test and verify the analysis, and discuss in Sect. 6.3.3 all known biases of the (simulated) data.

Throughout this work, we use the matter density  $\Omega_M = 0.3$ , the curvature constant  $\Omega_k = 0$ , the dark energy density  $\Omega_\Lambda = 0.7$ , the speed of light  $c = 2.998 \cdot 10^5$  km/s and the Hubble constant  $H_0 = 70$  (km/s)/Mpc.

### 6.1 Simulation of intrinsic properties of population

#### 6.1.1 r-band magnitude

The first step in the data simulation is to sample the absolute, unattenuated and intrinsic r-band magnitude (i.e. intrinsic magnitude  $M_{r,intr}$ ) of each galaxy. The Schechter function (Schechter, 1976) is the commonly used function to describe the luminosity distribution of galaxies and we adopt this function in this simulation

as well.

$$LF(L|\phi^*, \alpha, L^*) d(L/L^*) = \phi^* \cdot \left(\frac{L}{L^*}\right)^\alpha \cdot e^{-L/L^*} d(L/L^*). \quad (6.1)$$

The Schechter function parameters  $\phi^*$ ,  $\alpha$ , and  $L^*$  are the amplitude (in units of number per magnitude and volume), the faint end slope and the turnover luminosity, respectively.

Since the luminosity function in Eq. 6.1 is increasing exponentially for faint objects, this formulation is inconvenient for numerical integration or for drawing random samples. We therefore change variables from luminosity  $L$  to magnitude  $M$ , using the definition of the magnitude

$$M - M^* = -2.5 \log_{10} \left(\frac{L}{L^*}\right) \quad (6.2)$$

This results in the following form of the luminosity function:

$$LF'(M|\phi^*, \alpha, M^*) dM = 0.4 \cdot \ln(10) \cdot \phi^* \cdot (10^{0.4 \cdot (M^* - M)})^{(1+\alpha)} \cdot e^{(-10^{0.4 \cdot (M^* - M)})} dM \quad (6.3)$$

We sample the r-band intrinsic magnitudes  $M_{r,intr}$  from Eq. 6.3 using Monte Carlo sampling<sup>1</sup>.

For our simulation, we assume the Schechter function parameters  $\alpha = -0.94852566$ ,  $M^* = -21.239886$ , and  $\phi^* = 0.003$ , unless explicitly stated otherwise for a specific simulated data set. These values were derived from the fit of the luminosity function to the observed GAMA spiral galaxy data set (see appendix A)<sup>2</sup>

Fig. 6.1 (left) shows the distribution of intrinsic r-band magnitude as a function of redshift for all simulated galaxies (black), overplotted with those galaxies that pass the r-band detection limit (red). The sampling in redshift shells is clearly visible for the full simulated set.

Given the fact that both formulations of the Schechter function will diverge for faint galaxies, it is necessary to restrict the sampling to a magnitude interval. It is sufficient to choose a bright magnitude limit significantly brighter than the

---

<sup>1</sup>In practice: Two random numbers are drawn. One is a trial absolute magnitude within the sampling interval. The fraction of the probability (normalised LF) of this trial magnitude and a reference probability (arbitrary, but high enough that the fraction is always less than unity) is calculated. The other random number is between 0 and 1. Comparison of this number and the fraction of the probability and the reference probability decides between acceptance (random number is smaller than probability fraction) or rejection (random number is larger) of the trial magnitude. The overall acceptance rate can be tuned by the value of the reference probability.

<sup>2</sup>This is necessary as the spiral selection procedure of Grootes et al. (2014) preferentially selects larger and brighter galaxies (i.e. smaller and fainter galaxies are less likely to be included in the sample). Rather than adopting the actual values of  $M^*$  and  $\alpha$  found in the fit of the LF to the observed distribution of  $m_r$  and  $z$  for the GAMA spiral galaxy sample, we adopt a  $M^*$  which is slightly fainter and an  $\alpha$  which is slightly higher. This was done such that the observed distribution of  $m_r$  and  $z$  of the simulated spiral sample resembled those of the observed spiral sample.

turnover magnitude  $M^*$  (due to the shape of the function, the exact bright limit has a negligible effect on the integral value). On the bright end (high luminosities), the exponential term  $e^{-L/L^*}$  suppresses the power law term  $(\frac{L}{L^*})^\alpha$ . Therefore, for bright galaxies with  $M_{r,intr} \ll M^*$  Eq. 6.1 can be simplified:

$$LF_{bright} d(L/L^*) \approx \phi^* \cdot e^{-L/L^*} d(L/L^*). \quad (6.4)$$

If we choose our lower magnitude limit to be 2.5 magnitudes brighter than  $M^*$ , this translates to an upper luminosity limit of  $10 \cdot L^*$  (using Eq. 6.2). We can now estimate how many galaxies we expect to miss if we choose this upper luminosity limit (lower magnitude limit) instead of  $+\infty$ .

$$\begin{aligned} N_{missed} &= \int_{10}^{\infty} \phi^* \cdot e^{-L/L^*} d(L/L^*) \\ &= \phi^* \cdot (-e^{-L/L^*}) \Big|_{10}^{\infty} = \phi^* \cdot e^{-10} \sim \phi^* \cdot 4.53 \cdot 10^{-5} \end{aligned} \quad (6.5)$$

As reference, we calculate  $N_{ref}$ , the number of galaxies we expect to see in the interval  $M=[M^*-2.5, M^*-1]$ , i.e. the bright region of the magnitude range considered when choosing our lower magnitude limit.

$$\begin{aligned} N_{ref} &= \int_{10^{(1/2.5)}}^{10} \phi^* \cdot e^{-L/L^*} d(L/L^*) \\ &= \phi^* \cdot (-e^{-L/L^*}) \Big|_{10^{(1/2.5)}}^{10} = \phi^* \cdot (e^{-10^{(1/2.5)}} - e^{-10}) \sim \phi^* \cdot 8.11 \cdot 10^{-2} \end{aligned} \quad (6.6)$$

In other words, while the expected number of galaxies within the magnitude interval  $M = [M^* - 2.5, M^* - 1]$  is already significantly smaller than the expected number of galaxies in the whole sample (which is also dependent on  $\alpha$  and therefore not explicitly calculated here), we expect to miss a number of galaxies equal to  $\sim 0.056\%$  of the brightest galaxies if we choose our lower magnitude limit to be  $M^* - 2.5$  rather than  $-\infty$ . Compared with the whole data set, this percentage will be even less.

For the faint limit (upper magnitude limit), The most conservative approach is to choose the faintest possible magnitude that is still detectable at the distance of the closest galaxy, taking into account the effects of noise and dust attenuation. In practice, this faintest possible magnitude is only detectable over a very small redshift range. If we would sample out to this faint magnitude for the full redshift range, we would unnecessarily sample a huge amount of faint galaxies that not pass the detection limit in the end. Since sampling is computationally expensive, it is convenient to sample the galaxies in redshift intervals, with individual faint magnitude limits defined by each lower redshift limit plus a safety margin of one magnitude<sup>3</sup>.

The number of galaxies to be sampled in each redshift bin is calculated by integrating Eq.6.3 over the individual magnitude interval (thus obtaining the number of galaxies per unit volume) and multiplying it by the volume occupied by the redshift shell.

---

<sup>3</sup>For instance, for face-on galaxies with low dust content, dust will scatter more light towards the line of sight than is absorbed by it, thus leading to a small increase in apparent luminosity.

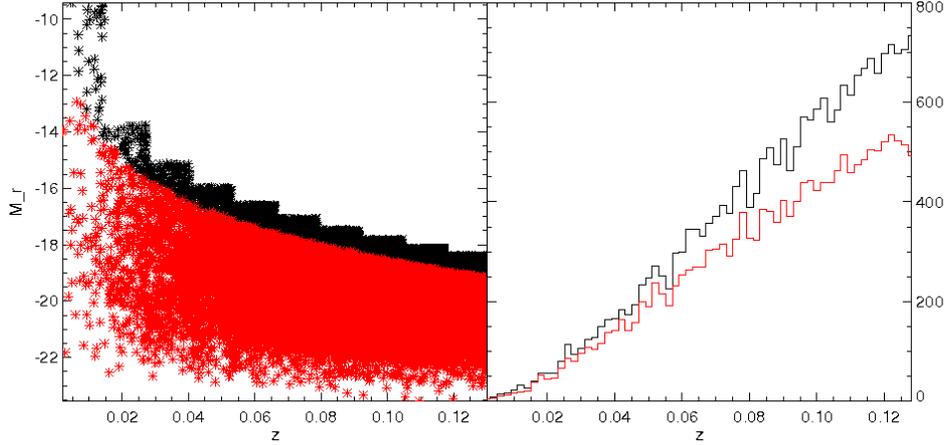


Figure 6.1: Intrinsic magnitude and redshift distribution

On the left, the distribution of absolute magnitude as a function of redshift is shown for data set A.1 (see Sect. 6.3.2), for all sampled galaxies (black) and only those that pass the r-band detection limit (red). The different faint magnitude limit (corresponding to  $m_r = 19.8\text{mag}$ ) for the different redshift shells is easily seen.

On the right, the redshift distribution is shown for all sampled galaxies (black) and only those that pass the r-band detection limit (red). Due to the different faint magnitude limit in the redshift cells, the total distribution in redshift does not follow Eq. 6.7 anymore. Furthermore, all data shown is subject to sampling noise.

### 6.1.2 Redshift

The redshift of a galaxy, even though an indirect observable via the galaxies' spectra, is a fundamental intrinsic quantity as it directly relates to the distance between the observer and the galaxy. It is crucial to be able to account for the change in flux of a galaxy due to its distance to the observer. In particular, we need to be able to convert the measured flux into luminosity (i.e. apparent into absolute magnitude) in order to make it comparable to other galaxies.

Assuming a Euclidian Universe homogeneously populated, the comoving distance can be easily derived from theory. The probability distribution of the comoving distance  $P(D_C)$  is a quadratic function in the comoving distance  $D_C$ .

$$P(D_C) dD_C \propto D_C^2 dD_C \propto d(D_C^3) \quad (6.7)$$

From this probability distribution, we randomly sample the comoving distance for each galaxy.

As the comoving distance is a direct function of the redshift  $z$ , we obtain the redshift  $z$  for each sampled  $D_C$  by solving the following equation numerically.

$$D_C(z) = \frac{c}{H_0} \int_0^z \frac{1}{\sqrt{\Omega_M(1+z')^3 + \Omega_K(1+z')^2 + \Omega_\Lambda}} dz' \quad (6.8)$$

The redshift is sampled in ten intervals over the range of  $[0.005, 0.13]$  (i.e. the comoving distance  $D_C$  is sampled within the corresponding interval). Since the faintest galaxies will only be visible in the nearest redshift interval, it is beneficial to simulate galaxies in redshift intervals and adjust the faint magnitude limit to the inner boundary of the redshift interval, as discussed above. Doing so, we avoid

simulating huge amounts of faint galaxies that would not pass the r-band detection limit (see Sect. 6.3.1), without introducing a bias. The redshift distribution of data set A.1 is shown in Fig. 6.1 (righthand side). Note that the redshift distribution of the observed galaxies differ from the theoretical distribution due to the detection limit.

The luminosity distance  $D_L$  is then calculated using

$$D_L = (1 + z) \cdot D_C(z) \quad (6.9)$$

as it is later used to calculate the apparent magnitude  $m$ .

$$m = M + 5[\log_{10}(D_L(z)) - 1] \quad (6.10)$$

It should be noted that we do not attempt to simulate the effect of large-scale flow of galaxies (local flow) on the redshift measurement.

For reasons of simplicity, we will in the following declare a function  $F$  as  $F(D_C)$  if it is a function of comoving distance or a function of both comoving and luminosity distance and specify the distance dependence in the equation defining of the function.

### 6.1.3 Stellar mass $M_s$

The integrated stellar mass is the most readily physical characteristic of galaxies which can be deduced from photometric data. Since the total metal content of the ISM in spiral galaxies is known to be statistically related to the stellar mass through empirically established scaling relations with metallicity (e.g. Tremonti et al., 2004) and gas mass (e.g. Peeples & Shankar, 2011), it is also natural to utilise  $M_s$  as a parameter of models predicting the dust content of galaxies.

Simple models for the optical emission from stellar populations in galaxies have shown that, for a range of plausible SF histories of galaxies, the stellar mass-to-light ratio should be tightly correlated with the optical colours of galaxies (Bell & de Jong, 2001, and many subsequent works). We predict  $M_s$  for each galaxy in our simulated sample using a slightly modified version of the relation between i-band magnitude and g-i colour by Taylor et al. (2011).

$$\log M_s = (1.15 + 0.7 \cdot c_{g-i,intr} - 0.4 \cdot M_{i,intr})/1.005 \quad (6.11)$$

The relation used by Taylor et al. (2011) to estimate  $M_s$  from the observed, dust-attenuated g-i colours and i-band absolute magnitude  $M_{i,d}$  of the GAMA galaxies is :

$$\log M_s = (1.15 + 0.7 \cdot c_{g-i} - 0.4 \cdot M_{i,d}) \quad (6.12)$$

The modification of Eq.6.12 due to the small factor 1.005 in Eq.6.11 accounts statistically for the small systematic effect of dust attenuation on the relation between mass-to-light and colour. It's value was adjusted such that the simulated relation between (attenuated)  $M_r$  and  $M_s$  matches that of the observed relation (see Fig.6.2).

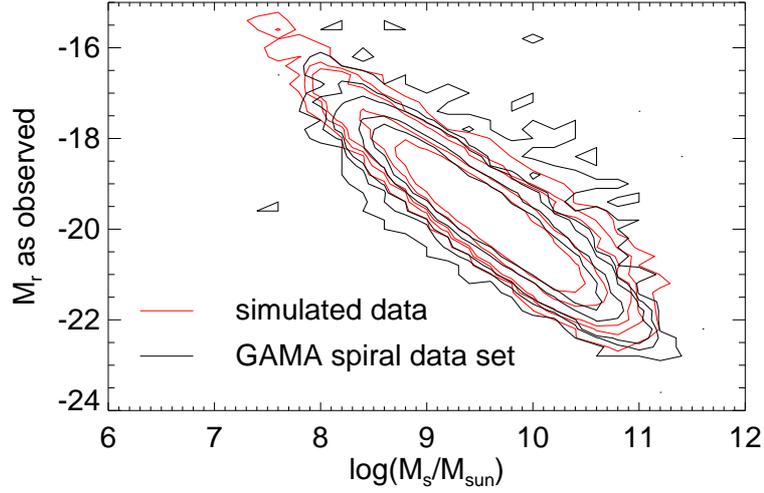


Figure 6.2: observed r-band mag vs. stellar mass  
Comparison between real GAMA spiral galaxy data set (black) and simulation (red; data set C.1). The contour levels are [1,5,10,50,100] per  $M_S - m_{NUV}$  bin.

### 6.1.3.1 Simulation of g-i colour

We derive the g-i colour from the simulated r-band magnitudes. Looking at the distribution of the intrinsic (i.e. dust de-attenuated) g-i colour  $c_{g-i,intr}$  versus the absolute intrinsic r-band magnitude  $M_{r,intr}$  for volume limited samples of real GAMA spiral galaxies<sup>4</sup>, one immediately sees a strong  $M_{r,intr}$ -dependent correlation with a moderate spread. This is modelled as a Gaussian distribution with mean  $\mu_{c_{g-i,intr}}$  and width  $\sigma_{c_{g-i,intr}}$ .

$$\mu_{c_{g-i,intr}} = -0.1 * M_{r,intr} - 1.25 \quad (6.13)$$

$$\sigma_{c_{g-i,intr}} = 0.18 \quad (6.14)$$

$c_{g-i,intr}$  is randomly sampled from this Gaussian distribution.

### 6.1.3.2 Simulation of i-band magnitude

The second quantity we need to simulate in order to derive the simulated  $M_s$  is the i-band magnitude  $M_{i,intr}$ .

Looking at the distribution in intrinsic g-i and r-i, one again sees a strong correlation with very small scatter. We model the r-i color distribution as function of the g-i color again as Gaussian distribution with mean  $\mu_{c_{r-i,intr}}$  and width  $\sigma_{c_{r-i,intr}}$

$$\mu_{c_{r-i,intr}} = 0.34 \cdot c_{g-i,intr} + 0.021 \quad (6.15)$$

For a given  $M_{r,intr}$ ,  $c_{r-i,intr}$  is already distributed according to a Gaussian due to the  $\sigma_{c_{g-i,intr}}$  term in Eq. 6.14. To account for the additional scatter in the colour-colour relation, we add in quadrature a further Gaussian scatter term with  $\sigma = 0.01$  and randomly sample from the combined distribution to determine  $c_{r-i,intr}$ .

<sup>4</sup>The absolute r-band magnitude and the g-i colour were de-attenuated assuming a dust content following Grootes et al. (2013) and using the radiation transfer model of (Popescu et al., 2011).

The i-band magnitude  $M_{i,intr}$  is then calculated using

$$M_{i,intr} = M_{r,intr} - c_{r-i,intr} \quad (6.16)$$

#### 6.1.4 Star formation rate (SFR)

Another quantity of relevance to the dust content of galaxies is the star formation rate - i.e. the rate of conversion of gas into stars. This is because young stars are the primary injectors of metals into the ISM, and the metals can subsequently condense into grains. Some of the metals injected by young stars may even already be in the form of grains. Furthermore, mechanical energy injected by young stars and supernovae into the ISM may lead to the destruction of interstellar grains. It is therefore important to investigate and test models for the dependency of grain content of galaxies on the SFR. In addition, we need to simulate the SFR of galaxies in order to simulate the UV emission of galaxies (see Sect. 6.1.5).

We simulate the SFR of galaxies from the simulated  $M_s$  values, by defining a main sequence relation between specific SFR ( $sSFR = \frac{SFR}{M_s}$ ) and  $M_s$ :

$$\log(sSFR) = -0.548 \cdot \log_{10}(M_s) - 4.358 + G(0.1) \quad (6.17)$$

where the function  $G(0.1)$  represents an intrinsic scatter term, defined as a Gaussian with mean 0 and standard deviation 0.1. The coefficient of  $\log_{10}(M_s)$  in this relation was obtained by fitting a first order polynomial to the corresponding relation of a volume-limited subsample of GAMA spiral galaxies, using the values of SFR and  $M_s$  calibrated in Sects. 5.3.2 and 5.3.1, respectively. In determining the SFR for the GAMA galaxies, the NUV photometry was corrected for dust attenuation using the  $\Sigma_s - \tau$  relation of Grootes et al. (2013).

In Eq. 6.17, the sSFR and  $M_s$  refer to simulated quantities. The value of sigma of 0.1 used in the scatter term in this equation was set such that the scatter in the apparent NUV magnitude for the simulated galaxies after including the effect of dust attenuation and measurement noise on the scatter<sup>5</sup> matched the observed scatter. A comparison of the scatter in the observed  $m_{NUV}$  with the scatter in the simulated  $m_{NUV}$  can be seen in Fig. 6.3. It can also be noted that the scatter in the simulated  $m_{NUV}$  is very much larger than the intrinsic scatter in the sSFR vs  $M_s$  scaling relation of Eq. 6.17, from which the  $m_{NUV}$  is derived. This is primarily due to the effects of NUV noise as well as dust attenuation and the dependence of dust attenuation on orientation, size and geometry of the simulated galaxies.

#### 6.1.5 Absolute intrinsic NUV magnitude $M_{UV,intr}$

The absolute intrinsic NUV magnitude  $M_{UV,intr}$  (i.e. corresponding to the luminosity of a galaxy in the absence of dust) is calculated from the simulated SFR from Eq. 6.17, using the relation between UV luminosity and SFR given by Kennicutt (1998):

$$M_{UV,intr} = -2.5 \cdot \log(SFR) - 18.5365 \quad (6.18)$$

---

<sup>5</sup>Apparent NUV magnitudes  $m_{NUV}$  for the simulated data in Fig. 6.3 were calculated from absolute intrinsic NUV magnitudes, incorporating the effects of dust attenuation, foreground extinction & k-corrections, and noise as detailed in Sect. 6.2. The dust attenuation model used was the stellar mass dependent model, with full treatment of orientation and geometry-dependence of the attenuation.

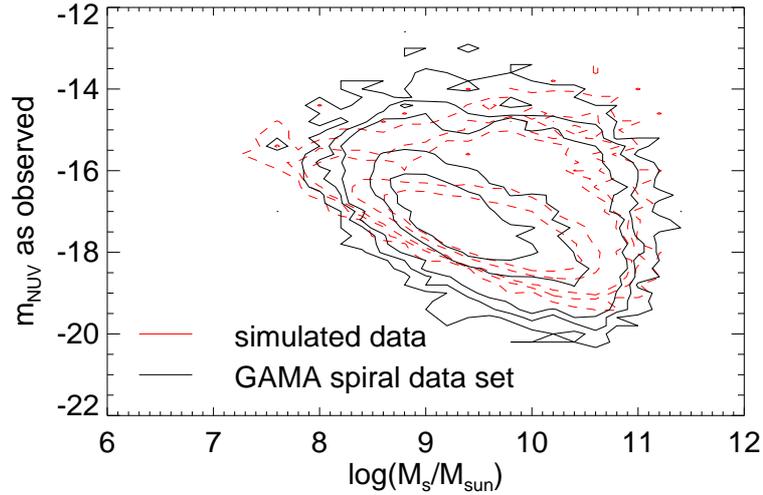


Figure 6.3: observed NUV mag vs. stellar mass

Comparison between the GAMA spiral galaxy data set (black) and the simulated data set C.1 (red). The sample is volume-limited for  $\log_{10}(M_s/M_\odot) > 9.0$ . The contour levels are [1,5,10,50,100] per  $M_s - m_{NUV}$  bin.

Substituting for the SFR in terms of  $M_s$  from Eq. 6.17, one obtains

$$M_{UV,intr} = -1.13059 \cdot \log M_s - G(0.25) - 7.64200 \quad (6.19)$$

In implementing Eq. 6.19 in the simulation, we note that the scatter between stellar mass and UV magnitude is already included in the simulation when generating the sSFR. Since the NUV magnitude directly probes the SFR, we do not need to include a further scatter term.

### 6.1.6 Dust content

As described in Sect. 5.4.1, the dust content is specified through the parameter  $\tau$  of the RT model of Popescu et al. (2011). The  $\tau$  distribution of the GAMA spiral galaxies, predicted as a function of other physical and/or directly observable quantities, is one of the desired outcomes of this work.

For the simulation, we make use of the same two models for  $\tau$  given in Sect. 5.4.2. The first mode, used for testing algorithms, is to assume the distribution of  $\tau$  to be a delta-function (with value 3 or 4). The second model postulates the stellar mass surface density ( $\Sigma_s$ ) dependence of  $\tau$  of Eq. 5.9, reproduced here for convenience.

$$\log_{10} \tau = 1.12 \cdot \log_{10} \Sigma_s - 8.6 \quad (6.20)$$

### 6.1.7 Intrinsic morphology

The intrinsic morphology, in terms of this simulation, consists of the intrinsic r-band radius, the intrinsic inclination, the intrinsic single Sérsic index, intrinsic bulge-to-disk ratio, and the galaxy classification as spiral galaxy. Those five attributes are incorporated into the simulations as described in the following:

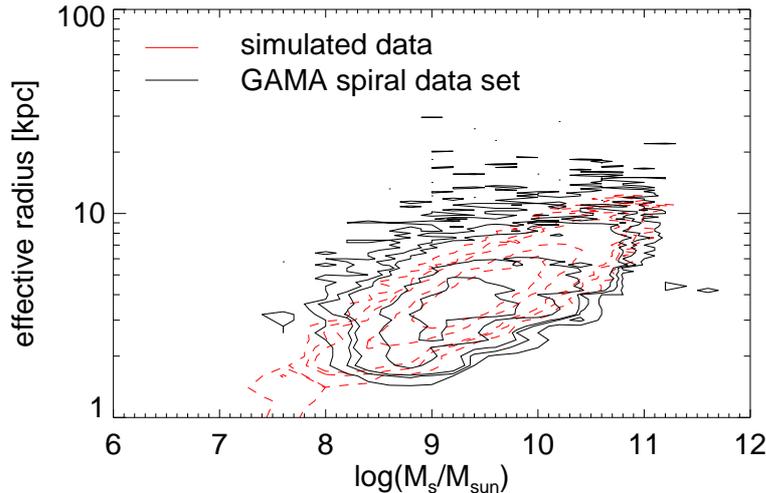


Figure 6.4: effective radius vs. stellar mass

Comparison between real GAMA spiral galaxy data set (black) and simulation (red; data set C.1). The contour levels are [1,5,10,50,100] per  $r_{eff} - M_S$  bin.

### 6.1.7.1 Intrinsic r-band radius

The size of a galaxy is empirically known to be related to its luminosity. We therefore fitted the relation of the logarithm of the observed effective r-band radius (in physical units: kpc) and the r-band absolute, de-attenuated magnitude<sup>6</sup> of the GAMA spiral data set (see Sects 2.1.3 and 5.1.2.1) with a linear function to derive a predictive relation of the intrinsic effective r-band radius:

$$\log r_{eff,intr}[kpc] = -0.118 \cdot M_{r,intr} - 1.69 + G(0.07) \quad (6.21)$$

Here, the function  $G(0.07)$  represents the scatter in this correlation, which we approximate as a Gaussian with standard deviation 0.07. The distribution of simulated and observed  $r_{eff}$  is shown in Fig. 6.4. The resulting distribution of simulated and observed stellar mass surface density is shown in Fig. 6.5.

It should be emphasized that, ideally, one should also correct the observed effective radius for the effect of dust attenuation as well as for the effect the bulge has on the observed radial profile and thus on the measurement of the effective radius<sup>7</sup> before deriving this relation since dust can markedly affect  $r_{eff}$  (Pastrav

<sup>6</sup>using the RT model and the  $\tau$  distribution according to Grootes et al. (2013)

<sup>7</sup>The measurement of the effective radius of disk galaxies using single-Sérsic fits is biased by the presence of a bulge as well as by the presence of dust in the disk. Adding to the brightness of the galaxy profile in the centre, the bulge will lead to higher Sérsic indices and thus smaller effective radii with increasing bulge-to-disk ratio.

At the same time, dust present in the disk biases the light profile we see from a galaxy, making it flatter in the centre. The single-Sérsic fits will prefer a lower Sérsic index and thus a larger effective radii for higher dust content. This effect is even more pronounced for high inclination galaxies than for low inclination galaxies, and it is in general very sensitive to the geometry of the dust distribution. Furthermore, the wavelength dependence of the dust attenuation will furthermore cause the bias of the effective radius to be a function of wavelength. This has been quantitatively predicted for pure disk systems (Möllenhoff et al., 2006; Pastrav et al., 2013a),

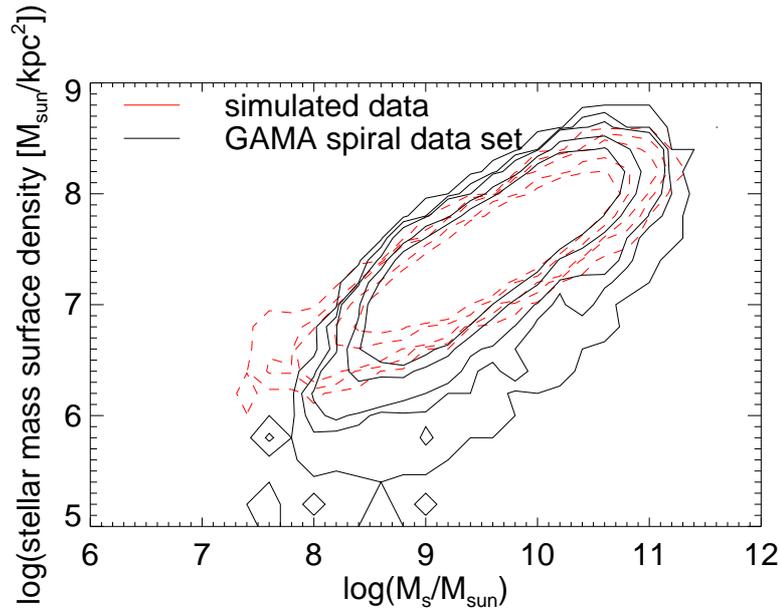


Figure 6.5: Stellar mass surface density vs. stellar mass  
 Comparison between real GAMA spiral galaxy data set (black) and simulation (red; data set C.1). The contour levels are [1,5,10,50,100] per stellar mass surface density– $M_S$  bin.

et al., 2013a,b). This is ignored in this work, both for the simulation and for the analysis of the real data in Chap. 7, but will be included in later work.

### 6.1.7.2 Intrinsic inclination

We take the 3D orientation of galaxies to be randomly distributed. The inclination angle  $i$  is defined as the angle between the normal vector of the galaxy’s disk and the line of sight. Since we observe only a 2D projection of this 3D distribution, the inclination angle  $i$  is not flat in  $i$  but instead is flat in  $1-\cos(i)$ , and we will derive this distribution in this section. For each simulated galaxy, the inclination  $i$  is randomly sampled from this uniform distribution in  $1-\cos(i)$ .

We consider a spherical coordinate system with coordinates  $r$ ,  $\varphi$ , and  $\theta$ , and define the  $\theta = 90^\circ$ -axis to be the line of sight. Thus,  $\theta$  is equal to the inclination angle  $i$  for any given parameter combination.

A given range of  $[\theta, \theta + d\theta]$  covers a circle on the  $(1, \varphi, \theta)$ -surface, with circumference  $C$ :

$$C(\theta)d\theta = 2\pi \sin \theta d\theta \quad (6.22)$$

As mentioned above, the galaxies are randomly distributed in the 3D space, i.e. their normal vector has no preferred position on the  $(1, \varphi, \theta)$ -surface. Therefore, the probability  $P$  of  $\theta$  is proportional to the area covered on the  $(1, \varphi, \theta)$ -surface:

$$P(\theta) \propto \sin(\theta)d\theta$$

---

and has been observed in the wavelength dependence of the size of galaxies (e.g. Kelvin et al., 2012; Häußler et al., 2013).

We now make a variable substitution in order to simplify  $P(\theta)$ :

$$\begin{aligned} u &= 1 - \cos \theta \\ \frac{du}{d\theta} &= \sin \theta \\ P(\theta) &\propto d(1 - \cos \theta) \end{aligned} \tag{6.23}$$

Subsequently, the inclination angle  $i$  has a probability distribution  $P(i)$  which is flat in  $d(1 - \cos(i))$ .

$$P(i) = \begin{cases} (1 - \cos i) & \text{if } 0 \leq i \leq \pi/2 \\ 0 & \text{otherwise} \end{cases} \tag{6.24}$$

An explicit normalisation constant  $a$  is unnecessary, as  $P(i)$  is already normalized.

$$\begin{aligned} \int_0^{\pi/2} a \cdot d(1 - \cos i) &= 1 \\ &= -a \int_0^{\pi/2} d(\cos i) = a [\cos(0) - \cos(\pi/2)] = a \end{aligned} \tag{6.25}$$

However, due to the inclination-dependent attenuation of a galaxy's light due to dust, the global detection limit introduces an inclination-dependent bias. As edge-on galaxies suffer more from attenuation than face-on galaxies, they are predicted to be under-represented in the final, detected sample (see Fig. 6.10).

In the simulation, the bias is generated simply by applying the detection limit to the dust-attenuated magnitudes (see sections below). The actual functional form of this bias is not of interest for the simulation, but is later needed in the analysis of the real data to recover the real  $\tau$  of galaxies and therefore derived in Sect.6.3.3.3. Although the morphology of a galaxy should strictly speaking be defined in terms of intrinsic properties, in parts this can only be done by selections based on apparent quantities.

### 6.1.7.3 Intrinsic single Sérsic index $n$

Imitating the GAMA spiral galaxies (see Fig. 6.6), we simulate the Sérsic index as a broken power law as function of absolute intrinsic r-band magnitude<sup>8</sup> and with Sérsic index  $n_s$  equal to 1 if the r-band magnitude is fainter than -20.5mag.

$$n_s = \max[\log(-2.16 \cdot M_{r,intr} - 4.4)/1.1, 1] + G(0.14) \tag{6.26}$$

where the function  $G(0.14)$  represents the scatter in this correlation and is defined as a Gaussian with mean 0 and standard deviation 0.14.

It should be emphasized that, ideally, one should also correct the measured Sérsic index for the effect of dust before deriving this relation since dust can markedly affect  $R_{eff}$  (Pastrav et al., 2013a,b). This is ignored in this work, both for the simulation and for the analysis of the real data in Chap. 7, but will be included in later work.

---

<sup>8</sup>The GAMA galaxies were corrected for dust attenuation using the RT model and the stellar mass dependent model for  $\tau$ .

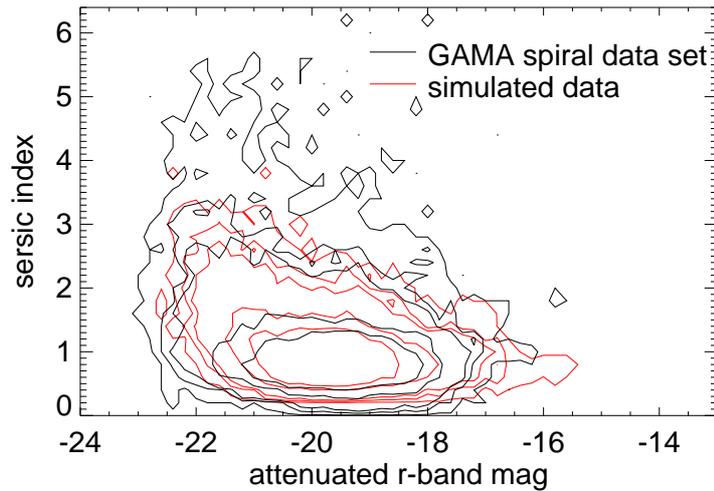


Figure 6.6: Sérsic index vs. observed r-band mag  
Comparison between the GAMA spiral galaxy data set (black) and the simulated data set C.1 (red). The contour levels are [1,5,10,50,100] per  $n - M_r$  bin.

#### 6.1.7.4 Intrinsic bulge-to-disk-ratio

We simulate a spiral galaxy population which we assume to follow a common distribution in colour, r-band magnitude and dust content, irrespective of their bulge-to-disk ratio. However, according to the RT model, dust attenuation (see Sect. 6.2.2.1 below) affects the bulge light differently than the disk light. Thus, the real data needs to consider the bulge-to-disk ratio and we account for this (but only in the calculation of the dust attenuation) in our simulation.

In Sect. 5.3.3.2, we presented the Sérsic index as a proxy for the B/D ratio. In our simulation we adopt this as a correlation, in order to enable the analysis to derive the input B/D from the simulated Sérsic index  $n$ .

#### 6.1.7.5 Spiral galaxy selection

We deliberately attempt to simulate not the full set of GAMA galaxies, but rather focus on the spiral subset<sup>9</sup>. In principle, this means that all simulated galaxies are "spiral galaxies". However, the spiral selection of Grootes et al. (2014) (see Sect. 5.2.1) was designed to be pure rather than complete. Thus, it will fail to include spiral galaxies if their parameter combination is in a region in the parameter space, which is also populated by elliptical or "undefined" galaxies.

In order to simulate the effect of the spiral selection on the distribution of the parameter of the selected spiral galaxies<sup>10</sup>, we apply the same spiral cut to the simulated data set.

<sup>9</sup>An alternative approach would be to consider all galaxies independent of morphology and assign a probability (as a function of measurements  $M_{i,d}$ , sérsic index  $n$  and  $r_{eff}$ ) of a galaxy being a spiral

<sup>10</sup>For instance, galaxies with both very bright and very faint i-band luminosity are disfavoured, as are small galaxies.

## 6.2 Simulation of apparent (directly observable) properties of the GAMA galaxies

### 6.2.1 Apparent morphology

Due to different effects, the intrinsic morphology cannot be observed directly. What we are able to observe is the apparent morphology. We include (some of) these effects in our simulation to make it more realistic and to verify the inference of the intrinsic attributes from the apparent morphology done by the analysis.

#### 6.2.1.1 Observed B/D, single Sérsic index $n$ , and size

The distribution of the simulated intrinsic B/D, single Sérsic index  $n$ , and size are derived in Sect. 6.1.7.2 in a way that they already resemble the observed distributions of the real spiral galaxy sample. In this work, we do not take into account the effect of dust attenuation on the measurements of these quantities and assume the intrinsic quantities to be the same as the observed ones<sup>11</sup>.

#### 6.2.1.2 Observed axis ratio $b/a$

The axis ratio  $b/a$  (i.e. observed semi-minor/semi-major axis ratio) is in practice never identical to  $\cos(i)$ . There are several effects that will cause the observed  $b/a$  to differ from the theoretical expectation of  $\cos(i)$ .

Firstly, the theoretical expectation is valid only for perfectly round disk galaxies. If a face-on disk is not circular but slightly elongated, it will lead to a  $b/a < 1$  and a derived inclination that is too high. As the attenuation varies only slightly with inclination for typical face-on galaxies, this is a minor effect which can be ignored.

Secondly, the intrinsic thickness of a galaxy's disk will lead to  $b/a > \cos(i)$  even for the perfect edge-on case. In fact, the intrinsic thickness  $d_z$  leads to a  $b/a$  defined by

$$b/a = \max[\cos(i), d_z/a] \quad (6.27)$$

thus equal to  $d_z/a$  for edge-on dustless galaxies. As the attenuation is very sensitive for small variations in inclination in the edge-on regime, this effect is much more severe than the first one. If  $b/a$  is equated to  $\cos(i)$ , one would infer a higher dust content than was actually present to explain the observed dimming of the light (see Sect. 7.5.4 for an illustration of this effect). Bearing in mind that we fit inclination and semi-major axis with single-component Sérsic fits to possibly two-component systems,  $d_z$  is also affected by potential bulges.

Thirdly, blurring due to the PSF will potentially cause  $b/a$  to be smaller than  $\cos(i)$ . In Sect. 5.3.3.1, Fig. 5.6, we showed that blurring has an effect on the determination of  $r_{eff}$  for very low mass galaxies close to the redshift limit. One might therefore expect blurring to have an effect on the measured semi-minor axis sizes over a much larger range of  $z$  and  $M_s$ .

---

<sup>11</sup>We refer to the outlook Sect. 8.6 for a description of the not yet incorporated corrections of these effects.

Bearing these considerations in mind, we first investigated a model to predict the observed  $b/a$  incorporating both PSF blurring and a finite intrinsic thickness. Assuming a fixed disk scale height (thickness of the disk)  $d_z=400\text{pc}$ , we calculated the intrinsic semi-minor axis (i.e. PSF-free)  $b_{intr}$  ratio via Eq. 6.27

$$b_{intr} = \max[\cos(i) \cdot a_{intr}, d_z] \quad (6.28)$$

where  $a_{intr}$  denotes the intrinsic semi-major axis (i.e. intrinsic effective radius) in units pc. Each axis was then transformed from linear to the angular size ( $b_{\theta,intr}$  and  $a_{\theta,intr}$ , in units of arcsec) and convolved with an assumed PSF<sup>12</sup> with  $\sigma_{PSF} = 0.7$  arcsec.

$$b/a_{obs} = \sqrt{\frac{b_{\theta,intr}^2 + \sigma_{PSF}^2}{a_{\theta,intr}^2 + \sigma_{PSF}^2}} \quad (6.29)$$

This model takes into account the stellar mass and redshift of each simulated galaxy, the intrinsic, stellar mass dependent linear size of galaxy (Eq. 6.21, via the mass-luminosity relation in Eq. 6.12), the true inclination randomly sampled from a flat distribution in  $\cos(i)$  (Eq. 6.24). Applying this for each galaxy in the simulated spiral sample yielded a simulated distribution of observed  $b/a$ , which could be compared with the true observed distribution for the real data. The effect of dust depressing the representation of truly edge-on galaxies (see Sect. 6.3.3.3) is explicitly taken into account, according to the stellar mass dependent dust model (see Sect. 5.4.2) used in the simulation. However, the simulation predicted a strong dependence of the distribution of  $b/a$  with  $M_s$  and redshift (in the sense that the distribution predicted to peak at lower  $b/a$  values for higher redshifts and lower stellar masses) which was not seen at all in the data (shown in the top panels of Fig. 6.7). This is puzzling as one would expect more massive galaxies to have higher ratios of scale-length to scale height rather than low mass galaxies. We speculate that this effect is nullified by the increasing prevalence of bulges in high mass galaxies affecting the single Sérsic fit parameters.

For this reason, a simpler model depending only on the specification of the disk thickness was adopted, as this fitted the data much better (see Fig. 6.7 for a comparison of the observed  $b/a$ , top panels, and the simulated  $b/a$ , bottom panels).

Specifically, we specify the observed axis ratio by including an intrinsic disk thickness with

$$d_z = a \cdot (0.2 + G(0.1)) \quad (6.30)$$

where  $G(0.1)$  is a Gaussian noise term with standard deviation of 0.1 and mean 0. This assumed thickness is slightly larger than what one would expect for pure disks (e.g. Xilouris et al. (1999) quote a disk scale-height to semi-major axis ratio of 0.074 for massive spiral galaxies). The scatter also represents both scatter in the intrinsic disk height, but predominantly the scatter in bulge-to-disk ratio.

---

<sup>12</sup>The PSF of the SDSS data used for the single Sérsic fits has a PSF of 1.4 arcsec. However, since Kelvin et al. (2012) did include a PSF convolution in their single Sérsic fit technique, we only expect to see effects from the difference between the assumed model of the PSF and the local PSF of the data.

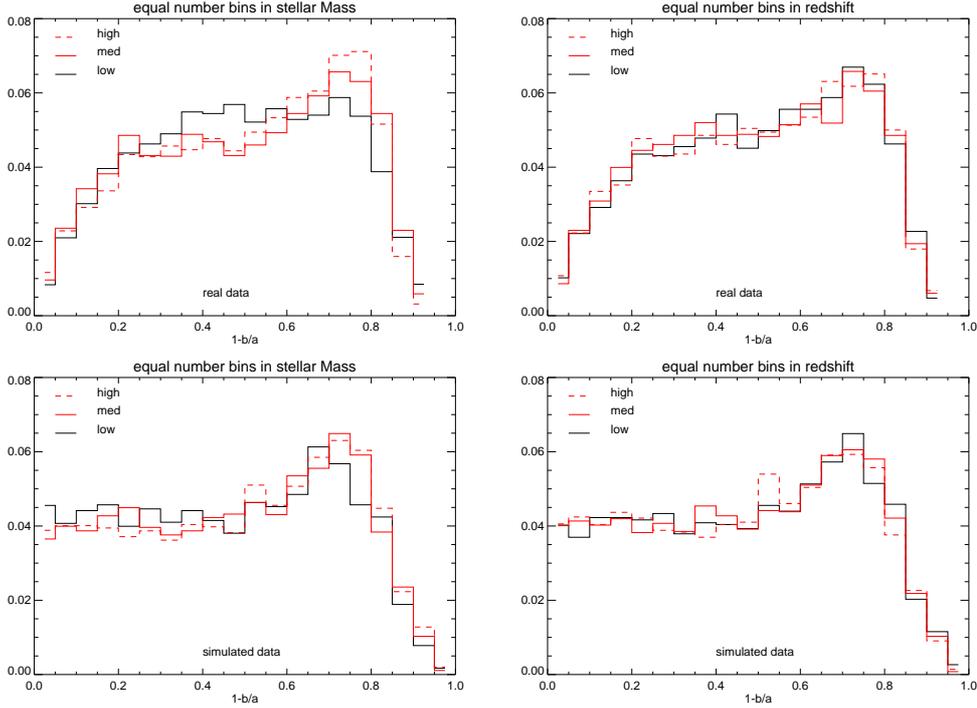


Figure 6.7:  $b/a$  distribution of real GAMA spiral galaxies and simulated data. The  $b/a$  distribution of real GAMA spiral galaxies (top row) and for a simulated data set F.1 are compared for three bins in stellar mass (left) and redshift (right). The shift of edge-on galaxies to slightly larger  $b/a$  values is very pronounced in both the real data and the simulation. The  $b/a$  distribution GAMA spiral galaxies is furthermore suppressed for face-on galaxies. See text for details.

Eq. 6.27 thus reads<sup>13</sup>

$$b/a = \max[\cos(i), 0.2 + G(0.1)] \quad (6.31)$$

Fig. 6.7 shows a comparison between the observed  $b/a$  ratio for the real GAMA spiral galaxy data set (top) and a simulated data set (bottom). Shown are histograms of  $b/a$  for three stellar mass bins (left) and three redshift bins (right). The shift of intrinsic edge-on galaxies to slightly lower inclination due to the intrinsic thickness of the disk is clearly noticeable by the lack of galaxies with  $1 - \cos(i) = 1$  and the pile up around  $1 - \cos(i) = 0.75$ . Both data sets show no trend with redshift. The GAMA data shows only a small trend with stellar mass, which is not present in the simulation. Because of the reasonably good prediction of the observed  $b/a$  distribution based on Eq. 6.30, the same model is used in the inference of  $i$  from  $b/a$  in the real data (see Sect. 7.2).

The GAMA data also show a suppression of perfectly face-on galaxies. While the reason for this is not known, other data sets suffer the from the same problem

<sup>13</sup>In theory,  $G(0.1)$  can produce values from  $-\infty$  to  $\infty$ .  $b/a$  is equal to  $\cos(i)$  if  $G(0.1)$  goes to  $-\infty$ , thus this case is not problematic. If a value larger than 0.8 is sampled from  $G(0.1)$ , then  $b/a > 1$  which is not defined (by definition  $b \leq a$ ). However, the probability of sampling a value from a gauss function that is more than 8 sigma away from the mean is negligible. In practice, one expects only  $\sim 1\%$  to have a  $d_z/a > 0.5$ .

(e.g. Masters et al. (2010) report a similar feature in the inclination distribution of Galaxy Zoo spiral galaxies). One possible explanation could be the effect of an elliptical bulge on the single Sérsic fits, biasing the ellipticity in the face-on view. However, lacking the real explanation for this bias, and since, as already mentioned, the dust attenuation is only weakly dependent on the exact inclination value for face-on galaxies, we do not include this effect when simulating the observed  $b/a$  values.

## 6.2.2 Noiseless apparent magnitudes

The main difference between (absolute) and apparent magnitude is of course the effect introduced by the distance between observer and galaxy (see Eq. 6.10). Apart from that, there are other effects that change the perceived luminosity density of a galaxy at any wavelength. For this work, we focus on the effects of dust attenuation, foreground extinction and redshift-dependent frequency shift (k-correction).

### 6.2.2.1 Dust attenuation

We use the results of the RT model of Popescu et al. (2011) to derive the attenuation values  $A_r(\tau, i)$  and  $A_{NUV}(\tau, i)$  for the r-band and the UV for each simulated galaxy using Eqs. 5.7 and 5.8, respectively, which are reproduced for convenience here:

$$A_r(\tau, i) = 2.5 \log_{10} \left( (1 - B/T) 10^{\frac{\Delta m_r^{disk}(\tau, i)}{2.5}} + B/T 10^{\frac{\Delta m_r^{bulge}(\tau, i)}{2.5}} \right) \quad (6.32)$$

$$A_{NUV}(\tau, i) = \Delta m_{NUV}^{disk}(\tau, i) - 2.5 \log_{10} (1 - F f_\lambda) \quad (6.33)$$

The inclination  $i$  of a given galaxy was sampled from a flat distribution in  $\cos(i)$  according to Eq. 6.24. The  $\tau$  value is, depending on the dust model used, either a delta function with  $\tau = 3$  or defined as a function of stellar mass surface density according to Eq. 5.9. The effective radius needed in Eq. 5.9 is defined in Eq. 6.21. The bulge-to-total ratio  $B/T$  corresponds to the bulge-to-disk ratio sampled via the Sérsic index from Eq. 6.26 (see Sects. 6.1.7.4 and 5.3.3.2).  $F$  and  $f_\lambda$  are fixed to 0.41 and 0.883, respectively, following Popescu et al. (2011), thus  $F f_\lambda = 0.362$  in the NUV (see Sect. 5.4.1). The values of the attenuation of the individual morphological components in the two bands  $\Delta m_r^{disk}(\tau, i)$ ,  $\Delta m_r^{bulge}(\tau, i)$ , and  $\Delta m_{NUV}^{disk}(\tau, i)$  are provided by Popescu et al. (2011) as a function of  $\tau$  and  $i$  as a look-up table with quite large gridding. In order to smooth the attenuation distribution, we interpolate the provided table using a spline fit in both dimensions,  $\tau$  and  $i$ .

The attenuation  $A(\tau, i)$  is then added to the intrinsic apparent magnitude  $m_{intr}$  in order to obtain the dust-attenuated apparent magnitude  $m_d$ . Note that, as explained in Sect. 5.4.1,  $\tau$  refers to the B-band central face-on opacity due to diffuse dust in the disk. Since the RT model also assumes an inclination independent, clumpy dust component around star forming regions, the predicted UV attenuation is  $> 0$  even for  $\tau = 0$ .

$$m_d = m_{intr} + A(\tau, i) \quad (6.34)$$

While the NUV magnitudes are subject to noise (see Sect. 6.2.3.2), we assume the real r-band data to have a high S/N. We therefore neglect the effect of noise on

the r-band magnitude and define the simulated dust-attenuated apparent r-band magnitude  $m_{r,d}$  to be equal to the observed apparent r-band magnitude  $m_{r,obs}$ .

### 6.2.2.2 Foreground extinction and k-correction

In our simulation, we employ a non-parametric model for the foreground extinction (A-correction) and the redshift-dependent frequency shift in the observed band (k-correction) by importing the corrections of the real GAMA spiral galaxies (see Sects. 5.1.2.2 and 5.1.2.3) as a four element vector array  $[A_{r,corr}, k_{r,corr}, A_{NUV,corr}, k_{NUV,corr}]$  (see Fig. 5.2 for a histogram of the distribution). For every simulated galaxy, we randomly sample a four element vector from the array, corresponding to a single real galaxy, thus preserving correlations between the two bands.

The dependence of the A-correction on the position on the sky for real galaxies (see Fig. 5.3 left) is thereby lost. However, the analysis is insensitive to galaxy position. The weak redshift dependence of the k-correction (see Fig. 5.3 right) of the GAMA spiral galaxies is however not included in the simulation, since it has only negligible effect on the analysis of galaxies with the redshift limit of  $z = 0.13$ .

Even though we assume the A- and k-corrections to be exact, i.e. the foreground extinction and the redshift dependent frequency shift in the observed band are assumed to be perfectly reversible, we need these effects in our simulation since the detection limit is applied to the observed data which includes these effects (see Sect. 6.3.1). We export the four element vector as an additional "measurement" for each simulated galaxy.

## 6.2.3 Apparent NUV magnitude with noise

While we assume the r-band magnitude to be bright enough to have negligible measurement errors due to the spectroscopic limit of 19.8 mag, the NUV source is included in the blind catalogue if it has a signal-to-noise ratio (S/N)  $\geq 2.5$ . As detailed in Sects. 3.5 and 3.4, magnitudes measured by COG or resulting from the flux redistribution in the advanced matching catalogue can even have lower S/N. In these regimes, the noise effect on the observed magnitude is especially severe. Therefore, we must simulate the NUV magnitude with noise. To do so, we must be aware of the NUV background and exposure time as they have an effect on the NUV noise.

### 6.2.3.1 UV background flux and exposure time

Background flux and exposure time are important characteristics of the UV flux measurement of the GALEX pipeline (see Sect. 3.1.3). In Fig. 3.4 we show the true UV background flux distribution of the area of sky covered by the GAMA survey. It does not show a huge spread and is in general very low compared to optical backgrounds. As GALEX-GAMA provides background maps, we import these as non-parametric model in our simulation and we randomly sample a background value from these maps. The existing dependence of the background on the position on the sky (i.e. [RA, DEC]) is ignored since the analysis treats the data as a single dataset and is insensitive to [RA, DEC].

The distributions of the NUV and FUV exposure times for the three GALEX-GAMA survey depths are shown in Fig. 3.2. This inhomogeneity within each depth affects the flux measurement, the noise level and the detection limit significantly (see following section). Therefore, it is crucial to include this inhomogeneity in our simulation by randomly sampling the exposure time from the provided exposure time maps (using GMC depth, as this is the deeper depth with reasonable coverage and will be used in the analysis of the real data).

Since both the background map and the exposure time map are provided at the same resolution, we choose to sample both quantities as pairs. In other words, we randomly sample a position on the maps from which we import both background and exposure time, preserving potential correlations<sup>14</sup>.

### 6.2.3.2 NUV noise

The noise properties of GALEX observations and the redefinition of the UV error were described in detail in Sect. 3.1.4. We adopt the noise definition of Eq. 3.5 for our simulations.

Hence, assuming circular sources with radius 3.4 arcsec (i.e. an area  $a = \pi \cdot 3.4''^2$ , which corresponds to unresolved point sources), we calculate the noise-free total count number from the noise-free source mag  $m_{NUV,d}$  corresponding to a noise-free flux. Given the sampled exposure time and background level, Eq. 3.5 calculates the standard deviation of the noise distribution. From this distribution, a random noise realization is drawn and added to the noise-free NUV flux.

### 6.2.4 Apparent NUV-r colour with noise

Having simulated the NUV and r-band magnitudes, with NUV noise, we can construct the observed NUV-r colour as a function of stellar mass. While the intrinsic NUV-r colour of the population is assumed to be distributed as a Gaussian with constant standard deviation and a stellar mass-dependent mean, the observed colour distribution differs slightly from this due to detection limits, dust attenuation and NUV noise (as discussed in detail in Sect. 6.3.3). Fig. 6.8 compares the actual observed colour -  $M_s$  relation for the real GAMA spiral galaxies with NUV detections and the corresponding simulation of the observed relation (data set C.1). It can be seen that the simulated galaxies have a somewhat steeper dependence of NUV-r colour on  $M_s$  than the observed. This may indicate that the simulated dependence of intrinsic UV luminosity on  $M_s$  is less steep than that given in Eq. 6.19 and/or that the dependence of dust mass on stellar mass is less strong than given by the Grootes et al. (2013) model.

### 6.2.5 Motivation of log-normal distribution of intrinsic luminosity

The colour - stellar mass diagram of the spiral GAMA galaxy sample with NUV detections is shown in Fig. 6.9 (left panel). The right panel shows a histogram of

---

<sup>14</sup>For instance, longer exposure times in regions with high background level might have been scheduled in order to be able to detect comparably faint objects as one would in low background regions. However, we could not see a significant correlation in the real data.

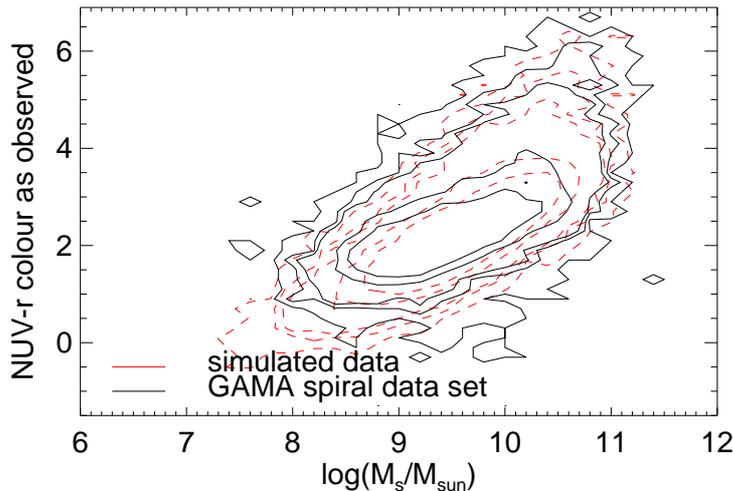


Figure 6.8: Observed NUV-r color vs. stellar mass

Comparison between the GAMA spiral galaxy data set with NUV detections (black) and the simulated data set C.1 (red). The contour levels are [1,5,10,50,100] per  $M_S - (m_{NUV} - m_r)$  bin.

the colour using only data with a small range of stellar mass ( $9.0 < \log(M_s/M_\odot) < 9.1$ ). Since the r-band flux-limited sample is complete at  $M_s \gtrsim 10^9 M_\odot$ , the form of the histogram should not be affected by the r-band flux limit. In addition  $\sim 98\%$  of the galaxies are detected in the NUV. Therefore, the histogram should correspond to the actual distribution of apparent colours of the population of spiral galaxies selected using morphological proxies. This distribution is well fitted by a Gaussian (see black line in the right hand panel of Fig. 6.9, corresponding to a Gaussian with mean  $\mu = 2.082$  mag and standard deviation  $\sigma = 0.415$  mag). Furthermore, the narrowness of the bin in stellar mass is such that the shape of the histogram will be unaffected by the systematic trend of increasing redness with increasing  $M_s$ . Because, for such low  $M_s$  (the lowest where we are still reasonably complete), dust attenuation may play only a subordinate role in the observed colour of galaxies, these results motivate the general use of Gaussian distributions in magnitude space to describe scatter in the simulated distributions of intrinsic (i.e. without dust attenuation) NUV and optical photometry (in Eqs. 6.19 and 6.14).

By the same token, this also motivates the assumption of a Gaussian distribution for the intrinsic distribution of colour fitted to the real data in the procedures described in chapter 7.

It should be noted that the Gaussian distribution in magnitude (log-normal in flux) may nevertheless be imperfect. A small tail can be seen towards red colours in the histogram of Fig. 6.9 (right). There is no way of knowing from the UV-optical photometry alone whether this is an effect in the intrinsic colours (for example due to a small minority of spirals being quiescent in SF activity) or due to an effect of high dust attenuation. This can in principle however be investigated using information on inclination as described in chapter 7.

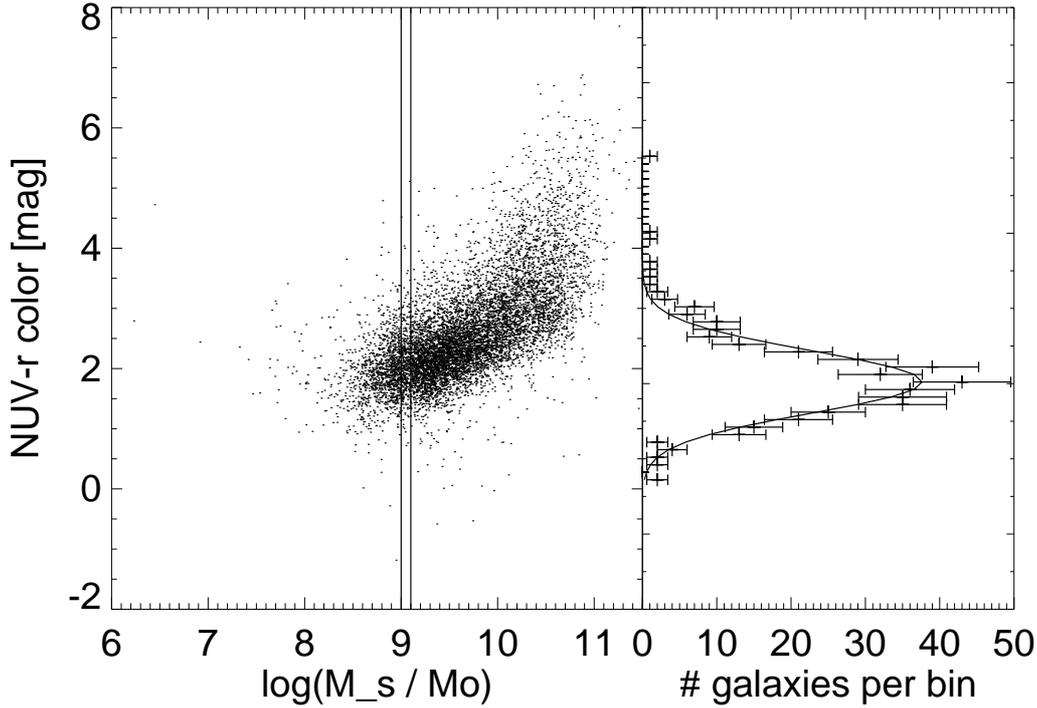


Figure 6.9: Observed NUV-r colour distribution for GAMA spiral galaxies.

The left panel displays the observed NUV-r colour distribution for GAMA spiral galaxies with NUV detections as a function of stellar mass. The dependence of colour on stellar mass is easily seen. The right panel shows a histogram of the colours, using only data with a small range of stellar mass ( $9.0 < \log(M_s/M_\odot) < 9.1$ ; indicated by the solid line on the left panel). The solid line on the right panel represents the best fitting Gauss function ( $\mu = 2.02$ ,  $\sigma = 0.414$ ) to this histogram. Note the histogram of the colours is slightly skewed which may either be due to a skew in the distribution of intrinsic colours, or in the distribution of reddening due to dust.

## 6.3 Simulation of flux-limited samples

### 6.3.1 Detection limits

The real data inherits a detection limit in the r-band from the spectroscopic limit of the GAMA survey. Furthermore, the UV catalogues have their own detection limit. While galaxies must pass the r-band detection limit in order to be in the sample, the properties of the sample are unaffected by the NUV detection limit (i.e. NUV non-detections are included in the sample). Nevertheless, we need to accurately simulate the classification of UV counterparts to galaxies as being detections or non-detections, as the inference of intrinsic colour distribution and dust content described in chapter 7 needs to make use of the upper limits. In this section, we repeat the detection limit definitions, which we must also apply to the simulated data in order to have a realistic simulation.

#### 6.3.1.1 r-band detection limit

The spectroscopic limit of the GAMA survey, which represents the r-band magnitude limit of data set, is 19.8 mag after foreground extinction correction (see

Sect. 2.1). This translates to

$$m_{r,obs} - A_{r,corr} < 19.8 \text{ mag} \quad (6.35)$$

where  $m_{r,obs}$  is the simulated apparent r-band magnitude including the effects of dust (see Eq. 6.34). We note that k-corrections are not included in this definition as the spectroscopic limit of the GAMA survey purposely did not include k-corrections.

Real galaxies with fainter apparent magnitudes are not included in the GAMA data set and therefore simulated galaxies fainter than this detection limit are reported as "not observed" and are excluded from the simulated data set.

### 6.3.1.2 NUV detection limit

The UV detection limit of the blind catalogue (and for 1-to-1 matches passed on to the advanced match catalogue) is defined via the signal-to-noise ratio (S/N). Sources with  $S/N < 2.5$  are not included in the data set. We derive the UV detection limit in detail in Sect. 3.8 using the redefined measurement errors, and only quote here Eqs. 3.9 and 3.8:

$$m_{lim} = ZP - 2.5 \cdot \log_{10}(s_{lim})$$

and

$$s_{lim} = \frac{(S/N)^2 + \sqrt{(S/N)^4 + 4 \cdot t_e \cdot 1.5 \cdot bg \cdot a \cdot (S/N)^2}}{2 \cdot t_e}$$

Note that this definition of the detection limit leads to an individual detection limit for each source / position on the sky. Since this is the case for the real data, it is crucial to include this individual detection limit in the simulation. This is done by using the individual background and exposure time values sampled as described in Sect. 6.2.3.1.

## 6.3.2 Simulated data sets

In the following chapters, we develop our analysis for deriving the intrinsic colour distribution and the dust model (parameters) of the GAMA spiral galaxies. Doing so, we will use simulated data sets for demonstration and validation of the analysis. In this section, we present the simulated data sets used for these purposes and describe them in detail. In the following chapters we only quote the specific data set and refer to this section for the details of the simulation.

We will refer to data set X for all realisations combined and to data set X.1 if we mean only the first of multiple realisations.

The distributions from which the quantities of the simulated galaxies are sampled from are described in detail in Sects. 6.1 and 6.2. All simulations assume the same luminosity function parameters for the r-band:  $\alpha = -0.94852566$ ,  $M^* = -21.239886$ , and  $\phi^* = 0.003$ . The effects which can be turned on and off (indicated with "yes"/"no") or which can have individual features for an individual simulated data set are:

- Intrinsic scatter

- Dust model and its parameter
- NUV noise
- Vertical disk thickness
- Number of data set realizations

### 6.3.2.1 Data set A

- **Intrinsic scatter:** yes
- **Dust model and its parameter:** simple dust model with  $\tau = 3$
- **NUV noise:** both with and without
- **Disk thickness:** infinitely flat disk, i.e.  $b/a = \cos(i)$
- **Number of data set realizations:** 100

### 6.3.2.2 Data set B

Data set B is a data set used to test the derivation of the colour parameters in Sect. 7.3. It uses a constant rather than a stellar mass dependent mean colour ( $\mu = 1.4$ ) and Gaussian standard deviation of  $\sigma = 0.6$ .

In order to have a large and a small sample, we generated the data set with two different values for the Schechter function amplitude:  $\phi^* = 0.003$  and  $\phi^* = 0.0003$ .

- **Dust model and its parameter:** simple dust model with  $\tau = 0$
- **NUV noise:** no
- **Disk thickness:** infinitely flat disk, i.e.  $b/a = \cos(i)$
- **Number of data set realizations:** 100

### 6.3.2.3 Data set C

- **Intrinsic scatter:** yes
- **Dust model and its parameter:** stellar mass-dependent dust model with  $\beta = 1$
- **NUV noise:** both with and without
- **Disk thickness:** infinitely flat disk, i.e.  $b/a = \cos(i)$
- **Number of data set realizations:** 100

### 6.3.2.4 Data set D

- **Intrinsic scatter:** no (i.e. the widths of the Gauss distributions used to represent the scatter of the relations used in the simulation are set to zero, causing the intrinsic colour to be a delta function, which however is still a function of stellar mass)
- **Dust model and its parameter:** simple dust model with  $\tau = 3$
- **NUV noise:** no
- **Disk thickness:** infinitely flat disk, i.e.  $b/a = \cos(i)$
- **Number of data set realizations:** 1

### 6.3.2.5 Data set E

- **Intrinsic scatter:** yes
- **Dust model and its parameter:** simple dust model with  $\tau = 3$
- **NUV noise:** both with and without
- **Disk thickness:** finite disk height, i.e.  $b/a \neq \cos(i)$
- **Number of data set realizations:** 100

Data set E is absolute identical to data set A except for the disk thickness.

### 6.3.2.6 Data set F

- **Intrinsic scatter:** yes
- **Dust model and its parameter:** stellar mass-dependent dust model with  $\beta = 1$
- **NUV noise:** both with and without
- **Disk thickness:** finite disk height, i.e.  $b/a \neq \cos(i)$
- **Number of data set realizations:** 100

Data set F is absolute identical to data set C except for the disk thickness.

## 6.3.3 Sample biases

The colour distribution suffers from a bias caused by the applied detection limits as well as by the NUV noise. Using the simulated data sets, we discuss the effect qualitatively here and will discuss the impact of this bias on the inference of the intrinsic colours and dust content later in Sect. 7.1.

Furthermore, we quantitatively discuss the bias of the inclination distribution of flux limited samples.

### 6.3.3.1 Colour bias caused by the r-band

We have defined the colour of a galaxy as a function of stellar mass. For a given stellar mass, there is a small scatter in the r-band magnitude - stellar mass relation. That is, for a given stellar mass, the r-band magnitude of a galaxy will be distributed with some spread of values. In the presence of detection limits, only part of this distribution would pass the detection limit. Specifically, the faint tail of the distribution will be excluded from the sample. This will cause the remaining data to have a narrower distribution (i.e. narrower colour distribution) and to be r-band brighter. Because the NUV luminosity is simulated as function of stellar mass (i.e. indirect luminosity-colour relation) and because surviving data points are r-band brighter, the data will also have redder NUV-r intrinsic colours. As the detection limit is placed on the observed r-band magnitudes, subsamples of the data set with high attenuation (i.e. high  $\tau$  values or high inclination) suffer more from this bias.

In general, the bias caused by the r-band detection is prominent for low stellar mass galaxies, as they are fainter in the r-band than high stellar mass galaxies.

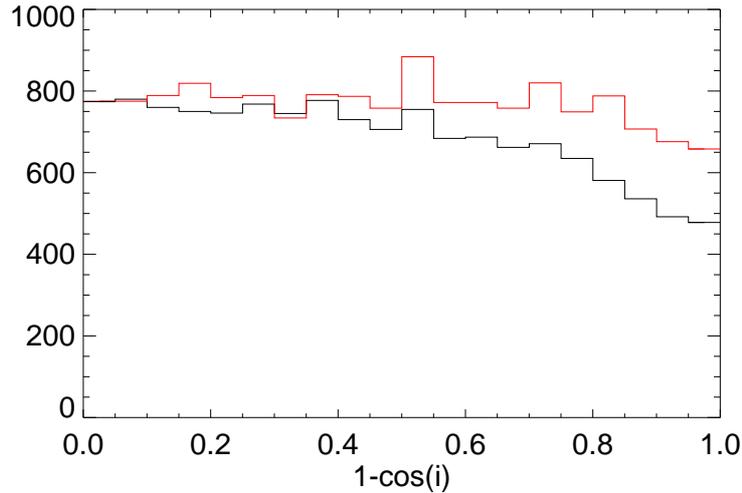


Figure 6.10: Inclination bias

Shown are the intrinsic inclination distribution of the simulated flux limited samples A.1 (black,  $\tau=3$ ) and C.1 (red, stellar mass surface density-dependent  $\tau$ ). The presence of the r-band detection limit biases the intrinsic inclination distribution against higher inclination galaxies. This bias is an increasing function of attenuation, thus it is more pronounced for higher dust content  $\tau$ .

### 6.3.3.2 Colour bias caused by NUV noise

As discussed in detail in Sect. 3.1.4, the NUV noise can be well approximated as a Gaussian function in flux. For any given r-band magnitude and stellar mass, the noise-free UV magnitude is modelled as a Gaussian distribution (through the Gaussian probability functions used in the definition of the SFR, Eq. 6.17, and  $M_s$ , Eq. 6.11). Thus, the noise-free UV flux has a log-normal distribution (which has dominant tails on the bright side).

Convolving the log-normal distribution of the noise-free UV flux with the Gaussian noise distribution, we obtain the distribution of noisy flux. This distribution will be slightly wider than the original log-normal distribution of the noise-free flux. Moreover, it will have a stronger red tail. Transforming this back into magnitude space, the noisy colour distribution is also wider and has a strong red tail, with the prominence of the red tail increasing with lower S/N.

Since edge-on galaxies suffer higher dust attenuation than face-on galaxies, they will appear at lower S/N and will therefore have a more prominent red tail in the distribution of measured colours in magnitude space than is the case for their face-on counterparts. Thus, the observed colour distribution of edge-on galaxies will not only be broadened, but also be skewed to redder values compared to face-on galaxies, even if the systematic effect of dust attenuation has been perfectly corrected.

### 6.3.3.3 Inclination bias

As shown in Sect. 5.4.1, dust attenuation is predicted to be much more severe for galaxies viewed close to the edge-on orientation than galaxies viewed more face-on

(see Fig. 6.10). This affects the relative number of more edge-on and more face-on galaxies in a flux-limited sample.

Specifically, the r-band detection limit is applied to attenuated magnitudes of the galaxies. As attenuation is inclination-dependent, the face-on population is less affected by dust dimming than the edge-on population. More faint galaxies are included in the face-on sample compared to the edge-on sample.

Fig. 6.10 shows the simulated inclination distribution for data set C.1 (i.e. only galaxies that passed the r-band detection limit; red line; stellar mass dependent dust model). The distribution is clearly suppressed for edge-on galaxies, the so-called "inclination bias". The effect becomes more severe for more opaque populations of galaxies. This can be seen for the black line, representing the simulated inclination distribution for data set A.1 (constant  $\tau = 3$ ). For a volume-limited sample, this distribution would be uniformly flat.

It is important to show that the inclination bias is also sensitive to the form of the luminosity function. This can be explicitly seen from Eq. A.25 in the Appendix, which defined the probability of seeing a galaxy with dust-attenuated absolute r-band magnitude  $M_d$  at a comoving distance  $D_C$

$$P(M_d, D_C) \propto \int \int D_C^2 \cdot P_{SF}(M_d - A(\tau, i) - A_{corr} - k_{corr}|\phi^*, \alpha, M^*) \cdot \underbrace{\eta(M_d + 5[\log_{10}(D_L) - 1])}_{m_d} \cdot P_A(A(\tau, i)) d\tau di \quad (6.36)$$

where  $P_{SF}$  is the normalized Schechter function as defined in Eq. A.4,  $\eta$  is the Heavyside function of the r-band detection limit (Eq. A.16), and  $P_A(A(\tau, i)) = P_\tau(\tau) \cdot P(i)$  is the probability distribution of the dust attenuation as a function of inclination  $i$  and dust content  $\tau$  (Eq. A.14).

To derive the inclination distribution of the flux-limited data set, we modify Eq. A.25, such that we do not integrate over  $i$ , but over  $\tau$ ,  $M_d$ , and  $D_C$ .

$$P(i|\alpha, M^*) \propto \int \int \int D_C^2 \cdot P_{SF}(M_d - A(\tau, i) - A_{corr} - k_{corr}|1, \alpha, M^*) \cdot \underbrace{\eta(M_d + 5[\log_{10}(D_L) - 1])}_{m_d} \cdot P_\tau(\tau) \cdot P(i) dM_d D_C d\tau \quad (6.37)$$

The simulated data set automatically has this observed inclination distribution due to the forward modelling. No explicit constraint needs to be made to generate the observed inclination distribution.

Note that the inclination distribution of any flux-limited sample will differ from the intrinsic inclination distribution due to any detection limit, as soon as the luminosity function is not flat. Moreover, even a volume-limited sample (i.e. for which the whole dataset will follow the flat intrinsic inclination distribution) will have a flux-dependent inclination distribution if it has a non-flat luminosity function. Assuming a Schechter luminosity function (with  $\alpha < -1$ ), considering any given observed magnitude in the range of the data sample, we will see more low-inclined galaxies because they experienced less attenuation than high-inclined galaxies. In other words, at the same observed magnitude, low-inclination galaxies

have fainter intrinsic magnitudes and the Schechter function predicts more galaxies with fainter intrinsic magnitudes<sup>15</sup>.

---

<sup>15</sup>For  $\alpha > -1$ , one can make a similar argument, but must analyse separately the bright and the faint end of the Schechter function, as one has a positive and the other a negative slope.

# Chapter 7

## Statistical inference of the attenuation of starlight by dust in spiral galaxies

In this chapter we develop the principal elements of a statistical method for the elucidation of the attenuation of UV starlight by dust in statistical samples of morphologically-selected disk galaxies. Although here applied to the GALEX-GAMA measurements of GAMA galaxies, the method will be applicable to the determination of attenuation by dust at any UV or optical wavelength and for any flux-limited optical survey of disk galaxies for which redshifts, measurement of optical size and ellipticity via single Sérsic fits are available.

### 7.1 Method

The method operates on disk galaxies, here selected according to the photometric proxies for the morphology described in Sect. 5.2.1 (i-band attenuated magnitude  $M_{i,d}$ , effective radius  $r_{eff}$ , and Sérsic index  $n$ ). The method calculates, as a function of stellar mass, the predicted distributions of NUV-r colour in the observed flux-limited sample, for simple hypotheses for the probability distribution of the intrinsic colour of the parent population of disk galaxies at fixed  $M_s$  and for simple hypotheses for the attenuation of starlight by dust. Values for  $M_s$  are taken from Taylor et al. (2011), see Sect. 5.3.1.

As will be described in Sect. 7.3.1, the hypothesis for the intrinsic NUV-r colour is taken to be a log-normal distribution (normal in magnitude space) at fixed  $M_s$ , with a mean colour taken to be a function of  $M_s$ . The two dust models considered are both parameterised through the B-band face-on optical depth  $\tau$  of the radiation transfer model of Popescu et al. (2011), as described in Sect. 5.4, which uses information on morphology from the optical single Sérsic fits as constraints. The first dust model, used for reference, assumes a fixed value of  $\tau$  for all galaxies, whereas the second model predicts  $\tau$  from the observed surface density of  $M_s$ .

For trial parameter sets (for predicting intrinsic colour and  $\tau$ ) for each model, the likelihood of the observed colour of a galaxy is calculated, given the observed size and axis ratio of the galaxy. A joint likelihood function for the whole GAMA flux-limited sample is then constructed, and maximised to find preferred values of the model parameters and to intercompare the two models.

### 7.1.1 Utilisation of ellipticities

A critical feature of the model is that the measured ellipticity is used to predict the true inclination of a galaxy as a probability function self-consistently with the dust model considered, and taking into account the flux limit of the survey. In principle, this allows the fundamental degeneracy between dust reddening and intrinsic colour to be broken. Our technique is a generalisation of previous analyses which have analysed the attenuation-inclination relation for optical samples of galaxies to infer dust attenuation (e.g. Giovanelli & Haynes, 1983; Driver et al., 2007; Masters et al., 2010). However, the reddening-colour degeneracy is only broken if attenuation is a strong function of inclination, which may not always be the case for real galaxies. For example, galaxies which are optically thick in the r-band as well as in the UV display only a small systematic shift in observed colour from face-on to edge-on. Because of this, it is desirable to introduce a further criterion to differentiate between intrinsic colour and colour shift due to dust.

### 7.1.2 Utilisation of scatter in the intrinsic colours

A very powerful approach, which we adopt, is to favour hypotheses for dust attenuation which predict the smallest scatter in the intrinsic NUV-r distribution of the parent population of disk galaxies. This is motivated by the simulations which show a large increase in the predicted dispersion in the observed NUV-r colour (as compared with the scatter in the intrinsic colour) at fixed  $M_s$  due to the effects of dust (see Sect. 6.2.4 and Fig. 6.8). This increased observed scatter is seen at all inclinations. Thus, potentially, optimizing parameters of the dust model to minimize the inferred spread in intrinsic colour of the parent population of galaxies can also directly constrain the attenuation of light in the face-on orientation, which is not constrained by the attenuation-inclination studies.

We note that this approach is only viable if galaxy evolution has proceeded such that there is only a small dispersion in sSFR at fixed stellar mass. This would require to a very tight main sequence relation for disk galaxies between intrinsic sSFR and  $M_s$ , in which the intrinsic scatter is smaller than the scatter imparted to the observed relation due to dust. Evidence that this is the case has been found by Grootes et al. (2013) (see also Fig. 7.1).

One might, in principle, expect this technique to break down at low stellar masses where the SF history over the previous few hundred Myr may have taken a more impulsive nature, increasing the intrinsic dispersion of the NUV emission compared to massive galaxies, and dust attenuation may be less severe. For these objects, the constraints imposed by ellipticities may still be the most effective method for breaking the colour reddening degeneracy, since at more moderate opacities one expects a stronger variation of colour with inclination than for galaxies which are optically thick at both optical and UV wavelength. These considerations provide an additional motivation for allowing the dispersion term in the intrinsic colour distribution to be determined separately for subcategories of galaxies divided according to stellar mass.

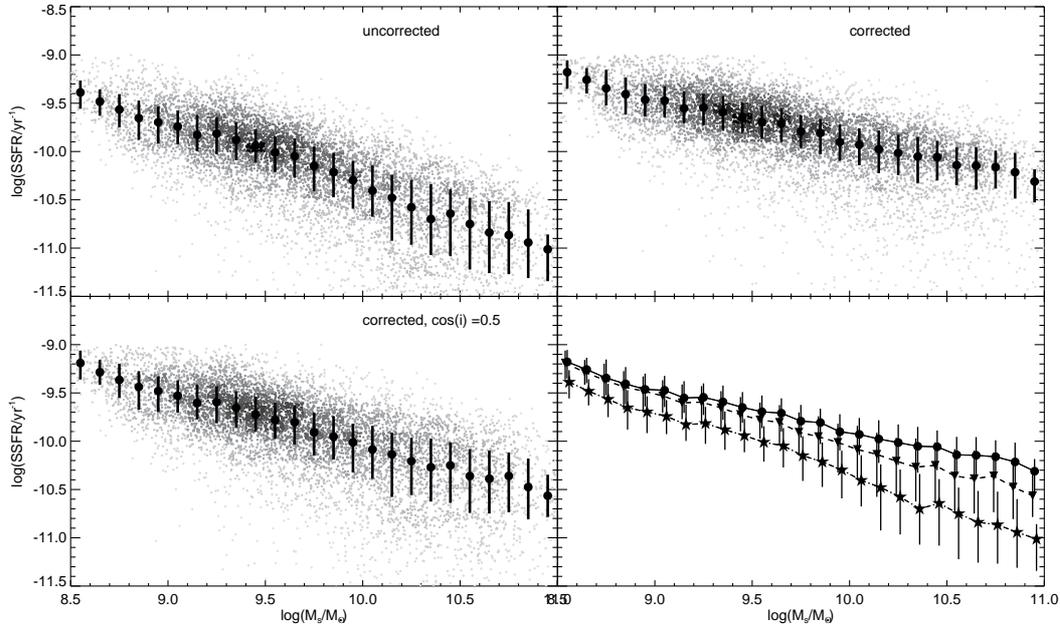


Figure 7.1: Specific star formation rate as a function of stellar mass  $M_s$  for GAMA spiral galaxies

The relation is shown before correction for attenuation by dust (top left panel), after the full inclination-dependent correction, described in Sect. 5.4, using the radiation transfer model in conjunction with the  $\Sigma_s$ - $\tau$  relation and the inclination derived from the observed axis ratio (top right panel), and after a partial correction artificially setting a uniform inclination  $i$  with  $\cos(i) = 0.5$  for all galaxies (lower left panel). The sources are binned in 15 bins of equal size in  $M_s$ , with the median depicted by a filled circle, and the bars showing the interquartile range. The scatter in the relation due to the scatter in the NUV is reduced from the uncorrected to the fully corrected case. The intrinsic values of sSFR are shifted upwards w.r.t. the uncorrected values. The linear gray-scale shows the number density of sources at that position, with the same scale having been applied to all samples. The median values and interquartile ranges are shown together in the bottom right panel. The uncorrected values are depicted by stars and a dash-dotted line, the values corrected at a fixed inclination of  $\cos(i) = 0.5$  are shown as inverted triangles and a dashed line, and the fully corrected values are shown as circles and a solid line. The bin centres have been offset by 0.01 in  $\log(M_s)$  for improved legibility. Figure taken from Grootes et al. (2013).

### 7.1.3 Present limitations

At present, the implementation of the model is not yet complete, resulting in limitations to its use and predictive power. Firstly, the use of information on measurement errors has not yet been incorporated into the analysis. For this reason (as will be motivated in Sect. 7.5.3), the analysis of the GAMA galaxies is limited to a bright subsample, selected for  $m_{r,obs} < 17$  mag. This limits the sample to 1155 galaxies.

Secondly, the model does not yet take into account the r-band selection limit in a fully self-consistent way. Specifically, the fact that the r-band flux-limited sample will preferentially be biased against galaxies with low intrinsic r-band luminosities is presently ignored. This amounts to neglecting the weighting term  $P(m_{r,intr}|m_{r,obs}, \phi)$  in Eq. 4.1 of Sect. 4.

Despite these limitations, the results from the model in the present incomplete form are still meaningful when compared with results of the same incomplete model applied to simulated data. This is done in Sect. 7.7.

In preceding Sections, we described the inference of the intrinsic inclination  $i$  from the observed  $b/a$  (Sect. 7.2), the derivation of the likelihood function as function of the intrinsic colour parameters (Sect. 7.3), and the derivation of the likelihood function for both colour and dust parameters (Sect. 7.4). Tests of the method on simulated data are shown in Sect. 7.5, and the ability to discriminate between different dust models is verified, also using simulated data, in Sect. 7.6. The application to the GAMA spiral galaxy sample is presented and the results are discussed in Sect. 7.7.

## 7.2 Inference of inclination from observed $b/a$

In Sect. 6.2.1.2, we discussed how we model the observed axis ratio  $b/a$  for a given intrinsic inclination  $i$ . For the analysis however, we face the opposite problem. We have a measured  $b/a$  and want to know the intrinsic inclination of that galaxy.

In most published works, either the simplified assumption of infinitely thin disk galaxies without an intrinsic height, bulges and dust content is made, thus  $b/a = \cos i$ , or the Hubble formula is used, taking into account a finite disk thickness (e.g. Driver et al., 2007).

In this work, we make use of our simulations, which shows us how significantly the observed  $b/a$  differs from the inclination. Specifically, we use our simulation to constrain a lookup table mapping a given  $b/a$  to a probability distribution of intrinsic inclination.

We derive a 2D histogram of  $b/a$  and intrinsic inclination for a simulated data set (using 100 times as many galaxies as usual to reduce the sampling noise). As we require a normalized probability distribution of the intrinsic inclination for a given  $b/a$ , we normalize each  $b/a$  row of the 2D histogram independently<sup>1</sup>. Thus, we obtain a table with which we can reconstruct the probability distribution of the intrinsic inclination for a given galaxy.

---

<sup>1</sup>It should be noted that this lookup table will not give a properly normalized probability distribution of  $b/a$  for a given inclination.

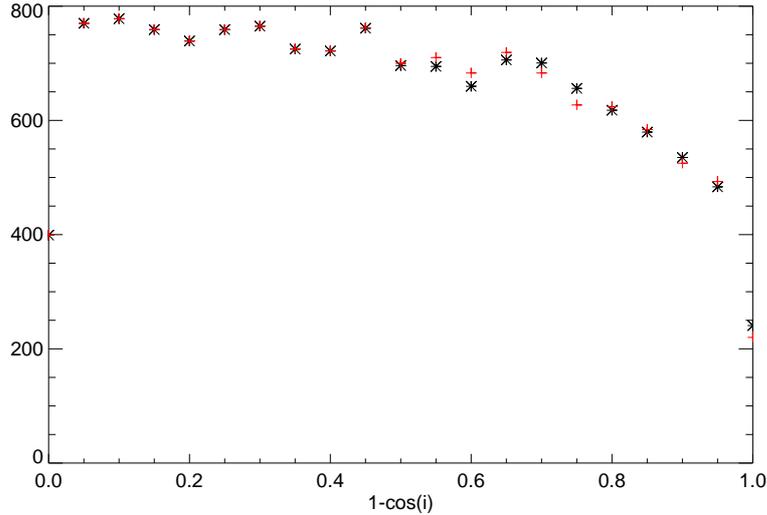


Figure 7.2: Recovering the inclination from the  $b/a$

This is a comparison between the distribution of the observed inclination  $i$  recovered from the  $b/a$  values of each galaxy (red) with the distribution of the intrinsic inclination (black), made for the simulated data set E.1. Both distributions agree very well. The small deviations are caused by the sampling noise of the considered simulated data set. Note that only half of the bin range of the first and last bin are in the region with defined inclination values. Thus they are only half populated.

Figure 7.2 displays the intrinsic inclination distribution (black) and the inclination distribution recovered from the  $b/a$  values of the individual galaxies (using the 2D lookup table) for the simulated data set E.1 (defined in Sect. 6.3.2.5). Both distributions agree very well. The sampling noise in data set E.1 causes only minor deviations. Note that both distributions are not expected to be flat in  $1 - \cos(i)$  due to the inclination bias. The edge-on galaxies are suppressed in a flux-limited sample, the exact shape of the inclination distribution  $P(i)$  being a function of the derived shape of the luminosity function, the detection limit and the attenuation (see Sect. 6.3.3.3).

Alternatively, one can use an analytic expression for the probability distribution of the intrinsic inclination which we want to derive in the following. Unfortunately, the evaluation of the analytic expression is very expensive due to the need to calculate  $P(i)$ , and was therefore not implemented in this work. Nevertheless, it is instructive to derive this expression as follows:

As defined in Eq. 6.31, an expression for the observed axis ratio  $b/a$  which successfully predicts the observed probability distribution of  $b/a$  is

$$b/a = \max[\cos(i), 0.2 + G(0.1)]$$

We reformulate this as a probability distribution  $P(b/a|i)$ , and distinguish three cases:  $b/a < \cos(i)$ ,  $b/a > \cos(i)$ , and  $b/a = \cos(i)$ . The first case,  $b/a < \cos(i)$ , has by definition the probability zero, as Eq. 6.31 forces  $b/a \geq \cos(i)$ . The second case,  $b/a > \cos(i)$ , is the case where  $b/a$  is defined via the intrinsic scale height, i.e. it is sampled from the distribution  $0.2 + G(0.1)$ . The probability of  $b/a$  is

a Gaussian function with mean 0.2 and standard deviation 0.1. We will use the symbol  $g(x, 0.2, 0.1)$  to refer to the value of the Gaussian function with mean 0.2 and standard deviation 0.1 at the position  $x$ . The third case,  $b/a = \cos(i)$ , is the case where  $\cos(i)$  is larger or equal to the sampled  $dz/a$  value from  $0.2 + G(0.1)$ . Thus, the probability of sampling a value  $x \leq \cos(i)$  from the Gaussian function with mean 0.2 and standard deviation 0.1 is equal to  $\int_{-\infty}^{\cos(i)} g(x, 0.2, 0.1) dx$ .

Expressed as a single equation, this reads<sup>2</sup>:

$$P(b/a|i) = \begin{cases} g(x, 0.2, 0.1) & \text{if } b/a > \cos(i) & \text{high inclination case} \\ \int_{-\infty}^{\cos(i)} g(x, 0.2, 0.1) dx & \text{if } b/a = \cos(i) & \text{low inclination case} \\ 0 & \text{if } \cos(i) > b/a & \text{forbidden case} \end{cases} \quad (7.1)$$

Eq. 7.1 is normalized by definition:

$$\int_{-\infty}^{\infty} P(b/a|i) d(b/a) = \int_{-\infty}^{\cos(i)} g(x, 0.2, 0.1) dx + \int_{\cos(i)+\epsilon}^{\infty} g(x, 0.2, 0.1) dx = 1 \quad (7.2)$$

where  $\epsilon$  is a positive number close to zero.

We now employ Bayes' theorem to infer the probability distribution if the intrinsic inclination given the observed  $b/a$ :

$$P(i|b/a) = \frac{P(b/a|i) \cdot P(i)}{P(b/a)} \quad (7.3)$$

Since there are no constraints on  $b/a$ ,  $P(b/a)$  is flat (in the range of  $0 \leq b/a \leq 1$ ) and we can focus on the numerator. In Eq. 6.27, the sampling prescription of  $b/a$  is defined.

While the  $P(i)$  of the population is flat in  $d \cos(i)$ , the inclination distribution is biased in the presence of attenuation and an detection limit. For the flux-limited sample,  $P(i)$  is defined in Eq. 6.37.

$$P(i|b/a) = c(b/a) \cdot \begin{cases} g(b/a, 0.2, 0.1) \cdot P(i) & \text{if } b/a > \cos(i) & \text{edge-on case} \\ \int_{-\infty}^{\cos(i)} g(x, 0.2, 0.1) dx \cdot P(i) & \text{if } b/a = \cos(i) & \text{low inclination case} \\ 0 & \text{if } \cos(i) > b/a & \text{forbidden case} \end{cases} \quad (7.4)$$

where  $c(b/a)$  is the  $b/a$  dependent normalization constant of  $P(i|b/a)$ .

### 7.3 Likelihood of the intrinsic colour distribution

We define a probabilistic model for the intrinsic colour distribution in Sect. 7.3.1. The optimization for the intrinsic colour parameters is described in Sect. 7.3.2.

---

<sup>2</sup>It should be noted that, in principle,  $b/a$  is limited to the interval  $[0,1]$ . Therefore, in Eq. 6.31, instead of a Gaussian distribution,  $G(0.1)$  should be a truncated Gaussian function normalized over the range  $[0,1]$ . In practice however, as already discussed in Sect. 6.2.1.2, the probability of  $b/a$  larger than 1 is negligible and the case of sampling negative axis ratio is prohibited since  $\cos(i) > 0$ . We therefore refrain from defining  $G(0.1)$  as a truncated Gaussian. Nevertheless, the outlined derivation is also valid for  $G(0.1)$  representing a normalized truncated Gaussian function.

Test of the method and a derivation of the uncertainty of the intrinsic colour parameters are presented in Sect. 7.3.3.

### 7.3.1 Intrinsic colour distribution

Motivated by the observed colour distribution of the GAMA spiral galaxy sample (see Fig. 6.9 and Sect. 6.2.5), we hypothesize that the intrinsic colour probability distribution for a galaxy of stellar mass  $M_s$ ,  $P(c|M_s)$  (with  $c = M_{NUV,intr} - M_{r,intr}$ ) can be expressed as a Gaussian distribution:

$$P(c|M_s) = \frac{1}{\sqrt{2\pi}\sigma_g(M_s)} e^{-\frac{1}{2}\left(\frac{c-\mu_g(M_s)}{\sigma_g(M_s)}\right)^2} \quad \text{with} \quad \int_{-\infty}^{\infty} P(c) dc = 1 \quad (7.5)$$

Here,  $\mu_g(M_s)$  and  $\sigma_g(M_s)$  denote the expectation value and standard deviation of the distribution at any fixed value of  $M_s$ . Since the dependence of  $\mu_g$  and  $\sigma_g$  on  $M_s$  is unknown (this being completely degenerate with the to be determined dependence of disk opacity on  $M_s$ ) we seek to derive it empirically in a non-parametric way by dividing the galaxy sample into bins of  $M_s$  and solving for constant values of  $\mu$  and  $\sigma$  with in each bin:  $\mu = \mu_g(M_s)$  and  $\sigma = \sigma_g(M_s)$  for a bin centred at stellar mas  $M_s$ . The PDF of the colour distribution can then be written

$$P(c|M_s) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{1}{2}\left(\frac{c-(\mu+f(M_s))}{\sigma}\right)^2} \quad \text{with} \quad \int_{-\infty}^{\infty} P(c) dc = 1 \quad (7.6)$$

Here, the function  $f(M_s)$  accounts for the stellar mass dependence of the mean colour over the bin in  $M_s$ . This has to be included since the change in the mean colour over the bin would otherwise be misinterpreted as increase in the standard deviation  $\sigma$ .  $f(M_s)$  is treated as prior knowledge. In practice, the binwise intrinsic colour fit is done iteratively in two steps. In a first step,  $f(M_s)$  is equated to 0 in all bins and a first estimate for  $\mu$  is found for each bin using galaxies with low inclination only<sup>3</sup>. The function  $f(M_s)$  is then approximated by a quadratic fit to the values of  $\mu$  found for all stellar mass bins in the first step. In the second step, we solve for  $\mu$  and  $\sigma$  with this constraint on  $f(M_s)$ . In essence, this amounts to a colour correction for the derived values of  $\mu$  and  $\sigma$ . In the following, it will be convenient to refer to these colour corrected values as  $\mu^c$  and  $\sigma^c$ . The iterative procedure is outlined in detail in Sect. 7.4.

To derive the likelihood that a galaxy has an observed colour  $c_{obs}$ , we first derive the intrinsic magnitudes in the r-band and the NUV by correcting the observed magnitudes for dust attenuation (using the hypothesis for  $\tau$ ), and applying A- and k-corrections (see Sect. 6):

$$M_{intr} = M_{obs} - A(\tau, i) - A_{corr} - k_{corr} \quad (7.7)$$

In terms of the intrinsic colour  $c$ :

$$\underbrace{M_{NUV,intr} - M_{r,intr}}_c = \underbrace{M_{UV,obs} - M_{r,obs}}_{c_{obs}} - A_{NUV}(\tau, i) - A_{NUV,corr} + A_r(\tau, i) - k_{NUV,corr} + A_{r,corr} + k_{r,corr} \quad (7.8)$$

<sup>3</sup>In practice, we order the galaxies according to their observed  $b/a$  ratio and only consider the first 1/3 of the galaxies (i.e. most face-on).

The dust attenuation  $A(\tau, i)$  for each galaxy is fully defined by the considered dust model (parameters, i.e.  $\tau$ ) and the inclination.

Since the corrections applied Eq. 7.8 only shift the colour, but do not change the shape of the PDF, and in the absence of NUV noise, the probability distribution of the observed colour is:

$$P(c_{obs}|\mu, \sigma, i, \tau, M_s) = P(c|\mu, \sigma, M_s) \quad (7.9)$$

where  $\mu$  and  $\sigma$  refer to the values of each stellar mass bin,  $\tau$  is calculated individually for each galaxy according to its stellar mass, and  $P(c|\mu, \sigma, i, \tau, M_s)$  is defined in Eq. 7.6.

However, the inclination of each galaxy is expressed as a probability distribution  $P(i|b/a, \tau)$  as a function of the observed axis ratio  $b/a$  using the lookup-table presented in Sect. 7.2. Therefore, we need to convolve the likelihood as defined in Eq. 7.9 with the probability distribution of the inclination  $P(i|b/a, \tau)$ .

$$P(c_{obs}|b/a, \mu, \sigma, \tau, M_s) = \int_0^{\pi/2} P(c_{obs}|\mu, \sigma, i, \tau, M_s) \cdot P(i|b/a, \tau) di \quad (7.10)$$

We can adapt Eq. 7.10 easily for NUV non-detections. While a detected galaxy has a known value for the observed colour  $c_{obs}$ , the NUV non-detections have an “observed” NUV magnitude fainter than the individual detection limit  $M_{detlim}$ .

$$P(c_{obs}|b/a, \mu, \sigma, \tau, M_s) = \int_{c_{detlim}}^{\infty} \int_0^{\pi/2} P(c_{obs}|\mu, \sigma, i, \tau, M_s) \cdot P(i|b/a, \tau) di dc_{obs} \quad (7.11)$$

with  $c_{detlim}$  being the “observed” colour corresponding to the NUV detection limit given the observed r-band magnitude.

Applying Bayes theorem and assuming no further constraints on  $P(\mu, \sigma)$  and  $P(c_{obs}, b/a)$ <sup>4</sup> (i.e. flat priors), we can use

$$P(\mu, \sigma|c_{obs}, b/a, \tau, M_s) = \int_0^{\pi/2} P(c_{obs}|b/a, \mu, \sigma, \tau, M_s) \cdot P(i|b/a, \tau) di \quad (7.12)$$

$$P(\mu, \sigma|c_{detlim}, b/a, \tau, M_s) = \int_{c_{detlim}}^{\infty} \int_0^{\pi/2} P(c_{obs}|b/a, \mu, \sigma, \tau, M_s) \cdot P(i|b/a, \tau) di dc_{obs} \quad (7.13)$$

as the likelihood of the intrinsic colour parameters  $\mu$  and  $\sigma$  given the observed data for NUV detections and NUV non-detections respectively.

As will be shown in Sect. 7.5.3, it is desirable to account for the NUV noise in the intrinsic colour probability calculation. For this, we refer the reader to the outlook section B.

---

<sup>4</sup>In case of the observed colour  $c_{obs}$ , this assumption is over-simplified, as the r-band detection limit introduces a bias against blue galaxies, which is a function of the r-band luminosity function, the detection limit, and the attenuation. Furthermore, the observed  $b/a$  distribution depends via  $P(i)$  (Eq. 6.37) on the dust attenuation, the r-band LF, the r-band detection limit, the r-band apparent magnitude, and the distance of the considered galaxy.

### 7.3.2 Maximum likelihood fit of the colour

Eq. 7.12 defines the likelihood function of the intrinsic colour distribution for an individual galaxy. The best-fitting parameter combination of  $\mu$  and  $\sigma$  is the one which maximizes the joint likelihood  $p(x)$  of the considered galaxy sample  $x$  (i.e. all galaxies in the considered stellar mass bin).

$$p(\mu, \sigma|x) = \prod_i P(\mu, \sigma|x_i) \quad (7.14)$$

where  $x_i$  represents the data of a single galaxy. For convenience and numerical stability, we use the logarithm of the likelihood rather than the likelihood itself.

$$\log p(\mu, \sigma|x) = \sum_i \log P(\mu, \sigma|x_i) \quad (7.15)$$

Eq. 7.15 is then maximized with respect to  $\mu$  and  $\sigma$  using a gradient method.

A first estimate of the parameters can be obtained directly from calculating the first and second moment of the observed colour distribution, i.e. assuming  $c = M_{UV,obs} - M_{r,obs}$ . This estimate is of course biased due to the noise in the UV and the dust attenuation in both bands changing the colour distribution. Nevertheless, it provides a good first estimate of the parameter values and is thereby an excellent starting point for the optimization routine. Applying A- and k-corrections to both bands before fitting the observed distribution, and assuming a mean attenuation from literature (e.g.  $\tau = 3.8 \pm 0.7$  found by Driver et al. (2007)), the starting point can even be improved.

It should be noted that the joint likelihood combining all galaxies in all stellar mass bins can be easily derived from the joint likelihood  $p_j(\mu_j, \sigma_j|x_j)$  of each individual stellar mass bin  $j$  as defined in Eq. 7.14.

$$p(\underline{\mu}, \underline{\sigma}|x) = \prod_j p(\mu_j, \sigma_j|x_j) = \sum_j \log p(\mu_j, \sigma_j|x_j) \quad (7.16)$$

where  $\underline{\mu}$  and  $\underline{\sigma}$  are the vectors containing the individual  $\mu_j$  and  $\sigma_j$  values of the stellar mass bin  $j$ .

### 7.3.3 Tests on simulated data and uncertainty of $\mu$ and $\sigma$

To demonstrate the accuracy of the fit parameter derived by maximising the joint likelihood in Eq. 7.14, we fit simulated data set B (consisting of 200 individual data sets, 100 with 10 times more sources than the other 100, see Sect. 6.3.2.2) and show the resulting best fitting colour parameter  $\mu$  and  $\sigma$  in Fig. 7.3. The input values of the simulation are shown as black lines. One can easily see that the fit is able to recover the input parameter.

The black asterisks represent the data sets containing  $\sim 1540$  galaxies each, the red asterisks represent the data sets containing  $\sim 15400$  galaxies each. The mean values and their uncertainties are  $\mu = 1.4000 \pm 0.0051$  and  $\sigma = 0.6003 \pm 0.0035$  for the large data sets (red) and  $\mu = 1.4000 \pm 0.0145$  and  $\sigma = 0.5992 \pm 0.0105$  for the small data sets (black). The uncertainty of the derived fit parameters are completely dominated by the sampling noise of the data and scale as expected with  $1/\sqrt{N}$  (the number of galaxies  $N$  in each sample). Thus, the uncertainties can be scaled to the size of any data set.

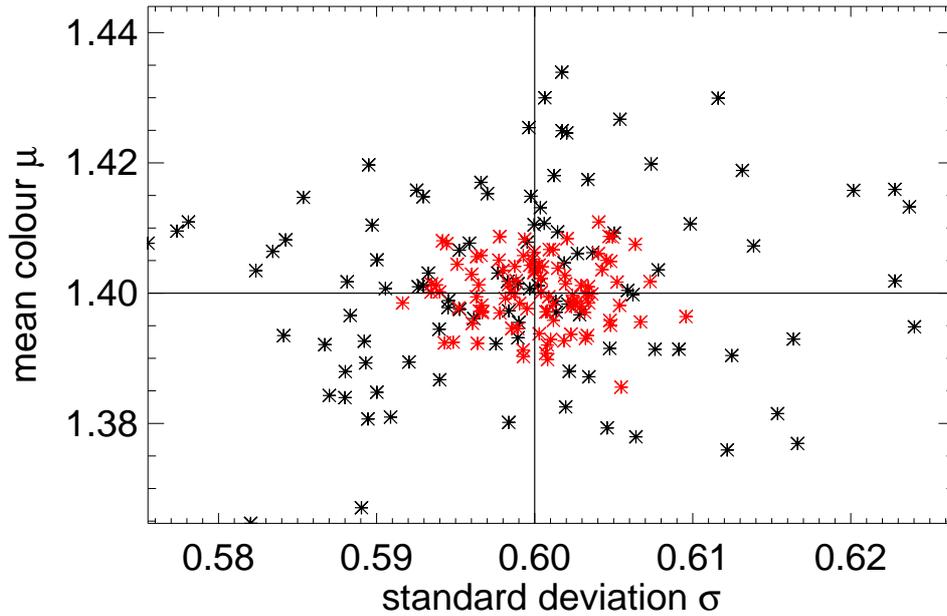


Figure 7.3: Intrinsic colour fit results for noise-free simulated data set B

## 7.4 Deriving the best-fitting parameters of a given dust model

In Sect. 7.3, we described how the intrinsic colour parameter  $\mu^c$ ,  $\sigma^c$  and the likelihood are derived for a given dust model specified by  $\tau$ . To derive the best fitting parameters of a given dust model, this procedure is repeated for different hypotheses for  $\tau$  and the resulting likelihoods are compared.

In principle, this procedure can be done on the basis of a single likelihood value generated from the whole data set. However, in practice, we do not know whether we are considering a correct model for the transfer of radiation in galaxies. If our model was incorrect, the maximum likelihood solution would likely correspond to wrong values for the mean and the standard deviation in intrinsic colour  $\mu^c$  and  $\sigma^c$ . Therefore, in practice, we divide the data sample into three ranges of inclination<sup>5</sup>, and derive separate solutions for  $\mu^c$  and  $\sigma^c$  in each inclination range. A wrong model can then be identified if the solutions for  $\mu^c$  and  $\sigma^c$  vary systematically with inclination. Specifically, we compute maximum likelihood values for  $\mu^c$  and  $\sigma^c$  over a 40 x 3 grid of 40 bins in  $M_s$  and 3 inclination bins, with the bin boundaries set such that there is roughly equal occupancy of data in each bin. This calculation is performed for a given list of trial values for the dust model parameter(s) from the second step described above.<sup>6</sup> In theory, the correct dust model parameter(s) will cause the mean colour  $\mu^c$  to scatter around 0, the standard deviation  $\sigma^c$  to be reduced to the intrinsic standard deviation and the likelihood to be maximized, simultaneously for all inclination bins. A wrong parameter (or a wrong model for

<sup>5</sup>Inclination bins were separated according to the observed  $b/a$  ratio.

<sup>6</sup>It should be noted that since the uncertainties of the derived intrinsic colour parameter scale with number of data points used, all 120 bins have the same parameter uncertainties and we do not need to assign different weights to different bins during the analysis.

attenuation) will cause an inclination-dependent trend in these quantities (increase in  $\sigma^c$  / decrease in likelihood with inclination and a deviation of the mean colour  $\mu^c$  of the higher inclined bin from the face-on bin).

A further aspect of the determination of dust parameters is that, while the maximisation of the likelihood function makes effective use of the information we have on the inclination-dependent component of attenuation, it does not strongly discriminate between different prescriptions for the face-on component of attenuation. In order to address this, we make use of the standard deviation  $\sigma^c$  in intrinsic colour, for the identification of dust parameters. Specifically, as motivated in Sect. 7.1.2, we seek to minimize  $\sigma^c$  as a function of the trial dust parameters. This makes use of a prior expectation of the existence of fundamentally tight scaling relations in nature, linking intrinsic emission, or colour, to physical quantities such as stellar mass.

Bearing all these considerations in mind, we therefore identify three criteria, by which we judge the best-fitting dust model parameter:

- Maximization of the joint (logarithmic) likelihood over all bins:  $\sum \log(\text{likelihood})$ . In the following, we plot this criterion as function of the dust model parameter using  $(-1) \cdot \sum \log(\text{likelihood})$ . Thus, we will minimize the negative joint logarithmic likelihood for easier visual comparison of the position of the minimum with the other two criteria.
- Minimization of the standard deviation  $\sigma^c$  averaged over all bins:  $\langle \sigma^c \rangle = \frac{\sum \sigma^c}{n_{bin}}$ . This criterion is sensitive to the change of attenuation with  $\tau$  in both the face-on and the inclination-dependent component.
- Minimization of the rms of the stellar mass trend-removed mean colour  $\mu^c$  over all bins:  $\mu_{rms}^c = \sqrt{\frac{\sum (\mu^c)^2}{n_{bin}}}$ . This criterion is only sensitive to the inclination-dependent component of the attenuation, as the face-on mean colour is subtracted, and with it the face-on component of the attenuation. It is useful to provide a visual check on the inclination-dependence of the solutions.

Finally, we note that our grid search technique allows a non-parametric determination of the dust parameters as a function of  $M_s$ . However, since in the practical application to real data we cannot yet account for noise in the NUV, we are presently limited to an evaluation of bright galaxies (see Sect. 7.1.3 and 7.5.3) with insufficient statistics for such a non-parametric evaluation of  $\tau(M_s)$ .

## 7.5 Tests on simulated data

As mentioned above, the criteria used to identify the best-fitting dust model parameters are subject to different biases. In this section, we illustrate both the method itself and its biases using different simulated data sets. With each test presented, we increase the complexity of the simulated data in order to demonstrate which bias is caused by which feature of the simulated data.

### 7.5.1 Parameter estimation for simulated data with scatter-free intrinsic colour

To demonstrate the bias caused by the dependence of the mean colour  $\mu$  on the stellar mass across the stellar mass bin, we derive the intrinsic colour parameters of the noise-free data set D (see Sect. 6.3.2.4), which has an intrinsic colour distribution with the stellar mass-dependent mean colour  $\mu$  and standard deviation  $\sigma$  set to 0 (i.e. a delta function in colour for a given stellar mass). Fig. 7.4 illustrates the two steps of the colour fit by showing the intrinsic colour parameters  $\sigma$ ,  $\mu$ , and likelihood derived in the first step (black) and  $\sigma^c$ ,  $\mu^c$ , and the likelihood derived in the second step (red) as a function of stellar mass and inclination. The trend of  $\mu$  with stellar mass is removed in the second step. The parameters  $\sigma$  and likelihood are clearly biased for large stellar mass bins, even for the small bin width adopted. This bias is almost entirely removed for  $\sigma^c$  and the likelihood in the second step. Since also the mean colour  $\mu$  of each stellar mass bin is slightly biased in the first step and these biased values were used for the fit of the stellar mass dependency, a small bias remains in the second step as well. However, the remaining bias is negligible compared to the intrinsic scatter and the scatter due to UV noise, as shown in the following section.

Fig. 7.5 shows the results of the second step for the input dust model parameter  $\tau = 3$  (red) and a false  $\tau = 2$  (black). As expected, the intrinsic colour parameters show the same trend with stellar mass for all three inclination bins if the trial value for  $\tau$  is set to the value used in generating the simulation. As is also expected, there is a clear trend with inclination (additional to the stellar mass trend) towards high  $\sigma^c$ , redder  $\mu^c$  and lower likelihood if the wrong trial value  $\tau = 2$  is used. The difference in the standard deviation  $\sigma^c$  between the high inclination bins and the low inclination bins ( $\sim 0.055$ ) is quite small compared to the intrinsic standard deviation used in the other simulated data sets and expected for the real data ( $\sim 0.4\text{mag}$ ; see following section). Even though the attenuation of a single band varies strongly with  $\tau$  and inclination, the attenuation of the NUV-r colour shows a much smaller dependency since the attenuation curves for the individual bands are roughly parallel over the width of our inclination bins (see Fig. 5.9). Thus, the scatter in the colour introduced by a wrong  $\tau$  value is relatively small<sup>7</sup>. The stronger dependence with  $\tau$  and inclination is seen in the reddening of the NUV-r colour as shown in the middle of Fig. 7.5. Nevertheless, both effects contain extractable information, as shown in the following Sect. 7.5.2.

Fig. 7.6 evaluates the three criteria for identifying the best-fitting dust model parameters (see Sect. 7.4) for a list of trial  $\tau$  values. All three criteria clearly favour the input value of  $\tau = 3$  (blue line) for the intrinsic scatter free data set D.

---

<sup>7</sup>If the data follows the stellar mass-dependent model (see Sect. 5.4.2), the increase in scatter for a false scaling factor  $\beta$  is slightly higher as the model predicts a distribution of  $\tau$  values as function of stellar mass surface density. Thus, the scatter in sizes for a given stellar mass leads to a scatter in attenuation, which is only perfectly accounted for if the true dust model parameter is used.

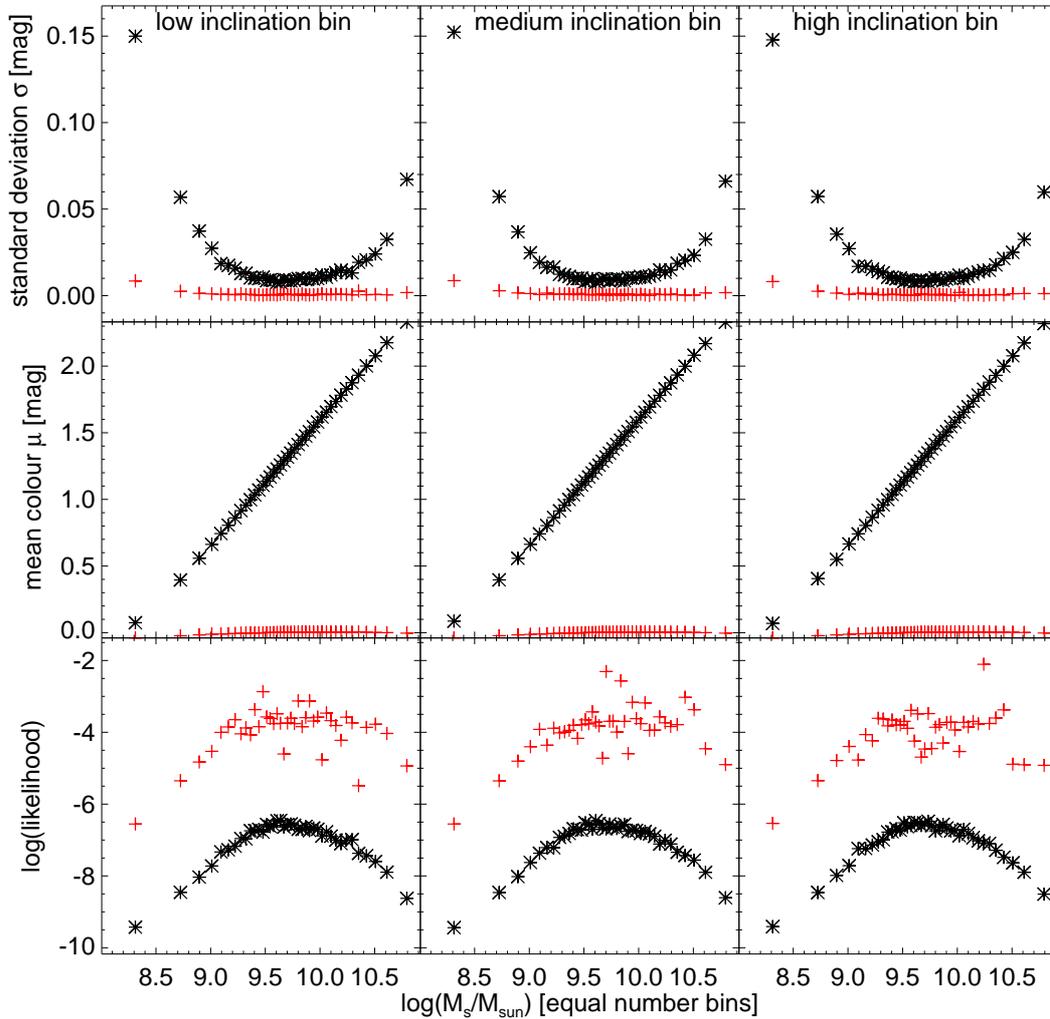


Figure 7.4: Best-fitting parameter combination for first and second step of the intrinsic colour fit.

For a data set without intrinsic scatter and without UV noise (data set D), using the correct dust model (parameter), the best-fitting intrinsic colour parameters of the first fitting step for inclination and stellar mass bins are plotted in black. The parameters found in the second step including the stellar mass dependency of the intrinsic colour are overplotted in red. As is to be expected, the strong stellar mass dependency of the intrinsic colour is removed in the second step, resulting in mean colour fluctuating around 0. The standard deviation  $\sigma$  and the likelihood exhibit especially strong biases in the first step for wide stellar mass bins. These biases are almost entirely removed in the second step.

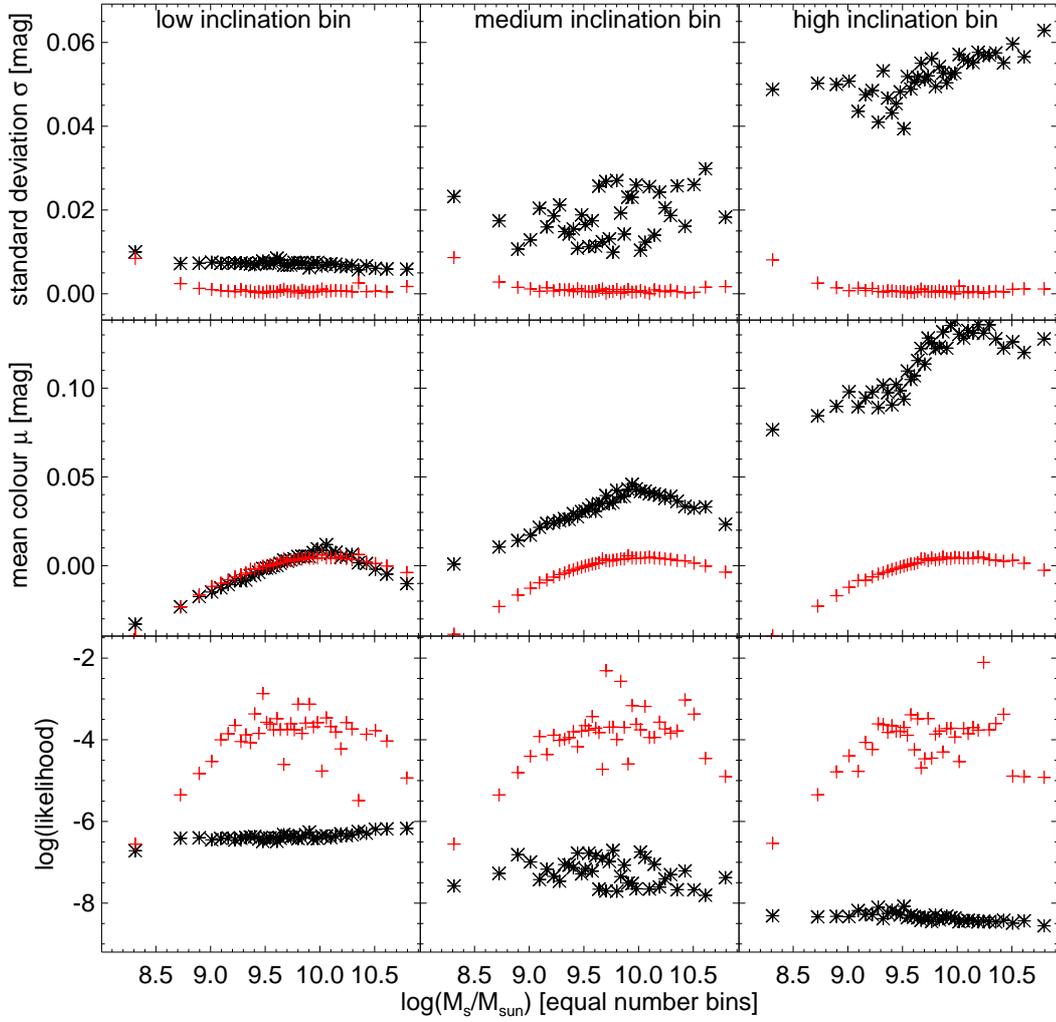


Figure 7.5: Comparison of the intrinsic colour parameters recovered for different assumed  $\tau$  for intrinsic scatter free data.

The recovered intrinsic colour parameters  $\mu^c$ ,  $\sigma^c$ , and joint likelihood are plotted for each stellar mass and inclination bin assuming either a trial value of  $\tau = 2$  (black) or a trial value  $\tau = 3$  (red; the same value used in generating this simulated data set D). The data is free of intrinsic scatter. One can see very well the inclination trend of all three parameters if a false trial value for  $\tau$  is assumed.

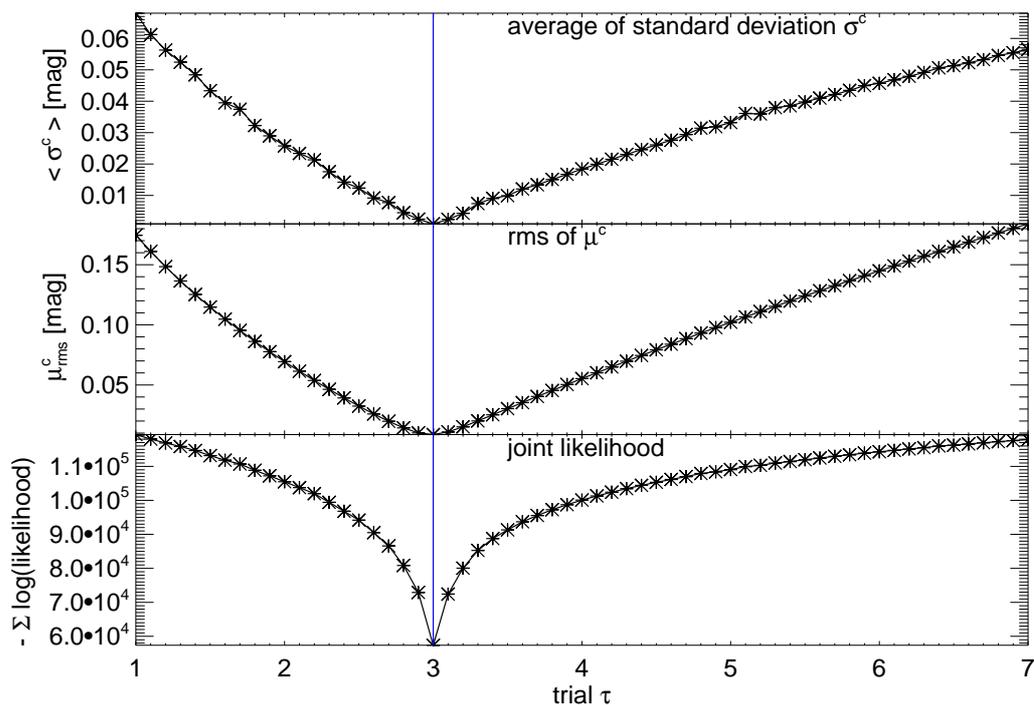


Figure 7.6: Identifying best-fitting  $\tau$  value for noise-free data without intrinsic scatter. For data set D (NUV noise-free, no intrinsic scatter, input value  $\tau = 3$ ), the values of  $\langle \sigma^c \rangle$ ,  $\mu_{rms}^c$ , and  $-\sum \log(\text{likelihood})$  are plotted as function of the trial  $\tau$  value. The position of the minimum of each of the three curves (the  $\tau$  values favoured by the particular criterion) perfectly reproduces the input value (blue line).

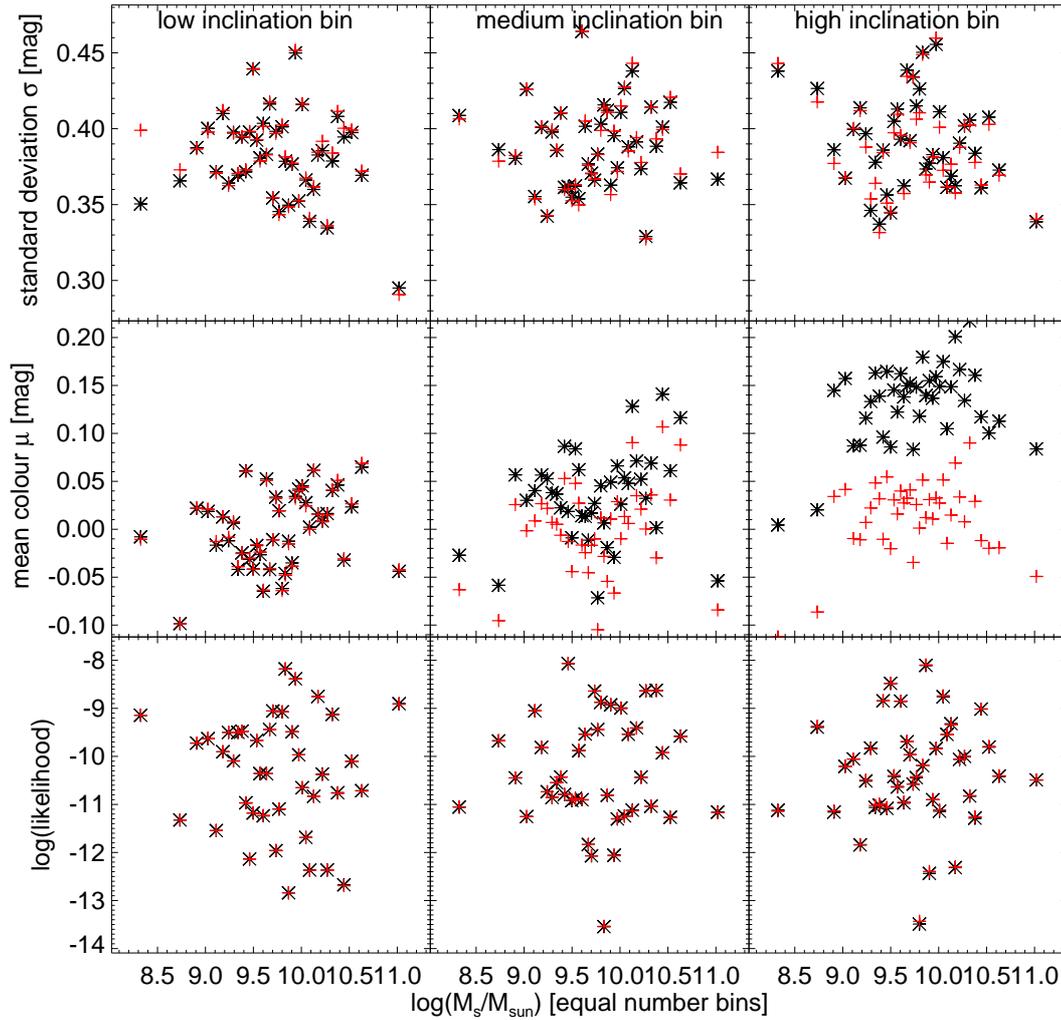


Figure 7.7: Comparison of the intrinsic colour parameters derived assuming different  $\tau$  for simulated data with intrinsic scatter

The recovered intrinsic colour parameters  $\mu^c$ ,  $\sigma^c$ , and joint likelihood for each stellar mass and inclination bin assuming a  $\tau = 2$  (black) and assuming the input  $\tau = 3$  (red) are shown for the noise-free simulated data set A.1 (with intrinsic scatter). One can clearly see the inclination trend of the  $\mu^c$  parameter if a false  $\tau$  value is assumed.

### 7.5.2 The effect of intrinsic scatter on the parameter estimation

The simulated data set A (defined in Sect. 6.3.2.1) incorporates a non-zero standard deviation of the intrinsic colour distribution ( $\sim 0.4$  mag, see Sects. 6.1 and 6.2). Thus, even for the true dust model (parameters), we expect not be able to reduce  $\sigma^c$  to (close to) zero, but only to the value of the intrinsic standard deviation. Furthermore, the data in each bin is subject to sampling noise, causing an uncertainty in the fitted intrinsic colour parameters.

The effects of intrinsic scatter in data set A are shown in Fig. 7.7, which compares the intrinsic colour parameters found in the second step (i.e. colour - stellar mass trend removed) for the true input  $\tau = 3$  (red) with the false  $\tau = 2$  (black).

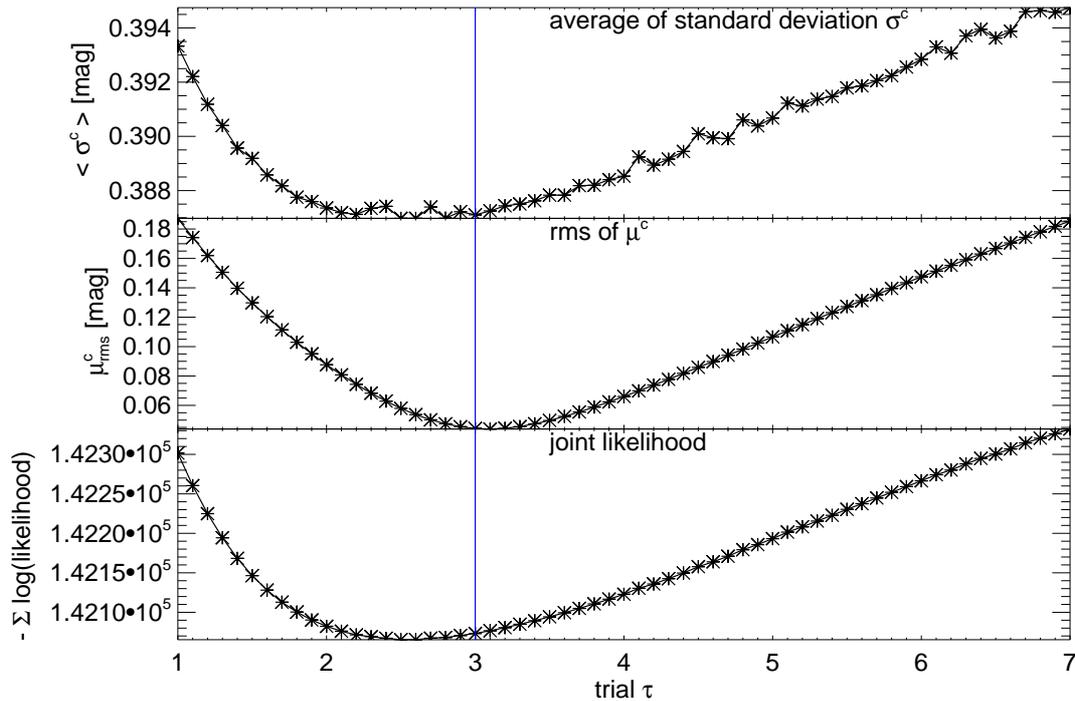


Figure 7.8: Identifying best-fitting  $\tau$  value for noise-free data

For data set A.1 (NUV noise-free, intrinsic scatter, input value  $\tau = 3$ ), the values of  $\langle \sigma^c \rangle$ ,  $\mu_{rms}^c$ , and  $-\sum \log(\text{likelihood})$  are plotted as function of the assumed  $\tau$  value. The position of the minimum of each of the three curves is the  $\tau$  values favoured by the particular criterion. As expected, the criteria focussing on the mean colour  $\mu^c$  are lightly biased towards higher  $\tau$  values, while the  $\langle \sigma^c \rangle$  and the joint likelihood criteria are slightly biased towards low  $\tau$  values (the input value indicated as blue line, see text for detail).

Just like in the previous section, the false (too low) value for  $\tau$  leads to a bias high in  $\sigma^c$ , a red bias in  $\mu^c$  and a bias low in likelihood, with increasing strength with increasing inclination. However, the intrinsic scatter is already so large that only the bias in  $\mu^c$  is visible by eye, while the bias in  $\sigma^c$  and especially in the likelihood is barely noticeable. It should be noted that this is a consequence of there being no scatter in  $\tau$  for given stellar mass in the simulated data set. We expect the real data to have such as scatter due to the  $\Sigma_s - \tau$  relation of Grootes et al. (2013) (see Sect. 5.4.2).

Nevertheless, even  $\sigma^c$  and the likelihood contain extractable information about the best-fitting dust model parameter. Fig. 7.8, which displays the three criteria as function of the dust model parameter  $\tau$ , shows that all three criteria are sensitive to the change in dust model parameter.

Fig. 7.9 shows a histogram of the best-fitting  $\tau$  values for 100 simulated, noise-free data sets (data set A, see Sect. 6.3.2.1) using the three different criteria. While the criterion based on the rms of the mean colour  $\mu_{rms}^c$  reproduces the input  $\tau$  value of 3 well (being only slightly biased towards high  $\tau$  values), the criteria based on the average standard deviation  $\langle \sigma^c \rangle$  and the likelihood show a small bias towards low  $\tau$  values.

The bias in the  $\tau$  estimate using the mean colour  $\mu^c$  is attributed to the fact that the present algorithm does not take into account the dust-dependent incomplete-

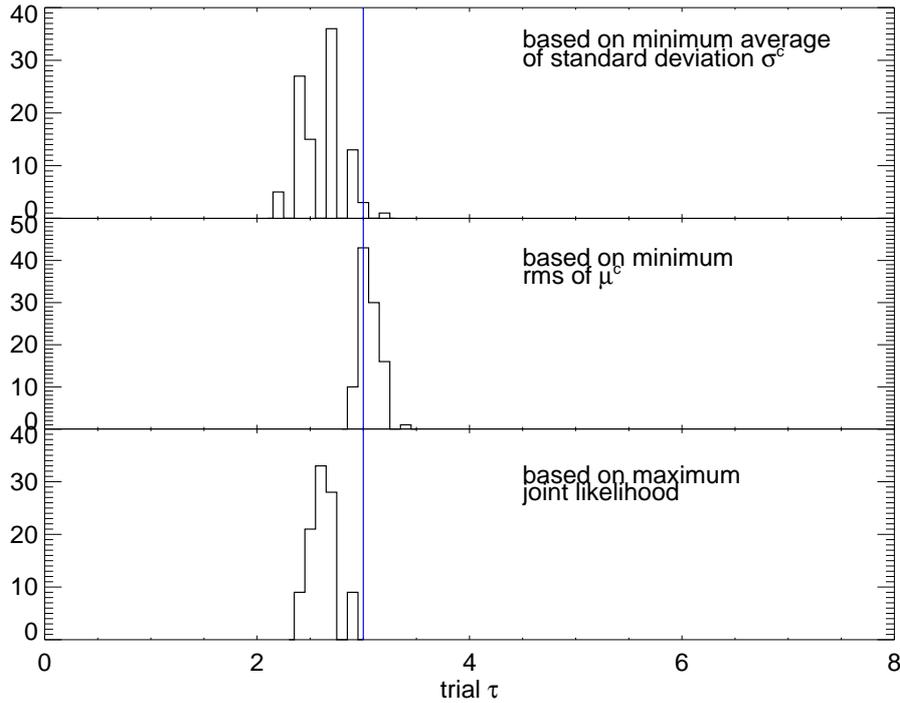


Figure 7.9: Best-fitting  $\tau$  value for noise-free data

The histogram of best-fitting  $\tau$  values identified with the three different criteria for data set A (100 noise-free data realisations) is shown. The  $\langle \sigma^c \rangle$  and the joint likelihood criteria are slightly biased towards low  $\tau$  values (the input value indicated as blue line), while the  $\mu_{rms}^c$  criterion is only slightly biased towards higher  $\tau$  values. The width of the histograms is caused by the sampling noise of the simulated data and represents the uncertainty of the best-fitting  $\tau$  value (see text for detail). The vertical blue line denotes the value  $\tau = 3$  used in the generation of the simulated data sets.

ness of the intrinsic r-band magnitude. This is the missing term  $P(m_{r,intr}|m_{r,obs}, \phi)$  of Eq. 4.1. In detail, this can be explained as follows: For a given stellar mass, the distribution of intrinsic NUV magnitude is independent of the distribution of the intrinsic r-band magnitude. Both are tightly correlated with stellar mass and both relations have intrinsic scatter. These two scatter components together are described by the standard deviation of the intrinsic colour distribution. Now, the detection limit is applied to the r-band flux, removing low-flux galaxies from the sample. Statistically speaking, these low-flux galaxies will also have fainter intrinsic r-band magnitudes. For a given stellar mass bin, the detection limit therefore removes fainter r-band galaxies, leaving brighter ones in the sample. Since the UV magnitude is a function of the given stellar mass and not of the r-band luminosity, the distribution of the NUV remains unchanged. Therefore, by removing fainter r-band galaxies, we remove bluer galaxies and bias the mean colour  $\mu$  and  $\mu^c$  of the remaining (i.e. “observed”) sample towards the red. As the detection limit is applied to the attenuated flux, this bias will be more extreme for high inclination bins, as the galaxies passing the r-band detection limit must be even more intrinsically r-band bright for higher attenuation values (leading to even redder intrinsic colours of the “observed” sample in the high inclination bin). The redder colour in high inclination is misinterpreted as stronger dust attenuation, thus the bias in colour leads to a bias high in the estimated dust content for given

$M_s$ . For the real data, we expect a similar bias.

The  $\tau$  estimate via the average standard deviation  $\langle \sigma^c \rangle$  or the joint likelihood are less affected by this bias as it is overcompensated by another bias. The removal of blue galaxies leads to a narrowing of the observed colour distribution for the detected sample. The de-attenuation with the true dust model parameter will restore the intrinsic colour distribution for the detected sample on the red side, but at the same time it will cause the formerly sharp cut off in the attenuated r-band caused by the detection limit to be smeared out in the de-attenuated r-band. Smearing out this formerly sharp cut edge however slightly broadens the colour distribution. Thus, the fitted average standard deviation  $\langle \sigma^c \rangle$  will be minimized for a slightly too low  $\tau$  value, where the remaining broadening of the intrinsic colour distribution of the detected sample due to the imperfectly de-attenuated colour is smaller than the broadening of the intrinsic colour distribution due to the smearing out of the formerly sharp cut edge caused by the detection limit.

### 7.5.3 The effect of UV noise on the parameter estimation

In this section we explore the consequences of there being NUV noise in the data, when the algorithm in its present form assumes no noise. We use the results to determine how bright a galaxy needs to be for inference of parameters to be made with the present algorithm.

In the presence of UV noise (Gaussian in flux), the intrinsic colour distribution (log-normal in flux) is convolved with the noise distribution. As the noise is proportional to the square root of the counts, bright NUV sources have a smaller relative error<sup>8</sup> than faint sources.

Generally ignoring the NUV noise, or at least the non-symmetry of the magnitude errors, as it is usually done in the literature (e.g. Treyer et al., 2005; Robotham & Driver, 2011), leads the fit of intrinsic colour distribution to be biased towards redder mean colour  $\mu$  and larger standard deviation  $\sigma$ , especially in the edge-on bin, since dust attenuation causes this bin to contain galaxies with fainter apparent magnitudes than those galaxies in the face-on bin (see Fig. 7.10, bottom panel). The inclination-dependence of the intrinsic colour parameters thereby caused is undistinguishable from the inclination dependence caused by dust attenuation which we use as a diagnostic.

In the absence of a rigorous treatment of noise, the solution is to use only bright objects (i.e. apply a sharp detection limit  $m_{lim}$ ), such that the relative error in the UV detection can be ignored. However, applying this cut on the UV data would bias our sample against red galaxies. For moderate cases, this could be accounted for by integrating the probability distribution according to Eq. 7.13. However, when the UV detection limit is so bright, that a significant fraction of galaxies in a given bin are non-detections, the fit of the intrinsic colour distribution would become unstable. It would be biased towards the blue, as there are only blue galaxies detected<sup>9</sup>.

---

<sup>8</sup>of course, the absolute flux error is larger. But since we are interested in the intrinsic colour distribution, which is a Gaussian in magnitude, the relative error is of interest.

<sup>9</sup>This can also be seen through considerations of an example scenario for a data set with galaxies all at the same distance: The detection limit is one  $\sigma$  bluewards of the intrinsic colour. Thus, we have 84% non-detections of which we only know that they are redder than the detection

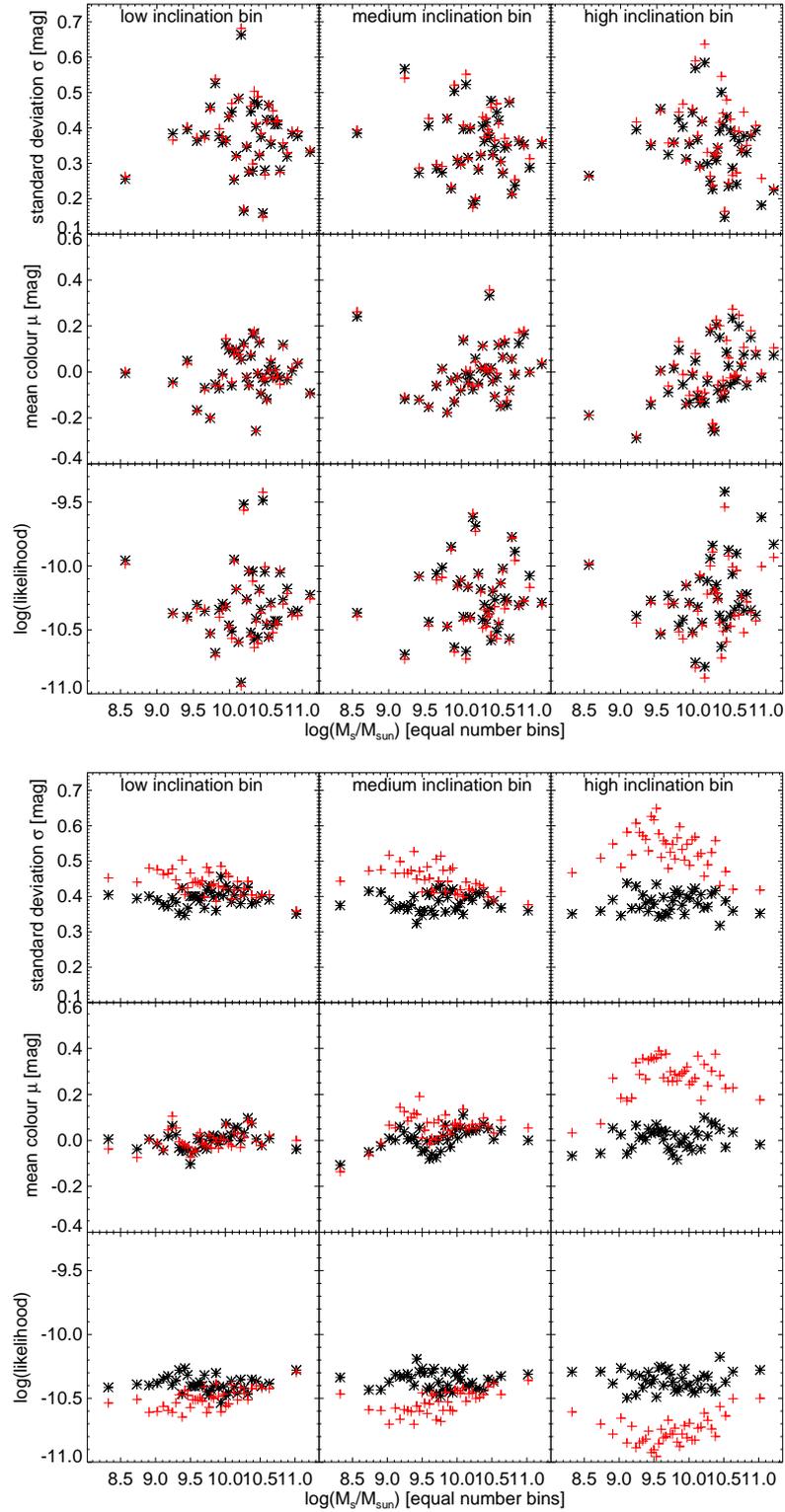


Figure 7.10: Comparison of the intrinsic colour parameter derived from noisy data with and without imposing a bright r-band flux limit.

Shown are the intrinsic colour parameters  $\mu^c$ ,  $\sigma^c$ , and joint likelihood derived for the data set A.1 with (red) and without (black) NUV noise. The lower panel shows the simulations for the  $m_r = 19.8$  mag detection limit and the upper panel shows the same simulations for the bright cut at  $m_r = 17$  mag. The recovered parameters for the noise-free data show no trend with inclination. The noisy data causes a strong bias with inclination when the  $m_r = 19.8$  mag detection limit is used (bottom). This bias is almost removed when the bright cut at  $m_r = 17$  mag is applied (top). However, the significant reduction in number of galaxies increases the sampling noise significantly.

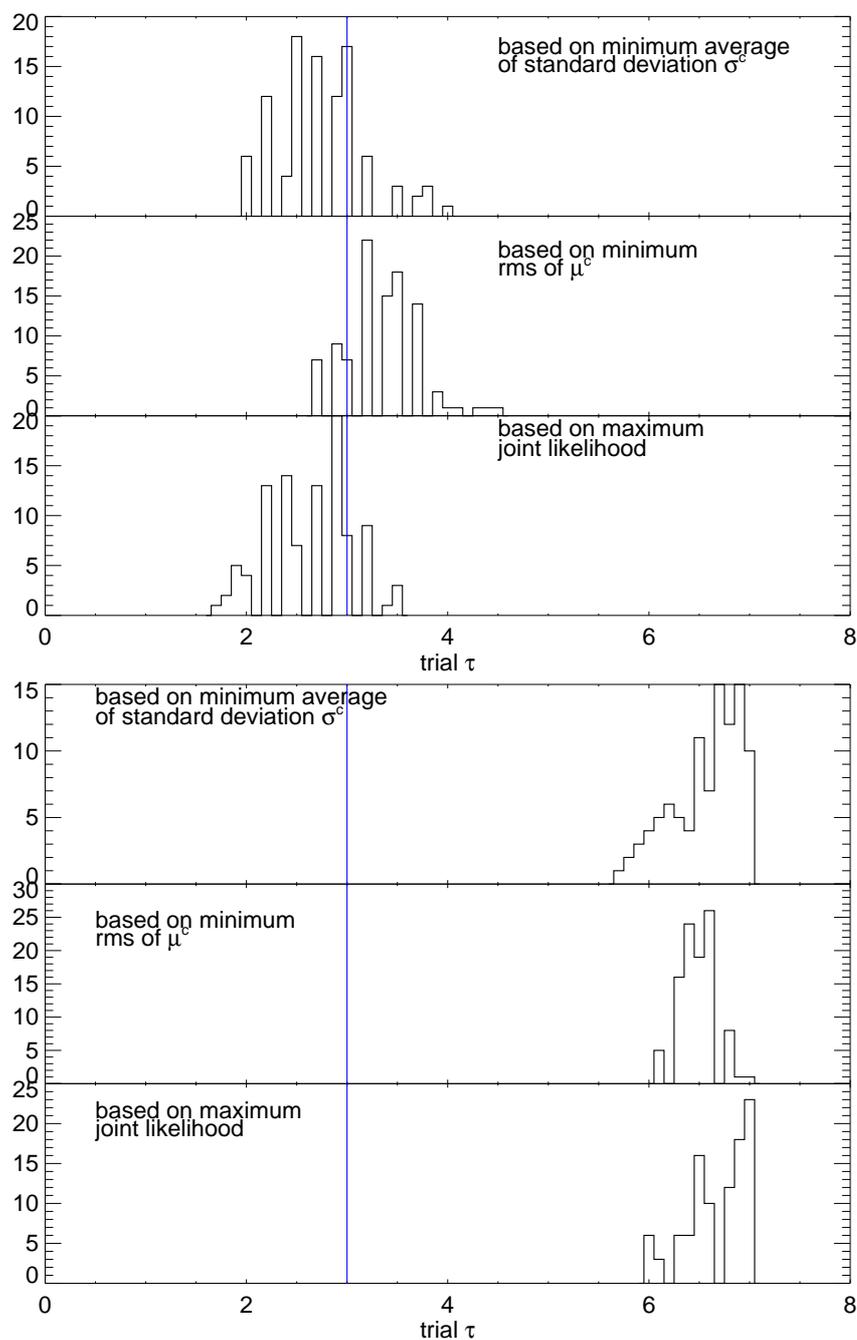


Figure 7.11: Comparison of best-fitting  $\tau$  values of noisy data set A analysed with and without imposing a bright r-band flux limit.

For 100 simulated noisy data sets (i.e. data set A), the histograms of the best-fitting  $\tau$  values analysed with (top) and without (bottom) the 17 mag r-band cut are shown. While all three criteria roughly reproduce the input value of  $\tau = 3$  (indicated by the blue line) for the bright cut, a strong bias in  $\tau$  is visible if analysed using the 19.8 mag r-band cut.

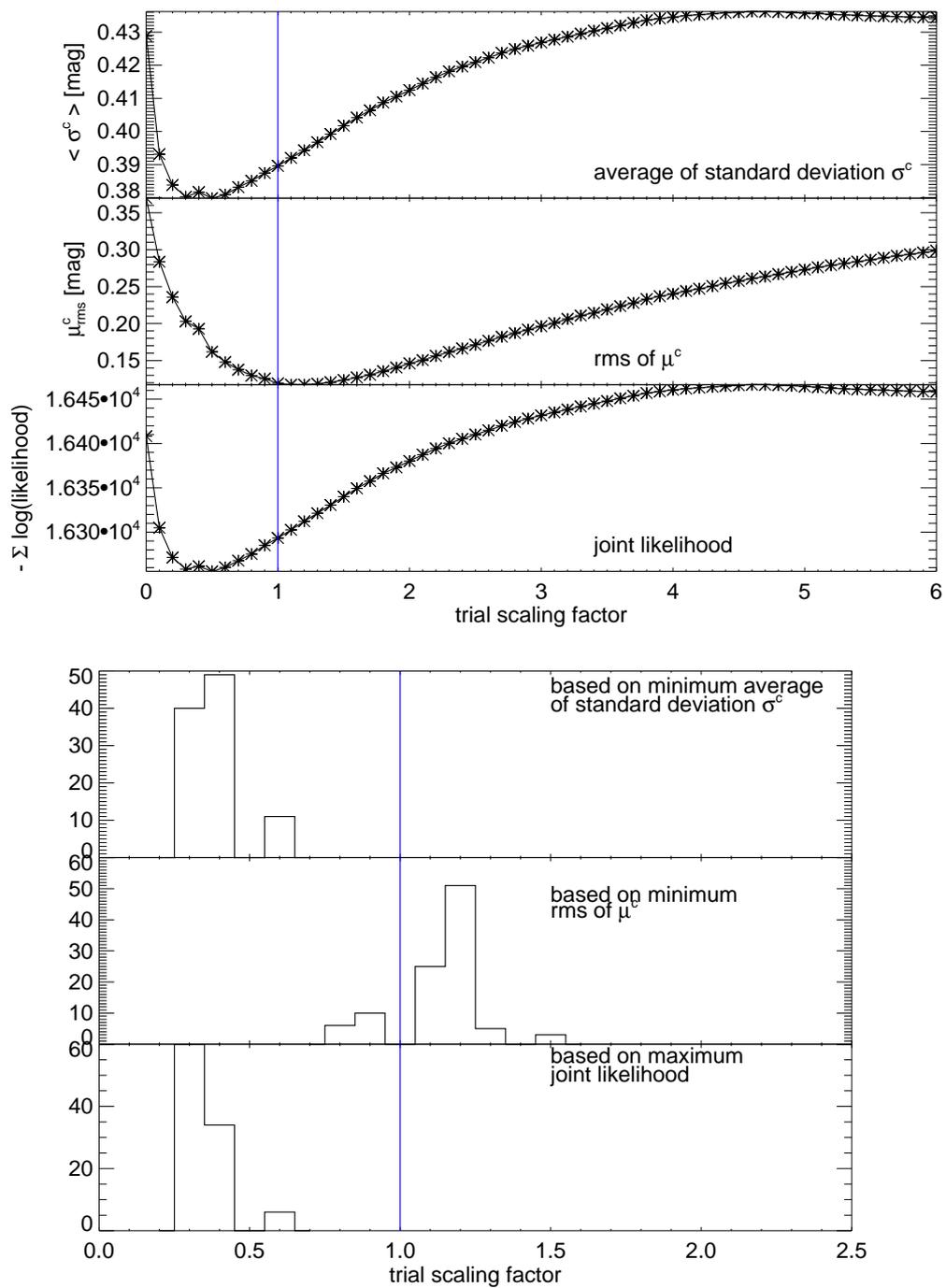


Figure 7.12: Identifying the best-fitting scaling factor  $\beta$  between  $\tau$  and the stellar mass surface density for the stellar mass-dependent dust model.

For the simulated data set C.1 (single realisation, stellar mass-dependent model, NUV noise) and imposing a bright r-band flux limit, the values of  $\langle \sigma^c \rangle$ ,  $\mu_{rms}^c$ , and  $-\sum \log(\text{likelihood})$  are plotted as function of the assumed scaling factor  $\beta$  in the top panel (input value indicated by the blue line). The bottom panel shows the histograms of the best-fitting amplitude values recovered for the 100 simulations contained in data set C judged by the three criteria. The vertical blue line denote the value  $\beta = 1$  used in the generation of the simulated data sets.

The alternative, which we adopt here, is to apply a rigorous detection limit to the r-band. Doing so, the full UV-r colour distribution is preserved, while excluding objects with faint apparent magnitudes in the r-band and subsequently in the UV. The intrinsic colour fit will not be biased<sup>10</sup>. Nevertheless, this rigorous r-band detection limit significantly reduces the sample size. In this work, we require a galaxy to have an r-band magnitude brighter than 17mag, and thus reduce our sample to the brightest 12%. This sample size reduction is at this stage a necessary evil to avoid a bias in the results of the analysis due to noise. In appendix B we outline a method of a rigorous treatment of the NUV noise distribution in the inference of the dust parameters.

Fig. 7.10 compares the fit results for noise-free data (black) and noisy data (red) using the bright r-band cut (top panel) and the 19.8mag cut (bottom panel) for the same data set. The bottom panel shows a clear trend in the recovered parameters with inclination for the noisy data, and no trend for noise-free data. In other words, the bottom panel illustrates the bias introduced to the fit results by NUV noise. In the top panel, the trend in the found parameter with inclination for the noisy data is significantly reduced, such that it can be neglected. The other striking difference is the increased scatter of the found values if the bright detection limit is used. This is easily understood as this fit only uses 12% of the data compared to the fit which uses the 19.8mag limit.

Fig. 7.11 shows a histogram of the best-fitting  $\tau$  values using the three criteria for 100 simulated noisy data sets (input value  $\tau = 3$ ) with the bright detection limit (top) and with the 19.8mag limit (bottom). Using the bright limit of 17mag, the best-fitting  $\tau$  distribution shows the same trends as those of noise-free data shown in Fig. 7.9. The increase in scatter is a result of the sample size reduction. Using the 19.8mag r-band limit, the best-fitting  $\tau$  distribution is heavily biased to high values due to the NUV noise.

We have also performed this test (including the bright r-band cut) for the data set C, which uses the stellar mass-dependent dust model, in order to judge the biases caused by the r-band detection limit for this dust model. Fig. 7.12 (upper panel) shows for the single simulation of data set C.1 the recovered values of the three criteria for different trial values of the model parameter (i.e. the scaling factor  $\beta$  linking  $\tau$  and the stellar mass surface density as defined in Eq. 5.10). Fig. 7.12 (lower panel) shows the distribution of best-fitting values of  $\beta$  for the 100 simulations contained in data set C. We observe a similar bias as already for the simple dust model. The criterion based on the mean colour of the sample (i.e.  $\mu_{rms}^c$ ) is biased toward higher scaling factors  $\beta$  (which translates to higher  $\tau$  values). The criteria based on the likelihood and the standard deviation of the colour distribution are biased towards low scaling factor  $\beta$  values, i.e. towards

---

limit. Combinations of  $\mu$  and  $\sigma$ , which satisfy  $\mu - \sigma = m_{lim}$  have the same joint likelihood for all non-detections as the true parameter combination has. (The integral of a normalized Gauss function from  $\mu - \sigma$  to  $\infty$  has the constant value of 0.84, independent of the actual values of  $\mu$  and  $\sigma$ .) The remaining 16% are very blue. Thus, they will favour a blue  $\mu$  over a red  $\mu$ . In combination, the fitted colour distribution will be too blue and too narrow.

<sup>10</sup>In our simulation, this is only true to first order, since the UV-r colour is defined via the stellar mass. An r-band detection limit will remove r-band faint galaxies, i.e. blue galaxies, resulting in a narrower and redder colour distribution. This effect can be seen when analysing noise-free data. However, this effect is significantly smaller than the bias introduced by untreated UV noise.

lower  $\tau$  values.

#### 7.5.4 Effect of equating $b/a$ with $\cos(i)$

As already noted many studies of inclination dependence of attenuation have equated  $b/a$  with  $\cos(i)$ . Even though our algorithm contains a fully self-consistent treatment of the inference of  $i$  from  $b/a$  (see Sect. 6.2.1.2), it is of interest to explore the consequences of falsely equating  $b/a$  with  $\cos(i)$ .

Contrary to the simulated data, the intrinsic inclination is not known for real data. The GAMA data set provides only observed axis ratios ( $b/a$  values, see Sect. 5.3.3.1). As we have discussed in Sect. 6.2.1.2,  $b/a = \cos(i)$  is only valid for very flat disks. Fig. 7.13 illustrates the severe bias of the derived dust model parameter if  $b/a$  is falsely equated to  $\cos(i)$  for data set F using the stellar mass-dependent dust model (top) and data set E using the simple dust model (bottom). All three criteria are biased towards high values for the dust content compared with the parameter distribution shown in Fig. 7.11 and Fig. 7.12, respectively, for the analysis using the known  $\cos(i)$ . The criterion based on the joint likelihood seems to be least affected by this bias. This may possibly be due to the coincidental equal strength of the bias caused by the r-band detection limit, roughly counterbalancing the bias caused by assuming  $b/a = \cos(i)$ . Apart from the mean of the parameter distribution, also their width is significantly increased to a level that they lose much of their predictive power. We conclude that a realistic treatment of the inference of  $i$  from  $b/a$  is essential in any statistical analysis of dust content. We have included a full treatment as described in Sect. 7.2, especially Eq. 7.4.

The effect of including this full treatment is shown in Fig. 7.14, which shows the distribution of the best-fitting dust model parameter for data set F using the stellar mass-dependent dust model (top) and data set E using the simple dust model (bottom). There are still systematic biases in the recovered parameters due to the r-band detection limit as already discussed. However, the parameter estimates have much less scatter, and are closer to the true values. Indeed, the distributions in recovered values in Fig. 7.14 are similar to the distribution obtained when the values  $i$  used in the parameter derivation are set to the input values in the simulation (see Figs. 7.11 and 7.12).

## 7.6 Distinguishing between dust models

In this section, we investigate the ability of the analysis technique to distinguish between different functional forms for the dust model. To do this, we compare the parameters recovered by the algorithm assuming the two different functional forms for  $\tau$  (simple model with delta function for  $\tau$  and stellar mass-dependent model) when analysing a simulated data set using the stellar mass-dependent model with  $\beta = 1$ . Fig. 7.15 (top panel) shows computed values for  $\langle \sigma^c \rangle$ ,  $\mu_{rms}^c$ , and the joint likelihood as a function of the  $\beta$  parameter of the stellar mass-dependent model, while Fig. 7.15 (bottom panel) shows the same quantities computed as a function of the  $\tau$  value for the simple dust model. One can clearly see that all three criteria prefer the input model, i.e. the stellar mass-dependent dust model

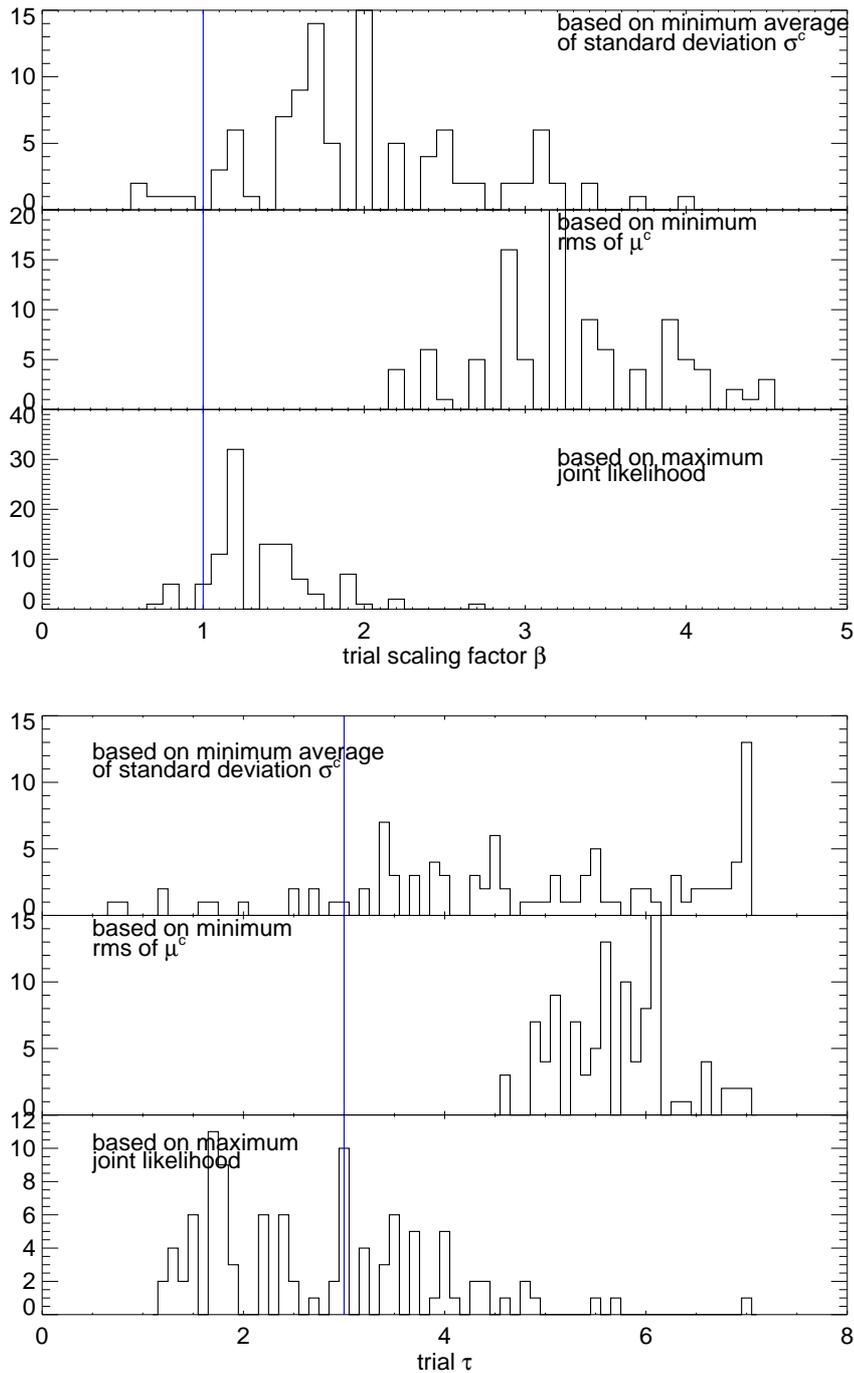


Figure 7.13: Best-fitting dust model parameters falsely assuming  $b/a = \cos(i)$ . Shown are the histograms of the derived dust model parameter falsely assuming  $b/a = \cos(i)$ . Top: The values for the best-fitting dust model parameter  $\beta$  of the stellar mass-dependent dust model for data set F are  $2.05 \pm 0.70$ ,  $3.34 \pm 0.56$ , and  $1.38 \pm 0.33$ , based on the criteria  $\min(\langle \sigma^c \rangle)$ ,  $\min(\mu_{rms}^c)$ , and  $\max(\text{joint likelihood})$ , respectively. Bottom: For data set E, the best-fitting dust model parameter  $\tau$  of the simple dust model using the same criteria are  $4.87 \pm 1.65$ ,  $5.71 \pm 0.56$ , and  $2.82 \pm 1.13$ . The vertical blue lines denote the value of the dust model parameter  $\tau = 3$  or  $\beta = 1$  used in the generation of the simulated data sets.

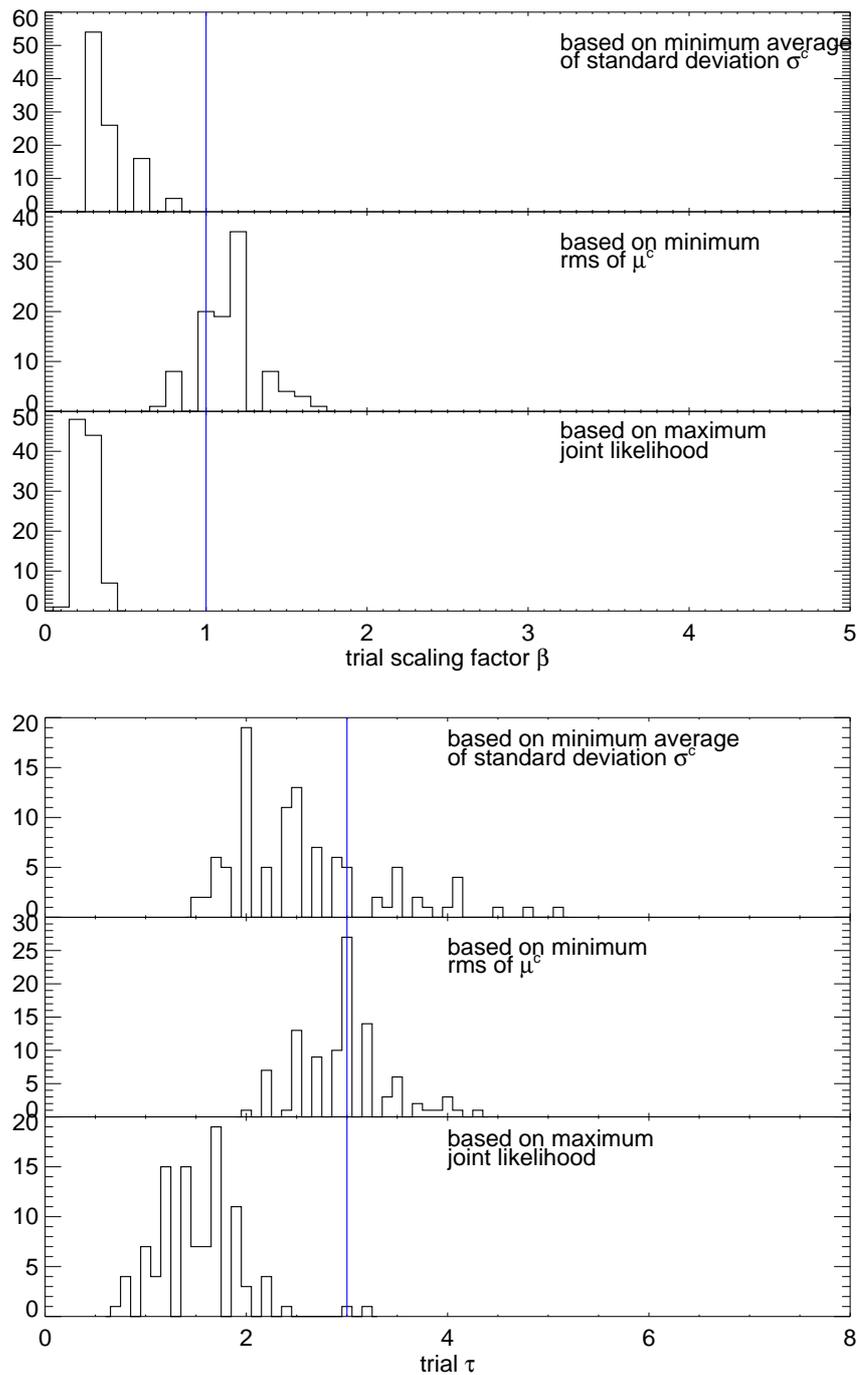


Figure 7.14: Best-fitting dust model parameter derived with  $b/a$  values

Top: The values for the best-fitting dust model parameter  $\beta$  of the stellar mass-dependent dust model for data set F are  $0.47 \pm 0.12$ ,  $1.17 \pm 0.19$ , and  $0.25 \pm 0.07$ , based on the criteria  $\min(\langle \sigma^c \rangle)$ ,  $\min(\mu_{rms}^c)$ , and  $\max(\text{joint likelihood})$ , respectively. Bottom: For data set E, the best-fitting dust model parameter  $\tau$  of the simple dust model using the same criteria are  $2.62 \pm 0.76$ ,  $3.02 \pm 1.17$ , and  $1.55 \pm 0.42$ . The vertical blue lines denote the value of the dust model parameter  $\tau = 3$  or  $\beta = 1$  used in the generation of the simulated data sets.

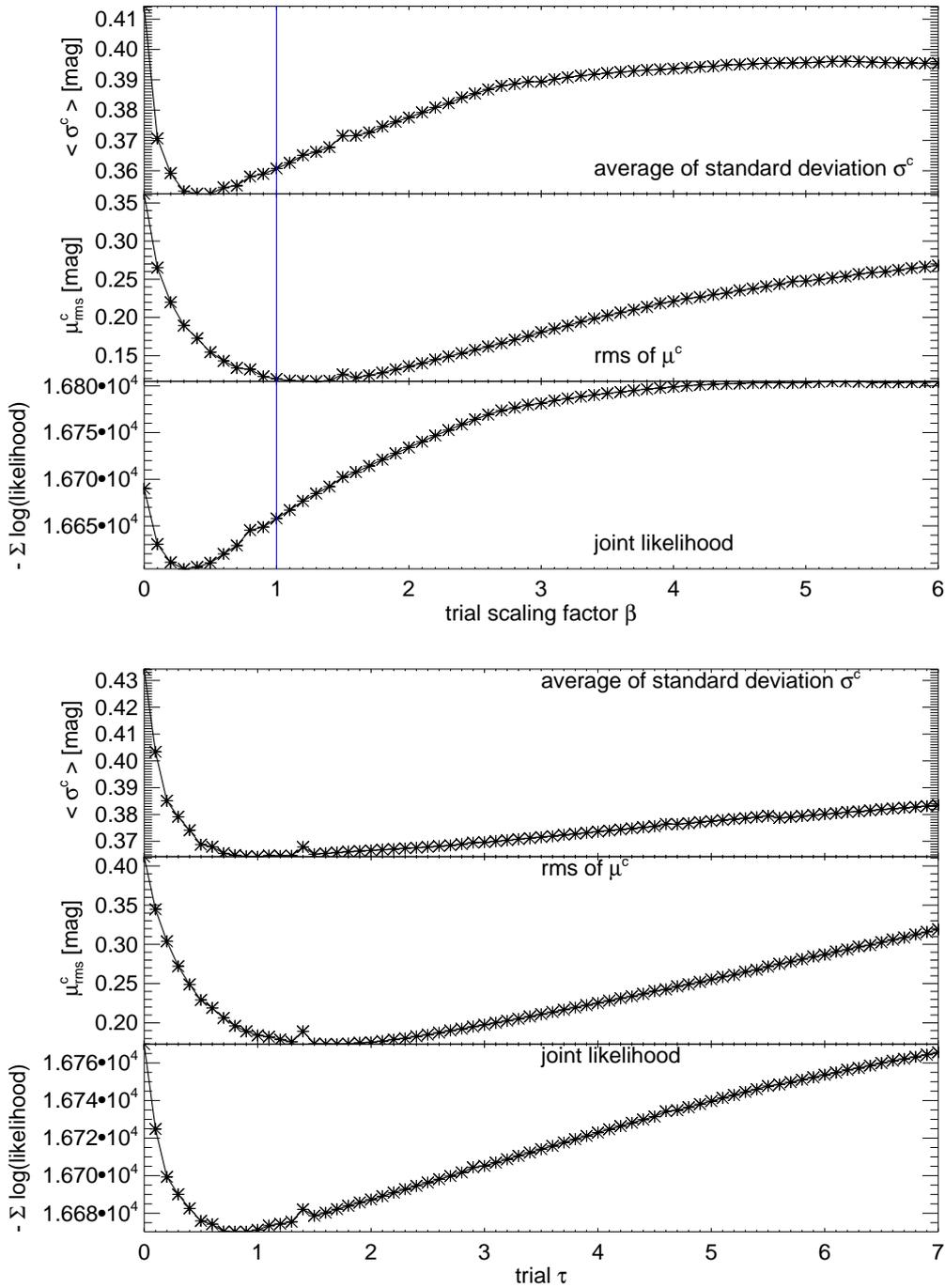


Figure 7.15: Comparison of the application of different dust models operating on simulated data set E.1 (created using stellar mass-dependent model)

The values of  $\langle \sigma^c \rangle$ ,  $\mu_{rms}^c$ , and  $-\sum \log(\text{likelihood})$  are plotted as function of the trial scaling factor  $\beta$  for stellar mass-dependent model in the top panel (input value indicated by the blue line) and as function of the trial parameter  $\tau$  of the simple dust model (false model). The simulated data is flux-limited at  $m_r=17$  mag.

for the considered data set. The difference in the maximum log-likelihood (at the best-fitting values for  $\beta$  and  $\tau$ ) is  $\sim 80$  between the two models.

Note that this approach only identifies the best model out of the given set of models. It compares the given models, but it is not capable of judging the objective applicability of a given model.

## 7.7 Application to real data

Bearing in mind the biases on the derived dust parameters discussed in the previous sections, we now apply the analysis technique to the real GAMA spiral galaxy sample. We apply the same bright r-band detection limit of  $m_r < 17$  mag as discussed in Sect. 7.5.3. We present results for the recovered dust parameter in Sect. 7.7.1. A technical discussion of these results is presented in Sect. 7.7.3 and we discuss the results scientifically in Sect. 7.7.4.

### 7.7.1 Best-fitting dust model (parameter) for the GAMA spiral galaxy sample

Figure 7.16 (top) shows the three criteria average standard deviation  $\langle \sigma^c \rangle$ , rms of the stellar mass-colour trend removed residual mean colour  $\mu_{rms}^c$ , and the joint likelihood as function of the stellar mass-dependent dust model parameter  $\beta$ . The spread between the favoured  $\beta$  values of the three criteria is large at first glance. However, if we compare these values with the distribution of the favoured scaling factors for simulated data set F in Fig. 7.14, we see that the favoured values for the real data are almost three times the value favoured for the simulated data set F for each criterion. The bias induced by the lack of consideration of the r-band detection limit is likely to be responsible for the systematic difference of the favoured  $\beta$  value for the three criteria. If we assume that both the biases and the error in the derived dust parameter can be taken from the results of simulation data set F, then the biases and errors for the real data set can be obtained by scaling the biases and error from the simulation by the same factor. Thus, the favoured values of the scaling factor  $\beta$  are  $1.3 \pm 0.4$  ( $\min(\langle \sigma^c \rangle)$ ),  $3.2 \pm 0.6$  ( $\min(\mu_{rms}^c)$ ), and  $0.7 \pm 0.2$  ( $\max(\text{joint likelihood})$ ).

Fig. 7.16 (bottom) shows the three criteria as function of the simple dust model parameter  $\tau$ . The favoured values of the  $\tau$  are  $1.7 \pm 0.76$  ( $\min(\langle \sigma^c \rangle)$ ),  $4.6 \pm 0.46$  ( $\min(\mu_{rms}^c)$ ), and  $1 \pm 0.43$  ( $\max(\text{joint likelihood})$ ), adopting the uncertainties of the derived dust model parameter from data set E in Sect. 7.14. Again, the three criteria show a similar trend for the real data as they did for the simulated data set F, but prefer a significantly higher parameter value.

For all three criteria, the minimum of the best-fitting dust model parameter is deeper for the stellar mass-dependent dust model than for the simple dust model. Thus, all three criteria favour the stellar mass-dependent dust model for the attenuation correction of the GAMA spiral galaxy data set. The difference in log of the maximum likelihood for the two dust models is  $\sim 90$ , favouring the stellar mass-dependent model.

For both dust models, the three criteria show rather broad minima (especially  $\mu_{rms}^c$ , which we consider to be least affected by biases). Thus, the uncertainty of

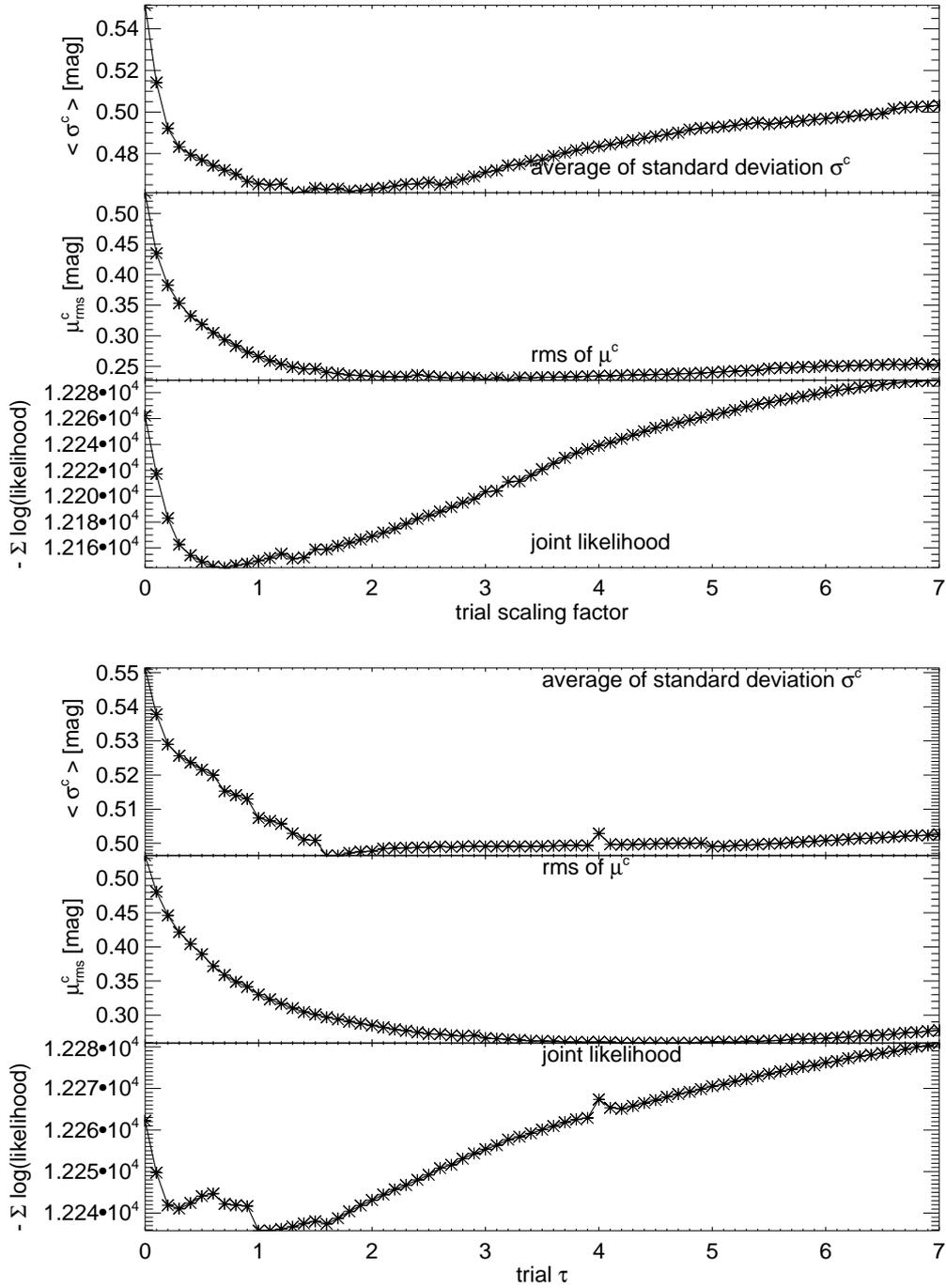


Figure 7.16: Derived dust model parameter for GAMA spiral galaxies. The top and lower panels respectively show the results obtained using the stellar mass-dependent dust model and the simple dust model. The values of  $\langle \sigma^c \rangle$ ,  $\mu_{rms}^c$ , and  $-\sum \log(\text{likelihood})$  are plotted as a function of trial values for the parameters scaling factor  $\beta$  in the top panel and  $\tau$  in the lower panel.

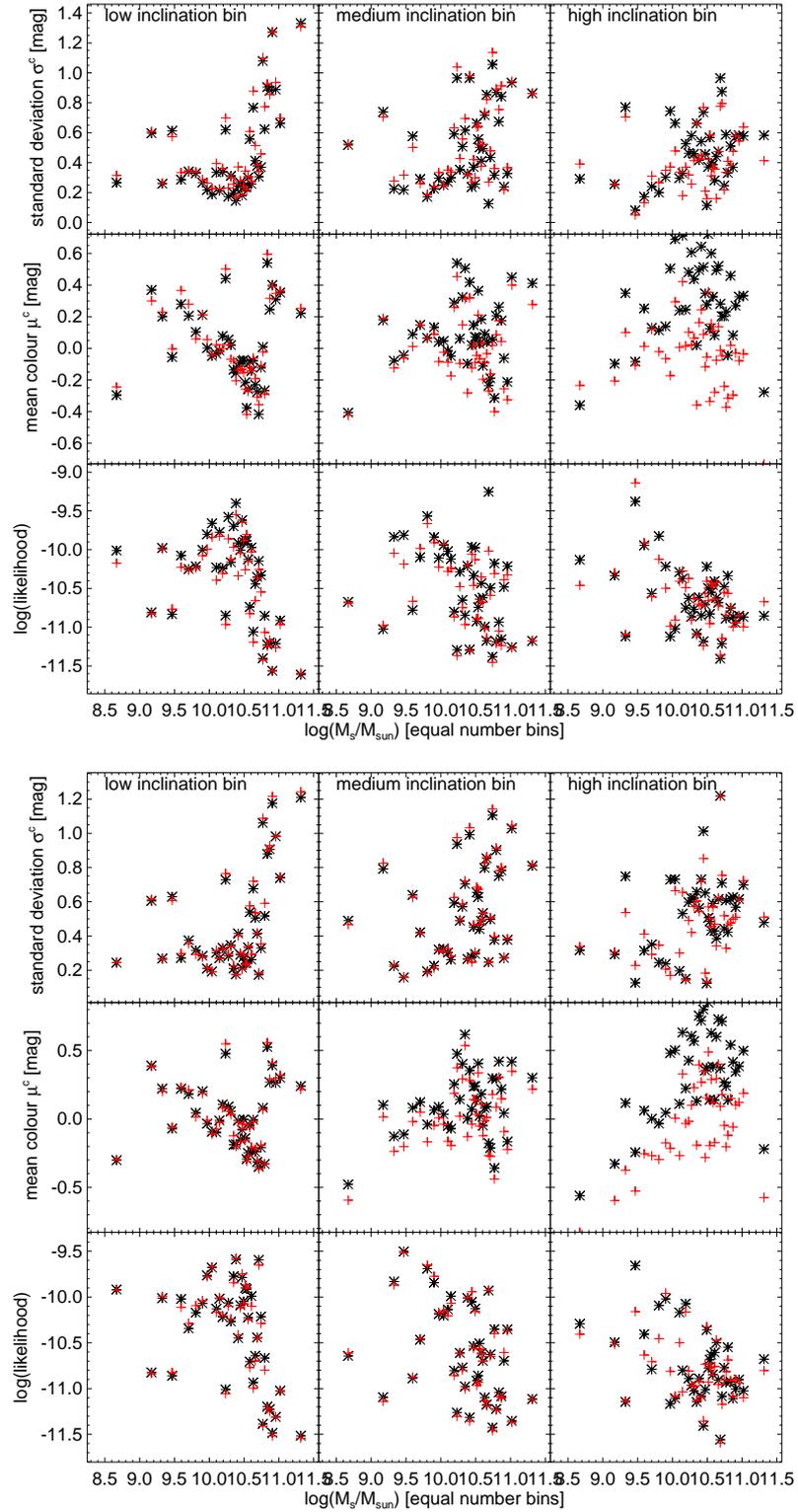


Figure 7.17: Derived intrinsic colour parameters for GAMA spiral galaxies.

The top two rows of each panel show the recovered values for standard deviation  $\sigma^c$  and mean colour  $\mu^c$  as a function of the stellar mass  $M_s$  for three ranges of  $b/a$  (corresponding to low, medium, and high inclination) calculated for different fixed values of the dust model parameters. In the top panel, black and red points are for values of  $\beta = 0.7$  and  $3.2$ , respectively (corresponding to the global solutions favoured by the maximum joint likelihood and the  $\mu_{rms}^c$  criteria, respectively). In the lower panel, the black and red points are for values of  $\tau = 1$  and  $4.6$ , respectively (again corresponding to the global solutions favoured by the maximum joint likelihood and the  $\mu_{rms}^c$  criteria). The lower row in each panel gives the joint likelihood for galaxies in each  $M_s$  bin for the same global solutions for  $\beta$  (top panel) and  $\tau$  (lower panel).

the best-fitting dust model parameter must be considered rather large, as we have already discussed for the distribution of best-fitting dust model parameters for the simulated data sets. Furthermore, it should be noticed that the depth of the minima is less than those for the simulated data. In the simulation, only the NUV magnitude is a noisy quantity. For all other quantities, the exact value is known. For the real data, all quantities have measurement uncertainties. Even though most of them are small enough to be neglected in the analysis, they nevertheless will slightly broaden the minima for the best-fitting dust parameter and make them less deep (see Sects. 7.7.3.2 and 7.7.4 for the discussion of uncertainties of the B/D ratio and the dust content and their potential effect on the analysis).

Figure 7.17 compares the intrinsic colour parameters derived using the best-fitting dust parameters based on the maximum joint likelihood (black) with the ones derived using the best-fitting dust parameters based on the minimum of  $\mu_{rms}^c$  (red) for the stellar mass-dependent dust model (top) and the simple dust model (bottom). For both dust models, one can see that the dust model parameter value favoured by the  $\mu_{rms}^c$  criterion (red) cause the mean colour  $\mu^c$  to scatter around  $\mu^c = 0$  for all three inclination bins. The dust model parameter values favoured by the maximum joint likelihood (black), however, do not fully remove the inclination trend in  $\mu^c$  (i.e. does not perfectly reverse the attenuation), causing the high inclination bin (edge-on) to have redder de-attenuated colours than the low inclination bin (face-on). This is likely to be mainly due to the bias introduced by the r-band detection limit, as already discussed in Sect. 7.5.2. For the other two intrinsic colour parameters ( $\sigma^c$  and  $\log(\text{likelihood})$ ), the inclination dependence is difficult to evaluate by eye, as the change with inclination is small compared to the face-on value. Their information content becomes visible to the eye by combining all stellar mass and inclination bins to the average standard deviation  $\langle \sigma^c \rangle$  and the joint logarithmic likelihood as shown in Fig. 7.16.

A direct comparison of the values of the standard deviation  $\sigma^c$ , the rms of the stellar mass - colour trend removed residual mean colour  $\mu^c$ , and the logarithmic likelihood at the position of the favoured dust model parameter for both dust models is shown in Fig. 7.18 (scaling factor  $\beta = 3.2$ , red, and  $\tau = 4.6$ , black, respectively, which are the solutions obtained using the  $\mu_{rms}^c$  criterion). Again, the slightly smaller values for  $\sigma^c$  and larger values for the logarithmic likelihood using the stellar mass-dependent dust model are difficult to recognize by eye. For  $\mu^c$ , only the high inclination bin clearly shows that the stellar mass-dependent dust model (red) is able to reduce the rms further than the simple dust model (black). Furthermore, one can see a trend of  $\mu^c$  with stellar mass in the high inclination bin (edge-on) using the simple dust model, indicating that the simple dust model is not able to remove the stellar mass trend in attenuation<sup>11</sup>.  $\mu^c$  does not show a trend with stellar mass noticeable by eye, indicating a good agreement between the stellar mass dependence of the attenuation in the real data and the one assumed in the stellar mass-dependent dust model.

---

<sup>11</sup>The intrinsic colour trend with stellar mass of the face-on bin was removed from all three bins in the first step. Since we do not expect the intrinsic colour to be a function of inclination, the remaining stellar mass trend visible in the edge-on bin must be a trend of the attenuation with stellar mass.

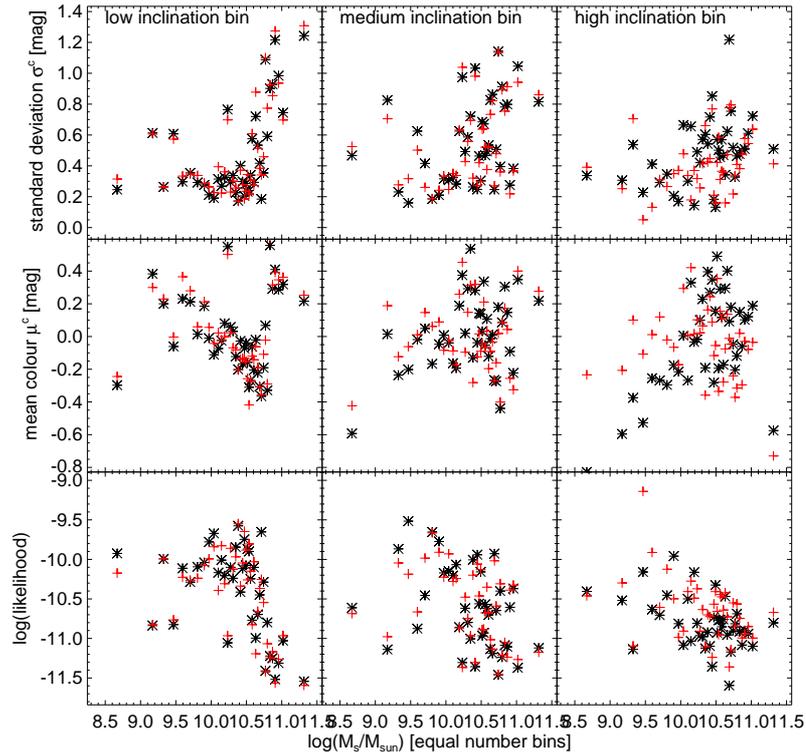


Figure 7.18: Best-fitting intrinsic colour parameters for GAMA spiral galaxies. Shown is the comparison of the intrinsic colour parameters  $\mu^c$ ,  $\sigma^c$ , and joint likelihood derived in the second step (i.e. with stellar mass - colour trend removal) assuming the best-fitting scaling factor  $\beta = 3.2$  (red) and the best-fitting  $\tau = 4.6$  (black), obtained from the global solution using the  $\mu_{rms}^c$  criterion.

## 7.7.2 Derived intrinsic colours

Finally, we note the implications for the derived distribution of intrinsic colour with  $M_s$  that the best-fit dust models discussed above have. The variation in colour  $\mu$  with  $M_s$  as deduced from the first step of the analysis (where the primary colour trends in  $M_s$  are fitted to the data; see Sect. 7.3.1) are shown for the best-fitting simple and stellar mass-dependent dust models (with best fit parameters  $\tau = 4.6$  and  $\beta = 3.2$ , respectively, as recovered using the  $\mu_{rms}^c$  criterion) in Fig. 7.19. The black points in the middle row of the upper panel show the primary colour trend with  $M_s$  returned by the analysis made using the stellar mass-dependent model, for the three bins in  $b/a$ . The corresponding result for the simple dust model is shown by the black points in the middle row of the lower panel. In the case of the favoured stellar mass-dependent model, it can be seen that the deduced variation in intrinsic colour is not continuous over stellar mass, but instead is flat for  $M_s \lesssim 10^{10} M_\odot$  and steepens rapidly redwards for  $M_s > 10^{10} M_\odot$ . It should be noted that this result is for disk-like spirals, so can't readily be explained in terms of an increasing fraction of red spheroids entering the sample for higher  $M_s$ . It is interesting to note, too, that the reddening in intrinsic colour at high  $M_s$  is accompanied by an increase in the intrinsic scatter of the colours and by a decrease in the joint likelihood. One potential cause of this could be the emergence of a bimodal distribution in intrinsic colour for high mass spiral galaxies.

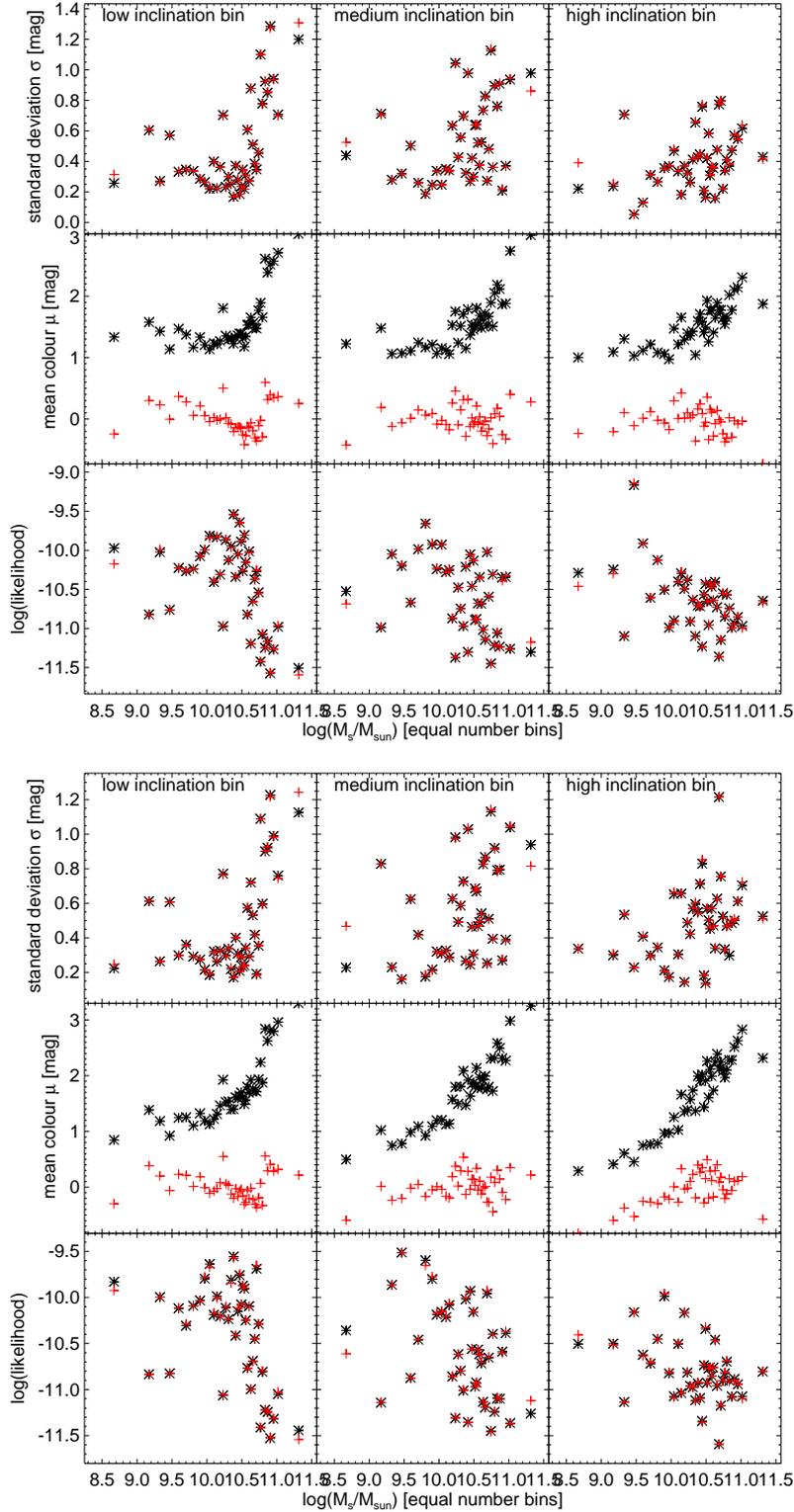


Figure 7.19: Best-fitting intrinsic colour parameters for GAMA spiral galaxies. Shown is the comparison of the intrinsic colour parameters derived in the first step (i.e. without stellar mass - colour trend removal, black) and derived in the second step (i.e. with stellar mass - colour trend removal, red) assuming the best-fitting scaling factor  $\beta = 3.2$  (top) and the best-fitting  $\tau = 4.6$  (bottom).

### 7.7.3 Technical discussion of the results

In this section, we discuss the effect on recovered dust parameters of the known technical deficiencies of the model and effects of the known deficiencies in the data.

#### 7.7.3.1 r-band detection limit

We have discussed in Chap. 4, the full probability of seeing a galaxy with the observed quantities is a combination of the probability of seeing its observed colour and the probability of seeing the r-band magnitude. The later probability function describes the selection effects causing the flux-limited sample to differ from the parent population (i.e. it describes the effect of the detection limit on the sample characteristics).

The colour incompleteness for a given stellar mass range caused by the r-band detection limit leads to a strong bias in the derived dust parameters. For the simulated data sets, we have shown its effect on the derived dust model parameter in detail (see Sects. 7.5.2, 7.5.3, and 7.5.4). While we are not able to isolate this effect for the real data, we do expect a similar bias for the GAMA spiral galaxy sample.

#### 7.7.3.2 Bulge-to-disk proxy

Since the GAMA data set does not (yet) have bulge-to-disk decompositions available, we relied on using the single Sérsic index  $n$  as proxy for the bulge-to-disk ratio B/D (see Sect. 5.3.3.2). Although carefully derived, this proxy however is nevertheless only able to predict the most likely B/D value for a given  $n$ . Thus, the scatter between the physical B/D distribution for a given  $n$  and the value predicted by the proxy will result in a scatter in the r-band attenuation. Since the bulge is not luminous in the NUV, the NUV attenuation is only calculated from the disk component, see Sect. 5.4.1. However, the bulge has a strong effect on inferred disk sizes, since there are inferred from the r-band data. The disk sizes have a strong influence on attenuation predicted by the radiation transfer model.

Even though the scatter in the attenuation is small, it will cause the depth of the minima of the dust model parameter criteria to be shallower compared to the simulation.

#### 7.7.3.3 Effects of dust attenuation on size and inclination estimates

The effect of dust attenuation and projection effects due to the inclination on size measurements of galaxy disks was studied in great detail by Pastrav et al. (2013a,b). They find that the effective radii of highly inclined disks are slightly overestimated even for dust-free galaxies. The change of the observed light distribution due to dust attenuation results in an even stronger bias towards high radii<sup>12</sup>. Consequently, galaxies with higher dust attenuation and/or higher inclination have overpredicted radii. In principle, this may have an effect on inference

---

<sup>12</sup>It should be noted that there is also a bias for the derived Sérsic index. At least to some extent can the bias in effective radius be explained by the different light profile for different Sérsic indices.

of all quantities dependent on disk size<sup>13</sup>. One effect would be on the inference of inclination from  $b/a$  from highly inclined galaxies.

Accounting for the bias of the size estimate in our analysis would allow us to isolate and remove the effect that this bias has on the final results. However, we expect this the bias caused by the overpredicted sizes to be small compared to the bias caused by the r-band detection limit.

#### 7.7.3.4 NUV noise

We have discussed the effect of NUV noise on the derived dust model parameter in detail in Sect. 7.5.3. For this work, we were forced to restrict our analysis to a very small, bright subsample of the actual GAMA spiral galaxy data set in order to avoid biases in the parameter estimation. In the appendix B we sketch a technique allowing us to account for NUV noise inside of the intrinsic colour probability. This will enable us to repeat our analysis for the full GAMA data set (increasing the data set by a factor  $\sim 10$ ). This will allow us to be much more precise in the dust model parameter estimation since we will grossly reduce the sampling noise, and extend the range of stellar mass addressed.

#### 7.7.3.5 Form of parametric distribution in colour

As described in Sect. 6.2.5, the choice of a log-normal form for the distribution in intrinsic colour was empirically motivated through the consideration of the observed colour distribution of  $M_s = 10^9 M_\odot$  galaxies, combined with simulations of the effect of dust attenuation on the observed colour distribution. Nevertheless, we have no real prior knowledge that a log-normal distribution is still valid at high stellar masses ( $M_s > 10^{10} M_\odot$ ) where dust attenuation is severe. As already commented on in Sect. 7.7.2, there is a sudden turn towards the red of the inferred intrinsic colour of high mass galaxies. This is visible for the face-on orientations (see black points in middle left-hand panel of the upper set of panel in Fig. 7.19, which shows the best fitting intrinsic colour parameters for stellar mass-dependent dust model favoured by the  $\mu_{rms}^c$  criterion). This suggests that quiescent red disk galaxies may be entering the population at higher  $M_s$ . At the same time, the standard deviation  $\sigma$  is larger (and the likelihood is smaller) for the high mass bins. We see this red excess for high stellar mass in the colour - stellar mass plot in Fig. 6.9 as well. Biasing the lower inclination bins towards the red would cause an underestimation of the dust attenuation judged by the  $\mu_{rms}^c$  criterion.

In addition to possible bimodality in intrinsic colour due to quiescent spirals, we must also consider whether contamination of the sample by ellipticals may be an issue. The spiral selection technique by Grootes et al. (2014) is designed to be pure and has only an elliptical contamination of  $\sim 2\%$ . However, these 2% are more likely to be in the high stellar mass regime. Furthermore, they are not likely to be in the high inclination range. In other words, we expect the elliptical contamination to affect only the high stellar mass bins for the face-on and the intermediate inclination bin.

---

<sup>13</sup>There will, however, not be a direct effect on the derivation of the  $\tau$  from stellar mass surface densities using the method of Grootes et al. (2013), which explicitly corrected for this effect

It remains to be investigated whether this sudden turn towards redder colours is caused by elliptical contamination, and to what extent it is due to quiescent disk galaxies. In the outlook, we will sketch a technique which accounts for the probability of a galaxy being a spiral or an elliptical galaxy rather than making a hard cut on the sample.

It should be noted that the expected bias caused by the emergence of a population of high mass red disk galaxies or elliptical contamination would have the opposite effect on the estimated dust parameters than the bias caused by the r-band detection limit, which is clearly the dominant bias for our analysis. Thus, an analysis accounting for the r-band detection limit in conjunction with an explicitly consideration of bimodality in intrinsic colour of disk galaxies is essential for investigating the incidence and properties of intrinsically red galaxies.

#### 7.7.4 Scientific discussion of the results

The analysis hinted that there might be a significantly larger scaling factor  $\beta$  for the stellar mass surface density ( $\Sigma_s$ )- $\tau$  relation than that derived by Grootes et al. (2013) on the basis of measurements of dust masses from submm emission. We have discussed how technical issues affect the favoured values in the previous section. Although the results for the GAMA spiral galaxy sample must be regarded as provisional pending a full quantitative treatment of the systematic bias still affecting the current method, it is instructive to consider possible astrophysical reasons for this result and identify hypotheses to be investigated using a future improved version of the analysis presented here.

The models we have considered for predicting  $\tau$  do not allow for any intrinsic scatter in dust mass at fixed  $M_s$ . We must therefore consider whether our results could be better explained using models for  $\tau$  with intrinsic scatter in dust mass. A Gaussian scatter in the logarithmic dust mass at given stellar mass would cause the dust mass (and with it the  $\tau$  value) to have an asymmetric distribution with a longer high tail for a given stellar mass. Thus, galaxies with an intrinsic  $\tau$  value on the upper tail (i.e. which have an underpredicted  $\tau$  value from the stellar mass-dependent model) will more strongly favour a higher  $\beta$  value than galaxies with an intrinsic  $\tau$  value on the lower tail will favour a lower one. Even if the scatter in  $\tau$  were symmetrical, an asymmetrical scatter in the attenuation would also be introduced into the data, and thus introduce a bias in our present analysis. This is because the change of the NUV-r inclination-dependent attenuation (see Fig. 5.9), especially the difference in the extreme cases of face-on attenuation and edge-on attenuation, does not scale linearly with  $\tau$ .

There may be some evidence for systematic variations of  $\tau$  at fixed  $M_s$ . Grootes et al. (2013) find evidence for redder galaxies to have lower dust masses, and a relation between attenuation and SFR may be one reason for the long established dependence of the ratio of FIR to UV luminosities of galaxies on SFR (Buat & Burgarella, 1998). It would not be surprising if galaxies with higher SFRs did have higher dust surface densities, since it is empirically well established that the surface density of SFR is dependent on the surface density of gas in galaxies, and one would expect the surface density of dust to be also correlated with the surface density of gas. To investigate these effects it would be desirable to construct a model for  $\tau$  with a specific dependence on SFR as well as stellar mass surface

density, and investigate whether the fits to the data were improved. It would furthermore be quite straight forward to further constrain such a model for dust content using metallicity measurements available from the GAMA spectroscopy.

By utilising the full  $M_r < 19.8\text{mag}$  GAMA sample, such an analysis could also extend to much lower values of  $\Sigma_s$  than were addressed in the analysis of the  $\Sigma_s$ - $\tau$  relation by Grootes et al. (2013). It will be particularly interesting to probe the dust content of low mass galaxies in an unbiased flux-limited sample bearing in mind recent results showing evidence for a non-linear variation of the dust-to-gas mass relation with metallicity (e.g. Draine et al., 2007; Galametz et al., 2011). This may throw light on mechanisms for dust injection and growth in low mass galaxies.

Finally, we note that one possible cause of the discrepancy between the value of  $\beta$  found in the present analysis and the original work of Grootes et al. (2013) is that while the latter depends on the value of grain emissivity in the submm regime, our analysis depends on the optical properties of grains in the UV and optical range. The analysis technique we have introduced will therefore be sensitive to systematic changes in the chemical composition of galaxies as a function of stellar mass.

In summary, the bias introduced by an (asymmetric) scatter in the  $\Sigma_s$  -  $\tau$  relation is most likely the cause of the high  $\beta$  value favoured for the GAMA spiral galaxy sample. However, this bias needs to be further investigated through detailed simulations to make a definite statement.



# Chapter 8

## Summary and outlook

In the first part of this thesis, we have presented the GALEX-GAMA survey. We first introduced the GAMA spectroscopic survey and the GALEX mission in Sects. 2.1 and 2.2, respectively. In Sect. 3 we described in detail how we incorporate the GALEX data into the GAMA spectroscopic survey. Specifically, we presented the individual data products and their derivation, with special focus on the matched catalogue using the advanced matching technique (see Sect. 3.4) and the catalogue containing the curve-of-growth re-measurements of the NUV (see Sect. 3.5) as these techniques were developed in this thesis.

In the second part of this thesis, we presented an application of the derived NUV photometry: a probabilistic analysis of the dust content of GAMA spiral galaxies via the intrinsic NUV-r colour, incorporating for the first time inclination and stellar mass-dependent attenuation corrections. We developed a sophisticated simulation of the spiral galaxy population (see chapter 6) including detailed features constrained by the comparison of the predicted observables of the simulation and the actual observations of the real data sample. We presented a method for the analysis of the dust content via the stellar mass-dependent intrinsic colour distribution in detail in Sect. 7.1. In Sect. 7.5 we use our simulated data to demonstrate the severity of the biases caused by the r-band detection limit, NUV noise and finite disk thickness if not accounted for in the analysis. We presented a probabilistic treatment of the finite disk thickness and identified an r-band flux range which allowed us to reduce the bias caused by NUV noise to a negligible level while preserving the colour distribution. We qualitatively discussed the influence of the remaining bias caused by the r-band detection limit on the results of the analysis.

In Sect. 7.7, we applied our analysis to the GAMA spiral galaxies data set and found that much higher dust masses were favoured by the analysis than expected from the literature. We discussed in detail potential technical and scientific reasons for this result in Sects. 7.7.3 and 7.7.4, respectively. We concluded that, rather than being a physical result contradicting with previous literature, this result demonstrates the necessity of using a model which does not only consider a fixed relation between dust content and stellar mass, but also an intrinsic scatter in this relation.

As discussed in the technical and scientific discussion of the results in Sects. 7.7.3 and 7.7.4, respectively, it is necessary to account for several further effects potentially influencing the analysis results before we can make definite inferences about

the detailed physical dependencies of the dust content in galaxies and the resulting attenuation of the galaxy’s emission. In the following sections, we describe the immediate steps necessary to reach that goal.

## 8.1 Accounting for r-band detection limit and NUV noise

The two major biases identified in Sect. 7.5 are the bias caused by the r-band detection limit and the bias caused by the NUV noise. We were able to minimize the latter effect to a negligible level, but paid a high price for it in form of reducing our data sample size to one tenth. The former bias was not accounted for at all.

In Appendix C, we outline the mathematical framework for accounting for the colour incompleteness caused by the r-band detection limit in the likelihood calculation. In other words, we do not use the probability distribution of the intrinsic colour of the population (see Eq. 7.5), we rather modify Eq. 7.5 such that it represents the probability distribution of the intrinsic colour of the detected spiral sample (i.e. we identify the probability  $P(m_{r,intr}|m_{r,obs}, \phi)$  in Eq. 4.1). This is done by accounting for the probability of having a galaxy with the observed r-band magnitude at the observed distance, given its stellar mass (see chapter 4, Eq. 4.1).

As discussed in Sects. 6.3.3.2 and 7.5.3, the NUV noise causes the observed NUV-r colour distribution to be broader, to be skewed towards redder colours and to have a more prominent red tail in comparison with the intrinsic NUV-r colour distribution. In Appendix B, we present the mathematics of the convolution of the intrinsic colour probability with the NUV noise distribution. Fig. B.1 compares the observed distribution of the NUV magnitude with the intrinsic distribution of the NUV magnitude convolved with the noise distribution for simple toy data. Both distributions agree very well and illustrate the broadening, the skewness and the prominent faint tail compared with the intrinsic NUV magnitude distribution.

## 8.2 Maximum likelihood with prior

In the current version of the algorithm, we do not put constraints on the parameter  $\mu$  and  $\sigma$  derived via the maximum likelihood method. In order to make use of the prior assumption that scaling relations in nature are narrow, we artificially defined a criterion for the identification of the best-fitting dust model parameter which is based on the minimisation of the scatter of the intrinsic colour. This prior knowledge can also be included in the maximum likelihood derivation by designing a prior for  $\sigma$  which punishes large values for  $\sigma$  (e.g. an exponential function falling off rapidly and thus suppressing high  $\sigma$  values).

## 8.3 Probabilistic treatment of spiral selection

In principle, we should consider the probability of a galaxy being identified as spiral galaxy, because the spiral selection slightly changes the shape of the r-band

luminosity function and thereby the effect of the r-band detection limit. Since in this work we employed a hard cut using the Grootes et al. (2014) cell-based spiral selection, the probability of a galaxy being identified as spiral galaxy is not identical with the probability of being a spiral galaxy. True spiral galaxies which have  $M_i$ ,  $r_{eff}$  and  $n$  combinations that fall in cells dominated by non-spiral galaxies will always be excluded from the samples using a hard cut.

Alternatively, since the distribution of spiral and non-spiral galaxies overlap in the parameter space, it is desirable to assign probabilities of being a spiral to each galaxy. The analysis then fits spiral and non-spiral galaxies simultaneously, fitting different parameters for each population. The likelihood of seeing a galaxy with the given observables  $\Phi_{obs}$  would then be the combination of the likelihood of seeing a spiral galaxy  $P(\Phi_{obs}|s)$  and the likelihood seeing a non-spiral galaxy  $P(\Phi_{obs}|ns)$ , weighted by the probability of the considered galaxy being a spiral  $P(s|M_i, r_{eff}, n)$  or an non-spiral  $P(ns|M_i, r_{eff}, n)$  respectively.

$$P(\Phi_{obs}) = P(s|M_i, r_{eff}, n) \cdot P(\Phi_{obs}|s) + P(ns|M_i, r_{eff}, n) \cdot P(\Phi_{obs}|ns) \quad (8.1)$$

This, however, requires to optimize for twice as many parameters, as the spiral and the non-spiral population are not expected to have the same luminosity function or intrinsic colour distribution.

## 8.4 Iterative approach to the derivation of the dust model and of the attenuation corrected r-band luminosity function

The derivation of the dust model and its parameter is highly sensitive to the colour incompleteness of our flux-limited sample. In Appendix C, we describe how to include the colour incompleteness due to the r-band detection limit into the statistical derivation of the dust model. The correct description of the incompleteness depends on the intrinsic r-band luminosity function, which we derive in Appendix A. However, the derivation of the r-band LF itself is depending on the dust model (and its parameters) assumed. Therefore, an iterative approach is necessary, which feeds the results of the LF derivation into the derivation of the dust model (parameters) and the results of the dust model derivation back into the LF derivation, until we obtain a consistent derivation of both.

## 8.5 Self-consistent calculation of stellar mass

In this work, we have made used the stellar mass estimates of Taylor et al. (2011). As described in more detail in Sect. 5.3.1, the stellar Mass estimate already contains an intrinsic dust attenuation correction assuming single foreground dust sheets, i.e. they were derived using a different attenuation model than we investigate in this work. A self-consistent derivation of the stellar mass using the same attenuation model considered in the analysis is desirable.

## 8.6 Accounting for the effects of dust attenuation on measured size, axis ratio, and Sérsic index

The measurement of the effective radii of disk galaxies using single-Sérsic fits is strongly biased by the presence of a bulge as well as by the presence of dust in the disk.

Adding to the brightness of the galaxy profile in the centre, the bulge will lead to higher Sérsic indices and thus smaller effective radii with increasing bulge-to-disk ratio.

At the same time, dust present in the disk biases the radial light profile we see from a galaxy, making it flatter in the centre. The single-Sérsic fits will prefer a lower Sérsic index and thus a larger effective radii for higher dust content. This effect is even more pronounced for high inclination galaxies than for low inclination galaxies, and it is in general very sensitive to the geometry of the dust distribution. Furthermore, the wavelength dependence of the dust attenuation will cause the bias of the effective radius to be a function of wavelength. This has been quantitatively predicted for pure disk systems (Möllenhoff et al., 2006; Pastrav et al., 2013a), and has been observed in the wavelength dependence of the size of galaxies (e.g. Kelvin et al., 2012; Häußler et al., 2013).

Pastrav et al. (2013a,b) studied in detail the effect of dust attenuation on shape fitting of composite systems and provide correction terms for the change in the semi-major axis as a function of inclination, dust content, and bulge-to-disk ratio in form of lookup tables, which we can use to derive the intrinsic  $r_{eff}$  from the observed  $r_{eff}$ . Furthermore, Pastrav et al. (2013a,b) also provide corrections for the effect of dust attenuation on the Sérsic index.

## 8.7 Empirical correlations of dust content with physical quantities

In the work presented here, we used the radiation transfer model from Popescu et al. (2011) as an attenuation model and modelled the dust content either as constant for all galaxies or following the stellar mass - dust mass relation of Grootes et al. (2014). We investigated the hypothesis that the dust mass correlates with the total stellar mass. Alternative models could correlate the dust mass with the mass of young stars or with the mass of old stars independently (e.g. Cortese et al., 2008, found a strong dependence of dust mass with the age of the stellar population). For example a dust mass correlation with SFR (total SFR or SFR per area) or r-band luminosity (total or per area) could be used as tracer of the young and old stellar mass, respectively. More sophisticated, physically-based hypotheses predicting dust content of the diffuse ISM as a function of intrinsic SFR and intrinsic stellar mass could readily be tested using the algorithm developed in this work. In particular, different physical prescriptions for grain injection, growth and propagation in the ISM, and grain destruction could be tested using this technique. With the present parametrisation involving SFR and  $M_s$  parameters, it will be straight forward to test what the grain injection is from the young stellar population. It will also be possible to test models using metallicity data to constrain the available metals for grains in galaxies.

## 8.8 Comparing subsamples of galaxies

Another interesting aspect of modelling the dust content of galaxies would be to model the dust content for subsamples of disk galaxies defined by any observable property. Doing so, we can potentially investigate a range of factors influencing dust content and/or intrinsic colour of disk galaxies, such as environment (field galaxies vs. group galaxies), AGN activity, metallicity, or proximity to other galaxies.

## 8.9 Alternative attenuation models

The method can be used to test any physically or empirically based model for the attenuation of starlight by dust. A large variety of model for dust attenuation need to be compared and contrasted. This should range from simple empirical descriptions to radiation transfer models with physically motivated prescription of dust content in galaxies.

## 8.10 Multi-wavelength approach

The NUV-r colour is one of the GAMA colours most affected by attenuation and is thus the ideal first choice for our study of dust attenuation. However, as the RT model makes predictions of dust attenuation for a wide range of wavelengths, it is desirable to extend our analysis to further wavelengths / colours. Doing so, we can test the predictions of the RT model and possibly optimize the physical assumptions made. Since we can deduce  $\tau$  independently at each wavelength, it will in principle be possible to measure the extinction law ( $\tau$  vs. wavelength) for statistical samples of disk galaxies as a function of  $M_s$  and any of the other properties listed in Sect. 8.8.



Part III  
Appendices



# Appendix A

## r-band Luminosity Function

The GAMA spectroscopic survey is a multi wavelength survey, with an input catalogue based on the SDSS r-band. This means that the r-band plays a key role in the survey, especially regarding the r-band detection limit and effects that influence this detection limit.

Analysing the measured fluxes of galaxies, we need to be aware of several effects, e.g. dust attenuation or other systematic shifts, which introduce biases to the measurements. These biases cause the observed luminosity distribution to differ from the intrinsic luminosity distribution.

In order to better understand, simulate and compensate completeness biases of the GAMA spiral galaxy sample, as well as for its own sake, we want to derive the intrinsic r-band luminosity distribution of the GAMA spiral galaxies.

Previous work (e.g. Loveday et al., 2012; Driver et al., 2012), either derived a luminosity function for the full GAMA galaxy population, or employed different spiral selection criteria and simplistic models for the dust content.

In this appendix, we present a Bayesian model to calculate the probability of observing a given data set. We start with calculating the probability of seeing an individual galaxy, including the intrinsic r-band luminosity function in Sect. A.1.1, the distance distribution in Sect. A.1.2, and the detection limit in Sect. A.1.3. We then present the treatment of dust attenuation in Sect. A.1.4, and the compensation of the bias introduced by the detection limit in combination with the attenuation and the k-correction (Sect. A.1.5).

We then describe the fit of the Schechter function parameters by maximizing the joint probability of the data set in Sect. A.2. Tests of this method on simulated data and illustrations of some biases are presented in Sect. A.3.

The final results for the fit to the GAMA spiral galaxy data set are described in Sect. A.4.

### A.1 Functional form of the luminosity function

#### A.1.1 Luminosity function of the intrinsic absolute magnitude

The commonly used function to describe the luminosity distribution is the Schechter function (Schechter, 1976), with model parameter faint-end slope  $\alpha$ , turnover magnitude  $M_*$ , and amplitude  $\phi^*$ . For the discussion of the functional form we refer

to Sect.6.1.1 and only quote Eqs.A.1 and A.2 here:

$$LF(L|\phi^*, \alpha, L_*) d(L/L_*) = \phi^* \cdot \left(\frac{L}{L_*}\right)^\alpha \cdot e^{-L/L_*} d(L/L_*) \quad (\text{A.1})$$

$$LF'(M|\phi^*, \alpha, M_*) dM = 0.4 \cdot \ln(10) \cdot \phi^* \cdot (10^{0.4 \cdot (M_* - M)})^{(1+\alpha)} \cdot e^{(-10^{0.4 \cdot (M_* - M)})} dM \quad (\text{A.2})$$

For our analyses, we are interested in calculating the probability of measuring a galaxy at a given absolute magnitude  $M$ . We therefore need to normalize Eq. A.2 over the considered magnitude range.

$$c_1 \cdot LF(M|\phi^*, \alpha, M_*) = P_{SF}(M) \quad \text{with} \quad \int_{M_{min}}^{M_{max}} P_{SF}(M) dM = 1 \quad (\text{A.3})$$

The upper magnitude limit (i.e. the faint end side) must be chosen so that a galaxy with this upper-limit-magnitude at the smallest distance possible does not meet the detection limit anymore (see Sect. A.1.2 and A.1.3 for the description of the distance distribution and the detection limit, respectively).

The integration limit on the bright end should in principle be  $-\infty$ . However, both for the normalisation and for the simulation of data samples, the value  $-\infty$  for the lower limit is inconvenient. It is sufficient to choose a lower limit significantly brighter than the turnover magnitude  $M_*$  (due to the shape of the function, the exact bright limit has negligible effect of the integral value; see Sect 6.1.1).

$$P_{SF}(M|\phi^*, \alpha, M_*) dM = c_1 \cdot 0.4 \cdot \ln(10) \cdot \phi^* \cdot (10^{0.4 \cdot (M_* - M)})^{(1+\alpha)} \cdot e^{(-10^{0.4 \cdot (M_* - M)})} dM \quad (\text{A.4})$$

In practice, we are actually interested in the probability of detecting a galaxy with magnitude  $M$  given a number  $N$  of detected galaxies in an area on of sky  $\Delta\Omega$ . We will include these information in Sect. A.2 in the calculation of the Schechter function amplitude  $\phi^*$ .

## A.1.2 Distance distribution

In order to be able to analyse intrinsic quantities of galaxies, it is crucial to be able to account for the different appearance of a galaxy due to its distance to the observer. In particular, we need to convert the measured flux into luminosity (i.e. apparent into absolute magnitude) in order to make it comparable to other galaxies. To do this, we need the distribution of the luminosity distance  $D_L$ , the actual distance the light has travelled on its journey to the observer.

The luminosity distance  $D_L$  is defined in Eq. 6.9 as

$$D_L = (1 + z) \cdot D_C(z) \quad (\text{A.5})$$

Assuming an euclidean universe, the probability distribution of the comoving distance  $D_C$  is a quadratic function. We briefly discuss the cosmology in Sect. 6.1.2.

$$P_{D_C}(D_C) dD_C = c_2 \cdot D_C^2 dD_C = 3 \cdot c_2 \cdot d(D_C^3) \quad (\text{A.6})$$

where  $c_2$  is the normalisation constant defined by

$$\int_{D_{C,min}}^{D_{C,max}} P_{D_C}(D_C) dD_C = 1 \quad (\text{A.7})$$

The integration limits  $D_{C,min}$  and  $D_{C,max}$  are calculated from the lower and upper redshift limit of the considered volume  $z_{min}$  and  $z_{max}$  using Eq. 6.8.

It is convenient to choose  $z_{min} > 0$ , in order to avoid the lower integration limit of the normalisation of the Schechter function in Eq. A.4 to become infinity. Choosing  $z_{min} = 0.002$  for this work, we gain numerical stability in the analysis and reject only a handful of galaxies from the GAMA sample which have even lower redshifts. Furthermore, galaxies in such a small volume will suffer from clustering effects. While these effects average out in magnitude ranges detectable over a larger volume, this effect might introduce a bias on the faintest magnitudes detectable only in this small volume.

For simplicity reasons, we will declare a function  $F$  as  $F(D_C)$  if it is a function of distance and specify in the definition whether it is a function of the comoving distance directly or indirectly via the luminosity distance defined in Eq. 6.9.

### Probability of measuring a galaxy

Combining the probability of having a galaxy with intrinsic absolute magnitude  $M_{intr}$  defined in Eq. A.4 with the probability of this galaxy having a comoving distance  $D_C$  to the observer, we obtain the joint probability of having a galaxy with  $M_{intr}$  and  $D_C$ . Since  $P_{D_C}(D_C)$  and  $P_{SF}(M_{intr}|\phi^*, \alpha, M_*)$  are independent, their joint probability  $P'(M_{intr}, D_C)$  becomes the product of the individual probabilities.

$$P'(M_{intr}, D_C|\phi^*, \alpha, M_*) = P_{D_C}(D_C) * P_{SF}(M_{intr}|\phi^*, \alpha, M_*) \quad (\text{A.8})$$

### A.1.3 Detection limit

The GAMA spectroscopic survey is based on the SDSS survey with a foreground attenuation removed r-band apparent magnitude detection limit of  $m_{lim} = 19.8$ . Therefore, we need to be aware of the fact that a galaxy of a given intrinsic absolute magnitude  $M_{intr}$  is observable only out to a certain distance (see subsection below on "Malmquist bias").

Especially for the analysis of binned data, it is very popular to correct for this bias by weighting the observed data using the  $1/V_{max}$  method introduced by Efstathiou et al. (1988). This approach however is dependent on observing at least one galaxy at a given absolute magnitude, to which the weight can then be assigned. Furthermore, this approach suffers from Poisson noise from counting the actually detected galaxies, which is then amplified by the assigned weight. We can avoid these disadvantages by designing our model to account for this bias rather than to correct the data for it.

In first order, this is already done by evaluating the joint probability  $P'(M_{intr}, D_C)$  of a galaxy having a given  $M_{intr}$  and  $D_C$ . Rather than averaging over all galaxies with similar  $M_{intr}$ , disregarding  $D_C$ , we perform our analysis keeping both information. In this 2D space, a combination of  $M_{intr}$  and  $D_C$  leads either to a magnitude brighter or fainter than the detection limit, no weighting needed.

Given a hard cut in apparent magnitude<sup>1</sup> and using the conversion function from absolute to apparent magnitude in Eq. 6.10, this can be formulated as

$$\eta(m) = \Theta(m_{lim} - m) \quad (\text{A.9})$$

where  $\Theta$  is the Heaviside function, becoming unity if the detection limit  $m_{lim}$  is fainter than or equal to the considered  $m$ , and zero otherwise.

However, since we want to include effects in our model that change the apparent magnitude (e.g. dust attenuation, see Sect. A.1.4, especially Sect. A.1.4), the above mentioned detection limit is not applied to the intrinsic magnitude, but rather to the dust-attenuated and frequency shifted observed magnitude, which will be defined in the following sections. We therefore used  $M$  rather than  $M_{intr}$  in Eqs. A.9.

### Probability of measuring a galaxy

Thus, the joint probability  $P'(M, D_C)$  as defined in Eq. A.8 needs to be multiplied by  $\eta(m)$  and re-normalized.

$$P''(M, D_C) = c_3 \cdot P'(M_{intr}, D_C) \cdot \underbrace{\eta(M + 5[\log_{10}(D_L) - 1])}_m \quad (\text{A.10})$$

with the normalization constant  $c_3$  defined by

$$\int_{D_{C,min}}^{D_{C,max}} \int_{M_{min}}^{M_{max}} P''(M, D_C) dM dD_C = 1 \quad (\text{A.11})$$

### Malmquist bias

The tendency of a luminosity limited survey to be more likely to detect intrinsically brighter objects than intrinsically fainter objects is referred to as Malmquist bias Malmquist (1920). The bias results from the fact that an intrinsically brighter object can be detected in a larger volume than an intrinsically fainter object, therefore more intrinsically brighter objects will be overall detected than intrinsically fainter ones. This effect is accounted for in our model by applying the apparent magnitude detection limit.

---

<sup>1</sup>This only holds under the assumption of no absorption / scattering of the light, e.g. assuming the absence of dust in the galaxy itself and no foreground extinction. Otherwise Eq. 6.10 needs to be extended to include these effects.

### A.1.4 Dust attenuation

The observed distribution of the magnitude is different to the distribution of the intrinsic magnitude due to the attenuation caused by the dust in the galaxy's own disk. The light produced by the stars of a galaxy travels through dust on its way escaping the galaxy. Some fraction of this light is absorbed (and scattered) by the dust and the energy is re-radiated in the far infrared.

Using the attenuation model described in Sect. 5.4.1, we can derive the absorbed fraction of light and subtract the magnitude of the dust attenuation  $A(\tau, i)$  from the observed, dust-attenuated absolute magnitude  $M_d$  in order to obtain the intrinsic absolute magnitude  $M_{intr}$ <sup>2</sup>.

$$M_d = M_{intr} + A(\tau, i) \quad \Rightarrow \quad P''(M_{intr}, D_C) = P''(M_d - A(\tau, i), D_C) \quad (\text{A.12})$$

Since  $A(\tau, i)$  is a function of the dust  $\tau$  and the inclination angle  $i$ , we need to convolve the probability distribution with the probability distribution of the attenuation  $P_A(A(\tau, i))$ .

$$P'''(M_d, D_C) = \int \int P''(M_d - A(\tau, i), D_C) P_A(A(\tau, i)) d\tau di \quad (\text{A.13})$$

The inclination angle  $i$  and the dust  $\tau$  are assumed to be independent variables<sup>3</sup>. The probability  $P_A(A(\tau, i))$  therefore factories into the individual probabilities  $P_\tau(\tau)$  and  $P_i(i)$ .

$$P_A(A(\tau, i)) d\tau di = P_\tau(\tau) d\tau \cdot P_i(i) di \quad (\text{A.14})$$

Since galaxies are randomly oriented in 3D space, the distribution of the intrinsic inclination (i.e. without any potential bias introduced by the measurement of the inclination) is flat in  $d(1 - \cos i)$ :

$$P_i(i) = \begin{cases} (1 - \cos i) & \text{if } 0 \leq i \leq \pi/2 \\ 0 & \text{otherwise} \end{cases} \quad (\text{A.15})$$

and is derived in detail in Sect. 6.1.7.2. The treatment of potential biases is discussed in Sect. 6.3.3.3. (Note that we do not use the actual inclination measurements here. Therefore, we are not suffering from the bias caused by the difference between  $b/a$  and  $\cos(i)$ .)

The distribution of  $\tau$  for spiral galaxies is one of the outcomes of this work. For the LF, we assumed a delta function of  $\tau=4$  following Driver et al. (2007), which in retrospect turned out to be too large for the GAMA spiral galaxy data set.

---

<sup>2</sup>Note that  $M_D$  is not the absolute magnitude "as observed", since we also must consider the A- and k-corrections described in Sects. A.1.5, respectively Sects. 5.1.2.2 and 5.1.2.2.

<sup>3</sup>In practice, the inclination measurement of an individual galaxy is biased by the morphology (e.g. intrinsic thickness of the disk or bulge to disk ratio), which in turn might also influence the dust content. Furthermore, the inclination can influence the size estimate which is used in the Grootes et al. (2014) model to derive the dust content (see Sect. 6.2.1 for further discussion). However, for the r-band LF, we consider the full galaxy sample as ensemble rather than the attenuation of the individual galaxy, which justifies the assumption of independence.

The need of a renormalisation of Eq. A.13 is discussed in detail in the following subsection.

In the future, an iterative approach is desirable, feeding the  $\tau$ -distribution resulting from the analysis of the intrinsic colour back into the r-band LF fit (see outlook Sect. 8). Furthermore an object-to-object defined attenuation value would also improve the LF fit, even though we still need  $P_A(A(\tau, i))$  of the full sample for normalization reasons (see below).

### Accounting for the inclination bias caused by the detection limit

The attenuation of a galaxy's light due to the presence of dust increases with the galaxy's inclination since the path of the light through the galaxy is longer. Therefore, the observed magnitudes of edge-on galaxies are statistically fainter than those of face-on galaxies.

In combination with a detection limit, this leads to an inclination bias: The observed sample will have less edge-on galaxies. The observed inclination distribution will differ from a flat distribution in  $d \cos i$  derived in Sect. 6.1.7.2.<sup>4</sup> Looking at this bias in terms of the distribution of the intrinsic magnitude: galaxies with faint intrinsic magnitudes will be underrepresented in the data sample as they pass the detection limit only for low inclination, but fail it for high inclinations.

It is therefore important to apply the detection limit to the dusty magnitudes including the full attenuation distribution. While this is more or less automatically done for the real data, we need to pay attention to our model, especially to the normalization of it, since assuming a mean attenuation for all galaxies will not account for the inclination bias and thus it will introduce a bias to the analysis.

Contrary to the first impression, Eq. A.13 is not properly normalized. Even though we convolved the normalized probability distribution  $P''(M_{intr}, D_C)$  with another normalized probability distribution  $P_A(A)$ , we also shifted the distribution by the attenuation and we need to account for this shift in the detection limit function  $\eta$ .

In theory, we only need to make sure that it is the actually observed magnitude, i.e. the dusty apparent magnitude  $m_d$ , is the one compared with the detection limit  $m_{lim}$ . We can therefore adapt Eq. A.9 by substituting

$$\begin{aligned} \eta(m_{intr}) &= \Theta(m_{lim} - m_{intr}) && \text{with} \\ \eta(m_d) &= \Theta(m_{lim} - m_d) = \Theta(m_{lim} - m_{intr} - A(\tau, i)) \end{aligned} \quad (\text{A.16})$$

However, normalizing the probability distribution for each trial parameter combination by integrating over the dusty absolute magnitude  $M_D$ , we would need to calculate the full distribution of  $M_{intr}$  for each  $M_D$  used in the integration.

$$\int \int c_4 \cdot P'''(M_d, D_C) dM_d dD_C = 1 \quad (\text{A.17})$$

---

<sup>4</sup>Note: The same line of argument holds for the distribution of  $\tau$ . Galaxies with high  $\tau$  are biased low in the data sample. Even though we focus in the text on the bias in inclination  $i$ , the formula automatically account for the joint dependence on  $\tau$  and  $i$ .

$$= \int \int c_4 \cdot \int \int P''(M_d - A(\tau, i), D_C) P_A(A(\tau, i)) d\tau di dM_d dD_C \quad (\text{A.18})$$

$$= \int \int c_4 \cdot \int \int c_3 \cdot P'(M_d - A(\tau, i), D_C | \phi^*, \alpha, M_*) \cdot \quad (\text{A.19})$$

$$\underbrace{\eta(M_d + 5[\log_{10}(D_L) - 1])}_{m_d} \cdot P_A(A(\tau, i)) d\tau di dM_d dD_C$$

Evaluating  $P'(M_d - A(\tau, i), D_C)$  for each combination of  $\tau$  and  $i$  for each trial parameter combination of  $\phi^*$ ,  $\alpha$ , and  $M_*$  is costly and can be avoided.

Rather than integrating over the dusty absolute magnitude  $M_D$  for each parameter combination, in practice, it is more convenient to integrate over the intrinsic magnitude  $M_{intr}$ .<sup>5</sup>

$$\int \int c_4 \cdot P'''(M_{intr} + A(\tau, i), D_C) dM_{intr} dD_C = \int \int \int \int c_4 \cdot c_3 \cdot P'(M_{intr}, D_C | \phi^*, \alpha, M_*) \cdot \underbrace{\eta(M_{intr} + A(\tau, i) + 5[\log_{10}(D_L) - 1])}_{m_d} \cdot P_A(A(\tau, i)) d\tau di dM_{intr} dD_C \quad (\text{A.20})$$

We identify

$$F_{det}(M_{intr}, D_C) = \int \int \underbrace{\eta(M_{intr} + A(\tau, i) + 5[\log_{10}(D_L) - 1])}_{m_d} \cdot P_A(A(\tau, i)) d\tau di \quad (\text{A.21})$$

as Schechter function parameter independent, thus as a constant given  $M_{intr}$  and  $D_C$ . We only need to calculate  $F_{det}(M_{intr}, D_C)$  once for each  $(M_{intr}, D_C)$  combination in the integral and use it as a look-up table during the fitting process. Thus, the normalization is therefore

$$\int \int c_4 \cdot P'''(M_{intr} + A(\tau, i), D_C) dM_{intr} dD_C = 1 \\ = \int \int c_4 \cdot c_3 \cdot P'(M_{intr}, D_C | \phi^*, \alpha, M_*) \cdot F_{det}(M_{intr}, D_C) dM_{intr} dD_C \quad (\text{A.22})$$

$F_{det}(M_{intr}, D_C)$  can be interpreted as the detection fraction, the fraction of galaxies with  $M_{intr}$  and  $D_C$  that will be detected (illustrated in Fig. A.1 for simulated data). For bright galaxies, the attenuated apparent magnitudes will be brighter than the detection limit for all inclinations and  $\tau$ -values, thus  $F_{det}(M_{intr}, D_L) = 1$ . For faint galaxies, galaxies with a higher  $\tau$  and higher inclination will fall below the detection limit, causing  $F_{det}(M_{intr}, D_L) < 1$ .

Note: When calculating the probability of detecting an actually observed galaxy, we know by definition that  $m_d = M_{intr} + 5[\log_{10}(D_L) - 1] + A(\tau, i) < m_{lim}$ , i.e. that  $m_d$  passed the detection limit<sup>6</sup>. The correction term is only needed for the normalisation (respectively the calculation of the amplitude  $\Phi^*$ ).

<sup>5</sup> $dM_{intr} = dM_d$ , integration boundaries change according to Eq. A.12.

<sup>6</sup>For any trial combination of  $(M_d, D_C)$ , the detection limit is a hard cut requiring  $m_d < m_{lim}$ .

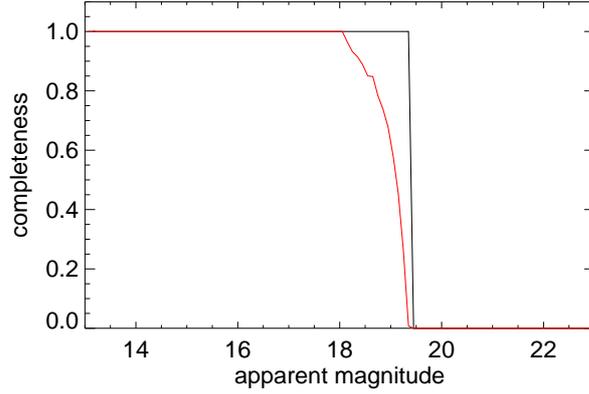


Figure A.1: Completeness in presence of dust

A comparison of the completeness of the sample as function of the intrinsic and the attenuated, dusty apparent magnitude for simulated data with  $\tau = 4$ , is shown. The completeness in the actually observed, dusty magnitude  $m_d$  (black line) is 100% up to the detection limit and then drops to 0, mirroring the hard cut of the detection limit. The completeness of the intrinsic apparent magnitude  $m_{intr}$  (i.e. the magnitude the galaxy would have without dust present; red line) drops already significantly for brighter magnitudes than the detection limit. Due to the distribution of dust attenuation, the detection limit becomes smeared out (since  $m_{intr} + A = m_d$ ).

### A.1.5 Restframe magnitude

The observed light is not only attenuated by intrinsic and foreground dust, it is also shifted in wavelength. Thus, we need to include the k-corrections (see Sect. 5.1.2.3) as another link between intrinsic and observed magnitude.

If we define the 'actual' apparent magnitude as  $m = m_{obs} - k_{corr} - A_{corr}$  with  $m_{obs}$ ,  $k_{corr}$ , and  $A_{corr}$  being the observed magnitude, the k-correction, and the A-correction, respectively, all galaxies are observed which meet the detection limit<sup>7</sup>

$$m_{obs} - A_{corr} = m + k_{corr} \leq m_{lim} \quad (\text{A.23})$$

Besides correcting each data point using its individual k-correction, we need to be aware of the incompleteness in  $m$  of the sample caused by this shift in combination with the detection limit. We need to derive an incompleteness correction in order to properly normalize our model, even though  $k_{corr}$  and  $A_{corr}$  is known for each individual galaxy. Fig. 5.2, left panel, shows a histogram of the r-band k-correction  $k_{corr}$  distribution for the spiral sample of the GAMA galaxies (black dashed line). The distribution is fairly narrow and more or less symmetrical. Furthermore, beyond  $z=0.04$ , the scatter in  $k_{corr}$  dominates the redshift dependence over the considered redshift range (see Fig. 5.3, right panel).

For the normalization of the likelihood, we therefore simplify the distribution to be a delta function at  $\overline{k_{corr}}$ , the mean value of the k-correction distribution. We then redefine the detection limit for the normalization to be

$$m \leq m_{lim} - \overline{k_{corr}} \quad (\text{A.24})$$

<sup>7</sup>Remember, the detection limit is applied to the A-corrected observed magnitude.

Fig. A.3 shows the best-fitting parameter combination for 200 data sets of simulated galaxies with dust (bottom panel) and without dust (top panel). Both datasets were analysed correcting each data point with its individual  $k_{corr}$  and  $A_{corr}$ . Once, the analyses ignored the incompleteness of the rest frame magnitude (blue), and once the incompleteness was corrected for by the redefined detection limit (black). For both the dusty and the dust-free data sets, the best-fitting parameters found by the analyses ignoring the incompleteness are biased and do not reproduce the input parameter (black lines). The analyses using the redefined detection limit is able to reproduce the input parameter with the same precision as for simulations without rest frame magnitude incompleteness, thus demonstrating that our simplification of the k-correction distribution is adequate.

### A.1.6 Full r-band model

To summarize all steps discussed above, we present the full model of the r-band luminosity distribution here. In Eq.A.13, the probability of measuring a dust-attenuated absolute magnitude  $M_D$  at a comoving distance  $D_C$  is defined as

$$P(M_d, D_C) = \int \int P''(\underbrace{M_d - A(\tau, i)}_{M_{intr}}, D_C) P_A(\tau, i) d\tau di$$

We substitute the actual probability distributions as defined in Eqs. A.10, A.8, A.6, and A.17.

$$P(M_d, D_C) = c_3 \cdot c_2 \cdot c_4 \cdot \int_0^{\pi/2} \int_0^\infty D_C^2 \cdot P_{SF}(M_d - A(\tau, i) | \phi^*, \alpha, M_*) \cdot \underbrace{\eta(M_d + 5[\log_{10}(D_L) - 1])}_{m_d} \cdot P_A(A(\tau, i)) d\tau di \quad (\text{A.25})$$

with  $P_{SF}$ ,  $\eta$ , and  $P_A$  defined in Eqs. A.4, A.16 and A.24, and A.14 respectively.

## A.2 Deriving the best-fitting luminosity function parameters

The luminosity function parameters in our analysis are the amplitude  $\phi^*$ , the faint-end slope  $\alpha$  and the turn-over magnitude  $M^*$ . While  $\alpha$  and  $M^*$  are non-linear parameters, the amplitude  $\phi^*$  is a linear parameter which can be calculated analytically for a given combination  $\alpha$  and  $M^*$  as presented in Sect.A.2.1. In Sect. A.2.2 we then present the actual maximum likelihood fit for  $\alpha$  and  $M^*$  and discuss the derivation uncertainties of the best-fitting parameters.

### A.2.1 Calculating the amplitude $\phi^*$

Our model as described in Eq. A.25 is defined as the probability of finding a galaxy with given  $(M_d, D_C)$  and is normalized such that the expectation value of finding a galaxy within the considered magnitude and redshift interval in the area of 1

steradian is unity. We deliberately chose this formulation because this definition allows us to derive our model as probability function rather than as number density. In practice however, the expectation value is  $N$  galaxies in the considered magnitude and redshift interval, measured in the area  $\Delta\Omega$  (in steradian) on the sky. Using the real expectation value, the normalization constant  $c_1$  in Eq. A.3 would be defined as

$$\int_{M_{min}}^{M_{max}} c_1 \cdot c_4 \cdot LF(M|\phi^*, \alpha, M_*) \cdot F_{det} \cdot \Delta\Omega dM = N \quad (\text{A.26})$$

taking into account the detection limit applied to the dust-attenuated magnitude and marginalized over the distance. However, since the Schechter function amplitude  $\phi^*$  and the normalization constants  $c_1, c_2, c_3$ , and  $c_4$  are 100% degenerate (e.g.  $c_1 \cdot LF(M|\phi^*, \alpha, M_*) = c_1 \cdot \phi^* \cdot LF(M|1, \alpha, M_*)$ ), it is possible to use the model description as probability function as defined in Eq. A.25 and to account for the real expectation value when calculating  $\phi^*$  during the fit.

For a given fit parameter combination  $(\alpha, M_*)$ , we calculate the value of  $\Phi^*$  that satisfies

$$\int \int P(M_d, D_C|\Phi^*, \alpha, M_*) dM_d dD_C = 1 \quad (\text{A.27})$$

and derive the Schechter function amplitude  $\phi^*$  with

$$\phi^* = \frac{N}{\Delta\Omega} \cdot \Phi^* \quad (\text{A.28})$$

### A.2.2 Deriving best-fitting $\alpha$ and $M^*$ using maximum likelihood method

Traditionally, the best-fitting parameters of a luminosity function are derived by binning the data in absolute magnitude bins and comparing the bin population with the prediction of the model (see e.g. Loveday et al., 2012; Hill et al., 2011; Driver et al., 2012, for LF fits to GAMA galaxies). As this requires the model to represent a number density integrated over the bin width and furthermore marginalizes over the distance (i.e. averaging over an important piece of information), we decided to use a maximum likelihood method as alternative to fitting the number density to binned data.

Sandage et al. (1979) first introduced the maximum likelihood approach to galaxy luminosity function estimation. Following Johnston (2011), the probability of a set of absolute magnitudes  $M_i$  given the model parameter is

$$L(\text{data}|\alpha, M_*) = \prod_i P(M_i, D_C|\alpha, M_*) \quad (\text{A.29})$$

where  $P(M_i, D_C|\alpha, M_*)$  is the probability of measuring a dust-attenuated absolute magnitude  $M_D$  at a comoving distance  $D_C$  as defined in Eq. A.25.

Since this probability distribution has very small values, it is more convenient to formulate Eq. A.29 as logarithmic likelihood<sup>8</sup>.

$$\log L(data|\alpha, M_*) = \sum_i \log(P_M(M_i|\alpha, M_*)) \quad (\text{A.30})$$

Since the maximum of  $L$  will always be the maximum of  $\log(L)$ , we can choose the more convenient formulation of Eq. A.30 to find the best-fitting parameter values. For that, we can use algorithms like gradient-method (used in this work) or random walk to maximize the likelihood of the data.

### Error estimation

The error of the derived best-fitting LF parameters can in principle be calculated directly from the shape of the likelihood surface in the parameter space (see e.g. Efsthathiou et al., 1988; Johnston, 2011). However, a easier approach is to evaluate the distribution of the best-fitting LF parameters for different simulated data sets with the same input parameters (i.e. different realisations of the same parent distribution, differing only due to sampling noise). The best-fitting parameters for the different data sets effectively sample the likelihood function around the "true" best-fitting parameters (i.e. in case of the simulation around the input parameters). We can fit a simple Gauss function to the parameter distribution and thus derive the standard deviation  $\sigma$  of the LF parameters.

---

<sup>8</sup>Multiplying a large number of very small values with one another, we will quickly run into numerical instabilities. Adding up large, negative number is much more stable.

## A.3 Tests on simulated data

### A.3.1 Best-fitting parameters for dust-free simulation

For 100 simulated dust-free data sets (including A- and k-corrections), the best-fitting parameters are shown in Fig. A.2. The maximum likelihood method is able to reproduce the input parameters correctly.

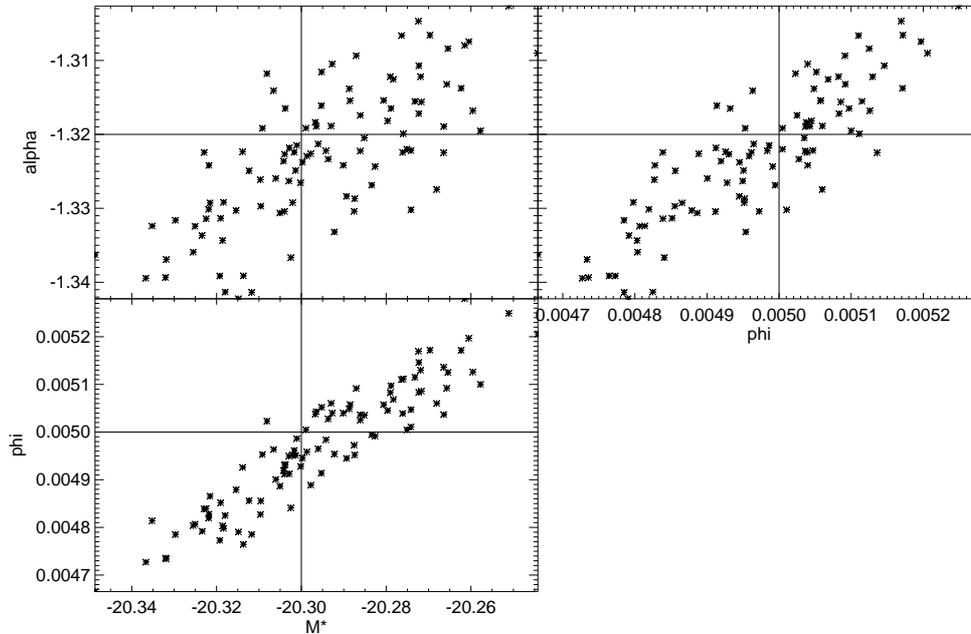


Figure A.2: Fit precision for dust-free galaxies

Shown are the distributions of the best-fitting parameter combinations for 100 simulated datasets, containing  $\sim 21.7$ k dust-free galaxies. The true input parameters of the simulation ( $\alpha = -1.32$ ,  $M_* = -20.3$ ,  $\phi^* = 0.005$ ,  $\tau = 0$ ) are indicated by the black lines.

### A.3.2 Normalization of the probability distribution containing dust

For Fig. A.3, 200 dust-free data sets (with dust,  $\tau = 4.0$ , displayed in the bottom panel,  $\sim 14.8$ k galaxies each; without dust in the top panel,  $\sim 20.3$ k galaxies each) which suffer from incompleteness due to the k-correction were analysed. Once without any incompleteness correction in the normalization (blue) and once correcting for the incompleteness by shifting the detection limit in the normalization by the mean of the k-correction. In both cases, each data point was accurately corrected by its exact k-correction value, i.e. restoring the true intrinsic magnitude. The black lines indicate the input Schechter function parameter ( $\alpha = -1.32$ ,  $M_* = -20.3$ ,  $\phi^* = 0.005$ ). While the blue cloud of best-fitting parameter is clearly biased, the shift of the detection limit is sufficient to remove the bias and to reproduce the input parameter (black cloud). The larger scatter in the best-fitting parameter of the dusty data sets compared with the dust-free data sets is due to the fact that dust attenuation causes the fainter galaxies to fall below the detection limit and thus reduces the size of each data set.

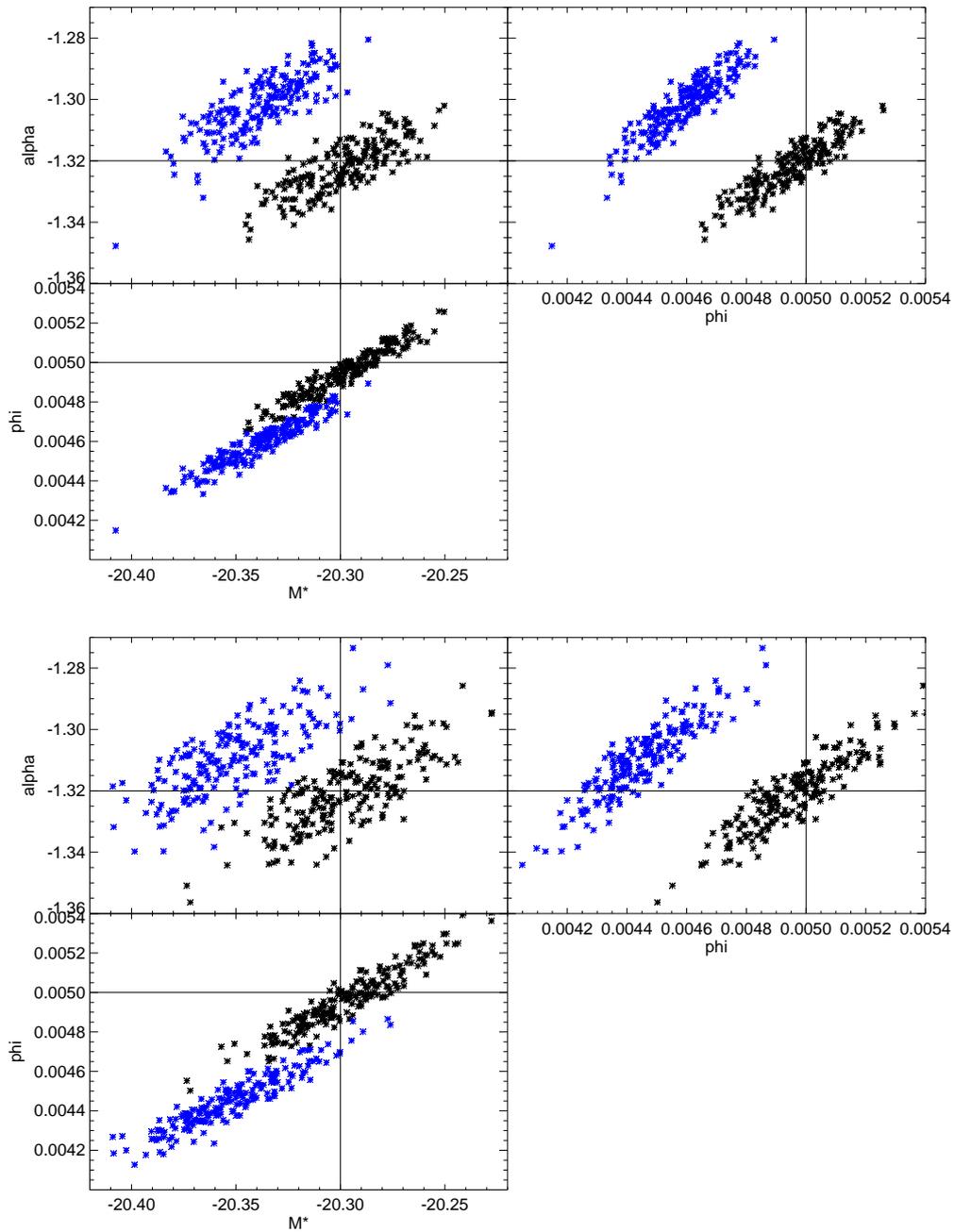


Figure A.3: Effect of k-correction incompleteness on the fit with and without bias removal. Shown are the best-fitting parameter combinations for dust-free data (top) and dust attenuated data (bottom) using the completely modelled likelihood (black) and without accounting for the effective shift of the detection due to the k-correction (blue). As is to be expected, the blue parameter cloud shows a significant bias for all three parameters. The black cloud reproduces the input values of the simulation nicely for both, the dust-free and the dusty case.

100 simulated data sets of  $\sim 16\text{k}$  dusty galaxies each (input parameter:  $\alpha = -1.32$ ,  $M_* = -20.3$ ,  $\phi^* = 0.005$ ,  $\tau = 4.0$ ) were analysed ignoring the effect of the inclination bias on the normalization of the likelihood. The best-fitting parameter combination for the fit parameter  $\alpha$ ,  $M_*$ , and  $\phi^*$  (given the true  $\tau$  and inclination values for each galaxy) are shown in Fig. A.4. There is a significant bias in the turnover magnitude  $M_*$  and the amplitude  $\phi^*$ , and a smaller bias in the faint-end slope  $\alpha$ .

This bias illustrates the necessity of accounting for the dust attenuation not only for the individual r-band magnitude, but also for the normalisation of the likelihood.

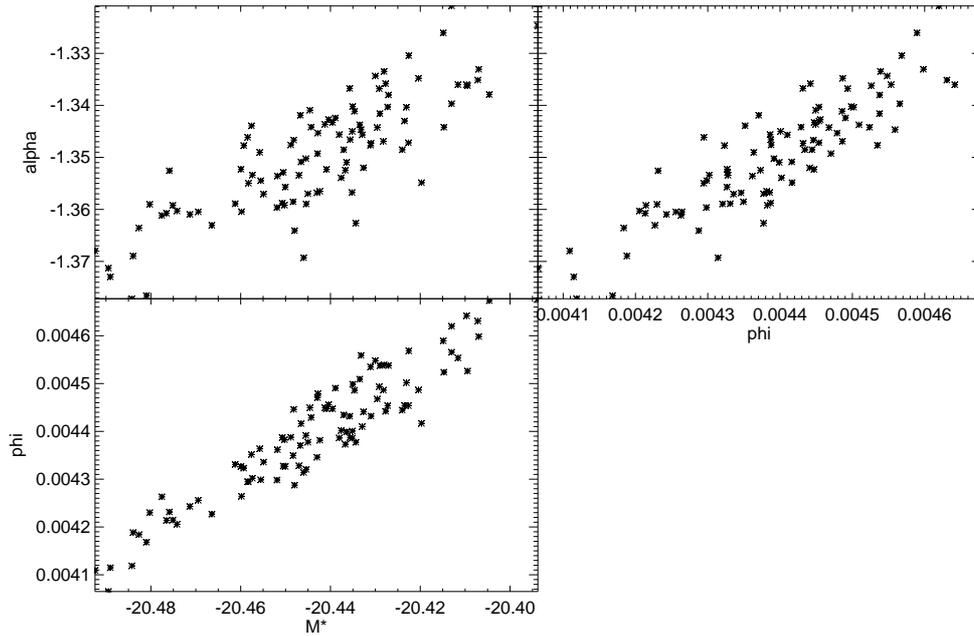


Figure A.4: Effect of ignored inclination bias on the best-fitting parameters. Shown are the best-fitting parameter as found if the likelihood normalized does NOT account for the inclination bias due to dust attenuation. All parameters are significantly biased.

### A.3.3 Effect of starting point of the fit

The influence of the starting point of the fit (or better the lack of it) is illustrated in Fig. A.5. Shown are the distributions of the best-fitting parameter combinations for 100 simulated datasets, containing  $\sim 21.7\text{k}$  dust-free galaxies each. The colour indicates their different starting points of the fit: exact input values (black), slightly high  $\alpha$  and faint  $M_*$  (blue), and slightly low  $\alpha$  and bright  $M_*$  (red). The difference in the fit results show no clear trend with starting point, and even the non-systematic differences are negligible compared with the sampling noise of the data sets (i.e. the spread of the best-fitting parameter between each data set using the same starting point). True input parameter of the simulation are indicated by the black lines.

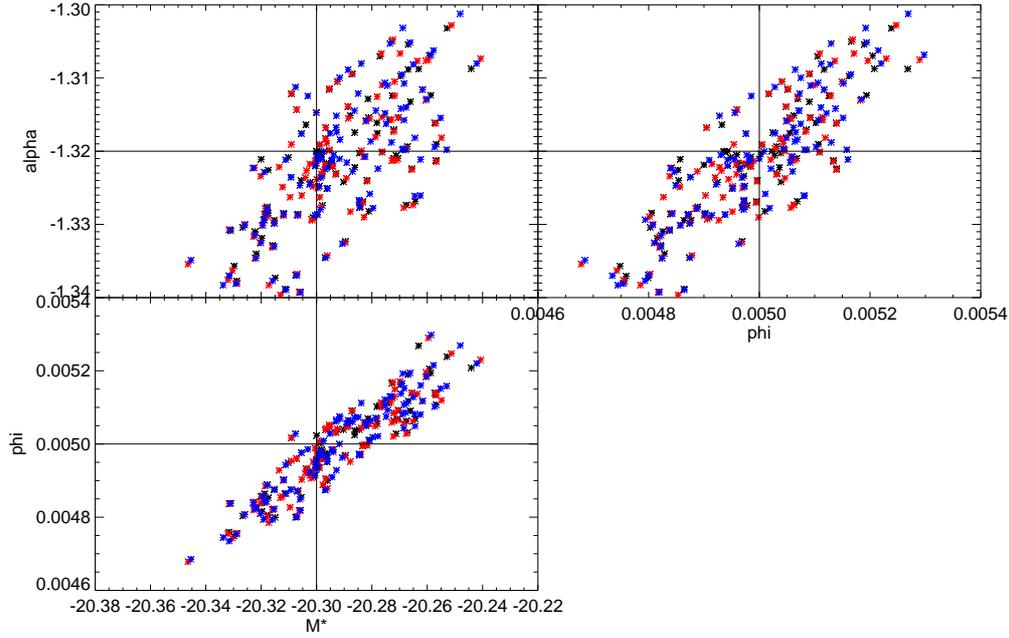


Figure A.5: Effect of the fit starting point on the accuracy of the fit parameter. Shown are the best-fitting parameter for 100 simulated data sets (input parameter values indicated by the black lines) for different starting points of the fit. No trend with starting point of the fit can be seen. Colour coding: exact input values (black), slightly high  $\alpha$  and faint  $M_*$  (blue), and slightly low  $\alpha$  and bright  $M_*$  (red).

### A.3.4 Effect of sample size

The size of the galaxy sample to be analysed has an influence on the precision of the found fit parameters. The sampling noise of the data set introduces an uncertainty to the fit parameter which scales with 1 over the square root of the number of data points.

To quantify and illustrate this effect, we generated three data sets with different sample sizes. In order to keep the same parameter value for the amplitude  $\phi^*$ , the different sample sizes were generated by changing the are of sky covered ( $\Delta\Omega$ ) by the simulated data set. Fig. A.6 shows the best-fitting parameters for three different sample sizes (black:  $\Delta\Omega = 50^\circ$ , blue:  $\Delta\Omega = 100^\circ$ , red:  $\Delta\Omega = 200^\circ$ ; input values  $\alpha = -1.32$ ,  $M_* = -20.3$ ,  $\phi^* = 0.005$  indicated by the black lines), with 200 data sets each. The  $1\sigma$  uncertainty of each parameter (marginalized

	$\Delta\Omega = 50^\circ$	$\Delta\Omega = 100^\circ$	$\Delta\Omega = 200^\circ$
$\sigma_{M^*}$	0.049293934	0.031969209	0.026446990
$\sigma_\alpha$	0.019054071	0.016828856	0.011029107
$\sigma_{\phi^*}$	0.00030609388	0.00020855353	0.00016457985

Table A.1: Luminosity function parameter uncertainties

Listed are the  $1\sigma$  errors of each luminosity function parameter, marginalized over all values of the other two. As to be expected, they roughly scale with the square root of the number of data points used for the fit (in this case with the area  $\Delta\Omega$ ). For the correlation of the parameter see Fig. A.6.

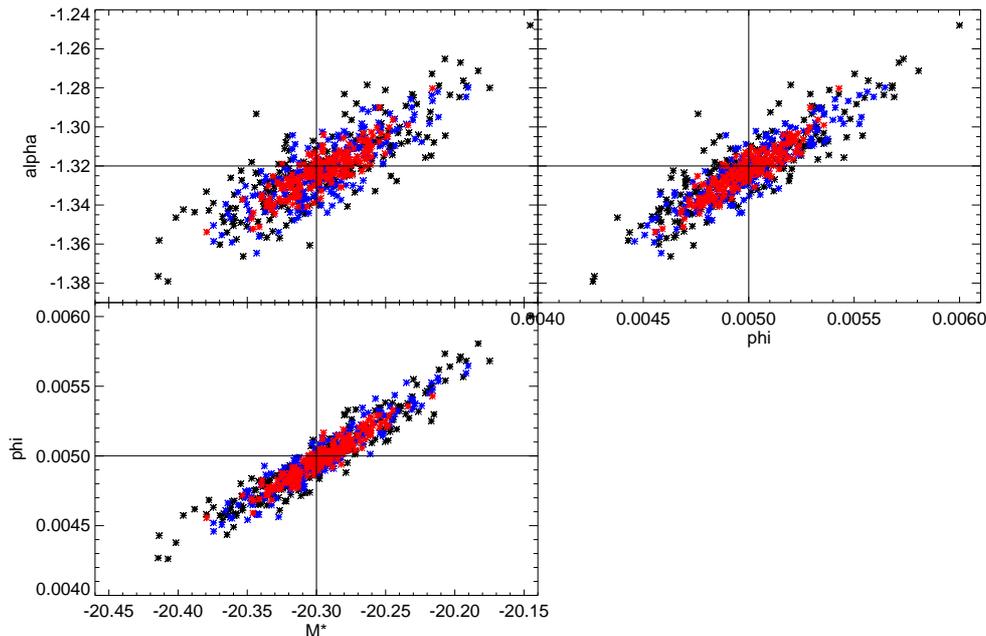


Figure A.6: Scatter in the best-fitting parameter for different sample sizes. Displayed are the best-fitting parameter combination for  $\alpha$ ,  $M_*$ , and  $\phi^*$  (given the true  $\tau$  value) for three different sampling sizes using different angular coverage:  $\Delta\Omega = 50^\circ$  (red),  $\Delta\Omega = 100^\circ$  (blue), and  $\Delta\Omega = 200^\circ$  (black). 100 simulated data sets were analysed for each angular coverages, containing each  $\sim 4.4\text{k}$ ,  $\sim 8.9\text{k}$ , and  $\sim 17.8\text{k}$  dusty galaxies for  $\Delta\Omega = 50^\circ$ ,  $\Delta\Omega = 100^\circ$ , and  $\Delta\Omega = 200^\circ$ , respectively, with input parameter:  $\alpha = -1.32$ ,  $M_* = -20.3$ ,  $\phi^* = 0.005$ ,  $\tau = 4.0$  (indicated by the black lines). As is to be expected, the sampling noise of the data increases the width of the distribution of the best-fitting parameter for smaller data sets, while the mean values of the distributions are unchanged.

over all values of the other two parameters) are listed in table A.1. As is to be expected, the parameter uncertainties roughly scale with  $1/\sqrt{N}$  over the square root of the number of data points. We will use this dependency on the sample size to derive the parameter uncertainties for the GAMA spiral galaxy sample by rescaling the parameter uncertainties found for the simulated data sets.

## A.4 fit parameter for GAMA galaxies

We presented the LF fit method in the previous sections and discussed its biases and their compensation. We now apply the method to the GAMA spiral galaxy sample (equatorial fields only). The best-fitting parameters are  $\alpha = -0.867 \pm 0.015$ ,  $M_* = -21.410 \pm 0.028$ ,  $\phi^* = 0.00236 \pm 0.00018$  (assuming a  $\tau = 4.0$ ; parameter errors are the parameter errors of the  $\Delta\Omega = 100^\circ$  data set rescaled for  $N=11173$ ).

The spread in techniques, sample selection and thus LF parameter values in literature is remarkable. Many authors have reported luminosity functions for galaxies in the SDSS r-band, using different techniques to derive the model parameters  $\alpha$ ,  $M_*$ , and  $\phi^*$ . For instance, Blanton et al. (2001) derive the LF parameters  $\alpha = -1.20 \pm 0.03$ ,  $M_* = -20.83 \pm 0.03$ , and  $\phi^* = 0.0146 \pm 0.0012$  for SDSS galaxies (without making a distinction between spiral/elliptical or blue/red). Shao et al. (2007) derive their LF parameters ( $\alpha$  and  $M_*$  only) for SDSS spiral galaxies as a

function of inclination (assuming  $b/a = \cos(i)$ ) and use a model to correct for the inclination-dependent component of the dust attenuation (i.e. they do not correct for the face-on attenuation component). They report for the face-on spiral sample the parameter  $\alpha = -1.25 \pm 0.01$ ,  $M_* = -20.54 \pm 0.01$ .

A more direct comparison allow the LF estimates reported for GAMA galaxy samples. Loveday et al. (2012) report  $\alpha = -1.39 \pm 0.01$ ,  $M_* = -20.23 - 5 \log h \pm 0.06$ , and  $\phi^* = 0.0058 \pm 0.0004 \cdot h^3$  for the parameters of their single Schechter function fitted to their blue GAMA galaxy sample. Guanawardhana et al. (2014, in prep.) report  $\alpha = -1.32 \pm 0.05$ ,  $M_* = -20.80 \pm 0.07$ , and  $\phi^* = 0.0034 \pm 0.0005$  for the parameters of their local ( $z < 0.1$ ), blue sample. Driver et al. (2012) report for their non-elliptical galaxy sample the parameters  $\alpha = -1.20 \pm 0.01$ ,  $M_* = -20.68 - 5 \log h \pm 0.04$ , and  $\phi^* = 0.0096 \pm 0.0001 \cdot h^3$ , using the same radiation transfer dust model as we do to correct for dust attenuation of the spiral population (assuming  $\tau = 3.8$ ).

Comparing the literature values to our results, one notices that the  $M_*$  value is generally larger in literature and the  $\alpha$  value is smaller. This translates to a fainter distribution, which has more emphasis on the faint-end galaxies than the distribution we obtain for our GAMA spiral galaxy sample<sup>9</sup>. On the one hand, this is due to the fact that most of the studies do not account for dust attenuation, leading to an overall fainter magnitude and a steeper (i.e. smaller) faint-end slope  $\alpha$ . However, the main reason for the differences between our findings and those in literature is the spiral selection used in our work. The spiral selection from Grootes et al. (2013) was designed with the emphasis on generating a pure spiral sample at the cost of completeness. Especially the criterion effective radius preferentially selects large galaxies ( $\sim$  Milky Way size and above). This leads to a selection against fainter galaxies due to the size-luminosity relation.

---

<sup>9</sup>The differences in amplitude  $\phi^*$  are to a large extent caused by the difference in the other parameter. The other effect is the fraction of all galaxies included in the considered sample (i.e. relative numbers of galaxies selected by the selection criterion used in the study).



# Appendix B

## Theoretical framework for the treatment of NUV noise

While the r-band flux measurements are very precise and have a high signal-to-noise ratio, the NUV has a much higher noise contribution. Especially when we include the NUV measurement obtained by the COG method for all GAMA galaxies (see Part II, especially Sects. 3.5 and 3.6.3), we can have S/N values of 1 or even lower.

The bias on the estimate of dust model parameter using the attenuation inclination relation caused by asymmetric noise in the NUV-r colour (i.e. symmetric noise in NUV flux) was shown in Sect. 7.5.3 and Fig. 7.11. We discussed that without properly accounting for the NUV noise in our analysis, we rely on using only data with high S/N values, where the noise distribution is narrow enough to be neglected. This lead to a significant reduction of our data sample, which is highly regrettable. In this section, we therefore present the theory for accounting for the NUV noise in the calculation of the intrinsic colour probability. More precisely, we need to convolve the intrinsic colour probability distribution with the NUV noise probability distribution.

A detailed description of the NUV noise and a motivation and derivation of the Gaussian noise model were presented in Sects. 2.2.1 and 3.1.4. For the noise convolution, we adopt the standard deviation of the Gaussian noise model given in Eq. 3.5:

$$\sigma_{err} = \sqrt{(s_f \cdot 0.01)^2 + \frac{(s_f + 1.5 \cdot bg) \cdot cuj}{t_e}} \quad (\text{B.1})$$

where the source flux  $f$  and the background level are in total counts.

Thus, we define a likelihood for the intrinsic colour distribution in Eq. 7.12 by defining the probability of the noise-free NUV source flux  $s_{nn}$  given the observed, noisy NUV flux  $s_{obs}$

$$P(s_{nn}|s_{obs}) = \frac{1}{\sqrt{2\pi}\sigma_{err}^2} e^{-\frac{1}{2}\left(\frac{s_{nn}-s_{obs}}{\sigma_{err}}\right)^2} \quad (\text{B.2})$$

It should be noted that a negative noise-free source flux  $s_{nn}$  is unphysical (e.g. the A- or k-corrections, see Sect. 6.2.2.2, are multiplicative changes in flux, thus

unable to cause negative noise-free flux). We assume a Gaussian distribution in intrinsic colour. Negative UV flux would give us a colour of infinity, resulting in an intrinsic colour probability of 0. Therefore, we do not need to correct for this unphysical flux value explicitly, as the possibility of negative noise-free source flux is automatically excluded when the probability of the intrinsic colour  $P(c)$  is calculated.

Combining the likelihood as defined in Eq. 7.12 with the likelihood for the NUV noise defined in Eq. B.2, we obtain the likelihood of the intrinsic colour model parameters for a galaxy, given its observed r-band flux, its observed NUV flux, its distance and its stellar mass.

$$P(\mu, \sigma) = \int_0^{90} P(c(M_{r,intr}, M_{NUV,intr}, M_s)) \cdot P(i) \cdot P(s_{nn}|s_{obs}) di \quad (\text{B.3})$$

Since we do not know  $s_{nn}$ , we need to marginalize over it.

$$P(\mu, \sigma) = \int_0^\infty \int_0^{90} P(c) \cdot P(i) \cdot P(s_{nn}|s_{obs}) di ds_{nn} \quad (\text{B.4})$$

It should be noted that the colour  $c$  is the noise-free, A- & k-corrected intrinsic NUV-r colour, which is a function of the noise-free NUV magnitude corresponding to the NUV flux  $s_{nn}$ .

For NUV non-detections, we marginalize Eq. B.4 over the possible noisy NUV flux from  $-bg$  to the detection limit  $s_{detlim}$ .

$$P(\mu, \sigma) = \int_{-bg}^{s_{detlim}} \int_0^\infty \int_0^{90} P(c) \cdot P(i) \cdot P(s_{nn}|s_{obs}) di ds_{nn} ds_{obs} \quad (\text{B.5})$$

### Toy data

For simplistic toy data, we want to compare the intrinsic NUV distribution and the noisy colour distribution of the toy data sample (i.e. subject to sampling noise) with our according model probabilities. Therefore, we simulate 10000 galaxies with a Gaussian intrinsic NUV distribution with mean 21mag and standard deviation of 0.1mag. We deliberately choose a rather faint magnitude and a narrow intrinsic magnitude distribution to demonstrate both the effect of the noise on the distribution as well as the ability of the noise convolution to account for it. The intrinsic magnitudes were converted to counts (using the input values of the exposure time equals 1500s and the background  $bg$  equals 2000 counts in total for all galaxies). Then, a noise term was added to the counts  $f$ , sampled from a Gaussian with mean 0 and standard deviation of  $\sigma = \sqrt{(0.01 * f)^2 + (f + 1.5 * bg)}$  (i.e. slightly different standard deviation for each galaxy, depending on the intrinsic NUV flux).

Fig. B.1 shows a histogram of the intrinsic NUV distribution (blue) overplotted with the Gaussian model (black). The distribution of the noisy NUV magnitude is shown in red and is very well modelled by the noise convolved intrinsic colour distribution (black). The prominent faint tail and the broadening of the noisy NUV magnitude distribution can be well seen in both the data and the noise-convolved model.

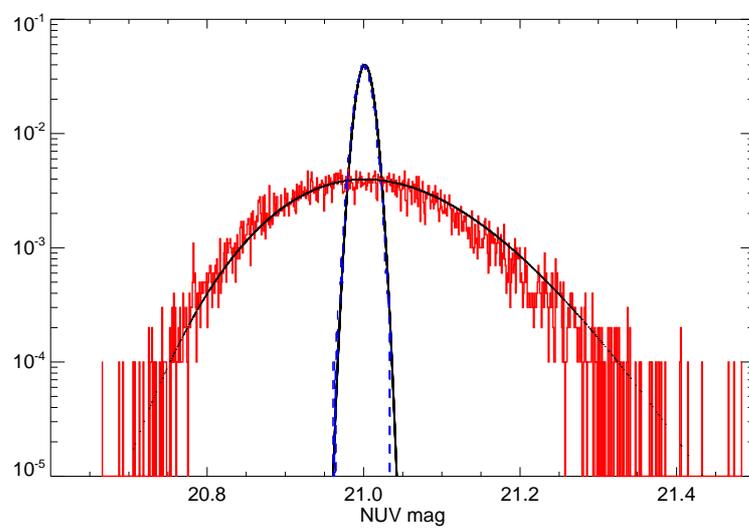


Figure B.1: Noise convolution of the colour model for toy data

The histogram of the intrinsic NUV distribution (blue; Gaussian with mean 21mag and standard deviation of 0.1mag) is very narrow compared to the noisy NUV distribution (red). Noteworthy is that the noise causes the formerly symmetric distribution to have a dominant faint tail. Both, the model for the intrinsic NUV magnitude as well as the noise convolved model (both overplotted as black lines) agree well with the data.



# Appendix C

## Accounting for r-band detection limit

We have demonstrated in Sect. 7.5 how severely the observed colour distribution differs from the intrinsic one. Even in the absence of (NUV) noise, the intrinsic scatter of the colour distribution in combination with a detection limit in one band is enough to cause biases in the derived colour parameters, and thus in the derived dust model parameter.

In Chap. 4, we have already mentioned that the probability of the intrinsic colour is only one part of the probability of seeing a galaxy with its observed colour. We also need to consider the probability of having a galaxy with the apparent r-band magnitude at the given distance<sup>1</sup>.

Eq. A.25 defined the probability of seeing a galaxy with dust-attenuated absolute r-band magnitude  $M_d$  at a comoving distance  $D_C$

$$P(M_d, D_C) \propto \int \int D_C^2 \cdot P_{SF}(M_d - A(\tau, i) - A_{corr} - k_{corr}|\phi^*, \alpha, M^*) \cdot \underbrace{\eta(M_d + 5[\log_{10}(D_L) - 1])}_{m_d} \cdot P_A(A(\tau, i)) d\tau di \quad (\text{C.1})$$

where  $P_{SF}$  is the normalized Schechter function as defined in Eq. A.4,  $\eta$  is the Heavyside function of the r-band detection limit (Eq. A.16), and  $P_A(A(\tau, i)) = P_\tau(\tau) \cdot P(i)$  is the probability distribution of the dust attenuation as function of the probability distribution of the intrinsic inclination  $i$  (i.e. flat distribution in  $(\cos(i))$ ) and the probability distribution of dust content  $\tau$  (dust content model dependent). Since we analyse the intrinsic colour and the dust features as function of stellar mass  $M_s$ , this equation must be adapted such that it represents the probability of a galaxy with observed  $M_d$  and  $D_C$  given its  $M_s$ .

Combining both probabilities as described in Chap. 4 will account for the colour bias caused by the r-band detection limit.

---

<sup>1</sup>We will not consider measurement noise for the r-band magnitude or the redshift measurement.



# Appendix D

## GALEX-GAMA DMU product details

In this appendix, we present an overview over the content of the GALEX-GAMA DMU (Data Management Unit). For all catalogues of the DMU, we list the column names and a brief description of the columns. Even though a catalogue might exist in several versions (for individual GAMA fields or UV survey depths), it is listed here once. For detail about the matching procedure to the GAMA catalogues see Sect. 3. The default value for non-measurements is -99.0 in all catalogues.

### D.1 Blind catalogue

Column name	Unit	Description
GGOID	-	GALEX Global Object ID
NAME	-	Object IAU format name
RA	deg	Right Ascension (J2000)
DEC	deg	Declination (J2000)
GLON	deg	Galactic longitude
GLAT	deg	Galactic latitude
POSERR	arcsec	Position error
TILENUM	-	GALEX tile number
TILE	-	GALEX tile name
FOV_RADIUS	deg	Distance of object from centre of FoV
HP_PIXEL	-	HEALPix pixel on sky coverage maps on GALEX-GAMA web site (nside = 16384; nested)
MANFLAG	-	manual flag for extreme, extended artifacts
EXPTIME_NUV	s	NUV exposure time of tile
EFF_EXPTIME_NUV	s	NUV effective exposure time (relative response)
EXPTIME_FUV	s	FUV exposure time of tile
EFF_EXPTIME_FUV	s	FUV effective exposure time (relative response)
FLUX_NUV	$\mu\text{Jy}$	NUV calibrated flux
FLUXERR_NUV	$\mu\text{Jy}$	NUV flux error

Table D.1: Blind catalogue: column names & description

Column name	Unit	Description
MAG_NUV	mag	NUV calibrated AB Kron magnitude
MAGERR_NUV	mag	NUV calibrated AB Kron magnitude error
S2N_NUV	-	NUV signal-to-noise from flux value
ARTIFACT_NUV	-	NUV artifact flag
BKGRND_FLUX_NUV	$\mu\text{Jy}/\text{arcsec}^2$	NUV sky background surface brightness
BKGRND_MAG_NUV	$\text{mag}/\text{arcsec}^2$	NUV sky background surface brightness
FLUX_FUV	$\mu\text{Jy}$	FUV calibrated flux using NUV aperture
FLUXERR_FUV	$\mu\text{Jy}$	FUV flux error using NUV aperture
MAG_FUV	mag	FUV calibrated AB Kron magnitude using NUV aperture
MAGERR_FUV	mag	FUV calibrated AB Kron magnitude error using NUV aperture
S2N_FUV	-	FUV signal-to-noise from flux value (NB: can become negative for negative flux)
ARTIFACT_FUV	-	FUV artifact flag
BKGRND_FLUX_FUV	$\mu\text{Jy}/\text{arcsec}^2$	FUV sky background surface brightness
BKGRND_MAG_FUV	$\text{mag}/\text{arcsec}^2$	FUV sky background surface brightness
FLUX_D3p0_NUV	$\mu\text{Jy}$	NUV flux in 3.0" (2 pixel) diameter aperture
FLUX_D4p5_NUV	$\mu\text{Jy}$	NUV flux in 4.5" (3 pixel) diameter aperture
FLUX_D7p5_NUV	$\mu\text{Jy}$	NUV flux in 7.5" (5 pixel) diameter aperture
FLUX_D12p0_NUV	$\mu\text{Jy}$	NUV flux in 12.0" (8 pixel) diameter aperture
FLUX_D18p0_NUV	$\mu\text{Jy}$	NUV flux in 18.0" (12 pixel) diameter aperture
FLUX_D25p5_NUV	$\mu\text{Jy}$	NUV flux in 25.5" (17 pixel) diameter aperture
FLUX_D34p5_NUV	$\mu\text{Jy}$	NUV flux in 34.5" (23 pixel) diameter aperture
FLUXERR_D3p0_NUV	$\mu\text{Jy}$	NUV flux error in 3.0" (2 pixel) diameter aperture
FLUXERR_D4p5_NUV	$\mu\text{Jy}$	NUV flux error in 4.5" (3 pixel) diameter aperture
FLUXERR_D7p5_NUV	$\mu\text{Jy}$	NUV flux error in 7.5" (5 pixel) diameter aperture
FLUXERR_D12p0_NUV	$\mu\text{Jy}$	NUV flux error in 12.0" (8 pixel) diameter aperture
FLUXERR_D18p0_NUV	$\mu\text{Jy}$	NUV flux error in 18.0" (12 pixel) diameter aperture
FLUXERR_D25p5_NUV	$\mu\text{Jy}$	NUV flux error in 25.5" (17 pixel) diameter aperture
FLUXERR_D34p5_NUV	$\mu\text{Jy}$	NUV flux error in 34.5" (23 pixel) diameter aperture

Table D.2: Tab. D.1 continued

Column name	Unit	Description
MAG_D3p0_NUV	mag	NUV AB magnitude in 3.0" (2 pixel) diameter aperture
MAG_D4p5_NUV	mag	NUV AB magnitude in 4.5" (3 pixel) diameter aperture
MAG_D7p5_NUV	mag	NUV AB magnitude in 7.5" (5 pixel) diameter aperture
MAG_D12p0_NUV	mag	NUV AB magnitude in 12.0" (8 pixel) diameter aperture
MAG_D18p0_NUV	mag	NUV AB magnitude in 18.0" (12 pixel) diameter aperture
MAG_D25p5_NUV	mag	NUV AB magnitude in 25.5" (17 pixel) diameter aperture
MAG_D34p5_NUV	mag	NUV AB magnitude in 34.5" (23 pixel) diameter aperture
MAGERR_D3p0_NUV	mag	NUV AB magnitude error in 3.0" (2 pixel) diameter aperture
MAGERR_D4p5_NUV	mag	NUV AB magnitude error in 4.5" (3 pixel) diameter aperture
MAGERR_D7p5_NUV	mag	NUV AB magnitude error in 7.5" (5 pixel) diameter aperture
MAGERR_D12p0_NUV	mag	NUV AB magnitude error in 12.0" (8 pixel) diameter aperture
MAGERR_D18p0_NUV	mag	NUV AB magnitude error in 18.0" (12 pixel) diameter aperture
MAGERR_D25p5_NUV	mag	NUV AB magnitude error in 25.5" (17 pixel) diameter aperture
MAGERR_D34p5_NUV	mag	NUV AB magnitude error in 34.5" (23 pixel) diameter aperture
FLUX20_RADIUS_NUV	arcsec	NUV 20 percent fraction-of-light radius
FLUX50_RADIUS_NUV	arcsec	NUV 50 percent fraction-of-light radius
FLUX80_RADIUS_NUV	arcsec	NUV 80 percent fraction-of-light radius
FLUX90_RADIUS_NUV	arcsec	NUV 90 percent fraction-of-light radius
FLUX95_RADIUS_NUV	arcsec	NUV 95 percent fraction-of-light radius
FLUX_MAX_NUV	$\mu$ Jy	NUV peak flux above background
X_IMAGE	pixel	Object pixel position on intensity maps
Y_IMAGE	pixel	Object pixel position on intensity maps
XPEAK_IMAGE	pixel	Pixel position of the brightest NUV pixel on intensity maps
YPEAK_IMAGE	pixel	Pixel position of the brightest NUV pixel on intensity maps
SEMIMAJOR	arcsec	Semimajor axis
SEMIMINOR	arcsec	Semiminor axis

Table D.3: Tab. D.1 continued

Column name	Unit	Description
SEMIMAJORERR	arcsec	Semimajor axis error
SEMIMINORERR	arcsec	Semiminor axis error
POSANG	deg	Position angle (east of north)
POSANGERR	deg	Position angle error
FWHM_NUV	arcsec	NUV FWHM assuming a gaussian core
FLAGS_NUV	-	NUV SExtractor extraction flags
CLASS_STAR	-	NUV SExtractor star/galaxy classifier output
FLUX_D3p0_FUV	$\mu$ Jy	FUV flux in 3.0" (2 pixel) diameter aperture
FLUX_D4p5_FUV	$\mu$ Jy	FUV flux in 4.5" (3 pixel) diameter aperture
FLUX_D7p5_FUV	$\mu$ Jy	FUV flux in 7.5" (5 pixel) diameter aperture
FLUX_D12p0_FUV	$\mu$ Jy	FUV flux in 12.0" (8 pixel) diameter aperture
FLUX_D18p0_FUV	$\mu$ Jy	FUV flux in 18.0" (12 pixel) diameter aperture
FLUX_D25p5_FUV	$\mu$ Jy	FUV flux in 25.5" (17 pixel) diameter aperture
FLUX_D34p5_FUV	$\mu$ Jy	FUV flux in 34.5" (23 pixel) diameter aperture
FLUXERR_D3p0_FUV	$\mu$ Jy	FUV flux error in 3.0" (2 pixel) diameter aperture
FLUXERR_D4p5_FUV	$\mu$ Jy	FUV flux error in 4.5" (3 pixel) diameter aperture
FLUXERR_D7p5_FUV	$\mu$ Jy	FUV flux error in 7.5" (5 pixel) diameter aperture
FLUXERR_D12p0_FUV	$\mu$ Jy	FUV flux error in 12.0" (8 pixel) diameter aperture
FLUXERR_D18p0_FUV	$\mu$ Jy	FUV flux error in 18.0" (12 pixel) diameter aperture
FLUXERR_D25p5_FUV	$\mu$ Jy	FUV flux error in 25.5" (17 pixel) diameter aperture
FLUXERR_D34p5_FUV	$\mu$ Jy	FUV flux error in 34.5" (23 pixel) diameter aperture
MAG_D3p0_FUV	mag	FUV AB magnitude in 3.0" (2 pixel) diameter aperture
MAG_D4p5_FUV	mag	FUV AB magnitude in 4.5" (3 pixel) diameter aperture
MAG_D7p5_FUV	mag	FUV AB magnitude in 7.5" (5 pixel) diameter aperture
MAG_D12p0_FUV	mag	FUV AB magnitude in 12.0" (8 pixel) diameter aperture
MAG_D18p0_FUV	mag	FUV AB magnitude in 18.0" (12 pixel) diameter aperture
MAG_D25p5_FUV	mag	FUV AB magnitude in 25.5" (17 pixel) diameter aperture
MAG_D34p5_FUV	mag	FUV AB magnitude in 34.5" (23 pixel) diameter aperture

Table D.4: Tab. D.1 continued

Column name	Unit	Description
MAGERR_D3p0_FUV	mag	FUV AB magnitude error in 3.0" (2 pixel) diameter aperture
MAGERR_D4p5_FUV	mag	FUV AB magnitude error in 4.5" (3 pixel) diameter aperture
MAGERR_D7p5_FUV	mag	FUV AB magnitude error in 7.5" (5 pixel) diameter aperture
MAGERR_D12p0_FUV	mag	FUV AB magnitude error in 12.0" (8 pixel) diameter aperture
MAGERR_D18p0_FUV	mag	FUV AB magnitude error in 18.0" (12 pixel) diameter aperture
MAGERR_D25p5_FUV	mag	FUV AB magnitude error in 25.5" (17 pixel) diameter aperture
MAGERR_D34p5_FUV	mag	FUV AB magnitude error in 34.5" (23 pixel) diameter aperture
FLUX20_RADIUS_FUV	arcsec	FUV 20 percent fraction-of-light radius
FLUX50_RADIUS_FUV	arcsec	FUV 50 percent fraction-of-light radius
FLUX80_RADIUS_FUV	arcsec	FUV 80 percent fraction-of-light radius
FLUX90_RADIUS_FUV	arcsec	FUV 90 percent fraction-of-light radius
FLUX95_RADIUS_FUV	arcsec	FUV 95 percent fraction-of-light radius
FWHM_FUV	arcsec	FUV FWHM assuming a gaussian core
FLAGS_FUV	-	FUV SExtractor extraction flags

Table D.5: Tab. D.1 continued

## D.2 Simple match catalogue

Column name	Unit	Description
CATAID	-	Unique ID of GAMA object
OBJID	-	SDSS objid of GAMA object
GGOID	-	Global Object ID of nearest neighbour GALEX object
FUVMATCH	-	Flag indicating whether matched GALEX source has FUV flux with S/N $\geq 2.5$ (-1=no FUV data)
NN_DIST	arcsec	Distance between GAMA object and nearest neighbour GALEX object
NMATCH4	-	Number of GALEX objects found within 4 arcsec of GAMA object
MANY2ONE	-	Number of GAMA objects the GALEX object is matched to
GNEIGHBOR_DIST	-	String list of distances to all GALEX sources within 20 arcsec
GNEIGHBOR_NAME	-	String list of names of all GALEX sources within 20 arcsec

Table D.6: Simple matching catalogue: column names & description

## D.3 Advanced match catalogue

Column name	Unit	Description
CATAID	-	Unique ID of GAMA object
OBJID	-	SDSS objid
NMATCHUV	-	Number of GALEX objects involved in the matching
NMATCHOPT	-	Number of GAMA objects involved in the matching
UVIDXLIST	-	ID list of GALEX object used in the flux redistribution calculation using NAME from GalexPhot; string always ends with a comma
OPTIDXLIST	-	ID list of GAMA objects used in the flux redistribution calculation using SDSS OBJID; string always ends with a comma
NUVFLAG	-	Combination of the NUV artifact flags of all GALEX objects involved in the matching
FUVFLAG	-	Combination of the FUV artifact flags of all GALEX objects involved in the matching
FLUXTOT_NUV	$\mu\text{Jy}$	Total NUV flux of all GALEX objects associated with GAMA object (unweighted sum of the flux)
FLUXTOTERR_NUV	$\mu\text{Jy}$	Total NUV flux error
MAGTOT_NUV	mag	Total NUV AB magnitude of all GALEX objects associated with GAMA object (unweighted sum of the flux)
MAGTOTERR_NUV	mag	Total NUV AB magnitude error
FLUXSPLIT_NUV	$\mu\text{Jy}$	Redistributed NUV flux of all GALEX objects associated with GAMA object
FLUXSPLITERR_NUV	$\mu\text{Jy}$	Redistributed NUV flux error
MAGSPLIT_NUV	mag	Redistributed NUV AB magnitude of all GALEX objects associated with GAMA object
MAGSPLITERR_NUV	mag	Redistributed NUV AB magnitude error
FLUXTOT_FUV	$\mu\text{Jy}$	Total FUV flux of all GALEX objects associated with GAMA object (unweighted sum of the flux)
FLUXTOTERR_FUV	$\mu\text{Jy}$	Total FUV flux error
MAGTOT_FUV	mag	Total FUV AB magnitude of all GALEX objects associated with GAMA object (unweighted sum of the flux)
MAGTOTERR_FUV	mag	Total FUV AB magnitude error
FLUXSPLIT_FUV	$\mu\text{Jy}$	Redistributed FUV flux of all GALEX objects associated with GAMA object
FLUXSPLITERR_FUV	$\mu\text{Jy}$	Redistributed FUV flux error
MAGSPLIT_FUV	mag	Redistributed FUV AB magnitude of all GALEX objects associated with GAMA object
MAGSPLITERR_FUV	mag	Redistributed FUV AB magnitude error

Table D.7: Advanced matching catalogue: column names &amp; description

## D.4 Curve-of-Growth catalogue

Column name	Unit	Description
CATAID_I	-	Unique GAMA I ID of GAMA object
CATAID	-	Unique GAMA II ID of GAMA object
CPS_NUV	counts/s	NUV counts per second
CPSERR_NUV	counts/s	NUV counts per second error
FLUX_NUV	$\mu\text{Jy}$	NUV flux
FLUXERR_NUV	$\mu\text{Jy}$	NUV flux error
MAG_NUV	mag	NUV magnitude
MAGERR_NUV	mag	NUV magnitude error
RE_CRIT_NUV	arcsec	NUV integration radius (multiple of 3 arcsec)
HP_RAD_NUV	arcsec	NUV radius enclosing 50% of the total flux
CONFSOURCE_NUV	-	NUV flag indicating whether masking occurred
CPS_FUV	counts/s	FUV counts per second
CPSERR_FUV	counts/s	FUV counts per second error
FLUX_FUV	$\mu\text{Jy}$	FUV flux
FLUXERR_FUV	$\mu\text{Jy}$	FUV flux error
MAG_FUV	mag	FUV magnitude
MAGERR_FUV	mag	FUV magnitude error
RE_CRIT_FUV	arcsec	FUV integration radius (multiple of 3 arcsec)
HP_RAD_FUV	arcsec	FUV radius enclosing 50% of the total flux
CONFSOURCE_FUV	-	FUV flag indicating whether masking occurred

Table D.8: COG catalogue: column names & description

## D.5 Best-mag catalogue

Column name	Unit	Description
CATAID	-	Unique ID of GAMA object
OBJID	-	SDSS objid
RA	deg	Right Ascension of GAMA object (J2000)
DEC	deg	Declination of GAMA object (J2000)
EFF_EXPTIME_NUV	s	NUV effective exposure time (relative response)
EFF_EXPTIME_FUV	s	FUV effective exposure time (relative response)
BEST_FLUX_NUV	$\mu\text{Jy}$	Best estimate of the GAMA object's NUV flux
BEST_FLUXERR_NUV	$\mu\text{Jy}$	Best estimate NUV flux error
BEST_MAG_NUV	mag	Best estimate of the GAMA object's NUV magnitude
BEST_MAGERR_NUV	mag	Best estimate NUV magnitude error
BEST_FLUX_FUV	$\mu\text{Jy}$	Best estimate of the GAMA object's FUV flux
BEST_FLUXERR_FUV	$\mu\text{Jy}$	Best estimate FUV flux error
BEST_MAG_FUV	mag	Best estimate of the GAMA object's FUV magnitude
BEST_MAGERR_FUV	mag	Best estimate FUV magnitude error
BEST_METHOD	-	Method used for the best flux estimates (Advanced-Match or CoG)
NMATCHUV	-	Number of GALEX objects involved in advanced matching procedure
NMATCHOPT	-	Number of GAMA objects involved in advanced matching procedure
NUVFLAG	-	Combination of the NUV artifact flags of all GALEX objects involved in advanced matching procedure
FUVFLAG	-	Combination of the FUV artifact flags of all GALEX objects involved in advanced matching procedure
TOT_FLUX_NUV	$\mu\text{Jy}$	Total NUV flux of all GALEX objects associated with GAMA object (unweighted sum of the flux)
TOT_FLUXERR_NUV	$\mu\text{Jy}$	Error on the total NUV flux of all GALEX objects associated with GAMA object
TOT_MAG_NUV	mag	Total NUV AB magnitude of all GALEX objects associated with GAMA object (unweighted sum of the flux)
TOT_MAGERR_NUV	mag	Error on total NUV AB magnitude of all GALEX objects associated with GAMA object
TOT_FLUX_FUV	$\mu\text{Jy}$	Total FUV flux of all GALEX objects associated with GAMA object (unweighted sum of the flux)

Table D.9: Best-mag catalogue: column names &amp; description

Column name	Unit	Description
TOT_FLUXERR_FUV	$\mu\text{Jy}$	Error on total FUV flux of all GALEX objects associated with GAMA object
TOT_MAG_FUV	mag	Total FUV AB magnitude of all GALEX objects associated with GAMA object (un-weighted sum of the flux)
TOT_MAGERR_FUV	mag	Error on total FUV AB magnitude of all GALEX objects associated with GAMA object
SPLIT_FLUX_NUV	$\mu\text{Jy}$	Redistributed NUV flux of all GALEX objects associated with GAMA object
SPLIT_FLUXERR_NUV	$\mu\text{Jy}$	Error on redistributed NUV flux of all GALEX objects associated with GAMA object
SPLIT_MAG_NUV	mag	Redistributed NUV AB magnitude of all GALEX objects associated with GAMA object
SPLIT_MAGERR_NUV	mag	Error on redistributed NUV AB magnitude of all GALEX objects associated with GAMA object
SPLIT_FLUX_FUV	$\mu\text{Jy}$	Redistributed FUV flux of all GALEX objects associated with GAMA object
SPLIT_FLUXERR_FUV	$\mu\text{Jy}$	Error on redistributed FUV flux of all GALEX objects associated with GAMA object
SPLIT_MAG_FUV	mag	Redistributed FUV AB magnitude of all GALEX objects associated with GAMA object
SPLIT_MAGERR_FUV	mag	Error on redistributed FUV AB magnitude of all GALEX objects associated with GAMA object
COG_FLUX_NUV	$\mu\text{Jy}$	CoG NUV flux at the position of the GAMA object
COG_FLUXERR_NUV	$\mu\text{Jy}$	CoG NUV flux error
COG_MAG_NUV	mag	CoG NUV magnitude
COG_MAGERR_NUV	mag	CoG NUV magnitude error
COG_FLUX_FUV	$\mu\text{Jy}$	CoG FUV flux at the position of the GAMA object
COG_FLUXERR_FUV	$\mu\text{Jy}$	CoG FUV flux error
COG_MAG_FUV	mag	CoG FUV magnitude
COG_MAGERR_FUV	mag	CoG FUV magnitude error
COG_RE_CRIT_NUV	arcsec	CoG NUV integration radius (multiple of 3 arcsec)
COG_HP_RAD_NUV	arcsec	CoG NUV radius enclosing 50% of the total flux

Table D.10: Tab. D.9 continued

Column name	Unit	Description
COG_RE_CRIT_FUV	arcsec	CoG FUV integration radius (multiple of 3 arcsec)
COG_HP_RAD_FUV	arcsec	CoG FUV radius enclosing 50% of the total flux
COG_CONFSOURCE_NUV	-	CoG NUV flag indicating whether masking occurred
COG_CONFSOURCE_FUV	-	CoG FUV flag indicating whether masking occurred
NN_GGOID	-	Global Object ID of nearest neighbour GALEX object
NN_RA_GALEX	deg	Right Ascension of nearest neighbour GALEX object (J2000)
NN_DEC_GALEX	deg	Declination of nearest neighbour GALEX object (J2000)
NN_DIST	arcsec	Distance between GAMA object and nearest neighbour GALEX object
NN_NMATCH4	-	Number of GALEX objects found within 4 arcsec of GAMA object
NN_MANY2ONE	-	Number of GAMA objects the nearest neighbour GALEX object is matched to
NN_FLUX_NUV	$\mu\text{Jy}$	NUV flux of nearest neighbour GALEX object
NN_FLUXERR_NUV	$\mu\text{Jy}$	NUV flux error of nearest neighbour GALEX object
NN_MAG_NUV	mag	NUV AB Kron magnitude of nearest neighbour GALEX object
NN_MAGERR_NUV	mag	NUV AB Kron magnitude error of nearest neighbour GALEX object
NN_FLUX_FUV	$\mu\text{Jy}$	FUV flux of nearest neighbour GALEX object using NUV aperture
NN_FLUXERR_FUV	$\mu\text{Jy}$	FUV flux error of nearest neighbour GALEX object using NUV aperture
NN_MAG_FUV	mag	FUV AB Kron magnitude of nearest neighbour GALEX object using NUV aperture
NN_MAGERR_FUV	mag	FUV AB Kron magnitude error of nearest neighbour GALEX object using NUV aperture
NN_ARTIFACT_NUV	-	NUV artifact flag of nearest neighbour GALEX object
NN_ARTIFACT_FUV	-	FUV artifact flag of nearest neighbour GALEX object
NN_SFLAGS_NUV	-	NUV Sextraxtor extraction flags of nearest neighbour GALEX object
NN_SFLAGS_FUV	-	FUV Sextraxtor extraction flags of nearest neighbour GALEX object

Table D.11: Tab. D.9 continued

Column name	Unit	Description
NN_FLUX50_RADIUS_NUV	arcsec	NUV 50% fraction-of-light radius of nearest neighbour GALEX object
NN_FLUX50_RADIUS_FUV	arcsec	FUV 50% fraction-of-light radius of nearest neighbour GALEX object
NN_SEMIMAJOR_NUV	arcsec	NUV semimajor axis of nearest neighbour GALEX object
NN_SEMIMINOR_NUV	arcsec	NUV semiminor axis of nearest neighbour GALEX object
NN_SEMIMAJORERR_NUV	arcsec	NUV semimajor axis error of nearest neighbour GALEX object
NN_SEMIMINORERR_NUV	arcsec	NUV semiminor axis error of nearest neighbour GALEX object
NN_POSANG_NUV	deg	NUV position angle (east of north) of nearest neighbour GALEX object
NN_POSANGERR_NUV	deg	NUV position angle error of nearest neighbour GALEX object

Table D.12: Tab. D.9 continued

## D.6 ObsInfo catalogue

Column name	Unit	Description
CATAID	-	Unique ID of GAMA object
OBJID	-	SDSS objid
EFF_EXPTIME_NUV_GSC	s	GSC NUV effective exposure time (relative response)
EFF_EXPTIME_FUV_GSC	s	GSC FUV effective exposure time
BKGRND_FLUX_NUV_GSC	$\mu\text{Jy}/\text{arcsec}^2$	GSC NUV sky background surface brightness
BKGRND_FLUX_FUV_GSC	$\mu\text{Jy}/\text{arcsec}^2$	GSC FUV sky background surface brightness
NUVFLUXLIMIT2p5_GSC	$\mu\text{Jy}$	GSC S/N=2.5 NUV flux detection limit for a circular source at this position
FUVFLUXLIMIT2p5_GSC	$\mu\text{Jy}$	GSC S/N=2.5 FUV flux detection limit for a circular source at this position
NUVMAGLIMIT2p5_GSC	mag	GSC S/N=2.5 NUV flux detection limit for a circular source at this position
FUVMAGLIMIT2p5_GSC	mag	GSC S/N=2.5 FUV flux detection limit for a circular source at this position
NUVFLUXLIMIT5_GSC	$\mu\text{Jy}$	GSC S/N=5 NUV flux detection limit for a circular source at this position
FUVFLUXLIMIT5_GSC	$\mu\text{Jy}$	GSC S/N=5 FUV flux detection limit for a circular source at this position
NUVMAGLIMIT5_GSC	mag	GSC S/N=5 NUV flux detection limit for a circular source at this position
FUVMAGLIMIT5_GSC	mag	GSC S/N=5 FUV flux detection limit for a circular source at this position
NUVFLUXLIMIT10_GSC	$\mu\text{Jy}$	GSC S/N=10 NUV flux detection limit for a circular source at this position
FUVFLUXLIMIT10_GSC	$\mu\text{Jy}$	GSC S/N=10 FUV flux detection limit for a circular source at this position

Table D.13: ObsInfo catalogue: column names &amp; description

Column name	Unit	Description
NUVMAGLIMIT10_GSC	mag	GSC S/N=10 NUV flux detection limit for a circular source at this position
FUVMAGLIMIT10_GSC	mag	GSC S/N=10 FUV flux detection limit for a circular source at this position
EFF_EXPTIME_NUV_GMC	s	GMC NUV effective exposure time (relative response)
EFF_EXPTIME_FUV_GMC	s	GMC FUV effective exposure time (relative response)
BKGRND_FLUX_NUV_GMC	$\mu\text{Jy}/\text{arcsec}^2$	GMC NUV sky background surface brightness
BKGRND_FLUX_FUV_GMC	$\mu\text{Jy}/\text{arcsec}^2$	GMC FUV sky background surface brightness
NUVFLUXLIMIT2p5_GMC	$\mu\text{Jy}$	GMC S/N=2.5 NUV flux detection limit for a circular source at this position
FUVFLUXLIMIT2p5_GMC	$\mu\text{Jy}$	GMC S/N=2.5 FUV flux detection limit for a circular source at this position
NUVMAGLIMIT2p5_GMC	mag	GMC S/N=2.5 NUV flux detection limit for a circular source at this position
FUVMAGLIMIT2p5_GMC	mag	GMC S/N=2.5 FUV flux detection limit for a circular source at this position
NUVFLUXLIMIT5_GMC	$\mu\text{Jy}$	GMC S/N=5 NUV flux detection limit for a circular source at this position
FUVFLUXLIMIT5_GMC	$\mu\text{Jy}$	GMC S/N=5 FUV flux detection limit for a circular source at this position
NUVMAGLIMIT5_GMC	mag	GMC S/N=5 NUV flux detection limit for a circular source at this position
FUVMAGLIMIT5_GMC	mag	GMC S/N=5 FUV flux detection limit for a circular source at this position
NUVFLUXLIMIT10_GMC	$\mu\text{Jy}$	GMC S/N=10 NUV flux detection limit for a circular source at this position
FUVFLUXLIMIT10_GMC	$\mu\text{Jy}$	GMC S/N=10 FUV flux detection limit for a circular source at this position
NUVMAGLIMIT10_GMC	mag	GMC S/N=10 NUV flux detection limit for a circular source at this position

Table D.14: Tab. D.13 continued

Column name	Unit	Description
FUVMAGLIMIT10_GMC	mag	GMC S/N=10 FUV flux detection limit for a circular source at this position
EFF_EXPTIME_NUV_GDC	s	GDC NUV effective exposure time
EFF_EXPTIME_FUV_GDC	s	GDC FUV effective exposure time
BKGRND_FLUX_NUV_GDC	$\mu\text{Jy}/\text{arcsec}^2$	GDC NUV sky background surface brightness
BKGRND_FLUX_FUV_GDC	$\mu\text{Jy}/\text{arcsec}^2$	GDC FUV sky background surface brightness
NUVFLUXLIMIT2p5_GDC	$\mu\text{Jy}$	GDC S/N=2.5 NUV flux detection limit for a circular source at this position
FUVFLUXLIMIT2p5_GDC	$\mu\text{Jy}$	GDC S/N=2.5 FUV flux detection limit for a circular source at this position
NUVMAGLIMIT2p5_GDC	mag	GDC S/N=2.5 NUV flux detection limit for a circular source at this position
FUVMAGLIMIT2p5_GDC	mag	GDC S/N=2.5 FUV flux detection limit for a circular source at this position
NUVFLUXLIMIT5_GDC	$\mu\text{Jy}$	GDC S/N=5 NUV flux detection limit for a circular source at this position
FUVFLUXLIMIT5_GDC	$\mu\text{Jy}$	GDC S/N=5 FUV flux detection limit for a circular source at this position
NUVMAGLIMIT5_GDC	mag	GDC S/N=5 NUV flux detection limit for a circular source at this position
FUVMAGLIMIT5_GDC	mag	GDC S/N=5 FUV flux detection limit for a circular source at this position
NUVFLUXLIMIT10_GDC	$\mu\text{Jy}$	GDC S/N=10 NUV flux detection limit for a circular source at this position
FUVFLUXLIMIT10_GDC	$\mu\text{Jy}$	GDC S/N=10 FUV flux detection limit for a circular source at this position
NUVMAGLIMIT10_GDC	mag	GDC S/N=10 NUV flux detection limit for a circular source at this position
FUVMAGLIMIT10_GDC	mag	GDC S/N=10 FUV flux detection limit for a circular source at this position

Table D.15: Tab. D.13 continued



# Bibliography

- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, *ApJS*, 182, 543
- Baldry, I. K., Glazebrook, K., Brinkmann, J., et al. 2004, *ApJ*, 600, 681
- Baldry, I. K., Robotham, A. S. G., Hill, D. T., et al. 2010, *MNRAS*, 404, 86
- Baldry, I. K., Driver, S. P., Loveday, J., et al. 2012, *MNRAS*, 421, 621
- Baldry, I. K., Alpaslan, M., Bauer, A. E., et al. 2014, *ArXiv e-prints*, arXiv:1404.2626
- Bell, E. F., & de Jong, R. S. 2001, *ApJ*, 550, 212
- Bertin, E., & Arnouts, S. 1996, *A&AS*, 117, 393
- Bianchi, L., Conti, A., & Shiao, B. 2014, *Advances in Space Research*, 53, 900
- Blanton, M. R., & Roweis, S. 2007, *AJ*, 133, 734
- Blanton, M. R., Dalcanton, J., Eisenstein, D., et al. 2001, *AJ*, 121, 2358
- Brewer, B. J., Foreman-Mackey, D., & Hogg, D. W. 2013, *AJ*, 146, 7
- Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, *MNRAS*, 351, 1151
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, 344, 1000
- Buat, V., & Burgarella, D. 1998, *A&A*, 334, 772
- Budavári, T., & Szalay, A. S. 2008, *ApJ*, 679, 301
- Calzetti, D. 2001, *PASP*, 113, 1449
- Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *ApJ*, 533, 682
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, 345, 245
- Chabrier, G. 2003, *PASP*, 115, 763
- Cortese, L., Boselli, A., Franzetti, P., et al. 2008, *MNRAS*, 386, 1157

- de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013, *Experimental Astronomy*, 35, 25
- Draine, B. T., Dale, D. A., Bendo, G., et al. 2007, *ApJ*, 663, 866
- Driver, S. P., Popescu, C. C., Tuffs, R. J., et al. 2008, *ApJ*, 678, L101
- . 2007, *MNRAS*, 379, 1022
- Driver, S. P., Norberg, P., Baldry, I. K., et al. 2009, *Astronomy and Geophysics*, 50, 050000
- Driver, S. P., Hill, D. T., Kelvin, L. S., et al. 2011, *MNRAS*, 413, 971
- Driver, S. P., Robotham, A. S. G., Kelvin, L., et al. 2012, *MNRAS*, 427, 3244
- Duffy, A. R., Moss, A., & Staveley-Smith, L. 2012, *PASA*, 29, 202
- Eales, S., Dunne, L., Clements, D., et al. 2010, *PASP*, 122, 499
- Efstathiou, G., Ellis, R. S., & Peterson, B. A. 1988, *MNRAS*, 232, 431
- Fabian, A. C. 1994, *ARA&A*, 32, 277
- Fioc, M., & Rocca-Volmerange, B. 1997, *A&A*, 326, 950
- Galametz, M., Madden, S. C., Galliano, F., et al. 2011, *A&A*, 532, A56
- Giovanelli, R., & Haynes, M. P. 1983, *AJ*, 88, 881
- Grootes, M. W., Tuffs, R. J., Popescu, C. C., et al. 2014, *MNRAS*, 437, 3883
- . 2013, *ApJ*, 766, 59
- Häußler, B., Bamford, S. P., Vika, M., et al. 2013, *MNRAS*, 430, 330
- Heinis, S., Budavári, T., & Szalay, A. S. 2009, *ApJ*, 705, 739
- Heymans, C., Van Waerbeke, L., Miller, L., et al. 2012, *MNRAS*, 427, 146
- Hill, D. T., Kelvin, L. S., Driver, S. P., et al. 2011, *MNRAS*, 412, 765
- Hopkins, A. M., Driver, S. P., Brough, S., et al. 2013, *MNRAS*, 430, 2047
- Humason, M. L., Mayall, N. U., & Sandage, A. R. 1956, *AJ*, 61, 97
- Johnston, R. 2011, *A&A Rev*, 19, 41
- Kelvin, L. S., Driver, S. P., Robotham, A. S. G., et al. 2012, *MNRAS*, 421, 1007
- Kennicutt, R. C., & Evans, N. J. 2012, *ARA&A*, 50, 531

- Kennicutt, Jr., R. C. 1983, *ApJ*, 272, 54
- . 1998, *ARA&A*, 36, 189
- Kron, R. G. 1980, *ApJS*, 43, 305
- Kroupa, P., Tout, C. A., & Gilmore, G. 1993, *MNRAS*, 262, 545
- Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, *MNRAS*, 379, 1599
- Le Fèvre, O., Vettolani, G., Garilli, B., et al. 2005, *A&A*, 439, 845
- Lintott, C., Schawinski, K., Bamford, S., et al. 2011, *MNRAS*, 410, 166
- Loveday, J., Norberg, P., Baldry, I. K., et al. 2012, *MNRAS*, 420, 1239
- Malmquist, K. 1920, *Medd. Lund. Astron. Obs*, 22, 1
- Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, *ApJ*, 619, L1
- Masters, K. L., Nichol, R., Bamford, S., et al. 2010, *MNRAS*, 404, 792
- Molinari, S., Bally, J., Glover, S., et al. 2014, *ArXiv e-prints*, arXiv:1402.6196
- Möllenhoff, C., Popescu, C. C., & Tuffs, R. J. 2006, *A&A*, 456, 941
- Morrissey, P., Conrow, T., Barlow, T. A., et al. 2007, *ApJS*, 173, 682
- Morrissey, P., Schiminovich, D., Barlow, T. A., et al. 2005, *ApJ*, 619, L7
- Naylor, T., Broos, P. S., & Feigelson, E. D. 2013, *ApJS*, 209, 30
- O'Donnell, J. E. 1994, *ApJ*, 422, 158
- Pastrav, B. A., Popescu, C. C., Tuffs, R. J., & Sansom, A. E. 2013a, *ArXiv e-prints*, arXiv:1301.5602
- . 2013b, *A&A*, 557, A137
- Peeples, M. S., & Shankar, F. 2011, *MNRAS*, 417, 2962
- Pierre, M., Picaud, F., Juin, J. B., et al. 2011, *MNRAS*, 414, 1732
- Popescu, C. C., Misiriotis, A., Kylafis, N. D., Tuffs, R. J., & Fischera, J. 2000, *A&A*, 362, 138
- Popescu, C. C., Tuffs, R. J., Dopita, M. A., et al. 2011, *A&A*, 527, A109+
- Rana, N. C., & Basu, S. 1992, *A&A*, 265, 499
- Robotham, A., Driver, S. P., Norberg, P., et al. 2010, *PASA*, 27, 76

- Robotham, A. S. G., & Driver, S. P. 2011, *MNRAS*, 413, 2570
- Salim, S., Rich, R. M., Charlot, S., et al. 2007, *ApJS*, 173, 267
- Salpeter, E. E. 1955, *ApJ*, 121, 161
- Sandage, A., Tammann, G. A., & Yahil, A. 1979, *ApJ*, 232, 352
- Schechter, P. 1976, *ApJ*, 203, 297
- Schinnerer, E., Smolčić, V., Carilli, C. L., et al. 2007, *ApJS*, 172, 46
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, 500, 525
- Sérsic, J. L. 1963, *Boletín de la Asociación Argentina de Astronomía La Plata Argentina*, 6, 41
- Shao, Z., Xiao, Q., Shen, S., et al. 2007, *ApJ*, 659, 1159
- Simard, L., Mendel, J. T., Patton, D. R., Ellison, S. L., & McConnachie, A. W. 2011, 2011arXiv1107.1518S, arXiv:1107.1518
- Smith, D. J. B., Dunne, L., da Cunha, E., et al. 2012, *MNRAS*, 427, 703
- Spergel, D. N., Verde, L., Peiris, H. V., et al. 2003, *ApJS*, 148, 175
- Stoughton, C., Lupton, R. H., Bernardi, M., et al. 2002, *AJ*, 123, 485
- Strauss, M. A., Weinberg, D. H., Lupton, R. H., et al. 2002, *AJ*, 124, 1810
- Sutherland, W., & Saunders, W. 1992, *MNRAS*, 259, 413
- Taylor, E. N., Hopkins, A. M., Baldry, I. K., et al. 2011, *MNRAS*, 418, 1587
- Tonry, J. L., Blakeslee, J. P., Ajhar, E. A., & Dressler, A. 2000, *ApJ*, 530, 625
- Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, *ApJ*, 613, 898
- Treyer, M., Wyder, T. K., Schiminovich, D., et al. 2005, *ApJ*, 619, L19
- Tuffs, R. J., Popescu, C. C., & Völk, H. J. 2004, *Astronomische Nachrichten Supplement*, 325, 119
- Weingartner, J. C., & Draine, B. T. 2001, *ApJ*, 548, 296
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, 140, 1868
- Wyder, T. K., Treyer, M. A., Milliard, B., et al. 2005, *ApJ*, 619, L15
- Wyder, T. K., Martin, D. C., Schiminovich, D., et al. 2007, *ApJS*, 173, 293
- Xilouris, E. M., Byun, Y. I., Kylafis, N. D., Paleologou, E. V., & Papamastorakis, J. 1999, *A&A*, 344, 868
- York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, *AJ*, 120, 1579