mocknotes

April 4, 2022

1 Fitting Mock Abundances with One-Zone Models

If we generate mock abundance data using one-zone models and artifically induced scatter, how well do fitting routines built around one-zone models perform in re-deriving the known parameters of the sample? How are the accuracy and precision of this procedure affected by factors such as sample size and measurement uncertainy?

1.1 The Mock Samples

We begin with a mock sample generated from a relatively standard one-zone model. Assuming the IMF-averaged yields of $y_O^{CC} = 0.015$, $y_{Fe}^{CC} = 0.0012$, $y_O^{Ia} = 0$, and $y_{Fe}^{Ia} = 0.0017$ from Weinberg, Andrews & Freudenburg (2017), we adopt an exponential infall history

$$\dot{M}_{in} \propto e^{-t/\tau_{in}}$$

where $\tau_{in} = 2$ Gyr and the overall normalization of the accretion history is inconsequential to the chemical evolution as the initial gas supply is 0. We assume the star formation efficiency (SFE) timescale $\tau_\star \equiv M_{gas}/\dot{M}_\star = 10$ Gyr and the outflow mass loading factor $\eta = \dot{M}_{out}/\dot{M}_\star = 25$. These values are chosen such that the population might reflect an at least semi-typical dwarf galaxy. For demonstrating this procedure in a paper, it may be good to motivate the values from the literature, but this will do for now.

1.1.1 The Fiducial Sample

To generate mock observations from this model, we sample 500 individual stars by taking random draws from the star formation history. We then artificially perturb the [Fe/H], [O/Fe], and log(age) of each individual star by $\sigma([Fe/H]) = \sigma([O/Fe]) = 0.05$ and $\sigma(\log(age)) = 0.1$.

We illustrate this sample below by plotting them in the [O/Fe]-[Fe/H] plane, color coded according to their ages.
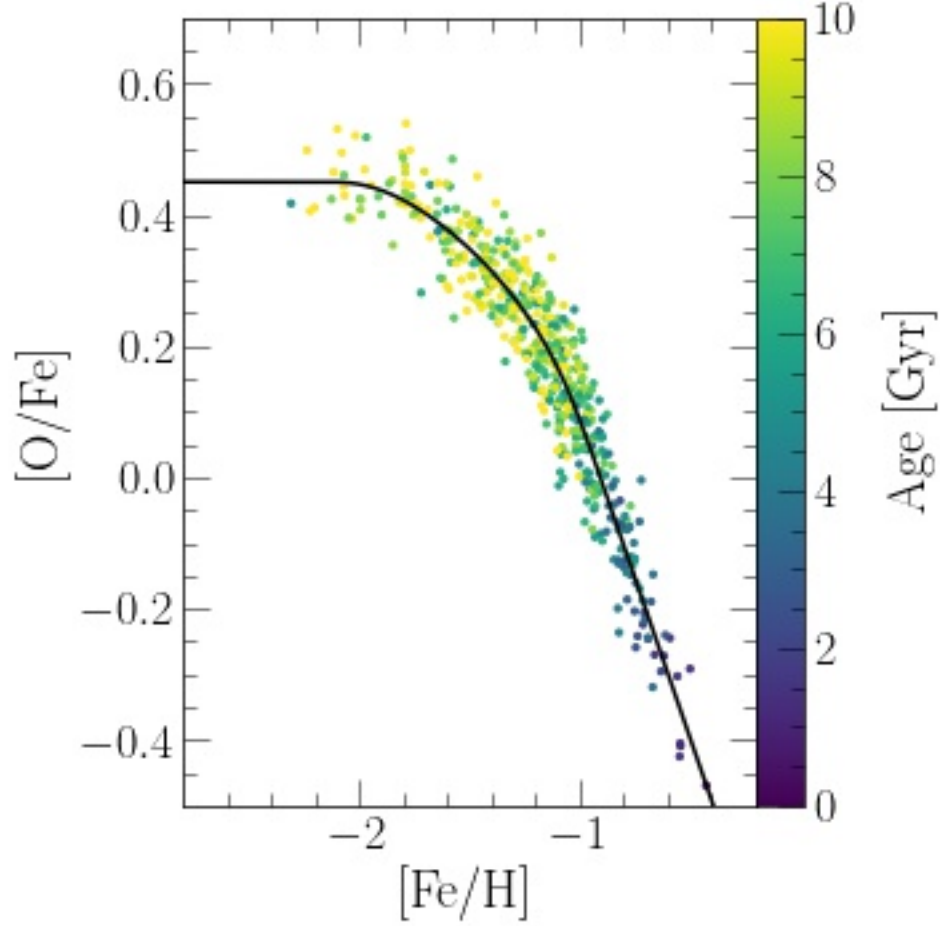
1.1.2 Alternative Samples

We additionly create variations of the fiducial sample, defined according to:

- **Precise**: This variation has $\sigma([Fe/H]) = \sigma([O/Fe]) = 0.01$ and $\sigma(\log(age)) = 0.05$ instead of 0.05 and 0.1, respectively.

- **Imprecise**: This variation has $\sigma([Fe/H]) = \sigma([O/Fe]) = 0.1$ and $\sigma(\log(age)) = 0.3$ instead of 0.05 and 0.1, respectively.

- **Small**: This variation has $N = 100$ individual stars rather than 500.

- **Large**: This variation has $N = 2000$ individual stars rather than 500.

- **No Ages**: In this variation, we do not record age information on any stars at all, and leverage the likelihood of a given set of parameters over their predicted [Fe/H] and [O/Fe] values only.

1.2 Fitting the Samples

To fit the samples, we use Markov Chain Monte Carlo methods with the emcee python package from Forman-Mackey et al. (2013). With 50 walkers starting in the volume defined by $\vec{m} \equiv (\tau_{in}, \tau_{\star}, \eta) =$

$(2, 10, 25) \pm (0.2, 1, 2.5)$ (i.e. 10% gaussian scatter around the known value), we conduct 50 steps of burn-in on each walker, and record their next 100 as the $N = 50(100) = 5000$ Markov Chain.

For a given realization of the model with a defined set of parameters $\vec{m}$ and one-zone model predictions $\vec{\mu} \equiv ([\text{Fe/H}], [\text{O/Fe}], \text{age})$, the likelihood of the data given the model is equal to the product of the likelihoods of each individual point given the model. That is:

$$L(d|m) \equiv \prod_i L(d_i|m)$$

$$\implies \ln L(d|m) = \ln \left(\prod_i L(d_i|m) \right)$$

$$= \sum_i \ln L(d_i|m)$$

For a given model m, however, there is no guaranteed way of knowing which point m_j along the [O/Fe]-[Fe/H] track should correspond to the data point d_i. We therefore marginalize over the entire track for every data point d_i by summing the likelihoods from all m_j model vectors. That is:

$$L(d_i|m) = \sum_j L(d_i|m_j)$$

Using the relation $L(d_i|m_j) \propto e^{-\chi^2/2}$ with $\chi^2 = \Delta_{ij} C_i^{-1} \Delta_{ij}^T$, where $\Delta_{ij} \equiv \vec{\mu}_{i,\text{data}} - \vec{\mu}_{j,\text{model}}$ and C_i^{-1} is the inverse covariance matrix of the ith data point, we arrive at the following expression $L(d_i|m_j)$, up to a normalization constant which we can safely ignore.

$$L(d_i|m_j) = \exp\left(-\frac{1}{2} \Delta_{ij} C_i^{-1} \Delta_{ij}^T \right)$$

However, periods of heightened star formation should contribute proportionally higher likelihood values to the fit simply because more stars are forming. Put differently, if we were assigning individual model points m_j to data points d_i, the point m_{j+1} could be a more suitable point to pair d_i with if the star formation rate was higher at time m_{j+1} than m_j. We therefore introduce weights w_j to each model vector m_j, where w_j is given by:

$$w_j = \xi \dot{M}_{\star,j}$$

where $\xi \equiv \left(\sum_k \dot{M}_{\star,k} \right)^{-1}$ over all timesteps k is included to ensure that $\sum_j w_j = 1$ (this proved important in practice, though we're not super clear on why). $L(d_i|m_j)$ then takes on the following form:

$$L(d_i|m_j) \to w_j \exp\left(-\frac{1}{2} \Delta_{ij} C_i^{-1} \Delta_{ij}^T \right)$$

This results in the following expression for $L(d|m)$:

$$\ln L(d|m) = \sum_i \ln \left(\sum_j w_j \exp \left(-\frac{1}{2} \Delta_{ij} C_i^{-1} \Delta_{ij}^T \right) \right)$$

In principle, inverting this expression to determine the model likelihood given the data $\ln L(m|d)$ requires applying Bayes' Theorem:

$$L(m|d) = \frac{L(d|m)L(m)}{L(d)}$$

Where $L(m)$ is the *prior* and $L(d)$ is the *evidence*. My understanding is the $L(d)$ is a constant of proportionality, while $L(m)$ is a distribution that we would convolve with our likelihood function $L(d|m)$ based on some motivation to do so to determine the posterior $L(m|d)$. The only true constraints on our fit parameters, τ_{in}, $\tau_\star$, and η, are that they are positive definite (though η can in principle be zero if we're taking the closed-box regime). We therefore take a flat prior in all parameters, effectively assuming

$$L(m|d) \approx L(d|m) = \sum_i \ln \left(\sum_j w_j \exp \left(-\frac{1}{2} \Delta_{ij} C_i^{-1} \Delta_{ij}^T \right) \right)$$

1.3 The Fit to the Fiducial Mock Sample

In the figure below, we show the 2-dimensional likelihood functions for each of the fit parameters in the fiducial model. The red "cross-hairs" denote the true, known values, while the blue stars denote the position of the maximum likelihood element of the Markov Chain and the full likelihood distributions in the background.
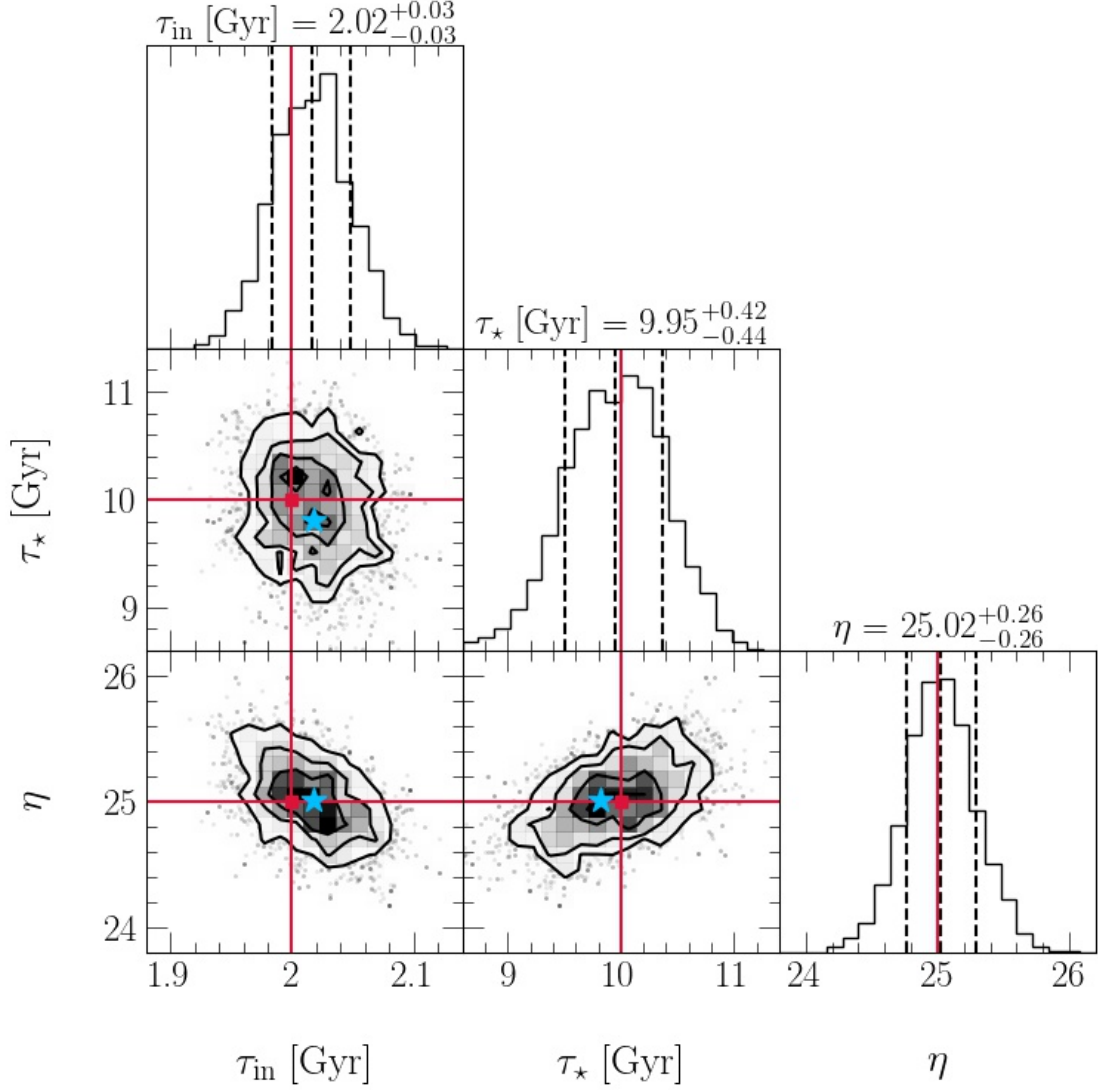
This fitting routine performs well in recovering the parameters of the known sample. The maximum likelihood parameters are within $1\,\sigma$ of the known values.

In the figure below, we show the [O/Fe]-[Fe/H] tracks of the known and best-fit models as well as the mock sample generated from the known evolutionary history, color-coded by the ages of the stars. In the remaining panels, we should the same comparison for the 1-dimensional distributions in age, [Fe/H], and [O/Fe].
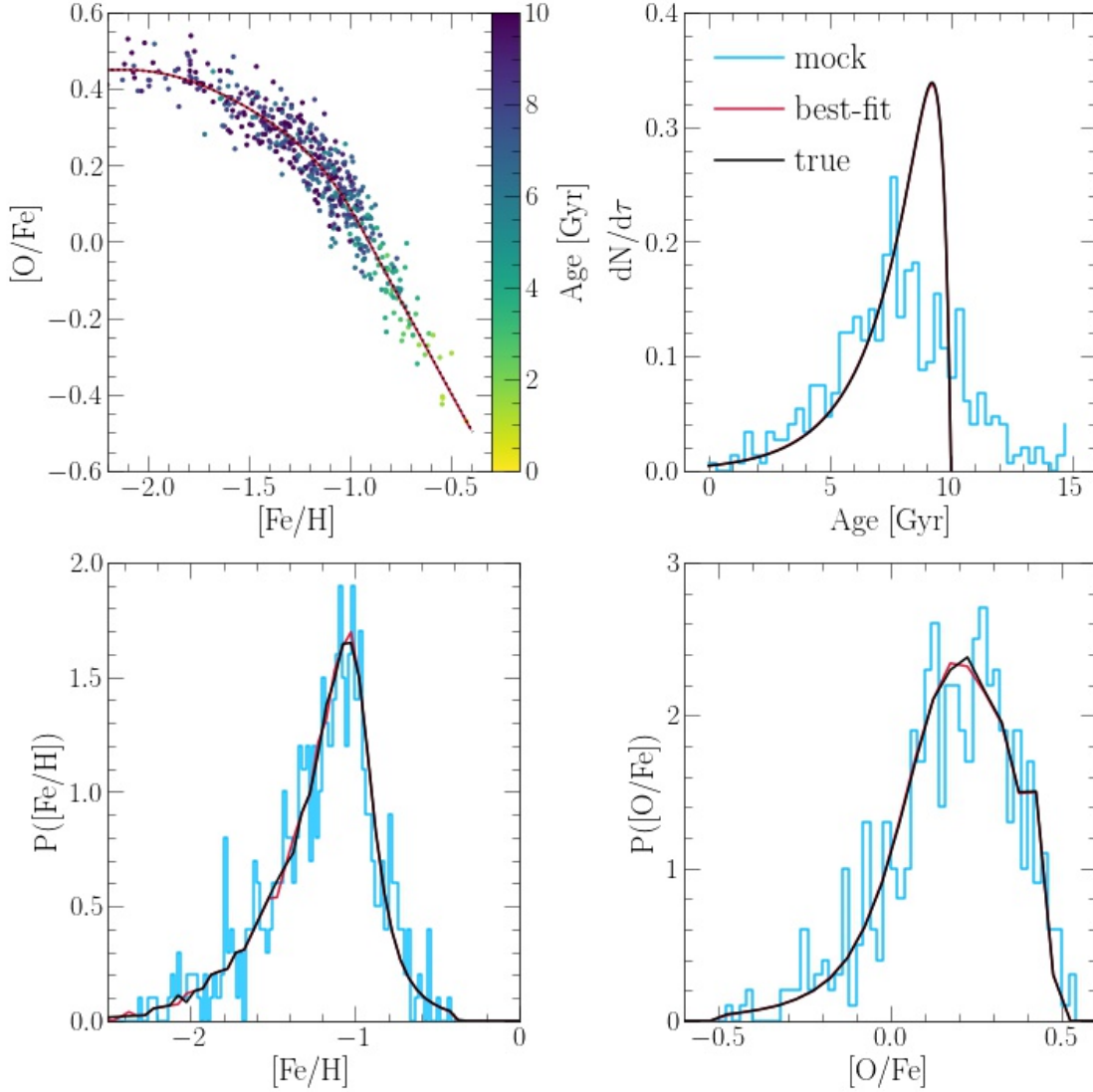
This fitting method recovers the distributions in these quantities well, largely because of the weights introduced in the fit. This ensures that regions of parameter space are assigned to be a good fit not only because they pass through the same regions of chemical space as the data, but also because they form the right number of stars in those regions. The fit is accurate enough that in multiple panels above, the known distribution/track and the best-fit distribution/track overlap visually.

The best-fit age distribution relative to the mock is interesting. The mock sample distribution appears much closer to a symmetric gaussian than the intrinsic distribution, with a peak at significantly lower ages than in the intrinsic distribution. This is due to the log-normal errors in this quantity in the mock data. This suggests that obviously unphysical ages (10-15 Gyr in this case since that precedes the onset of star formation, though observationally anything up to ~ 13.8 Gyr can be physical), still contain information that may be valuable in fitting chemical evolution models. This is due to them being a part of the upward scatter in age.

Remaining questions:

1. Let's say we eliminated all of the stars with ages > 11 Gyr from out sample. To what extent

does this systematically impact the derived age distribution?

2. What happens if we additionally fit for the time that we run the simulation for (i.e. the maximum value of the intrinsic age distribution)? How well does it recover this value?

1.4 What happens when we adjust precision and sample size?

| Which Sample	η	$\tau_\star$ (Gyr)	$\tau_{\rm in}$ (Gyr)
Imprecise	$25.25^{+0.75}_{-0.80}$	$9.84^{+1.72}_{-1.49}$	$2.04^{+0.10}_{-0.08}$
Precise	$24.99^{+0.06}_{-0.07}$	$9.93^{+0.10}_{-0.10}$	$2.01^{+0.01}_{-0.01}$
Fiducial	$25.02^{+0.26}_{-0.26}$	$9.95^{+0.42}_{-0.44}$	$2.02^{+0.03}_{-0.03}$
Large	$25.06^{+0.14}_{-0.14}$	$10.29^{+0.23}_{-0.27}$	$2.00^{+0.01}_{-0.01}$
Small	$25.77^{+0.68}_{-0.67}$	$11.70^{+1.19}_{-1.12}$	$1.84^{+0.08}_{-0.08}$
No Ages	$24.74^{+0.37}_{-0.35}$	$10.46^{+0.62}_{-0.63}$	$1.97^{+0.08}_{-0.09}$

- With a factor of 5 increase in precision in [Fe/H] and [O/Fe] and a factor of 2 in log(age), the likelihood distributions tightened by factors of ~ 4 in η, ~ 4 in $\tau_\star$, and ~ 3 in $\tau_{\rm in}$.
- With a factor of 2 loss in precision in [Fe/H] and [O/Fe] and a factor of 3 in log(age), the likelihood distributions widened by factors of ~ 3 in η, ~ 4 in $\tau_\star$, and ~ 3 in $\tau_{\rm in}$.
- With a factor of 4 increase in sample size, the likelihood distributions tightened by factors of ~ 2 in η, ~ 2 in $\tau_\star$, and ~ 3 in $\tau_{\rm in}$.
- With a factor of 5 decrease in sample size, the likelihood distributions widened by factors of ~ 2.5 in η, ~ 3 in $\tau_\star$, and ~ 2.5 in $\tau_{\rm in}$.
- Neglecting age information entirely widened the likelihood distributions by factors of ~ 1.5 in η, ~ 1.5 in $\tau_\star$, and ~ 3 in $\tau_{\rm in}$.

With the exception of the small sample, there is only one re-derived parameter with a $\gtrsim 1\sigma$ discrepancy with its known value: the value of $\tau_\star$ in the large sample is *marginally* inconsistent with the known value by only 20 Myr.

In the case of the small sample, however, there are ~ 2, ~ 1.5, and $\sim 1\sigma$ discrepancies in *each of* $\tau_{\rm in}$, $\tau_\star$, and η. This suggests that for small sample sizes, the quality of the fit suffers quickly

Folding all results together, to first order it seems that the precision of the fit is determined by the precision of the measurements while the accuracy of the fit is determined by the sample size. In general, all fits are accurate provided that the sample contains $\gtrsim 200$ stars, but there are no diminishing returns in measurement precision within the range probed here.

Interestingly enough, the fit performs more accurately and more precisely when *no age information is available at all* than when the sample size is too small but contains age information. We demonstrate this in the figure below, same as the one immediately above but for the case where the sample is only $N = 100$ stars,
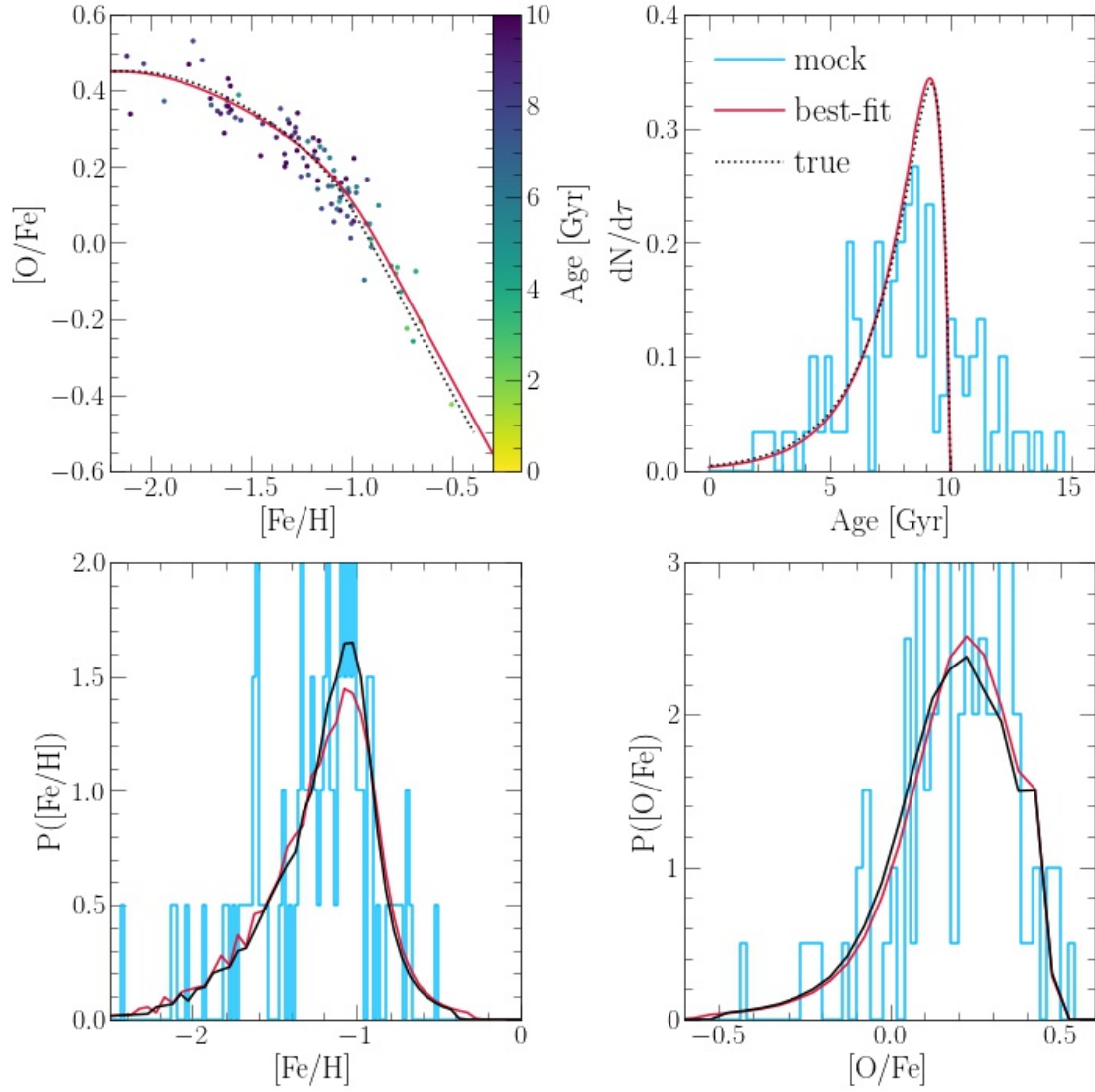
and in the next one as well showing the same thing for the sample with $N = 500$ but no age information.
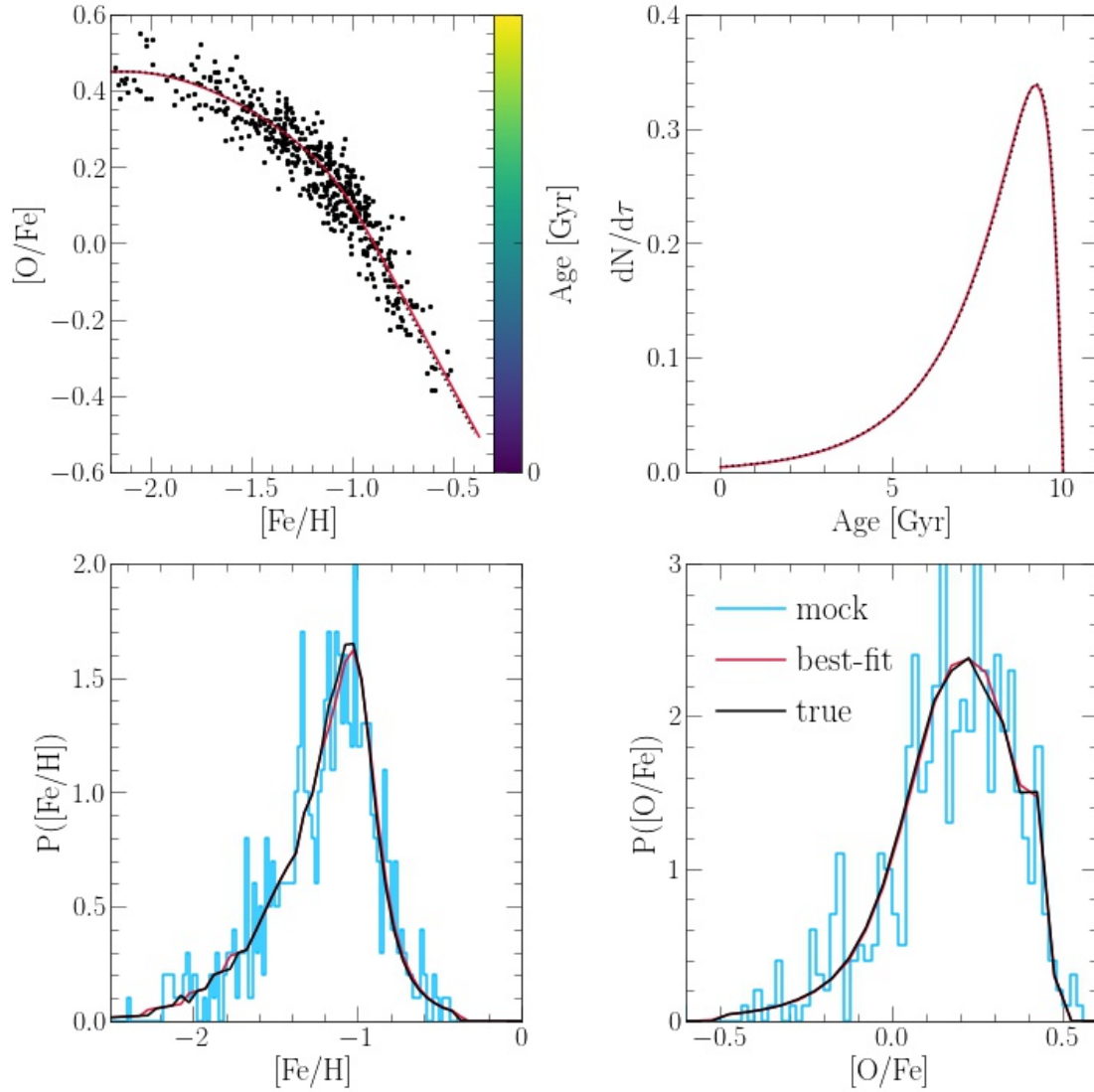
Although the best-fit model to the small sample isn't a terribly inaccurate description of the known, underlying distributions, the best-fit case to the ageless sample is noticeably more accurate and precise.

[ ]: