University of Groningen

Physical drivers of the cosmic star formation history

Pearson, William James

DOI:
[10.33612/diss.101445849](https://doi.org/10.33612/diss.101445849)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Pearson, W. J. (2019). *Physical drivers of the cosmic star formation history*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.101445849>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Physical Drivers of the Cosmic Star Formation History

William J. Pearson

rijksuniversiteit
 groningen

Physical Drivers of the Cosmic Star Formation History

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. C. Wijmenga
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

vrijdag 22 november 2019 om 16.15 uur

door

William James Pearson

geboren op 17 november 1992
te Leicester, Verenigd Koninkrijk

Promotor

Prof. dr. F.F.S. van der Tak

Copromotor

Dr. L. Wang

Beoordelingscommissie

Prof. dr. P.D. Barthel

Prof. dr. H. Dole

Prof. dr. S. Eales

The research leading to this thesis was funded by SRON, The Netherlands Institute for Space Research

Cover Design: William J. Pearson

Cover Background: gri composite image using data from the Kilo Degree Survey

Front Cover Image: Cartoon of the *Herschel* Space Observatory, William J. Pearson

Rear Cover Image Credit: NGC 6872, ESA/Hubble & NASA

Rear Cover Acknowledgement: Judy Schmidt (geckzilla.com)

Printed by: Ipskamp Printing

ISBN 978-94-034-2128-5 (Printed Version)

ISBN 978-94-034-2127-8 (Electronic Version)

Contents

List of Figures	ix
List of Tables	xxi
Nederlandse samenvatting	xxv
English summary	xxxi
1 Introduction	1
1.1 Star formation	2
1.1.1 Star formation rate tracers	3
1.1.2 Star formation over cosmic time	5
1.1.3 The mass dependence of star formation rates	7
1.1.4 Galaxy mergers and their influence on star formation	9
1.2 Data used within this thesis	12
1.2.1 Far infrared for star formation rates	12
1.2.2 Identifying mergers	13
1.3 Methodology and tools employed within this thesis	14
1.3.1 Dealing with <i>Herschel</i> confusion	14
1.3.2 Machine learning	16
1.4 This thesis	18
2 De-blending deep <i>Herschel</i> surveys:	
A multi-wavelength approach	27
2.1 Introduction	28
2.2 Data	28
2.3 Methodology	30
2.3.1 Tools	30
2.3.2 Extracting the SPIRE flux densities	30
2.4 Results	31
2.4.1 CIGALE and XID+ performance	31
2.4.2 Comparison of ALMA data to CIGALE SEDs	32
2.4.3 Model dependance	35
2.5 Conclusions	38
2.6 Appendix	42
2.6.1 Homogeneity of COSMOS	42

2.6.2	CIGALE Parameters	44
3	Main sequence of star forming galaxies beyond the <i>Herschel</i> confusion limit	47
3.1	Introduction	48
3.2	Data sets	50
3.3	Methodology	52
3.3.1	Tools	52
3.3.2	Science pipeline	53
3.3.3	Mass completeness and redshift binning	54
3.3.4	Forward modelling	55
3.4	Results	57
3.5	Discussion	63
3.5.1	Form of the main sequence	63
3.5.2	Contamination by quiescent galaxies	63
3.5.3	Comparisons with other works	65
3.6	Conclusions	68
3.7	Appendix	74
3.7.1	CIGALE Parameters	74
3.7.2	XID+ prior width	78
3.7.3	Further comparisons with Schreiber et al. (2015)	78
4	Identifying galaxy mergers in observations and simulations with deep learning	83
4.1	Introduction	84
4.2	Image data	85
4.2.1	SDSS images	85
4.2.2	EAGLE images	86
4.3	Deep learning	88
4.3.1	Convolutional neural networks	88
4.3.2	Architecture	90
4.3.3	Training, validation and testing	94
4.4	Results and discussion	94
4.4.1	Observation trained network	95
4.4.2	Simulation trained network	97
4.4.3	Cross application of the networks	107
4.4.4	Differences in network merger identification	120
4.5	Conclusions	122
4.6	Appendix	128
4.6.1	Image colour scaling	128
5	Effect of galaxy mergers on star formation rates	131
5.1	Introduction	132
5.2	Data	134
5.2.1	SDSS Data Release 7	135
5.2.2	KiDS	135
5.2.3	CANDELS	141

5.2.4	Mass completeness	146
5.3	Tools	146
5.3.1	Convolutional neural networks	146
5.3.2	Forward modelling of the galaxy main sequence	148
5.4	Results	149
5.4.1	Performance of the CNN	149
5.4.2	SDSS	150
5.4.3	KiDS	153
5.4.4	CANDELS	153
5.5	Discussion	154
5.5.1	Merger influence on SFR	155
5.5.2	Merger fractions	157
5.5.3	Starburst merger fraction	160
5.6	Conclusions	160
5.7	Appendix	170
5.7.1	De-blending the SPIRE data	170
5.7.2	CNN performance definitions	173
5.7.3	Example non-mergers and mergers	173
6	Conclusions and Future Prospects	179
6.1	Conclusions	180
6.1.1	Improvement of existing far-infrared flux density measurement - Chapter 2	181
6.1.2	The relation between star formation rate and stellar mass - Chapter 3	181
6.1.3	Detecting galaxy mergers with deep learning - Chapter 4	182
6.1.4	The relation between galaxy mergers and star formation rate - Chapter 5	183
6.1.5	Overall conclusions	184
6.2	Future prospects	184
6.2.1	Improvements for far infrared flux density measurement	184
6.2.2	The main sequence of star forming galaxies	185
6.2.3	Identifying galaxy mergers	185
	Acknowledgements	189

List of Figures

- 1 Impresie van de *Herschel* Space Observatory. Afbeelding van: ESA/AOES Medialab xxvii
- 2 Samengesteld beeld van de 250 μm , 350 μm en 500 μm *Herschel* SPIRE-gegevens in het COSMOS-veld. Gemaakt met behulp van gegevens van de Herschel Multi-tiered Extragalactic Survey. xxviii
- 3 Een paar fuserende sterrenstelsels (NGC 4676, of ook wel: De Muizen), vastgelegd door de Hubble Space Telescope. Afbeelding van: NASA, H. Ford (JHU), G. Illingworth (UCSC/LO), M.Clampin (STScI), G. Hartig (STScI), de ACS Science Team en ESA xxix
- 4 Artist impression of the *Herschel* Space Observatory. Image credit: ESA/AOES Medialab xxxii
- 5 Composite image of the 250 μm , 350 μm and 500 μm *Herschel* SPIRE data in the COSMOS field. Created using data from the Herschel Multi-tiered Extragalactic Survey. xxxiv
- 6 A pair of merging galaxies NGC 4676, or The Mice, imaged with the Hubble Space Telescope. Image credit: NASA, H. Ford (JHU), G. Illingworth (UCSC/LO), M.Clampin (STScI), G. Hartig (STScI), the ACS Science Team and ESA xxxiv
- 1.1 Dust attenuation (A_{λ}/A_V) as a function of wavelength (λ) (Lo Faro et al. 2017, Fig. 1). Shown are attenuation curves from Calzetti et al. (2000, red), modified Calzetti et al. (2000) from Buat et al. (2011, salmon) and Charlot & Fall (2000, CF00; solid non-salmon). Different power law slopes (δ), as noted in the legend, are shown for Calzetti et al. (2000) as dot-dashed black lines and for Charlot & Fall (2000) as solid coloured lines. All curves have been normalised to unity at $\lambda \approx 0.55\mu\text{m}$ 4
- 1.2 Co-moving star formation rate density against redshift for rest frame far UV (green, blue and purple), and rest frame far IR (red) star formation rate densities. The UV star formation rate densities are uncorrected for dust attenuation. See Madau & Dickinson (2014) Table 1 for a full list of the data used. (Madau & Dickinson 2014, Fig. 8) 6

- 1.3 One of the first plots of the star formation rate (SFR) vs stellar mass (M_*) plane presented in Brinchmann et al. (2004, Fig. 17). The tight correlation can be clearly seen. The red dashed line is the average star formation rate at a given stellar mass while solid blue line is the mode star formation rate. The dashed blue lines show the limits that contain 95% of the galaxies at a given stellar mass. 8
- 1.4 The ‘consensus’ main sequence of star forming galaxies from Speagle et al. (2014, Fig. 8). This is the main sequence derived when combining the results from a large number of pre 2014 main sequence studies. As can be seen, the normalisation increases with redshift while the evolution of the slope is less clear. 8
- 1.5 Examples of galaxy mergers from the Sloan Digital Sky Survey (SDSS; York et al. 2000). The images are gri band composites, as the blue, red and green components respectively, and created using the SDSS SkyServer JPEG service. Each image is 256×256 pixels, corresponding to an angular size of 101×101 arcsecs. 10
- 1.6 Star formation rate (SFR) of merging galaxies as a fraction of the SFR of non-merging galaxies as a function of time since the first passage (t) for merging galaxies in the FIRE simulation (Moreno et al. 2019, Fig. 6). The vertical dashed lines are (left to right) the time of first passage, second passage and coalescence. As can be seen, merging galaxies spend the majority of a merger period with only slightly enhanced star formation rates and only short periods as starbursts. 11
- 1.7 1×1 arcmin tiles from COSMOS imaged with different telescopes at different wavelengths: (top row, left to right) Subaru 0.8 μm , IRAC 3.6 μm , MIPS 24 μm , PACS 100 μm , (middle row, left to right) SPIRE 250 μm , SPIRE 350 μm , SPIRE 500 μm , SCUBA-2 450 μm , (bottom row, left to right) SCUBA-2 800 μm , AzTEC 1.1 mm, MAMBO 1.2 mm and VLA 20 cm. (Casey et al. 2014) 15
- 1.8 Plot from Hurley et al. (2017, Fig. 3) showing the 16th, 50th and 84th percentiles for flux density accuracy from DESPHOT (dashed lines) with the 50th percentile for XID+ (coloured line) and corresponding 16th to 84th percentile region (shaded region) as a function of source flux density. The results from SPIRE 250 μm (blue, left), 350 μm (green, centre) and 500 μm (red, right) are shown. It can be seen that XID+ performs better than DESPHOT and produces more realistic errors. 16
- 1.9 Hubble Space Telescope imaged galaxies that have been classified by a convolutional neural network in Huertas-Company et al. (2015, Fig. 13) and have not been used for training. These galaxies are classified as spheroid dominated (left), disc dominated (centre left), irregular (centre right) or a point source (right). 18

2.1	Scatter plot of the predicted CIGALE flux densities against the flux densities from the COSMOS2015 catalogue for the $250\mu\text{m}$ (blue points), $350\mu\text{m}$ (green points), and $500\mu\text{m}$ (red points). The errors for the catalogue flux are instrumental noise plus confusion noise. The one-to-one line (magenta) and $y = 0.4x$ (cyan) illustrate that the data better fits the idea that bright sources in catalogues are overestimated.	32
2.2	Extracted $250\mu\text{m}$ flux densities using the informed Gaussian (F_{250}^{Info}) against the $250\mu\text{m}$ flux density priors from CIGALE (F_{250}^{Pri}) for the ALMA selected sources (blue). The fPDF of all objects extracted from the same tiles as the ALMA sources are also shown and presented from high (light red) to low (dark red) density. The red line is the one-to-one line.	33
2.3	Distribution of the residuals between the measured ALMA flux densities and the flux densities inferred from the SEDs at the ALMA wavelengths expressed as a multiple of the error on the ALMA flux density (σ) for the flat prior (red) and the informed Gaussian prior (blue). The informed prior reduces the mean from 3.44σ to 1.83σ , the standard deviation from 12.21σ to 7.95σ , and the skew from 11.50 to 7.97	34
2.4	Example plots showing the best fitting SEDs (curves) for the extracted SPIRE flux densities (red and blue points) using the flat prior (red) and informed Gaussian prior (blue). The other multi-wavelength data from the COSMOS2015 catalogue (purple) are overplotted. Panel (a) is an example of Group A, illustrating how the informed prior can increase the agreement between the best fitting SED and the measured ALMA flux density (green). Panel (b) is for Group B, illustrating how the informed and flat priors can give equal agreement between the best fitting SED and the measured ALMA flux density. Panel (c) is for Group C, showing how the flat prior can occasionally give better agreement.	36
2.5	Plot of the ratio of the $250\mu\text{m}$ flux densities extracted using the informed Gaussian prior (F_{250}^{Info}) to those extracted with the flat prior (F_{250}^{Flat}) against (F_{250}^{Info}). The ALMA sources are in dark blue, while the statistical average number density is from light red (high) to dark red (low).	37
2.6	Scatter plot of the infrared (IR, $3\mu\text{m} - 1100\mu\text{m}$) luminosity vs. redshift for the results from XID+ with the flat prior (red points/bars) and the informed Gaussian prior (blue points/bars) with a histogram of the IR luminosities. The green line corresponds to the $250\mu\text{m}$ 5σ confusion limit.	37
2.7	The noise map for the SPIRE $250\mu\text{m}$ band in the (a) COSMOS field (2 deg^2) and (b) XMM-LSS field (11 deg^2) from the HerMES DR4 (Oliver et al. 2012) and downloaded from HeDAM. For the COSMOS field, the outline of the study area is shown in red, while the ALMA sources are shown in blue.	43
3.1	Image of the SPIRE $250\mu\text{m}$ COSMOS coverage (Oliver et al. 2012, 7.84 deg^2 , red) with the overlayed MIPS $24\mu\text{m}$ COSMOS coverage (Sanders et al. 2007, 2.23 deg^2 , blue). The data used in this work were cut to match the MIPS $24\mu\text{m}$ coverage.	51
3.2	Brief summary of the science pipeline.	53

3.3	Scatter of the 250 μm residual map using different depth cuts on the prior list. The blue and red points are with and without 3σ clipping, respectively, while the green line is the 1σ instrument noise for the COSMOS field.	54
3.4	Masses of all objects detected in the Ks band (blue) are shown against redshift along with the faintest 20% in each redshift bin (red) for the (a) deep and (b) ultra-deep regions. The 90% completeness limit is shown by the thick black lines, while the dashed black lines show the edges of each redshift bin.	56
3.5	Example of the data generated at one step of the MCMC routine for the lowest ($0.2 \leq z < 0.5$) redshift bin, shown as number density from high (dark red) to low (dark blue). The MS being tested at this step, in this case the most likely step, is shown as the red line, while the contours for number density of the observed data are shown in orange. The size of the average observed error on SFR and $M_\star$ is also shown as a blue cross. . . .	57
3.6	Corner plot (Foreman-Mackey 2016) of the marginalised posterior of the forward modelling routine applied to a mock data set with known slope (0.6), normalisation ($0.7 \log(M_\odot/\text{yr}^{-1})$), and scatter (0.3 dex), shown as the blue lines. Panels on the diagonal show the 1D marginalised posteriors for the slope, normalisation and scatter (left to right). Off-diagonal panels show the combined 2D posteriors as labelled by their axes. The recovered 16th, 50th, and 84th percentiles are shown by the dashed vertical lines; all the input parameters are recovered, within error.	58
3.7	Comparison of fitting with Eq. 3.7 (orange dashed line) and Eq. 3.8 (red line) in the $0.5 \leq z < 0.8$ redshift bin. The galaxies are shown as a number density plot, with dark red being high number density and dark blue low number density, and the size of the average error on SFR and $M_\star$ is shown as a blue cross. There is very little difference in shape between the two fits, demonstrating that Eq. 3.8 is the preferred form of the MS.	60
3.8	Fitting of Eq. 3.8 to the objects in the redshift bins as labelled. The solid line is the most likely MS across the fitted $M_\star$ range, while the dotted line is an extrapolation across the $M_\star$ range of the plot. The galaxies are shown as a number density plot, with dark red being high number density and dark blue low number density. The vertical density discontinuities are the result of the two depths of data used: the deep mass limit is the yellow dashed line and the ultra-deep mass limit is the dot-dashed yellow line. Each panel also indicates the χ^2 of the most likely MS ($\chi^2_{\min}$) and the number of degrees of freedom in the fitting (n_{dof}), and shows the size of the average SFR and $M_\star$ errors as a blue cross.	64
3.9	Fitted MS trends using Eq. 3.8. The solid and dashed lines are the MS across the fitted $M_\star$ range. The dotted lines are extrapolations across the $M_\star$ range of the plot. The normalisation of the MS clearly increases as redshift increases. The slight decrease in slope at low redshift can be seen along with the increase above $z = 1.1$. The very steep slopes at $1.8 \leq z < 2.9$ can also be seen.	65

- 3.10 Comparison of the UVJ selected MS results of this work (magenta) with the observational MS from Speagle et al. (2014, FUV data in blue; IR data in green; and radio, hydrogen lines, and UV SED fitting in yellow), Schreiber et al. (2015) low mass MS (red), and the MS from the Illustris Simulation (Vogelsberger et al. 2014; Sparre et al. 2015, black diamonds). The α and β parameters from Eq. 3.8 are in panels (a) and (b), respectively. As Schreiber et al. (2015) hold their low-mass slope constant at unity, this is indicated in panel (a) as a flat red line. The redshifts shown for this work are the mean redshift in each redshift bin, while the horizontal error bars show the width of the redshift bin. A version of this plot using SFRs derived from IR template fitting can be found in Appendix 3.7.3. 66
- 3.11 Intrinsic scatter about the MS found from the MCMC fitting of the data in each redshift bin to Eq. 3.8. This work (magenta) is shown along with the observed scatters from Speagle et al. (2014, FUV data in blue, IR data in green, and other data in yellow) and Schreiber et al. (2015, red) as well as the median scatter found at each redshift in the Illustris Simulation (Genel et al. 2014; Vogelsberger et al. 2014; Sparre et al. 2015, black diamonds). The redshifts shown for this work are the mean redshift in each redshift bin while the horizontal error bars show the width of the redshift bin. Also shown is the best fit to this work's intrinsic scatter assuming no redshift evolution (dashed magenta line). The intrinsic scatter found in this work is consistent with existing literature, although above $z \approx 1.8$ our intrinsic scatter is smaller than is found in works that use IR SFR traces. 67
- 3.12 Distributions for the difference between the observed ALMA 870 μm flux densities and the predicted 870 μm flux densities from CIGALE using an error expansion factor of 2 (a), 3 (b), and 4 (c) in XID+. The Gaussian distribution for each expansion factor is also shown in orange, along with the means (μ) and standard deviations (σ) of the distributions. All the distributions have an approximately consistent σ so the best expansion factor was deemed to be that with μ closest to zero: 2 times the error from CIGALE. 79
- 3.13 Comparisons of Schreiber et al. (2015, red crosses) SFR- $M_\star$ positions with redshift to different methods of deriving SFR in this work. The mean CIGALE SFR in each mass bin is in magenta, the fitting of the de-blended SPIRE data with Chary & Elbaz (2001, CE01) templates to determine the infrared luminosity (L_{IR}) derived SFR is indicated with blue triangles, the IAS Stacking Library derived SFR with yellow stars, the full Schreiber et al. (2015) MS trends with red dashed lines, and our own fits to the Schreiber et al. (2015) data with orange dot-dashed lines. Horizontal error bars are omitted for ease of examination, but are all 0.25 dex. The Schreiber et al. (2015) data is complete to lower masses due to their use of the deeper GOODS data. As can be seen, the SFRs from SED fitting are systematically lower than those from stacking or template fitting. However, our stacked and CE01 data points are consistent with Schreiber et al. (2015) within error. 80

- 3.14 Comparison of the UVJ selected MS results of this work using the Chary & Elbaz (2001, CE01) template derived SFR (dark purple triangles) with the Schreiber et al. (2015) low-mass MS (red). The α and β parameters from Eq. 3.8 are in panels (a) and (b), respectively, and the trends from Fig. 3.10 are in magenta. As Schreiber et al. (2015) hold their low-mass slope constant at unity, this is indicated in panel (a) as the flat red line. The orange crosses in panel (b) are the normalisations for the full mass range (including turn-over) from Schreiber et al. (2015). The redshifts shown for this work are the mean redshift in each redshift bin while the horizontal error bars show the width of the redshift bin. The good agreement between the purple triangles and orange crosses and poor agreement between the purple triangles and red circles demonstrates how forcing a low-mass slope of unity results in normalisations that are too high at lower redshifts. 81
- 4.1 Examples of the central 64×64 pixels of SDSS gri, as blue, green and red respectively, galaxy images, corresponding to an angular size of 25.3×25.3 arcsec. The top row shows merging galaxies from the Darg et al. (2010a,b) catalogue while the bottom row shows non-merging galaxies. 87
- 4.2 Examples of the raw (first and third rows, linear colour scaling) and processed (second and fourth rows, Lupton et al. (2004) non-linear colour scaling) EAGLE images for merging (first and second rows) and non-merging (third and fourth rows) systems. The raw images shown are 128×128 pixels and imaged at 10 Mpc, corresponding to a *physical* size of 30×30 kpc or an *angular* size of 621×621 arcsec, while the processed images are 64×64 pixel images corresponding to an *angular* size of 25.3×25.3 arcsec. The redshifts are those that the EAGLE images have been projected to. 89
- 4.3 ROC curve for the observation network used on visually classified SDSS images at validation (blue) and testing (yellow). The area under each curve is 0.966. The dashed red line shows the position of a truly random network. 95
- 4.4 Examples of FP galaxies from the observation network for (a) a chance projection, (b) a galaxy filling the image, (c) a galaxy filling the image with a chance projection and (d) an isolated, non-interacting galaxy. Panels (e) to (h) show TN galaxies that are visually similar to those shown in (a) to (d). 98
- 4.5 Examples of FN galaxies from the observation network for (a) a galaxy with its merging companion outside the image, (b) a galaxy with its merging companion on the edge of the image, (c) a merging system and (d) a merging system where the minor galaxy is almost point-like. Panels (e) to (h) show TP galaxies that are visually similar to those shown in (a) to (d). 98

4.6	ROC curve for the simulation networks at validation (purple, light blue, yellow) and testing (dark blue, green, orange) for 100 Myr (solid), 200 Myr (dashed) and 300 Myr (dot-dashed) from the merger event. The areas under the curves can be found in Table 4.5. The dashed red line shows the position of a truly random network.	99
4.7	Distributions for the correctly (blue) and incorrectly (orange) identified EAGLE objects in the simulation network for (a) mergers and (b) non-mergers as a function of simulation snapshot redshift. Merging objects with low snapshot redshifts are disproportionately assigned a non-merger classification while non-merging objects with high simulation redshifts are often seen as mergers.	103
4.8	Distribution of EAGLE galaxies from the simulation network of the correctly (green) and incorrectly (brown) mergers (a) and non-mergers (b) as a function of EAGLE sSFR. Merging galaxies with low sSFR are often misclassified as non-merging while high sSFR non-mergers are often identified as mergers.	104
4.9	Distributions for the correctly (purple) and incorrectly (yellow) identified objects for (a) mergers and (b) non-mergers as a function of g-band magnitude for EAGLE galaxies classified by the simulation network. Faint mergers are preferentially classified as non-mergers while the distribution of misclassified non-mergers is at intermediate magnitudes.	105
4.10	Examples of FP EAGLE galaxies from the simulation network for (a) a chance projection, (b) a galaxy where the chance projection from the SDSS noise has resulted in the EAGLE galaxy appearing faint in the image, (c) a galaxy at low projection redshift and (d) an isolated, non-interacting galaxy. Panels (e) to (h) show TN galaxies that are visually similar to those shown in (a) to (d).	106
4.11	Examples of FN EAGLE galaxies from the simulation network for (a) a galaxy where the chance projection from the SDSS noise has resulted in the EAGLE galaxy appearing faint in the image, (b) a merging system that appears as a single, smooth galaxy, (c) a galaxy with a clearly identifiable counterpart and (d) as asymmetric galaxy. Panels (e) to (h) show TP galaxies that are visually similar to those shown in (a) to (d).	107
4.12	ROC curve for the SDSS images classified by the simulation network (blue) and the EAGLE images classified by the observation network (yellow). The area under the EAGLE through observation network is 0.502 while the area under the SDSS through simulation network is 0.689. The dashed red line shows the position of a truly random network.	109
4.13	Distributions for the correctly (blue) and incorrectly (orange) identified EAGLE objects for (a) mergers and (b) non-mergers as a function of redshift used for projection after being classified by the observation network. High redshift, merging systems are preferentially classified as non-merging while low redshift non-merging systems are preferentially classified as merging.	111

- 4.14 Distributions for the correctly (purple) and incorrectly (yellow) identified EAGLE objects for (a) mergers and (b) non-mergers as a function of g-band magnitude after being classified by the observation network. Faint, merging systems are preferentially classified as non-merging while bright non-merging systems are preferentially classified as merging. 113
- 4.15 Examples of EAGLE FP galaxies (a to d) from the observation network for (a) an isolated, non-interacting galaxy, (b) a chance projection, (c) a galaxy at low projection redshift and (d) a galaxy where the chance projection from the SDSS noise has resulted in the EAGLE galaxy appearing faint in the image. Panels (e) to (h) show TN galaxies that are visually similar to those shown in (a) to (d). 114
- 4.16 Examples of EAGLE FN galaxies (a to d) from the observation network for (a) an apparent single object, (b) a galaxy with a counterpart, either a merger counterpart from EAGLE or a chance projection from the SDSS noise, (c) an unambiguous merger and (d) a galaxy where the chance projection from the SDSS noise has resulted in the EAGLE galaxy appearing faint in the image. Panels (e) to (h) show TP galaxies that are visually similar to those shown in (a) to (d). 114
- 4.17 Distributions for the correctly (green) and incorrectly (brown) identified SDSS objects for (a) mergers and (b) non-mergers as a function of sSFR after being classified by the simulation network. Low sSFR merging systems are preferentially classified as non-merging while high sSFR non-merging systems have a higher misclassification rate than low sSFR non-merger. 117
- 4.18 Distributions for the correctly (purple) and incorrectly (yellow) identified SDSS objects after being classified by the simulation network for (a) mergers and (b) non-mergers as a function of z-band magnitude. Bright mergers are preferentially classified as non-mergers while the distribution of misclassified non-mergers is skewed towards the faint end of the distribution. This trend becomes less pronounced as the bands become more blue, from z to u-band. 118
- 4.19 Examples of SDSS FP galaxies from the simulation network for (a) a galaxy with a close (in projection) companion, (b) a non-interacting, isolated galaxy, (c) a galaxy showing asymmetry or morphological disturbance and (d) a galaxy with a non-physical artefact within the image. Panels (e) to (h) show TN galaxies that are visually similar to those shown in (a) to (d). 119
- 4.20 Examples of SDSS FN galaxies from the simulation network for (a) a galaxy with a clear merging counterpart, (b) a clearly disturbed system, (c) a galaxy whose merger companion is outside of the 64×64 pixel image and (d) the larger 256×256 pixel image showing the merger companion outside of panel (c). Panels (e) to (h) show TP galaxies that are visually similar to those shown in (a) to (d). 120

- 4.21 Heat maps to demonstrate how the observation (top row) and simulation (bottom row) networks detect an example merging SDSS system. Panels (a) and (d) show the original image of the galaxy being classified, panels (b) and (e) show the regions that most effect the merger classification and panels (c) and (f) show the heat maps where regions with darker colours have a greater affect on the classification (lower merger class output). Panels (b) and (e) are created by stretching the heat map between zero and one and multiplying this with the original image. 121
- 4.22 Heat maps to demonstrate how the observation (top row) and simulation (bottom row) networks detect an example merging EAGLE system. Panels (a) and (d) show the original image of the galaxy being classified, panels (b) and (e) show the regions that most effect the merger classification and panels (c) and (f) show the heat maps where regions with darker colours have a greater affect on the classification (lower merger class output). Panels (b) and (e) are created by stretching the heat map between zero and one and multiplying this with the original image. 122
- 5.1 Rest frame $g - r$ colour vs. absolute r magnitude (M_r) for SDSS DR7. The colour cut is shown as a red line where galaxies below the line are considered to be star-forming. 136
- 5.2 Comparison of $M_\star$ from this work (y-axis) with $M_\star$ from GAMA (x-axis). The red line denotes the 1-to-1 relation. The two data sets are in reasonable agreement with the average stellar masses within 0.2 dex and remain the same with and without the inclusion of SPIRE data. The typical statistical error on $M_\star$ is 0.1 dex. 137
- 5.3 Comparison of SFR from this work (y-axis) with SFR from GAMA (x-axis). The red line denotes the 1-to-1 relation. The two data sets are within 0.2 dex on average and are consistent within the typical error of 0.26 dex. Both the GAMA SFRs and the SFRs from this work are derived from SED fitting. 137
- 5.4 Rest frame U-V colour vs. rest frame V-J colour for KiDS. The colour cut is shown as a red line where galaxies below and to the right of the line are considered to be star-forming. 138
- 5.5 Gini vs. M_{20} for KiDS-GAMA GZ galaxies binned by Gini and M_{20} . The average `merger_neither_frac` from GZ within each bin is shown from low (red) to high (blue). The green line is the Lotz et al. (2004) split between merging and non-merging galaxies while the yellow line is the Lotz et al. (2008) split. Regions with low `merger_neither_frac` are visually identified as merging galaxies. Panel (b) includes the visually confirmed mergers from Darg et al. (2010a,b) as purple stars. 140

- 5.6 Asymmetry (A) vs. Smoothness (S) for KiDS-GAMA GZ galaxies binned by A and S. The average `merger_neither_frac` from GZ within each bin is shown from low (red) to high (blue). Regions with low `merger_neither_frac` are visually identified as merging galaxies. The orange line denotes the Conselice (2003) split between merging and non-merging galaxies. Panel (b) includes the visually confirmed mergers from Darg et al. (2010a,b) as purple stars. 142
- 5.7 Rest frame $U - V$ colour vs. rest frame $V - J$ colour for CANDELS-z000. The colour cut is shown as a red line where galaxies below and to the right of the line are considered to be star-forming. 144
- 5.8 Schematic representation of the architecture of the CNN used with input (three or one colour, 64×64 pixel image) on the left and output (binary classification of merger or non-merger) on the right. The sizes of the kernels (red) and layers are shown with the input layer having a depth of three for the SDSS (gri) and CANDELS ($1.6 \mu\text{m}$, $1.25 \mu\text{m}$, and 814 nm) data and one for the KiDS (r-band) data. 147
- 5.9 SFR- $M_\star$ plane populated with (a) non-merging galaxies and (b) merging SDSS galaxies. The colour indicates the number density from low (light yellow) to high (dark purple). Overlaid in red is the MS that has been fitted to all star-forming galaxies. As can be seen, the distributions of the merging and non-merging galaxies are similar with respect to the plotted MS. 151
- 5.10 Distribution of MS subtracted SFR for star-forming SDSS non-merging galaxies (blue) and merging galaxies (red). As can be seen, the merging star-forming population has a slightly higher mean MS subtracted SFR. . 152
- 5.11 Distribution of MS subtracted SFR for star-forming KiDS non-merging galaxies (blue) and merging galaxies (red). As can be seen, the merging star-forming galaxies have a similar mean MS subtracted SFR to the non-merging galaxies. 153
- 5.12 Distribution of MS subtracted SFR for the $0.85 < z \leq 1.21$ redshift bin for CANDELS non-merging galaxies (blue) and merging galaxies (red). This is the only data that is fitted with a double Gaussian distribution due to the clear multi-modal population. As can be seen, the main and secondary populations have a slightly higher mean MS subtracted SFR than the non-merging galaxies. 154
- 5.13 Average MS subtracted SFR of star-forming galaxies for SDSS merging (purple circle) and non-merging (dark blue diamond); KiDS merging (light blue circle) and non-merging (green diamond); and CANDELS merging (orange circles) and non-merging (red diamonds) galaxies. As can be seen, the change in SFR between the merging and non-merging galaxies is typically small. 156

5.14	Total merger fraction as a function of redshift for SDSS (dark blue circle), KiDS (light blue circle), and CANDELS (red circles) by redshift bin. Also plotted are the mass limited merger fractions with $\log(M_\star/M_\odot) > 10.0$ from Conselice et al. (2003, green stars), Cotini et al. (2013, lilac diamonds), Lotz et al. (2011) magnitude limited merger fractions with $M_B > -19.2$ (orange crosses), and the Duncan et al. (2019) lower mass ($9.7 < \log(M_\star/M_\odot < 10.3$, L, purple left triangles) and higher mass ($\log(M_\star/M_\odot > 10.3$, H, brown right triangles) merger fractions. The SDSS data are slightly higher than would be expected and the KiDS and CANDELS merger fractions are approximately a factor of two higher than the other results.	158
5.15	Merger fraction of quiescent (diamonds) and star-forming (circles) as a function of redshift for SDSS (purple and blue), KiDS (yellow and green), and CANDELS (orange and red). There is no overall trend with redshift with SDSS having a lower merger fraction for the star-forming galaxies, KiDS having a higher merger fraction for star-forming galaxies, and CANDELS star-forming galaxies having a higher merger fraction at all redshifts.	159
5.16	Merger fraction for star-forming galaxies with SFRs above indicated distances above the MS for SDSS and KiDS data (top panel) and CANDELS data (bottom panel). To avoid low number statistics, only thresholds above which there are 50, or more, galaxies are shown. The SDSS (top panel, purple), KiDS (top panel, blue), and all CANDELS data show a trend of increasing merger fraction as the distance to the MS increases, although the CANDELS-z000 drops again above $0.62 \log(M_\odot \text{yr}^{-1})$	161
5.17	Fraction of star-forming, merging galaxies with SFRs above given distances above the MS (solid lines) and fraction of star-forming, non-merging galaxies with SFRs above given distances above the MS (dashed lines). The top panel contains the SDSS and KiDS data sets while the bottom panel contains the CANDELS data. To avoid low number statistics, only thresholds above which there are 50, or more, galaxies are shown. The SDSS (top panel, purple), KiDS (top panel, blue), and all CANDELS data show that there is a higher fraction of the total number of merging galaxies above nearly all distance above the MS.	162
5.18	Examples of non-merging galaxies (top row) and merging galaxies (bottom nine rows) for SDSS data set as defined by the CNN. Images are gri composite with a size of 64×64 pixel or 13.7×13.7 arcsec.	175
5.19	Examples of non-merging galaxies (top row) and merging galaxies (bottom nine rows) for KiDS data set as defined by the CNN. Images are greyscale r-band with a size of 64×64 pixel or 25.3×25.3 arcsec.	176
5.20	Examples of non-merging galaxies (top row) and merging galaxies (bottom nine rows) for CANDELS data set as defined by the CNN. Images are $1.6 \mu\text{m}$, $1.25 \mu\text{m}$, 814 nm composite with a size of 64×64 pixel or 3.8×3.8 arcsec.	177

List of Tables

2.1	Absolute residuals ($ \gamma $) between the measured ALMA flux densities and the fluxes of the best fitting SEDs at the ALMA wavelengths expressed as a multiple of the error on the ALMA flux density. The values in Cols. 2 and 3 are the number (percentage) of objects with a $ \gamma $ less than the value in Col. 1.	33
2.2	Absolute residuals ($ \gamma $) between the measured ALMA flux densities and the fluxes of the best fitting MBB templates and MAGPHYS SEDs at the ALMA wavelengths expressed as a multiple of the error on the ALMA flux density. The values in Cols. 3 and 4 are the number (percentage) of objects with a $ \gamma $ less than the value in Col. 2.	38
2.3	Parameters used for the various properties in the CIGALE model SEDs where they different from the default values. All ages and times are in Gyr.	45
3.1	Telescopes and associated bands that were used for CIGALE spectral energy distribution fitting from the COSMOS2015 catalogue.	51
3.2	Parameters from fitting Eq. 3.8 to the UVJ selected star forming galaxies, with α the slope and β the normalisation at $\log(M_{\star}/M_{\odot}) = 10.5$. The intrinsic scatter is found during the MCMC fitting.	62
3.3	Parameters used for the various properties in the CIGALE model SEDs where they differ from the default values. All ages and times are in Gyr.	76
4.1	Terms used when describing the performance of neural networks	91
4.2	Architecture of the CNN. The first column in the type of layer while the second column contains the associated properties. The input is a 64×64 pixel, three channel image and the output is two probabilities, one for the probability the input is a merger and one for the probability the input is a non-merger. Further details on what the properties of the layers mean can be found in Sect. 4.3.2.	93
4.3	Confusion matrix for SDSS images classified by the observation network.	96
4.4	KS-test statistic, $D_{N,M}$, and the critical value, $\text{Crit}_{N,M} = c(\alpha) \sqrt{\frac{n+m}{nm}}$, for the SDSS images classified by the observation network. If $D_{N,M} > \text{Crit}_{N,M}$, the null hypothesis that the two distributions are the same is rejected at level $\alpha = 0.05$. Here, $c(\alpha) = 1.224$ for $\alpha = 0.05$ and n and m are the sizes of samples N and M	97

4.5	Statistics for the SDSS and the 100 Myr, 200 Myr and 300 Myr EAGLE trained networks at testing.	100
4.6	Confusion matrix for EAGLE images classified by the simulation network.	100
4.7	KS-test statistic, $D_{N,M}$, and the critical value, $\text{Crit}_{N,M} = c(\alpha) \sqrt{\frac{n+m}{nm}}$, for the EAGLE images classified by the simulation network. If $D_{N,M} > \text{Crit}_{N,M}$, the null hypothesis that the two distributions are the same is rejected at level $\alpha = 0.05$. Here, $c(\alpha) = 1.224$ for $\alpha = 0.05$ and n and m are the sizes of samples N and M	102
4.8	Statistics for the network trained with partially processed EAGLE images at testing. C is convolving the EAGLE image with the SDSS PSF, N is injecting the EAGLE image into the real SDSS noise, R is matching the EAGLE resolution to that of SDSS and Z is changing the EAGLE magnitude to apparent from absolute.	108
4.9	Confusion matrix for EAGLE images classified by the observation network.	109
4.10	KS-test statistic, $D_{N,M}$, and the critical value, $\text{Crit}_{N,M} = c(\alpha) \sqrt{\frac{n+m}{nm}}$, for the EAGLE images classified by the observation network. If $D_{N,M} > \text{Crit}_{N,M}$, the null hypothesis that the two distributions are the same is rejected at level $\alpha = 0.05$. Here, $c(\alpha) = 1.224$ for $\alpha = 0.05$ and n and m are the sizes of samples N and M	110
4.11	Statistics for the EAGLE images classified by the observation network and the SDSS images classified by the simulation network.	115
4.12	Confusion matrix for SDSS images classified by the simulation network. .	116
4.13	KS-test statistic, $D_{N,M}$, and the critical value, $\text{Crit}_{N,M} = c(\alpha) \sqrt{\frac{n+m}{nm}}$, for the SDSS images classified by the simulation network. If $D_{N,M} > \text{Crit}_{N,M}$, the null hypothesis that the two distributions are the same is rejected at level $\alpha = 0.05$. Here, $c(\alpha) = 1.224$ for $\alpha = 0.05$ and n and m are the sizes of samples N and M	119
4.14	Statistics for the 100 Myr, 200 Myr and 300 Myr EAGLE trained networks at testing using EAGLE images created with a linear colour scaling. The fourth column presents the cross application of passing the linear colour scaled EAGLE images through the observation network while the fifth column presents the cross application of passing the SDSS images through the 100 Myr linearly scaled EAGLE trained network.	129
5.1	Summary of data used. The SDSS and KiDS limiting magnitudes are in r-band while the CANDELS limiting magnitude in H-band.	145
5.2	Statistics for trained CNNs. Definitions of terms can be found in Appendix 5.7.2	150
5.3	Best fit parameters for skewed Gaussian distributions fitted to star-forming SDSS data. μ and σ are in units of $\log(M_{\odot}\text{yr}^{-1})$	152
5.4	Best fit parameters for skewed Gaussian distributions fitted to star-forming KiDS data. μ and σ are in units of $\log(M_{\odot}\text{yr}^{-1})$	153

5.5	Best fit parameters for skewed or double Gaussian distributions fitted to star-forming CANDELS data. For the $0.85 < z \leq 1.21$ bin, where a double Gaussian is used, the star-forming component is the component with the lowest μ . μ , and σ are in units of $\log(M_{\odot}\text{yr}^{-1})$	155
5.6	Merger fraction by redshift and data set for quiescent, star-forming, and total galaxy populations. Errors are derived from correcting for the precision of the network.	157
5.7	Parameters used for various properties in CIGALE model SEDs where they differ from default values. All ages and times are in Gyr.	171
5.8	Terms used when describing neural network performance from Pearson et al. (2019)	174

Nederlandse samenvatting

Beknopte samenvatting

Kennis over de opbouw van sterren is cruciaal voor het begrijpen waarom ons universum er vandaag de dag zo uit ziet: het zijn de sterren die de nachtelijke hemel verlichten. Dit proefschrift onderzoekt enkele van de drijfveren van deze stervorming over de geschiedenis van het universum.

Het verband tussen de snelheid van stervorming en de massa van sterren in sterrenstelsels wordt bestudeerd, tezamen met het verband tussen stervorming en fusies van sterrenstelsels. Hiervoor zijn nieuwe technieken nodig om nauwkeurige schattingen van de stervormingssnelheid te maken en fusies van sterrenstelsels te detecteren. Een belangrijk onderdeel van de schattingen van de stervormingssnelheid, de emissie in het ver-infrarood, heeft te lijden onder een relatief lage resolutie waardoor individuele sterrenstelsels in elkaar op lijken te gaan en niet langer van elkaar te onderscheiden zijn.

In dit proefschrift zijn de bestaande technieken voor het onderscheiden van zulke sterrenstelsels verbeterd, waardoor een betere extractie van de helderheid in het ver-infrarood, en daarmee van de stervormingssnelheid, mogelijk is. Dit proefschrift maakt gebruik van de nieuwste technieken om fuserende sterrenstelsels te identificeren in zowel simulaties als echte waarnemingen.

Onderzoek naar de relatie tussen de stervormingssnelheid en de massa van sterren in sterrenstelsels heeft uitgewezen dat in het vroege heelal hoge en lage massa sterrenstelsels met dezelfde snelheid sterren vormden. Naarmate het heelal ouder werd, konden de hoge massa sterrenstelsels minder goed nieuwe sterren vormen. Daarnaast bleek dat fusies van sterrenstelsels niet van grote invloed zijn op stervormingssnelheid, maar wel tot uitbarstingen van stervorming kunnen leiden.

Samenvatting

Dit proefschrift richt zich op de evolutie van sterrenstelsels gedurende de geschiedenis van het heelal en op wat invloed heeft op de snelheid waarmee sterren worden gevormd. Ik ontwikkel een techniek om betrouwbaardere en nauwkeurigere schattingen van de stervormingssnelheid in sterrenstelsels te genereren. Verder gebruik ik ‘deep learning’, een soort kunstmatige intelligentie, om methoden te ontwikkelen om fusies van sterrenstelsels te detecteren: gebeurtenissen die tot enkele miljarden jaren kunnen duren en de mogelijkheid hebben om de snelheid van stervorming in sterrenstelsels sterk te beïnvloeden. Met deze nieuwe technieken en instrumenten onderzoek ik hoe de stervormingssnelheid van sterrenstelsels wordt beïnvloed door de massa in sterren en hoe deze zich ontwikkelt naarmate het heelal ouder wordt. Ik bestudeer ook hoe fusies van sterrenstelsels de snelheid waarmee sterrenstelsels sterren vormen veranderen.

Sterrenstelsels zijn grote structuren van sterren, gas en stof, die in alle delen van het elektromagnetische spectrum straling uitzenden. Astronomen gebruiken verschillende golflengten van licht om verschillende eigenschappen van sterrenstelsels te karakteriseren of juist soortgelijke eigenschappen op verschillende manieren. Om de stervormingssnelheid van sterrenstelsels nauwkeurig te bepalen, zijn gedetailleerde ver-infrarood gegevens nodig, met golflengten tussen ongeveer 50 en 500 μm . De grootte van het kleinste kenmerk dat we kunnen zien, ook wel resolutie genoemd, is evenredig met de golflengte van het licht dat we waarnemen, zodat we steeds minder details kunnen zien naarmate de golflengte toeneemt. Dit probleem wordt erger naarmate de sterrenstelsels die we bekijken verder weg liggen, en kan resulteren in sterrenstelsels die zich met elkaar vermengen in de waargenomen beelden terwijl ze fysiek gescheiden zijn. Licht van zulke niet gescheiden sterrenstelsels, zo ver weg dat ze niet eens als puntbronnen verschijnen, hoopt zich op in de vorm van achtergrondruis. Dit achtergrondruis, bekend als ‘confusion noise’ in de Engelstalige literatuur, is een bekend probleem voor het SPIRE-instrument op de *Herschel* ruimtetelescoop, getoond in Fig. 1, en kan niet worden opgelost door langer te observeren, in tegenstelling tot de meeste andere vormen van ruis.

Om dit probleem van slechte (in vergelijking met optische data) resolutie en niet gescheiden ruis te omzeilen, is de techniek XID+ ontwikkeld. XID+ maakt gebruik van Bayesiaanse statistieken om ver-infrarood beelden te ont mengen en te ontruisen, waardoor een nauwkeurige waarde van de ver-infrarood emissie van elk sterrenstelsel wordt verkregen. Als onderdeel van dit doctoraat is deze techniek verbeterd. Door gebruik te maken van informatie van andere golflengten is het mogelijk om een redelijke schatting te maken van de helderheid van een sterrenstelsel in het ver-infrarood. Deze schattingen kunnen dan gebruikt worden om XID+ te sturen en de waarden van de ver-infrarood straling te verbeteren. Hoofdstuk 2 gebruikt gegevens van de Atacama Large Millimeter/submillimeter Array (ALMA) om aan te tonen dat deze waarden beter zijn dan de waarden die XID+ zonder deze schattingen genereert. Kenmerken van ALMA zijn dat het zeer gevoelig is, een zeer goede resolutie heeft en weinig confusion noise. Het is echter niet ontworpen om grote oppervlaktes aan de hemel te onderzoeken die niet gericht zijn op specifieke sterrenstelsels, iets wat wel met *Herschel* kan worden gedaan, en het levert meestal alleen beelden op van een handvol sterrenstelsels per keer. Dit maakt ALMA een slechte keuze voor grote, niet doelgerichte, statistische studies. Met behulp van een klein aantal objecten die zowel door *Herschel* als door ALMA zijn afgebeeld, laat hoofdstuk 2

Figuur 1: Impressie van de *Herschel* Space Observatory. Afbeelding van: ESA/AOES Medialab

zien dat de verbeterde *XID+* beter presteert dan de oorspronkelijke *XID+*.

Hoofdstuk 3 is een vervolg op deze werkzaamheden en past deze nieuwe techniek toe op een grote dataset. De verbeterde *XID+* werd toegepast op het 2 vierkante graad grote COSMOS hemeloppervlak, weergegeven in Fig. 2. Op basis van deze afbeelding zijn van meer dan 200 000 sterrenstelsels de ver-infrarood helderheden bepaald, die vervolgens zijn gebruikt om zeer betrouwbare schattingen van hun stervormingssnelheid te genereren. Deze sterrenstelsels beslaan 90% van de geschiedenis van het heelal, waarbij de oudste sterrenstelsels zijn afgebeeld toen het heelal nog geen 1 miljard jaar oud was. Schattingen van de stellaire massa van deze sterrenstelsels, dat wil zeggen de totale massa van de sterren, zijn ook afgeleid. Met de stervormingssnelheden en stellaire massa's van deze sterrenstelsels is het mogelijk om het verband tussen stellaire massa en stervorming, bekend als de hoofdreeks van stervormende sterrenstelsels, over een veel langere periode te bestuderen dan eerdere werken.

Op basis van deze studie is vastgesteld dat naarmate de stellaire massa van een sterrenstelsel toeneemt, ook de stervormingssnelheid toeneemt. De normalisatie van deze relatie bleek lineair toe te nemen naarmate we verder in de geschiedenis van het heelal kijken. Dit geeft aan dat sterrenstelsels in het vroege universum sneller sterren vormden dan nu het geval is, mogelijk omdat gas opraakt naarmate het universum ouder wordt. Dit resultaat komt overeen met andere studies, zowel aan de hand van waarnemingen als simulaties. Een soortgelijke verandering werd gevonden in de helling van de relatie, waarbij de helling snel afneemt in de eerste drie miljard jaar van het heelal en daarna ongeveer constant blijft. De helling van de relatie verbindt de hoge massa met de lage massa sterrenstelsels. Een helling van ongeveer één, gevonden in het vroege universum, geeft aan dat in het jonge universum de hoge massa en de lage massa sterrenstelsels met dezelfde snelheid sterren vormden, in verhouding tot hun massa. Naarmate de helling met de tijd afneemt, worden de hoge massa sterrenstelsels minder efficiënt in het vormen van sterren ten opzichte van de lage massa sterrenstelsels.

Figuur 2: Samengesteld beeld van de 250 μm , 350 μm en 500 μm *Herschel* SPIRE-gegevens in het COSMOS-veld. Gemaakt met behulp van gegevens van de *Herschel* Multi-tiered Extragalactic Survey.

Figuur 3: Een paar fuserende sterrenstelsels (NGC 4676, of ook wel: De Muizen), vastgelegd door de Hubble Space Telescope. Afbeelding van: NASA, H. Ford (JHU), G. Illingworth (UCSC/LO), M.Clampin (STScI), G. Hartig (STScI), de ACS Science Team en ESA

Dit proefschrift gaat vervolgens verder met het bestuderen van hoe fuserende sterrenstelsels, zoals NGC 4674 in Fig. 3, invloed hebben op de stervormingssnelheden, waardoor een groot aantal fuserende sterrenstelsels geïdentificeerd moet worden. Hiervoor werd deep learning gebruikt. Echter, vanwege het relatief kleine aantal vooraf geïdentificeerde fuserende sterrenstelsels beschikbaar voor het trainen van een deep learning netwerk, werd onderzoek gedaan aan de hand van afbeeldingen van gesimuleerde sterrenstelsels. Om dit te doen, werd in hoofdstuk 4 twee keer een deep learning architectuur ontwikkeld en getraind, één keer met echte waarnemingen van Sloan Digital Sky Survey (SDSS) en één keer met gesimuleerde afbeeldingen van de Evolution and Assembly of GaLaxies and their Environments (EAGLE)-simulatie, verwerkt om eruit te zien als SDSS-beelden. Deze netwerken werden vervolgens toegepast, ook kruislings: de waarnemingen werden dus ook aan het EAGLE getrainde netwerk doorgegeven en de simulaties ook aan het SDSS getrainde netwerk.

Het SDSS getrainde netwerk dat is toegepast op SDSS-beelden presteerde goed, waarbij bijna alle sterrenstelsels correct zijn geïdentificeerd als fuserend of niet-fuserend. Er waren geen fysieke parameters gevonden die tot een verkeerde classificatie leidden. Het EAGLE-getrainde netwerk dat werd toegepast op EAGLE-beelden presteerde niet zo goed, waardoor slechts twee derde van de sterrenstelsels correct geïdentificeerd werden. De verkeerd geïdentificeerde EAGLE-sterrenstelsels waren voornamelijk lichtzwak of hadden een lage roodverschuiving in de simulatie, wat overeenkomt met het universum zoals het nu is. Het netwerk zelf bleek te vertrouwen op diffuse kenmerken voor het maken van een correcte classificatie.

De kruislingse toepassing van de getrainde deep learning netwerken was minder succesvol. Het doorgeven van de gesimuleerde beelden aan het SDSS getrainde netwerk resulteerde in de correcte identificatie van slechts de helft van de sterrenstelsels. Het algoritme blijkt voorkeur te geven aan classificaties die duiden op niet-fuserende sterren-

stelsels. Het doorgeven van SDSS-beelden aan het EAGLE getrainde netwerk was echter succesvoller, met een nauwkeurigheid van 65%. Dit suggereert dat de latere fases van de fusies, het merendeel van de SDSS fuserende sterrenstelsels, ontbreken in de simulaties, terwijl de SDSS-beelden niet de minder voor de hand liggende fusies bevatten die zich wel onder de EAGLE-beelden bevinden. De verkeerd geclassificeerde sterrenstelsels bleken meestal een hoge stervormingssnelheid te hebben in het geval van niet-fuserende sterrenstelsels die geclassificeerd zijn als wel fuserende sterrenstelsels, of juist lage stervormingssnelheden in het geval van fuserende sterrenstelsels die geclassificeerd zijn als niet-fuserende sterrenstelsels. Dit suggereert dat fuserende sterrenstelsels in simulaties een toegenomen stervormingssnelheid hebben, iets wat misschien niet in echte sterrenstelsels te zien is.

In het verlengde hiervan is in hoofdstuk 5 een studie uitgevoerd naar het effect van fuserende sterrenstelsels op de stervormingssnelheid. Dit hoofdstuk combineert het werk van de vorige drie. De hoofdreeks van stervormende sterrenstelsels werd geïdentificeerd in drie datasets, SDSS, de Kilo Degree Survey (KiDS) en de Cosmic Assembly Near-infrared Assembly Deep Extragalactic Legacy Survey (CANDELS), waarbij de stervormingssnelheid in KiDS met behulp van ver-infraroodgegevens is afgeleid met behulp van de in hoofdstuk 2 ontwikkelde technieken. Met behulp van technieken die vergelijkbaar zijn met die beschreven in hoofdstuk 4, werden fusies van sterrenstelsels geïdentificeerd in deze drie datasets en werden ze in kaart gebracht in een diagram dat de stervormingssnelheid vergelijkt met de stellaire massa's van de sterrenstelsels. Vervolgens werd de verandering in stervormingssnelheid voor de fuserende sterrenstelsels, vergeleken met de niet-fuserende sterrenstelsels, onderzocht, waarbij de hoofdreeks van stervormende sterrenstelsels als nulpunt werd gebruikt.

Het effect van fuserende sterrenstelsels op de stervormingssnelheid is klein. Er is zelden meer dan een factor twee verandering in stervormingssnelheid en meestal is de factor 1,2. De SDSS- en KiDS-gegevens toonden een lichte toename aan, terwijl de CANDELS gegevens een afname in het oudere universum en een toename in het jongere universum lieten zien. In plaats van direct de fractie van sterrenstelsels met hoge stervormingssnelheden die ook fuserende sterrenstelsels zijn te onderzoeken, onderzoeken we deze fractie als een functie van de afstand boven de hoofdreeks van het sterrenstelsel. Voor SDSS, KiDS en CANDELS neemt de fractie van de fusies toe naarmate de afstand boven de hoofdreeks van stervormende sterrenstelsels toeneemt. Dit bewijst dat fusies perioden van versterkte stervorming kunnen veroorzaken, hoewel deze kortstondige uitbarstingen van stervorming in de minderheid van de fuserende sterrenstelsels worden gevonden.

“Fysieke drijfveren van de kosmische stervormingsgeschiedenis” is een enorm vakgebied dat men niet in één scriptie, of zelfs maar één leven lang, volledig kan behandelen. Hopelijk heeft dit proefschrift ten minste één steentje bijgedragen en heeft het, na het werk van mijn voorgangers, een stap in de goede richting gezet, zodat anderen het kunnen gebruiken als basis voor de route die zij af zullen leggen.

English summary

Abstract

The build up of stars over the history of the universe is important for understanding why our universe looks the way it does today: it is the stars that create the energy that lights up the night sky. This thesis looks into some of the drivers of this star formation over the history of the universe.

The connection between star formation rate and the mass of stars in galaxies is studied along with the connection between star formation and galaxy mergers. To do this, new techniques and tools are required for generating accurate star formation rate estimates and detecting galaxy mergers. A key component of star formation rate estimates, emission in the far-infrared, suffers from relatively low resolution which causes galaxies to blend with one another.

In this thesis, existing de-blending tools have been improved, allowing better extraction of far-infrared luminosities and hence better estimates of star formation rates. This thesis also employs the latest deep learning techniques to identify merging galaxies in both simulations and observations of our universe.

Studying the relation between the star formation rate and the existing mass of stars in galaxies found that in the early universe, high mass and low mass galaxies formed stars at similar rates. As the universe aged, high mass galaxies become less able to form new stars. For the influence of galaxy mergers on star formation rates, this thesis found that on average, galaxy mergers do not notably influence star formation rates but can trigger starbursts.

Figure 4: Artist impression of the *Herschel* Space Observatory. Image credit: ESA/AOES Medialab

Summary

This thesis focuses on galaxy scale star formation over the history of the Universe and what influences the rate at which stars are formed in their host galaxies. I develop a technique to generate more reliable and accurate estimates of the star formation rate in galaxies. Furthermore, I use deep learning to develop methods to detect galaxy mergers: events that can last up to several billion years and have the potential to greatly influence the star formation rates in galaxies. With these new techniques and tools, I examine how the star formation rates of galaxies are influenced by the stellar masses and how this evolves as the universe ages. I also study how galaxy mergers change the rate at which galaxies form stars.

Galaxies are huge structures of stars, gas and dust, which emit in all parts of the electromagnetic spectrum. Astronomers use different wavelengths of light to characterise different properties of galaxies, or similar properties in different ways. To accurately determine the star formation rates of galaxies, detailed far-infrared data are needed, with wavelengths between approximately 50 and 500 μm . The size of the smallest feature we can see is proportional to the wavelength of the light we are observing, so we can see less and less detail as the wavelength increases. This issue becomes worse as the galaxies we look at become further away, with galaxies becoming blended with each other in the images when they are physically distinct. Light from unresolved galaxies, so distant they do not even appear as point like sources, accumulates to form background noise. This unresolved background noise, known as confusion noise, is a well known issue for the SPIRE instrument on the *Herschel* Space Observatory, shown in Fig. 4, and cannot be suppressed by taking a longer exposure, unlike most other forms of noise.

To work around this issue of poor resolution, when compared to optical data, and confusion noise, the tool XID+ has been developed. XID+ uses Bayesian statistics to de-blend and de-confuse far-infrared images, providing an accurate value of the far-infrared

emission from each galaxy. As part of this PhD, this tool has been enhanced. By using information from other wavelengths, it is possible to generate a reasonable estimate of how bright a galaxy will be in the far-infrared. These estimates can then be used to guide XID+ and improve the values of the far-infrared emission. Chapter 2 uses data from the Atacama Large Millimeter/submillimeter Array (ALMA) to prove that these values are better than the values generated by XID+ without these estimates. The setup of ALMA means that it is very sensitive with a very high resolution and low confusion noise. However, it is not designed to perform large area surveys that do not target specific galaxies, which *Herschel* is able to conduct, and typically only images a hand full of galaxies at a time. This results in it being a poor choice for large, un-targeted statistical studies. Using a small number of objects that have been imaged by *Herschel* and ALMA, Chapter 2 shows that the enhanced XID+ performs better than the original XID+.

Chapter 3 leads on from this work and applies this new technique to a large data set. The enhanced XID+ was applied to the 2 square degree COSMOS field, shown in Fig. 5. From this image, over 200 000 galaxies had their far-infrared emission determined, which were then used to generate very reliable estimates of their star formation rates. These galaxies cover 90% of the history of the universe, with the oldest being imaged when the universe was less than 1 billion years old. Estimates of the stellar mass of these galaxies, that is the total mass of stars, were also derived. With the star formation rates and stellar masses of these galaxies, it is possible to study the relation between stellar mass and star formation, known as the main sequence of star forming galaxies, over a far longer time range than previous works.

From this study, it was determined that as the stellar mass of a galaxy increases, so does its star formation rate. The normalisation of this relation was found to increase linearly as we look further into the history of the universe. This indicates that galaxies in the early universe were forming stars faster than they are today, possibly as gas is depleted as the universe ages. This result is consistent with other studies, both using observations and simulations. A similar change was found with the slope of the relation, with the slope decreasing rapidly in the first three billion years of the universe before becoming approximately constant. The slope of the relation links the high mass and low mass galaxies. A slope of approximately one, found in the early universe, indicates that in the young universe high mass and low mass galaxies formed stars at similar rates, with respect to their mass. As the slope becomes shallower with time, the high mass galaxies become less efficient at forming stars with respect to the low mass galaxies.

This thesis then moves on to study how galaxy mergers, such as NGC 4674 shown in Fig. 6, influence star formation rates, requiring a large number of merging galaxies to be identified. For this, deep learning (a kind of artificial intelligence) was employed. However, due to the relatively small number of pre-identified merging galaxies available for training a deep learning network, using images of simulated galaxies was explored. To do this, a deep learning architecture was developed and trained twice in Chapter 4, once with real observations from Sloan Digital Sky Survey (SDSS) and once with simulated galaxy images from the Evolution and Assembly of GaLaxies and their Environments (EAGLE) simulation, processed to look like SDSS images. These networks were then applied to the data set they were not trained on (cross-applied): the observations were passed through the EAGLE trained network and the simulations were passed through the SDSS trained network.

Figure 5: Composite image of the 250 μm , 350 μm and 500 μm *Herschel* SPIRE data in the COSMOS field. Created using data from the *Herschel* Multi-tiered Extragalactic Survey.

Figure 6: A pair of merging galaxies NGC 4676, or The Mice, imaged with the Hubble Space Telescope. Image credit: NASA, H. Ford (JHU), G. Illingworth (UCSC/LO), M. Clampin (STScI), G. Hartig (STScI), the ACS Science Team and ESA

The SDSS trained network applied to SDSS images performed well, with nearly all of the galaxies being correctly identified as merging or not merging and no physical parameters were obviously causing mis-classification. The EAGLE trained network applied to EAGLE images did not perform as well, resulting in only two thirds of the galaxies correctly classified. The misclassified EAGLE galaxies were primarily faint or at low redshift in the simulation, equivalent to the universe as it is today. The network itself was also found to rely on diffuse features of the galaxies to correctly identify mergers.

Cross application of the trained deep learning networks was less successful. Passing the simulated images through the SDSS trained network resulted in only half of the galaxies being correctly identified, with preferential assignment to the non-merger classification. Passing SDSS images through the EAGLE trained network, however, was more successful, achieving an accuracy of 65%. This suggests that simulations are lacking in the later stage mergers that make up the majority of the SDSS mergers while the SDSS images do not have the less obvious mergers that the EAGLE images contain. The misclassified galaxies were typically found to have high star formation rates, for non-mergers classified as mergers, or low star formation rates, for mergers classified as non-mergers. This suggests that merging galaxies in simulations have an increase in star formation rate that may not be seen in real galaxies.

Leading from this, a study into the effect galaxy mergers have on star formation rates was undertaken in Chapter 5. This chapter combines the work done in the previous three. The galaxy main sequence was identified within three data sets, SDSS, the Kilo Degree Survey (KiDS) and the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS), with the star formation rates in KiDS using far-infrared data de-blended using the techniques developed in Chapter 2. Using deep learning techniques similar to those in Chapter 4, galaxy mergers were then identified in these three surveys and were plotted on the star formation rate versus stellar mass diagram. The change in star formation rate for the merging galaxies compared to the non-merging galaxies was then examined, using the galaxy main sequence of star forming galaxies as a zero point.

The effect of merging galaxies on the star formation rate is small, rarely over a factor of two change in star formation rate and typically a factor of 1.2. The SDSS and KiDS data showed this change to be a slight increase while the CANDELS data showed this change to be a decrease in the older universe and an increase in the younger universe. Instead of directly examining the fraction of starburst galaxies that are mergers, we examine the merger fraction as a function of distance above the galaxy main sequence. For SDSS, KiDS and CANDELS the fraction of mergers increases as the distance above the galaxy main sequence increases. This is evidence that mergers can cause periods of enhanced star formation although these short-lived starbursts are found in the minority of merging galaxies.

“Physical Drivers of the Cosmic Star Formation History” is an enormous subject area that one cannot hope to completely cover within one thesis, or indeed one lifetime. However, this thesis has hopefully laid at least one stone along the path, to carry on from those who came before and provide a foundation for those who come after.

Approximately 14 Gyr ago, the Universe started with the Big Bang. Within a few hundred million years, the first galaxies had formed and were filled with the first generation of stars (e.g. Rees 1993; Barkana & Loeb 2001; Bromm & Yoshida 2011). These early galaxies formed the seeds for the galaxies in the local Universe, assuming that galaxies grow hierarchically with smaller galaxies merging to form larger galaxies (e.g. Bromm 2009; Bromm & Yoshida 2011). Throughout this time, the older stars within the galaxies died and new stars were formed from the remnants of the older stars and the as yet unused gas left from the Universe's birth.

The emission that we observe from galaxies originates from stars or the accretion of material onto active galactic nuclei (AGN). The shape of the emission, spatially and spectrally, is also the basis for the majority of classification systems for galaxies. This results in the stellar population being crucial to our understanding of galactic structure; knowing the history of the stellar population will allow an understanding of the formation and evolution of galaxies. Thus, the rate at which gas is transformed into stars, the star formation rate, is central to our theoretical understanding of the formation and evolution of galaxies. As a result, star formation is one of the most important areas of research into galaxy formation and evolution.

This thesis is an exploration into the star formation rates of galaxies and what influences these rates. This will be done using state of the art observations and numerical simulations coupled with cutting edge data analysis techniques. A technique will be developed to overcome some of the limitations of far infrared (IR) observations and used to improve star formation rate estimates. From this, the mass dependence of star formation rates will be explored. Machine learning techniques will be employed to identify galaxy mergers and generate merger catalogues using data from a number of surveys. Using these catalogues, I will explore the influence of galaxy mergers on star formation rates. These machine learning techniques will also be used to study the differences that are present between observations and simulations of galaxy mergers. This will be done with a view to identifying where simulated mergers do not appear similar to observed mergers as well as identifying the type of mergers current observational identification techniques can miss.

This chapter provides the background for the work undertaken within this thesis. As this thesis primarily focuses on star formation, this introduction will briefly discuss how stars form before moving onto how star formation rates of galaxies can be estimated and a brief discussion on how the physical and environmental properties of galaxies can affect this. This chapter will also discuss the data used, how to get the most out of these data as well as the basics of the techniques and tools employed.

1.1 Star formation

At the basic level, stars form inside of clouds of hydrogen known as giant molecular clouds. If a region of a giant molecular cloud becomes dense enough such that the self-gravity of the region is greater than the sum of any outward pressure forces, for example magnetic or thermal, the region collapses. Thus it is easier to form stars inside cooler gas as there is lower thermal pressure. If the region is sufficiently massive, the gas can achieve sufficient density at its core to begin fusion of the hydrogen; it forms a star (e.g. Shu et al. 1987; McKee & Ostriker 2007). The stellar population affects the star formation rate

locally within its giant molecular cloud. Radiation pressure, supernovae, stellar winds, photoionisation, photoelectric heating and black hole activity, as examples, all have effects on the rate that the giant molecular cloud can collapse to form new stars (Hopkins et al. 2014).

The giant molecular clouds are small scale structures when compared to an entire galaxy. While it is possible to examine star formation in other galaxies on sub-galactic scales with high resolution observations, for this thesis I am concerned with global star formation rates, that is the average star formation rate across the entire galaxy. The star formation of concern here is therefore the combined activity of all of these giant molecular clouds within a galaxy. On these galactic scales, the conditions can be influenced by the larger scale environment that the galaxy itself is in. Galaxy interactions and mergers can drive the gas within a galaxy together and cause an increase in gas density inside the galaxies. This can result in a period of enhanced star formation (Kennicutt & Evans 2012).

1.1.1 Star formation rate tracers

There are a number of different ways of calculating the star formation rate of a galaxy, each with their own strengths, weaknesses and timescales that they probe. Star formation rates can be estimated by examining the spectral energy distribution and emission lines of a galaxy. The luminosity of a band can be used to determine the star formation rate using the equation (Kennicutt & Evans 2012)

$$\log\left(\frac{SFR}{M_{\odot}yr^{-1}}\right) = \log(L_x) - \log(C_x) \quad (1.1)$$

where L_x is the luminosity of the band, C_x is the star formation rate calibration constant and $1 M_{\odot}$ is the mass of the Sun. These relations, however, need luminosities that are corrected for dust extinction, as shown in Fig. 1.1, or they will underestimate the star formation rate. They are also all sensitive to the form of the initial mass function (e.g. Kennicutt 1998), that is the assumed distribution of low and high mass stars when they are born. Equivalent equations relate the emission from a spectral line to the star formation rate with the hydrogen line $H\alpha$ being commonly used to trace star formation on time scales of 10^7 years (e.g. Kennicutt 1983; Calzetti 2013). With the relations between the $H\alpha$ line and other hydrogen recombination lines, for example Bry , known from models and laboratory tests, the $H\alpha$ to star formation rate relation can be extended to other hydrogen lines in other ranges of the spectrum (e.g. Smajić et al. 2014).

The ultra-violet (UV) emission may initially appear to be a good choice of band to use for the estimation of the star formation rate. Integrated UV emission traces young stars, which emit brightly into this part of the spectrum, over approximately the last 200 Myr (e.g. Kennicutt & Evans 2012; Madau & Dickinson 2014). However, the UV suffers from acute sensitivity to dust absorption (e.g. Calzetti et al. 2000; Charlot & Fall 2000; Kennicutt & Evans 2012), as can be seen in Fig. 1.1. There is also sensitivity to the metallicity of the stars, with lower metallicity stars producing more UV emission (e.g. Madau & Dickinson 2014). UV based star formation rate estimates are also valid for time periods of less than 10^8 years as massive stars remain luminous in this part of the spectrum for an extended period (e.g. Madau & Dickinson 2014; Guidi et al. 2015).

Figure 1.1: Dust attenuation (A_λ/A_V) as a function of wavelength (λ) (Lo Faro et al. 2017, Fig. 1). Shown are attenuation curves from Calzetti et al. (2000, red), modified Calzetti et al. (2000) from Buat et al. (2011, salmon) and Charlot & Fall (2000, CF00; solid non-salmon). Different power law slopes (δ), as noted in the legend, are shown for Calzetti et al. (2000) as dot-dashed black lines and for Charlot & Fall (2000) as solid coloured lines. All curves have been normalised to unity at $\lambda \approx 0.55 \mu\text{m}$

Emission in the UV that is absorbed by dust causes the dust to heat and emit thermally in the IR. Thus, the IR emission can be used to trace star formation. As a result of the IR emission being due to UV absorption, star formation rates from IR tracers are also valid for time periods of less than 10^8 years. The mid-to-far IR (approximately $20\text{--}60\ \mu\text{m}$) is mainly emission from transiently heated, small grains while in the far IR ($\approx 60\text{--}500\ \mu\text{m}$) emission from steady-state, large grains dominates (e.g. Kennicutt & Evans 2012). Dust in galaxies is also heated by AGN activity and older stellar populations, so the derived star formation rate is sensitive to estimates of the age distribution of the stellar population (e.g. Kennicutt & Evans 2012). IR derived star formation rates also suffer from lacking the unobscured stellar emission, the amount of which can vary between galaxies (e.g. Kennicutt & Evans 2012).

Radio emission also offers a method to estimate star formation rates (e.g. Condon 1992; Kennicutt & Evans 2012). This arises from the tight correlation between radio and the far IR emission that holds over 5 orders of magnitude in luminosity and out to redshifts of at least 6 (e.g. Helou et al. 1985; Yun et al. 2001; Pannella et al. 2015; Delhaize et al. 2017). This correlation is believed to be rooted in massive star formation. The young stars emit strongly in the UV, which is reprocessed by dust as above and re-emits in the IR, and quickly die as supernovae. These supernovae accelerate cosmic rays and contribute to the non-thermal synchrotron emission in the radio (e.g. Condon 1992; Condon et al. 2002; Delhaize et al. 2017). Thermal radio emission can trace the production of ionising photons around younger stars via bremsstrahlung (free-free) emission (e.g. Tabatabaei et al. 2007, 2017; Mahajan et al. 2019). Radio has the advantage that observations with interferometric facilities can achieve much higher angular resolution than far IR facilities as well as the radio emission being free from dust extinction (e.g. Madau & Dickinson 2014). Non-thermal radio emission traces star formation over time scales of 10^8 years while thermal radio emission traces star formation over much shorter scales of 10^6 years (Condon et al. 2002; Bell 2003).

X-ray emission can trace X-ray binaries, supernovae and their remnants and massive stars, all of which are associated with recent star formation, across the integrated 2–10 keV range. Like the radio emission, the X-ray tracer of star formation rate is linked to the far IR emission, although the X-ray emission is also contaminated by AGN which emit strongly in this part of the spectrum. (e.g. Kennicutt & Evans 2012; Madau & Dickinson 2014)

1.1.2 Star formation over cosmic time

The average star formation rate of galaxies has not been constant across the history of the universe. Madau (2000) showed that the co-moving star formation rate density¹ peaked at a redshift of $z = 1.9$ and has been decreasing ever since. Using data from a number of surveys, Madau & Dickinson (2014) demonstrated that approximately half of the star formation happened before $z = 1.3$ with a quarter before the peak at $z = 1.9$ and a further quarter after $z = 0.7$. The results of their analysis is shown in Fig. 1.2 for both rest frame UV and rest frame IR star formation rate densities. The UV data are uncorrected for dust attenuation. Data used in this plot are mainly taken from post 2006 galaxy surveys with

¹The co-moving distance between two co-moving objects remains constant as the universe expands while the proper (physical) distance increases. Thus, the co-moving volume in which the co-moving star formation rate density is calculated remains constant as the universe expands while the proper volume increases.

Figure 1.2: Co-moving star formation rate density against redshift for rest frame far UV (green, blue and purple), and rest frame far IR (red) star formation rate densities. The UV star formation rate densities are uncorrected for dust attenuation. See Madau & Dickinson (2014) Table 1 for a full list of the data used. (Madau & Dickinson 2014, Fig. 8)

the redshift range for each survey varying from being local ($z = 0.01 - 0.1$) to high redshift ($z = 3.8 - 7.9$). Full information and references for the data sets used in Fig. 1.2 can be found in Madau & Dickinson (2014) Table 1.

The UV star formation rate densities presented in Fig. 1.2 are uncorrected for dust attenuation. At high redshift, $z \gtrsim 4$, the UV and IR star formation rate densities are in reasonable agreement but as the redshift decreases, the agreement becomes worse. As can be seen, the UV star formation rate densities are below the IR values by over 0.5 dex at low redshift, $z \approx 0$. This demonstrates the need to correct for dust attenuation to generate accurate star formation rate estimates. Hence the desire to get the most out of our existing far IR observations to capture the dust obscured emission, as will be done within this thesis.

Simulations offer the ability to study the causes of the evolution of the co-moving star formation rate density by tweaking parameters within the simulations to match observations. It has been found that both feedback from AGN and supernovae driven winds can play important roles in creating the shape of the high redshift ($z \gtrsim 3$) evolution seen in Fig. 1.2 (Katsianis et al. 2017). If the feedback from wind is not present, simulations tend to overestimate the star formation rate density at these redshifts (Katsianis et al. 2017). Some simulations have been able to reproduce the evolution of the star formation rate density without AGN feedback (Chiosi et al. 2017) while other simulations have not (Katsianis et al. 2017). The reasons for this are unclear. At lower redshifts, the decline in

the star formation rate density is a result of gas depletion. Star formation up to the peak of the star formation rate density has used the majority of the free gas in the universe, resulting in a steady decline in the star formation rate density from $z \approx 1.9$ to the present day (Katsianis et al. 2017).

The mass distribution of galaxies that contribute to the star formation rate density has also been studied. At high and low redshifts in simulations, the main contribution to the star formation rate density comes from lower mass galaxies (stellar mass $< 10^{10} M_{\odot}$) while at intermediate redshifts there are equal contributions from low and intermediate mass galaxies. At all redshifts, however, high mass galaxies have a lower contribution than low and intermediate mass galaxies (Chiosi et al. 2017). This is in disagreement with observations, at least at low redshift, which show the majority of star formation is contained in galaxies with the total mass of stars (stellar mass) between $10^{10} M_{\odot}$ and $10^{11} M_{\odot}$ (Brinchmann et al. 2004). Our own galaxy, the Milky Way, falls within this intermediate mass range with a stellar mass of $\approx 5 \times 10^{10} M_{\odot}$ (e.g. Licquia & Newman 2015; McMillan 2017). The majority of the star formation is also in galaxies that are steadily forming stars and lie on what has become known as the main sequence of star forming galaxies (Kauffmann et al. 2003).

1.1.3 The mass dependence of star formation rates

Galaxies are seen to exhibit a tight correlation between the star formation rate and stellar mass, what has (somewhat controversially) become known as the main sequence of star forming galaxies. Early works in the low redshift universe ($z < 0.2$) found that the star formation rate of a galaxy increases as the stellar mass of the galaxy increases with a scatter in this trend of approximately 0.3 dex as can be seen in Fig. 1.3 (Brinchmann et al. 2004). This relation was studied again by Noeske et al. (2007), from whom the name “main sequence of star forming galaxies” originated, and Elbaz et al. (2007) who both find a tight correlation between star formation rate and stellar masses in galaxies out to $z = 1.2$.

Since these early works, a large number of further studies have been undertaken. The main sequence of star forming galaxies has been found to exist at least out to a redshift of 6 with the scatter consistently found to be between 0.2 and 0.5 dex (Speagle et al. 2014, and references therein). This has allowed the exploration of how the main sequence of star forming galaxies evolves over the history of the universe. The majority of studies agree that the normalisation of the main sequence, what can loosely be considered to be the average star formation rate, increases with redshift approximately linearly with look-back time. The evolution of the slope of the main sequence, the connection between high and low mass galaxies, is less clear with a large spread of values at different redshifts (Speagle et al. 2014). The ‘consensus’ main sequence from pre-2014 works has been derived by Speagle et al. (2014) and demonstrates how the normalisation increases while the evolution of the slope is less well defined as is presented in Fig. 1.4. There is also no single, self-consistent study across a large redshift range, something that will be addressed by this thesis in Chapter 3.

The shape that the main sequence of star forming galaxies takes is still a point of contention. Early works found that the relation between star formation rate and stellar mass is linear in log space, although there were indications that the slope of the main

Figure 1.3: One of the first plots of the star formation rate (SFR) vs stellar mass (M_*) plane presented in Brinchmann et al. (2004, Fig. 17). The tight correlation can be clearly seen. The red dashed line is the average star formation rate at a given stellar mass while solid blue line is the mode star formation rate. The dashed blue lines show the limits that contain 95% of the galaxies at a given stellar mass.

Figure 1.4: The ‘consensus’ main sequence of star forming galaxies from Speagle et al. (2014, Fig. 8). This is the main sequence derived when combining the results from a large number of pre 2014 main sequence studies. As can be seen, the normalisation increases with redshift while the evolution of the slope is less clear.

sequence could flatten at high masses (Noeske et al. 2007). Subsequent works have found that the main sequence turns over at stellar masses above $\approx 10^{10} M_{\odot}$ (e.g. Whitaker et al. 2014; Lee et al. 2015; Tacchella et al. 2016). A small number of other works, however, still find and use a linear main sequence relation (e.g. Speagle et al. 2014; Renzini & Peng 2015; McPartland et al. 2019). The presence of a turnover is potentially due to contamination of the star forming galaxy sample by quiescent galaxies (Johnston et al. 2015; McPartland et al. 2019). If a turnover does exist it suggests that high mass galaxies have an additional quenching mechanism that partially suppresses star formation. Chapter 3 of this thesis will also explore this main sequence turnover.

The tight scatter of the main sequence has existed since at least redshift 6 and is found to be independent of the stellar mass (Whitaker et al. 2012; Speagle et al. 2014; Tomczak et al. 2016). The scatter is a result of minor fluctuations in the flow of material into galaxies (Tacchella et al. 2016; Mitra et al. 2017). As gas is compacted in the centre of a galaxy, star formation is triggered and the galaxy rises through the main sequence. Once this gas is depleted, the star formation rate is reduced and the galaxy sinks back down towards the bottom of the main sequence. New gas then falls into the galaxy, triggering further star formation and the cycle repeats until the time it takes to replenish the gas is longer than the depletion time of the galaxy. At this point the galaxy becomes quiescent and falls out of the main sequence of star forming galaxies (Tacchella et al. 2016).

Large cosmological simulations have also been used to study the main sequence of star forming galaxies, including Evolution and Assembly of GaLaxies and their Environments (EAGLE; Schaye et al. 2015), Illustris (Genel et al. 2014; Vogelsberger et al. 2014) and Illustris The Next Generation (IllustrisTNG; Pillepich et al. 2018). These simulations are typically able to recover the observed slope and scatter of observations (Furlong et al. 2015; Sparre et al. 2015; Donnari et al. 2019). At low redshifts, EAGLE, Illustris and IllustrisTNG are all consistent with observations although issues arise at $1 \lesssim z \lesssim 2$. The EAGLE simulation's main sequence has a normalisation that is ≈ 0.2 dex too high (Furlong et al. 2015), when compared with observations, while Illustris produces normalisations ≈ 0.2 dex too low (Sparre et al. 2015) and IllustrisTNG produces normalisations that are ≈ 0.5 dex too low (Donnari et al. 2019). The reason for this disparity in normalisations between simulations and observations is unclear. It is possible that the feedback mechanisms used in the simulations do not accurately represent what is going on in the Universe or it could be that contamination in star formation rate traces is causing the differences (Donnari et al. 2019). All three simulations also find a high mass turnover at stellar masses of $\approx 10^{10.5} M_{\odot}$, although the stellar mass where the turnover occurs reduces as redshift increases in EAGLE, which is not seen in observations (Furlong et al. 2015).

1.1.4 Galaxy mergers and their influence on star formation

Galaxy mergers are an important part of our understanding of how galaxies form and evolve and can potentially cause galaxies to migrate across the star formation rate vs stellar mass plane. In the cold dark matter paradigm, dark matter halos merge under hierarchical growth: smaller halos join together to form larger halos. As baryonic matter that forms visible galaxies lies within the potential well of the dark matter halos, this results in the galaxies colliding and merging with one another (e.g. Conselice 2014; Somerville &

Figure 1.5: Examples of galaxy mergers from the Sloan Digital Sky Survey (SDSS; York et al. 2000). The images are gri band composites, as the blue, red and green components respectively, and created using the SDSS SkyServer JPEG service. Each image is 256×256 pixels, corresponding to an angular size of 101×101 arcsecs.

Davé 2015), as can be seen in the examples in Fig. 1.5. The result of this interaction can cause stars within a galaxy’s disc to migrate to the central spheroidal component or mergers can result in disc galaxies becoming elliptical or spheroidal (Somerville & Davé 2015). Galaxy mergers are also thought to help feed the super massive black holes in the centre of galaxies causing them to become AGN (e.g. Sanders & Mirabel 1996), although more recent work has begun to cast doubt on this assertion (e.g. Darg et al. 2010; Weigel et al. 2018).

In addition, galaxy mergers are believed to be the triggering mechanism behind some of the brightest infrared objects known: (ultra) luminous infrared galaxies (Sanders & Mirabel 1996). High IR luminosity is linked with high star formation rates, as discussed in Sect. 1.1.1, which resulted in a prevailing interpretation from early work on mergers that most mergers go through a starburst phase, an increase in star formation rate of a factor of 5 or more (e.g. Joseph & Wright 1985; Schweizer 2005). More modern studies have also shown that mergers can induce starbursts. Typically, these studies find that the starburst population contains a larger proportion of galaxy mergers than non-mergers, with mergers making up 50% or more of the starburst population (Luo et al. 2014; Knapen & Cisternas 2015; Cibinel et al. 2019)

Observations have also shown that merger induced starbursts are found in the minority of galaxy mergers. Studies with a large number of merging galaxies have found that the typical increase seen in star formation rate while a galaxy is undergoing a merger is at most a factor of two, far too low to be considered to be a starburst and resulting in the merging galaxies remaining on the main sequence (Knapen et al. 2015; Silva et al. 2018). The majority of galaxy mergers appear to result in a reduction in star formation rate when compared to non-merging galaxies of the same stellar mass (Knapen et al. 2015), resulting in a migration to the bottom of the main sequence of star forming galaxies. In total, approximately 10-20% of galaxies on the main sequence are found to be undergoing a merger (Luo et al. 2014; Cibinel et al. 2019).

Simulations of merging galaxies have been conducted such as those in Feedback In Realistic Environments (FIRE; Hopkins et al. 2014). These allow us to study the star formation rates of the merging galaxies throughout the entire merger sequence. These simulations have shown that the star formation rate is briefly highly enhanced when the galaxies first pass each other, as can be seen in Fig. 1.6. The star formation rate then reduces but remains higher by approximately a factor of two than non-interacting galaxies before being enhanced again, and for a longer period, at the second pass and coalescence.

Figure 1.6: Star formation rate (SFR) of merging galaxies as a fraction of the SFR of non-merging galaxies as a function of time since the first passage (t) for merging galaxies in the FIRE simulation (Moreno et al. 2019, Fig. 6). The vertical dashed lines are (left to right) the time of first passage, second passage and coalescence. As can be seen, merging galaxies spend the majority of a merger period with only slightly enhanced star formation rates and only short periods as starbursts.

The majority of the merger period, approximately 2.5 Gyr out of 3.5 Gyr, is spent not undergoing a starburst but only slightly enhanced star formation (Moreno et al. 2019). This would explain why so few galaxy mergers are seen to be starbursts as the period during which the galaxies have highly enhanced star formation is short with respect to the entire merger sequence. The burst during passing and coalescence is strongest for head on collisions and reduces in strength as the approach of the two galaxies becomes more oblique. However, the strength of this starburst is also dependent on the resolution of the simulation, with lower resolution simulations finding fewer starburst mergers, typically less than observed in the real universe (Sparre & Springel 2016).

While observational evidence for the change in star formation rates as merger progresses is difficult to obtain due to the long timescales involved, there is evidence that starbursts are triggered on the first close passage of two merging galaxies with a second burst at coalescence that both appear to last 10^7 to 10^8 years. Early stage mergers, before first close passage, show a period of reduced star formation although this may be due to the merging galaxies in that study being relatively gas poor (Cortijo-Ferrero et al. 2017).

Gas rich merging galaxies are able to support high star formation rates as they have the fuel readily available to create new stars (e.g. Lin et al. 2008; Perez et al. 2011; Athanasoulas et al. 2016). Gas poor mergers, on the other hand, do not have gas readily available for star formation and so it is more difficult for these systems to host bursts of star formation (e.g. Bell et al. 2006; Naab et al. 2006; Lin et al. 2008). The gas poor mergers are typically found in denser environments than gas rich mergers, unsurprisingly as denser environments contain a larger number of gas poor and low star formation rate galaxies, and dominate over gas rich mergers in these environments (Lin et al. 2010). The fraction of gas poor mergers is also increasing as the universe ages (Lin et al. 2008). This thesis includes a large, statistical study into the effect galaxy mergers have on star formation

rates in Chapter 5 although this does not study the differences in gas rich and gas poor mergers.

1.2 Data used within this thesis

1.2.1 Far infrared for star formation rates

As discussed in Sect. 1.1.1, far IR data are key to accurately determining the star formation rates of galaxies. Thus, to get the best estimates of star formation rates we need to use the best available far IR data: data from *Herschel* (Pilbratt et al. 2010). *Herschel* is a prime choice of telescope as it covered the far IR to sub-mm part of the spectrum with spectral range of 55-671 μm (Pilbratt et al. 2010) allowing observation of the peak of the far IR emission. As *Herschel* was a space telescope, it also did not suffer from atmospheric absorption that causes issues with ground based telescopes that cover similar spectral ranges. Far IR observations are key for this thesis as they provide a strong constraint on the star formation rate while UV emission suffers from absorption and optical emission is a poor tracer of star formation.

The use of *Herschel* data also allows easy access to the popular multi-wavelength extragalactic fields. The *Herschel* Multi-tiered Extragalactic Survey (HerMES; Oliver et al. 2012) was designed to cover many of the popular multi-wavelength fields, such as The Cosmic Evolution Survey (COSMOS; Scoville et al. 2007) and The Great Observatories Origins Deep Survey (GOODS; Giavalisco et al. 2004), so there is a large amount of supplementary data that can be used to complement the *Herschel* data. These extra data will allow the examination of galactic properties that may be otherwise hidden in the far IR. For example, the UV data mentioned above can be used to complement the far IR data when examining star formation by removing the need to assume the fraction of UV emission that is unobscured.

However, *Herschel* data suffer from confusion (see Sect. 1.3.1) and thus need to be de-confused. To ensure that the de-confusion techniques developed and employed within this thesis provide improvements over existing methods, it is necessary to verify the de-confused *Herschel* data. Data from the Atacama Large Millimeter/sub-millimeter Array (ALMA) will be employed to do this. ALMA provides highly sensitive and high spatial resolution data of objects of interest with wavelengths ranging from 0.32 mm to 3.6 mm and plans to expand the wavelength coverage out to 8.5 mm. Observing at 870 μm , ALMA can achieve a resolution of 0.03 arcsec (ALMA Partnership et al. 2015) compared to the 5.2 arcsec resolution of *Herschel's* shortest observation wavelength of 70 μm (Poglitsch et al. 2010). The high altitude and low humidity position of this ground based facility in the Atacama Desert makes it a great place to observe the far IR and sub-mm regimes (the wavelengths that ALMA can observe that are less than 1 mm), although issues with atmospheric water vapour still provide complications at these wavelengths. Despite this, the wavelengths available with ALMA are perfect for constraining the star formation rates of galaxies when seeing conditions are good. ALMA is not designed to perform surveys of the sky, so data are typically collected on a galaxy by galaxy basis. As a result, use of the ALMA Science Archive² is required to perform large statistical studies although it is

²<http://almascience.eso.org/aq/>

currently not large enough to provide sample sizes similar to those from *Herschel*.

1.2.2 Identifying mergers

To be able to study galaxy mergers, it is first necessary to identify a sample of galaxy mergers. For this, optical and near IR imaging will be used, choosing surveys that already have small samples of merging galaxies. These small samples will be used to train machine learning tools to increase the sample size (see Sect. 1.3.2). Data of different depths will also be used, with wide and shallow surveys coupled with narrow and deep, so any changes with redshift can be studied. Sloan Digital Sky Survey (SDSS; York et al. 2000) and Kilo Degree Survey (KiDS; de Jong et al. 2013a,b) optical images as well as near IR Hubble Space Telescope images from the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011) will be used as all three of these surveys already have small samples of identified merging galaxies within small regions that can be expanded to cover larger areas of these surveys.

SDSS is an optical imaging and spectroscopy survey originally designed to survey half of the northern sky using a dedicated 2.5 m telescope. The survey began in earnest in 2000 and to date has had 15 data releases (Aguado et al. 2019) with the 16th scheduled for the end of 2019. The data from SDSS now includes the extended Baryon Oscillation Spectroscopic Survey (eBOSS; Dawson et al. 2016), which surveys galaxies and quasars at redshifts $z \approx 0.6 - 3.5$, IR spectra from the Apache Point Observatory Galactic Evolution Experiment (APOGEE; Majewski et al. 2017) and integral field unit spectroscopy for $\approx 10\,000$ galaxies from Mapping Nearby Galaxies at Apache Point Observatory (MaNGA; Bundy et al. 2015).

KiDS is another optical imaging survey that aims to image 1500 square degrees in four optical bands: ugr_i at 358 nm, 476 nm, 632 nm and 759 nm respectively. The survey is undertaken using the OmegaCAM instrument on the Very Large Telescope Survey Telescope (de Jong et al. 2013a,b) providing higher spatial resolution than SDSS. Currently, there have been four data releases covering a total area of 1006 square degrees (Kuijken et al. 2019). These data releases have been coordinated with the data releases of the companion Visible and Infrared Survey Telescope for Astronomy (VISTA) Kilo-degree INfrared Galaxy survey (VIKING; Edge et al. 2013) that provides a further five near IR band images in ZYJHKs at 889 nm, 1.0 μm , 1.3 μm , 1.6 μm and 2.2 μm respectively.

CANDELS is the largest Hubble Space Telescope project ever undertaken. This near IR survey aimed to image over 250 000 galaxies with redshifts between $z = 1.5$ and $z = 8.0$ in the HST bands from 0.6 μm to 1.6 μm . The survey covers a total of 800 square arcmins within five popular extragalactic fields: COSMOS, GOODS South, GOODS North, Extended Groth Strip (EGS; Davis et al. 2007) and the United Kingdom Infrared Telescope (UKIRT) Infrared Deep Sky Survey (UKIDSS) Ultra-Deep Survey (UDS; Lawrence et al. 2007; Cirasuolo et al. 2007). This survey was designed to characterise the evolution of galaxies in the young universe but the data also have applications in the low redshift universe.

Simulations of our universe will also be used to study galaxy mergers. This will be done to compare real and simulated galaxies to determine any biases that may exist in our observed merger selection and identify places where simulations do not accurately represent the Universe. The Evolution and Assembly of GaLaxies and their Environments

(EAGLE; Schaye et al. 2015) simulation is a cosmological hydrodynamical simulation designed to study how galaxies form and evolve and so is ideal for this work. It uses $\approx 7 \times 10^9$ particles to simulate the dark and baryonic matter and begins when the universe is uniform and without stars or galaxies. Initial conditions are set based on observations of the cosmic microwave background by the Planck Satellite (Tauber et al. 2010). The EAGLE simulation achieves a resolution of 0.7 pc for gravity and so must rely on sub-grid models for processes that function at scales below this, for example the evolution of individual stars from birth to death, the formation of and accretion onto black holes and the distribution of metals and energy in the interstellar medium (Crain et al. 2009; Schaye et al. 2010; Le Brun et al. 2014; Schaye et al. 2015). These sub-grid models are calibrated to match observations or are tweaked by hand to produce realistic results. The outputs of the simulation are saved at specific redshift points, creating what are known as snapshots, for lower time resolution between 90 and 950 Myr (McAlpine et al. 2016), or snipshots, for higher time resolution between 10 and 60 Myr (Rosas-Guevara et al. 2016).

1.3 Methodology and tools employed within this thesis

1.3.1 Dealing with *Herschel* confusion

For *Herschel* Spectral and Photometric Imaging REceiver (SPIRE; Griffin et al. 2010) data (250-500 μm), there is a large amount of confusion in the data (e.g. Nguyen et al. 2010). Confusion is additional noise caused by signals from faint, unresolved sources on scales that are comparable to the beam of the instrument and is a major source of noise in far IR and sub-mm observations (Condon 1974; Blain et al. 2002). Confusion is a problem that becomes more prevalent at longer wavelengths, so is rarely an issue in optical and near IR observations, and is exacerbated with smaller apertures or higher sensitivity instruments. *Herschel* observations also suffer from source blending. With the smallest SPIRE beam size of 18 arcsec and a typical separation of galaxies on the sky of ≈ 4 arcsec, two or more objects can reside within the same beam and become difficult, or impossible, to distinguish. The result is a pixel which contains flux density from more than one object, making it hard to accurately assign the received flux to the correct object. In many cases, it can also result in two, or more, objects blending and appearing as a single object. Figure 1.7 shows how confusion and blending can become more of an issue as the wavelength increases.

Source confusion and blending have an influence on science results. A recent study by Wang et al. (2019) has explored the far IR luminosity function, the number of galaxies as a function of luminosity in the far IR, using data *Herschel* data that has been de-blended using the techniques developed in Chapter 2 of this thesis. The new de-blended *Herschel* data are compared to older *Herschel* results and results from facilities that suffer less from confusion and blending, such as the Submillimetre Common-User Bolometer Array 2 (SCUBA-2; Holland et al. 2013). Typically, the longer wavelength *Herschel* far IR luminosity functions were above those of SCUBA-2 at high flux densities. However, with the latest de-blending techniques the *Herschel* luminosity function is brought into agreement with the SCUBA-2 data, highlighting the importance of correctly dealing with confusion and blending. Thus, Chapters 2, 3, and 5 of this thesis will employ Bayesian inference alongside Markov Chain Monte Carlo (MCMC) methods to aid assigning flux

Figure 1.7: 1×1 arcmin tiles from COSMOS imaged with different telescopes at different wavelengths: (top row, left to right) Subaru $0.8 \mu\text{m}$, IRAC $3.6 \mu\text{m}$, MIPS $24 \mu\text{m}$, PACS $100 \mu\text{m}$, (middle row, left to right) SPIRE $250 \mu\text{m}$, SPIRE $350 \mu\text{m}$, SPIRE $500 \mu\text{m}$, SCUBA-2 $450 \mu\text{m}$, (bottom row, left to right) SCUBA-2 $850 \mu\text{m}$, AzTEC 1.1 mm , MAMBO 1.2 mm and VLA 20 cm . (Casey et al. 2014)

Figure 1.8: Plot from Hurley et al. (2017, Fig. 3) showing the 16th, 50th and 84th percentiles for flux density accuracy from DESPHOT (dashed lines) with the 50th percentile for XID+ (coloured line) and corresponding 16th to 84th percentile region (shaded region) as a function of source flux density. The results from SPIRE 250 μm (blue, left), 350 μm (green, centre) and 500 μm (red, right) are shown. It can be seen that XID+ performs better than DESPHOT and produces more realistic errors.

density values to each source. There are already tools that are able to do this, such as DESPHOT (Roseboom et al. 2010) and XID+ (Hurley et al. 2017).

DESPHOT is designed to work with SPIRE data from *Herschel*. The program works in map space and directly uses the images. It splits up the image based on the signal to noise ratio (SNR); areas of high SNR are created which are bounded by low SNR. An input source list is used and DESPHOT assumes that the image can be made from the fluxes associated to the sources. The sources are ‘activated’ one at a time until the simulated image matches the observations. Using this method, sources that do not contribute to the observed image can be removed to prevent oversampling. (Roseboom et al. 2010; Wang et al. 2014)

XID+ is also designed to work with *Herschel* SPIRE data. The program takes a prior catalogue and makes a source cut at a defined flux threshold, for example a 100 μm catalogue may have sources with a flux density of less than 50 μJy removed. A flux value for each SPIRE band is assigned to the remaining sources using Bayesian statistics and a SPIRE observation is simulated. The resulting map is compared to a real observation and MCMC is employed to converge the flux values with the observed map. (Hurley et al. 2017).

This thesis makes extensive use of XID+. XID+ is preferred over DESPHOT as it provides the entire posterior distribution for each source and so allows a better estimate of the uncertainty to be produced, as shown in Fig. 1.8 (Hurley et al. 2017). Through the course of the thesis, the code of XID+ will be modified to tailor it to the needs of the project. This will involve including a more informed flux density prior through spectral energy distribution information as well as determining the ideal depth of a prior catalogue.

1.3.2 Machine learning

For the work studying galaxy mergers, a fast and reliable method is required to accurately identify merging galaxies. While it is possible, classification by hand is time consuming and unreliable, with the same galaxy potentially being given different classifications when viewed different times. As an alternative that is much faster and repeatable, state of the

art machine learning techniques will be employed within this thesis

These algorithms are designed to allow computers to perform specific tasks without pre-defined instructions. Instead of rules defined by a researcher, machine learning employs pattern recognition routines to derive trends in data that can be applied to new, unseen information. The subset of machine learning used within this thesis is known as neural networks. This type of machine learning is designed to loosely mimic the way an organic brain interprets information as it is passed through layers of biological neurons. In neural networks, the biological neurons are replaced with layers of non-linear mathematical functions, dubbed neurons, of the form

$$x' = f(\mathbf{w}\mathbf{x} + b). \quad (1.2)$$

Here, x' is the output from the neuron, $\mathbf{x}$ are the inputs from other neurons, $\mathbf{w}$ are known as the weights, b is the bias term and f is the activation function. The weights and biases are the terms that are ‘learnt’ during the training of a neural network, typically using stochastic gradient descent (Robbins & Monro 1951) to minimise the loss of the network, whichever way the user has defined the loss. The activation function is where the neuron gains its non-linearity. These functions scale $\mathbf{w}\mathbf{x} + b$ or force x' to a certain value, depending on the value of $\mathbf{w}\mathbf{x} + b$ and the activation function used. Examples of activation functions include Regularised Linear Units (ReLU; Nair & Hinton 2010), which is $\max(0, y)$, or tanh, which is $\tanh(y)$.

For this thesis, a specific type of neural network known as a convolutional neural network will be used. Convolutional neural networks are designed to work on multi-dimensional data, such as audio or images, and employ layers with kernels that are convolved with the output from lower layers, giving this structure its name. The kernels are small groups of neurons that are scanned across the input, the size and shape of which are predetermined before the network is trained. As with any other neural network, the weights and bias terms of the neurons in a kernel are learnt during training. The upper layers of a convolutional neural network are of a more typical form: they are 1 dimensional and fully connected, that is to say all neurons in one layer are connected to all the neurons in the layer below. A more thorough and complete description of convolutional neural networks can be found in Lecun et al. (1998).

For training I employ supervised learning. Here, the input data are labelled and the loss of the network can be considered to be the difference between the label assigned to the data by the network and the true label of the data. This differs from unsupervised learning where a network is not given any pre-defined groups or labels and must generate these itself.

Convolutional neural networks are already being employed in astronomy, typically to classify objects. For example, convolutional neural networks have seen use in star-galaxy classification (e.g. Kim & Brunner 2017), galaxy morphology classification (e.g. Huertas-Company et al. 2015), gravitational lens identification (e.g. Petrillo et al. 2017; Davies et al. 2019) and to generate photometric redshift estimates (e.g. Hoyle 2016). They have also been used to mimic the distributions of a large number of human classifications (Dieleman et al. 2015) and galaxy merger identifications (e.g. Ackermann et al. 2018), the latter of which will also be done within this thesis. Examples of galaxies that were not used for training and classified by a convolutional neural network in Huertas-Company et al. (2015) are shown in Fig. 1.9.

Figure 1.9: Hubble Space Telescope imaged galaxies that have been classified by a convolutional neural network in Huertas-Company et al. (2015, Fig. 13) and have not been used for training. These galaxies are classified as spheroid dominated (left), disc dominated (centre left), irregular (centre right) or a point source (right).

1.4 This thesis

Throughout this thesis, I will study how the properties and environment of galaxies affect their star formation rates. Primarily, this thesis will aim to address the following questions:

- **How is the star formation rate of a galaxy related to its stellar mass and how does this evolve with time?**
- **How do galaxy mergers influence the star formation rates of the colliding galaxies?**
- **Can we use images from simulations to identify merging galaxies in the real universe?**

To address these questions, the thesis requires large samples of galaxies over a large redshift range and with reliable star formation rate estimates. The former requirement raises problems with identifying merging galaxies, a task that can be both time consuming and unreliable. To combat this, this thesis will apply state of the art deep learning techniques, convolutional neural networks, to identifying galaxy mergers in a number of different optical and near IR data sets (Chapters 4 and 5). To address the latter requirement, this thesis will attempt to improve state of the art de-blending tools to allow more reliable measurements of far IR flux densities (Chapter 2). These will allow better constraints on star formation rate, as discussed in Sects. 1.1.1 and 1.3.1, and allow a detailed study into the relation between star formation rate and stellar mass (Chapter 3).

These cutting edge tools will be coupled with state of the art observations. Data from the far IR space telescope *Herschel* will be used to generate highly reliable star formation rate estimates. Images from SDSS, CANDELS and KiDS will be used to identify merging galaxies on a scale that has been difficult to match until recently.

The work presented in this thesis is work that is also published in peer-reviewed journals (Chapters 2, 3 and 4) or is close to submission (Chapter 5). A brief description of the content of each chapter is presented below.

Chapter 2 - In this chapter, I expand the Bayesian de-blending tool XID+ to include a more informed flux density prior. Doing so will allow more accurate far IR flux

density measurements which will help improve star formation rate estimates. I also present a pipeline that can be used to generate these priors. The improvement this gives over the standard XID+ flux density prior is demonstrated and quantified.

Chapter 3 - In this chapter, I apply the techniques developed in the previous chapter to de-blend *Herschel* SPIRE data in the 2 deg² COSMOS extragalactic field. I then use these data to create improved star formation rate estimates for over 200 000 galaxies covering the redshift range $0.2 < z < 6.0$. I then place these galaxies onto the star formation rate - stellar mass plane to find and study the galaxy main sequence and its evolution. This results in one of the largest redshift ranges for a self-consistent main sequence study.

Chapter 4 - In this chapter, I develop a convolutional neural network to identify galaxy mergers and train it twice: once with observational images from SDSS and once with simulation images from EAGLE. With these networks, I am able to identify the strengths and weaknesses of observational and simulation galaxy mergers. Also, by using cross application of the networks, it is possible to use simulated images of galaxies to train a network for use on real observations, something that is likely to be of great use in the large, upcoming surveys such as the Large Synoptic Survey Telescope (LSST Science Collaboration et al. 2009) and *Euclid* (Laureijs et al. 2011).

Chapter 5 - In this chapter, I train the architecture developed in Chapter 4 to detect galaxy mergers on two further data sets: CANDELS and KiDS. KiDS data sets provide shallow data that is higher resolution and deeper than SDSS and CANDELS provides much deeper data that can probe higher redshifts. The SDSS, KiDS and CANDELS data can therefore be used to study low, intermediate and high redshifts respectively. I then use the merger identifications from the SDSS, CANDELS and KiDS data to determine how, and if, galaxy mergers affect the star formation rates of galaxies. This is done by examining the differences in specific star formation rate between the merging and non-merging galaxies as well as seeing if the merging galaxies shift in the star formation rate - stellar mass plane with respect to the non-merging galaxies.

Chapter 6 - In this chapter, I provide the conclusions of the thesis and remark on future avenues of work.

Bibliography

- Ackermann, S., Schawinski, K., Zhang, C., Weigel, A. K., & Turp, M. D. 2018, *MNRAS*, 479, 415
- Aguado, D. S., Ahumada, R., Almeida, A., et al. 2019, *ApJS*, 240, 23
- ALMA Partnership, Brogan, C. L., Pérez, L. M., et al. 2015, *ApJ*, 808, L3
- Athanassoula, E., Rodionov, S. A., Peschken, N., & Lambert, J. C. 2016, *ApJ*, 821, 90
- Barkana, R., & Loeb, A. 2001, *Phys. Rep.*, 349, 125
- Bell, E. F. 2003, *ApJ*, 586, 794
- Bell, E. F., Naab, T., McIntosh, D. H., et al. 2006, *ApJ*, 640, 241
- Blain, A. W., Smail, I., Ivison, R. J., Kneib, J.-P., & Frayer, D. T. 2002, *Phys. Rep.*, 369, 111
- Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, *MNRAS*, 351, 1151
- Bromm, V. 2009, in *IAU Symposium*, Vol. 254, *The Galaxy Disk in Cosmological Context*, ed. J. Andersen, Nordströara, B. m, & J. Bland-Hawthorn, 337–342
- Bromm, V., & Yoshida, N. 2011, *ARA&A*, 49, 373
- Buat, V., Giovannoli, E., Heinis, S., et al. 2011, *A&A*, 533, A93
- Bundy, K., Bershady, M. A., Law, D. R., et al. 2015, *ApJ*, 798, 7
- 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
- Casey, C. M., Narayanan, D., & Cooray, A. 2014, *Phys. Rep.*, 541, 45
- Charlot, S., & Fall, S. M. 2000, *ApJ*, 539, 718
- Chiosi, C., Sciaratta, M., D’Onofrio, M., et al. 2017, *ApJ*, 851, 44
- Cibinel, A., Daddi, E., Sargent, M. T., et al. 2019, *MNRAS*, 485, 5631
- Cirasuolo, M., McLure, R. J., Dunlop, J. S., et al. 2007, *MNRAS*, 380, 585

- Condon, J. J. 1974, *ApJ*, 188, 279
- . 1992, *ARA&A*, 30, 575
- Condon, J. J., Cotton, W. D., & Broderick, J. J. 2002, *AJ*, 124, 675
- Conselice, C. J. 2014, *ARA&A*, 52, 291
- Cortijo-Ferrero, C., González Delgado, R. M., Pérez, E., et al. 2017, *A&A*, 607, A70
- Crain, R. A., Theuns, T., Dalla Vecchia, C., et al. 2009, *MNRAS*, 399, 1773
- Darg, D. W., Kaviraj, S., Lintott, C. J., et al. 2010, *MNRAS*, 401, 1552
- Davies, A., Serjeant, S., & Bromley, J. M. 2019, *MNRAS*, arXiv:1905.04303
- Davis, M., Guhathakurta, P., Konidaris, N. P., et al. 2007, *ApJ*, 660, L1
- Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, *AJ*, 151, 44
- de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013a, *Experimental Astronomy*, 35, 25
- de Jong, J. T. A., Kuijken, K., Applegate, D., et al. 2013b, *The Messenger*, 154, 44
- Delhaize, J., Smolčić, V., Delvecchio, I., et al. 2017, *A&A*, 602, A4
- Dieleman, S., Willett, K. W., & Dambre, J. 2015, *MNRAS*, 450, 1441
- Donnari, M., Pillepich, A., Nelson, D., et al. 2019, *MNRAS*, 485, 4817
- Edge, A., Sutherland, W., Kuijken, K., et al. 2013, *The Messenger*, 154, 32
- Elbaz, D., Daddi, E., Le Borgne, D., et al. 2007, *A&A*, 468, 33
- Furlong, M., Bower, R. G., Theuns, T., et al. 2015, *MNRAS*, 450, 4486
- Genel, S., Vogelsberger, M., Springel, V., et al. 2014, *MNRAS*, 445, 175
- Gialalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, *ApJ*, 600, L93
- Griffin, M. J., Abergel, A., Abreu, A., et al. 2010, *A&A*, 518, L3
- Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, *ApJS*, 197, 35
- Guidi, G., Scannapieco, C., & Walcher, C. J. 2015, *MNRAS*, 454, 2381
- Helou, G., Soifer, B. T., & Rowan-Robinson, M. 1985, *ApJ*, 298, L7
- Holland, W. S., Bintley, D., Chapin, E. L., et al. 2013, *MNRAS*, 430, 2513
- Hopkins, P. F., Kereš, D., Oñorbe, J., et al. 2014, *MNRAS*, 445, 581
- Hoyle, B. 2016, *Astronomy and Computing*, 16, 34
- Huertas-Company, M., Gravet, R., Cabrera-Vives, G., et al. 2015, *ApJS*, 221, 8

- Hurley, P. D., Oliver, S., Betancourt, M., et al. 2017, *MNRAS*, 464, 885
- Johnston, R., Vaccari, M., Jarvis, M., et al. 2015, *MNRAS*, 453, 2540
- Joseph, R. D., & Wright, G. S. 1985, *MNRAS*, 214, 87
- Katsianis, A., Tescari, E., Blanc, G., & Sargent, M. 2017, *MNRAS*, 464, 4977
- Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, *MNRAS*, 341, 33
- Kennicutt, R. C., & Evans, N. J. 2012, *ARA&A*, 50, 531
- Kennicutt, Jr., R. C. 1983, *ApJ*, 272, 54
- . 1998, *ApJ*, 498, 541
- Kim, E. J., & Brunner, R. J. 2017, *MNRAS*, 464, 4463
- Knapen, J. H., & Cisternas, M. 2015, *ApJ*, 807, L16
- Knapen, J. H., Cisternas, M., & Querejeta, M. 2015, *MNRAS*, 454, 1742
- Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, *ApJS*, 197, 36
- Kuijken, K., Heymans, C., Dvornik, A., et al. 2019, *A&A*, 625, A2
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints, arXiv:1110.3193
- Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, *MNRAS*, 379, 1599
- Le Brun, A. M. C., McCarthy, I. G., Schaye, J., & Ponman, T. J. 2014, *MNRAS*, 441, 1270
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998, in *Proceedings of the IEEE*, Vol. 86, Issue 11, 2278–2324
- Lee, N., Sanders, D. B., Casey, C. M., et al. 2015, *ApJ*, 801, 80
- Licquia, T. C., & Newman, J. A. 2015, *ApJ*, 806, 96
- Lin, L., Patton, D. R., Koo, D. C., et al. 2008, *ApJ*, 681, 232
- Lin, L., Cooper, M. C., Jian, H.-Y., et al. 2010, *ApJ*, 718, 1158
- Lo Faro, B., Buat, V., Roehlly, Y., et al. 2017, *MNRAS*, 472, 1372
- LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv e-prints, arXiv:0912.0201
- Luo, W., Yang, X., & Zhang, Y. 2014, *ApJ*, 789, L16
- Madau, P. 2000, *Physica Scripta* Volume T, 85, 156
- Madau, P., & Dickinson, M. 2014, *ARA&A*, 52, 415

- Mahajan, S., Ashby, M. L. N., Willner, S. P., et al. 2019, *MNRAS*, 482, 560
- Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, *AJ*, 154, 94
- McAlpine, S., Helly, J. C., Schaller, M., et al. 2016, *Astronomy and Computing*, 15, 72
- McKee, C. F., & Ostriker, E. C. 2007, *ARA&A*, 45, 565
- McMillan, P. J. 2017, *MNRAS*, 465, 76
- McPartland, C., Sanders, D. B., Kewley, L. J., & Leslie, S. K. 2019, *MNRAS*, 482, L129
- Mitra, S., Davé, R., Simha, V., & Finlator, K. 2017, *MNRAS*, 464, 2766
- Moreno, J., Torrey, P., Ellison, S. L., et al. 2019, *MNRAS*, 485, 1320
- Naab, T., Khochfar, S., & Burkert, A. 2006, *ApJ*, 636, L81
- Nair, V., & Hinton, G. E. 2010, in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 807–814
- Nguyen, H. T., Schulz, B., Levenson, L., et al. 2010, *A&A*, 518, L5
- Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, *ApJ*, 660, L43
- Oliver, S. J., Bock, J., Altieri, B., et al. 2012, *MNRAS*, 424, 1614
- Pannella, M., Elbaz, D., Daddi, E., et al. 2015, *ApJ*, 807, 141
- Perez, J., Michel-Dansac, L., & Tissera, P. B. 2011, *MNRAS*, 417, 580
- Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2017, *MNRAS*, 472, 1129
- Pilbratt, G. L., Riedinger, J. R., Passvogel, T., et al. 2010, *A&A*, 518, L1
- Pillepich, A., Springel, V., Nelson, D., et al. 2018, *MNRAS*, 473, 4077
- Poglitsch, A., Waelkens, C., Geis, N., et al. 2010, *A&A*, 518, L2
- Rees, M. J. 1993, *QJRAS*, 34, 279
- Renzini, A., & Peng, Y.-j. 2015, *ApJ*, 801, L29
- Robbins, H., & Monro, S. 1951, *Ann. Math. Statist.*, 22, 400
- Rosas-Guevara, Y., Bower, R. G., Schaye, J., et al. 2016, *MNRAS*, 462, 190
- Roseboom, I. G., Oliver, S. J., Kunz, M., et al. 2010, *MNRAS*, 409, 48
- Sanders, D. B., & Mirabel, I. F. 1996, *ARA&A*, 34, 749
- Schaye, J., Dalla Vecchia, C., Booth, C. M., et al. 2010, *MNRAS*, 402, 1536
- Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, *MNRAS*, 446, 521

- Schweizer, F. 2005, in *Astrophysics and Space Science Library*, Vol. 329, *Starbursts: From 30 Doradus to Lyman Break Galaxies*, ed. R. de Grijs & R. M. González Delgado, 143
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, 172, 1
- Shu, F. H., Adams, F. C., & Lizano, S. 1987, *ARA&A*, 25, 23
- Silva, A., Marchesini, D., Silverman, J. D., et al. 2018, *ApJ*, 868, 46
- Smajić, S., Moser, L., Eckart, A., et al. 2014, *A&A*, 567, A119
- Somerville, R. S., & Davé, R. 2015, *ARA&A*, 53, 51
- Sparre, M., & Springel, V. 2016, *MNRAS*, 462, 2418
- Sparre, M., Hayward, C. C., Springel, V., et al. 2015, *MNRAS*, 447, 3548
- Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, *ApJS*, 214, 15
- Tabatabaei, F. S., Beck, R., Krügel, E., et al. 2007, *A&A*, 475, 133
- Tabatabaei, F. S., Schinnerer, E., Krause, M., et al. 2017, *ApJ*, 836, 185
- Tacchella, S., Dekel, A., Carollo, C. M., et al. 2016, *MNRAS*, 457, 2790
- Tauber, J. A., Mandolesi, N., Puget, J.-L., et al. 2010, *A&A*, 520, A1
- Tomczak, A. R., Quadri, R. F., Tran, K.-V. H., et al. 2016, *ApJ*, 817, 118
- Vogelsberger, M., Genel, S., Springel, V., et al. 2014, *MNRAS*, 444, 1518
- Wang, L., Pearson, W. J., Cowley, W., et al. 2019, *A&A*, 624, A98
- Wang, L., Viero, M., Clarke, C., et al. 2014, *MNRAS*, 444, 2870
- Weigel, A. K., Schawinski, K., Treister, E., Trakhtenbrot, B., & Sanders, D. B. 2018, *MNRAS*, 476, 2308
- Whitaker, K. E., van Dokkum, P. G., Brammer, G., & Franx, M. 2012, *ApJ*, 754, L29
- Whitaker, K. E., Franx, M., Leja, J., et al. 2014, *ApJ*, 795, 104
- York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, *AJ*, 120, 1579
- Yun, M. S., Reddy, N. A., & Condon, J. J. 2001, *ApJ*, 554, 803

De-blending deep *Herschel* surveys: A multi-wavelength approach

W. J. Pearson, L. Wang, F. F. S. van der Tak, P. D. Hurley, D. Burgarella, S. J. Oliver
Astronomy & Astrophysics, 603, A102, 2017
Reproduced with permission, ©ESO

Abstract

Aims: Cosmological surveys in the far-infrared are known to suffer from confusion. The Bayesian de-blending tool, XID+, currently provides one of the best ways to de-confuse deep *Herschel* SPIRE images, using a flat flux density prior. This work is to demonstrate that existing multi-wavelength data sets can be exploited to improve XID+ by providing an informed prior, resulting in more accurate and precise extracted flux densities.

Methods: Photometric data for galaxies in the COSMOS field were used to constrain spectral energy distributions (SEDs) using the fitting tool CIGALE. These SEDs were used to create Gaussian prior estimates in the SPIRE bands for XID+. The multi-wavelength photometry and the extracted SPIRE flux densities were run through CIGALE again to allow us to compare the performance of the two priors. Inferred ALMA flux densities ($F_{\text{ALMA}}^{\text{infer}}$), at 870 μm and 1250 μm , from the best fitting SEDs from the second CIGALE run were compared with measured ALMA flux densities ($F_{\text{ALMA}}^{\text{meas}}$) as an independent performance validation. Similar validations were conducted with the SED modelling and fitting tool MAGPHYS and modified black-body functions to test for model dependency.

Results: We demonstrate a clear improvement in agreement between the flux densities extracted with XID+ and existing data at other wavelengths when using the new informed Gaussian prior over the original uninformed prior. The residuals between $F_{\text{ALMA}}^{\text{meas}}$ and $F_{\text{ALMA}}^{\text{infer}}$ were calculated. For the Gaussian priors these residuals, expressed as a multiple of the ALMA error (σ), have a smaller standard deviation, 7.95 σ for the Gaussian prior compared to 12.21 σ for the flat prior; reduced mean, 1.83 σ compared to 3.44 σ ; and have reduced skew to positive values, 7.97 compared to 11.50. These results were determined to not be significantly model dependent. This results in statistically more reliable SPIRE flux densities and hence statistically more reliable infrared luminosity estimates.

2.1 Introduction

Infrared (IR) radiation makes up approximately half of the total extragalactic emission that we observe (e.g. Dole et al. 2006; Burgarella et al. 2013). As a result, it is important to observe at these wavelengths to gain a better understanding of our Universe. The ESA *Herschel Space Observatory* (Pilbratt et al. 2010) Spectral and Photometric Imaging Receiver (SPIRE; Griffin et al. 2010) cosmological surveys probe the far-IR but are known to be affected by source confusion (e.g. Nguyen et al. 2010; Oliver et al. 2012), with object separation in the COSMOS2015 catalogue (Laigle et al. 2016) $\lesssim 8''$ and the SPIRE smallest beam size of $18''$ (Griffin et al. 2010). There have been a number of tools created to de-blend SPIRE images, such as DESPHOT (Roseboom et al. 2010, 2012; Wang et al. 2014) or T-PHOT (Merlin et al. 2015). These tools mostly use maximum likelihood estimation to generate flux density estimates using galaxy positions extracted from a shorter wavelength image.

More recently, the probabilistic, Bayesian de-blender XID+ (Hurley et al. 2017) has been developed, overcoming the main weaknesses of DESPHOT, namely its tendency to assign all the flux density to one source when many sources are within a single beam, as well as the poor estimation of variance and co-variance of sources. This is achieved by exploring the full posterior distribution (see Sect. 2.3.1) using the Bayesian statistical inference tool Stan (Carpenter et al. 2017). As a result of posterior exploration, XID+ produces much better flux density precision with more realistic associated uncertainties than DESPHOT across all three SPIRE bands (Hurley et al. 2017). However, the current, publicly available XID+ only uses a flat and uninformed flux density prior.

The aim of this paper is to show how the performance of XID+ can be advanced by exploiting the large amount of multi-wavelength data available in the popular COSMOS field (Scoville et al. 2007) as part of the continuing development of XID+, although the technique can be applied in any deep field with multi-wavelength data available. Throughout this paper, Wilkinson Microwave Anisotropy Probe year 7 cosmology (Komatsu et al. 2011; Larson et al. 2011) is followed: $\Omega_M = 0.27$, $\Omega_\Lambda = 0.73$, and $H_0 = 70.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

2.2 Data

For this work, a large multi-wavelength data set from a field covered by the *Herschel* Multi-tiered Extragalactic Survey (HerMES; Oliver et al. 2012) was required. The multi-wavelength data is needed to generate flux density priors for de-blending the SPIRE images from HerMES. A secondary data set is also required for validation of the extracted flux densities. For this we chose data from the Atacama Large Millimeter/submillimeter Array (ALMA), due to its higher resolution and because the $850 \mu\text{m}$ and $1100 \mu\text{m}$ bands are dominated by the same emission as the SPIRE bands (e.g. Scoville et al. 2014, 2016).

Thermal emission dominates at wavelengths of approximately $50 \mu\text{m}$ or more and contains the Rayleigh-Jeans (RJ) tail from approximately $100 \mu\text{m}$ to $1100 \mu\text{m}$ in the rest frame (e.g. Draine 2003; Draine & Li 2007). Thus, at least one SPIRE band lies within the RJ tail up to redshifts of 4 and all three SPIRE bands are dominated by thermal emission at all redshifts studied here. Similarly, the $870 \mu\text{m}$ ALMA band remains in the RJ tail for

the entirety of this study, while the 1250 μm ALMA band enters the RJ region at redshifts of 0.14 and so was chosen for high redshift objects only. Evidently, the SPIRE bands and the two ALMA bands chosen for this study are both dominated by the same physical processes, certainly up to redshifts of at least 4, making ALMA data a good choice to validate the SPIRE data.

The COSMOS field was chosen because of the prevalence of multi-wavelength data, and because the SPIRE images within COSMOS are considered to be reasonably homogeneous and of high quality (see Appendix 2.6.1). The COSMOS2015 catalogue (Laigle et al. 2016) was used as an ancillary data set as it contains data on over 1.2 million objects in over 30 wavelength bands. For this work, only bands from the ultraviolet to the mid-IR were used. The *Subaru* narrow bands were omitted as similar wavelengths are covered by the *Subaru* intermediate bands¹. The *Spitzer* Multiband Imaging Photometer (Rieke et al. 2004) 24 μm data were also not used as they are also noticeably affected by confusion, although not to the same extent as the SPIRE data (e.g. Dole et al. 2004; Rieke et al. 2004). It also lies in a region of the spectrum that is more complicated to model, with contributions from active galactic nuclei (AGN), thermal dust emission and non-thermal polycyclic aromatic hydrocarbons. The 3'' aperture data were used, where available, to match the *Spitzer* Infrared Array Camera data.

The latest SPIRE maps for the COSMOS field, Data Release 4 from HerMES (Oliver et al. 2012), from the *Herschel* Database in Marseille² were used for the SPIRE extraction. These maps have a beam size of 18.1'', 25.2'', and 36.6'' and 5σ confusion limits of 24.0, 27.5, and 30.5 mJy for the 250 μm , 350 μm , and 500 μm bands, respectively (Griffin et al. 2010; Nguyen et al. 2010).

The ALMA archive³ was searched for objects within the COSMOS field with positive flux densities at 870 μm or 1250 μm . A total of 214 objects were found that met these criteria: 192 objects with 870 μm data at $z < 3$ from Scoville et al. (2014, 2016) and 22 objects with 1250 μm data at $z > 3$ from Scoville et al. (2016). In the original studies, these objects were selected as they have stellar masses of approximately $10^{11} M_{\odot}$. Of the 214 objects, 43 (20%) have a signal-to-noise ratio of less than 2σ , all of which are from Scoville et al. (2014), due to the higher values of 1σ rms noise. The reader is referred to Scoville et al. (2014) and Scoville et al. (2016) for discussions on the signal-to-noise of the ALMA observations.

The ALMA data were matched to the COSMOS2015 sources. With a positional accuracy of 0.15'' for the COSMOS2015 catalogue (Laigle et al. 2016) and a pointing accuracy of 0.6'' for ALMA, matching within 1'' of the ALMA sources was deemed adequate. For one ALMA object there was more than one match to the COSMOS2015 catalogue, so the closest object was used. Once matched, the photometric redshifts from COSMOS2015 were used.

¹The *Subaru* narrow bands are used in the photometric redshift derivation in the COSMOS2015 catalogue.

²<http://hedam.lam.fr>

³<https://almascience.nrao.edu/alma-data/archive>

2.3 Methodology

2.3.1 Tools

XID+

XID+⁴ (Hurley et al. 2017) is a probabilistic de-blending tool created to extract source flux densities from photometry maps that are affected by source confusion. This is achieved by using Bayesian inference to explore the posterior distribution function. Once converged, the flux density is reported along with the upper and lower 1σ uncertainties. In the original version, XID+ uses a flat prior in parameter space, between zero and the brightest value in the map, along with the source positions on the sky. This work introduces a more informed Gaussian prior, again truncated between zero and the brightest value in the map. The mean and sigma for these Gaussian priors are generated by using CIGALE spectral energy distribution (SED) models to estimate the flux densities for the mean and using twice the error on these estimates as the sigma, to be conservative (see also Sect. 2.3.1 and 2.3.2).

To allow parallelisation, which reduces the time XID+ takes to de-blend the map, the map is split into tiles based on the Hierarchical Equal Area isoLatitude Pixelization of a sphere (HEALPix; Górski et al. 2005) system using order 11, which corresponds to an area of 2.95 arcmin^2 per tile. Order 11 was chosen as it is a compromise between the number of objects in a tile – more objects means a more reliable flux density extraction – and the time it takes a tile to run; here a tile was required to have a run time of less than one week. XID+ was run using four cores per tile on the 162 node Peregrine high performance computing cluster (HPC) at the University of Groningen.

CIGALE

Code Investigating GALaxy Emission⁵ (CIGALE; Noll et al. 2009) is a SED modelling and fitting tool with an improved fitting procedure by Serra et al. (2011). Here, the Python version 0.9.0 is used (Boquien et al., in prep; Burgarella et al., in prep) to generate SEDs and fit them to the data from COSMOS2015 to estimate the SPIRE $250 \mu\text{m}$, $350 \mu\text{m}$, and $500 \mu\text{m}$ flux densities. CIGALE was run using one node of the Peregrine HPC.

CIGALE models are based around three main components: stars, dust, and AGN. The SEDs generated by CIGALE used our choice of a double exponentially declining star formation history (SFH), Bruzual & Charlot (2003) stellar emission, Chabrier (2003) initial mass function, Calzetti et al. (2000) dust attenuation, the Draine et al. (2014) update of the Draine & Li (2007) IR dust emission, and Fritz et al. (2006) AGN models. A list of the parameters used, where they differ from the default values, can be found in Appendix 2.6.2, along with a brief justification.

2.3.2 Extracting the SPIRE flux densities

CIGALE and XID+ were used to extract the flux densities in the SPIRE bands. The ALMA data were not used in this process. The pipeline begins by using the multi-

⁴https://github.com/H-E-L-P/XID_plus

⁵<http://cigale.lam.fr/>

wavelength data from COSMOS2015 to generate estimates for the SPIRE flux densities and uncertainties using CIGALE. Approximately 1.3% of the objects in the COSMOS2015 catalogue have no known redshift. These objects were removed from the CIGALE run and assigned an arbitrary flux density of 7 mJy, along with an artificially large error to produce a functionally flat prior. A further 2.9% of the COSMOS2015 catalogue objects are classified as non-galaxies. These objects can be run through CIGALE, but care must be taken; they will be assigned flux density estimates with the assumption that they have the SED of a galaxy. To compensate for this, the errors were artificially inflated to create an almost flat prior. The objects with no redshift or those classified as non-galaxies cannot simply be discarded as they will likely appear in the SPIRE images. For the remaining objects in the COSMOS2015 catalogue with redshifts and classified as galaxies, the errors were multiplied by two to prevent overconstraint in XID+ (see also Sec. 2.4.1).

The SPIRE estimates from CIGALE, along with the added data for objects without redshifts, were used as the means in the priors for XID+, while the expanded errors were used as the standard deviations. XID+ was then run on the SPIRE images. As the priors for the objects that are not galaxies, or have no redshift, are effectively flat, the flux for these objects is free to change such that it will not interfere with the fitting of the more constrained sources.

For the flat prior flux densities, the tiles of interest were run through XID+ with a flat flux density prior. These flat prior flux densities were compared to the informed prior flux densities.

2.4 Results

2.4.1 CIGALE and XID+ performance

To check whether the CIGALE predictions for the SPIRE sources are reasonable, the flux densities for the 100 brightest sources at $250\ \mu\text{m}$ from the COSMOS2015 catalogue, which also have detections at 350 and $500\ \mu\text{m}$, were compared to the predicted flux densities from CIGALE. Blind extraction has been shown to overestimate $250\ \mu\text{m}$ flux densities by up to 150%, even for bright sources above the 5σ confusion limit (Scudder et al. 2016). Thus, we expect the flux densities from CIGALE to be systematically lower than those in COSMOS2015. We note, however, that the COSMOS2015 catalogue has SPIRE data extracted using the previous generation de-blending tool DESPHOT, not blind extraction, so the overestimation will not be the same, but a comparison is still valid (Scudder 2017 priv. comm.). It is also possible for the CIGALE predictions to underpredict for some sources, but this is accounted for in our increased error (see below).

Figure 2.1 shows that the CIGALE SPIRE predictions are indeed lower than the COSMOS2015 flux densities: the blue data points for the $250\ \mu\text{m}$ sources fall below the magenta one-to-one line. The cyan line is the locus where the CIGALE flux density is 40% the COSMOS2015 flux density: the fraction that is expected if the catalogue results are overestimated by 150%. This line splits the $250\ \mu\text{m}$ data approximately in half, with 52 objects below the line and 48 above, which would be expected if the catalogue results are 150% too high. However, there is a large scatter. The $350\ \mu\text{m}$ and $500\ \mu\text{m}$ bands are less well split, with 59 and 68 objects below the cyan line, respectively. With the

Figure 2.1: Scatter plot of the predicted CIGALE flux densities against the flux densities from the COSMOS2015 catalogue for the 250 μm (blue points), 350 μm (green points), and 500 μm (red points). The errors for the catalogue flux are instrumental noise plus confusion noise. The one-to-one line (magenta) and $y = 0.4x$ (cyan) illustrate that the data better fits the idea that bright sources in catalogues are overestimated.

lower resolution of the 350 μm and 500 μm bands, it is likely that multiplicity will have a greater impact, so it is not surprising that the longer wavelengths have more objects with predicted fluxes below the cyan line.

We can also calculate how many of the 100 bright objects have CIGALE flux densities consistent with 40% of the catalogue flux densities within CIGALE errors. This is 39, 44, and 45 for the 250 μm , 350 μm , and 500 μm bands, respectively. If the error is doubled the number of objects increases to 65, 70, and 64 for the 250 μm , 350 μm , and 500 μm bands and is consistent with what is expected for a 1σ uncertainty. Therefore, we decided to expand the errors from CIGALE by a factor of two.

To check that the results from XID+ with an informed prior are not just the prior itself, Fig. 2.2 shows a plot of the extracted 250 μm flux densities using the informed Gaussian against the 250 μm flux density priors from CIGALE for both the ALMA selected objects (dark blue) as well as the full probability density function derived from the posterior samples (fPDF) of all 63 701 objects in the same tiles as the ALMA objects, from high (light red) to low (dark red). If XID+ was simply returning the prior, all the points would lie close to or along the red one-to-one line. This is evidently not the case for the ALMA objects, with extracted flux densities up to 1.5 dex away from their flux density priors, or for the fPDF. The 350 μm and 500 μm results are similar.

2.4.2 Comparison of ALMA data to CIGALE SEDs

The SPIRE data constrain the peak of the cold dust emission in a galaxy's SED and, if extracted correctly, any data in the long wavelength tail should lie on the SED. To test this,

Figure 2.2: Extracted 250 μm flux densities using the informed Gaussian (F_{250}^{Info}) against the 250 μm flux density priors from CIGALE (F_{250}^{Pri}) for the ALMA selected sources (blue). The fPDF of all objects extracted from the same tiles as the ALMA sources are also shown and presented from high (light red) to low (dark red) density. The red line is the one-to-one line.

Table 2.1: Absolute residuals ($|\gamma|$) between the measured ALMA flux densities and the fluxes of the best fitting SEDs at the ALMA wavelengths expressed as a multiple of the error on the ALMA flux density. The values in Cols. 2 and 3 are the number (percentage) of objects with a $|\gamma|$ less than the value in Col. 1.

$ \gamma $	Flat Prior	Informed Prior
1	77 (36.2%)	98 (46.0%)
2	121 (56.8%)	142 (66.7%)
3	146 (68.5%)	166 (77.9%)
4	160 (75.1%)	184 (86.4%)
5	177 (83.1%)	192 (90.1%)

the SPIRE data from XID+, along with the data from the COSMOS2015 catalogue, were run through CIGALE to create the best fitting SED for all of the data. One object failed to converge, resulting in no output from CIGALE, and so had to be omitted. ALMA data were then compared to the best fitting SEDs.

For the 213 objects that successfully ran through CIGALE, the residuals between the measured ALMA flux densities and the ALMA flux densities inferred from the best fitting SEDs were calculated. These residuals (γ) are presented in Table 2.1 and Fig. 2.3 as multiples of the errors on the ALMA flux densities (σ), i.e. $\gamma = \text{residual} / \sigma$. A positive γ means that the measured ALMA flux density is greater than the inferred flux density, and the closer the value of γ is to zero, the better XID+ had performed.

Figure 2.3 demonstrates the improvement that the informed prior gives over the flat prior. The informed Gaussian prior has a reduced spread around zero, a standard deviation of 7.95σ compared to 12.21σ for the flat prior, indicating that the informed prior is providing a better fit to the ALMA data than the flat prior. A reduced skew for the

Figure 2.3: Distribution of the residuals between the measured ALMA flux densities and the flux densities inferred from the SEDs at the ALMA wavelengths expressed as a multiple of the error on the ALMA flux density (σ) for the flat prior (red) and the informed Gaussian prior (blue). The informed prior reduces the mean from 3.44σ to 1.83σ , the standard deviation from 12.21σ to 7.95σ , and the skew from 11.50 to 7.97 .

informed prior with a skew of 7.97 with respect to the flat prior with a skew of 11.50 , and a reduction in the mean of the distribution from 3.44σ for the flat prior to 1.83σ for the informed prior also demonstrates improved performance. With the reduced skew and lower mean, XID+ does not appear to underpredict the SPIRE flux densities as often with the informed prior. Table 2.1 shows the number of objects that have an absolute value of γ ($|\gamma|$) below a certain threshold. The number of objects below each $|\gamma|$ value is greater for the informed prior than it is for the flat prior, demonstrating that there is better agreement between the inferred ALMA flux densities of the best fitting SEDs and the measured data when using the informed prior than when using the flat prior. There is an increase of 27.3% in the number of sources at $|\gamma| < 1$ when using the informed prior over the flat prior. This reduces to 8.5% at $|\gamma| < 5$ and gives an average increase of 16.4% in the number of sources across all five ranges. The maximum $|\gamma|$ values follow the same trend with the flat prior giving a maximum $|\gamma|$ of 167.304 compared to 95.180 for the informed prior. The minimum $|\gamma|$ for the flat prior is that same as that of the informed prior.

The ALMA objects can be split into three groups: Group A, where the informed Gaussian prior provides a $|\gamma|$ value more than 5% smaller than the flat prior $|\gamma|$; Group B, where the informed prior $|\gamma|$ is within 5% of the flat prior $|\gamma|$; and Group C, where the informed $|\gamma|$ is more than 5% larger than the flat $|\gamma|$. Figure 2.4a provides an example of one of the 99 Group A objects. It can clearly be seen that the SED for the informed prior (blue line) is in much better agreement with the measured ALMA flux density (green point) than the SED for the flat prior (red line). There are 70 objects in Group B while Group C contains the remaining 44 objects, examples of which can be found in Fig.

2.4b and 2.4c, respectively. With Group A being the largest group, it is evident that the informed prior produces more accurate results than the flat prior. If the objects with less than 2 ALMA signal-to-noise (see Sect. 2.2) are removed, the number of objects in Groups A, B, and C are 89, 51, and 35, respectively. Thus, the removal of these objects does not change the conclusion that, on average, the informed prior produces more accurate results.

To see how the two XID+ priors compare with each other, Fig. 2.5 shows the ratio between the flux density extracted using the informed prior (F_{250}^{Info}) and flat prior (F_{250}^{Flat}) against F_{250}^{Info} . The ALMA sources (in dark blue) show that as F_{250}^{Info} increases for the ALMA objects, the ratio $F_{250}^{\text{Info}}/F_{250}^{\text{Flat}}$ increases, becoming unity at $F_{250}^{\text{Info}} \approx 3$ mJy, with smaller ratios below this flux density and higher ratios above. This implies that the flat prior overestimates flux densities for objects fainter than 3 mJy and underestimates flux densities for objects brighter than 3 mJy, with respect to the informed prior. This trend also appears for the fPDF, shown from high (light red) to low (dark red). There is an indication that the relation begins to flatten for high flux densities, which is expected as both priors should perform equally well at higher flux densities. However, $F_{250}^{\text{Info}}/F_{250}^{\text{Flat}}$ being above unity is surprising. This may be the high source density affecting the flat prior results: the flat prior may be “too democratic” as it assigns flux density to each object and so assigns too much to the faint objects and too little to the bright objects.

Figure 2.6 illustrates how the difference between the flat and informed Gaussian XID+ priors affects the IR luminosity (L_{IR}). The L_{IR} was calculated by integrating the best fitting SED between 3 and 1100 μm at rest. For the low ($z < 1$) and high ($z > 3$) redshift objects, the change in prior has little effect on L_{IR} . However, for the intermediate redshift ($1 < z < 3$) objects, there appears to be an increase in L_{IR} , although the number of objects in the sample is too small to draw any robust conclusions. As can be seen in the right panel of Fig. 2.6, this results in a smoother distribution of luminosities across the sample presented. For comparison, the green line in Fig. 2.6 shows the 5σ confusion limit for the 250 μm band, calculated by scaling the Wuyts et al. (2008) SED template to the 250 μm 5σ confusion limit of 24.0 mJy and integrating between 3 and 1100 μm for a range of redshifts. Twenty of the objects with an informed prior, and one with a flat prior, do not fall below this confusion limit.

2.4.3 Model dependance

To check whether the difference between the informed prior and flat prior results is caused by our CIGALE model, we also fitted the data from the COSMOS2015 catalogue and the extracted SPIRE data for the 213 ALMA sources with MAGPHYS⁶ (da Cunha et al. 2008), and we fitted modified black-body (MBB) functions to the SPIRE data from the 213 sources. MAGPHYS is a SED modelling and fitting tool that uses different models from CIGALE, while MBBs are commonly used to model thermal dust emission (e.g. Viero et al. 2013; Casey et al. 2014; Wang et al. 2016). Here we use a MBB of the form

$$S_{\nu} \propto \nu^{\beta} \times B_{\nu}(T), \quad (2.1)$$

where $B_{\nu}(T)$ is the black-body function for dust at temperature T . The emissivity β was fixed at 1.5, and T was allowed to vary between 10K and 150K. With these results, we

⁶<http://www.iap.fr/magphys/magphys/MAGPHYS.html>

Figure 2.4: Example plots showing the best fitting SEDs (curves) for the extracted SPIRE flux densities (red and blue points) using the flat prior (red) and informed Gaussian prior (blue). The other multi-wavelength data from the COSMOS2015 catalogue (purple) are overplotted. Panel (a) is an example of Group A, illustrating how the informed prior can increase the agreement between the best fitting SED and the measured ALMA flux density (green). Panel (b) is for Group B, illustrating how the informed and flat priors can give equal agreement between the best fitting SED and the measured ALMA flux density. Panel (c) is for Group C, showing how the flat prior can occasionally give better agreement.

Figure 2.5: Plot of the ratio of the 250 μm flux densities extracted using the informed Gaussian prior (F_{250}^{Info}) to those extracted with the flat prior (F_{250}^{Flat}) against (F_{250}^{Info}). The ALMA sources are in dark blue, while the statistical average number density is from light red (high) to dark red (low).

Figure 2.6: Scatter plot of the infrared (IR, 3 μm - 1100 μm) luminosity vs. redshift for the results from XID+ with the flat prior (red points/bars) and the informed Gaussian prior (blue points/bars) with a histogram of the IR luminosities. The green line corresponds to the 250 μm 5 σ confusion limit.

Table 2.2: Absolute residuals ($|\gamma|$) between the measured ALMA flux densities and the fluxes of the best fitting MBB templates and MAGPHYS SEDs at the ALMA wavelengths expressed as a multiple of the error on the ALMA flux density. The values in Cols. 3 and 4 are the number (percentage) of objects with a $|\gamma|$ less than the value in Col. 2.

Model	$ \gamma $	Flat Prior	Informed Prior
MBB	1	62 (29.1%)	70 (32.9%)
	2	106 (49.8%)	120 (56.3%)
	3	134 (62.9%)	155 (72.8%)
	4	152 (71.4%)	171 (80.3%)
	5	168 (78.9%)	181 (85.0%)
MAGPHYS	1	72 (33.8%)	79 (37.1%)
	2	119 (55.9%)	132 (62.0%)
	3	138 (64.8%)	160 (75.1%)
	4	156 (73.2%)	169 (79.3%)
	5	171 (80.3%)	177 (83.1%)

recreated Table 2.1 for the MBB and MAGPHYS results, and examined the populations of the three groups, A, B, and C.

For the MBB, qualitatively the results are the same, although the division between the flat and informed priors is less wide. The results, shown in Table 2.2, show that the number of objects within certain $|\gamma|$ thresholds is greater for the informed prior than for the flat prior, supporting the idea that the results from XID+ using the informed prior are more reliable than the results with the flat prior. There is an average increase of 12.4% in the number of sources across all five bins, which is less than the average over all five bins when compared with CIGALE. Also, looking at the distribution of objects in the three groups, we find that Group A again has the most objects (126), while Groups B and C have fewer (37 and 51 objects, respectively).

As with the results from the MBB, the results from MAGPHYS also have the same qualitative results as those using CIGALE. For MAGPHYS, the data from COSMOS2015 along with the extracted flux densities for the SPIRE bands were used, and SEDs were fitted to these data. Similar to CIGALE, the flux densities at $870\ \mu\text{m}$ and $1250\ \mu\text{m}$ were extracted from the best fitting SEDs and compared to the ALMA data. The results are presented in Table 2.2 and again show a greater number of objects within the values of $|\gamma|$ examined for the informed prior with respect to the flat prior, with an average increase of 9.7% in the number of sources across all five $|\gamma|$ ranges. As with the CIGALE and MBB results, Group A again contains the greatest number of objects (108), followed by Group C (65), and Group B (41). As both the MAGPHYS and the MBB results are consistent with the CIGALE results, we believe that there is little model dependence in our conclusions.

2.5 Conclusions

XID+ is one of the most advanced de-blending tools available. Here, we extend the flux density prior from a simple flat prior (Hurley et al. 2017) and explore the idea of using

supplementary data to provide a more physically motivated flux density prior. Using a rich set of multi-wavelength data, an informed Gaussian prior was introduced to XID+ and applied to 214 objects with ALMA detections in the COSMOS field. Using this method, it was shown that the agreement between the measured ALMA flux densities and the inferred flux densities (calculated with CIGALE) at the ALMA wavelengths improved by an average of 16.4% in the number of objects in each of the five $|\gamma|$ ranges examined in Sect. 2.4. If all three models used for the ALMA comparison are considered, there is an average increase of 12.8% demonstrating that, qualitatively, the results are not model dependent. However, the exact quantitative improvement seen is model dependent. This demonstrates that utilising supplementary data to provide a more physically motivated prior results in extracted flux density values that have improved agreement with data at other wavelengths.

Acknowledgements. We would like to thank the anonymous referee whose comments greatly improved this paper.

We would like to thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high performance computing cluster.

This research has made use of data from HerMES project (<http://hermes.sussex.ac.uk/>). HerMES is a Herschel Key Programme utilising Guaranteed Time from the SPIRE instrument team, ESAC scientists, and a mission scientist.

The HerMES data were accessed through the *Herschel* Database in Marseille (HeDaM - <http://hedam.lam.fr>) operated by CeSAM and hosted by the Laboratoire d’Astrophysique de Marseille.

HerMES DR4 was made possible through support of the Herschel Extragalactic Legacy Project, HELP (<http://herschel.sussex.ac.uk>).

The research leading to these results has received funding from the European Union Seventh Framework Programme FP7/2007-2013/ under grant agreement no. 607254. This publication reflects only the author’s view and the European Union is not responsible for any use that may be made of the information contained therein.

Bibliography

- Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
- Burgarella, D., Buat, V., Gruppioni, C., et al. 2013, A&A, 554, A70
- Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
- Carpenter, B., Gelman, A., Hoffman, M., et al. 2017, Journal of Statistical Software, 76, 1
- Casey, C. M., Narayanan, D., & Cooray, A. 2014, Phys. Rep., 541, 45
- Chabrier, G. 2003, PASP, 115, 763
- Ciesla, L., Charmandaris, V., Georgakakis, A., et al. 2015, A&A, 576, A10
- da Cunha, E., Charlot, S., & Elbaz, D. 2008, MNRAS, 388, 1595
- Dole, H., Rieke, G. H., Lagache, G., et al. 2004, ApJS, 154, 93
- Dole, H., Lagache, G., Puget, J.-L., et al. 2006, A&A, 451, 417
- Draine, B. T. 2003, ARA&A, 41, 241
- Draine, B. T., & Li, A. 2007, ApJ, 657, 810
- Draine, B. T., Aniano, G., Krause, O., et al. 2014, ApJ, 780, 172
- Fritz, J., Franceschini, A., & Hatziminaoglou, E. 2006, MNRAS, 366, 767
- Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
- Griffin, M. J., Abergel, A., Abreu, A., et al. 2010, A&A, 518, L3
- Hurley, P. D., Oliver, S., Betancourt, M., et al. 2017, MNRAS, 464, 885
- Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18
- Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, ApJS, 224, 24
- Larson, D., Dunkley, J., Hinshaw, G., et al. 2011, ApJS, 192, 16
- Merlin, E., Fontana, A., Ferguson, H. C., et al. 2015, A&A, 582, A15

- Mitchell, P. D., Lacey, C. G., Baugh, C. M., & Cole, S. 2013, *MNRAS*, 435, 87
- Nguyen, H. T., Schulz, B., Levenson, L., et al. 2010, *A&A*, 518, L5
- Noll, S., Burgarella, D., Giovannoli, E., et al. 2009, *A&A*, 507, 1793
- Oliver, S. J., Bock, J., Altieri, B., et al. 2012, *MNRAS*, 424, 1614
- Pilbratt, G. L., Riedinger, J. R., Passvogel, T., et al. 2010, *A&A*, 518, L1
- Puglisi, A., Rodighiero, G., Franceschini, A., et al. 2016, *A&A*, 586, A83
- Rieke, G. H., Young, E. T., Engelbracht, C. W., et al. 2004, *ApJS*, 154, 25
- Roseboom, I. G., Oliver, S. J., Kunz, M., et al. 2010, *MNRAS*, 409, 48
- Roseboom, I. G., Ivison, R. J., Greve, T. R., et al. 2012, *MNRAS*, 419, 2758
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, 172, 1
- Scoville, N., Aussel, H., Sheth, K., et al. 2014, *ApJ*, 783, 84
- Scoville, N., Sheth, K., Aussel, H., et al. 2016, *ApJ*, 820, 83
- Scudder, J. M., Oliver, S., Hurley, P. D., et al. 2016, *MNRAS*, 460, 1119
- Serra, P., Amblard, A., Temi, P., et al. 2011, *ApJ*, 740, 22
- Viero, M. P., Wang, L., Zemcov, M., et al. 2013, *ApJ*, 772, 77
- Wang, L., Viero, M., Clarke, C., et al. 2014, *MNRAS*, 444, 2870
- Wang, L., Norberg, P., Gunawardhana, M. L. P., et al. 2016, *MNRAS*, 461, 1898
- Wuyts, S., Labbé, I., Förster Schreiber, N. M., et al. 2008, *ApJ*, 682, 985

2.6 Appendix

2.6.1 Homogeneity of COSMOS

Figure 2.7a shows the noise map for the 2 square degree 250 μm SPIRE band in the COSMOS field with the area used for this study outlined in red and the position of the ALMA sources shown with blue points. As can be seen, the noise level within this area is approximately flat. For comparison, the noise map for the 250 μm SPIRE band in the 11 square degree XMM-LSS field is shown in Fig. 2.7b, which is much less homogenous. The noise maps for the 350 μm and 500 μm bands are very similar.

Figure 2.7: The noise map for the SPIRE 250 μm band in the (a) COSMOS field (2 deg^2) and (b) XMM-LSS field (11 deg^2) from the HerMES DR4 (Oliver et al. 2012) and downloaded from HeDAM. For the COSMOS field, the outline of the study area is shown in red, while the ALMA sources are shown in blue.

2.6.2 CIGALE Parameters

In this work, a double exponential SFH was used over the more commonly used exponentially declining or delayed exponential SFH as these two SFHs did not appear to reproduce the expected starburst population in the star formation rate vs. stellar mass plane. The e-folding time of the two stellar populations (old and young) in the SFH was roughly matched to that of Mitchell et al. (2013). As Mitchell et al. (2013) used a single declining exponential SFH, the e-folding times were split with the burst population taking values of 9 Gyr and above and the main population taking values of less than 9 Gyr. For the ages of the main population, the values were sampled between 1 and 13 Gyr with a log distribution. The mass fraction of the burst population follows Ciesla et al. (2015) along with the age of the young stellar population, which also had a lower age of 0.001 Gyr added. The Bruzual & Charlot (2003) stellar population model was used with a Chabrier (2003) initial mass function. As this study was not designed to explore the metallicity of galaxies, it was decided to leave the metallicity at solar.

The dust colour excess ($E(B-V)^*$) is often measured to be $\lesssim 0.6$, with a higher number of objects with lower excesses (e.g. Calzetti et al. 2000), so this region has been sampled for the SED models. However, it is possible that the $E(B-V)^*$ could be larger than this, so values of 1.1 and 2.0 are also included. For the $E(B-V)_{\text{old}}^*$ reduction factor, it has been shown that a value of 0.58 is needed for the local universe (Puglisi et al. 2016) instead of the Calzetti et al. (2000) value of 0.44. A compromise of 0.5 was therefore used.

For the dust emission, the polycyclic aromatic hydrocarbon (PAH) fraction had an increase in range around the default 2.5, so more fractions could be sampled while keeping the number of models reasonable. The minimum scaling factor of the radiation field was similarly given a range to sample with an increase in the smallest value from 1.0 to 5.0. The illuminated fraction was reduced to 0.02, following Ciesla et al. (2015).

The parameters in the AGN model were matched to those used by Ciesla et al. (2015), who undertook a detailed study of AGN host galaxy emission using CIGALE. The number of choices of frac_{AGN} was reduced from 14 to 9, while still covering the same range, to reduce the number of models created by CIGALE to decrease runtime.

A list of parameters, where they differ from default, can be found in Table 2.3.

Table 2.3: Parameters used for the various properties in the CIGALE model SEDs where they differ from the default values. All ages and times are in Gyr.

Parameter	Value	Star Formation History	Description
τ_{main}	1.0, 3.0, 5.0		e-folding time (main)
τ_{burst}	9.0, 13.0		e-folding time (burst)
f_{burst}	0.001, 0.01, 0.1, 0.2		Burst mass fraction
age	1.000, 1.329, 1.768, 2.351, 3.126, 4.157, 5.528, 7.352, 9.776, 13.000		Population age (main)
burst age	0.001, 0.010, 0.030, 0.100, 0.300		Population age (burst)
Stellar Emission			
Z	0.02		Metallicity (0.02 is solar)
Dust Attenuation			
$E(B - V)^*_{young}$	0.1, 0.18, 0.33, 0.6, 1.1, 2.0		$E(B - V)^*$ for young population
$E(B - V)^*_{old}$	0.5		Reduction factor in $E(B - V)^*$ for old population
Dust Emission			
q_{PAH}	1.12, 2.50, 3.19		Mass fraction of PAH
U_{min}	5.0, 10.0, 25.0		Minimum scaling factor of the radiation field intensity
γ	0.02		Illuminated fraction
AGN Emission			

τ	1.0, 6.0	Optical depth at 9.7 μm
γ	0.0	γ coefficient for the gas density function of the torus ^a
ψ	0.001 $^\circ$, 89.990 $^\circ$	Angle between equatorial axis and line of sight
$frac_{AGN}$	0.0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5 0.6, 0.7	AGN fraction

Notes. ^(a) Density function of the torus can be found in Fritz et al. (2006) as Equation 3.

Main sequence of star forming galaxies beyond the *Herschel* confusion limit

W. J. Pearson, L. Wang, P. D. Hurley, K. Mafek, V. Buat, D. Burgarella, D. Farrah,
S. J. Oliver, D. J. B. Smith, F. F. S. van der Tak
Astronomy & Astrophysics, 615, A146, 2018
Reproduced with permission, ©ESO

Abstract

Context: Deep far-infrared (FIR) cosmological surveys are known to be affected by source confusion, causing issues when examining the main sequence (MS) of star forming galaxies. In the past this has typically been partially tackled by the use of stacking. However, stacking only provides the average properties of the objects in the stack.

Aims: This work aims to trace the MS over $0.2 \leq z < 6.0$ using the latest de-blended *Herschel* photometry, which reaches ≈ 10 times deeper than the 5σ confusion limit in SPIRE. This provides more reliable star formation rates (SFRs), especially for the fainter galaxies, and hence a more reliable MS.

Methods: We built a pipeline that uses the spectral energy distribution (SED) modelling and fitting tool CIGALE to generate flux density priors in the *Herschel* SPIRE bands. These priors were then fed into the de-blending tool XID+ to extract flux densities from the SPIRE maps. In the final step, multi-wavelength data were combined with the extracted SPIRE flux densities to constrain SEDs and provide stellar mass ($M_\star$) and SFRs. These $M_\star$ and SFRs were then used to populate the SFR- $M_\star$ plane over $0.2 \leq z < 6.0$.

Results: No significant evidence of a high-mass turn-over was found; the best fit is thus a simple two-parameter power law of the form $\log(\text{SFR}) = \alpha[\log(M_\star) - 10.5] + \beta$. The normalisation of the power law increases with redshift, rapidly at $z \lesssim 1.8$, from 0.58 ± 0.09 at $z \approx 0.37$ to 1.31 ± 0.08 at $z \approx 1.8$. The slope is also found to increase with redshift, perhaps with an excess around $1.8 \leq z < 2.9$.

Conclusions: The increasing slope indicates that galaxies become more self-similar as redshift increases. This implies that the specific SFR of high-mass galaxies increases with redshift, from 0.2 to 6.0, becoming closer to that of low-mass galaxies. The excess in the slope at $1.8 \leq z < 2.9$, if present, coincides with the peak of the cosmic star formation history.

3.1 Introduction

It has been observed that there is a strong correlation between the stellar mass ($M_\star$) and the star formation rate (SFR) for the majority of star forming galaxies (SFG) (e.g. Brinchmann et al. 2004; Noeske et al. 2007; Elbaz et al. 2007; Speagle et al. 2014; Tomczak et al. 2016). This has become known as the main sequence (MS) of star forming galaxies.

The MS is notable due to its tight correlation; the scatter of the SFR- $M_\star$ relation has been found to be approximately 0.3 dex. This low scatter has been observed to exist for approximately the last 10 Gyr and has been found to be independent of $M_\star$ (Whitaker et al. 2012; Speagle et al. 2014; Tomczak et al. 2016). The consistency of the scatter is understood to be a result of every galaxy having star formation (SF) regulated by similar quasi-static processes (Lee et al. 2015), while the scatter arises as a result of minor fluctuations of the flow of material into galaxies (Tacchella et al. 2016; Mitra et al. 2017). Galaxies move to the upper MS as the gas is compacted in the centre of the galaxy which triggers SF. As the central gas is depleted, but before the galaxy moves above the MS, the SFR reduces and the galaxy falls to the lower MS. New gas then falls into the galaxy, replenishing the central gas reservoir before the galaxy becomes quiescent. This cycle repeats until the galaxy's gas replenishment time is longer than the depletion time and the galaxy becomes quiescent (Tacchella et al. 2016).

Two different schools of thought exist over the shape the MS takes. Some studies find the relation between SFR and stellar mass is a simple power law (e.g. Speagle et al. 2014), i.e. the log of the SFR increases with the log of the stellar mass as

$$\log(\text{SFR}/M_\odot\text{yr}^{-1}) = \alpha \log(M_\star/M_\odot) + \beta, \quad (3.1)$$

where α is the slope and β is the normalisation. However, other studies find there is a high-mass turn-over at $\log(M_\star/M_\odot) \approx 10.5$ with the slope of the MS being shallower above the turn-over than below (e.g. Lee et al. 2015; Tomczak et al. 2016). Recent studies have shown the two different forms of the MS may be a result of how the SFG population is separated from the quiescent galaxies (QG). Johnston et al. (2015) have shown that using a SFG-QG cut that is stricter in selecting the SFG population, in this case a 4000Å break index value that is lower, will result in a straight MS, while a cut that is less strict forms a MS with a turn-over. The less strict cut leaves in high-mass, low SFR objects, which lower the mean SFR at high mass.

Regardless of the form of the MS, the normalisation is found to increase with redshift (e.g. Speagle et al. 2014; Johnston et al. 2015; Schreiber et al. 2015; Tomczak et al. 2016). This increasing normalisation is not surprising. As redshift increases, the fraction of cold gas available for star formation increases (Tacconi et al. 2010; Dunne et al. 2011; Genzel et al. 2015; Scoville et al. 2016). Thus, with more gas available, more stars can form, raising the average SFR in all galaxies. This could be counterbalanced by a lower SFR per gas mass ($\text{SFR}/M_{\text{gas}}$) at higher redshift but existing works have shown that $\text{SFR}/M_{\text{gas}}$ is either relatively constant or rises with redshift (Tacconi et al. 2010; Scoville et al. 2016).

It has been argued that the slope of the MS should be unity for all mass ranges once the SFR has become stable in a galaxy (Pan et al. 2017). However, the slope of the MS is typically found to be lower than unity, with values between 0.4 and 1.0 (e.g. Whitaker et al. 2012; Speagle et al. 2014; Tomczak et al. 2016). This discrepancy is believed to be

the result of a combination of quenching and a reduction in the relative size of the cold gas reservoir as the M_* of a galaxy increases (Pan et al. 2017).

Far-infrared (FIR) emission is a key component for accurately determining the SFR of an object. Part of the ultraviolet (UV) emission from young stars heats dust in a galaxy, which then re-radiates in the infrared (IR). Observing in the UV would provide a direct measure of SFR, but could underestimate the total SFR as this absorption by dust obscures some of the UV emission (e.g. Meurer et al. 1999; Dale et al. 2009; Bourne et al. 2017; Dunlop et al. 2017). As a result, observing IR emission in combination with UV emission is key to providing a complete picture of the SFR.

However, FIR surveys, such as those conducted with the ESA *Herschel Space Observatory* (Pilbratt et al. 2010) Spectral and Photometric Imaging Receiver (SPIRE; Griffin et al. 2010), have relatively poor resolution with respect to optical, and also have source confusion (e.g. Nguyen et al. 2010; Oliver et al. 2012; Hurley et al. 2017). Previously, stacking was used to recover the flux densities of fainter sources (e.g. Pannella et al. 2015; Schreiber et al. 2015; Harris et al. 2016; Álvarez-Márquez et al. 2016). However, by its very nature, stacking only provides the average properties (e.g. mean, median) of the objects in the stack¹. As a result, the properties of individual objects cannot be determined, which results in the loss of information about the wider properties of a galaxy population.

To overcome confusion, it is necessary to de-blend the maps to generate individual flux density measurements for both faint and bright sources. For SPIRE, this was done with the De-blended SPIRE Photometry algorithm (DESPHOT; Roseboom et al. 2010, 2012; Wang et al. 2014), amongst other techniques (e.g. Laidler et al. 2007; Béthermin et al. 2010b; Kurczynski & Gawiser 2010; Viero et al. 2013; Merlin et al. 2015; Sarfzadeh et al. 2015; Wright et al. 2016). DESPHOT misassigns flux densities when more than one source is within a beam, and is also unable to realistically derive flux density errors of a given source (see discussion in Hurley et al. 2017). To improve source de-blending, XID+ (Hurley et al. 2017, see also Sect. 3.3.1) has been developed and subsequently expanded to improve flux density estimation by including more precise flux density priors (Pearson et al. 2017).

With XID+, it has been shown that more than 95% of blindly detected SPIRE 250 μm sources with a flux density greater than 30 mJy contain more than one object which contribute more than 10% of the source's total flux density. At least 70% of the flux density from these sources is assigned to the two brightest objects (Scudder et al. 2016). This suggests that many current SPIRE catalogues have too much flux density assigned to their objects, and any derived physical parameter that relies on the SPIRE emission, such as the total infrared luminosity or star formation rate (SFR), will be overestimated.

In this work, XID+ is used to de-blend the maps in the SPIRE bands, resulting in a catalogue of over 200 000 objects with SPIRE flux density measurements in the COSMOS field, compared to approximately 32 000 sources blindly detected at 250 μm (Roseboom et al. 2010; Wang et al. 2014; Sparre et al. 2015). These data, along with UV to *Herschel* Photodetector Array Camera and Spectrometer (PACS; Poglitsch et al. 2010) data from a multi-wavelength catalogue, are then used to examine the MS.

The paper is structured as follows. Section 3.2 describes where the data were collected. Section 3.3 explains the methodology and tools used to de-blend the SPIRE bands

¹ It is possible to recover the scatter of the objects in a stack. However, this requires strong assumptions on the distribution of the galaxies in the stack (Schreiber et al. 2015; Wang et al. 2016).

and find the MS. Section 3.4 explores the results of the analysis. Section 3.5 provides discussion, while Section 3.6 provides a summary of the conclusions. Where necessary, Wilkinson Microwave Anisotropy Probe year 7 (WMAP7) cosmology (Komatsu et al. 2011; Larson et al. 2011) is followed, with $\Omega_M = 0.272$, $\Omega_\Lambda = 0.728$, and $H_0 = 70.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$, to be consistent with CIGALE² (Noll et al. 2009, see also Sect. 3.3.1).

3.2 Data sets

For this work, the COSMOS field (Scoville et al. 2007) was chosen due to the wealth of multi-wavelength data available within its two square degree coverage. It also benefits from *Herschel* SPIRE data, which is considered to be homogenous and of high quality (e.g. Pearson et al. 2017).

For the multi-wavelength data, the public COSMOS2015 catalogue (Laigle et al. 2016) was used with CIGALE to generate flux density priors in the 250 μm , 350 μm , and 500 μm SPIRE bands for XID+ (see Sect. 3.3.1 and 3.3.2). COSMOS2015 contains photometric data in over 30 bands, narrow, medium, and broad, for approximately 1.2×10^6 objects in the COSMOS field. Here, we only use the bands that cover the UV to PACS 160 μm in CIGALE; a full list of the bands used can be found in Table 3.1. The aperture photometry was converted to total photometry and corrected for foreground extinction following Laigle et al. (2016). As we wanted to use the Multi-band Imaging Photometer for *Spitzer* 24 μm (MIPS24; Rieke et al. 2004) data for the $\approx 40\,000$ objects with MIPS24 data to help constrain the FIR SED, we used objects that only fall within the MIPS24 image in COSMOS. The MIPS24 coverage falls within the SPIRE coverage; the requirements are as follows (see also Fig. 3.1):

$$\begin{aligned}
 149.38^\circ &\leq RA \leq 150.86^\circ \\
 1.46^\circ &\leq Dec \leq 2.95^\circ \\
 Dec + (0.38 \times RA) &< 59.51^\circ \\
 Dec - (2.49 \times RA) &< -368.73^\circ \\
 Dec + (0.35 \times RA) &> 53.91^\circ \\
 Dec - (2.66 \times RA) &> -400.06^\circ
 \end{aligned} \tag{3.2}$$

The latest SPIRE images from the *Herschel* Database in Marseille³, the Data Release 4 maps from the *Herschel* Multi-tiered Extragalactic Survey (Oliver et al. 2012), were used in XID+ to extract the SPIRE flux densities. The 250 μm , 350 μm , and 500 μm band images have beam full widths at half maximum of 18.1'', 25.2'', and 36.6'' (Griffin et al. 2010) and 5σ confusion limits of 24.0, 27.5, and 30.5 mJy, respectively (Nguyen et al. 2010).

²<http://cigale.lam.fr/>

³<http://hedam.lam.fr>

Figure 3.1: Image of the SPIRE 250 μm COSMOS coverage (Oliver et al. 2012, 7.84 deg^2 , red) with the overlaid MIPS 24 μm COSMOS coverage (Sanders et al. 2007, 2.23 deg^2 , blue). The data used in this work were cut to match the MIPS 24 μm coverage.

Table 3.1: Telescopes and associated bands that were used for CIGALE spectral energy distribution fitting from the COSMOS2015 catalogue.

Telescope	Band(s)
GALEX	FUV, NUV
CFHT	u
Subaru	B, V, r, i+, z++, IB427, IB464, IA484, IB505, IA527, IB574, IA679, IB709, IA738, IA767, IB827
VISTA	Y, J, H, Ks
<i>Spitzer</i> (IRAC)	3.6 μm , 4.5 μm , 5.8 μm , 8.0 μm
(MIPS)	24 μm
<i>Herschel</i> (PACS)	100 μm , 160 μm

3.3 Methodology

3.3.1 Tools

CIGALE

Code Investigating GALaxy Emission (CIGALE; Noll et al. 2009) is a spectral energy distribution (SED) modelling and fitting tool with an improved fitting procedure by Serra et al. (2011). Here, the Python version 0.11.0 is used (Burgarella et al. 2005; Noll et al. 2009, Boquien et al., in prep.) to generate SEDs and to fit these SEDs to the UV to PACS data from COSMOS2015 to estimate the SPIRE 250 μm , 350 μm , and 500 μm flux densities for use as a flux density prior in XID+ (see Sects. 3.3.1 and 3.3.2). After the SPIRE band flux densities had been extracted, CIGALE was also used to calculate the physical parameters of each object, such as SFR and $M_\star$ as well as rest frame colours. CIGALE uses the energy balance between the attenuated UV emission by dust and the IR emission, allowing the estimation of the FIR flux densities. The reported values and errors for the SPIRE flux densities and physical parameters are created using CIGALE’s Bayesian probability density function analysis. CIGALE also gives the flux densities and physical parameters of the best fitting model for each object, but these are not used in this work.

The choices for the SED model components and parameters for the SPIRE band priors follow Pearson et al. (2017), with a different dust attenuation model and other minor changes, and will briefly be repeated here. We use a double exponentially declining star formation history (SFH), Bruzual & Charlot (2003) stellar emission, Chabrier (2003) initial mass function (IMF), Charlot & Fall (2000) dust attenuation, the updated Draine et al. (2014) version of the Draine & Li (2007) IR dust emission, and Fritz et al. (2006) AGN models. The dust attenuation model was changed from Calzetti et al. (2000), used in Pearson et al. (2017), to Charlot & Fall (2000) as recent work by Lo Faro et al. (2017) has shown that the Calzetti et al. (2000) dust model cannot accurately reproduce the attenuation seen in a sample of dusty galaxies at $z \approx 2$. A list of parameters used, where they differ from the default values, along with a justification can be found in Appendix 3.7.1.

XID+

XID+⁴ (Hurley et al. 2017) is a probabilistic de-blending tool used to extract source flux densities from photometry maps that suffer from source confusion. This is achieved by using Bayesian inference to explore the posterior. Once converged, the flux density is reported along with the upper and lower 1σ uncertainties. In its standard form, XID+ uses a flat prior in parameter space, between zero and the brightest value in the map, along with the source positions on the sky.

This work follows Pearson et al. (2017) by using a more informed Gaussian prior, again truncated between zero and the brightest value in the map. The mean and sigma for these Gaussian priors are generated by using CIGALE models to estimate the flux densities for the mean and using two times the error on these estimates as the sigma. To allow parallelisation, which reduces the time taken for XID+ to de-blend the map, the map is split up into tiles based on the Hierarchical Equal Area isoLatitude Pixelization

⁴https://github.com/H-E-L-P/XID_plus

Figure 3.2: Brief summary of the science pipeline.

of a sphere system (HEALPix; Górski et al. 2005) using order 11, which corresponds to an area of 2.95 arcmin^2 per tile. Order 11 was chosen as it is a compromise between the number of objects in a tile (more objects means a more reliable flux density extraction) and the time it takes a tile to run.

3.3.2 Science pipeline

The extraction of the flux densities in the SPIRE bands, which follows a similar pipeline to the one used in Pearson et al. (2017), is briefly summarised in Fig. 3.2 and is repeated here for completeness. We begin by using the far-UV (FUV) to PACS $160 \mu\text{m}$ data from COSMOS2015 to generate estimates for the flux densities in all three SPIRE bands simultaneously using CIGALE’s Bayesian analysis. All objects that were not classified as galaxies (TYPE flag in COSMOS2015 not set to 0, approximately 1.3%) were removed, as were objects without any photometric redshift ($\text{ZPDF} \leq 0$, approximately 2.9%). All predicted flux densities were then used in XID+ to extract the flux densities for the objects; we did not remove any objects with poor χ^2 values. Once the SPIRE flux densities were extracted, these SPIRE data were added to the FUV-PACS data and CIGALE was rerun, this time to get estimates for $M_\star$ and SFR. The same CIGALE models were used for the flux estimation and to obtain the physical parameters so the results from each CIGALE run should not be degenerate.

The first CIGALE run provides flux density estimates for all the objects in the catalogue. However, the flux density estimates for the faintest objects will be highly uncertain and hence result in extracted flux densities that are unreliable. To find the depth to which we can reliably run XID+, a number of different cut depths on the predicted flux densities at $250 \mu\text{m}$ were used, from 0.2 mJy to 20 mJy, and XID+ run on 25 tiles. A residual

Figure 3.3: Scatter of the $250\ \mu\text{m}$ residual map using different depth cuts on the prior list. The blue and red points are with and without 3σ clipping, respectively, while the green line is the 1σ instrument noise for the COSMOS field.

map was created by subtracting the replicated map from XID+ from the original image for each depth. In the ideal situation where all sources are correctly accounted for, the residual map should have a scatter consistent with 1σ instrument noise. Figure 3.3 shows how the scatter of the residual $250\ \mu\text{m}$ map changes with cut depth with and without 3σ clipping. As can be seen, as the cut gets deeper, from 20 mJy to 1 mJy, the scatter reduces; cut depths deeper than 1 mJy do not cause much of a change. All cut depths have a scatter greater than the instrument noise at $250\ \mu\text{m}$ (1.71 mJy). This is to be expected as our source list will miss the faintest sources and the flux densities assigned to the known sources will not exactly coincide with the true flux densities for all sources. As a result of all the deeper cuts having a scatter larger than the 1σ instrument noise, and wanting to keep as large a sample as possible while not risking creating too many degeneracies, a predicted $250\ \mu\text{m}$ cut of 0.7 mJy was used. For this cut, we ignore any uncertainty on the flux density estimate. This leaves 205 958 objects to be run through XID+.

To prevent an overly restrictive prior in XID+ and conservatively capture the uncertainty in SED modelling, the errors on the flux density estimates are expanded by a factor of two (see also Appendix 3.7.2). The estimates of all three SPIRE bands from CIGALE are then used as the priors for XID+. The flux density estimates are used as the means in the XID+ priors while the expanded errors are used as the standard deviations. XID+ was then run on the SPIRE images. The next step is to run the data from COSMOS2015 and XID+ through CIGALE to generate the required parameters for each object: $M_\star$, SFR, and U–V and V–J colours; the colours are needed for the QG/SFG separation.

3.3.3 Mass completeness and redshift binning

To study the MS, we removed all objects with a redshift below 0.2 as the mean error on the photometric redshifts for objects below $z = 0.2$ is approximately 0.2 (for discussion on the redshifts and their errors, see Laigle et al. 2016). The remaining objects were then

binned by redshift. Each bin's width was determined by using the average error on the redshift of the objects within a bin; the bin size was the average error of all the objects in the bin rounded up to the next decimal place. Using a bin size of twice the error was also briefly explored and provided consistent results.

Once binned by redshift, the galaxies were cut for completeness. This completeness cut was done empirically by following Pozzetti et al. (2010). The K_s band magnitude limits ($K_{s \text{ lim}}$) were set to 24.7, the 3σ limit for the objects within the UltraVISTA (McCracken et al. 2012) ultra-deep stripes, and 24.0, the 3σ limit for the rest of the COSMOS field, in the ultra-deep and deep regions of COSMOS, respectively (Laigle et al. 2016). For each galaxy with a redshift ($Z_{\text{PDF}} > 0$) and a K_s band detection the mass the galaxy would need (M_{lim}) to be observed at the magnitude limit was calculated using

$$\log(M_{\text{lim}}) = \log(M) - 0.4(K_{s \text{ lim}} - K_s), \quad (3.3)$$

where M is the galaxy's mass in $M_\odot$ and K_s is the galaxy's K_s band magnitude. In each redshift bin, the faintest 20% of objects were selected and the limiting mass was the M_{lim} value which 90% of these faintest objects lie below. As not all objects have a K_s band detection (the catalogue is zYJKs selected), these limiting mass data points were used for all objects. The mass limits for the deep and ultra-deep regions of COSMOS are shown in Fig 3.4.

3.3.4 Forward modelling

To determine the MS in each redshift bin, we use the Markov chain Monte Carlo (MCMC) sampler emcee (Foreman-Mackey et al. 2013) to sample the parameter space of the chosen MS model. For our routine, we create model SFRs using the observed $M_\star$, observed redshift, and the MS being tested at each step. For each $M_\star$, a random number is drawn from a Gaussian distribution centred on the SFR of the MS at that $M_\star$ and with the scatter of the MS as the standard deviation. The Gaussian distribution is also truncated such that it reproduces the observed upper and lower SFR limit. These upper and lower SFR limits are determined by finding the SFR in each redshift bin that 0.1% of the objects fall above or below, and then fitting to these values as a function of redshift using

$$S_{\text{lim}}(z) = [B_0 \times \log(B_1 + z)] + B_2, \quad (3.4)$$

where S is $\log(\text{SFR}/M_\odot \text{yr}^{-1})$ and B_n are the coefficients that are found. The parameters B_0 , B_1 , and B_2 , take the values 0.61 ± 0.32 , $-0.37 \pm 0.01 M_\odot \text{yr}^{-1}$, and $3.13 \pm 0.13 \log(M_\odot \text{yr}^{-1})$ for the upper limit and 2.86 ± 0.19 , $-0.17 \pm 0.05 M_\odot \text{yr}^{-1}$, and $0.12 \pm 0.10 \log(M_\odot \text{yr}^{-1})$ for the lower limit, respectively. The upper and lower limits are applied to each simulated object individually using the observed redshifts.

Once the model SFRs have been generated, both the SFR and $M_\star$ are perturbed by adding a second random number drawn from a Gaussian centred on zero and with a standard deviation equal to the error on the observed SFR or $M_\star$ of that individual object. An example of the MS generated at one step, assuming a linear power law, is shown in Fig. 3.5, along with the MS used to generate at that step and the contours of the observed data.

To compare the model data to the observed data at each step, the two data sets are binned by stellar mass into identical bins with a width of 0.25 dex. The mean and standard deviation of the SFRs in each mass bin are calculated and the mean and standard

Figure 3.4: Masses of all objects detected in the Ks band (blue) are shown against redshift along with the faintest 20% in each redshift bin (red) for the (a) deep and (b) ultra-deep regions. The 90% completeness limit is shown by the thick black lines, while the dashed black lines show the edges of each redshift bin.

Figure 3.5: Example of the data generated at one step of the MCMC routine for the lowest ($0.2 \leq z < 0.5$) redshift bin, shown as number density from high (dark red) to low (dark blue). The MS being tested at this step, in this case the most likely step, is shown as the red line, while the contours for number density of the observed data are shown in orange. The size of the average observed error on SFR and M_* is also shown as a blue cross.

deviations of the model data are compared to their counterparts from the observed data. The smaller the differences, the greater the likelihood that the model is a correct representation of the observed data.

All the parameters in the models used were treated as if they were uncorrelated, although this is not strictly true. However, this method was found to accurately recover input relations used to generate mock sets of data that assume a linear power law. Figure 3.6 shows an example of the posterior for a mock data set with known slope (0.6), normalisation ($0.7 \log(M_*/\text{yr}^{-1})$), and scatter (0.3 dex). As can be seen, the input parameters are recovered within error.

3.4 Results

The M_* found using CIGALE are compared to those found by Laigle et al. (2016) in COSMOS2015. We find that our M_* are higher, on average, by 0.15 dex and are consistent with COSMOS2015 within the COSMOS2015 average error of 0.15 dex, but not the 0.10 dex average error of our M_* . Both Laigle et al. (2016) and this work use the Chabrier (2003) IMF and the Bruzual & Charlot (2003) stellar population model, so this is not the cause of the slight discrepancy, but the dust attenuation models differ: this work uses the Charlot & Fall (2000) dust attenuation, while Laigle et al. (2016) uses the Calzetti et al. (2000) dust attenuation. Recently, Lo Faro et al. (2017) have investigated the effects of dust attenuation laws on the physical properties of dusty galaxies at $z \approx 2$. They find that using the Charlot & Fall (2000) dust attenuation results in higher M_* values than Calzetti et al. (2000), resulting in M_* higher by a factor of 1.4, or approximately 0.15 dex. This is consistent with the difference in M_* found between our current work and Laigle et al.

Figure 3.6: Corner plot (Foreman-Mackey 2016) of the marginalised posterior of the forward modelling routine applied to a mock data set with known slope (0.6), normalisation ($0.7 \log(M_{\odot}/\text{yr}^{-1})$), and scatter (0.3 dex), shown as the blue lines. Panels on the diagonal show the 1D marginalised posteriors for the slope, normalisation and scatter (left to right). Off-diagonal panels show the combined 2D posteriors as labelled by their axes. The recovered 16th, 50th, and 84th percentiles are shown by the dashed vertical lines; all the input parameters are recovered, within error.

(2016). Thus, we conclude that the difference in $M_\star$ is a result of the choice of dust attenuation model.

To remove the QG, we follow the Whitaker et al. (2011) UVJ colour cut

$$\begin{aligned}
 (U - V) &> 0.88 \times (V - J) + 0.69 & z < 0.5 \\
 (U - V) &> 0.88 \times (V - J) + 0.59 & z > 0.5 \\
 (U - V) &> 1.3, (V - J) < 1.6 & z < 1.5 \\
 (U - V) &> 1.3, (V - J) < 1.5 & 1.5 < z < 2.0 \\
 (U - V) &> 1.2, (V - J) < 1.4 & 2.0 < z < 3.5
 \end{aligned} \tag{3.5}$$

where the rest-frame $(U - V)$ and $(V - J)$ colours come from the second CIGALE run. If these conditions are not met, the object is SF. The $(U - V)$ and $(V - J)$ criteria for $2.0 < z < 3.5$ were expanded for all objects with a redshift greater than 2, such that the final line in Eq. 3.5 becomes

$$(U - V) > 1.2, (V - J) < 1.4 \quad 2.0 < z < 6.0. \tag{3.6}$$

Once the QG objects are removed, we fit two models to the data: the Lee et al. (2015) description, which contains a turn-over

$$S = S_0 - \log \left[1 + \left(\frac{M_\star}{M_0} \right)^{-\gamma} \right], \tag{3.7}$$

where S is $\log(\text{SFR}/M_\odot \text{yr}^{-1})$, S_0 is the maximum value of S that the function approaches at high mass, M_0 is the turn-over mass in $M_\odot$, and γ is the low-mass slope, and the Whitaker et al. (2012) single power law

$$S = \alpha[\log(M_\star) - 10.5] + \beta, \tag{3.8}$$

where α is the slope and β is the normalisation at $\log(M_\star/M_\odot) = 10.5$.

Fitting with Eq. 3.7 was reasonably unsuccessful. No redshift bins show significant evidence of a high-mass turn-over with all redshift bins being consistent with a simple power law, as can be seen in the example in Fig. 3.7. The majority of the turn-over masses found from the MCMC forward modelling are larger than the highest mass object within each redshift bin, approximately $\log(M_\star/M_\odot) = 12.5$. As a result, these cannot be considered reliable results as the turn-over position is unconstrained by the data. The turn-over positions at all redshifts, assuming they are reliable, are also much higher than those found in the literature ($\log(M_0/M_\odot) \approx 10.5$ e.g. Lee et al. 2015; Tomczak et al. 2016). We also found that when fitting Eq. 3.7 with γ held fixed at 1.17, which was shown by Lee et al. (2015) to be reasonable and also leaves the same number of free parameters as Eq. 3.8, the majority of the redshift bins have a best fit that is less probable than the best fit using Eq. 3.8. As a result, we conclude that Eq. 3.8 is a better description of our data.

With the knowledge that our data do not support a turn-over in the MS, Eq. 3.8 was also fitted to our data, as shown in Fig. 3.8. This form of the MS was much more successful, with well-constrained fitting parameters in all redshift bins (see Table 3.2 and Fig. 3.9). The β parameter (normalisation) increases rapidly with redshift out to $z \approx 2$ before increasing more slowly (see magenta points in Fig. 3.10b). The evolution of β was

Figure 3.7: Comparison of fitting with Eq. 3.7 (orange dashed line) and Eq. 3.8 (red line) in the $0.5 \leq z < 0.8$ redshift bin. The galaxies are shown as a number density plot, with dark red being high number density and dark blue low number density, and the size of the average error on SFR and M_* is shown as a blue cross. There is very little difference in shape between the two fits, demonstrating that Eq. 3.8 is the preferred form of the MS.

found, fitting using the SciPy (Jones et al. 2001) orthogonal distance regression (ODR; Boggs & Rogers 1990) package, to follow

$$\beta(z) = (1.10 \pm 0.07) + [(0.53 \pm 0.05) \times \ln\{0.03 \pm 0.11\} + z], \quad (3.9)$$

with errors in the coefficients from the ODR fitting.

The α parameter (slope) is slightly less smooth in its evolution (see Fig. 3.10a, magenta). Generally, the slope increases with redshift across the entire redshift range of this study. There is a potential rise between $z \approx 1.8$ and $z \approx 2.9$, although this is very marginal and may be a result of the SED model gridding causing a higher slope (see Fig. 3.8). A linear relation was used to find the slope evolution, which was determined to be

$$\alpha(z) = (0.38 \pm 0.04) + (0.12 \pm 0.02)z. \quad (3.10)$$

Short tests were also done using just the deep and ultra-deep data. At redshifts below 4.9, just using the deep or ultra-deep data made no significant difference to the α and β parameters: the values change within error, compared to using both the deep and ultra-deep data together. With just the deep data, the slope of the highest redshift bin becomes much lower, although it is consistent within error. For normalisation, the deep data has a lower normalisation, but again it is consistent within error. This is likely a result of low number statistics. There are only 16 UVJ selected objects with $z \geq 4.9$ in the deep coverage. With so few objects to fit with, the results are highly uncertain with just the deep data.

After determining the MS, we check the scatter to ensure that we have actually found the MS. The intrinsic scatter was determined during the MCMC fitting and was assumed to be independent of M_* , which is believed to be true (Whitaker et al. 2012; Speagle et al. 2014; Tomczak et al. 2016). Our intrinsic scatter is found to be consistent with other

literature (e.g. Whitaker et al. 2012; Speagle et al. 2014; Sparre et al. 2015; Schreiber et al. 2015; Tomczak et al. 2016, see Fig. 3.11), suggesting that we have indeed found the MS. We also fit a flat line to our intrinsic scatter, assuming that there is no redshift evolution, which gives an amplitude of 0.23 ± 0.03 dex, lower than the expected 0.3 dex typically found in observational works but consistent with the results from the Illustris Simulation (Genel et al. 2014; Vogelsberger et al. 2014; Sparre et al. 2015). Our small intrinsic scatter may be an underestimation resulting from the $M_\star$ and SFR being derived from the same SED fitting procedure. This results in some correlation between these two quantities and hence reduces the scatter about the MS.

Table 3.2: Parameters from fitting Eq. 3.8 to the UVJ selected star forming galaxies, with α the slope and β the normalisation at $\log(M_*/M_\odot) = 10.5$. The intrinsic scatter is found during the MCMC fitting.

Redshift Bin	Average Redshift	Number of Sources	α	α Error	β $\log(M_\odot \text{yr}^{-1})$	β Error $\log(M_\odot \text{yr}^{-1})$	Scatter dex	Scatter Error dex
$0.2 \leq z < 0.5$	0.37	11460	0.43	0.09	0.58	0.09	0.35	0.06
$0.5 \leq z < 0.8$	0.66	17615	0.50	0.10	0.92	0.08	0.33	0.06
$0.8 \leq z < 1.1$	0.95	24537	0.46	0.11	1.10	0.08	0.34	0.06
$1.1 \leq z < 1.4$	1.24	20712	0.48	0.09	1.22	0.07	0.28	0.06
$1.4 \leq z < 1.8$	1.59	19388	0.51	0.09	1.31	0.08	0.24	0.05
$1.8 \leq z < 2.3$	2.02	13166	0.74	0.14	1.39	0.19	0.29	0.08
$2.3 \leq z < 2.9$	2.59	6375	0.83	0.15	1.59	0.20	0.28	0.07
$2.9 \leq z < 3.8$	3.23	2338	0.70	0.09	1.77	0.08	0.10	0.05
$3.8 \leq z < 4.9$	4.34	590	0.93	0.22	1.87	0.20	0.15	0.07
$4.9 \leq z < 6.0$	5.18	72	1.00	0.22	1.92	0.21	0.08	0.05

3.5 Discussion

3.5.1 Form of the main sequence

In this work, we find little evidence of a turn-over. Thus, any cause of the reduction in specific SFR ($\text{SFR}/M_\star$, sSFR) with mass does not suddenly ‘turn on’, rather it is an effect that slowly increases with $M_\star$. An increasing slope with redshift indicates that the mass dependent reduction in sSFR becomes weaker as we look further back, meaning that the mechanism behind this was less strong in the past than it is now.

Second, at relatively low redshift we find that the slope is relatively shallow, which suggests that the ability of galaxies to form stars reduces as mass increases. This may be a result of a decrease in star formation efficiency, a reduction in the available cold gas, or some other mass-dependent quenching mechanism. We also see that the slope increases with redshift, implying that galaxies in the younger universe were more self-similar than they are today. Thus, in the modern universe, large galaxies have lower specific SFRs (sSFR) than smaller galaxies, while at high redshift, the sSFR s of large and small galaxies are similar.

Third, we find that the slope of the MS is very high between redshifts of 1.8 and 2.9 with respect to other epochs. We do note, however, that the significance of this rise is very marginal. A steep slope at these times suggests a higher sSFR in large galaxies compared to other times. This increase in the slope coincides with the peak of the cosmic star formation history (CSFH) at $z \approx 2 - 3$ (Madau & Dickinson 2014). It may also simply be a result of the SED model gridding causing a higher slope, indicated by the discontinuities in the two top right panels of Fig. 3.8.

When the normalisation is also considered, the low-mass end of the MS also sees changes in a galaxy’s ability to form stars. Equation 3.8 has a normalisation at $\log(M_\star/M_\odot) = 10.5$ and it does not evolve fast enough to maintain a constant crossing point of the MS at all redshifts; the intercept mass of the MS in one redshift bin with the MS in the lowest redshift bin increases out to $z \approx 1.8$ before becoming approximately constant, as can be seen in Fig. 3.9. Thus, the low-mass galaxies must have a decrease in sSFR as redshift increases, out to $z \approx 1.8$, while the high-mass galaxies have an increasing sSFR .

3.5.2 Contamination by quiescent galaxies

We find the main sequences in the highest redshift bins ($z > 1.8$) are identical regardless of whether we cut for SFG or not. This may be a result of the dusty SFG and QG at high redshift having their near-IR (NIR) observed frame emission suppressed below the detection limits of the NIR selected catalogue by dust obscuration in the galaxy. Alternatively, there is a greater mixing of SFG and QG at lower redshifts and higher masses, pulling down the MS when the QG classed objects are removed. The latter is likely the correct scenario. There is little evidence for a turn-over at lower redshifts after selection of the SFGs. We thus surmise that the turn-over in other works may be the result of a SFG selection that is not effective enough in removing the QG. This would result in an overabundance of low SFR galaxies at higher mass which act to pull the high-mass MS down, similar to the conclusions of Johnston et al. (2015). It is the lack of turn-off in the MS at lower redshifts that leads us to believe that the high- z colour cut employed (Eq.

Figure 3.8: Fitting of Eq. 3.8 to the objects in the redshift bins as labelled. The solid line is the most likely MS across the fitted M_* range, while the dotted line is an extrapolation across the M_* range of the plot. The galaxies are shown as a number density plot, with dark red being high number density and dark blue low number density. The vertical density discontinuities are the result of the two depths of data used: the deep mass limit is the yellow dashed line and the ultra-deep mass limit is the dot-dashed yellow line. Each panel also indicates the χ^2 of the most likely MS ($\chi^2_{\min}$) and the number of degrees of freedom in the fitting (n_{dof}), and shows the size of the average SFR and M_* errors as a blue cross.

Figure 3.9: Fitted MS trends using Eq. 3.8. The solid and dashed lines are the MS across the fitted M_* range. The dotted lines are extrapolations across the M_* range of the plot. The normalisation of the MS clearly increases as redshift increases. The slight decrease in slope at low redshift can be seen along with the increase above $z = 1.1$. The very steep slopes at $1.8 \leq z < 2.9$ can also be seen.

3.5) can be extended to all redshifts above $z = 2.0$ as we also recover the linear MS at these higher redshifts.

3.5.3 Comparisons with other works

The work by Speagle et al. (2014) collates a large number of MS studies and generates a consensus regarding the MS. This provides an excellent base line to which we can compare this work. As with the consensus result, this work finds little evidence of a high-mass turn-over. Broadly, our results agree with Speagle et al. (2014): both the slope and normalisation increase with redshift; the rate of change in normalisation decreases as redshift increases. However, quantitatively there are differences.

The normalisation found in this work is in reasonable agreement with the Speagle et al. (2014) collection (see Fig. 3.10b), but lies towards the bottom of the collection. This slight lowering of the normalisation, especially compared with studies using IR SFR indicators, may be due to the splitting of brighter objects into a number of fainter sources.

The slope of this work is consistent with the Speagle et al. (2014) results at all redshifts, rising from the bottom of the collection at low redshifts to the top of the collection at high redshifts. Between $z \approx 2$ and $z \approx 3$ we are approximately consistent with the far-ultraviolet SFR indicated MS, but are typically higher than the MS using other (radio, hydrogen lines and UV SED fitting) SFR indicators. This is potentially a result of the data in this work looking through the dust with the *Herschel* data, which will raise the SFR of the high-mass, dusty galaxies. However, this should also result in the FUV MS having a shallower slope than this work as well.

The work by Schreiber et al. (2015) looked at the MS using *Herschel*, but their study relied on stacking rather than de-blending. We convert the Schreiber et al. (2015) results from the Salpeter (1955) IMF to the Chabrier (2003) IMF used here, following Speagle

Figure 3.10: Comparison of the UVJ selected MS results of this work (magenta) with the observational MS from Speagle et al. (2014, FUV data in blue; IR data in green; and radio, hydrogen lines, and UV SED fitting in yellow), Schreiber et al. (2015) low mass MS (red), and the MS from the Illustris Simulation (Vogelsberger et al. 2014; Sparre et al. 2015, black diamonds). The α and β parameters from Eq. 3.8 are in panels (a) and (b), respectively. As Schreiber et al. (2015) hold their low-mass slope constant at unity, this is indicated in panel (a) as a flat red line. The redshifts shown for this work are the mean redshift in each redshift bin, while the horizontal error bars show the width of the redshift bin. A version of this plot using SFRs derived from IR template fitting can be found in Appendix 3.7.3.

Figure 3.11: Intrinsic scatter about the MS found from the MCMC fitting of the data in each redshift bin to Eq. 3.8. This work (magenta) is shown along with the observed scatters from Speagle et al. (2014, FUV data in blue, IR data in green, and other data in yellow) and Schreiber et al. (2015, red) as well as the median scatter found at each redshift in the Illustris Simulation (Genel et al. 2014; Vogelsberger et al. 2014; Sparre et al. 2015, black diamonds). The redshifts shown for this work are the mean redshift in each redshift bin while the horizontal error bars show the width of the redshift bin. Also shown is the best fit to this work’s intrinsic scatter assuming no redshift evolution (dashed magenta line). The intrinsic scatter found in this work is consistent with existing literature, although above $z \approx 1.8$ our intrinsic scatter is smaller than is found in works that use IR SFR traces.

et al. (2014, see Eq. 2) to correct the mass and Kennicutt & Evans (2012) to correct the SFR. Their results found a turn-over at high mass, unlike the results of this study. This high-mass turn-over is not a result of stacking; we performed our own stacking analysis and found no turn-over or UVJ selection because our UVJ selection is similar to that of Schreiber et al. (2015) (see Appendix 3.7.3).

As a result of their turn-over, just the low-mass slope of Schreiber et al. (2015) is used for comparison in this work. Schreiber et al. (2015) also enforce a slope of unity while the normalisation is allowed to change. A non-redshift dependant slope of unity is not found to be a good description of our data, with the slope remaining below this until the highest redshift bin. We also find that our normalisation is considerably lower at all redshifts, by approximately 0.4 dex, with the difference being larger at lower redshifts (see Fig. 3.10). However, this difference in normalisation is a result of the different methods used to derive the SFR of the galaxies, with our SED-based SFR producing values lower by 0.4 dex when compared to the SFRs derived from the IR luminosity from CIGALE. We also see the same 0.4 dex offset when deriving our own IR luminosity SFRs by fitting our SPIRE data to Chary & Elbaz (2001) templates (see Appendix 3.7.3).

We also compare our results to those from the Illustris Simulation (Genel et al. 2014; Vogelsberger et al. 2014; Sparre et al. 2015). We fit Eq. 3.8 to the SFR- $M_\star$ positions in Fig. 1 of Sparre et al. (2015) for the $z = 0, 1, 2$, and 4 redshift bins. The $z = 0$ bin shows clear evidence of a turn-over, but we still fit a simple power law to all the data in this redshift bin. The resulting α and β values can be found in Fig. 3.10 as black diamonds. Our work has slopes that are much shallower than the Illustris values below a redshift of ≈ 1.8 . At $z \approx 2$, this work and Illustris agree within error, while at $z \approx 4$ our slope appears to be consistent with Illustris. The normalisation of the MS from the Illustris simulation, along with other simulations, is known to be too low with respect to observations, certainly at $z = 1 - 2$ (Sparre et al. 2015; Furlong et al. 2015). Our normalisation results are not consistent with this picture; the normalisation we find is consistent with the results from Illustris. This may suggest that the low normalisations found in simulations may be the result of simulations not being affected by the source blending that observational studies can show.

3.6 Conclusions

The MS is a tight relation between $M_\star$ and the SFR of galaxies in the universe. The FIR is a very important part of the spectrum for determining the SFR of an object; however, current cosmological FIR surveys suffer from poor resolution. Here, the latest de-blending tool, XID+, and techniques have been used to break through the *Herschel* confusion limit, allowing better SFR estimates to be generated for nearly a quarter of a million objects in the COSMOS field. This has allowed the examination of the main sequence of star forming galaxies beyond the *Herschel* confusion limit.

The objects were binned by redshift between $0.2 \leq z < 6.0$ and fitted with both a power law and a power law with a high-mass turn-over. The high-mass turn-over form of the MS failed to provide reasonable values for its parameters. The simple power law, on the other hand, produced well-behaved parameters in every redshift bin. Thus, we conclude that the MS using multi-wavelength data and our de-blended *Herschel* SPIRE

data shows little evidence for a high-mass turn-over.

We find that the normalisation of the MS increases with redshift, rapidly out to $z \approx 2$ and more slowly thereafter. The slope has a different evolution, rising steadily across all redshifts, with perhaps a weak indication of a brief peak at $1.8 \leq z < 2.9$. An increase in slope with redshift is likely a result of high-mass galaxies having an increase in star formation efficiency with respect to the low-mass galaxies. This results in all galaxies becoming more self similar with increasing redshift, as the slope increases towards unity. The brief peak in slope, if present, appears to coincide with the peak of the CSFH.

Our normalisation results are consistent with the Speagle et al. (2014) MS compilation results within error, and with the Schreiber et al. (2015) stacked *Herschel* MS, once adjusted for IMF and SFR tracers. The slope of our MS is also consistent with the Speagle et al. (2014) compilation but moves from the lower regions at low redshift to the upper regions at higher redshifts. This may be a result of hidden star formation being revealed by the *Herschel* data, demonstrating the importance of the FIR/sub-mm for the MS.

Acknowledgements. We would like to thank the anonymous referee whose comments greatly improved this paper.

We would like to thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high performance computing cluster.

This research has made use of data from HerMES project (<http://hermes.sussex.ac.uk/>). HerMES is a Herschel Key Programme utilising Guaranteed Time from the SPIRE instrument team, ESAC scientists, and a mission scientist.

The HerMES data were accessed through the Herschel Database in Marseille (HeDaM - <http://hedam.lam.fr>) operated by CeSAM and hosted by the Laboratoire d'Astrophysique de Marseille.

HerMES DR4 was made possible through support of the Herschel Extragalactic Legacy Project, HELP (<http://herchel.sussex.ac.uk>). PDH and SJO would like to acknowledge the research leading to these results, which has received funding from the European Union Seventh Framework Programme FP7/2007-2013/ under grant agreement n° 607254. This publication reflects only the author's view and the European Union is not responsible for any use that may be made of the information contained therein.

Bibliography

- Álvarez-Márquez, J., Burgarella, D., Heinis, S., et al. 2016, *A&A*, 587, A122
- Bavouzet, N. 2008, *Theses*, Université Paris Sud - Paris XI
- Béthermin, M., Dole, H., Beelen, A., & Aussel, H. 2010a, *A&A*, 512, A78
- Béthermin, M., Dole, H., Cousin, M., & Bavouzet, N. 2010b, *A&A*, 516, A43
- Boggs, P. T., & Rogers, J. E. 1990, in *Contemp. Math.*, Vol. 112, Statistical analysis of measurement error models and applications (Arcata, CA, 1989) (Amer. Math. Soc., Providence, RI), 183–194
- Bourne, N., Dunlop, J. S., Merlin, E., et al. 2017, *MNRAS*, 467, 1360
- Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, *MNRAS*, 351, 1151
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, 344, 1000
- Burgarella, D., Buat, V., & Iglesias-Páramo, J. 2005, *MNRAS*, 360, 1413
- 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
- Chary, R., & Elbaz, D. 2001, *ApJ*, 556, 562
- Ciesla, L., Charmandaris, V., Georgakakis, A., et al. 2015, *A&A*, 576, A10
- Dale, D. A., Cohen, S. A., Johnson, L. C., et al. 2009, *ApJ*, 703, 517
- Draine, B. T., & Li, A. 2007, *ApJ*, 657, 810
- Draine, B. T., Aniano, G., Krause, O., et al. 2014, *ApJ*, 780, 172
- Dunlop, J. S., McLure, R. J., Biggs, A. D., et al. 2017, *MNRAS*, 466, 861
- Dunne, L., Gomez, H. L., da Cunha, E., et al. 2011, *MNRAS*, 417, 1510
- Elbaz, D., Daddi, E., Le Borgne, D., et al. 2007, *A&A*, 468, 33

- Foreman-Mackey, D. 2016, The Journal of Open Source Software, 24, doi:10.21105/joss.00024
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
- Fritz, J., Franceschini, A., & Hatziminaoglou, E. 2006, MNRAS, 366, 767
- Furlong, M., Bower, R. G., Theuns, T., et al. 2015, MNRAS, 450, 4486
- Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, MNRAS, 362, 41
- Genel, S., Vogelsberger, M., Springel, V., et al. 2014, MNRAS, 445, 175
- Genzel, R., Tacconi, L. J., Lutz, D., et al. 2015, ApJ, 800, 20
- Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
- Griffin, M. J., Abergel, A., Abreu, A., et al. 2010, A&A, 518, L3
- Harris, K., Farrah, D., Schulz, B., et al. 2016, MNRAS, 457, 4179
- Hurley, P. D., Oliver, S., Betancourt, M., et al. 2017, MNRAS, 464, 885
- Johnston, R., Vaccari, M., Jarvis, M., et al. 2015, MNRAS, 453, 2540
- Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open source scientific tools for Python
- Kennicutt, R. C., & Evans, N. J. 2012, ARA&A, 50, 531
- Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18
- Kurczynski, P., & Gawiser, E. 2010, AJ, 139, 1592
- Laidler, V. G., Papovich, C., Grogan, N. A., et al. 2007, PASP, 119, 1325
- Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, ApJS, 224, 24
- Larson, D., Dunkley, J., Hinshaw, G., et al. 2011, ApJS, 192, 16
- Lee, N., Sanders, D. B., Casey, C. M., et al. 2015, ApJ, 801, 80
- Lo Faro, B., Buat, V., Roehlly, Y., et al. 2017, MNRAS, 472, 1372
- Lutz, D., Poglitsch, A., Altieri, B., et al. 2011, A&A, 532, A90
- Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415
- McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, A&A, 544, A156
- Merlin, E., Fontana, A., Ferguson, H. C., et al. 2015, A&A, 582, A15
- Meurer, G. R., Heckman, T. M., & Calzetti, D. 1999, ApJ, 521, 64

- Mitchell, P. D., Lacey, C. G., Baugh, C. M., & Cole, S. 2013, *MNRAS*, 435, 87
- Mitra, S., Davé, R., Simha, V., & Finlator, K. 2017, *MNRAS*, 464, 2766
- Nguyen, H. T., Schulz, B., Levenson, L., et al. 2010, *A&A*, 518, L5
- Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, *ApJ*, 660, L43
- Noll, S., Burgarella, D., Giovannoli, E., et al. 2009, *A&A*, 507, 1793
- Oliver, S. J., Bock, J., Altieri, B., et al. 2012, *MNRAS*, 424, 1614
- Pan, Z., Zheng, X., & Kong, X. 2017, *ApJ*, 834, 39
- Pannella, M., Elbaz, D., Daddi, E., et al. 2015, *ApJ*, 807, 141
- Pearson, W. J., Wang, L., van der Tak, F. F. S., et al. 2017, *A&A*, 603, A102
- Pilbratt, G. L., Riedinger, J. R., Passvogel, T., et al. 2010, *A&A*, 518, L1
- Poglitsch, A., Waelkens, C., Geis, N., et al. 2010, *A&A*, 518, L2
- Pozzetti, L., Bolzonella, M., Zucca, E., et al. 2010, *A&A*, 523, A13
- Rieke, G. H., Young, E. T., Engelbracht, C. W., et al. 2004, *ApJS*, 154, 25
- Roseboom, I. G., Oliver, S. J., Kunz, M., et al. 2010, *MNRAS*, 409, 48
- Roseboom, I. G., Ivison, R. J., Greve, T. R., et al. 2012, *MNRAS*, 419, 2758
- Safarzadeh, M., Ferguson, H. C., Lu, Y., Inami, H., & Somerville, R. S. 2015, *ApJ*, 798, 91
- Salpeter, E. E. 1955, *ApJ*, 121, 161
- Sanders, D. B., Salvato, M., Aussel, H., et al. 2007, *ApJS*, 172, 86
- Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, *A&A*, 575, A74
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, 172, 1
- Scoville, N., Aussel, H., Sheth, K., et al. 2014, *ApJ*, 783, 84
- Scoville, N., Sheth, K., Aussel, H., et al. 2016, *ApJ*, 820, 83
- Scudder, J. M., Oliver, S., Hurley, P. D., et al. 2016, *MNRAS*, 460, 1119
- Serra, P., Amblard, A., Temi, P., et al. 2011, *ApJ*, 740, 22
- Sparre, M., Hayward, C. C., Springel, V., et al. 2015, *MNRAS*, 447, 3548
- Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, *ApJS*, 214, 15
- Tacchella, S., Dekel, A., Carollo, C. M., et al. 2016, *MNRAS*, 457, 2790
- Tacconi, L. J., Genzel, R., Neri, R., et al. 2010, *Nature*, 463, 781

- Tomczak, A. R., Quadri, R. F., Tran, K.-V. H., et al. 2016, *ApJ*, 817, 118
- Viero, M. P., Moncelsi, L., Quadri, R. F., et al. 2013, *ApJ*, 779, 32
- Vogelsberger, M., Genel, S., Springel, V., et al. 2014, *MNRAS*, 444, 1518
- Wang, L., Viero, M., Clarke, C., et al. 2014, *MNRAS*, 444, 2870
- Wang, L., Norberg, P., Bethermin, M., et al. 2016, *A&A*, 592, L5
- Whitaker, K. E., van Dokkum, P. G., Brammer, G., & Franx, M. 2012, *ApJ*, 754, L29
- Whitaker, K. E., Labbé, I., van Dokkum, P. G., et al. 2011, *ApJ*, 735, 86
- Wright, A. H., Robotham, A. S. G., Bourne, N., et al. 2016, *MNRAS*, 460, 765

3.7 Appendix

3.7.1 CIGALE Parameters

In this work we closely follow the CIGALE model selection of Pearson et al. (2017, Appendix A). A double exponential SFH was used over the more commonly used exponentially declining or delayed exponential SFH as these two SFHs did not appear to reproduce the expected starburst population in the star formation rate versus stellar mass plane. The e-folding time of the two stellar populations (old and young) in the SFH was roughly matched to that of Mitchell et al. (2013). As Mitchell et al. (2013) used a single declining exponential SFH, the e-folding times were split with the burst population taking values of 9 Gyr and above, and the main population taking values of less than 9 Gyr. For the ages of the main population, the values were sampled linearly between 1 and 11 Gyr with a wider sampling up to 13 Gyr. The mass fraction of the burst population follows Ciesla et al. (2015) along with the age of the young stellar population, which also had a lower age of 0.001 Gyr added. The Bruzual & Charlot (2003) stellar population model was used with a Chabrier (2003) initial mass function. As this study was not to explore the metallicity of galaxies, it was decided to leave the metallicity at solar.

Recent work by Lo Faro et al. (2017) has shown that a single power law is a poor representation of the attenuation that occurs in dusty galaxies. As such, we adopt a double power law, similar to that of Charlot & Fall (2000), with individual power laws for the birth clouds (BCs) and inter stellar medium (ISM). However, we use a smaller range of values for the V-band attenuation in the BCs than Charlot & Fall (2000), a maximum value of 3.8, as tests with higher attenuation values do not noticeably effect the results.

For the dust emission, the polycyclic aromatic hydrocarbon (PAH) fraction had an increase in range around the default 2.5 so more fractions could be sampled while keeping the number of models reasonable. The minimum scaling factor of the radiation field was similarly given a range to sample with an increase in the smallest value from 1.0 to 5.0. The illuminated fraction was reduced to 0.02, following Ciesla et al. (2015).

The parameters in the AGN model were matched to those used by Ciesla et al. (2015), who undertook a detailed study of AGN host galaxy emission using CIGALE. The number

of choices of frac_{AGN} was reduced from 14 to 5, while still covering the same range, to reduce the number of models created by CIGALE and hence decrease runtime.

A list of parameters, where they differ from default, can be found in Table 3.3.

Table 3.3: Parameters used for the various properties in the CIGALE model SEDs where they differ from the default values. All ages and times are in Gyr.

Parameter	Value	Description
Star Formation History		
τ_{main}	1.0, 1.8, 3.0, 5.0, 7.0	e-folding time (main)
τ_{burst}	9.0, 13.0	e-folding time (burst)
f_{burst}	0.00, 0.10, 0.20, 0.30	Burst mass fraction
Age	1.00, 1.50, 2.00, 2.50, 3.00, 3.50, 4.00, 4.50, 5.00, 5.50, 6.00, 6.50, 7.00, 7.50, 8.00, 8.50, 9.00, 9.50, 10.00, 10.50, 11.00, 12.00, 13.00	Population age (main)
Burst Age	0.001, 0.010, 0.030, 0.100, 0.300	Population age (burst)
Stellar Emission		
IMF	Chabrier (2003)	Initial Mass Function
Z	0.02	Metallicity (0.02 is Solar)
Separation Age	0.01	Separation between young and old stellar populations
Dust Attenuation		
A_V^{BC}	0.3, 1.2, 2.3, 3.3, 3.8	V-band attenuation of the birth clouds
Slope _{BC}	-0.7	Birth cloud attenuation power law slope
BC to ISM Factor	0.3, 0.5, 0.8, 1.0	Ratio of the birth cloud attenuation to ISM attenuation
Slope _{ISM}	-0.7	ISM attenuation power law slope
Dust Emission		

q_{PAH}	0.47, 1.12, 2.50, 3.9	Mass fraction of PAH
U_{min}	5.0, 10.0, 25.0	Minimum scaling factor of the radiation field intensity
α	2.0	Dust power law slope
γ	0.02	Illuminated fraction
AGN Emission		
r_{ratio}	60.0	Ratio of maximum to minimum radii
τ	1.0, 6.0	Optical depth at 9.7 μm
β	-0.5	β coefficient for the gas density function of the torus ^a
γ	0.0	γ coefficient for the gas density function of the torus ^a
Opening Angle	100.0°	Opening angle of the torus
ψ	0.001°, 89.990°	Angle between equatorial axis and line of sight
$frac_{AGN}$	0.0, 0.1, 0.3, 0.5, 0.7	AGN fraction

Notes. ^(a) Density function of the torus can be found in Fritz et al. (2006) as Equation 3.

3.7.2 XID+ prior width

We require the prior from CIGALE to be as informative as possible in XID+ while not being overly restrictive. To determine the preferred expansion factor of the SPIRE flux density estimate errors from CIGALE, we ran XID+ on the tiles that contain at least one of 178 objects detected at $870\mu\text{m}$ from Scoville et al. (2014, 2016) three times; expanding the flux density estimate error by a factor of 2, 3, and 4. The SPIRE flux densities from XID+ were then added to the data from COSMOS2015 and CIGALE was rerun to predict the Atacama Large Millimeter/submillimeter Array (ALMA) $870\mu\text{m}$ flux densities, for each of the three XID+ results. These predictions were compared to the ALMA $870\mu\text{m}$ observations by subtracting the CIGALE predictions from the observations and fitting a Gaussian to the resulting distribution. The histograms of the comparisons, along with the fitted Gaussian distributions, are found in Fig. 3.12. As the standard deviations are all approximately consistent, we chose to expand the errors from CIGALE by a factor of two as this provides the smallest deviation of the mean from zero.

3.7.3 Further comparisons with Schreiber et al. (2015)

As mentioned in Sect. 3.5.3, the normalisation presented in this work is consistent with that of Schreiber et al. (2015) once we account for IMF and SFR tracers. We also note that the high-mass turn-over found in Schreiber et al. (2015) is not a result of stacking. We discuss this here in further detail.

The disparity in normalisation is likely a result of the different methods of determining SFR. Schreiber et al. (2015) convert the UV and IR luminosities into UV and IR SFRs and then combine these values to generate a total SFR, while we use full SED fitting. As a result, our SFRs are lower as not all of the IR luminosity will be attributed to the young stellar population, but it also includes a component for the older stellar population. We have also fitted the Chary & Elbaz (2001) templates to our own de-blended SPIRE data, calculated the IR luminosities, and converted these into SFRs. From this fitting, we recover similar SFRs to those found by Schreiber et al. (2015) (Fig. 3.13, blue triangles). Hence, the SFRs from SED fitting are systematically lower than those found using IR luminosity conversion.

The forward modelling was re-done using the SFRs derived from the fitting of the Chary & Elbaz (2001) templates to our SPIRE data. The slope (α) and normalisation (β) were found to evolve as

$$\begin{aligned}\alpha(z) &= (0.58 \pm 0.05) + (0.11 \pm 0.03)z, \\ \beta(z) &= (1.31 \pm 0.11) + [(0.46 \pm 0.08) \times \ln(\{0.09 \pm 0.20\} + z)].\end{aligned}\tag{3.11}$$

As can be seen in Fig. 3.14, the slope using the Chary & Elbaz (2001) SFR evolves with redshift at the same rate as when the CIGALE SFRs are used but the slopes are steeper, a result of high-mass galaxies having more older stars (e.g. Gallazzi et al. 2005), which will lower the CIGALE SFR with respect to the Chary & Elbaz (2001) SFR for these high-mass objects. As expected, the normalisations are closer to those of the Schreiber et al. (2015) low-mass slope and become consistent, within error, at high redshift.

The remaining normalisation difference at low redshift is a result of the enforced slope of unity in Schreiber et al. (2015). We fit Eq. 3.8 to the data in Schreiber et al. (2015) Fig.

Figure 3.12: Distributions for the difference between the observed ALMA 870 μm flux densities and the predicted 870 μm flux densities from CIGALE using an error expansion factor of 2 (a), 3 (b), and 4 (c) in XID+. The Gaussian distribution for each expansion factor is also shown in orange, along with the means (μ) and standard deviations (σ) of the distributions. All the distributions have an approximately consistent σ so the best expansion factor was deemed to be that with μ closest to zero: 2 times the error from CIGALE.

Figure 3.13: Comparisons of Schreiber et al. (2015, red crosses) SFR- $M_{\star}$ positions with redshift to different methods of deriving SFR in this work. The mean CIGALE SFR in each mass bin is in magenta, the fitting of the de-blended SPIRE data with Chary & Elbaz (2001, CE01) templates to determine the infrared luminosity (L_{IR}) derived SFR is indicated with blue triangles, the IAS Stacking Library derived SFR with yellow stars, the full Schreiber et al. (2015) MS trends with red dashed lines, and our own fits to the Schreiber et al. (2015) data with orange dot-dashed lines. Horizontal error bars are omitted for ease of examination, but are all 0.25 dex. The Schreiber et al. (2015) data is complete to lower masses due to their use of the deeper GOODS data. As can be seen, the SFRs from SED fitting are systematically lower than those from stacking or template fitting. However, our stacked and CE01 data points are consistent with Schreiber et al. (2015) within error.

Figure 3.14: Comparison of the UVJ selected MS results of this work using the Chary & Elbaz (2001, CE01) template derived SFR (dark purple triangles) with the Schreiber et al. (2015) low-mass MS (red). The α and β parameters from Eq. 3.8 are in panels (a) and (b), respectively, and the trends from Fig. 3.10 are in magenta. As Schreiber et al. (2015) hold their low-mass slope constant at unity, this is indicated in panel (a) as the flat red line. The orange crosses in panel (b) are the normalisations for the full mass range (including turn-over) from Schreiber et al. (2015). The redshifts shown for this work are the mean redshift in each redshift bin while the horizontal error bars show the width of the redshift bin. The good agreement between the purple triangles and orange crosses and poor agreement between the purple triangles and red circles demonstrates how forcing a low-mass slope of unity results in normalisations that are too high at lower redshifts.

10 and find slopes below unity. As a result of this fitting, the normalisations from our fits to the Schreiber et al. (2015) data agree with the normalisations we find when fitting to our IR luminosity derived SFRs. Figure 3.14b also includes the normalisations for the full mass range from Schreiber et al. (2015), including their high-mass turn-over, as orange crosses. These normalisations are in good agreement with the normalisations found in this work, demonstrating how enforcing a low-mass slope of unity results in normalisations that are too high at lower redshifts.

Our own fitting to the Schreiber et al. (2015) MS also provides a slightly different interpretation of their data. In addition to fitting Eq. 3.8 to their data, we also fit Eq. 3.7. The result of this refitting is that we only find evidence of a high-mass turn-over in the two lowest redshift bins used by Schreiber et al. (2015); all the other redshift bins are consistent with a simple power law (Fig. 3.13 red dashed lines). Hence, their high-mass turn-over above $z = 1.2$ may be a result of forcing a low-mass slope of unity.

To explore whether our lack of a high-mass turn-over at low redshift is due to stacking, we stacked the SPIRE (from HerMES, Oliver et al. 2012) and PACS (from PACS Evolutionary Probe, Lutz et al. 2011) images of the COSMOS field on our source positions, using the same UVJ selection, mass bins, and redshift bins as Schreiber et al. (2015) to generate a fair comparison. We used two independent stacking codes: the IAS Stacking Library (Bavouzet 2008; Béthermin et al. 2010a) for luminosity stacking (Fig. 3.13 yellow stars), and Simstack (Viero et al. 2013) for flux density stacking. The stacked *Herschel* flux densities were then fitted with the Chary & Elbaz (2001) IR templates to calculate the IR luminosity and SFR. We find no evidence of a high-mass turn-over using either of the stacking tools.

Identifying galaxy mergers in observations and simulations with deep learning

W. J. Pearson, L. Wang, J. W. Trayford, C. E. Petrillo, F. F. S. van der Tak
 Astronomy & Astrophysics, 626, A49, 2019
 Reproduced with permission, ©ESO

Abstract

Context: Mergers are an important aspect of galaxy formation and evolution. With large upcoming surveys, such as *Euclid* and LSST, accurate techniques that are fast and efficient are needed to identify galaxy mergers for further study.

Aims: We aim to test whether deep learning techniques can be used to reproduce visual classification of observations, physical classification of simulations and highlight any differences between these two classifications. As one of the main difficulties of merger studies is the lack of a truth sample, we can use our method to test biases in visually identified merger catalogues. *Methods:* We developed a convolutional neural network architecture and trained it in two ways: one with observations from SDSS and one with simulated galaxies from EAGLE, processed to mimic the SDSS observations. The SDSS images were also classified by the simulation trained network and the EAGLE images classified by the observation trained network. *Results:* The observationally trained network achieves an accuracy of 91.5% while the simulation trained network achieves 65.2% on the visually classified SDSS and physically classified EAGLE images respectively. Classifying the SDSS images with the simulation trained network was less successful, only achieving an accuracy of 64.6%, while classifying the EAGLE images with the observation network was very poor, achieving an accuracy of only 53.0% with preferential assignment to the non-merger classification. This suggests that most of the simulated mergers do not have conspicuous merger features and visually identified merger catalogues from observations are incomplete and biased towards certain merger types.

Conclusions: The networks trained and tested with the same data perform the best, with observations performing better than simulations, a result of the observational sample being biased towards conspicuous mergers. Classifying SDSS observations with the simulation trained network has proven to work, providing tantalising prospects for using simulation trained networks for galaxy identification in large surveys.

4.1 Introduction

Galaxy-galaxy mergers are of fundamental importance to our current understanding of how galaxies form and evolve in cold dark matter cosmology (e.g. Conselice 2014). Dark matter halos and their baryonic counterparts merge under hierarchical growth to form the universe that we see today (Somerville & Davé 2015). Mergers play an important role in many aspects of galaxy evolution such as galaxy mass assembly, morphological transformation and growth of the central black hole (e.g. Johnston et al. 1996; Naab & Burkert 2003; Hopkins et al. 2006; Bell et al. 2008; Guo & White 2008; Genel et al. 2009). In addition, galaxy mergers are believed to be the triggering mechanism of some of the brightest infrared objects known: (ultra) luminous infrared galaxies (Sanders & Mirabel 1996). With bright infrared emission often comes high star formation rates (SFRs), hence a prevailing interpretation from early merger works is that most mergers go through a starburst phase (e.g. Joseph & Wright 1985; Schweizer 2005).

Recent studies are beginning to dismantle the claim that all galaxy mergers are starbursts. In a study of 1500 galaxies, within 45 Mpc of our own, Knapen et al. (2015) have found that the increase in SFR in merging galaxies is at most a factor of two, with the majority of galaxies showing no evidence of an increase in SFR, or even showing evidence of mergers quenching the star formation. Galaxy mergers do still cause starbursts and a higher fraction of starbursts are mergers than starbursting non-mergers (Luo et al. 2014; Knapen & Cisternas 2015; Cortijo-Ferrero et al. 2017). Claims about the importance of mergers depend critically on our ability to recognise galaxy interactions. A method to reliably identify complete merger samples among a large number of galaxies is clearly needed.

Existing automated techniques for detecting mergers include selecting close galaxy pairs or selecting morphologically disturbed galaxies. The close pair method finds pairs of galaxies that are close, both on the sky and in redshift (e.g. Barton et al. 2000; Patton et al. 2002; Lambas et al. 2003; Lin et al. 2004; De Propriis et al. 2005). This method requires highly complete, spectroscopic observations and, as a result, is observationally expensive. It can also be contaminated by flybys (Sinha & Holley-Bockelmann 2012; Lang et al. 2014). Selecting the morphologically disturbed galaxies using quantitative measurements of non-parametric morphological statistics, such as the Gini coefficient, the second-order moment of the brightest 20 percent of the light (Lotz et al. 2004) and the CAS system (e.g. Bershadsky et al. 2000; Conselice et al. 2000; Wu et al. 2001; Conselice et al. 2003), aims to detect disturbances such as strong asymmetries, double nuclei or tidal tails. This method relies on high-quality, high-resolution imaging to detect these features beyond the local universe and has a high percentage of misclassifications ($> 20\%$), especially at high redshift (Huertas-Company et al. 2015). There is also the option to classify galaxies through visual inspection. However, visual classifications are hard to reproduce and are time consuming. Large crowd sourced methods, such as Galaxy Zoo¹ (Lintott et al. 2008), are not scalable to the sizes of the data sets expected from upcoming surveys. Visual identification can also suffer from low accuracy and incompleteness (Huertas-Company et al. 2015).

Deep learning techniques have the potential to revolutionise galaxy classification. Once properly trained, the neural networks used in deep learning can classify thousands

¹<http://www.galaxyzoo.org/>

of galaxies in a fraction of the time it would take a human, or team of humans, to classify the same objects. The use of deep learning for galaxy classification was brought to wider attention after Galaxy Zoo lead a competition on the Kaggle platform², known as The Galaxy Challenge, to develop a machine learning algorithm to replicate the human classification of the Sloan Digital Sky Survey (SDSS; York et al. 2000) images. This competition was won by Dieleman et al. (2015) using a deep neural network, the architecture of which has formed the base for subsequent deep learning algorithms for galaxy classification (e.g. Huertas-Company et al. 2015; Petrillo et al. 2017). More recently, deep learning has been applied to SDSS images from Galaxy Zoo to classify objects as merging or non-merging systems using transfer learning, that is taking a pre-trained network and retraining the output layer to classify images into a different set of classifications (Ackermann et al. 2018). There has also been work using deep learning techniques to identifying mergers and tidal features in optical data from the Canada-France-Hawaii Telescope Legacy Survey (Gwyn 2012; Walmsley et al. 2019). These techniques will have an important application in classifying galaxies in large, upcoming surveys, such as the Large Synoptic Survey Telescope (LSST; LSST Science Collaboration et al. 2009) or *Euclid* (Laureijs et al. 2011).

In this work, we aim to develop a neural network architecture and independently train it with two different training sets. This will result in a trained neural network that can identify visually classified mergers from the SDSS data as well as one that can identify physically classified mergers from the Evolution and Assembly of GaLaxies and their Environments (EAGLE) hydrodynamical cosmological simulation (Schaye et al. 2015). Once trained, the networks will be cross applied: SDSS images through the EAGLE trained network and images of simulated galaxies from EAGLE through the SDSS trained network. Visually identified merger catalogues constructed from surveys, such as the SDSS, are biased towards mergers that produce conspicuous features but cosmological simulations include a wide variety of merging galaxies with different mass ratios, gas fractions, environments, orbital parameters etc. Therefore, through training our neural network separately with visual classifications of real observations, physical classifications in simulations and the cross-applications of the two, we can better understand any potential biases in observations and identify problems in simulations in terms of reproducing realistic merger properties.

The paper is structured as follows: Sect. 4.2 describes the data sets used, Sect. 4.3 covers the neural networks, Sect. 4.4 provides the results and discussion and Sect. 4.5 the concluding remarks. Where necessary, Wilkinson Microwave Anisotropy Probe year 7 (WMAP7) cosmology (Komatsu et al. 2011; Larson et al. 2011) is followed, with $\Omega_M = 0.272$, $\Omega_\Lambda = 0.728$ and $H_0 = 70.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

4.2 Image data

4.2.1 SDSS images

To train the neural network, a large number of images of merging and non-merging systems are required. For the training the observational network, we create our merger and

²<https://www.kaggle.com/c/galaxy-zoo-the-galaxy-challenge>

non-merger samples by following Ackermann et al. (2018) and combining the Darg et al. (2010a,b) merger catalogue with non-merging systems. The Darg et al. (2010a,b) catalogue contains 3003 merging systems selected by visually rechecking the visual classifications of all objects from Galaxy Zoo with the fraction of people who classified the object as merging greater than 0.4 and spectroscopic redshifts between 0.005 and 0.1. As a result of this thorough visual classification, the Darg et al. (2010a,b) catalogue is likely to be conservative and mainly contain galaxies with obvious signs of merger, that is two (or more) clearly interacting galaxies or obviously morphologically disturbed systems, and may miss more subtle mergers. The SDSS spectra were only taken for objects with apparent magnitude $r < 17.77$, or absolute magnitude $r < -20.55$ at $z = 0.1$ hence resulting in an effective mass limit of $\approx 10^{10} M_{\odot}$ at $z = 0.1$ (Darg et al. 2010a). For the non-merging systems, we generated a catalogue of all SDSS objects with spectroscopic redshifts in the same range as the Darg et al. (2010a,b) catalogue and the fraction of people who classified the object as merging in Galaxy Zoo less than 0.2 and then randomly selected 3003 of these to form the sample. As we also require spectroscopic redshifts, the non-merger sample will have the same effective mass limit of $\approx 10^{10} M_{\odot}$.

Cut-outs of the merging and non-merging objects were then requested from the SDSS jpeg cut-out server for data release 7³ (DR7) to create 6006 images with the gri bands as the blue, green and red colour channels respectively, each of 256×256 pixels. SDSS images created this way use a modified version of the Lupton et al. (2004) non-linear colour normalisation. These images were then cropped to the centre 64×64 pixels for use to reduce memory requirements while training. Larger image sizes were tested but showed no clear improvements over 64×64 pixel images. Examples of the central 64×64 pixels of merging and non-merging SDSS galaxies are given in Fig. 4.1.

The SFR and stellar mass ($M_{\star}$) for the SDSS objects were gathered from the MPA-JHU catalogue⁴; the $M_{\star}$ were created following the techniques of Kauffmann et al. (2003) and Salim et al. (2007), while the SFR were based on the Brinchmann et al. (2004) catalogue. The redshifts and ugriz magnitudes come from the SDSS DR7. Darg et al. (2010a) find that 54% are major mergers, defined as the ratio of the masses of the merging galaxies is between 1/3 and 3.

4.2.2 EAGLE images

For the simulation network, simulated gri images from EAGLE were used (Crain et al. 2015; McAlpine et al. 2016). These single channel images were generated using the `py-SPHviewer` code (Benítez-Llambay 2017). EAGLE star particles were treated as individual simple stellar populations in the imaging, assuming a Chabrier (2003) initial mass function. The luminosity of particles in each band was obtained using the Bruzual & Charlot (2003) population synthesis model via their zero-age mass, current age and smoothed particle hydrodynamics metallicity. Particle smoothing lengths were calculated based on the 64th nearest neighbour, as described in Trayford et al. (2017). We note that attenuation by dust was not accounted for, with these primary images functioning as maps of pure stellar emissivity.

³<http://cas.sdss.org/dr7/en/tools/chart/default.asp>

⁴<https://wwwmpa.mpa-garching.mpg.de/SDSS/DR7/>

Figure 4.1: Examples of the central 64×64 pixels of SDSS gri, as blue, green and red respectively, galaxy images, corresponding to an angular size of 25.3×25.3 arcsec. The top row shows merging galaxies from the Darg et al. (2010a,b) catalogue while the bottom row shows non-merging galaxies.

To provide enough galaxies to adequately train a neural network, EAGLE galaxies from the simulation snapshots with a redshift of less than 1.0 were used. Objects with $M_\star$ greater than $10^{10} M_\odot$ were selected while the merging partner of the merging systems must be larger than $10^9 M_\odot$. The merging partner must also be more than 10% of the $M_\star$ of the primary galaxy. Galaxies were deemed to have merged when they are tracked as two galaxies in one simulation snapshot and then tracked as one galaxy in the following snapshot in the EAGLE merger trees catalogue (Qu et al. 2017). This prevents the inclusion of chance flybys that may be selected as mergers if the EAGLE galaxies were selected based on proximity. Systems that are projected to merge, using a closing velocity extrapolation, within the next 0.3 Gyr (pre-merger) or are projected to have merged, again using a closing velocity extrapolation of the progenitors, within the last 0.25 Gyr (post-merger) were selected, along with a number of non-merging systems, and gri band images were created of these systems. Springel et al. (2005) have shown that the effects of a merger are visible for approximately 0.25 Gyr after the merger event while the pre-merger stage is much longer. However, we chose to have the pre and post merger period approximately equal as tests conducted with longer pre-merger times showed no improvement, see discussion in Sect. 4.4.2. We note, however, that the merger timing may suffer from imprecision as a result of the coarse time resolution of the EAGLE simulation, that is the time between snapshots, which becomes coarser at lower simulation redshift. Each galaxy was imaged at an assumed distance of 10 Mpc and each image contains all material within 100 kpc of the centre of the target galaxy and is 256×256 pixels, where 256 pixels corresponds to a physical size of 60 kpc. There are 537 pre-merger, 339 post-merger and 335 non-merging systems, each with six random projections to increase the size of the training set. Each of the six projections are treated as individual galaxies resulting in 3222 pre-merger, 2034 post-merger and 2010 non-merging galaxy images for training. The pre-mergers and post-mergers were combined to form the merger class, keeping the

pre-merger image if the same galaxy appears in both sets.

To make the raw EAGLE images look like SDSS images (processed EAGLE images), a number of operations were performed. For each projection of each system, a redshift was randomly chosen from the redshifts of the objects in the Darg et al. (2010a,b) catalogue and the surface brightness of the galaxy was corrected to match this redshift. The image was also re-binned using interpolation with the python `reproject` package (Robitaille 2018) so that the physical resolution of the EAGLE image matches that physical resolution of an SDSS galaxy at the selected redshift. The resulting apparent r-band magnitudes are less than 17.77 for all but 58 of the 10 134 galaxy projections, meaning that the brightness of the simulated galaxies is consistent with the observed SDSS galaxies. Once the surface brightness and physical resolution correction was completed, the observed SDSS point spread functions (PSFs) for the gri bands were created using the stand alone PSF tool⁵ and the simulated images were convolved with these PSFs. The EAGLE galaxies were then injected into real SDSS images to add realistic noise. Finally, red, green, blue (RGB) images were generated from the gri bands using a modified Lupton et al. (2004) non-linear colour normalisation to closely match the way the SDSS RGB images are made. A brief comparison using a simple linear colour scaling to generate the EAGLE images is discussed in Appendix 4.6.1.

To get real SDSS noise, the position of all known SDSS objects from DR7 in three SDSS images were collected. The noise images were generated by offsetting from the position of the objects in these images by a random distance between 6.329 and 18.986 arcsec (that is between 0.25 and 0.75 times the average separation of SDSS objects) and with a random angle in the RA-dec plane. Then 256×256 pixel cut-outs were made, centred on these offset positions, and were used as noise in the EAGLE images. The code used to make the EAGLE images SDSS like and get the noise cut-outs can be downloaded from GitHub⁶ while examples of the raw and processed EAGLE images can be found in Fig. 4.2.

The $M_\star$, SFR, ugriz absolute magnitudes, galaxy asymmetry, merger mass ratio and time to or since the merger event of the EAGLE galaxies are from the simulation. For the merging systems, $M_\star$ and SFR are calculated for the merger remnant. The galaxy asymmetry is the 3D asymmetry and was calculated as described in Trayford et al. (2019). Uniform bins of solid angle were created about the galaxy centre and the $M_\star$ within each bin is summed. The asymmetry is then the sum of the absolute mass difference between diametrically opposed bins divided by the total $M_\star$. Thus, the higher the asymmetry value, the more asymmetric the galaxy is. The minority of the EAGLE galaxies, 34%, are major mergers: the ratio of the masses of the major mergers is between 1/3 and 3.

4.3 Deep learning

4.3.1 Convolutional neural networks

Deep learning neural networks are a type of machine learning that aim to loosely mimic how a biological system processes information by using a series of layers of non-linear

⁵https://www.sdss.org/dr12/algorithms/read_psf/

⁶<https://github.com/wjpearson/SDSS-EAGLE-mergers>

Figure 4.2: Examples of the raw (first and third rows, linear colour scaling) and processed (second and fourth rows, Lupton et al. (2004) non-linear colour scaling) EAGLE images for merging (first and second rows) and non-merging (third and fourth rows) systems. The raw images shown are 128×128 pixels and imaged at 10 Mpc, corresponding to a *physical* size of 30×30 kpc or an *angular* size of 621×621 arcsec, while the processed images are 64×64 pixel images corresponding to an *angular* size of 25.3×25.3 arcsec. The redshifts are those that the EAGLE images have been projected to.

mathematic operations, known as neurons, each with its own weight and bias value. Here we use a type of deep learning known as convolutional neural networks (CNN). The lower layers of a CNN are known as convolutional layers and contain a defined number of kernels that are convolved with the output of the layer below. The kernels are groups of neurons that have a shape smaller in width and height than the input so must move across the input, performing the necessary matrix multiplications on the region of the input they cover at each step. The higher layers of a CNN are one-dimensional and fully connected, that is all the neurons are connected to all the neurons in the layer below. Dimensionality of the network can be reduced by applying pooling layers. This type of layer groups inputs into it and passes on the maximum or average value of each group to the next layer. When placed after an image input or a convolution layer, this grouping is done in the 2-dimensional height-width plane and not the depth/colour direction resulting in a reduction in the spatial size but not the depth. Each neuron in the kernels or fully connected layers of a network has an activation function to scale the result to pass to higher layers or force the output to a certain value, depending on the value passed into the activation function and the activation function used. The weights and biases in the neurons are trained by passing a large number of classified images through the network, in the case of supervised learning used here, such that the output classifications converge on the known input classifications. A thorough description of how CNNs work is beyond the scope of this paper; further information on CNNs can be found in Lecun et al. (1998).

When discussing neural networks, some terms are used whose definitions may differ from what is expected or be unfamiliar. Also, concepts have a number of different names. To prevent confusion, terms used in this paper are defined in Table 4.1, taking a positive result to mean a merger and a negative result to mean a non-merger.

4.3.2 Architecture

The CNN used in this work was built using the Tensorflow framework (Abadi et al. 2015). As the task we are attempting to complete is similar to that of The Galaxy Challenge, we based our network on the winning Dieleman et al. (2015) architecture but apply some tweaks. The input image was 64 by 64 pixels with three colour channels. We then applied a series of four, two dimensional convolutional layers with 32, 64, 128 and 128 kernels of 6×6 , 5×5 , 3×3 and 3×3 pixels for the first, second, third and fourth layers respectively. The strides of the kernels, how far the kernel is moved as it scans the input, was set at 1 pixel for all layers and the zero padding was set to ‘same’ to pad each edge of the image with zeros evenly (if required). Batch normalisation (Ioffe & Szegedy 2015) was applied after each layer, scaling the output between zero and one, and we used rectified linear units (ReLU; Nair & Hinton 2010) for activation. ReLU returns $\max(x, 0)$ when passed x . Dropout (Srivastava et al. 2014) was also applied after each activation, to help reduce overfitting, with a dropout rate of 0.2, randomly setting the output of neurons to zero 20% of the time during training. The output from the first, second and fourth convolutional layers had a 2×2 pixel max-pooling applied to reduce dimensionality. After the fourth convolutional layer, we used two one-dimensional, fully connected layers of 2048 neurons, again applying ReLU activation, batch normalisation and dropout. The output layer has two neurons⁷, one for each class, and uses a softmax output. For training,

⁷It is possible to do this with a single output but this setup makes it easier to add more classes in the future.

Table 4.1: Terms used when describing the performance of neural networks

Term	Definition	
Positive (P)	An object classified in the catalogues or identified by a network as a merger.	
Negative (N)	An object classified in the catalogues or identified by a network as a non-merger.	
True positive (TP)	An object classified in the catalogues as a merger that is identified by a network as a merger.	
False positive (FP)	An object classified in the catalogues as a non-merger that is identified by a network as a merger.	
True negative (TN)	An object classified in the catalogues as a non-merger that is identified by a network as a non-merger.	
False negative (FN)	An object classified in the catalogues as a merger that is identified by a network as a non-merger.	
Recall	Fraction of objects correctly identified by a network as a merger with respect to the total number of objects classified in the catalogues as mergers.	$TP / (TP+FN)$
Fall-out	Fraction of objects incorrectly identified by a network as a merger with respect to the total number of objects classified in the catalogues as mergers.	$FP / (TP+FN)$
Specificity	Fraction of objects correctly identified by a network as a non-merger with respect to the total number of objects classified in the catalogues as non-mergers.	$TN / (TN+FP)$
Precision	Fraction of objects correctly identified by a network as a merger with respect to the total number of objects identified by a network as a merger.	$TP / (TP+FP)$
Negative predictive value (NPV)	Fraction of objects correctly identified by a network as a non-merger with respect to the total number of objects identified by a network as a non-merger.	$TN / (TN+FN)$
Accuracy	Fraction of objects, both merger and non-merger, correctly identified by a network.	$(TP+TN) / (TP+FP+TN+FN)$

validation and testing samples with an equal number of each class, as done here, softmax output provides probabilities for each class, in the interval $[0, 1]$, that sum to one, that is softmax maps the un-normalised input into it into a probability distribution over the output classes. Thus there is one output that can be considered the probability the input image is of a merging system and one output that can be considered to be the probability the input image is of a non-merging galaxy. In this paper, we will use the output for the merger class, although with our binary classification the non-merger class can be considered equivalent as it is $1 - (\text{merger class output})$. The full network can be seen in Table 4.2. Loss of the network was determined using softmax cross entropy and was optimised using the Adam algorithm (Kingma & Ba 2015). A learning rate, that is how fast the weights and biases in the network can change, of 5×10^{-5} was used as it resulted in a more accurate network.

Table 4.2: Architecture of the CNN. The first column in the type of layer while the second column contains the associated properties. The input is a 64×64 pixel, three channel image and the output is two probabilities, one for the probability the input is a merger and one for the probability the input is a non-merger. Further details on what the properties of the layers mean can be found in Sect. 4.3.2.

Layer	Properties			
Input	64×64 pixels	3 channels		
Convolutional	32, 6×6 pixel kernels	1 pixel stride		
Dropout	Dropout rate of 0.2		same padding	ReLU activation
MaxPooling	2×2 pixel	2 pixel stride		
Convolutional	64, 5×5 pixel kernels	1 pixel stride		
Dropout	Dropout rate of 0.2		same padding	ReLU activation
MaxPooling	2×2 pixel	2 pixel stride		
Convolutional	128, 3×3 pixel kernels	1 pixel stride		
Dropout	Dropout rate of 0.2		same padding	ReLU activation
Convolutional	128, 3×3 pixel kernels	1 pixel stride		
Dropout	Dropout rate of 0.2		same padding	ReLU activation
MaxPooling	2×2 pixel	2 pixel stride		
Flatten				
Fully Connected	2048 neurons	Batch normalisation	ReLU activation	
Dropout	Dropout rate of 0.2			
Fully Connected	2048 neurons	Batch normalisation	ReLU activation	
Dropout	Dropout rate of 0.2			
Output	2 neurons	Softmax activation		

4.3.3 Training, validation and testing

If there were an unequal number of images in the two classes, the larger class size was reduced by randomly removing images until the classes are the same size. The images were then subdivided into three groups: 80% were used for training, 10% for validation and 10% for testing. The training set was the set used to train the network while the validation was used to see how well the network was performing as training progressed. Each network was trained for 200 epochs, an epoch is showing each image to the network once, and the epoch with simultaneously the highest accuracy and lowest loss with the validation set was selected for use. Using 200 epochs was long enough as by this point the loss for the validation set had begun to increase as the network starts to over-train and learn the training set, not the features in the training set. The testing set was used once, and once only, to test the performance of the network deemed to be the best from the validation. Testing images were not used for validation to prevent accidental training on the test data set. To reduce sensitivity to galaxy orientation, the images were also augmented as they were loaded for training (and only training): the images were randomly rotated by 0° , 90° , 180° or 270° . We also cropped the images to the centre 64×64 pixels and scale the images globally between zero and one, preserving the relative flux densities. The code used to create, train, validate and test the networks can be downloaded from GitHub⁸.

4.4 Results and discussion

We used the receiver operating characteristic (ROC) curve to determine how well the network has performed for a binary classification. The ROC curve is a plot of the recall against fall-out (see Table 4.1 for definitions) with each point along the curve corresponding to a different value for the output (threshold) above which an input image is considered to be of a merging system. Higher recall and lower fall-out means a better threshold while the (0,0) and (1,1) positions correspond to assigning all objects to the non-merger and merger classes respectively. The threshold with recall and fall-out closest to the (1,0) position, calculated as least squared difference, is the preferred threshold for splitting mergers from non-mergers. Also, the area under the ROC curve is unity for an infallible network, and close to unity for good networks, while a truly random network will have an area of 0.5.

The two-sample Kolmogorov-Smirnov test (KS-test; Smirnov 1939) is also used to compare the distributions of correctly and incorrectly identified objects to see if they are likely sampled from the same distribution. The null hypothesis that the two distributions are the same is rejected at level $\alpha = 0.05$ if the KS-test statistic, $D_{N,M}$, is greater than $\text{Crit}_{N,M} = c(\alpha) \sqrt{\frac{n+m}{nm}}$, where $c(\alpha) = 1.224$ for $\alpha = 0.05$ and n and m are the sizes of samples N and M .

⁸<https://github.com/wjpearson/SDSS-EAGLE-mergers>

Figure 4.3: ROC curve for the observation network used on visually classified SDSS images at validation (blue) and testing (yellow). The area under each curve is 0.966. The dashed red line shows the position of a truly random network.

4.4.1 Observation trained network

The 97th epoch of the network trained with SDSS images (observation network) was used. This epoch has an accuracy (see Table 4.1 for definition) at validation of 0.932 cutting at a threshold of 0.5 to separate mergers from non-merger classification. Using the validation set, we plot the ROC curve for this network in blue in Fig. 4.3. This has an area of 0.966 and provides an ideal cut threshold of 0.57. At this threshold, the accuracy of the validation set increases to 0.935. To determine the true accuracy of the network, we perform the same analysis for the test data set. The area under the ROC curve, shown in Fig. 4.3 in yellow, remains constant at 0.966. With the threshold set at 0.57, the final accuracy of the network is 0.915, with recall, precision, specificity and NPV of 0.920, 0.911, 0.910 and 0.919 respectively (see Table 4.1 for definitions). It is possible to increase the accuracy, and other cut dependent statistics, by changing the cut threshold for the training set. However, this risks accidentally using the test set for training and thus not giving a true representation of the network.

Our results can be compared to those of Ackermann et al. (2018), who performed a similar study using the same Darg et al. (2010a,b) merger catalogue. Ackermann et al. (2018) have a recall of 0.96, a precision of 0.97 and the area under the ROC curve is 0.9922. All these values are slightly larger than those we find, demonstrating that their network performs somewhat better. However, there are some differences between the two studies. The architecture of the CNN used here is different from that used by Ackermann et al. (2018), who use the Xception architecture (Chollet 2017), and they perform transfer learning: using a network pre-trained on the non-astronomical ImageNet images (Deng et al. 2009) and then continuing to train using the merger and non-merger images. The non-merger set is also different, with Ackermann et al. (2018) using 10 000 non-merging

Table 4.3: Confusion matrix for SDSS images classified by the observation network.

		Network Classification		Total
		Merger	Non-merger	
Catalogue Classification	Merger	276 TP	24 FN	300
	Non-merger	27 FP	273 TN	300
Total		303	297	

galaxies as opposed to our 3003. We use an equal number of mergers and non-mergers to prevent accidental bias against the class with fewer images. Finally, the Ackermann et al. (2018) study does not change the cut threshold from 0.5 to improve the recall or precision, suggesting that these values may be able to be improved.

Another study, by Walmsley et al. (2019), trains a CNN on data from the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS; Gwyn 2012). Here, they aim to identify galaxies with tidal features, which are likely due to galaxy interactions. The performance of our SDSS trained network is much better than that of the CFHTLS network: we achieve recall of 0.920 while Walmsley et al. (2019) achieve 0.760. However, the differences in data set and network architecture will have an effect on the results.

To determine if certain physical properties of the galaxies are the cause of the misclassification, the specific SFR ($\text{SFR}/M_\star$, $s\text{SFR}$), $M_\star$, redshift and ugriz band magnitudes of the misclassified objects have been compared to their correctly classified counterparts. This will allow us to determine if, for example, all of the high mass, non-mergers have been classified as mergers. There are no trends in any of these properties: the distribution of the misclassified objects is the same as the distribution for the correctly classified objects. The confusion matrix, showing the number of TP, FP, TN and FN, for the SDSS images classified by the observation network can be found in Table 4.3 while the KS-test statistics comparing the distributions of correctly and incorrectly identified galaxies with the physical properties can be found in Table 4.4. See Table 4.1 for definitions of TP, FP, TN and FN.

The images of the misclassified objects have also been visually inspected. Over half of the FP objects (16 of 27) have a close chance projection or a second galaxy projected into the disc of the primary galaxy, possibly fooling the network into believing that the two galaxies are merging. Four further galaxies fill the entire 64×64 pixel image, two of which also have a chance projection of a second galaxy into the disc of the primary galaxy. For six of the FP, there is no clear reason why they are misclassified: they appear to be isolated galaxies without signs of morphological disturbance. The final FP is a large

Table 4.4: KS-test statistic, $D_{N,M}$, and the critical value, $\text{Crit}_{N,M} = c(\alpha) \sqrt{\frac{n+m}{nm}}$, for the SDSS images classified by the observation network. If $D_{N,M} > \text{Crit}_{N,M}$, the null hypothesis that the two distributions are the same is rejected at level $\alpha = 0.05$. Here, $c(\alpha) = 1.224$ for $\alpha = 0.05$ and n and m are the sizes of samples N and M .

Physical Parameter	$D_{TP, FN}$	$\text{Crit}_{TP, FN}$	$D_{TN, FP}$	$\text{Crit}_{TN, FP}$
$M_{\star}$	0.144	0.261	0.091	0.247
sSFR	0.203	0.261	0.141	0.247
u-magnitude	0.190	0.260	0.195	0.247
g-magnitude	0.324	0.260	0.168	0.247
r-magnitude	0.236	0.260	0.178	0.247
i-magnitude	0.196	0.260	0.193	0.247
z-magnitude	0.199	0.260	0.197	0.247

grand design spiral that has been identified off centre in the original 256×256 pixel image. When cropped, the image contains only the arms of the spiral that appear like a disturbed system. Examples of the FP are shown in Fig. 4.4a-d. However, and unsurprisingly with so few misclassified objects, galaxies that are visually similar to the FP have also been correctly identified, as seen in Fig. 4.4e-h.

For the FN objects, six of the 24 have a merging companion that is either outside the 64×64 pixel image or on the very edge, indicating that a larger image may reduce the FP rate. The remaining images show a clear morphological disturbance or a clear merger companion. It is possible that these companions are being identified by the network as chance projections, especially the companions that are almost point-like in the image. Examples of the FN are shown in Fig. 4.5a-d. As with the FP objects, there are also example TP that are visually similar to the FN galaxies, presented in Fig. 4.5e-h.

4.4.2 Simulation trained network

The 26th epoch of the network trained with EAGLE images (simulation network) was used. This epoch has an accuracy of 0.672 at validation, cutting at a threshold of 0.5 to separate mergers from non-merger classification. Using the validation set, we plot the ROC curve for this network in dot-dashed yellow in Fig. 4.6. This has an area of 0.710 and provides an ideal cut threshold of 0.46. At this threshold, the accuracy of the validation set decreases to 0.644. To determine the true accuracy of the network, we perform the same analysis for the test data set. The area under the ROC curve, the dot-dashed orange curve in Fig. 4.6, increases to 0.726. With the threshold set at 0.46, the final accuracy of the network is 0.674, with recall, precision, specificity and NPV of 0.657, 0.680, 0.692 and 0.668 respectively.

The lower accuracy of the simulation trained network relative to the observation trained network (discussed in Sect. 4.4.1) is a result of the difference in the training sample. The SDSS merger sample has been thoroughly checked to verify there are visible indications of a merger, as can be seen in the examples in Fig. 4.1. The EAGLE

Figure 4.4: Examples of FP galaxies from the observation network for (a) a chance projection, (b) a galaxy filling the image, (c) a galaxy filling the image with a chance projection and (d) an isolated, non-interacting galaxy. Panels (e) to (h) show TN galaxies that are visually similar to those shown in (a) to (d).

Figure 4.5: Examples of FN galaxies from the observation network for (a) a galaxy with its merging companion outside the image, (b) a galaxy with its merging companion on the edge of the image, (c) a merging system and (d) a merging system where the minor galaxy is almost point-like. Panels (e) to (h) show TP galaxies that are visually similar to those shown in (a) to (d).

Figure 4.6: ROC curve for the simulation networks at validation (purple, light blue, yellow) and testing (dark blue, green, orange) for 100 Myr (solid), 200 Myr (dashed) and 300 Myr (dot-dashed) from the merger event. The areas under the curves can be found in Table 4.5. The dashed red line shows the position of a truly random network.

merger sample, however, contains physically classified mergers in the simulation without visual inspection to check whether there are any obvious signs of merging. As such, the EAGLE merger sample includes a wide variety of merger types (in terms of their mass ratios, orbital parameters, gas fractions, etc.) and hence some of the mergers are bound to have inconspicuous merging signs and will therefore be harder to discern, resulting in a lower accuracy. We have checked the merging galaxies misclassified by the network visually and confirm that most EAGLE mergers are indeed not as conspicuous as the ones in the SDSS catalogue, where the mergers from Darg et al. (2010a,b) have been selected to be conspicuous.

If the time before and after the merger event is decreased, the accuracy of the network increases. Performing the same analysis as with the full EAGLE data set, we find that using galaxies that are within 200 Myr of the merger event results in a network that has a test accuracy of 0.644 at a cut threshold of 0.40 in the 31st epoch. Using galaxies that are within 100 Myr of the merger event the network as a test accuracy of 0.652 at a cut threshold of 0.39 in the 52nd epoch. The full statistics for these two networks can be found in Table 4.5 and the ROC curves can be found in Fig. 4.6. As the 100 Myr network has the largest area under the ROC curve, the majority of the remainder of the paper will now focus on the 100 Myr network when discussing the simulation network. The confusion matrix for the 100 Myr network can be found in Table 4.6.

A similar study has been performed by Snyder et al. (2018) using simulated galaxy images from the Illustris simulation (Vogelsberger et al. 2014), although their technique is somewhat different. In their study, Snyder et al. (2018) train Random Forests using non-parametric morphology statistics, such as concentration, asymmetry, Gini and M_{20} ,

Table 4.5: Statistics for the SDSS and the 100 Myr, 200 Myr and 300 Myr EAGLE trained networks at testing.

	SDSS	100 Myr	200 Myr	300 Myr
Epoch used	97	52	31	26
Cut threshold	0.57	0.39	0.40	0.46
ROC area	0.966	0.787	0.704	0.726
Recall	0.920	0.632	0.562	0.657
Precision	0.911	0.658	0.673	0.680
Specificity	0.910	0.672	0.726	0.692
NPV	0.919	0.646	0.624	0.668
Accuracy	0.915	0.652	0.644	0.674

Table 4.6: Confusion matrix for EAGLE images classified by the simulation network.

		Network Classification		Total
		Merger	Non-merger	
Catalogue Classification	Merger	127 TP	74 FN	201
	Non-merger	66 FP	135 TN	201
Total		193	209	

as inputs, with these statistics derived from Illustris galaxies processed to look like Hubble Space Telescope images. They select galaxies that will, or have, merge within 250 Gyr. The recall of Snyder et al. (2018) is slightly higher than this work, they achieve ≈ 0.70 compared to our 0.632, but their precision is much lower, at ≈ 0.30 compared to 0.658. Comparing the Snyder et al. (2018) results to our 300 Myr trained network, a more fair comparison, shows similar results: Snyder et al. (2018) has higher recall, ≈ 0.70 compared to 0.657, but lower precision, ≈ 0.30 compared to 0.680.

As the galaxies are generated from a simulation, we know the physical properties of these systems. As with the SDSS objects, we can compare the physical properties of the galaxies that are correctly and incorrectly identified. KS-test statistics comparing the distributions of correctly and incorrectly identified galaxies with the physical properties can be found in Table 4.7.

Many FN objects appear to have low simulation snapshot redshifts when compared to the TP, see Fig. 4.7a, potentially a result of coarser time resolution of the simulation at low redshift. We note that the simulation snapshot redshift is different from the redshift used when making the EAGLE galaxies look like SDSS images. For the TN and FP populations, higher snapshot redshifts have a higher fraction of FP sources relative to the TN, see Fig. 4.7b. This suggests that simulated non-mergers in the local universe look different from simulated non-mergers in the higher- z universe.

The asymmetry of the non-merger population also has an effect: non-merging objects with higher asymmetry are preferentially being identified as merging systems. It is worth noting that the time to/from the merger event does not appear directly correlated with the asymmetry of the galaxy.

In M_* , there is a slight trend for the low mass, merging systems to be identified as non-mergers, although the non-merging galaxies are typically slightly lower mass than the merging systems so this is not overly unexpected. For sSFR, there is a splitting, with low sSFR merging systems being preferentially assigned the non-merger classification and the high sSFR non-merging galaxies preferentially identified as mergers, as shown in Fig. 4.8.

The apparent magnitude of the simulated galaxy after redshift projection has the largest effect, compared to the other parameters investigated, on the correct identification. For the merging systems, faint objects are preferentially classified as non-merging systems while the bright non-mergers are more likely to be misclassified as mergers. An example of this in the g -band is presented in Fig. 4.9. Misclassification for the merging systems is likely a result of the merging systems being brighter, on average, than the non-merger systems while the majority of the merging systems are fainter, hence the high misclassification rate for non-merging systems at these magnitudes.

As with the SDSS images, the misclassified EAGLE images have also been visually inspected. The majority of the FP galaxies, 43 of 66, contain a chance projection generated when the real SDSS noise is added. Three FP galaxies have a projected galaxy that is much brighter than the EAGLE galaxy, resulting in the EAGLE galaxy becoming extremely faint in the image and (almost) impossible to see by eye. There are also correctly identified galaxies that also suffer from the same image suppression, suggesting that this issue is not the sole cause of the misclassification. Of the remaining FP, three are at a low projection redshift, resulting in the features and inhomogeneities of the galaxy appearing as morphological disturbances, although, again, there are examples of these low

Table 4.7: KS-test statistic, $D_{N,M}$, and the critical value, $\text{Crit}_{N,M} = c(\alpha) \sqrt{\frac{n+m}{nm}}$, for the EAGLE images classified by the simulation network. If $D_{N,M} > \text{Crit}_{N,M}$, the null hypothesis that the two distributions are the same is rejected at level $\alpha = 0.05$. Here, $c(\alpha) = 1.224$ for $\alpha = 0.05$ and n and m are the sizes of samples N and M .

Physical Parameter	$D_{TP, FN}$	$\text{Crit}_{TP, FN}$	$D_{TN, FP}$	$\text{Crit}_{TN, FP}$
Projection	0.102	0.179	0.162	0.184
Redshift				
Simulation	0.272	0.179	0.260	0.184
Redshift				
Asymmetry	0.225	0.179	0.232	0.184
Time since	0.123	0.179	-	-
Merger				
u-magnitude	0.286	0.179	0.120	0.184
g-magnitude	0.259	0.179	0.075	0.184
r-magnitude	0.222	0.179	0.068	0.184
i-magnitude	0.216	0.179	0.065	0.184
z-magnitude	0.201	0.179	0.072	0.184
Mass ratio	0.164	0.179	-	-
$M_\star$	0.225	0.179	0.205	0.184
sSFR	0.338	0.179	0.207	0.191

Figure 4.7: Distributions for the correctly (blue) and incorrectly (orange) identified EAGLE objects in the simulation network for (a) mergers and (b) non-mergers as a function of simulation snapshot redshift. Merging objects with low snapshot redshifts are disproportionately assigned a non-merger classification while non-merging objects with high simulation redshifts are often seen as mergers.

Figure 4.8: Distribution of EAGLE galaxies from the simulation network of the correctly (green) and incorrectly (brown) mergers (a) and non-mergers (b) as a function of EAGLE sSFR. Merging galaxies with low sSFR are often misclassified as non-merging while high sSFR non-mergers are often identified as mergers.

Figure 4.9: Distributions for the correctly (purple) and incorrectly (yellow) identified objects for (a) mergers and (b) non-mergers as a function of g-band magnitude for EAGLE galaxies classified by the simulation network. Faint mergers are preferentially classified as non-mergers while the distribution of misclassified non-mergers is at intermediate magnitudes.

Figure 4.10: Examples of FP EAGLE galaxies from the simulation network for (a) a chance projection, (b) a galaxy where the chance projection from the SDSS noise has resulted in the EAGLE galaxy appearing faint in the image, (c) a galaxy at low projection redshift and (d) an isolated, non-interacting galaxy. Panels (e) to (h) show TN galaxies that are visually similar to those shown in (a) to (d).

projection redshift galaxies that have been correctly identified as non-merging. The other objects show no signs of asymmetry or morphological disturbances. Examples of these galaxies can be found in Fig. 4.10a-d while example TN galaxies that are visually similar can be found in Fig. 4.10e-h.

For the FN objects, nine of the 74 have a bright chance projection from the added SDSS noise that results in the EAGLE galaxy becoming (almost) impossible to see in the image. As these types of images are present in the FN, FP, TP and TN, it is unlikely that the bright counterpart is causing the misclassifications. 21 FNs do not appear morphologically disturbed or asymmetric. This is likely a result of the PSF convolution and redshift re-projection smoothing out the visual merger indicators resulting in what appears to be a single, smooth galaxy. The remaining objects do have clearly identifiable merger counterparts or asymmetry. Examples of these galaxies can be found in Fig. 4.11a-d. As with the EAGLE FP, Fig. 4.11e-h show there are also examples of visually similar galaxies that have been correctly identified by the simulation network.

The CNN architecture was also trained on simulation images that had only been partially processed to look like SDSS images. This will allow us to determine if there is a specific part of the process that results in the lower accuracy for the simulation network with respect to the observation network. For this, we use EAGLE galaxies that are within 100 Myr of the merger event and perform one of the following processes: convolve the EAGLE image with the SDSS PSF (C), inject the EAGLE image into the real SDSS noise (N), match the EAGLE resolution to that of the SDSS images (R), adjust the EAGLE magnitude to be the correct apparent magnitude for a chosen redshift (Z) or a combination of three (CNR, CNZ, CRZ, NRZ). We also train the network on the EAGLE images that have not been processed. As with training with SDSS or fully processed EAGLE

Figure 4.11: Examples of FN EAGLE galaxies from the simulation network for (a) a galaxy where the chance projection from the SDSS noise has resulted in the EAGLE galaxy appearing faint in the image, (b) a merging system that appears as a single, smooth galaxy, (c) a galaxy with a clearly identifiable counterpart and (d) as asymmetric galaxy. Panels (e) to (h) show TP galaxies that are visually similar to those shown in (a) to (d).

images, the epoch with simultaneously the lowest validation loss and accuracy is chosen and the cut threshold with fall-out and recall closet to (0,1) is used. The statistics are then calculated for the test set and are presented in Table 4.8. Individually, C and N do not notably change the accuracy of the trained network, remaining within one percentage point of the accuracy of the un-processed EAGLE images (87%). R and Z both cause a small reduction in the accuracy, reducing it to 82%, along with the combination of CNR. CNZ, CRZ and NRZ all show more significant reductions. Of these three, CRZ has the smallest effect, reducing accuracy to 78%, while the combinations CNZ and NRZ reduce the accuracy to 72% and 68% respectively. This suggests that correcting the apparent magnitude has the largest affect, especially when coupled with the inclusion of back ground noise. This is possibly because when changing from absolute to apparent magnitude, the fainter objects are becoming harder to discern from the background when injected into the real SDSS noise. We note, however, that only 58 of the original 10 134 processed EAGLE images have an apparent r-band magnitude greater than the limit applied to the SDSS images.

4.4.3 Cross application of the networks

Here we passed the images through the other network, that is we passed all 6006 SDSS images through the simulation network and all 4020 EAGLE images through the observation network. For this, we used the same cut threshold as for passing through the ‘correct’ images. This was done so that we can understand any biases and incompleteness in the two data sets. For example, the visually classified mergers from the SDSS data consist only of certain types of mergers with conspicuous merging signs, such as two massive

Table 4.8: Statistics for the network trained with partially processed EAGLE images at testing. C is convolving the EAGLE image with the SDSS PSF, N is injecting the EAGLE image into the real SDSS noise, R is matching the EAGLE resolution to that of SDSS and Z is changing the EAGLE magnitude to apparent from absolute.

Processing	None	C	N	R	Z	CNR	CNZ	CRZ	NRZ
Epoch used	183	167	191	135	116	167	43	145	36
Cut threshold	0.47	0.48	0.42	0.45	0.47	0.48	0.40	0.50	0.43
ROC area	0.943	0.948	0.933	0.903	0.893	0.911	0.803	0.866	0.763
Recall	0.841	0.841	0.866	0.786	0.791	0.776	0.692	0.667	0.657
Precision	0.904	0.889	0.888	0.845	0.846	0.867	0.735	0.859	0.688
Specificity	0.910	0.896	0.891	0.856	0.856	0.881	0.751	0.889	0.701
NPV	0.851	0.849	0.869	0.800	0.804	0.797	0.709	0.725	0.671
Accuracy	0.876	0.868	0.866	0.821	0.823	0.828	0.721	0.778	0.679

galaxies obviously interacting with strong tidal features. However, the EAGLE simulation contains a much more complete merger sample. So, one would expect the neural network trained with the visually classified SDSS merger sample to perform poorly on simulated images of EAGLE mergers. We also perform the cross application so that any SDSS objects classified as merging systems in the visual classification but not identified by the simulation trained network can be identified and help improve our understanding of the limitations of simulations so that they become more representative of the real universe in future developments.

EAGLE images through the observation network

Passing all the EAGLE images through the observation network resulted in an accuracy of 0.530, only slightly better than random assignment of objects at first glance. Precision and NPV are similarly close to random at 0.541 and 0.523. However, recall is low, at 0.387, and the specificity is high, at 0.673, demonstrating that the network preferentially assigns objects to the non-merger class but with each class containing just over half correctly identified objects. As to be expected, the area under the ROC curve is close to 0.5 at 0.502, depicted in yellow in Fig. 4.12. The confusion matrix can be found in Table 4.9.

As before, the physical properties of the EAGLE images can be examined to determine if they are affecting the classification by the network, a brief summary of which can be seen in the KS-test results in Table 4.10. One property that has an obvious splitting between correct and incorrect assignment is the redshift of projection. As is evident in Fig. 4.13, objects with high projection redshifts are preferentially being classified as non-merger systems, see Fig. 4.13a, while objects with low projected redshifts are classified as merging systems, see Fig. 4.13b. The distribution of redshifts used to re-project the EAGLE galaxies is nearly identical to the SDSS distribution: the redshifts used to re-project the galaxies were drawn randomly from the redshifts of the SDSS observations. Thus this effect is not a result of a mismatch in the redshift distributions between observations and simulations. The issue of misclassified mergers at high redshift may arise

Figure 4.12: ROC curve for the SDSS images classified by the simulation network (blue) and the EAGLE images classified by the observation network (yellow). The area under the EAGLE through observation network is 0.502 while the area under the SDSS through simulation network is 0.689. The dashed red line shows the position of a truly random network.

Table 4.9: Confusion matrix for EAGLE images classified by the observation network.

		Network Classification		Total
		Merger	Non-merger	
Catalogue Classification	Merger	777 TP	1352 FN	2010
	Non-merger	658 FP	1233 TN	2010
Total		1435	2585	

Table 4.10: KS-test statistic, $D_{N,M}$, and the critical value, $\text{Crit}_{N,M} = c(\alpha) \sqrt{\frac{n+m}{nm}}$, for the EAGLE images classified by the observation network. If $D_{N,M} > \text{Crit}_{N,M}$, the null hypothesis that the two distributions are the same is rejected at level $\alpha = 0.05$. Here, $c(\alpha) = 1.224$ for $\alpha = 0.05$ and n and m are the sizes of samples N and M .

Physical Parameter	$D_{TP, FN}$	$\text{Crit}_{TP, FN}$	$D_{TN, FP}$	$\text{Crit}_{TN, FP}$
Projection	0.306	0.056	0.5812	0.058
Redshift				
Simulation	0.100	0.056	0.086	0.058
Redshift				
Asymmetry	0.116	0.056	0.050	0.058
Time since	0.152	0.056	-	-
Merger				
u-magnitude	0.412	0.056	0.562	0.058
g-magnitude	0.436	0.056	0.605	0.058
r-magnitude	0.416	0.056	0.612	0.058
i-magnitude	0.403	0.056	0.612	0.058
z-magnitude	0.394	0.056	0.607	0.058
Mass ratio	0.074	0.056	-	-
$M_{\star}$	0.199	0.056	0.127	0.058
sSFR	0.110	0.056	0.049	0.060

while matching the physical resolution (that is kpc per pixel) of the EAGLE images to the SDSS images. At high redshift, this could result in a loss of finer detail that would be expected in merging systems, resulting in these systems being classified as non-mergers. The main misclassification of non-merging systems happens at low projection redshift. The physical resolution of EAGLE images matches the physical resolution of the SDSS images at $z \approx 0.03$. Objects assigned a redshift lower than this value are increased in physical resolution using a bicubic interpolation. This interpolation may result in the creation of artefacts that appear, to the CNN, like features of merging systems. Alternatively, it is possible that at low redshifts the individual particles of the simulation are detectably disturbing the light profile of the galaxies and resulting in misclassification.

There is also a trend with the mass ratio of the merging systems. Although the TP and FN do not split into two distinct distributions, the low mass ratio merger systems, that is major mergers, are more often misclassified as non-merging galaxies. This is the opposite to what would be expected: minor mergers would be expected to be misclassified more often as the disturbances from the smaller galaxy would be expected to be less obvious. Similarly, low mass mergers have a slight preference to be assigned the non-merger class and vice versa, although again this is unsurprising as the merger sample is typically higher mass than the non-merger sample for the EAGLE galaxies. The mass ranges for the EAGLE and SDSS data are not quite comparable: while the SDSS data has an effective mass limit of $10^{10} M_{\odot}$ at $z = 0.1$, there are lower mass galaxies in the sample, while the mass limit for EAGLE is $10^{10} M_{\odot}$ at all projected redshifts. For the sSFR, the high

Figure 4.13: Distributions for the correctly (blue) and incorrectly (orange) identified EAGLE objects for (a) mergers and (b) non-mergers as a function of redshift used for projection after being classified by the observation network. High redshift, merging systems are preferentially classified as non-merging while low redshift non-merging systems are preferentially classified as merging.

sSFR merging galaxies are often misclassified as non-mergers while there is no obvious misclassification for the non-merging galaxies.

As with the EAGLE images through the simulation network, the apparent magnitude of the object has the largest effect on the correct classification. Like the simulation network, the observation network also identifies faint, merging systems as non-merging systems and identifies bright, non-merging systems as merging systems. This is true for all five of the *ugriz* bands. An example in the *g*-band is presented in Fig. 4.14. This is consistent with the results seen with the projected redshift above and is likely a result of the re-projection to the projection redshift.

It is also more likely that the more complete classification for the EAGLE galaxies is causing the low accuracy. The SDSS classifications are for objects that are clearly visually merging systems while the EAGLE classifications will include systems that are not obviously, visually merging. Thus, the observation trained network has not been trained to identify merging systems that are not obviously, visually merging and hence assign these objects the non-merger classification, increasing the number of FN.

Visual inspection of a sub-sample of the FP shows that the majority of these objects ($\approx 64\%$) appear to be isolated, non-interacting systems. A further $\approx 18\%$ of the objects have a close chance projection that may be being mistaken for a merging partner by the CNN. $\approx 8\%$ of the FP objects are galaxies that have been projected into a larger angular size than the original, raw image from EAGLE. This often results in the internal structure of the galaxy being expanded and could appear to the network to be morphological disturbances or multiple galaxies. The remaining objects have a bright chance projection in the SDSS noise and, as a result, are (almost) impossible to see in the image. Examples of these galaxies can be found in Fig. 4.15a-d. With accuracy, specificity and NPV all being almost equivalent to 0.5, it is unsurprising to find examples of visually similar galaxies that have been correctly identified and are shown in Fig. 4.15e-h.

Images of the FN are more useful in understanding why the EAGLE images are poorly classified by the observation network. Inspecting a sub-sample, nearly half ($\approx 46\%$) appear to be a single object. This suggests that these objects have had the visible signatures of merger suppressed while being processed to look like SDSS images, likely by the re-projection and PSF matching, or that these mergers are not obvious, even without the processing steps to make them look like SDSS images. It could also be that the merging companion is hidden behind the galaxy it is merging with, as the angle the galaxy is viewed at is picked randomly, so it cannot be seen within the image. Of the remaining objects, $\approx 41\%$ had at least one counterpart, either from the simulation or random projections from the SDSS noise, that could potentially be merging with the central galaxy and $\approx 5\%$ were unambiguously merging systems. As with the FP, there are a number of images whose simulated galaxies have been suppressed by bright chance projections from the SDSS noise. Example FN galaxies can be found in Fig. 4.16a-d and their visually similar but correctly identified counterparts can be found in Fig. 4.16e-h.

There are limitations of the cross application due to the selection of the EAGLE galaxies. To increase the number of galaxies available to train the simulation network we use merging and non-merging galaxies with simulation redshifts out to $z = 1$ and re-project them to redshifts between 0.005 and 0.1. However, the EAGLE galaxies at $z = 1$ have approximately half the radii of a galaxy at $z = 0$ while the radii of EAGLE galaxies is similar to those observed in the real universe (Furlong et al. 2017). As a result, images of

Figure 4.14: Distributions for the correctly (purple) and incorrectly (yellow) identified EAGLE objects for (a) mergers and (b) non-mergers as a function of g-band magnitude after being classified by the observation network. Faint, merging systems are preferentially classified as non-merging while bright non-merging systems are preferentially classified as merging.

Figure 4.15: Examples of EAGLE FP galaxies (a to d) from the observation network for (a) an isolated, non-interacting galaxy, (b) a chance projection, (c) a galaxy at low projection redshift and (d) a galaxy where the chance projection from the SDSS noise has resulted in the EAGLE galaxy appearing faint in the image. Panels (e) to (h) show TN galaxies that are visually similar to those shown in (a) to (d).

Figure 4.16: Examples of EAGLE FN galaxies (a to d) from the observation network for (a) an apparent single object, (b) a galaxy with a counterpart, either a merger counterpart from EAGLE or a chance projection from the SDSS noise, (c) an unambiguous merger and (d) a galaxy where the chance projection from the SDSS noise has resulted in the EAGLE galaxy appearing faint in the image. Panels (e) to (h) show TP galaxies that are visually similar to those shown in (a) to (d).

Table 4.11: Statistics for the EAGLE images classified by the observation network and the SDSS images classified by the simulation network.

Images	EAGLE	SDSS
Network	Observation	Simulation
Cut threshold	0.57	0.39
ROC area	0.502	0.658
Recall	0.387	0.467
Precision	0.541	0.727
Specificity	0.673	0.825
NPV	0.523	0.608
Accuracy	0.530	0.646

the EAGLE galaxies can contain an object up to two times too small to be comparable in size to the observations, which may hamper the observation network’s ability to correctly identify merging and non-merging systems. If this limitation was reduced, or removed entirely, it is possible that the results of this cross application could improve.

SDSS images through the simulation network

Passing all the SDSS images through the simulation network was more successful than passing all the EAGLE images through the observation network. While still not as good as SDSS images through the observation network, the SDSS images classified by the simulation network had an accuracy of 0.646. Like the EAGLE images through the observation network, the SDSS images through the simulation network have a preference towards the non-merger assignment, demonstrated by a low recall of 0.467 and high specificity of 0.825. The area under the ROC curve is 0.658, see the blue line in Fig. 4.12. The statistics for the cross application of the networks can be found in Table 4.11. The confusion matrix, showing the number of correctly and incorrectly identified objects, can be found in Table 4.12.

As with the SDSS images identified by the observation network, we can examine the estimated physical parameters of the SDSS images that were classified by the simulation network. The KS-test statistics comparing the distributions of correctly and incorrectly identified galaxies with the physical properties can be found in Table 4.13. There is an obvious splitting in the distributions of $M_{\star}$ for correctly and incorrectly identified objects: the high mass merging objects are preferentially assigned the non-merger classification, while the intermediate mass non-merging objects are preferentially assigned the merger classification. Although no low mass mergers being assigned the non-merging class is reassuring as there are no low mass non-merging objects. This splitting may arise from the training sample having non-merging systems as preferentially high mass and merging objects as preferentially intermediate and low mass. A similar, but opposite, split is seen with sSFR: low sSFR mergers are identified as non-mergers, as seen in Fig. 4.17a, while high sSFR non-mergers have a higher misclassification rate than low sSFR non-mergers, as seen in Fig. 4.17b. This suggests that the EAGLE images for merging systems may

Table 4.12: Confusion matrix for SDSS images classified by the simulation network.

		Network Classification		Total
		Merger	Non-merger	
Catalogue Classification	Merger	1403 TP	1600 FN	3003
	Non-merger	526 FP	2477 TN	3003
Total		1929	4077	

preferentially show boosted sSFR.

The trend of the *ugriz* band magnitudes of the SDSS images is also interesting. As the band becomes more red, from *g* through to *z*, the distributions of correctly and incorrectly identified objects become more and more split, as can be seen by the increasing KS-test statistic in Table 4.13. Thus, as the band becomes redder, more and more bright mergers are classified as non-mergers while the faint objects are correctly classified more often. Similarly, as the bands become redder, the distribution of incorrectly identified non-mergers moves to the fainter end. An example of the *z*-band magnitude distribution is shown in Fig. 4.18. The trend that is seen for misclassification in the merging systems is the opposite of the effect seen in the EAGLE test set when classified by the simulation network.

A sub-sample of FP have been visually inspected. $\approx 42\%$ of the FP have at least one other galaxy that lie close to the primary galaxy but are not visually interacting with the primary. These secondary galaxies are likely being identified as a merging companion to the primary or they are possibly merging systems that appear in simulations but are not identified as such in Galaxy Zoo. A further $\approx 43\%$ are unambiguous, non-interacting, isolated galaxies. This is possibly a result of many merging systems in the EAGLE training set visually looking like single, undisturbed galaxies. However, that does not exclude these galaxies from being true mergers as the EAGLE training set should be more complete than the SDSS images. Approximately 8% of objects show signs of asymmetry or morphological disturbances. As with the misidentified chance projections, this may be a result of the strict selection for merging SDSS systems ignoring these galaxies but the more complete selection from EAGLE identifying these as mergers. The remaining galaxies contain a non-physical artefact, typically a single pixel width black line through the galaxy, although there are also a number of TN that also have similar artefacts, so this is unlikely to be causing the misclassification. Example FP galaxies can be found in Fig. 4.19a-d and the visually similar TN in Fig. 4.19e-h.

Alongside the 526 FP, there are 1600 FN of which we visually examine a sub-sample.

Figure 4.17: Distributions for the correctly (green) and incorrectly (brown) identified SDSS objects for (a) mergers and (b) non-mergers as a function of sSFR after being classified by the simulation network. Low sSFR merging systems are preferentially classified as non-merging while high sSFR non-merging systems have a higher misclassification rate than low sSFR non-merger.

Figure 4.18: Distributions for the correctly (purple) and incorrectly (yellow) identified SDSS objects after being classified by the simulation network for (a) mergers and (b) non-mergers as a function of z-band magnitude. Bright mergers are preferentially classified as non-mergers while the distribution of misclassified non-mergers is skewed towards the faint end of the distribution. This trend becomes less pronounced as the bands become more blue, from z to u-band.

Table 4.13: KS-test statistic, $D_{N,M}$, and the critical value, $\text{Crit}_{N,M} = c(\alpha) \sqrt{\frac{n+m}{nm}}$, for the SDSS images classified by the simulation network. If $D_{N,M} > \text{Crit}_{N,M}$, the null hypothesis that the two distributions are the same is rejected at level $\alpha = 0.05$. Here, $c(\alpha) = 1.224$ for $\alpha = 0.05$ and n and m are the sizes of samples N and M .

Physical Parameter	$D_{TP, FN}$	$\text{Crit}_{TP, FN}$	$D_{TN, FP}$	$\text{Crit}_{TN, FP}$
$M_\star$	0.472	0.045	0.417	0.059
sSFR	0.490	0.045	0.410	0.059
u-magnitude	0.186	0.045	0.130	0.059
g-magnitude	0.107	0.045	0.138	0.059
r-magnitude	0.285	0.045	0.268	0.059
i-magnitude	0.329	0.045	0.301	0.059
z-magnitude	0.379	0.045	0.334	0.059

Figure 4.19: Examples of SDSS FP galaxies from the simulation network for (a) a galaxy with a close (in projection) companion, (b) a non-interacting, isolated galaxy, (c) a galaxy showing asymmetry or morphological disturbance and (d) a galaxy with a non-physical artefact within the image. Panels (e) to (h) show TN galaxies that are visually similar to those shown in (a) to (d).

Figure 4.20: Examples of SDSS FN galaxies from the simulation network for (a) a galaxy with a clear merging counterpart, (b) a clearly disturbed system, (c) a galaxy whose merger companion is outside of the 64×64 pixel image and (d) the larger 256×256 pixel image showing the merger companion outside of panel (c). Panels (e) to (h) show TP galaxies that are visually similar to those shown in (a) to (d).

The majority of these systems ($\approx 79\%$) clearly show two interacting galaxies, which may be a result of the network identifying these as chance projections. A further $\approx 7\%$ show clear evidence of morphological disturbances or asymmetry. Approximately 14% of the FN have their counterpart of the edge or outside the image cut-out. This suggests, like the observation network, that a larger cut-out may help identify these objects. Examples of these objects can be found in Fig. 4.20a-d and examples of TP galaxies that look similar in Fig. 4.20e-h.

As mentioned in Sect. 4.4.3, the size of the galaxies from EAGLE may be impacting the results. The re-projection of the high snapshot redshift galaxies to lower redshifts can result in the EAGLE galaxies being too small by up to a factor of two in the most extreme cases (Furlong et al. 2017). As the simulation network is trained on these apparently smaller galaxies, it may have difficulty in correctly identifying the larger SDSS galaxies. As before, if this limitation was reduced or removed, the result may improve.

4.4.4 Differences in network merger identification

To further examine the differences in the observation and simulation networks, we examined the features that each network uses to identify a merging galaxy. A 12×12 pixel area within the images that were correctly identified by both networks was made black by setting the RGB values to zero. This region was moved across the image in steps of one pixel in both x and y directions, generating 2704 images with different 12×12 pixel areas masked. For each masked image, the output for the merger class was recorded. To determine which regions of the image have the greatest effect on the classification, we generated a heat map for each object where each pixel is the average of the merger class

Figure 4.21: Heat maps to demonstrate how the observation (top row) and simulation (bottom row) networks detect an example merging SDSS system. Panels (a) and (d) show the original image of the galaxy being classified, panels (b) and (e) show the regions that most effect the merger classification and panels (c) and (f) show the heat maps where regions with darker colours have a greater affect on the classification (lower merger class output). Panels (b) and (e) are created by stretching the heat map between zero and one and multiplying this with the original image.

outputs of the images where that pixel is masked, as seen in the example in Fig. 4.21c and f.

As can be seen with an example SDSS galaxy in Fig. 4.21, both networks require the secondary galaxy to correctly identify the system as merging. However, as can be seen clearly in Fig 4.21f, the simulation network is also affected by the edges of the primary galaxy, the more diffuse regions, and uses these to help determine the classification. This is likely a result of the EAGLE merging galaxies used to train the simulation network containing systems that are longer from the merger event and so have settled somewhat as well as alignments of the EAGLE image such that the secondary galaxy is in the line of sight with the primary galaxy, requiring closer examination of the primary galaxy to find perturbations that mark them as merger. As a result, the simulation network is more sensitive to smaller changes in the profile of the more diffuse material of a galaxy and more easily fooled into providing an incorrect classification.

A similar distinction between how the observation and simulation networks function is seen when doing a similar examination with the EAGLE galaxies. The simulation network is more easily fooled by masking pixels closer into the nucleus of the primary galaxy than the observation network. However, both networks can be fooled with many galaxy images requiring the background noise to be un-masked to generate a correct classification. This indicates that the EAGLE galaxies are not a true recreation of the SDSS images, helping to partially explain the poor performance of the cross application.

There are also a small minority of galaxies that the observation and simulation networks need the opposite pixels to correctly identify the system as merging. For example, the simulation network may not identify the system as merging if the primary galaxy is masked but the secondary galaxy is visible while the observation network may not identify the system correctly if the secondary is masked but the primary is visible, as shown

Figure 4.22: Heat maps to demonstrate how the observation (top row) and simulation (bottom row) networks detect an example merging EAGLE system. Panels (a) and (d) show the original image of the galaxy being classified, panels (b) and (e) show the regions that most effect the merger classification and panels (c) and (f) show the heat maps where regions with darker colours have a greater affect on the classification (lower merger class output). Panels (b) and (e) are created by stretching the heat map between zero and one and multiplying this with the original image.

in Fig 4.22.

4.5 Conclusions

Training and applying a CNN on SDSS images has been successful, achieving an accuracy of 91.5%. This clearly demonstrates that CNNs can be used to reproduce visual classification. There is no clear indication of a specific type of object that is incorrectly identified from the physical or observable parameters. Training and applying a CNN on the EAGLE images was also somewhat successful, with an accuracy of 65.2% when trained using mergers that will or have occurred within 100 Myr of the image snapshot. Using a longer time between the image snapshot and the merger reduces the accuracy of the network. This relatively lower accuracy suggests that some EAGLE mergers do not have visible merging features that can be picked up by the CNN. The incorrectly identified mergers are primarily at low simulation snapshot redshifts as well as faint apparent magnitude. The combination of real noise added to the EAGLE images and converting the absolute magnitude to apparent magnitude also reduces the effectiveness of the CNN, which demonstrates the importance of image quality (in terms of, for example, signal-to-noise and resolution) in merger identification. Within the image, chance projections result in a large number of non-merging galaxies being identified as mergers.

Examining the features of the merging galaxies that result in correct classifications, we find that the EAGLE trained network is more sensitive to features in the diffuse part of the galaxy, likely tidal structures and disturbances as well as the presence of a close companion. The SDSS trained network, however, primarily focuses only on the presence of a close companion galaxy.

The lower accuracy of the EAGLE trained network is most likely a result of the difference in the training sample. The SDSS merger sample has been selected to contain conspicuous mergers and so the features of a merger are more easily identified but will miss subtler mergers. Meanwhile, the EAGLE sample has fewer conspicuous mergers but should be more complete (including mergers with a wide range of mass ratios, gas fractions, viewing angles, environments, orbital parameters, etc.), resulting in less obvious merger features, in pixel space, that are harder for a CNN to recognise.

Passing the SDSS images through the EAGLE trained network has proven to work, although with only 64.6% accuracy. This relatively low accuracy appears to be a result of high mass or low sSFR objects being identified as non-mergers and low mass or high sSFR objects being identified as mergers. This could suggest that simulations show evidence of high sSFR in the merging systems when this may not necessarily be true. However, the EAGLE trained network may also be identifying merging systems that the visual classification missed. The EAGLE classification will be more complete, as we know which systems are merging, and so the EAGLE trained network may be identifying these objects in the SDSS images that have been missed by the less complete, but more visually obvious, SDSS classification. The result may be a lower specificity, that is a smaller fraction of non-mergers are being correctly identified, when using the SDSS classifications as the truth when in fact the EAGLE trained network is correctly identifying merging system missed by the human visual classification. However, the relatively low recall, the fraction of mergers correctly identified, suggests that EAGLE has relatively few conspicuous mergers.

This has a tantalising prospect for large upcoming surveys, such as LSST and *Euclid*. It is possible to train a CNN with images from simulations and apply it to observations of galaxies from the real universe. Presently, the simulation trained network could be used to generate a set of galaxy merger candidates, which would need to be checked by a human expert, for use in training an observation network. However, with further refinement to the training images from simulations it is not beyond the realm of possibility to reduce the need for an observation training set and apply a simulation trained CNN directly to images from an entire survey, massively speeding up identification.

Passing the EAGLE images through the SDSS trained network was unsuccessful, with the network preferentially assigning objects to the non-merger class. This suggests that some EAGLE mergers are not representative of the SDSS selected mergers, although this appears to be primarily due to how re-projecting the galaxies to their assigned redshift has been done, so it may not be that the EAGLE mergers themselves do not look like observable mergers. The mergers in EAGLE are also less conspicuous than those in the SDSS training set so the observational network has not been trained to identify these less obvious merger events, resulting in a large number of EAGLE mergers being identified as non-mergers.

Improvements for the simulation galaxies in future work would be to increase the mass resolution, which can affect the appearance of galaxies and galaxy mergers (Sparre et al. 2015; Torrey et al. 2015; Trayford et al. 2015; Sparre & Springel 2016), and exactly match the stellar mass distributions with those of observations. Increasing the time resolution, for example by using the snippets⁹ instead of snapshots from EAGLE, should also provide improvement along with improving the estimates of time to or since the merger

⁹High time resolution output from EAGLE.

event by tracing when the central black holes merge. It would also be informative to include the effects of dust attenuation.

As has been shown in this work, chance projections are a major problem in merger classification. In future work, we could train the network to recognise chance projections better by constructing training samples of galaxies which appear close together in the sky but are actually far away from each other. Another area to improve in the future is to come up with a more refined merger classification system, rather than just a binary classification of merger versus non-mergers. For example, have separate classes of early mergers, late mergers, minor mergers, major mergers etc.

Acknowledgements. We would like to thank the anonymous referee for their thoughtful comments that have improved the quality of this paper.

We would like to thank S. Ellison and L. Bignone for helpful discussions that have improved this paper.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is <http://www.sdss.org/>.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

Bibliography

- Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, software available from tensorflow.org
- Ackermann, S., Schawinski, K., Zhang, C., Weigel, A. K., & Turp, M. D. 2018, MNRAS, 479, 415
- Barton, E. J., Geller, M. J., & Kenyon, S. J. 2000, ApJ, 530, 660
- Bell, E. F., Zucker, D. B., Belokurov, V., et al. 2008, ApJ, 680, 295
- Benítez-Llambay, A. 2017, Py-SPHViewer: Cosmological simulations using Smoothed Particle Hydrodynamics, Astrophysics Source Code Library, ascl:1712.003
- Bershady, M. A., Jangren, A., & Conselice, C. J. 2000, AJ, 119, 2645
- Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151
- Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
- Chabrier, G. 2003, PASP, 115, 763
- Chollet, F. 2017, in 2017 IEEE Conference on Computer Vision and Pattern Recognition
- Conselice, C. J. 2014, ARA&A, 52, 291
- Conselice, C. J., Bershady, M. A., Dickinson, M., & Papovich, C. 2003, AJ, 126, 1183
- Conselice, C. J., Bershady, M. A., & Jangren, A. 2000, ApJ, 529, 886
- Cortijo-Ferrero, C., González Delgado, R. M., Pérez, E., et al. 2017, A&A, 607, A70
- Crain, R. A., Schaye, J., Bower, R. G., et al. 2015, MNRAS, 450, 1937
- Darg, D. W., Kaviraj, S., Lintott, C. J., et al. 2010a, MNRAS, 401, 1043
- . 2010b, MNRAS, 401, 1552
- De Propriis, R., Liske, J., Driver, S. P., Allen, P. D., & Cross, N. J. G. 2005, AJ, 130, 1516
- Deng, J., Dong, W., Socher, R., et al. 2009, in 2009 IEEE Conference on Computer Vision and Pattern Recognition
- Dieleman, S., Willett, K. W., & Dambre, J. 2015, MNRAS, 450, 1441

- Furlong, M., Bower, R. G., Crain, R. A., et al. 2017, *MNRAS*, 465, 722
- Genel, S., Genzel, R., Bouché, N., Naab, T., & Sternberg, A. 2009, *ApJ*, 701, 2002
- Guo, Q., & White, S. D. M. 2008, *MNRAS*, 384, 2
- Gwyn, S. D. J. 2012, *AJ*, 143, 38
- Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2006, *ApJS*, 163, 1
- Huertas-Company, M., Gravet, R., Cabrera-Vives, G., et al. 2015, *ApJS*, 221, 8
- Ioffe, S., & Szegedy, C. 2015, International Conference on Machine Learning
- Johnston, K. V., Hernquist, L., & Bolte, M. 1996, *ApJ*, 465, 278
- Joseph, R. D., & Wright, G. S. 1985, *MNRAS*, 214, 87
- Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, *MNRAS*, 341, 33
- Kingma, D. P., & Ba, J. 2015, 3rd International Conference for Learning Representations, arXiv:1412.6980
- Knapen, J. H., & Cisternas, M. 2015, *ApJ*, 807, L16
- Knapen, J. H., Cisternas, M., & Querejeta, M. 2015, *MNRAS*, 454, 1742
- Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, *ApJS*, 192, 18
- Lambas, D. G., Tissera, P. B., Alonso, M. S., & Coldwell, G. 2003, *MNRAS*, 346, 1189
- Lang, M., Holley-Bockelmann, K., & Sinha, M. 2014, *ApJ*, 790, L33
- Larson, D., Dunkley, J., Hinshaw, G., et al. 2011, *ApJS*, 192, 16
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, ArXiv e-prints, arXiv:1110.3193
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998, in *Proceedings of the IEEE*, Vol. 86, Issue 11, 2278–2324
- Lin, L., Koo, D. C., Willmer, C. N. A., et al. 2004, *ApJ*, 617, L9
- Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, *MNRAS*, 389, 1179
- Lotz, J. M., Primack, J., & Madau, P. 2004, *AJ*, 128, 163
- LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, ArXiv e-prints, arXiv:0912.0201
- Luo, W., Yang, X., & Zhang, Y. 2014, *ApJ*, 789, L16
- Lupton, R., Blanton, M. R., Fekete, G., et al. 2004, *PASP*, 116, 133
- McAlpine, S., Helly, J. C., Schaller, M., et al. 2016, *Astronomy and Computing*, 15, 72

- Naab, T., & Burkert, A. 2003, *ApJ*, 597, 893
- Nair, V., & Hinton, G. E. 2010, in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 807–814
- Patton, D. R., Pritchett, C. J., Carlberg, R. G., et al. 2002, *ApJ*, 565, 208
- Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2017, *MNRAS*, 472, 1129
- Qu, Y., Helly, J. C., Bower, R. G., et al. 2017, *MNRAS*, 464, 1659
- Robitaille, T. 2018, *reproject: Image reprojection (resampling)*, <https://reproject.readthedocs.io/en/stable/>, accessed: 2019-02-12
- Salim, S., Rich, R. M., Charlot, S., et al. 2007, *ApJS*, 173, 267
- Sanders, D. B., & Mirabel, I. F. 1996, *ARA&A*, 34, 749
- Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, *MNRAS*, 446, 521
- Schweizer, F. 2005, in *Astrophysics and Space Science Library*, Vol. 329, *Starbursts: From 30 Doradus to Lyman Break Galaxies*, ed. R. de Grijs & R. M. González Delgado, 143
- Sinha, M., & Holley-Bockelmann, K. 2012, *ApJ*, 751, 17
- Smirnov, N. V. 1939, *Bulletin Moscow University*, 2, 3
- Snyder, G. F., Rodriguez-Gomez, V., Lotz, J. M., et al. 2018, *arXiv e-prints*, arXiv:1809.02136
- Somerville, R. S., & Davé, R. 2015, *ARA&A*, 53, 51
- Sparre, M., & Springel, V. 2016, *MNRAS*, 462, 2418
- Sparre, M., Hayward, C. C., Springel, V., et al. 2015, *MNRAS*, 447, 3548
- Springel, V., Di Matteo, T., & Hernquist, L. 2005, *MNRAS*, 361, 776
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2014, *Journal of Machine Learning Research*, 15, 1929
- Torrey, P., Snyder, G. F., Vogelsberger, M., et al. 2015, *MNRAS*, 447, 2753
- Trayford, J. W., Frenk, C. S., Theuns, T., Schaye, J., & Correa, C. 2019, *MNRAS*, 483, 744
- Trayford, J. W., Theuns, T., Bower, R. G., et al. 2015, *MNRAS*, 452, 2879
- Trayford, J. W., Camps, P., Theuns, T., et al. 2017, *MNRAS*, 470, 771
- Vogelsberger, M., Genel, S., Springel, V., et al. 2014, *MNRAS*, 444, 1518
- Walmsley, M., Ferguson, A. M. N., Mann, R. G., & Lintott, C. J. 2019, *MNRAS*, 483, 2968

Wu, K. L., Faber, S. M., & Lauer, T. R. 2001, in *Deep Fields*, ed. S. Cristiani, A. Renzini, & R. E. Williams, 170

York, D. G., Adelmann, J., Anderson, Jr., J. E., et al. 2000, *AJ*, 120, 1579

4.6 Appendix

4.6.1 Image colour scaling

As well as using a modified Lupton et al. (2004) non-linear colour normalisation to create the RGB images of the EAGLE galaxies, we also used a simple linear scaling to generate RGB images, still using gri channels for the blue, green and red colours in the images. The difference in scaling has an effect on a network’s ability to accurately identify merging and non-merging galaxies.

As can be seen by comparing the results in Table 4.5 with those for the linear scaling in Table 4.14, the linear scaling provides a higher accuracy and larger ROC area for the 100 Myr and 200 Myr networks, while the 300 Myr network remain approximately constant. However, the cross application with passing the linear scaled EAGLE images through the observation network shows a slight decrease in accuracy to the result presented in Table 4.11 but the results still show a preference to assigning galaxies to the non-merger classification. Passing the SDSS images through the 100 Myr linearly scaled EAGLE trained network shows no notable difference in accuracy to the results in the main text but it does not show the tendency to assign the SDSS galaxies to the non-merger classification that the modified Lupton et al. (2004) trained network has.

Examination of the physical classifications and comparing the correctly and incorrectly identified objects, from both the self similar (EAGLE images through EAGLE trained network) and cross application, shows no qualitative differences between the modified Lupton et al. (2004) colour scaled and linearly scaled EAGLE images. Similarly, visual examination of the misclassified linearly scaled EAGLE images produced no major differences to the visual examination of the modified Lupton et al. (2004) scaled images.

Table 4.14: Statistics for the 100 Myr, 200 Myr and 300 Myr EAGLE trained networks at testing using EAGLE images created with a linear colour scaling. The fourth column presents the cross application of passing the linear colour scaled EAGLE images through the observation network while the fifth column presents the cross application of passing the SDSS images through the 100 Myr linearly scaled EAGLE trained network.

	100 Myr	200 Myr	300 Myr	Linear EAGLE images through observation	SDSS images through linear EAGLE trained
Epoch used	60	28	24	-	-
Cut threshold	0.37	0.39	0.46	0.57	0.37
ROC area	0.800	0.745	0.727	0.515	0.689
Recall	0.667	0.612	0.652	0.234	0.676
Precision	0.788	0.715	0.689	0.494	0.630
Specificity	0.821	0.756	0.706	0.761	0.603
NPV	0.711	0.661	0.670	0.498	0.651
Accuracy	0.744	0.684	0.679	0.497	0.639

Effect of galaxy mergers on star formation rates

W. J. Pearson, L. Wang, M. Alpaslan, I. Baldry, M. Bilicki, M. J. I. Brown, M. W. Grootes, B. W. Holwerda, T. D. Kitching, F. F. S. van der Tak
 Astronomy & Astrophysics, 631, A51, 2019
 Reproduced with permission, ©ESO

Abstract

Context: Galaxy mergers and interactions are an integral part of our basic understanding of how galaxies grow and evolve over time. However, the effect that galaxy mergers have on star-formation rates (SFR) is contested, with observations of galaxy mergers showing reduced, enhanced, and highly enhanced star formation.

Aims: We aim to determine the effect of galaxy mergers on the SFR of galaxies using statistically large samples of galaxies, totalling over 200 000, which is over a large redshift range from 0.0 to 4.0.

Methods: We trained and used convolutional neural networks to create binary merger identifications (merger or non-merger) in the SDSS, KiDS, and CANDELS imaging surveys. We then compared the SFR, with the galaxy main sequence subtracted, of the merging and non-merging galaxies to determine what effect, if any, a galaxy merger has on SFR.

Results: We find that the SFR of merging galaxies are not significantly different from the SFR of non-merging systems. The changes in the average SFR seen in the star-forming population when a galaxy is merging are small, of the order of a factor of 1.2. However, the higher the SFR is above the galaxy main sequence, the higher the fraction is for galaxy mergers.

Conclusions: Galaxy mergers have little effect on the SFR of the majority of merging galaxies compared to the non-merging galaxies. The typical change in SFR is less than 0.1 dex in either direction. Larger changes in SFR can be seen but are less common. The increase in merger fraction as the distance above the galaxy main sequence increases demonstrates that galaxy mergers can induce starbursts.

5.1 Introduction

Galaxy mergers and interactions form a key part of our understanding of how galaxies form and evolve over time. In cold dark matter cosmology, dark matter halos merge under hierarchical growth that results in the merger of the halos' baryonic counterparts (e.g. Conselice 2014; Somerville & Davé 2015). This interaction results in the disruption of the galaxies that lie at the centre of the dark matter halos. Tidal forces act to pull and distort the galaxies, subsequently moving stars within the galaxies from the disk to the spheroid component (e.g. Toomre & Toomre 1972; Somerville & Davé 2015, and references therein). Mergers can potentially increase the activity of an active galactic nucleus (e.g. Sanders & Mirabel 1996; Ellison et al. 2019), although more recent work suggests this may not always be the case (e.g. Darg et al. 2010b; Weigel et al. 2018).

Mergers are also thought to trigger periods of extreme star formation: starbursts. From simulations, these starbursts are believed to be a result of the tidal interactions of the galaxies that compress and shock the gas, which results in the rapid formation of stars (e.g. Barnes 2004; Kim et al. 2009; Saitoh et al. 2009). Such shock-induced star formation in mergers has also been observed (Schweizer 2009). These intense star-forming events are believed to be the cause of some of the brightest infrared objects, ultra luminous infrared galaxies (Sanders & Mirabel 1996; Niemi et al. 2012). This connection between starbursts and merging galaxies resulted in the prevailing theory that most merging galaxies go through a starburst phase (e.g. Joseph & Wright 1985; Schweizer 2005).

More recent observations have shown that merger induced starbursts are found in the minority of merging systems. These studies find that the typical increase in star-formation rate (SFR) of a merger is at most a factor of two, much lower than what would typically be considered a starburst (Ellison et al. 2013; Knapen et al. 2015; Silva et al. 2018). Work by Knapen et al. (2015) shows that the majority of galaxy mergers are found to cause a reduction in the SFR when compared to non-merging galaxies of comparable stellar masses. In total, approximately 10-20% of star-forming galaxies are found to be undergoing a merger (Luo et al. 2014; Cibinel et al. 2019) while this fraction increases with redshift (Berrier et al. 2006; Conselice et al. 2009; López-Sanjuan et al. 2010; Lotz et al. 2011; López-Sanjuan et al. 2015; Rodríguez-Puebla et al. 2017). Although acquiring observational evidence for the change of SFR as a function of time before and after a merger is difficult due to the long timescales involved, there is observational evidence for starbursts on the first and second close passes of two galaxies as well as coalescence. These bursts appear to last between 10^7 and 10^8 years (Cortijo-Ferrero et al. 2017). This is supported by observations that show close pairs have higher SFRs than more separated galaxies in mergers (Davies et al. 2015).

Gas rich (wet) mergers are able to support higher SFRs as there is an abundance of fuel available to create new stars (e.g. Lin et al. 2008; Perez et al. 2011; Athanassoula et al. 2016). Gas poor (dry) mergers, however, do not have gas readily available and so it is harder to form starbursts in these systems (e.g. Bell et al. 2006; Naab et al. 2006; Lin et al. 2008). As a result of dense environments containing a higher number of gas poor galaxies than gas rich galaxies, dry galaxy mergers dominate over wet mergers in dense environments (Lin et al. 2010). The fraction of dry mergers also increases with the age of the Universe (Lin et al. 2008). Due to gas poor galaxies dominating at high masses (stellar mass $\gtrsim 10^{10.7} M_{\odot}$), mergers of two high mass galaxies tend to be dry and, as a

result, can act to suppress star formation (Robotham et al. 2014).

A study by Davies et al. (2015) find that the merger ratio of the merging galaxies also influences the SFR. In major mergers (mass ratio $< 3:1$), the lower mass galaxy experiences a short period of enhanced star formation, while in minor mergers (mass ratio $> 3:1$) the star formation in the lower mass galaxy is suppressed. The more massive of the two merging galaxies, however, always experiences an increase in SFR regardless of whether the merger is major or minor.

Simulations of mergers have been conducted, allowing us to study the SFRs of the merging galaxies throughout the entire merger sequence from first passage to coalescence (e.g. Springel et al. 2005; Hopkins et al. 2006; Randall et al. 2008; Rupke et al. 2010). These simulations have shown that SFR is enhanced when the merging galaxies are close to one another at first pass, second pass, and coalescence (Moreno et al. 2019). The period between first and second passes also maintains a higher SFR than in an isolated galaxy, by approximately a factor of two. This period is the majority of the merger sequence, taking approximately 2.5 Gyr of the entire 3.5 Gyr merger timescale (Moreno et al. 2019). This can explain why so few galaxies are observed in the starburst phase of a merger as the period between starbursts is much longer than the starburst period of approximately 0.5 Gyr. The starburst caused by the close passage and coalescence is also found to be stronger for head on collisions and reduces in strength as the approach of the galaxies becomes more oblique. However, the strength of a starburst is also connected to the resolution of the simulation, with lower resolution simulations finding weaker starbursts (Sparre & Springel 2016).

A major observational challenge of merger studies is the difficulty in detecting a large sample of merging galaxies. Visually identifying galaxies is time consuming and hard to reproduce; different people can classify the same galaxy differently and the same classifier may assign different labels on different days. Some of this difficulty can be reduced by employing citizen science, such as Galaxy Zoo¹ (GZ; Lintott et al. 2008), to get many members of the public to classify images of galaxies. However, such approaches are not scalable to the volume of data we expect from upcoming large surveys. Using non-parametric statistics, such as concentration, asymmetry, smoothness (CAS; e.g. Ber-shady et al. 2000; Conselice et al. 2000; Wu et al. 2001; Conselice et al. 2003) or the Gini coefficient, a description of the relative distribution of flux within pixels, and the second-order moment of the brightest 20% of the light (M_{20} ; Lotz et al. 2004) avoids the issues with reproducibility, especially combined with detailed galaxy merger modelling to provide a classification baseline (Lotz et al. 2010a,b). However, merger detection with these non-parametric statistics is sensitive to image quality and resolution, and suffers from a high fraction of misidentifications (Huertas-Company et al. 2015). The close pair method is also often employed, finding pairs of galaxies that are close on the sky and in redshift (e.g. Barton et al. 2000; Lambas et al. 2003; De Propriis et al. 2005; Ellison et al. 2008; Rodrigues et al. 2018; Duncan et al. 2019). However, this method requires highly complete spectroscopic observations and can be contaminated with flybys (Sinha & Holley-Bockelmann 2012; Lang et al. 2014).

Deep learning has the potential to overcome some of these difficulties. Once trained, neural networks are able to perform visual like classifications of galaxies, and other astronomical objects, in a fraction of the time it takes a human, or team of humans, to

¹<http://www.galaxyzoo.org/>

classify the same objects. The classifications are also reproducible: if the same object is passed through the same neural network the result will always be the same. Deep learning techniques are becoming more commonplace in the astronomical community with uses including star-galaxy classification (e.g. Kim & Brunner 2017), galaxy morphology classification (e.g. Dieleman et al. 2015; Huertas-Company et al. 2015; Domínguez Sánchez et al. 2019), gravitational lens identification (e.g. Petrillo et al. 2017; Davies et al. 2019), and galaxy merger identification (e.g. Ackermann et al. 2018; Pearson et al. 2019).

In this work we aim to use deep learning techniques to identify merging galaxies within three data sets: the Sloan Digital Sky Survey (SDSS; York et al. 2000), the Kilo Degree Survey (KiDS; de Jong et al. 2013a,b), and the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011). These three data sets are employed so a large range of redshifts can be covered, with SDSS and KiDS at low redshift and CANDELS at high redshift. With these identifications, we compare the SFRs of the star-forming merging galaxies with the star-forming non-merging galaxies and determine if galaxy mergers have an effect on the SFR of the merging galaxies.

The paper is structured as follows. Section 5.2 discusses the data used and the merger selection for training our neural network. Section 5.3 describes the tools used in this study, including how we determined the galaxy main sequence through modelling and a brief description of the type of deep learning we employ: convolutional neural networks. This is followed by our results and discussion in Sects. 5.4 and 5.5 before we conclude in Sect. 5.6. Where necessary, Wilkinson Microwave Anisotropy Probe year 7 (WMAP7) cosmology (Komatsu et al. 2011; Larson et al. 2011) is adopted, with $\Omega_M = 0.272$, $\Omega_\Lambda = 0.728$, and $H_0 = 70.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

5.2 Data

To train the neural network, a large number of images of pre-classified merging and non-merging systems are required. We also collect images of unclassified images from the same surveys to classify with our networks to increase the sample size for this study. To determine if galaxy mergers affect star-formation rates in the galaxies, we also require stellar masses ($M_\star$), SFR, and redshifts for the pre-classified and unclassified galaxies. We gather these for SDSS, KiDS, and CANDELS.

These three data sets cover different redshift ranges for which merger detection is attempted: the SDSS data that we use covers the low redshift regime ($0.005 < z \leq 0.1$), along with the KiDS data ($0.00 < z \leq 0.15$), while the CANDELS data that we use goes to high redshift ($0.0 < z \leq 4.0$). The overlaps in the redshifts also allow us to examine differences due to resolution, depth, and other effects, by comparing the SDSS and KiDS results, as well as different wavelengths, by comparing the optical SDSS and KiDS with the near-infrared CANDELS. The CANDELS data also probes rest frame optical data at $z \approx 1.2$ with the three CANDELS bands used ($1.6 \mu\text{m}$, $1.25 \mu\text{m}$ and 814 nm) probing approximately the rest frame i, r, and g bands used in the SDSS data.

5.2.1 SDSS Data Release 7

For the SDSS data, we use the network trained on SDSS images from Pearson et al. (2019). The merging and non-merging galaxies used to train this network were collected following Ackermann et al. (2018). The 3003 merging galaxies are from Darg et al. (2010a,b), itself derived from classifications from the Galaxy Zoo (GZ) visual classification. These galaxies have GZ merger classification greater than 0.4 and were then visually checked again to ensure these galaxies are likely to be merging pairs. Approximately half (54%) of these merging galaxies are major mergers (Darg et al. 2010a), that is the ratio of the stellar masses of the two galaxies is less than three. For the non-merging galaxies, 3003 galaxies were randomly selected from galaxies that have their GZ merger classification less than 0.2. Cut-outs of the merging and non-merging objects were then requested from the SDSS cut-out server for data release 7² (DR7) to create 6006 images in the gri bands, each of 256×256 pixels and with Lupton et al. (2004) colour scaling. These images were then cropped to the central 64×64 pixels, corresponding to 25.3×25.3 arcsec or 46.5×46.5 kpc at $z = 0.1$, to reduce memory requirements while training. The merger fraction of the complete training sample, before randomly selecting the non-merging galaxies but after mass completeness cuts, is $\sim 1.0\%$.

To increase the sample for analysis, all SDSS galaxies with spectroscopic redshifts between 0.005 and 0.1 were selected, to match the redshift range of the training sample, and were then classified into merging and non-merging by the Pearson et al. (2019) network, a total of 206 037 galaxies once selected for mass completeness. Again, 256×256 pixel cutouts in the gri bands were collected for these galaxies from the SDSS DR7 cutout server and the central 64×64 pixels used for classification. The $M_\star$ and SFR for these objects were then collected from the MPA-JHU catalogue³, which uses the Kroupa (2001) initial mass function (IMF) (Kauffmann et al. 2003; Salim et al. 2007; Brinchmann et al. 2004). The $M_\star$ is therefore derived from spectral energy distribution (SED) fitting while the SFR is derived from H α observations.

For determining the galaxy main sequence (MS), the star-forming galaxies were selected by performing a cut in the $g-r$ - absolute r magnitude (M_r) plane, closely following Loveday et al. (2012), where we define star-forming galaxies as:

$$g - r < 0.08 - 0.03M_r. \quad (5.1)$$

The rest frame $g - r$ colour was determined by our own fitting of the five SDSS bands with CIGALE (Noll et al. 2009; Boquien et al. 2019). A UVJ colour cut, which is used for the KiDS and CANDELS data, is not used as the wavelength coverage is not sufficient to reliably constrain the J band magnitude. A plot of this colour cut can be seen in Fig. 5.1. The mass limit was determined to be $\log(M_\star/M_\odot) = 10.1$, see Sect. 5.2.4 for details.

5.2.2 KiDS

For our KiDS sample, we use the latest data release 4 (DR4; Kuijken et al. 2019). We match these catalogues with the Galaxy and Mass assembly (GAMA; Driver et al. 2009) GZ catalogue (Holwerda et al. 2019, Kelvin et al. in prep.) to determine the merging and

²<http://cas.sdss.org/dr7/en/tools/chart/default.asp>

³<https://wwwmpa.mpa-garching.mpg.de/SDSS/DR7/>

Figure 5.1: Rest frame $g-r$ colour vs. absolute r magnitude (M_r) for SDSS DR7. The colour cut is shown as a red line where galaxies below the line are considered to be star-forming.

non-merging galaxies and combine this classification with non-parametric statistics (see Sect. 5.2.2). For the KiDS data, we only use r -band images to train the network, using 64×64 pixel cutouts, corresponding to 13.7×13.7 arcsec or 25.2×25.2 kpc at $z = 0.1$, and with linear colour scaling. Tests comparing multi-channel, as used with SDSS and CANDELS, and single channel images, as used with KiDS, to identify galaxy mergers have shown that using a single channel does not notably affect the results. When applying the trained CNN to unclassified objects, we use objects that lie within the GAMA09 field. This region is large enough to provide a statistically significant sample size of galaxies and has the added benefit that it has *Herschel* Spectral and Photometric Imaging Receiver (SPIRE; Griffin et al. 2010) coverage to aid with determining SFRs.

The majority of the KiDS objects in DR4 do not have estimates of physical parameters, beyond photometric redshifts (Kuijken et al. 2019). Thus, to derive $M_\star$ and SFR, we use the 9-band catalogues combined with *Herschel* ATLAS (Eales et al. 2010; Smith et al. 2017) SPIRE data de-blended with XID+ (Hurley et al. 2017; Pearson et al. 2017, see also Appendix 5.7.1). From the 9-band catalogue we use the KiDS Gaussian aperture and point spread function (GAAP; Kuijken et al. 2015) flux densities for the $ugri$ optical bands and the VISTA Kilo-Degree Infrared Galaxy Survey (VIKING; Edge et al. 2013) GAAP flux densities for the ZYJHKs bands, all left uncorrected for foreground extinction. SEDs are fitted to these data using CIGALE and stellar populations with a Chabrier (2003) IMF. As can be seen in Fig. 5.2, the $M_\star$ from CIGALE are in good agreement, within 0.2 dex on average, with those from the GAMA survey (Wright et al. 2017) estimated using the MAGPHYS (da Cunha et al. 2008) SED fitting tool, which also uses the Chabrier (2003) IMF. A similar comparison is made with the SFR in Fig 5.3, showing good agreement with GAMA.

To select the star-forming galaxies for determining the MS, a UVJ colour cut was employed using the rest frame U-V and V-J colours, determined by CIGALE during the fitting to estimate $M_\star$ and SFR, and the photometric redshifts. For this, we follow Whitaker

Figure 5.2: Comparison of M_* from this work (y-axis) with M_* from GAMA (x-axis). The red line denotes the 1-to-1 relation. The two data sets are in reasonable agreement with the average stellar masses within 0.2 dex and remain the same with and without the inclusion of SPIRE data. The typical statistical error on M_* is 0.1 dex.

Figure 5.3: Comparison of SFR from this work (y-axis) with SFR from GAMA (x-axis). The red line denotes the 1-to-1 relation. The two data sets are within 0.2 dex on average and are consistent within the typical error of 0.26 dex. Both the GAMA SFRs and the SFRs from this work are derived from SED fitting.

Figure 5.4: Rest frame U-V colour vs. rest frame V-J colour for KiDS. The colour cut is shown as a red line where galaxies below and to the right of the line are considered to be star-forming.

et al. (2011):

$$\begin{aligned}
 (U - V) &> 0.88 \times (V - J) + 0.69 & z < 0.5 \\
 (U - V) &> 0.88 \times (V - J) + 0.59 & z > 0.5 \\
 (U - V) &> 1.3, (V - J) < 1.6 & z < 1.5 \\
 (U - V) &> 1.3, (V - J) < 1.5 & 1.5 < z < 2.0 \\
 (U - V) &> 1.2, (V - J) < 1.4 & 2.0 < z < 4.0
 \end{aligned} \tag{5.2}$$

where any galaxies that do not meet these criteria are determined to be star-forming. An example of the colour cut is shown in Fig. 5.4. The mass completeness limit for the KiDS galaxies was determined to be $\log(M_{\star}/M_{\odot}) = 9.6$, see Sect. 5.2.4 for details. This was determined using the magnitude limit from the GAMA survey of 19.8, for the r-band, as this is the limit imposed on the training sample.

KiDS-GAMA Galaxy Zoo

There are no pre-existing merger catalogues for the KiDS survey, although there are visual GZ classifications for 36 706 galaxies in the regions that overlap with the GAMA survey (KiDS-GAMA): the three GAMA equatorial fields. We can use this classification to help select a sample of merging galaxies to use with the KiDS data. As with other Galaxy Zoo (Lintott et al. 2008) projects, citizen scientists were asked to classify images of galaxies following a classification tree, as described in Holwerda et al. (2019), through the GZ web interface⁴ and we use the vote fractions that are weighted for user performance. These weighted vote fractions have votes from users that frequently disagree with the majority of other users weighted lower, reducing their influence on the overall vote fraction. These galaxies were selected to have redshifts between 0.002 and 0.15 and GAMA data quality

⁴<http://www.galaxyzoo.org/>

flags are used to ensure only science targets are shown. Of interest here is the question concerning galaxy interactions. This question asks the classifier to identify merging galaxies, galaxies with tidal tails, galaxies that are both merging and have tidal tails or galaxies that show neither of these features. The latter of these classifications, galaxies that have neither tidal tails nor show evidence of a merger, is what is used here to help identify galaxy mergers and will hence forth be referred to as `merger_neither_frac`. Galaxies that have `merger_neither_frac` less than 0.5, that is less than half the people who classified the galaxy thought it showed no tidal features or merger indications, is used here to for the basis of the merging galaxy sample with further refinements added.

KiDS merger selection

The visual GZ merger classifications require validation with other methods, as chance projections or star-galaxy overlaps can be misidentified as merging galaxies (Darg et al. 2010a,b). To do this, we use the Gini, the second-order moment of the brightest 20 percent of the light (M_{20}), concentration (C), asymmetry (A) and smoothness (S) non-parametric parameters (Lotz et al. 2004; Bershadsky et al. 2000; Conselice et al. 2000; Wu et al. 2001; Conselice et al. 2003). For each of the galaxies in the KiDS-GAMA sample, we derive these five non-parametric statistics using the python code `statmorph` (Rodriguez-Gomez et al. 2019).

There has been found to be a division between merging and non-merging galaxies using the Gini and M_{20} statistics: Lotz et al. (2004) found that galaxies can be considered to be non-mergers if

$$\text{Gini} < -0.115M_{20} + 0.384 \quad (5.3)$$

while Lotz et al. (2008) found a similar result with non-mergers defined as

$$\text{Gini} < -0.15M_{20} + 0.33. \quad (5.4)$$

We also populate the Gini- M_{20} parameter space, bin by Gini and M_{20} , and show the average `merger_neither_frac` of the galaxies inside each bin, as seen in Fig. 5.5. The `merger_neither_frac` is the fraction of GZ votes that say the galaxy has no indication of a galaxy merger or tidal tails. In doing this, we find that galaxies found to be mergers in the KiDS-GAMA GZ typically lie on or above these two lines. However, as can be seen in Fig. 5.5, there are also a large number of galaxies that lie above these lines that are classified by GZ as non-mergers: the merging galaxies appear to form a valley. Overlaying the visually confirmed merging galaxies from Darg et al. (2010a,b) that fall within the KiDS coverage, we also find that the majority of these galaxies lie below the Lotz et al. (2004) and Lotz et al. (2008) lines, as can be seen in Fig. 5.5b, suggesting that this is a poor choice to determine merger status for this KiDS data set.

This disparity may be a result of the different data used. The Gini and M_{20} statistics are calculated from the images and so depend on the resolution and signal-to-noise of the images (Lotz et al. 2004). The flux distribution of a lower resolution image will be different, the same flux will be spread across fewer pixels in a lower resolution images as well as removing smaller scale structures, which will increase the uncertainties in these statistics. Similarly, higher signal-to-noise images will reveal fainter features of a galaxy that will also affect the Gini and M_{20} . The Gini and M_{20} have been found to be reasonably consistent when the signal-to-noise is above two but M_{20} is particularly sensitive to

Figure 5.5: Gini vs. M_{20} for KiDS-GAMA GZ galaxies binned by Gini and M_{20} . The average `mergers_neither_frac` from GZ within each bin is shown from low (red) to high (blue). The green line is the Lotz et al. (2004) split between merging and non-merging galaxies while the yellow line is the Lotz et al. (2008) split. Regions with low `mergers_neither_frac` are visually identified as merging galaxies. Panel (b) includes the visually confirmed mergers from Darg et al. (2010a,b) as purple stars.

resolution (Lotz et al. 2004). The data used in Lotz et al. (2004) is lower resolution than KiDS while Lotz et al. (2008) uses Hubble Space Telescope data with higher resolution.

If instead we use the asymmetry (A) and smoothness (S) statistics, which have been found to be not overly sensitive to resolution as M_{20} (Lotz et al. 2004), we find a merging sample that agrees much better with the visual classification. It has been found, by Conselice (2003), that the merging galaxies lie above

$$A = 0.35S + 0.02. \quad (5.5)$$

As can be seen in Fig. 5.6, this classification is in good agreement with the visual classifications from GZ. Overlaying the Darg et al. (2010a,b) mergers, we find that the majority lie above Eq. 5.5. Based on this agreement, we select our merging sample to be those galaxies that have `merger_neither_frac` from GZ less than 0.5 and lie above Eq. 5.5, with non-merging galaxies defined as those with `merger_neither_frac` greater than 0.5 and lie below Eq. 5.5. This results in 1917 merging galaxies that we use to train the KiDS network. By matching these galaxies to the nearest galaxy within 3 arcsec in the full GAMA catalogue (Wright et al. 2017) and selecting the pairs that have redshifts within 0.05, we find that approximately half of these galaxies (6 of 14) are major mergers. The total number of matched pairs is very low, and misses pairs where the secondary galaxy is below the magnitude limit of the survey, but this fraction is in line with that seen by Darg et al. (2010a) in the SDSS data. We randomly select a further 1917 galaxies from the 20842 that lie below Eq. 5.5 and have `merger_neither_frac` greater than 0.5 to form the non-merging sample. With these classifications for merging and non-merging galaxies, and after mass completeness cuts, the merger fraction of the GZ galaxies is 8.4%.

5.2.3 CANDELS

To train the CANDELS network, we use the visual classifications for the Great Observatories Origins Deep Survey - South (GOODS-S; Giavalisco et al. 2004) from Kartaltepe et al. (2015). This catalogue contains galaxies with H magnitude less than 24.5 that have been classified by a small number of professional astronomers and we select objects with photometric redshift below 4.0. Of interest to this work are the classifications that identify mergers (`merger`), interaction within a segmentation map (`Int1`), interaction with a galaxy outside of the segmentation map (`Int2`), a non interacting companion (`Comp`) or no interaction (`NoInt`). During the classification, only one of these identifications may be chosen. The catalogue also contains an `Any_Int` category, which combines the `merger`, `Int1`, and `Int2` identifications.

We define galaxies as merging if the `Any_Int` classification is greater than 0.6 (that is more that 60% of people believe that the galaxy is interacting) and we define galaxies as non-merging if the `Any_Int` classification is less than 0.5. As with the KiDS galaxies, we match the merging galaxies to the rest of the CANDELS catalogue within 3 arcsec and selecting the pairs that have redshifts within 0.05, we find that approximately half of these galaxies (4 of 9) are major mergers. Again, the total number of matched pairs is very low, and this method misses pairs where the secondary galaxy is below the magnitude limit of the survey, but this fraction is in line with that seen in the SDSS data. Cutouts for these objects were created from the 1.6 μm , 1.25 μm , and 814 nm images. As the 814 nm

Figure 5.6: Asymmetry (A) vs. Smoothness (S) for KiDS-GAMA GZ galaxies binned by A and S. The average `merger_neither_frac` from GZ within each bin is shown from low (red) to high (blue). Regions with low `merger_neither_frac` are visually identified as merging galaxies. The orange line denotes the Conselice (2003) split between merging and non-merging galaxies. Panel (b) includes the visually confirmed mergers from Darg et al. (2010a,b) as purple stars.

images are twice the angular resolution of the other two bands, these images are reduced in size by averaging the flux density in 2×2 pixel groups. The $1.6 \mu\text{m}$, $1.25 \mu\text{m}$, and 814 nm bands are then used as the red, green, and blue channels in the images, with simple linear colour scaling. As with the SDSS and KiDS images, the CANDELS images are 64×64 pixels, corresponding to 3.8×3.8 arcsec or 32.7×32.7 kpc at $z = 1.5$. Objects with clear artefacts within the image were removed. This resulted in 694 merging galaxies and we randomly select a further 694 from the 4428 non-merging galaxies that meet our criteria. The merger fraction for the training sample using these criteria, and after mass completeness cuts, is 15.5%.

To increase the CANDELS sample for our analysis, we classified all CANDELS galaxies with H-magnitude < 24.5 and redshift between 0.0 and 4.0, to match the training sample, from the Cosmic Evolution Survey (COSMOS; Scoville et al. 2007), Extended Groth Strip (EGS; Davis et al. 2007), UKIRT Infrared Deep Sky Survey (UKIDSS) Ultra-Deep Survey (UDS; Lawrence et al. 2007; Cirasuolo et al. 2007) fields with the CANDELS network (once trained). Images for these galaxies were created as above. The H-magnitude and SED derived SFR and M_* , assuming a Chabrier (2003) IMF, come from Guo et al. (2013); Santini et al. (2015) for GOODS-S, Nayyeri et al. (2017) for COSMOS, Stefanon et al. (2017) for EGS, and Santini et al. (2015) for UDS. As these catalogues contain a number of different M_* and SFR values, the ‘M_med’ is used for M_* and we average ‘SFR_11a_tau’, ‘SFR_13a_tau’, ‘SFR_2a_tau’, ‘SFR_14a’, ‘SFR_14a_const’, ‘SFR_14a_delta’, ‘SFR_14a_lin’, ‘SFR_14a_tau’, ‘SFR_6a_delta’, ‘SFR_6a_invttau’, and ‘SFR_6a_tau’ for SFR, as these columns are common across all catalogues. These different SFR values assume different star-formation histories (SFH) where ‘cons’ is a constant SFH, ‘tau’ is an exponentially declining SFH, ‘delta’ is a delayed-exponential, ‘lin’ is linearly increasing, and ‘inctau’ is exponentially increasing. The numbers refer to the investigator within the CANDELS team who lead the determination of that SFR (Stefanon et al. 2017). For the redshift, the ‘z_best’ value in the catalogues were used. This value is the spectroscopic redshift, if available, or the best photometric redshift from six members within the CANDELS team (Guo et al. 2013; Santini et al. 2015; Nayyeri et al. 2017; Stefanon et al. 2017).

To determine which CANDELS galaxies are star-forming, we again apply the UVJ colour cuts defined in Eq. 5.2 using the rest frame U-V and V-J colours in the CANDELS catalogues, as shown in Fig. 5.7. Mass completeness limits were calculated to be $\log(M_*/M_\odot) = 8.3, 8.7, 9.1, 9.4$, and 9.9 within redshift bins with edges at $z = 0.0, 0.6, 0.85, 1.21, 1.66$, and 4.0 , see also Sect. 5.2.4 below. These redshift bins were selected so there are approximately 2000 galaxies within each bin after cutting for mass completeness. For ease of reference, these redshift bins shall be referred to as CANDELS-z000, CANDELS-z060, CANDELS-z085, CANDELS-z121, and CANDELS-z166. A summary of all data sets is presented in Table 5.1.

Figure 5.7: Rest frame $U - V$ colour vs. rest frame $V - J$ colour for CANDELS-z000. The colour cut is shown as a red line where galaxies below and to the right of the line are considered to be star-forming.

Table 5.1: Summary of data used. The SDSS and KiDS limiting magnitudes are in r-band while the CANDELS limiting magnitude in H-band.

Data	Resolution (arcsec)	Magnitude limit	Redshift range	Mass limit $\log(M_*/M_\odot)$	Training sample per class	Complete sample (Mass limited)
SDSS	1.4	17.77	$0.005 < z \leq 0.1$	10.1	3003	206037
KiDS	0.77	19.8 ^a	$0.00 < z \leq 0.15$	9.6	1917	1270
CANDELS-z000	0.15	24.5	$0.00 < z \leq 0.60$	8.3	694 ^b	2072
CANDELS-z060	0.15	24.5	$0.60 < z \leq 0.85$	8.7	694 ^b	2004
CANDELS-z085	0.15	24.5	$0.85 < z \leq 1.21$	9.1	694 ^b	2031
CANDELS-z121	0.15	24.5	$1.21 < z \leq 1.66$	9.4	694 ^b	2010
CANDELS-z166	0.15	24.5	$1.66 < z \leq 4.00$	9.9	694 ^b	1910

Notes. ^(a) As the training set is derived from GAMA classifications, the limiting magnitude is that of the GAMA survey not that of the KiDS survey. ^(b) The CANDELS network was trained with 694 galaxies per class for galaxies with $0.00 < z \leq 4.00$. The galaxies were split into the redshift bins shown after classification.

5.2.4 Mass completeness

Mass completeness limits were determined empirically by following Pozzetti et al. (2010) and using the galaxies identified as star-forming. For each galaxy, the mass the galaxy would need to have to be detected at the magnitude limit (M_{lim}) was calculated with

$$\log(M_{\text{lim}}) = \log(M) - 0.4(x_{\text{lim}} - x), \quad (5.6)$$

where x is the observed magnitude in the r-band (for SDSS and KiDS) or H-band (for CANDELS) and x_{lim} is the limiting magnitude of the observation. The limiting magnitudes for SDSS and CANDELS are 17.77 and 24.50 respectively. The KiDS limiting magnitude is 19.8, the limit of the GAMA survey. The faintest 20% of objects were selected and the limiting mass was the M_{lim} value that 90% of these faintest objects lie below. This was done as a function of redshift by binning the galaxies into redshift bins as described in Sect. 5.4 below. These completeness limits were then applied to the entire galaxy population.

5.3 Tools

5.3.1 Convolutional neural networks

Convolutional neural networks (CNNs) are a subset of deep learning (e.g. Lecun et al. 2015, and references therein). CNNs are used for image classifications and employ a series of non-linear mathematical functions, known as neurons, each with a weight and bias value. The structure of a CNN is built from a number of layers of these neurons. The lower layers are created from two-dimensional kernels that are convolved with the output of the layer below, giving CNN its name. Upper layers are one-dimensional and each neuron in these layers is connected to every neuron in the layer below. Forming a network in such a way can rapidly create a large number of neurons that require training resulting in many more free parameters within the network than there are data to train them. To reduce this dimensionality, pooling layers are employed between the lower convolutional layers. These pooling layers group the inputs into it and pass on the maximum or average value of the group, depending on the type of pooling used, with the grouping done in two-dimensions. The result is an output that is smaller in the width-height plane but has the same depth as the input. The weights and biases of the neurons within a network are trained, in the case of supervised learning used here, by passing labelled data through the network and requiring the output classification to converge on these labels. A complete and thorough description of CNNs is beyond the scope of this paper but further details are explained in Lecun et al. (1998).

This paper uses the definitions of Pearson et al. (2019) for the terms to describe the properties of CNNs. These terms may be an alternate nomenclature to other works or may be unfamiliar. To avoid confusion we reproduce these definitions in Appendix 5.7.2.

Architecture of the CNN

For this work, we use the architecture developed in Pearson et al. (2019) and use this to train on data from CANDELS and KiDS (the network trained in Pearson et al. (2019) is

Figure 5.8: Schematic representation of the architecture of the CNN used with input (three or one colour, 64×64 pixel image) on the left and output (binary classification of merger or non-merger) on the right. The sizes of the kernels (red) and layers are shown with the input layer having a depth of three for the SDSS (gri) and CANDELS (1.6 μm , 1.25 μm , and 814 nm) data and one for the KiDS (r-band) data.

used on the SDSS images). This network is built with *Tensorflow* (Abadi et al. 2015) and comprises of a series of four, two-dimensional convolutional layers followed by two one-dimensional, fully connected layers of 2048 neurons. The convolutional layers have 32, 64, 128, and 128 kernels of 6×6, 5×5, 3×3, and 3×3 pixels for the first, second, third, and fourth layers respectively with the stride, how far the kernel is moved as it scans the input, set at 1 pixel for all layers and the zero padding is set to ‘same’ to pad each edge of the image with zeros evenly (if required). 2×2 pixel max-pooling is applied after the first, second, and fourth convolutional layers to reduce the dimensionality of the network. Batch normalisation (Ioffe & Szegedy 2015) is applied after each layer, scaling the output between zero and one, and we use Rectified Linear Units (ReLU; Nair & Hinton 2010) for activation, which returns $\max(x, 0)$ when passed x . We also apply dropout (Srivastava et al. 2014) after each activation, to help reduce over-fitting, with a dropout rate of 0.2 to randomly set the output of neurons to zero 20% of the time during training. The loss of this network is determined using softmax cross entropy and optimised using the Adam algorithm (Kingma & Ba 2015) with a learning rate of 5×10^{-5} . The architecture can be seen schematically in Fig. 5.8.

The inputs to the network are 64×64 pixel images with either three (CANDELS) or one (KiDS) colour channels, that are globally scaled between 0 and 1, preserving the relative flux densities for multi colour images. The output layer has two neurons, one for each of the merging and non-merging classes, and uses a softmax output, providing the probability for each class in the range [0, 1] that sum to unity, that is softmax maps the unnormalised input into it to a probability distribution over the output classes if the training, test, and validation data sets have an equal number of each class. In the following, we use the output for the merger class (`frac_merger`) although this can be considered to be equivalent to using the output for the non-merger class as it is $1-(\text{frac_merger})$ in our binary classification.

Due to the limitations of this architecture, specifically the use of fully connected layers, it is not possible to have input images of different pixel sizes. As a result, it is not possible to use cutouts of different sizes that maintain the relative size of the galaxy within

the image. It is possible to resize the images but this risks losing small scale structure when downscaling galaxies or creating artefacts when upscaling. This may cause issues for large galaxies at low redshifts and low resolution as the primary galaxy may fill the image. However, it is possible to correctly identify the galaxies that fill, or are larger than, the image if the training set contains these types of galaxies, as has been shown in Pearson et al. (2019).

Training, validation, and testing

If there are an unequal number of images in the two classes, the larger class size is reduced by randomly removing images until the classes are the same size. The images were then subdivided into three groups: 80% were used for training, 10% for validation, and 10% for testing. The training set was the set used to train the network while the validation was used to see how well the network was performing as training progressed. The testing set was used once, and once only, to test the performance of the network deemed to be the best from the validation. Testing images are not used for validation to prevent accidental training on the test data set.

During training, the input images were augmented. To increase the number of galaxies to train with for CANDELS and KiDS, the images were randomly augmented in one of the following ways: the image was translated by up to 4 pixels in each direction, zoomed by up to 2 $\times$, rotated by a random angle about the galaxy, skewed by a random angle between 10° and 30° or passed without change. For the first four of these options, random Gaussian noise was added so the galaxy appeared ‘new’ to the network. These augmented images were then randomly rotated by 0°, 90°, 180° or 270° to reduce sensitivity to galaxy orientation. Once complete, we have one network for the CANDELS data and one for the KiDS data.

5.3.2 Forward modelling of the galaxy main sequence

In determining if galaxy mergers have an effect on the star-formation rates (SFR) of galaxies, it is necessary to remove the mass dependence from the star-formation rate. To do this we correct for the slope of the main sequence of star-forming galaxies (MS, e.g. Brinchmann et al. 2004; Noeske et al. 2007; Elbaz et al. 2007; Speagle et al. 2014; Pearson et al. 2018). The MS is an observed tight correlation between the $M_\star$ and SFR of star-forming galaxies that exists out to at least $z = 6$ with a scatter of ~ 0.3 dex that is mass and redshift independent. The slope of the MS is found to be less than one and to depend on the redshift (Pearson et al. 2018). To correct for this slope, it is necessary to first determine the MS of the galaxies.

To determine the MS, we follow Pearson et al. (2018) and use forward modelling. Assuming a linear MS, we use the Markov Chain Monte-Carlo (MCMC) sampler *emcee* (Foreman-Mackey et al. 2013) to simultaneously fit the slope, normalisation, and scatter of all the galaxies that are star-forming. At each sampled point in the parameter space, model SFRs are created using the observed stellar masses, observed redshifts, and the corresponding positions along the sampled MS. These SFRs are then perturbed by selecting a random number from a Gaussian distribution, with the standard deviation equal to the sampled scatter, truncated to match the observed upper and lower SFR limits. To include

observational uncertainties on SFR and $M_\star$, both the simulated SFR and the observed $M_\star$ of each galaxy are perturbed again by a random number sampled from a Gaussian distribution with the standard deviation equal to the error in the observed SFR or $M_\star$.

At each step, the mock SFR- $M_\star$ plane is compared to the observed SFR- $M_\star$ plane by binning the data into identical bins in $M_\star$. The means and standard deviations of the SFRs inside each bin are calculated and the results from the mock data are compared to the results from the observed data. The closer the means and standard deviations in the mock bins are to the observed bins, the more likely the model is a correct representation of the observed data. Due to the large number of objects, for the SDSS data we randomly select 50 000 star-forming objects within each redshift bin to determine the MS.

5.4 Results

5.4.1 Performance of the CNN

The network we use for the SDSS objects is that of Pearson et al. (2019). For completeness, we repeat these results here. The SDSS network achieves an accuracy of 0.932 at validation with a cut threshold of 0.5, that is any object with `frac_merger` > 0.5 is classified as a merging galaxy. We can alter the threshold value to find the threshold value that simultaneously minimises the fall out and maximises the recall. By doing this to set the threshold to 0.57, the accuracy increases to 0.935. Using the same threshold of 0.57 at testing, the final accuracy of the SDSS network is 0.915, with recall, precision, specificity, and negative predictive value (NPV) of 0.920, 0.911, 0.910, and 0.919 respectively. We apply the network to 256 497 SDSS galaxies with spectroscopic redshifts between 0.005 and 0.1, to match the training set, resulting in $28\,971 \pm 2578$ ($14.1 \pm 1.3\%$) galaxies being identified as mergers. The errors in these merger counts are derived from the precision of the network, see Appendix 5.7.2 for the definition. The number of merging galaxies is multiplied by the precision and the difference between this value and the original count is taken as the error for the number of mergers and number of non-mergers. This is likely an underestimate as the precision assumes equal population sizes of mergers and non-mergers, which is evidently not the case.

For the KiDS network, we use the CNN to identify galaxies that fall within the GAMA09 field. This network achieves an accuracy of 0.942 at validation with a cut threshold of 0.5. If we alter the threshold to 0.52, to simultaneously minimise the fall out and maximise the recall, the accuracy increases to 0.948. Using the same threshold of 0.52 at testing, the final accuracy of the KiDS network is 0.903, with recall, precision, specificity, and NPV of 0.942, 0.874, 0.864, and 0.938 respectively. Despite the resolution of the KiDS images being higher than that of the SDSS images, the same image size resulted in the best performance for the KiDS network: 64×64 pixels. Larger images were tried but these networks did not perform as well. Applying the KiDS network to all galaxies in the GAMA09 field with photometric redshifts below 0.15, a total of 1270 galaxies, we identify 436 ± 55 ($30.0 \pm 4.3\%$) merging galaxies.

The CANDELS network achieves an accuracy of 0.826 at validation with a cut threshold of 0.5. If we decrease the threshold to 0.47, the accuracy increases to 0.840. Using the same threshold of 0.47 at testing, the final accuracy of the CANDELS network is 0.818, with recall, precision, specificity, and NPV of 0.870, 0.789, 0.768, and 0.855 respectively.

Table 5.2: Statistics for trained CNNs. Definitions of terms can be found in Appendix 5.7.2

	SDSS ^a	KiDS	CANDELS
	0.005 < z < 0.1	0.00 < z ≤ 0.15	0.00 < z ≤ 4.0
Cut threshold	0.57	0.52	0.47
ROC area	0.966	0.957	0.861
Recall	0.920	0.942	0.870
Precision	0.911	0.874	0.789
Specificity	0.910	0.864	0.768
NPV	0.919	0.938	0.855
Accuracy	0.915	0.903	0.818

Notes. ^(a) The SDSS network is that of Pearson et al. (2019).

The poorer results for the CANDELS network is likely due to fewer pre-classified objects to train the network with, 694 per class for CANDELS compared to 3003 for SDSS, as well as the higher redshifts of the training objects. The CANDELS images also cover a much larger redshift range, resulting in a greater distribution of sizes in the image for galaxies at the same mass than the SDSS images. Ideally, it would be preferable to split the galaxies into redshift bins and train a network per redshift to minimise this effect, however with so few objects it is not feasible. We apply the CANDELS network to the objects with H-magnitude < 24.5 and 0.0 < z < 4.0 in the CANDELS COSMOS, EGS, and UDS fields and identify 3535 ± 746 merger candidates out of the 10027 galaxies in these three fields. This is a merger fraction of $35.3 \pm 7.4\%$, which is high. The statistics for all the networks are presented in Table 5.2 and examples of non-merger and mergers selected by the CNNs can be found in Appendix 5.7.3.

5.4.2 SDSS

To determine the effect of galaxy mergers on SFR, we determine the effect of mergers on the MS subtracted SFR. We fit the MS to all the star-forming galaxies, both mergers and non-mergers together, and the MS we have fitted to the SDSS data is shown overlaid onto all the non-merging and merging galaxies in Fig. 5.9. The MS subtracted SFR of the merging and non-merging galaxies are then compared by fitting a skewed Gaussian distribution, of the form

$$y = \frac{A}{\sigma} \exp\left(\frac{(x - \mu)}{2\sigma^2}\right) \left(1 + \operatorname{erf}\left[\frac{\alpha(x - \mu)}{\sqrt{2}\sigma}\right]\right), \quad (5.7)$$

to the distributions of the merging and non-merging galaxies, where A is the amplitude, μ , and σ are the mean and standard deviation of the Gaussian, α is the description of skewness, and erf is the error function.

To fit the skewed Gaussian we bin the MS subtracted SFR with bin sizes of 0.25 dex, between $-3 \log(\text{M}_\odot \text{yr}^{-1})$ and $2 \log(\text{M}_\odot \text{yr}^{-1})$ and fit the skewed Gaussian to the number of galaxies in each bin. The errors on these counts were determined by generating 100

Figure 5.9: SFR- M_* plane populated with (a) non-merging galaxies and (b) merging SDSS galaxies. The colour indicates the number density from low (light yellow) to high (dark purple). Overlaid in red is the MS that has been fitted to all star-forming galaxies. As can be seen, the distributions of the merging and non-merging galaxies are similar with respect to the plotted MS.

Figure 5.10: Distribution of MS subtracted SFR for star-forming SDSS non-merging galaxies (blue) and merging galaxies (red). As can be seen, the merging star-forming population has a slightly higher mean MS subtracted SFR.

Table 5.3: Best fit parameters for skewed Gaussian distributions fitted to star-forming SDSS data. μ and σ are in units of $\log(M_\odot \text{yr}^{-1})$

Parameter	Merger	Non-merger
μ	0.33 ± 0.02	0.25 ± 0.01
σ	0.43 ± 0.02	0.39 ± 0.01
α	-1.29 ± 0.18	-1.52 ± 0.16

realisations of the MS subtracted SFR by perturbing the SFR and $M_\star$ of each galaxy by a random number drawn from a Gaussian distribution centred on the observed SFR or $M_\star$ and with the error on the value as the standard deviation. Each realisation was then binned in the same way as the observations and the standard deviation of the counts in the bins of the 100 realisations were taken as the errors on the counts of the observations. The `scipy.optimize` package `curve_fit` was then used to fit the skewed Gaussian distribution to the counts in the bins and account for their errors. The distributions are presented in Fig. 5.10 with the parameters for the skewed Gaussian fits in Table 5.3.

Comparing the skewed Gaussian fits to the distributions, we find that the mean for the star-forming mergers and non-mergers are consistent within three times the error of the mean (σ_μ) and the merging galaxies have higher mean MS subtracted SFR. This suggests that the star-forming population has a slightly, but not significantly, increased SFR when undergoing a merger.

Figure 5.11: Distribution of MS subtracted SFR for star-forming KiDS non-merging galaxies (blue) and merging galaxies (red). As can be seen, the merging star-forming galaxies have a similar mean MS subtracted SFR to the non-merging galaxies.

Table 5.4: Best fit parameters for skewed Gaussian distributions fitted to star-forming KiDS data. μ and σ are in units of $\log(M_{\odot}\text{yr}^{-1})$

Parameter	Merger	Non-merger
μ	0.44 ± 0.1	0.13 ± 0.18
σ	0.47 ± 0.09	0.41 ± 0.08
α	-1.53 ± 0.91	-0.63 ± 0.78

5.4.3 KiDS

As with the SDSS data, we fit a skewed Gaussian distribution to the MS subtracted SFR of the star-forming galaxies. An example of the resulting MS subtracted SFR distributions for the KiDS is shown in Fig. 5.11. Table 5.4 shows that the merging star-forming galaxies have higher average SFRs. The differences not large, with the mean MS subtracted SFR being within $3\sigma_{\mu}$ of each other.

5.4.4 CANDELS

Due to the larger redshift coverage of the CANDELS data, we can examine if the impact of galaxy mergers on SFR changes as a function of redshift. To do this, we divided the data into redshift bins with edges at $z = 0.0, 0.6, 0.85, 1.21, 1.66$, and 4.0 , each with its own mass completeness limit and containing approximately 2000 galaxies after mass completeness cuts have been applied. Each redshift bin also had its own main sequence fitted as outlined in Sect. 5.3.2. For ease of reference, these redshift bins shall be re-

Figure 5.12: Distribution of MS subtracted SFR for the $0.85 < z \leq 1.21$ redshift bin for CANDELS non-merging galaxies (blue) and merging galaxies (red). This is the only data that is fitted with a double Gaussian distribution due to the clear multi-modal population. As can be seen, the main and secondary populations have a slightly higher mean MS subtracted SFR than the non-merging galaxies.

ferred to as CANDELS-z000, CANDELS-z060, CANDELS-z085, CANDELS-z121, and CANDELS-z166.

As before, we fit the distributions of the star-forming CANDELS galaxies with a skewed Gaussian function. However, there is an indication of a second, high SFR population in CANDELS-z085 ($0.85 < z \leq 1.21$), identifiable when the error on the skew of the both the merging and non-merging distributions are greater than 10^4 and so in that bin only, we fit a double Gaussian distribution and consider the lower mean to be the mean of the star-forming population. The distributions for the MS subtracted SFR for this redshift bin is shown in Fig. 5.12.

Using the best fitting values for the skewed and double Gaussian functions presented in Table 5.5, in the lowest redshift bin the merging galaxies act to suppress the SFR of galaxies, with the mean MS subtracted SFR for merging galaxies lower than non-merging galaxies by more than $3\sigma_\mu$. At redshifts above $z = 0.60$, the star-forming mergers have a higher mean than the star-forming non-mergers or are consistent within $3\sigma_\mu$.

5.5 Discussion

Here we present discussions of our results. We note that direct comparisons between the results of the three data sets is difficult due to the different definitions of mergers employed for the training data sets as well as difference in data quality, such as depth and resolution, which can also influence merger identification. While the merger definitions are similar, as they are all based on visual classification, the specifics of the definitions

Table 5.5: Best fit parameters for skewed or double Gaussian distributions fitted to star-forming CANDELS data. For the $0.85 < z \leq 1.21$ bin, where a double Gaussian is used, the star-forming component is the component with the lowest μ . μ , and σ are in units of $\log(M_\odot \text{yr}^{-1})$

Redshift	Parameter	Merger	Non-merger
$0.0 < z \leq 0.6$	μ	-0.37 ± 0.17	-0.27 ± 0.02
	σ	0.47 ± 0.08	0.29 ± 0.02
	α	9999.489 ± 87523220.85	1.21 ± 0.26
$0.6 < z \leq 0.85$	μ	0.10 ± 0.12	0.02 ± 0.11
	σ	0.24 ± 0.07	0.24 ± 0.06
	α	-0.878 ± 1.17	-0.79 ± 0.99
$0.85 < z \leq 1.21$	μ	-0.12 ± 0.01	-0.16 ± 0.01
	σ	0.20 ± 0.01	0.20 ± 0.01
$1.21 < z \leq 1.66$	μ	-0.43 ± 0.02	-0.42 ± 0.02
	σ	0.62 ± 0.03	0.52 ± 0.03
	α	5.058 ± 1.24	3.13 ± 0.69
$1.66 < z \leq 4.0$	μ	-0.47 ± 0.02	-0.5 ± 0.02
	σ	0.65 ± 0.03	0.61 ± 0.03
	α	3.241 ± 0.51	3.7 ± 0.61

differ. The classifications also cover both major and minor mergers, with approximately half of each training set comprising of major mergers. This likely results in a similar split for the mergers classified by our networks.

5.5.1 Merger influence on SFR

Across the SDSS, CANDELS, and KiDS data sets there is a difference between the SFRs of the merging and non-merging galaxies. However, the difference between the two is small and varies between the data sets as well as within the data sets. What is evident is that the merging systems are not only found as starburst galaxies but also as star-forming and quiescent systems.

Comparing the SDSS data with the KiDS data, we find little difference in how mergers are affecting the SFR. Both data sets show that star-forming merging galaxies have a slight increase in SFR. Within the CANDELS-z000 data the opposite is found: we find that there is a decrease in the MS subtracted SFR, suggesting that galaxy mergers are acting to reduce the SFR of the star-forming galaxies. The full comparison between the average SFRs for all data sets at all redshifts studied can be seen in Fig. 5.13.

The slight difference between the merging and non-merging SFRs is also not a result of the observation bands or methods used to derive the SFR. The mergers in all three surveys are detected using different bands: SDSS uses three optical bands (gri), CANDELS uses observed frame near infrared (Hubble $1.6 \mu\text{m}$, $1.25 \mu\text{m}$, and 814 nm bands), and KiDS uses a single optical band (r band). SFRs in the three surveys are also derived differently: SDSS uses $H\alpha$ based SFR while CANDELS and KiDS use SED derived SFR. The models used to derive the CANDELS and KiDS SFRs and $M_\star$ are also different.

Figure 5.13: Average MS subtracted SFR of star-forming galaxies for SDSS merging (purple circle) and non-merging (dark blue diamond); KiDS merging (light blue circle) and non-merging (green diamond); and CANDELS merging (orange circles) and non-merging (red diamonds) galaxies. As can be seen, the change in SFR between the merging and non-merging galaxies is typically small.

Thus, the small effect of merging galaxies on the SFRs seen in this study is robust.

Our results are qualitatively in line with previous work in that we only find small (less than a factor of two) changes in SFR. Lackner et al. (2014) and Knapen et al. (2015) find that mergers change the SFR by up to a factor of two. While we do not find that mergers always result in an increase in SFR, we do find that the change in SFR caused by a galaxy merger is typically small over the timescale of the entire merger. If an increase in SFR due to a galaxy merger is large but shorter lived, the effect will be hidden by the larger number of galaxies not undergoing such a burst of star formation. The changes in average MS subtracted SFR are small and typically found to be a factor of ~ 1.2 . Similarly, Silva et al. (2018) find that mergers produce no significant change to the SFR of galaxies, which is consistent with the results of our study. However, caution must be taken with this comparison as the work of Silva et al. (2018) uses mergers where the two merging galaxies are within 3-15 kpc of each other, something that this work does not take into account.

This study has its limitations. It is likely that we are observing different stages of galaxy mergers but our method is currently unable to determine at what stage the mergers are. As a result, it is not possible to say, from this study, if mergers cause a migrating of the merging galaxies across the SFR- M_* plane or if the merger only slightly affects the SFR resulting in the small changes we observe.

Table 5.6: Merger fraction by redshift and data set for quiescent, star-forming, and total galaxy populations. Errors are derived from correcting for the precision of the network.

Data Set	Total	Quiescent	Star-forming
SDSS	$14.1 \pm 1.3\%$	$14.3 \pm 1.3\%$	$13.4 \pm 1.2\%$
KiDS	$30.0 \pm 4.3\%$	$19.4 \pm 2.8\%$	$36.9 \pm 5.3\%$
CANDELS-z000	$32.0 \pm 6.8\%$	$30.0 \pm 6.2\%$	$32.4 \pm 6.8\%$
CANDELS-z060	$32.2 \pm 6.8\%$	$20.2 \pm 4.3\%$	$33.6 \pm 7.1\%$
CANDELS-z085	$32.6 \pm 6.9\%$	$24.4 \pm 5.1\%$	$33.3 \pm 7.0\%$
CANDELS-z121	$37.8 \pm 8.0\%$	$23.9 \pm 5.3\%$	$39.4 \pm 8.3\%$
CANDELS-z166	$42.1 \pm 8.9\%$	$28.5 \pm 5.9\%$	$44.3 \pm 9.3\%$

5.5.2 Merger fractions

The merger fractions for CANDELS, $35.3 \pm 7.4\%$, and KiDS, $36.9 \pm 5.3\%$, are notably higher than the merger fractions for the SDSS data at $14.1 \pm 1.3\%$, see Table 5.6 and Figs. 5.14 and 5.15. The errors in these merger fractions are derived from the precision of the network, see Appendix 5.7.2 for the definition. The number of merging galaxies is multiplied by the precision and the difference between this value and the original count is taken as the error for the number of mergers and number of non-mergers. The merging fraction for the precision corrected counts is then calculated and the difference between the original fraction and this precision corrected fraction is taken as the error. This is likely an underestimate as the precision assumes equal population sizes of mergers and non-mergers, which is evidently not the case.

Even only considering the lowest redshift bin for CANDELS, $0.00 < z \leq 0.60$, the merger fraction is much higher than the SDSS and higher redshift KiDS at $32.0 \pm 6.8\%$. It is unsurprising that this becomes a larger issue as the redshift increases because the pixel size of the galaxy within the image becomes smaller and the galaxies themselves become fainter, suppressing the features that the CNN looks for to identify a merging galaxy.

Comparing our merger fractions to other works shows that the CANDELS results are indeed much higher than would be expected. Figure 5.14 shows the comparison of this work with Conselice et al. (2003), who use CAS to identify mergers, Lotz et al. (2011), who use Gini and M_{20} , (Cotini et al. 2013), who use CAS, Gini and M_{20} , and Duncan et al. (2019), who use the close pair method. The results of Duncan et al. (2019) are the merger pair fraction (the number of pairs of merging galaxies divided by the total number of galaxies) and so we multiply their values by 0.6 to compare to our results (Lotz et al. 2011; Mundy et al. 2017).

The SDSS merger fraction is higher than the other works in the same redshift range but is consistent with the merger fractions of Conselice et al. (2003) and Lotz et al. (2011) at higher redshifts. The KiDS data has a merger fraction that is higher compared to the other works, both at similar and higher redshifts, similar to the merger fractions from CANDELS as discussed above.

We can compare the merger fractions of the quiescent and star-forming galaxies as shown in Fig. 5.15. The SDSS data has a slightly lower merger fraction for the quiescent

Figure 5.14: Total merger fraction as a function of redshift for SDSS (dark blue circle), KiDS (light blue circle), and CANDELS (red circles) by redshift bin. Also plotted are the mass limited merger fractions with $\log(M_*/M_\odot) > 10.0$ from Conselice et al. (2003, green stars), Cotini et al. (2013, lilac diamonds), Lotz et al. (2011) magnitude limited merger fractions with $M_B > -19.2$ (orange crosses), and the Duncan et al. (2019) lower mass ($9.7 < \log(M_*/M_\odot) < 10.3$, L, purple left triangles) and higher mass ($\log(M_*/M_\odot) > 10.3$, H, brown right triangles) merger fractions. The SDSS data are slightly higher than would be expected and the KiDS and CANDELS merger fractions are approximately a factor of two higher than the other results.

Figure 5.15: Merger fraction of quiescent (diamonds) and star-forming (circles) as a function of redshift for SDSS (purple and blue), KiDS (yellow and green), and CANDELS (orange and red). There is no overall trend with redshift with SDSS having a lower merger fraction for the star-forming galaxies, KiDS having a higher merger fraction for star-forming galaxies, and CANDELS star-forming galaxies having a higher merger fraction at all redshifts.

galaxies than the star-forming galaxies, although the difference is 0.2 percentage points, much less than the error on the merger fractions. KiDS data has a higher merger fraction for the quiescent galaxies than the star-forming galaxies. As these two data sets cover similar redshift ranges one would expect to see a similar trend in the merger fractions of these two populations. The difference in overall merger fractions may be a result of the SDSS and KiDS networks not being identical and the different selection criteria for the training sets.

This is qualitatively different to the CANDELS-z000 data that has a slightly lower quiescent merger fraction than star-forming merger fraction. This difference is more pronounced at higher redshifts resulting in a different conclusion from the KiDS data. The CANDELS data suggests that there is a higher fraction of star-forming galaxy mergers than quiescent galaxy mergers at all redshifts, implying that galaxy mergers do not often act to suppress star-formation rates.

The CNNs used in this work are not perfect as they misclassify mergers as non-mergers and non-mergers as mergers. The latter of these misclassifications may present issues with our analysis. As non-mergers are more prevalent than mergers, relatively high specificity of a network can still result in a large population of non-merging galaxies being added to the merging classification. If galaxy mergers do significantly change the SFR of the galaxies, the non-merging interlopers may act to suppress this effect in the statistical analysis used in this paper. However, as this work is primarily comparing the relative SFRs of merging and non-merging galaxies, we do not believe that this overly impacts our results as the differences we see in star-formation rates between the mergers

and non-mergers is small.

5.5.3 Starburst merger fraction

We avoid using a specific definition of a starburst galaxy and instead opt to study the merger fraction as a function of distance above the MS. For ease of reference, we will refer to the galaxies above a given SFR threshold as starbursting in this subsection, even if the threshold is the MS. To this end, we study the fraction of star-forming galaxies above a certain distance above the MS that are merging for all three data sets (number of merging galaxies above a certain threshold divided by total number of galaxies above the same threshold). These trends are presented in Fig. 5.16.

The SDSS and KiDS data show an increase in the merger fraction as the distance from the MS increases, with SDSS rising to $\sim 1.1 \log(\text{M}_\odot \text{yr}^{-1})$ and KiDS slowly declining above $\sim 0.8 \log(\text{M}_\odot \text{yr}^{-1})$. A similar trend is seen in the CANDELS data, with an increase in merger fraction as the distance from the MS increases. CANDELS-z000 rises to approximately $0.3 \log(\text{M}_\odot \text{yr}^{-1})$ while the other three CANDELS redshift bins rise to approximately 0.8, 1.2, 1.4, and 1.6 for CANDELS-z060, CANDELS-z085, CANDELS-z121, and CANDELS-z166 respectively. Thus, the merger fraction increases as the star-formation rate increases showing that mergers can act to trigger high star-formation rates and starbursts. We note, however, that the number of galaxies in Fig. 5.16 decreases as the distance above the MS increases meaning that the lower merger fraction at lower distance can contain more mergers than the higher merger fraction at larger distances. This allows for the small changes in SFR seen in the star-forming population despite the merger fraction increasing as the distance above the MS increases. This is qualitatively consistent with (Luo et al. 2014), who find approximately half of starburst systems (defined as an increase in SFR by a factor of five or more) are undergoing a merger while the fraction of mergers in non-starburst systems is lower.

We can also compare the fraction of star-forming, merging galaxies that have SFRs above a certain distance above the MS (number of merging galaxies above a certain threshold divided by total number of merging galaxies) with the fraction of star-forming, non-merging galaxies that have SFRs above a certain distance above the MS as shown in Fig. 5.17. The SDSS, KiDS, and CANDELS data all show that a higher fraction of starburst mergers are found than the fraction of starburst non-mergers, although this switches at larger distances for the KiDS data. This is clear evidence that the merging galaxies are causing an increase in SFR.

5.6 Conclusions

Galaxy mergers are an important part of how galaxies grow and evolve over the history of the universe. However, identifying galaxy mergers is a difficult and time-consuming task. Here we have employed deep learning techniques to identify galaxy mergers in SDSS, KiDS, and CANDELS imaging data. We have then used these classifications to explore how galaxy mergers affect SFRs.

We find that mergers do indeed influence the SFR in the merging galaxies. However, the resulting change in SFR is small, typically a factor of ~ 1.2 . Within the SDSS data, the

Figure 5.16: Merger fraction for star-forming galaxies with SFRs above indicated distances above the MS for SDSS and KiDS data (top panel) and CANDELS data (bottom panel). To avoid low number statistics, only thresholds above which there are 50, or more, galaxies are shown. The SDSS (top panel, purple), KiDS (top panel, blue), and all CANDELS data show a trend of increasing merger fraction as the distance to the MS increases, although the CANDELS-z000 drops again above $0.62 \log(M_{\odot}\text{yr}^{-1})$.

Figure 5.17: Fraction of star-forming, merging galaxies with SFRs above given distances above the MS (solid lines) and fraction of star-forming, non-merging galaxies with SFRs above given distances above the MS (dashed lines). The top panel contains the SDSS and KiDS data sets while the bottom panel contains the CANDELS data. To avoid low number statistics, only thresholds above which there are 50, or more, galaxies are shown. The SDSS (top panel, purple), KiDS (top panel, blue), and all CANDELS data show that there is a higher fraction of the total number of merging galaxies above nearly all distance above the MS.

star-forming objects have a slight increase, on average, that is also seen in the KiDS data within a similar redshift range. Between $0.0 < z \leq 0.6$, the CANDELS data shows a slight decrease in SFR for the star-forming population when examining the MS subtracted SFR. Continuing to higher redshifts with the CANDELS data, we again find slight increases SFR for the merging galaxies with. Overall, the change seen in the SFR of the star-forming population is small, with the majority of changes in the SFR in all data sets being less than $3\sigma_\mu$, a factor of ~ 1.2 .

The merger fraction of quiescent and star-forming galaxies also depends on the data set. The SDSS data has a slightly high merger fraction for quiescent galaxies compared to star-forming galaxies while the KiDS and CANDELS data is the opposite. Again, definite conclusions are difficult with the CANDELS and KiDS data showing that galaxy mergers are more common in star-forming galaxies at any redshift while the SDSS data does not.

Instead of directly examining the fraction of starburst galaxies that are mergers, we examine the merger fraction as a function of distance above the MS. For the SDSS, CANDELS, and KiDS the fraction of mergers increases as the distance above the MS increases. This is evidence that mergers can cause periods of enhanced star formation.

Our current work does not determine the stage of the galaxy merger but we can see by eye that our merger samples include mergers at different stages. Thus, it is possible that the period during which SFR is boosted significantly is very short during the merging process and missed within our more time averaged analysis. It could also be that SFR is only boosted significantly for a small fraction of merger types or a combination of both scenarios. Future work will aim to overcome these shortcomings by determining the merger stage.

Acknowledgements. We would like to thank the anonymous referee for their thoughtful comments that have improved the quality of this paper.

We would like to thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high performance computing cluster.

MB is supported by the Polish Ministry of Science and Higher Education through grant DIR/WK/2018/12.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is <http://www.sdss.org/>.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

Based on observations made with ESO Telescopes at the La Silla Paranal Observatory under pro-

gramme IDs 177.A-3016, 177.A-3017, 177.A-3018, and 179.A-2004, and on data products produced by the KiDS consortium. The KiDS production team acknowledges support from: Deutsche Forschungsgemeinschaft, ERC, NOVA, and NWO-M grants; Target; the University of Padova, and the University Federico II (Naples).

GAMA is a joint European-Australasian project based around a spectroscopic campaign using the Anglo-Australian Telescope. The GAMA input catalogue is based on data taken from the Sloan Digital Sky Survey and the UKIRT Infrared Deep Sky Survey. Complementary imaging of the GAMA regions is being obtained by a number of independent survey programmes including GALEX MIS, VST KiDS, VISTA VIKING, WISE, Herschel-ATLAS, GMRT, and ASKAP providing UV to radio coverage. GAMA is funded by the STFC (UK), the ARC (Australia), the AAO, and the participating institutions. The GAMA website is <http://www.gama-survey.org/>

This work is based on observations taken by the CANDELS Multi-Cycle Treasury Program with the NASA/ESA HST, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555.

This research made use of Astropy,⁵ a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013, 2018).

⁵<http://www.astropy.org>

Bibliography

- Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, software available from tensorflow.org
- Ackermann, S., Schawinski, K., Zhang, C., Weigel, A. K., & Turp, M. D. 2018, MNRAS, 479, 415
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
- Athanassoula, E., Rodionov, S. A., Peschken, N., & Lambert, J. C. 2016, ApJ, 821, 90
- Barnes, J. E. 2004, MNRAS, 350, 798
- Barton, E. J., Geller, M. J., & Kenyon, S. J. 2000, ApJ, 530, 660
- Bell, E. F., Naab, T., McIntosh, D. H., et al. 2006, ApJ, 640, 241
- Berrier, J. C., Bullock, J. S., Barton, E. J., et al. 2006, ApJ, 652, 56
- Bershady, M. A., Jangren, A., & Conselice, C. J. 2000, AJ, 119, 2645
- Boquien, M., Burgarella, D., Roehlly, Y., et al. 2019, A&A, 622, A103
- Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151
- Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
- Chabrier, G. 2003, PASP, 115, 763
- Charlot, S., & Fall, S. M. 2000, ApJ, 539, 718
- Cibinel, A., Daddi, E., Sargent, M. T., et al. 2019, MNRAS, 485, 5631
- Cirasuolo, M., McLure, R. J., Dunlop, J. S., et al. 2007, MNRAS, 380, 585
- Conselice, C. J. 2003, ApJS, 147, 1
- . 2014, ARA&A, 52, 291
- Conselice, C. J., Bershady, M. A., Dickinson, M., & Papovich, C. 2003, AJ, 126, 1183
- Conselice, C. J., Bershady, M. A., & Jangren, A. 2000, ApJ, 529, 886

- Conselice, C. J., Yang, C., & Bluck, A. F. L. 2009, *MNRAS*, 394, 1956
- Cortijo-Ferrero, C., González Delgado, R. M., Pérez, E., et al. 2017, *A&A*, 607, A70
- Cotini, S., Ripamonti, E., Caccianiga, A., et al. 2013, *MNRAS*, 431, 2661
- da Cunha, E., Charlot, S., & Elbaz, D. 2008, *MNRAS*, 388, 1595
- Darg, D. W., Kaviraj, S., Lintott, C. J., et al. 2010a, *MNRAS*, 401, 1043
- . 2010b, *MNRAS*, 401, 1552
- Davies, A., Serjeant, S., & Bromley, J. M. 2019, *MNRAS*, arXiv:1905.04303
- Davies, L. J. M., Robotham, A. S. G., Driver, S. P., et al. 2015, *MNRAS*, 452, 616
- Davis, M., Guhathakurta, P., Konidakis, N. P., et al. 2007, *ApJ*, 660, L1
- de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013a, *Experimental Astronomy*, 35, 25
- de Jong, J. T. A., Kuijken, K., Applegate, D., et al. 2013b, *The Messenger*, 154, 44
- De Propris, R., Liske, J., Driver, S. P., Allen, P. D., & Cross, N. J. G. 2005, *AJ*, 130, 1516
- Dieleman, S., Willett, K. W., & Dambre, J. 2015, *MNRAS*, 450, 1441
- Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., et al. 2019, *MNRAS*, 484, 93
- Draine, B. T., & Li, A. 2007, *ApJ*, 657, 810
- Draine, B. T., Aniano, G., Krause, O., et al. 2014, *ApJ*, 780, 172
- Driver, S. P., Norberg, P., Baldry, I. K., et al. 2009, *Astronomy and Geophysics*, 50, 5.12
- Duncan, K., Conselice, C. J., Mundy, C., et al. 2019, *ApJ*, 876, 110
- Eales, S., Dunne, L., Clements, D., et al. 2010, *PASP*, 122, 499
- Edge, A., Sutherland, W., Kuijken, K., et al. 2013, *The Messenger*, 154, 32
- Elbaz, D., Daddi, E., Le Borgne, D., et al. 2007, *A&A*, 468, 33
- Ellison, S. L., Mendel, J. T., Patton, D. R., & Scudder, J. M. 2013, *MNRAS*, 435, 3627
- Ellison, S. L., Patton, D. R., Simard, L., & McConnachie, A. W. 2008, *AJ*, 135, 1877
- Ellison, S. L., Viswanathan, A., Patton, D. R., et al. 2019, *MNRAS*, 487, 2491
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
- Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, *ApJ*, 600, L93
- Griffin, M. J., Abergel, A., Abreu, A., et al. 2010, *A&A*, 518, L3

- Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, *ApJS*, 197, 35
- Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, *ApJS*, 207, 24
- Holwerda, B. W., Kelvin, L., Baldry, I., et al. 2019, *AJ*, 158, 103
- Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2006, *ApJS*, 163, 1
- Huertas-Company, M., Gravet, R., Cabrera-Vives, G., et al. 2015, *ApJS*, 221, 8
- Hurley, P. D., Oliver, S., Betancourt, M., et al. 2017, *MNRAS*, 464, 885
- Ioffe, S., & Szegedy, C. 2015, International Conference on Machine Learning
- Joseph, R. D., & Wright, G. S. 1985, *MNRAS*, 214, 87
- Kartaltepe, J. S., Mozena, M., Kocevski, D., et al. 2015, *ApJS*, 221, 11
- Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, *MNRAS*, 341, 33
- Kim, E. J., & Brunner, R. J. 2017, *MNRAS*, 464, 4463
- Kim, J.-h., Wise, J. H., & Abel, T. 2009, *ApJ*, 694, L123
- Kingma, D. P., & Ba, J. 2015, 3rd International Conference for Learning Representations, arXiv:1412.6980
- Knapen, J. H., Cisternas, M., & Querejeta, M. 2015, *MNRAS*, 454, 1742
- Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, *ApJS*, 197, 36
- Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, *ApJS*, 192, 18
- Kroupa, P. 2001, *MNRAS*, 322, 231
- Kuijken, K., Heymans, C., Hildebrandt, H., et al. 2015, *MNRAS*, 454, 3500
- Kuijken, K., Heymans, C., Dvornik, A., et al. 2019, *A&A*, 625, A2
- Lackner, C. N., Silverman, J. D., Salvato, M., et al. 2014, *AJ*, 148, 137
- Lambas, D. G., Tissera, P. B., Alonso, M. S., & Coldwell, G. 2003, *MNRAS*, 346, 1189
- Lang, M., Holley-Bockelmann, K., & Sinha, M. 2014, *ApJ*, 790, L33
- Larson, D., Dunkley, J., Hinshaw, G., et al. 2011, *ApJS*, 192, 16
- Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, *MNRAS*, 379, 1599
- Lecun, Y., Bengio, Y., & Hinton, G. 2015, *Nature*, 521, 436
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998, in *Proceedings of the IEEE*, Vol. 86, Issue 11, 2278–2324
- Lin, L., Patton, D. R., Koo, D. C., et al. 2008, *ApJ*, 681, 232

- Lin, L., Cooper, M. C., Jian, H.-Y., et al. 2010, *ApJ*, 718, 1158
- Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, *MNRAS*, 389, 1179
- López-Sanjuan, C., Balcells, M., Pérez-González, P. G., et al. 2010, *A&A*, 518, A20
- López-Sanjuan, C., Cenarro, A. J., Varela, J., et al. 2015, *A&A*, 576, A53
- Lotz, J. M., Jonsson, P., Cox, T. J., et al. 2011, *ApJ*, 742, 103
- Lotz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2010a, *MNRAS*, 404, 590
- . 2010b, *MNRAS*, 404, 575
- Lotz, J. M., Primack, J., & Madau, P. 2004, *AJ*, 128, 163
- Lotz, J. M., Davis, M., Faber, S. M., et al. 2008, *ApJ*, 672, 177
- Loveday, J., Norberg, P., Baldry, I. K., et al. 2012, *MNRAS*, 420, 1239
- Luo, W., Yang, X., & Zhang, Y. 2014, *ApJ*, 789, L16
- Lupton, R., Blanton, M. R., Fekete, G., et al. 2004, *PASP*, 116, 133
- Moreno, J., Torrey, P., Ellison, S. L., et al. 2019, *MNRAS*, 485, 1320
- Mundy, C. J., Conselice, C. J., Duncan, K. J., et al. 2017, *MNRAS*, 470, 3507
- Naab, T., Khochfar, S., & Burkert, A. 2006, *ApJ*, 636, L81
- Nair, V., & Hinton, G. E. 2010, in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 807–814
- Nayyeri, H., Hemmati, S., Mobasher, B., et al. 2017, *ApJS*, 228, 7
- Niemi, S.-M., Somerville, R. S., Ferguson, H. C., et al. 2012, *MNRAS*, 421, 1539
- Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, *ApJ*, 660, L43
- Noll, S., Burgarella, D., Giovannoli, E., et al. 2009, *A&A*, 507, 1793
- Pearson, W. J., Wang, L., Trayford, J. W., Petrillo, C. E., & van der Tak, F. F. S. 2019, *A&A*, 626, A49
- Pearson, W. J., Wang, L., van der Tak, F. F. S., et al. 2017, *A&A*, 603, A102
- Pearson, W. J., Wang, L., Hurley, P. D., et al. 2018, *A&A*, 615, A146
- Perez, J., Michel-Dansac, L., & Tissera, P. B. 2011, *MNRAS*, 417, 580
- Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2017, *MNRAS*, 472, 1129
- Pozzetti, L., Bolzonella, M., Zucca, E., et al. 2010, *A&A*, 523, A13
- Randall, S. W., Markevitch, M., Clowe, D., Gonzalez, A. H., & Bradač, M. 2008, *ApJ*, 679, 1173

- Robotham, A. S. G., Driver, S. P., Davies, L. J. M., et al. 2014, *MNRAS*, 444, 3986
- Rodrigues, M., Puech, M., Flores, H., Hammer, F., & Pirzkal, N. 2018, *MNRAS*, 475, 5133
- Rodriguez-Gomez, V., Snyder, G. F., Lotz, J. M., et al. 2019, *MNRAS*, 483, 4140
- Rodríguez-Puebla, A., Primack, J. R., Avila-Reese, V., & Faber, S. M. 2017, *MNRAS*, 470, 651
- Rupke, D. S. N., Kewley, L. J., & Barnes, J. E. 2010, *ApJ*, 710, L156
- Saitoh, T. R., Daisaka, H., Kokubo, E., et al. 2009, *PASJ*, 61, 481
- Salim, S., Rich, R. M., Charlot, S., et al. 2007, *ApJS*, 173, 267
- Sanders, D. B., & Mirabel, I. F. 1996, *ARA&A*, 34, 749
- Santini, P., Ferguson, H. C., Fontana, A., et al. 2015, *ApJ*, 801, 97
- Schweizer, F. 2005, in *Astrophysics and Space Science Library*, Vol. 329, *Starbursts: From 30 Doradus to Lyman Break Galaxies*, ed. R. de Grijs & R. M. González Delgado, 143
- Schweizer, F. 2009, in *Globular Clusters - Guides to Galaxies*, ed. T. Richtler & S. Larsen (Berlin, Heidelberg: Springer Berlin Heidelberg), 331–338
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, 172, 1
- Silva, A., Marchesini, D., Silverman, J. D., et al. 2018, *ApJ*, 868, 46
- Sinha, M., & Holley-Bockelmann, K. 2012, *ApJ*, 751, 17
- Smith, M. W. L., Ibar, E., Maddox, S. J., et al. 2017, *ApJS*, 233, 26
- Somerville, R. S., & Davé, R. 2015, *ARA&A*, 53, 51
- Sparre, M., & Springel, V. 2016, *MNRAS*, 462, 2418
- Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, *ApJS*, 214, 15
- Springel, V., Di Matteo, T., & Hernquist, L. 2005, *MNRAS*, 361, 776
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2014, *Journal of Machine Learning Research*, 15, 1929
- Stefanon, M., Yan, H., Mobasher, B., et al. 2017, *ApJS*, 229, 32
- Toomre, A., & Toomre, J. 1972, *ApJ*, 178, 623
- Weigel, A. K., Schawinski, K., Treister, E., Trakhtenbrot, B., & Sanders, D. B. 2018, *MNRAS*, 476, 2308
- Whitaker, K. E., Labbé, I., van Dokkum, P. G., et al. 2011, *ApJ*, 735, 86

- Wright, A. H., Robotham, A. S. G., Driver, S. P., et al. 2017, *MNRAS*, 470, 283
- Wu, K. L., Faber, S. M., & Lauer, T. R. 2001, in *Deep Fields*, ed. S. Cristiani, A. Renzini, & R. E. Williams, 170
- York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, *AJ*, 120, 1579

5.7 Appendix

5.7.1 De-blending the SPIRE data

For de-blending the SPIRE data, we follow Pearson et al. (2017). CIGALE is used with the 9-band KiDS catalogue data to generate estimates of the SPIRE flux densities and we then select all objects with a predicted $250\ \mu\text{m}$ flux density above 1.1 mJy. The CIGALE flux density estimates are then used as a flux density prior inside XID+ and all three SPIRE bands in the GAMA09 field are de-blended.

For our CIGALE models, we follow Pearson et al. (2018) but remove the active galactic nuclei component, due to the limited wavelength coverage available, and increase the sampling of the age of the stellar population. Thus, we use a double exponentially declining star-formation history, Bruzual & Charlot (2003) stellar emission, Chabrier (2003) initial mass function (IMF), Charlot & Fall (2000) dust attenuation, and the updated Draine et al. (2014) version of the Draine & Li (2007) infrared dust emission. A list of parameters, where they differ from the default values, can be found in Table 5.7.1.

Table 5.7: Parameters used for various properties in CIGALE model SEDs where they differ from default values. All ages and times are in Gyr.

Parameter	Value	Star-Formation History	Description
τ_{main}	1.0, 1.8, 3.0, 5.0, 7.0		e-folding time (main)
τ_{burst}	9.0, 13.0		e-folding time (burst)
f_{burst}	0.00, 0.10, 0.20, 0.30, 0.40		Burst mass fraction
Age	0.50, 1.00, 1.50, 2.00, 2.50, 3.00, 3.50, 4.00, 4.50, 5.00, 5.50, 6.00, 6.50, 7.00, 7.50, 8.00, 8.50, 9.00, 9.50, 10.00, 10.50, 11.00, 12.00, 13.00		Population age (main)
Burst Age	0.001, 0.010, 0.030, 0.100, 0.300		Population age (burst)
Stellar Emission			
IMF	Chabrier (2003)		Initial Mass Function
Z	0.02		Metallicity (0.02 is Solar)
Separation Age	0.01		Separation between young and old stellar populations
Dust Attenuation			
A_V^{BC}	0.3, 1.2, 2.3, 3.3, 3.8		V-band attenuation of the birth clouds
Slope _{BC}	-0.7		Birth cloud attenuation power law slope
BC to ISM Factor	0.3, 0.5, 0.8, 1.0		Ratio of the birth cloud attenuation to ISM attenuation
Slope _{ISM}	-0.7		ISM attenuation power law slope
Dust Emission			

q_{PAH}	0.47, 1.12, 2.50, 3.9	Mass fraction of PAH
U_{min}	5.0, 10.0, 25.0	Minimum scaling factor of the radiation field intensity
α	2.0	Dust power law slope
γ	0.02	Illuminated fraction

5.7.2 CNN performance definitions

This paper uses the definitions of Pearson et al. (2019) for the terms to describe the properties of CNNs. These terms may be an alternate nomenclature to other works or may be unfamiliar. To avoid confusion we reproduce these definitions in Table 5.8.

5.7.3 Example non-mergers and mergers

Here we present example non-mergers and mergers as defined by the CNN. The images shown are all 64×64 pixel images with gri composite for SDSS, Fig. 5.18, greyscale r-band for KiDS, Fig. 5.19, and 1.6 μm , 1.25 μm , 814 nm composite for CANDELS, Fig. 5.20.

Table 5.8: Terms used when describing neural network performance from Pearson et al. (2019)

Term	Definition	
Positive (P)	An object classified in the catalogues or identified by a network as a merger.	
Negative (N)	An object classified in the catalogues or identified by a network as a non-merger.	
True Positive (TP)	An object classified in the catalogues as a merger that is identified by a network as a merger.	
False Positive (FP)	An object classified in the catalogues as a non-merger that is identified by a network as a merger.	
True Negative (TN)	An object classified in the catalogues as a non-merger that is identified by a network as a non-merger.	
False Negative (FN)	An object classified in the catalogues as a merger that is identified by a network as a non-merger.	
Recall	Fraction of objects correctly identified by a network as a merger with respect to the total number of objects classified in the catalogues as mergers.	$TP / (TP+FN)$
Fall-out	Fraction of objects incorrectly identified by a network as a merger with respect to the total number of objects classified in the catalogues as mergers.	$FP / (TP+FN)$
Specificity	Fraction of objects correctly identified by a network as a non-merger with respect to the total number of objects classified in the catalogues as non-mergers.	$TN / (TN+FP)$
Precision	Fraction of objects correctly identified by a network as a merger with respect to the total number of objects identified by a network as a merger.	$TP / (TP+FP)$
Negative Predictive Value (NPV)	Fraction of objects correctly identified by a network as a non-merger with respect to the total number of objects identified by a network as a non-merger.	$TN / (TN+FN)$
Accuracy	Fraction of objects, both merger and non-merger, correctly identified by a network.	$(TP+TN) / (TP+FP+TN+FN)$

Figure 5.18: Examples of non-merging galaxies (top row) and merging galaxies (bottom nine rows) for SDSS data set as defined by the CNN. Images are gri composite with a size of 64×64 pixel or 13.7×13.7 arcsec.

Figure 5.19: Examples of non-merging galaxies (top row) and merging galaxies (bottom nine rows) for KiDS data set as defined by the CNN. Images are greyscale r-band with a size of 64×64 pixel or 25.3×25.3 arcsec.

Figure 5.20: Examples of non-merging galaxies (top row) and merging galaxies (bottom nine rows) for CANDELS data set as defined by the CNN. Images are $1.6\ \mu\text{m}$, $1.25\ \mu\text{m}$, $814\ \text{nm}$ composite with a size of 64×64 pixel or 3.8×3.8 arcsec.

Conclusions and Future Prospects

6.1 Conclusions

Throughout this thesis, I have looked into the physical drivers of the cosmic star formation history. This has been done in a number of ways. Star formation rate estimates have been improved by improving the flux density measurements from existing *Herschel* SPIRE (Griffin et al. 2010) observations. From this, a detailed study of the relation between star formation rate and stellar mass has been undertaken which covers over 75% of the history of the universe. Using the latest machine learning techniques, how galaxy mergers affect star formation rates has also been studied.

This thesis began with improving existing de-blending methods so that far infrared (IR) flux density measurements can be more accurate and reliable (Chapter 2). The far IR part of the spectrum is a key component in determining star formation rates as it is emission from dust that has been heated by young stars. Thus, the pipeline that was developed allows better constraints on star formation rates in galaxies and allows us to explore fainter and more distant galaxies than was possible before.

The de-blending work became a key component in studying the galaxy main sequence (Chapter 3); the observed tight correlation between stellar mass and star formation rate for galaxies that are considered to be star forming. The improved star formation rate estimates have allowed an investigation over one of the largest redshift ranges for a self-consistent galaxy main sequence study ever undertaken.

Work then moved to study the relation between star formation rates and galaxy mergers. This required a method to identify a large sample of galaxy mergers, resulting in the employment of deep learning techniques. The original plan was to use data from one of the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011) fields to train a neural network to apply to the rest of the CANDELS fields. However, the small number of training sources resulted in this being a difficult task and plans were made to augment the training sample with galaxies from the cosmological hydrodynamic simulation Evolution and Assembly of GaLaxies and their Environments (EAGLE; Schaye et al. 2015). In doing so, this raised the question of whether the simulated galaxy mergers were representative of real galaxy mergers and lead to a study comparing observed galaxy mergers from Sloan Digital Sky Survey (SDSS; York et al. 2000) with galaxy mergers from EAGLE (Chapter 4). For this, a neural network was trained once with SDSS mergers and once with EAGLE mergers. With these networks, it was possible to examine how and why the merger classifications failed as well as attempt to apply the EAGLE trained network to SDSS data and the SDSS trained network to EAGLE data. The results of this cross application offer hope for using this technique on large upcoming surveys.

With this chapter completed, thoughts turned back to studying the relation between star formation and galaxy mergers (Chapter 5). This work brought together all the skills and methods grown throughout this thesis: the de-blending of far IR data, the galaxy main sequence and deep learning. The SDSS trained network from Chapter 4 could be used again in this work and the network was trained with two further data sets: CANDELS was returned to and Kilo Degree Survey (KiDS; de Jong et al. 2013a,b) was added. A large sample of merging galaxies were found, several hundred thousand, and were placed onto the galaxy main sequence, along with their non-merging counterparts, using de-blended far IR data to derive star formation rates. The conclusions for all four science chapters are

presented individually by chapter, followed by an overall concluding summary and future prospects.

6.1.1 Improvement of existing far-infrared flux density measurement - Chapter 2

In Chapter 2, I investigated how the flux density measurements from the three *Herschel* SPIRE bands can be made more accurate and reliable. This was done by expanding the bayesian de-blending tool XID+ (Hurley et al. 2017). A new flux density prior was introduced to XID+ that allows the inclusion of informative prior data where the original flux density prior was uninformative. A pipeline was developed that took existing, higher resolution photometry that did not suffer from blending and confusion and fitted spectral energy distributions to these data. From these spectral energy distributions, the flux densities of the three SPIRE bands could be predicted and used as a prior in XID+.

To test this methodology, *Herschel* data from the The Cosmic Evolution Survey (COSMOS; Scoville et al. 2007) extragalactic field was de-blended with and without the new informed prior. The results of both de-blendings were compared with Atacama Large Millimeter/submillimeter Array (ALMA) data at 870 and 1250 μm , again using spectral energy distribution fitting. It was found that the new, informed prior performed better, with a 27% increase in the number of galaxies within 1σ of the ALMA data point. This is clear evidence that de-blending techniques are improved with the inclusion of an informed flux density prior.

By improving the accuracy and reliability of the far IR flux density measurements it is possible to better constrain the total IR emission from a galaxy and hence provide more reliable star formation rate estimates. With improved star formation rate estimates, we can gain a better understanding of how galaxies have built up their stellar mass and evolved over the history of the universe. The main sequence of star forming galaxies is one such place these improved star formation rate estimates can help refine our understanding.

6.1.2 The relation between star formation rate and stellar mass - Chapter 3

In Chapter 3, I investigated the relation between star formation rate and stellar mass: the main sequence of star forming galaxies. For this, the techniques developed in Chapter 2 were employed and over 200 000 galaxies within the COSMOS field had SPIRE data de-blended. The star formation rates of these galaxies were then estimated with the aid of these data. With the de-blended data and depth of the ancillary COSMOS data, it was possible to study the form and evolution of the main sequence of star forming galaxies from $z = 0.2$ to $z = 6.0$, one of the longest redshift ranges for a self-consistent galaxy main sequence study. Forward modelling techniques were developed and used to find the main sequence of star forming galaxies.

Unlike a number of other works, we find no evidence of a high mass turnover in the main sequence of star forming galaxies. This suggests that the galaxy main sequence is linear in log space at all masses and there is no high mass quenching of star formation above stellar masses of $\approx 10^{10} M_{\odot}$. The normalisation of the main sequence was found to increase linearly with look-back time, consistent both qualitatively and quantitatively

with other works. The increasing normalisation indicates that galaxies at higher redshifts were forming stars faster than they are today, possibly as gas is depleted as the universe ages. For the slope of the main sequence, it was also found to increase with redshift. The slope of the main sequence of star forming galaxies connects the high mass and low mass galaxies. An increasing slope with redshift indicates that at earlier times (higher redshifts) high mass and low mass galaxies had similar specific star formation rates (star formation rate/stellar mass) but as the universe aged, the high mass galaxies' specific star formation rate reduced faster than their low mass counterparts.

An increase in the slope of the galaxy main sequence was found at $1.8 \leq z < 2.9$, although the evidence for this is tentative. If present, this increase in the slope coincides with the peak of the co-moving star formation rate density suggesting that the two may be connected. However, a more secure detection of this increase must be found before any further investigation into this link is warranted.

Through the work undertaken in this chapter, the effect of different methods of determining star formation rate was also briefly studied. This work used star formation rates derived from spectral energy distribution fitting and was compared to work that used ultraviolet plus infrared (UV+IR) derived star formation rate. Directly comparing the results of these two studies shows that the results presented in this thesis had a lower normalisation than the UV+IR galaxy main sequence. However, if the same UV+IR method was employed on the data used here, the two galaxy main sequences become equivalent, demonstrating that the results of a main sequence study, or any study that relies upon star formation rate, is sensitive to how the star formation rates are derived.

6.1.3 Detecting galaxy mergers with deep learning - Chapter 4

In Chapter 4, I investigated how deep learning, specifically convolutional neural networks, can be used to identify galaxy mergers. A convolutional neural network architecture was developed and trained twice, once with images from the SDSS and once with images from the EAGLE cosmological, hydrodynamical simulation. The quality of the results of the SDSS and EAGLE trained networks were compared along with cross application: passing SDSS imaged through the EAGLE trained network and vice versa.

The results from the SDSS trained network were very good, achieving an accuracy of 92% while the accuracy of the EAGLE trained network was lower at 65%. It was unclear why some SDSS objects were misclassified, there being no trend in physical or observable parameter. The misclassified EAGLE galaxies were primarily at low simulation snapshots and faint apparent magnitudes. The observed SDSS noise that is used when generating the EAGLE images causes galaxies with faint apparent magnitude to have very low signal to noise. This, coupled with the fact that the EAGLE network uses faint and diffuse features of a galaxy to assign a correct classification, results in the EAGLE network not finding the features it looks for in merging galaxies. The SDSS noise also creates a large number of chance projections for non-merging galaxies that then are misclassified as merging. The number of chance projections in the EAGLE images appears higher than in the SDSS images. Thus, in future work care would be needed to ensure the number of chance projections in the simulation images is more in line with observations.

Cross application of the networks was not as successful. Passing the EAGLE images through the SDSS trained network resulted in an accuracy of 53% with preferential assign-

ment to the non-merger class. This suggests that simulations are lacking in the later stage mergers that make up the majority of the SDSS mergers. Late stage mergers are where two galaxies are clearly interacting with each other and are beginning to coalesce. Passing SDSS images through the EAGLE trained network, however, was more successful, achieving an accuracy of 65%. The misclassified galaxies were typically found to have high specific star formation rate, for non-mergers classified as mergers, or low specific star formation rates, for mergers classified as non-mergers. This suggests that merging galaxies in simulations have an increase in star formation rate that may not be seen in real galaxies. The EAGLE training set is also more complete: it has more stages of merger than the observationally selected SDSS galaxies. Thus, it is possible that some of the misclassified non-mergers are in fact merging systems that were not selected by eye.

The results from running the SDSS images through the EAGLE trained network present a tantalising prospect for large upcoming surveys. The volume of data generated in future surveys will be so large that human classification will be difficult or at least incomplete. Being able to train a convolutional neural network to identify the type or morphology of a galaxy using simulations and applying it to observations will reduce the need for human classification to create a training sample, massively speeding up the classification of objects.

6.1.4 The relation between galaxy mergers and star formation rate - Chapter 5

Chapter 5 draws from the work undertaken in the three preceding chapters. The deep learning techniques developed in Chapter 4 were employed to detect galaxy mergers in the CANDELS and the KiDS data, as well as using the SDSS trained network to detect merging galaxies within the whole of SDSS. To improve the star formation rate estimates of the KiDS data, the de-blending technique developed in Chapter 2 was used to de-blend *Herschel* SPIRE data within the regions of KiDS that is covered by the Galaxy And Mass Assembly survey (GAMA; Driver et al. 2009). The forward modelling technique developed in Chapter 3 was used again in this chapter to derive the main sequence of star forming galaxies in all three surveys. Once mergers were identified in all three surveys, the mass corrected star formation rates of the merging and non-merging galaxies were compared. The mass correction was done in two ways: once by calculating the specific star formation rate and once by subtracting the star formation rate of the galaxy main sequence at a galaxy's stellar mass. These two corrections are equivalent if the slope of the galaxy main sequence is unity.

The effect of merging galaxies on the star formation rate is small, rarely over a factor of two change in star formation rate. However, whether mergers appear to cause an increase or decrease on the star formation rate depends on the data set employed. The SDSS and KiDS data cover similar redshift ranges ($0.0 < z \leq 0.15$) and show similar merger effects. The SDSS data show that merging star forming galaxies have a slight increase in star formation rate. This is consistent with the KiDS data which also shows an increase in the star formation rate for the merging star forming population. At higher redshifts with the CANDELS data, out to $z = 4.0$, there is no clear trend of increases or decreases in star formation rate with redshift.

As the distance above the galaxy main sequence increases, the fraction of galaxies

that are merging also increases for the SDSS, KiDS and CANDELS data. Generally, this supports the notion that galaxy mergers can cause highly enhanced star formation rates. However, these are the minority of galaxy mergers and it is more likely to observe a merging galaxy in the period between starbursts.

6.1.5 Overall conclusions

In the introduction to this thesis, three questions were posed that this thesis aimed to answer. It is now possible to answer these questions.

How is the star formation rate of a galaxy related to its stellar mass and how does this evolve with time?

The star formation rate is tightly correlated, with stellar mass with this relation being linear in log space. As redshift increases, the normalisation of this relation also increases as does the slope. Thus, at high redshift, high mass and low mass galaxies have similar specific star formation rates with the specific star formation rates of high mass galaxies reducing as the universe ages. The overall star formation rates of all galaxies at all stellar masses have also reduced as the universe has evolved.

How do galaxy mergers influence the star formation rates of the colliding galaxies?

Galaxy mergers do not have a large influence on the star formation rates of the colliding galaxies for the majority of merging systems. The star formation rate is changed by at most a factor of a few in the majority of galaxies, and this can be both an increase or decrease with different data sets providing different changes. However, a small number of merging galaxies do show a large increase in star formation rate and can be considered as starburst galaxies.

Can we use images from simulations to identify merging galaxies in the real universe?

It is possible to use images from hydrodynamical simulations to train a convolutional neural network to identify merging galaxies in the real universe. However, the performance is not as good as when identifying real galaxy mergers with a network trained on real galaxy mergers. It is likely that this performance can be improved by improving the realism of the simulated galaxy images by including effects of dust within the simulated galaxies. More careful selection of the position to inject the simulated galaxies into real images will also aid with reducing misclassifications. Using the simulated galaxy mergers to identify real mergers can also identify mergers that would not be selected by visual inspection.

6.2 Future prospects

6.2.1 Improvements for far infrared flux density measurement

The work undertaken to improve the flux densities from *Herschel* SPIRE is likely to provide some of the best flux density measurements in the FIR for the next decade or so. There is room for minor improvements, such as including the correlation between the

three SPIRE bands that has been developed at Sussex University since the publication of Chapter 2¹. However, these gains are marginal and at the cost of a large amount of computing time. It is therefore likely that the method used to de-blended *Herschel* data developed in this thesis will remain the best method to measure state of the art FIR flux densities until another FIR space mission is flown.

New FIR space missions would ideally have higher sensitivity instruments and have higher angular resolution, allowing deeper observations that are able to resolve individual galaxies out to higher redshifts than is currently possible. Ground based observations in this regime, such as those using the Submillimetre Common-User Bolometer Array 2 (SCUBA-2; Holland et al. 2013), suffer from atmospheric absorption, although they can achieve higher angular resolution than possible with *Herschel*. The proposed space mission SPICA (Roelfsema et al. 2018) is likely to be the next revolution for FIR observations. If selected, SPICA will provide spectroscopic and photometric observations that are over an order of magnitude more sensitive than *Herschel* (Roelfsema et al. 2018), allowing the observation of fainter or higher redshift galaxies. The proposed Origins Space Telescope (Battersby et al. 2018) will also offer improvements. This 9.1 or 5.9 m telescope aims to cover wavelengths from 5 to 600 μm and will offer up to three times the spatial resolution of *Herschel* resulting in fewer issues with confusion and blending.

6.2.2 The main sequence of star forming galaxies

The galaxy main sequence also still has open issues. There is contention about the shape of the galaxy main sequence, with some studies suggesting that the galaxy main sequence is linear in log space, such as the work presented here in Chapter 3, and others suggesting that there is a high mass turnover. Working to correct for the differences in star formation rate may help with this issue, a problem highlighted in Chapter 3, as well as developing a consistent method to select star forming galaxies. Methods of selecting star forming galaxies that are more generous in their definition will allow some quiescent objects into the star forming sample which may act to artificially create the high mass turnover seen in some works (Johnston et al. 2015).

6.2.3 Identifying galaxy mergers

The cross application of simulation trained neural networks has been shown in Chapter 4 to be promising for application with large surveys. However, this cross application does not perform as well as using an observationally trained network on observations, probably due to the realism of the simulated images. Improvements in the generation of the simulated images will help to mitigate this issue. Care with including all physical processes found within galaxies, such as the dust attenuation, will help increase the realism of the simulated images. As mentioned in Chapter 4, increasing the mass resolution will also help increase the realism of the simulated galaxies while using the black hole merger trees will help to improve the timing of the galaxy mergers, opening the way to use the simulations to identify the merger stage.

If the realism of the current EAGLE images is improved, for example by taking care to ensure the number coincidental galaxy projections in the EAGLE images matches the

¹ XID+ including the correlation between bands is available at https://github.com/H-E-L-P/XID_plus

number in real observations, the same study could be re-run. Once the images from existing simulations are as realistic as they can feasibly get without re-running the entire simulations, a study can be undertaken to compare the simulation mergers with observed mergers. This will allow us to better understand where current simulations are doing well at reproducing the real universe and where they are not performing so well. From this, the next generation of simulations will be able to be improved and produce more realistic results, furthering our understanding of galaxy evolution.

It would also be insightful to use other simulations to generate the images of galaxy mergers and perform similar analysis to that done in Chapter 4. For example, simulated galaxy images from Illustris (Genel et al. 2014; Vogelsberger et al. 2014) or Illustris The Next Generation (IllustrisTNG; Pillepich et al. 2018) could be used. If their performance differs from that of EAGLE in this thesis, it would allow us to explore the differences in the simulations, including the sub-grid physics, and compare them with a view to increasing the realism of subsequent large cosmological hydrodynamic simulations.

The study undertaken in Chapter 5 also ignores the stage of the merger. It is possible that a large increase in star formation rate is found in mergers at a specific point within the merger, as suggested by simulations. However, the work in Chapter 5 assumes that all merger stages are self-similar. Including the merger stage into this analysis could prove to be enlightening into how galaxy mergers affect star formation rates by understanding how star formation rates change, if they change at all, as a merger progresses.

Bibliography

- Battersby, C., Armus, L., Bergin, E., et al. 2018, *Nature Astronomy*, 2, 596
- de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013a, *Experimental Astronomy*, 35, 25
- de Jong, J. T. A., Kuijken, K., Applegate, D., et al. 2013b, *The Messenger*, 154, 44
- Driver, S. P., Norberg, P., Baldry, I. K., et al. 2009, *Astronomy and Geophysics*, 50, 5.12
- Genel, S., Vogelsberger, M., Springel, V., et al. 2014, *MNRAS*, 445, 175
- Griffin, M. J., Abergel, A., Abreu, A., et al. 2010, *A&A*, 518, L3
- Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, *ApJS*, 197, 35
- Holland, W. S., Bintley, D., Chapin, E. L., et al. 2013, *MNRAS*, 430, 2513
- Hurley, P. D., Oliver, S., Betancourt, M., et al. 2017, *MNRAS*, 464, 885
- Johnston, R., Vaccari, M., Jarvis, M., et al. 2015, *MNRAS*, 453, 2540
- Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, *ApJS*, 197, 36
- Pillepich, A., Springel, V., Nelson, D., et al. 2018, *MNRAS*, 473, 4077
- Roelfsema, P. R., Shibai, H., Armus, L., et al. 2018, *PASA*, 35, e030
- Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, *MNRAS*, 446, 521
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, 172, 1
- Vogelsberger, M., Genel, S., Springel, V., et al. 2014, *MNRAS*, 444, 1518
- York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, *AJ*, 120, 1579

Acknowledgements

As I am sure many are aware, I am not the most emotive person in the world. Despite this, there are people in the world who I am grateful for their existence in my life and people who I owe a debt of gratitude that I may never be able to repay. I probably also lack the ability to write this down onto paper in a coherent or eloquent way.

Lingyu. Thank you for letting me come over to the Netherlands and work with you, even if it was a bit odd to beat you in the move over here. This thesis was only possible through your guidance. I have appreciated all of your comments and criticisms (even if I was not necessarily very receptive at the time) and you have helped make me a better scientist. I don't think I could have wished for a better supervisor. I hope we can continue to collaborate in the future.

Floris. As with Lingyu, your guidance was invaluable. Your expertise lie in a slightly different area but your help was always useful. Your comments and suggestions were always insightful. As this project progressed we had fewer and fewer meetings and I hope this is because we didn't need them.

My office mates, old and new. Chris, Jorrit, Daniel, Davide, Aku, Samira and Saina. Thanks for putting up with my occasional moments of frustration, queries about random Dutch words and the odd correction of your English pronunciation (when asked). I have enjoyed the conversations we have shared over the last few years as well as the dinners. You'll have to wait five more minutes before we can go to lunch though. Also, big thanks to Jorrit for helping with my Dutch summary!

Members of the ISM Group (and Sub-mm group before that). Thank you for putting up with me talking about galaxies. I know your interests lie more in things inside our own galaxy but you always seemed interested when I prattled on about galaxy scale star formation or galaxy mergers. It was also interesting to see what people do closer to home and a different side of astrophysics.

The secretariat at both SRON and Kapteyn. Thank you for helping with the requests and queries I have had over the last four and a bit years. Without you, life here would be much harder.

All the PhD students, postdocs and staff. I hesitate to name you all or this will turn into one massive list. Your enlightening talks over lunch and the occasional coffee are always interesting. You make Kapteyn the fun and friendly place it is.

My family. Mum, Dad, Lizzie and Chris. Thanks for your support and for being there on the rare occasions I need to vent to someone. And to Grandma, for always believing in me. I am sad you could not see me finish but I know you would have been proud.

Last but not least you, dear reader. This work is pointless if no one else reads it, even if it's just the Acknowledgements section (you know who you are).