CASBI: Chemical Abundance Simulation Based Inference

Giuseppe Viterbo

June 28, 2024

Contents

Chapter 1

Abstract

Galaxies evolve trough merging events and destroy lower-mass systems over their lifetimes. The contribution that those lower-mass system brings to the modern picture has been freezes in stellar halos by the long orbital timescales, making the relicts of this objects retain part of their initial progenitor orbit. But dynamical information is not enough to disentangle these components, and the complementary chemical information helps to characterized these building blocks. In fact, merging events tend to quench the star formation rate of these objects, making the chemical abundance plane (iron abundance against α element abundance) a distinct imprint that retains information on the conditions of formation of their stars, like the total mass and the age of the system until the merging event. These theoretical background allows us to attempt to decompose the stellar halo into its components, unraveling the merging history. In the modern era of large N-body galaxy simulations, we recast this problem into an SBI pipeline to recover the properties of this building block, (e.g. total stellar mass, infall time, ...) using the chemical abundance plane as observables. We therefore present CASBI (Chemical Abundance Simulation Based Inferece), a python package to recover the posterior probability of properties of building blocks of Milky Way like galaxy's halo. Moreover, CASBI incorporate conditional neural network architectures as generator to obtain observables from parameters, smoothly interpolating on region of the parameters space that weren't fully covered during the N-body simulations.

Chapter 2

Previous work

2.1 The reconstruction of the Assembly history of the Milky Way

Inferring the assembly history of the Milky Way is a challenging task, even in the era of the astrometric Gaia mission and its 6 dimensional phase space data, and the complementary chemical information obtained from the wide-field spectroscopic programs such as the GALAH survey [6], the H3 survey [3], APOGEE [15], RAVE [21], SEGUE [26], and LAMOST [4]. The dynamical times of the accreted objects are far longer than the age of the host galaxy, allowing the phase space to retain part of the information on the original orbit parameters. On the other hand, the chemical space is dependent on the star formation history, in particular type II SNe produce α-elements and iron with a almost constant ratio, while type Ia SNe produce more efficiently iron. Another factor that governs the chemical space is the total mass of the galaxy, since the more massive galaxies are more capable to resist the expulsion of metals due to feedback mechanism. The crossmatch between Gaia and spectroscopic data allowed for the discovery of the "Gaia-Sausage-Enceladus" (GSE) ([2], [10]), a massive accretion event whose remnant now dominates the observation of the inner stellar halo of our Galaxy The GSE is described as major structure with mostly highly eccentric, retrograde orbit with a chemical abundance distribution of stars that highly distinct from the thin and thick disc star of the Milky Way, as it is possible to see in Fig 2.1.

Robustly identify distinct structure is challenging, and disentangle the components in fully phase mix situations is nearly impossible. In order to characterize the assembly history [5] propose to use the "CARDs", the chemical abundance ratio distributions of the stars, obtained from a subsample of accreted object candidate from the FIRE-2 zoom-in cosmological simulations of MW-mass galaxies [25]. Although similar to CASBI on how to leverage N-body simulations, this method do not recovers posteriors for the parameters of the accreted objects but rather considers the host halo as a linear combinations of templates CARDs

$$\mathrm{CARD}_{\mathrm{halo,\ model}}(x_d) = \sum_i \sum_j A_{ij} \mathrm{CARD}_{\mathrm{temp},ij}(x_d | M_{\mathrm{sat},i}, t_{100,j}), \tag{2.1}$$

treating each coefficient A_{ij} as the fraction of mass contribution from the accretion event of the template satellite with mass $M_{\mathrm{sat},i}$ and quenching time $t_{100,j}$, and tries to recover those coefficients by maximize a loss that comperes the observed CARDs with the combination of the templates. An example of template constructed from dwarf galaxies is presented in Fig. 2.2. The template

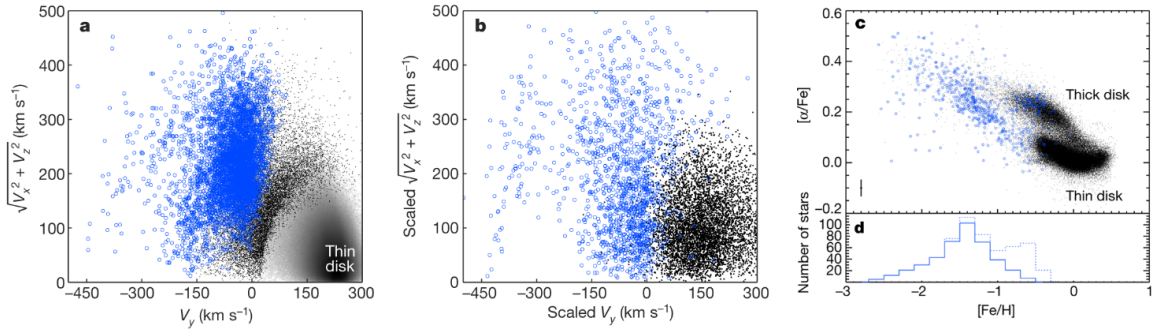

Figure 2.1: Left panel: Toomre diagram for Gaia DR2 data where in blue are stars selected to pick out the GSE structure. Middle panel: similar to the left panel but with simulated data of a minor merger, resulting in a less concentrated structure, mostly due to the fact that a more massive structure is more able to retain its original orbital properties. Right panel: the star in blue are the same of the left panel crossmateched with the APOGEE chemical information. [The figure is from [10]].

that were used belong to the catalog of star particles in the FIRE simulations belonging to dwarf galaxies, stellar streams and phase-mixed debris constructed in [17]. This method and CASBI share two more aspect: 1. Both of these methods are meant to be used on simulations, and the integrations of observational data is not yet implemented, even though theoretically possible. 2. Both rely on the assumption that the chemical space of accreted and isolated dwarf galaxies is very similar, due to ram pressure quenching the star formation history of the accreted object and hence 'freezing' these abundance ratios at the infall time.

Another approach is presented in [8], which takes advantage of the mass-metallicity relation to decompose the metallicity distribution functions (MDF) of the host galaxy as a mixture of accreted halo's MDF, assumed gaussian for each of these building blocks. This decomposition rely on [13] that demonstrated that at the dwarf mass scale, not only the average metallicity vary with the mass, but the width of the MDF also varies, with the lowest mass dwarf having a wider spread of metallicities. The Likelihood that is used in this work for the [Fe/H] distribution, indicated as $\mathbf{z}$ is then:

$$P(z|N, L_i, \mu_i, \sigma_i) = \frac{1}{\sum L_i} \sum_{i=1}^{N} L_i \mathcal{N}(z|\mu_i, \sigma_i), \tag{2.2}$$

where they have assumed that the number of stars in the sample scales linearly with galaxy luminosity L_i, that the μ_i follows a mass-metallicity relation and that the σ_i depend on the galaxy luminosity as described in [13]. Similarly to CASBI, this method has the problem of having a variable number of parameters, making it difficult to sample in practice, so to tackle this problem they decided to bin the luminosities L_j and count the number of contribution from each bin N_j. In order to perform the inference, they adopted a nested sampling scheme to obtain a posterior distribution for the number of galaxies in each luminosity bin, which can be considered a proxy for the star mass. The posterior probability for the Milky Way halo is reported in Fig. 2.3. The samples used in this posterior were obtained from different spectroscopic surveys after applying various cut to avoid contamination from thick disc stars. The cut were made based on parallax distance, radial distance, height with respect to the plane of the galaxy, and only stars with a retrograde orbit $v_\phi < 50$ were selected.

In CASBI we adopt the same superimposition of the components contribution, but we do not

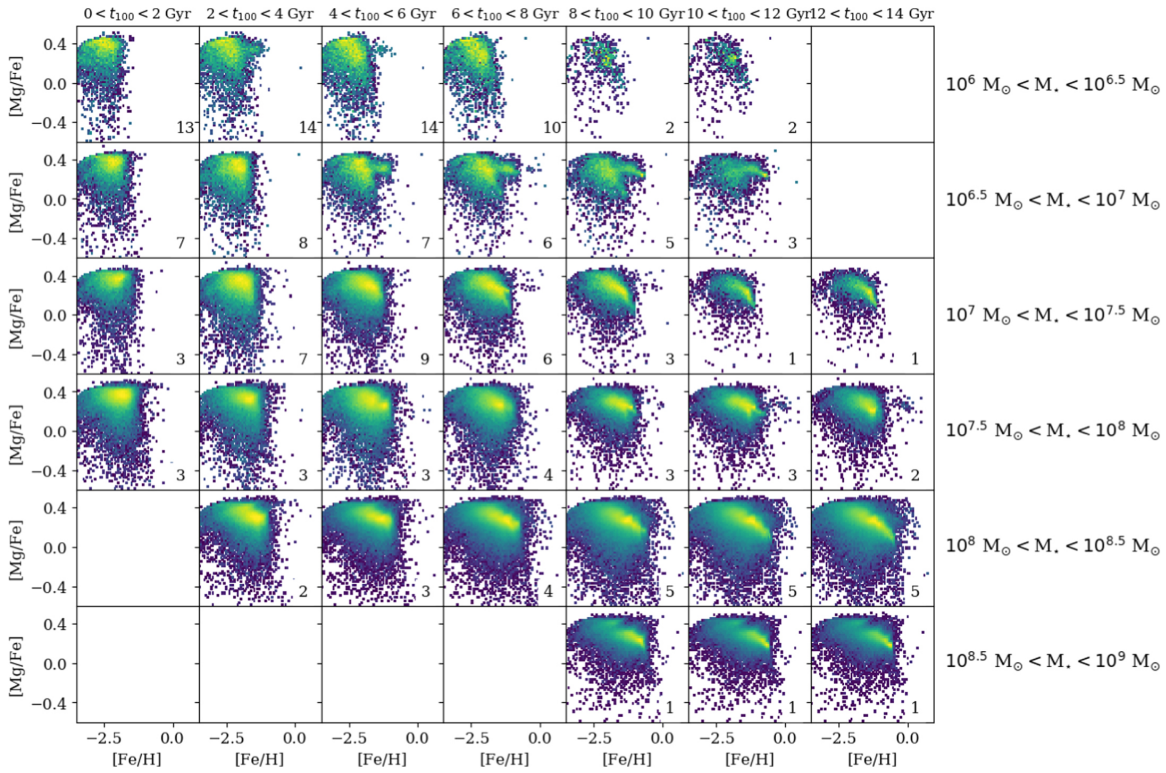

Figure 2.2: Template for accretion events constructed using dwarf galaxy in [5]. More massive dwarf galaxies have CARDs that extend to higher metallicities. At fixed stellar mass, galaxies that assemble more quickly (lower t_{100}) have more density at higher [Mg/Fe] than the component with a more extended star formation history.

assume neither a prefix or an analytical form for the joint distribution of the chemical abundances, relaxing these assumption and relying only on the available samples from the N-body simulations. Also a possible limitation of our method is that we do not take into consideration a distinction between destroyed and surviving accreted dwarf galaxy. This can lead, as described in [16] to a -0.3 dex offset of the mass metallicity [13] relation that might be necessary to take into account.

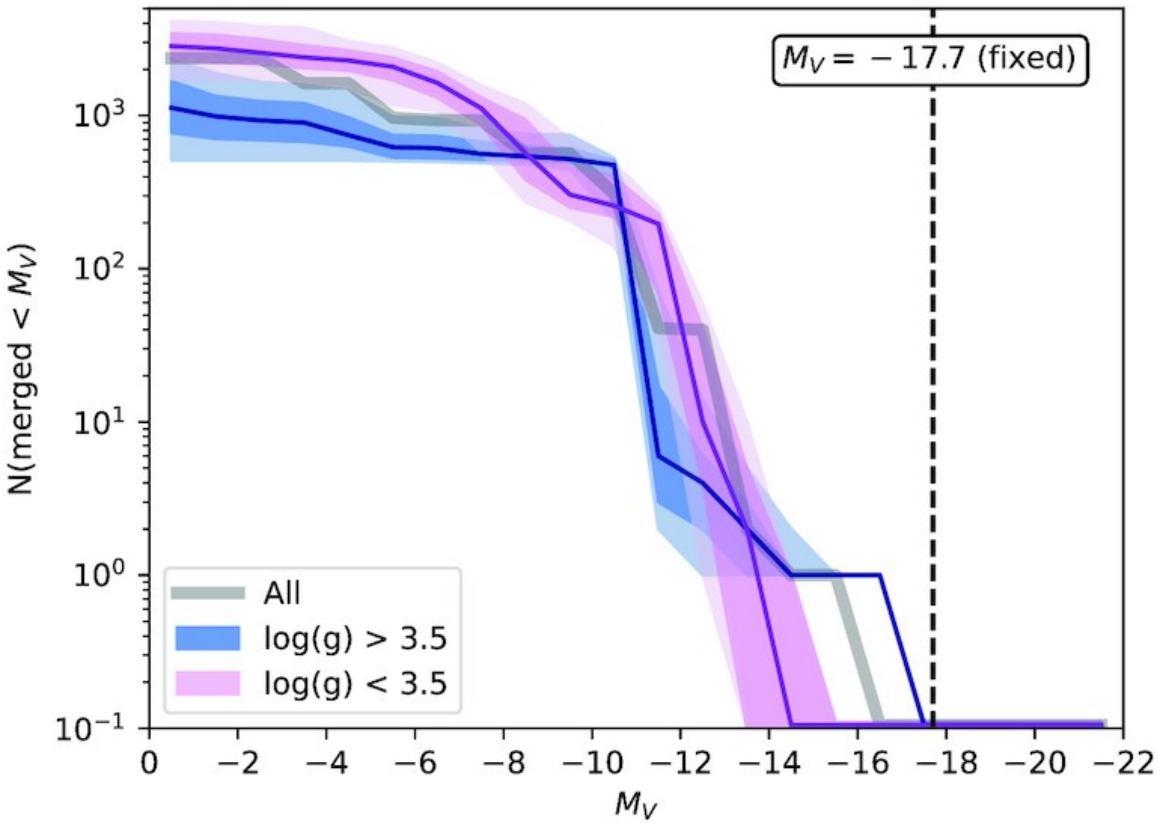

Figure 2.3: The estimated number of destroyed dwarf galaxy in the MW halo. The separation in two bin value of log(g) because dwarf stars and giants can have different metallicity biases.

Chapter 3

CASBI: Chemical Abundance Simulation Based Inference

3.1 Simulation Based Inference

CASBI is a Simulation Based Inference (SBI) package to recover the properties of building blocks of Milky Way like galaxy's halo from observations of the chemical abundance plane. The SBI framework has existed along side the more traditional likelihood based inference methods for quite some years already, and has its root in the Approximate Bayes Computation [20], and it has been used in a variety of fields, from cosmology to particle physics. The main difference between SBI and likelihood based methods, like MCMC, is that the former do not require the likelihood function to be known, but rather rely on a simulator to generate synthetic data $\mathbf{x}$ once the input parameters $\boldsymbol{\theta}$ are passed to it, and the inference pipeline is trained based on data-parameters pairs $(\mathbf{x}, \boldsymbol{\theta})$.

Recent advance of this technique was made possible by the use of machine learning models to emulate conditional probability distributions, a technique know as Neural Density Estimation (NDE) [18]. The NDE is achieved by training a Normalizing Flow architecture, a generative model that allows to obtain samples from a complex distribution $p(x)$ by constructing a series of **bijiective** transformations $f^i_{\phi_i}$ that map x to a latent space z that is distributed as a simple distribution, like a Gaussian. Accordingly to [12], implementing the transformations as Neural Network with parameters ϕ_i, in the end the models learns the following schema:

$$p(x) \sim x \equiv h_0 \overset{f^1_{\phi_1}}{\longleftrightarrow} h_1 \overset{f^2_{\phi_2}}{\longleftrightarrow} h_2 \dots \overset{f^K_{\phi_K}}{\longleftrightarrow} h_K \equiv z \sim \mathcal{N}(z; 0, \mathcal{I}), \tag{3.1}$$

by maximizing the negative log likelihood as loss function and using the change of variable formula as follows:

$$\begin{aligned} \log p(x) &= \log p(z) + \log \left| \det \left(\frac{\partial z}{\partial x} \right) \right| \\ &= \log p(z) + \sum_{i=1}^{K} \log \left| \det \left(\frac{\partial h_i}{\partial h_{i-1}} \right) \right| \\ &= \log p(z) + \sum_{i=1}^{K} \log \left| \det \left(\frac{\partial f^i_{\phi_i}(h_{i-1})}{\partial h_{i-1}} \right) \right|, \end{aligned} \tag{3.2}$$

where the last term is the sum of the log determinant of the Jacobian of the transformations $f^i_{\phi_i}$. Once the model is trained it is easy to sample from the distribution $p(x)$ by sampling from the latent space z and applying the inverse transformations $(f^1_{\phi_1})^{-1} \circ \cdots \circ (f^K_{\phi_K})^{-1}$. In order to keep the sum of log determinant tractable, the use of *Coupling layers* allows to split the input x along its dimensions and apply a transformation only to a subset of the dimensions, using the other as input for the transformation and keeping it fixed. The subset is than changed at each layers, allowing to have a permutation invariant transformation. The transformations $f^i_{\phi_i}$ are usually very simple invertible transformation like a translation and a scaling, or splines functions.

Following the discussion presented in [11], in Bayesian analysis we have the choice to approximate either the Posterior, the Likelihood or the Likelihood ratio, and this choice depend mostly on the problem that one wants to solve.

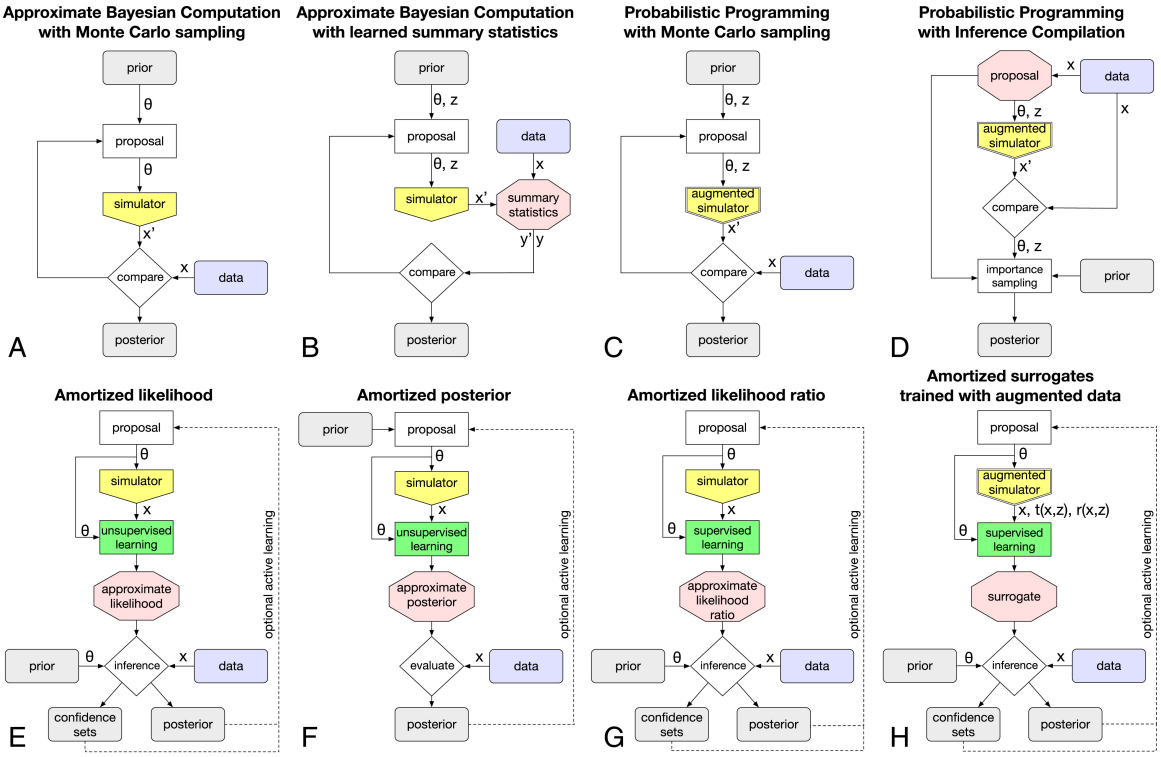

Figure 3.1: Different approaches to Simulation Based Inference, from [1].

In our case, due to the complexity of the Likelihood distribution of the chemical abundance space, we choose to approximate the Posterior distributions, and so we adopted the Neural Posterior Estimate (method **F** in Figure 3.1) that can be trained using the negative loglikelihood as loss function:

$$\begin{aligned} \mathcal{L}_{NPE}(\boldsymbol{\theta}) &= -\mathbb{E}_{\mathcal{D}_{train}} \log \hat{\mathcal{P}}(\theta_i | x_i) \\ &= -\mathbb{E}_{\mathcal{D}_{train}} \log \left(\frac{p(\theta)}{\tilde{p}(\theta)} q_\omega(\theta_i, x_i) \right), \end{aligned} \tag{3.3}$$

where our Posterior distribution $\hat{\mathcal{P}}(\theta_i|x_i)$ is approximated by the product of the ratio of the prior $p(\theta)$ and proposal distribution $\tilde{p}(\theta)$ and the neural conditional distribution $q_\omega(\theta_i, x_i)$, parametrized

by the parameters ω.

Many excellent framework for handling SBI analysis are already available, and CASBI is build on top of the `ltu-ili` python package [11]. In particular, CASBI analysis were performed relying on the `sbi` backend [22] to train a *Neural Posterior Estimate*[1] of the parameters' posteriors. The preprocessing of the data is described in Section 3.2, the details of the training of the NPE is described in Section 3.3.

3.2 NIHAO

The data-parameters pairs $(\mathbf{x}, \boldsymbol{\theta})$ are obtained from the **NIHAO** project [24]. The **NIHAO** (Numerical Investigation of a Hundred Astrophysical Objects) is a set of 100 cosmological zoom-in hydrodynamical simulations with halos that range from dwarf ($M_{star} \sim 5 \times 10^9 M_\odot$) to Milky Way like($M_{star} \sim 2 \times 10^{12} M_\odot$). In order to handle these simulations, in CASBI the preprocessing is done with the use of the functions available in `pynbody` [19]. In Fig. 3.2 we show face on samples of galaxies in the **NIHAO** simulations set.

Similarly to [5] and [8], we rely on the assumption that once the accreted object falls into the gravitational potential of the Milky Way like galaxy its star formation rate is halted, so we can treat each of the snapshot in this simulations as a possible building block of galactic halo. In order to create observables we construct 2D histogram, referred to as $\mathbf{x^i}$, of counts of the chemical abundance plane $[O/Fe]$ over $[Fe/H]$ (α element abundance over metallicity) for each of the snapshot available in **NIHAO**, after filtering in order to have galaxies with a total stellar mass $M_{star} < 6 \times 10^9 M_\odot$, which is the stellar mass of the Large Magellanic Cloud, the largest accreted object by the Milky Way. These 2D histogram have 64×64 pixels, they have minimum and maximum values set after filtering all the stars that were outside the 0.01 percentile in either metallicity or α element abundance, and they are linked to galaxy snapshot trough the `Galaxy_name` attribute. The actual observable $\mathbf{x^j} = \sum_i^{N^j_{sub}} x_i^j$ used in CASBI is then a super imposition of N_{sub} of these 2D histograms, where the N^j_{sub} is the number of accreted objects present in the j-th galaxy halo. In CASBI the inference is done on two set of parameters, the N^j_{sub} and the $\theta^{\mathbf{j}} = (M^j_{star,i}, M^j_{dm,i}, \tau^j_i)$ with $i = 2, ..., N_{sub}$, which are respectively the total stellar mass, the total dark matter mass, and the infall time (in Gyr) of the i-th accreted object in the j-th sample. The inference pipeline is further described in the following section.

3.3 Two step Inference

The objective of the inference is not trivial, because in order to recover the parameters of the building blocks of the Milky Way like galaxy we need to know how many parameters we should look for in order to fix the dimensionality of the priors. This is equivalent to have perfect knowledge on the number of substructure that are present in the galactic halo. In the case of not fully phase mixed structure, the dynamical information could be use to help to disentangle this structure, and also to separate them from the host halo background. In CASBI we do not leverage on this information because it would be hard to construct unbiased couples of observables-parameters, and we leave this task for future work. Instead we separate the inference in two steps:

1. **Inference of the number of substructure**: In this step we train a NPE to recover the posterior distribution of the number of substructure N_{sub}, by using the observable $\mathbf{x^j}$. The

[1] The `sbi` backed implement NPE using `nflows` [9]

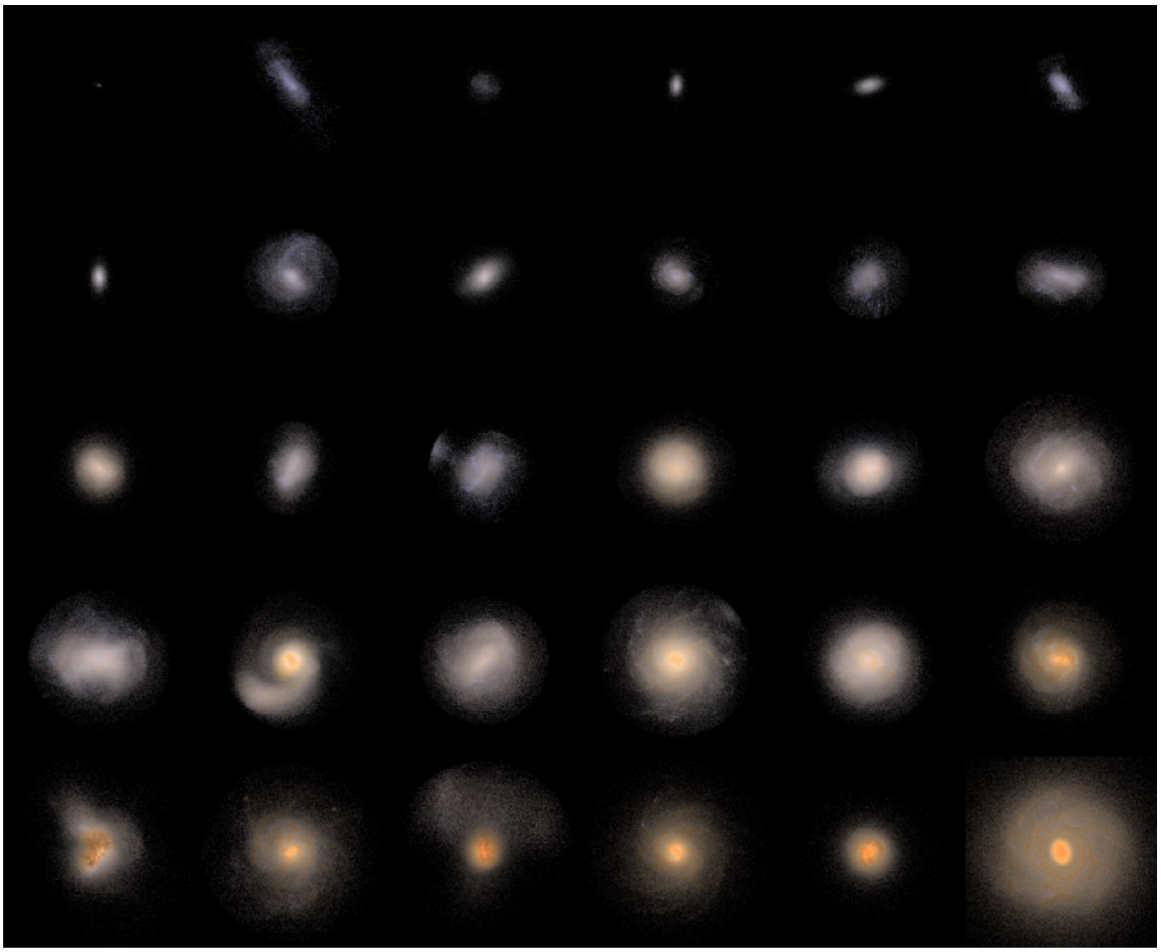

Figure 3.2: Face on **NIHAO** galaxies from [24].

prior for the parameter is assumed to be uniform between 2 and 100. The boundaries were selected in accordance to the order of magnitude of substructures found in [8]. For each of the possible N_{sub} we extract 1000 random samples of N_{sub} snapshots from the **NIHAO** simulations, and we construct the observable $\mathbf{x^j}$ as described in Section 3.2, so in total we have $1000 \times 100 = 10^5$ training samples, with 20 % used as validation, and we use the same process to generate $100 \times 100 = 10^4$ test set samples, making sure that the same combinations of `Galaxy_name` attribute weren't shown in training and test. The training of the NPE is done using the `sbi` backend, using 4 `nsf` (neural spline flow) with 10 layers and 100 neurons each. In order to take full advantage of the image-like structure of the data, we adopt a CNN embedding network to reduce the dimensionality of the input of the NPE from 64×64 to 128. The learning rate was set to ..., the batch size was set to .... The hyperparameter were tuned using the `optuna` package by maximizing the validation log likelihood. In this step we have not impose the constrain that the N_{sub} must be a discrete variable, and we have decide to approximate the sample estimate to an integer. To the knowledge of the author no SBI framework has implemented a way of dealing with the inference of discrete random variables, so

we leave a more precise implementation as a future work. We propose instead another method to obtain the number of substructure, by casting this inference as a classification problem. We use a SkipConnection CNN, considering the number of substructure as the label to assign to each x^j. The performance for both method are reported in Section 4.

2. **Inference of $\theta^{\mathbf{j}}$**: Once we have the estimate $\tilde{N}_{sub}$, whether using dynamical information, or the inference pipeline or the classification method, we can proceed to the inference of the parameters $\theta^{\mathbf{j}}$. The prior for the parameters are assumed to be uniform between the minimum and maximum values available for the galaxies that we have filtered from the NIHAO simulations. We extract 10^5 random samples of N_{sub} snapshots from the **NIHAO** simulations, and we construct the observable $\mathbf{x}^{\mathbf{j}}$ as described in Section 3.2, with 20 % used as validation and the rest as training. We repeat the same process to generate 10^3 test set samples, making sure that the same combinations of `Galaxy_name` attribute weren't shown in training and test to perform calibration of the inference model. The training of the NPE is done using the `sbi` backend, using 4 `nsf` (neural spline flow) with 10 layers and 100 neurons each. Once again we use a CNN as embedding for our observation x^j. The learning rate was set to ..., the batch size was set to .... The hyperparameter were tuned using the `optuna` package by maximizing the validation log likelihood.

The analysis with this two step inference did not satisfy the author, because the inference on the number of subhalos, or the classification, was not accurate enough to be generate a good foundation for the prior distribution of the second step. Moreover the inference pipeline for the parameters $\theta^{\mathbf{j}}$ did not converge using the fast `direct sampling`. This could happen because the Normalizing flow architecture do not constrain the support of the posterior, and the inference could get stack because the sampling happens in region outside the boundaries of the prior distribution. In order to overcome this limitation one might decide to switch to a traditional MCMC sampler, but due to the time consuming nature of this method we decided to rethink the inference pipeline, presented in the next subsection.

3.4 Realistic halo and 1 step Inference

In order to avoid the need for a two step inference and still retaining the possibility to access to the information on how many subhalos populate a given abundance plane, we have decided to condition the sbi model to retrieve the j-most massive subhalo. In this way the NPE is trained on $(x, \theta) = ((x_{j\in J},\, j), \theta_{j\in J})$ pairs, where J represent the J-th original mock galaxy halo, and j is the j-most massive subhalo of the J galaxy. In this way the J galaxy abundance plane is shown as many times as the number of subhalos present in it, but the embedding is conditioned to the j value by concatenating it to each input of the fully connected layer of the CNN used to embed the observations, and it is concatenate also before passing the embedded information to the normalizing flow.

In order to be more realistic in the creation of a mock galaxy halo we have decided to adopt a sampling scheme for the subhalos that is based on the luminosity function described in [14]. The luminosity function described the subhalo distribution in a range of luminosities that spans from $M_V = -2$ all the way to the luminosity of the Large Magellanic cloud:

$$\frac{dN}{dM_V} = 10 \times 10^{0.1(M_V+5)} \tag{3.4}$$

we can then manipulate this expression to the describe it as a function of the Luminosity in solar

luminosity L:

$$\begin{aligned}\frac{dN}{dL} &= \frac{dN}{dM_V} \times \frac{dM_V}{dL} \\ &= 10 \times 10^{0.1(M_V+5)} \times 0.4 \times 10^{0.4(M_V - M_{V,\odot})}\end{aligned} \tag{3.5}$$

which in the end can be integrated to obtain the cumulative distribution function $N(L)$ that we are going to use:

$$N(< L) = K \times L^{\alpha}, \tag{3.6}$$

where K represent a constant and $\alpha = -1.25$ is the single power law exponent obtained by [14]. Other work based not only on SDSS observations like [14] but also on ΛCDM N-body simulation set $\alpha = -1.9 \pm 0.2$ ([23]). We fix this value to -1.25 and we leave the analysis of the impact of this choice as a future work. After obtaining the relation 3.6 and assuming $L_{\odot} = M_{\odot}$, we normalize this relation after setting the support to be the interval of masses that we have available in our catalogue of NIHAO simulations ($10^5 M_{\odot} < M < 6 \cdot 10^9 M_{\odot}$) and we sample from this distributions using an inverse scheme. After obtaining the analytic samples we take the first and second Nearest Neighbors (NN) that are within a 10% of the analytic sampled mass as subhalo for our mock halos. We have also set a mass budget for our mock halo of $M = 1.4 \pm 0.2 \times 10^9 M_{\odot}$ based on [7], and each time a subhalo is sampled we reduce the total mass budget by the mass of the NN that we have used. During this iterative procedure we make sure to sample non repeated subhalo within the same mock halo and we avoid repetitions of the same combinations of subhalos between mock subhalos both within training and test set and across these two sets.

3.5 Surrogate Simulator

Chapter 4

Analysis

4.1 NIHAO UHD

Chapter 5

Conclusion

5.1 Future work

5.1.1 GRUMPY

5.1.2 Number halos has a free parameter: hierarchical sbi

5.1.3 True test: GAIA

Bibliography

[1] The frontier of simulation-based inference. https://www.pnas.org/doi/10.1073/pnas.1912789117.

[2] V. Belokurov, D. Erkal, N. W. Evans, S. E. Koposov, and A. J. Deason. Co-formation of the disc and the stellar halo. *Monthly Notices of the Royal Astronomical Society*, 478(1):611–619, July 2018.

[3] C. Conroy, A. Bonaca, P. Cargile, B. D. Johnson, N. Caldwell, R. P. Naidu, D. Zaritsky, D. Fabricant, S. Moran, J. Rhee, A. Szentgyorgyi, P. Berlind, M. L. Calkins, S. Kattner, and C. Ly. Mapping the Stellar Halo with the H3 Spectroscopic Survey. *The Astrophysical Journal*, 883(1):107, Sept. 2019.

[4] X.-Q. Cui, Y.-H. Zhao, Y.-Q. Chu, G.-P. Li, Q. Li, L.-P. Zhang, H.-J. Su, Z.-Q. Yao, Y.-N. Wang, X.-Z. Xing, X.-N. Li, Y.-T. Zhu, G. Wang, B.-Z. Gu, A.-L. Luo, X.-Q. Xu, Z.-C. Zhang, G.-R. Liu, H.-T. Zhang, D.-H. Yang, S.-Y. Cao, H.-Y. Chen, J.-J. Chen, K.-X. Chen, Y. Chen, J.-R. Chu, L. Feng, X.-F. Gong, Y.-H. Hou, H.-Z. Hu, N.-S. Hu, Z.-W. Hu, L. Jia, F.-H. Jiang, X. Jiang, Z.-B. Jiang, G. Jin, A.-H. Li, Y. Li, Y.-P. Li, G.-Q. Liu, Z.-G. Liu, W.-Z. Lu, Y.-D. Mao, L. Men, Y.-J. Qi, Z.-X. Qi, H.-M. Shi, Z.-H. Tang, Q.-S. Tao, D.-Q. Wang, D. Wang, G.-M. Wang, H. Wang, J.-N. Wang, J. Wang, J.-L. Wang, J.-P. Wang, L. Wang, S.-Q. Wang, Y. Wang, Y.-F. Wang, L.-Z. Xu, Y. Xu, S.-H. Yang, Y. Yu, H. Yuan, X.-Y. Yuan, C. Zhai, J. Zhang, Y.-X. Zhang, Y. Zhang, M. Zhao, F. Zhou, G.-H. Zhou, J. Zhu, and S.-C. Zou. The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST). *Research in Astronomy and Astrophysics*, 12(9):1197–1242, Sept. 2012.

[5] E. C. Cunningham, R. E. Sanderson, K. V. Johnston, N. Panithanpaisal, M. K. Ness, A. Wetzel, S. R. Loebman, I. Escala, D. Horta, and C.-A. Faucher-Giguère. Reading the CARDs: The Imprint of Accretion History in the Chemical Abundances of the Milky Way's Stellar Halo. *The Astrophysical Journal*, 934(2):172, Aug. 2022.

[6] G. M. De Silva, K. C. Freeman, J. Bland-Hawthorn, S. Martell, E. W. De Boer, M. Asplund, S. Keller, S. Sharma, D. B. Zucker, T. Zwitter, B. Anguiano, C. Bacigalupo, D. Bayliss, M. A. Beavis, M. Bergemann, S. Campbell, R. Cannon, D. Carollo, L. Casagrande, A. R. Casey, G. Da Costa, V. D'Orazi, A. Dotter, L. Duong, A. Heger, M. J. Ireland, P. R. Kafle, J. Kos, J. Lattanzio, G. F. Lewis, J. Lin, K. Lind, U. Munari, D. M. Nataf, S. O'Toole, Q. Parker, W. Reid, K. J. Schlesinger, A. Sheinis, J. D. Simpson, D. Stello, Y.-S. Ting, G. Traven, F. Watson, R. Wittenmyer, D. Yong, and M. Žerjal. The GALAH survey: Scientific motivation. *Monthly Notices of the Royal Astronomical Society*, 449(3):2604–2617, May 2015.

[7] A. J. Deason, V. Belokurov, and J. L. Sanders. The total stellar halo mass of the Milky Way. *Monthly Notices of the Royal Astronomical Society*, 490(3):3426–3439, Dec. 2019.

[8] A. J. Deason, S. E. Koposov, A. Fattahi, and R. J. J. Grand. Unravelling the mass spectrum of destroyed dwarf galaxies with the metallicity distribution function. *Monthly Notices of the Royal Astronomical Society*, 520(4):6091–6103, Feb. 2023.

[9] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios. Nflows: Normalizing flows in PyTorch. Zenodo, Nov. 2020.

[10] A. Helmi, C. Babusiaux, H. H. Koppelman, D. Massari, J. Veljanoski, and A. G. A. Brown. The merger that led to the formation of the Milky Way's inner stellar halo and thick disk. *Nature*, 563(7729):85–88, Nov. 2018.

[11] M. Ho, D. J. Bartlett, N. Chartier, C. Cuesta-Lazaro, S. Ding, A. Lapel, P. Lemos, C. C. Lovell, T. L. Makinen, C. Modi, V. Pandya, S. Pandey, L. A. Perez, B. Wandelt, and G. L. Bryan. LtU-ILI: An All-in-One Framework for Implicit Inference in Astrophysics and Cosmology, Feb. 2024.

[12] D. P. Kingma and P. Dhariwal. Glow: Generative Flow with Invertible 1x1 Convolutions, July 2018.

[13] E. N. Kirby, G. A. Lanfranchi, J. D. Simon, J. G. Cohen, and P. Guhathakurta. MULTI-ELEMENT ABUNDANCE MEASUREMENTS FROM MEDIUM-RESOLUTION SPECTRA. III. METALLICITY DISTRIBUTIONS OF MILKY WAY DWARF SATELLITE GALAXIES. *The Astrophysical Journal*, 727(2):78, Feb. 2011.

[14] S. Koposov, V. Belokurov, N. W. Evans, P. C. Hewett, M. J. Irwin, G. Gilmore, D. B. Zucker, H.-W. Rix, M. Fellhauer, E. F. Bell, and E. V. Glushkova. The Luminosity Function of the Milky Way Satellites. *The Astrophysical Journal*, 686(1):279–291, Oct. 2008.

[15] S. R. Majewski, R. P. Schiavon, P. M. Frinchaboy, C. A. Prieto, R. Barkhouser, D. Bizyaev, B. Blank, S. Brunner, A. Burton, R. Carrera, S. D. Chojnowski, K. Cunha, C. Epstein, G. Fitzgerald, A. E. G. Pérez, F. R. Hearty, C. Henderson, J. A. Holtzman, J. A. Johnson, C. R. Lam, J. E. Lawler, P. Maseman, S. Mészáros, M. Nelson, D. C. Nguyen, D. L. Nidever, M. Pinsonneault, M. Shetrone, S. Smee, V. V. Smith, T. Stolberg, M. F. Skrutskie, E. Walker, J. C. Wilson, G. Zasowski, F. Anders, S. Basu, S. Beland, M. R. Blanton, J. Bovy, J. R. Brownstein, J. Carlberg, W. Chaplin, C. Chiappini, D. J. Eisenstein, Y. Elsworth, D. Feuillet, S. W. Fleming, J. Galbraith-Frew, R. A. García, D. A. García-Hernández, B. A. Gillespie, L. Girardi, J. E. Gunn, S. Hasselquist, M. R. Hayden, S. Hekker, I. Ivans, K. Kinemuchi, M. Klaene, S. Mahadevan, S. Mathur, B. Mosser, D. Muna, J. A. Munn, R. C. Nichol, R. W. O'Connell, J. K. Parejko, A. C. Robin, H. Rocha-Pinto, M. Schultheis, A. M. Serenelli, N. Shane, V. S. Aguirre, J. S. Sobeck, B. Thompson, N. W. Troup, D. H. Weinberg, and O. Zamora. The Apache Point Observatory Galactic Evolution Experiment (APOGEE). *The Astronomical Journal*, 154(3):94, Sept. 2017.

[16] R. P. Naidu, C. Conroy, A. Bonaca, B. D. Johnson, Y.-S. Ting, N. Caldwell, D. Zaritsky, and P. A. Cargile. Evidence from the H3 Survey That the Stellar Halo Is Entirely Comprised of Substructure. *The Astrophysical Journal*, 901(1):48, Sept. 2020.

[17] N. Panithanpaisal, R. E. Sanderson, A. Wetzel, E. C. Cunningham, J. Bailin, and C.-A. Faucher-Giguère. The Galaxy Progenitors of Stellar Streams around Milky Way–mass Galaxies in the FIRE Cosmological Simulations. *The Astrophysical Journal*, 920(1):10, Oct. 2021.

[18] G. Papamakarios. Neural Density Estimation and Likelihood-free Inference, Oct. 2019.

[19] A. Pontzen, R. Roškar, G. Stinson, and R. Woods. Pynbody: N-Body/SPH analysis for python. *Astrophysics Source Code Library*, page ascl:1305.002, May 2013.

[20] D. B. Rubin. Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician. *The Annals of Statistics*, 12(4):1151–1172, Dec. 1984.

[21] M. Steinmetz, T. Zwitter, A. Siebert, F. G. Watson, K. C. Freeman, U. Munari, R. Campbell, M. Williams, G. M. Seabroke, R. F. G. Wyse, Q. A. Parker, O. Bienaymé, S. Roeser, B. K. Gibson, G. Gilmore, E. K. Grebel, A. Helmi, J. F. Navarro, D. Burton, C. J. P. Cass, J. A. Dawe, K. Fiegert, M. Hartley, K. S. Russell, W. Saunders, H. Enke, J. Bailin, J. Binney, J. Bland-Hawthorn, C. Boeche, W. Dehnen, D. J. Eisenstein, N. W. Evans, M. Fiorucci, J. P. Fulbright, O. Gerhard, U. Jauregi, A. Kelz, L. Mijović, I. Minchev, G. Parmentier, J. Peñarrubia, A. C. Quillen, M. A. Read, G. Ruchti, R.-D. Scholz, A. Siviero, M. C. Smith, R. Sordo, L. Veltz, S. Vidrih, R. Von Berlepsch, B. J. Boyle, and E. Schilbach. The Radial Velocity Experiment (RAVE): First Data Release. *The Astronomical Journal*, 132(4):1645–1668, Oct. 2006.

[22] A. Tejero-Cantero, J. Boelts, M. Deistler, J.-M. Lueckmann, C. Durkan, P. J. Gonçalves, D. S. Greenberg, and J. H. Macke. Sbi: A toolkit for simulation-based inference. *Journal of Open Source Software*, 5(52):2505, Aug. 2020.

[23] E. J. Tollerud, J. S. Bullock, L. E. Strigari, and B. Willman. Hundreds of Milky Way Satellites? Luminosity Bias in the Satellite Luminosity Function. *The Astrophysical Journal*, 688(1):277–289, Nov. 2008.

[24] L. Wang, A. A. Dutton, G. S. Stinson, A. V. Macciò, C. Penzo, X. Kang, B. W. Keller, and J. Wadsley. NIHAO project I: Reproducing the inefficiency of galaxy formation across cosmic time with a large sample of cosmological hydrodynamical simulations. *Monthly Notices of the Royal Astronomical Society*, 454(1):83–94, Nov. 2015.

[25] A. R. Wetzel, P. F. Hopkins, J.-h. Kim, C.-A. Faucher-Giguère, D. Kereš, and E. Quataert. RECONCILING DWARF GALAXIES WITH ΛCDM COSMOLOGY: SIMULATING A REALISTIC POPULATION OF SATELLITES AROUND A MILKY WAY–MASS GALAXY. *The Astrophysical Journal Letters*, 827(2):L23, Aug. 2016.

[26] B. Yanny, C. Rockosi, H. J. Newberg, G. R. Knapp, J. K. Adelman-McCarthy, B. Alcorn, S. Allam, C. A. Prieto, D. An, K. S. J. Anderson, S. Anderson, C. A. L. Bailer-Jones, S. Bastian, T. C. Beers, E. Bell, V. Belokurov, D. Bizyaev, N. Blythe, J. J. Bochanski, W. N. Boroski, J. Brinchmann, J. Brinkmann, H. Brewington, L. Carey, K. M. Cudworth, M. Evans, N. W. Evans, E. Gates, B. T. Gänsicke, B. Gillespie, G. Gilmore, A. N. Gomez-Moran, E. K. Grebel, J. Greenwell, J. E. Gunn, C. Jordan, W. Jordan, P. Harding, H. Harris, J. S. Hendry, D. Holder, I. I. Ivans, Ž. Ivezič, S. Jester, J. A. Johnson, S. M. Kent, S. Kleinman, A. Kniazev, J. Krzesinski, R. Kron, N. Kuropatkin, S. Lebedeva, Y. S. Lee, R. F. Leger, S. Lépine, S. Levine, H. Lin, D. C. Long, C. Loomis, R. Lupton, O. Malanushenko, V. Malanushenko, B. Margon, D. Martinez-Delgado, P. McGehee, D. Monet, H. L. Morrison, J. A. Munn, E. H. Neilsen, A. Nitta, J. E. Norris, D. Oravetz, R. Owen, N. Padmanabhan, K. Pan, R. S. Peterson, J. R. Pier, J. Platson, P. R. Fiorentin, G. T. Richards, H.-W. Rix, D. J. Schlegel, D. P. Schneider, M. R. Schreiber, A. Schwope, V. Sibley, A. Simmons, S. A. Snedden, J. A. Smith, L. Stark, F. Stauffer, M. Steinmetz, C. Stoughton, M. SubbaRao, A. Szalay, P. Szkody, A. R. Thakar, S. Thirupathi, D. Tucker, A. Uomoto, D. V. Berk, S. Vidrih, Y. Wadadekar, S. Watters, R. Wilhelm, R. F. G. Wyse, J. Yarger, and D. Zucker. SEGUE: A SPECTROSCOPIC SURVEY OF 240,000 STARS WITH g = 14-20. *The Astronomical Journal*, 137(5):4377–4399, May 2009.