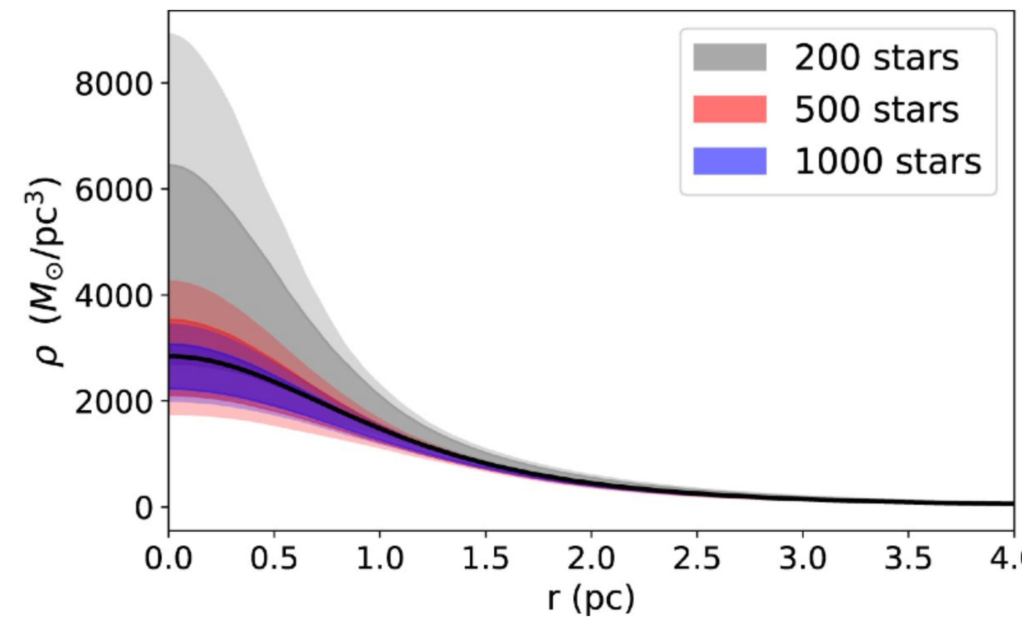


Hierarchical Bayesian Inference of Globular Cluster Properties

Robin Y. Wen, Joshua Speagle, Jeremy Webb & Gwendolyn Eadie
(2024, MNRAS, 527, 4193)

Raphael Oliveira, 13.mar.2024, Statistics Journal Club, OAUW

1 Introduction: evolution of GCs

- **Globular clusters:** $10^4 - 10^6$ stars, sharing roughly the same age (> 10 Gyr), chemical composition, distance and reddening
- Collisional system with two-body interactions, dynamically considered as a gravitating "gas" with stars as "molecules"
- Record of the dynamical and chemical conditions during the Galaxy formation and evolution → fossil relics, chemical clocks

NGC5024 / M53 (*HST*)

1 Introduction: evolution of GCs

- **Globular clusters:** $10^4 - 10^6$ stars, sharing roughly the same age (> 10 Gyr), chemical composition, distance and reddening

- Collisional system with two-body interactions, dynamically considered as a gravitating "gas" with stars as "molecules"

- Record of the dynamical and chemical conditions during the Galaxy formation and evolution → fossil relics, chemical clocks

- **Evolution of GCs stars:**
  - stellar evolution (HR, CMD diagrams)
  - dynamical processes (two-body relaxation, mass segregation, equipartition of energy, tidal disruption…) → spatial and kinematic distribution of stars

NGC5024 / M53 (*HST*)

1 Introduction: distribution functions (DFs)

* dynamical models

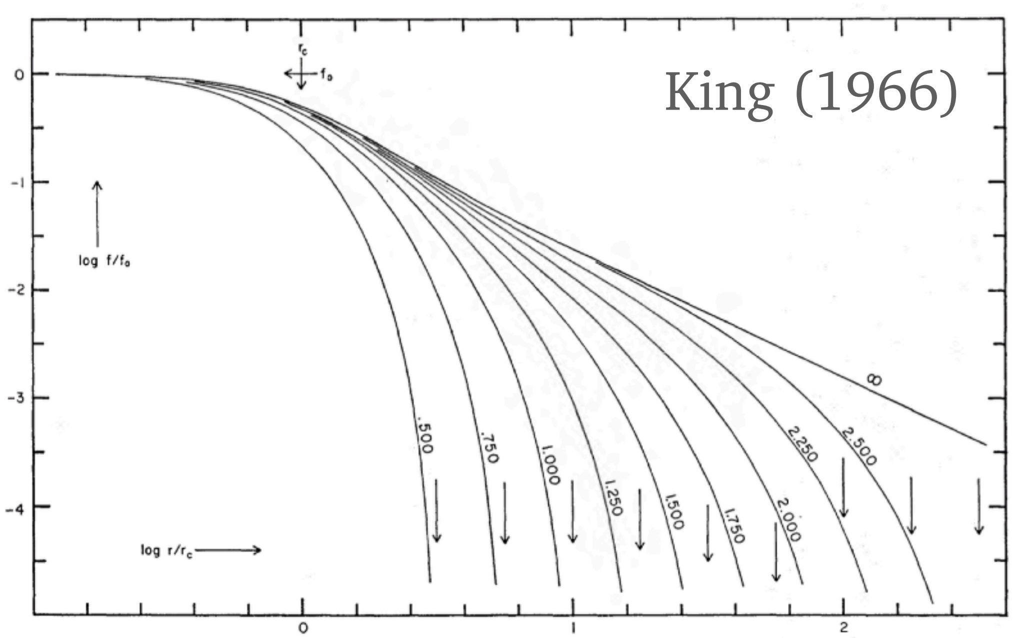
- **Timeline:**
  - Plummer (1905, $\sigma \propto r^{-4}$)
  - von Zeipel (1908, spherical mass of gas in isothermal equilibrium)
  - King (1962, $\sigma \propto r^{-2}$), King (1966)*
  - Michie (1963, radial anisotropy)
  - Wilson (1975)*
  - Elson, Fall & Freeman (1987)…

1 Introduction: distribution functions (DFs)

* dynamical models

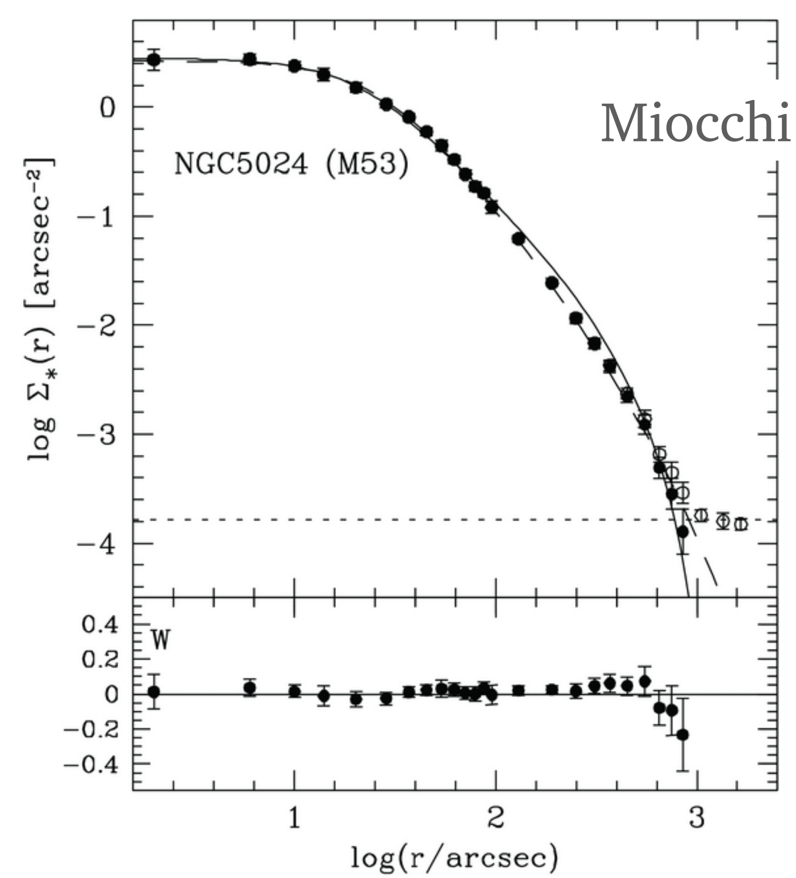
- **Timeline:**
  - Plummer (1905, $\sigma \propto r^{-4}$)
  - von Zeipel (1908, spherical mass of gas in isothermal equilibrium)
  - King (1962, $\sigma \propto r^{-2}$), King (1966)*
  - Michie (1963, radial anisotropy)
  - Wilson (1975)*
  - Elson, Fall & Freeman (1987)…

$$\sigma(r) = \sigma_\circ \left[\frac{1}{\sqrt{1+(r/r_c)^2}} - \frac{1}{\sqrt{1+(r_t/r_c)^2}} \right]^2 + \sigma_{bg}$$



Obs: for $\log_{10}(r_t/r_c) < 1.5 \sim$ King (1962)

Miocchi (2013)



1 Introduction: distribution functions (DFs)

* dynamical models

- **Timeline:**
  - Plummer (1905, $\sigma \propto r^{-4}$)
  - von Zeipel (1908, spherical mass of gas in isothermal equilibrium)
  - King (1962, $\sigma \propto r^{-2}$), King (1966)*
  - Michie (1963, radial anisotropy)
  - Wilson (1975)*
  - Elson, Fall & Freeman (1987)…

- Radial density, surface brightness, velocity dispersion profiles: in general obtained via radial binning, corrected for selection bias

- King models present a sharp cutoff of the density distribution close to r_t, whereas real data drops to the background level more smoothly.

  → **Limepy** (Gieles et al. 2015): lowered isothermal models

1 Introduction: Limepy code

- **Gieles et al. (2015):** family of lowered isothermal models, adopting either single or multi-mass. It is a generalization of all the other families of DFs: Woolley (1954, g=0), King (1966, g=1), Wilson (1975, g=2)

- $\boldsymbol{\theta_{lp} = (g, \Phi_0, M_{tot}, r_h)}$**,** where g and Φ_0 impact the shape and concentration of the GC profile, and M_{tot} and r_h are scaling parameters for mass and size

1 Introduction: Limepy code

- **Gieles et al. (2015):** family of lowered isothermal models, adopting either single or multi-mass. It is a generalization of all the other families of DFs: Woolley (1954, g=0), King (1966, g=1), Wilson (1975, g=2)

- **θ_{lp} = (g, Φ_0, M_{tot}, r_h),** where g and Φ_0 impact the shape and concentration of the GC profile, and M_{tot} and r_h are scaling parameters for mass and size

$$f\left(E(r,v)\right) = \begin{cases} A E_\gamma\left(g, -\frac{E-\phi(r_t)}{s^2}\right), & \text{for } E \leqslant \phi\left(r_t\right) \\ 0, & \text{for } E > \phi\left(r_t\right) \end{cases} \tag{1}$$

$$E(r,v) = v^2/2 + \phi(r). \tag{2}$$

$$E_\gamma(g,x) = \begin{cases} \exp(x), & g = 0 \\ \exp(x) P(g,x), & g > 0 \end{cases}, \tag{3}$$

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\phi}{dr}\right) = 4\pi G\rho, \text{ where } \rho = \int d^3v f(E(r,v)), \tag{4}$$

Notes:
- The total energy E is lowered by the potential at the truncation radius $\phi(r_t)$
- $P(g,x) = \gamma(g,x)/\Gamma(g)$ is the regularized lower incomplete gamma function
- Eq. 4 is the non-linear Poisson's equation

1 Introduction: Limepy code

- **Gieles et al. (2015):** family of lowered isothermal models, adopting either single or multi-mass. It is a generalization of all the other families of DFs: Woolley (1954, g=0), King (1966, g=1), Wilson (1975, g=2)
- **θ_{lp} = (g, Φ_0, M_{tot}, r_h),** where g and Φ_0 impact the shape and concentration of the GC profile, and M_{tot} and r_h are scaling parameters for mass and size

$$f\left(E(r,v)\right) = \begin{cases} A E_\gamma\left(g, -\frac{E-\phi(r_t)}{s^2}\right), & \text{for } E \leqslant \phi\left(r_t\right) \\ 0\,, & \text{for } E > \phi\left(r_t\right) \end{cases} \quad (1)$$

$$E(r,v) = v^2/2 + \phi(r). \quad (2)$$

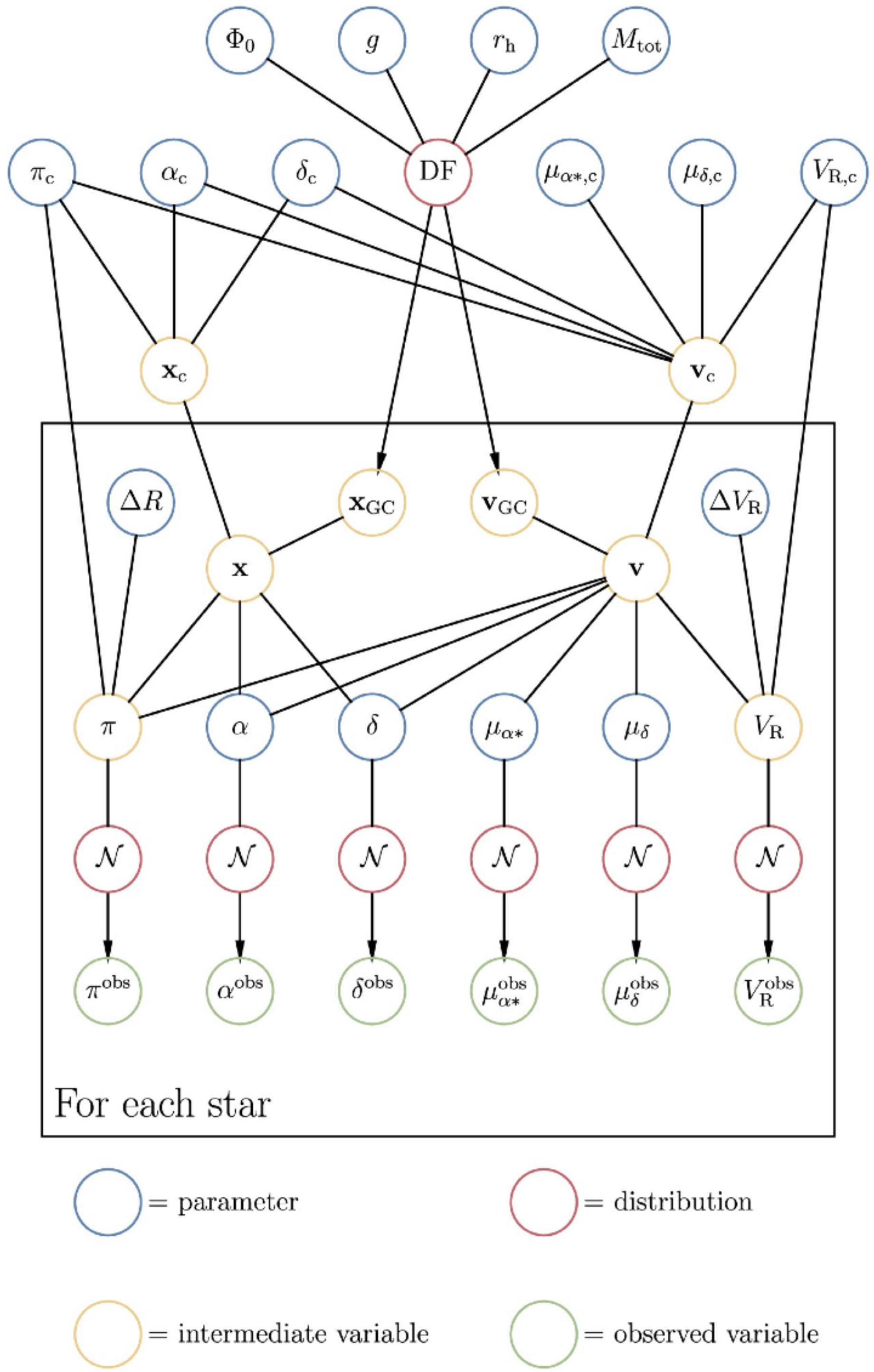
$$E_\gamma(g,x) = \begin{cases} \exp(x), & g = 0 \\ \exp(x) P(g,x), & g > 0 \end{cases}, \quad (3)$$

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2 \frac{d\phi}{dr}\right) = 4\pi G \rho, \text{ where } \rho = \int d^3 v f(E(r,v)), \quad (4)$$

Notes:
- The total energy E is lowered by the potential at the truncation radius $\phi(r_t)$
- $P(g,x) = \gamma(g,x)/\Gamma(g)$ is the regularized lower incomplete gamma function
- Eq. 4 is the non-linear Poisson's equation

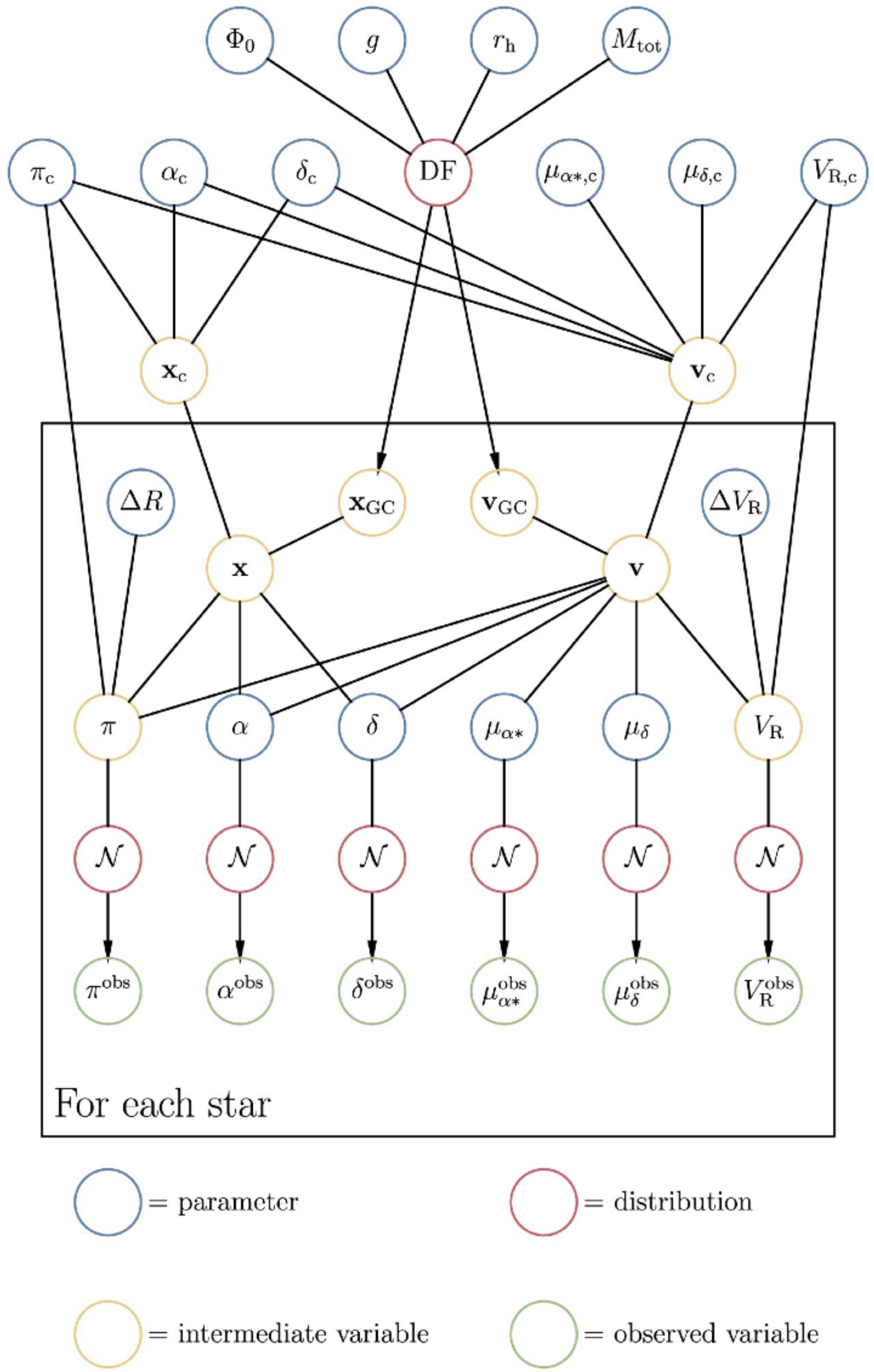
Limepy solves all these equations and gives DF $f(r, v)$ for any θ_{lp}

2 Methods: hierarchical bayesian model



- Eadie et al. (2022): MCMC with limepy neglecting the measurement errors in positions and velocities of stars
- Hierarchical bayesian inference

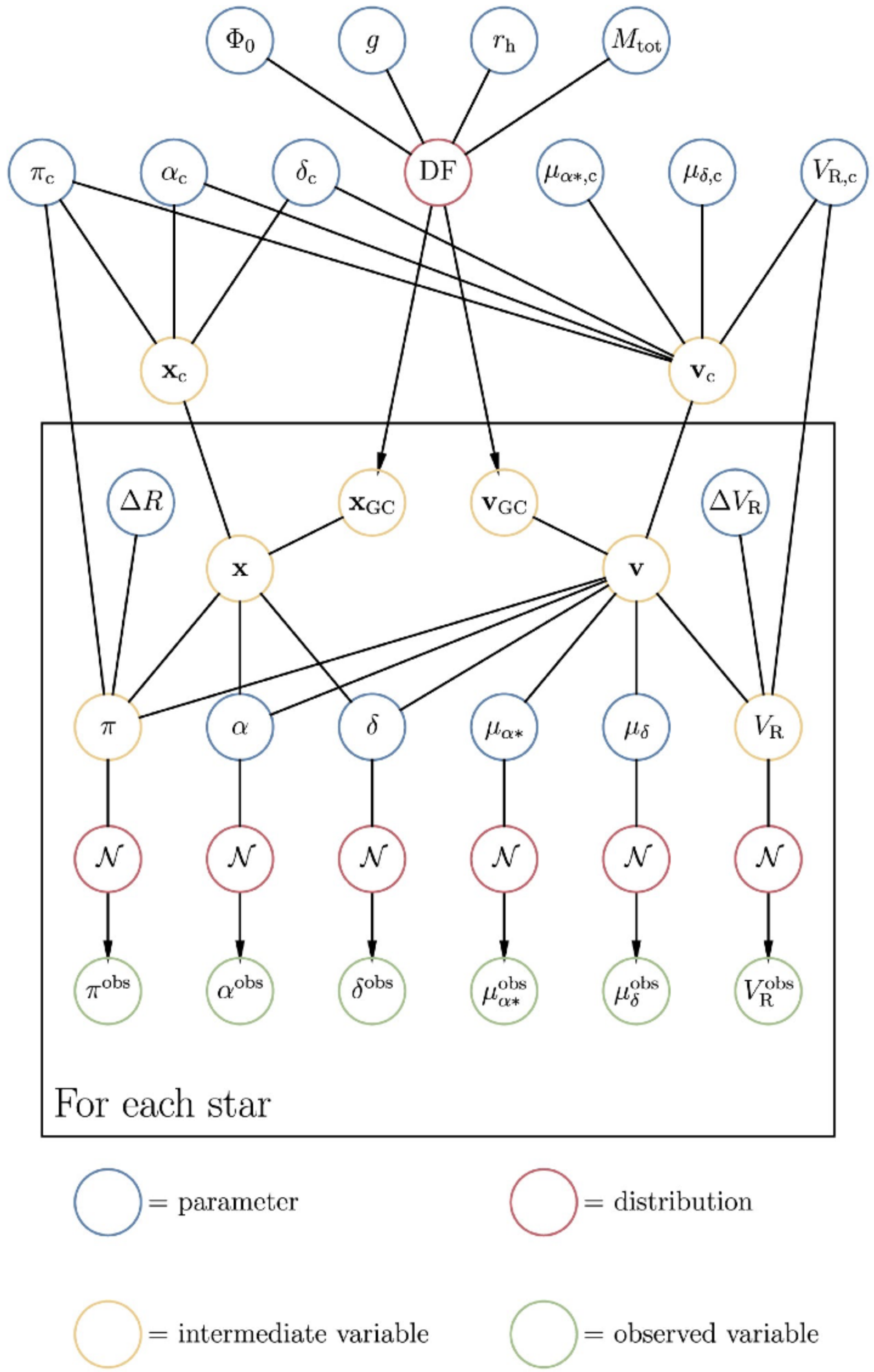
2 Methods: hierarchical bayesian model



- Eadie et al. (2022): MCMC with limepy neglecting the measurement errors in positions and velocities of stars
- Hierarchical bayesian inference
  - d_i^{obs}: phase space comp. of each star
  - these follow a **normal** distribution, with measurement errors (*Gaia*)
  - **equatorial** → **cartesian** coordinates
  - Likelihood combines true positions and velocities, GC center and structural parameters ($6N + 10$ params.):

$$p(\vec{\mathbf{x}}_{GC}, \vec{\mathbf{v}}_{GC}|\boldsymbol{\theta}_{lp}) = \prod_{i=1}^{N} \frac{f_{lp}\left(\mathbf{x}_{GC,i}, \mathbf{v}_{GC,i}|\boldsymbol{\theta}_{lp}\right)}{M_{total}}$$

2 Methods: hierarchical bayesian model



Posterior of the whole model:

$$
\begin{aligned}
& p\left(\boldsymbol{\theta}_{\mathrm{lp}}, \boldsymbol{\theta}_{\mathrm{c}}, \vec{\boldsymbol{s}}, \vec{\boldsymbol{t}} \mid \vec{\boldsymbol{q}}^{\mathrm{obs}}, \vec{\boldsymbol{p}}^{\mathrm{obs}}\right) \\
& \propto p\left(\vec{\boldsymbol{q}}^{\mathrm{obs}}, \vec{\boldsymbol{p}}^{\mathrm{obs}} \mid \boldsymbol{\theta}_{\mathrm{lp}}, \boldsymbol{\theta}_{\mathrm{c}}, \vec{\boldsymbol{s}}, \vec{\boldsymbol{t}}\right) p\left(\boldsymbol{\theta}_{\mathrm{lp}}, \boldsymbol{\theta}_{\mathrm{c}}, \vec{\boldsymbol{s}}, \vec{\boldsymbol{t}}\right) \\
& = p\left(\vec{\boldsymbol{q}}^{\mathrm{obs}}, \vec{\boldsymbol{p}}^{\mathrm{obs}} \mid \vec{\boldsymbol{q}}(\vec{\boldsymbol{s}}), \vec{\boldsymbol{p}}(\vec{\boldsymbol{t}})\right) p\left(\vec{\boldsymbol{s}}, \vec{\boldsymbol{t}} \mid \boldsymbol{\theta}_{\mathrm{lp}}, \boldsymbol{\theta}_{\mathrm{c}}\right) p\left(\boldsymbol{\theta}_{\mathrm{lp}}, \boldsymbol{\theta}_{\mathrm{c}}\right) \qquad (14)
\end{aligned}
$$

$$
\begin{aligned}
= \prod_{i=1}^{N} \Bigg\{ & p\left(\vec{\boldsymbol{q}}_i^{\mathrm{obs}}, \vec{\boldsymbol{p}}_i^{\mathrm{obs}} \mid \vec{\boldsymbol{q}}_i(\vec{\boldsymbol{s}}_i), \vec{\boldsymbol{p}}_i(\vec{\boldsymbol{t}}_i)\right) \\
& \frac{f_{\mathrm{lp}}\left(T\left(\boldsymbol{q}_i(\boldsymbol{s}_i), \boldsymbol{p}_i(\boldsymbol{t}_i)\right) - T\left(\boldsymbol{q}_{\mathrm{c}}, \boldsymbol{p}_{\mathrm{c}}\right) \mid \boldsymbol{\theta}_{\mathrm{lp}}\right)}{M_{\mathrm{total}}} \left| \frac{\partial T\left(\boldsymbol{q}_i(\boldsymbol{s}_i), \boldsymbol{p}_i(\boldsymbol{t}_i)\right)}{\partial(\boldsymbol{s}_i, \boldsymbol{t}_i)} \right| \Bigg\} \\
& p\left(\boldsymbol{\theta}_{\mathrm{lp}}, \boldsymbol{\theta}_{\mathrm{c}}\right), \qquad (15)
\end{aligned}
$$

2 Methods: hierarchical bayesian model

Table 1. `Limepy` model parameter values used to simulate stars in GCs. We consider all possible combinations of these four GC parameters. The parameter combination ($g = 2.0$, $\Phi_0 = 8.0$) is excluded due to its closeness to the cutoff boundary (shown in Fig. 2) that distinguishes realistic GC models from unrealistic ones (see Section 3.2 for a more detailed explanation). This leaves us with 32 sets of different parameters. We generate 10 simulations for each of the parameter sets and test our model on these simulations.

$\boldsymbol{\theta}_{\mathrm{lp}}$	Description	Possible values
g	Truncation parameter	1.2, 1.6, 2.0
Φ_0	Central gravitational potential	3.0, 5.0, 8.0
M_{tot}	Total mass ($M_\odot$)	10^5, 10^6
r_{h}	Half-light radius (pc)	3.0, 9.0

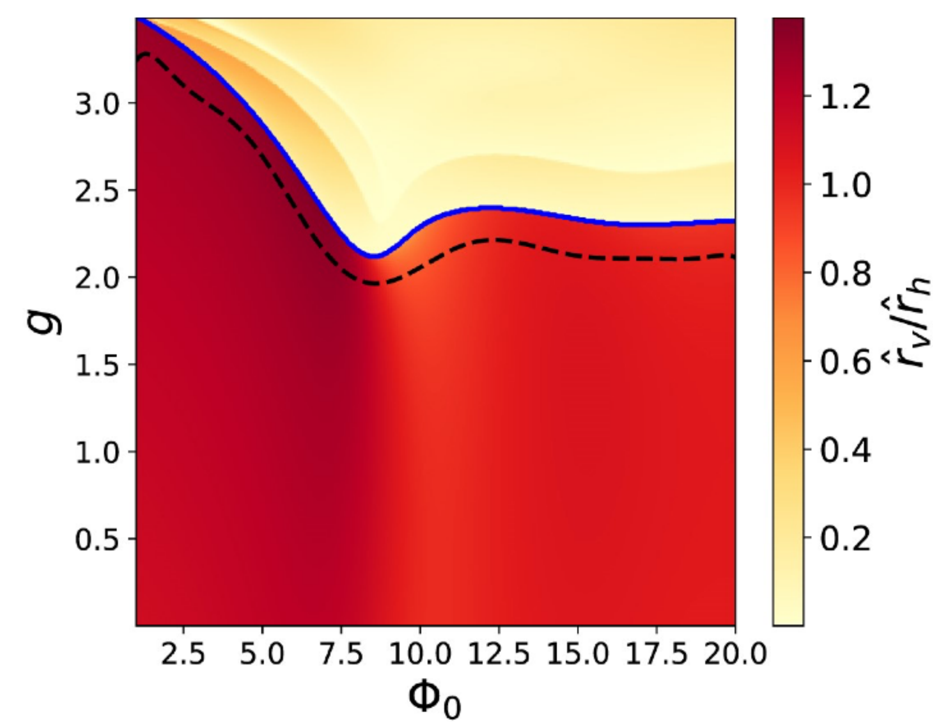
Table 2. Hyperparameter values used to simulate stars in GCs that reflect realistic observation conditions. We fix the structural GC parameters at (g, Φ_0, M_{tot}, r_{h}) = (2.0, 5.0, 10^5, 3) while changing the above hyperparameters of the simulations individually. In the third row, σ denotes the measurement errors for angular positions (mas), parallax (mas), and proper motions (mas yr^{-1}) for an individual star, where we assume the errors for all five components are the same. We also assume that every star in GC shares the same measurement uncertainties.

$\boldsymbol{\theta}$	Description	Possible values
N	Number of stars measured in cluster	100, 500, 1000
R_{c}	Distance of cluster centre to Earth (kpc)	1, 2, 5, 10
σ	Measurement uncertainties (mas, mas yr^{-1})	0.02, 0.1, 0.5

4 parameters, 3 hyperparameters

$3 \times 3 \times 2 \times 2 = 36 - 4 = 32$ sets excluding (g=2.0, Φ_0=8.0), close to the cutoff boundary $\hat{r}_v/\hat{r}_h \gtrsim 0.64$ to be relevant to modelling GCs

- Uniform priors, except for M_{tot}, r_{h}

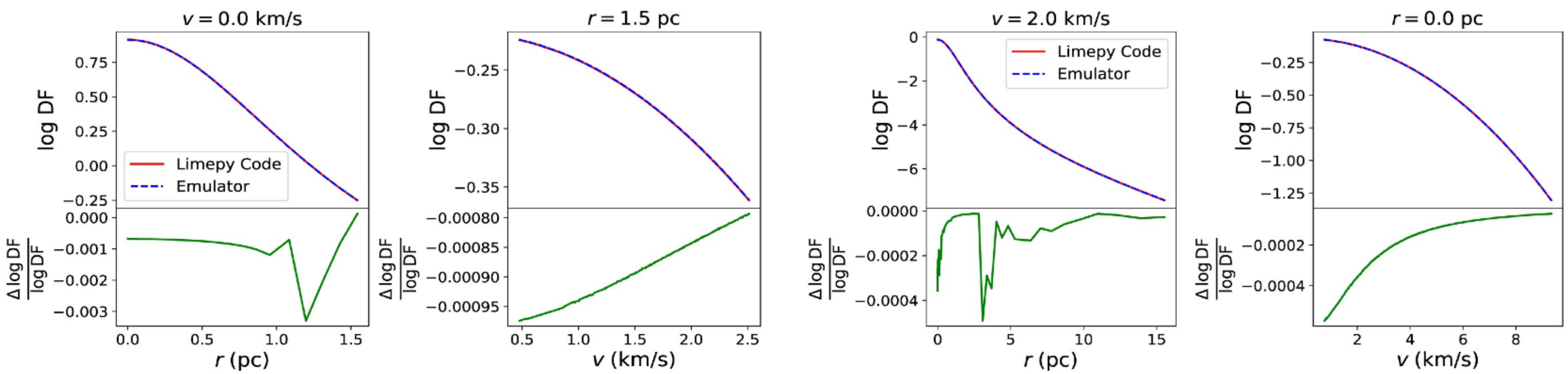


2 Methods: interpolation-based emulator

- Hamiltonian Monte Carlo: likelihood gradient, NUTS, automatic differentiation (ODE solver), JAX, pyMC → grid linear interpolations at fixed M_{tot}, $\mathrm{r_h}$

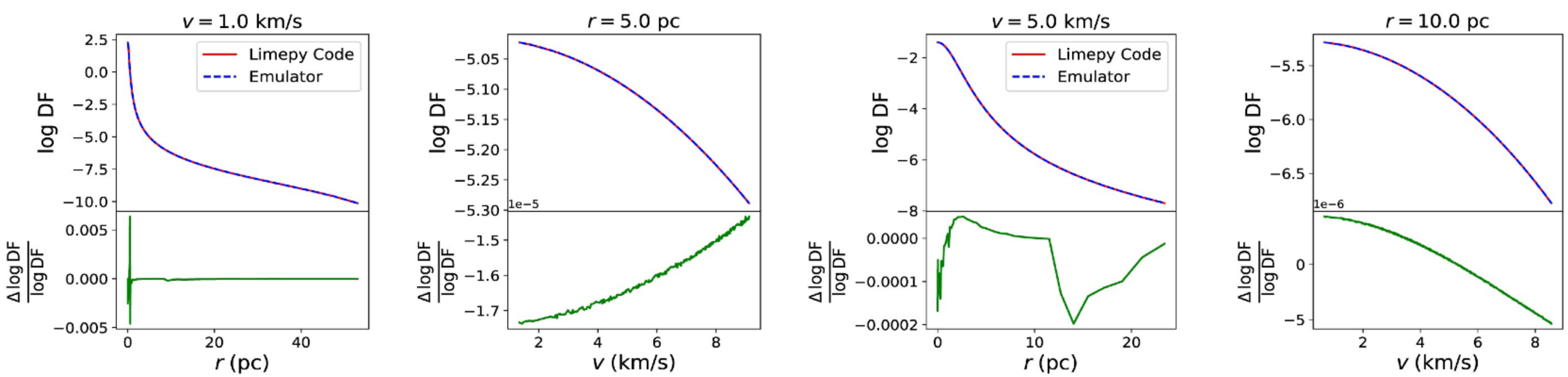
2 Methods: interpolation-based emulator

- Hamiltonian Monte Carlo: likelihood gradient, NUTS, automatic differentiation (ODE solver), JAX, pyMC → grid linear interpolations at fixed M_{tot}, r_h



(a) $(\Phi_0, g, \log_{10} M_{\mathrm{tot}}, r_h) = (1.5, 0.002, 4, 1)$

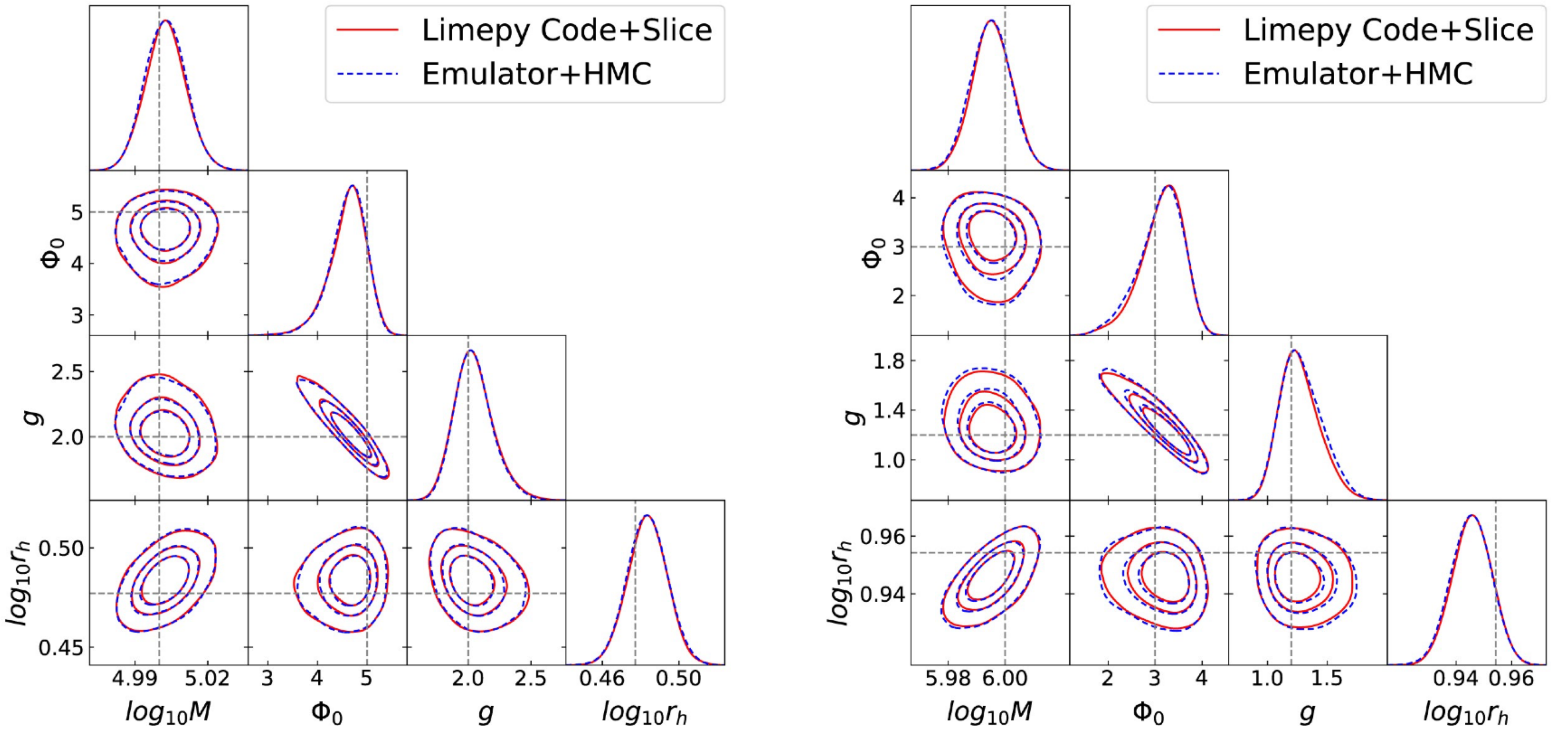
(b) $(\Phi_0, g, \log_{10} M_{\mathrm{tot}}, r_h) = (5, 2, 5, 3)$



(c) $(\Phi_0, g, \log_{10} M_{\mathrm{tot}}, r_h) = (10, 1, 6, 9)$

(d) $(\Phi_0, g, \log_{10} M_{\mathrm{tot}}, r_h) = (7.5, 0.5, 5.5, 12)$

2 Methods: interpolation-based emulator

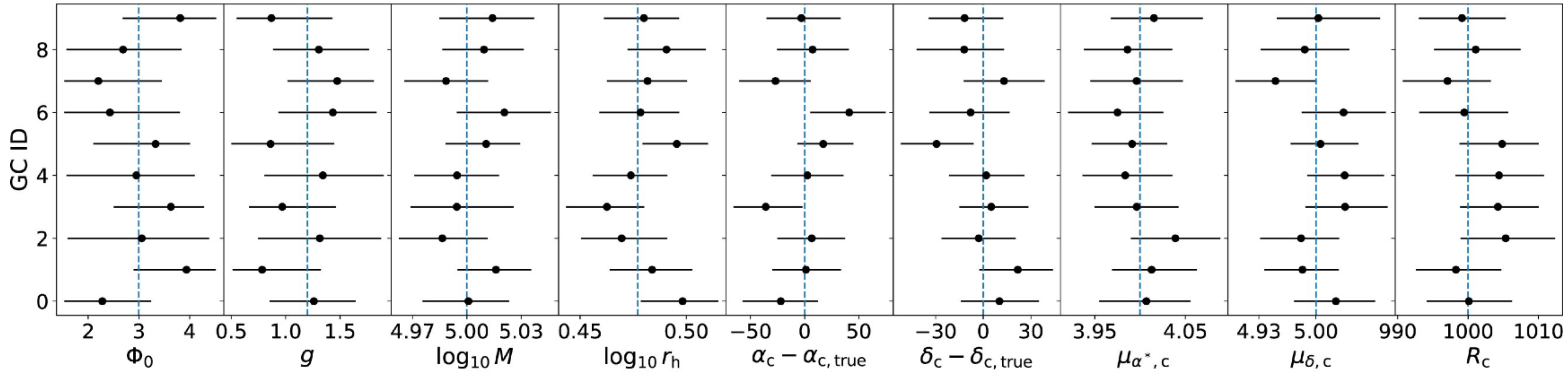


(a) True parameters $(\Phi_0, g, \log_{10} M_{\rm tot}, r_{\rm h}) = (5, 2, 5, 3)$

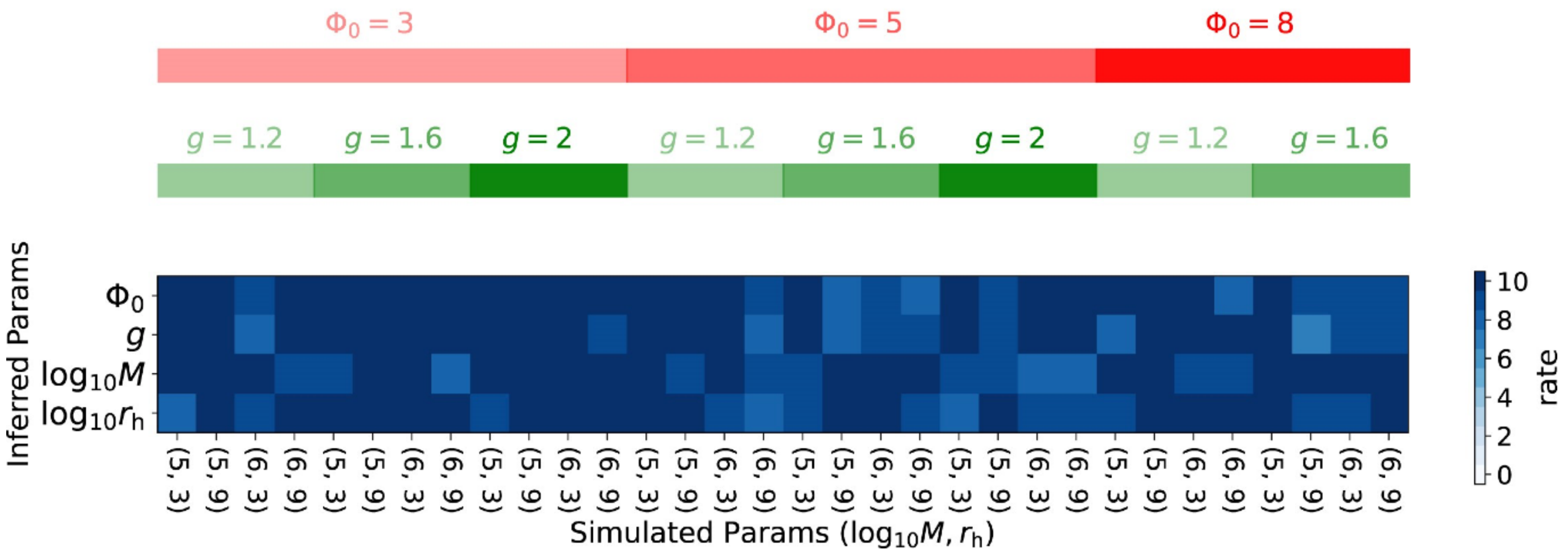
(b) True parameters $(\Phi_0, g, \log_{10} M_{\rm tot}, r_{\rm h}) = (3, 1.2, 6, 9)$

- No measurement errors assumed
- Emulator speeds-up 50x the likelihood evaluation: 1-2h instead of days (CPU?)

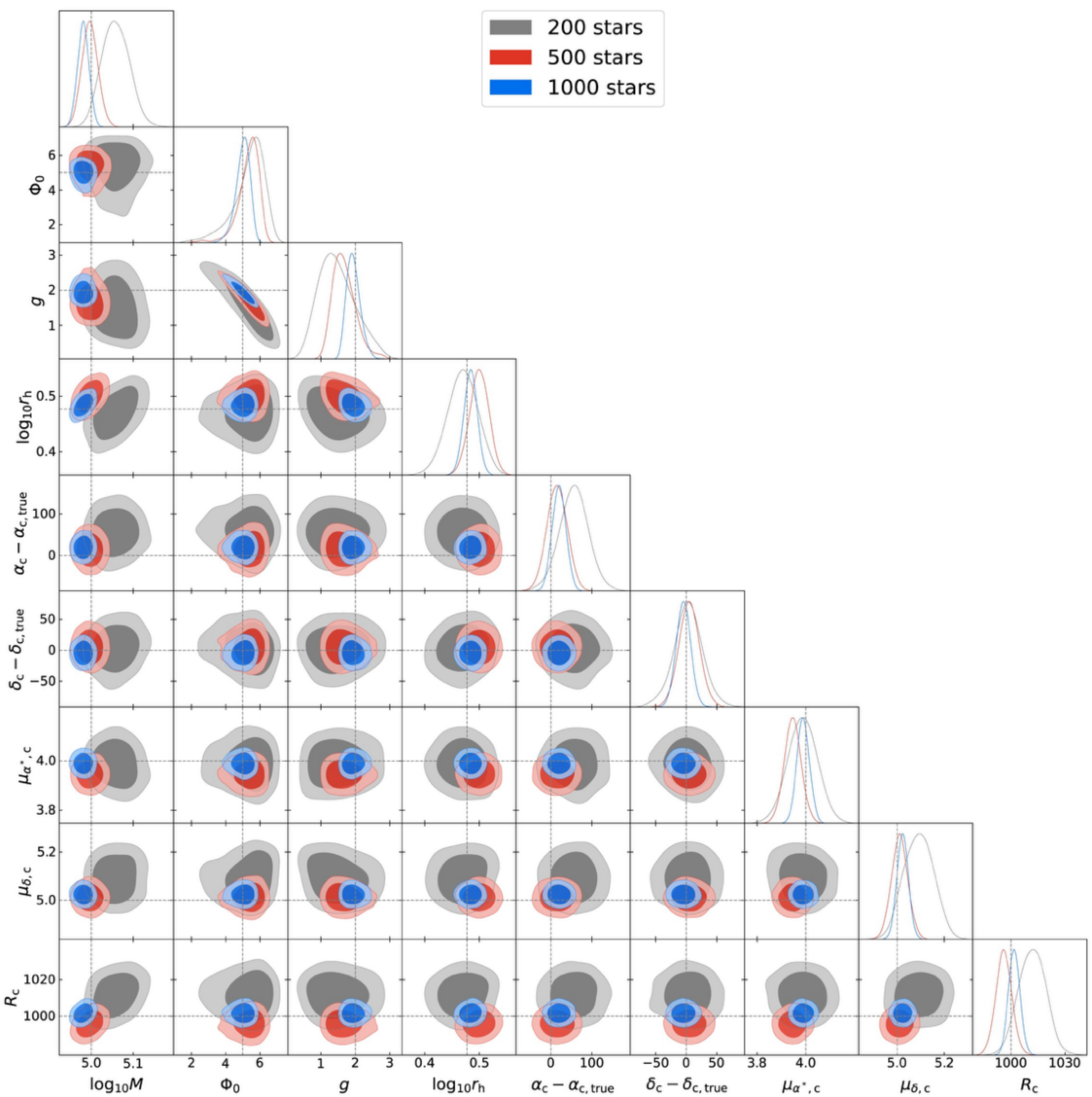
3 Results: structural parameters

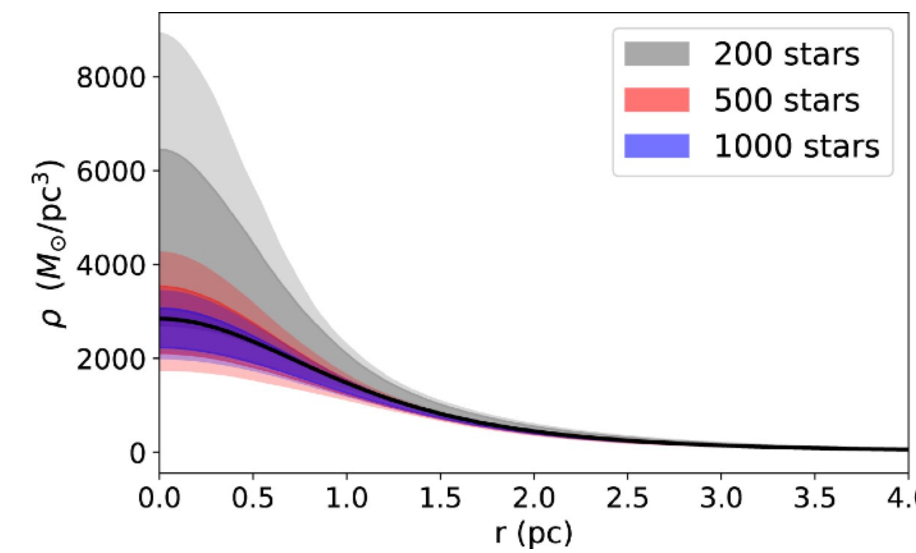


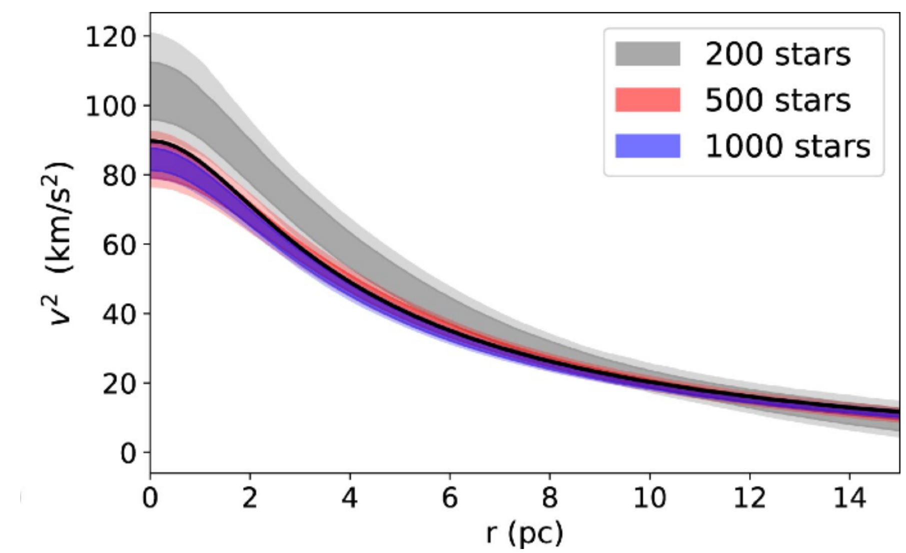
(a) True parameters $(\Phi_0, g, \log_{10} M_{tot}, r_h) = (3, 1.2, 5, 3)$



3 Results: different hyperparameters (N)

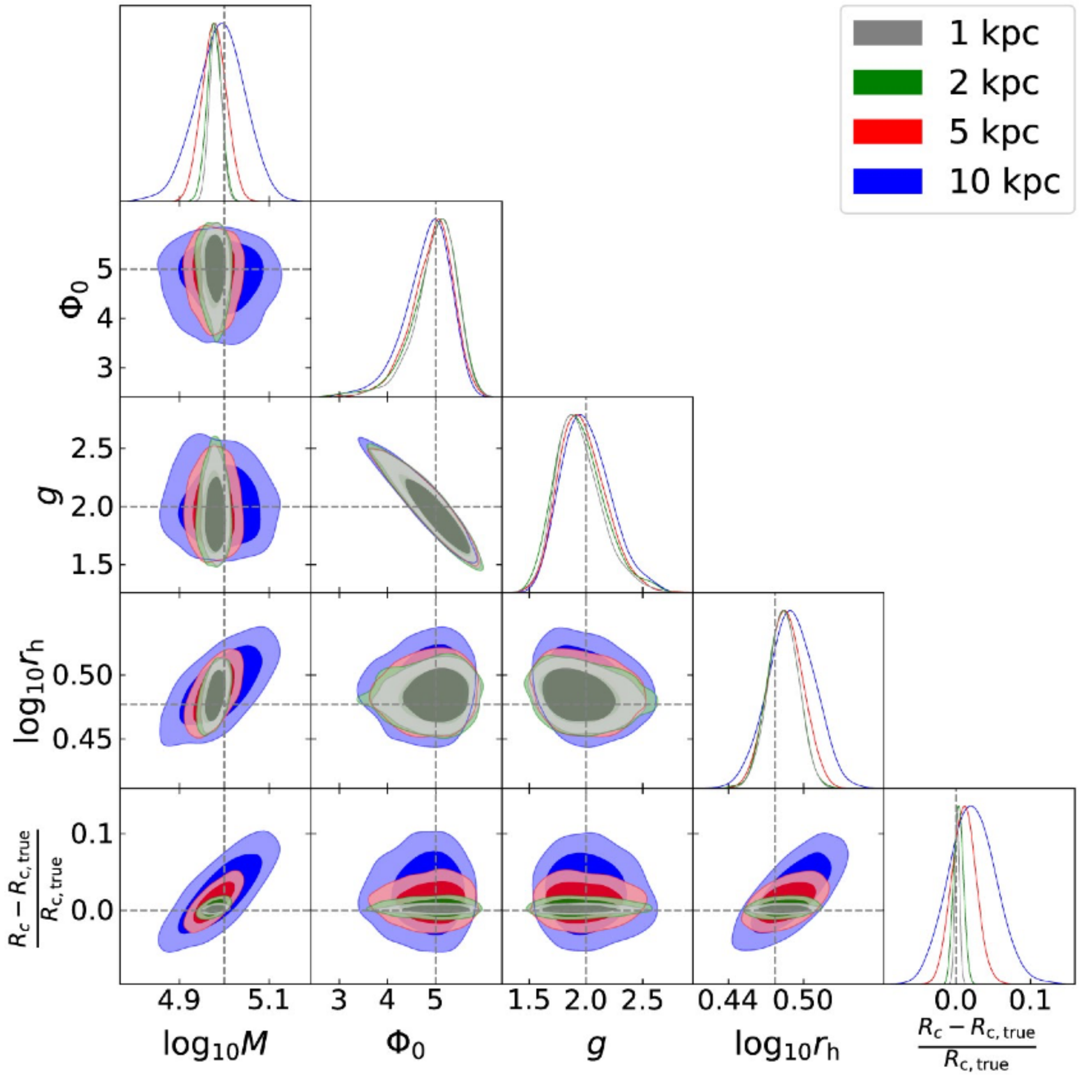




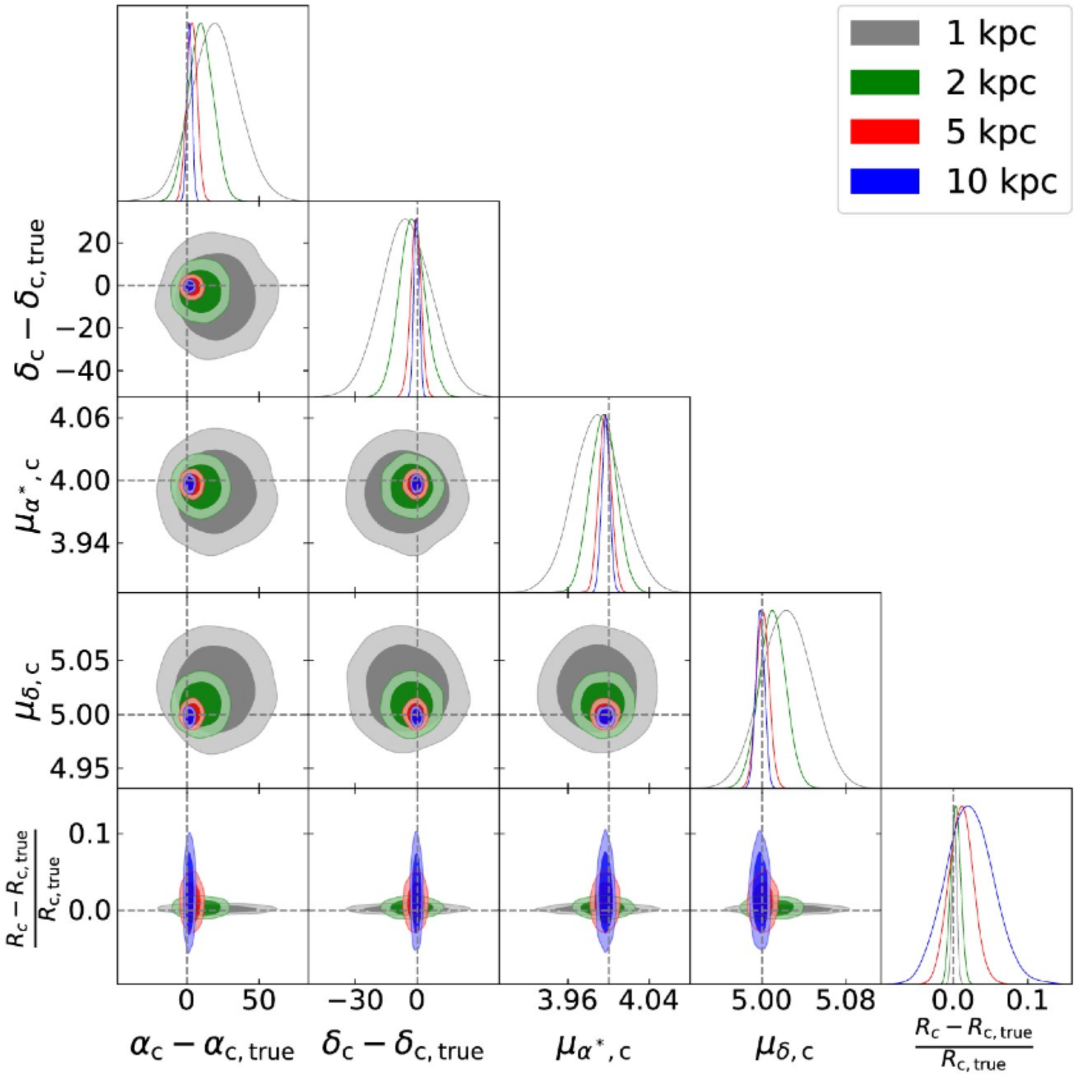


- Impact of different N stars
- 75 and 95 percentiles
- Larger N improve the results

3 Results: different hyperparameters (R_c)



(a) Estimates for the structural parameters and the radial distance

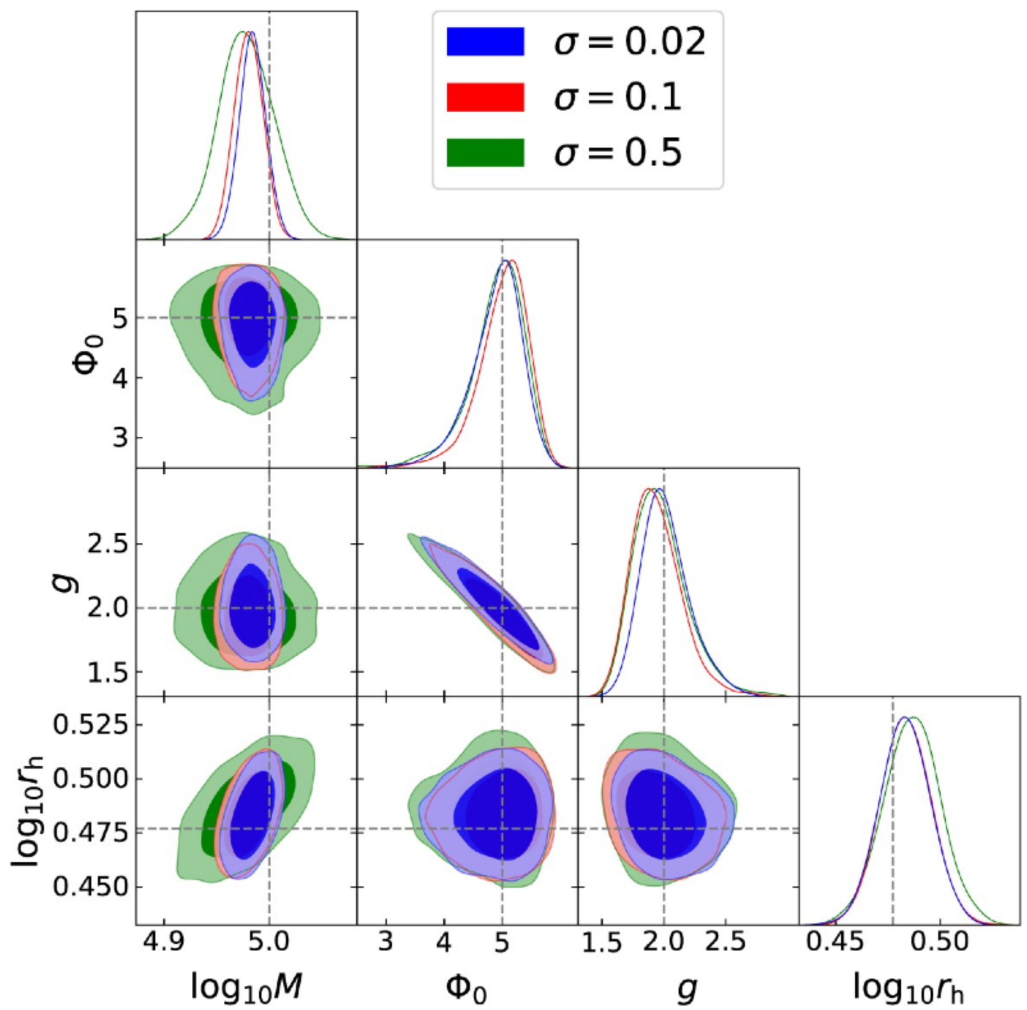


(b) Estimates for the GC center parameters

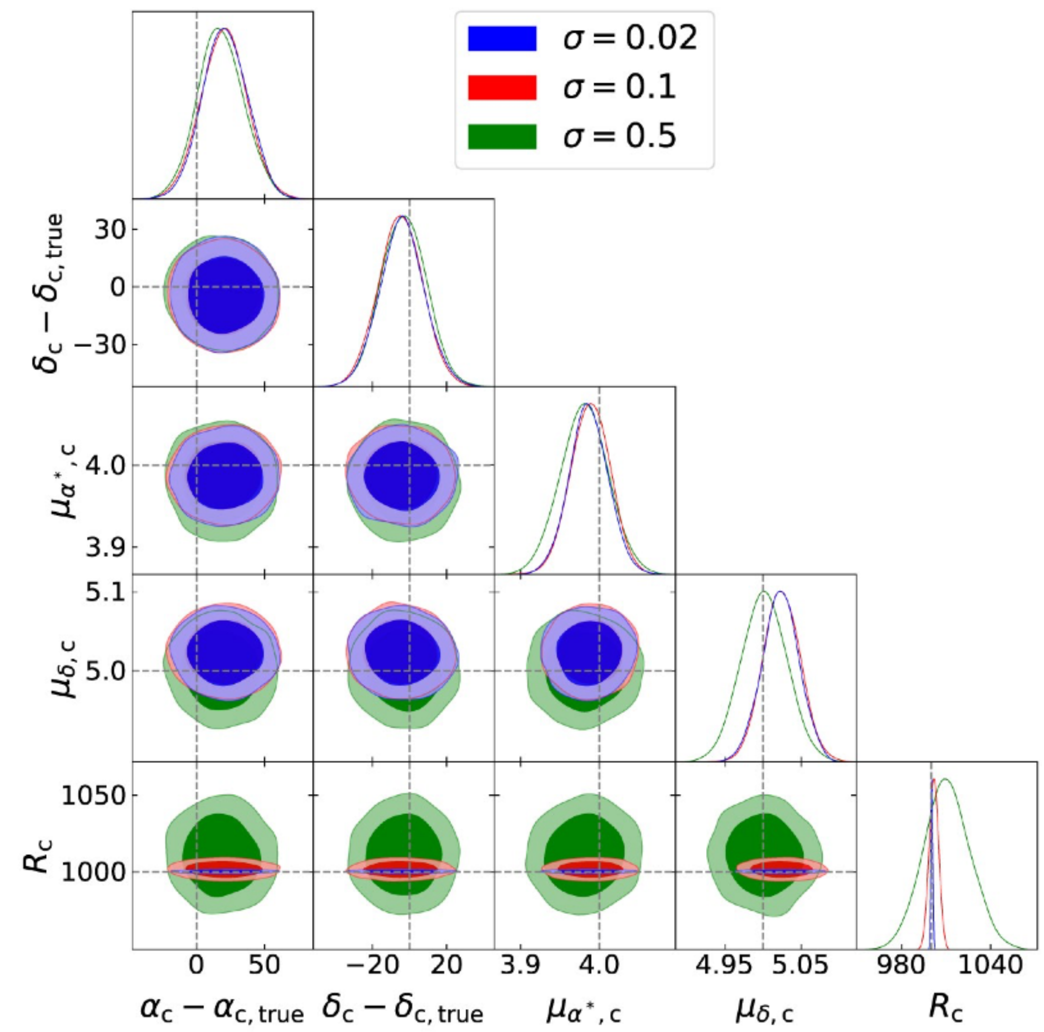
(Larger impacts in M and r_h)

(Uncertainties decrease with larger distances/smaller angular size)

3 Results: different hyperparameters (σ)



(a) Estimates for the structural parameters



(b) Estimates for the GC center parameters

(Uncertainty on parameters increase with σ. Radial distance is the most affected.)

4 Discussion and perspectives

- The new method (HBI including measurement errors) recover the parameter simulations (real values) within 95% of the credible interval

- Simplifications and perspectives:
  - single-mass, lowered isothermal DF (limepy) → multi-mass
  - technical problems with HMC, e.g. $g < 0.6$ and speed
  - simulations → observational effects, e.g. selection bias, crowding
  - uniform sampling of stars → different, heterogeneous datasets
  - compare with methods of radial binning
  - make limepy auto-differentiable…

- GitHub: limepy and hbmlimepy (source code and notebooks)