Draft version December 16, 2021
Typeset using LaTeX **twocolumn** style in AASTeX631

Inferring dark matter substructure with astrometric lensing beyond the power spectrum

Anonymous author(s)

ABSTRACT

Astrometry—the precise measurement of positions and motions of celestial objects—has emerged as a promising avenue for characterizing the dark matter population in our Galaxy. By leveraging recent advances in simulation-based inference and neural network architectures, we introduce a novel method to search for global dark matter-induced gravitational lensing signatures in astrometric datasets. Our method based on neural likelihood-ratio estimation shows significantly enhanced sensitivity to a cold dark matter population and more favorable scaling with measurement noise compared to existing approaches based on two-point correlation statistics, establishing machine learning as a powerful tool for characterizing dark matter using astrometric data.

Keywords: astrostatistics techniques (1886) — cosmology (343) — dark matter (353) — gravitational lensing (670) — convolutional neural networks (1938) — astrometry (80)

1. INTRODUCTION AND BACKGROUND

Although there exists plenty of evidence for dark matter (DM) on galactic scales and above (see Green (2021) for a recent overview), the distribution of DM clumps—subhalos—on sub-galactic scales is less well-understood and remains an active area of cosmological study. This distribution additionally correlates with and may provide clues about the underlying particle physics nature of dark matter (see e. g., Schutz (2020); Bode et al. (2001); Dalcanton & Hogan (2001)), highlighting its relevance across multiple domains.

While more massive dark matter subhalos can be detected and studied through their association with luminous tracers such as bound stellar populations, subhalos with smaller masses $\lesssim 10^9\,\mathrm{M}_\odot$ are not generally associated with luminous matter (Fitts et al. 2017; Read et al. 2017), rendering their characterization challenging. Gravitational effects provide one of the few avenues to probe the distribution of these otherwise-invisible subhalos (Buckley & Peter 2018). Gravitational lensing i. e., the bending of light from a background source due to a foreground mass, is one such effect and has been proposed in various incarnations as a probe of dark subhalos. Strong gravitational lensing, for example, has been used to infer the presence of dark matter substructure in galaxies outside of our own (Hezaveh et al. 2016; Vegetti et al. 2010; Gilman et al. 2020; Vegetti et al. 2012). Astrometric lensing, on the other hand, has recently emerged as a promising way to characterize the dark matter subhalo population within the Milky Way.

Astrometry refers to the precise measurement of the positions and motions of luminous celestial objects like stars and galaxies. Gravitational lensing of these background objects by a moving foreground mass, such as a dark matter subhalo, can imprint a characteristic, correlated signal on their measured kinematics (angular velocities and/or accelerations). Van Tilburg et al. (2018) introduced several methods for extracting this signature, including computing convolutions of the expected lensing signal on astrometric datasets and detecting local kinematic outliers. Mondino et al. (2020) applied the former method to data from the *Gaia* satellite, obtaining constraints on the abundance of dark compact objects in the Milky Way and showcasing the applicability of astrometric dark matter searches in a practical setting. Finally, Mishra-Sharma et al. (2020) proposed using the angular power spectrum of the astrometric field as an observable to infer the population properties of subhalos in our Galaxy, leveraging the collective, correlated signal of a large subhalo sample.

Astrometric datasets are inherently high-dimensional, consisting of positions and kinematics of potentially millions of objects. Especially when the expected signal consists of the collective imprint of a large number of lenses, characterizing their population properties involves marginalizing over all possible configurations of subhalos, rendering the likelihood intractable and usually necessitating the use of simplified data representations like the power spectrum. While effective, such simplification can result in loss of information compared to that contained in the original dataset when the ex-

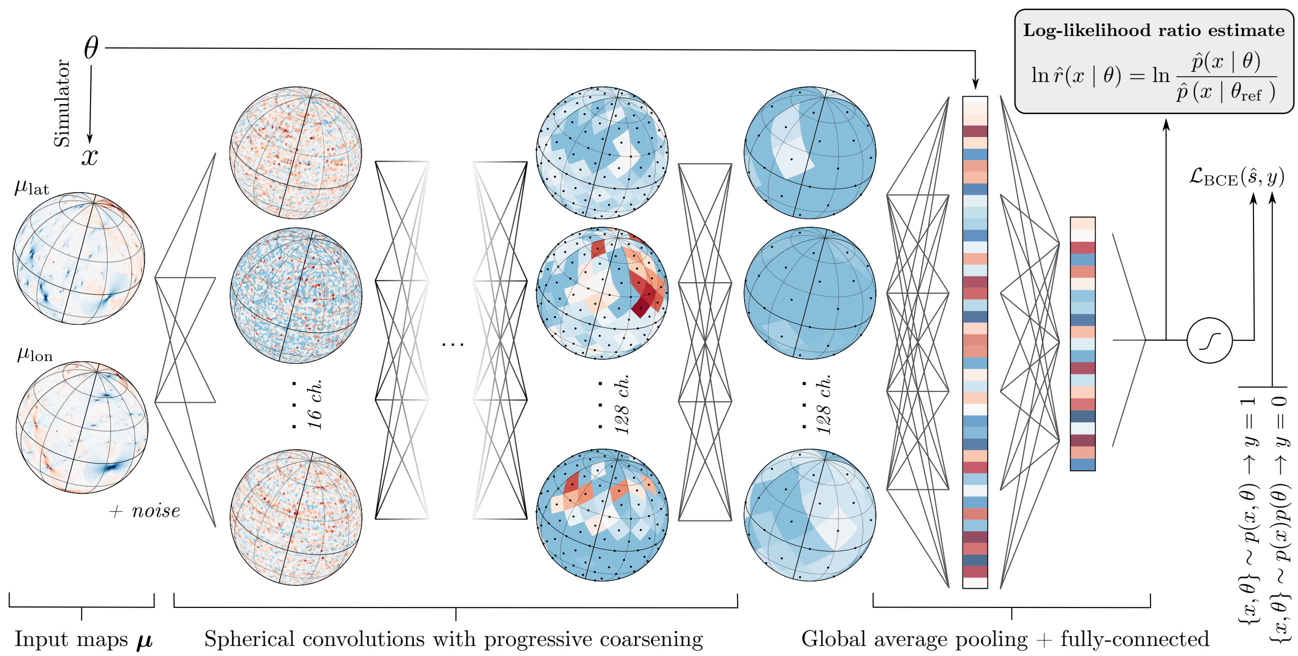

Figure 1. A schematic illustration of the method and neural network architecture used in this work.

pected signal is non-Gaussian in nature. The existence of systematic effects that are degenerate with a putative signal in the low-dimensional summary domain can further inhibit sensitivity.

The dawn of the era of precision astrometry, with the *Gaia* satellite (Gaia Collaboration 2016) having recently delivered the most precise astrometric dataset to-date (Gaia Collaboration 2018a; Lindegren et al. 2018; Gaia Collaboration 2021) and surveys including the Square Kilometer Array (SKA) (Fomalont & Reid 2004; Jarvis et al. 2015) and Roman Space Telescope (WFIRST Astrometry Working Group 2019) set to achieve further leaps in sensitivity over the next decade, calls for methods that can extract more information from these datasets than is possible using existing techniques. In this direction, Vattis et al. (2020) proposed using a binary classifier in order to detect either the presence or absence of a substructure signal in astrometric maps. In this paper, we introduce an *inference* approach that uses spherical convolutional neural networks—exploiting the symmetry structure of the signal and data domain—in conjunction with parameterized classifiers (Cranmer et al. 2015; Baldi et al. 2016) in order to estimate likelihood ratios associated with the abundance of a cold dark matter population directly from a binned map of the astrometric velocity field. We show that our method outperforms established proposals based on the two-point correlation statistics of the astrometric field, both in absolute sensitivity as well as its scaling with measurement noise. While we focus on the specific domain application to astrometric dark matter searches, we note that the method as presented here is broadly applicable to data sampled on the celestial sphere, which is ubiquitous in astrophysics and cosmology. More generally, the paper showcases how neural network architectures suited to processing real-world data structures—in our cases, pixelated vector fields of velocities on the celestial sphere—can be combined with advancements in simulation-based inference in order to directly perform inference on complex, high-dimensional datasets without resorting to the use of simplified summary statistics.

2. MODEL AND INFERENCE

2.1. *The forward model*

Our datasets consist of the 2-dimensional angular velocity map of background sources on the celestial sphere. In order to define the forward model we need to specify the properties of background sources as well as the population properties of dark matter subhalos acting as gravitational lenses. We focus in this work on subhalos within the canonical cold dark matter scenario; details of the population model along with a the prescription for computing the induced velocity signal are provided in App. A. The subhalo fraction $f_{\mathrm{sub}} \in \mathbb{R}$, quantifying the expected fraction of the mass of the Milky Way contributed by subhalos in the range 10^{-6}–$10^{10}\,\mathrm{M}_{\odot}$, is taken to be the parameter of interest.

We take our source population to consist of remote, point-like galaxies known as quasars which, due to their large distances from the Earth, are not expected to have significant intrinsic angular velocities. We assume the sources to be isotropically-distributed in the baseline configuration, and further study the effect of relaxing this assumption using an existing catalog of quasars from *Gaia*'s second data release (DR2). The velocity maps are assumed to be spatially binned, and we use a *HEALPix* binning (Gorski et al. 2005) with resolution parameter `nside=64`, corresponding to $N_{\rm pix} = 49{,}152$ pixels over the full sky with pixel area $\sim 0.8\,\deg^2$. The values within each pixel then quantify the average latitudinal and longitudinal velocity components of quasars within that pixel. An example of the induced velocity signal on part of the celestial sphere, projected along the Galactic latitudinal and longitudinal directions and exhibiting dipole-like structures, is shown in the leftmost column of Fig. 1. This pixelization level was motivated by the results of Mishra-Sharma et al. (2020), which showed the typical angular size of cold dark matter subhalos significantly contributing to the astrometric lensing signal to be much larger than the degree-scale pixel size used here. This can also be seen from the simulated signal realizations in Fig. 1

In order to enable a comparison with traditional approaches—which are generally not expected to be sensitive to a cold dark matter subhalo population with next-generation astrometric surveys (Van Tilburg et al. 2018; Mishra-Sharma et al. 2020)—we benchmark using an optimistic observational configuration corresponding to measuring the proper motions of $N_q = 10^8$ quasars over the fully sky with noise $\sigma_\mu = 0.1\,\mu{\rm as\,yr}^{-1}$. The final input maps $x \in \mathbb{R}^{2\,N_{\rm pix}}$ are obtained by combining the simulated signal with a realization of the noise model.

2.2. *The power spectrum approach*

Mishra-Sharma et al. (2020) introduced an approach for extracting the astrometric signal due to a dark matter subhalo population by decomposing the observed map into its angular (vector) power spectrum. The power spectrum is a summary statistic ubiquitous in astrophysics and cosmology and quantifies the amount of correlation contained at different spatial scales. In the case of data on a sphere, the basis of spherical harmonics is often used, and the power spectrum then encodes the correlation structure on different multipoles ℓ. The power spectrum effectively captures the linear component of the signal and, when the underlying signal is a Gaussian random field, captures *all* of the relevant information contained in the map(s) (Tegmark 1997). The expected signal in the power spectrum domain can be evaluated semi-analytically using the formalism described in Mishra-Sharma et al. (2020) and, assuming a Gaussian likelihood, the expected sensitivity can be computed using a Fisher forecasting approach. We use this prescription as a comparison point to the method introduced here.

While effective, reduction of the full astrometric map to its power spectrum results in loss of information; this can be seen from the fact that the signal in the leftmost column of Fig. 1 is far from Gaussian. Furthermore, the existence of correlations on large angular scales due to e. g., biases in calibration of celestial reference frames (Gaia Collaboration 2018b) or systematic variations in measurements taken over different regions of the sky introduces degeneracies with a putative signal and precludes their usage in the present context. For this reason multipoles $\ell < 10$ were discarded in Mishra-Sharma et al. (2020), degrading the projected sensitivity.

2.3. *Likelihood-ratio estimation using parameterized classifiers*

Recent advances in machine learning have enabled methods that can be used to efficiently perform inference on models defined through complex simulations; see Cranmer et al. (2020) for a recent review. Here, we make use of neural likelihood-ratio estimation (Cranmer et al. 2015; Baldi et al. 2016; Brehmer et al. 2018a, 2020, 2018b; Hermans et al. 2019), previously applied to the problem of inferring dark matter substructure using observations of strong gravitational lenses (Brehmer et al. 2019) and cold stellar streams (Hermans et al. 2020).

Given a classifier that can distinguish between samples $\{x\} \sim p(x \mid \theta)$ drawn from parameter points θ and those from a fixed reference hypothesis $\{x\} \sim p(x \mid \theta_{\rm ref})$, the decision function output by the optimal classifier $s(x, \theta) = p(x \mid \theta)/(p(x \mid \theta) + p(x \mid \theta_{\rm ref}))$ is one-to-one with the likelihood ratio, $r(x \mid \theta) \equiv p(x \mid \theta)/p(x \mid \theta_{\rm ref}) = s(x, \theta)/(1 - s(x, \theta))$, a fact appreciated as the likelihood-ratio trick (Cranmer et al. 2015; Mohamed & Lakshminarayanan 2017). The classifier $s(x, \theta)$ in this case is a neural network that can work directly on the high-dimensional data, and is parameterized by the parameter of interest θ by having it included as an additional input feature. In order to improve numerical stability and reduce dependence on the fixed reference hypothesis $\theta_{\rm ref}$, we follow Hermans et al. (2019) and train a classifier to distinguish between data-sample pairs from the joint distribution $\{x, \theta\} \sim p(x, \theta)$ and those from a product of marginal distributions $\{x, \theta\} \sim p(x)p(\theta)$ (defining the reference hypothesis and in practice obtained by shuffling samples within a batch) using the binary cross-entropy (BCE) loss as the optimization objective.

We briefly highlight the advantages of this method over traditional paradigms for simulation-based inference such as Approximate Bayesian Computation (ABC, Rubin (1984); Sisson et al. (2018)). In ABC, samples $\{x\}$ from the forward model are compared to a particular dataset x', with the approximate posterior defined through the set of parameters whose corresponding samples most closely match the dataset of interest according to a similarity metric. In our case, the curse of dimensionality would require a manual reduction of the raw datasets $x \in \mathbb{R}^{2\,N_\mathrm{pix}}$ into lower-dimensional summaries $f(x) \in \mathbb{R}^n$ with $n \ll 2\,N_\mathrm{pix}$ (e. g., the power spectrum), in order to enable tractable inference. A similarity metric and tolerance threshold $||f(x)-f(x')|| < \epsilon$ must additionally be specified in order to trade off between sample efficiency and inference precision. The machine learning-based method, on the other hand, uses neural networks in order to directly extract useful representations from high-dimensional datasets and can learn a continuous mapping from the data to the statistic of interest, in our case the likelihood ratio. Finally, ABC inference has to be performed anew for each dataset of interest. Our method, in contrast, is *amortized*—following an upfront computational cost associated with training the likelihood-ratio estimator, evaluation on a new sample can be performed almost instantaneously. This allows us to efficiently test our model and compute diagnostics such as statistical coverage over large data samples.

2.4. *Extracting information from high-dimensional astrometric maps*

Since our data consists of a velocity field sampled on a sphere, we use a spherical convolutional neural network in order to directly learn useful representations from these maps that are efficiently suited for the downstream classification task. Specifically, we make use of *DeepSphere* (Defferrard et al. 2020; Perraudin et al. 2019), a graph-based convolutional neural network tailored to data sampled on a sphere. For this purpose, the *HEALPix* grid can be cast as a weighted undirected graph with N_pix vertices and edges connecting each pixel vertex to its set of 8 neighboring pixels. The weighted adjacency matrix over neighboring pixels (i, j) is given by $A_{ij} = \exp\left(-\Delta r_{ij}^2/\rho^2\right)$ where Δr_{ij} specifies the 3-dimensional Euclidean distance between the pixel centers and the widths ρ are obtained from Defferrard et al. (2020). *DeepSphere* then efficiently performs convolutions in the spectral domain using a basis of Chebychev polynomials as convolutional kernels (Defferrard et al. 2016); here, we set $K = 4$ as the maximum polynomial order.

All inputs are normalized to zero mean and unit standard deviation across the training sample. Starting with 2 scalar input channels representing the two orthogonal (Galactic latitude and longitude) components of the velocity vector map, we perform a graph convolution operation, increasing the channel dimension to 16 followed by a batch normalization, ReLU nonlinearity, and downsampling the representation by a factor of 4 with max pooling into the next coarser *HEALPix* resolution. Pooling leverages scale separation, preserving important characteristics of the signal across different resolutions. Four more such layers are employed, increasing the channel dimension by a factor of 2 at each step until a maximum of 128, with maps after the last convolutional layer having resolution `nside=2` corresponding to 48 pixels. At this stage, we average over the spatial dimension (known as global average pooling (Lin et al. 2014)) in order to encourage approximate rotation invariance, outputting 128 features onto which the parameter of interest $f_\mathrm{sub} \in \mathbb{R}$ is appended. These features are passed through a fully-connected network with (1024, 128) hidden units and ReLU activations outputting the classifier decision $\hat{s}$ by applying a sigmoidal projection.

We note that the signal in our case does not respect strict rotation invariance—as described in App. A, the motion of the Sun relative to the frame of rest of the Milky Way induces a preferred direction in the velocities of dark matter subhalos relative to our frame of reference, breaking the rotation symmetry of the signal. The anisotropy in the data domain is further exacerbated in the case of a realistic noise model (as explored in Sec. 3.3 below) where measured uncertainties vary along different directions on the celestial sphere. We finally note that by representing the input angular velocity vector field in terms of two independent scalar channels, we explicitly break the rotation equivariance of spherical convolutions due to differences in how scalar and vector representations transform under rotations (see, e. g., Esteves et al. (2020)). Given these limitations, we leave a detailed study of the equivariance properties desired of the architecture in the context of our application to future work.

2.5. *Model training and evaluation*

10^5 maps from the forward model were produced, with 15% of these held out for validation. Samples containing between 0 and 300 subhalos in expectation over the mass range 10^8–$10^{10}\,\mathrm{M}_\odot$, approximately corresponding to substructure fractions f_sub between 0 and 0.4, were generated from a uniform proposal distribution. The estimator was trained using a batch size of 64 for up to 50 epochs with early stopping if the validation loss had not improved after 10 epochs. The ADAM optimizer (Kingma & Ba 2017) was used with initial learning

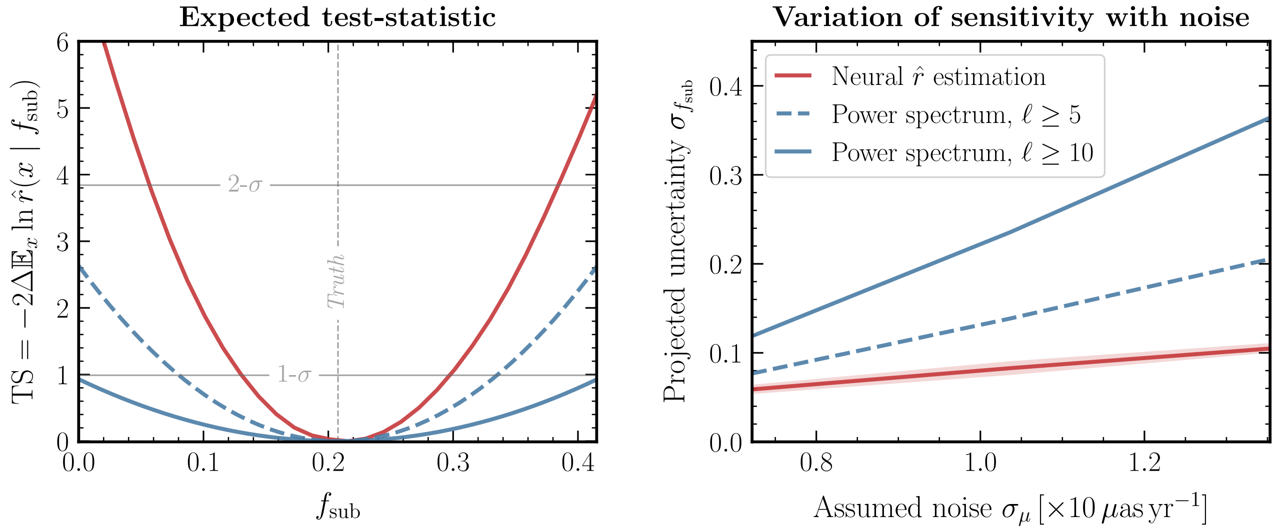

Figure 2. *(Left)* The expected log-likelihood ratio test-statistic (TS) profile for a cold dark matter population as a function of substructure fraction f_sub obtained using the neural likelihood-ratio estimation method introduced in this work (solid red line) compared with the corresponding profiles for existing approaches using power spectrum summaries with different multipole thresholds $\ell \gtrsim 5$ (dashed blue line) and $\ell \gtrsim 10$ (solid blue line). The vertical dotted line indicates the true benchmark value of the parameter f_sub in the test dataset. Our method shows enhanced sensitivity to a cold dark matter population compared to traditional approaches. *(Right)* Scaling of the expected sensitivities, quantified by the respective 1-σ uncertainties, with per-object instrumental noise. For the machine learning-based approach, the band quantifies the middle-95% containment of the inferred 1-σ uncertainty. Our method shows a more favorable scaling with assumed measurement noise.

rate 10^{-3} decayed through cosine annealing. A coarse grid search was used to inform the architecture and hyperparameter choices in this work. Experiments were performed on NVIDIA RTX8000 GPUs, taking ~ 10 minutes per training epoch for a total training time of $\sim$ 6–9 hours contingent on early stopping.

For a given test map, the log-likelihood ratio profile can be obtained by evaluating the trained estimator for different values of f_sub while keeping the input map fixed. The network output prior to the final sigmoidal projection directly gives the required log-likelihood ratio estimate: $\ln \hat{r} = S^{-1}(\hat{s})$, where S is the sigmoid function (Hermans et al. 2019, 2020). Figure 1 presents an illustrative summary of the neural network architecture and method used in this work.

3. EXPERIMENTS ON SIMULATED DATA

3.1. *Baseline results and diagnostics*

We evaluate our trained likelihood-ratio estimator on maps drawn from a benchmark configuration motivated by Hütten et al. (2016); Springel et al. (2008), containing 150 subhalos in expectation between 10^8–$10^{10}\,\mathrm{M}_\odot$ and corresponding to $f_\mathrm{sub} \simeq 0.2$. The left panel of Fig. 2 shows the expected log-likelihood ratio test-statistic (TS) as a function of substructure fraction f_sub for this nominal configuration. This is obtained by evaluating the trained estimator on 100 test maps over a uniform grid in f_sub and taking the point-wise mean. Corresponding curves using the power spectrum approach are shown in blue, using minimum multipoles of $\ell \geq 5$ (dashed) and $\ell \geq 10$ (solid). Thresholds corresponding to 1- and 2-σ significance assuming a χ^2-distributed TS are shown as the horizontal grey lines. We see that sensitivity gains of over a factor of ~ 2 can be expected for this particular benchmark when using the machine learning approach compared to the traditional power spectrum approach. No significant bias on the central value of the inferred DM abundance relative to the overall uncertainty scale is observed.

The right panel of Fig. 2 shows the scaling of expected 1-σ uncertainty on substructure fraction f_sub with assumed noise per quasar, keeping the number of quasars fixed (red, with the line showing the median and shaded band corresponding to the middle-95% containment of the uncertainty inferred over 50 test datasets) compared to the power spectrum approach (blue lines). A far more favorable scaling of the machine learning approach is seen compared to the power spectrum approach, suggesting that it is especially advantageous in low signal-to-noise regimes that are generally most relevant for dark matter searches.

Finally, we assess the quality of the approximate likelihood-ratio estimator through a test of statistical coverage. Within a hypothesis testing framework, this is necessary in order to ensure that the learned estimator is conservative over the parameter range of interest and does not produce overly confident or biased results (Hermans et al. 2021). We obtain the estimated TS profile for 1000 simulated samples with true substructure fraction values drawn from the range $f_{\rm sub} \in [0.1, 0.3]$. In doing so, we exclude parameter points towards the edges of our parameter space since the corresponding confidence intervals in these cases would extend outside of the tested parameter range, as can also be inferred from the baseline analysis shown in Fig. 2. For nominal confidence levels in the range $1 - \alpha \in [0.05, 0.95]$ we compute the empirical coverage over the set of samples, defined as the fraction of samples whose true parameter values fall within the TS confidence interval. The confidence level for a given nominal confidence interval is computed under the assumption that the TS is χ^2-distributed (Wilks 1938). The procedure is repeated for 10 different sets of 1000 samples in order to estimate the statistical uncertainty associated with the empirical coverage.

The results of the coverage test are shown in Fig. 3, illustrating the median (solid red line) and middle-68% containment (red band) of the empirical coverage. We see that the empirical coverage has the desired property of being conservative while still being close to the perfectly-calibrated regime (dashed grey line). We emphasize that this diagnostic tests the quality of the likelihood-ratio estimator over the entire evaluation parameter range $f_{\rm sub} \in [0.1, 0.3]$ rather than the baseline value $f_{\rm sub} \simeq 0.2$ in isolation.

3.2. *Experiments with unmodeled noise correlated on large scales*

Since the existence of measurement noise correlated on large spatial scales is a potential source of systematic uncertainty when working with astrometric maps, we test the susceptibility of our method to such effects by creating simulated data containing large-scale noise not previously seen by the trained estimator. Instead of assuming a scale-invariant noise power spectrum $C_\ell^{\rm noise} = 4\pi\sigma_\mu^2/N_q$ (Mishra-Sharma et al. 2020), in this case we model noise with an order of magnitude excess in power on scales $\ell \lesssim 10$, parameterized as $C_\ell^{\rm noise} = 4\pi\sigma_\mu^2/N_q \cdot (10 - 9S(\ell - 10))$ where S denotes the sigmoid function. The left panel of Fig. 4 illustrates this noise model (thicker green line) as well as the power spectrum of one simulated realization from this model (thinner green line, obtained using the *HEALPix* module `anafast`) contrasted with the standard scale-invariant noise case (red lines). The right panel of Fig. 4 shows the expected log-likelihood ratio test-statistic pro-

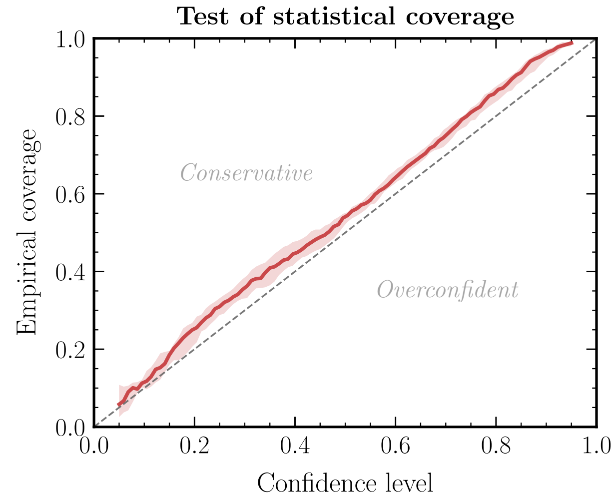

Figure 3. The empirical coverage of the baseline likelihood-ratio estimator as a function of nominal confidence level. The median (solid red line) and middle-68% containment (red band) over 10 sets of 1000 samples is shown. The estimator is seen to have the desired property of being conservative, while still closely tracing the perfectly-calibrated regime (dashed grey line).

file for the two cases. Although a bias in the maximum-likelihood estimate of $f_{\rm sub}$ is seen when the test data has unmodeled noise (green line), the true test parameter value (dashed vertical line) is seen to lie well within the inferred 1-σ confidence interval. This suggests that the method is only marginally susceptible to substantive amounts of correlated noise on large spatial scales.

3.3. *Experiments with a data-driven noise model*

We finally assess the performance of our method using a realistic noise model obtained from the astrometric catalog of quasars in *Gaia*'s Data Release 2 (DR2) (Gaia Collaboration 2018a; Lindegren et al. 2018). The catalog contains the measured 2-dimensional positions, proper motions, as well as proper motion uncertainties of 555,934 quasars. Although the measured uncertainties in this case are too large for the catalog to be viable for the current scientific use-case, they can be rescaled and used to construct a data-driven noise model for testing the viability of our method on forthcoming astrometric data.

We compute the pixel-wise proper motion uncertainties as the inverse-variance weighted values within each *HEALPix* pixel; $\sigma_\mu^{\rm pix} = \left(\sum_{q\in{\rm pix}} \sigma_{\mu,q}^{-2}\right)^{-1/2}$, where $\sigma_{\mu,q}^2$ are the provided variances of individual quasars within a given pixel. This results in a highly anisotropic noise model, shown in the left column of Fig. 5, additionally having different uncertainties in the latitudinal (top

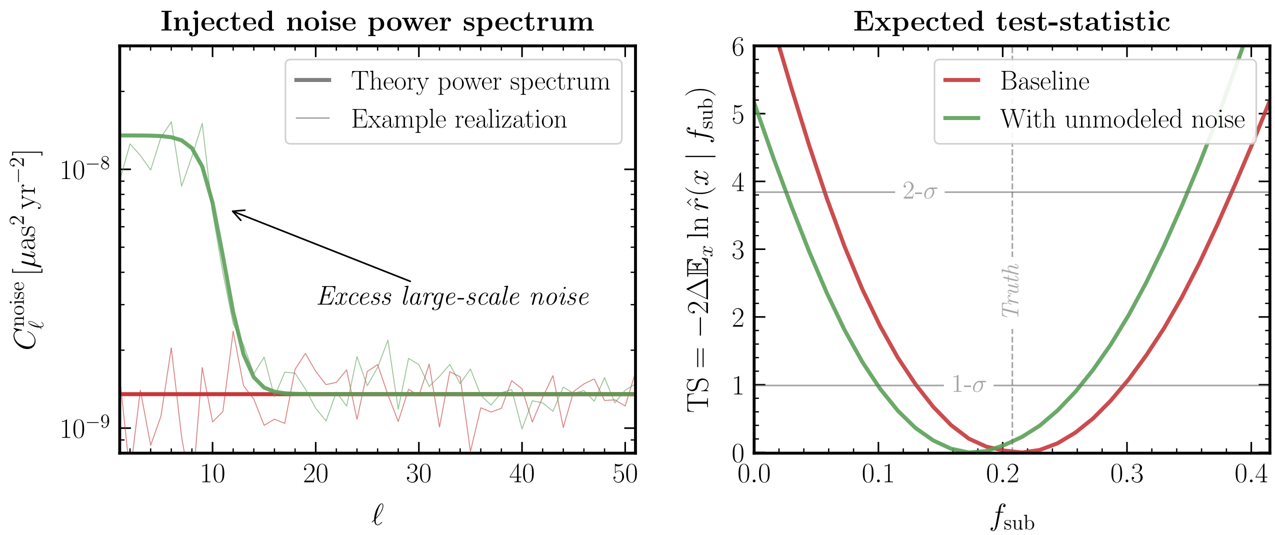

Figure 4. *(Left)* The power spectrum of the noise model (thicker green line) used to study the impact of correlated noise on large spatial scales, not modeled during training, on the performance of the likelihood-ratio estimator. The thinner green line shows the power spectrum of an example noise realization instantiated from this noise model. The red lines show corresponding power spectra for a scale-invariant noise model. *(Right)* The expected test-statistic profile for a model evaluated on maps containing excess large-scale noise (green line) compared to the model evaluated on maps with scale-invariant noise (red line). A bias in the maximum-likelihood estimate returned by the model is seen when substantial unaccounted-for noise is presented in the test maps.

row) and longitudinal (bottom row) directions. As expected due to occlusion from the Galactic disk, uncertainties are significantly higher towards the Galactic plane where the catalog has low completeness, additionally varying over the sky due to the scanning pattern and time-dependent instrumental response of the satellite. The region closest to the plane where no quasars are included in the catalog (shown in grey) is masked, testing the effect of partial sky coverage. In order to enable a direct comparison, the mean per-pixel variance for the data-driven noise model is normalized to that used in the baseline experiments in Sec. 3.1.

We evaluate the expected likelihood-ratio test statistic on samples generated with the data-driven noise model using two different estimators: *(i)* the baseline estimator trained using samples with isotropic noise, as describe in Sec. 3.1, and *(ii)* an estimator trained on samples generated with the correct, data-driven noise model used during evaluation. For *(ii)*, a larger training batch size of 512 was found to provide better results, with all other hyperparameters being the same as the baseline case. The expected likelihood-ratio test statistic profiles for these cases are shown in the right column of Fig. 5. Interestingly, evaluating the baseline estimator on samples with the data-driven noise model (dashed teal) produces results very similar to the baseline case (solid red). Using the correct, data-driven noise model on the other hand produces tighter constraints (solid teal) due to the fact that large portions of the sky in this case have smaller modeled uncertainties compared to the baseline case. In either case, successful recover of the astrometric lensing signal can be seen. These experiments demonstrate the viability of our method in the context of real-world applications, and we leave a more detailed study of our method in the context of forthcoming astrometric surveys and datasets to future work.

4. CONCLUSIONS AND OUTLOOK

We have introduced a method to analyze astrometric datasets over large regions of the sky using techniques based on machine learning with the aim of inferring the lensing signature of a dark matter substructure. We have shown our method to be significantly more sensitive to a cold dark matter subhalo population compared to established methods based on global summary statistics, with more favorable scaling as a function of measurement noise. Since the collection and reduction of astrometric data is an expensive endeavor, the use of methods that can take advantage of more of the available information can be equated to long periods of data-taking, underscoring their importance. Additionally, unlike the power spectrum approach, the current method does not require the construction of a numerically-expensive estimator to account for non-uniform exposure, selection effects, and instrumental noise in realistic datasets. These, as well as any other modeled observational effects, can be incorporated directly at the level of the forward model.

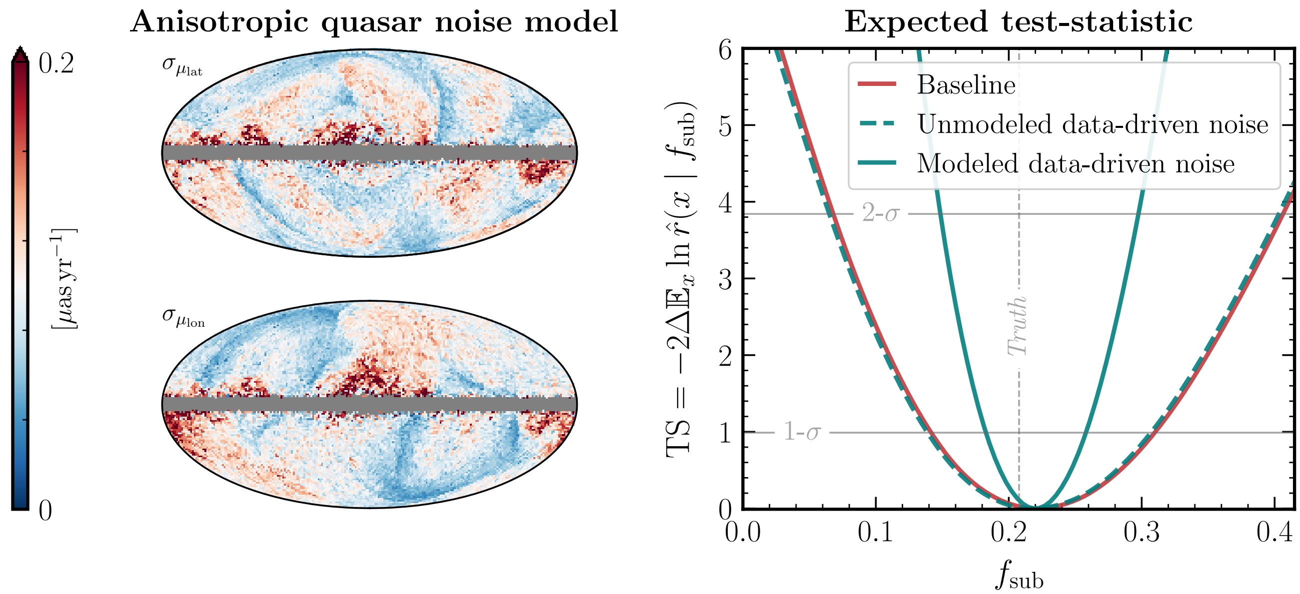

Figure 5. *(Left)* The data-driven noise model derived using the *Gaia* DR2 quasar catalog, showing maps of the effective per-quasar uncertainty within each pixel in the latitudinal (top) and longitudinal (bottom) directions. *(Right)* The expected likelihood-ratio test statistic profile evaluated on samples with data-driven noise using the baseline estimator (dashed teal line) and the estimator trained with data-driven noise (solid teal line), compared with the baseline estimator evaluated on samples with isotropic noise (solid red line).

We have focused in this work on assessing sensitivity to a cold dark matter-like subhalo population with quasar velocity astrometry, which is within the scope of upcoming radio surveys like the SKA (Fomalont & Reid 2004; Jarvis et al. 2015). Our method can also be applied in a straightforward manner to look for the *acceleration* lensing signal imprinted on Milky Way stars, in particular sourced by a population of more compact subhalos than those expected in the cold dark matter scenario. These features are expected to imprint a larger degree of non-Gaussianity compared to the signal explored here (as can be seen, e. g., from Fig. 1 of Mishra-Sharma et al. (2020)), and machine learning methods may provide larger relative sensitivity gains when deployed in that context. Such analyses are within purview of the upcoming Roman exoplanet microlensing survey (Pardo & Doré 2021) as well as future *Gaia* data releases. We note that when the region of interest covers a smaller fraction of the celestial sphere, as expected for the Roman microlensing survey of the Galactic bulge, the use of conventional convolutional architectures may be preferred as more efficient compared to spherical convolutions when the flat-sky approximation is valid.

Several improvements and extensions to the method presented in this paper are possible. The use of architectures that can equivariantly handle vector inputs (Esteves et al. 2020) may aid in learning more efficient representations of the astrometric map. Using convolutions based on fixed rather than learned filters can additionally reduce model complexity and produce more interpretable representations (Cheng et al. 2020; Ha et al. 2021; Saydjari & Finkbeiner 2021; McEwen et al. 2021; Valogiannis & Dvorkin 2021). The use of methods for likelihood-ratio estimation that can leverage additional latent information in the forward model can significantly enhance the sample efficiency of the analysis (Brehmer et al. 2018a, 2020, 2018b; Stoye et al. 2018). We leave the study of these extensions as well as application of our method to other dark matter population scenarios to future work.

Astrometric lensing has been established as a promising way to characterize the Galactic dark matter population, with theoretical progress in recent years going in step with advances on the observational front. While this work is a first attempt at bringing principled machine learning techniques to this field, with the availability of increasingly complex datasets we expect machine learning to be an important general-purpose tool for future astrometric dark matter searches.

(Acknowledgments anonymized for review)

Software: *Astropy* (Robitaille et al. 2013; Price-Whelan et al. 2018), *healpy* (Gorski et al. 2005; Zonca et al. 2019), *IPython* (Pérez & Granger 2007), *Jupyter* (Kluyver et al. 2016), *Matplotlib* (Hunter 2007), *MLflow* (Chen et al. 2020), *NumPy* (Harris et al. 2020), *PyGSP* (Defferrard et al. 2017), *PyTorch* (Paszke et al. 2019), *PyTorch Geometric* (Fey & Lenssen 2019), *PyTorch Lightning* (Falcon et al. 2020), *sbi* (Tejero-Cantero et al. 2020), *SciPy* (Virtanen et al. 2020), and *seaborn* (Waskom et al. 2017).

APPENDIX

A. ADDITIONAL DETAILS ON THE FORWARD MODEL

We consider a population of Navarro-Frenk-White (NFW) (Navarro et al. 1996) subhalos following a power-law mass function, $\mathrm{d}n/\mathrm{d}m \propto m^{\alpha}$, with slope $\alpha = -1.9$ as expected if the population is sourced from nearly scale-invariant primordial fluctuations in the canonical Λ Cold Dark Matter (ΛCDM) scenario. The concentration-mass relation from Sánchez-Conde & Prada (2014) is used to model the concentrations associated with density profiles of individual subhalos. Subhalos between 10^7–$10^{10}\,\mathrm{M}_\odot$ are simulated, assuming the influence of lighter subhalos to be too small to be discernable (Mishra-Sharma et al. 2020).

The spatial distribution of subhalos in the Galactocentric frame is modeled using results from the Aquarius simulation following Hütten et al. (2016); Springel et al. (2008). Since this spatial distribution accounts for the depletion of subhalos towards the Galactic Center due to gravitational tidal effects, the angular number density of subhalos looking out from the Sun frame can be considered to be effectively isotropic.

The asymptotic velocities of subhalos in the Galactocentric frame are taken to follow a truncated Maxwell-Boltzmann distribution (Chandrasekhar 1939; Lisanti 2017) $f_{\mathrm{Gal}}(\mathbf{v}) \propto e^{-\mathbf{v}^2/v_0^2} \cdot H(v_{\mathrm{esc}} - |\mathbf{v}|)$, where $v_{\mathrm{esc}} = 550\,\mathrm{km\,s^{-1}}$ is the Galactic escape velocity (Piffl et al. 2014), $v_0 = 220\,\mathrm{km\,s^{-1}}$ (Kerr & Lynden-Bell 1986), and H is the Heaviside step function. Once instantiated, the positions and velocities of subhalos are transformed into the Galactic frame, assuming $R_\odot = 8.2\,\mathrm{kpc}$ to be the distance of the Sun from the Galactic Center (Gravity Collaboration 2019; Bovy 2020) and $\mathbf{v}_\odot = (11, 232, 7)\,\mathrm{km\,s^{-1}}$ its Galactocentric velocity (Schönrich et al. 2010). Note that the asymmetry in the direction of motion of the Sun in the Milky Way introduces a preferred direction for the Sun-frame velocities of subhalos, breaking strict rotation invariance in the forward model. Although not explicitly pursued here, the expected characteristic form of this asymmetry can be used as an additional distinguishing handle for the lensing signal, as was done in Mishra-Sharma et al. (2020).

Once a subhalo population has been instantiated using the forward model, the induced velocity lensing signal at different positions on the celestial sphere can be computed. Given a spherically-symmetric subhalo lens moving with transverse velocity $\mathbf{v}_l$, the expected lens-induced velocity for a background source at impact parameter $\mathbf{b}$ is given by (Van Tilburg et al. 2018)

$$\boldsymbol{\mu}(\mathbf{b}) = 4G_{\mathrm{N}} \left\{ \frac{M(b)}{b^2} \left[2\hat{\mathbf{b}} \left(\hat{\mathbf{b}} \cdot \mathbf{v}_l \right) - \mathbf{v}_l \right] - \frac{M'(b)}{b} \hat{\mathbf{b}} \left(\hat{\mathbf{b}} \cdot \mathbf{v}_l \right) \right\} \tag{A1}$$

where $M(b)$ and $M'(b)$ are the projected mass of the subhalo at a given impact parameter distance $b = |\mathbf{b}|$ and its gradient. In the context of our spatially-binned velocity map, $\mathbf{b}$ represents the vector from the center of the subhalo to the center of the respective *HEALPix* pixel.

REFERENCES

Baldi, P., Cranmer, K., Faucett, T., Sadowski, P., & Whiteson, D. 2016, Eur. Phys. J. C, 76, 235, arXiv: 1601.07913, doi: 10.1140/epjc/s10052-016-4099-4

Bode, P., Ostriker, J. P., & Turok, N. 2001, Astrophys. J., 556, 93, arXiv: astro-ph/0010389, doi: 10.1086/321541

Bovy, J. 2020, arXiv: 2012.02169

Brehmer, J., Cranmer, K., Louppe, G., & Pavez, J. 2018a, Phys. Rev. D, 98, 052004, arXiv: 1805.00020, doi: 10.1103/PhysRevD.98.052004

—. 2018b, Phys. Rev. Lett., 121, 111801, arXiv: 1805.00013, doi: 10.1103/PhysRevLett.121.111801

Brehmer, J., Louppe, G., Pavez, J., & Cranmer, K. 2020, Proc. Nat. Acad. Sci., 117, 5242, arXiv: 1805.12244, doi: 10.1073/pnas.1915980117

Brehmer, J., Mishra-Sharma, S., Hermans, J., Louppe, G., & Cranmer, K. 2019, Astrophys. J., 886, 49, arXiv: 1909.02005, doi: 10.3847/1538-4357/ab4c41

Buckley, M. R., & Peter, A. H. G. 2018, Phys. Rept., 761, 1, arXiv: 1712.06615, doi: 10.1016/j.physrep.2018.07.003

Chandrasekhar, S. 1939, An introduction to the study of stellar structure

Chen, A., et al. 2020, in Proceedings of the fourth international workshop on data management for end-to-end machine learning, 1–4

Cheng, S., Ting, Y.-S., Ménard, B., & Bruna, J. 2020, Mon. Not. Roy. Astron. Soc., 499, 5902, arXiv: 2006.08561, doi: 10.1093/mnras/staa3165

Cranmer, K., Brehmer, J., & Louppe, G. 2020, Proc. Nat. Acad. Sci., 117, 30055, arXiv: 1911.01429, doi: 10.1073/pnas.1912789117

Cranmer, K., Pavez, J., & Louppe, G. 2015, arXiv: 1506.02169

Dalcanton, J. J., & Hogan, C. J. 2001, Astrophys. J., 561, 35, arXiv: astro-ph/0004381, doi: 10.1086/323207

Defferrard, M., Bresson, X., & Vandergheynst, P. 2016, arXiv: 1606.09375

Defferrard, M., Martin, L., Pena, R., & Perraudin, N. 2017, PyGSP: Graph Signal Processing in Python, v0.5.0, Zenodo, doi: 10.5281/zenodo.1003158. https://doi.org/10.5281/zenodo.1003158

Defferrard, M., Milani, M., Gusset, F., & Perraudin, N. 2020, arXiv: 2012.15000

Esteves, C., Makadia, A., & Daniilidis, K. 2020, in Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual, ed. H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, & H. Lin. https://proceedings.neurips.cc/paper/2020/hash/6217b2f7e4634fa665d31d3b4df81b56-Abstract.html

Falcon, W., et al. 2020, PyTorchLightning/pytorch-lightning: 0.7.6 release, 0.7.6, Zenodo, doi: 10.5281/zenodo.3828935. https://doi.org/10.5281/zenodo.3828935

Fey, M., & Lenssen, J. E. 2019, in ICLR Workshop on Representation Learning on Graphs and Manifolds

Fitts, A., et al. 2017, Mon. Not. Roy. Astron. Soc., 471, 3547, arXiv: 1611.02281, doi: 10.1093/mnras/stx1757

Fomalont, E. B., & Reid, M. 2004, New Astron. Rev., 48, 1473, arXiv: astro-ph/0409611, doi: 10.1016/j.newar.2004.09.037

Gaia Collaboration. 2016, Astron. Astrophys., 595, A1, arXiv: 1609.04153, doi: 10.1051/0004-6361/201629272

—. 2018a, Astron. Astrophys., 616, A1, arXiv: 1804.09365, doi: 10.1051/0004-6361/201833051

—. 2018b, Astron. Astrophys., 616, A14, arXiv: 1804.09377, doi: 10.1051/0004-6361/201832916

—. 2021, Astron. Astrophys., 649, A1, arXiv: 2012.01533, doi: 10.1051/0004-6361/202039657

Gilman, D., Birrer, S., Nierenberg, A., et al. 2020, Mon. Not. Roy. Astron. Soc., 491, 6077, arXiv: 1908.06983, doi: 10.1093/mnras/stz3480

Gorski, K. M., Hivon, E., Banday, A. J., et al. 2005, Astrophys. J., 622, 759, arXiv: astro-ph/0409513, doi: 10.1086/427976

Gravity Collaboration. 2019, Astron. Astrophys., 625, L10, arXiv: 1904.05721, doi: 10.1051/0004-6361/201935656

Green, A. M. 2021, in Les Houches summer school on Dark Matter

Ha, W., Singh, C., Lanusse, F., Upadhyayula, S., & Yu, B. 2021, arXiv: 2107.09145

Harris, C. R., et al. 2020, Nature, 585, 357, doi: 10.1038/s41586-020-2649-2

Hermans, J., Banik, N., Weniger, C., Bertone, G., & Louppe, G. 2020, arXiv: 2011.14923

Hermans, J., Begy, V., & Louppe, G. 2019, arXiv: 1903.04057

Hermans, J., Delaunoy, A., Rozet, F., Wehenkel, A., & Louppe, G. 2021, Averting A Crisis In Simulation-Based Inference, arXiv: 2110.06581

Hezaveh, Y. D., et al. 2016, Astrophys. J., 823, 37, arXiv: 1601.01388, doi: 10.3847/0004-637X/823/1/37

Hunter, J. D. 2007, Computing In Science & Engineering, 9, 90

Hütten, M., Combet, C., Maier, G., & Maurin, D. 2016, JCAP, 09, 047, arXiv: 1606.04898, doi: 10.1088/1475-7516/2016/09/047

Jarvis, M. J., Bacon, D., Blake, C., et al. 2015, arXiv: 1501.03825

Kerr, F. J., & Lynden-Bell, D. 1986, Mon. Not. Roy. Astron. Soc., 221, 1023

Kingma, D. P., & Ba, J. 2017, Adam: A Method for Stochastic Optimization, arXiv: 1412.6980

Kluyver, T., et al. 2016, in ELPUB

Lin, M., Chen, Q., & Yan, S. 2014, Network In Network, arXiv: 1312.4400

Lindegren, L., Hernández, J., Bombrun, A., et al. 2018, A&A, 616, A2, arXiv: 1804.09366, doi: 10.1051/0004-6361/201832727

Lisanti, M. 2017, in Theoretical Advanced Study Institute in Elementary Particle Physics: New Frontiers in Fields and Strings, 399–446

McEwen, J. D., Wallis, C. G. R., & Mavor-Parker, A. N. 2021, arXiv: 2102.02828

Mishra-Sharma, S., Van Tilburg, K., & Weiner, N. 2020, Phys. Rev. D, 102, 023026, arXiv: 2003.02264, doi: 10.1103/PhysRevD.102.023026

Mohamed, S., & Lakshminarayanan, B. 2017, Learning in Implicit Generative Models, arXiv: 1610.03483

Mondino, C., Taki, A.-M., Van Tilburg, K., & Weiner, N. 2020, Phys. Rev. Lett., 125, 111101, arXiv: 2002.01938, doi: 10.1103/PhysRevLett.125.111101

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, Astrophys. J., 462, 563, arXiv: astro-ph/9508025, doi: 10.1086/177173

Pardo, K., & Doré, O. 2021, arXiv: 2108.10886

Paszke, A., et al. 2019, in Advances in Neural Information Processing Systems 32, ed. H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Curran Associates, Inc.), 8024–8035

Pérez, F., & Granger, B. E. 2007, Computing in Science and Engineering, 9, 21, doi: 10.1109/MCSE.2007.53

Perraudin, N., Defferrard, M., Kacprzak, T., & Sgier, R. 2019, Astronomy and Computing, 27, 130, arXiv: 1810.12186, doi: 10.1016/j.ascom.2019.03.004

Piffl, T., et al. 2014, Astron. Astrophys., 562, A91, arXiv: 1309.4293, doi: 10.1051/0004-6361/201322531

Price-Whelan, A., et al. 2018, Astron. J., 156, 123, arXiv: 1801.02634, doi: 10.3847/1538-3881/aabc4f

Read, J. I., Iorio, G., Agertz, O., & Fraternali, F. 2017, Mon. Not. Roy. Astron. Soc., 467, 2019, arXiv: 1607.03127, doi: 10.1093/mnras/stx147

Robitaille, T. P., et al. 2013, Astron. Astrophys., 558, A33, arXiv: 1307.6212, doi: 10.1051/0004-6361/201322068

Rubin, D. B. 1984, The Annals of Statistics, 12, 1151 , doi: 10.1214/aos/1176346785

Sánchez-Conde, M. A., & Prada, F. 2014, Mon. Not. Roy. Astron. Soc., 442, 2271, arXiv: 1312.1729, doi: 10.1093/mnras/stu1014

Saydjari, A. K., & Finkbeiner, D. P. 2021, arXiv: 2104.11244

Schönrich, R., Binney, J., & Dehnen, W. 2010, Mon. Not. Roy. Astron. Soc., 403, 1829, arXiv: 0912.3693, doi: 10.1111/j.1365-2966.2010.16253.x

Schutz, K. 2020, Phys. Rev. D, 101, 123026, arXiv: 2001.05503, doi: 10.1103/PhysRevD.101.123026

Sisson, S. A., Fan, Y., & Beaumont, M. 2018, Handbook of approximate Bayesian computation (CRC Press)

Springel, V., Wang, J., Vogelsberger, M., et al. 2008, Mon. Not. Roy. Astron. Soc., 391, 1685, arXiv: 0809.0898, doi: 10.1111/j.1365-2966.2008.14066.x

Stoye, M., Brehmer, J., Louppe, G., Pavez, J., & Cranmer, K. 2018, arXiv: 1808.00973

Tegmark, M. 1997, Phys. Rev. D, 55, 5895, arXiv: astro-ph/9611174, doi: 10.1103/PhysRevD.55.5895

Tejero-Cantero, A., et al. 2020, Journal of Open Source Software, 5, 2505, doi: 10.21105/joss.02505

Valogiannis, G., & Dvorkin, C. 2021, arXiv: 2108.07821

Van Tilburg, K., Taki, A.-M., & Weiner, N. 2018, JCAP, 07, 041, arXiv: 1804.01991, doi: 10.1088/1475-7516/2018/07/041

Vattis, K., Toomey, M. W., & Koushiappas, S. M. 2020, arXiv: 2008.11577

Vegetti, S., Koopmans, L. V. E., Bolton, A., Treu, T., & Gavazzi, R. 2010, Mon. Not. Roy. Astron. Soc., 408, 1969, arXiv: 0910.0760, doi: 10.1111/j.1365-2966.2010.16865.x

Vegetti, S., Lagattuta, D. J., McKean, J. P., et al. 2012, Nature, 481, 341, arXiv: 1201.3643, doi: 10.1038/nature10669

Virtanen, P., et al. 2020, Nature Methods, 17, 261, doi: 10.1038/s41592-019-0686-2

Waskom, M., et al. 2017, mwaskom/seaborn: v0.8.1 (September 2017), v0.8.1, Zenodo, doi: 10.5281/zenodo.883859. https://doi.org/10.5281/zenodo.883859

WFIRST Astrometry Working Group. 2019, Journal of Astronomical Telescopes, Instruments, and Systems, 5, 044005, arXiv: 1712.05420, doi: 10.1117/1.JATIS.5.4.044005

Wilks, S. S. 1938, The Annals of Mathematical Statistics, 9, 60 , doi: 10.1214/aoms/1177732360

Zonca, A., Singer, L., Lenz, D., et al. 2019, Journal of Open Source Software, 4, 1298, doi: 10.21105/joss.01298