The Andromeda Optical and Infrared Disk Survey

by

Jonathan Sick

A thesis submitted to the
Department of Physics, Engineering Physics and Astronomy
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen's University
Kingston, Ontario, Canada
February 2018

Abstract

The spectral energy distributions of galaxies inform us about a galaxy's stellar populations and interstellar medium, revealing stories of galaxy formation and evolution. How we interpret this light depends in part on our proximity to the galaxy. For nearby galaxies, detailed star formation histories can be extracted from the resolved stellar populations, while more distant galaxies feature the contributions of entire stellar populations within their integrated spectral energy distribution. This thesis aims to resolve whether the techniques used to investigate stellar populations in distant galaxies are consistent with those available for nearby galaxies. As the nearest spiral galaxy to us, the Andromeda Galaxy (M31) is the ideal testbed for the joint study of resolved stellar populations and panchromatic spectral energy distributions (SEDs). We present the Andromeda Optical and Infrared Disk Survey (ANDROIDS), which adds new near-UV to near-IR ($u^{*}g'r'i'JK_{s}$) imaging using the MegaCam and WIRCam cameras at the Canada-France-Hawaii telescope to the available M31 panchromatic dataset. In order to accurately subtract photometric background from our extremely wide-field (14 square degree) mosaics, we present integrated observing and data reduction techniques with sky-target nodding, optimization of image-to-image surface brightness, and a novel hierarchical Bayesian model to trace the background signal while modelling the astrophysical SED. We model the spectral energy distributions of M31 pixels with MAGPHYS (da Cunha et al. 2008) and compare those results to resolved stellar population models of the same pixels from the Panchromatic Hubble Andromeda Treasury (PHAT) survey (Williams et al. 2017). We find substantial (0.3 dex) differences in stellar mass estimates despite a common use of the Chabrier (2003) initial mass function. Stellar mass estimated from the resolved stellar population is larger than any mass estimate from SED models or colour-$\mathcal{M}_*/L$ relations (CMLRs). There is also considerable diversity among CMLR estimators, largely driven by differences in the star formation history prior distribution. We find broad consistency between the star formation history estimated by integrated spectral energy distributions and resolved stars. Generally, spectral energy distribution models yield a stronger inside-out radial metallicity gradient and bias towards younger mean ages than resolved stellar population models.

Co-Authorship

Loic Albert (CFHT), Stéphane Courteau (Queen's University), Jean-Charles Cuillandre (CFHT), Julianne Dalcanton (UW), Puragra Guhathakurta (UCSC), Roelof de Jong (AIP), Mike McDonald (MIT), Luca Rizzi (Joint Astronomy Centre), Brent Tully (UH) co-authored the CFHT WIRCam and MegaCam telescope time proposals; those datasets are featured in Chapters 2 and 3. Stéphane Courteau, Jean-Charles Cuillandre, Michael McDonald, Roelof de Jong, and Brent Tully co-authored Sick et al. (2014); that work is presented here as Chapter 2. Jean-Charles Cuillandre developed the Elixir-LSB and applied it to the CFHT/MegaCam observations as described in Chapter 3. The idea of using a hierarchical Bayesian model to correct spatially-varying background in the ANDROIDS spectral energy distribution was developed with David Hogg (NYU) at the AstroData Hack Week at the University of Washington in 2014; those discussions inspired Chapter 4. Stéphane Courteau created the isophotal profile of the ANDROIDS panchromatic SED, as presented in Chapter 5. Joel Roediger (NRC-HIA) provided a modified version of MAGPHYS and Christine Hall (Queen's) provided insight into mid-IR mass-to-light ratio relations that aided Chapter 6. I also thank Ben Williams (UW) and the committee for providing comments that improved the final version of this manuscript. All other work is the result of the author's own efforts.

Acknowledgments

This research made use of Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013), as well as the NumPy[1] and Scipy[2], Matplotlib (Hunter 2007) and APLpy (Robitaille & Bressert 2012) open-source Python packages. This research has made use of NASA's Astrophysics Data System Bibliographic Services. This research is based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/DAPNIA, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Science de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. Additionally, this research is based on observations obtained with WIRCam, a joint project of CFHT, Taiwan, Korea, Canada, France, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique of France, and the University of Hawaii. This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and

[1] `http://www.numpy.org`
[2] `https://www.scipy.org/scipylib/index.html`

Space Administration and the National Science Foundation. The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain. Computations were performed on CANFAR, a scientific data centre operated by CANARIE, the Canadian Astronomy Data Centre, Compute Canada, and the National Research Council of Canada. The Large Synoptic Survey Telescope project, through the Association of Universities for Research in Astronomy, provided additional computational support. The author was supported by an NSERC Alexander Graham Bell Postgraduate Scholarship.

I am indebted to by supervisor, Stéphane Courteau, for conceiving of this ambitious research project and helping me bring it to fruition. I thank the community of professors and fellow students at Queen's University for excellent discussions. I extend special thanks to Joel Roediger (NRC-HIA) for sharing a modified version of MAGPHYS and Christine Hall (Queen's) for discussions about mid-IR mass-to-light ratio relations. I also thank Ben Williams (UW) and the committee for providing comments that improved this manuscript.

Finally, and most significantly, I couldn't have done this without the support of my family, especially my parents Carmel and Gordon, and my fiancé Amanda Murdoch.

Contents

List of Tables

List of Figures

List of Symbols

$\langle A \rangle$ The mass-weighted mean age of a stellar population. Age is measured as a lookback time from the present.

g SDSS bandpass centred at 470 nm, calibrated in the AB magnitude system. g' is the related CFHT/MegaCam bandpass.

i SDSS bandpass centred at 749 nm, calibrated in the AB magnitude system. i' is the related CFHT/MegaCam bandpass.

J Near-IR bandpass centred at 1.2 μm. For this study, the AB magnitude system is assumed unless stated otherwise.

K_s Near-IR bandpass centred at 2.2 μm. For this study, the AB magnitude system is assumed unless stated otherwise.

L_λ Luminosity (typical in units of solar luminosity, $L_\odot$) at a specific wavelength or integrated over a certain bandpass, designated by the subscript (for example, L_i is the i-band luminosity).

$\mathcal{M}_*$ Stellar mass, typically in units of solar masses ($\mathrm{M}_\odot$).

$\log_{10} \mathcal{M}_*/L_i$ The i-band (SDSS AB magnitude) stellar mass-to-light ratio.

r SDSS bandpass centred at 616 nm, calibrated in the AB magnitude system. r' is the related CFHT/MegaCam bandpass.

R_e De-projected effective radius. Half of M31's light is contained by this radius. We adopt R_e based on the disk scale length measured by Courteau et al. (2011), $R_d = 5.3$ kpc. Thus $R_e = 1.678 * 5.3$ kpc.

R_{M31} De-projected M31 disk radius. That is, the distance of a pixel to M31's center if the disk is viewed face on.

u SDSS bandpass centred at 355 nm, calibrated in the AB magnitude system. u^* is the related CFHT/MegaCam bandpass.

$\log Z/Z_\odot$ The metallicity. Z is the mass fraction of elements heavier than hydrogen and helium in a stellar population. $Z_\odot$ is the metallicity of the Sun.

List of Acronyms

2MASS Two Micron All Sky Survey.

PSC Point Source Catalog. Typically, the 2MASS PSC in the context of this work.

AGB Asymptotic giant branch. Post-main sequence phase of stellar evolution for low and intermediate mass stars where Helium burning occurs in a shell.

ANDROIDS Andromeda Optical and Infrared Disk Survey.

CANFAR Canadian Advanced Network for Astronomy Research.

CFHT Canada-France-Hawaii Telescope.

CMD Colour-magnitude diagram.

CMLR Colour-mass-to-light ratio. A relation between a colour (difference of magnitudes) and mass.

DN Digital number. The native intensity unit of an image before absolute calibration.

FSPS Flexible Stellar Population Synthesis. Population synthesis code (Conroy et al. 2009, 2010; Conroy & Gunn 2010).

HELGA Herschel Exploitation of Local Galaxy Andromeda (Fritz et al. 2012).

IMF Initial mass function.

IRAC Infrared Array Camera. An instrument on the *Spitzer* space telescope.

ISM Interstellar medium.

IQR Inter-quartile range. Interval from the 25% percentile to 75% percentile of a distribution.

LSB Low surface brightness.

LGGS Local Group Galaxy Survey (Massey et al. 2006).

MAGPHYS Multi-wavelength Analysis of Galaxy Physical Properties. SED modelling software (da Cunha et al. 2008).

MCMC Markov Chain Monte Carlo. A Bayesian estimation method.

NGS Nearby Galaxy Survey (Thilker et al. 2005).

NIR Near-infrared. Generally defined as the wavelength range of 0.7–2.5 μm.

NM Nelder & Mead (1965). Multi-dimensional simplex-based optimization algorithm.

RGB Red giant branch. Post-main sequence phase of stellar evolution for low and intermediate mass stars where Hydrogen fusion occurs in a shell around a Helium core.

PACS Photodetecting Array Camera and Spectrometer. An instrument on the *Herschel* space telescope.

pdf Probability distribution function.

PHAT Panchromatic Hubble Andromeda Treasury (Dalcanton et al. 2012).

SB Surface brightness. Measured in units: mag arcsec^{-2}.

SDSS Sloan Digital Sky Survey.

SED Spectral energy distribution. A set of flux measurements and a range of effective wavelengths.

SPIRE Spectral and Photometric Imaging Receiver. An instrument on the *Herschel* space telescope.

SPS Stellar population synthesis. The category of algorithms for generating a model SED for a population of stars that combines individual stellar spectra, isochrones, IMF, star formation history, chemical evolution, and dust effects.

SSP Simple stellar population. A population of stars, generated from a given IMF, with a single age and a single metallicity.

ST Sky-target, as in "sky-target nodding."

TRILEGAL TRIdimensional modeL of thE GALaxy (Girardi et al. 2005).

TP-AGB Thermally-pulsating asymptotic giant branch. Post-main sequence phase of stellar evolution for low and intermediate mass stars where a Hydrogen burning shell sporadically re-ignites an inner Helium burning shell.

WCS World coordinate system (Greisen & Calabretta 2002; Calabretta & Greisen 2002).

WIRCam Wide-field Infrared Camera. An instrument on CFHT.

WISE Wide-field Infrared Survey Explorer. A space-based telescope.

Chapter 1

Introduction

Measuring the structure of galaxies, and understanding their formation and evolution, is a key endeavour of modern astronomy. Galaxies are environments that bridge cosmology and our understanding of dark matter and dark energy with the baryonic physics of star formation and stellar astrophysics. The shapes, ages, kinematics, metallicities and relative fractions of galaxy components (bulge, disk, halo) in large spiral galaxies like our own reveal precious information about their formation, accretion, and merging histories (see the review of Kormendy & Kennicutt 2004).

The spectral energy distributions (SEDs) of stellar populations tell us about their formation histories, metallicities, and masses. How we observe starlight and infer physical properties about the stellar population depends in large part on distance to the galaxy, and trade-offs in details versus observational efficiency. Within our own Milky Galaxy, and the Local Group, individual stars are resolvable with both ground and space-based telescopes. This enables measure the individual photometry of populations stars and fit their distributions in color-magnitude diagrams with stellar evolutionary models. Beyond the Local Group (distances much greater than 1 Mpc), individual stars are no longer resolvable, and instead we must model the integrated light of populations of stars. Modelling an integrated stellar population is difficult because instead of estimating a single age, metallicity and dust extinction parameter, we now model distributions of ages, metallicities, and dust attenuation from a single spectrum or SED.

Despite the difficulty, there is tremendous value in accurately estimating the properties of distant, unresolved stellar populations. With a greater observational volume, we increase the diversity of galaxy types and environments that we can study. And by observing high-redshift galaxies, we directly observe galaxies earlier in their evolution. Thus, in extragalactic astronomy we are confronted by trade-offs in the accuracy with which we can understand the Local Group, and the expanded laboratory afforded by distance.

We can confront this trade-off directly by simultaneously applying observation and modelling methods suited for nearby and distant galaxies alike. Our laboratory is the Andromeda Galaxy, M31. As a disk galaxy with a similar mass, M31 is an approximate analogue to our own Milky Way. At a distance of $D_{\mathrm{M31}} = 785$ kpc (McConnachie et al. 2005), $1'' = 3.7$ pc. Ground-based telescopes, with typical resolutions of $0\farcs6$, then, can study discrete neighbourhoods in M31. Further, it is possible to resolve individual stars.

Thus our objective in this work is to apply and compare an array of modelling methods to the M31 galaxy. In doing so, we can calibrate biases and uncertainties in the modelling of more distant stellar populations.

1.1 Stellar Population Estimation Methods

How we estimate the properties of a galaxy's stellar populations depends on the proximity of the galaxy and the observation instrumentation being used. In an ideal case, we would obtain a high-resolution spectrum of each star in a galaxy to learn about its luminosity (L), temperature (T_{eff}), surface gravity ($\log g$), and composition. However, this type of dataset is impractical to obtain. Even with massive multi-object spectrographs and wide-field integral field units becoming increasingly available, such spectra are far less efficient to produce than broadband images of resolved stars. And in most cases, individual stars in distant galaxies are either too faint or too crowded to obtain resolved spectra in external galaxies. Thus we typically work with compressed datasets when estimating the properties of a stellar population. This compression occurs on both spectral and spatial axes. For example, if stars are not resolved, their light is compressed into a single pixel. Or rather than measuring a high-resolution SED (a spectrum) we might image a galaxy in multiple broadband filters, thereby compressing spectral information. Different instrumentation and intrinsic factors about a galaxy (such as its proximity) cause us to make different compromises about the way that a stellar population's information is compressed. Corresponding to each choice, different estimation methods are available. In this section we briefly review those methods, recalling that the aim of this work is to directly compare stellar population estimation methods.

When a stellar population is resolvable (a Local Group galaxy or an equidistant Milky Way stellar population), we typically image the star field in broadband filters, measure the brightness of individual stars in each filter with PSF-fitting photometry software (Stetson 1987; Dolphin 2000, 2016), and plot the distribution of measured stars on a colour-versus-magnitude diagram. The "colour" aspect of a colour-magnitude diagram (CMD) is the difference of magnitudes in a pair of relatively blue and red broadband filters. Such a colour is a measurement of the shape of the overall SED of a star, which is in turn related to a star's temperature. Thus the observed CMD can be modelled with a synthesized stellar population to fit the distributions of stellar ages, metallicities, and even dust extinction, in the observed stellar population. This type of modelling is done with codes such as *StarFISH* (Harris & Zaritsky 2001) and *MATCH* Dolphin (2002, 2012, 2013). The main attraction of CMD modelling is that stellar evolution is plainly measurable in the CMD plane (as a proxy to a temperature-luminosity plane). The colour and magnitude of well-defined features such as the main sequence turn off and red giant branch are clear indicators of a stellar population's age and metallicity.

A drawback of CMD modelling is that only two bandpasses of SED information are used at once. Unfortunately, different colours are sensitive to different aspects of stellar evolution. One could model the same stellar population in several CMD planes, but it is not obvious how to do so properly since different CMD planes give different results (Hills et al. 2015). A promising technique is to model the hyperdimensional SED of individual stars with a Bayesian analysis (van Dyk et al. 2009). von Hippel et al. (2006) applied this method to model simple stellar populations (SSP) such as globular clusters. Recently, Dalcanton et al. (2015) and Gordon et al. (2016) applied a Bayesian approach to map dust extinction of M31's resolved stellar populations.

With more distant galaxies it is impossible to measure individual stars, and instead entire stellar populations are compressed into one imaging pixel. Since imaging is observationally efficient, it is typical to image a galaxy with many bandpasses, even many instruments and telescopes, and sample the entire UV-to-IR SED of a galaxy in low-spectral-resolution broadband measurements (see Walcher et al. 2011, for a broad review of galaxy SED modelling). Stellar population synthesis software, such as GALEV (Bruzual & Charlot 2003) and FSPS (Conroy et al. 2009), model these observed broadband SEDs. The main challenge is that broadband SEDs provide insufficient information to non-parametrically model such properties as metallicity or star formation history, as one can do with CMD modelling. Instead, the common practice is to assume a parameterized stellar population (such as a single metallicity and an exponentially declining star formation rate) that has far fewer parameters than CMD-based stellar population models. A bonus of working in a constrained model space is that the parameter space may be adequately sampled to pre-compute a library of SED models. Thus SED modelling is not only observationally efficient, but also computationally efficient. At a minimum, an SED can be "fit" by finding the model in the library that has the smallest χ^2 difference to the observed fluxes. Taylor et al. (2011) found that χ^2 best-fits result in stellar mass-to-light ratios that are systematically biased by +0.1 dex. A more accurate way of modelling an SED given a library of models is by marginalizing across the likelihood of all models, a Bayesian method used by MAGPHYS (da Cunha et al. 2008) and Prospector (Leja et al. 2017), among many other implementations.

When observational resources are limited, stellar population parameters can be estimated from single colours, that is, a pair of images in different bandpasses. Bell & de Jong (2001) pioneered the use of colours for estimating stellar mass-to-light ratios, with the optical–near-IR $g - i$ colour being commonly used. Star formation rates can also be estimated with UV-IR colours. Different colours are sensitive to the star formation rates at different lookback times (e.g., Calzetti 2013). Given two colours, with images in three or four different bandpasses, the characteristic age and metallicity of a stellar population can even be estimated from its location in a colour-colour diagram.

Unfortunately, colour-based stellar population estimation in particular, and SED-based estimation in general, is thwarted by a three-fold degeneracy between stellar age (A), metallicity (Z) and ISM dust. The reddening caused by increased age is degenerate with the reddening caused by increased metallicity (enhanced absorption lines at shorter wavelengths in stellar atmospheres), and of course by dust attenuation. A promising method of reducing this degeneracy is with a combination of optical and infrared images, along with realistic dust models (de Jong 1996; MacArthur et al. 2004; Pforr et al. 2012).

Compounding these challenges are fundamental uncertainties in modern stellar population synthesis, particularly uncertainties in the interpretation of near-infrared (NIR) light. For instance, optical-NIR SEDs yield unreliable population synthesis fits compared to optical-only SED fits. This failure is largely attributable to inadequate stellar population synthesis recipes for NIR bands and naive parameterization of star formation histories (Taylor et al. 2011; Courteau et al. 2014).

First, spectral energy distribution (SED) fitting often relies on simplistic star formation history (SFH) parameterizations. Because NIR colours lift age-metallicity-dust degeneracies, modelling of NIR bands may require additional sophistication, namely composite star formation and metal enrichment histories. The appropriate form of SFH models cannot be constrained from the integrated light of galaxies alone (as is typically attempted); resolved color-magnitude diagrams (CMDs) are both more effective and, in fact, essential for deriving non-parametric stellar population histories.

Second, asymptotic giant branch (AGB) stars from intermediate-aged stellar populations heavily influence the NIR light (Maraston 1998). Modelling AGB stars is most challenging due to their complex dredge-up cycles that change surface chemistry and temperature (the M- to C-type transition), and circumstellar winds that further perturb an AGB star's location in the CMD. A proper calibration of NIR stellar population synthesis models (*e.g.* Maraston 2005; Marigo et al. 2008, Charlot & Bruzual in prep.) may yield a 30%–50% improvement in the estimation of stellar masses and ages of high redshift systems (*e.g.* Maraston et al. 2006; Bruzual 2007; Conroy & Gunn 2010; Conroy 2013).

In this work, we have proposed a remedy by observing M31's entire bulge and disk ($R \leq 20$ kpc) in both resolved and integrated stellar light at J and K_s wavelengths. In doing so, we can directly relate a NIR stellar population's decomposition in the color-magnitude plane to the panchromatic SED of M31. Though such a calibration could be made with other galaxies, M31 is unique in its proximity so that even ground-based instrumentation can resolve its bright stellar population.

Furthermore, both resolved (CMD fitting) and unresolved (SED and colour-based models) stellar population models can be applied simultaneously at M31, thanks to its proximity. Thus we can directly establish systematics of different estimation methods when applied to a massive galaxy with a potentially complex stellar population.

1.2 Towards a Comprehensive Picture of M31

Elements of the M31 narrative are being assembled, with recent contributions from large surveys. Initially discovered by Guhathakurta et al. (2005), the CFHT Pan-Andromeda Archaeological Survey (PAndAS; McConnachie et al. 2009) comprehensively mapped the vast stellar halo of M31 and M33 and confirmed our basic picture of hierarchical galaxy formation in a ΛCDM cosmology. Deep observations by the *Hubble Space Telescope* have resolved the star formation histories at discrete points across the outer disk and inner halo of M31. Colour-magnitude diagrams produced by Brown et al. (2003), Ferguson et al. (2005) and Bernard et al. (2012), among others, can resolve the red clump and main sequence turn-off populations that are sensitive indicators of stellar population ages and metallicites. Superimposed on the old metal-poor population expected in a galaxy halo, those authors find metal-rich intermediate-age (6–8 Gyr) populations that can only be explained by mass accretion. Richardson et al. (2008) find that the HST fields can be described as combinations of disk-like and stream-like colour-magnitude diagrams CMDs. The proportions of these populations is consistent with N-body simulations by Fardal et al. (2007) of the giant stream interaction with M31.

The bulge of M31 also tells a complex story. Bulge-disk-halo decompositions of surface brightness profiles by Courteau et al. (2011) yield a bulge with a Sérsic index of $n \simeq 2.2 \pm 0.3$, and a bulge-to-disk ratio of $R_e/R_d \sim 0.2$. Courteau et al. (1996) find that this shape and size are consistent with a bulge formed classically by mergers (*e.g.,* Searle & Zinn 1978) followed by secular evolution where angular momentum tends to increase central concentration while also spreading the disk outwards. Stellar populations deduced from the long-slit spectra of Saglia et al. (2010) confirm this story, with the bulge forming early ($\sim$ 12 Gyr ago) and rapidly (an α-element enhanced population), while a 4–8 Gyr old population in the inner arcseconds of the M31 nucleus indicate secular evolution. Despite its classical shape, NIR images from the 2MASS survey (Beaton et al. 2007) show that the bulge has boxy isophotes, with a major axis that is mis-aligned with the disk. Athanassoula & Beaton (2006) find this consistent with a bar embedded in a classical bulge, although Saglia et al. (2010) find no kinematic evidence for a bar. A more complete 2D kinematic analysis coupled with stellar populations and light profiles will be needed to resolve the structure of M31's bulge.

Our knowledge of M31's disk also remains incomplete. Using de-projected *Spitzer* IRAC and MIPS mosaics, tracing hot dust in star-forming regions, Gordon et al. (2006) and Block et al. (2006) find that M31 has heavily-disturbed spiral arms that are best described as two pseudo-rings with radii of 10 kpc and 1.5 kpc. Both groups suggest that these star-forming pseudo-rings could have been formed by a head-on interaction with M32, with Block et al. specifically modelling an interaction occurring 250 Myr ago, while Gordon et al. (2006) models an interaction 20 Myr ago. On the other hand, rings can also be formed by orbital resonances with a bar (Athanassoula & Beaton 2006).

The timescales of galaxy formation and evolution scenarios can be arbitrated by stellar population evidence. HST has been very useful in providing CMDs of resolved stars. Williams (2002) collected imagery for 27 WFPC2 fields from the HST archive that cover the M31 disk inside, along, and outside the 10 kpc star forming ring. He found that the star forming rate was universally high, 2–20 $M_\odot$ yr^{-1}, between 10 and 13 Gyr ago. Roughly 1 Gyr ago the star forming rate declined. In some fields, even in the outer disk at $R \sim 20$ kpc, 100 Myr young star burst populations are detected. Specific star formation rate indicators, such as $L(\mathrm{H}\alpha)$, and $L(8\mu\mathrm{m})$ give estimates of 0.4 $\mathrm{M}_\odot\ \mathrm{yr}^{-1}$ (Barmby et al. 2006), while Kang et al. (2009) infer 0.6–0.7 $\mathrm{M}_\odot\ \mathrm{yr}^{-1}$ from $L(\mathrm{NUV})/L(\mathrm{FUV})$. Both UV observations and HST CMDs indicate a recent peak in star formation roughly 100 Myr ago, which is consistent with an interaction that could have formed the 10 kpc ring.

A map of of M31's star forming rate, however, cannot be interpolated from the sparse HST sampling of the disk. Recent surveys have begun to fill in this picture. The first project to consistently map point sources across the M31 disk is the Local Group Galaxy Survey (LGGS) of Massey et al. (2006). LGGC compiled $UBVRI$ mosaics using the Mayall 4-m telescope under modest, $0\farcs8$ – $1\farcs2$ seeing. Exposure depths were 50 minutes in U, 5 minutes in BVR and 12.5 minutes in I. Williams (2003) used the LGGS BV data set to map the global star formation of M31's resolved stellar disk in the last 250 Myr, segmented in $4' \times 4'$ regions. That map is a tantalizing tell-tale of M31's recent past, where star formation propagates asymmetrically about the 10 kpc ring. In the vicinity of M32, star formation peaked between 2^7–2^8 Myr ago, suggestive of triggering by an interaction. Unfortunately the LGGS star catalog could not be used to fit star formation histories older than 250 Myr. $B-V$ CMDs are contaminated by foreground Milky Way dwarfs and red giant stars, and the arcsecond seeing is too poor for resolving the dense fields of M31 red giant stars.

A first step towards globally mapping M31's older stellar populations came from wide-field mapping with the Sloan Digital Sky Survey (SDSS). Tempel et al. (2011) assembled $ugriz$ images of M31's entire disk out to $R_{\mathrm{M31}} \sim 20$ kpc. Those authors took advantage of the SDSS's drift-scanning camera to interpolate and subtract background from long, continuous imaging stripes across M31's disk. Unfortunately, the mosaics assembled by Tempel et al. (2011) are relatively shallow — photometric uncertainty exceeds 0.1 mag at $R_{\mathrm{M31}} > 10$ kpc. Tamm et al. (2012) used these mosaics to model M31's optical SED with three Blanton & Roweis (2007) composite spectral templates and found that M31's disk is characterized as uniformly old (7–12 Gyr) and near-solar metallicity ([Fe/H] = 0.03). Given their use of just three characteristic spectral templates, it is difficult to read further into M31's evolution from those results.

These limited results have given way to the Panchromatic Hubble Andromeda Treasury (PHAT: Dalcanton et al. 2012). PHAT has provided a transformative view of M31's resolved stellar populations that covers roughly a third of the M31 bulge and disk out to 20 kpc with six-band HST/WFC3 UV–NIR imaging (Williams et al. 2014). With this dataset, Lewis et al. (2015) re-measured the recent star formation across M31's disk by modelling M31's main sequence stellar populations in F475W-F814W CMDs. One of the most important results from that work is that M31's 10 kpc ring has existed, and has remained stationary, for at least 400 Myr. Thus the 10 kpc ring is more likely a long-lived dynamical resonance than a product of a collision with M32. Gregersen et al. (2015) also used PHAT photometry and modelled the colour of M31's red giant branch to confirm that M31 has a near-solar to slightly sub-solar metallicity. Stellar metallicity declines with radius, which is consistent with inside-out galaxy evolution where the bulge and central disk form first and evolve to become metal-rich, while the outer parts form later with characteristically lower metallicities and younger ages. Finally, Williams et al. (2017) fit the full M31 CMD, to model M31's entire range of stellar populations. Those results have confirmed the earlier studies by Lewis et al. (2015) and Gregersen et al. (2015), and determined that the bulk of M31's disk formed earlier than 8 Gyr ago and that this old stellar population is well-mixed across M31's disk.

Space-based observatories, meanwhile, have provided a comprehensive mapping of M31's spectral energy distribution with sub-kpc scale resolution. Herschel observations (Fritz et al. 2012; Draine et al. 2014) in the mid and far-IR, along with the aforementioned Spitzer IRAC and MIPS images, and GALEX UV images (Gil de Paz et al. 2007) have enabled full-SED modelling of M31 to understand both its stellar populations and interstellar medium (ISM). Viaene et al. (2014) used such a panchromatic dataset in conjunction with MAGPHYS (da Cunha et al. 2008), a tool commonly used for distant galaxies, to model the stellar populations and ISM in individual pixels across M31's disk. A key goal of that study was to understand the star formation scaling relationships, or the relationship between star formation and stellar mass or gas density.

This approach, of applying panchromatic SED modelling tools to individual M31 pixels, is promising and can be applied more broadly In this thesis we show how SED modelling can be used to characterize all aspects of M31's stellar populations including stellar ages, and metallicities. In doing so, the broadband SED perspective of M31's entire disk can be rationalized with the detailed view of one quadrant of M31 already characterized in detail with resolved stellar populations by the PHAT survey.

1.3 The Andromeda Optical and Infrared Disk Survey

We have established, in the previous sections, the Andromeda Galaxy's immense value in understanding the structure and evolution of massive spiral galaxies, and serving as a testing ground for stellar population estimation methods. The PHAT survey has provided the state-of-the-art near-UV to near-IR resolved star catalog of a third of M31's bulge and disk. However, we still lack a complementary near-UV to near-IR surface brightness dataset.

The challenge of building a well-calibrated, high-resolution, and high signal-to-noise optical-to-near-IR M31 dataset is the large area that must be imaged. The diameter of M31's disk, within 20 kpc, is 3 degrees on the sky. Even with the state-of-the-art wide-field images available from medium-scale (4-m) telescopes, observing M31 requires imaging multiple fields. While making wide-field mosaics is common for surveying large areas of distant galaxies or point source objects, there is little precedent for wide, multi-field surface brightness imaging. The key challenge in this case is that photometric background from the Earth's atmospheric emission and scattered light cannot be directly measured from the field being observed, since M31 light covers the entire field as well. Previous wide-field optical (Tempel et al. 2011) and near-IR (Beaton et al. 2007) M31 datasets have taken advantage of scanning imagers in the SDSS and 2MASS telescopes to interpolate backgrounds sampled from the beginning and end of long scans across M31's disks to subtract background from the disk itself. In practice, this method has provided limited success. The 2MASS JHK_s mosaics made by Beaton et al. (2007) have obvious background subtraction issues that prevent their use beyond the bulge. And the SDSS mosaics of Tempel et al. (2011) are comparatively shallow, with uncertainties in surface brightness profiles exceeding

0.1 mag by 10 kpc.

In this work, we present the Andromeda Optical and Infrared Disk Survey (ANDROIDS) with the goal of building state-of-the-art surface brightness maps in near-UV to near-IR bands of the M31 galaxy. ANDROIDS uses the Canada-France-Hawaii Telescope (CFHT) observatory. Its square-degree optical camera, MegaCam, is proven to be suitable for wide-field surface photometry thanks to development for the Next Generation Virgo Cluster Survey (Ferrarese et al. 2012). Beyond having a larger mirror aperture than SDSS, CFHT/MegaCam also has greater near-UV sensitivity in its u^* band than SDSS's u band. CFHT's WIRCam camera is particularly unique as a wide-field near-IR camera. Altogether, CFHT enables us to build a deep and well-calibrated near-UR to near-IR ($u^* g' r' i' J K_s$) image set of M31's entire bulge and disk with excellent seeing ($0\farcs65$).

In this work, we present the Andromeda Optical and Infrared Disk Survey as follows. In Chapter 2, we present our CFHT/WIRCam near-IR observations and develop calibration methods to address background subtraction. Then in Chapter 3, we present the CFHT/MegaCam optical maps. In Chapter 4, we address the issue of background calibration in our wide-field optical and near-IR mosaics with a novel hierarchical Bayesian model of pixel spectral energy distributions. In Chapter 5, we combine our CFHT MegaCam and WIRCam dataset with datasets available in the literature to yield a homogeneous UV to far-IR SED dataset for M31. We model M31's stellar populations and dust with this SED dataset in Chapter 6 and compare SED modelling to other methods. In Chapter 7 we explore resolved stellar population modelling with the Panchromatic Hubble Andromeda Treasury dataset. Finally in Chapter 8 we synthesize results pertaining to wide-field dataset calibration, stellar population modelling, and characterize M31's stellar populations.

Chapter 2

WIRCam Observations and Reduction

This chapter is based upon the publication Sick et al. (2014).

2.1 Introduction

ANDROIDS uses the WIRCam instrument (Puget et al. 2004) on the Canada-France-Hawaii Telescope (CFHT), which is among the first generation of wide-field ground-based NIR detector arrays, covering a $21'.5 \times 21'.5$ field of view. Indeed, the advent of detectors such as WIRCam makes such a wide-field, high-resolution survey of an object as vast as M31 possible. The excellent natural seeing ($0''.65$) on Mauna Kea is sufficient for resolving giant-branch stars throughout the disk of M31.

Recovering the true NIR surface brightness map of M31 is, however, technically challenging. The NIR background (from both atmospheric fluorescence and instrumental thermal emission) is $\sim 10^3\times$ brighter than the NIR surface brightness of M31 at $R = 20$ kpc, demanding exceptional background characterization. Whereas most NIR galaxy surveys can measure the instantaneous background from blank sky pixels surrounding the galaxy on a detector array, M31's extended size requires physically nodding the telescope away from the galaxy by 1°–3° to sample blank sky (called Sky-Target, or ST, nodding). That we can never observe the instantaneous background on the disk of M31, but rather sample the sky at both a different location and time, introduces additional complications. Adams & Skrutskie (1996) clearly showed, with $9^\circ \times 9^\circ$ movies of the sky, that NIR sky emission has coherent spatial structure that moves across the sky over those scales, akin to a cirrus cloud system. This implies that background sampled from a sky field *will not* correspond directly to the sky affecting disk observations.

Besides background level, another concern is the accuracy of surface brightness shapes across individual WIRCam fields of view. Spatial structures in the NIR sky can leave residual shapes in background-subtracted disk images that ultimately affect our ability to produce a seamless NIR mosaic of M31. Vaduvescu & McCall (2004) also found that detector systems themselves, in their case the (now decommissioned) CFHT-IR camera, can add a time-varying background signal whose strength may be comparable to the NIR surface brightness of the outer M31 disk.

Because such a large mosaic has never before been assembled in a ST nodding WIRCam program, we focus this contribution on engineering the best practices for this type of observing. This includes: finding the optimal ST nodding cadence, defining the appropriate data reduction procedures for a WIRCam surface brightness reduction, and finally presenting an analysis of the surface brightness accuracy in wide-field WIRCam mosaics.

§ 2.2 describes the novel observational strategies used to reduce background subtraction uncertainties. § 2.3 presents the image reduction pipeline; with night sky flat fielding and median background subtraction in § 2.4, and zeropoint calibration

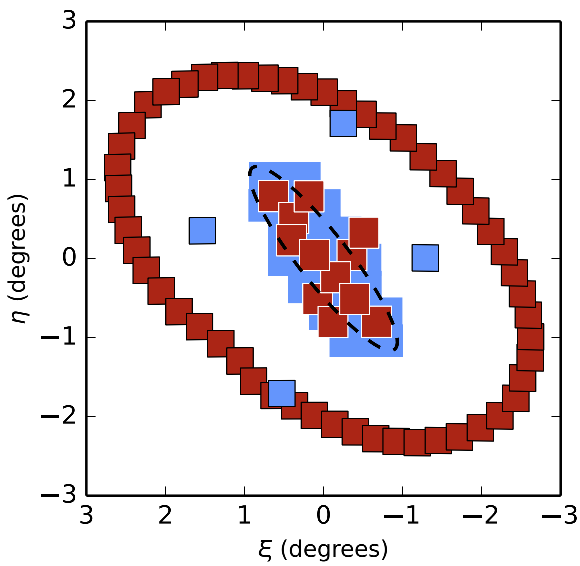


Figure 2.1: ANDROIDS WIRCam field positions on M31. Central blue tiles are the 27 disk fields observed in 2007B, surrounded by four sky fields. Dark red tiles at center are the 12 disk fields observed in 2009B. The outer ring of 53 fields is the 2009B sky sampling ring. The dashed ellipse marks the M31 disk at $R = 20$ kpc along the major axis. Coordinates are centred on the nucleus of M31 with North up, and East to the left. See the online version for color versions of this and all subsequent figures.

practises in § 2.5. The accuracy of our WIRCam image calibrations are analyzed in § 2.6. In § 2.7 we present our method for recovering the galaxy surface brightness by minimizing image-to-image differences across the mosaic, while in § 2.8 we analyze the results of this algorithm. We estimate the systematic uncertainties in our mosaic solution in § 2.9, where we also compare our technique to the Montage package (Berriman et al. 2008) and the Spitzer/IRAC mosaics. Finally in § 2.10 we summarize the uncertainty of NIR background subtraction on the scale of M31 and outline our ideal observation and reduction method.

Table 2.1: Summary of WIRCam observing programs. N is the number of WIRCam fields covering the M31 disk in each semester (see Figure 2.1). ST Nods give the nodding sequence between sky and target. For example, "$[S^3T^8]^2S^3$" means that a sky field is observed three times, followed by eight target observations. That pattern is repeated a second time and followed by a sequence of three sky observations. T_{int} is the total integration time per disk field while T_{exp} is the integration time per WIRCam exposure. Eff. is the observing efficiency, or percentage of time in a program allocated to integrating the disk of M31, compared to nodding, read out and sky overheads. μ_{bkg} gives the range (min-max) of background surface brightnesses seen in each band. PSF reports the distribution seeing as measured from the full width at half maximum (FWHM) of stellar point spread functions in the uncrowded sky images.

Semester	Band	N	ST Nods	$\frac{T_{\mathrm{int}}}{\mathrm{field}}$ (min)	T_{exp} (s)	Eff. (%)	μ_{bkg} $(\frac{\mathrm{mag}}{\mathrm{arcsec}^2})$	PSF (arcsec) 25th	50th	75th
2007B	J	27	$[S^3T^8]^2S^3$	12.5	47	49	(15.4, 16.7)	0.68	0.75	0.84
	K_s		$[S^5T^{13}]^2S^5$	10.8	25	42	(13.4, 14.2)	0.60	0.65	0.73
2009B	J	12	$[ST^2]^{20}S$	13.3	20	26	(15.0, 16.5)	0.61	0.69	0.83
	K_s						(13.4, 14.3)	0.60	0.66	0.76

2.2 Observations

M31 was observed in the NIR using the WIRCam instrument, mounted on the 3.6-meter Canada-France-Hawaii Telescope (CFHT), at the summit of Mauna Kea in Hawaii over multiple runs between 2007 and 2009. Observations were carried out in the NIR J ($\lambda_0 \sim 1.2\mu$m) and K_s ($\lambda_0 \sim 2.2\mu$m) bands.

WIRCam is an array of four HgCdTe HAWAII-RG2 detectors (Puget et al. 2004). Each detector comprises 2048×2048 pixels, with a scale of $0\farcs3$ pix^{-1}, which critically samples CFHT's typical seeing of $0\farcs65$. WIRCam's detectors are arranged in a 2×2 grid with $45''$ gaps, so that the entire instrument covers $21\farcm5 \times 21\farcm5$ of sky. It is truly the advent of NIR focal plane arrays, like WIRCam, that has enabled wide-field NIR studies of M31.

The ANDROIDS WIRCam survey is designed to simultaneously resolve stars and recover the integrated surface brightness of the entire M31 disk. As discussed in § 2.1, NIR observations require frequent monitoring of the background. Vaduvescu & McCall (2004) found, for example, that the NIR background intensity can vary by 0.5% per minute, yet the low surface brightness of M31's NIR disk at $R = 20$ kpc requires the background level to be constrained to within approximately 0.01% flux accuracy (equivalently, ~ 0.01 mag). With a $190' \times 60'$ optical disk, M31 is much larger than the WIRCam fields of view, and monitoring of the background zeropoint is only possible by periodically pointing the telescope away from M31 towards blank sky through ST nodding. The fundamental compromise of ST nodding observation programs is to balance the cadence of sky sampling with the efficiency of observing the target itself. Although studies such as Vaduvescu & McCall (2004), and references therein, provide good guidelines for NIR background behaviour, no program has attempted to construct a near-IR surface brightness mosaic covering an area as large and faint as M31's disk.

We now have the opportunity to experiment with different ST nodding strategies since observations were taken over the 2007B and 2009B semesters. An objective of this study is to determine how observational design can improve the construction of a wide-field NIR mosaic by comparing the performance of two pre-defined observing strategies.

2.2.1 2007B Semester

The initial survey was carried out in the 2007B semester by the CFHT Queue Service Observing under photometric conditions. Here M31 is covered with 27 contiguous WIRCam fields out to the optical radius where $\mu_V = 23$ mag arcsec^{-2} at $R = 20$ kpc.

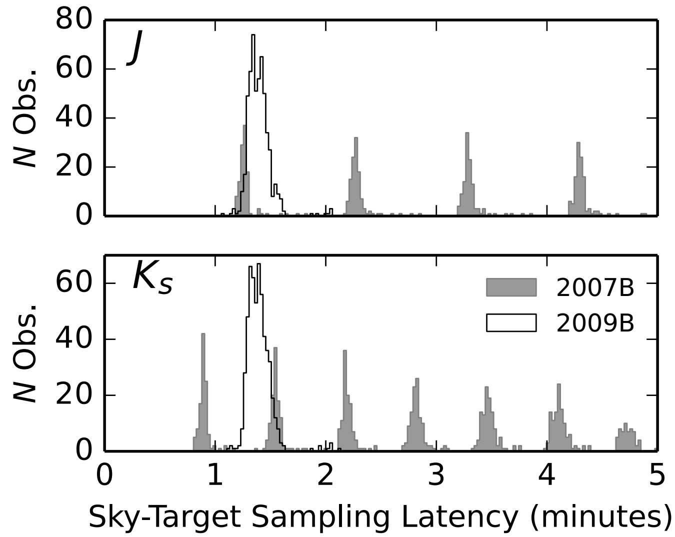


Figure 2.2: Time latency between target observations and sky field sampling in the 2007B and 2009B WIRCam observing runs. The 2009B program was designed to ensure that no disk sample would be removed by more than 1.5 minutes from a sky sample by using a STTS nodding pattern.

The fields are arranged with at least $1'$ overlap in declination, and approximately $5'$ overlap in right ascension. The field configuration is shown in Figure 2.1.

As shown in Table 2.1, each field was integrated for 16×47 s $= 12.5$ minutes in J and 26×25 s $= 10.8$ minutes in K_s. These integrations are sufficiently deep for resolved stellar photometry to reach at least 1 mag below the tip of the red giant branch, a crucial requirement for decomposing the contributions of red giant and AGB stars to the NIR light.

The 2007B ST nodding strategy was motivated by a canonical understanding of NIR background behaviour, since ST nodding background subtraction had never been attempted on this scale before. Our initial sky ST nodding strategy ensured that a sky sample would be no more than 5 minutes removed from an M31 target image, although the mean latencies would be 2–3 minutes. Given the respective

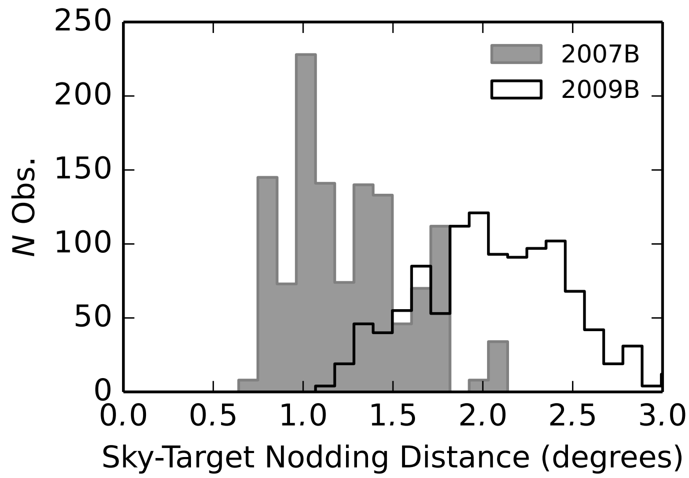


Figure 2.3: Distance between sky and target observations in the 2007B and 2009B WIRCam observing runs. The larger nodding distance of 2009B is a consequence of sky ring sampling. The maximum nodding distance across the sky ring was purposely set to $\sim 3°$ to avoid excessive time overheads (see Figure 2.1). As such, a given disk field only samples roughly half of the full sky ring.

exposure times (chosen so as not to saturate with the background flux), this implied an ST observing sequence of $S^3T^8S^3$ in J and $S^5T^{13}S^5$ in K_s to minimize telescope repositioning overhead.[1] Four sky fields were chosen (Figure 2.1) to avoid bright foreground stars. Each disk field was associated with a single sky field to minimize telescope slewing distance.

2.2.2 2009B Semester

Analysis of the 2007B data revealed that the adopted sky-target nodding strategy was not sufficient for recovering the M31 surface brightnesses due to uncertainties in the background. Repeatedly sampling one of only four sky fields also proved not ideal. This motivated the 2009B observing campaign.

[1]Superscripts here denote the number of times an observation is repeated in sequence for a given target disk field.

Rather than replicate the 28-field footprint of the 2007B campaign, we observed 12 new fields in 2009B (see red boxes in Figure 2.1) that overlap each other and all of the 2007B footprints, to form a network of well-subtracted fields. Thus the 2009B observations augment and calibrate the 2007B NIR mapping.

To improve background subtraction fidelity, we recognized challenges not fully appreciated in the 2007B survey design. Not only does the background change rapidly in time, it possesses a significant spatial structure on the scale of WIRCam fields and larger. This has two ramifications: the background level sampled at a sky field *will not* necessarily reflect the background present at the disk, and the background in each WIRCam frame has a 2D shape, not simply a scalar level.

This resulted in three principal changes to the observing strategy. First, we chose to minimize latency between sky and target observations with a ST^2S pattern. That is, each target observation was directly paired with a sky observation taken within 1.5 minutes (Figure 2.2).

Second, we also increased the number of repetitions so that each field is observed 40 times in each band in a $[ST^2]^{20}S$ pattern. This repetition enables averaging over spatial sky background structures on the scale of WIRCam fields.

Finally, we employ a pseudo-randomized sky-target nodding pattern where no sky field is used repeatedly for a disk field. In order to maintain rapid telescope nods, only northern sky fields serviced the northern disk, and similarly for the southern fields; the maximum offset on the sky was 3° (see Figure 2.3). This non-repetitive sampling of sky fields yielded two possible advantages: 1) when a median background image is constructed, many *background shapes* are combined, possibly yielding an intrinsically flatter background image (see § 2.4.1), and 2) if there is a coherent structure in the

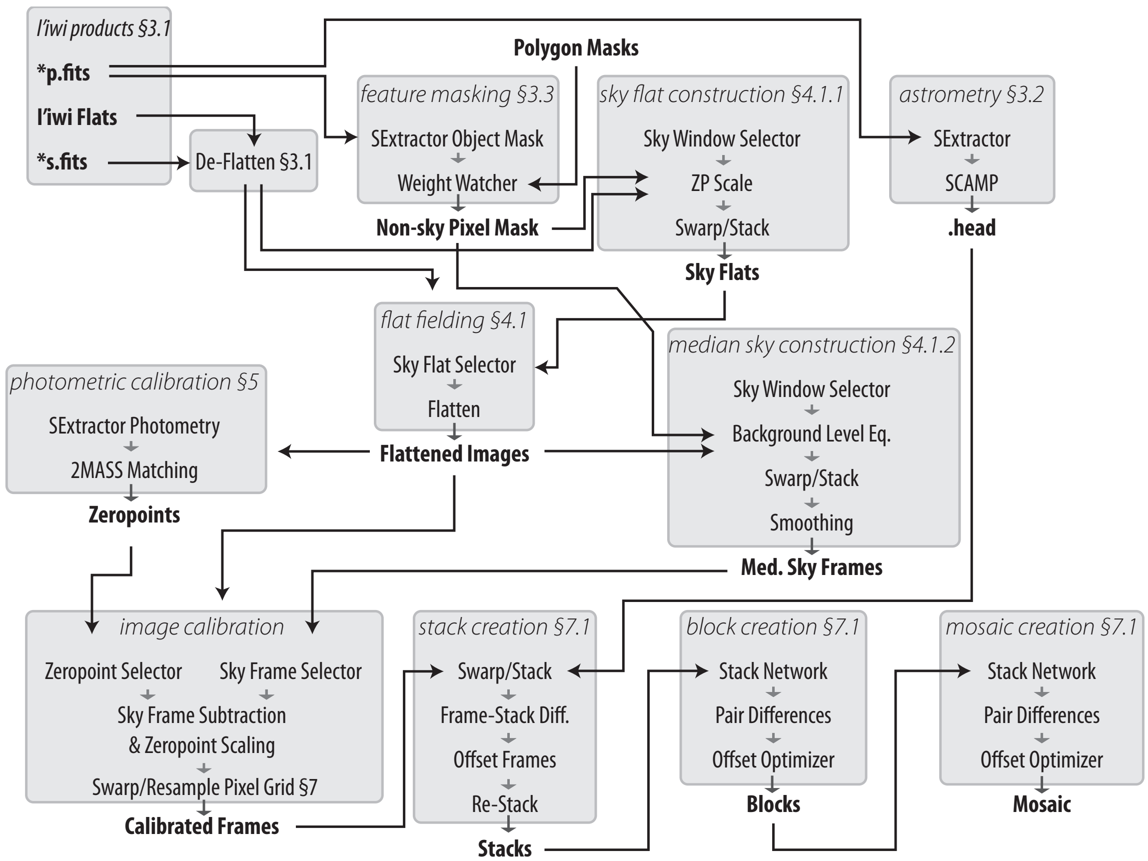


Figure 2.4: Flowchart representation of the ANDROIDS WIRCam pipeline, from receipt of CFHT *'I'iwi* data products to rendering of M31 mosaics.

NIR background, sampling sky fields degrees apart in rapid succession should average out these systematic biases in estimating the background level *on the galaxy disk*. Given these observations, we now consider how to properly construct a wide-field NIR surface brightness mosaic of M31.

2.3 Image Preparation

While CFHT distributes calibrated WIRCam data products, we haven chosen to replace much of their data reduction recipes with our own to optimize and explore the limitations of wide-field NIR surface brightness maps. An overview of the pipeline is shown in Figure 2.4; the principal steps are 1) astrometry, 2) source masking, 3) night sky flat fielding, 4) zeropoint estimation against 2MASS sources, 5) median sky frame construction, 6) image calibration with zeropoints and median sky subtraction and 7) background optimization and mosaic production in three hierarchical steps.

2.3.1 Choice of Starting Point

CFHT offers WIRCam data in two degrees of preprocessing with the *'I'iwi* pipeline: an image that has been corrected for nonlinearity, dark-subtracted and flat-fielded (`*s.fits`); and an image that has been background-subtracted, in addition to all the previous treatments (`*p.fits`). In order to implement our own calibration strategy, our mosaics stem from `*s.fits` products (though we note that `*p.fits` products are still used for astrometry and source masking, see below). Nonetheless, two *'I'iwi* processing stages included in `*s.fits` products must be handled carefully.

Crosstalk correction WIRCam integrations prior to March 2008 (which includes the 2007B data set, not the 2009B data) suffered from electronic crosstalk within the detector. This cross-talk is manifested in repeating rings above and below saturated stars.[2] By default, the *'I'iwi* pipeline removes this crosstalk by subtracting a median of the 32 amplifier slices. Unfortunately, this algorithm fails in cases where the background has a surface brightness gradient (such as on the disk of M31) and produces a brightness gradient that is stronger than the galaxy surface brightness itself. Loic Albert (then at CFHT) kindly re-processed our 2007B data set with the cross-talk correction omitted.

[2]See `http://cfht.hawaii.edu/Instruments/Imaging/WIRCam/WIRCamCrosstalks.html`.

Flat fielding A peculiarity of `*s.fits` images is that even though they are flat fielded using dome flats by CFHT, those products still exhibit strong non-uniformity, dust artifacts and surface defects. We note that CFHT produces dome flats (for each queue run) from median stacks of 15 images taken under a tungsten lamp, subtracted from images of the same integration time taken with the lamp off. This procedure should remove the additive thermal background from the flat, ensuring that the flat field is a purely multiplicative calibration. Despite this, the presence of dust artifacts betrays the fact that dome flats do not reproduce the same illumination pattern as sky photons. Similarly, the presence of surface defects in dome-calibrated images could be caused by disparities in both the optical path and the color of the tungsten lamp versus the night sky background. In this work, we find that WIRCam images can be adequately flat fielded using night sky flats. We give a visual demonstration of the superiority of night sky flat fielding in Figure 2.5: surface defects left by dome flat fielding are removed with night sky flat fielding.

One interesting feature of `*s.fits` images is the appearance of horizontal banding corresponding to the 32 amplifiers that service independent horizontal bands of each WIRCam detector (seen in Figure 2.5). It is odd that a dome flat failed to calibrate such electronic structures in a detector, and one might expect that such banding should be calibrated with an additive correction. Indeed, this banding is absent from fully-processed *ʻIʻiwi* images due to median sky frame *subtraction*. However, we maintain that flat fielding is the correct treatment for these structures since they appear to be proportional to the background throughout the night (which can vary by 10% during a night), yet are still corrected with a single night sky flat.

Our ANDROIDS pipeline thus begins with `*s.fits` data that have been *uncorrected* for dome flat fielding. That is, we multiply the `*s.fits` image with its associated dome flat.[3] The result is an image that retains *'I'iwi*'s prescription for dark subtraction, bad pixel masking and non-linearity correction, ready for our own sky flat-fielding (to be described in § 2.4).

2.3.2 Astrometry

Early in the pipeline we build a unified astrometric frame for our image set using *SCAMP* (Bertin 2006). *SCAMP* matches stars in *Source Extractor* (Bertin & Arnouts 1996) catalogs of each WIRCam frame both internally (to $\sigma_{\rm int} = 0\farcs10$), and against the 2MASS Point Source Catalog (Skrutskie et al. 2006) to a precision of $\sigma_{\rm ref} = 0\farcs15$. By processing all 4286 frames in the ANDROIDS/WIRCam survey simultaneously, *SCAMP* allows an accurate and internally consistent coordinate frame for our mosaic. *SCAMP* handles this data volume gracefully provided we cull the input star catalogs for stars with $S/N > 100$, and by using the `SAME_CRVAL` astrometry assumption that the WIRCam focal plane geometry is stable. Also note that we build our *Source Extractor* catalogs using the fully-processed `*p.fits` *'I'iwi* images since those are adequate for source detection and astrometry. While *SCAMP* is capable of also fitting a photometric solution for each frame, we choose to establish photometric zeropoints later in our pipeline using a combination of background flux observed across the detector array, and bootstrapping against 2MASS sources observed in uncrowded 2MASS images (see § 2.4.1 and § 2.5).

[3]Dome and twilight flats are made available by CFHT, `http://limu.cfht.hawaii.edu:80/detrend/wircam/`.

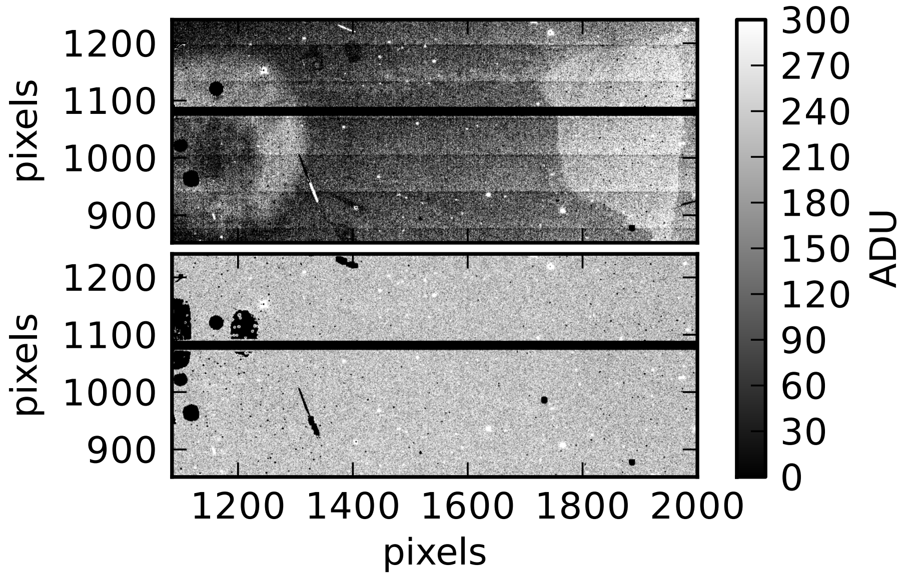


Figure 2.5: Comparison of a WIRCam frame cutout processed with dome flats by the *'I'iwi* pipeline (top), and with sky flats (bottom). Both images are shown in linear counts with identical level ranges. No median background subtraction has been applied. Dome flats leave WIRCam images with dust artifacts (left) and detector surface defects (right). Furthermore, the 64-pixel high horizontal amplifier bands are clearly visible. Simply using sky flats eliminates these artifacts.

2.3.3 Non-Sky Pixel Masking

A second preliminary pipeline stage is source masking. For each sky image we build masks that yield only blank pixels to aid with background level estimation, sky flat construction (§ 2.4) and median sky frame construction (§ 2.4.1). These masks are built by a combination of *Source Extractor* object maps (detected in `*p.fits` images) and hand-drawn polygon regions that cover the diffraction spikes and halos of very bright foreground stars. These masks, along with the *'I'iwi* bad pixel mask, are combined by using *WeightWatcher* (Marmo & Bertin 2008).

2.4 Sky Flat Fielding and Median Sky Subtraction

As we mentioned previously in § 2.3, dome flats fail to properly calibrate dust, amplifier gain and surface defects in WIRCam data (see Figure 2.5). Sky flats are an appropriate alternative, both because of the abundant background photons (any NIR imaging program can use its own images to build sky flats), and because sky flats directly match the illumination path of observations. Sky flats also have the advantage of being contemporaneous with observations: if the WIRCam flat field is variable then sky flats can be built to track such variability. This is a distinct advantage over dome flats, which CFHT builds at the beginning of every WIRCam queue run, or even twilight flats that can only be built once per night.

Despite these advantages, sky flats are built on the assumption that all background illumination in a night sky image is proportional to the flat field function. Several contaminants prevent this from being true: thermal emission from the detector or telescope structures can add a significant background in the K_s band, and scattered light (*e.g.,* off the camera's cold pupil stop) further perturbs the proportionality of flat field images. Ideally one would subtract these contaminants from images before constructing sky flats. Then science images could be flat fielded, and additive contaminants in science images would be automatically removed in subsequent median background subtraction. Note that both dome flats and twilight flat fields can distinguish additive contaminants from the multiplicative flat field function. Dome flats are built from the differences of images taken with the lamp on and off, directly removing any thermal component from the flat field. Twilight flats can also treat additive contamination by capturing images at different levels of sky illumination so that linear fits to each pixel allow any additive bias to be removed. In the case of sky flat fielding,

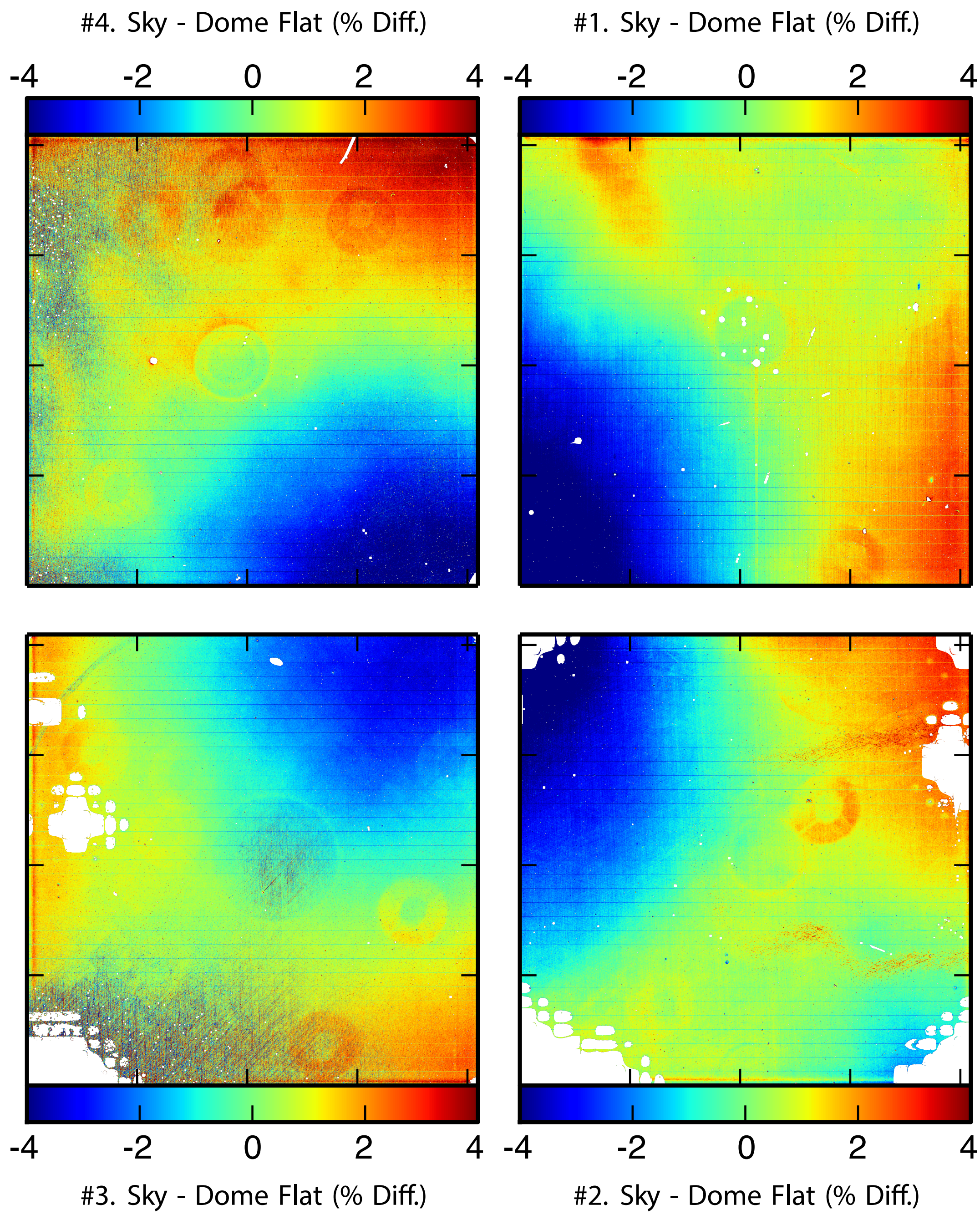


Figure 2.6: Ratio of corresponding K_s band `NIGHT` sky flat and dome sky flats. This ratio provides an approximate upper limit on the systematic uncertainty of flat fielding with WIRCam. Note that the `NIGHT` sky flats have been renormalized to remove the chip-to-chip zeropoint correction described in § 2.4.1.

however, we *cannot* disentangle additive from multiplicative processes in images so that the sky flat field and median background subtraction steps presented here are not in fact separable and independent operations. Figure 2.6 shows a ratio of sky flats (using the `NIGHT` prescription; see below) and dome flats. This ratio effectively sets an approximate upper limit on the systematic accuracy of a WIRCam flat field. Thus we can expect our flats to be correct within a few percent. To proceed, we must accept that our sky flats are not purely proportional calibrations and instead are a first step in our combined flux calibration and background subtraction pipeline. In § 2.6 we analyze the performance of our night sky flat fielding and median background subtraction procedure to find that frame-by-frame surface brightness shape errors are dominated by rapid variations in the background itself.

A second assumption built into sky flat construction is that skyglow is uniform across the detector. Wide-field images of the NIR night sky show that skyglow has rich spatial and temporal variations. However, by marginalizing over a large number of sky images, any illumination bias in the sky can be mitigated. The 2009B observing program even took this marginalization process further by sampling pseudo-random sites on the sky while building sky flats. One degree of control that can be exerted over sky flat construction is the time window that sky images are drawn from. Using a long window, such as the full length of a queue observing run, ensures that the intrinsic WIRCam flat field function is stable over several days, while producing the statistically flattest residual sky illumination pattern. Shorter windows make the opposite assertion that the WIRCam flat field is unstable, and that any bias in sky shapes can be tolerated.

We investigate three sky flat designs in this study, labelled `QRUN`, `NIGHT` and

`FW100K`. `QRUN` flats are built from all sky integrations taken during a queue run, and through a given filter. For the ANDROIDS program, 25–637 (typically $\sim$ 140) sky images, obtained over a sequence of $\leq$ 10 days, are composed into a `QRUN` flat. `NIGHT` flats are made from all sky integrations taken during a single night, through a given filter. Because observations are observed in queue service observing mode, the ensemble of sky images typically sample 0.5–3 hours of a night. Last, we introduce *real-time* `FW100K` sky flats that are rapidly updated throughout the night in case the WIRCam illumination function and gain structure are unstable. `FW100K` flats are designed such that the pool of sky images reaches cumulative background levels of at least 100,000 ADU, or that the time span from first to last sky integration is no longer than two hours. Given the 07B J-band ST nodding pattern, 15 sky integrations are accumulated in 50 minute windows, whereas the more frequent nodding in the 09B campaign shortened this window to 20 minutes (though as long as 50–90 minutes in dark sky conditions). The brighter K_s sky calls for just 7–13 integrations in 07B, or 10–20 integrations in the 09B campaign. This number of K_s sky samples was accumulated within 10–30 minutes in 07B, or 10–70 minutes in 09B.

2.4.1 Implementation of Sky Flat Fielding and Median Background Subtraction

We now describe the technical details of sky flat fielding and median background subtraction steps. Recall from Figure 2.4 that the inputs of flat field construction are 'de-flattened' images that retain the linearity and dark-current subtraction of the *'I'iwi* pipeline.

Sky Flat Construction

According to the type of sky flat being constructed, `QRUN`, `NIGHT` and `FW100K`, ensembles of sky images are formed. Given an ensemble of sky integrations, our next task is to scale the intensity of each image according to three requirements: 1) each image frame in the median stack is at the same level, 2) each WIRCam detector has a unified zeropoint, and 3) the sky flat across the whole array is flux normalized.[4] This scaling is determined by the median pixel level measured on each detector for each sky integration—let us denote these median levels as $\alpha_{i,j}$ for the ith sky image's level in detector j ($j \in \{1, 2, 3, 4\}$). To avoid bias in the background estimate, we mask any pixels that do not sample blank sky (see § 2.3.3).

From the ensemble of images produced by an individual WIRCam detector, we compute the median background level: $\beta_j = \text{median}(\alpha_{1,j}, \alpha_{2,j} \ldots \alpha_{n,j})$. Further, we also compute S, the median of all median detector levels: $S = \text{median}(\beta_1, \beta_2, \beta_3, \beta_4)$. Then each sky image is scaled by the factor $f_{ij} = \beta_j/(\alpha_{ij}S)$. Note that the factor α_{ij}^{-1} normalizes each image to the same level for stacking, while the ratio β_j/S adjusts the level of each detector according to detector-to-detector zeropoint offsets.

The flat itself is built by median combination. Median combination of a stack of hundreds of 2048×2048 pixel images, each with a weightmap masking astronomical sources, is computationally intensive. A convenient solution is to use *Swarp* (an image-mosaicing software package, Bertin et al. 2002) in a mode that combines images pixel-to-pixel. Once the sky flat is built, it is divided from the appropriate science images to produce a flat-fielded data set.

[4]It is also acceptable to establish chip-to-chip zeropoint offsets using differential 2MASS photometry, rather than from background surface brightness. In § 2.6.2 we establish the equivalence of the two methods.

Median Background Subtraction

Since M31 is much larger than individual WIRCam fields, background is subtracted (to first order) using the background levels found in contemporary sky images. In § 2.2, we described the sky-target nodding sequences chosen for the 2007B and 2009B observing campaigns. Although a scalar background level can be estimated from a sky image, and subtracted from the paired target images, it is common to construct a median background image and subtract this from target images.

Independent median background images for each WIRCam detector are produced by choosing a sky image (the primary sky image) and four other sky images taken at adjacent times. Across each image, the median background intensity is recorded. A Source Extractor object mask, as used in § 2.4.1 for flat fielding, removes bias from astrophysical sources. Each sky image is additively scaled to a common intensity level to compensate for background level variations. As described in § 2.4.1, *Swarp* is used to median-combine the sky images with non-sky pixel masks. Since the background has only low-frequency spatial information, these median background images are smoothed with a Gaussian kernel (note this is quite different from the function of median background images applied to dome-flat processed WIRCam data, where median background subtraction also removed pixel-to-pixel artifacts). This median background image is then additively scaled back to the original level of the primary sky image. Finally, to background-subtract a science image, we apply the concurrent median background image.

2.5 Photometric Calibration

Our flat field procedure necessitates a revision of photometric zeropoints. Since our program is observed in short (~ 1 hour) blocks in CFHT's queue service observing, we do not have the necessary airmass baseline to solve for nightly zeropoint and atmospheric extinction terms for each band. Instead, we estimate photometric zeropoints by directly bootstrapping against sources from the Two Micron All Sky Survey (2MASS) Point Source Catalog (PSC; Skrutskie et al. 2006). Although these are not standards, the ensemble of 2MASS stars may be treated as such. Since the disk of M31 is crowded, and 2MASS has low resolution ($1''$ pix^{-1}), we choose to directly estimate zeropoints only in the sky images. We estimate the zeropoints of M31's images from a sliding window average of zeropoints from adjacent sky images (analogous to the median background subtraction procedure, described in § 2.4.1).

Specifically, instrumental photometry of stars in the uncrowded sky fields is obtained with Source Extractor (Bertin & Arnouts 1996). We use the `AUTO` photometry mode to capture the full stellar light without necessitating aperture corrections. 2MASS PSC objects are matched to our Source Extractor detections by position using the author's *Mo'Astro*[5] Python package, that manages the full 2MASS PSC in a MongoDB database. The 2MASS PSC contains many galaxies, and many 2MASS sources are saturated in our deeper WIRCam images. Thus we select sources with $J < 14$ or $K_s < 15$ magnitudes, and FWHM $< 1''$ according to our Source Extractor photometry. Additionally, we select sources with $J - K_s < 0.8$ (typical of foreground Milky Way stars) as we observe larger zeropoint residuals in redder stars. After filtering, ~ 200 matched 2MASS sources remain in typical WIRCam images. Given a

[5] `http://moastro.jonathansick.ca`

joined catalog of 2MASS and instrumental photometry (in ADU) in a specific sky image, we estimate an instrumental zeropoint as the median photometric offset:

$$m_0 = \langle m_{\mathrm{2MASS}} + 2.5\log_{10}(\mathrm{ADU}/T_{\mathrm{exp}})\rangle. \tag{2.1}$$

Our data show no trend in zeropoint versus $J - K_s$ color index. Hence, following practice at CFHT, we do not apply a color transformation between 2MASS and WIRCam bandpasses. Internal testing at CFHT with synthetic photometry indicates that color transformation coefficients may be $A_J = 0.05$ and $A_{K_s} = -0.005$ (K. Thanjuvar, priv. comm.) For typical M31 RGB stars with $J - K_s \sim 1$, this color transformation would be a < 0.1 mag effect.

Given that 2MASS stars in each image have photometric uncertainties $0.05 \lesssim \sigma_{\mathrm{2MASS\ mag}} \lesssim 0.3$, the typical statistical zeropoint uncertainty, σ_{m_0}, is 0.1 mag in a single image. We reduce this random uncertainty to < 0.01 mag by smoothing the zeropoint time series with a sliding window average.

2.6 Analysis of Sky Flat Fielding and Background Subtraction Methods

Near-infrared sky flat fields are fraught with *additive* contaminants from thermal emission and scattered light. Although we regard a perfect near-infrared flat field as unattainable, we can test which flat field prescription (`QRUN`, `NIGHT` or `FW100K`) performs best, and assess whether the final quality of our NIR M31 mosaics are limited by uncertainties from flat fielding or from ST-nodding background subtraction uncertainties.

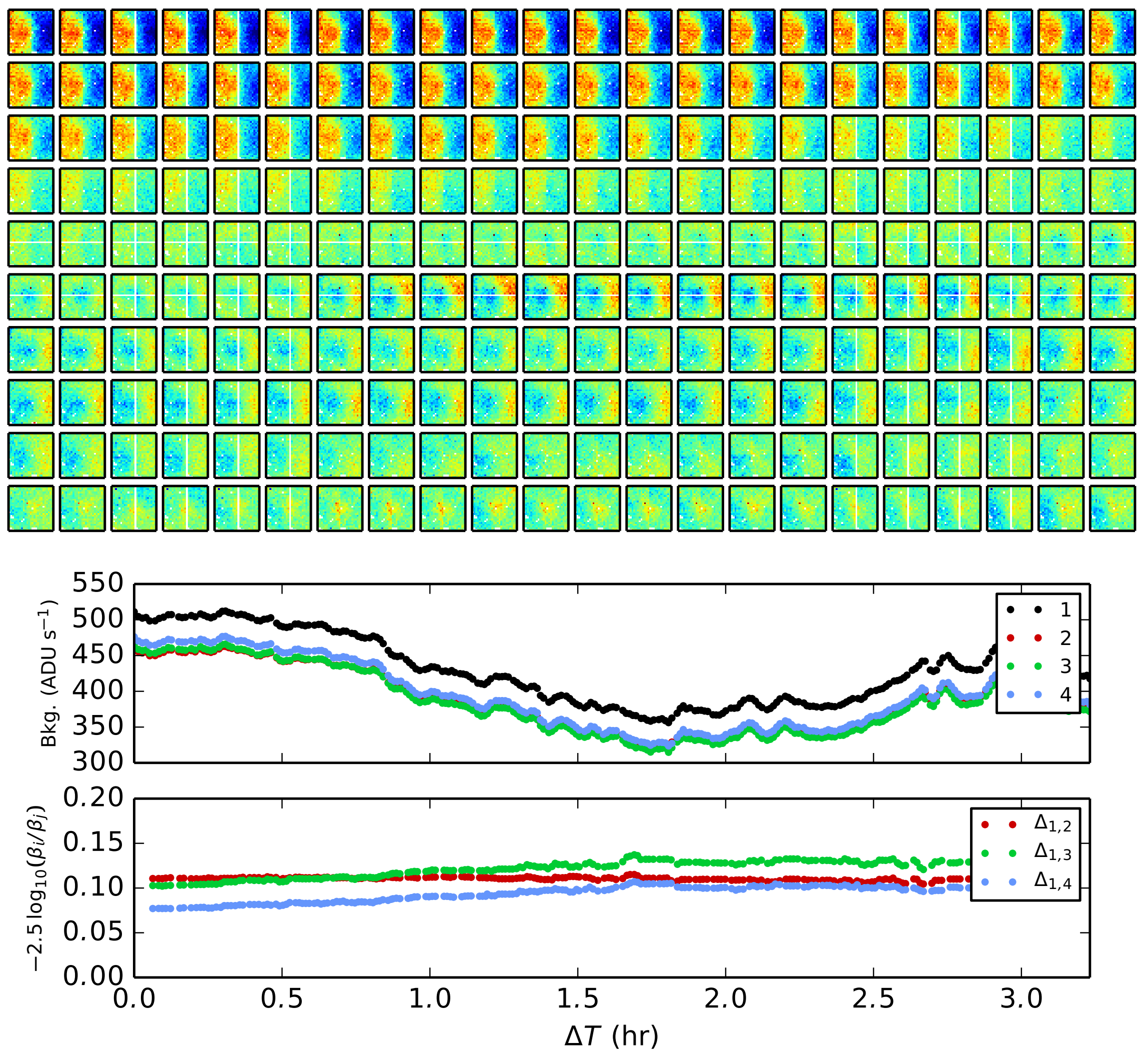


Figure 2.7: Evolution of FW100K K_s-band sky flats over the course of three hours. Percent difference maps of FW100K sky flats relative to the NIGHT sky flat are shown in the upper grid (time evolves left to right, from the top row). Colors in the percent difference maps show ±2% variation. The middle panel shows the background level observed in each detector as a function of time. The bottom panel shows zeropoint differences computed for each FW100K sky flat as a function of time since the first FW100K flat between detector #1 and detectors #2, 3 and 4 respectively.

2.6.1 Evolution of Real Time Sky Flats

A key advantage of sky flats is their close temporal correspondence to the data. Taken to the extreme, our `FW100K` flats are updated with sliding windows of approximately 30 minutes. Here we investigate the nature of evolution in the 'real-time' `FW100K` flats throughout a night. In Figure 2.7 we show the evolution of `FW100K` sky flats relative to a single `NIGHT` sky flat over the course of three hours on a single night. This special sequence of engineering observations consisted of consecutive integrations on sky fields, without nodding to the M31 disk, so that an uninterrupted view of sky flat evolution could be visualized. Over the course of three hours, we see a large-scale shape perturbation move across the detectors from left to right. On the same timescale, the background *level* has changed by as much as 30% (Figure 2.7, middle panel).

Although Figure 2.7 clearly demonstrates that real-time `FW100K` sky flats evolve smoothly, it does not distinguish whether this evolution is driven by proportional effects or by additive contamination such as a thermal background or scattered light. However, we do note that the patterns are similar to those observed in the CFHT-IR camera by Vaduvescu & McCall (2004), who report a thermal background contamination.

An alternative interpretation is that these sky flat deviations are instabilities in the WIRCam detector electronics. The dominant macroscopic electronic features in WIRCam flat fields are the amplifier bands. Each WIRCam detector is divided into 32 horizontal bands (each 64 pixels high) that are read out into independent amplifiers. These amplifiers have gains that result in levels that differ by 10% in flat field images. However, we find that the gain of each amplifier band is stable throughout the night,

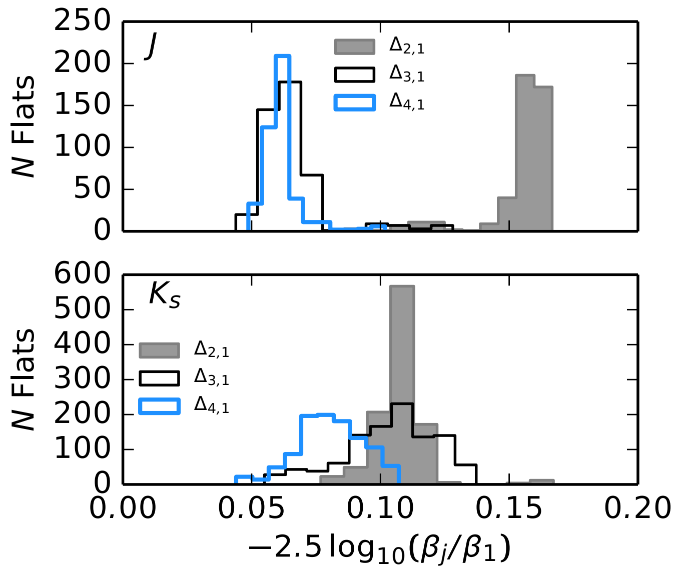


Figure 2.8: Distribution of real-time sky flat scaling factors, measuring detector-to-detector zeropoint differences relative to detector no. 1 (gray: no. 2, black outline: no. 3, blue outline: no. 4) in J and K_s bands.

at a level of $< 0.1\%$ relative to other amplifiers. Thus sky flat evolution is not driven by WIRCam gain instabilities, but by large-scale variations in the flat-field function or an additive background component.

2.6.2 Detector-to-Detector Zeropoint Evolution

We can test if real-time sky flats are tracking evolution in the intrinsic WIRCam flat field function, or merely a background contamination, by examining the detector-to-detector photometric consistency against 2MASS standard photometry. Recall that our sky flats are designed to unify the zeropoints of the four WIRCam detectors by scaling according to the median background levels seen in each detector. Any additive background contamination will introduce detector-to-detector zeropoint offsets. In Figure 2.8 we examine zeropoint differences implied by the median background

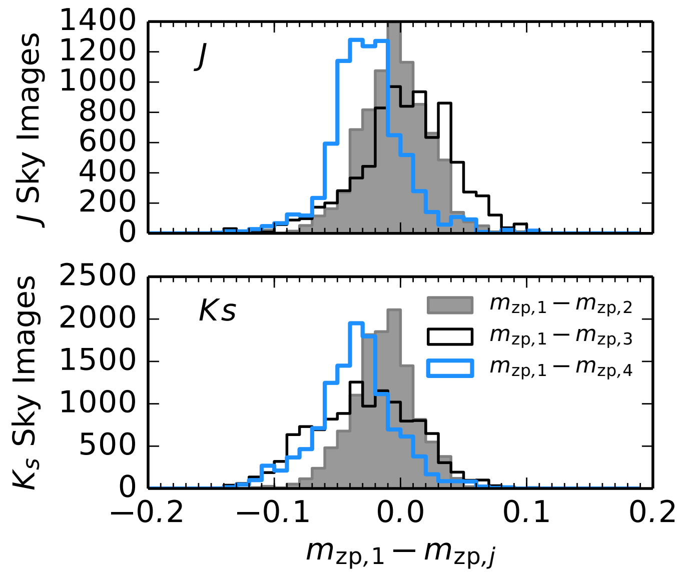


Figure 2.9: Distribution of mean detector-to-detector zeropoint offsets for sky images processed by real-time sky flats. Zeropoint offsets between detectors no. 1 and no. 2, no. 1 and no. 3, and no. 1 and no. 4 are plotted as gray, black outline and blue outline histograms, respectively, for the J band (top) and K_s band (bottom).

values in real-time `FW100K` sky flats, which are computed as $-2.5\log_{10}(\beta_i/\beta_1)$ from the discussion in § 2.4.1. From Figure 2.8 we see that the estimated zeropoint differences between detectors can be variable over a range of 0.05 mag in the K_s-band. This variability is more prominent in K_s-band sky flats than in J-band.

We can test the validity of these zeropoint transformations by verifying the photometric zeropoints of individual detectors against 2MASS stars, as was done in § 2.5. Figure 2.9 shows the distribution of mean detector-to-detector zeropoint offsets observed in images processed by `FW100K` sky flats. We find that zeropoints are consistent within ±0.1 mag, though we detect a small possible systematic bias between detectors #1 and #4 at the level of 0.03 mag. The origin of this zeropoint bias can be seen in the bottom panel of Figure 2.7, which tracks the detector-to-detector zeropoint evolution estimated from each real-time `FW100K` sky flat over the course of 3 hours. The relative zeropoints slowly shift by $\lesssim 0.05$ mag in concert with the evolution in the shapes of `FW100K` sky flats due to time-varying additive contamination (upper panel of Figure 2.7). These results confirm that our sky flats *are* contaminated by a thermal background, albeit at a small level. The systematic photometric bias at the level of 0.03 mag is negligible compared to the photometric uncertainty of individual stars. This result also suggests that our flat fields are significantly better than the $\pm 4\%$ systematic error upper limit established by comparing sky and dome flat fields (Figure 2.6).

2.6.3 Frame Residuals Shapes

We have established (§ 2.6.1–§ 2.6.2) the presence of an additive contamination in WIRCam sky flats that varies over the course of a night and has a slight (< 0.1 mag) influence on photometric calibration. Here we demonstrate how contamination in sky flats influences our observations of M31's surface brightness by examining the residual shapes of individual frames against the median shape of the disk (as assembled in our wide-field mosaic, § 2.7). This also provides a test of the timescale over which the intrinsic WIRCam flat field function is stable. If the residuals of datasets treated by `QRUN` or `NIGHT` sky flats vary systematically with time, in correspondence with the results of § 2.6.1, then the flat field function of WIRCam truly would be variable throughout a night. In this case, `FW100K` sky flats should be most appropriate. This effect should be exacerbated in signal- (not sky-) dominated fields as flat field errors grow in proportion to signal strength.

Our 2009B observations of the field M31-37 in the K_s band are ideal for this experiment: a single detector in that field covers the core of M31, and observations were taken in two blocks, covering a total window of 2 hours (most blocks for this program are observed by the CFHT queue in a half hour). Both the high surface brightness and wide time baseline of this field should highlight flat field bias and variation. In Figure 2.10 we show the residual shapes of individual WIRCam frames against the median shape of the mosaic, given `FW100K`, `NIGHT` and `QRUN` sky flattening of K_s-band images at the M31-37 field. To analyze the shapes of these difference images we marginalize along their rows (left side of Figure 2.10), and columns (right side of Figure 2.10). Note that these marginalizations are done for each detector in the 2×2 WIRCam array; the core of M31 resides in detector #2 (lower-right). In that high surface brightness region, there are strong surface brightness residuals that clearly point out flaws in the flat field itself.

Comparing panels in Figure 2.10, we see that both `FW100K` and `NIGHT` sky flats have similar performance, where frames vary in surface brightness by $\pm 0.5\%$ at the core of M31. The exception are `QRUN`-treated frames that show an evolution on the order of $\pm 1\%$ of the K_s-band sky brightness over a similar time scale as indicated in Figure 2.7. This indicates that `QRUN` sky flats, which are built over several nights of data, are unsuitable for capturing the WIRCam flat field function. While this indicates that the intrinsic WIRCam flat field function varies detectably from night-to-night, the performance equivalence of `FW100K` and `NIGHT` sky flats indicates that the WIRCam flat *does not* vary throughout the night.

It is useful to contrast the frame shape residuals seen in detector #2 with those in other detectors, where the disk surface brightness is lower. There, `FW100K` (Figure 2.10a), `NIGHT` (Figure 2.10b) and `QRUN` (Figure 2.10c) show similar residual distributions, on the order of $\lesssim 0.2\%$ of the NIR background brightness. Further, the results are not monotonically varying in time, as they are in detector #2, and indeed appear to vary essentially randomly. We interpret this behaviour as being caused by random additive background processes, distinct from flat field biases that are proportional to surface brightness. We extend this analysis across the entire data set below.

2.6.4 Distributions of Frame Shape Residuals

Figure 2.11 shows the distribution of frame-block residual shape amplitudes, measured at the 95% difference interval to quantify the reliability of recovering surface brightness shapes in individual WIRCam frames. As in our test of median sky frame flatness (§ 2.6.5), we see that the consistency of frame shapes is $\sim 0.3\%$ of the background level. This result is seen universally amongst the `QRUN`, `NIGHT` and `FW100K` sky flat pipelines and for 2007B and 2009B observing schemes, agreeing with our observation in § 2.6.3 that in background-dominated regimes, frame shape consistency is *not* correlated with flat field bias. Rather, we interpret Figure 2.11 as measuring the amplitudes of *additive* stochastic background shapes originating either from the sky, or associated with the instrumentation itself. Effectively, Figure 2.11 illustrates the *flatness limit* of WIRCam frames observed with large sky-target nods, sky flat fielding, and median sky subtraction.

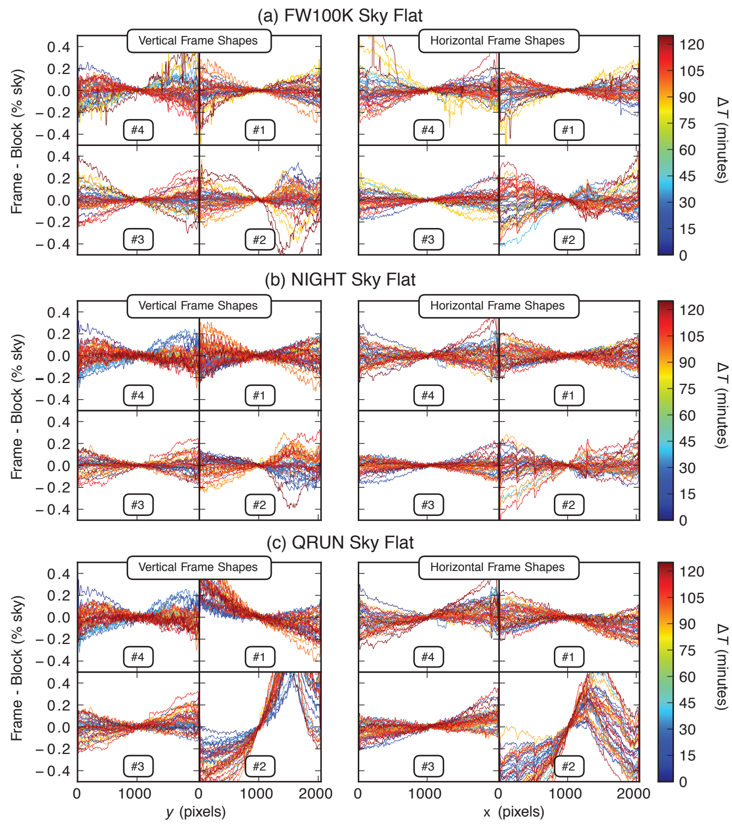


Figure 2.10: Comparison of frame shape residual variations for images processed with (a) `FW100K`, (b) `NIGHT` and (c) `QRUN` sky flats. Residuals are the difference of an individual frame to the median (mosaic) shape. Images are from the M31-37 field, K_s-band, observed in the 2009B semester. Residuals have been marginalized across the x (left) and y (right) axes to provide 1D views. Axes match WIRCam's 2×2 detector footprint. Individual integrations are colored by their time after the first disk integration. The centre of M31 is located in the lower-right detector (#2); surface brightness bias in these regions betray the presence of flat field bias. Lower surface brightness regions are dominated by shape variations on the order of $\pm 0.2\%$ of background level, interpreted as additive uncertainties associated either with the detector, skyglow, or both. `NIGHT` sky flats reliably capture the disk shape in the signal-dominated detector #2 as well, if not better, than `FW100K` flats. Hence the WIRCam flat field function is stable over a night. `QRUN` sky flats introduce large biases in the bulge-dominated surface brightness in detector #2 (lower-right). In the more sky-dominated regions of the image (detector #4), `QRUN` sky flats produce images with similar stability to `FW100K` and `NIGHT` sky flats, indicating that the limit of additive uncertainties associated with sky or instrumental background variations is reached here.

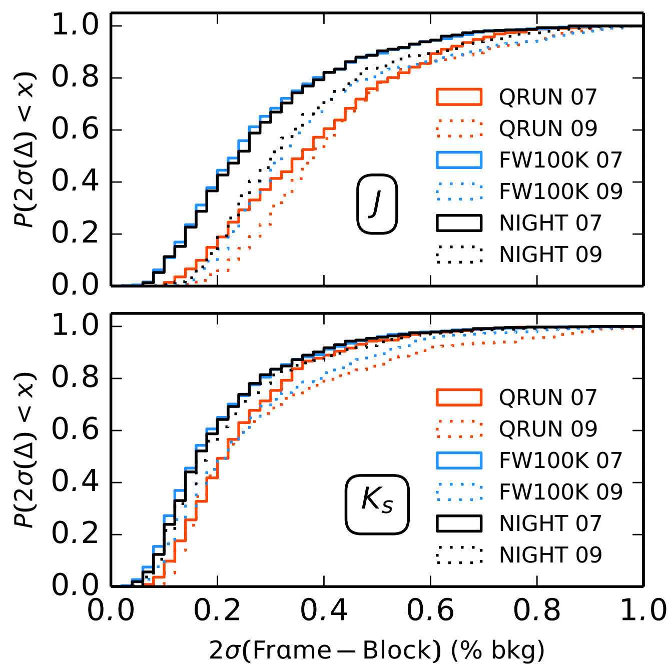


Figure 2.11: Cumulative distributions of scalar difference amplitudes between individual frames and blocks in the J (top) and K_s (bottom) mosaics, measured as a dispersion of pixel differences at the 95% level. Whether processed with `QRUN` (orange), `NIGHT` (black) or `FW100K` (blue) sky flats, or observed in 2007B (solid lines) or 2009B (dotted lines), the residual amplitude differences between frames and blocks are similarly distributed. The mean amplitude difference is 0.3% of the J background brightness (0.2% in K_s).

2.6.5 Shapes of Median Background Frames

Another test of sky flats is their ability to produce an unbiased sky background, up to the level of intrinsic sky variations. This test can be made by examining the amplitudes of median background frames (§ 2.4.1) produced by `QRUN`, `NIGHT` and `FW100K`-processed data sets. We measure the amplitude of shapes across the $10' \times 10'$ WIRCam frame as the 2-standard deviation interval (95%) of each median sky image's pixel distribution: 2σ(med sky). In Figure 2.12 the cumulative distribution functions of the background shape amplitudes are presented for each set of flat fielded data, in each band and for each semester.

`QRUN` sky flats, which are known to be incorrect (§ 2.6.3), produce median backgrounds with the largest amplitudes—as much as 3% of the background. Compared to the other flats, the `QRUN` sky flats are biasing the background shape, relying upon the median background subtraction process to effectively flatten the sky background. `FW100K` sky flats have the opposite effect. The temporal windows from which `FW100K` sky flats are constructed have nearly the same span as those for median background images, so the background amplitude is unsurprisingly (and unrealistically) flat. Thus `FW100K` sky flats are incorrect as they divide all structure in images, be they multiplicative or additive in origin. `NIGHT` flats produce median backgrounds with moderate, realistic amplitudes.

A surprising result from Figure 2.12 is that backgrounds in 2009B images are not flatter than in 2007B. Recall that median sky images are composed of five sky frames taken closest to a disk frame. In 2007B, all sky frames were sampled from the same coordinate on the sky and span a 12 minute window covering sky integrations taken before and after a disk image (for both J and K_s sky-target nods). In 2009B, sky frames were sampled from randomly chosen sites along the sky field ring (Figure 2.1) with a window typically spanning 15 minutes. Thus both 2007B and 2009B median sky images span similar time windows, although the 2009B strategy attempts to marginalize over five distinct sites on the sky (and thus sky background shapes) while 2007B median sky images do not. From Figure 2.12 we conclude that this marginalization does not effectively occur, implying that the background shapes sampled from distinct sites on the sky *are* correlated. Since wide-field movies of the NIR sky (Adams & Skrutskie 1996) suggest that this should not be true, much of the

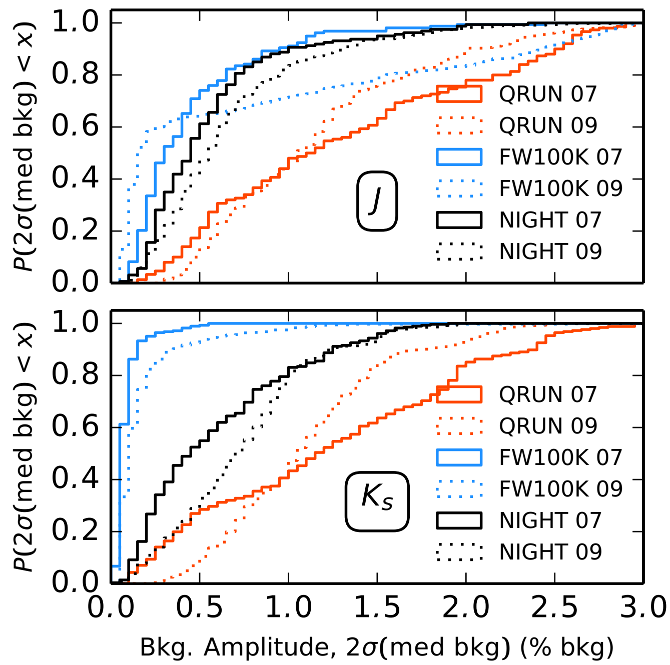


Figure 2.12: Cumulative distribution function of background level amplitudes across median sky images processed with `QRUN` (orange), `NIGHT` (black) and `FW100K` (blue) sky flats for the J (top) and K_s (bottom) bands. Real-time (`FW100K`) sky flats produce much flatter median sky frames, with mean shape amplitudes of 0.3% (J) to 0.1% (K_s) of the NIR sky level, while the mean amplitude of `QRUN`-flat processed sky images is 1.5% of the background level, and as high as 3% of the background level. `NIGHT` flat-processed data generate mean shape amplitudes of $\sim 0.6\%$ of the background level. We interpret this as `QRUN` flats introducing a bias in background shapes, while `FW100K` flats create unrealistically flat backgrounds by dividing skyglow structure that should be left for background subtraction. `NIGHT` flats have realistic background amplitudes that are similar in both bandpasses.

structure in the median background images is an instrumental background. Similar backgrounds are seen by Vaduvescu & McCall (2004) in the CFHT-IR camera.

2.6.6 Section Summary

In this section we have analyzed the performance of our flat field, median background subtraction and photometric calibration procedures outlined in Figure 2.4. Here we summarize our findings on the accuracy of surface brightness shapes reproduced by WIRCam in a sky-target nodding observing program.

Night sky flat fields are preferable to dome flats because sky flats properly calibrate dust, detector surface, and gain structures thanks to proper illumination of the WIRCam entrance pupil. Such flats can be easily made from programme data. Specifically, we advocate `NIGHT` flats, which are built from all sky images taken over a single night, because they capture night-to-night variations in the WIRCam flat field not captured by `QRUN` flats (compare Figure 2.10b and Figure 2.10c).

Although we confirm that night sky flats are afflicted by an *additive* contamination (*i.e.,* thermal background, or scattered light) that varies on sub-hour timescales (Figure 2.7), this contamination impacts the photometric zeropoint at a level of 0.03 mag. This *does not* impact the overall surface brightness fidelity of our mosaic. Since the ANDROIDS/WIRCam mosaic of M31 is almost entirely background dominated, we are principally limited by background subtraction, which is non-trivial since the background cannot be directly measured on the M31 disk given our current field of view. Ultimately we find that the *shape* of the background on the disk can be known to within 0.3% of the NIR background levels (Figure 2.11). By comparison, the typical amplitude of median sky images is $\sim 0.6\%$ of both the J and K_s background levels. Note that it is impossible for a NIR observing program with large ST nods to subtract background shapes better than the 0.3% we find here: the shape of the background at the target will always be distinct from background shapes measured at designated sky fields.

To summarize the first half of this chapter, we have demonstrated that `NIGHT` sky flats are appropriate for our application, and the impact of thermal background and scattered light on flat fields and zeropoints are negligible compared to background uncertainties due to NIR skyglow. In the following sections we assemble NIR mosaics of M31 using images processed according to the preceding sections, using `NIGHT` sky flats, and show that the single most important calibration for wide-field NIR mosaicing is sky offset optimization.

2.7 Sky Offset Optimization

In Figure 2.13a, we plot mosaics (assembled using *Swarp*) from ANDROIDS frames processed with the pipeline discussed in § 2.3–§ 2.5 and adopting the `NIGHT` sky flat prescription favored in § 2.6.6. Although these image preparations can constrain the surface brightness *shape* of a WIRCam frame to $\lesssim 0.3\%$ of the sky level, Figure 2.13a demonstrates that the true level of the background on M31's disk is lost by the temporal and spatial variation of skyglow between disk and sky field observations. This results in the distinct field-to-field surface brightness discontinuities that are seen in Figure 2.13a. However, we can use the constraint that all overlapping pairs of images composed in our mosaic should have equal surface brightnesses in their intersections. To enforce this constraint, we solve for a *sky offset* for each image: a small scalar nudge of intensity that can be added or subtracted from each image so that all images in the mosaic have continuous surface brightnesses. Note that we use the term "sky offset" for these intensity adjustments, but in practice these offsets are agnostic of the cause of background subtraction error that they correct. Since our mosaics are made from many inter-connected images (3924 J and 4972 K_s image frames), our optimization of sky offsets can provide powerful constraints on the true level of the background at the M31 disk.

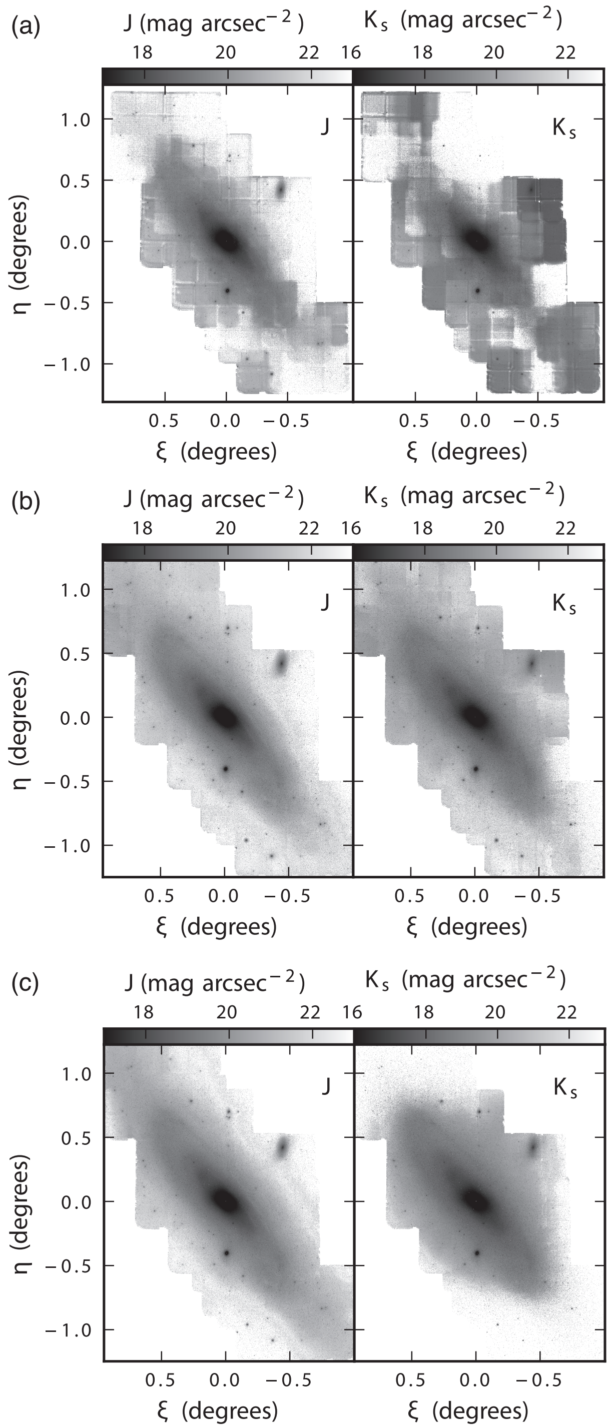


Figure 2.13: ANDROIDS/WIRCam mosaics of M31 in J (left panels) and K_s (right panels). (a) Median background-subtracted mosaics processed according to § 2.3–§ 2.5 using `NIGHT` sky flats, (b) mosaics after scalar sky offsets are applied, as described in § 2.7 and (c) mosaics generated by *Montage* using planar sky offsets.

Montage is a FITS mosaicing package (Berriman et al. 2008) originally written for the 2MASS survey that includes sky offset estimation (background rectification, in their terminology) functionality. *Montage* can solve sky offsets either as scalar levels, or as planes. Sky offsets are then chosen iteratively by looping through each image pair and choosing the offset needed to minimize the difference image of that pair, counting previous sky offset estimates. Sky offsets are refined over several loops through the entire set of overlapping image pairs until convergence is reached (that is, once incremental adjustments to sky offsets diminish below a user-specified threshold). Although this iterative implementation of sky offset optimization is elegant, its accuracy has never been formally analyzed in literature, to our knowledge. In particular, we are interested in the robustness of *Montage* sky offsets against local minima in the N-dimensional solution space of sky offsets, given a mosaic of N independent images. Further, the optimization is slow, given the several thousand frames in our mosaics. Thus we decided to implement our own sky offset algorithm, although a comparison to the *Montage* solution is given in § 2.9.

2.7.1 Sky Offset Implementation

Our sky offset algorithm is based on two features that distinguish it from the *Montage* implementation. First, the optimization is carried out in three hierarchical stages to accommodate the large number of images. Second, we use a downhill simplex algorithm (Nelder & Mead 1965, hereafter, NM) with re-convergence checks rather than the iterative approach of the *Montage* sky offset solver. We begin with `NIGHT`-sky-flat-calibrated, median-sky-subtracted, and photometrically calibrated image sets that are resampled using *Swarp* to a common pixel in an Aitoff equal-area projection with the native WIRCam pixel scale of $0\farcs3$ pix^{-1}.

We address the sky offset optimization hierarchically by considering the geometry of the WIRCam detectors in a 2×2 grid and the arrangement of 39 WIRCam fields on the M31 disk (See Figure 2.1). The first stage of optimization is to stack all *detector frames* (images taken with a given detector, at a given field) into *stacks*. The offsets applied to WIRCam frames to build stacks are labelled Δ_F. Next, we solve for the offsets Δ_S to ensure surface brightness continuity across the four stacks in a WIRCam field. We call the combined unit of four stacks a *block*. The last stage of optimization solves for the offsets Δ_B applied to each block to ensure surface brightness continuity across the mosaic. The net scalar sky offset applied to each frame is thus

$$\Delta_\Sigma = \Delta_F + \Delta_S + \Delta_B. \tag{2.2}$$

After each stage of optimization, the scalar sky offsets are added to the WIRCam frames, stacks, and blocks, and Swarp is used to median-coadd the images to generate stacks, blocks and a mosaic, respectively. Using this hierarchical scheme ensures that, at worst, the number of dimensions in our optimizations is 39 as opposed to the number of WIRCam frames (a factor 10^2 reduction).

We use two algorithms for solving sky offsets. Solving Δ_F offsets in the first stage is trivial since all frames simultaneously overlap. Thus it is sufficient to simply compute a mean surface brightness across all frames, and directly compute offsets (Δ_F) between between the levels of each frame and the mean level. In the last two stages, stacks and blocks, respectively, are arranged in networks of overlapping pairs. For this case we introduce our NM simplex-based offset optimization algorithm.

We identify overlaps between images in a brute-force fashion according to their frames in the mosaic pixel space, defined by the `CRPIX`, `NAXIS1` and `NAXIS2` header values of the resampled images. For each overlapping image pair, we compute a difference image, and ultimately a median difference, $\langle \boldsymbol{I}_i - \boldsymbol{I}_j \rangle$. While computing the median difference, we mask bad pixels using weight maps (propagated by *Swarp*) and expand this mask with sigma clipping. Along with a difference estimate, we also record the area A_{ij} of unmasked pixels in the overlap, and the standard deviation of the difference, σ_{ij}.

We can estimate the optimal set of scalar sky offsets, Δ_i, for each image i by minimizing the objective function:

$$\mathcal{F}(\Delta_1, \ldots, \Delta_n) = \sum_{i,j} \mathcal{W}_{ij} \left(\langle \boldsymbol{I}_i - \boldsymbol{I}_j \rangle - \Delta_i + \Delta_j \right)^2 . \quad (2.3)$$

Note that each coupled image pair is its own term in the objective summation, and that there are as many degrees of freedom (Δ_i) as there are images in the mosaic. Each coupling is tempered by a weighting term $\mathcal{W}_{ij}$:

$$\mathcal{W}_{ij} = \frac{A_{ij}}{\sigma_{ij}}, \quad (2.4)$$

so that more priority is given to couplings of larger areas (A_{ij}), and small standard deviations of their difference images (σ_{ij}).

The objective function in Eq. 2.3 puts no constraint on the net sky offset: $\sum \Delta_i$. Assuming that background subtraction errors are normally distributed, and not biased, sky subtraction offsets should not add a net amount of flux to the mosaic. Fortunately, it is possible to impose this constraint *post facto* by subtracting the mean offset from the sky offsets:

$$\Delta_i^* = \Delta_i - n^{-1} \sum_{j=1}^{n} \Delta_j. \tag{2.5}$$

In the limit that sky offsets Δ_i are drawn from a Gaussian distribution, with standard deviation σ_Δ, the absolute brightness of the whole mosaic will be uncertain by $\sigma_\Delta / \sqrt{N_{\mathrm{images}}}$. The consequences of this uncertainty are revisited in § 2.9.

Given the image coupling records, we optimize the set of Δ_i by applying the object function (Eq. 2.3) to the NM downhill simplex algorithm. The NM algorithm is naturally multi-dimensional and does not require knowledge of the gradient of the objective function. Instead, the NM algorithm operates by constructing a geometric simplex of $N + 1$ dimensions that samples the sky offset parameter space. By evaluating the objective function at each vertex of the simplex, the NM algorithm adapts the simplex shape to ultimately contract upon a minimum.

The NM algorithm will converge into any local minimum without necessarily seeking the global minimum of the objective function. We resolve this issue with two methods: ensuring re-convergence, and sampling different starting conditions.

The practice of ensuring reconvergence in a downhill optimization is suggested by Press et al. (2007). Upon each convergence, the optimal point in the simplex, $\boldsymbol{p}$, is recorded. A new simplex is then generated where one vertex is $\boldsymbol{p}$, and the rest are $\boldsymbol{p} + \boldsymbol{\delta}$ where $\boldsymbol{\delta}$ is a normal random variable of mean zero, and standard deviation $\sigma_{\mathrm{restart}}$. That is, the simplex of the restart retains one vertex upon the previously found minimum, while the other vertices surround that minimum. We set $\sigma_{\mathrm{restart}}$ to $2\times$ the dispersion of image-to-image differences. Our optimization iterative converges and re-converges simplexes until the same minimum is consecutively arrived upon, indicating that the NM algorithm has arrived upon a robust solution. Our sky offset optimizations for 39 blocks typically require ~ 1000 restarts before converging definitively.

Besides ensuring reconvergence, we also start several independent NM simplex optimizations from random starting points in parameter space to seek a globally optimal sky offset solution. We find that $N_s = 50$, and possibly fewer, starts are quite sufficient for an optimization with 39 sky offset parameters (such as the fitting of Δ_B block offsets in the mosaic). For each start, an initial simplex is generated randomly. Since each point in the N by $N+1$ simplex is a suggested sky offset for a given field, each offset is randomly sampled from a normal distribution whose dispersion is $3\times$ the standard deviation of image-to-image differences to ensure the parameter space is well covered. Note that each simplex start and series of subsequent restarts can be performed in parallel. Once all simplex runs are complete, the set of sky offsets belonging to the run that yielded the smallest value of the objective function is adopted.

2.8 Analysis of Scalar Sky Offsets

Figure 2.13b presents the fruits of our WIRCam pipeline and sky offset optimization. Compared to our mosaics without sky offsets, Figure 2.13a, the sky offset optimization is clearly essential for assembling wide-field NIR mosaics. These mosaics are not yet perfect; field-to-field discontinuities at a level of 0.05% of background remain, and large-scale background residuals perturb the outer M31 disk.

2.8.1 Amplitudes of Sky Offsets

The distribution of scalar sky offsets provides an excellent characterization of background subtraction uncertainties when using sky-target nodding. Recall that sky offsets are optimized hierarchically: WIRCam frames are fitted to stacks, stacks are fitted into blocks of four contemporaneously-observed WIRCam detector fields, and these blocks are fitted into a mosaic. Table 2.2 lists the standard deviations of these offset distributions with respect to the typical background level observed in the J and K_s bands.

Note that the sky offsets, as a percentage of background level, are comparable in the J and K_s bands, despite the background being $\sim 4\times$ brighter in K_s than J. This indicates that spatio-temporal variations in the NIR background are monochromatic.

Within the hierarchy of background fitting, simply fitting frames to a stack (with Δ_F) is a correction on the order of 2% of the background intensity. Fitting blocks into a mosaic (Δ_B) is a further $\sim 1\%$ correction. Overall, the temporal and spatial lags of sky-target nodding induce a 2% uncertainty in the background level at the target (see Table 2.2). It is this level of uncertainty that sky offset optimization must diminish to transform uncorrected mosaics (Figure 2.13a) into ones that reproduce the disk with fidelity (Figure 2.13b).

Note that offsets to fit a stack into a block (Δ_S) of four detector field stacks are smallest: 0.1% of the background level. This suggests that on the scale of the 2×2 WIRCam array, the contemporaneously observed detector frames are subjected to nearly identical biases in background. Stack offsets, then, arise from uncertainties in the pipeline's measurement of the background level from single frames in two stages: estimating detector-to-detector zeropoint offsets from frame background levels

Table 2.2: Hierarchy of scalar sky offsets (using `NIGHT` flat fielding, and median background subtraction). Each level of sky offset is defined in Eq. 2.2. $\langle I_{\rm bkg}\rangle$ is taken as the instantaneous background level for the images being sampled.

Offset Type	Sem.	J $\frac{\sigma_\Delta}{\langle I_{\rm bkg}\rangle}$ (%)	K_s $\frac{\sigma_\Delta}{\langle I_{\rm bkg}\rangle}$ (%)
Δ_F	07B	2.54	2.29
	09B	1.88	1.87
Δ_S	07B	0.08	0.05
	09B	0.05	0.03
Δ_B	07B	1.25	0.94
	09B	0.70	1.14
Δ_Σ	07B	2.73	2.44
	09B	1.98	1.88

(§ 2.4.1) and again when subtracting a median background frame (§ 2.4.1). Indeed, in § 2.6 we showed that median background images have shape amplitudes of 0.3% of the background level and that individual frames have surface brightness shapes that are uncertain at a level of 0.2%; Δ_S sky offsets are thus a consequence of the limited surface brightness flatness across a WIRCam frame.

2.8.2 Acceptability of Sky Offsets

Recall that scalar sky offsets were initially introduced as intensity increments to overcome uncertainty in the background level of detector field stacks. For sky offsets to be considered acceptable, we demand that the offsets applied to blocks, Δ_B, be consistent with the background level uncertainty of the blocks themselves. We can conservatively measure the background uncertainty as the dispersion of Δ_F frame offsets in a stack: σ_{Δ_F}. If sky offsets fitted between blocks are statistically permissible, then $\Delta_B \lesssim \sigma_{\Delta_F}$. In Figure 2.14, we plot field maps (in the same spatial configuration

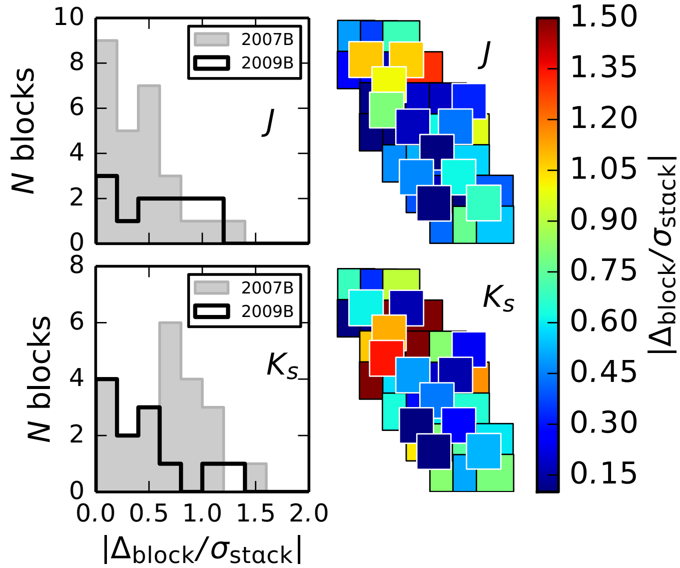


Figure 2.14: Acceptability of J and K_s scalar sky offsets between blocks, as measured by the ratio of $\Delta_B/\sigma_{\Delta_F}$, plotted as histograms and field maps. Shaded and black-outlined histograms distinguish blocks observed in 2007B and 2009B, respectively. Scalar sky offsets required for blocks are consistent with the background level uncertainties of single frames, given sky-target nodding background subtraction.

as Figure 2.1) painted with the values of $\Delta_B/\sigma_{\Delta_F}$ for each block in the J and K_s mosaics. The sky offsets are indeed distributed within the uncertainty budgeted by σ_{Δ_F}: the sky offsets are statistically acceptable.

One can also see the veracity of these sky offsets by plotting a time series of both directly measured background levels and background levels interpolated on disk observations via sky offsets. In such plots we see remarkable continuities of the background level as estimated from sky offsets. Through the sky target nodding and sky offset optimization, we *have* effectively measured the background level on M31.

2.8.3 Residual Image Level Differences

Although scalar sky offsets are statistically valid, they are imperfect prescriptions against the background subtraction uncertainties of each image stack—that much is visually true. We quantify the limited effectiveness of sky offset fitting as the image differences between coupled blocks i and j, $(\boldsymbol{I}_i - \Delta_{B,i}) - (\boldsymbol{I}_j - \Delta_{B,j})$, after the sky offsets $\Delta_{B,i}$ have been optimally fitted to each block. Table 2.3 lists distributions of both image level differences before and after the application of scalar sky offsets. Uncorrected, the ensemble of coupled blocks have a mean intensity difference of $\sim 1\%$ of the typical background intensity. Scalar sky offsets decrease the differences between overlapping fields to $\sim 0.2\%$.

Figure 2.15a shows the block-to-block residual differences as a fraction of the local surface brightness. Note that throughout the bright inner disk of M31, block-to-block residuals are negligible compared to the disk signal; at the mosaic periphery ($R \sim 20$ kpc), field-to-field residuals become comparable to, or greater than, the disk surface brightness. The poor fit is driven primarily by diminishing disk signal, rather than poor convergence of sky offsets. This can be seen by plotting the magnitude of block-to-block residuals (in units of background brightness) in Figure 2.15b. There, significant residuals are distributed throughout the disk, rather than the low-SB periphery of the mosaic.

The inability of scalar sky offset optimization to eliminate residual image differences should not be interpreted as a failure to detect the global minimum; the sky offset optimization algorithm (§ 2.7.1) appears robust in yielding this offset solution set. Evidence of this can be seen in Figure 2.15c, where block-to-block network connections are colored by the ratio of the residual block-to-block intensity difference to

Table 2.3: Coupled block intensity differences and residual intensity differences after application of scalar sky offsets: 25th, 50th and 75th percentiles of distribution. Differences are presented as a percent of the mean background level seen by observations in each band.

	Coupled Block $\langle I_i - I_j \rangle / \langle I_{\mathrm{bkg}} \rangle$ (%)		
	25th	50th	75th
J, uncorrected	0.49	0.91	1.76
J, scalar offset	0.05	0.10	0.17
K_s, uncorrected	0.44	0.91	1.41
K_s, scalar offset	0.02	0.04	0.08

the uncertainty in the block-to-block difference image. The sky offsets solved by the NM simplex algorithm are within the uncertainties of the difference images themselves; better scalar sky offsets *cannot* be made with the WIRCam blocks that our pipeline has produced. Our ability to produce a continuous NIR mosaic is fundamentally limited by our ability to the subtract the true background shape seen at the M31 disk. As described in § 2.6, sky-target nodding on the scale of M31 introduces an intrinsic shape uncertainty of 0.3% of the NIR background levels.

2.8.4 The Growth of Sky Offsets in Time and Space and Effectiveness of the 2009B Strategy

In § 2.8.1–§ 2.8.3, we established the usefulness of sky offsets. We now address the value of the 2007B and 2009B observing programs, with the sky-target nodding observing strategies (introduced in § 2.2). A key question is whether the 2009B strategy to minimize background level uncertainty by minimizing the latency between sky and target samples to ∼1.2 minutes is justified. Overall the 2009B semester reduced sky offsets by 30% (Table 2.2), at the cost of a roughly 50% reduction in

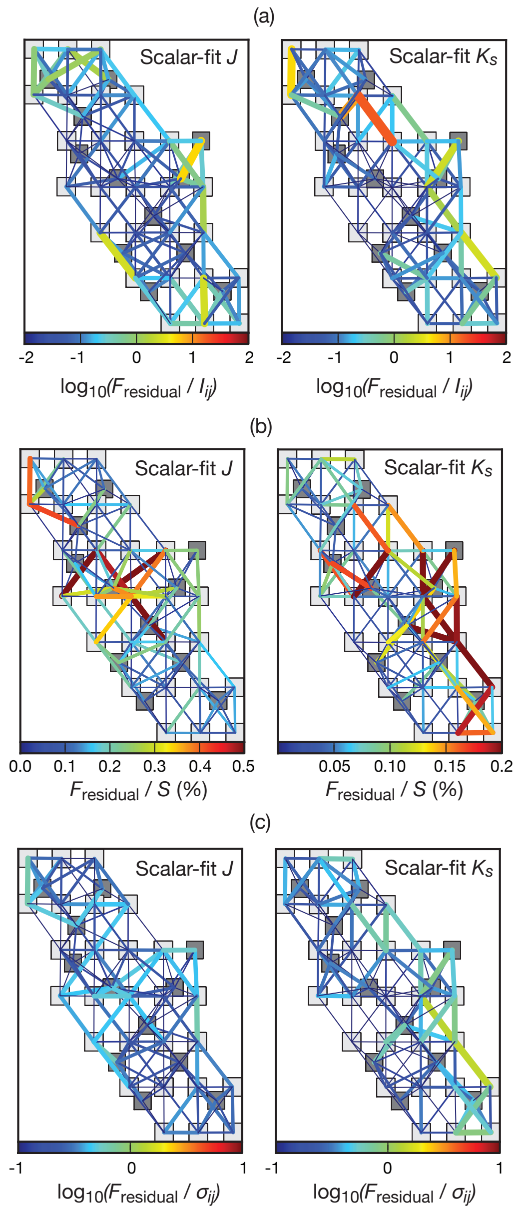


Figure 2.15: Map of residual block-to-block surface brightness differences after sky offsets: (a) as a fraction of the mean local surface brightness, (b) as a fraction of background level, (c) as a fraction of the standard deviation of the difference image. Thicker lines denote larger residual differences between overlapping fields (see also the color mapping). These graphs mimic the spatial distribution of the 2007B and 2009B WIRCam fields (Figure 2.1), with the footprints exploded to allow room for lines to connect coupled blocks. Fields observed in 2009B are plotted as darker squares than those observed in 2007B.

observational efficiency (see Table 2.1). Here we show that the background certainty of any sky-target nodding campaign is limited both by this minimum sky sampling latency, and also by the spatial structure of the sky background.

To test and distinguish temporal and spatial variations in the sky background, we plot the growth of background level variations versus time in Figure 2.16. As a fiducial for the intrinsic behaviour of background variations, we measure the mean and 95% background change as a function of time at a stationary site on the sky (that is, a single sky field, without telescope nodding). In agreement with Vaduvescu & McCall (2004), we see a mean background level variation of $\sim 0.5\%$ in 1 minute for both J and K_s bands. After 5 minutes, the intrinsic background level variation typically grows to 2%. At worst, we see background variations (measured at the 95% level of the sample distribution) of 5% in 5 minutes.

Individual points in Figure 2.16 are net sky offsets of disk images plotted against the time latency to the paired sky sample. The periodic time structure in Figure 2.16 is a consequence of the 2007B and 2009B sky-target nodding schemes (see Table 2.1); the circles and 'x' marks in Figure 2.16 denote the mean and 95% level of background variation, respectively, in each cluster. Recall that all 2009B disk integrations have equal sky sample latency due to the *sky-target-target-sky* nodding pattern. In both J and K_s bands, we see that nodding the telescope between sky and target *generates* additional background level uncertainty beyond that expected from strictly temporal sky background evolution. This makes sense in the context of spatial sky variations (Adams & Skrutskie 1996). As shown in Figure 2.16, the process of sky-target nodding can inflate sky background variations by 1.5–2 times the background variability expected at a stationary site on the sky on short time scales. On longer time scales, the nodding and stationary sky variance converge, perhaps indicative of the timescales that NIR skyglow structures move across a nodding distance (1°–2° on the sky).

This analysis underscores the challenge of accurately recovering surface brightness in a wide-field NIR mosaic. Sky-target nodding with CFHT implies typical time latencies of 60–70 seconds, and nodding distances of 1°–2°. Both of these elements prevent the true level of the background on M31's disk, in any single frame, from being known to an accuracy greater than 2%.

This also highlights why the 2009B program could only reduce the distribution of sky offsets by 30%. The rather shallow slopes of the mean background variance seen in sky-target nodding demonstrate the modest gain in background certainty by capping sky sampling latency at 1.2 minutes in 2009B compared to allowing latencies of 5 minutes in 2007B (see also Table 2.2). We also note that the expected 2009B sky offsets, at ~ 1.2 minute latency, are 10% larger than those from 2007B in both J and K_s bands. While this could indicate different physical behaviours in the background between the 2007B and 2009B semesters, we also note that the nods employed in 2009B were larger than in 2007B (Figure 2.3). Had the 2009B campaign used the same sky fields as in 2007B, rather than the pseudo-random sky ring, the performance of the 2009B sky offsets listed in Table 2.2 could be more impressive. Ultimately, future NIR observers with similar large sky-target nodding programs may choose to implement 2007B- or 2009B-like observing strategies depending on the priority of observational efficiency or absolute certainty of the background level. Figure 2.16 should be useful in planning such programs.

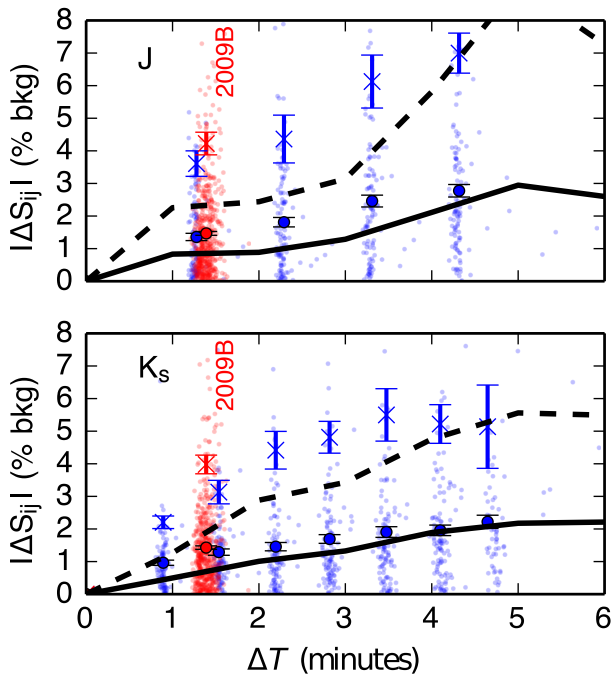


Figure 2.16: Temporal growth of background level variations observed at stationary sites on the sky versus with sky target nodding. Solid and dashed back lines mark the mean and 95% levels, respectively, of background variation observed in a single field without telescope nodding. Small blue and red dots show the net sky offset levels in 2007B and 2009B, effectively indicating the variation in background sky level seen as the telescope is nodded. Mean and 95% levels of sky offsets in the 2007B semester are plotted as large blue circles and 'x' symbols, respectively, while the same for the 2009B semester is plotted as red symbols where sky latency was constrained to 1.2 minutes in both bands.

2.9 Systematic Uncertainties in Surface Brightness Reconstruction

Sky offsets produce a mosaic that is rigorously optimal only in the sense of field-to-field surface brightness continuity—not absolute background subtraction. In this section, we attempt to gauge the systematic surface brightness error inherent in the sky offset technique.

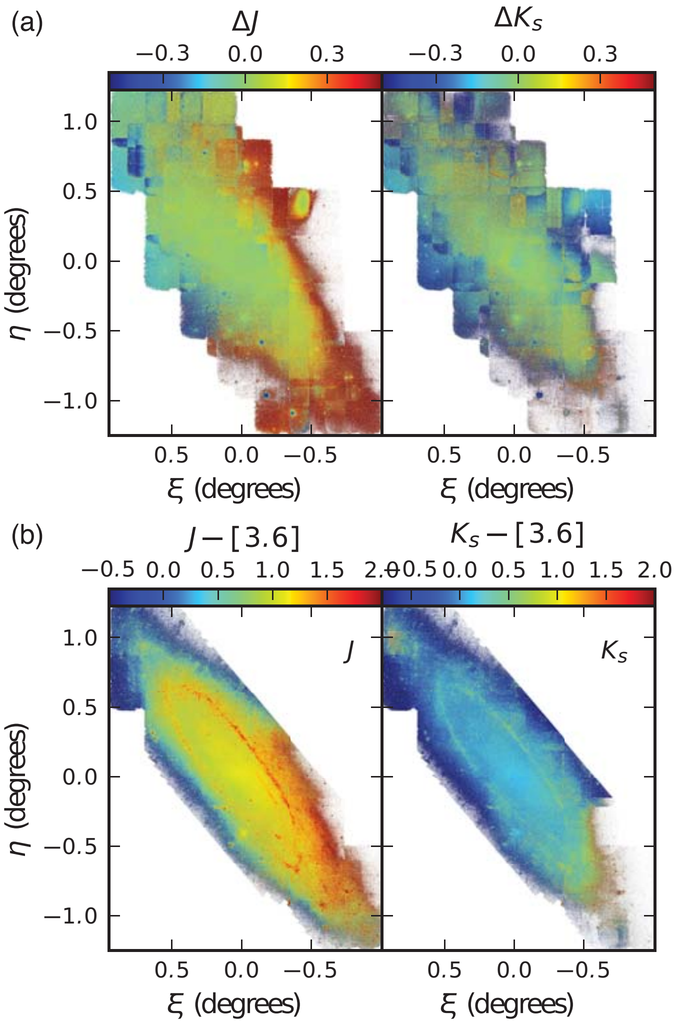


Figure 2.17: Surface brightness difference maps, showing systematic uncertainties in surface brightness reconstruction. (a) difference maps between our simplex scalar-fit mosaics (Figure 2.13b) and *Montage* scalar-fit mosaics. (b) Maps of $J - [3.6]$ and $K_s - [3.6]$ surface color inferred from the simplex scalar-sky fitted WIRCam mosaics and Spitzer/IRAC 3.6 μm image (Barmby et al. 2006). Note that the IRAC map crops the ANDROIDS/WIRCam footprint.

2.9.1 Comparison to Montage-fitted images

Besides the simplex method developed in § 2.7.1 and analyzed in detail in § 2.8, we also tested the *Montage* code that uses an iterative algorithm to solve either scalar or planar (two-dimensional linear gradient) sky offsets. Figure 2.17a shows the surface brightness difference between our simplex solution and the iterative *Montage* mosaic solution assuming scalar sky offsets. Despite an identical dataset, the two methods yield systematic differences of up to ~ 0.5 mag arcsec^{-2} at 20 kpc, though the solutions are consistent in the inner disk. Although the simplex and *Montage* scalar-offset mosaics *appear* comparable to the eye, a unique and optimal sky offset solution either does not exist, or is extremely difficult for our optimization algorithms to find.

Montage is also capable of fitting planar sky offsets to images, which is a tempting solution to the field-to-field discontinuities that persist between scalar-offset blocks. The result of planar fitting is shown in Figure 2.13c. We see that planar sky offsets, in this case, do little to improve the mosaics, and indeed, have a dramatic effect on the systematic surface brightness of the mosaic (by more than 1 mag arcsec^{-2} in the K_s band). Since our uncertainty on the shape of individual frames (0.3% of the background level across a WIRCam frame) is significant compared to the uncertainty in background level, fitting planar offsets to each WIRCam field introduces additional systematic error propagation compared to scalar sky offsets. *We thus recommend against using planar, or higher-order, sky offsets in wide-field WIRCam mosaics.*

2.9.2 Comparison to Spitzer/IRAC Images

We also explore systematic uncertainties in our WIRCam mosaics with comparisons against well-calibrated images of M31. A template for the NIR disk is the 3.6 μm Spitzer/IRAC map, presented in Barmby et al. (2006). Note that although Spitzer data avoid background subtraction issues caused by the NIR sky, planar sky offsets

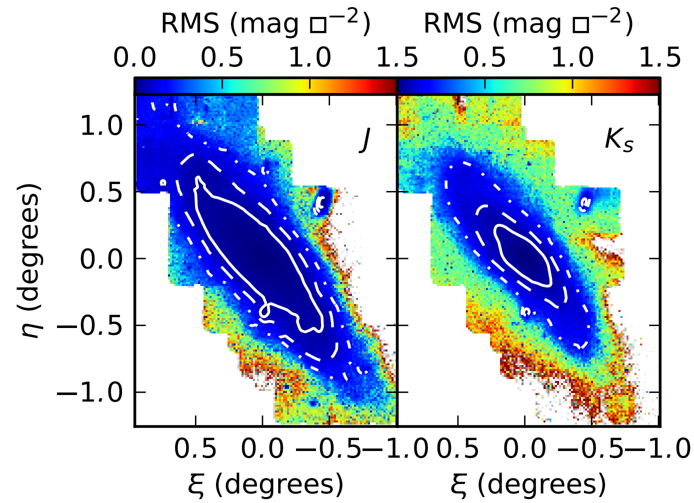


Figure 2.18: Mosaic maps of bootstrap RMS surface brightness in J (left) and K_s (right). White contours identify RMS levels of 0.05 (solid), 0.1 (dashed) and 0.2 (dash-dot) mag arcsec^{-2}.

were used by Barmby et al., though presumably of a smaller magnitude than our WIRCam sky offsets. In Figure 2.17b, we compare our simplex scalar-fitted mosaics against the 3.6 μm image. Generally the $J - [3.6]$ and $K_s - [3.6]$ colors decrease with disk radius, but increase in the star-forming regions due to hot dust emission. However both color maps (coincidentally) become redder in the south western disk beyond the 10 kpc star forming ring. We interpret this as a systematic over-subtraction of background in these regions on the order of $\gtrsim 1$ mag arcsec^{-2}. Evidently, our scalar sky offset mosaics are not systematically reliable beyond the bright disk of M31 with $R > 15$ kpc.

2.9.3 Monte Carlo Analysis of Systematic Surface Brightness Uncertainties

The difference images presented in the previous section illustrate how the surface brightness reconstructions of identical data can vary depending on the optimization algorithm. Here we pose a slightly different question: how are reconstructions affected by the initial conditions of background errors? That is, given the possible sets of background level biases affecting the blocks, what is the distribution of surface brightness reconstructions? We answer this with a realistic Monte Carlo (MC) analysis.

Our MC realization is generated by perturbing the surface brightness of the corrected blocks with a background error drawn (with replacement) from the ensemble of block sky offsets observed in the original mosaic (Figure 2.13b). Using the scalar-sky fitting procedure, sky offsets are optimized against the known sky background perturbations; 100 such realizations are made to compile an ensemble of mosaics in both bands. Figure 2.18 shows the RMS deviation of MC mosaic surface brightness against the original scalar-fitted mosaics. Reconstructed surface brightness in the outer disk can vary by ~ 1 mag arcsec^{-2}, consistent with color biases in the $J-[3.6]$ and $K_s-[3.6]$ maps.

We can ultimately understand the source of these systematic surface brightness errors by examining the standard deviations in the residual between expected and realized sky offsets in each Monte Carlo iteration. This residual dispersion is 0.15% of the J-background (0.17% of the K_s background); we find this dispersion to be constant across all fields in the mosaics. If mosaic surface brightness uncertainty is caused by flexure in the mosaic—where blocks on the mosaic periphery are forced to conform to the surface brightness of more central and tightly coupled blocks—then outer blocks would have higher offset dispersion. This is not the case.

Rather than mosaic flexure, a better model for Figure 2.18 involves uncertainties in the *post priori* adjustment for zero net offset (Eq. 2.5). Since block sky offsets have approximately Gaussian distributions with dispersions given in Table 2.2, the uncertainty in the net offset correction is simply $\sigma(\mathrm{block})/\sqrt{n_{\mathrm{blocks}}}$, where $n_{\mathrm{blocks}} = 39$ in the combined 2007B and 2009B mosaic. Given that $\sigma_{\Delta_B} \sim 1\%$, the expected uncertainty in the net offset correction is 0.16%: in perfect correspondence to the observed mosaic uncertainty. The dominant source of uncertainty shown in the MC simulations, Figure 2.18, is the use of an arithmetic mean of offsets to set an absolute zeropoint, not flexure or uncertainty in the network of offsets. This suggests that external zeropoints could be very useful in replacing Eq. 2.5. We pursue this in Chapter 4.

2.10 Conclusions

In this chapter, we have presented near-infrared (J and K_s) images of M31's entire bulge and disk with CFHT/WIRCam. These maps surpass the 2MASS (Beaton et al. 2007) and Spitzer (Barmby et al. 2006) mosaics with superior resolution that permits the identification of individual stars throughout M31's mid and outer disk. The dataset is also complementary to the HST/WFC3 PHAT survey (Dalcanton et al. 2012) by providing complete coverage of M31's entire disk within $R = 22$ kpc, and by offering a broader NIR color baseline ($J - K_s$) than is offered by WFC3 (approximately $J - H$). NIR mosaics of M31 have crucial applications for studies of the nearly attenuation-free stellar structure of our nearest spiral neighbour and for tests of stellar population synthesis models in NIR regimes.

Our focus in this chapter has been the establishment of procedures for accurately recovering the NIR surface brightness across 3 sq. deg. of the M31 disk using a sky-target nodding observing strategy with WIRCam on CFHT. We have compared two different observing methods to study the effects of sky target nodding cadences and patterns on sky subtraction uncertainties. We have also developed and tested flat fielding, zeropoint estimation, median sky subtraction, and sky offset optimization procedures in our WIRCam pipeline.

The surface brightness accuracy of a WIRCam frame is affected by both flat field uncertainties and additive background uncertainties. We recommend using sky flats built from images collected every night to calibrate WIRCam images since these capture the gain structure of WIRCam detectors (unlike dome flats). We also tested sky flats built on longer time spans (across a WIRCam queue run) or shorter spans (updated every half hour), but find that these either introduce flat fielding biases or become responsive only to changes in additive background contamination, respectively. Although an additive background (*e.g.,* thermal or scattered light) contaminates these flats, the influence of flat fielding errors on surface brightness shapes is minimal across the background-dominated mosaic. Instead, we find that the surface brightness across WIRCam frames is uncertain by 0.3% of the background intensity due to variations in the background between sky and target fields.

The necessity of nodding between sky and target fields limits our direct knowledge of the background level on the disk by $\gtrsim 2\%$ of the background level. Strictly minimizing latency between sky and disk integrations (as in the 2009B program) provides a 30% reduction in sky offsets, but is ultimately limited by overheads in nodding the telescope, and spatial structure in the NIR skyglow itself. Sky offset optimization is successful in reducing block-to-block surface brightness differences to $< 0.1\%$ of the background level. Our optimization algorithm reliably finds a consistent sky offset solution, so any errors in surface brightness shape across the mosaic are caused by errors in the shapes of individual blocks. There is, however, an uncertainty in the zero point of sky offsets, of order $\sim \sigma_{\Delta_B}/\sqrt{N_{\mathrm{blocks}}}$; 0.16% of the background level. In the next chapter we present CFHT/MegaCam imaging of M31's bulge and disk in $u^* g' r' i'$ bands. Background calibration, though still challenging, is more manageable in optical bands. Thus in Chapter 4 we show how well-constrained optical surface brightness maps can be used to resolve the background zeropoint uncertainty of the NIR surface brightness maps through a novel hierarchical SED modelling approach.

Our experience suggests that wide-field NIR programs that require large sky-target nods must tune their observing strategies with sky offset optimization in mind, besides simply minimizing sky sampling latency. Sky offset optimization is aided by having many independent blocks covering the target to decrease the statistical zero-point uncertainty. Increasing the number of independent blocks (observed hours or even a night apart to decouple sky and instrumental biases) is the most reliable way to establish the absolute surface brightness accuracy of the mosaic. Since sky offsets are further biased by any surface brightness shape errors in blocks (realized as our inability to diminish block-to-block offsets below $\sim 0.1\%$ of background brightness), we propose that blocks be interlaced by 50% (so that one detector completely overlaps a detector from an adjacent block). This interlacing pattern would enable the marginalization of shape errors across the entire detector frame. By doubling the number of blocks, each with individually halved exposure times, the mosaic could be reproduced with an equivalent net integration time.

Chapter 3

MegaCam Observations and Reduction

3.1 Introduction

The optical, $u^*g'r'i'$, portion of the ANDROIDS survey was observed with the MegaCam detector array on CFHT. MegaCam is an ideal instrument for M31 photometry since is covers a 1 deg. sq. field in a single integration across 36 detectors. Furthermore, CFHT/MegaCam is one of the most sensitive ground-based telescopes to the u^*-band, thanks to reduced UV absorption on Mauna Kea and optimized optics.[1] In this chapter, we describe the CFHT/MegaCam dataset acquired by ANDROIDS.

3.2 Elixir-LSB Observations

For the purposes of high-accuracy surface brightness (SB) imaging, the intent of this chapter, the drawback of MegaCam is the scattered light manifested around bright stars and across the field in general due to the wide-field correction optics. The latter manifests itself as radial gradient in the background, at a level of $\sim 10\%$ of the sky background. This background is also variable in time and telescope pointing.

[1] `http://www.cfht.hawaii.edu/Instruments/Imaging/Megacam/specsinformation.html`

To monitor and subtract the background, J. C. Cuillandre (*private communication*) developed the Elixir-LSB observing and processing method for MegaCam. Elixir-LSB borrows from NIR observing practice by building a real-time background image from a set of sky integrations drawn from a sliding time window. In applications for distant galaxies, there is sufficient blank sky around objects to build a background model from widely dithered science integrations—this is the approach taken by the Next Generation Virgo Cluster Survey (NGVS, Ferrarese et al. 2012). M31, however, is significantly larger than a single MegaCam field so that blank sky must be imaged from specially designated fields away from M31's disk.

Mapping the M31 disk out to the disk-halo transition at $R_{\mathrm{maj}} \sim 40$ kpc requires 14 MegaCam fields (Figure 3.1 and Table 3.1). Around these 14 target fields we chose eight sky fields devoid of bright stars. A sequence of observations begins with a sky field, then the telescope is nodded to two target fields in sequence before returning to a sky field, in a repeating sky-target-target pattern. Over an approximately hour-long observing block the entire ANDROIDS footprint is imaged. During these continuous observing blocks a real-time median background image is built.

This imaging cycle is repeated six times for each bandpass, potentially over different nights, with the telescope pointing offset to build a six-point dither pattern at each field. In g' and r' bands each integration is 160 seconds, giving a 16 minute integration depth per field (Table 3.2). Since M31 is predominantly red, the i' integrations are shorter (12 minute integration per field), while u^*-band integrations are longer (18 minute integration per field). Those longer integrations also compensate for CFHT/MegaCam's lower u^* bandpass throughput relative to other bands.

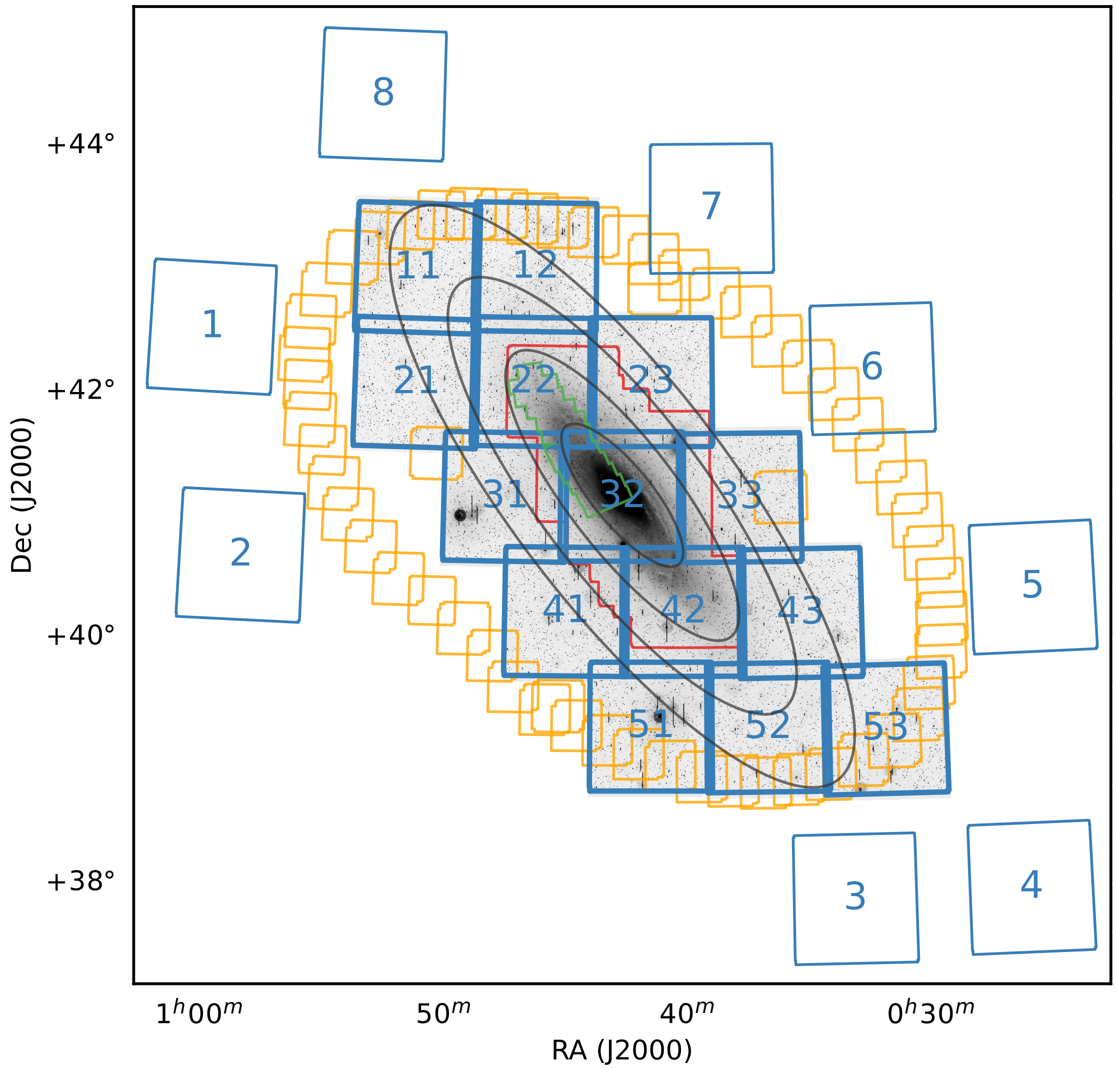


Figure 3.1: Androids Elixir-LSB MegaCam fields on M31. Individual MegaCam fields are drawn in blue, labelled by numbers corresponding to field names in Table 3.1. That is, the MegaCam footprint consists of 14 sky-subtracted "M31_SB" fields, calibrated using the eight surrounding "M31_SKY" fields. For comparison, the androids/WIRCam footprint is shown in red (disk) and orange (sky) fields (see also Figure 2.1). The HST/phat (Dalcanton et al. 2012) footprint is drawn in green. Grey ellipses approximate projected M31 disk radii of 10, 20, 30, and 40 kpc.

Table 3.1: Locations of MegaCam Elixir-LSB fields. Disk fields have names starting with "M31_SB", while background reference fields have names starting with "LSB_SKY." Relative coordinates are east (ξ) and north (η) tangent-plane offsets from the nucleus of M31.

Field	Sky Coordinate		Relative Coordinate	
	α	δ	χ (°)	η (°)
M31_SB_11	$00^h51^m56.01^s$	+43°06′05.10″	+1.68	+1.86
M31_SB_12	$00^h46^m40.73^s$	+43°07′47.02″	+0.72	+1.87
M31_SB_21	$00^h51^m51.20^s$	+42°09′48.00″	+1.69	+0.92
M31_SB_22	$00^h46^m40.52^s$	+42°11′25.68″	+0.73	+0.93
M31_SB_23	$00^h41^m29.60^s$	+42°11′45.98″	−0.23	+0.93
M31_SB_31	$00^h47^m50.52^s$	+41°14′47.58″	+0.96	−0.02
M31_SB_32	$00^h42^m44.99^s$	+41°15′32.17″	+0.00	−0.01
M31_SB_33	$00^h37^m37.68^s$	+41°14′47.57″	−0.96	−0.02
M31_SB_41	$00^h45^m08.86^s$	+40°18′52.80″	+0.46	−0.95
M31_SB_42	$00^h40^m06.75^s$	+40°18′51.25″	−0.50	−0.95
M31_SB_43	$00^h35^m04.80^s$	+40°17′39.73″	−1.46	−0.96
M31_SB_51	$00^h41^m29.64^s$	+39°22′36.90″	−0.24	−1.89
M31_SB_52	$00^h36^m31.83^s$	+39°21′46.95″	−1.20	−1.90
M31_SB_53	$00^h31^m34.38^s$	+39°19′50.93″	−2.16	−1.91
LSB_SKY_1	$01^h00^m58.35^s$	+42°32′47.27″	+3.36	+1.37
LSB_SKY_2	$00^h59^m11.53^s$	+40°41′47.97″	+3.12	−0.50
LSB_SKY_3	$00^h33^m00.80^s$	+37°57′04.48″	−1.92	−3.30
LSB_SKY_4	$00^h25^m42.83^s$	+37°59′50.98″	−3.36	−3.20
LSB_SKY_5	$00^h25^m04.78^s$	+40°26′47.48″	−3.36	−0.74
LSB_SKY_6	$00^h31^m42.48^s$	+42°16′48.68″	−2.04	+1.04
LSB_SKY_7	$00^h38^m45.63^s$	+43°36′50.15″	−0.72	+2.35
LSB_SKY_8	$00^h53^m45.33^s$	+44°30′24.07″	+1.97	+3.27

3.3 Mosaicing and Sky Offset Solutions

Our pipeline for producing mosaics of M31 with MegaCam Elixir-LSB data is similar to that developed for WIRCam, and described in Chapter 2 (Sick et al. 2014). The Elixir-LSB pipeline provides multi-extension FITS images, that are flat-fielded, have a photometric zero point solution, and possess a flat background. The remaining

Table 3.2: MegaCam Elixir-LSB observing pattern per field.

Bandpass	RUNID	N Integrations	Total Integration (minutes)	$\langle\mu_{\mathrm{background}}\rangle$ (mag arcsec^{-2})
u^*	12BC23	6	18.0	22.8 ± 0.2
g'	10BC97	6	16.0	22.1 ± 0.1
r'	10BC97	6	16.0	21.2 ± 0.2
i'	10BD99, 12BC23	6	12.0	20.2 ± 0.2

steps over which we have control are: 1) subtracting an absolute background level, 2) resampling the images to a common pixel frame, and 3) stacking images in a mosaic while adjusting the background level in each image to minimize surface brightness differences across images in the mosaic.

We subtract the absolute background level from each disk frame by using a background level estimated from the sky image taken most closely in time. To cleanly estimate the scalar background level in sky images we first build object masks with *Source Extractor* (Bertin & Arnouts 1996). With objects masked, we estimate the background level for each MegaCam chip from a 3σ-clipped median of the remaining pixels.

To build an astrometric solution prior to mosaicing we follow the standard *SCAMP* (Bertin 2006) workflow. First we build a *Source Extractor* object catalog for each disk image. These *Source Extractor* catalogs are inputs to *SCAMP*, which computes astrometric solutions, including higher-order distortions, for the images that are both consistent among MegaCam observations, and consistent to an external star catalog. To provide sufficient numbers of sources for the u-band images, we tie into the USNO-B1 star catalog (Monet et al. 2003). Although we use the `SAME_CRVAL` setting in conjunction with a global header to constrain the chip-to-chip MegaCam geometry, we find that *SCAMP* could reliably converge on a simultaneous astrometric solution for all 14 disk fields and four bandpasses. As a compromise, we instead build field-by-field astrometric solution sets so that all images in a single field are accurately registered. This exposes us to small (sub-arcsecond) astrometric offsets between fields due to the limited accuracy of the USNO-B1 external star catalog.

With astrometric solutions and background estimates, we build mosaics with a modified version of the WIRCam mosaicing pipeline introduced in Chapter 2 (Sick et al. 2014). First, we subtract the background (B, in DN) from each chip (F, as DN) and apply the CFHT Elixir photometric calibration so that each image has a zeropoint of 30 mag:

$$F' = (F - B) T_{\mathrm{exp}}^{-1} 10^{-0.4[\mathtt{PHOT\_C} + \mathtt{PHOT\_K}(\mathtt{AIRMASS} - 1) - 30]}. \tag{3.1}$$

T_{int} is the integration time of the image in seconds, and `PHOT_C` and `PHOT_K` are the photometric zeropoint constants and airmass terms, respectively, estimated by CFHT's Elixir pipeline. `AIRMASS` is, of course, the airmass, which is approximately $\sec z$, where z is the angle of the observation from the zenith. The readout noise and gain are also scaled by the same factor, and its inverse, respectively. We then project these uniformly calibrated images, for all MegaCam bands and fields, to a common pixel grid (i.e., equal pixel scales and `CRVAL`) by using *Swarp* (Bertin et al. 2002) in a resample-only mode.

Assembling these calibrated and resampled images into a pan-M31 mosaic would be straightforward with *Swarp* if not for background subtraction uncertainties inherent with sky-target nodding. Such uncertainties appear as surface brightness discontinuities between individual images in the mosaic. In Chapter 2 (Sick et al. 2014) we addressed this problem, in the same spirit as other codes such as *Montage* (Berriman et al. 2008), by solving for a set of scalar sky offsets that minimize image-to-image differences. Our *Skyoffset* code,[2] used here and in Chapter 2, differs from *Montage* in that we use *Swarp* as a mosaicing engine, use a downhill simplex algorithm to optimize the set of sky offsets, and finally use the concept of hierarchical image groups, such as individual detector chips and telescope fields (*stacks* and *blocks* in the terminology of Chapter 2). In Chapter 2 we took the approach that each WIRCam frame (of which four are imaged simultaneously by the WIRCam detector array) would have independent sky subtraction uncertainties, such that sky offsets were solved for each frame to build stacks, and a second set of sky offsets were solved to build blocks from groups of four stacks. Thanks to the relative stability of the optical sky, with Elixir-LSB we find that background residuals are flat across all 36 frames in a single MegaCam integration. We take advantage of this by building blocks, mosaics of MegaCam fields, in a *locked mode*. Here we solve frame offsets for each stack as usual, but instead of using those frame sky offsets directly we measure the median of all 36 frame offsets estimated for a given integration. This median frame offset is used to directly build MegaCam blocks from individual frames. From the 14 MegaCam blocks we again solve for block-level sky offsets that build the final ANDROIDS/MegaCam mosaics.

3.4 Discussion

We show the result of this observing and reduction program in Figure 3.2. The reduction approach is successful in producing mosaics that both free of field-to-field discontinuities and obvious large-scale background structures. Indeed, in § 5.5 we show that

[2] https://github.com/jonathansick/skyoffset

the ANDROIDS/MegaCam surface brightness profiles are stable out to the footprint limit at $R_{\mathrm{M31}} = 40$ kpc. The exquisite nature of the images is well-demonstrated in Figure 3.3, where we see M31's northern spur in unresolved light. To our knowledge, this disk warp has only been previously seen in resolved stellar density maps (e.g., Ferguson et al. 2002).

The main artifacts in these mosaics are halos around bright foreground stars. Our only recourse is to mask those foreground stars and their halos, which we do in Chapter 5.

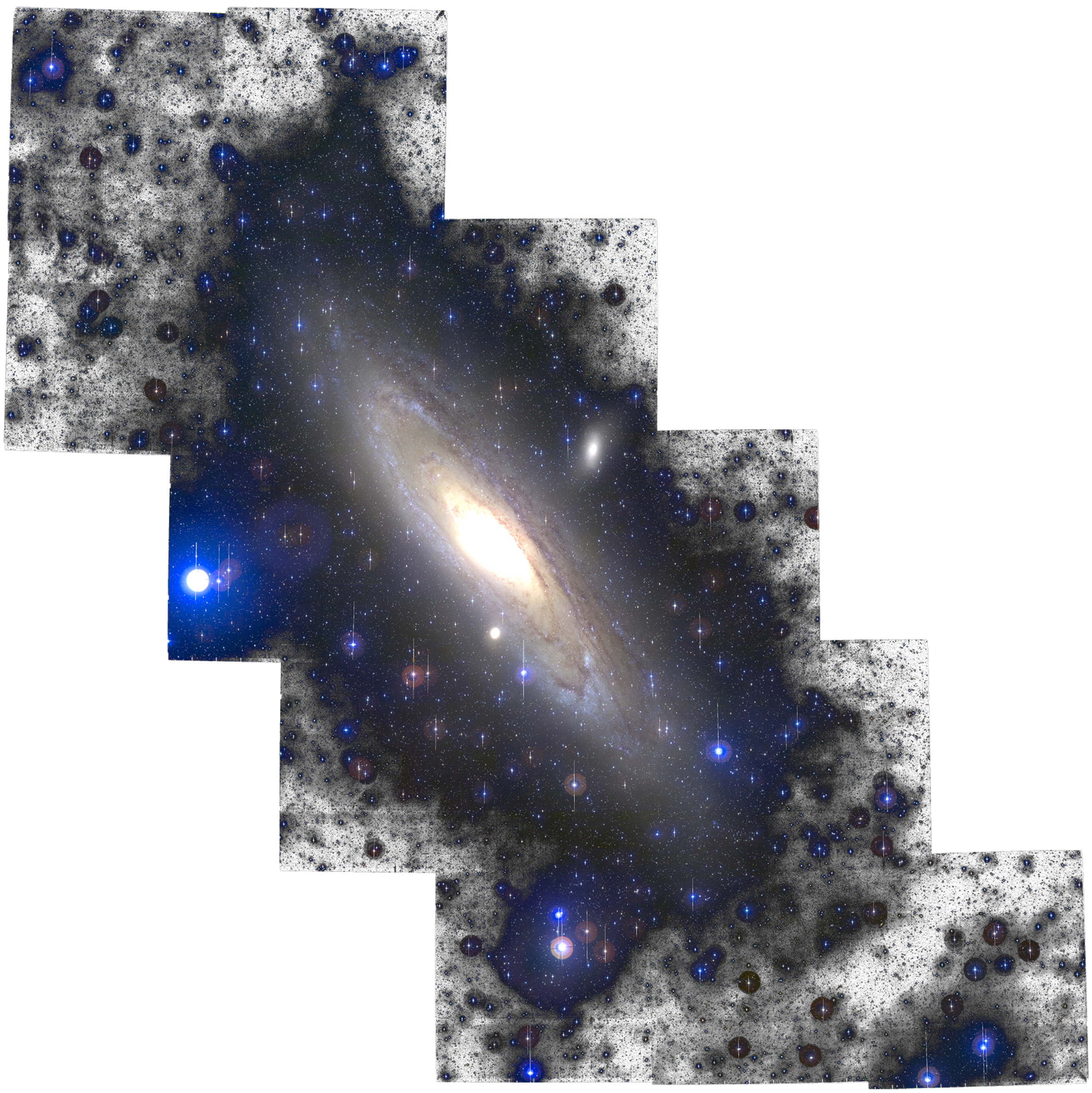


Figure 3.2: Androids $u^*g'r'$ mosaic of M31. The low surface brightness regions are shown in reversed r' light for clarity.

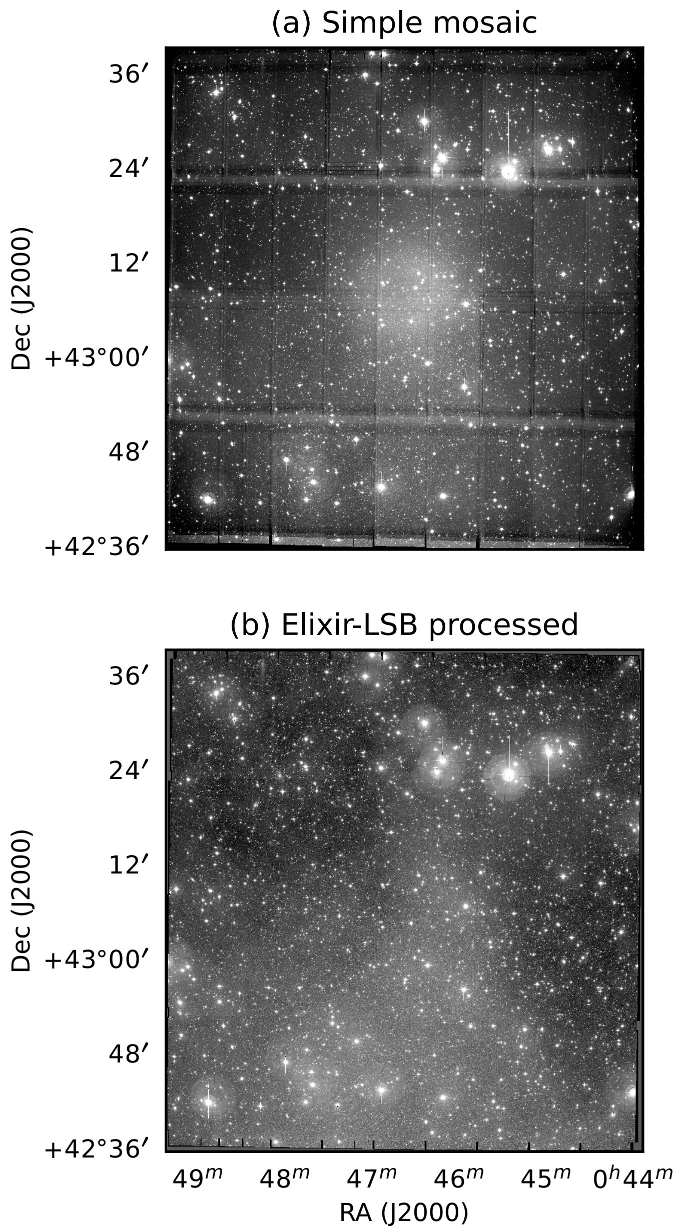


Figure 3.3: Comparison of (a) regular and (b) Elixir-LSB MegaCam processing in the `M31_SB_21` g' block. In regular MegaCam mosaicing, individual exposures that are astrometrically and photometrically calibrated are coadded with *Swarp*. Typically, local sky subtraction is performed with *Swarp* to remove sky background and instrumental scattered light; this cannot be done in ANDROIDS because the background consists of the unresolved M31 stars. Thus panel (a) shows MegaCam mosaicing where local sky subtraction cannot be performed. Central scattered light from MegaCam's optics is a main feature, along with image-to-image sky level variations seen in chip edges. With Elixir-LSB, the real-time background is subtracted from each image, eliminating both sky background and scattered light. With this procedure, the faint features like the M31 northern spur (panel b) are plainly visible.

Chapter 4

Background Calibration with Hierarchical Stellar Population Modelling

4.1 Introduction

Photometric background calibration is ANDROIDS's most substantial challenge. Extragalactic imaging surveys typically observe galaxies that fall within a single instrument field, allowing a survey's reduction pipeline to subtract a background measured directly from pixels surrounding an image. Such *direct* background measurement is not possible in ANDROIDS MegaCam and WIRCam imaging where the entire image plane is covered by M31 in any exposure. Both the WIRCam (Chapter 2) and MegaCam (Chapter 3) surveys used a first-order approach to background estimation through sky-target nodding, where sky backgrounds for a "target" M31 field are measured directly from a different "sky" field observed closely in time. With WIRCam J and K_s imaging, we found that such sky estimates are only accurate to $> 2\%$ of the sky background, which for near-infrared observations is unacceptably large. Taking advantage of image-to-image overlaps, we optimize background corrections that reduce image-to-image surface brightness differences in mosaics. While this is successful in producing *continuous* mosaics, the correction technique leaves the mosaic's overall surface brightness zeropoint (as a flux) uncertain by $\sigma_{\Delta\mathrm{B}}/\sqrt{N_{\mathrm{images}}}$. This limitation, combined with the possibility that background optimization can permit spatially-varying background biases, means that ANDROIDS mosaics, and WIRCam mosaics in particular, cannot readily be used for spectral energy distribution (SED) modelling.

In this chapter we address these issues and attempt to produce ANDROIDS MegaCam and WIRCam mosaics with minimal absolute background calibration biases. First, in § 4.2 we explore the possibility of using reference surface brightness profiles from the literature to set the absolute calibration of ANDROIDS mosaics. As this is not feasible, in § 4.3 we develop an algorithm for modelling stellar populations in ANDROIDS mosaics to estimate a background per image. In § 4.4 we describe how a hierarchical Gibbs algorithm can estimate parameters in this model. Finally in § 4.5 we apply this background modelling approach to the ANDROIDS mosaics and describe the results.

4.2 Reference M31 Surface Brightness Datasets

Other studies of M31's surface brightness profiles have been published before, often resulting in decompositions of its bulge, disk, and stellar halo components. See Courteau et al. (2011), hereafter C11, and references therein. It is tempting to solve the ANDROIDS background calibration issue by bootstrapping against the absolute calibrations of existing datasets. This section surveys such datasets in the literature to discern whether or not they provide sufficient constraints to calibrate ANDROIDS

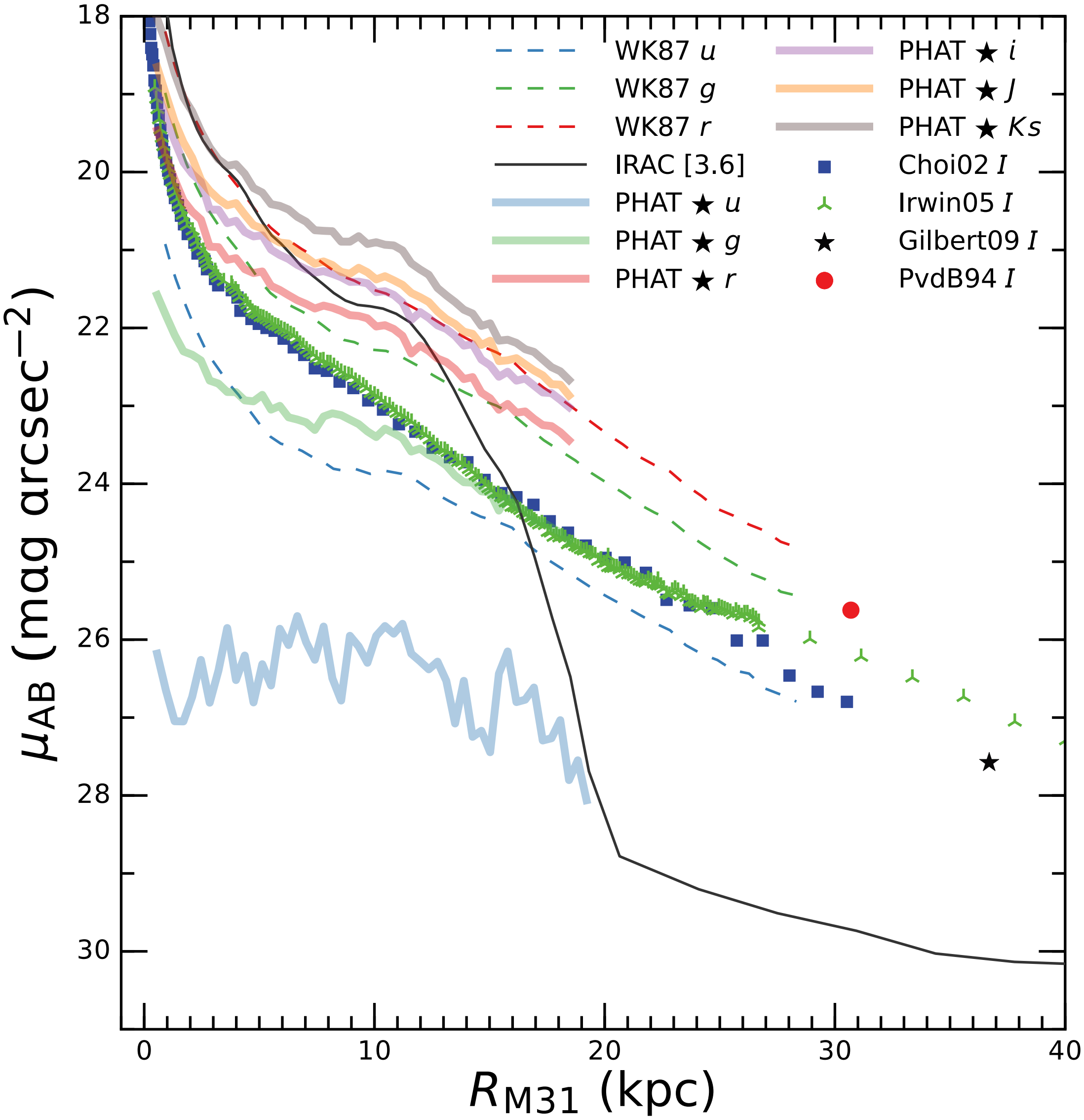
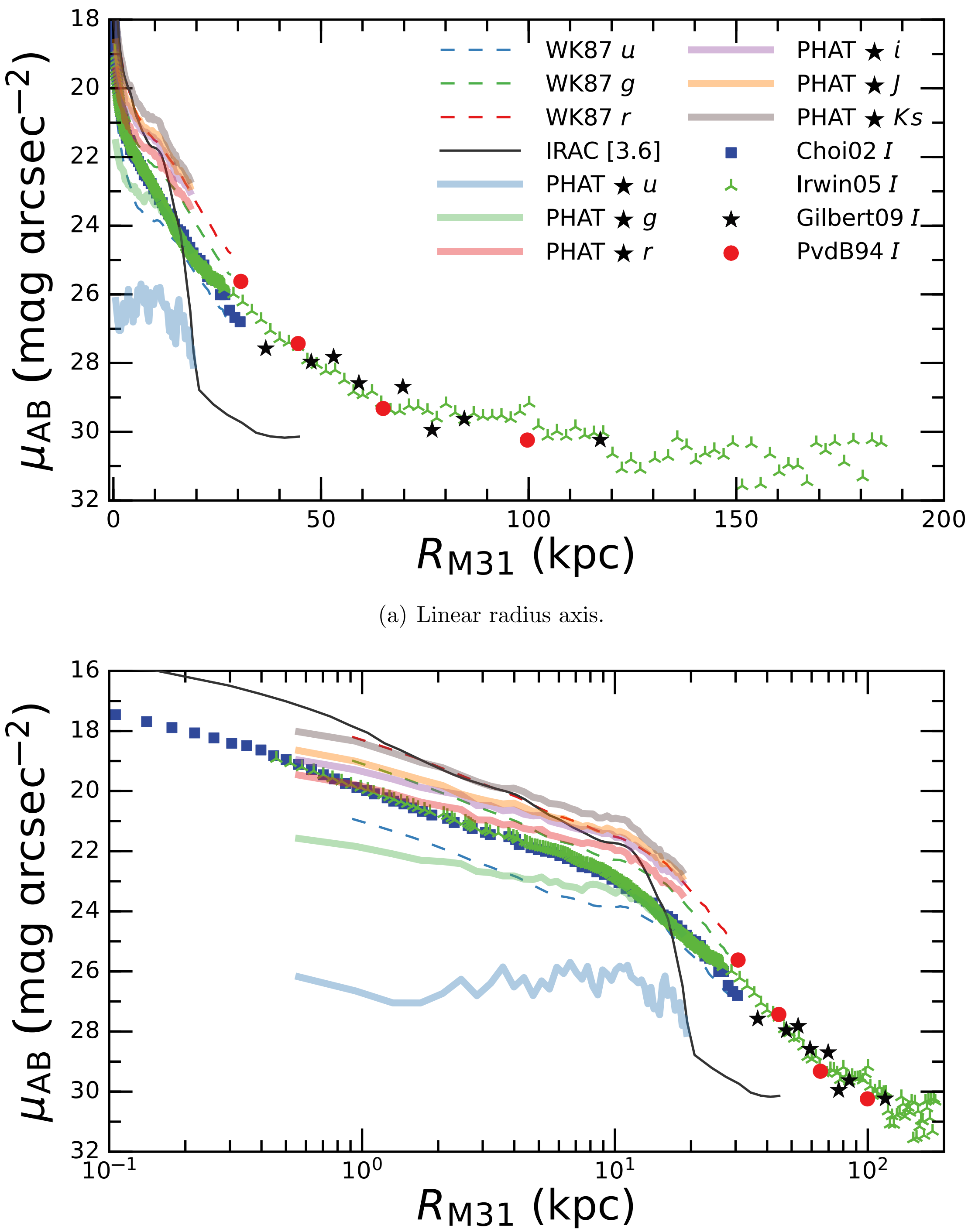


Figure 4.1: Reference M31 surface brightness profiles within 40 kpc from the literature, in AB magnitudes and projected along the major axis. WK87 profiles are measured from photographic plates by Walterbos & Kennicutt (1987) and transformed here into SDSS AB magnitudes (see § 4.2.1). The "Choi02", "Irwin05", "Gilbert09", and "PvdB94" datasets are I-band profiles calibrated by C11 and converted to an AB zeropoint. "IRAC [36]" is an IRAC 3.6 μm profile measured by Rafiei Ravandi et al. (2016) and converted to the AB magnitude system. "PHAT" profiles are synthesized from PHAT (Dalcanton et al. 2012) resolved stellar catalogs and transformed into $ugriJK_s$ (see § 4.2.2). Figure 4.2 is an alternative version of this plot extending to $R_{\mathrm{M31}} = 200$ kpc.



(a) Linear radius axis.

(b) Logarithmic radius axis.

Figure 4.2: Reference M31 surface brightness profiles at large radii. See Figure 4.1 for a discussion of data sources.

mosaics and surface brightness profiles.

4.2.1 Existing Optical and Near-Infrared Surface Brightness Profiles

Walterbos & Kennicutt 1987

The Walterbos & Kennicutt (1987), hereafter WK87, $UBVR$ images are a dataset with enduring value. WK87 observed M31 with the 0.94 meter Burrell Schmidt telescope, which has a 5.16×5.16 degree^2 field that encompasses the entire M31 stellar disk. WK87 defined a "sky" region covering 35% of the digitized plate and fitted the spatially-varying background with a second-order two-dimensional polynomial to narrow rectangles oriented perpendicular to M31's major axis. This technique enabled them to model complex background variations with a simple function, and without being too sensitive to small-scale effects created by resolved sources. In Figure 4.1 we show surface brightness profiles extracted from these observations by WK87, though converted from Johnson-Cousins Vega magnitudes to the SDSS ugr filter system in AB magnitudes following Fukugita et al. (1996):

$$g = V + 0.56(B - V) - 0.12, \tag{4.1}$$

$$r = V - 0.49(B - V) + 0.11, \tag{4.2}$$

$$u = 1.38(U - B) + 1.14 + g. \tag{4.3}$$

These profiles show that M31's stellar disk is exponential out to at least 30 kpc. Despite their reliability, the WK87 profiles are not useful here since they do not provide direct information about profile shapes in redder iJK_s bands where ANDROIDS background calibration is most uncertain.

Courteau et al 2011 and Choi et al 2002

C11 is a significant analysis of M31's bulge, disk, and halo shapes. The authors gathered multiple M31 surface brightness datasets, spanning M31's full radial extent, and consistently calibrated and measured surface brightness profiles. Of those, the Choi et al. (2002) profile is most useful since it is an I-band profile that covers M31's bulge and disk out to 30 kpc — similar to, though smaller than, the ANDROIDS/MegaCam footprint.

We transformed C11's minor axis profiles into a major axis profile to enable comparison to the ANDROIDS profiles. Note that this transformation is an approximation as it assumes a flat circular disk. Given an inclination of 77.5° (Walterbos & Kennicutt 1988), surface brightnesses measured across the minor axis are re-projected to the major axis:

$$\mu_{\mathrm{major}}(R_{\mathrm{M31}}) = \mu_{\mathrm{minor}}(R_{\mathrm{projected}}) + 2.5 \log_{10}(\cos 77.5^{\circ}). \quad (4.4)$$

Projected radii across the minor axis are converted to de-projected radii along the major axis with:

$$R_{\mathrm{M31}} = \frac{R_{\mathrm{projected}}}{\cos 77.5^{\circ}}. \quad (4.5)$$

The C11 profiles (labeled as "Choi02", "Irwin05", and "Gilbert09") are shown in Figs. 4.1 and 4.2. This profile agrees with the outer disk profile slope observed by WK87, and further shows that M31's light profile remains exponential out to $R_{31} = 30$ kpc; there is no break in M31's disk.

Rafiei Ravandi et al 2016

All M31 reference profiles considered to this point have traced M31's optical light. These are less useful for calibrating ANDROIDS's near-infrared datasets, which have the most substantial background uncertainties. The 2MASS mosaics of M31 in JHK_s, assembled by Beaton et al. (2007) as part of the '6X' deep 2MASS project, are the closest analogue to ANDROIDS's WIRCam imaging. That dataset is different from WIRCam imaging since the disk was observed with a drift scanning technique, and thus background levels are correlated from one side of the disk to the other. Despite this, C11 found the 2MASS mosaics to be unsuitable for profile fitting because of background variations.

Space-based observations, which avoid effects of an *atmospheric* background, can in principle be more productive. With Spitzer IRAC, Barmby et al. (2006) mapped the M31 bulge disk and disk in four channels, covering 3.6 μm – 8.0 μm wavelengths. Rafiei Ravandi et al. (2016) expanded upon the original survey with additional IRAC 3.6 μm and 4.5 μm imaging along the major and minor axes. This footprint extends to $R_{\mathrm{M31}} = 45$ kpc along the major axis, covering M31's inner stellar halo.

Rafiei Ravandi et al. measured profiles with complementary integrated light and star counting approaches in the inner and outer disk regions, respectively. Spitzer IRAC is affected by a Zodiacal light background, which Rafiei Ravandi et al. measure directly from blank sky regions at the edge of the mosaic. They interpolate this background across M31's disk with a 2D first-order polynomial (see their Figs. 2 and 3). Using the same major-axis logarithmic wedge binning scheme as C11, Rafiei Ravandi et al. (2016) measure an integrated light surface brightness profile to $R_{M31} =$ 20 kpc.

Beyond 20 kpc, Rafiei Ravandi et al. synthesize a surface brightness profile from star counts in wedge bins. The zeropoint of the star count profile is fit to the integrated light profile at the transition point, $R_{\mathrm{M31}} = 20$ kpc. While resolved stellar photometry is an excellent tracer for low-surface brightness features, Milky Way foreground and stellar crowding both affect the conversion of observed stellar density to an M31 surface brightness profile. To correct for crowding, Rafiei Ravandi et al. used an artificial star testing method: synthetic stars of different magnitudes are added to images in different stellar density environments. These images are photometered, and the rate that known artificial stars are recovered is p_C. This recovery rate is often binned by the magnitude m and source density ρ of the artificial stars: $p_C(m, \rho)$. The recovery rate can be interpreted as the probability that a star is observed in real data. Thus one can correct for the incompleteness of an observed star catalog by treating each observed star as $p_C^{-1}(m, \rho) \geq 1$ stars.

To correct for Milky Way foreground stars, Rafiei Ravandi et al. assembled an observational of model of a Milky Way $[3.6] - [4.5]$ Hess diagram from WISE imaging (Wright et al. 2010), and from TRILEGAL (Girardi et al. 2005), a Milky Way stellar population modelling code. These models provide a probability that an observed star is an M31 star given near-infrared photometry and location in the sky, $p_{\mathrm{M31}}([3.6], [4.5], \alpha, \delta)$ Observed stars are then again re-weighted by a factor p_{M31}.

Altogether, the combined IRAC integrated light and resolved star count profile is shown in Figs. 4.1 and 4.2. Compared to the C11 profile, the Rafiei Ravandi et al. (2016) profile has multiple profile breaks. Just outside the 10 kpc ring the [3.6] profile breaks downwards and does not flatten out until $R_{\mathrm{M31}} = 20$ kpc. Given the lack of corroboration in other profiles, we propose that the breaks are driven by background calibration errors. An over-subtracted integrated light profile could produce the downward profile break. That another break occurs at the transition between the integrated and resolved stellar profiles again reflects an incorrect background subtraction in the integrated light profile.

Despite a space-based instrument and detailed calibration methods, a reliable near-infrared surface brightness profile does not yet exist for calibrating ANDROIDS WIRCam images.

4.2.2 Synthetic Surface Brightness Samples from Hubble Space Telescope Star Catalogs

Given the dearth of reliable M31 surface photometry, particularly coinciding with ANDROIDS WIRCam imaging, we investigated the viability of synthesizing surface photometry from resolved star catalogs. Transforming resolved stellar photometry catalogs into surface photometry is discussed above in conjunction with the Rafiei Ravandi et al. (2016) IRAC dataset, but this investigation differs in three important ways from the method used by Rafiei Ravandi et al. (2016). First, we shall use observations from Hubble Space Telescope (HST), which has a smaller PSF that reduces crowding in the dense stellar disk. This allows the technique to be used *in* the M31 disk, rather than outside $R_{\mathrm{M31}} = 20$ kpc. Second, HST's ACS and WFC3 cameras provide optical and near-infrared photometry that is more analogous to ANDROIDS's $u^{*}g'r'i'JK_{s}$ filter set. Finally, we attempt to generate an *absolute* surface brightness calibration by combining completeness measurements with the measured photometry of individual stars. Such an absolute calibration avoids having to bootstrap a resolved stellar photometry profile from an integrated light profile, meaning that the surface photometry synthesized from star catalogs stands on its own.

The Panchromatic Hubble Andromeda Treasury (PHAT, Dalcanton et al. 2012) is ideal for this effort since it covers nearly the same radial extent as the WIRCam mosaic footprint (out to $R_{\mathrm{mag}} \sim 20$ kpc, see Figure 7.1, though across only 0.5 degree^2 of the NE M31 quadrant). PHAT observed M31 in six bandpasses: F225W and F336W with WFC3/UV; F475W and F814W with ACS; F110W and F160W with WFC3/IR. F110W and F160W are similar to the J and H bands. In this study we use the `V2` PHAT photometric catalog made available by Williams et al. (2014), hereafter W14.

PHAT is divided into 23 *bricks*, each covering $6' \times 12'$. Within each brick are 18 fields, corresponding to $2' \times 2'$ WFC3/IR footprints. Those fields become the pixels of the surface brightness computation.

For each field in each brick, we compute the surface brightness in a given band, X, as follows. First, we determine the effective area (A, arcsec^2), of the field from mask images published by W14. Then we determine completeness for stars. W14 performed artificial star testing in six fields across the PHAT footprint (see Figure 7.1). Thus we adopt the artificial star testing field closest to the field being considered. Artificial star testing consists of injecting mock stars into images, and measuring their observed photometry (and whether the star is recovered at all). Thus we compute a completeness *luminosity function* by binning mock stars by magnitude and measuring the fraction of mocks that are recovered in each bin. Then for each observed star in a field, we adopt the completeness fraction c read from that completeness luminosity function. See W14 for further details on PHAT's completeness functions. Finally, we compute the surface brightness in bandpass X for a given PHAT field by adding up the light of individual stars and correcting for completeness:

$$\mu_X = -2.5 \log_{10} \left(\frac{\sum_i c_{i,X}^{-1} 10^{-0.4 m_{i,X}}}{A} \right) \tag{4.6}$$

where A is the area in arcsec^2 of the field, $c_{i,X}$ is the completeness fraction estimated for star i in the X band, and $m_{i,X}$ is the observed magnitude of star i in band X. To create surface brightness *profiles* from per-field surface brightness scalars, we bin each PHAT field by de-projected galactocentric radius (R_{M31}) and compute the sample median within each radial bin.

To facilitate comparison of these profiles to those from the literature, discussed in § 4.2.1, we convert the Vega-based magnitudes in F225W, F336W, F475W, F814W, F110W, and F160W to the AB system (Sirianni et al. 2005, and WFC3 Instrument Handbook [1]):

$$\mathrm{F275W_{AB}} = \mathrm{F275W_{Vega}} + 1.4983, \tag{4.7}$$

$$\mathrm{F336W_{AB}} = \mathrm{F336W_{Vega}} + 1.1846, \tag{4.8}$$

$$\mathrm{F475W_{AB}} = \mathrm{F475W_{Vega}} - 0.495, \tag{4.9}$$

$$\mathrm{F814W_{AB}} = \mathrm{F814W_{Vega}} + 0.436, \tag{4.10}$$

$$\mathrm{F110W_{AB}} = \mathrm{F110W_{Vega}} + 0.7595, \tag{4.11}$$

$$\mathrm{F160W_{AB}} = \mathrm{F160W_{Vega}} + 1.2514. \tag{4.12}$$

Then we transform magnitudes in HST filters into an $ugriJK_s$ set using 2nd order polynomials fit to a library of FSPS (Conroy et al. 2009, 2010; Conroy & Gunn 2010)

[1] http://www.stsci.edu/hst/wfc3/phot_zp_lbn

model stellar population photometry:

$$u = \mathrm{F336W_{AB}} - 0.0912 - 0.3435(\mathrm{F275W_{AB}} - \mathrm{F336W_{AB}}) + 0.1381(\mathrm{F275W_{AB}} - \mathrm{F336W_{AB}})^2, \tag{4.13}$$

$$g = \mathrm{F475W_{AB}} + 0.0001 + 0.0196(\mathrm{F475W_{AB}} - \mathrm{F814W_{AB}}) + 0.0062(\mathrm{F475W_{AB}} - \mathrm{F814W_{AB}})^2, \tag{4.14}$$

$$r = \mathrm{F814W_{AB}} + 0.0120 + 0.4331(\mathrm{F475W_{AB}} - \mathrm{F814W_{AB}}) - 0.0705(\mathrm{F475W_{AB}} - \mathrm{F814W_{AB}})^2, \tag{4.15}$$

$$i = \mathrm{F814W_{AB}} + 0.0014 + 0.1046(\mathrm{F475W_{AB}} - \mathrm{F814W_{AB}}) - 0.0162(\mathrm{F475W_{AB}} - \mathrm{F814W_{AB}})^2, \tag{4.16}$$

$$J = \mathrm{F160W_{AB}} + 0.0053 + 0.6740(\mathrm{F110W_{AB}} - \mathrm{F160W_{AB}}) - 0.0461(\mathrm{F110W_{AB}} - \mathrm{F160W_{AB}})^2, \tag{4.17}$$

$$K_s = \mathrm{F160W_{AB}} + 0.3200 - 0.8219(\mathrm{F110W_{AB}} - \mathrm{F160W_{AB}}) - 0.4217(\mathrm{F110W_{AB}} - \mathrm{F160W_{AB}})^2, \tag{4.18}$$

These $ugriJK_s$ surface brightness profiles synthesized from PHAT star catalogs are shown in Figure 4.1. Outside $R_{\mathrm{M31}} > 10$ kpc, these profiles have similar exponential slopes as the WK87 and C11 profiles. And though the bandpasses are transformed, the combination of F110W and F160W predict the WIRCam J and K_s profiles.

Nonetheless, the synthetic PHAT surface brightness profiles do have invalidating qualities. Though we attempted to compute absolutely calibrated profiles, the overall zeropoints of the u, g and r profiles do disagree with WK87. Second, these profiles cover only one quadrant of M31, making a direct calibration of full-disk mosaics less straightforward. In summary, absolutely-calibrated surface brightness profiles synthesized from HST resolved stellar photometry are not adequate for calibrated ANDROIDS photometry.

4.2.3 Summary

In this section we have investigated the possibility of using M31 surface brightness datasets available in literature to calibrate photometric background biases in ANDROIDS MegaCam and WIRCam mosaics. All of these datasets are unsuitable, either on grounds that they do not constrain near-infrared J and K_s profiles, or that they themselves have suspect calibration. Since no literature profile is useful for calibrating ANDROIDS photometry, we also attempted to synthesize absolute surface brightness profiles from PHAT star catalogs. Although the synthesized PHAT profiles match the mid- and outer-disk shape observed by WK87 and C11, their absolute calibrations do not match WK87's *ugr* profiles. This failure precludes us from directly using the synthesized PHAT profiles to bootstrap ANDROIDS WIRCam surface brightness calibrations.

Short of external datasets, we must seek a method of calibrating surface brightness profiles from physical modelling alone. In the following section we propose and describe such an approach, where stellar population modelling yields background estimates as marginalized model parameters.

4.3 Modelling Pixel SEDs to Calibrate Backgrounds

The strong backgrounds in the ANDROIDS mosaics, along with a dearth of reliable M31 SB datasets in the literature noted in § 4.2 makes the task of calibrating the ANDROIDS mosaics daunting. Nevertheless we can proceed by realizing that background subtraction residuals are detectable as non-physical SEDs. If we can model a pixel's SED as a combination of M31's light and a background residual, we can in principle remove the background bias that plagues the present analysis.

4.3.1 Bayesian Modelling of SEDs

SEDs can be modelled with stellar population synthesis software that combines stellar isochrones (itself a product of stellar evolution and atmosphere models) with prescriptions of the stellar initial mass functions, star formation history, and dust attenuation. As we will discuss in more detail in following sections, we use FSPS (Conroy et al. 2009, 2010; Conroy & Gunn 2010) to generate model SEDs.

Let us write the model SED produced by a stellar population synthesis code as $f(\{\theta\})$, where $\{\theta\}$ is the set of stellar population parameters (e.g., stellar mass, star formation history, metallicity, attenuation, initial mass function, distance).

Since observed SEDs have Normally-distributed uncertainties, via the Central Limit Theorem of Poisson photon statistics, the likelihood of the observed SED given the model can be written as

$$\log \mathcal{L}(F|\{\theta\}) = -\sum_{X} \left(\frac{F_X - f_X(\{\theta\})}{\sigma_X^2} \right), \tag{4.19}$$

where F and σ^2 are observed flux and variance, respectively, and X denotes specific bandpasses in the SED (e.g., u, g, r, i, J, K_s in the case of ANDROIDS).

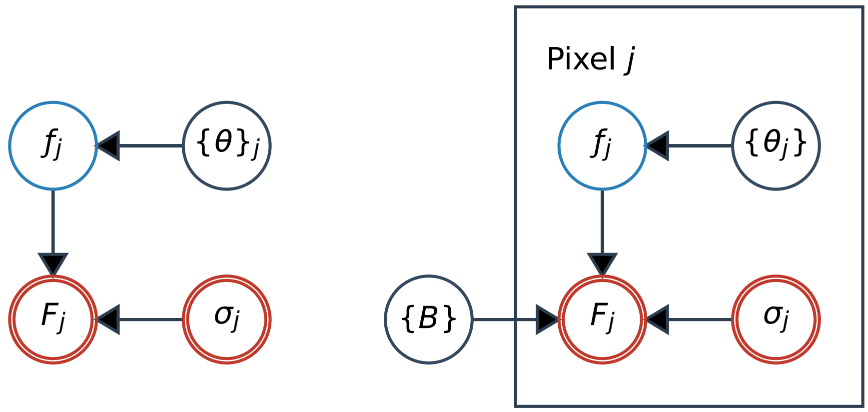
In classical, frequentist, statistics, it would be standard practise to estimate the stellar population parameters $\{\theta\}$ as ones that maximize the likelihood $\log \mathcal{L}(F|\{\theta\})$. However, one notes an awkwardness in this likelihood: it refers to the probability of the observations being explained by the model, rather than vice versa. Bayes' Theorem allows us to invert this expression via

$$p(\{\theta\}|F) = \frac{\mathcal{L}(F|\{\theta\})p(\{\theta\})}{p(F)}. \tag{4.20}$$

This equation forms the basis of Bayesian SED inference. The probability of model parameters reflecting the observations, $p(\{\theta\}|F)$, is termed the *posterior probability*. The term $p(\{\theta\})$ is the prior probability of the model parameters, which is at the discretion of the practitioner to assert belief in the probability of the true parameters occupying any region of parameter space. Typically the prior probabilities of each parameter are independent so that the prior may be written as the joint probability $p(\{\theta\}) = \prod_i p(\theta_i)$. Note that the denominator of Eq. 4.20 is a model-independent constant. Bayesian inference is typically concerned only with ratios of posterior probability, so that we may ignore the $p(F)$ factor. Thus:

$$\log p(\{\theta\}|F) \propto -\sum_X \left(\frac{F_X - f_X(\{\theta\})}{\sigma_X^2} \right) + \sum_i \log p(\theta_i) \tag{4.21}$$

is the posterior probability of a stellar population model given an observed SED, written in full. The model specified by Eq. 4.21 is also visualized as a probabilistic graphical diagram in Figure 4.3(a).



(a) Single-pixel SED model.

(b) Multiple-pixel SED model with a global background term.

Figure 4.3: Probabilistic graphical SED model. PGMs describe generative Bayesian models; panel (a) represents Eq. 4.21 and panel (b) represents Eq. 4.23. Specifically, in (a), the observed flux F (double circles represent observations) can be generated from both the observed uncertainties σ_F and the stellar population SED model, f. The model flux, in turn, is generated from information in the stellar population parameters, $\{\theta\}$. In the hierarchical model, the box (termed, *plate*) indicates there are several pixels j in the model. All generative nodes *within* the plate act independently for each pixel. In other words, each pixel has an independent stellar population to accommodate, for example, stellar population gradients across a galaxy disk. However, the background term acts globally, generating all pixel observations in the model.

4.3.2 A Hierarchical Model for Multiple Pixel SEDs with a Background

Returning to the issue of background bias, we realize that an observed SED, F, with a strong background will always yield extremely low posterior likelihood models for *any* set of physical parameters. However, we can add a background parameter to our SED model so that the background bias SED is effectively decoupled from the galaxy SED model. With a background term, the posterior likelihood can be re-written as

$$\log p(\{\theta\}, \{B\}|F) \propto -\sum_X \left(\frac{F_X - B_X - f_X(\{\theta\})}{\sigma_X^2} \right) + \sum_i \log p(\theta_i) + \sum_X p(B_X). \tag{4.22}$$

If we can model both the stellar population and residual background in a pixel, we can effectively achieve our calibration and astrophysical goals simultaneously. Of course independently modelling the stellar population and background in each pixel is ill-determined: background corrections for several bandpasses and stellar population parameters are completely degenerate. However, if we assume that the background bias varies on much larger scales than the stellar population properties (which is true in this case) then we may model a background correction in each bandpass, while modelling independent stellar populations in each pixel. Thus the *global* posterior likelihood, p_G, can be written as

$$\begin{aligned} \log p_G(\{\theta\}_1, \dots \{\theta\}_N, \{B\}|F_1, \dots F_N) \propto &- \sum_j \sum_X \left(\frac{F_{X,j} - B_X - f_X(\{\theta\}_j)}{\sigma_{X,j}^2} \right) \\ &+ \sum_j \sum_i \log p(\theta_{i,j}) + \sum_X p(B_X). \end{aligned} \tag{4.23}$$

Requiring that a single background correction yield physically sensible SEDs across a diverse range of stellar populations makes this a tractable, unique, modelling problem. This is a hierarchical modelling approach. The model specified by Eq. 4.23 is also visualized as a probabilistic graphical diagram in Figure 4.3(b). In this diagram, it becomes clear that the stellar population parameters, $\{\theta\}$, exist at a lower level in the model, and the background terms, $\{B\}$, are hyperparameters that act globally.

4.3.3 A Metropolis-Hastings in Gibbs MCMC Sampler

At first glance, the problem with this hierarchical modelling approach is the abundance of parameters. If a single pixel's stellar population can be modelled with N parameters (typically $N \sim 9$ for simple models), six bandpasses, and M pixels, then the total number of parameters being fit is $6NM$. It is easy to quickly confront models with 10^3–10^6 parameters. A popular Bayesian approach to hierarchical modelling is the Gibbs Markov Chain Monte Carlo (MCMC) sampler (Geman & Geman 1984).

MCMC allows us to estimate model parameters by sampling from parameter space in proportion to the posterior likelihood. Specifically, this is done by generating a chain of parameter samples, where a new parameter is proposed from the latest state, and the proposed step is accepted depending on the posterior likelihoods of the current and proposed states. A popular method for generating and testing MCMC step proposals is the Metropolis-Hastings algorithm (Metropolis et al. 1953; Hastings 1970).

Rather than proposing a step in all parameters simultaneously, a Gibbs sampler allows us to sample from parameters in each hierarchical level independently, and in alternation. That is, a Gibbs sampler allows us to step in the background hyperparameter space, then assume that background state when stepping in the stellar population parameter spaces of pixels. Since each pixel's stellar population is treated independently, we can further simplify the computation by sampling the stellar populations of pixels in parallel.

Let us now walk through the operation of Gibbs sampler for stellar population modelling. Note that the specific implementation is discussed in § 4.4.

Sampler Initialization

We begin by initializing the state of all parameters. For each image in the SED we initialize a background, $\{B\}^{(0)} = \{B_1, B_2, \ldots B_N\}^0$, of 0 μJy arcsec^{-2}. The stellar population of each pixel, j, is also initialized with a set of physically-motivated parameters, $\{\theta\}_j^{(0)}$.

We also define a Metropolis-Hastings step distribution for each stellar population parameter. These distributions are simply defined by standard deviations of Gaussian distributions with zero mean. Note that all pixels share the same step distribution, and these step distributions are tuned to yield a 30%–50% step acceptance rate.

Stepping in Pixel Stellar Population Parameter Space

The first phase of a Gibbs step, k, is to advance the chain for each pixel. We consider one parameter at a time, and propose an updated parameter value from the step distribution based on the current state. Then we compute the posterior likelihood given this proposed parameter value and the remaining set of current parameters. In the Metropolis-Hastings algorithm we compare the proposed, $p^{(*)}(\{\theta_1^{(*)}, \theta_2^{(k)}, \theta_3^{(k)}, \ldots \theta_n^{(k)}\}, \{B\}^{(k)}$, and current, $p^{(k)}(\{\theta\}^{(k)}, \{B\}^{(k)})$, posterior likelihoods for the pixel:

$$x = \log_{10} p^{(*)}(\{\theta_1^{(*)}, \theta_2^{(k)}, \theta_3^{(k)}, \ldots \theta_n^{(k)}\}, \{B\}^{(k)}) - \log_{10} p^{j}(\{\theta\}^{(k)}, \{B\}^{(k)}) \qquad (4.24)$$

If $x > 0$ then the proposed parameter is always accepted, otherwise the update is only accepted randomly with a probability of 10^x. In case a proposal is rejected, the current parameter value is retained for the chain's next state. This process is repeated for all remaining stellar population parameters. For example, the Metropolis-Hastings test of the proposal for the second stellar population parameter can be written as:

$$x = \log_{10} p^{(*)}(\{\theta_1^{(k+1)}, \theta_2^{(*)}, \theta_3^{(k)}, \ldots \theta_n^{(k)}\}, \{B\}^{(k)}) - \log_{10} p^{(k)}(\{\theta\}^{(k)}, \{B\}^{(k)}). \quad (4.25)$$

Note that this stellar population parameter update step occurs in parallel for each pixel.

Stepping in Global Background Parameter Space

Once the chain is advanced with respect to the stellar populations of the pixels, the last phase of the Gibbs step, k, is to update the background correction for each image (i.e., each bandpass, X), holding all stellar population parameters fixed. One could do this with a Metropolis-Hastings sampler, but realize that the background correction can also be estimated linearly from the difference of observed $F_{X,j}$, and modelled, $f_X(\{\theta\}_j)$, flux for each image j:

$$\langle B \rangle_{X,j} = F_{X,j} - f_X(\{\theta\}_j) \quad (4.26)$$

where $\langle B \rangle_{X,j}$ is the estimated background correction for pixel j in the bandpass X. The mean and variance of these background estimates define a Normal distribution; from that Normal distribution we sample a new background estimate:

$$B_X^{(k+1)} \leftarrow \mathrm{N}\left(\frac{\sum_j \frac{\langle B\rangle_{X,j}}{\sigma^2_{X,j}}}{\sum_j \frac{1}{\sigma^2_{X,j}}}, \frac{1}{\sum_j \sigma^{-2}_{X,j}}\right). \tag{4.27}$$

Parameter Estimation

In a Gibbs MCMC sampler the process of stepping through parameter space is repeated, typically more than 10^3 times. Each step of the sampler yields an instance of both the pixel stellar population parameters, $\{\theta\}_j^{(k)}$, and background estimates, $\{B\}^{(k)}$. To estimate any parameter, one simply computes the mean value of the parameter values from the ensemble of steps, often called the *chain*. This is equivalent to computing the integral

$$\langle\theta\rangle = \int \theta p(\theta|F)\mathrm{d}\theta. \tag{4.28}$$

Recall that MCMC samples the parameter space proportionately to the posterior likelihood. Thus taking the mean of the parameter chain is equivalent to computing a posterior likelihood-weighted integral of the parameter space. This takes advantage of the MCMC method's ability to trivially marginalize dimensions in a parameter space by simply omitting those dimensions while analyzing the chain.

One caveat is that an MCMC will initially take many steps to transition from the initialized parameter space to the true distribution. We omit this burn-in period by omitting the first half of the chain from parameter estimation.

To estimate the uncertainty of an estimated parameter value, one could also compute the sample standard deviation of the chain. However, to consider non-Gaussian distributions, it is more appropriate to construct a two-sided *confidence interval* from predefined quantiles. For this study, we define confidence intervals within the 20%–80% percentiles of the distribution.

4.4 A Gibbs Markov Chain Monte Carlo Sampler with FSPS

4.4.1 Implementation with Flexible Stellar Population Synthesis

In the previous section we laid out a hierarchical Bayesian model for estimating backgrounds in a set of images by modelling the SEDs of many pixels simultaneously. Here, we detail our choices for implementing that algorithm in the case of the ANDROIDS data set. Then in § 4.5.2 we will review the results.

An implementation of this algorithm is available in our open-source `sedbot` Python package[2] (in particular, see the '`multipix`' sub-package). Our code is designed such that one can construct and estimate a model for any arbitrary pixel dataset.

4.4.2 Population Synthesis

Central to our implementation of the hierarchical background modelling approach is the Flexible Stellar Population Synthesis package (FSPS, Conroy et al. 2009, 2010; Conroy & Gunn 2010), which is used to compute a model SED, $f(\{\theta\})$, for each pixel during each Gibbs step. FSPS distinguishes itself from other population synthesis packages for several reasons. First, FSPS includes modern calibrations of AGB light (see the aforementioned series of publications), which is crucial for the optical-NIR baseline of ANDROIDS. Second, FSPS was built as an *application programming interface* (API) so that it is not only easy, but also necessary, to directly call FSPS's

[2] Available at `https://github.com/jonathansick/sedbot`.

functions when computing synthetic stellar populations. This has an enormous performance benefit since all computations and communication are done in memory. In fact, FSPS even caches SSP spectra in memory from one computation to the next. Besides this performance benefit, FSPS's API-oriented design is also extremely convenient. We have built the `python-fsps` (Foreman-Mackey et al. 2014) Python wrapper for FSPS using the `F2PY` code (Peterson 2009). Thanks to this wrapper, `sedbot` is able to use FSPS as it is were a regular Python object.

Isochrones and IMF We use FSPS with the default set of 'Padova' isochrones (Marigo et al. 2008) with the BaSeL spectral library (Westera et al. 2002). The initial mass function (IMF) is set by Chabrier (2003), which is appropriate for Milky Way-like disk galaxies. Note that the IMF primarily serves to normalize the stellar mass, though in principle IMF differences can affect the shape of the SED as well by changing the rate that stars are fed into luminous post main-sequence phases (Conroy 2013).

Photometric Systems FSPS supplies model absolute magnitudes in the AB system (Oke & Gunn 1983). We convert these magnitudes into fluxes (adopting a distance modulus to M31 of 24.45 mag, McConnachie et al. 2005):

$$f_{\mathrm{FSPS}} = 3631 \times 10^{6-0.4(m_{\mathrm{FSPS}}+24.45)}\ [\mu\mathrm{Jy}]. \tag{4.29}$$

Dust FSPS computes SEDs with dust attenuation and emission self-consistently. We use the power-law dust attenuation curve calibrated by Conroy & Gunn (2010).

Following Charlot & Fall (2000), dust attenuation for young and old stellar populations are parameterized separately since young populations are typically more embedded in dust than older stars. We use `dust1` and `dust2` to parameterize the optical depth for young and old stars, respectively.

Metallicity FSPS, with the Padova isochrones and BaSeL spectral library, provides 23 discrete stellar metallicity values, in the range $\log Z/Z_{\odot} = (-1.98, +0.20)$. Each synthetic stellar population is characterized by a single metallicity; no metallicity evolution is considered within a pixel. To approximate a continuous metallicity distribution, `sedbot` computes a given stellar population twice, at discrete metallicity levels bracketing the desired metallicity. Then the synthetic SED, along with derived quantities such as mass-to-light ratio, are linearly interpolated to the intended metallicity.

Star Formation History We parameterize the star formation history as a linear combination of constant and exponentially-declining star formation rates:

$$\mathrm{SFR(t)} = (1-C)\,\frac{e^{-t/\tau}}{\int_{t=t_0}^{t_f} e^{-t/\tau}\mathrm{d}t} + \frac{C\,(t_f - t_0)^{-1}}{\int_{t=t_0}^{t_f} C\,(t_f - t_0)^{-1}\,\mathrm{d}t}, \tag{4.30}$$

where τ is the e-folding time of the star formation rate (in Gyr) and C is the fraction of stellar mass formed at a constant rate. The times after the Big Bang when star formation started and the age of the universe when the stellar population is observed are denoted by t_0 and t_f, respectively. We adopt $t_f = 13.7$ Gyr, though the results are not sensitive to the exact age of the Universe due to the t_0 parameter. Note that FSPS can also add a burst mode of star formation at an arbitrary time with an arbitrary amplitude. We choose not to consider bursts in this Gibbs modelling for simplicity.

SEDs computed by FSPS are normalized such that 1 $\mathrm{M}_{\odot}$ of stars are formed between t_0 and t_f. Therefore the total stellar mass in a pixel, $\mathcal{M}^*$, is an additional parameter to the star formation history. Some stellar population modellers linearly solve for the $\mathcal{M}^*$ factor that minimizes the χ^2 metric between the observed and modelled SED (e.g. Taylor et al. 2011). However, we simply include M_* in the set of stellar population parameters for each pixel, $\{\theta\}_j$.

In that case, the modelled flux f_X of Eq. 4.23 can be re-written for this specific implementation as

$$f_X(\{\theta\}_j) = \mathrm{M}_{*,j} f_{X,\mathrm{FSPS}}\left(\{t_0, t_f, \tau, C, \mathrm{dust1}, \mathrm{dust2}, \log Z/Z_{\odot}\}_j\right). \tag{4.31}$$

4.4.3 Model Initialization and Priors

For every pixel, our Bayesian model implemented with FSPS requires seven parameters, in addition to a background parameter for each bandpass. Each of these parameters requires an initial value, and with the exception of background level, also requires a prior probability distribution and a Metropolis-Hastings step distribution.

Our choices are summarized in Table 4.1. In general the priors are intended to be *uninformative*; the uniform distribution serves primarily to provide hard bounds on the parameter space. The initial values, likewise, are chosen to be reasonable. Most interesting is our treatment of stellar mass. Since stellar mass is strongly correlated to surface brightness, and ANDROIDS pixels span a wide range of surface brightnesses, we choose to set all prior distributions and initialization values from the i-band mass-to-light ratio. The parameter is still sampled, however, in the space of stellar mass

Table 4.1: Parameters of the FSPS hierarchical background model. U indicates a uniform distribution, bounded by the arguments. N indicates a normal distribution with mean zero and standard deviation set by the argument in the units of the parameter. Note that the prior and initialization value for stellar mass is from the i-band stellar mass-to-light ratio, $\mathcal{M}^*/L_i$; the step distribution remains in units of stellar mass.

Parameter		Prior	Initialization	Step
$\mathcal{M}_*$	Stellar Mass ($M_\odot$)	$U_{\log \mathcal{M}^*/L_i}(-0.5, 1.5)$	$\log \mathcal{M}^*/L_i = 0.5$	$N(0.1)$
$\log Z/Z_\odot$	Stellar Metallicity	$U(-1.98, 0.2)$	-0.1	$N(0.3)$
t_0	Start of SF (Gyr)	$U(0.5, 10)$	3	$N(1.0)$
$\log \tau$	E-folding time of SF (log Gyr)	$U(-1, 2)$	1	$N(0.6)$
C	Mass fraction from constant SFR	$U(0, 1)$	0.1	$N(0.5)$
`dust1`	Young star dust optical depth	$U(0, 5)$	1.0	$N(0.8)$
`dust2`	Old star dust optical depth	$U(0, 3)$	0.2	$N(0.3)$
B_X	Background per band (μJy arcsec^{-2})	—	0	—

itself.

The Metropolis-Hastings step distributions for each parameter were chosen to be normal distributions with a mean value of zero. For each parameter, we tuned the step size so that the sampler would move freely in parameter space while maintaining an overall ~ 0.3–0.5 acceptance fraction.

4.5 Application to the ANDROIDS Dataset

We apply the hierarchical SED modelling algorithm, as implemented in § 4.4 to the ANDROIDS MegaCam and WIRCam datasets. First, we describe how the input SEDs are prepared (§ 4.5.1). Then in § 4.5.2 we describe the results of background modelling to the ANDROIDS photometry.

4.5.1 SED Dataset Preparation

We model ANDROIDS MegaCam $u^* g' r' i'$ and WIRCam JK_s mosaics to model their photometric background biases. This section describes how these datasets are reduced into SEDs that can be modelled in the framework of Sections 4.3 and 4.3.3.

Input Mosaics

We begin with full-resolution WIRCam and MegaCam mosaics described in Chapters 2 and 3, respectively.

The ANDROIDS WIRCam J and K_s mosaics consist of images from the 27 fields of the 2007B survey and 11 fields of the 2009B survey. The footprint extends to 20 kpc on M31's major axis. As described in Chapter 2, backgrounds in these mosaics are initially calibrated with a sky-target nodding method. Then a sky-offset optimation method further improved the *consistency* of background calibration by reducing image-to-image surface brightness discontinuities.

ANDROIDS MegaCam u^*, g', r', and i' mosaics are composed of images covering 14 pointings, extending out to 40 kpc along M31's major axis. The Elixir-LSB observing strategy provided excellent background subtraction through sky-target nodding. Sky offset optimization was still used to reduce image-to-image surface brightness discontinuities, although less necessary than with WIRCam J and K_s mosaics.

We resampled all six mosaics to a common pixel grid, while also reducing the pixel scale to $3''$. Higher resolution photometry would not benefit this background modelling experiment. This resampling was performed with *Swarp* (Bertin 2007) using a procedure described in § 5.3.3.

Masking

We mask all pixels containing non-M31 light. In particular, we mask Milky Way halo stars based on 2MASS Point Source Catalog (Skrutskie et al. 2006) sources with $J - K_s < 0.8$. Metal poor Milky Way stars are readily distinguished from more metal rich M31 stars with such a selection. Brighter Milky Way stars also leave diffraction spikes and internal reflection halos in MegaCam images. We manually built polygon masks that mark areas of mosaics contaminated by these effects. All masked pixels are excluded from aggregation in the mosaic binning, described next.

Wedge Segmentation

Recall that a central requirement of the hierarchical SED model is that many pixels must be modelled simultaneously to obtain a unique constraint on the background. Specifically, we require that the pixels share a *common* background bias, while the pixels also span a wide range of stellar populations. The diversity of stellar populations guards against model degeneracies, while the commonality of background bias is a limitation of the scalar background model in Eq. 4.26. At the same time, processing times scale roughly linearly with the number of pixels as each pixel introduces a new round of θ-level Gibbs steps.

We designed a radial wedge binning scheme for the ANDROIDS M31 images as a compromise to all of these factors. This segmentation strategy traces the projected M31 disk, dividing the image into 36 projected angular wedges, divided into radial segments of variable width. Along each wedge, we measure the deprojected galactocentric radius of pixels in the disk. Given that M31 has central coordinates (α_0, δ_0), inclination i, and position angle ϕ, we can measure galactocentric distances of an arbitrary pixel in the disk with coordinates (α, δ) by first projecting into a rectangular sky coordinate system:

$$X = (\alpha_0 - \alpha)\cos\left(\frac{\delta + \delta_0}{2}\right) \tag{4.32}$$

$$Y = \delta_0 - \delta. \tag{4.33}$$

We then rotate into a frame aligned with the disk major axis:

$$\phi = \phi_0 - \frac{\pi}{2} + \tan^{-1}\left(\frac{y}{x}\right) \tag{4.34}$$

$$X' = r_{\rm sky} \times \cos\phi \tag{4.35}$$

$$Y' = r_{\rm sky} \times \sin\phi \tag{4.36}$$

where $r_{\rm sky}$ is the angular separation between (α, δ) and (α_0, δ_0) in the plane of sky.

Then we de-project the rotated y' coordinate:

$$y'' = \frac{y'}{\cos i} \tag{4.37}$$

and measure a galactocentric distance on the disk:

$$R_{\rm maj} = d_{\rm M31}\tan\left(\sqrt{(x')^2 + (y'')^2}\right) \ \text{[kpc]}. \tag{4.38}$$

Within each wedge, pixels are binned along a radial axis with bin widths specified by

$$\Delta R_{\mathrm{maj}}\ [\mathrm{kpc}] = \begin{cases} 0.1 & \text{if } R_{\mathrm{maj}} < 0.1\ \mathrm{kpc} \\ 0.5 & \text{if } \leq 1\ \mathrm{kpc} \leq R_{\mathrm{maj}} < 20\ \mathrm{kpc} \\ 1 & \text{if } \leq 20\ \mathrm{kpc} \leq R_{\mathrm{maj}} < 50\ \mathrm{kpc}. \end{cases} \tag{4.39}$$

In the multipixel hierarchical model, we find that restricting the radial domain to 30 kpc is an effective compromise between including a wide span of stellar populations, and ensuring sufficient stellar light contribution to each pixel. As such, our multipixel models of single wedges through M31 are computed with 58 SEDs.

Flux Measurement

Within each segmented region, we extract flux measurements in each band (mosaic). The area, A_{bin}, is the area covered by pixels in the segmentation region, whether masked or not. Within each region, for a given mosaic, we compute the median among unmasked pixels, $\overline{\mathrm{DN}}_{\mathrm{bin}}$. The uncertainty is estimated by combining uncertainties of individual pixels: $\sigma_{\mathrm{bin}} = \sqrt{\sum \sigma_i^2 / n_{\mathrm{pixels}}}$. These measurements, in the DN units of mosaics, are converted to fluxes via:

$$F_\nu = A_{\mathrm{bin}} \times \overline{\mathrm{DN}}_{\mathrm{bin}} \times 10^{-0.4(m_{\mathrm{zp}}+48.6)} \times 10^{23} \times 10^{6}\ [\mu\mathrm{Jy}], \tag{4.40}$$

where m_{zp} is the mosaic's zeropoint. This practice of multiplying the median pixel intensity by the total area of pixels effectively interpolates over masked pixels. These fluxes are further corrected for Milky Way foreground extinction according to the prescription given by Schlafly & Finkbeiner (2011) with $R_V = 3.1$ and $E(B-V) = 0.07$ (Schlegel et al. 1998). Section 5.3.5 provides additional details.

4.5.2 Background Modelling Results

We have used the hierarchical multipixel SED model to estimate scalar backgrounds associated with the ANDROIDS $ugriJK_s$ images. Models are computed independently for each wedge (Figure 4.7), yielding background estimates for 36 position angles radiating from M31's centre in the ANDROIDS data.

Radial surface brightness profiles of disk galaxies like M31 tend to have well-ordered profiles for each bandpass; from blue as the dimmest to the near-infrared as the brightest. Although these profiles have different shapes, implying colour (and thus stellar population) gradients, sudden changes in colour gradients are evidence of background subtraction errors. Our first modelling attempts showed that the hierarchical model can be successful at restoring well-ordered, physically plausible profiles through scalar background correction. This success, however, comes at the cost of large systematic changes to M31's light profiles.

Fixing the g' Background to Zero

To prevent systematic surface brightness drift, we ran models where one profile would be considered 'correct' and have a fixed-0 background correction. Ideally one could use an externally-calibrated SB profile to act as a reference, but none such are available (§ 4.2). Instead, we postulated that the ANDROIDS g'-band could itself be a reasonable description of M31's true SB profile since the ANDROIDS g'-band agrees with WK87 (see Figure 4.1).

An example application of the modelling with $B_{g'}$ is shown in Figure 4.5. In the original profile, Figure 4.5(a), the K_s profile clearly has an over-subtracted background, as does the i' profile at $\mu_{i'} \sim 24$ mag arcsec^{-2}. By comparison, the g'

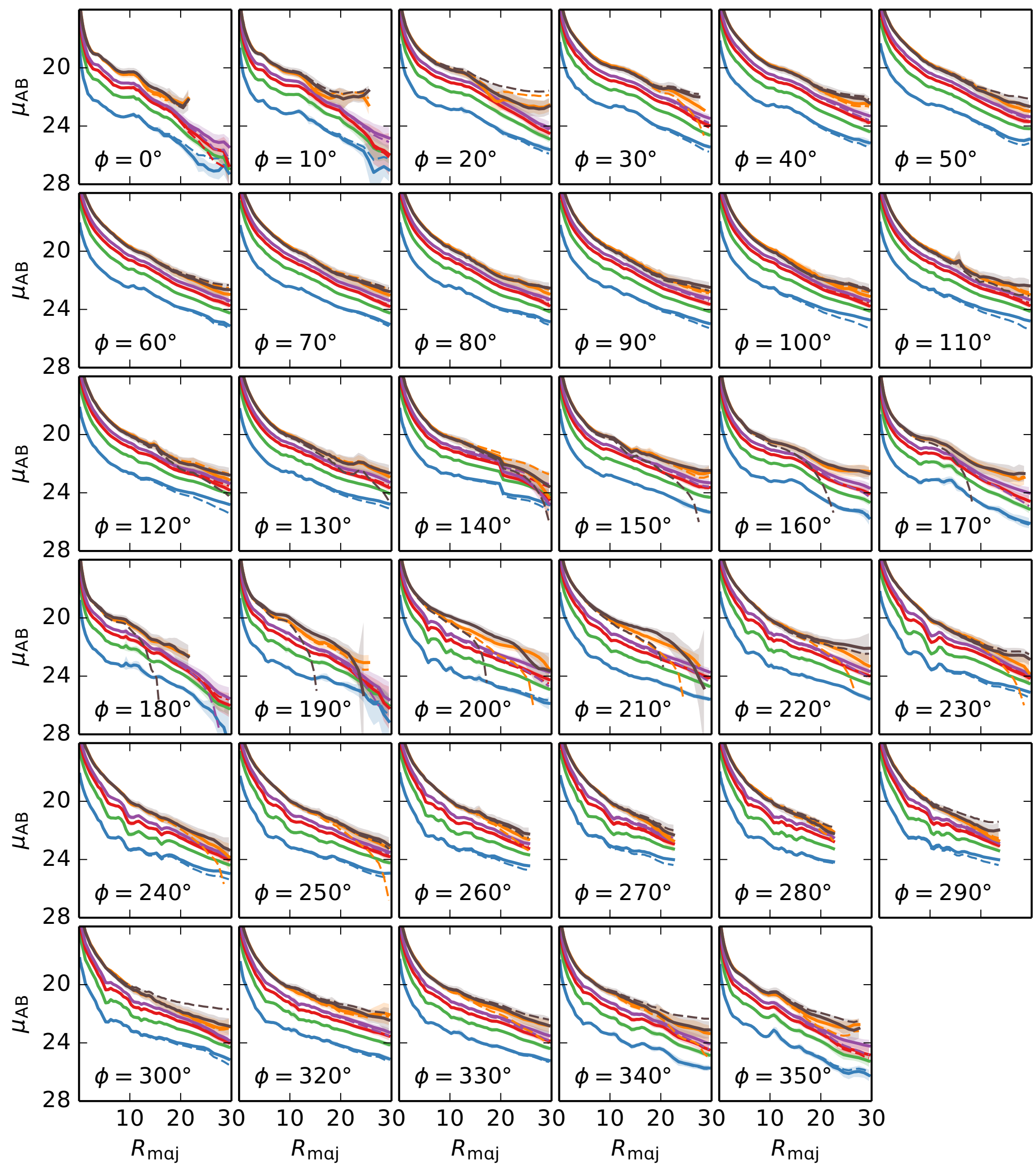
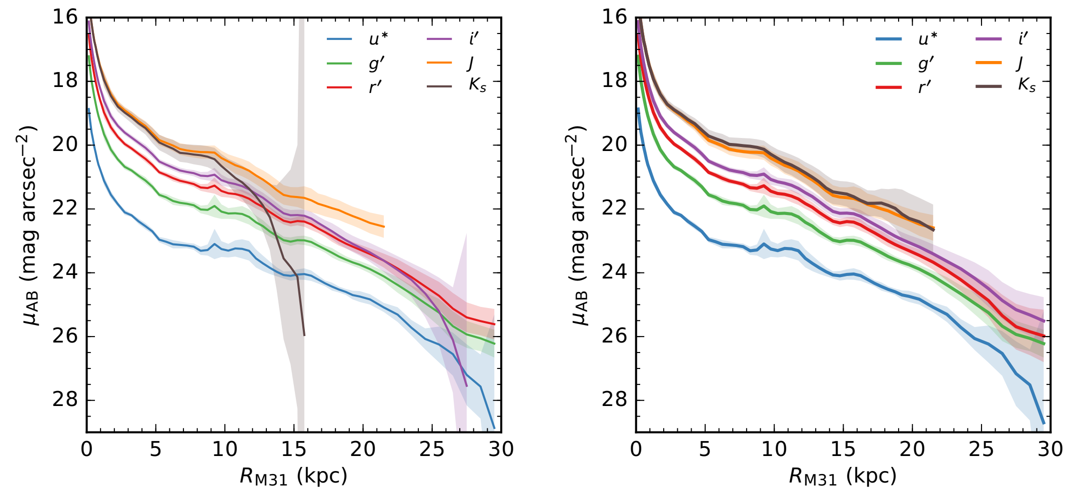


Figure 4.4: Surface brightness cuts given hierarchical background modelled with a fixed g'-band. Original (dashed) and calibrated (solid) surface brightness cuts are shown at all position angles around M31 in u^* (blue), g' (green), r' (red), i' (purple), J (orange), and K_s (dark grey). Horizontal axis is de-projected disk radius, in kpc. Background modelling alleviates some suspicious profile shapes, particularly in J and K_s. This figure also shows that the modelling does not impose flat colour gradients, but rather ensures that colours are permitted by stellar population models.



(a) Original SW major axis surface brightness profile

(b) Corrected SW major axis surface brightness profile with background where $B_{g'} = 0$.

Figure 4.5: SW major axis surface brightness profiles without (left) and with (right) background correction via the hierarchical background modelling algorithm. In these models the g' surface brightness profile is fixed to a null surface brightness correction.

profile does not show any suspicious behaviour with a near-exponential profile out to $R_{\mathrm{maj}} = 30$ kpc. Assuming $B_{g'} = 0$, we applied the hierarchical background model to these surface brightness profiles to yield Figure 4.5(b). We find that this background correction is stable, evidenced by the MCMC parameter chains shown in Figure 4.6. In this corrected dataset, all bandpasses have well-ordered profiles, with approximately equal gradients. This case clearly exhibits the potential of hierarchical modelling with even a scalar background correction for self-calibrating galaxy light profiles.

We applied the hierarchical background models (where $B_{g'} = 0$) to all 36 wedges around M31. A grid of corrected profiles is shown in Figure 4.4. In almost all cases the light profiles are improved by the calibration, though the absolute calibrations cannot be guaranteed correct.

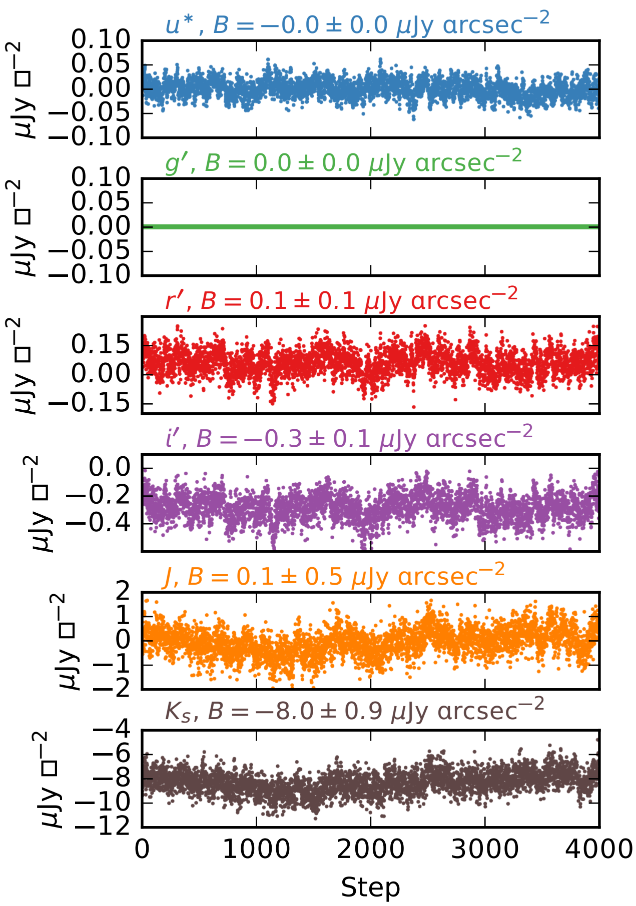


Figure 4.6: Gibbs MCMC chain of background parameters corresponding to Figure 4.5. The g' background is fixed to 0.

Fixing the $u^*g'r'$ Backgrounds to Zero

In the previous section, we found that setting $B_{g'} = 0$ is useful for constraining hierarchical models for bands where the background biases are believed to be negligible. Effectively, the constraint ensures that stellar population models match the constrained g' band, thereby reducing the available parameter space for backgrounds in other bands. We can improve this by constraining other blue ANDROIDS bands that we are confident in being background bias-free. The u^* and r' bands already

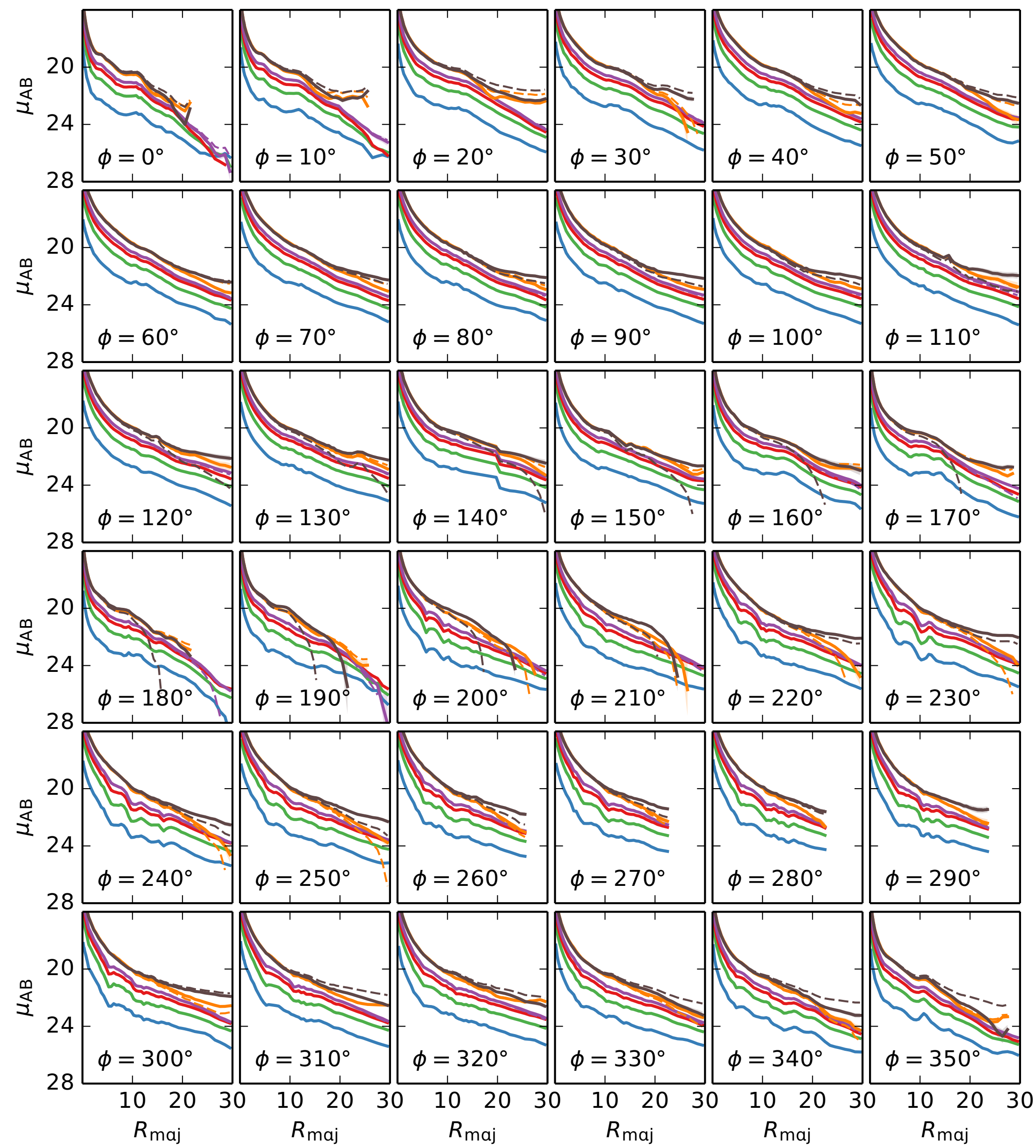


Figure 4.7: Surface brightness cuts given hierarchical background modelled with fixed $u^*g'r'$ bands. Original (dashed) and calibrated (solid) surface brightness cuts are shown at all position angles around M31 in u^* (blue), g' (green), r' (red), i' (purple), J (orange), K_s (dark grey).

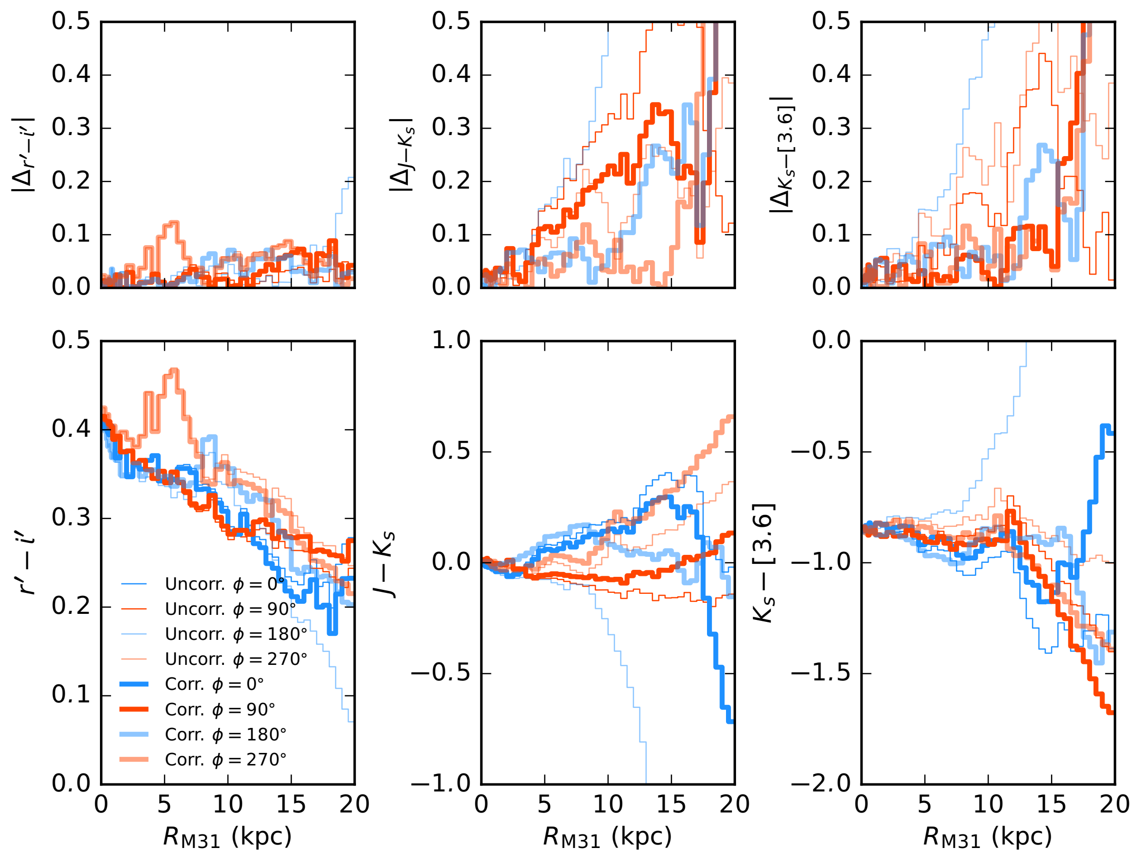


Figure 4.8: M31 radial colour profiles before and after background corrections, corresponding to colour maps shown in Figure 4.9. Across the bottom row, each line is a colour profile computed along a wedge (§ 4.5.1) following the major axis (blue, $\phi = 0°$), minor axis (orange, $\phi = 90°$), the anti-major axis (light blue, $\phi = 180°$), or anti-minor axis (light orange, $\phi = 270°$). The top row shows absolute residuals, compared to the major axis profile. Uncorrected profiles are plotted with thin lines, while profiles corrected according to § 4.5.2 (where $u^*g'r'$ are fixed with no background corrections in the models) are plotted with thick lines. Assuming that the M31 disk is azimuthally symmetric, to first order, colour profiles should be similar at each position angle (particularly the major and anti-major axes, which are less affected by dust attenuation than minor axis cuts). Although the background modelling described in § 4.5.2 substantially improves colour profile symmetry, the amount of profile-to-profile colour variation that persists after correction reflects an upper-limit to the accuracy of ANDROIDS colour profiles, particularly in the NIR in the outer disk.

have minimal estimated background corrections in Figure 4.4, and neither do we expect these bands to require substantial background corrections (the optical blue sky is dimmer than the near-IR sky).

In this section, we discuss hierarchical SED modelling where we constrain $B_{\{u^* g' r'\}} = 0$. The outcome of this modelling is shown in Figure 4.7. Compared to the profiles with $B_{g'} = 0$ models shown in Figure 4.4, profiles modelled with $B_{\{u^* g' r'\}} = 0$ have qualitatively smaller uncertainty intervals. Some J and K_s profiles continue to unexpectedly cross profiles of bluer bands. Backgrounds corrected using hierarchical modelling are an improvement, but not a complete solution.

We can more quantitatively assess the surface brightness accuracy of the ANDROIDS mosaics by analyzing the azimuthal symmetry of colour profiles after background corrections have been applied. M31 surface brightness, and colour, profiles should be symmetric, to first order, around M31's disk if we assume that a disk galaxy's structure and stellar populations are primarily radial functions. Any deviations from symmetry may reflect a possible background calibration bias. Here we specifically use colour (magnitude difference) profiles rather than surface brightness profiles as they are both slowly varying radial gradients, and obviate the need to deproject minor wedge profiles to match the major axis. In Figure 4.8, we have plotted $r' - i'$, $J - K_s$, and $K_s - [3.6]$ colour profiles corresponding to wedges (§ 4.5.1) along the major and anti-major axes, and the minor and anti-minor axes (the [3.6] photometry is obtained from Spitzer IRAC, and formally introduced in Chapter 5). Figure 4.8 also shows absolute residuals of these profiles against the major axis's colour.

The $r' - i'$ profile in Figure 4.8 is an example of a well-calibrated colour profile. At each disk position angle, the profiles are consistent within $\lesssim 0.1$ mag. The background correction also had a minimal effect on the i' profile as the before-and-after-correction $r' - i'$ profiles are similar, though the background correction does successfully resolve the 0.2 mag asymmetry that appears across the major axis near $R_{\rm M31} \sim 20$ kpc, seen in Figure 4.8.

The $J - K_s$ profiles reflect the challenging nature of ANDROIDS background correction. Before background corrections, the $J - K_s$ profile is unusable: at $R_{\rm M31} = 10$ kpc, the major and anti-major axis $J - Ks$ asymmetry is 0.3 mag. After background correction, the major-axis $J - K_s$ colour asymmetry is resolved to be < 0.1 mag within $R_{\rm M31} < 10$ kpc. However, there is still asymmetry between the major and minor axes at a level of 0.2 mag at $R_{\rm M31} = 10$ kpc. Since the IRAC 3.6 μm image is free of background biases, the $K_s - [3.6]$ colour profile helps distinguish the reliability of J and K_s images. In Figure 4.8, $K_s - [3.6]$ is symmetric at a level of < 0.1 mag across all axes within $R_{\rm M31} < 12$ kpc. Generally, beyond $R_{\rm M31} = 10$ kpc all $J - K_s$ colours are unreliable at a level of 0.2 mag or greater after background correction.

An assumption made in the background calibration method is that the background is constant within a wedge, provided that backgrounds vary at large scales across the mosaics. Figure 4.9 shows $r' - i'$, $J - K_s$, and $K_s - [3.6]$ colour maps before and after background correction. Figure 4.9(a) bears out this assumption as colour maps of uncorrected mosaics have large scale bias tends. The uncorrected $J - K_s$ and maps are biased blueward approximately parallel to the southern major axis (similarly, $K_s - [3.6]$ maps are biased redward, suggesting that the K_s mosaic is primarily responsible for the colour bias). Qualitatively, this background bias is resolved with hierarchically-modelled background corrections (Figure 4.9(b)).

Figure 4.10 shows the background corrections computed for each wedge in the

i', J and K_s bands. The azimuthal distribution of these background corrections explains Figure 4.9 effectively. K_s-band corrections are oriented about the major axis, while J-band corrections are smaller and oriented independently. Most importantly, Figure 4.10 shows continuity in background corrections between adjacent wedges. This continuity is not enforced in modelling, and arises spontaneously from the data. Backgrounds in the ANDROIDS mosaics *do* vary on large scales, making the wedge-based approach to background calibration effective.

4.6 Quantifying Surface Brightness Accuracy Requirements

The previous section demonstrated the application of the hierarchical stellar population modelling method for calibrating the backgrounds in ANDROIDS mosaics. On the basis of disk symmetry, the method reduces background bias in surface brightness profiles. Within $R_{\mathrm{M31}} < 10$ kpc, $J - K_s$ azimuthal asymmetry is reduced to less than 0.2 mag. This analysis does not address the more fundamental question: are ANDROIDS mosaics now sufficiently well-calibrated to support astrophysically-motivated stellar population estimation?

Figure 4.11 provides one perspective on this, where we show colour-colour diagrams of $u'g'r'i'JK_s[3.6]$ SEDs extracted from the wedge segmentation, before and after background correction. The Spitzer IRAC 3.6 μm dataset is formally introduced in Chapter 5, but it is useful here since it is a near-infrared dataset with a well-calibrated background that provides a stable fiducial for near-IR colours. As expected, Figure 4.11 shows that WIRCam $J-K_s$ colours benefit most from background corrections. Before background corrections, $J - K_s$ colours varied by ~ 0.5 mag at

$R_{\mathrm{M31}} = 10$ kpc. With background corrections, the $J - K_s$ colour variations are reduced to ≤ 0.2 mag at $R_{\mathrm{M31}} = 10$ kpc. Colours composed of r', i' and [3.6] colours, alone, are largely unchanged by background modelling. For reference, these colour distributions are plotted against grids of simple stellar populations (SSPs) modelled with FSPS. It is plainly visible that constraining colour distributions is important for interpreting the stellar populations of SEDs, given how rapidly stellar populations shift in colour compared to the distribution of observations.

In detail, the accuracy of stellar population estimation is non-trivial to understand. However, to first order we can consider a stellar population's colour (such as $J - K_s$) as a function of one stellar population parameter, all other parameters being fixed. In such a constrained environment, stellar population estimation is done by measuring the stellar population's colour and looking-up the stellar population parameter that corresponds to that colour. Now we can ask how much the looked-up parameter is shifted if the measured colour is shifted, to simulated a background subtraction bias.

Results from such an experiment are shown in Figure 4.12. In the top panel, the $J - K_s$ colours of SSPs with fixed $A = 6$ Gyr age for a grid of $\log Z/Z_{\odot}$ metallicities with FSPS. All other stellar population parameters are fixed. There is no dust attenuation in these models. For each metallicity $(\log Z/Z_{\odot})_i$ we identify the $(J - K_s)_i$ colour. Then we bias the colour by a fixed amount $\Delta_{J-K_s} = (J - K_s)_j - (J - K_s)_j$ and find, via numerical optimization, the corresponding metallicity $(\log Z/Z_{\odot})_j$. The shift $\Delta_{\log Z/Z_{\odot}} = (\log Z/Z_{\odot})_j - (\log Z/Z_{\odot})_i$ is representative of the metallicity estimation bias induced by a photometric calibration bias. The same experiment is repeated in the bottom panel of Figure 4.12, but for SSPs of different ages with a fixed solar $(\log Z/Z_{\odot} = 0)$ metallicity.

Stellar populations of different ages have similar $J - K_s$ colours. Figure 4.12 underscores that any realistic uncertainty in observed $J - K_s$ colour renders age estimation impractical.

Metallicity (Figure 4.12, top) is a more interesting case. The near-infrared is dominated by light from asymptotic giant branch (AGB) and red giant branch (RGB) stars. The $J - K_s$ colour of the AGB and RGB becomes monotonically redder with increased metallicity. Combined with near-infrared colours being minimally reddened by dust, $J - K_s$ should be an ideal stellar population metallicity indicator. However, even with $J - K_s$ photometry calibrated to within 0.05 mag, metallicity is uncertain to at least ± 0.2 dex at near-solar, or ± 0.5 dex at low metallicities. Androids $J - K_s$ photometry at $R_{\mathrm{M31}} = 10$ kpc is likely only certain to ± 0.2 mag. At high metallicities this represents a metallicity uncertainty of ± 0.9 dex and ± 1.5 dex (or indeed, is limited physically to the metallicity domain). This result suggests that accurate stellar population estimation requires surface brightness calibration that is better than even the current hierarchical stellar population and background modelling method can provide.

Of course, the stellar populations are not estimated using the approach described here. Stellar populations are multi-variate, making this experiment a lower bound on the stellar population estimation accuracy that can be expected given $J - K_s$ photometry. Real stellar population estimation also uses information from multiple colours. While this alleviates the case of $J - K_s$ alone being a poor predictor for metallicity, this experiment still places limits on how informative near-infrared photometry can be expected to be in estimating metallicity, where $J - K_s$ colour *is* a strong predictor.

4.7 Conclusions

Understanding the background in our wide-field optical and near-infrared mosaics of M31 is a substantial challenge for ANDROIDS, and is prerequisite for applying this dataset towards understanding M31's stellar populations. The challenge that ANDROIDS faces is niche: the M31 disk fills 14 MegaCam fields and 27 WIRCam fields, preventing direct background estimation and subtraction that is typically used for extragalactic observations where the galaxy is much smaller than the detector's field. Despite unprecedented sky-target nodding and optimization of image-to-sky surface brightness variations with sky offsets, ANDROIDS mosaics, primarily in near-infrared JK_s, are not calibrated sufficiently for stellar population modelling. See Figure 4.9(a) for a clear demonstration.

In this chapter we have explored how to resolve the background calibration of ANDROIDS mosaics, building upon existing sky-target nodding and sky-offset optimization calibrations. First, in § 4.2 we attempted to find reference surface brightness profiles from the literature that could externally-calibrate ANDROIDS mosaics. We also explored synthesizing surface brightness maps from Hubble PHAT star catalogs. None of these datasets are suitable for calibrating ANDROIDS images, having either a lack of near-infrared coverage, or suspect surface brightness calibrations themselves.

Instead, we developed an alternative method of calibrating backgrounds by taking advantage of stellar population synthesis modelling. This method, developed in § 4.3, supposes that the SEDs of a group of pixels can be modelled as an independent stellar population in each pixel and a shared background brightness per bandpass. Such a model can be formally written as a hierarchical Bayesian model, with one level being the common scalar backgrounds per band, and the other being stellar population synthesis parameters. In § 4.3.3 we found that such a hierarchical model, while possessing a large number of dimensions, can be estimated with a Gibbs Markov Chain Monte Carlo sampling method. The backgrounds are sampled with a linear estimator, and thus we found that the model chains converged quickly (Figure 4.6).

Correcting mosaics with backgrounds estimated through this method, we found a definite reduction in ANDROIDS J and K_s surface brightness bias as measured through azimuthal symmetry of colour profiles (Figure 4.8). Although the method scales to an arbitrary number of bandpasses, we found that the models are better constrained if that background model parameter is fixed to zero in bands where any background bias is likely negligible. We ultimately configured the models to assume that backgrounds in the blue u^*, g', and r' bands is fixed at zero. A background is modelled in i', though it is a nearly negligible correction.

A key assumption in the hierarchical background modelling method is that a group of pixels with diverse stellar populations must share a common background bias, but have diverse stellar populations and surface brightnesses. Otherwise, parameters for a uniform stellar population are degenerate with background parameters. Our solution was to segment the M31 disk into 36 projected wedges, and model backgrounds in each wedge with radially-segmented pixels. Though sampling pixels across a radial cut ensures a variety of stellar populations and surface brightnesses are included, this approach assumes that the backgrounds vary across ANDROIDS mosaics on scales larger than the wedges. This assumption was validated since the estimated backgrounds in each wedge are highly correlated from one position angle to the next (Figure 4.10).

Using azimuthal symmetry of disk colour profiles as an estimator for surface brightness calibration bias, we found that at $R = 10$ kpc, the ANDROIDS $J - K_s$ colour is only constrained to within ± 0.2 mag after background corrections (Figure 4.8). Through a simplified stellar population estimation experiment described in § 4.6, we found that such a bias translates into a metallicity estimation bias of about 1 dex. Stellar population estimation with a panchromatic SED, not just $J - K_s$ colour, is certainly more powerful. However, this experiment does reflect a sobering reality that ANDROIDS near-infrared photometry may not be informative of stellar populations in the mid and outer M31 disk. On the other hand, ANDROIDS $u^* g' r' i'$ photometry appears very well calibrated.

With this caveat in mind, in the next chapter we augment the six-band ANDROIDS dataset with additional datasets. Then in Chapter 6 we model M31's stellar populations with panchromatic integrated light photometry, and attempt to explore the limitations of galaxy SED modelling in greater detail.

4.8 Appendix: SED Library-in-Gibbs Sampling

As an addendum to the Metropolis-Hastings in Gibbs approach to hierarchical background modelling, we also propose an alternative algorithm that is not yet implemented. Notice that the stepping algorithm described in § 4.3.3 re-synthesizes an SED for each parameter, repeated for each pixel, in a single Gibbs step. Though a technically correct algorithm, this approach is somewhat computationally taxing. In Chapter 6 we explore library-based stellar population estimation, which retains the advantages of a Bayesian framework, while effectively caching and re-using SED computations for later re-use. Briefly, library-based stellar population estimation works by pre-computing a library of stellar population models. The distribution of models in the library reflects our prior knowledge of the stellar population. For example, the prior probability distribution functions listed in Table 4.1 could be used to create the model library. Note that the stellar population library omits any background parameters; the SEDs in the library assume $B = 0$. The role of the stellar population library is to replace Metropolis-Hastings steps.

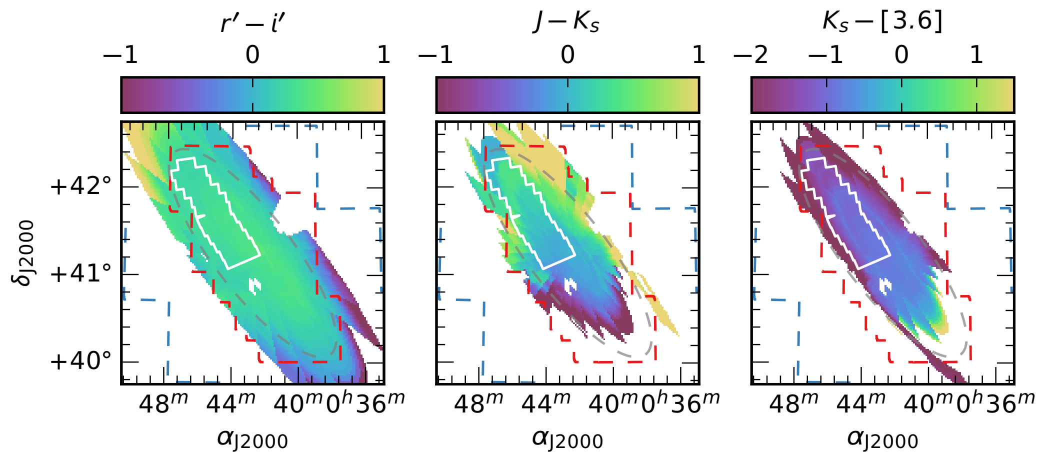
The library-based stellar population is integrated as follows in a Gibbs sampler to estimate backgrounds in an observed SED:

1. Initialize the background SED for all pixels as $\{B\}^{(0)} = 0$.
2. Estimate the stellar population of each pixel. This is done by marginalizing each pixel's SED adjusted by the current background estimate, $\{F\}_j - \{B\}^{(k)}$, against the model SED library. Marginalization yields an estimate of stellar population parameters for each pixel, $\theta_j^(k)$ along with uncertainties.
3. A background is estimated in each band of each pixel as the difference of the modelled SED and the observed SED, as in Eq. 4.26. As in Eq. 4.27, these estimates define a normal distribution from which a new background SED is sampled, $B^{(k+1)}$.
4. Steps 2–4 are repeated a set number of times with the background being estimated as the mean of the background estimate chain, see Eq. 4.28.

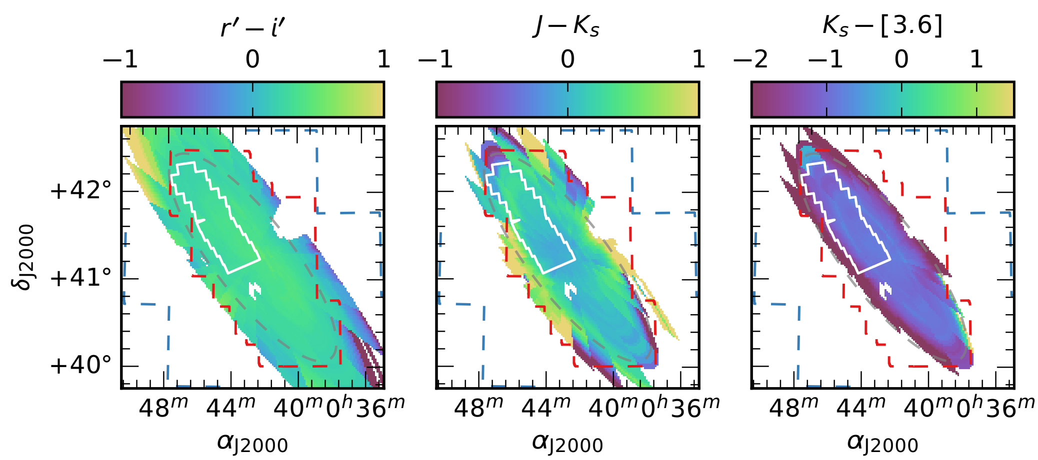
This SED library-in-Gibbs sampler is clearly very similar to the Metropolis-Hastings in Gibbs sampler, with the exception that the stellar population model for each pixel is estimated via a library marginalization rather than a series of Metropolis-Hastings steps for each stellar population parameter.

One subtlety involves choosing the *estimated* SED at each step (e.g., $\{f\}(\theta_j^{(k)})$ from Eq. 4.26). An option is to find the stellar population model in the library that most closely matches the estimated $\theta_j^{(k)}$ and adopt its SED as $\{f\}(\theta_j^{(k)})$. Another option is to compute f directly from the estimated stellar population parameters. However, the most correct definition of f in a Bayesian framework is the marginalization of all library model SEDs given the observed SED. That is, f is estimated similarly to stellar population parameters.

In fact, this method does not require the individual stellar population parameters to be estimated at all since only the marginalized SED is needed to estimate $\{f\}_j^{(k)}$. An interesting consequence, then, is that the details of stellar population models are entirely abstracted from this background estimator through the use of a pre-built library that encodes stellar population priors. This is turn means that the library does not even need to be built from synthetic SEDs; an unbiased sample of galaxy SEDs that are representative of the object being studied could also be used as an SED library.



(a) Before correction.



(b) With background correction.

Figure 4.9: ANDROIDS colour maps before and after background corrections. Maps are in units of mag$_{\mathrm{AB}}$ arcsec^{-2}. Panel (a) shows the original colour maps, while (b) shows M31 colours after background corrections are applied to the $i'JK_s$ images following § 4.5.2 (where $u^*g'r'$ are fixed with no background corrections in the models). In the background-corrected colour maps, the M31 disk's colour profile is more azimuthally symmetric. Qualitatively, this suggests that the hierarchical SED modelling is successful at identifying background biases. Note that these maps use the wedge pixelization scheme (§ 4.5.1), and the HST PHAT survey's footprint is shown for spatial reference.

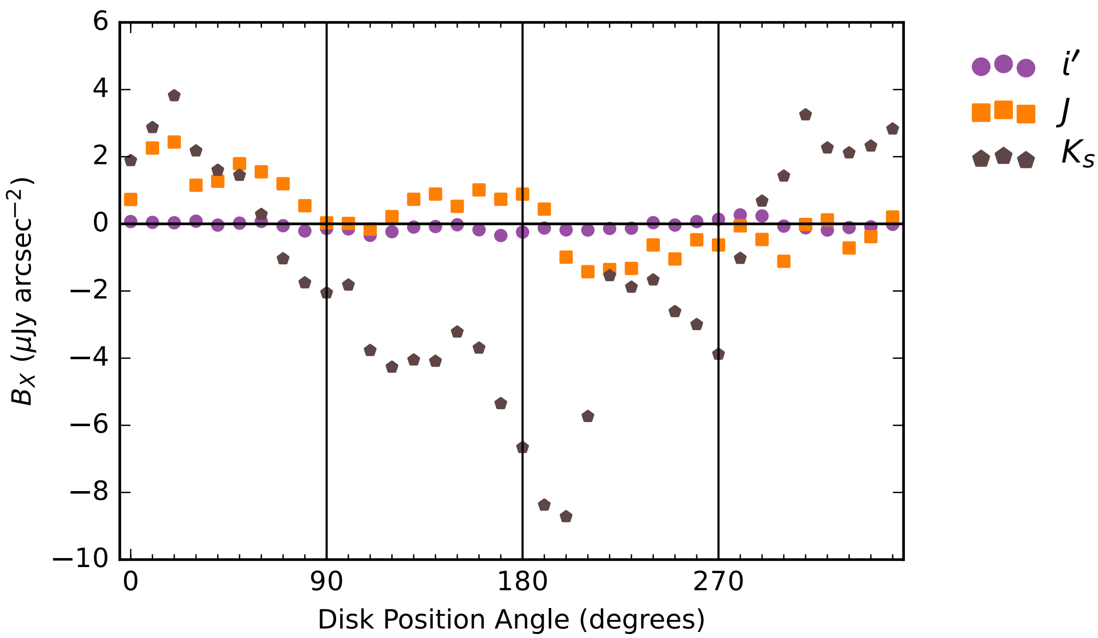
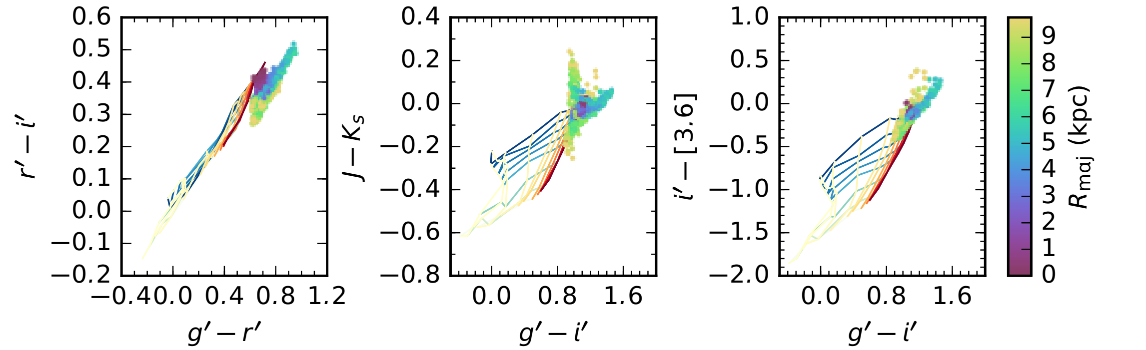
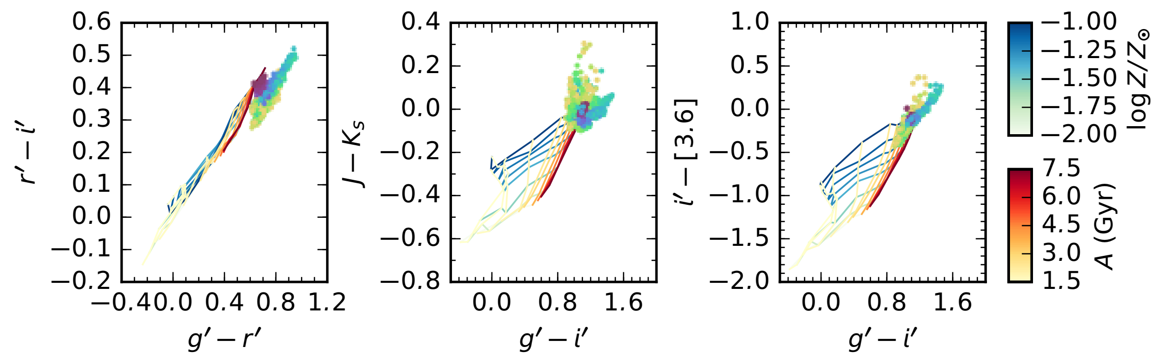


Figure 4.10: Background estimates for individual disk position angles modelled with fixed $B_{u^*,g',r'} = 0$ backgrounds. The Northern major axis is 0° and south-eastern minor axis is 90°. The K_s-band background corrections show a clear large-scale background bias approximately parallel to M31's major axis. In all cases the background estimates are highly correlated, suggesting that the sky background bias across ANDROIDS mosaics varies smoothly at large scales. Thus the assumption that the background is constant, to first order, within a wedge is valid.



(a) Before correction.



(b) With background correction.

Figure 4.11: ANDROIDS colour-colour diagrams (a) before and (b) after background corrections. Individual dots are pixels in the wedge segmentation (§ 4.5.1) coloured by M31 disk radius (R_{M31}). Colours are computed as AB magnitudes. For reference, simple stellar population (SSP) grids produced with FSPS (Conroy et al. 2009) show colours of dust-free stellar populations with ages between 100 Myr and 13 Gyr and $\log Z/Z_{\odot}$ metallicities between -2.0 and 0.2.

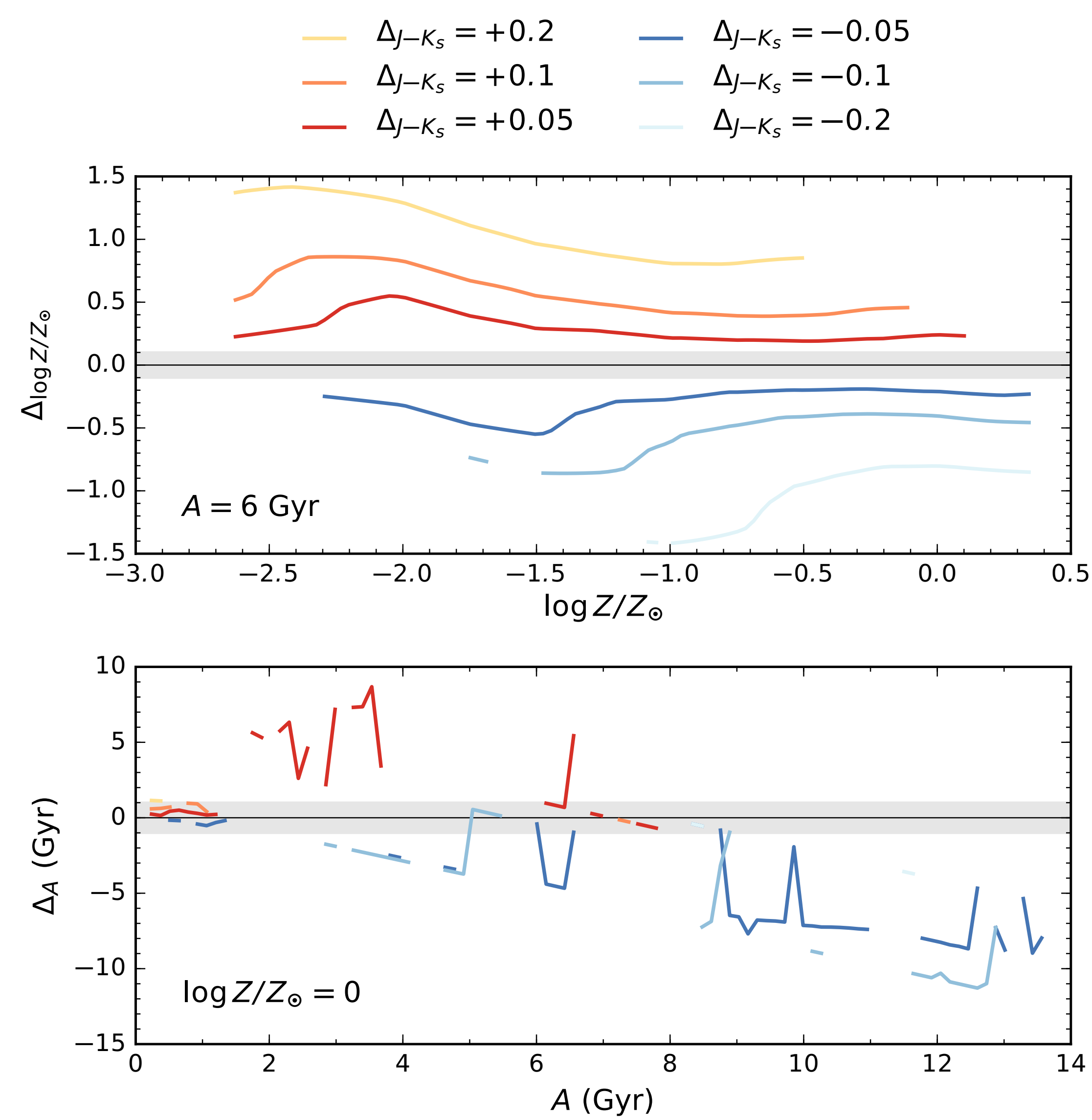


Figure 4.12: Each coloured line depicts how much a stellar population's metallicity (top) or age (bottom), when estimated entirely from $J - K_s$ colour, is biased when the $J - K_s$ colour is itself biased by -0.2, -0.1, -0.05, +0.05, +0.1, or +0.2 mag. In the top panel, SSPs with 6 Gyr ages are used. In the bottom panel, SSPs with solar metallicity are shown. All SSPs are dust-free, and modelled by FSPS (see § 4.6). This figure shows that a $J - K_s$ systematic bias of even ±0.05 mag can bias metallicity estimates by ±0.2 dex. Age estimation from $J - K_s$ colour is impractical. The $J - K_s$ biases modelled here can be compared to those seen in Figure 4.8.

Chapter 5

Panchromatic Spectral Energy Distribution Dataset

5.1 Introduction

The cornerstone of the ANDROIDS survey is novel low-surface brightness imaging in $u^* g' r' i' J K_s$ bands with CFHT WIRCam (Chapter 2) and MegaCam (Chapter 3). Critically, this dataset provides high resolution, low-surface brightness imaging in optical and near-infrared bandpasses suitable for spectral energy distribution (SED) modelling where previously lower resolution and more crudely calibrated datasets were used, like SDSS (used by Tamm et al. 2012; Viaene et al. 2014), and 2MASS (Beaton et al. 2007, though not used for SED modelling). However, the ANDROIDS optical-NIR dataset *alone* is insufficient to constrain stellar population and dust models of M31 pixels. The core ANDROIDS dataset misses light from extremely young stars emitted predominantly in UV. It also completely misses light that is absorbed by dust and re-emitted in the mid and far-infrared. SED modelling can only approach a true picture of M31 when a fully panchromatic UV, optical, near-IR, mid-IR and far-IR dataset is considered.

This chapter describes the comprehensive ANDROIDS dataset as it is used by subsequent SED modelling in Chapter 6. First, we introduce published datasets obtained from GALEX, Spitzer IRAC, and Herschel PACS/SPIRE in § 5.2. Note that all of these observatories are space-based, where wide-field imaging of M31 is not subject to atmospheric variations, and thus these datasets can be used predominantly as-is, without background modelling of the type described in Chapter 4. Second, in § 5.3 we describe how the panchromatic SED dataset is processed into a form suitable for SED modelling, including: PSF normalization, masking, and pixelization. Finally, in § 5.4 we describe and analyze the basic characteristics of the SED dataset.

5.2 Datasets

5.2.1 ANDROIDS CFHT MegaCam and WIRCam Dataset

The core of the ANDROIDS panchromatic dataset is unique CFHT MegaCam $u^*g'r'i'$ and WIRCam JK_s mosaics. The ANDROIDS MegaCam mosaics have deeper integrations and better background control than previous M31 disk optical mosaics made by SDSS (Tamm et al. 2012), while the WIRCam observations provide the first usable near-infrared J and K_s-band photometric measurements of M31.

CFHT/MegaCam We gathered the MegaCam datasets in the fall 2010 (g', r', and i') and 2012 (u^* and i') semesters. Fourteen MegaCam fields map the entire M31 disk out to 40 kpc on the major axis (Figure 3.1). To control background we used the Elixir-LSB observational strategy to map M31. To implement Elixir-LSB, the MegaCam observing sequence cycled through the 14 disk fields with eight sky fields (a sky-disk-disk-sky pattern). The moving window of sky integrations allowed us to create median background images to directly subtract the sky from M31 disk integrations. In each band we repeated the observing sequence six times for 16 minute total integrations per field in g' and r', 12 minute total integration per field in i' and 18 minutes per field in u^*.

In this Chapter we use the u^*, g', r', and i' mosaics that are processed as described in Chapter 3, specifically as images that are fully mosaiced with Swarp (Bertin et al. 2002) at their native $0\farcs18$ pixel scale. Flag maps (defining pixels that have flux, or not) and noise maps also accompany the full-resolution mosaics.

In Chapter 4 we modelled the ANDROIDS $u^*g'r'i'JKs$ SED to estimate background level corrections. We determined that the u^*, g', and r' images do not require any background correction, and in fact constrained those background corrections to zero in fitting exercises. However, we did model and estimate i' background corrections along 36 position angles around the M31 disk, though those corrections are nearly negligible. In this Chapter we will apply those background corrections to the i' image.

CFHT/WIRCam The near-infrared J and K_s datasets were introduced in Chapter 2 and Sick et al. (2014). Briefly, we gathered the WIRCam datasets during the 2007B and 2009B semesters. In the 2007B we mapped 27 disk fields, covering M31's bulge and disk out to 20 kpc. The total integration depth is 12.5 minutes per field in J and 10.8 minutes per field in K_s. In 2009B we observed 11 fields to a depth of 13.3 minutes in both J and K_s bands. These 2009B fields do not extend the WIRCam imaging footprint on M31, but rather improve background estimation statistics.

Like the MegaCam and Elixir-LSB observations, we used a sky-target nodding approach to subtract a real-time sky estimate from disk integrations. We found these sky estimates uncertain to about 2% of the sky background level (Table 2.2). To address this issue, we calculated sky offsets for each image that minimized image-to-image surface brightness differences in overlaps. These sky offsets enabled us to create full-resolution ($0\overset{\prime\prime}{.}3 \times 0\overset{\prime\prime}{.}3$) J and K_s images, which we use in this Chapter. In Chapter 4 we further estimated a set of background level corrections for the J and K_s mosaics along 36 disk positions angles based on hierarchical modelling of the ANDROIDS $u^*g'r'i'JK_s$ SED. In this chapter we combine those final background and the full-resolution WIRCam mosaics to produce the J and K_s components of the final SED.

5.2.2 GALEX UV Dataset

Ultraviolet imaging provides valuable information about on-going star formation in galaxies. OB stars emit most of their light in the UV, making UV luminosity an indicator of star formation rate in the last ~ 300 Myr timescale, which is the lifetime of B stars. A UV SFR relation can be written as (Calzetti 2013),

$$\frac{\mathrm{SFR(UV)}}{\mathcal{M}\ \mathrm{yr}^{-1}} = 3.0 \times 10^{-47} \frac{\lambda}{\mathring{A}} \frac{L(\lambda)}{\mathrm{erg\ s}^{-1}} \tag{5.1}$$

though Calzetti notes that this relation for the mean *star formation rate* is highly dependent on the recent *star formation history*. Besides this, UV light is highly attenuated by dust. In summary, UV information is valuable for stellar population and dust models.

GALEX mapped M31 in UV bands as part of its Nearby Galaxy Survey (NGS; Thilker et al. 2005).[1] Using 15 pointings, the GALEX NGS mosaic covers the bulge and disk out to $R_{\mathrm{M31}} = 27$ kpc (comparable to the ANDROIDS/WIRCam footprint).

[1] GALEX FITS mosaics were obtained through the NASA Extragalactic Database `http://ned.ipac.caltech.edu/cgi-bin/imgdata?objname=MESSIER+031`.

GALEX provides images in Far-UV (FUV, 1350–1750 Å) and Near-UV (NUV, 1750–2750 Å) bands. The GALEX PSF is $5''$, which is well-sampled by 1.5 arcsec^2 pix^{-1} pixels.

5.2.3 Spitzer IRAC Near- and Mid-Infrared Dataset

The Spitzer Space Telescope surveyed M31 with both IRAC (Barmby et al. 2006) and MIPS (Gordon et al. 2006). This study incorporates IRAC imaging calibrated and mosaiced by Barmby et al. (2006).

These IRAC observations cover the bulge and disk out to $R_{\mathrm{M31}} \sim 20$ kpc, similar to the ANDROIDS/WIRCam footprint. Pixels are 1.2 arcsec^2 pix^{-1}. Observations in four bandpasses sample the near- and mid-IR SED: 3.6 μm, 4.5 μm, 5.6 μm, and 8.0 μm. Like the WIRCam K_s band, IRAC 3.6 μm is nearly pure stellar emission, with neither dust attenuation nor emission. The redder bands progressively transition from stellar to dust, particularly polycyclic aromatic hydrocarbon (PAH) line, emission.

5.2.4 HELGA Herschel PACS and SPIRE Mid- and Far-Infrared Dataset

The final additional dataset is provided by Herschel space telescope imaging in the mid- to far-infrared. This dataset captures dust emission, including cold ISM dust.

The Herschel Exploitation of Local Galaxy Andromeda (HELGA: Fritz et al. 2012) assembled five mosaics of M31 with the PACS (100 μm, 160 μm) and SPIRE (250 μm, 350 μm, 500 μm) cameras. The HELGA footprint covers the entire M31 bulge and disk, again providing a good match to the ANDROIDS dataset. Of all the data sources we have aggregated, Herschel has the largest pixel scales and point spread functions. PACS images at 100 μm have a $8''$ FWHM, which increases to $36.4''$ in the reddest

SPIRE 500 μm images. Since SPIRE 500 μm has the broadest PSF, and also largest pixel scale (36 arcsec pix^{-1}), we use this image as both a pixel grid and PSF template for the combined SED (see § 5.3.3 and § 5.3.2).

We obtained mosaics as reduced by Viaene et al. (2014), hereafter HELGA IV. These images are fully reduced and calibrated, though we did find an error in the zeropoint stated in the FITS headers for SPIRE images. The SPIRE zeropoint is for flux units of μJy, rather than Jy as stated in the header.

5.3 Panchromatic SED Reduction

The panchromatic ANDROIDS SED consists of 17 images from GALEX (Thilker et al. 2005), CFHT/MegaCam, CFHT/WIRCam, Spitzer IRAC (Barmby et al. 2006), Herschel PACS and Herschel SPIRE (HELGA IV). This section describes the pipeline that combines these datasets. Broadly, this pipeline applies the following steps:

1. Masks foreground Milky Way stars (§ 5.3.1).
2. Matches PSFs (§ 5.3.2).
3. Resamples pixels to a common grid (§ 5.3.3).
4. Corrects for Milky Way foreground dust extinction (§ 5.3.5).
5. Applies background corrections (§ 5.3.5).

The end products of this pipeline are SED measurements in pixels that match the Herschel SPIRE 500 μm PSF and the HELGA IV pixel grid.

5.3.1 Milky Way Foreground Source Masking

We prepare and apply Milky Way foreground masks to full resolution images to prevent Milky Way light from leaking into downsampled and convolved mosaics in later processing steps. The means of identifying Milky Way foreground stars is tailored to each instrument's bandpass and PSF.

GALEX Milky Way Source Masking

To mask foreground Milky Way sources in GALEX images we follow the HELGA IV methodology to use *Source Extractor* (Bertin & Arnouts 1996) to detect sources and use colour criteria to select Milky Way sources. We created independent source catalogs for both the FUV and NUV GALEX images. Using *Source Extractor*, sources 5σ above the noise level were detected and analyzed (settings `DETECT_THRESH=5.0`, `ANALYSIS_THRESH=5.0` and `THRESH_TYPE=RELATIVE`). We combined the FUV and NUV source catalogs (using a maximum $5''$ spatial match criterion) and computed colours from *Source Extractor*'s standard isophotal magnitudes. Following the methodology of Gil de Paz et al. (2007), Milky Way foreground sources are identified as sources with:

$$|\mathrm{FUV} - \mathrm{NUV}| > 0.75. \tag{5.2}$$

Given these classifications, we replaced the pixel footprints of Milky Way sources with the *Source Extractor*-generated interpolated background image. 3529 sources were masked in the GALEX FUV and NUV images.

MegaCam Milky Way Source Masking

MegaCam foreground masks are built from a combination of two methods: manually-created stellar halo masks and colour-selected 2MASS stellar sources.

Reflected light, primarily in the wide-field corrector, cause stellar light to leak out of the typical Gaussian PSF. Around brighter stars, donut-like reflections surround stars. The geometry of these reflections, including translations with respect to the star, depend on the location of the star on the focal plane. Although there have been efforts to model MegaCam's internal reflections and point-spread function from optical principles, these methods have not yielded a means to reliably mask or subtract bright stellar halos. Instead, we took a simpler approach and created polygon regions surround all visible halos and central stars in the 14 ANDROIDS disk fields. The masks also include diffraction spikes and bleeds. Since only Milky Way stars are bright enough to create easily detectable halos, these mask effectively mask brighter Milky Way foreground sources.

For dimmer sources we select Milky Way stars with the 2MASS source catalog (Skrutskie et al. 2006). We classified all sources with colour

$$J - K_s < 0.9. \tag{5.3}$$

Luminous M31 stars on the Red Giant Branch and Asymptotic Giant Branch tend to have distinctly redder near-infrared colours, $J - K_s > 1.0$, owing to their higher metallicity than the foreground of predominantly Milky Way thick disk and halo stars. This classification is illustrated in Figure A.2 on page 339. We exclude 2MASS sources inside a de-projected M31 disk radius $R_{\mathrm{M31}} < 15$ kpc since the 2MASS Point Source Catalog can become unreliable in the crowded Milky Way disk, and

M31 light dominates the Milky Way foreground. For all selected 2MASS sources we create mask regions with radii of 12″ for sources brighter than $J \leq 14.0$ mag, 10″ for $14.0 < J < 15.0$, and 5″ for $J \geq 15$. These circular regions effectively mask the PSF of these stars in CFHT/MegaCam images.

With *WeightWatcher* (Marmo & Bertin 2008) we rendered mask images from both the manually-drawn polygon regions covering bright stars and the algorithmically-drawn regions over 2MASS detected stars. Unlike the GALEX flag processing, we did not replace flagged regions with a median background in the highest resolution images. Instead, we used the convolution algorithm to interpolate over flagged pixels.

WIRCam Milky Way Source Masking

Milky Way source masking of the WIRCam was accomplished similarly to the procedure described above for MegaCam. We only used the 2MASS Point Source Catalogs, not the halo masks since WIRCam is not afflicted by internal reflections like MegaCam is.

5.3.2 PSF Matching

Each of the 17 images in the panchromatic dataset has a distinct point spread function, meaning that light from sources are distributed differently in each image. This effect is unimportant when the SED pixelization is much larger than the PSF. However, when the PSF FWHM is similar to the region size, differences in PSF cause sources to contribute different proportions of their light to a given pixel. Extracting SEDs from SPIRE-sized (36 arcsec pix^{-1}) pixels is such a regime. The solution is to match the PSFs of all datasets by convolving each image to the broadest PSF, which is SPIRE 500 μm for the present dataset.

We use Aniano et al. (2011) kernels[2] to convolve an input image to match the SPIRE 500 μ image. Each convolution kernel is specifically designed for the input and output PSF pair (here, SPIRE 500 μm). The convolution itself was implemented with *Astropy's* `astropy.convolution.convolve_fft` function. The following are notes specific to each dataset:

MegaCam and WIRCam While specific PSFs are known for GALEX, IRAC and Herschel images, the MegaCam and WIRCam images are convolved with kernels designed for a Gaussian PSF with 3.0″ FHWM. In this pipeline, we create intermediate-stage MegaCam and WIRCam mosaics with a 3 arcsec pix^{-1} scale (as used by § 4.5). Since the convolution kernel itself is necessarily downsampled to the same pixel scale,[3] there is no effective difference between the mosaic PSFs and the kernel input PSF.

GALEX, Spitzer IRAC, and Herschel PACS and SPIRE Note that GALEX and Spitzer IRAC images are processed from a 3.0 arcsec pix^{-1} scale to reduce computational memory requirements. Herschel PACS and SPIRE images are processed from their original pixel scale, which is larger than 3.0 arcsec pix^{-1}.

In summary, this pipeline stage yields mosaics with a common PSF corresponding to the SPIRE 500 μm PSF.

---

[2] Available online at `http://www.astro.princeton.edu/~ganiano/Kernels/Ker_2012_May/Kernels_fits_Files/Low_Resolution/`

[3] Before applying the convolution, we resample the kernel to the input dataset's pixel scale using Scipy's `scipy.interpolate`

5.3.3 Pixel resampling

Next, we resample the PSF-matched images to the HELGA SPIRE 500 μm image's native pixel grid and 36 arcsec pix^{-1} scale. This resampling accomplishes two things. First, like PSF matching, this resampling matches all images to the lowest resolution dataset (Herschel SPIRE). Second, using the HELGA pixel grid specifically makes direct comparison of ANDROIDS and HELGA SED modelling straightforward. Chapter 6 takes advantage of this feature.

This pixel resampling step is built upon *Swarp* (Bertin 2007), and operates in a similar fashion as pixel resampling already employed in § 2.7.1, § 3.3 and § 4.5.1. Specifically, *Swarp* is operated in a resample-only mode, without coaddition, where all 17 images in the dataset are processed simultaneously. The HELGA SPIRE 500 μm used by *Swarp* to define the output world coordinate system (WCS). For this, we use a slightly obscure *Swarp* feature where an existing image can be used to define a target WCS by symbolically linking its file on the filesystem to the name of the output path, but with a `.head` extension. Although *Swarp* properly resamples images to the target WCS, we find it often adds padding to the resampled images so that their dimensions are mismatched. This is straightforward to solve by post-processing the resampled images to add or subtract padding from each image edge to ensure that the `NAXIS1`/`NAXIS2` (image dimensions) and `CRPIX1`/`CRPIX2` (pixel containing the reference RA and Declination coordinate) FITS headers are consistent to the target image. We have made a Python package, called *Skyoffset*,[4] that wraps `Swarp` to automate a resampling pipeline, including correcting any padding mis-matches.

[4]`https://github.com/jonathansick/skyoffset`

5.3.4 Flux Measurement

From these resampled and convolved images we extract a 17-band SED at each pixel. Absolute fluxes are measured in each pixel given an intensity I DN as:

$$\frac{F_X}{\mu\mathrm{Jy}} = \left[f_{\mathrm{MW}} \frac{I_X}{\mathrm{DN}} 10^{-0.4(m_0+48.6)+23+6}\right] - B_X. \tag{5.4}$$

Three calibration terms are included in this expression: a zeropoint (m_0), a correction for Milky Way foreground extinction ($f_{\mathrm{MW},X}$), and a final background correction (B_X). The latter two terms are described in the following sections.

5.3.5 Milky Way Extinction Correction

Pixel fluxes are corrected for foreground Milky Way dust extinction using the Schlafly & Finkbeiner (2011) dust reddening map. Specifically, we assume a uniform colour excess in the direction of M31 of $E(B-V) = 0.07$ (Schlegel et al. 1998). Given an $R_V = 3.1$ extinction law,[5] Table 5.1 lists the absolute Milky Way foreground extinction.

Note that Schlafly & Finkbeiner (2011) provides extinction coefficients in SDSS bandpasses. Thus we initially compute total attenuations for SDSS *ugri* bandpasses, then transform those magnitudes to the CFHT/MegaCam $u^* g' r' i'$ filter system using analytical fits provided by CADC MegaPipe:[6]

[5] $R_V \equiv A_V / E(B-V)$ describes the shape of the optical extinction law.
[6] `http://www3.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/megapipe/docs/filt.html`

Table 5.1: Milky Way foreground dust attenuation in ANDROIDS bandpasses, given Schlafly & Finkbeiner (2011) with $R_V = 3.1$ and $E(B-V) = 0.07$ (Schlegel et al. 1998). GALEX attenuation is given by Gil de Paz et al. (2007).

Filter	A_X (mag)
GALEX FUV	0.55
GALEX NUV	0.56
u^*	0.28
g'	0.22
r'	0.16
i'	0.12
J	0.05
K_s	0.02

$$u^* = u - 0.241(u - g), \tag{5.5}$$

$$g' = g - 0.153(g - r), \tag{5.6}$$

$$r' = r - 0.024(g - r), \tag{5.7}$$

$$i' = i - 0.003(r - i). \tag{5.8}$$

Then the flux correction coefficient is

$$f_{\mathrm{MW},X} = 10^{-0.4 A_X}. \tag{5.9}$$

Background Calibration

As we extract SEDs from individual pixels we apply final background corrections computed either through hierarchical background SED modelling (Chapter 4) or direct background pixel measurement.

By measuring median fluxes in several regions of pixels in "blank" peripheral regions around M31, we found that the IRAC images (Barmby et al. 2006) and HELGA IV Herschel PACS and SPIRE images do not require background corrections. The intensity of the background structure in GALEX mosaics is less than the background noise in the sampled blank regions, and any background structure that can seen visually is more spatially complex than a simple scalar level or plane that could be fitted. For this reason, we do not attempt to subtract any background from the GALEX mosaics. This is contrary to HELGA IV who fit and subtract background planes from both GALEX and IRAC mosaics. Note that this study directly uses Herschel PACS and SPIRE mosaics from HELGA IV, so we inherit those background corrections.

We do apply background corrections to the MegaCam i' and WIRCam JK_s images based on hierarchical background modelling presented in Chapter 4. Recall that in Chapter 4 we modelled a SEDs of pixels in 36 wedges covering the M31 disk. To correct the background from an individual pixel we compute that pixel's position angle on the M31 disk and linearly interpolate the background estimated between the two bounding wedges. Background corrections are dependent on disk radius, though in practice background corrections are only meaningful for pixels in the outer disk; in the inner disk the background correction is much less than the M31 flux.

Summary

In this section we have described a uniform processing of the panchromatic ANDROIDS dataset into a set of 17 image slices that make a sample M31's SED from 0.15 μm (GALEX FUV) to 500 μm (Herschel SPIRE). The images are both spatially matched and PSF-matched to the SPIRE 500 μm image. We have also presented a best effort to mask foreground Milky Way sources and prevent their light from leaking into the mosaics. Finally, in these mosaics we have applied background corrections modelled in Chapter 4 and corrected for foreground Milky Way dust extinction.

In § 5.4 we will briefly plot and characterize these images. Then in § 5.5 we reduce the mosaics further into radial isophotal light profiles.

5.4 The Panchromatic Image Set

In Figure 5.1 we show the set of 17 images we have compiled into the panchromatic ANDROIDS SED and processed as discussed in § 5.3. The image grid shows the varying faces of M31 in each bandpass. Seen in the GALEX FUV, NUV, and MegaCam u^* bands, M31 is characterized by young star formation in 5 and 10 kpc rings. The redder optical and near-infrared bands trace the galaxy's smooth stellar disk. The 10 kpc ring is still visible in the near-infrared, possibly due to an enhancement of asymptotic giant branch (AGB) stars that are a main feature in 1 Gyr old stellar populations (see § 7.6 for a detailed discussion of stellar mass build-up and AGB populations seen in resolved stellar catalogs). Flux from the stellar disk subsides in the mid-infrared IRAC [4.5] to [8.0] and the Herschel PACS and SPIRE images trace the dust emission.

The colours of pixels give some intuition into the stellar populations and dust. Recall that colour indices are logarithmic ratios of fluxes in two images, F_1 and F_2:

$$m_1 - m_2 = -2.5 \log_{10} F_1/F_2 \tag{5.10}$$

Figure 5.2 shows colours in several indices of GALEX, MegaCam, WIRCam, IRAC and Herschel bands.

Star-Forming Arm Structures The GALEX FUV – NUV (Figure 5.2(a)) and MegaCam $u^* - g'$ (Figure 5.2(b)) maps trace recent star formation along ring structures as blue regions. Most star formation is occurring in the 10 kpc ring, but also in the 5 kpc ring, around the hole in the southern 10 kpc ring (related to the passage by M32, see Gordon et al. 2006; Block et al. 2006), and additional arms along the northern major axis that extend to 15 kpc.

Stellar Population Gradients The optical and near-infrared colours trace the characteristics of the bulk of M31's stellar mass. In general, the inner regions of M31 are redder than the outer disk. This is most clearly seen in the $u^* - g'$ (Figure 5.2(b)) and $g' - i'$ (Figure 5.2(c)) maps. Redder colours in the centre can be due to any combination of higher age, higher metallicity, and higher dust content. Inside-out galaxy formation models (Matteucci & Francois 1989) predict that a spiral galaxy's bulge and inner disk should be both older and more enriched than the more recently-formed outer disk.

The ANDROIDS dataset also allows us, for the first time, to examine the near-infrared $J - K_s$ colour gradient across M31. Figure 5.2(e) shows that the $J - K_s$ colour is nearly constant throughout the disk, with a small inside-out *reddening* (as opposed to the outer disk being more blue). Extinction in the 5 kpc and 10 kpc arms on the near-side minor axis is also perceptible, underscoring that the J-band is not extinction-free.

Differential Near-Far Inclined Disk Extinction The $g' - i'$ (Figure 5.2(c)) and $i' - [3.6]$ (Figure 5.2(d)) maps demonstrate differential dust extinction across the M31 disk due to its inclination. In these maps, the north-eastern minor axis is redder compared to the south-western minor axis. Elmegreen & Block (1999) reproduced this effect by modelling light absorption scatter in an inclined stellar disk and bulge with an embedded dust disk. That modelling confirmed that the *near* side of an inclined stellar disk is reddened compared to the far side because a line of sight though the near disk has more light *behind* the dust disk compared to a line of sight through the far disk that has more stellar light *in front* of the dust disk.

5.5 Radial Profiles with Elliptical Isophotes

In this section we reduce the 2D ANDROIDS SED maps into a set of 1D flux and surface brightness profiles that are projected along M31's major axis. Taking advantage of the first-order azimuthal symmetry of M31's bulge and disk, surface brightness profiles are convenient for quantifying the radial structure of galaxies.

5.5.1 Elliptial Isophote Fitting

We used the XVISTA software package to fit elliptical isophotes in ANDROIDS mosaics. Rather than fitting isophotes individually to each image, we fit isophotes only to the r' image. With this common ellipse geometry we are able to measure surface brightness profiles in all bands.

We specifically use the ANDROIDS r' image since it is both representative of M31's stellar mass distribution (tracing lower mass that make up the bulk of M31 total stellar mass) and has the best background subtraction of the MegaCam set. Due to memory constraints, we measured the r' mosaic resampled to $5''$ pixel^{-1}, though the image is not otherwise convolved from its native PSF.

In Figure 5.3 we plot this r' profile, as major axis position angle, ellipticity, and

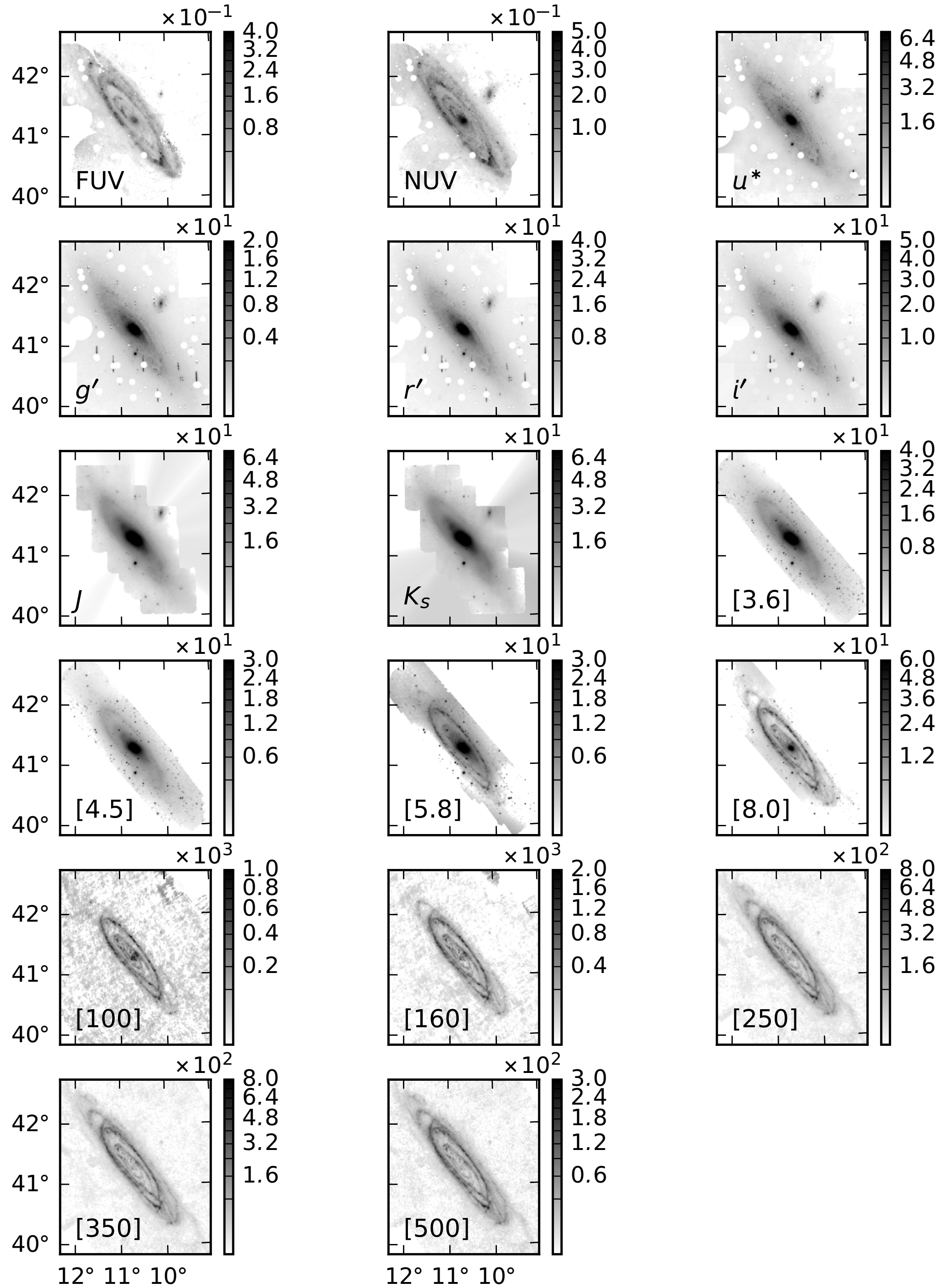
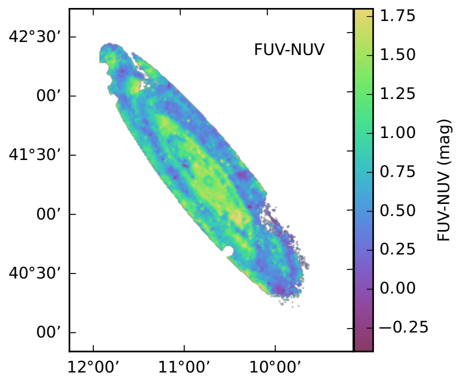
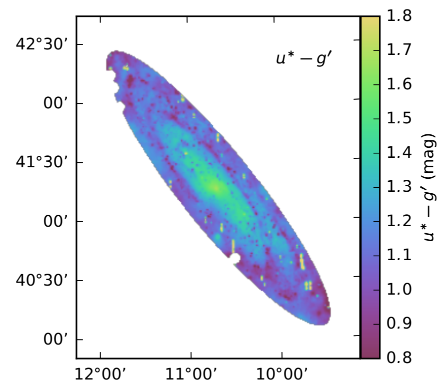


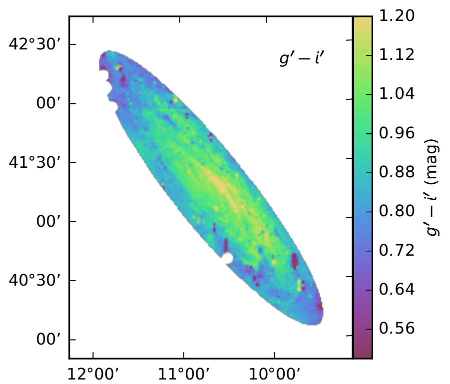
Figure 5.1: Images in the 17 ANDROIDS SED bands, as used for per-pixel SED analysis in Chapter 6. Color bars are in units of μJy arcsec^{-2}, with a square-root stretch to show M31's stellar disk. Spatial axes are equatorial right ascension and declination, in units of degrees. The processing for these images is described in § 5.3, and includes masking, pixel resampling, and convolution to the SPIRE 500 PSF and pixel scale.



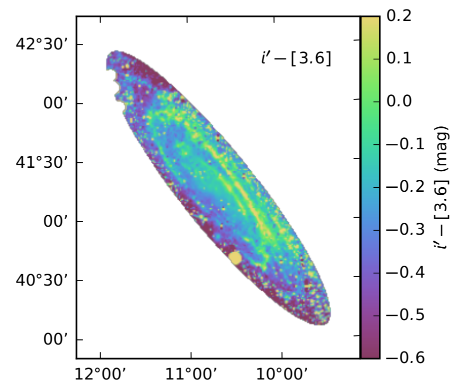
(a) FUV − NUV.



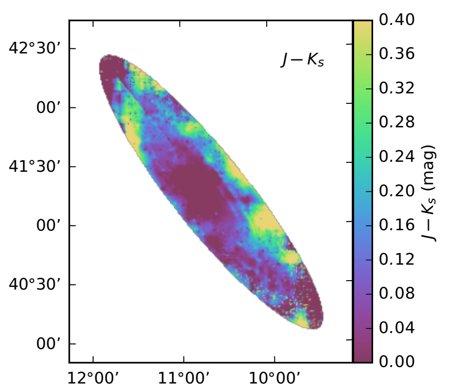
(b) $u^* - g'$.



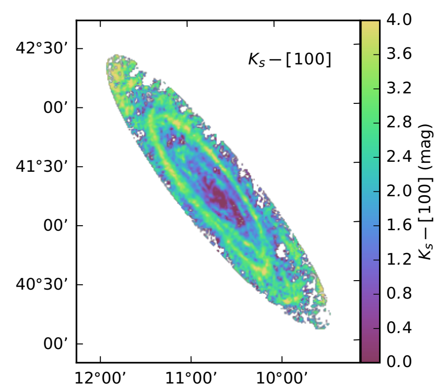
(c) $g' - i'$.



(d) $i' - [3.6]$.



(e) $J - K_s$.



(f) $K_s - [100]$.

Figure 5.2: Colour index maps of M31, truncated to $R_{\mathrm{M31}} = 20$ kpc. The colour scale is an AB magnitude difference of the two images, $m_1 - m_2 = -2.5 \log F_1/F_2$. Spatial axes are equatorial right ascension and declination, in units of degrees. Each image is processed according to § 5.3, and thus is convolved to the SPIRE 500 PSF and pixel scale.

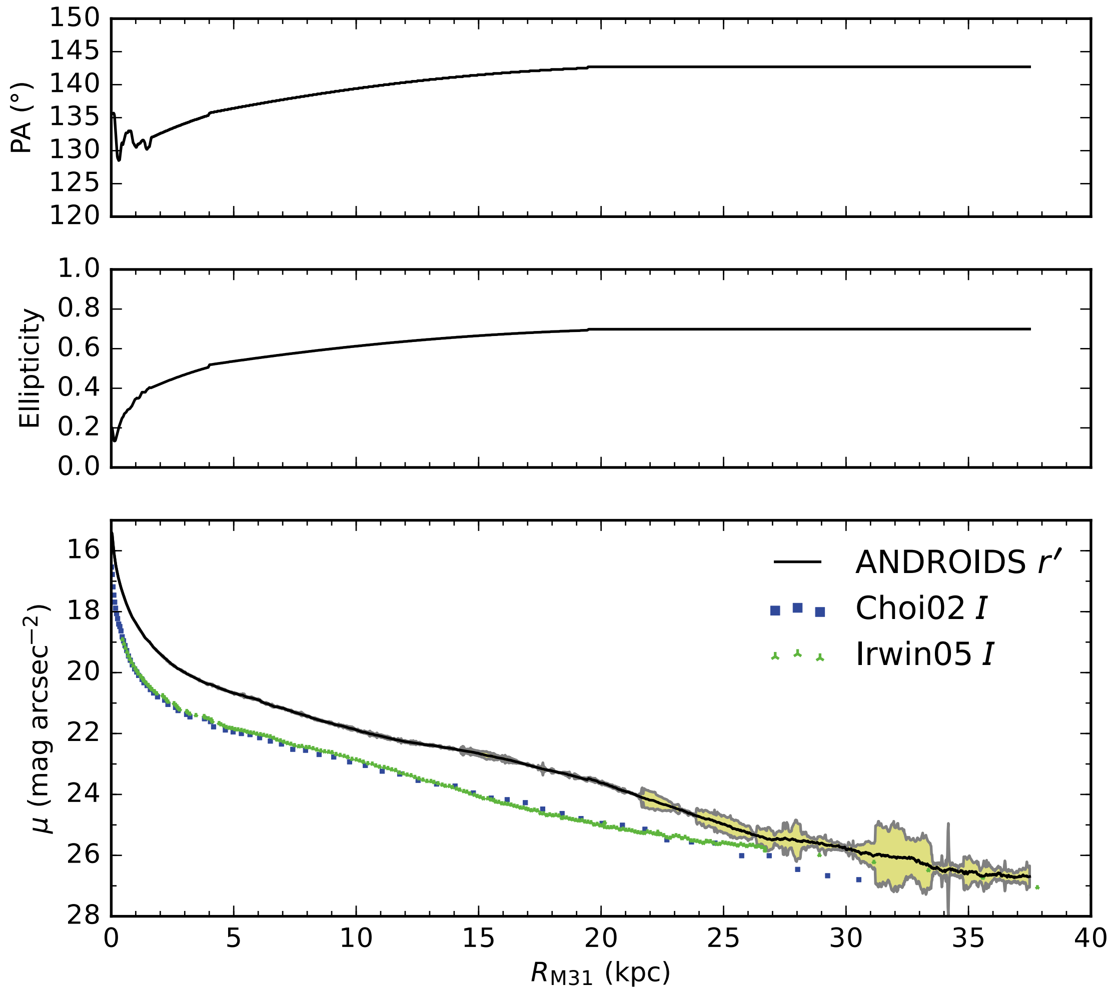


Figure 5.3: M31 r'-band surface brightness profile measured with elliptical isophote fitting of the ANDROIDS r'-band mosaic. Top: the position angle of the isophote major axis. Middle: ellipticity of the isophote. Bottom: surface brightness profile with observed variance of pixels within the isophote's annulus. For comparison, we also show the I major axis surface brightness profiles measured by Choi et al. (2002) and Irwin et al. (2005).

surface brightness as functions of M31's major axis radius. Ellipticity is defined as

$$\epsilon(r) = 1 - \frac{R_{\mathrm{minor}}(r)}{R_{\mathrm{M31}}}. \tag{5.11}$$

We find that ellipticity increases from 0.2 at M31's centre to 0.7 by $R_{\mathrm{M31}} = 20$ kpc. That the ellipticity is not zero at the centre reflects M31's boxy bulge (Beaton et al. 2007), and the high ellipticity in the disk reflects its 77.5° inclination (Walterbos & Kennicutt 1988).

As the ellipticity grows from bulge to disk, so does the position angle of isophote ellipses. This is because although M31's boxy bulge is mostly aligned with the M31 disk, it is not entirely aligned. The mean position angle of the bulge is 132°, while the orientation of the disk is 142.7°.

The r' surface brightness profile is essentially exponential across the entire M31 disk, out to the limiting radius of the observations. The shape of the disk's surface brightness profile is similar to that seen in the I-band by Choi et al. (2002). We do not see any flattening in the surface brightness profile at $R_{\mathrm{M31}} = 25$, as seen by Irwin et al. (2005).

5.5.2 Panchromatic Surface Brightness Profiles

We used the r' profile geometry described in the previous section to measure surface brightness profiles in other bands. Given the profile geometry, we determined the deprojected disk radius of each pixel. Then we created radial bins — annuli — and measured the median flux within each annulus on each band of the convolved and resampled panchromatic image set described in § 5.3. Section 5.5.5 describes the panchromatic SB profiles.

We also used the ellipses at each radius as apertures for measuring the integrated flux of M31. This is described in § 5.5.3, next.

5.5.3 Integrated Photometry and Spectral Energy Distributions

Elliptical aperture photometry makes M31 comparable to more distant galaxies where spatial information is unavailable. In Tables 5.2 and 5.3 we list total fluxes within several isophotal ellipses of increasing radii, out to $R_{\mathrm{M31}} = 20$ kpc, measured from the ANDROIDS dataset described in § 5.4. The geometries of the elliptical apertures at each radius are defined by the fitted r' isophotal ellipses (Figure 5.3). Figure 5.4 is a graphic counterpart to the tables, showing the shape of M31's SED within each projected isophotal disk radius. Note in particular how both the UV and mid to far-infrared domains of the SED gain most of their flux only within the 10 kpc ring. Figure 5.5 elaborates this point by showing continuous radial growth curves of flux in each ANDROIDS SED bandpass.

5.5.4 Comparison of Integrated M31 Flux to HELGA IV

Integrated photometry is a convenient means of validating the ANDROIDS SED and reduction methodology (§ 5.3). HELGA IV published an SED of M31's "global" light. Although the HELGA IV "global" aperture, with semi-major axis $R_{\mathrm{M31}} = 22$ kpc, is slightly different than the largest aperture measured with the ANDROIDS dataset ($R_{\mathrm{M31}} = 20$ kpc), the flat flux growth curves (Figure 5.5) suggest that the two measurements are comparable.

Figure 5.6 shows the HELGA IV "global" SED alongside the ANDROIDS 20 kpc integrated SED. Both surveys share the same GALEX, IRAC, and Herschel PACS and SPIRE observations, though they are independently measured and reduced according to § 5.3. Generally the flux measurements are in excellent correspondence, particularly the IRAC measurements. The one exception is the SPIRE 250 μm flux measurement,

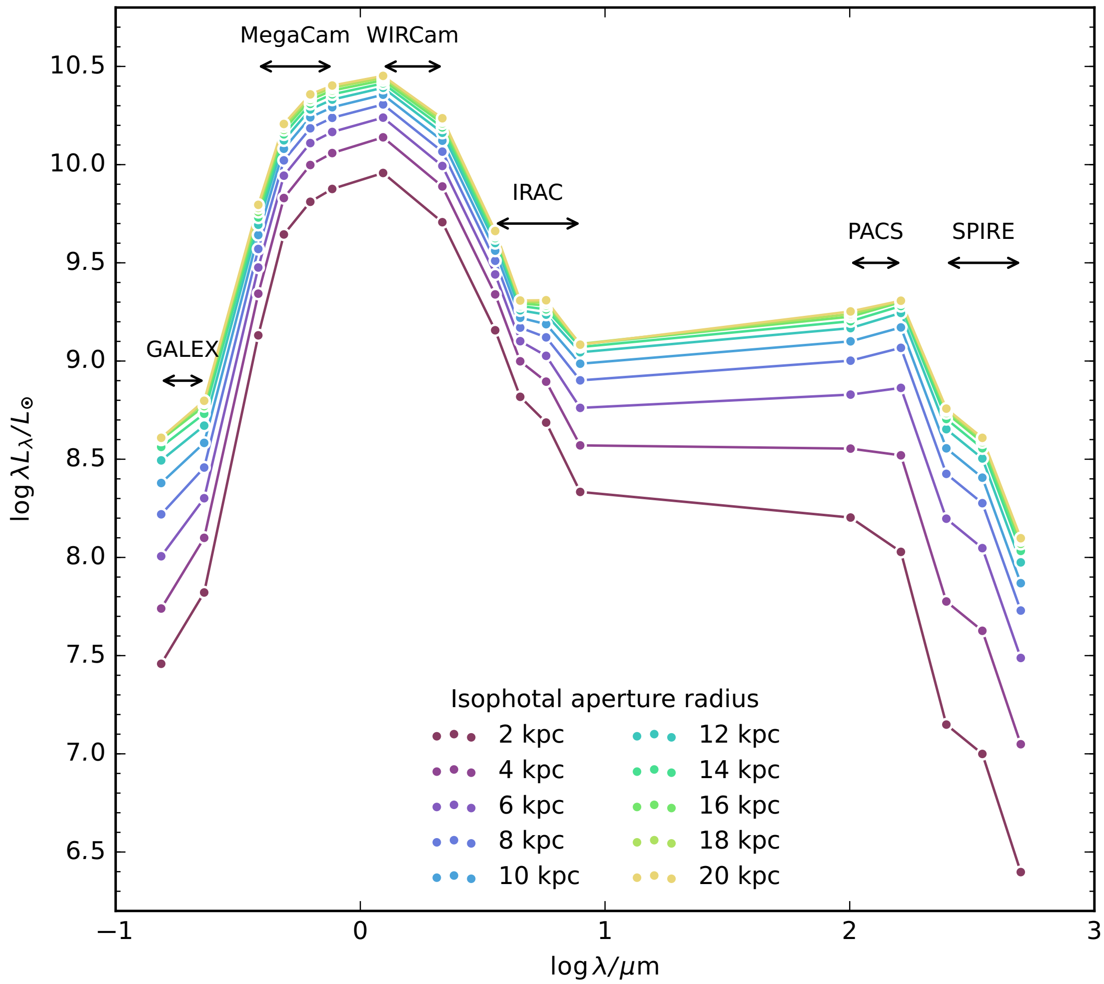


Figure 5.4: Integrated M31 SED sampled in a series of isophotal apertures. Each SED is coloured by isophotal radius, from 2 kpc to 10 kpc along M31's major axis. Fluxes are integrated within isophotal ellipses defined in § 5.5 of the given major axis radius. Labels highlight the wavelength domains of each instrument incorporated by ANDROIDS.

which is smaller in ANDROIDS measurements. We do not know why this disagreement exists.

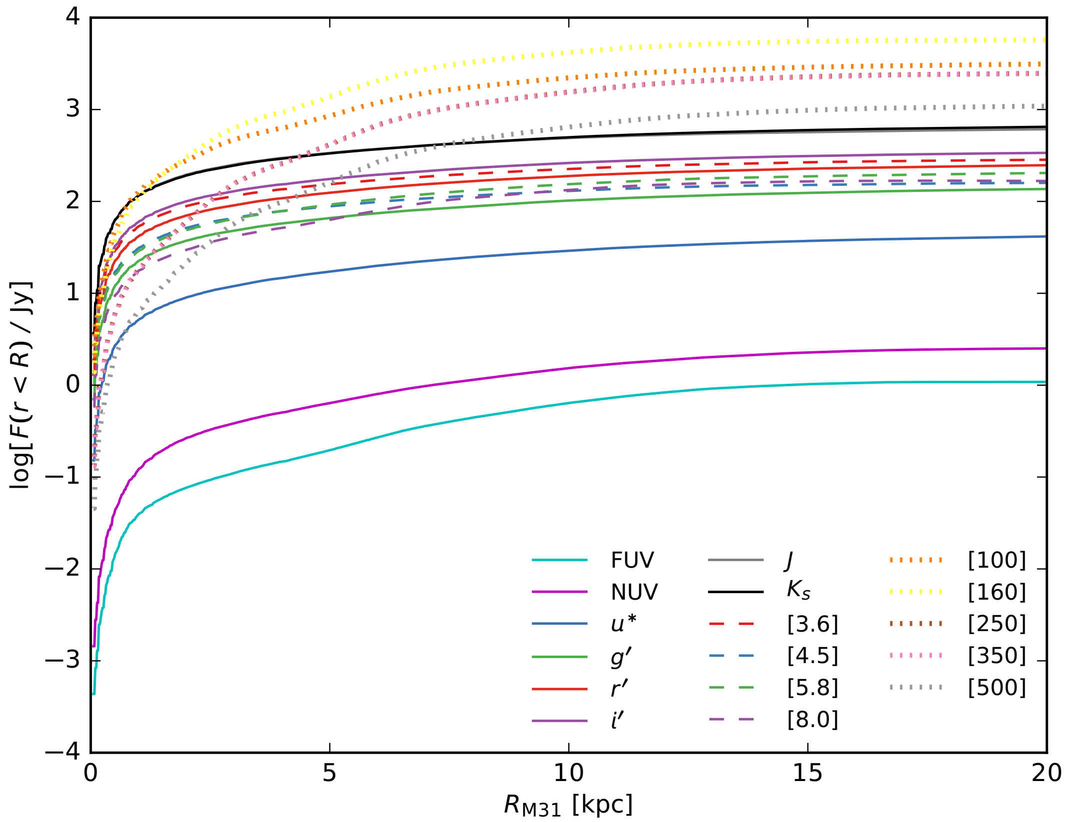


Figure 5.5: Cumulative flux profiles measured in isophotal ellipse apertures (§ 5.5).

5.5.5 Radial Surface Brightness Profiles

In this section we present radial surface brightness profiles for all 17 ANDROIDS bands in Figs. 5.7–5.10. These profiles are measured with the r' isophotal ellipse geometry (§ 5.5.1) against the ANDROIDS mosaics convolved to the SPIRE 500 μm PSF (§ 5.3). These figures surpass WK87, Choi et al. (2002), Irwin et al. (2005), C11, Tempel et al. (2011), and Rafiei Ravandi et al. (2016) as the most comprehensive set of photometric profiles for M31's bulge and disk.

We show the ANDROIDS MegaCam and WIRCam $u^* g' r' i' J K_s$ profiles in Figure 5.7.

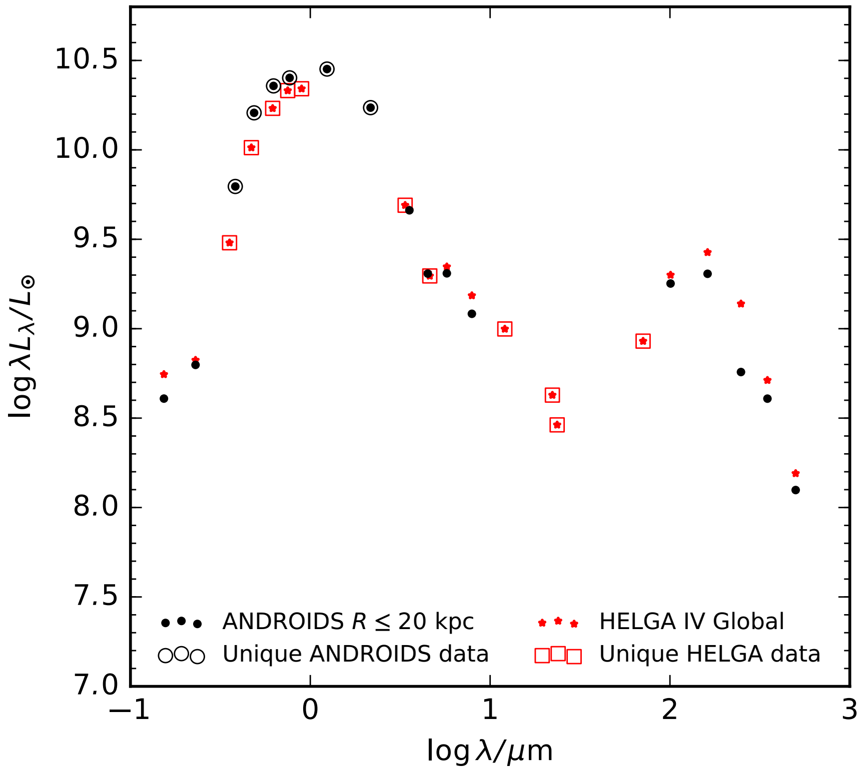


Figure 5.6: Comparison of ANDROIDS and HELGA IV global M31 SEDs. The ANDROIDS SED is shown as black dots while the HELGA IV global SED is shown at red stars. Bandpasses unique to ANDROIDS are highlighted with outer circle markers, while those unique to HELGA IV highlighted with red squares. The datasets unique to HELGA IV are SDSS, WISE, and Spitzer MIPS. The GALEX, IRAC, and Herschel fluxes allow a direct comparison between ANDROIDS and HELGA IV measurement methods with identical source datasets. Note that the HELGA IV "global" SED is measured in an elliptical region with a 22 kpc major axis radius, whereas the ANDROIDS SED is measured in an elliptical isophote with a 20 kpc major axis radius. The difference in regions yields only a minimal visible differences in the SEDs given the diminishing curve of growth in the flux at large radii (see Figs. 5.4 and 5.5).

In all bands we see a level plateau in the surface brightness profiles at $R_{\mathrm{M31}} = 10$ kpc, corresponding to the star forming ring. That this ring is measurable in all bands, even the near-infrared iJK_s, hints that the ring structure must be long lived enough to have a concentration of evolved lower mass stars on the red giant and asymptotic giant branches seen in the near-infrared, not just very young O- and B-type stars that dominate bluer bands. See Maraston (1998), particularly their Fig. 7, for a discussion of what stellar phases contribute to a stellar population's light at each age.

Beyond $R_{\rm M31} = 10$ kpc, photometry in all bands follow featureless exponential profiles. The WIRCam profiles are truncated to $R_{\rm M31} = 20$ kpc due to spatial coverage. The exponential profiles measured in ANDROIDS data are slightly shallower than that measured by Choi et al. (2002) and Irwin et al. (2005). Beyond $R_{\rm M31} = 25$ kpc the ANDROIDS MegaCam profile shows signs of becoming shallower. This change in slope could be either due to a systematic under-subtraction of background (including Milky Way light and background galaxies) or the transition from the M31 disk to halo.

The Irwin et al. (2005) I-band profile also shows a flattening at $R_{\rm M31} > 25$ kpc. This is significant because this portion of the Irwin et al. (2005) profile is measured from resolved star counts, calibrated to match their inner integrated light surface brightness profile. Resolved star counts should either be less prone to — or respond differently to — foreground Milky Way and background galaxy light contamination than the integrated light ANDROIDS profiles.

Rafiei Ravandi et al. (2016) is another study that produced an extended surface brightness profile of M31 using integrated photometry across the inner disk and star counts in the outer disk. Their profile of the IRAC 3.6 μm is shown in Figure 5.8 alongside this study's measurements of the WIRCam and IRAC images. The Rafiei Ravandi et al. (2016) profile diverges sharply from ANDROIDS measurements at $R >$ 15 kpc, even before their transition to resolved star counts at $R_{\rm M31} = 20$ kpc. This suggests that the background was over-subtracted in the Rafiei Ravandi et al. (2016)

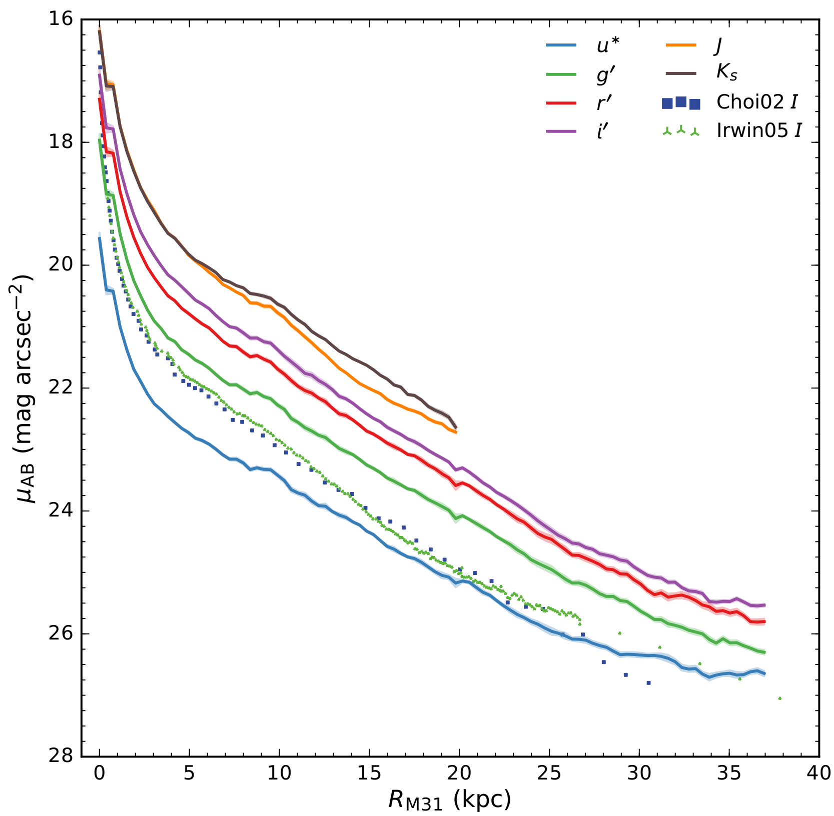


Figure 5.7: Surface brightness profiles of ANDROIDS MegaCam and WIRCam images processed as described in § 5.3 and measured with isophotal ellipses fit to the r' image (§ 5.5.1). For reference, we also plot the I profiles of Choi et al. (2002) and Irwin et al. (2005).

study, making that study's profile unusable for validation.

5.5.6 Radial Colour Profiles

In Figure 5.11 we show radial colour profiles for a selection of pairs of ANDROIDS bands. Colour profiles are the differences in surface brightness profiles.

In these profiles the trends seen in the 2D colour maps (Figure 5.2) become more

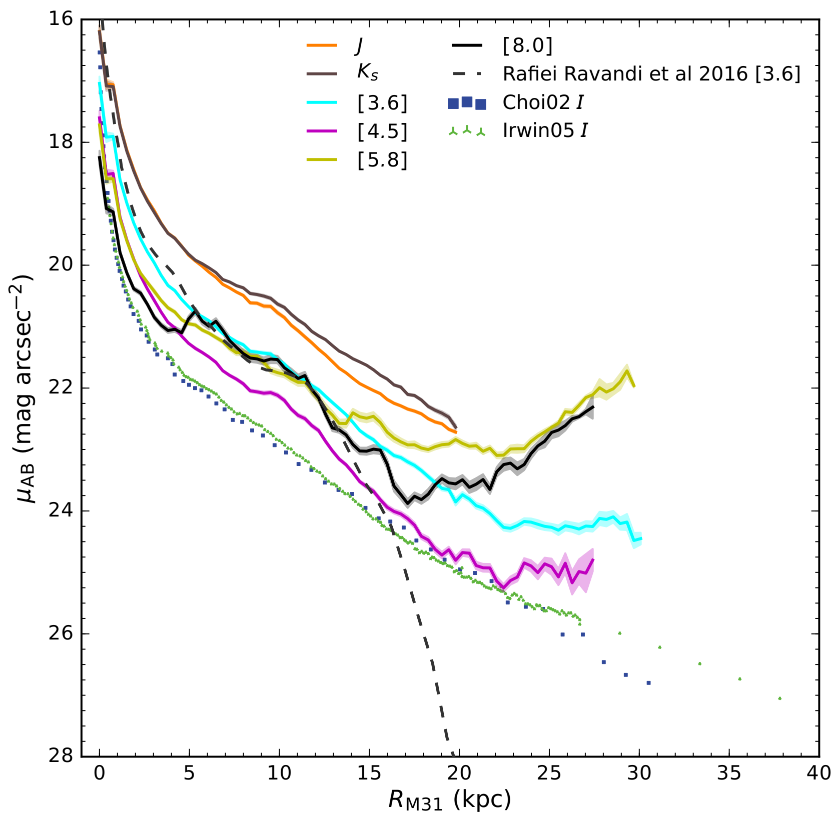


Figure 5.8: Surface brightness profiles of near-infrared WIRCam and IRAC (Barmby et al. 2006) images processed as described in § 5.3 and measured with isophotal ellipses fit to the r' image (§ 5.5.1). For reference, we also plot the I profiles of Choi et al. (2002) and Irwin et al. (2005), and the extended 3.6 μm profile measured by Rafiei Ravandi et al. (2016).

clear. In optical colours, M31's centre is redder than the outer disk. The $g' - i'$ colour exemplifies this well: we see a nearly linear radial decrease from $g' - i' = 1.1$ at the centre to $g' - i' = 0.6$ at $R_{\rm M31} = 30$ kpc. What is remarkable about the radial $g' - i'$ colour trend is that it is unaffected by M31's stellar structures, such as the central bulge or the 10 kpc ring.

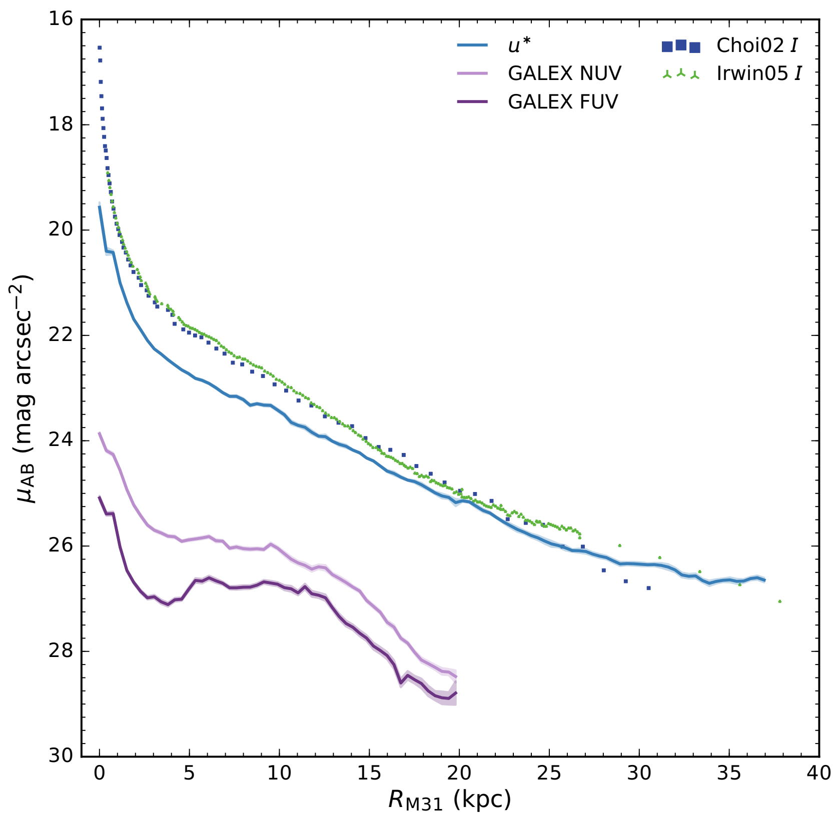


Figure 5.9: Surface brightness profiles of GALEX (Thilker et al. 2005) images processed as described in § 5.3 and measured with isophotal ellipses fit to the r' image (§ 5.5.1). For reference, we also plot the I profiles of Choi et al. (2002) and Irwin et al. (2005), and the ANDROIDS u^* profile.

5.6 Summary

In this Chapter we presented the ANDROIDS panchromatic image dataset. The dataset includes 17 individual bandpasses, spanning wavelengths of 0.15 μm to 500 μm. Besides the core CFHT MegaCam and WIRCam mosaics compiled originally for this study, we have compiled existing images from GALEX, Spitzer IRAC and Herschel. We have processed these images into a uniform set, suitable for extracting per-pixel

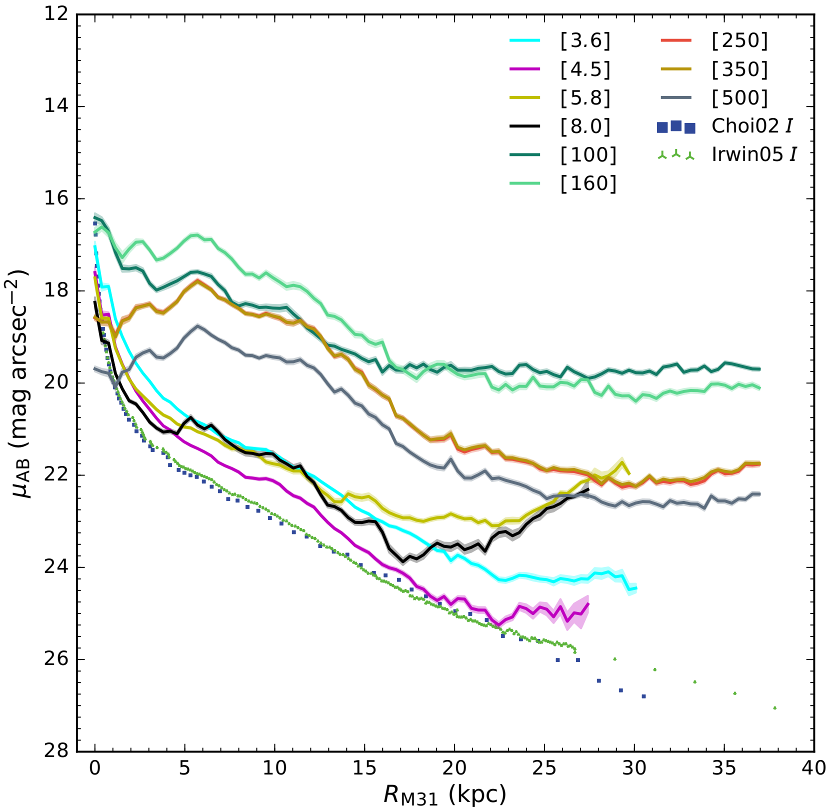


Figure 5.10: Surface brightness profiles of infrared IRAC (Barmby et al. 2006) and Herschel (Viaene et al. 2014) images processed as described in § 5.3 and measured with isophotal ellipses fit to the r' image (§ 5.5.1). For reference, we also plot the I profiles of Choi et al. (2002) and Irwin et al. (2005).

SEDs. Specifically, all mosaics are resampled to the PACS 500 μm PSF and 36 arcsec pixel scale. We have also characterized this dataset in terms of integrated photometry, maps, and radial profiles. We have notably found that the optical light profile follows a single exponential beyond the 10 kpc ring, and that the $g' - i'$ colour profile follows a single slope across both the bulge and disk. These photometric results call for detailed stellar population and dust modelling of M31's spatially-resolved SEDs.

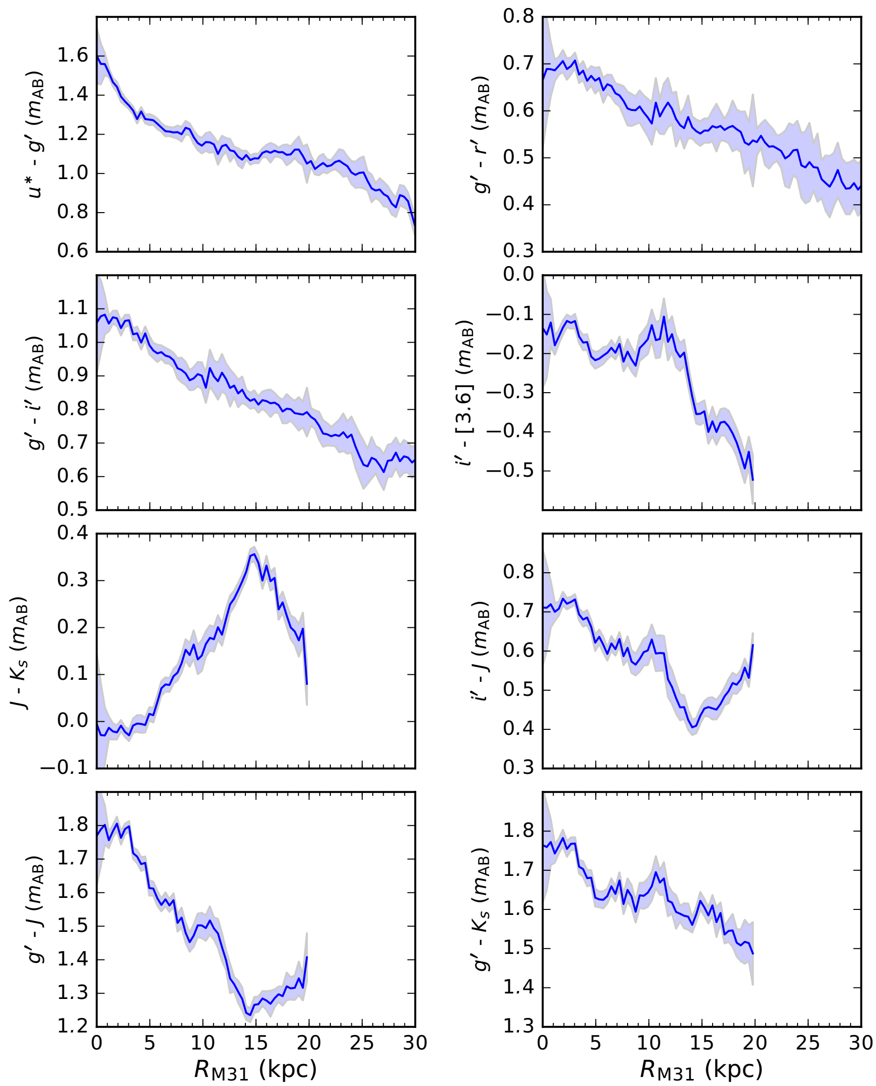


Figure 5.11: Radial colour profiles.

We pursue this in the next chapter.

Table 5.2: Integrated UV, optical and NIR fluxes of M31, presented in units of Jansky (Jy). These fluxes are integrated within the isophotal ellipse (§ 5.5) of the given radius using images calibrated according to § 5.3. See Table 5.3 for comparable integrated fluxes in IRAC and Herschel bands.

R_{31} (kpc)	FUV	NUV	u^*	g'	r'	i'	J	K_s
2	0.1	0.3	9.0	37.4	70.3	100.4	195.4	191.7
5	0.2	0.6	17.3	66.1	124.6	175.3	337.6	332.8
10	0.7	1.5	29.2	102.1	189.2	261.7	489.2	498.5
15	1.0	2.2	37.1	124.3	227.3	311.8	569.4	597.1
20	1.1	2.5	41.6	136.5	247.6	337.6	610.2	648.4

Table 5.3: Integrated IRAC and Herschel fluxes within isophotal apertures. These fluxes are integrated within the isophotal ellipse (§ 5.5) of the given radius using images calibrated according to § 5.3. See Table 5.2 for comparable integrated fluxes in GALEX, MegaCam, and WIRCam bands.

R_{31} (kpc)	[3.6]	[4.5]	[5.8]	[8.0]	[100]	[160]	[250]	[350]	[500]
2	88.8	51.5	48.6	29.7	279.9	300.6	60.8	60.5	21.7
5	153.8	88.9	92.1	62.9	853.3	1379.8	415.3	412.6	159.8
10	225.8	130.1	154.0	133.5	2209.1	4176.3	1552.1	1544.2	643.4
15	266.8	151.9	187.4	165.1	2884.0	5500.4	2274.0	2264.2	981.6
20	284.2	159.2	203.9	166.9	3137.3	5712.0	2472.4	2463.1	1089.9

Chapter 6

Stellar Populations Inferred from Spectral Energy Distribution Modelling

In this Chapter we model the observed SEDs of pixels in the ANDROIDS mosaic dataset (Chapter 5) to infer information about the spatially-resolved stellar populations and dust distribution in M31.

In § 6.1 we review the problem of SED modelling and the approaches explored in other studies. We then look in § 6.2 at the MAGPHYS SED modelling package (da Cunha et al. 2008) in greater detail in order to understand the impact of its library composition and algorithm ANDROIDS SED modelling. We further present the basic set of MAGPHYS models and examine the goodness of fit, broken down by individual bandpasses in the SED. In the remaining sections we analyze individual model parameters: stellar mass (§ 6.3), metallicity (§ 6.4), age (§ 6.5), and dust mass and attenuation (§ 6.6).

6.1 Introduction

SED modelling is an attempt to infer the properties of stellar populations and the interstellar medium in a galaxy through the observed luminosity in several bandpasses. A number of statistical methods have been proposed to allow this inference. In this section we will explore the main methods and explain why Bayesian marginalization with a library of pre-computed SEDs is well-suited for ANDROIDS.

In its most basic form, SED modelling involves an observed SED consisting of fluxes $\{F\}_X$ in a series of bandpasses, X. This observed SED has observational uncertainties $\{\sigma\}_X$. We can produce an astrophysical model of the stellar and dust flux in each bandpass, $f(\{\theta\})$ where $\{\theta\}$ is a set of stellar population and dust parameters. Stellar population synthesis codes normalize the cumulative stellar mass of model realizations to 1 $M_\odot$. Thus we write the modelled flux as $\mathcal{M}_* f(\{\theta\})$, where $\mathcal{M}_*$ is a model parameter for the total stellar mass.

Given that in large numbers, photon fluxes follow a Normal distribution, we can write a χ^2 goodness-of-fit estimator for the model SED to the observed SED given model parameters as

$$\chi^2(\mathcal{M}_*, \{\theta\}) = \sum_X \left(\frac{F_x - \mathcal{M}_* f_X(\{\theta\})}{\sigma_X^2} \right)^2 . \tag{6.1}$$

The usual practice (followed by da Cunha et al. 2008; Taylor et al. 2011) is to estimate $\mathcal{M}_*$ by analytically minimizing Eq. 6.1, giving

$$\mathcal{M}_* = \frac{\sum_X \frac{F_X f_X(\{\theta\}}{\sigma_X^2}}{\sum_X \left(\frac{f_X(\{\theta\}}{\sigma_X} \right)^2}. \tag{6.2}$$

While mass estimation is a linear normalization, the real difficulty is estimating the remaining high-dimensional parameter space in $\{\theta\}$.

6.1.1 The Markov Chain Monte Carlo Estimation Approach

One approach to solving complex functions that depend on multiple, possibly correlated, parameters is the Markov Chain Monte Carlo (MCMC) method, which we previously explored in Chapter 4. Indeed, the MCMC samples the posterior likelihood space of the model, $\log \mathcal{L}(\{\theta\}|F)$. Equation 6.1, though, is proportional to $\log \mathcal{L}(F|\{\theta\})$, the probability of the observed SED given the model. With Bayes' Theorem (Eq. 4.20), we can invert this likelihood to write the posterior likelihood as

$$\log \mathcal{L}(\{\theta\}|F) \propto -\sum_X \left(\frac{F_X - \mathcal{M}_* f_X(\{\theta\})}{\sigma_X^2} \right) + \sum_i \log p(\theta_i). \tag{6.3}$$

Here, $\log p(\theta_i)$ is the prior probability of a model parameter value θ_i. Prior probability distributions are set for each parameter to instill astronomical knowledge and constrain a parameter to a physically permissible domain.

In the MCMC algorithm, we generate samples of the parameters, $\{\theta\}$, in proportion to the underlying $\mathcal{L}(\{\theta\}|F)$ posterior probability distribution. The Metropolis-Hastings sampler (Metropolis et al. 1953; Hastings 1970), for example, does this by proposing a random step in the $\{\theta\}$ parameters from a given state $\{\theta\}_i$ to a new *candidate* state, $\{\theta\}_{i+1}$. The Metropolis-Hastings algorithm chooses whether to accept the candidate state by computing an acceptance ratio,

$$R = \frac{\mathcal{L}(\{\theta\}_{i+1}|F)}{\mathcal{L}(\{\theta\}_i|F)}. \tag{6.4}$$

If $R > 1$ the candidate state is automatically accepted. However, if $R < 1$, the new candidate state is accepted only with a random probability of R (if the candidate is not accepted, the current state is maintained). The Metropolis-Hastings algorithm repeats this process of randomly stepping through parameter space for hundreds, and possibly thousands, of steps to build up an ensemble of model parameters. Given this ensemble, parameter estimation is as straightforward as taking the mean of a parameter across the entire ensemble. Furthermore, the distribution of parameters can be interpreted directly as the posterior probability distribution of that parameter.

While MCMC is a powerful method because it not only produces point estimates, but also defines the full posterior likelihood space of all parameters, it is not a computationally efficient approach. Each pixel requires several hundred, to several thousand stellar population synthesis computations, yielding tens of millions of population synthesis computations to model all of M31.

The Model Library Maximum Likelihood Approach

An alternative to computing synthetic stellar population and interstellar medium models at each step in an MCMC chain is to pre-compute a grid of models and their SEDs, f. A particularly straightforward approach is to generate a hyperdimensional grid of models, where model parameters are regularly spaced over their respective ranges. Then given a particular SED, F, one computes the χ^2 (Eq. 6.1) of each model and selects the model with the minimum χ^2 as the best fit. This is a maximum likelihood approach.

This maximum likelihood estimate can be a biased estimator, though. Taylor et al. (2011) found that the maximum likelihood estimate, when made this way, can bias mass-to-light ratios, $\log \mathcal{M}_*/\mathcal{L}_i$, upwards by 0.1 dex compared to a Bayesian estimator. For reference, 0.1 dex is the typical 1 σ uncertainty of the $\log \mathcal{M}_*/\mathcal{L}_i$ estimates themselves.

Another issue demonstrated by Taylor et al. (2011) is that maximum-likelihood estimation with a library of models introduces discretization in estimated stellar population parameters. Metallicity can take on only one of a few tens of metallicity values, for example, included in most stellar isochrone libraries (such as Marigo et al. 2008). This itself is a source of bias in estimates.

The Model Library Marginalized Posterior Likelihood Approach It is clear that a library-based modelling approach is more efficient because it allows us to precompute and cache stellar population synthesis calculations. It is also clear that using a frequentist approach of choosing a single best-fitting model in the library can lead to biased estimates. The advantages of both can be combined in a Bayesian method where parameters are estimated by marginalizing over the likelihood of models in a library. This is the approach taken by Kauffmann et al. (2003), Taylor et al. (2011), and implemented in the MAGPHYS software package (da Cunha et al. 2008). Here we will review the basic principles of the library marginalization SED modelling method.

The foundation of the method is the construction of the model library. The library consists of a fixed number of model realizations. MAGPHYS, for instance, consists of 50000 stellar population models (da Cunha et al. 2008). For each realization, each model parameter is randomly sampled from a prior distribution. Recall that the Bayesian posterior likelihood (Eq. 6.3) includes a prior probability term, $p(\{\theta\})$. This prior probability is used by the MCMC algorithm to guide the chain, in convolution with the likelihood, to sample the posterior probability distribution. In the library-based estimation method, the sampling of models in the parameters in the model space serves the role of prior probability. Indeed, the random sampling of model parameters can use exactly the same prior probability function as would be used by MCMC.

Given an observed SED, with fluxes $\{F\}$ and uncertainties $\{\sigma\}$, the algorithm iterates through each library model i and computes a posterior likelihood, $\log \mathcal{L}_i(\{\theta\}_i|\{F\})$. This is a two-step process. First we linearly estimate the stellar mass (Eq. 6.2) given that each model is normalized to 1 $M_\odot$. Then that mass is used in the calculation of the posterior probability, which is proportional to $\exp(-\chi^2)$ where χ^2 is given in Eq. 6.1.

For each model parameter, θ_j, one can bin the library models and in each bin add up the posterior probabilities of models. Such a histogram, when normalized to unity, is the *marginalized posterior probability distribution function (pdf)* of that parameter. To estimate a parameter, one computes the mean of the marginalized posterior pdf, which is equivalent to

$$\widehat{\theta_j} = \frac{\sum_{i=0}^{N} \theta_{i,j} e^{-\chi^2(\mathcal{M}_{*,\rangle},\{\theta\}_i|F)}}{\sum_{i=0}^{N} e^{-\chi^2(\mathcal{M}_{*,\rangle},\{\theta\}_i|F)}} \tag{6.5}$$

for N models in the library. Similarly, to establish confidence intervals one computes confidence intervals from percentiles of the marginalized posterior pdf.

A variant on this algorithm is colour-colour diagram look-ups. There, one again computes a library of models (sampled according to realistic priors) and then bins those diagrams into a colour-colour (Hess) diagram. In each Hess pixel, one computes the mean value of model parameters, along with the sample standard deviation of model parameters as a measure of uncertainty. Given an observed SED, one estimates model parameters and uncertainties by finding the corresponding pixel in the Hess diagram. Zibetti et al. (2009) demonstrated this approach for computing stellar mass-to-light ratios in pixels of nearby galaxies. This algorithm is still *Bayesian*, and works when the observed SED has three or four bandpasses. However, it may be more biased than the full library marginalization algorithm described earlier because it only marginalizes against models in a single Hess pixel; the Hess look-up algorithm does not accommodate uncertainties in the observed colour, for example.

6.2 The MAGPHYS Library and SED Estimation Method

MAGPHYS (da Cunha et al. 2008) uses a library-based approach (described previously in § 6.1) to estimate stellar population and dust models in panchromatic SEDs. Given the intrinsic efficiency of the library-based estimation approach, we will use MAGPHYS in this Chapter to model SEDs in the ANDROIDS dataset. In this section we review the MAGPHYS package and understand its application to ANDROIDS data. First in § 6.2.1 we review the MAGPHYS model construction and the unique estimation method that balances stellar and dust models in § 6.2.2. We find that MAGPHYS needs to be modified to work with local galaxies and M31 pixels, and these modifications are outlined in § 6.2.4. Finally in § 6.2.3 we compare the space of MAGPHYS library SED colours with observed M31 colours to verify that M31 SEDs are encompassed by the MAGPHYS model library.

6.2.1 The MAGPHYS Model Library

The composition of a stellar population and interstellar medium model library is critically important for determining and understanding parameter estimates. da Cunha et al. (2008) describes these parameters in detail. Here we review the most relevant aspects of the MAGPHYS library to support later interpretations.

Stellar Population Models

We use MAGPHYS with a base (unpublished) Charlot & Bruzual (2007) stellar population synthesis code. This code is an updated version of Bruzual & Charlot (2003) that takes advantage of the newer Marigo & Girardi (2007) isochrones for thermally pulsating asymptotic giant branch (TP-AGB) stars. This TP-AGB phase primarily affects the NIR luminosity function for intermediate age, $\sim 1-2$ Gyr stellar populations (see Fig. 7 of Maraston 1998, for an excellent visualization of the luminosity contribution expected from each stellar phase over time). MAGPHYS uses the Chabrier (2003) initial mass function (IMF) in conjunction with these isochrones. The Chabrier (2003) IMF is tuned for the Milky Way disk, which is applicable to the M31 disk as well.

The star formation histories of the MAGPHYS library are based on that developed by Kauffmann et al. (2003). Recall that the distribution of model parameters defines a Bayesian prior. The Kauffmann et al. (2003) library distribution is designed to mirror the distributions of galaxies in the Sloan Digital Sky Survey Data Release 1, which is to say that the priors are not highly informative towards a single type of galaxy.

The star formation histories of models in the MAGPHYS library are a combination of two components: a continuous exponentially-declining star formation rate and a set of stochastic star bursts. The continuous component is characterized as

$$\mathrm{SFR}(t) \propto \exp\left\{-\frac{t_0 - t}{\tau}\right\}, \tag{6.6}$$

where parameters are: τ, the e-folding timescale of star formation in Gyr; t_0, the time when star formation began after the Big Bang in Gyr; and t, the time of star formation defined for $(t \geq t_0)$ in Gyr. In the model library, t_0 is sampled uniformly in the library, 0.1 Gyr $< t_0 <$ 13.5 Gyr. The inverse of the star forming rate's e-folding time, τ^{-1}, is sampled according to a distribution $p(\tau^{-1}) \propto 1 - \tanh(8\tau^{-1} - 6)$, which ensures that galaxies in the library are still actively forming stars.

The second star formation component is stochastic star formation bursts. For each library model, MAGPHYS uniformly samples burst events at times between t_0 and the present. The total number of bursts associated with each model was normalized so that half of the models in the library experienced a burst within the last 2 Gyr. The amplitude of bursts is set such that the mass fraction of stars formed in a burst to that formed by the continuous (exponentially declining) mode is logarithmically distributed according with limits $0.03 < \mathcal{M}^{*}{}_{\mathrm{burst}}/\mathcal{M}^{*}{}_{\mathrm{continuous}} < 4.0$. In general, this allows MAGPHYS star formation histories to be more stochastic by virtue of having multiple burst epochs.

Models have single metallicities that are linearly distributed in the range $-1.7 \leq \log Z/Z_{\odot} \leq 0.3$. As is common in the current state-of-the-art, there is no chemical evolution associated the star formation history.

For starlight attenuation by dust, MAGPHYS uses the Charlot & Fall (2000) two-component dust model, which assigns different effective optical depths to stars younger than 10 Myr that reside in a stellar birth cloud ($\widehat{\tau}_{\lambda}^{\mathrm{BC}}$) from the effective dust optical depth in the ISM that applies to older stars ($\widehat{\tau}_{\lambda}^{\mathrm{ISM}}$). The total V-band effective optical depth ($\widehat{\tau}_V = \widehat{\tau}_V^{\mathrm{BC}} + \widehat{\tau}_V^{\mathrm{ISM}}$) is sampled such that $p(\widehat{\tau}_V)$ is approximately constant in the interval (0, 4) and approaches 0 at $\widehat{\tau}_V > 6$. MAGPHYS samples an additional parameter, μ:

$$\mu \equiv \frac{\widehat{\tau}_V^{\mathrm{ISM}}}{(\widehat{\tau}_V^{\mathrm{BC}} + \widehat{\tau}_V^{\mathrm{ISM}})} \tag{6.7}$$

to determine the individual birth cloud and ISM effective optical depths for a given model. The prior distributions for both μ and $\widehat{\tau}_V$ match those seen in SDSS galaxy samples by Brinchmann et al. (2004) and Kong et al. (2004).

Dust Emission Models

Independent from the library of stellar population models, MAGPHYS also includes a library of dust emission models. The infrared dust emission SED is computed as a sum of three components: polycyclic aromatic hydrocarbon dust (PAH) emission, continuum emission from hot dust near stellar birth clouds, and cold continuum dust emission in the interstellar medium.

PAH molecules emit broad emission lines in the near and mid-infrared at 3.3, 6.2, 7.7, 8.6, 11.3 and 12.7 μm. As such, PAH lines are key sources in Spitzer IRAC (with bands centred at 3.6, 4.5, 5.8 and 8.0 μm). da Cunha et al. (2008) build the PAH emission template from a combination of ISO/ISOCAM observations of the M17 star forming region (Madden et al. 2006) along with analytical lines for the 3.3 μm line and a modified blackbody ($T = 880$ K) continuum that appears to be associated with PAH-emission.

Mid-infrared dust emission is characterized by small dust grains that are heated by UV photons in star forming regions. da Cunha et al. (2008) model this hot dust continuum emission as a equal combination of 130 K and 250 K blackbodies.

The far-infrared dust emission is composed of two blackbodies associated with warm dust in star forming regions and cold dust of the ISM that is heated only by the ambient interstellar radiation field. Here da Cunha et al. (2008) provide greater flexibility in the model since the temperatures of these blackbodies depend strongly on the strength of the stellar radiation fields. Models in the MAGPHYS library have warm dust in stellar birth clouds with temperatures, $T_{\mathrm{W}}^{\mathrm{BC}}$, sampled between 30 and 60 K. Models have cold dust with temperatures, $T_{\mathrm{C}}^{\mathrm{ISM}}$, sampled between 15 and 25 K.

Besides these two temperatures, the MAGPHYS dust emission model consists of scaling factors for the relative luminosity from each dust emission component. In total, the dust model consists of seven parameters (Table 1 of da Cunha et al. 2008). Of these, a key parameter is f_{ν}^{IR}, which defines the fraction of the infrared luminosity that is emitted by cold dust in the ISM. In other words, a luminosity of $f_{\nu}^{\mathrm{IR}} L_{\mathrm{d}}^{\mathrm{tot}}$ is emitted by dust heated by a galaxy's ambient interstellar radiation field and $(1 - f_{\nu}^{\mathrm{IR}}) L_{\mathrm{d}}^{\mathrm{tot}}$ is emitted by dust heated by stellar birth clouds.

Like the stellar model library, MAGPHYS includes a pre-generated library of these infrared dust emission models. As described in the next section, these two libraries are combined to yield a library of panchromatic SEDs.

6.2.2 The Combined MAGPHYS Stellar and Dust Model Library

A novel aspect of MAGPHYS is its ability to pair the optical SED with an infrared dust emission SED based on energy balance. In MAGPHYS, the optical SED (in which dust *absorbs* luminosity) and infrared SED (where dust emits energy) libraries are created completely independently.

The parameters that determine dust absorption in the optical models are the optical depths to the ISM, τ_{ISM}, and birth clouds, τ_{BC}, via the Charlot & Fall (2000) model. Combined with the star formation rate, SFR(t), and stellar radiation field, $S_\lambda(t)$, parameters allow MAGPHYS to compute the stellar luminosity absorbed by dust:

$$L_{\mathrm{d}}^{\mathrm{BC}}(t) = \int_0^\infty \mathrm{d}\lambda \left(1 - e^{\hat{\tau}_\lambda^{\mathrm{BC}}}\right) \int_0^{10^{\mathrm{Myr}}} \mathrm{d}t'\mathrm{SFR}(t-t')S_\lambda(t'), \tag{6.8}$$

$$L_{\mathrm{d}}^{\mathrm{ISM}}(t) = \int_0^\infty \mathrm{d}\lambda \left(1 - e^{\hat{\tau}_\lambda^{\mathrm{ISM}}}\right) \int_{10^{\mathrm{Myr}}}^{\mathrm{t}} \mathrm{d}t'\mathrm{SFR}(t-t')S_\lambda(t'), \tag{6.9}$$

$$L_{\mathrm{d}}^{\mathrm{tot}}(t) = L_{\mathrm{d}}^{\mathrm{BC}}(t) + L_{\mathrm{d}}^{\mathrm{ISM}}(t). \tag{6.10}$$

Thus the fraction of stellar energy absorbed by dust in the ambient ISM is

$$f_\mu^* \equiv \frac{L_{\mathrm{d}}^{\mathrm{ISM}}}{L_{\mathrm{d}}^{\mathrm{tot}}}. \tag{6.11}$$

This f_μ^* is equivalent to the f_μ^{IR} parameter for dust emission in the ambient ISM. This allows MAGPHYS to match each optical SED model to all infrared models with that are identical within 15%: $f_\nu^* = f_\nu^{IR} \pm 0.15$. After the 50000 optical and 50000 infrared SED models are matched, the total MAGPHYS SED library consists of 600 million SED instances.

6.2.3 Correspondence of MAGPHYS and Observed M31 Colour Spaces

The zeroth-order acceptance test for a SED modelling package is whether it can reproduce observed galaxy colours. For library-based stellar population packages this is straightforward to test since all permitted colours are pre-computed. If MAGPHYS can successfully model the ANDROIDS SEDs, the MAGPHYS library must have colours that span the domain of observed colours, at a minimum.

In Figure 6.1 we show the distribution of MAGPHYS in four colour-colour diagrams. These colour-colour diagrams span common colour indices from u^* to IRAC 3.6 μm bands. Note that these colours are confined to MAGPHYS's stellar emission model library, which is independent of the dust emission library. Against these library colour distributions, we plot in Figure 6.1 the corresponding distributions of ANDROIDS pixel colours.

For most colour-colour diagrams, the ANDROIDS SED is compatible with the MAGPHYS model space. In $g' - i'$ versus $J - K_s$, the ANDROIDS SEDs occupy the edge of the MAGPHYS model space, albeit an edge that is highly probable. On the other hand, in $u^* - g'$ versus $g' - i'$ the observed colour locus is too red in $u^* - g'$ (or too blue in $g' - i'$) by 0.5 mag. In $g' - r'$ versus $r' - i'$, the observed ANDROIDS colours are outside the MAGPHYS model space entirely, also by 0.5 mag.

It is surprising that MAGPHYS library is more limited in reproducing *optical* colours than *near-infrared* galaxy colours. Compared to the near-infrared, and the impact of luminous TP-AGB stars, the optical colours of stars are typically regarded as better known. One possibility is that the MAGPHYS attenuation model, which affects optical colours more strongly than near-infrared colours, is too limited to reproduce reddening patterns seen in M31.

6.2.4 Adapting MAGPHYS for M31 Pixel Fitting

MAGPHYS was originally intended for fitting the SEDs of entire distant galaxies. This application allowed MAGPHYS to make practical assumptions that prevent it from being used in high-resolution studies of nearby galaxies.

First, MAGPHYS normalizes its model SEDs according to luminosity distances computed from the observed galaxy's redshift (and a cosmological model). This is appropriate for distant galaxies, but inappropriate for galaxies in the Local Group where redshifts are not driven by cosmological expansion. Specifying an effective $z = 0$ redshift cannot work either since this does not reflect M31's known distance. Fixing an accurate distance is necessary for estimating the stellar and dust masses in M31 pixels; distance and mass are both proportional to the observed flux. For this study, we resolve this issue by hard-coding M31's known distance of 785 ± 25 kpc (McConnachie et al. 2005) into the MAGPHYS source, by-passing luminosity distance calculations.

The second necessary modification is less obvious. Recall that MAGPHYS estimates individual model parameters by marginalizing over a library of stellar population and dust realizations. For each model in its library, MAGPHYS computes the likelihood that a model corresponds to the observed SED. Then MAGPHYS adds that probability to the corresponding bin of a histogram of each parameter. The bounds and binning of these histograms are fixed in MAGPHYS's source code. While this approach works for parameters like stellar metallicity, where the minimum and maximum stellar metallicity are fixed by the stellar population synthesis code, it is inappropriate for mass-dependent quantities, including stellar mass, dust mass, and dust luminosity. M31 pixels have an entirely different mass scale than whole galaxies meaning that mass and luminosity estimates from individual models would fall outside MAGPHYS's predefined marginalization histograms. We solved this issue, for this study, by shifting the histogram bounds to appropriate domains for Herschel SPIRE-sized pixels on M31. Ideally, a stellar population code should be able to do this dynamically at runtime.

6.2.5 MAGPHYS Models of the Full and Subset ANDROIDS SED

We have applied MAGPHYS modelling to each pixel in the ANDROIDS dataset, as constructed in Chapter 5. In this section we outline the fitting campaigns and characterize the general goodness of fit.

In addition to fitting the full SED, we have also produced models where single instruments were dropped from the fitted SED. That is, in one fitting campaign the two GALEX bands were dropped (the "No GALEX" campaign), in another the MegaCam bands were dropped (the "No MegaCam" campaign), and so on. Thus for every pixel we report seven models: one with the full SED and six more with individual instruments dropped. This experiment allows us to characterize the influence of different segments of the SED on the overall model. For example, we can establish the importance of WIRCam SEDs in constraining stellar metallicity, or whether systematics associated with individual instruments are biasing results.

6.2.6 Goodness of Fit

Before we begin interpreting the astrophysical parameter estimates of the MAGPHYS models, in this section we quantify the goodness of fit of the MAGPHYS models. In general, we quantify goodness of fit by the minimization of flux residuals between the model and observations. These residuals can either be scalars (such as the χ^2 metric, Eq. 6.1) or broken down by individual bandpasses.

In Figure 6.2 we show the distribution of model χ^2 values, binned by the M31 disk radius of each pixel. The median χ^2 of the full SED fits is 12, which is substantially larger than seen in the MAGPHYS fits performed by HELGA IV, $\langle\chi^2\rangle < 1$ (their Fig. 2). This χ^2 statistic is improved in fitting campaigns where we jettison the

Herschel SED bands. In the inner M31 region, eliminating Herschel PACS SEDs improves the χ^2 mode to 6. Re-introducing PACS but removing SPIRE improves the χ^2 further to a mode of 1. This suggests issues with our construction of the Herschel SED despite using the same reduced dataset as HELGA IV, as discussed in Chapter 5.

MAGPHYS fits the core optical and near-infrared SED well for all modelling campaigns. In Figure 6.3 we show surface brightness residuals (units of magnitudes) in the u^* through 3.6 μm bands for each fitting campaign. For MegaCam bands these residuals are unbiased at a level of < 0.1 mag, with a similar level of scatter. This result is better than might be expected from Figure 6.1, where observed ANDROIDS colours did not coincide with the MAGPHYS library space at all on a $g' - r' - r' - i'$ colour-colour diagram. Recall, though, that MAGPHYS fits an SED by first estimating a mass normalization (Eq. 6.2) and marginalizing over the resultant library of likelihoods (Eq. 6.3), which is different from fitting colours in a colour-colour space, as done by Zibetti et al. (2009). This mass normalization, and working in the full SED space, effectively provides flexibility to the SED modelling that directly computing colours does not.

Interestingly, the residuals in these optical and near-infrared bands are relatively unaffected by dropping infrared bands, particularly the PACS and SPIRE bands that negatively impact the overall χ^2 of the fits. This result shows that mid- and far-infrared bands do not significantly leverage the optical and near-infrared stellar emission SEDs fit by MAGPHYS. The stellar and dust emission SEDs are isolated despite the energy-balance that MAGPHYS enforces (this is not true of the UV stellar emission SED, see below).

Another interesting result is that the WIRCam SED does not compromise fits to the optical MegaCam SED. Taylor et al. (2011), with a similar library-based SED fitting algorithm, found that their optical SDSS and near-infrared UKIRT SEDs could not be fit simultaneously. We do not find any evidence of that phenomenon here: MegaCam residuals in Figure 6.3 are not minimized with the "No WIRCam" fitting campaign. At the same time, WIRCam fits *are* biased at a level of 0.1 mag. This bias, though, seems driven by the need to fit Spitzer IRAC SEDs. In the "No IRAC" fitting campaign the WIRCam SEDs are minimized.

In Figure 6.4 we gain a panchromatic perspective of the MAGPHYS fitting residuals and find that the goodness of fit of the optical and near-infrared bands is an exception, not the norm. In GALEX bands we see typical residuals of 50% of the flux, or more. Dropping PACS and SPIRE bands from the dataset does improve the GALEX fits, however. Here we do see a coupling between dust absorption in GALEX UV bands and dust emission in Herschel bands, though this dataset does not appear compatible with the observations.

Of the IRAC bands, the 5.8 μm band has the greatest systematic fitting bias (25% of the flux). The 5.8 μm image is dominated by PAH emission lines, suggesting that the MAGPHYS model of PAH emission is not compatible with that seen in M31.

Finally, the Herschel PACS 160 μm and SPIRE bands all show significant systematic fitting residuals (10% – 50%). As we have seen in the improvement of other bandpasses when the Herschel fluxes are dropped, here we see directly that the Herschel SEDs are poorly modelled. Given the success that HELGA IV have in fitting these same Herschel datasets with MAGPHYS, we can only speculate that our Herschel SED extraction process is faltering and yielding an erroneous SED for these fits.

6.3 Stellar Mass Estimation

Stellar mass is a key parameter in our understanding of the formation and evolution of galaxies (see the recent review of stellar mass in galaxies by Courteau et al. 2014). By observing population distributions of galaxy stellar mass as a function of redshift, we witness the formation of galaxies through a combination of gas accretion and mergers. Velocity-luminosity relationships (Tully & Fisher 1977) and the Baryonic Tully-Fisher Relation between a spiral galaxy's baryonic (stellar and gas) mass to its total dynamical mass underscore how tightly coupled star formation physics is to dark matter cosmology. In this section we seek to use M31, which is a well-studied galaxy, as a platform to identify systematic biases in different stellar mass estimation methods.

Perhaps distinct from other physical stellar population and interstellar medium parameters, stellar mass is ideally estimated from SED datasets. With SEDs, stellar mass is driven by the flux in the SED and the distance to the object. Specifically, MAGPHYS estimates stellar mass in individual model pixels by marginalizing over linearly-fit flux scaling terms (Eq. 6.2) that appear in the SED likelihood expression (Eq. 6.3). Distance is fixed to a prior (785 kpc, McConnachie et al. 2005, and see also § 6.2.4). Thus MAGPHYS's stellar mass estimates are different in nature than other parameters, where it marginalizes over a library of pre-computed models. Overall, stellar mass estimates are less affected by SED shape (or colour) than age, metallicity, or dust attenuation parameters.

Since stellar population properties are a secondary factor in interpreting the stellar mass of an SED, it is common to model stellar mass-to-light ratios ($\mathcal{M}_*/L_\lambda$) as a function of stellar population's broadband colour, known as colour-mass-to-light-ratio

(CMLR) relations (for more recent reviews, see Roediger & Courteau 2015; Zhang et al. 2017, and references therein). Taylor et al. (2011) find that the uncertainty of $\log_{10}\mathcal{M}_*/L_i$ given a $g-i$ colour is ± 0.1 dex. The advantage of CMLRs is that they are an economical method for estimating stellar masses of galaxies. Only two photometric measurements are required (to establish a colour), and one of those photometric measurements also establishes the luminosity factor. In this section, we describe ANDROIDS stellar mass estimates made with MAGPHYS (§ 6.3.1), and compare those estimates to optical (§ 6.3.3) and mid-IR (§ 6.3.4) CMLRs, and results from literature (§ 6.3.5).

6.3.1 MAGPHYS Stellar Mass Estimates

We used MAGPHYS to model the mass of M31 in individual ANDROIDS pixels sampled in the HELGA IV pixel frame. In Figure 6.7, we plot the cumulative stellar mass as a function of deprojected disk radii. The deprojected radius of each pixel is known from isophotal fitting (§ 5.5). Table 6.1 summarizes cumulative stellar masses at key radii. We find that, within 20 kpc, the stellar mass of M31 is 5.4×10^{10} $M_\odot$. This estimate includes M32 since its light is projected within M31 isophotes between 8 kpc and 12 kpc. NGC 205 *is not* included in the mass estimate since it is outside the 20 kpc isophote.

The statistical uncertainty of the 20 kpc stellar mass estimate is $\pm 3.1 \times 10^8$ $M_\odot$, or 0.6% of the total mass. This uncertainty is established by bootstrapping from the pixel mass sample variance. That is, within narrow isophotal annuli we computed the sample standard deviation of individual pixel mass estimates. We assigned these sample standard deviations as the mass uncertainty to all pixels within each isophotal annulus. Then, we generated 10^3 new mass maps by randomly sampling from the uncertainty distribution of each pixel, and computed a cumulative stellar mass from each map realization. Note that MAGPHYS generates internal stellar mass uncertainty estimates from the marginalization of M_* over the model library (Eq. 6.5), though in practice we found the mass probability distribution function (pdf) was too under-sampled to be usable. The MAGPHYS mass pdf could be sampled more finely, but due to the internal limitations in MAGPHYS we would need to compile versions of MAGPHYS tuned for individual isophotal radii across M31. The bootstrap method is a flexible alternative, and in fact, provides an upper limit on statistical mass estimation uncertainty since it incorporates intrinsic stellar population variance within an isophotal band.

That the statistical uncertainty is so small confirms that systematic errors are the true source of mass estimation uncertainty. We explore systematic uncertainties several ways: through the SED bandpass jackknife analysis, comparison to CMLR estimates, and comparison to wholly independent M31 mass estimates from the literature.

6.3.2 Bandpass Dependence of Stellar Mass Estimation

As described in § 6.2.5, we ran seven MAGPHYS fitting campaigns. One campaign is the full ANDROIDS SED fit with 17 bands, and the other six campaigns omit a single instrument, each, from the fit. This jackknife-like approach allows us to assess the influence of individual instruments on the overall parameter estimate.

We find that, for our MAGPHYS fits, the SED composition *is* important. As we show in Figure 6.5, modelling campaigns that omitted an instrument from the SED yield total stellar masses within 20 kpc that range from 4.9×10^{10} M$_\odot$ to 5.9×10^{10} M$_\odot$. The stellar mass estimate based on the full ANDROIDS SED is within this range: 5.4×10^{10} M$_\odot$. Campaigns that omitted UV, optical or near-IR bands tended to underestimate stellar mass (the MegaCam bands are most significant). On the other hand, campaigns that omitted bands in the mid- and far-IR SED overestimated stellar mass. This result is surprising because the PACS and SPIRE fluxes were not expected to have a substantial effect on stellar mass estimation. PACS has no stellar emission, so it can only contribute to mass estimates through dust attenuation modelling.

Figure 6.6 shows these results in terms of the i-band stellar mass-to-light ratio, $\log_{10} \mathcal{M}_*/L_i$. Note that the systematic $\log_{10} \mathcal{M}_*/L_i$ biases associated with missing bandpasses is less than the random estimation uncertainty (dashed lines in Figure 6.6, left), which is 0.05 dex within $R_{\mathrm{M31}} = 10$ kpc and 0.1 dex at $R_{\mathrm{M31}} = 15$ kpc. The salient difference is that the random estimation uncertainties are unbiased, while the small biases in $\log_{10} \mathcal{M}_*/L_i$ associated with missing bandpasses accumulate into systematic stellar mass estimation errors of $\pm 0.5 \times 10^{10}$ dex for the full M31 bulge and disk.

6.3.3 Comparison to Optical Mass-to-Light Ratio Estimators

Colour $\mathcal{M}_*/L_\lambda$ ratio (CMLR) relations, pioneered by Bell & de Jong (2001), enable one to estimate a $\log_{10} \mathcal{M}_*/L$ (and thus a mass) given two photometric measurements. As mentioned previously, this economical mass estimation method is powerful when applied to optical surveys where large samples of galaxies are observed with very few bandpasses. Bell et al. (2003) and Kauffmann et al. (2003) are two early studies that took this approach with the SDSS. CMLRs are compelling to study with M31, and ANDROIDS in particular, because their estimates can be readily compared to other estimates, including full-SED modelling. This provides a useful validation for CMLR estimates of distant galaxies where CMLRs are the only practical stellar mass estimation methods.

For the purposes of this analysis we restrict ourselves to CMLRs based on optical $g - i$ (SDSS system) colours, despite the availability of CMLRs for other optical and near-infrared bandpass combinations. This $g - i$ colour is favoured by many studies, including Zibetti et al. (2009), for its known constructive degeneracies with stellar age, metallicity and dust attenuation (Bell & de Jong 2001). That is, as a stellar population dims due to dust attenuation, the reddening vector pushes the colour along the CMLR to a higher stellar mass-to-light ratio. Zhang et al. (2017) found that other optical colours are just as accurate, also with an uncertainty of 0.18 dex (interquartile range, IQR), see their Fig. 17. Also note that both Zibetti et al. (2009) and Zhang et al. (2017) advocate two-colour mass-to-light ratio estimators, such as $g-i$ and $i-H$. Zhang et al. (2017) estimates an IQR uncertainty of such an estimator to be 0.1–0.2 dex. Again, we do not test two-colour $\mathcal{M}/L_\lambda$ ratio relations in this work.

Review of Optical Colour $\mathcal{M}_*/L_i$ Ratio Relations

In this section we study existing optical CMLRs from Zibetti et al. (2009), Taylor et al. (2011), Into & Portinari (2013), Roediger & Courteau (2015), and Zhang et al. (2017). These CMLRs, and their relationship to the stellar mass distribution estimated by MAGPHYS, are summarized in Figure 6.8.

Zibetti et al. (2009) constructed CMLRs by synthesizing a library of stellar populations and fitting a linear relationship between colour indices and $\log_{10} \mathcal{M}_*/L$. They use a Charlot & Bruzual (2007) stellar population synthesis engine with a Chabrier (2003) IMF and Charlot & Fall (2000) attenuation model. Since MAGPHYS also uses Charlot & Bruzual (2007) SPS and Chabrier (2003) IMF, the Zibetti et al. (2009) CMLR may yield masses similar to MAGPHYS models. Their CMLR, using SDSS (AB) magnitudes in g and i bands is (Zibetti et al. 2009, Table B.1):

$$\log_{10} \mathcal{M}_*/L_i(g-i) = -0.963 + 1.032(g-i). \tag{6.12}$$

The Taylor et al. (2011) CMLR is constructed from a library of stellar populations modelled with the Bruzual & Charlot (2003) synthesis code, which is an earlier version of the code used by MAGPHYS and Zibetti et al. (2009) that does not include the updated TP-AGB treatment described in Marigo & Girardi (2007) and Marigo et al. (2008). Taylor et al. (2011) does, however, adopt a Chabrier (2003) IMF and the Calzetti et al. (2000) attenuation law. Thus the Taylor et al. (2011) CMLR is interesting because it uses the same IMF (mass normalization) as MAGPHYS and Zibetti et al. (2009), but a different stellar population synthesis prescription. Their CMLR is (Taylor et al. 2011, Eq. 7):

$$\log_{10} \mathcal{M}_*/L_i(g-i) = -0.68 + 0.7(g-i). \tag{6.13}$$

We also consider the Into & Portinari (2013) CMLRs. Unlike the other studies, Into & Portinari (2013) implemented their own population synthesis code rather than relying directly upon Bruzual & Charlot (2003), Charlot & Bruzual (2007), or equivalent, in order to maintain flexibility. Their stellar populations were synthesized from Marigo & Girardi (2007) and Marigo et al. (2008) isochrones, which include TP-AGB stars. For this analysis we consider their disk galaxy models, which involve a chemically evolving stellar population with exponential star formation histories. This disk model, described in Portinari et al. (2004), connects chemical evolution to the evolving stellar population and includes gas infall. Thus unlike the Zibetti et al. (2009), Taylor et al. (2011), and indeed, MAGPHYS CMLRs, the Into & Portinari (2013) CMLRs are not based on single-metallicity stellar populations. The disk model we use here is also dust-free, though this is not entirely significant since optical CMLRs (based of $V-I$ or $g-i$) are degenerate with dust attenuation such that the dimming of light is effectively compensated by the reddening of the population.

Into & Portinari (2013) used a Kroupa (2001) IMF, which is similar to the Chabrier (2003) IMF used in the above studies. Their chemical stellar population models motivated this IMF: with bottom-light IMFs there are fewer long-lived low-mass stars to abate metal return to the ISM. The Into & Portinari (2013) $g-i$ CMLR is (their Table 5):

$$\log_{10} \mathcal{M}_*/L_i(g-i) = -0.669 + 0.985(g-i). \tag{6.14}$$

For a Salpeter (1955) IMF:

$$\log_{10} \mathcal{M}_*/L_i(g-i) = -0.4685 + 0.985(g-i). \tag{6.15}$$

Roediger & Courteau (2015) studied the uncertainties and systematics of CMLR stellar mass estimation in comparison to SED fitting. For the purposes of this study, they generated two groups of CMLRs fit from a library of synthetic stellar populations. Both CMLRs are based on the library of models in MAGPHYS, which is useful for constraining the effects of different prior choices in comparing full SED fits to CMLR-based mass estimates. Roediger & Courteau (2015) used the Chabrier (2003) IMF. One CMLR is based on the library of models in MAGPHYS da Cunha et al. (2008), though generated with Bruzual & Charlot (2003). The first CMLR is based on the MAGPHYS model library, generated with Bruzual & Charlot (2003):

$$\log_{10} \mathcal{M}_*/L_i(g-i) = -0.831 + 0.979(g-i). \tag{6.16}$$

We refer to this CMLR as the "RC15 BC03" model. Note that the full SED fits described in § 6.3.1 use the Charlot & Bruzual (2007) models, rather than the Bruzual & Charlot (2003) models, so the comparison is not entirely direct.

The second CMLR is also based on the MAGPHYS model library, but generated with FSPS (Conroy et al. 2009), which includes revised TP-AGB population synthesis:

$$\log_{10} \mathcal{M}_*/L_i(g-i) = -0.597 + 0.831(g-i). \tag{6.17}$$

We refer to this CMLR as the "RC15 FSPS".

The last set of models we consider are those by Zhang et al. (2017). A unique aspect of the Zhang et al. (2017) study is their use of realistic star formation and chemical evolution histories, as opposed to analytical star formation histories. Their base sample is based on resolved stellar populations fit in 40 Local Group galaxies with HST observations by Weisz et al. (2014). Using both Bruzual & Charlot (2003) and

FSPS population synthesis models, they converted these models — as sums of simple stellar populations of different mass, age, metallicity and extinction combinations — into broadband SEDs. The IMF is Chabrier (2003). From this set of 40 SEDs they fit CMLRs that we label "LG:"

$$\text{Z17 BC03 LG} : \log_{10} \mathcal{M}_*/L_i(g-i) = -0.505 + 0.709(g-i) \tag{6.18}$$

$$\text{Z17 FSPS LG} : \log_{10} \mathcal{M}_*/L_i(g-i) = -0.471 + 0.720(g-i). \tag{6.19}$$

Zhang et al. (2017) expanded this set of 40 SEDs to a standard sample of 86046 synthetic SEDs by randomly sampling new mass-weighted metallicities to apply to the basic star formation and metallicity evolution patterns seen in the LG SED set. Note these models maintain a fixed attenuation towards young stars: $A_{V,\text{young}} = 0.5$. We label these CMLRs as Model A:

$$\text{Z17 BC03 A} : \log_{10} \mathcal{M}_*/L_i(g-i) = -0.416 + 0.630(g-i) \tag{6.20}$$

$$\text{Z17 FSPS A} : \log_{10} \mathcal{M}_*/L_i(g-i) = -0.351 + 0.578(g-i). \tag{6.21}$$

Finally, Zhang et al. (2017) generated another set of synthetic SEDs where the attenuation towards young stars is allowed to vary according to a Charlot & Fall (2000) model, rather than being fixed to 0.5 mag. We refer to the CMLRs built from this set as Model B:

$$\text{Z17 BC03 B} : \log_{10} \mathcal{M}_*/L_i(g-i) = -0.406 + 0.552(g-i) \tag{6.22}$$

$$\text{Z17 FSPS B} : \log_{10} \mathcal{M}_*/L_i(g-i) = -0.330 + 0.494(g-i). \tag{6.23}$$

Note that all the Zhang et al. (2017) models assume a Small Magellanic Cloud Gordon et al. (2003) extinction curve.

Conversions Between the MegaCam and SDSS Photometry Systems

The CMLRs described above are valid for photometry in the SDSS AB magnitude system ($ugrizy$), while ANDROIDS includes optical photometry in the native system of MegaCam (these bands are denoted as $u^*g'r'i'$). The MegaCam system is not interchangeable with the SDSS system. For $\log_{10} \mathcal{M}_*/L_i$ we find that using MegaCam photometry in the CMLRs above results in biases of -0.1 dex. To effectively compare $\log_{10} \mathcal{M}_*/L_i$ with the published CMLRs, we convert the MegaCam photometry using bandpass transformations determined from the CFHT Legacy Survey.[1] The transformations are:

$$u = 1.318u^* - 0.377g' + 0.060r' \tag{6.24}$$

$$g = 1.186g' - 0.186r' \tag{6.25}$$

$$r = 0.981r' + 0.027g' \tag{6.26}$$

$$i = i'. \tag{6.27}$$

[1] `http://www1.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/community/CFHTLS-SG/docs/extra/filters.html`

Note that the SDSS i is also dependent on both MegaCam g' and r', though the transformation coefficients are on the order of 10^{-5}.

Results

Optical CMLRs, through the assumptions in their calibrations, show diverse behaviours and estimates of a galaxy's stellar mass. As summarized in Figure 6.7 and Tables 6.1–6.2, the optical CMLRs we describe here suggest total M31 bulge and disk stellar masses from 4.9×10^{10} $M_\odot$ to 9.9×10^{10} $M_\odot$, a range of 0.3 dex. The MAGPHYS full SED stellar mass estimate lies at the lower end of this range: 5.4×10^{10} $M_\odot$. Several studies, including Zibetti et al. (2009), Taylor et al. (2011) and Roediger & Courteau (2015), find that the stellar mass uncertainty when using a *single* CMLR is 0.1 dex. In this section, we seek to understand the diversity of CMLR estimates, and if any class of CMLR might be more correct than others.

We begin the comparison of CMLRs in the $(g - i) - \log_{10} \mathcal{M}_*/L_i$ plane, Figure 6.8. As discussed earlier, most CMLRs cover the locus of per-pixel MAGPHYS $\log_{10} \mathcal{M}_*/L_i$ estimates, though most CMLRs lie on the high-mass tail of the distribution. Of all the CMLRs, Zibetti et al. (2009) is the closest match to the ANDROIDS/MAGPHYS pixel distribution. Recall that Zibetti et al. (2009) use a nearly identical stellar library as MAGPHYS: including Charlot & Bruzual (2007) population synthesis, Chabrier (2003) isochrones, Charlot & Fall (2000) dust attenuation, and library parameter prior distributions based on Kauffmann et al. (2003). This result, in the context of highly resolved SED modelling of M31, agrees well with the conclusion of Roediger & Courteau (2015): given the minimal observation requirements (g and i-band photometry) and minimal computation requirements (the CMLR

Table 6.1: Cumulative stellar mass internal to deprojected disk radii, estimated by MAGPHYS ("Full SED") fits and other $\mathcal{M}_*/L$ indicators with ANDROIDS data. Estimators are: Z09 (Zibetti et al. 2009); T11 (Taylor et al. 2011); E12 (Eskew et al. 2012); IP13 (Into & Portinari 2013); R15 (Roediger & Courteau 2015) using Bruzual & Charlot (2003) and FSPS (Conroy et al. 2009) models (see § 6.3.3). W17 (Williams et al. 2017) is based on resolved stellar population models transformed into a $\mathcal{M}_*/L_i$ profile (see § 6.3.5). R_{M31} is in units of kpc and corresponds to the major axis radius of elliptical isophotes (§ 5.5). Masses are in units of 10^{10} $\mathrm{M}_\odot$. See Tables 6.2 and 6.3 for additional estimates.

R_{M31}	Full SED	Z09	T11	IP13	R15 BC03	R15 FSPS	W17
1	0.5	0.6	0.4	0.6	0.6	0.7	1.4
2	1.2	1.2	0.9	1.8	1.4	1.5	2.9
3	1.7	1.8	1.4	2.7	2.0	2.2	4.1
4	2.2	2.3	1.7	3.5	2.6	2.9	5.0
5	2.6	2.7	2.1	4.4	3.2	3.5	5.8
6	2.9	3.2	2.4	5.1	3.6	4.1	6.4
7	3.2	3.5	2.7	5.7	4.0	4.5	6.9
8	3.5	3.8	2.9	6.2	4.4	4.9	7.4
9	3.7	4.1	3.1	6.7	4.7	5.3	7.8
10	4.0	4.3	3.3	7.1	5.0	5.7	8.2
11	4.2	4.6	3.5	7.5	5.3	6.0	8.6
12	4.4	4.8	3.7	8.0	5.6	6.3	8.9
13	4.6	5.0	3.9	8.3	5.8	6.6	9.2
14	4.8	5.2	4.0	8.6	6.0	6.9	9.5
15	4.9	5.3	4.2	8.9	6.2	7.1	9.8
16	5.1	5.5	4.3	9.2	6.4	7.3	10.1
17	5.2	5.7	4.4	9.4	6.6	7.5	10.3
18	5.3	5.8	4.7	9.6	6.8	7.7	10.5
19	5.4	5.9	4.8	9.8	6.9	7.9	10.7
20	5.4	6.0	4.9	9.9	7.0	8.0	10.9

Table 6.2: Cumulative stellar mass internal to deprojected disk radii, estimated by the Zhang et al. (2017) $g - i$ $\mathcal{M}_*/L$ estimators with ANDROIDS data. The "BC03" estimates are based on Bruzual & Charlot (2003) population synthesis models, while "FSPS" are based on the Conroy et al. (2009) models. The "LG" estimates use CMLRs calibrated to the star formation histories of 40 Local Group dwarf galaxies. Model "A" estimates use CMLRs calibrated with a constant $A_{V,\mathrm{young}} = 0.5$ dust attenuation. Model "B" estimates use CMLRs calibrated with a variable $A_{V,\mathrm{young}}$ dust attenuation. See § 6.3.3. R_{M31} is in units of kpc and corresponds to the major axis radius of elliptical isophotes (§ 5.5). Masses are in units of 10^{10} $\mathrm{M}_{\odot}$. See Tables 6.1 and 6.3 for additional estimates.

R_{M31}	BC03 LG	BC03 A	BC03 B	FSPS LG	FSPS A	FSPS B
1	0.6	0.6	0.6	0.7	0.6	0.6
2	1.4	1.4	1.2	1.5	1.4	1.2
3	2.0	2.0	1.7	2.2	2.0	1.7
4	2.6	2.6	2.2	2.8	2.6	2.2
5	3.1	3.1	2.6	3.5	3.1	2.6
6	3.6	3.6	3.0	4.0	3.6	3.0
7	4.0	4.0	3.3	4.4	4.0	3.4
8	4.4	4.3	3.6	4.8	4.4	3.7
9	4.7	4.7	3.9	5.2	4.7	4.0
10	5.1	5.0	4.2	5.6	5.0	4.2
11	5.4	5.3	4.4	5.9	5.3	4.5
12	5.6	5.6	4.7	6.2	5.6	4.8
13	5.9	5.9	4.9	6.5	5.9	5.0
14	6.1	6.1	5.1	6.8	6.1	5.2
15	6.3	6.3	5.3	7.0	6.4	5.4
16	6.6	6.5	5.4	7.2	6.6	5.6
17	6.7	6.7	5.6	7.5	6.8	5.8
18	7.0	7.0	5.8	7.7	7.0	6.0
19	7.1	7.1	5.9	7.9	7.1	6.1
20	7.2	7.2	6.0	8.0	7.2	6.2

is pre-computed), CMLRs are the preferred approach for estimating the stellar mass of a galaxy. Full SED fitting may not be necessary for stellar mass estimation.

The Roediger & Courteau (2015) CMLRs are similar to Zibetti et al. (2009) since Roediger & Courteau also adopted the MAGPHYS stellar population library. Their use of stellar population synthesis models is, however, different; Roediger & Courteau (2015) rely on the Bruzual & Charlot (2003) models rather than the Charlot & Bruzual (2007) models with revised TP-AGB treatments used in the Zibetti et al. (2009) CMLR. The full SED MAGPHYS fitting is, of course, different. Figure 6.8 shows that the Roediger & Courteau (2015) "BC03" CMLR has a slope comparable to Zibetti et al. (2009)'s CMLR, but is heavier by 0.1 dex.

The Roediger & Courteau (2015) "FSPS" CMLR demonstrates the effect of changing the population synthesis models on $\log_{10} \mathcal{M}_*/L_i$ whilst maintaining the same IMF and library parameter distributions. Overall the Roediger & Courteau (2015) "FSPS" CMLR has a higher zeropoint (equivalent to 0.1 dex at $g - i \sim 1$), but also a shallower slope in Figure 6.8. Compared to the Zibetti et al. (2009) IMF, the Roediger & Courteau (2015) "FSPS" CMLR is 0.2 dex heavier. This result is remarkable because both CMLRs are built from roughly identical model parameter distributions and IMF. Both use population synthesis models that incorporate revised TP-AGB treatments (Charlot & Bruzual 2007 and FSPS, respectively). At face value, we interpret this as an indication that FSPS is calibrated to effectively estimate masses 0.2 dex heavier than Charlot & Bruzual (2007).

Substantial differences between FSPS and BC03-based population synthesis models do not appear in the Zhang et al. (2017) models, though. In Figure 6.8, both the "Z17 FSPS A" and "BC03 A" CMLR models are identical at a level of < 0.05 dex, though the FSPS-based CMLR has a shallower slope (as also seen by Roediger & Courteau). Systematic differences between BC03 and FSPS stellar population models, in terms of CMLR mass estimation, may therefore be specific to the underlying model parameter distributions. The MAGPHYS, Zibetti et al. (2009), and Roediger & Courteau (2015) model libraries are biased towards recent star formation, in addition to being monometallic. By comparison, the Zhang et al. (2017) models are design to be unbiased, while also featuring physically-based chemical evolution.

The Taylor et al. (2011) and Into & Portinari (2013) CMLRs established lower and upper bounds, respectively, as seen in Figure 6.8 for the $(g - i) - \log_{10} \mathcal{M}_*/L_i$ plane. In their study, Zhang et al. (2017) found that younger stellar populations tend to be interpreted by CMLRs as having lower masses. Taylor et al. (2011) built their stellar population library for galaxies with redshifts up to 0.65. Zhang et al. found that the model library is thus biased to stellar populations with mass-weighted ages younger than 7 Gyr, and thus the Taylor et al. CMLR tends to underestimate stellar masses when applied to older stellar populations (of which M31 is an example).

Similar to Zhang et al. (2017), Into & Portinari (2013) considered stellar populations with realistic chemical evolution histories. The Into & Portinari (2013) CMLR also incorporates revised TP-AGB handling, like Charlot & Bruzual (2007) and FSPS. It is interesting, then, that the Into & Portinari CMLR has the same slope as both the Roediger & Courteau (2015) and Zibetti et al. (2009) CMLRs that use Bruzual & Charlot (2003) and Charlot & Bruzual (2007) population synthesis models, respectively. Compared to the other CMLRs, Into & Portinari (2013) is unique for not including dust attenuation. We do see evidence that different attenuation prescriptions affect mass estimation in the Zhang et al. (2017) CMLRs. Their "A" models maintain a constant $A_V = 0.5$ mag toward young (< 40 Myr) stars, while the "B" models allow for the attenuation towards young stars to vary from 0 to 5 mag. The latter models have masses typically lower by 0.1 dex. That CMLR also has a shallower slope.

The effects of these CMLRs on the overall M31 stellar mass estimate is apparent in the radial $\mathcal{M}_*/L_i$ profile (Figure 6.9) and total mass profiles (Figure 6.7, Tables 6.1–6.2).

Summary of Optical Mass-to-Light Ratio Estimators

Overall, common CMLRs available in the literature show a diverse range of behaviours resulting in a 0.3 dex range of stellar mass estimates. Further study, with true controls, is necessary to fully explain all the observed behaviours. The choice of stellar population synthesis model is not as important as may be expected as both BC03 and newer FSPS and CB07 models with revised TP-AGB treatments can have similar behaviours. The most determinant factor appears to be the construction of the stellar population model library from which the CMLR is fit. Models with chemical evolution, such as Into & Portinari (2013) and Zhang et al. (2017), have higher overall mass estimates. Likewise, model libraries that do not favour more recent star bursts, like Zhang et al. (2017), also have higher overall mass estimates. Finally, models with less dust attenuation towards young stars may yield higher overall mass estimates. As we mentioned earlier, we find that a CMLR can reproduce the stellar mass estimate of a full Bayesian SED estimator. This should encourage the use of CMLRs for stellar mass estimation, provided an appropriately calibrated CMLR is used.

6.3.4 Comparison to Mid-Infrared Mass-to-Light Ratio Estimators

Like optical CMLRs based on $g - i$ colours, the mid-IR colour of a galaxy is also commonly used as a $\log_{10} \mathcal{M}_*/L$ proxy. With Spitzer, and especially WISE, many galaxies have $[3.6] - [4.5]$ colour measurements (or equivalently, $W1 - W2$), enabling survey-scale stellar mass estimation. In this section we review common Mid-IR-based stellar $\log_{10} \mathcal{M}_*/L$ estimators and compare those estimates to the MAGPHYS mass estimate based on full SED modelling.

Review of Mid-IR Mass Estimators

Eskew et al. (2012) empirically modelled a relationship between $[3.6] - [4.5]$ colour and $\log_{10} \mathcal{M}_*/L$ in the context of the Large Magellanic Cloud:

$$\log_{10} \mathcal{M}_*/L_{[3.6]} = -0.7([3.6]_{\mathrm{Vega}} - [4.5]_{\mathrm{Vega}}) - 0.23 \tag{6.28}$$
$$\text{for } -0.12 < [3.6]_{\mathrm{Vega}} - [4.5]_{\mathrm{Vega}} < 0.34.$$

Note that this relation is calibrated for magnitudes in the Vega zeropoint system. Androids photometry is natively calibrated in AB magnitudes. Thus we note that the conversion from AB magnitudes to Vega magnitudes for the Spitzer 3.6 μm and 4.5 μm bandpasses, based on data compiled in FSPS (Conroy et al. 2009), is

$$[3.6]_{\mathrm{Vega}} = [3.6]_{\mathrm{AB}} - 2.78, \tag{6.29}$$

$$[4.5]_{\mathrm{Vega}} = [4.5]_{\mathrm{AB}} - 3.25. \tag{6.30}$$

Meidt et al. (2014) established an empirical $\log_{10}\mathcal{M}_*/L$ relation from the S^{4}G Spitzer survey (Sheth et al. 2010). A unique aspect of this relation is that dust emission is subtracted from the Mid-IR fluxes used to calibrate this CMLR. Thus the Meidt et al. (2014) CMLR is idealized for purely stellar emission in Mid-IR light:

$$\begin{aligned}&\log_{10}\mathcal{M}_*/L_{[3.6]} = 3.98(\pm 0.98)([3.6]_{\mathrm{Vega}} - [4.5]_{\mathrm{Vega}}) + 0.13(\pm 0.06) \\ &\text{for } -0.14 < [3.6]_{\mathrm{Vega}} - [4.5]_{\mathrm{Vega}} < -0.04.\end{aligned} \tag{6.31}$$

The Cluver et al. (2014) $\log_{10}\mathcal{M}_*/L$ relation uses WISE $W1-W2$ photometry rather than Spitzer $[3.6]-[4.5]$. In practice, these filter sets are equivalent, with a negligible zeropoint difference between the two. Thus in this work we treat the Cluver et al. (2014) $W1-W2$ CMLR as a $[3.6]-[4.5]$ CMLR:

$$\begin{aligned}&\log_{10}\mathcal{M}_*/L_{[3.6]} = -2.54([3.6]_{\mathrm{Vega}} - [4.5]_{\mathrm{Vega}}) - 0.17 \\ &\text{for } -0.12 < [3.6]_{\mathrm{Vega}} - [4.5]_{\mathrm{Vega}} < 0.17.\end{aligned} \tag{6.32}$$

The valid domain is reported by Cluver (2017, private communication).

Finally, we can consider the assumption that quiescent galaxies have a constant $\log_{10}\mathcal{M}_*/L_{[3.6]}$ implying that 3.6 μm flux can be universally scaled into a stellar mass. Meidt et al. (2014) raised this possibility and suggested $\mathcal{M}_*/L_{[3.6]} = 0.6$ is the typical mid-IR mass-to-light ratio seen in quiescent galaxies. Kettlety et al. (2017) confirmed this by comparing stellar mass estimates from panchromatic SED modelling (Taylor et al. 2011) to $W1-W2$ CMLRs. Kettlety et al. found that a fixed $\mathcal{M}_*/L_{[3.6]} = 0.6$

reproduced the stellar mass estimated by full SED fits with an accuracy of 25%.

Results

Figure 6.10 provides an overview of the mid-IR CMLRs described previously and their comparison to the $\log_{10} \mathcal{M}_*/L_{[3.6]}$ estimated by MAGPHYS full SED fitting. First, note that both the Cluver et al. (2014) and Meidt et al. (2014) CMLRs are highly sensitive to $[3.6]-[4.5]$ colour (given their large slopes in Eq. 6.32 and 6.31). Overall, the Cluver et al. (2014) operates in bluer colours while Meidt et al. (2014) applies to redder $[3.6]-[4.5]$. Furthermore, the slopes of these two CMLRs are opposite. The key difference between the Cluver et al. (2014) and Meidt et al. (2014) CMLRs is that Meidt et al. (2014) fit their CMLR to dust-emission-free SEDs, whereas Cluver et al. (2014) include dust emission. Older stellar populations naturally become redder as hot massive stars die off, leaving low mass and less luminous red stars. This stellar evolution describes the Meidt et al. (2014) CMLR (as well as the optical CMLRs, Figure 6.8).

Cluver et al. (2014), on the other hand, includes the influence of mid-IR dust emission on the $\log_{10} \mathcal{M}_*/L_{[3.6]}$. Mid-IR dust emission comes from very hot dust in star formation regions, and from PAH (polycyclic aromatic hydride) lines. With this dust emission considered, as the $[3.6]-[4.5]$ colour increases, the mid-IR light is increasingly dominated by dust emission, which drives the mid-IR stellar mass-to-light ratio downwards.

Eskew et al. (2012), using LMC SEDs, also includes the influence of dust emission on the CMLR. Like Cluver et al. (2014), the Eskew et al. (2012) $\log_{10} \mathcal{M}_*/L_{[3.6]}$ decreases with increasing $[3.6]-[4.5]$, however the sensitivity to this colour is smaller.

Table 6.3: Cumulative stellar mass internal to deprojected disk radii, estimated by mid-IR $\mathcal{M}_*/L$ estimators with ANDROIDS data. Estimators are: E12 (Eskew et al. 2012); M14 (Meidt et al. 2014); and C14 (Cluver et al. 2014). $\Upsilon_{[3.6]} = 0.6$ is the constant 3.6 μm mass-to-light ratio advocated by Meidt et al. (2014) and Kettlety et al. (2017). $\Upsilon_{[3.6]} = 0.46$ is the constant 3.6 μm mass-to-light ratio fitted to reproduce the MAGPHYS Full SED fit at $R_{\rm M31} = 20$ kpc. $R_{\rm M31}$ is in units of kpc and corresponds to the major axis radius of elliptical isophotes (§ 5.5). Masses are in units of 10^{10} $M_\odot$. See Tables 6.1 and 6.2 for additional estimates.

$R_{\rm M31}$	E12	M14	C14	$\Upsilon_{[3.6]} = 0.6$	$\Upsilon_{[3.6]} = 0.46$
1	0.8	0.5	1.4	0.6	0.5
2	1.6	1.0	3.0	1.3	1.0
3	2.3	1.4	4.4	1.9	1.4
4	2.9	1.8	5.7	2.4	1.8
5	3.5	2.8	6.9	2.9	2.2
6	4.0	3.8	7.9	3.3	2.5
7	4.5	4.1	8.9	3.7	2.8
8	4.9	4.3	9.8	4.1	3.1
9	5.3	4.6	10.8	4.4	3.3
10	5.7	4.9	11.6	4.8	3.6
11	6.1	5.3	12.5	5.1	3.9
12	6.5	6.5	13.3	5.4	4.1
13	6.9	6.8	14.2	5.7	4.3
14	7.3	7.1	15.2	6.0	4.6
15	7.6	7.3	16.2	6.2	4.7
16	7.9	8.2	17.2	6.5	4.9
17	8.2	8.9	18.5	6.7	5.1
18	8.4	9.7	20.1	6.9	5.2
19	8.7	10.3	22.8	7.1	5.3
20	8.9	10.6	35.2	7.2	5.4

Finally, the $\mathcal{M}_*/L_{[3.6]} = 0.6$ estimator is consistent with the valid domains of each of the Eskew et al. (2012), Cluver et al. (2014), and Meidt et al. (2014) CMLRs.

We find, in Figure 6.10 that the MAGPHYS full SED-fit mass estimates are most consistent with the Meidt et al. (2014) CMLR. This result is counter-intuitive since the ANDROIDS SEDs *include* dust emission, unlike the premise of the Meidt et al. (2014) calibration. With Figure 6.10 we also see that only only pixels with $R_{\mathrm{M31}} < 3$ kpc have $[3.6] - [4.5]$ colours reliably consistent with the Eskew et al. (2012) and Meidt et al. (2014) valid domains. At larger radii, $[3.6] - [4.5]$ become scattered. If this scatter is due to common data calibration issues, such as improper background subtraction or inclusion of unmasked foreground stars, the high S/N central pixels of the ANDROIDS dataset should be the most reliable samples of M31's intrinsic $[3.6] - [4.5]$ and $\log_{10} \mathcal{M}_*/L_{[3.6]}$. Focusing on pixels within the central 3 kpc, then, the MAGPHYS full SED fits are quite comparable to the universal $\mathcal{M}_*/L_{[3.6]} = 0.6$ advocated by Kettlety et al. (2017) for quiescent stellar populations. This constant $\log_{10} \mathcal{M}_*/L_{[3.6]}$ is useful for ANDROIDS data since the accuracy of the ANDROIDS per-pixel $[3.6] - [4.5]$ colours is apparently too poor given the high sensitivity of the Meidt et al. (2014) and Cluver et al. (2014) estimators.

The difficulty of applying the Meidt et al. (2014) and Cluver et al. (2014) relations to ANDROIDS pixels is further demonstrated in Figures 6.11 and 6.7, that show radial profiles of $\log_{10} \mathcal{M}_*/L_{[3.6]}$ and total stellar mass, respectively. Because of the colour sensitivity of the Meidt et al. (2014) and Cluver et al. (2014) relations, the estimated $\log_{10} \mathcal{M}_*/L_{[3.6]}$ varies widely within each isophotal annulus, as shown by the $1\,\sigma$ bands in Figure 6.11. This uncertainty propagates into total mass estimates that are 200% to 600% larger than the MAGPHYS estimate. Although the Eskew et al. (2012)

$\log_{10}\mathcal{M}_*/L_{[3.6]}$ is systematically higher than the MAGPHYS estimates (Figure 6.10), it yields a total mass ($8.8 \times 10^{10}\mathrm{M}_\odot$) that is more consistent with the MAGPHYS estimate because the Eskew et al. (2012) relation's colour sensitivity is lower, leading to less extreme per-pixel mass estimates.

An alternative to these sensitive mid-IR CMLRs is to instead assume a constant $\log_{10}\mathcal{M}_*/L_{[3.6]}$ irrespective of colour, and directly estimate stellar mass from the 3.6 μm flux. Meidt et al. (2014), in particular, identified a physical basis for asserting a constant $\log_{10}\mathcal{M}_*/L_{[3.6]}$ by observing that the effect of metallicity on the relationship between $[3.6] - [4.5]$ and $\log_{10}\mathcal{M}_*/L_{[3.6]}$ is orthogonal to the effect of age. Thus metal-rich, young populations naturally have similar $\log_{10}\mathcal{M}_*/L_i$ to old, metal-poor populations. The universal value advocated by Meidt et al. (2014) is $M_*/L_{[3.6]} = 0.6 \pm 0.1$. Kettlety et al. (2017) found that a value of $M_*/L_{[3.6]} = 0.6$ reproduced the stellar mass of the GAMA galaxy sample of quiescent galaxies with a scatter of 12% compared to full SED mass estimates (produced by Taylor et al. 2011).

For M31, we find that the total stellar mass estimated within $R = 20$ kpc by using a constant $M_*/L_{[3.6]} = 0.6$ is 7.1×10^{10} $\mathrm{M}_\odot$ (Table 6.3), which is 0.1 dex larger than the MAGPHYS full SED estimate. In order to fully reproduce the total MAGPHYS full SED-fit stellar mass estimate of 5.4×10^{10} $\mathrm{M}_\odot$, we fit $M_*/L_{[3.6]} = 0.46$. However, a constant $M_*/L_{[3.6]}$ does not reproduce the shape of the MAGPHYS-fit stellar mass profile. The MAGPHYS fit shows that $\log_{10}\mathcal{M}_*/L_{[3.6]}$ decreases across the disk (Figure 6.11). Within $R_{\mathrm{M31}} < 3$ kpc, the MAGPHYS fit implies $M_*/L_{[3.6]} = 0.53$. Outside that region, the MAGPHYS fit implies $M_*/L_{[3.6]} = 0.41$ (see Table 6.4).

Again, the 3.6 μm mass-to-light ratio fit to the ANDROIDS M31 dataset and full SED fits is 20% (-0.1 dex) smaller than the value advocated by Meidt et al. (2014).

This discrepancy may simply reflect the approach, and inherent uncertainties therein, by which dust emission is handled. We do not subtract emission from the 3.6 μm image, unlike Meidt et al. (2014) whose independent component analysis (ICA) separates light from old stellar populations from dust emission. The 3.6 μm dust emission in the ANDROIDS dataset would naturally drive $\log_{10} \mathcal{M}_*/L_{[3.6]}$ downwards. This does bring into question the efficacy of the mid-IR mass-to-light ratio approach for estimating the stellar mass of galaxies like M31 that do host ongoing star formation and possess hot dust. Whereas mid-IR mass-to-light ratios require that dust emission be subtracted to be unbiased, optical CMLRs (based on $g - i$, for example) do not. In fact, dust attenuation is highly degenerate with optical mass-to-light ratios, meaning that optical mass-to-light ratios are largely independent of the effects of dust. For this reason, optical CMLRs should be regarded as more effective than mid-IR mass-to-light ratios.

6.3.5 Comparison to Literature

Our best estimate for the stellar mass of the M31 bulge and disk, within 20 kpc, is 5.4×10^{10} M$_\odot$, based on MAGPHYS modelling of the full ANDROIDS SED. The random uncertainties of this estimate are negligible, of the order 10^8 M$_\odot$. Instead, systematic effects dominate the stellar mass uncertainty. Standard optical and mid-IR CMLRs permit M31's stellar mass to vary from 4.9×10^{10} M$_\odot$ to 8.9×10^{10} M$_\odot$. In this section we compare these estimates, based on ANDROIDS SEDs, to other direct estimates of M31's stellar mass mass. Table 6.5 summarizes total bulge and disk stellar masses found in the literature and our study.

The Geehan et al. (2006), Seigar et al. (2008), Chemin et al. (2009), and Corbelli

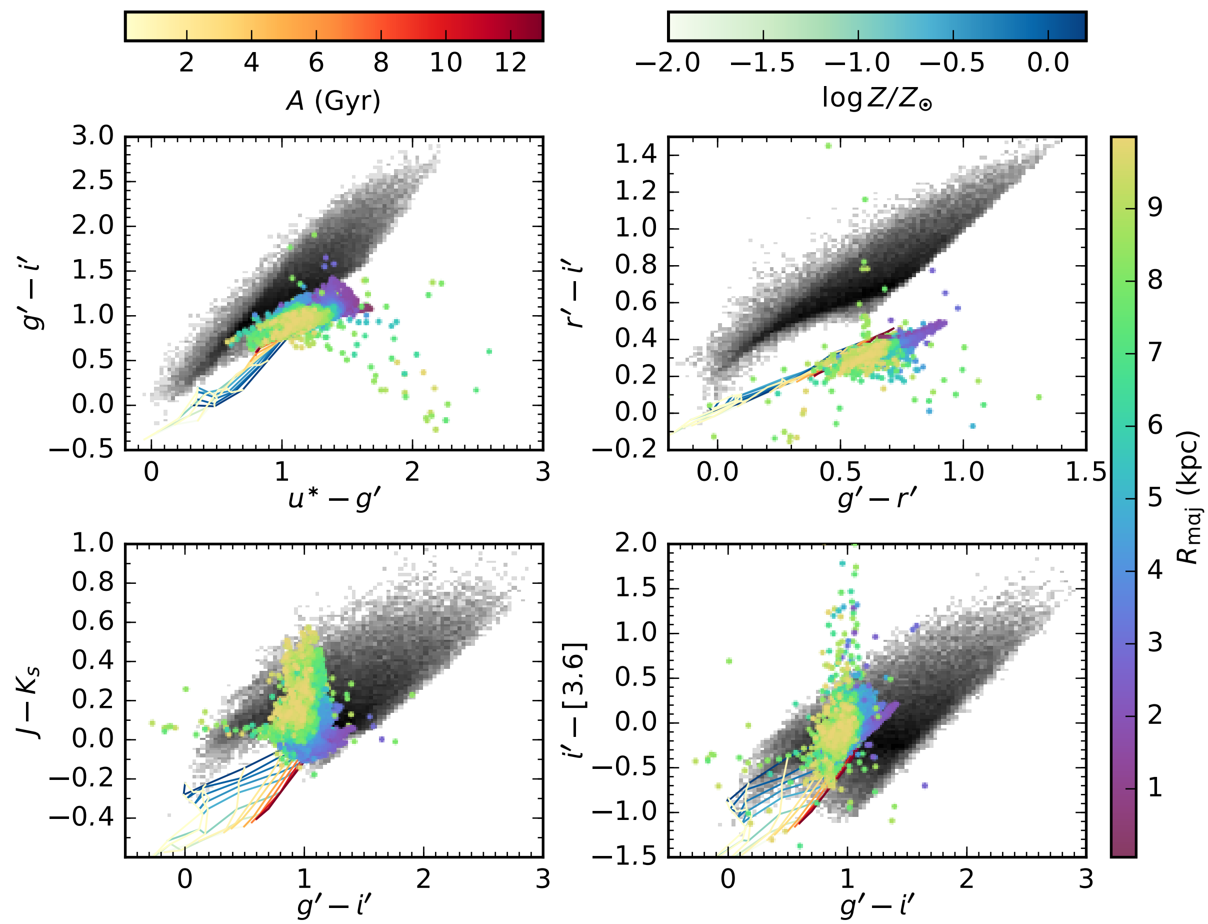


Figure 6.1: Comparison of the MAGPHYS stellar SED library colour space and observed ANDROIDS SEDs. In each colour-colour diagram panel, the grey scale locus shows the density of MAGPHYS library realizations. As described in § 6.2.1, these models use the (unpublished) Charlot & Bruzual 2007 (CB07) SED library and sample different dust attenuation. The dots are individual SEDs from the ANDROIDS dataset, coloured by M31 disk radius. To provide reference, we also plot colours of dust-free simple stellar populations (SSPs) for a grid of ages and metallicities. These SSPs are computed with FSPS (Conroy et al. 2009, 2010; Conroy & Gunn 2010), which is different from the CB07 population synthesis library, and thus is potentially shifted in colour from MAGPHYS. Dust reddening, though, is a more significant driver of colour differences between the reference SSP grids and the observed SEDs. Overall, M31 SEDs are consistent with old, metal-rich, and possibly dust-attenuated stellar populations.

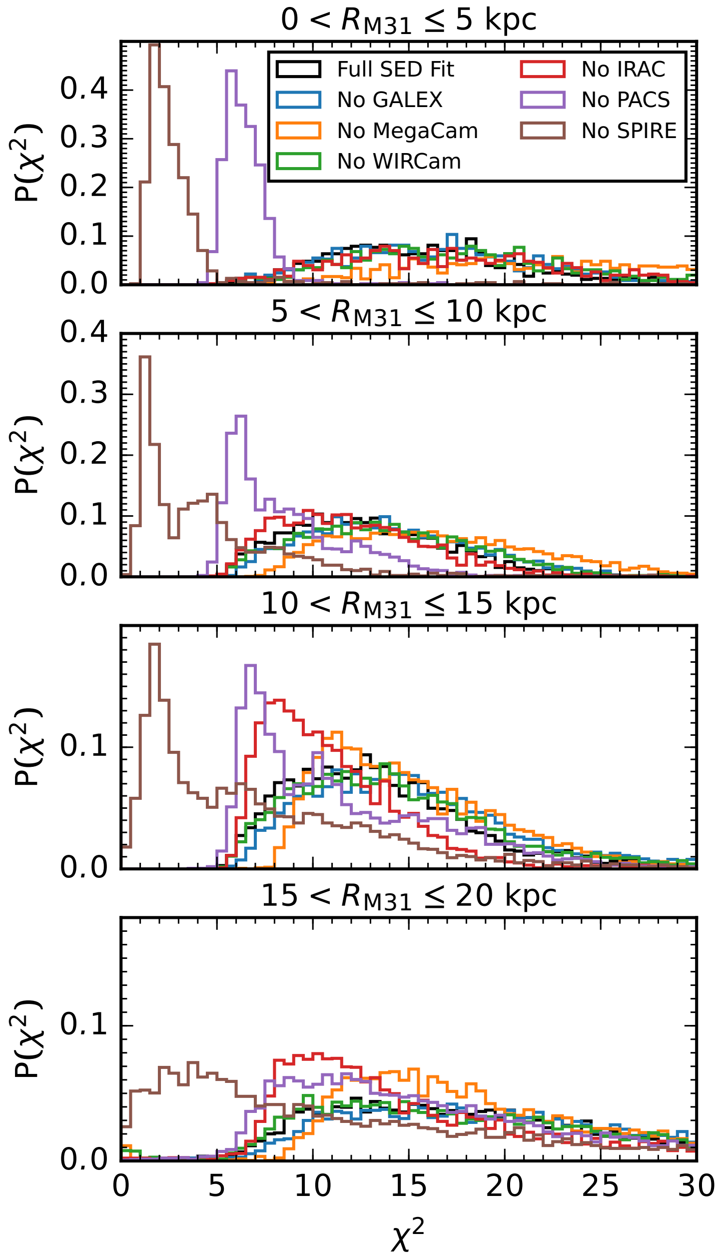


Figure 6.2: The distributions of MAGPHYS model fitting χ^2 for different MAGPHYS fitting campaigns and radial regions. The χ^2 statistic (Eq. 6.1) describes the overall goodness-of-fit of MAGPHYS model of each pixel. Smaller χ^2 values reflect better fits in terms of reduced cumulative residuals across fitted SED fluxes. In each panel, pixels are binned by M31 radius since disk radius is a proxy for pixel S/N.

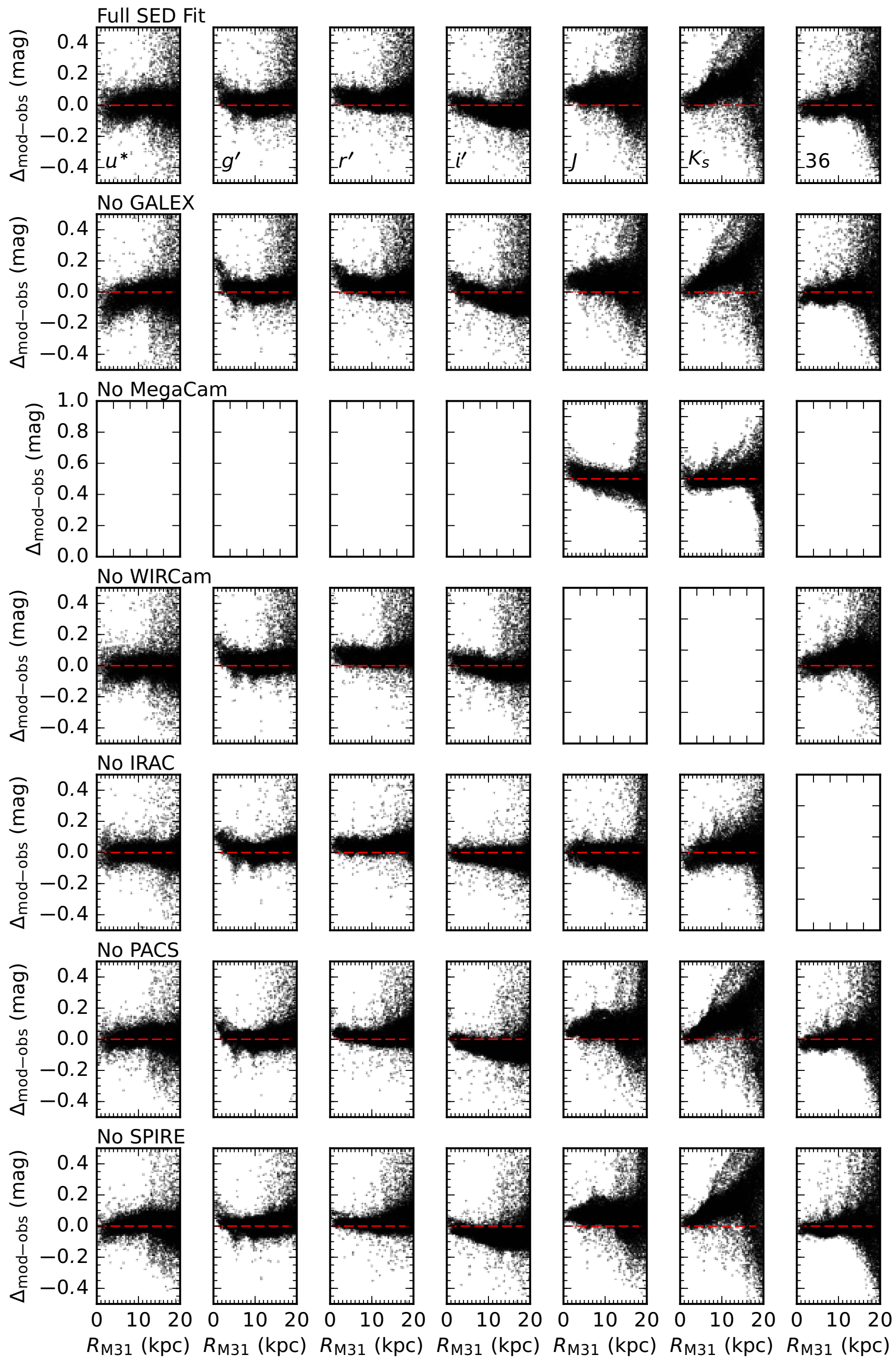


Figure 6.3: Optical and Near-IR MAGPHYS SED fitting residuals. Each row corresponds to a different MAGPHYS fitting campaign. The top row is the full SED fit, while lower rows are fitting campaigns where a specified instrument was omitted from the SED. SED modelling residuals are shown for each of the u^*, g', r', i', J, K_s, and IRAC 3.6 μm bands for each mosaic pixel, as functions of M31 disk radius.

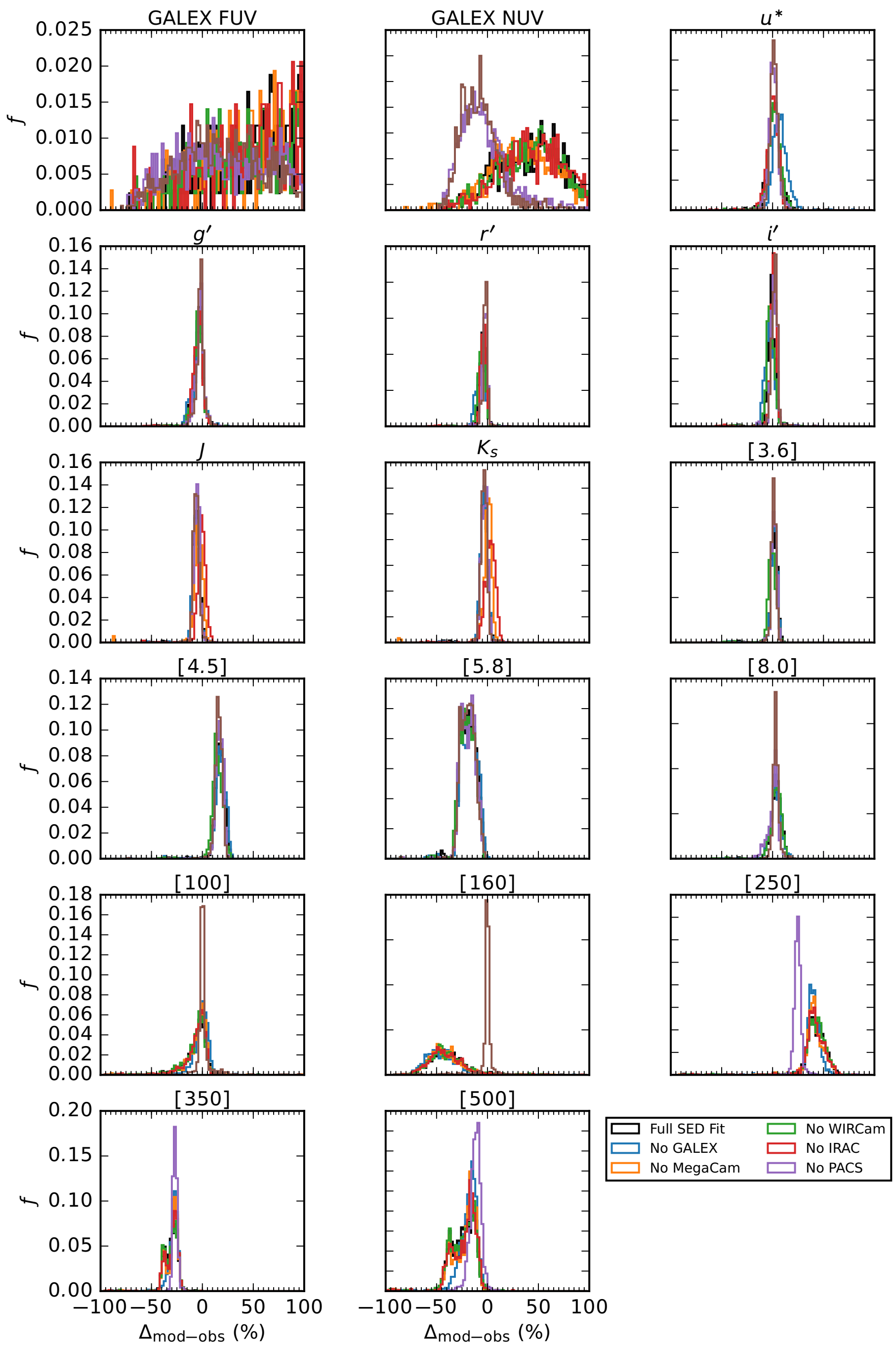


Figure 6.4: Distributions of MAGPHYS fitting residuals for all bands in the ANDROIDS SED, and for each fitting campaign. An alternative view to Figure 6.3, this plot shows fitting residuals only for the pixels in the inner 5 kpc of M31, and is in units of percent uncertainty of the flux: $\Delta_{\rm mod-obs} = (F_{\rm model}/F_{\rm obs} - 1) \times 100$. This view, however, allows us to compare fitting residuals across all 17 bands in the ANDROIDS SED.

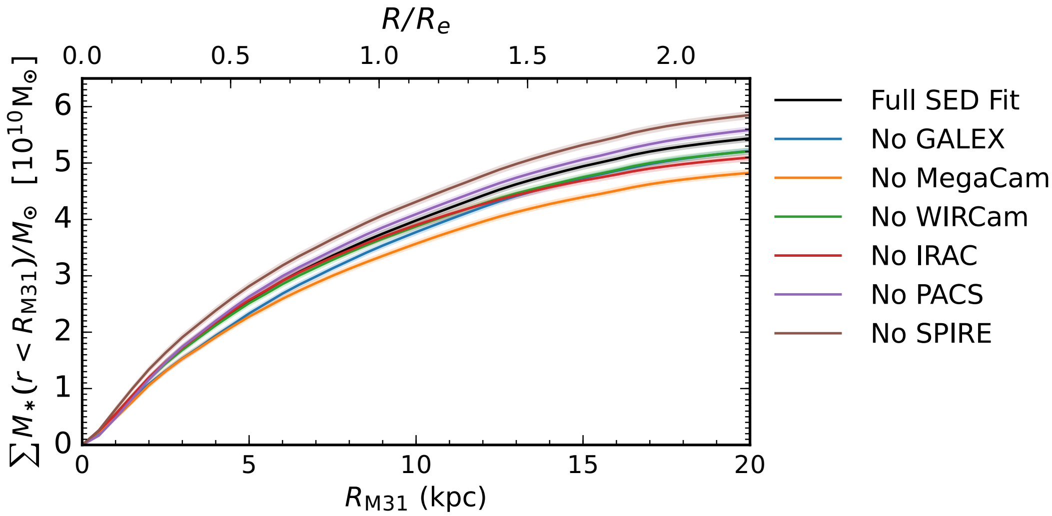


Figure 6.5: Radial stellar mass profiles estimated by MAGPHYS with different SED sets (jackknife analysis): the full SED fit, and fits where a given instrument is omitted ("No GALEX," *etc.*). Radii correspond to the elliptical isophotes (§ 5.5). M32's stellar mass *is included*, partially in the 10 kpc mass and fully included in the 20 kpc mass.

Table 6.4: Mid-IR mass-to-light ratios fit by MAGPHYS Full SED modelling. The R_{M31} column indicates the radial domain over which the mass-to-light ratio is computed. For comparison, Meidt et al. (2014) and Kettlety et al. (2017) advocate $\mathcal{M}_*/L_{[3.6]} = 0.6$ for quiescent galaxies.

R_{M31} (kpc)	$\langle \mathcal{M}_*/L_{[3.6]} \rangle$
$(0, 3)$	0.53
$(3, 20)$	0.41
$(0, 20)$	0.46

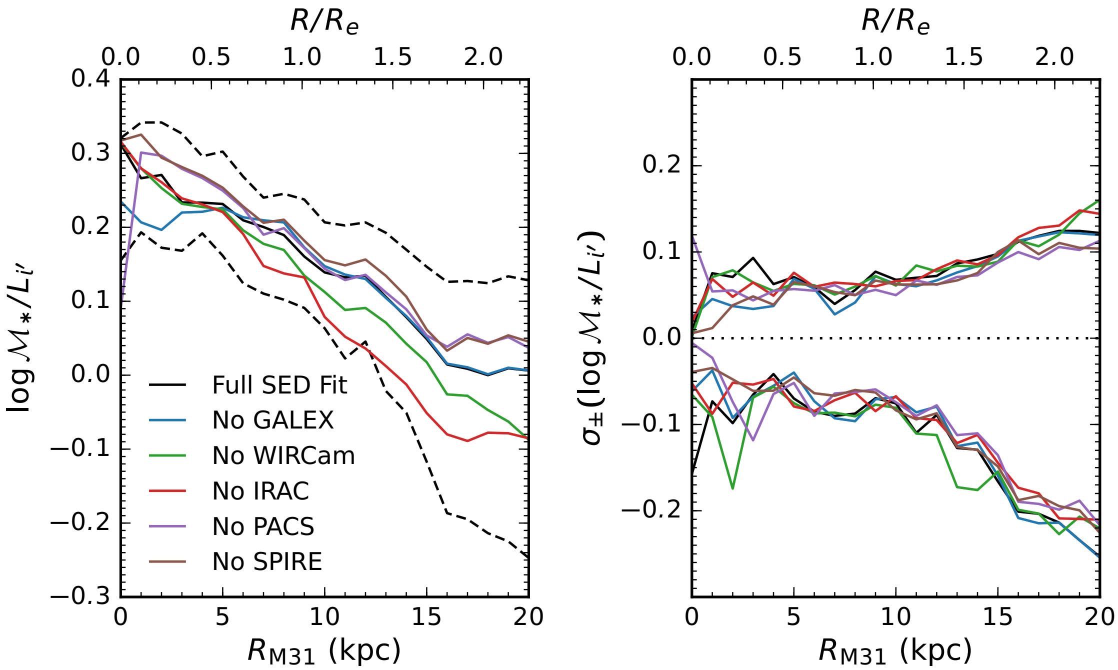


Figure 6.6: Right: radial profile of $\log_{10} \mathcal{M}_*/L_i$ for MAGPHYS fitting campaigns. The 16% and 85% (1σ) percentiles of the radial annulus sample distribution of the full SED fit are shown in dashed black lines. This confidence interval provides scale for the left panel: the $84\% - 50\%$ and $50\% - 16\%$ ($\pm 1\sigma$) sample standard deviations as a function of M31 disk radius.

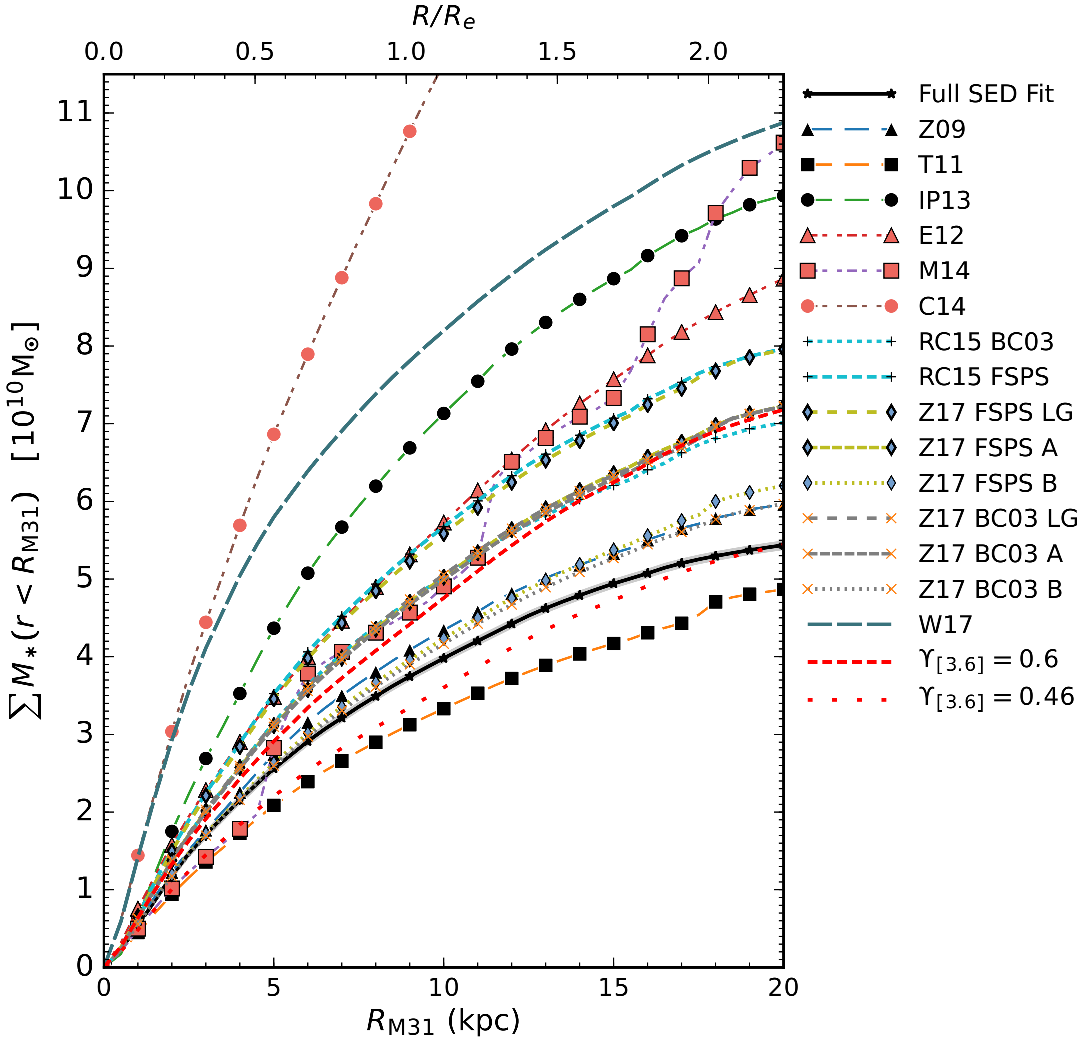


Figure 6.7: Radial stellar mass profiles estimated by MAGPHYS, colour M/L indicators, and other studies. The Full SED fit is made by MAGPHYS with the full ANDROIDS dataset. Profiles made by optical $(g-i)$ colour-mass-to-light ratio (CMLR) estimators are: Z09 (Zibetti et al. 2009), T11 (Taylor et al. 2011), IP13 (Into & Portinari 2013), Roediger & Courteau (2015) estimators with Bruzual & Charlot (2003) and FSPS population synthesis models, and Zhang et al. (2017) estimators (see § 6.3.3). Profiles made by mid-IR CMLRs are: E12 (Eskew et al. 2012), M14 Meidt et al. (2014), and C14 (Cluver et al. 2014) (see § 6.3.4). $\Upsilon_{[3.6]} = 0.6$ and 0.46 are based on constant IRAC 3.6 μm mass-to-light ratios. A mass estimate based on W17 (Williams et al. 2017) resolved stellar mass estimates is also shown (see § 6.3.5). Radii correspond to the elliptical isophotes (§ 5.5). M32's stellar mass *is included*, partially in the 10 kpc mass and fully included in the 20 kpc mass.

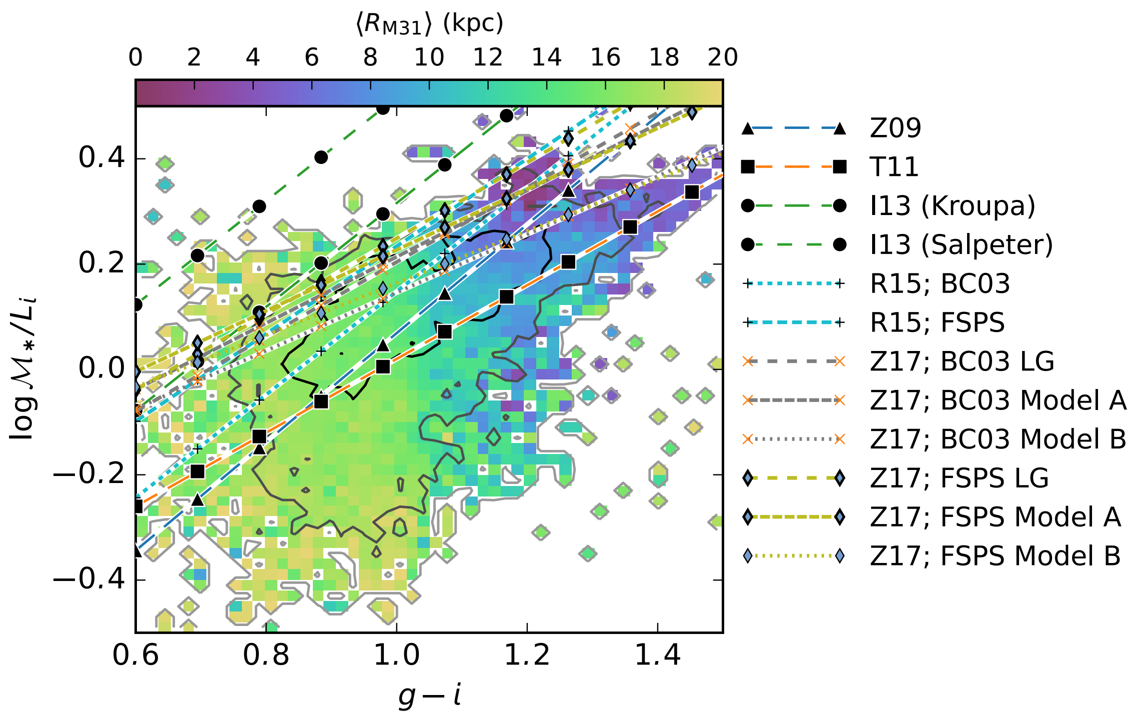


Figure 6.8: Comparison of optical colour-$\mathcal{M}_*/L_i$ ratio estimators, described in § 6.3.3, to the distribution of pixel $\mathcal{M}_*/L_i$ estimated by MAGPHYS. Pixel are annotated by mean radius in the $g - i - \log_{10} \mathcal{M}_*/L_i$ plane. Black and grey contours show the 1σ, 2σ, and 3σ density distribution of pixels in this plane. See Figure 6.7 for a description of the models.

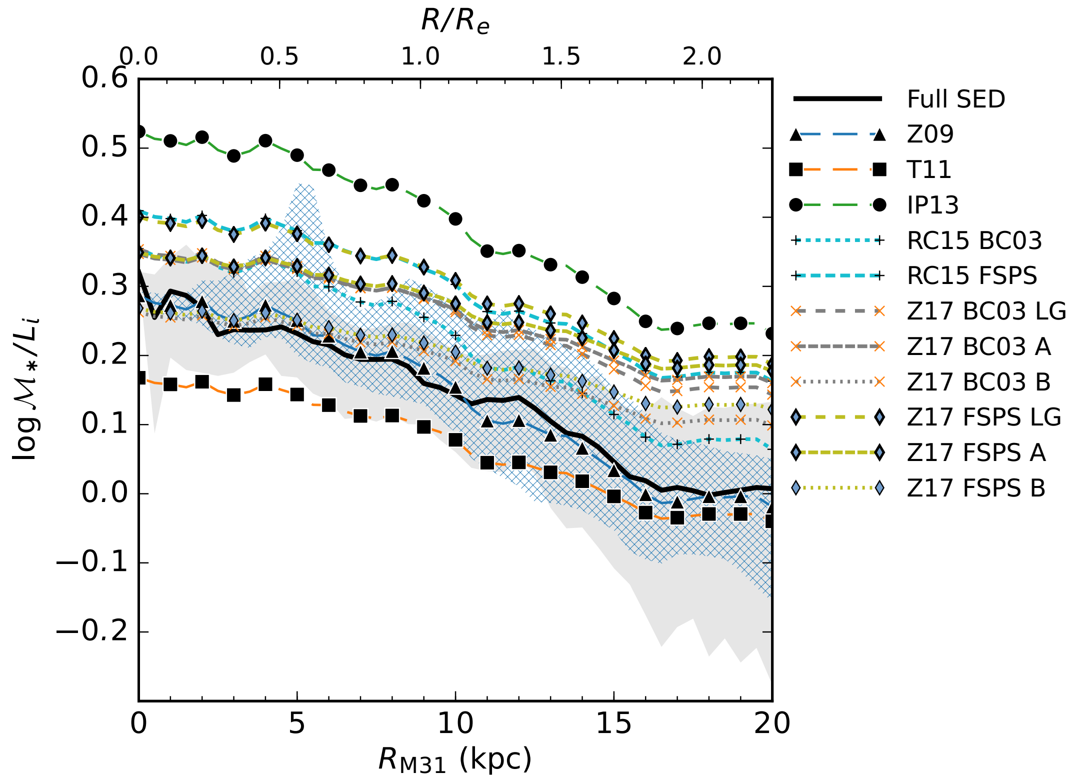


Figure 6.9: Comparison of $\log_{10}\mathcal{M}_*/L_i$ radial profiles estimated by MAGPHYS and optical colour mass-to-light ratio (CMLR) relations applied to the ANDROIDS dataset. The mean $\log_{10}\mathcal{M}_*/L_i$ estimated MAGPHYS Full SED fit is also shown as a thick black line along with the $\pm1\sigma$ sample distribution within isophotal annuli shown as the grey area. Similarly, the same annular $\pm1\sigma$ sample distribution of Z09 (Zibetti et al. 2009) is shown as the blue diagonal hatched area. While fluctuations in the MAGPHYS estimates increase with radius, the uncertainties based on CMLRs do not.

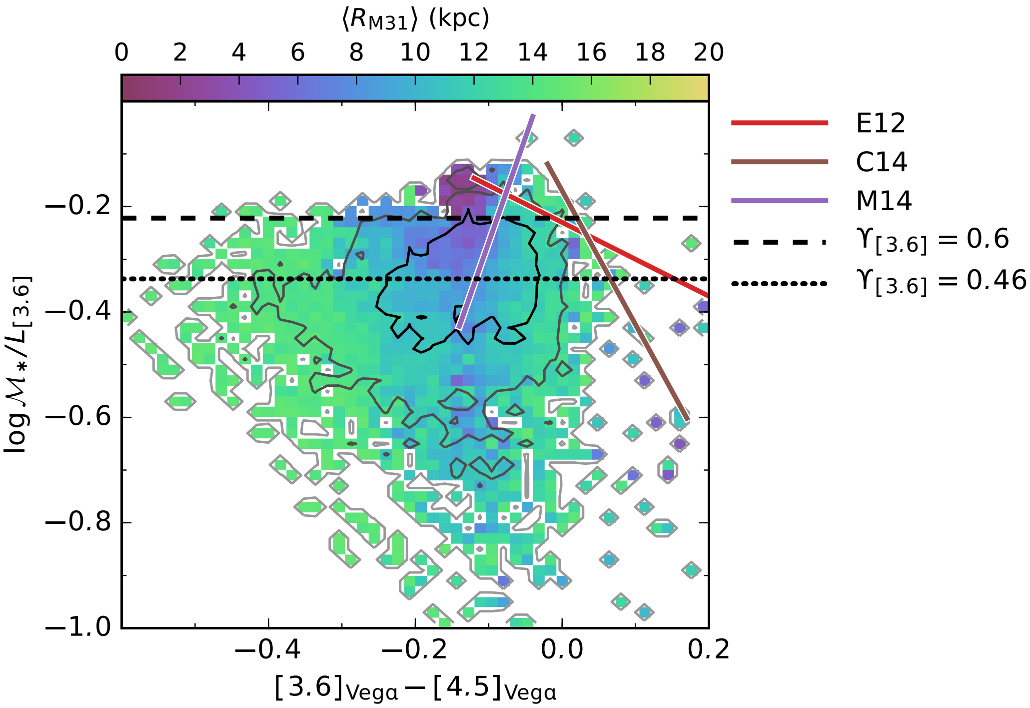


Figure 6.10: Relationship between the $[3.6]_{\rm Vega} - [4.5]_{\rm Vega}$ colour and $\log_{10} \mathcal{M}_*/L_{[3.6]}$ for MAGPHYS full SED models and CMLRs. The background map is the mean radius of M31 pixels in the $[3.6]_{\rm Vega} - [4.5]_{\rm Vega}$ colour and $\log_{10} \mathcal{M}_*/L_{3.6}$ plane. Black and grey contours show the 1σ, 2σ, and 3σ density distribution of pixels in this plane. The constant $\mathcal{M}_*/L_{[3.6]} = 0.6$ estimator is advocated by Meidt et al. (2014) and Kettlety et al. (2017). The constant $\mathcal{M}_*/L_{[3.6]} = 0.46$ value matches the MAGPHYS mass estimate of M31 with the full ANDROIDS SED.

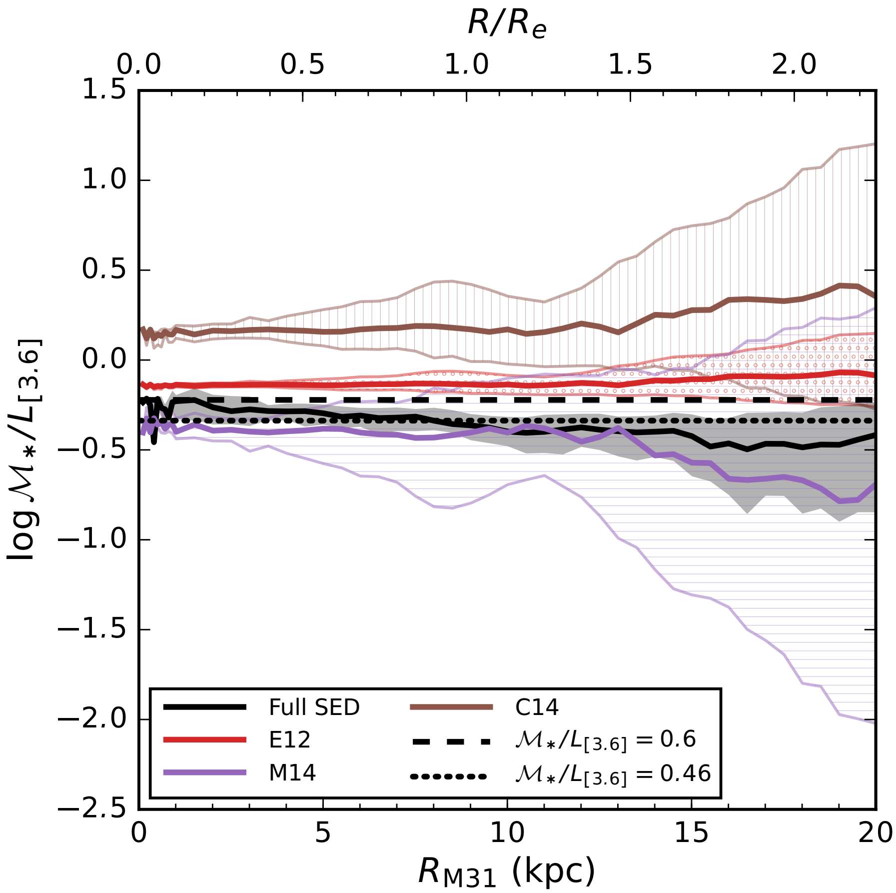


Figure 6.11: Radial profile of mid-IR mass-to-light ratios estimated by CMLRs and MAGPHYS full SED fits (see also Figure 6.10). Solid lines are the same median per-pixel $\log_{10} \mathcal{M}_*/L_{[3.6]}$ at each radius while shaded regions describe the 1σ sample standard deviation of pixels in each isophotal band (§ 5.5).

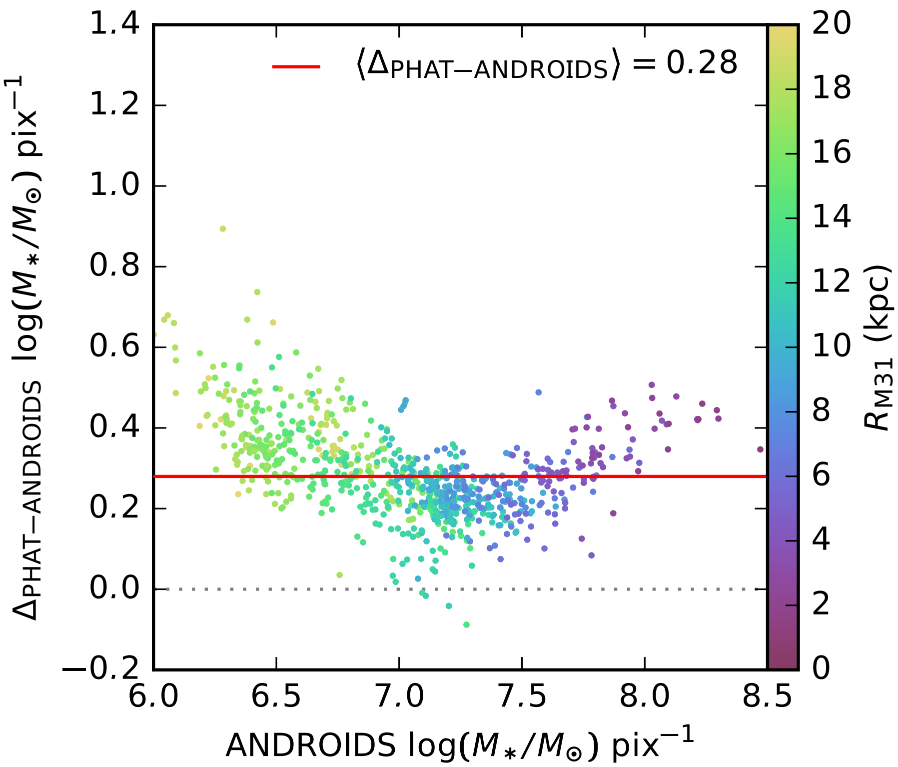


Figure 6.12: Differences of stellar mass estimates between PHAT (Williams et al. 2017) and this study (full SED MAGPHYS fit) as a function of stellar mass. Each pixel in this diagram is $83'' \times 83''$ (0.3×1.4 kpc), the scale of the Williams et al. (2017) map. The solid red line corresponds to the mean difference.

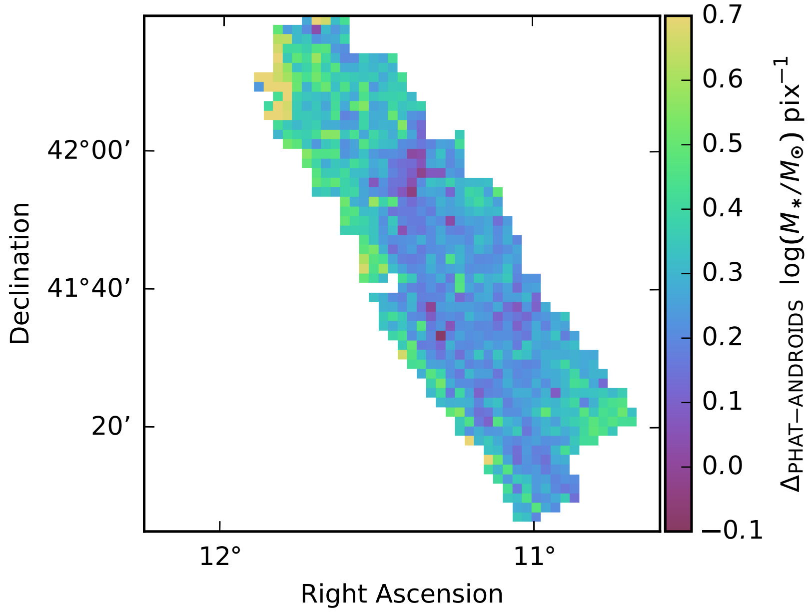


Figure 6.13: Map of differences of stellar mass, $\log_{10}(M_*/M_\odot)$ estimated by PHAT (Williams et al. 2017) and this study (full SED MAGPHYS fitting campaign). Each pixel in this diagram is $83'' \times 83''$ (0.3×1.4 kpc), the scale of the Williams et al. (2017) map. ANDROIDS pixels were binned to correspond to this scale.

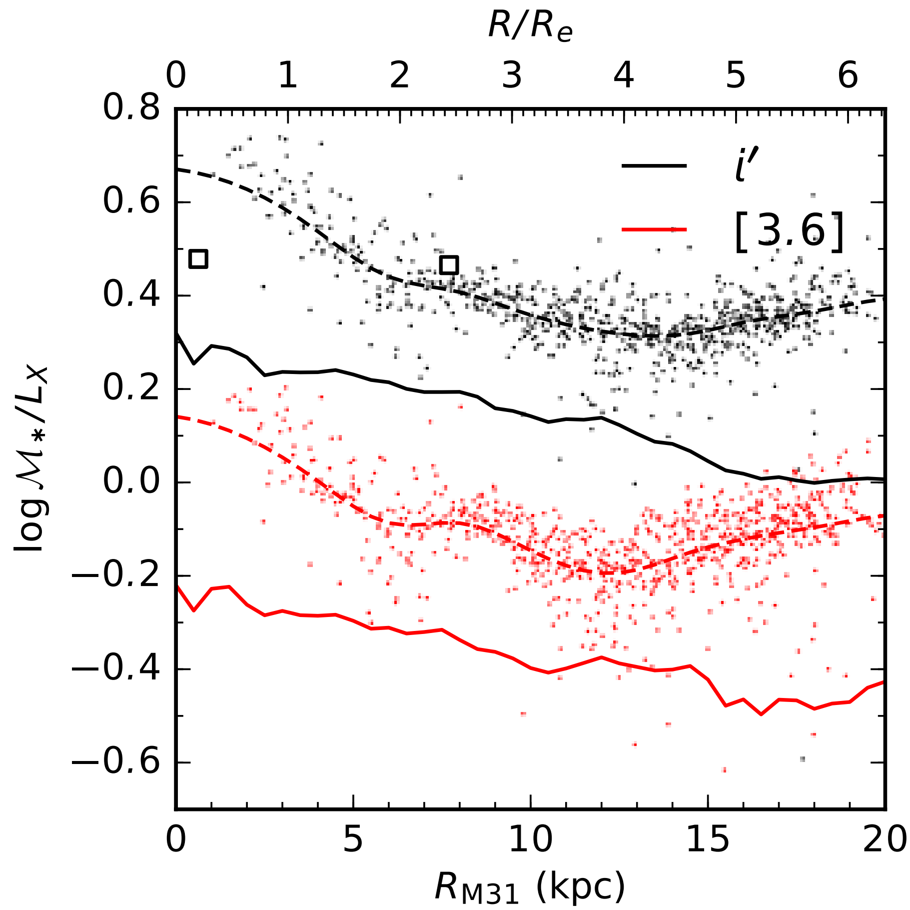


Figure 6.14: Radial i' (black) and [3.6]-band (red) mass-to-light ratio profiles fit from Williams et al. (2017) resolved stellar mass estimates and ANDROIDS photometry compared to ANDROIDS/MAGPHYS mass-to-light profiles and Tamm et al. (2012) $\log_{10} \mathcal{M}_*/L_i$. Individual points are stellar mass-to-light ratios fit in individual $83'' \times 83''$ (0.3×1.4 kpc) PHAT regions. Dashed lines are non-parametric stellar mass-to-light ratio profiles fit to the estimates in each region. For comparison, we show the MAGPHYS-fit stellar mass-to-light ratio profiles as solid lines. White squares are the mean $\log_{10} \mathcal{M}_*/L_i$ fit by Tamm et al. (2012) to the stellar bulge and disk.

Table 6.5: M31 bulge and disk stellar mass estimates. This table is based on an original literature compilation by Tamm et al. (2012). All ANDROIDS-based measurements (including Williams et al. 2017) encompass all stellar mass within an $R_{\rm M31} = 20$ kpc isophotal ellipse. Other studies integrate integrate stellar mass across fitted analytic stellar bulge and disk components.

Source	Notes	$\mathcal{M}_*$ (10^{10} $\mathrm{M}_\odot$)
Geehan et al. (2006)	Best-fit model	11.7
Geehan et al. (2006)	Maximum-disk model	17.0
Seigar et al. (2008)	Without adiabatic contraction	9.3
Seigar et al. (2008)	With adiabatic contraction	10.8
Chemin et al. (2009)	Hybrid model	9.42
Corbelli et al. (2010)	NFW with constant $\mathcal{M}_*/L$	12.6
Tamm et al. (2012)	Blanton & Roweis (2007) SPS	10.1
Tamm et al. (2012)	Maximum-disk	15.2
Williams et al. (2017)	With ANDROIDS i'	11.2
Williams et al. (2017)	With ANDROIDS [3.6]	11.5
ANDROIDS	MAGPHYS full SED	5.4
ANDROIDS	Zibetti et al. (2009) CMLR	6.0
ANDROIDS	Taylor et al. (2011) CMLR	4.9
ANDROIDS	Into & Portinari (2013) CMLR	8.9
ANDROIDS	Roediger & Courteau (2015) BC03 CMLR	7.0
ANDROIDS	Roediger & Courteau (2015) FSPS CMLR	8.0
ANDROIDS	Zhang et al. (2017) FSPS LG	8.0
ANDROIDS	Zhang et al. (2017) FSPS Model A	7.2
ANDROIDS	Zhang et al. (2017) FSPS Model B	6.2
ANDROIDS	Zhang et al. (2017) BC03 LG	7.2
ANDROIDS	Zhang et al. (2017) BC03 Model A	7.2
ANDROIDS	Zhang et al. (2017) BC03 Model B	6.0
ANDROIDS	$\mathcal{M}_*/L_{[3.6]} = 0.6$	7.2

et al. (2010) models are all primarily driven by dynamical modelling using rotation curves (Widrow et al. 2003; Chemin et al. 2009; Corbelli et al. 2010) and extended kinematic tracers such as planetary nebulae, globular clusters, and satellites. Stellar mass-to-light ratios are either fitting parameters (Geehan et al. 2006; Seigar et al. 2008) or based on CMLRs (Bell & de Jong 2001; Bell et al. 2003) with typical M31 $B - R$ colours observed by Walterbos & Kennicutt (1988), as is the case for Chemin et al. (2009) and Corbelli et al. (2010). These studies all find a total stellar mass (including bulge and disk) in the range of 9.3–12.6 $\times 10^{10} \mathrm{M}_{\odot}$—higher than any total stellar mass estimated from ANDROIDS photometry (Table 6.5).

Tamm et al. (2012) used SDSS *ugrizy* and Spitzer 3.6 μm imaging to find a total stellar mass of $(10-15)\times 10^{10}$ $\mathrm{M}_{\odot}$. Their lower-bound estimate is based on modelling their *ugrizy* pixel SEDs as a linear combination of three synthetic model spectra computed by Blanton & Roweis (2007) that are themselves composed of nebular line emission models and Bruzual & Charlot (2003) models for ranges of stellar ages and metallicities. Tamm et al. (2012)'s upper-bound bulge and disk stellar mass estimate is based on a maximum-disk model, where the full rotation curve is explained by baryonic, as opposed to dark, matter. This maximal disk stellar mass is similar to the Geehan et al. (2006) maximal disk estimate of 17×10^{10} $\mathrm{M}_{\odot}$.

We have reasons to question the Tamm et al. (2012) sub-maximal stellar mass estimate, which is significantly at odds with our ANDROIDS result. Tamm et al. (2012) modelled the SEDs of individual pixels as the best-fitting linear combination of three synthetic model spectra (see their Table 1). These composite spectra have ages of 0.7 Gy, 0.4–1 Gyr, and 7–12 Gyr, with monometallicities of [Fe/H]=0.40, 0.05, and 0.03. The effective $\log_{10} \mathcal{M}_*/L_i$ of these models are -0.14, -0.25, and 0.49, respectively.

Crucially, they find that 98% of the stellar mass is associated to the old, high mass-to-light ratio composite spectrum. As shown in Figure 6.14, this high $\log_{10}\mathcal{M}_*/L_i$ is seen in both the bulge and disk. Such a high $\log_{10}\mathcal{M}_*/L_i$ is incompatible with ANDROIDS photometry for two reasons. First, the ANDROIDS $g-i$ colours would have to be be redder by at least 0.3 mag to be compatible with $\log_{10}\mathcal{M}_*/L_i = 0.5$ (Figure 6.8). Second, we find a clear radial $g-i$ gradient (Figure 5.11) that results in a linear $\log_{10}\mathcal{M}_*/L_i$ gradient that drops from 0.3 at the centre to 0.0 at 20 kpc. On these grounds, the Tamm et al. (2012) stellar mass estimate is untenable.

Williams et al. (2017) estimated the stellar mass of M31 disks using models of resolved stellar populations measured by the Panchromatic Hubble Andromeda Treasury (PHAT) survey. This method is distinct from the broadband SED fitting that is the basis for all the other stellar mass estimates already described. We discuss resolved stellar population fitting in detail in Chapter 7, but the method works roughly as follows. The authors build photometric catalogs of individual stars in multiple bandpasses, and also use artificial star testing to model both photometric uncertainties and completeness — the rate that a star of a given brightness and colour can be successfully observed in a star field. Then the authors model that photometric catalog (using in colour- magnitude diagram projections) by fitting a linear combination of isochrones of different ages, metallicities, and affected by different dust screens. That optimized set of isochrones is effectively a physical model of the stellar population.

In isochrone fitting, stellar mass estimates are determined from the sum of masses assigned to each isochrone. A peculiarity of this process, though, is that isochrones are normalized by their birth stellar mass. As a stellar population evolves, massive stars die and return mass to the interstellar medium. MAGPHYS and the CMLR methods are calibrated to the instantaneous stellar mass of living stars and remnants (like white dwarfs), while isochrone fitting produces the total mass of stars ever formed. Williams et al. (2017), citing an M31 stellar population that is older than 8 Gyr, find that the current stellar mass is a factor 0.6 less than the total mass of stars formed, given a Kroupa (2001) IMF. This being a significant factor, one wishes that the current stellar mass would be computed integrally with the isochrone fitting, rather than being estimated afterwards. Nonetheless, we adopt the 0.6 conversion factor estimated by Williams et al. (2017) when comparing their stellar mass estimates to those of MAGPHYS and other studies.

With Figures 6.12 and 6.13 we compare the stellar mass estimates between Williams et al. (2017) and the full SED ANDROIDS/ MAGPHYS fit in each $83'' \times 83''$ PHAT region. On average, the Williams et al. (2017) stellar mass estimates (including the correction for dead stars) are $0.2 - -0.3 \log(M_*/M_\odot)$ heavier than the MAGPHYS estimates. In the 10 kpc star forming regions, in particular, the differences are minimal (Figure 6.13).

Another useful comparison of the resolved mass estimates by Williams et al. (2017) and the broadband SED fits with MAGPHYS is through their stellar mass-to-light ratios. We estimated mass-to-light ratios by matching the Williams et al. (2017) stellar mass estimates to photometry integrated in ANDROIDS mosaics over the same apertures as the PHAT regions. In Figure 6.14 we show these radial mass-to-light ratio profiles in comparison to those estimated from the full SED MAGPHYS fits. The resolved mass-to-light profile has generally the same gradient as that estimated from SEDs, though with a 0.2 dex higher normalization. As we shall see in § 6.5, such a normalization difference may by driven by differences in estimated stellar population ages.

At both the high and low mass regimes (bulge and outer M31 disk) the differences are inflated. This behaviour in the inner disk and bulge is consistent with the Williams et al. (2017) artificial star tests generally overcompensating for incompleteness in the resolved stellar catalog in crowded environments. On the other hand, the outer disk is an ideal environment for resolved stellar photometry so the Williams et al. (2017) stellar mass estimates should be deemed reliable.

The Williams et al. (2017) estimate for the current stellar mass in the M31 disk assumes azimuthal symmetry. Therefore, the mass measured directly in the observed PHAT survey footprint can be multiplied by a factor of 3.0 to yield a full disk mass, which they find to be $(9 \pm 2) \times 10^{10} \mathrm{M}_{\odot}$ of stars and remnants within $R = 20$ kpc. Furthermore, this mass estimate excludes M31's inner bulge, where stellar crowding is too high to reliably model a stellar population from resolved stellar photometry. To make a total stellar mass estimate that can be more directly compared to our full SED ANDROIDS models, we combined the Williams et al. (2017) radial $\log_{10} \mathcal{M}_*/L_i$ profile (Figure 6.14) with our ANDROIDS i'-band mosaic. Note that this estimate automatically includes the stellar mass of M31's bulge by extrapolating the Williams et al. (2017) mass estimates inwards. Altogether, we find that the total stellar mass of M31 bulge and stellar disk estimated through the resolved stellar population models of Williams et al. (2017) is 11.2×10^{10} $\mathrm{M}_{\odot}$.

The comparison with literature analogs (Table 6.5) shows this stellar mass estimate from resolved stellar populations by Williams et al. (2017) to be higher than any mass estimate that can be built from optical CMLRs, mid-IR CMLRs, or the full SED MAGPHYS fit. It is more consistent with Geehan et al. (2006), Seigar et al.

(2008), Chemin et al. (2009), Corbelli et al. (2010), and Tamm et al. (2012) models. The Williams et al. (2017) total bulge and disk stellar mass estimate is still 70% of the mass implied by the maximum-disk (minimal dark matter) fits Geehan et al. (2006) and Tamm et al. (2012).

6.4 Metallicity Estimation

Metallicity ($\log Z/Z_{\odot}$) is the mass ratio of metals, elements heavier than helium, to hydrogen in stellar atmospheres. Along with age and dust attenuation, metallicity is a key parameter in determining a stellar population's SED. Metallicity broadly describes the evolution of a galaxy: stars formed from relatively pristine gas infall have low metallicities, while stars formed after multiple star formation cycles have higher metallicities.

Note that MAGPHYS makes a key simplification that a single metallicity is associated with an SED. In reality, a galaxy's SED is a linear combination of SEDs from stars of multiple ages and metallicities. This is a typical limitation of broadband SED modelling, where stellar age, metallicity, and dust attenuation effects are degenerate in colour space. At most, broadband stellar population synthesis models may introduce a parameterized chemical evolution model tied to the star formation history (for example, Portinari et al. 2004). Studies of resolved stellar populations (see Chapter 7), on the other hand, are often able to measure a chemical evolution history by fitting distinct features in colour-magnitude diagram space (for example, Brown et al. 2006). For this section, recall that the monometallic estimates being quoted effectively represent a mass-weighted mean metallicity. Finally, note that MAGPHYS parameterizes metallicity as $\log Z/Z_{\odot}$ so that "0" is solar metallicity,

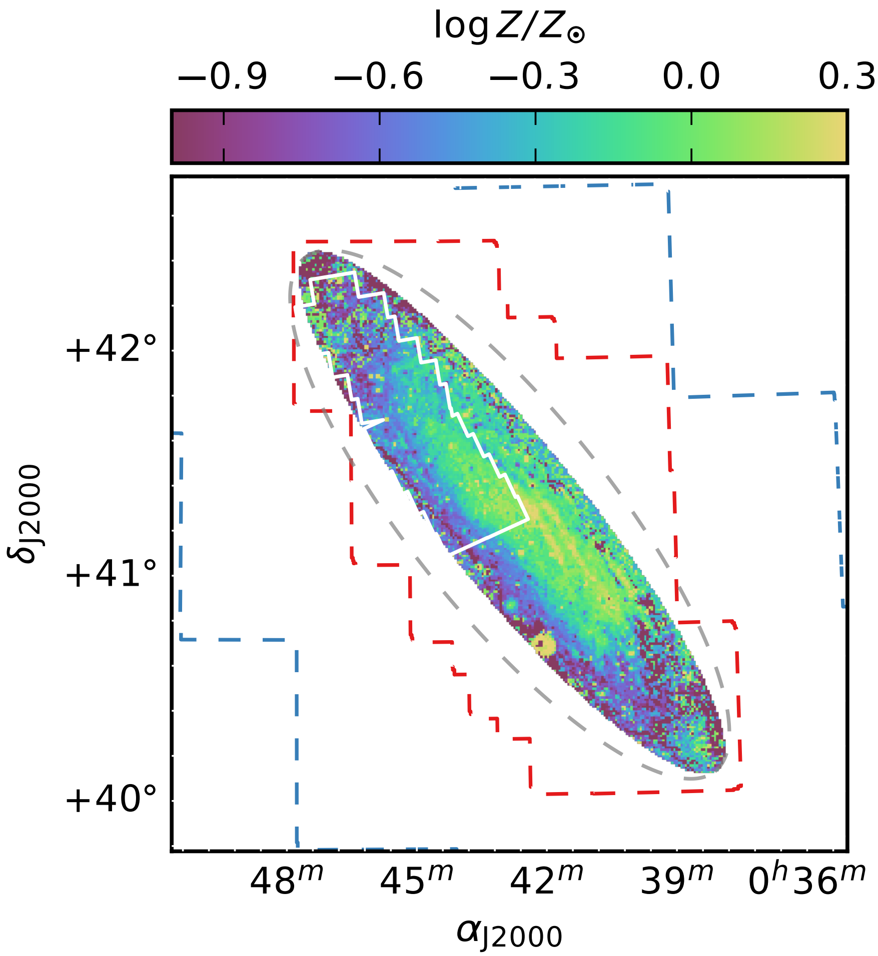


Figure 6.15: Map of ANDROIDS/MAGPHYS full SED-fit metallicity estimates within $R_{\mathrm{M31}} = 20$ kpc. For reference, the white polygon is the PHAT footprint (Dalcanton et al. 2012). The red dashed polygon is the footprint of the ANDROIDS WIRCam survey (Chapter 2) while the dashed blue polygon shows the ANDROIDS MegaCam survey footprint, which extends beyond the axes.

following standard practice.

6.4.1 MAGPHYS Metallicity Estimates

Figure 6.15 shows metallicity across M31's bulge and disk, modelled by MAGPHYS using the full ANDROIDS SED. The key feature of this map is M31's smooth radial

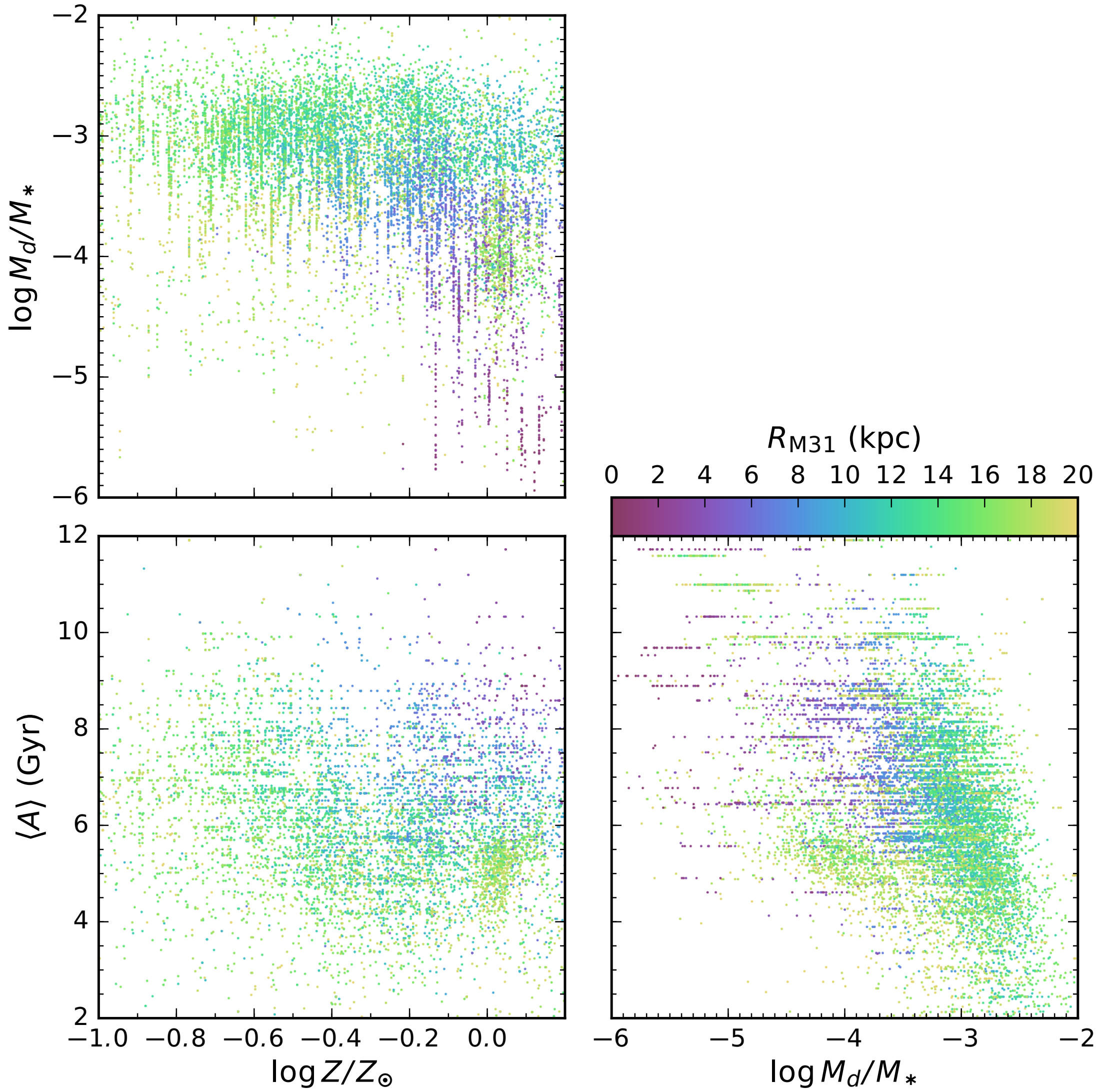


Figure 6.16: Relationship between estimated dust mass, stellar metallicity, and mean stellar age in the ANDROIDS/MAGPHYS full SED fit; each dot corresponds to the stellar population of a $36'' \times 36''$ pixel. $\log M_d/M_*$ is the ratio of dust to stellar mass, $\log Z/Z_\odot$ is the stellar metallicity, and $\langle A \rangle$ is the mass-weighted mean stellar age. Dots are coloured by their deprojected M31 disk radius, $R_{\rm M31}$.

metallicity gradient. M31's bulge and inner disk are metal-rich, reminiscent of highly evolved early-type galaxy stellar populations, while the outer disk is increasingly metal-poor with radius. Overall, Figure 6.15 is qualitatively consistent with inside-out disk formation.

A second-order effect in Figure 6.15 is near-far azimuthal asymmetry, which may be attributed to dust attenuation. The Northwest minor axis of M31 (top right corner of Figure 6.15) is closer to us than the disk along the Southwest minor axis (lower left corner). As explained by Binney & Merrifield (1998), the overall dust attenuation is higher through the near side than the far side because more stellar light originates behind the thin dust disk on the near side of a titled galaxy disk. This is due to the exponentially declining radial surface brightness profiles of disk galaxies. On the near side of an inclined disk galaxy, the starlight in front of the dust disk is at a larger deprojected disk radius than the starlight behind the dust disk so that the higher surface brightness light in that line of sight is behind the dust disk. On the far side of the inline disk the opposite is true.

That the metallicity is maximized along the 5 kpc and 10 kpc dust arms on the minor axis is further evidence of this effect. This connection between dust attenuation and metallicity is a manifestation of the age-metallicity-dust degeneracy, where age, metallicity, and dust reddening all introduce similar colour shifts in a galaxy's optical SED (Bell & de Jong 2001). To further explore the possibility of an age-metallicity-dust degeneracy in the MAGPHYS estimates, we show in Figure 6.16 the distribution of pixels in the estimated age-metallicity-dust parameter space. In that plot, the ratio of dust mass to stellar mass ($\log M_d/M_*$) is a proxy for overall dust attenuation (see § 6.6). Figure 6.16 illustrates that higher metallicities are and higher ages alike are found in pixels with lower dust content. These results, of course, cannot be separated from true stellar population trends. Indeed, deprojected radius (R_{M31}) is also corrected with age, dust, and metallicity. We do, however, see in § 6.5 and Figure 6.19 a near-far effect on stellar age estimates: where younger ages are seen on the near side of the disk. Thus we find evidence of a parameter degeneracy between age and metallicity estimates due to dust attenuation. This effect should result in a broader dispersion of metallicity estimates along isophotes, as seen in radial plots (Figure 6.17 and 6.18).

6.4.2 Bandpass Dependence

In Figure 6.17 we illustrate how different instruments affect stellar metallicity estimation. Besides the full SED fit, other estimates are based on modelling campaigns where individual instruments were omitted from SED fitting (§ 6.2.5). All the fitting campaigns are similar, finding that M31's bulge has solar or 0.1 dex above solar metallicity. Then $\log Z/Z_\odot$ decreases linearly with disk radius. The two outliers are the "No IRAC" and "No MegaCam" campaigns. The optical SED is clearly of central importance for constraining the stellar metallicity.

Without the NIR and mid-IR points, the estimated metallicity is both higher (by 0.2 dex at $R_{\mathrm{M31}} = 1$ kpc) and has a shallower radial gradient. NIR light, including the 3.6 μm and 4.5 μm bands, is dominated by red giant branch (RGB) and asymptotic giant branch (AGB) stars at intermediate ages. The NIR colours of these branches are strong metallicity indicators. Thus this analysis underscores the value of NIR SEDs in stellar metallicity estimation.

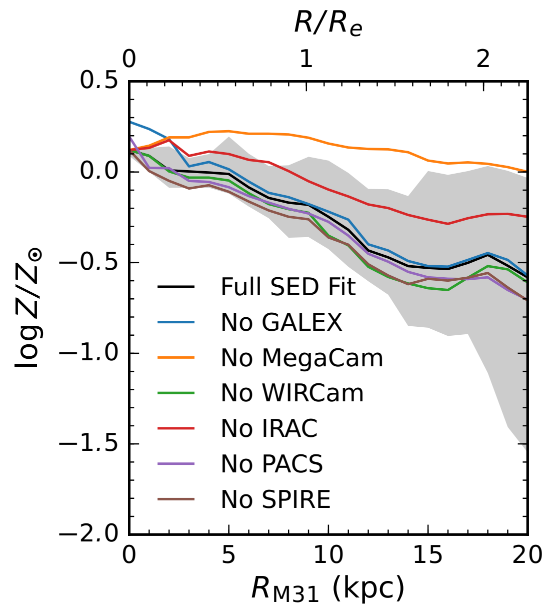


Figure 6.17: Jackknife analysis of stellar metallicity estimation bandpass dependence. *Left panel:* radial metallicity profiles estimated by MAGPHYS with different SED sets (jackknife analysis): the full SED fit, and fits where a given instrument is omitted ("No GALEX," *etc.*). The shaded region is the 1σ sample distribution of Full SED-fit estimates within isophotal annuli. Radii correspond to the elliptical isophotes (§ 5.5). *Right panel:* the $\pm 1\sigma$ sample distributions within isophotal annuli for each modelling campaign.

6.4.3 Comparison to Literature

Gregersen et al. (2015) measured stellar metallicity across the PHAT footprint in M31's northeastern sector. Those authors estimated metallicities of individual RGB stars by interpolating across a grid of isochrones in F475W − F814W colour-magnitude diagrams. In Figure 6.18 we directly compare our ANDROIDS/MAGPHYS full SED-fit metallicity estimates to those estimated by Gregersen et al. (2015) in each pixel of their study. Overall, the Gregersen et al. (2015) metallicity estimates are both higher (by 0.1 to 0.5 dex), and have a shallower radial gradient ($\log Z/Z_{\odot} =$ -0.02 dex kpc^{-1}). Unfortunately we cannot compare metallicity estimates inside $R_{\mathrm{M31}} \leq 4$ kpc due to

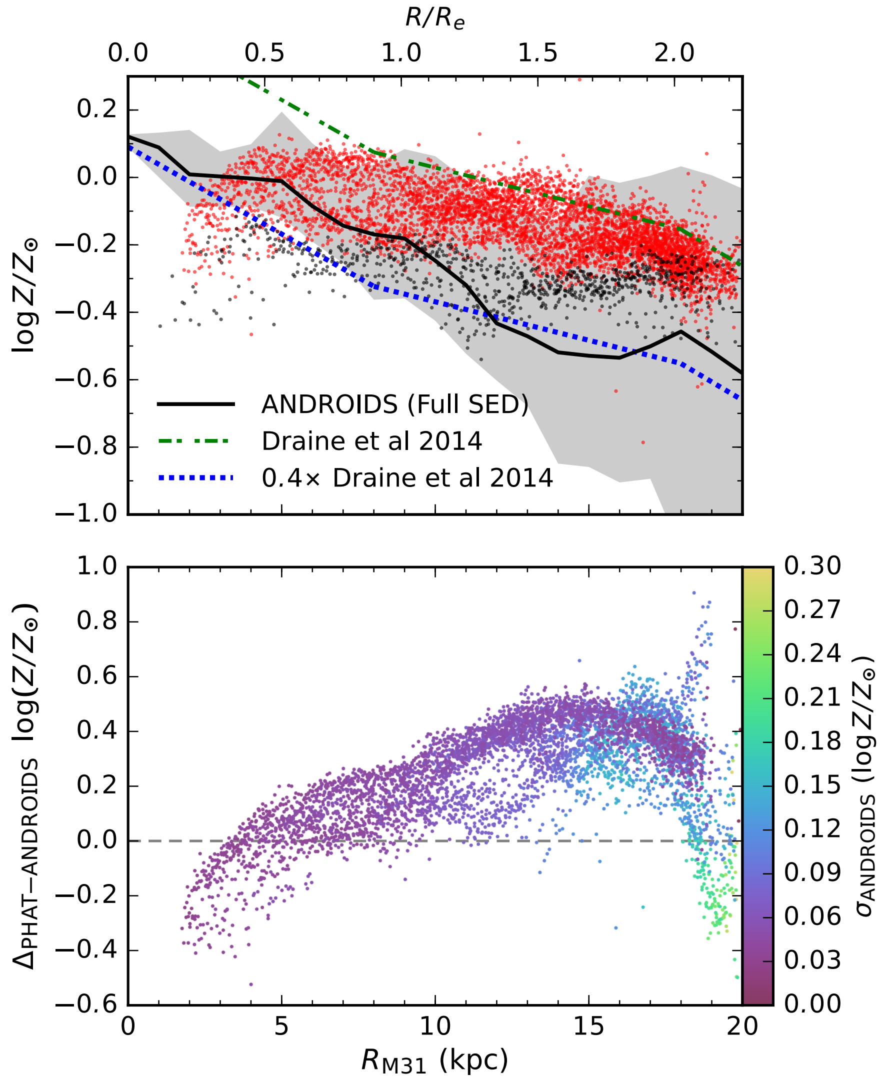


Figure 6.18: Radial metallicity estimate profiles. *Upper panel:* The black line is the median ANDROIDS/MAGPHYS Full SED fit and the shaded region is the 1σ sample distribution within isophotal annuli. Red and black dots are the Gregersen et al. (2015) and Williams et al. (2017), respectively, mean metallicity estimates in individual PHAT regions. The dash-dotted green line is the metallicity profile estimated by Draine et al. (2014) using dust mass mapping and $M_{\rm d}/M_{\rm H} = 0.009 Z/Z_{\odot}$ and the blue dashed line is the Draine et al. (2014) estimate scaled according to Dalcanton et al. (2015). *Lower panel:* radial metallicity estimate residual profile, showing the difference PHAT - ANDROIDS/MAGPHYS Full SED-fit metallicity estimates for each PHAT region. Individual points are the uncertainties of the MAGPHYS metallicity estimate (the 1σ confidence interval of the posterior distribution, Eq. 6.5).

crowding affects on the Williams et al. (2014) resolved stellar photometry.

Since the F475W − F814W colour of the RGB is affected not only by metallicity but also stellar age, a key assumption in Gregersen et al. (2015)'s metallicity estimate is that M31's disk stellar populations have a fixed age of 4 Gyr. Williams et al. (2017), who also modelled PHAT resolved stellar photometry later found that the mean age of M31's disk is much older (> 8 Gyr). As we show in § 6.5, we find a mean stellar age of 6 Gyr with MAGPHYS modelling. Gregersen et al. (2015) modelled the effects of stellar age on their metallicity estimates. For $\langle A \rangle = 6$ Gyr, $\log Z/Z_{\odot}$ decreases by 0.1 dex and by 0.2 dex for 8 Gyr. Neither an overall shift in age, or introduction of a radial age gradient, to the Gregersen et al. (2015) profile can reproduce our mean metallicity gradient.

Williams et al. (2017) also estimated metallicity with PHAT resolved photometry, though using a more conventional isochrone fitting approach. Their results are shown in Figure 6.18. The Williams et al. (2017) metallicity estimates are consistent with the Gregersen et al. (2015) estimates given the age effect discussed above. As well, the Williams et al. (2017) metallicity estimates are consistent with the mean metallicity found by our MAGPHYS fits in the mid-disk, though the steeper gradient we infer is not seen.

Finally, we also draw a comparison with the metallicity profile inferred by Draine et al. (2014) from their dust mass models of M31 (see § 6.6 for further discussion). Since metals in stellar atmospheres and dust in the interstellar medium are both products of previous generations of stellar evolution, it is possible to correlate the mean metallicity of a stellar population with the measured dust in a galaxy. Draine et al. (2014) estimate

$$\frac{\mathcal{M}_{\rm d}}{\mathcal{M}_{\rm H}} \approx 0.0091 \frac{Z}{Z_\odot} \tag{6.33}$$

where $\mathcal{M}_{\rm d}/\mathcal{M}_{\rm H}$ is the ratio of dust to gas mass. Figure 6.18 shows Draine et al. (2014)'s metallicity profile and we find it predicts a $\log Z/Z_\odot$ that is 0.2–0.3-dex higher than other estimates. However, as we discuss in § 6.6, Dalcanton et al. (2015) find that the Draine et al. (2014) $\mathcal{M}_d$ is a factor 2.5× too large. After applying this correction factor, we find that Draine et al. (2014) metallicity profile based has both a similar zeropoint and slope as our own MAGPHYS profile.

6.4.4 Summary

In summary, we have considered four distinct approaches for estimating the radial stellar metallicity profile of M31. After considering corrections, all four methods find that the mid-disk of M31 has a sub-solar metallicity ($\log Z/Z_\odot \sim -0.2$). Estimates based on resolved stellar populations preduct a much flatter metallicity gradient than our own based on SED modelling and that based on $\mathcal{M}_{\rm d}/\mathcal{M}_{\rm H}$.

6.5 Mean stellar age estimation

As described in § 6.2.1, star formation consists of a continuous exponentially-declining star formation rate and stochastic star bursts. Since several parameters simultaneously contribute to the star formation history of MAGPHYS models, in this section we use the mass-weighted mean stellar age, $\langle A \rangle$, as proxy for the overall star formation history.

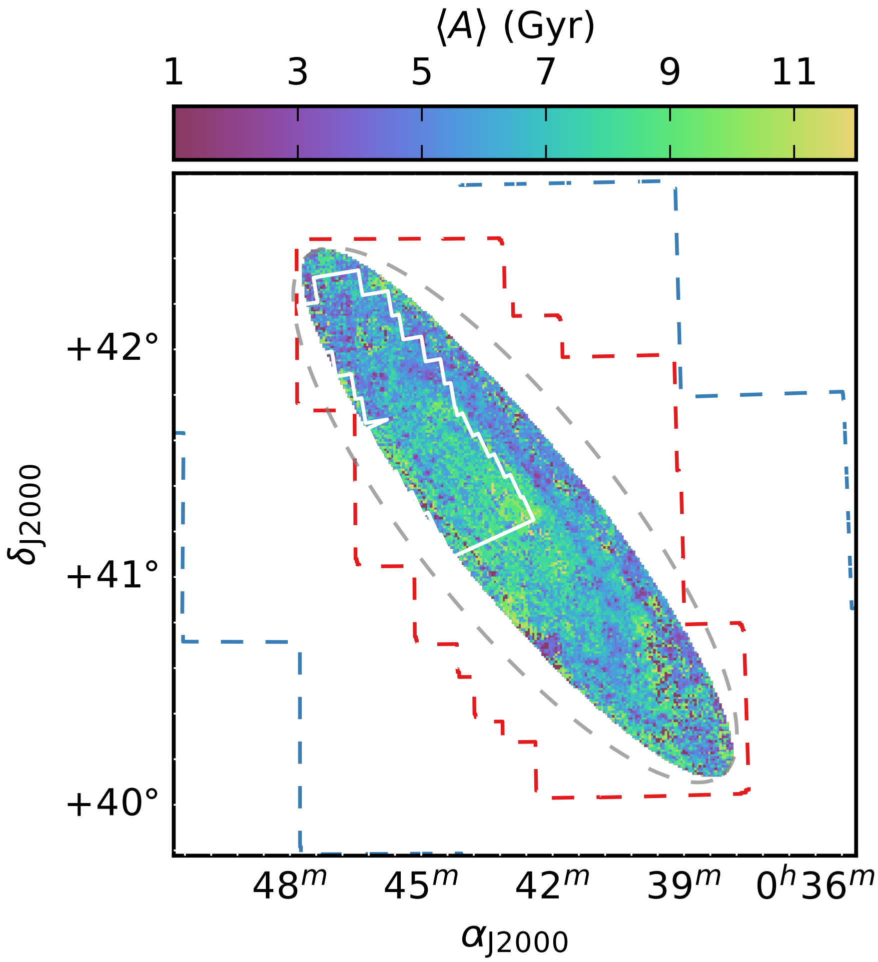


Figure 6.19: Map of ANDROIDS/MAGPHYS Full SED-fit mass-weighted mean stellar age estimates. See Figure 6.15 for information about the reference footprints.

6.5.1 MAGPHYS Mean Age Estimates

In Figure 6.19 we plot the mean stellar age estimated by ANDROIDS/MAGPHYS full SED modelling. Overall, the bulge of M31 is quite old, $\langle A \rangle \approx 9$ Gyr. The mean stellar age drops to 6 Gyr in the mid and outer disk.

As previously discussed in § 6.4.1, the SED-based age estimates are biased by near-far dust attenuation effect. The more heavily attenuated near side of the disk

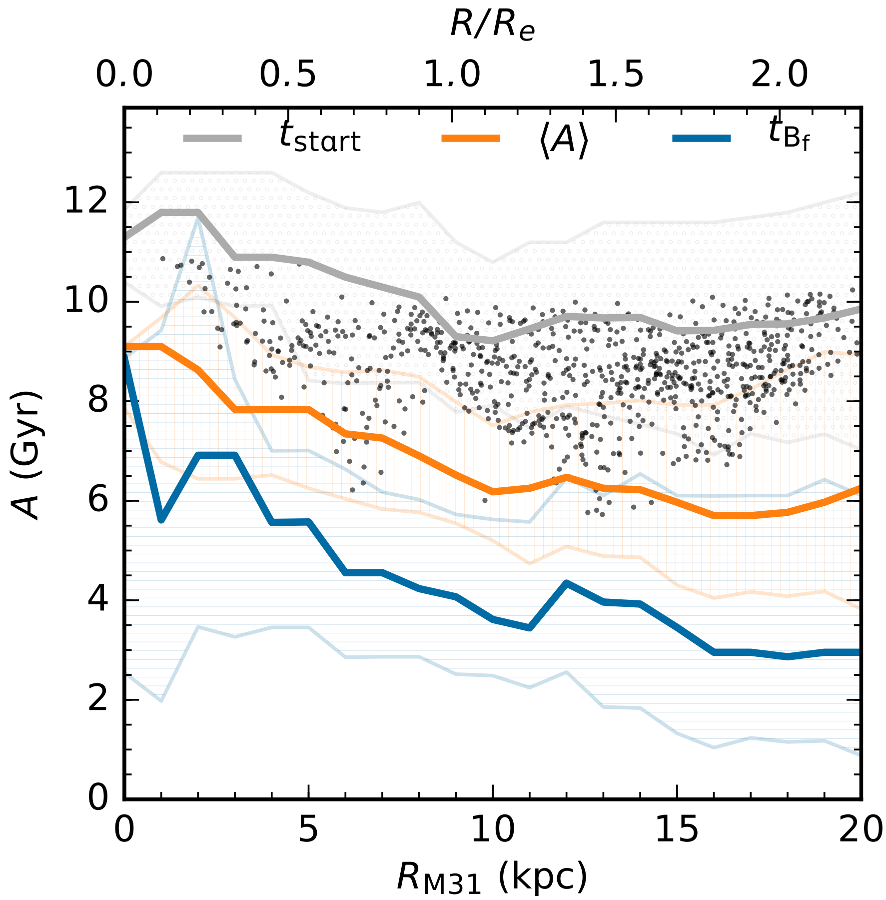


Figure 6.20: Radial profile of key ages in M31's stellar assembly estimated by the MAGPHYS full SED fit: age of the start of star formation ($t_{\rm start}$), mean age ($\langle A \rangle$), and age of the last starburst (t_{B_f}). Lines are median radial profiles of these parameters, while circle, vertical line, and horizontal line-hatched regions describe the 1σ sample distributions within isophotal annuli of each parameter, respectively. Black dots are the mean ages estimated by Williams et al. (2017) in individual PHAT fields.

(northwestern minor axis; top right side of Figure 6.19) is estimated as $\sim$ 2 Gyr younger than the far side. The youngest age estimates appear along the 10 kpc dust ring on the near side. Aside from the near-far effect, the 10 kpc ring is not prominent in the mean age map. This corroborates resolved stellar population studies whereby the 10 kpc ring has persisted for ¿1 Gyr (Lewis et al. 2015) and most of M31's stellar

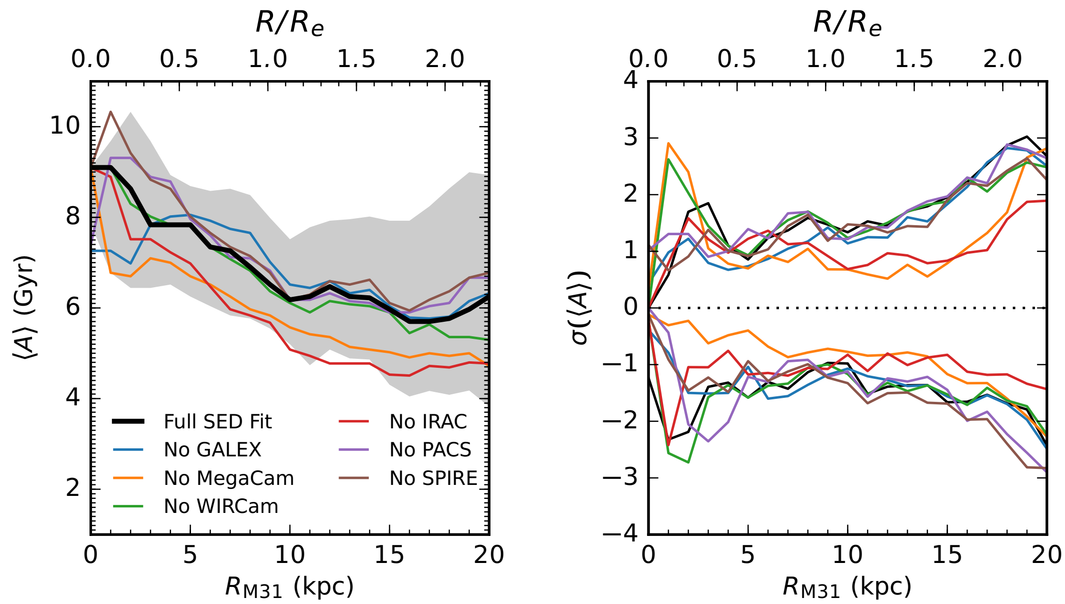


Figure 6.21: Jackknife analysis of mean stellar age estimation bandpass dependence. *Left panel:* radial mean age profiles estimated by MAGPHYS with different SED sets (jackknife analysis): the full SED fit, and fits where a given instrument is omitted ("No GALEX," *etc.*). The shaded region is the 1σ sample distribution of Full SED-fit estimates within isophotal annuli. Radii correspond to the elliptical isophotes (§ 5.5). *Right panel:* the $\pm 1\sigma$ sample distributions within isophotal annuli for each modelling campaign.

mass had already formed about 8 Gyr ago and has radially mixed (Williams et al. 2017).

In Figure 6.20 we reduce this map to a 1D radial profile. From this profile, we can identify at least three distinct segments. First, the inner bulge of M31 is on average 9 Gyr old. The mean stellar age decreases steeply with -0.45 Gyr kpc^{-1} to approximately $R_{\rm M31} = 5$ kpc. Between 5 and 10 kpc, the age gradient is shallower, -0.15 Gyr kpc^{-1}. Finally, beyond 10 kpc, the mean age gradient is flat at $\langle A \rangle = 6$ Gyr.

6.5.2 Bandpass Dependence

We use multiple MAGPHYS modelling campaigns where different instruments are omitted from the fitted SED to assess the bandpass sensitivity of our mean age estimates. As we show in Figure 6.21, similar mean age profiles are recovered from each modelling campaign within the 1σ uncertainty envelope of the full SED fit. Omitting either the Herschel PACS and SPIRE far-IR bands tends to raise the mean age of the bulge to 10 Gyr. Omitting the MegaCam optical SED results in a flat age profile with a mean age of 7 Gyr. As we found in § 6.4.2 with respect to metallicity estimation, IRAC's near and mid-IR bandpasses are influential. Without IRAC, the mean age in the mid disk drops by 1 Gyr to 5 Gyr compared to the full SED fit.

All modelling campaigns exhibit similar estimate distributions along isophotal annuli. The typical mean age uncertainty in the bulge is ± 1 Gyr, and grows to ± 3 Gyr towards the outer disk.

6.5.3 Discussion

Thus far we have focused on the mean age metric. In Figure 6.20 we elaborate on the stellar assembly timeline indicated by our MAGPHYS modelling by showing estimated radial profiles for the lookback times to the start of star formation (t_0) and the last stochastic star burst (t_{B_f}), in addition to mean age. The estimated lookback time to the start of star formation follows the mean age: star formation began in the bulge is 12 Gyr ago. In the mid and outer disk, star formation started 10 Gyr ago. The ANDROIDS SEDs also favour models where the most recent star formation burst occurred at least 2 Gyr ago. Qualitatively, this picture of a very old bulge and a moderately old and quiescent disk is similar to results in the literature.

Saglia et al. (2010), using long-slit spectroscopy, found that M31's bulge is 12 Gyr old and above solar metallicity. This result is older than the ANDROIDS/MAGPHYS result, where the mean age in the bulge is 8–9 Gyr. Williams et al. (2017), using PHAT resolved stellar photometry, found that most stellar mass in M31's disk had formed at least 8 Gyr ago, which is older than the $\langle A \rangle = 6$ Gyr estimated with MAGPHYS. Further, they found that M31 was quiescent from its original formation over 8 Gyr ago to a most recent star formation episode 2–4 Gyr ago. It is conceivable that the MAGPHYS models are biased to younger mean ages to accommodate this second star formation episode. Finally, Williams et al. (2017) found that stellar populations are constant throughout M31's disk, presumably due to radial mixing of the old stellar population. This flat age profile is compatible with the ANDROIDS/MAGPHYS results. ANDROIDS only finds a significant age gradient in the transition from the bulge to disk stellar populations.

6.6 Dust Estimation

The panchromatic ANDROIDS SED is ideal for modelling M31's dust content. The IRAC and Herschel observations feature the dust emission corresponding to attenuation in the GALEX and CFHT MegaCam and WIRCam observations. With MAGPHYS (da Cunha et al. 2008), we self-consistently model M31's dust ISM by requiring energy conservation across dust attenuation and emission.

In this section we review the properties of M31's dusty interstellar medium inferred from MAGPHYS modelling of the ANDROIDS at a $36'' \times 36''$ (137 pc $\times$ 608 pc) scale. As before, our goals are both to characterize M31's dust and stellar populations, and to understand estimation systematics associated with broadband SED modelling. First,

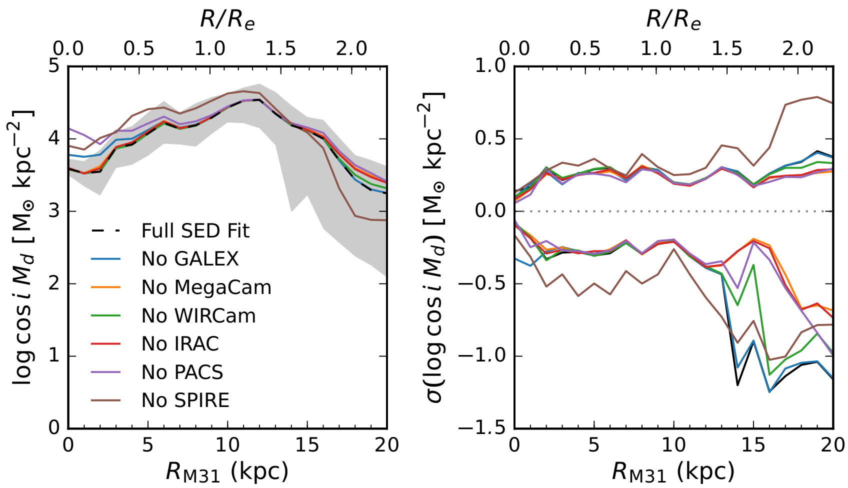


Figure 6.22: Jackknife analysis of dust mass ($\log M_{\rm d}$) estimation bandpass dependence. *Left panel:* radial $\log M_{\rm d}$ profiles estimated by MAGPHYS with different SED sets (jackknife analysis): the full SED fit, and fits where a given instrument is omitted ("No GALEX," *etc.*). The shaded region is the 1σ sample distribution of Full SED-fit estimates within isophotal annuli. Radii correspond to the elliptical isophotes (§ 5.5). *Right panel:* the $\pm 1\sigma$ sample distributions within isophotal annuli for each modelling campaign.

in § 6.6.1 we consider the relative contributions of different instruments to the models with a jackknife-like analysis. Then in § 6.6.2 we directly compare our ANDROIDS models with results from Viaene et al. (2014), hereafter HELGA IV, a study that also used MAGPHYS and similar SED data as this study.

6.6.1 Bandpass dependence of dust modelling

By repeatedly modelling the ANDROIDS dataset with different instruments systematically omitted from the SED, we understand how different bandpasses affect MAGPHYS's dust mass estimation. In Figure 6.22, we find that a complete UV and IR SED is required to minimize bias in $M_{\rm d}$ estimates. At all radii, the full SED fit is on or within the bounds of estimates where an instrument was omitted. Note that the sign of the bias in M_d estimation is not consistent and appears to depend on the nature of the SED and indeed, the environment in the galaxy. For example, model campaigns that omit the Herschel SPIRE SED points overestimate M_d by 0.3 dex in inner regions, but underestimate M_d by 0.5 dex in outer regions. Given the importance of a well-sampled SED for dust mass estimation, recall that the ANDROIDS SED dataset does not include mid-IR WISE and Spitzer MIPS observations. In the next section we determine the potential effect of those omissions through direct comparison to other studies.

6.6.2 Dust estimates and literature comparison

In this section we present maps of MAGPHYS parameter estimates related to M31's dust and compare these estimates to other studies. In doing so we both gain an understanding of M31's dust and stellar populations, while also understanding systematics association with different observational and modelling approaches.

As discussed in Chapter 5, this study is designed to follow the approach taken by Viaene et al. (2014) (HELGA IV) to permit direct pixel-to-pixel parameter comparison. The HELGA IV SED consists of observations from GALEX, SDSS, Spitzer IRAC and MIPS, WISE, and Herschel PACS and SPIRE (indeed, we have used their reduced Herschel images in this study). Thus they have better SED sampling in the mid-IR (through MIPS and WISE), though ANDROIDS's MegaCam and WIRCam images should be an improvement over the SDSS mosaics. Finally, recall that we convolved the ANDROIDS SED set to the Herschel SPIRE 500 μm PSF and resampled to HELGA IV's $36'' \times 36''$ pixel grid to enable direct pixel-to-pixel comparisons. HELGA IV used MAGPHYS in a manner similar to ours in this Chapter, with the exception of implementing an enhanced dust model library compared to the standard MAGPHYS distribution. Whereas MAGPHYS normally includes cold ISM dust models with temperatures ranging from 15 K $< T_C^{\rm ISM} <$ 25 K and warm dust with 30 K $< T_W^{\rm BC} <$ 60 K, HELGA IV's custom model library spans 10 K $< T_C^{\rm ISM} <$ 30 K and warm dust with 30 K $< T_W^{\rm BC} <$ 70 K. Thus it is possible that the ANDROIDS results are biased by using a library of dust models with narrower temperature distributions. Overall, HELGA IV is an excellent verification of our own SED calibrations (Chapters 2–5) and application of MAGPHYS (§ 6.2).

In addition to HELGA IV, we also draw comparisons with two other studies of M31's global dust distribution: Draine et al. (2014) and Dalcanton et al. (2015). Draine et al. (2014) modelled M31's infrared SED (composed of Spitzer and Herschel maps) with their Draine et al. (2007) dust model. This work is distinct from the MAGPHYS-based ANDROIDS and HELGA IV studies both because they do not directly model attenuation of the stellar SED and the Draine et al. (2007) dust models differ from the Dunne et al. (2000) dust emission model used by MAGPHYS. Dalcanton et al. (2015) took an entirely different approach to modelling M31's dust by measuring the extinction seen towards individual stars in PHAT resolved stellar photometry.

Dust Luminosity

Let us first compare the luminosity models. Dust luminosity estimates reflect on the overall quality of the input SED data since $L_{\rm d}$ has a straightforward relationship to the SED itself. Dust mass is useful as a proxy for more complex dust physics. In Table 6.6 we report cumulative profiles of both $M_{\rm d}$ and $L_{\rm d}$ as modelled by this study,

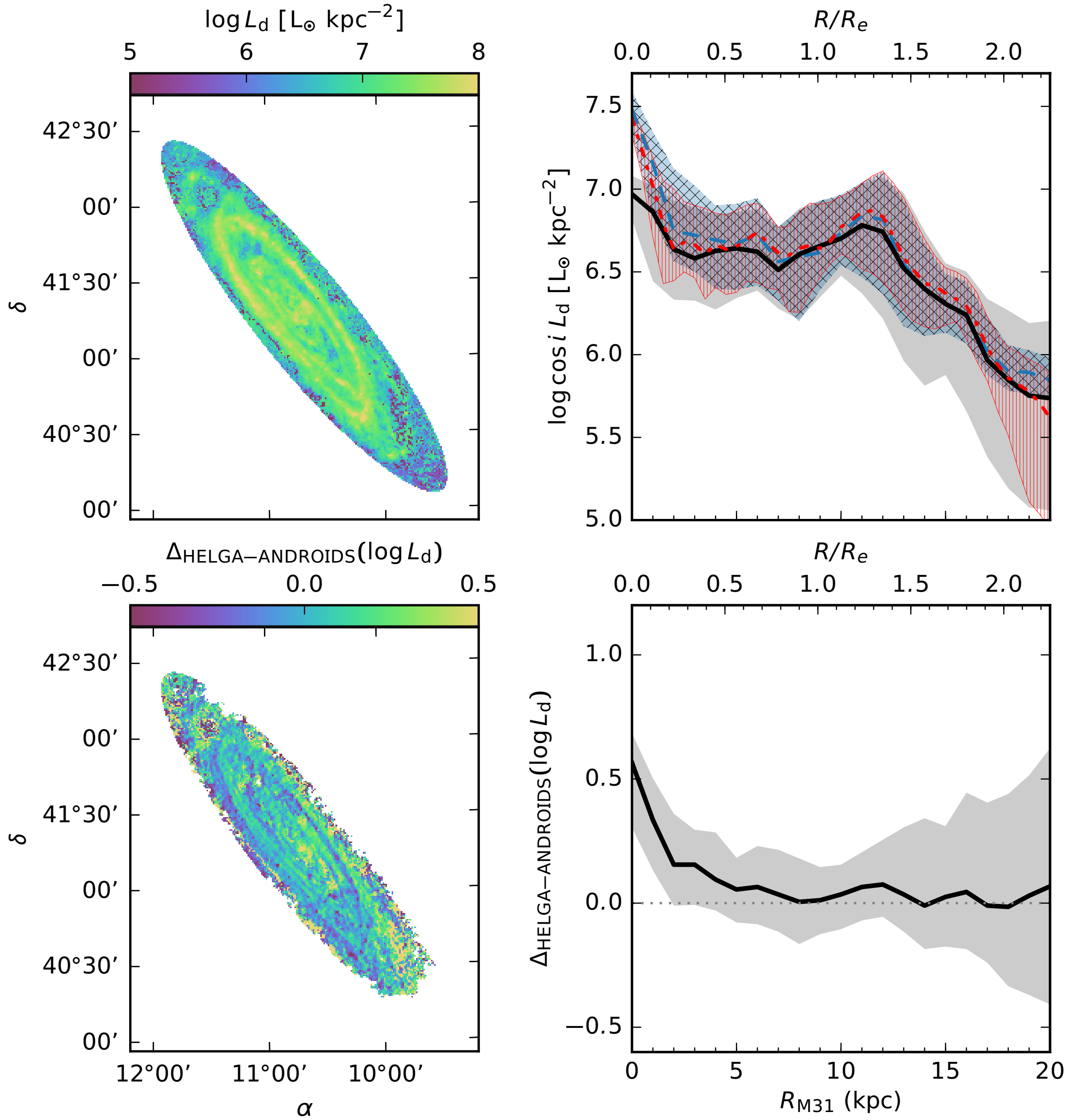


Figure 6.23: Estimated total dust luminosity $\log L_\mathrm{d}$. *Top:* map and radial profile (solid black line) of $\log L_\mathrm{d}$ estimated by MAGPHYS models of the full ANDROIDS SED. The dashed blue line is the HELGA IV profile while the dashed-dotted red line is the estimate by Draine et al. (2014). The profile is deprojected for M31's inclination given $\cos i = 0.213$. *Bottom:* map and radial profile of the difference between $\log(L_\mathrm{d}/L_\odot)$ estimated by HELGA IV and ANDROIDS.

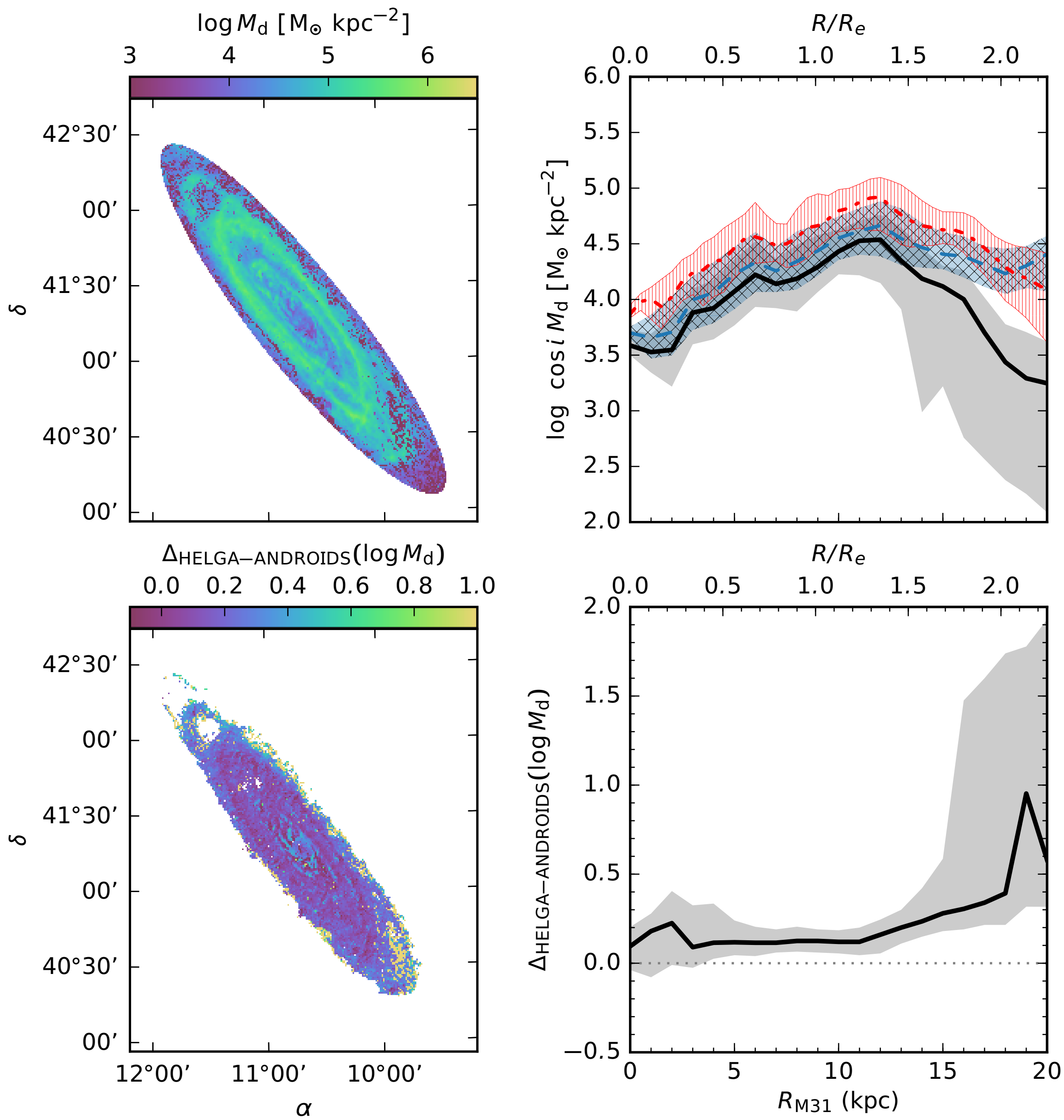


Figure 6.24: Estimated dust mass, $\log M_\mathrm{d}$. *Top:* map and radial profile (solid black line) of $\log M_\mathrm{d}$ estimated by MAGPHYS models of the full ANDROIDS SED. The dashed blue line is the HELGA IV profile while the dashed-dotted red line is the estimate by Draine et al. (2014). The profile is deprojected for M31's inclination given $\cos i = 0.213$. *Bottom:* map and radial profile of the difference between $\log M_\mathrm{d}$ estimated by HELGA IV and ANDROIDS.

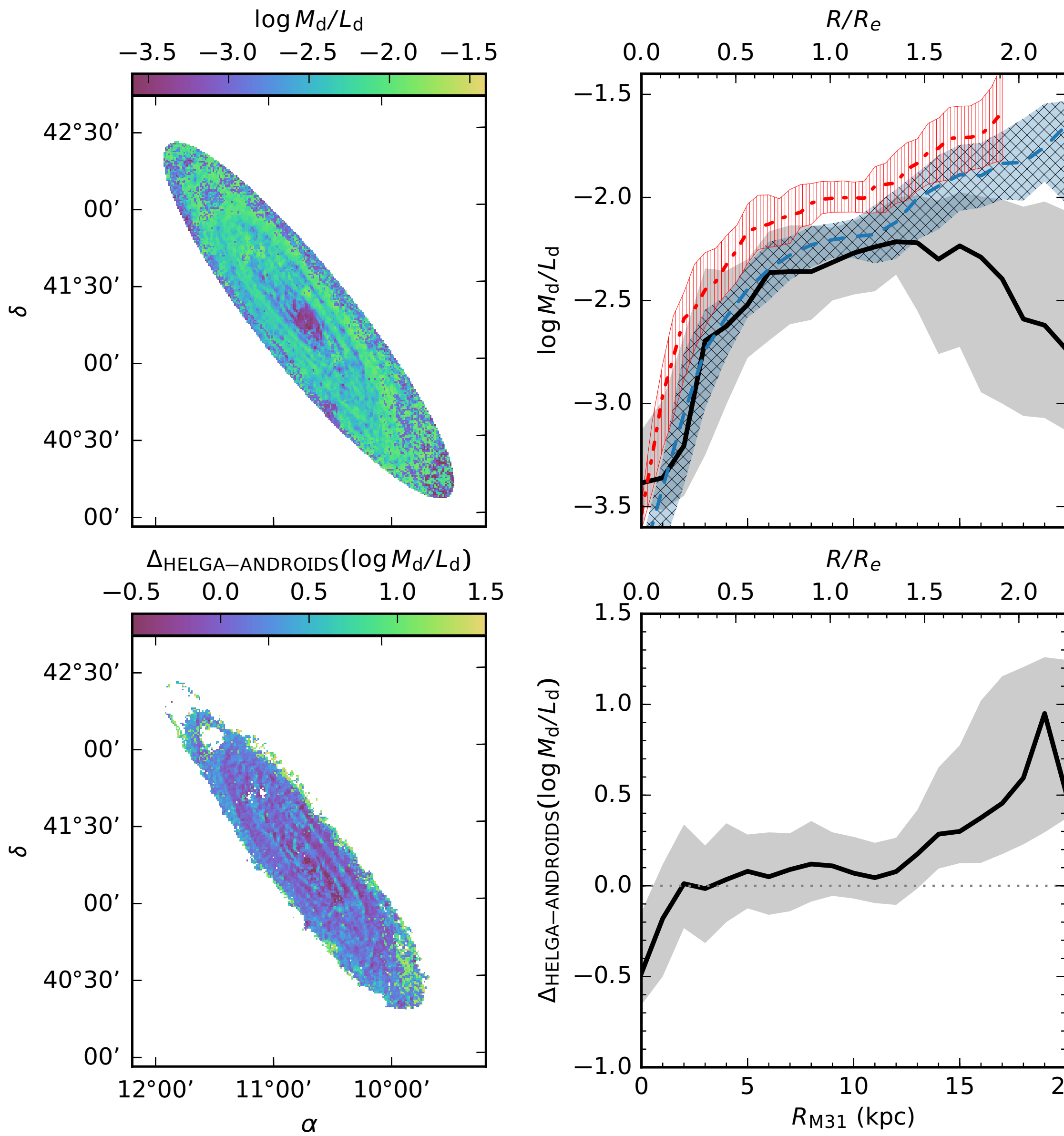


Figure 6.25: Estimated dust mass-to-light ratio, $\log(M_\mathrm{d}/L_\mathrm{d})$. *Top:* map and radial profile (solid black line) of $\log(M_\mathrm{d}/L_\mathrm{d})$ estimated by MAGPHYS models of the full ANDROIDS SED. The dashed blue line is the HELGA IV profile while the dashed-dotted red line is the estimate by Draine et al. (2014). *Bottom:* map and radial profile of the difference between $\log(M_\mathrm{d}/L_\mathrm{d})$ estimated by HELGA IV and ANDROIDS.

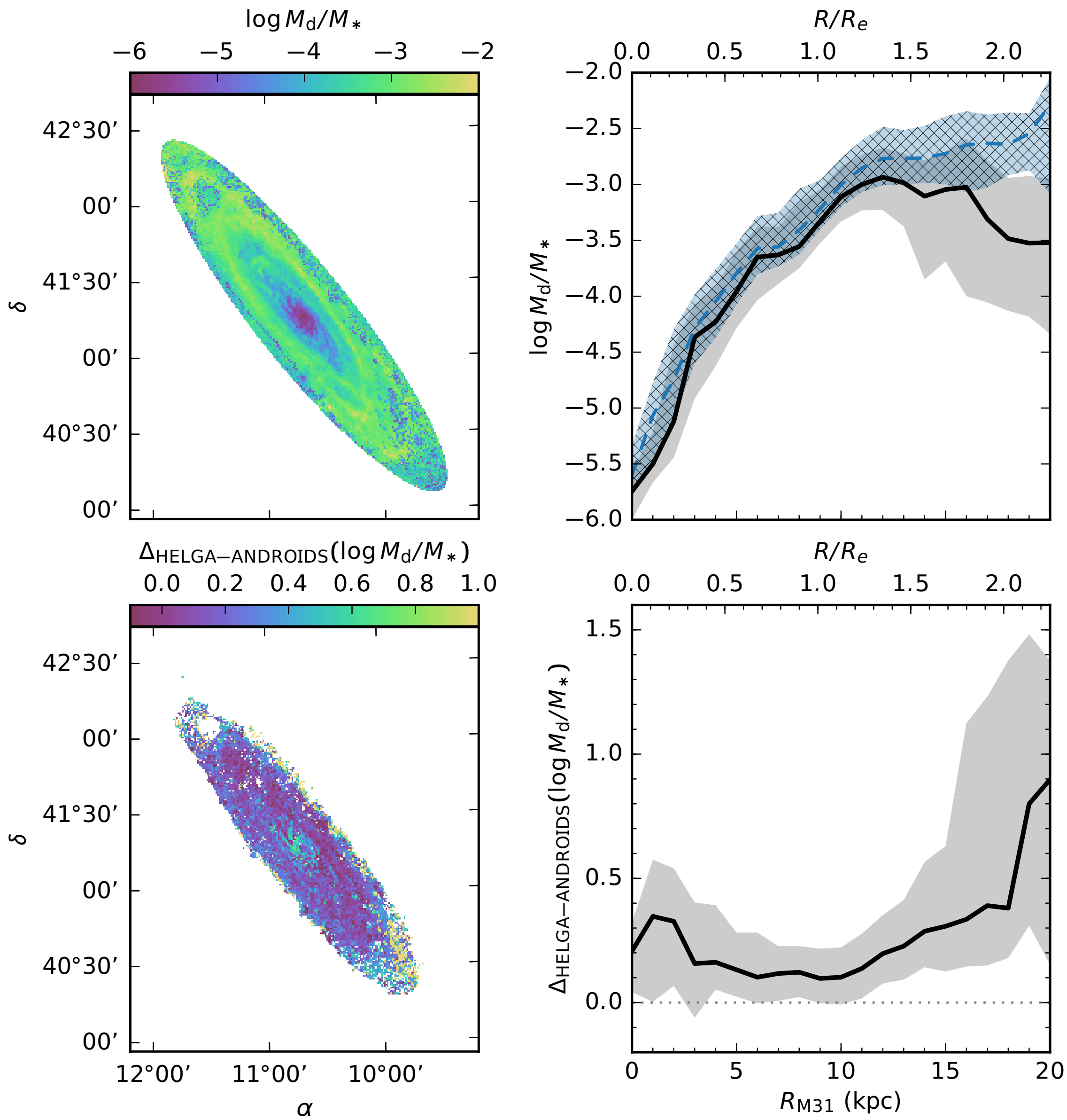


Figure 6.26: Estimated dust-to-stellar mass ratio, $\log(M_\mathrm{d}/M_*)$. *Top:* map and radial profile (solid black line) of $\log(M_\mathrm{d}/M_*)$ estimated by MAGPHYS models of the full ANDROIDS SED. The dashed blue line is the HELGA IV profile. *Bottom:* map and radial profile of the difference between $\log(M_\mathrm{d}/M_*)$ estimated by HELGA IV and ANDROIDS.

Table 6.6: Integrated radial profiles of dust mass (M_d) and luminosity (L_d) modelled by this study, HELGA IV, and Draine et al. (2014). Values are integrated within isophotal ellipses (§ 5.5) of the specified radius, except for values by Draine et al. (2014) that are taken from their Table 4. Uncertainties are provided in parentheses.

R_M31 kpc	$M_\mathrm{d,ANDROIDS}$	10^7 M$_\odot$	$M_\mathrm{d,HELGAIV}$	10^7 M$_\odot$	$M_\mathrm{d,D14}$ 10^7 M$_\odot$	$L_\mathrm{d,ANDROIDS}$	10^9 $L_\odot$	$L_\mathrm{d,HELGAIV}$	10^9 $L_\odot$	$L_\mathrm{d,D14}$ 10^9 $L_\odot$
1	0.00	(0.00)	0.00	(0.00)	·	0.03	(0.00)	0.07	(0.00)	·
2	0.00	(0.00)	0.01	(0.00)	·	0.08	(0.00)	0.17	(0.01)	·
3	0.02	(0.00)	0.02	(0.00)	·	0.16	(0.01)	0.28	(0.01)	·
4	0.04	(0.00)	0.05	(0.00)	·	0.27	(0.01)	0.41	(0.01)	·
5	0.07	(0.00)	0.10	(0.00)	·	0.41	(0.01)	0.57	(0.01)	·
6	0.13	(0.00)	0.18	(0.00)	·	0.58	(0.01)	0.78	(0.01)	·
7	0.21	(0.00)	0.28	(0.00)	·	0.77	(0.01)	0.98	(0.01)	·
8	0.28	(0.00)	0.38	(0.00)	·	0.96	(0.01)	1.17	(0.01)	·
9	0.40	(0.00)	0.54	(0.01)	·	1.26	(0.01)	1.44	(0.02)	·
10	0.56	(0.01)	0.75	(0.01)	·	1.60	(0.02)	1.79	(0.02)	·
11	0.78	(0.01)	1.04	(0.01)	·	2.06	(0.02)	2.26	(0.02)	·
12	1.05	(0.01)	1.39	(0.01)	·	2.58	(0.02)	2.83	(0.02)	·
13	1.31	(0.01)	1.78	(0.01)	·	3.08	(0.03)	3.36	(0.02)	·
14	1.49	(0.01)	2.07	(0.01)	·	3.47	(0.03)	3.70	(0.03)	·
15	1.62	(0.01)	2.33	(0.01)	·	3.72	(0.03)	3.94	(0.03)	·
16	1.74	(0.01)	2.59	(0.01)	·	3.94	(0.03)	4.15	(0.03)	·
17	1.83	(0.01)	2.78	(0.01)	4.50	4.11	(0.03)	4.31	(0.03)	3.98
18	1.87	(0.01)	2.84	(0.01)	·	4.23	(0.03)	4.39	(0.03)	·
19	1.90	(0.01)	2.86	(0.01)	·	4.33	(0.03)	4.45	(0.03)	·
20	1.93	(0.01)	2.88	(0.01)	5.03	4.44	(0.03)	4.48	(0.03)	4.16

HELGA IV, and Draine et al. (2014).

Overall, ANDROIDS reproduces HELGA IV's M31 dust emission luminosity quite well, at 4.6×10^9 $L_\odot$ within $R = 20$ kpc. Draine et al. (2014), using a slightly different set of Herschel observations, found a similar value at 4.2×10^9 $L_\odot$. All three L_d estimates appear to have nearly identical radial profiles in Figure 6.23. Again, this reassures us that the infrared SED we assembled in Chapter 5 is correct. As Figure 6.23 shows, L_d declines exponentially beyond the 10 kpc ring.

Dust Mass

The dust mass estimates are more diverse, though. Overall, we find the $M_{\rm d}$ within 20 kpc is 1.9×10^7, while the HELGA IV estimate is 50% higher (Table 6.6). Draine et al. (2014)'s $M_{\rm d}(20~\text{kpc}) = 5.0 \times 10^7$ is higher still. The Draine et al. (2014) mass is suspect, though, as Dalcanton et al. (2015) find that a dust mass 2.5× lower than Draine et al. (2014) (or $M_d(20~\text{kpc}) = 2.5~\text{M}_\odot$) is compatible with their A_V extinction maps. Thus it appears that the Dunne et al. (2000) dust emission model used by MAGPHYS is, overall, more suitably calibrated for M31 than the Draine et al. (2007) model.

While the Draine et al. (2007) dust model is heavier than Dunne et al. (2000), it is proportionately so. Figure 6.24 shows that Draine et al. (2014)'s $M_{\rm d}$ estimate follows our own, albeit with a +0.4 dex zeropoint shift. For comparison, the HELGA IV $M_{\rm d}$ profile is +0.1 dex higher than ours. Overall, this suggests that all three models are rooted in similar-quality observations, but that the dust models are slightly different in a way that is predominantly uniform across all radii.

Although the ANDROIDS and HELGA IV $M_{\rm d}$ estimates are similar, we find a striking difference in the dust mass profiles at large radii. As shown in Figure 6.24, the dust surface density modelled by ANDROIDS peaks at the outer radius of the 10 kpc dust rings. On the other hand, the HELGA IV dust mass surface density has a much shallower drop beyond 10 kpc. In relative terms, both the dust mass-to-light ratio (Figure 6.25) and dust-to-stellar mass ratio (Figure 6.26) increase with radius, while the same quantities as modelled by ANDROIDS are relatively constant beyond 10 kpc. An increasing M_d/M_* ratio with radius is consistent with dust disks having longer exponential scale lengths than stellar disks. By measuring overlapping disk galaxies (occultations), White et al. (2000) found that the stellar and dust disks have similar scale lengths, consistent with ANDROIDS. On the other hand, Xilouris et al. (1998) found that the scale length of the dust disk is 1.5× that of stellar disk in the edge-on galaxy NGC 891. Finally, note that HELGA IV's $M_{\rm d}/M_{*}$–stellar mass surface density is consistent with that found by the Herschel Reference Survey (Cortese et al. 2012), which makes the ANDROIDS dust mass result an outlier.

ISM and Birth Cloud Dust Components

A key feature of MAGPHYS's dust treatment is the Charlot & Fall (2000) model that handles the opacity towards "birth clouds" (stellar populations younger than 10 Myr) from the interstellar medium (ISM) that affects all stellar populations, as we reviewed in § 6.2. A key parameter is f_{μ}, defined as the ratio of infrared luminosity emitted by ISM dust to the aggregate infrared luminosity, including birth clouds (Eq. 6.11). We find that ANDROIDS overestimates the contribution of dust in birth clouds relative to the HELGA IV models. As we show in Figure 6.27, HELGA IV estimated that the ISM contributes 80%–95% of the infrared light, while the ANDROIDS models find $f_{\mu} \approx 0.6$. In addition, we find that ANDROIDS models exchange lower optical depth in the ISM (Figure 6.28) with very high optical depth towards birth clouds (Figure 6.29). The optical depth towards birth clouds, τ_{V}, may mask enhanced recent star formation modelled by ANDROIDS (Figure 6.30).

These results hint at a modelling degeneracy between star formation rate and the modelling of ISM versus birth cloud dust components. One reason for this may be the lack of Spitzer MIPS mid-IR SED data our ANDROIDS SED that could otherwise help constrain the hot dust components associated with birth clouds. Another issue

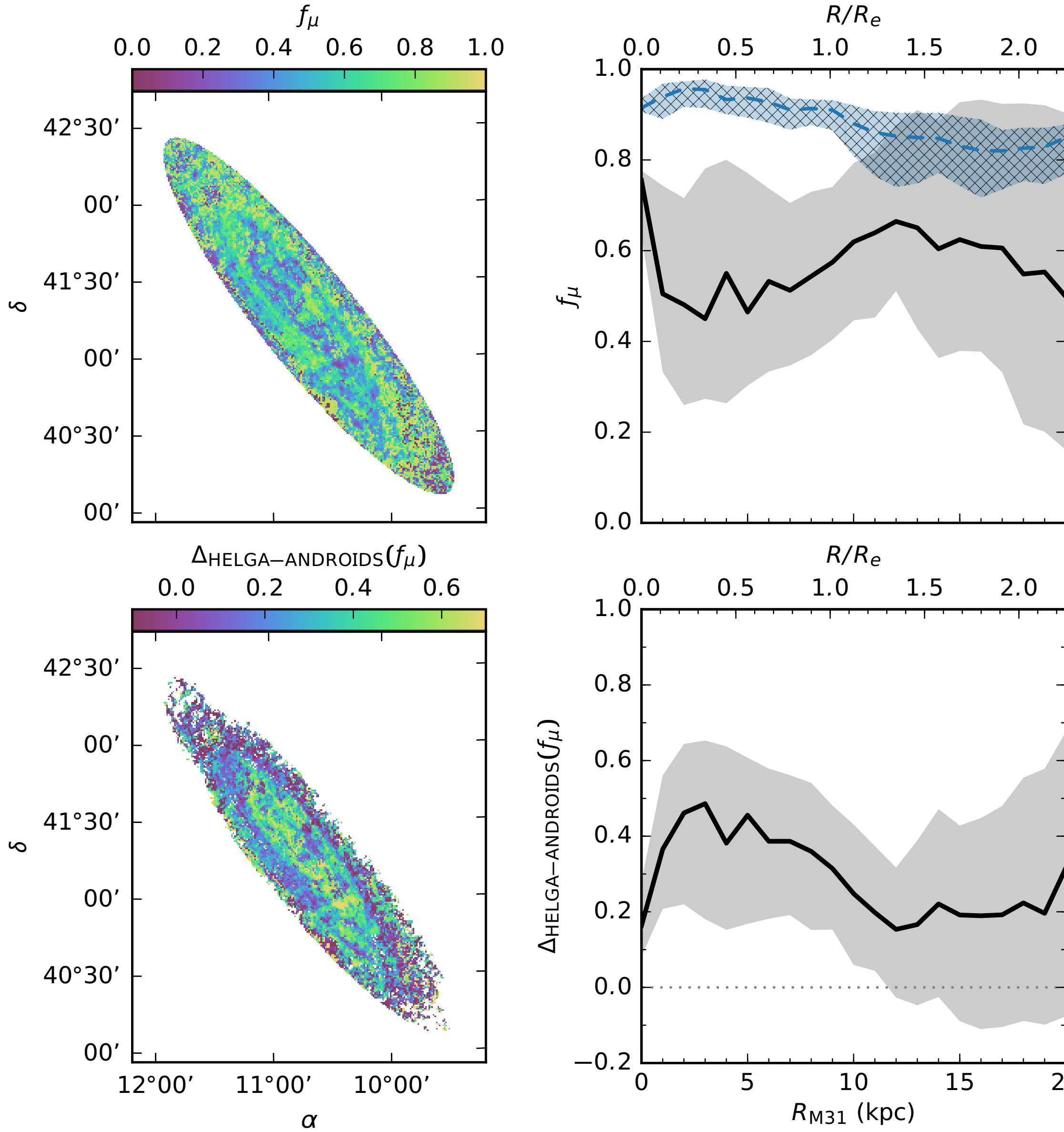


Figure 6.27: Estimated ratio of ISM-to-total dust luminosity, f_μ. *Top:* map and radial profile (solid black line) of f_μ estimated by MAGPHYS models of the full ANDROIDS SED. The dashed blue line is the HELGA IV profile. *Bottom:* map and radial profile of the difference between f_μ estimated by HELGA IV and ANDROIDS.

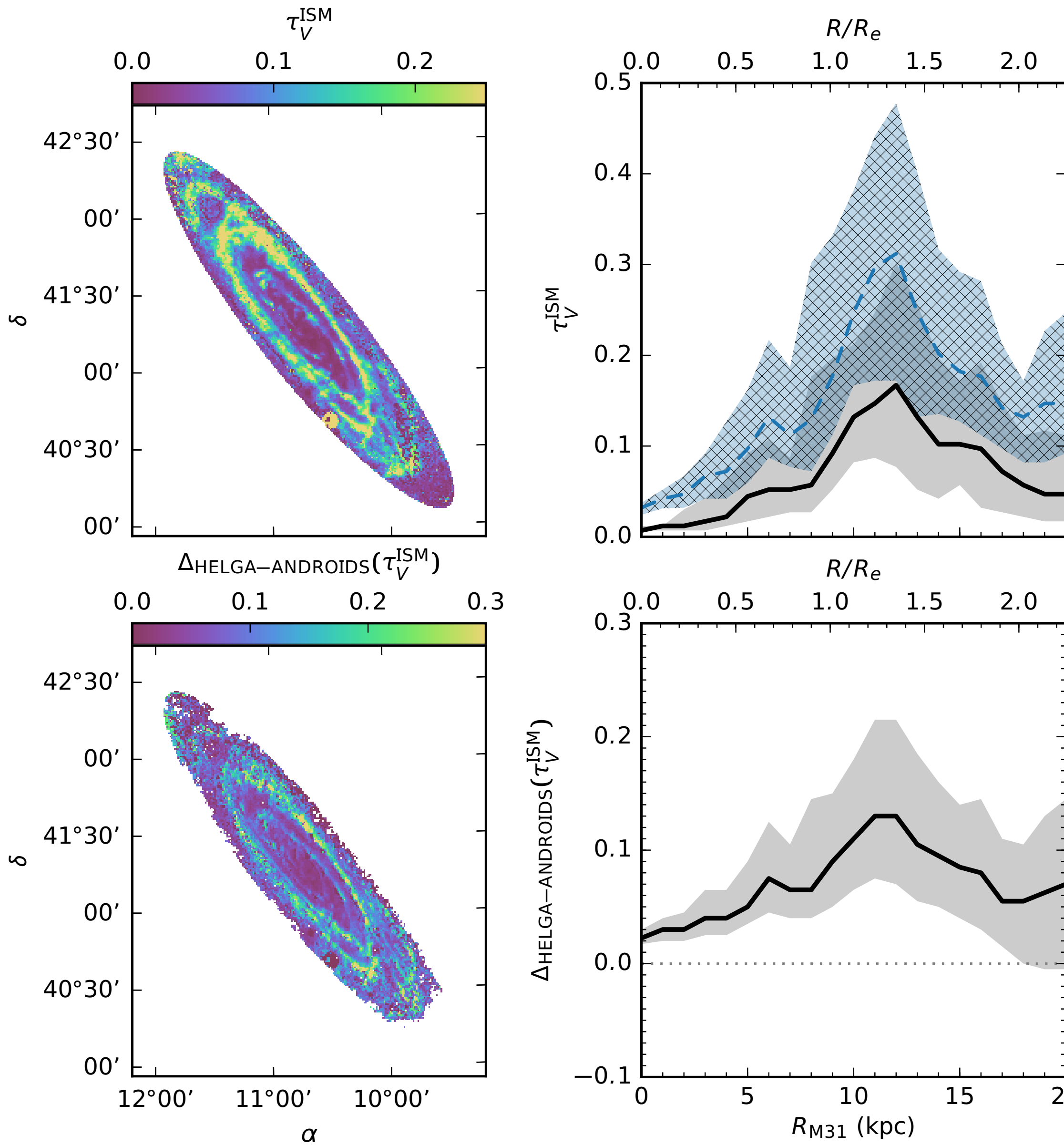


Figure 6.28: ISM dust optical depth ($\tau_V^{\rm ISM}$). *Top:* map and radial profile (solid black line) of $\tau_V^{\rm ISM}$ estimated by MAGPHYS models of the full ANDROIDS SED. The dashed blue line is the HELGA IV profile. *Bottom:* map and radial profile of the difference between $\tau_V^{\rm ISM}$ estimated by HELGA IV and ANDROIDS.

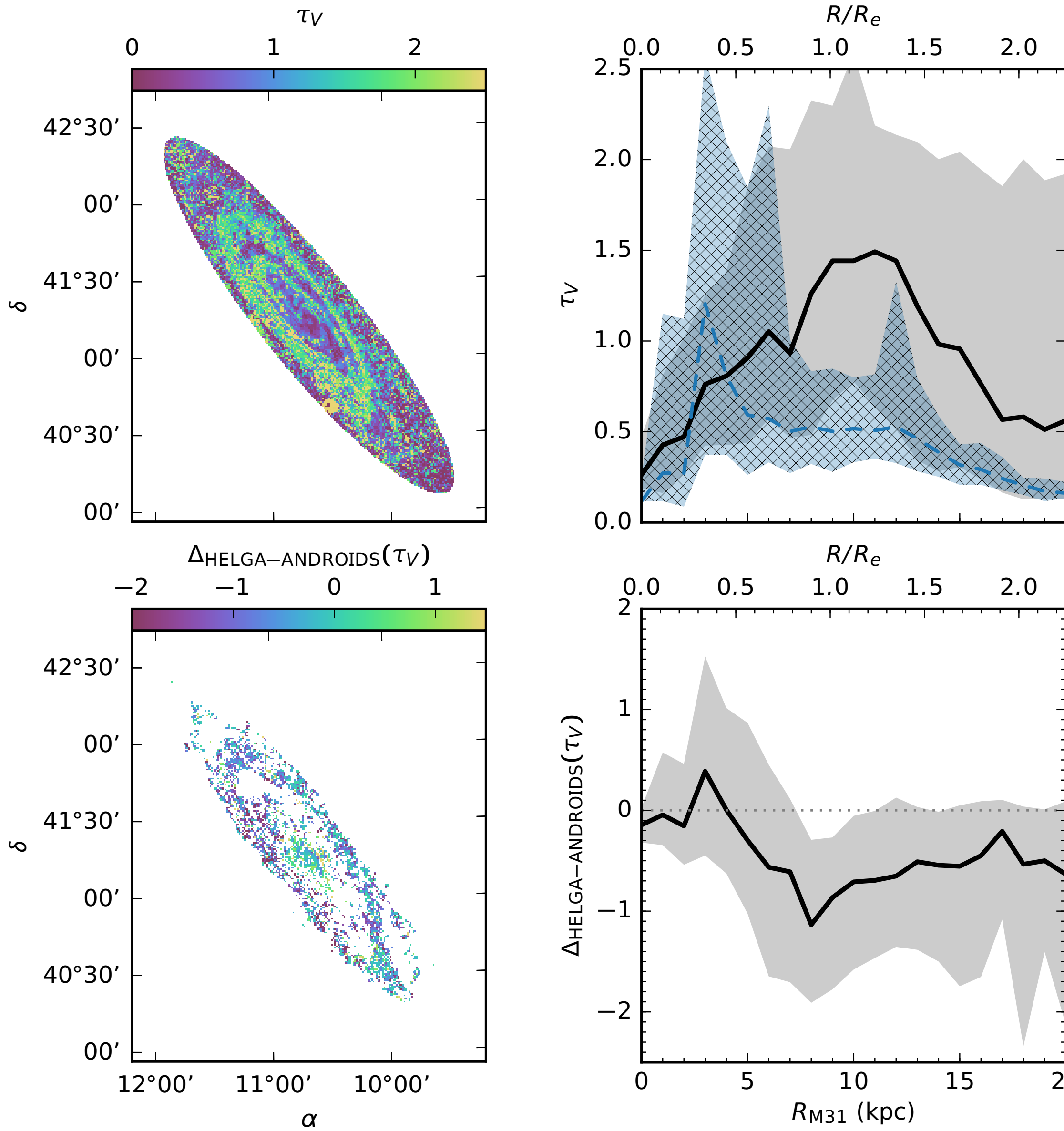


Figure 6.29: Total dust optical depth (τ_V) that combines ISM dust (Figure 6.28) and additional dust around birth clouds. *Top:* map and radial profile (solid black line) of τ_V estimated by MAGPHYS models of the full ANDROIDS SED. The dashed blue line is the HELGA IV profile. *Bottom:* map and radial profile of the difference between τ_V estimated by HELGA IV and ANDROIDS.

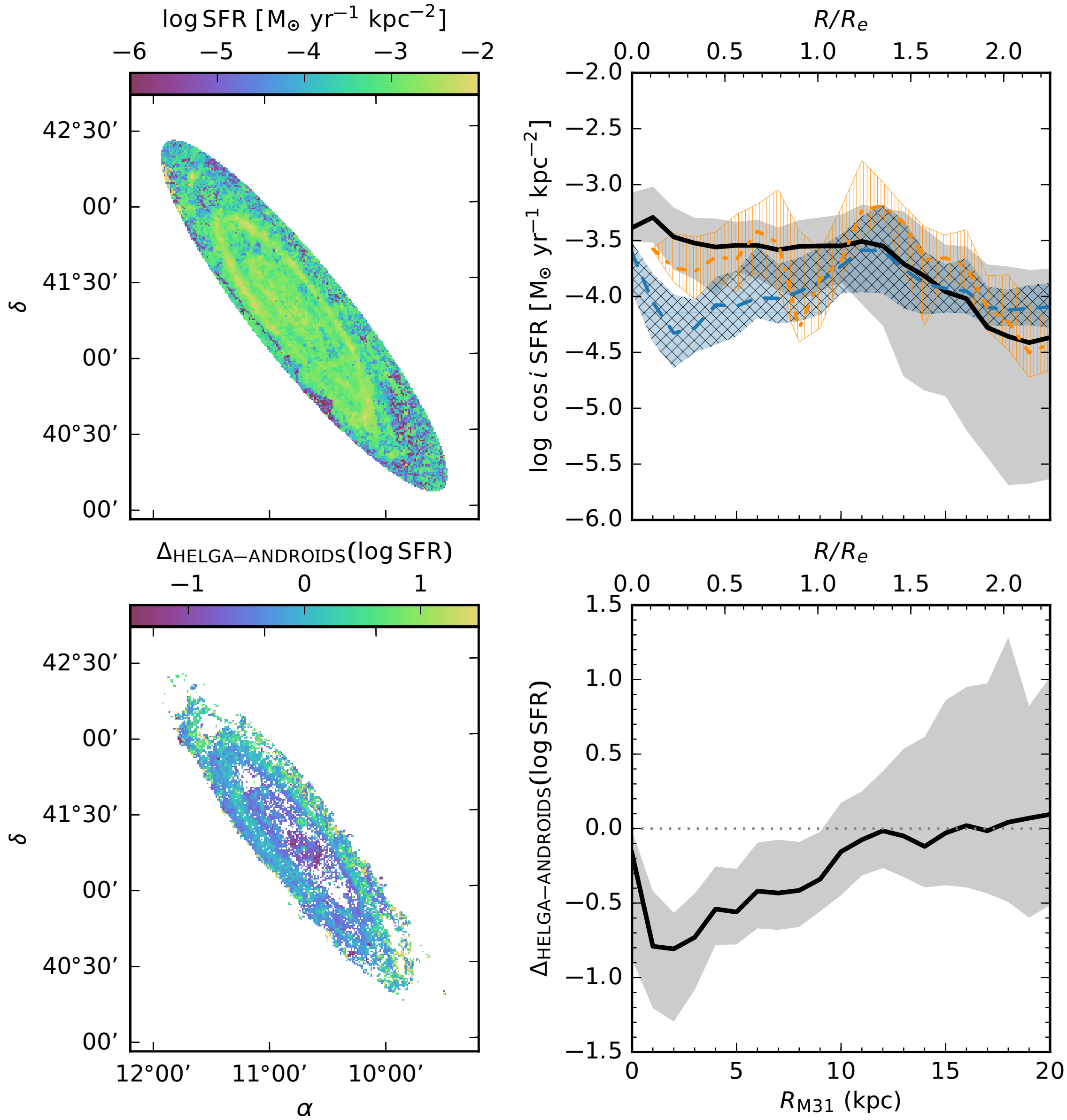


Figure 6.30: Star formation rates (SFR) modelled by ANDROIDS HELGA IV and Williams et al. (2017). The ANDROIDS and HELGA IV estimates are MAGPHYS's 100 Myr SFR estimate. The Williams et al. (2017) SFR is based on their youngest isochrone bin (0–300 Myr). *Top:* map and radial profile (solid black line) of SFR estimated by MAGPHYS models of the full ANDROIDS SED. The dashed blue line is the HELGA IV profile while the dash-dotted orange line is the Williams et al. (2017) profile estimated with PHAT resolved stellar populations. The profile is deprojected for M31's inclination given $\cos i = 0.213$. *Bottom:* map and radial profile of the difference between the SFR estimated by HELGA IV and ANDROIDS.

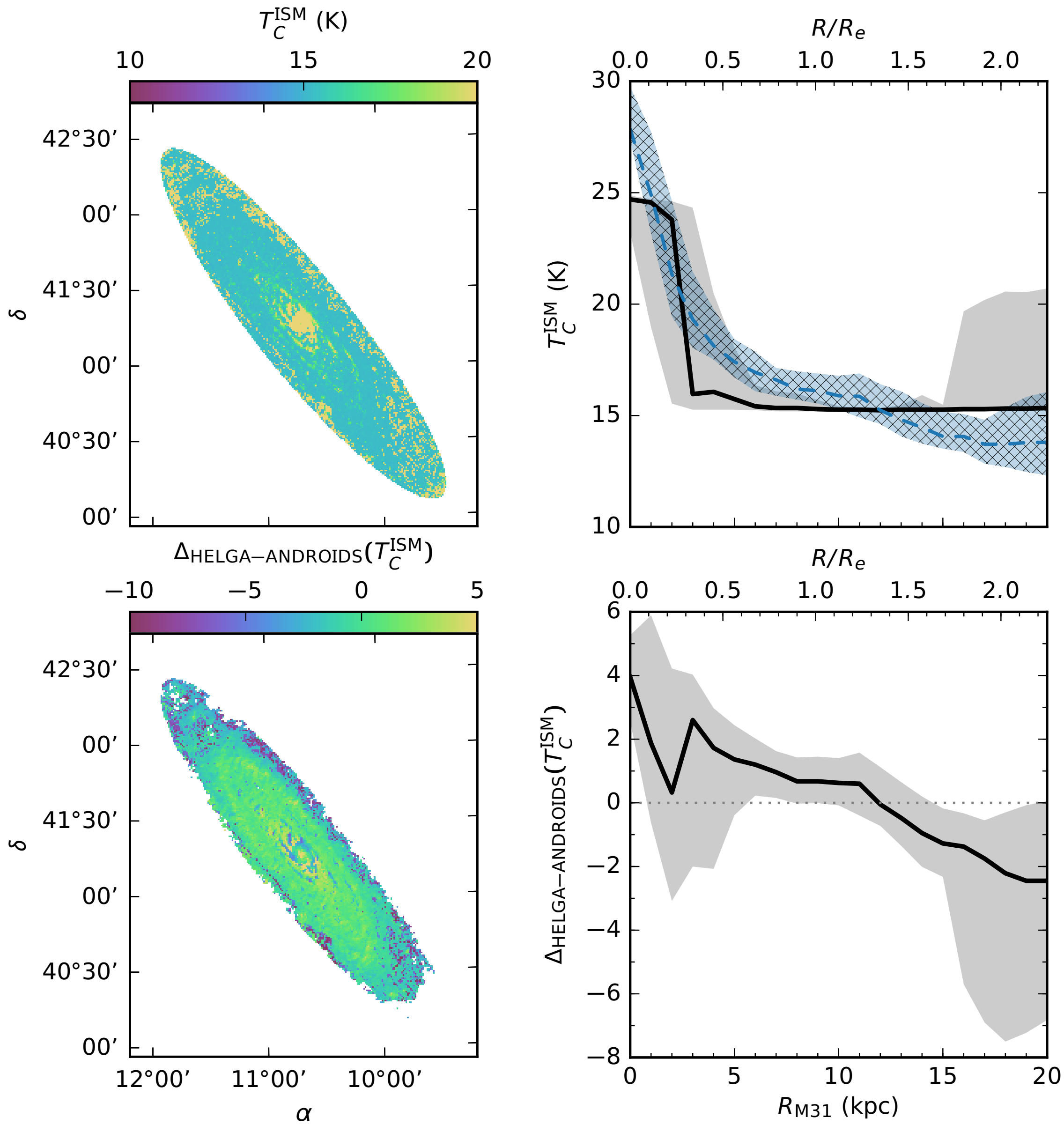


Figure 6.31: ISM dust temperature ($T_C^{\rm ISM}$) modelled by ANDROIDS and HELGA IV. *Top:* map and radial profile (solid black line) of $T_C^{\rm ISM}$ estimated by MAGPHYS models of the full ANDROIDS SED. The dashed blue line is the HELGA IV profile. *Bottom:* map and radial profile of the difference between the $T_C^{\rm ISM}$ estimated by HELGA IV and ANDROIDS.

is that the ANDROIDS modelling lacks the extended MAGPHYS dust model library used by HELGA IV to include both lower and higher ISM dust temperatures. Indeed, we see that our models are constrained by the 15 K lower temperature limit of the standard MAGPHYS dust library (Figure 6.31).

On the other hand, we show in Figure 6.30 that the ANDROIDS SFR is comparable to that measured by Williams et al. (2017) from CMD fitting of resolved stellar populations. With such independent confirmation, ANDROIDS confirmation of ISM and birth cloud dust optical depths may be accurate.

Chapter 7

Stellar Populations Inferred from Resolved Stellar Photometry

7.1 Introduction

Modelling distributions of stars in colour-magnitude diagram (CMD) planes is an effective technique for stellar population estimation. Lewis et al. (2015) considered CMD fitting a 'gold standard' in stellar population estimation since precise features in stellar evolution, such as the main sequence turn off or red clump, are more directly tied to the age, metallicity and attenuation attributes of a stellar population than their unresolved spectrum or SED. Therefore resolved stellar populations offer an ideal standard for assessing the accuracy of SED modelling. In this Chapter we describe the methods for modelling M31's stellar populations from CMD modelling, including an analysis of M31's star formation history.

7.2 The Panchromatic Hubble Andromeda Treasury (PHAT) Dataset

CMD modelling demands an accurate stellar photometric catalog with well-understood uncertainties and completeness. Although we have attempted to build a photometric catalog from ground based ANDROIDS imaging, we find it more practical at present to take advantage of the Panchromatic Hubble Andromeda Treasury (PHAT). Thanks to HST's intrinsic resolution, PHAT is far more effective for crowded photometry than ANDROIDS, resulting in a deeper and more accurate stellar catalog.

PHAT offers M31 stellar photometry in six bandpasses: F225W and F336W with WFC3/UV; F475W and F814W with ACS; F110W and F160W with WFC3/IR. Thus PHAT is an ideal complement to ANDROIDS since it covers roughly the same SED baseline, although it is slightly bluer in both the UV and NIR (equivalent to H-band rather than K_s) extremes. PHAT observed M31 in 23 bricks across the northwest third of the M31 disk, extending from the centre to roughly $R_{\mathrm{mag}} \sim 20$ kpc (see Figure 7.1). In this study we use the `V2` PHAT photometric catalog made available by Williams et al. (2014). Specifically, we use the brick catalogs where photometry from each instrument (WF3/UV, ACS, WF3/IR) has been pre-combined.

To assess photometric uncertainties and incompleteness we use the artificial star photometry catalog published by Williams et al. (2014). This artificial star catalog was produced for specific test fields by repetitively adding artificial stars to the observed image stack whose coordinates and SED were selected from a representative parameter space; the photometry pipeline was run, and the artificial star's recovered magnitudes, if any, were recorded. Thus artificial star testing empirically measures both completeness (probability a star would be observed) and uncertainties as a function of location on color-magnitude diagram planes. Since artificial star tests are computationally expensive (the photometry pipeline must effectively be re-run thousands of times), Williams et al. produced artificial star catalogs for only six PHAT fields with representative stellar densities, shown in Figure 7.1. For the work

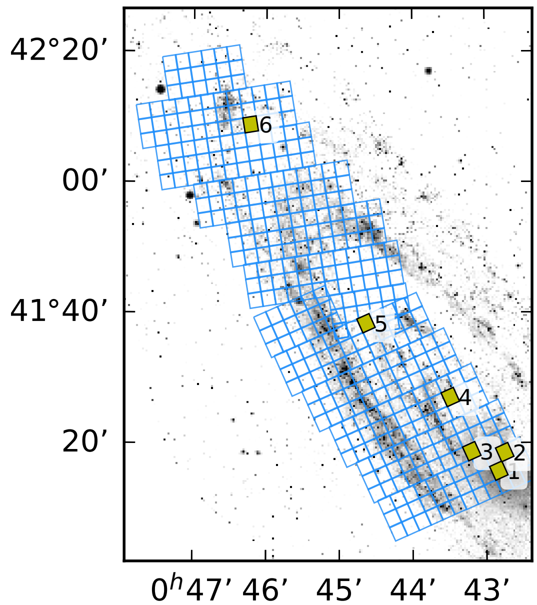


Figure 7.1: Phat fields on the NE sector of M31. Fields where Williams et al. (2014) performed artificial star tests are highlighted in yellow. A GALEX NUV map (Gil de Paz et al. 2007) is shown to spatial reference of young star forming regions.

below, we thus adopt artificial star catalogs from the field most akin to the field being modelled.

7.3 Modelling the PHAT Hess Planes with StarFISH

We adopt an optimization-based approach to resolved stellar population modelling where an observed Hess diagram is fit with a linear combination of synthetic Hess diagram planes, each one corresponding to a distinct population age and metallicity. A Hess diagram is a density plot of the distribution of objects (specifically, stars) on a two-dimensional colour-magnitude plane (axes can be either a magnitude or a colour, i.e., difference of magnitudes). Since synthetic Hess diagrams are normalized to a total stellar mass of 1 $M_\odot$, the fitted linear coefficients of each synthetic Hess diagram can be interpreted as the total mass of stars formed in each age and metallicity bin. Dividing this mass by the time duration that a given Hess plane corresponds to yields a star formation rate.

This approach for fitting linear combinations of Hess diagrams has been implemented in the *StarFISH* (Harris & Zaritsky 2001), *MATCH* (Dolphin 2002) and *IAC-STAR/IAC-Pop* (Aparicio & Gallart 2004; Aparicio & Hidalgo 2009) packages. Of these, only *StarFISH* is available under a free and open source license, and therefore this is the implementation we use.[1] *StarFISH* itself is a suite of Fortran programs, each requiring tedious and repetitive configuration files to be written. To enhance the usability of *StarFISH*, particularly within a fitting pipeline, we have built the *starfisher* python package[2] to represent *StarFISH* configuration, pipelines, and outputs in an object-oriented framework.

In the following sections we describe our specific process for simulating and fitting the PHAT star catalog; the next chapter will explore the application of this method.

7.3.1 Isochrones

An *isochrone* describes the effective temperatures and luminosities of a set of stars for a range of masses born at the same time, given parameter choices. With a bolometric correction library, these temperatures and luminosities are converted into observed magnitudes and colours. For this work we use the latest iteration of the PARSEC version `1.2S` isochrones, described in Bressan et al. (2012), Tang et al. (2014), and

[1] *StarFISH* may be downloaded from `http://www.noao.edu/staff/jharris/SFH/`, although it is also maintained under version control at `http://github.com/jonathansick/starfish`, with patches included.

[2] Available online at `https://github.com/jonathansick/starfisher`.

Chen et al. (2014). Note that these isochrones *exclude* the thermally-pulsating AGB stellar phase at the time of writing. The bolometric corrections are described in Chen et al. (2014) for normal stars (using `PHOENIX BT-Settl` atmospheres for very cool stars, and `ATLAS9` otherwise), and Aringer et al. (2009) for Carbon stars.

We obtain these isochrones through the `CMD 2.7` web site[3], using our `padova` package for programatically requesting and caching isochrone sets.[4] `CMD` provides isochrones in grids, and we request isochrones for specific metallicities (see below) in grids of $6.6 \leq \log(A~\mathrm{Gyr}^{-1}) \leq 10.13$ with spacing of $\Delta_{\log(A~\mathrm{yr}^{-1})} = 0.05$. This resolution provides an adequate coverage of isochrones on our Hess planes. With `padova` we splice together isochrones in the PHAT bandpasses in the `wfc3_wide` and `acs_wide` filter sets to match that PHAT photometry system. All isochrones and photometry are in the Vega system.

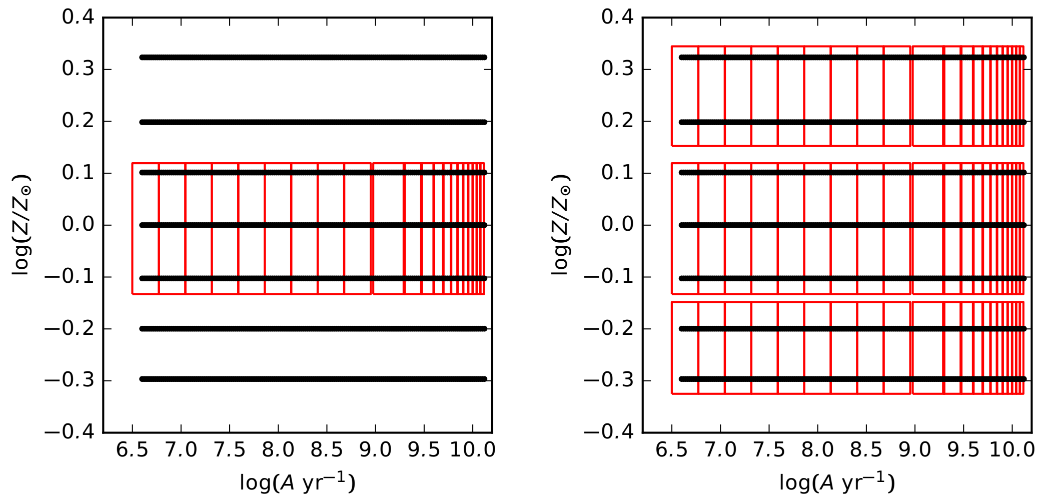
Finally, remember that population synthesis codes, such as *StarFISH*, use an initial mass function (IMF) to populate simulated stars along isochrones. *StarFISH* only permits IMFs described by a single exponential, which requires us to use a Salpeter (1955) IMF, where

$$\xi(m)\Delta m = \xi_0 \left(\frac{m}{M_\odot}\right)^{-2.35} \left(\frac{\Delta m}{M_\odot}\right). \tag{7.1}$$

Recall from § 6.2.1 that the MAGPHYS stellar population model library opted to use a Chabrier (2003) IMF since it is tuned to disk stellar populations. This Chabrier IMF is described by a log-normal distribution that *StarFISH* cannot compute. In practice, our use of Salpeter for CMD fitting is acceptable since the commonly used IMF prescriptions typically differ only for sub-$M_\odot$ stellar masses, with the Salpeter (1955)

[3] `http://stev.oapd.inaf.it/cgi-bin/cmd`

[4] Available online at `https://github.com/jonathansick/padova`.



(a) Fixed-solar metallicity isochrone binning (b) Three-metallicity isochrone lock file.

Figure 7.2: *StarFISH* isochrone lock file specifications. Each black marker (seen as dots that cannot be resolved) corresponds to an isochrone that is potentially incorporated into synthetic Hess diagrams. Red boxes, spanning age and metallicity, define each isochrone group. Synthetic Hess diagrams from isochrones within each box are combined into a single star formation rate amplitude when fitting an observed CMD.

IMF being 'bottom-heavy' in overpopulating low-mass stars. This difference primarily manifests itself in altering a stellar population's mass-to-light ratio; a Chabrier IMF yields a M/L ratio 1.8 times small than a Salpeter IMF (Chabrier 2003). IMFs will also manifest in slightly different occupation frequencies of post-main sequence tracks by old, low-mass stars, though we expect this effect to be small (otherwise it would be common practice to fit IMFs from CMD fitting).

7.3.2 Isochrone Locking

The *StarFISH* approach to fitting star formation histories was initially described as one of finding the sum of simple stellar population planes that reproduce the observed

Hess planes. Formally, a simple stellar population consists of a *single* isochrone specifying a singular age and metallicity. In practise, true SSP Hess planes are not useful because at the age and metallicity grid density required to continuously sample the Hess diagram space there would be far too many free parameters in the Downhill Simplex optimization. *StarFISH* addresses this issue with *lock files* that allow simulated Hess planes for a group of isochrones to be binned together, effectively reducing the number of star formation amplitudes that must be fit.

In our lock file implementation we bin isochrones in both age and metallicity space. For ages less than 1 Gyr, the bin width is $\Delta_{\log(A\ \mathrm{yr}^{-1})} = 0.25$. For ages older than 1 Gyr, isochrones are binned in linearly-spaced boxes 1 Gyr wide. This allows our Hess planes to capture the dynamic changes in young main sequence populations, while also covering the lifetime of the Universe, in just 21 age bins.

In metallicity, we use fewer bins. Our 'solar metallicity' bin consists of the set of $\log(Z/Z_\odot) = \{-0.1, 0.0, +0.1\}$ isochrone sets. We also have a sub-solar metallicity tracking consisting of the set $\log(Z/Z_\odot) = \{-0.3, -0.2\}$, and a super-solar metallicity set consisting of $\log(Z/Z_\odot) = \{+0.2, +0.3\}$. The latter is considered our 'three-metallicity' isochrone configuration. Both of these lock file configurations are visualized in Figure 7.2.

7.3.3 Simulated Hess Planes

For each set of isochrones grouped together via a lock file, *StarFISH* renders synthetic Hess planes. Although Hess planes can be colour-colour (e.g., F475W $-$ F814W vs F110W$-$F160W), we work in colour-magnitude Hess planes since the distance to M31 is well calibrated and the luminosity function provides precious information about the

Table 7.1: Hess planes available for fitting star formation histories.

Name	x mag	x range	y mag	y range	Comments
`ACS-MS`	F475W – F814W	$(-0.5, 1.0)$	F814W	$(26, 21)$	Main sequence in ACS bands; emulates Lewis et al. (2015).
`ACS-RGB`	F475W – F814W	$(1.2, 5.0)$	F814W	$(23.5, 20.0)$	Red giant branch and asymptotic giant branch sequence in ACS bands.
`ACS-ALL`	F475W – F814W	$(-1.0, 5.0)$	F814W	$(25.5, 20.0)$	Full Hess plane in ACS bands.
`OIR-ALL`	F475W – F160W	$(-0.8, 8.0)$	F160W	$(25.0, 17.5)$	Full Hess plane covering wide color baseline, approximately $V - H$.
`NIR-RGB`	F110W – F160W	$(0.3, 1.3)$	F106W	$(24.0, 16.5)$	WFC3/IR plane covering only RGB and AGB populations. Blue Core Helium Burning stars are masked.

stellar population. We prepared several Hess planes that use different combinations of bandpasses and limits to isolate, or not, different phases of stellar evolution. All planes are built with colour bin sizes of $\Delta m = 0.05$ mag, and are populated with 10×10^6 stars covering a mass range of 0.08 $M_\odot$ to 150 $M_\odot$. Table 7.1 specifies Hess planes used in fitting PHAT photometry.

We note that the `ACS-MS` plane, specified in Table 7.1 is intended to replicate the young main sequence stellar population fits performed by Lewis et al. (2015). Figure 7.4 shows the PARSEC 1.2S isochrones against the `ACS-MS` Hess plane. The `ACS-MS` plane is clearly limited to stars less than 1 Gyr old, with stellar population ages being determined predominantly by the luminosity of the main sequence turn-off.

The `ACS-ALL` plane retains the same F475W – F814W colour baseline as `ACS-MS`, but covers the entire visible stellar population, including red giant branch (RGB) and asymptotic giant branch (AGB) populations.

We also attempt to use the NIR photometry obtained with the WFC3/IR camera by PHAT. Lewis et al. report that the F160W photometry is 2 mag shallower than the ACS optical catalogs. In particular, F160W photometry loses faint blue stars on the main sequence. Nonetheless, using an expanded color F475W − F160W baseline (the `OIR-ALL` plane) may offer some resolution to degeneracies in extinction and metallicity in the temperature of the red giant branch. Figure 7.3 provides insight into the age and evolutionary stages of stars that occupy the `OIR-ALL` plane.

Finally, the `ACS-RGB` and `NIR-RGB` planes are presented as comparisons to the `ACS-MS` approach, where only the older stellar populations on the AGB and RGB are included.

7.3.4 Dust Extinction Distributions

StarFISH treats dust extinction as an input parameter, such that an assumed extinction distribution is used to generate synthetic Hess planes. It is only the linear coefficients of each synthetic Hess plane that are actually fit to estimate a star formation history.

In principle, the extinction distribution could be fit as a hyperparameter of the star formation history. As in the Gibbs sampler described in § 4.3, the dust distribution model space would be explored alternately with fitting the star formation history. With *StarFISH*, however, this is an extremely computationally intensive procedure since new synthetic Hess planes must be produced each round. Potentially the *StarFISH* code could be modified to allow different amplitudes of an extinction parameter by allowing the synthetic Hess planes to be translated along reddening vectors. Alternatively, hierarchical Bayesian stellar population fitting codes could more straightforwardly incorporate extinction distributions as a fitting parameter since they do not precompute Hess planes, but would work star-by-star.

Instead we follow a tack suggested by Lewis et al. (2015) and use far-infrared dust maps to construct extinction distributions for any arbitrary line of sight through M31. Specifically, we use the Draine et al. (2014) model of dust surface mass density at the SPIRE 350 resolution (FWHM = $24.9''$; $10''$ pixel^{-1}).[5] Draine et al. (2014) constructed this map by modelling the SED from Spitzer MIPS (Gordon et al. 2006) and Herschel PACS and SPIRE cameras (Groves et al. 2012) spanning 60 μm to 2.1 mm wavelengths. In their fits of PHAT main sequence stellar populations, Lewis et al. (2015) explored a grid of possible dust distribution to find to find a best-fitting dust model. They modelled extinction distributions through a disk as a uniform distribution of A_V values to individual stars, thus this grid search allowed them to find the maximum A_V associated with the uniform A_V distribution in each pixel. In their Figure 17, Lewis et al. (2015) plotted the ratio of maximum A_V from their grid search to Σ_{dust}, the dust surface mass density estimated by Draine et al. (2014). This ratio appears to be approximately Normally-distributed with a scatter of 2 dex at low star formation rates, but a clear mean value of

$$\left\langle \log \frac{A_{V,\mathrm{max}}}{\Sigma_{\mathrm{dust}}} \right\rangle = -5.4. \tag{7.2}$$

Thus for any field of M31, we set $A_{V,\mathrm{max}}$ from the mean value of the dust surface mass density in the field footprint. Meanwhile, the minimum extinction in the field is

[5]Available at `http://www.astro.princeton.edu/~draine/m31dust/M31_S350_110_SSS_110_Model_All_SurfBr_Mdust.fits.gz`.

set by the Milky Way foreground, $A_{V,\mathrm{min}} = 0.17$ (Schlafly & Finkbeiner 2011). From $A_{V,\mathrm{min}}$ and $A_{V,\mathrm{max}}$ we generate a uniform distribution of extinction values. Along with extinction ratios, A_X/A_V for the PHAT bandpasses (Schlafly & Finkbeiner 2011, Table 6), *StarFISH* generates synthetic Hess diagrams that are both reddened and broadened according to this extinction model.

This method of treating extinction *is* simplistic and relies heavily on the relation found by Lewis et al. (2015) between their best-fit $A_{V,\mathrm{max}}$ and dust mass maps. Dust maps constructed from Bayesian fits to individual star SEDs will provide more realistic dust treatments. In the meantime, this is a reasonable technique that adds no additional degrees of freedom to the modelling.

7.3.5 Interpretation of StarFISH Fits

The result of a *StarFISH* fit is an array of coefficients corresponding to the number of simulated stars associated with each lockfile bin: $n_{*,i}$. Following examples in *StarFISH* reduction scripts (Harris & Zaritsky 2001), $n_{*,i}$ can be converted into a stellar mass by multiplying by the mean stellar mass for an initial mass function, $\langle M_{IMF} \rangle$. Again, *StarFISH* requires that we use a Salpeter (1955) IMF, so that $\langle M_{\mathrm{IMF}} \rangle = 1.68\ \mathrm{M}_{\odot}$. *StarFISH* also estimates a 1-σ confidence interval, $[n_{*,i}^{-\sigma}, n_{*,i}^{+\sigma}]$, in each amplitude.

Thus we obtain a stellar mass and uncertainty for each fitted lockfile bin as

$$m_i = \langle M_{\mathrm{IMF}} \rangle n_{*,i}, \tag{7.3}$$

$$\sigma_{m,i}^{-} = \langle M_{\mathrm{IMF}} \rangle (n_{*,i} - n_{*,i}^{-\sigma}), \tag{7.4}$$

$$\sigma_{m,i}^{+} = \langle M_{\mathrm{IMF}} \rangle (n_{*,i}^{+\sigma} - n_{*,i}). \tag{7.5}$$

The star formation rate corresponding to each lockfile bin is derived from the bin's time duration:

$$\mathrm{SFR}(t_i) = \frac{m_i}{\Delta t_i}, \tag{7.6}$$

$$\sigma^-_{\mathrm{SFR}}(t_i) = \frac{\sigma^-_i}{\Delta t_i}, \tag{7.7}$$

$$\sigma^+_{\mathrm{SFR}}(t_i) = \frac{\sigma^+_i}{\Delta t_i}. \tag{7.8}$$

Mean Age of StarFISH Fits

The mean age for a stellar population fitted by *StarFISH* is then the mean of each lockfile bin's age, weighted by the bin's mass:

$$\langle A \rangle = \frac{\sum m_i A_i}{\sum m_i}, \tag{7.9}$$

where m_i is the stellar mass associated with each SSP of age A_i.

To estimate the uncertainty of $\langle A \rangle$, we use bootstrap resampling. In this algorithm, we draw a new set of resampled mass weights, $m_i' = m_i + N(0, \sigma_{m,i})$ to compute a resampled $\langle A \rangle'$. The uncertainty of $\langle A \rangle$ is estimated directly from the sample standard deviation of the resampled mean ages.

7.4 Exploratory Fits to Brick 23

To test the extent to which M31's stellar populations can be fit with *StarFISH*, and thereby provide a fiducial by which to compare SED-fit stellar populations, we begin with an exploration of the simplest PHAT dataset: Brick 23. Brick 23 is PHAT's outermost field, lying at $R_{\mathrm{maj}} = 20$ kpc on the northern M31 major axis. As such, Brick 23 has the least crowding, and therefore best photometry, of any PHAT brick.

A key motivation for this experiment with Brick 23 is to determine which Hess planes are most useful for fitting star formation histories across M31. Lewis et al. (2015) fit only stars in the optical main sequence, but a full star formation history can only be inferred by incorporating stellar phases representing all ages of stars. In this section we determine the viability of such modelling by performing fits against all Hess planes specified in Table 7.1 and described in § 7.3.3.

As well, we must determine what metallicity treatment is necessary for modelling M31's disk stellar population. In this section we perform fits with both fixed solar metallicity isochrone sets, and with a three-metallicity isochrone set (Figure 7.2(b); § 7.3.2).

7.4.1 Fits with a Fixed Solar Metallicity Track

Fitting a Hess plane with a single metallicity track is advantageous in reducing the number of free fitting parameters, but concedes flexibility in explaining the observed population of stars. In Figure 7.5 and Figure 7.6 we show an overview of fits and predictions with a fixed solar metallicity isochrone set for different Hess planes. Unfortunately, *StarFISH* could not converge on a solution while fitting the `ACS-MS` Hess plane. This defeats a direct comparison of the present experiment to Lewis et al. (2015), who also used a similar metallicity configuration for fitting PHAT main sequence populations. This failure to fit the `ACS-MS` plane is possibly due to a star formation history likelihood surface that is too flat so that there is no decisive direction for *StarFISH's* downhill simplex algorithm to follow. Indeed, PHAT observations of the upper main sequence do not reveal a dominant main sequence turn-off feature, that would, for example, defeat degeneracies in a star formation history.

We measure goodness-of-fit with both a per-Hess-pixel χ^2 metric given observed (H_i) and modelled ($\hat{H}_i$) Hess pixel values:

$$\chi_i^2 = \left(\frac{H_i - \hat{H}_i}{\sqrt{H_i}} \right)^2, \tag{7.10}$$

and a reduced χ_r^2 metric across the entire plane:

$$\chi_\mathrm{r}^2 = \frac{\sum_{i=1}^{N} \chi_i^2}{N - N_\mathrm{amp}} \tag{7.11}$$

where N_amp is the number of fitted synthetic Hess planes and N is the number of Hess pixels. Uncertainties in the observed Hess diagram are assumed to be purely due to Poisson sampling of stars populating each Hess pixel.

Given this metric, the `ACS-RGB` plane is most able to fit itself (with a reduced $\chi_\mathrm{r}^2 = 4.6$). The RGB populations observed exclusively by WFC3/IR imaging alone, `NIR-RGB`, have a much worse fit, $\chi_\mathrm{r}^2 = 36.1$. In the fit residual plane, Figure 7.6, the model is redder by approximately F110W − F160W ~ 0.2 mag, while the luminosity function at F160W < 22 mag is more strongly populated in the observations than the models.

In both the optical-only and optical-NIR color baselines, we see that the upper main sequence is reliably fit, with the models becoming too under-populated at the fainter end of the MS luminosity function, F814W < 23.5 mag and F160W < 24 mag. For giant branch stars, the models are consistently bluer by ~ 0.5 mag, with the model population becoming too strong at the fainter luminosities. Most importantly, the red clump is too weak in the model. The *AGB bump* at $F160W \sim 22$ mag, however, does occur at the same luminosity in both the model and observations, albeit with differing colours.

7.4.2 Fits with a Three-Metallicity Model

In addition to the fixed-solar metallicity isochrone set, we also fit Hess planes with a three-metallicity isochrone set (§ 7.3.2). χ^2 and residual difference planes are shown in Figure 7.7 and Figure 7.8. This configuration triples the number of degrees of freedom in the models, potentially allowing more flexibility in matching the observations' colours. Despite this additional complexity, *StarFISH* was able to converge to optimal fits.

Compared to the fixed-solar metallicity fits, the three-metallicity fits generally have smaller χ^2_{r} values. Nevertheless, the fitted planes still suffer from the same issues described for the fixed-solar metallicity fits, including modelled giant branch sequences that are too blue. Considering this, the three-metallicity fits have smaller χ^2_{r} values than their fixed-solar counterparts not because the correct metallicity can be determined, but because the additional metallicity tracks allow for broader giant branch colours and additional degrees of freedom.

7.4.3 Predictive Capability of Different Hess Planes

Given the fits to individual Hess planes, shown as the highlighted blue panels in Figures 7.5–7.8, it is also interesting to consider how those star formation histories predict other Hess planes.

Consider, for example the fit to `ACS-MS` shown in Figure 7.7. That main sequence stellar population fit is able to predict the main sequence in the `OIR-ALL` plane, but in all cases, severely over-predicts the older stellar populations on the red giant branch.

Fits to both `ACS-ALL` and `OIR-ALL` successfully predict the RGB in `ACS-RGB`, but *do not* predict the `NIR-RGB` plane. This suggests a fundamental issue with fitting the red giant branch with near-IR isochrones. Both the `ACS-ALL` and `OIR-ALL` fits tend to predict a main sequence that is too red. Ignoring the giant branches, as Lewis et al. (2015) did, is thus a reasonable choice if one is willing to consider only recent star formation.

7.4.4 Dependence of Star Formation History Estimates on Fitting Method

The true test of a stellar population fit is not the degree to which Hess planes can be reproduced, but how well and consistently the star formation history is estimated. In Figure 7.9 we see fitted star formation histories from each Hess plane, for both the fixed-solar and three-metallicity isochrone sets.

Among these fits, we find considerable diversity in estimates for very young populations (younger than 100 Myr) and old populations (older than 1.5 Gyr). Partially, this inconsistency is due to the differing sensitivities of each stellar population plane. Fits of only the main sequence (`ACS-MS`) tend to have low recent star formation rates and high older star formation rates, and vice versa for fits to only the red giant branch (`ACS-RGB`, `NIR-RGB`). The `ACS-ALL` and `OIR-ALL` fits to the entire observable stellar population are both consistent with each other, and present a much smoother star formation history in Brick 23 where the star formation rate has only declined by 1 dex over the last 10^9 years.

All fits are consistent in their estimates of star formation rates in the intermediate regime: $8 < \log(t/\mathrm{yr}) < 9.2$. In this regime (see Figure 7.3) stars are occupying the faint main sequence, which is better populated in the Brick 23 PHAT dataset. Additionally, this population includes Early-AGB stars that dominate the bright-end of the red giant luminosity function. Thus the well-populated faint main sequence and the bright post-main sequence are readily fit by current models with the present data. The populations most difficult to fit, as mentioned before, include older stars on the red giant branch, particularly in the red clump, and stars younger than 100 Myr on the bright main sequence.

From Fig. 7.9(b), we see that the `ACS-ALL` and `OIR-ALL` star formation histories are $\sim$ 0.5–1 Gyr older for three-metallicity models than the fixed solar metallicity models. Hence there is a slight covariance of age and metallicity results that depends on modelling. This result does not imply that the three-metallicity models are true estimates of stellar metallicity. In Figure 7.10, star formation history fits with the three-metallicity isochrone set are shown decomposed into individual metallicity groups. The stochasticity of decomposed star formation histories in these plots does not readily reveal chemical evolution within M31. The usefulness of the three-metallicity isochrone set, then, is in providing additional flexibility, but this flexibility should be marginalized before interpretation (*i.e.,* Figure 7.9(b) is a fairer description of star formation in Brick 23 than Figure 7.10).

7.4.5 Ideal Hess Fitting Planes and Metallicity Configurations

From these attempts to fit the PHAT Brick 23 photometric catalog, we find that no Hess plane can both be fit without systematic residuals, while also being sensitive to stellar populations across the full history of M31. While young main sequence

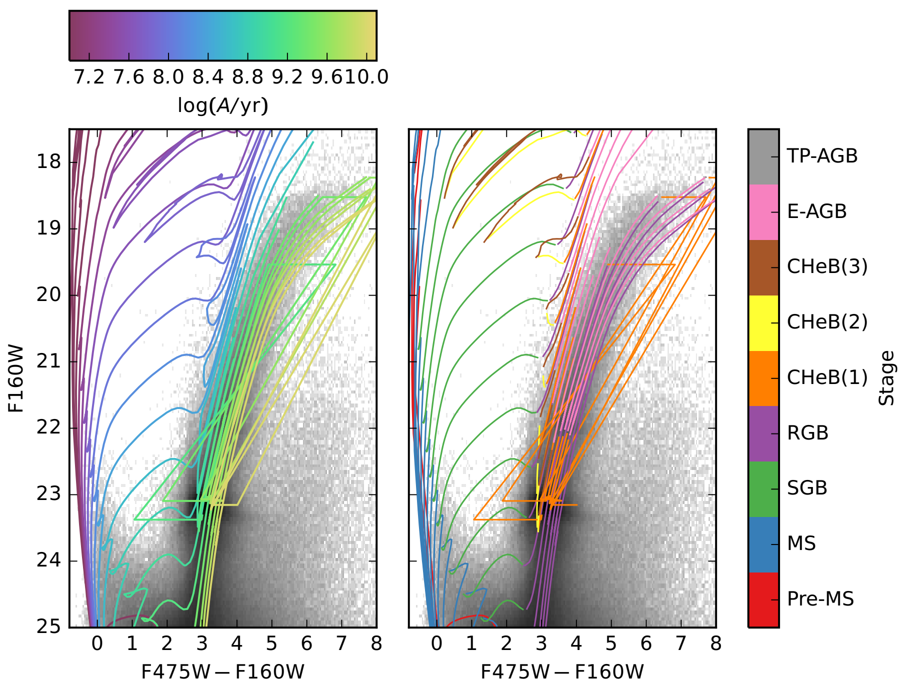


Figure 7.3: Solar-metallicity PARSEC 1.2S isochrones (§ 7.3.2) spanning a grid of ages superimposed on PHAT Brick 23 photometry in the `OIR-ALL` CMD plane. Left: isochrones labelled by age. Right: isochrones are labelled with phases of stellar evolution (MS: Main Sequence; SGB: Sub-Giant Branch; RGB: Red Giant Branch; CHeB: Core Helium Burning, split into three phases; E-AGB: Early Asymptotic Giant Branch); TP-AGB: Thermally Pulsating Asymptotic Giant Branch (not present in PARSEC 1.2S isochrones).

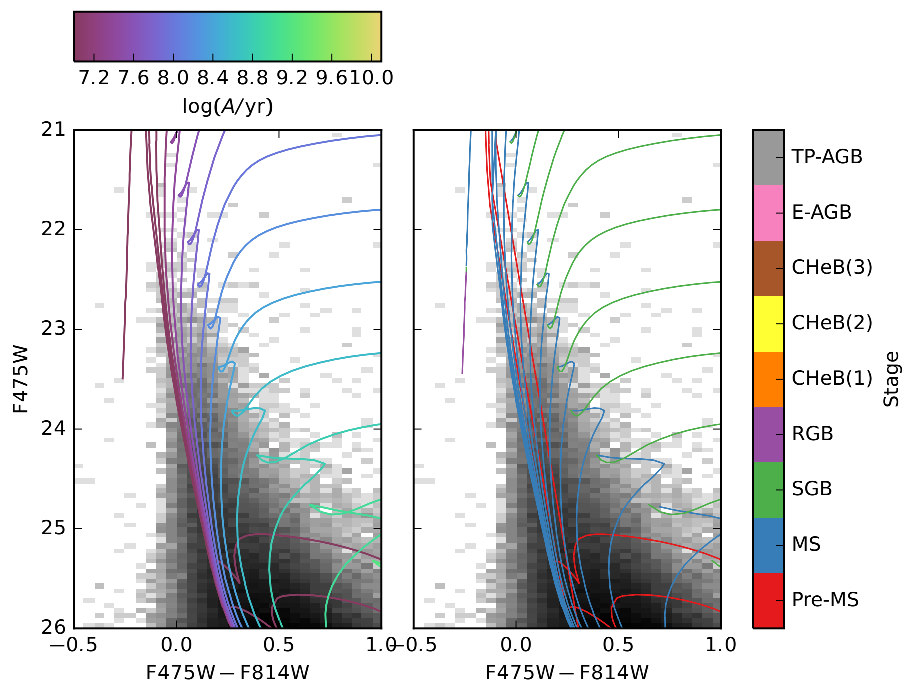


Figure 7.4: Same as Figure 7.3, though in the `ACS-MS` plane.

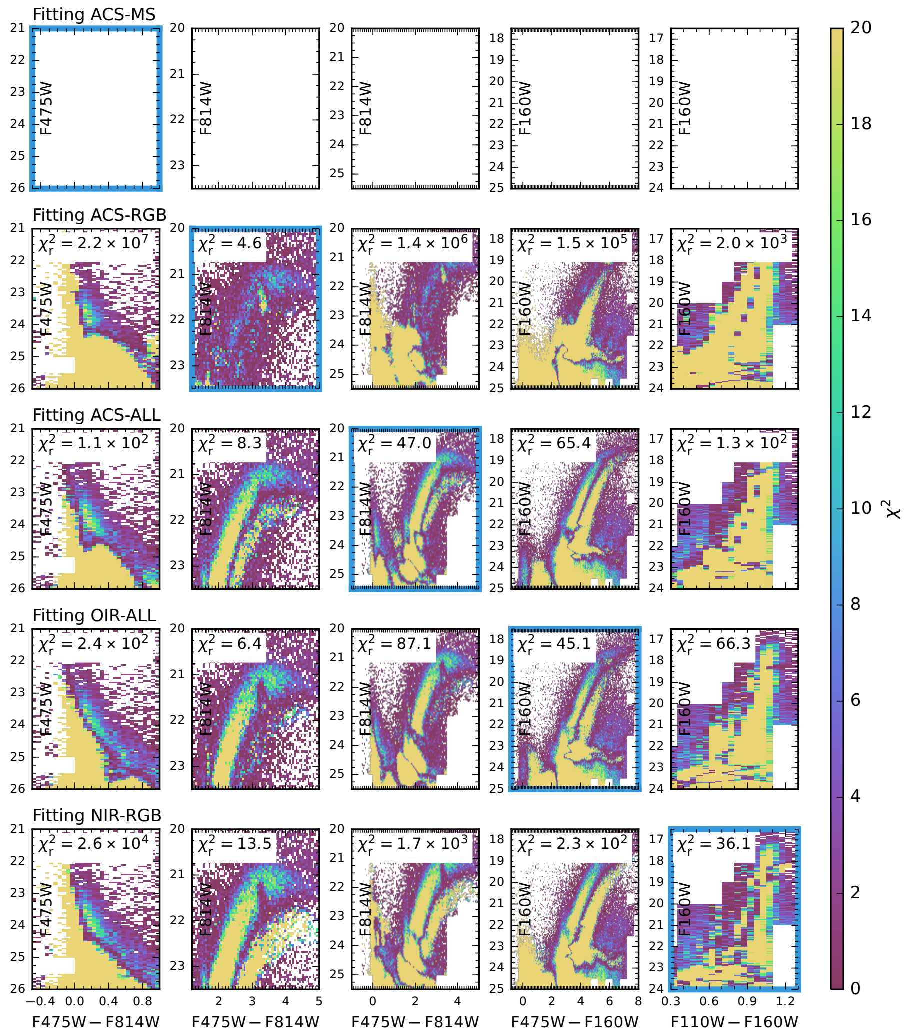


Figure 7.5: Hess diagrams of χ^2 residuals for fixed solar metallicity *StarFISH* models of PHAT Brick 23 photometry, as described in § 7.4.1. Five different Hess planes were fit, corresponding to the five rows: `ACS-MS`, `ACS-RGB`, `ACS-ALL`, `OIR-ALL`, and `NIR-RGB`. In each row the fitted Hess plane is highlighted with a blue border. The other columns are χ^2 residuals of the *predicted* Hess diagram given the row's fitted star formation history. In each panel, the fitted or predicted reduced χ^2_r is supplied, ith Hess pixels coloured by a per-pixel χ^2 statistic. Empty or masked pixels are white. Results for the `ACS-MS` fit are absent since *StarFISH* could not achieve convergence. Although fitting the entire stellar population is difficult, it is still necessary for fits that are representative of the whole star formation history (*e.g.*, the `ACS-RGB` prediction of `ACS-MS`). `OIR-ALL` presents an ideal Hess plane for fitting both itself, and predicting features in other Hess planes. Note that the `ACS-MS` plane could not be successfully fit with *StarFISH*.

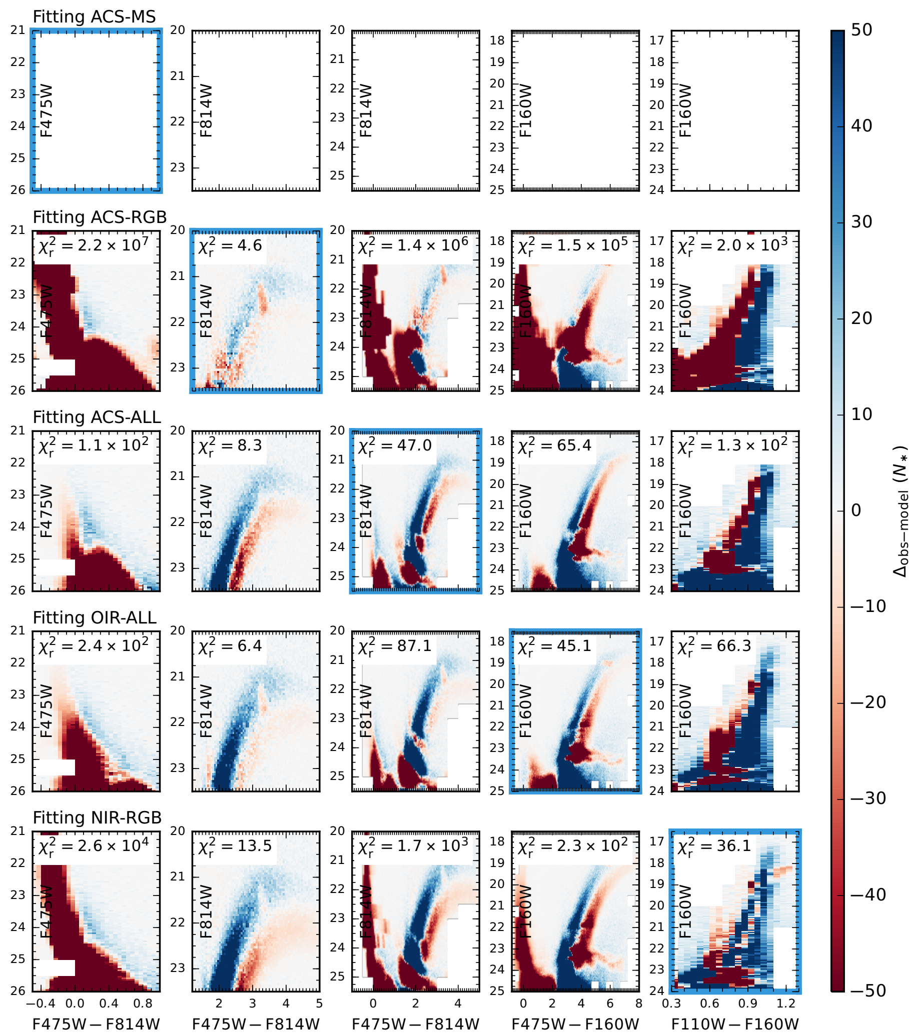


Figure 7.6: Hess diagrams of observed-model residuals for *StarFISH* models of PHAT Brick 23 photometry with a fixed solar metallicity, analogous to Figure 7.5. Differences are computed in units of stars per Hess pixel (0.05 mag^2) for the entire Brick 23 catalog. Note that the `ACS-MS` plane could not be successfully fit with *StarFISH*.

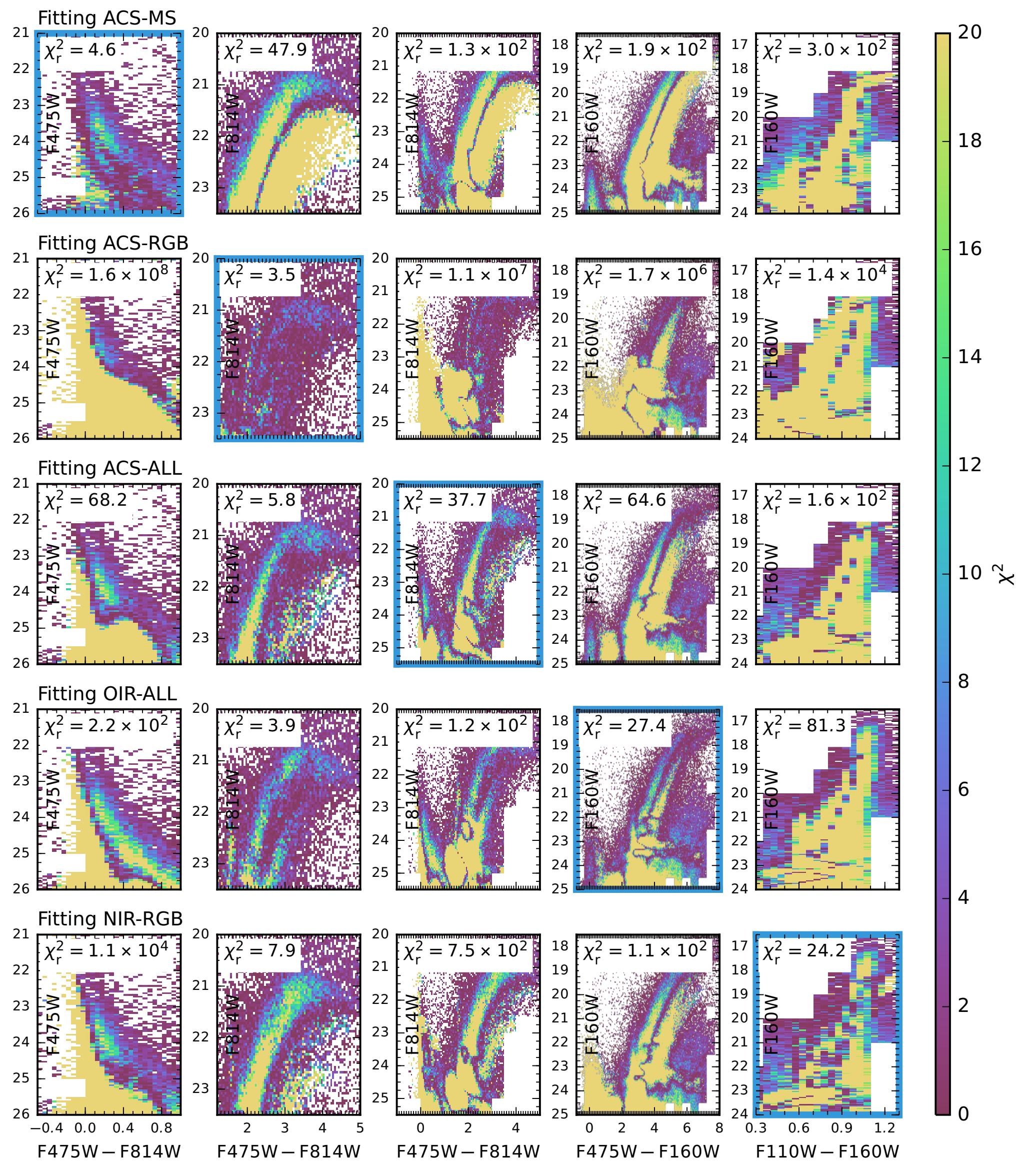


Figure 7.7: Hess diagrams of χ^2 residuals for *StarFISH* models of PHAT Brick 23 photometry with three-metallicity isochrone sets, as described in § 7.4.2. The Figure's mechanics are otherwise described in the caption of Figure 7.5. As in the fixed solar metallicity experiment (Figure 7.5), `OIR-ALL` is an ideal plane for fitting full star formation histories. The χ^2_r statistics are lower using the three-metallicity model, suggesting it is a more viable model.

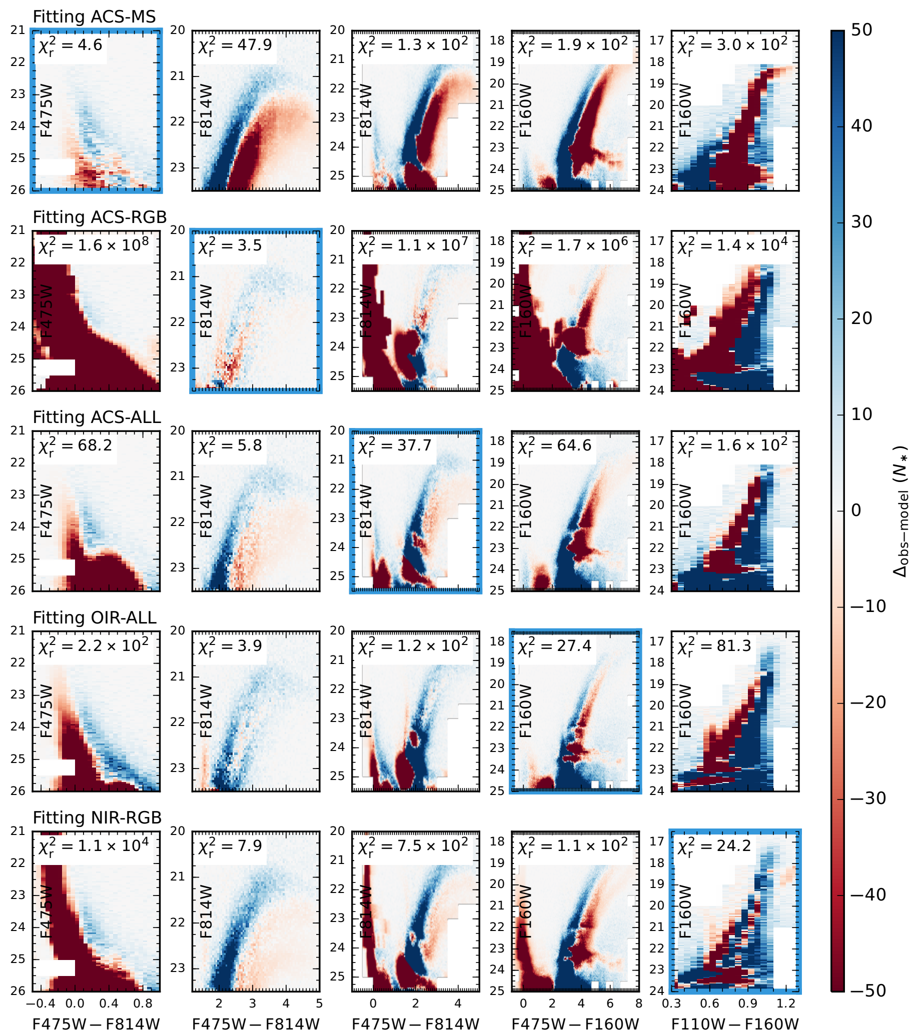
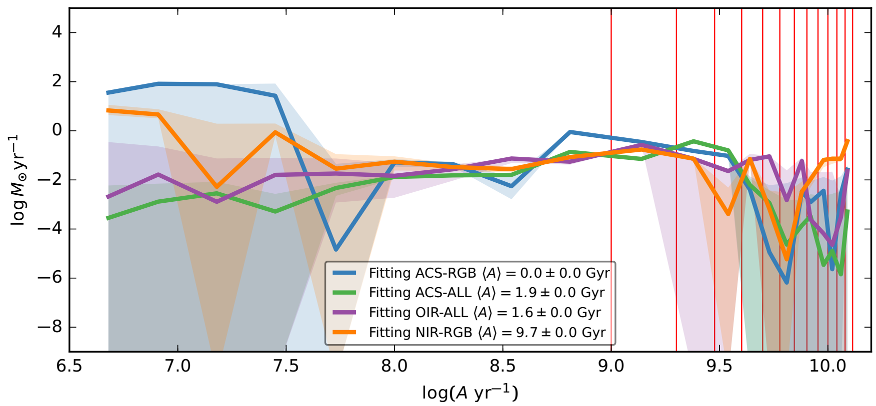
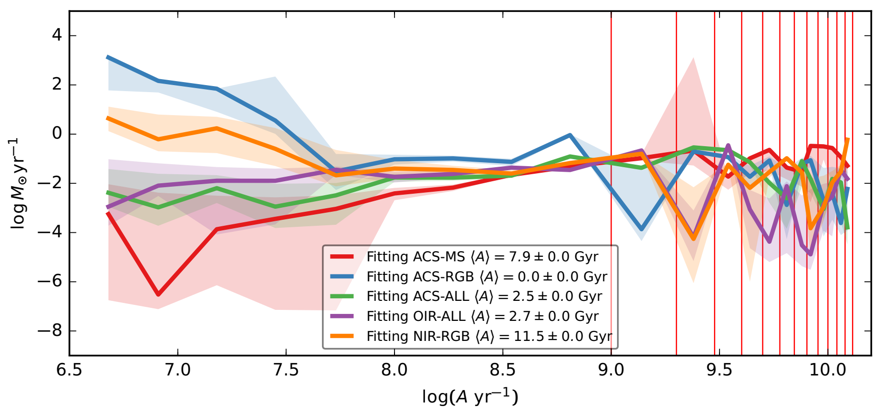


Figure 7.8: Hess diagrams of observed-model residuals for *StarFISH* models of PHAT Brick 23 photometry with three-metallicity isochrone sets, analogous to Figure 7.7. Differences are computed in units of stars per Hess pixel (0.05 mag^2) for the entire Brick 23 catalog.

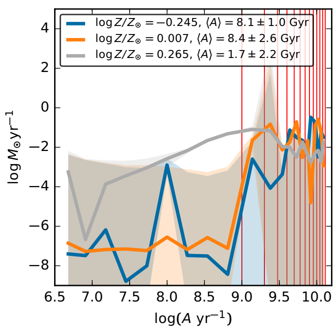


(a) Fixed Solar Metallicity Isochrone Set.

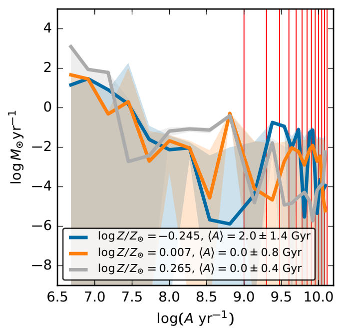


(b) Three-Metallicity Isochrone Set.

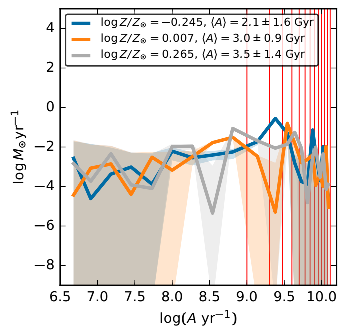
Figure 7.9: Star formation histories in PHAT Brick 23 with Fixed Solar and Three Metallicity Isochrone Sets for `ACS-MS`, `ACS-ALL`, `OIR-ALL`, and `NIR-RGB` fitting planes. Star formation rates are cumulative across the entire brick. Vertical red lines mark multiples of 1 Gyr. For (b), star formation rates from each metallicity track have been marginalized to produce a single star formation history; see Figure 7.10 for SFH as a function of metallicity.



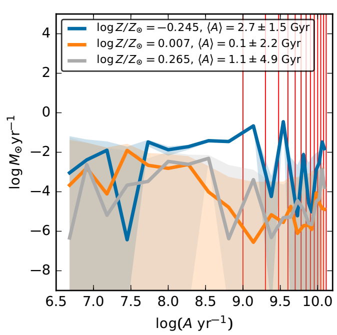
(a) ACS-MS.



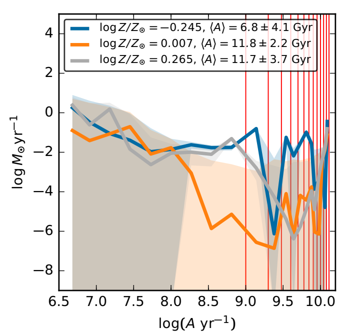
(b) ACS-RGB.



(c) ACS-ALL.



(d) OIR-ALL.



(e) NIR-RGB.

Figure 7.10: Star formation histories in PHAT Brick 23 decomposed into three metallicity components (see § 7.3.2) for ACS-MS, ACS-RGB, ACS-ALL, OIR-ALL, and NIR-RGB fitting planes. Star formation rates are cumulative across the entire brick. Vertical red lines mark multiples of 1 Gyr.

populations *can* be fit in isolation (Figure 7.7), both the main sequence and post-main sequence populations cannot be simultaneously fit. This result mirrors the experience reported by Lewis et al. (2015), although they did not explicitly show their attempts to fit post main sequence populations.

Since our goal in this project is to compare stellar population estimates from both unresolved SEDs and resolved star catalogs, we are compelled to make Hess fits that are sensitive to M31's full star formation history. That is, we must use either the `ACS-ALL` or `OIR-ALL` planes. Considering the χ^2_{r} quality of fits and predictions, we find that `OIR-ALL` combined with a flexible three-metallicity isochrone set is an ideal compromise.

7.5 Assessment of Star Formation History Estimation with Mock Populations

In the previous section we tuned a method for resolved stellar population fitting of PHAT photometry with *StarFISH*. One of the key decisions was the choice of fitting plane, and we found that fitting all visible stellar phases is necessary for understanding the full star formation history. However, the modelling technology is also fundamentally limited in reproducing the observed CMDs.

This limitation arises from two sources (likely simultaneously): flaws in stellar population modelling, and flaws in the star formation history fitting algorithm. In this section we study the latter issue in isolation by producing and fitting *mock* stellar populations. Mock tests determine how well *StarFISH* is able to use stellar population models to fit a CMD produced by those same models. Effectively, these mock tests develop an upper-limit on the accuracy of star formation histories estimated with

Table 7.2: PHAT artificial star test fields used for mock testing. See Figure 7.1 for locations of the fields relative to the GALEX map of M31's star forming regions.

Number	Brick	Field	R.A.	Dec.
1	1	10	0:42:50	41:15:40
2	1	5	0:42:45	41:18:31
3	3	15	0:43:12	41:18:42
4	5	10	0:43:29	41:27:00
5	9	2	0:44:38	41:38:22
6	21	15	0:46:13	42:08:48

StarFISH.

Specifically, we develop three experiments to understand how photometric uncertainties and the intrinsic star formation history affect *StarFISH's* ability to accurately recover a star formation history. First, in § 7.5.1 we use realistic PHAT photometric uncertainties from artificial star experiments in six fields to generate mock datasets with simple stellar populations. Such an experiment can show how PHAT uncertainties, incompleteness, and our choices of fitting planes impact the recovered star formation history. We evolve this experiment in § 7.5.2 by fitting mock photometry of simple stellar populations that is perfectly complete and without error, and now assess the accuracy of star formation histories fit from photometric catalogs far superior to those available. Finally, in § 7.5.3 we study mock exponentially declining star formation histories to understand how *StarFISH* recovers continuous star formation histories.

7.5.1 Mock Tests of Simple Stellar Populations with Realistic Uncertainties and Incompleteness

We perform separate mock star formation tests with uncertainty and incompleteness measurements corresponding to each of the artificial star testing fields that Williams et al. (2014) studied. See Table 7.2 and Figure 7.1 for reference. PHAT AST-6 corresponds to the outer disk, outside any major star forming regions. AST-5 is in a region of higher crowding, but also outside major star forming regions. AST-4 lies on a star forming region at $R_{\mathrm{maj}} = 5$ kpc, and AST-1 – AST-3 lie on the M31 bulge.

For this first series of tests we generate mock simple stellar populations (SSPs). Mock tests with SSPs are straightforward to interpret since the mock and modelled ages can be directly compared. We produced mock SSPs with ages that correspond to ages of isochrone groups: 0.05, 0.10, 0.18, 0.35, 0.65, 1.4, 3.5, 5.4, and 9.5 Gyr (see § 7.3.2). In each case, the SSP has solar metallicity. Since we assume, rather than fit, extinction, the mock population is extincted according to the model described in § 7.3.4. Thus there is inherently no mis-match in the extinction of the mock and model Hess planes. If the observed population is extincted with a dust distribution *different* from our assumed model, then the systematic errors described in these mock trials will only be a lower limit.

The mean age of the fitted SFH is an excellent measure of *StarFISH's* accuracy since SSPs are defined by their singular, well-defined age. Figure 7.11 shows how reliably the ages of mock SSPs, fitted with `ACS-MS` and `OIR-ALL` planes and in different AST fields, can be predicted as a function of their age.

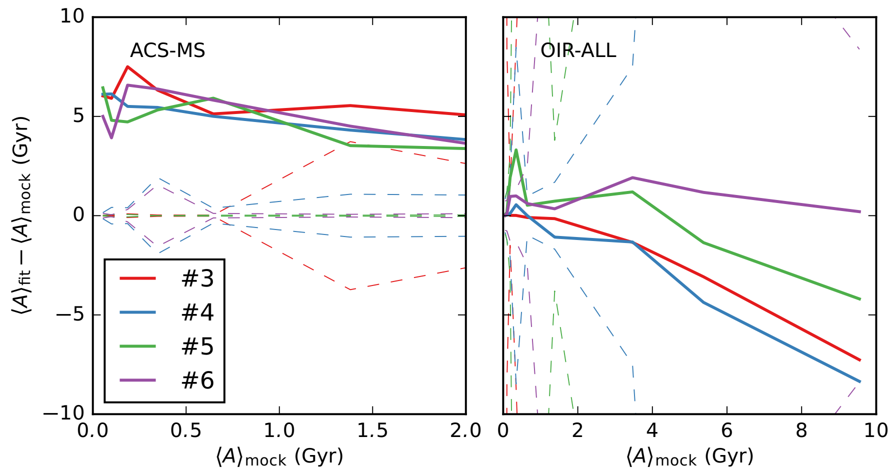


Figure 7.11: Mock SSP age estimation accuracy in the `ACS-MS` (left) and `OIR-ALL` (right) planes. SSP age error is plotted against the mock SSP age. The `ACS-MS` age axis is truncated since the `ACS-MS` plane inherently has no predictive power at ages older than 1 Gyr (see Figure 7.4). Mean age errors for each AST field (Table 7.2) are shown as solid lines, while dashed lines show the uncertainty of the age estimate from bootstrap resampling.

Poor Fits in the ACS-MS Plane

The most striking result is the poorness of `ACS-MS`-based CMD fits at all ages. We would expect the `ACS-MS` fits to break down at ages of 1 Gyr or older when the `ACS-MS` plane is devoid of stellar population information (Figure 7.4). Instead, we see that `ACS-MS` fits are biased towards *older* ages by 5 Gyr for all tested mock SSP ages. Note that the poorness of these fits *is not* reflected in Hess fitting residuals. For example, Figure 7.12 shows the mock CMD, fitted CMD, and residual CMDs for the 53.7 Myr mock SSP. In the CMD plane, the fit is excellent, with minimal per-pixel χ^2 residuals. However, since the `ACS-MS` plane is completely unconstrained for ages older than 1 Gyr, the fitting algorithm can arbitrarily introduce older stellar populations to the fitted stellar population model that biase the estimated mean age

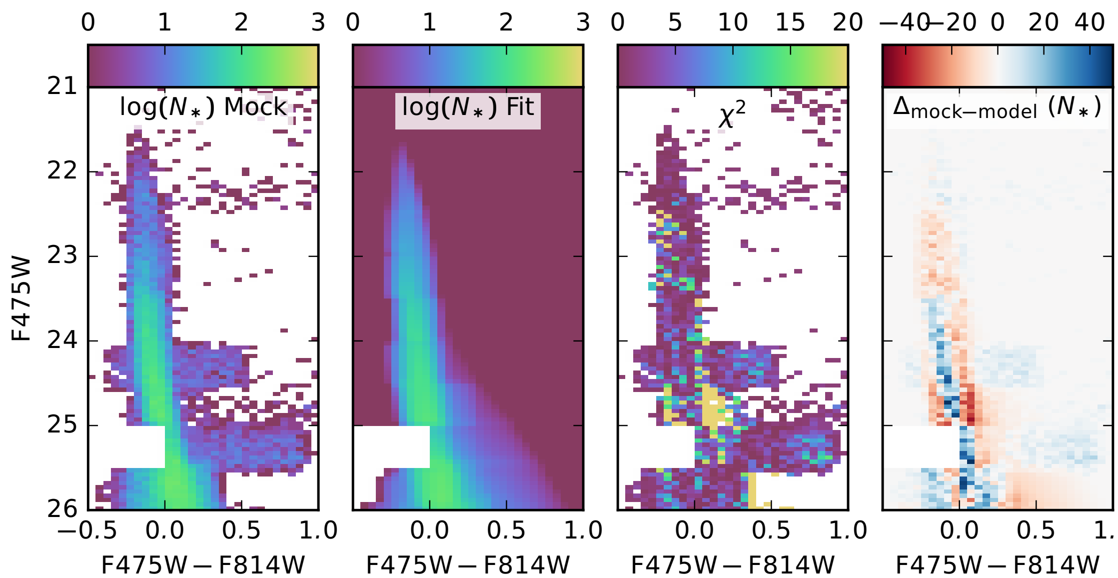
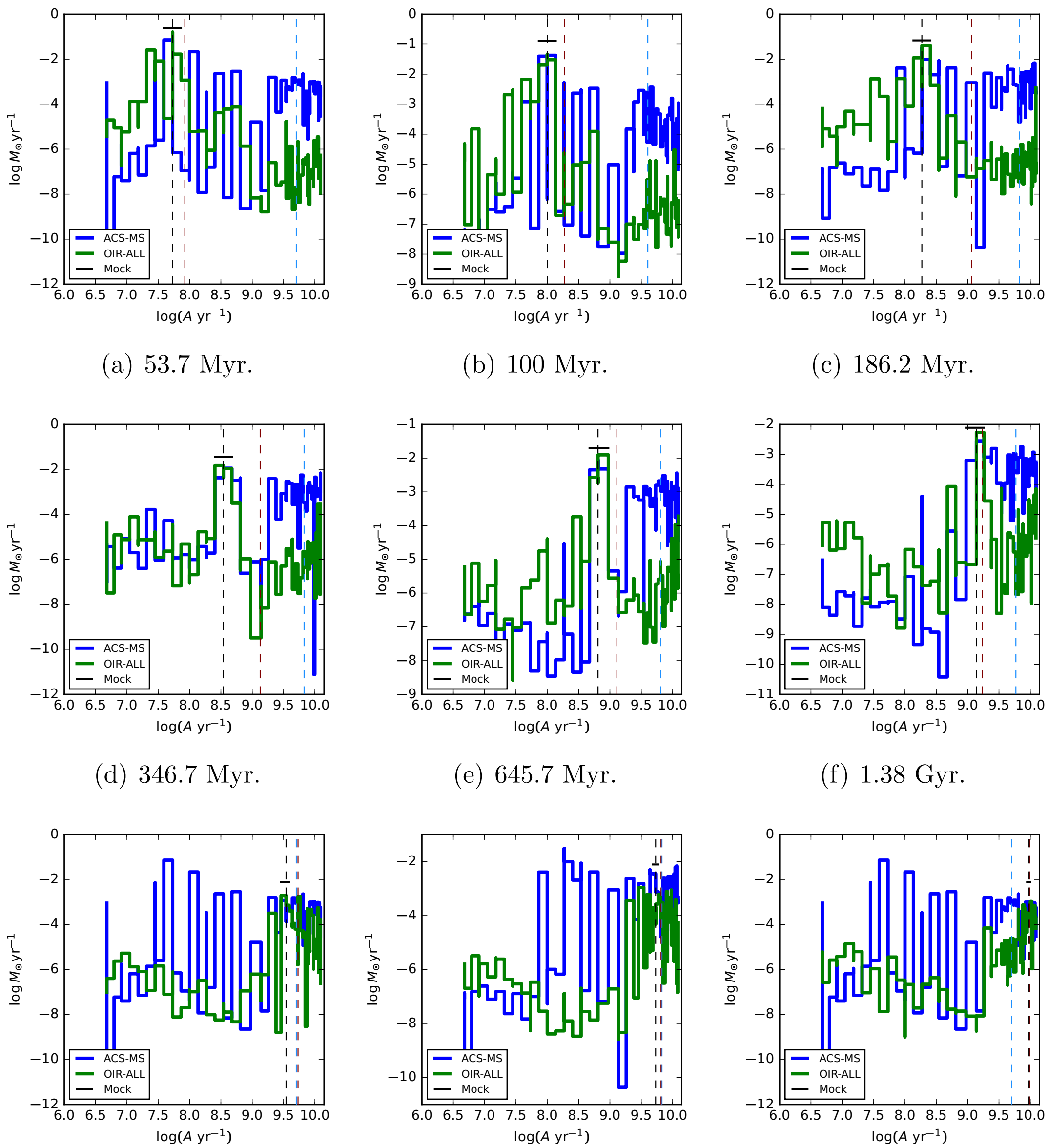


Figure 7.12: Residual `ACS-MS` Hess diagram of the 53.7 Myr mock SSPs generated with photometric errors corresponding to AST field #6. Panels correspond to, from left to right: mock Hess diagram, fitted Hess diagram, χ^2 residual between mock and fitted Hess diagrams, and difference of mock and fitted Hess diagrams in units of number of stars per Hess pixel.

(see Figure 7.13a). Compared to the `OIR-ALL` plane, Figure 7.13 shows that `ACS-MS` fits to young stellar populations have significant old stellar population components (the fact that fitted star formation histories in Figure 7.13 do not resemble SSPs is considered later in this section). Stellar population fitting in the `ACS-MS` plane fails because *StarFISH* appears unable to omit Hess planes from the fit that contain no information. Unfortunately, this apparent software limitation thwarts any comparison of `ACS-MS` fits to the Lewis et al. (2015) results.



(a) 53.7 Myr. (b) 100 Myr. (c) 186.2 Myr.

(d) 346.7 Myr. (e) 645.7 Myr. (f) 1.38 Gyr.

(g) 3.46 Gyr. (h) 5.37 Gyr. (i) 9.50 Gyr.

Figure 7.13: Star formation histories of mock SSPs generated with photometric errors corresponding to AST field #6. Each panel corresponds to an SSP of age 53.7 Myr, 100 Myr, 186.2 Myr, 346.7 Myr, 645.7 Myr, 1.38 Gyr, 3.46 Gyr, 5.37 Gyr and 9.50 Gyr. Star formations histories fitted with the `ACS-MS` (blue) and `OIR-ALL` (green) planes are shown against the mock SSP, shown as a horizontal black line at the appropriate age and star formation rate. See § 7.5.1 for discussion.

Effect of Age and Field Crowding on OIR-ALL SFH Fits

Increasing SSP age and increasing crowding both bias the `OIR-ALL` fits to *younger* mean fitted ages. Figure 7.11 shows that these biases are fundamentally driven by crowding effects. In the outer AST field, #6, age estimation bias is minimal at all ages. Mock SSPs reflecting photometric quality in progressively more crowded fields (AST #5, #4, #3) showed progressively larger degrees of age bias with increasing mock SSP age. These age biases are significant. A 10 Gyr mock SSP in AST #5 is estimated to be 5 Gyr younger. Even more extreme, a 10 Gyr mock SSP in AST #4 or #3 is estimated to be 9 Gyr younger.

To understand these trends, in Figure 7.14 we show fitting residuals in each AST field for (a) 1 Gyr and (b) 5 Gyr mock SSPs. The Hess diagrams plainly show that as stellar crowding increases, completeness in the photometric catalog decreases. It follows that the observed stellar population becomes sparser, with observations biased to the upper end of the RGB. In particular, the red clump at $m_{\mathrm{F160W}} \sim 23.5$ mag disappears in crowded fields.

Age introduces a similar bias, with the red clump becoming increasingly more sparsely populated. Figure 7.15 demonstrates this behaviour with plots of mock SSP `OIR-ALL` Hess diagrams for each lockfile bin age, along with a fit residual Hess diagram. One can think of this plot as a re-expression of Figure 7.11 through Hess diagrams rather than a mean age metric. Mock SSPs generated with errors for AST field # 4—Figure 7.15(a)—represent fitting behaviour in crowded fields, while Figure 7.15(b) demonstrates fitting relatively uncrowded photometry in AST field #6.

In the uncrowded case, Figure 7.15(b), the upper main sequence is visible up to age 1.3 Gyr. At the same time, the sub giant branch feature becomes less populated.

In the crowded case, the upper main sequences disappears by 0.6 Gyr. Once the upper main sequence disappears, or is only tenuously visible (e.g., at 347 Myr), *StarFISH* is quite poor at finding an appropriate star formation history; often using younger SSPs.

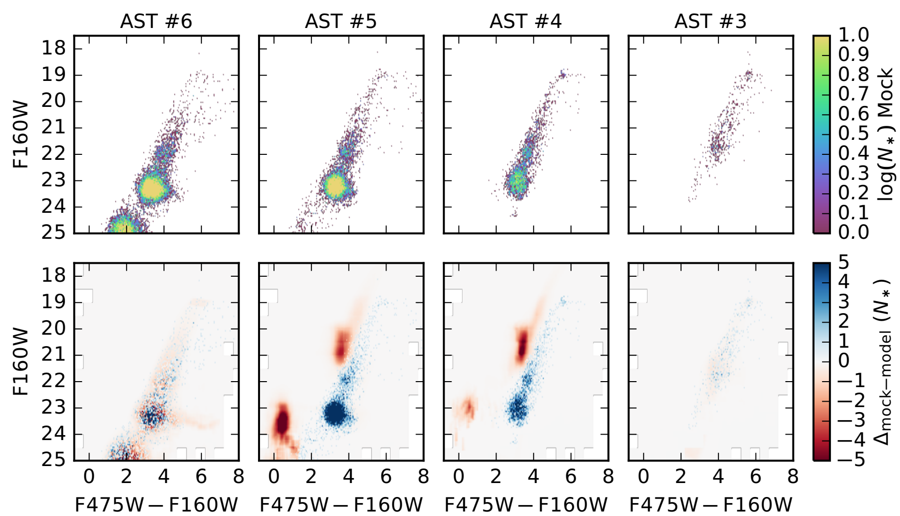
One reason that *StarFISH* may prefer younger SSPs in crowded fits is that information in the synthetic Hess diagrams for younger lockfile bins is less affected by incompleteness, judging by stellar density in the mock Hess diagrams. The optimization algorithm may find well-sampled synthetic Hess diagrams a more effective basis set for reproducing an observed Hess diagram, even if more technically correct Hess diagrams are available.

At very old ages, age estimation is difficult because age is captured entirely by the giant branch luminosity function, rather than the evolving positions of the main sequence turn off and subgiant branch. Under crowded conditions, the shape of the giant branch luminosity function is heavily affected—even the red clump is invisible. In this case, age is impossible to estimate correctly.

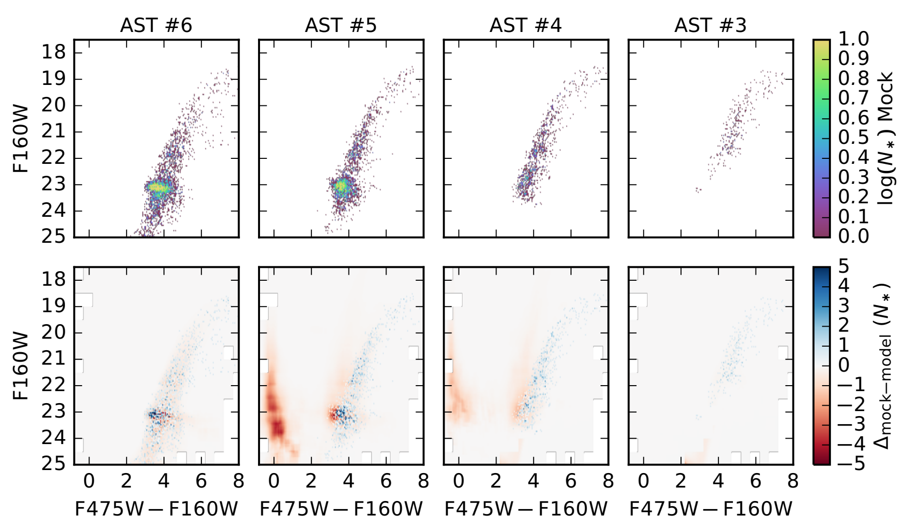
In summary, older SSPs in crowded fields are difficult to fit because they lack distinctive features in their Hess diagrams, such as the main sequence, the main sequence turn-off and red clump. *StarFISH* cannot reliably fit a stellar population to an incomplete upper RGB—even if the observational incompleteness, metallicity and dust extinction distributions are precisely known *a priori*.

Mean Age Uncertainty Estimates are not Backed by Observation

While we have so far considered only the amount of bias in the mean age estimate, Figure 7.11 also shows the sensitivity of statistical uncertainty in the mean age to

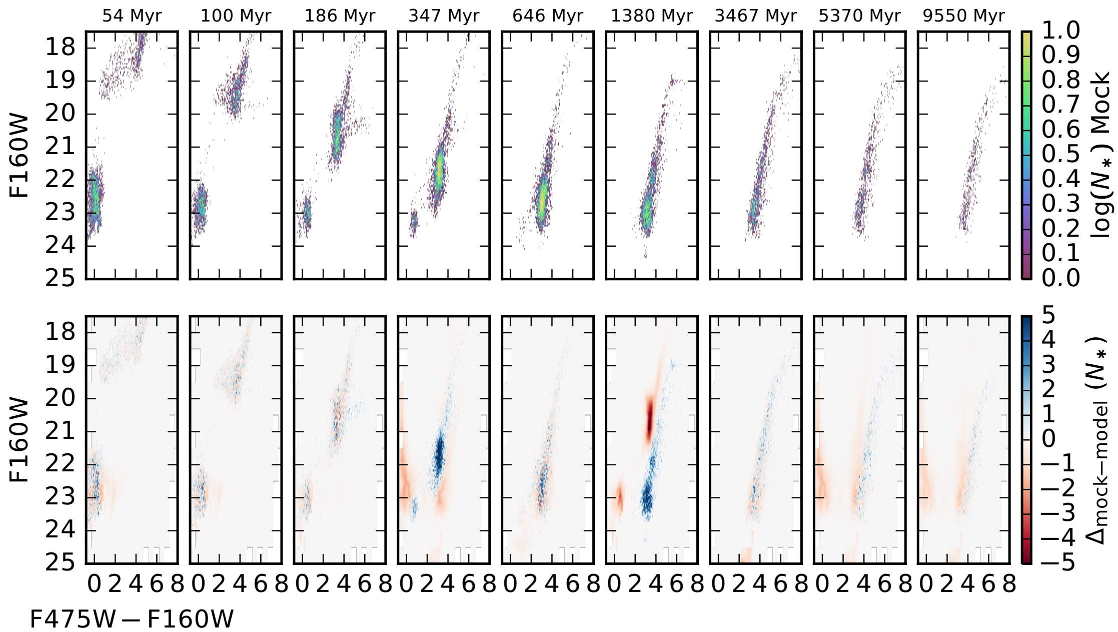


(a) $A_{\mathrm{mock}} = 1.38$ Gyr.

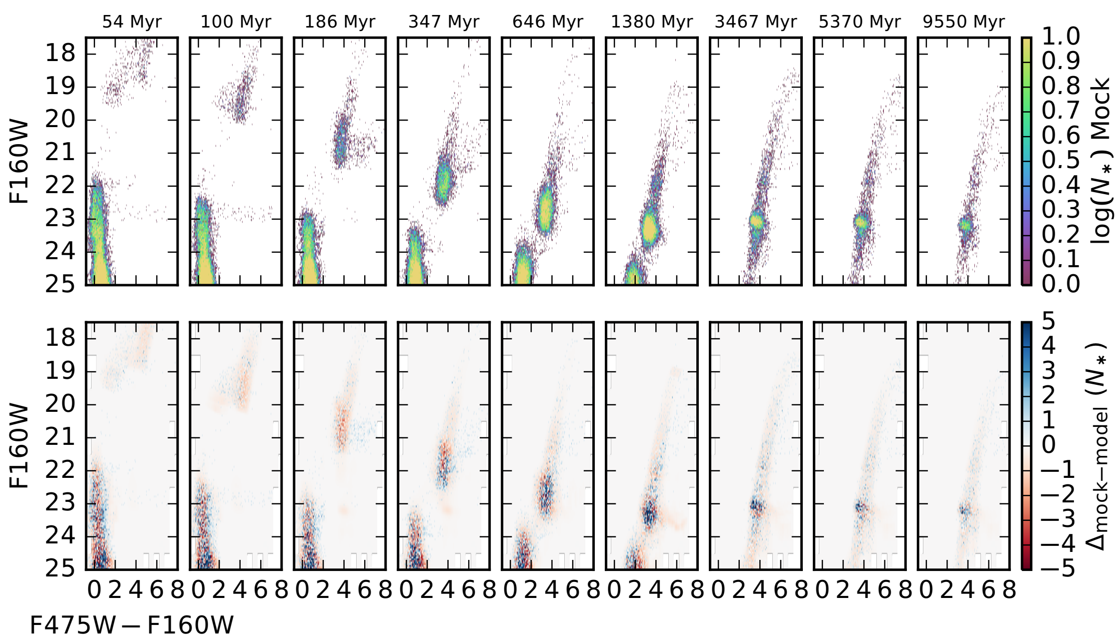


(b) $A_{\mathrm{mock}} = 5.37$ Gyr.

Figure 7.14: Mock SSPs aged (a) 1.38 Gyr and (b) 5.37 Gyr generated with errors for each artificial star test field. For each subplot, the top row shows the mock Hess diagram, and the bottom rows the *StarFISH* fitting residuals. Crowding in the artificial star test fields increases from left to right.



(a) AST #4



(b) AST #6

Figure 7.15: Mock SSP Hess diagrams (top rows) and fitting residuals for each lockfile time bin in the #4 and #6 PHAT artificial star test fields.

the mock SSP's age. Recall that we estimate the mean age's uncertainty through bootstrap resampling of the weights when computing the mass weighted age (§ 7.3.5). Individual weights are resampled from an uncertainty distribution of each lockfile's mass contribution to the SFH that are computed by *StarFISH*.

The statistical uncertainty of the mean age is larger than the bias in the mean age. In such a case, we would expect to *observe* the large statistical uncertainty through a dispersion in mean age estimates that dominates over systematic effects. This is not the case.

One possible explanation for this behaviour might be that since the SFHs that *StarFISH* fits to mock SSPs are not well peaked (see below), the mean age is naturally uncertain. This cannot be true since the mean age estimates *are* stable from age-to-age, with systematic variation overall. The most likely explanations are that either *StarFISH's* reported uncertainties are unrealistic, or our application software is misinterpreting and incorrectly converting *StarFISH*'s reported uncertainties into mass uncertainties.

SSPs Cannot be Recovered by StarFISH

Though reproducing an SSP's age through a mean age measure is useful, CMD fitting is intended to fit the star formation history of a stellar population. *If the intrinsic SFH is an SSP, we expect that the fitted SFH should resemble an SSP.* In this regard, *StarFISH* fails. The SFH models fitted to the mock SSPs studied here, such as in Figure 7.13, have non-negligible SF occurring at all ages.

As described previously, *StarFISH* uses a downhill simplex optimization to find the mass fraction associated with each isochrone, such that the linear combination of synthetic Hess diagrams built from those isochrones reproduces the observed Hess diagram. An underlying assumption promoted by literature that uses CMD fitting for stellar population science (such as Brown et al. 2003, 2006; Lewis et al. 2015, for M31 alone) is that these fits are *unique*. Uniqueness states that only one solution exists for a given fitting problem. Testing uniqueness with solutions found by downhill simplex-based optimizers is difficult because the optimization algorithm simply converges on a minimum, rather than mapping out the entire likelihood surface (as MCMC does, for example). However, these Mock tests give indirect proof that *StarFISH's* fitted SFHs are not unique. The correct fitted SFH corresponds to the mock SFH itself (all stellar mass is associated with a single isochrone), however, our fitted CMDs are composed of star formation contributions from *all* isochrones.

This is a fundamental flaw in the CMD fitting implementation of *StarFISH*; the algorithm is unable to judge the information content of each isochrone in attempting to fit a star formation history with a model containing the fewest degrees of freedom. *StarFISH* will always use the full complexity of the isochrones (binned by lockfiles) that the user made available. Compare this to other modelling techniques, where *model selection* includes Occam's Razor-like provisions for preferring simpler models (such as the Odds Ratio in Bayesian model selection, in particular).

The proper use of CMD fitting codes, like *StarFISH*, calls for vigilance when choosing the available model space based in part on the user's own interpretation of the astrophysical priors and results from previous fitting trials.

Table 7.3: Additional Hess planes for mock SFH estimation experiments. C.f. Table 7.1.

Name	x mag	x range	y mag	y range	Comments
`ACS-MS-28`	F475W – F814W	$(-0.5, 1.0)$	F814W	$(28, 21)$	Main sequence in ACS bands
`OIR-ALL-28`	F475W – F160W	$(-0.8, 8.0)$	F160W	$(28.0, 17.5)$	Full Hess plane covering wide color baseline, approximately $V - H$.

7.5.2 Mock SSP Tests with Idealized Photometry

The previous mock experiments explored the accuracy of fitting data with realistic errors, reflecting different crowding conditions across the M31 disk. Stellar crowding played a dominant role in the recovery of SSP stellar ages, however we also saw hints of intrinsic weakness in the Hess fitting optimization algorithm where a single SSP was fit with non-negligible contributions from SSPs of all ages (e.g., Figure 7.13).

To disentangle effects of photometric uncertainties and intrinsic limitations in isochrone fitting, we repeat the mock fits, but now using a mock dataset generated with no photometric scatter or incompleteness. The mock datasets are still reddened, so we use extinction estimated for the AST #6 field (Table 7.2).

Mock datasets constructed without errors are not impacted by incompleteness, so it is also possible to produce deeper Hess planes that capture more of the stellar luminosity function. We are still limited in the depth of these Hess planes since many orders more simulated stars must be sampled from the luminosity function to capture both the well-populated main sequence and comparatively populated post-main sequence features. To provide a sense of the performance improvements that are possible from improved data, we simulated and fit `OIR-ALL` Hess planes that are 2 mag deeper (`OIR-ALL-28`). These deeper planes are specified in Table 7.3. We *do*

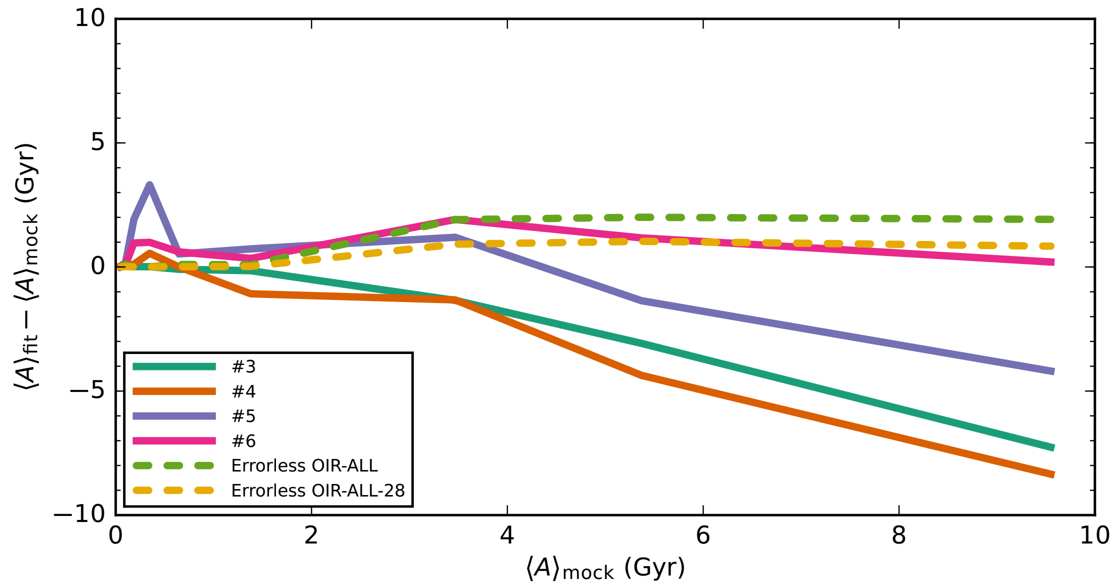


Figure 7.16: Mean age estimation accuracy for mock SSPs generated with no errors versus mock SSPs generated with errors corresponding to PHAT artificial star testing (AST) fields. Dashed lines show mean age estimation residuals for mock SSPs generated without photometric errors or incompleteness (but with dust extinction corresponding to the PHAT AST #6 field) in the `OIR-ALL` (red) and the deeper `OIR-ALL-28` (blue, see Table 7.3) Hess planes. Solid lines correspond to the mean age estimation residuals for mock SSPs generated with errors corresponding to the PHAT AST field, listed in Table 7.2 (all in the `OIR-ALL` Hess plan). For reference, AST field #6 is in the outer M31 disk, with the lowest crowding.

not consider the `ACS-MS` plane here since *StarFISH's* inability to reliably fit SSPs to the main sequence alone (see previous) makes any discussion of mean ages with idealized mock photometry moot.

Figure 7.16 summarizes this experiment by plotting mean age estimation residuals for the errorless mock photometry against results previously described in § 7.5.1. Mock SSPs modelled without photometric errors are modelled with similar reliability to trials that used photometric errors measured in PHAT AST #6. In the oldest age bins, ~ 10 Gyr, the errorless mock SSP fit with an `OIR-ALL` plane was estimated to be

2 Gyr too old; fits with the slightly deeper `OIR-ALL-28` plane reduced the maximum ages estimation error to 1 Gyr. By comparison, the oldest mock SSP modelled with photometric error consistent with PHAT AST #6 has even smaller age estimation errors. The limiting factor for accuracy in *StarFISH* age estimation of SSPs, then, is not photometric errors or incompleteness. Instead, the ultimate fitting limitation is the inability for the optimization algorithm to fit an SSP with a single isochrone, rather than a combination of several isochrone bins.

7.5.3 Mock Stellar Populations with a Declining Exponential Star Formation History

The previous two sections demonstrated that *StarFISH* is ill-suited for fitting SSPs (without specific tuning) since its Hess diagram optimization algorithm does not specifically attempt to reproduce an observed Hess diagram with the fewest isochrones possible. At the same time, M31 is not a simple stellar population. To understand *StarFISH's* fitting behaviour with *continuous* star formation histories, we have prepared a series of mock stellar populations with the exponentially declining star formation rate model,

$$\mathrm{SFR}(t_{\mathrm{start}} - A) \propto e^{-(t_{\mathrm{start}}-A)/\tau}, \tag{7.12}$$

where A is the *age* being considered and t_{start} is the age when star formation began. The shape of the star formation history is characterized by the e-folding time, τ (Gyr). For $\tau < 1$ Gyr, star formation is completed very early in a population's history, yielding an old population seen today. When the e-folding time increases, $\tau > 1$ Gyr, there is less early star formation, and overall the population is seen as

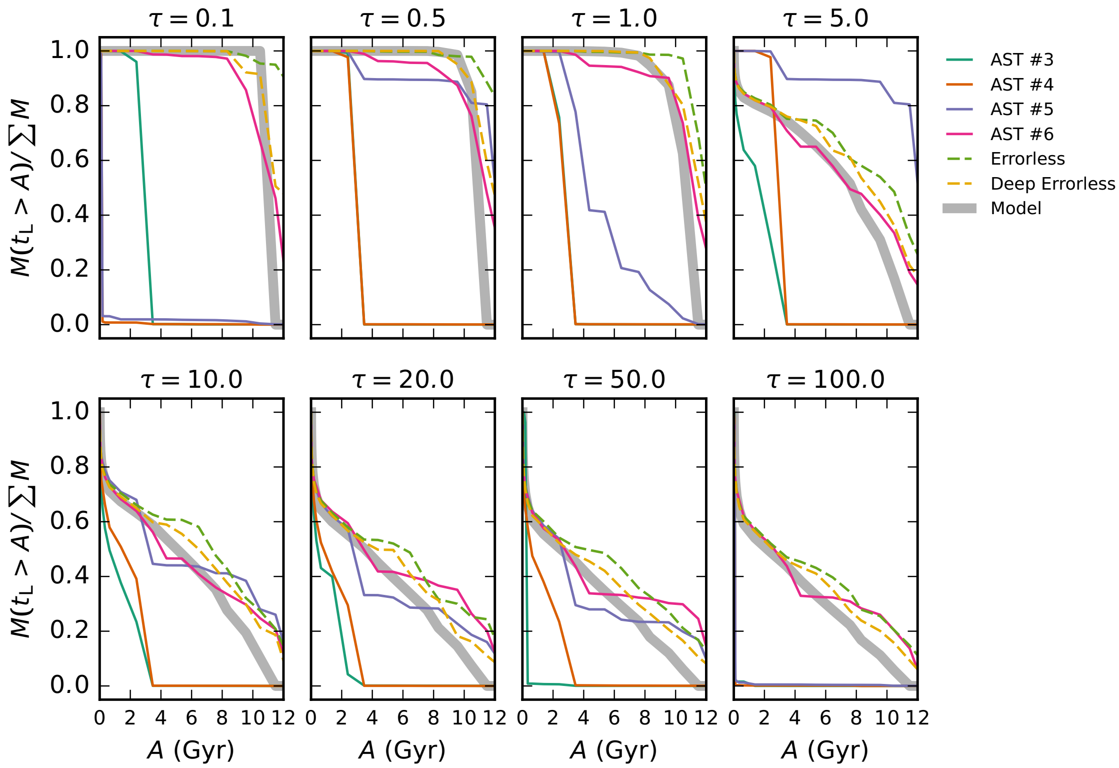


Figure 7.17: Mock and *StarFISH*-fitted cumulative mass growth functions for a series of exponential declining (τ) mock star formation histories. The e-folding time of each model's star formation rate, τ, is listed in units of gigayears. The thick grey line shows the mass growth curve of the mock stellar population, solid lines show fitted mass growth given mock photometry modelled photometric errors corresponding to PHAT AST fields, and dashed lines show fits with errorless mock photometry. All fits are done in the `OIR-ALL` plane, except for 'Deep Errorless,' which is done in the deeper `OIR-ALL-28` plane.

being younger.

Mean stellar age is a less relevant metric for measuring the accuracy of fitting a continuous star formation history. Instead we visualize the quality of fits through comparison of mock and fitted cumulative stellar mass growth function: $M(t_L - A)/\sum M$. That is, the total stellar mass formed across look-back times (t_L) before a given age (A), normalized by the total stellar mass formed by today. Since the `ACS-MS` plane is ill-suited for fitting entire star formation histories, we focus entirely on fitting in the `OIR-ALL` plane. In addition to producing mock stellar populations with errors drawn from PHAT AST fields, we also produce and fit errorless mock photometry (as in § 7.5.2).

Figure 7.17 shows cumulative mass growth functions for eight 'τ' star formation models, ranging from very old ($\tau = 0.1$ Gyr) to young ($\tau = 100$ Gyr). Mirroring results from estimating SSP ages, we also see that *StarFISH's* accuracy in reproducing continuous star formation histories is highly dependent on photometric quality. AST #6 is the only field where a continuous star formation history is reliably reproduced for any underlying star formation history. At most the fitted cumulative mass growth function deviates by $\pm 20\%$ of the mock stellar population. The errorless mock stellar populations are fit with equivalent accuracy to trials with mock photometry with AST #6 error models. In Figure 7.17 there is a hint that these error mock models are fit with systematically older stellar populations, which mirrors what was seen for SSPs (Figure 7.16).

7.5.4 Summary of Mock Fitting Results

In this section we have studied the fidelity of star formation histories estimated by *StarFISH* by fitting Hess diagrams that emulate the PHAT catalog. Mock trials allow us to isolate and compare several factors that influence the quality of fits: fitting plane (`ACS-MS` and `OIR-ALL`), star formation history (SSP and τ distributions with a range of intrinsic ages), and degree of photometric uncertainty and incompleteness by modelling Hess diagrams for each PHAT AST field along with trials that are free of all photometric uncertainty and incompleteness.

Fits in the `ACS-MS` plane were affected by unconstrained older isochrones, making `ACS-MS` an unsuitable tool for general-purpose star formation history fitting. Using the `ACS-MS` plane successfully would require trimming the isochrone basis set to manually truncate the fitted star formation history to only those that are seen in the Hess plane. Since our objective is to compare full star formation histories modelled by SEDs and Hess diagrams, rather than measuring the recent star formation history (Lewis et al. 2015), the `OIR-ALL` will be our primary tool when fitting the PHAT dataset.

We have also seen that photometry quality and completeness, driven by stellar crowding, ultimately limits the ability for *StarFISH* to fit a full star formation history. *Only the mock trials with error corresponding to the outer most* PHAT *fields, or trials without any photometric errors accurately reproduced star formation histories of all ages.* Star formation histories fit in the 10 kpc ring and within can be potentially unreliable (e.g., Figure 7.17).

These lessons from fitting mock photometry compel us to be cautious when interpreting the *StarFISH* analysis of M31's full star formation history, particularly in the mid and inner M31 disk, as we attempt in the next section.

7.6 Fitting Star Formation Histories Across the PHAT Footprint

With the procedures established in the previous sections, we can proceed to fit star formation histories across M31. Our specific goals are twofold. First, we seek to verify the results of Lewis et al. (2015), who fit main sequence stellar populations across the entire PHAT footprint, yielding star formation histories that extend to 600 Myr. Then with the same infrastructure, we attempt to extend the analysis to include all stellar populations. It is with this fit that we are able to directly compare star formation histories estimated from resolved stellar populations and unresolved spectral energy distributions (Chapter 6).

7.6.1 Fitting Method

We attempt to follow the methodologies of Lewis et al. (2015) closely, with three notable differences. The first two we have discussed at length: fits are made with *StarFISH* (Harris & Zaritsky 2001) rather than MATCH (Dolphin 2002), and the dust distribution is based entirely on Eq. 7.2 rather than a local optimization.

The third difference is that we fit star formation in PHAT fields, rather than subsections thereof. Lewis et al. (2015) divided the PHAT star catalog into 100×100 pc patches for fitting star formation histories. While this granularity represents the finest scale on which the star catalog is still sufficiently sampled, the computations are prohibitively expensive for us to repeat those fits. Instead, we fit PHAT's natural segmentation into fields, which represent coverage of single ACS or WFC3 camera exposures. Each PHAT brick consists of 18 fields; PHAT is composed of 23 bricks. PHAT fields cover 300 arcmin^2, or 16.8×10^3 pc^2 (projected).

For each field, we select stars based on both spatial criteria and that have `GST` flags in both bandpasses used in the CMD (*i.e.,* F475W and F814W or F475W and F160W). We associate that photometry to artificial star test catalogs from the artificial star test field in closest proximity (recall that Williams et al. computed artificial star tests in six selected fields across the disk). We also compute the mean dust mass surface density within each field to construct a localized attenuation distribution (§ 7.3.4). In order to achieve our goals of validating Lewis et al. (2015) and extending the analysis to all stellar populations, we fit with both the `ACS-MS` and

`OIR-ALL` planes.

Because of the location dependence of the attenuation distribution and artificial star tests, each field requires unique synthetic Hess planes to be generated with *StarFISH*. Each field requires roughly 12 hr of CPU time to generate synthetic Hess diagrams and perform star formation history optimizations in each plane. To cope with this load, we process fields as independent jobs on the CANFAR[6] computing grid.

7.6.2 Results

Spatially-resolved star formation histories are inherently multi-dimensional datasets, and we will consider them through multiple views. First we will consider only the distribution of mean ages in individual PHAT fields, and in the following section study star formation rates at discrete epochs.

Mean Age

Since we used mean age metrics to understand the accuracy of *StarFISH* in mock SFH fitting trials, it makes sense to first examine mean ages fitted to PHAT field photometry. In this section we refer only to fits made to the `OIR-ALL` Hess plane, since the `ACS-MS` is not sensitive to all ages and thus cannot be practically used for ages estimation (c.f. Figure 7.11).

Figure 7.18 shows a map of mean ages fitted to PHAT fields, while Figure 7.19 shows the mean ages of fields projected onto a M31 galactocentric disk radius axis. Broadly, Figure 7.18 shows that the *StarFISH* fits, at least qualitatively, capture the $R_{\mathrm{M31}} = 10$ kpc star forming ring where the mean stellar age is $\langle A \rangle \sim 2$ Gyr.

[6] `http://www.canfar.net/en/`

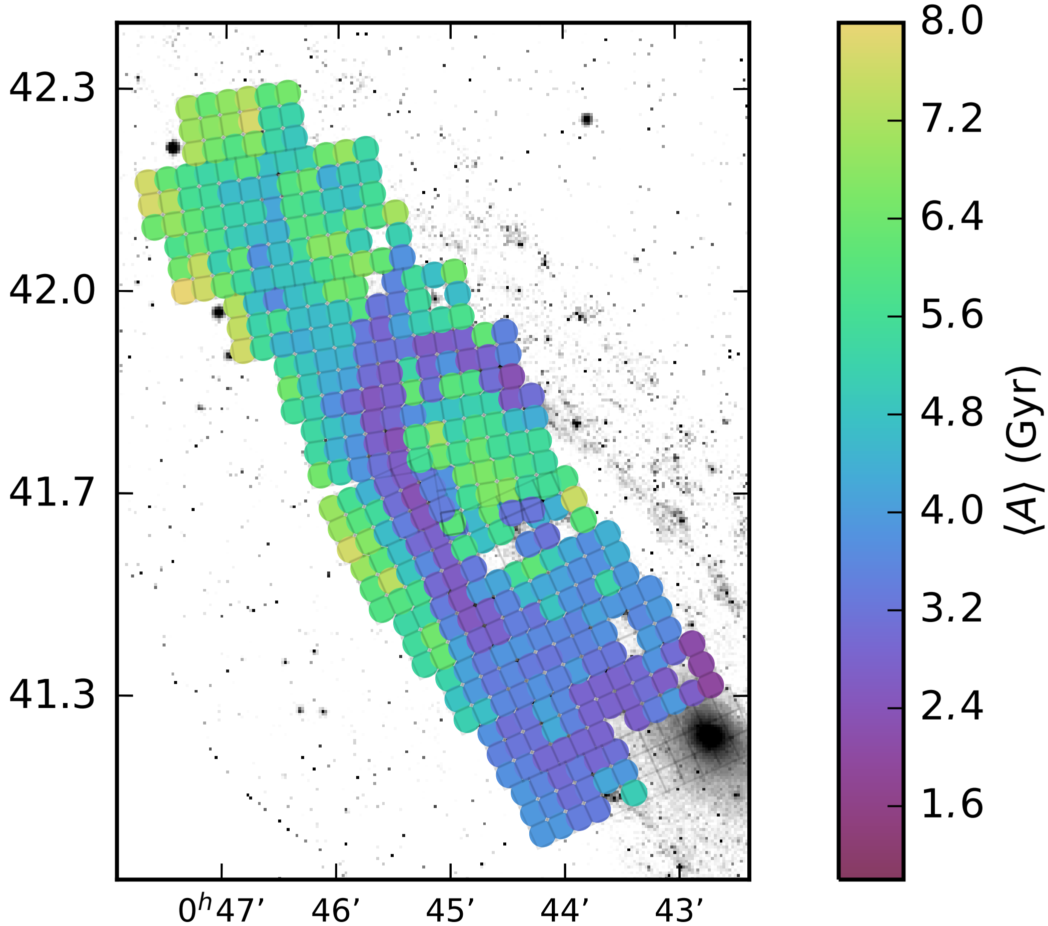


Figure 7.18: Mean stellar age of individual PHAT fields fitted with *StarFISH* using the `OIR-ALL` Hess plane. Each dot corresponds to a single PHAT field, the extents of which are defined by the grey box outlines. The GALEX NUV map of M31 (Gil de Paz et al. 2007) shows the locations of recent star formation in M31, for reference.

The *StarFISH* fits even suggest a secondary outer star forming ring, matching the GALEX UV map, where the mean age is estimated to be ~ 4 Gyr. In the inter-ring disk, outside of the ring structures, the mean stellar age is ~ 6 Gyr.

Besides revealing an arm and inter-arm age decomposition, Figure 7.18 also shows a radial age gradient, with the inner regions of M31 being youngest. Figure 7.19

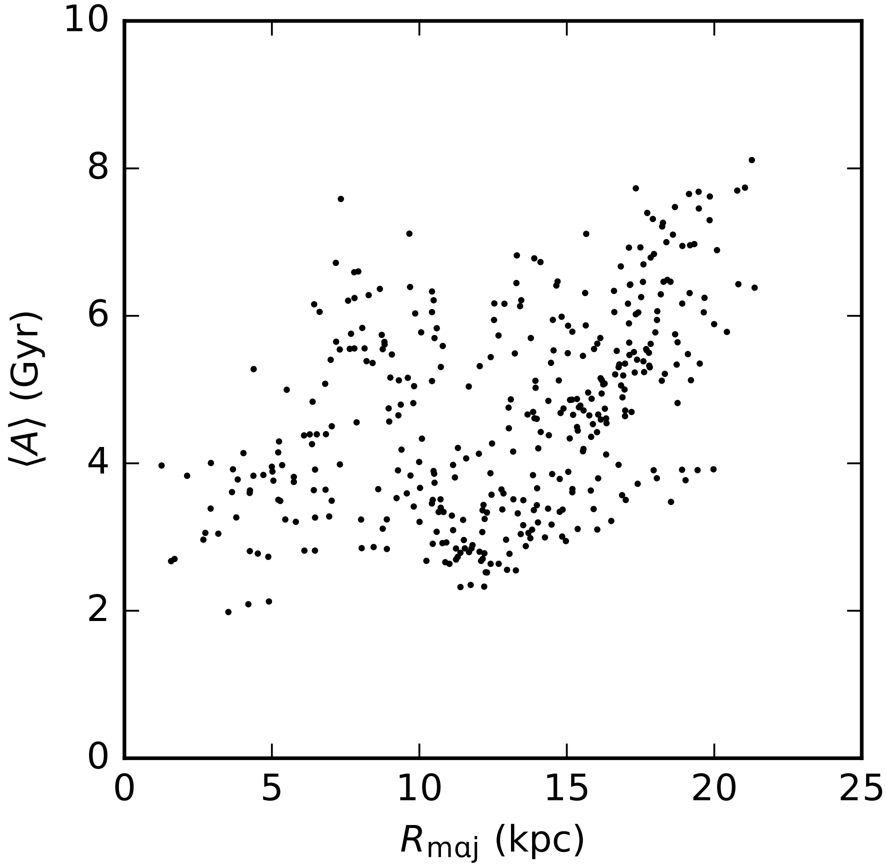


Figure 7.19: Radial profile of mean ages of individual PHAT fields fit with `OIR-ALL` planes.

shows that this radial gradient has considerable scatter, $\sim$ 2 Gyr, but PHAT fields are generally fit to mean ages of 2–4 Gyr in the inner regions of M31 and 6–8 Gyr in the outer regions. Fits to bulge regions inside $R_{\mathrm{M31}} < 5$ kpc are as young as fields on the 10 kpc ring. Of course, this result for a young M31 bulge disagrees with spectroscopic studies (e.g., Saglia et al. 2010) and our own SED fits (Chapter 6). Section 7.5.1 (particularly Figure 7.11) showed that our *StarFISH* fits tend to under-estimate stellar ages in crowded stellar regimes. In the most crowded test field, we found that old, 10 Gyr, SSPs would be fit with a mean age of 2 Gyr by *StarFISH*. The fits to PHAT fields presented here are entirely consistent with this bias; the M31 bulge *could* intrinsically be $\langle A \rangle \sim 10$ Gyr.

The potential for age estimation biases by *StarFISH* begs us to question what aspects of Figure 7.18 and Figure 7.19 *are* realistic. In § 7.5.1 we found that *StarFISH* fits the `OIR-ALL` plane in AST fields #5 and #6 accurately. Recall the positions of these AST fields from Figure 7.1; AST #5 is well within the 10 kpc star forming ring, lying at ~ 5 kpc. Hence we should not dismiss the mean age results for both the rings and the inter-ring for radii beyond $R_{\mathrm{M31}} > 5$ kpc. Brown et al. (2006) fit deep F606W-F814W HST CMDs that reached the main sequence turn-off in an outer disk field at $R_{\mathrm{M31}} \sim 20$ kpc on the northern major axis and found an age distribution very consistent with our mean age result for the same region (see their Fig. 18).

Assessing the accuracy of the mean age result for the 10 kpc star-forming ring is more complex. Younger mean ages are consistent with enhanced crowding specifically in the rings (§ 7.5.1). If *accurate*, our *StarFISH* fits imply that half of the stellar mass associated with the ring regions was created in the last 2 Gyr; supporting the idea that the rings are a long lived structure, and not simply a 'frosting' of recent star formation. In the next section we expand upon the current mean stellar age analysis by examining the full stellar mass growth curve estimated by *StarFISH*.

Stellar Mass Build-up in M31

M31 is a complex stellar population, with star formation having occurred throughout the galaxy's formation. Although mean age is a convenient metric to visualize, it reduces the *StarFISH* fitting results into a single metric. As we did in § 7.5.3, the cumulative stellar mass growth curve is a useful way to plot the full star formation history of a galaxy.

Through Figure 7.20 we show cumulative mass growth functions, fitted by *StarFISH* with the `OIR-ALL` plane, for individual PHAT fields located along M31's major axis. The galactocentric radius of each field is encoded via colour. As Figure 7.19 summarized earlier, the *StarFISH* fits illustrate an overall outside-in picture of M31's star formation. Through the cumulative mass growth functions, we now see that these fits to M31's inner regions mirror the mock fits of 'τ' formation histories, Figure 7.17. For both older ($\tau < 1$) and younger ($\tau > 1$) mock 'τ' star formation histories, mock trials of crowded fields tended to fail catastrophically by delaying star formation until the $A \sim 4$ Gyr lockfile bin. That the same behaviour is seen here in fits to PHAT fields with $R_{\mathrm{maj}} \lesssim 5$ kpc indicates that our fits to M31's inner disk and bulge CMDs should be discounted completely.

Interpreting the veracity of fits to fields in the 10 kpc ring is more complex since no PHAT AST field coincided with the 10 kpc ring for direct investigation with mock fitting trials. In Figure 7.21 we show maps of the ages when each PHAT field is estimated to have formed a specific fraction of mass. The 10 kpc ring stands out in these plots, with star formation only starting earnestly (as indicted by the age when $M(t_L > A)/\sum M \geq 0.2$) in the last 3 Gyr, compared to 5–6 Gyr for neighbouring regions in the outer disk. For this to be possible, the 10 kpc ring would need to be a massive and long-lived structure in the M31 disk so that late stellar mass growth in the 10 kpc ring would dominate the stellar mass fraction over any smooth stellar disk component that began to form 6 Gyr ago. The surface brightness profiles—Figure 4.7, particularly in the near-IR—show no evidence of a massive stellar component at

10 kpc. We must conclude that the fitted star formation histories of PHAT fields in the 10 kpc ring are also biased to younger fitted ages by the same stellar crowding-driven process that affects the inner disk.

Mock trials allow us to be more confident in *StarFISH* fits to the outer M31 disk, beyond the 10 kpc ring. In these fields, star formation is estimated to have started very early, up to 12 Gyr ago, with 20% of stellar mass in place since 5–6 Gyr ago. Between 5 Gyr ago and 2 Gyr ago most of the mass in the outer disk had formed. The rapid rise in stellar mass growth between 4–2 Gyr in Figure 7.20 is also seen in mock results for 'τ' formation histories; see the $\tau = 10, 20, 50, 100$ panels in Figure 7.17. Coincident to this inflection epoch is a fundamental shift in the structure of the stellar population on the `OIR-ALL` fitting plane, shown in Figure 7.15(b). Populations younger than 3 Gyr include both the red clump and upper main sequence in `OIR-ALL` plane, while older stellar populations include only the red clump (and giant branch).

Star Formation Rates

Whereas the mean age and cumulative mass growth history analyses considered the entire star formation history, star formation rates (SFRs) are a measure of a stellar population's instantaneous behaviour. Examining recent star formation rates, for example, allows us to finally use the `ACS-MS` fits, which are known to be otherwise unreliable for ages older than ~ 1 Gyr (§ 7.5.1). This allows us to tie the current *StarFISH*-based analysis with the work of Lewis et al. (2015) who used a plane similar to `ACS-MS` for their fits.

In Figure 7.22 we show SFRs of all PHAT fields fit with both `ACS-MS` and `OIR-ALL` fitting planes. Each SFR is coded by the field's galactocentric radius.

This plot shows how fits to the `OIR-ALL` and `ACS-MS` planes disagree significantly. The *StarFISH* fit with the `ACS-MS` plane has declined, on average, by 2 dex in the last 13 Gyr. As discussed in § 7.5.1, this result is problematic since the overall SFR(t) gradient is driven by the bulk of star formation being assigned to ages that the `ACS-MS` plane is not sensitive to. Alternately, the `OIR-ALL` fit has low star formation rates in early years, peaking at ~ 1 Gyr, and declining by 1 dex since, consistent with the mass accumulation picture of Figure 7.20. Again, we question whether this result is real, or the result of new features appearing in the `OIR-ALL` plane for stellar populations ~ 1 Gyr and younger.

Figure 7.22 does not clearly show radial separation of SFR behaviours. Instead we refer to Figure 7.23 and Figure 7.24, which show maps of star formation rates at specific ages for `ACS-MS` and `OIR-ALL` fits, respectively. In `OIR-ALL` fits, the 10 kpc ring only became gradually distinct from an overall inside-out SFR gradient over the last 1 Gyr (with the appearance of significant star formation in the M31 bulge around 1 Gyr likely related to the crowding effects seen in mock tests, discussed in § 7.5.3). In `ACS-MS` fits the 10 kpc ring is seen at all times in SFR maps. This is another example of how the existence of a young stellar population is driving inferred star formation at ages older than are directly observed in the `ACS-MS` plane. But if we examine the relative star formation rate maps since 600 Myr, for which `ACS-MS` is sensitive, we see that the 10 kpc ring had active star formation since at least 600 Myr ago with star formation generally declining since. Recently (within the last 10 Myr), the *StarFISH* fits show a resurgence of star formation in the 10 kpc ring, although the rate of star formation is still ~ 1 dex lower than 600 Myr ago.

The ubiquity of the 10 kpc ring as a star formation feature in recent times is reiterated in Figure 7.25, which shows SFRs at discrete times against galactocentric radius for PHAT fields found on the M31 major axis. This Figure is analogous to Figure 6 of Lewis et al. (2015). In very recent times, `OIR-ALL` and `ACS-MS` fits both find that the 10 kpc ring is the dominant feature, with the 5 kpc and 15 kpc regions having slightly lower levels of current star formation.

Both fits also show that the 5, 10, and 15 kpc star formation regions are stationary. Lewis et al. also found that the 10 kpc ring was stationary over the last 600 Myr, and thus unlikely to be an expanding shockwave created by a face-on collision between M31 and M32, originally suggested by Block et al. (2006). Instead, these star formation regions (including the 5 and 15 kpc) might be related to a long-term resonance set up by M31's asymmetric bulge (Athanassoula & Beaton 2006; Beaton et al. 2007).

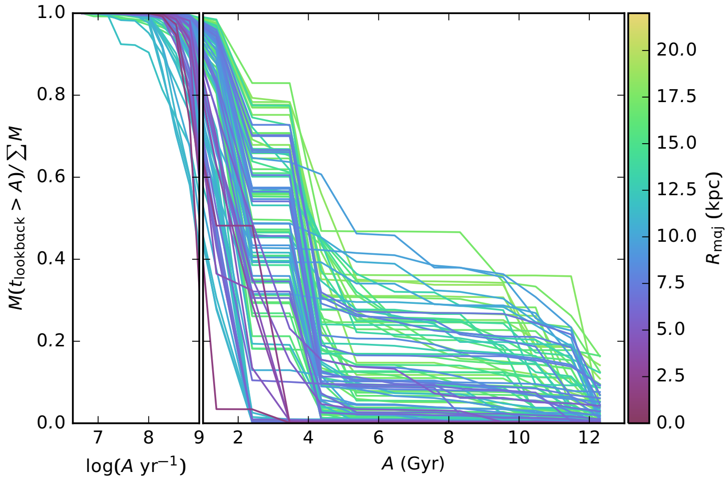


Figure 7.20: Mass accumulation distribution functions for PHAT fields within $\pm 20°$ of the northern M31 major axis fit by *StarFISH* with the `OIR-ALL` CMD plane. Mass is normalized to the total mass modelled for each PHAT field. The time axis is split between a logarithmic projection for ages less than 1 Gyr (left panel) and linear for older ages (right panel). Mass accumulation functions for each PHAT field are coloured by galactocentric radius.

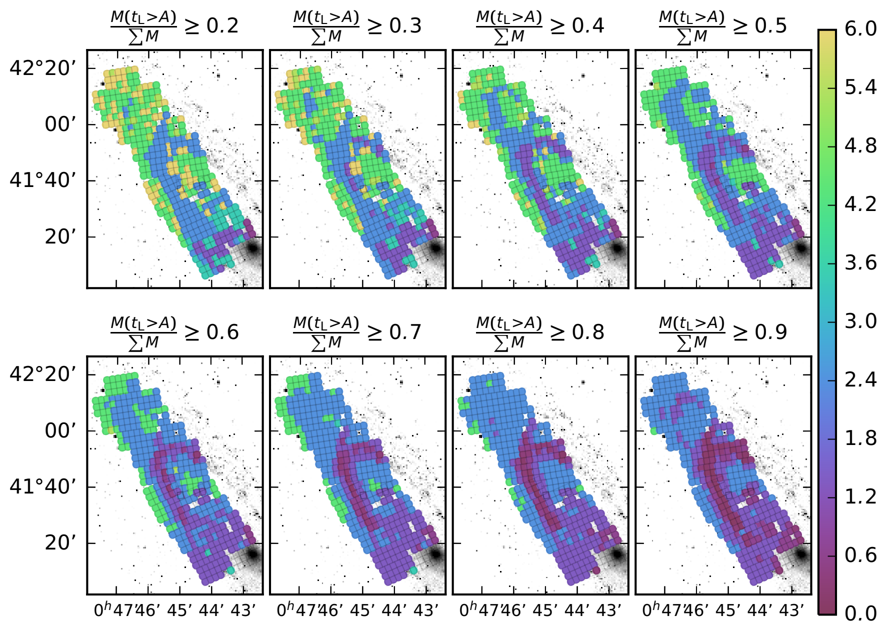


Figure 7.21: Maps of ages when the specified fraction of stellar mass has formed in each PHAT field. Each panel is labelled by a threshold fraction of stellar mass formation. Older ages at each threshold fraction indicate that mass build-up occurred earlier for the stellar population in that PHAT field.

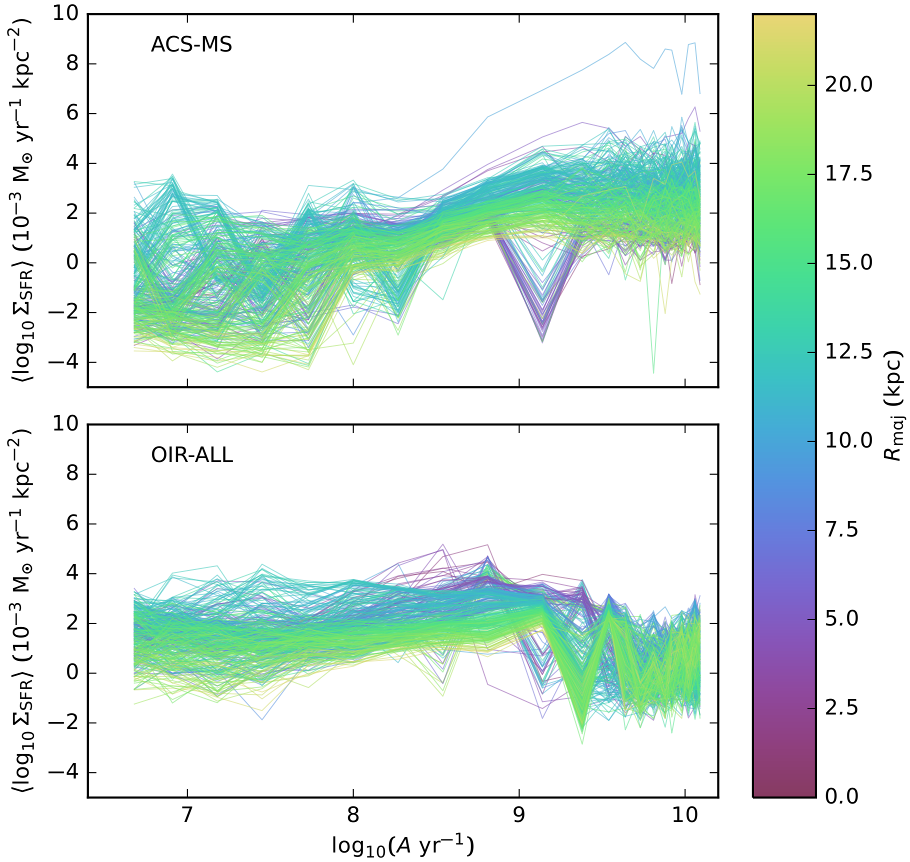


Figure 7.22: Star formation histories of individual PHAT fields, coloured by their galactocentric radius. Histories fit with the `ACS-MS` (top) and `OIR-ALL` (bottom) planes are shown.

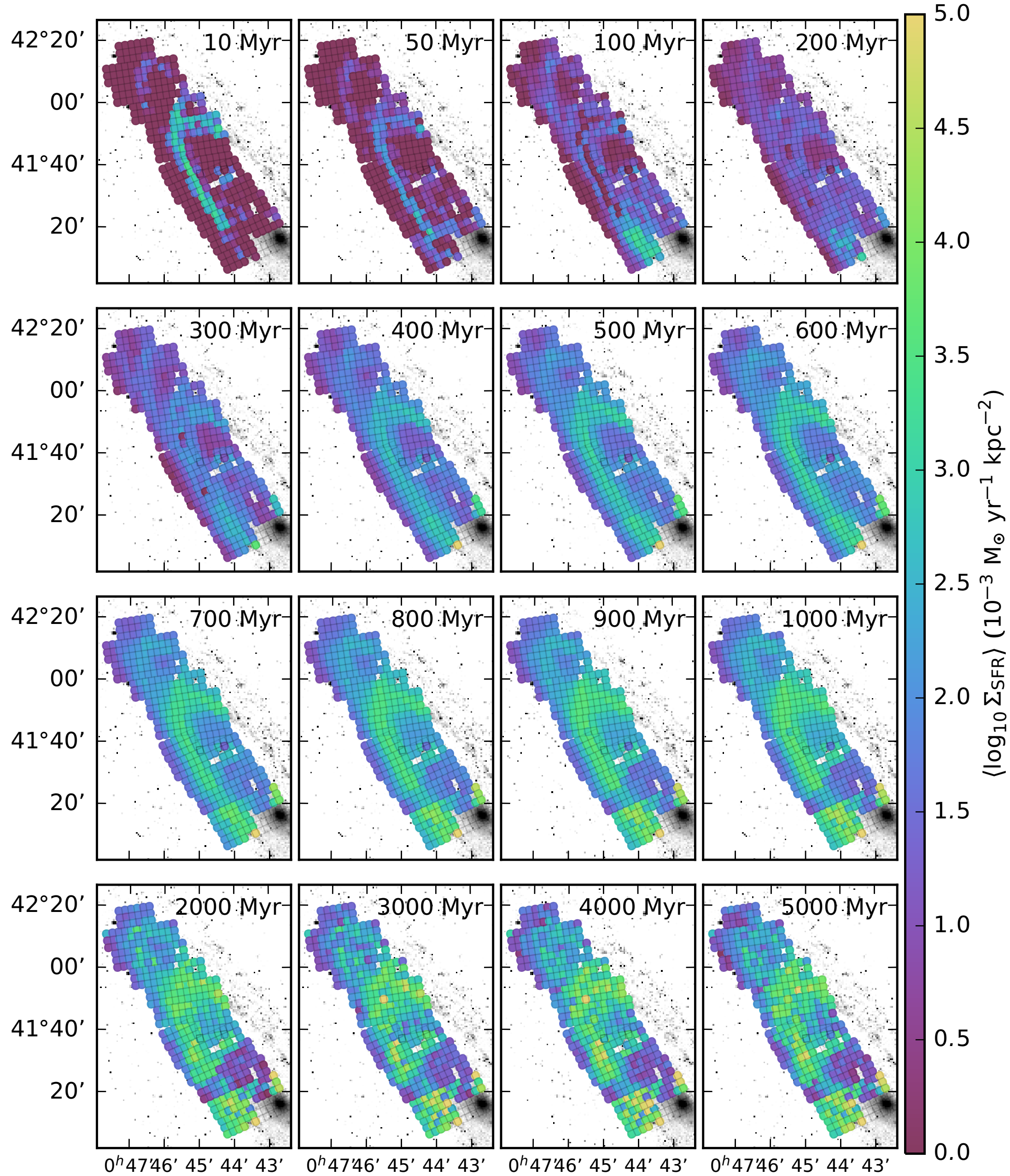


Figure 7.23: Star formation rates of individual PHAT fields fit in the `ACS-MS` plane. Each panel corresponds to a look-back time specified in the upper right corner. The GALEX NUV background map (Gil de Paz et al. 2007) shows regions of current star formation.

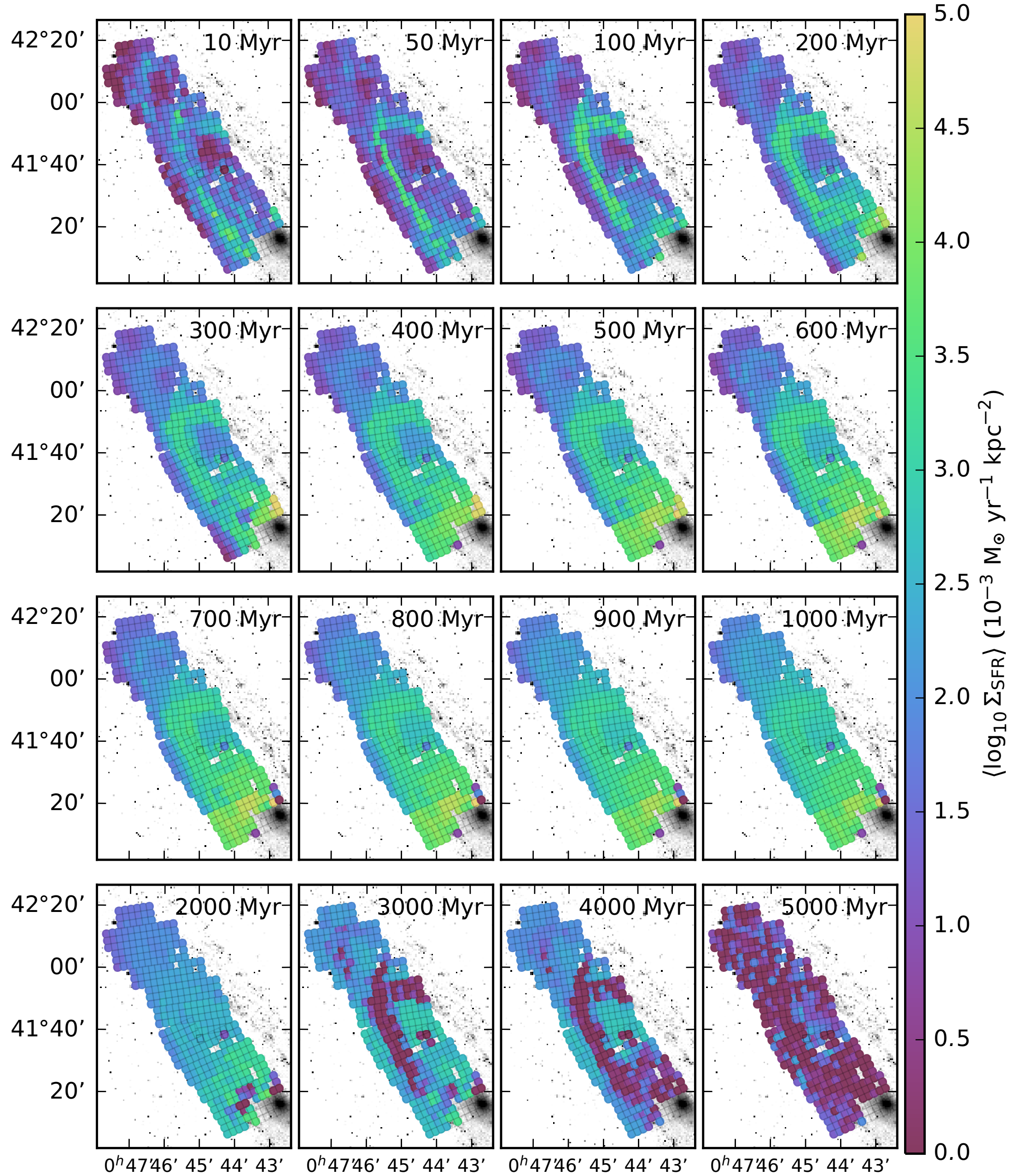


Figure 7.24: Star formation rates of individual PHAT fields fit in the `OIR-ALL` plane. Each panel corresponds to a look-back time specified in the upper right corner. The GALEX NUV background map (Gil de Paz et al. 2007) shows regions of current star formation.

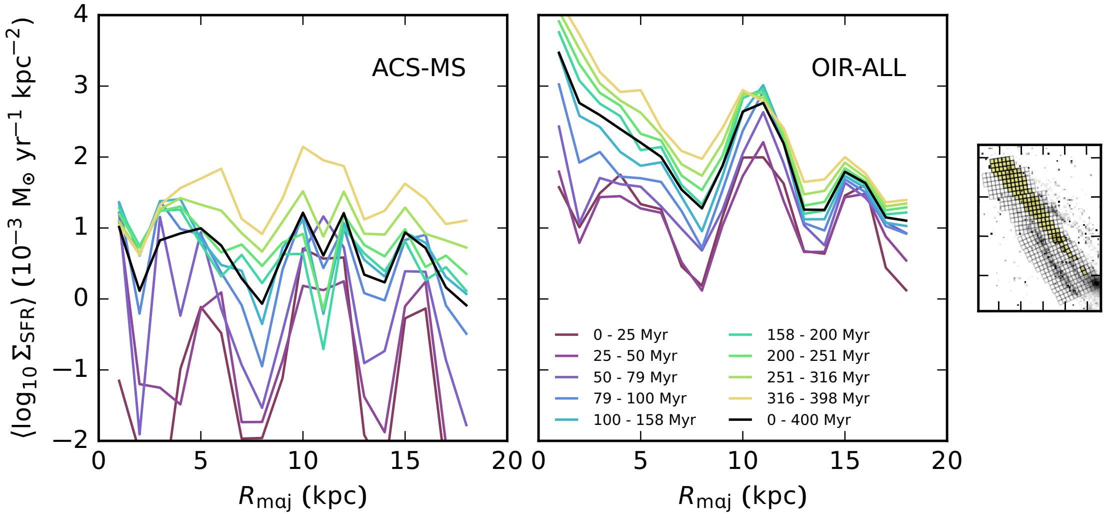


Figure 7.25: Mean star formation rates at discrete look-band times for fields along the M31 major axis. Star formation rates estimates from the `ACS-MS` and `OIR-ALL` planes are shown in the left and centre panels, respectively. PHAT fields, highlighted in yellow on the map at right, are selected around a projected $\pm 20°$ wedge around the northern major axis. This figure emulates Figure 6 of Lewis et al. (2015).

Chapter 8

Discussion

In this work we have presented ANDROIDS, the Andromeda Optical and Infrared Disk Survey, a comprehensive photometric study of M31 with original imaging in six optical and near-IR bands. ANDROIDS is a foil for multiple diverse studies that span the technical aspects of wide-field photometry and spectral energy distribution (SED) modelling to the astrophysics of the Andromeda Galaxy's assembly and evolution.

In this Chapter, we review what we have learned throughout the course of this study. First, in § 8.1 we briefly review the ANDROIDS dataset. Then in § 8.2 we review the technical steps needed to photometrically calibrate ANDROIDS's wide-field photometric dataset. In § 8.3 we summarize our new results about M31's stellar populations and interstellar medium through the full ANDROIDS dataset. Finally in § 8.4 we compare the performance of multiple stellar population estimation methods as they apply to M31 data in this work and in the literature.

8.1 Review of the Andromeda Optical and Infrared Disk Survey

ANDROIDS is a panchromatic survey of spatially-resolved SEDs across M31. The survey is constructed from a combination of new optical and near-IR images that improve greatly upon existing datasets, with existing datasets.

We used CFHT/MegaCam, along with the Elixir-LSB method to assemble accurate maps of M31's $u^*g'r'i'$ surface brightness out to 40 kpc from M31's centre (down to $\mu_{g'} = 26.5$ mag arcsec^{-2}). Using CFHT/WIRCam we assembled the first well-calibrated near-IR J and K_s mosaics of M31 out to 20 kpc, reaching $\mu_{K_s} = 22.5$ mag arcsec^{-2}. With both instruments, we used a sky-target nodding observing strategy to model and subtract the real-time background.

To complement this dataset, we incorporated existing GALEX UV (Gil de Paz et al. 2007), Spitzer IRAC mid-IR (Barmby et al. 2006), and Herschel PACS and SPIRE images covering the mid and far-IR (Fritz et al. 2012). For most analyses we combined these datasets with a normalized PSF and $36'' \times 36''$ pixel scale (equivalent to 136 pc $\times$ 608 pc) to match the Spitzer 500 μm image.

8.2 Wide-field Photometric Background Calibration

Before we could address astrophysically-motivated questions about M31's stellar populations and the comparative effectiveness of different modelling approaches, we had to reckon with the challenge of calibrating a six-band surface brightness dataset that covered the entire M31 bulge and disk obtained from the ground-based CFHT. The crux of the problem is that both the $1^\circ \times 1^\circ$ and $20' \times 20'$ fields of the CFHT MegaCam and WIRCam cameras, respectively, are filled entirely by M31 light. This prevents us from employing the typical background calibration procedure for extragalactic observations where background is measured directly from the blank sky surrounding the galaxy on the same CCD integration. Instead, we leverage the sky-target nodding technique to sample background level from sky proxy observations interspersed with target observations. Sky-target nodding is difficult to employ in M31 because of the large, 1°–3° nods required to avoid M31 light.

Background calibration with WIRCam is further complicated by the bright and variable near-IR atmospheric background and by uncertainty in the overall best practices for photometric calibration. We find that the previously-standard use of dome flats is unsuitable, and instead we calibrate the ANDROIDS/WIRCam images with sky flats built nightly from our own sky observations. Regarding sky-target nodding, we demonstrated multiple cadences with maximum sky sampling latencies of X and Y minutes. Given the 1°–3° nods, we find that regardless of the latency we are limited to knowing the absolute background level on the M31 to no better than 2% — an unacceptably large uncertainty given that the M31 disk is up to 10^4 times dimmer than the background in J and K_s bands. Furthermore, the edge-to-edge background level across a WIRCam image is uncertain at a level of 0.3% due to differences in the background at target and sky fields.

To solve this, we find it is essential to calculate and apply sky offsets that minimize surface brightness differences in overlapping images. The most significant calibrations are made by computing the offset necessary for individual WIRCam frames at the same field to have consistent surface brightness, followed by the offsets required for the image of one WIRCam block (a field) to have consistent surface brightness with the network of neighbouring blocks. WIRCam frames within a block do not require significant offsets to have consistent surface brightnesses. Overall, the sky offset technique is able to reduce the 1σ distribution of block-to-block surface brightness differences from $\sim 2\%$ to $< 0.1\%$ of the background level.

We therefore find that, even more important than minimizing the latency between sky and target observations, wide-field surface photometry observing programmes should optimize for sky-offset optimization-based calibration. Observing many overlapping blocks improves the effectiveness of sky offset optimization.

In assembling ANDROIDS/MegaCam optical $u^*g'r'i'$ mosaics, we did not experience the same level of complication as with the near-IR observed with the WIRCam instrument. The sky-target nodding observational approach in conjunction with the Elixir-LSB method for subtracting a model of the background results in well-calibrated block mosaics. We did fit sky offsets to make the full-disk MegaCam mosaics, though these offsets are minor compared to MegaCam. Overall, the surface brightness calibration of the MegaCam mosaics is footprint-limited. In r', for example, the MegaCam images trace an exponential surface brightness profile $\mu_{r'} = 26$ mag arcsec^{-2} at the edge of the ANDROIDS footprint. The main photometric limitation in our MegaCam images is scattered light around bright Milky Way foreground stars. We resolve this issue by manually masking these regions.

We developed a hierarchical Bayesian model to refine the surface brightness calibration of ANDROIDS images. In this method, we modelled the SEDs of multiple pixels along a wedge of M31 simultaneously, while also modelling a constant background correction per bandpass in each wedge. Using Gibbs sampling, these models converged quickly, and we found that the necessity of modelling many pixels with diverse stellar populations simultaneously with the same background correction effectively constrained the background correction. This hierarchical background modelling is critical for resolving background subtraction uncertainties in our WIRCam images. At the same time, this modelling was not necessary for the MegaCam images; indeed we constrained the $u^*g'r'$ images to have no background correction.

Overall, we have successfully demonstrated the construction of an extremely wide-field surface brightness map of a galaxy in both optical and near-IR bands where a sky-target nodding approach is required. Our ANDROIDS maps are stable to $\mu_{g'} = 26.2$ mag arcsec^{-2} at $R_{\mathrm{M31}} = 40$ kpc and $\mu_J = 22.5$ mag arcsec^{-2} at $R_{\mathrm{M31}} = 20$ kpc. In short, sky-target nodding for assembling a background model is useful for both optical and near-IR imaging. Sky offset optimization compensates for the absolute background level uncertainties that remain with sky-target nodding. To be suitable for sky offset optimization, observational programs must be designed with many overlapping blocks. Finally, hierarchical SED modelling is a promising technique for refining remaining systematic background levels.

8.3 The Stellar Populations and Interstellar Medium of M31

By modelling the SED of ANDROIDS CFHT imaging in conjunction with GALEX, Spitzer, and Herschel images available in the literature, we have drawn a comprehensive picture of M31's stellar and ISM content across its entire bulge and disk.

ANDROIDS photometry shows clear gradients in colours across M31's disk. M31's bulge is redder ($g' - i' = 1.1$ mag) than its outer disk ($g' - i' = 0.7$ mag). This result alone is consistent with inside-out disk galaxy formation, where the bulge forms first and the disk grows with gas infall. Furthermore, the radial profile gradients in both $g' - r'$ (0.008 mag kpc^{-1}) and $g' - i'$ (0.015 mag kpc^{-1}) are constant across the entire radial span, meaning that M31's disk cannot be treated as a uniform stellar population.

By modelling the full UV – far-IR SEDs of $83'' \times 83''$ (0.3 kpc $\times$ 1.4 kpc) pixels with MAGPHYS (da Cunha et al. 2008), we find that the total stellar mass within

20 kpc is $5.4(\pm 0.031) \times 10^{10}$ $M_\odot$. 68% of this mass lies within 1 R_e. This stellar mass estimate is approximately 30% of the stellar mass required by maximal disk rotation curve decompositions (Geehan et al. 2006; Tamm et al. 2012). Our sub-maximal disk mass thus underscores the importance of dark matter in the dynamics of galaxy interiors, not just the halo. Finally, our stellar mass estimate at 20 kpc is approximately 50% smaller than many similar mass estimates previously published. Given our MAGPHYS models, M31 is similar to the Milky Way in stellar mass.

The inside-out formation scenario suggested by the colour profiles is borne out by our per-pixel SED modelling. M31's bulge is the oldest and most metal-rich stellar component of M31. We find that the bulge began to form 12 Gyr ago, and the mass-weighed average stellar age is $\langle A \rangle = 9 \pm 2$ Gyr. The metallicity is slightly super-solar, with $\log Z/Z_\odot = 0.10 \pm 0.05$.

In the disk, stars began to form 10 ± 2 Gyr ago, with a mass-weighted averaged age $\langle A \rangle = 6 \pm 2$ Gyr. Within uncertainties, we find that the stellar age at, and beyond, $R_{\mathrm{M31}} \geq 10$ kpc is constant. These results suggest an inflection in the stellar age gradient at $R_{\mathrm{M31}} = 10$ kpc, coincident with M31's main star forming ring structure. A plausible scenario, suggested by Athanassoula & Beaton (2006), is that the 10 kpc ring is a resonant structure associated with M31's boxy (non-spherical) bulge. Thus the 10 kpc ring may divide two distinct mixing zones: stars outside the ring are well-mixed and have a uniform age, while stars interior to 10 kpc are migrating between the bulge and the ring.

This stellar population inflection is *not* seen in metallicity maps, however. Instead, the median radial metallicity gradient is relatively mild, -0.13 dex kpc^{-1}, though the metallicity profile comes with rather large error bars. Part of this uncertainty is driven by a near-far effect that is related to the classic age-metallicity-dust parameter degeneracy. Nonetheless, in the median profile behaviour, we find that the mid-disk has sub-solar metallicities (-0.2 ± 0.2 dex at 10 kpc). Again, this metallicity gradient supports the inside-out formation model for M31 where the younger outer regions of M31 were originally built from relatively pristine gas-infall and have had fewer generations of star formation to enrich the ISM.

By leveraging UV and IR data from the literature, the ANDROIDS dataset is directly sensitive to both the absorption and emission effects of M31's dusty ISM. The principle advantages of ANDROIDS over similar studies (Viaene et al. 2014; Draine et al. 2014) is improved optical $u^* g' r' i'$ imaging over SDSS, and the presence of J and K_s band photometry that is missing from other studies. We find that the total dust mass within 20 kpc is $\mathcal{M}_\mathrm{d} = 1.93 \pm 0.01 \times 10^7$ $\mathrm{M}_\odot$. The bulge is essentially dust-free, both in terms of mass ($\log_{10} \mathcal{M}_\mathrm{d}/\mathcal{M}_* < 5$) and optical depth $\tau_V^\mathrm{ISM} \sim 0.01$. $\log_{10} \mathcal{M}_\mathrm{d}/\mathcal{M}_*$ peaks at the 10 kpc ring, where we find that ISM has an optical depth of $0.2 < \tau_V^\mathrm{ISM} < 0.3$, but current star formation in the 10 kpc ring is shrouded in optically-thick dust ($1 < \tau_V < 2.5$).

8.4 Comparison of Stellar Population Modelling Methods

Beyond improving our understanding of M31, our overall motivation for this work is to understand how stellar populations are interpreted by different modelling methods. In doing so, we can understand how distant stellar populations, where only techniques like broadband SED modelling operate, would contrast with pictures derived by other methods such as CMD modelling of resolved stellar populations.

In Chapter 6 we modelled spatially-resolved broadband SEDs across M31's bulge and disk, and contrasted those results with matched results from the literature. In Chapter 7 we additionally attempted to model CMDs with resolved stellar populations. Here we summarize and contrast these various perspectives.

8.4.1 Stellar Mass Estimation

We find that stellar mass estimates, though quoted rather precisely, are somewhat inaccurate with modern methods yielding a range of total M31 stellar masses from 4.9×10^{10} $M_{\odot}$ to 11.2×10^{10} $M_{\odot}$ (Table 6.5). In general, the largest absolute uncertainties in stellar mass estimation come from the assumed IMF and from distance estimation (the latter has not been relevant in this study of M31, but can significantly impact extragalactic surveys in general). Nonetheless, the > 0.3 dex spread in stellar mass estimates seen in this work is irrespective of IMF variations, and instead depends largely on the modelling method and assumptions in the stellar population modelling.

We see differences between stellar mass estimates made with resolved stellar populations (Williams et al. 2017) and those made by integrated SEDs (MAGPHYS, da Cunha et al. 2008) or colour mass-to-light ratio relations (CMLRs, Zibetti et al. 2009; Taylor et al. 2011; Into & Portinari 2013; Roediger & Courteau 2015; Zhang et al. 2017). Overall the Williams et al. $\log_{10} \mathcal{M}_*/L_i$ profile is 0.2 dex higher than the ANDROIDS profile estimated with MAGPHYS. Both profiles have a similar gradient across the mid disk, with the MAGPHYS $\log_{10} \mathcal{M}_*/L_i$ turning upwards near $R_{\mathrm{M31}} = 20$ kpc. The latter is most interesting because CMLRs rely on a non-degenerate relationship between colour and stellar mass-to-light ratio: the Williams et al. (2017) radial $\log_{10} \mathcal{M}_*/L_i$ profile does not track colour gradients seen by ANDROIDS. ANDROIDS photometry shows a 0.1 mag decrease in $g-i$ colour from 10 kpc to 20 kpc, which leads to a decrease in $\log_{10} \mathcal{M}_*/L_i$ of 0.06 dex (for shallower CMLRs such as Zhang et al. 2017, FSPS Model A) to 0.10 (for steeper CMLRs Zibetti et al. 2009).

Multiple scenarios could reconcile these two perspectives. One scenario is for the ANDROIDS colour profile to be incorrect at a level of 0.1 mag arcsec^{-2}. Another scenario is that the colours of integrated SEDs do not map as closely to $\mathcal{M}_*/L$ as most CMLRs portray, and that the intrinsic accuracy of CMLRs is much larger than the 0.1 dex often quoted. A third scenario is that the varying levels of crowding across M31's disk bias CMD modelling in ways that artificial star tests cannot fully compensate for. While the ANDROIDS survey enables us to recognize this discrepancy, resolving it likely requires CMD and SED fitting to realistic mock galaxy datasets where the true stellar population is known.

The most important effect though is the overall normalization difference between the Williams et al. (2017) and MAGPHYS/ANDROIDS $\log_{10} \mathcal{M}_*/L_i$ profiles. This is related to the older overall stellar age identified from resolved stellar populations compared to those from parameterized star formation histories fit with SEDs. In other words, the accuracy SED-based stellar mass estimates are fundamentally limited by the ability for parameterized star formation histories to reflect the true stellar populations of galaxies.

CMLRs provide a useful illustration of biases induced by choices of stellar population synthesis parameterization. Given a consistent IMF, CMLRs differ in their population synthesis model, whether chemical evolution is allowed, and overall distribution of star formation histories considered. Surprisingly, we did not see a consistent difference in stellar mass estimated by CMLRs based on Bruzual & Charlot (2003) population synthesis and the newer Charlot & Bruzual (2007) or Conroy et al. (2009) (FSPS) models. CMLRs that allow for chemical evolution across the star formation history, such as Into & Portinari (2013) and Zhang et al. (2017), tend to have higher $\log_{10} \mathcal{M}_*/L_i$. CMLRs built from model libraries that do not emphasize recent star formation also tend to have higher $\log_{10} \mathcal{M}_*/L_i$ (Zhang et al. 2017). Once again, these trends are not well-defined. The Roediger & Courteau (2015) CMLR built with FSPS population synthesis but using the MAGPHYS star formation history library behaved similarly to the Zhang et al. (2017) CMLR that also used FSPS but instead used star formation histories observed in Local Group dwarf galaxies. The best way to resolve this conundrum is with a fully controlled experiment, where population synthesis models and the distribution of model star formation histories are systematically varied.

Full SED fits that assume a prior distribution of model SEDs (MAGPHYS modelling, for example) are also affected. Indeed, the stellar mass estimation from our full SED MAGPHYS models was nearly identical to that obtained from the Zibetti et al. (2009) CMLR, which was constructed from the same underlying population synthesis model and library distribution. In other words, stellar mass can be estimated equally well with integrated light as with a single optical colour, such as $g - i$. The main limitation, in both cases, is in the distributions of assumed star formation histories.

8.4.2 Stellar Population Estimation

Next we compare the relative performance of broadband SED modelling methods and resolved stellar population modelling for characterizing a galaxy's age, metallicity, and dust properties. Note that CMD modelling methods benefit from richer input data to model the star formation histories in greater detail. In this section we primarily concerned with comparing the general characterizations provided by each method, such as mean age and mean metallicity.

Overall, our MAGPHYS SED modelling yields mean stellar ages that are about 2 Gyr younger than those estimated by Williams et al. (2017) with PHAT resolved photometry. Further, the stellar metallicity estimated by MAGPHYS is statistically equivalent to the mass-weighted mean metallicity estimated from Williams et al. (2017) CMD fitting, though the MAGPHYS metallicity is highly uncertain (ranging between -0.4 and 0.1 at 10 kpc). Thus modelling with broadband SEDs should be treated as broad characterizations of a stellar population (for example categorizing metallicities as super-solar, sub-solar, and metal-poor metallicity, and characterizing ages as old, intermediate, or young).

Metallicity is the most finicky parameter to estimate with integrated SEDs. Despite a well-sampled SED, the metallicity map measured with MAGPHYS shows a clear near-far bias indicative of a degeneracy between metallicity and attenuation parameters. Additionally, the median MAGPHYS metallicity radial profile is a steeper gradient than that estimated from resolved stellar populations by both Gregersen et al. (2015) and Williams et al. (2017). At the same time, assuming metallicity parallels dust mass, the MAGPHYS metallicity profile is well-predicted by the dust mass profile, also fit from SEDs.

8.4.3 Discussion

Androids has provided a unique perspective on stellar population modelling with either resolved stellar populations or broadband SEDs. We have validated that broadband SED modelling is a useful method, though it is afflicted by substantial systematic uncertainties to star formation history and metallicity that also induce significant biases in stellar mass estimates.

SED modelling could be further developed on multiple fronts. We found that the colours predicted by the MAGPHYS model library did not completely overlap the space of colours observed with androids photometry. Secondly, results from most modern SED modelling methods are limited by prior assumptions in the distribution of star formation histories. These priors include the distributions of star formations histories and the allowance of chemical evolution (as opposed to the conventional approximation of monometallic stellar populations). At the same time, broadband SED modelling benefits from having fewer free parameters and well-sampled model libraries.

A way forward might be to develop SED modelling with priors that are tuned to a particular galaxy, or at least a given type of galaxy. The MAGPHYS model library, for example, is relatively uninformative and accommodates a universal perspective of galaxy evolution. By incorporating narrowband imaging of nebular emission (such as the Hα line), for example, it may be possible to better constrain the very recent star formation history, which might in turn improve estimates of dust and other parameters (Leja et al. 2017). In addition, results from SED modelling benefit from expanded broadband filter sets as we see from our own testing. Incorporating improved and expanded UV and near-IR datasets (including WISE) can better constrain both the recent (10 Myr) and time-integrated star formation. It might also be possible to leverage galaxy formation and evolution modelling to develop priors of star formation histories and chemical evolution, similarly to how Zhang et al. (2017) used the empirical distribution of star formation histories and chemical evolution in Local Group dwarf galaxies to establish realistic colour-$\mathcal{M}_*/L$ relations. This expansion in the parameterization of galaxy star formation histories, combined with continued improvements in stellar isochrone modelling, should ultimately eliminate discrepancies in observed and modelled galaxy colours, and reduce the bias in broadband SED models compared to resolved stellar population modelling.

At the same time, there is great opportunity to improve star catalog modelling methods. By attempting to independently reproduce star formation models from the PHAT star catalog, we found that CMDs have limited inference power. In Chapter 7 we attempted to independently reproduce star formation history models from the PHAT star catalog and found little success. Our mock trials showed that CMD fits are greatly affected by crowding, even when artificial star testing is used. We could only recover the star formation history of the outermost PHAT brick, 20 kpc from the centre of M31. With increased crowding, our mock fits tended to be biased towards younger mean ages. Our attempts to independently reproduce star formation models from the PHAT star catalog in Chapter 7 might be characterized as simplistic compared to the studies by the PHAT team, including Lewis et al. (2015) and Williams et al. (2017). For example, in Chapter 7 we were generally unable to constrain a chemical evolution history, unlike Williams et al. (2017). While these results reflect poorly on our adopted method, they do so as well on the state of resolved stellar population modelling in general whereby, given the same data, results are not readily reproducible. We found that CMD fits require a great deal of tuning, including choosing how to bin the Hess plane, bin isochrones, and even distribute dust extinctions. Further, most modern CMD fitting studies use the venerable *MATCH* code Dolphin (2002), which is neither open source nor readily available upon request.

Considering this, we believe that resolved stellar population studies can benefit greatly from the development of modelling technology that is generally more usable and portable. An important step in this direction is the replacement of downhill simplex-based optimization of Hess planes with statistically-justified Bayesian modelling. Examples of this work are van Dyk et al. (2009) and Dalcanton et al. (2015). An advantage of these Bayesian methods, which do not fit Hess planes, is that they take full advantage of multi-band photometric catalogs that are commonly made with modern surveys. We found that that our mock fits to one Hess plane would not reproduce the mock plane of the same stellar catalog projected on a different Hess plane. Bayesian fitting methods would possibly prevent this issue.

8.5 Future work

In this work, we have focused on using the ANDROIDS dataset for integrated SED modelling. Our CFHT MegaCam and WIRCam imaging also have sufficient image quality to support resolved stellar population studies. While the PHAT star catalog (Williams et al. 2014) authoritatively covers the optical disk of M31, an ANDROIDS star catalog would provide complementary coverage of M31's outer disk in similar bandpasses. An example of one such study, a survey of Carbon AGB stars, is described in Appendix A.

Bibliography

Adams, J. D., & Skrutskie, M. F. 1996, Airglow and 2MASS Survey Strategy, `http://astsun.astro.virginia.edu/~mfs4n/2mass/airglow/airglow.html`

Aniano, G., Draine, B. T., Gordon, K. D., & Sandstrom, K. 2011, PASP, 123, 1218

Aparicio, A., & Gallart, C. 2004, AJ, 128, 1465

Aparicio, A., & Hidalgo, S. L. 2009, AJ, 138, 558

Aringer, B., Girardi, L., Nowotny, W., Marigo, P., & Lederer, M. T. 2009, A&A, 503, 913

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33

Athanassoula, E., & Beaton, R. L. 2006, MNRAS, 370, 1499

Barmby, P., Ashby, M. L. N., Bianchi, L., et al. 2006, ApJL, 650, L45

Battinelli, P., Demers, S., & Letarte, B. 2003, AJ, 125, 1298

Battinelli, P., Demers, S., & Mannucci, F. 2007, A&A, 474, 35

Beaton, R. L., Majewski, S. R., Guhathakurta, P., et al. 2007, ApJL, 658, L91

Bell, E. F., & de Jong, R. S. 2001, ApJ, 550, 212

Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. D. 2003, ApJS, 149, 289

Bernard, E. J., Ferguson, A. M. N., Barker, M. K., et al. 2012, MNRAS, 420, 2625

Berriman, G. B., Good, J. C., Laity, A. C., & Kong, M. 2008, in Astronomical Society of the Pacific Conference Series, Vol. 394, Astronomical Data Analysis Software and Systems XVII, ed. R. W. Argyle, P. S. Bunclark, & J. R. Lewis, 83

Bertin, E. 2006, in Astronomical Society of the Pacific Conference Series, Vol. 351, Astronomical Data Analysis Software and Systems XV, ed. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique, 112

Bertin, E. 2007, SWarp, Terapix

Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393

Bertin, E., Mellier, Y., Radovich, M., et al. 2002, in Astronomical Society of the Pacific Conference Series, Vol. 281, Astronomical Data Analysis Software and Systems XI, ed. D. A. Bohlender, D. Durand, & T. H. Handley, 228

Binney, J., & Merrifield, M. 1998, Galactic Astronomy (Princeton University Press)

Blanton, M. R., & Roweis, S. 2007, AJ, 133, 734

Block, D. L., Bournaud, F., Combes, F., et al. 2006, Nature, 443, 832

Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127

Brinchmann, J., Charlot, S., Heckman, T. M., et al. 2004, ArXiv Astrophysics e-prints, astro-ph/0406220

Brown, T. M. 2009, in Astronomical Society of the Pacific Conference Series, Vol. 419, Galaxy Evolution: Emerging Insights and Future Challenges, ed. S. Jogee, I. Marinova, L. Hao, & G. A. Blanc, 110

Brown, T. M., Ferguson, H. C., Smith, E., et al. 2003, ApJL, 592, L17

Brown, T. M., Smith, E., Ferguson, H. C., et al. 2006, ApJ, 652, 323

Bruzual, G. 2007, in IAU Symposium, Vol. 241, Stellar Populations as Building Blocks of Galaxies, ed. A. Vazdekis & R. Peletier, 125

Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000

Calabretta, M. R., & Greisen, E. W. 2002, A&A, 395, 1077

Calzetti, D. 2013, Star Formation Rate Indicators, ed. J. Falcón-Barroso & J. H. Knapen, 419

Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682

Chabrier, G. 2003, PASP, 115, 763

Charlot, S., & Fall, S. M. 2000, ApJ, 539, 718

Chemin, L., Carignan, C., & Foster, T. 2009, ApJ, 705, 1395

Chen, Y., Girardi, L., Bressan, A., et al. 2014, MNRAS, 444, 2525

Choi, P. I., Guhathakurta, P., & Johnston, K. V. 2002, AJ, 124, 310

Cioni, M.-R. L. 2009, A&A, 506, 1137

Cluver, M. E., Jarrett, T. H., Hopkins, A. M., et al. 2014, ApJ, 782, 90

Conroy, C. 2013, ARAA, 51, 393

Conroy, C., & Gunn, J. E. 2010, ApJ, 712, 833

Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486

Conroy, C., White, M., & Gunn, J. E. 2010, ApJ, 708, 58

Corbelli, E., Lorenzoni, S., Walterbos, R., Braun, R., & Thilker, D. 2010, A&A, 511, A89

Cortese, L., Ciesla, L., Boselli, A., et al. 2012, A&A, 540, A52

Courteau, S., de Jong, R. S., & Broeils, A. H. 1996, ApJL, 457, L73

Courteau, S., Widrow, L. M., McDonald, M., et al. 2011, ApJ, 739, 20

Courteau, S., Cappellari, M., de Jong, R. S., et al. 2014, Rev. Mod. Phys., 86, 47

da Cunha, E., Charlot, S., & Elbaz, D. 2008, MNRAS, 388, 1595

Dalcanton, J. J., Williams, B. F., Lang, D., et al. 2012, ApJS, 200, 18

Dalcanton, J. J., Fouesneau, M., Hogg, D. W., et al. 2015, ApJ, 814, 3

de Jong, R. S. 1996, A&A, 313, 377

Demers, S., Battinelli, P., & Letarte, B. 2003, AJ, 125, 3037

Dolphin, A. 2016, DOLPHOT: Stellar photometry, Astrophysics Source Code Library, ascl:1608.013

Dolphin, A. E. 2000, PASP, 112, 1383

—. 2002, MNRAS, 332, 91

—. 2012, ApJ, 751, 60

—. 2013, ApJ, 775, 76

Draine, B. T., Dale, D. A., Bendo, G., et al. 2007, ApJ, 663, 866

Draine, B. T., Aniano, G., Krause, O., et al. 2014, ApJ, 780, 172

Dunne, L., Eales, S., Edmunds, M., et al. 2000, MNRAS, 315, 115

Elmegreen, B. G., & Block, D. L. 1999, MNRAS, 303, 133

Eskew, M., Zaritsky, D., & Meidt, S. 2012, AJ, 143, 139

Fardal, M. A., Guhathakurta, P., Babul, A., & McConnachie, A. W. 2007, MNRAS, 380, 15

Feast, M. W., Abedigamba, O. P., & Whitelock, P. A. 2010, MNRAS, 408, L76

Ferguson, A. M. N., Irwin, M. J., Ibata, R. A., Lewis, G. F., & Tanvir, N. R. 2002, Astronomical Journal, 124, 1452

Ferguson, A. M. N., Johnson, R. A., Faria, D. C., et al. 2005, ApJL, 622, L109

Ferrarese, L., Côté, P., Cuillandre, J.-C., et al. 2012, ApJS, 200, 4

Foreman-Mackey, D., Sick, J., & Johnson, B. 2014, python-fsps: Python bindings to FSPS (v0.1.1)

Fritz, J., Gentile, G., Smith, M. W. L., et al. 2012, A&A, 546, A34

Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996, AJ, 111, 1748

Geehan, J. J., Fardal, M. A., Babul, A., & Guhathakurta, P. 2006, MNRAS, 366, 996

Geman, S., & Geman, D. 1984, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6, 721

Gil de Paz, A., Boissier, S., Madore, B. F., et al. 2007, ApJS, 173, 185

Girardi, L., Groenewegen, M. A. T., Hatziminaoglou, E., & da Costa, L. 2005, A&A, 436, 895

Gordon, K. D., Clayton, G. C., Misselt, K. A., Landolt, A. U., & Wolff, M. J. 2003, ApJ, 594, 279

Gordon, K. D., Bailin, J., Engelbracht, C. W., et al. 2006, ApJL, 638, L87

Gordon, K. D., Fouesneau, M., Arab, H., et al. 2016, ApJ, 826, 104

Gregersen, D., Seth, A. C., Williams, B. F., et al. 2015, AJ, 150, 189

Greisen, E. W., & Calabretta, M. R. 2002, A&A, 395, 1061

Groves, B., Krause, O., Sandstrom, K., et al. 2012, MNRAS, 426, 892

Guhathakurta, P., Ostheimer, J. C., Gilbert, K. M., et al. 2005, ArXiv Astrophysics e-prints, arXiv:astro-ph/0502366

Harris, J., & Zaritsky, D. 2001, ApJS, 136, 25

Hastings, W. K. 1970, Biometrika, 97

Hills, S., von Hippel, T., Courteau, S., & Geller, A. M. 2015, AJ, 149, 94

Hunter, J. D. 2007, Computing in Science and Engineering

Ibata, R., Chapman, S., Ferguson, A. M. N., et al. 2005, ApJ, 634, 287

Into, T., & Portinari, L. 2013, MNRAS, 430, 2715

Irwin, M. J., Ferguson, A. M. N., Ibata, R. A., Lewis, G. F., & Tanvir, N. R. 2005, ApJL, 628, L105

Kacharov, N., Rejkuba, M., & Cioni, M.-R. L. 2012, A&A, 537, A108

Kang, Y., Bianchi, L., & Rey, S.-C. 2009, ApJ, 703, 614

Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, MNRAS, 341, 33

Kettlety, T., Hesling, J., Phillipps, S., et al. 2017, MNRAS, 473, 776

Kong, X., Charlot, S., Brinchmann, J., & Fall, S. M. 2004, MNRAS, 349, 769

Kormendy, J., & Kennicutt, Jr., R. C. 2004, ARAA, 42, 603

Kroupa, P. 2001, MNRAS, 322, 231

Leja, J., Johnson, B. D., Conroy, C., van Dokkum, P. G., & Byler, N. 2017, ApJ, 837, 170

Lewis, A. R., Dolphin, A. E., Dalcanton, J. J., et al. 2015, ApJ, 805, 183

MacArthur, L. A., Courteau, S., Bell, E., & Holtzman, J. A. 2004, ApJS, 152, 175

Madden, S. C., Galliano, F., Jones, A. P., & Sauvage, M. 2006, A&A, 446, 877

Maraston, C. 1998, MNRAS, 300, 872

—. 2005, MNRAS, 362, 799

Maraston, C., Daddi, E., Renzini, A., et al. 2006, ApJ, 652, 85

Marigo, P., & Girardi, L. 2007, A&A, 469, 239

Marigo, P., Girardi, L., Bressan, A., et al. 2008, A&A, 482, 883

Marmo, C., & Bertin, E. 2008, in Astronomical Society of the Pacific Conference Series, Vol. 394, Astronomical Data Analysis Software and Systems XVII, ed. R. W. Argyle, P. S. Bunclark, & J. R. Lewis, 619

Massey, P., Olsen, K. A. G., Hodge, P. W., et al. 2006, AJ, 131, 2478

Matteucci, F., & Francois, P. 1989, MNRAS, 239, 885

McConnachie, A. W., Irwin, M. J., Ferguson, A. M. N., et al. 2005, MNRAS, 356, 979

McConnachie, A. W., Irwin, M. J., Ibata, R. A., et al. 2009, Nature, 461, 66

Meidt, S. E., Schinnerer, E., van de Ven, G., et al. 2014, ApJ, 788, 144

Melbourne, J., & Boyer, M. L. 2013, ApJ, 764, 30

Melbourne, J., Williams, B. F., Dalcanton, J. J., et al. 2012, ApJ, 748, 47

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, The Journal of Chemical Physics, 21, 1087

Monet, D. G., Levine, S. E., Canzian, B., et al. 2003, AJ, 125, 984

Nelder, J. A., & Mead, R. 1965, The Computer Journal, 7, 308

Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713

Peterson, P. 2009, International Journal of Computational Science and Engineering (IJCSE), 4

Pforr, J., Maraston, C., & Tonini, C. 2012, MNRAS, 422, 3285

Portinari, L., Sommer-Larsen, J., & Tantalo, R. 2004, MNRAS, 347, 691

Press, W. H., Tuekolsky, S. A., Vetterling, W. T., & Flannery, B. P. 2007, Numerical recipes: the art of scientific computing, 3rd edn. (Cambridge, UK: Cambridge University Press)

Puget, P., Stadler, E., Doyon, R., et al. 2004, in Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, Vol. 5492, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, ed. A. F. M. Moorwood & M. Iye, 978

Rafiei Ravandi, M., Barmby, P., Ashby, M. L. N., et al. 2016, MNRAS, 459, 1403

Richardson, J. C., Ferguson, A. M. N., Johnson, R. A., et al. 2008, AJ, 135, 1998

Robitaille, T., & Bressert, E. 2012, APLpy: Astronomical Plotting Library in Python, Astrophysics Source Code Library, ascl:1208.017

Roediger, J. C., & Courteau, S. 2015, MNRAS, 452, 3209

Saglia, R. P., Fabricius, M., Bender, R., et al. 2010, A&A, 509, A61

Salpeter, E. E. 1955, ApJ, 121, 161

Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Searle, L., & Zinn, R. 1978, ApJ, 225, 357

Seigar, M. S., Barth, A. J., & Bullock, J. S. 2008, MNRAS, 389, 1911

Sheth, K., Regan, M., Hinz, J. L., et al. 2010, PASP, 122, 1397

Sibbons, L. F., Ryan, S. G., Cioni, M.-R. L., Irwin, M., & Napiwotzki, R. 2012, A&A, 540, A135

Sick, J., Courteau, S., Cuillandre, J.-C., et al. 2014, AJ, 147, 109

Sirianni, M., Jee, M. J., Benítez, N., et al. 2005, PASP, 117, 1049

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163

Stetson, P. B. 1987, PASP, 99, 191

Tamm, A., Tempel, E., Tenjes, P., Tihhonova, O., & Tuvikene, T. 2012, A&A, 546, A4

Tang, J., Bressan, A., Rosenfield, P., et al. 2014, MNRAS, 445, 4287

Taylor, E. N., Hopkins, A. M., Baldry, I. K., et al. 2011, MNRAS, 418, 1587

Tempel, E., Tuvikene, T., Tamm, A., & Tenjes, P. 2011, A&A, 526, A155

Thilker, D. A., Hoopes, C. G., Bianchi, L., et al. 2005, ApJL, 619, L67

Tully, R. B., & Fisher, J. R. 1977, A&A, 54, 661

Vaduvescu, O., & McCall, M. L. 2004, PASP, 116, 640

van Dyk, D. A., DeGennaro, S., Stein, N., Jefferys, W. H., & von Hippel, T. 2009, Annals of Applied Statistics, 3, 117

Viaene, S., Fritz, J., Baes, M., et al. 2014, A&A, 567, A71

von Hippel, T., Jefferys, W. H., Scott, J., et al. 2006, ApJ, 645, 1436

Walcher, J., Groves, B., Budavári, T., & Dale, D. 2011, Ap&SS, 331, 1

Walterbos, R. A. M., & Kennicutt, Jr., R. C. 1987, A&AS, 69, 311

—. 1988, A&A, 198, 61

Weisz, D. R., Dolphin, A. E., Skillman, E. D., et al. 2014, ApJ, 789, 148

Westera, P., Lejeune, T., Buser, R., Cuisinier, F., & Bruzual, G. 2002, A&A, 381, 524

White, III, R. E., Keel, W. C., & Conselice, C. J. 2000, ApJ, 542, 761

Widrow, L. M., Perrett, K. M., & Suyu, S. H. 2003, ApJ, 588, 311

Williams, B. F. 2002, MNRAS, 331, 293

—. 2003, AJ, 126, 1312

Williams, B. F., Lang, D., Dalcanton, J. J., et al. 2014, ApJS, 215, 9

Williams, B. F., Dolphin, A. E., Dalcanton, J. J., et al. 2017, ApJ, 846, 145

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868

Xilouris, E. M., Alton, P. B., Davies, J. I., et al. 1998, A&A, 331, 894

Zhang, H.-X., Puzia, T. H., & Weisz, D. R. 2017, ApJS, 233, 13

Zibetti, S., Charlot, S., & Rix, H. 2009, MNRAS, 400, 1181

Zijlstra, A. A., Matsuura, M., Wood, P. R., et al. 2006, MNRAS, 370, 1961

Appendix A

Future Study: Asymptotic Giant Branch Star Classification with a Narrow-band TiO and CN Filter Survey of M31

In this Appendix, we describe an extension of the ANDROIDS survey for classifying resolved asymptotic giant branch (AGB) stars in M31 with CFHT narrowband imaging.

A.1 Motivation

Near-infrared (near-IR) photometry has become an essential tool for lifting degeneracies between dust and the effects of stellar ages and metallicities on the optical colours of galaxies, and for estimating stellar masses. Meanwhile, $\sim 25\%$ of the K-band light (LMC/SMC, Melbourne & Boyer 2013) is contributed by giant stars undergoing rapid ($\sim 10^6$ yr) evolutionary phases: luminous red helium burning stars, and thermally-pulsating asymptotic giant branch (TP-AGB) stars. As the AGB sequence itself evolves, the third dredge-up increases the carbon/oxygen (C/O) ratio in these stellar atmospheres, transitioning a star from an M-type to become a *redder*

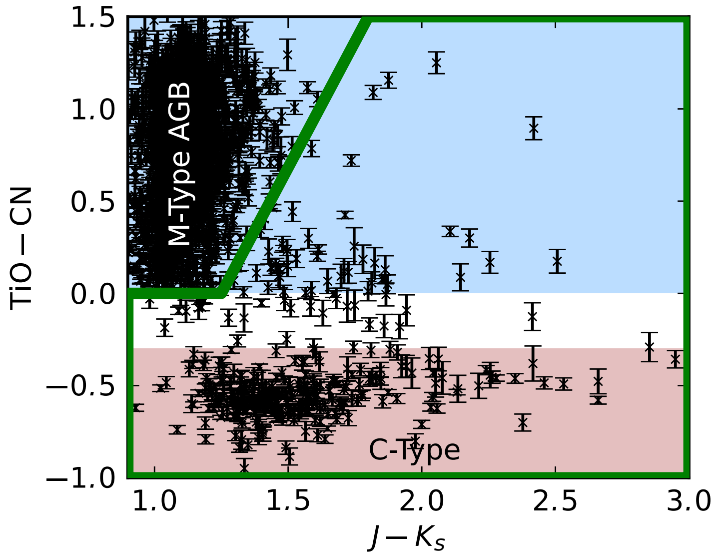


Figure A.1: Classification of M- and C-type AGB stars using ANDROIDS $J - K_s$ photometry combined with CFH12K TiO − CN (kindly provided by S. Demers) in the Battinelli et al. (2003) SW1 field. Here only M31 stars brighter than the tip of the RGB are selected. Red and blue boxes show canonical TiO − CN selection criteria for C- and M-types. The green boundary separates likely C stars from the main locus of bluer M stars. Colour functions of these populations, with both selection criteria, are shown in Figure A.3.

C-type AGB star. Lower metallicity stars can more rapidly become C-types since fewer dredge up cycles are required to alter the photospheric C/O ratio, thus linking a population's C/M ratio to metallicity (*e.g.,* Cioni 2009); there may also be an age dependence to the C/M ratio (Feast et al. 2010). As pointed out by Conroy (2013), while some stellar population models (*e.g.,* Marigo & Girardi 2007; Marigo et al. 2008) are well-tuned to the TP-AGB of the LMC, they fail to predict the near-IR light in other metallicity and age regimes. For instance, Melbourne et al. (2012) find that younger populations in local dwarfs can have TP-AGB fluxes overestimated by 2×, yielding equally wrong estimates of age, metallicity, and stellar mass. Calibrating stellar population models at both optical and near-IR wavelengths over a wide

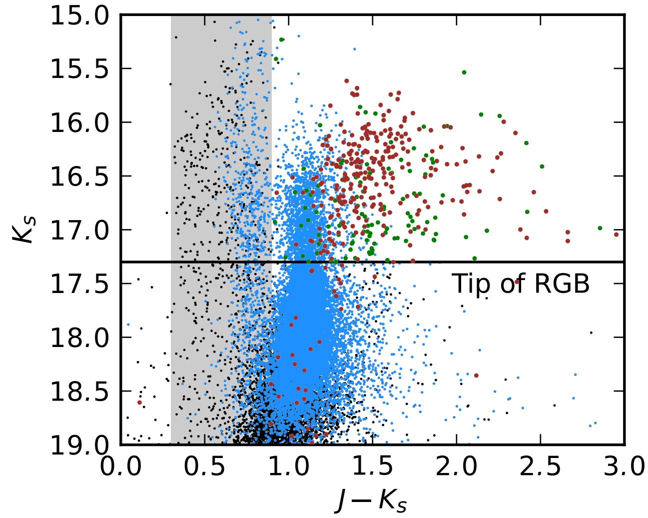


Figure A.2: ANDROIDS CMD of stars cross-correlated with the Battinelli et al. (2003) SW1 dataset. Blue stars are identified by Battinelli et al. as M-type giants, and red stars are identified as C-type AGB stars using canonical TiO − CN criteria (blue and red boxes in Figure A.1). Green stars are additional C-star candidates given the green selection region in Figure A.1. The horizontal line denotes the tip of the red giant branch. Foreground Milky Way stars (in grey region) cleanly separate from M31 giants with $J - K_s$ photometry.

range of environments is a tremendous challenge largely in light of the uncertain AGB contribution to the near-IR light.

Our goal in this study is to provide observational constraints on the composition of near-IR light by accurately decomposing populations of M- and C-type AGBs, thereby providing crucial additional physical constraints for stellar population modellers. Because of the extreme stochasticity in the resolved TP-AGB populations of dwarfs, we look to M31 which hosts large populations of C and M-type AGB stars in various local age and metallicity regimes, also representative of stellar populations in other galaxies.

Calibrating the contribution of AGB stars to the near-IR right requires parallel

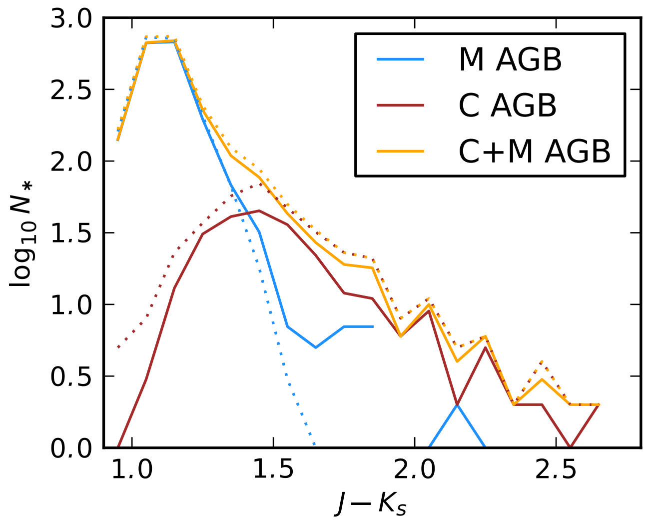


Figure A.3: $J - K_s$ colour function of C- and M-type AGB stars from Figure A.1 and Figure A.2. Solid lines are based on canonical TiO − CN-only identifications, while dotted lines use our hybrid narrow-band and near-IR criteria (green region in Figure A.1). Note the overlap between M- and C-type AGB stars: M- and C-type AGB populations cannot be decomposed from $J - K_s$ photometry alone; TiO − CN classifications are required. Likewise, TiO − CN criteria alone find too many extremely-red M-type AGBs Zijlstra et al. (2006). The proper colour function of C/M stars ultimately requires our hybrid narrow-band and near-IR selection (yielding the dotted functions).

datasets that describe both spatially-resolved surface brightness (spectral energy distribution, SED) and resolved photometry of the stars that contribute to that SED. The core ANDROIDS dataset, described in Chapters 2–5, is well-suited for the surface brightness aspect. For our Andromeda Optical and Infrared Disk Survey (ANDROIDS) with CFHT MegaCam/WIRCam, we have assembled a panchromatic dataset covering the entire disk of M31. In Sick et al. (2014) (Chapter 2), we solved many issues in wide-field WIRCam surface photometry, allowing us to construct the first first authentic near-IR surface brightness profiles of M31 out to $R_{\mathrm{M31}} = 20$ kpc. These surface brightness maps allow us to *directly* measure the total near-IR flux in $i'JK_s$

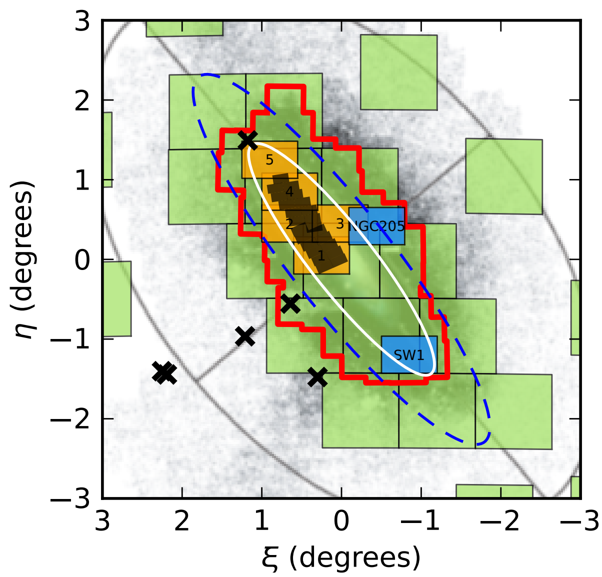


Figure A.4: Map of narrow-band MegaCam fields (orange) compared to the existing CFH12K SW1 and NGC 205 fields (blue) and the ANDROIDS WIRCam (red outline) and MegaCam (green) footprints. The narrow-band fields cover the PHAT survey (dark grey region) and reach out to 25 kpc (white ellipse) to measure the C/M ratio in a range of metallicity regimes. Black crosses are Brown (2009) deep HST fields. These fields are plotted against the star count map of Ibata et al. (2005).

bands, unlike Melbourne et al. (2012) and others who must integrate their modelled stellar population.

The remaining challenge is detecting and robustly classifying C- and M-type AGB stars to directly measure their contribution to the M31 SED.

A.2 AGB Classification with $J - K_s$ and Narrowband TiO and CN Photometry

To identify C- and M-type AGB stars, we required resolved stellar photometry. The CFHT/WIRCam dataset described in Chapter 2 was obtained under ideal seeing conditions, $\sim 0.6''$ seeing, making point-source J and K_s photometry possible in M31's mid and outer disk. The fundamental difficulty of establishing homogeneous C/M ratios is defining efficient M- and C-type AGB selection criteria. Carbon stars are typically detected with low resolution spectroscopy, broadband near-IR photometry, or TiO − CN narrow-band imaging as a proxy for the photospheric C/O ratio; the former being untenable in a wide-field survey.

Using TiO − CN data graciously provided by S. Demers, we have cross-correlated our ANDROIDS $J - K_s$ WIRCam photometry with a portion of the SW1 CFH12K field from Battinelli et al. (2003). Figure A.1 shows our photometry in the $J - K_s$ vs TiO − CN plane, while Figure A.2 shows classified M- and C-Type stars in our $J - K_s$ CMD. Note the substantial overlap in the $J - K_s$ colours of C- and M-types: only with TiO − CN imaging can C- and M-type AGB stars be decomposed. Note also how Figure A.1 exposes a third population of objects that, given their red $J - K_s$ colours, are likely carbon stars Kacharov et al. (2012). These objects are clearly visible for the first time because of the larger statistics offered by M31 AGB stars (cf., Battinelli et al. 2007). Thus we propose a new C-type selection criteria in Figure A.1 and plot these additional C-types as green points in Figure A.2. This result mirrors the conclusion of Sibbons et al. (2012), whereby tight C/M selection criteria can only be achieved with combined TiO–CN and near-IR data sets. Figure A.1 represents the first major step in this direction.

A.3 The Extended ANDROIDS TiO and CN Survey

To pursue this study, we obtained narrowband TiO and CN observations with the CFHT/MegaCam during the 2013B semester in fields shown in Figure A.4. Note that since CFH12K narrow-band filters are used, these MegaCam fields cover only

$42' \times 28'$. The radial distribution of these fields allows the metallicity dependence of the C- and M-type colour function to be constrained locally (a crucial aspect of this program!). Moreover, these fields overlap the PHAT fields, whose HST photometry becomes especially valuable in crowded fields of the inner disk and bulge. The very deep PHAT CMDs also provide the most accurate estimates on local age and metallicity. One of our fields partially overlaps the NGC 205 field by Demers et al. (2003) in order to enable photometric uniformity across all fields.

Each field is observed with 3×10 min = 30 minute integrations in both the TiO and CN filters We will use a 3-point medium dither pattern to cover small gaps between the chips. This procedure follows the successful CFH12K programs of Battinelli et al. (2003) and Demers et al. (2003).

Zeropoints for the TiO and CN images will be internally determined, rather than requiring dedicated calibration fields. Following Battinelli et al. (2003), and others, the zeropoint of the TiO − CN index are set by hot stars which, by convention, have TiO − CN = 0.

A.4 Summary

This dataset enables a future study of AGB stars. With these additional TiO − CN fields, combined with our existing ANDROIDS near-IR photometry and PHAT's very deep CMDs, we will unveil several key constraints for AGB modelling as functions of the ages and metallicities of the local population such as: a) local C/M ratios b) luminosity functions of the M- and C-type AGB populations, and c) direct measurement of the fractional iJK_s light contributed by M- and C-type AGB stars. These will empirically constrain mass-loss rates, efficiency of convection, and the role played by circumstellar envelopes in stellar models. Determining AGB luminosity contributions, as functions of age and metallicity, is one of the most pressing quests in modern stellar population modelling.