Artificial Intelligence Assisted Galactic Archaeology

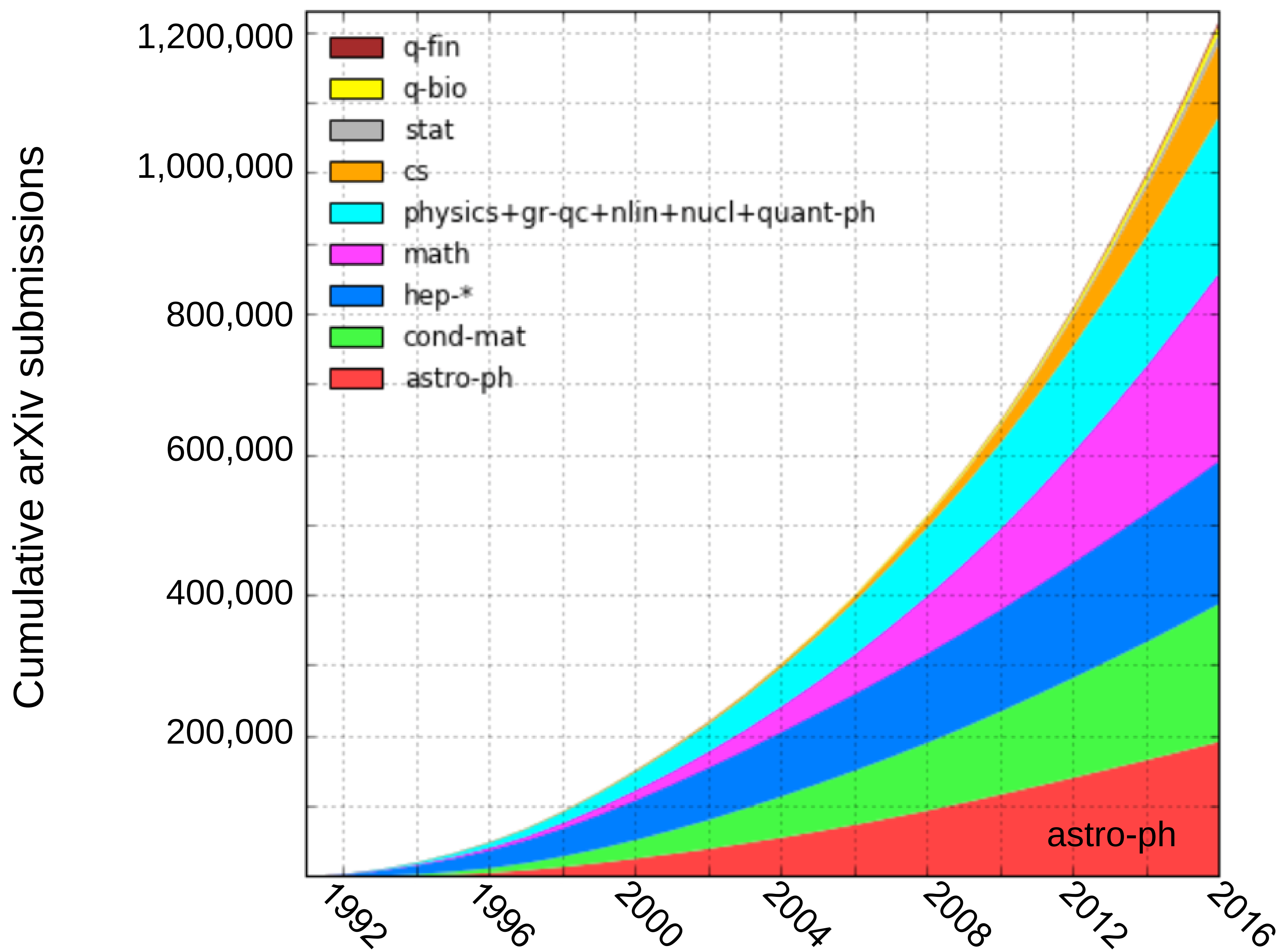
Student: Tom Crossland

Supervisor Team:

Daisuke Kawata (Primary), Tom Kitching, Jason McEwen
(MSSL, UCL)

Pontus Stenetorp, Sebastian Riedel
(Machine Reading group, Computer Science, UCL)

How can we reveal a global picture of the Milky Way from increasing number publications?



Measurement Extraction

ABSTRACT

We have compiled a new sample of 240 halo objects with accurate distance and radial velocity measurements, including globular clusters, satellite galaxies, field blue horizontal branch stars and red giant stars from the Spaghetti survey. The new data lead to a significant increase in the number of known objects for Galactocentric radii beyond 50 kpc, which allows a reliable determination of the radial velocity dispersion profile out to very large distances. The radial velocity dispersion shows an almost constant value of 120 km s^{-1} out to 30 kpc and then continuously declines down to 50 km s^{-1} at about 120 kpc. This fall-off puts important constraints on the density profile and total mass of the dark matter halo of the Milky Way. For a constant velocity anisotropy, the isothermal profile is ruled out, while both a dark halo following a truncated flat model of mass $1.2^{+1.8}_{-0.5} \times 10^{12} M_{\odot}$ and an NFW profile of mass $0.8^{+1.2}_{-0.2} \times 10^{12} M_{\odot}$ and c =18 are consistent with the data. The significant increase in the number of tracers combined with the large extent of the region probed by these has allowed a more precise determination of the Milky Way mass in comparison to previous works. We also show how different assumptions for the velocity anisotropy affect the performance of the mass models.

Key words: dark-matter – Galaxy: halo, dynamics, structure

https://arxiv.org/abs/astro-ph/0506102

Tabulated Data

Table 1 Parameters derived with the ARES+MOOG method for the provided test spectra.

File	Star	Teff (K)	log g (dex)	[Fe/H] (dex)	v_{tur} (km/s)
TestA.fits	HD128620	5832	4.33	0.23	1.11
TestB.fits	HD128621	5234	4.40	0.16	0.90
TestC.fits	HD179949	6287	4.54	0.21	1.36

https://arxiv.org/abs/1711.01839

Range Data, Units, etc.

ABSTRACT

Recent observations of Cepheids in the Virgo cluster have bolstered the evidence that supports a Hubble constant in 70 − 90 km sec^{-1} Mpc^{-1} range. This evidence, by and large, probes the expansion of the Universe within 100 Mpc. We investigate the possibility that the expansion rate within this region is systematically higher than the true expansion rate due to the presence of a local, large underdense region or void. We begin by calculating the expected deviations between the locally measured Hubble constant and the true Hubble constant for a variety of models. The calculations are done using linear perturbation theory and are compared with results from N-body simulations wherever possible. We also discuss the expected correlations between these deviations and mass fluctuation for the sample volume. We find that the fluctuations are small for the standard cold dark matter as well as mixed dark matter models but can be substantial in a number of interesting and viable nonstandard scenarios. However, deviations in the Hubble flow for a region of radius 200 Mpc are small for virtually all reasonable models. Therefore, methods based on supernovae or the Sunyaev-Zel'dovich effect, which can probe 200 Mpc scales, will be essential in determining the true Hubble constant.

We discuss, in detail, the fluctuations induced in the cosmic background radiation by voids at the last scattering surface. In addition, we discuss the dipole and quadrupole fluctuations one would expect if the void enclosing us is aspherical or if we lie off-center.

https://arxiv.org/pdf/astro-ph/9506120

Multi-Sentence Data

2.2. *Pulsar Timing Capabilities*

Pulsar timing capabilities are provided by the pulsar timing backend (PST), a sub-element of CSP. For pulsars with DM in the range 0 to 3000 pc cm^{-3} and for periods as fast as 0.4 ms, it can perform observations with durations up to 5 hours. The beamformer produces up to 16 polarisation-calibrated voltage beams for pulsar timing purposes (8 for SKA1-Mid Band 5). The bandwidth of these beams is the entire 300-MHz band for SKA1-Low; for SKA1-Mid it is the full band for Bands 1 and 2, and 2.5 GHz for Band 5a/b. These data are coherently dedispersed in PST and folded according to the known ephemerides. The process of 'timing' to check and update pulsar ephemerides, and react accordingly, is performed in SDP. One can also perform pulsar searches using the timing backend using the "dynamic spectrum mode", essentially a mode where PST doesn't fold the data, and reduces the time and/or frequency resolution to keep the data rate within limits, and with coherent dedispersion optional. This mode is appropriate for long targetted search pointings, e.g. Sgr A*, globular clusters, supernova remnants etc. Noteworthy, but on the negative side, is that Band 3 (S-band) is not covered in the initial construction budget — an upgrade focused on pulsar timing would no doubt prioritise this (Cordes *et al.* 2016; Shannon & Cordes 2017). In terms of timing precision, there is uncertainty in the common delay centre of each SKA sub-array in the amount of 2 ns (1σ), with the SKA timescale linked to UTC with an uncertainty of 5 ns (1σ) for SKA1-Mid; for SKA1-Low it is 10 ns (1σ). These effects combine in quadrature to give 5.4 ns and 10.2 ns for SKA1-Mid and -Low respectively. TOA uncertainty decreases with increasing S/N until this intrumental limit is reached, if the limit due to intrinsic pulse jitter (which is pulsar specific) is not reached first.

https://arxiv.org/abs/1711.01910

Conditional Data

ABSTRACT

We present observations with the Submillimeter Array of the continuum emission at $\boxed{\lambda = 1.3\ \mathrm{mm}}$ from 62 young stars surrounded by a protoplanetary disk in the Serpens star-forming region. The typical angular resolution for the survey in terms of beam size is $3.5'' \times 2.5''$ with a median rms noise level of 1.6 mJy beam^{-1}. These data are used to infer the dust content in disks around low-mass stars $\boxed{(0.1\text{–}2.5\,M_\odot)}$ at a median stellar age of 1–3 Myr. Thirteen sources were detected in the 1.3 mm dust continuum with inferred dust masses of $\approx$10–260 $M_\oplus$ and an upper limit to the median dust mass of $\boxed{5.1^{+6.1}_{-4.3}\,M_\oplus}$, derived using survival analysis. Comparing the protoplanetary disk population in Serpens to those of other nearby star-forming regions, we find that the populations of dust disks in Serpens and Taurus, which have a similar age, are statistically indistinguishable. This is potentially surprising since Serpens has a stellar surface density two orders of magnitude in excess of Taurus. Hence, we find no evidence that dust disks in Serpens have been dispersed as a result of more frequent and/or stronger tidal interactions due its elevated stellar density. We also report that the fraction of Serpens disks with $M_{\mathrm{dust}} \geq 10\,M_\oplus$ is less than 20%, which supports the notion that the formation of giant planets is likely inherently rare or has substantially progressed by a few Myrs.

https://arxiv.org/abs/1711.01266

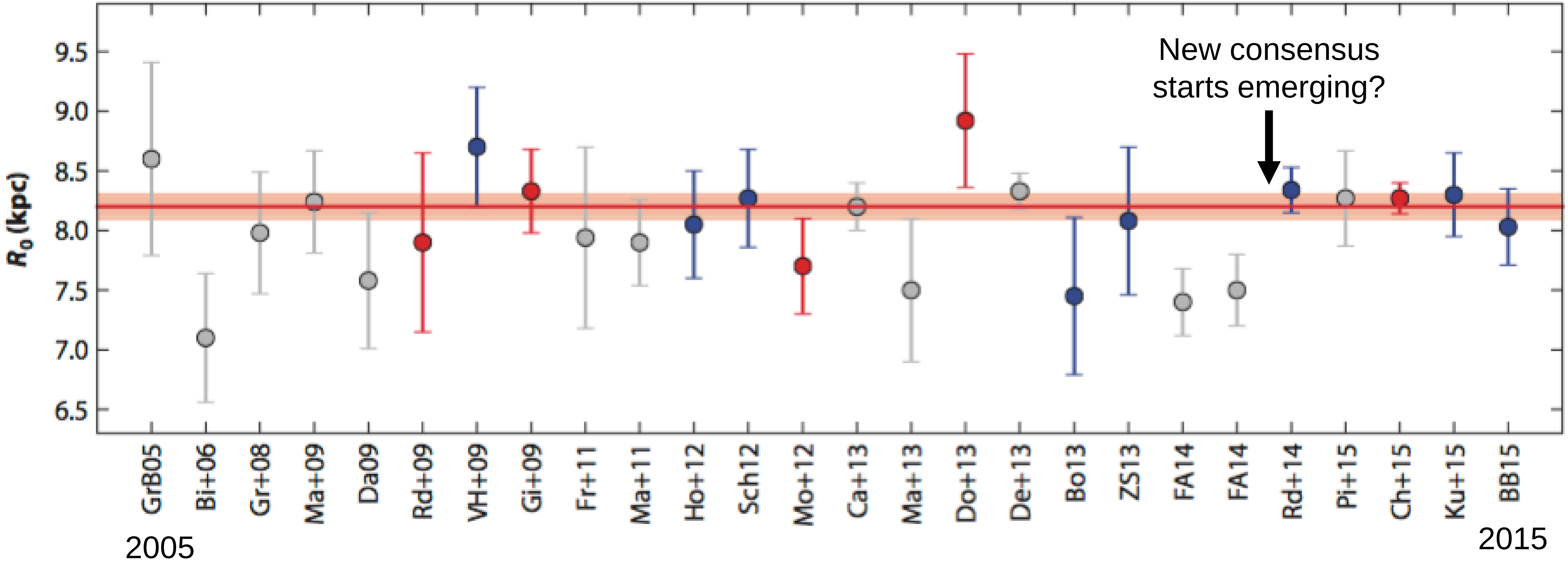
Units

5. Summary

In this work we combined interferometric observations at milliarcsec resolution in the mid-IR with a multi-wavelength dataset at subarcsec resolution. This approach allowed us to probe the physical scales associated with the nuclear emission in IC 3639. This Seyfert 2 galaxy shows a strong point-like source at the resolution of VISIR (~ 320 mJy within $\lesssim 340$ mas), but surprisingly most of the emission is spatially resolved in the MIDI data, showing correlated fluxes of only ~ 90 mJy, that is a factor 3 to 4 times lower when compared to the single dish measurements.

$$1\,Jy = 10^{-16}\,\mathrm{W}\,m^{-2}\,Hz^{-1} = 10^{-23}\,erg\,s^{-1}\,cm^{-2}\,Hz^{-1}$$

https://arxiv.org/abs/1711.01268



Last 10 years of published distance to the Galactic centre, R_0.
Bland-Hawthorn & Gerhard (2016)

Data Pre-Processing

- Extract `.tex` (and `.tab`, `.cls`, etc.)

- Convert `.tex` to `.txt`

  $\longrightarrow$ delatex

- Sentence split and tokenize

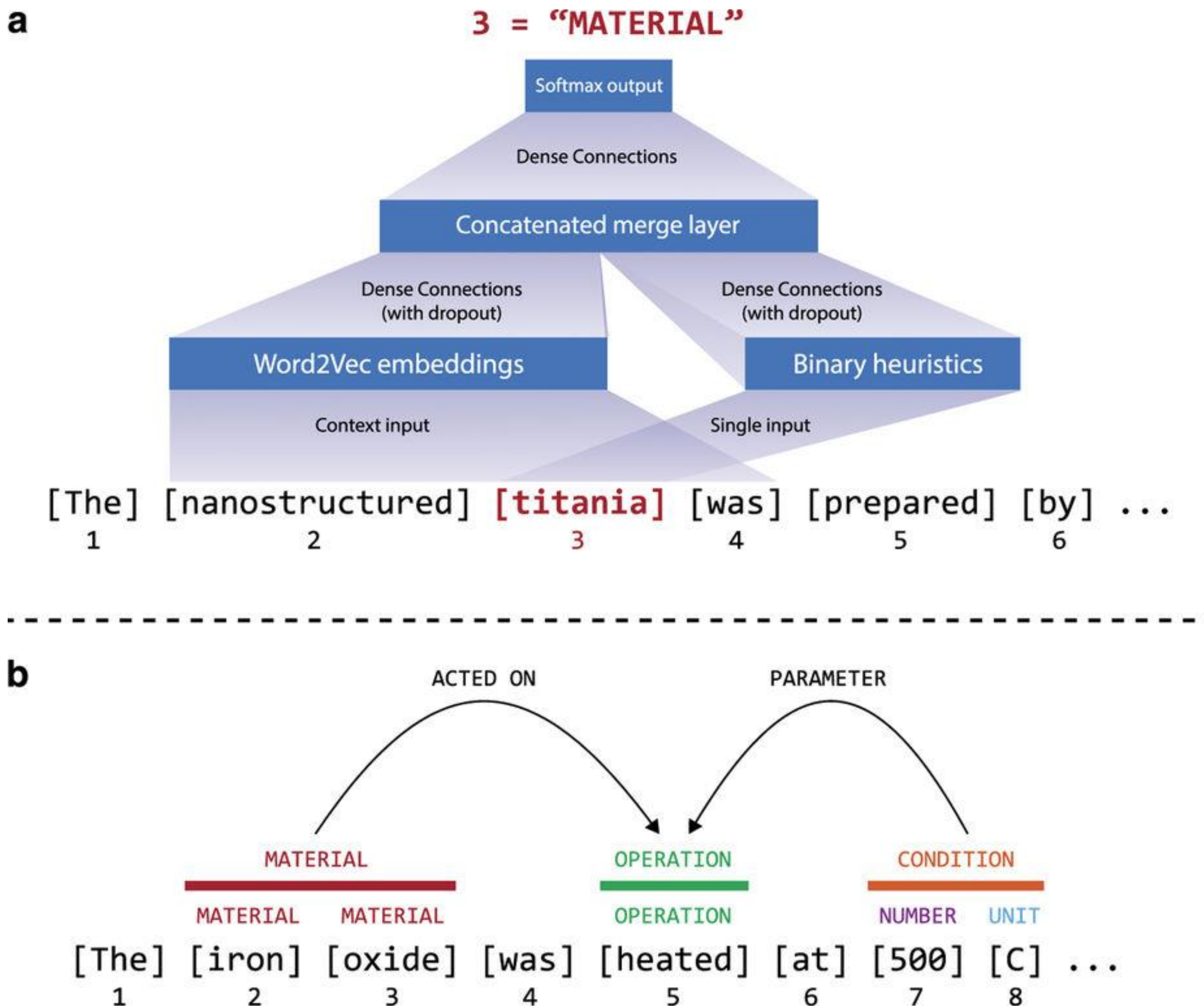
Next Steps

- Begin with rule-based system

  Identify closest numeric token to keyword

- Progress to grammar parsing (hopefully)

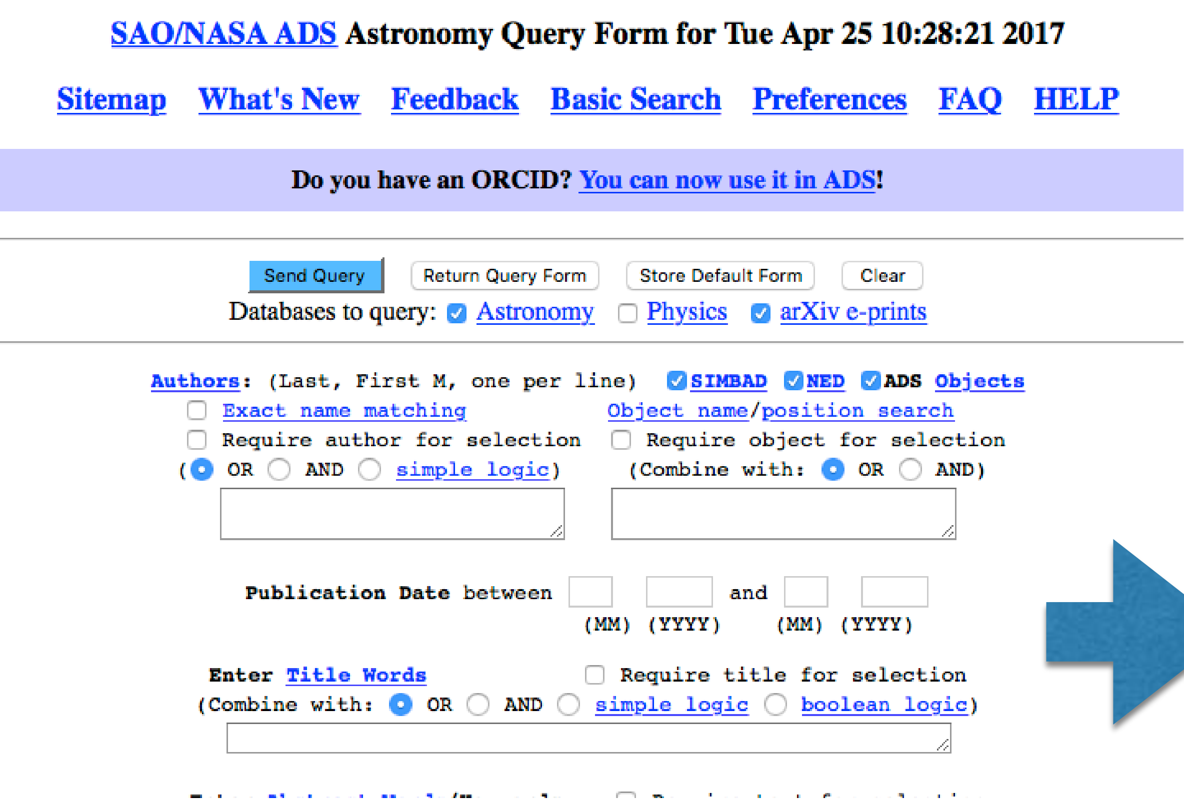
  Find “parameter” tokens and associated numeric token



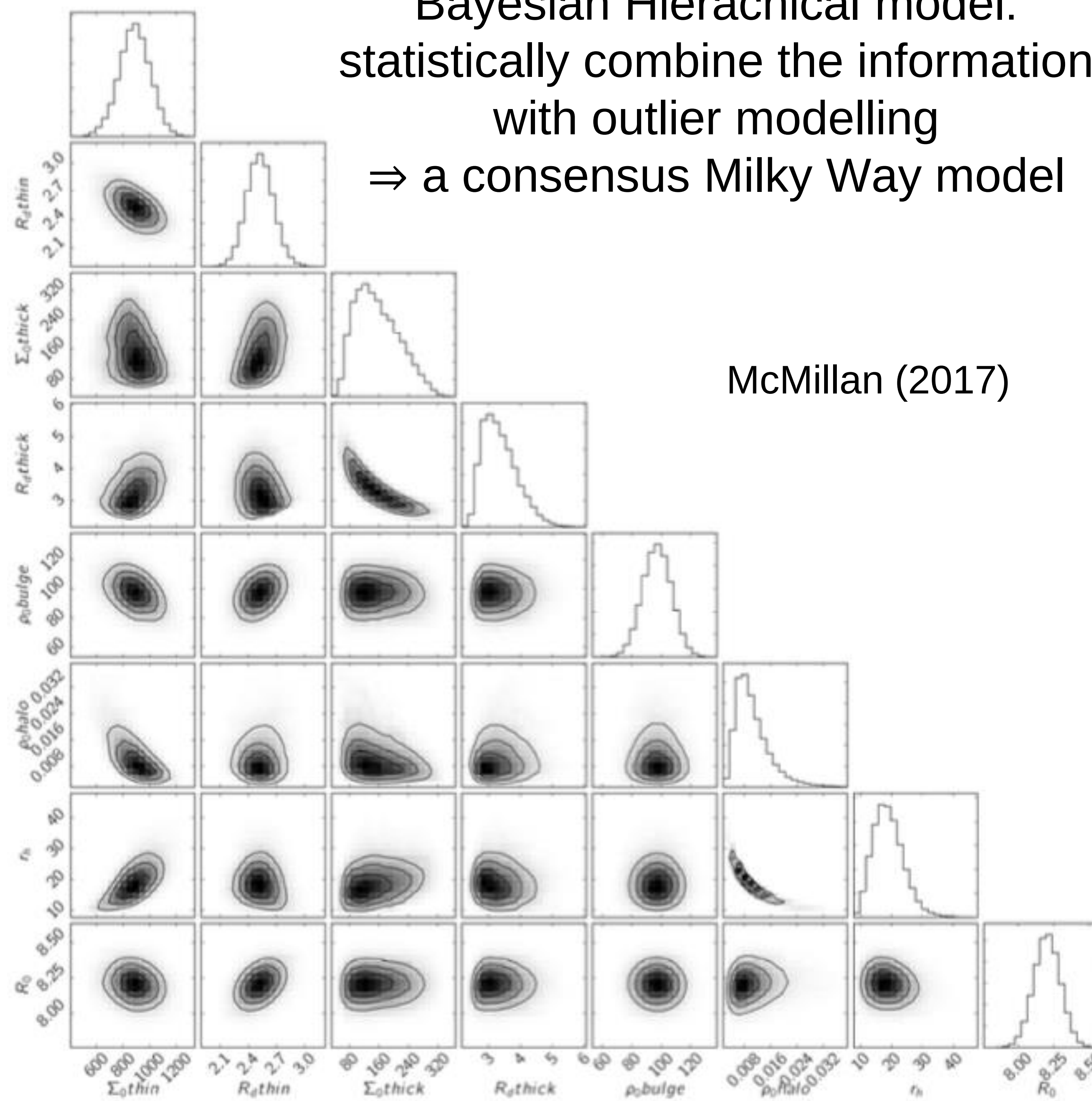
Kim, *et al.* (2017) Machine-learned and codified synthesis parameters of oxide materials, *Scientific Data,* ***4***

Fully-automated self-updating Galaxy model!

Machine Reading:
extracting measured values
by different methods and/or data
from all the published papers



Bayesian Hierachical model:
statistically combine the information
with outlier modelling
⇒ a consensus Milky Way model

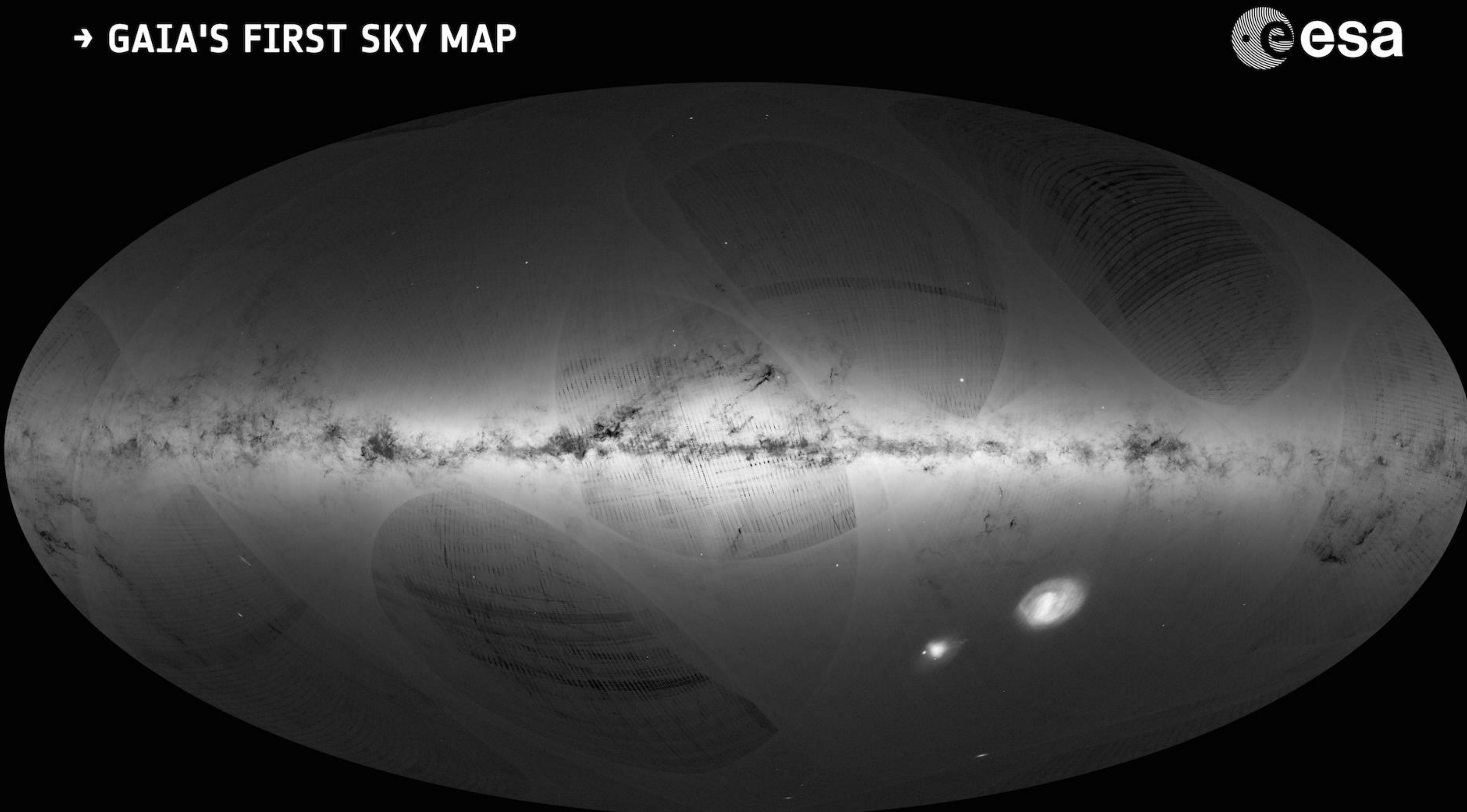


McMillan (2017)

w/ UCL Machine Reading Group!

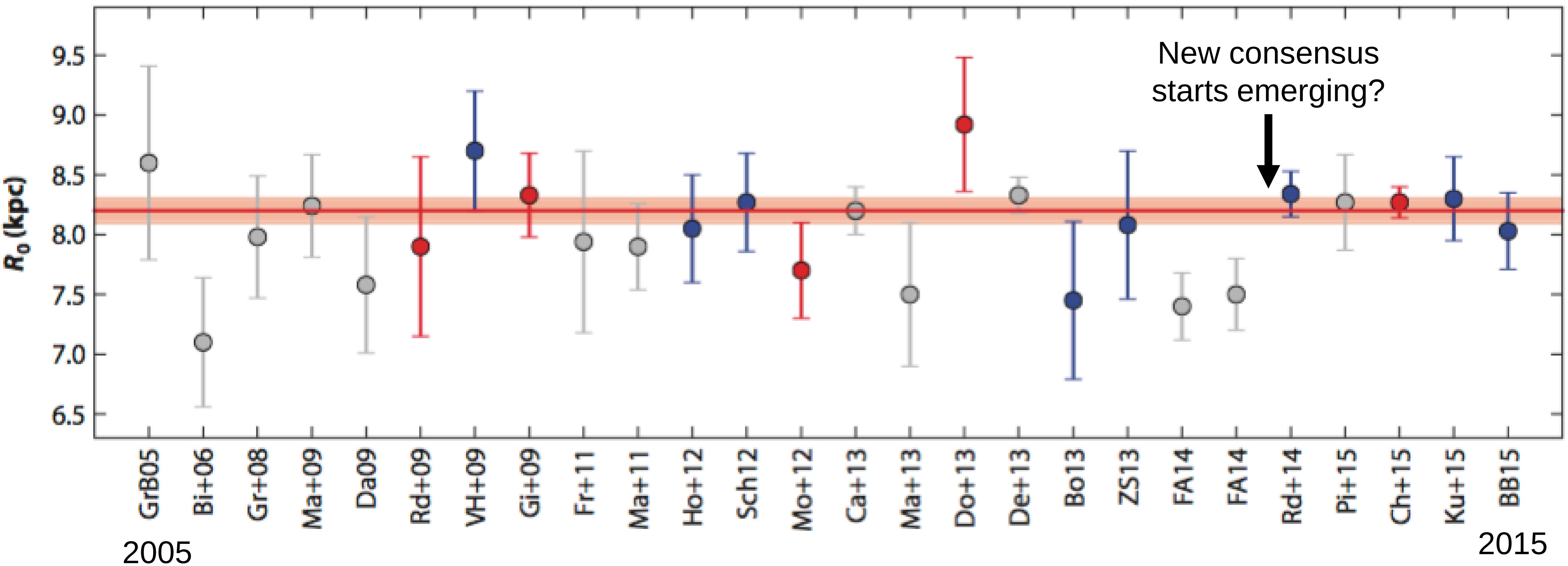

ESA’s Gaia satellite: Launched 19th Dec. 2013
mapping the positions and velocities of over 1 billion stars!
MSSL:
CCD and electronics
RVS (Radial Velocity Spectrograph)
Spectroscopic data processing

1st Data Release: 14 September 2016
~1/1000 of stars with 20 times lower accuracy,
still produced about 200 related papers in 8 months!



Credit: ESA/Gaia/DPAC

Comparing the Galaxy model with or without Gaia publications.
⇒ impact of the Gaia mission.
Identifying what kind of data and/or analysis made
stronger impact on the model? or new consensus?
⇒ what kind of the data are still missing after Gaia?



Last 10 years of published distance to the Galactic centre, R_0.
Bland-Hawthorn & Gerhard (2016)

Applicable to any research topic
to capture trends and holistic view of the research area!