

Set-based Implicit Likelihood Inference of Galaxy Cluster Mass

Bonny Y. Wang^{1 2 3 4} Leander Thiele^{3 4}

Abstract

We present a set-based machine learning framework that infers posterior distributions of galaxy cluster masses from projected galaxy dynamics. Our model combines Deep Sets and conditional normalizing flows to incorporate both positional and velocity information of member galaxies to predict residual corrections to the $M-\sigma$ relation for improved interpretability. Trained on the Uchuu-UniverseMachine simulation, our approach significantly reduces scatter and provides well-calibrated uncertainties across the full mass range compared to traditional dynamical estimates.

1. Introduction

Galaxy clusters are powerful cosmological probes that constrain fundamental parameters of the Universe (Voit, 2005; Allen et al., 2011; Dodelson et al., 2016; Haiman et al., 2001). Given the current S_8 tension, accurately characterizing the cluster mass function is critical for testing and assessing the standard Λ CDM model.

Currently, we use a wide range of observational strategies for cluster mass estimation. These include X-ray measurements of intra-cluster gas (e.g., Mantz et al., 2016; Giles et al., 2017), thermal Sunyaev-Zel’dovich (tSZ) effect, (e.g., Hilton et al., 2021), and richness information (e.g., Yee & Ellingson, 2003). More direct methods include weak gravitational lensing (e.g., Applegate et al., 2014; McClintock et al., 2019) and galaxy kinematics (e.g., Zwicky, 1933; Old et al., 2014; Geller et al., 2013), though these approaches are either limited by low signal-to-noise ratios, projection effects, or fail to fully capture baryonic modeling uncertainties.

¹McWilliams Center for Cosmology and Astrophysics, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA ²Entertainment Technology Center, Carnegie Mellon University, Pittsburgh, PA 15213, USA ³Kavli IPMU (WPI), UTIAS, The University of Tokyo, Japan ⁴Center for Data-Driven Discovery, Kavli IPMU (WPI), Japan. Correspondence to: Bonny Y. Wang <bonnyw@uchicago.edu>.

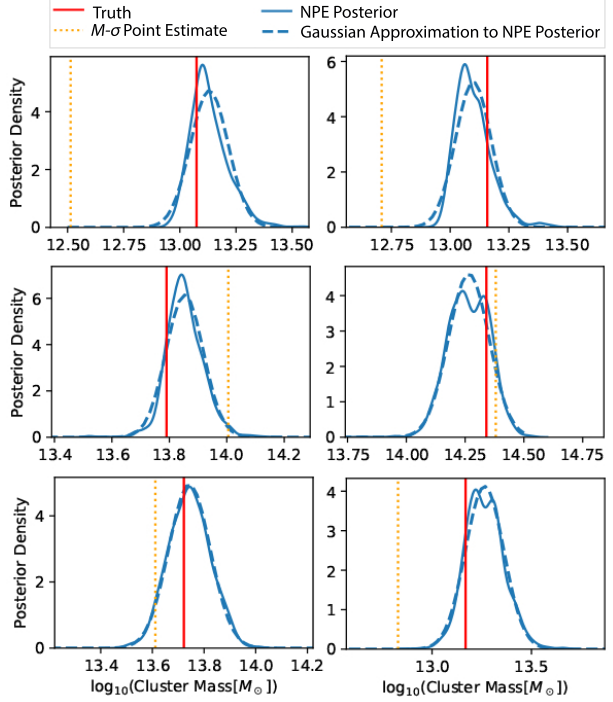


Figure 1. Posterior distributions of galaxy cluster mass estimates. The solid blue lines show posteriors predicted by the Neural Posterior Estimator (NPE), while the blue dashed lines indicate Gaussian approximations to these generated distributions. The red vertical lines mark the true cluster masses, and the orange dotted lines denote baseline $M-\sigma$ mass estimates. The left column presents examples of unimodal distributions, whereas the right column highlights bimodal cases.

The classic $M-\sigma$ relation provides a simple dynamical mass approximation based on the velocity dispersion of member galaxies. However, it relies on idealized assumptions that clusters are perfectly spherical, composed of equal-mass galaxies, and in a state of virial equilibrium. In reality, galaxy clusters are more dynamically complex and often far from equilibrium. To overcome these limitations, recent work has incorporated deep learning approaches (e.g., Ntampaka et al., 2016; 2019; Ho et al., 2019; 2021; 2022; 2023; Yan et al., 2020; Kodi Ramanah et al., 2021; Gupta & Reichardt, 2020; 2021; Garuda et al., 2024) to model nonlinear patterns and extract information from high-dimensional

galaxy dynamics. These previous methods often omit spatial and morphological information, rely mostly on point estimates, or compress the data.

Here, we employ Deep Sets architecture that incorporates the full projected phase-space information of member galaxies, including both their velocities and positions, alongside global cluster morphology. This architecture is well suited for astronomical problems involving unordered collections with varying numbers of objects (see e.g., Thiele et al., 2022; Wang et al., 2023; Jung et al., 2023; de Santi et al., 2023).

Our method fits within the broader context of simulation-based inference (SBI) (Cranmer et al., 2020). We specifically build on recent advances in neural posterior estimation (NPE) (Greenberg et al., 2019; Lueckmann et al., 2017; Papamakarios & Murray, 2016) by using neural spline flows (Durkan et al., 2019; Dolatabadi et al., 2020). This enables accurate uncertainty quantification by producing expressive posterior distributions for galaxy cluster masses. As shown in Figure 1, the predicted posteriors flexibly capture asymmetries, multimodality, and varying uncertainty structure, while also generating the true cluster mass more accurately than the $M\text{--}\sigma$ estimate.

Furthermore, we incorporate physical knowledge into the architecture by predicting residual corrections to the standard $M\text{--}\sigma$ relation, rather than directly predicting the cluster mass. This design enhances the interpretability of our model, allowing us to explicitly isolate information beyond what is captured by equilibrium-based assumptions.

2. Method

Here, we describe our dataset, model architecture and selected features to infer posterior distributions of galaxy cluster masses from member galaxy dynamics. The full implementation is available at our GitHub repository¹.

2.1. Dataset

To model realistic galaxy cluster dynamics, we use the Uchuu-UniverseMachine dataset (Aung et al., 2023), a large-volume $(2 \text{ Gpc}/h)^3$ N -body simulation with $12,800^3$ particles. We analyze galaxy clusters from the Rockstar finder (Behroozi et al., 2013), and galaxy properties from UniverseMachine (Behroozi et al., 2019). This provides robust and realistic representations of both galaxies and galaxy clusters. The baseline $M\text{--}\sigma$ relation for galaxy clusters in this dataset is

$$\log_{10}(\sigma_v) = 0.33 \log_{10}(M_{\text{vir}}) - 1.95 + \gamma, \quad (1)$$

where M_{vir} is the virial mass, σ_v is the member galaxy velocity dispersion, and γ is a random variable drawn from

¹<https://github.com/BonnyWang/Cluster-Mass-DeepSet-NF>

a distribution that depends on the mass range and other properties of the galaxy cluster. We split the data into 70% for training, 15% for validation, and 15% for testing.

2.2. Model Architecture

To probe whether galaxy dynamics contain additional information beyond the standard $M\text{--}\sigma$ relation, we train a machine learning model to predict the residual between the true cluster mass and a baseline $M\text{--}\sigma$ estimate. The corrected mass prediction is:

$$\log_{10} M_{\text{vir}}^{\text{pred}} = \log_{10} M_{\text{vir}}^{M\text{--}\sigma} + \Delta_{\text{ML}}, \quad (2)$$

where Δ_{ML} is a learned correction modeled probabilistically. Our architecture consists of two components: a stack of three Deep Sets blocks for feature extraction and a conditional normalizing flow for posterior generation.

Each Deep Sets block here consists of a pointwise encoder ρ and a context updater ϕ , both modeled as multilayer perceptrons (MLPs), along with a permutation-invariant aggregation $\oplus$. Notably, we use mean pooling to avoid incorporating information related to galaxy richness. Given input galaxy features $\vec{p}_i$ for member galaxies in a cluster, along with an initial global cluster-level feature vector $\vec{u}^{(0)}$, the architecture performs the following operations:

$$\vec{h}_i^{(k)} = \rho^{(k)}([\vec{p}_i, \vec{u}^{(k-1)}]), \quad \vec{u}^{(k)} = \phi^{(k)}\left(\bigoplus \vec{h}_i^{(k)}\right), \quad (3)$$

where $\vec{h}_i^{(k)}$ is the latent feature of the i -th galaxy after the k -th encoder, and $\vec{u}^{(k)}$ is the global context vector summarizing the cluster state at block k .

As we stack three Deep Sets blocks, the final output $\vec{u}^{(3)}$ serves as a latent context vector summarizing the internal galaxy dynamics of each cluster. This vector conditions a normalizing flow that models the residual Δ_{ML} as a flexible, non-Gaussian distribution. We implement this using the conditional spline transformation provided by PYRO (Bingham et al., 2018). Finally, all components are jointly optimized by maximizing the conditional log-likelihood of the posterior:

$$\mathcal{L} = -\log p(\Delta_{\text{ML}} | \vec{u}^{(3)}). \quad (4)$$

The progression of training and validation loss values is shown in Appendix A.

In addition, to have an equal number of samples across all mass ranges, we augmented the dataset to enforce a uniform mass distribution across training samples.

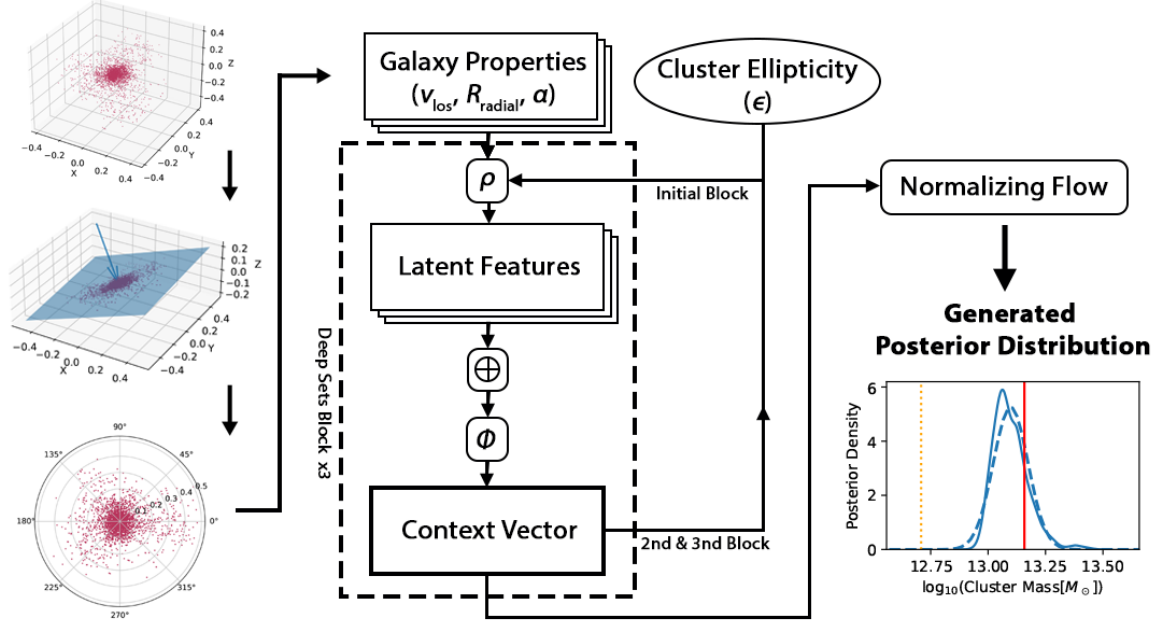


Figure 2. Schematic overview of our model for probabilistic cluster mass estimation. Galaxy member properties—line-of-sight velocity v_{los} , projected radial distance R_{radial} , and angular position α —are obtained by projecting galaxies onto randomly selected planes. These features are processed through three Deep Sets blocks to produce permutation-invariant latent representations. The initial block incorporates a global cluster-level feature, ellipticity ϵ , which is updated and passed through subsequent blocks. The final latent vector serves as context for a normalizing flow, which outputs a full posterior distribution over cluster mass.

2.3. Observational Feature

To mimic the observational process, we construct our dataset by first selecting a random line of sight and projecting member galaxies onto a plane perpendicular to it. To preserve the cluster’s overall morphology while ensuring rotational invariance, we fit an ellipse to the projected distribution and take its semi-major axis as a reference. Based on this setup, we select the following features as inputs to our machine learning model: Line-of-sight velocity (v_{los}), radial distance to the halo center (R_{radial}), angle relative to the fitted ellipse axis (α). In addition, we include a global feature to characterize the galaxy distribution: the ellipticity (ϵ), computed from the fitted ellipse’s semi-major and semi-minor axes.

These features serve as inputs to the machine learning framework described in the previous section. The full pipeline, including the construction of observational features and the machine learning architecture, is illustrated in Figure 2.

3. Result

In this section, we present the predictions from our machine learning framework, followed by an analysis of how galaxy cluster properties influence these predictions.

3.1. Machine Learning Correction

We show our main result comparing machine learning predicted and true galaxy cluster masses in the left panel of Figure 3. True masses are on the x-axis; predicted masses are on the y-axis. Orange points indicate traditional M - σ estimates, while violin plots visualize the model’s posterior distributions. The black diagonal marks perfect agreement. Most posteriors are centered closer to this line than M - σ predictions.

The middle panel of Figure 3 illustrates the bias and standard deviation in cluster mass predictions as a function of true mass. The standard deviation of the NPE empirical scatter (blue) closely matches the predicted posterior standard deviation (green) across most mass ranges, indicating well-calibrated uncertainty estimates. Compared to the standard deviation of the M - σ point estimate (orange), the NPE empirical standard deviation exhibits approximately twice the reduced scatter to the truth. This highlights a great improvement in mass accuracy. The model shows little bias across all masses, as evidenced by the near-zero purple line.

Moreover, we assess uncertainty calibration using two methods: the highest posterior density (HPD) and TARP (Tests of Accuracy with Random Points, Lemos et al., 2023). In the right panel of Figure 3, both HPD (orange) and TARP (green) closely follow the black dashed identity line, again,

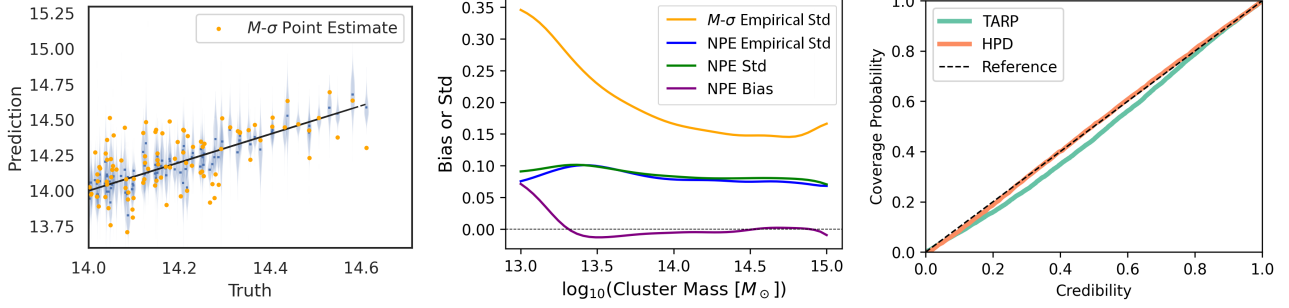


Figure 3. Left: Comparison between true and predicted galaxy cluster masses. The x-axis shows true masses; the y-axis shows predictions. Orange points represent baseline M - σ estimates, and violin plots indicate the machine learning posterior distributions. The black diagonal denotes perfect agreement. Middle: Bias and standard deviation in cluster mass as a function of true mass. The orange and blue lines show the standard deviation of M - σ and NPE mean relative to true mass, respectively; the green line shows the mean predicted posterior standard deviation; and the purple line shows the mean prediction bias. The dashed black line indicates zero bias. Right: Uncertainty calibration assessment using highest posterior density (HPD, orange) and Tests of Accuracy with Random Points (TARP, green).

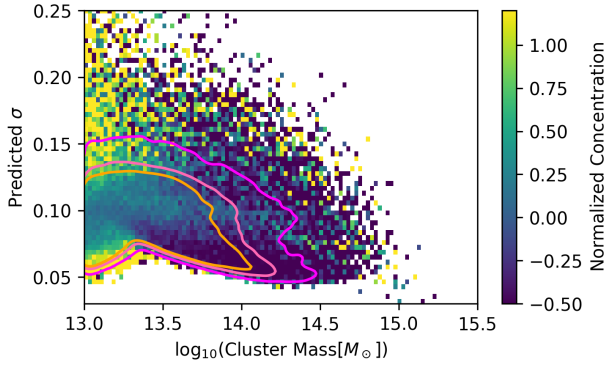


Figure 4. Predicted posterior standard deviation as a function of true cluster mass, color-coded by the normalized cluster concentration. The overlaid contours mark the 90%, 94%, and 98% levels of the number density distribution, suggesting a bimodal pattern: the upper peak is associated with higher concentrations, while the lower peak corresponds to lower concentrations.

indicating well-calibrated posteriors.

3.2. Galaxy Property Analysis

To gain further insight into the behavior of our machine learning model, we examine how other cluster properties relate to its predictions. We observe a noticeable impact of cluster concentration on the predicted posterior uncertainty.

In Figure 4, we show the predicted posterior standard deviation as a function of true cluster mass, color-coded by normalized cluster concentration. The normalization removes the mass dependence of concentration. The overlaid contours represent the number density distribution of clusters. We notice a bimodal structure in the contours: the

upper peak corresponds to higher concentrations, while the lower peak aligns with lower concentrations. This suggests that the model yields tighter constraints for less concentrated, likely more unrelaxed, clusters. This contradicts our expectation that relaxed systems are easier to constrain under equilibrium assumptions. Several possible explanations exist: overconcentrated clusters may be in fact out of equilibrium or in a transient phase of their dynamical evolution (Ludlow et al., 2012; Sereno & Covone, 2013), and the fraction of dynamically unrelaxed clusters increases at the high-mass end (Ludlow et al., 2013).

4. Conclusion and Discussion

We present a machine learning framework that combines Deep Sets and conditional normalizing flows to infer full posterior distributions of galaxy cluster masses from projected galaxy dynamics. Using the Uchuu-UniverseMachine simulation and a physically motivated set of observational features, we explore residual corrections beyond the standard M - σ relation. This residual information may reflect complex dynamical processes such as galaxy merger histories, motivating further investigation.

In terms of methodology, we explicitly incorporate projected positional information through a set-based architecture. This allows the model to learn directly from the full phase-space configuration of cluster member galaxies. By correcting the M - σ mass estimate via learned residuals, our framework utilizes physical priors from analytic models and adds more interpretability to the prediction. Finally, the use of conditional normalizing flows enables full posterior estimation.

Compared to the baseline M - σ estimates, our results significantly reduce scatter by approximately a factor of two and produce well-calibrated uncertainties across the full

mass range. Notably, we find that the predicted uncertainty varies somewhat with cluster concentration: the model yields slightly tighter constraints for less concentrated and likely more dynamically unrelaxed clusters. While this trend deviates from expectations based on equilibrium assumptions, it may reflect subtle dynamical effects captured by the model. This finding also highlights the potential of machine learning not only to improve predictive performance but also to reveal nuanced patterns in cluster physics. Further investigation is needed to understand the origin of this concentration-dependent behavior and its implications for cluster physics.

Finally, for application to real observations such as DESI (DESI Collaboration et al., 2016), PFS (Tamura et al., 2016), Rubin/LSST (Ivezić et al., 2019), and Euclid (Laureijs et al., 2011), a few additional steps are needed to test the robustness of our model. First, we need to evaluate its reliability with respect to membership selection, including completeness and interloper contamination. Second, we should assess its performance across different simulations, since it is currently trained within a specific cosmology. Although the highly nonlinear and largely virialized nature of cluster dynamics suggests limited sensitivity to cosmological variations, future work is necessary to confirm these effects.

Acknowledgements

We thank Hy Trac, Matthew Ho, and Michelle Ntampaka for valuable discussions. We thank the reviewers for their valuable and constructive feedback, which helped improve the quality of our work. This research used computing resources at Kavli IPMU and the Flatiron Institute. The Kavli IPMU is supported by the WPI (World Premier International Research Center) Initiative of the MEXT (Japanese Ministry of Education, Culture, Sports, Science and Technology). Leander Thiele was supported by JSPS KAKENHI Grant 24K22878.

Impact Statement

This paper presents work whose goal is to advance the field of applying Machine Learning in astrophysical problems. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

- Allen, S. W., Evrard, A. E., and Mantz, A. B. Cosmological Parameters from Observations of Galaxy Clusters. *Annu. Rev. Astron. Astrophys.*, 49(1):409–470, September 2011. doi: 10.1146/annurev-astro-081710-102514.
- Applegate, D. E., von der Linden, A., Kelly, P. L., Allen, M. T., Allen, S. W., Burchat, P. R., Burke, D. L., Ebeling, H., Mantz, A., and Morris, R. G. Weighing the Giants - III. Methods and measurements of accurate galaxy cluster weak-lensing masses. *Mon. Not. Roy. Astron. Soc.*, 439(1):48–72, March 2014. doi: 10.1093/mnras/stt2129.
- Aung, H., Nagai, D., Klypin, A., Behroozi, P., Abdullah, M. H., Ishiyama, T., Prada, F., Pérez, E., López Cacheiro, J., and Ruedas, J. The Uchuu-universe machine data set: galaxies in and around clusters. *Mon. Not. Roy. Astron. Soc.*, 519(2):1648–1656, February 2023. doi: 10.1093/mnras/stac3514.
- Behroozi, P., Wechsler, R. H., Hearin, A. P., and Conroy, C. UNIVERSEMACHINE: The correlation between galaxy growth and dark matter halo assembly from $z = 0$ –10. *Mon. Not. Roy. Astron. Soc.*, 488(3):3143–3194, September 2019. doi: 10.1093/mnras/stz1182.
- Behroozi, P. S., Wechsler, R. H., and Wu, H.-Y. The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. *Astrophys. J.*, 762(2):109, January 2013. doi: 10.1088/0004-637X/762/2/109.
- Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., Singh, R., Szerlip, P., Horsfall, P., and Goodman, N. D. Pyro: Deep Universal Probabilistic Programming. *arXiv e-prints*, art. arXiv:1810.09538, October 2018. doi: 10.48550/arXiv.1810.09538.
- Cranmer, K., Brehmer, J., and Louppe, G. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020. doi: 10.1073/pnas.1912789117. URL <https://www.pnas.org/doi/abs/10.1073/pnas.1912789117>.
- de Santi, N. S. M., Shao, H., Villaescusa-Navarro, F., Abramo, L. R., Teyssier, R., Villanueva-Domingo, P., Ni, Y., Anglés-Alcázar, D., Genel, S., Hernández-Martínez, E., Steinwandel, U. P., Lovell, C. C., Dolag, K., Castro, T., and Vogelsberger, M. Robust Field-level Likelihood-free Inference with Galaxies. *Astrophys. J.*, 952(1):69, July 2023. doi: 10.3847/1538-4357/acd1e2.
- DESI Collaboration, Aghamousa, A., Aguilar, J., Ahlen, S., Alam, S., Allen, L. E., Allende Prieto, C., Annis, J., Bailey, S., Balland, C., Ballester, O., Baltay, C., Beaufore, L., Bebek, C., Beers, T. C., Bell, E. F., Bernal, J. L., Besuner, R., Beutler, F., Blake, C., Bleuler, H., Blomqvist, M., Blum, R., Bolton, A. S., Briceno, C., Brooks, D., Brownstein, J. R., Buckley-Geer, E., Burden, A., Burtin, E., Busca, N. G., Cahn, R. N., Cai, Y.-C., Cardiel-Sas, L., Carlberg, R. G., Carton, P.-H., Casas, R., Castander, F. J., Cervantes-Cota, J. L., Claybaugh, T. M., Close, M., Coker, C. T., Cole, S., Comparat, J., Cooper, A. P., Cousinou, M. C., Crocce, M., Cuby, J.-G., Cunningham, D. P.,

- Davis, T. M., Dawson, K. S., de la Macorra, A., De Vicente, J., Delubac, T., Derwent, M., Dey, A., Dhungana, G., Ding, Z., Doel, P., Duan, Y. T., Ealet, A., Edelstein, J., Eftekharzadeh, S., Eisenstein, D. J., Elliott, A., Escoffier, S., Evatt, M., Fagrellius, P., Fan, X., Fanning, K., Farahi, A., Farihi, J., Favole, G., Feng, Y., Fernandez, E., Findlay, J. R., Finkbeiner, D. P., Fitzpatrick, M. J., Flaugh, B., Flender, S., Font-Ribera, A., Forero-Romero, J. E., Fosalba, P., Frenk, C. S., Fumagalli, M., Gaensicke, B. T., Gallo, G., Garcia-Bellido, J., Gaztanaga, E., Pietro Gentile Fusillo, N., Gerard, T., Gershkovich, I., Giannantonio, T., Gillet, D., Gonzalez-de-Rivera, G., Gonzalez-Perez, V., Gott, S., Graur, O., Gutierrez, G., Guy, J., Habib, S., Heetderks, H., Heetderks, I., Heitmann, K., Hellwing, W. A., Herrera, D. A., Ho, S., Holland, S., Honscheid, K., Huff, E., Hutchinson, T. A., Huterer, D., Hwang, H. S., Illa Laguna, J. M., Ishikawa, Y., Jacobs, D., Jeffrey, N., Jelinsky, P., Jennings, E., Jiang, L., Jimenez, J., Johnson, J., Joyce, R., Jullo, E., Juneau, S., Kama, S., Karcher, A., Karkar, S., Kehoe, R., Kennamer, N., Kent, S., Kilbinger, M., Kim, A. G., Kirkby, D., Kisner, T., Kitanidis, E., Kneib, J.-P., Koposov, S., Kovacs, E., Koyama, K., Kremin, A., Kron, R., Kronig, L., Kueter-Young, A., Lacey, C. G., Lavever, R., Lahav, O., Lambert, A., Lampton, M., Landriau, M., Lang, D., Lauer, T. R., Le Goff, J.-M., Le Guillou, L., Le Van Suu, A., Lee, J. H., Lee, S.-J., Leitner, D., Lesser, M., Levi, M. E., L’Huillier, B., Li, B., Liang, M., Lin, H., Linder, E., Loebman, S. R., Lukić, Z., Ma, J., MacCrann, N., Magneville, C., Makarem, L., Manera, M., Manser, C. J., Marshall, R., Martini, P., Massey, R., Matheson, T., McCauley, J., McDonald, P., McGreer, I. D., Meisner, A., Metcalfe, N., Miller, T. N., Miquel, R., Moustakas, J., Myers, A., Naik, M., Newman, J. A., Nichol, R. C., Nicola, A., Nicolati da Costa, L., Nie, J., Niz, G., Norberg, P., Nord, B., Norman, D., Nugent, P., O’Brien, T., Oh, M., and Olsen, K. A. G. The DESI Experiment Part I: Science, Targeting, and Survey Design. *arXiv e-prints*, art. arXiv:1611.00036, October 2016. doi: 10.48550/arXiv.1611.00036.
- Dodelson, S., Heitmann, K., Hirata, C., Honscheid, K., Roodman, A., Seljak, U., Slosar, A., and Trodden, M. Cosmic Visions Dark Energy: Science. *arXiv e-prints*, art. arXiv:1604.07626, April 2016. doi: 10.48550/arXiv.1604.07626.
- Dolatabadi, H. M., Erfani, S., and Leckie, C. Invertible Generative Modeling using Linear Rational Splines. *arXiv e-prints*, art. arXiv:2001.05168, January 2020. doi: 10.48550/arXiv.2001.05168.
- Durkan, C., Bekasov, A., Murray, I., and Papamakarios, G. Neural Spline Flows. *arXiv e-prints*, art. arXiv:1906.04032, June 2019. doi: 10.48550/arXiv.1906.04032.
- Garuda, N., Wu, J. F., Nelson, D., and Pillepich, A. Estimating Dark Matter Halo Masses in Simulated Galaxy Clusters with Graph Neural Networks. *arXiv e-prints*, art. arXiv:2411.12629, November 2024. doi: 10.48550/arXiv.2411.12629.
- Geller, M. J., Diaferio, A., Rines, K. J., and Serra, A. L. Measuring the mass distribution in galaxy clusters. *The Astrophysical Journal*, 764(1):58, Jan 2013. doi: 10.1088/0004-637X/764/1/58. URL <https://dx.doi.org/10.1088/0004-637X/764/1/58>.
- Giles, P. A., Maughan, B. J., Dahle, H., Bonamente, M., Landry, D., Jones, C., Joy, M., Murray, S. S., and van der Pyl, N. Chandra measurements of a complete sample of X-ray luminous galaxy clusters: the luminosity-mass relation. *Mon. Not. Roy. Astron. Soc.*, 465(1):858–884, February 2017. doi: 10.1093/mnras/stw2621.
- Greenberg, D., Nonnenmacher, M., and Macke, J. Automatic posterior transformation for likelihood-free inference. In Chaudhuri, K. and Salakhutdinov, R. (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2404–2414. PMLR, 09–15 Jun 2019. URL <https://proceedings.mlr.press/v97/greenberg19a.html>.
- Gupta, N. and Reichardt, C. L. Mass Estimation of Galaxy Clusters with Deep Learning. I. Sunyaev-Zel’dovich Effect. *Astrophys. J.*, 900(2):110, September 2020. doi: 10.3847/1538-4357/aba694.
- Gupta, N. and Reichardt, C. L. Mass Estimation of Galaxy Clusters with Deep Learning II. Cosmic Microwave Background Cluster Lensing. *Astrophys. J.*, 923(1):96, December 2021. doi: 10.3847/1538-4357/ac32d0.
- Haiman, Z., Mohr, J. J., and Holder, G. P. Constraints on Cosmological Parameters from Future Galaxy Cluster Surveys. *Astrophys. J.*, 553(2):545–561, June 2001. doi: 10.1086/320939.
- Hilton, M., Sifón, C., Naess, S., Madhavacheril, M., Oguri, M., Roza, E., Rykoff, E., Abbott, T. M. C., Adhikari, S., Aguena, M., Aiola, S., Allam, S., Amodeo, S., Amon, A., Annis, J., Ansarinejad, B., Aros-Bunster, C., Austermann, J. E., Avila, S., Bacon, D., Battaglia, N., Beall, J. A., Becker, D. T., Bernstein, G. M., Bertin, E., Bhandarkar, T., Bhargava, S., Bond, J. R., Brooks, D., Burke, D. L., Calabrese, E., Carrasco Kind, M., Carretero, J., Choi, S. K., Choi, A., Conselice, C., da Costa, L. N., Costanzi, M., Crichton, D., Crowley, K. T., Dünner, R., Denison, E. V., Devlin, M. J., Dicker, S. R., Diehl, H. T., Dietrich, J. P., Doel, P., Duff, S. M., Duivenvoorden, A. J., Dunkley, J., Everett, S., Ferraro, S., Ferrero, I., Ferté, A., Flaugh, J.,

- B., Frieman, J., Gallardo, P. A., García-Bellido, J., Gaztanaga, E., Gerdes, D. W., Giles, P., Golec, J. E., Gralla, M. B., Grandis, S., Gruen, D., Gruendl, R. A., Gschwend, J., Gutierrez, G., Han, D., Hartley, W. G., Hasselfield, M., Hill, J. C., Hilton, G. C., Hincks, A. D., Hinton, S. R., Ho, S. P. P., Honscheid, K., Hoyle, B., Hubmayr, J., Huppenberger, K. M., Hughes, J. P., Jaelani, A. T., Jain, B., James, D. J., Jeltema, T., Kent, S., Knowles, K., Koopman, B. J., Kuehn, K., Lahav, O., Lima, M., Lin, Y. T., Lokken, M., Loubser, S. I., MacCrann, N., Maia, M. A. G., Marriage, T. A., Martin, J., McMahon, J., Melchior, P., Menanteau, F., Miquel, R., Miyatake, H., Moodley, K., Morgan, R., Mroczkowski, T., Nati, F., Newburgh, L. B., Niemack, M. D., Nishizawa, A. J., Ogando, R. L. C., Orlowski-Scherer, J., Page, L. A., Palmese, A., Partridge, B., Paz-Chinchón, F., Phakathi, P., Plazas, A. A., Robertson, N. C., Romer, A. K., Carnero Rosell, A., Salatino, M., Sanchez, E., Schaan, E., Schillaci, A., Sehgal, N., Serrano, S., Shin, T., Simon, S. M., Smith, M., Soares-Santos, M., Spergel, D. N., Staggs, S. T., Storer, E. R., Suchyta, E., Swanson, M. E. C., Tarle, G., Thomas, D., To, C., Trac, H., Ullom, J. N., Vale, L. R., Van Lanen, J., Vavagiakis, E. M., De Vicente, J., Wilkinson, R. D., Wollack, E. J., Xu, Z., and Zhang, Y. The Atacama Cosmology Telescope: A Catalog of ~ 4000 Sunyaev-Zel'dovich Galaxy Clusters. *Astrophys. J. Suppl.*, 253(1): 3, March 2021. doi: 10.3847/1538-4365/abd023.
- Ho, M., Rau, M. M., Ntampaka, M., Farahi, A., Trac, H., and Póczos, B. A Robust and Efficient Deep Learning Method for Dynamical Mass Measurements of Galaxy Clusters. *Astrophys. J.*, 887(1):25, December 2019. doi: 10.3847/1538-4357/ab4f82.
- Ho, M., Farahi, A., Rau, M. M., and Trac, H. Approximate Bayesian Uncertainties on Deep Learning Dynamical Mass Estimates of Galaxy Clusters. *Astrophys. J.*, 908(2): 204, February 2021. doi: 10.3847/1538-4357/abd101.
- Ho, M., Ntampaka, M., Rau, M. M., Chen, M., Lansberry, A., Ruehle, F., and Trac, H. The dynamical mass of the Coma cluster from deep learning. *Nature Astronomy*, 6: 936–941, June 2022. doi: 10.1038/s41550-022-01711-1.
- Ho, M., Soltis, J., Farahi, A., Nagai, D., Evrard, A., and Ntampaka, M. Benchmarks and explanations for deep learning estimates of X-ray galaxy cluster masses. *Mon. Not. Roy. Astron. Soc.*, 524(3):3289–3302, September 2023. doi: 10.1093/mnras/stad2005.
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., Abel, B., Acosta, E., Allsman, R., Alonso, D., AlSayyad, Y., Anderson, S. F., Andrew, J., Angel, J. R. P., Angeli, G. Z., Ansari, R., Antilogus, P., Araujo, C., Armstrong, R., Arndt, K. T., Astier, P., Aubourg, É., Auza, N., Axelrod, T. S., Bard, D. J., Barr, J. D., Barrau, A., Bartlett, J. G., Bauer, A. E., Bauman, B. J., Baumont, S., Bechtol, E., Bechtol, K., Becker, A. C., Becla, J., Beldica, C., Bellavia, S., Bianco, F. B., Biswas, R., Blanc, G., Blazek, J., Blandford, R. D., Bloom, J. S., Bogart, J., Bond, T. W., Booth, M. T., Borgland, A. W., Borne, K., Bosch, J. F., Boutigny, D., Brackett, C. A., Bradshaw, A., Brandt, W. N., Brown, M. E., Bullock, J. S., Burchat, P., Burke, D. L., Cagnoli, G., Calabrese, D., Callahan, S., Callen, A. L., Carlin, J. L., Carlson, E. L., Chandrasekharan, S., Charles-Emerson, G., Chesley, S., Cheu, E. C., Chiang, H.-F., Chiang, J., Chirino, C., Chow, D., Ciardi, D. R., Claver, C. F., Cohen-Tanugi, J., Cockrum, J. J., Coles, R., Connolly, A. J., Cook, K. H., Cooray, A., Covey, K. R., Cribbs, C., Cui, W., Cutri, R., Daly, P. N., Daniel, S. F., Daruich, F., Daubard, G., Daues, G., Dawson, W., Delgado, F., Dellapenna, A., de Peyster, R., de Val-Borro, M., Digel, S. W., Doherty, P., Dubois, R., Dubois-Felsmann, G. P., Durech, J., Economou, F., Eifler, T., Eracleous, M., Emmons, B. L., Fausti Neto, A., Ferguson, H., Figueroa, E., Fisher-Levine, M., Focke, W., Foss, M. D., Frank, J., Freemon, M. D., Gangler, E., Gawiser, E., Geary, J. C., Gee, P., Geha, M., Gessner, C. J. B., Gibson, R. R., Gilmore, D. K., Glanzman, T., Glick, W., Goldina, T., Goldstein, D. A., Goodenow, I., Graham, M. L., Gressler, W. J., Gris, P., Guy, L. P., Guyonnet, A., Haller, G., Harris, R., Hascall, P. A., Haupt, J., Hernandez, F., Herrmann, S., Hileman, E., Hoblitt, J., Hodgson, J. A., Hogan, C., Howard, J. D., Huang, D., Huffer, M. E., Ingraham, P., Innes, W. R., Jacoby, S. H., Jain, B., Jammes, F., Jee, M. J., Jenness, T., Jernigan, G., Jevremović, D., Johns, K., Johnson, A. S., Johnson, M. W. G., Jones, R. L., Juramy-Gilles, C., Jurić, M., Kalirai, J. S., Kallivayalil, N. J., Kalmbach, B., Kantor, J. P., Karst, P., Kasliwal, M. M., Kelly, H., Kessler, R., Kinnison, V., Kirkby, D., Knox, L., Kotov, I. V., Krabbendam, V. L., Krughoff, K. S., Kubánek, P., Kuczewski, J., Kulkarni, S., Ku, J., Kurita, N. R., Lage, C. S., Lambert, R., Lange, T., Langton, J. B., Le Guillou, L., Levine, D., Liang, M., Lim, K.-T., Lintott, C. J., Long, K. E., Lopez, M., Lotz, P. J., Lupton, R. H., Lust, N. B., MacArthur, L. A., Mahabal, A., Mandelbaum, R., Markiewicz, T. W., Marsh, D. S., Marshall, P. J., Marshall, S., May, M., McKercher, R., McQueen, M., Meyers, J., Migliore, M., Miller, M., and Mills, D. J. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *Astrophys. J.*, 873 (2):111, March 2019. doi: 10.3847/1538-4357/ab042c.
- Jung, G., Ravenni, A., Baldi, M., Coulton, W. R., Jamieson, D., Karagiannis, D., Liguori, M., Shao, H., Verde, L., Villaescusa-Navarro, F., and Wandelt, B. D. Quijote-PNG: The Information Content of the Halo Mass Function. *Astrophys. J.*, 957(1):50, November 2023. doi: 10.3847/1538-4357/acf70.
- Kodi Ramanah, D., Wojtak, R., and Arendse, N. Simulation-

- based inference of dynamical galaxy cluster masses with 3D convolutional neural networks. *Mon. Not. Roy. Astron. Soc.*, 501(3):4080–4091, March 2021. doi: 10.1093/mnras/staa3922.
- Laureijs, R., Amiaux, J., Arduini, S., Auguères, J. L., Brinchmann, J., Cole, R., Cropper, M., Dabin, C., Duvel, L., Ealet, A., Garilli, B., Gondoin, P., Guzzo, L., Hoar, J., Hoekstra, H., Holmes, R., Kitching, T., Maciaszek, T., Mellier, Y., Pasian, F., Percival, W., Rhodes, J., Saavedra Criado, G., Sauvage, M., Scaramella, R., Valenziano, L., Warren, S., Bender, R., Castander, F., Cimatti, A., Le Fèvre, O., Kurki-Suonio, H., Levi, M., Lilje, P., Meylan, G., Nichol, R., Pedersen, K., Popa, V., Rebolo Lopez, R., Rix, H. W., Rottgering, H., Zeilinger, W., Grupp, F., Hudelot, P., Massey, R., Meneghetti, M., Miller, L., Paltani, S., Paulin-Henriksson, S., Pires, S., Saxton, C., Schrabback, T., Seidel, G., Walsh, J., Aghanim, N., Amendola, L., Bartlett, J., Baccigalupi, C., Beaulieu, J. P., Benabed, K., Cuby, J. G., Elbaz, D., Fosalba, P., Gavazzi, G., Helmi, A., Hook, I., Irwin, M., Kneib, J. P., Kunz, M., Mannucci, F., Moscardini, L., Tao, C., Teyssier, R., Weller, J., Zamorani, G., Zapatero Osorio, M. R., Boulade, O., Foumond, J. J., Di Giorgio, A., Guttridge, P., James, A., Kemp, M., Martignac, J., Spencer, A., Walton, D., Blümchen, T., Bonoli, C., Bortoletto, F., Cerna, C., Corcione, L., Fabron, C., Jahnke, K., Ligorì, S., Madrid, F., Martin, L., Morgante, G., Pampolona, T., Prieto, E., Riva, M., Toledo, R., Trifoglio, M., Zerbi, F., Abdalla, F., Douspis, M., Grenet, C., Borgani, S., Bouwens, R., Courbin, F., Delouis, J. M., Dubath, P., Fontana, A., Frailis, M., Grazian, A., Koppenhöfer, J., Mansutti, O., Melchior, M., Mignoli, M., Mohr, J., Neissner, C., Noddle, K., Poncet, M., Scodreggio, M., Serrano, S., Shane, N., Starck, J. L., Surace, C., Taylor, A., Verdoes-Kleijn, G., Vuerli, C., Williams, O. R., Zacchei, A., Altieri, B., Escudero Sanz, I., Kohley, R., Oosterbroek, T., Astier, P., Bacon, D., Bardelli, S., Baugh, C., Bellagamba, F., Benoist, C., Bianchi, D., Biviano, A., Branchini, E., Carbone, C., Cardone, V., Clements, D., Colombi, S., Conselice, C., Cresci, G., Deacon, N., Dunlop, J., Fedeli, C., Fontanot, F., Franzetti, P., Giocoli, C., Garcia-Bellido, J., Gow, J., Heavens, A., Hewett, P., Heymans, C., Holland, A., Huang, Z., Ilbert, O., Joachimi, B., Jennins, E., Kerins, E., Kiessling, A., Kirk, D., Kotak, R., Krause, O., Lahav, O., van Leeuwen, F., Lesgourgues, J., Lombardi, M., Magliocchetti, M., Maguire, K., Majerotto, E., Maoli, R., Marulli, F., Maurogordato, S., McCracken, H., McLure, R., Melchiorri, A., Merson, A., Moresco, M., Nonino, M., Norberg, P., Peacock, J., Pello, R., Penny, M., Pettorino, V., Di Porto, C., Pozzetti, L., Quercellini, C., Radovich, M., Rassat, A., Roche, N., Ronayette, S., and Rossetti, E. Euclid Definition Study Report. *arXiv e-prints*, art. arXiv:1110.3193, October 2011. doi: 10.48550/arXiv.1110.3193.
- Lemos, P., Coogan, A., Hezaveh, Y., and Perreault-Levasseur, L. Sampling-Based Accuracy Testing of Posterior Estimators for General Inference. *40th International Conference on Machine Learning*, 202:19256–19273, January 2023. doi: 10.48550/arXiv.2302.03026.
- Ludlow, A. D., Navarro, J. F., Li, M., Angulo, R. E., Boylan-Kolchin, M., and Bett, P. E. The dynamical state and mass-concentration relation of galaxy clusters. *Mon. Not. Roy. Astron. Soc.*, 427(2):1322–1328, December 2012. doi: 10.1111/j.1365-2966.2012.21892.x.
- Ludlow, A. D., Navarro, J. F., Boylan-Kolchin, M., Bett, P. E., Angulo, R. E., Li, M., White, S. D. M., Frenk, C., and Springel, V. The mass profile and accretion history of cold dark matter haloes. *Mon. Not. Roy. Astron. Soc.*, 432(2):1103–1113, June 2013. doi: 10.1093/mnras/stt526.
- Lueckmann, J.-M., Goncalves, P. J., Bassetto, G., Öcal, K., Nonnenmacher, M., and Macke, J. H. Flexible statistical inference for mechanistic models of neural dynamics. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/addfa9b7e234254d26e9c7f2af1005cb-Paper.pdf.
- Mantz, A. B., Allen, S. W., Morris, R. G., von der Linden, A., Applegate, D. E., Kelly, P. L., Burke, D. L., Donovan, D., and Ebeling, H. Weighing the giants—v. galaxy cluster scaling relations. *Monthly Notices of the Royal Astronomical Society*, 463(4):3582–3603, 09 2016. ISSN 0035-8711. doi: 10.1093/mnras/stw2250. URL <https://doi.org/10.1093/mnras/stw2250>.
- McClintock, T., Varga, T. N., Gruen, D., Roza, E., Rykoff, E. S., Shin, T., Melchior, P., DeRose, J., Seitz, S., Dietrich, J. P., Sheldon, E., Zhang, Y., von der Linden, A., Jeltema, T., Mantz, A. B., Romer, A. K., Allen, S., Becker, M. R., Bermeo, A., Bhargava, S., Costanzi, M., Everett, S., Farahi, A., Hamaus, N., Hartley, W. G., Hollowood, D. L., Hoyle, B., Israel, H., Li, P., MacCrann, N., Morris, G., Palmese, A., Plazas, A. A., Pollina, G., Rau, M. M., Simet, M., Soares-Santos, M., Troxel, M. A., Vergara Cervantes, C., Wechsler, R. H., Zuntz, J., Abbott, T. M. C., Abdalla, F. B., Allam, S., Annis, J., Avila, S., Bridle, S. L., Brooks, D., Burke, D. L., Carnero Rosell, A., Carrasco Kind, M., Carretero, J., Castander, F. J., Crocce, M., Cunha, C. E., D’Andrea, C. B., da Costa, L. N., Davis, C., De Vicente, J., Diehl, H. T., Doel, P., Drlica-Wagner, A., Evrard, A. E., Flaugh, B., Fosalba,

- P., Frieman, J., García-Bellido, J., Gaztanaga, E., Gerdes, D. W., Giannantonio, T., Gruendl, R. A., Gutierrez, G., Honscheid, K., James, D. J., Kirk, D., Krause, E., Kuehn, K., Lahav, O., Li, T. S., Lima, M., March, M., Marshall, J. L., Menanteau, F., Miquel, R., Mohr, J. J., Nord, B., Ogando, R. L. C., Roodman, A., Sanchez, E., Scarpine, V., Schindler, R., Sevilla-Noarbe, I., Smith, M., Smith, R. C., Sobreira, F., Suchyta, E., Swanson, M. E. C., Tarle, G., Tucker, D. L., Vikram, V., Walker, A. R., Weller, J., and DES Collaboration. Dark Energy Survey Year 1 results: weak lensing mass calibration of redMaPPer galaxy clusters. *Mon. Not. Roy. Astron. Soc.*, 482(1):1352–1378, January 2019. doi: 10.1093/mnras/sty2711.
- Ntampaka, M., Trac, H., Sutherland, D. J., Fromenteau, S., Póczos, B., and Schneider, J. Dynamical Mass Measurements of Contaminated Galaxy Clusters Using Machine Learning. *Astrophys. J.*, 831(2):135, November 2016. doi: 10.3847/0004-637X/831/2/135.
- Ntampaka, M., ZuHone, J., Eisenstein, D., Nagai, D., Vikhlinin, A., Hernquist, L., Marinacci, F., Nelson, D., Pakmor, R., Pillepich, A., Torrey, P., and Vogelsberger, M. A Deep Learning Approach to Galaxy Cluster X-Ray Masses. *Astrophys. J.*, 876(1):82, May 2019. doi: 10.3847/1538-4357/ab14eb.
- Old, L., Skibba, R. A., Pearce, F. R., Croton, D., Muldrew, S. I., Muñoz-Cuartas, J. C., Gifford, D., Gray, M. E., von der Linden, A., Mamon, G. A., Merrifield, M. R., Müller, V., Pearson, R. J., Ponman, T. J., Saro, A., Sepp, T., Sifón, C., Tempel, E., Tundo, E., Wang, Y. O., and Wojtak, R. Galaxy cluster mass reconstruction project - I. Methods and first results on galaxy-based techniques. *Mon. Not. Roy. Astron. Soc.*, 441(2):1513–1536, June 2014. doi: 10.1093/mnras/stu545.
- Papamakarios, G. and Murray, I. Fast ϵ -free inference of simulation models with bayesian conditional density estimation. In Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/file/6aca97005c68f1206823815f66102863-Paper.pdf.
- Sereno, M. and Covone, G. The mass-concentration relation in massive galaxy clusters at redshift ~ 1 . *Mon. Not. Roy. Astron. Soc.*, 434(1):878–887, September 2013. doi: 10.1093/mnras/stt1086.
- Tamura, N., Takato, N., Shimono, A., Moritani, Y., Yabe, K., Ishizuka, Y., Ueda, A., Kamata, Y., Aghazarian, H., Arnouts, S., Barban, G., Barkhouser, R. H., Borges, R. C., Braun, D. F., Carr, M. A., Chabaud, P.-Y., Chang, Y.-C., Chen, H.-Y., Chiba, M., Chou, R. C. Y., Chu, Y.-H., Cohen, J., de Almeida, R. P., de Oliveira, A. C., de Oliveira, L. S., Dekany, R. G., Dohlen, K., dos Santos, J. B., dos Santos, L. H., Ellis, R., Fabricius, M., Ferrand, D., Ferreira, D., Golebiowski, M., Greene, J. E., Gross, J., Gunn, J. E., Hammond, R., Harding, A., Hart, M., Heckman, T. M., Hirata, C. M., Ho, P., Hope, S. C., Hovland, L., Hsu, S.-F., Hu, Y.-S., Huang, P.-J., Jaquet, M., Jing, Y., Karr, J., Kimura, M., King, M. E., Komatsu, E., Le Brun, V., Le Fèvre, O., Le Fur, A., Le Mignant, D., Ling, H.-H., Loomis, C. P., Lupton, R. H., Madec, F., Mao, P., Marrara, L. S., Mendes de Oliveira, C., Minowa, Y., Morantz, C., Murayama, H., Murray, G. J., Ohyama, Y., Orndorff, J., Pascal, S., Pereira, J. M., Reiley, D., Reinecke, M., Ritter, A., Roberts, M., Schwochert, M. A., Seiffert, M. D., Smee, S. A., Sodre, L., Spergel, D. N., Steinkraus, A. J., Strauss, M. A., Surace, C., Suto, Y., Suzuki, N., Swinbank, J., Tait, P. J., Takada, M., Tamura, T., Tanaka, Y., Tresse, L., Verducci, O., Vibert, D., Vidal, C., Wang, S.-Y., Wen, C.-Y., Yan, C.-H., and Yasuda, N. Prime Focus Spectrograph (PFS) for the Subaru telescope: overview, recent progress, and future perspectives. In Evans, C. J., Simard, L., and Takami, H. (eds.), *Ground-based and Airborne Instrumentation for Astronomy VI*, volume 9908 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pp. 99081M, August 2016. doi: 10.1117/12.2232103.
- Thiele, L., Cranmer, M., Coulton, W., Ho, S., and Spergel, D. N. Predicting the thermal Sunyaev-Zel’dovich field using modular and equivariant set-based neural networks. *Machine Learning: Science and Technology*, 3(3):035002, September 2022. doi: 10.1088/2632-2153/ac78c2.
- Voit, G. M. Tracing cosmic evolution with clusters of galaxies. *Rev. Mod. Phys.*, 77:207–258, Apr 2005. doi: 10.1103/RevModPhys.77.207. URL <https://link.aps.org/doi/10.1103/RevModPhys.77.207>.
- Wang, B. Y., Pisani, A., Villaescusa-Navarro, F., and Wandelt, B. D. Machine-learning Cosmology from Void Properties. *Astrophys. J.*, 955(2):131, October 2023. doi: 10.3847/1538-4357/aceaf6.
- Yan, Z., Mead, A. J., Van Waerbeke, L., Hinshaw, G., and McCarthy, I. G. Galaxy cluster mass estimation with deep learning and hydrodynamical simulations. *Mon. Not. Roy. Astron. Soc.*, 499(3):3445–3458, December 2020. doi: 10.1093/mnras/staa3030.
- Yee, H. K. C. and Ellingson, E. Correlations of Richness and Global Properties in Galaxy Clusters. *Astrophys. J.*, 585(1):215–226, March 2003. doi: 10.1086/345929.
- Zwicky, F. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6:110–127, January 1933.

A. Training and validation loss curves

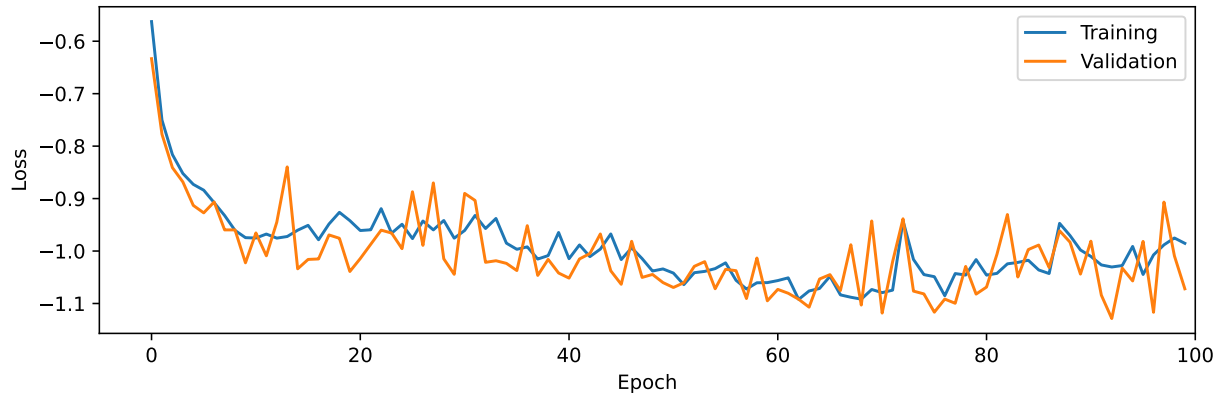


Figure 5. Conditional log-likelihood loss during model training. Both training and validation losses generally decrease over epochs, indicating improved model fit.