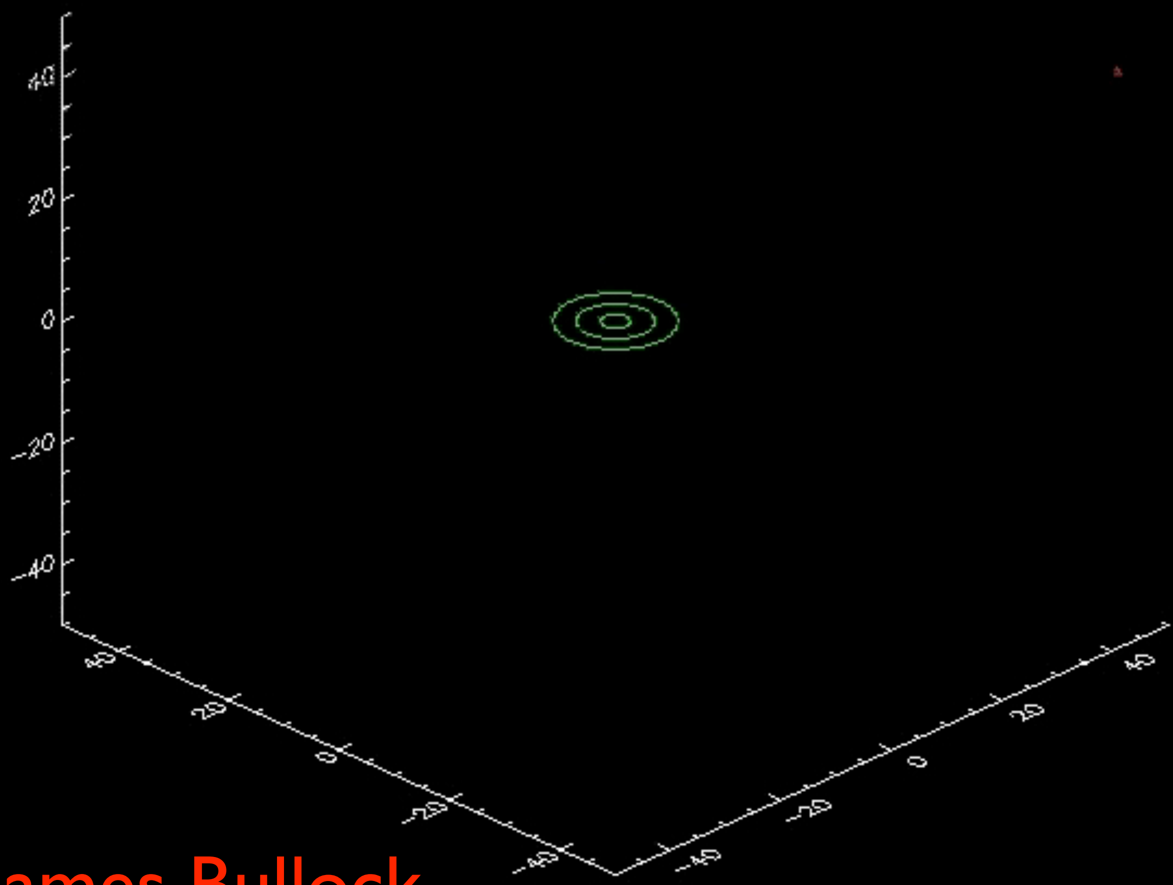
Reconstructing the Galaxy's accretion history: A finite mixture model approach

Duane Lee, Will Jessop, Kathryn Johnston & Bodhisattva Sen
Columbia University

Credit: Adam Evans (Wikipedia)

GISS 2011

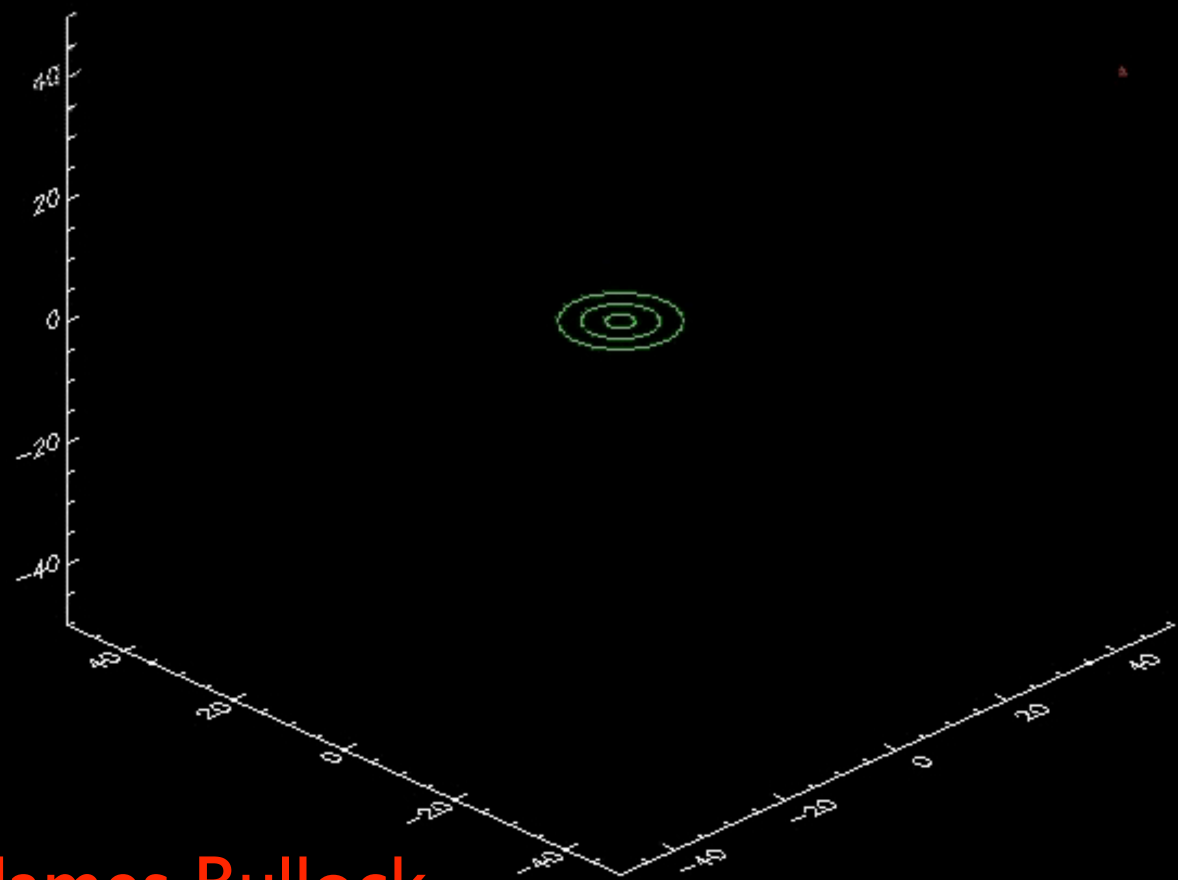
Reconstructing the Galaxy's accretion history



Credit: James Bullock

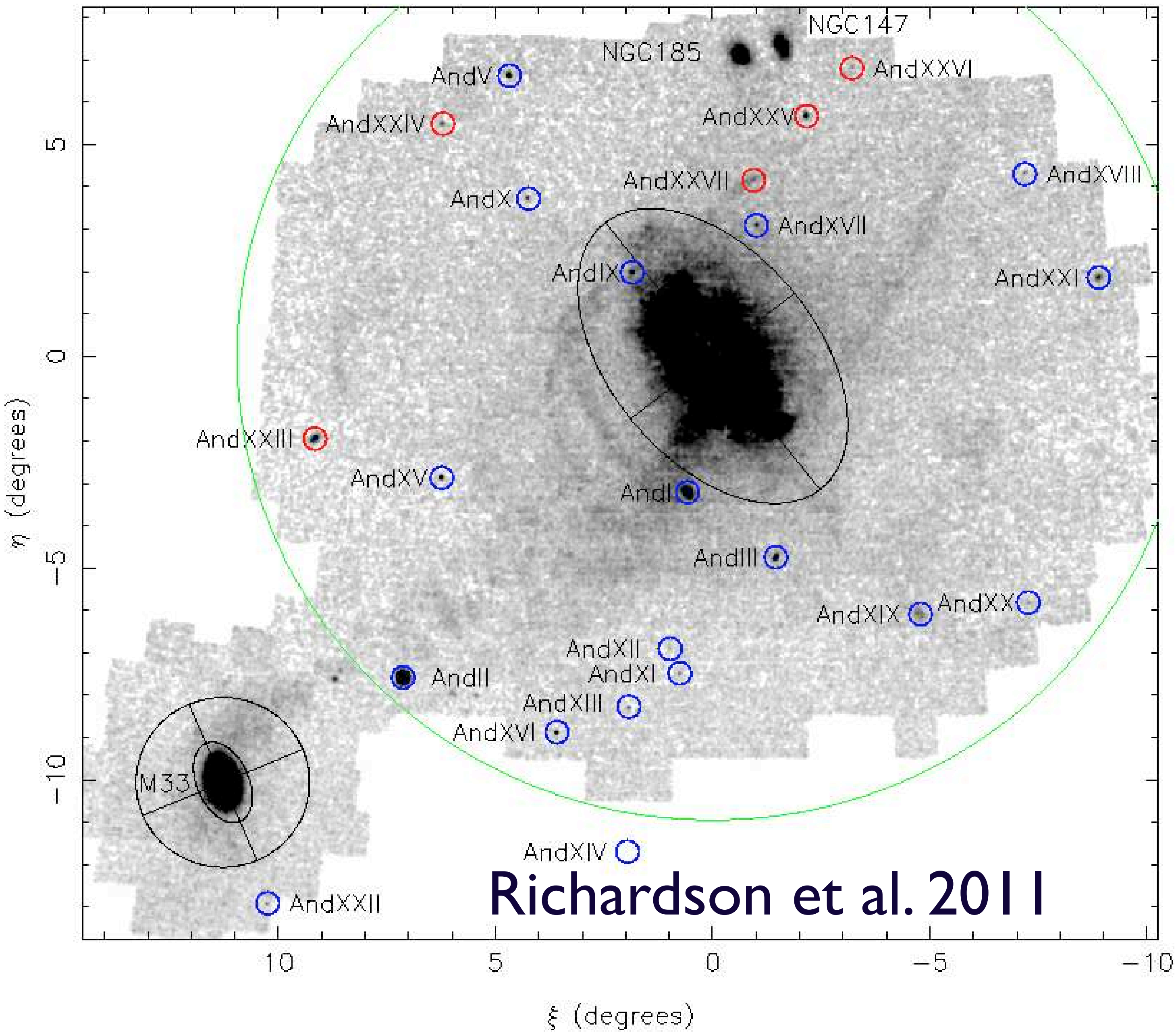
GISS 2011

Reconstructing the Galaxy's accretion history



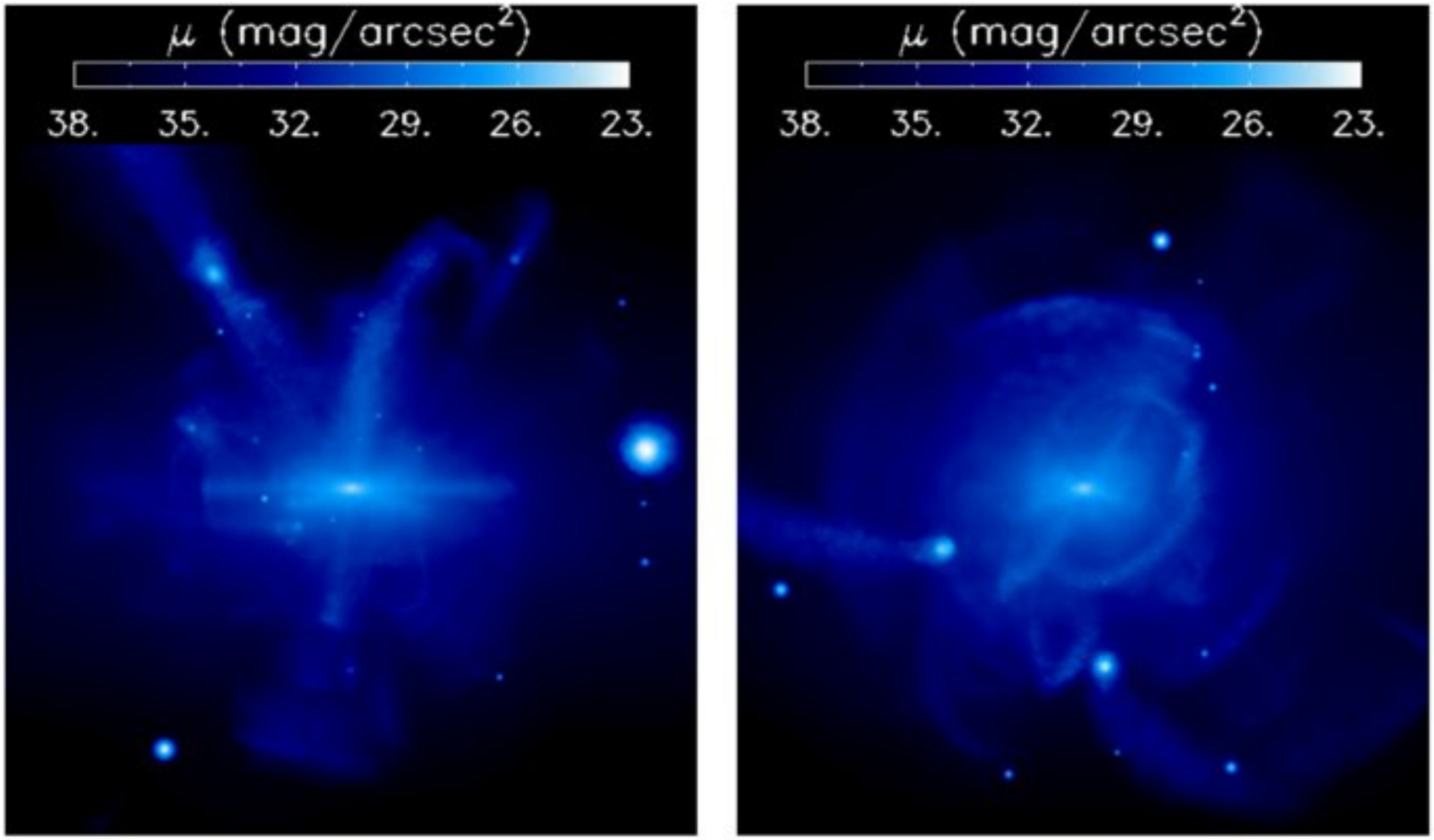
Credit: James Bullock

Reconstructing the Galaxy's accretion history



- Evidence for Hierarchical Merging
- Stellar halo "substructure" found using star counts
- Dynamical models can be applied to "extract" recent accretion history
- "Phase mixing" limits the scope of dynamical modeling (no streams)

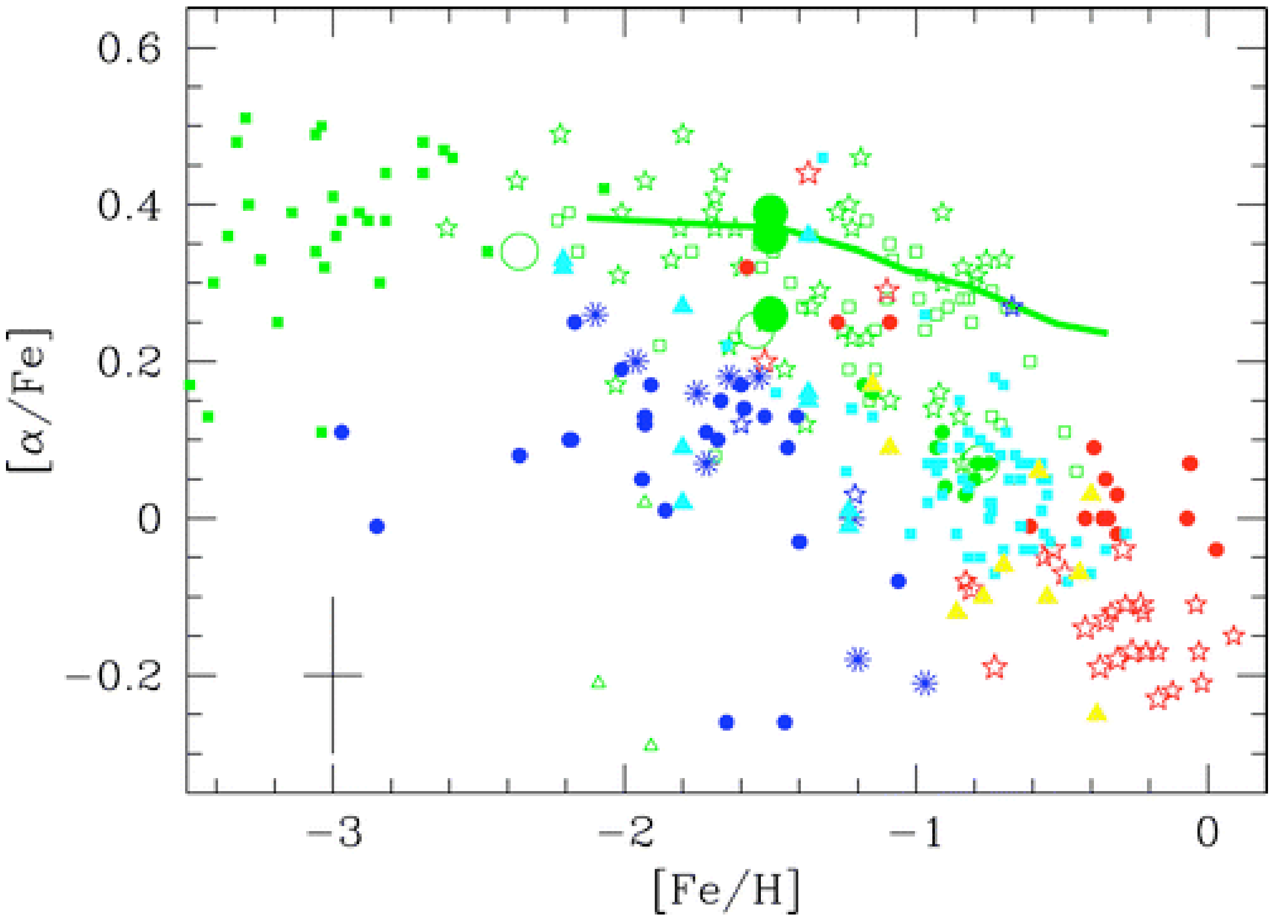
Reconstructing the Galaxy's accretion history



Bullock & Johnston 2005

- "Cosmologically"-motivated simulations track bulk stellar distributions

Reconstructing the Galaxy's accretion history



- Galactic Genealogy

➡Stars "remember" their "genetic" ancestry - that is, chemical abundances inherited from previous generations of stars

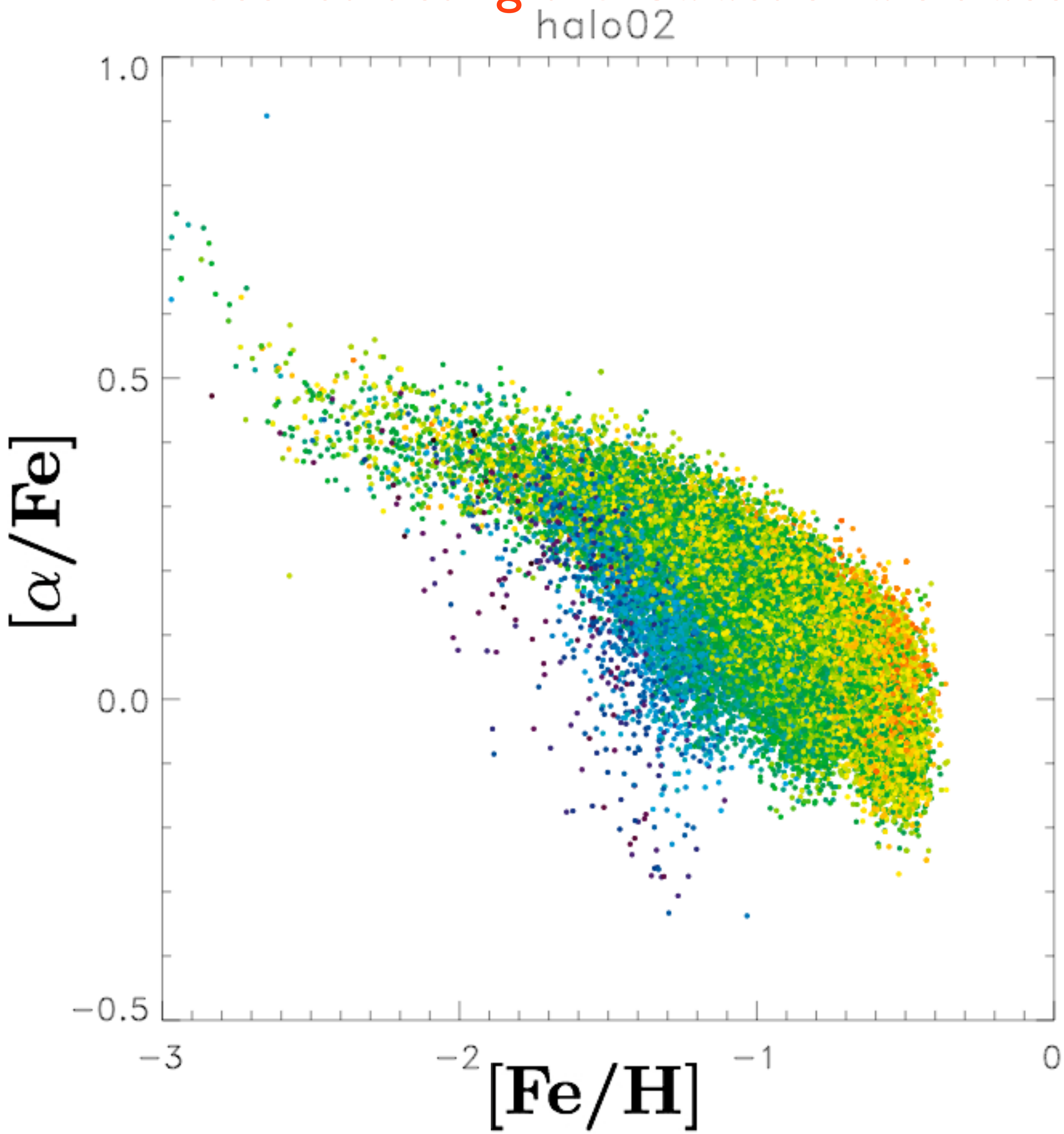
green - halo
blue - low mass dSph
yellow - dIrr
red - Sgr
cyan - LMC

(data compilation from Geisler et al, 2007)

- Observations reveal trends in 2-D metallicity-space

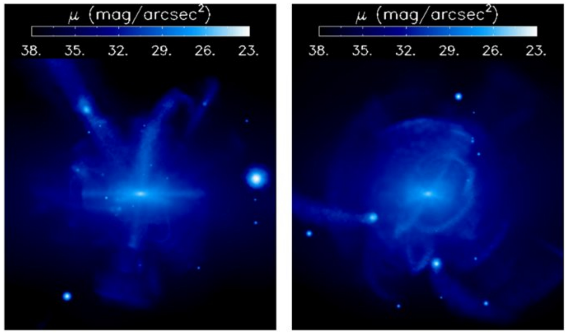
➡Metallicity distribution of satellites are correlated with their accretion time & mass

Reconstructing the Galactic halo's accretion history



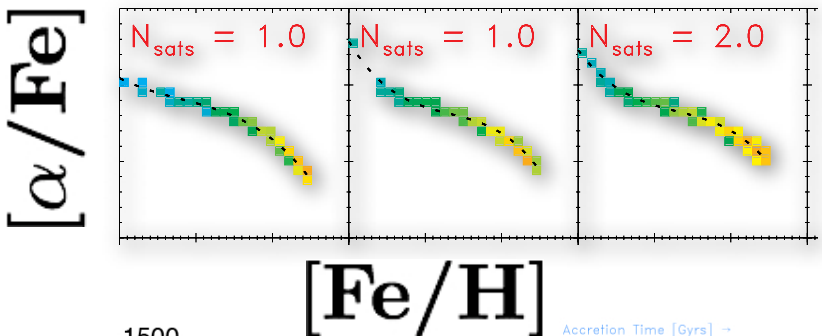
- The simulated halos also include a semi-analytical treatment of metal-enrichment via star formation

Reconstructing the Galaxy's accretion history

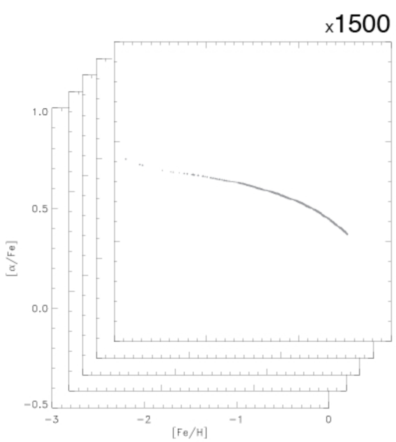


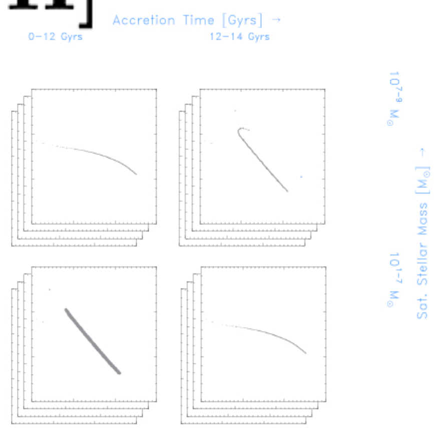
. . . } 11 Halos

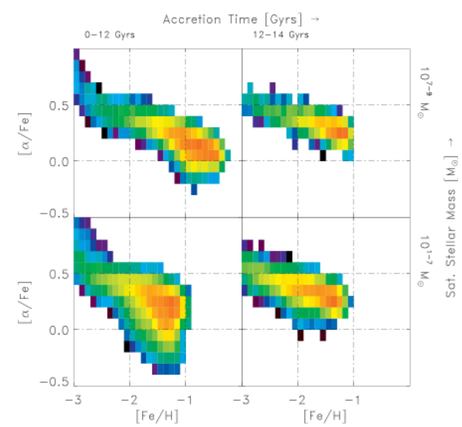
. . .



. . . } ~1500 sim. accreted dwarf galaxies

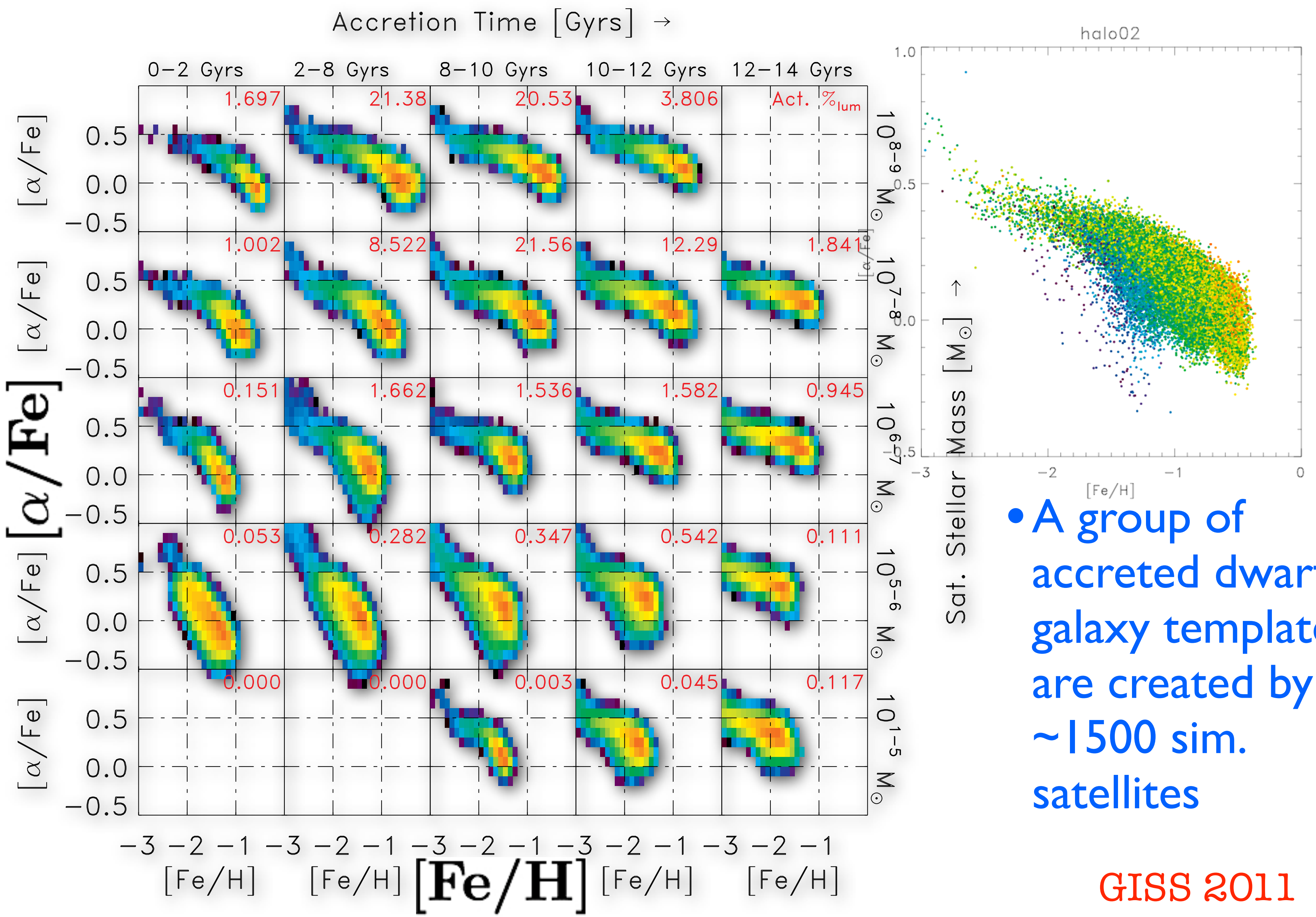






GISS 2011

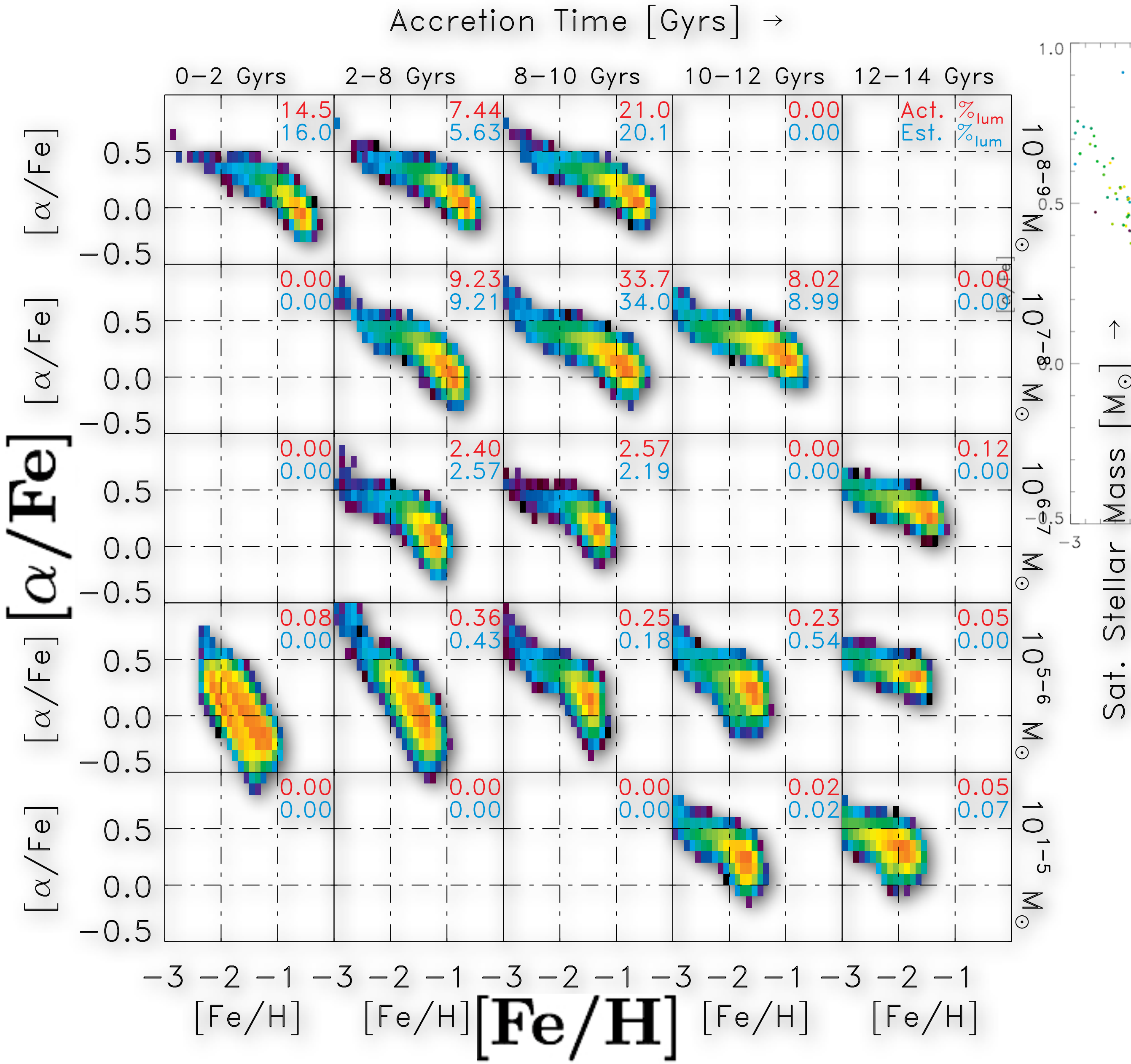
Reconstructing the Galaxy's accretion history

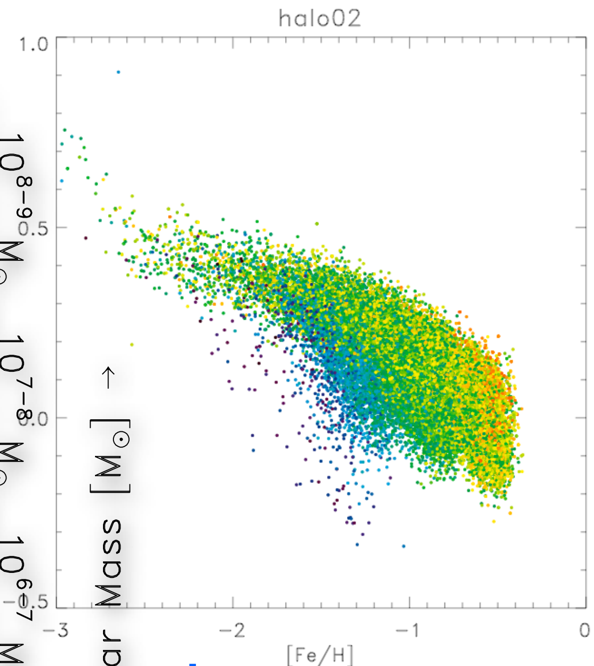


- A group of accreted dwarf galaxy templates are created by ~1500 sim. satellites

GISS 2011

Reconstructing the Galaxy's accretion history



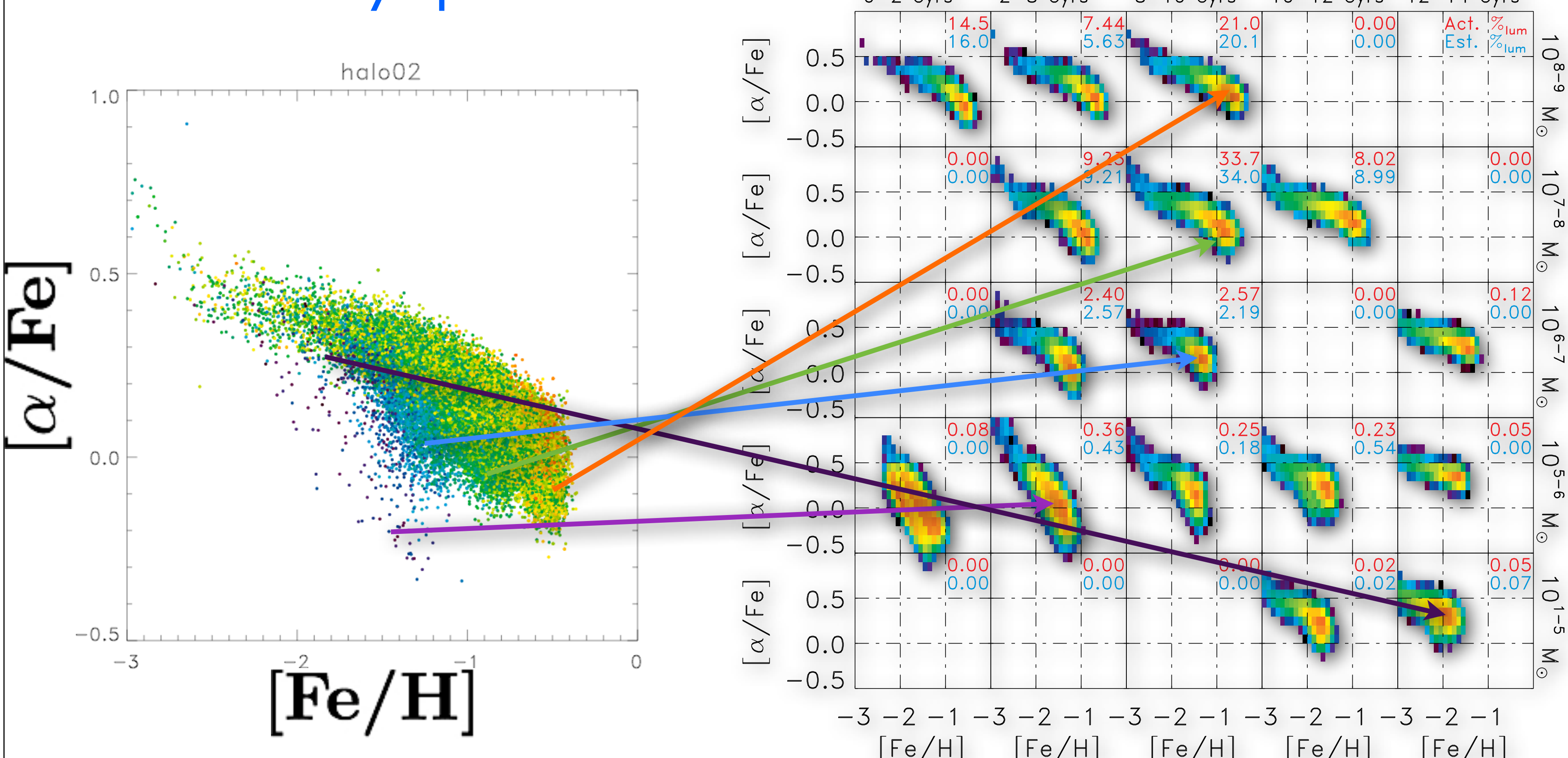


- Luminosity contributions are then ascertained via some fitting algorithm, e.g., the EM algorithm

GISS 2011

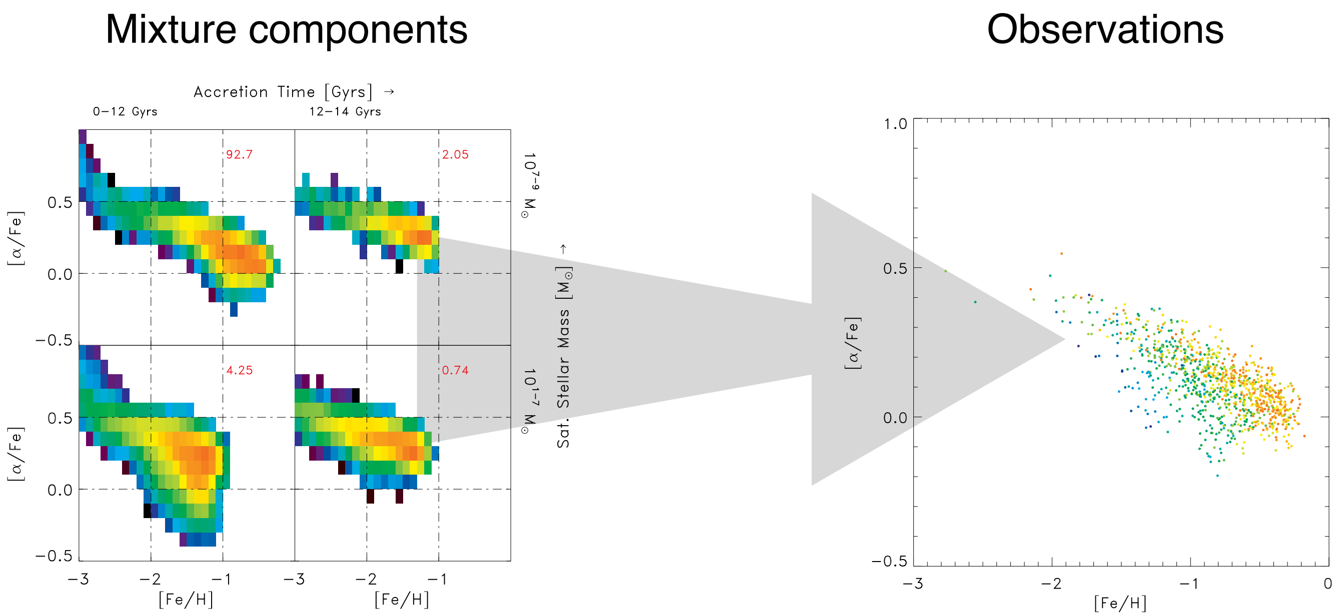
Reconstructing the Galaxy's accretion history

- Can we reconstruct the accretion history of the Galactic halo from stellar distributions in 2-D metallicity-space?



GISS 2011

A generative finite mixture model



$$\left[\frac{Fe}{H}, \frac{\alpha}{Fe}\right]_{i=1}^{N} \text{i.i.d} \sim f(x,y) = \sum_{j=1}^{m} \pi_j f_j(x,y)$$

Where the mixing proportions, $\boldsymbol{\pi}$, give the formation history.

Model definition

$$\text{Let} \qquad x = \frac{\alpha}{Fe}, \qquad y = \frac{Fe}{H}$$

Given m mixture components, we propose that the density from which observations are generated is

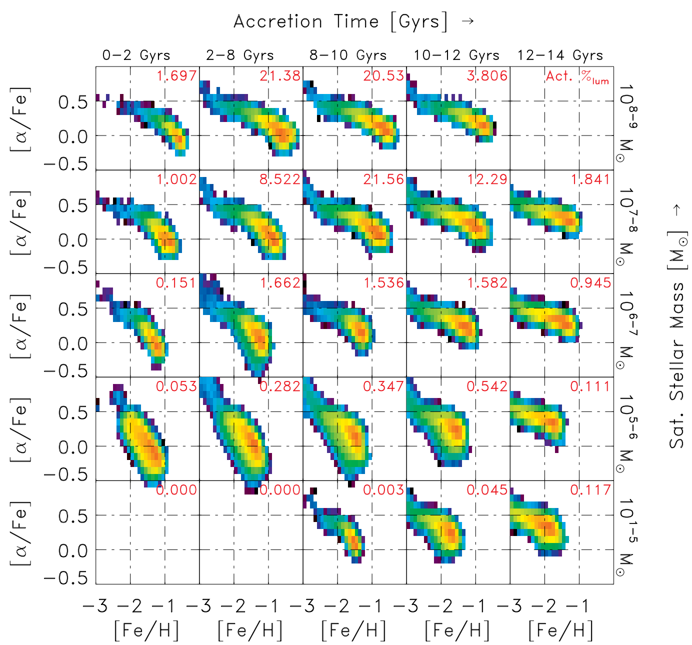
$$f(x, y) = \sum_{j=1}^{m} \pi_j \, f_j(x, y) \tag{1}$$

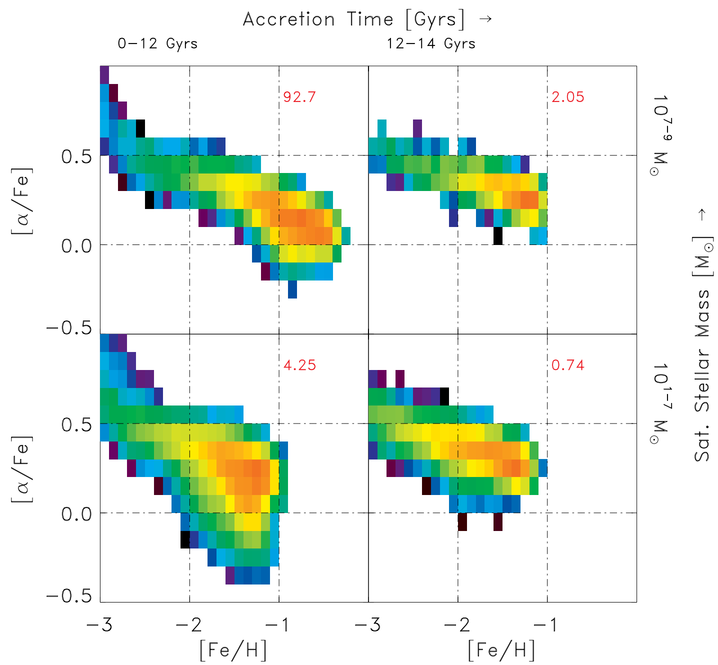
- Mixing proportion
- Mixture component j

$$\text{where} \qquad \sum_{j=1}^{m} \pi_j = 1, \qquad \pi_j \geq 0, \qquad j = 1, \ldots, m$$

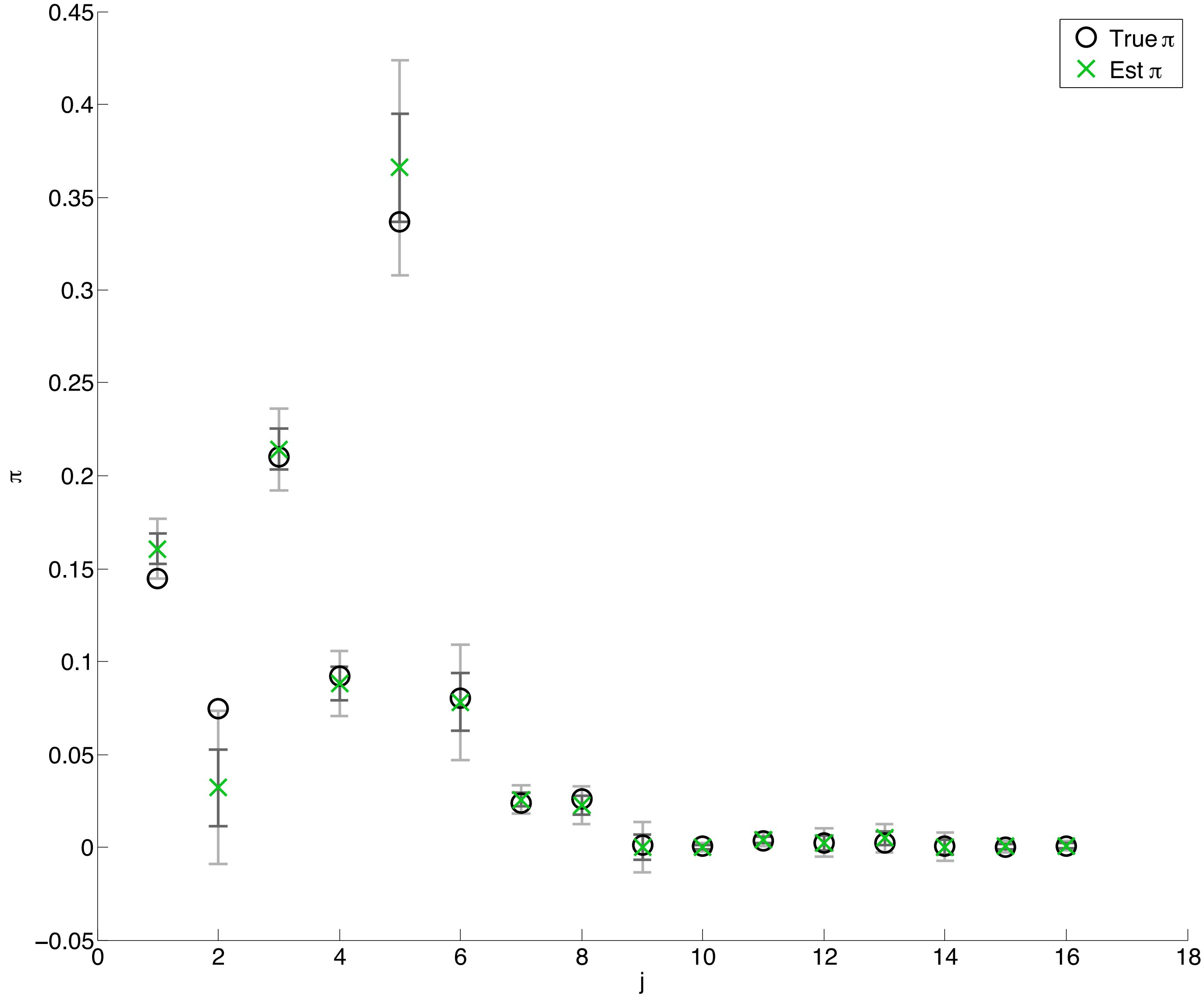
Simulations

- Generated observations from 11 realizations of halos
- Generated mixing components for these halos
- Used a 5x5 grid ($m = 25$), and several 2x2 grids ($m = 4$)

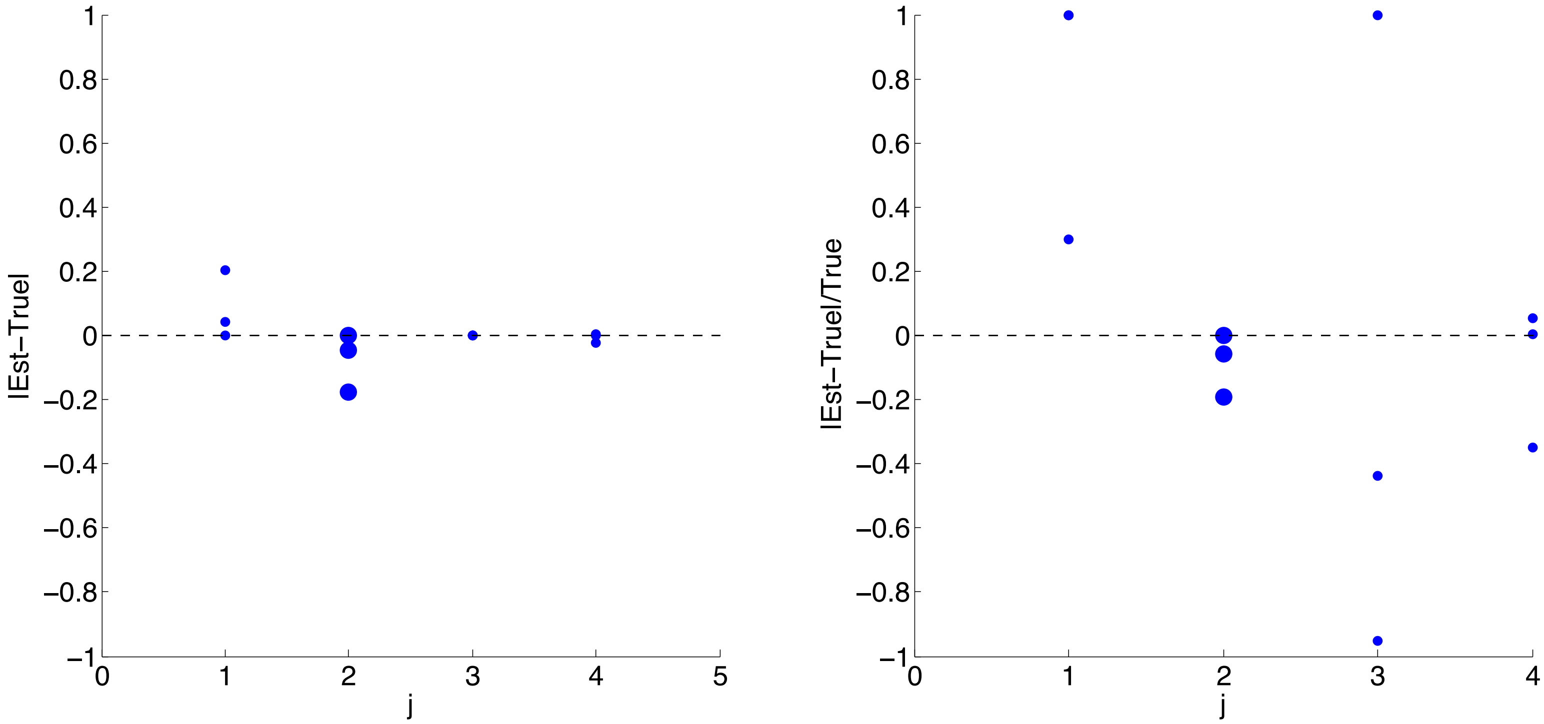




Results for a 5x5 grid ($m = 25$), for one halo realization



2x2: All 11 halos



Estimating the mixing proportions $\boldsymbol{\pi}$

To estimate the mixing proportions, we can use a maximum likelihood approach

$$\hat{\boldsymbol{\pi}}_{\mathrm{MLE}} = \underset{\boldsymbol{\pi}}{\operatorname{argmax}} \, \mathcal{L}(\boldsymbol{\pi})$$

where $$\mathcal{L}(\boldsymbol{\pi}) = \sum_{i=1}^{n} \log \left(\sum_{j=1}^{m} \pi_j f_j(x_i, y_i) \right)$$

Unfortunately the standard MLE procedure for estimating $\boldsymbol{\pi}$ is intractable with this likelihood.

The Expectation Maximization (EM) algorithm provides an alternative way to estimate $\hat{\boldsymbol{\pi}}_{\mathrm{MLE}}$

Expectation Maximization

Suppose we knew which mixture component f_j each observation came from:

$$z_{ij} = \mathbf{1}(x_i, y_i \sim f_j) = \begin{cases} 1 & (x_i, y_i) \sim f_j \\ 0 & \text{otherwise} \end{cases}$$

The log likelihood can then be expressed as

$$\ell(\boldsymbol{\pi}) = \sum_{i=1}^{n} \sum_{j=1}^{m} z_{ij} \log \left\{ \pi_j f_j(x_i, y_i) \right\}$$

The addition of the latent variable $\mathbf{z}$ actually makes things easier because it is easily differentiable in $\boldsymbol{\pi}$.

Estimating $\hat{\boldsymbol{\pi}}$ using expectation maximization

We don't know **z**, so we replace **z** with the expected value of **z**, conditioned on the data and the last known $\hat{\boldsymbol{\pi}}$:

$$\hat{\boldsymbol{\pi}}^{(t)} = \operatorname*{argmax}_{\boldsymbol{\pi}} \mathbb{E}\left[\ell(\boldsymbol{\pi}) \middle| \mathbf{x}, \mathbf{y}, \hat{\boldsymbol{\pi}}^{(t-1)}\right]$$

Starting with some random initial value for $\hat{\boldsymbol{\pi}}^{(0)}$, we iteratively

- Find the expected value of $\ell(\boldsymbol{\pi})$ using the current expected values of the latent variable **z**
- Set $\hat{\boldsymbol{\pi}}^{(t)}$ to the $\operatorname*{argmax}_{\boldsymbol{\pi}}$ of this expectation, which is simple to compute

And repeat until $\ell(\boldsymbol{\pi})$ stabilizes to a range $< 10^{-4}$

Find the expected value of $\ell(\boldsymbol{\pi})$ using the current expected value of the latent variable

The expected value of $\ell(\boldsymbol{\pi})$, with respect to the conditional distribution of $\mathbf{z}$, given observed data and $\hat{\boldsymbol{\pi}}^{(t-1)}$ is

$$\mathbb{E}_{\boldsymbol{\pi}}\left[\ell(\boldsymbol{\pi})\middle|\mathbf{x}, \mathbf{y}, \hat{\boldsymbol{\pi}}^{(t-1)}\right] = \sum_{i=1}^{n}\sum_{j=1}^{m}\mathbb{E}_{\boldsymbol{\pi}}\left[z_{ij}|x_i, y_i\right]\left\{\log f_j(x_i, y_i) + \log \pi_j\right\}$$

Since z_{ij} is an indicator, its expected value is simply the probability that data point i comes from model j

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\pi}}\left[z_{ij}|x_i, y_i\right] &= \Pr_{\boldsymbol{\pi}}(z_{ij}|x_i, y_i)\\ &= \frac{p(x_i, y_i|z_{ij}=1)p(z_{ij}=1)}{p(x_i, y_i)}\\ &= \frac{\pi_j f_j(x_i, y_i)}{\sum_{k=1}^{m}\pi_k f_k(x_i, y_i)}\end{aligned}$$

Find the $\underset{\boldsymbol{\pi}}{\text{argmax}}$ of this expectation

Now that we have the expected value of $\ell(\boldsymbol{\pi})$ with respect to the conditional distribution of $\mathbf{z}$, we need only evaluate

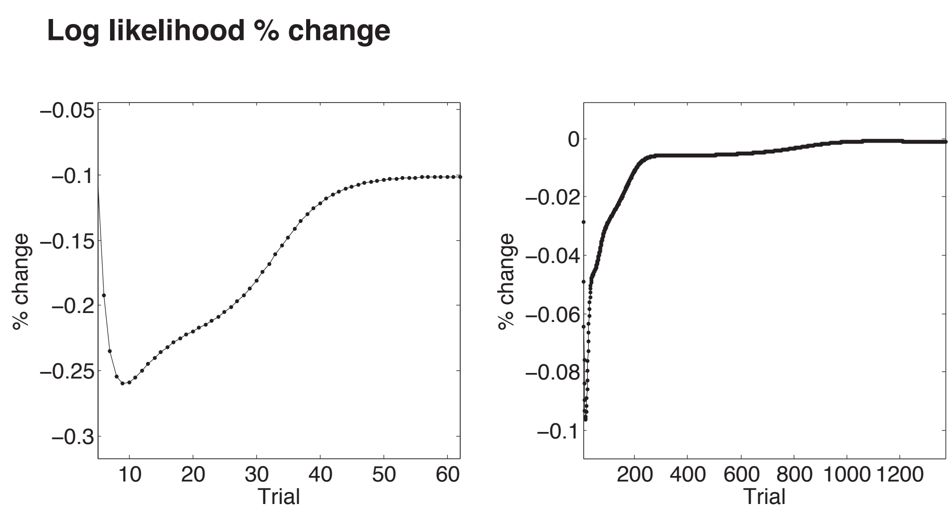
$$\hat{\boldsymbol{\pi}}^{(t)} = \underset{\boldsymbol{\pi}}{\text{argmax}}\, \mathbb{E}\left[\ell(\boldsymbol{\pi}) \middle| \mathbf{x}, \mathbf{y}, \hat{\boldsymbol{\pi}}^{(t-1)}\right]$$

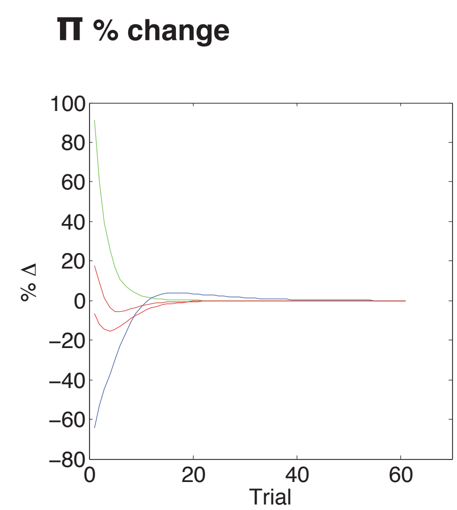
Which can be analytically specified, at each time t, as:

$$\hat{\pi}_k^{(t)} = \frac{\sum_{i=1}^{n} w_{ij}^{(t-1)}}{n}$$

$$\text{where} \qquad w_{ij}^{(t+1)} = \frac{\pi_j^{(t)} f_j(x_i, y_i)}{\sum_{k=1}^{m} \pi_k^{(t)} f_k(x_i, y_i)}$$

Convergence and minimum observation size





- Produces reasonable results few as 1,000 observations
- Confidence intervals narrow with more data
- Insensitive to initialization of π
- Large weights identified after 10 iterations
- $\ell(\pi)$ stops changing appreciably after 60 (m=4) or 600 (m=25) iterations
- Always converges

Covariance and confidence intervals

The asymptotic covariance matrix of $\hat{\boldsymbol{\pi}}$ can be approximated by the inverse of the observed Fisher information matrix, I:

$$I(\boldsymbol{\pi}'|\mathbf{x},\mathbf{y}) = -\frac{\partial^2 \ell(\boldsymbol{\pi}')}{\partial \boldsymbol{\pi}' \partial \boldsymbol{\pi}'^T}$$

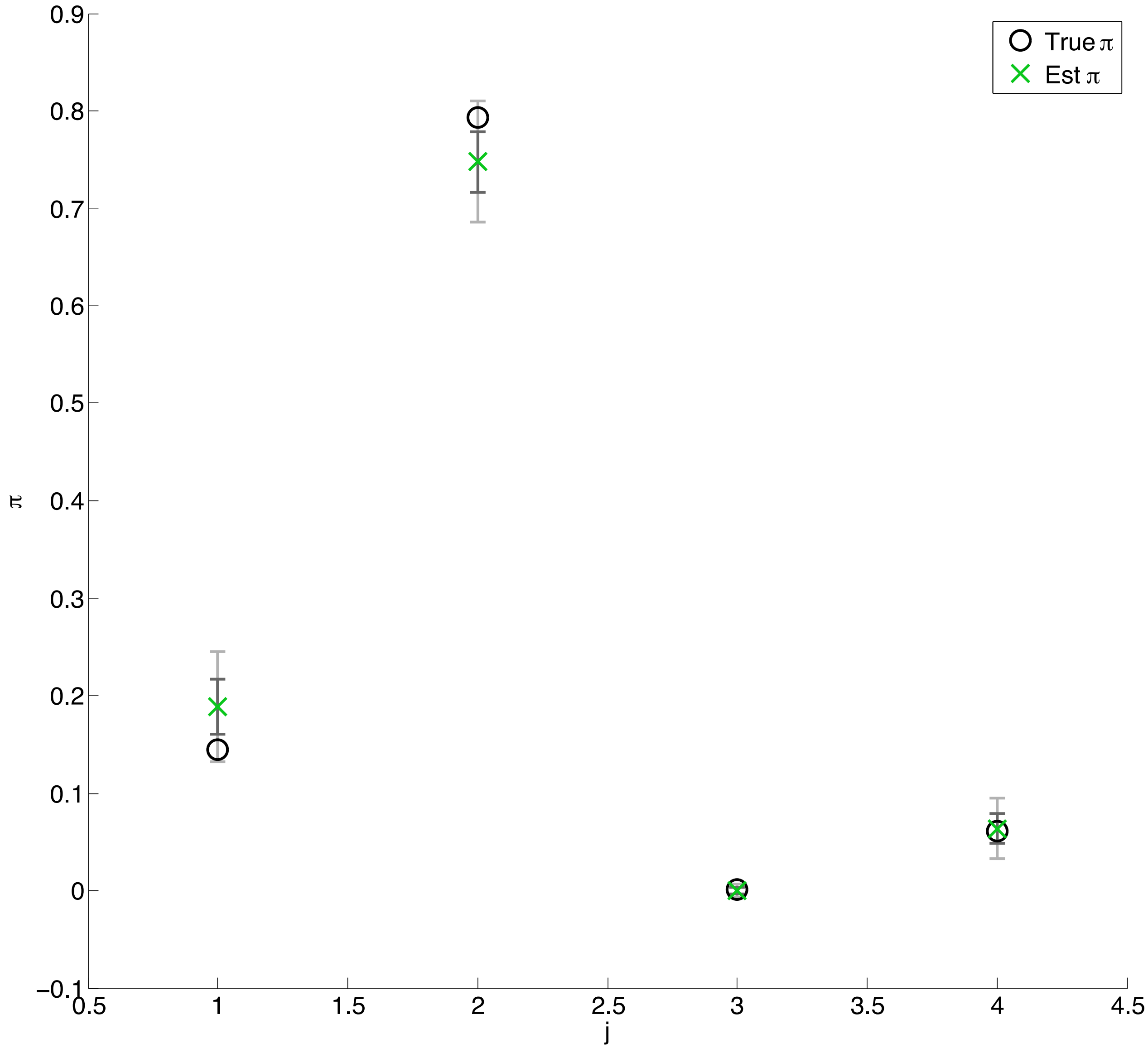
$$\mathrm{Cov}(\hat{\pi}_p, \hat{\pi}_q) = \left[I^{-1}(\hat{\boldsymbol{\pi}}') \right]_{pq}$$

with variance and correlation given by

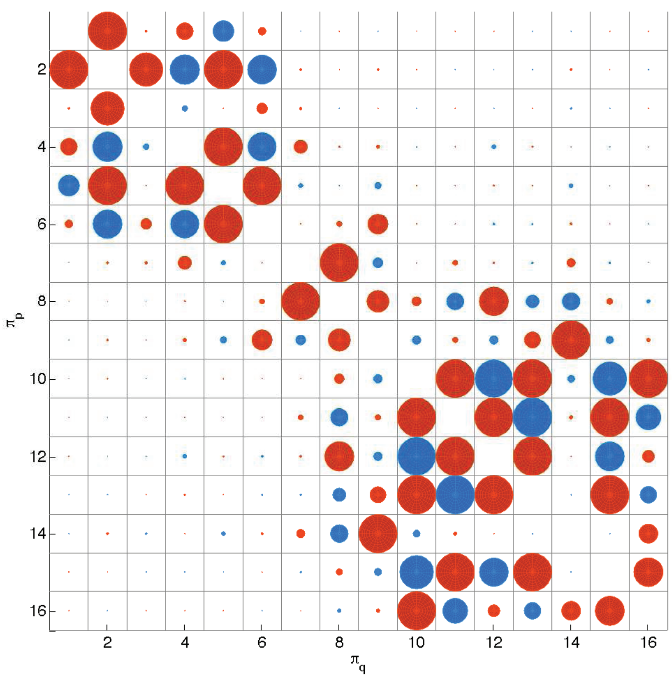
$$\mathrm{Var}(\hat{\pi}_j) = \sigma_j^2 = \left\{ \mathrm{Cov}(\hat{\boldsymbol{\pi}}) \right\}_{jj}$$

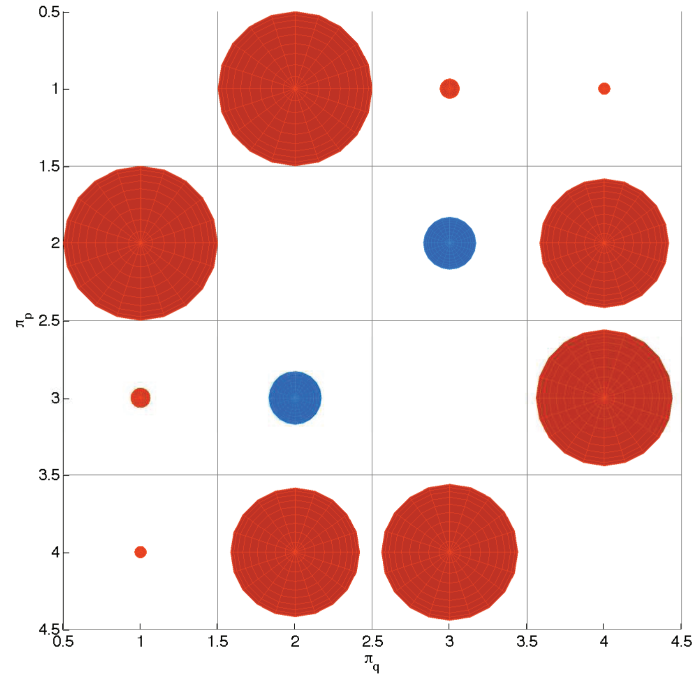
$$\mathrm{Corr}(\hat{\pi}_p, \hat{\pi}_q) = \frac{\mathrm{Cov}(\hat{\pi}_p, \hat{\pi}_q)}{\sqrt{\sigma_p^2 \sigma_q^2}}$$

Confidence Intervals: 2x2 results



Correlation between $\hat{\pi}$





Conclusion

- We were able to reconstruct the formation history
  - For multiple halo realizations
  - With a single finite mixture model
  - With good accuracy on a 2x2 grid
  - Relatively quickly
  - Equally well for large and small values of π_j
  - With more data, more granular grids could be used
- We found confidence intervals and covariance matrices for the mixing proportions
  - Fisher information and n-out-of-n bootstrapping produced nearly identical results

Future work

- Adaptive partitioning of mass and time since accretion
- Mixing components from smoothed metallicity curves

• Final Comments

- 2-D metallicities are a start... more dimensions to come!
- Our method should be strengthened by larger data sets (also necessary when expanding to higher dimensions)!
- Adding a more physical interpretation of models to sorting galaxies may help increase the effectiveness of the model templates in "extracting" reliable accretion histories