Draft version July 28, 2017
Typeset using LaTeX **preprint** style in AASTeX61

LEGACY SURVEY LARGE GALAXY ATLAS

John Moustakas[1] and Dustin Lang[2]

[1]*Department of Physics & Astronomy, Siena College, 515 Loudon Road, Loudonville, 12211*
[2]*Dunlap Institute for Astronomy & Astrophysics and Department of Astronomy & Astrophysics University of Toronto*

ABSTRACT

I like cheese.

jmoustakas@siena.edu
dstndstn@gmail.com

1. INTRODUCTION

The Dark Energy Spectroscopic Instrument (DESI) will select spectroscopic targets using data from three precursor ground-based optical imaging surveys—DECaLS (DECam Legacy Survey), MzLS (Mayall z-band Legacy Survey), and BASS (Beijing-Arizona Sky Survey).[1] The positions, shapes, sizes, colors, and other observable properties of the galaxies and stars in these datasets are being measured by the `legacypipe`[2] photometric pipeline, which uses the `Tractor`[3] to build a probabilistically justified model of each source.

However, large galaxies—defined here to be galaxies with large (angular) sizes projected on the sky, typically larger than $5-10$ arcsec—are not being modeled properly by the production version of the pipeline, which will have significant implications for DESI target selection, especially the Bright Galaxy Survey (BGS). A closely related issue is that each DESI pointing of 5000 fibers will contain at least one galaxy 30 arcsec or larger, so by analyzing large galaxies we will be able to identify and select the desired fiber positions in fields containing large galaxies (e.g., along the major axis or on the most prominent star-forming regions). Finally, large galaxies should be studied in their own right because of the tremendous insight into galaxy formation they provide.

The purpose of this *TechNote* is to report on the (ongoing) work we have undertaken to carry out a custom analysis of all the large galaxies in the approximately $14,000$ deg^2 DESI footprint. Focusing first on the DECaLS Data Release 2 (DECaLS/DR2), we select an angular diameter limited sample of $10,654$ galaxies with existing grz imaging and we assess how the current version of the pipeline handles sky-subtraction, deblending, and the photometry of one test case, UGC04203. We present the results of preliminary code written to deal with some of the unique challenges posed by large galaxies (see Section 1.1), and we conclude with a discussion of key goals moving forward.

All the code written as part of this analysis is publicly accessible within the `legacypipe` Github repository[4], while this *TechNote* itself can be found in the DESI SVN repository at https://desi.lbl.gov/svn/docs/technotes/imaging/large-galaxies.

1.1. *Challenges Posed by Large Galaxies*

Detecting, deblending, and modeling the surface brightness profiles of large galaxies poses several key challenges for ground-based optical imaging surveys. Here, we highlight some of these specific issues, in no particular order of importance:

- *Sky subtraction*—Galaxies do not have sharp, or truncated edges, making it difficult to ascertain where the galaxy ends and the sky begins. Most standard sky-subtraction algorithms, which work perfectly well for stars and galaxies with typical (apparent) sizes, tend to subtract the light contained in the outer parts of galaxies. Perhaps surprisingly, these outer, low surface-brightness isophotes may contain 50% or more of the integrated light of the galaxy, depending on its morphological type, mass-assembly history, and large-scale environment.
- *Morphological complexity*—Almost by definition, large galaxies are much better resolved spatially than their smaller (typically more distant) counterparts. For example, visual inspection of large galaxies may reveal distinct bulge and disk components, bars and rings, spiral arms,

[1] legacysurvey.org
[2] https://github.com/legacysurvey/legacypipe
[3] http://thetractor.org
[4] The code is currently in the `largegalaxies` branch although it will be merged eventually into the `master` branch.

dust lanes, tidally stripped material, azimuthal asymmetries, and other unique morphological characteristics (which is largely what makes large galaxies so fascinating to study!). However, it is clear that single-component, azimuthally symmetric models (e.g., exponential or single-Sérsic) cannot fully capture this complexity, requiring more sophisticated and flexible galaxy models. This need is especially acute when dealing with multi-band data, since many galaxy types exhibit radial (occasionally, non-monotonic!) color gradients.

- *Detecting and deblending*—Because of their size, one of the most commonly encountered issues when analyzing large galaxies is *shredding*, which is when a single galaxy is (improperly) shredded into multiple disjoint components. A related issue is when individual (resolved) components *within* the galaxy (e.g., star-forming regions) are identified as distinct objects. Similarly, even when the galaxy is not shredded, it is still challenging to efficiently detect and measure the colors of galaxies and stars located on top of or near large galaxies, owing to the non-uniform "background" these sources find themselves on.

In this *TechNote* we begin to investigate these and other issues, with the ultimate goal of implementing solutions to all of them.

2. SAMPLE SELECTION

Our goal is to define a parent sample of large galaxies—galaxies with large angular diameters— in Legacy Survey imaging. Although there are many different published catalogs of large galaxies, we have found the `HyperLeda`[5] extragalactic database to be the most comprehensive and homogeneous. We select our parent sample by executing the following two SQL queries:[6]

```
SELECT
  pgc, objname, objtype, al2000, de2000, type,
  multiple, logd25, logr25, pa, bt, it, v
WHERE
  logd25 > 0.05 AND logd25 < 0.5 and (objtype='G' or objtype='g' or
  objtype='M' or objtype='M2' or objtype='M3' or objtype='MG' or
  objtype='MC')
```

and

```
SELECT
  pgc, objname, objtype, al2000, de2000, type,
  multiple, logd25, logr25, pa, bt, it, v
WHERE
  logd25 > 0.5 and (objtype='G' or objtype='g' or
  objtype='M' or objtype='M2' or objtype='M3' or objtype='MG' or
  objtype='MC')
```

[5] http://leda.univ-lyon1.fr

[6] See http://leda.univ-lyon1.fr/leda/fullsql.html for the SQL interface and http://leda.univ-lyon1.fr/leda/ldoc.html for documentation on each of the stored quantities. Note that two separate queries are necessary to stay within the time and memory limits of the SQL server.

The resulting two output tables, consisting of 2, 143, 628 unique galaxies (or unresolved galaxy pairs and triplets, e.g., mergers), are parsed and combined into a single FITS catalog which is angular-diameter limited to 0.11 arcmin, or approximately 6.7 arcsec (subject to the surface brightness completeness of HyperLeda). Following historical precedent, the angular diameter is given by $D(25)$, the diameter of the galaxy at the 25 mag arcsec^{-2} isophote in the B-band (Johnson, Vega).

Next, we reduce this sample further by applying two additional cuts. First, we restrict the sample to have $0.5 < D(25) < 10$ arcmin. The lower limit was somewhat arbitrarily chosen to reduce the sample to a more manageable size, while the upper cut was applied to remove galaxies like Andromeda=M31 from the sample, which require even more care and effort to analyze properly. We note here that 10 arcmin is roughly the diameter of a single DECam CCD. And finally, we remove galaxies which do not have grz imaging in DECaLS/DR2, leaving a final sample of 10, 654 galaxies.

In Figure ?? we plot the positions of all the galaxies in the parent sample (blue density map) and the galaxies in our final sample (red squares). Obviously, the size of the sample with three-band imaging will grow significantly in subsequent DECaLS data releases, as well as once we begin to incorporate data from the BASS and MzLS imaging surveys.

Finally, in Figure ?? we plot $D(25)$ versus B-band magnitude[7] for both samples. On this figure we annotate two particularly large galaxies—M31 and the Small Magellanic Cloud—which are excluded by our angular diameter cuts (horizontal dashed lines). We plot our final sample of galaxies using small red points.

3. MODELING LARGE GALAXIES

To highlight our progress modeling the sample of large galaxies, we focus on one test case, UGC04203 (also known as the Phoenix galaxy), a face-on Sa galaxy with an angular diameter of 0.85 arcmin. In Section 3.1 we summarize the performance of the production (DR2) version of the legacypipe pipeline, in Section 3.2 we discuss masking and sky-subtraction around large galaxies, and in Section 3.3 we show preliminary results from a version of the pipeline which is customized for large galaxies.

3.1. *Production Pipeline*

In Figure ?? we illustrate the performance of the DR2 version of the legacypipe pipeline. The left, middle, and right panels show the data, model, and residual (data minus model) color mosaics. The small circles in each panel identify detected sources, where the color of the circle corresponds to the final morphological classification of the object: PSF=white, SIMP=red, EXP=orange, DEV=cyan, and COMP=yellow.[8] Sources are identified as 5σ peaks above the rms noise of the image after convolving with a Gaussian kernel whose full-width at half-maximum (FWHM) is given by the median seeing of the data.

Overall, we find that the DR2 pipeline is doing a terrible job of modeling the large, central galaxy. Although UGC04203 has been successfully classified as a COMP (linear combination of an exponential plus de Vaucouleurs surface-brightness profile), its center has been significantly over-subtracted and the outer envelope has been under-subtracted. Moreover, the central panel (model) shows clear aliasing in Fourier space. Finally, we find that several significant sources have not been identified,

[7] The B-band magnitudes are tabulated in HyperLeda and are only meant to be indicative of the galaxy brightness. Many of these magnitudes are based on decades-old photographic plates.

[8] See http://legacysurvey.org/dr2/catalogs for a description of these classifications.

including two objects in the outer envelope of UGC04203 (lower-right quadrant) and one source in the bottom-left part of the footprint. Failing to detect these sources creates several significant issues when the pipeline attempts to optimize the parameters (shapes, sizes, and surface-brightness profiles) of the sources it *does* detect.

3.2. *Improved Masking and Sky Subtraction*

Unlike the typical, “small” stars and galaxies which appear in Legacy Survey imaging, large galaxies have extended outer envelopes which may include a significant amount of light. Consequently, both object masking and sky-subtraction are critically important. Masking and sky-subtraction, of course, must occur at the CCD-level data, since the `Tractor` operates from the unremapped, unstacked, pixel-level data.

To illustrate the available multi-band imaging for UGC04203, in Figure **??** we plot the positions of the g-band (left panel), r-band (middle panel), and z-band (right panel) exposures, where each individual CCD has been color-coded as indicated in the legend. For reference, the size of the dashed square is five times the angular diameter of the galaxy, and the footprint covered by the square corresponds to the sky region shown in Figure **??**. Finally, the small circle at the center of each panel in Figure **??** indicates the angular diameter of UGC04203.

In Figure **??** we focus on one of these CCDs—`ccd04`, an r-band exposure with UGC04203 positioned reasonably close to the center of the field—in more detail. In this figure we show the image (including the sky), the object mask constructed by the DR2 pipeline (one and zero indicate masked and unmasked pixels, respectively), the object mask generated by the large-galaxy pipeline, the DR2 model of the sky, and finally the model of the sky based on the customized large-galaxy pipeline. The white circle in each panel shows the position and angular extent of UGC04203.

Figure **??** shows two key ideas. First, the large-galaxy pipeline is much more aggressive at masking pixels containing astrophysical sources, especially the outer isophotes of UGC04203. And second, the production version of the pipeline clearly over-subtracts the sky in and around UGC04203, whereas the large-galaxy pipeline (currently) subtracts a uniform sky background from the data. Other diagnostic plots (not shown here) indicate that the low-resolution spline sky-subtraction implemented in the DR2 version of `legacypipe` is overly aggressive (flexible) around large galaxies, and can lead to the kind of over-subtraction systematics shown in Figure **??** (fourth panel from the left).

3.3. *Large-Galaxy Pipeline*

Figure **??** shows the final result of running the customized large-galaxy pipeline on UGC04203. As in Figure **??** (see also Section 3.1), we show—from left to right—a color montage of the data, the model, and the residuals, and we show the detected sources and their classification using colored circles. Overall, we find a significant improvement in the two-dimensional model of UGC04203, which as before is classified as a composite galaxy. In particular, the residual image shows astrophysically interesting features, including spiral structure and low-level shells at large galactocentric radius, the latter of which is generally cited as evidence of a merger-driven origin for spheroidal and bulge-dominated galaxies.

One outstanding issue shown in Figure **??**, however, is the source in the lower-left quadrant which fails to be detected. To compensate for this extraneous flux, the pipeline chooses a thin, almost needle-shaped galaxy profile for the adjacent source. Although this issue does not specifically affect our model of UGC04203, it is a recurring issue which we hope to address in a future version of the `legacypipe` pipeline.

4. DISCUSSION

Discuss.

5. CONCLUSIONS & NEXT STEPS

This *TechNote* reports on our ongoing effort to develop a modified version of the `legacypipe` photometric pipeline which is optimized for large galaxies, with the ultimate goal of producing a Legacy Survey Large Galaxy Atlas (LSLGA). In addition to enabling detailing multi-wavelength analyses of large, spatially well-resolved galaxies, this effort will also significantly improve DESI target selection for the Bright Galaxy Survey and in pointings containing large galaxies.

In addition to addressing the outstanding issues discussed above, in the near future we intend to work on the following additional threads, in no particular order of importance:

- We will revisit the parent galaxy sample selection to ensure that low surface brightness galaxies are not being excluded (e.g., by incorporating H I-selected galaxies).
- We will run the large-galaxy pipeline on the full sample of $10,654$ galaxies, in order to identify the most common problems and failure modes.
- We will build a web-based interface on the `legacysurvey.org` web-server (work that we have already begun), in order to make it easy to inspect the data and model outputs for the large-galaxy sample.
- Using existing code, we will use image simulations—whereby we insert simulated large galaxies into the CCD-level data—to test the performance of the large-galaxy pipeline under different initial conditions and in different regimes.
- We will incorporate more sophisticated galaxy models, including multi-component Sérsic profiles, into the library of possible two-dimensional galaxy models.
- We will investigate ways of accounting for second-order variations in the galaxy models, such as radial color gradients or azimuthal asymmetries.
- And finally we will engage with the Data Systems and Target Selection Working Group to incorporate the catalogs constructed using the large-galaxy pipeline into DESI target selection and fiber assignment.