CASBI – Chemical Abundance Simulation-Based Inference for Galactic Archeology

Giuseppe Viterbo
Interdisciplinary Center
for Scientific Computing
University of Heidelberg
D-69120 Heidelberg
giuseppe.viterbo@iwr.uni-heidelberg.de

Tobias Buck
Interdisciplinary Center
for Scientific Computing
University of Heidelberg
D-69120 Heidelberg
tobias.buck@iwr.uni-heidelberg.de

Abstract

Galaxies evolve hierarchically through merging with lower-mass systems and the remnants of destroyed galaxies are a key indicator of the past assembly history of our Galaxy. However, accurately measuring the properties of the accreted galaxies and hence unraveling the Milky Way's (MW) formation history is a challenging task. Here we introduce CASBI (Chemical Abundance Simulation Based Inference), a novel inference pipeline for Galactic Archeology based on Simulation-based Inference methods. CASBI leverages on the fact that there is a well defined mass-metallicity relation for galaxies and performs inference of key galaxy properties based on multi-dimensional chemical abundances of stars in the stellar halo. Hence, we recast the problem of unraveling the merger history of the MW into a SBI problem to recover the properties of the building blocks (e.g. total stellar mass and infall time) using the multi-dimensional chemical abundances of stars in the stellar halo as observable. With CASBI we are able to recover the full posterior probability of properties of building blocks of Milky Way like galaxies. We highlight CASBI's potential by inferring posteriors for the stellar masses of completely phase mixed dwarf galaxies solely from the 2d-distributions of stellar abundance in the iron vs. oxygen plane and find accurate and precise inference results.

1 Motivation: The reconstruction of the assembly history of the Milky Way

Inferring the assembly history of the Milky Way (MW) 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 wide-field spectroscopic programs such as GALAH [10], H3 [6], APOGEE [20], RAVE [24], SEGUE [27], or LAMOST [8]. While the dynamical times of the accreted objects are on the order of the age of the host galaxy, phase mixing of accreted and in-situ stars will occur and the phase space only retains part of the information on the original infall parameters. Hence, robustly identifying distinct structures is challenging, and disentangling the components in fully phase mixed situations is nearly impossible. On the other hand, stellar chemical abundances are unchanged over the lifetime of a star serving as unique labels to tag stars. Additionally, the chemical abundance space is dependent on the star formation history and the total stellar mass of the galaxies leading to distinct differences in the abundance distribution of different galaxies [e.g. 4]. Very recently, the crossmatch between astrometric data from Gaia with spectroscopic data allowed for the discovery of the "Gaia-Sausage-Enceladus" [GSE, 1, 14], a massive accretion event whose remnant now dominates the observation of the inner stellar halo of our Galaxy. The discovery of this massive structure opened the door to unveil the merging history of our Galaxy. This work aims to exploit the information available in modern cosmological simulations to guide an automatic and data driven approach to

reconstruct the properties of these objects with the use of the Simulation Based Inference [SBI, 7]] technique.

2   Related Work: Galactic Archaeology with Simulations

In order to characterize the assembly history of the Milky Way (MW) from stellar chemical abundances alone [9] proposed to use the "CARDs", the chemical abundance ratio distributions of the stars, obtained from a subsample of accreted objects from the FIRE-2 zoom-in cosmological simulations of MW-mass galaxies [26]. This method models the host halo as a linear combination of 2d templates for the joint iron and oxygen abundance: $\text{CARD}_{\text{halo, model}}(x_d) = \sum_i \sum_j A_{ij}\text{CARD}_{\text{temp},ij}(x_d|M_{\text{sat},i}, t_{100,j})$, treating each coefficient A_{ij} as the fraction of mass contribution from the accretion event of the template satellite with mass $M_{\text{sat},i}$ and quenching time $t_{100,j}$. CARDs tries to recover those coefficients by maximizing a loss that compares the observed abundance distribution with the combination of the templates. Additionally, this method does not recover full posterior for the parameters of the accreted objects but rather point estimates. Another approach is presented in [12], which takes advantage of the mass-metallicity relation to decompose the metallicity distribution function (MDF) of the host galaxy as a mixture of accreted halo's MDFs, assuming Gaussian shape for each of these building blocks. This decomposition relies on the observational evidence that at the dwarf galaxy mass scale, not only the average metallicity varies with the mass, but also the width of the MDF [16] where lower mass dwarfs have a wider spread of metallicities. This method has the fundamental problem of having a variable number of model parameters, making it difficult to sample in practice. To circumvent this, [12] decided to bin the accreted dwarf galaxies in luminosity and count the number of contributions from each luminosity bin adopting a nested sampling scheme to obtain a posterior distribution for the number of galaxies in each luminosity bin.

3   CASBI: Likelihood-free inference of MW's accretion history

Simulation-based inference   The SBI framework has existed along side the more traditional likelihood based inference methods for quite some years already, having its root in Approximate Bayes Computation [23]. The main difference between SBI and traditional sampling methods, like Markov Chain Monte Carlo (MCMC), is that the former do not require the likelihood function to be known, but rather rely on a simulator to generate synthetic data $\mathbf{x}$ from input parameters $\boldsymbol{\theta}$. Inference networks are then trained based on data-parameters pairs $(\mathbf{x}, \boldsymbol{\theta})$. Recent advances of this technique were made possible by the use of Normalizing Flow models to emulate conditional probability distributions, a technique know as Neural Density Estimation (NDE) [21]. Following the discussion presented in [15], in SBI we have the choice to approximate either the Posterior, the Likelihood or the Likelihood ratio, and this choice depends mostly on the problem that one wants to solve, and in particular on the dimensionalities of $\mathbf{x}$ and $\boldsymbol{\theta}$. 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 (NPE). Another strong choice would have been to perform Neural Ratio Estimate (NRE), which enables one to avoid specifying an approximating distribution model and trains a classifier to target the Likelihood ratio. This kind of approach is quite capable of dealing with cases where both the Likelihood and the Posterior exhibit complicated shapes, but for our case the simple Gaussian-like shape of the posterior and the high dimensionality of the observations $\mathbf{x}$ led to our choice of using NPE, leaving NRE for future testing.

Simulator model in CASBI   CASBI is an SBI package to recover the properties of building blocks of MW's stellar halo from observations of the chemical abundance plane. Hence, the data $\mathbf{x}$ are multi-dimensional distributions of chemical abundances. In this proof-of-concept work we restrict ourselves to only use iron and oxygen. The parameters in CASBI ($\boldsymbol{\theta}$) can be anything describing the infalling objects, e.g. infall time, stellar mass before infall, gas mass or dark matter mass, etc.. Again, we restrict ourselves to inferring only stellar mass, M_{star}, and infall time, τ of the infalling dwarf galaxies for this first application, so in the end the single accreted object parameter is a vector $\boldsymbol{\theta_i} = (M_{star}, \tau)$. As the simulator to generate the data-parameters pairs $(\mathbf{x}, \boldsymbol{\theta})$ needed to train the NPE we rely on the NIHAO cosmological hydrodynamical simulations [25, 2, 3, 5, 4]. **NIHAO** 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 MW ($M_{star} \sim 2 \times 10^{12} M_\odot$) mass. Similar to [9] and [12], we rely on the assumption that once an accreted object falls into the gravitational potential of the host galaxy its star formation is halted and its abundance distribution is frozen. This means, accreted dwarf galaxies

will evolve the same as isolated dwarf galaxies prior to infall. Under this assumption, we treat each dwarf galaxy snapshot as a possible configuration shortly before infall into the host galaxy. In order to create a subhalo template observation we construct 2D histograms (64×64 pixels), referred to as $\mathbf{x_i}$, by binning the chemical abundance plane ([O/Fe], [Fe/H])[1] for each of the snapshots available in **NIHAO**. We additionally filter the galaxies to only consider galaxies with a total stellar mass lower than the Large Magellanic Cloud ($M_{star} < 6 \times 10^9 M_\odot$), the largest accreted object in the MW. The set of all possible subhalos is defined as 'Template Library'. The actual stellar halo observable $\mathbf{x^j} = \sum_i^{N_{sub}^j} x_i^j$ used in CASBI is then a superposition of N_{sub} of these 2D histograms, where the N_{sub}^j is the number of accreted objects present in the j-th galaxy halo. The actual choice of how to sample from the 'Template Library' can be adapted and we use a physically informed approach by using a analytic luminosity function, taken from [17] and a total stellar mass budget taken from [11][2]. The choice of the total stellar mass budget put a constraint on the maximum subhalo mass. With this the final goal of CASBI is to recover $\theta_\mathbf{i}^\mathbf{j}$ for each of the subhalos in the galactic halo from the observable $\mathbf{x^j} = \sum_i x_i^j$, and gaining insight on how many subhalos there are.

The CASBI pipeline
Many excellent frameworks for handling SBI analysis are already available, and CASBI is build on top of the `ltu-ili` python package [15]. In Fig. 1 we show the CASBI pipeline. The modularity of the SBI technique is fully integrated, allowing to change all the components of this pipeline. The 'Template Library' can be adapted to any other suite of simulated galaxies (e.g. [22]), the sampling scheme can incorporate different luminosity functions and stellar halo budgets, the NPE and embedding network architecture and hyper-parameters can be modified to allow for higher accuracy and posterior coverage thanks to the `optuna` grid search implementation, and surrogate models[3] (e.g. Free Form Flow FFF [13] surrogates, or semi-analytical galaxy formation models such as GRUMPY [18]) can be implemented to allow for the sequential version of the NPE.
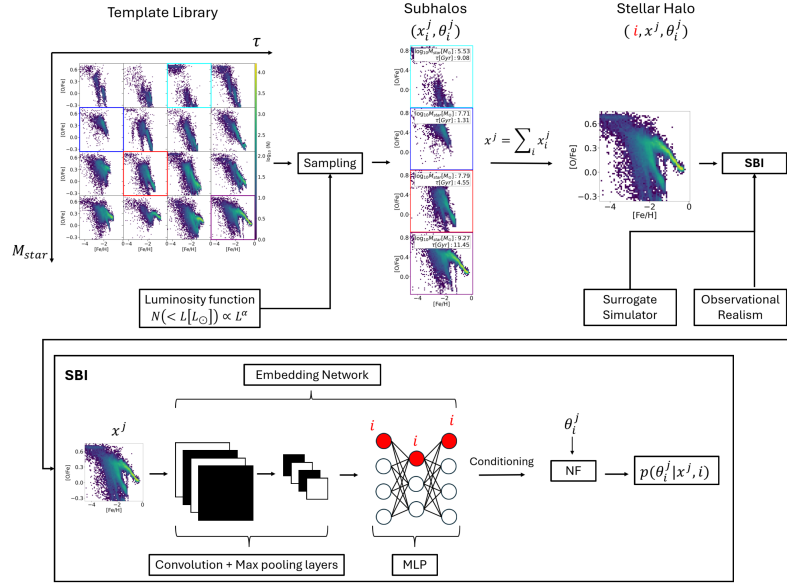

Figure 1: CASBI pipeline. In our analysis the **Template Library** is fixed to the **NIHAO** simulations. The choice of the Template Library incorporates all the assumptions that we make on the chemical enrichment histories of Galaxies, the dynamical effects that accreted objects undergo, and the cosmology, making this part the principal cause of possible misspecification. We **sample** non-repeated subhalos aiming to reproduce a **Luminosity function** $N(< L)$, [17] with fixed total stellar mass budget of the halo [11]. The **SBI** pipeline can incorporate a **Surrogate Simulator** to perform the sequential version of NPE. The **Observational Realism** encapsulates all effects bridging the gap between simulation and observation (e.g. observational uncertainties, selection functions, realistic in-situ stars background, etc.).

Final CASBI model architecture and training data After an extensive hyperparameter tuning and model evaluation phase using `optuna` Multi-objective Optimization, we have found that the 'Neural Spline Flow' (nsf) model with 100 hidden units and 20 transformations available in the `lampe` back-end of `ltu-ili`, achieves both the highest calibration and log posterior. As described in

---

[1]They are respectively proxy for α elements abundance and metallcity

[2]After obtaining the analytic samples we take the first and second Nearest Neighbors (NN) that are within 10% of the sampled mass as our mock subhalo and reduce the total mass budget accordingly.

[3]We leave the integration of the surrogate model for future work

[15], a single NDE tends to be overconfident, so it underestimates uncertainty; in order to deal with this problem one can ensamble multiple models and take a weighted sampling scheme accordingly to their Posterior value. In our case we ensambled four nsf models to obtain our final results. The embedding network is a Convolutional Neural Network (CNN) with 3 convolutional, 3 Max Pooling and 4 fully connected layers with a final embedding dimension of 32 which was obtained during the hyperparameter tuning phase. We have generated 1000 training set stellar halo and 100 test set stellar halo with a maximum of 100 subhalos each. The training took ~ 1 hour on a single NVIDIA A100.

4 Results

In Fig. 2 we present the posterior probability of subhalo masses for a test set stellar halo. In this example we can appreciate that the true value for each of the subhalos falls close to the posterior estimate (i.e. the posterior mean) and that posterior predictions are relatively peaked around the true values. This finding makes us confident that CASBI is able to infer infall properties of completely phase mixed dwarf galaxies solely from chemical abundances with relatively high precision and accuracy. In order to correctly evaluate the constraining power and the calibration
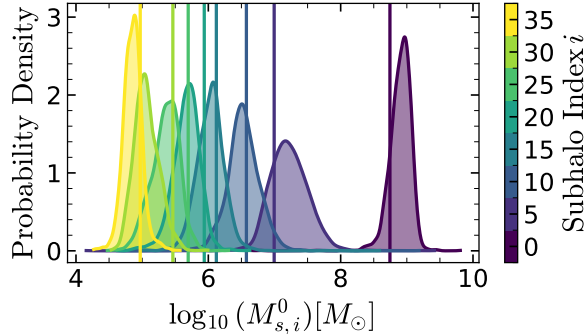

Figure 2: Posterior of the stellar mass M_s for subhalos with index i belonging to the first stellar halo of the test set ($j = 0$). The vertical lines show the true values for each subhalo.
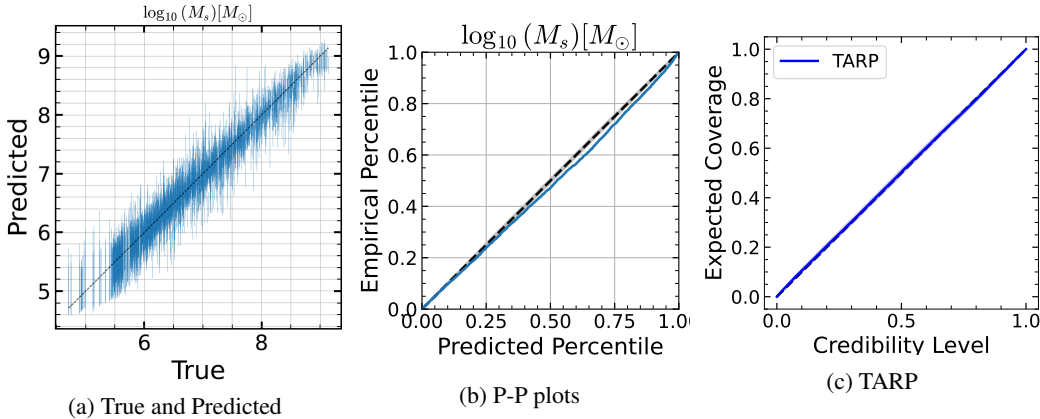
of the model we report in Fig. 3 the comparison between true and predicted values of the parameters in the test simulations, the coverage tests using both the percentile-percentile (P-P) plots and the TARP test [19][4]. Looking at 3a, we find good predictive performance on the stellar mass M_s across the entire mass range probed. Similarly, from Fig. 3b and Fig. 3c we see that the model results appear to be well calibrated both in the marginal and in the full posterior. We have further tested CASBI's ability to infer infall times, τ, from chemical abundances alone but found poor performance for this task. The lack of predictability of the infall times can be associated to the degeneracy of the chemical abundances between a massive system that was accreted early on and a less massive system that was accreted more recently. The integration of the orbital information or using more α elements (e.g. Mg, Si, Mn) could alleviate this limitation of the model. Both of this integration are left as future work.

(a) True and Predicted

(b) P-P plots

(c) TARP

Figure 3: Inference results on the test set. We show the True vs Prediced values for the stellar mass M_s part of the $\boldsymbol{\theta}$, with error bars representing the 68% confidence interval in (a), the marginal posterior coverage for the stellar mass M_s in (b), and full posterior coverage in (c).

---

[4]TARP constructs, in the limit of sufficient samples, an estimate of the untractable joint posterior coverage which is guaranteed to converge to the true posterior coverage. In our case, since the infall time τ component of $\boldsymbol{\theta}$ is not constrained and its posterior resamples the prior distribution, the TARP value is less meaningful than the marginal coverage plot

5   Conclusion and limitations

We find that SBI can be applied to the complex task of parameter estimation for accreted dwarf galaxies in the MW, leveraging only on the chemical information. This allows for the inference of infall properties even in the case of fully phase mixed objects in the Galactic halo and opens up a completely new avenue for Galactic Archeology. We find that the information of multi-dimensional chemical abundance distributions is able to guide the model towards a well calibrated and accurate reconstruction of the stellar mass of accreted dwarf galaxies. Furthermore, our method avoids the need to bin the subhalo information as was done in [12]. In this way with CASBI it is possible to directly perform Bayesian inference on the parameters of each individual subhalo. One limitation of our method is the simplified scenario in which we have tested CASBI. We are asuming to be able to perfectly remove background stars and perform inference on a clean halo sample of stars. Additionally, we have not yet assumed any observational selection function nor observational uncertainties on stellar abundances. But all this will be implemented in the 'Observational Realism' component of CASBI and will be evaluated in future work (see Fig. 1). Our code for CASBI including extensive documentation is available on GitHub as well as the data to reproduce the results of this paper.

Broader impact statement

The authors are not aware of any immediate ethical or societal implications of this work. This work purely aims to aid scientific research and proposes to apply novel SBI techniques to unravel the formation history of the Milky Way.

Acknowledgments and Disclosure of Funding

References

References

[1] 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.

[2] Tobias Buck. On the origin of the chemical bimodality of disk stars: A tale of merger and migration. *Monthly Notices of the Royal Astronomical Society*, 491(4):5435–5446, February 2020.

[3] Tobias Buck, Aura Obreja, Andrea V. Macciò, Ivan Minchev, Aaron A. Dutton, and Jeremiah P. Ostriker. NIHAO-UHD: the properties of MW-like stellar discs in high-resolution cosmological simulations. *Monthly Notices of the Royal Astronomical Society*, 491(3):3461–3478, January 2020.

[4] Tobias Buck, Aura Obreja, Bridget Ratcliffe, Yuxi(Lucy) Lu, Ivan Minchev, and Andrea V. Macciò. The impact of early massive mergers on the chemical evolution of Milky Way-like galaxies: insights from NIHAO-UHD simulations. *Monthly Notices of the Royal Astronomical Society*, 523(1):1565–1576, July 2023.

[5] Tobias Buck, Jan Rybizki, Sven Buder, Aura Obreja, Andrea V. Macciò, Christoph Pfrommer, Matthias Steinmetz, and Melissa Ness. The challenge of simultaneously matching the observed diversity of chemical abundance patterns in cosmological hydrodynamical simulations. *Monthly Notices of the Royal Astronomical Society*, 508(3):3365–3387, December 2021.

[6] Charlie Conroy, Ana Bonaca, Phillip Cargile, Benjamin D. Johnson, Nelson Caldwell, Rohan P. Naidu, Dennis Zaritsky, Daniel Fabricant, Sean Moran, Jaehyon Rhee, Andrew Szentgyorgyi, Perry Berlind, Michael L. Calkins, ShiAnne Kattner, and Chun Ly. Mapping the Stellar Halo with the H3 Spectroscopic Survey. *The Astrophysical Journal*, 883(1):107, September 2019.

[7] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020.

[8] Xiang-Qun Cui, Yong-Heng Zhao, Yao-Quan Chu, Guo-Ping Li, Qi Li, Li-Ping Zhang, Hong-Jun Su, Zheng-Qiu Yao, Ya-Nan Wang, Xiao-Zheng Xing, Xin-Nan Li, Yong-Tian Zhu, Gang Wang, Bo-Zhong Gu, A-Li Luo, Xin-Qi Xu, Zhen-Chao Zhang, Gen-Rong Liu, Hao-Tong Zhang, De-Hua Yang, Shu-Yun

Cao, Hai-Yuan Chen, Jian-Jun Chen, Kun-Xin Chen, Ying Chen, Jia-Ru Chu, Lei Feng, Xue-Fei Gong, Yong-Hui Hou, Hong-Zhuan Hu, Ning-Sheng Hu, Zhong-Wen Hu, Lei Jia, Fang-Hua Jiang, Xiang Jiang, Zi-Bo Jiang, Ge Jin, Ai-Hua Li, Yan Li, Ye-Ping Li, Guan-Qun Liu, Zhi-Gang Liu, Wen-Zhi Lu, Yin-Dun Mao, Li Men, Yong-Jun Qi, Zhao-Xiang Qi, Huo-Ming Shi, Zheng-Hong Tang, Qing-Sheng Tao, Da-Qi Wang, Dan Wang, Guo-Min Wang, Hai Wang, Jia-Ning Wang, Jian Wang, Jian-Ling Wang, Jian-Ping Wang, Lei Wang, Shu-Qing Wang, You Wang, Yue-Fei Wang, Ling-Zhe Xu, Yan Xu, Shi-Hai Yang, Yong Yu, Hui Yuan, Xiang-Yan Yuan, Chao Zhai, Jing Zhang, Yan-Xia Zhang, Yong Zhang, Ming Zhao, Fang Zhou, Guo-Hua Zhou, Jie Zhu, and Si-Cheng Zou. The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST). *Research in Astronomy and Astrophysics*, 12(9):1197–1242, September 2012.

[9] Emily C. Cunningham, Robyn E. Sanderson, Kathryn V. Johnston, Nondh Panithanpaisal, Melissa K. Ness, Andrew Wetzel, Sarah R. Loebman, Ivanna Escala, Danny Horta, and Claude-André 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, August 2022.

[10] G. M. De Silva, K. C. Freeman, J. Bland-Hawthorn, S. Martell, E. Wylie 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.

[11] Alis J. Deason, Vasily Belokurov, and Jason L. Sanders. The total stellar halo mass of the Milky Way. *Monthly Notices of the Royal Astronomical Society*, 490(3):3426–3439, December 2019.

[12] Alis J Deason, Sergey E Koposov, Azadeh Fattahi, and Robert 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, February 2023.

[13] Felix Draxler, Peter Sorrenson, Lea Zimmermann, Armand Rousselot, and Ullrich Köthe. Free-form Flows: Make Any Architecture a Normalizing Flow, April 2024.

[14] Amina Helmi, Carine Babusiaux, Helmer H. Koppelman, Davide Massari, Jovan Veljanoski, and Anthony 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, November 2018.

[15] Matthew Ho, Deaglan J. Bartlett, Nicolas Chartier, Carolina Cuesta-Lazaro, Simon Ding, Axel Lapel, Pablo Lemos, Christopher C. Lovell, T. Lucas Makinen, Chirag Modi, Viraj Pandya, Shivam Pandey, Lucia A. Perez, Benjamin Wandelt, and Greg L. Bryan. LtU-ILI: An All-in-One Framework for Implicit Inference in Astrophysics and Cosmology, February 2024.

[16] Evan N. Kirby, Gustavo A. Lanfranchi, Joshua D. Simon, Judith G. Cohen, and Puragra Guhathakurta. MULTI-ELEMENT ABUNDANCE MEASUREMENTS FROM MEDIUM-RESOLUTION SPECTRA. III. METALLICITY DISTRIBUTIONS OF MILKY WAY DWARF SATELLITE GALAXIES. *The Astrophysical Journal*, 727(2):78, February 2011.

[17] 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, October 2008.

[18] Andrey Kravtsov and Viraj Manwadkar. GRUMPY: A simple framework for realistic forward-modelling of dwarf galaxies. *Monthly Notices of the Royal Astronomical Society*, 514(2):2667–2691, June 2022.

[19] Pablo Lemos, Adam Coogan, Yashar Hezaveh, and Laurence Perreault-Levasseur. Sampling-Based Accuracy Testing of Posterior Estimators for General Inference.

[20] Steven R. Majewski, Ricardo P. Schiavon, Peter M. Frinchaboy, Carlos Allende Prieto, Robert Barkhouser, Dmitry Bizyaev, Basil Blank, Sophia Brunner, Adam Burton, Ricardo Carrera, S. Drew Chojnowski, Kátia Cunha, Courtney Epstein, Greg Fitzgerald, Ana E. García Pérez, Fred R. Hearty, Chuck Henderson, Jon A. Holtzman, Jennifer A. Johnson, Charles R. Lam, James E. Lawler, Paul Maseman, Szabolcs Mészáros, Matthew Nelson, Duy Coung Nguyen, David L. Nidever, Marc Pinsonneault, Matthew Shetrone, Stephen Smee, Verne V. Smith, Todd Stolberg, Michael F. Skrutskie, Eric Walker, John C. Wilson, Gail Zasowski, Friedrich Anders, Sarbani Basu, Stephane Beland, Michael R. Blanton, Jo Bovy, Joel R. Brownstein, Joleen Carlberg, William Chaplin, Cristina Chiappini, Daniel J. Eisenstein, Yvonne Elsworth, Diane Feuillet, Scott W. Fleming, Jessica Galbraith-Frew, Rafael A. García, D. Aníbal García-Hernández, Bruce A. Gillespie, Léo Girardi, James E. Gunn, Sten Hasselquist, Michael R. Hayden, Saskia Hekker, Inese

Ivans, Karen Kinemuchi, Mark Klaene, Suvrath Mahadevan, Savita Mathur, Benoît Mosser, Demitri Muna, Jeffrey A. Munn, Robert C. Nichol, Robert W. O'Connell, John K. Parejko, A. C. Robin, Helio Rocha-Pinto, Matthias Schultheis, Aldo M. Serenelli, Neville Shane, Victor Silva Aguirre, Jennifer S. Sobeck, Benjamin Thompson, Nicholas W. Troup, David H. Weinberg, and Olga Zamora. The Apache Point Observatory Galactic Evolution Experiment (APOGEE). *The Astronomical Journal*, 154(3):94, September 2017.

[21] George Papamakarios. Neural Density Estimation and Likelihood-free Inference, October 2019.

[22] Annalisa Pillepich, Diego Sotillo-Ramos, Rahul Ramesh, Dylan Nelson, Christoph Engler, Vicente Rodriguez-Gomez, Martin Fournier, Martina Donnari, Volker Springel, and Lars Hernquist. Milky Way and Andromeda analogs from the TNG50 simulation, March 2023.

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

[24] 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, October 2006.

[25] Liang Wang, Aaron A. Dutton, Gregory S. Stinson, Andrea V. Macciò, Camilla Penzo, Xi Kang, Ben W. Keller, and James 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, November 2015.

[26] Andrew R. Wetzel, Philip F. Hopkins, Ji-hoon Kim, Claude-André Faucher-Giguère, Dušan Kereš, and Eliot 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, August 2016.

[27] Brian Yanny, Constance Rockosi, Heidi Jo Newberg, Gillian R. Knapp, Jennifer K. Adelman-McCarthy, Bonnie Alcorn, Sahar Allam, Carlos Allende Prieto, Deokkeun An, Kurt S. J. Anderson, Scott Anderson, Coryn A. L. Bailer-Jones, Steve Bastian, Timothy C. Beers, Eric Bell, Vasily Belokurov, Dmitry Bizyaev, Norm Blythe, John J. Bochanski, William N. Boroski, Jarle Brinchmann, J. Brinkmann, Howard Brewington, Larry Carey, Kyle M. Cudworth, Michael Evans, N. W. Evans, Evalyn Gates, B. T. Gänsicke, Bruce Gillespie, Gerald Gilmore, Ada Nebot Gomez-Moran, Eva K. Grebel, Jim Greenwell, James E. Gunn, Cathy Jordan, Wendell Jordan, Paul Harding, Hugh Harris, John S. Hendry, Diana Holder, Inese I. Ivans, Željko Ivezič, Sebastian Jester, Jennifer A. Johnson, Stephen M. Kent, Scot Kleinman, Alexei Kniazev, Jurek Krzesinski, Richard Kron, Nikolay Kuropatkin, Svetlana Lebedeva, Young Sun Lee, R. French Leger, Sébastien Lépine, Steve Levine, Huan Lin, Daniel C. Long, Craig Loomis, Robert Lupton, Olena Malanushenko, Viktor Malanushenko, Bruce Margon, David Martinez-Delgado, Peregrine McGehee, Dave Monet, Heather L. Morrison, Jeffrey A. Munn, Eric H. Neilsen, Atsuko Nitta, John E. Norris, Dan Oravetz, Russell Owen, Nikhil Padmanabhan, Kaike Pan, R. S. Peterson, Jeffrey R. Pier, Jared Platson, Paola Re Fiorentin, Gordon T. Richards, Hans-Walter Rix, David J. Schlegel, Donald P. Schneider, Matthias R. Schreiber, Axel Schwope, Valena Sibley, Audrey Simmons, Stephanie A. Snedden, J. Allyn Smith, Larry Stark, Fritz Stauffer, M. Steinmetz, C. Stoughton, Mark SubbaRao, Alex Szalay, Paula Szkody, Aniruddha R. Thakar, Sivarani Thirupathi, Douglas Tucker, Alan Uomoto, Dan Vanden Berk, Simon Vidrih, Yogesh Wadadekar, Shannon Watters, Ron Wilhelm, Rosemary F. G. Wyse, Jean Yarger, and Dan Zucker. SEGUE: A SPECTROSCOPIC SURVEY OF 240,000 STARS WITH $g = 14$-20. *The Astronomical Journal*, 137(5):4377–4399, May 2009.