MNRAS **000**, 1–7 (2015) Preprint 30 April 2016 Compiled using MNRAS LATEX style file v3.0

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*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

1 INTRODUCTION

Many objects in the universe,from stellar to extragalactic scales, are variable on timescales less than a few hundred years. By repeated observations we can characterize the physical properties and relevant timescales of very faint objects. However, since all astronomical observations suffer from a detection threshold, a variable object may be undetectable for certain epochs, so that lightcurve becomes incomplete. This can be circumvented by co-addition of images, allowing to detect an object, but in this process loosing all time domain information. We restore the lightcurve performing at each epoch forced photometry performed at a precise location of a variable object known from the coadded images. A challenge inherent to forced photometry is an interpretation of faint, noise-dominated flux measurement. Hitherto a standard practice has ignored the wealth of information present at the faint end, applying a cutoff 1-2 magnitudes above the detection threshold magnitude. In order to fully utilize information present in time-domain surveys, such as Large Scale Synoptic Telescope, Palomar Transient Factory, or Sloan Digital Sky Survey, we need to address the issue of characterizing faint variable object, and how their properties are affected by our handling of the faint signal. A new methodology would allow an unbiased study of such faint variable objects, including quasars, RR Lyrae, Cepheids, and a wealth of other variable sources. With the advent of precision time-domain astronomy surveys it is crucial to apply the best possible faint forced photometry algorithms and thus make full use of the data.

2 DATA ANALYSIS

2.1 Data Overview

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 $\pm 1.27\,\mathrm{deg}$, extending from R.A. $\approx 20^h(320\,\mathrm{deg})$ to R.A. $\approx 4^h(55\,\mathrm{deg})$ (Sesar+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 $-40\,\mathrm{deg}\,(+320\,\mathrm{deg}) < RA < +10\,\mathrm{deg}$ and IN2P3 with $+5\,\mathrm{deg} < RA < +55\,\mathrm{deg}$. There is a $5\,\mathrm{deg}$ 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 Ivezic+2007 catalog (similar to Jiang+2014).

2.2 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 $3\,\sigma$. 8 million of those are brighter than 23^{rd} 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`).

2.3 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 23^{rd} mag.

Since an image for each epoch is background-subtracted, and the noise level fluctuates about zero, the flux

reported in an aperture at some locations may be smaller than zero.

This can be understood realizing that background in the optical bands is bright. We can calculate the likelihood of false detections, i.e. the background noise being confused for a source flux, in a following way. If we assume that the measured number of background counts oscillates around the value B_0, then the measured background count B is distributed as a Gaussian of width σ_B : $B - B_0$ $\mathcal{N}(0, \sigma_B)$. The noise is Poissonian, i.e. depends on the number of counts, and since for optical mesaurements the number of counts is large, $\sigma_B = \sqrt{B}$. Therefore, on a 4kx4k CCD, with 16 Mpix, at the 5σ level (corresponding to 1 false detection in a million), we would expect about 16 false detections.

For each patch the raw lightcurves contain the `id`, `objectId`, `exposure_id`, `mjd`, `psfFlux`, `psfFluxErr`, sorted by `objectId`, measuring flux in $[ergs/cm^2/sec/Hz]$ (`rawDataFPSplit/bandPatchStart_PatchEnd.csv`).

2.4 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:

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

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

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

$$\mathrm{WeightedMean} = \mathrm{WVar}\sum \frac{y}{\sigma_y^2} \tag{3}$$

$$\mathrm{WeightedStdCorr} = \left[\frac{\mathrm{WVar}}{N-1}\sum \frac{(y - \mathrm{WeightedMean})^2}{\sigma_y^2}\right]^{1/2} \tag{4}$$

From these metrics, we can calculate the catalog photometry :

$$\texttt{median\_mag} = -2.5\log_{10}\mathrm{median} - 48.6 \tag{5}$$

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`).

2.5 Faint Sources

When performing forced photometry we encounter sources that are very dim ($< 2\sigma$) in a single epoch image. Consider the distribution of the likelihood of flux measurement $L(F|data)$. For bright sources it is a very narrow Gaussian centered on the measured F_S, with width σ_F on the level of $1-2\% \approx 0.01 - 0.02$ mag. However, for faint sources that have larger measurement uncertainties, the Gaussian likelihood has much wider tails. When a tail extends below zero, it means that there is a nonzero likelihood of a negative flux measurement, which is unphysical.

As an example of prior choice, we consider a sinusoidally varying flux. If the flux of an object over many epochs is expected to vary in a sinusoidal fashion, i.e. $F(t) = F_{min} + sin(t)$, the probability of a given flux measurement is a cosine, from F_{min} to F_{max}. Thus, if no prior measurements were taken, we can confidently say that the flux of our object is most likely $(F_{min} + F_{max})/2$, i.e. at the peak of the cosine likelihood function. However, as soon as one measurement is taken of (F_i, e_i), the probability distribution of a flux of an object at that epoch becomes a convolution of our cosine prior information with the Gaussian curve of width e_i, centered on F_i (assuming Gaussian errors). However, without any a-priori information about the variability pattern of the considered object, the least informative Bayesian prior we can impose is a flat one : $p(F) = 0$ for $F < 0$, and 1 elsewhere. Thus the posterior probability is

$$p(F|data) \propto L(F|data)p(F) \tag{6}$$

Such flat prior would only affect the faint sources, by moving the posterior distribution above zero flux. The measurement of flux for bright sources would not be affected. We chose to apply the Bayesian prior for faint measurements where $\langle F_L \rangle < k\sigma$, with $k = 2$. This corresponds to the 2% probability of $F_L < 0$ assuming Gaussian likelihood.

To test our method we generate fiducial lightcurves (DRW / sinusoidal), with a uniform sampling ($N = 100 \div 1000$). Based on the generated flux (F_{true}) we define the 5σ level as the robust 25-th percentile (or median) of the ensemble F_{true} distribution : $\sigma_F = (1/5)F_{25\%}$ (in reality, σ increases for fainter observations, but this is a good approximation). We define $F_{obs} = F_{true} + F_{noise}$, where the Gaussian noise $F_{noise} = \sigma_F \mathcal{N}(0,1)$ was added to each point. For a weak signal, defined as $F^i_{obs} < 2\sigma F$, we consider $p(F)$ - a Gaussian likelihood associated with i-th measurement: $p_i(F) = \mathcal{N}(\mu = F^i_{obs}, \sigma = \sigma_F)$. Each measurement F_{obs} is a mean of this likelihood: $F_{obs} = \langle p_i(F) \rangle$. We call it $p(F)$ for short :

$$p(F) = \frac{1}{\sqrt{2\pi\sigma_F^2}}\exp\left(-\frac{(F-\mu)^2}{2\sigma_F^2}\right) \tag{7}$$

After imposing the Bayesian uniform prior $p(F)$ becomes a truncated Gaussian (without the negative part),

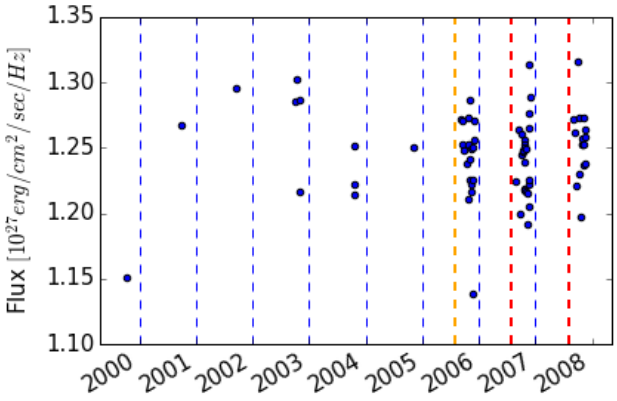

Figure 1. A plot showing an example lightcurve for an object id 217720894888346422. Jan 1st of each year (blue), August 1st of 2005 (orange) and August 1st of each subsequent year (red) is indicated by vertical dashed lines. Observations prior to August 1st of 2005 have sparser cadence, whereas those 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.

centered on F^i_{obs}, with a width σ_F. Thus for the truncated $p(F)$ the mean ceases to be F^i_{obs}, but it can be defined as

$$F_{mean} = \frac{\int_0^\infty F\,p(F)dF}{\int_0^\infty p(F)dF} \tag{8}$$

remembering to normalize by integrating over the Gaussian likelihood.

We also define the median as

$$\int_0^{F_{median}} p(F)dF = \int_{F_{median}}^\infty p(F)dF \tag{9}$$

Finally, since for a Gaussian distribution the area contained between $\mu \pm \sigma$ is 95.5% of the total area under the curve, for the truncated Gaussian we define the 2σ level as two areas $B = 0.05A$, or :

$$\int_{F_{2\sigma}}^\infty p(F)dF = 0.05 * \int_0^\infty p(F)dF \tag{10}$$

In those cases, normalization constant on both sides of equations cancels out.

(see Appendix)

2.6 Photometric colors

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. 2.6).

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

$$< F >= \frac{\sum w_i F_i}{\sum w_i} \qquad \sigma_{<F>} = \left(\sum w_i\right)^{-1/2} \tag{11}$$

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 lightcurves added in quadrature

2.7 Variability

Many lightcurves display intrinsic variability, in addition to the error-induced noise. A lightcurve consists of a set of N measurements of the object brightness x_i and associated 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, so that the distribution of measurements is a Gaussian. In this framework μ describes the median brightness, which for non-variable objects is the true brightness. Using the Bayesian approach, we find μ by maximizing the posterior probability distribution function (pdf) of μ given x_i and e_i : $p(\mu|x_i,\sigma_i)$. We can proceed analoguously to find the width of the distribution, σ, which describes the departure from the mean.

To find μ and σ, we follow Ivezic+2014, with the two-step approach. 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. The maximum of the 2D likelihood becomes our full solution - σ_{full} and μ_{full}, as shown on Fig. 2.7 (see Appendix B for the detailed calculation).

For each lightcurve, 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) :

$$\chi^2_{dof} = \frac{1}{N-1}\sum\left(\frac{x_i - < x_i >}{e_i}\right)^2 \tag{12}$$

and

$$\chi^2_R = 0.7414(Z_{75\%} - Z_{25\%}) \tag{13}$$

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

Initially, we evaluate μ_{full}, σ_{full}, χ^2_{dof}, and χ^2_R for the entire lightcurve. Then, only if either $\sigma_{full} > 0$ or $\chi^2 > 1$, which hints some intrinsic variability, we also calculate mu_{full}, σ_{full}, and χ^2 for the seasonally-binned portions of the lightcurve.

On Fig. 2.7 we plot the stages of calculating μ and σ. Left and middle panels compare two methods of calculating our variability parameters. The initial approximation (dashed) is based on bootstrapped resampling of x_i, e_i points from the lightcurve. With bootstrapping we resample the lightcurve M times, so that instead of a single sample with $N \approx 10-70$ points we have M samples. Random resampling means that points may be chosen multiple times in a random

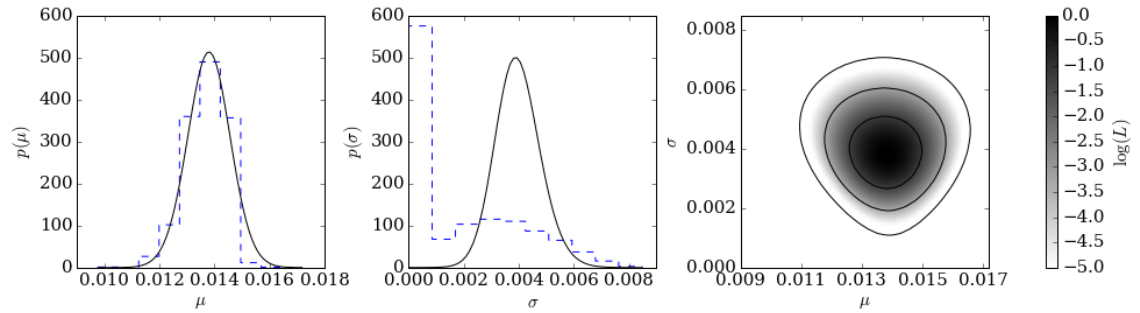

Figure 2. Two-step approach to finding μ and σ via μ_0 and σ_0 for an object 217720894888346446. In this calculation we use raw psf flux, before employing the faint source treatment outlined in Section 2.5. On the left and middle panels, solid lines trace marginalized posterior pdfs for μ and σ , while 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.

fashion. This allows us to calculate M approximate values of μ,σ. Dashed lines trace the histogram for $M = 1000$ resamples of our variability parameters. The approximate values are in turn used to provide bounds for the $200x70$ grid of μ, σ on which we evaluate the full posterior likelihood density function (right panel). This ensures that we are sampling the peak of the underlying distribution.

2.8 Extinction Correction

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]

3 RESULTS

4 CONCLUSIONS

ACKNOWLEDGEMENTS

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

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

APPENDIX A: TREATMENT OF FAINT SOURCES

Using our definition, the mean is :

$$F_{mean} = \frac{\int_0^\infty F\, p(F) dF}{\int_0^\infty p(F) dF} = I_0 / I_1 \tag{A1}$$

We evaluate

$$I_0 = \int_0^\infty \frac{F}{\sqrt{2\pi\sigma_F^2}} \exp\left(-\frac{(F-F_{obs})^2}{2\sigma_F^2}\right) dF = \frac{\sigma_F}{\sqrt{2\pi}} \exp\left(-\frac{F_{obs}^2}{2\sigma_F^2}\right) + F_{obs}\, \mathrm{sf}\left(\frac{-F_{obs}}{\sigma_F}\right) \tag{A2}$$

and

$$I_1 = \int_0^\infty p(F) dF = \int_0^\infty \frac{\exp\left(-\frac{(F-F_{obs})^2}{2\sigma_F^2}\right)}{\sqrt{2\pi\sigma_F^2}} dF = \mathrm{sf}(-F_{obs}/\sigma_F) \tag{A3}$$

so that

$$x_{mean} = \frac{\exp(-x_{obs}^2/2)}{\mathrm{sf}(-x_{obs})\sqrt{2\pi}} + x_{obs} \tag{A4}$$

where we scaled F_{obs} by σ_F (i.e. $F_{mean} = x_{mean} \cdot \sigma_F$).

To find the median and the 2σ level we transform from F space to x space , scaling by σ_F, so that $x = F/\sigma_F$, and thus the likelihood $p(x) \sim \mathcal{N}(x_{obs}, 1)$ is :

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-x_{obs})^2}{2}\right) \tag{A5}$$

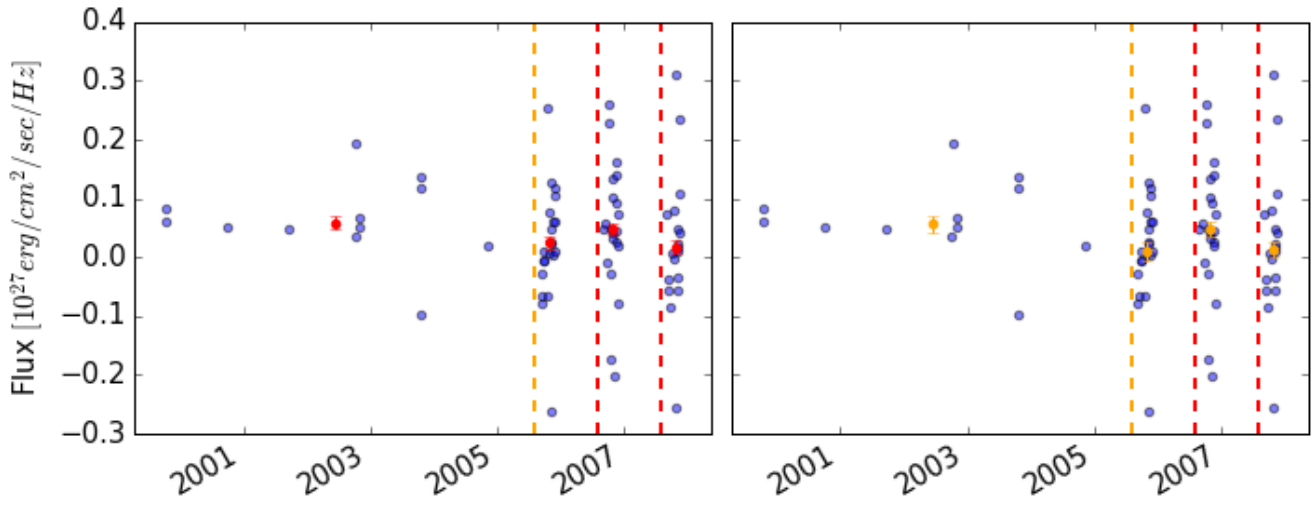

Figure 3. A plot showing an outcome of seasonal averaging for an object id 217720894888346425. The left panel (red dots) shows (mean, meanErr), and the right panel (orange) shows (median,medianErr), instead of seasonal points (blue). Vertical dashed lines as on Fig. 2.6

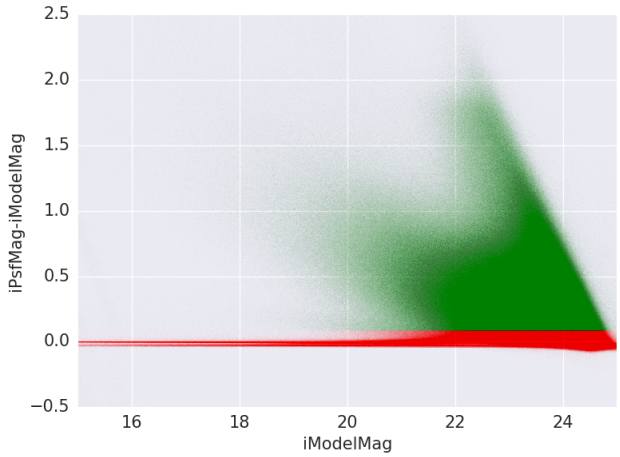

Figure 4. A plot showing NCSA sources detected in coadds, removing the outliers beyond the edges of the plot. The coloring corresponds to the `extendedness` parameter calculated in the pipeline based on the iPsfMag-iModelMag : red being 0 (compact), and green being 1 (extended). As iModelMag increases, the separation becomes less certain, as more distant galaxies are more compact.

Then we transform from x to z space, with a translation by $x_{obs} = F_{obs}/\sigma_F$: $z = x - x_{obs}$, so that now $p(z) \sim \mathcal{N}(0,1)$:

$$p(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(z)^2}{2}\right) \tag{A6}$$

In z-space, the median from

$$\int_0^{x_{med}} p(x)dx = \int_{x_{med}}^{\infty} p(x)dx \tag{A7}$$

becomes

$$\int_{x_0}^{z_{med}} p(z)dz = \int_{z_{med}}^{\infty} p(z)dz \tag{A8}$$

with $x_0 = -x_{obs}$

This expression for z_{med} can be evaluated : rhs is a survival function (sf) = 1 - cumulative density function (cdf) :

$$\int_{z_{med}}^{\infty} p(z)dz = \mathrm{sf}(z_{med}) \tag{A9}$$

and the lhs, assuming $z_{med} > x_0$, is :

$$\int_{x_0}^{z_{med}} p(z)dz = \int_{-\infty}^{z_{med}} p(z)dz - \int_{-\infty}^{x_0} p(z)dz = \mathrm{cdf}(z_{med}) - \mathrm{cdf}(x_0) \tag{A10}$$

Rearranging, and using the percent point function (ppf) - the inverse of the cdf, we find:

$$z_{med} = \mathrm{ppf}\left(\frac{1+\mathrm{cdf}(x_0)}{2}\right) \tag{A11}$$

and transforming back to F space:

$$F_{med} = F_{obs} + \sigma_F \ \mathrm{ppf}\left(\frac{1+\mathrm{cdf}(x_0)}{2}\right) \tag{A12}$$

with x_0 and x_{obs} as above. We also normalize this expression by $\int_0^{\infty} p(F)dF$ and σ_F:

$$x_{med}^{norm} = x_{obs} \tag{A13}$$

In z space , the 2σ areas A and B are :

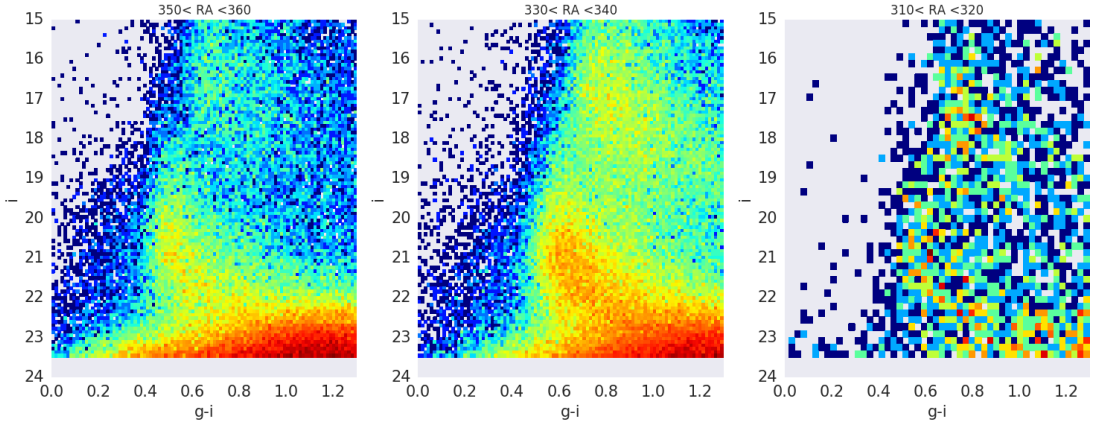

Figure 5. A color-magnitude plot , reproducing the results of Sesar+2010 , Fig.23 . We show here only NCSA-processed sources, which is why certain RA ranges are omitted or have less sources. We only select sources with `extendedness=0` parameter (stars). The scale is showing the $\log_{10}$ of count. All sources have their colors corrected for extinction. On first two panels the features of Sagittarius Stream are clearly visible.

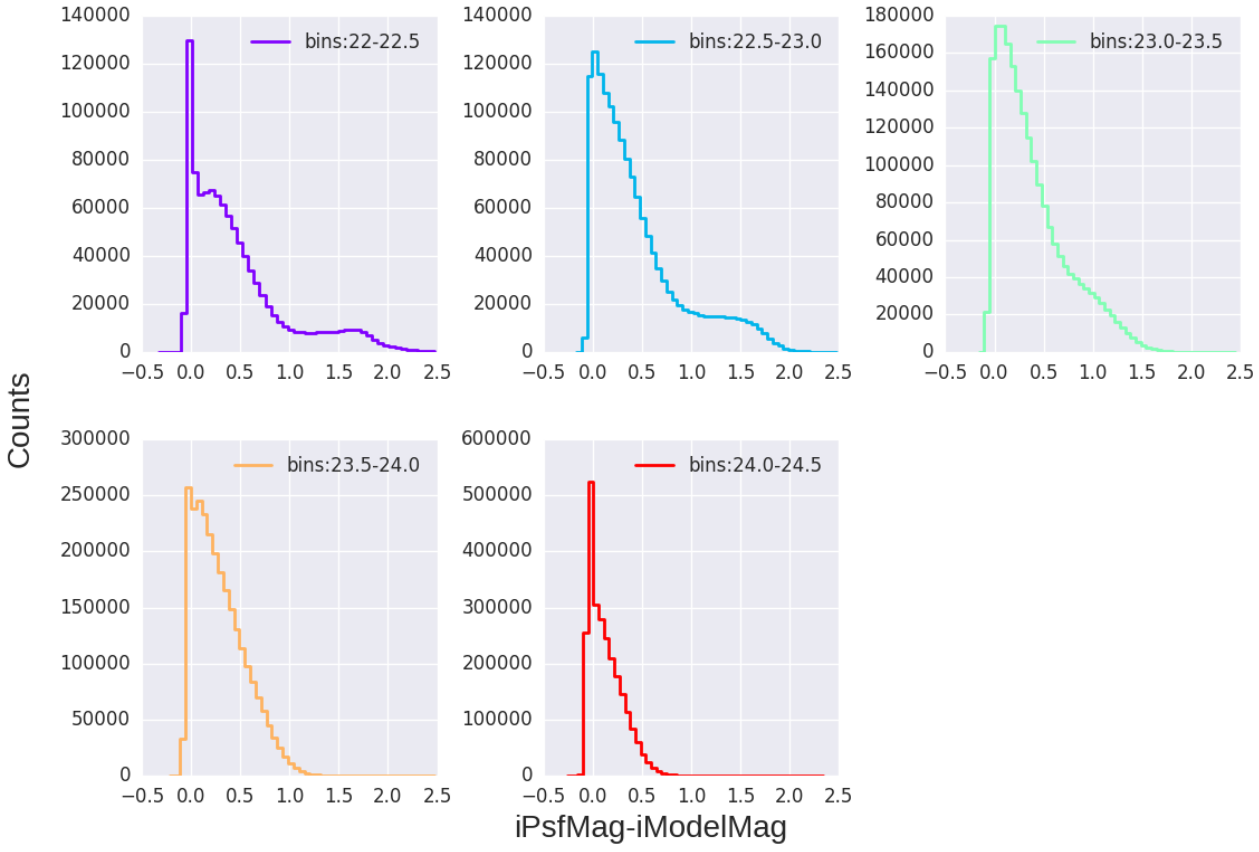

Figure 6. The histograms show the count of sources in 5 magnitude bins, corresponding to the vertical cut through Fig. 2.8. It helps to verify how well can the extended and compact sources be separated based solely on the iPsfMag-iModelMag

A = sf(x_0) and B = sf(z_B), so to find z_B we use the inverse survival function isf : z_B = isf(0.05A). Thus transforming back to F-space we have:

$$F_{2\sigma} = F_{obs} + \sigma_F\,(\,\mathrm{isf}(\,0.05\,\mathrm{sf}(x_0)) \tag{A14}$$

APPENDIX B: CHARACTERIZING VARIABILITY

We further characterize the variability of lightcurves 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 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 :

$$\sigma_0 = (variance_{approx})^{1/2} = (\zeta^2 \cdot \sigma_G^2 - e_{50}^2)^{1/2} \tag{B1}$$

where

$$\zeta = \frac{median(\widetilde{\sigma_i})}{mean(\widetilde{\sigma_i})} \tag{B2}$$

and

$$\widetilde{\sigma_i} = (\widetilde{variance})^{1/2} = (\sigma_G^2 + e_i^2 - e_{50}^2)^{1/2} \tag{B3}$$

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

$$\log L = -0.5 \sum \left(\ln(\sigma^2 + e_i^2) + \frac{(x_i - \mu)^2}{(\sigma^2 + e_i^2)} \right) \tag{B4}$$

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

$$L = e^{\log L - max(\log L)} \tag{B5}$$

We then marginalize over μ or σ :

$$p(\sigma) = \sum_{\mu} (L_{\sigma,\mu}) \qquad p(\mu) = \sum_{\sigma} (L_{\sigma,\mu}) \tag{B6}$$

and normalize the probability :

$$p_{norm}(\sigma) = \frac{p(\sigma)}{\int p(\sigma) \mathrm{d}\sigma} \qquad p_{norm}(\mu) = \frac{p(\mu)}{\int p(\mu) \mathrm{d}\mu} \tag{B7}$$

To characterize lightcurve variability we first calculate for the entire lightcurve of an object the approximate μ_0 and σ_0 using bootstrapped resampling of the lightcurve. 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. 2.7).

APPENDIX C: 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 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. 3 for an example).

APPENDIX D: 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 $< 2\sigma$: $F_{2\sigma}$.

This paper has been typeset from a TEX/LATEX file prepared by the author.