MNRAS **000**, 1–12 (2018) Preprint 19th March 2019 Compiled using MNRAS LATEX style file v3.0

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

Timothy K. Lingard,[1]★ K. L. Masters,[12] C. M. Krawczyk,[1] Everybody Else,[?]
[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]*Department, Institution, Street Address, City Postal Code, Country*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT
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. This paper presents a novel interface, built inside the Zooniverse citizen science platform, which enables users to help to create detailed photometric models of galaxies from r-band SDSS images. We examine the consistency of this method to changes in number of and population of citizen scientists and compare it to more traditional automated fitting pipelines. 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

One of the cornerstones of all areas of empirical science over the past few centuries has been the sifting-through and sorting into types of objects of interest; 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. 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 advantage of this switch in methodology is the removal of subjective opinion from a catalog of classification: 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). It also enables the selection of morphology-based sample sets from which physical processes and characteristics can be probed.

Until the late 20$^{\text{th}}$ 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 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-cost 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 col-

★ E-mail: tim.lingard@port.ac.uk

our 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.

1.1.1 *Citizen Science*

A promising solution to 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 data improvement methodologies (i.e. Hart et al. 2016), has produced a large catalogue of detailed morphological classifications which are in good agreement with other studies (Willett et al. 2013, Simmons et al. 2014, Willett et al. 2016).

The questions presented to Galaxy Zoo volunteers have evolved over time. Originally asking if a galaxy was one of: smooth, a "Z"-wise spiral, an "S"-wise spiral, an edge-on spiral or a star or unknown or a merger event; the decision tree has now expanded to a more complex array of possibilities to better encompass the variety of galaxy morphologies (including bars, spirals and other, rarer morphological features).

However, while it is possible to draw quantitative measures from these classifications using vote fractions, they are predominantly descriptors of the visible structure present in the galaxy and don't provide a clear quantitative measurement of the relative size of different components, or allow the light from each to be isolated for further analysis.

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 eaxample, the amount of 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 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, which is often seen as a record of its host physical processes and subsequent evolution. [[**More on Sandor's work here**]]

A large part of the problem of performing these detailed decompositions is the tendency for fitting functions to wander away from physical results in the chase for the best possible residual. An example of this would be a Sérsic bulge swapping places with an exponential disc component, as the extra degree of freedom of the Sérsic profile will result in 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.

A third 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, Vika et al. (2014) eye-balled the resulting model and residual images for all 163 of their parametric fits to ensure physical results.

The end result of most of these problems is that researchers will have to eyeball many of their fits to ensure they have converged on a physical model and we are faced with a similar problem to that discussed in subsection 1.1. This paper proposes an analogous solution to that introduced by Lintott et al. (2008), inside the ecosystem that their research set in motion, where we leverage citizen scientists to pick model components and perform rough model fitting in a web-browser environment.

We describe our method in Section.2, including details of the images and ancilliary data from SDSS. 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 is a citizen-science project built on the Zooniverse[1] web platform. It asks volunteers to perform detailed photometric modelling of spiral galaxies (including the bulge, disk, bar and spiral arm components). As a project of this kind, allowing users to interact with and model data, had never been attempted inside the 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.

The closest relative to this project within the Galaxy Zoo 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. 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 developed over one and a half-years, and was built inside the Zooniverse's Panoptes-Front-End[2] codebase, using Facebook's React.js[3] framework, as well as WebGL[4] 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 users 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 presents a volunteer with three images, which they can switch between at any time: an *r*-band cutout of a spiral galaxy (for details of the images used 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 in Fig.1.

The interface prompts volunteers to work through the step-by-step creation of a photometric model of a galaxy (described in detail in Section 2.3). At each step volunteers are asked to first draw a simple isophote (ellipse for disk and bulge, rectangle for bar and polygon-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).

The workflow is designed so that volunteers slowly subtract increasing amounts of light from the galaxy, as can be seen in Fig.2. A tutorial is present which contains a step-by-step guide to completing a classification.

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

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

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 calculate from the residual image shown to voluteers (which was Arcsinh-scaled in a manner described by Lupton et al. 2004), however it is overly sensitive to small deviations of the model from the galaxy. Based on experience running the project, we don't think results are highly sensitive to what "score" was shown but would recommend a simpler error measure to future similar projects.

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[5]). 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. We do experience problems with some computers and browsers not being able to render models, which we detect and prompt volunteers to use a compatible browser. Our choice also has the impact of limiting precision of the model 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.

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

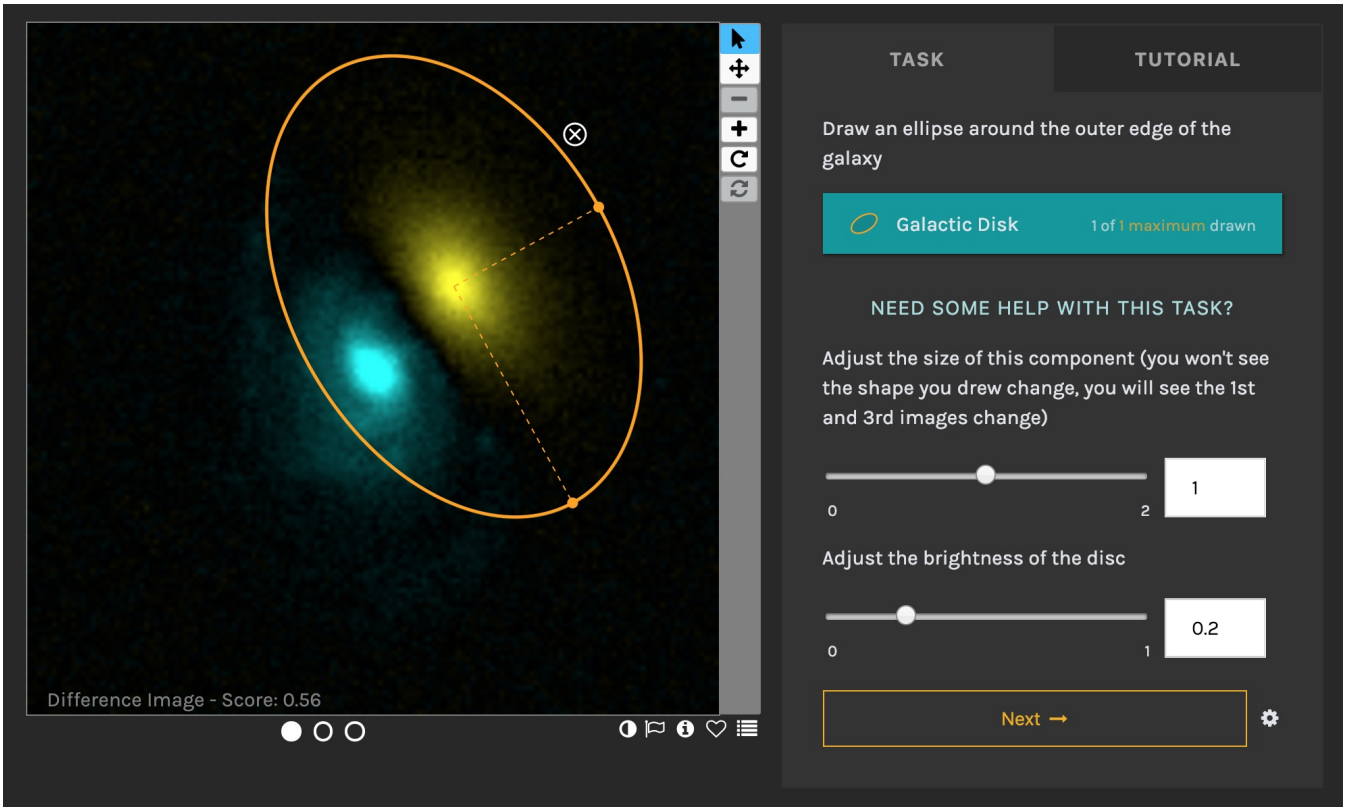

Figure 1. The galaxy builder interface. The residual image is being shown, and the volunteer is on the "Disc" task. The drawn disk component (yellow) is offset from the galaxy image (blue) to demonstrate the positive and negative residuals. Where the image equals the model the residual is black.

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). 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 $14 < m_r < 17$
- Computation of stellar mass completeness limits using the method of Pozzetti et al. (2009) and limits calculated by Hart et al. (2017)
  - Exclusion of galaxies outside these limits
    $\log(M_*/M_\odot) > 2.07\log(z) + 12.64$
    $\log(M_*/M_\odot) < 2.45\log(z) + 14.05.$
- A volume correction, excluding galaxies outside $9.45 < \log(M_*/M_\odot) \leq 11.05$

This selection criteria results in the *stellar mass-complete sample* of 6222 spiral galaxies. We do not obtain volunteer classifications for all of these, but a sub-selection of 198 (the selection of these is detailed in section 2.4).

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 data release was created. To combine multiple FITS images, we made use of the MONTAGE[6] software package.[7]

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

[7] Two out of the two-hundred selected galaxies failed to run

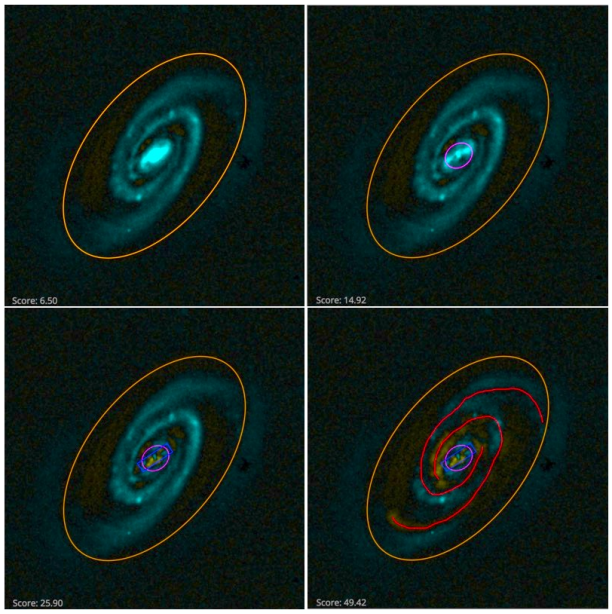

Figure 2. Figure demonstrating the ideal result of each step of the classification process. The top left panel shows the galaxy after only a disk component has been added: the top right contains a disk and a bulge; the bottom left has a disk, bulge and bar; the bottom right is the finished model with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

- 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 it's higher signal-to-noise than other bands.

Once a sub-sample had been created, the Zooniverse's panoptes-python-client[8] (Krawczyk et al. 2019) 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 chosen model components were: An exponential, elliptical disk; an elliptical Sérsic bulge; a Sérsic bar with a "boxiness" modifier on the isophote; spiral arms drawn using freehand lines.

Each spiral arm is modelled using a polygon-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 polygon-line, with users able to choose the Gaussian spread and peak brightness using sliders. Radial falloff was added by multiplying by the value of the previously added exponential disk, though volunteers could change the half-light radius of this falloff disk.

The decision to render spirals in this manner was chosen to accommodate irregular spiral patterns drawn on by volunteers in a simple but acceptably realistic manner. It was decided that the actual spiral profiles were less important to the reduction and analysis process than the positions of the drawn polygon-lines.

The modelling code correctly ignores masked regions identified in the images. It over-samples the bulge, disk and bar components to a factor of five and performs PSF convolution using a PSF obtained in the subject set creation process.

2.4 Choice of Retirement limit

As we were unsure as to the rate at which volunteers would be able to classify images, 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.

After collecting data for two of these sub-samples, preliminary analysis of the classifications indicated that, in order to be able to reliably calculate an aggregate model, we needed more classifications per galaxy than the 10 we had originally envisioned.

A hand-picked sample of 56 galaxies was then reactivated 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 preliminary aggregation methodology development.

Once this hand-picked sample was completed, the remaining galaxies from the initial two sub-samples were reactivated, as well as a repeat of the first sub-sample (hereafter the *validation subset*) to measure volunteer consistency. At the time of writing, only these three batches of galaxies have been run through the project.

2.5 Classification Aggregation Methodology

We will use galaxy UGC 4721 as an example galaxy for the illustration of the data reduction and aggregation methodology. For UGC 4721 we received 32 classifications, containing 28 disks, 24 bulges, 17 bars and 47 drawn spiral arm polylines. These annotations can be seen in Figure 3.

2.5.1 Disk, Bulge and Bar Aggregation

Clustering the disk, bulge and bar components was done purely on the shapes drawn by volunteers, as preliminary

through the image preparation process, due to an error when attempting to montage multiple frames. The root cause of this error is unknown.

[8] https://github.com/zooniverse/aggregation-for-caesar

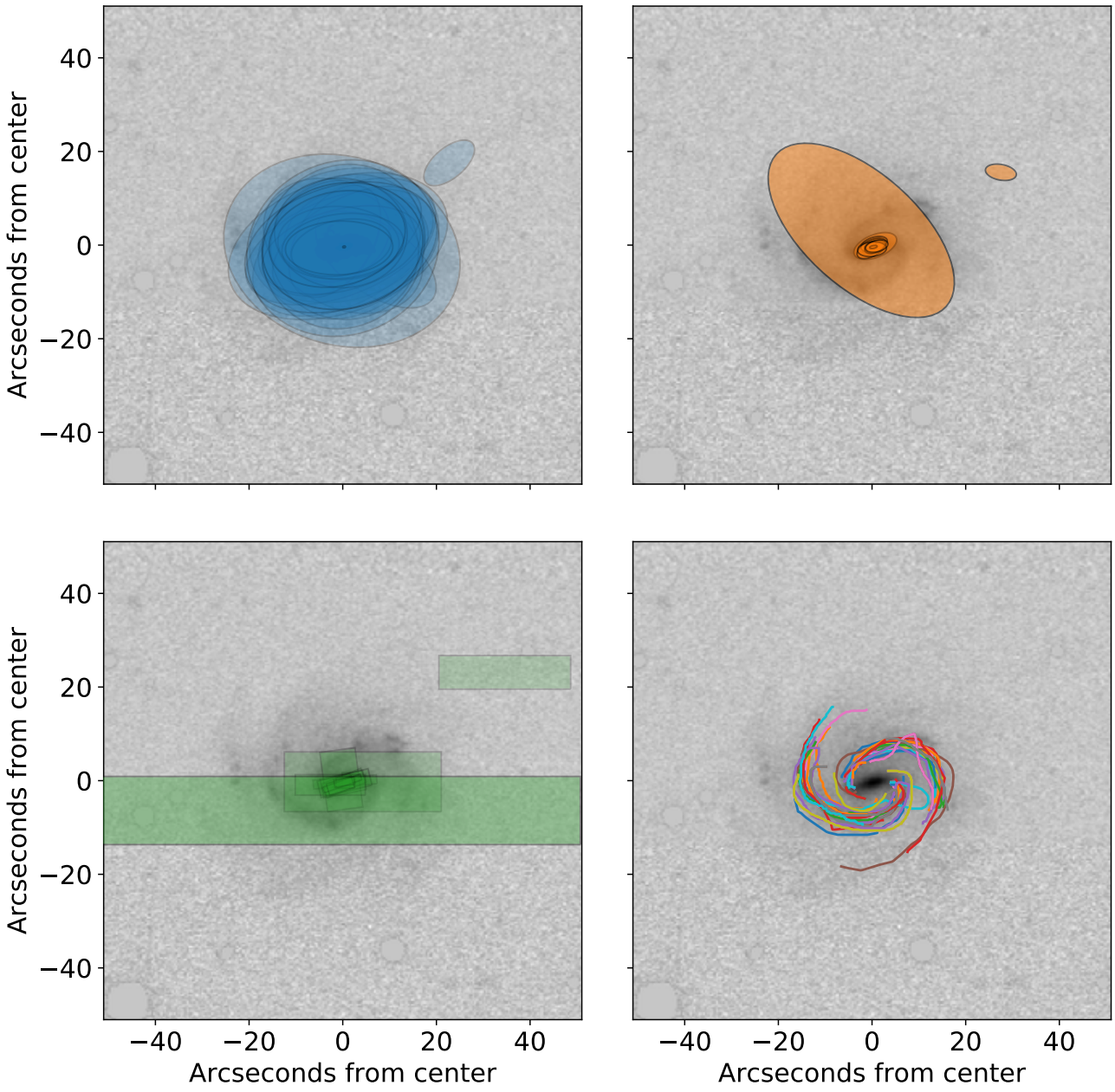

Figure 3. Components drawn by volunteers for UGC 4721. The top left panel shows drawn disks (blue), top right shows drawn bulges (orange), bottom left shows drawn bars (green) and bottom right shows drawn spiral arms.

data exploration suggested that some users simply left model parameters at their default values, or moved them to an extreme of the allowed range. These shapes can be used as the starting point for numerical fits.

Clustering was done 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:

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

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 their mean 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, all drawn disks, bulges, bars and spiral arms can be seen in Fig.3. Resulting mean components can be seen in Fig.4 and the final galaxy model (without spiral arms) can be seen in 5

To cluster drawn spiral arms, we define a custom separ-

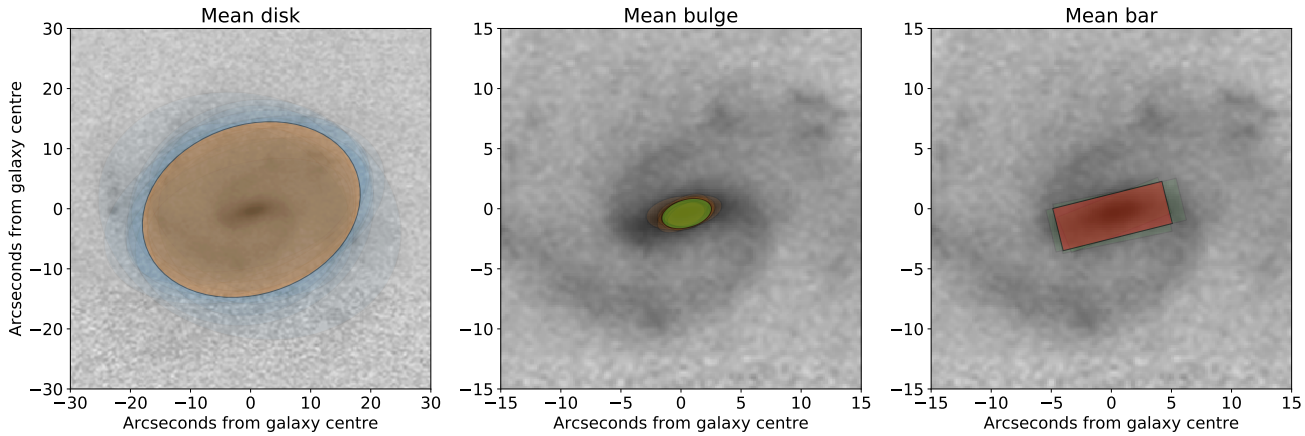

Figure 4. Calculated mean components for UGC 4721. The mean disk is shown in orange in the first panel, the mean bulge in green in the second panel and the mean bar in red in the third panel.

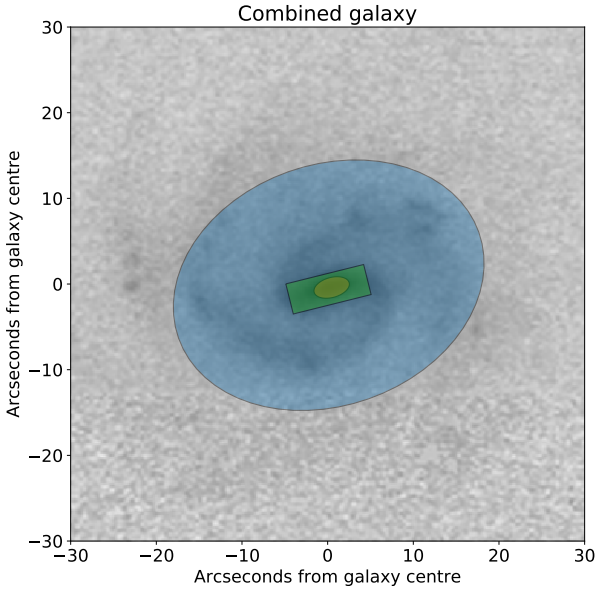

Figure 5. Resulting bulge + disk + bar model for UGC 4721. The disk is shown in blue, the bulge in orange and the bar in green.

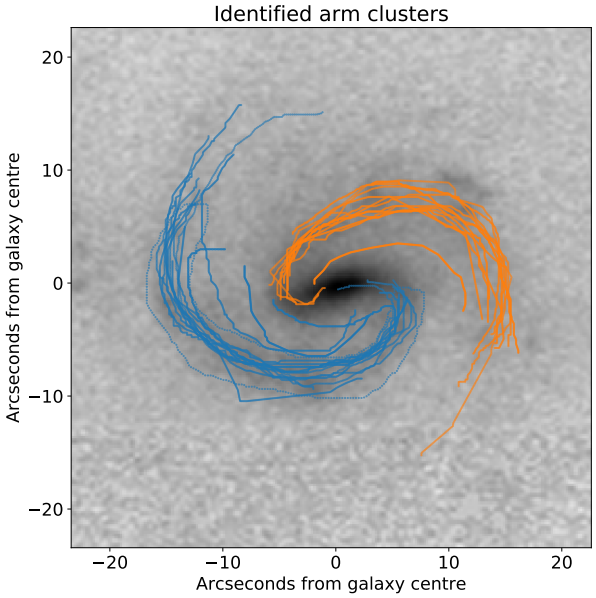

Figure 6. The result of clustering drawn poly-lines for UGC 4721 using our separation measure and the DBSCAN clustering algorithm.

ation measure to represent how far away one poly-line 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 B.

We now 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). The results of this clustering for our example galaxy can be seen in Fig.6.

This clustering method sometimes picks up a small cluster of lines which are a subset of another arm. This is a feature of the separation measure used and is a necessity at this step due to the presence of classifications encompassing both arms with one line.

Once spiral classifications on a galaxy have been clustered into the physical arms they represent, the points are deprojected using the result of a 2D, single-component Sérsic fit in r-band, provided in the NASA Sloan-atlas value-added catalogue. Once 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.

For each arm cluster in each galaxy, a logarithmic spiral model was fitted using Bayesian Ridge Regression, per-

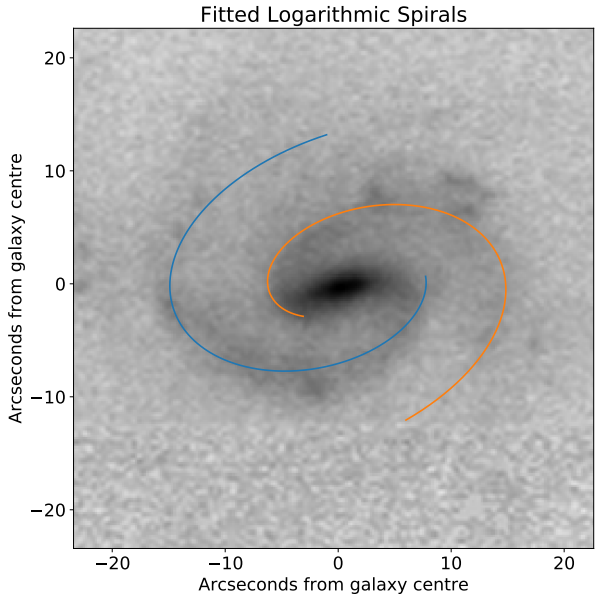

Figure 7. The fitted logarithmic spirals for UGC 4721

formed using the Scikit-learn python package. Hyperpriors on the noise parameter were chosen by fitting a truncated gamma distribution [[**citation needed**]] to the spiral width slider values returned by volunteers (ignoring sliders left at the default or moved to the extremes of allowed values). The logarithmic spiral fit for our example galaxy is shown in Fig.7.

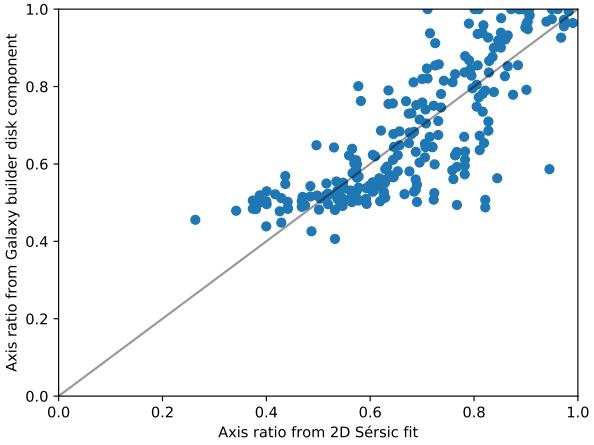

Figure 8. Comparison between the axis ratios of the disk components of aggregated Galaxy Builder models to the results of an r-band Sérsic profile fit.

3 RESULTS

3.1 Examination of volunteer consistency

3.2 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 disks 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 (Fig.8). We note that the results from galaxy builder tend to prefer a slightly more elliptical axis ratio, but plateau at an axis ratio of 0.5. This could be due to the drawing tool ellipse having a default axis ratio of 0.5, and biasing volunteer classification or aggregate shapes.

We can also make comparisons to existing measures of morphology for individual galaxies. Using morphology results from GZ2, we can investigate how likely a volunteer is to incorporate a bulge or bar component in their model relative to its existing morphological classification obtained through citizen science.

Mirroring Kruk et al. (2017b), we define a galaxy as disk dominated if the debiased GZ2 fractions $p_{\text{no bulge}} + p_{\text{just noticeable}} > p_{\text{obvious bulge}} + p_{\text{dominant bulge}}$, or bulge dominated if the converse. Grouping our galaxies based on their type, we find no significant difference in the probability of bulge being present in their model[9]. Inspection of resulting aggregated models from galaxy builder suggests that many volunteers were using a highly elliptical bulge component to include both a galaxy's bulge and its bar, it is possible that enforcing a circular bulge may have provided more consistent classifications and less ambiguity for volunteers.

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

[9] The probability of a classification of a disk dominated galaxy having a bulge was [[0.7293 ± 0.0781]], whereas for a bulge dominated galaxy it was [[0.7707 ± 0.0806]].

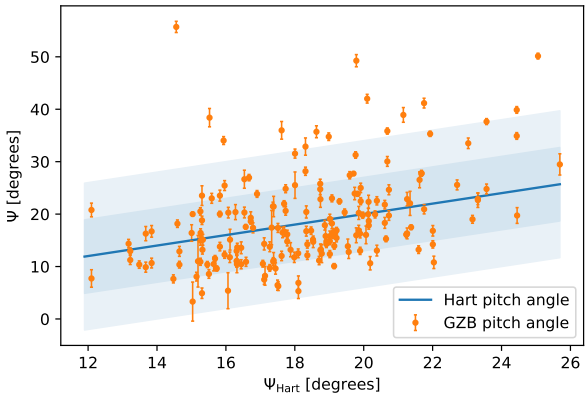

Figure 9. A comparison of Pitch angle obtained by Hart et al. (2016) with measured pitch angles for the aggregated model results in galaxies in the Galaxy Zoo Builder sample. Errors on the Galaxy Builder pitch angles come from dispersion in the arms drawn by volunteers.

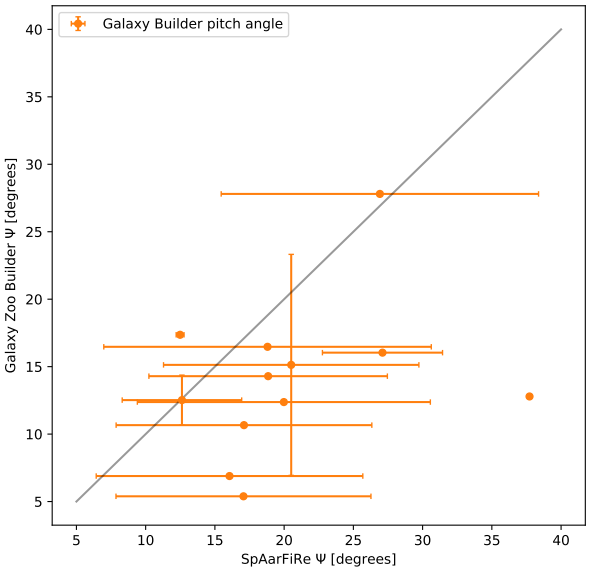

Figure 10. A comparison between pitch angles obtained by Hart et al. (2016) (blue line) to those obtained through Galaxy Builder and SpArcFiRe. In this plot, unlike Fig.9, errors are the sample error of arms used in the averaging and is a measure of inter-arm variation pitch angle.

et al. 2010), we see a significant difference: classifications of strongly-barred galaxies ($p_{\rm bar} > 0.5$) had a [[0.5713 ± 0.1549%]] chance of containing a bar, vs [[0.3666 ± 0.1147%]] 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 [[0.751]].

The strongest consistency check of our results is to make comparisons between galaxy builder models and the result of many-component photometric fits. 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. We compare details of those photometric models to the models obtained in Galaxy Builder in table.3.2. As some of the galaxies received two independent sets of classifications (due to being in the *validation subset*), we list 19 comparisons in the table.

[[**Some text on the results of the comparison needed (plots?)**]]

3.3 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 (Fig.9). Their fit was obtained using a leading automated spiral arm detection and fitting tool, SpArcFiRe (Davis & Hayes 2014). We find good agreement, though the large error bars on the GZ2-produced pitch angle (along with the caveat in the paper that these pitch angles should not be used for individual galaxies) make this comparison rather inconclusive.

We can also directly compare the pitch angles from Galaxy Builder models to the output of the SpArcFiRe software, as shown in Fig.10. The SpArcFiRe pitch angle used is the length-weighted mean of all arms of the dominant chirality of a galaxy.

[[**Show results of sparcfire / GZ2 comparison**]]

3.3.1 Model selection

[[**This is more "Are log spirals right?", should it go in a different paper?**]]

Group K-fold cross validation is used to perform model selection, where points are grouped according to the polyline they were drawn in. Five folds is chosen as a compromise to minimise variance on the resulting scores, while also minimising the bias from removing too many classifications from our training set. To score the goodness-of-fit of a model, we use negative median absolute error.

In order to confidently describe a spiral arm as not a logarithmic spiral, we [[**TODO**]]

Table 1. Table comparing the result of guided photometric three-component fits performed by Kruk et al. (2017b) to the aggregated model obtained from Galaxy Builder. All length units are in arcseconds. [[**Change from using zooniverse subject ids to IAU name from NSA**]] [[**Needs updating**]]

Galaxy ID	Model Type	Fit components	$r_{\mathrm{eff,disk}}$	$(b/a)_{\mathrm{disk}}$	$r_{\mathrm{eff,bar}}$	$(b/a)_{\mathrm{bar}}$
20901993	Galaxy Builder	disk+bar+bulge	15.78	0.68	7.15	0.41
	Photometric	disc+bar	25.40	0.68	11.35	0.17
20902000	Galaxy Builder	disk+bulge	21.63	0.55	NaN	NaN
	Photometric	disc+bar+bulge	29.71	0.58	16.21	0.21
20902013	Galaxy Builder	disk+bar+bulge	15.58	0.83	4.67	0.39
	Photometric	disc+bar	28.33	0.89	7.45	0.14
20902032	Galaxy Builder	disk+bulge	23.29	0.57	NaN	NaN
	Photometric	disc+bar	36.32	0.63	11.53	0.44
20902040	Galaxy Builder	disk+bar+bulge	22.73	0.65	9.05	0.41
	Photometric	disc+bar	38.77	0.73	6.91	0.20
20902049	Galaxy Builder	disk+bulge	8.24	0.51	NaN	NaN
	Photometric	disc+bar+bulge	18.66	0.73	9.46	0.35
20902053	Galaxy Builder	disk+bar+bulge	9.82	0.67	6.83	0.27
	Photometric	disc+bar+bulge	16.47	0.77	9.34	0.27
20902057	Galaxy Builder	disk+bulge	10.78	0.85	NaN	NaN
	Photometric	disc+bar+bulge	20.31	0.96	15.41	0.22
20902058	Galaxy Builder	disk+bulge	14.68	0.89	NaN	NaN
	Photometric	disc+bar+bulge	24.88	0.94	5.62	0.20
20902064	Galaxy Builder	disk+bar+bulge	13.85	0.63	7.93	0.29
	Photometric	disc+bar	17.16	0.84	13.68	0.20
21686510	Galaxy Builder	disk+bulge	18.89	0.52	NaN	NaN
	Photometric	disc+bar+bulge	29.71	0.58	16.21	0.21
21686523	Galaxy Builder	disk+bar+bulge	16.60	0.67	7.90	0.38
	Photometric	disc+bar	33.31	0.76	9.86	0.27
21686542	Galaxy Builder	disk+bar+bulge	15.30	0.87	6.76	0.56
	Photometric	disc+bar	28.33	0.89	7.45	0.14
21686549	Galaxy Builder	bar	NaN	NaN	6.56	0.32
	Photometric	disc+bar	43.44	0.75	11.09	0.48
21686569	Galaxy Builder	disk+bar+bulge	24.52	0.72	9.98	0.40
	Photometric	disc+bar	38.77	0.73	6.91	0.20
21686578	Galaxy Builder	bulge	NaN	NaN	NaN	NaN
	Photometric	disc+bar+bulge	18.66	0.73	9.46	0.35
21686582	Galaxy Builder	disk+bar+bulge	10.17	0.70	6.61	0.26
	Photometric	disc+bar+bulge	16.47	0.77	9.34	0.27
21686587	Galaxy Builder	disk+bar+bulge	15.62	0.89	6.08	0.35
	Photometric	disc+bar+bulge	24.88	0.94	5.62	0.20
21686593	Galaxy Builder	disk+bar+bulge	13.47	0.62	9.25	0.23
	Photometric	disc+bar	17.16	0.84	13.68	0.20

4 CONCLUSIONS

Galaxy Zoo and citizen science has its strengths, but 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 this.

References

Abazajian K. N., et al., 2009, ApJS, 182, 543

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

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

Fall S. M., Romanowsky A. J., 2018

Gadotti D. A., 2010, Monthly Notices of the Royal Astronomical

Society, 415, 3308
Hart R. E., et al., 2016
Hart R. E., et al., 2017
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
Kalinova V., et al., 2017, MNRAS, 469, 2539
Kormendy J., Drory N., Bender R., Cornell M. E., 2010, The Astrophysical Journal, 723, 54
Krawczyk C., McMaster A., Allen C., 2019, zooniverse/aggregation-for-caesar: Add a citable DOI to the repository, doi:10.5281/zenodo.2562861, https://doi.org/10.5281/zenodo.2562861
Kruk S. J., et al., 2017a
Kruk S. J., et al., 2017b, Monthly Notices of the Royal Astronomical Society, 473, 4731
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
Mendez-Abreu J., et al., 2016
Naim A., et al., 1995, Monthly Notices of the Royal Astronomical Society
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
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
Simard L., et al., 2002a, The Astrophysical Journal Supplement Series, 142, 1
Simard L., et al., 2002b, ASTROPHYS J SUPPL S, 142, 1
Simmons B. D., et al., 2014, Monthly Notices of the Royal Astronomical Society, 445, 3466
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., 2016, Monthly Notices of the Royal Astronomical Society, 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.

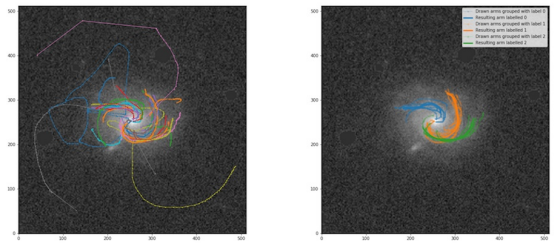

Figure C1. An example of the resulting spiral classifications for a galaxy in the Galaxy Builder sample

APPENDIX A: TECHNICAL OVERVIEW OF THE SOFTWARE

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

First, define a poly-line containing n pairs of x- and y-values (vertices) as

$$A : \{i \in \mathbb{N};\ i < n\} \longrightarrow \mathbb{R}^2 \tag{B1}$$

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 $\vec{w}$):

$$t(\vec{p},\ \vec{v},\ \vec{w}) \equiv \frac{(\vec{p} - \vec{v}) \cdot (\vec{v} - \vec{w})}{|\vec{w} - \vec{v}|^2}. \tag{B2}$$

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

$$d(\vec{p},\ \vec{v},\ \vec{w}) = \| (\vec{v} + \min(\max(t(\vec{p},\ \vec{v},\ \vec{w}),\ 0),\ 1)\ (\vec{w} - \vec{v})) - \vec{p}\| \tag{B3}$$

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

$$D(A,\ B) \equiv \sum_{i=0}^{n} \min\{j \in \mathbb{N}_0,\ j < m;\ d(A_i,\ B_j,\ B_{j+1})^2\}. \tag{B4}$$

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

$$distance(A,\ B) \equiv D(A,\ B) + D(B,\ A). \tag{B5}$$

[[Some figures showing how this distance is calculed would be useful]]

APPENDIX C: EXAMPLE RESULTING SPLINES FROM GALAXY BUILDER

APPENDIX D: MODEL RENDERING PROFILE

Standard Sèrsic [[**citation**]]:

$$I(R) = I_e \exp\left(-b_n \left[\left(\frac{R}{R_e}\right)^{1/n} - 1\right]\right) \tag{D1}$$

As implemented in Galaxy Builder (for a galaxy with axis ratio E, drawn isophotal major axis r_e, boxyness c and Sèrsic index n):

$$I(R) = \frac{I_0}{2} \exp\left(-b_n \left[D^{1/n} - 1\right]\right), \tag{D2}$$

where

$$\begin{aligned} D &= 3\sqrt{\left(\frac{E}{r_e}\right)^c |x' - \mu_x|^c + \left(\frac{1}{r_e}\right)^c |y' - \mu_y|^c} \\ &= \frac{3}{r_e^{c/2}} \sqrt{E^c |x' - \mu_x|^c + |y' - \mu_y|^c} \end{aligned} \tag{D3}$$

and

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \vec{\mu} - \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \vec{\mu}. \tag{D4}$$

Therefore the effective radius is given by

$$R_e = \frac{r_e^{c/2}}{3} \tag{D5}$$

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