MNRAS **000**, 1–12 (2019) Preprint 9th August 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 Masters[1,2], Coleman Krawczyk[1], Chris Lintott[3], Steven Bamford,[4], Sandor Kruk,[5], Brooke Simmons,[6], Robert Simpson,[7], Bob 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]*ESA, Leyden, NL* [6]*Department, Institution, Street Address, City Postal Code, Country* [7]*Department, Institution, Street Address, City Postal Code*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT
This paper presents a novel interface, built inside the Zooniverse citizen science platform, which enables volunteers to help to create detailed photometric models of galaxies from SDSS images. The value of citizen science to obtain classifications of galaxy morphology on a large scale has been widely demonstrated in the literature, however such morphologies can lack the quantitative power obtained through the fitting of photometric profiles. Multi-component modelling of complex galaxies is plagued by issues with convergence, model selection and parameter degeneracies, which we demonstrate can be adressed using citizen science. We examine the consistency of this new method with changes in number and population of citizen scientists and compare it to more traditional automated fitting pipelines [[**Chris: how does it get on?**]]. 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

1.1 Galaxy Morphology

Test. One of the cornerstones of all areas of empirical science over the past few centuries has been the sifting-through and sorting of objects of interest into types; this has been true of the field of galaxy classification since the mid-1930s, when Hubble began developing his tuning-fork model of the galaxies observed in photographic images of the sky. (Hubble 1936).

It is commonly accepted that this tuning fork, even with its subsequent expansions (for example that used in Sandage 1961 and de Vaucouleurs et al. 1991), is too simplistic and subjective a measure for systems as complex as galaxies. One of the tuning fork's greatest strengths is that it provides a framework for basic ideas, including the notion that galaxies are comprised of distinct components such as discs, bulges, spiral arms and bars. Recently, it appears that expert classification of galaxies has shifted away from Hubble's original spiral tightness classification for spiral galaxies and towards bulge size, which appears not to be correlated with spiral tightness (Masters et al. 2019).

Researchers have begun to explore alternate methods of classification, including the use of rotation curves and internal dynamics (Cappellari et al. 2011, Kalinova et al. 2017, Fall & Romanowsky 2018).

One major problem encountered in morphological classification is that it is difficult to avoid subjectjivity in a classification scheme: Naim et al. (1995) noted that between their sample of six experts, a galaxy would only receive a 50% consensus on Hubble type (more than three identical, independent classifications from the experts) 55% of the time, with all experts agreeing independently on a classification for only 8 of their sample of 354 galaxies.

Grouping galaxies based on their morphology also results in the clustering of other physical parameters such as star formation rate or gas fraction (Roberts & Haynes 1994 provides a detailed discussion of many such correlations [[**A more recent citation would be better**]]). It also enables the selection of morphology-based sample sets from which physical processes and characteristics can be probed.

Until the late 20th Century, it was possible for small teams to band together and compile catalogues of classifications of many of the well-observed galaxies at the time (i.e

★ E-mail: tim.lingard@port.ac.uk

The Third Reference Catalog of Bright Galaxies, de Vaucouleurs et al. 1991, containing 18,000 classifications, or the ESO catalog of galaxies, Lauberts & Valentijn 1989, containing 15,000 classifications). However, with the beginning of the era of large sky surveys such as the Sloan Digital Sky Survey (hereafter SDSS, Blanton et al. 2017, Abazajian et al. 2009, over 50 million galaxies), the time required for classifying galaxies grew to unsustainable levels for most research teams. See Nair & Abraham (2010), Schawinski et al. (2007) for the largest expert efforts. Naim et al. (1995) was one of the first to discuss the need to move to automated photometric classification, and investigated possible methods of automation, including the use of an Artificial Neural network.

One approach to solving the problem of large-scale classification was to identify proxies for morphology, such as colour or concentration index (for example, the "Zest" method of Scarlata et al. (2007)). These methods could then be easily rolled-out to the required scales. However, the use of proxies introduces some unknown bias in the resulting classifications (i.e. not all spiral galaxies have blue outer regions). Different sample sets created using different morphological proxies will differ in some difficult-to-quantify manner, due to the underlying nature of the galaxies being classified. The usage of catalogues created using some form of morphological proxy may also be statistically unsound for particular science cases; for instance studying star formation rates using a catalogue compiled using optical colours will obviously produce a biased result.

The trend towards use of colour, or automated proxies for morphology (such as bulge size), has also resulted in a subtle shift in what we mean by certain morphological terms (like late- or early-type), so care must be taken comparing with older morphological based classifications (see Masters et al. 2019 for a longer discussion of this point).

1.2 Light Distribution Modelling

Another way to characterise an image of a galaxy is to make use of analytic functions to model the components of that galaxy. These fully quantitative methods allow researchers to obtain structural parameters of galaxy sub-components, which can be 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 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. 2002a), 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. (2002b) 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 (2010) 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. (2002b). 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. 2017a).

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. The extra degree of freedom of the Sérsic profile allows a more desirable residual (Kruk et al. 2017a). 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 manually checking many of their fits to ensure they have converged on a physical model, meaning we are faced with a similar problem of scale discussed in Section 1.1.

1.2.1 Citizen Science

A promising solution to the problem of large-scale classification 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.

This paper proposes an analogous solution to that introduced by Lintott et al. (2008) for the issues faced by galaxy parametric modelling inside the ecosystem that their research set in motion; leveraging citizen scientists to pick

model components and perform model optimization in an online, web-browser environment. 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 $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

2 METHOD

2.1 The Galaxy Builder Zooniverse project

Galaxy Builder[1] is a citizen-science project built on the Zooniverse[2] 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, we had to design and implement a (comparatively basic) model rendering suite inside the existing Zooniverse front-end code-base. We had to not only consider the accuracy of the resulting model, but also user experience and engagement in our design decisions. Feedback from expert users was essential to this process as part of the the typical beta trial process for Zooniverse projects [[**citation needed**]].

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.

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.

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).

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

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

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

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.

The rendering 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[6]). This enables the model (and residual) to be calculated at a speed which allows low-latency feedback to a

[1] https://www.zooniverse.org/projects/tingard/galaxy-builder/
[2] https://www.zooniverse.org
[3] http://github.com/zooniverse/Panoptes-Front-End
[4] https://reactjs.org/
[5] https://www.khronos.org/webgl/
[6] http://regl.party/

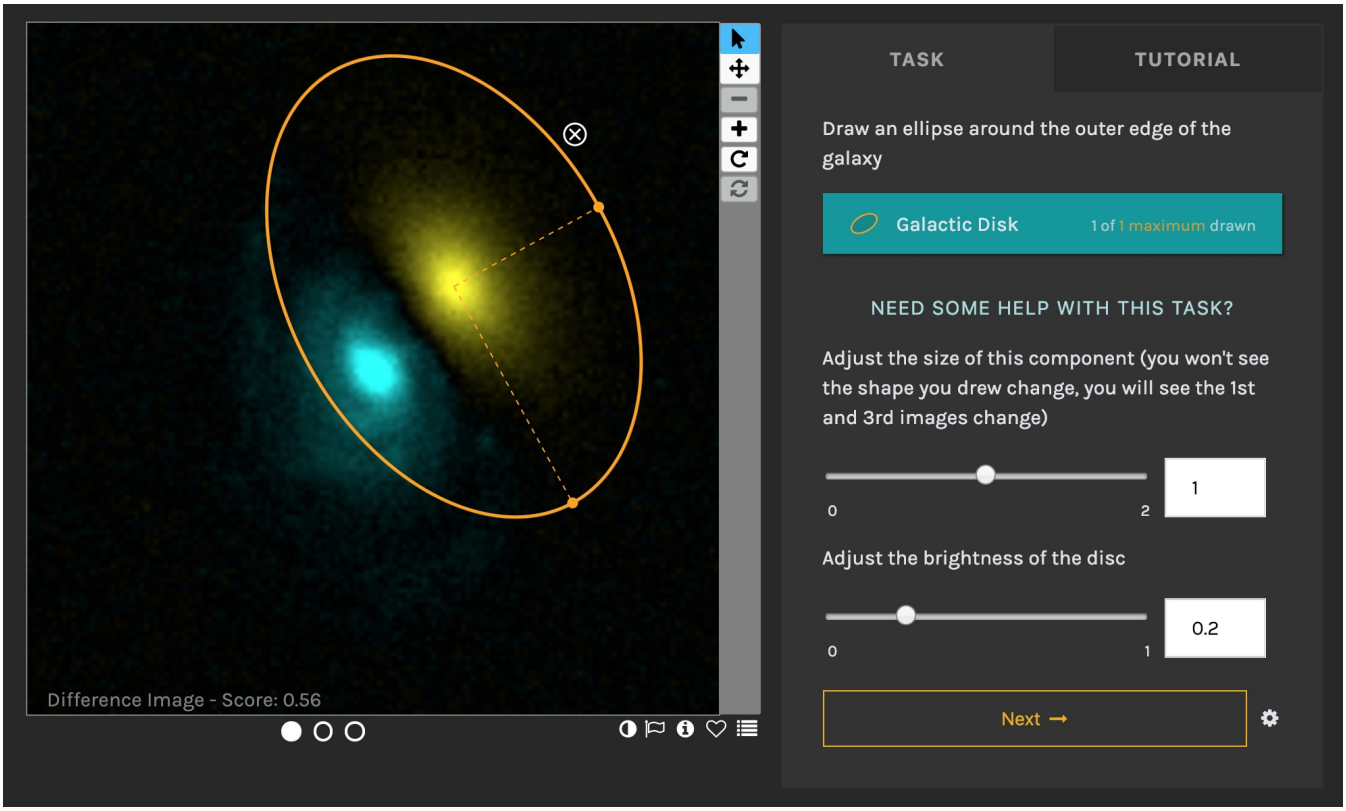

Figure 1. The Galaxy Builder interface. The residual image is being shown, and the volunteer is on the "Disc" task. The drawn disc component (yellow) is offset from the galaxy image (blue) to demonstrate the positive and negative residuals. Where the image equals the model the residual is black. The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

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. As these models are envisioned as a starting point for a later numerical fit, it was determined this precision was sufficient.

2.2 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 following:

- GZ2 $p_{\rm features} \cdot p_{\rm not\ edge\ on} \cdot p_{\rm spiral} \geq 0.5$
- $0.02 < z < 0.055$
- g-band axial ratio $a/b > 0.4$
- r-band magnitude $14 < m_r < 17$
- A volume correction, excluding galaxies outside $9.45 < \log(M_*/M_\odot) \leq 11.05$
- Computation of stellar mass completeness limits using the method of Pozzetti et al. (2009) and limits calculated by Hart et al. (2017)

$9.45 < \log(M_*/M_\odot) \leq 11.05$
$\log(M_*/M_\odot) > 2.07\log(z) + 12.64$
$\log(M_*/M_\odot) < 2.45\log(z) + 14.05.$

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 fascilitate the unknown rate at which volunteers would provide classifications. Two out of the two-hundred selected galaxies failed to run through the image preparation process, due to an error when attempting

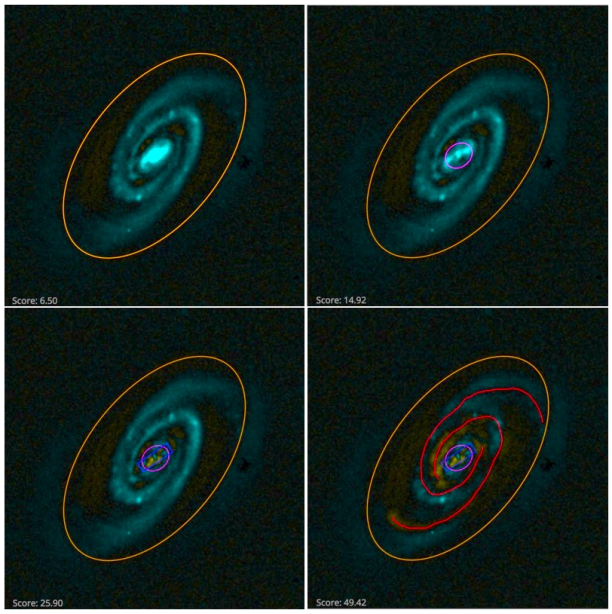

Figure 2. Figure demonstrating the desired result of each step of the modelling process, as seen from the residual image provided to volunteers. The top left panel shows the galaxy after only a disc component has been added: the top right contains a disc and a bulge; the bottom left has a disc, bulge 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.

to montage multiple frames. The root cause of this error is unknown.

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 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 foreground stars 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 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 asinh-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 other bands.

Once a sub-sample had been created, the Zooniverse's PANOPTES-PYTHON-CLIENT[7] was used to upload them as a subject-set to the Zooniverse.

2.3 The Galaxy Model

The model we chose was largely based off 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 allowed to vary from 0.5 to 5); a Sérsic bar with a "boxiness" modifier (as described in Peng et al. 2002); freehand poly-line spiral arms.

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 correctly ignores masked regions identified in the images. 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 volume of data 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, the remaining galaxies from the initial two subsamples 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 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.

2.5 Classification Aggregation Methodology

In this Section, we will use galaxy UGC 4721 as an example galaxy for the illustration of the data reduction and aggreg-

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

ation 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).

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

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 robustness and speed. We made use of Scikit-learn (Pedregosa et al. 2011) to implement the algorithm. The parameters `eps` and `min_points` were chosen by visually inspecting the resulting clustering results.

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). 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 to the spiral width slider values returned by volunteers (ignoring sliders left at the default or moved to the extremes of allowed values).

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). Parameter bounds were chosen to be as uninformative as possible, with bulge Sérsic index restricted to between 0.1 and 10. Tuning the best individual models did not significantly improve residuals, but due to the large presence of parameters left at their defaults / extrema [[**give a number / fraction here**]], tuning was neccessary for aggregate models. Models and residuals for the tuned best individual and aggregate models is shown in Figure 7.

2.6.1 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 [[**citation needed showing similar method in GZ2**]]. Figure 4 illustrates the variance in clustered shapes for the example galaxy, we recover.

3 RESULTS

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

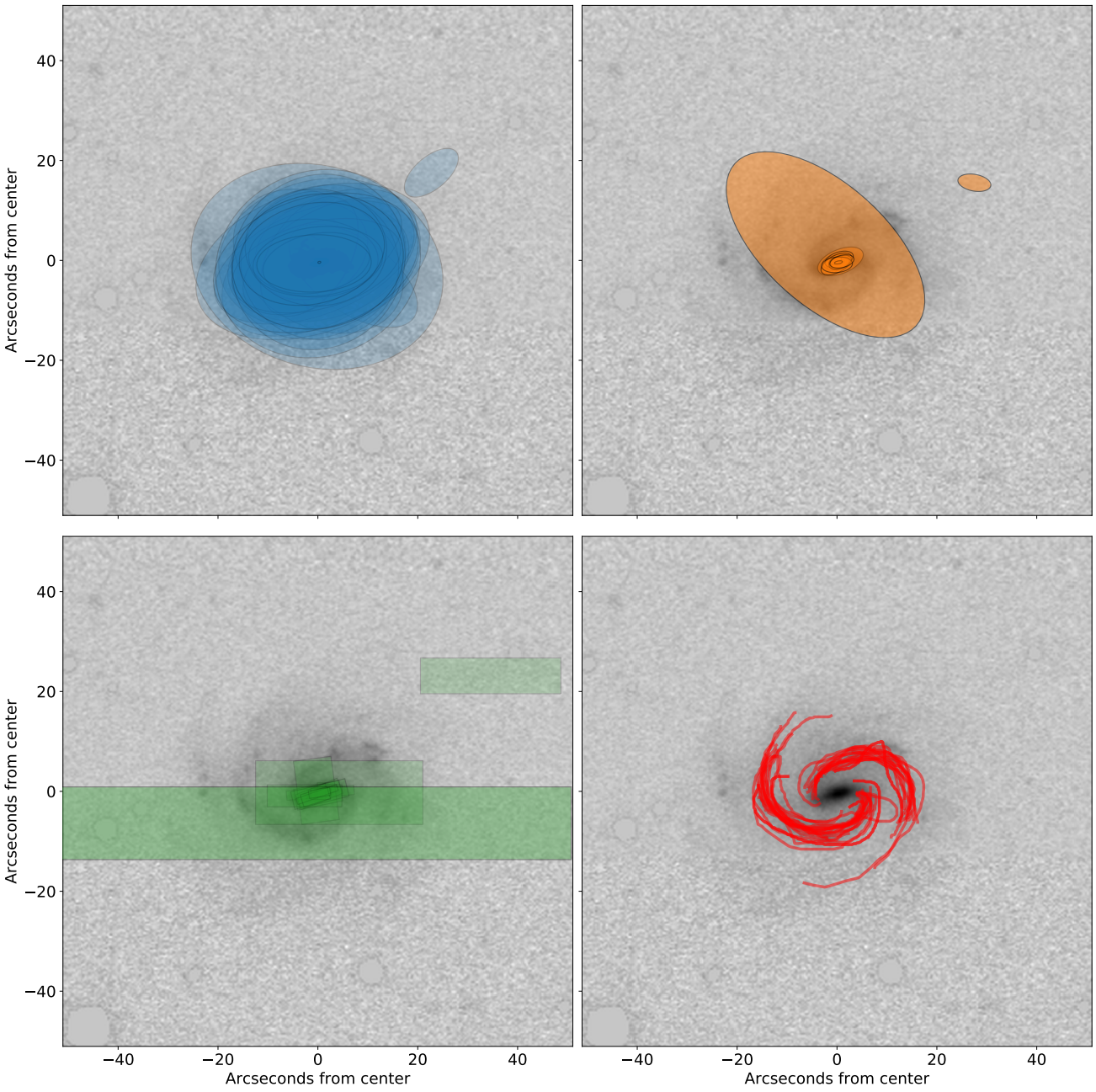

Figure 3. Components drawn by volunteers for UGC 4721. The top left panel shows drawn discs (blue), top right shows drawn bulges (orange), bottom left shows drawn bars (green) and bottom right shows drawn spiral arms.

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 (error 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 good consistency in isophotal shape and size, as shown in Figure 9. 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,

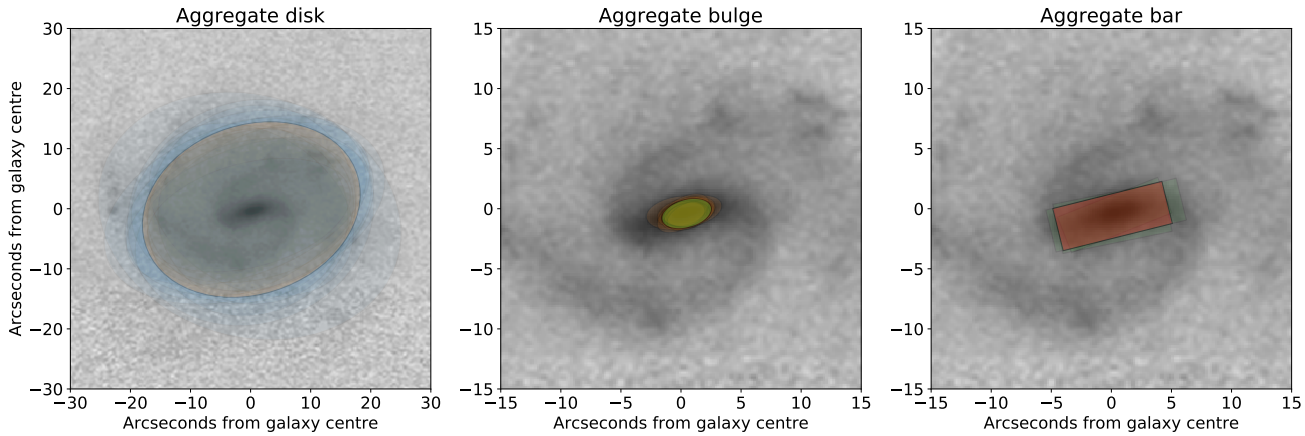

Figure 4. Calculated aggregate components for UGC 4721. The aggregate disc is shown in orange in the first panel, the aggregate bulge in green in the second panel and the aggregate bar in red in the third panel.

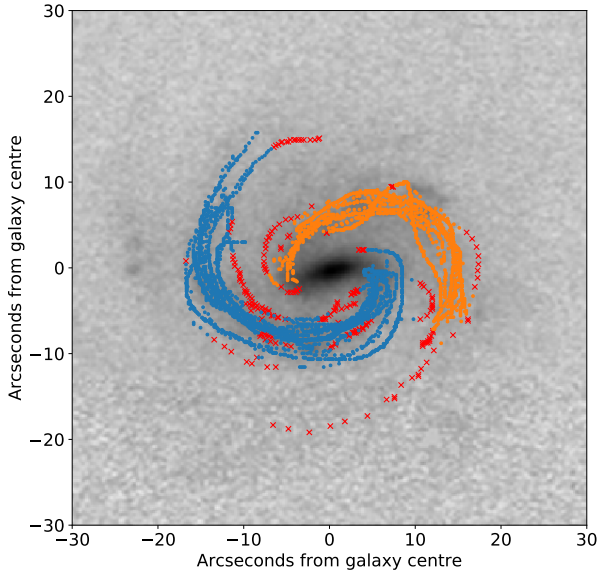

Figure 5. Point cleaning for the spiral arm clusters for UGC 4721. Points identified as outliers are displayed as red crosses, points used to fit log spirals are orange and blue dots.

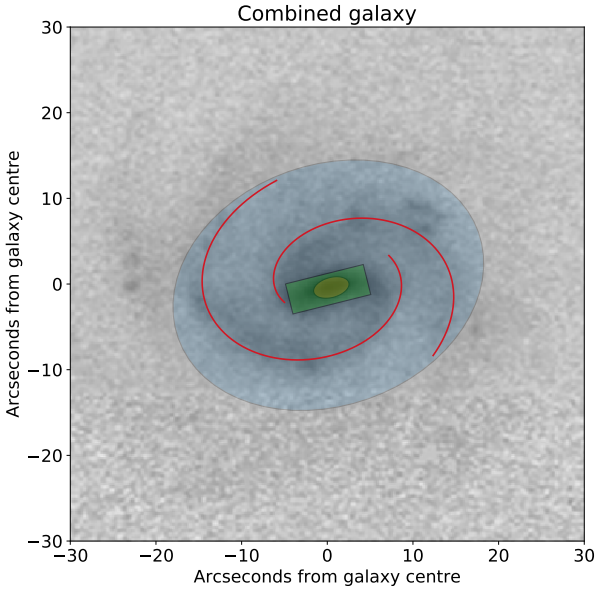

Figure 6. Resulting aggregate bulge + disc + bar + spiral arms components for UGC 4721. The disc is shown in blue, the bulge in orange, the bar in green and the spiral arms in red.

even for a strongly barred galaxy we effectively go from receiving 30 classifications to around 12 for the bar

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 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.

[[**Should we make a big deal about how this is true even for expert classifiers? Though I am comparing raw volunteer models to tuned aggregate models, and we don't know if tuning the expert volunteer models would make a big difference**]]

3.3 Comparison to results in the literature

3.3.1 Disk, Bulge and Bar

After having obtained aggregated models for our galaxies, we examine the reliability of our models through comparison to other results in the literature. For instance, if 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 NASA-Sloan Atlas), we see excellent agreement (Figure 10), with an error of ~ 0.1, consistent with our expected errors (derived in Section 2.6.1). We observe a large number of measurements outside 2σ around a Galaxy Builder axis ratio of 0.5, which

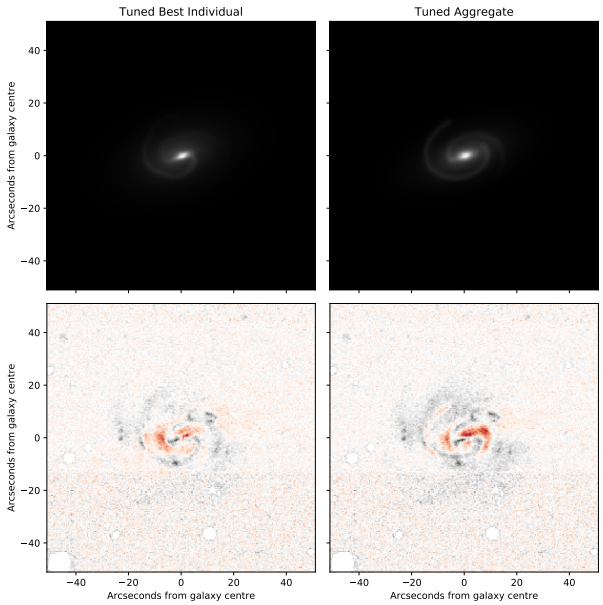

Figure 7. Tuned best individual and aggregate models, and their residuals for UGC 4721. The top two panels show the models, with the tuned best individual model on the left and the tuned aggregate model on the right. Bottom panels show the corresponding residuals, where red indicates oversubtraction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

is likely to be due to the drawing tool ellipse having a default axis ratio of 0.5, and biasing volunteer classification or aggregate shapes.

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.

One of the largest catalogs of 2D many-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 disk and exponential bulge, and an exponential disk and de Vaucouleurs bulge. 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 consistency with others, where such consistency is to be expected. As bulge measurements are very sensitive to central substructure and model choice (Gao & Ho 2017), when comparing bulge to total ratio we see good agreement to models where bulge Sérsic index is similar to that most often chosen by Galaxy Builder volunteers (who consistently preferred to use bulges with low Sérsic indices), as bulge measurements are very sensitive to central substructure and model choice (Gao & Ho 2017). The correlation coefficients between our measurments and Simard et al. (2011) and Lackner & Gunn (2012) can be seen in Figure 11.

Kruk et al. (2017b) performed many-component, multi-band decompositions of a selection of Sloan galaxies, 12 of which were also classified in Galaxy Builder. For each of these galaxies we obtain the "best" model provided by volunteers (scored using mean squared error, in units of nanomaggies) and futher optimize the slider parameters available to volunteers, as well as the effective radius and axis ratio of all components. Figure 12 compares the axis ratios and effective radii of bulges, discs and bars in Kruk et al. (2017b) to those produced here.

3.3.2 Spiral Arms

In order to benchmark the reliability of this method of spiral parameter 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). We find good agreement, although there are large error bars on the GZ2-produced pitch angle (and a caveat in the paper that these pitch angles should not be used for individual galaxies).

4 CONCLUSIONS

This paper has explored how citizen science could impact the problem of photometric modelling of spiral galaxies using complex models. We detailed a novel project inside the popular and successful Zooniverse citizen science platform that enables galaxy photometric decomposition in near-realtime.

We have demonstrated that it is possible to obtain models from volunteer classifications, using both the best individual classification and an aggregate model from clustering, which show good agreement with existing results in the literature.

The complexity of the tasks involved in this project coupled with the difficulties in clustering and aggregating data suggest that this is one problem space the power of the crowd hasn't quite overcome yet. Despite this, we remain of the opinion that better user experience design & in-browser computer optimization will help improve the quality of classifications and thus that of the recovered models.

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.

5 ACKNOWLEDGEMENTS

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

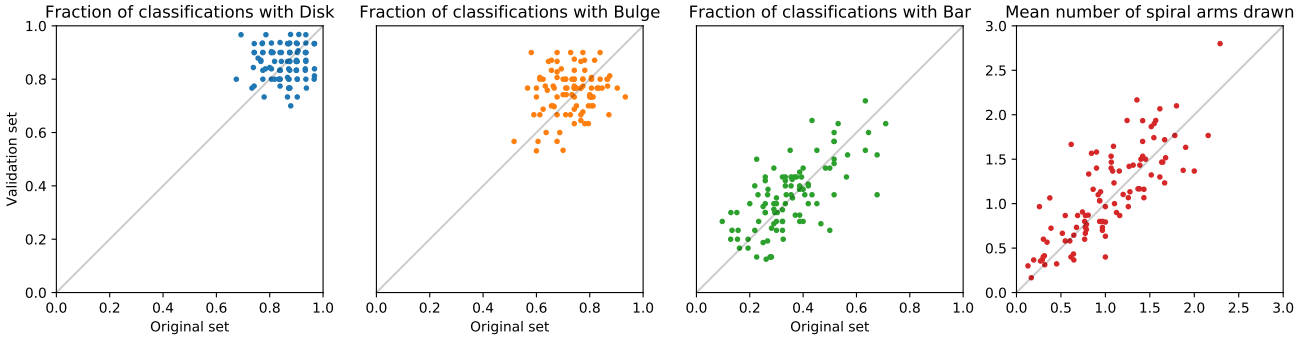

Figure 8. Comparison of frequency of use of component in volunteer models between the original and validation sets of classifications.

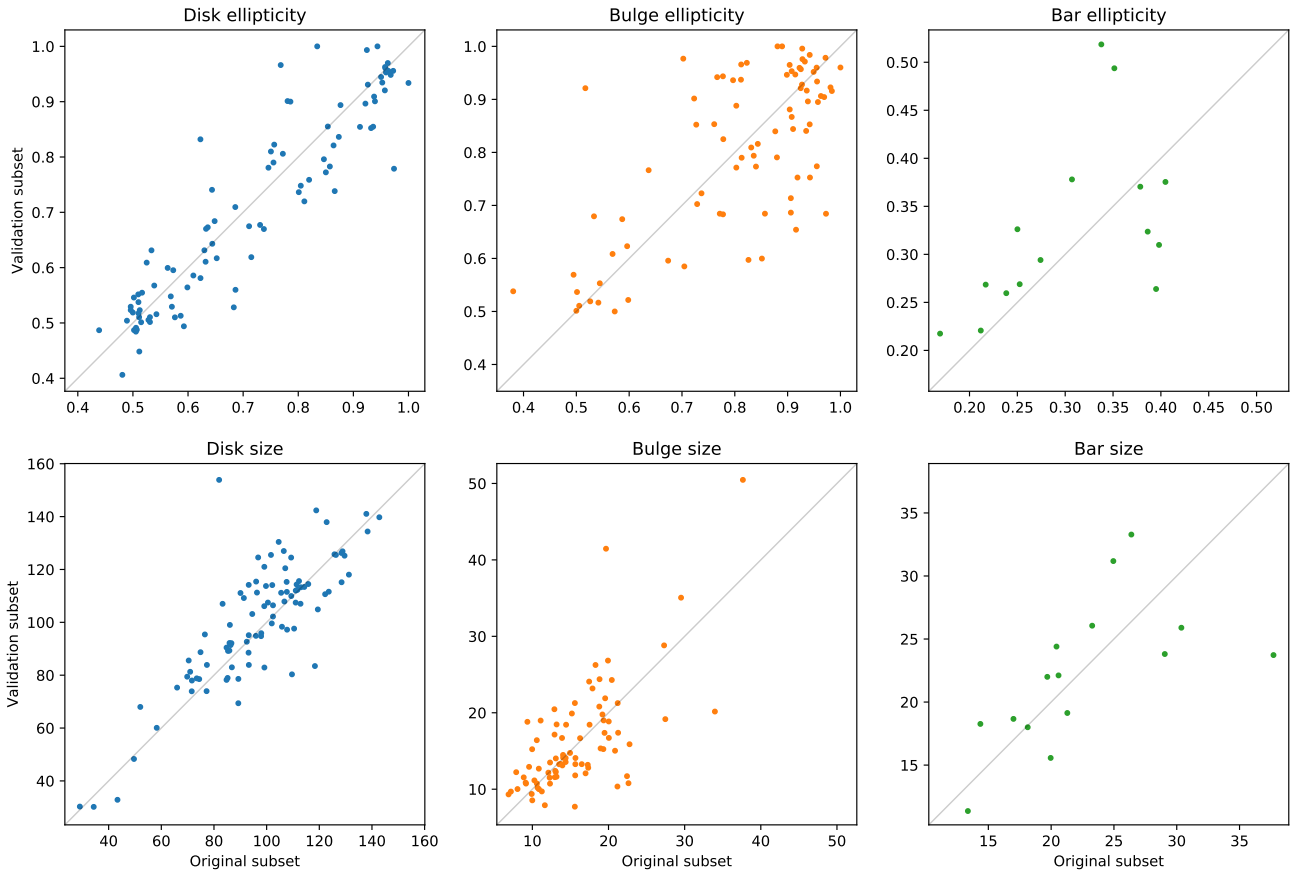

Figure 9. Comparison of component shape in aggregate models between the original and validation sets.

Earth Science Technology Office, Computation Technologies 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
Cappellari M., et al., 2011, MNRAS, 413, 813
Davis D. R., Hayes W. B., 2014, ApJ, 790, 87
Dobbs C., Baba J., 2014, Publ. Astron. Soc. Aust., 31

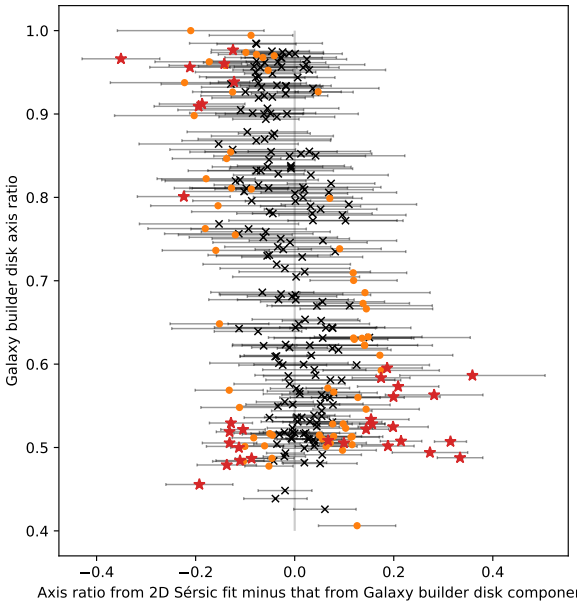

Figure 10. Difference between the axis ratios of the disc components of aggregated Galaxy Builder models to the results of an r-band Sérsic profile fit. Points outside 1- and 2σ are highlighted in orange and red.

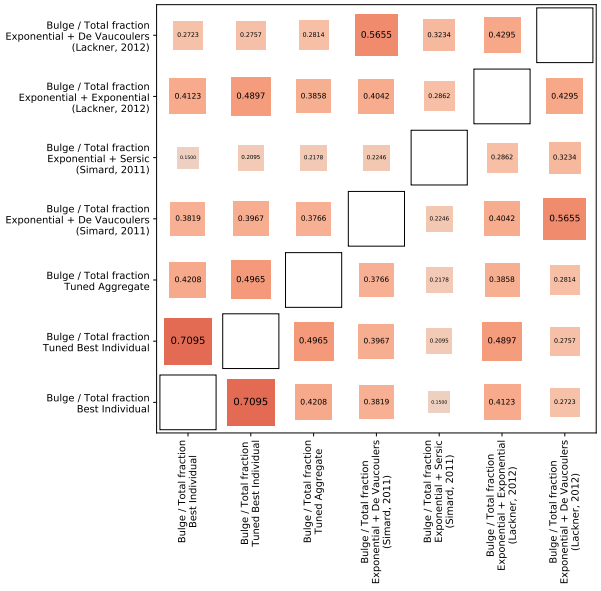

Figure 11. Correlation matrix showing Kendall rank correlation coefficient between measures of Bulge to Total fraction from Galaxy Builder results and other models fitted in Simard et al. (2011) and Lackner & Gunn (2012). Colours and box size indicate the strength of the correlation.

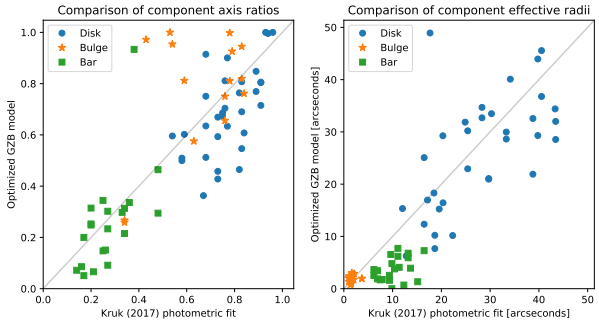

Figure 12. Comparison between optimized Galaxy Builder models and the result of 3component, multiwavelength fits performed by Kruk et al. (2017b).

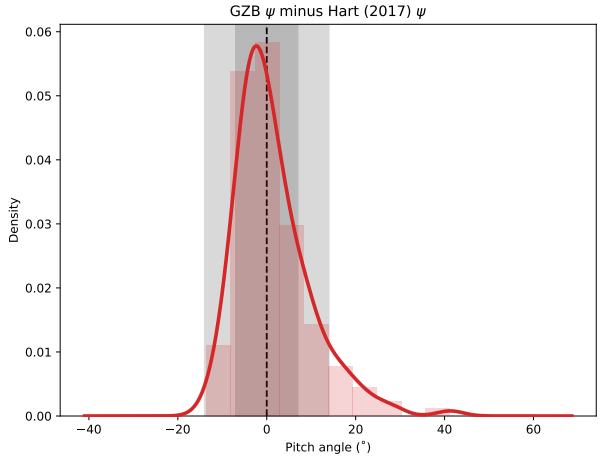

Figure 13. A comparison of Pitch angle obtained by Hart et al. (2016) with measured pitch angles for the aggregated model results in galaxies in the Galaxy Zoo Builder sample. The grey regions show 1- and 2σ errors from Hart et al. (2016). Errors on Galaxy Builder-measured pitch angles are not accounted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

Fall S. M., Romanowsky A. J., 2018
Gadotti D. A., 2010, Monthly Notices of the Royal Astronomical Society, 415, 3308
Gao H., Ho L. C., 2017, The Astrophysical Journal, 845, 114
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
Hubble E. P., 1936, Realm of the Nebulae
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/
Kalinova V., et al., 2017, MNRAS, 469, 2539
Kormendy J., Drory N., Bender R., Cornell M. E., 2010, The Astrophysical Journal, 723, 54
Kruk S. J., et al., 2017a, MNRAS, 469, 3363
Kruk S. J., et al., 2017b, Monthly Notices of the Royal Astronomical Society, 473, 4731
Kruk S. J., et al., 2018, MNRAS, 473, 4731
Lackner C. N., Gunn J. E., 2012, MNRAS, 421, 2277
Lauberts A., Valentijn E. A., 1989, The Messenger, 56, 31
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
Masters K. L., et al., 2019
Mendez-Abreu J., et al., 2016
Naim A., et al., 1995, Monthly Notices of the Royal Astronomical Society

Nair P. B., Abraham R. G., 2010, VizieR Online Data Catalog, p. J/ApJS/186/427
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
Roberts M. S., Haynes M. P., 1994, ARA&A, 32, 115
Robotham A. S. G., Taranu D. S., Tobar R., Moffett A., Driver S. P., 2016
Sanchez S. F., et al., 2011, Astronomy and Astrophysics, 538, A8
Sandage A., 1961, The Hubble Atlas of Galaxies
Scarlata C., et al., 2007, The Astrophysical Journal Supplement Series, 172, 494
Schawinski K., et al., 2007, ApJS, 173, 512
Simard L., et al., 2002a, The Astrophysical Journal Supplement Series, 142, 1
Simard L., et al., 2002b, ASTROPHYS J SUPPL S, 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
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
de Vaucouleurs G., de Vaucouleurs A., Corwin Jr. H. G., Buta R. J., Paturel G., Fouqué P., 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.

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.

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

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

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

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

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

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

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