The Astrophysical Journal, 900:178 (18pp), 2020 September 10 https://doi.org/10.3847/1538-4357/ab9d83

Galaxy Zoo Builder: Four-component Photometric Decomposition of Spiral Galaxies Guided by Citizen Science

Timothy K. Lingard[1], Karen L. Masters[2], Coleman Krawczyk[1], Chris Lintott[3], Sandor Kruk[4], Brooke Simmons[5], Robert Simpson[6], Steven Bamford[7], Robert C. Nichol[1], and Elisabeth Baeten[8]
[1] Institute of Cosmology and Gravitation, University of Portsmouth Dennis Sciama Building, Burnaby Road, Portsmouth, PO1 3FX, UK; tklingard@gmail.com
[2] Haverford College, 370 Lancaster Avenue, Haverford, PA 19041, USA
[3] Oxford Astrophysics, Denys Wilkinson Building, Keble Road, Oxford, OX1 3RH, UK
[4] European Space Agency, ESTEC, Keplerlaan 1, NL-2201 AZ, Noordwijk, The Netherlands
[5] Physics Department, Lancaster University, Lancaster, LA1 4YB, UK
[6] Google UK, Six, Pancras Square, London N1C 4AG, UK
[7] Centre for Astronomy & Particle Theory, School of Physics & Astronomy University of Nottingham, Nottingham, NG7 2RD, UK
[8] Independent Zooniverse Volunteer, UK
Received 2020 May 11; revised 2020 June 15; accepted 2020 June 15; published 2020 September 14

Abstract

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

Unified Astronomy Thesaurus concepts: Galaxy classification systems (582); Spiral galaxies (1560); Disk galaxies (391); Barred spiral galaxies (136)

1. Introduction

Disk galaxies are complex objects containing many different components, including a disk, disk phenomena (i.e., spiral arms, bars, and rings), and central structures (e.g., bulges; bars; active galactic nuclei, AGN; and inner rings; Carollo et al. 2002). Decomposing disk galaxies into their component structures has become an important tool for extragalactic astronomers seeking to understand the formation and evolution of the galaxy population, ranging from analyzing bulge and bar structure (Elmegreen & Elmegreen 1985; de Jong 1996; Gadotti 2011; Gao & Ho 2017; Mendez-Abreu et al. 2017; Kruk et al. 2018) to the secular evolution of disk galaxies (Lilly et al. 1998; Barden et al. 2005, Allen et al. 2006) and general galaxy assembly and evolution (Simard et al. 2002; Bamford et al. 2011; Lackner & Gunn 2012; Rampazzo et al. 2019).

These fully quantitative methods allow researchers to obtain structural parameters of galaxy subcomponents, which has been used in a variety of astrophysical and cosmological research. For example, the stellar mass found in disks and bulges places strong constraints on the galaxy merger tree from ΛCDM *N*-body simulations (Parry et al. 2009; Hopkins et al. 2010; Rodrigues et al. 2018); the strength of a galaxy's classical bulge is thought to be tied to the strength of merger events in its past (Springel & Hernquist 2005; Kormendy et al. 2010); some work has found a correlation between bulge mass relative to that of the disk and the mass of a central black hole (Davis et al. 2019; Sahu et al. 2019); and different spiral arm formation theories vary in their predictions of spiral morphology (Dobbs & Baba 2014; Pour-Imani et al. 2016; Hart et al. 2017).

The usefulness of obtaining parametric models of a galaxy has motivated the creation of many image modeling and fitting suites, including GIM2D (Simard et al. 2002), GALFIT (Peng et al. 2002), MEGAMORPH (Bamford et al. 2011), PROFIT (Robotham et al. 2017), and PROFILER (Ciambur 2016), to name a few. Using these tools, researchers have built large catalogs of model fits to galaxies. One of the largest photometric model catalogs is that of Simard et al. (2011), who performed automated 2D, two-component (bulge + disk) decomposition of 1,123,718 galaxies from the Legacy imaging of the Sloan Digital Sky Survey (hereafter SDSS) Data Release 7 (Abazajian et al. 2009).

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

A prominent issue when performing these detailed decompositions is the tendency for fitting functions to converge on

unphysical results when not properly guided or constrained. For instance, in a two-component model containing a Sérsic (Sérsic 1963; Graham & Driver 2005) bulge and an exponential disk, the bulge may grow to encompass the galaxy's disk, as its extra parameter allows for a closer fit (as observed by Graham 2001 and Kruk et al. 2018). It is also the case that often, without near-optimal starting points, detailed model fits will fail to converge at all (Lange et al. 2016).

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

Another problem that 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 rigid algorithmic design (e.g., a "logical filter," Allen et al. 2006), or 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 5282, respectively) to ensure physical results with the correct components present. The end result of most of these problems is that researchers will have to invest time to individually check many of their fits to ensure they have converged on a physical model. In the era of large sky surveys such as the SDSS, the time required to do this becomes unsustainable and introduces concerns over human error if done by only a single individual or by small numbers of individuals.

A demonstrably successful solution to the similar problem of galaxy classification in the era of large surveys was to find a new source of person-power: Lintott et al. (2008) invited large numbers of people to classify SDSS-images of galaxies over the internet in the Galaxy Zoo project. The resulting classifications (a mean of 38 per galaxy) were then weighted and averaged to create a morphological catalog of 893,212 galaxies. This hugely successful project, including its subsequent iterations and expansions (i.e., Willett et al. 2013, 2017; Hart et al. 2016; Simmons et al. 2017), has produced a large catalog of detailed morphological classifications that are in good agreement with other studies, and have been used in a wide variety of studies of the local galaxy population (see Masters & the Galaxy Zoo Team 2020 for a recent review).

In this paper we explore an analogous solution to the solution that Lintott et al. (2008) brought to galaxy classification for the issues faced by galaxy parametric modeling, inside the ecosystem that Galaxy Zoo set in motion (namely *The Zooniverse*[9]). We leverage citizen scientists to pick model components and perform model optimization in an online, web-browser environment.[10] We describe our method in Section 2, including details of the images and ancillary data from SDSS as well as the strategy used to obtain scientifically useful models from volunteer classifications. We provide consistency checks within our infrastructure and to other methods in Section 3 and discuss the efficacy and potential impact of our new method relative to existing methodologies in Section 4.

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

2. Method

This section describes the *Galaxy Builder* project, and the sample of galaxies and synthetic images used in this paper to examine the project's output. The entire process is summarized in flowchart form in Figure 1, with appropriate sections referenced therein.

2.1. *The* Galaxy Builder *Zooniverse Project*

Galaxy Builder is a citizen-science project built on the Zooniverse web platform. It asks volunteers to perform detailed photometric modeling of spiral galaxies (potentially including bulge, disk, bar, and spiral arm components). A project of this kind, allowing volunteers to interact with and model data, has never been attempted inside the current Zooniverse web platform before, so this project involved designing and implementing a model rendering[11] suite inside the existing Zooniverse front-end codebase. As with all citizen-science solutions, we had to not only consider the accuracy of the resulting model, but also user experience and engagement in our design decisions.

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

2.1.1. *Project Timeline and Development*

The *Galaxy Builder* project was built inside the Zooniverse's (Simpson et al. 2014) PANOPTES-FRONT-END[12] codebase, using the REACT.JS[13] framework, as well as WebGL[14] to enable low-latency photometric galaxy model rendering. *Galaxy Builder* entered a Zooniverse beta in 2017 late November, and after some user-experience improvements and code refactoring, the project was launched as an official Zooniverse project on 2018 April 24.

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

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

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

[11] 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.

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

[13] https://reactjs.org/

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

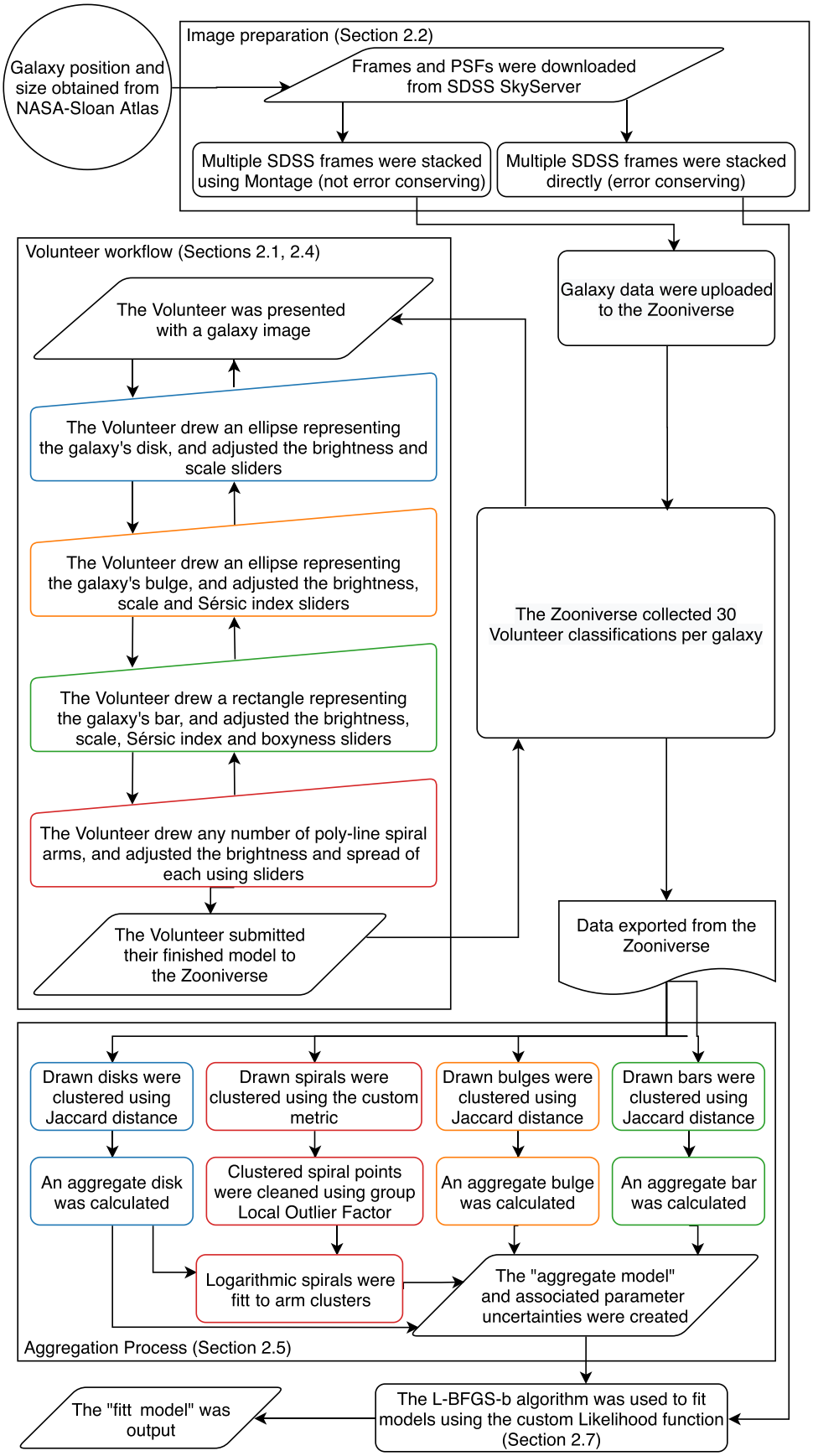

Figure 1. Flowchart detailing the entire *Galaxy Builder* process, from image creation, through classification collection using the Zooniverse, to model aggregation and fitting. Processes, manual input, data inputs and exports, and document exports are displayed distinctly. Colors distinguish between component-specific processes (disk in blue, bulge in orange, bar in green, and spiral in red). Black nodes relate to the galaxy as a whole.

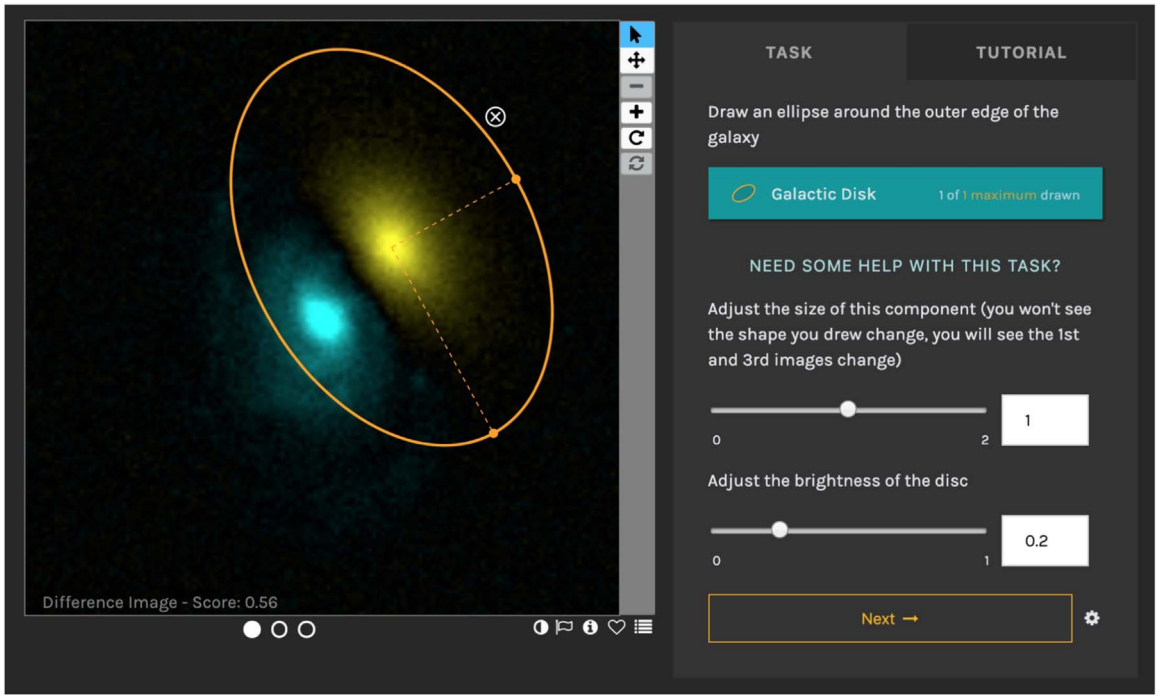

Figure 2. The *Galaxy Builder* interface. The residual image is being shown, and the volunteer is on the "Disk" task. The drawn disk component (yellow) is offset from the galaxy image (blue) to demonstrate the positive and negative residuals. Where the image equals the model, the residual is black. The 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 color 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 favorite 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.

essential to this process as part of the typical beta trial process for Zooniverse projects.[15]

2.1.2. The Project Interface

The *Galaxy Builder* project prompts volunteers to work through the step-by-step creation of a photometric model of a galaxy (described in detail in Section 2.4). A screenshot of the interface can be seen in Figure 2, where a residual image is shown. The interface presents a volunteer with three views, which they can switch between at any time: an *r*-band cutout image of a spiral galaxy (see Section 2.2), the galaxy model they have created so far, and the residual between their model and image (shown in blue and yellow).

The workflow is designed so that volunteers slowly subtract increasing amounts of light from the galaxy, as is illustrated in Figure 3. A tutorial is available that contains a step-by-step guide to completing a classification. At each step, volunteers are asked to first draw a simple isophote, and then make use of a series of sliders to adjust the parameters of the model component (see Section 2.4 for more information).

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

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

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

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

2.2. Sample Selection: Images and Ancillary Data

As a proposed solution to the problem of fitting multi-component and complex galaxies, *Galaxy Builder* finds a niche with a sample of disk galaxies with spiral features. One such sample is the stellar mass-complete sample in Hart et al. (2017), which is a sample of relatively face-on spiral galaxies ($b/a > 0.4$) with and without bars and selected to be complete across stellar masses $9.45 < \log(M_\star/M_\odot) < 11.05$. The test sample we use for the Galaxy Builder project was therefore selected from the Hart et al. (2017) sample of relatively face-on spiral galaxies.

The morphological information required to select spiral galaxies came from the public data release of Galaxy Zoo 2 (Willett et al. 2013, hereafter GZ2). Each response to a GZ2 morphology question is allocated a p value ranging from 0 to 1, where 0 indicates that no volunteers responded positively to that question, and 1 indicates all volunteers who classified that galaxy responded positively (i.e., $p_{\mathrm{bar}} = 0.5$ would indicate 50% of volunteers said a bar was present in a galaxy). Photometric measurements used for selection came from the

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

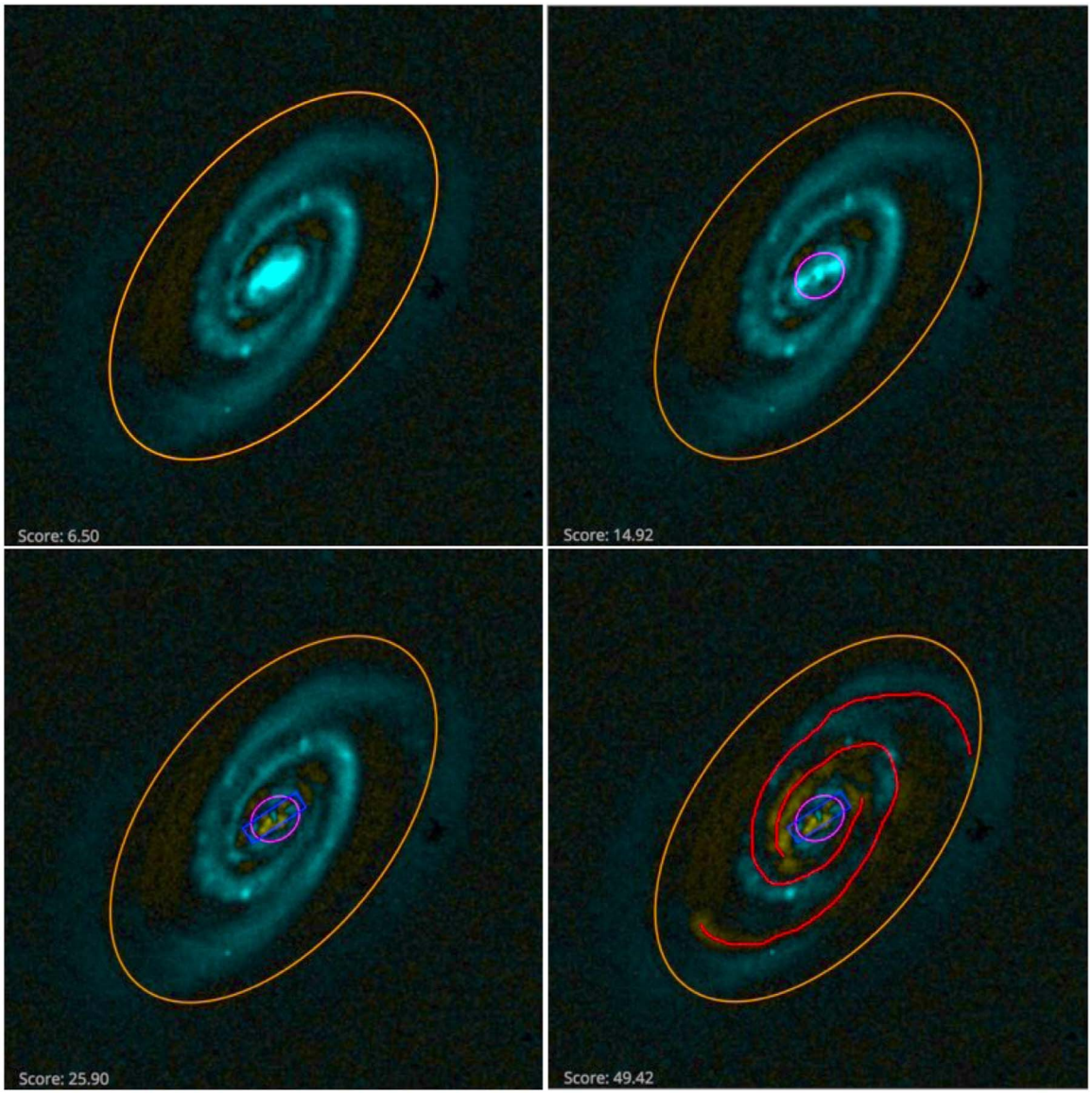

Figure 3. Figure demonstrating the desired result of each step of the modeling process, as seen from the residual image provided to volunteers. The top left panel shows the galaxy after only a disk component has been added: the top right contains a disk and a bulge; the bottom left has a disk, bulge, and bar; and the bottom right is the finished model with a disk, bulge, bar, and spiral arms. The image shown is SDSS J104238.12+235706.8. The brightness and contrast of this image have been edited to improve visibility in print.

Table 1
The Selection Criteria Used in Hart et al. (2017) to Create the *Stellar Mass-complete Sample* of 6222 Spiral Galaxies

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

Notes. Computed using the method of Pozzetti et al. (2009) and limits calculated by Hart et al. (2017).
[a] Stellar masses from Mendel et al. (2014).

NASA-Sloan Atlas (Blanton et al. 2011, hereafter NSA). The stellar mass-complete sample is constructed using the set of criteria detailed in Table 1.

The stellar mass-complete sample was split into smaller subsamples, each containing 100 galaxies. In an iterative process, each subsample was chosen to contain the 60 unclassified galaxies with the lowest redshift, and 40 random unclassified galaxies. This was done to ensure that we would have an early sample to work with given the a priori unknown rate at which volunteers would provide classifications. Due to time constraints, classifications were only collected for two unique subsamples. The mass-redshift relation of galaxies in

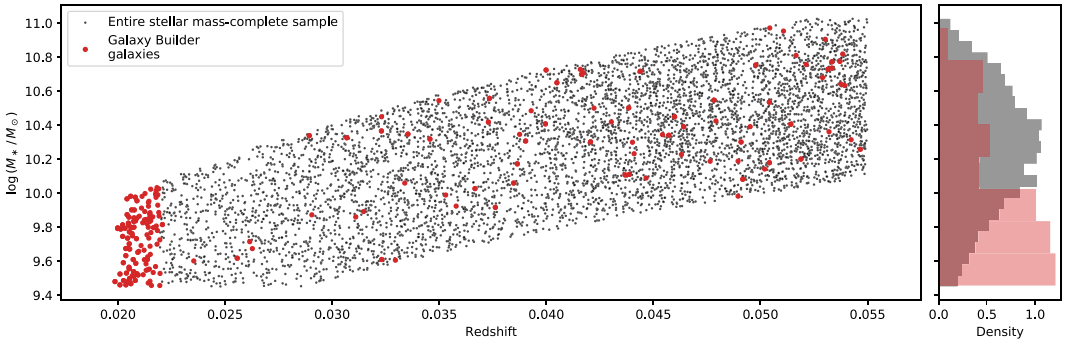

Figure 4. Redshift against total galaxy stellar mass for all galaxies in the stellar mass-complete sample, with the 198 galaxies considered in this paper highlighted in red. The distribution of stellar masses is shown in the right panel for the total sample and for the galaxies considered here. It is evident that the galaxies for which we collected classifications are not complete in stellar mass, but it is possible to select a further subset that would be.

the stellar mass-complete sample from Hart et al. (2017) can be seen in Figure 4, with galaxies present in this work highlighted in red. Stellar masses were calculated by Mendel et al. (2014).

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

2.2.1. Image and Modeling 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:

1. A montage of multiple *r*-band corrected frames from the SDSS DR13 (Albareti et al. 2017) data release was created. To combine multiple FITS images, we made use of Astropy (Astropy Collaboration et al. 2018) and of the MONTAGE (Jacob et al. 2010) software package.
2. This montage was cropped to four times the Petrosian radius of the galaxy.
3. The SEXTRACTOR software (Bertin & Arnouts 1996) was used to identify regions containing secondary sources (foreground statistics, other galaxies) and generate a mask.
4. A point-spread function (PSF) was obtained from the relevant Sloan *r*-band `psField` file, extracted at the central position of the galaxy (Stoughton et al. 2002).
5. The JSON file was written containing the cutout data and the 2D boolean mask obtained from the source extraction process. This file also contained other metadata needed for the rendering process (PSF, the size of the PSF array, and the width and height of the image).
6. Another JSON file containing simply the information used to render the volunteer's model (image size and PSF) was created.
7. An arcsinh-stretch was applied to the masked cutout (as described by Lupton et al. 2004). It was then saved as a gray-scale image.

We decided to use *r*-band images in our subject set because its signal-to-noise ratio was higher than other bands.

Once a subsample had been created, the Zooniverse's PANOPTES-PYTHON-CLIENT[16] was used to upload them as a subject-set to the Zooniverse.

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

2.3. Choice of Retirement Limit

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

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

We also created nine synthetic images of galaxies that contained various combinations of components available to volunteers and a spread of possible parameters. These synthetic galaxies were based on a set of target galaxies from *Galaxy Builder* and were designed to be as realistic as possible, including the addition of realistic noise and pixel masks. This set of synthetic images is shown in Figure 5 and was used to calibrate our aggregation and fitting methodology. It is referred to as the calibration subset.

2.4. The Galaxy Model

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

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

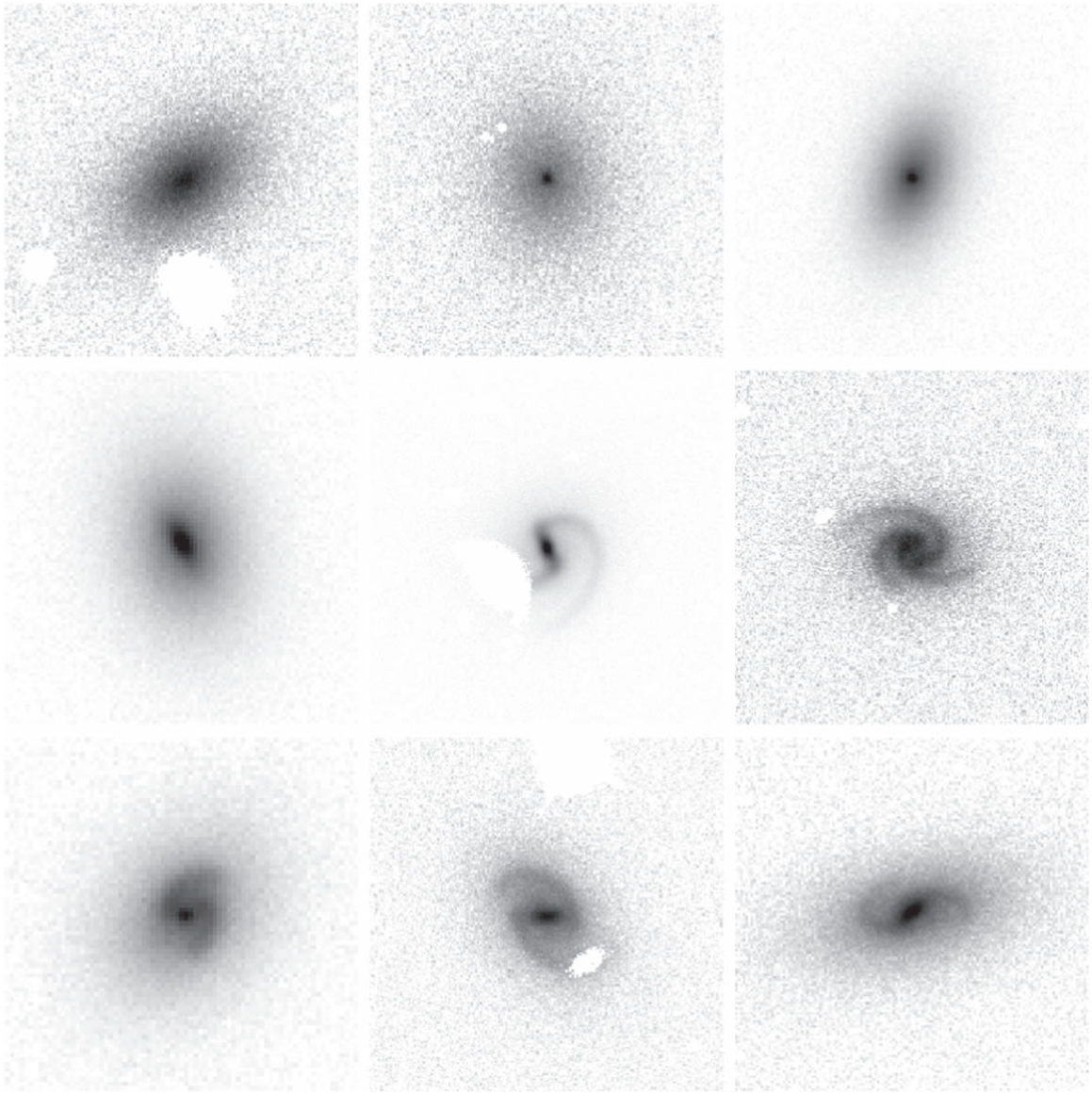

Figure 5. Arcsinh-stretched images of the synthetic galaxies present in the calibration subset. These galaxies were designed to look as realistic as possible, while being described perfectly by the model available to volunteers.

1. One exponential, ellipsoidal disk.
2. One ellipsoidal Sérsic bulge, with n chosen by volunteers.
3. One Sérsic bar with a "boxiness" modifier (as described in Peng et al. 2002), with n and c chosen by volunteers.
4. Any number of freehand polyline[17] spiral arms, as described below.

2.4.1. Spiral Arm Model

Each spiral arm is modeled using a polyline drawn by the volunteer. The brightness of a spiral arm at any point is given by the value of a Gaussian centered at the nearest point on any drawn polyline, 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 disk, although volunteers could change the half-light radius of this falloff disk.

[17] A polyline, or polygonal chain, is a series of connected line segments.

2.5. Classification Aggregation Methodology

In this section, we use the galaxy UGC 4721, a two-armed barred spiral galaxy at $z = 0.02086$ classified by de Vaucouleurs et al. (1991) as SBcd, to illustrate the data reduction and aggregation methodology. For UGC 4721 we received 32 classifications, containing 28 disks, 24 bulges, 17 bars, and 47 drawn spiral arm polylines (4 classifications did not contain spirals, 7 contained one spiral arm, 14 contained two arms, 6 contained three arms, and one contained four arms). These annotations can be seen in Figure 6, overlaid on the gray-scale r-band image of the galaxy.

2.5.1. Aggregation of Volunteer Models

Aggregate model calculation was done on a component-by-component basis, rather than per classification, i.e., clustering of disks was performed independently to that of bulges, bars, and spirals. We did not take into account any slider values, only the shape drawn by the volunteers. Disk classifications were

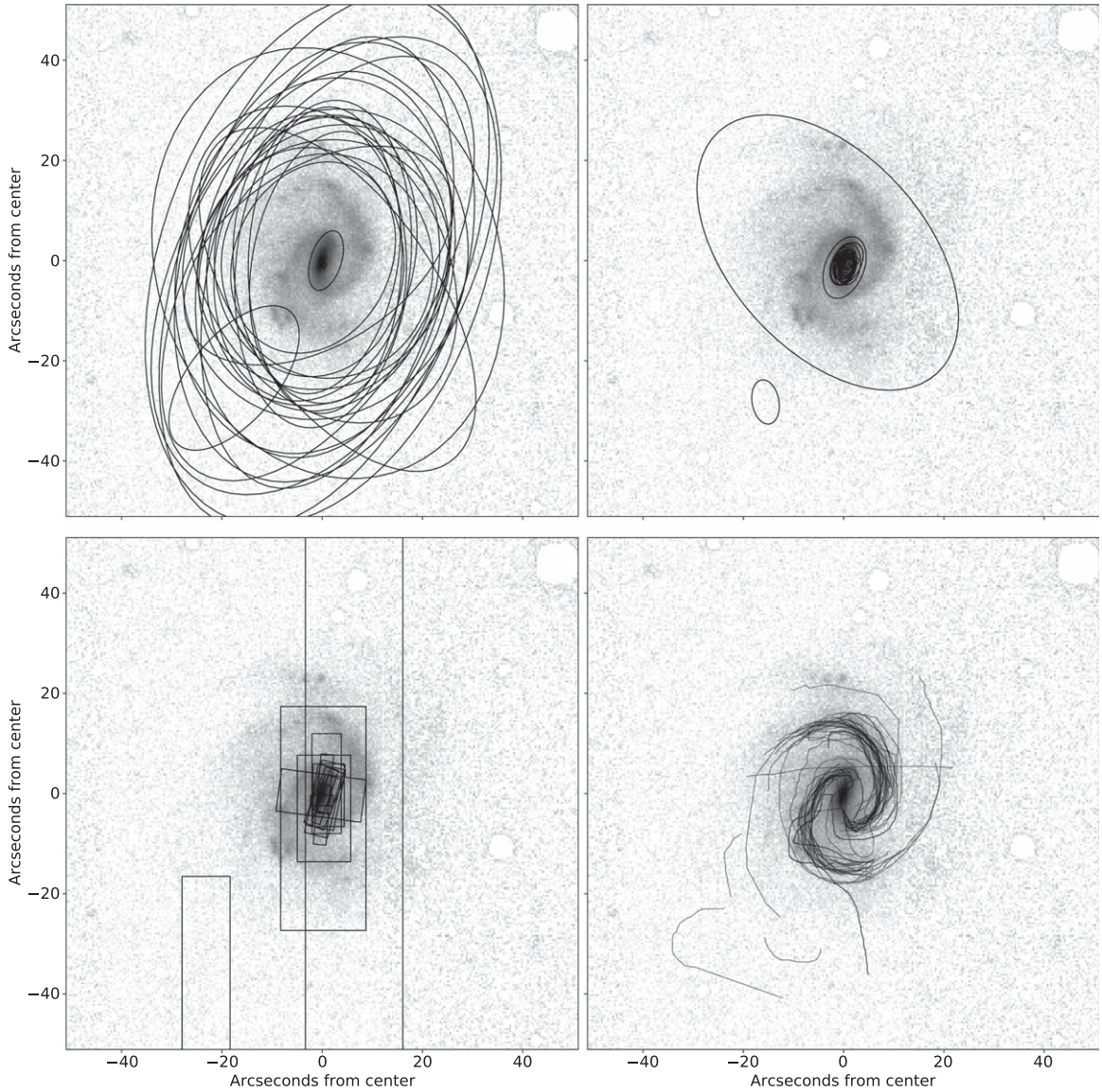

Figure 6. Components drawn by volunteers for UGC 4721. The top left panel shows drawn disks, top right shows drawn bulges, bottom left shows drawn bars, and bottom right shows drawn spiral arms. Disks, bulges, and bars are displayed at twice their effective radii. These raw marks are subsequently aggregated to produce a consensus value for each galaxy component.

doubled in effective radius to correct for a systematic error in disk size observed in the classifications received for the calibration subset. Model parameters were restricted to be within the limits shown in Table 2 (deemed to be the physically acceptable bounds). All components were transformed from the coordinate space of the MONTAGE-created images to the more accurate stacked images created for fitting. Clustering was performed using the Jaccard distance measure (also known as the intersect-over-union distance, or IOU distance), which is a simple metric determining the relative shared area of two sets:

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

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. In DBSCAN, the core of a cluster is defined as a group of at least $N_{\min}$ items that are all within a distance ϵ of each other. Additionally, any points within a distance ϵ of a cluster's core are also associated with the cluster.

2.5.2. Disk, Bulge, and Bar Clustering

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

We define the aggregate component to be the shape that minimizes the sum of Jaccard distances to each of the members of the cluster. For our example galaxy, UGC 4721, clustered and aggregate components can be seen in Figure 7.

Table 2
The Maximum, Minimum, and Default Values for Model Parameters

Component	Parameter	Tuning Minimum Bound	Tuning Maximum Bound
disk	μ_x	−inf	inf
	μ_y	−inf	inf
	ψ	−inf	inf
	q	0.25	1.2
	R_e	0	inf
	I_e	0	inf
bulge	μ_x	−inf	inf
	μ_y	−inf	inf
	ψ	−inf	inf
	q	0.6	1.2
	$R_e/R_{e,\mathrm{disk}}$	0.01	1
	$(B/T)_r$	0	0.99
	n	0.5	5
bar	μ_x	−inf	inf
	μ_y	−inf	inf
	ψ	−inf	inf
	q	0.05	0.5
	$R_e/R_{e,\mathrm{disk}}$	0.05	1
	$(B/T)_r$	0	0.99
	n	0.3	5
	c	1	6
spiral	I_s	0	inf
	A	0	inf
	spread	0	inf
	ϕ	−85	85
	θ_{min}	−inf	inf
	θ_{max}	−inf	inf

Note. Model parameters are defined in the Appendix. Note that some parameters were allowed to overflow when fitting, for instance, an axis ratio greater than 1 (signifying a swap of major and minor axis) was allowed, and corrected for when the fitting reached completion. This helped avoid the optimizer encountering parameter bounds and failing to converge. Component position angle (ψ) and spiral pitch angle (ϕ) were similarly unconstrained.

2.5.3. Spiral Arm Clustering

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

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

Deprojected points within each drawn polyline are converted into polar coordinates and unwound to allow model fitting. These unwound points are then cleaned using the local-outlier-factor algorithm (LOF, Breunig et al. 2000). For each drawn polyline 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 their grouped nature. The points removed as outliers for the example galaxy are shown in the bottom right panel of Figure 7.

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

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

The galaxy model for UGC 4721 obtained through aggregation can be seen in the bottom left panel of Figure 8.

2.6. Error Estimation of Aggregate Models

As all components in a cluster can be viewed as volunteers' attempts at modeling the true underlying component, the sample variance of the parameters of these shapes can be used as a measure of confidence in the parameters present in the aggregate result. These are highly sensitive to clustering hyperparameters, and are only valid for a component's position, size, and shape. Figure 7 illustrates the variance in clustered shapes for our example galaxy (UCG 4721); we see a large variation in the clustered disks, and much closer agreement on the size and shape of the bulge and bar.

2.7. Model Fitting

The final step in creating *Galaxy Builder* models is a numerical fit to fine-tune parameters. This fitting was performed using the L-BFGS-b algorithm (Byrd et al. 1995), implemented in SCIPY (Virtanen et al. 2020). We minimize a custom likelihood function that assumes Gaussian error on pixel values and incorporates the priors on parameters we obtain from clustering. The full fitting model and likelihood is detailed in the Appendix. We use the same model as used by volunteers in the online interface (with altered limits), with spiral arms restricted to being logarithmic spirals relative to the disk, and without the ability to change the relative falloff of spiral arms.

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

We initially fit only for the brightnesses of components, and then simultaneously for all free parameters of all components. The result of the fit, including the final photometric model for UGC 4721, can be seen in Figure 8. The secondary components have been accounted for well, and the model has a sensible reduced chi-squared value of 1.176, where we have approximated degrees of freedom as the number of unmasked pixels present in the galaxy image (similar to GALFIT).

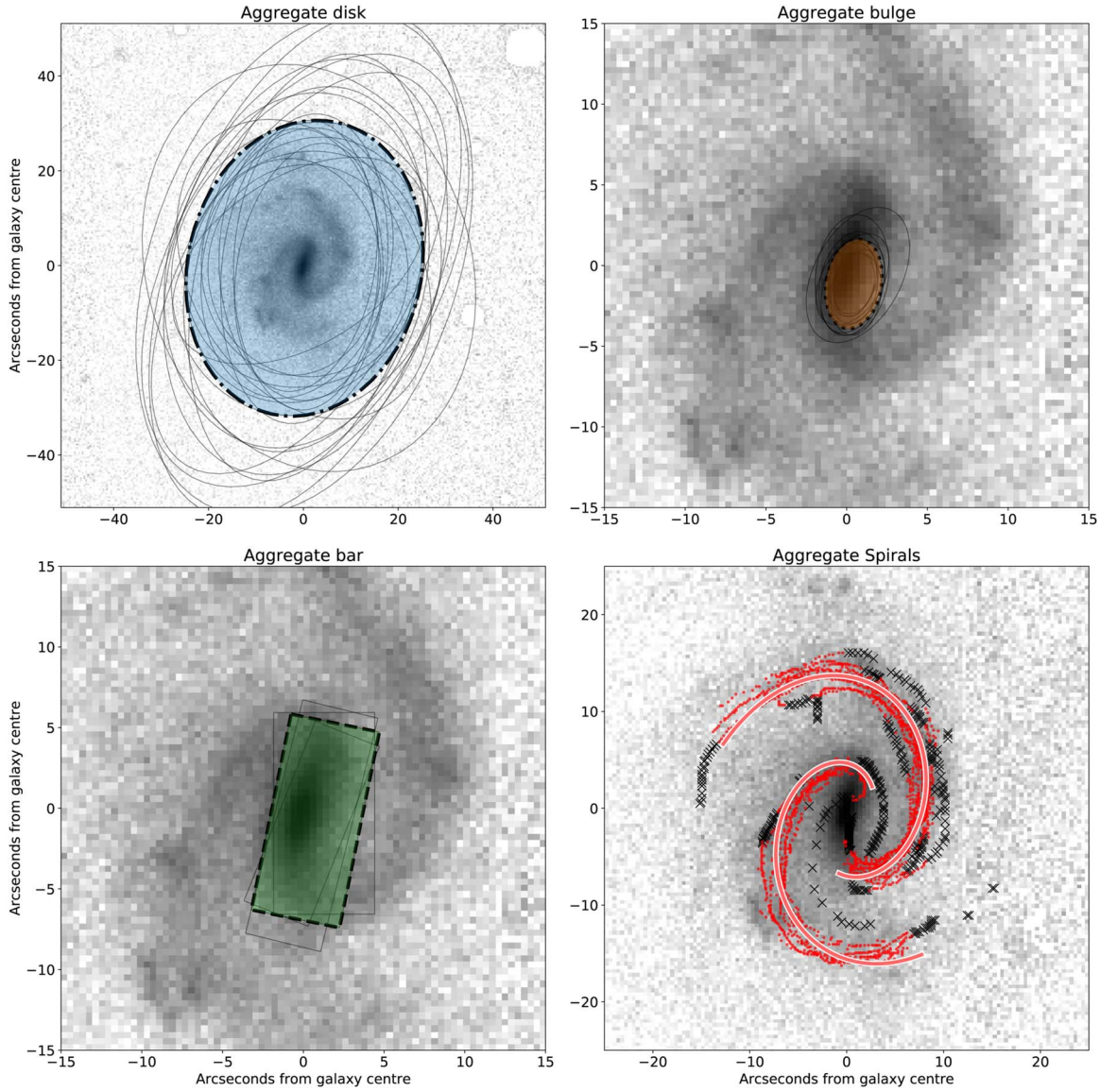

Figure 7. Calculated aggregate components for UGC 4721. The aggregate disk is shown using a dot–dashed line and blue shading in the upper left panel, the aggregate bulge with a dotted line and orange shading in the upper right panel, the aggregate bar using a dashed line and green shading in the lower left panel, and the aggregate spiral arms are plotted as red lines in the lower right panel. Sérsic components are displayed at twice their effective radii. Black crosses in the lower right panel indicate spiral arm points that were identified as outliers and removed during cleaning (described in Section 2.5.3). The aggregated components agree well with the underlying morphology, despite the noisiness of the classifications received.

We use the errors described in Section 2.6 as parameter uncertainties, as we feel an approach based on the local curvature of the likelihood space (as used by `Galfit`) would likely fall foul of the issues described in the introduction and thus be an underestimate. This decision means that we do not have uncertainties for some parameters.

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

3. Results

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

3.1. The Calibration Set

The calibration subset was a set of nine synthetic galaxy images created from *Galaxy Builder* models, which were then rerun through the *Galaxy Builder* process. These galaxies were used to fine-tune clustering and fitting hyperparameters (See Section 2.5.1), as the ground truth was known. Our ability to recover morphology accurately is an essential validation for our ability to recover good photometric models of galaxies.

The scatter between true and measured parameters is shown in Figure 9; these results highlight the importance of good priors to obtain accurate fits of complex photometric models. In more detail, the models recovered for the nine synthetic galaxy images demonstrate that

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

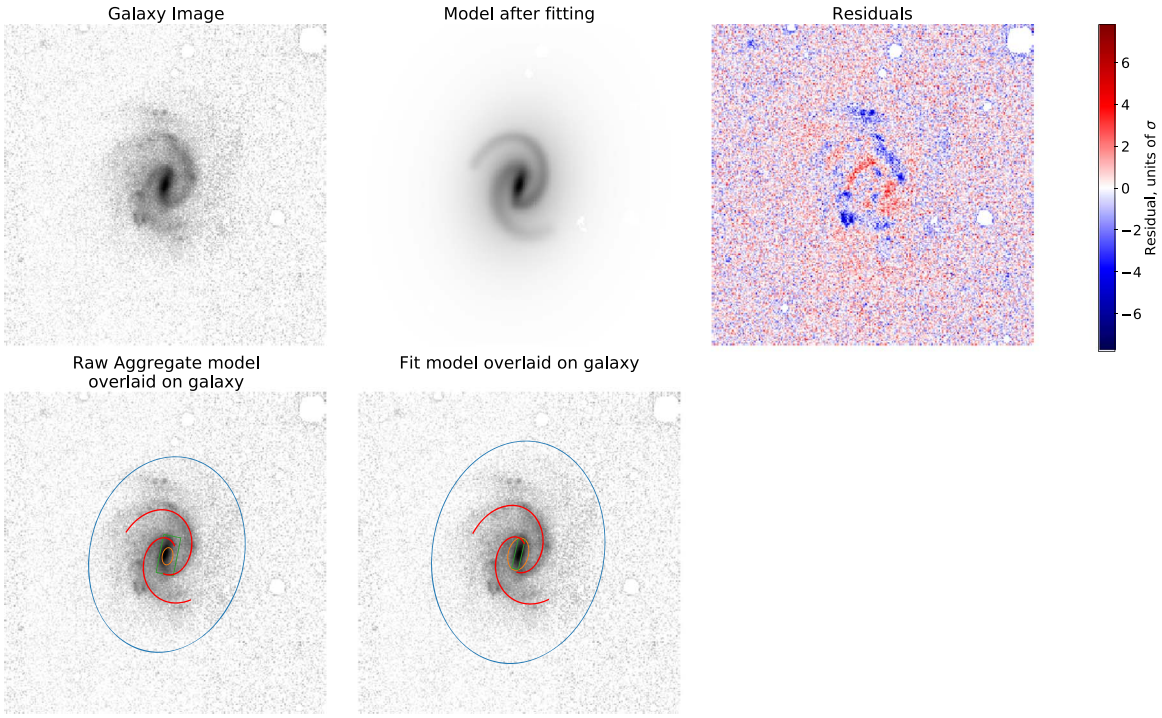

Figure 8. Effect of fitting on the aggregated models. The top left panel shows an Arcsinh-scaled image of the galaxy being fit (UGC 4721), the top middle shows the final model obtained (with the same limits and scaling as the galaxy image), and the top right shows the difference between the two images in units of pixel uncertainty. The bottom panels show a simple representation of the model before and after tuning, overlaid on the galaxy image from the top left panel. With minimal change to the aggregated components, we recover a detailed model that matches the galaxy exceptionally well, as evident in the residuals.

2. We successfully recover all spiral arms present, and do not receive any false positives. The spiral pitch angles obtained through aggregation vary by $<9°$ from the true values, with fitting improving this error slightly.
3. Volunteers systematically use elongated bulges to model bar components. This resulted in two false positives for bulge presence in the aggregate models. This feature (switching light between model components) is a common issue in all photometric fitting methods (Kruk et al. 2018).
4. The Jaccard metric is unstable to small changes in rotation for highly elliptical components (i.e., bars). This resulted in one false negative of bar presence in the aggregate model.

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

3.2. Examination of Volunteer Consistency

We aggregate two independent models for a set of 98 galaxies based on "original" or repeat ("validation") classifications, obtained with the same retirement limit (see Section 2.2 for more on this selection).

One of the simplest choices the volunteers have is whether to include a model component. Figure 10 illustrates the consistency with which volunteers made use of a component in their model for a galaxy. We see that volunteer classification is very consistent, with scatter as predicted by the binomial uncertainty on the mean. Volunteers almost always make us of a disk and bulge (as seen in the calibration subset), and bulge, bar, and spiral arm usage is consistent within the binomial error. One common challenge is that some volunteers used a very ellipsoidal bulge and the ends of spiral arms to model light that other users modeled with a bar. This caused some scatter in aggregate models.

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

After selecting a component, the volunteer sets its shape and size. The variation in axial ratios and effective radii for the aggregate disks, bulges, and bars is shown in Figure 11. The aggregate disks and bulges are consistent within errors, but bars show more scatter. Bars are one of the most challenging components to aggregate consistently. This is partly because even a strongly barred galaxy with 30 classifications overall might receive only 15 or so drawn bars, and lower numbers of classifications result in more scatter. In addition, the aggregation method is more sensitive to rotation of highly elongated shapes. Both factors probably contribute to lower consistency in bar components.

3.3. Comparison to Results in the Literature

After aggregating and fitting models for our galaxies, we examine how our models compare to other results in the literature. Part of the motivation for exploring the *Galaxy*

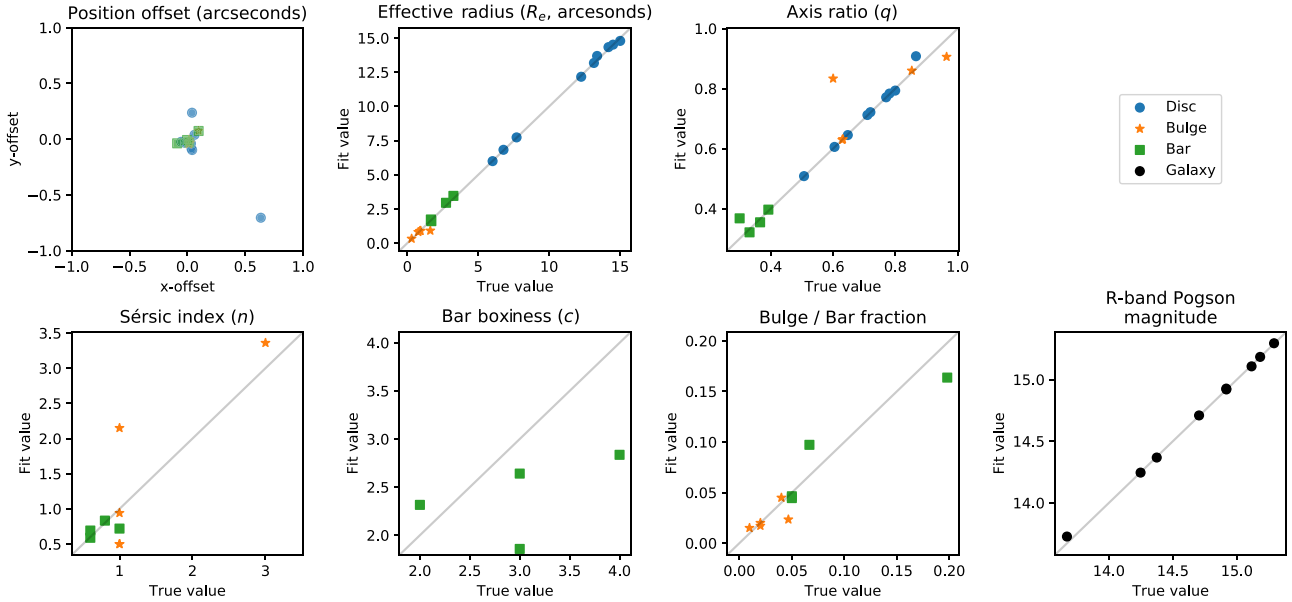

Figure 9. Plots examining the accuracy of fit parameters for the calibration subset of galaxies. Most parameters are recovered to a high degree of accuracy, however Sérsic index and boxiness are difficult to determine using the gradient descent alone, as they do not significantly impact the goodness of fit (Lackner & Gunn 2012). The error in the fit values reflects this problem.

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

3.3.1. Comparison to Galaxy Zoo Morphology

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

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

We also compare the number of spiral arms aggregated by *Galaxy Builder* with the responses to the GZ2 "number of arms" question (of which the possible responses were one, two, three, four, more than four, or "can't tell"). We attempt to account for the spread in volunteer answers to this question by binning responses, rather than using the mean or modal response. The results of this comparison can be seen in Figure 12. The area of each circle can be seen as the level of agreement between *Galaxy Builder* aggregate models and GZ2 classifiers, it is defined as

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

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

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

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

3.3.2. Comparison to One-component Fit—Axis Ratio

We compare the axis ratios of the disks of *Galaxy Builder* aggregate models (without fitting) to the axis ratio of a 2D Sérsic fit to the *r*-band SDSS image of each galaxy (as provided in the NSA catalog, Blanton et al. 2011). The resulting scatter is shown in Figure 13; for these untuned models there is an error of ∼0.1, consistent with our expected errors (derived in Section 2.6).

We observe a clustering of outlying values around $b/a = 0.5$. This is almost certainly due to the drawing tool ellipse having a default axis ratio of 0.5. Where this default is a "good enough" fit, we hypothesize that volunteers are less likely to

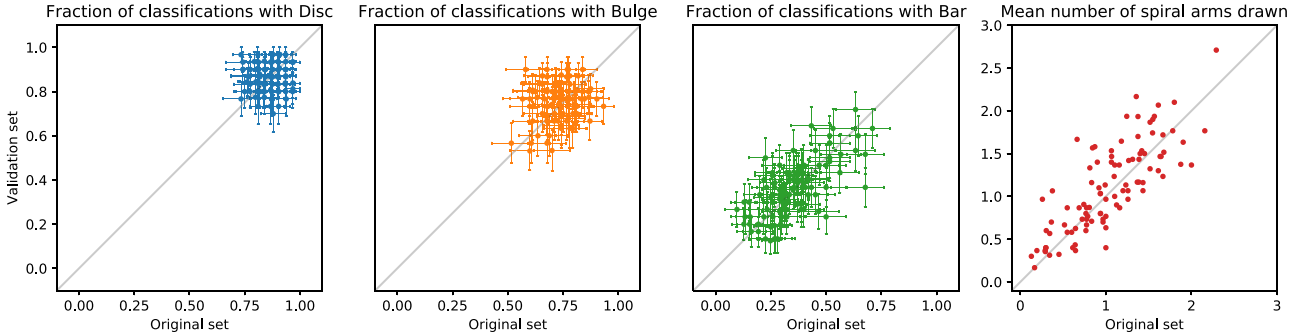

Figure 10. Comparison of the frequency of use of components in volunteer models between the original and validation sets of classifications. Errors shown on the disk, bulge, and bar arise from binomial error estimation. We see that classifications are generally consistent within the errors, which validates our assumption of volunteer independence.

modify it, while if it needs to move a long way, they find a more refined value. Overall, we see that 36% of all disk components drawn by volunteers were left at the default axis ratio. We recommend that future projects should carefully consider their interface design to minimize this bias (e.g., forcing volunteers to draw both the major and minor axis), but the fitting process we implement on the aggregate models successfully removes the bias, and the overall scatter does not change significantly.

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

3.3.3. Comparison to Disk–Bulge Models

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

One of the largest catalogs of 2D multicomponent fits has been compiled by Simard et al. (2011), who 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 fit: a pure Sérsic model, an exponential disk and de Vaucouleurs bulge model (hereafter exp+deV), and an exponential disk and a Sérsic bulge model (exp+S). Fitting was performed using the Metropolis algorithm, which is resilient to local minima and therefore suitable for the complex likelihood space of galaxy photometric modeling. Lackner & Gunn (2012) similarly fit two models to SDSS main-sample galaxies: an exponential disk and exponential bulge (exp+exp), and an exponential disk and de Vaucouleurs bulge. They used a Levenberg–Marquardt gradient-descent algorithm, with initial parameters taken from previous SDSS analysis.

We compare our central component fraction (the flux of the bulge and bar relative to the total model flux) to the bulge fraction from Simard et al. (2011), where their analysis indicated genuine bulge+disk systems ($P_{pS} \leqslant 0.32$). We compare to the Lackner & Gunn (2012) bulge fractions only when their model selection criteria determined that a model was the best-fit model. We see a strong correlation with significant scatter (Figure 14). The relationship to exp+deV models appears to be lower than 1:1, while the relationship to exp+exp models is greater than 1:1, highlighting the dependence of the bulge fraction on the Sérsic index. Taking *Galaxy Builder* results as ground truth implies that exp+deV places too much light into the bulge, while exp+exp places too little light there.

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

Another comprehensive catalog of 2D two-component fits is that of Meert et al. (2015), who fit identical models to Simard et al. (2011) on $\sim 7 \times 10^5$ galaxies imaged by SDSS, using GALFIT and PYMORPH (Vikram et al. 2010). They made use of a set of logical filters to distinguish between model fits, allowing them to identify cases where the model did not converge to a physically meaningful result. There is an overlap of 86 *Galaxy Builder* galaxy models with their "intermediate catalog," and we see some scatter between measured parameters (see Figure 15). The modeling of spiral arms does not appear to impact the measured disk parameters, with disk size and ellipticity showing strong agreement between the catalogs. We see significant scatter in the bulge Sérsic index, especially when a bar is present. The total luminosity is not strongly affected by the addition of detail to the model.

3.3.4. Comparison to Disk–Bulge–Bar Models

Kruk et al. (2018) performed multicomponent (up to three) multiband decompositions of a selection of SDSS galaxies, 23 of which were also classified in *Galaxy Builder* (with 16 in the repeated validation subset). Figure 16 compares the axis ratios and effective radii of bulges, disks, and bars in Kruk et al. (2018) to those present in the fit models. We see good consistency in effective radii of all components in the majority of galaxies. There is more scatter in the fit axis ratios of components. In particular, we observe that many of the *Galaxy Builder* bulges reach the imposed lower boundary. Comparing

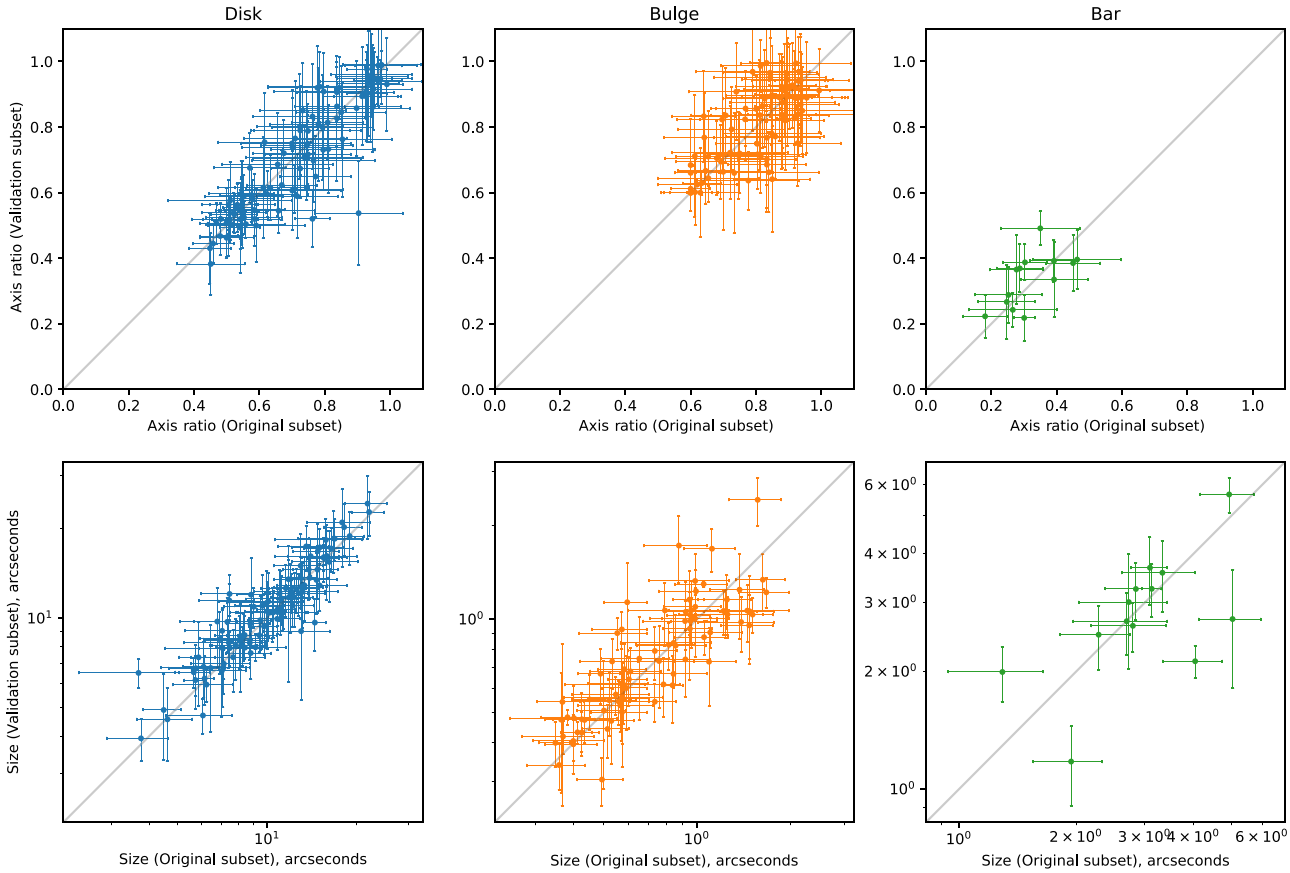

Figure 11. Comparison of the component shape in aggregate models between the original and validation sets. Errors are obtained through the sample variance of clustered components, as detailed in Section 2.6. We see close agreement between aggregate components from the two sets, suggesting that the clustering method is robust to the scatter in classifications.

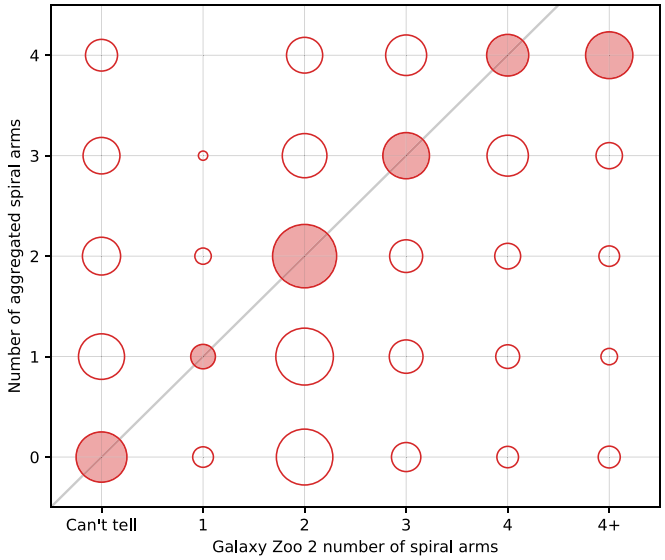

Figure 12. Density plot of GZ2 vote counts for spiral arm number vs. the number of spiral arms obtained through aggregation. The area of each circle can be seen as the level of agreement between *Galaxy Builder* aggregate models and GZ2 classifiers, and is defined by Equation (3). The circle with the largest area for each possible GZ2 response is highlighted by shading. The 1:1 relationship suggests that the clustering method correctly recovers the behavior of volunteers.

the central component fraction between *Galaxy Builder* models and those in Kruk et al. (2018), we see next to no scatter.

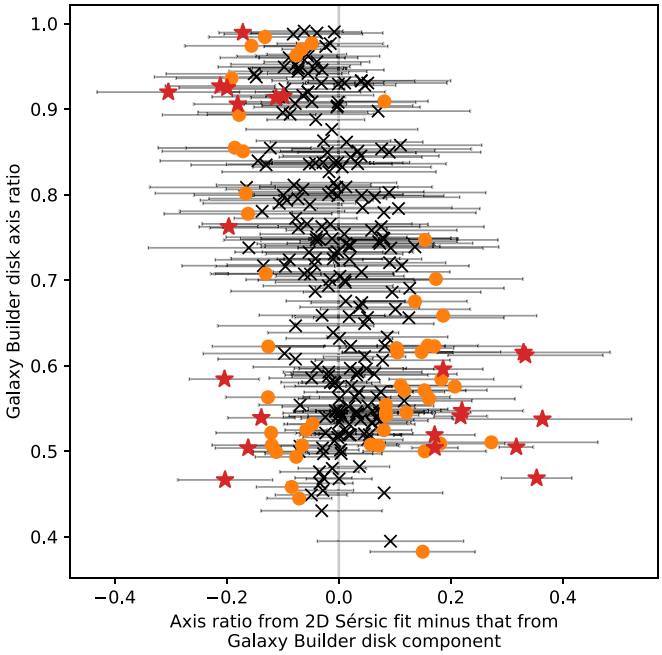

Figure 13. Difference between the axis ratios of the aggregated disk component (before fitting) to the results of an *r*-band Sérsic profile fit. Points between 1σ and 2σ are highlighted as orange squares, and points outside 2σ are shown as red stars. While the overall relationship is good, the increase prevalence of points outside 2σ is a clear indication of bias caused by the *Galaxy Builder* online user interface.

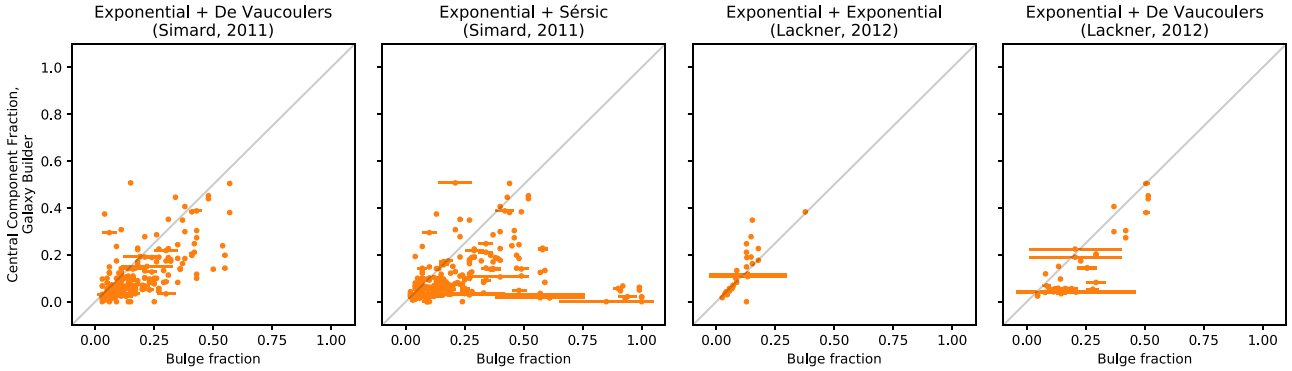

Figure 14. Scatter plots comparing the flux ratio from central components (bulge and bar) to the total flux between fit models from *Galaxy Builder* and two-component models in the literature. Our models are broadly consistent with their results, but should be more accurate for complex galaxies because we account for galaxy bars.

3.3.5. Comparison to Disk–Bulge–Bar–Spiral Models

To the best of our knowledge, no photometric models exist for the *Galaxy Builder* sample that contain a spiral arm structure. The closest comparable result is that produced by Gao & Ho (2017), but the galaxies they used are not in the Sloan footprint.

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

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

4. Summary and Conclusions

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

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

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

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

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

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

Upcoming survey missions such as the LSST (Ivezić et al. 2019) and Euclid (Laureijs et al. 2011; Amiaux et al. 2012) present a rich source of astrophysical data. However, the approach detailed in this paper will not be sufficient to deal with the volume of galaxies that these surveys will image (twenty billion and two billion, respectively, although a large proportion of these will not benefit from detailed photometric modeling). Tools such as *Galaxy Builder* may serve an important role in the generation of training catalogs for scalable machine-learning techniques, in an analogous manner to that currently employed for visual morphological classification in Galaxy Zoo: Enhanced (Walmsley et al. 2020).

We were able to obtain aggregate models for 296 images with an average rate of one galaxy per day, and fit photometric models for 294 images. At the time of writing and to the best of our knowledge, the number of photometric models obtained here is still significantly larger than the largest sample obtained through purely computational photometric fitting of a disk, bulge, bar, and spiral arms in galaxies (10 galaxies, Gao &

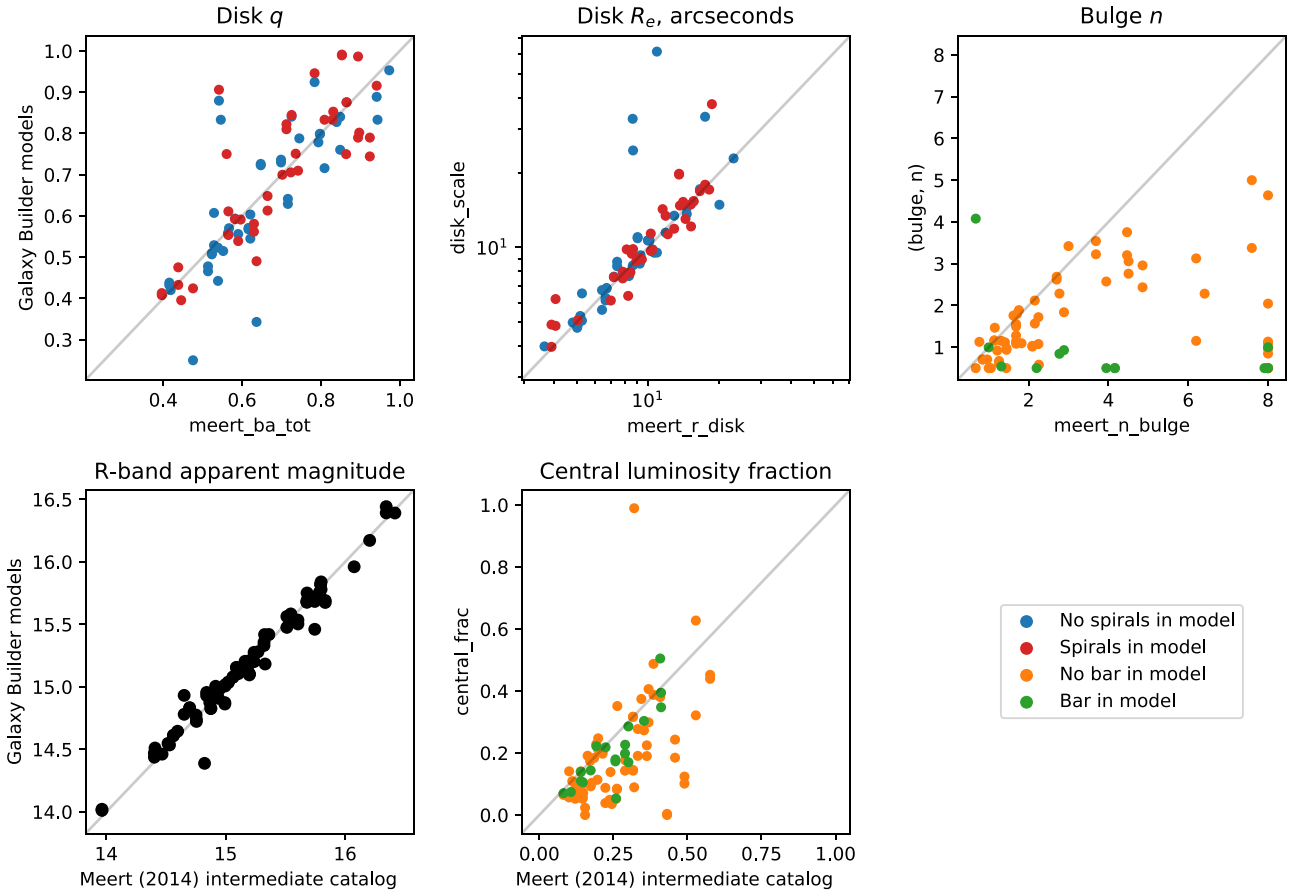

Figure 15. Scatter plots comparing measured model parameters between Meert et al. (2015, *x*-axis) and *Galaxy Builder* (*y*-axis). We note that adding spirals to a model does not strongly impact the disk parameters, but the presence of a bar has a significant impact on the bulge Sérsic index measurement.

Ho 2017, who also included rings, disk-breaks, and further components).

The software used to generate image cutouts, perform clustering and aggregation of volunteer models, and fit photometric models is available under a GNU general public licence on GitHub.[18] We hope that publishing this code with the paper promotes transparency and accountability in astrophysical software development. All models created as part of the *Galaxy Builder* project will be available on the Galaxy Zoo website.[19]

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

We would like to thank our referee for their input; suggesting numerous ways to improve the clarity and content of this paper, and providing avenues of discussion that we had not previously covered. We would also like to thank Ross Hart for providing the target catalog for the stellar mass-complete sample.

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

This publication uses data generated via the Zooniverse.org platform, development of which is funded by generous support, including a Global Impact Award from Google, and by a grant from the Alfred P. Sloan Foundation. We would also like to thank the 2340 volunteers who have submitted classifications to the *Galaxy Builder* project, especially user EliabethB, whose presence on the *Galaxy Builder* forum on top of a large number of galaxies modeled has been a huge help.

Montage is funded by the National Science Foundation under grant No. 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.

This project was partially funded by a Google Faculty Research Award to Karen Masters (https://ai.google/research/outreach/faculty-research-awards/), and Timothy Lingard acknowledges studentship funding from the Science and Technology Facilities Council (ST/N504245/1).

[18] http://github.com/tingard/gzbuilder_analysis
[19] https://data.galaxyzoo.org

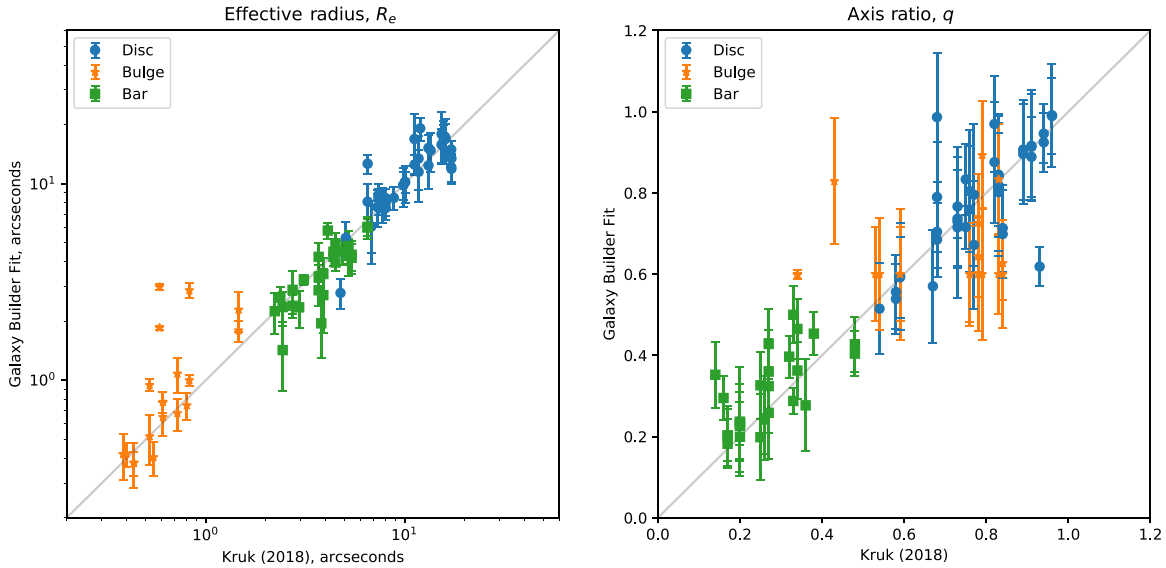

Figure 16. Comparison between *Galaxy Builder* fit models and the result of three-component multiwavelength fits performed by Kruk et al. (2018). Disks, bulges, and bars are shown as blue circles, orange stars, and green squares, respectively. The left panel compares the effective radii of the components, and the right panel compares the component axis ratio. The components match well, with bulges showing the most scatter. Bulges in *Galaxy Builder* fit models are often stuck at the lower allowed value, even though the initial conditions were physically motivated.

Appendix
Model Fitting

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

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

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

,where

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

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

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

$$I_{\rm pix}(\boldsymbol{P}) = \frac{1}{\delta_x\delta_y}\int_{-\delta_y/2}^{\delta_y/2}\int_{-\delta_x/2}^{\delta_x/2} dxdy\ I_{\rm sersic} \times \left(\boldsymbol{P} + \begin{pmatrix}\delta_x\\ \delta_y\end{pmatrix}\right). \tag{A3}$$

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

An inclined rotated logarithmic spiral requires parameters brightness I_s, spread s, minimum and maximum θ ($\theta_{\rm min}$ and $\theta_{\rm max}$), an amplitude A, pitch angle ϕ, position $\boldsymbol{\mu}$, position angle ψ, and axis ratio q, where $\boldsymbol{\mu}$, ψ, and q are inherited from the disk component.

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

$$D_{\rm s}(\boldsymbol{P}) = \min_{\theta\in[\theta_{\rm min},\theta_{\rm max}]} \times \left|\left|\boldsymbol{P} - \boldsymbol{\mu} - Ae^{\theta\tan\phi}\begin{pmatrix}\cos\psi & \sin\psi\\ -\sin\psi & \cos\psi\end{pmatrix}\begin{pmatrix}q\cos\theta\\ \sin\theta\end{pmatrix}\right|\right|^2. \tag{A4}$$

In practice, the spiral distance was approximated using the distance to a polyline with 200 vertices, as solving the above minimization for each pixel at each fitting step is computationally intractable. We also adjust A, $\theta_{\rm min}$, and $\theta_{\rm max}$ to account for the rotation of the disk component from its starting value, in order to prevent spirals inadvertently moving far from starting locations for face-on disks (which have poorly constrained position angles). These adjustments are

$$\begin{aligned} A' &= Ae^{\Delta\psi\tan\phi},\\ \theta'_{\rm min} &= \theta_{\rm min} - \Delta\psi,\\ \theta'_{\rm max} &= \theta_{\rm max} - \Delta\psi. \end{aligned} \tag{A5}$$

The pixel brightness is then calculated as

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

For the fit, we parameterize disk I_e as the Sérsic total luminosity, given by

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

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

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

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

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

ORCID iDs

Timothy K. Lingard https://orcid.org/0000-0002-0529-7427
Karen L. Masters https://orcid.org/0000-0003-0846-9578
Coleman Krawczyk https://orcid.org/0000-0001-9233-2341
Chris Lintott https://orcid.org/0000-0001-5578-359X
Sandor Kruk https://orcid.org/0000-0001-8010-8879
Brooke Simmons https://orcid.org/0000-0001-5882-3323
Robert Simpson https://orcid.org/0000-0003-2896-2503
Steven Bamford https://orcid.org/0000-0001-7821-7195
Robert C. Nichol https://orcid.org/0000-0003-0939-6518

References

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
Albareti, F. D., Allende Prieto, C., Almeida, A., et al. 2017, ApJS, 233, 25
Allen, P. D., Driver, S. P., Graham, A. W., et al. 2006, MNRAS, 371, 2
Amiaux, J., Scaramella, R., Mellier, Y., et al. 2012, Proc. SPIE, 8442, 84420Z
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Bamford, S. P., Häußler, B., Rojas, A., & Borch, A. 2011, in ASP Conf. Ser. 442, Astronomical Data Analysis Software and Systems XX, ed. I. N. Evans et al. (San Francisco, CA: ASP), 479
Barden, M., Rix, H.-W., Somerville, R. S., et al. 2005, ApJ, 635, 959
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
Boonchoo, T., Ao, X., & He, Q. 2018, arXiv:1801.06965
Bradbury, J., Frostig, R., Hawkins, P., et al. 2018, JAX: Composable Transformations of Python+NumPy Programs v0.1.55 http://github.com/google/jax
Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. 2000, in Proc. 2000 ACM SIGMOD Int. Conf. on Management of Data, SIGMOD '00, ed. M. Dunham et al. (New York: ACM), 93, doi:10.1145/342009.335388
Byrd, R., Lu, P., Nocedal, J., & Zhu, C. 1995, SIAM J. Sci. Comput., 16, 1190
Carollo, C. M., Stiavelli, M., Seigar, M., de Zeeuw, P. T., & Dejonghe, H. 2002, AJ, 123, 159
Ciambur, B. C. 2016, PASA, 33, e062
Davis, B. L., Graham, A. W., & Cameron, E. 2019, ApJ, 873, 85
Davis, D. R., & Hayes, W. B. 2014, ApJ, 790, 87
de Jong, R. S. 1996, A&AS, 118, 557
de Vaucouleurs, G., de Vaucouleurs, A., Corwin, H. G., Jr., et al. 1991, Third Reference Catalogue of Bright Galaxies Vol. I, II, & III (New York: Springer)
Díaz-García, S., Salo, H., Knapen, J. H., & Herrera-Endoqui, M. 2019, arXiv:1908.04246
Dobbs, C., & Baba, J. 2014, PASA, 31, 35
Elmegreen, B. G., & Elmegreen, D. M. 1985, ApJ, 288, 438
Gadotti, D. A. 2011, MNRAS, 415, 3308
Gao, H., & Ho, L. C. 2017, ApJ, 845, 114
Gao, H., Ho, L. C., Barth, A. J., & Li, Z.-Y. 2018, ApJ, 862, 100
Graham, A. W. 2001, AJ, 121, 820
Graham, A. W., & Driver, S. P. 2005, PASA, 22, 118
Hart, R. E., Bamford, S. P., Hayes, W. B., et al. 2017, MNRAS, 472, 2263
Hart, R. E., Bamford, S. P., Willett, K. W., et al. 2016, MNRAS, 461, 3663
Head, J. T. C. G., Lucey, J. R., & Hudson, M. J. 2015, MNRAS, 453, 3729
Holincheck, A. J., Wallin, J. F., Borne, K., et al. 2016, MNRAS, 459, 720
Hopkins, P. F., Croton, D., Bundy, K., et al. 2010, ApJ, 724, 915
Ivezić, Ž, Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Jacob, J. C., Katz, D. S., Berriman, G. B., et al. 2010, arXiv:1005.4454
Kelvin, L. S., Driver, S. P., Robotham, A. S. G., et al. 2012, MNRAS, 421, 1007
Kormendy, J., Drory, N., Bender, R., & Cornell, M. E. 2010, ApJ, 723, 54
Kruk, S. J., Lintott, C. J., Bamford, S. P., et al. 2018, MNRAS, 473, 4731
Kruk, S. J., Lintott, C. J., Simmons, B. D., et al. 2017, MNRAS, 469, 3363
Lackner, C. N., & Gunn, J. E. 2012, MNRAS, 421, 2277
Lange, R., Moffett, A. J., Driver, S. P., et al. 2016, MNRAS, 462, 1470
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv:1110.3193
Lilly, S., Schade, D., Ellis, R., et al. 1998, ApJ, 500, 75
Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, MNRAS, 389, 1179
Lupton, R., Blanton, M. R., Fekete, G., et al. 2004, PASP, 116, 133
Masters, K. L. & the Galaxy Zoo Team 2020, in Proc. IAU Symp. 353, Galactic Dynamics in the Era of Large Surveys, ed. M. Valluri & J. A. Sellwood (Cambridge: Cambridge Univ. Press), 205
Masters, K. L., Nichol, R. C., Hoyle, B., et al. 2010, MNRAS, 411, 2026
Meert, A., Vikram, V., & Bernardi, M. 2015, MNRAS, 446, 3943
Mendel, J. T., Simard, L., Palmer, M., Ellison, S. L., & Patton, D. R. 2014, ApJS, 210, 3
Mendez-Abreu, J., Ruiz-Lara, T., Sanchez-Menguiano, L., et al. 2017, A&A, 598, 32
Parry, O. H., Eke, V. R., & Frenk, C. S. 2009, MNRAS, 396, 1972
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, J. Mach. Learn. Res., 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., et al. 2016, ApJL, 827, L2
Pozzetti, L., Bolzonella, M., Zucca, E., et al. 2009, A&A, 523, A13
Rampazzo, R., Uslenghi, M., Giorgiev, I. Y., et al. 2019, arXiv:1910.02664
Robotham, A. S. G., Taranu, D. S., Tobar, R., Moffett, A., & Driver, S. P. 2017, MNRAS, 466, 1513
Rodrigues, M., Puech, M., Flores, H., Hammer, F., & Pirzkal, N. 2018, MNRAS, 475, 5133
Sahu, N., Graham, A. W., & Davis, B. L. 2019, ApJ, 876, 155
Sérsic, J. L. 1963, BAAA, 6, 41
Simard, L., Mendel, J. T., Patton, D. R., Ellison, S. L., & McConnachie, A. W. 2011, ApJS, 196, 11
Simard, L., Willmer, C. N. A., Vogt, N. P., et al. 2002, ApJS, 142, 1
Simmons, B. D., Lintott, C., Willett, K. W., et al. 2017, MNRAS, 464, 4420
Simpson, R., Page, K. R., & De Roure, D. 2014, in Proc. 23rd Int. Conf. on World Wide Web, WWW '14 Companion, ed. C.-W. Chung (New York: ACM), 1049, doi:10.1145/2567948.2579215
Springel, V., & Hernquist, L. 2005, ApJL, 622, L9
Stoughton, C., Lupton, R. H., Bernardi, M., et al. 2002, AJ, 123, 485
van der Wel, A., Bell, E. F., Häussler, B., et al. 2012, ApJS, 203, 24
Vika, M., Bamford, S. P., Häußler, B., & Rojas, A. L. 2014, MNRAS, 444, 3603
Vikram, V., Wadadekar, Y., Kembhavi, A. K., & Vijayagovindan, G. V. 2010, MNRAS, 409, 1379
Virtanen, P., Gommers, R., & Oliphant, T. E 2020, NatMe, 17, 261
Wales, D., & Doye, J. 1998, arXiv:cond-mat/9803344
Walmsley, M., Smith, L., Lintott, C., et al. 2020, MNRAS, 491, 1554
Willett, K. W., Galloway, M. A., Bamford, S. P., et al. 2017, MNRAS, 464, 4176
Willett, K. W., Lintott, C. J., Bamford, S. P., et al. 2013, MNRAS, 435, 2835
Zaninetti, L. 2014, arXiv:1401.0287