servian_

Dr. Elaina A. Hyde
Western Sydney University, Adjunct Fellow
Consultant/GC Engineer, Servian
Twitter @AstroHyde
elaina.hyde@servian.com.au

Data Science in Astrophysics

servian_

Established in 2008, Servian has offices in Sydney, Melbourne, Adelaide, Canberra, Brisbane and Bangalore. Our mission is to enable our customers to use their data and analytics to build competitive advantage.

200+ onshore employees

50+ customers

7 locations

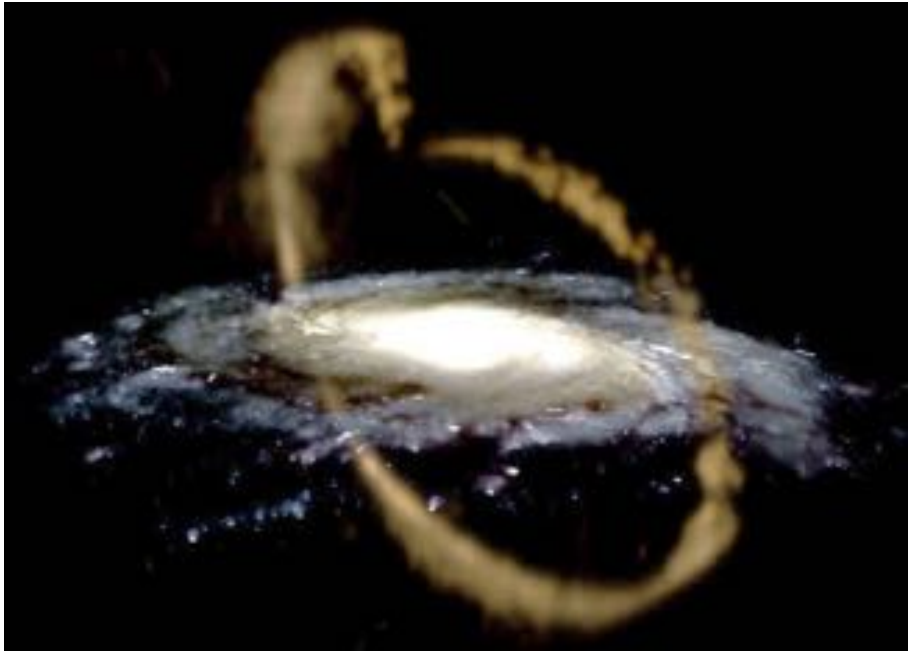
Acknowledgement of Country

Before we begin I would like to acknowledge and pay respect to the traditional custodians of the land, Elders past and present of the Gadigal of the Eora Nation. It is upon their ancestral lands that this building stands.

Data Science in Astrophysics and Beyond

20,000 stars in the Sagittarius (Sgr) Dwarf and Stream. Uncover the velocity and metallicity distribution.

A new selection of probable Sgr member stars in the stream using a statistical selection method and present their metallicity distribution.



The Local Group and What We Expect to See Today?

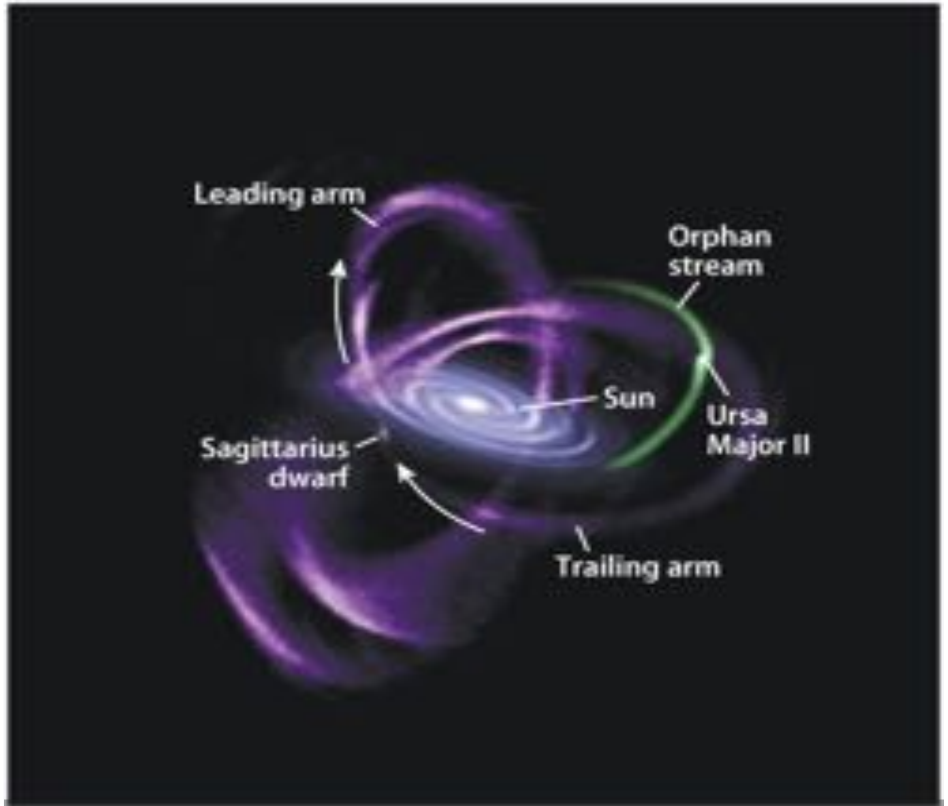
The Milky Way
M31, M33
30+ Dwarf Galaxies

Theory: Hierarchical Scenario (small + small ➔bigger…)

We should see:

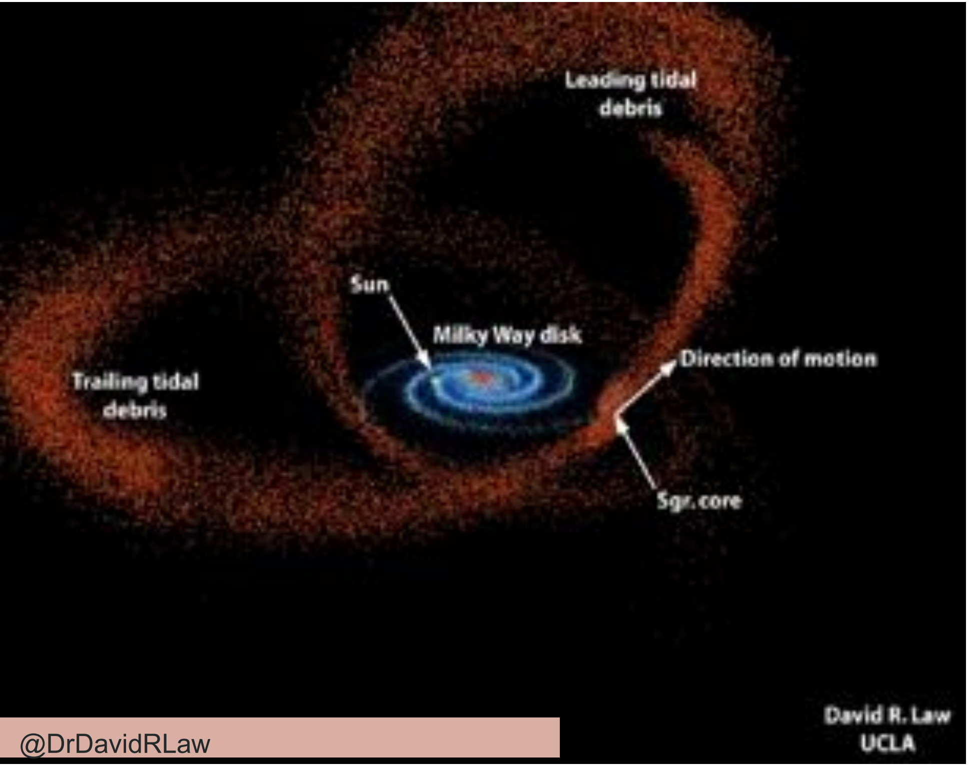
Many stellar streams
Many surviving satellites
small satellite galaxies (hundreds?) predicted in the Local Group

Missing Satellites?



Vasily Belokurov: Sagittarius Stream

The Sagittarius (SGR) Stream

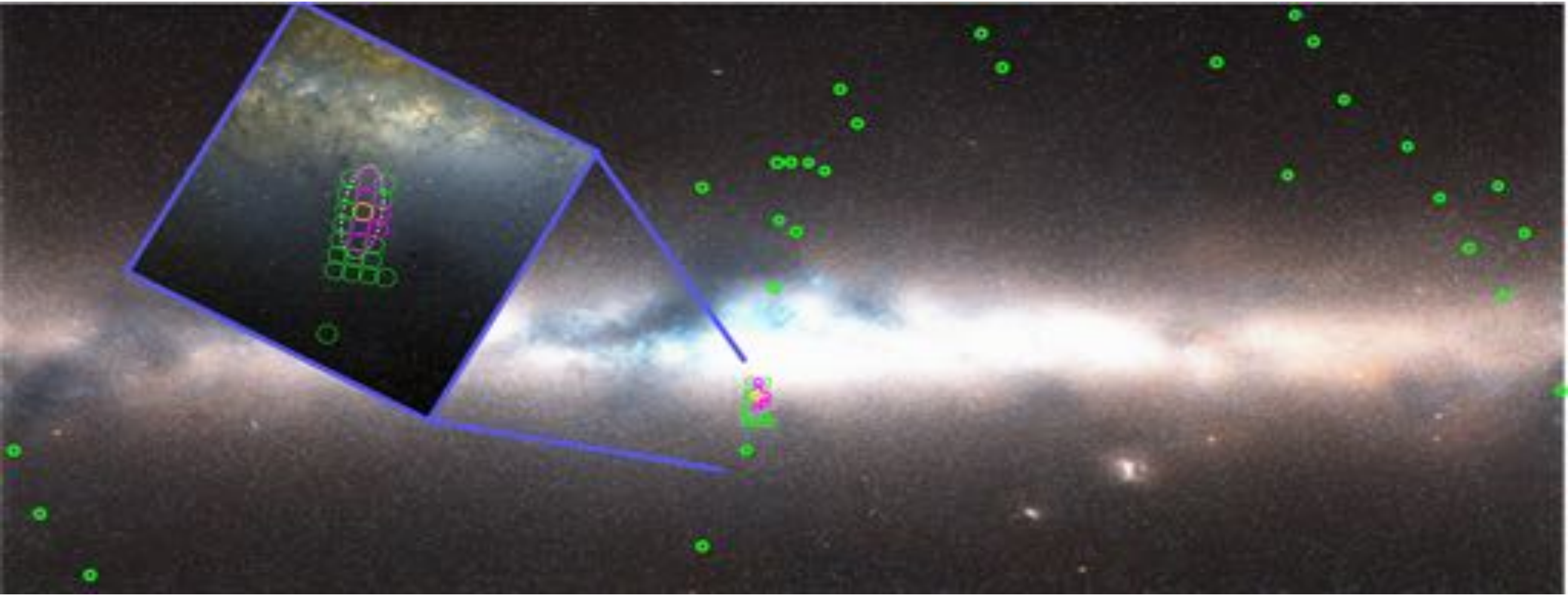


Discovered 1994

Sagittarius Dwarf, a Milky Way satellite in the process of being torn apart by our Galaxy's gravity.

The original Sgr dwarf galaxy has been thought to be a non-rotating, spheroidal galaxy.

24,110 Spectra



Background image constructed from online data Mellinger (2009).

How do we get the data?

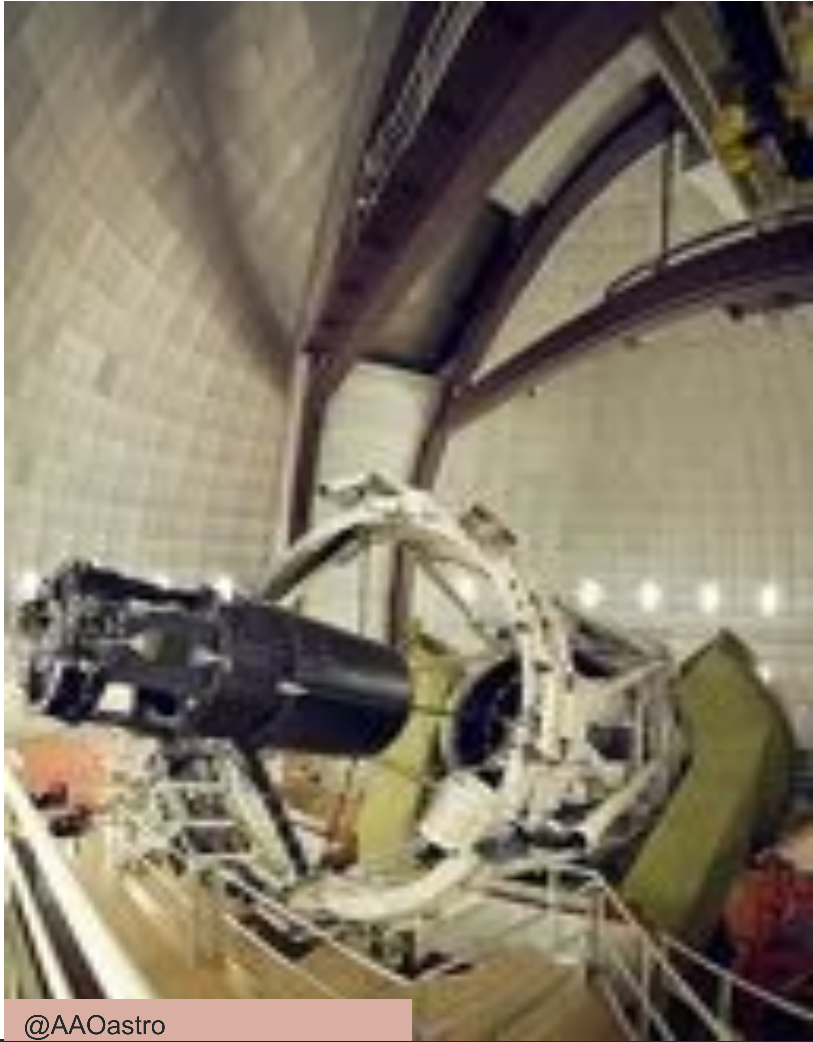
The Data

AAT multi-object spectra from AAOmega

The general reduction was done with the 2dfdr data reduction program of the Australian Astronomical Observatory (AAO)

The IRAF task fxcor to extract radial velocity information. Due to the large number of fibers, 400 per field, the velocities were measured by using a template spectrum fitted to the Ca II triplet.

Modified RAVE analysis



@AAOastro

The Robot


DANGER
KEEP HEAD AND LIMBS CLEAR
ROBOT MAY MOVE AT ANY TIME

Parameters of Interest

- Position: multiple observations of the same star
- Velocity: stars grouped together
- Color: 2MASS catalog
- Gravity=Log(g)
- Temperature=Teff
- Metallicity – self measured
- RAVE pipeline
- Model values

The Problem

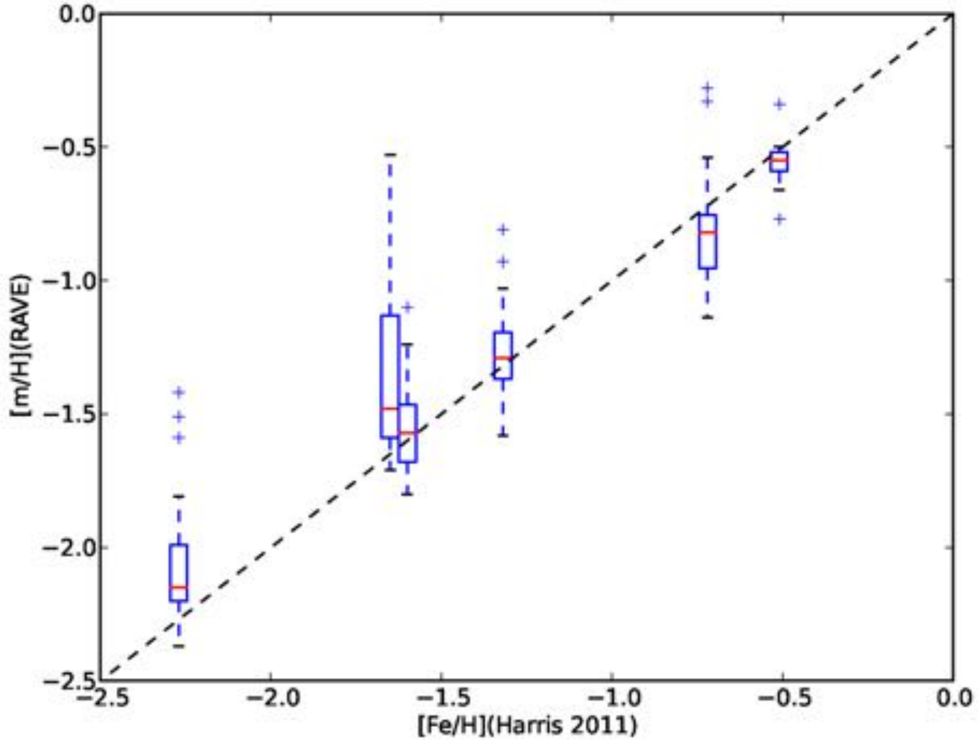
How to select best data, and what we mean when we say `best' is an issue still very much under discussion in astronomical data analysis. From Hyde et al. (2015) we used a statistical selection technique to identify potential member stars for the Sagittarius Dwarf stream based on membership likelihood.

This type of selection is geared towards a statistical likelihood that, after eliminating all Galactic halo stars in our sample, whatever left will be non-halo stars. After selecting for Sagittarius members photometrically we used generalized Gaussian histograms to compare the data to the smooth Galactic halo and to define a selection criteria for likely Sgr members from the population of stars in a field

Metallicity Calibration

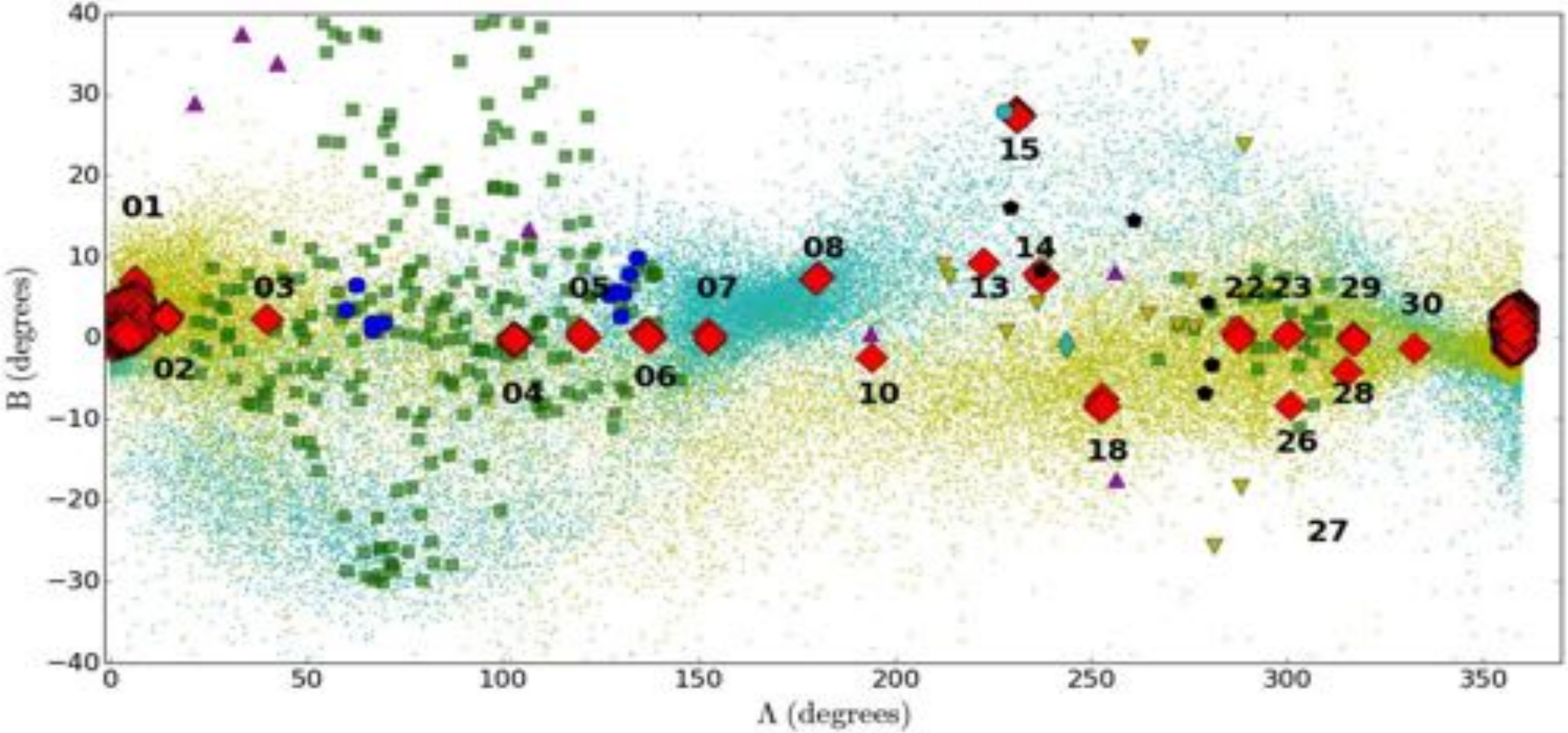
Metallicity measurements, a new calibration based on six known clusters

RAVE results processed by A. Siebert
M and K giants



RAVE Calibrations (Siebert et al. 2011)

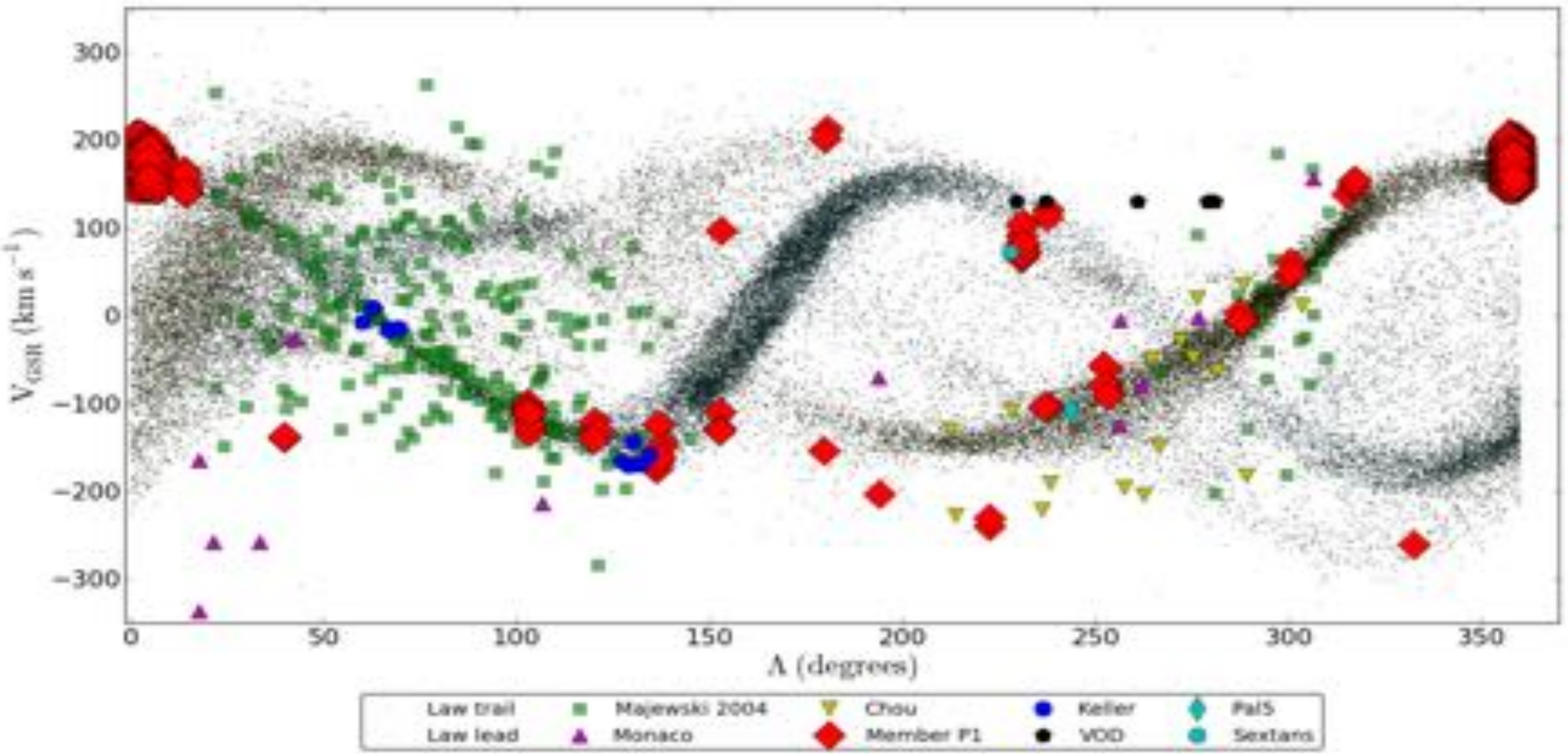
Λ,B



Sgr models (Law et al. 2004,2010)

Λ,B Coordinates (Majewski et al. 2003)

Velocity Distribution



The Selection

Generalized Gaussian histograms compare the data to the smooth Galactic halo and define a selection for likely Sgr members.

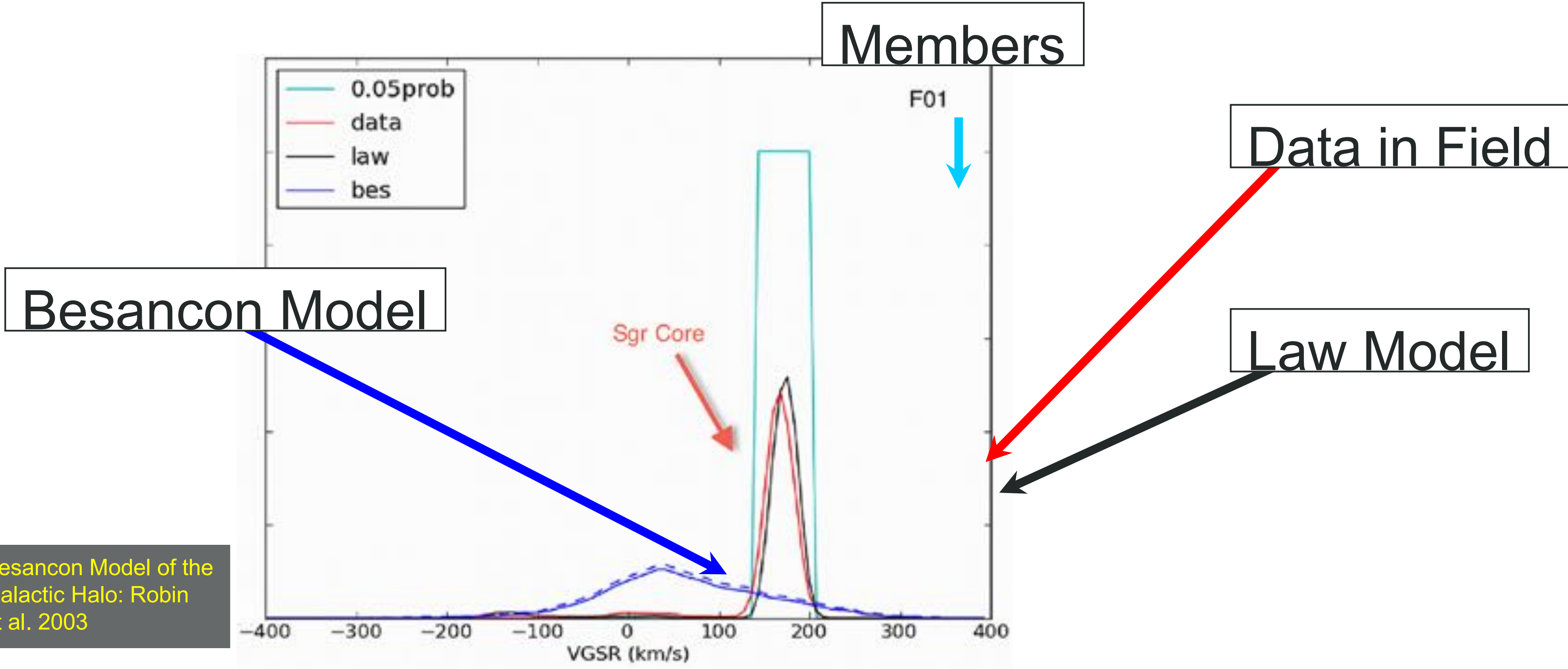
$$D_{\text{field}} = \sum_{i}^{N} D[i] = \frac{1}{(\sigma \times \sqrt{2\pi})} \times e^{-(x-\mu)^2/(2\sigma^2)}$$

where i is each measurement VGSR for each star in the field in question. The value μ is given by each measurement of VGSR in the field. N is the number of stars in the field, σ = Verr = 2.7 km s−1 and μ = VGSR and produces the velocity distributions. Stars with Dcand ≤ DB to have a PNotHalo = 0 and likewise if DB ≤ 0.001 we let PNotHalo = 1. In the region where Dcand ≥ DB then we need to consider the difference in the distributions. To keep an accurate scale of likelihood between zero and one we will set a confidence level at C=0.996 (cut off where Dcand ≥ DB).

$$P_{\text{NotHalo}} = \left[\frac{D_{\text{cand}}}{D_B} - 1\right] / (C_{\text{bes}}/D_B)$$

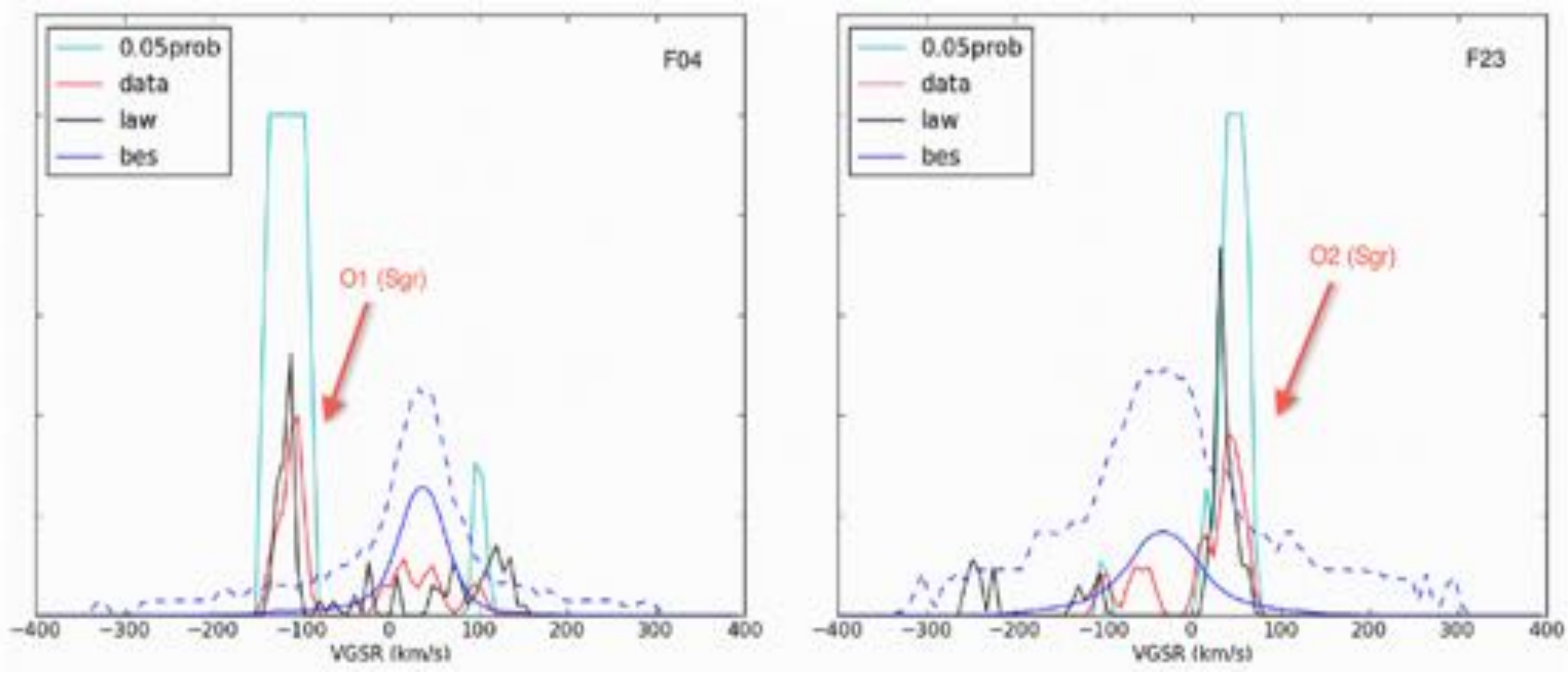
➔106 potential stream members

SGR Member Selection for the Stream

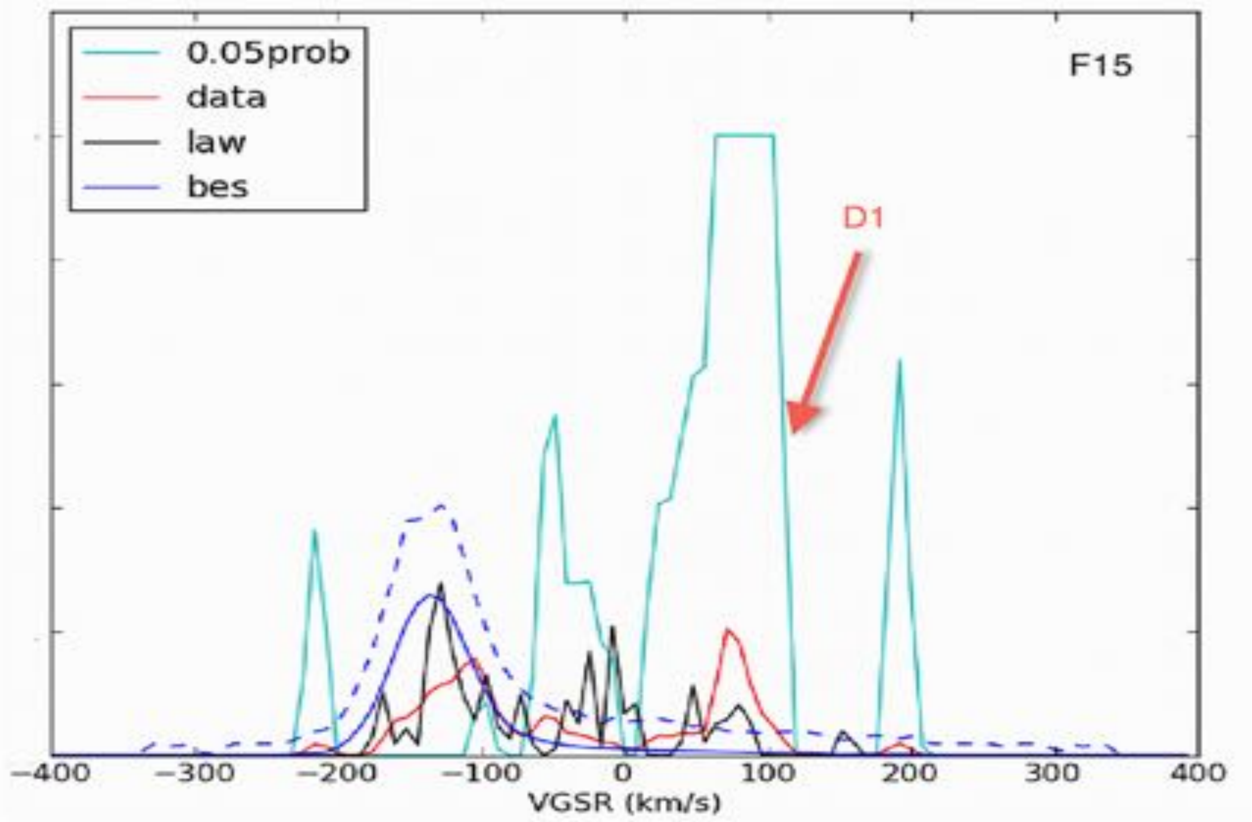


Besancon Model of the Galactic Halo: Robin et al. 2003

SGR Member Selection for the Stream



SGR Member Selection for the Stream



Besancon Model of the Galactic Halo: Robin et al. 2003

Triaxial model Law et al. 2012

Significant populations?

The Approximate Position and Velocity of P1 Overdensities

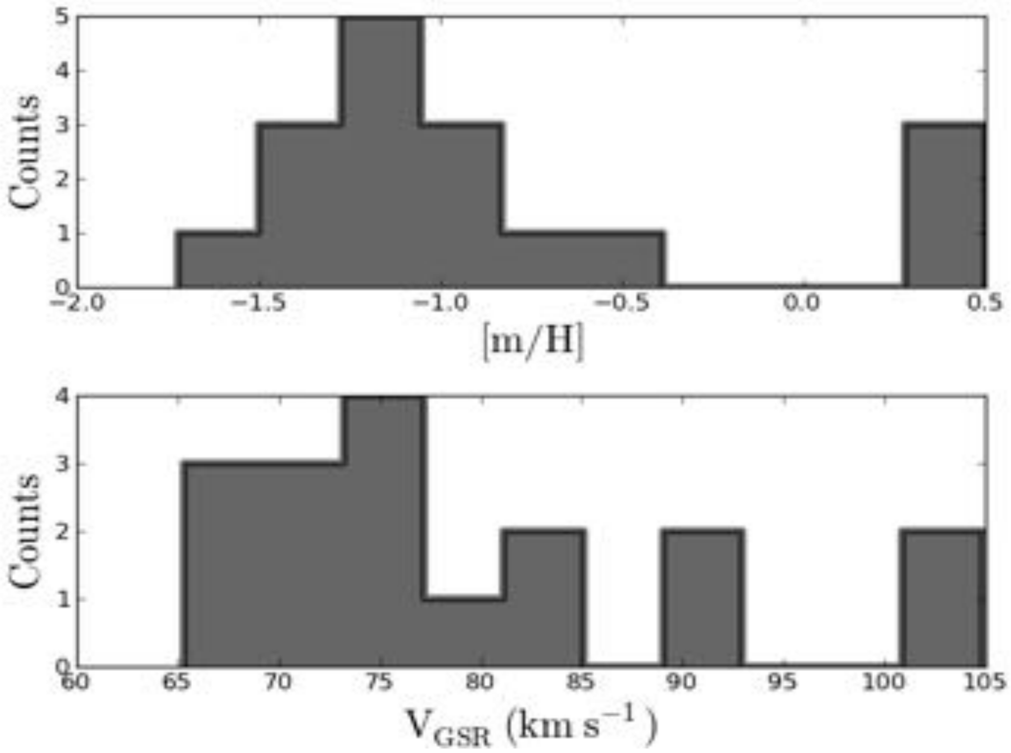
Field	Stars	Λ_{avg}	B_{avg}	RA_{avg}	DEC_{avg}	$[m/H]_{avg}$	σ_{mH}	V_{GSR}(avg)	σ_v	Category
		(degrees)		(degrees)		(dex)	(dex)	(km s^{-1})	(km s^{-1})	
01	5513	154.32	1.65	19.0	−30.8	−0.59	0.34	168.76	12.15	Core F01 region of Sgr
04	23	103.05	−0.22	02.0	−01.9	−0.65	0.2	−112.61	10.90	Potential Sgr stream clump
15	17	231.02	27.52	10.2	−01.4	−1.10, +0.48	0.65	79.45	11.49	Potential Sextans detection
23	3	300.31	0.23	15.0	−06.9	0.38	0.02	50.71	5.39	Potential Sgr stream clump

Defining close groups as significant populations in the stream we get the above table of significant results, within 6 degrees and +/- 20km/s while being 3 or greater. Other measures of significance are possible, but in terms of the stream it's hard to resolve them.

What's next?

- Other types of significance?
- Different population requirements and selections?

New Discoveries: The Sextans Dwarf

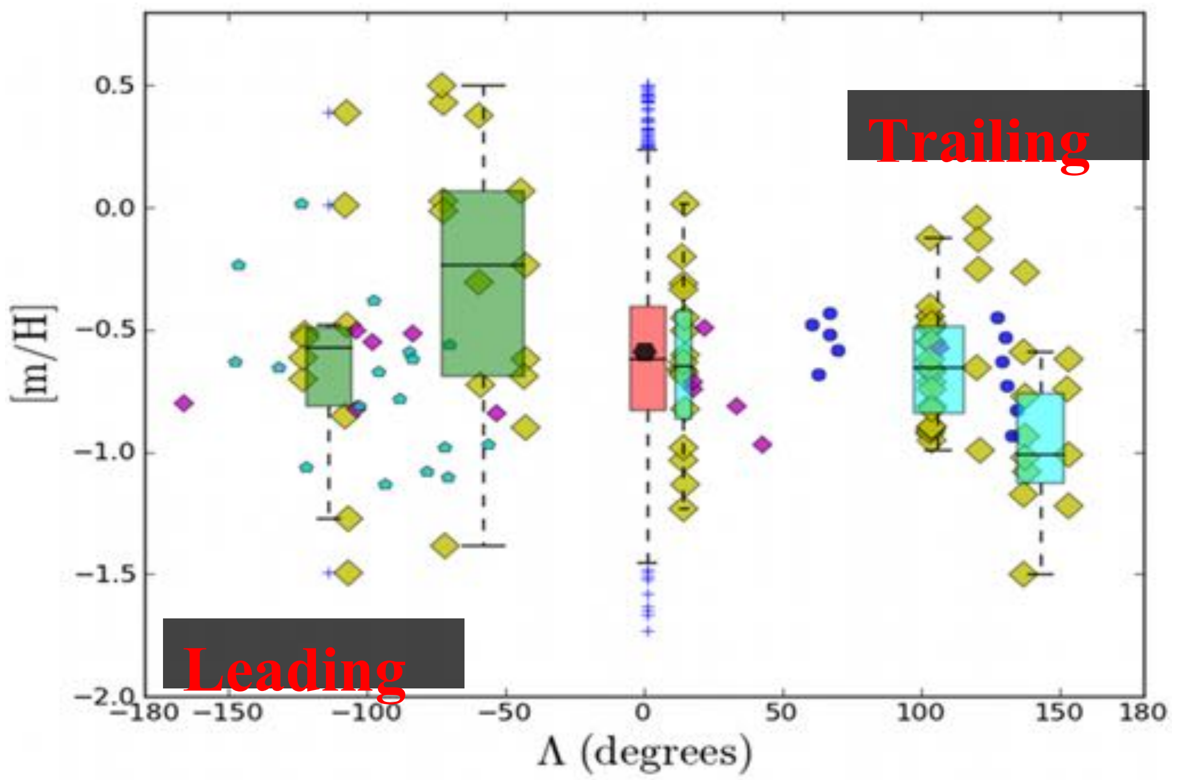


We note a potential detection of the Sextans dwarf in our F15 region.

➔at RA, DEC ≃ (10.2, −01.4),
➔Peaks in the metallicity distribution at [m/H] = −1.10 and +0.46 dex
➔Average velocity of VGSR = 78.48 km s−1.

Metallicity Distribution

P1 data are yellow diamonds, Keller et al. (2010) blue circles, Monaco et al. (2007) magenta diamonds, and Chou et al. (2007) light blue pentagons. The Sgr core is given in black at [Fe/H] ~ −0.5 dex (Monaco et al. 2005).



Core region; mean [m/H] of −0.59 dex (dispersion 0.34).

The black line in each box is the median , the 25th percentile is given by the upper edge , and the 75th percentile by the lower.

Stream bins have :
→Λ ≃ [−59, −112, 142, 106, 13]
→Λmin − Λmax≃[30, 17, 17, 18, 1]
→[13, 10, 11, 24, 16] stars
→mean [m/H]~[−0.27, −0.60, −0.97, −0.63, −0.64]

Stream Conclusions

→Selection technique to separate Sgr member stars

→We identify 106 likely members of the Sgr stream

→The Sgr stream VGSR evolution with Λ has general agreement with the triaxial model of Law & Majewski (2010) .

→Intrinsic MDF gradient in the progenitor supported

→Potential detection of the Sextans dwarf at F15.

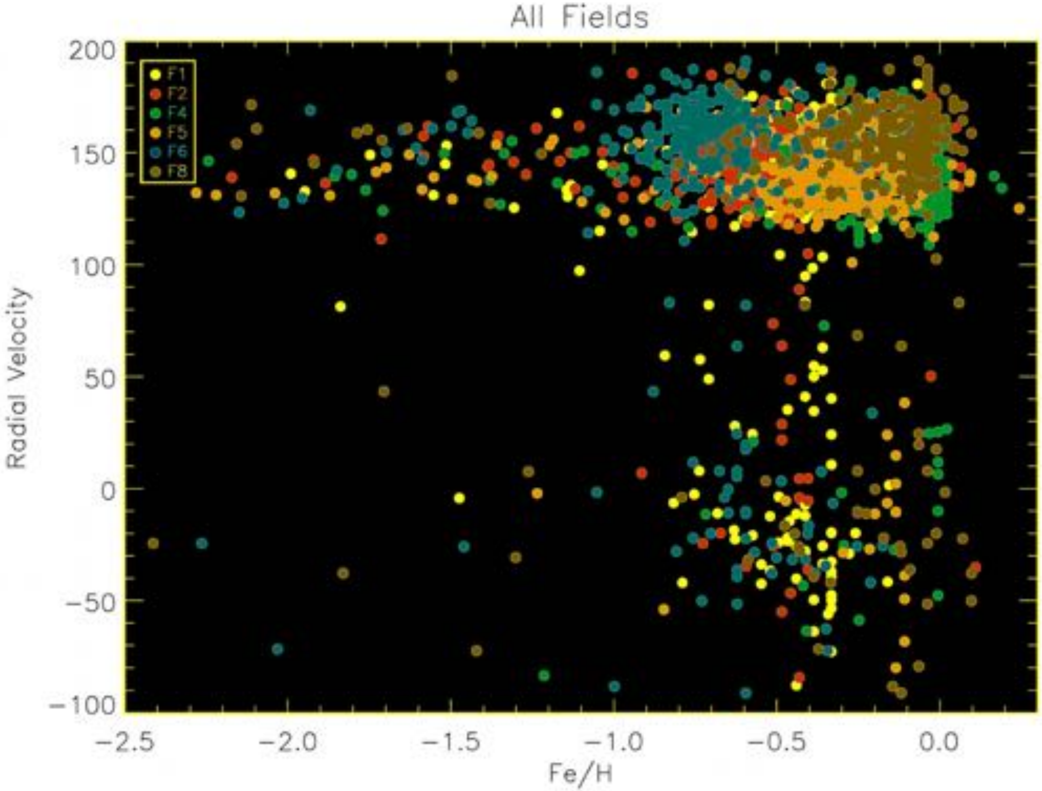
The Core

If Sgr was in fact a rotating disc galaxy
Residual rotation with a measurable amplitude (20 km/s).

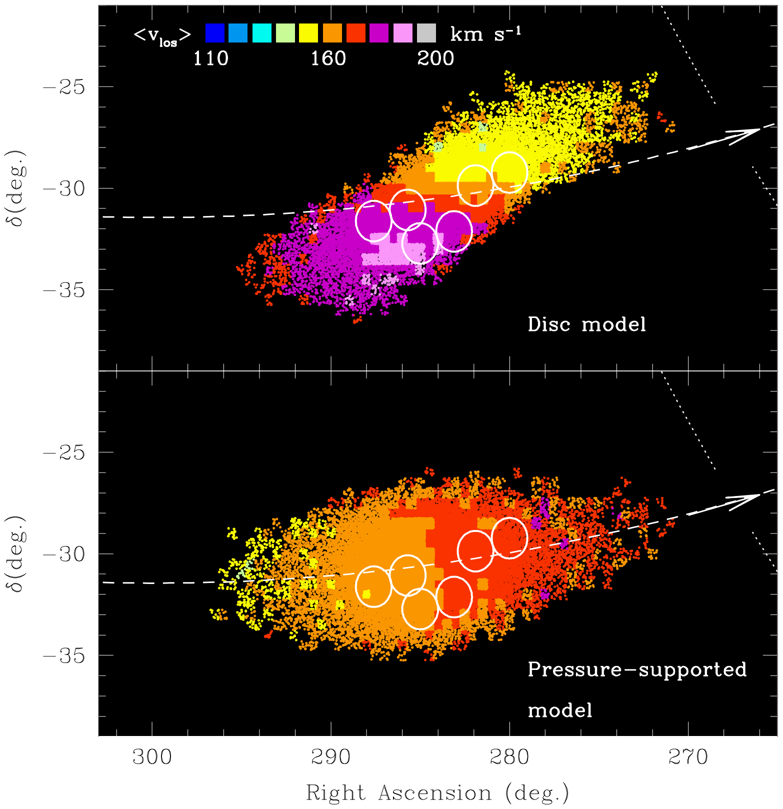
Our original data set included spectra of over 2000 stars near the core of Sgr in order to test the disc-galaxy hypothesis…

Is it rotating or not?

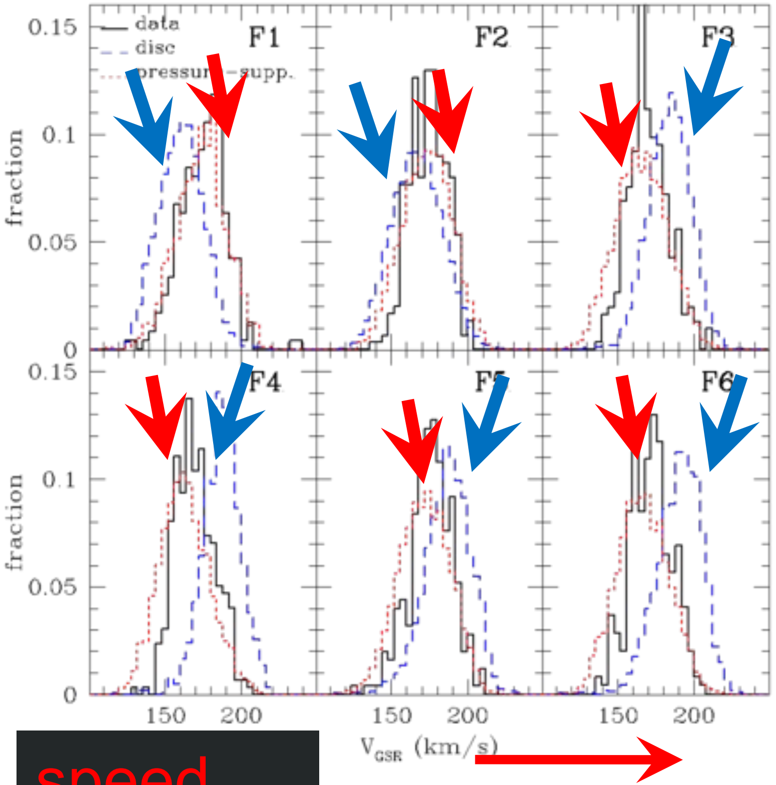
Sgr Members clump together near the core



What do we see?



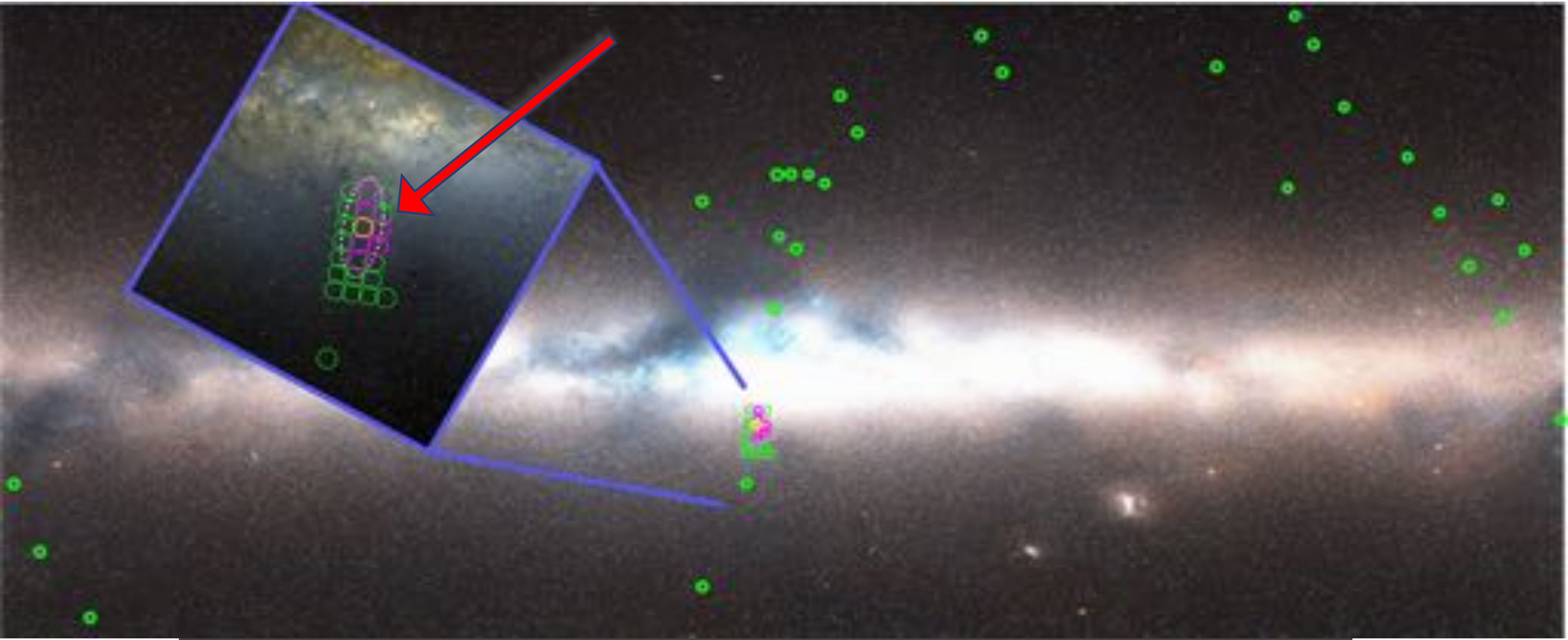
Non-Rotating matches better!



speed

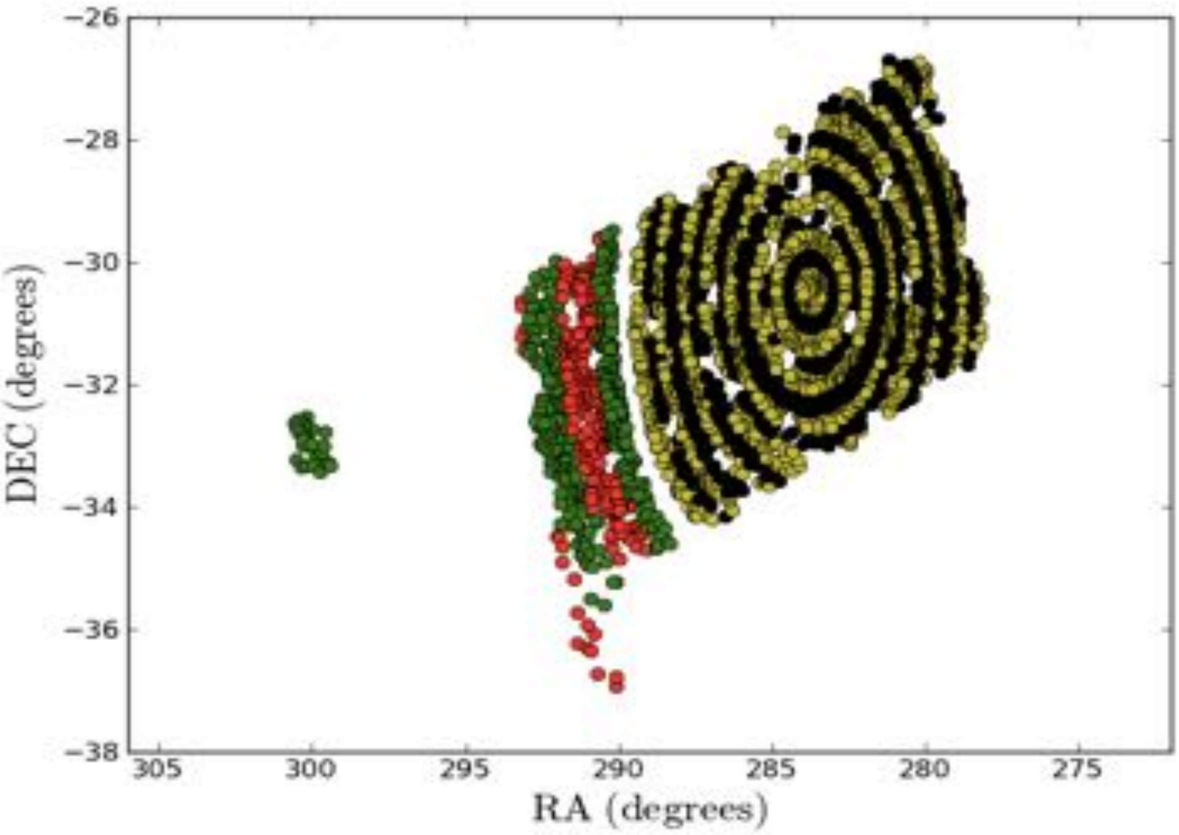
Peñarrubia et al. (2011 ApJ)

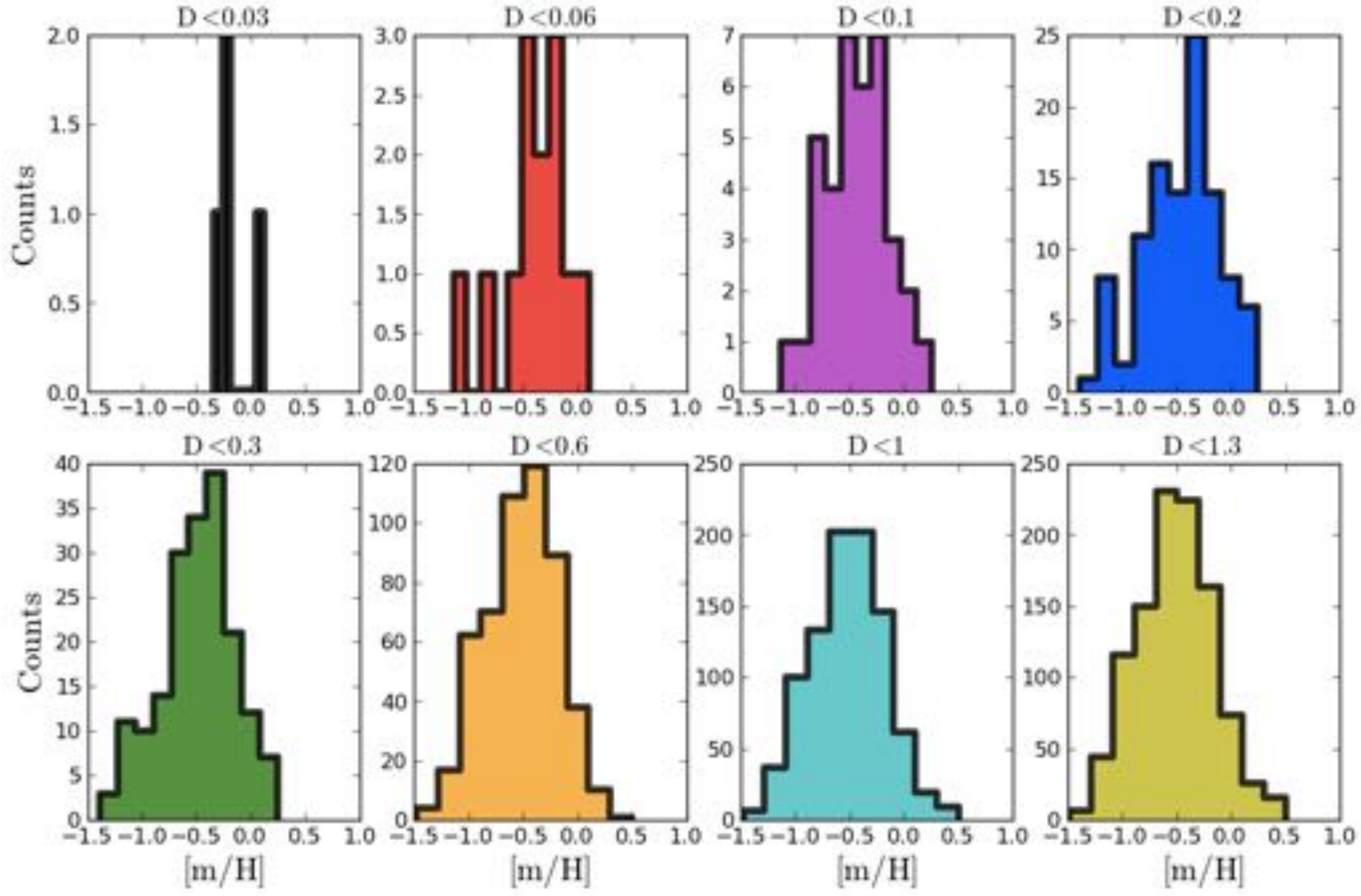
New Core Subset → 9086 Spectra



Background image constructed from online data Mellinger (2009).

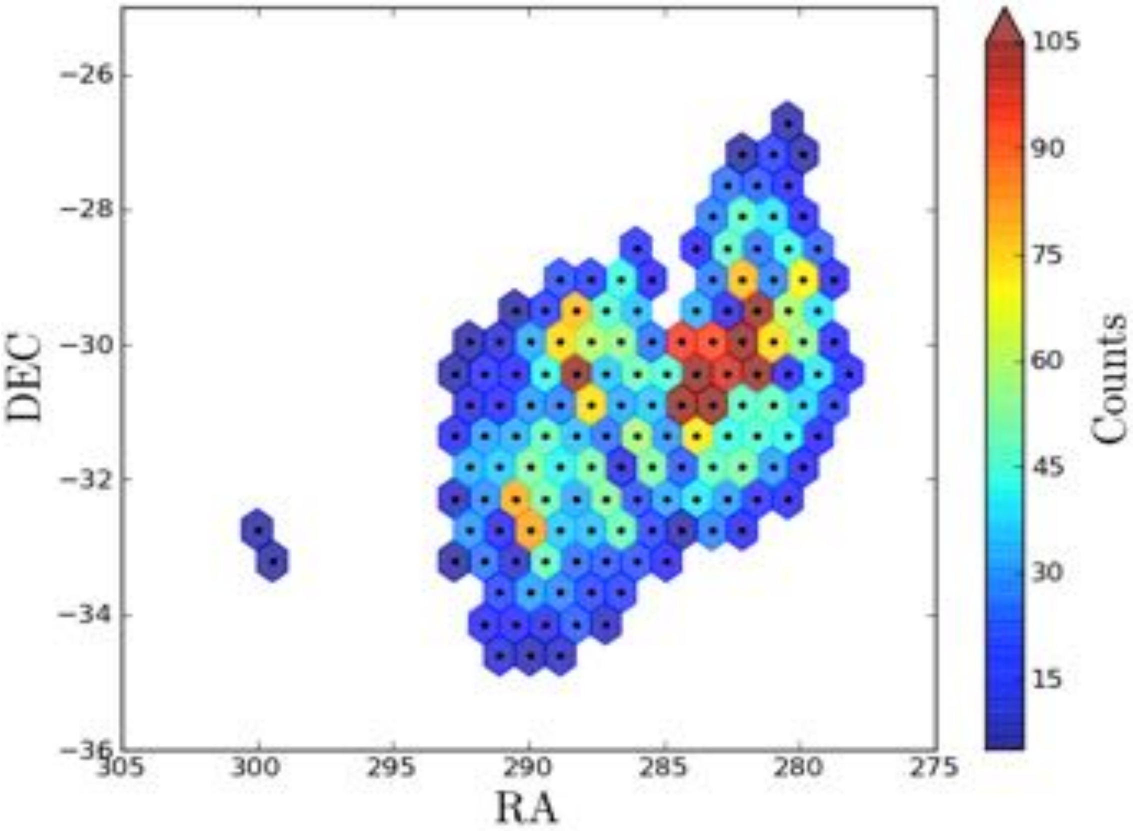
Radial Annuli Distribution





M54 Cluster in the Core?

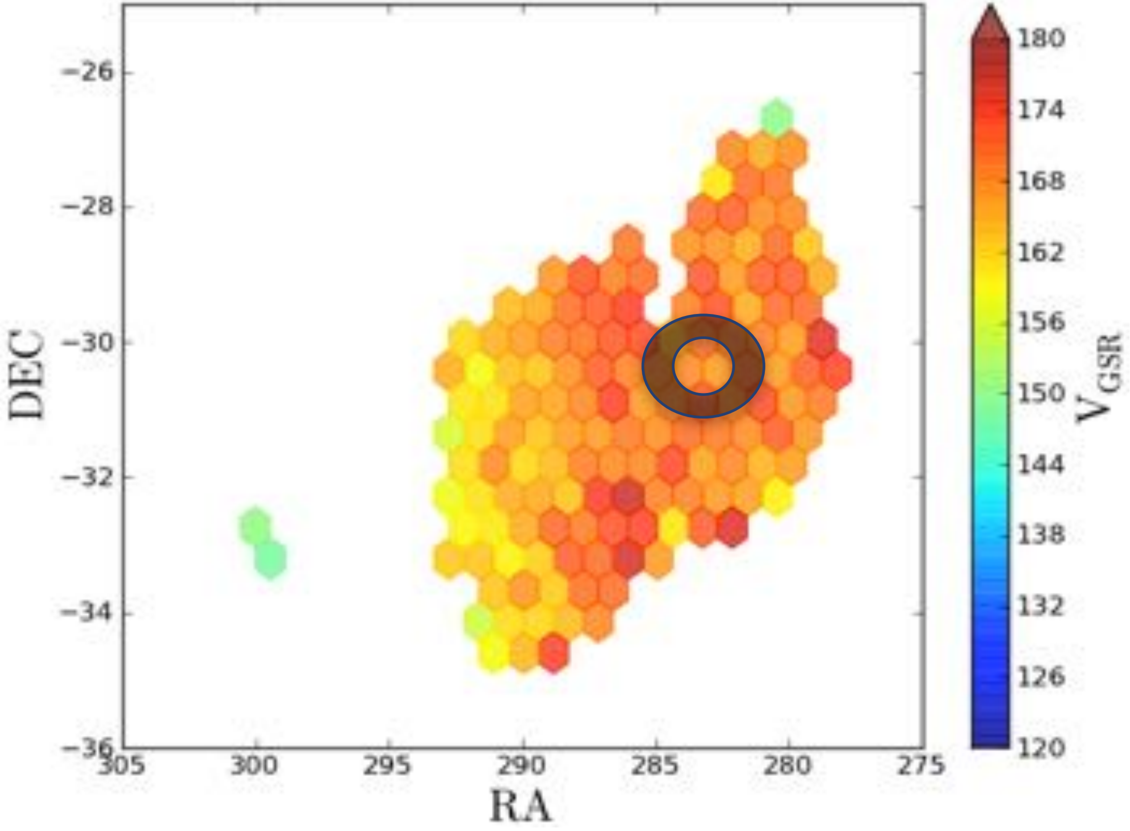
Hexagon Based Distributions



Velocity Spatial Distribution

Hexagonal binning
Sum of the std VGSR /N_hexagons gives a typical velocity standard deviation of σtyp= 12.7 km s−1.

VGSR gradient from about 156 to 170 km s−1 from east to west in the RA direction. This is consistent with a gradient of ~ 15 km s−1 as expected from Peñarrubia et al. (2011).

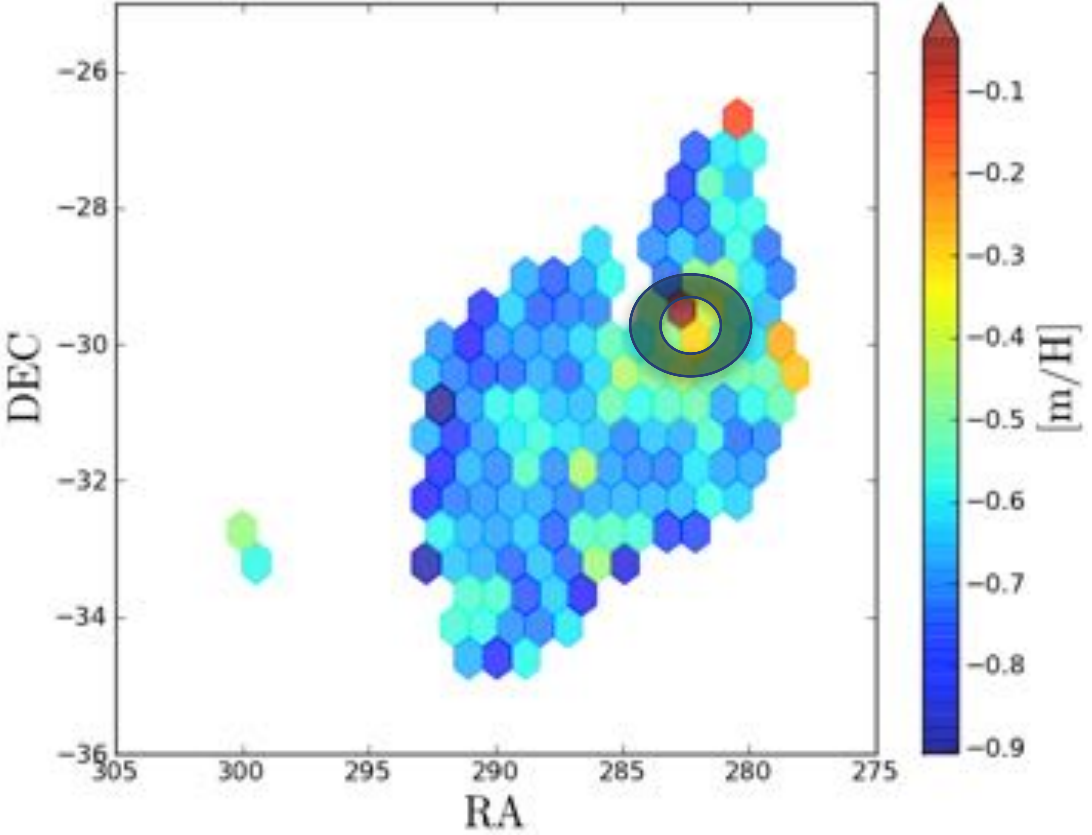


Metallicity Spatial Distribution

M54 0.06◦ from the centre. Overwhelmed by Sgr!

Average [m/H] likely higher 2◦ from centre of Sgr. Overlapping MDFs of inner and outer regions make for large uncertainty.

Concentration of metal-rich stars in centre significant at a 40% level. Found for some, but not all, other dwarf spheroidals in the Local Group (Koch et al. 2006; Harbeck et al. 2001).



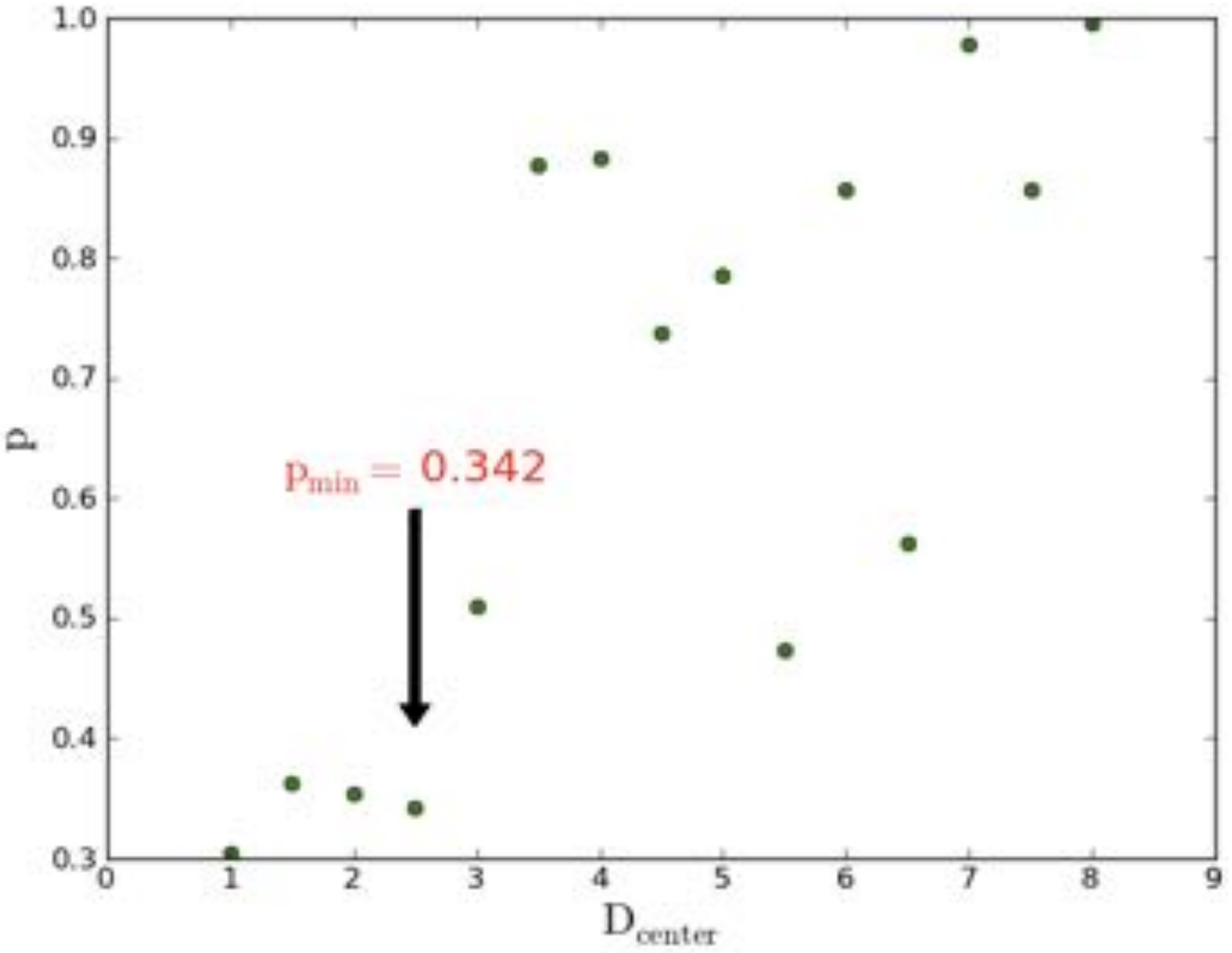
Kolmogorov-Smirnov test

Define a cut off for inner and outer regions that are 'most significant' using KS test. The 2 sample KS statistic is:

$$D = sup_x \, |A(x) \; - \; B(x)|$$

Where supx is the supremum of the set of distances and A(x) and B(x) are the empirical distribution functions for the samples. This is a 2-sided test for the null hypothesis where the 2 samples are drawn from the same continuous distribution.

The null hypothesis can be rejected when the p-value is less than the significance level.



We can reject the null hypothesis at 40%

Core Conclusions

NO EVIDENCE FOR INTERNAL ROTATION IN THE REMNANT CORE OF THE SAGITTARIUS DWARF!
See: Peñarrubia, J., Zucker, D.B., Irwin, M.J. and Hyde, E.A. et al. (2011ApJ...727L...2P)

Decrease in mean metallicity from −0.59 dex to −0.97 in Sgr Stream →TIDAL DISRUPTION FROM A PROGENITOR DWARF HAD INTERNAL METALLICITY GRADIENT
See: E.A. Hyde et al. 2015

- Low metallicity stars throughout Sgr Dwarf core and stream.
- Expanding our characterization into the stream and combining with future GALAH observations will lead to a unprecedented level of characterization for this system.

What’s Next?

The parameter space we use for this study includes 2MASS, the modified RAVE pipeline log(g), Teff, [m/H] metallicity, and velocity.

Expanded Parameters:
→The ppmxl (Roeser et al. 2010) catalog, the Sloan Digital Sky Survey (SDSS) (York et al. 2000) observations, the GALAH survey, Skymapper, and more!

Smarter selections:
→Use additional statistical tests to check for population differences

Faster Computing/Larger Datasets on Cloud

The End?

Dr. Elaina A. Hyde
Western Sydney University, Adjunct Fellow
Consultant/GC Engineer, Servian
Twitter @AstroHyde
elaina.hyde@servian.com.au

Why Cloud Platforms?

- On-premise to cloud migration and integration
- Cloud-to-cloud migration and integration
- Cloud data warehouse
- Big data platform (on the cloud)
- Big data analytics (on the cloud)
- Native cloud dev and micro-services
- Container orchestration (kubernetes, mesos)
- Containers (docker, packer)

Durability, reliability, global reach

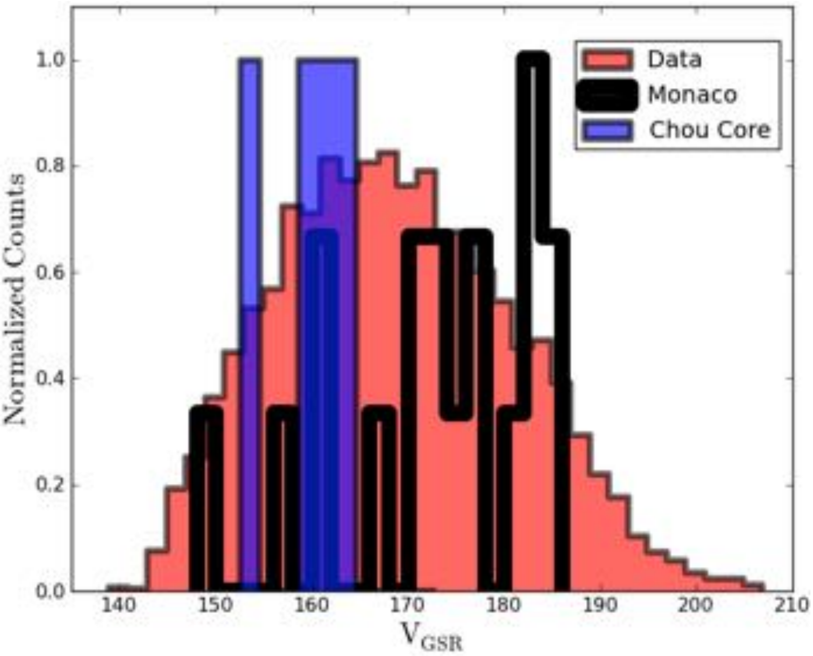
Distributed Computing, Networking, Data Storage, Compute Power

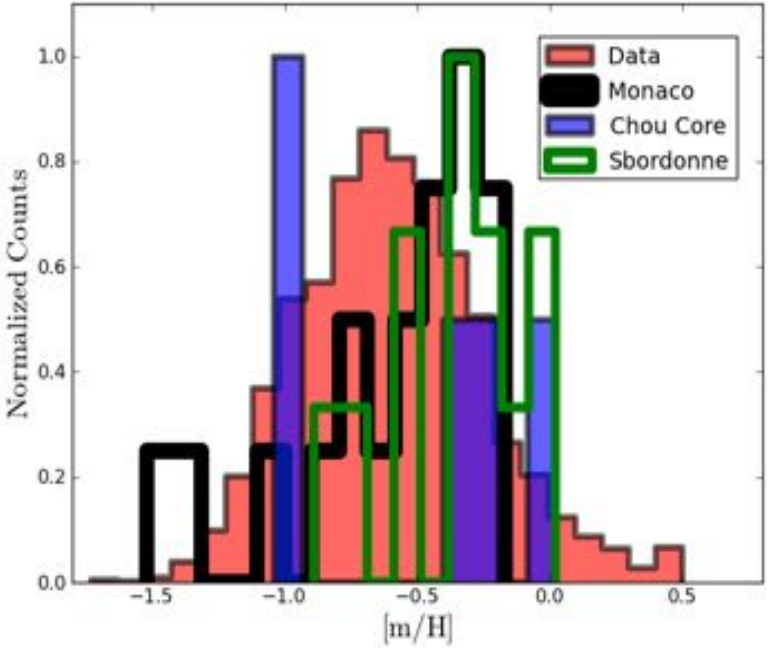
cloud + technology

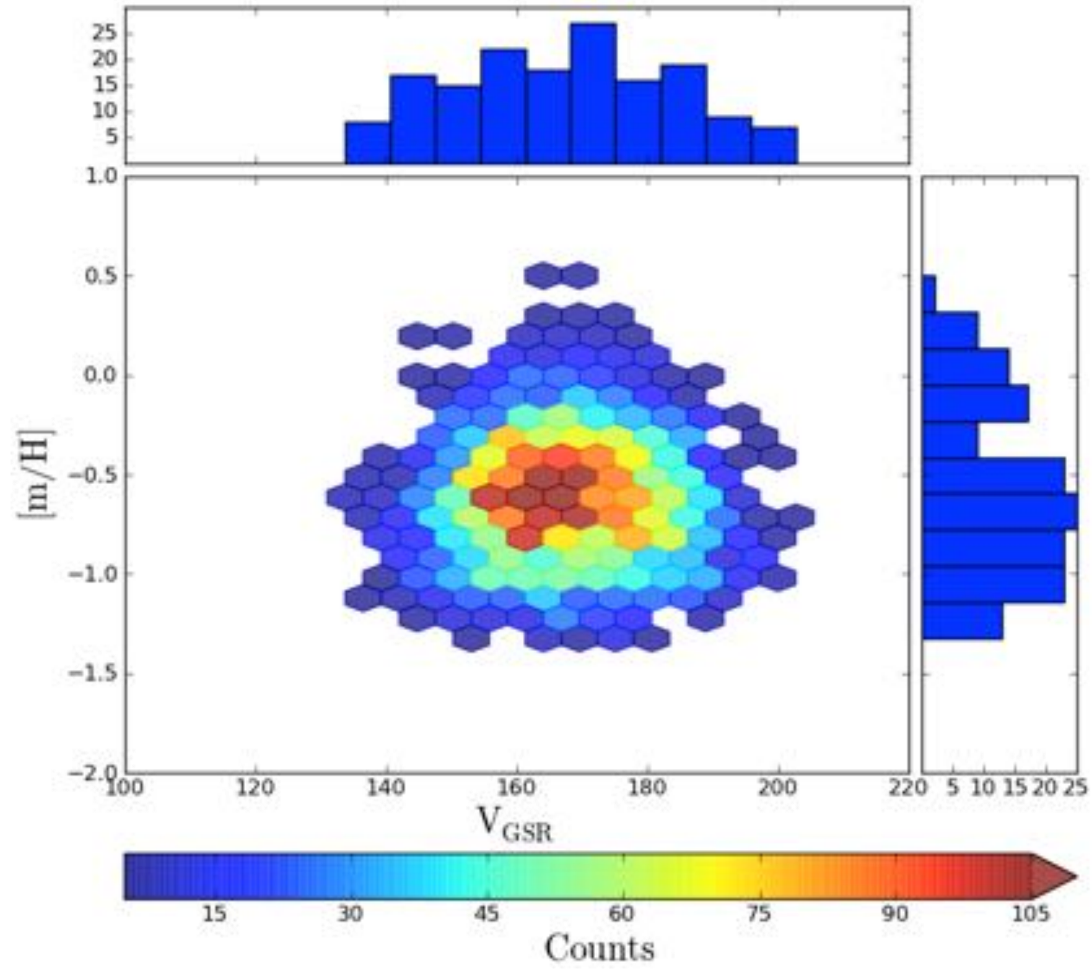
Servian has a proven track record with cloud migrations, cloud native development, DevOps and data/analytics platforms across all major cloud providers. Servian deliver customer outcomes with a unique set of people, accelerators and reference models which lowers the risk of transition and maximises capability outcomes on cloud infrastructure. The continuous commitment we have to training keeps our people and approach fresh in this rapidly changing area.

The Core MDF and VDF

⟨[Fe/H]⟩ ≃ ⟨[m/H]⟩ ≃ −0.59 dex with a standard deviation of σ[m/H] ≃ 0.34 dex
VGSR=168.76 km s−1 with a variation of σall= 12.2 km s−1.





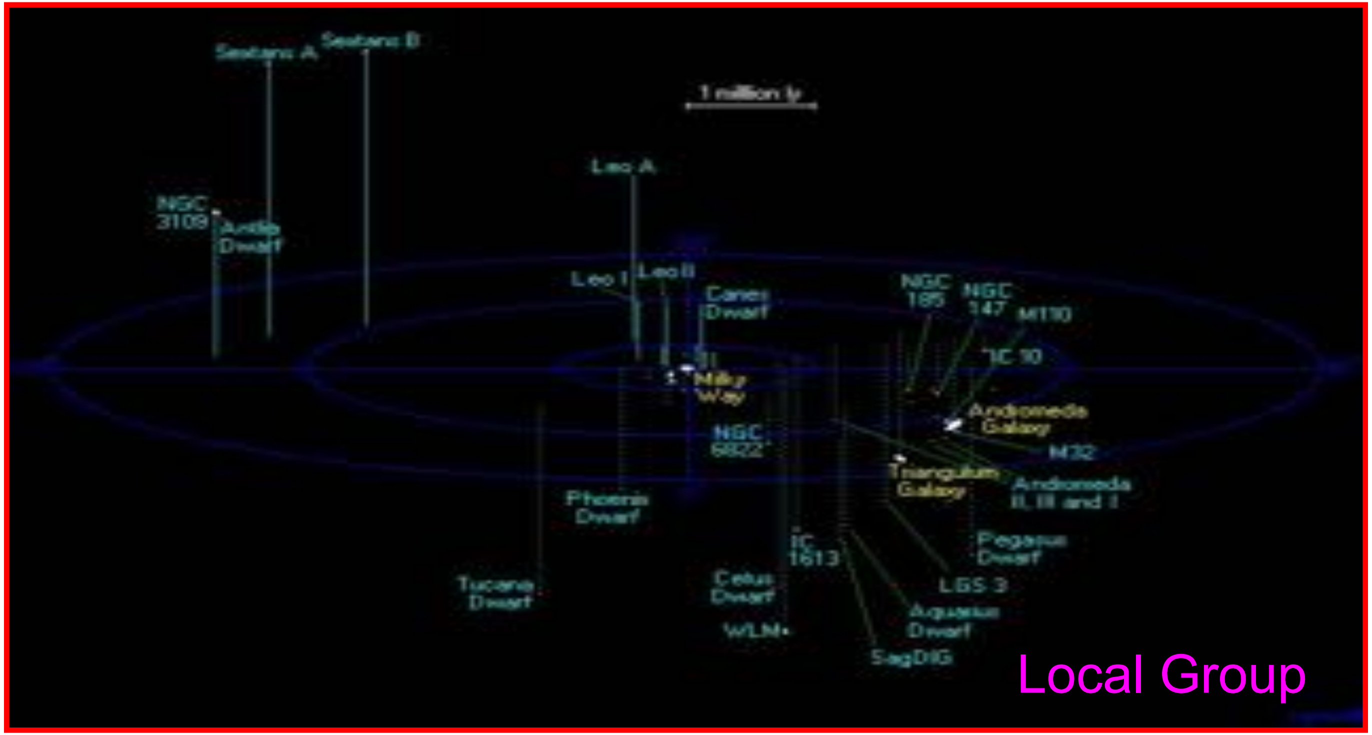


25
20
15
10
5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
[m/H]
100
120
140
160
180
200
220
5 10 15 20 25
V_GSR
15
30
45
60
75
90
105
Counts

The Local Group of Galaxies

The Milky Way
M31, M33
30+ Dwarf Galaxies

Theory: Hierarchical Scenario (small + small ➔bigger…)



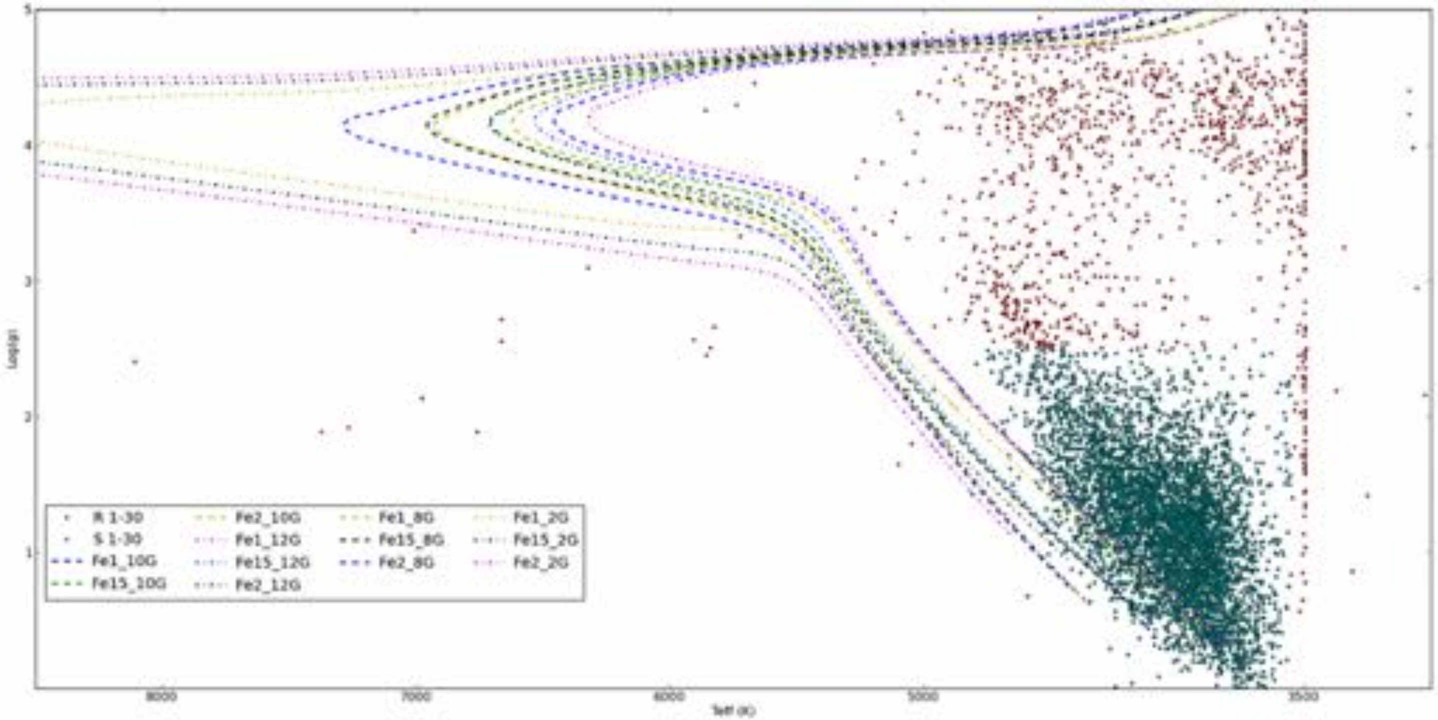
Richard Powell: http://www.atlasoftheuniverse.com/

The Data: 24,110 stars

Table Browser for 2: 15g.dat

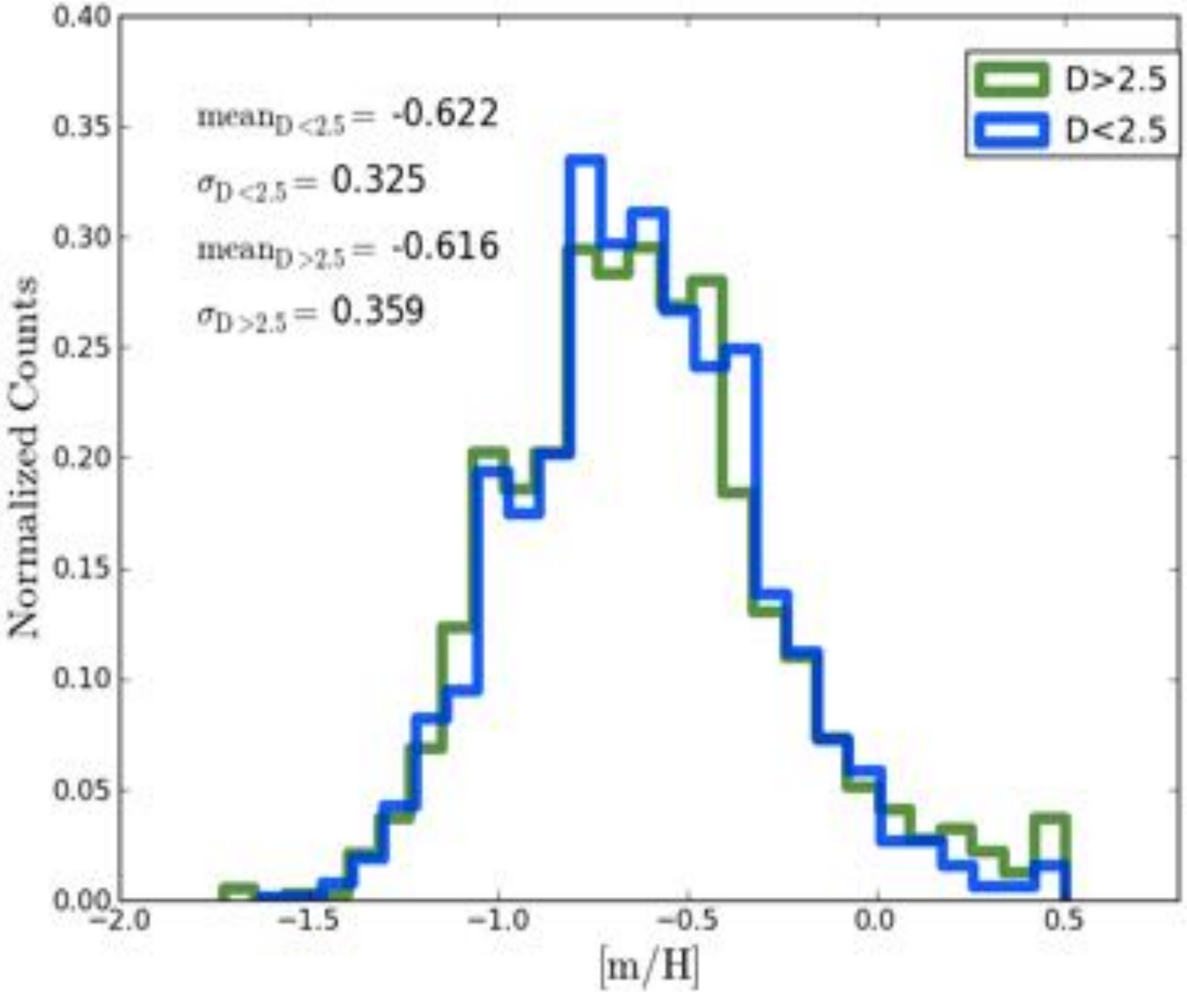
	num	object	Radeg	Decdeg	Mag3	Vhelavg	Verravg	FEHplavg	FEHmlavg	loggavg	Teffavg	mhavg	alphafeavg	Rcoeffavg
1	1	P_RGB_00475	152.6877	-2.1617	16.61	-9.02	4.12	-1.579	-1.0942	1.76	4000.	0.01	0.	15.62
2	1	P_RGB_00408	152.6865	-2.0027	17.06	107.18	6.79	-0.6373	-0.6182	2.5	3751.	0.42	0.	16.09
3	1	P_RGB_00190	152.6064	-1.5496	17.66	57.93	4.96	-1.1534	-1.1182	2.31	4192.	0.18	0.2	20.66
4	1	P_RGB_00199	153.2672	-1.2466	18.52	51.77	6.65	-0.2335	0.0469	1.81	4012.	0.01	0.	28.13
5	1	P_RGB_00106	153.504	-1.1467	17.07	5.08	2.69	-1.1728	-1.1111	2.47	4235.	0.2	0.	21.74
6	1	P_RGB_00257	155.5742	-0.6652	17.51	-10.65	8.09	-1.5653	-1.504	2.38	4046.	0.05	0.14	17.25
7	1	P_RGB_00054	154.7795	-0.9945	17.22	42.05	7.08	-0.5285	-0.5511	0.94	4000.	0.24	0.26	15.45
8	1	P_RGB_00061	154.6309	-0.9957	17.29	90.39	7.86	-0.207	-1.3191	2.24	3667.	0.46	0.	33.26
9	1	P_RGB_00017	154.635	-0.7979	17.65	114.74	7.3	-0.7747	-0.6932	1.43	4187.	0.12	0.14	20.42
10	1	P_RGB_00044	154.5355	-0.8506	17.	35.35	5.58	-0.7902	-0.8775	1.59	3863.	0.15	0.16	19.07
11	1	P_RGB_00173	154.0972	-0.6196	18.39	339.98	6.05	-0.5823	-0.5773	2.18	3540.	0.41	0.	31.75
12	1	P_RGB_00365	153.7972	-0.4404	18.18	34.75	7.96	-0.6867	-1.5182	2.5	3922.	0.5	0.23	25.09
13	1	P_RGB_00106	154.2649	-0.4265	18.42	94.79	8.11	-0.7106	-0.9692	1.96	4613.	0.44	0.35	24.58
14	1	P_RGB_00174	154.4399	-0.0324	18.11	42.72	4.97	-0.6693	-0.8705	1.97	4191.	0.23	0.11	17.12
15	1	P_RGB_00162	154.567	-0.0119	17.26	7.85	2.56	-1.0619	-1.0122	2.48	3697.	-0.21	0.21	20.27
16	1	P_RGB_00332	154.821	0.3238	17.36	13.53	7.73	-0.7749	-1.1133	1.77	4111.	0.	0.2	21.41
17	1	P_RGB_00379	155.049	0.3422	18.36	-68.78	4.87	0.1052	-0.293	1.43	3873.	-0.09	0.07	23.01
18	1	P_RGB_00214	155.2202	-0.0409	17.11	60.93	3.98	-1.1579	-1.157	2.31	4482.	-0.24	0.27	18.88
19	1	P_RGB_00103	155.1493	-0.2604	18.54	121.48	6.25	-1.3115	-1.1952	1.95	4512.	0.5	0.	16.78
20	1	P_RGB_00373	155.702	-0.2999	18.39	193.4	14.43	-0.9541	-1.6007	2.02	3642.	0.4	0.	33.65
21	1	P_RGB_00066	152.1314	-2.6482	16.83	56.69	7.3	-1.1556	-1.2001	2.48	4157.	0.13	0.	31.08
22	1	P_RGB_00230	152.4869	-2.9469	16.98	29.2	6.03	-0.8784	-0.8364	2.01	4009.	0.26	0.01	17.38
23	1	P_RGB_00222	152.451	-2.9613	16.52	-0.42	3.35	-1.5719	-1.3488	2.31	4177.	0.03	0.11	17.97
24	1	P_RGB_00126	152.1605	-2.9912	18.08	54.01	8.3	0.1044	0.5909	1.73	3609.	0.44	0.	26.39
25	1	P_RGB_00069	151.6405	-2.9676	18.09	55.94	9.53	-0.7831	-0.8371	1.63	3842.	-0.	0.14	27.68
26	1	P_RGB_00159	151.3043	-3.0411	17.42	22.06	6.74	-0.6889	-0.87	1.83	4216.	0.24	0.11	21.31
27	1	P_RGB_00086	151.3992	-2.869	17.64	38.05	6.5	-0.9208	-1.141	2.	4122.	0.25	0.	29.53
28	1	P_RGB_00050	151.4427	-2.5294	18.6	197.7	23.03	-0.7631	-0.7274	0.92	3511.	0.48	0.39	21.62
29	1	P_RGB_00116	150.0709	-2.2744	17.72	47.05	9.27	-1.2039	-1.0597	2.14	3947.	0.17	0.14	18.71
30	1	P_RGB_00271	150.963	-2.2501	17.06	70.83	3.62	-1.5649	-1.6606	2.49	3853.	0.5	0.21	24.29
31	1	P_RGB_00249	151.0799	-2.1303	17.07	12.97	2.77	-1.1834	-1.2028	2.48	4126.	-0.01	0.	21.04

Log(g) $\leq$ 2.5 and 3500 $\leq$Teff$\leq$ 5000



2MASS color cuts as defined by Majewski et al. (2004)

Dartmouth Isochrones: (Dotter et al. 2008)



$\mathrm{mean}_{D<2.5}$ = -0.622
$\sigma_{D<2.5}$ = 0.325
$\mathrm{mean}_{D>2.5}$ = -0.616
$\sigma_{D>2.5}$ = 0.359
D>2.5
D<2.5
Normalized Counts
[m/H]
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−2.0
−1.5
−1.0
−0.5
0.0
0.5

$$D_{\mathrm{center}} = R = \sqrt{x^2 + y^2}$$

$$x = (\cos(D_{\mathrm{core}}) \cdot \sin(R_s - R_{\mathrm{core}}))/Cc$$

$$y = ((\cos(D_s) \cdot \sin(D_{\mathrm{core}})) - (\sin(D_s) \cdot \cos(D_{\mathrm{core}}) \cdot \cos(R_s - R_{\mathrm{core}})))/Cc$$

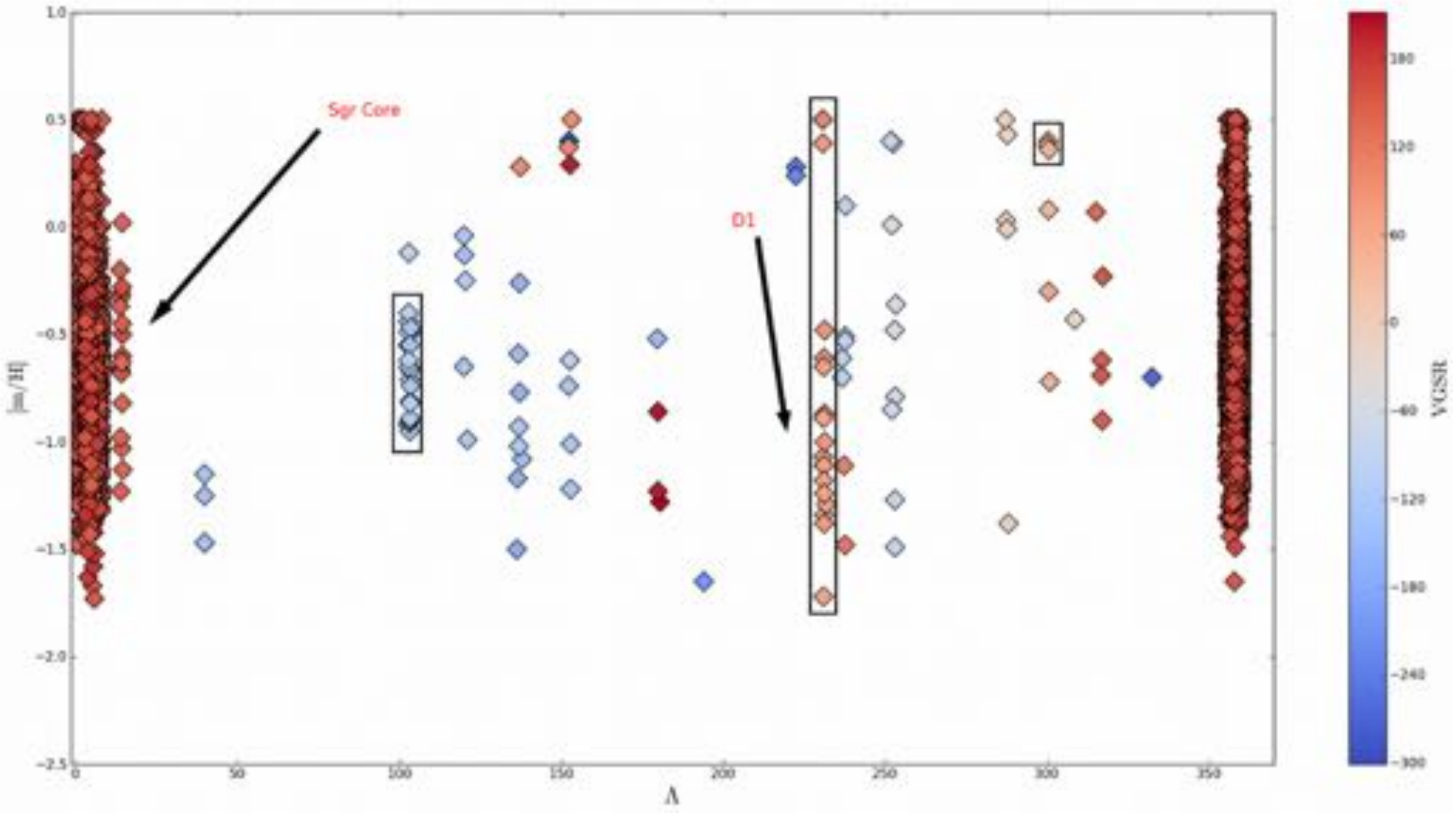
$$f = N_p/N_{\mathrm{all}}$$

$$Cc = (\sin(D_s) \cdot \sin(D_{\mathrm{core}})) + (\cos(D_s) \cdot \cos(D_{\mathrm{core}}) \cdot \cos(R_s - R_{\mathrm{core}}))$$

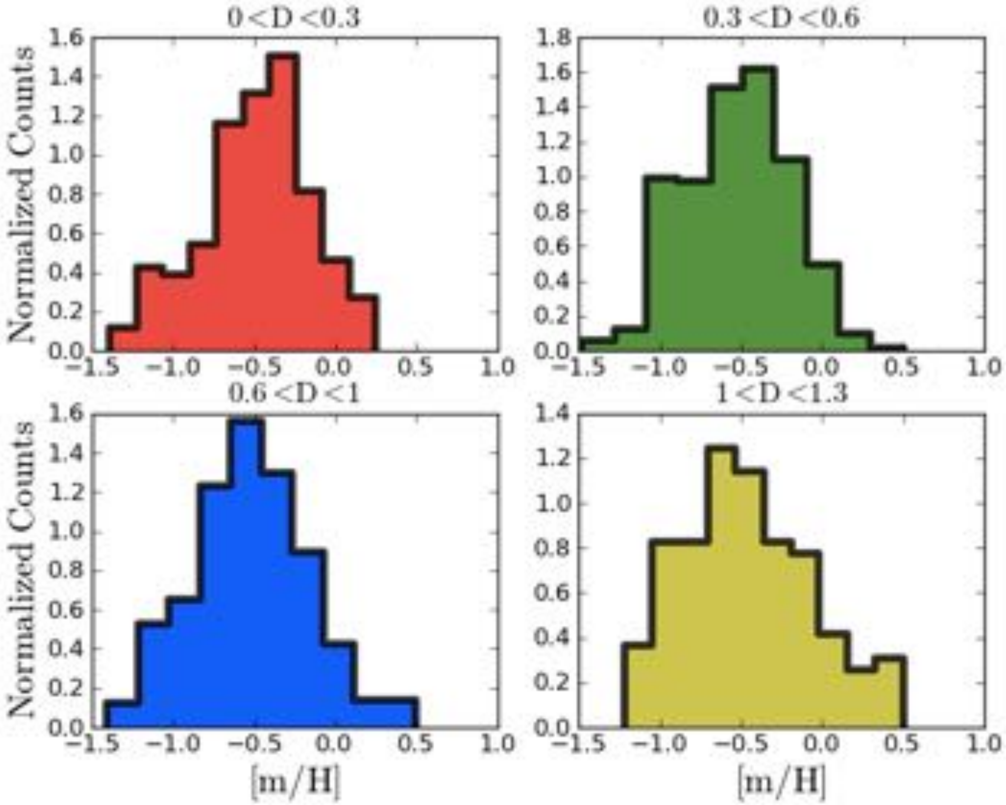
$$f = N_p/N_{\mathrm{all}}$$

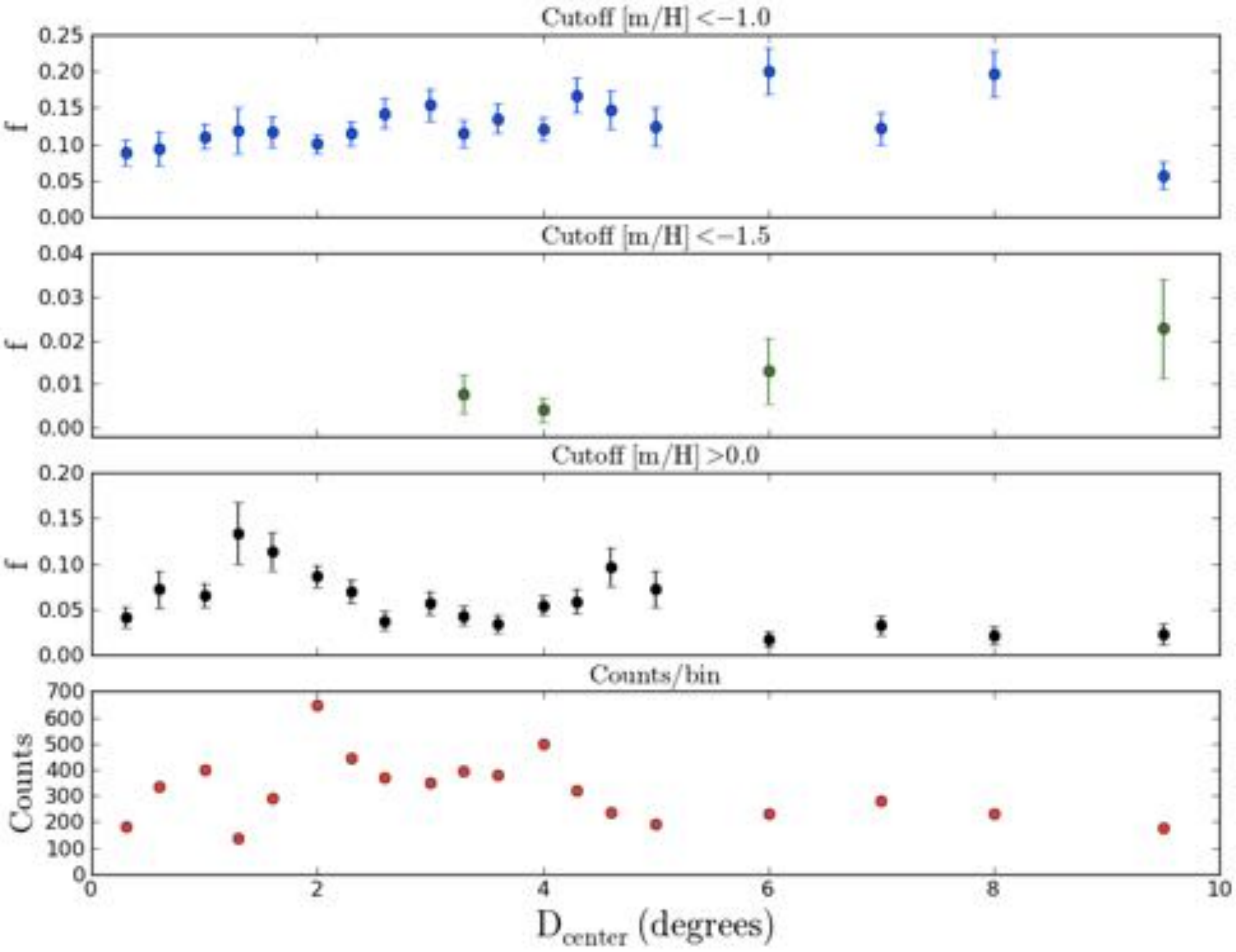
$$f_{\mathrm{err}} = f \cdot \sqrt{\left(\sqrt{N_p}/N_p\right)^2 + \left(\sqrt{N_{\mathrm{all}}}/N_{\mathrm{all}}\right)^2}.$$

Grouping



M54 Cluster in the Core?





Cutoff [m/H] <−1.0
Cutoff [m/H] <−1.5
Cutoff [m/H] >0.0
Counts/bin
f
Counts
D_{center} (degrees)

Galactic Archaeology Questions:

-Stellar streams and satellites in Local Group halos: relics of galaxy formation

-Properties of stream progenitors?

-Depth/shape of main galaxy's potential?

-Surviving satellites: is there a lower mass limit for galaxies?

-Are there numerous subhalos with low stellar content ("missing satellites")?

Contact Us

Sydney
Level 36, 60 Margaret Street
Sydney NSW 2000
+61 2 9376 0700

Melbourne
Level 11, 45 William Street
Melbourne VIC 3000
+61 3 9081 3700

Adelaide
The Hub, Suite 10, 5 Peel Street
Adelaide SA 5000
+61 414 458 763

Brisbane
Level 3, 200 Mary Street
Brisbane QLD 4000
+61 458 290 389

Managing Consultant
Andrew Pym
andrew.pym@servian.com
+61 414 971 680

Head of Digital & UX
Timothy Mannah
timothy.mannah@servian.com
+ 61 416 149 461

Head of Sales
Kate Voisey-Youldon
kate@servian.com
+61 406 040 338