Galaxy Zoo Builder: Four Component Photometric decomposition of Spiral Galaxies Guided by Citizen Science

Timothy K. Lingard [iD],[1] Karen L. Masters [iD],[2] Coleman Krawczyk [iD],[1] Chris Lintott [iD],[3] Sandor Kruk [iD],[4] Brooke Simmons [iD],[5] Robert Simpson [iD],[6] Steven Bamford,[7] Robert C. Nichol [iD],[1] and Elisabeth Baeten[8]

[1]*Institute of Cosmology and Gravitation, University of Portsmouth*
Dennis Sciama Building, Burnaby Road, Portsmouth, PO1 3FX, UK
[2]*Haverford College, 370 Lancaster Ave., Haverford, PA 19041, USA*
[3]*Oxford Astrophysics, Denys Wilkinson Building, Keble Road, Oxford, OX1 3RH, UK*
[4]*European Space Agency, ESTEC, Keplerlaan 1, NL-2201 AZ, Noordwijk, The Netherlands*
[5]*Physics Department, Lancaster University, Lancaster, LA1 4YB, UK*
[6]*Google UK, Six, Pancras Square, London N1C 4AG*
[7]*Centre for Astronomy & Particle Theory, School of Physics & Astronomy*
University of Nottingham, Nottingham, NG7 2RD, UK
[8]*Independent Zooniverse Volunteer*

Abstract

Multi-component modelling of galaxies is a valuable tool in the effort to quantitatively understand galaxy evolution, yet the use of the technique is plagued by issues of convergence, model selection and parameter degeneracies. These issues limit its application over large samples to the simplest models, with complex models being applied only to very small samples. We attempt to resolve this dilemma of "quantity or quality" by developing a novel framework, built inside the Zooniverse citizen science platform, to enable the crowdsourcing of model creation for SDSS galaxies. We have applied the method, including a final algorithmic optimisation step, on a test sample of 198 galaxies, and examine the robustness of this new method. We also compare it to automated fitting pipelines, demonstrating that it is possible to consistently recover accurate models which either show good agreement with, or improve on, prior work. We conclude that citizen science is a promising technique for modelling images of complex galaxies, and release our catalogue of models.

1. INTRODUCTION

Disc galaxies are complex objects, containing many different components, including a disc, disc phenomena (i.e. spiral arms, bars and rings) and central structures (bulges, bars). Decomposing disc galaxies into their component structures has become an important tool for extragalactic astronomers seeking to understand the formation and evolution of the galaxy population, ranging from analysing bulge and bar structure (Elmegreen & Elmegreen 1985, de Jong 1996, Gadotti 2011, Mendez-Abreu et al. 2016, Gao & Ho 2017, Kruk et al. 2018) to the secular evolution of disc galaxies (Lilly et al. 1998, Barden et al. 2005) and general galaxy assembly and evolution (Simard et al. 2002a, Bamford et al. 2011, Lackner & Gunn 2012, Rampazzo et al. 2019).

These fully quantitative methods allow researchers to obtain structural parameters of galaxy sub-components, which has been used in a variety of astrophysical and cosmological research. For example, the stellar mass found in discs and bulges places strong constraints on the galaxy merger tree from ΛCDM N-body simulations (Parry et al. 2009, Hopkins et al. 2010, Rodrigues et al. 2018); the strength of a galaxy's classical bulge is thought to be tied to the strength of

Corresponding author: Timothy K. Lingard
tklingard@gmail.com

merger events in its past (Springel & Hernquist 2005, Kormendy et al. 2010); different spiral arm formation theories vary in their predictions of spiral morphology (Dobbs & Baba 2014, Pour-Imani et al. 2016, Hart et al. 2017).

The usefulness of obtaining parametric models of a galaxy has motivated the creation of many image modelling and fitting suites, including GIM2D (Simard et al. 2002b), GALFIT (Peng et al. 2002), MEGAMORPH (Bamford et al. 2011) and PROFIT (Robotham et al. 2016) to name a few. Using these tools, researchers have built large catalogues of model fits to galaxies. One of the largest photometric model catalogs is that of Simard et al. (2011), who performed automated 2D, two-component (bulge + disc) decomposition of 1,123,718 galaxies from the Legacy imaging of the Sloan Digital Sky Survey (hereafter SDSS) Data Release 7 (Abazajian et al. 2009).

Many other large catalogues of photometric fits exist (i.e. Lackner & Gunn 2012, Kelvin et al. 2012, van der Wel et al. 2012), but despite the usefulness of photometric fitting, and the presence of analytic profiles and methods for modelling more complex galaxy sub-components, relatively few studies have attempted to perform large-scale (1000s of galaxies) parametric decomposition of galaxies using more complicated models than that of Simard et al. (2011) (two axisymmetric Sérsic components representing a bulge and disc). Not properly taking into account these "secondary" morphological features (such as a bar, ring and spiral arms) can impact detailed measurements of a galaxy's bulge (Gao & Ho 2017). Proper decomposition of secondary morphological features allows investigation into mechanisms behind the secular evolution of galaxies (Head et al. 2015, Kruk et al. 2018, Gao et al. 2018) and exploration of environmental effects on morphology, such as offset bars (Kruk et al. 2017).

A prominent issue when performing these detailed decompositions is the tendency for fitting functions to converge on unphysical results when not properly guided or constrained, for instance in a two-component model containing a Sérsic bulge and an exponential disc, the bulge may grow to encompass the galaxy's disc, as its extra parameter allows for a closer fit (as observed by Kruk et al. 2018). It is also the case that often, without near-optimal starting points, detailed model fits will fail to converge at all (Lange et al. 2016).

Compounding this, uncertainties reported by many software fitting packages (i.e. GALFIT and MEGAMORPH from the above list) are often lower estimates on the real uncertainty, due to secondary sources not being modelled, flat-fielding errors and incorrect models (including the possibly incorrect assumption of Poisson noise) (Peng et al. 2010). Other packages such as GIM2D and PROFIT attempt to fully model posterior distributions and so produce more representative uncertainties, however this comes with a larger computational cost and configuration complexity. Formal uncertainties are measures of the likelihood space and therefore underestimates of the true error as an analytic model will rarely capture the nuanced light profile of a galaxy.

Another problem which needs to be addressed is whether a component should be present in the model at all. An automated fit will generally attempt to add as many components as possible to produce the closest-matching model. Many studies therefore need to select the most appropriate model by visual inspection of the resulting residuals or recovered parameters. For example, both Vika et al. (2014) and Kruk et al. (2018) inspected the resulting model and residual images for all of their parametric fits (163 and 5,282 respectively) to ensure physical results with the correct components present. The end result of most of these problems is that researchers will have to invest time to individually check many of their fits to ensure they have converged on a physical model. In the era of large sky surveys such as the SDSS, the time required to do this becomes unsustainable and introduces concerns over human error if done by only a single, or small number of individuals.

A demonstrably successful solution to the similar problem of galaxy classification in the era of large surveys, was to find a new source of person-power: Lintott et al. (2008) invited large numbers of people to classify SDSS-images of galaxies over the internet in the Galaxy Zoo project. The resulting classifications (a mean of 38 per galaxy) were then weighted and averaged to create a morphological catalogue of 893,212 galaxies. This hugely successful project, including its subsequent iterations and expansions (i.e. Willett et al. 2013, Hart et al. 2016, Willett et al. 2017, Simmons et al. 2017), has produced a large catalogue of detailed morphological classifications which are in good agreement with other studies, and have been used in a wide variety of studies of the local galaxy population (see Masters & the Galaxy Zoo Team 2019 for a recent review).

In this paper we explore an analogous solution to that Lintott et al. (2008) brought to galaxy classification for the issues faced by galaxy parametric modelling, inside the ecosystem that Galaxy Zoo set in motion (namely *The Zooniverse*[1]). We leverage citizen scientists to pick model components and perform model optimization in an online,

---

[1] https://www.zooniverse.org

web-browser environment[2]. We describe our method in Section 2, including details of the images and ancillary data from SDSS as well as the strategy used to obtain scientifically useful models from volunteer classifications. We provide consistency checks within our infrastructure and to other methods in Section 3 and discuss the efficacy and potential impact of our new method relative to existing methodologies in Section 4.

Where necessary, we make use of $H_0 = 70$ km s^{-1} Mpc^{-1}.

2. METHOD

2.1. *The Galaxy Builder Zooniverse project*

Galaxy Builder is a citizen-science project built on the Zooniverse web platform. It asks volunteers to perform detailed photometric modelling of spiral galaxies (potentially including bulge, disc, bar and spiral arm components). A project of this kind, allowing volunteers to interact with and model data, had never been attempted inside the current Zooniverse web platform before, so this project involved designing and implementing a model rendering[3] suite inside the existing Zooniverse front-end code-base. As with all citizen science solutions, we had to not only consider the accuracy of the resulting model, but also user experience and engagement in our design decisions.

The closest relative to this project within the Zooniverse ecosystem was the Galaxy Zoo: Mergers project (Holincheck et al. 2016). This project asked volunteers to help match the morphological properties of an image of merging galaxies to a plethora of restricted three-body simulations. Galaxy Zoo: Mergers required volunteers to download a Java applet to take part in model selection, while *Galaxy Builder* operates purely inside a web page.

2.1.1. *Project Timeline and Development*

The *Galaxy Builder* project was built inside the Zooniverse's (Simpson et al. 2014) Panoptes-Front-End[4] code-base, using the React.js[5] framework, as well as WebGL[6] to enable low-latency photometric galaxy model rendering. *Galaxy Builder* entered a Zooniverse beta in late November 2017 and after some user experience improvements and code refactoring, the project was launched as an official Zooniverse project on the 24th of April 2018.

A major challenge during development of the project was finding the right balance between keeping a simple and intuitive interface and workflow while also allowing the freedom and versatility to properly model galaxies. It was also a significant challenge to develop a compelling and simple tutorial for what is one of the most complex projects attempted on the Zooniverse platform. Feedback from expert users was essential to this process as part of the typical beta trial process for Zooniverse projects[7].

2.1.2. *The project interface*

The *Galaxy Builder* project prompts volunteers to work through the step-by-step creation of a photometric model of a galaxy (described in detail in Section 2.4). A screenshot of the interface can be seen in Figure 1, where a residual image is shown. The interface presents a volunteer with three views, which they can switch between at any time: a *r*-band cutout image of a spiral galaxy (see Section 2.2), the galaxy model they have created so far, and the residual between their model and image (shown in blue and yellow).

The workflow is designed so that volunteers slowly subtract increasing amounts of light from the galaxy, as is illustrated in Figure 2. A tutorial is available which contains a step-by-step guide to completing a classification. At each step volunteers are asked to first draw a simple isophote, and then make use of a series of sliders to adjust the parameters of the model component (see Section 2.4 for more information).

Volunteers are also guided by a "score", which is tied to the residuals and chosen to increase from zero to some arbitrary value depending on the galaxy; a less noisy and more easily modelled galaxy will have a higher maximum score. To map a residual image to a final score shown to volunteers we used

$$S = 100 \exp\left(\frac{-A}{N} \sum_{i=0}^{N} \frac{\text{arcsinh}^2 \left(|y_i - M_i| \, / \, 0.6 \right)}{\text{arcsinh}\, 0.6} \right), \tag{1}$$

---

[2] https://www.zooniverse.org/projects/tingard/galaxy-builder

[3] We use the term rendering in a similar manner to that used for computer graphics: to calculate an image from a model or set of rules.

[4] http://github.com/zooniverse/Panoptes-Front-End

[5] https://reactjs.org/

[6] https://www.khronos.org/webgl/

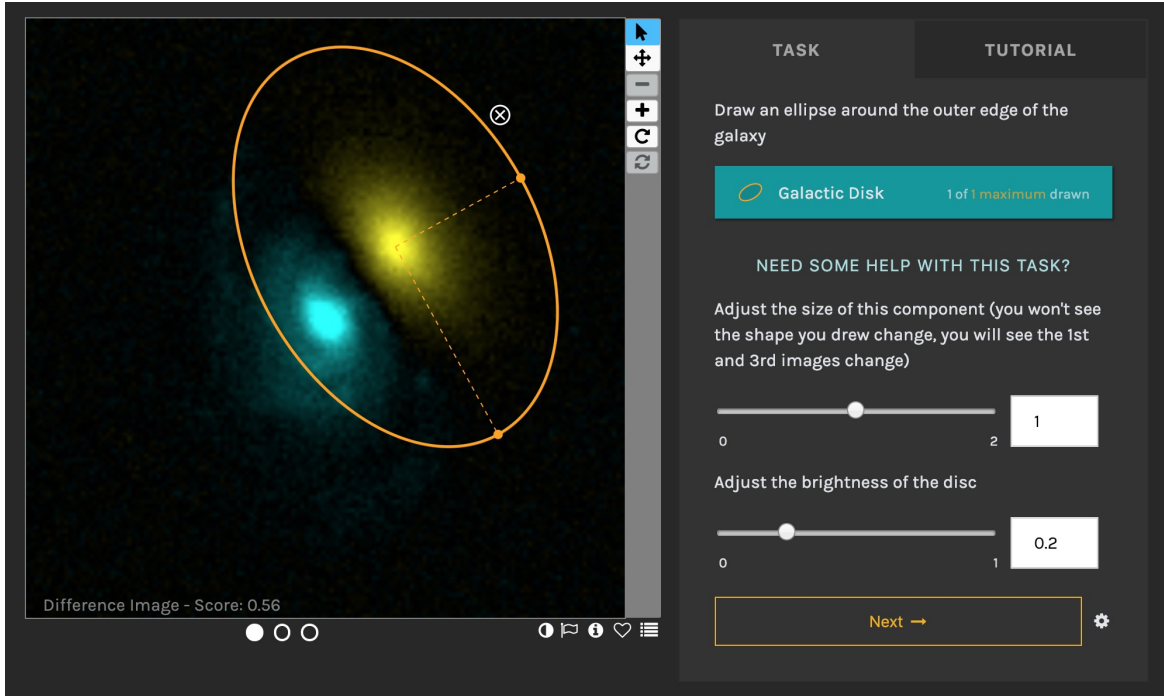[7] https://help.zooniverse.org/best-practices/

Figure 1. The *Galaxy Builder* interface. The residual image is being shown, and the volunteer is on the "Disc" task. The drawn disc component (yellow) is offset from the galaxy image (blue) to demonstrate the positive and negative residuals. Where the image equals the model the residual is black. The dots below the residual image allow the user to switch images. The icons to the right allow panning and zooming of the image (rotation was not functional for this project). The icons to the bottom right of the image allow colour inversion of the galaxy cutout, flagging of the image as inappropriate, inspection of galaxy metadata (i.e. sky position, link to SDSS SkyServer), ability to save the image as a favourite and to add to a Zooniverse "collection". The Score shown in the bottom left of the image is calculated using Equation 1 and is a rough goodness-of-fit measure.

where N is the total number of pixels, y is the cutout of the galaxy, normalized to a maximum value of 1 ($y =$ cutout/max(cutout)), M is the model calculated by volunteers and $A = 300$ is an arbitrary choice of scaling chosen based on a handful of test galaxies.

This score has the advantage of being easy (and fast) to generate from the residual image shown to volunteers (which was Arcsinh-scaled in a manner described by Lupton et al. 2004), however it is overly sensitive to small deviations of the model from the galaxy.

2.2. *Sample Selection: Images and Ancillary Data*

As a proposed solution to the problem of fitting multi-component and complex galaxies, *Galaxy Builder* finds a niche with a sample of disc galaxies with spiral features. One such sample is the *stellar mass-complete sample* in Hart et al. (2017), which is a sample of face-on spiral galaxies with and without bars and selected to be complete across stellar masses $9.45 < \log(M_\star/M_\odot) < 11.05$. The test sample we use for the Galaxy Builder project was therefore selected from the Hart et al. (2017) sample of face-on spiral galaxies.

The morphological information required to select spiral galaxies came from the public data release of Galaxy Zoo 2 (Willett et al. 2013, hereafter GZ2). Each response to a GZ2 morphology question is allocated a p value ranging from 0 to 1, where 0 indicates no volunteers responded positively to that question and 1 indicates all volunteers who classified that galaxy responded positively (i.e. $p_{\rm bar} = 0.5$ would indicate 50% of volunteers said a bar was present in a galaxy). Photometric measurements used for selection came from the NASA-Sloan Atlas (Blanton et al. 2011, hereafter NSA). The *stellar mass complete sample* is constructed using the set of criteria detailed in Table 1.

The *stellar mass-complete sample* was split into smaller sub-samples, each containing 100 galaxies. In an iterative process, each sub-sample was chosen to contain the 60 lowest redshift unclassified galaxies, and 40 random unclassified galaxies. This was done to ensure we would have an early sample to work with given the *a priori* unknown rate at which volunteers would provide classifications. Due to time constraints, classifications were only collected for two unique sub-samples. The mass-redshift relation of galaxies in the *stellar mass-complete sample* from Hart et al. (2017)
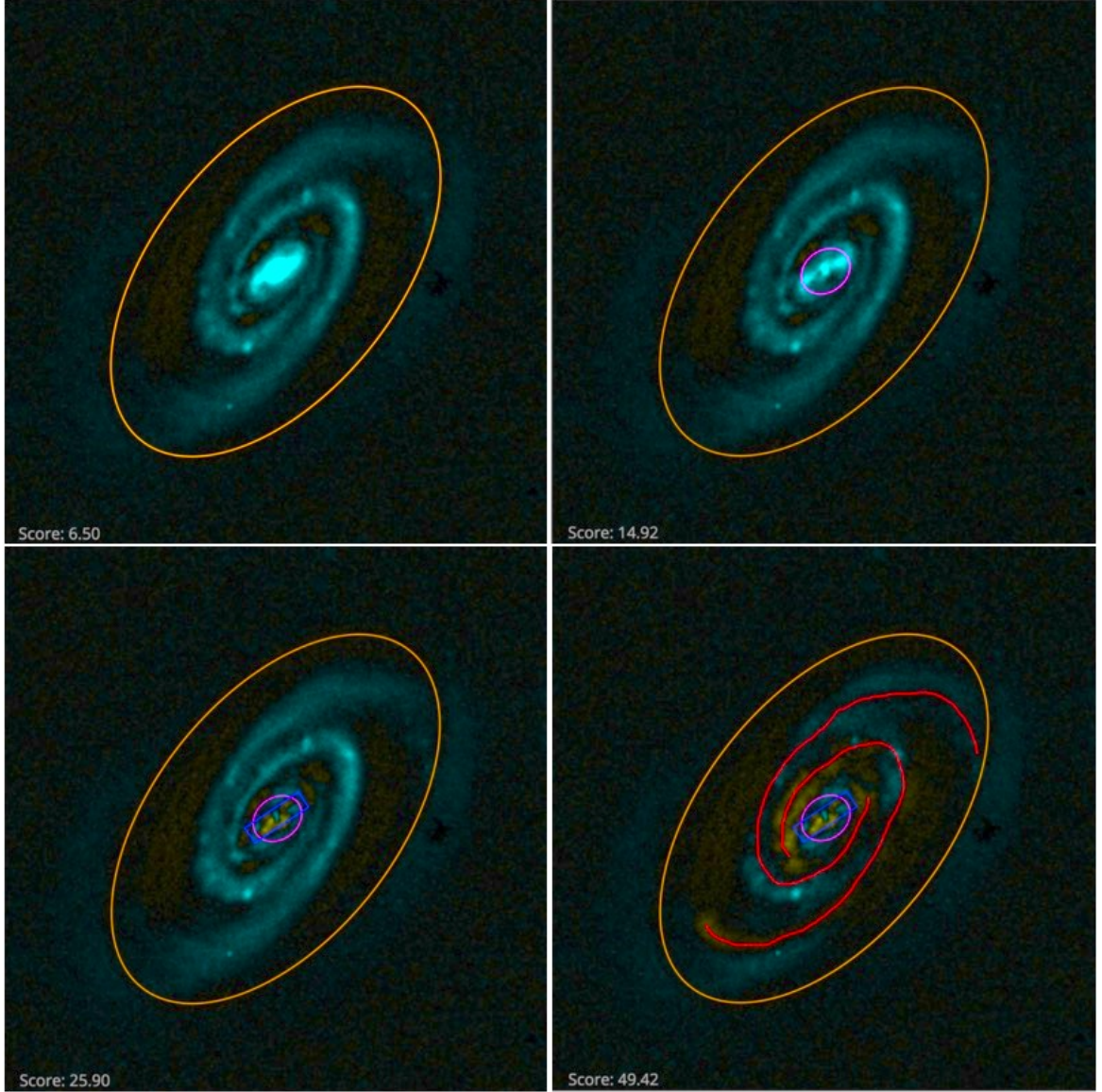
Figure 2. Figure demonstrating the desired result of each step of the modelling process, as seen from the residual image provided to volunteers. The top left panel shows the galaxy after only a disc component has been added: the top right contains a disc and a bulge; the bottom left has a disc, bulge and bar; the bottom right is the finished model with a disc, bulge, bar and spiral arms. The image shown is SDSS J104238.12+235706.8. This brightness and contrast of this image have been edited to improve visibility in print.

can be seen in Figure 3, with galaxies present in this work highlighted in red. Stellar Masses were calculated by Mendel et al. (2014).

In the first two sets of 100 galaxies, 1% of galaxies (i.e. 2 images) failed to run through the image preparation process, due to an error when attempting to montage multiple frames. The root cause of this error is unknown, but it leaves a sample of 198 galaxies with images (the *test sample*, 98 of which are repeated in a validation subset) that are considered in this paper, in order to explain the method used and test the reliability of the models obtained.

2.2.1. *Image and modelling metadata extraction*

The galaxy data shown to volunteers in the *Galaxy Builder* project came in two forms: A gray-scale image cutout of the galaxy and a JSON file containing rendering information for the web-interface.

Both forms of data were obtained using a similar process:

Table 1. The selection criteria used in Hart et al. (2017) to create the *stellar mass-complete sample* of 6222 spiral galaxies.

Description	Value
Face-on spiral morphological selection.	GZ2 $p_{\text{features}} \times p_{\text{not edge on}} \times p_{\text{spiral}} \geq 0.5$
Redshift limits.	$0.02 < z < 0.055$
Face-on galaxy selection using g-band axial ratio.	$(b/a)_g > 0.4$
Mass limits for rough volume limited sample.	$9.45 < \log(M_*/M_\odot) \leq 11.05$
Mass limits for complete sample[a]	$2.07 \log(z) + 12.64 < \log(M_*/M_\odot) < 2.45 \log(z) + 14.05$

[a]Stellar masses from Mendel et al. (2014)

[a]Using the method of Pozzetti et al. (2009) and limits calculated by Hart et al. (2017)
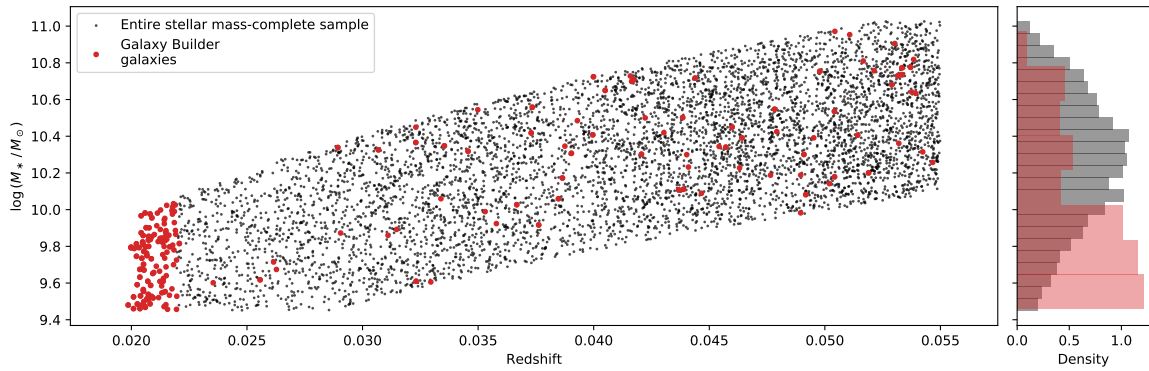
Figure 3. Redshift against total galaxy stellar mass for all galaxies in the *stellar mass-complete sample*, with the 198 galaxies considered in this paper highlighted in red. The distribution of stellar masses is shown in the right panel for the total sample and for the galaxies considered here. It is evident that the galaxies for which we collected classifications are not complete in stellar mass, but it is possible to select a further subset which would be.

1. A montage of multiple r-band corrected frames from the SDSS DR13 (Albareti et al. 2017) data release was created. To combine multiple FITS images, we made use of Astropy (Astropy Collaboration et al. 2018), and the MONTAGE (Jacob et al. 2010) software package.

2. This montage was cropped to four times the Petrosian radius of the galaxy.

3. The SEXTRACTOR software (Bertin & Arnouts 1996) was used to identify regions containing secondary sources (foreground stats, other galaxies) and generate a mask.

4. A PSF was obtained from the relevant Sloan r-band `psField` file, extracted at the central position of the galaxy (Stoughton et al. 2002).

5. The JSON file was written containing the cut-out data and the 2D boolean mask obtained from the source extraction process. This file also contained other metadata needed for the rendering process (PSF, the size of the PSF array, and the width and height of the image).

6. Another JSON file containing simply the information used to render the volunteer's model (image size and PSF) was created.

7. An arcsinh-stretch was applied to the masked cutout (as described by Lupton et al. 2004). It was then saved as a grey-scale image.

The decision to use r-band images in our subject set was due to its higher signal-to-noise than other bands.

Once a sub-sample had been created, the Zooniverse's panoptes-python-client[8] was used to upload them as a subject-set to the Zooniverse.

The reprojection performed by Montage has a smoothing effect on the data, and thus does not conserve errors. We therefore create a separate stacked image, sigma image and corresponding pixel mask, using the same r-band corrected frames present in the montage. These images were not shown to volunteers but were used for model fitting and comparison.

2.3. *Choice of Retirement limit*

The number of independent answers needed to create reliable and reproducible aggregate classifications was not known at the start of this project. An initial experiment with collecting 10 classifications per galaxy demonstrated that this was insufficient; further experimentation with a diverse range of galaxy types (most with prominent spiral features including grand-design and flocculent arms) revealed 30 classifications per galaxy was sufficient.

The entire *test sample* of 198 galaxies was then presented to users, with 30 classifications collected per galaxy. In addition, one of the subsets was presented a second time, thus providing a validation subset to measure consistency between sets of 30 classifications on the same galaxies.

We also created 9 synthetic images of galaxies, containing various combinations of components available to volunteers and a spread of possible parameters. These synthetic galaxies were based off of a set of target galaxies from *Galaxy Builder* and designed to be as realistic as possible, including the addition of realistic noise and pixel masks. This set of synthetic images is shown in Figure 4 and was used to calibrate our aggregation and fitting methodology and thus is referred to as the *calibration subset*.

2.4. *The Galaxy Model*

Our chosen galaxy model was largely based on components described in Peng et al. (2002). The modelling code ignores masked regions identified as secondary sources by SExtractor. It over-samples the bulge, disc and bar components by a factor of five and performs PSF convolution using a PSF obtained from the relevant Sloan r-band `psField` file, extracted at the central position of the galaxy (Stoughton et al. 2002). The model created by a volunteer could be chosen from

1. One exponential, ellipsoidal disc.

2. One ellipsoidal Sérsic bulge, with n chosen by volunteers.

3. One Sérsic bar with a "boxiness" modifier (as described in Peng et al. 2002), with n and c chosen by volunteers.

4. Any number of freehand poly-line[9] spiral arms, as described below.

2.4.1. *Spiral Arm Model*

Each spiral arm is modelled using a poly-line drawn by the volunteer. The brightness of a spiral arm at any point is given by the value of a Gaussian centred at the nearest point on any drawn poly-line, with volunteers able to choose the Gaussian width and peak brightness using sliders. Radial falloff was added by multiplying by the value of the previously added exponential disc, though volunteers could change the half-light radius of this falloff disc.

2.5. *Classification Aggregation Methodology*

In this Section, we will use the galaxy UGC 4721, a two-armed barred spiral galaxy at $z = 0.02086$ classified by de Vaucouleurs et al. (1991) as SBcd, to illustrate the data reduction and aggregation methodology. For UGC 4721 we received 32 classifications, containing 28 discs, 24 bulges, 17 bars and 47 drawn spiral arm poly-lines (four classifications did not contain spirals, seven contained one spiral arm, fourteen contained two arms, six contained three arms and one contained four arms). These annotations can be seen in Figure 5, overlaid on the greyscale r-band image of the galaxy.

---

[8] https://github.com/zooniverse/panoptes-python-client

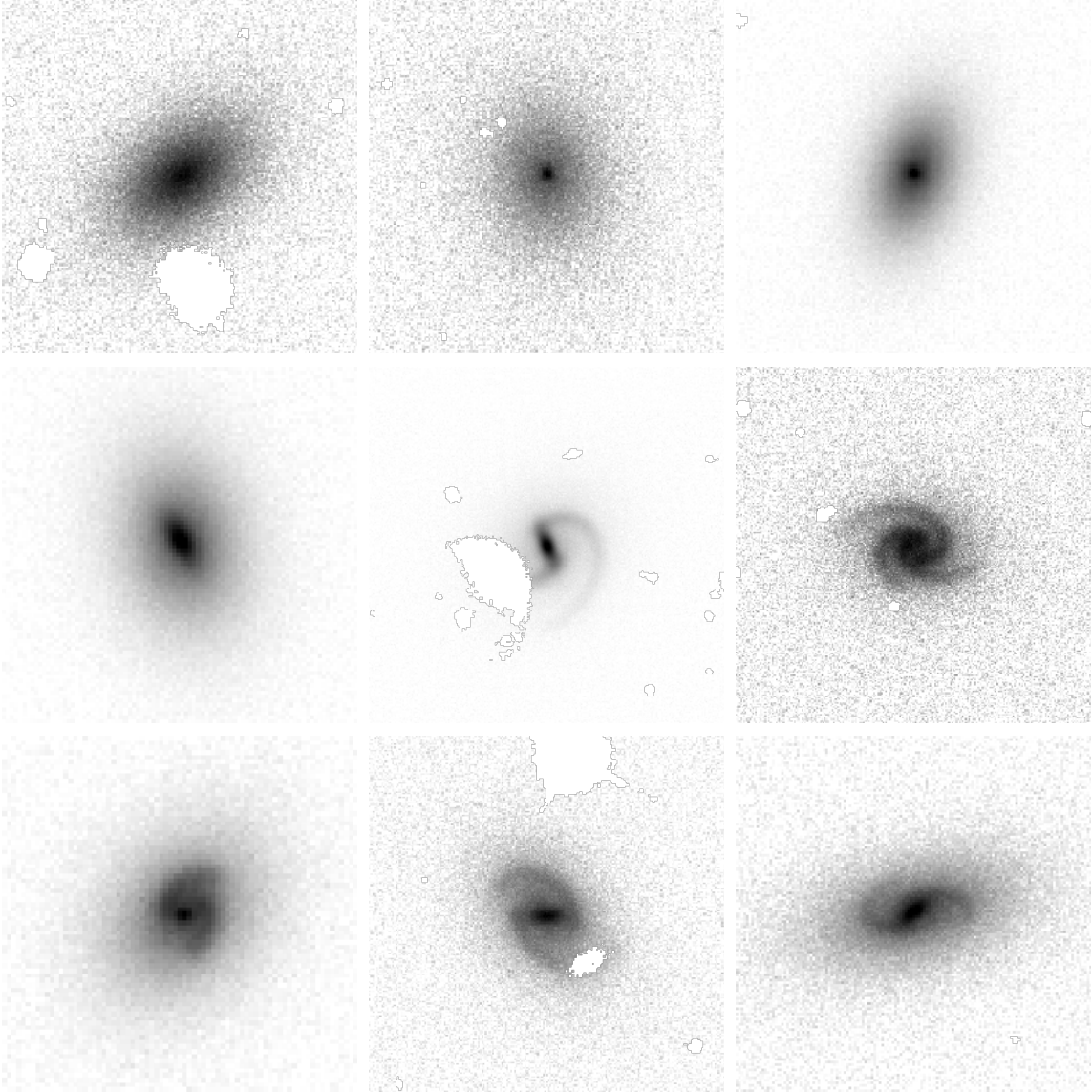[9] a poly-line, or polygonal chain, is a series of connected line segments.

Figure 4. Arcsinh-stretched images of the synthetic galaxies present in the *calibration subset*. These galaxies were designed to look as realistic as possible, while being described perfectly by the model available to volunteers.

2.5.1. *Aggregation of Volunteer Models*

Aggregate model calculation was done on a component-by-component basis, rather than per classification, i.e. clustering of discs was performed independently to that of bulges, bars and spirals. We did not take into account any slider values, only the shape drawn by the volunteers. Disk classifications were doubled in effective radius to correct for a systematic error in disk size observed in the classifications received for the *calibration subset*. Models parameters were clipped to be within the limits shown in Table 2 (deemed to be the physically acceptable bounds). All components were transformed from the coordinate space of the MONTAGE-created images to the more accurate stacked images created for fitting. Clustering was performed using the Jaccard distance measure (also known as the intersect-over-union distance, or IOU distance), which is a simple metric determining the relative shared area of two sets:

$$d_J(A, B) = 1 - \frac{|A \cap B|}{|A \cup B|}. \tag{2}$$

The algorithm chosen to perform clustering was the density-based spatial clustering of applications with noise (DBSCAN, Boonchoo et al. 2018) algorithm, due to its robustness and speed. We made use of Scikit-learn (Pedregosa
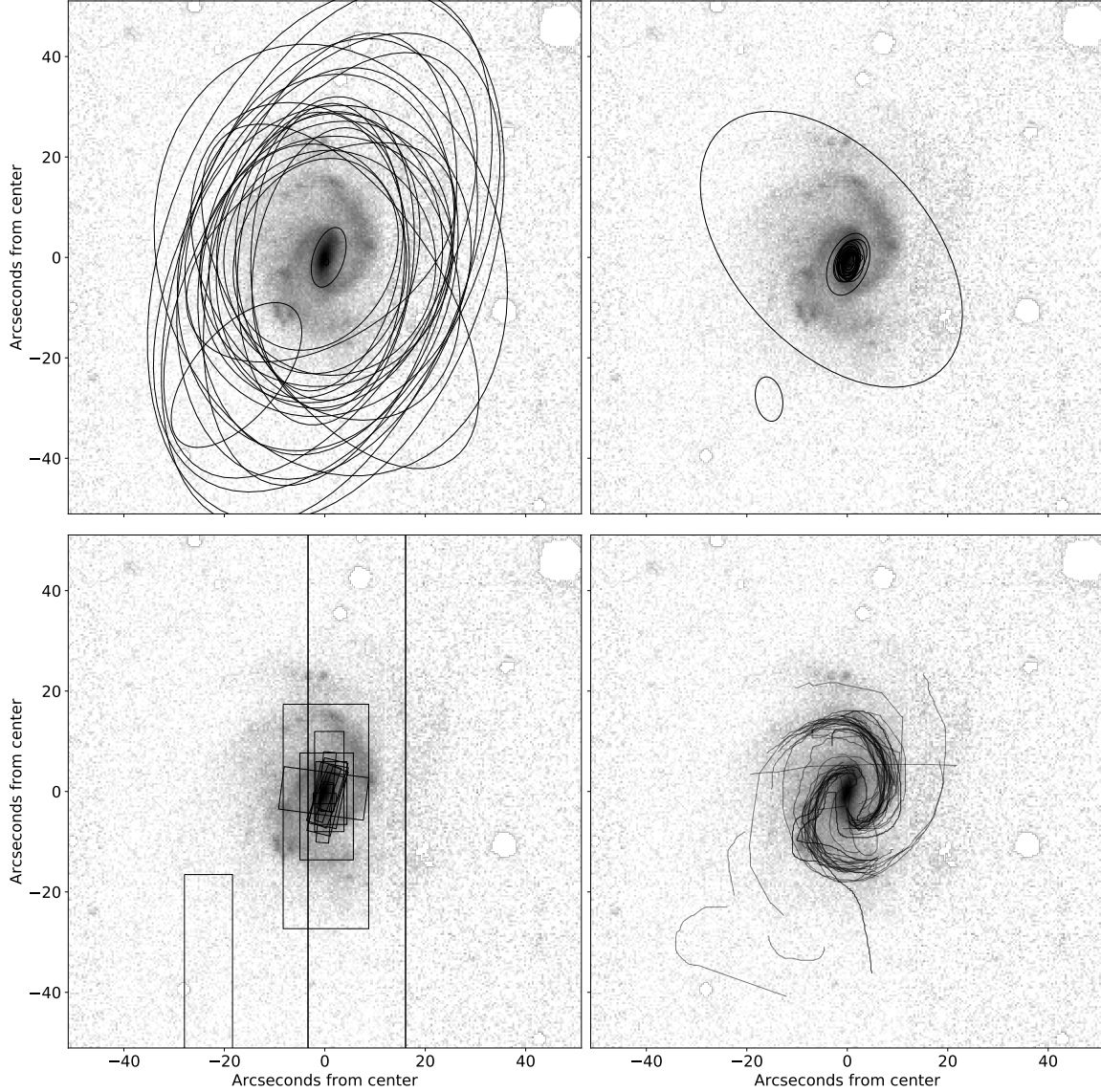
Figure 5. Components drawn by volunteers for UGC 4721. The top left panel shows drawn discs, top right shows drawn bulges, bottom left shows drawn bars and bottom right shows drawn spiral arms. Discs, bulges and bars are displayed at twice their effective radii. These raw marks are subsequently aggregated to produce a consensus value for each galaxy component.

et al. 2011) to implement the algorithm. In DBSCAN the core of a cluster is defined as a group of at least `min_points` that are all within a distance `eps` of each other. Additionally, any points within a distance `eps` of a cluster's core are also associated with the cluster.

2.5.2. *Disc, Bulge and Bar Clustering*

We select the disc clustering hyperparameters such that a disc is clustered for all galaxies, and the bulge hyperparameters to most successfully recover the morphology of galaxies in the *calibration subset*. The value of `eps` used to cluster bars was tuned such that the aggregate model best agreed with GZ2 p_{bar} ($p_{bar} < 0.2$ implying no bar and $p_{bar} > 0.5$ implying a definite bar). The values chosen for `eps` were 0.3, 0.4, 0.478 for the disc, bulge and bar; `min_points` was set to 4 for all three of these components.

We define the aggregate component to be the shape which minimises the sum of Jaccard distances to each of the members of the cluster. For our example galaxy, UGC 4721, clustered and aggregate components can be seen in Figure 6.

2.5.3. *Spiral Arm Clustering*

To cluster drawn spiral arms, we define a custom separation measure to represent how far away one poly-line is from another. This measure was chosen to be the mean of the squared distances from each vertex in a poly-line to the nearest point (vertex or edge) of another poly-line, added to the mean of the squared distances from the second poly-line to the first. We make use of this separation measure inside the DBSCAN algorithm to cluster these drawn lines, after removing any self-intersecting drawn arms (as this was deemed an easy method to filter out "bad" classifications). Values of 0.001 and 4 were used for the `eps` and `min_samples` hyper-parameters respectively.

Once spiral classifications on a galaxy have been clustered into the physical arms they represent, the points are deprojected using the axis ratio and position angle of the aggregated disc. The deprojection method assumes a thin disc and stretches the ellipsoidal minor axis to match the major axis.

Deprojected points within each drawn poly-line are converted to polar coordinates and unwound to allow model fitting. These unwound points are then cleaned using the Local-outlier-factor algorithm (LOF, Breunig et al. 2000). For each drawn poly-line in the cluster, the LOF algorithm was trained on all points not in that arm, and then used to predict whether each point in the arm should be considered an outlier. In this way we clean our data while respecting its grouped nature. The points removed as outliers for the example galaxy are shown in the bottom right panel of Figure 6.

For each arm cluster in each galaxy, a logarithmic spiral model was fitted using Bayesian Ridge Regression, performed using the Scikit-learn python package. A logarithmic spiral was chosen due to its simple form with constant pitch angle. Hyperpriors on the noise parameter were chosen by fitting a truncated gamma distribution (Zaninetti 2014) to the spiral width slider values returned by volunteers (ignoring sliders left at the default or moved to the extremes of allowed values). Any logarithmic spirals within a distance of 0.0005 (given by the clustering metric) were deemed to be from the same arm and thus their classifications were merged and a log-spiral recalculated.

We do not assume that every arm in a galaxy has the same pitch angle. To obtain a single value for the pitch angle of a galaxy, we take the length-weighted average pitch angle of all arms detected in the galaxy (as used by Davis & Hayes 2014).

The galaxy model for UGC 4721 obtained through aggregation can be seen in the bottom left panel of Figure 7.

2.6. *Error Estimation of Aggregate models*

As all components in a cluster can be viewed as volunteers' attempts at modelling the true underlying component, the sample variance of the parameters of these shapes can be used as a measure of confidence in the parameters present in the aggregate result. These are highly sensitive to clustering hyperparameters, and are only valid for a component's position, size and shape. Figure 6 illustrates the variance in clustered shapes for our example galaxy (UCG 4721); we see a large variation in the clustered discs, and much closer agreement on bulge and bar size and shape.

2.7. *Model Fitting*

The final step in creating *Galaxy Builder* models is a numerical fit to fine-tune parameters. This fitting was performed using the L-BFGS-b algorithm (Byrd et al. 1995), implemented in SCIPY (Jones et al. 2001). We minimize a custom likelihood function that assumes gaussian error on pixel values and incorporates the priors on parameters we obtain from clustering. The full fitting model and likelihood is detailed in Appendix A. We use the same model as used by volunteers in the online interface (with altered limits), with spiral arms restricted to being logarithmic spirals relative to the disc, and without the ability to change the relative falloff of spiral arms.

The model rendering and fitting code was written up using Google's JAX package (Bradbury et al. 2018), which allows GPU-optimization and automatic gradient calculation, enabling quick and accurate calculation of the jacobian matrix needed for the L-BFGS-b minimization algorithm.

We initially fit only for the brightnesses of components, and then simultaneously for all free parameters of all components. The result of the fit, including the final photometric model for UGC 4721, can be seen in 7. The secondary components have been accounted for well, and the model has a sensible reduced chi-squared value of 1.176, where we have approximated degrees of freedom as the number of unmasked pixels present in the galaxy image (similar to GALFIT).

We use the errors described in Section 2.6 as parameter uncertainties, as we feel an approach based on the local curvature of the likelihood-space (as used by `Galfit`) would likely fall foul of the issues described in the introduction and thus be a under-estimate. This decision means we do not have uncertainties for some parameters.
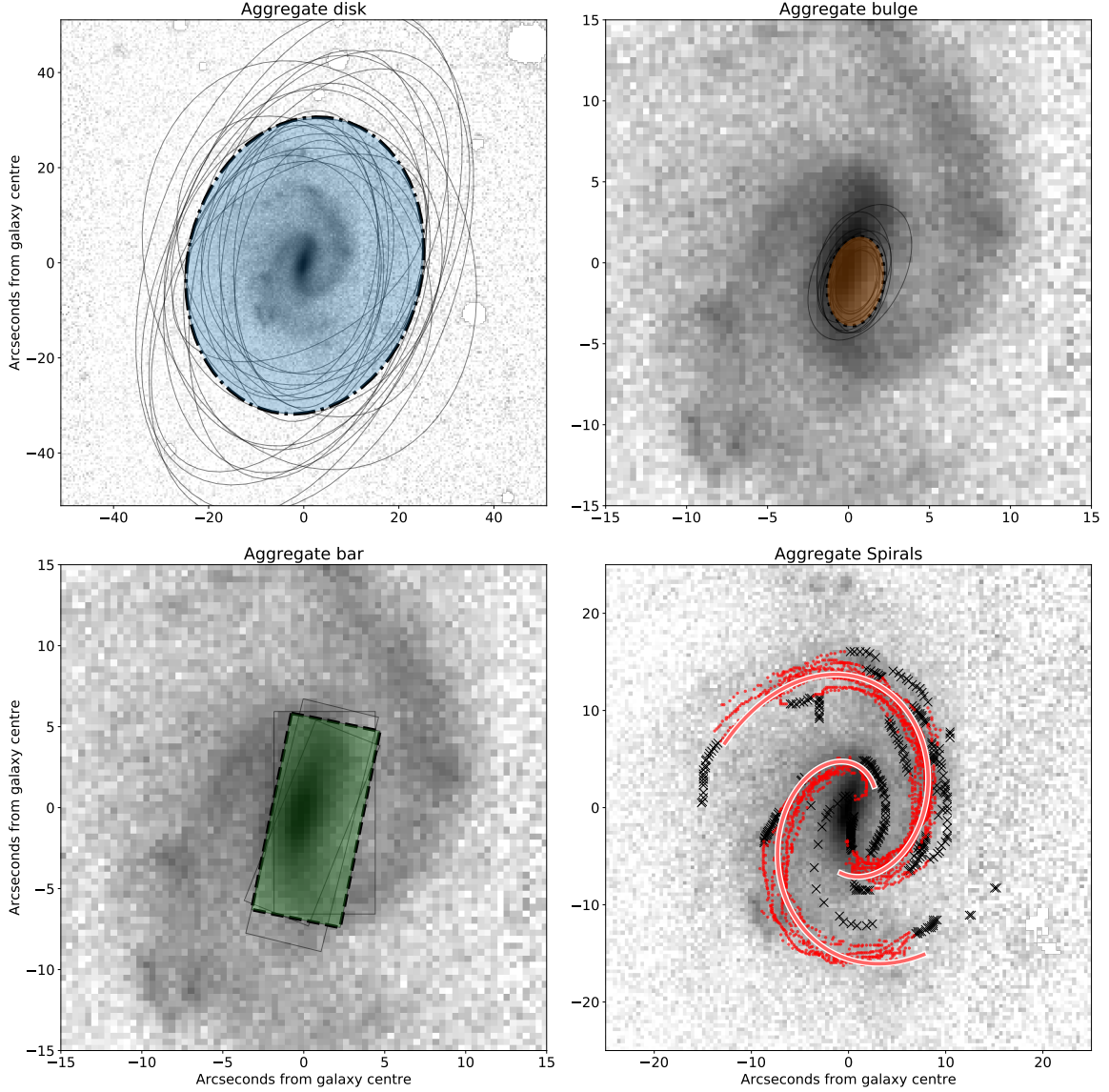
Figure 6. Calculated aggregate components for UGC 4721. The aggregate disc is shown using a dot-dashed line and blue fill in the upper left panel, the aggregate bulge with a dotted line and orange fill in the upper right panel, the aggregate bar using a dashed line and green fill in the lower left panel and the aggregate spiral arms are plotted as red lines in the lower right panel. Sérsic components are displayed at twice their effective radii. Black crosses in the lower right panel indicate spiral arm points that were identified as outliers and removed during cleaning (described in Section 2.5.3). The aggregated components agree well with the underlying morphology, despite the noisiness of the classifications received.

We remove two models for which a fit did not converge.

3. RESULTS

In this Section we present *Galaxy Builder* models for 198 galaxies, both from simple aggregation of user classifications (aggregate models), and with parameters fine tuned by a numerical fit (fitted models). We explore the consistency with which volunteers modelled galaxies, the accuracy of the aggregate models, and compare the aggregate and fitted models to comparable results in the literature.

3.1. *The Calibration Set*

The calibration subset was a set of nine synthetic galaxy images created from *Galaxy Builder* models, which were then re-run through the *Galaxy Builder* process. These galaxies were used to fine-tune clustering and fitting hyperparameters
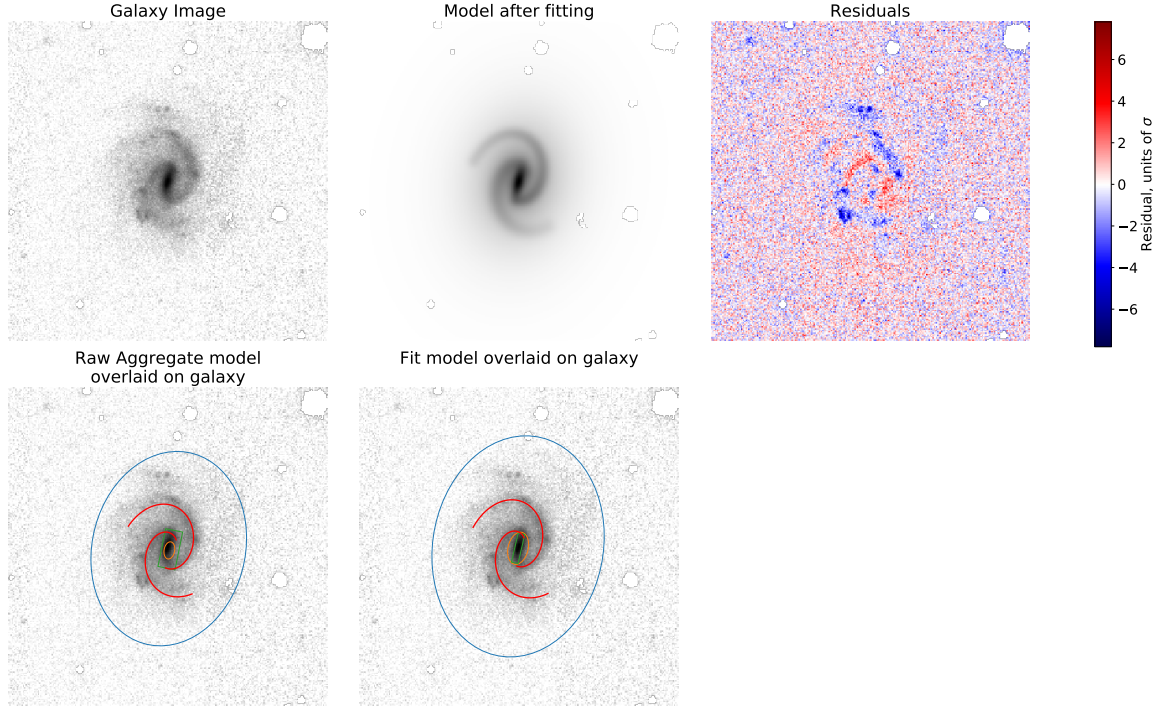
Figure 7. Effect of fitting on the aggregated models. The top left panel shows an Arcsinh-scaled image of the galaxy being fit (UGC 4721), the top middle shows the final model obtained (with the same limits and scaling as the galaxy image) and the top right shows the difference between the two images, in units of pixel uncertainty. The bottom panels show a simple representation of the model before and after tuning, overlaid on the galaxy image from the top-left panel. With minimal change to the aggregated components, we recover a detailed model that matches the galaxy exceptionally well, as evident in the residuals.

(See Section 2.5.1), as the ground truth was known. Our ability to recover morphology accurately is essential validation for our ability to recover good photometric models of galaxies.

Overall, these results highlight the importance of good priors to obtain accurate fits of complex photometric models. In more detail, the models recovered for the nine synthetic galaxy images demonstrate that:

1. Model parameters were generally recovered to a high degree of accuracy

2. We successfully recover all spiral arms present, and do not receive any false positives The spiral pitch angles obtained through aggregation vary by $< 9°$ from the true values, with fitting improving this error slightly.

3. Volunteers systematically use elongated bulges to model bar components. This resulted in two false positives for bulge presence in the aggregate models. This feature (switching light between model components) is a common issue in all photometric fitting methods (Kruk et al. 2018).

4. The Jaccard metric is unstable to small changes in rotation for highly elliptical components (i.e. bars). This resulted in one false negative of bar presence in the aggregate model.

The fitting step for this subset of images highlighted the benefit of obtaining a rough starting point through clustering of user classifications; the method struggled to recover structural parameters for which we did not obtain such a starting point (Sérsic index and bar boxyness). These parameters are difficult to identify using gradient descent (Lackner & Gunn 2012), suggesting future work should attempt to obtain priors on these parameters from volunteers and make use of a more robust fitting algorithm.

3.2. *Examination of Volunteer consistency*

We aggregate two independent models for a set of 98 galaxies based on "original" or repeat ("validation") classifications, obtained with the same retirement limit (see Section 2.2 for more on this selection).
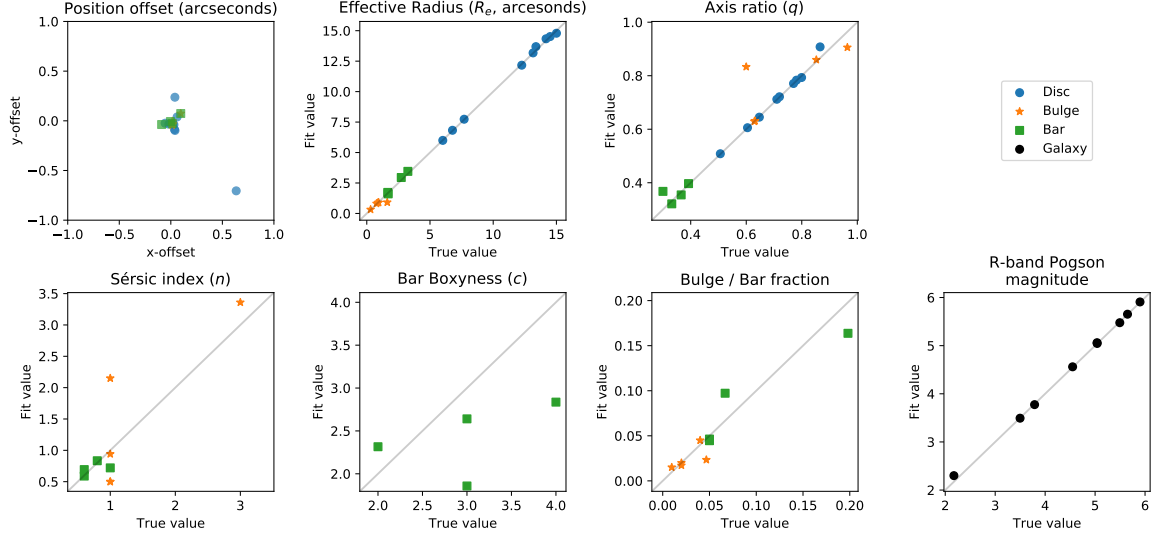
Figure 8. Plots examining the accuracy of fit parameters for the calibration subset of galaxies. Most parameters are recovered to a high degree of accuracy, however Sérsic index and boxyness are difficult to determine only using gradient descent, and the error in the fit values reflects this problem.
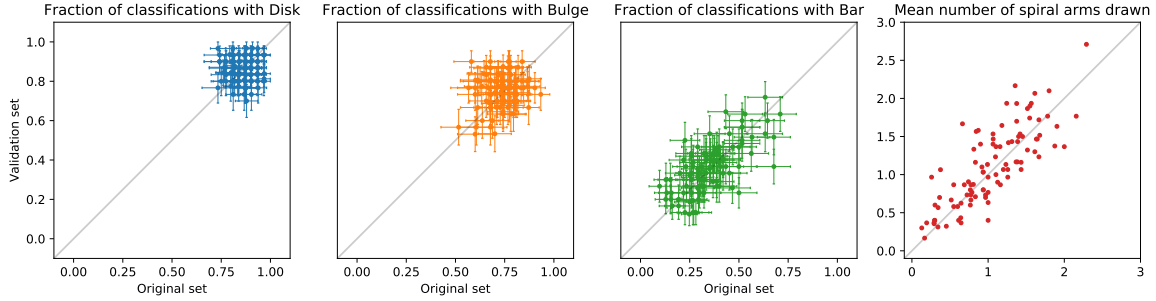

Figure 9. Comparison of frequency of use of component in volunteer models between the original and validation sets of classifications. Errors shown on the disc, bulge and bar arise from Binomial error estimation. We see that classifications are generally consistent within errors, validating our assumption of volunteer independence.

One of the simplest choices the volunteers have is whether to include a model component or not. Figure 9 illustrates the consistency with which volunteers made use of a component in their model for a galaxy. We see that volunteer classification is very consistent, with scatter as predicted by the Binomial uncertainty on the mean. Volunteers almost always make us of a disc and bulge (as seen in the *calibration subset*), and bulge, bar and spiral arm usage is consistent within Binomial error. One common challenge is that some volunteers used a very ellipsoidal bulge and the ends of spiral arms to model light that other users modelled with a bar. This caused some scatter in aggregate models.

In the end, the aggregated validation model is identical to the original aggregated model in around 40% of galaxies. The most common changes are a missing bar component or a missing single spiral arm. This may suggest that more than 30 classifications should be collected per galaxy, or may be an artefact of the lack of consensus among volunteers for galaxies with difficult-to-determine components.

After selecting a component, the volunteer sets its shape and size. The variation in axial ratios and effective radii for the aggregate discs, bulges and bars are shown in Figure 10. The aggregate discs and bulges are consistent within errors, however bars show more scatter. Bars are one of the most challenging components to aggregate consistently. This is partly because even a strongly barred galaxy with 30 classifications overall might receive only 15 or so drawn bars, and lower numbers of classifications result in more scatter. In addition the aggregation method is more sensitive to rotation of highly elongated shapes. Both factors probably contribute to lower consistency in bar components.
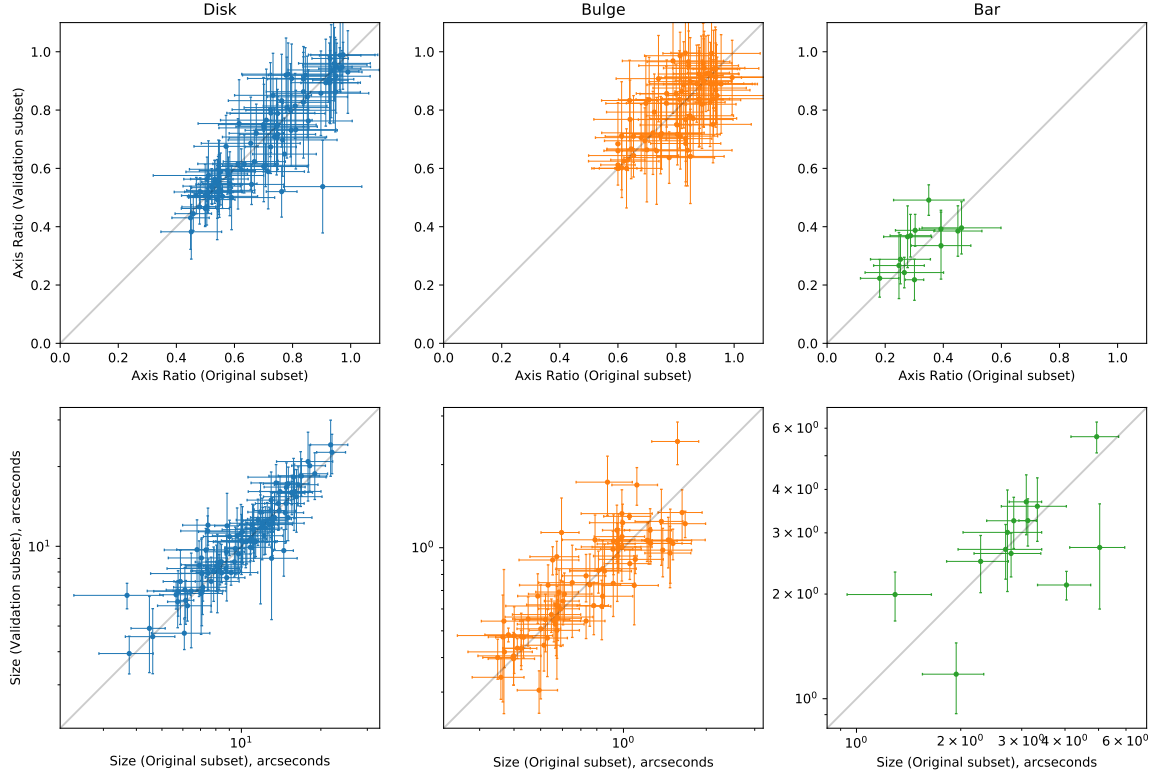
3.3. *Comparison to results in the literature*

Figure 10. Comparison of component shape in aggregate models between the original and validation sets. Errors are obtained through the sample variance of clustered components, as detailed in Section 2.6. We see close agreement between aggregate components from the two sets, suggesting that the clustering method is robust to the scatter in classifications.

After having aggregated and fitted models for our galaxies, we examine how our models compare to other results in the literature. Part of the motivation for exploring the *Galaxy Builder* method was that there exists no published large sample of galaxies with four-component photometric fits. This means we can only make comparisons for individual or subsets of model components (e.g. just disc and bulge) and by design *Galaxy Builder* models will differ as we have attempted to fit bulge-disc-bar-spiral models to all our galaxies. The reader is therefore cautioned against treating literature models as any kind of "ground truth" since deviation from these simple models is part of the goal of this project. We provide these comparisons not to check how well our models work, but to provide data on how they compare with other well known, but much simpler photometric models.

3.3.1. *Comparison to Galaxy Zoo morphology*

The simplest comparison we can make to external results is to examine whether our models respect the existing morphological classifications present in the literature. We make use of Galaxy Zoo 2 (GZ2, Willett et al. 2013) results, including the redshift debiasing described in Hart et al. (2016) and spiral properties calculated in Hart et al. (2016).

When comparing the probability of a volunteer's classification containing a bar component against a galaxy being classed as strongly-barred or as having no bar (as defined in Masters et al. 2010), we see reasonable agreement. Classifications of GZ2 strongly-barred galaxies ($p_{\rm bar} > 0.5$) are more likely to contain a bar than GZ2 unbarred galaxies (0.47 ± 0.15 vs. 0.29 ± 0.11). While there is some overlap in these probabilities, the Pearson correlation between GZ2's $p_{\rm bar}$ and the bar likelihood in *Galaxy Builder* is 0.56, implying a significant correlation. We also note that GZ2 bar classifications exclude most weak bars (Kruk et al. 2017).

We also compare the number of spiral arms aggregated by *Galaxy Builder* with the responses to the GZ2 "number of arms" question (of which the possible responses were one, two, three, four, more than four or "Can't tell"). We attempt to account for the spread in volunteer answers to this question by binning responses, rather than using the
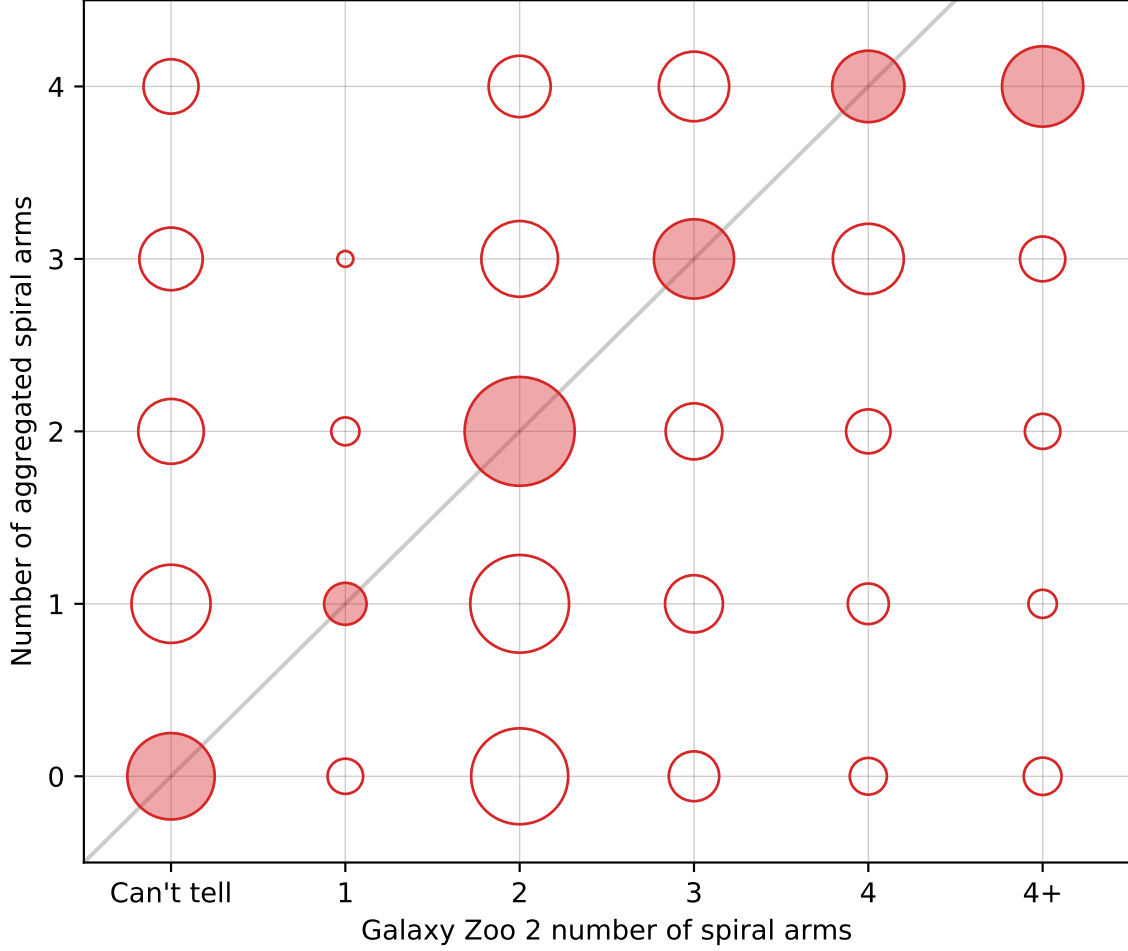
Figure 11. Density plot of GZ2 vote counts for spiral arm number vs the number of spiral arms obtained through aggregation. The area of each circle can be seen as the level of agreement between *Galaxy Builder* aggregate models and GZ2 classifiers, and is defined by Equation 3. The circle with the largest area for each possible GZ2 response is highlighted by shading. The 1:1 relationship suggests the clustering method is correctly recovering the behaviour of volunteers.

mean or modal response. The results of this comparison can be seen in Figure 11. The area of each circle can be seen as the level of agreement between *Galaxy Builder* aggregate models and GZ2 classifiers, it is defined as

$$A_{i,j} \propto \sum_k^{N_g} \frac{1}{M_k} \sum_m^{M_k} \begin{cases} 1, & \text{if } n_k = i \text{ and } C_{k,m} = j \\ 0, & \text{otherwise} \end{cases}, \tag{3}$$

where n_k is the number of aggregate arms for galaxy k (out of N_g galaxies), $C_{k,m}$ is the m-th answer for galaxy k (out of M_k answers).

The circle with the largest area for each possible GZ2 response is highlighted, and agrees with the number of spiral arms aggregated here for $m = 1, 2, 3, 4$. No aggregate model contained more than four spiral arms, and when galaxies have an uncertain number of spiral arms (the "Can't tell" GZ2 response) we mostly do not aggregate any spiral arms.

It is not uncommon in *Galaxy Builder* for one spiral arm to have been broken into two smaller segments. We also occasionally identify two distinct clusters that represent the same physical arm. These two reasons account for a majority of cases where GZ2 classifications suggest a galaxy has two spiral arms and we have clustered a larger number. Improved project user experience would be crucial in correcting these errors.

3.3.2. *Comparison to One-component fit - axis ratio*

We compare the axis ratios of the discs of *Galaxy Builder* aggregate models (without fitting) to the axis ratio of a 2D Sérsic fit to the *r*-band SDSS image of each galaxy (as provided in the NSA catalog, Blanton et al. 2011). For these untuned models there is an error of ∼ 0.1, consistent with our expected errors (derived in Section 2.6).

We observe a clustering of outlying values around $b/a = 0.5$. This is almost certainly due to the drawing tool ellipse having a default axis ratio of 0.5. Where this default is a "good enough" fit we hypothesise that volunteers are less likely to modify it, while if it needs to move a long way they find a more refined value. Overall we see that 36% of all disc components drawn by volunteers were left at the default axis ratio. We recommend that future projects should carefully consider their interface design to minimize this bias (e.g. forcing volunteers to draw both the major and minor axis), however the fitting process we implement on the aggregate models successfully removes the bias, and the overall scatter does not change significantly.

As we account for light in spiral arms and bars, we expect that disc axis ratios fit by *Galaxy Builder* should be more physical than those from models that do not account for how these non axisymmetries can bias measurements of ellipticity.

3.3.3. *Comparison to Disc-Bulge models*

A strong motivation for performing multi-component modelling is the desire to measure the fraction of a galaxy's light being emitted by its central components (such as bulge fraction, defined as the ratio of bulge luminosity to total luminosity). Gao & Ho (2017) demonstrate that modelling secondary central components is essential for recovering an accurate measure of bulge fraction. The difficulty of measuring bulge fraction is further compounded by the complex degeneracies present in even two-component fits, meaning that many gradient-descent based solvers often fail to find the globally optimum solution (Robotham et al. 2016), especially when bulge Sérsic index is a free parameter.

One of the largest catalogs of 2D multi-component fits is Simard et al. (2011), which performed simultaneous, two-bandpass decompositions of 1,123,718 galaxies in the Legacy area of the SDSS DR7 using GIM2D. Three variations of models were fitted: a pure Sérsic model, an Exponential disc and de-Vaucouleurs bulge model (hereafter exp+deV), and an Exponential disc and a Sérsic bulge model (exp+S). Fitting was performed using the Metropolis algorithm, which is resilient to local minima and therefore suitable for the complex likelihood space of galaxy photometric modelling. Lackner & Gunn (2012) similarly fitted two models to SDSS main-sample galaxies: an exponential disc and exponential bulge (exp+exp), and an exponential disc and de Vaucouleurs bulge. They used a Levenberg-Marquadt gradient descent algorithm, with initial parameters taken from previous SDSS analysis.

We compare our central component fraction (the flux of the bulge and bar relative to the total model flux) to bulge fraction from Simard et al. (2011) where their analysis indicated genuine bulge+disc systems ($P_{pS} \leq 0.32$). We compare to Lackner & Gunn (2012) bulge fractions only when their model selection criteria determined that model was the best-fit model. We see a strong correlation with significant scatter (Figure 13). The relationship to exp+deV models appears to be less than 1:1, while the relationship to exp+exp models is greater than 1:1, highlighting the dependence of bulge fraction on Sérsic index. Taking *Galaxy Builder* results as ground truth implies that exp+deV puts too much light into the bulge, while exp+exp puts too little.

The amount of scatter (and lack of consistent 1:1 relationships) between bulge fractions between any two of the published two-component models is comparable to the scatter we see between any one of them and our more complex model. Bulge fractions for complex multi-component galaxies fit with any method should be used with caution.

3.3.4. *Comparison to Disc-Bulge-Bar models*

Kruk et al. (2018) performed multi-component (up to three), multi-band decompositions of a selection of SDSS galaxies, 23 of which were also classified in *Galaxy Builder* (with 16 in the repeated validation subset). Figure 14 compares the axis ratios and effective radii of bulges, discs and bars in Kruk et al. (2018) to those present in the fitted models. We see good consistency in effective radii of all components in the majority of galaxies. There is more scatter in the fit axis ratios of components. In particular we observe many of the *Galaxy Builder* bulges reaching the imposed lower boundary. Comparing the central component fraction between *Galaxy Builder* models and those in Kruk et al. (2018), we see next to no scatter.

3.3.5. *Comparison to Disc-Bulge-Bar-Spiral models*

To the best of our knowledge, no photometric models exist for the *Galaxy Builder* sample which contain spiral arm structure. The closest comparable result is that produced by Gao & Ho (2017), however the galaxies they used are not in the Sloan footprint.
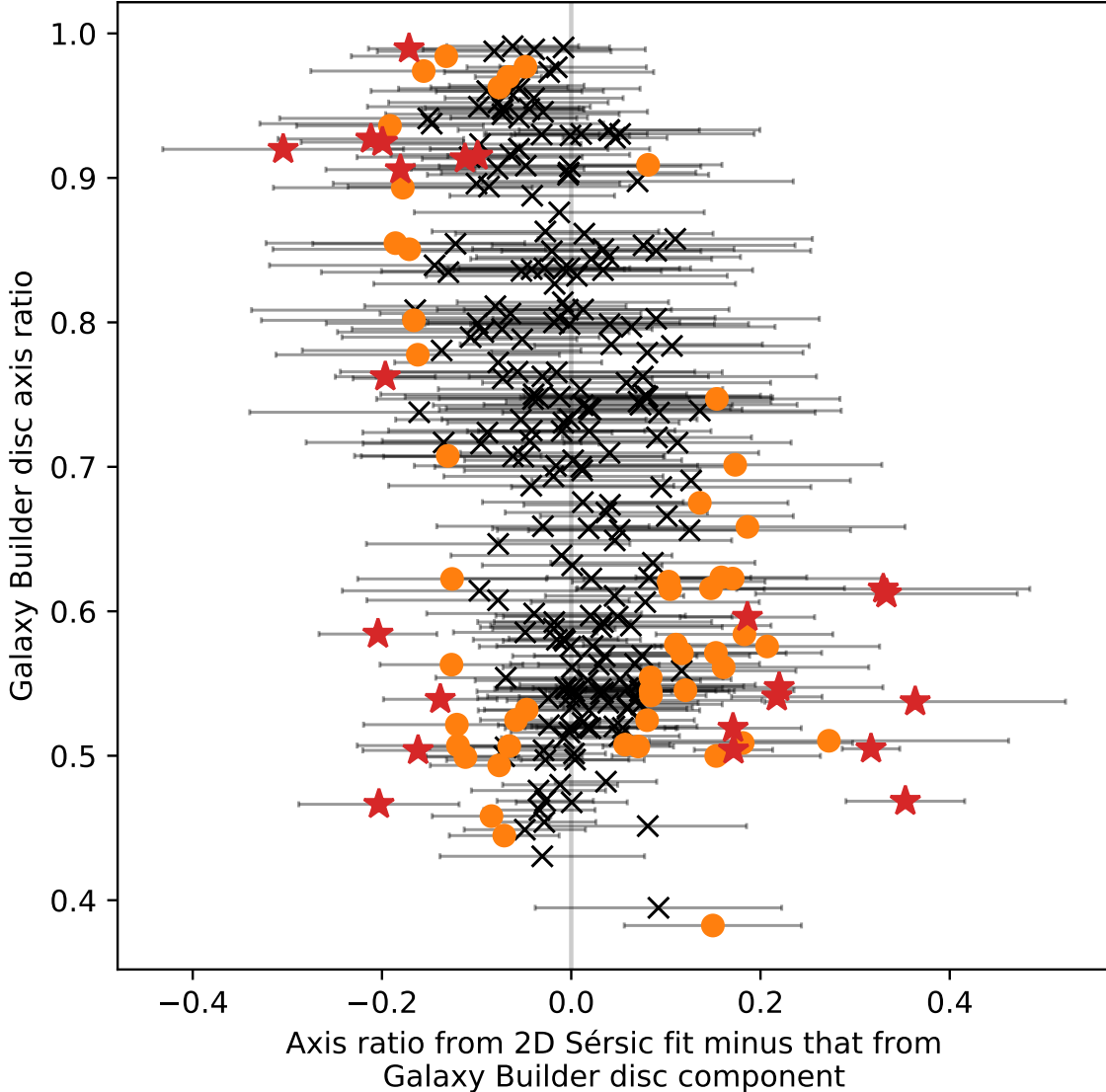
Figure 12. Difference between the axis ratios of the aggregated disc component (before fitting) to the results of an *r*-band Sérsic profile fit. Points between one- and two-sigma are highlighted as orange squares, points outside 2σ are shown as red stars. While the overall relationship is good, the increase prevalence of points outside 2σ is a clear indication of bias caused by the *Galaxy Builder* online user interface.

In order to provide a comparison for our novel method of spiral parameter (pitch angle and amplitude) extraction, we compare the result of our galaxy length-weighted pitch angles to the relationship obtained by Hart et al. (2016) between GZ2 classification and galaxy pitch angle. Their fit was obtained by using the Zooniverse project *Spiral Spotter* to filter good vs bad spiral arm segments identified using an automated spiral arm detection and fitting tool, SPARCFIRE (Davis & Hayes 2014), whereas *Galaxy Builder* asks volunteers to provide their own opinion on spiral arm number, location and tightness. *Galaxy Builder* pitch angles are within the (large) uncertainties on the Hart et al. (2016) fit.

Many researches (Davis & Hayes 2014, Díaz-García et al. 2019 to name a few) have noted that many galaxies show large inter-arm variations in pitch angle, suggesting that obtaining a single value of a galaxy's pitch angle is highly dependent on which arms have been identified. We plan to further explore this issue in future work.

4. SUMMARY AND CONCLUSIONS
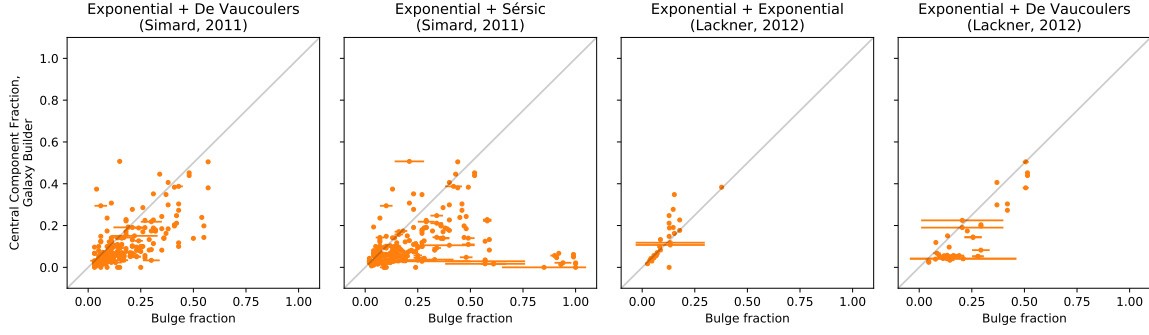
Figure 13. Scatter plots comparing the ratio of flux from central components (bulge and bar) to the total flux between fitted models from *Galaxy Builder* and two-component models in the literature. Our models are broadly consistent with their results, but should be more accurate for complex galaxies, as we account for galaxy bars.
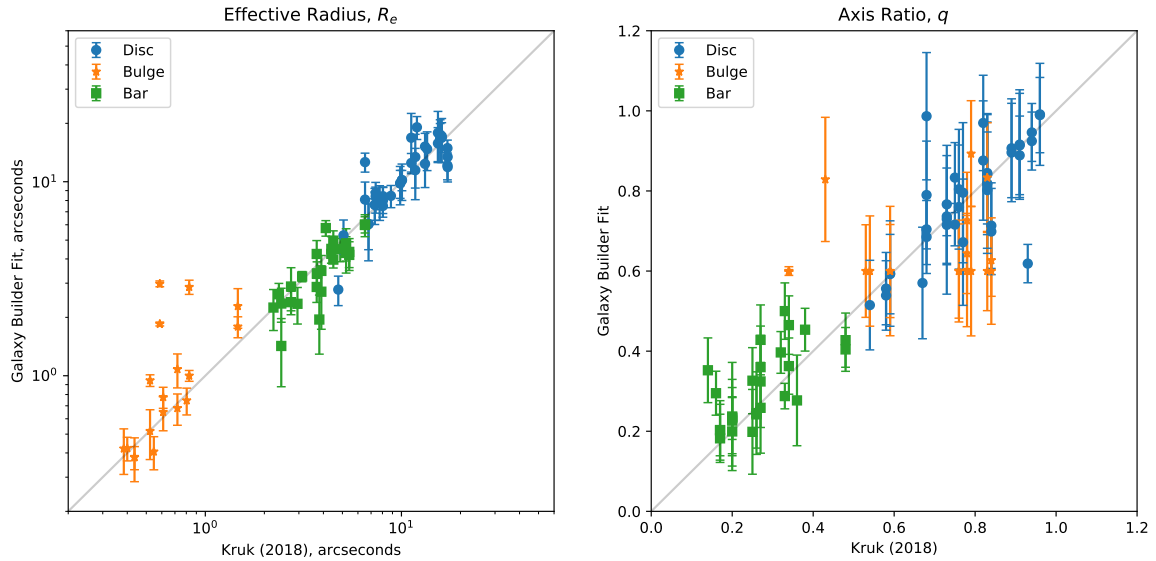

Figure 14. Comparison between *Galaxy Builder* fitted models and the result of 3-component, multiwavelength fits performed by Kruk et al. (2018). Discs, Bulges and Bars are shown as blue circles, orange stars and green squares respectively. The left panel compares component effective radius, the right panel compares the component axis ratio. The components match well, with bulges showing the most scatter. Bulges in *Galaxy Builder* fit models often get stuck at the lower allowed value, despite the physically motivated initial conditions.

In this paper we present a novel method for modelling of galaxy images, *Galaxy Builder*, which was conceived with the goal of solving the "quality or quantity" dilemma facing galaxy image modelling, which, despite advances in computation, still typically requires significant human interaction to achieve quality fits. We release our catalogue of models to the community, and in future work we use this sample to investigate spiral arm formation mechanisms.

Galaxy Builder leverages the power of crowdsourcing for the hardest to automate parts of image fitting, namely determining the appropriate number of model components to include, and finding regions of parameter space close to the global optima.

The use of a small sample of synthetic images to calibrate and test our model clustering and fitting code has demonstrated our ability to recover galaxy morphology in the majority of cases. For example, our spiral arm fitting recovered spiral pitch angles to within 9 deg. This set of 9 synthetic images revealed a systematic tendency for volunteers to incorporate more bulges and fewer bars than necessary for photometric models of strongly barred spirals. Future work might implement an improved clustering algorithm (and improved user interface) to address the failures of bar model clustering we observed in a small fraction of galaxies.

Some of our parameters do not cluster well (bulge and bar Sérsic n, bar boxyness), we hypothesise that this is because a wide range of values fit the light profile well. As a result we are unable to obtain reliable physical results with our optimization algorithm (gradient descent-based methods are subject to being trapped in local minima, or not converging for parameters with flat likelihoods). A solution to this would be performing a full Bayesian optimization with priors obtained from volunteer input, or using a more robust algorithm (such as Basin-Hopping; Wales & Doye 1998). This work is beyond the scope of the current study.

We have demonstrated our ability to obtain physically motivated models with comparable reduced chi-squared values (between 1 and 5) to results in the literature. We obtain errors on parameters where possible through the sample standard deviation of component clusters, which is less likely to be an under-estimate than approximations using the local curvature of the Likelihood-space.

We compare these new models to existing results in the literature. We find good agreement where the models or parameters are comparable, and suggest that where differences are found, *Galaxy Builder* should generally provide superior models because of the more realistic modelling of the galaxy morphologies.

We were able to obtain aggregate models for 296 images with an average rate of one galaxy per day, and fit photometric models for 294 images. User experience and task simplification will need to be considered if significantly larger numbers of these models are to be obtained. At the time of writing and to the best of our knowledge, the number of photometric models obtained here is still significantly larger than the largest sample obtained through purely computational photometric fitting of disc, bulge, bar and spiral arms in galaxies (10 galaxies, Gao & Ho 2017, who also included rings, disc-breaks and further components).

Any citizen science project is only as good as the volunteers who generously donate their time to it. We were incredibly fortunate to be able to make use of the wonderful pool of volunteers built by the Zooniverse, who in some cases contributed hundreds of detailed galaxy classifications to this project. We are optimistic about the potential of projects like *Galaxy Builder* to dramatically increase the ability of researchers to perform complex, labour-intensive modelling of galaxy photometry, leveraging the power of the crowd to perform the complex tasks best suited to humans, and computer algorithms for the final optimization.

5. ACKNOWLEDGEMENTS

APPENDIX

A. MODEL FITTING

Assume Normal priors on component parameters determined from clustering (μ_x, μ_y, q, Re), with the spread given by the spread in the clustered values. We therefore have that our final log-likelihood (to be maximised) is the sum of the gaussian log-likelihood of the residuals given the pixel uncertainty and the gaussian log-likelihood of the variation in parameters, given their uncertainty.

The model being rendered is the PSF-convolved sum of the separate components and outputs an (N_x, N_y) image. The disc, bulge and bar are variations on the boxy Sérsic profile:

$$I_{\mathrm{sersic}}(\vec{P}) = I_e \exp\left\{-b_n\left[\left(\frac{r(\vec{P})}{R_e}\right)^{1/n} - 1\right]\right\} \tag{A1}$$

where

$$r(\vec{P}) = \left\|\begin{pmatrix} \frac{1}{q} & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{pmatrix}(\vec{\mu} - \vec{P})\right\|_c. \tag{A2}$$

The disc is resticted to $n = 1; c = 2$, bulge to $n \in (0.5, 6); c = 2$ and bar to $n \in (0.5, 6); c \in (0.5, 6)$.

The Sérsic components are actually rendered at 5x the image resolution, and downsampled using the mean pixel brightness. This is a widely used method of approximating the true pixel value, which is an integration over the area of sky inside the pixel. I.e. for a pixel of size (δ_x, δ_y)

$$I_{\mathrm{pix}}(\vec{P}) = \frac{1}{\delta_x\delta_y}\int_{-\delta_y/2}^{\delta_y/2}\int_{-\delta_x/2}^{\delta_x/2} \mathrm{d}x\mathrm{d}y\, I_{\mathrm{sersic}}\left(\vec{P} + \begin{pmatrix} \delta_x \\ \delta_y \end{pmatrix}\right). \tag{A3}$$

Spiral arms were restricted to be logarithmic with respect to the inclined, rotated disc. They were rendered in a similar manner to the online interface; using the nearest distance from a pixel to a calculated logarithmic spiral.

An inclined, rotated log spiral requires parameters brightness I_s, spread s, minimum and maximum θ (a and b), an amplitude A, pitch angle ϕ, position $\vec{\mu}$, position angle ψ and axis ratio q, where $\vec{\mu}$, ψ and q are inherited from the disc component.

The distance from a pixel to a logarithmic spiral is given by

$$D_{\mathrm{s}}(\vec{P}) = \min_{\theta \in [a,b]}\left\|\vec{P} - \vec{\mu} - Ae^{\theta\tan\phi}\begin{pmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{pmatrix}\begin{pmatrix} q\cos\theta \\ \sin\theta \end{pmatrix}\right\|^2. \tag{A4}$$

In practise the spiral distance was approximated using the distance to a poly-line with 200 vertices, as solving the above minimization for each pixel at each fitting step is computationally intractable. We also adjust A, a and b to account for the rotation of the disc component from its starting value, in order to prevent spirals inadvertently moving far from starting locations for face-on discs (which have poorly constrained position angles). These adjustments are

$$\begin{aligned} A' &= Ae^{\Delta\psi\tan\phi}, \\ a' &= a - \Delta\psi, \\ b' &= b - \Delta\psi. \end{aligned} \tag{A5}$$

The pixel brightness is then calculated as

$$I_{\mathrm{spiral}}(\vec{P}) = I_{\mathrm{disc}}(\vec{P}) \times I_s \exp\left(\frac{-D_{\mathrm{s}}(\vec{P})}{2s^2}\right). \tag{A6}$$

For the fit, we parametrize disc I_e as the Sérsic total luminosity, given by

$$L_{\mathrm{tot}} = I_e R_e^2\, 2\pi n \frac{e^{b_n}}{(b_n)^{2n}}\Gamma(2n). \tag{A7}$$

Bulge (bar) I_e is reparametrized as "bulge (bar) fraction", which we define as

$$F_{\mathrm{bulge}} = \frac{L_{\mathrm{bulge}}}{L_{\mathrm{disc}} + L_{\mathrm{bulge}}}, \tag{A8}$$

and is limited to be between 0 and 1. Disc luminosity is allowed to take any value greater than or equal to zero.

Similarly, bulge and bar effective radius are reparametrized as their scale relative to the disc ($R_e = R_e/R_{e,\,\mathrm{disc}}$). Bulge and bar are also restricted to have the same position.

Table 2. The maximum, minimum and default values for model parameters. Note that some parameters were allowed to overflow when fitting, for instance an axis ratio greater 1 (signifying a swap of major and minor axis) was allowed, and corrected for once fitting reached completion. This helped avoid the optimizer encountering parameter bounds and failing to converge. Component roll and spiral pitch angle was similarly unconstrained.

Component	Parameter	Tuning Minimum Bound	Tuning Maximum Bound
disc	μ_x	-inf	inf
	μ_y	-inf	inf
	roll	-inf	inf
	q	0.25	1.2
	R_e	0	inf
	Σ_e	0	inf
bulge	μ_x	-inf	inf
	μ_y	-inf	inf
	roll	-inf	inf
	q	0.6	1.2
	R_e/R_{disc}	0.01	1
	$(B/T)_r)$	0	0.99
	n	0.5	5
bar	μ_x	-inf	inf
	μ_y	-inf	inf
	roll	-inf	inf
	q	0.05	0.5
	R_e/R_{disc}	0.05	1
	$(B/T)_r)$	0	0.99
	n	0.3	5
	c	1	6
spiral	Σ_e	0	inf
	spread	0	inf
	ϕ	-85	85

B. ANCILLARY TABLES

REFERENCES

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543, doi: 10.1088/0067-0049/182/2/543

Albareti, F. D., Allende Prieto, C., Almeida, A., et al. 2017, ApJS, 233, 25, doi: 10.3847/1538-4365/aa8992

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123, doi: 10.3847/1538-3881/aabc4f

Bamford, S. P., Häußler, B., Rojas, A., & Borch, A. 2011, in Astronomical Society of the Pacific Conference Series, Vol. 442, Astronomical Data Analysis Software and Systems XX, ed. I. N. Evans, A. Accomazzi, D. J. Mink, & A. H. Rots, 479

Barden, M., Rix, H.-W., Somerville, R. S., et al. 2005, ApJ, 635, 959, doi: 10.1086/497679

Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393, doi: 10.1051/aas:1996164

Blanton, M. R., Kazin, E., Muna, D., Weaver, B. A., & Price-Whelan, A. 2011, AJ, 142, 31, doi: 10.1088/0004-6256/142/1/31

Boonchoo, T., Ao, X., & He, Q. 2018, CoRR, abs/1801.06965. https://arxiv.org/abs/1801.06965

Bradbury, J., Frostig, R., Hawkins, P., et al. 2018, JAX: composable transformations of Python+NumPy programs, 0.1.55. http://github.com/google/jax

Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. 2000, in Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, SIGMOD '00 (New York, NY, USA: ACM), 93–104, doi: 10.1145/342009.335388

Byrd, R., Lu, P., Nocedal, J., & Zhu, C. 1995, SIAM Journal on Scientific Computing, 16, 1190, doi: 10.1137/0916069

Davis, D. R., & Hayes, W. B. 2014, ApJ, 790, 87

de Jong, R. S. 1996, A&AS, 118, 557. https://arxiv.org/abs/astro-ph/9601002

de Vaucouleurs, G., de Vaucouleurs, A., Corwin, Jr., H. G., et al. 1991, Third Reference Catalogue of Bright Galaxies. Volume I: Explanations and references. Volume II: Data for galaxies between 0^h and 12^h. Volume III: Data for galaxies between 12^h and 24^h.

Díaz-García, S., Salo, H., Knapen, J. H., & Herrera-Endoqui, M. 2019, arXiv e-prints, arXiv:1908.04246. https://arxiv.org/abs/1908.04246

Dobbs, C., & Baba, J. 2014, Publ. Astron. Soc. Aust., 31

Elmegreen, B. G., & Elmegreen, D. M. 1985, ApJ, 288, 438, doi: 10.1086/162810

Gadotti, D. A. 2011, MNRAS, 415, 3308, doi: 10.1111/j.1365-2966.2011.18945.x

Gao, H., & Ho, L. C. 2017, ApJ, 845, 114, doi: 10.3847/1538-4357/aa7da4

Gao, H., Ho, L. C., Barth, A. J., & Li, Z.-Y. 2018, ApJ, 862, 100, doi: 10.3847/1538-4357/aacdac

Hart, R. E., Bamford, S. P., Willett, K. W., et al. 2016

Hart, R. E., Bamford, S. P., Hayes, W. B., et al. 2017, MNRAS, 472, 2263, doi: 10.1093/mnras/stx2137

Head, J. T. C. G., Lucey, J. R., & Hudson, M. J. 2015, MNRAS, 453, 3729, doi: 10.1093/mnras/stv1662

Holincheck, A. J., Wallin, J. F., Borne, K., et al. 2016

Hopkins, P. F., Croton, D., Bundy, K., et al. 2010, The Astrophysical Journal, 724, 915

Jacob, J. C., Katz, D. S., Berriman, G. B., et al. 2010, arXiv e-prints, arXiv:1005.4454. https://arxiv.org/abs/1005.4454

Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open source scientific tools for Python. http://www.scipy.org/

Kelvin, L. S., Driver, S. P., Robotham, A. S. G., et al. 2012, MNRAS, 421, 1007, doi: 10.1111/j.1365-2966.2012.20355.x

Kormendy, J., Drory, N., Bender, R., & Cornell, M. E. 2010, The Astrophysical Journal, 723, 54

Kruk, S. J., Lintott, C. J., Simmons, B. D., et al. 2017, MNRAS, 469, 3363, doi: 10.1093/mnras/stx1026

Kruk, S. J., Lintott, C. J., Bamford, S. P., et al. 2018, MNRAS, 473, 4731, doi: 10.1093/mnras/stx2605

Lackner, C. N., & Gunn, J. E. 2012, MNRAS, 421, 2277, doi: 10.1111/j.1365-2966.2012.20450.x

Lange, R., Moffett, A. J., Driver, S. P., et al. 2016, MNRAS, 462, 1470, doi: 10.1093/mnras/stw1495

Lilly, S., Schade, D., Ellis, R., et al. 1998, ApJ, 500, 75, doi: 10.1086/305713

Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, Monthly Notices of the Royal Astronomical Society, 389, 1179

Lupton, R., Blanton, M. R., Fekete, G., et al. 2004, PUBL ASTRON SOC PAC, 116, 133

Masters, K. L., & the Galaxy Zoo Team. 2019, arXiv e-prints, arXiv:1910.08177. https://arxiv.org/abs/1910.08177

Masters, K. L., Nichol, R. C., Hoyle, B., et al. 2010, Monthly Notices of the Royal Astronomical Society, 411, 2026

Mendel, J. T., Simard, L., Palmer, M., Ellison, S. L., & Patton, D. R. 2014, ApJS, 210, 3, doi: 10.1088/0067-0049/210/1/3

Mendez-Abreu, J., Ruiz-Lara, T., Sanchez-Menguiano, L., et al. 2016

Parry, O. H., Eke, V. R., & Frenk, C. S. 2009, MNRAS, 396, 1972, doi: 10.1111/j.1365-2966.2009.14921.x

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825

Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266, doi: 10.1086/340952

—. 2010, AJ, 139, 2097, doi: 10.1088/0004-6256/139/6/2097

Pour-Imani, H., Kennefick, D., Kennefick, J., et al. 2016, The Astrophysical Journal, 827, L2

Pozzetti, L., Bolzonella, M., Zucca, E., et al. 2009, Astronomy & Astrophysics, 523, A13

Rampazzo, R., Uslenghi, M., Giorgiev, I. Y., et al. 2019, arXiv e-prints, arXiv:1910.02664. https://arxiv.org/abs/1910.02664

Robotham, A. S. G., Taranu, D. S., Tobar, R., Moffett, A., & Driver, S. P. 2016

Rodrigues, M., Puech, M., Flores, H., Hammer, F., & Pirzkal, N. 2018, MNRAS, 475, 5133, doi: 10.1093/mnras/sty098

Simard, L., Mendel, J. T., Patton, D. R., Ellison, S. L., & McConnachie, A. W. 2011, ApJS, 196, 11, doi: 10.1088/0067-0049/196/1/11

Simard, L., Willmer, C. N. A., Vogt, N. P., et al. 2002a, ASTROPHYS J SUPPL S, 142, 1

—. 2002b, The Astrophysical Journal Supplement Series, 142, 1

Simmons, B. D., Lintott, C., Willett, K. W., et al. 2017, MNRAS, 464, 4420, doi: 10.1093/mnras/stw2587

Simpson, R., Page, K. R., & De Roure, D. 2014, in Proceedings of the 23rd International Conference on World Wide Web, WWW '14 Companion (New York, NY, USA: ACM), 1049–1054, doi: 10.1145/2567948.2579215

Springel, V., & Hernquist, L. 2005, ApJL, 622, L9, doi: 10.1086/429486

Stoughton, C., Lupton, R. H., Bernardi, M., et al. 2002, AJ, 123, 485, doi: 10.1086/324741

van der Wel, A., Bell, E. F., Häussler, B., et al. 2012, ApJS, 203, 24, doi: 10.1088/0067-0049/203/2/24

Vika, M., Bamford, S. P., Häußler, B., & Rojas, A. L. 2014, Monthly Notices of the Royal Astronomical Society, 444, 3603

Wales, D., & Doye, J. 1998, arXiv e-prints, cond. https://arxiv.org/abs/cond-mat/9803344

Willett, K. W., Lintott, C. J., Bamford, S. P., et al. 2013

Willett, K. W., Galloway, M. A., Bamford, S. P., et al. 2017, MNRAS, 464, 4176, doi: 10.1093/mnras/stw2568

Zaninetti, L. 2014, arXiv e-prints, arXiv:1401.0287. https://arxiv.org/abs/1401.0287