Anomaly detection in ZTF DR3 arXiv:2012.01419

Konstantin Malanchev (UIUC, SAI MSU)

on the behalf of the SИAD+ team:

Maria Pruzhinskaya, Vladimir Korolev, Patrick Aleo, Matwey Kornilov, Emille Ishida, Vadim Krushinsky, Florian Mondon, Sreevarsha Sreejith, Alina Volnova, Aleksander Belinski, Aleksander Dodin, Andrey Tatarnikov, Sergey Zheltoukhov

DESC SN subgroup meeting, 2020.12.15

SNAD

International research group for anomaly detection

2018, Moscow

2020, Zoom

2019, Clermont-Ferrand

http://snad.space

Methodology

Data

Zwicky Transient Facility Data Release 3:

- ~3 billion light curves in *gri* in 1020 overlapping fields, ~$2 \cdot 10^{11}$ detections
- If an object was observed in p passbands and f fields it will be represented as $p \cdot f$ individual light curves with different object identifiers (OIDs)

Our data set:

- The first 9 months of the survey due to better cadence
- 3 fields and *zr* passband only, ~2.25 million light curves total

Field name	ZTF field	Centre (α, δ)	Centre (l, b)	FOV	N	Object count
M 31	Field 695, ccdid 11	$10\overset{\circ}{.}72, 41\overset{\circ}{.}34$	$121\overset{\circ}{.}20, -21\overset{\circ}{.}51$	3 deg^2	≥ 100	57 546
DEEP	Field 795	$246\overset{\circ}{.}72, 54\overset{\circ}{.}88$	$83\overset{\circ}{.}91, 42\overset{\circ}{.}45$	55 deg^2	≥ 100	406 611
DISK	Field 807	$19\overset{\circ}{.}99, 62\overset{\circ}{.}05$	$126\overset{\circ}{.}27, -0\overset{\circ}{.}63$	55 deg^2	≥ 500	1 790 565

Pre-processing

Feature extraction + normalisation

- 42 light-curve features, e. g. Lomb-Scargle Periogoram-based features, magnitude statistics, irregular time-series features like η^e or Stetson K
- New feature extraction code written on Rust: https://github.com/hombit/light-curve

  Has Python pip-installable binding `light-curve-python`, which is an order magnitude faster than popular Python codes like Feets.

  Currently it is a part of the ANTARES broker pipeline for each new ZTF alert
- Standard scaling: removing the mean and scaling to unit variance

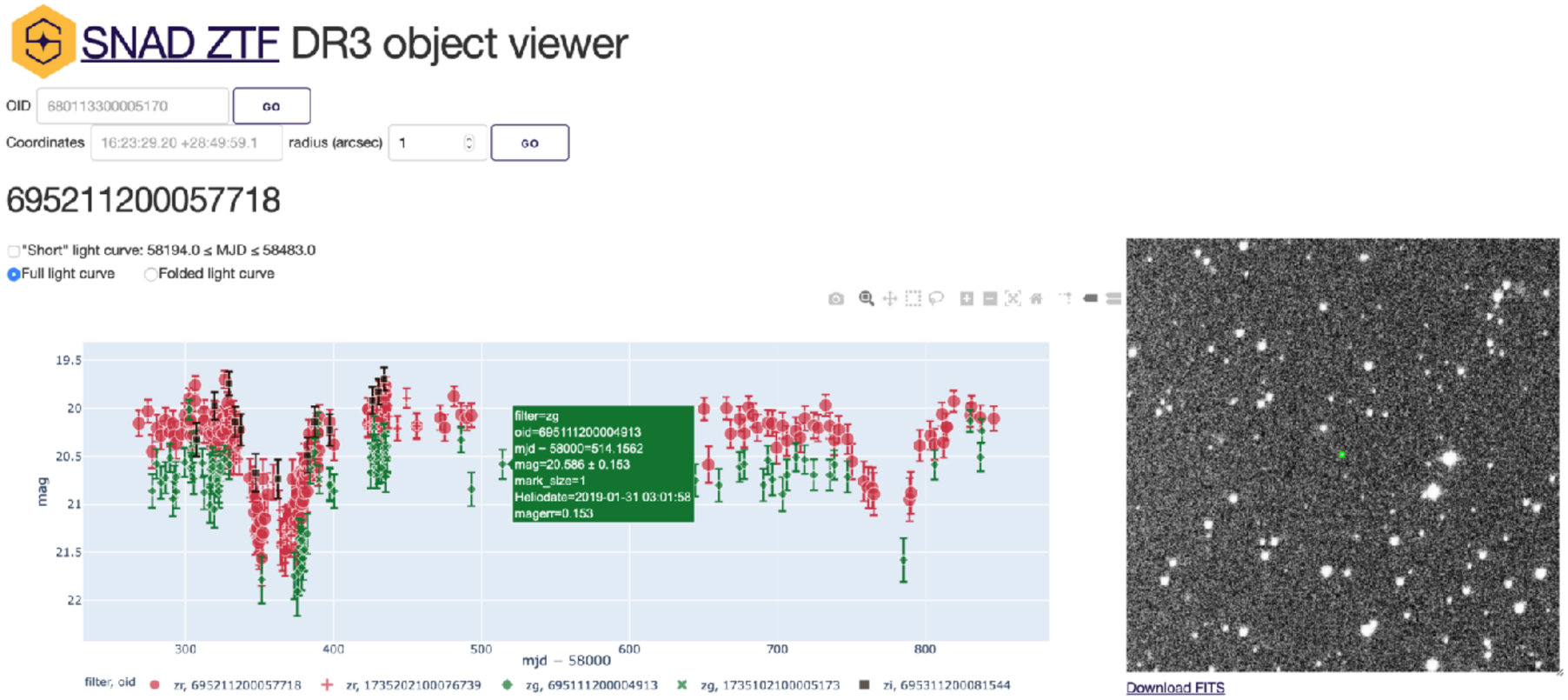
Anomaly Detection Pipeline

https://github.com/snad-space/zwad

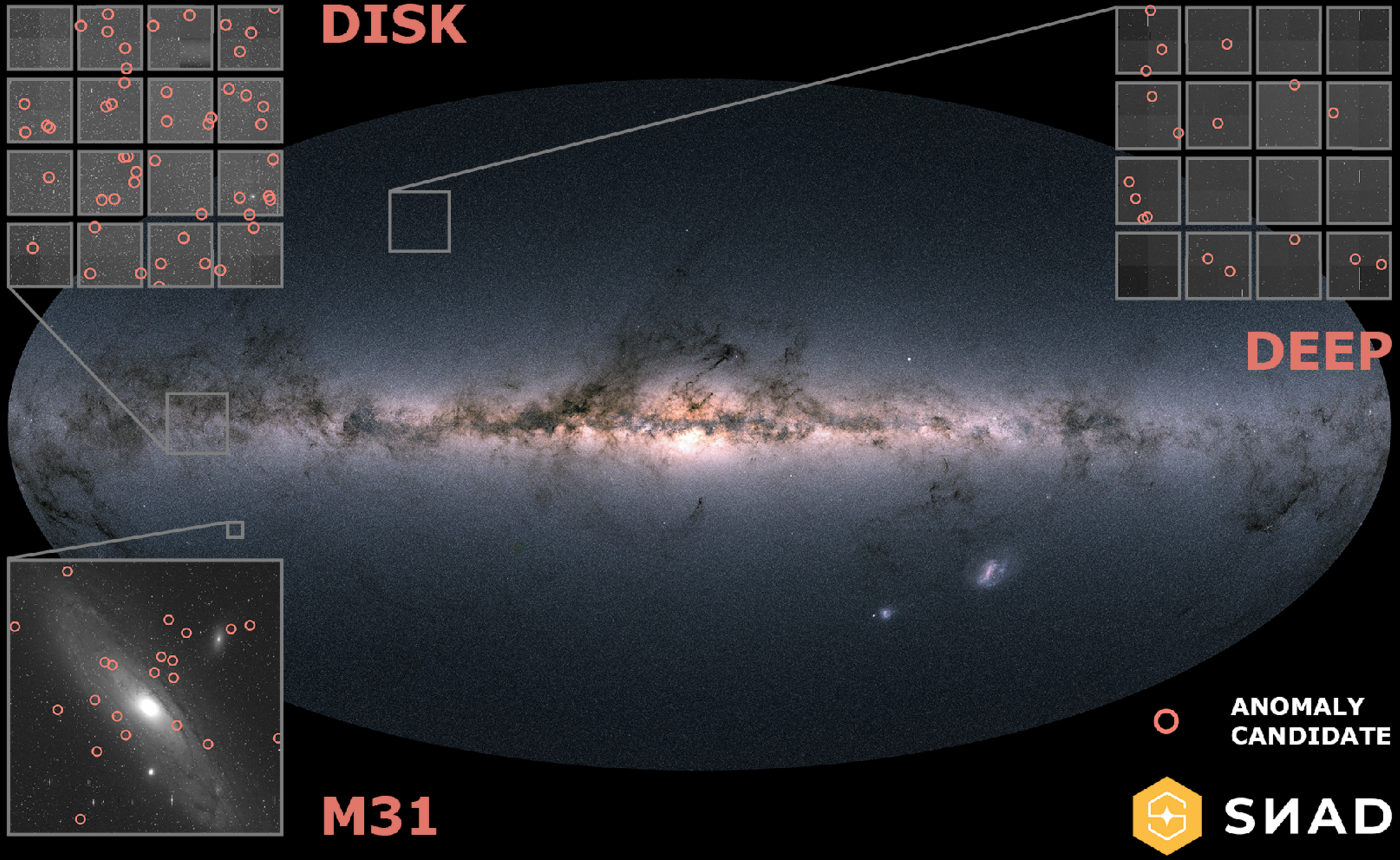
- Def. *Outlier* is an object located in a sparse region of the feature space
- Def. *Anomaly* is an astrophysical source having unusual properties for its class or a representative of some rare class
- Four outlier detection algorithms: Isolation Forest, Gaussian Mixture Model, Local Outlier Factor, One-class Support Vector Machine
- Each field data are a separate data set
- Each algorithms is asked for 40 objects with the greatest outlier score for each field.
- Total 277 objects are examined using catalogs, literature, simulations, photometry redoing, and additional photometric and spectroscopic observations

SNAD ZTF web-viewer

https://ztf.snad.space



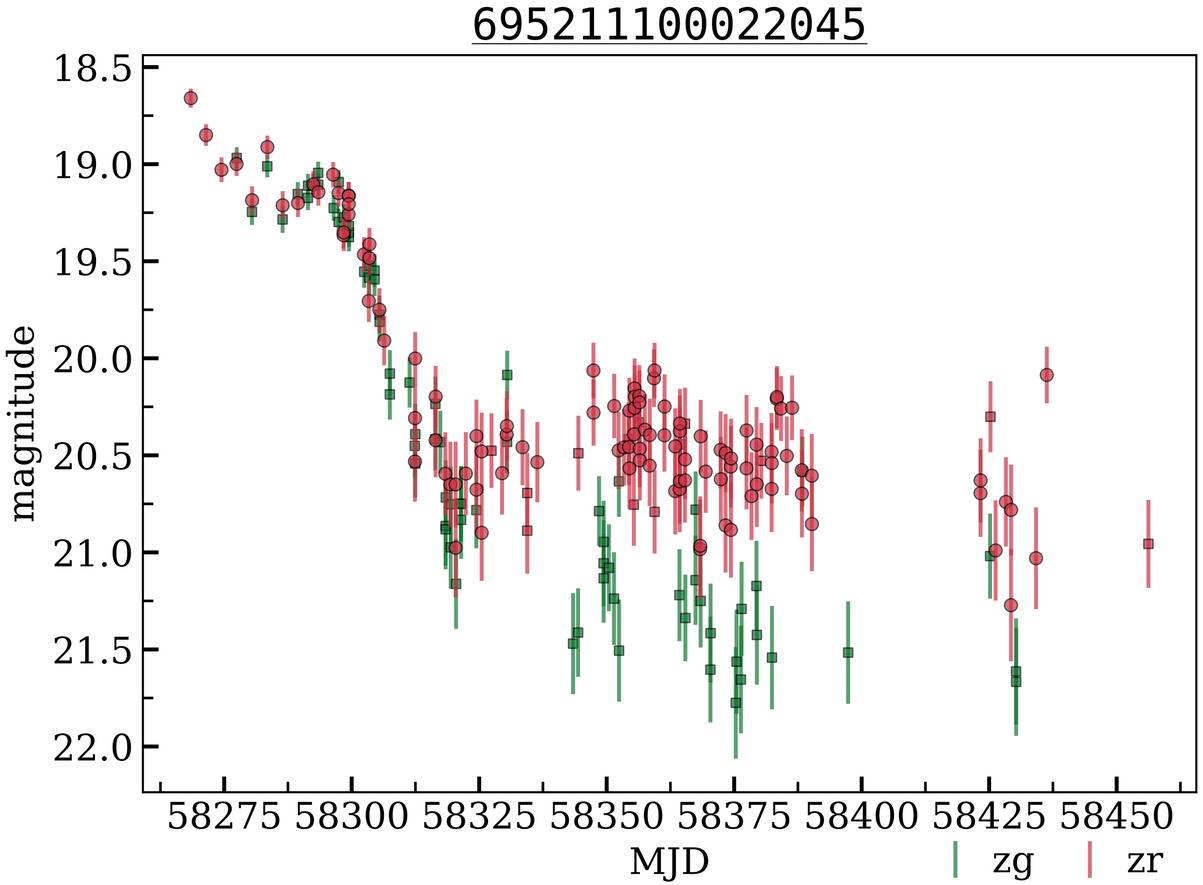
Anomaly Candidates — 89 / 277



Recurrent Nova Candidate

AT 2017ixs from M31 field, it had had two outbursts in 2017, before ZTF

- Red circles are the considering by the anomaly detection pipeline light curve
- Squares are other light curves of the same source but in different fields and/or passbands

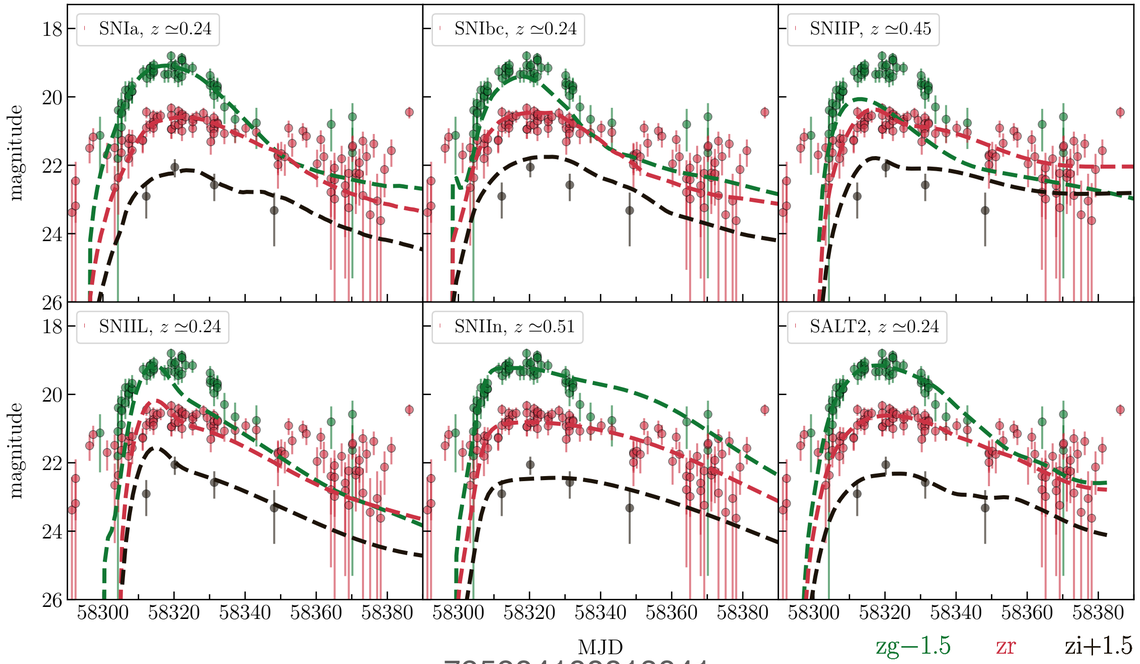


Supernova Candidates

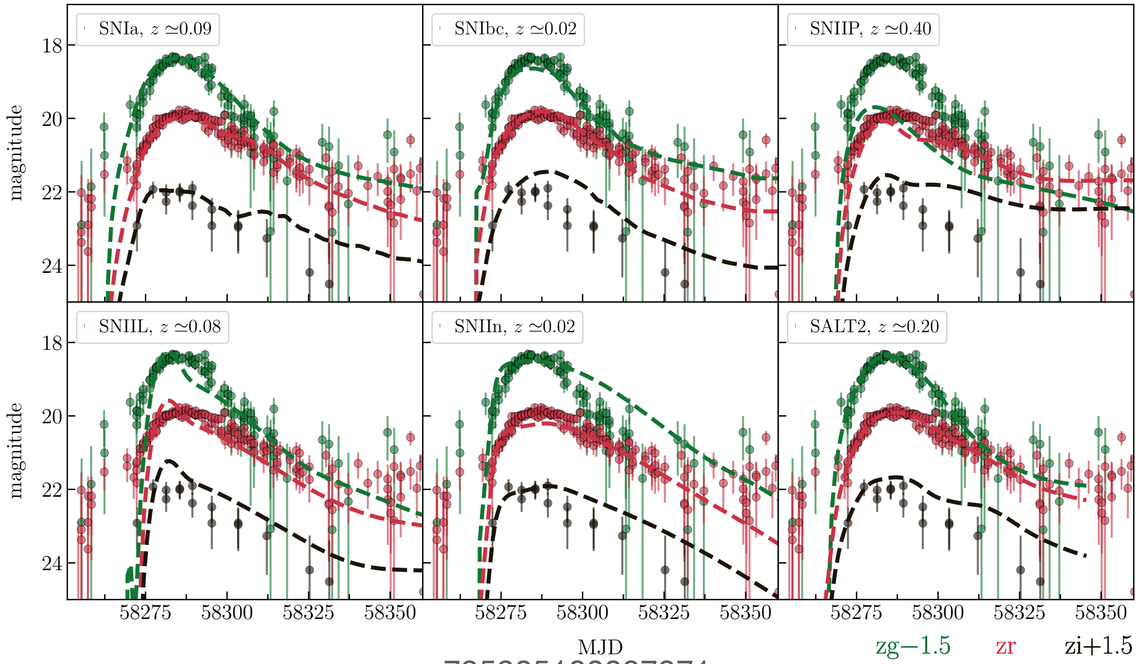
Six candidates from DEEP field (400 000 objects), only one is in TNS

Table 2. Results of the light curve fit with the SALT2 model for supernova candidates from the DEEP field.

OID	Host galaxy*	z_{ph}	z	t_0	x_1	c	Comments†
795202100005941/ZTF18aanbnjh	SDSS J163437.92+521642.2	0.424 ± 0.103	—	—	—	—	Blazar
795204100013041/ZTF18abgvctp	SDSS J160913.83+521251.3	0.375 ± 0.138	~0.24	58320.9336±0.4389	1.71±0.51	−0.044±0.035	—
795205100007271/ZTF18aayatjf	—	—	~0.20	58285.8334±0.1810	−0.54±0.18	−0.075±0.021	SN Ia
795209200003484/ZTF18abbpebf	—	—	~0.11	58299.7269±0.0008	0.60±0.12	−0.013±0.012	SN Ia
795212100007964/ZTF18aanbksg	SDSS J161144.90+555740.7	0.288 ± 0.122	~0.18	58214.4470±0.0002	0.40±0.20	−0.282±0.020	Blazar
795213200000671/ZTF18aaincjv	—	—	—	—	—	—	AGN-I



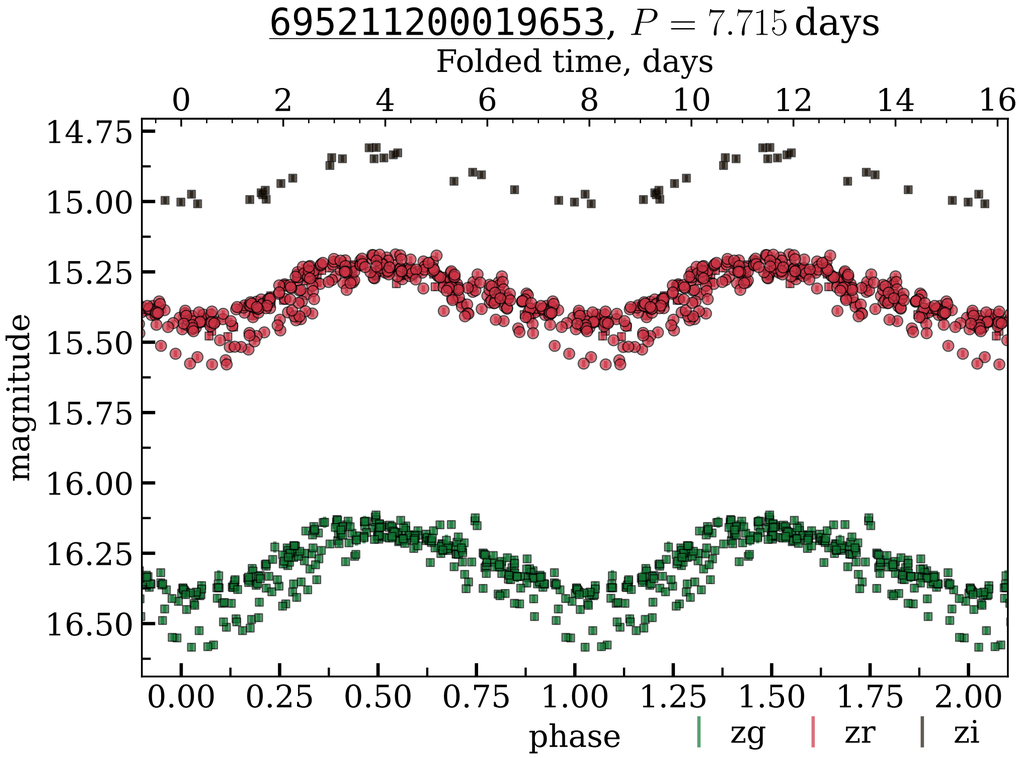
795204100013041

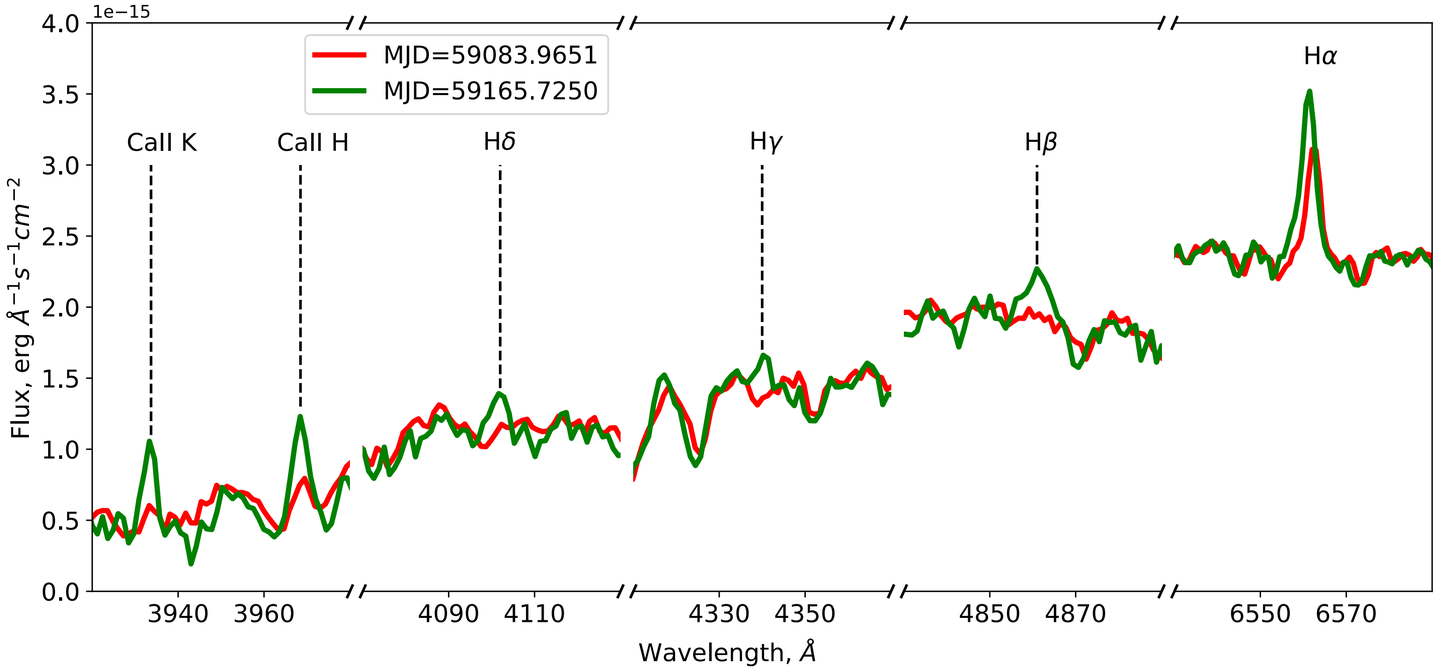


795205100007271

Classification of RS CVn (binary w/ spots)

Our spectra + period change + flare activity

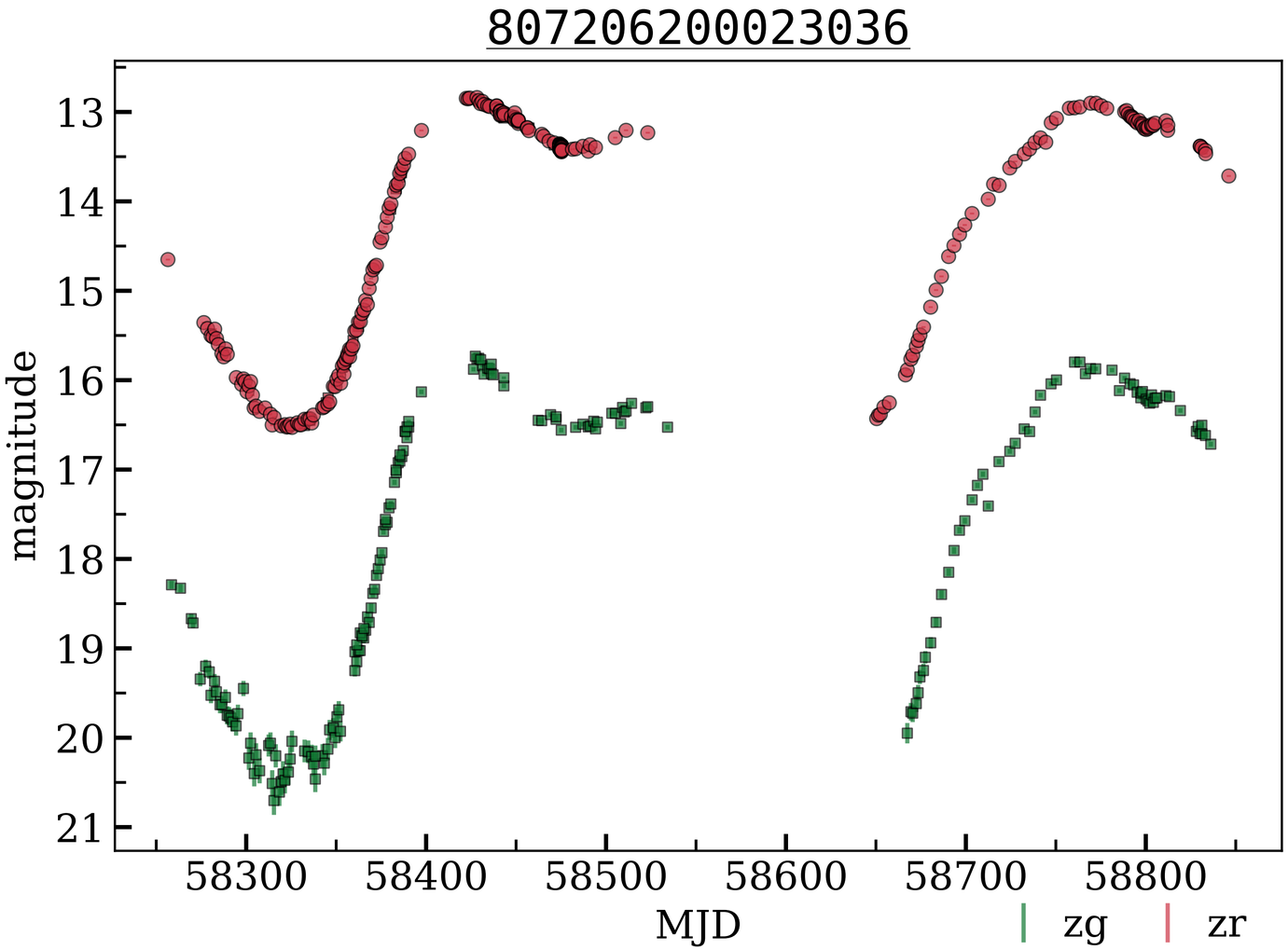




2.5m CMO SAI MSU spectra of the object at different phases of the orbital cycle

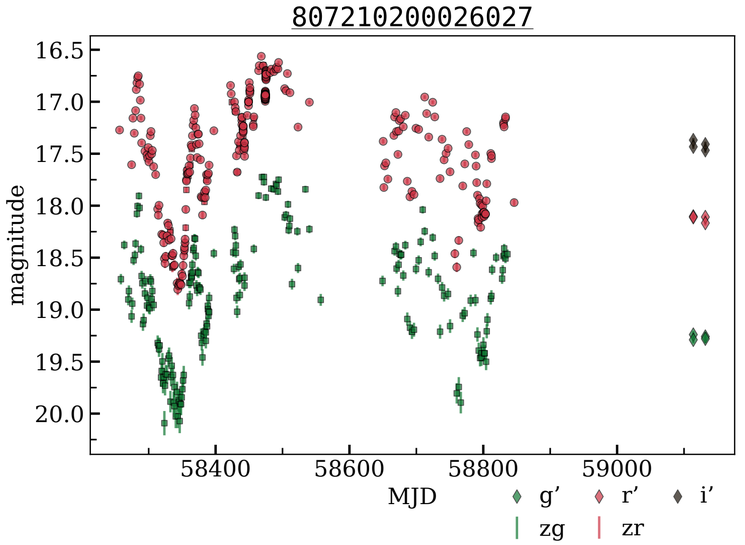
Mira Binary Candidate

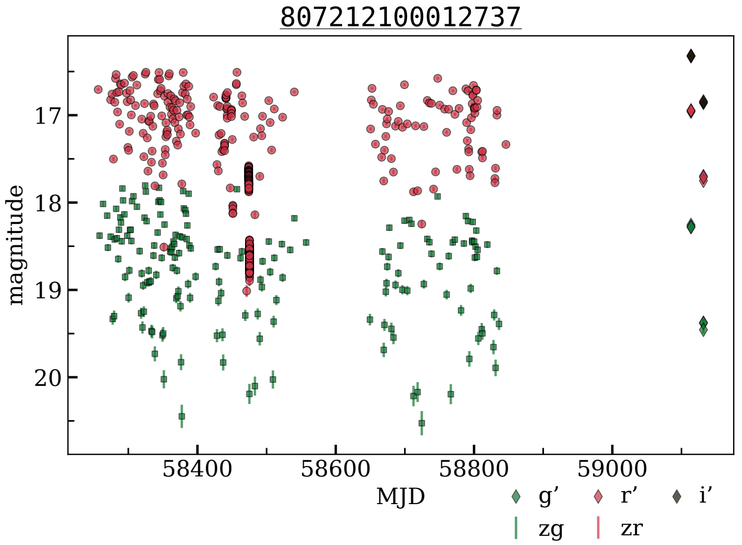
Light curve may indicate the presence of a companion

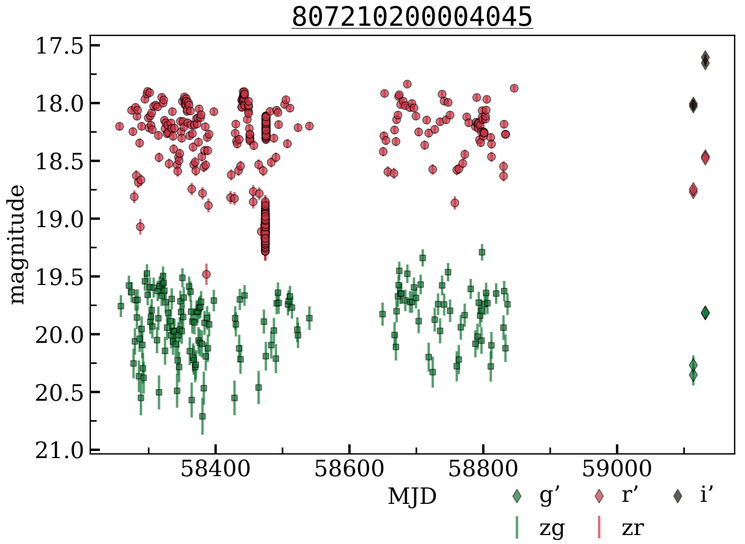


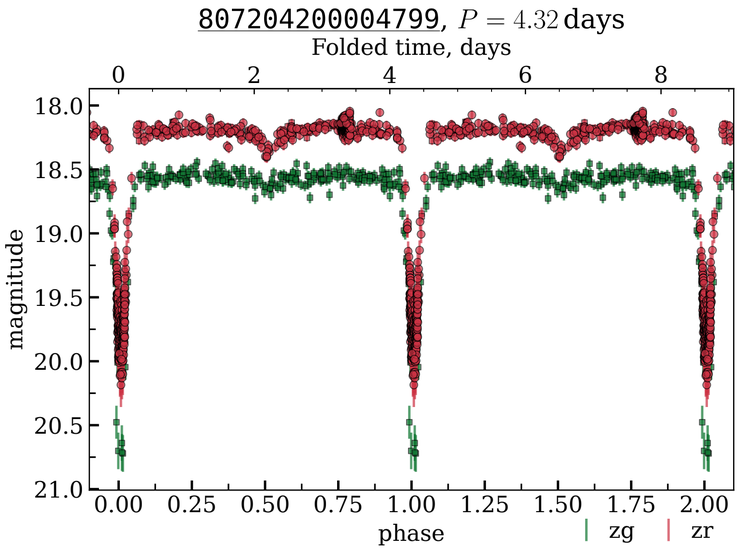
Non-catalogued Sources — 23 / 277

- Diamonds are our observations (the right group of observations)
- Red circles are the considering by the anomaly detection pipeline light curve
- Squares are other light curves of the same source









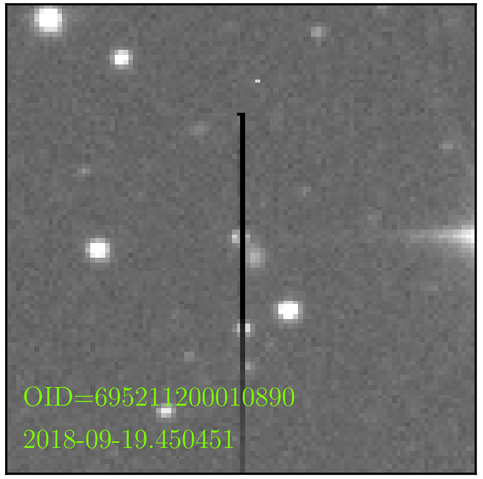
Bogus Light Curves — 188 / 277

Bogus Detections

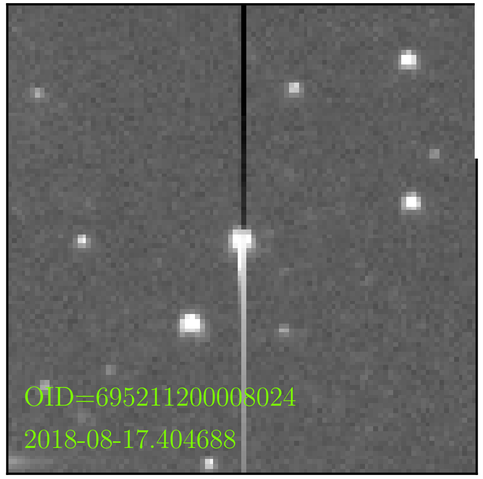
The objects are in the frame centers

Satellite or Plane Track

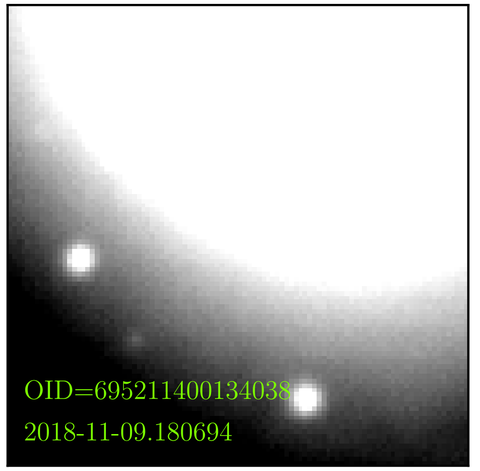
Bad Column



Bad Column

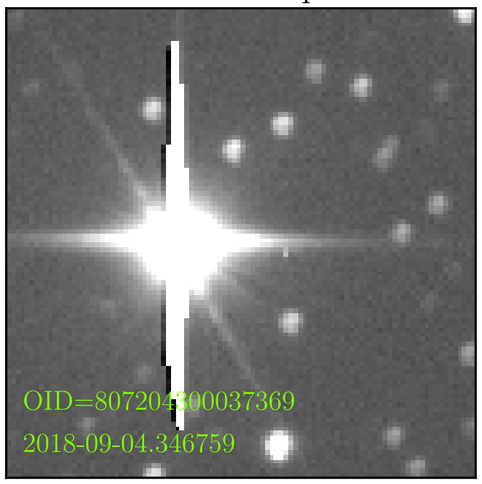


Close to M31 Centre

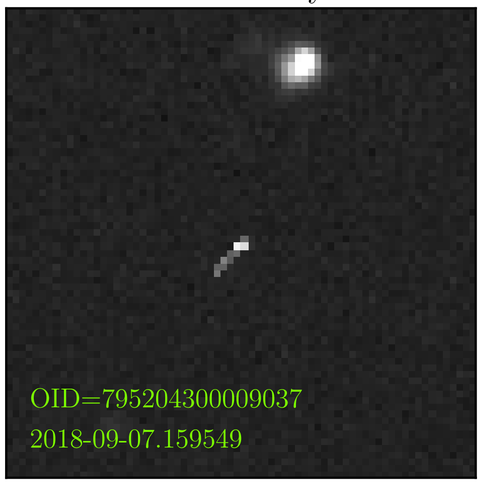


Ghost

Diffraction Spike



Cosmic Ray

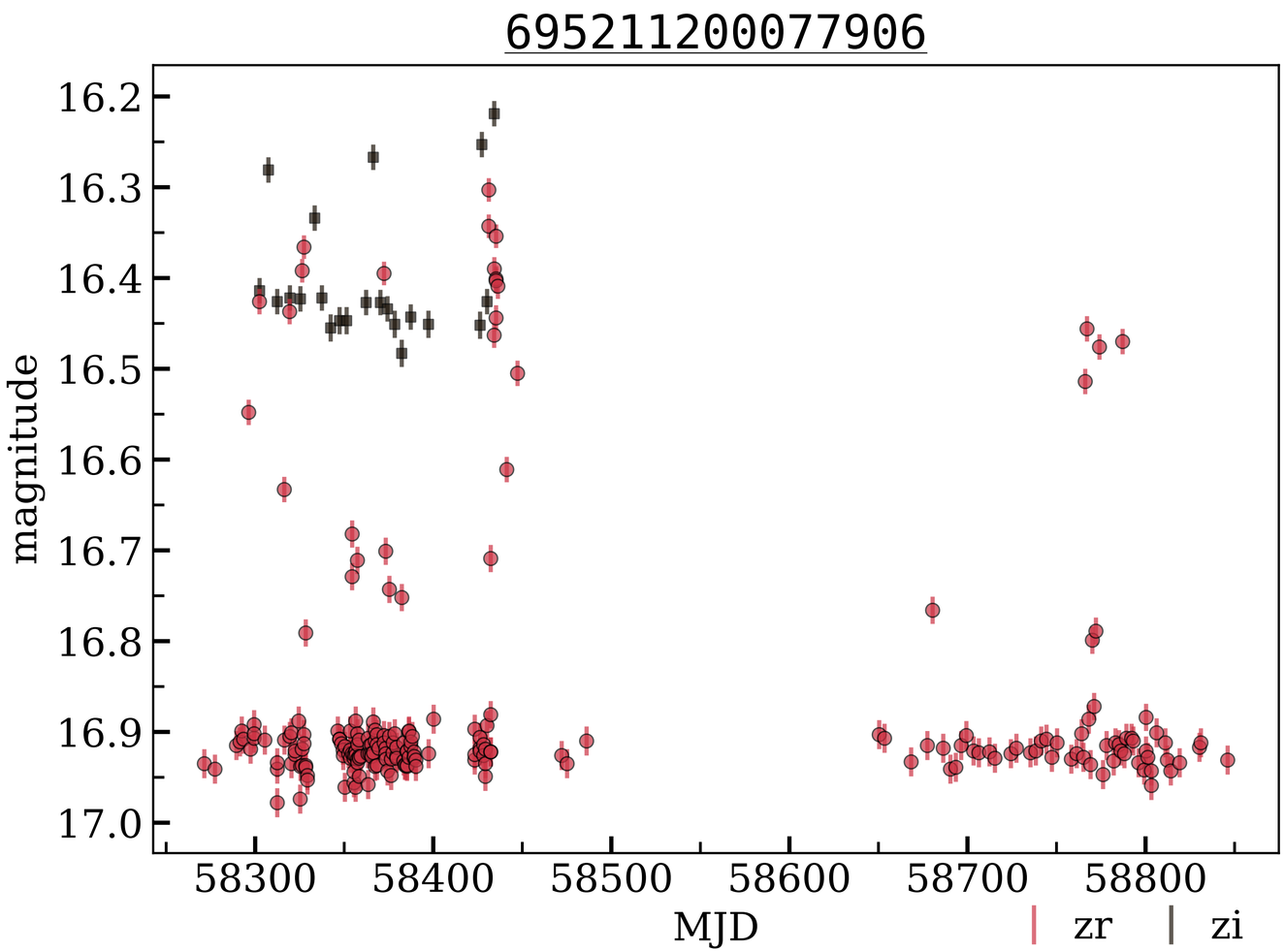


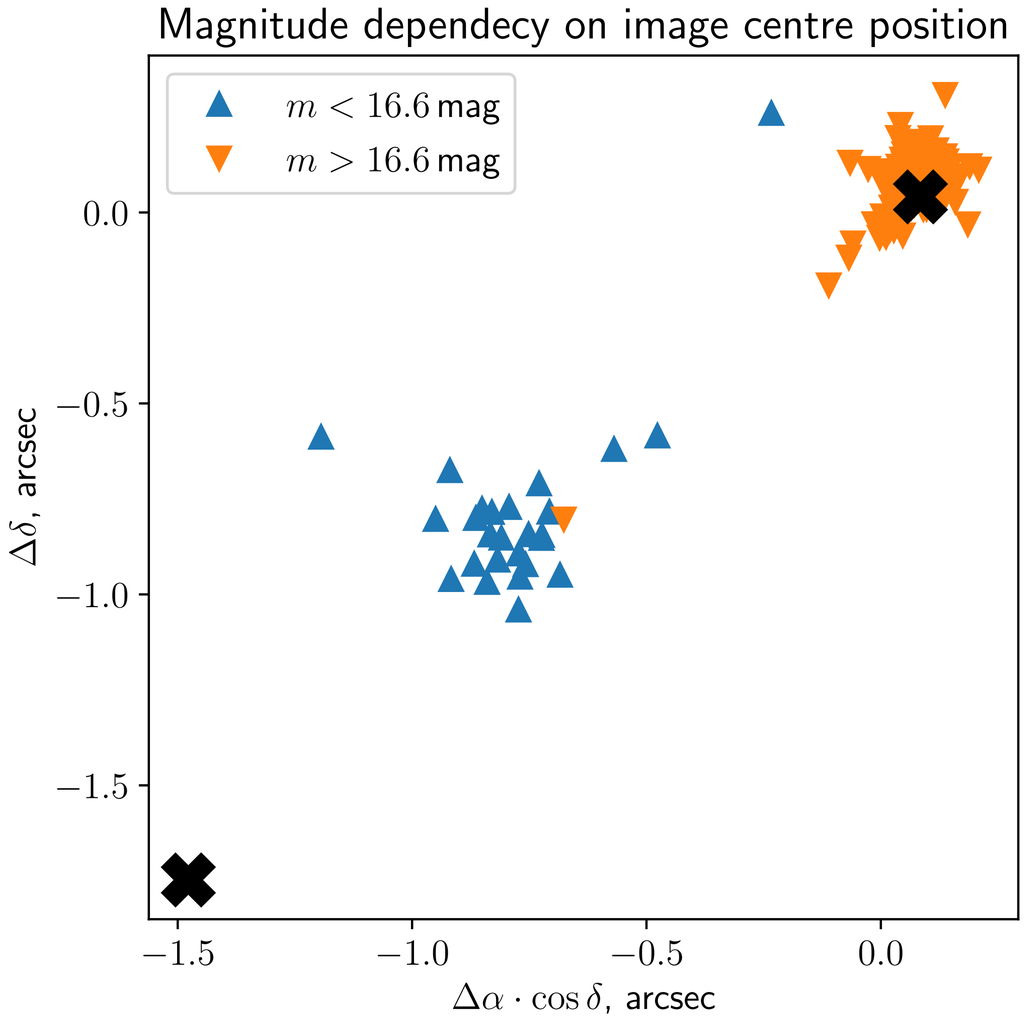
Defocusing

Ghost

Double Stars Defocusing

If a separation is about 2” (typical FWHM) then defocusing can cause the false variability

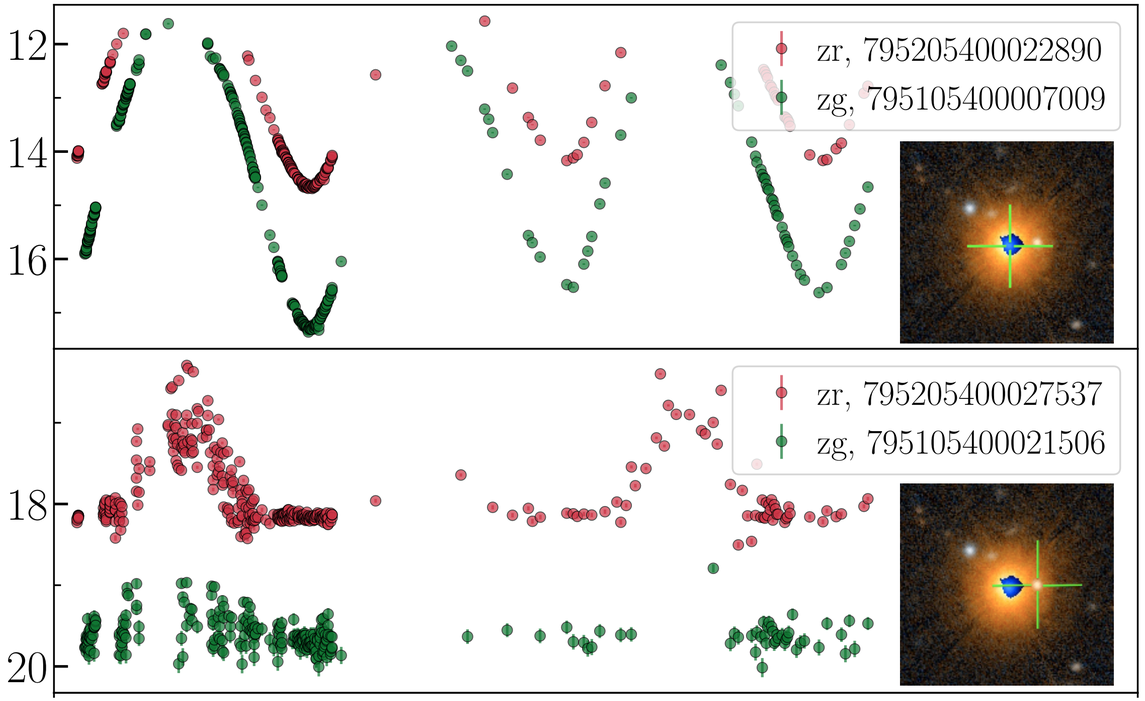




IW Dra "Echos"

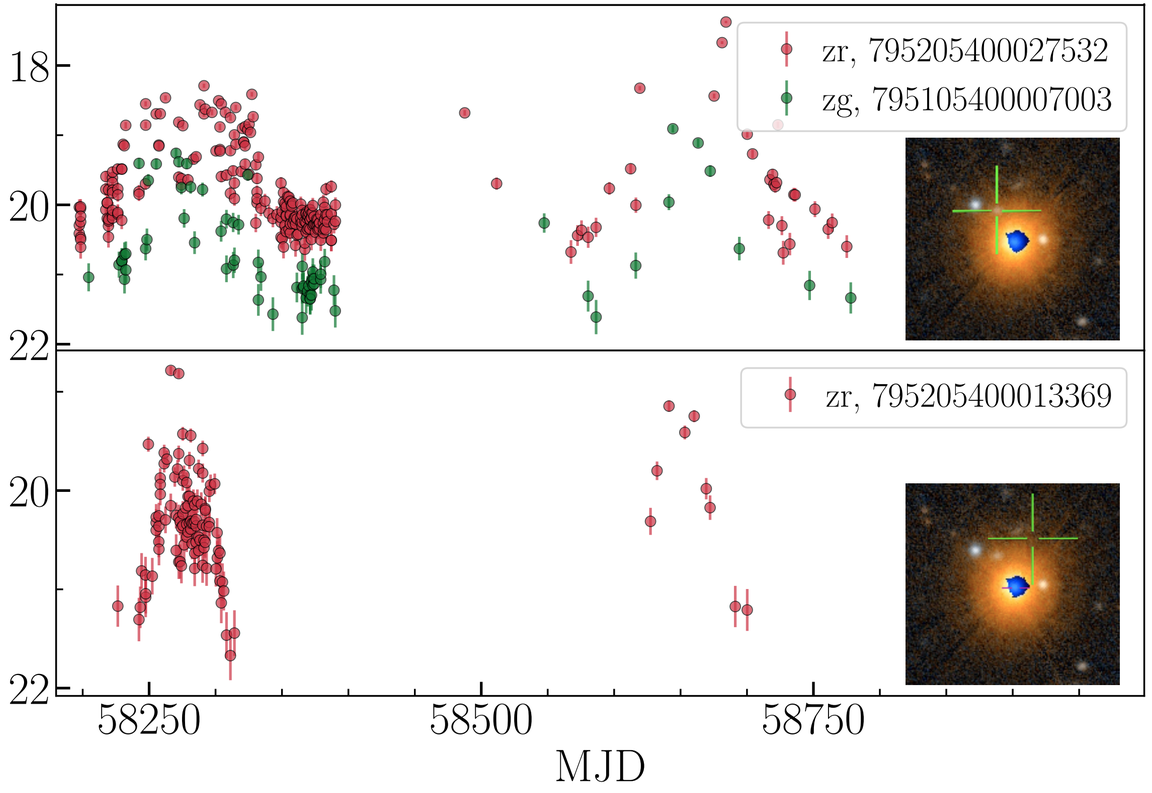
All four objects are found by the outlier detection pipeline

Mira-like IW Dra



Echo 1

Echo 2

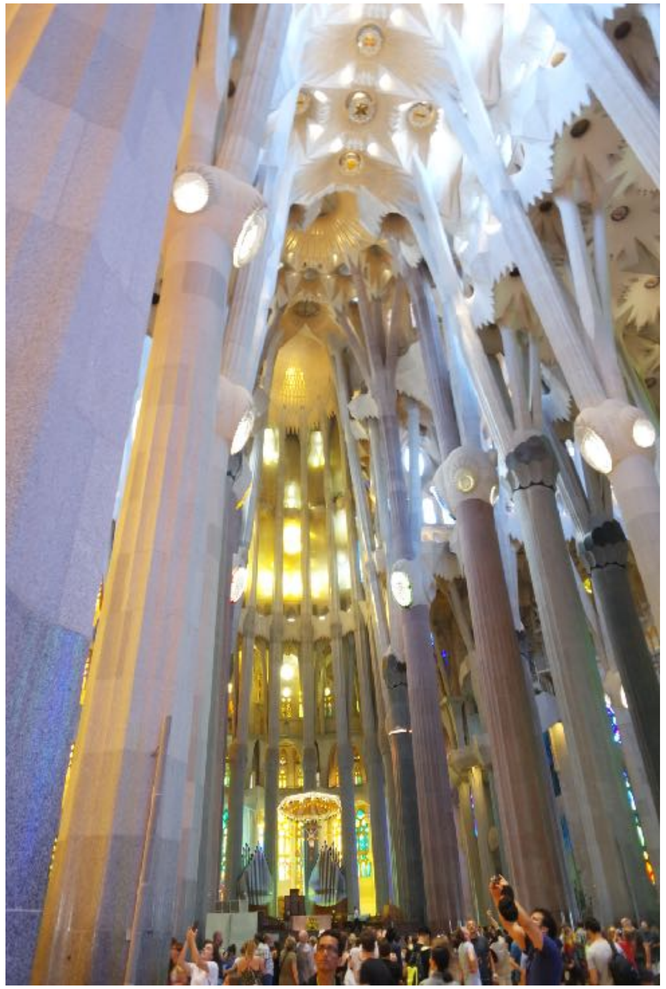


Echo 3

Barcelona

I bet you’ve never been to Barcelona

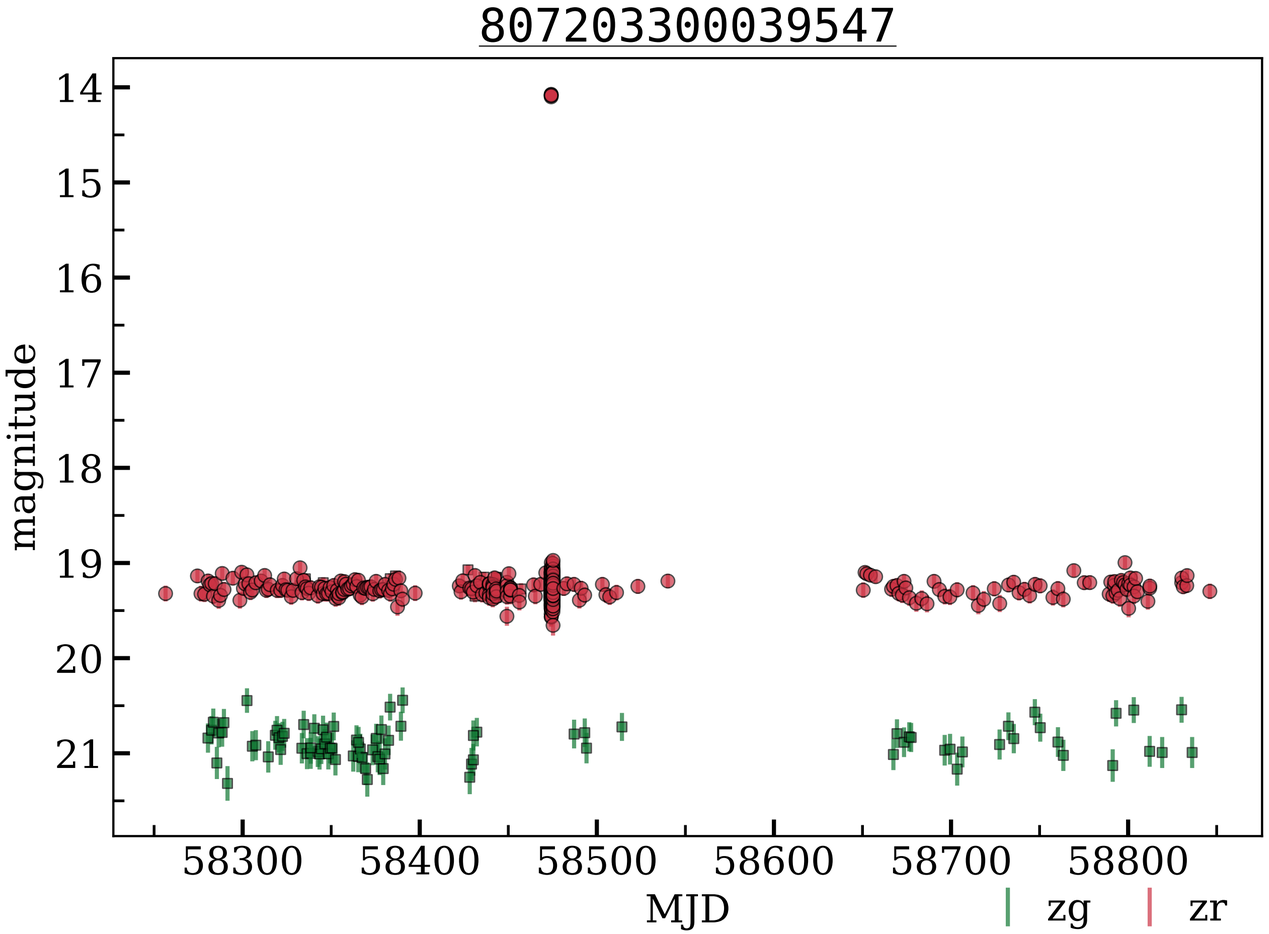
m0n5t3r @ Flickr

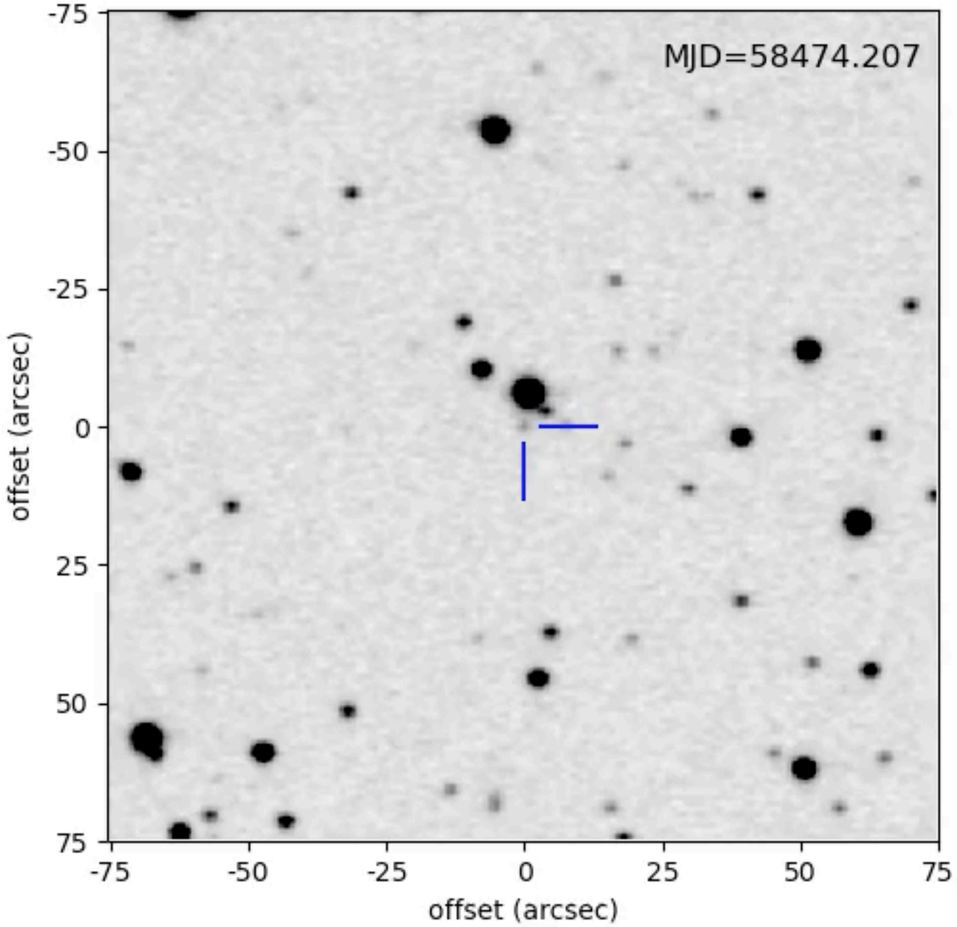


Sergey Lisakov

Barcelona

Asteroid 945

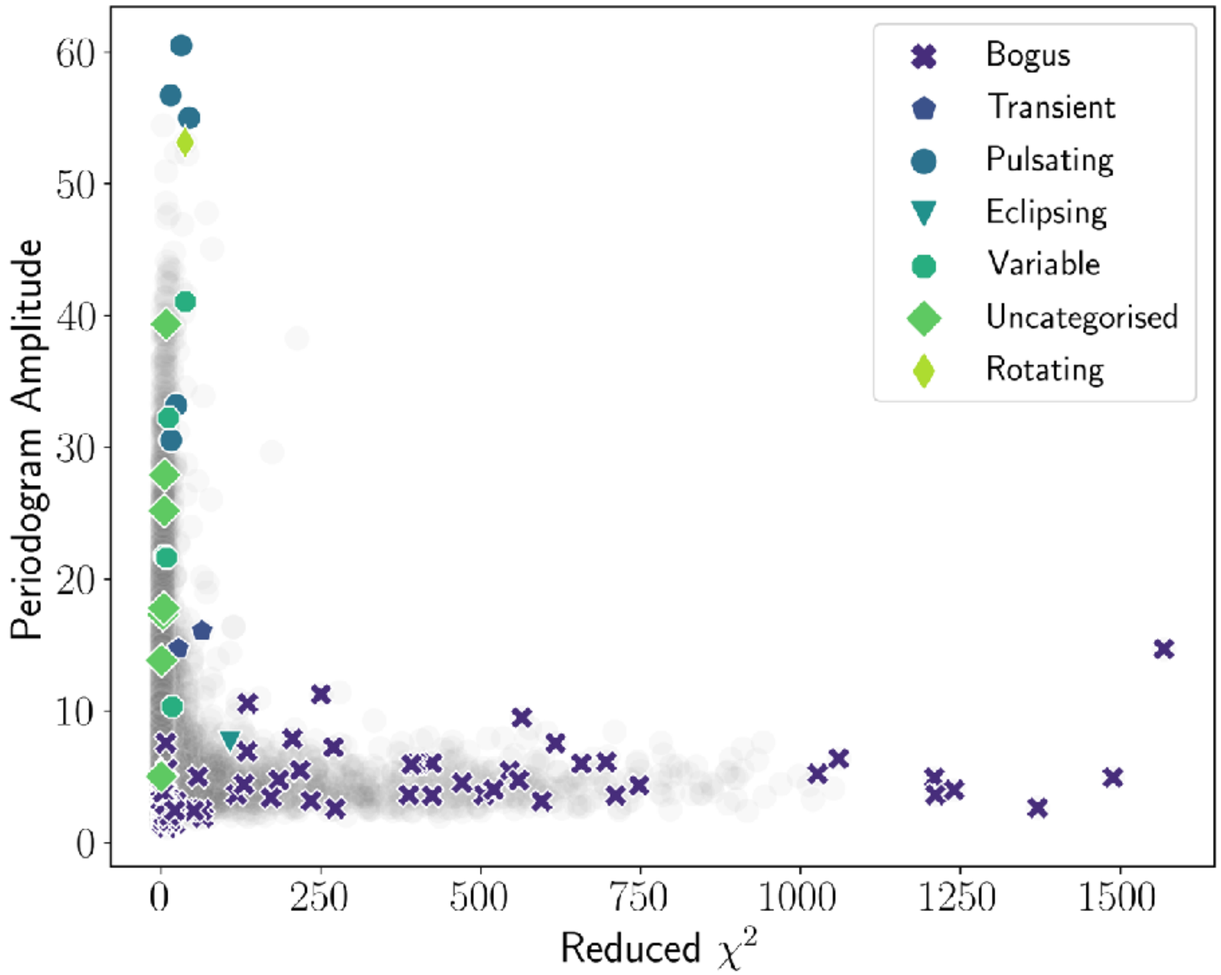


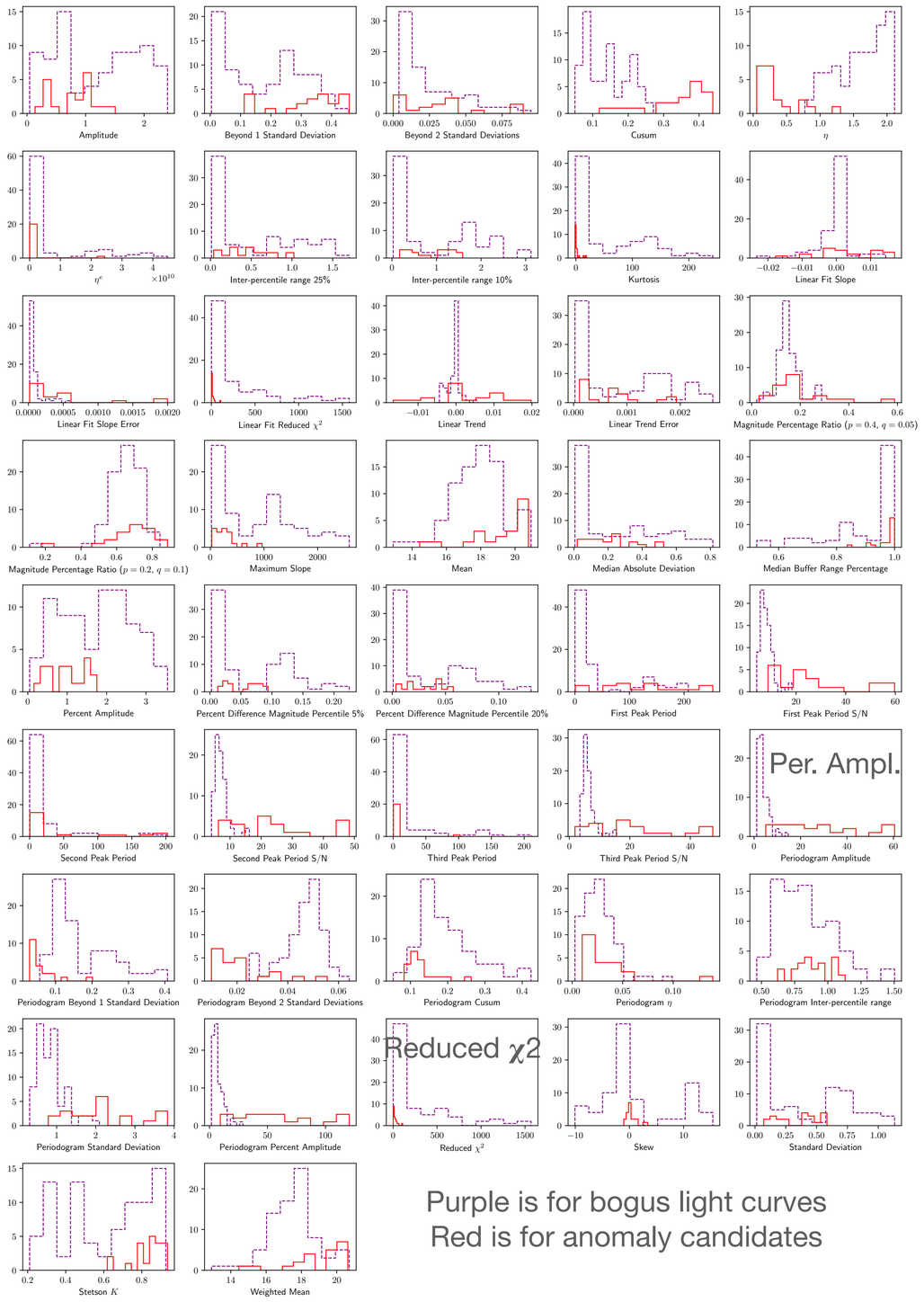


Exploratory Data Analysis

Bogus vs Real

M31 field

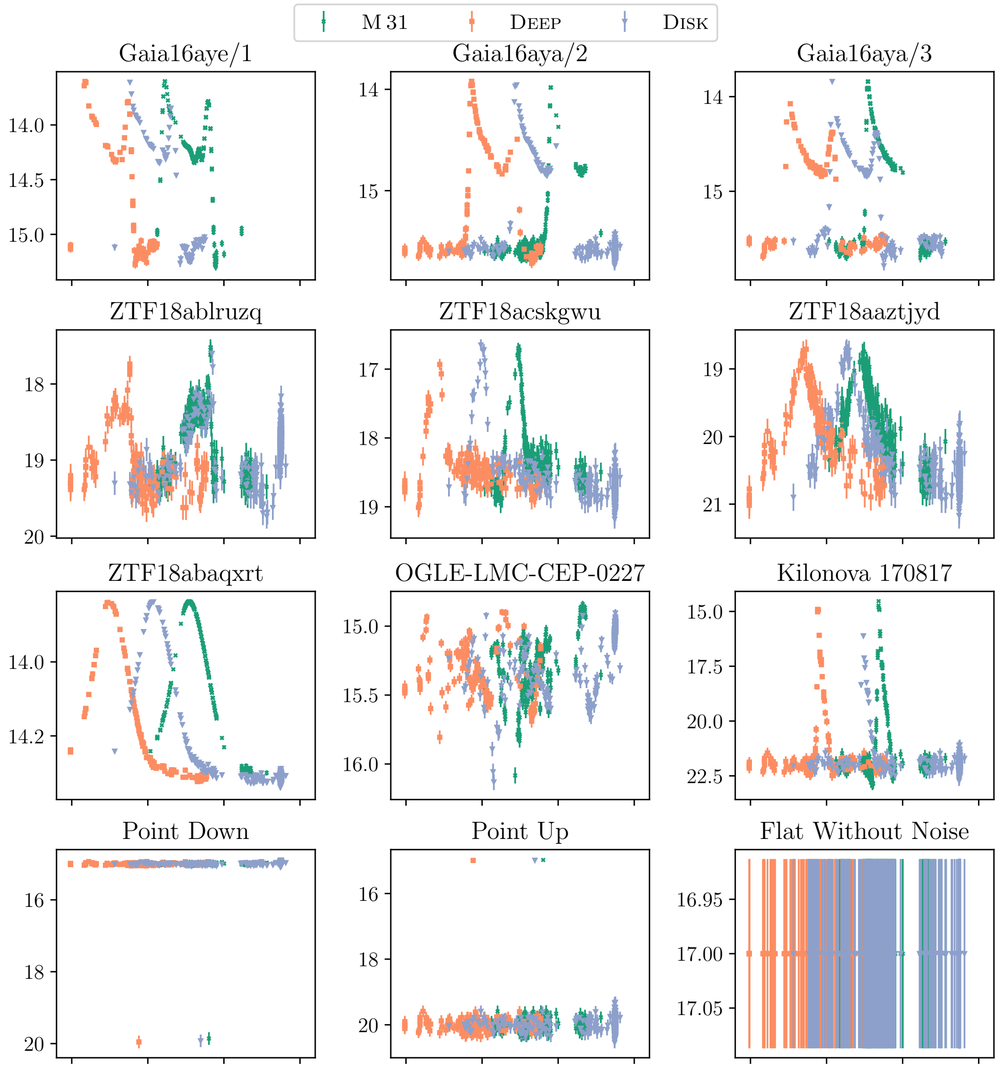


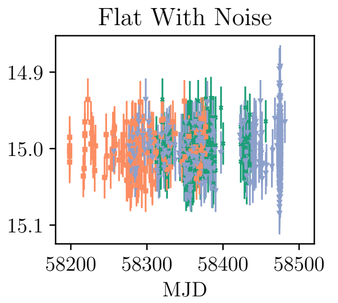


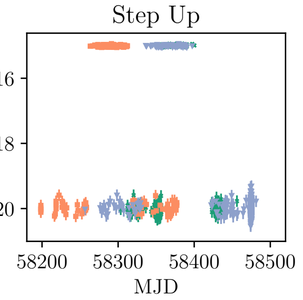
Purple is for bogus light curves
Red is for anomaly candidates

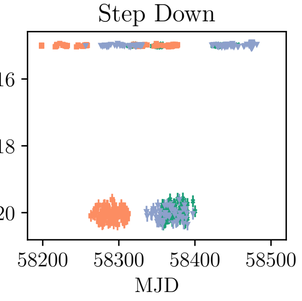
Outlier Detection Pipeline Verification

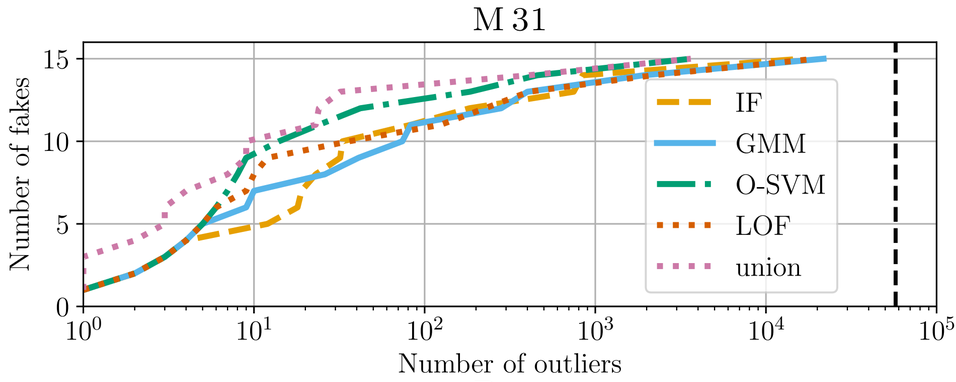
Fake light curves

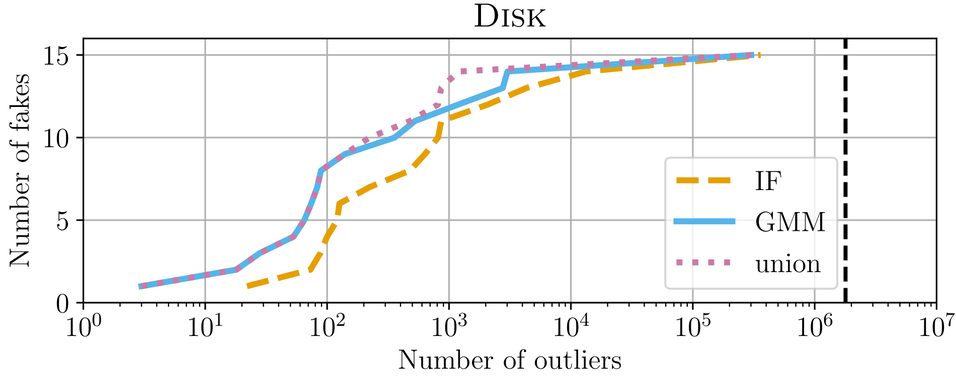


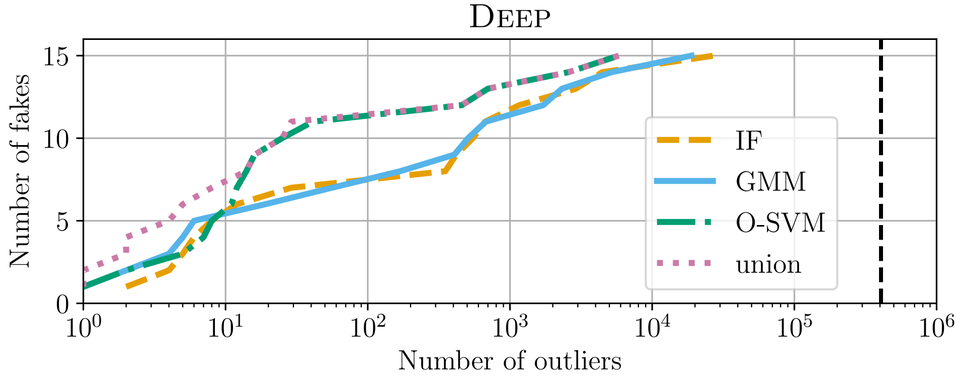




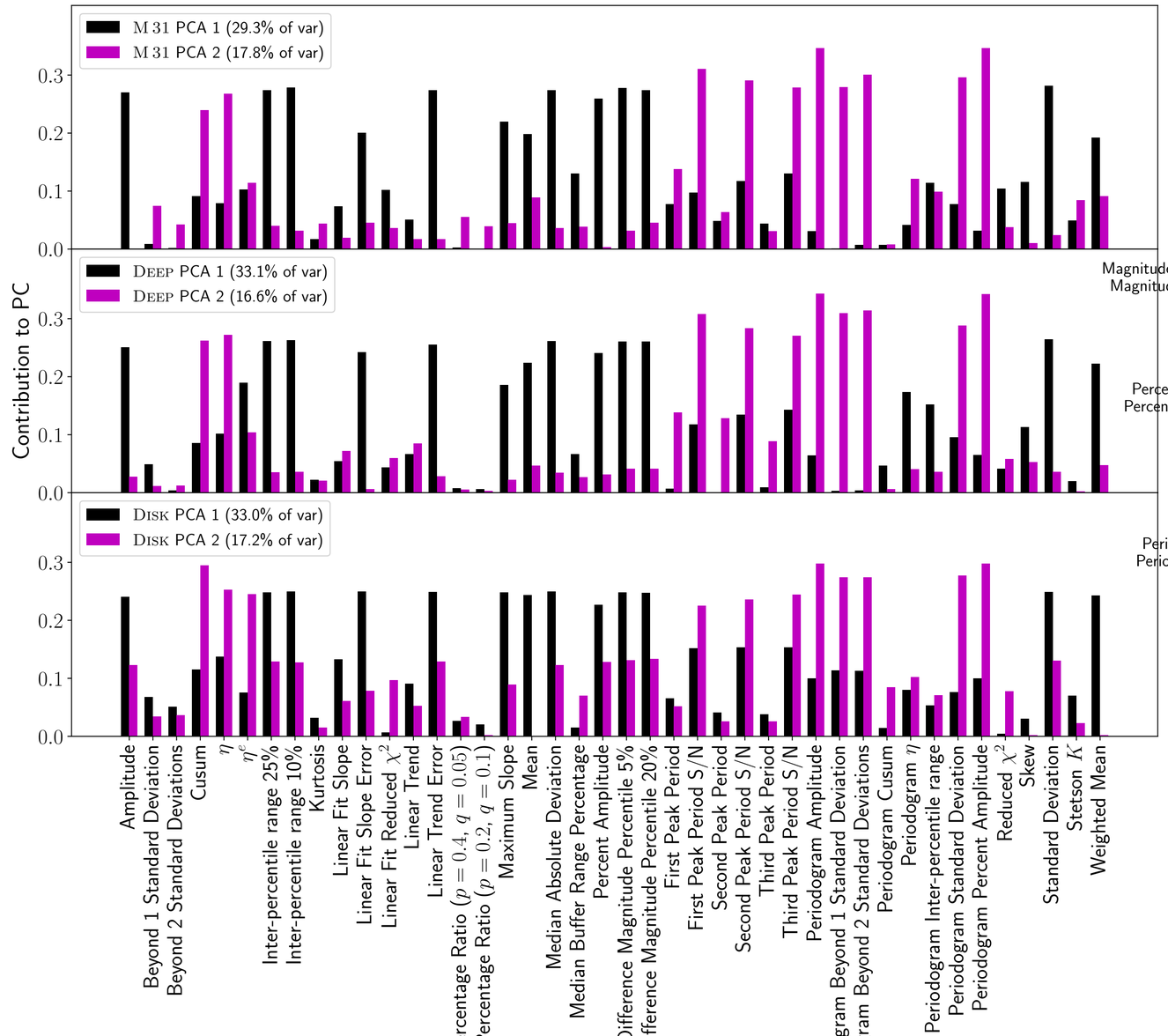


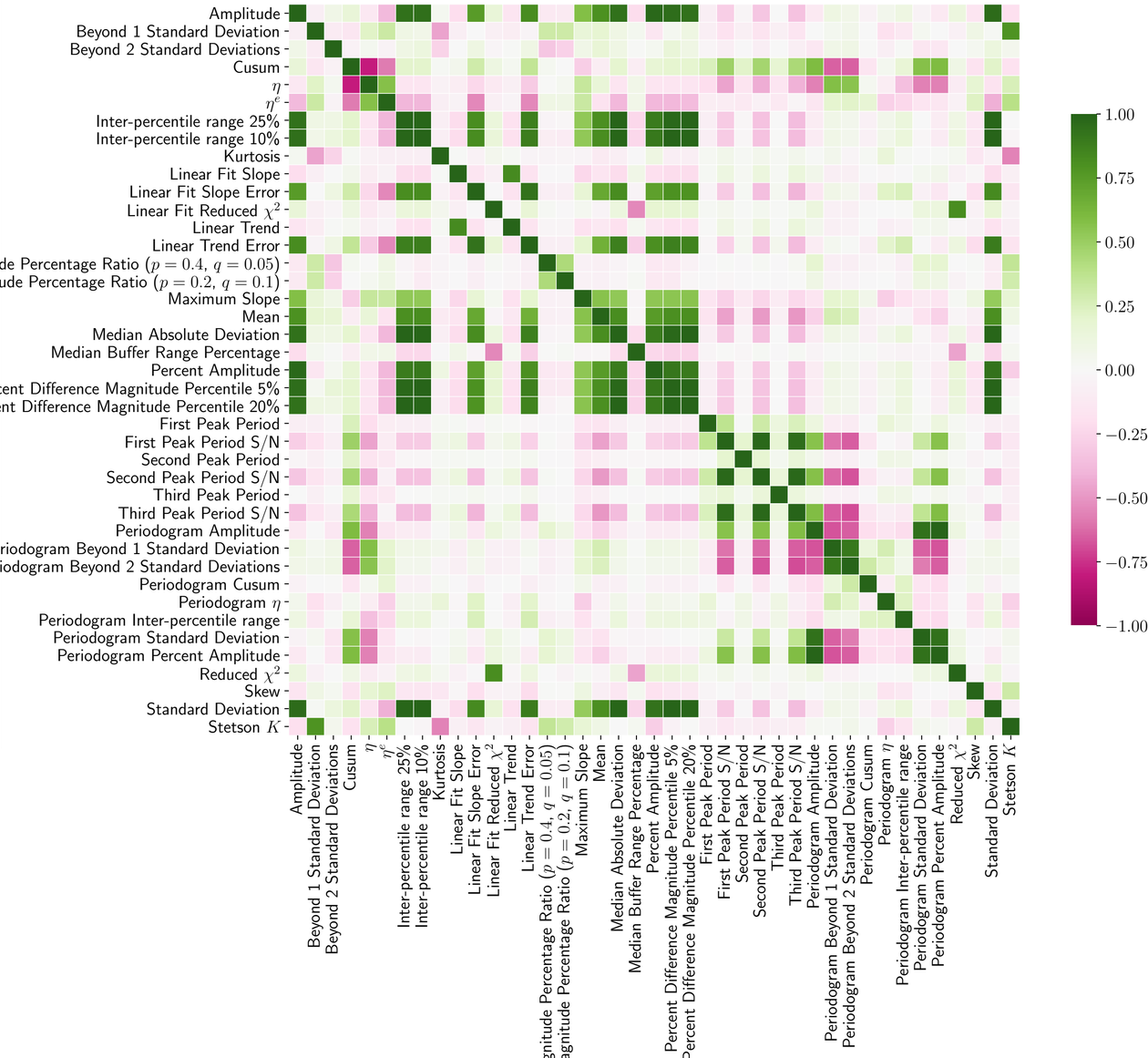






Principal Component Analysis





Non-catalogued Anomaly Candidates

Three largest contributions to the three principal components