SDSS Stripe 82 : quasar variability from forced photometry

Krzysztof Suberlak,[1]* Željko Ivezić, [1] Yusra AlSayyad[1]
[1]*Department of Astronomy, University of Washington, Seattle, WA, United States*

ABSTRACT

1 INTRODUCTION

Introduction : why S82 is great, why we need Quasars, what has been done so far with S82, why we can do better with reprocessed data, our stab at successfully selecting quasars (among other objects) , what we do with our 'gold' sample.

2 METHODS

3 DATA OVERVIEW

3.1 Stripe 82

We use data from all SDSS runs up to an including run 7202 (Data Release 7), including all 6 SDSS camera columns. Stripe 82 survey covered an equatorial strip of the sky, defined by declination limits of ± 1.27 deg, extending from R.A. $\approx 20^h (320 \deg)$ to R.A. $\approx 4^h (55 \deg)$ (Sesar et al. 2007; Sesar et al. 2010). Observations conducted prior to September 2005 (part of SDSS I-II) had a more sparse sampling than SDSS-III, and the SDSS Supernova Survey, which ran between September 1st - November 30th each year between 2005-2007.

The SDSS Stripe 82 DR7 data was processed in two data centers : NCSA (National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, IL) and IN2P3 (Institut national de physique nucléaire et de physique des particules in Paris, France). NCSA processed data from $-40 \deg (+320 \deg) < RA < +10 \deg$ and IN2P3 with $+5 \deg < RA < +55 \deg$. There is a 5 deg overlap, used to confirm that the data processing pipeline in both data centers yields identical data products. The entire strip was split into smaller patches.

All epochs (individual images) were background-subtracted, and then scaled from the Digital Unit counts to fluxes by comparing standard objects against the (Ivezić et al. 2007) catalog (similar to Jiang+2014).

3.2 Source Detection

Sources were detected in the i-band coadds. Each detection in the coadded images was assigned a deepSourceId (elsewhere called objectId). Considering a dense region with clumped stars and/or galaxies, the entire clump was considered as one parent source (with single ParentSourceId). For an object which is a parent (eg. a galaxy), ParentSourceId

is null. Solitary sources which are not blended in clumps are their own parents. The result of this procedure were 40 million i-band detections down to 3σ. 8 million of those are brighter than 23rd mag. Part of Stripe82 processed in NCSA yielded 20978391 detections (`iCoaddAll.csv`). The part that does not overlap with IN2P3 has 16520093 sources (`iCoaddPhotometryAll.csv`), of which 16514187 are brighter than 30^{mag} (5906 less) (`DeepSourceNCSA_i_lt300.csv`). [1]

3.3 Forced Photometry

On positions specified by the detection data AlSayyad+2015 performed forced photometry in all SDSS photometric bands, on the individual epoch images. It is different from image differences technique, where the photometry is done on a difference between a coadd and an individual epoch image. The total number of photometric measurements (combining NCSA and IN2P3) was (40 million i-band detections) x (80 epochs) x (5 filters) = 16 billion measurements, including (8 million i-band detections i < 23) x (80 epochs) x (5 filters) = 3.2 billion measurements brighter than 23rd mag. For each patch the raw light curves contain the `id`, `objectId`, `exposure_id`, `mjd`, `psfFlux`, `psfFluxErr`, sorted by `objectId`, measuring flux in $[ergs/cm^2/sec/Hz]$ (`rawDataFPSplit/bandPatchStart_PatchEnd.csv`).

4 ANALYSIS

We developed a new pipeline that was applied to all forced photometry light curves. The main steps involve:

- selecting faint epochs (where S/N is less than a selected threshold),
- applying the Bayesian treatment (see an accompanying paper for details) and replacing the flux for faint epochs
- calculating a number of standard flux-based features (mean, median, skewness, χ^2_{DOF}, etc., as well as applying the full Bayesian likelihood to parametrize the probability that the object is intrinsically variable
- calculating flux-based magnitudes
- calculating seasonal averages per light curve

---

[1] Code used to create Summer 2013 re-processed summary files is available on GitHub at https://github.com/imcgreer/QLFz4

- merging the light curve aggregates across filters

[the rest moved to FAINT PIPELINE REPORT]

4.1 Variability

In what follows, for faint measurements we choose to replace F with F_{mean} (see Faint Pipeline Report). For all objects we calculate statistics based on entire light curves. We denote (F_{obs}, σ_{obs}) as (x_i, σ_i) :

- mean weighted by $w_i = (\sigma_i)^{-2}$

$$\bar{x} = \frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i} \qquad (1)$$

- weighted mean error

$$\sigma_{\bar{x}} = \left(\sum w_i\right)^{-1/2} \qquad (2)$$

- weighted standard deviation

$$\sigma_{st.dev.w.} = \left(\frac{\sum w_i (x_i - \bar{x})^2}{\sum w_i}\right)^{1/2} \qquad (3)$$

- error on standard deviation

$$\sigma_s = \frac{1}{\sqrt{2}} \sqrt{\frac{N}{N-1}} \sigma_{\bar{x}} \qquad (4)$$

- robust interquartile width

$$\sigma_G = 0.7414 * (75\% - 25\%) \qquad (5)$$

- median as the 50-th percentile , median error

$$\sigma_{median} = \sqrt{\pi/2} \sigma_{\bar{x}} \qquad (6)$$

Variations in object brightness have two main origins: an error-induced noise, and an intrinsic variability. A light curve consists of a set of N measurements of brightness x_i with errors e_i. In this analysis we assume that x_i are drawn from a Gaussian distribution $\mathcal{N}(\mu, \sigma)$, and that errors e_i are homoscedastic. We describe this distribution with two parameters : mean μ, and width σ. To increase efficiency, we employ a two-step approach after Ivezic+2014. First, we find approximate values of μ_0 and σ_0, and then we evaluate the full logarithm of the posterior pdf in the vicinity of the approximate solution. With a Bayesian approach, we find $\mu_{full}, \sigma_{full}$ by maximizing the posterior probability distribution function (pdf) of μ, σ given x_i and e_i : $p(\mu|x_i, \sigma_i)$ (see Fig. 1, and Appendix B for the detailed calculation).

For each light curve, we also calculate mean-based χ^2_{DOF} and median-based χ^2_R (the latter is more robust against any outliers in the distribution) :

$$\chi^2_{dof} = \frac{1}{N-1} \sum \left(\frac{x_i - <x_i>}{e_i}\right)^2 \qquad (7)$$

and

$$\chi^2_R = 0.7414(Z_{75\%} - Z_{25\%}) \qquad (8)$$

with $Z = (x_i - median(x_i))/e_i$.

On Fig. 1 we plot the stages of calculating μ and σ. Left and middle panels compare the two methods of calculating variability parameters. The initial approximation (dashed)

is based on bootstrapped resampling of (x_i, e_i) points from the light curve. By randomly resampling the light curve M times, instead of a single sample with $N \approx 10 - 70$ points we have M samples. The histogram of $M = 1000$ values for μ, σ from resampling is plotted with dashed lines. We use the approximate values to provide bounds for the 200x70 grid of μ, σ, used to evaluate the full posterior likelihood density function (right panel). This ensures that, despite using a coarse grid to improve computational speed, we still resolve the peak of the underlying distribution.

All variability parameters describe in a certain way the light curve variability. σ_{full} corresponds to the spread of the flux distribution. For a non-variable source, χ^2 would be centered about 1, with a width σ of $\sqrt{2/N}$, where N is the number of points per light curve. Thus we would expect that for any distribution 50% of sources would have $\chi^2 > 1$. Therefore a "3σ" variability detection would require $\chi^2 > \chi^2_{limit}$, with $\chi^2_{limit} = 1 + 3\sqrt{2/N}$ We call a source a 'robust variability candidate' if $\sigma_{full} > 0$ and ($\chi^2_{DOF} > \chi^2_{limit}$ or $\chi^2_R > \chi^2_{limit}$), i.e. we require the full σ to be larger than 0, and either robust or DOF χ^2 to be larger than the limiting χ^2.

Initially, we evaluate variability parameters μ_{full}, σ_{full}, χ^2_{dof}, and χ^2_R based on all points of the light curve. Only for variable sources, as defined above, we calculate the variability parameters for the seasonally-binned portions of the light curve. The χ^2_{DOF} vs. χ^2_R plotted on Fig. ... shows that these distributions are similar for seasons as well as full light curves.

4.2 Colors

Since the reported fluxes are not extinction-corrected, we use a table of E(B-V) in a direction of a given source to correct for the galactic extinction. We use the formula $x_{corr} = x_{obs} + A_x * E(B - V)$, where x is u,g,r,i,z , and A_x is 5.155, 3.793, 2.751, 2.086, 1.479 for each filter respectively [Schlegel 98, Av are for RV = 3.1, also suggested by Eddie Schlafly]

Colors $x - y$ for an object with observations over many epochs are defined as the difference in magnitudes $m_x - m_y$. To find m_x, we need to define the average brightness of an object in a given filter. With a special treatment of faint sources, substituting (F_{obs}, σ_F) for each faint observation by $(<F_{exp}>, rms)$, we analyse updated lightcurves, addressing sparse sampling (see Fig. 4).

Thus for a given object we average all sparser observations prior before SDSS-III, and calculate annual averages for all subsequent years. We calculate weighted mean and the rms as

$$<F> = \frac{\sum w_i F_i}{\sum w_i} \qquad \sigma_{<F>} = \left(\sum w_i\right)^{-1/2} \qquad (9)$$

with weights as $w_i = 1/\sigma_i^2$. We also calculate the robust median and the median error : $\sqrt{\pi/2} \sigma_F$ [robust $\sigma_G = 0.7414 * (75\% - 25\%)$, based on the interquartile range] . Then lightcurve for a given object is reduced to one (F_i, σ_i) point prior to March 2006, and a single point per every subsequent year, where (F_i, σ_i) is (mean, meanErr) or (median, medianErr).

The resulting average flux is converted to magnitude, and the color is $c = m_x - m_y$, with combined errors of band light curves added in quadrature
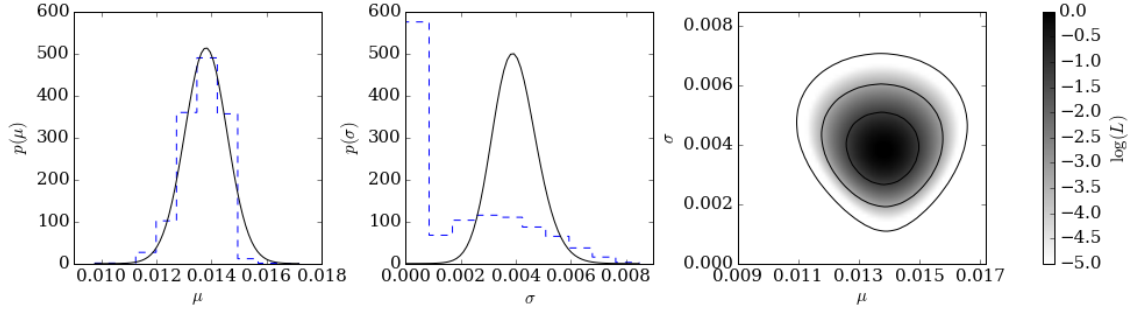
Figure 1. Two-step approach to finding μ and σ via μ_0 and σ_0 for an object 217720894888346446. In this calculation we use raw psf flux, before employing the faint source treatment outlined in Section 4. On the left and middle panels, solid lines trace marginalized posterior pdfs for μ and σ , while dashed lines depict histogram distributions of 10,000 bootstrap resamples for μ_0 and σ_0. The right panel shows the logarithm of the posterior probability density function for μ and σ.
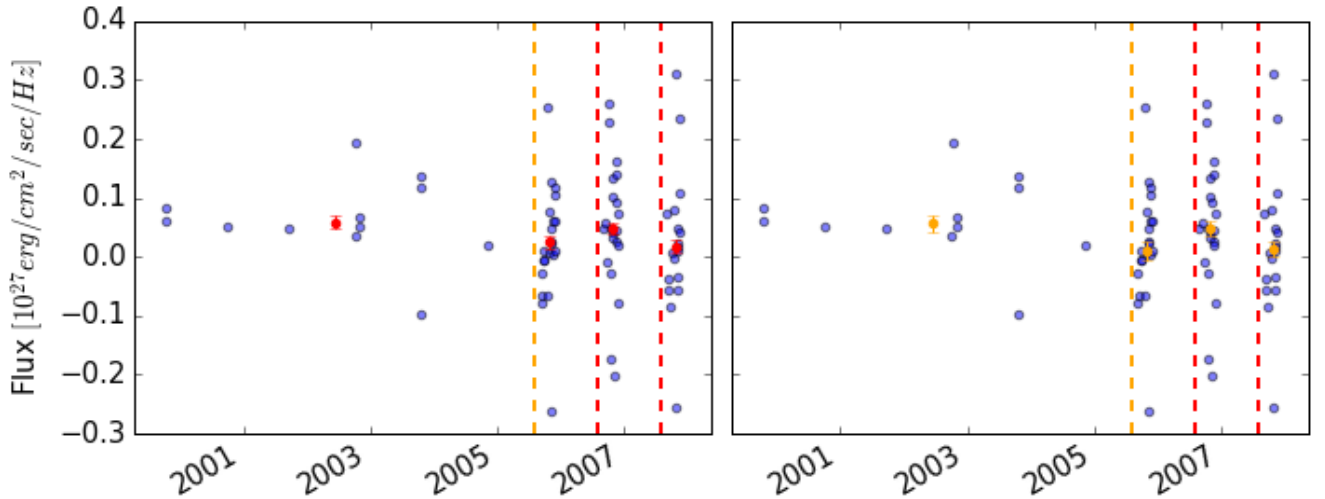
Figure 2. A plot showing an outcome of seasonal averaging for an object id 217720894888346425. The left panel (red dots) shows (mean, meanErr), and the right panel (orange) shows (median,medianErr), instead of seasonal points (blue). Vertical dashed lines as on Fig. 4

5   RESULTS

6   CONCLUSIONS

APPENDIX A: CHARACTERIZING VARIABILITY

We further characterize the variability of light curves by calculating σ_0 (the approximate value), and σ_{full}, following Ivezic+2014, chapter 5.

σ_0 is found in the following way: if by (x_i, e_i) we denote the measurement and associated error, then the bootstrapped sampling of (x_i, e_i) is sampling each vector at a
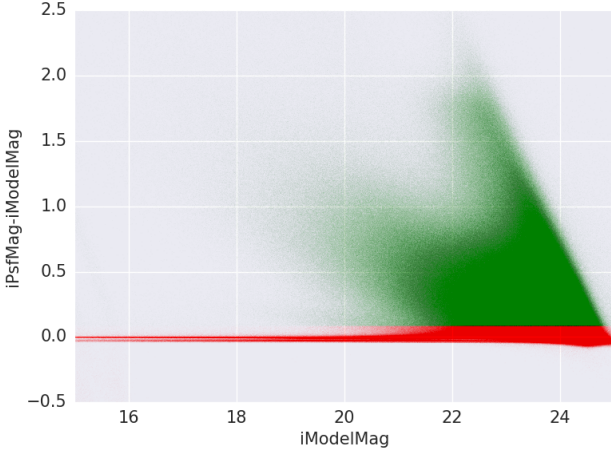
Figure 3. Scatter plot of all NCSA sources detected in coadds (from Deep Source catalog); The coloring corresponds to the `extendedness` parameter calculated in the pipeline based on the value of iPsfMag-iModelMag : red being 0 (compact), and green being 1 (extended). The threshold iPsfMag-iModelMag = 0.085645 (indeed, the red points appear to be cut sharply at y=0.085645 line). Initially, for bright objects, on the left side of the plot, iPsfMag-iModelMag is very much larger for green (extended) as opposed to red (compact) objects. As iModelMag decreases (and correspondingly, so does iPsfMag), the objects become fainter, and less extended. This is why the green swaths approach the red horizontal group. Indeed, for fainter objects, the separation becomes less certain, as more distant galaxies appear to be more compact.
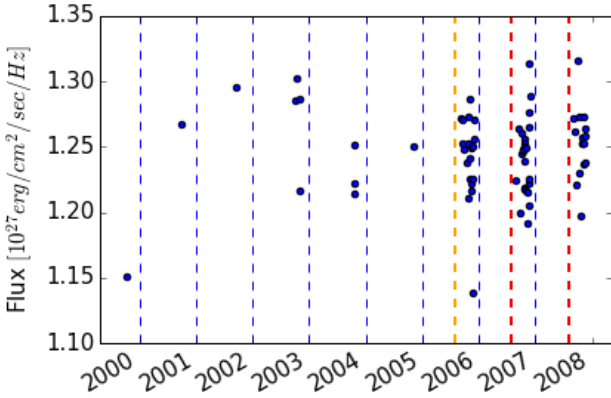

Figure 4. A plot showing an example light curve for an object id 217720894888346422. Jan 1st of each year (blue), August 1st of 2005 (orange) and August 1st of each subsequent year (red) is indicated by vertical dashed lines. Observations prior to August 1st of 2005 have sparser cadence, whereas those after that date have more frequent observations. This is due to the SDSS-III Supernova Survey which begun Sept 1st 2005. All points to the left of August 1st 2005 (orange line) are averaged together. Points to the right of August 1st 2005 are seasonally averaged.

number of N random indices (eg. N=1000). Thus instead of x_i which may include only N=10 measurements, we have $x_{i,boot}$ which has N=1000 random samples. Median is the 50-th percentile of any sample. Following [Ivezic+2014], chapter 5, we use the sample median to estimate $\mu_0 = median(x_{i,boot})$, and an interquartile range width estimator to estimate the standard deviation : $\sigma_G = 0.7413(X_{75\%} - X_{25\%})$ for $X = x_{i,boot}$. With the median error $e_{50} = median(e_{i,boot})$, we estimate σ_0 as :

$$\sigma_0 = (variance_{approx})^{1/2} = (\zeta^2 \cdot \sigma_G^2 - e_{50}^2)^{1/2} \tag{A1}$$

where

$$\zeta = \frac{median(\widetilde{\sigma_i})}{mean(\widetilde{\sigma_i})} \tag{A2}$$

and

$$\widetilde{\sigma_i} = (\widetilde{variance})^{1/2} = (\sigma_G^2 + e_i^2 - e_{50}^2)^{1/2} \tag{A3}$$

For the marginalized σ_{full}, we calculate logarithm of the posterior probability distribution for the grid of μ and σ values as:

$$\log L = -0.5 \sum \left(\ln(\sigma^2 + e_i^2) + \frac{(x_i - \mu)^2}{(\sigma^2 + e_i^2)} \right) \tag{A4}$$

We shift the maximum value of logL by subtracting the maximum value of logL, thus calculating the likelihood :

$$L = e^{\log L - max(\log L)} \tag{A5}$$

We then marginalize over μ or σ :

$$p(\sigma) = \sum_\mu (L_{\sigma,\mu}) \qquad p(\mu) = \sum_\sigma (L_{\sigma,\mu}) \tag{A6}$$

and normalize the probability :

$$p_{norm}(\sigma) = \frac{p(\sigma)}{\int p(\sigma)\mathrm{d}\sigma} \qquad p_{norm}(\mu) = \frac{p(\mu)}{\int p(\mu)\mathrm{d}\mu} \tag{A7}$$

To characterize light curve variability we first calculate for the entire light curve of an object the approximate μ_0 and σ_0 using bootstrapped resampling of the light curve. This yields the boundaries for the more exact calculation of the full 2D log-likelihood performed on a grid of μ and σ values. Thus the more accurate σ_{full} and μ_{full} are found as a maximum of the 2D log-likelihood distribution (see Fig. 1).

APPENDIX B:
MAKING OF UGRIZ METRICS

Colors can be calculated in two ways: using the median of forced photometry over all epochs (object detected in coadded i-band has photometry in all epochs: `ugrizMetrics.csv`), or the median over single-epoch detections (only when an object was above the detection threshold for a single epoch : `medianPhotometry.csv`). The median
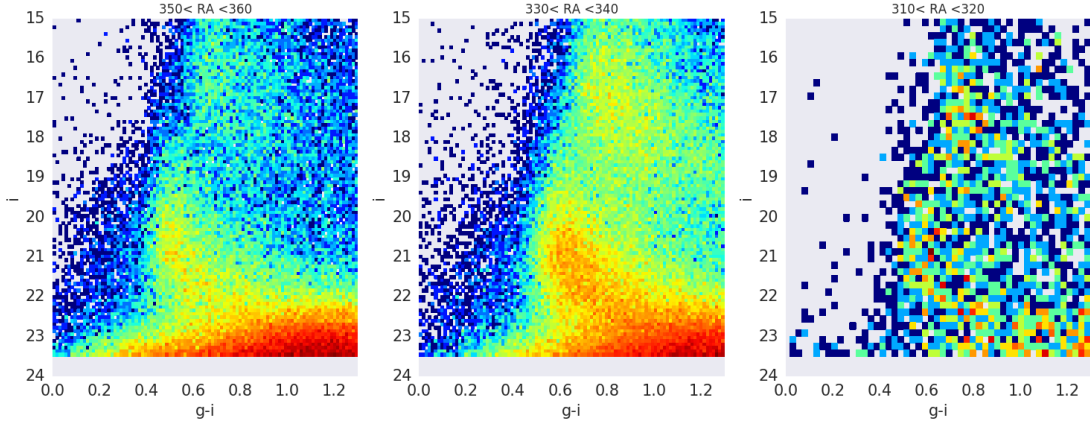
Figure 5. A color-magnitude plot , reproducing the results of (Sesar et al. 2010) , Fig.23 . We show here only NCSA-processed sources, which is why certain RA ranges are omitted or have less sources. We only select sources with `extendedness=0` parameter (stars). The scale is showing the $\log_{10}$ of count. All sources have their colors corrected for extinction. On first two panels the features of Sagittarius Stream are clearly visible.
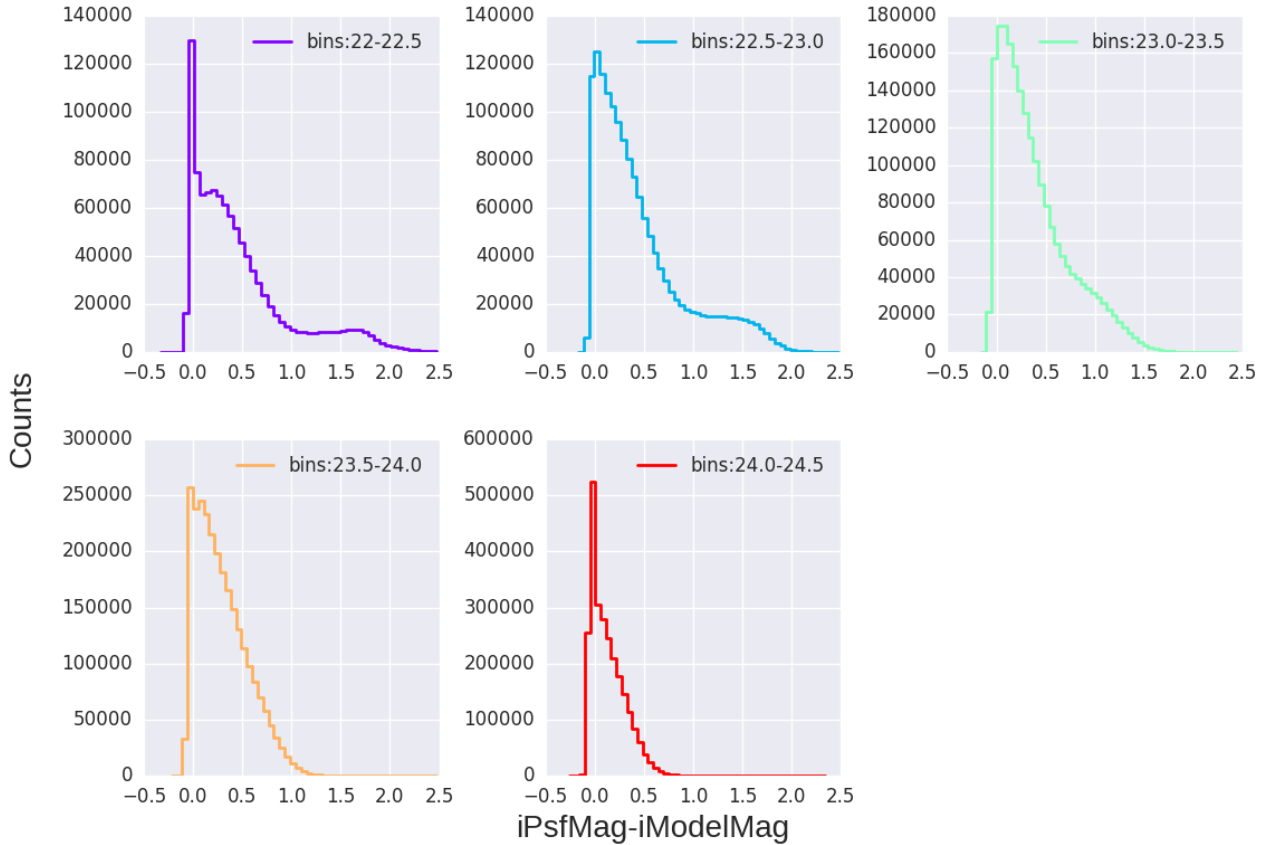

Figure 6. The histograms show the count of sources in 5 i-magnitude bins, corresponding to the vertical cut through Fig. 3. It helps to verify how well can the extended and compact sources be separated based solely on the iPsfMag-iModelMag. In the S82 reprocessing pipeline, the Deep Source catalogs were made assuming that whenever iPsfMag-iModelMag > 0.085645, extendedness = 1. In other words, when the object appears brighter by 0.085645 magnitudes with the model fit rather than psf magnitude, then it is considered to be extended. See Palanque-Delabrouille+2016 Fig.2 for more on how iPsfMag-iModelMag can be used to select quasars. Specifically, nearby quasars may still have the host galaxy visible, which would in some cases not pass this criterion and be assigned extendedness 0.

over all detections will be biased (especially for faint sources) towards higher brightness. On the other hand, the median over all epochs will be more representative of the true brightness of an object in a given filter. If a median brightness is negative, we can use zero point magnitudes and in such cases median over all epochs will be an upper limit on brightness, but still less biased than median over all detections. Therefore we choose to use median over all epochs to calculate colors (see Fig. 5 for an example).

APPENDIX C:
ZERO FLUX MAGNITUDES

If the median flux of an object over all epochs is negative (an outcome of forced photometry on fluctuating noise), we cannot define its magnitude in that filter. In such situation one can revert to using for each negative flux the zero point magnitude (m_1) - the magnitude for a source with a flux of 1 count per sec, different for each exposure. The zero point magnitude for each exposure with negative flux is calculated from the Flux of 0 magnitude source, F_0, as $m_1 = 2.5 \log_{10} F_0$. For that object the new median magnitude in that filter will be the upper limit. We did not use this method, since a better way is to calculate the $2 - \sigma$ flux limit for each flux measurement $< 2\sigma : F_{2\sigma}$.

APPENDIX D:
LIGHTCURVE METRICS

For each object we calculate lightcurve-derived metrics. Denoting `psfFlux` and `psfFluxErr` as y and σ_y, we find the number of measurements per lightcurve (N), the mean flux, the median flux (the 50th quartile), median flux error `e50`, mean flux error `e_mean`, σ_G (based on interquartile flux range 0.7413(q75-q25)), χ^2:

$$\frac{1}{N-1} \sum \left(\frac{y - mean(y)}{\sigma_y} \right)^2 \tag{D1}$$

mean weighted by the `WVar` - the inverse variance (`WeightedMean`), and the standard deviation weighted by inverse variance and corrected for intrinsic scatter (`WeightedStdCorr`):

$$\text{WVar} = \left(\sum \frac{1}{\sigma_y^2} \right)^{-1} \tag{D2}$$

$$\text{WeightedMean} = \text{WVar} \sum \frac{y}{\sigma_y^2} \tag{D3}$$

$$\text{WeightedStdCorr} = \left[\frac{\text{WVar}}{N-1} \sum \frac{(y - \text{WeightedMean})^2}{\sigma_y^2} \right]^{1/2} \tag{D4}$$

From these metrics, we can calculate the catalog photometry :

$$\text{median\_mag} = -2.5 \log_{10} \text{median} - 48.6 \tag{D5}$$

There are 5892054 sources with catalog photometry brighter than 23 mag (`ugrizMetrics.csv`).

We can also calculate median photometry over all individual epochs detections, cross-matched by extinction tables [HOW ?] There are 12373162 sources with median photometry, matched with E(B-V) data (`medianPhotometry.csv`) .

For each band we calculate metrics describing the lightcurve behavior for a given band, including the Butler & Bloom classifier, which can for high S/N objects, where it has a good discriminating power. It's advantage over the full DRW analysis for each lightcurve is that by assuming a range of τ, amplitude, expected for a DRW for a QSO, we calculate the likelihood of a given lightcurve belonging to a QSO (`i_metrics.csv`).

APPENDIX E:
VARIANCE OF χ^2_{DOF}

Detecting variability is similar to asking whether the data are consistent with constant signal with noise. We simulate simple sinusoidal time-series to investigate what is the minimum amplitude that can be detected given measurement errors. Our time series is sourced from a parent distribution $y(t) = A(\sin(\omega t)$ by a sample of N 100 points, with homoscedastic Gaussian errors : $\varepsilon \sim \mathcal{N}(0, \sigma)$. Thus each point of the sample distribution $x_i = y_i + \varepsilon_i$. We call the theoretical variance of the parent distribution a variance over time series. Variance over realizations of the time series (samples) with errors is an estimator of the parent variance. For a continuous smoothly varying probability function that describes the parent distribution, the mean is the first raw moment:

$$\mu = \int_{-\infty}^{\infty} x p(x) dx \tag{E1}$$

and the variance is the second raw moment:

$$Var = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 p(x) dx = \int_{-\infty}^{\infty} x^2 p(x) dx - \mu^2 \tag{E2}$$

We denote by $\langle x \rangle$ the expectation value of x. Thus the above can be written as : $Var = \sigma^2 = \langle x^2 \rangle - \langle x \rangle^2$.

For our parent distribution, $Var(x) = Var(y + \varepsilon) = Var(y) + Var(\varepsilon) = Var(A \sin(x)) + \sigma^2$. Now from the definition, $Var(A \sin(x)) = \langle A^2 \sin^2(x) \rangle - \langle A \sin(x) \rangle^2 = A^2/2$, because $\langle \sin^2(x) \rangle = 1/2$ and $\langle sin(x) \rangle = 0$ due to symmetry. Thus $Var(x) = A^2/2 + \sigma^2$.

Now, similarly to variance, chi-square can be considered either for the theoretical parent distribution, or for the sample, where the latter may be a good estimator of the parent chi-square if sufficient sampling is available. The chi-square is often defined as a sum of squares of departures from the mean over the variance:

$$\chi^2 = \sum \frac{(x_i - \mu)^2}{\sigma^2} \tag{E3}$$

where μ is the predicted mean, and σ^2 the expected variance. When divided by the number of degrees of freedom, it

is $\chi^2_{DOF} = \chi^2/N$ (also called χ^2_{PDF} : per degree of freedom). After Bevington&Robinson, χ^2 can be also understood as:

$$\chi^2 = \sum \frac{[h(x_J) - NP(x_J)]^2}{\sigma_J(h)^2} \quad (E4)$$

with $h(x_j)$ being the observed frequencies of x, $y(x_J) = NP(x_J)$ the predicted distribution, and $\sigma_J(h)^2$ the variance of the parent population. It is equivalent to our definition.

Thus using the first definition, for x_i we have $\chi^2 = 1/N \sum (x_i - \mu/\sigma)^2$. Since $\mu(x) = 0$, we have:

$$\chi^2_{DOF} = \frac{1}{N\sigma^2} \sum x_i^2 = 1/\sigma^2 \langle x_i^2 \rangle \approx Var(x_i)/\sigma^2 \quad (E5)$$

since for $Var(x_i) = \langle x_i^2 \rangle$ (zero mean). Thus we approximate the parent variance by the sample variance:

$$\chi^2_{DOF} \approx \frac{A^2/2 + \sigma^2}{\sigma^2} \quad (E6)$$

Now if we assume that there is no amplitude, i.e. $A = 0$, $\chi^2_{DOF} = 1$, and the expectation value $\langle \chi^2_{DOF} \rangle = 1$. With zero mean, the variance of χ^2_{DOF} is :

$$Var(\chi^2_{DOF}) = Var\left(\frac{1}{N} \sum \frac{x_i^2}{\sigma^2}\right) = Var\left(\frac{v}{N\sigma^2}\right) \quad (E7)$$

Now $Var(\alpha x) = \alpha^2 Var(x)$. Thus

$$Var\left(\frac{v}{N\sigma^2}\right) = \frac{Var(v)}{N^2\sigma^4} \quad (E8)$$

$Var(v) = \langle v^2 \rangle - \langle v \rangle^2$. We find first the second component: the mean $\langle v \rangle = \langle \sum x_i^2 \rangle = \sum \langle x_i^2 \rangle$ (no cross terms). And since $Var(x_i) = \langle x_i^2 \rangle = \sigma^2$, $\langle v \rangle = \sum \sigma^2 = N\sigma^2$, so that $\langle v \rangle^2 = N^2\sigma^4$.

The first component is less straightforward:

$$\langle v^2 \rangle = \langle (\sum x_i^2)^2 \rangle = \langle \sum x_i^2 \sum x_j^2 \rangle = \sum \sum \langle x_i^2 x_j^2 \rangle \quad (E9)$$

The expectation value of $x_i^2 x_j^2$ involves N center terms and $N^2 - N$ cross terms :
$(x_1^2 + x_2^2 + x_3^2 + ...)(x_1^2 + x_2^2 + x_3^2 + ...) = x_1^4 + x_2^4 + ... + x_1^2 x_2^2 + x_1^2 x_3^2...$

Thus $\langle x_i^2 x_j^2 \rangle = (N^2 - N) \int x^2 P(x) dx \int x^2 P(x) dx + N \int x^4 P(x) dx$.

Since $A = 0$, $P(x)$ is a Gaussian.

We evaluate

$$\int_{-\infty}^{\infty} \frac{x^2}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} dx \quad (E10)$$

substituting $u = x/\sqrt{2}\sigma$ and use the standard result $\int u^2 \exp(-\alpha u^2) du = \sqrt{(\pi)}/2\alpha^{3/2}$, so that : $\int x^2 P(x) dx = \sigma^2$.

similarly, for

$$\int_{-\infty}^{\infty} \frac{x^4}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} dx \quad (E11)$$

with identical substitution and the standard result $\int u^4 \exp(-\alpha u^2) du = 3\sqrt{(\pi)}/4\alpha^{5/2}$ we obtain : $\int x^4 P(x) dx = 3\sigma^4$.

Therefore:

$$\langle v^2 \rangle = (N^2 - N)\sigma^4 + N3\sigma^4 = N^2\sigma^4 + 2N\sigma^4 \quad (E12)$$

Finally,

$$Var\left(\frac{v}{N\sigma^2}\right) = \frac{Var(v)}{N^2\sigma^4} = \frac{1}{N^2\sigma^4}(N^2\sigma^4 + 2N\sigma^4 - N^2\sigma^4) = \frac{2}{N} \quad (E13)$$

REFERENCES

Ivezić Ž., et al., 2007, AJ, 134, 973
Sesar B., et al., 2007, The Astronomical Journal, 134, 2236
Sesar B., et al., 2010, ApJ, 708, 717

This paper has been typeset from a TeX/LaTeX file prepared by the author.