LSST & Lessons from Gaia

N W Evans (Cambridge)

14th May 2019, Cardiff

Lessons from Gaia

Lessons from Gaia

Be Big & Well-Organised

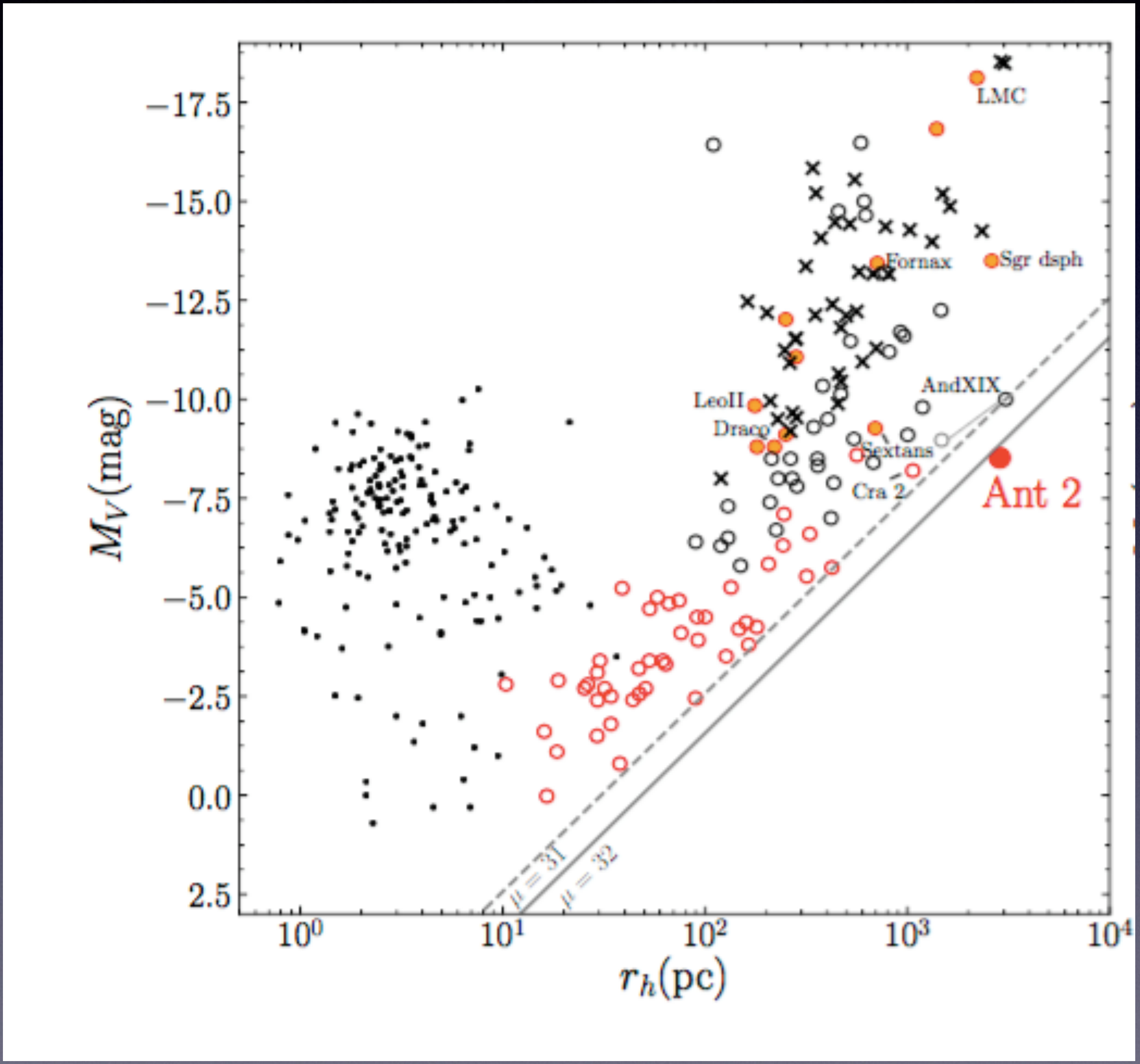
- Satellites and stellar streams of the Milky Way
- Galaxy dynamics with stellar ages
- AGN catalogues (Gaia & unWISE)

Milky Way Satellites

- The Southern sky has so far been surveyed by DES down to $\mu \approx 31$ mag arcsec^{-2}. DES has only observed about a quarter of it. There may be lower surface brightness galaxies hidden in the DES footprint. Most of the rest of the Southern sky is unexplored at these faint surface brightness limits.

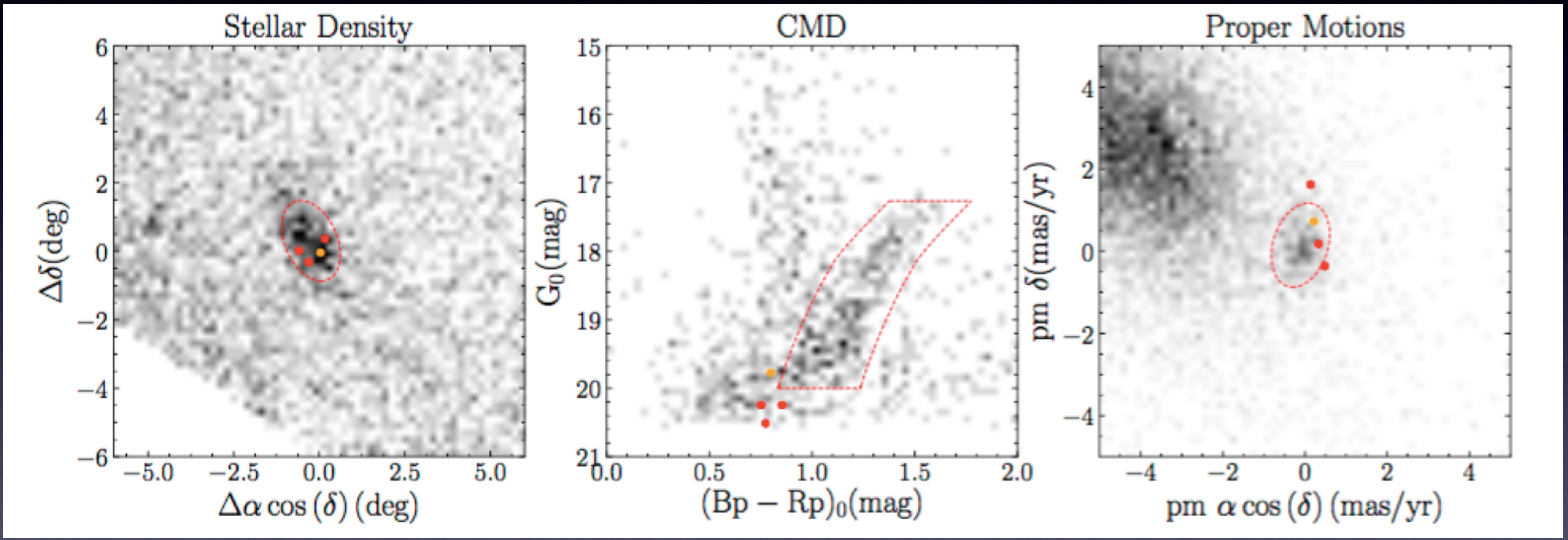
- Models suggest that there should be between 50 and 300 satellites with $M_V<0$ within 100 kpc of the Milky Way. Extrapolating to the viral radius increase the number to many hundreds.

Milky Way Satellites



Gaia discovers Antlia II with μ ≈ 32.3 mag arcsec^{-2}.

Milky Way Satellites



Torrealba et al 2019

The satellite was identified using a combination of astrometry, photometry and variability data from Gaia DR2 (its nature was confirmed with deep archival DECam imaging).

Milky Way Satellites

- Antlia 2 is located behind the Galactic disc at a latitude of b ~ 11^{o} and spans 1.26^{o}, or ~ 2.9 kpc at its distance of 130 kpc. The same size as the LMC, Antlia 2 is orders of magnitude fainter with M_V = −8.5 mag.It is by far the lowest surface brightness system known — 100 times more diffuse than the so-called ultra diffuse galaxies.

- Using the AAT, its velocity dispersion is ~ 5.7 kms^{-1}. Antlia 2 has one of the least dense Dark Matter halos found to date. Modelling suggests that a combination of a cored DM profile and strong tidal stripping may explain the observed properties.

Stellar streams

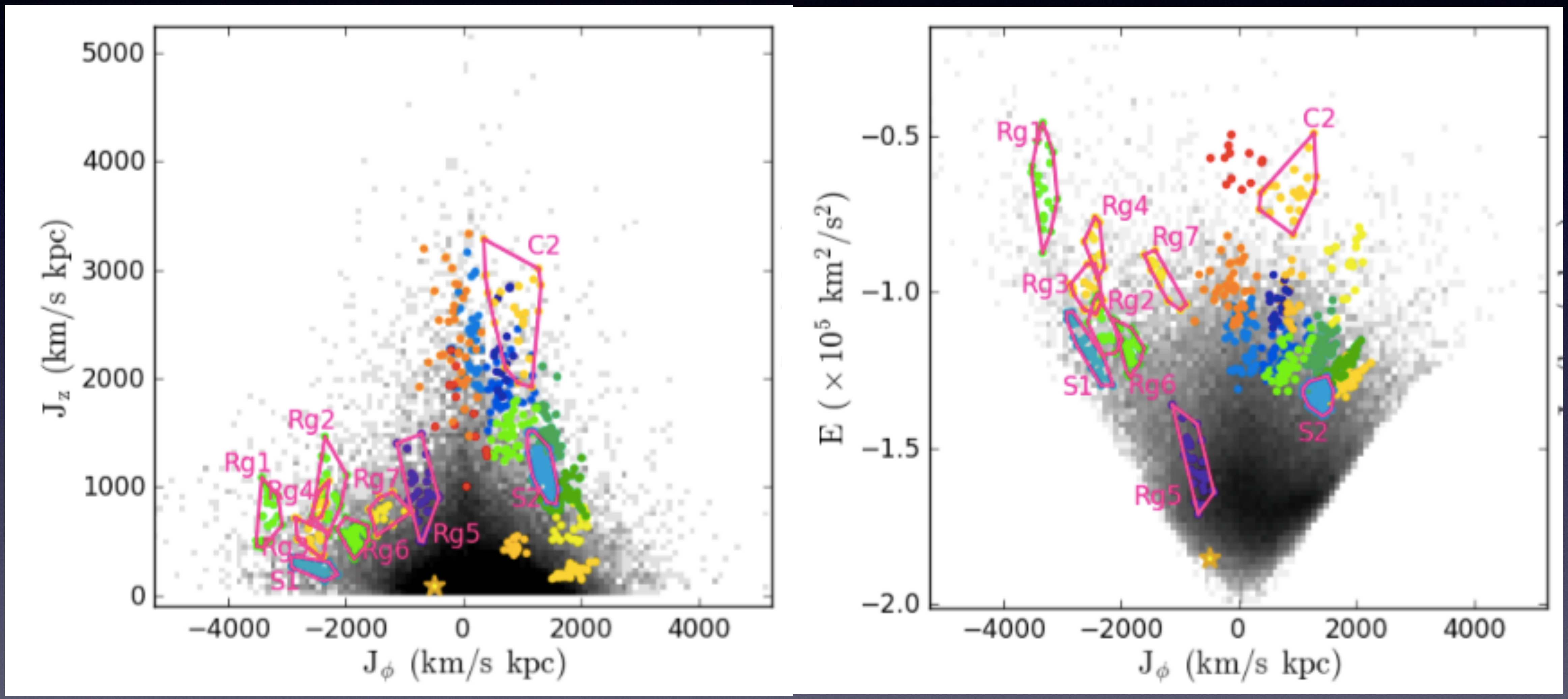
- While Gaia has discovered only on new ultrafaint dwarf galaxy, it has found abundant stellar streams.
- Streams remain kinematically cold and identifiable as substructure in phase space long after they have ceased to be recognisable in star counts against the stellar background of the galaxy (Tremaine 1993, Johnston 1998).
- ΛCDM suggests that there should be hundreds of partially phase-mixed structures in the inner halo.

STREAMFINDER

- STREAMFINDER (Malhan & Ibata 2018) shoots trial orbits within a realistic Galactic potential, using the astrometric and photometric measurements of the stars to select initial conditions for the orbits. The likelihood of this stream model is compared to a no-stream model (or the background).

- STREAMFINDER has found 8 new high-significance structures found at heliocentric distances between 1 and 10 kpc and at Galactic latitudes $|b| > 20^{o}$, named Slidr, Sylgr, Ylgr, Fimbulthul, Svöl, Fjörm, Gjöll, and Leiptr.

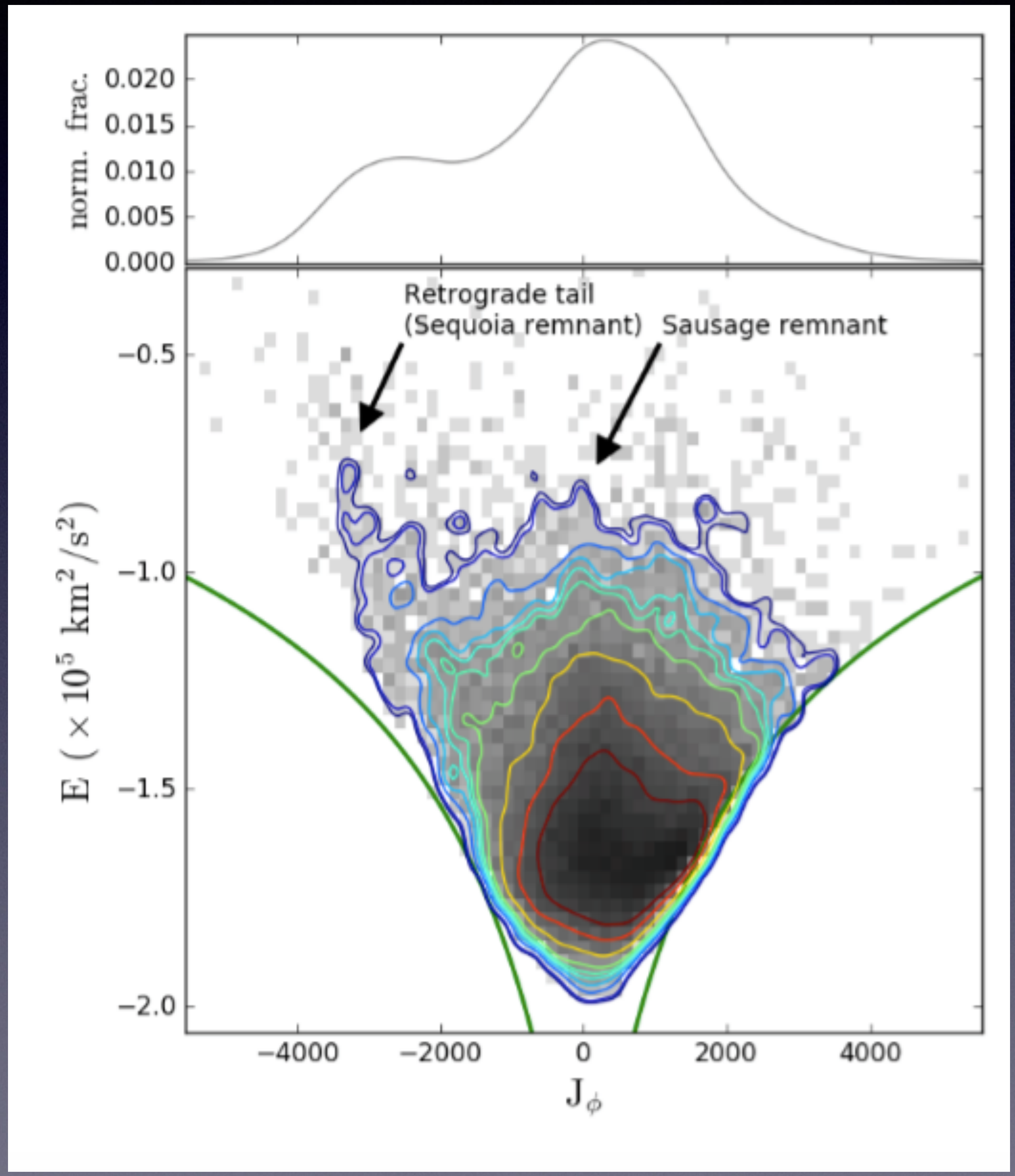
Action searches



Myeong et al. 2018

A KDE is used to provide a smooth background, and then over-densities in action space identified.

The Sequoia Event



Myeong et al 2019

LSST Streams

- Resolved stellar streams in imaging data and phase space substructure can be found by direct application of existing techniques to LSST photometry/astrometry.

- LSST can probe an entirely new regime. Stellar halo substructure in the Local Universe has been found using observations with low-aperture telescopes (e.g. Martinez-Delgado et al 2010, Abraham & van Dokkum 2014). We are developing algorithms for LSST so that we can also search for faint, unresolved tidal features out to ~ 10s Mpc.

LSST Streams

- No wide-area automated survey of such unresolved tidal substructures has been run. Kado- Fong et al (2018) made the first step in automated tidal feature identification as applied to ~200 square degs of HSC imaging & found 1000s faint tidal streams around galaxies with $0.05 < z < 0.45$
- The algorithm is a variation of unsharp masking, where the smoothed version of the image is subtracted to emphasize high-frequency features. It performs the convolution with a library of kernels and combines the high-frequency detections into one master image, delivering a clean set of the tidal features in the outskirts of the host.

Machine Learning

- Gaia offers huge potential for machine learning

AGN Catalogues

- Cross-match the Gaia DR2 optical source catalogue of 1.7 billion sources with the unWISE mid-IR catalogue of 2 billion sources.

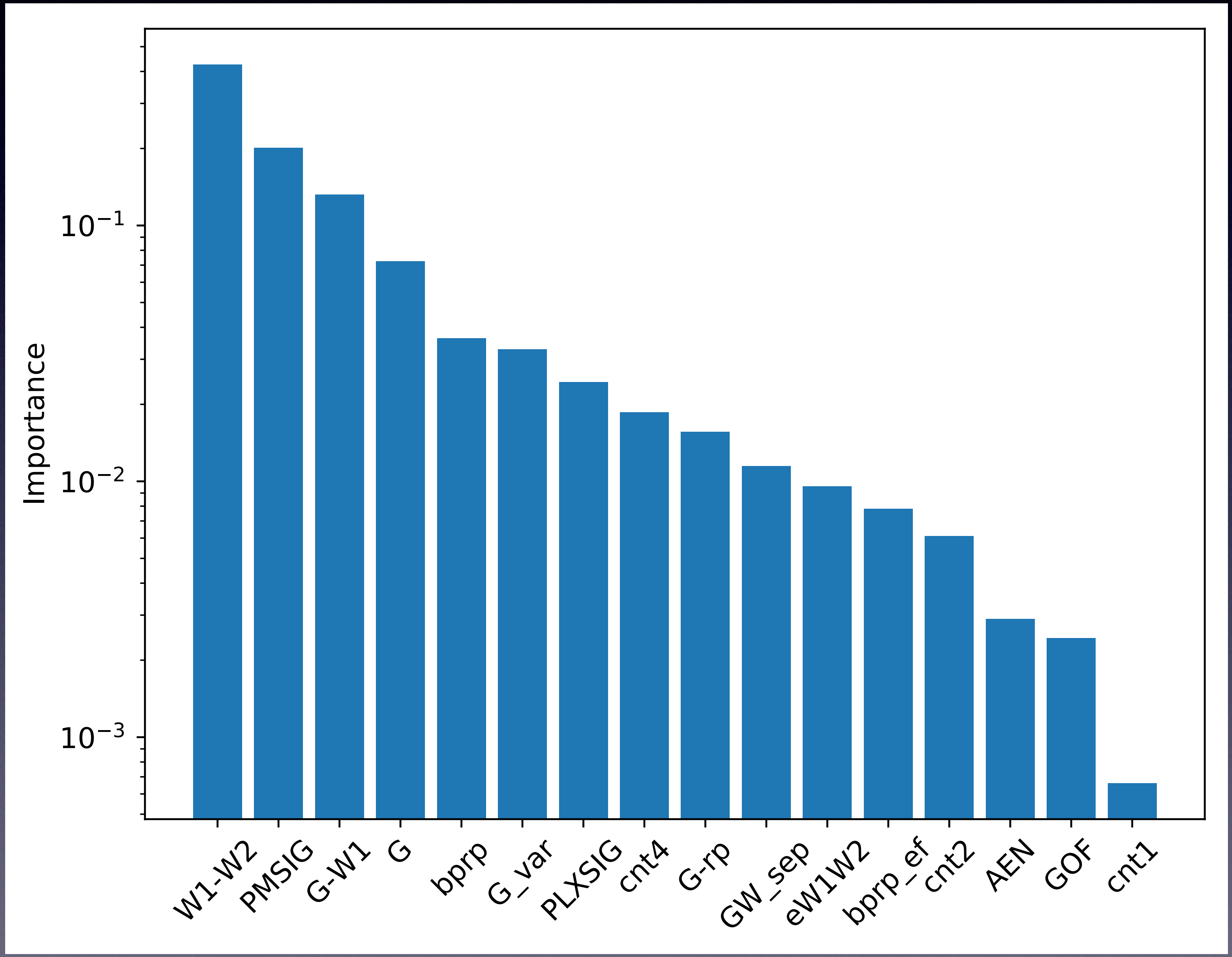
- The Gaia-unWISE sample consists of 641,266,352 sources. Only point-like sources can be detected by Gaia, so these objects are either stars, or AGN, or bright, compact star-forming regions in external galaxies.

AGN Catalogues

- We build the AGN data set for a RF classifier from the largest spectroscopically confirmed QSO sample --- the SDSS DR14 QSO catalogue. Requiring unWISE and Gaia counterparts results in 339,194 QSOs.

- To build the non-AGN data set, we randomly select 10 million objects from the Gaia-PS1 and cross-match them with the unWISE catalogue. We clean this non-AGN data set by identifying and removing as many known AGNs as possible using the million quasar catalog & the ALLWISE two-colour selected AGN catalogues. This give us 2,340,541 objects.

AGN Catalogues



Shu et al 2019

Use a RF classifier based on 16 features

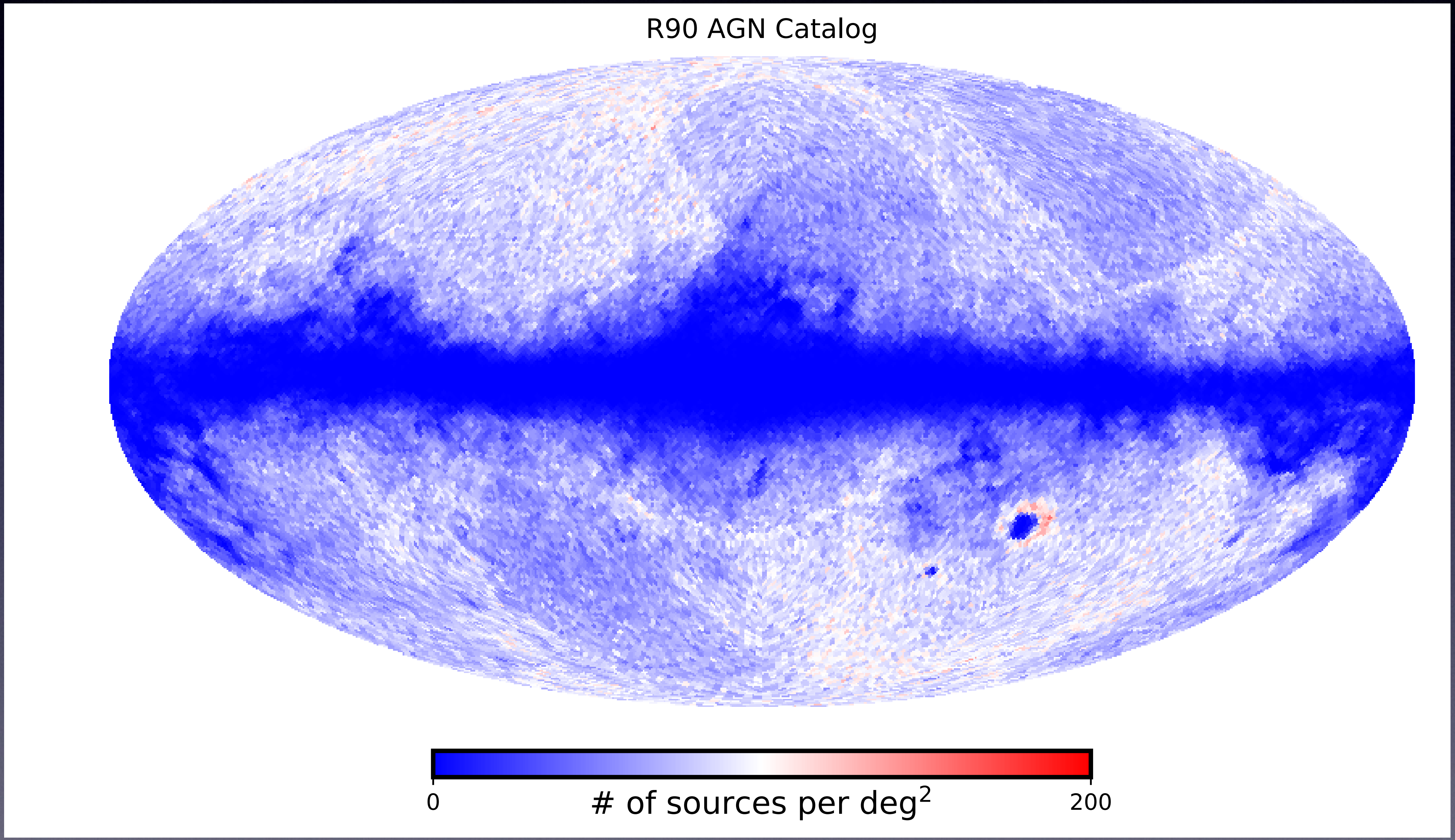
AGN Catalogues

- We build two catalogues, C95 and R90. They are constructed by applying different p_{AGN} thresholds to achieve 95 % completeness and 90 % reliability respectively.

- The C95 AGN catalog contains 2,873,393 sources, and the R90 AGN catalog contains 2,344,069 sources. At least 1,021,962 and 608,610 AGN candidates in our C95 and R90 catalogues are previously unknown.

- In Gaia, there are ~ 100 AGN per square degree.

AGN Catalogues



Shu et al 2019

R90 distribution — the candidate AGNs

Stellar Ages

This is the ultimate revenge fantasy

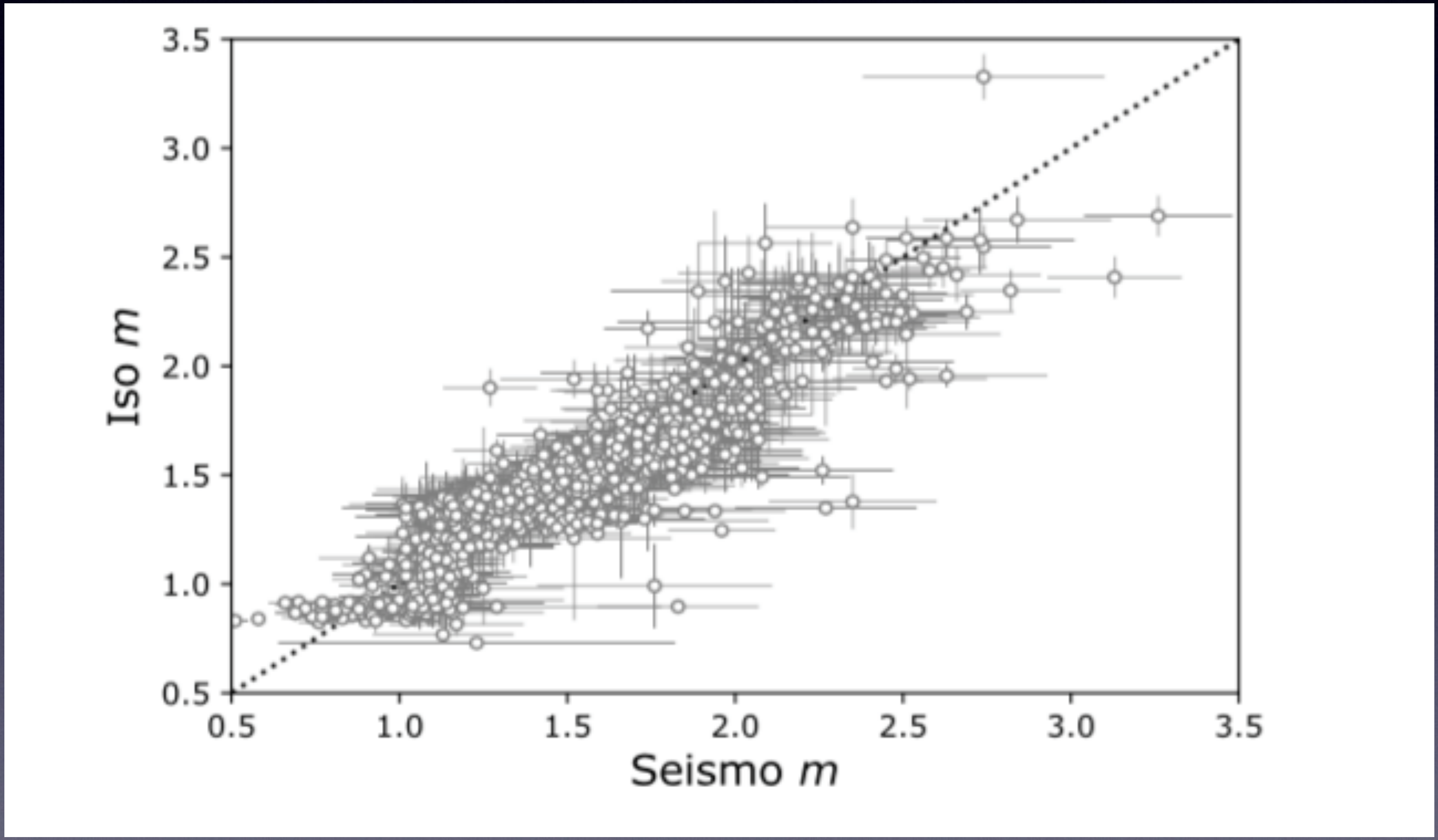
- Asteroseismological ages (needs high cadence, high precision stellar photometry, restricted to nearby stars).
- Isochrone ages (restricted to subparts of the HR diagram)
- Chemical ages (RC, RGB)

Stellar Ages

- The feed-forward ANN is trained using a sample of red giant stars with mass estimates from asteroseismology. A Bayesian pipeline uses the astrometric, multi-band photometric, spectroscopic and asteroseismology data to determine posterior distributions for the training outputs: mass, age, and distance.

- The ANN inputs are *H*, *J-K*, parallax, log g, effective temperature, [M/H], [α/H], [C/M] and [N/M].

Stellar Ages



Sanders et al 2019

Stellar Ages

- The ANN was applied ~ 10 000 red giants in the overlap between the 14th data release from the APOGEE and TGAS. The ANN was able to reduce the uncertainty on mass, age, and distance estimates for training-set stars. The fractional uncertainties on mass are < 10% and on age are between 10 to 25%.

- There is room for much more work on this fundamental problem with LSST/Weave/4MOST.

Conclusions

- The Gaia data is an excellent testing ground to develop algorithms for subsequent extension to LSST, especially of streams and faint satellites. We have already seen with the discovery of Antlia II & the Sausage/Sequioa hints of what the LSST will uncover.

- The Gaia data is an excellent testing ground for machine learning projects, in preparation for LSST (e.g., automatic AGN identification, stellar ages).

We are Pavlov's Dog

AGN Catalogues