PHOTOMETRIC AND SPECTROSCOPIC FOOTPRINT CORRECTIONS IN THE
SLOAN DIGITAL SKY SURVEY’S 6TH DATA RELEASE

MIKE A. SPECIAN AND ALEX S. SZALAY
Department of Physics and Astronomy, The Johns Hopkins University, 3701 San Martin Drive,
Baltimore, MD 21218, USA; mspecial@jhu.edu, szalay@jhu.edu

Received 2015 November 3; accepted 2016 February 18; published 2016 March 29

ABSTRACT

We identify and correct numerous errors within the photometric and spectroscopic footprints (SFs) of the Sloan Digital Sky Survey’s (SDSS) 6th data release (DR6). Within the SDSS’s boundaries hundreds of millions of objects have been detected. Yet we present evidence that the boundaries themselves contain a significant number of mistakes that are being revealed for the first time within this paper. Left unaddressed, these can introduce systematic biases into galaxy clustering statistics. Using the DR6 Main Galaxy Sample (MGS) targets as tracers, we reveal inconsistencies between the photometric and SF definitions provided in the Catalog Archive Server (CAS), and the measurements of targets therein. First, we find that 19.7 deg² of the DR6 photometric footprint are devoid of MGS targets. In volumes of radii 7 h⁻¹ Mpc, this can cause errors in the expected number of galaxies to exceed 60%. Second, we identify seven areas that were erroneously included or excluded from the SF. Moreover, the tiling algorithm that positioned spectroscopic fibers during and between DRs caused many areas along the edge of the SF to be significantly undersampled relative to the footprint’s interior. Through our corrections, we increase the completeness 2.2% by trimming 3.6% of the area from the existing SF: The sum total of these efforts has generated the most accurate description of the SDSS DR6 footprints ever created.

Key words: catalogs – large-scale structure of universe – surveys

1. INTRODUCTION

Well-defined galaxy survey footprints are a necessity for precision cosmology. Accurate boundaries limit our focus to regions where data have been collected and processed in a consistent way. A footprint’s area can define the average angular density of objects, a necessary statistic for calculating the expected numbers of galaxies in volumes of space. They are used to determine spectroscopic completeness along lines-of-sight and can guide decisions regarding how to account for targets without spectra. They help us to better understand geometry-dependent effects like zero-point photometric offsets, and much more.

One of the most fruitful galaxy surveys of all time is the Sloan Digital Sky Survey (SDSS; York et al. 2000). The SDSS, which began its observations from Apache Point, New Mexico in 2000, is a multi-color, multi-fiber imaging, and spectroscopic survey purposed with measuring the positions and properties of hundreds of millions of celestial objects over a large fraction of the sky.

The SDSS employed a drift scanning approach by which photometric imaging was gathered in over 2000 rectangular scanlines as the sky rotated overhead. The union of these scanlines formed a mostly contiguous area referred to as the photometric footprint (PF). Subsets of the PF can be defined as well, e.g., the area that only contains primary observations of objects.

The properties of detected objects, of which there are approximately 500 million, were fed into a data pipeline that selected a small subset of those for spectroscopic observation. Spectroscopy was gathered by overlapping circular tiles, the union of which is referred to as the spectroscopic footprint (SF). Approximately one million objects were observed in this fashion. The SF is always smaller than and subsumed by the PF.

For both footprints, certain minimum criteria should be maintained. All imaged objects must exist within the PF, and all areas within the PF must be imaged. Likewise, all galaxies that are observed spectroscopically must lie within the SF. Because of limitations in fiber allocation, not all objects within the SF will have their spectra taken. However, the tiling algorithm that helps determine which objects are assigned fibers is designed to maximize the number of targets that have fibers assigned to them. This necessarily means that spectroscopic completeness, or the fraction of targets with measured spectra, will vary with direction. It is clear, though, that the footprint of a uniform spectroscopic survey should not contain areas where targets’ spectra are either grossly undersampled or completely nonexistent.

All of the information needed to test whether these criteria were met is stored on SDSS’s publicly accessible Catalog Archive Server (CAS),1 aka the SDSS database (Thakkar et al. 2000; Stoughton et al. 2002). Here one can access the geometric descriptions of scanlines and tiles needed to define the extents of the PF and SF. The database also contains imaging and spectroscopic data for three high-priority samples of galaxies—the Main Galaxy Sample (MGS; Strauss et al. 2002), Luminous Red Galaxies (LRGs; Eisenstein et al. 2001), and QSOs (Richards et al. 2002). The final calibrated collection of these objects is known as the Legacy Survey.

Checking these criteria involves comparing the distribution of objects with the footprint definitions as presented in the SDSS database. We have opted to use MGS targets as tracers since they outnumber both the LRGs and the QSOs by about an order of magnitude. The higher surface density of MGS targets best enables the comparisons needed to find gaps and excesses in coverage.

1 http://cas.sdss.org/dr6
Figure 1. SDSS DR6 imaging and spectroscopic footprints. Areas in dark gray mark the footprint of the Legacy Survey through DR5. Areas in light gray are new to DR6. Three STRIPEs in the southern Galactic cap are also part of the Legacy Survey. The sinuous line between the northern Galactic cap and the STRIPEs in the southern Galactic cap denotes the Galactic plane. Areas in red, blue, and green represent other areas observed in DR6 for projects unrelated to the Legacy Survey. This map is presented in units of R.A. and decl. in J2000.0 equatorial coordinates. This figure was first published in Adelman-McCarthy et al. (2008).

About once a year the SDSS issued a data release (hereafter: DR) containing the latest results with each new release subsuming earlier versions. With the release of DR6 in 2006 July, the imaging for the Legacy Survey was substantially complete—230 million unique objects had been imaged over an area of 8417 deg$^2$ (Adelman-McCarthy et al. 2008). Spectra had been taken for 790,220 of those over an area of 6860 deg$^2$ (see Figure 1).

DR6 was very much a survey in transition. Decisions regarding footprint geometry and spectroscopic coverage were made with an eye toward DR7. That is, the choice to collect spectra in some regions but not others was made with the anticipation that the resulting gaps would be filled in during the next DR. We find that this strategy complicates the measurements of certain quantities, like the expected number of galaxies in select regions of space. In this way, one cannot study DR6 effectively without also incorporating the geometry of DR7. It turns out that some of the geometrical region definitions relevant to DR6 are only specified in DR7. We will show a number of cases where this occurs and offer suggestions for improvement.

DR6 was the sixth of eight official DRs for the Legacy Survey. These data have been made public in a series of yearly DRs (Stoughton et al. 2002; Abazajian et al. 2003, 2004, 2005; Adelman-McCarthy et al. 2006, 2007, 2008; Abazajian et al. 2009; Aihara et al. 2011, hereafter the EDR, DR1, DR2, DR3, DR4, DR5, DR6, DR7, and DR8 papers, respectively).

We limit the majority of our analysis to DR6 for a few reasons. First, the survey is decidedly complete and unlikely to experience many future adjustments. Second, the geometry of the survey is provided in a way that allows its photometric and SFs to be studied in detail. Furthermore, by DR6 the Legacy Survey had probed the majority of its ultimate volume. Redshifts were known for hundreds of thousands of galaxies, enough to draw meaningful statistical conclusions. By falling in the midst of other DRs, the survey was dynamic. New objects would be added to those already observed. Given that the most fruitful period for surveys is likely to be while they are still active, it is important to note how intra-release assumptions and observational protocols can leave errors behind. DR6 was also the final DR to photometrically calibrate each scanline independently using the PT method (Smith et al. 2002). This provides a test bed for understanding systematic errors due to offsets in photometric zero-points, an analysis we take up in M. A. Specian & S. A. Szalay (2016, in preparation).

For the vast majority of areas, consistency between the distribution of the MGS targets and the PF and SF definitions has been met. However, as we shall demonstrate, there are a number of areas of both the PF and SF where consistency does not exist. The purpose of this paper is threefold: (1) lay out a methodology by which consistency checks can take place, (2) identify and correct for regions that have been erroneously included or excluded from the footprints, and (3) identify regions of the SF with starkly different spectroscopic completeness properties than the majority of the survey.

Much of the background information on SDSS geometry is provided in Section 2. In Sections 3 and 4 we reveal mistakes in the PF and SF, respectively. These mistakes can be characterized into three classes—regions that should have been included but were not, regions that should have been excluded, and unreported sampling anisotropies. In Section 5 we summarize these results and offer suggestions for improving future survey design, or at least the ways in which DRs are presented to the greater scientific community.

2. SLOAN DIGITAL SKY SURVEY GEOMETRY

Discovering and correcting errors in the SDSS footprints requires a firm understanding of the survey’s geometric properties. In this section we outline the key aspects of SDSS’s observational strategy as it relates to generating the survey footprints. Descriptions and visualizations of regions of interest are provided. We summarize the region algebra system used to define each region and to determine which points lie within it.

The SDSS contains a number of official region types, examples of which include TILE and SECTOR. Many of these region names have generic definitions in addition to their official designations. When referring to the SDSS regions specifically, their names will be fully capitalized for clarity. Extracting the geometric definitions of these regions is accomplished by querying the SDSS database. The SQL scripts used to do so are provided in Appendix A.

To visualize SDSS regions and MGS target distributions we utilize the VisIVO software package. VisIVO uses colored filters to represent points in three-dimensional space. It is optimized to allow desktop computers to display hundreds of millions of elements at a time. Our typical method of visualizing regions was to generate high density, uniformly distributed random points on the unit sphere, then filter them through the region definitions using the region algebra. The “noisiness” present in some of the images represents the limited
resolution of these simulations, but enables the shapes of overlapping regions to be illustrated more clearly.

We measured the areas of regions using a Monte Carlo process. A random number of points $N$ was generated within a known area $A$ surrounding the region. The number of points $n$ that also satisfied the region definitions was counted, and the region’s area was taken to be $nA/N$. Enough randoms were used to determine those areas to a precision of $<0.03\%$.

2.1. Observation Strategy

Of the many tables in the database, a few deserve special mention. The first of these is named PhotoObjAll. This table contains all objects’ imaging data, including unique identifiers, spatial information, fluxes, brightness profiles, when the data were collected, and more. The second table of interest is SpecObjAll, which contains all the spectroscopic information gathered during the survey including, unique identifiers, links to the objects’ photometric data, redshifts, and spectral classification. The HalfSpace table contains the boundaries (i.e., geometry) of SDSS’s many regions. Several tables of derived photometric redshifts are available as well.

During the Legacy Survey, the SDSS’s main telescope used a drift scanning technique. As the sky rotated overhead, light cascaded onto the detectors, producing up to 200 GB of data in a single night. The telescope was positioned such that each of the six camera columns, or camcols, mapped the sky in great circles $10–12\text{arcmin}$ wide and up to $130^\circ$ long. These six nonoverlapping scanlines collectively formed a strip. A second strip, slightly offset from the first, filled in the areas between scanlines and formed a contiguous stripe of width $2^\circ.5$ (Gunn et al. 1998; Stoughton et al. 2002). Scanlines in adjacent strips overlapped slightly, producing multiple unique observations of objects within the overlapping areas.

Before spectra were taken, a target selection algorithm determined whether an imaged object qualified for spectroscopic observations. Each object type (e.g., MGS, LRG, QSO) had its own selection criteria. Once targets were determined, fibers aligned along two slitheads transmitted light from the focal plane of the telescope to the spectrographs. Each spectrograph could accept a maximum of 320 $3''$ diameter fibers for a total of 640 (Gunn et al. 1998; Uomoto et al. 1999).

At the focal plane, the fibers were manually inserted into holes in a circular disk known as a tile. Each tile was a 1 m diameter, $1/4$ inch thick circular disk of aluminum which, once inserted into the main telescope, subtended an angular radius of $1^\circ.49$. The position of each hole/fiber on the tile corresponded to the position of a spectroscopic target. Because each region of the sky offered its own unique set of targets, a couple thousand tiles needed to be manufactured. The holes were drilled off-site and the tiles were then transported to Apache Point where the fibers were subsequently inserted.

Of the 640 available fibers, 48 were reserved for observing sky backgrounds and spectrophotometric standards, leaving 592 fibers for spectroscopic targets. Due to the size of their fiber claddings, no two fibers on a single tile could be positioned closer than $55''$ to one another. This constraint is frequently referred to as fiber collisions. The effect of fiber collisions was partially mitigated by the fact that tiles’ footprints were permitted to overlap. This allowed pairs of objects with separations less than $55''$ to be spectroscopically observed as long as their respective fibers were placed on separate tiles.

Obtaining spectra was both financially expensive and time consuming. Not every target selected for spectroscopic observation could ultimately have its spectra taken. Deciding which targets would be assigned fibers was accomplished through the use of the tiling algorithm of Blanton et al. (2001). Using a set of photometric criteria (Strauss et al. 2002), certain objects were identified as targets worthy of spectroscopic observation. Targets were identified within a rectangular area called a chunk and prioritized based on type. Brown dwarfs, by virtue of their rarity, received the highest priority followed by QSO’s, then MGS galaxies and LRGs. The latter two types received equal treatment, meaning that when fiber placement between the two came into conflict, the one whose spectrum was measured was determined at random.

Once targets were selected, tiles were projected nearly uniformly on the sky. To determine which targets to assign to...
each tile, a maximal set of targets separated by at least 55° was identified. This became known as the *decollided set*. Next, the algorithm iteratively perturbed the centers of the tiles with the goal of maximizing the number of quality spectra. A cost function was applied and minimized until an optimal orientation was found that assigned fibers to >99% of decollided targets.

To distinguish between types of MGS candidates, we employ a specific lexicon. All objects that meet MGS photometric criteria will be referred to as MGS *targets*. All

---

**Figure 3.** Visual representation of PRIMARYs belonging to the CHUNKs in Figure 2. Top: a CHUNK (cyan) encloses its PRIMARY (purple). A PRIMARY’s area is always less than or equal to the area of its CHUNK. Bottom: the non-overlapping PRIMARYs are visualized in random colors.

**Figure 4.** Visualization of DR6’s 111 PRIMARY regions. No PRIMARYs overlap. Each is assigned a random color to distinguish it from its neighbors. An average of two to three PRIMARYs compose each STRIPE.
targets that have their spectra measured will be referred to as MGS galaxies, while those that do not will be referred to as MGS objects. When used in this context, targets, galaxies, and objects will be emphasized for clarity. The query used to extract the MGS is included in Appendix A.

2.2. Geometry

This section reviews the geometric properties of the SDSS photometric and SFs. In much the same way as a map of the United States can be described by the boundaries formed by state lines, congressional districts, electric interconnections, lakes, and many more, the SDSS footprints are defined in terms of regions, of which there are over 20 types.

The first part of this section focuses on the extent of the PF and the sets of regions that comprise it. Some of the more common photometric regions types are STRIPEs, STRIPs, and CAMCOLs, yet there are many second-order regions of considerable importance. Some define the difference between what was intended to be observed as part of the Legacy Survey versus what was actually observed. Another region type, PRIMARY SEGMENTs, are not explicitly defined in the database but play a prominent role in the propagation of errors due to photometric zero-points.

Next, we review the main principles that define the SF. These include TILEs, which are the spatial manifestations of the tiling algorithm, and SECTORs, which are formed by the thousands of Venn diagram-like intersections of TILEs and other masks. We conclude with an explanation of halfspaces, which both define regions and facilitate rapid searches within them.

2.2.1. Photometric Geometry

The total region of the sky imaged by SDSS is the FOOTPRINT, or to distinguish it from spectroscopic coverage, the PF. By virtue of the SDSS’s drift scanning approach, the PF is the union of areas enclosed within sets of great circle segments. These circles are defined with respect to an imaginary axis that passes through two stationary poles at (R.A., decl.) = (95°, 0°) and (275°, 0°).

During a single drift scan, each of SDSS’s six cameras observed its own, nonoverlapping scanline covering an abstract region known as a CAMCOL. The union of those six CAMCOLs is a region known as a STRIP. The region covered by two (slightly overlapping) interlocking STRIPs is known as a STRIPE.

STRIPEs are separated by 2.5° and span from pole to pole, though in practice observations never spanned this full distance. Each is defined and indexed by its inclination relative to the equator such that a STRIPE of index \( n \) has an inclination of \( -25 + 2.5n \). For example, STRIPE 10 lies along the equator and STRIPE 11 lies 2.5° above it in the northern hemisphere. The highest possible latitude STRIPE is at \( n = 46° \), or 90°, though only STRIPEs 1 through 45 are formally defined in the database. In DR6 there are also three STRIPEs—76, 82, and 86—defined in the southern hemisphere.

STRIPEs have their own coordinate system known as “great circle coordinates” (Stoughton et al. 2002). If the center line of each STRIPE (1.25° from each boundary) acts as its own equator, then the SDSS coordinates \( \mu / \nu \) act as R.A./decl. for this truncated region of the sky. While STRIPEs are, in principle, abstract regions spanning pole to pole, they are not defined as such within the database. Rather, they are assigned mu limits that more closely align the abstract ideal of a STRIPE to what was actually observed. The definitions of STRIPs and CAMCOLs are similarly limited.

There were instances in which the actual survey geometry differed from the idealized survey geometry. For example, in the early DRs there were slight deviations of a few arcseconds in latitudinal pointing from what was planned. In other cases,
the SDSS telescope concluded observing before or after reaching the limit of a STRIPE.

The SDSS database reports the true survey geometry through regions called CHUNKs. While there were only 48 STRIPEs intended for DR6, there are 111 CHUNKs, meaning the average STRIPE was “broken up” into two to three separate pieces during observing runs. If the survey had been conducted “perfectly,” then CHUNKs and STRIPEs would have been identical.

Because STRIPEs are rectangular objects projected onto a sphere, they begin to overlap as they approach the poles. Similarly, CHUNKs overlap as illustrated in Figure 2. Targets observed within CHUNK overlap regions are usually imaged at least twice, once for each scan of the region. Once CHUNKs are resolved by the pipeline, objects are assigned primary or secondary status depending on what side of the line bisecting the CAMCOLs it falls.

The region that exclusively contains a CHUNK’s primary objects is referred to as a PRIMARY. In DR6 there are 111 CHUNKs and therefore 111 PRIMARYs. Each pair shares a unique chunkID, which can be found in the Segment table. No CHUNK’s area beyond the limit of its corresponding STRIPE is permitted to lie within a PRIMARY region. A visualization of CHUNKs and PRIMARYs that extends the example of Figure 2 is provided in Figure 3. Figure 4 offers a full sky view of all DR6 PRIMARYs. Our investigation focuses exclusively on targets with primary status (i.e., those within PRIMARYs), therefore nothing outside the union of PRIMARYs will be considered here. Hereafter, the term PF should be considered synonymous with the union of all PRIMARYs.

Just as each STRIPE is comprised of 12 CAMCOLs, each CHUNK is comprised of 12 SEGMENTs. In this way, SEGMENTs can be thought of as the realized observations of the abstract CAMCOLs. There are 48 STRIPEs, which means that under ideal observing conditions only $48 \times 12 = 576$ distinct CAMCOLs would exist.

Of course, ideal observing conditions are the exception rather than the rule. Due to effects such as the deterioration of seeing conditions during the night, full STRIPEs were rarely observed in a single run. The complete imaging of DR6 required 171 runs, which created $171 \times 12 = 2052$ SEGMENTs as illustrated in Figure 5. As with CHUNKs,
CAMCOLs are redefined in the database such that their angular limits match those of their corresponding runs.

Just as PRIMARYs are the non-overlapping portions of CHUNKs that contain primary observations, PRIMARY SEGMENTs are the non-overlapping portions of SEGMENTs that contain the same. Figure 6 offers a visual description of how SEGMENTs are cropped to form PRIMARY SEGMENTS. Unlike the other regions previously described, PRIMARY SEGMENTs are not explicitly defined in the database. Instead, their geometric properties are derived by applying the PRIMARY constraints atop the SEGMENT definitions (see Appendix A). The union of PRIMARY SEGMENTs occupies a combined area of 8304 deg², or 98.7% of the 8417 deg² covered by the full PF (Adelman-McCarthy et al. 2008).

2.2.2. Spectroscopic Geometry

The area of the sky observed with a physical metal tile is referred to as a TILE region. While tiles can only be inserted into the spectrograph one at a time, TILES may overlap to increase the effective density of available fibers (Blanton et al. 2003). The number of TILES overlapping an area of the sky is referred to as that area’s depth. Greater depth generally implies greater spectroscopic completeness.

Multiple TILES are generated during each tiling run. Such runs are contained within tiling masks. The area within the tiling boundaries but outside the tiling masks is referred to as the tiling region. DR6 has 1520 TILES with regionID’s between 1839 and 3358, as shown in Figure 7. The gap in the equatorial declinations of the northern galactic cap represents an area that was imaged as of DR6, but not yet tiled. Those regions were “filled in” with other TILES during DR7.

SECTORs are non-overlapping intersections of TILES. Under the simplest circumstance a SECTOR would be a single circle corresponding to its TILE. In practice, the application of tiling boundaries, tiling masks, and intersections with other TILES creates thousands of additional intersections, each one of which is its own SECTOR. A visual example of the SECTORs within a randomly selected TILE is provided in Figure 8. There are 9464 distinct SECTORs defined within the DR6 database. Each is provided its own regionID. In DR7, new SECTORs were introduced as spectroscopic observations continued. This enlargement of the SF did not change the definitions of any of the DR6 SECTORs, but it did change their unique indices.

Because spectroscopic observations only occur within SECTORs, hereafter SF should be considered equivalent to the union of all SECTORs. As verified by Figure 9, the SF lies entirely within the boundaries of the PF. There are also a considerable number of “holes” in the SF as compared to the TILE footprint. This is due to a number of effects, including the introduction of tiling masks and differences between the intended and realized spectroscopies.

Figure 7. Spatial representation of the TILES defined within the SDSS DR6 database. Each circular TILE is assigned a color based upon its regionID. The order of the regionID’s does not convey the order in which spectroscopic observations were conducted within and between data releases.
2.2.3. Region Algebra

While the SDSS uses a few different methods to describe a region’s geometry, we exclusively utilize the so-called constraint conditions. The basic idea is that regions like SECTORs and SEGMENTs have multiple sides, each of which can be considered a constraint. Any object that satisfies all of a region’s constraints must lie within it. (For more, read Gray et al. 2004.)

SDSS treats each constraint as a planar intersection of the sky’s unit sphere, as in Figure 10. The resulting small or large circle is described by four parameters: the three Cartesian components of \( \hat{n} \), the unit vector that points toward its center, and \( c = \cos \theta \) where \( \theta \) is the circle’s angular radius. This area is referred to as a halfspace since the plane divides three-dimensional space in half.

A point \( \hat{x} \) on the unit sphere lies within the halfspace if \( \hat{n} \cdot \hat{x} - c > 0 \), an inequality referred to as a constraint condition or a halfspace constraint. Circles with small angular radii have \( c \approx 1 \). With great circles, \( c = 0 \). Halfspaces with \( c < 0 \) correspond to areas greater than a hemisphere. Each constraint condition is uniquely identified in the database and defined through the four-vector \( [n_x, n_y, n_z, c] \).

More complicated areas are created by intersecting multiple halfspaces. In general, these intersections are called convexes. For example, a SEGMENT is a convex with four halfspace constraints (i.e., four sides). A point that simultaneously satisfies all four of those constraints lies within the boundaries of the SEGMENT. Convex constraints are extracted from the SDSS database’s Region, RegionConvex, Segment, and HalfSpace tables.

The halfspaces and their associated inequalities comprise a region algebra and provide a convenient framework for determining whether or not an object occupies a region. In general, for a point \( \hat{x} \) to lie within a convex with \( m \) constraints,

\[
\hat{n}_i \cdot \hat{x} - c_i > 0 \quad \forall \ i = 1, 2, \ldots m.
\]

Note that this region algebra does not require trigonometric functions, but only relatively inexpensive dot products. The increase in speed this algebra provides was indispensable for efficiently executing many of the simulations discussed in the pages to come. An example that uses four constraints to represent PRIMARY 208 is shown in Figure 11.

SEGMENTs, PRIMARYs, CHUNKs, and TILEs are both regions and convexes, while SECTORs are regions formed from the union of one or more convexes. Each of a SECTOR’s convexes received a convexid in the DR6 HalfSpace table that ranges between 0 (the 1st convex) and 11 (the 12th convex). A point lies within a SECTOR as long as it satisfies all of the constraint conditions of any of its convexes.

We remind the reader that the original DR6 PF occupies an area of 8417 deg\(^2\). However, the PF analyzed in these pages is defined to be the union of PRIMARY SEGMENTs. Our new PF is created by applying the PRIMARY constraints atop the original PF definition. This process reduces the original PF by approximately 113 deg\(^2\).
In this section, we discuss errors found in the DR6 PF and the steps taken to remedy them. Our detection method involved filtering millions of uniformly distributed angular random points into sets that resided in regions defined to be PRIMARY SEGMENTs. By superimposing the positions of MGS targets, it was possible to locate regions where the survey geometry was inconsistent with the observations. This process yielded a number of discoveries, including five areas that ought to be removed from the PF and an area where PRIMARY SEGMENT definitions overlap. We conclude this section by quantifying the effect these errors have on the numbers of galaxies expected within volumes on different length scales and on the galaxy overdensity power spectrum. We find that failure to correct these errors leads to systematic underestimations of overdensities within these volumes and an amplified variance in power on large scales.

3.1. Locating and Correcting Footprint Problems

Figure 12 displays four circular TILEs superimposed upon a sea of targets. A closer view of this area is shown in Figure 13. The areas boxed in red lie within the union of PRIMARY SEGMENTs, yet contain no targets. While an absence of targets does not necessarily indicate a problem with the PF, there are three observations in this case that strongly suggest an error in the six PRIMARY SEGMENT definitions. First, given the ambient surface density of targets, the probability that areas of this size would be empty due to cosmic variance alone is very low. Second, the shapes of the empty regions align perfectly with the SDSS geometry. Each boxed region corresponds to a single SEGMENT. These regions comprise a group of six within a single STRIP. Finally, these six areas are missing from the SF, suggesting that whatever caused the lack of targets was reflected in the SECTOR definitions, but
not the SEGMENT definitions. Together, these observations provide compelling evidence that the six SEGMENTs in Figure 13 were included in the PF by mistake. We refer to their union as “P1.”

The PF can be corrected if the constraint conditions that define the edges of P1 can be identified. Then, points that lie within the union of PRIMARY SEGMENTs could be filtered through six additional searches over these SEGMENTs. Any points that lie inside any of those six regions would be summarily classified as residing outside the PF. The constraint conditions for the edges of footprint errors are not always defined in the DR6 geometry. In our experience, and for reasons elaborated upon in Section 4.2, it is preferable to define the boundaries of troublesome regions using constraints reported in the DR7 region definitions.

The first step in defining P1 is identifying the TILEs surrounding the six SEGMENT portions. Once approximate angular limits of the TILEs are found (lines of constant R.A. and decl. can be generated by the user and superimposed onto one’s visualization), a query such as the following can pick out the region IDs of the colored TILEs:

```
SELECT *
FROM BestDR7.dbo.fRegionsContainingPointEq
(126, 42.5, “TILE”, 0)
```

The goal is to discover the smallest set of TILEs whose interiors contain all boundaries of these six SEGMENTs. We recommend that the user pre-compute for each TILE a table that contains a list of its SECTORs and their definitions. This table should be organized such that one’s visualization software is able to display SECTORs individually, as in Figure 14. Here SECTOR 92487, colored in blue, is revealed to share the same boundaries as the bottom and right edges of the top SEGMENT in P1.

The other boundary conditions are identified by examining SECTORs within TILE 2499. Figure 15 shows that the bottom edge of SECTOR 90846 is the same as the upper edge of the top SEGMENT. SECTOR 91521, pictured in Figure 16, possesses a comb-shaped border that traces every other boundary in P1. Once the minimum number of SECTORs that share all of the SEGMENTs’ boundaries are singled out, their
Figure 12. Wide view of P1. Starting at the top and going clockwise, the colored circles are DR7 TILES 2660, 2500, 2818, and 2499. They are superimposed atop MGS targets, represented by pink pixels, that indicate the extent of the DR6 PRIMARY SEGMENT footprint. The curved lines mark R.A.’s of 145° and 155° (left to right) and decl. of 55° and 60° (bottom to top).

Figure 13. Close view of P1. The colored circles contain the portions of DR7 TILES 2660, 2500, 2818, and 2499 that lie within the DR6 spectroscopic footprint. The positions of MGS targets are marked by pink pixels. The areas boxed in red are the portions of six SEGMENTS within the PRIMARY SEGMENT footprint that contain no targets.
Each constraint condition is drawn from a pre-computed table of each SECTOR’s halfspace constraints.

Each constraint marks the intersection of a great or small circle with the unit sphere. Points lying on one of the sides of a constraint occupy at least a full hemisphere and can be computationally expensive to search over. To speed up computations, the best method is to perform a preliminary filtering, perhaps by limiting points to those within a single TILE, and then applying the halfspace constraints one at a time.

Ultimately, 24 constraint conditions are needed to define P1, or four for each of its six SEGMENTS. (Because these SEGMENTS share the same left and right boundaries, only 14 of these constraints are unique, however.) The aim is to report each SEGMENT’s constraints such that any point that satisfies all four must lie within it.

Figure 17 illustrates one such constraint for the upper boundary of the fourth SEGMENT from the top. Points that satisfy this constraint are colored in blue and lie on the interior side of the SEGMENT, as desired. If the points had lain on the opposite side of the boundary, all four constraint components \([n_x, n_y, n_z, c]\) would have been multiplied by \(-1\) to flip the condition.

### 3.2. Census of PF Errors

The last section offered a detailed example of how to locate, identify, and characterize area P1. In this section, we provide a census of four other problem areas and one “ambiguous STRIP” located in the SDSS’s southern hemisphere.

In Figure 18 we reveal the location of P2. A zoomed-in version is shown in Figure 19. The combined area of these regions is sufficiently small that cosmic variance could plausibly explain the absence of targets; however, these five rectangular areas are also missing from the union of SECTORS, so we find it more likely that they reflect a problem with the PF.

The location of P3 is revealed in Figure 20. A zoomed-in version is shown in Figure 21. Unlike P1 and P2, this area lies outside the SF, making it impossible to use SECTOR constraints to define its shape. Instead, DR6 SEGMENT definitions are used to find the boundaries of the long edges, i.e., those that run roughly parallel to the lines of R.A. The one exception is SEGMENT 1770’s extreme edge, which is bounded by PRIMARY constraint condition 1225.

The lower declination side of P3 is bounded by the edge of a DR6 PRIMARY given by constraint condition 1228. The database does not appear to contain any constraint for the opposite side, so it must be approximated by trial and error. This edge is roughly parallel to condition 1228, so a modification of its \(c\) parameter is sufficient to shift the boundary. We found the appropriate constraint four-vector to be:

\[
\begin{align*}
    n_x &= 0.122972102, \\
    n_y &= -0.720394564, \\
    n_z &= -0.682575662, \\
    c &= 0.021.
\end{align*}
\]

Because P3 lies so far away from the SF, the area is unlikely to be sampled in a galaxy clustering analysis that requires targets in a region to possess redshifts. From this perspective, P3 is the least troublesome of PF errors.

The location of P4 is circled in Figure 22 and shown in detail in Figure 23. The location of P5 is shown in Figure 24. Both were visually identified relatively easily by finding areas without targets that coincided with holes in the SF.

To summarize, area P3 is defined by DR6 SEGMENTS 1766–1770 and the constraint four-vector given above. The geometric descriptions of the other four areas are fully provided by the constraint conditions present in the following DR7 SECTORS:

- P1: 90846, 92485, 92487
- P2: 91430, 91439, 91440, 92430, 92438
- P4: 92220, 92658, 92673
- P5: 85193, 85549, 85560, 85563, 85642, 85654.

We refer to the final area of the PF that requires correction as the “ambiguous STRIP.” It is pictured in Figure 25. PRIMARY SEGMENTS are defined to be strictly nonoverlapping, but we see here that SEGMENTS 5344–5349 and SEGMENTS 6874–6879 do overlap at the edge of PRIMARY 308.

The STRIP complementary to SEGMENTS 6874–6879 does contain galaxies, yet is undefined in both the photometric and spectroscopic region geometries. This suggests that its
omissions from the SECTOR and SEGMENT definitions are errors. There are a couple possible explanations. The first is that the existence of SEGMENTs 6874–6879 is a mistake, meaning that SEGMENTs 5344–5349 are defined correctly, but its complimentary STRIP either does not extend far enough (i.e., all the way to the next PRIMARY) or it was legitimately truncated early. Another possibility is that SEGMENTs 6874–6879 are real and SEGMENTs 5344–5349 extend too
far beyond their true boundary. Either way, this introduces significant ambiguity regarding what is actually happening in this area. If one’s research concerns the effect of photometric zero-points, for example, this mangling of region definitions complicates efforts to handle targets within them with fidelity.

One conservative solution would be to consider SEGMENTS 6874–6879 as real and manufacture its six complimentary SEGMENTS. While these 12 SEGMENTS might truly belong on the edge of PRIMARY 308, splitting them off would merely reduce the statistical knowledge that could be gained by knowing that the measured magnitudes in the adjacent regions are correlated.

We ultimately decided to remove the regions covered by SEGMENTS 6874–6879 from the PF entirely. This slightly reduces its area but also removes any worry that the objects in this region should have been excluded for a legitimate reason. The contiguous area covered by this portion of PRIMARY 308 is defined by constraint conditions 8466, 15572, 15573, and 15574.

Table 1 summarizes the sizes of each region removed from the PF. The sum of P1–P5 plus the ambiguous STRIP is 19.7474 deg$^2$. This is 0.237807% of the original PRIMARY SEGMENT footprint and 0.047872% of the total sphere. In the end, the area of our improved PF becomes 8284.21 deg$^2$.

---

**Figure 17.** Illustration of a constraint condition that defines the upper boundary of one of P1’s SEGMENTS. Points that lie within the union of DR7 TILE 2499’s SECTORs and satisfy the constraint condition are colored in blue. Were these points not confined to TILE 2499, they would fill an entire hemisphere. Clockwise from the top, the union of SECTORs within DR7 TILES 2600, 2500, 2818, and 2499 are also pictured.

**Figure 18.** Location of P2 circled in red. The approximate position of the region is [R.A., decl.] ≈ [152°, 58°]. MGS targets are represented by magenta pixels.

---
Figure 19. Close-up view of P2. The five PRIMARY SEGMENT portions that require removal from the PF are boxed in red. MGS targets are represented by white pixels.

Figure 20. Location of P3 circled in red. The region’s position is [R.A., decl.] ≈ [269°, 47°] where the curved lines are decl. of 45° (bottom) and 55° (top). The DR6 improved spectroscopic footprint is colored in cyan. MGS targets, which mark the true extent of the PF, are colored in magenta.
Figure 21. Close-up view of P3. PRIMARY SEGMENTS that require removal are boxed in red. These are portions of DR6 SEGMENTS 1766 (lower left) through 1770 (upper right). MGS targets, represented by white pixels, are conspicuously absent from these regions. The boundary of the PRIMARY SEGMENT footprint is outlined in blue. Galaxies outside this boundary are secondary targets and are not considered in our analysis.

Figure 22. Location of P4 circled in red. The approximate position of the region is [R.A., decl.] ≈ [255°, 37°]. MGS targets are colored in magenta.
reduction of about 133 deg$^2$ or 1.58\% from the reported value of 8417 deg$^2$ (Adelman-McCarthy et al. 2008).

3.3. Impact on Expected Number Count

We quantify the impact of PF errors on galaxy clustering statistics as a function of scale. To do so, we populate the volume of the DR6 SF with closely packed spherical cells of radii 7, 11, and 16 $h^{-1}$ Mpc (hereafter R7, R11, and R16) using the hexagonal closest packing (HCP) arrangement (Conway & Sloane 1993). The volume is defined within the redshift limits $\zeta \in [0.02, 0.22]$ for R7 and R11, and $\zeta \in [0.02, 0.30]$ for R16. To ensure that at least half of the targets in cells (on average) are galaxies, we require that $\beta_{\text{spec}}$—the fraction of a cell’s volume within the SF—is at least 62\%. The radial distribution of galaxies within the SF is used to parameterize the Schechter luminosity function, which in turn is integrated over all absolute magnitudes between $M_{\text{min}}(\zeta)$ and $M_{\text{max}}(\zeta)$ to yield the radial selection function $S(\zeta)$. Assuming a standard flat cosmology where $\Omega_m = 0.3$, $\Omega_k = 0$, $\Omega_{\Lambda} = 0.7$, and $h = 0.7$, the selection function yields $\langle n \rangle$, or the expected number of galaxies in a cell in the absence of clustering.

By assuming that galaxies can exist in areas P1–P5, the expected number of galaxies $\langle n \rangle$ within cells that intersect those areas will be overestimated. In turn, the overdensity $\delta$ will be underestimated. This type of error is bound to be most pronounced in high-redshift cells with small angular radii. However, low-redshift cells can potentially intersect multiple PRIMARY SEGMENTS in the corrected areas so the effect is worth examining empirically.

We begin by letting $\beta_{\text{ps}}^0$ equal the fraction of a cell’s volume that lies within the union of PRIMARY SEGMENTS and $\beta_{\text{ps}}$ equal the fraction of a cell’s volume within the improved PF. The number of expected galaxies will be overestimated by the fraction $(\beta_{\text{ps}}^0 - \beta_{\text{ps}})/\beta_{\text{ps}}$.

All cells that intersect areas P1–P5 were identified, and using a Monte Carlo method $\beta_{\text{ps}}^0$ and $\beta_{\text{ps}}$ were estimated for each. We present the fractional overestimates of $\langle n \rangle$ in Figure 26. Because no cells intersect P3, it is omitted from the Figure. The ambiguous STRIP is likewise omitted since its problem is not

![Figure 23](image.png) Close-up view of P4. The four SEGMENT portions that require removal from the PF are boxed in red. MGS targets are represented by white pixels.

![Figure 24](image.png) View of P5. The union of the four rectangular regions that require removal from the PF are outlined in red. MGS targets are represented by magenta pixels. The left and right edges of P5 lie at [R.A., decl.] $\approx [230^{\circ}.5, 30^\circ]$ and [250^{\circ}.5, 19^\circ] respectively.
an absence of expected galaxies, but rather an ambiguity regarding PRIMARY SEGMENT definitions.

The error in \( \langle n \rangle \) can be significant, exceeding 60% for some R7 cells. The maximum possible error tends to increase with redshift since the areas of the most distant cells decrease while the angular extent of P1–P5 remains fixed. Due to its larger area, the errors are greatest in P5. By virtue of their size, R16 cells have smaller fractions of their volumes affected by P1–P5. It follows that their average fractional error in \( \langle n \rangle \) is smaller than that of the R7 cells.

More relevant to the estimation of cosmological parameters is the MGS overdensity power spectrum. Using a fiducial signal model, we simulate overdensities in the R7, R11, and R16 cells defined above. We represent the power before the PF corrections as \( P(k) \). The differences between these spectra on all three scales are plotted in Figure 27.

PF errors do not have an appreciable effect on small scales. On large scales, the footprint errors introduce an additional variance. From our simulations, it is not clear whether the PF errors amplify or suppress power. The 1\( \sigma \) error bars are consistent with zero change in the power. However, it is clear that the errors introduce a variance in the large-scale power that would not have existed otherwise.

4. SF CORRECTIONS

As the SDSS progressed, new TILEs were periodically placed. The intersections between new and existing TILEs created hundreds of new SECTORs at a time, many obtaining new spectroscopic properties. In this way, the SF became larger and more complex with each DR. Tiling runs for DR7 were anticipated during DR6. This "forward-looking" approach facilitated the continuous evolution of the survey, but also introduced inconsistencies at the times of DRs, including that of DR6.

For example, consider a DR7 TILE partially overlapping the southern portion of a DR6 TILE. The tiling algorithm could restrict fiber placement to the northern portion of the DR6 TILE with the expectation that the remainder would be filled in during DR7. The entirety of the DR6 TILE would be included within the DR6 SF even though the completeness was overwhelmingly nonhomogeneous. This effect, which was largely limited to the edges of the SF, could fool the user into believing that the DR6 footprint was larger than it actually is.

Assuming one was even aware of the existence of troublesome SECTORs, searches for them are not easy. Their region definitions are complicated and not well documented. Statistical tools to root them out might be developed, but settling on decision criteria is challenging. The TILEs within which they reside still receive their full compliments of fibers.

---

Table 1 Adjusted PF Regions

| Region            | Area (deg\(^2\)) | Percentage of Original PRIMARY SEGMENT Footprint | Percentage of Total Sphere |
|-------------------|------------------|-----------------------------------------------|---------------------------|
| Original PRIMARY  | 8303.96          | 100.0000                                      | 20.1308                   |
| P1                | 1.9805           | 0.0239                                        | 0.00480                   |
| P2                | 0.5084           | 0.0061                                        | 0.00123                   |
| P3                | 0.7008           | 0.0109                                        | 0.00170                   |
| P4                | 1.3334           | 0.0521                                        | 0.00323                   |
| P5                | 10.3819          | 0.0703                                        | 0.00251                   |
| Ambiguous STRIP   | 4.8419           | 0.0675                                        | 0.00174                   |

Note. The areas of the ambiguous STRIP and P1–P5 were derived empirically using \( 3.85 \times 10^8 \) full sky angular randoms filtered through each region.
Searching for empty regions alone is insufficient because troublesome SECTORs were routinely assigned a nonzero number of fibers, though certainly not enough to achieve representative spectroscopic completeness. Moreover, it is difficult to distinguish between regions that are systematically undersampled and those that are legitimately underdense.

The DR6 SF is defined to be the union of all DR6 SECTORs. However, there are hundreds of individual regions...
where this simple definition fails. This section addresses the challenge of repairing the SF to the point where its completeness is not subject to gross undersampling biases or incorrectly placed boundaries.

We begin by identifying five regions that contain galaxies but lie outside the union of SECTORs. We provide constraint conditions to reintroduce these areas to the footprint. Next, we show how a large, low declination CHUNK and a SEGMENT portion made it into the union of SECTORs even though no galaxies lie within them. We continue by showing how strongly undersampled regions of the SF are defined, located, and removed. We conclude by visualizing the improved SF (ISF) and reporting the stark statistical differences between targets inside its area and those trimmed from it.

4.1. Inclusion/Exclusion Regions

There are five areas of the PF that contain galaxies but which were not included in the SF. All five lie in the northern hemisphere on the lowest declination edge of the PF. We label these alphabetically A through E, and display them in Figures 28 and 29. The presence of galaxies in these regions proves that they were spectroscopically observed and therefore deserve to be included in the SF. It is possible to reintroduce them using the constraint conditions supplied in Table 2.

| Area | Constraint 1 | Constraint 2 | Constraint 3 | Constraint 4 |
|------|--------------|--------------|--------------|--------------|
| A    | −33111       | −33094       | 33101        | 5264         |
| B    | 269          | 33111        | 5252         | 5251         |
| C    | −33111       | −33094       | 5265         | 33092        |
| D    | 269          | 270          | 5255         | 5256         |
| E    | 269          | 270          | 5259         | 5260         |

Note. DR6 constraint conditions that define regions A through E. A negative sign indicates that the \( [x, y, z, c] \) coordinates had their signs reversed to ensure the constraint was on the correct side of the halfspace. Points that satisfy all four of an area’s constraints lie within that rectangular region.
The most prominent of these regions is CHUNK 113, which appears as a cyan-colored rectangle in a low-declination region of Figure 30. The next largest areas appear near the edges of the SF and are addressed separately in Section 4.2. The tiniest of these areas appear as cyan speckles in the survey's interior. These result from the small areas between TILEs and are an expected byproduct of the SDSS survey strategy and tiling algorithm.

The second area we remove from the SF is the low-R.A. portion of DR7 SEGMENT 5417. This region is pictured bounded in red in Figure 31. The red pixels mark the presence of targets. None of these on the left side of the pictured DR7 TILE had their spectra taken, indicating that their areas should not be part of the corrected SF. This region is defined by the intersection of SEGMENT 5417 with any of the following DR7 SECTORS: 39201, 39205, or 39212.

Finally, the portion of the STRIPE spanned by SEGMENTs 6874–6879 in Figure 25 is removed. Because the SF is a subset of the PF, areas removed from the latter must also be removed from the former. The ambiguous, overlapping region definitions that prompted this removal were covered in Section 3.1.

4.2. Undersampled SECTORS

The SF can be separated into two kinds of areas. The first, which lies mostly in the interior of the survey, was observed in such a way that the percentage of MGS targets lacking spectra is approximately 6% for those that are fiber collided (Strauss et al. 2002) and approximately 20% overall. The second, which lies mostly on the edges of the survey, is comprised of regions where further spectroscopic measurements were planned for DR7. Here, the percentage of targets lacking spectra routinely ranged between 50% and 100%. These two types of areas have vastly different statistical properties and deserve specialized handling.

An example of the second type of area is shown in Figure 32. The portion of the survey pictured lies on the edge of the DR6 SF. Most of the visible gaps in spectroscopic coverage were eventually filled in during DR7. Areas shaded in gray lie within the union of DR6 SECTORs yet contain no galaxies.

The shapes of these areas appear to be formed from the intersections of circular TILEs. Upon overlaying TILEs placed during DR7, we find this is indeed the case. The TILE in Figure 32 contains five DR7 SECTORs that clearly overlap the unsampled area. We emphasize that the boundaries of this area cannot be defined through the geometric descriptions provided in DR6. This indicates that research published using DR6 data prior to the release of DR7 would almost certainly have been unable to properly account for regions like these.
Through a tedious process of visually comparing the positions of galaxies against the union of DR6 SECTORs, we were able to identify 183 areas covered by DR7 SECTORs that appear to have been either unsampled or significantly undersampled during DR6. The removal of these regions resulted in what we refer to as the better SF.

The selection process operated according to a few principles. First, because undersampled regions were located exclusively near the edges of the DR6 spectroscopic survey, the majority of our attention was focused here. This limited view helped differentiate between regions that were undersampled and those that were merely underdense. Second, we overlaid suspect areas with DR7 TILEs and SECTORs. If the shapes of these areas visually overlapped the empty areas of the DR6 SF, the constraint conditions of the appropriate DR7 SECTORs were gathered so those SECTORs could be removed. Third, SECTORs that lay within contiguous “gray areas” as in Figure 32 were removed as a group. It is possible that some of the smaller SECTORs within that group were adequately sampled, but contained a paucity of galaxies through cosmic variance. We give such SECTORs no benefit of the doubt and assume that those in the “gray areas” were similarly treated by the tiling algorithm.

Fourth, there were borderline cases in which it was unclear whether an area was undersampled or just underdense. Under these circumstances we default toward removing SECTORs because (1) our motive in repairing the SF is to ensure that its entire area has consistent spectroscopic completeness and (2) measures of galaxy surface density are less sensitive to the exclusion of good SECTORs than to the inclusion of bad ones. The extent of MGS galaxies is vast, and while a handful may be unnecessarily excluded, this exclusion comes with the peace of mind that the remaining SF has relatively uniform sampling properties and can be treated similarly.

Once the initial 183 SECTORs were removed from the SF, we embarked on a second round of corrections using the MGS objects. Just as the absence of galaxies reflects undersampled regions, so too does an overabundance of objects. In a uniformly sampled survey, the number density of objects averaged over sufficiently large areas should be relatively constant. Yet as we see in Figure 33, numerous areas of the better SF—largely along its edges—have uncharacteristically high object densities. Careful exploration for these regions uncovered an additional 120 DR7 SECTORs that we designated for removal. In total, the areas covered by 303 DR7 SECTORs were removed to produce our final product—the improved spectroscopic footprint (hereafter ISF). A complete list of the removed DR7 SECTORs is given in Appendix B.

4.3. Improved Spectroscopic Footprint

In summary, the final DR6 ISF is created through the following steps:
1. Include the union of all DR6 SECTORs,
2. Include the areas in regions A, B, C, D, and E,
3. Remove CHUNK 113,

Figure 32. Illustration of undersampled SECTORs in the DR6 spectroscopic footprint. The red lines mark the boundary of the union of DR6 SECTORs, while the blue pixels indicate the positions of MGS galaxies. In the top panel, select regions within the SF that contain no galaxies are shaded in gray. In the bottom panel a DR7 TILE is superimposed. Five of its SECTORs overlap the unsampled area in the SF. The region pictured lies approximately in the range R.A. ∈ [205°, 220°] and decl. ∈ [25°, 35°].
4. Remove the area within the intersection of DR6 SEGMENT 5417 and DR6 SECTOR 39201 or 39205 or 39212 (all yield the same result),
5. Remove the STRIPE containing DR6 SEGMENTS 6874–6879, and
6. Remove the area in all 303 undersampled DR7 SECTORs.

The improvements to the SF reduce its area from 6860 deg^2 (Adelman-McCarthy et al. 2008) to 6621.3 ± 1.8 deg^2. This constitutes an adjustment of about 239 deg^2, or 3.6%, from the commonly reported value. Two perspectives of the new footprint are provided in Figure 34.

Table 3 summarizes the impact that improving the SF has on the number counts and densities of MGS targets within its boundary. For this table we separate the MGS galaxies into two mutually exclusive sets. The first, “low-quality galaxies,” are targets with spectral measurements of poor quality. To fall into this set, a galaxy must have a redshift confidence $z_{\text{conf}} < 0.9$, or have a $z_{\text{Status}}$ of cross-correlation with low confidence, redshift determined from em-lines with low confidence, or redshift determined “by hand” with low confidence. The second set contains all other galaxies.

We find there are substantial differences in the distributions of galaxies and objects inside the ISF and in the regions discarded from the original SF. The trimming process reduced the total number of galaxies by 5045 (approximately 1%) and the number of objects by 14,074 (approximately 12.3%). Overall, our 3.6% reduction in the SF area led a 2.2% increase in the spectroscopic completeness of the entire survey.

The number of galaxies per deg^2 is about 72 inside the ISF, but only about 20 in the discarded region. The low-quality galaxies have a nearly identical ratio; there are about 3.6 times fewer of them per unit area in the discarded region than in the ISF. The similarity of these ratios suggests that a connection between the quality of observed spectra and the spectroscopic completeness anisotropies near the footprint’s edges is unlikely.

The ISF is useful for those who wish to discretize space by counting galaxies in cells. For a given minimum $\beta_{\text{spec}}$, the absolute number of objects within the cells’ projections has decreased. Trimming the footprint removed almost three times more objects than galaxies, even though the latter are four times as prevalent overall. This improvement reduces the uncertainty in galaxy counts per cell (particularly for those cells

---

4 The fields $z_{\text{conf}}$ and $z_{\text{Status}}$ can be found in the database table SpecObj.
on the survey’s edge) without eliminating an undesirably high amount of useful redshift information.

The final distributions of MGS targets within the ISF are illustrated in Figure 35. We note that the most significant anisotropies due to the tiling algorithm have been removed. This is not to say that our resulting footprint is perfect. There still exist some small border regions where objects appear systematically overdense. The same holds for a low-R.A., high-decl. CHUNK in the northern hemisphere. However, anisotropies in these regions are relatively minor and we feel that they do not warrant the removal of what would end up being a large chunk of the northern sky.

5. SUMMARY AND DISCUSSION

In this article we have explored numerous issues related to the SDSS photometric and SFs. In short, imaged objects must lie within the PF, and all areas of the PF must have been imaged. Similarly, all spectroscopically observed galaxies should lie within the SF, and all areas within the SF must have had fibers allocated to them. By comparing the distribution of MGS targets against the extent of the SDSS footprints, we were able to discover numerous inconsistencies.

This process was facilitated by the review of the SDSS geometry provided in Section 2. We expanded upon the explanations of SDSS regions provided in the literature and through SkyServer documentation. We did this by pairing definitions of idealized photometric regions (e.g., STRIPE, CAMCOL), realized photometric regions (e.g., CHUNK, SEGMENT), and spectroscopic regions (e.g., TILE, SECTOR) with unique visualizations. These original views of SDSS’s structure revealed problems that might have otherwise gone undiscovered.

Table 3
Spectroscopic Footprint Statistics

| Target Type          | Number of Targets | Number per deg² |
|----------------------|-------------------|-----------------|
|                      | Union of DR6 SECTORs | Improved Spectroscopic Footprint | Union of DR6 Spectroscopic Footprint | Discarded Spectroscopic Footprint |
| Galaxies             | 479, 419          | 474, 651        | 69.886                          | 71.69 ± 0.02                       | 19.975 ± 0.005                    |
| Galaxies (Low-Quality) | 22, 794          | 22, 517         | 3.3227                          | 3.4007 ± 0.0009                    | 0.9510 ± 0.0003                   |
| Objects              | 114, 038          | 99, 964         | 16.624                          | 15.097 ± 0.004                     | 58.961 ± 0.016                    |

Spectroscopic Completeness

|                          | 0.778             | 0.795            | 0.250                           |

24

Figure 35. Angular distribution of MGS targets within the ISF. Pictured are the MGS galaxies (upper left), MGS objects (upper right), and MGS low-quality galaxies (bottom).
Chief among these was the realization that key aspects of the DR6 geometry could only be adequately expressed using DR7 region definitions. Descriptions of the latter were indispensable, as they offered a way to distinguish whether regions were undersampled due to survey error or ordinary cosmological variance. Moreover, visually toggling between footprints and target distributions revealed that problems in one footprint were sometimes reflected in the other. This was particularly useful in correcting the PF, as many of its regions that required removal had already been deleted from the SF.

This was the case for P1–P5, four areas of the original PF that contained zero MGS targets, and which were ultimately removed from the improved PF. Area P3 met a similar outcome, though without the benefit of SECTOR constraint conditions. The “ambiguous STRIP,” a region of the southern hemisphere contaminated by overlapping region definitions (see Table 1), was deleted as a conservative measure. Removing these regions from the union of PRIMARY SEGMENTs reduced its area by 19.7 deg², or 0.24% of the total footprint. We learned that these problematic regions, despite their small aggregate size, could induce an error in the expected number of galaxies that exceeded 60% in cells of radius 7 h⁻¹ Mpc. Furthermore, we found that a failure to correct the PF introduces additional variance to the galaxy power spectrum on scales $k \lesssim 4 \times 10^{-2} h$ Mpc⁻¹.

A larger number of corrections were required to improve the SF. Five low-declination areas, which we labeled A–E, had been excluded from the SF even though they contained galaxies. Conversely, CHUNK 113 and a portion of DR6 SEGMENT 5417 had to be removed from the original SF since they contained no galaxies. The area covered by the “ambiguous STRIP” was also removed to maintain the requirement that the SF must be a subset of the PF.

In addition, the areas covered by 303 DR7 SECTORs were removed to reduce significant anisotropies in the survey’s spectroscopic completeness. The discarded regions, which lie almost entirely on the edges of the original SF, have vastly different MGS galaxy and object distributions as summarized in Table 3. Ultimately, a 3.6% reduction in the SF area led to a 2.2% increase in the spectroscopic completeness of the entire survey. Changes like these can dramatically increase the accuracy of overdensity measurements within cells on the survey’s boundary, without simultaneously removing an inordinate amount of useful spectroscopic data.

We recommend that the process of conducting visual consistency checks between survey geometry and target/galaxy distributions be extended to DR7, but especially to DR8. Because DR8 is the final official release of the Legacy Survey, optimizing its geometric definitions can produce a more definitive accounting of the MGS, LRG, and QSO samples.

Our findings highlight some of the difficulties that can arise when maintaining the geometric descriptions of large, active surveys. To avoid the sort of systematic errors described in this article, a protocol should be established by which changes made to one footprint are guaranteed to be reflected in the other, and efforts should be made to increase the transparency of such changes. It is easy to overlook errors in surveys as complicated as the SDSS. For this reason, survey operators should be mindful that the results of private discussions and deliberations, many conducted only through email or in-person conversations, should be made public if doing so would facilitate understanding and usage of the survey data. In addition, SQL queries to generate all region types (preferably with examples) ought to be provided.

Efforts to root out problems in a survey can be facilitated through a dedicated visualization interface. SDSS currently has the Finding Chart Tool,⁵ which is available through SkyServer. This tool offers a number of features, including the abilities to view a JPEG centered on a user-specified [R.A., decl.] and zoom in and out on that image. Photometric and spectroscopic objects can be identified, and the boundaries of bounding boxes, fields, masks, and plates can be overlaid.

However, the features of the Finding Chart Tool leave much to be desired. In developing a visual interface to a survey, we would suggest that the functions of such a tool could be determined by community input, similar to how the SDSS collaboration was originally asked to identify the top twenty questions they wanted the data to answer so that the CAS database could be structured in an optimal fashion.

In that same spirit, our recommendation would be to identify the most important “visual questions” that would need to be asked of the data. Such inquiries should include the ability to display objects as a function of type (e.g., MGS, LRG, stars, brown dwarfs, etc.) and DR. A user should be able to distinguish between targets with spectra and those without. Geometrically, one should be able to visualize the boundaries of each region within the survey. Because SDSS regions are defined with a relatively simple region algebra through the projection of small or great circles onto a unit sphere, visualizations such as these should not be prohibitively taxing on computing resources.

In addition to error identification, the ability to view overlapping fields (like object type and geometry) can offer insights that data processing alone might not reveal. At present, users may export region definitions and data to perform their own visualizations, but such an approach is time consuming, prone to errors, and will likely introduce inconsistencies between researchers.

Finally, we reemphasize that the footprint corrections made here were done with the Legacy Survey’s most populous sample—the MGS. The scope of future investigations should be widened to consider the impact on other sets, like the LRGs and QSOs.

APPENDIX A

SQL QUERIES FOR SDSS GEOMETRIC REGIONS

This appendix contains explicit SQL queries that can be used to reproduce the samples referenced within this article. Scripts to return the geometric properties for all SEGMENTs, PRIMARYs, PRIMARY SEGMENTs, TILES, and SECTORs are followed by the query used to extract the MGS.

A.1. SEGMENTs

The RegionConvex table identifies the SEGMENTs, and the HalfSpace table returns their constraint condition four-vectors $(n_x, n_y, n_z, c)$. A distribution of SEGMENT lengths is shown in Figure 36. A sample query output is provided in

⁵ http://skyserver.sdss.org/dr12/en/tools/chart/chartinfo.aspx
A.2. PRIMARYs, CHUNKs

Like SEGMENTs, PRIMARYs (and CHUNKs) are defined through four constraint conditions stored in the HalfSpace table. A sample query output is provided in Table 5.

```
    SELECT regionid, constraintid, x, y, z, c
    FROM HalfSpace
    WHERE regionID IN (SELECT regionID
                        FROM RegionConvex
                        WHERE type = "SEGMENT")
```

A.3. PRIMARY SEGMENTs

Each of the 2052 SEGMENTs belongs to one of the 111 PRIMARYs. This query identifies which SEGMENTs are associated with which PRIMARYs. A sample of the output is provided in Table 6.

```
    SELECT r.regionid as SegmentID,
           s.chunkID, c.regionID as PrimaryID
    FROM Segment s, Region r, Region c
    WHERE r.type = "SEGMENT" AND
          r.id = s.segmentID AND
          c.type = "PRIMARY" AND
          s.chunkID = c.id
    ORDER BY r.regionID
```

A point within the primary portion of a SEGMENT must meet the constraint conditions of both the SEGMENT and its corresponding PRIMARY, for a total of eight constraint conditions in all. The SQL procedure listed below returns a table with 2052 * 8 = 16,416 rows. Eight rows corresponding to a single PRIMARY SEGMENT are presented
CREATE TABLE PrSegConstraints
    (regionID int NOT NULL,
    constraintID int NOT NULL,
    x float NOT NULL,
    y float NOT NULL,
    z float NOT NULL,
    c float NOT NULL)

INSERT INTO PrSegConstraints
SELECT r regionid as RegionID,
    h constraintID, h x, h y, h z, h c
FROM Region r, Segment s, Region c, HalfSpace h
WHERE r.type = "SEGMENT" AND
    r.id = s.segmentID AND
    c.type = "PRIMARY" AND
    c.id = s.chunkid AND
    h.regionid = r.regionid

INSERT INTO PrSegConstraints
SELECT rc RegionID, h constraintid, h x, h y, h z, h c
FROM(SELECT r regionid as RegionID,
    s.chunkID
FROM Region r, Segment s, Region c
WHERE r.type = "SEGMENT" AND
    r.id = s.segmentID AND
    c.type = "PRIMARY" AND
    c.id = s.chunkid) rc,
    Region g, HalfSpace h
WHERE rc.chunkID = g.ID AND
    g.type = "PRIMARY" AND
    g.regionID = h.regionid

The distributions of PRIMARY SEGMENT and SEGMENT lengths are plotted in Figure 36. The figure shows that there are more PRIMARY SEGMENTS of shorter lengths and fewer of intermediate and long lengths. This verifies that the addition of PRIMARY constraints has shortened the SEGMENTs by cropping the areas that lie outside the SEGMENTs’ PRIMARYs.

The speed of searches over PRIMARY SEGMENTs depends on the order in which constraint conditions are applied. Because SEGMENTs are smaller than PRIMARYs, fewer points will survive the former’s constraints than the latter’s. This suggests that searching over SEGMENTs first will reduce the number of required mathematical operations. Assuming the constraint conditions are applied in the order they exist in the output table, a reverse ordering of the constraintID’s is preferable:

SELECT *
FROM PrSegConstraints
ORDER BY regionID, constraintID DESC

Such a sorting was used to measure the lengths of the regions in Figure 36. High-density, uniformly distributed angular randoms were placed in the vicinity of the regions, then the constraints were applied. The length of each region was set equal to the largest angular separation among the points that survived the filtering. We found that 116 PRIMARY SEGMENTs have lengths equal or approximately equal to zero.

### A.4. TILES

Each tile is circular and, thus, has only one constraint condition.

SELECT h regionid, h x, h y, h z, h c
FROM Region r, HalfSpace h
WHERE r.type = "TILE" AND
    r.regionid = h.regionid

The following query returns the SECTORs associated with each TILE. The section in green excludes SECTORs that are masks for a particular TILE. These masks are exclusively outside their associated TILE’s boundaries and were only used in the past to serve as masks for other TILES. The final column returned by this query contains the number of SECTORs in each TILE. An example output for TILE 112 is provided in

### Table 5
Partial Results for PRIMARYs/CHUNKs Query

| Regionid | Constraintid | x       | y       | z       | c         |
|----------|--------------|---------|---------|---------|-----------|
| 97       | 385          | -0.6421 | 1716    | -0.7659 | 98579     |
| 97       | 387          | -0.0343 | 5340    | -0.0038 | 0168     |
| 97       | 388          | 0.0343  | 5340    | 0.9990  | 4822     |
| 97       | 386          | 0.2584  | 9476    | 0.9658  | 9745     |
| 98       | 389          | -0.3568 | 5829    | 0.9340  | 8188     |
| 98       | 390          | -0.1590 | 9796    | -0.9872 | 0804     |

Note. Full query returns 888 rows, 4 for each of the 111 PRIMARYs and CHUNKs. Significant figures are truncated in final four columns.

### Table 6
Partial Results for PRIMARY/SEGMENTs Association Query

| SegmentID | ChunkID | PrimaryID |
|-----------|---------|-----------|
| 4870      | 579289762671475565 | 213       |
| 4871      | 579289762671475565 | 213       |
| 4872      | 579289762671475565 | 213       |
| 4873      | 579289762671475565 | 213       |
| 4874      | 579289762671475565 | 213       |
| 4875      | 579289762671475565 | 213       |

Note. Full query returns 2052 rows, one for each PRIMARY SEGMENT.
Table 8. Partial Results for TILEs Query

| Tile  | sectorID | ns  |
|-------|----------|-----|
| 112   | 33778    | 7   |
| 112   | 33852    | 7   |
| 112   | 33853    | 7   |
| 112   | 33854    | 7   |
| 112   | 33855    | 7   |
| 112   | 33856    | 7   |
| 112   | 33857    | 7   |

Note. Full query returns 19,646 rows.

A.5. SECTORs

SECTORs are the complex intersections of multiple TILEs and tile masks, and can be unions of anywhere between 1 and 12 constraint conditions. To search over SECTORs, the user must first determine the number of associated convexes. The constraint conditions for each convex are applied one at a time. Points that satisfy all of any convex’s constraints lie within the SECTOR.

The following query returns all the information needed to determine whether a point lies within a SECTOR. The eighth and third columns contain, respectively, the number of convexes associated with the SECTOR and the number of constraint conditions per convex. An example for a single SECTOR is provided in Table 9.

```sql
SELECT b.tile, b.regionID as sectorID, ns
FROM(SELECT tile count(*) as ns
     FROM SELECT
     FROM Sector2Tile
     WHERE type SECTOR
     AND isMask = 0) a
GROUP BY tile)

(SELECT *
     FROM Sector2Tile
     WHERE type SECTOR
     AND isMask = 0) b
WHERE a.tile = b.tile
```

A.6. MAIN GALAXY SAMPLE

We use the following query to extract MGS galaxies from the SDSS database. The first condition involving primaryforces the target to have at least one of the following three flags set: “TARGET_GALAXY,” “TARGET_GALAXY_BIG,” and/or “TARGET_GALAXY_BRIGHT_CORE.” The
The Astrophysical Journal, 820:115 (29pp), 2016 April

Table 9
Partial Results for SECTORs Query

| Regionid | Convexid | nc | x     | y     | z       | ε     | neconv |
|----------|----------|----|-------|-------|---------|-------|--------|
| 33267    | 0        | 3  | -0.45629748 | -0.77420675 | 0.43863028 | 0.999661875 | 2   |
| 33267    | 0        | 3  | -0.4175120 | -0.75420242 | 0.45405088 | 0.999661875 | 2   |
| 33267    | 1        | 4  | -0.45629748 | -0.77420675 | 0.43863028 | 0.999661875 | 2   |
| 33267    | 1        | 4  | -0.67301938 | 0.05888157 | 0.73727734 | -0.0146444924 | 2   |
| 33267    | 1        | 4  | -0.67301938 | 0.05888157 | -0.73727734 | 0.0146444924 | 2   |
| 33267    | 1        | 4  | -0.65683681 | 0.05746577 | 0.75183981 | 0     | 0      |
| 33267    | 1        | 4  | -0.59218453 | 0.64012213 | 0.48944984 | 0     | 0      |

Note. Full query returns 42,921 rows. Significant figures are truncated in columns 4–7.

spectral classification algorithm must also identify the target as a galaxy. We establish a lower, extinction-corrected r-band magnitude limit of 15. To stabilize the selection function we restrict galaxies to the redshift range 0.02 ≤ z ≤ 0.22. Several additional status checks are made to ensure the redshifts are of good quality. Every galaxy returned by this query possess an absolute magnitude ≥−17. As written, this query returns 480, 569 galaxies.

SELECT p.*, s.specObjID, s.zStatus, s.z, s.conf, s.specClass
FROM PhotoPrimary p, SpecObj s
WHERE (p.primtarget & 448) = 0 AND
(p.petromag_r - p.extinction_r) > 15 AND
p.specObjID = 0 AND
p.specObjID = s.specObjID AND
s.specClass = 2 AND
s.z <= 0.22 AND
s.z >= 0.02 AND
s.zStatus NOT IN (0, 1, 10, 2, 5, 8) AND
s.conf >= 0.9

The query for MGS objects is similar, but because these targets do not possess spectral information, all but the top three non-redshift-dependent conditions in the WHERE clause are omitted.

APPENDIX B
DR7 SECTORS REMOVED FROM ISF

The following is a complete list of the 303 DR7 SECTORS whose areas are deleted from the original SF to create the ISF.

Abazajian, K., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2003, AJ, 126, 2081
Abazajian, K., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2007, AJ, 123, 1755
Adelman-McCarthy, J. K., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
Adelmann-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2006, ApJS, 162, 38
Adelmann-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2007, ApJS, 172, 634
Adelmann-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2008, ApJS, 175, 297
Aihara, H., Allende Prieto, C., An, D., et al. 2011, ApJS, 193, 29
Blanton, M. R., Dalcanton, J., Eisenstein, D., et al. 2001, AJ, 121, 2358
Blanton, M. R., Lin, H., Lupton, R. H., et al. 2003, AJ, 125, 2276
Conway, J. H., & Sloan, N. J. 1993, Sphere Packings, Lattices, and Groups (2nd ed.; New York: Springer)
Eisenstein, D. J., Annis, J., Gunn, J. E., et al. 2001, AJ, 122, 2267
Gray, J., Szalay, A. S., Thakar, A. R., et al. 2004, arXiv:cs/040831
Gunn, J. E., Carr, M., Rockosi, C., et al. 1998, AJ, 116, 3040
Richards, G. T., Fan, X., Newberg, H. J., et al. 2002, AJ, 123, 2945
Smith, J. A., Tucker, D. L., Kent, S., et al. 2002, AJ, 123, 2121
Stoughton, C., Lupton, R. H., Bernardi, M., et al. 2002, AJ, 125, 485
Strauss, M. A., Weinberg, D. H., Lupton, R. H., et al. 2002, AJ, 123, 1810
Thakar, A. A., Kunz, P. Z., Szalay, A. S., & Szokoly, G. P. 2000, in ASP Conf. Ser. 216, Astronomical Data Analysis Software and Systems IX, ed. N. Mandel, C. Veillet, & D. Crabtree (San Francisco, CA: ASP), 231
Uomoto, A., Smeé, S., Rockosi, C., et al. 1999, BÃKS, 31, 1501
York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579

REFERENCES

Abazajian, K., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2003, AJ, 126, 2081
Abazajian, K., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2004, AJ, 128, 502
Abazajian, K., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2005, AJ, 129, 1755
Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
Adelmann-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2006, ApJS, 162, 38
Adelmann-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2007, ApJS, 172, 634
Adelmann-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2008, ApJS, 175, 297
Aihara, H., Allende Prieto, C., An, D., et al. 2011, ApJS, 193, 29
Blanton, M. R., Dalcanton, J., Eisenstein, D., et al. 2001, AJ, 121, 2358
Blanton, M. R., Lin, H., Lupton, R. H., et al. 2003, AJ, 125, 2276
Conway, J. H., & Sloan, N. J. 1993, Sphere Packings, Lattices, and Groups (2nd ed.; New York: Springer)
Eisenstein, D. J., Annis, J., Gunn, J. E., et al. 2001, AJ, 122, 2267
Gray, J., Szalay, A. S., Thakar, A. R., et al. 2004, arXiv:cs/040831
Gunn, J. E., Carr, M., Rockosi, C., et al. 1998, AJ, 116, 3040
Richards, G. T., Fan, X., Newberg, H. J., et al. 2002, AJ, 123, 2945
Smith, J. A., Tucker, D. L., Kent, S., et al. 2002, AJ, 123, 2121
Stoughton, C., Lupton, R. H., Bernardi, M., et al. 2002, AJ, 125, 485
Strauss, M. A., Weinberg, D. H., Lupton, R. H., et al. 2002, AJ, 123, 1810
Thakkar, A. A., Kunz, P. Z., Szalay, A. S., & Szokoly, G. P. 2000, in ASP Conf. Ser. 216, Astronomical Data Analysis Software and Systems IX, ed. N. Mandel, C. Veillet, & D. Crabtree (San Francisco, CA: ASP), 231
Uomoto, A., Smeé, S., Rockosi, C., et al. 1999, BÃKS, 31, 1501
York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579