UNIVERSITY OF TORONTO

The HERBAL model:

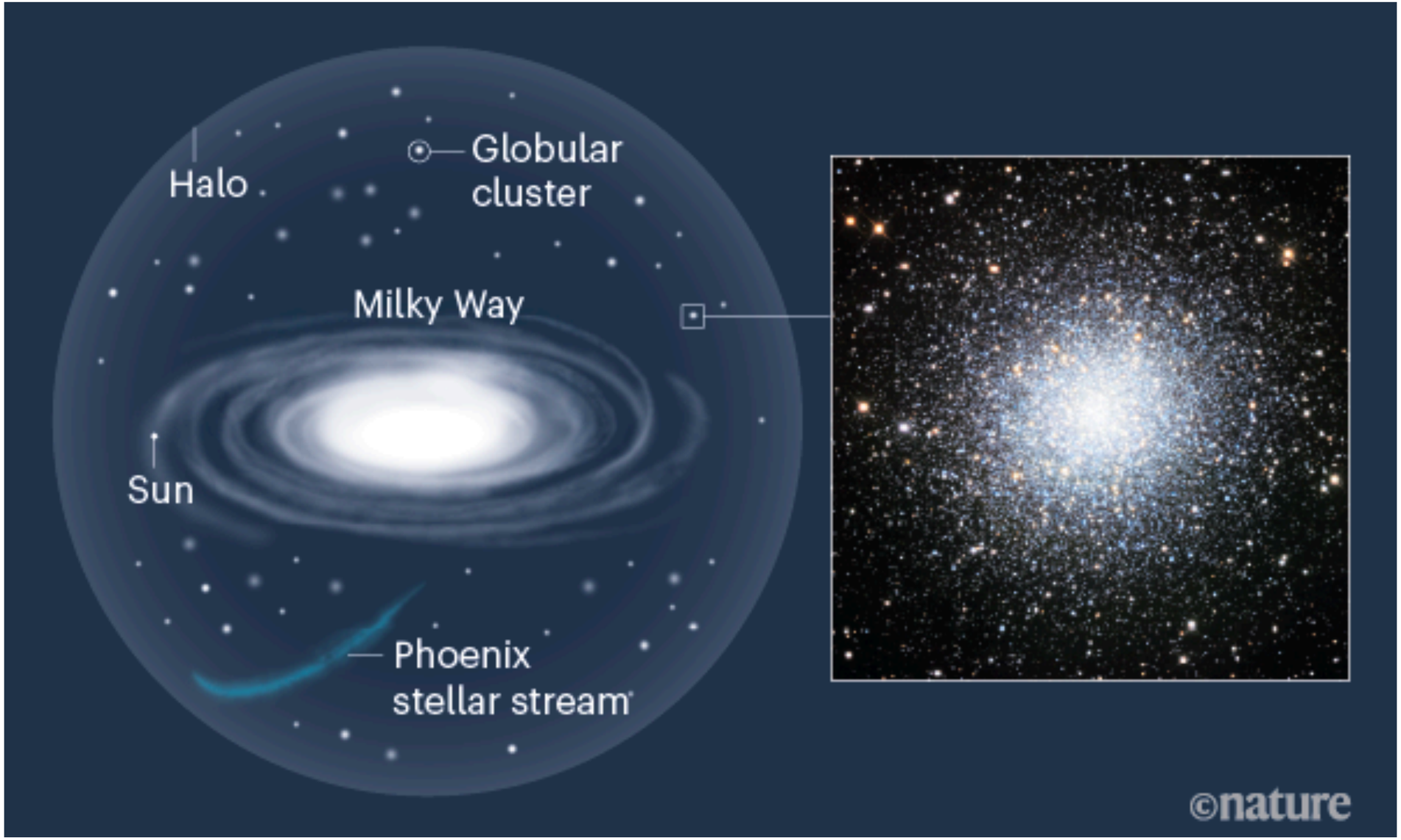
A hierarchical errors-in-variables Bayesian lognormal hurdle model for the galaxy mass — globular cluster system mass scaling relation

Sam Berek, with Gwen Eadie, Josh Speagle, and Bill Harris

BIRS astrostatistics workshop

October 31, 2023

Globular clusters (GCs) are large, dense clusters of stars with masses ~ $10^2 - 10^7 M_\odot$ (solar masses).



Globular clusters (GCs) are large, dense clusters of stars with masses ~ $10^2 - 10^7 M_\odot$ (solar masses).

GCs are found in **all but the smallest galaxies.**

Globular clusters (GCs) are large, dense clusters of stars with masses ~ $10^2 - 10^7 M_\odot$ (solar masses).

GCs are found in **all but the smallest galaxies.**

Globular clusters formed early in the universe and are some of the oldest stellar objects that exist.

Their formation and evolution is deeply intertwined with that of their host galaxies

NASA/ESA Hubble

Globular clusters (GCs) are large, dense clusters of stars with masses ~ $10^2 - 10^7 M_\odot$ (solar masses).

GCs are found in **all but the smallest galaxies.**

Globular clusters formed early in the universe and are some of the oldest stellar objects that exist.

Their formation and evolution is deeply intertwined with that of their host galaxies

Studying GC populations can tell us a lot about how galaxies form and evolve

NASA/ESA Hubble

Studying GC populations can tell us a lot about how galaxies form and evolve

Galaxy formation:

- Trace early star formation histories
- Determine gas feedback's influence on star formation
- Understand how galaxies assembled

Studying GC populations can tell us a lot about how galaxies form and evolve

Galaxy formation:

- Trace early star formation histories
- Determine gas feedback's influence on star formation
- Understand how galaxies assembled

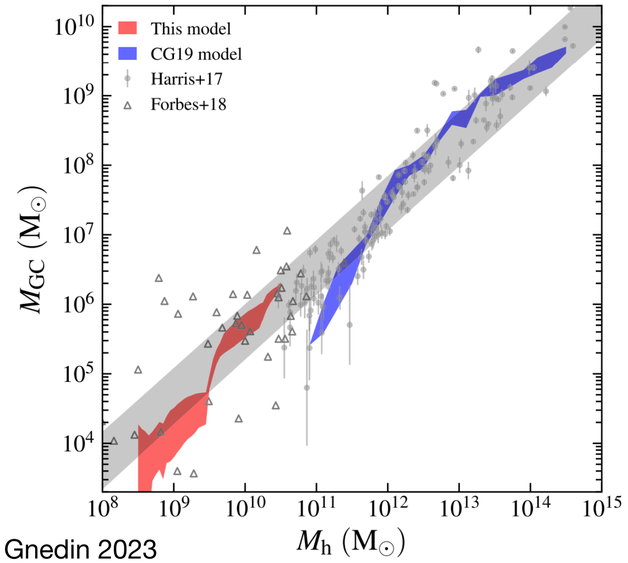
Galaxy evolution:

- Trace galactic potentials
- Trace galaxy merger histories
- Indicate intense, bursty star formation

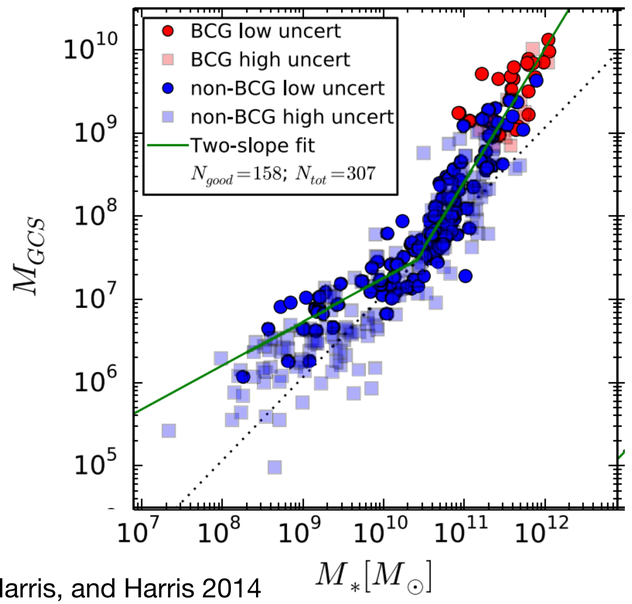
But how best can we study globular cluster systems?

One of the most fundamental relationships between a galaxy and its cluster system is the **scaling relation** between the **mass of the galaxy** and the **mass of its combined GC system**

The mass scaling relation

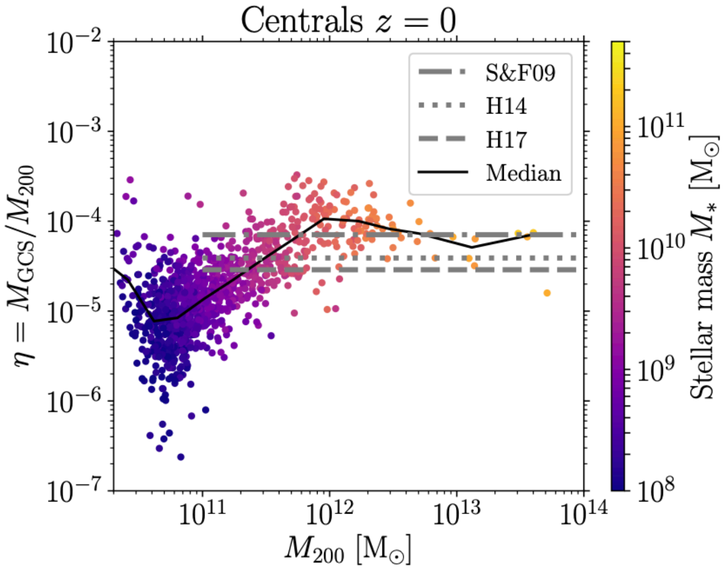


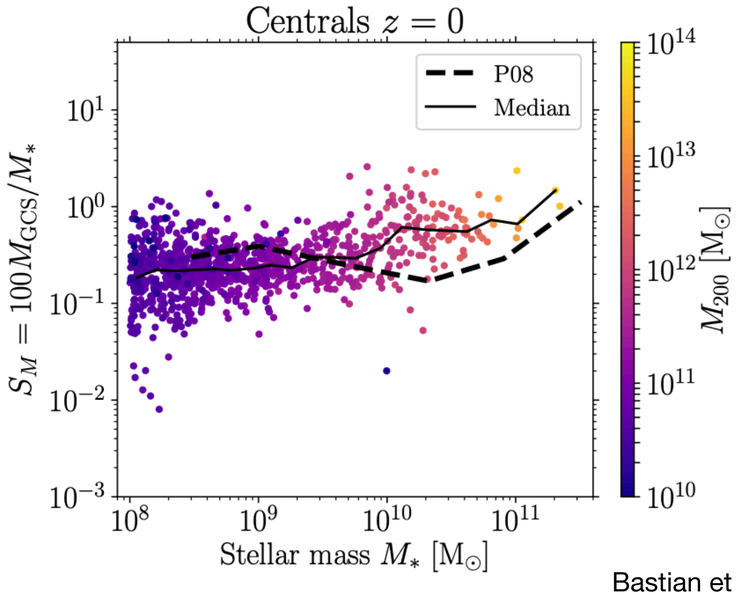
Chen and Gnedin 2023



Hudson, Harris, and Harris 2014

There is an approximately linear relationship between galaxy mass and GC system mass





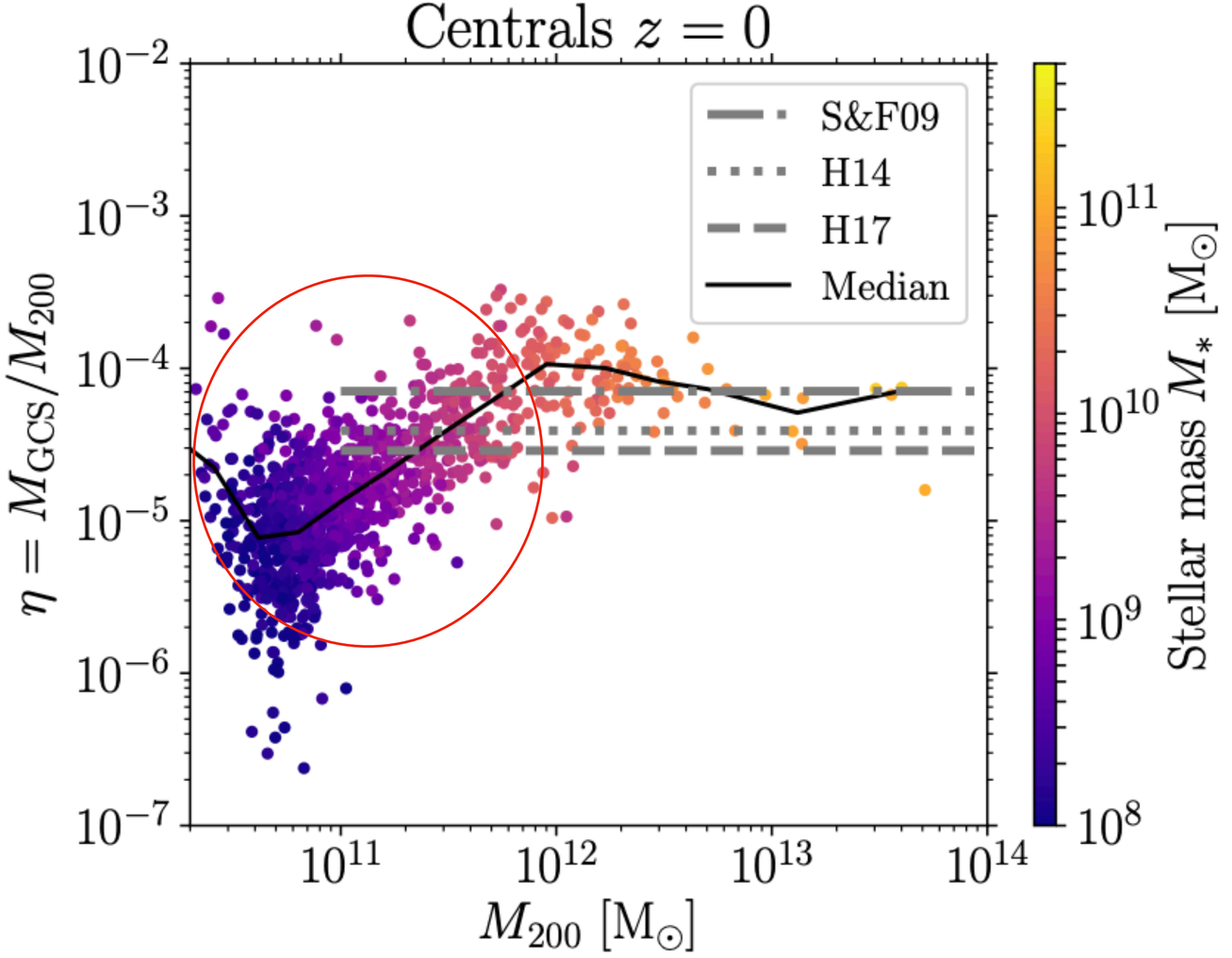
Bastian et al 2020

There is considerable uncertainty about the shape of the relation at the **low mass end**

There is considerable uncertainty about the shape of the relation at the **low mass end**

Some studies find the linear relationship continuing, others find a downturn.

WHY?



Bastian et al 2020

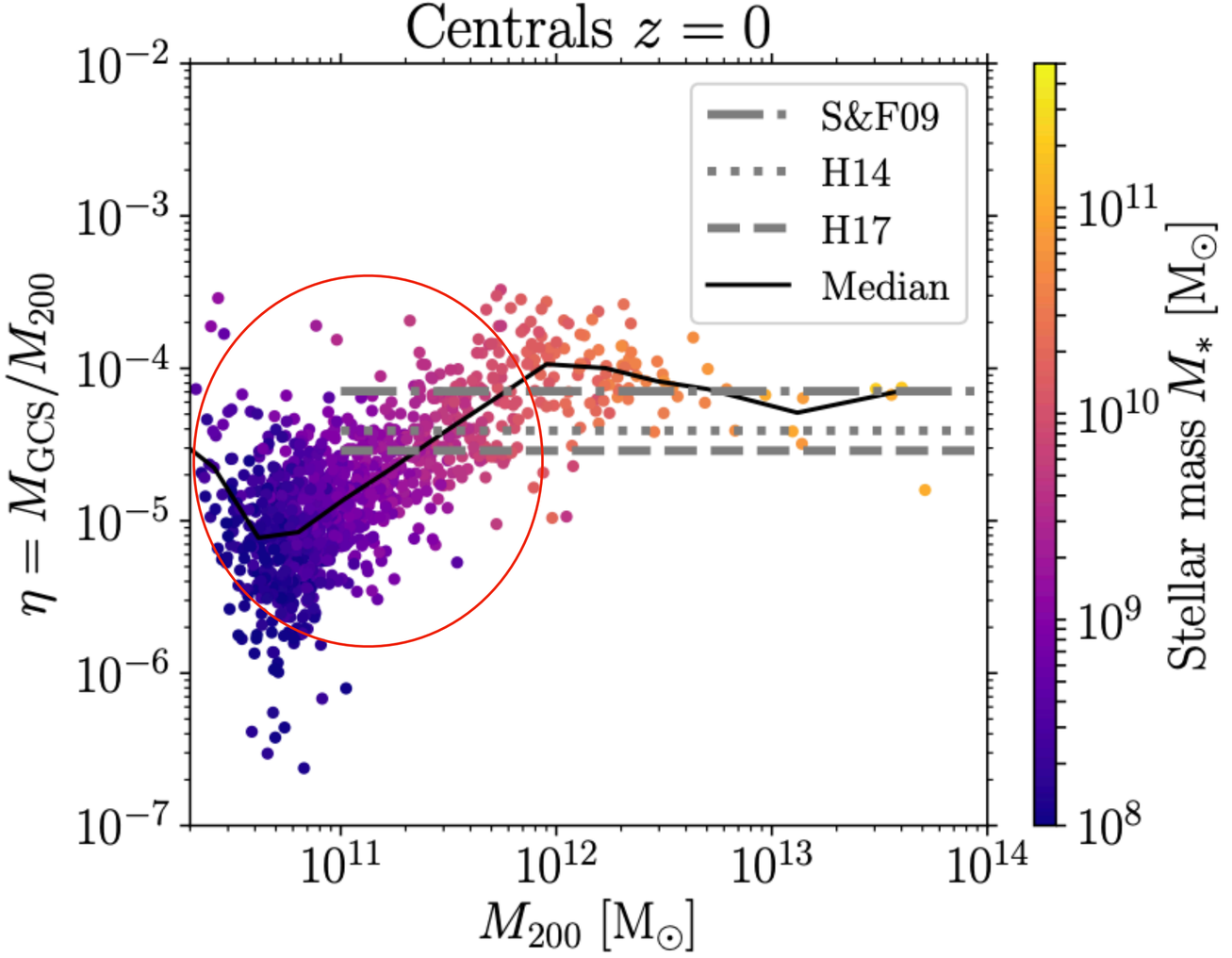
There is considerable uncertainty about the shape of the relation at the **low mass end**

Some studies find the linear relationship continuing, others find a downturn.

WHY?

1. Uncertainty in the stellar mass halo mass relation

2. Generally no inclusion of galaxies without GCs.

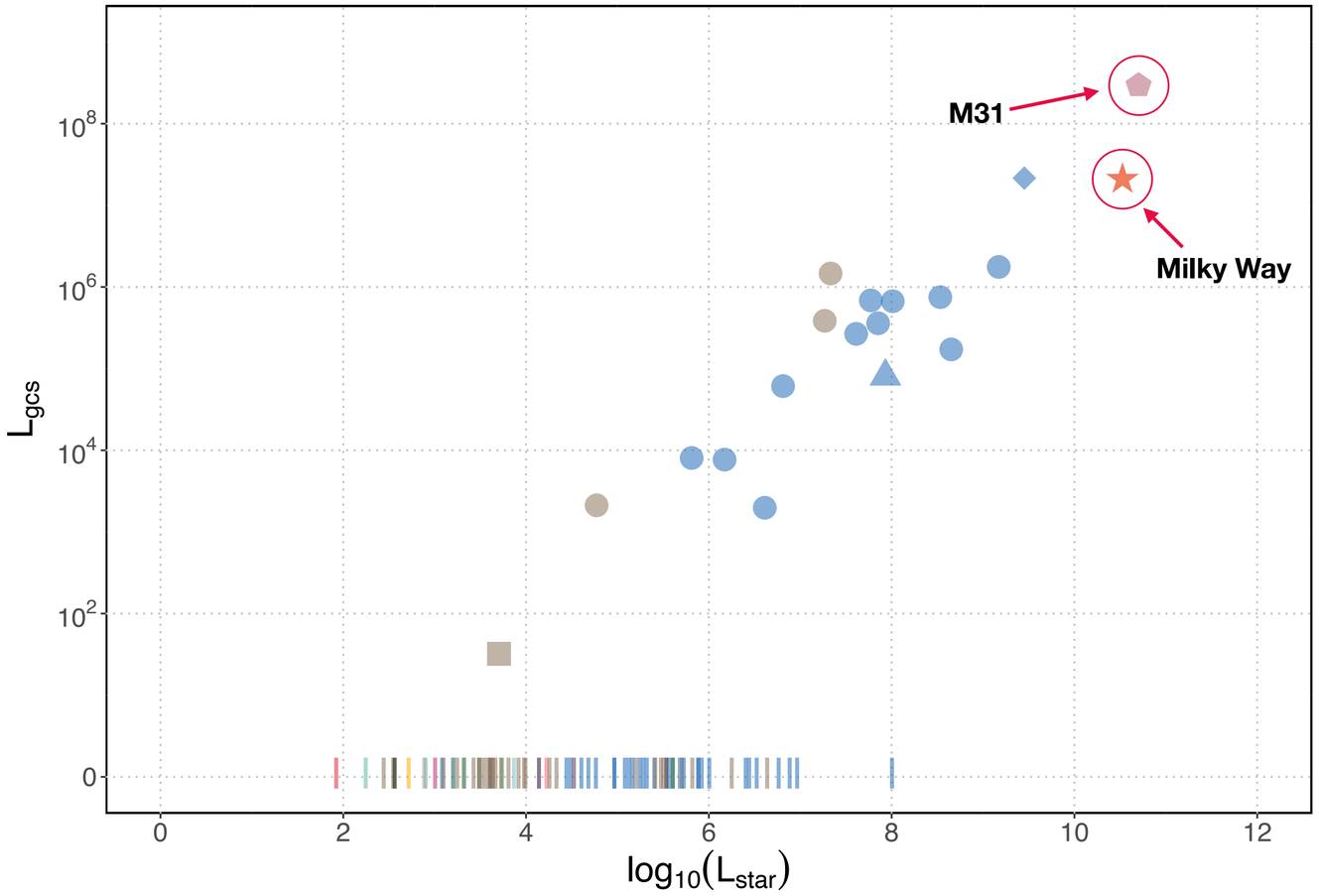


Bastian et al 2020

We use the Local Group to constrain the low-mass regime

The Local Group: The group of galaxies containing the Milky Way, Andromeda (M31), and many smaller galaxies

Closest galaxies to us -> most high quality GC data available



Available in Berek et al 2023: https://arxiv.org/abs/2306.14945

We aim to:

1. Extend the mass relation down to low mass, dwarf galaxies

2. Incorporate galaxies that don’t have GC populations

We aim to:

1. Extend the mass relation down to low mass, dwarf galaxies

2. Incorporate galaxies that don't have GC populations

To do this, we need a model that can handle zeros.

We use the **hurdle model.**

This model was originally proposed by Eadie et al (2022)

The lognormal hurdle model

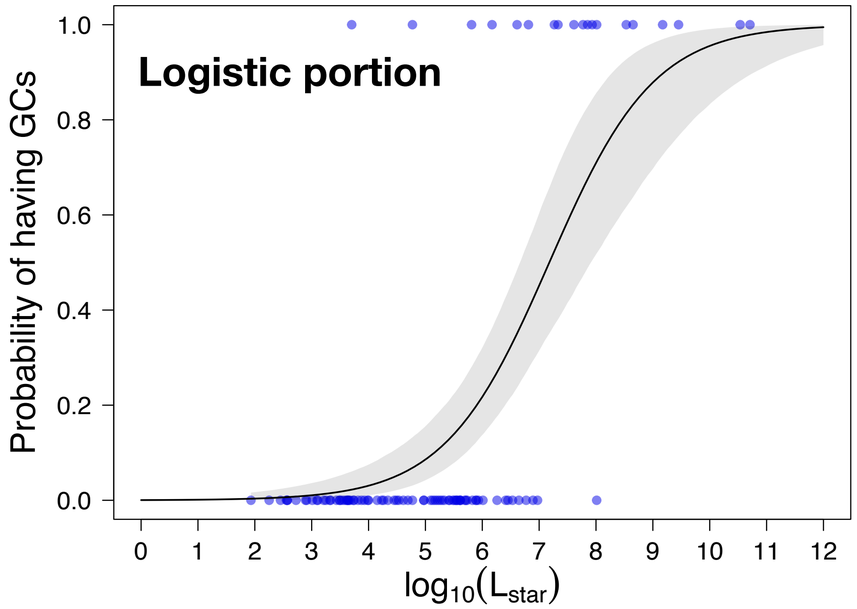
This model is a combination of a logistic and linear model:

$$
\begin{aligned}
I &\sim \text{Bern}(p(x)) \\
Y|(I=0) &= 0 \\
Y|(I=1) &\sim \mathcal{N}(\mu(x), \sigma)
\end{aligned}
$$

where:

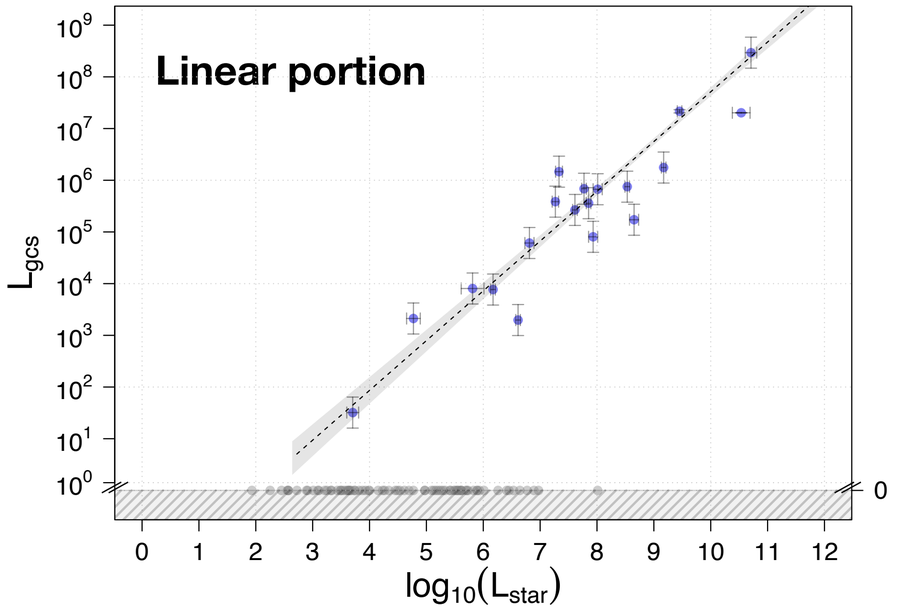
$$
\begin{aligned}
p(x) &= \frac{1}{1+e^{-(\beta_0+\beta_1 x)}} \\
\mu(x) &= \gamma_0 + \gamma_1 x
\end{aligned}
$$

This model is a combination of a logistic and linear model:



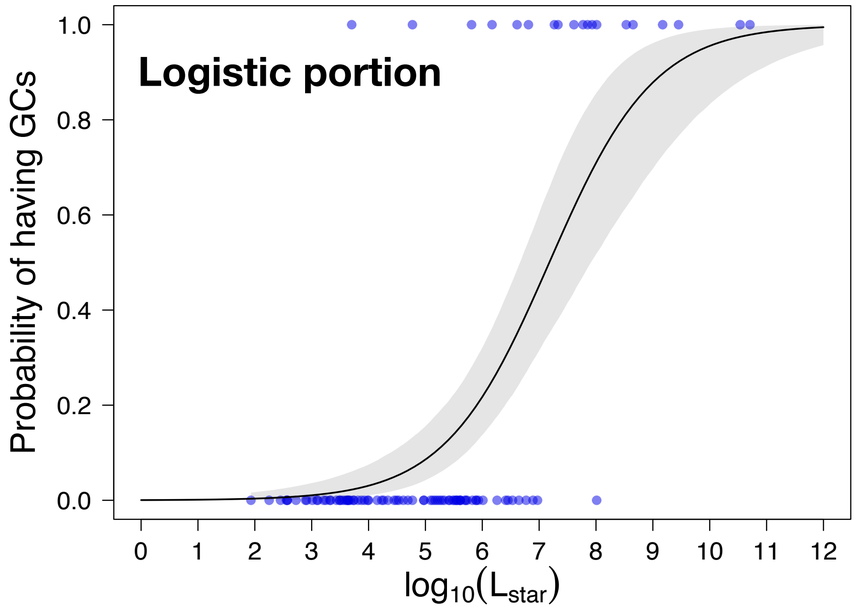
- Probability model
- Binary response variable: does or does not have GCs

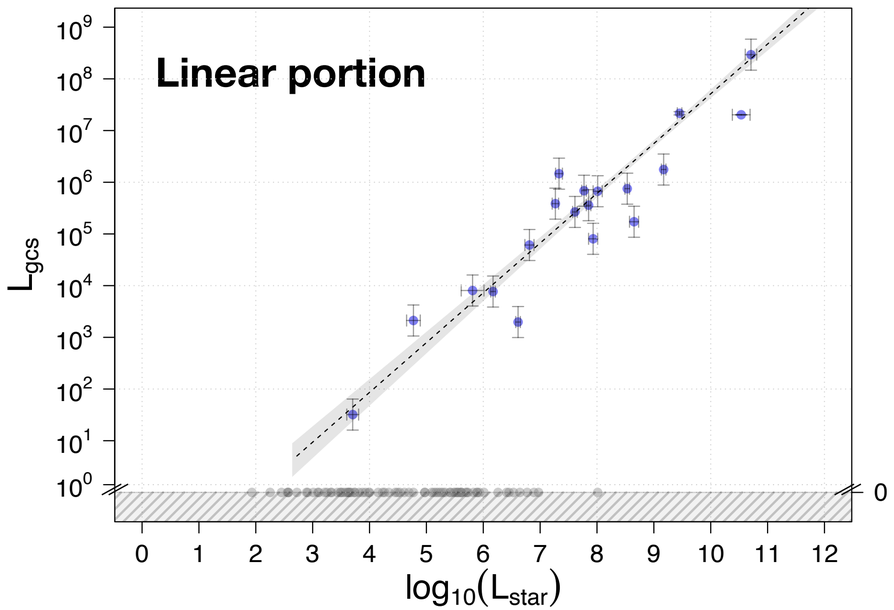
This model is a combination of a logistic and linear model:



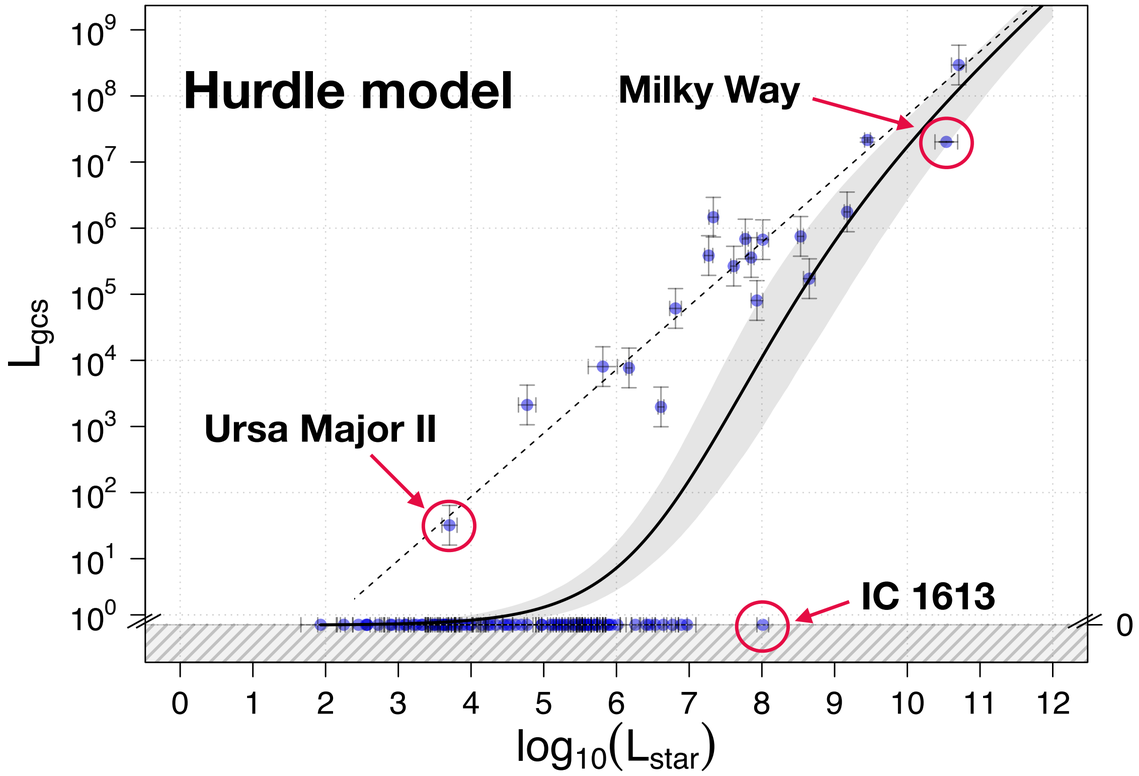
- Continuous response variable - GC luminosity/mass
- Zeros cannot be fit with non-zeros
- Log version cannot handle zeros

This model is a combination of a logistic and linear model:



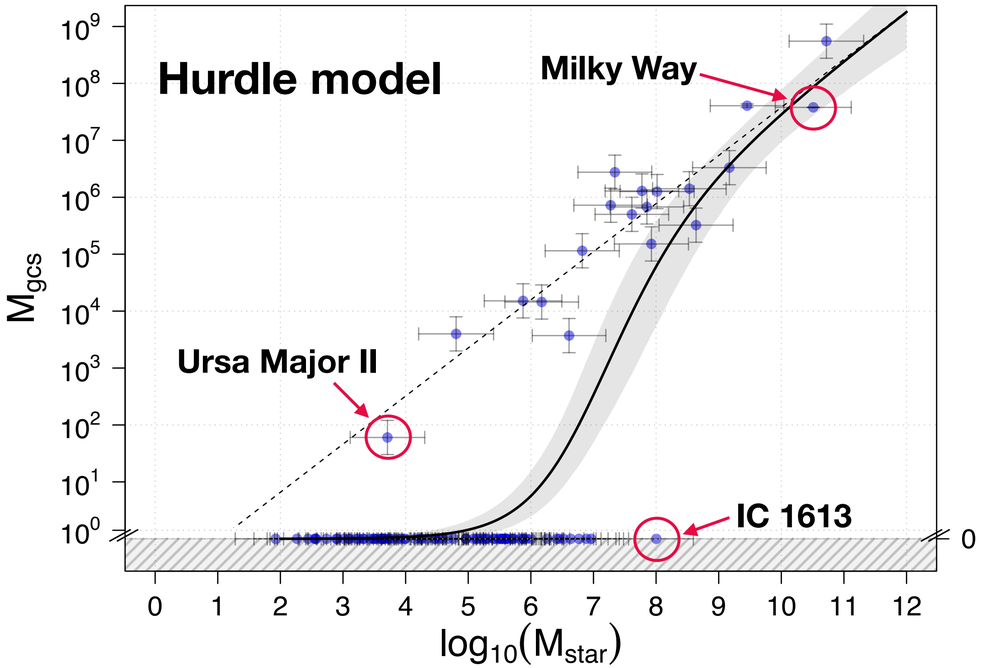


X



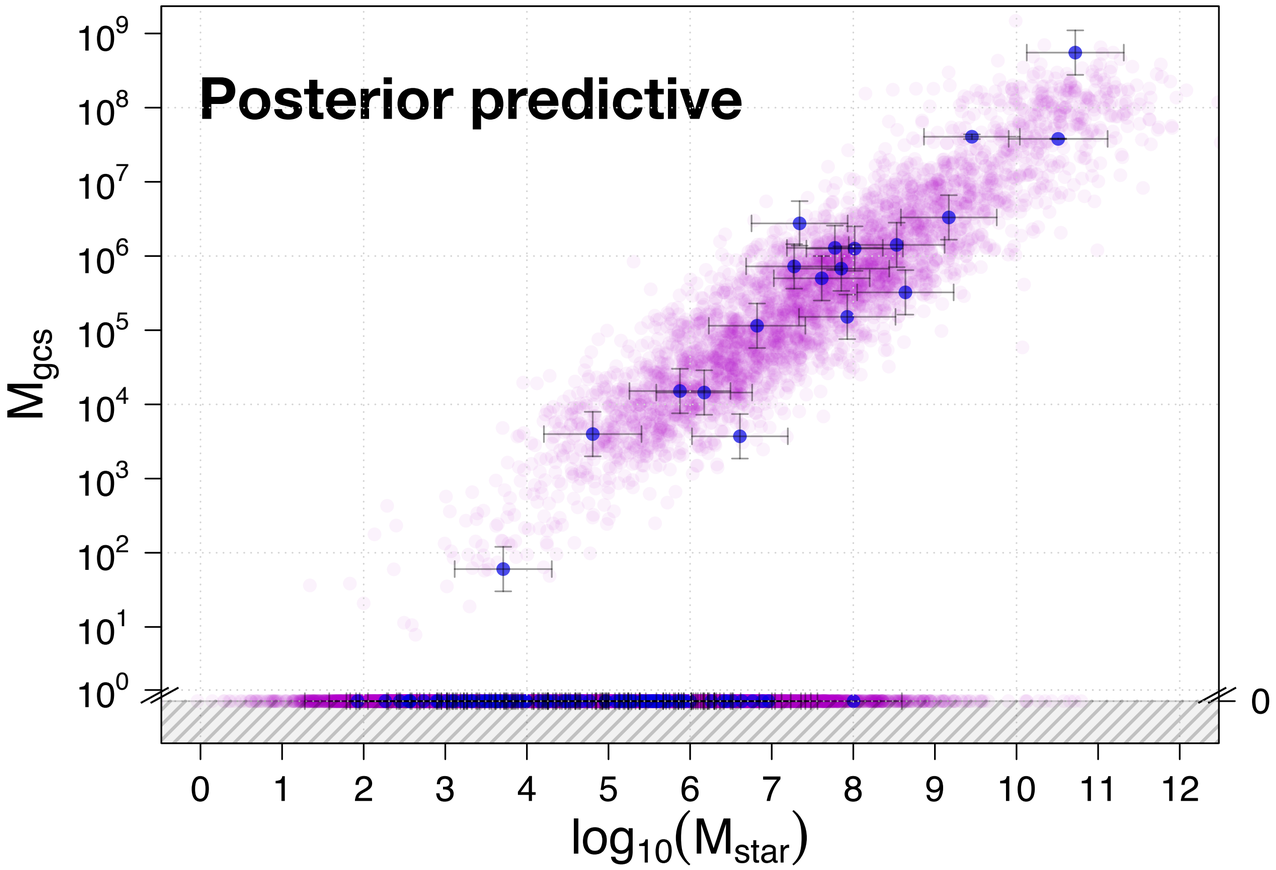
Berek et al 2023: https://arxiv.org/abs/2306.14945

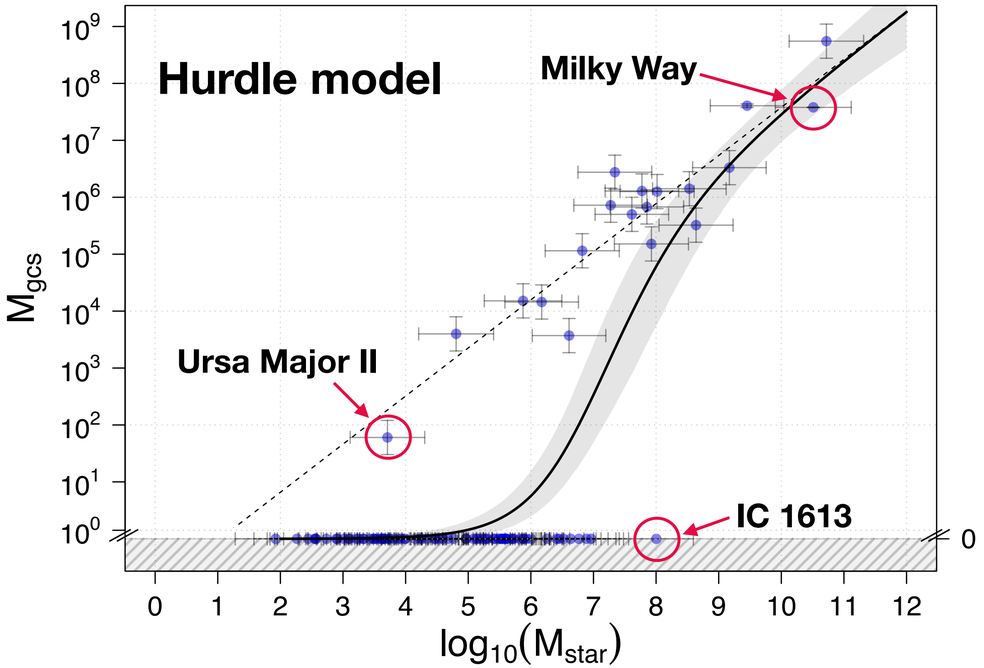
Results



Berek et al 2023: https://arxiv.org/abs/2306.14945

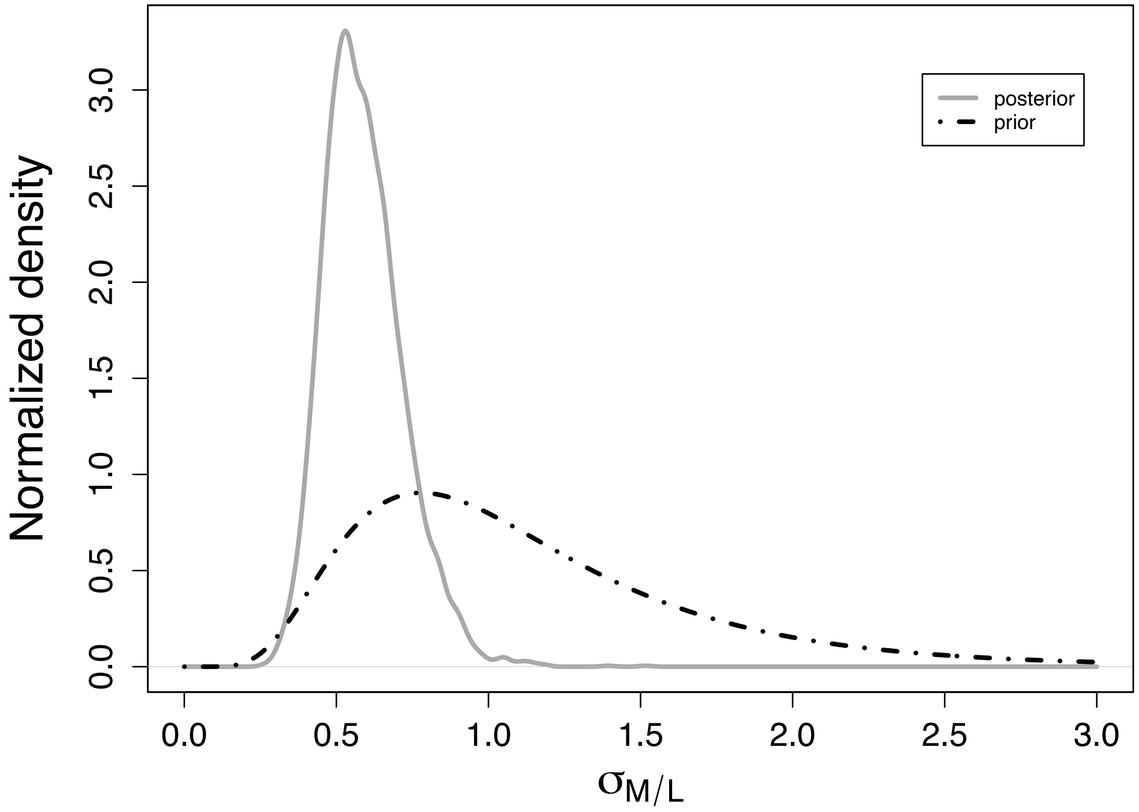
Results



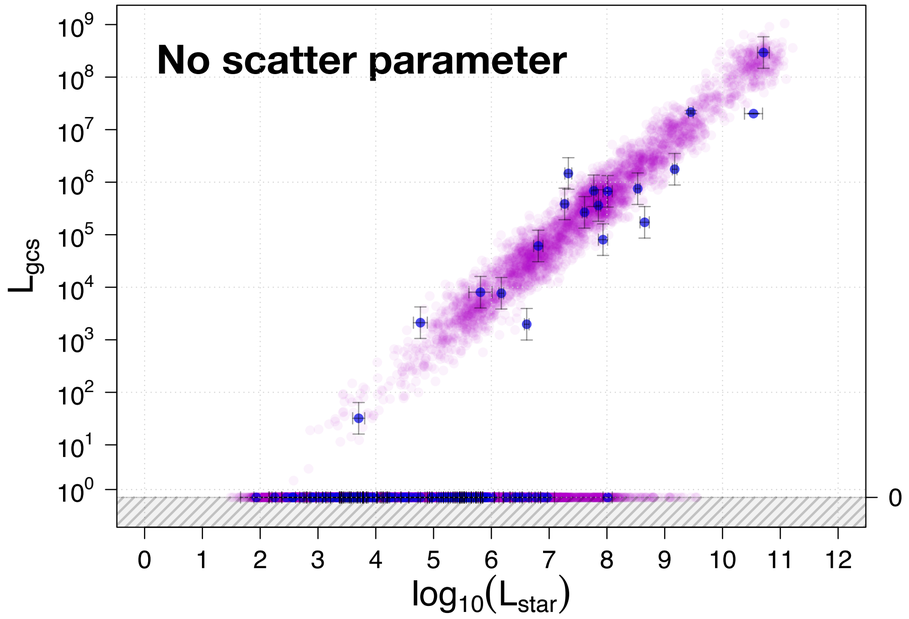


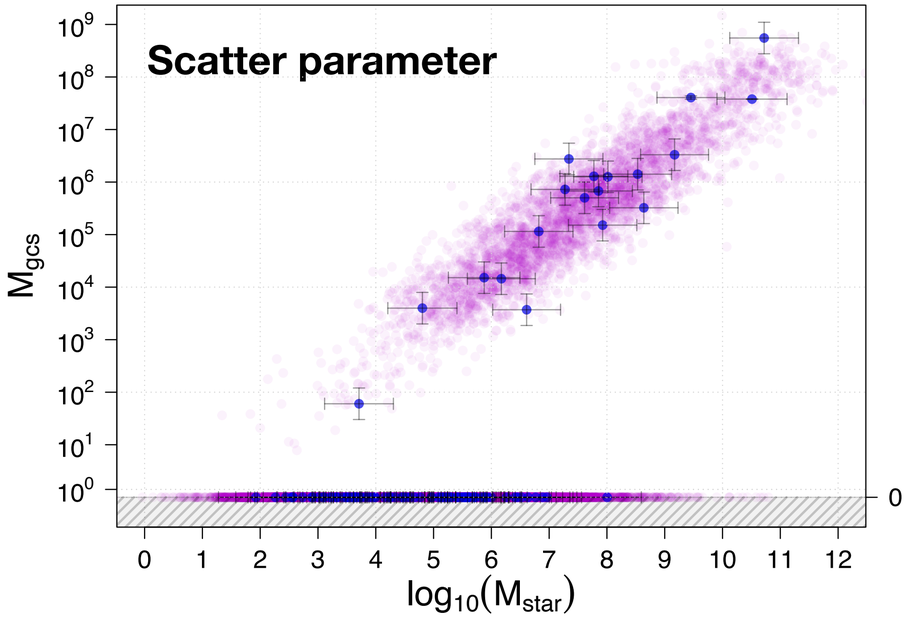
Berek et al 2023: https://arxiv.org/abs/2306.14945

There is non-zero **intrinsic scatter** in the $M_* - M_{GC}$ relation

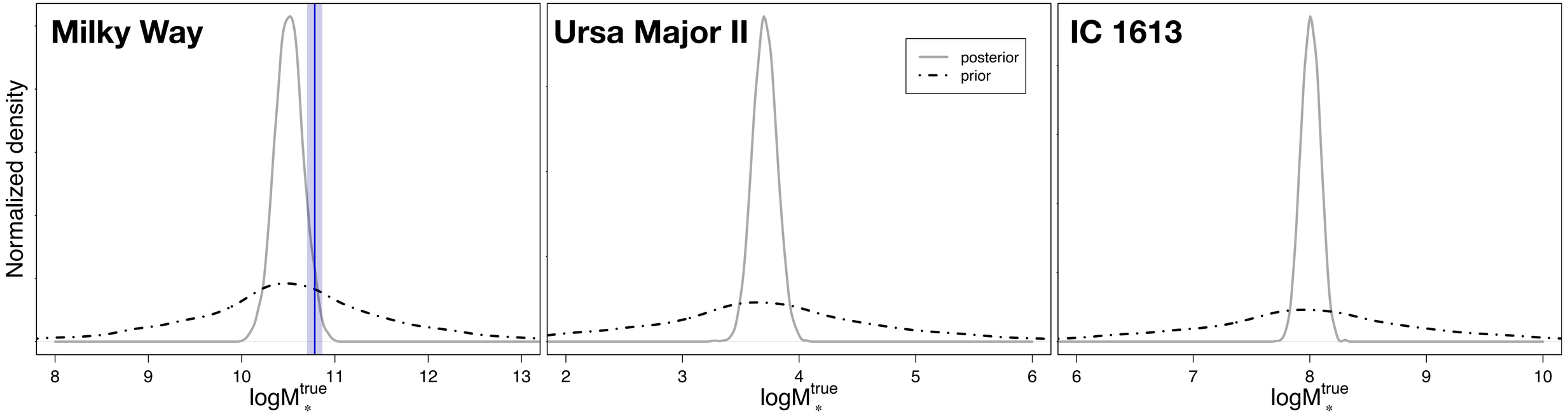


best fit value: $\sigma_{M/L} = 0.59$



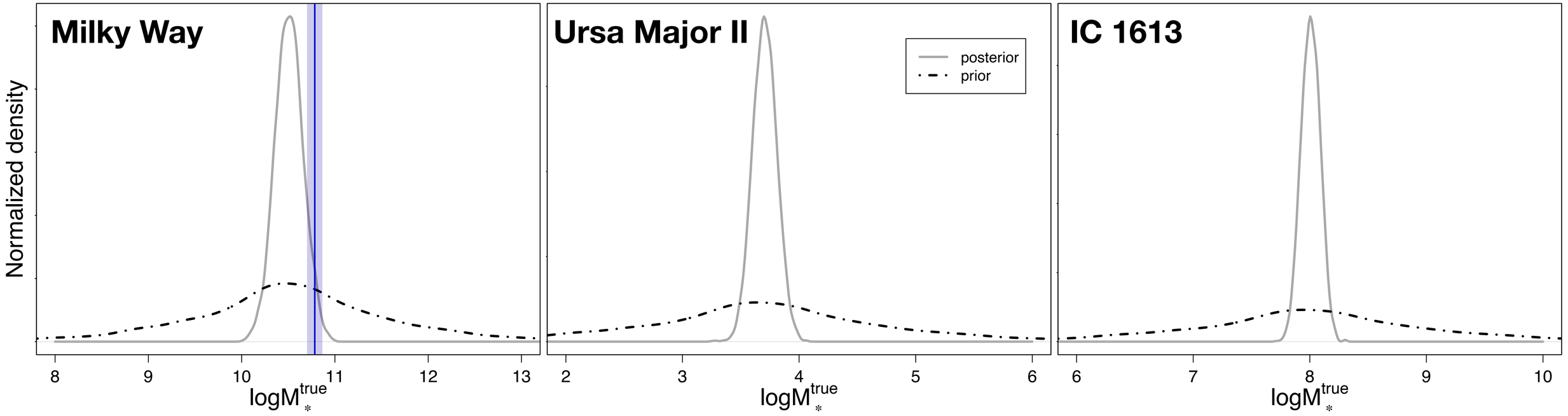


The hierarchical model gives us mass estimates for the galaxies, as well as estimates for their mass-to-light ratios



MW mass estimate from Licquia and Newman (2015)

The hierarchical model gives us mass estimates for the galaxies, as well as estimates for their mass-to-light ratios



Our estimate of the Milky Way's mass is not nearly as thorough as other studies, but is in agreement with the best estimates from literature

MW mass estimate from Licquia and Newman (2015)

The HERBAL model is expandable due to its hierarchical structure.

The HERBAL model is expandable due to its hierarchical structure.

In the future, we plan to…

- Incorporate a variety of other prior information (galaxy environments, types, etc)
  - *How to make this into a mixed model?*

Berek et al 2023: https://arxiv.org/abs/2306.14945

The HERBAL model is expandable due to its hierarchical structure.

In the future, we plan to…

- Incorporate a variety of other prior information (galaxy environments, types, etc)
- Test other error distributions
  - *Any specific recommendations?*

Berek et al 2023: https://arxiv.org/abs/2306.14945

The HERBAL model is expandable due to its hierarchical structure.

In the future, we plan to…

- Incorporate a variety of other prior information (galaxy environments, types, etc)
- Test other error distributions
- Add mixing between populations
  - *Methods to incorporate possibility of observed zeros being real non-zeros?*

Berek et al 2023: https://arxiv.org/abs/2306.14945

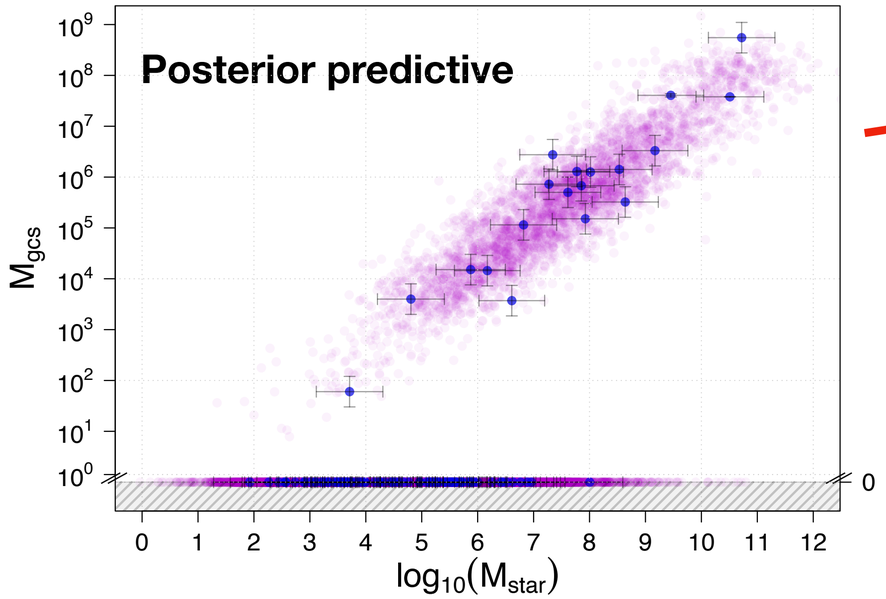
The HERBAL model is expandable due to its hierarchical structure.

In the future, we plan to…

- Incorporate a variety of other prior information (galaxy environments, types, etc)
- Test other error distributions
- Add mixing between populations
- Incorporate selection effects
  - *We often can't see faint GCs, causing us to underestimate our system masses. How can we account for this complex selection function?*

Berek et al 2023: https://arxiv.org/abs/2306.14945

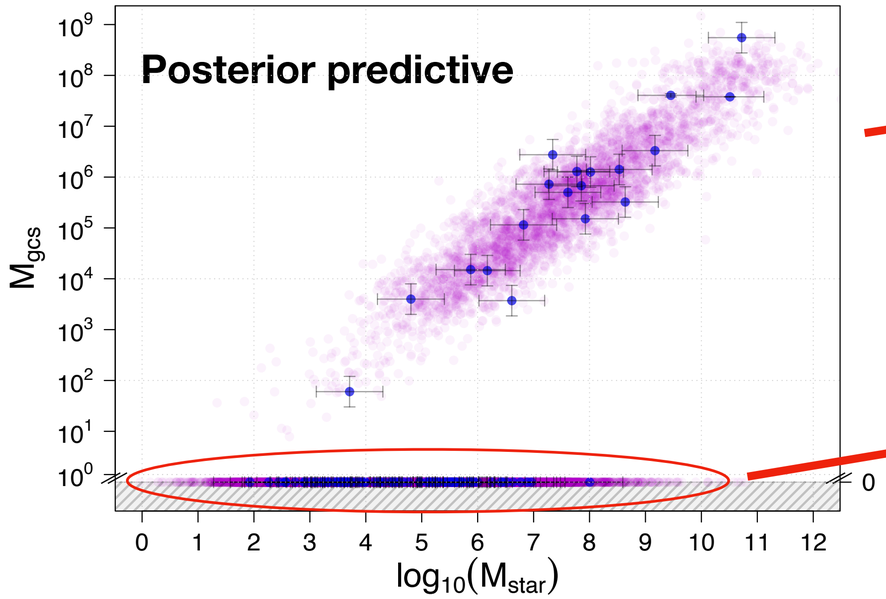
Conclusions:



The HERBAL model is a hierarchical model for the relationship between galaxy masses and their GC systems

Berek et al 2023: https://arxiv.org/abs/2306.14945

Conclusions:

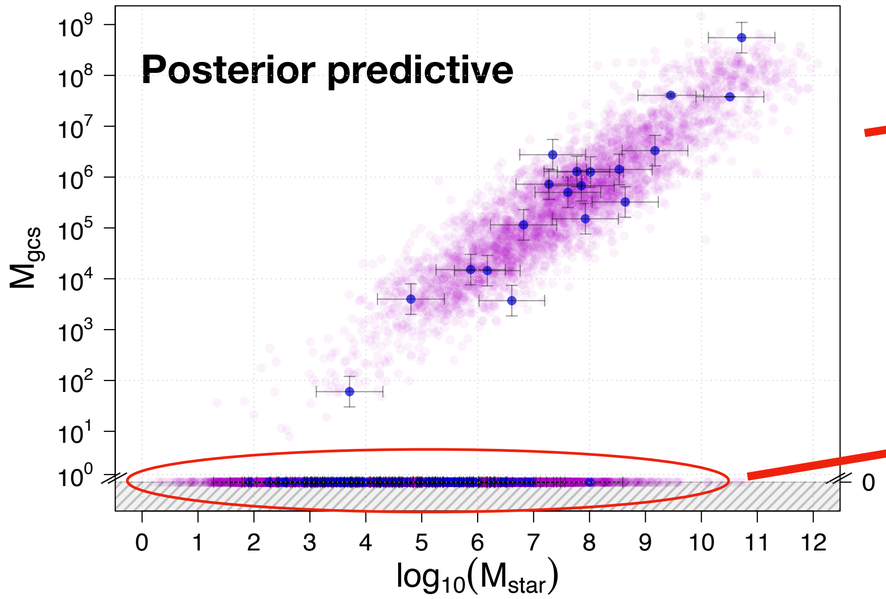


The HERBAL model is a hierarchical model for the relationship between galaxy masses and their GC systems

It incorporates zeros (galaxies without GC populations)

Berek et al 2023: https://arxiv.org/abs/2306.14945

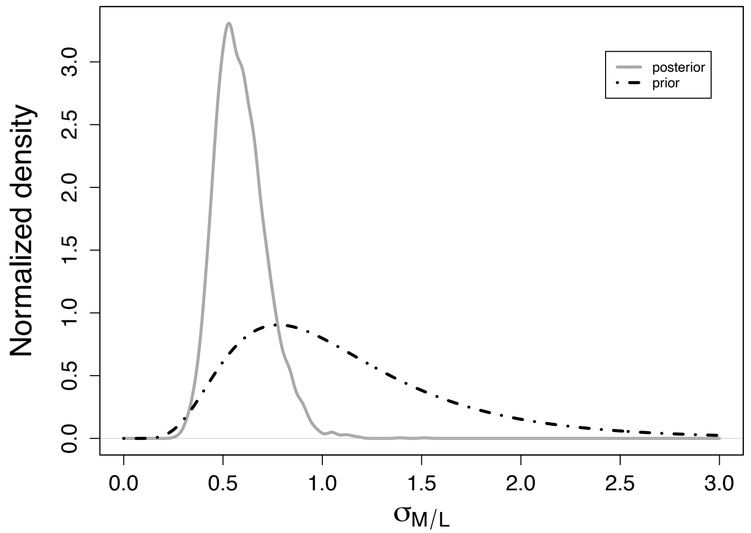
Conclusions:



The HERBAL model is a hierarchical model for the relationship between galaxy masses and their GC systems

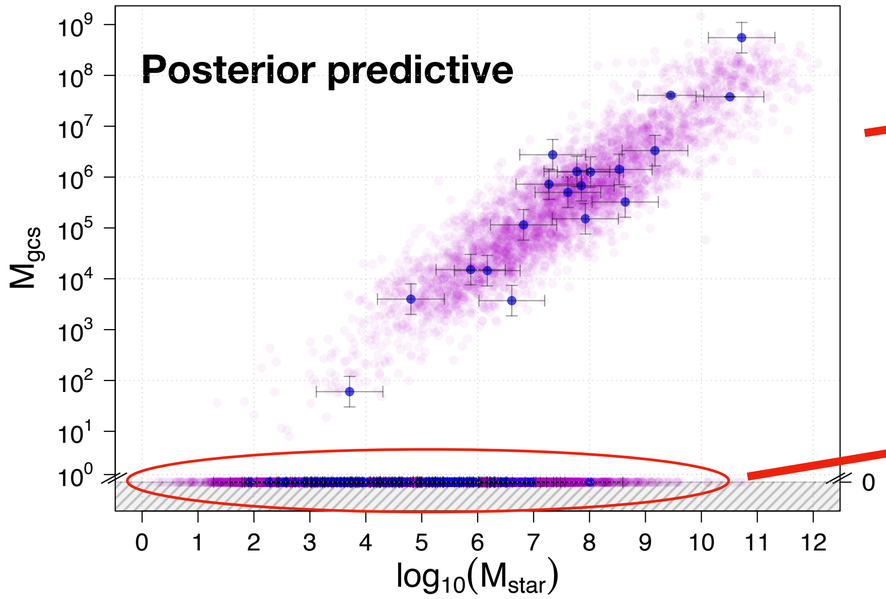
It incorporates zeros (galaxies without GC populations)

It accounts for uncertainties in all observational data and measures other scatter



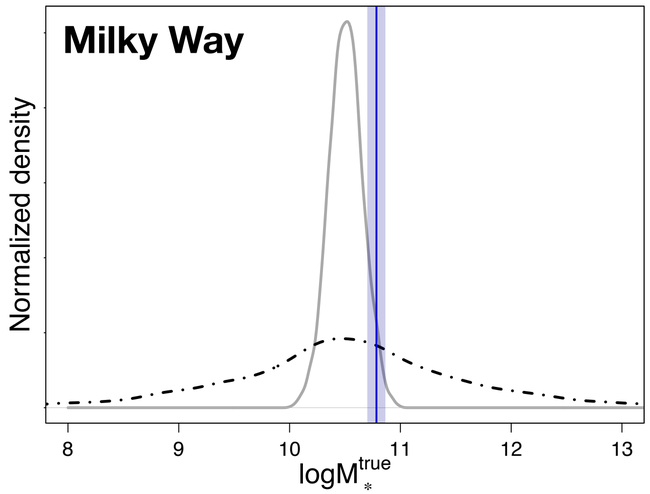
Berek et al 2023: https://arxiv.org/abs/2306.14945

Conclusions:



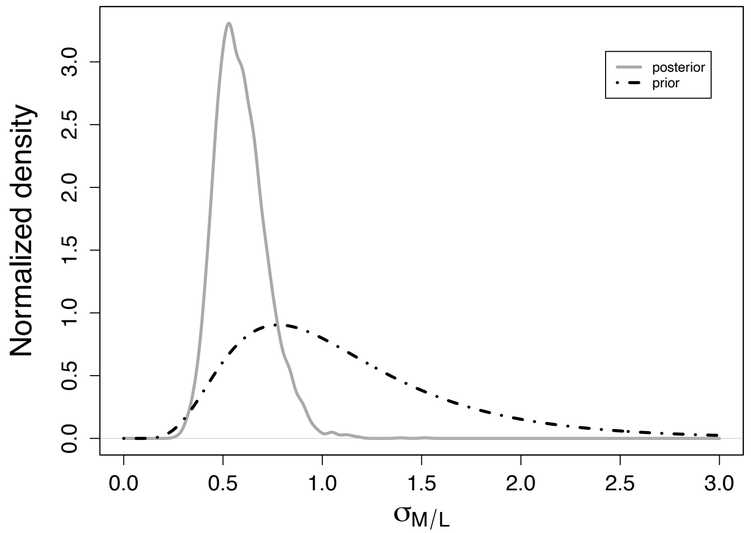
The HERBAL model is a hierarchical model for the relationship between galaxy masses and their GC systems

It incorporates zeros (galaxies without GC populations)



It accounts for uncertainties in all observational data and measures other scatter

It estimates galaxy masses and M/L ratios

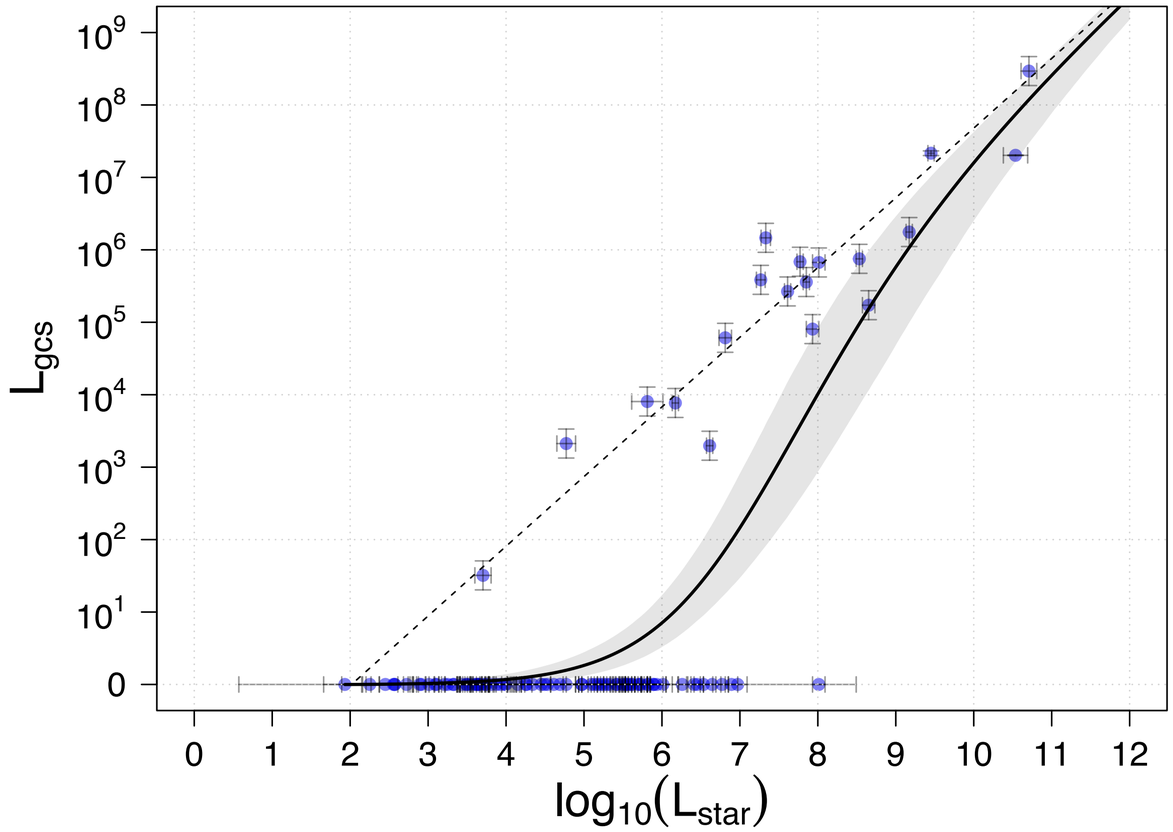


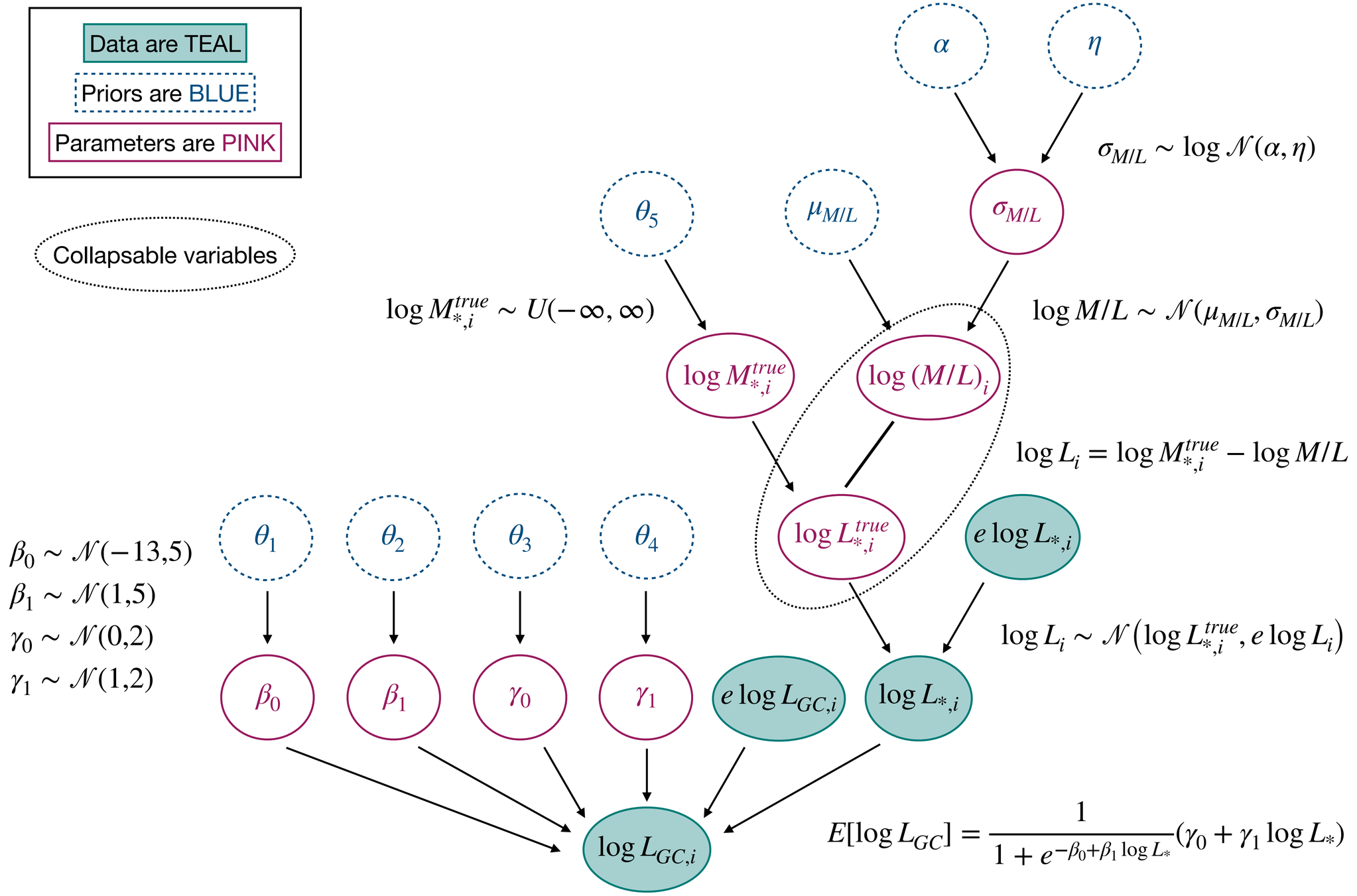
Berek et al 2023: https://arxiv.org/abs/2306.14945

The lognormal hurdle model

This leads to the expectation value:

$$E[\log L_{GC}] = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \log L_*)}}(\gamma_0 + \gamma_1 \log L_*)$$




Data are TEAL
Priors are BLUE
Parameters are PINK
Collapsable variables
α
η
$\sigma_{M/L} \sim \log \mathcal{N}(\alpha, \eta)$
θ_5
$\mu_{M/L}$
$\sigma_{M/L}$
$\log M_{*,i}^{true} \sim U(-\infty, \infty)$
$\log M/L \sim \mathcal{N}(\mu_{M/L}, \sigma_{M/L})$
$\log M_{*,i}^{true}$
$\log (M/L)_i$
$\log L_i = \log M_{*,i}^{true} - \log M/L$
θ_1
θ_2
θ_3
θ_4
$\log L_{*,i}^{true}$
$e \log L_{*,i}$
$\beta_0 \sim \mathcal{N}(-13,5)$
$\beta_1 \sim \mathcal{N}(1,5)$
$\gamma_0 \sim \mathcal{N}(0,2)$
$\gamma_1 \sim \mathcal{N}(1,2)$
$\log L_i \sim \mathcal{N}\left(\log L_{*,i}^{true}, e \log L_i\right)$
β_0
β_1
γ_0
γ_1
$e \log L_{GC,i}$
$\log L_{*,i}$
$\log L_{GC,i}$
$E[\log L_{GC}] = \frac{1}{1 + e^{-\beta_0 + \beta_1 \log L_*}} (\gamma_0 + \gamma_1 \log L_*)$