Machine Astronomers to Discover the Unexpected in Astronomical Surveys

Ray Norris, Western Sydney University &
CSIRO Astronomy & Space Science,

WESTERN SYDNEY UNIVERSITY

CSIRO

Overview

1. The “Machine Learning in Astronomy” Collaboration
2. The process of astronomical discovery
3. WTF: discovering the unexpected in radio surveys

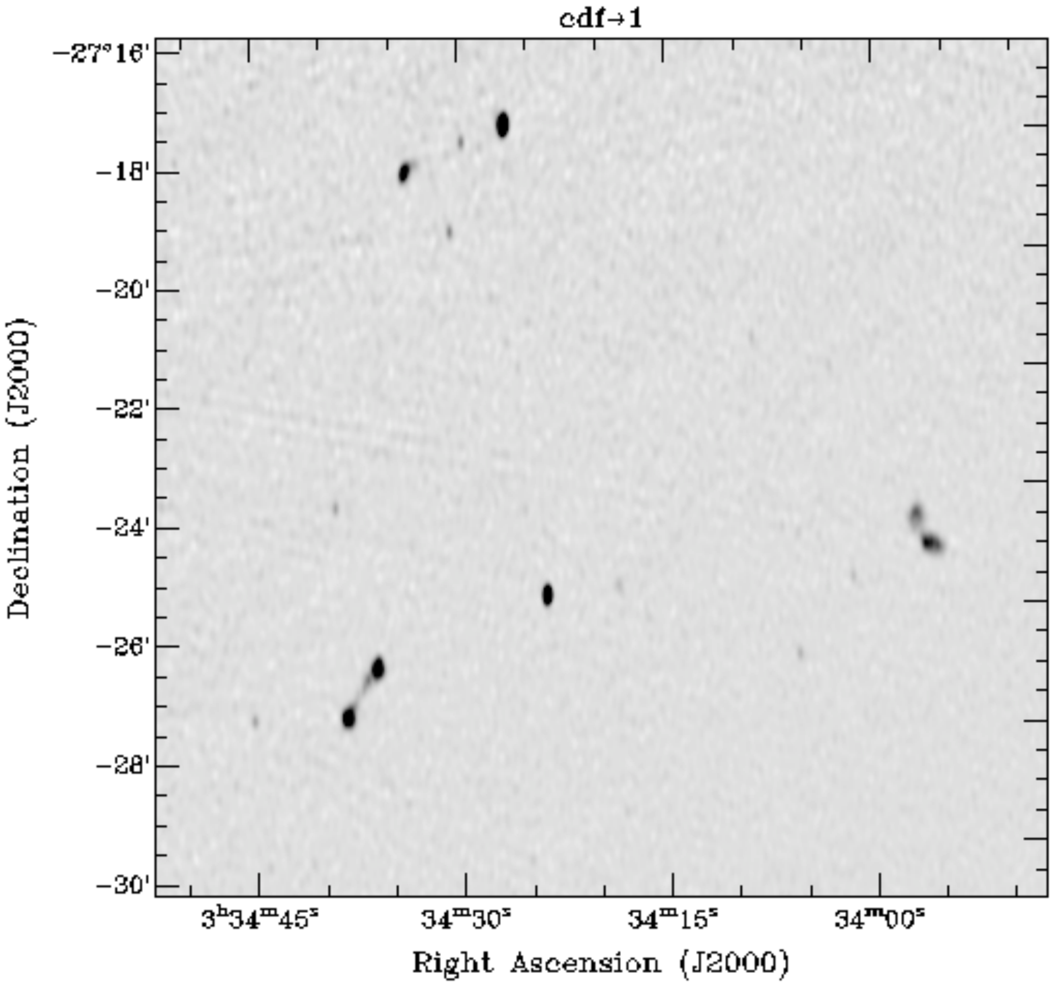
“Machine learning in astronomy” collaboration

- An open collaboration – see mlprojects.pbworks.com
- Participants from several Australian Universities
  + CSIRO Data61, EMU project, etc.
- Regular informal zoom research meetings take place every Thursday at 11.00 - 12.00 AEST, 09.00-10.00 AWST, on https://uws.zoom.us/j/2319669070
- Provide data sets and training sets for experimenting
  - E.g. ATLAS DR3, synthetic data sets, etc
- Other resources

WSU “Machine learning in astronomy” projects

Focus on:

- Source cross-identification and self-identification
- Source classification
- Photometric redshifts
- Building data sets
- WTF



WSU "Machine learning in astronomy" projects

- Ray Norris+: building training/test sets (placed in public domain)
- Laurence Park+: radio source morphology classification techniques
- Gary Segal (PhD)+: Anomaly detection for WTF
- Pero Manojlovic (PhD)+: Finding bent-tail galaxies with CNN
- Nathan Kayani (M. Res): Evaluating ML techniques for classifying radio source morphology
- Kieran Luken (M. Res) et al.: photometric redshifts
- Nick Ralph (M. Res)+: Using ML to use ASKAP monitoring data to improve images
- Katherine James (vac. stud.)+: radio source classification with CNN

- See also
  - Baron+Poznanski 2016, The weirdest SDSS galaxies: results from an outlier detection algorithm, **arXiv:1611.07526**
  - Aniyan+Thorat 2017 Classifying Radio Galaxies with Convolutional Neural Network **arXiv:1705.03413**

We have found that much of our ML work is limited by the lack of good training sets

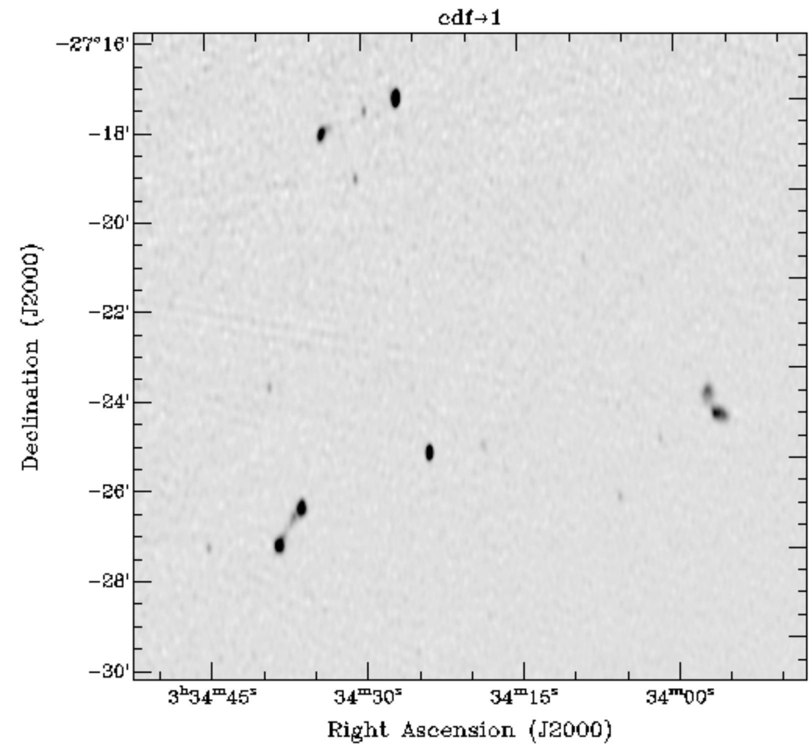
We are making the following training sets publicly available on mlprojects.pbworks.com:

Available now/very soon:

- **ATLAS DR3 enhanced training set (Swan+Norris)**
  4500 simple radio sources, 4500 complex
  with IR cross-IDs, labels, etc
- **Synthetic training set**
  20,000 simple, 20,000 complex
  (tricky – beware of Russian tanks!)

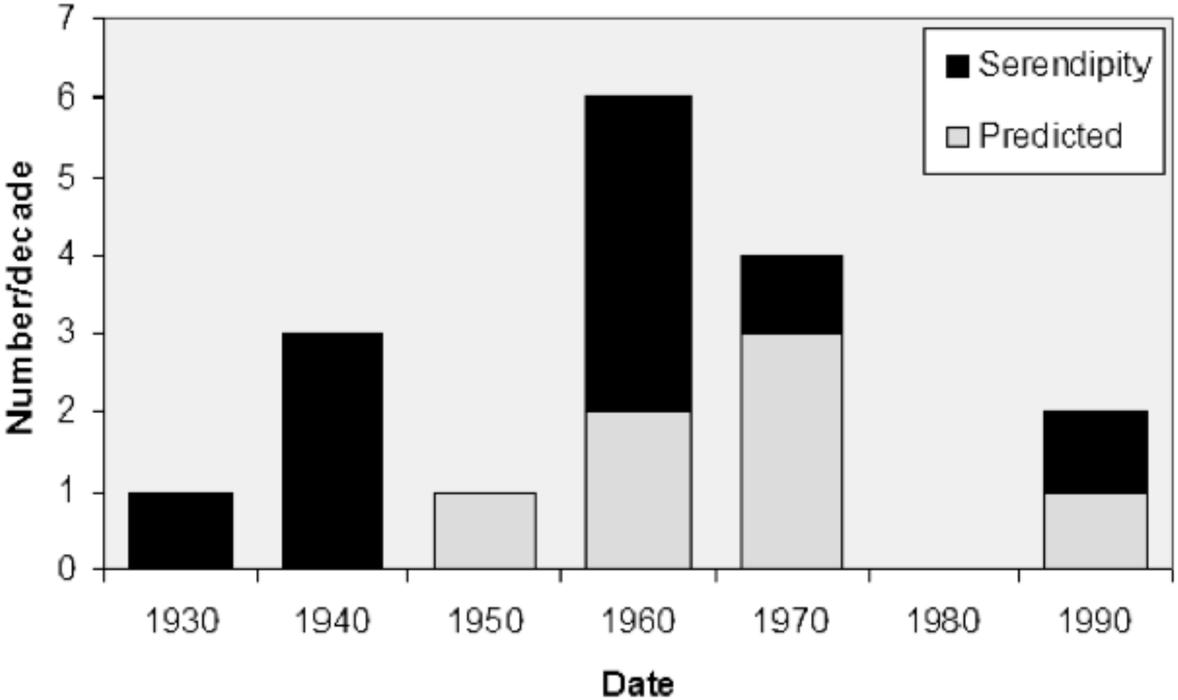
Available in future:

- Bent tails
- WTF/SETI (tricky – limited by our imaginatic
- Redshifts (SEDs + spec-z's)



The Process of Astronomical Discovery

What fraction of discoveries in astronomy were "Popperian"?



(b) Predicted v Serendipity

Serendipity: 10
Predicted: 7

From Ekers (2009) PoS(sps5)007

See also:

- *Harwit(1981), Cosmic Discovery*
- *Kellermann(2009) PoS(sps5), 44*
- *Wilkinson et al.(2004), New Astr. Rev., 48, 1551 45*
- *Wilkinson(2007) the Modern Radio Universe, 144*
- *Wilkinson(2015) (AASKA14), 65*

Discoveries with HST

Project	Key project	Planned?	Nat. Geo. top ten?	Highly cited?	Nobel prize?
Use Cepheids to improve value of H0	✔	✔	✔	✔	
study intergalactic medium with uv spectroscopy	✔	✔			
Medium-deep survey	✔	✔			
Image quasar host galaxies		✔	✔		
Measure SMBH masses		✔	✔		
Exoplanet atmospheres		✔	✔		
Planetary Nebulae		✔	✔		
Discover Dark Energy			✔	✔	✔
Comet Shoemaker-Levy			✔		
Deep fields (HDF, HDFS, UDF, FF, etc)			✔	✔	
Proplyds in Orion			✔		
GRB Hosts			✔		

from Norris arXiv:1611.05570

Discoveries with HST (see e.g. Lallo: *arXiv:1203.0002)*

Project	Key	Planned?	Nat.	Highly	Nobel prize?
Use Cepheids to					
study intergalac uv spectroscopy					
Medium-deep s					
Image quasar ho					
Measure SMBH					
Exoplanet atmos					
Planetary Nebul					
Discover Dark Er					✔
Comet Shoemak					
Deep fields (HDF					
Proplyds in Orio…					
GRB Hosts			✔		

Summary:

Of the "top ten" HST discoveries:

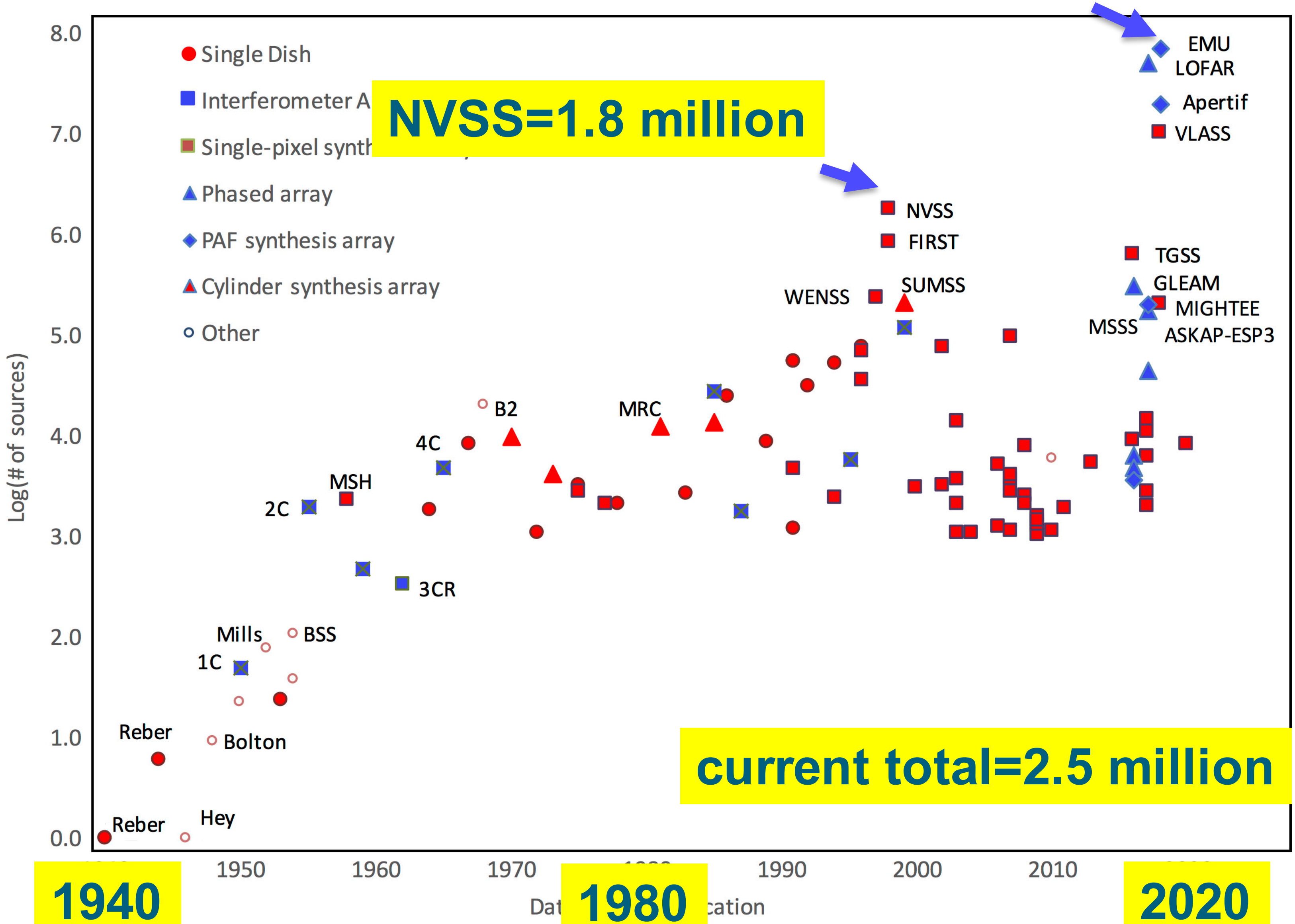
- **1 was a key project**
- **4 were planned by astronomers but were not key projects**
- **5 were totally unexpected (e.g. dark energy)**

CSIRO

How to make unexpected discoveries with next-generation surveys (e.g ASKAP-EMU)?

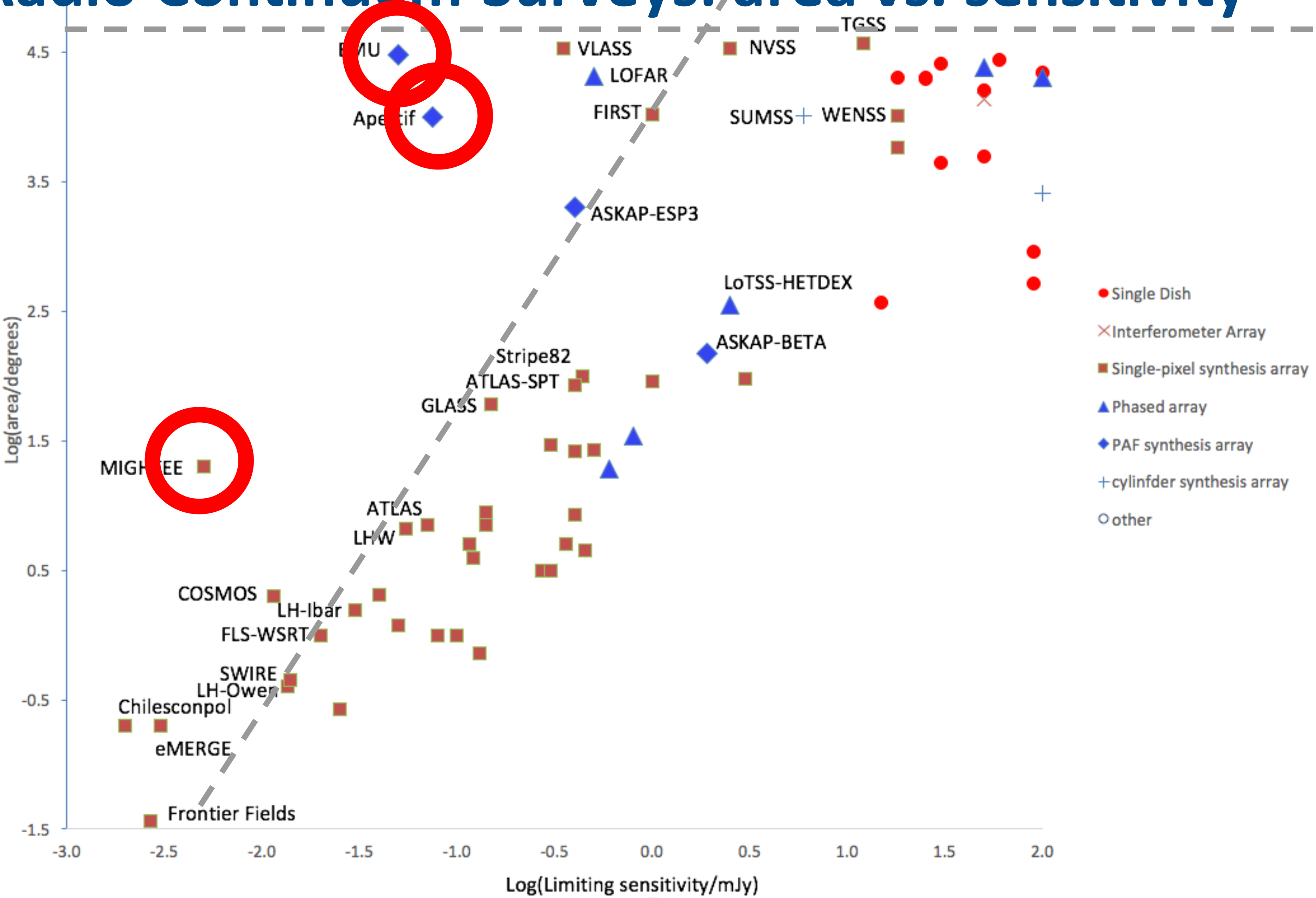
Size of radio continuum surveys over time

ASKAP Radio Continuum survey: EMU = 70 million



From Norris (2017) Nature Astronomy, in press.

Radio Continuum Surveys: area vs. sensitivity



From Norris (2017) Nature Astronomy, in press.

Can ASKAP Discover the Unexpected?

- Data volumes are huge – cannot sift by eye
- Instrument is complex – no single individual will be familiar with all possible artifacts
- ASKAP will be superb at answering well-defined questions (the "known unknowns")
- Humans won't be able to find the "unknown unknowns"
- Can we mine data for the unexpected, by rejecting the expected?

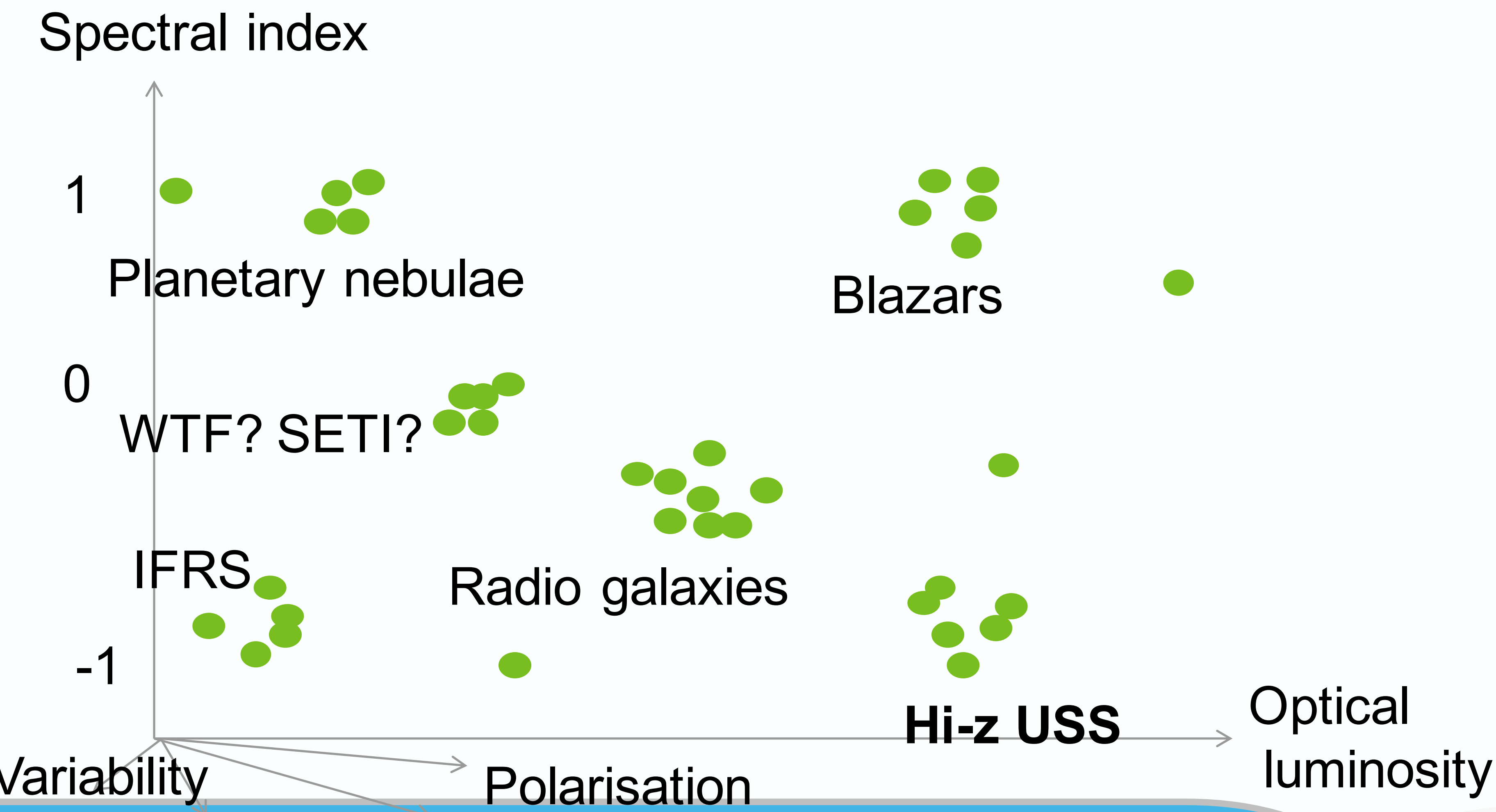
If not, ASKAP will not reach its full potential
i.e. it will not deliver value for money

mining radio survey data for the unexpected

WTF?

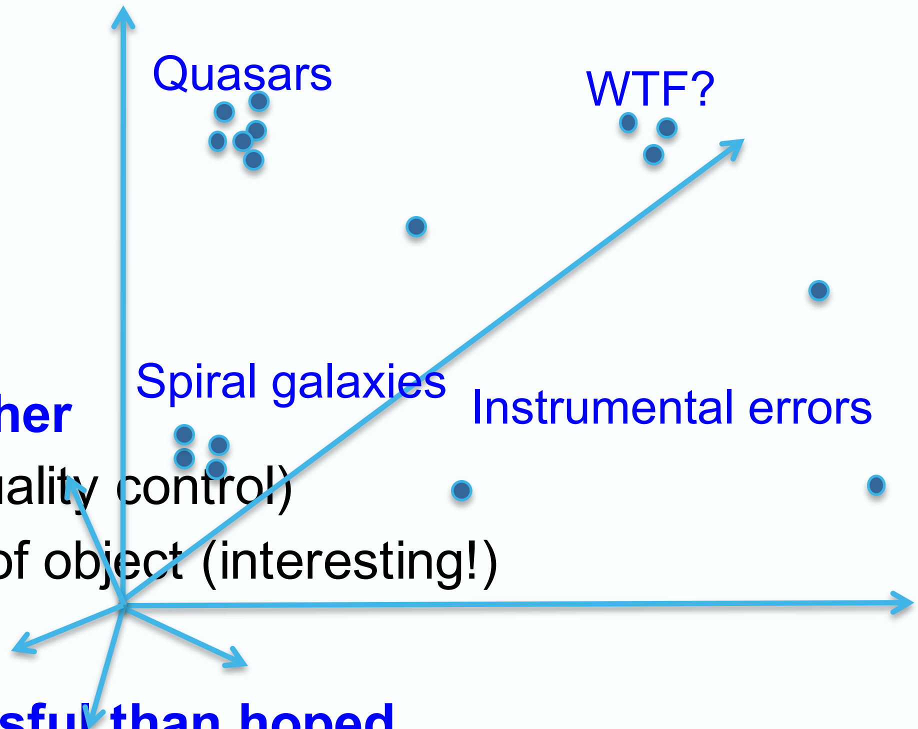
WTF = Widefield ouTlier Finder

An example of what a WTF/SETI detection might look like



Mining large data sets for the unexpected

WTF will work by searching the n-dimensional (large n) phase space of observables, using techniques (both supervised and unsupervised) such as



- kNN (k-nearest-neighbours)
- Neural nets/deep learning
- Self-organised maps
- Support vector machine
- Random forest

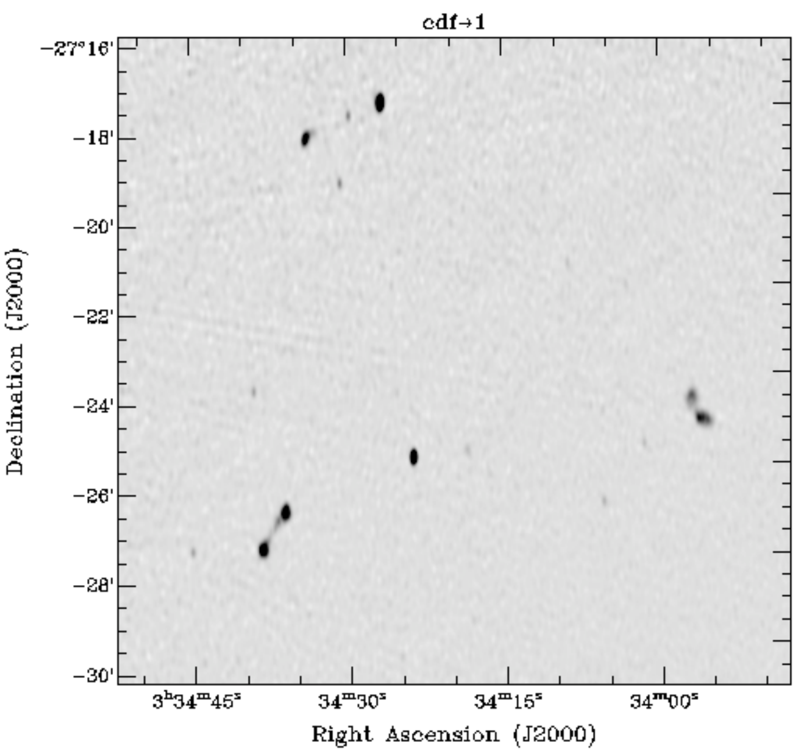
Identified objects/regions will be either

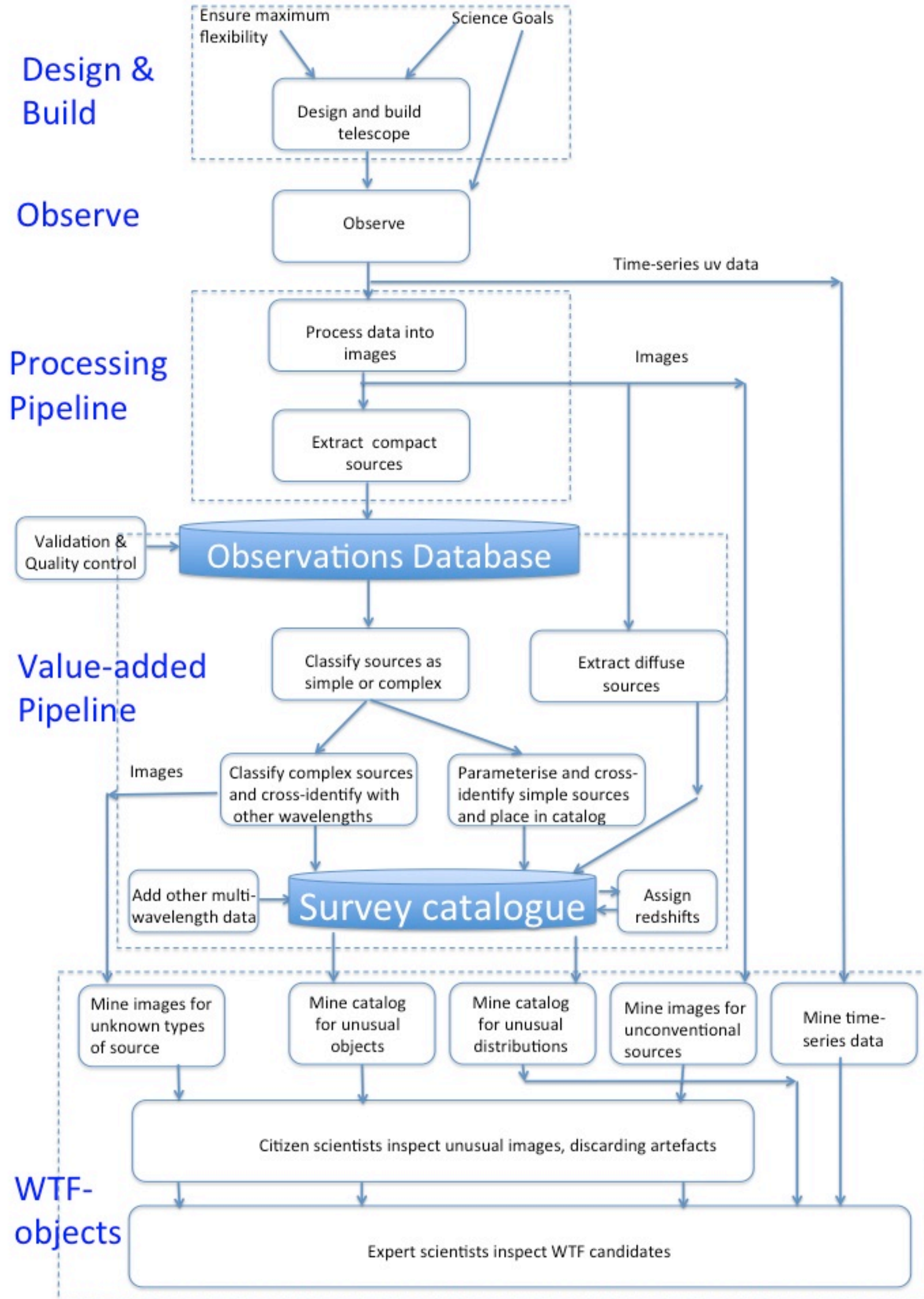
- processing artifacts (important for quality control)
- statistical outliers of known classes of object (interesting!)
- New classes of object (WTF)

Earlier public challenge less successful than hoped

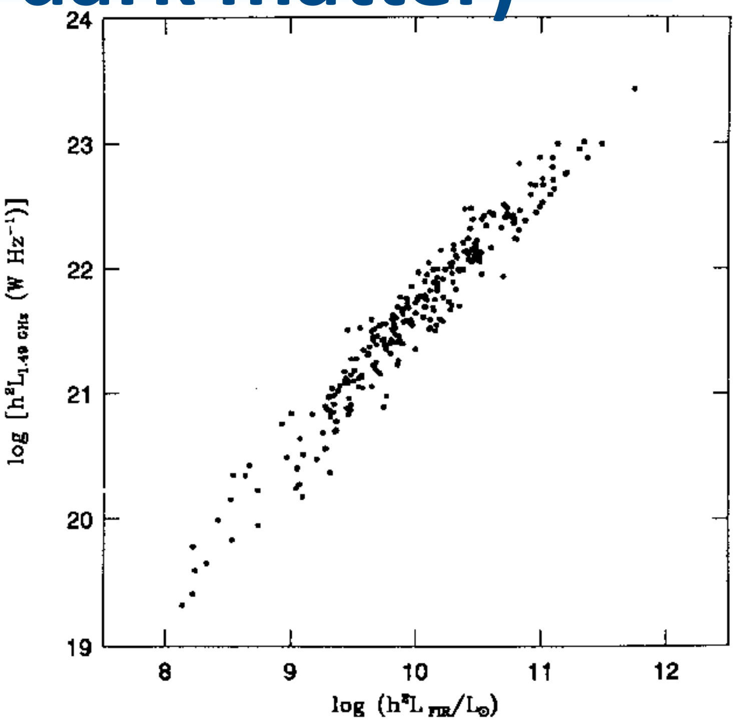
- Partly because of difficulty of providing a good WTF training set (e.g. face)
- But attracted collaborators from computer science community

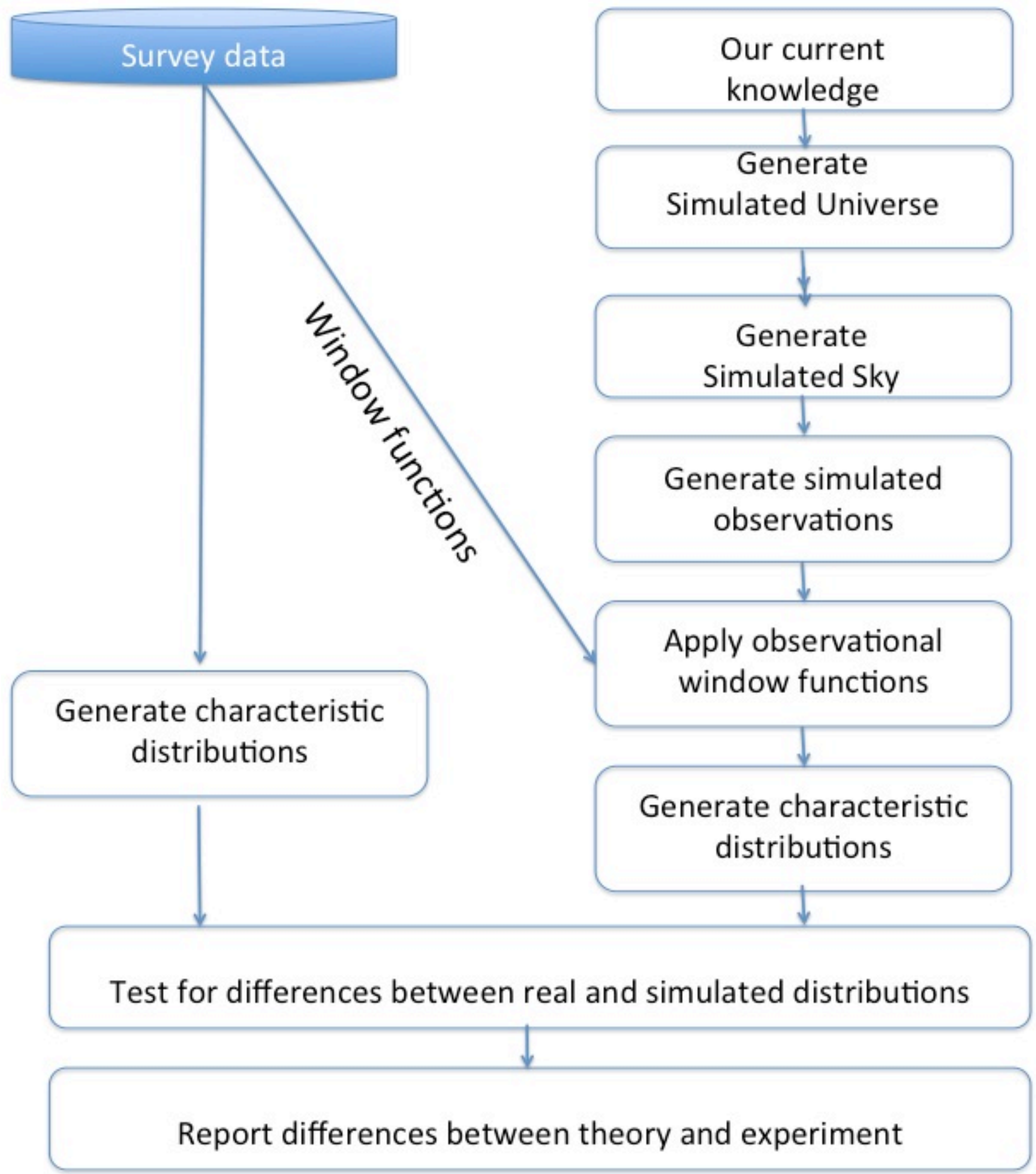
Type 1 discoveries: Unexpected objects (e.g. pulsars, quasars) Simple anomaly detection – right?



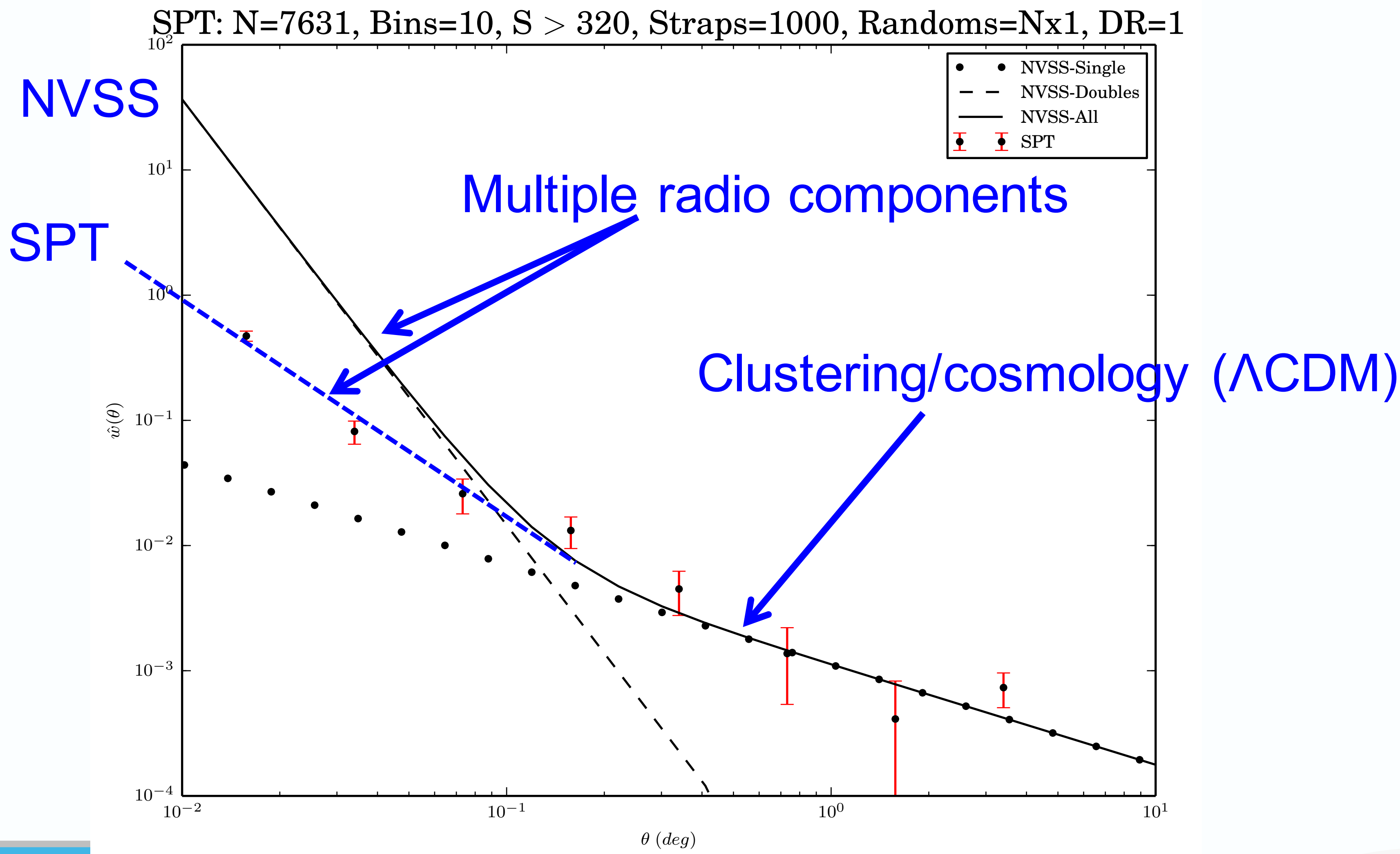


Type 2 discoveries: unexpected phenomena (e.g. Hubble expansion, dark energy, dark matter)

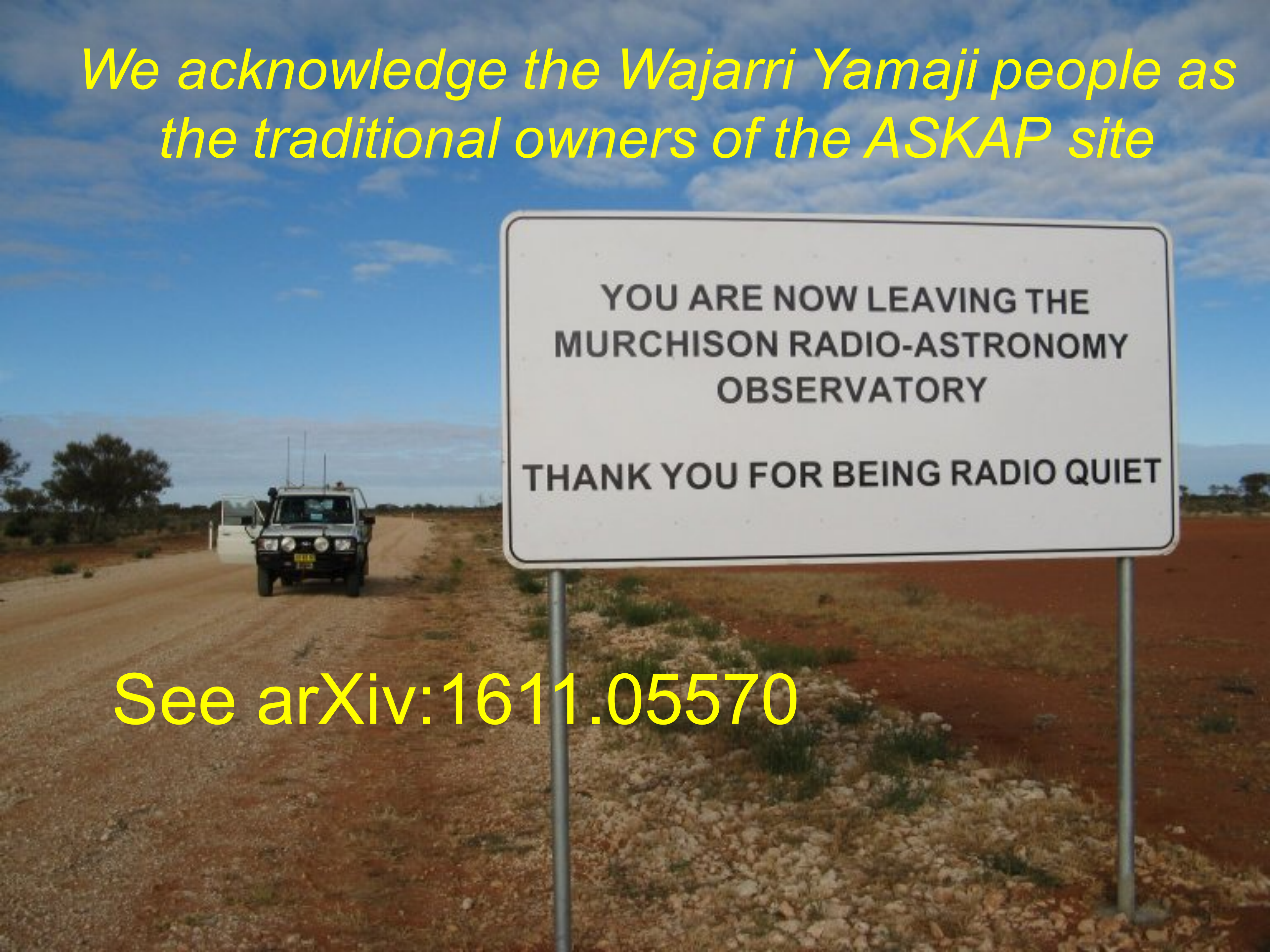




E.g. Glen Rees PhD thesis
Spatial autocorrelation function on ATLAS-SPT



We acknowledge the Wajarri Yamaji people as the traditional owners of the ASKAP site



See arXiv:1611.05570