MNRAS **000**, 1–14 (2019) Preprint 3rd October 2019 Compiled using MNRAS LATEX style file v3.0

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

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

[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] *Centre for Astronomy & Particle Theory, School of Physics & Astronomy, University of Nottingham, Nottingham, NG7 2RD, UK*
[5] *European Space Agency, ESTEC, Keplerlaan 1, NL-2201 AZ, Noordwijk, The Netherlands*
[6] *Physics Department, Lancaster University, Lancaster, LA1 4YB, UK*
[7]

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Multi-component modelling of complex galaxies is a valuable tool in the quantitative understanding of galaxy evolution, however it is plagued by issues with convergence, model selection and parameter degeneracies, which either limits it to simple models for large samples, or complex models in very small samples (a dilemma we summarize as a choice between "quality or quantity"). This paper presents a novel framework, built inside the Zooniverse citizen science platform, which enables volunteers to help crowdsource the creation of multiple component photometric models of galaxies from FITS images. We test if this method can help solve the quandry over choosing "quality" or "quantity" for complex galaxy image modelling.

We have run the method (including a final algorithmic optimization from the best crowd-sourced solution) on a sample of 296 images from the Sloan Digital Sky Survey (SDSS). The results from these data demonstrate that using citizen science to make selections on number of model parameters to include, as well as in some parameters identifying best fits, is a promising technique for modeling the images of complex galaxies. We examine the robustness of this new method to variation in number and population of citizen scientists, as well as compare it to automated fitting pipelines. We demonstrate that it is possible to consistently recover accurate models which show good agreement with, or improve on previous models in the literature. These results will be used in future work to investigate spiral arm formation mechanisms and we release our catalogue of models to the community.

Key words: galaxies: evolution – galaxies: spiral – galaxies: photometry

1 INTRODUCTION

Disc galaxies are complex objects, containing many different components, including a disc and disc phenomena (i.e. spiral arms, bars and rings), as well as central more spheroidal structures (bulges, nuclear bulges). 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 (e.g. Simard et al. 2002a, Simard et al. 2011, Lackner & Gunn 2012, Kruk et al. 2017, Bamford et al. 2011, Gadotti 2011, Mendez-Abreu et al. 2016, Park et al. 2007, Salo et al. 2015).

These fully quantitative methods allow researchers to obtain structural parameters of galaxy sub-components, which are useful 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 (Hopkins et al. 2010); the strength of a galaxy's classical bulge is thought to be tied

⋆ E-mail: tim.lingard@port.ac.uk

to the strength of a merger event in its past (Kormendy et al. 2010); different spiral arm formation theories slightly 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. Perhaps most notably Simard et al. (2011) performed two-dimensional, Point-Spread-Function (PSF) convolved, two-component (bulge + disc) decomposition of 1,123,718 galaxies from the Legacy area of the SDSS DR7. Other large catalogues of photometric fits exist: Gadotti (2011) made use of parametric multi-band light distribution modelling to model stellar bars in 300 galaxies, Mendez-Abreu et al. (2016) made use of a human-supervised approach to perform multi-component decomposition of 404 galaxies from the CALIFA survey (Sanchez et al. 2011).

However, despite the usefulness of this technique and the presence of analytic profiles and methods for modelling more complex galaxy sub-components, relatively few studies have attempted to perform large-scale parametric decomposition of galaxies using more complicated models than that of Simard et al. (2011). 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 (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 simply when not properly guided or constrained, for instance a Sérsic bulge swapping places with an exponential disc component. It is also the case that often, without near-optimal starting points, detailed model fits will fail to converge at all.

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 Sloan Digital Sky Survey (hereafter SDSS, Blanton et al. 2017, Abazajian et al. 2009, which in total imaged over 50 million galaxies), 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, Willett et al. 2017, Simmons et al. 2017, Hart et al. 2016), 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 [[**add citation**]] Masters 2019 for a recent review).

In this paper we explore an analogous solution to that Galaxy Zoo brought to galaxy classification (Lintott et al. 2008) for the issues faced by galaxy parametric modelling inside the ecosystem that Galaxy Zoo set in motion (namely *The Zooniverse*[1]; leveraging citizen scientists to pick model components and perform model optimization in an online, web-browser environment[2]). We describe our method in Section 2, including details of the images and ancilliary 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.

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 (including the bulge, disc, bar and spiral arm components). As a project of this kind, allowing volunteers to interact with and model data, had never been attempted inside the current Zooniverse web platform before, this project involved designing and implementing a model rendering 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, in an attempt to identify the possible initial conditions resulting in the observed morphology. For part of the project, volunteers downloaded a Java applet, which would run restricted three-body simulations and generate output images. Volunteers could mainpulate the models used in the simulation. The volunteer then voted on simulations which matched a given galaxy merger image, or shared important tidal features. A new batch of simulations would then be run.

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

2.1.1 Project Timeline and Development

The *Galaxy Builder* project was built inside the Zooniverse's (Simpson et al. 2014) Panoptes-Front-End[3] codebase, using Facebook's React.js[4] framework, as well as WebGL[5] to enable realtime photometric galaxy model rendering. *Galaxy Builder* entered a Zooniverse beta in late November 2017 and after some user experience improvements and significant code reworking to meet internal standards, 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 the interface and instructions simple enough for volunteers to understand intuitively, 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 the typical beta trial process for Zooniverse projects.

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 2.3). The interface presents a volunteer with three images, which they can switch between at any time: a r-band cutout 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). A screenshot of the interface can be seen in Figure 1, where the residual image is shown.

The workflow is designed so that volunteers slowly subtract increasing amounts of light from the galaxy, as can be seen in Figure 2. A tutorial is present which contains a step-by-step guide to completing a classification. At each step volunteers are asked to first draw a simple isophote (ellipse for disc and bulge, rectangle for bar and poly-lines for spiral arms), and then make use of a series of sliders to adjust the parameters of the model component (i.e. brightness, Sérsic index and boxyness). A list of parameters, their default values and extrema can be found in Table B1.

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

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

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 voluteers (which was Arcsinh-scaled in a manner described by Lupton et al. 2004), however it is quite sensitive to small deviations of the model from the galaxy.

2.1.3 Rendering the model

The rendering[6] code used in the *Galaxy Builder* project was designed to run on a computer's GPU, using the WebGL rendering API (via the high-level Javascript interface regl[7]). This enables the model (and residual) to be calculated at a speed which allows low-latency feedback to a change in the model made by the volunteer in the browser. Using WebGL has the effect of limiting precision of the model rendered in the browser to that allowed by WebGL, which is 0.00097656 relative precision. however the use of arcsinh scaling resulted in this precision loss being slighly more impactful than if we had rendered without scaling. As these models are envisioned as a starting point for a later numerical fit, it was determined this precision was sufficient.

2.2 Sample Selection: Images and Ancillary Data

The original sample proposed for the *Galaxy Builder* project aimed to mirror the *stellar mass-complete sample* in Hart et al. (2017). This was a sample of face-on spiral galaxies, with and without bars, complete in stellar mass.

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 responed 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.

This selection criteria results in the *stellar mass-complete sample* of 6222 spiral galaxies from Hart et al. (2017).

We split the *stellar mass-complete sample* into smaller sub-samples, each containing 100 galaxies. In an iterative process we chose each sub-sample to contain 60 of the lowest redshift galaxies and 40 random galaxies of those remaining in the sample. This was done to account for the unknown rate at which volunteers would provide classifications (we wanted some results to test as early as possible). 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

[3] `http://github.com/zooniverse/Panoptes-Front-End`
[4] `https://reactjs.org/`
[5] `https://www.khronos.org/webgl/`

[6] 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.
[7] `http://regl.party/`

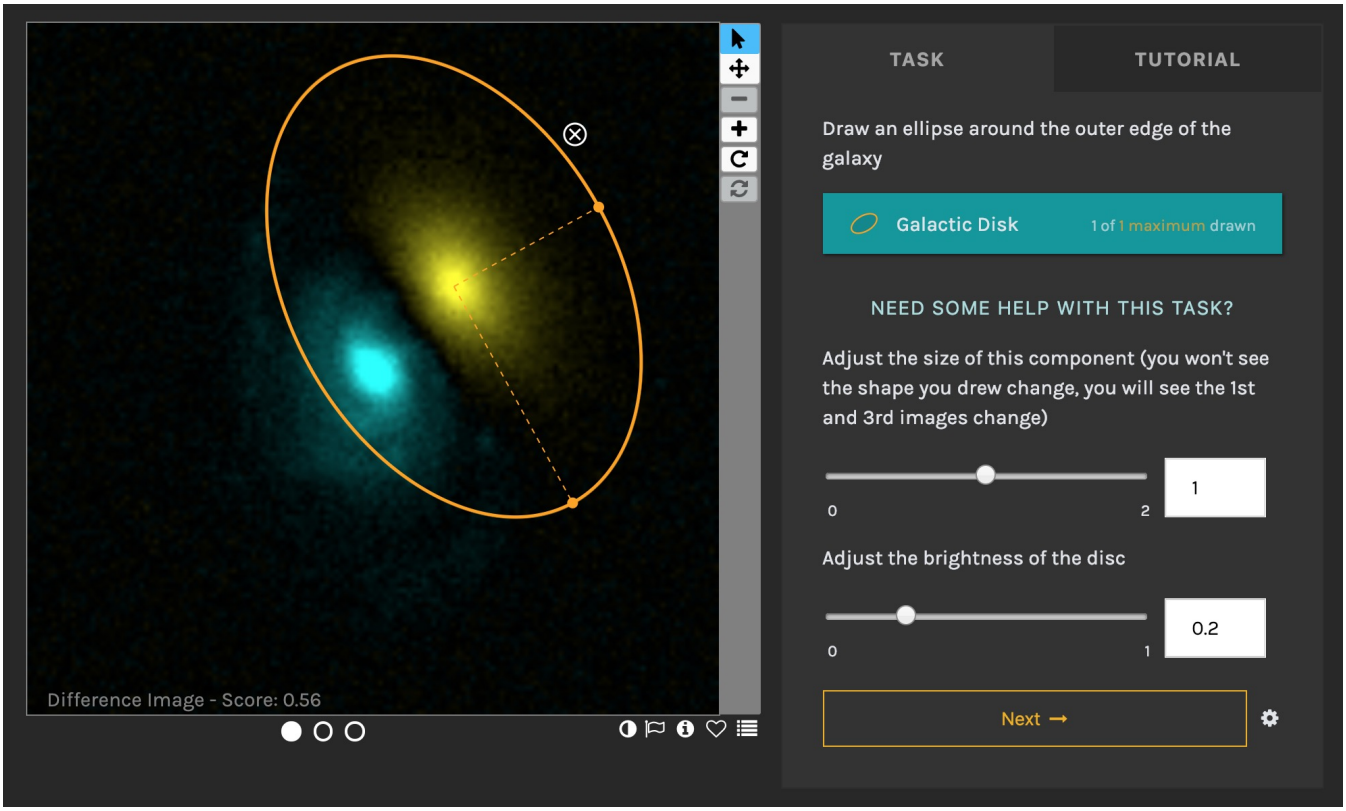

Figure 1. The *Galaxy Builder* interface. The residual image is being shown, and the volunteer is on the "Disc" task. The drawn disc component (yellow) is offset from the galaxy image (blue) to demonstrate the positive and negative residuals. Where the image equals the model the residual is black. The 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.

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 $p_{\text{features}} \cdot p_{\text{not edge on}} \cdot p_{\text{spiral}} \geq 0.5$
Low-redshift cutoff.	$0.02 < z < 0.055$
Face-on galaxy selection using g-band axial ratio.	$(b/a)_g > 0.4$
Volume correction.	$9.45 < \log(M_*/M_\odot) \leq 11.05$
Computation of stellar mass completeness limits using the method of Pozzetti et al. (2009) and limits calculated by Hart et al. (2017).	$2.07\log(z) + 12.64 < \log(M_*/M_\odot) < 2.45\log(z) + 14.05$

cause of this error is unknown, but it leaves a sample of 198 galaxies with images we consider in this paper.

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:

- A montage of multiple r-band observations from the SDSS DR13 (Albareti et al. 2017) data release was created. To combine multiple FITS images, we made use of Astropy (The Astropy Collaboration et al. 2018), and the MONTAGE (Jacob et al. 2010) software package.
- This montage was cropped to four times the Petrosian radius of the galaxy.
- The SEXTRACTOR software (Bertin & Arnouts 1996) was used to identify regions containing secondary sources (foreground stats, other galaxies) and generate a mask.
- 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

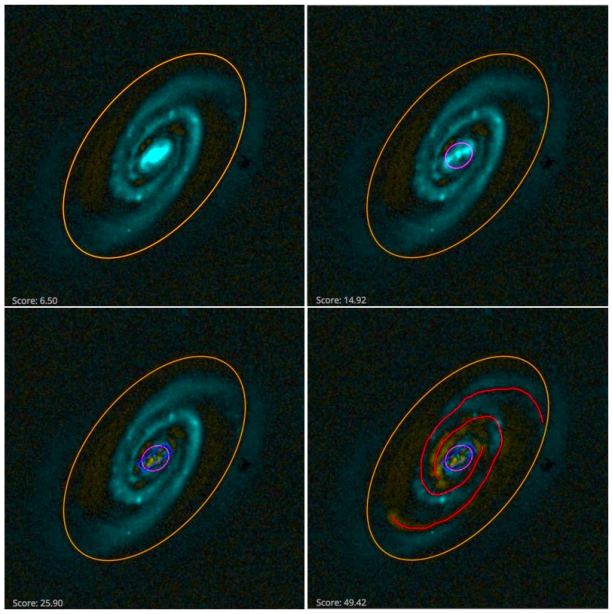

Figure 2. Figure demonstrating the desired result of each step of the modelling process, as seen from the residual image provided to volunteers. The top left panel shows the galaxy after only a disc component has been added: the top right contains a disc and a bulge; the bottom left has a disc, bulge and bar; the bottom right is the finished model with a disc, bulge, bar and spiral arms. The images shown is SDSS J104238.12+235706.8. This brightness and contrast of this image have been edited to improve visibility in print.

for the rendering process (PSF, the size of the PSF array, and the width and height of the image).

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

• 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.

We chose to use r-band images in our subject set due to its higher signal-to-noise than shorter-wavelength bands, but clearer internal structure than the longer-wavelength i-band.

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.

2.3 The Galaxy Model

The model we chose was largely based of components and methodology described in Peng et al. (2002). The allowed model components consisted of:

• An exponential, elliptical disc;

• an elliptical Sérsic bulge (with n chosen by volunteers and allowed to vary from 0.5 to 5);

• a Sérsic bar with a "boxiness" modifier (as described in Peng et al. 2002);

• any number of freehand poly-line spiral arms.

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

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.

The modelling code ignores masked regions identified as secondary sources by SExtractor. It over-samples the bulge, disc and bar components to 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).

2.4 Choice of Retirement limit

We initially collected 10 classifications per galaxy, however preliminary analysis indicated this number of independent answers would be insufficient to create reliable and reproducable aggregate classifications. For this reason, a hand-picked sample of 56 galaxies was then re-activated with a retirement limit of 30 classifications per galaxy. This sample was chosen by eye to be a relatively diverse set of galaxies, most with prominent spiral features including grand-design and flocculent arms. Its purpose was to allow the development of the aggregation methodology.

Once this hand-picked sample was completed, and we had determined that 30 classifications per galaxy was sufficient (having trialled 10 and concluding more was needed for a reliable aggregation), the remaining galaxies from the initial two sub-samples were re-activated, as well as a repeat of the first sub-sample (hereafter the *validation subset*) to measure volunteer consistency. This paper focuses on these 198 galaxies (present in *Galaxy Builder* as 296 separate images) in order to explain the method used, and test the reliability of the models obtained. The *Galaxy Builder* project is still active on the Zooniverse website as of the time of writing and continues to collect classifications for further samples of galaxies.

2.5 Classification Aggregation Methodology

In this Section, we will use galaxy UGC 4721 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. These annotations can be seen in Figure 3, overlaid on the greyscale r-band image of the galaxy.

2.5.1 Best Individual Classification

As *Galaxy Builder* is primarily asking volunteers to solve a complicated regression problem, it is possible to identify the classification provided for each galaxy with the best residual, and assume that this classification has roughly found the globally optimal model. We make use of the common mean squared error metric, in units of nanomaggies, to score the volunteer models (note that this is not the score shown to volunteers).

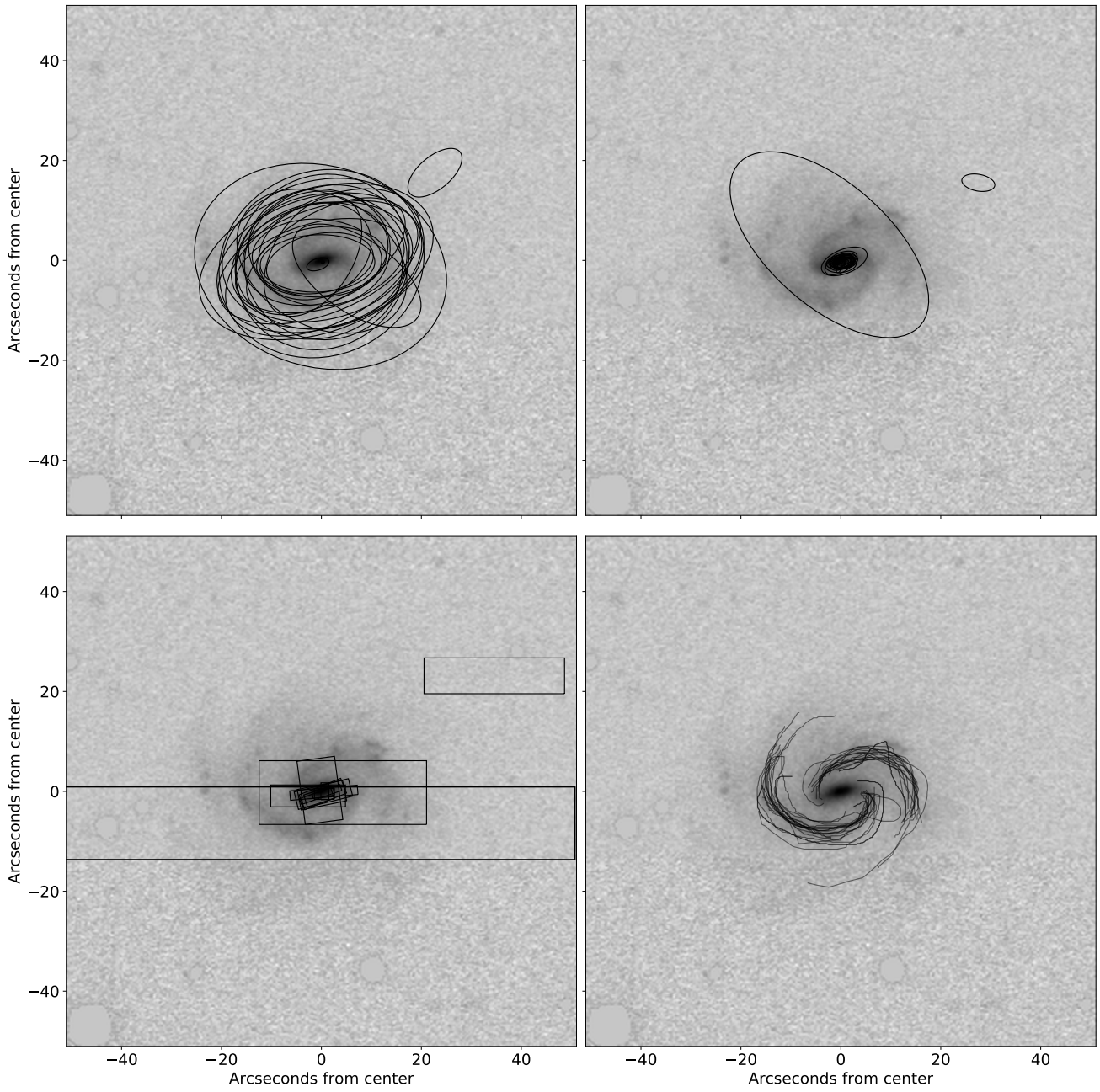

Figure 3. Components drawn by volunteers for UGC 4721. The top left panel shows drawn discs, top right shows drawn bulges, bottom left shows drawn bars and bottom right shows drawn spiral arms.

Once the "best" classification for a galaxy has been identified, we computationally optimize the model using a linear fitting algorithm, fixing the position of each component but allowing all other parameters to vary. This optimized model for each galaxy we refer to as the "best individual classification".

2.5.2 Aggregation of Volunteer Models

As we have multiple independent answers for each parameter and component, we can also combine these to find an aggregated answer. 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. 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 shapes:

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

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 ro-

Table 2. The clustering parameters used for the disc, bulge and bar components present in *Galaxy Builder*

component	`eps`	`min_points`
Disc	0.3	5
Bulge	0.3	3
Bar	0.385	3

bustness and speed. We made use of Scikit-learn (Pedregosa 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. `eps` and `min_points` were chosen by visually inspecting the resulting clustering results and can be seen in Table 2. Bar with axis ratios greater than 0.6 were removed from the pool.

For shapes clustered in this way, we define the aggregate component to be the shape which minimises the sum of Jaccard distances to each of the shapes in the cluster. For our example galaxy, UGC 4721, clustered and aggregate components can be seen in Figure 4.

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. A mathematical description of this measure can be found in Appendix A. 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).

Once spiral classifications on a galaxy have been clustered into the physical arms they represent, the points are deprojected using the axial ratio from a 2D, single-component Sérsic fit in r-band, provided in the NSA catalogue (Blanton et al. 2011). The deprojection method assumes a thin disc and stretches the elliptical minor axis to match the major axis.

Deprojected points within each drawn poly-line are converted to polar coordinates and unwound using `numpy.unwrap` to allow model fitting. These unwound points are then cleaned using the Local-outlier-factor algorithm (LOF, Breunig et al. 2000). For each arm 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 Figure. 5.

For each arm cluster in each galaxy, a logarithmic spiral model was fitted using Bayesian Ridge Regression, performed using the Scikit-learn python package. 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). 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 Figure 6.

2.6 Model Tuning

As mentioned above, we anticipated the need for a numerical fit to fine-tune parameters of *Galaxy Builder* models. This tuning was performed using the L_BFGS-b algorithm (Byrd et al. 1995), implemented in Scipy (Jones et al. 01), with a mean-squared error metric. Parameter bounds were chosen to be as uninformative as possible, while preventing catastrophic fitting failure where possible. All parameter bounds can be found in Table B1. Tuning the best individual models did not significantly improve residuals, but due to the large presence of parameters left at their defaults / extrema (36% of all disc components drawn by volunteers were left at the default axis ratio, and over half of all bars were left with their default rotation, 12% of the best individual galaxies had disc axis ratios left at the default values), tuning was neccessary as an attempt to correct for this bias. Models and residuals for the tuned best individual and aggregate models for our example galaxy are shown in Figure 7;

Both forms of tuned model generally accounted for roughly 50% of the light present in a cutout, and reduced the peak pixel brightness to 25% that of the original value. However, both of these measures are hugely dependent on the signal-to-noise of each galaxy and whether all secondary sources where correctly identified and masked.

2.7 Error Estimation

The uncertainties reported by many software fitting packages (Galfit and MegaMorph from the above list) are in fact lower estimates, due to secondary sources not being modelled, flat-fielding errors and incorrect models (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.

As all shapes in a cluster can be viewed as volunteers' attempts at modelling the underlying component, the sample variance of the parameters of these shapes can be used as a good approximation of the underlying variance of the component. Figure 4 illustrates the variance in clustered shapes for our example galaxy (UCG 4721). Effective radii generally had relative errors of around 15% and axis ratios show an absoulute error of around 0.13. Parameters dictated by sliders show much larger and less consistent errors, potentially due to their impact being conceptually harder for volunteers. It is therefore possible that slider errors should be treated as a measure of volunteer confidence rather than a measure of the posterior. A pivot table detailing all errors on parameters is provided in the appendix (Table B2).

3 RESULTS

In this Section we explore the consistency with which volunteers modelled galaxies, the variance of the aggregate

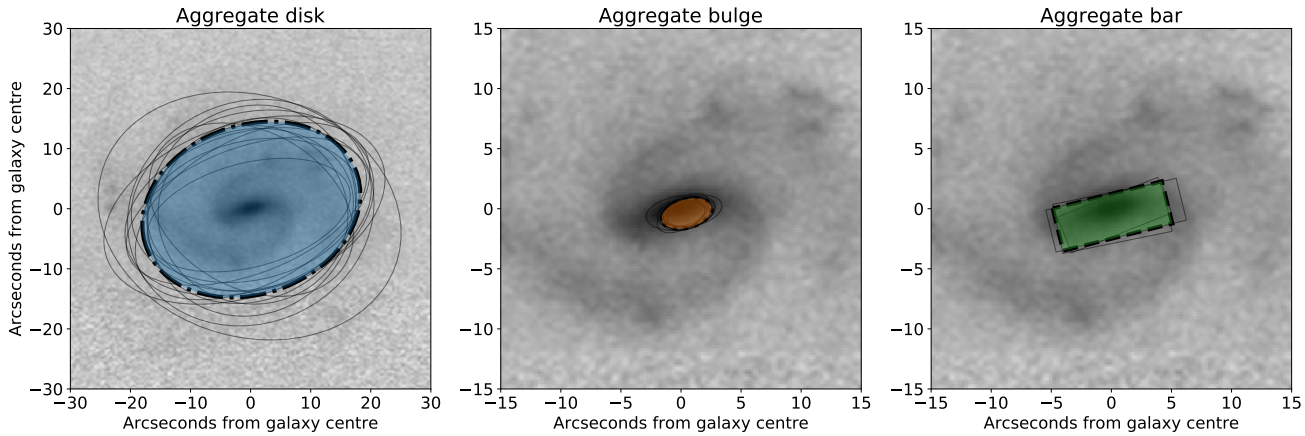

Figure 4. Calculated aggregate components for UGC 4721. The aggregate disc is shown using a dashed line and blue fill in the first panel, the aggregate bulge with a dash-dot line and orange fill in the second panel and the aggregate bar using a dotted line and blue fill in the third panel.

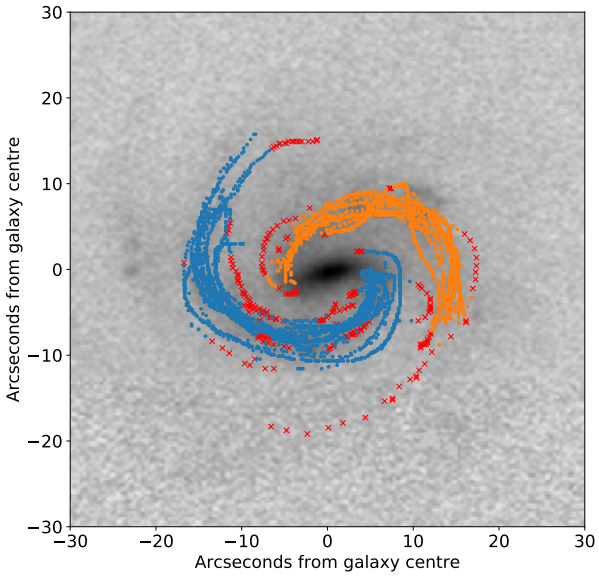

Figure 5. Point cleaning for the spiral arm clusters for UGC 4721. Points identified as outliers are displayed as red crosses, points used to fit log spirals are orange and blue dots.

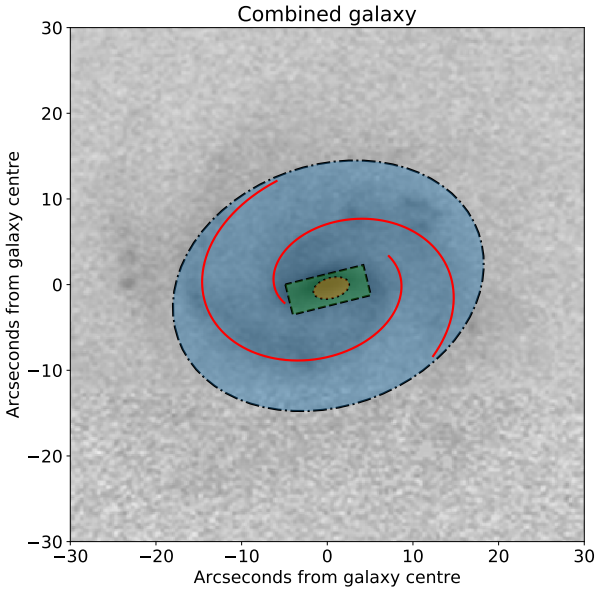

Figure 6. Resulting aggregate bulge + disc + bar + spiral arms components for UGC 4721. The disc is shown in blue with a dashed outline, the bulge in orange with a dash-dot outline, the bar in green with a dotted outline, and the spiral arms in red.

model recovered and how well our recovered models agree with other results in the literature.

3.1 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 2.4 for more on this selection).

One of the simplest choices the volunteers have is whether to include a model component or not. Figure 8 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 (scatter in fraction of 0.1), with volunteers almost always using a disc and bulge, and consistent proportions agreeing on the presence of a bar and the number of spiral arms.

After selecting a component, the volunteer sets its shape and size. We see generally good consistency in isophotal shape and size, with the least consistent component is the bar, which may be caused by the lower proportion of volunteers incorporating one into their model. Visual inspection suggests that many volunteers used a very elliptical bulge and drawn spirals to capture the light from the bar. Fewer bars having been drawn by volunteers also has the effect of making clustering more difficult and more uncertain, even for a strongly barred galaxy we effectively go from receiving 30 classifications to around 12 for the bar. The variation in axial ratios and effective radii for the aggregate discs, bulges and bars are shown in Figure 9: the sample error in disc size

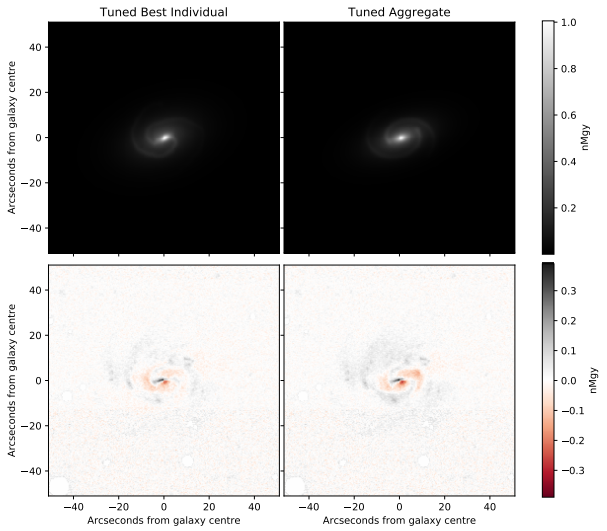

Figure 7. Tuned best individual and aggregate models, and their residuals for UGC 4721. The top two panels show the models, with the tuned best individual model on the left and the tuned aggregate model on the right. Bottom panels show the corresponding residuals, where red indicates oversubtraction of the galaxy and black indicates undersubtraction.

is 3.78 arcseconds, bulge size is 1.63 arcseconds and bar size is 1.52 arcseconds. There is very little consistency in axial ratio for the bulge and bar, however the disc axial ratio shows good consistency, with a sample error of 0.064.

3.2 Best Individual vs Aggregate Model

For each galaxy in the sample we compare the tuned aggregate model to the tuned best individual model, and find that the best individual model consistently outperforms the aggregate (around 70% of the time). However, there is a very small (less than a 5%) probability that any individual classifier will beat the aggregate model for more than half of the galaxies they model, with most volunteers outperforming the aggregate less than one in ten times. In summary, for an individual galaxy it is usually the case that the best individual model is best, but these models come from a wide variety of users, not a small number of super-users. It is notable how close many volunteers' models came to the optimial solution without any tuning.

Our recommendation is to make use of the tuned aggregate model for scientific analysis, though users may wish to instead make use of a number of volunteer models as starting points for their own numerical fits.

3.3 Comparison to results in the literature

After having obtained aggregated models for our galaxies, we examine how our models compare to other results in the literature. There exists no published comparison sample with four-component fits, so we are only able to do this separate components (or pairs of components).

We also make comparisons to existing measures of morphology for individual galaxies: When comparing the probability of a 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 a significant difference: classifications of strongly-barred galaxies ($p_{\rm bar} > 0.5$) had a 0.47 ± 0.14 chance of containing a bar, vs 0.30 ± 0.11 for galaxies classed as having no bar ($p_{\rm bar} < 0.2$). The Spearman correlation between GZ2's $p_{\rm bar}$ and the bar likelihood in *Galaxy Builder* is 0.56, which implies a significant correlation.

3.3.1 Comparison to One-component fit - axis ratio

We compare the axis ratios of the discs recovered from *Galaxy Builder* 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). We see excellent agreement (Figure 10), with an error of ~ 0.1, consistent with our expected errors (derived in Section 2.7). We find that 23.42% measurements with axis ratio less than 0.6 are outside 2σ, significantly higher than the expected 5%. This is opposed to 4.97% of measurements with axis ratio greater than 0.6. There is a clustering of outlying values at $b/a = 0.5$ which is almost certainly due to the drawing tool ellipse having a default axis ratio of 0.5. Where this is a "good enough" fit we hypothesise that volunteers are less likely to tweak it, while if it needs to move a long way they find a more refined value. We see similar behaviour in the cluster of values at $b/a = 1$ representing volunteers hitting the maximum value and not refining further.

3.3.2 Comparison to Disc-Bulge models

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, and an exponential disc and a Sérsic bulge model. 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 (exp+dV). They used a Levenberg-Marquadt gradient descent algorithm, with initial parameters taken from previous SDSS analysis.

Comparing between these catalogues and to *Galaxy Builder* models, we see that our models show excellent agreement with others, (when the models which are fit are comparable). Comparing bulge to total fraction, we see closest agreement to the exp+exp model, where bulge Sérsic index is similar to that most often chosen by *Galaxy Builder* volunteers, who consistently preferred to fit bulges with low Sérsic indices. Bulge measurements are very sensitive to central sub-structure and model choice (Gao & Ho 2017), so comparing B/T between models with very different bulge profiles would be expected to show a lot of scatter. The Kendall rank correlation coefficients between our measurments and all the different fits of Simard et al. (2011) and Lackner & Gunn (2012) (as well as those compared with each other) can be seen in Figure 11. Comparing GZB with the exp+exp models of both Simard and Lakner & Gunn we see the best agreement, while comparisons between exp+deV and exp+exp models show very poor agreements. The low correlation coefficients in general, even when comparing the

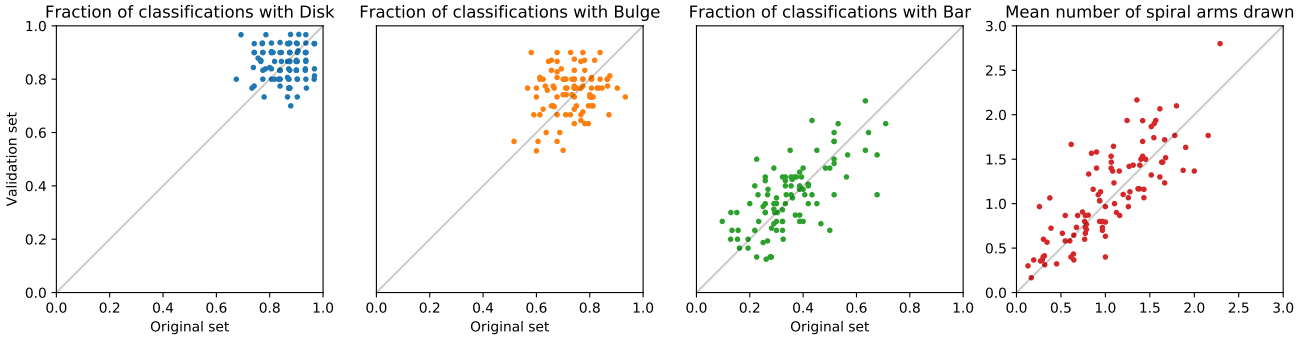

Figure 8. Comparison of frequency of use of component in volunteer models between the original and validation sets of classifications.

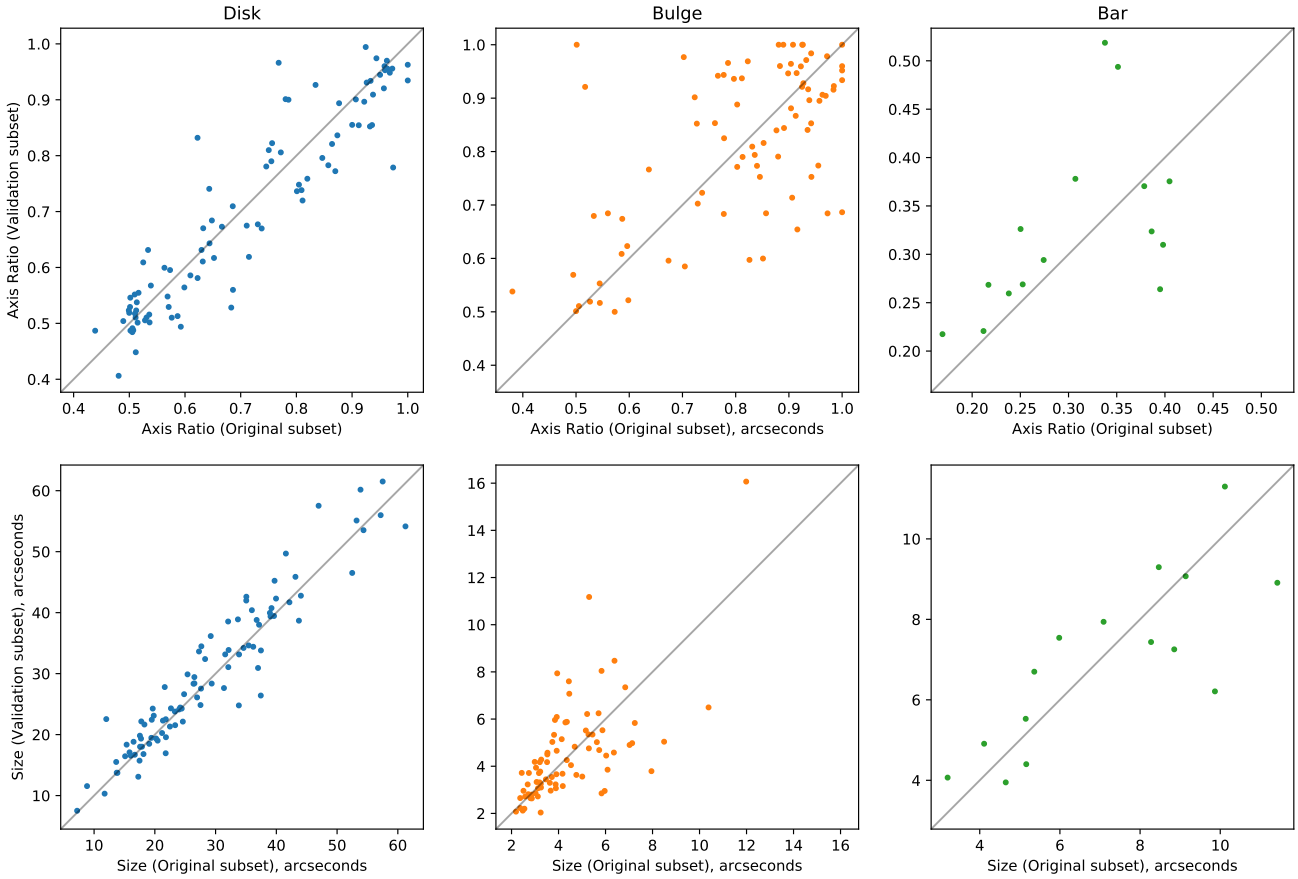

Figure 9. Comparison of component shape in aggregate models between the original and validation sets.

results of fitting identical models (the two exp+dV models), illustrate the difficulty of calculating reliable bulge to total ratios for galaxies. We conclude that *Galaxy Builder* models are able to agree with more conventional photometrically fit models as much as those models agree with each other, and provide a physically motivated B/T ratio.

3.3.3 Comparison to Disc-Bulge-Bar models

Kruk et al. (2018) performed multi-component, multi-band decompositions of a selection of SDSS galaxies, 12 of which were also classified in *Galaxy Builder*. Figure 12 compares the axis ratios and effective radii of bulges, discs and bars in Kruk et al. (2018) to those present in the tuned best individual classification. We see strong agreement in effective radius, with disc, bulge and bar sample standard deviations being 9.88“, 2.65“ and 3.07“ respectively. We do note that *Galaxy Builder* bulges generally have a larger effective radius than the bulges in Kruk et al. (2018). This discrepancy could be caused by the differences in bulge Sérsic index, as bulges present in Kruk et al. (2018) tend to have lower Sérsic indices than those in the *Galaxy Builder* models: lower Sérsic indices generally result in a higher best-fit effective radius, though this is highly dependent on the PSF and resolution of the

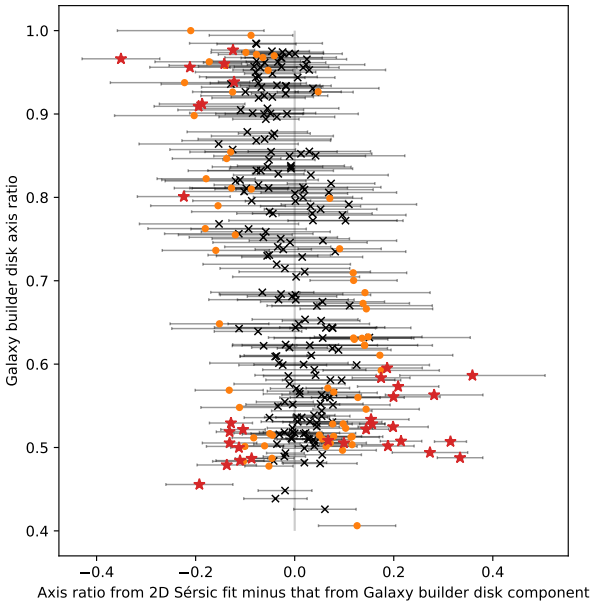

Figure 10. Difference between the axis ratios of the disc components of aggregated *Galaxy Builder* models to the results of an r-band Sérsic profile fit. Points outside 1- and 2σ are highlighted in orange and red.

image. We observe a large scatter in axis ratio, though an overall correlation.

3.3.4 Comparison to Disc-Bulge-Bar-Spiral models

To the best of our knowledge, no photometic 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 and therefore not available to this project.

In order to provide a comparison for our novel method of spiral parameter (pitch angle and amplitude) extraction, we compare the result of our logarithmic spiral fit to the relationship obtained by Hart et al. (2016) between GZ2 classification and galaxy pitch angle (Figure 13). Their fit was obtained by using the Zooniverse to filter good vs bad spiral arm segments identified using a leading 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 uncertainties on the Hart et al. (2016) fit, even when not accounting for error on our measurements.

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 hope to further explore this issue in a future work.

4 SUMMARY AND CONCLUSIONS

In this paper we present a novel method for modelling of galaxy images, *Galaxy Builder*[9], which was conceived with the goal of solving the "quality of quantity" dilema facing galaxy image modelling, which, despite advances in computation, still typically requires significant human interaction to achieve quality fits.

Galaxy Builder leverages the power of crowd sourcing for the hardest to automate parts of image fitting, namely determining the appropriate number of model components to include, and finding parameters which are close to the global optima.

We have demonstrated that we are able to obtain models with residuals with a median maximum value of 20% of the maximum galaxy brightness, using either the best individual classification provided by volunteers or an aggregate model from clustering. We obtain errors on parameters through the sample standard deviation of component clusters, which better respects the complex curvature of the likelihood space than simple jacobian approximations.

We compare these new models to existing results in the literature where available. We find good agreement where the models or parameters are comparable, and comment on instances where *Galaxy Builder* should provide superior models.

We were able to obtain models for 296 images with a rate of one galaxy per day. We note that user experience and task simplification will need to be considered if significantly larger numbers of these models are to be obtained.

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.

We release our catalogue of models to the community, and in future work we use this sample to investigate spiral arm formation mechanisms (T. Lingard et al. in prep.).

5 ACKNOWLEDGEMENTS

This publication made use of SDSS-I/II data. Funding for the SDSS and SDSS-II was 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/`.

This project was partially funded by a Google Faculty Research Award to Karen Masters (`https://ai.google/research/outreach/faculty-research-awards/`)

Montage is funded by the National Science Foundation under Grant Number ACI-1440620, and was previously funded by the National Aeronautics and Space Administration's Earth Science Technology Office, Computation Technologies

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

Figure 11. Correlation matrix showing Kendall rank correlation coefficient between measures of Bulge to Total fraction from *Galaxy Builder* results and other models fitted in Simard et al. (2011) and Lackner & Gunn (2012). Colours and box size indicate the strength of the correlation.

Project, under Cooperative Agreement Number NCC5-626 between NASA and the California Institute of Technology.

References

Abazajian K. N., et al., 2009, ApJS, 182, 543
Albareti F. D., et al., 2017, ApJS, 233, 25
Bamford S. P., Häußler B., Rojas A., Borch A., 2011, in Evans I. N., Accomazzi A., Mink D. J., Rots A. H., eds, Astronomical Society of the Pacific Conference Series Vol. 442, Astronomical Data Analysis Software and Systems XX. p. 479
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Blanton M. R., Kazin E., Muna D., Weaver B. A., Price-Whelan A., 2011, AJ, 142, 31
Blanton M. R., et al., 2017, AJ, 154, 28
Boonchoo T., Ao X., He Q., 2018, CoRR, abs/1801.06965
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. ACM, New York, NY, USA, pp 93–104, doi:10.1145/342009.335388, `http://doi.acm.org/10.1145/342009.335388`
Byrd R., Lu P., Nocedal J., Zhu C., 1995, SIAM Journal on Scientific Computing, 16, 1190
Davis D. R., Hayes W. B., 2014, ApJ, 790, 87
Díaz-García S., Salo H., Knapen J. H., Herrera-Endoqui M., 2019, arXiv e-prints, p. arXiv:1908.04246
Dobbs C., Baba J., 2014, Publ. Astron. Soc. Aust., 31
Gadotti D. A., 2011, MNRAS, 415, 3308
Gao H., Ho L. C., 2017, ApJ, 845, 114

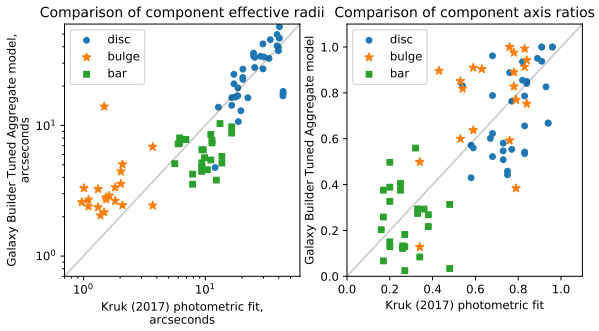

Figure 12. Comparison between *Galaxy Builder* tuned aggregate models and the result of 3-component, multiwavelength fits performed by Kruk et al. (2018). Discs, Bulges and Bars are shown as blue circles, orange stars and green squares respectively. The left panel compares component effective radius, the right panel compares the component axis ratio.

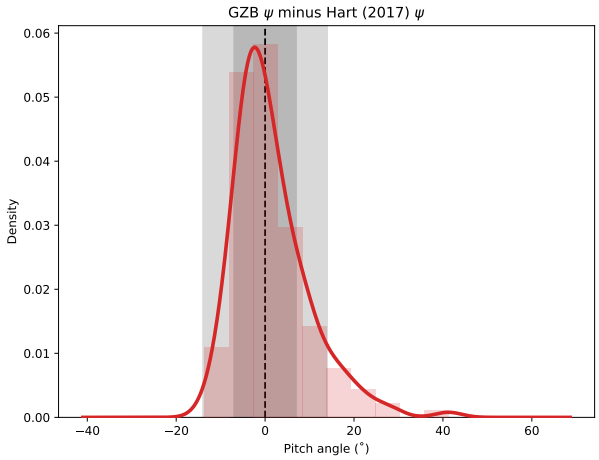

Figure 13. A comparison of Pitch angle obtained by Hart et al. (2016) with measured pitch angles for the aggregated model results in galaxies in the Galaxy Zoo Builder sample. The grey regions show 1- and 2σ errors from Hart et al. (2016). Errors on *Galaxy Builder*-measured pitch angles are not accounted for.

Gao H., Ho L. C., Barth A. J., Li Z.-Y., 2018, ApJ, 862, 100
Hart R. E., et al., 2016
Hart R. E., et al., 2017, MNRAS, 472, 2263
Holincheck A. J., et al., 2016
Hopkins P. F., et al., 2010, The Astrophysical Journal, 724, 915
Jacob J. C., et al., 2010, arXiv e-prints, p. arXiv:1005.4454
Jones E., Oliphant T., Peterson P., et al., 2001–, SciPy: Open source scientific tools for Python, `http://www.scipy.org/`
Kormendy J., Drory N., Bender R., Cornell M. E., 2010, The Astrophysical Journal, 723, 54
Kruk S. J., et al., 2017, MNRAS, 469, 3363
Kruk S. J., et al., 2018, MNRAS, 473, 4731
Lackner C. N., Gunn J. E., 2012, MNRAS, 421, 2277
Lintott C. J., et al., 2008, Monthly Notices of the Royal Astronomical Society, 389, 1179
Lupton R., Blanton M. R., Fekete G., Hogg D. W., O'Mullane W., Szalay A., Wherry N., 2004, PUBL ASTRON SOC PAC, 116, 133
Masters K. L., et al., 2010, Monthly Notices of the Royal Astronomical Society, 411, 2026
Mendez-Abreu J., et al., 2016
Park C., Choi Y.-Y., Vogeley M. S., III J. R. G., Blanton M. R., the SDSS collaboration F., 2007, ApJ, 658, 898
Pedregosa F., 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
Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2010, AJ, 139, 2097
Pour-Imani H., Kennefick D., Kennefick J., Davis B. L., Shields D. W., Abdeen M. S., 2016, The Astrophysical Journal, 827, L2
Pozzetti L., et al., 2009, Astronomy & Astrophysics, 523, A13
Robotham A. S. G., Taranu D. S., Tobar R., Moffett A., Driver S. P., 2016
Salo H., et al., 2015, ApJS, 219, 4
Sanchez S. F., et al., 2011, Astronomy and Astrophysics, 538, A8
Simard L., et al., 2002a, ASTROPHYS J SUPPL S, 142, 1
Simard L., et al., 2002b, The Astrophysical Journal Supplement Series, 142, 1
Simard L., Mendel J. T., Patton D. R., Ellison S. L., McConnachie A. W., 2011, ApJS, 196, 11
Simmons B. D., et al., 2017, MNRAS, 464, 4420
Simpson R., Page K. R., De Roure D., 2014, in Proceedings of the 23rd International Conference on World Wide Web. WWW '14 Companion. ACM, New York, NY, USA, pp 1049–1054, doi:10.1145/2567948.2579215, `http://doi.acm.org/10.1145/2567948.2579215`
Stoughton C., et al., 2002, AJ, 123, 485
The Astropy Collaboration et al., 2018, AJ, 156, 123
Vika M., Bamford S. P., Häußler B., Rojas A. L., 2014, Monthly Notices of the Royal Astronomical Society, 444, 3603
Willett K. W., et al., 2013
Willett K. W., et al., 2017, MNRAS, 464, 4176
Zaninetti L., 2014, arXiv e-prints, p. arXiv:1401.0287

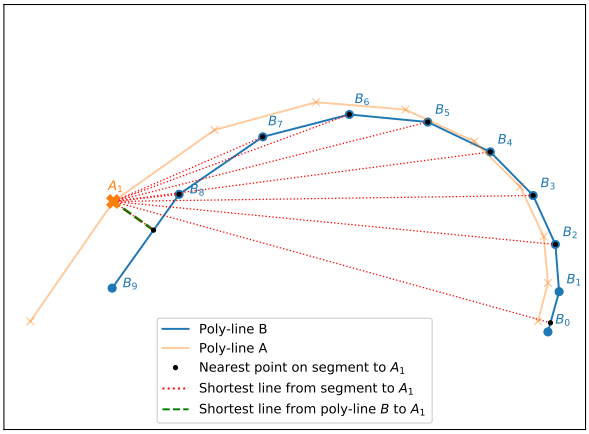

Figure A1. Illustration of the metric used. For each point in line $\boldsymbol{A}$, the shortest distance to each segment in line $\boldsymbol{B}$ is calculated (shown as dotted red lines). The minimum of these distances (corresponding to the line shown in dashed green) is squared.

APPENDIX A: MATHEMATICAL DESCRIPTION OF THE POLY-LINE SEPARATION MEASURE

This appendix details the metric used in Section 2.5.2 to cluster polylines used by volunteers to model spirals arms. It can be seen as a variant of the Fréchet distance. The metric is illustrated in Figure A1.

First, define a poly-line containing n 2D cartesian coordinates (vertices) as

$$A : \{i \in \mathbb{N};\ i < n\} \longrightarrow \mathbb{R}^2 \tag{A1}$$

We also define a function, t, which calculates how far a point $\vec{p}$ is along the line between two other points ($\vec{v}$ and $\vec{w}$):

$$t(\vec{p},\ \vec{v},\ \vec{w}) \equiv \frac{(\vec{p} - \vec{v}) \cdot (\vec{v} - \vec{w})}{|\vec{w} - \vec{v}|^2}. \tag{A2}$$

The minimum distance from $\vec{p}$ to the line segment between $\vec{v}$ and $\vec{w}$ is given by

$$d(\vec{p},\ \vec{v},\ \vec{w}) = \| \left(\vec{v} + \min(\max(t(\vec{p},\ \vec{v},\ \vec{w}),\ 0),\ 1)\ (\vec{w} - \vec{v})\right) - \vec{p}\| \tag{A3}$$

We then define a "squared distance" from the poly-line A (containing n vertices) to the poly-line B (containing m vertices):

$$D(A,\ B) \equiv \frac{1}{n} \sum_{i=0}^{n} \min\{j \in \mathbb{N}_0,\ j < m;\ d(A_i,\ B_j,\ B_{j+1})^2\}. \tag{A4}$$

The choice to square the distances and penalize large deviations from other lines was a data-driven choice to improve the results of clustering.

Finally, we define our separation measure between two drawn poly-lines as

$$distance(A,\ B) \equiv D(A,\ B) + D(B,\ A). \tag{A5}$$

APPENDIX B: ANCILLARY TABLES

This paper has been typeset from a TEX/LATEX file prepared by the author.

Table B1. 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 was similarly unconstrained.

Component	Parameter	Default	Volunteer allowed Minimum	Volunteer allowed Maximum	Tuning Minimum Bound	Tuning Maximum Bound
disc	axRatio	0.5	0.0	inf	0.01	100
	scale	2.0	0.0	1.00	0	inf
	i0	1.0	0.0	0.20	0	inf
bulge	axRatio	0.5	0.0	inf	0.01	100
	scale	2.0	0.0	1.00	0	inf
	i0	2.0	0.0	0.50	0	inf
	n	5.0	0.5	1.00	0.1	10
bar	roll	0.0	-inf	inf	0.01	100
	scale	2.0	0.0	1.00	0	inf
	i0	1.0	0.0	0.20	0	inf
	n	2.0	0.3	0.50	0.1	10
	c	3.0	1.5	2.00	0.01	10
spiral	i0	1.0	0.0	0.75	0	inf
	spread	2.0	0.0	1.00	0	inf
	falloff	2.0	0.0	1.00	0.01	inf

Table B2. Pivot table of reported errors on parameters for our sample of 297 galaxies. All errors apart from b/a are relative errors.

component	parameter	count	mean	std	min	25%	50%	75%	max
disc	b/a	292	0.09	0.04	0.01	0.06	0.09	0.12	0.20
	Σ_e	292	0.51	0.25	0.06	0.30	0.47	0.69	1.20
	r_e	292	0.18	0.05	0.04	0.14	0.17	0.21	0.34
bulge	b/a	273	0.10	0.05	0.00	0.07	0.10	0.14	0.23
	Σ_e	273	0.50	0.25	0.00	0.32	0.51	0.68	1.31
	n	273	0.69	0.27	0.00	0.47	0.76	0.90	1.26
	r_e	273	0.15	0.07	0.00	0.10	0.14	0.19	0.44
bar	b/a	81	0.07	0.03	0.01	0.05	0.07	0.09	0.14
	c	81	0.13	0.08	0.00	0.08	0.13	0.20	0.35
	Σ_e	81	0.56	0.37	0.00	0.23	0.53	0.87	1.37
	n	81	0.47	0.30	0.00	0.22	0.51	0.73	0.97
	r_e	81	0.18	0.09	0.02	0.13	0.17	0.23	0.44