



Discovering the Unknown in the LSST era

Organizers: Ashish Mahabal, Xiaolong Li, Federica Bianco

#day4-thu-1600-discovering-unknown

Get Notifications for @ MentionsHuddle



Reminder - Code of Conduct



Project & Community Workshop 2023 7-11, August 2023 | Marriott University Park Tucson | Tucson, AZ

Agenda Register Travel & Venue Code of Conduct

Harassment and unprofessional conduct (including the use of offensive language) of any kind is not permitted at any time and should be reported to:

- Andrew Connolly (ajc@astro.washington.edu),
- John Franklin Crenshaw (jfc20@uw.edu), and/or
- Alysha Shugart (<u>ashugart@lsst.org</u>).



Rubin Observatory adheres to the principles of kindness, trust, respect, diversity, and inclusiveness in order to provide a learning environment that produces rigor and excellence. Check name-tags for these contact comfort level stickers. Wear a mask if you want to!

Elbow/Fist Bump OK

Use the confidential email <u>rubin2023-covid@lists.lsst.org</u> to request a test, report your test results, or ask questions.

Need My Space



Handshakes OK

If someone is wearing a pin like this, and it indicates a low social battery, please give them their space or offer to restart the conversation at a later time.

If you feel unsafe at any time send an email to rubin2023-helpline@lists.lsst.org

2



Reminder - Virtual Participation



Virtual participants should be muted when they're not speaking.



In-person participants should speak into the room microphone(s), or the chair should repeat all questions into the microphone, so that the virtual participants can hear what is said.



In the Rubin2023_PCW Slack Space, all participants can use the session's channel for Q&A and discussion. The channel name convention is, e.g.: #day1-mon-slot3a-intro-to-rubin



In Zoom, use the chat to:

- request to unmute to ask a question, or
- type your question so someone can speak it aloud.

The Zoom "raise hand" feature is generally harder for moderators to track, and is not preferred, but may be used at the discretion of the session chair.

3



- Federica Bianco: Rules of engagement, intro, TVS anomalies subgroup (<=7 min)
- Ashish Mahabal: new ISSC anomaly detection interest group (<=7 min)
- Xiaolong Li: Time domain perspective (<=10 min)
- Will Clarkson: Proper motion perspective (<=10 min)
- Ari Heinze: Solar System Perspective (<=10 min)
- Agnieszka Pollo: Extragalactic Perspective (<=7min)
- Konstantin Malanchev: Precursors: detection of anomalies in ZTF (<=10 min)
- Ashley Villar (remote): detection of time domain anomalies with AI (<=10 min)

1 sliders:

Chris Lintott (video); Ashish Mahabal; Abi Saha; Patrick Aleo (by proxy)

Anyone else who wants to speak and panel discussion!



Includes members from TVS with similar interest

Patrick Aleo Federica Bianco Eric Feigelson Bryce Kalmbach Xiaolong Li Ashish Mahabal Alex Malz Konstantin Malanchev Anais Moller Gautham Narayan Priscila Pessi Fabio Ragosta Monika Soraisam Ashley Villar

https://docs.google.com/document/d/1llu11fCLXt4wFk-1a7wS41tsyyj36cV7f8pG9_MkwsU/edit

Mahabal | TVS anomaly subgroup group

There is also a TVS anomaly subgroup, but it's not very active... contact Federica if you want to be the new chair!

Federica Bianco Neven Caplar Sid Chaini Tatiana Acero Cuellar Shar Daniels James Davenport Tansu Daylan Rafael Martínez-Galarza Swayamtrupta Panda Kaylee de Soto Elahe'e Khalouei Andrew Levan Ashish Mahabal Keming Zhang

https://docs.google.com/document/d/1llu11fCLXt4wFk-1a7wS41tsyyj36cV7f8pG9_MkwsU/edit



Intended meeting frequency: once a month

Talks Discussions Datatypes, Datasets Types of anomalies Methods Cross fertilization





Intended meeting frequency: once a month

Talks Discussions Datatypes, Datasets Types of anomalies Methods Cross fertilization

Rare events In-class outliers (extreme) Inter-class objects New phenomena/classes

Artifacts -> improvements



Mahabal | ISSC anomaly interest group





Possible first project (proposed by Alex Malz)

Analyze audio data from stress testing the camera system at SLAC Filters, motions, ...

Look for anomalies

Alex Boucaud



Xiaolong Li



Acronyms & Glossary

Unknowns Unknowns with Rubin LSST: Time Domain and Proper motion

We wrote two papers aimed at optimizing the survey cadence for the discovery of never-before seen phenomena Xiaolong Li, Fabio Ragosta, Will Clarkson, Federica Bianco



https://iopscience.iop.org/article/10.3847/1538-4365/ac3bca

Published in ApJS Rubin Survey Strategy Focus Issue

To be submitted to ApJS Rubin Survey Strategy Focus Issue

Rubin LSST PCW 2023 August 10

Unknown unknowns with Rubin LSST : Time domain

Feature space

Designing algorithms to discover the unknown

Recommended Observing Strategy

Unknown unknowns with Rubin LSST : Proper motion

Confusion Index

Likelihood Score

Transient proper motion



Time domain novelties

II. Unknown Unknowns with Rubin LSST: Time domain

Unknown unknowns with Rubin LSST: Metrics

Goal: evaluate the ability of the LSST survey to discover "true novelties": phenomena never before seen nor predicted from theory

Model-independent approach: to measure completeness in the space of features that Rubin can observe



Unknown unknowns with Rubin LSST: Metrics



Time gaps in observations taken

Unknown unknowns with Rubin LSST: Metrics



Unknown unknowns with Rubin LSST: Figure of Merit



Observing strategy for unknown unknowns



Credit: Lynne Jones https://epyc.astro.washington.edu/~lynnej/opsim_downloads/ender_a1_v3.1_10yrs__N_Visits.mp4 <u>PSTN-055</u> Official Rubin Cadence Recommendation 1/2023 : Rolling is recommended on the WFD, with the sky split into two 1/2-sky regions defined by declination limits and with a 0.9 rolling weight.

Opsim 3.0 and 3.2 - 2023





Will Clarkson

Vera C. Rubin Observatory | Operations PCW | 7-11 August 2023

Acronyms & Glossary

Unknown unknowns with Rubin LSST : Time domain

Feature space

Designing algorithms to discover the unknown

Recommended Observing Strategy

Unknown unknowns with Rubin LSST : Proper motion

Confusion Index

Likelihood Score

Transient proper motion

Presenter: Will Clarkson

LSST: proper motions ~1 mas/yr @ r=24 (full 10y)

E.g. Ivezic, Monet et al. 2008:

"LSST will smoothly extend Gaia's error-vs-magnitude curve to a four-magnitude fainter level."

LSST will attach proper motion to objects which...

- ... are too intrinsically faint for Gaia, and/or
- ... lie in regions too highly extincted for Gaia, and/or
- ... are only bright enough for proper motion measurement for a short interval during the LSST survey

Thanks to the **rubin_sim** team, a metric for the predicted proper motion uncertainty for LSST observations is available, calculating proper motion uncertainty as straight-line gradient uncertainty incorporating individual position-measurement uncertainty.



Proper motions for true anomalies

Led by Fabio Ragosta (now at INAF/Rome), we have been investigating the ability of LSST to add proper motion evidence to previously-unexpected astrophysical objects it may discover, using the **rubin_sim** framework to assess candidate observational strategies.

Three figures of merit:

- Confusion Index (CI);
- Likelihood Score (LS);
- Transient Proper Motion (TPM);

In general, stellar populations are simulated in 6D (position, velocity) following standard prescriptions for the "known" Galactic components, and the capability of LSST to isolate objects of interest against this "known" background is assessed using the metadata for candidate observing strategies ("opsims")..



fabio.ragosta@inaf.it

Proper motion: Confusion Index (CI)

The CI parameterizes the ability of LSST to kinematically identify comoving populations via their locations in the diagram of reduced proper motion (RPM) vs (g-r) color.

For each simulated object, measurement uncertainty in the RPM/g-r diagram is parameterized by the "area" of the uncertainty ellipse in this space (right, below).

This is then used as the input uncertainty for mixture models for Sgr-dwarf-like objects injected at various distances from the observer. The result is a map in RPM/g-r space of objects that are considered confused (i.e. ambiguous membership probability to the field or the desired population).

$$H_g = g + 5 \log_{10} \left\lfloor \frac{\mu}{\text{mas/yr}} \right\rfloor - 10 - A_g$$

$$\pi \times \det(C_{H,gr}) = \pi \left(\sigma_g^2 \sigma_r^2 + \frac{4.715\sigma_\mu^2}{\mu^2} \left[\sigma_g^2 + \sigma_r^2 \right] \right)$$

Proper motion: Confusion Index (CI)



Proper motion: Likelihood Score (LS)

The LS parameterizes LSST's ability to kinematically separate a point source from the "background" due to known Galactic components using LSST proper motions.

Galactic populations are simulated for each sight-line (the "known" population) and an "unknown" (i.e. anomalous population) is simulated on top of them. Objects are then recovered as "unusual" if they are **measured** to fall outside the inter-quartile range corresponding to the Galactic population.

Finally, the fraction of injected "unusual" objects that are recovered as "unusual" is evaluated for the entire simulated survey.

Figures from Ragosta et al. (in prep.)



Proper motion: Transient Proper Motion Metric (TPM)

The TPM parameterizes LSST's ability to attach proper motion information to transients (particularly those that are only bright enough for proper motion measurement during the transient event).

A population of transients is simulated (for a few example regimes). Then the Transients metric (of Li et al. 2023) is assessed to determine which objects LSST would detect as transients.

For the subset that LSST would detect as transients, the proper motion uncertainty is evaluated over those observations for which the transient is sufficiently bright for LSST detection at 5 sigma.



Microlensing lightcurve from the OGLE-III website

Proper motions for True Anomalies: status

Three figures of merit:

- Confusion Index (CI) Implemented.
- Likelihood Score (LS) Implemented.
- Transient Proper Motion (TPM) Finalizing implementation.

Next steps:

- Finish implementing and testing TPM
- Run on the latest **opsim**s to determine which candidate strategies are best for proper motion anomaly detection.

Ragosta et al. 2023 in prep for ApJS special issue on LSST strategy.

Input is welcome!



fabio.ragosta@inaf.it



Ari Heinze

Vera C. Rubin Observatory | Operations PCW | 7-11 August 2023

Acronyms & Glossary



Finding the Unexpected in the Solar System

Ari Heinze

DiRAC Institute and the Department of Astronomy, University of Washington



AURA

SLAC

NOIR Lab



Possible surprising discoveries

New types of objects New types of orbits New types of behavior New associations of objects New objects in the solar system (interstellar) objects)

How to find/recognize the unexpected

Ari Heinze

Image credit: Bruno C. Quint

New types of objects Unlikely to see objects of an entirely different composition Might see new types of fragments from known bodies (e.g., nitrogen icebergs) but still unlikely.

Ari Heinze

Image credit: Bruno C. Quint

New types of orbits



New types of orbits

Unlikely: objects are already known in nearly all physically possible solar system

orbits

Might find more in very rare orbit types: e.g. retrograde asteroids
New types of behavior Likely – the only question is what type Objects suddenly appearing (emission of dust, ??) Objects suddenly disappearing (slow, high-amplitude rotation, clearing of dust, ??) Change in orbit (gas emission, Yarkovsky)

Ari Heinze

effect, ??)

Image credit: Bruno C. Quint

New associations of objects

- Guaranteed to be found
- Collisional families in almost the same orbits
- Asteroid pairs from YORP fission
- Will reveal recent exciting events in the asteroid belt
- Will cause me a lot of trouble (but it's worth it)

Ari Heinze

New associations of objects



Asteroids 87887 – 415992: the youngest known asteroid pair?

J. Žižka¹, A. Galád^{2,}, D. Vokrouhlický¹, P. Pravec³, P. Kušnirák³, and K. Hornoch³

Ari Heinze

Interstellar objects Rates are highly uncertain, but no major obstacle to detection

Ari Heinze

Image credit: Bruno C. Quint

How to find the unexpected Only humans can discover Computers can sort through the normal stuff, show us anomalies we should check out Experts and beginners may see different things Write fast, efficient code Search areas of parameter space you know are empty

Ari Heinze

Image credit: Bruno C. Quint



Agnieszka Pollo

Vera C. Rubin Observatory | Operations PCW | 7-11 August 2023

Acronyms & Glossary

Search for anomalous objects in the ALLWISE survey

Agnieszka Pollo

National Centre for Nuclear Research (Warsaw) & Jagiellonian University (Cracow); Poland with

Aleksandra (Ola) Solarz; Maciek Bilicki, Mariusz Gromadzki, M., Anna Durkalec, Michal Wypych

Tucson, 7.08.2023



Based on the paper: "Automated novelty detection in the WISE survey with one-class support vector machines", Solarz et al. 2017 Astronomy & Astrophysics, Volume 606, id.A39

Angieszka Pollo

Wide-field Infrared Survey Explorer (WISE)

All-Sky survey in IR

- Detected over 747 mln sources
- (15 PB of data; tables + images)
- Publicly available (position, photometry in 4 bands (3.6-22 um))
- Low angular resolution (~6")
- No redshift information



(http://wise2.ipac.caltech.edu/docs/release/allsky

Angieszka Pollo

Exploration of parameter spaces

 \rightarrow The usual first approach to selection: CC diagrams

→ Novel or anomalous
 sources deviate from
 expectations but could mimic
 a behaviour of a 'normal'
 object

 \rightarrow need for new ML approaches based on larger parameter spaces



WISE: first step towards ML novel source detection

Training set: \rightarrow AllWISE x SDSS (α , δ) with secure spectro-z



Solarz et al. 2017



WISE: first step towards ML novel source detection

Solarz et al. 2017





Parameter space: → Brightness W1 →Color W1-W2 →Compactness W1mag13 = Angieszka Poll@mpro(5") - w1mpro(11")

WISE: accounting for unknown Solarz et al. 2017 Unknowns



Novelty detection with One-Class Support Vector Machines



- Create one 'known' class (mix of AllWISE x SDSS galaxies, stars, QSOs)
- Maps input data to a higher D parameter space (based on Kernel methods)
- Hypersurface hugging the expected sources
- Anything with 'unknown' patterns falls outside the hypersurface => novelties



Solarz et al. 2017

Angieszka Pollo

Spurious sources

- W1-W2 ~ -1 ; 80%; Spitzer GLIMPSE: IRAC I1 [3.6 um], IRAC I2 [4.5 um]
- Low WISE resolution (6") in crowded fields
 => blends
- OCSVM: good tool for selecting hidden artefacts





Solarz et al. 2017



AGN candidates?

- 30,000 sources (those in Galactic Plane: still mostly blends)
- 76% undetected at other wavelengths!
- ~7 000 objects with SDSS photometry (no spectro-z), follow-up observations
- Peculiar (dusty) QSOs
- Low-z very dusty galaxies
- Very dusty Galactic objects like YSO

Solarz et al. 2020

6

4

2

0

 $^{-2}$

6

8

mag

[Vega

W1-W2



Conclusions

- With sufficiently big data (like LSST) anomalies may come in big numbers
- Parameter space is important
- What we find is mostly trash but
 - valuable trash if not found by other methods \rightarrow novelty search as a method for additional cleaning of catalogs
 - some genuine atypical sources
 - nedeed methods of separation (clustering in different parameter spaces/with DL)

Angieszka Pollo



Konstantin Malanchev







SNAD – personalized anomaly detection for astronomy

Konstantin Malanchev (LINCC Frameworks @ CMU) & SNAD Team

Rubin Project and Community Workshop, 10 August 2023

Konsta Malanchev

Anomaly Detection is Discovery



"An anomaly is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism" - Hawkins, 1980

Stages of Discovery in Astronomy:

- Detection
- Interpretation
- Understanding
- Acceptance

Which mechanism? Is it something we are familiar with but fail to proper model or recognize? Is it something we have never seen before?

Is there something new for us to learn?

In order to identify the unusual we need to have a clear ideal of what is usual ...

.. and that is a social construct. It changes and adapts with time!

Konsta Malanchev

Light Potentially interesting anomalies: **Machine learning** curves Candidate 1 Sandidate 2 e.g. outlier Candidate 3 ... Preprocessin detector . . . Active anomal detection loop **SNAD Viewer** Not interesting Get more Metadata data Images Catalogs Interesting and **Simulations** M Publication Follow up Very interesting! Konsta Malafichev Literature

Pipeline

ML produces recommendations only

G

Active anomaly detection

From outlier to anomaly detection algorithm

How to discriminate annoying non-anomalies sources and bogus light curves?

- We can ask an expert interactively about each new outlier
- If it is not an anomaly, show similar objects less often
- Retrain, ask the expert again



Some interesting objects

144 objects reported to TNS, mostly PSNe and AGNs https://snad.space/catalog

Two SLSNe (one spec.-confirmed by other group afterwards)

Various variable stars: RS CVn (spec. confirmed by us), binary Mira, YSOs

Hundreds of well-sampled ~hour long flares





Objects found at SNAD-VI (June 2023) 2023RNAAS...7..155V



µ-lens candidate

Optical counterpart of radio source NVSS J080730+755017



Artifacts and bogus variability



Photometric contamination from a bright variable https://snad.space/art/ Konsta Malanchev





remote

Ashley Villar



Acronyms & Glossary



Anomaly Detection for LSST transients

V. Ashley Villar Harvard University

Vera C. Rubin Observatory | Operations PCW | 7-11 August 2023

Ashley Villar

Acronyms & Glossary



Out-of-distribution anomalies are "easy" to find





Out-of-distribution anomalies are "easy" to find



Vera C. Rubin Observatory **Operations PCW** 7-11 August 2023



But how do we choose the axes..?





But there is no physical model to describe ALL physics





And so, we become data-driven











Such methods can be extended to the ~135 million variable stars LSST will observe

Chen, Cheung, VAV, Ho 2021; 2022 **Ashley Villar**


Our latent space is physics-unaware





That is because our neural net wants to optimize one thing*



The NN's objective:

Minimize the difference between the model and data

*This is a lie. The variational AE I showed earlier also encourages a specific (Gaussian) distribution.

Ashley Villar



But we know about other, reasonable distance metrics

E.g.,

Minimize the difference between the model and data

And

Make signals of similar types cluster together (called contrastive learning)

And

Make the latent features independent of redshift (e.g., group-invariant learning)

Ashley Villar



Thank you!

Vera C. Rubin Observatory | Operations PCW | 7-11 August 2023

Ashley Villar



Lightening talks









Vera C. Rubin Observator





Abi Saha | EARLY detection of rare TD phenomena

THE ASTROPHYSICAL JOURNAL, 892:112 (19pp), 2020 April 1 © 2020. The American Astronomical Society. All rights reserved.

https://doi.org/10.3847/1538-4357/ab7b61



A Classification Algorithm for Time-domain Novelties in Preparation for LSST Alerts. Application to Variable Stars and Transients Detected with DECam in the Galactic Bulge

Monika D. Soraisam^{1,2}, Abhijit Saha³, Thomas Matheson³, Chien-Hsiu Lee³, Gautham Narayan², A. Katherina Vivas⁴, Carlos Scheidegger⁵, Niels Oppermann⁶, Edward W. Olszewski⁷, Sukriti Sinha⁸, and Sarah R. DeSantis⁸ (ANTARES collaboration)



Southern Horizons in Time-Domain Astronomy Proceedings IAU Symposium No. 339, 2018 E. Griffin, B.D. Editor & C.E. Editor, eds.

© 2018 International Astronomical Union DOI: 00.0000/X0000000000000000X

Recognition of Rare and Peculiar Temporal Phenomena from LSST Alert Streams





Figure 2. The frequency distribution of magnitude differences for 2 measurements of a sine wave (left) and a sawtooth (right) light curve with period 0.5 days, when sampled at three epochs separated successively by 0.25 and then 0.12 days. See description in § 3.2 for a complete description.

VERA C. RUBIN OB SERVATORY **Patrick Aleo** | **Real-time Anomaly Detection in ZTF**

