



---

LARGE SYNOPTIC SURVEY TELESCOPE

---

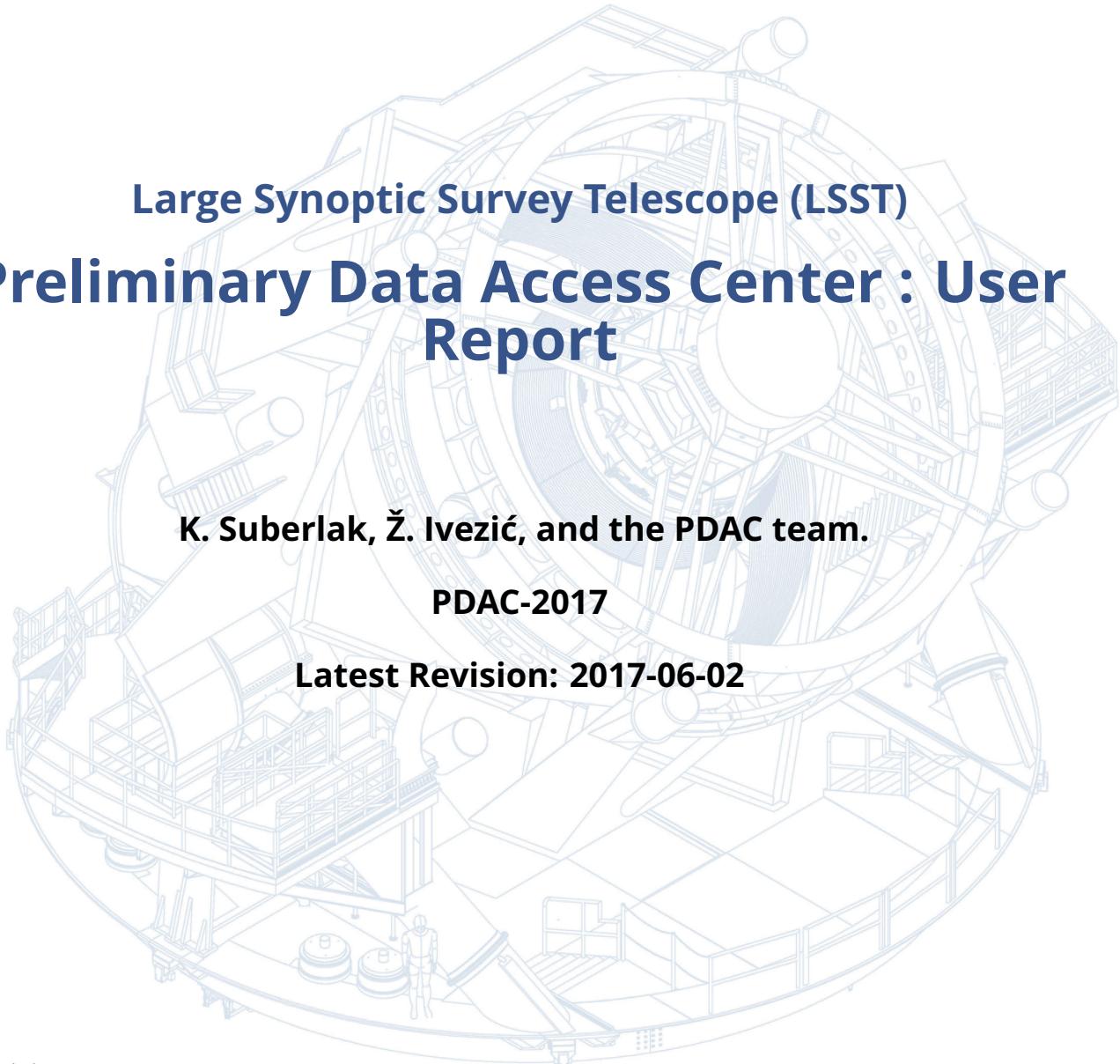
Large Synoptic Survey Telescope (LSST)

# Preliminary Data Access Center : User Report

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

PDAC-2017

Latest Revision: 2017-06-02



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
4	2017-04-11	Used new format for User Interface tests.	Krzysztof Suberlak
5	2017-04-20	Edited infrastructure tests.	Krzysztof Suberlak
6	2017-05-02	Added External Images and Catalogs sections.	Krzysztof Suberlak
7	2017-05-04	Edited time domain tests.	Krzysztof Suberlak
8	2017-05-20	Added single object tests.	Krzysztof Suberlak
9	2017-05-28	Added box query tests.	Krzysztof Suberlak

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Overview of performed tests</b>	<b>2</b>
<b>3</b>	<b>User Interface: what we see</b>	<b>2</b>
3.1	UI overview . . . . .	2
<b>4</b>	<b>Infrastructure : what is available and how to get it</b>	<b>4</b>
4.1	Overview . . . . .	6
4.2	Tests Performed . . . . .	7
4.2.1	Identification numbers in DeepSource and DeepForcedSource . . . . .	7
4.2.2	Postage Stamp Miniatures . . . . .	8
4.2.3	Database linkage : obtaining magnitudes . . . . .	10
4.2.4	External Images . . . . .	12
4.2.5	External Catalogs . . . . .	13
4.3	Time Series : periodogram . . . . .	22
<b>5</b>	<b>Database Ingestion : is what we get what we expected to get?</b>	<b>26</b>
5.1	Single forced photometry light curve . . . . .	34
5.2	Positional comparison : box query . . . . .	43
5.3	Light curve comparison : multiple cone queries . . . . .	48
<b>6</b>	<b>Conclusions</b>	<b>50</b>

<b>A Appendix : SQL queries</b>	<b>59</b>
A.1 Single source cone query, calibrated flux . . . . .	59
A.2 Single source cone query, calibrated magnitude . . . . .	60
A.3 Box query, $i < 23.5$ . . . . .	60

## 1 Introduction

This is a document to report on the user experience testing of the Preliminary Data Access Center on macOS Sierra Version 10.12.3 (16D32), using the Safari Version 10.0.3 (12602.4.8).

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 detailed description of technical requirements and a vision for the Science User Interface and Tools is outlined in documents LDM-130 (SUIT requirements)<sup>1</sup> and LDM-492 (SUIT Vision)<sup>2</sup>, as well as confluence pages<sup>3</sup>. Current Prototype Data Access Center implementation of SUIT is described on the Guide to PDAC<sup>4</sup>.

An online tool that can serve as an interface between the user and the data is not a new idea. There is indeed a number of websites that serve such purpose, such as Aladin<sup>5</sup>, Sloan Digital Sky Survey (SDSS) Sci Server including CasJobs<sup>6</sup>, NASA/IPAC InfraRed Science Archive (IRSA)<sup>7</sup>, Mikulsky NASA Archive<sup>8</sup>, NASA Extragalactic Databased (NED)<sup>9</sup>. These allow a user to query for data (either via SQL query, or with the interactive user interface), returning the data table. Some user interfaces (eg. IRSA) have some rudimentary plotting capabilities. SUIT stands at the forefront of user interface development, that would go beyond displaying the time series, spectrum, and other static information, but would also allow to analyze the data on the cloud. For instance, a jupyter notebook could be employed to run machine learning models, and use the output to further select objects based on feature analysis. Currently National Optical Astronomy Observatory (NOAO) DataLab<sup>10</sup> is an example of a tool that can help explore the data in an interactive way, using allocated computational resources. Similar developments are advocated at the MIT Astroinformatics / Geoinformatics group (eg. Mahali <http://mahali.mit.edu>)<sup>11</sup>

<sup>1</sup><https://docushare.lsstcorp.org/docushare/dsweb/ServicesLib/LDM-130/History>

<sup>2</sup><https://docushare.lsstcorp.org/docushare/dsweb/ServicesLib/LDM-492/History>

<sup>3</sup><https://confluence.lsstcorp.org/display/DM/Science+User+Interface+and+Tools>

<sup>4</sup><https://confluence.lsstcorp.org/display/DM/Guide+to+PDAC+version+1>

<sup>5</sup><http://aladin.u-strasbg.fr>

<sup>6</sup><http://www.sciserver.org/tools/>

<sup>7</sup><https://irsa.ipac.caltech.edu/frontpage/>

<sup>8</sup><https://archive.stsci.edu>

<sup>9</sup><https://ned.ipac.caltech.edu>

<sup>10</sup><http://datalab.noao.edu>

<sup>11</sup>Also, Prof. Victor Pankratius talk "Computer-Aided Discovery: Towards Scientific Insight Generation with Machine Support" <https://sites.google.com/site/victorpankratius/>

This report details tests performed on the PDAC implementation of SUIT, 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](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, focusing on the Sloan Digital Sky Survey Stripe 82 Forced Photometry LSST reprocessed dataset. In Section 3 we describe the functionality available through user interface. In Section 4 we describe the structure of available data : both datasets available directly from the 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 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

### 3.1 UI overview

In order to access PDAC we follow the directions<sup>4</sup> 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<sup>12</sup> Infrared Science Archive(IRSA) publicly accessible catalogs, including GAIA, WISE, 2MASS, SPITZER, etc. (see Fig. 2).

1. Title : Range of input accepted by 'Name or Position' box

- Description : We test what is the range of RA, dec values and types of names accepted by the 'Name or Position' box (see Fig. 1)
- Input : As input we use a set of coordinates, and names : first in  $-180^\circ < \text{RA} < 180^\circ$  convention  $(\text{ra},\text{dec}) = (-7.530128^\circ, -1.171239^\circ)$ , then the same but in  $0^\circ < \text{RA} < 360^\circ$  convention :  $(\text{ra},\text{dec})=(352.469872^\circ, -1.171239^\circ)$ . Finally we use an objectId = 216471849679198456, present in the DeepSource table when querying this location with 2 arcsec radius. We Cone query DeepSource table with 2" search radius.
- Results : Negative RA is not resolved (Fig. 3). Unless we use a name from NED / Simbad, the objectId is not resolved, even though it is present in the queried table.)
- Date: 2/15/2017

2. Title : Are all search options available in 'Method Search' ?

- Description: The 'Method Search' dialog box contains 'Cone', 'Elliptical', 'Box', 'Polygon', 'Multi-Object', 'All Sky' options. We test whether each method works with simple input.
- Input: We employ coordinates  $(\text{ra},\text{dec})=(352.469872^\circ, -1.171239^\circ)$  as search region center. We use 'Cone' radius 2 arcsec, 'Elliptical' semi-major axis of 2 arcsec, 'Box' side of 2 arcsec, 'Polygon' default vertices  $(352.48041^\circ, -1.18156^\circ), (352.45985^\circ, -1.18156^\circ), (352.45984^\circ, -1.16073^\circ), (352.48040^\circ, -1.16073^\circ)$ , 'Multi-Object': a list of two ra, dec submitted as a text file in the format  $\text{ra}, \text{dec} : (352.469872^\circ, -1.171239^\circ) | 342.469872^\circ, -1.101239^\circ$ . 'All Sky' : we add and SQL constraint  $\text{id} = 216471849679198456$ . For all search methods we query DeepSource catalog.
- Results: 'Cone', 'Elliptical', 'Box', and 'Polygon' search methods return a list of objects in coadds for a given search region. 'Multi-Object' provided with the `radec.txt` file returns an error 'Fail to load table. Error: `edu.caltech.ipac.firefly.server.query.DataAccessException`:

<sup>12</sup>Infrared Processing and Analysis Center, <http://www.ipac.caltech.edu/project/lsst>

DataAccessException:ERROR:Could not do Multi Object search, internal configuration wrong.: table should be a post search not a get from:unknown'. It would be more informative if an information about unavailability of this search method was on the main page rather than returning a result after a user uploaded and prepared an radec.txt file. The 'All Sky' method returns correct output (a list of selected fields where DeepSource.id =216471849679198456).

- Date: 3/28/2017

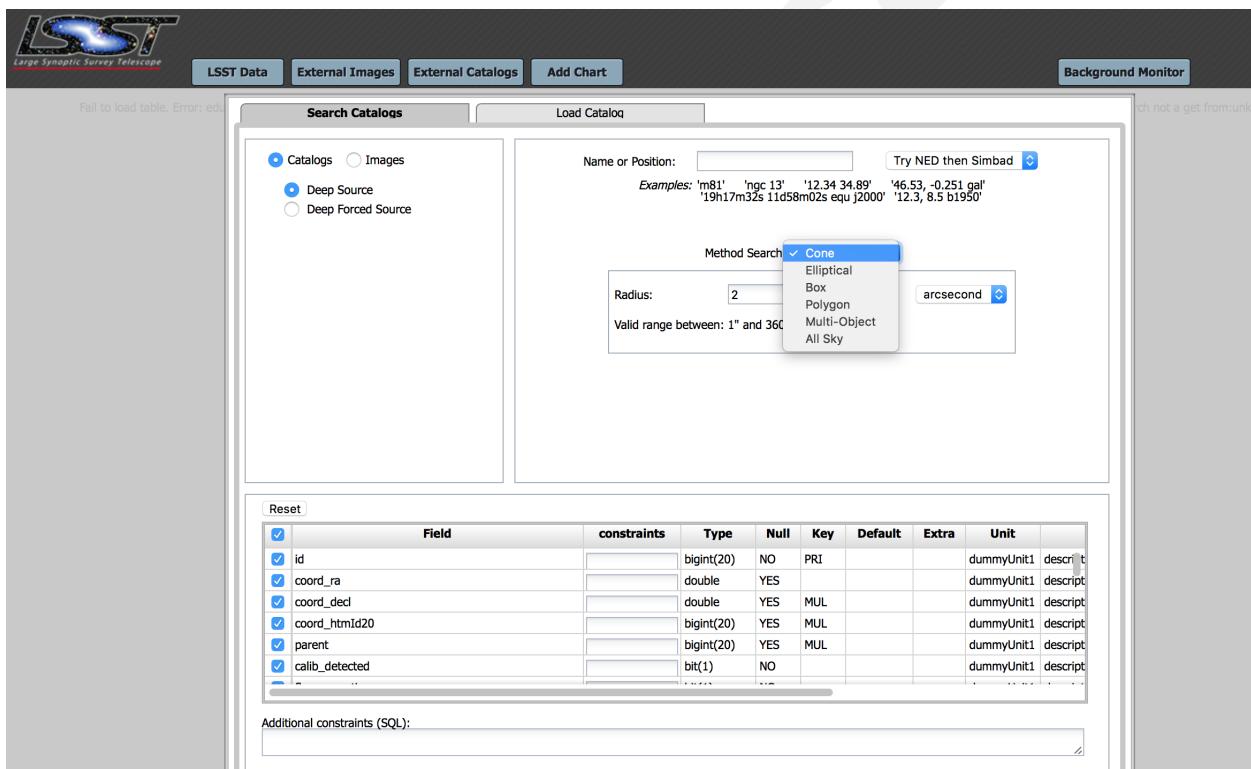


FIGURE 1: The main user interface of PDAC ver. 1. As of April 2017, Multi-Object and All Sky queries are not available. The 'Name or Position' only resolves positive RA ( $0^\circ < \text{RA} < 360^\circ$ ), while using direct SQL query resolves both positive and negative RA ( $-180^\circ < \text{RA} < 180^\circ$ ). 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.

## 4 Infrastructure : what is available and how to get it

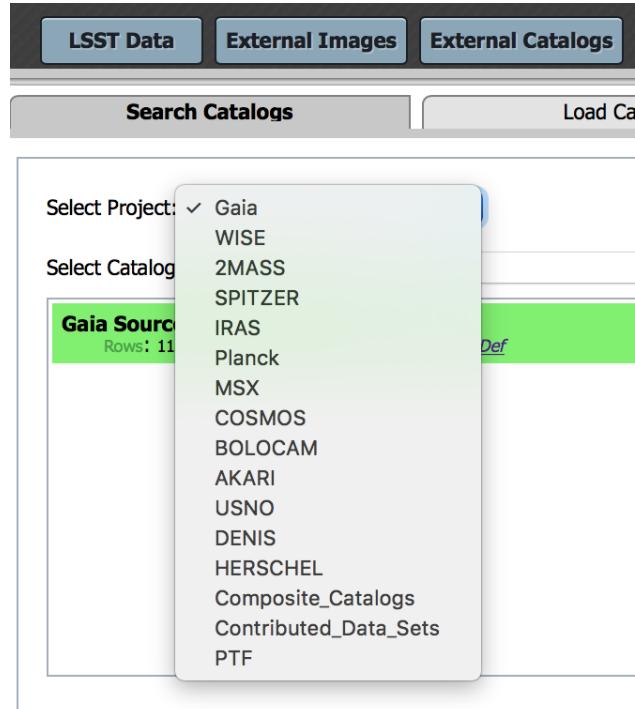


FIGURE 2: IPAC- hosted catalogs , accessible via IRSA.

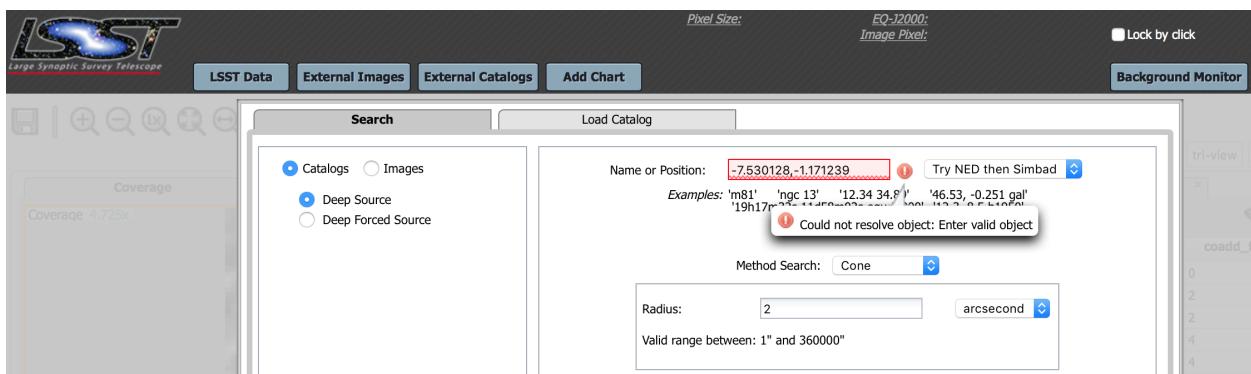


FIGURE 3: Testing range of input accepted by the 'Name or Position' box. The negative RA is not resolved, even though direct SQL query accepts both positive and negative values.

## 4.1 Overview

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<sup>13</sup> as part of Data Challange with the continuously developed LSST Stack<sup>14</sup>.

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, accessible 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' .
- 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<sup>15</sup>.

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

<sup>13</sup><https://confluence.lsstcorp.org/display/DM/Properties+of+the+2013+SDSS+Stripe+82+reprocessing>

<sup>14</sup><https://pipelines.lsst.io/index.html>

<sup>15</sup>[https://lsst-web.ncsa.illinois.edu/schema/index.php?t=DeepForcedSource&sVer=S12\\_lsstsim](https://lsst-web.ncsa.illinois.edu/schema/index.php?t=DeepForcedSource&sVer=S12_lsstsim)

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. 5).

## 4.2 Tests Performed

### 4.2.1 Identification numbers in DeepSource and DeepForcedSource

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  $\text{ra}, \text{dec} = 0.283437^\circ, 1.178522^\circ$ , 2" search radius (this is the RR Lyrae ID=13350 also investigated in Sec. 4.3). 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 DeepForcedSource.id - we recommend to highlight this in the metadata for it is a potential area for confusion. The only way to currently recover a light curve 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)

#### 1. Title : Ease of selecting a light curve for a single object

- Description : we test how difficult it is to select forced photometry light curve for only one object.
- Input : Cone query DeepForcedSource catalog for  $\text{ra}, \text{dec} = 0.283437^\circ, 1.178522^\circ$ , with 2" search radius. Select forced photometry data products for a single object within the search radius.
- Result : There is no option to show which objectId's are present in a given Deep-

ForcedSource query, and somehow select only one. The simplest workaround is to query DeepSource against given coordinates, adding SQL constraint 'coadd\_filter\_id = 3' to return only i-band id's (because id's corresponding to coadds in other filters were not a seed of forced photometry, and only for i-band there is a correspondence DeepSource.id = DeepForcedSource.objectId). The i-band DeepSource.id is 3588818166880604. We then query DeepForcedSource against the same coordinates, adding SQL constraint 'objectId' = 3588818166880604'. This is quite confusing (id means something different for DeepSource and DeepForcedSource), and may cause some problems to users willing to download forced photometry for only one object within a given space region.

- Date : 4/3/2017

#### 4.2.2 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 the origin of the miniature images.

1. Title : comparing the miniature images to SDSS DR13 Sky Server for a point source
  - Description : we test how well does a coverage image reflect the queried region.
  - Input : perform cone query against the Deep Source table, using coordinates of ra,dec = 23h30m57.31s, +1d1m13.8s (or 352.73878°, 1.02049°), with search radii of 2'', 10'', 100'', or 1000''. Each time, compare the coverage to the SDSS DR13 Sky Server <http://skyserver.sdss.org/dr13/en/tools/chart/navi.aspx>.
  - Result : we expected that the coverage image would be centered on a star, and indeed it was at 2'', 10'', 100'' search radii. However, at 1000'' radius (and larger), the image miniature drastically switches to using a different imaging survey without informing the user. It is much more blurry than before, impossible to recognize

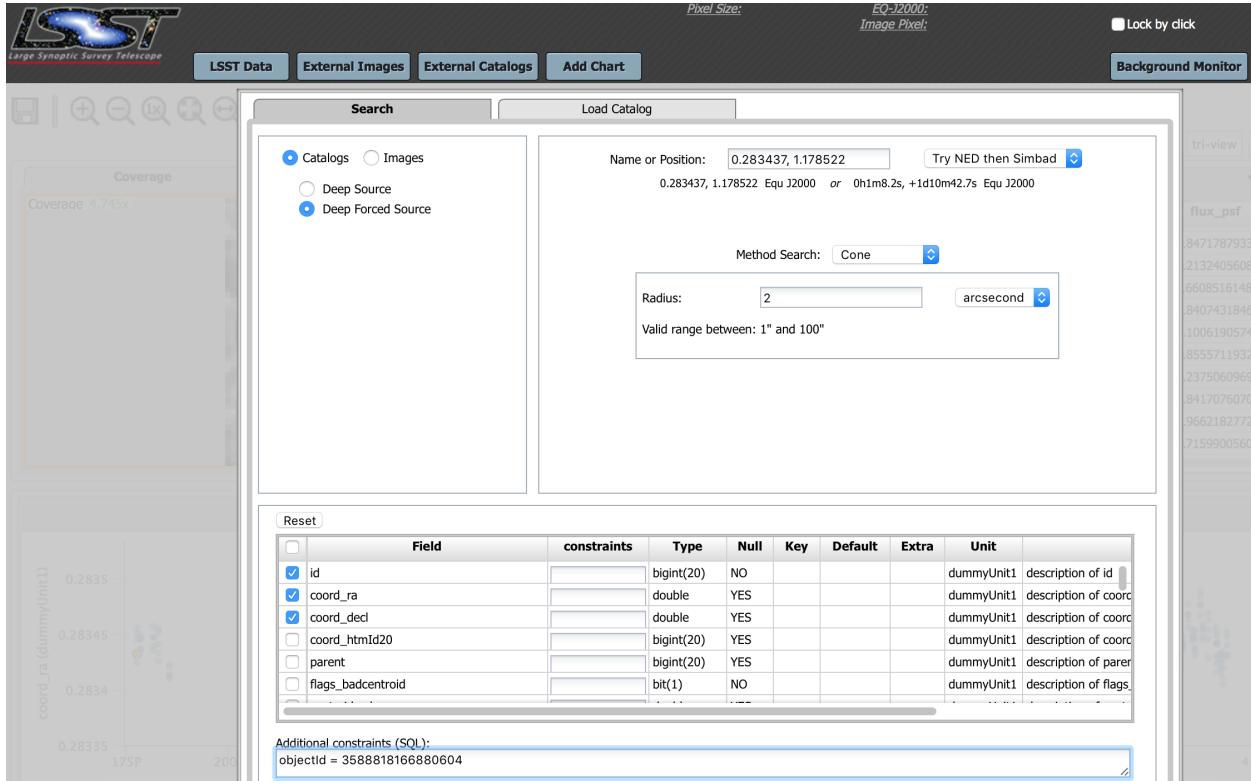


FIGURE 4: The correct way to select a light curve for a particular object from the forced photometry DeepForcedSource catalog. Here we first queried against the DeepSource catalog to find id's for objects detected in coadds in a small region within 2" from ra,dec = 0.283437°, 1.178522°. For i-band there 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.

features that should be present at that scale as compared to the SDSS DR13 Sky Server.

- Date: 3/7/2017
2. Title : comparing the miniature images to SDSS DR13 Sky Server for an extended source
- Description : we test how well does a coverage image reflect the queried region for an extended source (a galaxy)
  - Input : perform cone query against the Deep Source table, using coordinates of a Galaxy  $\text{ra},\text{dec} = 40.433^\circ, 0.449''$ . Each time, compare the coverage to the SDSS DR13 Sky Server <http://skyserver.sdss.org/dr13/en/tools/chart/navi.aspx>.
  - Result : we expected that the coverage image would be centered on a star, and indeed it was at  $2'', 10'', 100''$  search radii. However, at  $1000''$  radius (and larger), the image miniature drastically switches to using a different imaging survey without informing the user. It is much more blurry than before, impossible to recognize features that should be present at that scale as compared to the SDSS DR13 Sky Server.
  - Date: 3/7/2017

#### 4.2.3 Database linkage : obtaining magnitudes

A user querying a database would be very likely interested in obtaining calibrated magnitudes. We test the ease of obtaining coadd magnitudes for sources in a given location, and for obtaining light curves for a particular object.

1. Title : obtaining the calibrated magnitudes for coadd images (Deep Source catalog)
- Description : we test how easy it is to access with the user interface calibrated magnitudes for coadds for sources within a certain region.
  - Input : perform cone query against the Deep Source table, using coordinates of  $\text{ra},\text{dec} = 23\text{h}30\text{m}57.31\text{s}, +1\text{d}1\text{m}13.8\text{s}$  ( or  $352.73878^\circ, 1.02049^\circ$ ), with search radii of  $10''$ . Seek to select magnitudes from the available field.
  - Result : negative. We expected to find a field 'g magnitude', or similar, to find magnitudes in a given filter. However, such field is not present. Currently, one can only obtain magnitudes via a direct SQL query<sup>16</sup>

<sup>16</sup><https://confluence.lsstcorp.org/display/DM/PDAC+sample+queries+and+test+cases>

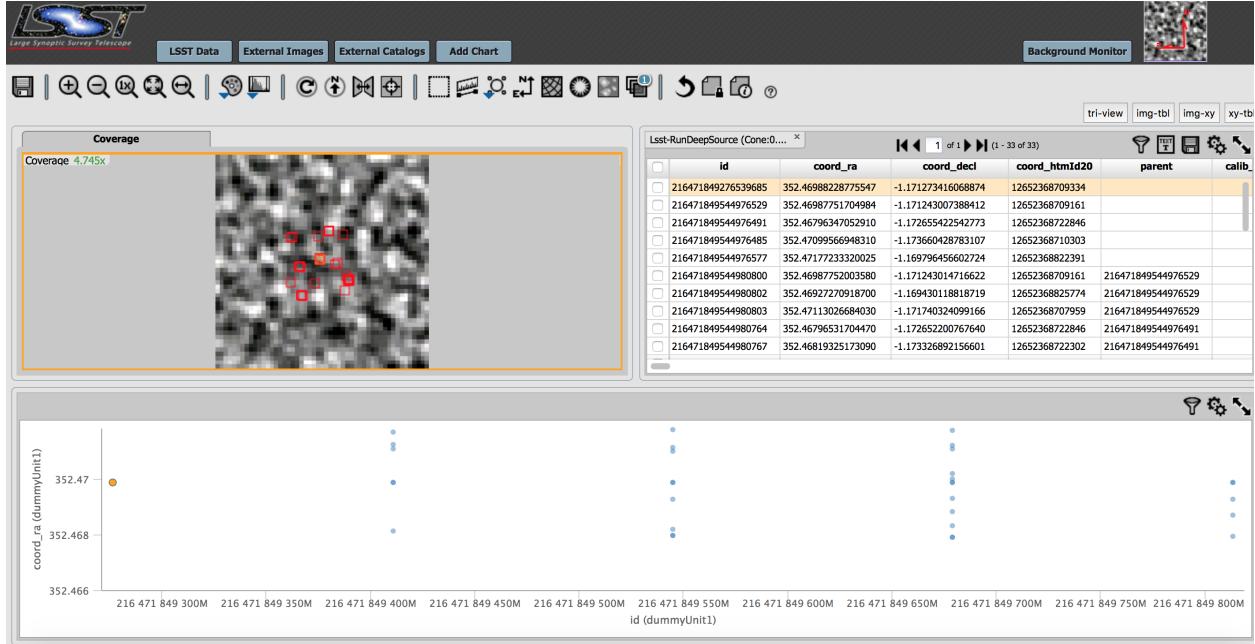


FIGURE 5: Example cone query against Deep Source table, returning all objects detected within a radius of 10'' from the position  $ra=352.469872^\circ$ ,  $dec=-1.171239^\circ$ . Note that the background image (postage stamp miniature) does not show the actual S82 coadds. This particular feature is described further in Sec. 4.2.2

- Date: 4/27/2017
2. Title : obtaining the calibrated magnitudes for forced photometry images (Deep Forced Source catalog)
- Description : we test how easy it is to access with the user interface calibrated magnitudes for forced photometry light curves for sources within a certain region.
  - Input : perform cone query against the Deep Source table, using coordinates of  $ra, dec = 23h30m57.31s, +1d1m13.8s$  ( or  $352.73878^\circ, 1.02049^\circ$ ), with search radii of 10''. Seek to select magnitudes from the available field.
  - Result : negative. We expected to find a field 'g magnitude', or similar, to find magnitudes in a given filter. However, such field is not present. As in the case of coadd images, we can only obtain magnitudes via a direct SQL query<sup>16</sup>
  - Date: 4/27/2017

#### 4.2.4 External Images

Apart from Stripe 82 reprocessed data, the user can also access via External Images tab image data from 2MASS, WISE, SDSS, MSX, DSS, IRAS (see Fig. 6)

1. Title : are the miniatures from External Images properly centered ?
  - Description : we test how well the object is rendered when querying for an image from External Images catalogs
  - Input : perform a NED resolved query for M3 globular cluster in Extended Images. Select 'Create New Plot'. Query SDSS u-band.
  - Result : we expected that the image would be centered on M3. In fact, it isn't - M3 appears off the center (see Fig. 11).
  - Date: 4/25/2017
2. Title : are the miniatures from External Images properly rotated ?
  - Description : we test how well the large-scale extended object is rendered when querying for an image from various External Images catalogs
  - Input : perform a NED resolved query for M81 in External Images. Select 'Create New Plot'. Query WISE, 2MASS, SDSS and DSS surveys.
  - Result : we expected that all miniatures would be centered on M81. The miniatures are only approximately centered, and the SDSS image is rotated with respect to WISE or 2MASS. A different angular scale is seemingly shown in tiled view. (see Fig. 7). Update: selecting WCS option (see Fig. 8) shows that images are approximately properly rotated, but are still not oriented in the same fashion.
  - Date: 4/23/2017
3. Title : does the option of creating new color image from selected RGB frames work properly?
  - Description : we test the ease of use of External Images - 'Create New Plot - 3 Colors' user interface.
  - Input : in Extended Images: Create New Plot - 3 Colors, query for M81, and select SDSS R,G,U bands to stand for red, green, blue colors.

- Result : the frames are properly added and the ability to select frames from other surveys (such as infrared) is really useful. However, the SDSS image still appears off-center (see Fig. 9 and 10)
- Date: 4/24/2017

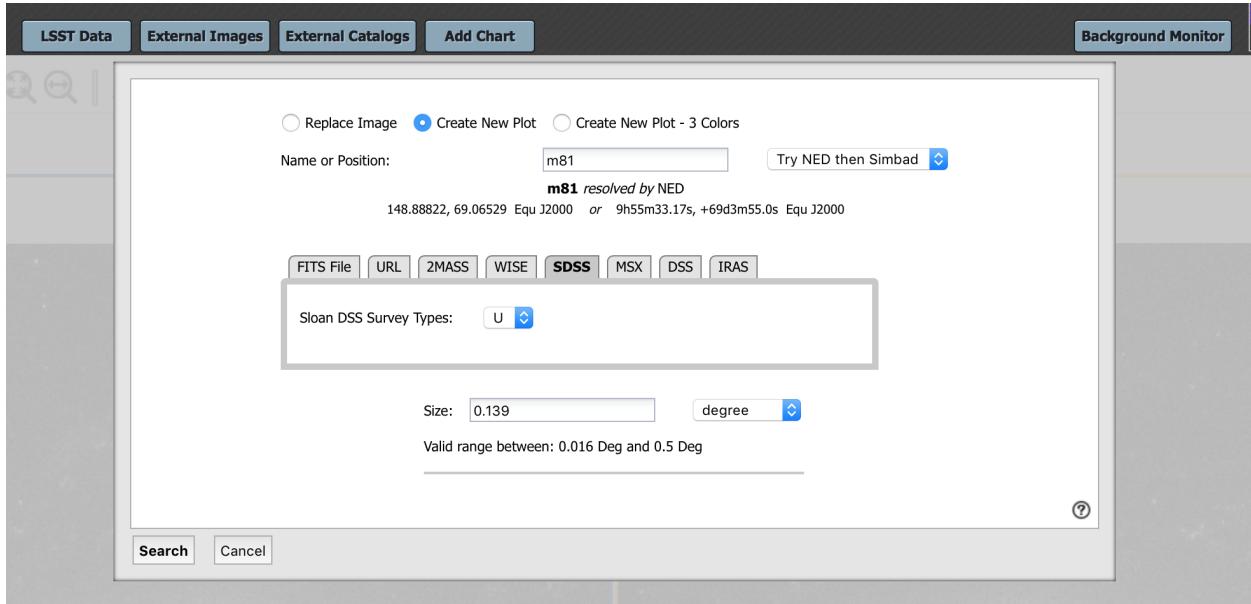


FIGURE 6: External Images User Interface. The radio buttons on the top allow to create a new image, add another panel to an image created by the previous query, or create a three-color image. In this example, described in Sec. 4.2.4, we query for an image of M81 , with the expected size of the image of 0.139 degrees.

#### 4.2.5 External Catalogs

The existing interface allows access to data from External Catalogs, that is not physically present at the NCSA, but parsed through IRSA. We test the ease of obtaining Gaia and WISE data for simple NED-resolved locations.

1. Title : can we obtain data from Gaia for a region specified by the cone query around a NED-resolved object?
  - Description : we test whether the data obtained from Gaia matches the location specified by the user. We choose M81 as our target for the ease of comparison to the miniature image shown in the results.

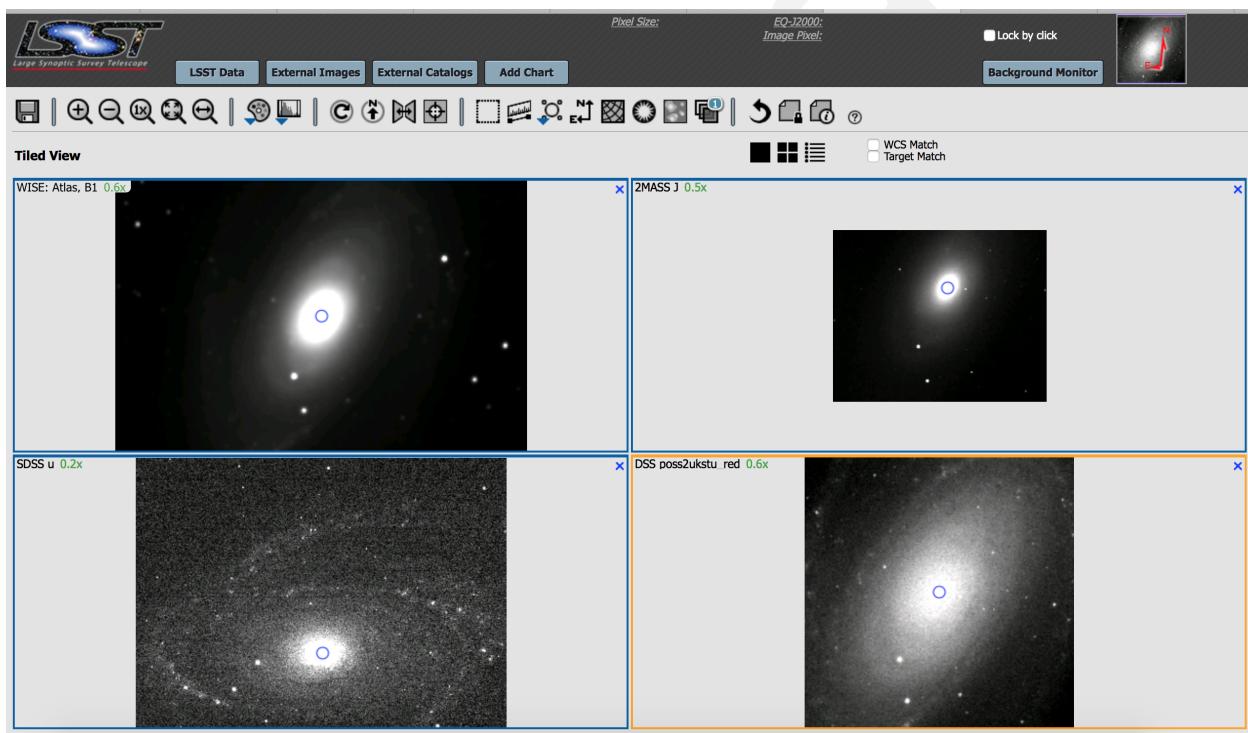


FIGURE 7: External Images User Interface : a succession of queries against WISE, 2MASS, SDSS and DSS catalogs for an image miniature of M81. The SDSS miniatures are rotated with respect to the other catalogs.

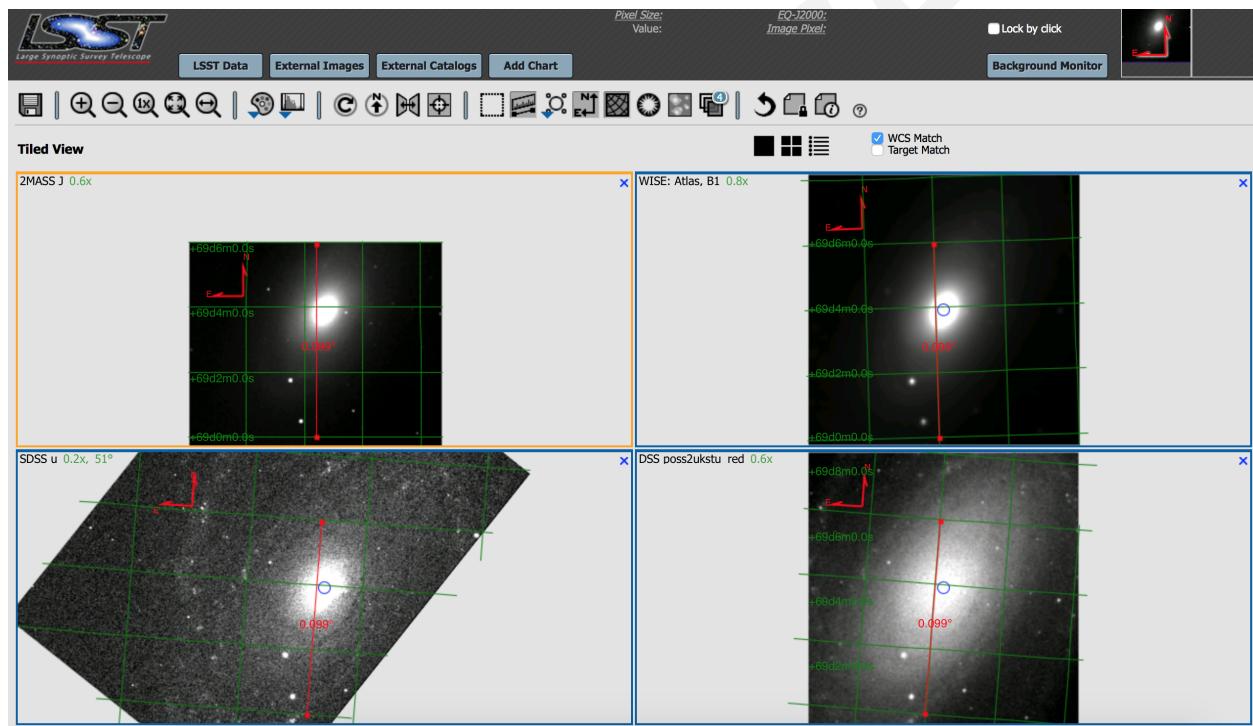


FIGURE 8: External Images User Interface : a follow-up to Fig. 7, showing tiles of WISE, 2MASS, SDSS u, DSS images. Selecting the 'WCS match' option, and clicking buttons to add grid layer to the image, and show the directions of Equatorial J2000 North and East, we see that the images are not all oriented in the same fashion. Using the ruler button we show the same distance on each tile to give a sense of scale.

Replace Image  
  Create New Plot  
  Create New Plot - 3 Colors

Name or Position:  Try NED then Simbad

**m81 resolved by NED**  
148.88822, 69.06529 Equ J2000 or 9h55m33.17s, +69d3m55.0s Equ J2000

**▼ RED**

Sloan DSS Survey Types:

**▼ GREEN**

Sloan DSS Survey Types:

**▼ BLUE**

Sloan DSS Survey Types:

Size:  degree

Valid range between: 0.016 Deg and 0.5 Deg



FIGURE 9: External Images : 'Create New Plot - 3 Colors' main menu.

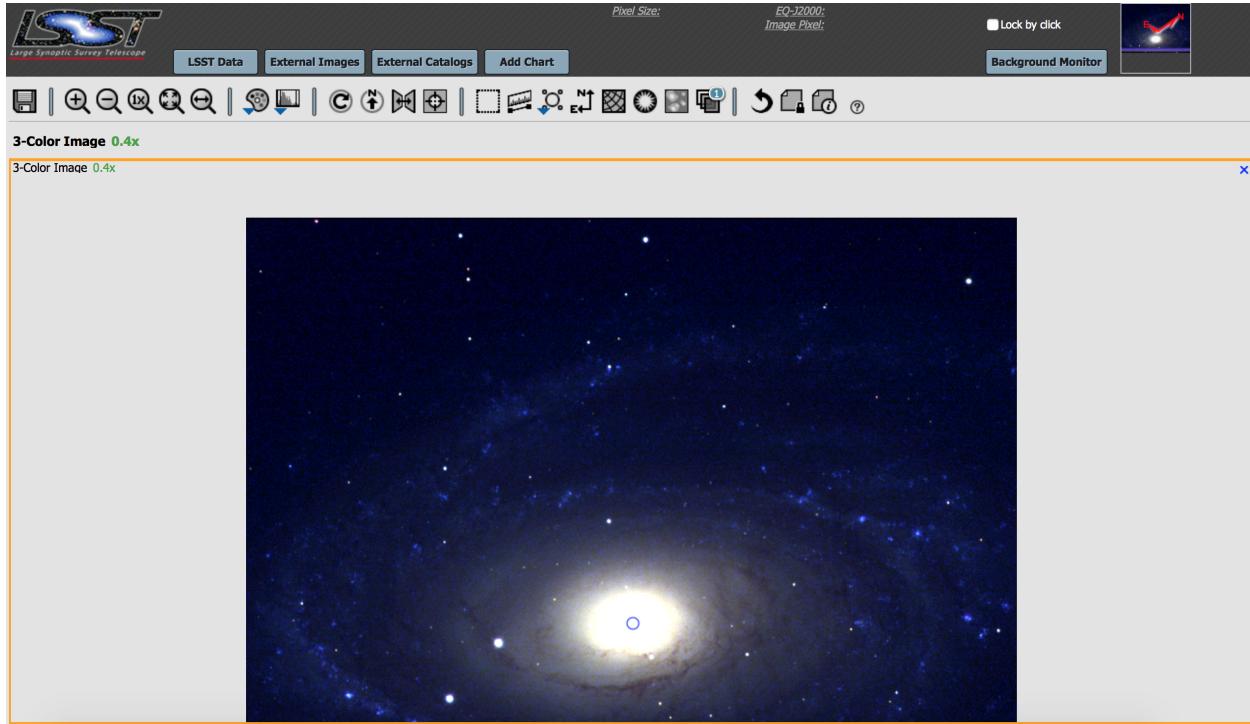


FIGURE 10: External Images : the result of query for M81 in SDSS R,G,U filter data shown as RGB colors.

- Input : perform a NED resolved query for M81 galaxy in External Catalog, with 100 arcsec search radius.
  - Result : positive. We see detections around M81 , and the miniature at this scale properly displays the queried region (see Fig. 13).
  - Date: 5/3/2017
2. Title : can we obtain WISE data from External Catalogs for a crowded field ?
- Description : we test how the WISE data can be accessed via the UI.
  - Input : perform a NED resolved cone query for M33 in Extended Catalogs, selecting AllWISE Source Catalog, and 100 arcsec search radius.
  - Result : positive. The miniature image is approximately centered on M33, and it is in the infrared (as we can see from comparing Fig. 14 to 2MASS image from SDSS SkyServer on Fig. 15)
  - Date: 5/3/2017

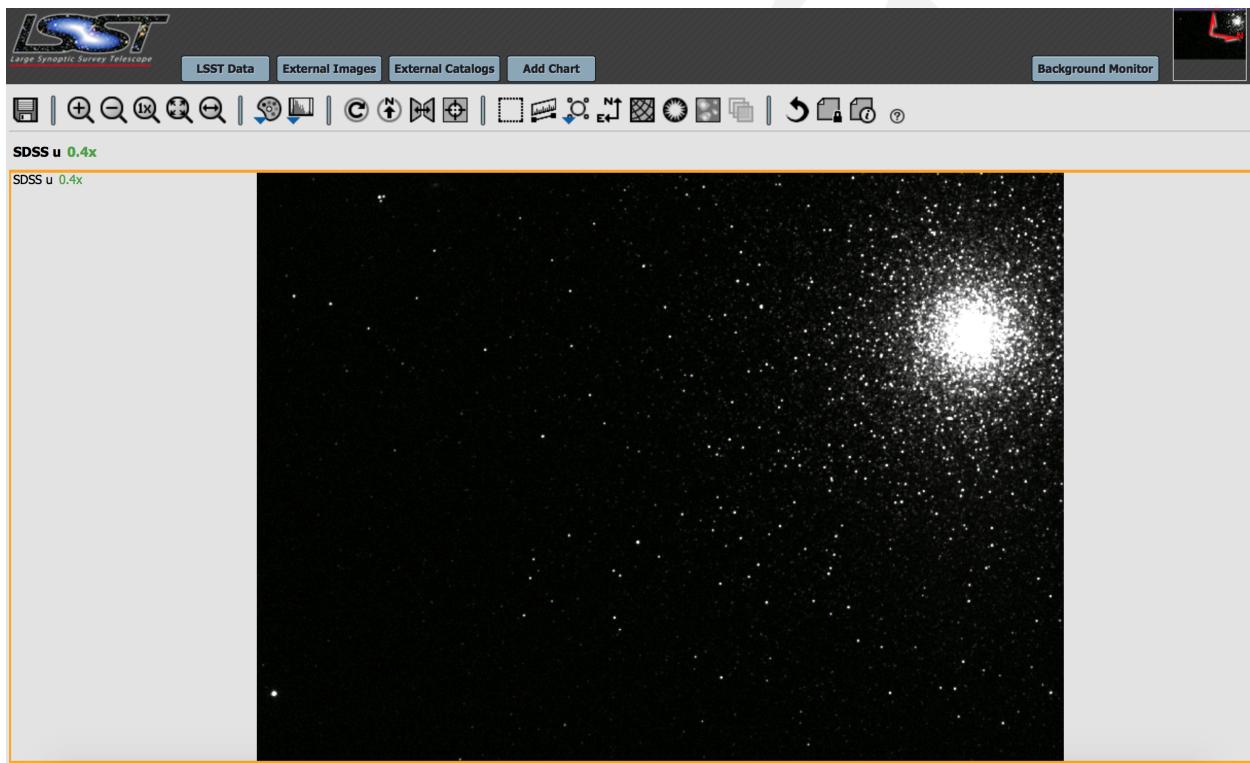


FIGURE 11: External Images : querying for an SDSS image of M3 Globular Cluster. It appears off the image center.

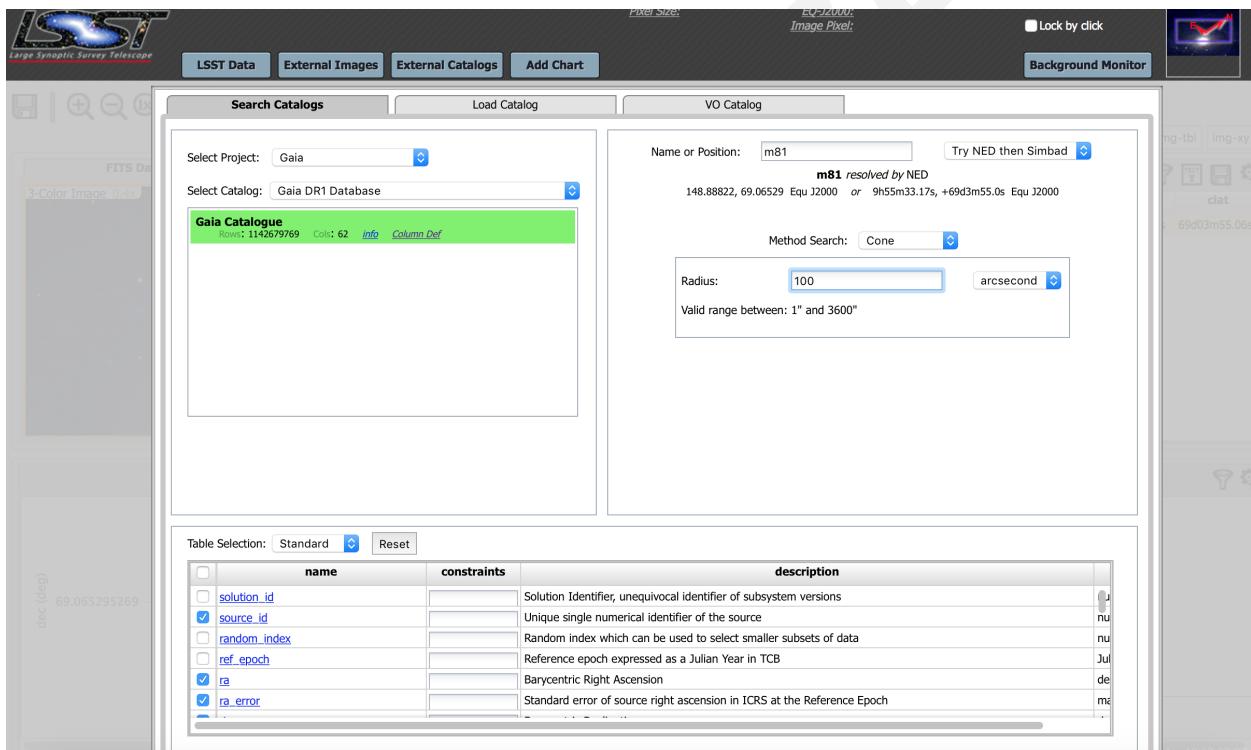


FIGURE 12: External Catalogs: main menu. The user can select one of many projects (see Fig. 2) : Gaia, WISE, 2MASS, SPITZER, IRAS< Planck, MSX, COSMOS, BOLOCAM, AKARI, USNO, DENIS, HERSCHEL, PTF, and others. Some projects have more than one associated catalog (eg WISE includes AllWISE, NeoWISE, etc).

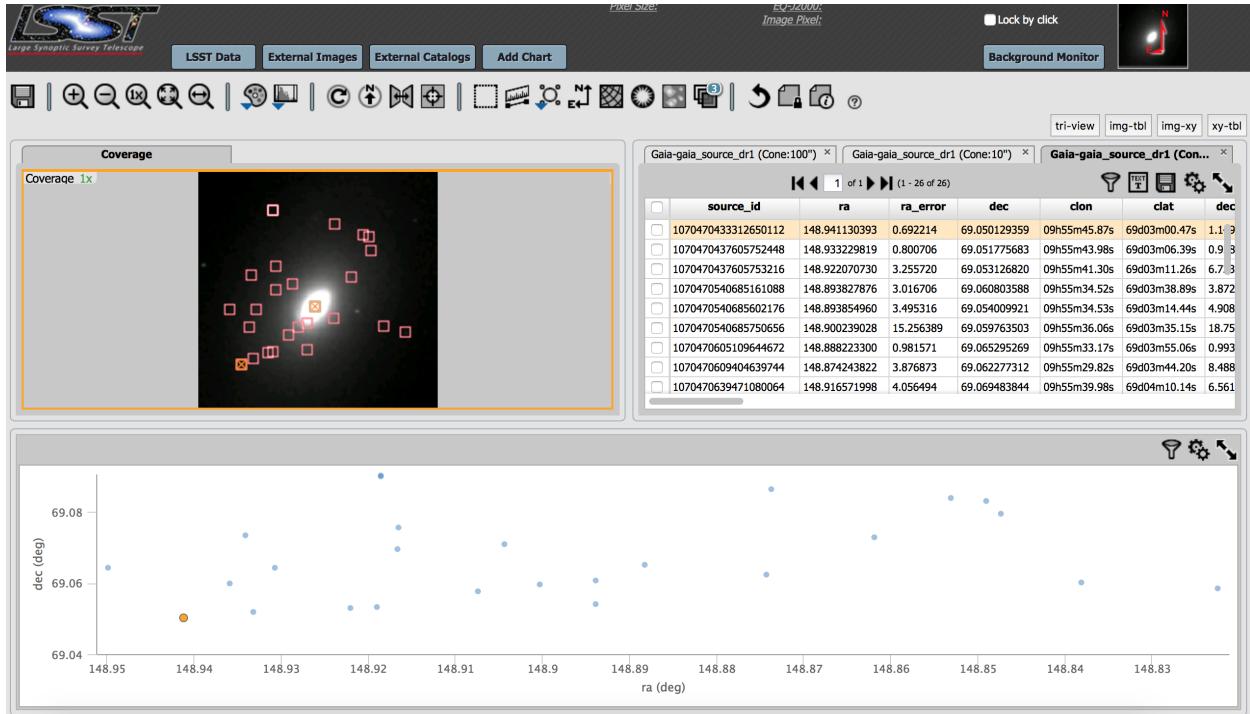


FIGURE 13: External Catalogs: the result of querying for Gaia detections around the location of M81.

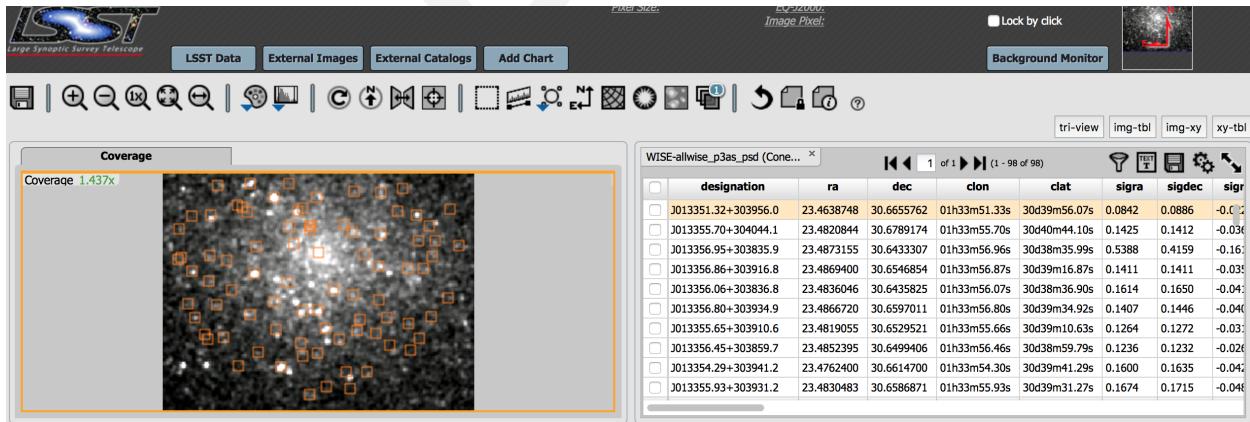


FIGURE 14: External Catalogs: the result of querying AllWISE Source Catalog in the cone query around location of M33 - the Triangulum Galaxy. Note that the miniature image may be in infrared, which is why the spiral features are not easily discernible (compare to Fig. 15 - the 2MASS cutout from SDSS Sky Server).

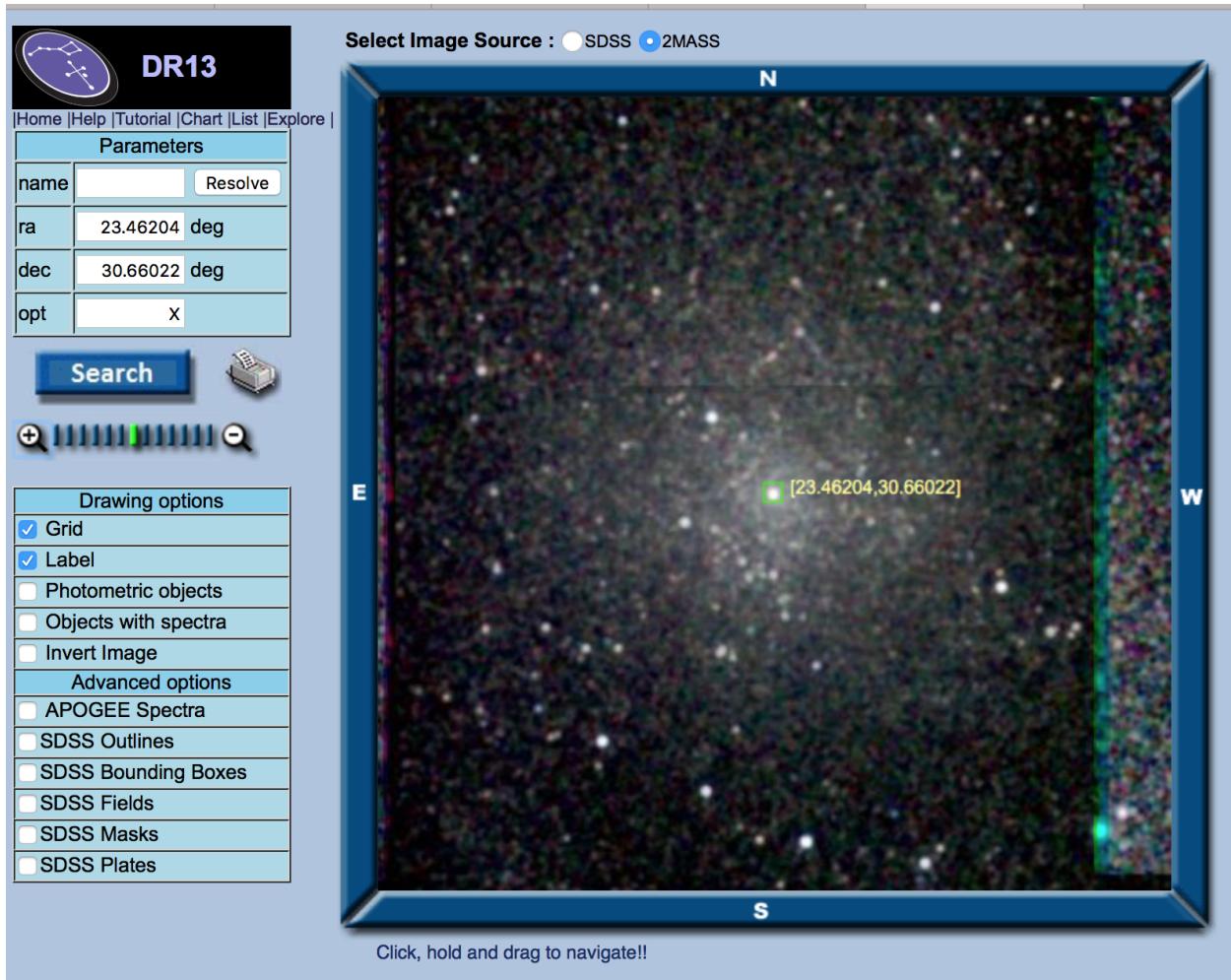


FIGURE 15: SDSS DR13 Sky Server : image showing the 2MASS image of the region around the position of M33 galaxy : ra,dec = 23.46204, 30.66022. Compare that to the query against the location of M33 for AllWISE detections shown on Fig. 14.

### 4.3 Time Series : periodogram

Ability to view the time series of any object is very useful for exploring its variability. The result of a query into Forced Deep Source table allows a simple 2D plot of any two data columns, and that could include time vs flux. However, the advantage of the Time Series View in the PDAC User Interface is that it also performs the cross-match against the Science CCD Exposure table to find the calibrated magnitudes, and allows calculating Lomb-Scargle periodogram to find the best fitting period.

Here we test the overall navigation experience in the Time Series User Interface. This includes finding the best periodicity with the Lomb-Scargle periodogram, folding the light curve on an accepted period, and changing the band in which the light curve is folded. We test the Time Series View on known RR Lyrae stars [2].

#### 1. Title : can we access the Time Series View easily?

- Description : we test the access to the Time Series View: is the button always responsive, do we get the same result if querying twice?
- Input : cone query the Deep Source catalog for a known RR Lyrae (ID=13350 in the catalog of [2] ), at ra,dec = 0.283437°, 1.178522°, with 2 arcsec search radius, to find objectIds at this location in i-band (since only i-band coadd was the seed for forced photometry in all bands). To select only i-band coadd we need to add an SQL constraint coadd\_filter\_id = 3. We find that the only objectId in i-band is 'id=3588818166880604'. We cone query Run Deep Forced Source table at the same position and search radius, adding an SQL constraint 'objectId = 3588818166880604' (since RunDeepSource.id == RunDeepForcedSource.objectId). We then select 'View Time Series' button (see Fig. 16).
- Result : positive. In most cases the Time Series View appears, with the three bottom windows missing any plots : a message 'Plot Failed = Could not create plot' appears (see Fig. 17). It is not clear what the three windows were meant to display. It is easy to select a light curve in a different band for the same image with the radio buttons in the upper left corner.
- Date: 4/7/2017

#### 2. Title : can we recover the period of an RR Lyrae star using the Lomb-Scargle periodogram with default settings?

- Description : the tool becomes useful if it can provide meaningful results. We test whether it is possible to find the true period of a known RR Lyrae star using the default settings.
  - Input : as in the ‘can we access the Time Series View easily?’ test, we cone query Deep Forced Source table for  $\text{ra}, \text{dec} = 0.283437^\circ, 1.178522^\circ$ , with 2 arcsec search radius, adding SQL constraint ‘objectId = 3588818166880604’. Click ‘View Time Series’, select SDSS band ‘g’ (the default is u), click ‘Find Period’. Select Periodogram Type ‘Lomb-Scargle’, Period Step Method ‘Fixed Frequency’. Leave Number of Peaks at 50. Click ‘Periodogram Calculation’ (Fig. 18).
  - Result : negative. The default grid stretches too many possible periods, and does not recover the true period of 0.546 days (see Fig. 19). We recommend implementing Periodograms for Multiband Astronomical Time Series (as in [4]) to utilize the presence of near-simultaneous multi-band observations. Another way to improve is to allow several different present grid spacing settings, eg. AstroML-like, super-Nyquist, etc.
  - Date: 4/7/2017
3. Title : can we recover the period of an RR Lyrae star using the Lomb-Scargle periodogram with custom settings?
- Description : we test whether it is possible to find the true period of a known RR Lyrae star using the custom settings. Given our knowledge of the range of true RR Lyrae periods, we constrain the maximum and minimum periods (frequency) for Lomb-Scargle Periodogram calculation.
  - Input : as in the ‘can we access the Time Series View easily?’ test, we cone query Deep Forced Source table for  $\text{ra}, \text{dec} = 0.283437^\circ, 1.178522^\circ$ , with 2 arcsec search radius, adding SQL constraint ‘objectId = 3588818166880604’. Click ‘View Time Series’, select SDSS band ‘g’ (the default is u), click ‘Find Period’. Select Periodogram Type ‘Lomb-Scargle’, Period Step Method ‘Fixed Frequency’. Leave Number of Peaks at 50, enter Period Min = 0.229, and Period Max = 0.998 [days]. Click ‘Periodogram Calculation’. Try also 0.29-0.9 range.
  - Result : positive. One of the major periodogram peaks corresponds to the true period of 0.546 days (see Fig. 20). By judiciously constraining the minimum and maximum periods searched (based on our knowledge of periods for this type of RR Lyrae stars) we can recover the correct period. However, the method is very sensitive to chosen frequency grid - small change in bounds leads to dramatic improvement / degradation in usefulness of periodogram (see Fig. 21). We also recommend

that the method used ('Fixed Frequency' or 'Fixed Period') should be displayed in the Viewer.

- Date: 4/7/2017

4. Title : is there a documentation describing the algorithms used to calculate the periodogram ?

- Description : we look for documentation so that the user can better understand what type of Lomb-Scargle algorithm is used, since it would affect the outcome of the calculation
- Input : search the Time Series view, or the results of the periodogram calculation, for any documentation.
- Result : negative. We find no reference to which specific algorithms were used. From the PDAC team we found that 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, and the original documentation of NASA Exoplanet Archive is available at [http://exoplanetarchive.ipac.caltech.edu/docs/pgram/pgram\\_parameters.html](http://exoplanetarchive.ipac.caltech.edu/docs/pgram/pgram_parameters.html). We recommend that there be a link to this page providing the user with this crucial information.
- Date: 5/4/17

5. Title : does the period selection bar correctly update limits and values?

- Description : we test in selecting minimum and maximum periods in the Periodogram calculation whether the x-ticks of the period selection slider bar update correctly, and whether it is easy to change the values manually.
- Input : as in "can we recover the period of an RR Lyrae star using the Lomb-Scargle periodogram with custom settings?". Check whether the x-ticks are appropriately updated (selected minimum on the far left of the slider bar, and selected maximum on the far right). Enter a different minimum value : 0.4 instead of 0.229, and check the behavior.
- Result : negative. As in Fig. 19, the minimum and maximum of the slider bar is always 0.001, 365 days. Furthermore, as in Fig. 20, we see that the period slider bar x-ticks do not correctly update to the custom-chosen period range.
- Date: 5/4/17

6. Title : is it easy to fold the light curve on an accepted period in a different band ?

- Description : we test whether it is possible to fold the light curve on an accepted period found with one band (eg.u) but using the data from a different band (eg. g). This is useful because there should be approximately the same period detected throughout bands for many sources (eg. RR Lyrae).
- Input : follow the steps as in "can we recover the period of an RR Lyrae star using the Lomb-Scargle periodogram with custom settings?". Choose the period closest to 0.5467 days. Click 'Accept period' in the lower right corner. In the new 'Viewer' window, click on a radio button corresponding to a different band (eg. r). Observe if the light curve in r-band is folded on the accepted period.
- Result : negative. Once the period is accepted, we are brought to a 'Viewer' window, which shows the light curve folded in the original band on an accepted period. Clicking on r-band instead of folding the light curve in r-band on that period (as seems intuitive), shows the raw photometry, with no way of folding it on any period (see Fig. 23). Furthermore, the actual value of the accepted period is not displayed anywhere in the 'Viewer' window. We recommend: a) once the user clicks 'accept period', clicking in the 'Viewer' window on radio buttons of u,g,r,i or z band would simply fold the light curve using the data from the selected band on the chosen period . b) the accepted period ought be displayed somewhere in the 'Viewer' window, eg. on top of the plot, saying 'Accepted period = .... [days]'. c) the windows on the bottom do not show anything informative - it would be useful if the user was informed what are they supposed to show. d) the parameters used for periodogram calculation (minimum, maximum period) to be remembered within a single object Time Series View.
- Date: 5/3/17

7. Title : is the periodogram tool very sensitive to boundary values?

- Description : we test whether the current implementation of the periodogram tool is sensitive to the boundary values chosen for period.
- Input : as in "can we recover the period of an RR Lyrae star using the Lomb-Scargle periodogram with custom settings?". Cone query the Deep Forced Source table for  $ra,dec = 0.283437^\circ, 1.178522^\circ$ , with 2 arcsec search radius, adding SQL constraint 'objectId = 3588818166880604'. Click 'View Time Series', select SDSS band 'g' (the default is u), click 'Find Period'. Select Periodogram Type 'Lomb-Scargle', leave Number of Peaks at 50, set Period Step Method to 'Fixed Frequency'. Enter Period

Min = 0.22, and Period Max = 0.9 [days]. Click 'Periodogram Calculation'. Try also 0.3-0.9, and 0.29-0.9 range. Try changing to 'Fixed Period', using the same period bounds (0.22-0.9, 0.3-0.9, 0.29-0.9).

- Result : positive. The periodogram tool is very sensitive to changes in boundaries: compare Figs. 20 and 21 : 0.229-0.998 vs 0.29-0.9. We find that with the same bounds, the period step method yield also different results: compare pairs of 0.22-0.9 and 0.3-0.9 (left panel is 'Fixed Frequency', right is 'Fixed Period'). Surprisingly, even if the range itself is smaller, we may be less sensitive to the true period: compare 0.29-0.9 to 0.22-0.9 (Figs. 21 and 25) - the latter has a more dense grid, and yet the performance is strikingly different. We recommend to warn the user that at best the periodogram tool may help confirm the period, and that the default periodogram number of peaks may be insufficient.

- Date: 5/5/17

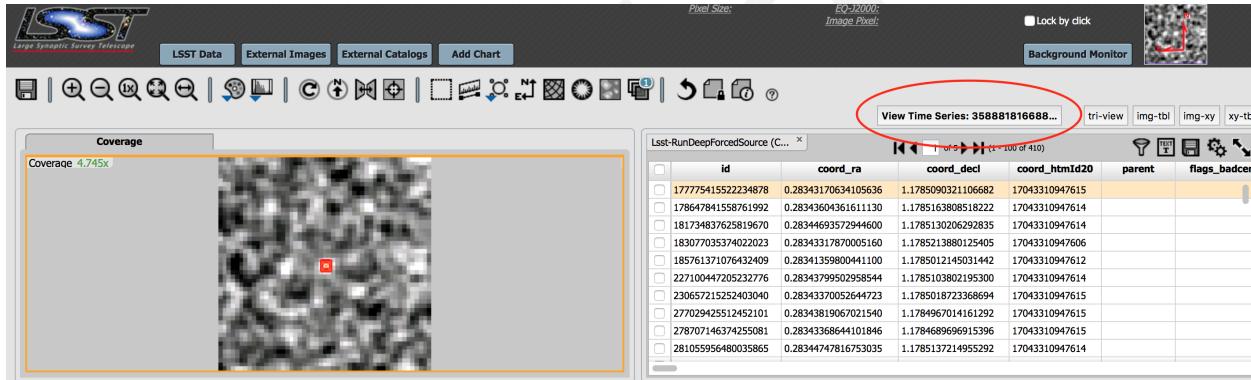


FIGURE 16: A result of cone query against Run Deep Forced Source table (containing S82 S13 data), with (ra,dec,radius = 0.283437, 1.178522, 0.00055 degrees). We circle the 'View Time Series' button that links to the Time Series UI shown on Fig. 17.

Finally, we compare the PDAC Time Series User Interface to that of the NASA Exoplanet Archive Periodogram<sup>17</sup> (see Fig. 26). Using few RR Lyrae PDAC g-band light curves, each calculation is allocated a time slot of approximately 15 seconds. Also see Table 3 for a summary of results.

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

<sup>17</sup><http://exoplanetarchive.ipac.caltech.edu/cgi-bin/Pgram/nph-pgram>

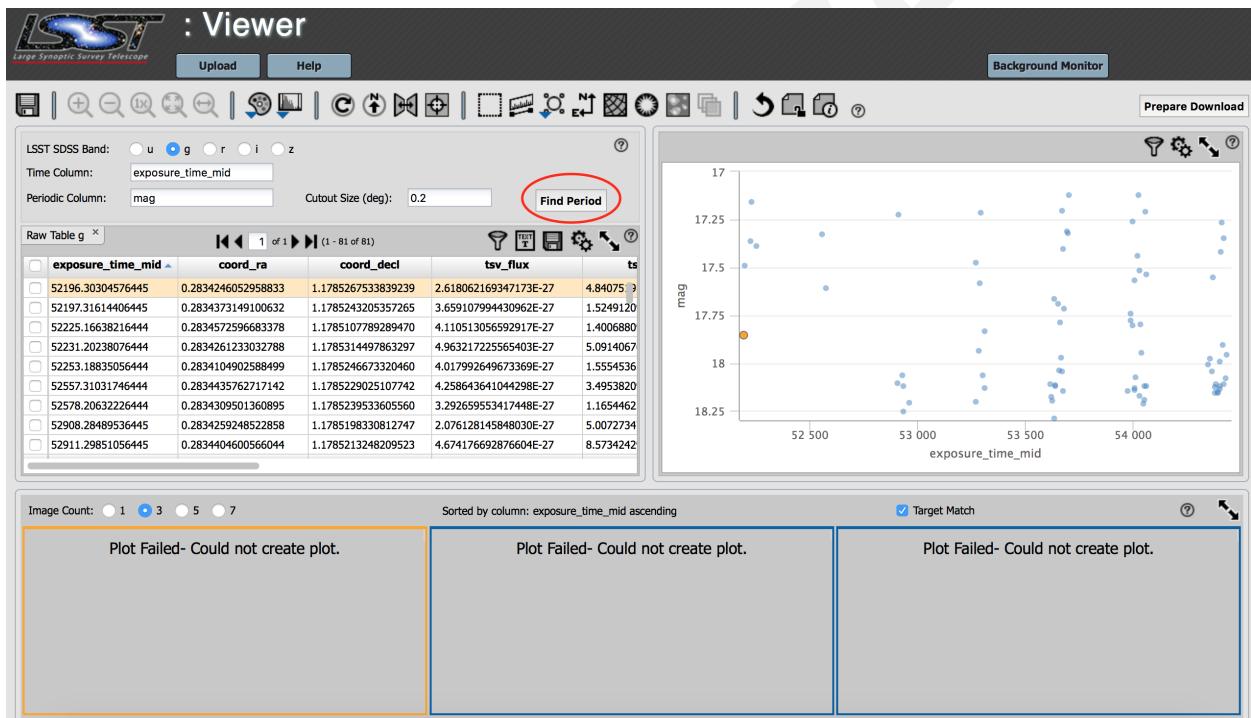


FIGURE 17: Time Series Viewer for an objectId=3588818166880604, at ra, dec = 0.283437°, 1.178522°. Note that initially on the bottom there are three empty panels. The radio buttons in the upper left corner allow intuitive selection of the SDSS filter for the periodogram calculation. We choose g-band, and click 'Find Period', marked with the red oval, to calculate Lomb-Scargle periodogram for that band (this takes the user to Fig 18)

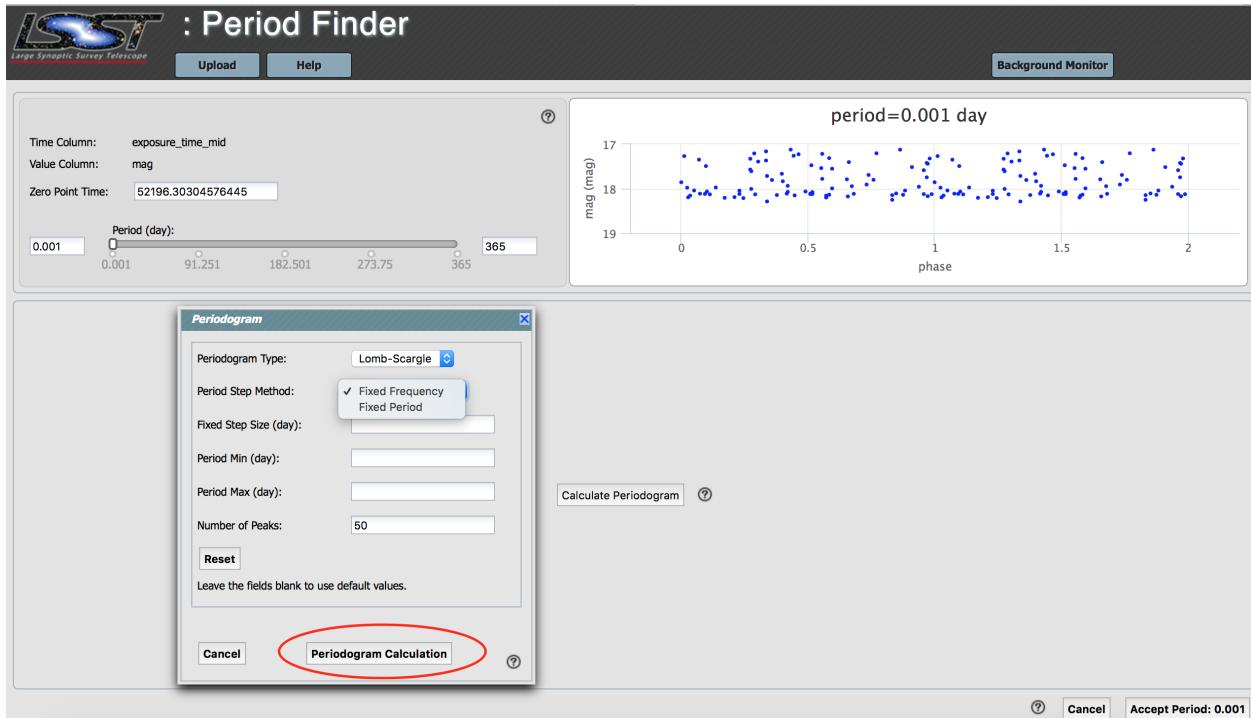


FIGURE 18: Calculating Lomb-Scargle periodogram for an objectId=3588818166880604, at ra, dec = 0.283437°, 1.178522°. The slider in the upper left corner allows to fold the light curve on a chosen period. Clicking 'Calculate Periodogram' button opens the dialog window 'Periodogram'. Currently it contains only the Lomb-Scargle Periodogram Type. Period Step Method include Fixed Frequency or Fixed Period, similar to the NASA Exoplanet Periodogram Tool (Fig. 26). If we don't choose anything for maximum and minimum periods, the calculation will proceed with defaults, which for this RR Lyrae fails to detect the true period (Fig. 19). 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 ( 20), although the method itself is very sensitive to frequency grid. Clicking on 'Periodogram Calculation' proceeds with evaluating Lomb-Scargle periodogram with chosen Period Step Method using the default frequency grid (see Fig. 19)



FIGURE 19: Calculating Lomb-Scargle periodogram for an objectId=3588818166880604, at ra, dec = 0.283437°, 1.178522°. Using the default settings does not recover the true underlying period of 0.547987 days. See Fig. 20 for a more appropriate choice of period range.

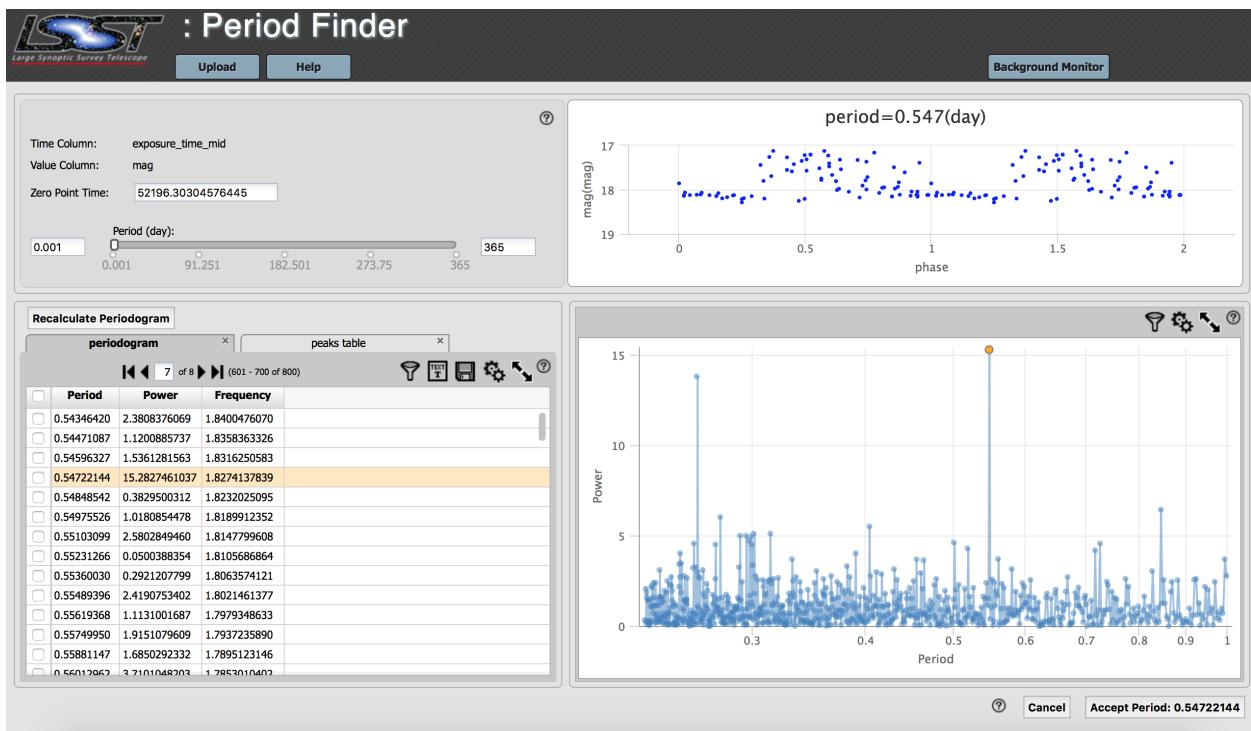


FIGURE 20: Calculating Lomb-Scargle periodogram for an objectId=3588818166880604 at ra, dec = 0.283437°, 1.178522°. When we appropriately constrain the frequency grid on which the powers of periodogram are evaluated, we recover the period close to the true period of 0.547987 days. On this figure we choose fixed frequency method, with  $P_{min} = 0.229$  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 light curve on any period does not update to the values used in the Periodogram search. However, slight change of frequency range can drastically improve the calculation: see Fig 21.

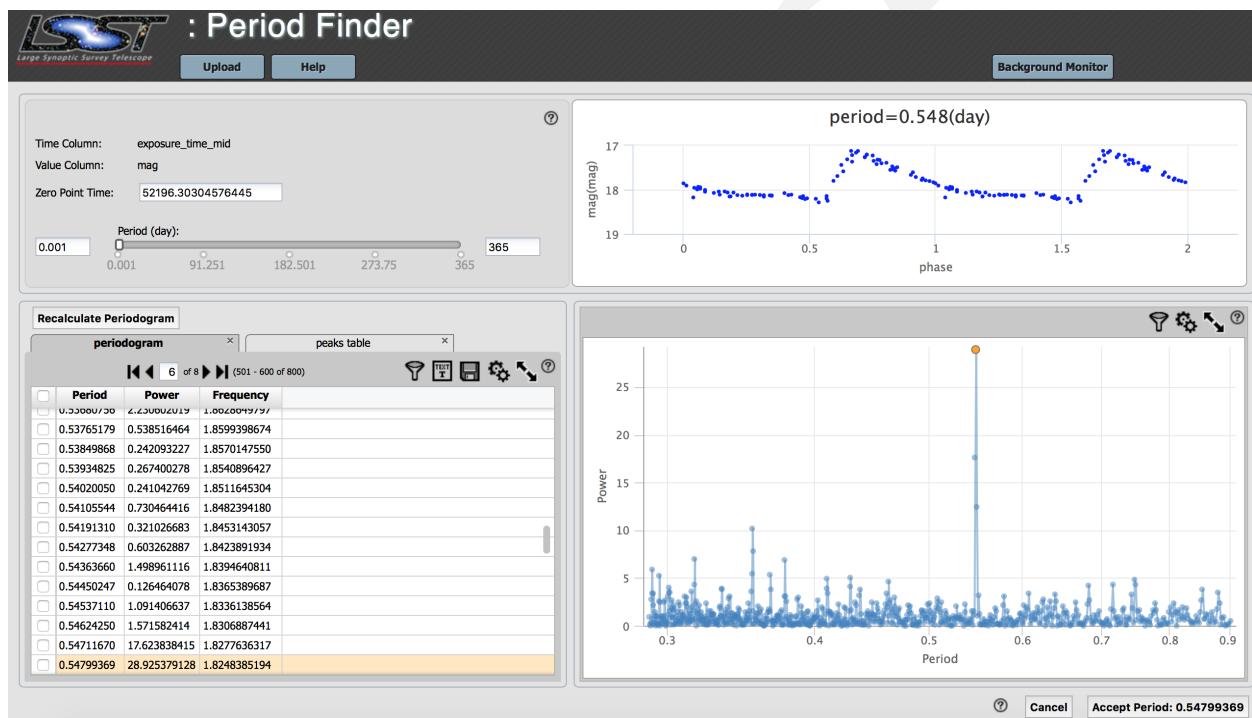


FIGURE 21: Calculating Lomb-Scargle periodogram for an objectId=3588818166880604 at ra, dec = 0.283437°, 1.178522°. Slight change of frequency grid can heavily affect the outcome : here we use fixed frequency, with period chosen between 0.29 and 0.9 days. Clicking 'Accept Period' takes the user to Fig. 22.

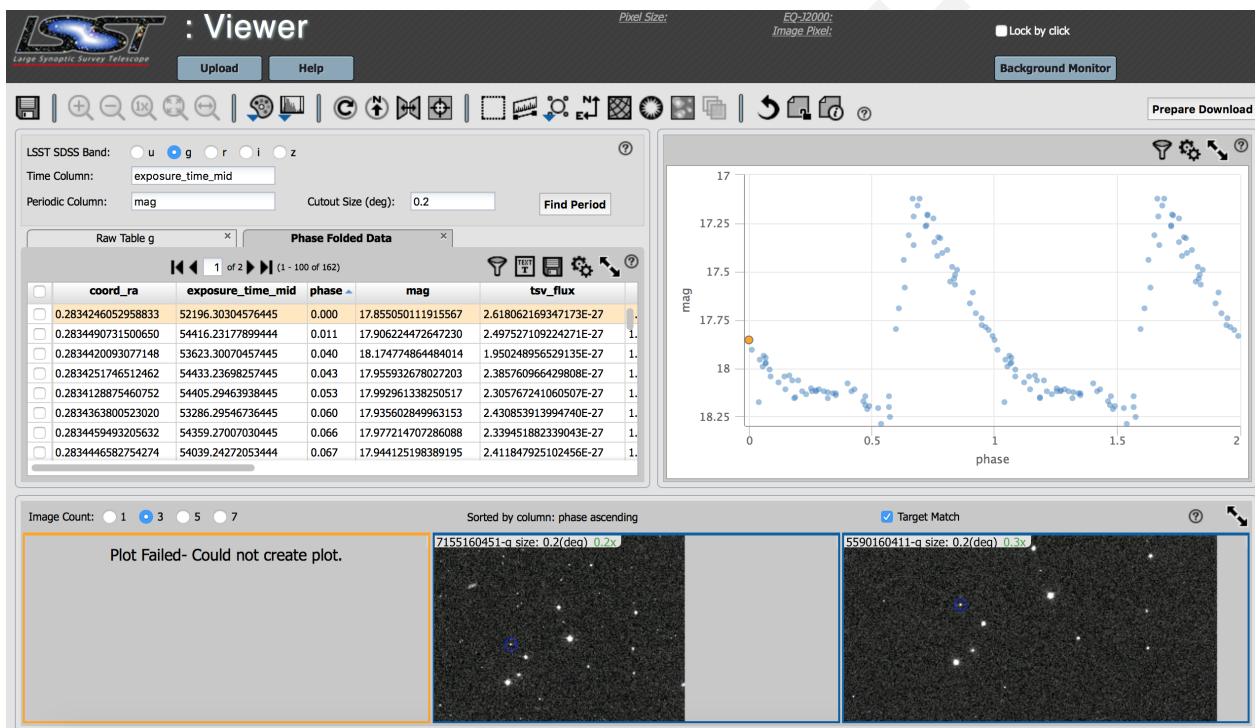


FIGURE 22: The result of accepting the period found by limiting the periodogram bounds by  $P_{min} = 0.29$  and  $P_{max} = 0.9$  days, for an RR Lyrae ID=13350 (objectId = 3588818166880604, at ra, dec =  $0.283437^\circ, 1.178522^\circ$ ). It is not clear what the miniatures show : different epochs? The light curve is correctly folded on g-band data. A surprising behavior here is that selecting a different band (eg 'r'), instead of folding the light curve in that band on the accepted period, the raw data is displayed (Fig. 23)

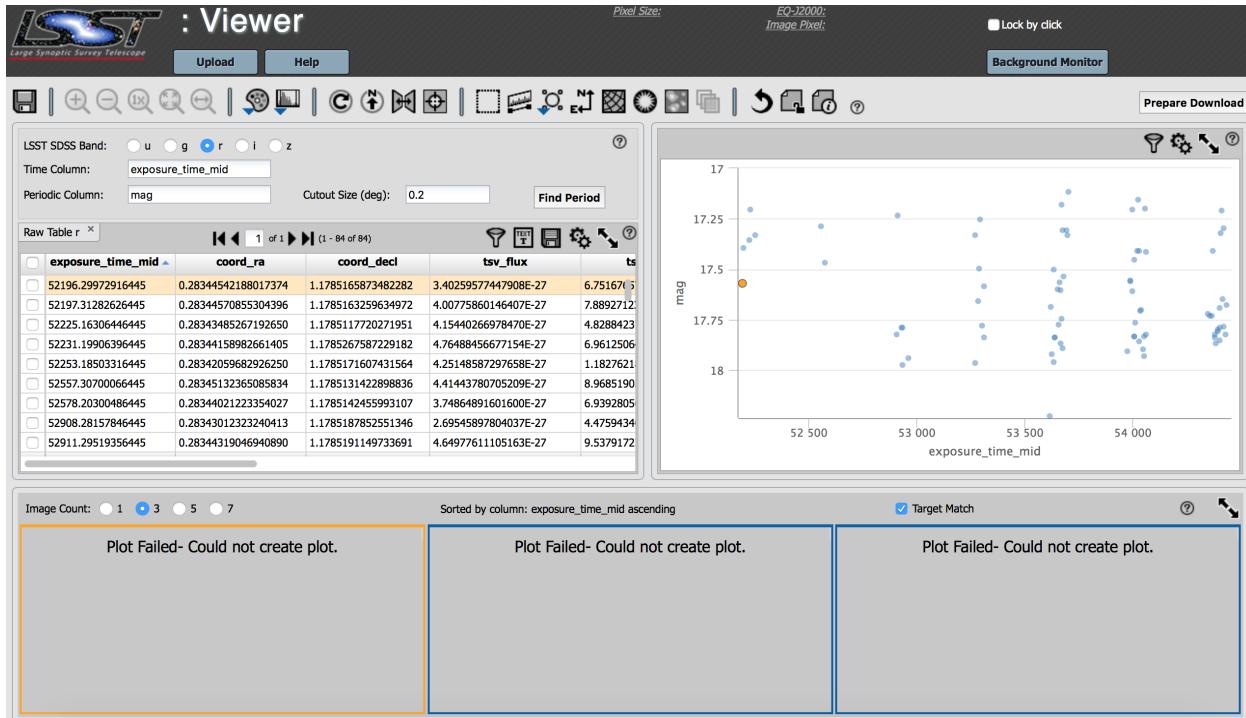


FIGURE 23: The result of clicking on the radio button for the r-band data. Instead of showing the r-band light curve folded on the accepted period, we see the raw r-band data. Clicking the 'Find Period' button does not 'remember' the result of the previous search on the same data.

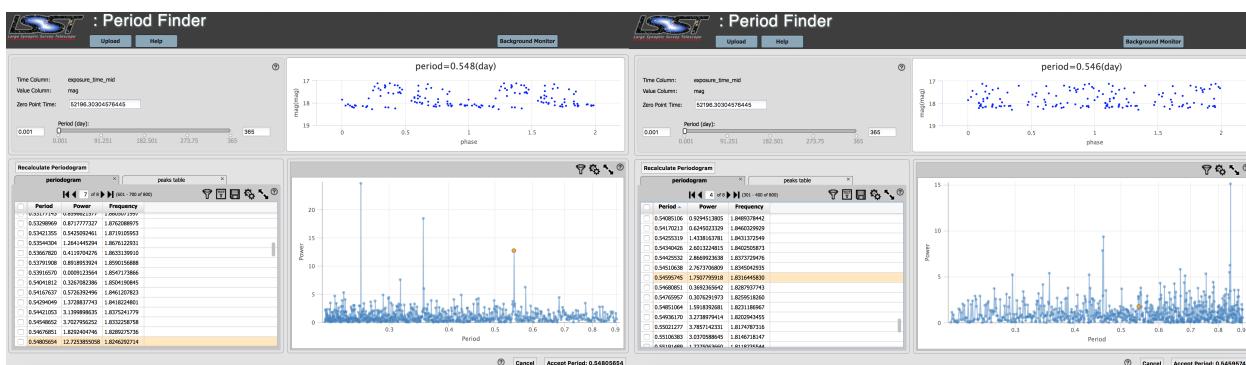


FIGURE 24: Calculating Lomb-Scargle periodogram for an objectId=3588818166880604 at ra, dec = 0.283437°, 1.178522°. Bounding period between 0.22-0.9 days, the left panel used 'Fixed Frequency', and the right panel 'Fixed Period'.

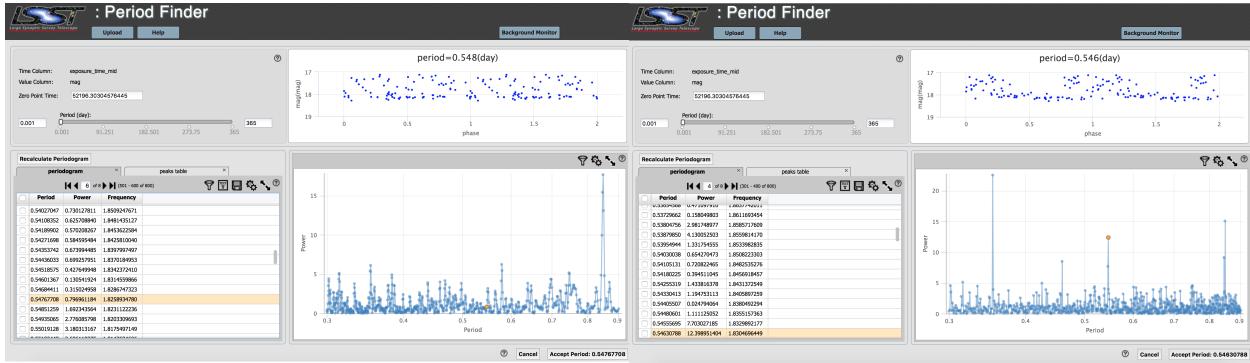


FIGURE 25: Calculating Lomb-Scargle periodogram for an objectId=3588818166880604 at ra, dec = 0.283437°, 1.178522°. Bounding period between 0.3-0.9 days, the left panel used 'Fixed Frequency', and the right panel 'Fixed Period'.

## 5.1 Single forced photometry light curve

We compare a forced photometry light curve for a single object from the S82 dataset between the PDAC copy and the locally stored version (at the University of Washington - UW). We used an objectId = 216172782516437336, in patch g00\_22, with 72 forced photometry epochs in g-band. From available detection data (analogous to Deep Source table at PDAC) we find ra, dec = 359.974019436°, -1.25626927667°, extendedness= 1.

### 1. Title : can we query Deep Source table by objectId ?

- Description : we have both the ra, dec coordinates and objectId of a source. We test if it is possible to query solely by objectId, without having to specify ra, dec in cone query.
- Input : cone query the Deep Source catalog for an objectId = 216172782516437336 at ra, dec = 359.974019436°, -1.25626927667°, search radius 3 arcsec. Add an SQL constraint coadd\_filter\_id = 3, objectId = 216172782516437336. Try to query without specifying ra,dec.
- Result : negative. It is necessary to specify ra,dec, even if the objectId is known. However, we find the sought object. We recommend that querying by objectId be implemented in the next release of the UI .
- Date: 5/22/2017

### 2. Title : is the number of photometric points between UW and PDAC copy of S82 dataset identical?

# NASA EXOPLANET ARCHIVE

## NASA EXOPLANET SCIENCE INSTITUTE

**Home    About Us    Data    Tools    Support    Login**

**Periodogram Inputs    Edit Input Table    Plot Input    Results**

Periodogram Inputs

Input File Options		Algorithm and Period Settings	Reset
<b>Upload Data File:</b> <a href="#">?</a> <input type="button" value="Choose File"/> <input type="text" value="13350_g.txt"/> <input type="button" value="Upload"/>  <b>Current Periodogram Data File:</b> Name: <b>13350_g.txt</b> Source: <i>user uploaded file</i> <input type="button" value="Edit Input Table"/>  <b>Select Column Names:</b> Time Column: <input type="text" value="col1"/> <input type="button" value="▼"/> Data Column: <input type="text" value="col2"/> <input type="button" value="▼"/> <input type="button" value="Plot Time vs. Data Columns"/>  <b>Input File Information:</b> Points used: 58 of 58 Time range: 51075.302311 to 54412.235925 Data range: 17.113 to 18.242		<b>Select Algorithm:</b> <a href="#">?</a> Algorithm: <input type="text" value="Lomb-Scargle"/> <input type="button" value="▼"/>  <b>Period Range:</b> Minimum Period: <input type="text" value="0.228731"/> Maximum Period: <input type="text" value="0.998246"/>  <b>Period Step Method:</b> <a href="#">?</a> Select Method: <input type="text" value="Fixed Frequency"/> <input type="button" value="▼"/> Fixed Step Size: <input type="text" value="0.0001226"/>	
Default(s) calculated successfully.			
<input type="button" value="Calculate Periodogram"/> <input type="button" value="Start New Session"/>		Calculation Name: <input type="text" value="13350_g.txt"/> <input type="button" value="?"/>	
<i>Estimated processing time: 15 seconds</i>			

FIGURE 26: The same object as Fig. 39, and Fig 42, 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 light curve (compare to few milliseconds of Astroml code naive single-sinusoid approach that gave the same result for this particular object)

- Description : we compare the local copy of the g-band photometry for objectId = 216172782516437336 vs PDAC hosted version.
- Input: cone query the Deep Forced Source catalog for an objectId = 216172782516437336 at ra, dec = 359.974019436, -1.25626927667 , search radius 3 arcsec. Add an SQL constraint id = 216172782516437336 and exposure\_filter\_id = 1.
- Result : positive. This query yields 73 epochs as compared to UW 72 epochs, but upon examination 1 epoch of the PDAC result(exposureid 6518110430, exposure\_time\_mid = 54040.248144284444 ) has flux\_psf\_err flux\_psf = None. Recommendation : forced flux measurements where flux is not available should not be returned to the user without warning, or an appropriate information (eg. 'for this object there are also N epochs where the reported flux\_psf is None, available here'). After removing the non-detection, the number of epochs for forced photometry measurements in PDAC-hosted light curve is identical to the local copy.
- Date: 5/24/2017

### 3. Title : are the time stamps for each forced photometry epoch the same ?

- Description : we compare the light curve resulting from test above, focusing on the reported time per exposure : are the times identical between PDAC and UW ?
- Input: light curve from PDAC and light curve from UW for objectId = 216172782516437336.
- Result : negative. The UW light curve has a mean offset of 26.95 seconds vs PDAC light curve for the same object : approximately half of the mean exposure time. UW data records the beginning of the exposure, whereas PDAC : the middle of the exposure. Although metadata for UW does not allow to guess that information (calling time column 'mjd'), the PDAC column name makes this distinction more clear by calling time 'exposure\_time\_mid'. Recommendation : given that it causes a small discrepancy in coincidence measurements, it would be prudent to perhaps provide both the beginning of the exposure and the middle of the exposure, unless providing the time of the beginning is a very nonstandard practice. Note that the mjd data stored at PDAC has more significant digits (12-13 in PDAC data compared to 6-7 at the UW data).
- Date: 5/25/2017

### 4. Title : are the units for flux measurements between PDAC and UW the same ?

- Description : we test whether the flux units returned by PDAC query of Run Deep Forced Source table allows for direct comparison with the UW data, and if not, whether it is possible to make them comparable using only the UI.

- Input : query against the objectId = 216172782516437336, using raw flux.
- Result : negative. Data at UW is already calibrated for zero-point, thus is stored in ergs/sec/cm<sup>2</sup>/Hz, whereas the PDAC data is stored in ADUs. It is currently impossible to obtain calibrated fluxes using the PDAC UI : an example of an SQL query to obtain calibrated fluxes, requiring a join between RunDeepForcedSource table, for the information about the individual observations, and the Science\_Ccd\_Exposure table (for a per-visit-image photometric zero point), is on the PDAC Sample Queries. Recommendation : allow the user to query RunDeepForcedSource and offer options to return uncalibrated flux measurements, or a calibrated flux making a join query with Science\_Ccd\_Exposure in the background.
- Date: 5/27/2017

5. Title : are the calibrated fluxes from PDAC identical as those at the UW ?

- Description : we compare the calibrated fluxes obtained by SQL query against RunDeepForcedSource and Science\_Ccd\_Exposure table in g-magnitude for the objectId = 216172782516437336, and the UW hosted forced photometry data.
- Input : calibrated flux light curve for objectId = 216172782516437336, using SQL scisql\_dnToFlux and scisql\_dnToFluxSigma functions to convert raw to calibrated fluxes (see Sec. A.1 for the SQL query).
- Result : positive. If we remove the missing photometry datapoint from PDAC, then there is an identical number of epochs. The mean residual difference between PDAC and UW calibrated flux data is zero, and all the difference can be ascribed to floating point arithmetic (different accuracy of data stored at UW and PDAC). See Fig. 27 for the histogram of difference between UW and PDAC fluxes. Mean flux ratio is 1.0000003263832988 . On Fig. 28 we plot the PDAC and UW data as time series, with the residuals representing flux difference between the two sources of data.
- Date: 5/27/2017

Forced photometry measurements may report flux values that are below zero, or are very noisy ( $S/N < 2$ ). For instance, for objectId=216172782516437336, 24 epochs have  $S/N < 2$ . We developed a pipeline<sup>18</sup> that imposes a uniform prior on each flux measurement in accordance with the non-physicality of negative flux values. Thus we provide re-estimate of the faint flux

<sup>18</sup>[https://github.com/suberlak/Faint\\_pipeline\\_report](https://github.com/suberlak/Faint_pipeline_report)

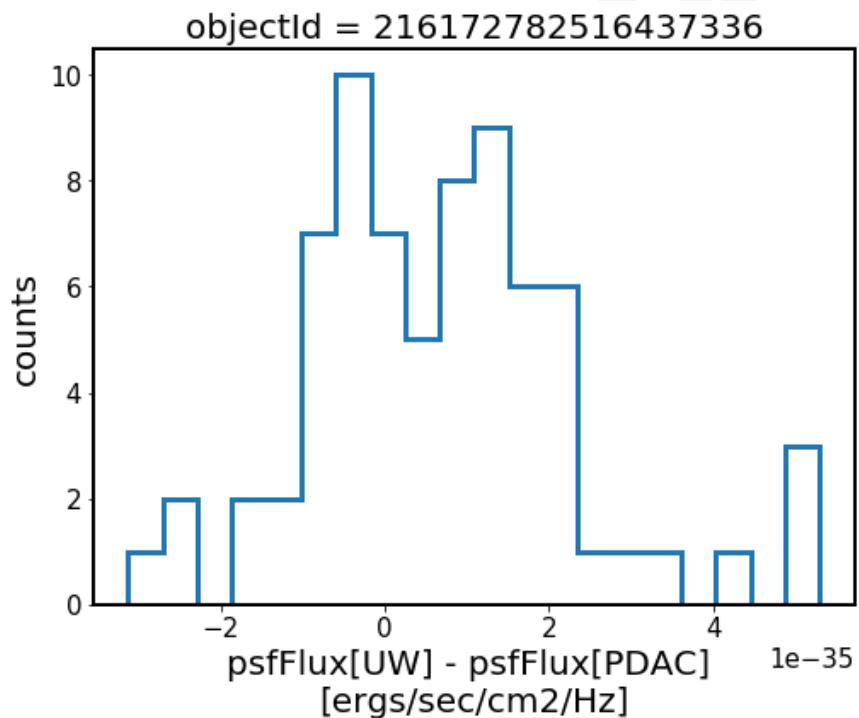


FIGURE 27: Histogram of flux difference between data stored at PDAC vs UW for objectId = 216172782516437336. The difference is less than 1E-35, i.e. on the level of the floating point error. This proves that once fluxes are correctly calibrated (using the zero point magnitudes from Science\_Ccd\_Exposure table), light curves at PDAC represent approximately identical information as those stored at UW.

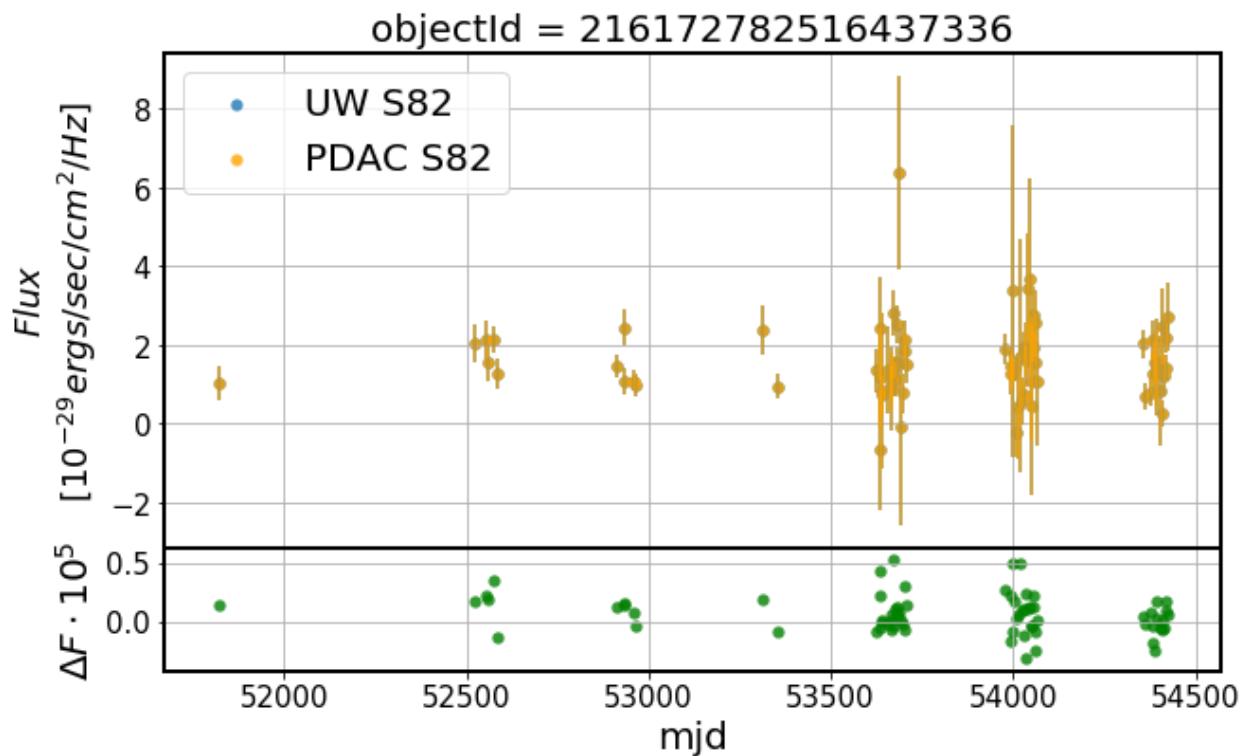


FIGURE 28: Calibrated flux for objectId= 216172782516437336, same as Fig. 27. Light curves at PDAC and UW are identical once brought to a common unit system. The bottom panel shows the difference between PDAC and UW light curve - the values are multiplied by 1E5 to show the difference (in units of  $10^{-29} \text{ ergs/sec/cm}^2/\text{Hz}$ , so that the order of flux difference is 1E-34).

(S/N < 2) measurement, and a magnitude for each forced-photometry epoch. We constrain that with the PDAC scisql\_dnToAbMag function which simply removes the points with negative flux from the light curve. The impact of this pipeline on the UW data is shown in Fig. 29

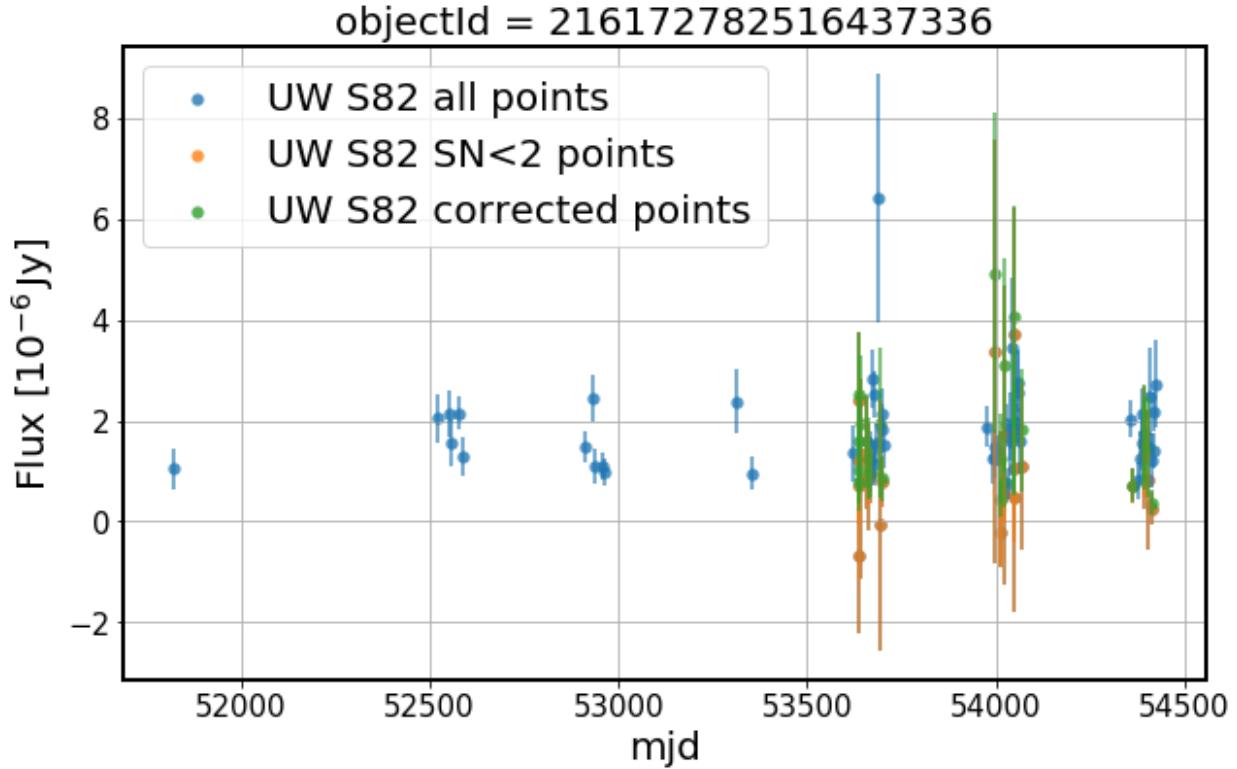


FIGURE 29: Light curve from UW for objectID=216172782516437336, highlighting the faint points, with S/N < 2. The green points show the effect of applying a non-negative Bayesian prior on forced photometry.

Given that as we have shown, the flux measurements contained in light curves at PDAC and UW are identical, we show the impact of correction of faint flux measurement on PDAC data on Fig. 30.

Given the impact of correction on flux, magnitudes would also be affected. We show the impact of corrected-flux magnitudes vs the scisql function magnitudes. There is no change where S/N > 2, but for faint points, the corrected fluxes yield smaller magnitudes. On Fig. 31 we juxtapose the PDAC SQL-computed magnitudes (using SQL query in Sec. A.2), vs. magnitudes based on PDAC SQL-computed fluxes, corrected using custom pipeline wherever S/N < 2.

In conclusion, based on the case of objectID = 216172782516437336, light curves from UW

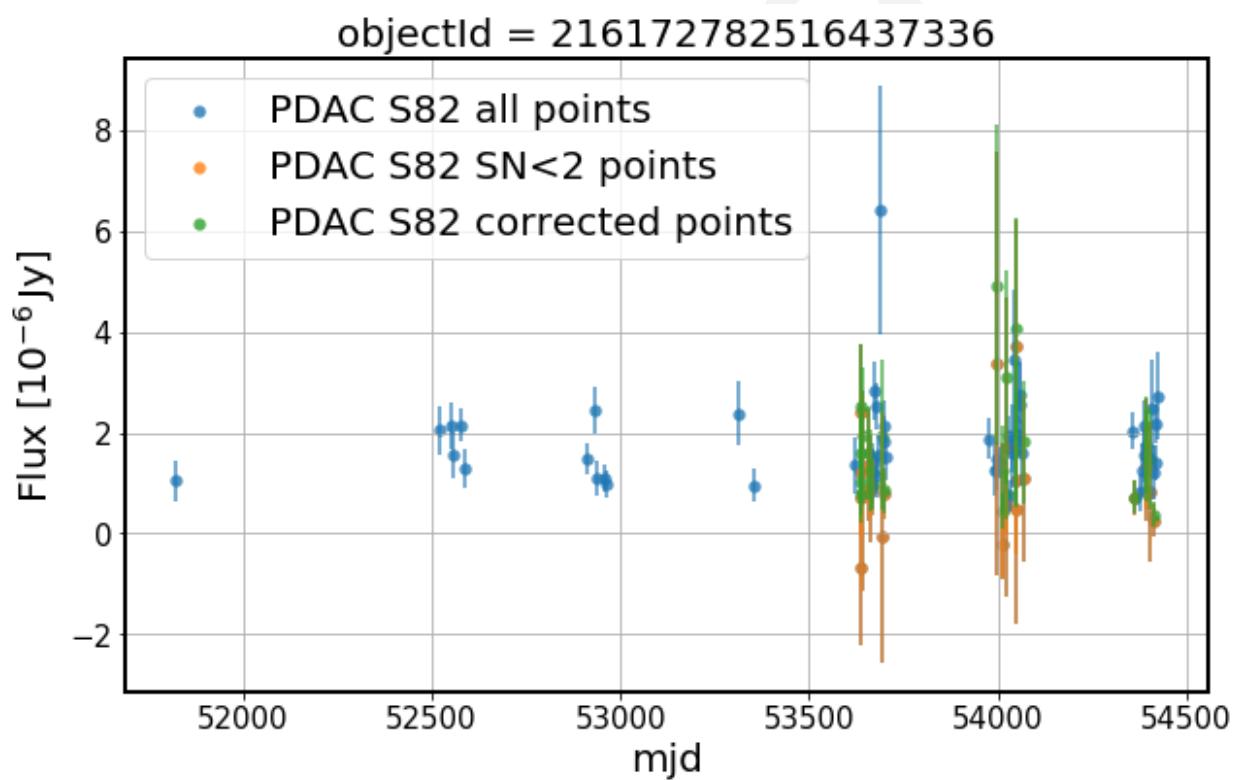


FIGURE 30: Same as Fig. 29, but using PDAC data.

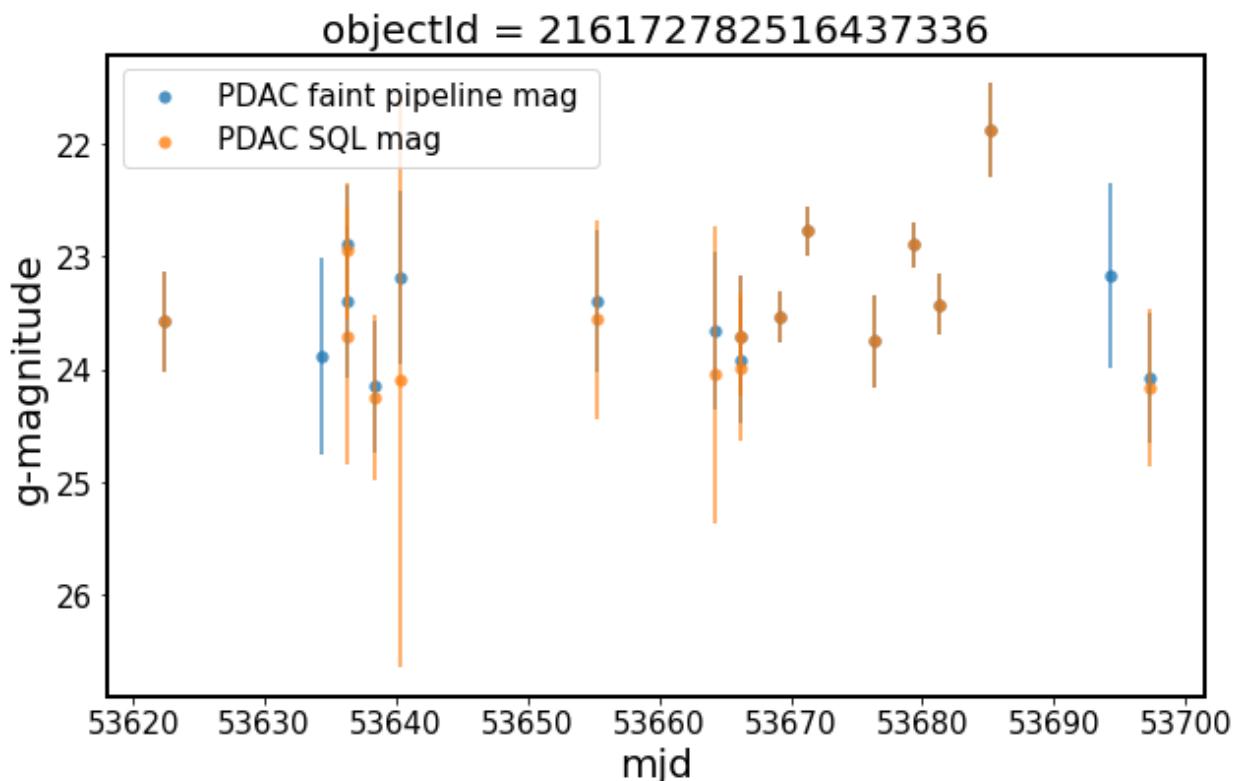


FIGURE 31: A section of light curve for objectId=216172782516437336, comparing the magnitudes for PDAC data calculated with the built-in `scisql_dnToAbMag` and `scisql_dnToAbMagSigma` functions (orange), and the magnitudes based on PDAC calibrated fluxes, treated with the faint fluxes pipeline (blue). Note that in cases where PDAC flux was below zero, the built-in `scisql` pipeline ignores these points in magnitude calculation, since magnitude is undefined for negative flux. Using our Bayesian pipeline, we put a better estimate on the upper limit of the flux measurement.

and PDAC are identical to the precision afforded by the floating point error. However, accessing the PDAC calibrated magnitudes is nontrivial : we recommend that the functionality of executing a joint query between DeepForcedSource and Science\_CCD\_Exposure to return calibrated fluxes and magnitudes be accessible via the UI. In addition, assumptions that come into computing magnitudes from fluxes should all be displayed and available to the user (eg. where the zero point magnitudes are stored to calculate calibrated fluxes, how only non-negative flux values are used to calculate magnitudes, what formulae are used in that conversion). This would greatly improve the educational value of SUIT. Furthermore, user experience would significantly improve if it were possible to query either DeepSource or DeepForcedSource directly by the objectId , without the need to specify ra, dec.

## 5.2 Positional comparison : box query

We compare the source density between the PDAC and UW S82 datasets. This is a simple test that can help confirm whether the two datasets are identical, as we would expect. Any irregularities and information on how they differ would highlight a particular aspect of the PDAC S82 dataset that should be brought to the attention of the user. Steps required to make this comparison also highlight direct ways to improve the user experience. The UW data is a deep source coadd catalog, limited at  $i < 23.5$ , containing all detected sources in i-band. The PDAC Run Deep Source catalog contains coadds in each of the five bands, and is not limited in magnitude depth.

### 1. Title : can we apply the magnitude cutoff in a box query within the UI ?

- Description : limiting the maximum magnitude of sources returned in a query is an important functionality - given the understanding about the limiting magnitude of the survey, the user may wish to select a good quality measurements by simply selecting the cutoff magnitude below the limiting magnitude of the survey. Although coadds greatly increase the survey depth, there is a lot of faint sources for which data quality may be insufficient for scientific purpose. The magnitude cutoff also speeds up the query, since the number of sources grows exponentially with magnitude.
- Input : query Run Deep Source using box query with ra, dec limited by (325, -0.7) and (-0.6, 325.1). Use SQL constraint coadd\_filter\_id = 3 . Seek to limit the i-band magnitude to 23.5.

- Result : negative. We recommend that in the future version of the PDAC, obtaining calibrated fluxes and magnitudes for objects in both Deep Source and Deep Forced Source tables be available via the UI. An additional feature could be allowing the user to limit the magnitude of returned objects. This is currently only possible with an SQL query (see Sec. A.3, using HAVING  $i < 23.5$  clause, where  $i$  magnitude is the outcome of join between DeepForcedSource and Science\_CCD\_Exposure).
- Date: 5/30/2017

## 1. Title: is the source density the same between UW and PDAC data?

- Description: we compare the source density in the S82 dataset using PDAC Run Deep Source catalog and the locally hosted (UW) Deep Source catalog. The local catalog is cut at  $i < 23.5$  mag. The two should agree on the number of sources in the same area of the sky.
- Input: Use box query against the PDAC S82 data. The query box in the SQL `qserv_areaspec_box` function requires minimum and maximum ra,dec bounds. Use  $325^\circ$ ,  $325+D^\circ$ , and  $-0.7^\circ$ ,  $-0.7+D^\circ$ , with  $D=[0.1, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4]$ . This specifies box queries with increasing area (see Fig. 32). Measure the query speed including the download of parent id and object id, i-band coadd magnitude and error. Limit the query depth to  $i < 23.5$  to mirror the UW data. Select only i-band coadd by specifying `RunDeepSource.coadd_filter_id = 3`. Measure the time it takes to query the QSERV and download the data. Count the number of unique id in the query. Compare to the number of objects in the same area from the UW S82 Deep Source catalog.
- Result: negative. Number of sources in PDAC per chosen area, constrained in the same way as the UW catalog, is  $\approx 15\%$  higher (see Tab. 2 for detailed results). This is the result of choosing whether to include deblender parents. In all cases, we find that the 'id' of extra sources at the same location as sources matched to UW catalog are the same as 'parentDeepSourceId' in the UW catalog. We recommend that this issue be explained to the user in more detail, in particular the meaning of parent, or deblend\_n\_child column names. Otherwise searches for objects within a certain area may yield unexpected results. As predicted, the measured query time increases with the size of the queried area (Fig. 33), allowing within less than a minute to investigate the area of a few square degrees. See Sec. A.3 for an example of an SQL query used.
- Date: 5/29/2017

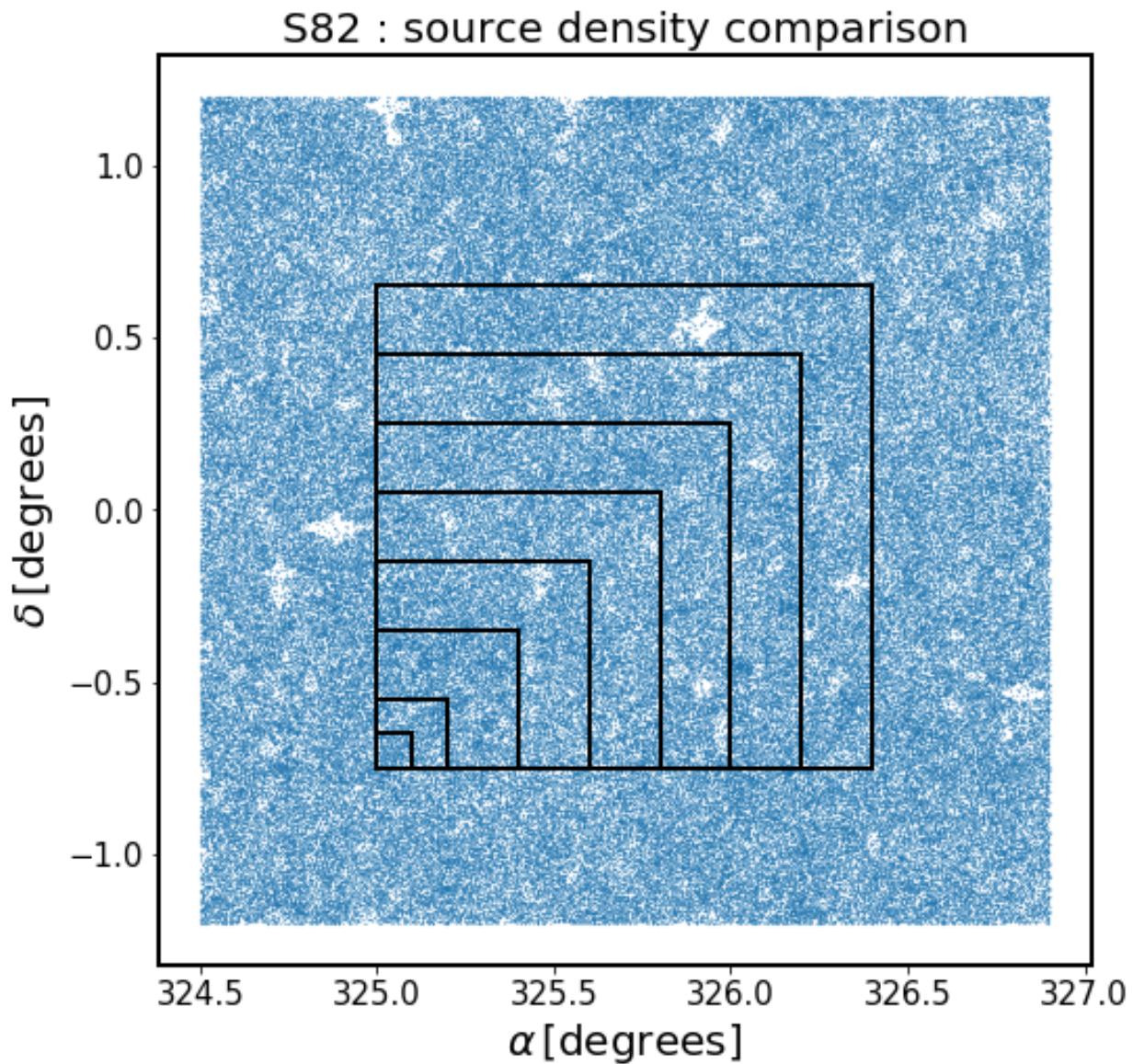


FIGURE 32: The area of comparison between the PDAC and UW S82 Deep Coadd catalogs. In the background the S82 source counts from the UW catalog. Each black square illustrates the query area.

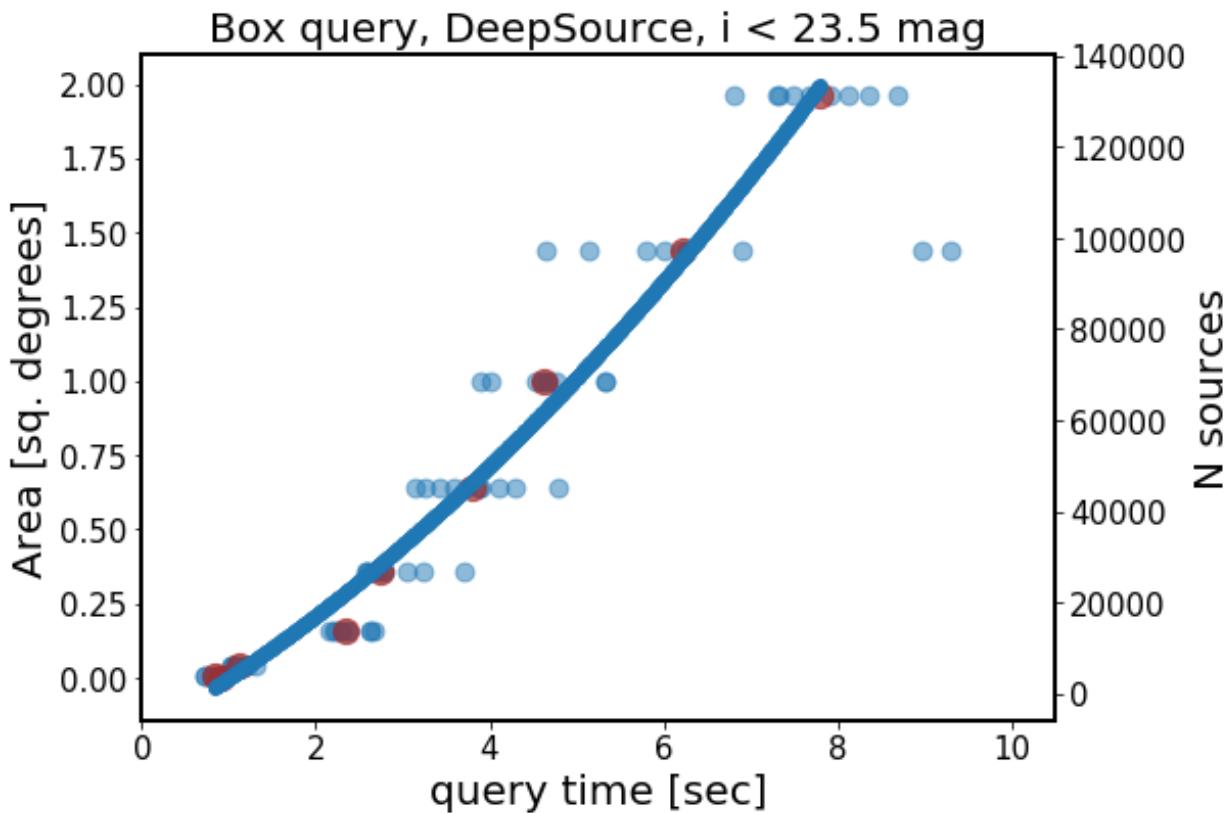


FIGURE 33: Result of a box query against PDAC S82 dataset with increasing box size. We plot query area in square degrees and the number of sources in PDAC contained in that area against query time. There is a scatter in query time corresponding to fluctuating network speed, thus we fit the polynomial to the median query time to highlight the increasing trend. This also emphasizes the importance of allowing the user to limit the query magnitude : otherwise the number of sources resulting from a query may be prohibitively large.

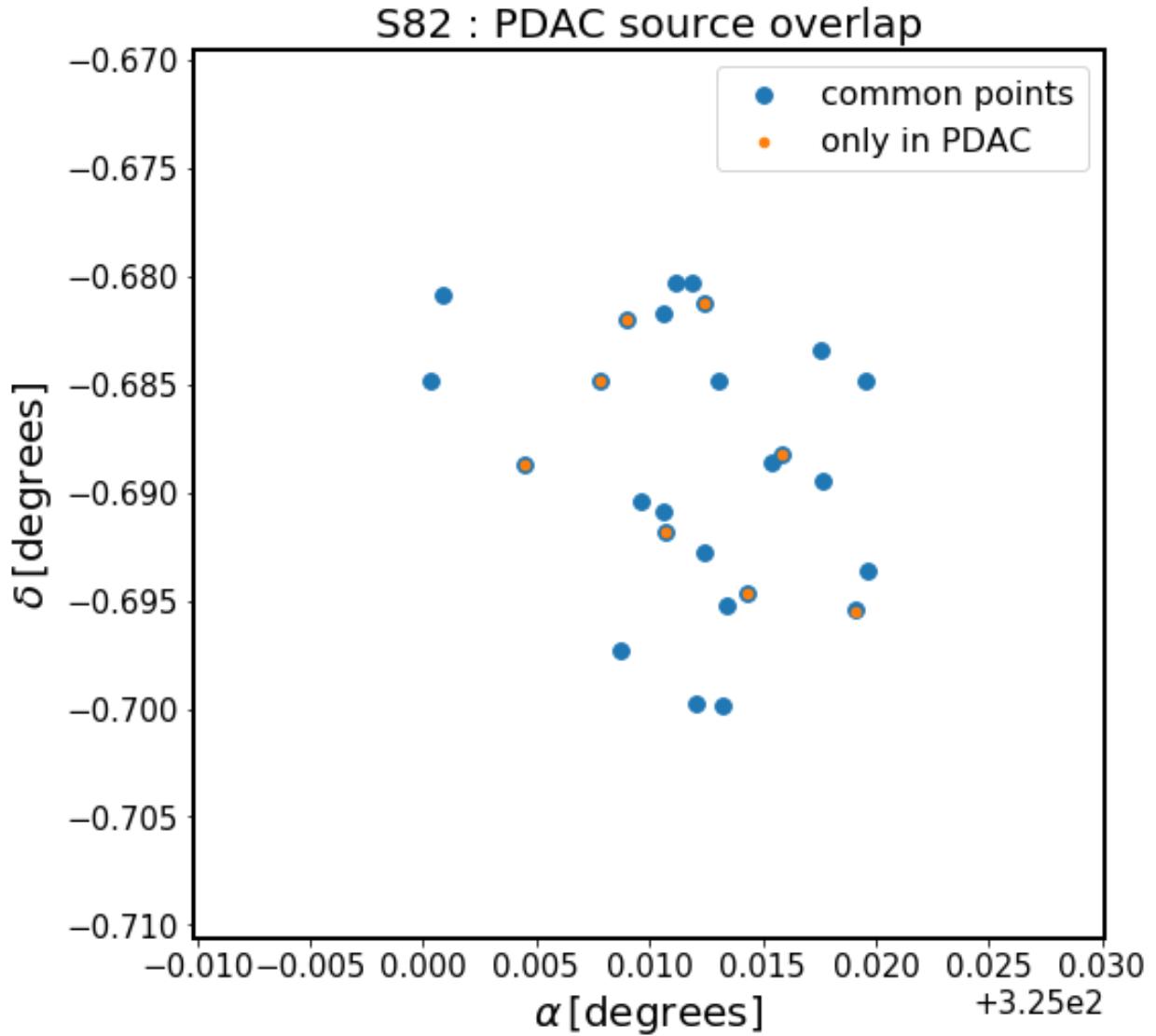


FIGURE 34: Comparing sources in a very small area (1.44 square arcminutes) between PDAC and UW hosted Deep Source catalogs, bounded by a Polygon with vertices (325.0 -0.7, 325.02 -0.7, 325.02 -0.68, 325.0 -0.68). Blue dots represent sources with id shared between the two catalogs. Orange dots represent sources with id only present in the PDAC sources catalog. These are child sources with a parent at the same location. For example, consider sources id = 217571363454454701 and 217571363454457757, both located within 1 arcsec from ra,dec = 325.008953599611, -0.6820032359336851 . Both are brighter than  $i = 23.5$  mag, but 217571363454454701 is the parent, and also has 4 other deblender children (deblend\_nchild = 5 for 217571363454454701), that are fainter than  $i = 23.5$  mag.

TABLE 2: Comparison of the number of sources in the same area between PDAC and UW hosted versions of the Deep Source Coadd catalogs. The number of unique id's is considered to be a proxy for the number of sources. If we exclude objects that are parents of the UW sources at the same location, we come closer to resolving the discrepancy (we remove rows where 'id' is the same as a 'parent' of another object). However, there are still objects which may or may not have any match in the UW catalog. Those could be 'parent' sources of 'child' sources that are too faint to be included in the query ( $i < 23.5$  limit).

Area [sq.deg]	N(PDAC)	N(UW)	N(PDAC) no parents
0.01	731	562	583
0.04	3020	2273	2350
0.16	12320	9170	9533
0.36	25854	19154	19909
0.64	44880	33074	34371
1.00	69374	51077	52989
1.44	98871	72876	75506
1.96	133916	99007	102630

### 5.3 Light curve comparison : multiple cone queries

We assess whether the Lomb-Scargle periodogram can be safely used to determine periods for the data similar in cadence and sparseness to the S82 S13. This may inform the choice of preset settings for the Period Search tool in the Time Series View in PDAC UI. We use as the 'true' period the best-fit results from detailed template fitting [2]. Both fit parameters and SDSS DR2 light curves used in [2] are publicly accessible in the online version of the journal.

We obtain the S82 S13 calibrated g-magnitude light curves from PDAC. This requires simultaneous cone query of Deep Forced Source catalog at the given position, and cross-matching against Science CCD Exposure, currently only available via SQL query. As shown on Fig. 37, only 343 of 483 RR Lyrae are at locations within the PDAS S82 S13 dataset. For each light curve, using the *astroML* python module [3], we calculate the Lomb-Scargle periodogram powers on a uniform frequency grid of 5000 frequencies spanning between  $\omega_{min} = 0.9(2\pi/P_{max})$ ,  $\omega_{max} = 1.1(2\pi/P_{min})$ , where  $P_{min}$  and  $P_{max}$  correspond to 90% of the smallest and 110% of the largest periods in the catalog. Thus we are constraining the periodicities searched exactly where we expect them to be. For periodogram we assess the significance of the peak by performing 500 bootstrap resamplings<sup>19</sup>. We use the generalized Lomb-Scargle as the calculation mode, following the defaults in *astroML* (see also Eq.20 in [6], and Section 10.3.2 in [1]).

<sup>19</sup>[http://www.astroml.org/book\\_figures/chapter10/index.html](http://www.astroml.org/book_figures/chapter10/index.html)

TABLE 3: Comparison of RR Lyrae periods obtained with different methods. First column - 'true', is the 'ground truth' - period resulting from detailed template fitting by [2]. Second column - DR2 LS, is the period corresponding to the most prominent frequency in the Lomb-Scargle periodogram computed on the SDSS DR2 light curve from [2]. Third column - DR2 EXO, shows the period found for SDSS DR2 data of [2] with the NASA Exoplanet Archive Periodogram service. Fourth column - PDAC LS, is the period found using Lomb-Scargle periodogram on PDAC S82 S13 data.

ID	true	DR2 LS	DR2 EXO	PDAC LS
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

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 light curve folded on the 'true' period look physical?
2. Using naive Lomb-Scargle, can we find this period using the SDSS DR2 data originally used by Sesar+2010 ?
3. Using naive Lomb-Scargle, can we find this period using the S82 S13 PDAC data?

We show that even with such a densely sampled Lomb-Scargle periodogram (100 times denser than the default 50 powers calculated in the PDAC Time Series Viewer), choosing a very appropriate frequency range we are able to recover the true period only for about a half of the tested stars (see Fig. 36)

We summarize the results of Lomb-Scargle calculation on SDSS DR2 data into following groups: where with the LS we find the same period (Fig. 39), a smaller period (Fig. 40), , or a bigger period (Fig. 41) than the ground truth. This proves that with SDSS DR2 data it would be hard to use solely LS for period finding.

Using the PDAC S82 S13 data we also calculate LS periodogram for the same objects - see Figs. 42, 43 and 44.

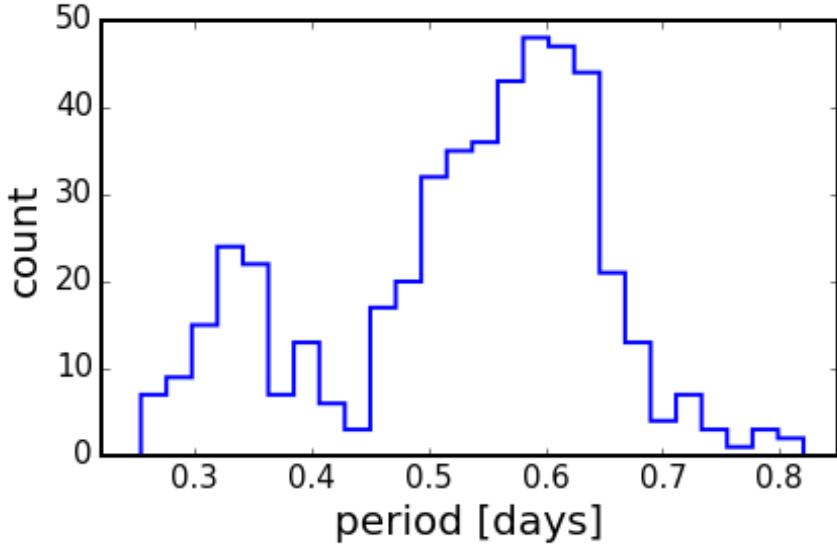


FIGURE 35: Distribution of RR Lyrae periods for 483 objects in [2]. Note the bimodal distribution, reflecting two main RR Lyrae types : 309 RRab (right) and 104 RRc (left) (see also Fig.16 in [2]).

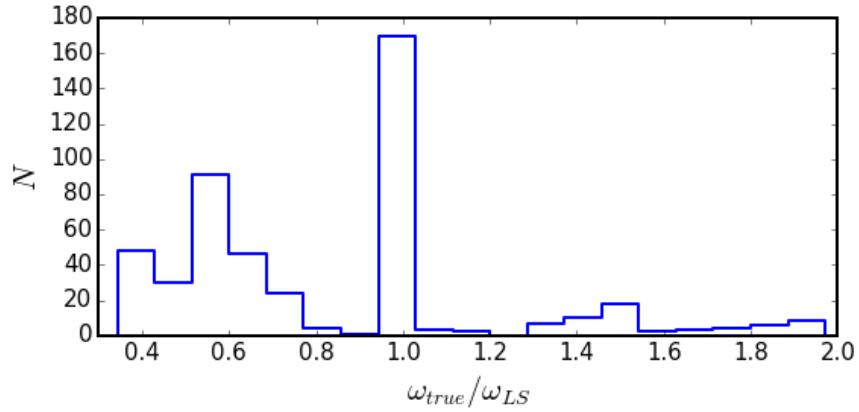


FIGURE 36: The distribution of the ratio of  $\omega_{true}$  to  $\omega_{fit}$ .  $\omega_{true} = 2\pi/P_{true}$ , is the frequency corresponding to the ‘true’ period as found by Sesar+2010 (see Table 2 in [2]).  $\omega_{fit} = 2\pi/P_{fit}$  corresponds to the highest peak in the Lomb-Scargle periodogram evaluated on Sesar+2010 data - SDSS DR2.  $\omega_{true}/\omega_{fit}$  is approximately equal to 1, when the naive LS approach is able to recover the ‘true’ period. When this ratio is smaller or greater than 1, it means that the period recovered from the LS method is respectively shorter or longer than the ‘true’ period. This may be caused by the inherent simplicity of the simple single-term Fourier Series fitting. Indeed, some RR Lyrae light curves may have shapes that are insufficiently described by a single sinusoid (as on Fig.10.18 in [1]). See Figs. 39, 40 and 41 for details of evaluating LS on SDSS DR2 light curves.

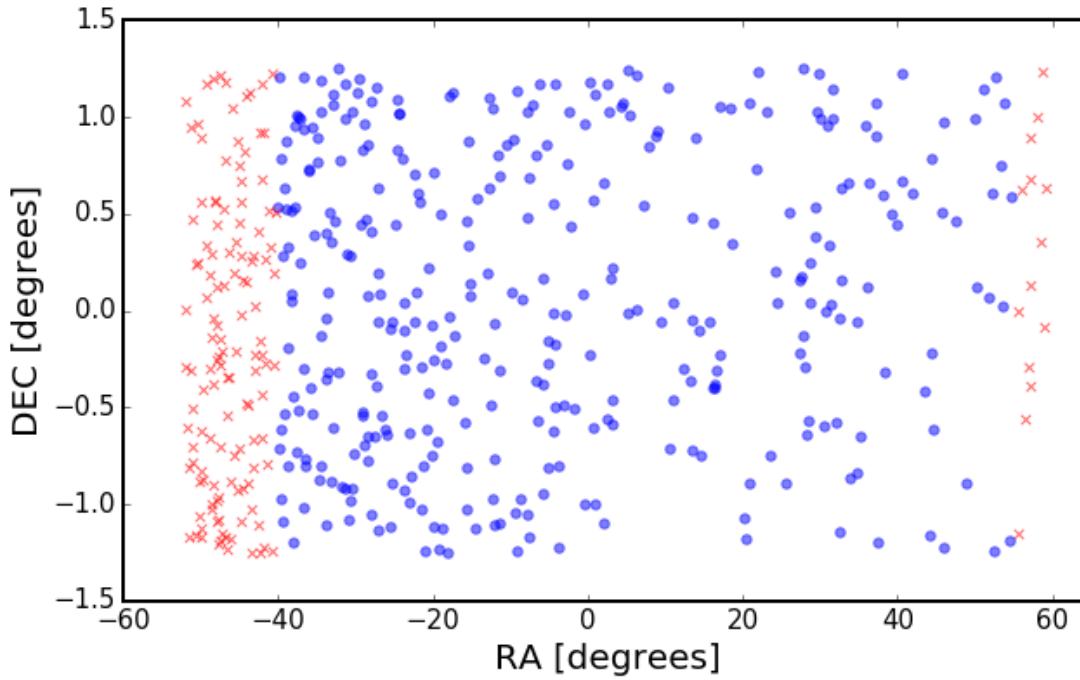


FIGURE 37: Results of positional query against 483 RR Lyrae stars from [2], using their RA, Dec. Blue dots are 343 stars that have a match in the PDAC S82 dataset within 2 arc-sec, and red crosses are 140 stars that did not. Increasing the search radius to 3 arcsec does not alter this result.

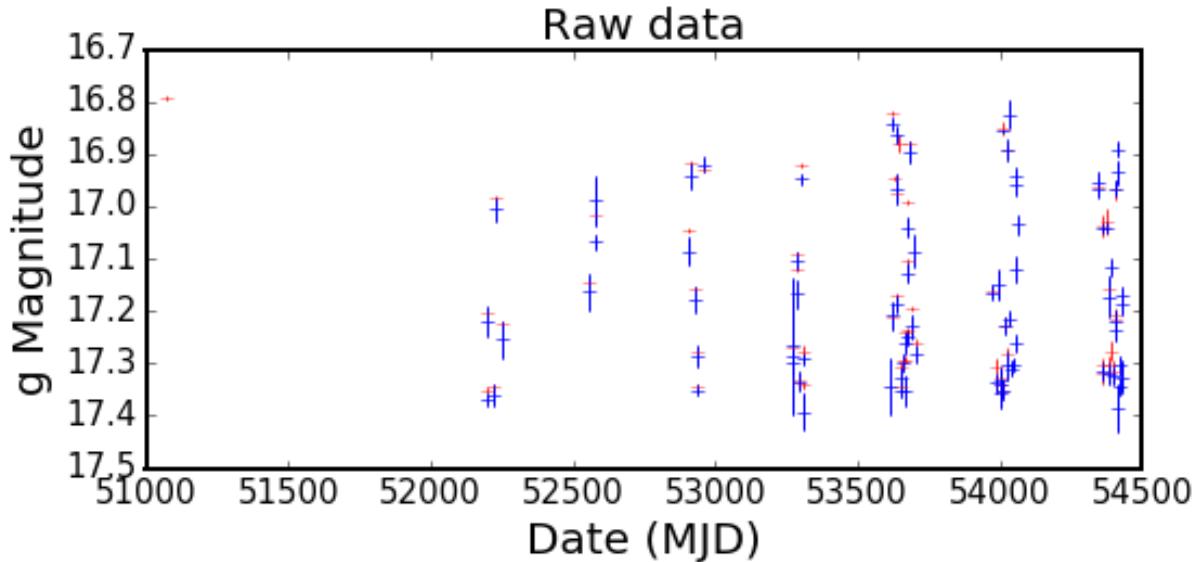


FIGURE 38: Comparison of RR Lyr ID=4099 from [2] (red crosses), and PDAC (blue crosses). The two light curves have different length : 59 vs 162 points, respectively.

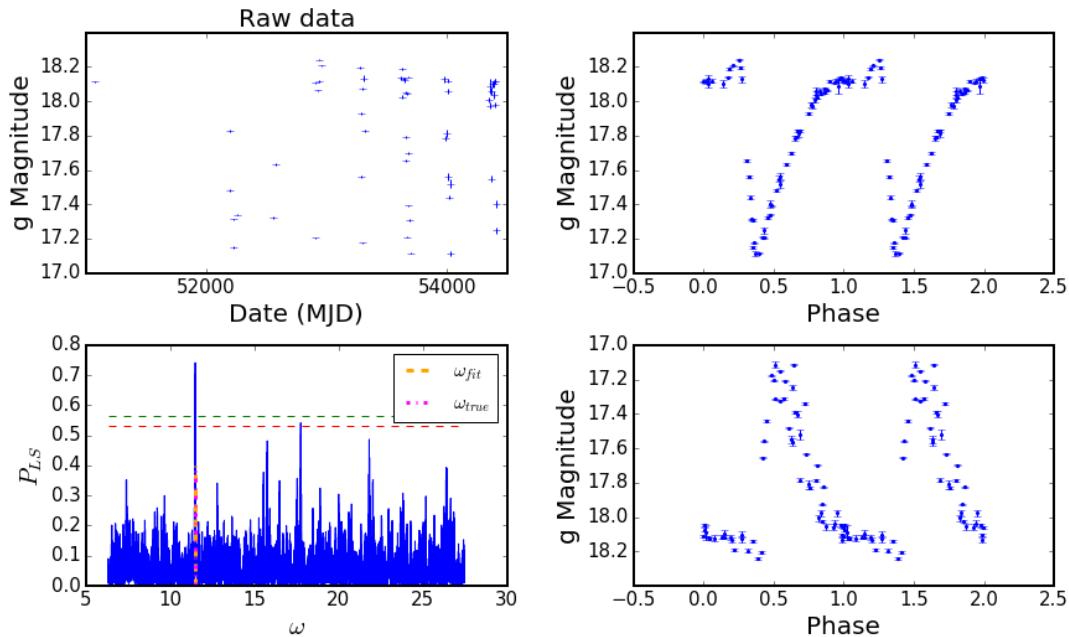


FIGURE 39: An example of the *astroML* Lomb Scargle periodogram performance, calculated for RR Lyr ID=13350 in SDSS g band (following Table 2 in [2]), using the SDSS data from [2]. It took 18.6 miliseconds on a laptop to calculate this periodogram. The upper left panel depicts the raw data. The upper right panel shows the phased light curve constructed with the 'true' period of 0.547987 days (' $P_{true}$ '). The lower left panel shows the Lomb Scargle periodogram, 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 light curve 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 ). The same object, but using the PDAC S82 S13 data is shown on Fig. 42

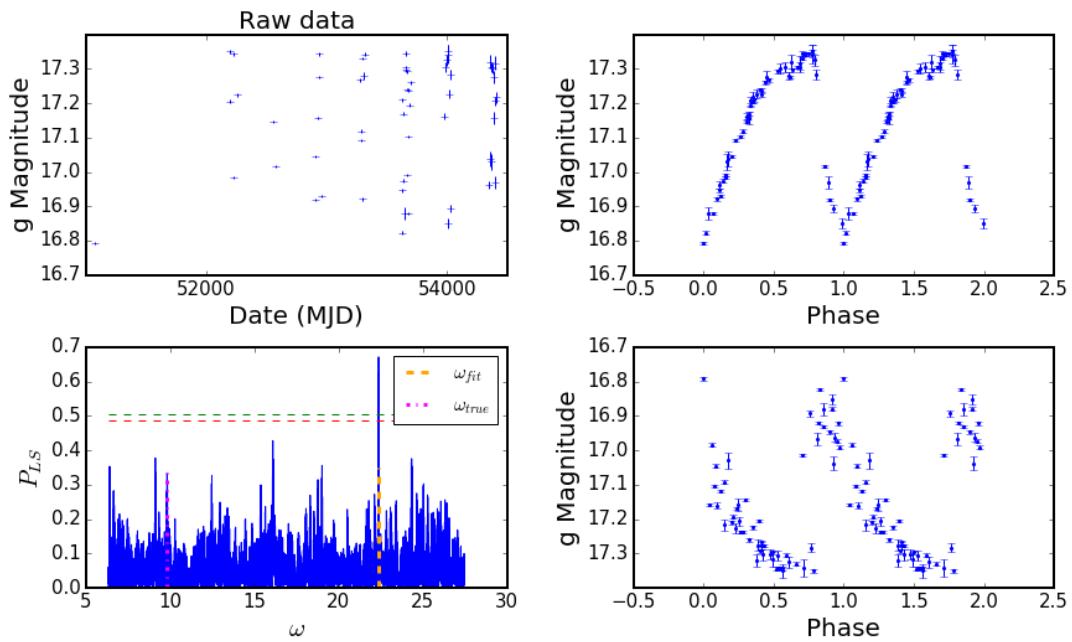


FIGURE 40: An example of a failure of naive single Lomb Scargle periodogram performance - the ratio of  $\omega_{true}/\omega_{fit} = 0.437$ . In these four panels we use SDSS DR2 data for RR Lyr ID =4099 from [2]. Upper-left panel: raw data. Upper-right panel : raw data folded on the true period. Bottom-left panel : Lomb-Scargle periodogram with significance levels. Bottom-right panel : the raw data folded on the LS period corresponding to the highest peak. The 'true' period from [2] is 0.641754 days, whereas the naive Lomb-Scargle periodogram approach yields the 'fit' period of 0.280827 days. Here  $\omega_{fit}$  and  $\omega_{true}$  significantly differ for this RR Lyr, and the 'true' frequency appears as only one of many insignificant periodogram peaks. We show the PDAC S82 S13 data used for this object on Fig. 43). Everything else as on Fig. 39.

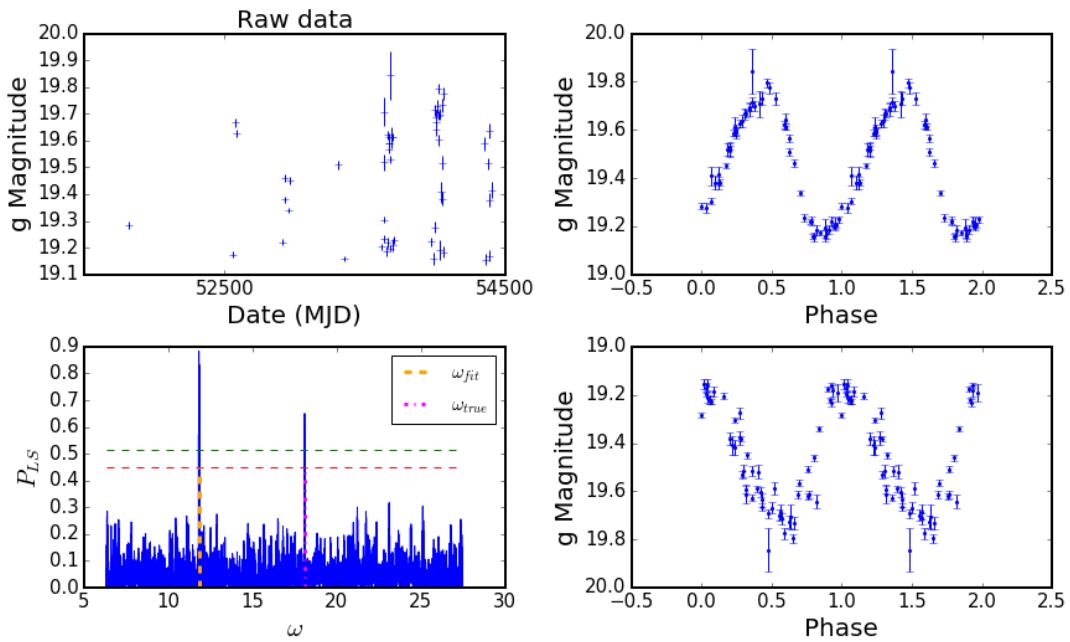


FIGURE 41: Same as Fig. 40, using the SDSS DR2 data from [2]. This RR Lyr ID=470994, has a cited period of 0.346794 days (' $P_{true}$ '), whereas the period derived from the Lomb-Scargle periodogram is 0.531667 days (' $P_{fit}$ '). Thus  $\omega_{true}/\omega_{fit} = 1.53$ . It may be a good example of aliasing.

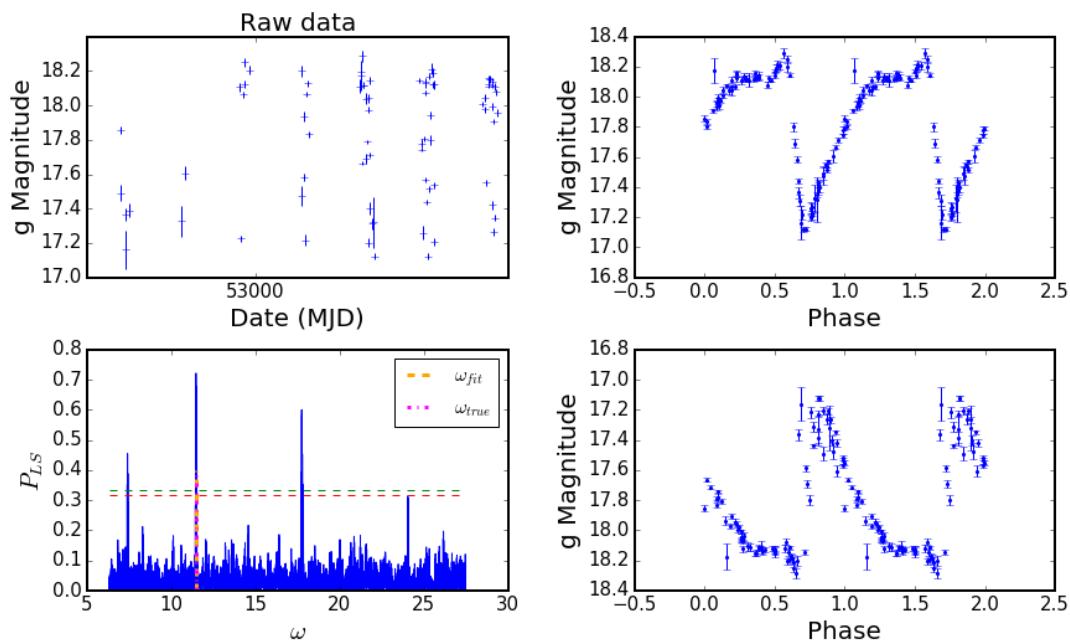


FIGURE 42: The same object as Fig. 39, but using data downloaded using PDAC. Using PDAC data, the RR Lyr ID=13350 has a best-fit period of 0.547969 days, almost identical to true period of 0.547987 from [2]. Panels the same as on Fig. 40

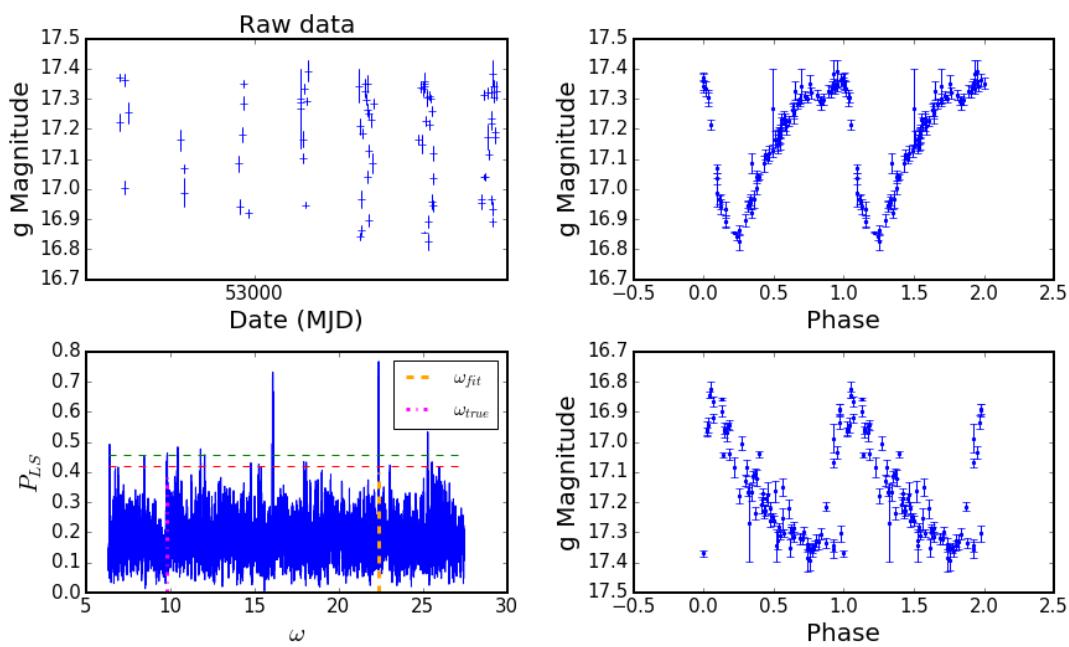


FIGURE 43: The same object as Fig. 40, but using data downloaded using PDAC. Calculating a naive LS periodogram using PDAC data for RR Lyr ID=4099 we find the best-fit period (frequency with highest power) of 0.280827 days, almost identical to the period found using LS periodogram on the SDSS [2] data 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. 39

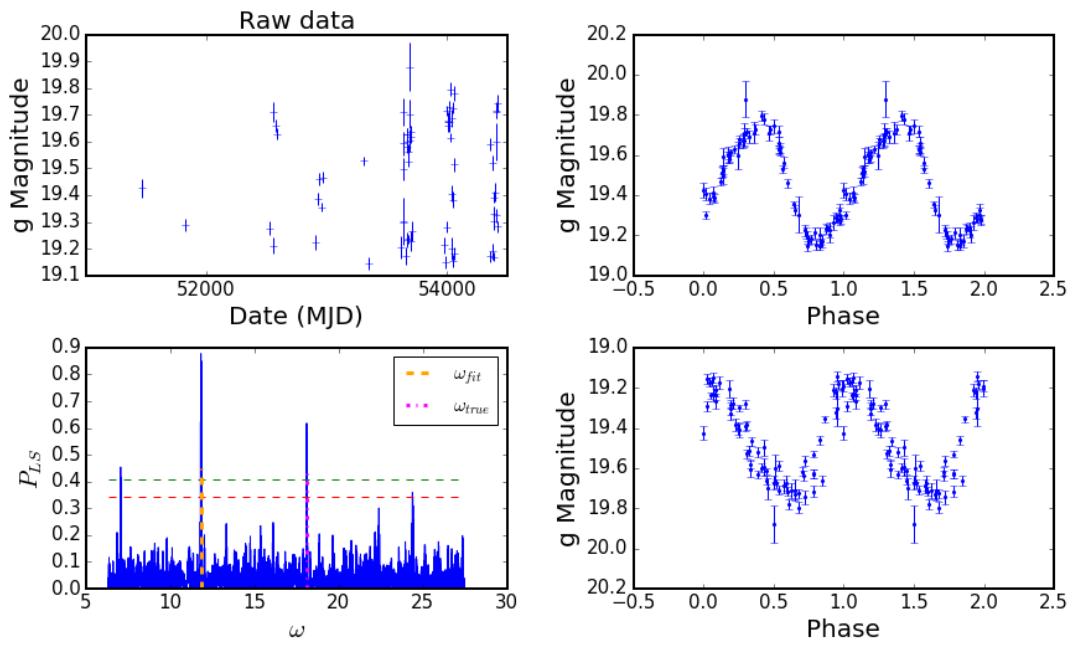


FIGURE 44: The same object as Fig. 41, but using data downloaded from PDAC. Calculating a naive LS periodogram using PDAC data for RR Lyr ID=470994 we find the best-fit period (frequency with highest power) of 0.531667 days, almost twice as high as the ‘true’ period of 0.346794 days from [2]. For this 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. 39

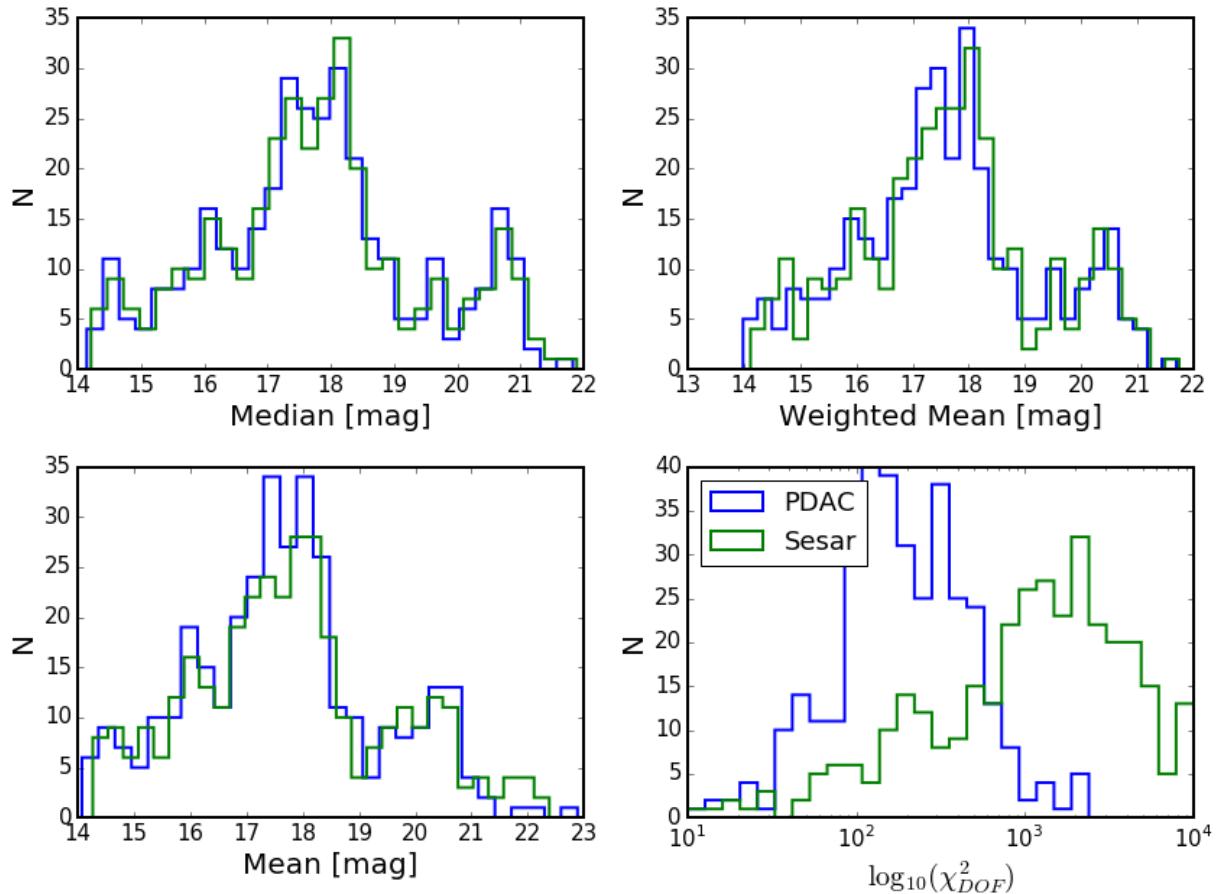


FIGURE 45: Comparison of the original [2] light curves (green) against data for the same objects pulled from PDAC (blue). For each of the 383 light curves in SDSS  $g$ -band, without any pre-processing or clipping, we calculated the median, weighted mean, mean, and  $\chi^2_{DOF}$ .

## 6 Conclusions

The Preliminary Data Access Center readily implements many requirements for SUIT, as outlined in the LDM-130 document. We find that the overall structure of the interface is very intuitive, building on the strengths of NASA/IPAC IRSA and SDSS CasJobs. Multiple ways of defining the query region, and resolving names of objects with NED helps quickly access the data, and adding SQL constraints makes it possible to reduce the amount of displayed information. An innovative Time Series View, building on the strengths of the NASA Exoplanet Archive Periodogram Tool has a great potential to aid quickly analysis of time series data. Easy access to external image catalogs allow overlaying information across a range of wavelengths, including a very well toolbox to make an RGB image. The QSERV responds to complicated SQL queries, and scisql functions help to execute very quick joins between catalogs to obtain needed information.

There is a number of areas with possible improvements, including aspects of the User Interface, as well as the SQL query engine. First, certain information, such as calibrated fluxes or magnitudes is only available via Time Series view or a direct SQL query, while a user may be interested in accessing this information more directly. Furthermore, allowing query by objectId, without specifying the coordinates, would make it easier to find information for a particular object. Search capabilities will become more robust once AllSky and Multiple object queries are supported. It will be also very time efficient if the user is able to select limiting magnitude when querying either catalog, to avoid many faint, and not dependable, measurements. We also recommend expanding the educational aspect of the user interface : providing more explanatory metadata about column content, relating the available data to the image processing pipeline, and explaining the details of algorithms used in calculation of calibrated fluxes and magnitudes, as well as Lomb-Scargle periodogram (like [5]). On the level of the QSERV engine, errors currently provided experience a degradation of the error message quality, and it would be very beneficial to provide a meaningful error to the user interacting with PDAC. This would make it easier to debug the SQL queries, and streamline the data access.

This report outlined a system that is in active development. Current improvements in PDAC based on the user feedback are captured in the epic DM-10432. JIRA tickets specifically resulting from this work include DM-7990 , DM-10477, DM-10431, DM-10433, DM-10463 , DM-10465, DM-10466.

## A Appendix : SQL queries

### A.1 Single source cone query, calibrated flux

```
curl -o calib_lightcurve.json -d 'query= \
SELECT \
    id, fsrcc.exposure_time_mid , \
    scisql_dnToFlux(fsrcc.flux_psf, exp.fluxMag0) AS psfFlux, \
    scisql_dnToFluxSigma(fsrcc.flux_psf, fsrcc.flux_psf_err, \
        exp.fluxMag0, exp.fluxMag0Sigma) AS psfFluxErr \
FROM \
    RunDeepForcedSource AS fsrcc, \
    Science_Ccd_Exposure AS exp \
WHERE \
    exp.scienceCcdExposureId = fsrcc.exposure_id \
    AND fsrcc.exposure_filter_id = 1 \
    AND objectId = 216172782516437336 \
    ORDER BY exposure_time_mid' \
http://lsst-qserv-dax01.ncsa.illinois.edu:5000/db/v0/tap-sync
```

### A.2 Single source cone query, calibrated magnitude

```
curl -o gmag_lightcurve.json -d 'query=\
SELECT \
    objectId, fsrcc.exposure_time_mid AS mjd, \
    scisql_dnToAbMag(fsrcc.flux_psf,exp.fluxMag0) AS g, \
    scisql_dnToAbMagSigma(fsrcc.flux_psf, fsrcc.flux_psf_err, \
        exp.fluxMag0, exp.fluxMag0Sigma) AS gErr \
FROM \
    RunDeepForcedSource AS fsrcc, \
    Science_Ccd_Exposure AS exp \
WHERE \
    exp.scienceCcdExposureId = fsrcc.exposure_id \
    AND fsrcc.exposure_filter_id=1 \
    AND objectId IN (216172782516437336)' \
http://lsst-qserv-dax01.ncsa.illinois.edu:5000/db/v0/tap-sync
```

### A.3 Box query, i<23.5

```
curl -o deepSourceBox.json -d 'query= \
SELECT \
dsrc.parent AS parent, dsrc.id AS id, \
scisql_dnToAbMag(dsrc.flux_psf,coadd.fluxMag0) AS i, \
scisql_dnToAbMagSigma(dsrc.flux_psf, dsrc.flux_psf_err, \
coadd.fluxMag0, coadd.fluxMag0Sigma) AS iErr \
FROM \
RunDeepSource AS dsrc, \
DeepCoadd AS coadd \
WHERE \
qserv_areaspec_box(325,-0.7, 325.1, -0.6)\
AND dsrc.coadd_id = coadd.deepCoaddId \
AND dsrc.coadd_filter_id = 3 \
HAVING i < 23.5' \
http://lsst-qserv-dax01.ncsa.illinois.edu:5000/db/v0/tap-sync
```

## References

- [1] Ivezić, Ž., Connolly, A. J., VanderPlas, J. T., & Gray, A. 2014, Statistics, Data Mining, and Machine Learning in Astronomy
- [2] Sesar, B., et al. 2010, ApJ, 708, 717
- [3] Vanderplas, J., Connolly, A., Ivezić, Ž., & Gray, A. 2012, in Conference on Intelligent Data Understanding (CIDU), 47
- [4] VanderPlas, J., & Ivezić, Ž. 2015, The Astrophysical Journal, 812, 18
- [5] VanderPlas, J. T. 2017, ArXiv e-prints
- [6] Zechmeister, M., & Kürster, M. 2009, A&A, 496, 577