Kinematic Sunyaev-Zel′dovich pairwise velocity reconstruction with machine learning

Yulin Gong● and Rachel Bean

Department of Astronomy, Cornell University, Ithaca, New York 14853, USA

We demonstrate that pairwise peculiar velocity correlations for galaxy clusters can be directly reconstructed from the kinematic Sunyaev-Zel'dovich (kSZ) signature imprinted in the cosmic microwave background (CMB) using a machine learning model with a gradient boosting algorithm trained on high-fidelity kSZ simulations. The machine learning model is trained using six to seven cluster features that are directly related to observables from CMB and large-scale structure surveys. We validate the capabilities of the approach in light of the presence of primary CMB, detector noise, and potential uncertainties in the cluster mass estimate and cluster center location. The pairwise velocity statistics extracted using the techniques developed here have the potential to elicit valuable cosmological constraints on dark energy, modified gravity models, and massive neutrinos with kSZ measurements from upcoming CMB surveys, including the Simons Observatory, CMB-S4 and CCAT, and the DESI and SDSS galaxy surveys.

I. INTRODUCTION

The origin of accelerated cosmic expansion remains a critical outstanding problem in physics. Measurements of the cosmic microwave background (CMB) radiation [1–8], baryon acoustic oscillations (BAO) (e.g. [9–14]), and type 1a supernovae (e.g. [15–18]) together provide exquisite constraints on the expansion history of the Universe. This expansion history is consistent with the standard cosmological model, which assumes general relativity (GR) and a cosmological constant, Λ, the simplest form of dark energy, as the component of the cosmic matter density proposed to explain the accelerated expansion of the Universe (e.g. [19–22]). Given the fine-tuning and coincidence problems [23–31] related to the discordance between the observed value of Λ and those naturally predicted from theory, modifications of gravity, beyond GR, have also been actively considered as alternative explanations for the accelerated expansion (e.g. see Ref. [32] for a review). Such modifications can be developed to match a ΛCDM expansion history but concurrently predict differences in the growth and dynamical properties of inhomogeneities, probed through the clustering and dynamical properties of large-scale structure (LSS), galaxies, and clusters of galaxies (e.g. [32–38]).

In the context of this paper, we focus on the use of the dynamics of galaxy clusters as a cosmological tracer of the underlying gravitational field. Galaxy clusters can be detected observationally through unique signatures left in CMB photons when they interact with the hot gas of a galaxy cluster, the Sunyaev-Zel'dovich (SZ) effect. The SZ effect occurs when CMB photons interact with electrons in galaxy clusters and can be separated into two principle components: the thermal Sunyaev-Zel'dovich (tSZ) effect and the kinematic Sunyaev-Zel'dovich effect (kSZ) [39–41] (and see reviews [42,43]). The tSZ is caused by the hot electrons with random velocities boosting the blackbody spectrum imprinting a characteristic frequency-dependent signature that facilitates its isolation from the CMB through multifrequency measurements. The kSZ is produced by the peculiar (bulk) line of sight motion of a galaxy cluster creating a Doppler shift of the CMB spectrum, and is one order of magnitude smaller than the tSZ effect and largely frequency independent making it harder to extract.

The kSZ signature is an observational tracer of the underlying peculiar velocities of clusters and, in turn, the gravitational potential [44–47]. The gravitational attraction between pairs of clusters creates inherent infall toward each other. This gravitational attraction leads to a pairwise correlation statistic that can provide a potentially sensitive measurement of the large-scale velocity field [48–51].

Despite the comparative challenges in kSZ versus tSZ detection, the prospect of extracting the peculiar velocities of galaxy clusters from the kSZ effect is actively investigated as it offers a potentially powerful technique for cosmological inference (e.g. [52–59]) and complementary constraints on the properties of gravity on cosmic scales and the neutrino mass sum to those from galaxy lensing and clustering measurements [60–63].

The pairwise cluster momentum has been a principal estimator used to extract out the kSZ signal. The mean pairwise cluster momentum estimator was first detected from the kSZ signal in [64] with the Atacama Cosmology

Telescope (ACT) CMB observations [65] and the LSS surveys of Sloan Digital Sky Survey (SDSS) [66,67]. Measurements from subsequent ACT and SDSS data releases, covering a greater survey area and expanded catalogs, have also been made [68,69]. The kSZ signal has also been detected using the same estimator, by the Planck Collaboration with SDSS [70] and the South Pole Telescope (SPT) Collaboration [71] using a cluster catalog from the Dark Energy Survey (DES) [72,73].

The kSZ signal can also be utilized with techniques other than pairwise momentum, including projected fields [74–78], velocity reconstruction [79–82], kSZ tomography [83,84], individual cluster measurement [85,86], CMB temperature dispersion [59], 21 cm intensity mapping [87], signatures from CMB anisotropies [88], and Fourier space analysis [89].

Typically, the extraction of pairwise velocity statistics is undertaken by combining the pairwise kSZ momentum estimators with observational estimation of the optical depths for the same cluster sample. This requires a clear understanding of how the kSZ signal and optical depth measurements are related in terms of their respective sampling of the cluster properties [90,91]. Optical depth estimates can be obtained using one or more observables such as the tSZ effect and x-ray observations; however, the optical depth estimation presents challenges in introducing additional uncertainties and potential biases [69,73,86,92–95].

In this work, we demonstrate the efficacy of an alternative approach focused on extracting cluster velocity information from the kSZ measurements directly, using a machine learning technique. These velocity estimates are then used to infer the pairwise velocity, rather than calculating it from separate measurements of the pairwise momentum and optical depth.

Machine learning, a subfield of artificial intelligence and computer science, is a powerful tool that uses statistical techniques to efficiently analyze and identify patterns in massive and complex datasets to acquire knowledge that may be computationally intractable using other approaches. With the growth in complexity and volume of astronomical data, machine learning methods have been used for a range of applications, including cluster mass estimation [96–105], cluster scaling relations [106–109], simulation and maps [110–113], cosmological parameter constraint and estimation [114–118], cosmic structure formation [119–122], reconstruction of the cosmological density and velocity fields [123–126], strong gravitational lensing [127–130], and weak gravitational lensing [131–137], amongst others (e.g. see Refs. [138–141]). Algorithms are normally designed without specific programming of physics, hence, machine learning provides an alternative data-driven method to the physics model-driven analysis paradigm.

Deep learning techniques have been used to recover galaxy cluster peculiar velocities [81,142] using 2D SZ images and 3D galaxy distributions respectively with a neural network. In this work, we use an alternative machine learning technique, the gradient boosting algorithm that has successfully been applied to other areas of astronomy ([143–146]), to reconstruct the pairwise peculiar velocity using 1D features that are commonly distilled from observations for use in kSZ analyses. This includes the disk kSZ temperatures, redshifts, and halo mass estimates. We test the capability of the model under conditions including realistic primordial CMB and detector noise for upcoming CMB observations as well as the impact of some key potential systematic effects. We train our machine learning algorithm on one kSZ simulation and apply it to test samples from two simulation datasets that consider similar astrophysical effects and cosmological models. Both datasets provide us with complete halo information including the redshift, mass, peculiar line-of-sight velocity, kSZ temperatures, and other key properties. This information allows us to efficiently train our model to reconstruct an unbiased cluster's peculiar velocity.

The following is a summary of the paper's outline: The theory for the kSZ signal extraction techniques and pairwise estimator is introduced in Sec. II, along with a description of the kSZ simulation datasets and halo catalogs that we used in this work. In Sec. III, we describe our feature engineering process, the structure of our machine learning model, and covariance estimates. In Sec. IV, we present the findings, including the sensitivity to various modeling assumptions and to systematic errors that might arise from cluster center mislocation and scattering in the cluster mass estimation. We cross-validate the model by training on one set of simulations and inferring the velocities from completely distinct kSZ and halo simulations. In Sec. V, we conclude with a summary of the approach, the key findings, and implications for future study.

II. BACKGROUND

The primary goal is to recover an estimate of the pairwise peculiar velocity correlations for a cosmological galaxy cluster sample using the information from galaxy surveys and kSZ temperatures measured from CMB observations. In Sec. II A, we discuss the simulated galaxy and CMB datasets used in this work. In Sec. II B, we describe the kSZ signal extraction techniques, and the pairwise statistical estimator is described in Sec. II C.

A. Datasets

In this work, we use the simulated kSZ maps from [147][1] (the Flender simulation) to train and test our machine learning model. The kSZ maps were generated to represent the signal expected from the passage through dark matter large-scale structures in a single realization from the Mira

---

[1]https://www.hep.anl.gov/cosmology/ksz.html.

Universe simulation suite [148] using the HACC (Hardware/Hybrid Accelerated Cosmology Code N-body simulation) framework [149]. The simulation adopts WMAP7 cosmological parameters [150] and is simulated with 3200^3 particles in a $(2.1 \text{ Gpc})^3$ volume and a mass resolution of $10^{10} M_\odot$. A friends-of-friends algorithm is used to identify dark matter halos. The simulation provides a full-sky halo catalog spanning redshifts $0 < z < 1$ and masses $10^{10} M_\odot < M < 3 \times 10^{15} M_\odot$.

The kSZ signal is simulated under three different models of the intracluster gas: Model 1 assumes baryons trace the dark matter at all scales (i.e. gas traces mass) and includes the diffuse (nonhalo) component. Model 2 follows an intracluster gas prescription from [91] and ignores the diffuse kSZ component. Model 3 (herein FL3) is a combination of Model 2 and the diffuse kSZ components from Model 1. As Model 3 is the most realistic kSZ model, it is the principal kSZ signal used in this work. The Flender kSZ maps are simulated at a resolution of $0.43'$ which is comparable to the recent CMB experiments such as the ACT data [151].

We also consider a second simulated kSZ map from [152][2] (the Websky simulation) to cross-validate the model that is trained on the Flender simulation. The kSZ maps in this case were generated from a lightcone realization of N-body with 6144^3 particles in a $(7.7 \text{ Gpc})^3$ volume that covers redshift $0 < z < 4.6$ over the full sky. The astrophysical effects and cosmological model considered in Websky are very similar to the Flender simulation such as star formation, feedback, and nonthermal pressure simulated using parametric models informed by hydrodynamical simulations. A primary distinction between the two simulations is that the resolution of the Websky simulation, $0.87'$, is lower than the Flender simulation.

We also test the feasibility of our model to anticipate CMB observations by simulating the detector noise and primordial CMB. We run CAMB [153] to generate the angular power spectrum of the primary CMB anisotropies with best-fit Planck cosmological parameters for a zero-curvature universe [154]. The instrument noise is simulated in accordance with the upcoming observation from Simons Observatory (SO) [155] at 145 GHz. Specifically, we simulate the instrument noise as white noise at a 6.3 μK-arcmin noise level. A Gaussian beam (FWHM = 1.4 arcmin) is also considered in accordance with the SO instrument beam at 145 GHz.

B. kSZ effect and signal extraction

The CMB temperature change resulting from the kSZ effect is given by

$$\frac{\delta T_{\text{kSZ}}}{T_0} = -\int_{\text{los}} \sigma_T n_e \frac{v_{\text{los}}}{c} dl, \tag{1}$$

where σ_T is the Thomson cross section, n_e is the electron number density, c is the speed of light, v_{los} is the line-of-sight peculiar velocity, and $T_0 = 2.726$ K is the average CMB temperature.

1. Aperture photometry

Aperture photometry is a filtering method that measures an average disk temperature by averaging the temperatures within the central disk and subtracting the average temperature in a surrounding ring annulus of equal area to mitigate potential background contamination. For a disk located at position $\hat{\mathbf{n}}$, the aperture photometry temperature can be written as

$$T_{\text{AP}}(\hat{\mathbf{n}}, \theta_{\text{AP}}) = T_{\text{disk}}(\hat{\mathbf{n}}, \theta_{\text{AP}}) - T_{\text{ring}}(\hat{\mathbf{n}}, \theta_{\text{AP}}), \tag{2}$$

where angular radius θ_{AP} is the filter scale. Note that a correction factor that takes into consideration the kSZ signal in the annulus being subtracted is necessary for aperture photometry to relate T_{AP} to the disk kSZ temperature [95].

2. Matched filter

Given a predefined template profile embedded under a noisy signal, a matched filter can be constructed to detect the presence of the template profile with minimum variance. For the kSZ signal template, we use the projected Navarro-Frenk-White (NFW) profile [156] τ which is written as

$$\tau(x) = \frac{A}{x^2 - 1} \begin{cases} 1 - \frac{2}{\sqrt{1-x^2}} \tanh^{-1} \sqrt{\frac{1-x}{x+1}} & 0 < x < 1 \\ 0 & x = 1 \\ 1 - \frac{2}{\sqrt{x^2-1}} \tan^{-1} \sqrt{\frac{x-1}{x+1}} & x > 1, \end{cases} \tag{3}$$

where $x = \theta/\theta_s$, and θ_s is the scale angle. In Fourier space, the matched filter can be written as

$$\Psi(k) = \sigma^2 \frac{\tau(k)B(k)}{P(k)}, \tag{4}$$

where $\tau(k)$ is the signal template profile in Fourier space, $B(k)$ is the instrument beam, and σ^2 denotes the filter variance that

$$\sigma^2 = \left[\int \frac{|\tau(k)B(k)|^2}{P(k)} \frac{d^2k}{(2\pi)^2} \right]^{-1}, \tag{5}$$

where $P(k)$ is the noise power spectrum.

We implement the matched filter in Fourier space by filtering a postage cutout centered at the cluster, and the

---

[2]https://mocks.cita.utoronto.ca/index.php/WebSky_Extragalactic_CMB_Mocks.

matched filtered temperature, $T_{\mathrm{MF}}(\hat{\mathbf{n}}, \theta_s)$, is measured from the filtered cutout.

Matched filters are commonly used to determine the central peak amplitude; however in this work, we implement the matched filter differently. Following [95], a matching filter, as opposed to aperture photometry, can directly extract an unbiased disk kSZ temperature using a carefully calibrated NFW signal template profile parametrized solely with the scale radius θ_s. Specifically, the matched filter is built in a way similar to minimizing the mean sum of squares (MSS) between the difference of the matched filtered amplitude and the true amplitude of the kSZ radial profile at each position r within a certain angular scale θ, denoted as

$$\mathrm{MSS} = \frac{1}{N} \sum_{r=0'}^{r=\theta'} (T_r - T_{\mathrm{MF},r})^2, \tag{6}$$

where T_r is the signal amplitude at r. This is done by tuning the filter's signal temple profile τ through θ_s. In this work, θ_s is set by selecting a subsample of 20,000 halos from the sample at random and searching for the value of θ_s that minimizes the difference of the average disk temperature between the true and matched filtered estimate for this subsample.

C. Pairwise estimator

The pairwise velocity of clusters given each of their line of sight peculiar velocity v_i, is given by [50]

$$\hat{\mathbf{V}}(r) = -\frac{\sum_{ij}(v_i - v_j)c_{ij}}{\sum_{ij}c_{ij}^2}, \tag{7}$$

where the sum is over all cluster pairs with separation $r_{ij} = |r_i - r_j|$ that fall in a radial bin centered around distance r. c_{ij} is the geometric weight projected along the line of sight given by

$$c_{ij} = \hat{\mathbf{r}}_{ij} \frac{\hat{\mathbf{r}}_{\mathbf{i}} \dot{-} \hat{\mathbf{r}}_{\mathbf{j}}}{2} = \frac{(r_i - r_j)(1 + \cos\theta)}{2\sqrt{r_i^2 + r_j^2 - 2r_i r_j \cos\theta}}, \tag{8}$$

where r_i, r_j are the comoving distance, and θ is the angle between vectors r_i and r_j.

Replacing the velocity with the kSZ temperature of each cluster gives the pairwise kSZ momentums estimator and is written as [64]

$$\hat{\mathbf{p}}(r_i, z_i) = -\frac{\sum_{ij}(\delta T_i - \delta T_j)c_{ij}}{\sum_{ij}c_{ij}^2}, \tag{9}$$

where

$$\delta T_i(\hat{\mathbf{n}}_i, z_i) = T(\hat{\mathbf{n}}_i) - \bar{T}(\hat{\mathbf{n}}_i, z_i, \sigma_z). \tag{10}$$

$T(\hat{\mathbf{n}}_i)$ is the kSZ signal, that can be T_{disk}, T_{AP}, or T_{MF}. The subtraction of $\bar{T}(\hat{\mathbf{n}}_i, z_i, \sigma_z)$ given by

$$\bar{T}(\hat{\mathbf{n}}_i, z_i, \sigma_z) = \frac{\sum_j T(\hat{\mathbf{n}}_i)w(z_i, z_j, \sigma_z)}{\sum_j w(z_i, z_j, \sigma_z)} \tag{11}$$

mitigates a potential redshift-dependent systematic that could contaminate the pairwise signal where

$$w(z_i, z_j, \sigma_z) = \exp\left(-\frac{(z_i - z_j)^2}{2\sigma_z^2}\right), \tag{12}$$

with $\sigma_z = 0.01$, following [157].

We measure the pairwise signal in 15 comoving pair separation bins of equal width between 0 and 150 Mpc, and four additional bin edges on 200, 250, 315, and 395 Mpc in accordance with [69]. The covariance of the pairwise signal is estimated with the bootstrap resampling analysis during which we randomly replace the peculiar velocities (or temperature decrements), v (or δT), of galaxy positions with replacement. We then estimate the covariance matrix, C_{ij}, by repeating this process 1000 times and then calculating the covariance of the list of pairwise momentum estimators calculated from each of the new samples.

III. FEATURE ENGINEERING AND LEARNING MODEL

In this section, we present the machine learning model used to recover the pairwise peculiar velocity. Machine learning, or statistical learning, is a field of study that utilizes a large amount of data for the computer to learn and understand the inherent patterns and interconnections in data and thereby make predictions and inferences without knowledge of the underlying physical laws or processes. The inputs that machine learning models utilize for training and inference to make predictions are known as model features. Model features are important as the accuracy of a machine learning model depends on the precise selection and composition of the features. The process of extraction and transformation of the features from raw data is called feature engineering. The motivation of feature engineering is to use these features to enhance the quality of the inference results as opposed to simply providing raw data to the machine learning algorithm.

One central goal of predictive modeling is to build a model that can make accurate predictions on new data that the model has never seen. In machine learning, data leakage refers to the situation when the model training process uses information that would not be expected to be available at the time of prediction. This happens when the model includes information on test data in the training process, and as a result, it achieves high performance while testing but might perform poorly for unseen data.

In this work, we use machine learning to predict the peculiar velocity of each individual cluster using features measured from the kSZ signal. We consider a sample of 300,000 cluster-mass halos randomly extracted from the Flender simulations spanning $0 < z < 1$ and $10^{13} M_\odot < M_{500} < 3 \times 10^{15} M_\odot$, and associated kSZ maps, for training and testing purposes. To avoid data leakage, we use a train-test split procedure in which we train our model with 70% of the sample (the "training sample" of 210,000 halos) and use the remaining 30% of them ("Flender test sample I" of 90,000 halos) [142,158], which the training model has never seen, for testing and model validation. In this way, we simulate how the model would perform on new, unseen data.

For velocity reconstruction when detector noise and primordial CMB are present, we examine a second testing dataset from the Flender data ("Flender test sample II") with a higher minimum mass threshold, as will be motivated in the next section. This sample consists of 86,866 Flender simulated halos $0 < z < 1$, $8 \times 10^{13} M_\odot < M_{500} < 3 \times 10^{15} M_\odot$ drawn separately from test sample I from the Flender simulation.

To facilitate the assessment of the approach beyond purely training and testing on a single simulation dataset, we consider a dataset distinct from the Flender data, using the Websky simulations. Specifically, we use a randomly extracted sample of 300,000 cluster-mass halos spanning $0 < z < 1$ and $10^{13} M_\odot < M_{500} < 3 \times 10^{15} M_\odot$ split 70%/30% into distinct training and testing subsamples. We also consider a fourth separate test sample for analysis with primary CMB and noise of 90,000 Websky simulated halos (the "Websky test sample"), with $0 < z < 1$ and $8 \times 10^{13} M_\odot < M_{500} < 3 \times 10^{15} M_\odot$.

These studies, and the use of ~90,000 cluster test samples in our analysis, are motivated in the anticipation of baryon oscillation spectroscopic survey (BOSS) and dark energy spectroscopic instrument (DESI) data

providing spectroscopic redshifts for photometrically selected galaxy clusters in the SDSS survey. Specifically, we use clusters of the Wen Han Liu (WHL) catalog [159] as a guide. This catalog covers a redshift range of $0.05 < z < 0.75$ with a minimum mass of $5 \times 10^{13} M_\odot$ and consists of 158,103 galaxy clusters with photometric redshifts. Spectroscopic redshifts from BOSS are available for 121,103 (77%) of the galaxy cluster sample, and of these 102,033 have $M_{500} > 8 \times 10^{13} M_\odot$, with the expectation that DESI will provide additional spectroscopic redshifts. The full-sky Planck temperature maps would naturally have the same sky coverage as this catalog but the Planck map does not have enough resolution to do a matched filter [95], a key process for our model that will be discussed below. On the other hand, the ACT maps with higher resolution do allow a matched filter approach, but they have a smaller sky coverage overlap with the WHL catalog. The motivation for developing the machine learning approaches here is to apply the approach to a spectroscopic catalog similar to WHL in combination with an upcoming multifrequency CMB data, such as from SO, CMB-S4, and CCAT, that is of higher resolution and higher sensitivity than Planck but will also survey a significant fraction of the sky.

For each cluster, we create seven features for training and testing purposes, summarized in Fig. 1. Given the peculiar velocity is directly proportional to the kSZ temperature, as in (1), we consider the kSZ temperatures as our key features: T_{Disk} in the training data (and T_{MF} in the testing data) and T_{AP}. For T_{disk} (training dataset) and T_{MF} (test dataset) the mean and the standard deviation of the temperatures of the pixels within the aperture are used as input features.

These temperature features are directly measured from the observation map at the location of each cluster. For the T_{disk} and T_{MF} data we consider two different aperture sizes for the disk kSZ temperatures: where the signal-to-noise ratio (SNR) peaks at 2.1′ [95,147] and also a 1.3′ aperture
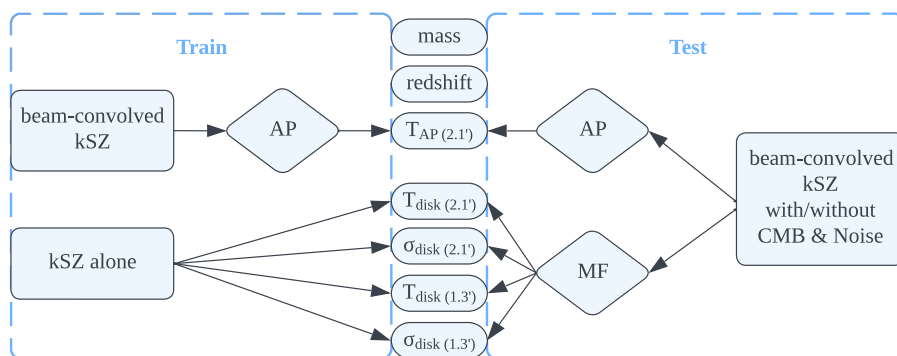

FIG. 1. Feature engineering process and the features used for our machine learning model for (left) the training and (right) the testing datasets. We consider seven features for each halo object, i.e. the mass and redshift, z, of each halo, the mean and standard deviation of disk kSZ temperatures, T_{disk} and σ_{disk}, and the aperture photometry (AP) temperature, T_{AP}, within different angular scales (i.e. 1.3 and 2.1 arcminutes) from the halo center. For the test data, we use a matched filter (MF) approach to estimate the disk temperature properties and consider scenarios with both pure kSZ and also data which includes primary CMB and detector noise.

size that focuses more on the central peak kSZ signals. We also include a spectroscopic redshift and mass estimate for each halo in the set of features. In this work, we use the cluster mass estimate itself as the feature rather than a feature related to the mass proxy. Observationally, the cluster mass estimate could be inferred from one or more potential observable proxies of cluster mass, such as using richness [160,161], x-ray [162–164], and weak lensing observations [165–167]. We discuss the potential effect of inaccuracies in the mass estimate on the modeling in more detail in Sec. IV C 2.

For training purposes, one can choose to train the model with either noise-free or noisy data. Noise-free data enables the machine learning models to concentrate on the underlying structures and patterns we seek to extract. We have considered models trained on noise-free data and data including primary CMB and noise but find the latter perform significantly less well in reconstructing the features. Given this, and following the approach frequently employed with the machine learning technique, as listed in Sec. I, we train our model on the kSZ signal alone, without primary CMB and noise. Aperture photometry features are measured from the beam-convolved kSZ map, and the disk features are measured directly from the unfiltered pure kSZ map as summarized in the left of Fig. 1. Note that we convolve the kSZ map with a beam for AP to account for the beam effect in real data analysis.

We test our training model in two scenarios, one with a pure kSZ signal and the other when CMB anisotropies and detector noise are present. Similar to the training process, aperture photometry features are again measured from the kSZ map with beam convolution. Different from the training process, the disk temperature features are measured using the matched filter from the kSZ map with beam convolution as summarized in the right of Fig. 1. The intent of using the matched filter and the beam-convolved kSZ map during the testing process is to mimic the signal extraction process when analyzing the real observation data. When adapting to real observations (e.g. when detector noise and CMB anisotropies are present), following [81], the CMB-dominated signals above 30 arcminutes are filtered out, and the small-scale kSZ-related signals below 15 arcminutes are preserved, before we measure kSZ temperatures with AP and MF.

This work aims to recover the pairwise peculiar velocity of the cluster sample through a one-step machine learning inference of estimating the individual peculiar velocity for each cluster in the testing sample. We train the model to predict a peculiar velocity for each cluster based on the seven features described above. This process is akin to a regression problem, making an inference based on the numerical model features. We adopt a gradient boosting algorithm, LightGBM [168],[3] that has demonstrated

stability and out-performance in making regression inferences with a set of one-dimensional numerical model features. The gradient boosting model is in the form of an ensemble of weak learners, where a weak learner usually means a simple model that is slightly more accurate than random prediction. The model is trained sequentially, and in each iteration, a new weak learner is trained and added to the ensemble to correct for the previous model. In this process, several weak learners are combined into strong learners, in which each new weak learner is trained to minimize the prior model's loss; specifically the mean squared error is used here.

One of the main challenges during the training process is to avoid overfitting while achieving model accuracy and convergence. Below, we will briefly discuss our choice of two of the main hyperparameters of the LightGBM model, i.e. the learning rate and the number of iterations. The learning rate determines how much the model corrects for the error at each iteration toward achieving a minimum of the loss. While a learning rate that is too low will take a long training time to converge and stuck in an unwanted local minimum, a learning rate that is too high will cause the learning to bounce over minima and never converge. The number of iterations, on the other hand, controls how many times the model corrects for the error. Too many iterations will usually cause an overfitting of the training data that the model provides precise forecasts for training data, but not for unseen data, while too few iterations will not let the model learn sufficiently. To achieve higher accuracy, a common strategy is to let the model learn over more iterations and have a low learning rate. This is equivalent to finding a good combination of iterations and learning rates. In this work, we choose a learning rate value of 0.03 which is commonly used in many applications. We find that the model begins to converge after 3,000 iterations and starts to overfit after 5,000 iterations. We stopped our models at 3,600 iterations to avoid overfitting while also achieving high accuracy. The other hyperparameters of the model are determined using a grid search cross-validation technique running Scikit-Learn's GridSearchCV [169]. Cross-validation is a technique that splits the dataset into subsets, or folds, with the purpose of using each fold as a validating set and the remaining folds for training. We evaluate the model performance on the validation set, where each fold is used exactly once as the validation set during the iterations of this process. The results from each iteration are then averaged to obtain the performance of the model. In the GridSearchCV algorithm, a grid of hyperparameters with possible values is first defined. Cross-validation is then used to train the model and assess its performance for each combination of hyperparameters as it explores the hyperparameter space. In this way, it helps in determining the ideal hyperparameter combination that yields the best model performance.

In assessing the efficacy of the machine learning velocity reconstruction, we utilize the following statistics.

---

[3]https://lightgbm.readthedocs.io/en/stable/.

We measure the strength of linear correlation between variables using the Pearson correlation coefficient. For a pair of random variables (X, Y), the correlation coefficient can be written as

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}, \quad (13)$$

where x_i, y_i are the individual sample data points, n is the sample size, and $\bar{x}, \bar{y}$ are the sample mean.

The SNR for the machine learning model predicted pairwise velocity signal, is evaluated as

$$\text{SNR} = \sqrt{\sum_{ij}\hat{\mathbf{V}}_{i,\text{pred}}C_{ij}^{-1}\hat{\mathbf{V}}_{j,\text{pred}}} \quad (14)$$

with $\hat{\mathbf{V}}_{i,\text{pred}}$ the predicted pairwise velocity estimator using the machine learning model for the ith pair separation bin, and C_{ij}^{-1} is the inverse of the covariance matrix estimated off a bootstrap of the dataset as discussed in Sec. II C.

We determine how well the pairwise velocity predicted by the model fits to that obtained directly from the true velocities in the simulation with the χ^2,

$$\chi^2 = \sum_{ij}\Delta\hat{\mathbf{V}}_i C_{ij}^{-1}\Delta\hat{\mathbf{V}}_j, \quad (15)$$

with $\Delta\hat{\mathbf{V}}_i = \hat{\mathbf{V}}_{i,\text{true}} - \hat{\mathbf{V}}_{i,\text{pred}}$, where $\hat{\mathbf{V}}_{i,\text{true}}$ is the true pairwise velocity estimator for the test data. The best-fit model is achieved with the minimum χ^2.

It is important to note how the χ^2 should be interpreted in this analysis. Usually, we are comparing the fit of a theoretical model to observational survey data which can be viewed as a random realization of that model, modulated by cosmic variance and measurement uncertainties. In that case, a good fit to the data is characterized by a reduced χ^2, χ^2 per degree of freedom, value of around one as long as the covariance has been accurately estimated. In this analysis, however, we are comparing how well the machine learning model predicted velocity statistics for the test data compared with that true velocity correlation. In the limit of perfect reconstruction, the χ^2 would be zero. The degree of variation between the predicted and true signals is affected by the intrinsic uncertainties in the *training* model data set and the efficacy of the training algorithm to recreate the test dataset. The training data set used here is over twice the size of the test data and will have commensurately smaller statistical uncertainties than the test data. For the model to work well, the prediction uncertainties, $\Delta\hat{\mathbf{V}}_i$, should be substantially smaller than the statistical uncertainties in the test data, characterized by the covariance C_{ij}. As such, the prediction from a well-performing machine learning algorithm should have a χ^2 per degree of freedom that is well below one.

IV. ANALYSIS AND RESULTS

In Sec. IV A we consider the relation between the model features and peculiar velocities, and test the performance of the machine learning model to recover the individual halo velocities. In Sec. IV B we consider the pairwise velocity predicted using test data with solely the kSZ signal and, more akin to real observations, when detector noise and primordial CMB are included. We discuss the effects of systematic errors from cluster miscentering and mass misestimation in Sec. IV C. Finally, in Sec. IV D we consider the performance of the machine learning model when trained and tested using different kSZ simulations.

A. Individual halo velocity reconstruction

In advance of developing the machine learning model, we first establish how the key model features relate to the halo's peculiar velocity. In Fig. 2, we show the correlation coefficient between two of the key features, T_{MF} (2.1′) and T_{AP} (2.1′), and the true peculiar velocity of each halo, v_t, as a function of the halo mass for four kSZ datasets that include solely the kSZ, and combinations of the primary CMB and detection noise.

When just the kSZ is included the correlation of the MF and AP temperatures with the true velocity are both high and consistent with one another. For the highest mass halos, the correlation is nearly perfect. For lower mass halos the correlation falls, to just above 0.6 for masses of $10^{13}M_\odot$. This is consistent with the kSZ signal being smaller for lower mass clusters and thus the relative contribution from other kSZ sources along the line of sight, but unrelated to the cluster mass halo, becomes more important, creating the noisier signal.

When detector noise and/or primordial CMB are included in the test data, there is, as is to be expected, a weaker correlation between the measured model features and the true peculiar velocity. We find that, in general, kSZ signals extracted with matched filter have a tighter correlation with the cluster peculiar velocity than those extracted with aperture photometry. The inclusion of the primary CMB has the largest impact on weakening the correlation for both signal extraction methods.

Using a correlation of around 0.5 as a guiding threshold [170], we find that an overall correlation of 0.46 between T_{MF} and the true peculiar velocities is maintained for a test sample with minimum mass threshold $M_{500} = 8 \times 10^{13}M_\odot$. By contrast, a sample needs to have $M_{500} > 3 \times 10^{14}M_\odot$ for T_{AP} to retain a correlation of 0.5 of higher. We seek to include as many halos as possible to minimize statistical uncertainties and maximize performance. Given the poorer correlation of the AP features, and the restrictive mass threshold this would impose when the detector noise and primordial CMB are present, we exclude the AP temperature features (i.e. T_{AP}) and preserve only the MF-filtered kSZ temperature features (i.e. T_{MF}).
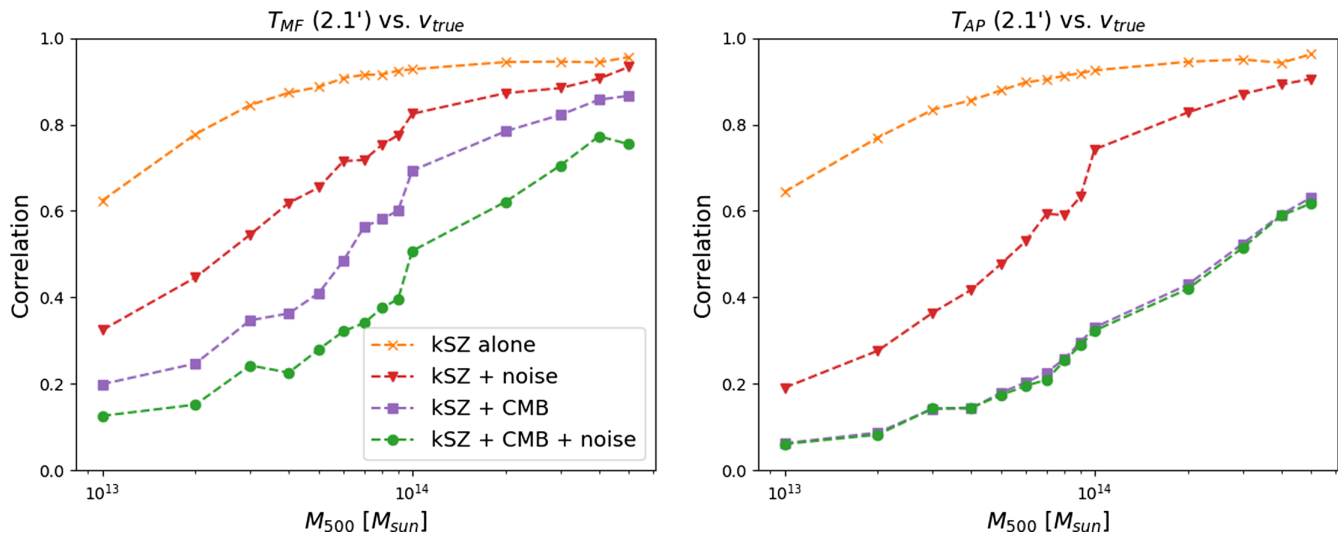
FIG. 2. Correlation between two of the key features the matched filter, T_{MF} (2.1′) (left) and aperture photometry T_{AP} (2.1′) (right), and the true peculiar velocity of each halo in Flender test sample I as a function of halo mass, when the test data includes just pure kSZ data (orange, cross), kSZ plus detector noise (red, triangle), kSZ plus primary CMB (purple, square) and kSZ, primary CMB and detector noise (green, circle). Note large-scale signals are filtered out before employing the MF and AP filters.
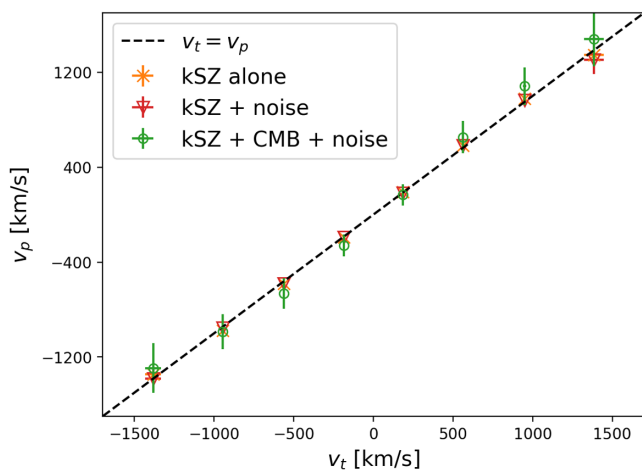
FIG. 3. Comparison of the machine learning predicted peculiar velocities, v_p, to the true peculiar velocities, v_t, for the Flender test sample II halo catalog, when the test data includes just kSZ (orange, cross), kSZ and detector noise (red, triangle), and kSZ and both primary CMB and detector noise (green, circle). The mean signal binned by v_t is shown with 1σ standard errors.

FIG. 4. Correlation between the predicted and true velocities, $|v_p|$ and $|v_t|$, as a function of halo mass for Flender test sample II, when the test data includes only kSZ (orange, cross), kSZ + detector noise (red, triangle), kSZ + primary CMB (purple, square) and kSZ, primary CMB and detector noise (green, circle).

Motivated by this, we create a second halo testing sample, Flender test sample II with $M_{500} > 8 \times 10^{13} M_\odot$, containing 86,866 objects distinct from those used in the training model. Figure 3 shows the comparison of the predicted velocity using the machine learning model with only the matched-filtered kSZ temperature inputs from the test sample to the true velocity for this sample. We find that the application of the machine learning model allows us to recover the peculiar velocities without bias for both kSZ-only signals and when the detector noise and primordial CMB are added.
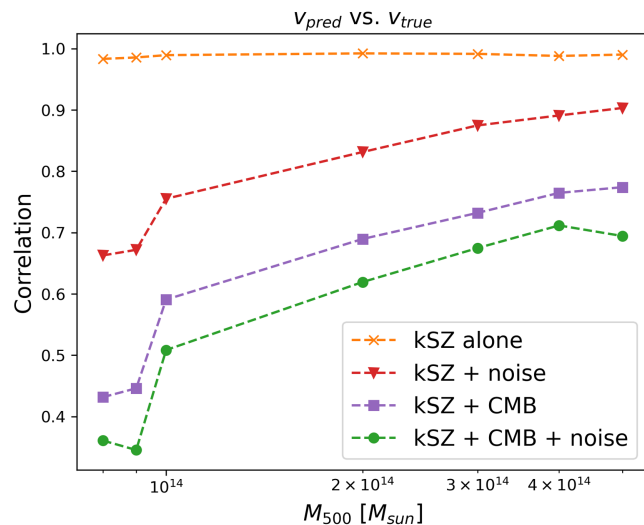
In Fig. 4 we show the correlation between the halo velocity predicted by the machine learning model, $|v_p|$, and the true velocity, $|v_t|$, as a function of mass for Flender test sample II with four different kSZ datasets: kSZ alone, and when noise and primary CMB are included. We find that the trends in correlation are consistent with those observed between kSZ temperatures and the true velocity shown in Fig. 2. The correlation for the overall correlation across all mass bins is 0.43, reinforcing that $8 \times 10^{13} M_\odot$ is a reasonable minimum mass threshold for the analysis.

In the rest of the paper, we will focus the analysis on the samples with $M_{500} > 8 \times 10^{13} M_{\odot}$ and also exclude the AP temperature features as discussed above.

B. Pairwise peculiar velocity reconstruction

Using Flender test sample II, we start by testing how well we can recover the pairwise peculiar velocity using the pure kSZ signals in the test set model features. Figure 5 shows the pairwise velocity statistics associated with the predicted individual halo velocities described in Sec. IV A. We find that the predicted statistic, $\hat{\mathbf{V}}_p$, recovers the true estimator, $\hat{\mathbf{V}}_t$, well within the statistical errors estimated from bootstrapping the sample. Comparing the reconstructed pairwise signal to the true one gives a best-fit $\chi^2 = 1.6$ for 19 bins covering all the pair separations to 395 Mpc.

We then test the feasibility of our machine learning model to extract the estimator from more realistic observations, when both primary CMB and detector noise are included which all together can be orders of magnitude larger than the kSZ signal alone.

We generate 10 realizations of both the detector noise and primary CMB as described in Sec. II A. We generate independent primary CMB and detector noise realizations from each of their individual power spectra without kSZ included. Thus, there will be no correlation between these 10 realizations. For each of these 10 realizations, primary CMB is added to the Flender model 3 map before a beam convolution of FWHM = 1.4′, after which the detector noise is included. As mentioned in Sec. III, the large-scale

FIG. 5. Top: pairwise peculiar velocity estimator calculated for the true velocity ($\hat{\mathbf{V}}_t$) (blue, square) and predicted velocity ($\hat{\mathbf{V}}_p$) with pure kSZ signals (orange, cross), the mean of 10 realizations when detector noise and primary CMB are included (green, circle), and the null test of using just the CMB and detector noise (and no kSZ) (red, star) for Flender test sample II. Bottom: difference between prediction and expectation (e.g. $\Delta\hat{\mathbf{V}} = \hat{\mathbf{V}}_p - \hat{\mathbf{V}}_t$) scaled relative to the respective uncertainty, σ, for each sample.

CMB-dominated signals are filtered out before the model features are created for each halo object. The mean pairwise peculiar velocity estimator across the 10 realizations is calculated and errors are estimated by conducting a bootstrap on one realization.

The results are shown in Fig. 5. We first conduct a null test in which the mean signal is obtained for purely the 10 maps of detector noise and primary CMB with no kSZ. The reconstructed signal is consistent with $\hat{\mathbf{V}} = 0$. When the kSZ signal is included, the machine learning model recovers an unbiased estimate of the pairwise peculiar velocity for cluster pair separations below 200 Mpc. For the 16 bins with $r < 200$ Mpc, we find a best-fit $\chi^2 = 9$ for the mean pairwise signal over the 10 realizations and SNR = 11. For separations greater than 200 Mpc, the intrinsic amplitude of the pairwise signal is very small. While the signal is able to be recovered for the pure kSZ inputs, once detector noise and primary CMB anisotropies are included, and dwarf the kSZ signal in the test features, the machine learning algorithm is unable to recover a robust prediction of these low amplitude signals; the inferences are systematically biased for these large pair separations. For this reason, the analyses in the rest of the paper focus on the pairwise separations below 200 Mpc.

The machine learning inference presents an alternative to the approach in which the pairwise velocity estimator is reconstructed by combining the pairwise kSZ momentum with associated optical depth estimates. The pairwise kSZ momentum estimator calculated from the MF temperature features has SNR = 11. If we consider an optical depth estimate using tSZ or x-ray observations with an assumed 18% uncertainty [93,94], we would estimate a pairwise velocity reconstruction with SNR = 7, compared to SNR = 11 using the machine learning method in this work.

To understand how the model learns to break the degeneracy between velocity and optical depth from the kSZ temperature, we have tested the feature importance by excluding one feature at a time and evaluating the model performance after the exclusion of each feature. Note that each time, we retrain and test the model using solely the selected features. In Fig. 6, we show the 1σ standard errors for $40 < r < 80$ Mpc for the pairwise velocity statistics derived from the predicted peculiar velocity when each feature is excluded from the model. The exclusion of the halo mass in the training data gives the largest increase in the 1σ errors on pairwise velocity statistics. This suggests that the mass information plays an important role in velocity estimation model consistent with the expectation that the mass feature serves as a proxy for the optical depth estimate and can therefore improve the model accuracy through an implicit cluster $M_{500} - \tau$ relation.

C. Effects of uncertainties in the test data

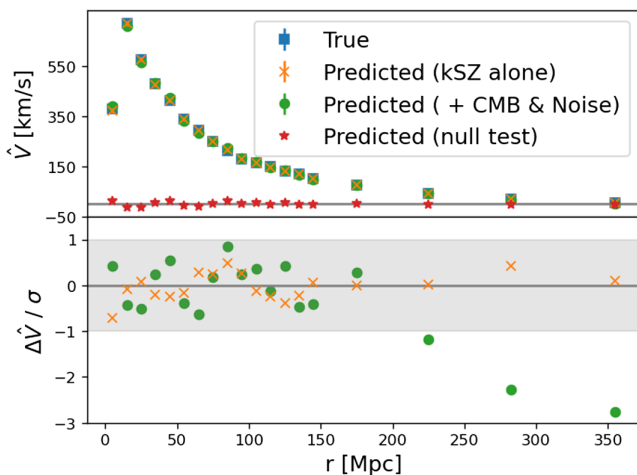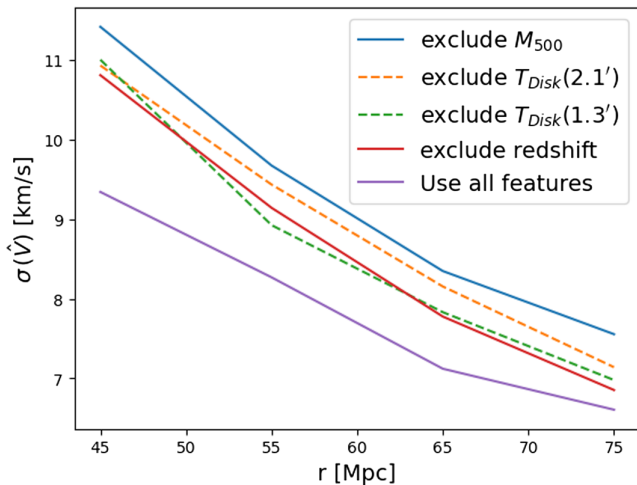In this section, we discuss how the machine learning model performs when accounting for potential systematic

FIG. 6. Sensitivity of the 1σ standard errors on the pairwise peculiar velocity momentum derived from the machine learning algorithm when trained and tested without one of the features included. The cluster feature set that excludes M_{500} (blue), $T_{\text{Disk}}(2.1')$ (orange), $T_{\text{Disk}}(1.3')$ (green), redshift (red), and no feature exclusion (purple) are shown for Flender test sample II.

FIG. 7. Fractional difference $(\Delta\hat{\mathbf{V}}/\hat{\mathbf{V}}_t)$ between the estimator $(\hat{\mathbf{V}}_t)$ derived from the true velocity and the prediction $(\hat{\mathbf{V}}_p)$ derived from the machine learning model for Flender test sample II when the test data is kSZ alone. Predictions when the test model features are calculated assuming cluster centers modeled from the true center (blue, square), from displacements based on Saro *et al.* (red, star), and when the peak tSZ signal is used as the cluster center (orange, cross). The mean signal deviation across the whole sample is also shown for each case (dashed line).

errors arising due to cluster miscentering and mass mis-estimation.

1. Miscentering bias

While in simulations we can identify the gravitational center of a cluster halo, observationally this is not the case. Typically, following the "central galaxy paradigm" [171], the most massive, and brightest, galaxy in the cluster will be the most centrally located. As such, the brightest galaxy identified in a cluster will commonly be used as the proxy to define the center of a cluster. The position of the brightest galaxy may not always be the gravitational potential minimum of the cluster, however, and the brightest galaxy may also not be the closest cluster member to the center. Roughly 30% of cluster mass systems, in halos 10^{13}–$10^{14.5} M_\odot$, are found to not have the brightest galaxy as the central most member [172–176]. Hence, the cluster center estimated from the brightest galaxy can be biased from the true cluster center. This can lead to biases in the calculated pairwise signal kSZ [177] as well as other cluster measurements such as in the x ray [178].

Saro *et al.* [175] compared the cluster center derived from the tSZ profile to the position of the brightest galaxy and found that 63% of clusters have a miscentering of $\sigma_0 = 0.07h^{-1}$ Mpc, and the rest have a miscentering of $\sigma_1 = 0.25h^{-1}$ Mpc. This miscentering can be expressed in terms of a bimodal Gaussian distribution,

$$P(x) = 2\pi x\left(\frac{\rho_0}{2\pi\sigma_0^2}\exp\left\{-\frac{x^2}{2\sigma_0^2}\right\} + \frac{1-\rho_0}{2\pi\sigma_1^2}\exp\left\{-\frac{x^2}{2\sigma_1^2}\right\}\right),$$
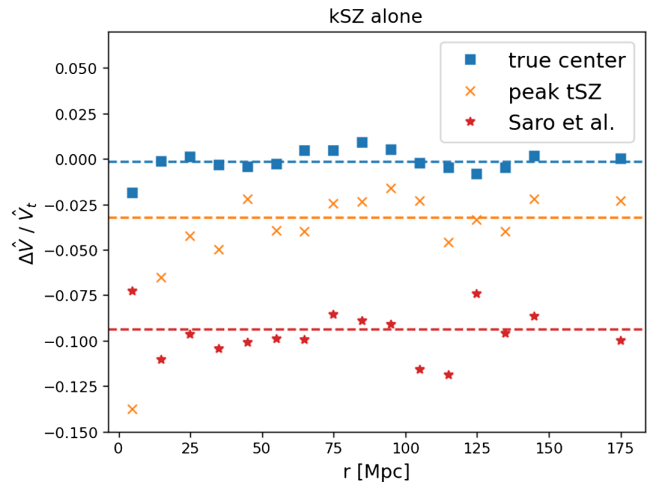
(16)

with $\rho_0 = 0.63$.

Following Saro *et al.*, we create a modified version of Flender test sample II in which the test model features are created using halo centers which have a random offset applied using (16). In Fig. 7, we present the resulting predicted pairwise velocity statistics when test features are just for the kSZ alone (with no detector noise and CMB included). Miscentering at the level of the Saro model reduces the predicted pairwise peculiar velocity $(\hat{\mathbf{V}})$ by 9% relative to that derived using the true centers to derive the test data model features.

To alleviate the suppression of the amplitude of the pairwise velocity estimator caused by miscentering, we consider an alternative centering approach in which the center is located using the tSZ signal profile. The process is summarized as follows. We first assign each cluster a new center following the Saro *et al.* model. We then use the tSZ map and search around the new center to locate the peak tSZ signal. By following these steps we reflect the practical process that we would do with real data in which the initial guess of the center is determined by the Saro *et al.* model. We finally assign a center for each cluster based on the location of the identified peak tSZ signal and calculate model features using that assumed center.

Using the tSZ-derived center, the predicted pairwise peculiar velocity $(\hat{\mathbf{V}})$ is reduced by only 3% on average; the associated results are shown in Fig. 7. We conclude that using the cluster center derived from the peak tSZ signal could alleviate the suppression of the amplitude of the pairwise velocity estimator caused by miscentering if the optically brightest galaxies are purely used as the cluster center proxy.
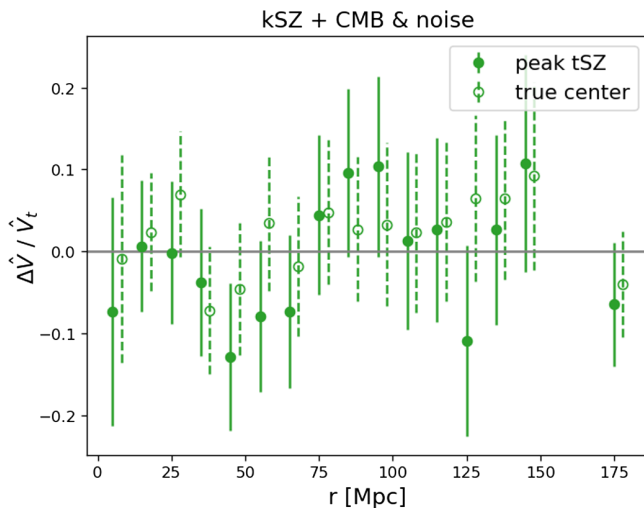
FIG. 8. Fractional difference ($\Delta\hat{\mathbf{V}}/\hat{\mathbf{V}}_t$) between the pairwise velocity estimator from the true velocity ($\hat{\mathbf{V}}_t$) and that from the machine learning prediction ($\hat{\mathbf{V}}_p$) derived from kSZ and a single realization of the primary CMB and detector noise for Flender test sample II when the location of the peak tSZ signal is used as the proxy for the cluster center (filled circle) and the true cluster center from the simulations is used (empty circle). 1σ error bars estimated from a bootstrap of the test data are shown.

In Fig. 8, using one realization of kSZ, CMB, and noise, we compare the machine learning pairwise $\hat{\mathbf{V}}$ prediction derived from the peak tSZ center model to the true $\hat{\mathbf{V}}$ derived from the true halo velocity provided in the catalog. We find that the inclusion of detector noise and CMB anisotropies does not additionally suppress the amplitude from the miscentering and that the offset due to miscentering is smaller than the sample's statistical errors. A best-fit χ^2 of 11 is found for the predicted pairwise signal from the single CMB and noise realization, relative to the underlying true pairwise velocities when the centers are derived from the tSZ peak. This just is slightly larger than the best-fit $\chi^2 = 10$ when the true cluster centers are used for the test feature for the same noise and CMB realization.

2. Mass scatter bias

In addition to uncertainties related to the location of the cluster center, the cluster mass is also a model feature that will need to be estimated from observational proxies. For a galaxy survey, a common approach, as used in [69,94], is to estimate the stellar mass from the luminosity of the brightest galaxy using the mass-to-light ratio [179–182], and then the halo mass is derived from the stellar-to-virial mass relation, $M_\star - M_{\rm vir}$ [180]. Other approaches to mass estimation include using the mass-richness relation, and calibration from combinations of x-ray, weak lensing, and tSZ observations. Any uncertainty in these relationships could lead to added dispersion or bias between the halo mass estimated as the input feature and the true halo mass.

We consider a scatter in the logarithmic mass estimate of 20% for each halo mass, reflective of mass uncertainty estimates using mass-richness and weak lensing calibration approaches (e.g. [183–186]). We use the Flender test sample II and augment it with a random subsample of lower mass halos down to $10^{13} M_{\rm sun}$. Using the sample size of Flender test sample II as a reference, the attached subsample is selected with number distribution following the halo-mass function of the Flender simulation. In this way, we avoid attaching too few or too many lower-mass halos, and the combined subsample can have the same halo mass distribution as the whole catalog. For each cluster, we assign a new mass based on relation $\ln M_{\rm new} = \ln M_{500} + \Delta$, where Δ is drawn from a Gaussian distribution with zero mean and standard deviation $\sigma = 0.2$.

We first note that introducing this uncertainty has implications for the minimum mass threshold imposed to create the sample. The scattering results in some clusters in the original sample with true masses $M_{500} > 8 \times 10^{13} M_\odot$ being excluded based on the mass estimate; we find 87% of the original sample objects are retained in the new sample. The new "augmented Flender test sample II" will also include some clusters with true masses $M_{500} < 8 \times 10^{13} M_\odot$ for which the inferred mass now exceeds the threshold. Because the cluster mass function decreases as one goes to larger masses, we find that the overall test sample size, with the scatter included, would increase by 8% as more lower-mass halos will move above the threshold than higher-mass ones will fall below. Given our model takes the cluster mass as an input feature, we start by testing how the mass scatter would affect the model prediction with pure kSZ signals when each cluster is assigned the "wrong" mass and this is used as an input feature.

In Fig. 9, we compare the predicted $\hat{\mathbf{V}}$ with the true $\hat{\mathbf{V}}$ for the new augmented Flender test sample II. We find that mass scatter reduces the predicted pairwise signal by 3% on average for the new sample. This is understandable given the sample with the mass scatter will include lower mass halos, that had previously fallen below the mass threshold, for which the mass is overestimated. As such, with the higher wrong input mass as an input feature along with the kSZ temperature, with an amplitude consistent with its true lower mass, the machine learning model will predict a smaller absolute velocity on average.

In Sec. IV A, we demonstrated that the correlation between velocity and the input features falls below 0.5 for cluster samples with $M_{500} < 8 \times 10^{13} M_\odot$ when primary CMB and detector noise are included. Given the halos in our sample selected based on mass estimates incorporating scatter have true mass $M_{500} < 8 \times 10^{13} M_\odot$, we want to consider the effect of mass scatter when these lower mass halos and primary CMB and detector noise are present in the input features. In Fig. 10, we find that the mass scatter does produce a systematic ∼3% reduction in the estimated

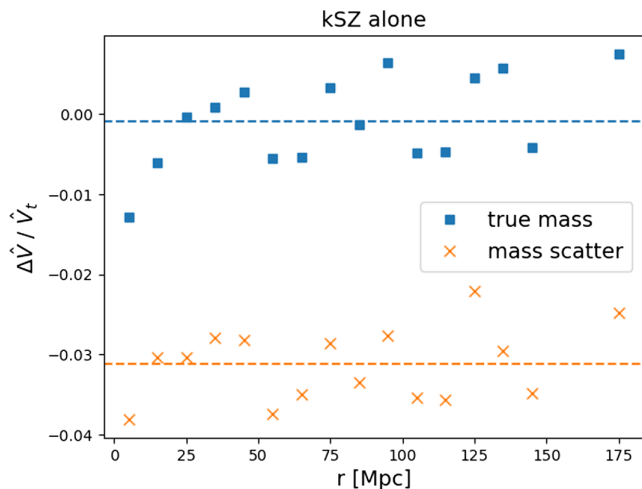FIG. 9. Fractional difference ($\Delta \hat{\mathbf{V}}/\hat{\mathbf{V}}_t$) between the true estimator ($\hat{\mathbf{V}}_t$) derived from the true velocity and the prediction ($\hat{\mathbf{V}}_p$) derived from the machine learning model for pure kSZ test data for the augmented Flender test sample II. The predicted velocity when the true cluster masses are used as input features (blue, square) is shown along with that when the mass features incorporate scatter to reflect misestimation (orange, cross). The mean fractional difference (dashed line) is also given for each sample.
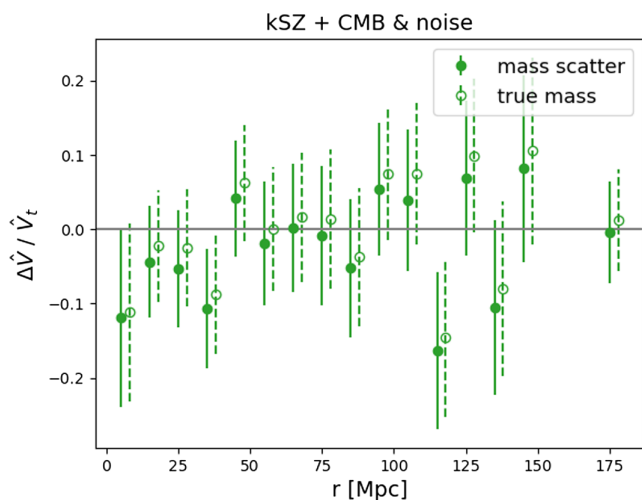

FIG. 10. Fractional difference ($\Delta \hat{\mathbf{V}}/\hat{\mathbf{V}}_t$) between the pairwise velocity estimator from the true velocity ($\hat{\mathbf{V}}_t$) and that predicted by the machine learning model ($\hat{\mathbf{V}}_p$) derived from kSZ + a single realization of the primary CMB and detector noise for the augmented Flender test sample II. The velocity when the true mass is used as a model feature (empty circle) is compared to that when the mass feature is modified to account for mass misestimation (filled circle). 1-σ error bars estimated using a bootstrap of the test data are also shown.

pairwise velocity when the CMB and noise are included, relative to that inferred if the true halo masses are used as input features. The deviations are largely captured within the 1σ bootstrap uncertainties.

The best-fit χ^2 for the predicted pairwise signal from the single CMB and noise realization, relative to the underlying true pairwise velocities, is 11 for the 16 bins when the mass scatter is included, versus 10 when we assume perfect knowledge of the cluster masses.

Note here, to demonstrate the machine learning prediction efficacy, the comparisons above are between the predicted and true velocity correlations for the augmented Flender test sample II sample. Separately from this, the true pairwise velocity statistic for the augmented sample is in itself intrinsically 2.4% lower, on average, than that for the original Flender test sample II, because the augmented sample has a lower mean mass. Both effects, the difference in the two samples and the bias in the machine learning prediction, each arising from the mass misestimation, would have to be factored in if the derived pairwise velocity is to be compared to theoretical cosmological predictions that assume a given cluster mass distribution.

D. Application to other simulations

To broaden beyond the Flender simulation, we also test the machine learning model on a second kSZ simulation, the Websky simulation [152]. We upgrade the Websky kSZ map to the same resolution as the Flender simulation before we measure the model features, and we use the same signal template profile (parameters), calibrated based on FL3, for the matched filter across different simulations.

Similar to the Flender simulation, we find that training on 70% of the pure kSZ signals on the Websky simulation and testing it on the remaining 30% of the Websky simulation gives an unbiased reconstruction of the pairwise peculiar velocity estimator. Note that we find the Websky trained machine learning model performs poorly for scales above 200 Mpc, because the systematic errors dominate, as it did for the Flender trained one.

We then cross-validate our model by applying the model trained on the Flender simulation to predict the velocities in the Websky test sample with a mass sample $M_{500} > 8 \times 10^{13} M_{\odot}$, as outlined in Sec. II A. The 10 realizations of the detector noise and primary CMB used on the Flender analysis were also included and a bootstrap analysis of one realization was performed to estimate the signal covariance. We find that the machine learning algorithms respectively trained on the Flender simulations and Websky simulations can both recover the true $\hat{\mathbf{V}}$ well for the Websky training sample as shown in Fig. 11.

When the model is trained on the Flender simulations, the mean predicted pairwise signal over the 10 realizations for the Websky test data has a best-fit χ^2 of 11 and an SNR of 9, for the 16 bins for $r < 200$ Mpc. When trained on the Websky simulations, using the same 10 realizations of CMB and noise, the predicted signal has a best-fit χ^2 of 11 and SNR of 10. The two predictions, trained on the different simulation sets, are therefore very consistent, showing that the approach is not limited to simply
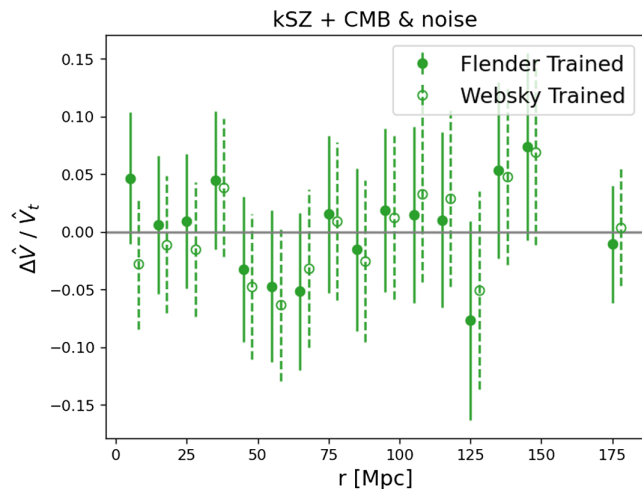
FIG. 11. Average fractional difference ($\Delta\hat{\mathbf{V}}/\hat{\mathbf{V}}_t$) between the mean pairwise velocity estimator from the true velocity ($\hat{\mathbf{V}}_t$) and that from the machine learning prediction ($\hat{\mathbf{V}}_p$) for the Websky test sample over 10 realizations of the detector noise and primary CMB, using the machine learning model trained on data from the Flender simulation (filled circle/full line) and trained on the Websky simulation (open circle/dashed line).

reconstruction of test data derived from the same simulations as the training set. This demonstrates the potential for using this machine learning model, trained on the Flender simulations (or other future high fidelity kSZ and galaxy simulations), as a suitable approach for the analysis of real survey data.

V. CONCLUSION

In this work, we have demonstrated that a machine learning model using the gradient boosting algorithm trained on high-fidelity simulations can be used to recover an unbiased estimate of pairwise peculiar velocity from kSZ and halo catalog data. This provides a tantalizing alternative to the two-step process currently used to extract the velocity estimate, combining a pairwise kSZ momentum estimator with a separate estimate of the cluster optical depths, with the additional uncertainties that involves.

For the machine learning approach, we consider seven halo features for training and prediction. Following [95], we consider two signal extraction techniques, matched filter and aperture photometry, to create the kSZ temperature features. We measure kSZ signals after filtering out CMB-dominated signals ($\ell < 360$) when adapting to real observations.

We first test the feasibility of our machine learning model to recover the halo individual peculiar velocity. We find that the model is efficient at recovering an unbiased measurement of the individual peculiar velocity for both pure kSZ signals and when detector noise and primary CMB are added. Given that the primary CMB dominates at large scales, and the detector noise dominates at small

scales, the tSZ anisotropies and foreground contaminations are subdominant than the primary CMB and detector noise. We thus did not test the presence of these noises in this work. We suggest testing the deprojection and the presence of these foreground contaminations in future work. While the presence of the CMB and noise makes the correlation weaker between temperature features and peculiar velocities, we find that a minimum mass threshold of $M_{500} > 8 \times 10^{13} M_{\odot}$ is practical for confidently recovering the halo velocities with the model using a matched filter. The correlation is lower when using aperture photometry, which translates into a far higher minimum mass threshold. For this reason, aperture photometry features are not incorporated in the main analysis. At large separations, of $r > 200$ Mpc, the kSZ signal becomes very small, and non-kSZ signals dominate leading to inaccurate and biased predictions. For separations $r < 200$ Mpc, the machine learning approach provides an unbiased prediction of the velocities with modeling uncertainty levels that are significantly smaller than the statistical uncertainties estimated to be inherent in the test sample itself. We also test the model feature importance by excluding different features from the model and evaluating the model performance. We find that the mass information is the most important feature that break the degeneracy between velocity and optical depth from the kSZ signals. The mass serves as a proxy for optical depth through an implicit cluster mass—τ relation that allows the model to differentiate among the different levels of kSZ temperature and velocity combinations.

We consider the effects of two potentially significant systematic effects: cluster location miscentering and mass misestimation. We find that using the location of the peak tSZ signal as the proxy to define the cluster center could reduce the systematic error in the positional offset introduced by using the brightest galaxy as a proxy for the halo center. The pairwise velocity amplitude suppression from miscentering is reduced from 9%, when modeling centering using a brightest galaxy, to 3% when using the peak tSZ signal as the proxy center. In considering the impact of cluster mass misestimation, we find that a 20% mass scatter would reduce the predicted pairwise velocity estimator by 3% for both pure kSZ signals and when detector noise and primary CMB are added. Such 3% offsets fall well within the 1σ statistical sample uncertainty estimates.

While the main analysis uses the Flender simulations to train and test the machine learning model, we also consider its performance with a second kSZ simulation to demonstrate its robustness and ability to be generalized. We use the Websky simulation which includes similar astrophysical effects to the Flender simulation while at a lower resolution. We find that the model trained on the Websky simulation is efficient at reconstructing the pairwise velocity estimator when applied to the Websky test data. Furthermore, the velocity statistics in the Websky test data

are also able to be accurately reconstructed with the Flender-trained model.

In this work, we show the feasibility of using a machine learning model that uses cluster properties measurable with survey data as features to recover the pairwise velocity estimator with the intent to demonstrate its applicability to upcoming real observations. We focus on one approach using a gradient-boosting algorithm, LightGBM. Given other machine learning applications to the analysis of galaxy clusters, for example, neural networks for the study of mass estimation [104], further opportunities for improved performance could well be realized by using additional model features together with a more complex feature engineering process. This could include enhancing the kSZ temperature features by including additional velocity information using velocity reconstruction with the galaxy number density field [187,188] or additional cluster features, such as the tSZ, weak lensing, x-ray, and richness statistics related to target clusters, to augment the cluster fitting model. Improvement of the mass calibration might also be achieved through machine learning approaches trained on large-scale simulations to refine calibration/scaling relations. Novel multifrequency techniques [189–192] have been developed to deproject the Compton-y and foreground contaminations, such as the cosmic infrared background (CIB), from the CMB observations. Such techniques have been used, for example, to provide a cleaned $kSZ + CMB - only$ map as the data products of ACT DR4 [189], analyzed in [69,94]. It would be interesting to consider how the presence of residual tSZ or CIB foreground components, or emission from radio and dusty star-forming galaxies, might affect, or be identified through, the machine learning approach. We leave the integration of such extensions to future work.

In summary, this work provides a promising method to measure the pairwise peculiar velocity estimator using a machine learning model. The model utilizes the relationships between cluster observables derived from high-fidelity simulations to predict the velocity directly from the kSZ data rather than measuring additional optical depth information derived from observations. The approach will be valuable in analyses of the upcoming CMB experiments, from the CCAT, Simons Observatory, and CMB-S4 surveys, along with galaxy surveys from DESI, SDSS, Euclid, and LSST to use cluster velocity correlations to constrain the properties of gravity, dark energy, and neutrinos.

[1] P. A. R. Ade *et al.* (Planck Collaboration), Astron. Astrophys. **571,** A16 (2014).

[2] R. Adam *et al.* (Planck Collaboration), Astron. Astrophys. **594,** A1 (2016).

[3] N. Aghanim *et al.* (Planck Collaboration), Astron. Astrophys. **641,** A6 (2020); **652,** C4(E) (2021),

[4] C. L. Bennett *et al.* (WMAP Collaboration), Astrophys. J. Suppl. Ser. **148,** 1 (2003).

[5] E. Komatsu *et al.* (WMAP Collaboration), Astrophys. J. Suppl. Ser. **192,** 18 (2011).

[6] G. Hinshaw *et al.* (WMAP Collaboration), Astrophys. J. Suppl. Ser. **208,** 19 (2013).

[7] E. Calabrese *et al.*, Phys. Rev. D **87,** 103012 (2013).

[8] K. T. Story *et al.*, Astrophys. J. **779,** 86 (2013).

[9] D. J. Eisenstein *et al.* (SDSS Collaboration), Astrophys. J. **633,** 560 (2005).

[10] W. J. Percival, S. Cole, D. J. Eisenstein, R. C. Nichol, J. A. Peacock, A. C. Pope, and A. S. Szalay, Mon. Not. R. Astron. Soc. **381,** 1053 (2007).

[11] W. J. Percival *et al.* (SDSS Collaboration), Mon. Not. R. Astron. Soc. **401,** 2148 (2010).

[12] N. Padmanabhan, X. Xu, D. J. Eisenstein, R. Scalzo, A. J. Cuesta, K. T. Mehta, and E. Kazin, Mon. Not. R. Astron. Soc. **427,** 2132 (2012).

[13] E. A. Kazin *et al.*, Mon. Not. R. Astron. Soc. **441,** 3524 (2014).

[14] S. Alam *et al.* (eBOSS Collaboration), Phys. Rev. D **103,** 083533 (2021).

[15] A. G. Riess *et al.* (Supernova Search Team Collaboration), Astron. J. **116,** 1009 (1998).

[16] S. Perlmutter *et al.* (Supernova Cosmology Project Collaboration), Astrophys. J. **517,** 565 (1999).

[17] T. de Jaeger *et al.*, Astrophys. J. **815,** 121 (2015).

[18] B. R. Zhang, M. J. Childress, T. M. Davis, N. V. Karpenka, C. Lidman, B. P. Schmidt, and M. Smith, Mon. Not. R. Astron. Soc. **471,** 2254 (2017).

[19] P. J. E. Peebles and B. Ratra, Rev. Mod. Phys. **75,** 559 (2003).

[20] J.-P. Uzan, Gen. Relativ. Gravit. **42,** 2219 (2010).

[21] B. Jain *et al.*, arXiv:1309.5389.

[22] A. Joyce, B. Jain, J. Khoury, and M. Trodden, Phys. Rep. **568,** 1 (2015).

[23] S. Weinberg, Rev. Mod. Phys. **61,** 1 (1989).

[24] C. A. Egan and C. H. Lineweaver, Phys. Rev. D **78,** 083528 (2008).

[25] M. Quartin, M. O. Calvao, S. E. Joras, R. R. R. Reis, and I. Waga, J. Cosmol. Astropart. Phys. 05 (2008) 007.

[26] D. J. Shaw and J. D. Barrow, Phys. Rev. D **83,** 043518 (2011).

[27] H. E. S. Velten, R. F. vom Marttens, and W. Zimdahl, Eur. Phys. J. C **74,** 3160 (2014).

[28] A. Del Popolo and M. Le Delliou, Galaxies **5,** 17 (2017).

[29] F. Cunillera and A. Padilla, J. High Energy Phys. 10 (2021) 055.

[30] M. Shimon, arXiv:2204.02211.

[31] S. Jalalzadeh, A. Mohammadi, and D. Demir, Phys. Dark Universe **40,** 101227 (2023).

[32] T. Clifton, P. G. Ferreira, A. Padilla, and C. Skordis, Phys. Rep. **513,** 1 (2012).

[33] E. Bellini and I. Sawicki, J. Cosmol. Astropart. Phys. 07 (2014) 050.

[34] I. de Martino, M. De Laurentis, and S. Capozziello, Universe **1,** 123 (2015).

[35] G. Cusin, M. Lewandowski, and F. Vernizzi, J. Cosmol. Astropart. Phys. 04 (2018) 005.

[36] B. L'Huillier, H. A. Winther, D. F. Mota, C. Park, and J. Kim, Mon. Not. R. Astron. Soc. **468,** 3174 (2017).

[37] L. Perenon, J. Bel, R. Maartens, and A. de la Cruz-Dombriz, J. Cosmol. Astropart. Phys. 06 (2019) 020.

[38] A. Jiang, W. Liu, B. Li, C. Barrera-Hinojosa, Y. Zhang, and W. Fang, Phys. Rev. D **109,** 083537 (2024).

[39] R. A. Sunyaev and Y. B. Zeldovich, Astrophys. Space Sci. **7,** 3 (1970).

[40] R. A. Sunyaev and Y. B. Zeldovich, Comments Astrophys. Space Phys. **4,** 173 (1972), https://ui.adsabs.harvard.edu/abs/1972CoASP...4..173S/abstract.

[41] R. A. Sunyaev and Y. B. Zeldovich, Mon. Not. R. Astron. Soc. **190,** 413 (1980).

[42] M. Birkinshaw, Phys. Rep. **310,** 97 (1999).

[43] J. E. Carlstrom, G. P. Holder, and E. D. Reese, Annu. Rev. Astron. Astrophys. **40,** 643 (2002).

[44] A. Kosowsky and S. Bhattacharya, Phys. Rev. D **80,** 062003 (2009).

[45] M. A. Mitchell, C. Arnold, C. Hernández-Aguayo, and B. Li, Mon. Not. R. Astron. Soc. **501,** 4565 (2021).

[46] T. Okumura and A. Taruya, Phys. Rev. D **106,** 043523 (2022).

[47] J. Hou, J. Bautista, M. Berti, C. Cuesta-Lazaro, C. Hernández-Aguayo, T. Tröster, and J. Zheng, Universe **9,** 302 (2023).

[48] P. J. E. Peebles, *Large-Scale Structure of the Universe* (Princeton University Press, 1980), ISBN 978-0-691-08240-0, https://ui.adsabs.harvard.edu/abs/1980lssu.book.....P/abstract.

[49] A. Diaferio, R. A. Sunyaev, and A. Nusser, Astrophys. J. Lett. **533,** L71 (2000).

[50] P. G. Ferreira, R. Juszkiewicz, H. A. Feldman, M. Davis, and A. H. Jaffe, Astrophys. J. Lett. **515,** L1 (1999).

[51] N. Aghanim, K. M. Gorski, and J. L. Puget, Astron. Astrophys. **374,** 1 (2001).

[52] C. Hernandez-Monteagudo, L. Verde, R. Jimenez, and D. N. Spergel, Astrophys. J. **643,** 598 (2006).

[53] S. DeDeo, D. N. Spergel, and H. Trac, arXiv:astro-ph/0511060.

[54] S. Bhattacharya and A. Kosowsky, Astrophys. J. Lett. **659,** L83 (2007).

[55] S. Bhattacharya and A. Kosowsky, Phys. Rev. D **77,** 083004 (2008).

[56] P. Fosalba and O. Dore, Phys. Rev. D **76,** 103523 (2007).

[57] P. A. R. Ade *et al.* (Planck Collaboration), Astron. Astrophys. **561,** A97 (2014).

[58] B. Soergel, A. Saro, T. Giannantonio, G. Efstathiou, and K. Dolag, Mon. Not. R. Astron. Soc. **478,** 5320 (2018).

[59] N. Aghanim *et al.* (Planck Collaboration), Astron. Astrophys. **617,** A48 (2018).

[60] E.-M. Mueller, F. de Bernardis, R. Bean, and M. D. Niemack, Phys. Rev. D **92,** 063501 (2015).

[61] E.-M. Mueller, F. de Bernardis, R. Bean, and M. D. Niemack, Astrophys. J. **808,** 47 (2015).

[62] J. Kuruvilla, N. Aghanim, and I. G. McCarthy, Astron. Astrophys. **644,** A170 (2020).

[63] Y. Zheng, Astrophys. J. **904,** 48 (2020).

[64] N. Hand *et al.*, Phys. Rev. Lett. **109,** 041101 (2012).

[65] D. S. Swetz *et al.*, Astrophys. J. Suppl. Ser. **194,** 41 (2011).

[66] J. E. Gunn *et al.* (SDSS Collaboration), Astron. J. **116,** 3040 (1998).

[67] J. E. Gunn *et al.* (SDSS Collaboration), Astron. J. **131,** 2332 (2006).

[68] F. De Bernardis *et al.*, J. Cosmol. Astropart. Phys. 03 (2017) 008.

[69] V. Calafut *et al.*, Phys. Rev. D **104,** 043502 (2021).

[70] P. A. R. Ade *et al.* (Planck Collaboration), Astron. Astrophys. **586,** A140 (2016).

[71] J. E. Carlstrom *et al.*, Publ. Astron. Soc. Pac. **123,** 568 (2011).

[72] B. Soergel *et al.* (DES, SPT Collaborations), Mon. Not. R. Astron. Soc. **461,** 3172 (2016).

[73] E. Schiappucci *et al.* (SPT-3G, DES Collaborations), Phys. Rev. D **107,** 042004 (2023).

[74] J. C. Hill, S. Ferraro, N. Battaglia, J. Liu, and D. N. Spergel, Phys. Rev. Lett. **117,** 051301 (2016).

[75] S. Ferraro, J. C. Hill, N. Battaglia, J. Liu, and D. N. Spergel, Phys. Rev. D **94,** 123526 (2016).

[76] A. Kusiak, B. Bolliet, S. Ferraro, J. C. Hill, and A. Krolewski, Phys. Rev. D **104,** 043518 (2021).

[77] B. Bolliet, J. C. Hill, S. Ferraro, A. Kusiak, and A. Krolewski, J. Cosmol. Astropart. Phys. 03 (2023) 039.

[78] R. Patki, N. Battaglia, and S. Ferraro, Phys. Rev. D **108,** 043507 (2023).

[79] M. Li, R. E. Angulo, S. D. M. White, and J. Jasche, Mon. Not. R. Astron. Soc. **443,** 2311 (2014).

[80] E. Schaan *et al.* (ACTPol Collaboration), Phys. Rev. D **93,** 082002 (2016).

[81] H. Tanimura, N. Aghanim, V. Bonjean, and S. Zaroubi, Astron. Astrophys. **662,** A48 (2022).

[82] M. Mallaby-Kay *et al.* (DES Collaboration), Phys. Rev. D **108,** 023516 (2023).

[83] K. M. Smith, M. S. Madhavacheril, M. Münchmeyer, S. Ferraro, U. Giri, and M. C. Johnson, arXiv:1810.13423.

[84] G. Sato-Polito, J. L. Bernal, K. K. Boddy, and M. Kamionkowski, Phys. Rev. D **103,** 083519 (2021).

[85] J. Sayers *et al.*, Astrophys. J. **778**, 52 (2013).

[86] A. Mittal, F. de Bernardis, and M. D. Niemack, J. Cosmol. Astropart. Phys. 02 (2018) 032.

[87] D. Li, H.-M. Zhu, and U.-L. Pen, Phys. Rev. D **100**, 023517 (2019).

[88] E. M. George *et al.*, Astrophys. J. **799**, 177 (2015).

[89] S. Li, Y. Zheng, Z. Chen, H. Xu, and X. Yang, Astrophys. J. Suppl. Ser. **271**, 30 (2024).

[90] N. Battaglia, J. Cosmol. Astropart. Phys. 08 (2016) 058.

[91] L. D. Shaw, D. Nagai, S. Bhattacharya, and E. T. Lau, Astrophys. J. **725**, 1452 (2010).

[92] K. Dolag and R. Sunyaev, Mon. Not. R. Astron. Soc. **432**, 1600 (2013).

[93] S. Flender, D. Nagai, and M. McDonald, Astrophys. J. **837**, 124 (2017).

[94] E. M. Vavagiakis *et al.*, Phys. Rev. D **104**, 043503 (2021).

[95] Y. Gong, R. Bean, P. A. Gallardo, E. M. Vavagiakis, N. Battaglia, and M. Niemack, Phys. Rev. D **109**, 023513 (2024).

[96] S. B. Green, M. Ntampaka, D. Nagai, L. Lovisari, K. Dolag, D. Eckert, and J. A. ZuHone, Astrophys. J. **884**, 33 (2019).

[97] D. Kodi Ramanah, R. Wojtak, and N. Arendse, Mon. Not. R. Astron. Soc. **501**, 4080 (2021).

[98] Z. Yan, A. J. Mead, L. Van Waerbeke, G. Hinshaw, and I. G. McCarthy, Mon. Not. R. Astron. Soc. **499**, 3445 (2020).

[99] M. Ho, M. M. Rau, M. Ntampaka, A. Farahi, H. Trac, and B. Poczos, Astrophys. J. **887**, 25 (2019).

[100] M. Ho, A. Farahi, M. M. Rau, and H. Trac, Astrophys. J. **908**, 204 (2021).

[101] N. Gupta and C. L. Reichardt, Astrophys. J. **900**, 110 (2020).

[102] D. de Andres, W. Cui, F. Ruppin, M. De Petris, G. Yepes, I. Lahouli, G. Aversano, R. Dupuis, and M. Jarraya, EPJ Web Conf. **257**, 00013 (2022).

[103] A. Ferragamo *et al.*, Mon. Not. R. Astron. Soc. **520**, 4000 (2023).

[104] D. de Andres *et al.*, Nat. Astron. **6**, 1325 (2022).

[105] M. Ho, J. Soltis, A. Farahi, D. Nagai, A. Evrard, and M. Ntampaka, Mon. Not. R. Astron. Soc. **524**, 3289 (2023).

[106] H. Shao, F. Villaescusa-Navarro, S. Genel, D. N. Spergel, D. Angles-Alcazar, L. Hernquist, R. Dave, D. Narayanan, G. Contardo, and M. Vogelsberger, Astrophys. J. **927**, 85 (2022).

[107] A. M. Delgado, D. Wadekar, B. Hadzhiyska, S. Bose, L. Hernquist, and S. Ho, Mon. Not. R. Astron. Soc. **515**, 2733 (2022).

[108] D. Wadekar *et al.*, Mon. Not. R. Astron. Soc. **522**, 2628 (2023).

[109] D. Wadekar, L. Thiele, F. Villaescusa-Navarro, J. C. Hill, M. Cranmer, D. N. Spergel, N. Battaglia, D. Anglés-Alcázar, L. Hernquist, and S. Ho, Proc. Natl. Acad. Sci. U.S.A. **120**, e2202074120 (2023).

[110] A. C. Rodriguez, T. Kacprzak, A. Lucchi, A. Amara, R. Sgier, J. Fluri, T. Hofmann, and A. Réfrégier, Comput. Astrophys. Cosmol. **5**, 4 (2018).

[111] A. Mishra, P. Reddy, and R. Nigam, arXiv:1908.04682.

[112] F. Villaescusa-Navarro *et al.* (CAMELS Collaboration), Astrophys. J. **915**, 71 (2021).

[113] T. Rothschild, D. Nagai, H. Aung, S. B. Green, M. Ntampaka, and J. ZuHone, Mon. Not. R. Astron. Soc. **513**, 333 (2022).

[114] S. Pan, M. Liu, J. Forero-Romero, C. G. Sabiu, Z. Li, H. Miao, and X.-D. Li, Sci. China Phys. Mech. Astron. **63**, 110412 (2020).

[115] S. Ravanbakhsh, J. Oliva, S. Fromenteau, L. C. Price, S. Ho, J. Schneider, and B. Poczos, arXiv:1711.02033.

[116] A. Lazanu, J. Cosmol. Astropart. Phys. 09 (2021) 039.

[117] C. Bengaly, M. A. Dantas, L. Casarini, and J. Alcaniz, Eur. Phys. J. C **83**, 548 (2023).

[118] L. A. Perez, S. Genel, F. Villaescusa-Navarro, R. S. Somerville, A. Gabrielpillai, D. Anglés-Alcázar, B. D. Wandelt, and L. Y. A. Yung, Astrophys. J. **954**, 11 (2023).

[119] L. Lucie-Smith, H. V. Peiris, A. Pontzen, and M. Lochner, Mon. Not. R. Astron. Soc. **479**, 3405 (2018).

[120] S. He, Y. Li, Y. Feng, S. Ho, S. Ravanbakhsh, W. Chen, and B. Póczos, Proc. Natl. Acad. Sci. U.S.A. **116**, 13825 (2019).

[121] L. Lucie-Smith, Ph.D. thesis, University Coll. London (2020).

[122] A. Petulante, Ph.D. thesis, 0000-0002-7627-5444, Vanderbilt University, 2022.

[123] Z. Wu, Z. Zhang, S. Pan, H. Miao, X. Wang, C. G. Sabiu, J. Forero-Romero, Y. Wang, X.-D. Li, and X. Luo, Astrophys. J. **913**, 2 (2021).

[124] Z. Wu, L. Xiao, X. Xiao, J. Wang, X. Kang, Y. Wang, X. Wang, L. Z. Le Zhang, and X.-D. Li, Mon. Not. R. Astron. Soc. **522**, 4748 (2023).

[125] F. Qin, D. Parkinson, S. E. Hong, and C. G. Sabiu, J. Cosmol. Astropart. Phys. 06 (2023) 062.

[126] Z. Wang, F. Shi, X. Yang, Q. Li, Y. Liu, and X. Li, Sci. China Phys. Mech. Astron. **67**, 219513 (2024).

[127] C. E. Petrillo *et al.*, Mon. Not. R. Astron. Soc. **472**, 1129 (2017).

[128] L. Perreault Levasseur, Y. D. Hezaveh, and R. H. Wechsler, Astrophys. J. Lett. **850**, L7 (2017).

[129] Y. D. Hezaveh, L. Perreault Levasseur, and P. J. Marshall, Nature (London) **548**, 555 (2017).

[130] W. R. Morningstar, Y. D. Hezaveh, L. Perreault Levasseur, R. D. Blandford, P. J. Marshall, P. Putzky, and R. H. Wechsler, arXiv:1808.00011.

[131] M. Tewes, T. Kuntzer, R. Nakajima, F. Courbin, H. Hildebrandt, and T. Schrabback, Astron. Astrophys. **621**, A36 (2019).

[132] A. Gupta, J. M. Z. Matilla, D. Hsu, and Z. Haiman, Phys. Rev. D **97**, 103515 (2018).

[133] O. M. Springer, E. O. Ofek, Y. Weiss, and J. Merten, Mon. Not. R. Astron. Soc. **491**, 5301 (2020).

[134] D. Ribli, B. A. Pataki, J. M. Zorrilla Matilla, D. Hsu, Z. Haiman, and I. Csabai, Mon. Not. R. Astron. Soc. **490**, 1843 (2019).

[135] J. Fluri, T. Kacprzak, A. Lucchi, A. Refregier, A. Amara, T. Hofmann, and A. Schneider, Phys. Rev. D **100**, 063514 (2019).

[136] N. Jeffrey, F. Lanusse, O. Lahav, and J.-L. Starck, Mon. Not. R. Astron. Soc. **492**, 5023 (2020).

[137] Z. Zhang, H. Shan, N. Li, C. Wei, J. Yao, and R. Li, Astron. Astrophys. **683,** A209 (2024).

[138] D. Baron, arXiv:1904.07248.

[139] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, Phys. Rep. **810,** 1 (2019).

[140] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Rev. Mod. Phys. **91,** 045002 (2019).

[141] M. Ntampaka *et al.*, arXiv:1902.10159.

[142] Y. Wang, N. Ramachandra, E. M. Salazar-Canizales, H. A. Feldman, R. Watkins, and K. Dolag, Mon. Not. R. Astron. Soc. **506,** 1427 (2021).

[143] A. M. Darya, I. Fernini, M. Vellasco, and A. Hussain, in *2023 International Joint Conference on Neural Networks (IJCNN)* (IEEE, New York, 2023), 10.1109/IJCNN54540.2023 .10191163.

[144] N. Sahakyan, V. Vardanyan, and M. Khachatryan, Mon. Not. R. Astron. Soc. **519,** 3000 (2022).

[145] A. Tolamatti, K. K. Singh, and K. K. Yadav, Mon. Not. R. Astron. Soc. **523,** 5341 (2023).

[146] J. Coronado-Blázquez, Mon. Not. R. Astron. Soc. **521,** 4156 (2023).

[147] S. Flender, L. Bleem, H. Finkel, S. Habib, K. Heitmann, and G. Holder, Astrophys. J. **823,** 98 (2016).

[148] K. Heitmann *et al.*, Astrophys. J. **820,** 108 (2016).

[149] S. Habib *et al.*, New Astron. **42,** 49 (2016).

[150] D. Larson *et al.*, Astrophys. J. Suppl. Ser. **192,** 16 (2011).

[151] S. Naess *et al.*, J. Cosmol. Astropart. Phys. 12 (2020) 046.

[152] G. Stein, M. A. Alvarez, J. R. Bond, A. van Engelen, and N. Battaglia, J. Cosmol. Astropart. Phys. 10 (2020) 012.

[153] A. Lewis and A. Challinor, arXiv:1102.026.

[154] P. A. R. Ade *et al.* (Planck Collaboration), Astron. Astrophys. **594,** A13 (2016).

[155] P. Ade *et al.* (Simons Observatory Collaboration), J. Cosmol. Astropart. Phys. 02 (2019) 056.

[156] M. Bartelmann, Astron. Astrophys. **313,** 697 (1996), https://ui.adsabs.harvard.edu/abs/1996A%26A...313 .697B/abstract.

[157] Y.-C. Li, Y.-Z. Ma, M. Remazeilles, and K. Moodley, Phys. Rev. D **97,** 023514 (2018).

[158] A. Gholamy, V. Kreinovich, and O. Kosheleva (2018), https://api.semanticscholar.org/CorpusID:7467506.

[159] Z. L. Wen and J. L. Han, Astrophys. J. **807,** 178 (2015).

[160] S. Andreon, Astron. Astrophys. **587,** A158 (2016).

[161] J. E. Geach and J. A. Peacock, Nat. Astron. **1,** 795 (2017).

[162] R. Stanek, A. E. Evrard, H. B. Bohringer, P. Schuecker, and B. Nord, Astrophys. J. **648,** 956 (2006).

[163] S. Ettori, A. Donnarumma, E. Pointecouteau, T. H. Reiprich, S. Giodini, L. Lovisari, and R. W. Schmidt, Space Sci. Rev. **177,** 119 (2013).

[164] S. Amodeo, S. Ettori, R. Capasso, and M. Sereno, Astron. Astrophys. **590,** A126 (2016).

[165] D. E. Applegate, A. von der Linden, P. L. Kelly, M. T. Allen, S. W. Allen, P. R. Burchat, D. L. Burke, H. Ebeling, A. Mantz, and R. G. Morris, Mon. Not. R. Astron. Soc. **439,** 48 (2014).

[166] T. McClintock *et al.* (DES Collaboration), Mon. Not. R. Astron. Soc. **482,** 1352 (2019).

[167] C. Murray, J. G. Bartlett, E. Artis, and J.-B. Melin, Mon. Not. R. Astron. Soc. **512,** 4785 (2022).

[168] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, Adv. Neural Inf. Process. Syst. **30,** 3146 (2017), https://papers.nips.cc/paper_files/paper/2017/ hash/6449f44a102fde848669bdd9eb6b76fa-Abstract .html.

[169] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, J. Mach. Learn. Res. **12,** 2825 (2011), https://www.arxiv.org/abs/1201.0490.

[170] B. Ratner and J. Targeting, Meas. Anal. Mark. **17,** 139 (2009), https://www.ncbi.nlm.nih.gov/pmc/articles/ PMC8795012/.

[171] F. C. van den Bosch, S. M. Weinmann, X.-H. Yang, H. J. Mo, C. Li, and Y. P. Jing, Mon. Not. R. Astron. Soc. **361,** 1203 (2005).

[172] R. A. Skibba, F. C. v. d. Bosch, X. Yang, S. More, H. Mo, and F. Fontanot, Mon. Not. R. Astron. Soc. **410,** 417 (2011).

[173] E. Rozo, E. S. Rykoff, M. Becker, R. M. Reddick, and R. H. Wechsler, Mon. Not. R. Astron. Soc. **453,** 38 (2015).

[174] H. Hoshino *et al.*, Mon. Not. R. Astron. Soc. **452,** 998 (2015).

[175] A. Saro *et al.* (DES Collaboration), Mon. Not. R. Astron. Soc. **454,** 2305 (2015).

[176] J. U. Lange, F. C. van den Bosch, A. Hearin, D. Campbell, A. R. Zentner, A. Villarreal, and Y.-Y. Mao, Mon. Not. R. Astron. Soc. **473,** 2830 (2018).

[177] V. Calafut, R. Bean, and B. Yu, Phys. Rev. D **96,** 123529 (2017).

[178] R. Seppi, J. Comparat, K. Nandra, K. Dolag, V. Biffi, E. Bulbul, A. Liu, V. Ghirardini, and J. Ider-Chitham, Astron. Astrophys. **671,** A57 (2023).

[179] G. Chabrier, Publ. Astron. Soc. Pac. **115,** 763 (2003).

[180] A. Kravtsov, A. Vikhlinin, and A. Meshscheryakov, Astron. Lett. **44,** 8 (2018).

[181] E. F. Bell and R. S. de Jong, Astrophys. J. **550,** 212 (2001).

[182] E. F. Bell, D. H. McIntosh, N. Katz, and M. D. Weinberg, Astrophys. J. Suppl. Ser. **149,** 289 (2003).

[183] D. Gruen, G. M. Bernstein, T. Y. Lam, and S. Seitz, Mon. Not. R. Astron. Soc. **416,** 1392 (2011).

[184] F. Köhlinger, H. Hoekstra, and M. Eriksen, Mon. Not. R. Astron. Soc. **453,** 3107 (2015).

[185] T. Schrabback *et al.* (SPT Collaboration), Mon. Not. R. Astron. Soc. **474,** 2635 (2018).

[186] E. J. Baxter, E. Rozo, B. Jain, E. Rykoff, and R. H. Wechsler, Mon. Not. R. Astron. Soc. **463,** 205 (2016).

[187] B. R. Guachalla, E. Schaan, B. Hadzhiyska, and S. Ferraro, arXiv:2312.12435.

[188] B. Hadzhiyska, S. Ferraro, B. R. Guachalla, and E. Schaan, arXiv:2312.12434.

[189] M. S. Madhavacheril *et al.*, Phys. Rev. D **102,** 023534 (2020).

[190] D. Han and N. Sehgal, Phys. Rev. D **105,** 083516 (2022).

[191] F. McCarthy and J. C. Hill, Phys. Rev. D **109,** 023528 (2024).

[192] C. T. Pratt, Z. Qu, J. N. Bregman, and C. J. Miller, Astrophys. J. **964,** 122 (2024).