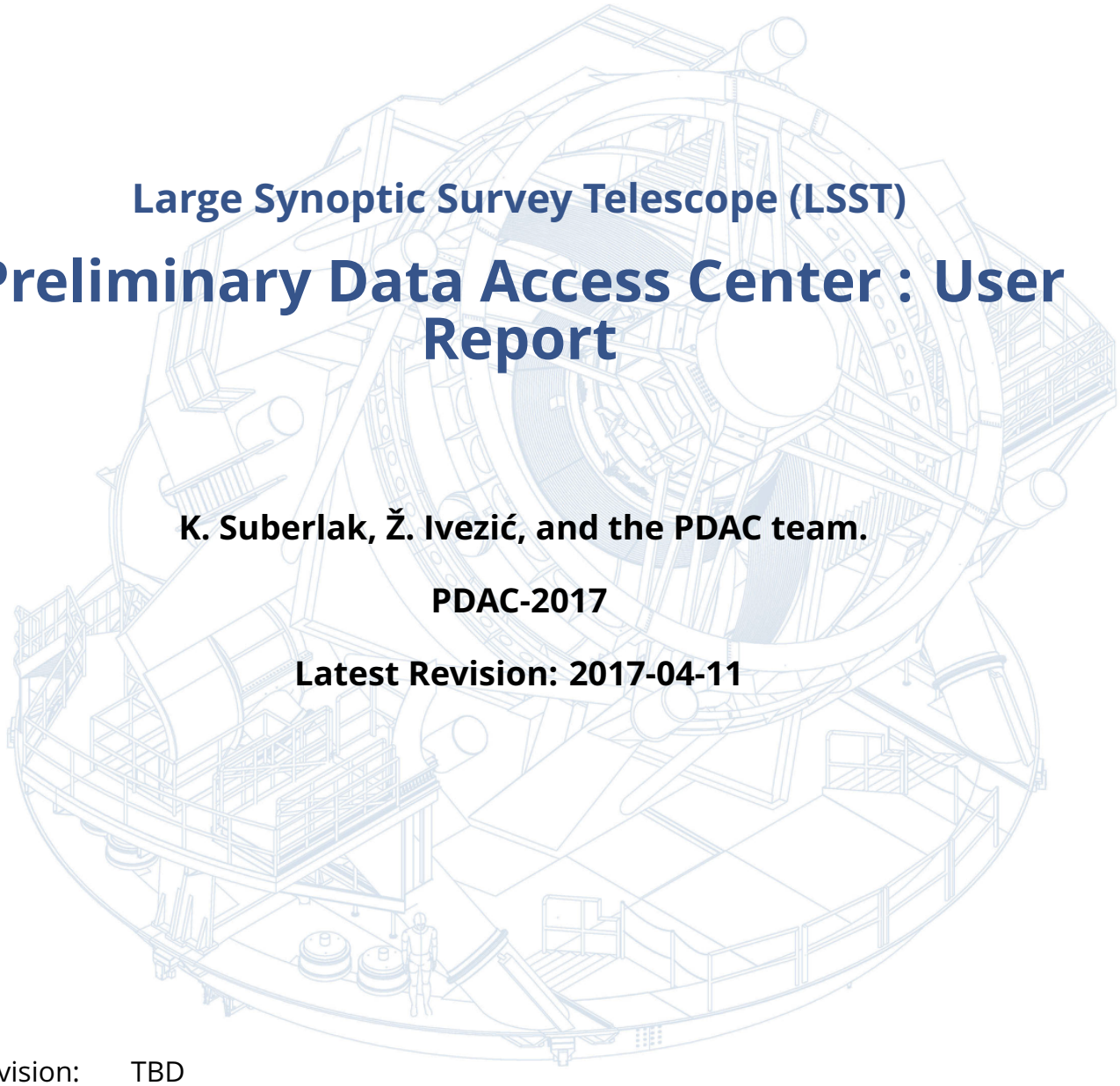
LARGE SYNOPTIC SURVEY TELESCOPE



Large Synoptic Survey Telescope (LSST)

Preliminary Data Access Center : User Report

K. Suberlak, Ž. Ivezić, and the PDAC team.

PDAC-2017

Latest Revision: 2017-04-11

revision: TBD
status: draft

Abstract

A report on user experience of the Preliminary Data Access Center (PDAC). We test the quality and ease of access to the data. PDAC will pave the way to the Science User Interface and Tools (SUIT). We employ both in-detail study of individual objects, and a statistical study of an ensemble of objects. We evaluate user-friendliness of the current interface, and make recommendations for its future improvements.

Change Record

Version	Date	Description	Owner name
1	2017-02-15	First draft.	Krzysztof Suberlak
2	2017-03-10	Reordered sections.	Krzysztof Suberlak
3	2017-04-03	Added Time Series UI description.	Krzysztof Suberlak

Contents

1 Introduction

This is a document to report on the user experience testing of the Preliminary Data Access Center. The Large Scale Synoptic Telescope (LSST) will produce a big volume of data. Such unprecedented data stream poses new challenges to provide an easy access for users, in such a way that they can quickly find what they need, and thus be able to focus on the science goal that they would like to achieve. The detail description of such online user-interface called Science User Interface and Tools is outlined in documents LDM-130 (SUIT requirements) and LDM-492 (SUIT Vision). An idea of having an interface to the data is not new : there exists Aladin, SDSS CAS jobs, IPAC IRSA, Mikulsky NASA Archive, NED, and many other archives. These allow a user to query for data (either via SQL query, or interface), returning the data table. Some user interfaces (eg. IRSA) have some rudimentary plotting capabilities. There have been ideas of a new interface, that would not only eg. plot the lightcurve and display the spectrum, but also allow the user to run some machine learning algorithms, or simple models that can help narrow down the query, or obtain science results in the browser. Namely, Victor Pankratius, from MIT, in his talk "Computer-Aided Discovery: Towards Scientific Insight Generation with Machine Support" outlined the idea of an ipython notebook - access to data, which lives in the cloud, is allocated some CPU share and memory, and allows one to upload / download the data and run the model in real time, which is especially helpful to geoscientists doing fieldwork, where new data acquisition conditions their next step.

These requirements and the vision for SUIT have been further described on confluence pages[1]. Some technical notes about current implementation of SUIT by PDAC are also available via confluence pages[2].

This report details tests and queries employed, including screenshots and data-based plots. A shorter summary of monthly progress is released every month at the github repository of the LSST System Science Team : `https://github.com/lsst-dmsst/PDAC_report`.

2 Overview of performed tests

We test a variety of aspects of PDAC : the user interface, infrastructure, and database ingestion. The user interface is similar to IRSA, which aids the ease of access. In Section 3 we

[1] `https://confluence.lsstcorp.org/display/DM/Science+User+Interface+and+Tools`
[2] `https://confluence.lsstcorp.org/display/DM/Guide+to+PDAC+version+1`

describe the functionality available through user interface. It is a work in progress, hence any deficiency outlined may become updated in real-time, whereas some recommendations, if met with approval, may have a longer implementation timescale. In Section 4 we describe the structure of available data : both data that is available directly from NCSA (internal catalogs) , and data that is available from IRSA (external catalogs). In that section we also provide an overview of query and analysis methods available directly through the User Interface, as well as through SQL. Finally, in Section 5 we consider the quality of database ingestion, answering the question of how well was a given dataset loaded into PDAC. In particular we compare the S82 forced photometry dataset, an outcome of the Summer 2013 reprocessing, to the same data stored locally at the University of Washington.

3 User Interface: what we see

In order to access PDAC we follow the directions[2] that include logging to NCSA via VPN `https://vpn.ncsa.illinois.edu/` using Cisco AnyConnect Secure Mobile Client, and opening in the web browser `http://lsst-sui-proxy01.ncsa.illinois.edu/suit`. This opens the main interface screen, which allows to select the database, and perform the desired query.

Currently, PDAC v1, in the upper-left corner of the interface, under tab 'LSST Data' (see Fig. 1) includes the Summer 2013 DM-stack reprocessed SDSS Stripe 82 data (database `sdss_stripe82_00`), hosted at the NCSA on the LSST prototype ("integration cluster") hardware, in Qserv [Gregory Dubois-Felsmann, priv.comm. 02-20-2017, slack]. The only other locally stored database (as of March 2017), is WISE catalog, that is not yet accessible via the graphical user interface (it can be queried as Data Base `wise_00`, with catalogs 'Object' containing objects (like DeepSource in S82 above), and 'ForcedSource' containing forced photometry (like DeepForcedSource in S82)).

The upper-left corner of the interface also leads to 'External Images' and 'External Catalogs'. The Catalogs are all NASA/IPAC[3] Infrared Science Archive(IRSA) publicly accessible catalogs, including GAIA, WISE, 2MASS, SPITZER, etc. (see Fig. 2).

[3] Infrared Processing and Analysis Center, `http://www.ipac.caltech.edu/project/lsst`

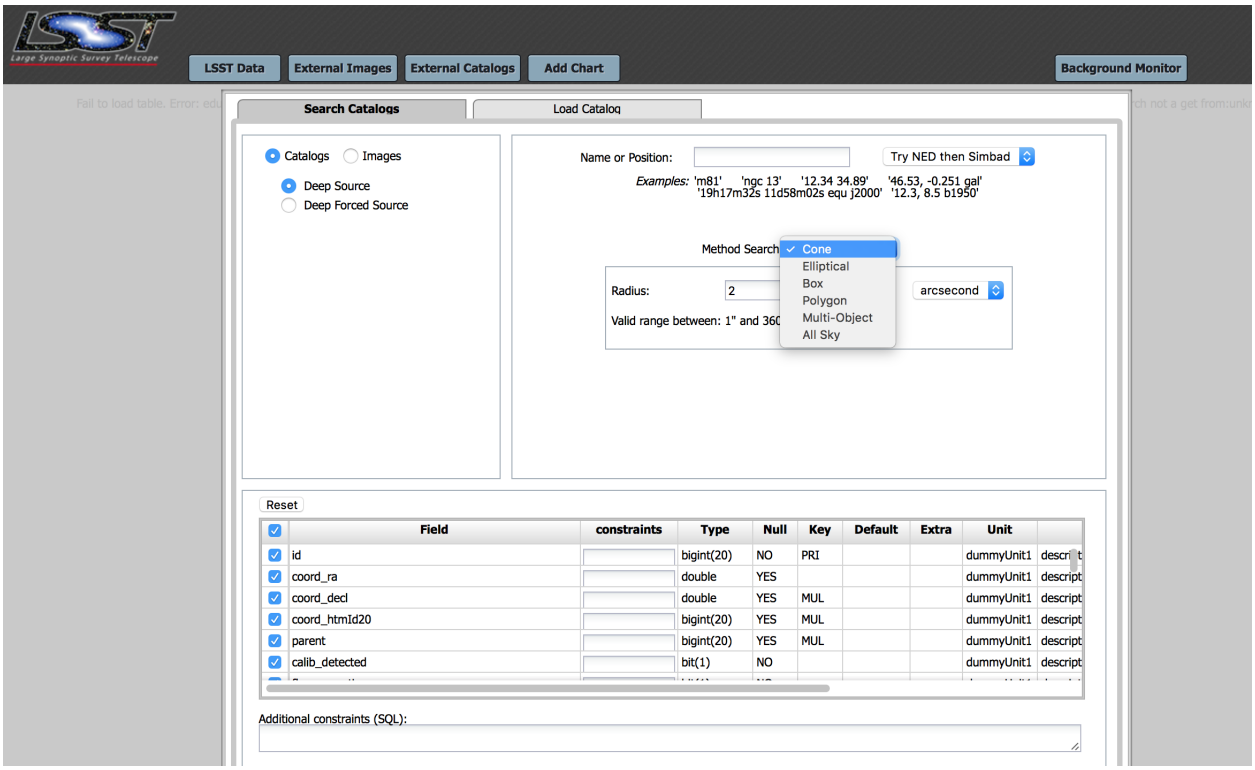


FIGURE 1: The main user interface of PDAC ver. 1. As of April 2017, Multi-Object and All Sky queries are not yet available. The 'Name or Postion' only resolves positive RA (0<RA<360), while using direct SQL query resolves both positive and negative RA (-180 < RA < 180). Currently this is an inconsistency that we recommend to be addressed in the future. Furthermore, the names resolved have to be consistent with those present in NED or Simbad databases - any id's from the database queried (eg. 'id' in RunDeepSource, or 'objectId' in RunDeepForcedSource) are not yet resolved.

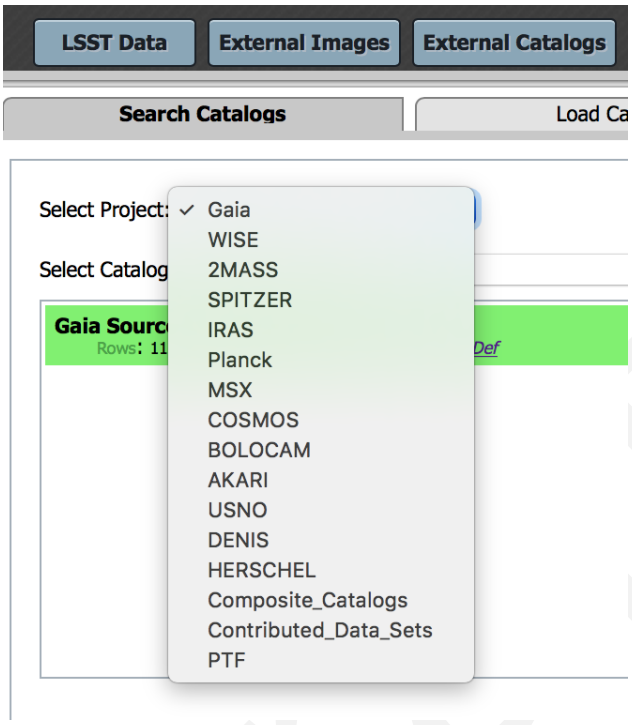


FIGURE 2: IPAC- hosted catalogs , accessible via IRSA.

4 Infractructure : what is available and how to get it

As we described in Section 3, the main user interface allows access to the internally stored (at NCSA) SDSS Stripe 82 data reprocessed during the Summer 2013 [4] as part of Data Challange with the continuously developed LSST Stack [5].

The reprocessing included:

- coadding the data from all epochs in each of the ugriz SDSS filters. Measurements on coadds (per object) are available as `RunDeepSource` table, accessible via Catalogs –> 'DeepSource' . The single-band coadded images with MariaDB metadata are available as `DeepCoadd` table, accessibla via Images –> 'DeepCoadd' .
- using i-band detections to seed forced photometry on all epochs in all bands. The results of photometry are available as `RunDeepForcedSource` table, accessible via Catalogs –> 'Deep Forced Source' .

[4] `https://confluence.lsstcorp.org/display/DM/Properties+of+the+2013+SDSS+Stripe+82+reprocessing`
[5] `https://pipelines.lsst.io/index.html`

- For reference , the individual calibrated single epoch images are available as `Science_Ccd_Exposure` table, accessible via Images –> 'Science CCD Exposure'

Additional details of the schema are also outlined in the LSST Data Challenge Report [Shaw, Juric, Becker, Krughoff et al. 2013], and the LSST Database Schema Browser [6].

Spatial queries that can be directly executed from the PDAC interface, called 'Method Search', include cone, box, elliptical and polygon (See Fig. 1). Spatial queries allow to choose a certain region of the sky by the object ra,dec coordinates. Cone, elliptical, and box queries return objects in a region of the sky bound by a geometrical shape centered on given coordinates (ra,dec). Cone is the most useful type of query, allowing to find objects within a certain radius from the coordinate query. Elliptical search allows to define the shape by an ellipse with a given semi-major axis, position angle and the axis ratio. A box is a square centered on the query coordinates, with a given side size. A polygon allows to define the search region by between 3 and 15 coordinate pairs (vertices of the polygon). Note : Multi-object query is listed in the drop-down menu, but has not yet been implemented (March 2017) - in the future it will allow the user to upload a list of ra,dec and search radii, finding 1-to-1 matches in the existing catalog. An All-Sky option (no spatial constraints) has not been tested given the size of the database.

Any query returns a list of all objects within the given region (Fig. 3).

A certain limitation of the main UI is inability to resolve id's from the database itself (see Fig. 1). Indeed, the only way to find which objects have been detected in a certain small region in DeepSource coadds, and download light curves only for one of them from DeepForcedSource forced photometry catalog, is to use an SQL constraint. For example, we performed cone query against DeepSource table for ra,dec = 0.283437, 1.178522, 2 arcsec search radius (this is the RR Lyrae ID=13350 also investigated in Sec. 4.2). Limiting the results to [`id , coord_ra , coord_decl, flux_psf , coadd_id , coadd_filter_id`], we find that there is a coadd for each filter (denoted with `coadd_filter_id`). The identification in i-band coadd (`coadd_filter_id=3`) is id=3588818166880604. Note that while DeepSource has a separate id for a coadd in each band, only id's for i-band coadd are inherited by DeepForcedSource catalog. The DeepSource.id == DeepForcedSource.objectId, because DeepForcedSource.id stands for forced photometry detection id, which is unique for each epoch. Therefore a single object has one DeepSource.id, equal to DeepForcedSource.objectId, but multiple Deep-

[6] `https://lsst-web.ncsa.illinois.edu/schema/index.php?t=DeepForcedSource&sVer=S12_lsstsim`

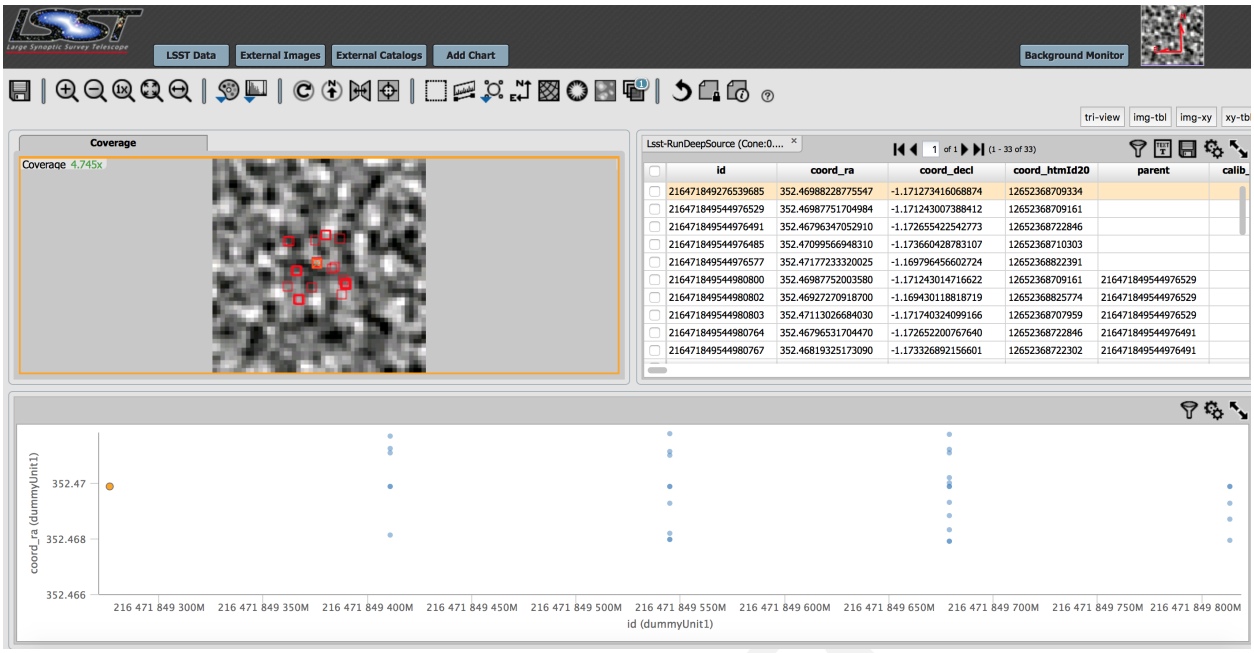


FIGURE 3: Example cone query against Deep Source table, returning all objects detected within a radius of 10 arcseconds from the position ra=352.469872 , dec=-1.171239. Note that the background image (postage stamp miniature) does not show the actual S82 coadds. This particular feature is described further in Sec. 4.1

ForcedSource.id - we recommend to highlight this in the metadata for it is a potential area for confusion. The only way to currently recover a lightcurve for a single object from DeepForcedSource is to first select the detection id in DeepSource, and use that as a constraint when using cone query on DeepForcedSource (see Fig. 4)

4.1 Postage Stamp Miniatures

We compared the postage stamp miniatures showing the overview of the region against which a given query was performed. We find that the miniature image does not always come from the catalog we query against. In fact, the "coverage" image comes from IRAS, DSS, 2MASS, or WISE - the survey is chosen depending on the size of the region needed to be shown [Xiuqin Wu, priv.comm., 2017]. Indeed, as the query region is increased, the shown image changes unexpectedly from DSS to IRAS or WISE, without issuing a relevant information to the user. A recommendation is to display information about which sky survey a particular image is coming from.

As a concrete example, we execute a cone query against the Deep Source table, expecting to

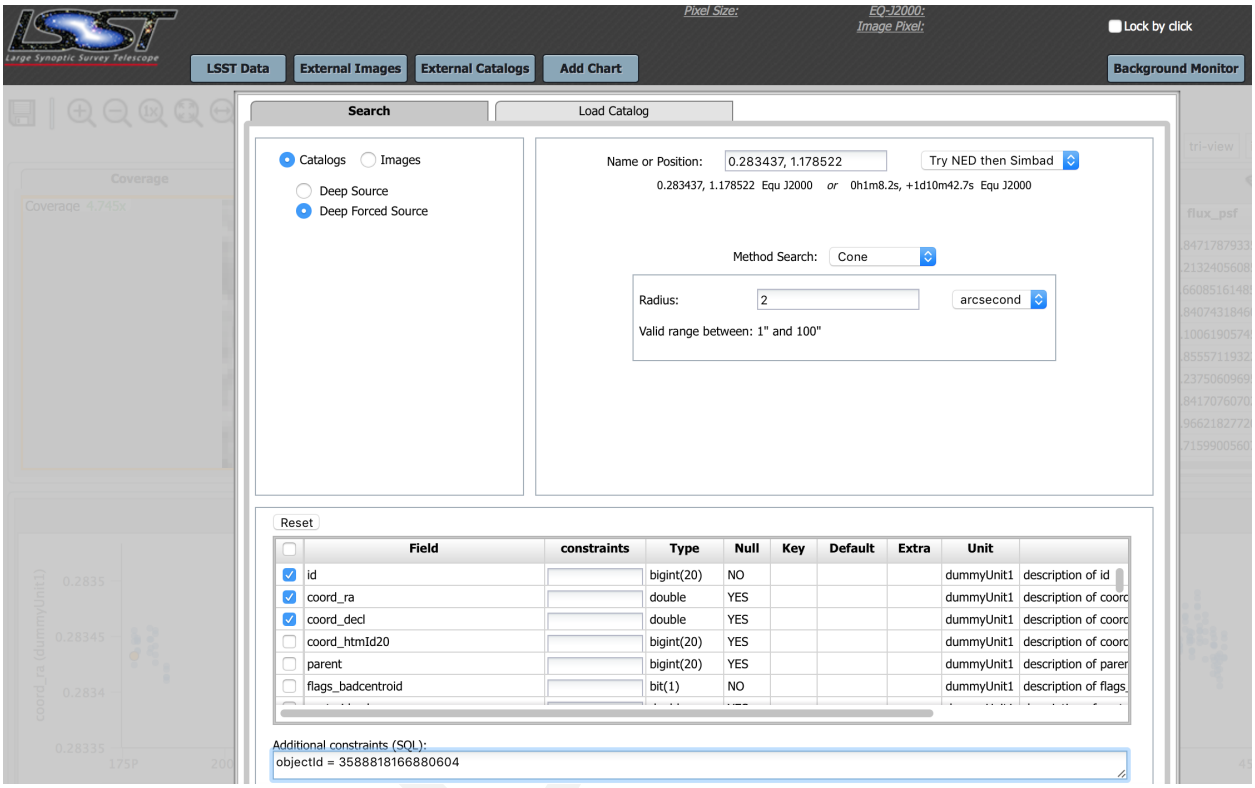


FIGURE 4: The correct way to select a light curve for a particular object from the forced photometry DeepForcedSource catalog. Here we first queried against the DeepSource catalog to find id's for objects detected in coadds in a small region within 2 arcsec from ra,dec = 0.283437, 1.178522. For i-band there is only one id : 3588818166880604. Since DeepSource.id = DeepForcedSource.objectId, we require objectId to be equal to 3588818166880604. Thus we are able to acces forced photometry for precisely one object. Otherwise, obtaining a light curve from a direct spatial query of DeepForcedSource would provide all photometry for all objects detected in coadds within the search radius, which may not be the desired beavior for analysis of Time Series. We recommend that the result of spatial quary against RunDeepForcedSource should contain a summary of which unique objectId's are present, with an ability to select only one object (with multi-band photometry), if more than one is present in the search region. Otherwise it becomes a long-winded process to first find what id's were detected in coadds (DeepSource), to then select id for i-band coadd, and select only rows corresponding to that objectId in RunDeepForcedSource.

find a point source: a star at ra,dec = 23h30m57.31s, +1d1m13.8s (or 352.73878967464526 , 1.020496777615987 degrees), with search radius of 2, 10, 100 and 1000 arcseconds . We find that the queried region only sometimes has the postage stamp miniature aligned with the query region.

We also consider a galaxy located at ra,dec = 40.433, 0.449. A 10 arcsec cone query only sometimes is centered on the galaxy.

4.2 Time Series : periodogram

We compare the g-band time-series of a few objects, taking as the ground truth periodicities reported by [2]. We record the execution time, and describe in detail the experience of exploring the periodogram tool.

As a test case we consider a RR Lyrae star ID=13350, located at ra, dec = 0.283437°, 1.178522°. [2] did detailed template fitting, finding the period to be 0.547987 days. To analyze the light curve of this object, we Cone query the Deep Source catalog to find objectId within this location. A cone query with search radius of 2 arcseconds brings us to a lightcurve view, from where it is possible to open the Time Series view (see Fig.5). Figs. 6 and 7 guide through the steps allowed by this UI to calculate the correct period of this RR Lyrae star. The backend of the periodogram tool is a clone of the NASA Exoplanet Archive [Xiuqin Wu 2017, priv.comm.]. The PDAC implementation has been internally tested to comply with that original toolset, but there are is no documentation about details of the algorithm used to compute the periodogram powers beyond `http://exoplanetarchive.ipac.caltech.edu/docs/pgram/pgram_parameters.html`. We recommend that there would be a link to this page to provide the user of the periodogram information about the content of what is plotted in the UI.

Using the periodogram tool we find that there are a few glitches awaiting improvement, eg. the period slider ticklabels not adapting well to changing the period limits. Also, setting the Periodogram minimum and maximum period does not automatically set the slider bounds.

To find the correct period for the test case RR Lyrae star we needed to guide the Periodogram tool by entering the minimum and maximum periods - using the defaults we fail to recover the correct period (Fig. 8) . We set the periodogram minimum and maximum search period using the known range of periods for RR Lyrae in [2]. In this study we set P_{min} as 90 % of

the smallest period, and P_{max} as 110 % of the largest period : $P_{min} = 0.9 * 0.254 = 0.229$ and $P_{max} = 1.1 * 0.907 = 0.998$ days. Using these bounds we recover the correct period with the Time Series View periodogram tool (see Fig. 9).

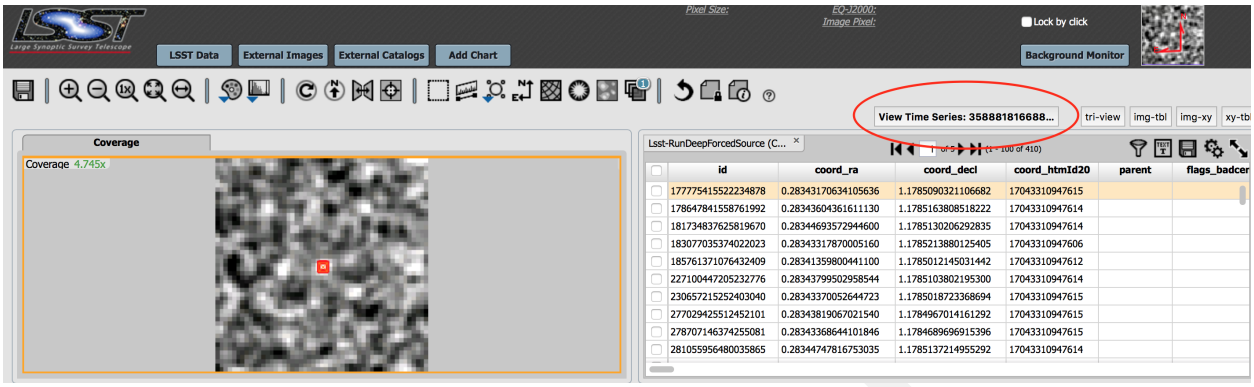


FIGURE 5: A result of cone query against Run Deep Forced Source table (containing S82 S13 data), with (ra,dec,radius = 0.283437, 1.178522, 0.00055 degrees). We circle the 'View Time Series' button that links to the Time Series UI shown on Fig. 6.

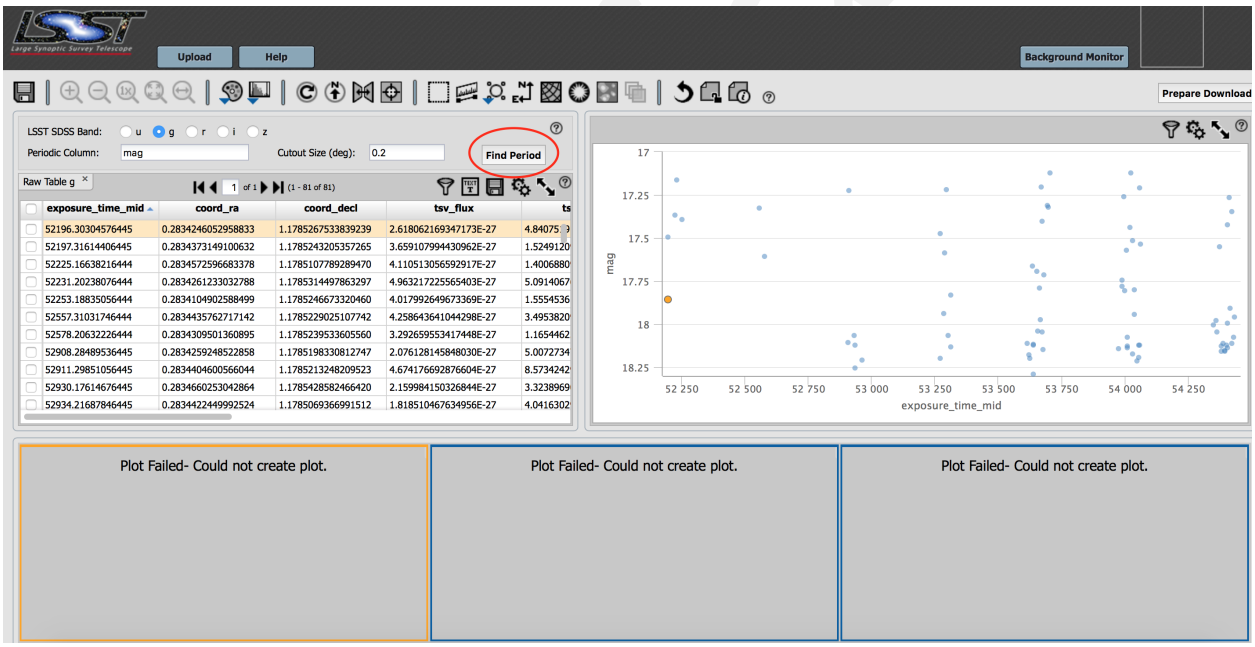


FIGURE 6: Time Series view for an RR Lyrae ID=13350 star at ra, dec = 0.283437°, 1.178522°. Note that initially on the bottom there are three empty panels. The radio buttons in the upper left corner allow intuitively to select SDSS filter for lightcurve periodogram calculation (multi-band periodogram as in [4].) We select 'Find Period', marked with red oval, to calculate Lomb-Scargle periodogram for that band (this takes the user to Fig 7)

Finally, we compare the PDAC Time Series User Interface to that of the NASA Exoplanet

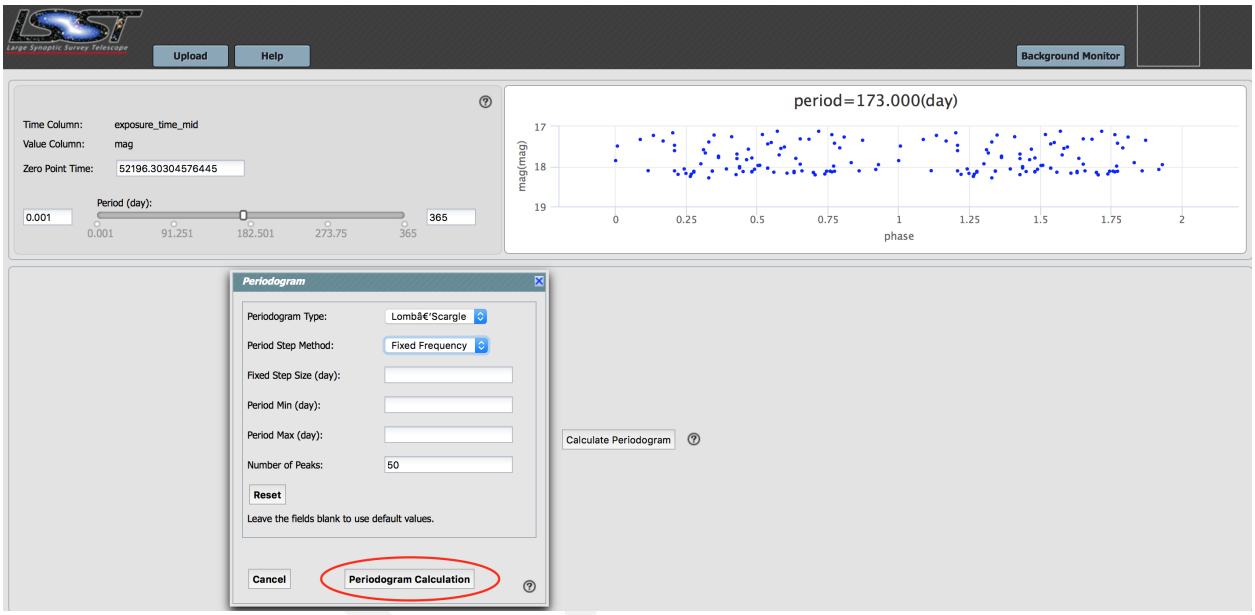


FIGURE 7: Calculating Lomb-Scargle periodogram for RR Lyrae ID=13350 at ra, dec = 0.283437°, 1.178522°. The slider in the upper left corner allows to fold the lightcurve on the chosen period. Clicking 'Calculate Periodogram' button opens the dialog window 'Periodogram'. Currently it contains only the Lomb-Scargle as Periodogram Type. Period Step Method include Fixed Frequency or Fixed Period, similar to the NASA Exoplanet Periodogram Tool (Fig. 12). If we don't choose anything for maximum and minimum periods, the calculation will proceed with defaults, which for this RR Lyra fail to detect the true period (Fig. 8). If we choose the minimum and maximum periods knowing what period to expect for a given class of object, we are more likely to detect the true period (9). For this particular RR Lyrae we chose 0.229, 0.998 days as limits on period, which corresponds to the range of periodicities in [2] sample of 483 RR Lyrae. Clicking on 'Periodogram Calculation' proceeds with evaluating Lomb-Scargle periodogram with chosen Period Step Method

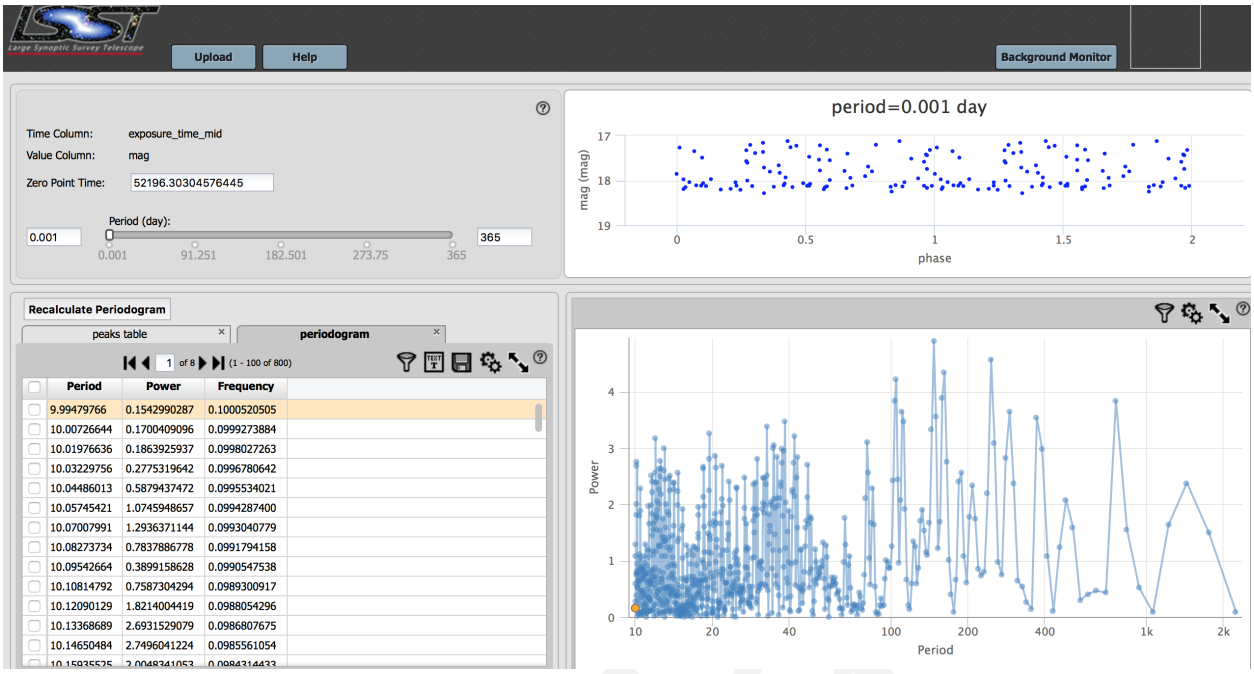


FIGURE 8: Calculating Lomb-Scargle periodogram for RR Lyrae ID=13350 at ra, dec = 0.283437°, 1.178522°. Not choosing the minimum and maximum, but letting the backend choose the defaults, does not recover the true underlying period of 0.547987 days.

Archive Periodogram[7] (see Fig. 12). Using few RR Lyrae PDAC g-band lightcurves, each calculation is allocated time slot of approximately 15 seconds. Also see Table 2 for a summary of results.

5 Database Ingestion : is what we get what we expected to get?

5.1 Positional comparison : box query

We compare the ra,dec of objects found within a given region in PDAC S82 to the analoguous data stored at UW (DeepSource tables). Regardless of photometry, for a given objectId, the ra,dec should agree.

[7] http://exoplanetarchive.ipac.caltech.edu/cgi-bin/Pgram/nph-pgram

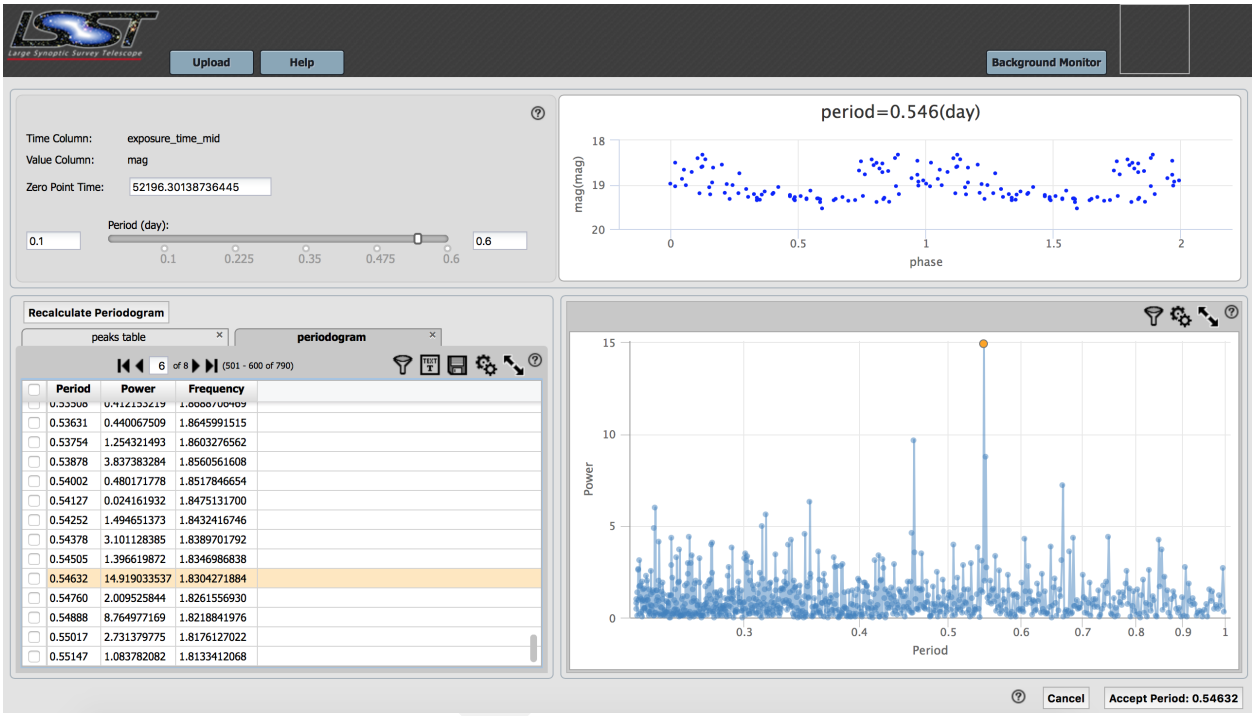


FIGURE 9: Calculating Lomb-Scargle periodogram for RR Lyrae ID=13350 at ra, dec = 0.283437°, 1.178522°. When we appropriately constrain the periodicities for which the LS power is calculated, we recover the true period of 0.547987 days. One choice of bounds is to set $P_{min} = 0.229$ and $P_{max} = 0.998$ days, which are 90% of the smallest and 110 % of the largest RR Lyrae periods in [2] sample. Note that as of April 2017, the minimum and maximum value of a slider allowing to interactively fold the lightcurve on any period does not update to the values used in the Periodogram search. Clicking 'Accept Period' takes the user to Fig. 10

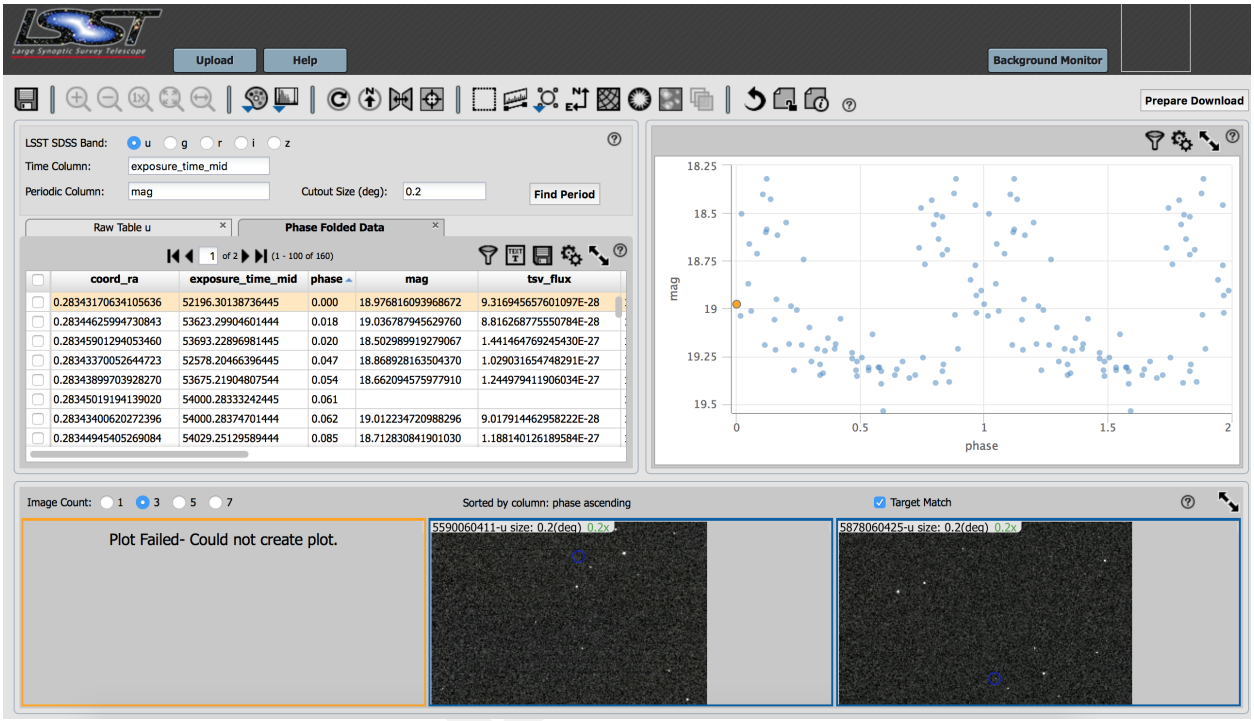


FIGURE 10: The result of accepting the period found by limiting the periodogram bounds by $P_{min} = 0.229$ and $P_{max} = 0.998$ days, for an RR Lyrae ID=13350 (RunDeepForcedSource.objectId = 3588818166880604). Note that the image coverage does not show the search region - we recommend improvements in this area. Furthermore, this view shows the light curve folded on u-band data, even though the period was found using g-band data. We recommend that once period is accepted, the lightcurve should fold on the same band as what was used to calculate the periodogram. A surprising behavior here is that clicking on one of the radio buttons instead of folding the lightcurve in that band on the accepted period, it displays the raw data for that band (Fig. 11)

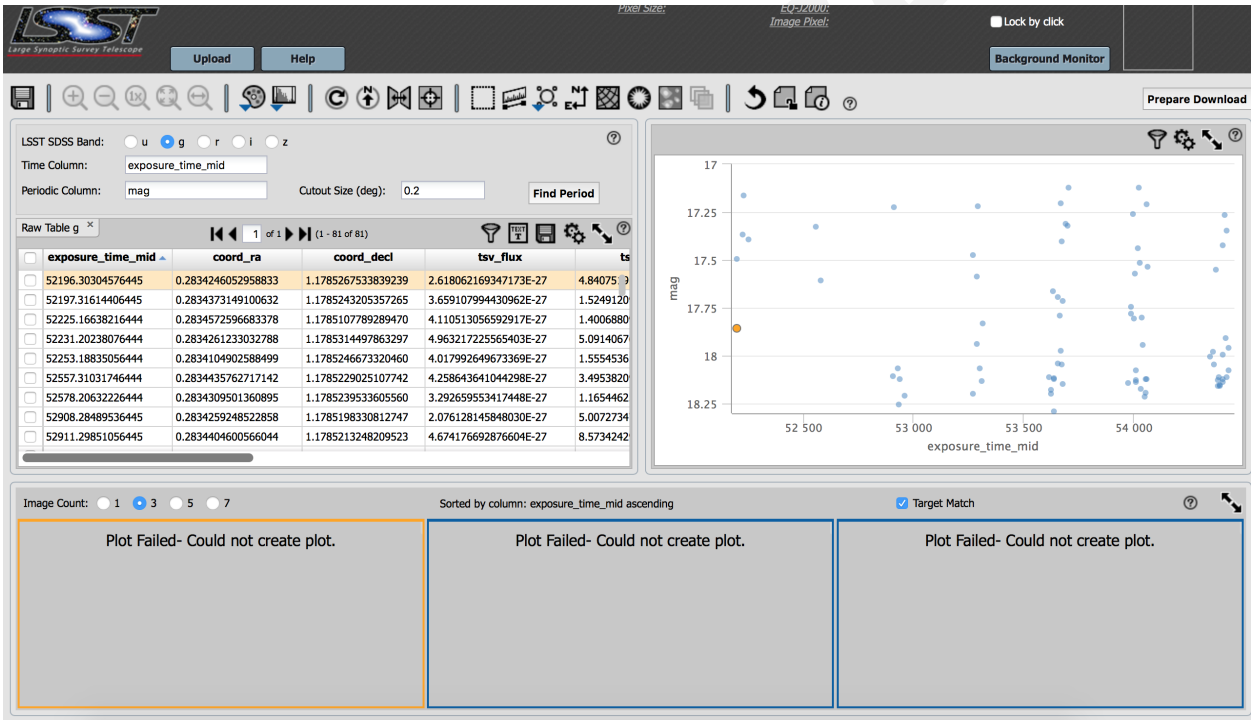


FIGURE 11: The result of clicking on the radio button for the g-band data. Instead of showing the light curve folded on the accepted period in the g-band, we see the raw g-band data. Clicking the 'Find Period' button does not 'remember' the result of the previous search on the same data. We recommend the parameters used for periodogram calculation to be remembered within a single object Time Series View.

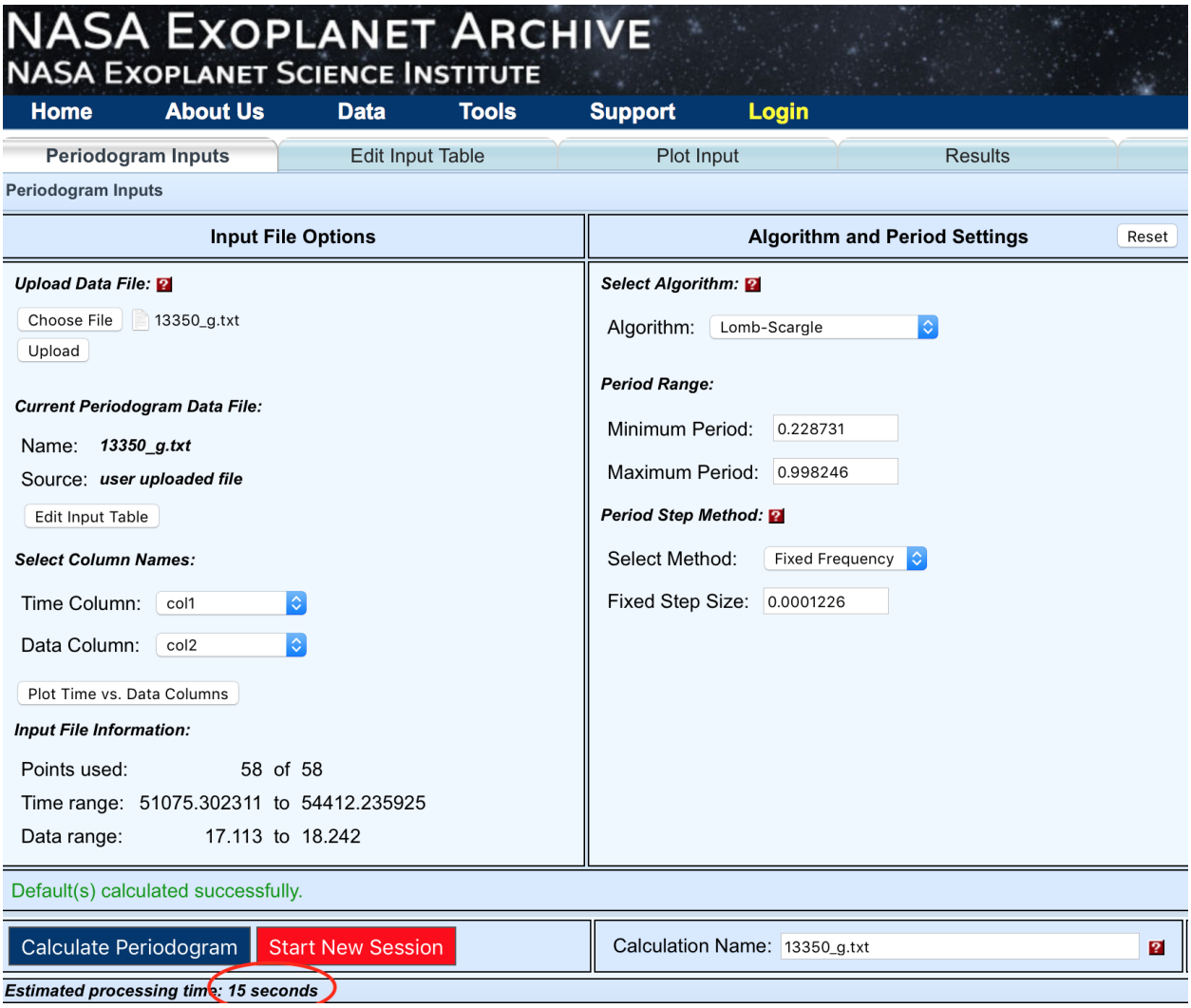


FIGURE 12: The same object as Fig. 14, and Fig 20, using the SDSS data from [2]. The highest significance frequency peak (power 21.58) corresponds to a period of 0.35365194 days. Only the second in significance peak (power 20.62) corresponds to the 'true' period of 0.547969 [2]. Note the bottom-left corner : the calculation took 15 secs for one lightcurve (compare to few miliseconds of AstroML code naive single-sinusoid approach that gave the same result for this particular object)

5.2 Light curve comparison : multiple cone queries

We perform multiple cone queries given known coordinates of well-classified sources. We take as ground truth 484 RR Lyrae stars from [2] , which were identified by color cuts, Lomb-Scargle periodogram and then confirmed by template fitting. Thus the periods from that work are considered to be reliable (see RR Lyrae 'ground truth' period distribution on Fig.13). Both fit parameters and SDSS lightcurves used in [2] are publicly accessible in the online version of the journal.

We perform individual SQL cone queries against PDAC for each RR Lyrae, allowing a generous 2 arcsec search radius (see Appendix A). Location of these RR Lyrae on the sky is shown on Fig. 18.

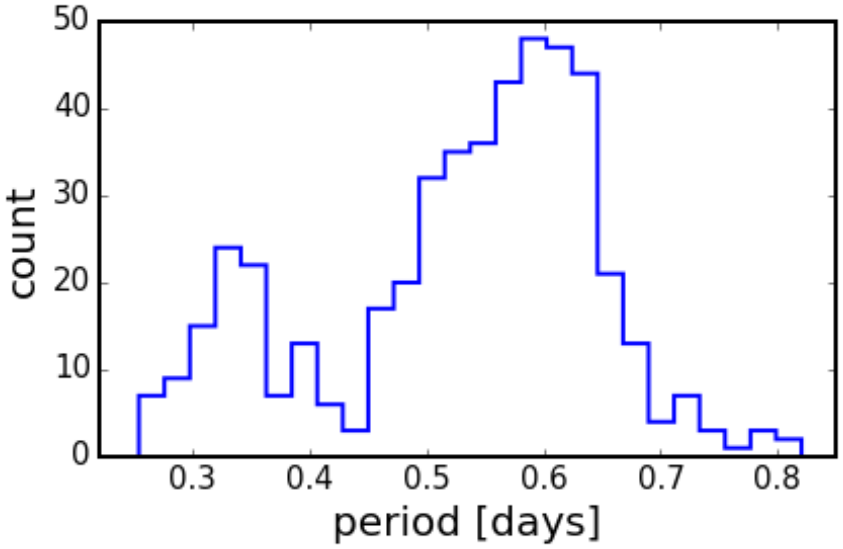


FIGURE 13: Distribution of RR Lyrae periods for 483 objects in [2]. Note the bimodal distribution, reflecting two main RR Lyrae types : 309 RRab (right) and 104 RRc (left) (see also Fig.16 in [2]).

Lomb-Scargle periodogram does not always find the 'true' period - it is subject to non-uniform sampling, aliasing, and necessity of choosing well the frequency sample on which periodogram powers are evaluated (see [5] for a recent overview). We nevertheless perform few simple sanity checks :

1. Does the PDAC lightcurve folded on 'true' period look real?
2. Using naive Lomb-Scargle, do we find the same period with the S82 Summer 2013

locally-stored data, and PDAC-hosted equivalent dataset?

3. Using naive Lomb-Scargle, do we find the same period with the [2] data and PDAC data?

Question 3 is less direct than questions 1 and 2 , since [2] used an earlier SDSS Data Release than what was used to create the S82 S13 reprocessed dataset. However, question 1 test solely whether the same object is stored in PDAC as we would assume (since it is unlikely that a random object would be well-represented by an RR Lyrae period). Similarly, Question 2 employs datasets that should be identical, and therefore we would assume that Lomb-Scargle tests would yield identical periodograms.

Assuming that the periods for the 483 RR Lyrae in [2] are correct, we attempt to download PDAC data for these objects, compute periodograms, and fold them on the true periods. This serves both to confirm that the periods are correct, and that the objects themselves are indeed RR Lyrae (see Sec. 5.2)

Using *astroML* python module [3], we sample the uniformly spaced frequency grid with N=5000 samples span between the smallest and the largest frequency reported in Table 1 of [2] $\pm 10\%$, i.e. $\omega_{min} = 0.9(2\pi/P_{min})$, $\omega_{max} = 1.1(2\pi/P_{min})$. We use the default *astroML* Lomb Scargle periodogram settings, namely generalized LS (see Eq.20 in [6], and Section 10.3.2 in [1]).

Using the same frequency grid for all 483 RR Lyrae, we compute Lomb-Scargle periodograms, and determine the best-fit period from the highest frequency peak (see eg. Fig. 15). We find that for about half of the SDSS lightcurves from [2], the Lomb-Scargle periodogram fitting single-term Fourier Series (LS) is sufficient to find the 'true' period, (see Fig. 17). We illustrate examples of RR Lyrae falling into each group : where with the naive LS we find the same period (Fig. 14), a smaller period (Fig. 15), , or a bigger period (Fig. 16) than the ground truth. For the same objects we also show PDAC lightcurves for which we also computed LS periodogram - see Figs. 20, 21 and 22, respectively.

Using the $\mathrm{RA}, \mathrm{Dec}$ for the RR Lyrae from [2] we positionally query the PDAC `RunDeepForcedSource` database (Cone Search), to find objects within 2 arcsec radius. As shown on Fig. 18, not all objects have a match. For the 343 stars with a PDAC match, we obtain calibrated g-magnitude lightcurves querying the `RunDeepForcedSource` and `Science_Ccd_Exposure` for the zero point magnitudes per exposure. For these PDAC lightcurves we also calculate Lomb-Scargle periodogram and find the frequency with most-significant power.

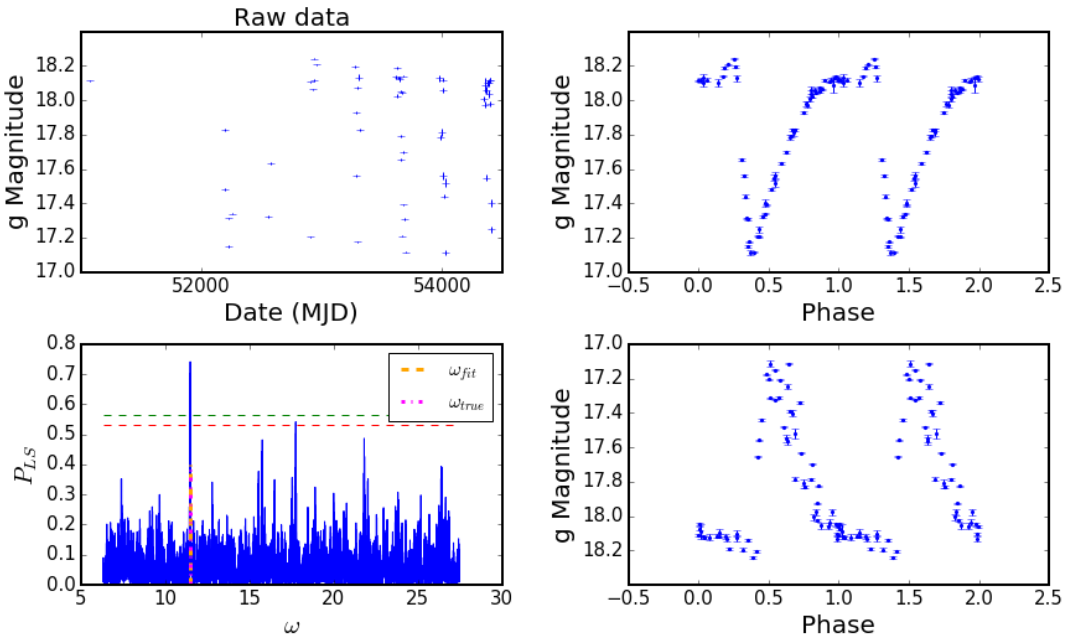


FIGURE 14: An example of the *astroML* Lomb Scargle periodogram performance, calculated for RR Lyr ID=13350 in SDSS g band (following Table 2 in [2]), using the SDSS data from [2]. It took 18.6 miliseconds on a laptop to calculate this periodogram. The upper left panel depicts the raw SDSS lightcurve data. The upper right panel shows the phased lightcurve constructed with a cited period of 0.547987 days ('P_{true}'). The lower left panel shows the Lomb Scargle periodogram on a uniform frequency grid (5000 bins), where the orange and magenta vertical lines mark the location of the highest periodogram peak, and the frequency based on the reported period ($\omega_{true} = 2\pi/P_{true}$). The lower right panel shows the phased lightcurve constructed with the Lomb-Scargle Periodogram period of 0.547161 days, corresponding to the highest peak, $P_{fit} = 2\pi/\omega_{fit}$. The horizontal red and green lines mark the 5% and 1% significance levels for the highest peak, as found from 500 bootstrap resamplings (See `http://www.astroml.org/book_figures/chapter10/index.html`). The same object, but pulling the data from PDAC, is shown on Fig. 20

TABLE 2: Comparison of RR Lyrae periods obtained with different methods. First, P(S) is the 'ground truth' - period resulting from detailed template fitting by [2]. Second, P(LS) is the period corresponding to the most prominent frequency in the Lomb-Scargle periodogram (LS) computed on the SDSS lightcurve for a given object pulled from online journal data in [2]. Third, P(EXO) uses the same SDSS data from [2] in g-band, to find the best period with the NASA Exoplanet Archive Periodogram service. Fourth, P(PDAC) uses the data pulled from PDAC, for which we find the best period using LS periodogram (same method as P(LS)).

ID	P(S)	P(LS)	P(EXO)	P(PDAC)
4099	0.641754	0.280827	0.64175	0.280827
13350	0.547987	0.547161	0.35365	0.547969
470994	0.346794	0.531667	0.34679	0.531667

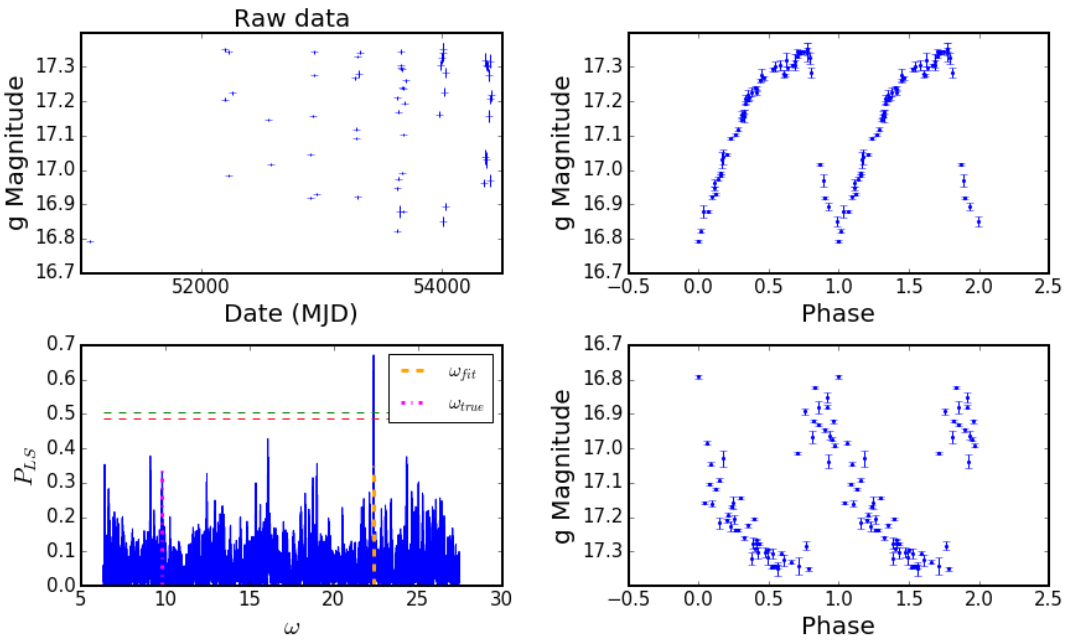


FIGURE 15: A periodogram, raw and folded lightcurve using SDSS data from [2] for RR Lyr ID =4099 (the same object with PDAC data is shown on Fig. 21). It is an example of a failure of naive single Lomb Scargle periodogram performance - the ratio of $\omega_{true}/\omega_{fit} = 0.437$. The 'true' period from [2] is 0.641754 days, whereas the naive Lomb-Scargle periodogram approach yields the 'fit' period of 0.280827 days. Note that here ω_{fit} and ω_{true} significantly differ for this RR Lyr, and the 'true' frequency, backed-up by the full lightcurve fitting of [2], appears as only one of insignificant periodogram peaks. Everything else as on Fig. 14.

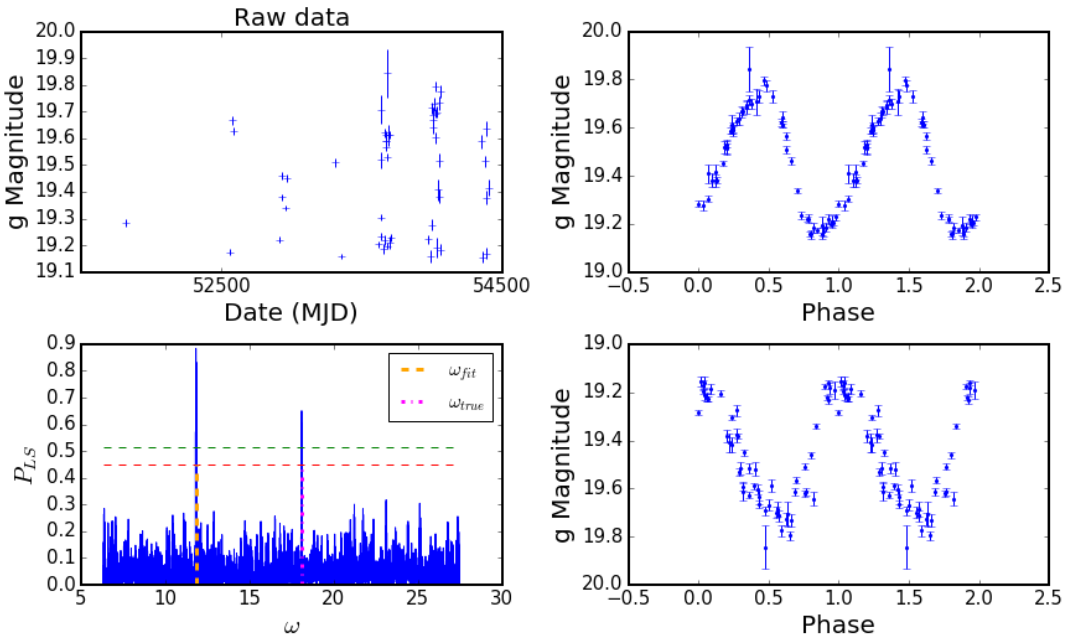


FIGURE 16: Same as Fig. 15, using the SDSS data pulled from [2], with $\omega_{true}/\omega_{fit} = 1.53$. Here RR Lyr ID=470994 has a cited period of 0.346794 days ('P_{true}'), whereas period derived from the Lomb-Scargle periodogram is 0.531667. It may be a good example of aliasing.

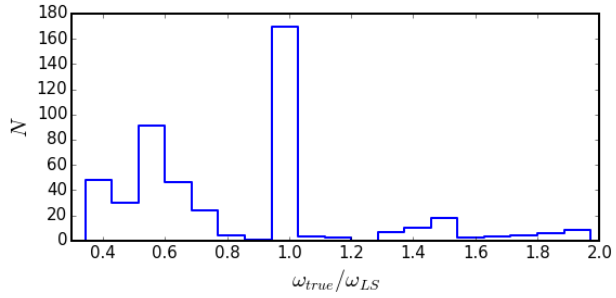


FIGURE 17: The distribution of the ratio of ω_{true} to ω_{fit}, where ω_{true} is inferred directly from the 'ground truth' - period cited in Table 2 of [2]. We take the same SDSS data from the paper (Table 1 in [2]), and calculate the Lomb-Scargle single-term generalized periodogram. The frequency corresponding to the highest peak is ω_{fit}. Thus, wherever this ratio is approximately equal to 1, this means that the naive LS approach is able to recover the 'true' period. However, where the highest frequency peak is not the same as ω_{true}, the ratio will be smaller or bigger from 1. This may be caused by the inherent simplicity of the simple single-term Fourier Series fitting. Indeed, some RR Lyrae lightcurves may have shapes that are insufficiently described by a single sinusoid (as on Fig.10.18 in [1]).

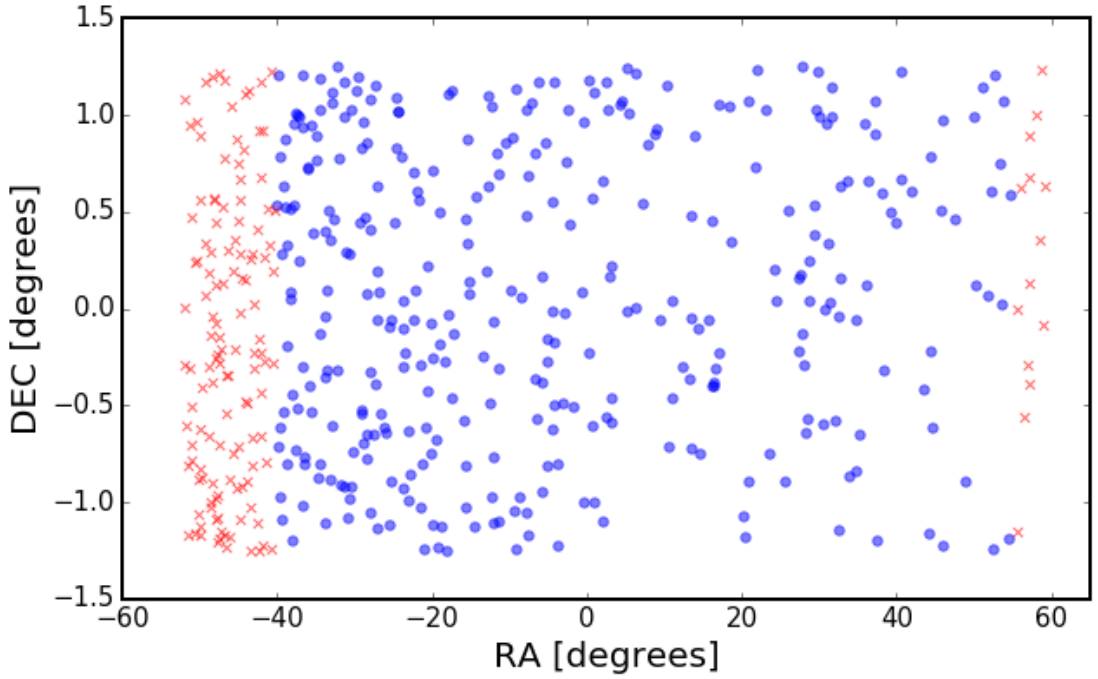


FIGURE 18: Results of positional query against 483 RR Lyrae stars from [2], using their RA, Dec. Blue dots are 343 stars that have a match in the PDAC S82 dataset within 2 arcsec, and red crosses are 140 stars that did not. Increasing the search radius to 3 arcsec does not alter this result.

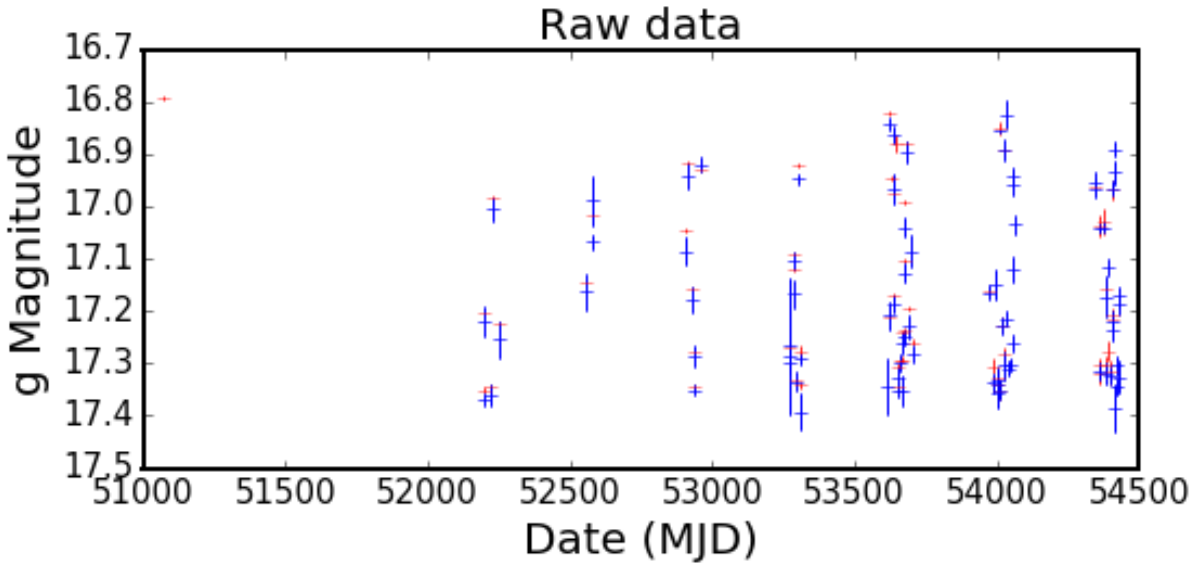


FIGURE 19: Comparison of RR Lyr ID=4099 from [2] (red crosses), and PDAC (blue crosses). The two lightcurves have different length : 59 vs 162 points, respectively.

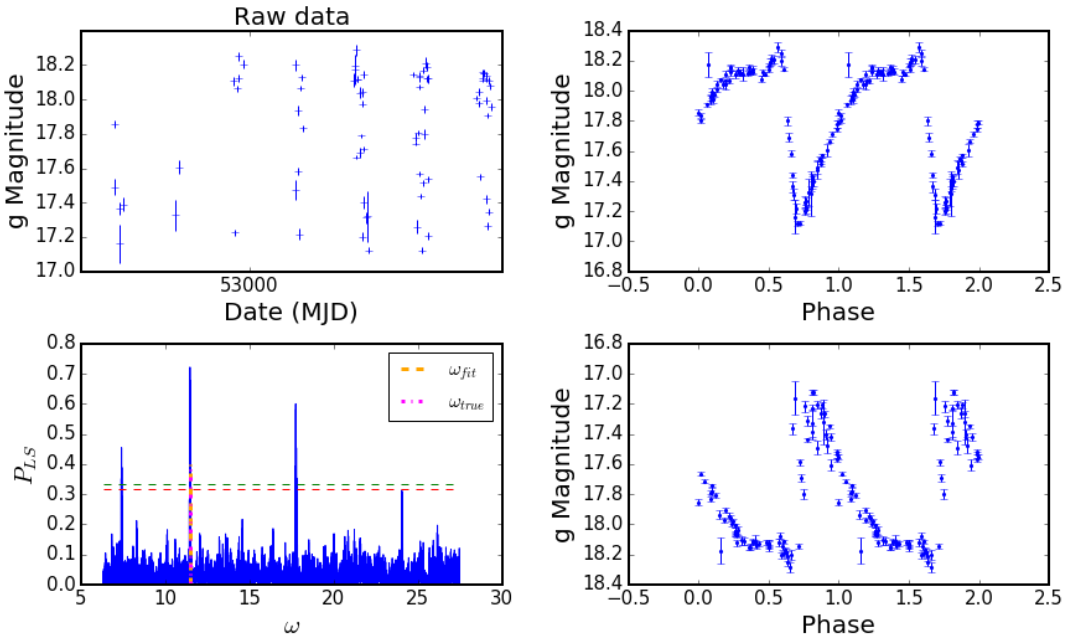


FIGURE 20: The same object as Fig. 14, but using data downloaded using PDAC. Using PDAC data, the RR Lyr ID=13350 has a best-fit period of 0.547969 days, almost identical to true period of 0.547987 from [2]. Panels the same as on Fig. 15

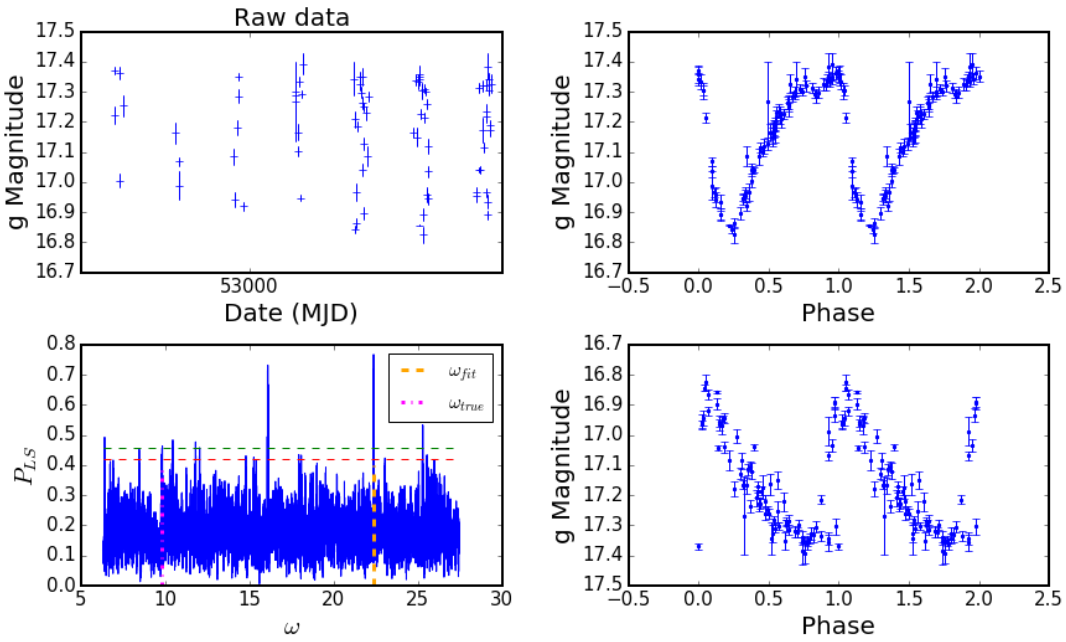


FIGURE 21: The same object as Fig. 15, but using data downloaded using PDAC. Calculating a naive LS periodogram using PDAC data for RR Lyr ID=4099 we find the best-fit period (frequency with highest power) of 0.280827 days, almost identical to the period found using LS periodogram on the SDSS [2] data of 0.280827 days. Both are discrepant with respect to the 'true' period of 0.641754 days from [2]. Panels the same as on Fig. 14

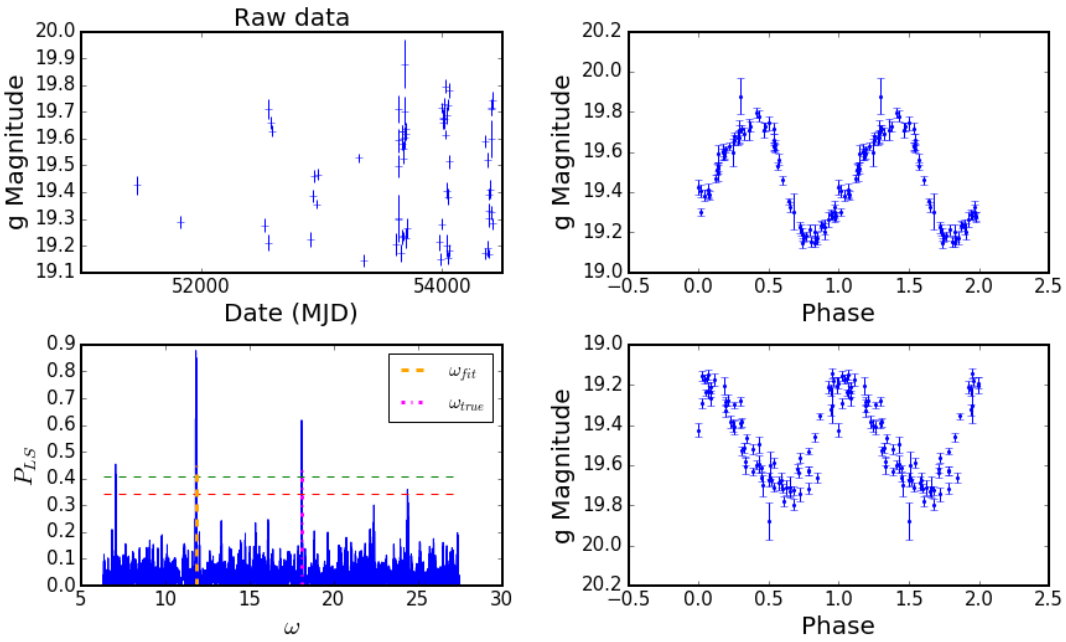


FIGURE 22: The same object as Fig. 16, but using data downloaded from PDAC. Calculating a naive LS periodogram using PDAC data for RR Lyr ID=470994 we find the best-fit period (frequency with highest power) of 0.531667 days, almost twice as high as the 'true' period of 0.346794 days from [2]. For this star we get an identical period if we use LS periodogram on SDSS data from [2] as opposed to PDAC. Panels the same as on Fig. 14

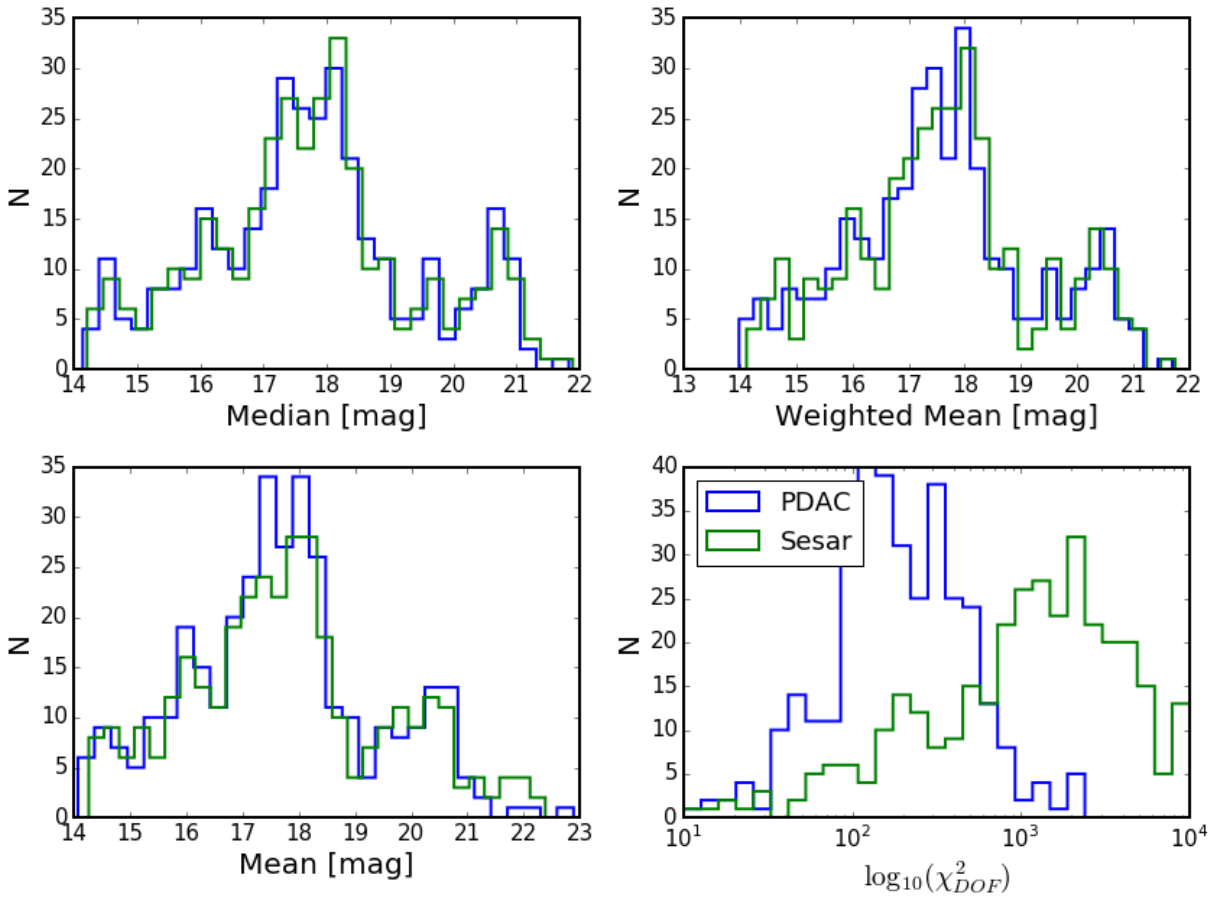


FIGURE 23: Comparison of the original [2] lightcurves (green) against data for the same objects pulled from PDAC (blue). For each of the 383 lightcurves in SDSS g-band, without any pre-processing or clipping, we calculated the median, weighted mean, mean, and chi^2_{DOF}.

6 More Tests

7 Conclusions

Acknowledgements

Thank you !

References

[1] Ivezić, Ž., Connolly, A. J., VanderPlas, J. T., & Gray, A. 2014, Statistics, Data Mining, and Machine Learning in Astronomy

[2] Sesar, B., Ivezić, Ž., Grammer, S. H., et al. 2010, ApJ, 708, 717

[3] Vanderplas, J., Connolly, A., Ivezić, Ž., & Gray, A. 2012, in Conference on Intelligent Data Understanding (CIDU), 47 –54

[4] VanderPlas, J., & Ivezić, Ž. 2015, The Astrophysical Journal, 812, 18

[5] VanderPlas, J. T. 2017, ArXiv e-prints, arXiv:1703.09824

[6] Zechmeister, M., & Kürster, M. 2009, A&A, 496, 577