1. THE QUASAR TARGET SELECTION FRAMEWORK

In this section, we describe the overarching framework of the BOSS quasar target selection algorithm. The al-

gorithm is complex, having developed from a philosophy that attempts to allow BOSS to optimize it's main goal—the baryon acoustic feature in the Lyα Forest —while maintaining a more generally useful sample for the community.

1.1. *The CORE and the BONUS*

The primary science for which BOSS is assembling a large sample of quasar spectra is to detect and measure the baryon acoustic feature in large-scale structure as traced by the Lyα Forest. As the Lyα Forest is unchanged on large scales by being illuminated in absorption by background quasars, a complex angular quasar coverage has no bearing on the power of the survey to measure structure.

In addition to the Lyα Forest survey, BOSS will assemble an unparalleled collection of quasars at $2 < z < 4$, potentially making a range of new science possible in the moderate-redshift universe. Many quasars in the BOSS target range are easy to find in single-epoch $ugriz$ imaging without any detrimental effect on measurements of clustering in the Lyα Forest.

To address both of these considerations, the BOSS quasar targeting framework was effectively divided into two algorithms. The first algorithm takes `calibObj` objects from the datasweeps and only uses single-epoch $ugriz$ data from within the SDSS to implement target selection decisions. The sample that results from this algorithm is called the "CORE" sample, and is intended to be useful for statistical studies of quasars. The second algorithm is free to incorporate multi-epoch data from within SDSS and multi-wavelength and multi-epoch data from any external survey that might help improve the ultimate signal derived by BOSS In the Lyα Forest. The sample that results from this second algorithm is called the "BONUS" sample. We caution the reader that the targeting algorithm that defines the BONUS sample may continue to change—sufficiently to render this document obsolete.

Over Year 1 of the survey, the definition of the CORE algorithm was altered several times to attempt to optimize return within the statistical sample of BOSS quasars. Here, we describe the final definition of the CORE, derived near the end of Year 1 which uses the Extreme Deconvolution method of Bovy et al. (2011)/Likelihood method of Kirkpatrick et al. (2011). We caution the reader that prior to Year 1, the quasar CORE did not reflect a statistical sample.

1.2. *Single-Epoch and Multi-Epoch Photometry*

Certain areas of the SDSS SGC (e.g., stripe 82) were deliberately targeted multiple times for imaging, and therefore comprise a sample in which multi-epoch imaging can be used to improve target selection, either through variation in information across multiple epochs, or by coaddition to reach a deeper flux limit. It is less well known that, due to the drift-scan nature of the survey, the SDSS NGC, and wider footprint, contains many areas with multi-epoch coverage where multiple scan lines overlapped. This is particularly a feature near the edges of the survey.

The quasar targeting algorithm incorporates two types of input fluxes. First, the standard imaging datasweeps compiled in, e.g., the NYU Value Added Galaxy Catalog.

Second, a different set of imaging that co-adds fluxes where there is multi-epoch information from overlapping scans. Single-epoch $ugriz$ imaging from SDSS is always sent to the CORE algorithm and multi-epoch imaging is always sent to the BONUS algorithm. Adopting only single-epoch imaging from SDSS in the CORE ensures that quasars selected in the CORE have a simple survey footprint that is easier to statistically reproduce.

Maybe include exact equation for coadd....it's in sdss_fcombine.pro

1.3. *Logic Cuts*

Along with the CORE and the BONUS algorithms, two other main algorithms are used in BOSS quasar target selection. The first includes or excludes sources based on whether they are known quasars or stars from other surveys. The second includes objects based on whether they have a match to an object in the FIRST radio survey. We discuss each of these algorithms subsequently and will refer to them as CORE, BONUS, KNOWN and FIRST.

Each algorithm is sent a different set of logic flags. All logic cuts are applied using single-epoch data with one exception: color cuts made in the FIRST selection algorithm are made on coadded, multi-epoch data. This is because FIRST objects are not considered to be part of the statistical sample that CORE is designed for, unless they also meet the criteria of CORE.

The logic cuts are made in the form of an overall bitmask, based on the combination of a subset of bitmasks. The choice of a bitmask allows any cut to be recorded as a unique bit, then combined logic cuts can be made based on whether combinations of those bits are non-zero. The 7 subset bitmasks are `RESOLVE_BITMASK`, `CALIB_BITMASK`, `BOUNDS_BITMASK`, `GMAG_BITMASK`, `GMAG_BITMASK_NOB`, `FIRST_COLOR_BITMASK`, `FLAG_BITMASK` and are discussed in full in Appendix A. Note that `CALIB_BITMASK`, which cuts to objects with top-quality photometry in SDSS imaging, is *not, in general, populated* but its definition is retained for backwards-compatibility.

Both CORE and BONUS are sent `BOUNDS_BITMASK` + `RESOLVE_BITMASK` + `CALIB_BITMASK` + `FLAG_BITMASK` + `GMAG_BITMASK`. Again, see Appendix A for definitions of these bitmasks.

KNOWN objects are sent `BOUNDS_BITMASK` + `RESOLVE_BITMASK` + `CALIB_BITMASK`. In essence, this means that all known objects are considered regardless of the quality of SDSS images, provided that they are primary objects in the SDSS footprint.

FIRST objects are sent `BOUNDS_BITMASK` + `RESOLVE_BITMASK` + `CALIB_BITMASK` + `FLAG_BITMASK` + `GMAG_BITMASK_NOB` + `FIRST_COLOR_BITMASK`. `GMAG_BITMASK_NOB` is applied to FIRST rather than `GMAG_BITMASK` because the main quasar contaminant at bright magnitudes are Galactic stars, which are radio-quiet. The `FIRST_COLOR_BITMASK` contains a cut on $u - g$ color to restrict to radio-loud sources that have colors consistent with $z > 2.2$ quasars. This is the only logic cut that we allow to be made on (coadded) multi-epoch photometry.

1.4. *Known objects*

The KNOWN algorithm takes objects passed to quasar targeting and matches them against a database of spec-

troscopically confirmed stars and quasars assembled from a range of surveys. An initial coordinate match between target imaging and the database of known objects is made at $1.5''$. Objects that don't already match a known object are then matched again at $2.0''$, with the additional caveat that they must muse also be within 0.5 luptitudes in the $g-band$. Objects are *always* targeted if they match a known quasar at $z \geq z_{\text{cut}}$ and are *never* targeted if they match a known quasar at $z < z_{\text{cut}}$ or a known star. Through the end of Year 1 of BOSS $z_{\text{cut}} = 2.15$. It became clear that the blue response of the spectrograph left little signal in the Lyα Forest for $z \sim 2.15$, so from Year 1 onwards, $z_{\text{cut}} = 2.2$

The database of known objects consists of confirmed stars and quasars from the 2dF QSO Redshift Survey (2QZ;), the 2dF-SDSS LRG and QSO (2SLAQ;); the SDSS DR7 quasar catalog (); the Sloan Extension for Galactic Understanding and Exploration survey (SEGUE;); the AAω-UKIDSS-SDSS Survey (AUS;); and a deep commissioning survey of BOSS quasars in SDSS stripe 82 carried out with the MMT/Hectospec (A. D. Myers, private communication). Objects observed through Year 1 of BOSS were also placed in the database of known objects, so that additional signal could be obtained from duplicate observations of quasars in overlapping plates confirmed for the first time by BOSS.

ADM...I'm never gonna publish the MMT survey, maybe just add unique objects from it as a table in an appendix in this paper...

1.5. *Sources that match the FIRST Survey*

The FIRST algorithm takes objects passed to quasar targeting and matches them against the FIRST survey source catalog (Faint Images of the Radio Sky at Twenty-Centimeters;). A coordinate match of $1''$ is made, with no additional cuts (beyond those mentioned in Section 1.3) on the basis of flags in the FIRST catalog. In general, only extragalactic objects are radio-loud. As most of the unresolved objects classified as "star-like" in SDSS imaging are not galaxies, the majority of unresolved radio-loud objects are quasars provided a close coordinate match is made to reduce chance superpositions (e.g., McGreer paper).

As noted in Appendix A, a reverse UV-excess (UVX) cut is made on FIRST matches to mainly limit to quasars at $z > 2.2$. As the FIRST matches are not considered part of the statistical ("CORE") sample, this reverse UVX cut is made on co-added colors, where available, using the information in the multiple-epoch coadded flux files (section 1.2).

1.6. *The CORE Algorithm*

The CORE algorithm takes objects passed to quasar targeting and applies the Extreme Deconvolution method of Bovy et al. (2011)/Likelihood method of Kirkpatrick et al. (2011) processed using single-epoch fluxes. Ideally, the CORE algorithm would be designed to reflect a nearly constant fiber density of 20 deg^{-2} across the entire BOSS survey. But, in actuality the CORE density fluctuates significantly across the survey. This is due to two major effects: First, because single-epoch fluxes have large measurement errors near the magnitude limit of the BOSS survey—and Galactic stars have similar colors to quasars at $z \sim 3$; and second, because the number of contaminating stars is a strong function of position in BOSS, while the number of quasars remains constant. From the end of Year 1, a threshold on (RELEVANT PARAMETER) of ($> X$) was chosen to define the BOSS sample. This fixes the fiber density for the core to 20 deg^{-2} across the NGC region of the SDSS.

The CORE algorithm is designed to create a statistical sample of quasars. But, over Year 1 of the survey, the algorithm was altered several times to attempt to find the best targeting approach, using many of the techniques that currently inform the BONUS algorithm (see Section 1.7). In addition, the CORE algorithm over the commissioning period, and through the end of Year 1, was tuned in a variety of ways to attempt to meet a fiber density of 20 deg^{-2}. We therefore caution readers not to consider the CORE a statistical sample during year 1 of the survey.

1.7. *The BONUS Algorithm*

KDE
Likelihood
ExD
NN
Combinator
uses multi-epoch fluxes
Will continue to change, it's a living algorithm
Could incorporate IR, GALEX etc.
Tuned to 40 by filling in difference on top of core.

1.8. *Quasar Target Selection Flags*

What comes out of this mess and in what combination (QSOFIRST+QSOKNOWN+ etc.). Currently put table in Appendix, could be here instead.

APPENDIX

QUASAR TARGETING LOGIC CUTS

All objects initially classified as a star (i.e. `OBJC_TYPE = 6`) are passed to the quasar target selection code. Subsequently, 6 different sets of combinations of possible logic cuts are combined and applied to sources that are passed to quasar targeting, depending on the reasons that source could be targeted. For instance, sources targeted because they are known quasars undergo a different set of logic cuts than sources targeted because they match the FIRST radio survey. The flag cuts are quicker to process than the more complex main quasar targeting algorithms, so are carried out at the start of the algorithm to reduce runtime.

Several individual masks of bits are populated according to logic cuts, and combined into a parent mask of bits (henceforth referred to as the *bitmask*). The *bitmask* makes it possible to determine why a source was not passed through a particular portion of the quasar targeting algorithm. For example, the bit 2^0 was reserved for extended objects, and so is never populated in the *bitmask*. In this supplement we describe these 6 parent logic cuts and how

the *bitmask* is populated. Note that the magnitude logic contains two sets of possible parent logic cuts. In the main text, we refer to various combinations of the individual masks listed in the individual subsections of this supplement.

Resolve Logic

The `RESOLVE_BITMASK` records whether a source is a primary target in the SDSS photometry. The *bitmask* is populated with a value of 2^{13} for objects that are not primary. In the latest version of the target selection code, only sources that are primary are passed through targeting.

Calibration Logic

The `CALIB_BITMASK` records whether a target was observed in photometric conditions. The *bitmask* is populated with a value of 2^{14} for objects that were not observed in photometric conditions. In the latest versions of the target selection code, no cut is made on calibration logic but we retain the `CALIB_BITMASK` information for backwards compatibility with pieces of code written earlier in the survey.

Boundary Logic

The `BOUNDS_BITMASK` records whether a source is within the SDSS geographical target footprint. The *bitmask* is populated with a value of 2^{16} for objects that are outside of the footprint. Only sources that are within the footprint are passed through targeting.

FIRST Color Logic

The `FIRST_COLOR_BITMASK` is a special color cut reserved for objects that are targeted because they match a radio source from the FIRST survey. To better limit to high redshift quasars within the FIRST sample, we impose a reverse UVX cut at $u - g > 0.4$ (extincted PSF luptitudes) on all FIRST matches. Sources that fail this cut (but might otherwise match FIRST sources) are flagged in the *bitmask* with a value of 2^{17}.

Magnitude Logic

The `GMAG_BITMASK` records whether a target meets the default magnitude limits required to be targeted as a quasar. Magnitude cuts are made on extincted PSF luptitudes. In the latest version of target selection, the default magnitude limits are $(g \leq 22 \parallel r \leq 21.85)$ && $i \geq 17.8$. The *bitmask* is populated with a value of 2^{11} for objects that lie outside of this magnitude range. A second incarnation, `GMAG_BITMASK_NOB`, is used when no bright cut is required—such as when retargeting known quasars. Magnitude cuts are then simply based on the faint limit $(g \leq 22 \parallel r \leq 21.85)$. The *bitmask* is populated with a value of 2^{12} for objects that fail this logic.

Flag Logic

By far the most complicated set of logic cuts is based on flags for sources that had problems when processed through the photometric reduction pipeline, which populate the `FLAG_BITMASK`. The *bitmask* contains rich information on which of these flag cuts a potential target failed, as listed in Table 1. The raw flag cut definitions used below are detailed in Stoughton et al. (2002) except for gerr, rerr and ierr, which denote errors on extincted PSF luptitudes in the g, r and i bands, , and for the column and row movement information.

First, two sets of flag cuts are defined from the raw flag cuts, denoting whether the photometric pipeline had difficulty deblending or interpolating a source:

`INTERP_PROBLEMS` = (`PSF_FLUX_INTERP` && (gerr > 0.2 $\parallel$ rerr > 0.2 $\parallel$ ierr > 0.2)) $\parallel$ `BAD_COUNTS_ERROR` $\parallel$ (`INTERP_CENTER` && `CR`)

`DEBLEND_PROBLEMS` = `PEAKCENTER` $\parallel$ `NOTCHECKED` $\parallel$ (`DEBLEND_NOPEAK` && (gerr > 0.2 $\parallel$ rerr > 0.2 $\parallel$ ierr > 0.2))

where symbols (&&, $\parallel$, !) denote standard Boolean logic. Then, a source is determined to pass the flag logic (is `GOOD`) if it passes a range of flag cuts:

`GOOD` = `BINNED1` && !`BRIGHT` && !`SATURATED` && !`EDGE` && !`BLENDED` && !`NODEBLEND` && !`NOPROFILE` && !`INTERP_PROBLEMS` && !`DEBLEND_PROBLEMS` && !`MOVED`

where `MOVED` is set if the raw flag `DEBLENDED_AS_MOVING` is set *and* if the value and error on the row or column position in the SDSS imaging pipeline (rowv, colv, rowverr, colverr) suggests a $> 3\sigma$ move across a row or a column.

QUASAR TARGET FLAGS

Flag	$bitmask$ value	Flag	$bitmask$ value
EXTENDED	2^0	NODEBLEND	2^9
INTERP_PROBLEMS	2^1	NOPROFILE	2^{10}
DEBLEND_PROBLEMS	2^2	GMAG_RANGE	2^{11}
NOT_BINNED1	2^3	GMAG_RANGE_NO_BRIGHT_CUT	2^{12}
EDGE	2^4	NOT_PRIMARY	2^{13}
BRIGHT	2^5	NOT_PHOTOMETRIC	2^{14}
SATUR	2^6	EXTRA_LOGIC	2^{15}
MOVED	2^7	OUT_OF_BOUNDS	2^{16}
BLENDED	2^8	FIRST_MATCH_UMG_TOO_BLUE	2^{17}

TABLE A1

POSSIBLE IMAGING PROBLEMS THAT CAUSE A SOURCE TO BE FLAGGED IN THE $bitmask$. SOURCES WITH NON-ZERO VALUES IN THE $bitmask$ ARE NOT PASSED THROUGH CERTAIN SECTIONS OF THE QUASAR TARGETING CODE. ALL VALUES IN THE $bitmask$ ARE DESCRIBED IN THE APPENDIX, EXCEPT FOR EXTRA_LOGIC WHICH IS USED FOR TESTING ALTERATIONS TO THE ALGORITHM.

Flag	$bitmask$ value	Flag	$bitmask$ value
QSO_CORE	2^{10}	QSO_LIKE	2^{17}
QSO_BONUS	2^{11}	QSO_FIRST_BOSS	2^{18}
QSO_KNOWN_MIDZ	2^{12}	QSO_KDE	2^{19}
QSO_KNOWN_LOHIZ	2^{13}	QSO_CORE_MAIN	2^{40}
QSO_NN	2^{14}	QSO_BONUS_MAIN	2^{41}
QSO_UKIDSS	2^{15}	QSO_CORE_ED	2^{42}
QSO_KDE_COADD	2^{16}	QSO_CORE_LIKE	2^{43}

TABLE B1

QUASAR TARGET FLAGS.