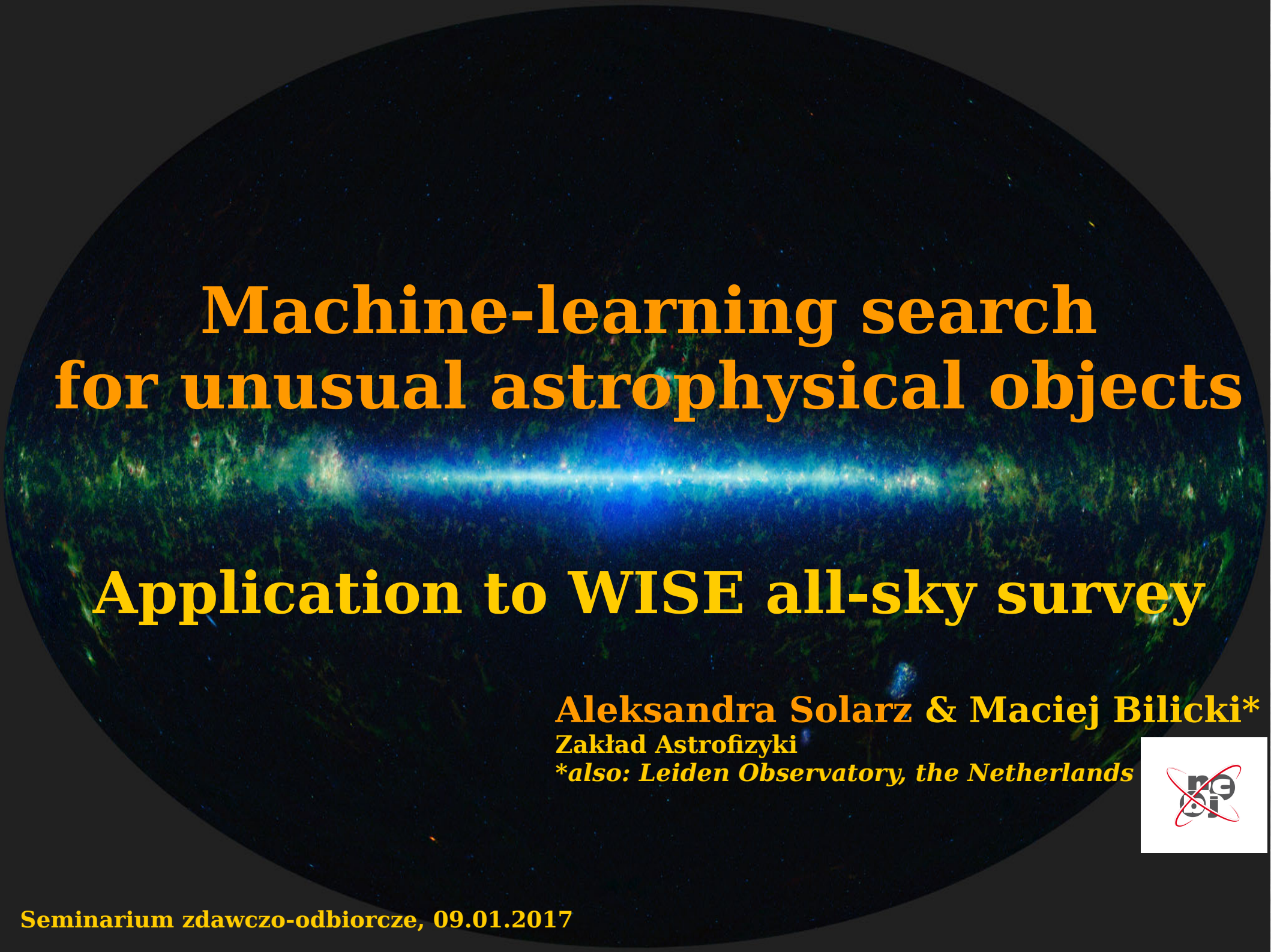


Machine-learning search for unusual astrophysical objects

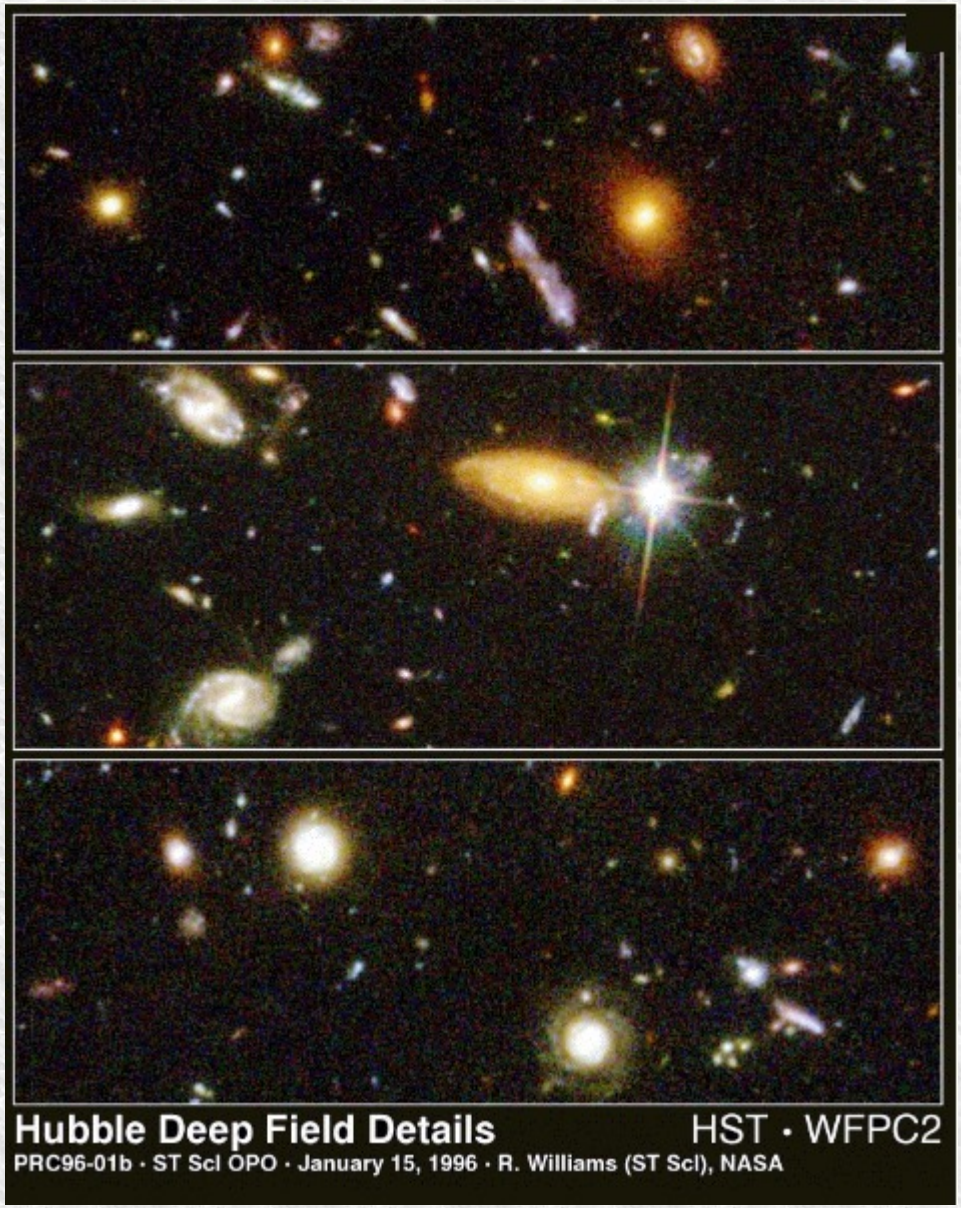
Application to WISE all-sky survey

Aleksandra Solarz & Maciej Bilicki*
Zakład Astrofizyki
also: Leiden Observatory, the Netherlands

Seminarium zdawczo-odbiorcze, 09.01.2017

Digital Sky Surveys

- As large and as deep as possible
- Large statistical samples of celestial objects
- Starting point for more detailed information about the sources
- Include rare and/or unusual objects

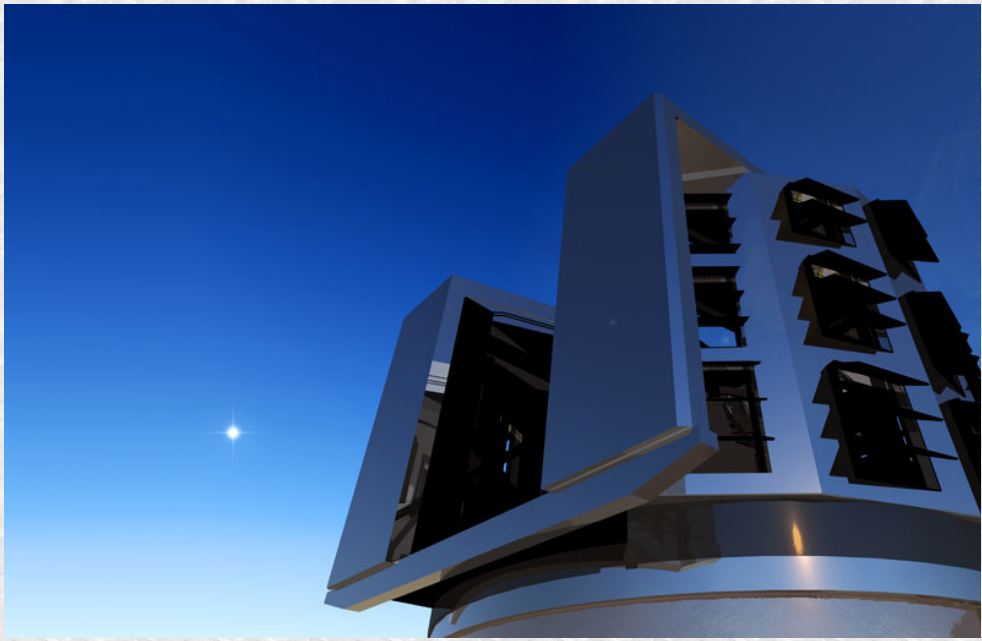


The data avalanche in modern astronomy

- SDSS: ~115 TB in total
- Zwicky Transient Facility (ZTF; start 2018)
  → 1 PB of imaging data, ~1 billion objects
- Large Synoptic Survey Telescope (LSST)
  first light ~2020, five years of operation

  → 30 TB PER NIGHT
- The Square Kilometer Array (SKA)

  → ~4.6 Zettabytes



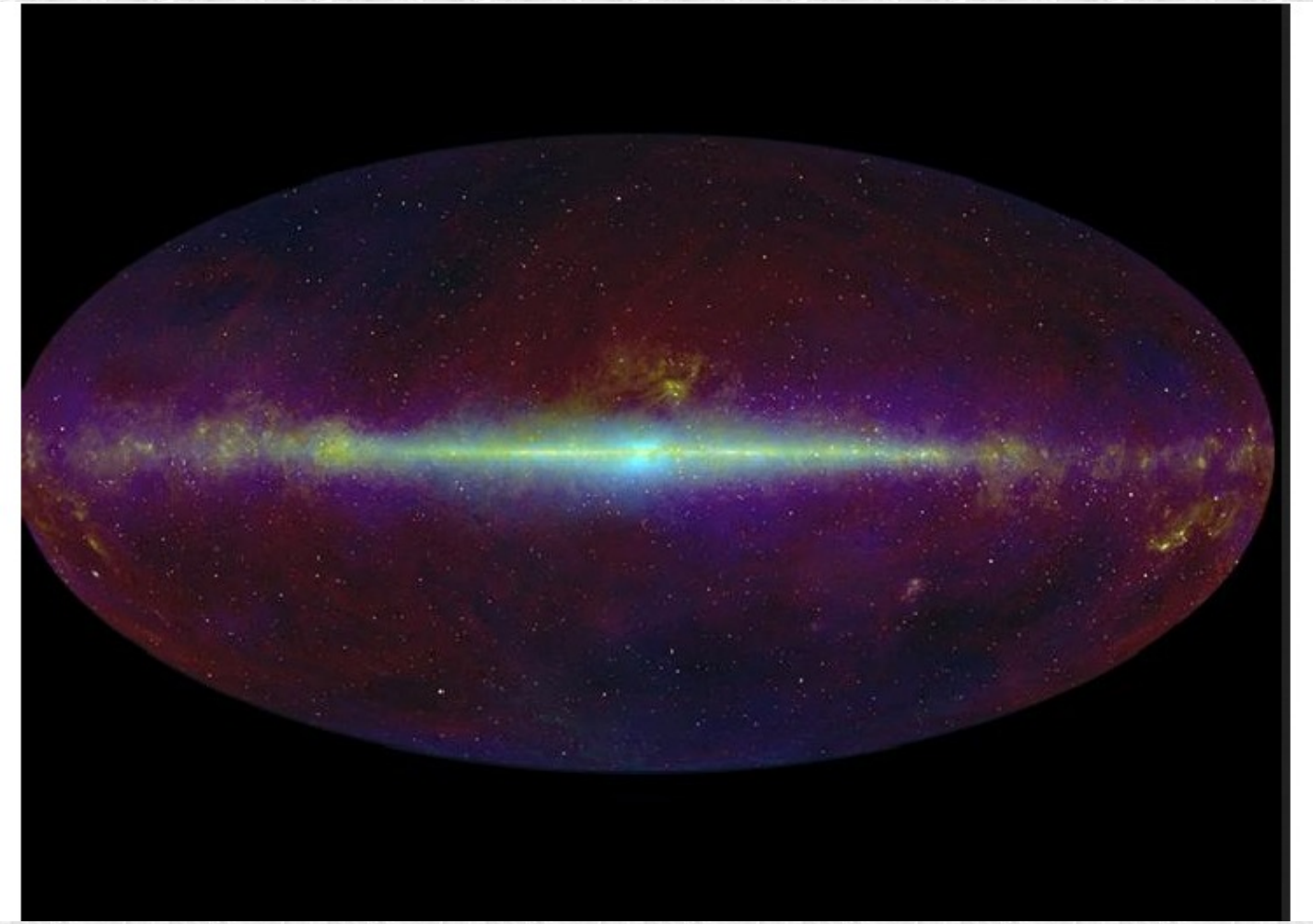
https://www.lsst.org/lsst

SKA; South Africa

Need for automated tools to detect, characterize and classify observed sources

Wide-field Infrared Survey Explorer (WISE)

- All-Sky survey in IR (4 bands, 3.6-22 μm)
- Cataloged ~750 mln sources
- Publicly available (positions, photometry in four bands, etc.)
- Low angular resolution (>6”)
- No spectroscopic information for most of the sources
- No extended vs. point source classification in the database



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

Exploration of parameter spaces

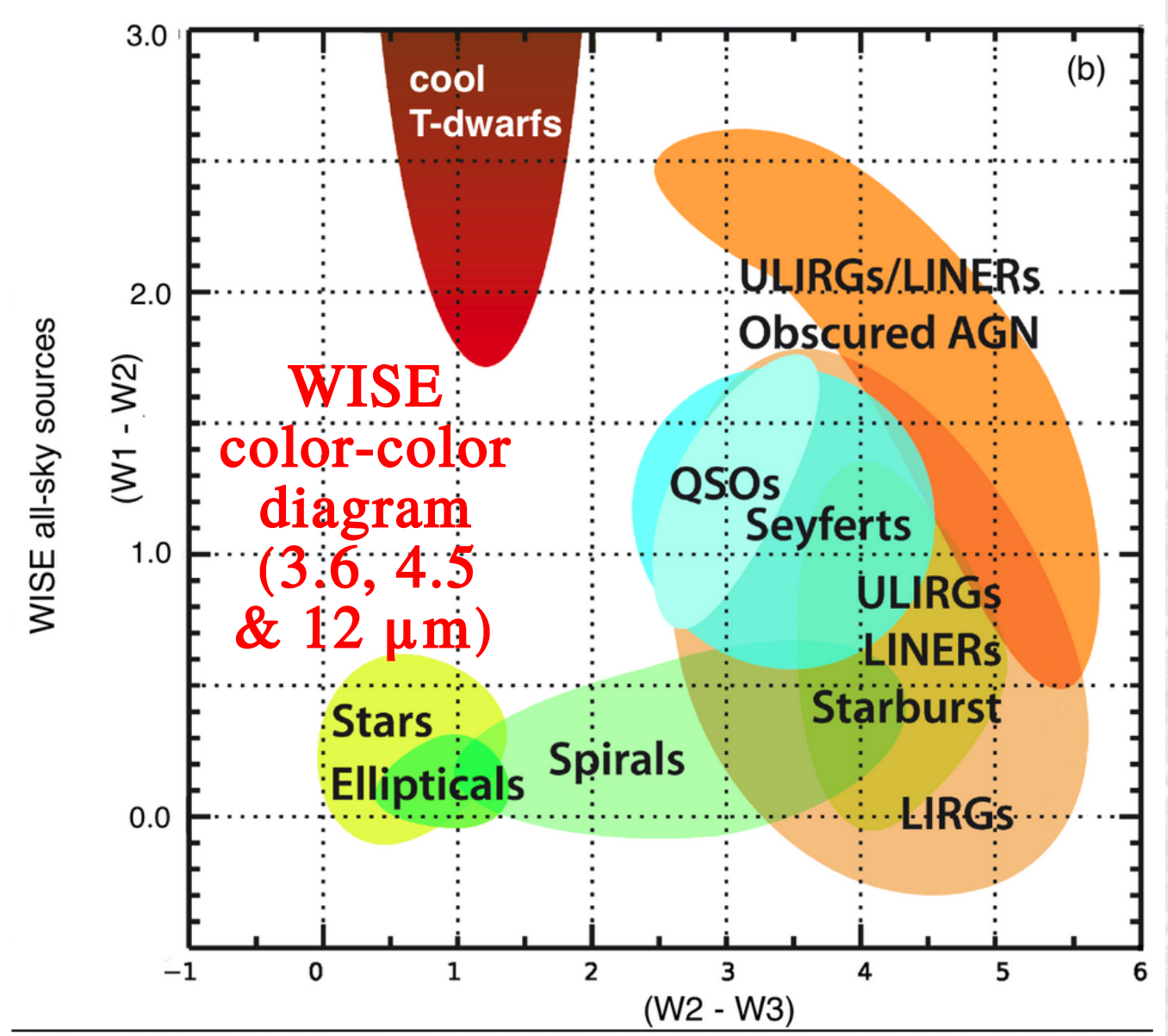
The usual approach to selection of desired sources without spectroscopy:

Color-Color diagrams

Much information is lost/unseen by human eye

→ Computer can be more precise and deal with a lot of data at once; not restricted to three dimensions

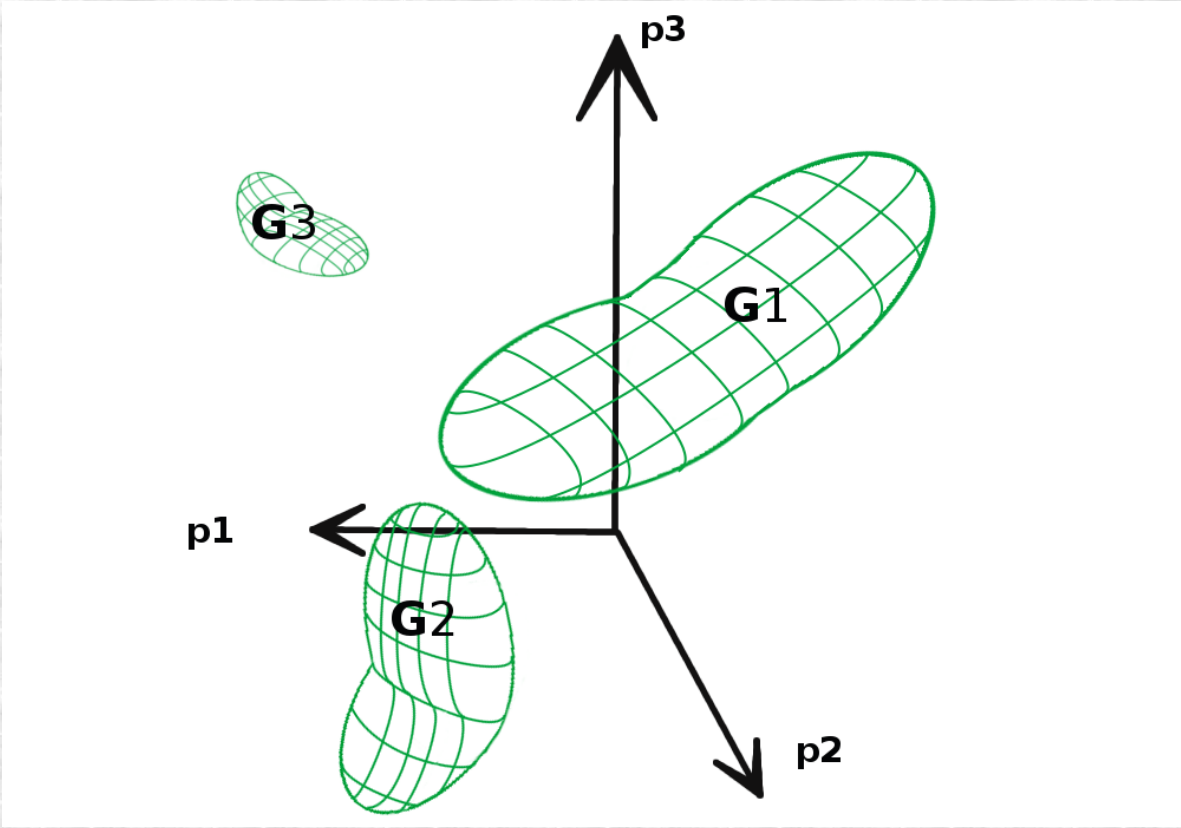
→ Machine learning!



Credit: Wright et al. 2010

Machine learning for source classification: a supervised approach

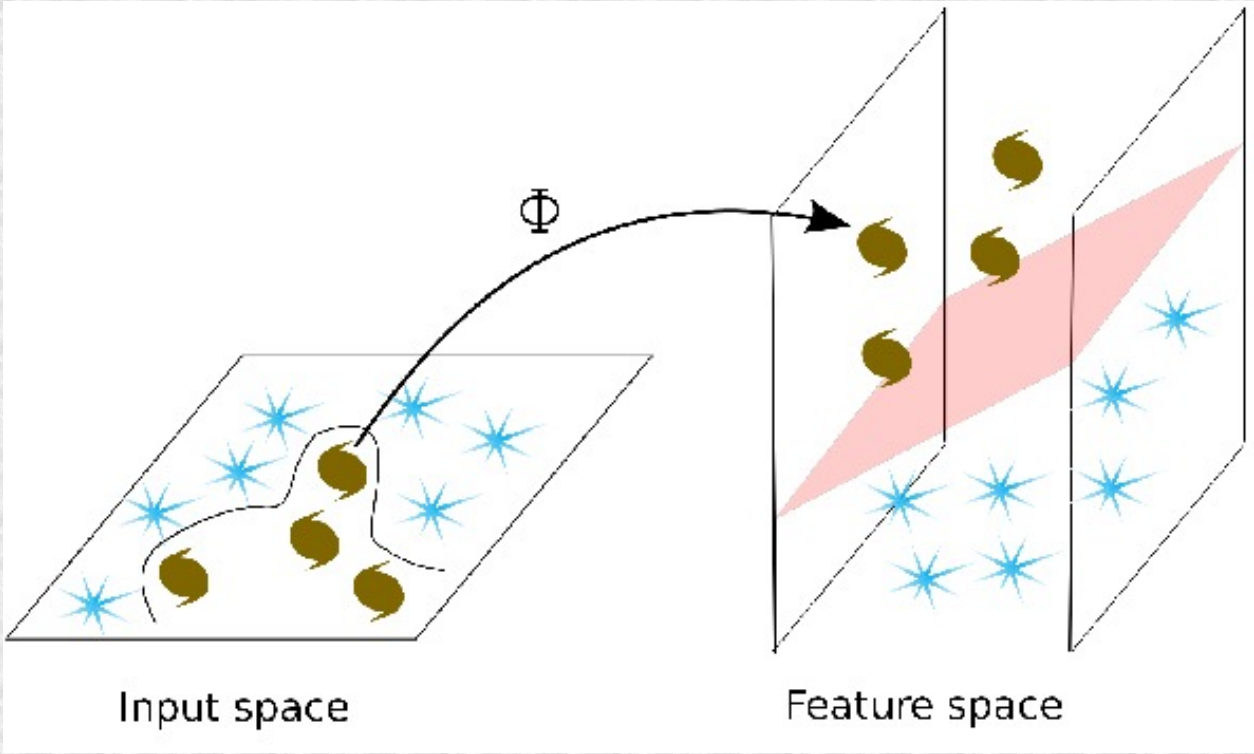
- ML algorithm learns how to recognize different types of data (G) based on training examples
- ML works in a parameter space (p) constructed from discriminating properties of the data
- If needed uses more dimensions than those of the input data
- Our approach: Support Vector Machines (SVM)



Support Vector Machines

SVM: segregate data into categories based on training examples

- Use kernel functions to map input data into higher-dimensional feature space
- Find a hyperplane separating two classes in the feature space
- Output source classes assigned based on their position relative to the boundary

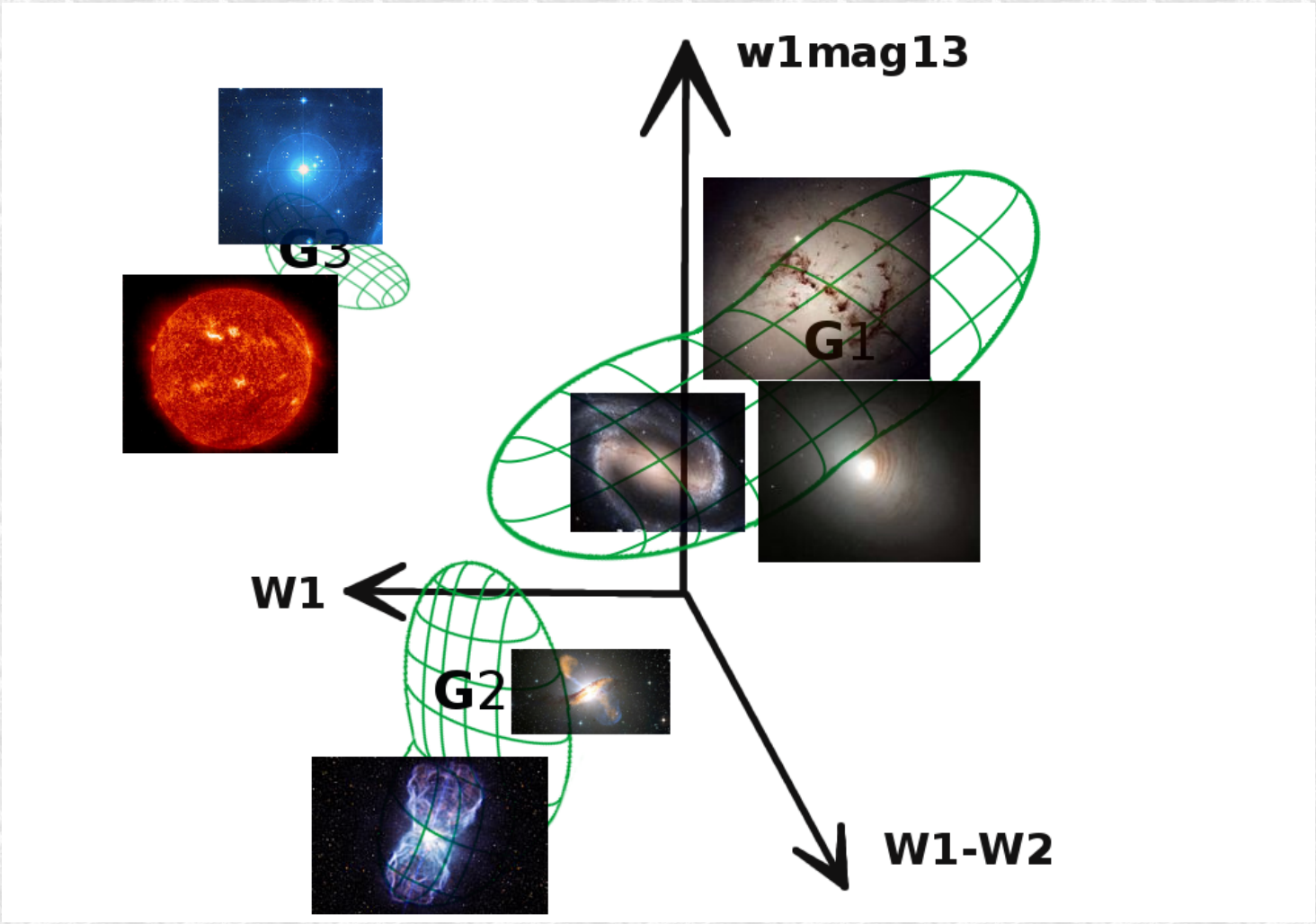


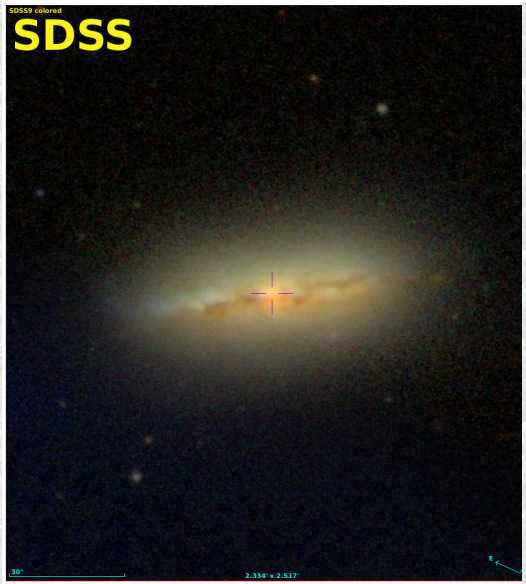
Małek, Solarz and VIPERS team, 2013

First steps towards ML source classification in WISE

What we expect ← → Training set : WISE x SDSS spectroscopic

Three classes: stars, galaxies, quasars

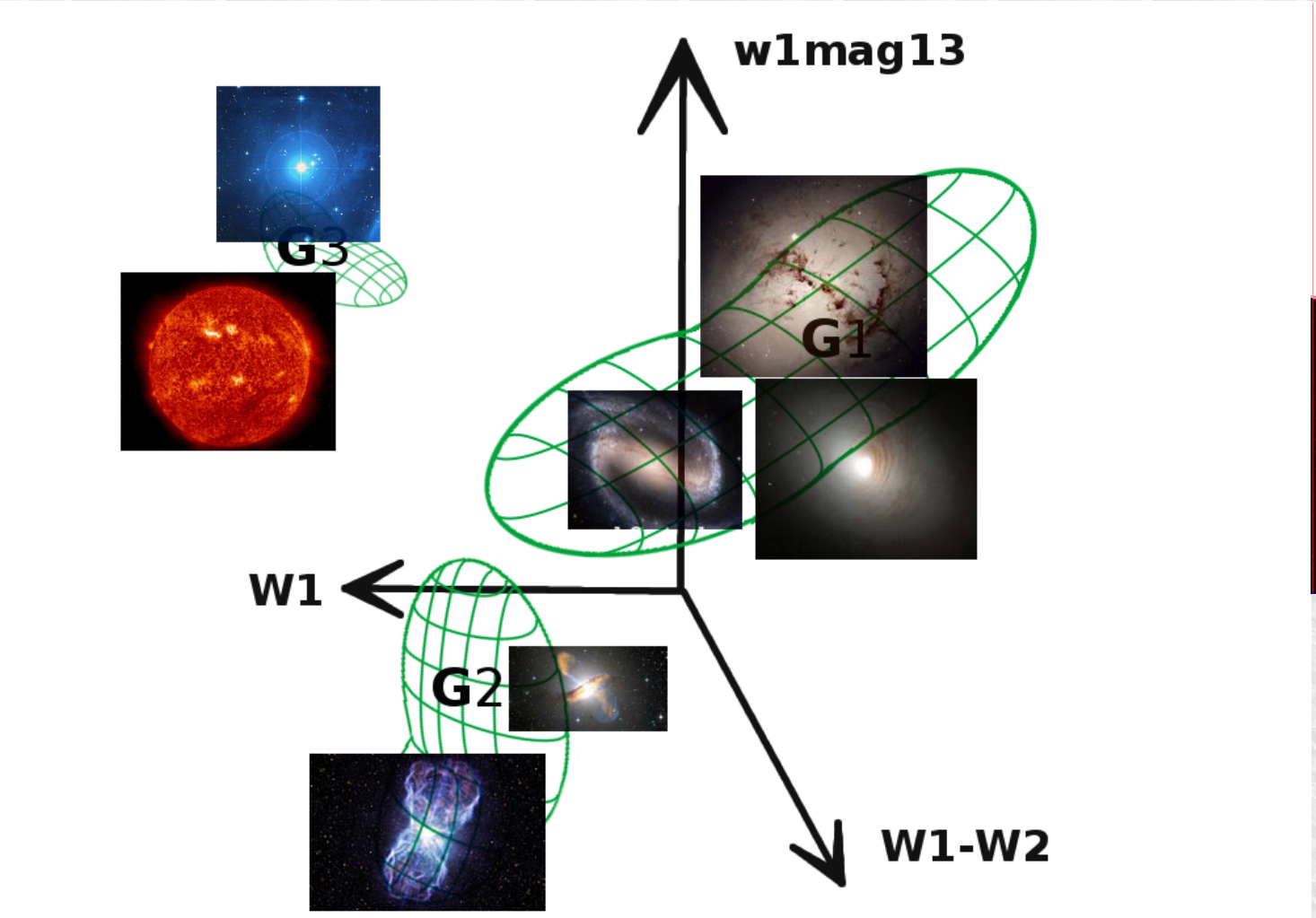


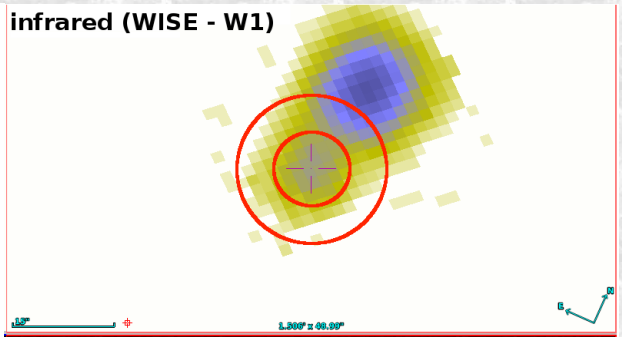


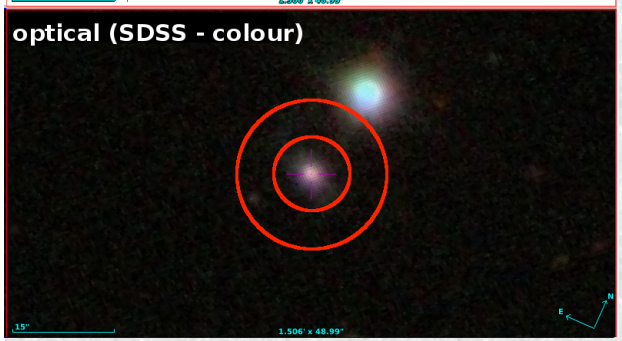
Kurcz, Bilicki, Solarz et al., 2016, A&A, arXiv:1604.04229
Krakowski, Małek, Bilicki et al., 2016, A&A, arXiv:1607.01188

First steps towards ML source classification in WISE

Where to expect it? (input parameter space)





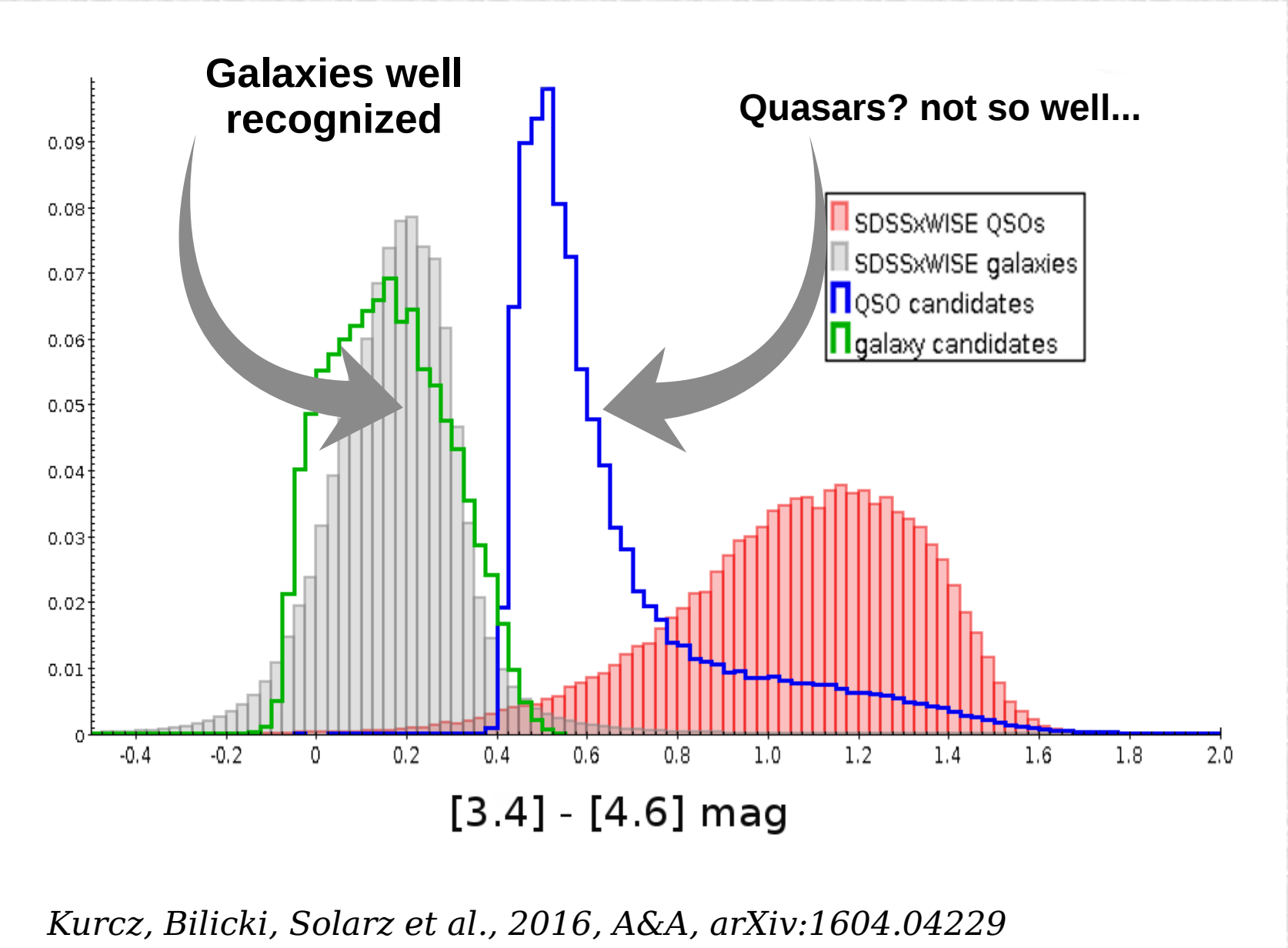


Compactness parameter

W1mag13 =
= w1mpro(5”) -w1mpro(11”)

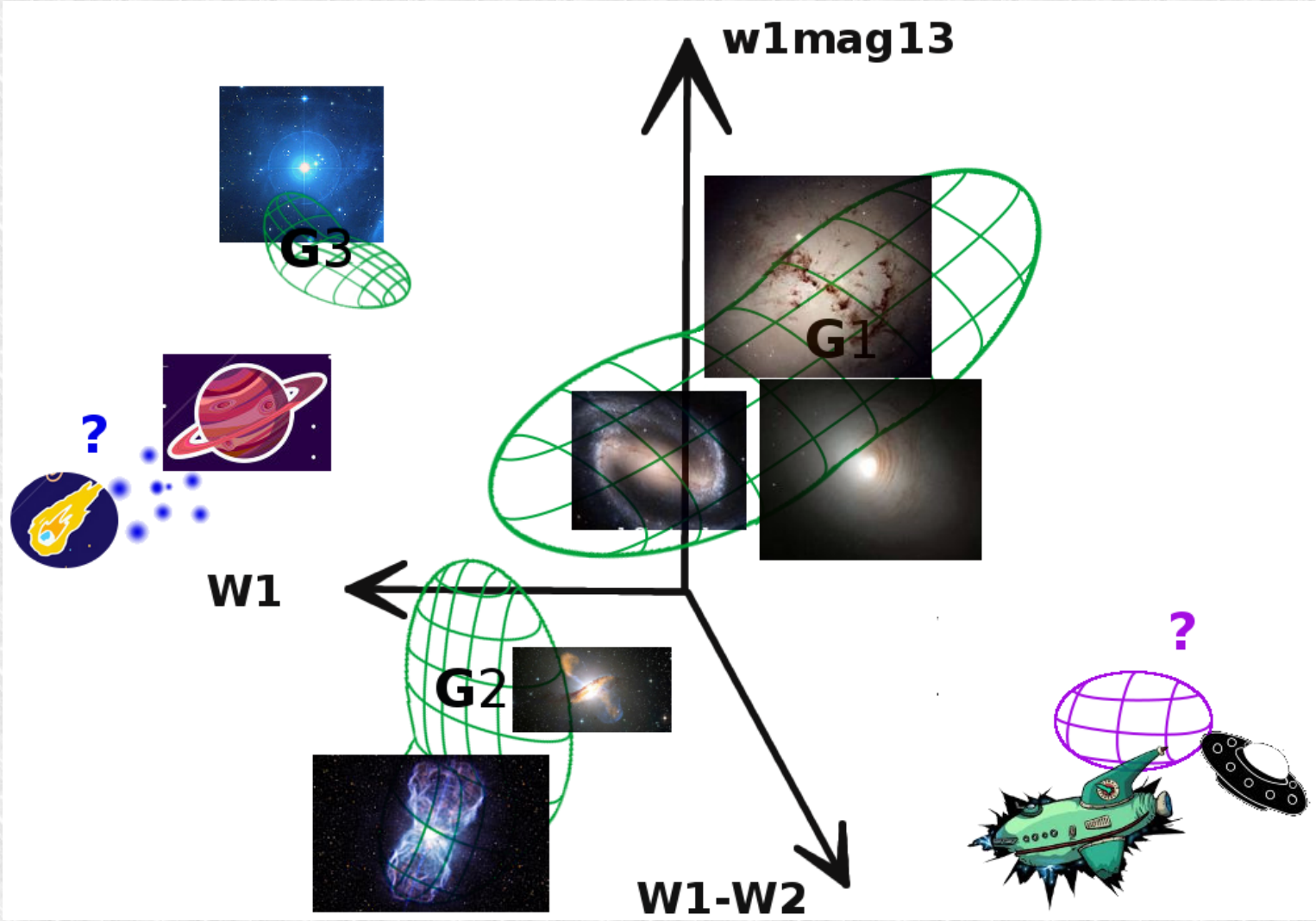
Kurcz, Bilicki, Solarz et al., 2016, A&A, arXiv:1604.04229
Krakowski, Małek, Bilicki et al., 2016, A&A, arXiv:1607.01188

First attempt at ML source classification in WISE



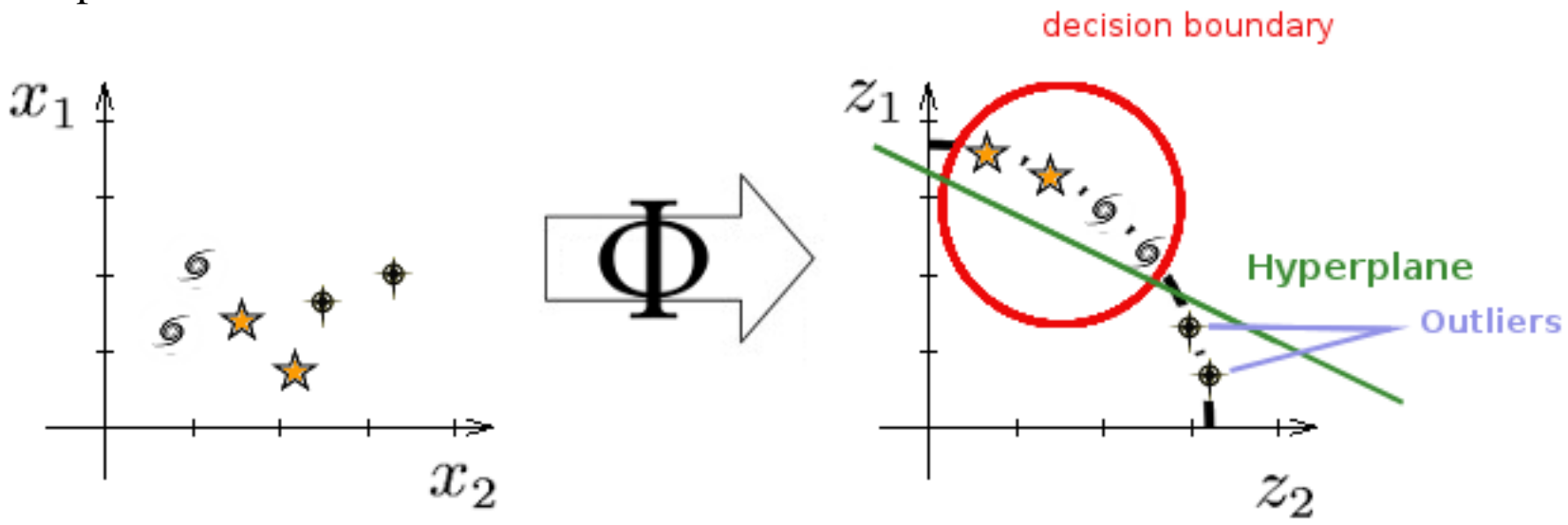
Kurcz, Bilicki, Solarz et al., 2016, A&A, arXiv:1604.04229

WISE: accounting for unknown unknowns



Novelty detection with One-Class Support Vector Machines

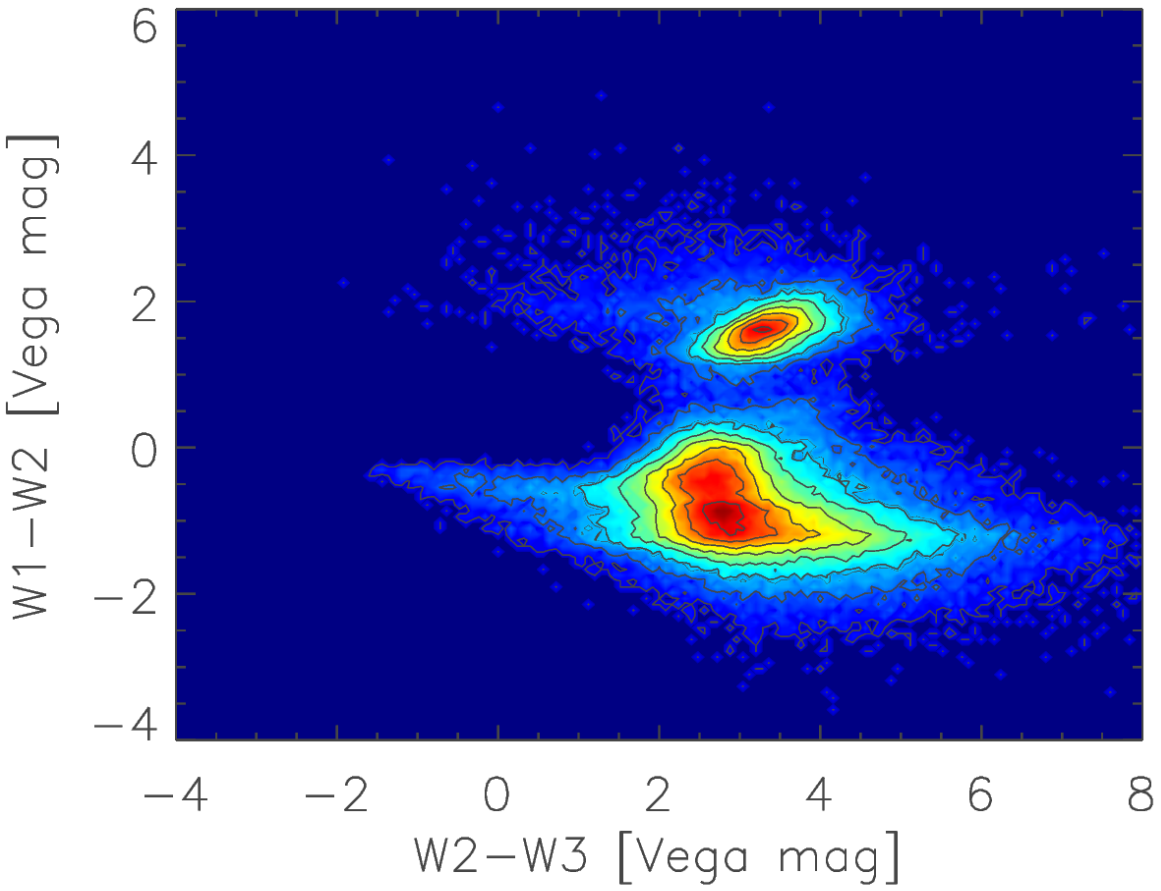
The Principle:



- Create one ‘known’ class (mix of WISE x SDSS galaxies, stars, quasars)
- Map input data to a higher-dimensional parameter space
- Define a hypersurface encapsulating the expected sources
- Anything with ‘unknown’ patterns falls outside the hypersurface → Novelties

Solarz, Bilicki, et al., 2017, A&A, arXiv:1706.06389

OCSVM applied to WISE data



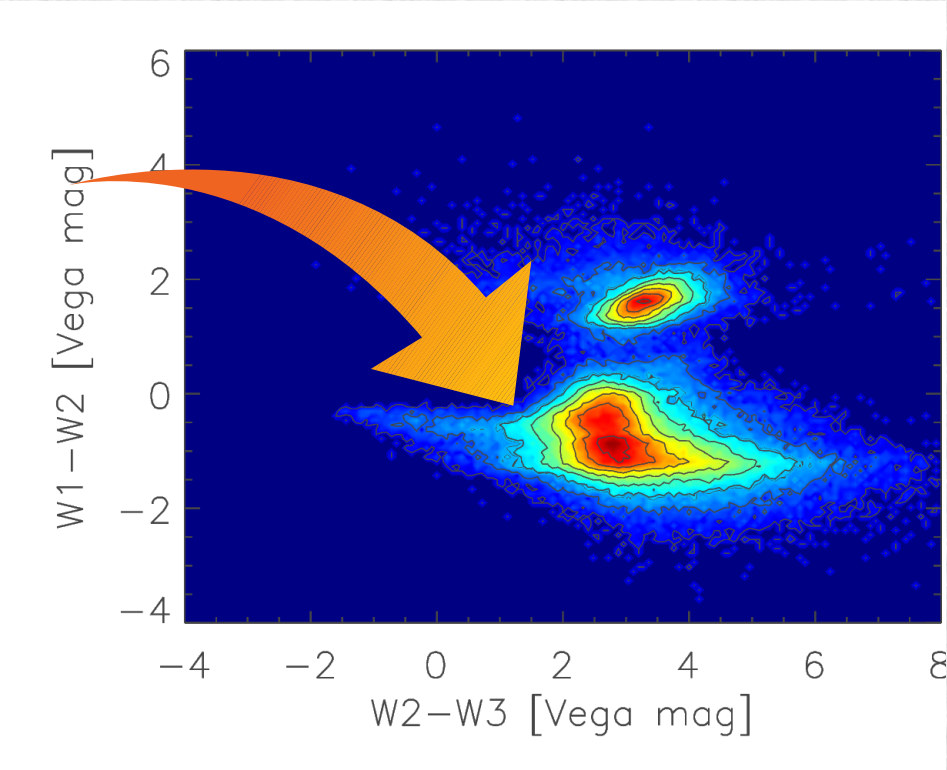
~650,000 "anomalous" sources

What is their nature?

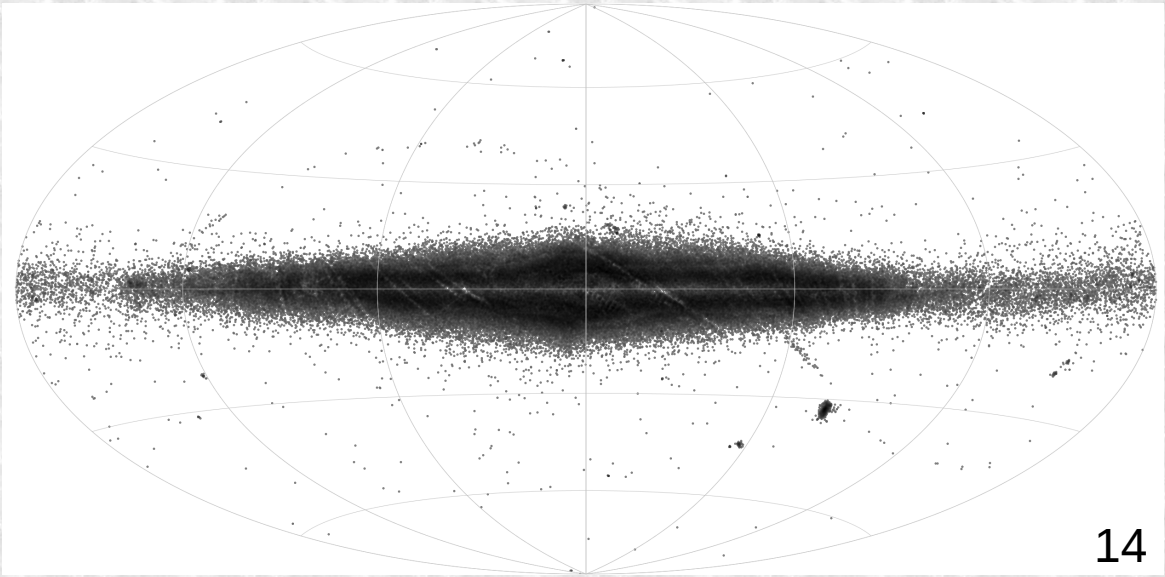
Solarz, Bilicki, et al., 2017, A&A, arXiv:1706.06389

Most novelties are “spurious”…

- Unphysical W1-W2 ~ -1 ; 80%
- Low WISE resolution (6”) → blends in crowded fields
- Spitzer GLIMPSE: double sources

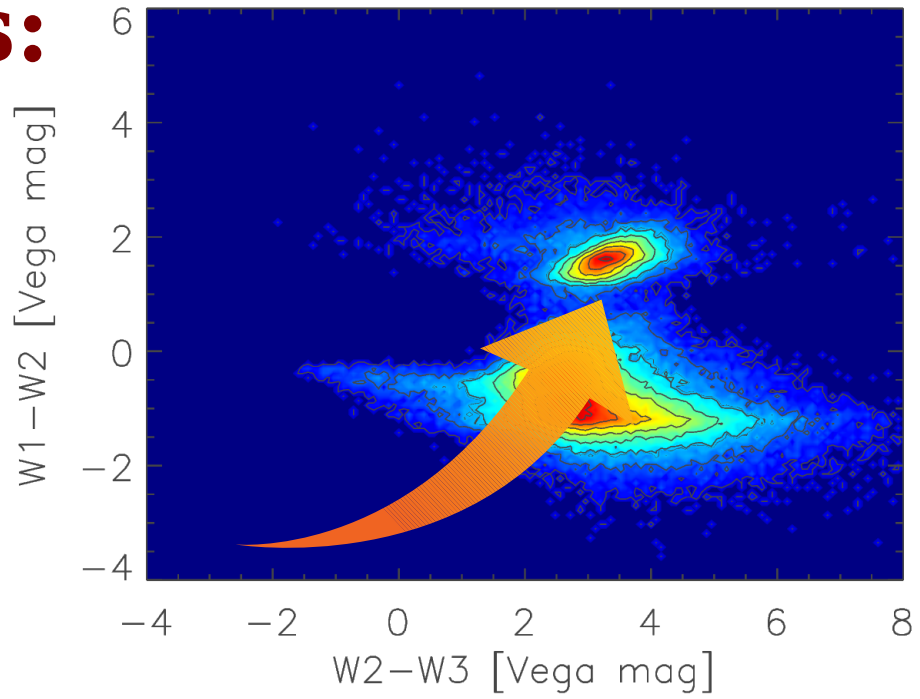


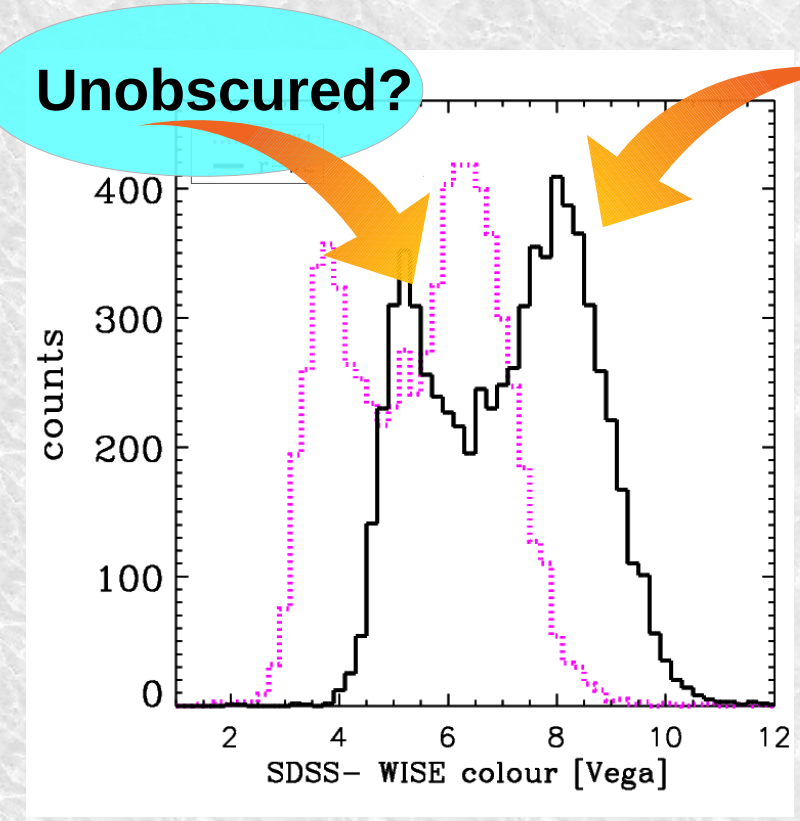
Solarz, Bilicki et al. 2017



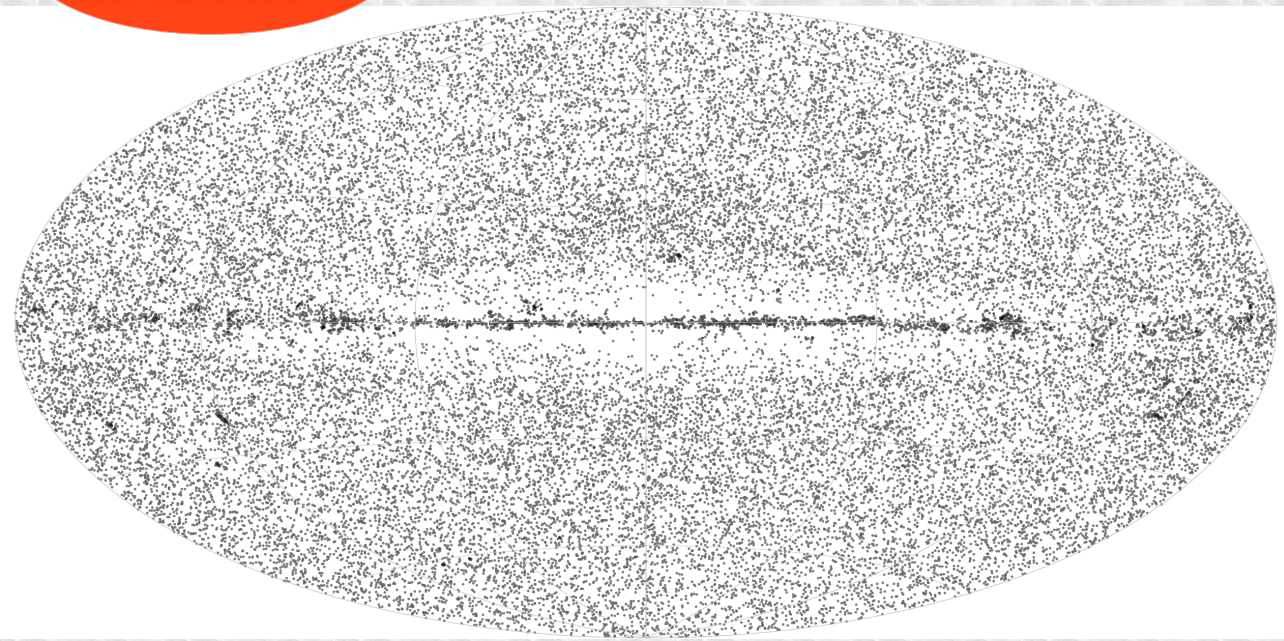
...but not only artifacts: AGN candidates

- ~30,000 sources all over the sky
- Warm dust / PAH emission – colors consistent with active galactic nuclei (AGN)
- 75% undetected at other wavelengths!
- ~7,000 objects with SDSS photometry





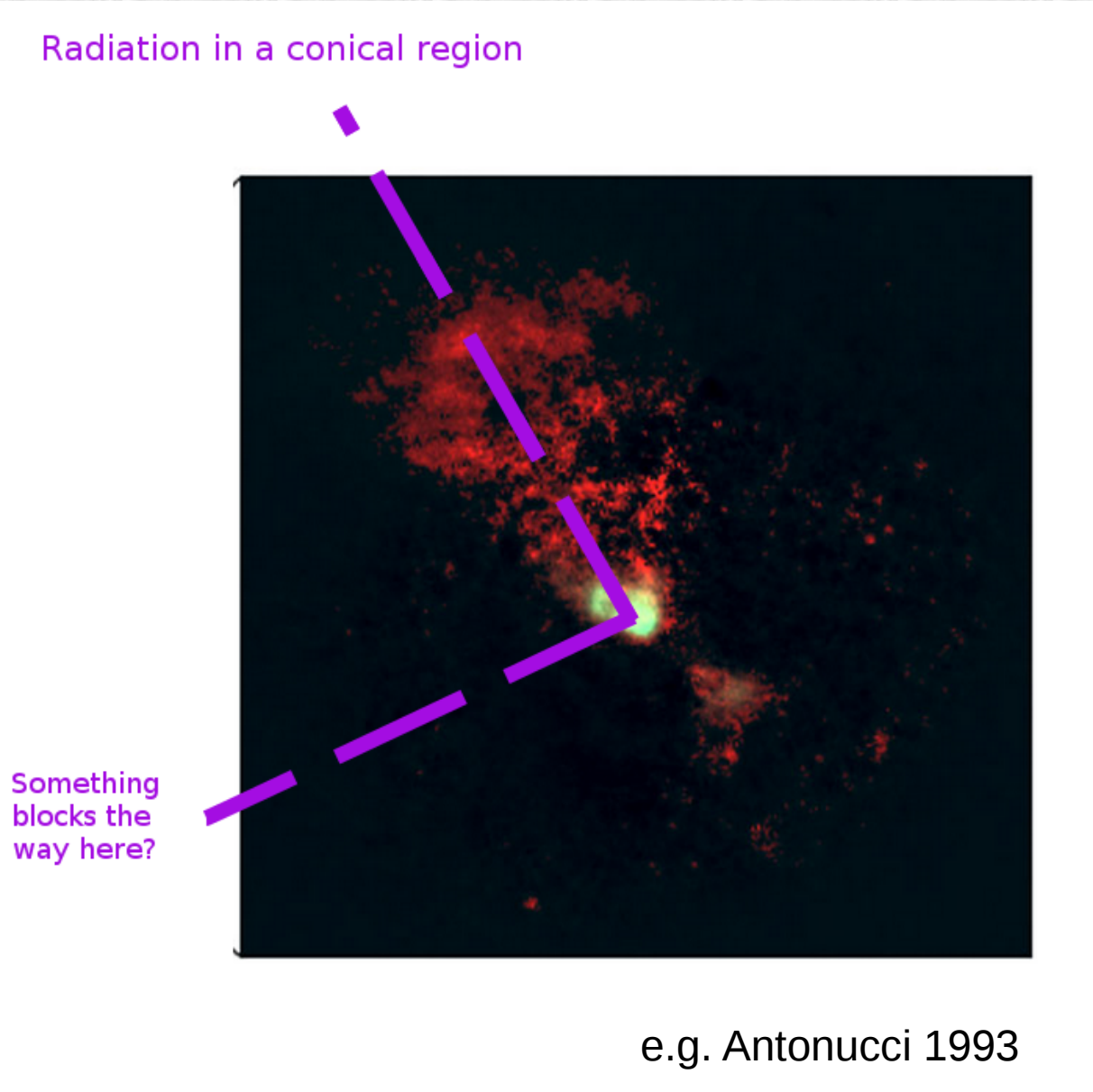
Solarz, Bilicki et al. 2017



Obscured/unobscured AGNs: unified model

Different classes of active galactic nuclei (“engines” of quasars)

→ different orientations to the observer’s line of sight of intrinsically similar systems



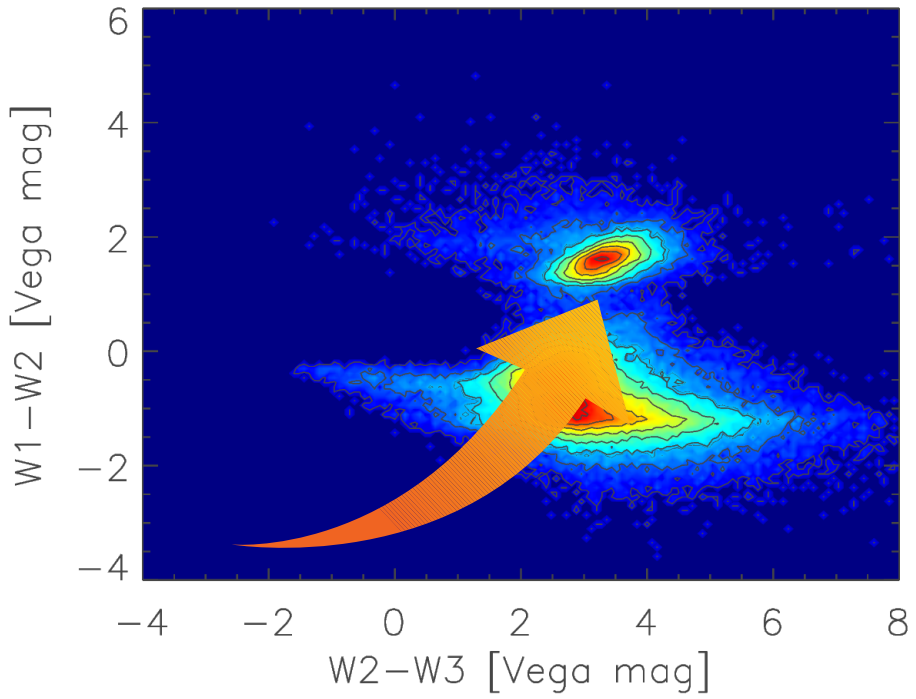
e.g. Antonucci 1993

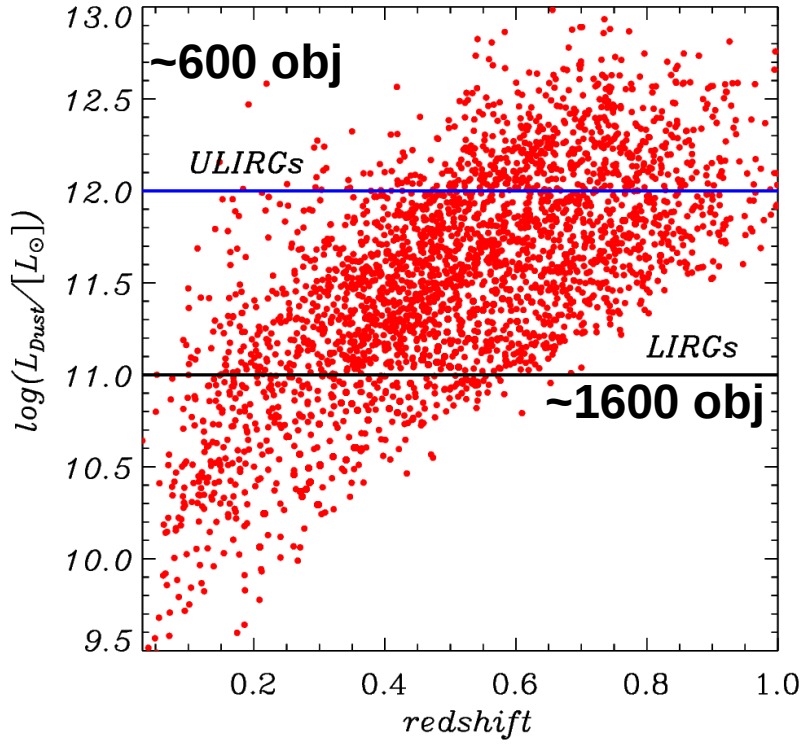
WISE novelties: AGN candidates, (U)LIRGS

- Some sources have SDSS + WISE photometry
- Spectral Energy Distribution with CIGALE
- **RESULTS:**

  85% (Ultra)Luminous Infrared Galaxies

  LIRGs (L_{dust} >10^{11} L_{sun})
  and ULIRGs (L_{dust} >10^{12} L_{Sun})



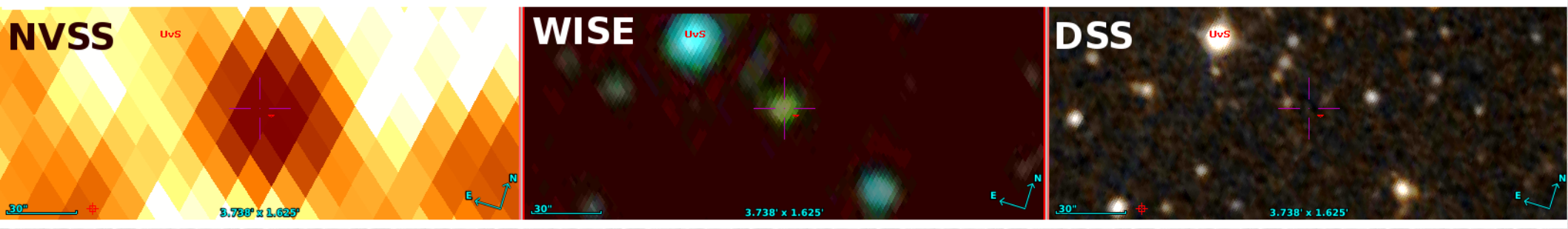


https://cigale.lam.fr/

WISE novelties: looking for more information

Very few counterparts of our novelties in other available catalogs

- ~1,300 in NVSS (radio, 20 cm)



- Far-IR: ~ 65 in Herschel (250 um); 13 in AKARI AllSky Survey FIS (60 μm – 160 μm)
- Soft X-rays:~ 38 in XMM-Newton + Chandra;
- Hard X-rays & Gamma-rays: Fermi/Swift, Integral/IBIS: 0

To verify the astrophysical nature:

- Follow-up optical photometry needed (ongoing and planned surveys)
- Spectroscopy would be best (proposals: EFOSC2/SOFI @ La Silla observatory: time allocated!!!, ; GEMINI-FLAMINGOS2: time allocated!!!; ALMA next in line)