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

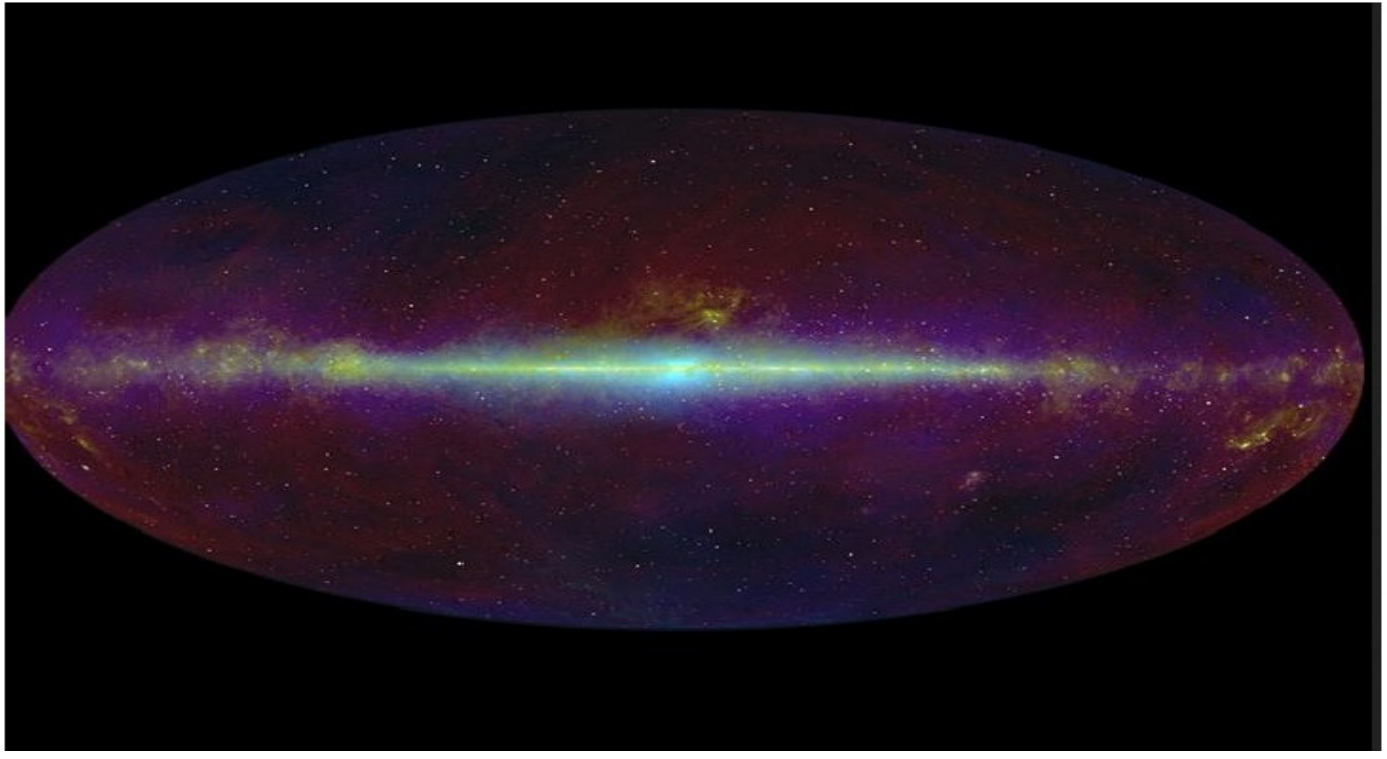
Based on the paper:
”Automated novelty detection in the WISE survey with one-class support vector machines”,
Solarz et al. 2017
Astronomy & Astrophysics, Volume 606, id.A39

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

Solarz et al. 2017



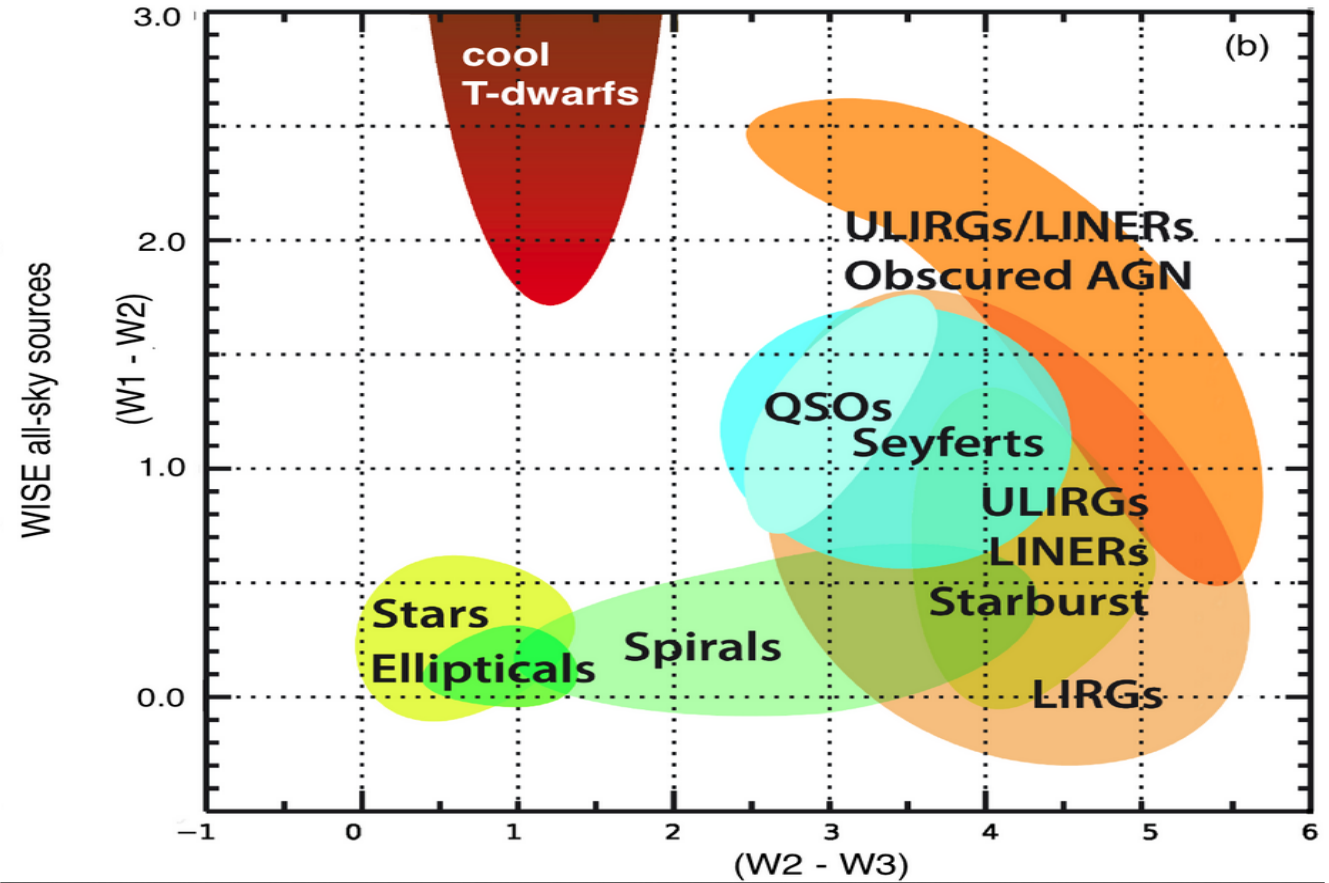
(http://wise2.ipac.caltech.edu/docs/release/allsky/)

Exploration of parameter spaces

→ The usual first approach to selection: CC diagrams

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

→ need for new ML approaches based on larger parameter spaces

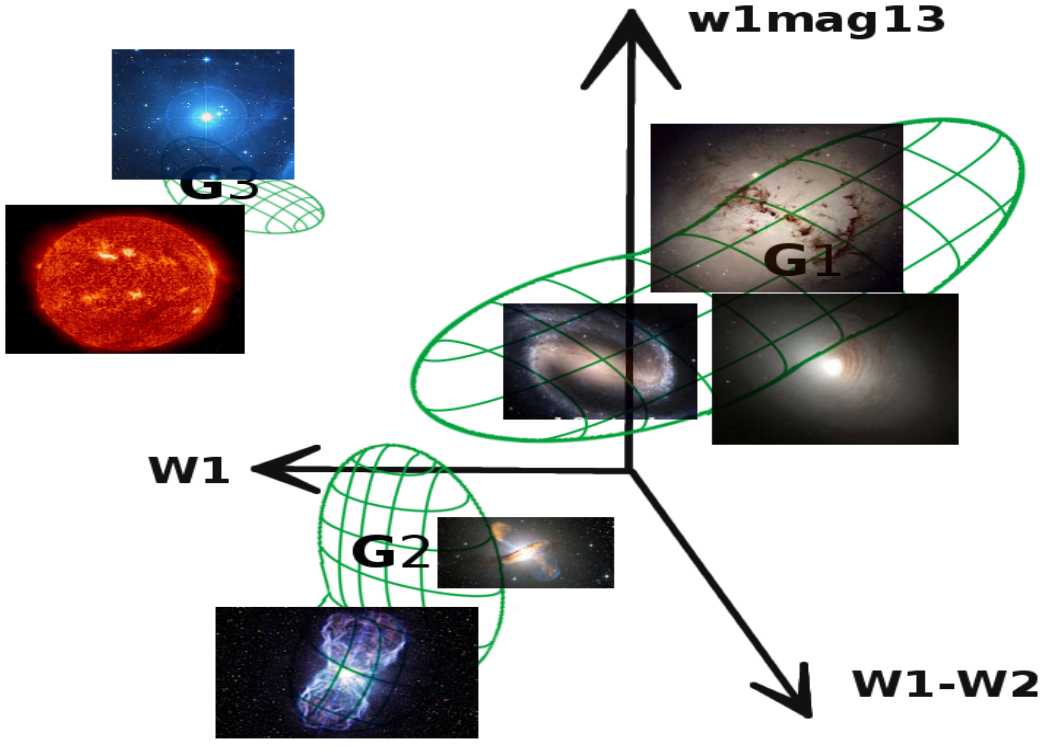


Credit: Wright+10

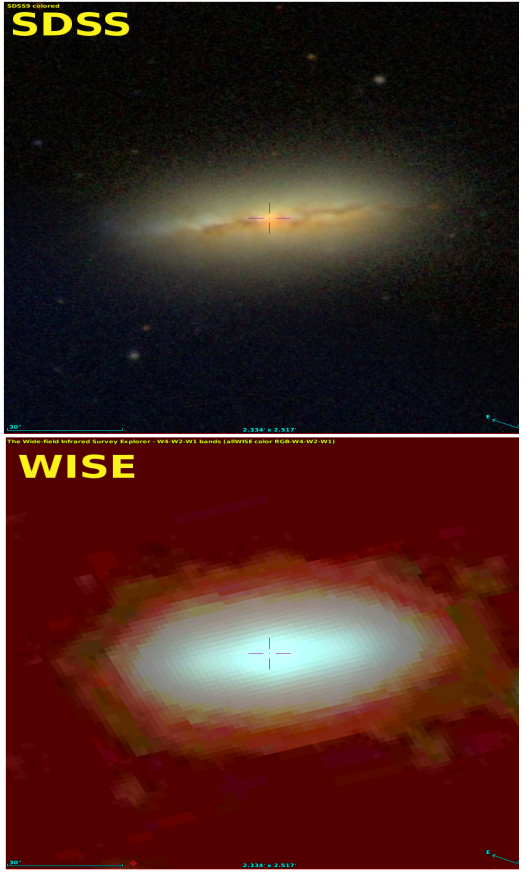
WISE: first step towards ML novel source detection

Training set:

→ AllWISE x SDSS (α,δ) with secure spectro-z

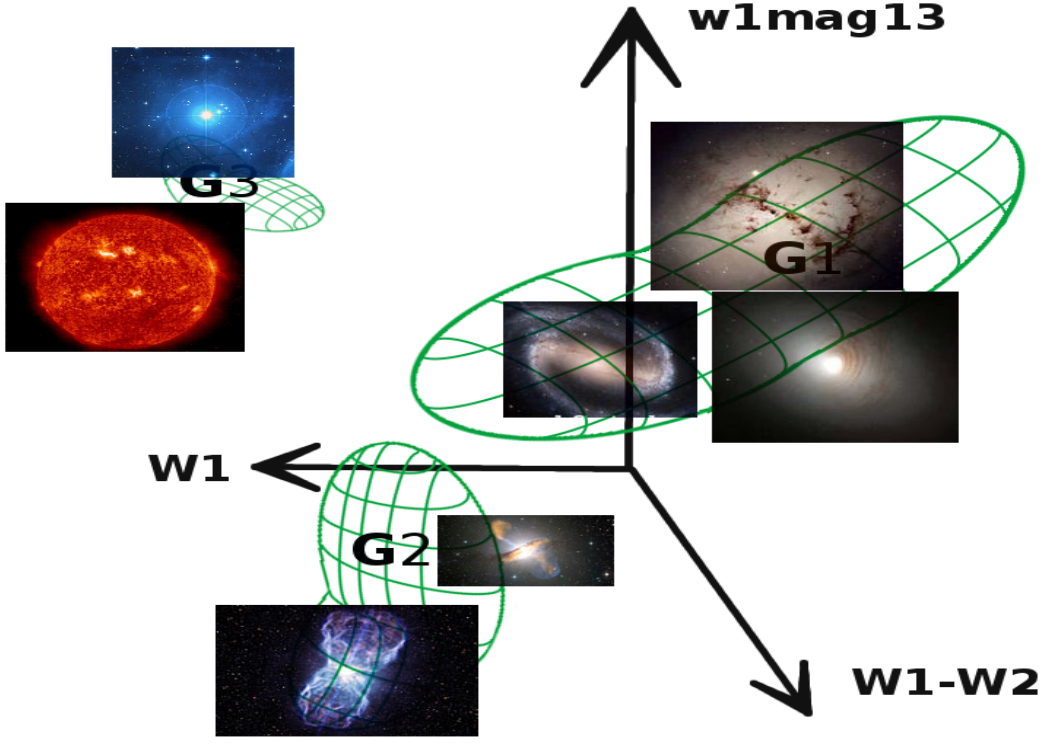


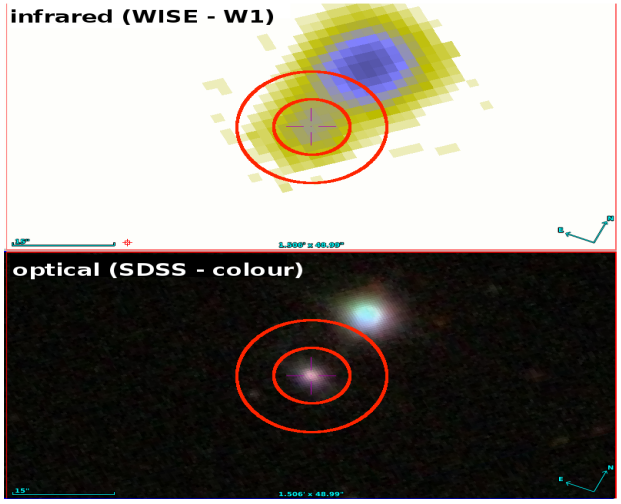
Solarz et al. 2017



WISE: first step towards ML novel source detection

Solarz et al. 2017





Parameter space:

→ Brightness W1

→ Color W1-W2

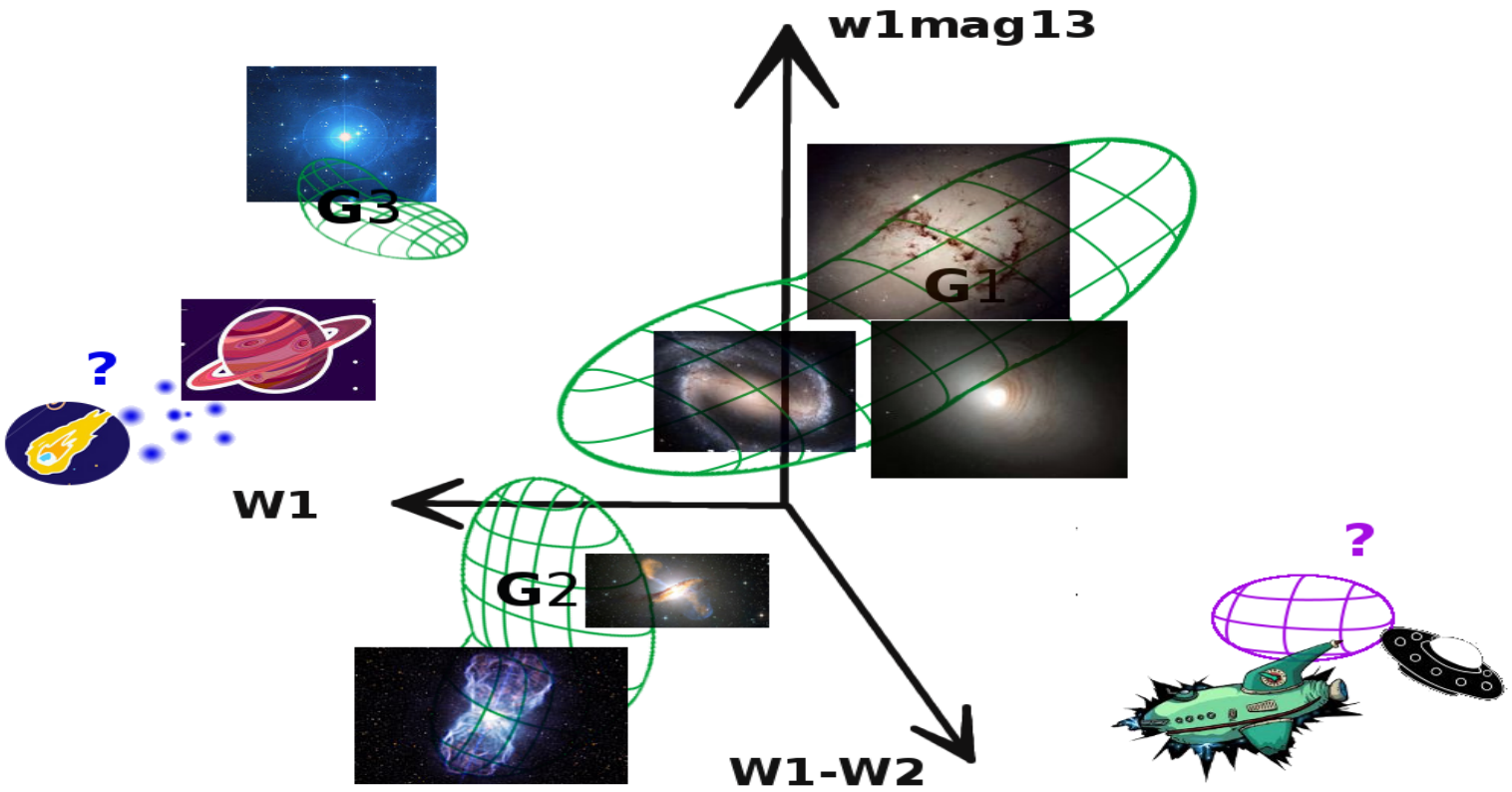
→ Compactness

W1mag13 =

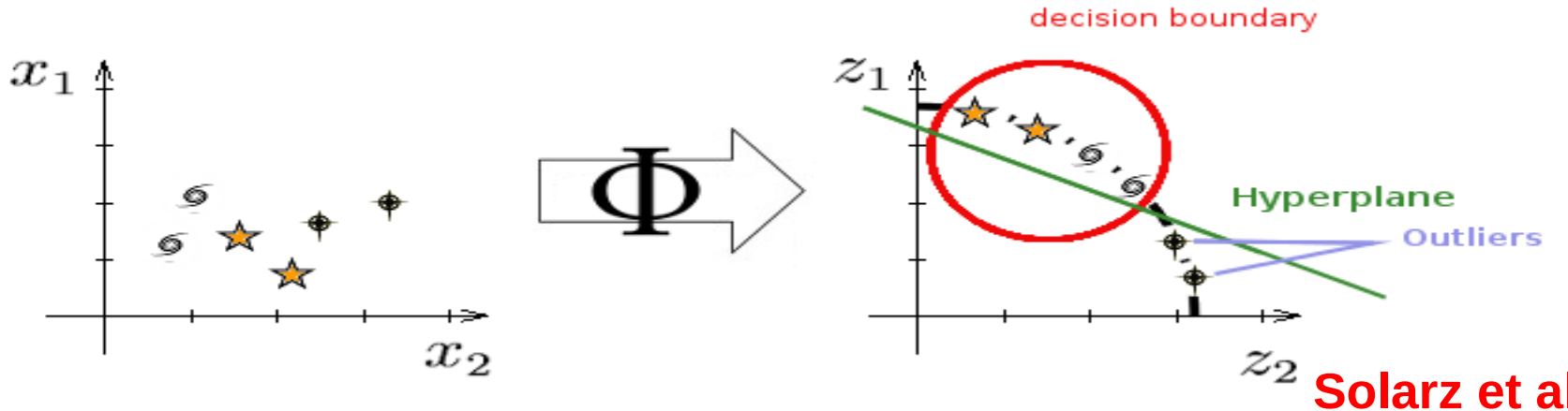
w1mpro(5”) - w1mpro(11”)

WISE: accounting for unknown unknowns

Solarz et al. 2017



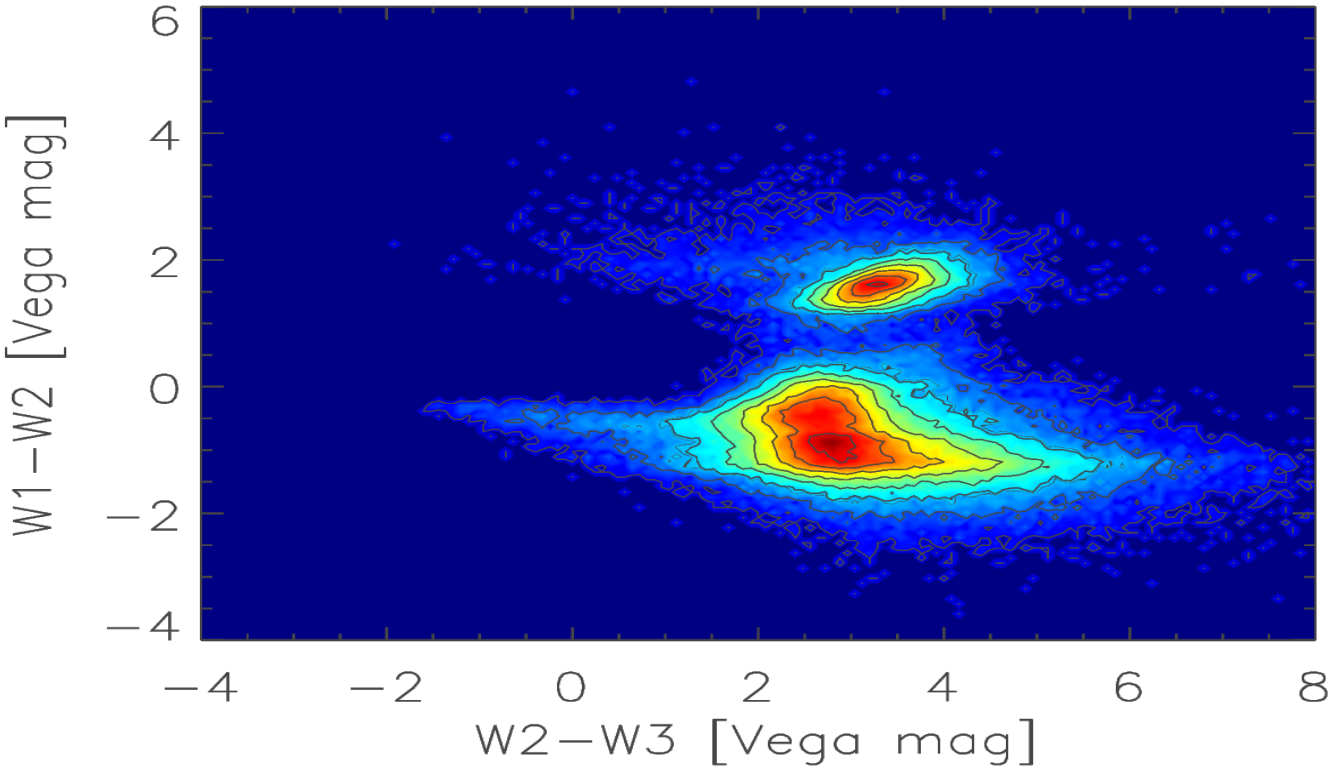
Novelty detection with One-Class Support Vector Machines



Solarz et al. 2017

- 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

Results:



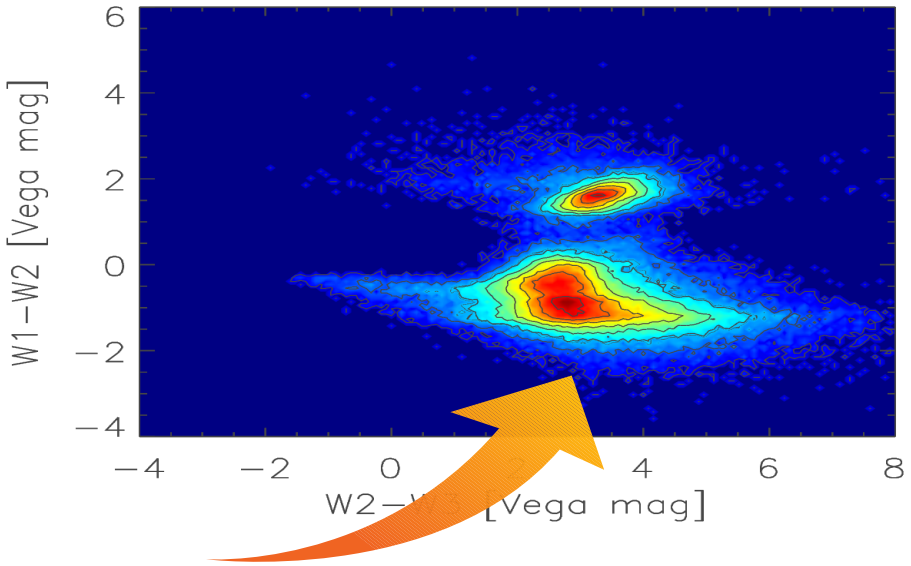
~650,000 anomalous sources

Solarz et al. 2017

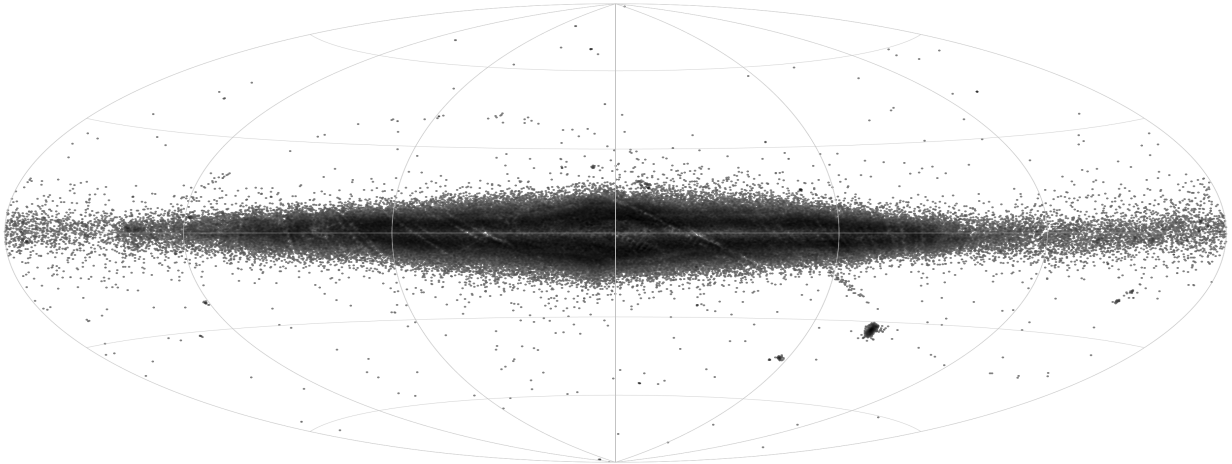
What are they?

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



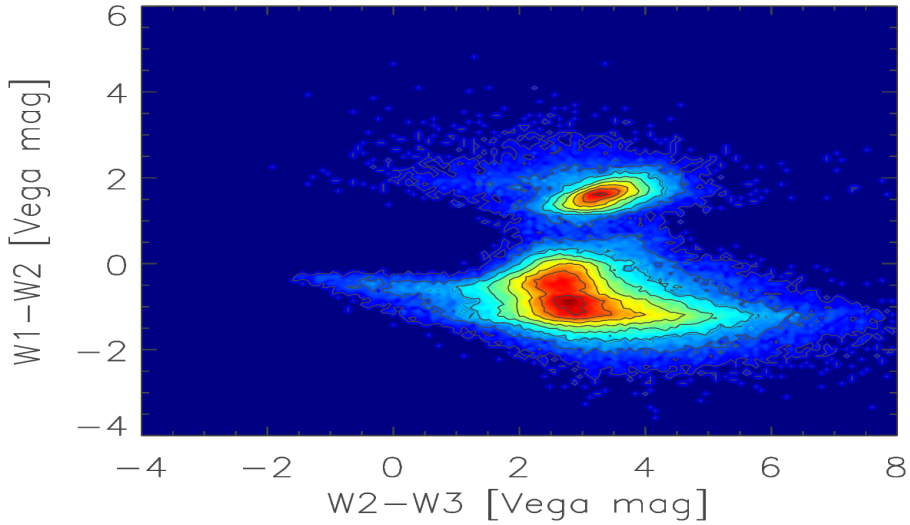
Solarz et al. 2017

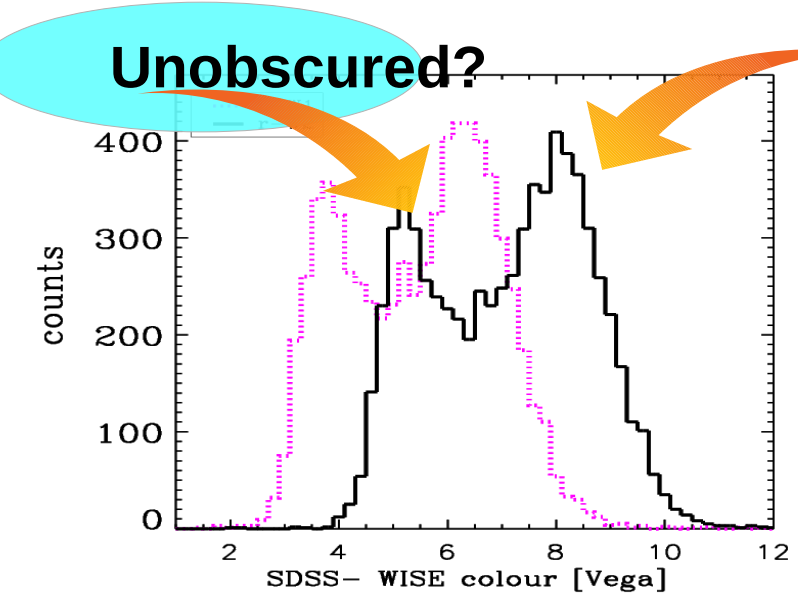


AGN candidates?

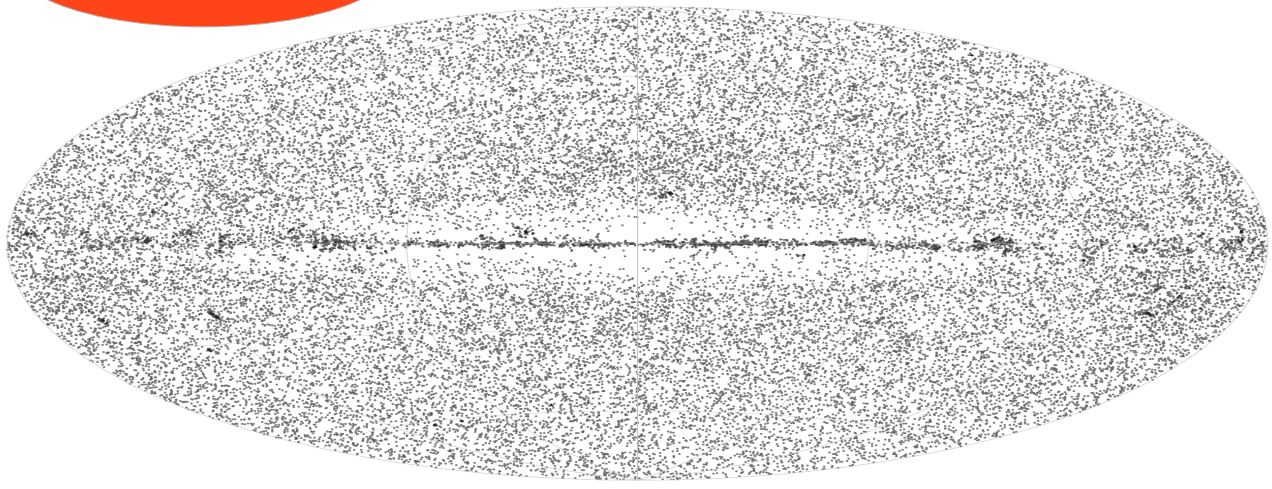
- 30,000 sources (those in Galactic Plane: still mostly blends)
- 76% undetected at other wavelengths!
- ~7 000 objects with SDSS photometry (no spectro-z), follow-up observations
  - Peculiar (dusty) QSOs
  - Low-z very dusty galaxies
  - Very dusty Galactic objects – like YSO

Solarz et al. 2020





Solarz et al. 2017



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 → 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)