Lawrence Berkeley National Laboratory
LBL Publications

Title
TARDIS. II. Synergistic Density Reconstruction from Lyα Forest and Spectroscopic Galaxy Surveys with Applications to Protoclusters and the Cosmic Web

Permalink
https://escholarship.org/uc/item/89z6v91d

Journal
The Astrophysical Journal, 906(2)

ISSN
0004-637X

Authors
Horowitz, Benjamin
Zhang, Benjamin
Lee, Khee-Gan
et al.

Publication Date
2021

DOI
10.3847/1538-4357/abca35

Peer reviewed

TARDIS Paper II: Synergistic Density Reconstruction from Lyman-α Forest and Spectroscopic Galaxy Surveys with Applications to Protoclusters and the Cosmic Web

Benjamin Horowitz,[1, 2] Benjamin Zhang,[3, 4] Khee-Gan Lee,[4] and Robin Kooistra[4]

[1]*Princeton Department of Astrophysical Sciences, Princeton, NJ 08544, USA*
[2]*Lawrence Berkeley National Lab, 1 Cyclotron Road, Berkeley, CA 94720, USA*
[3]*Department of Astronomy, University of California at Berkeley, Berkeley, CA 94720, USA*
[4]*Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan*

Abstract

In this work we expand upon the Tomographic Absorption Reconstruction and Density Inference Scheme (TARDIS) in order to include multiple tracers while reconstructing matter density fields at Cosmic Noon ($z \sim 2-3$). In particular, we jointly reconstruct the underlying density field from simulated Lyα forest observations at $z \sim 2.5$ and an overlapping galaxy survey. We find that these data are synergistic, with the Lyman Alpha forest providing reconstruction of low density regions and galaxy surveys tracing the density peaks. We find a more accurate power spectra reconstruction going to higher scales when fitting these two data-sets simultaneously than if using either one individually. When applied to cosmic web analysis, we find performing the joint analysis is equivalent to a Lyman Alpha survey with significantly increased sight-line spacing. Since we reconstruct the velocity field and matter field jointly, we demonstrate the ability to evolve the mock observed volume further to $z = 0$, allowing us to create a rigorous definition of "proto-cluster" as regions which will evolve into clusters. We apply our reconstructions to study protocluster structure and evolution, finding for realistic survey parameters we can provide accurate mass estimates of the $z \approx 2$ structures and their $z = 0$ fate.

1. INTRODUCTION

Starting with the discovery by Gunn & Peterson (1965) of the photo-ionized intergalactic medium (IGM), the absorption in the rest frame Lyα by intervening neutral hydrogen in the spectra of luminous quasars, the so-called Lyα forest, has given crucial insights to the high redshift universe. The power spectra of these fluctuations has been used to constrain exotic physics models (Viel et al. 2005, 2013) as well as ΛCDM cosmology (Hernquist et al. 1996; Seljak et al. 2006; de Sainte Agathe et al. 2019), while individual features of the Lyα flux[1] have been used to constrain reionization models (Becker et al. 2015; Fan et al. 2006) and find cosmic structures (Wakker et al. 2015; Cai et al. 2017a).

In abstract, a sufficiently spatially dense collection of quasars at high redshift could be used to reconstruct the three dimensional map of the intergalactic medium, as first

bhorowitz@berkeley.edu

[1] In this paper, the term 'flux' implicitly refers to the Lyα forest transmitted flux.

discussed in Pichon et al. (2001) and Caucci et al. (2008). However, the relative rarity of luminous quasars have made making three dimensional maps of the flux field quite difficult (although see early attempts by Rollinde et al. 2003 and D'Odorico et al. 2006). Recently, there has been increased interest on tomographic reconstructions by also incorporating Lyα forest absorption features observed in UV-emitting star-forming galaxies at z > 2, often referred to as 'Lyman-Break Galaxies' (LBGs). The observed forest flux features can be reconstructed in three dimensional space via a Wiener filtering algorithm (Lee et al. 2014a) which can weight each sight-line based on its noise characteristics and impose a signal covariance based on the line of sight separations.

The ongoing COSMOS Lyman Alpha Mapping And Tomographic Observations (CLAMATO) survey (Lee et al. 2018) has used 240 LBG sightlines covering a ~ 600 square arcmin footprint (a sightline density of 1455 deg^{-2}), reconstructing a 3D tomographic map of the $2.05 < z < 2.55$ Lyα forest. This has allowed the probing of the three-dimensional (3D) structure of the optically thin IGM gas at $z > 2$ on scales of several comoving Mpc. Another survey aimed at

IGM tomography is the Lyα Tomography IMACS Survey (LATIS, Newman et al. 2020a), which has observed a wider area than CLAMATO, albeit at slightly lower spatial sampling.

Recently, Ravoux et al. (2020) carried out tomographic Lyα forest reconstructions using data from the SDSS eBOSS survey. As opposed to CLAMATO, which used sightlines from both quasars and Lyman-break galaxies, the authors used only quasar sightlines from eBOSS. This resulted in a sightline density of 37 deg^{-2}, but on a much larger footprint of 220 deg^2. With a mean sightline separation of $13\,h^{-1}\,\mathrm{Mpc}$, the authors were unable to resolve individual features of the cosmic web, but were able to identify large protocluster candidates and voids.

The limited number of sight-lines within a given magnitude limit restricts the mean separation between sources and therefore the scales which can be reconstructed accurately. One possible approach to reconstruct smaller scales is to add additional constraints on the reconstruction beyond the assumed correlation function and spectroscopic noise properties used in direct Wiener Filtering. This was explored in Horowitz et al. (2019a, , hereafter TARDIS-I), where the authors created a dynamical forward model from the initial density field (at $z \sim \infty$) to the observed data. This was then cast as a Bayesian inference problem and optimized using standard solvers. Applying this technique to mock catalogs, TARDIS-I was found to outperform Wiener filtering both in terms of flux map reconstruction and overall cosmic web classification. However, as the Lyα flux saturates in extremely dense environments, there was significant residual noise bias leading to an inability to accurately reconstruct protoclusters in the observed volume. Porqueres et al. (2019) utilized a Hamiltonian Monte Carlo solver with a strong cosmological prior to reconstruct the density field with a similar forward model.

One possibility to improve Lyα forest reconstructions in overdense small-scale regions is to use additional information to jointly reconstruct the underlying density field. A natural choice for an additional probe is the observed 3D galaxy field overlapping the reconstruction volume, much of which will be collected incidental to any Lyα tomographic spectroscopic survey. These galaxy populations are expected to be biased towards the more massive halos and provide a complementary probe of the cosmic structures.

When incorporating both galaxies and Lyα flux within a dynamic forward modelling framework, one would expect synergy where both fields inform the reconstruction of each other. When viewed within the peak-background split framework (Sheth & Tormen 1999), we expect galaxies to fall in dark matter halos which will collapse preferentially in more dense environments (i.e. on the peak of long wavelength modes). This would mean that the positions of galaxies should inform the positions of more diffuse structures (as traced by the Lyα forest) and visa versa. Alternatively, within a perturbation theory framework, one can view this as dynamical mode coupling (Jain & Bertschinger 1994; Crocce & Scoccimarro 2006) which would allow one to infer the relative power of longer wavelength modes based on observations of short wavelength modes (and visa versa).

This joint reconstruction takes on additional significance due to upcoming surveys probing both the galaxy positions and Lyα forest features over a wide sky area, including the Subaru Prime Focus Spectrograph Subaru-PFS (PFS; Takada et al. 2014a) and Maunakea Spectroscopic Explorer (MSE; McConnachie et al. 2016). These telescopes are expected to have IGM tomography programs similar in noise properties and density to CLAMATO, but over a significantly wider area of sky ($\sim 50\times$ in the case of PFS, or 12-15 deg^2). The PFS program in particular is planned to specifically target galaxies in the coeval region with the tomography program in order to study the interplay between galaxy properties and cosmic structures (see the Appendix in Nagamine et al. 2020).

Farther into the future, thirty-meter class facilities such as the Thirty Meter Telescope (TMT; Skidmore et al. 2015), Giant Magellan Telescope (GMT; Johns et al. 2012), and European Extremely Large Telescope (EELT; Evans et al. 2014), have the potential to drastically improve sensitivity for faint background sources allowing for significant increases in sight line density and probe spatial scales of ~ 1 cMpc and below (Japelj et al. 2019). The proposed 6.5m class MegaMapper facility (Schlegel et al. 2019), on the other hand, would push into the full-sky regime to obtain spectra for $\sim 10^8$ galaxies between $z = 2$ and $z = 5$ over a hemisphere, which would immediately enable joint Lyα tomography and galaxy analysis over $\sim$Gpc volumes.

Performing this joint reconstruction within a forward modeling framework has a number of significant advantages for cosmological analysis. Assuming no primordial non-gaussianity (as suggested in the latest Planck cosmological analysis (Planck Collaboration et al. 2018)) the power-spectrum of the initial density field should provide a lossless statistic, capturing all possible cosmological information. The entire family of higher order correlations and topological measures (such as three-point functions, density peak counts, voids, Minkowski functionals, etc.) arise due to gravitational evolution of a density field described by this primordial power spectrum. This encapsulation of a wide variety of cosmological probes has motivated a number of works in the field (, e.g., Seljak et al. 2017; Modi et al. 2018).

Beyond cosmological analysis, reconstructions from the Lyα forest and galaxy field data can help inform the galaxy - cosmic structure relationship at high redshift. One of the original goals of the CLAMATO survey was to map out protoclusters and voids in their volume. Stark et al. (2015a)

pointed out, using N-body simulations and mock Lyα forest data, that the properties of high-redshift protoclusters observed with IGM tomography, such as their spatial extent, can be linked to properties of their virialized $z = 0$ cluster descendants (but see Cai et al. 2016, Cai et al. 2017b, and Miller et al. 2019 for an alternative approach). This technique was applied by Lee et al. (2016) to study a $z \sim 2.4$ protoclusters detected in early CLAMATO data. However, accurate initial density reconstructions from joint galaxy-Lyα forest data allow the possibility of directly modelling the gravitational evolution of *individual* observe high-redshift protoclusters all the way to $z = 0$ and the fate of their constituent galaxies. Analogous techniques could also be used to study the evolution of individual void regions detected in IGM tomographic maps (Stark et al. 2015b; Krolewski et al. 2018).

Direct comparisons between galaxy positions and their Lyα forest environment have also been an active avenue for investigation in recent years (Font-Ribera et al. 2013; Mukae et al. 2017; Bielby et al. 2017; Momose et al. 2020a,b), in an effort to use the Lyα forest flux as a proxy for the underlying density field. The assumption of Lyα flux as a density tracer is, however, often complicated by hydrodynamical and radiative effects that can complicate the relationship between matter density and Lyα flux, as well as the fact that it is an exponential redshift-space tracer of the density field. Reconstruction methods like TARDIS-I, on the other hand, offer a way to compare galaxy properties with their surrounding large-scale (>1 Mpc) environments as defined directly by the matter density field. At the same time they also offer, at least in principle, a framework to simultaneously explore uncertainties in the flux-density mapping although we leave such investigations to future work.

In this work, we explore the possibilities of a joint reconstruction between an overlapping Lyα forest and galaxy spectroscopic survey within the forward modelling framework. In Section 2 we summarizes additions and changes to the technique described in (Horowitz et al. 2019a). In Section 3, we explain our mock generation procedure, in particular how we populate the density field with galaxies. We model our number densities and noise properties off of the upcoming Prime Focus Spectrograph. In Section 4, we show how the joint reconstructions improve the accuracy of the recovered density field and cosmic web classifications. In Appendix A, we demonstrate that our reconstruction quality does not deteriorate appreciable when applied to mock catalogs including fully modeled hydrodynamical effects.

For our mock catalogs, we use the best fit cosmology from Planck Collaboration et al. (2016).

2. METHODOLOGY

For a review of the optimization framework, see Horowitz et al. (2019a, hereafter "TARDIS-I"), as well as more general works (Seljak et al. 2017; Horowitz et al. 2019b). Here we specify how we expand our model to include galaxy fields and perform a joint optimization.

2.1. *Joint Likelihood*

The goal of our likelihood analysis is to maximize the combined likelihood of a prior term, a galaxy field term, and a Lyα term. This can be expressed as likelihood on the initial density field, δ_i conditional on the observed flux, $\delta_{F,\mathrm{obs}}$, and observed galaxy field, $\delta_{g,\mathrm{obs}}$. This can be expressed as the sum of three individual log-likelihoods, a composite likelihood (Lindsay 1988; Varin et al. 2011), as

$$\mathcal{L}_{comb}(\delta_i|\delta_{g,\mathrm{obs}}, \delta_{\mathrm{Ly}\alpha,\mathrm{obs}}) = \mathcal{L}_{\mathrm{prior}}(\delta_i)$$
$$+\mathcal{L}_{gal}(\delta_i|\delta_{g,\mathrm{obs}}) + \mathcal{L}_{\mathrm{Ly}\alpha}(\delta_i|\delta_{\mathrm{Ly}\alpha,\mathrm{obs}}). \quad (1)$$

Here, we choose $\mathcal{L}_{\mathrm{prior}}$ to enforce positivity of the power-spectra band-powers, but we do not impose any strong cosmological constraint on the fields. In full generality, we could also include a cross term between the galaxy field and the Lyα forest field, for example if one expects strong feedback effects where the galaxies effect the observed flux field. At the scales of interest in this work ($\sim 1\, h^{-1}\,\mathrm{Mpc}$) there is no indication of a strong proximity effect nearby galaxies, but it could be included e.g. if known quasars are in the volume (see Schmidt et al. 2019).

2.1.1. *Lyα Likelihood*

The Lyα likelihood term is a comparison between the observed Lyα forest flux and the reconstructed flux. This reconstructed flux is derived from the initial density field via a four step process, which is identical to that in TARDIS-I. The initial density field is evolved to the target redshift via flowPM, a GPU implementation of the fastPM framework (Feng et al. 2016) The evolved density is transformed to Lyα optical depth, τ, via the fluctuating Gunn Peterson approximation, $\tau = A\delta^\gamma$, with parameters $A = 0.226$ and $\gamma = 1.5$. The optical depth is transformed to red-shift space using the evolved velocity field at the target redshift. The optical depth is transformed to flux, $F = \exp(-\tau)$, and the observed lines of sight are selected.

We can be expressed the log-likelihood in the Gaussian approximation as a chi-squared comparison of the observed flux with our forward modelled flux

$$\mathcal{L}_{\mathrm{Ly}\alpha}(\delta_i|\delta_{\mathrm{Ly}\alpha,\mathrm{obs}}) = \sum_n \frac{(\delta_{\mathrm{Ly}\alpha,obs}(n) - \delta_{\delta_{\mathrm{Ly}\alpha,rec}}(n))^2}{\sigma_{obs}(n)^2}, \quad (2)$$

where the sum over n is over all individual pixel data and $\sigma_{obs}(n)$ is a noise level estimate from the spectra data reduction. We could also in full generality include off-diagonal elements in this likelihood to model additional effects such as continuum fitting errors.

2.1.2. *Galaxy Likelihood*

For the reconstruction we want to roughly take into account the mapping from density to galaxies. To fully model this would require a differentiable (and fast) friends-of-friends halo finder and a framework to deal with the stochasticity of the galaxy population model. For simplicity, we will instead use a linear and quadratic bias term to model our real space galaxy field given the matter field at each step of the reconstruction. We will work in Lagrangian space , i.e.

$$\delta_g(\boldsymbol{x}) = b_1 \delta_L(\boldsymbol{x}) + b_2^2 \delta_L^2(\boldsymbol{x}). \quad (3)$$

The parameters b_1 and b_2 could be fit jointly, or calibrated via fitting the biased matter power spectra at some fiducial cosmology to the observed galaxy field. In addition, one could use a variant of Lagrangian halo bias estimators as explored in Modi et al. (2017). In addition to this model, we explored a more standard Eulerian bias scheme, but found better agreement across a range of scales with a Lagrangian bias (see Desjacques et al. 2018 for a thorough description of biasing schemes).

In analogy with the flux field described in Section 2.1.1, we convolve the galaxies with the peculiar velocity field to map to redshift space for comparison with the data. Here we only use objects with a spectroscopically confirmed redshift in our likelihood, and assume no error on that redshift. In practice, galaxy redshift errors ($\lesssim 100 \text{ km s}^{-1}$ or $\lesssim 1 \ h^{-1} \text{ Mpc}$) from near-infrared spectroscopy are small compared to our forward model particle resolution thus we expect it to have minimal effect on our final reconstruction.

We can compare our reconstructed galaxy field to the true galaxy field through a simple χ^2 calculation with Poisson error-bars derived from the mock observed number counts. For a real survey, the noise covariance would also include a sizable contribution from the selection function, derivable from the spectrograph properties and survey strategy.

2.2. *Changes to Optimization and Forward Model*

We have also further optimized our reconstructions with the following additional changes and refinements in comparison with the method first presented in TARDIS-I.

1. Our code has been completely rewritten in TensorFlow, allowing future easy integration with machine learning techniques and allowing optimization to utilize Graphical Processing Units (GPUs) for faster optimization. The underlying optimization is still performed with SciPy's LBFGS implementation.

2. In order to implement our code on a GPU, we are using FlowPM Modi et al. (2020)[2] for our dynamical forward model. This is a version of FastPM (Feng et al.
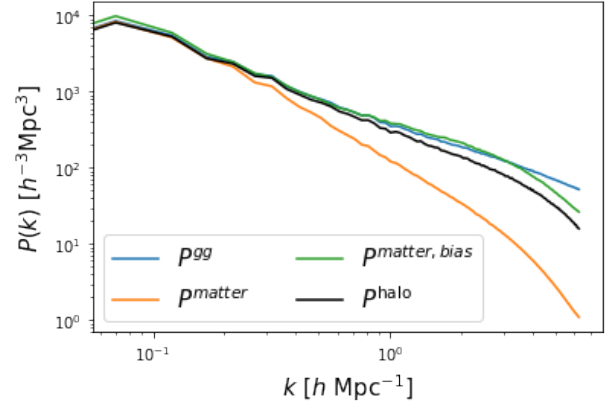
Figure 1. Calibration of galaxy bias by adjusting parameters b_1 and b_2 of the matter density field power spectra to match the observed galaxy field power spectra. While agreement breaks down at small scales, this is well below the $\sim 2.0 \ h^{-1} \text{ Mpc}$ scale of interest in this work.

2016) written in TensorFlow to allow easy automatic differentiation since the forward model is a composition of analytical functions.

3. We have adjusted our optimization to perform a multi-scale optimization where at each step in the optimization we smooth the resulting density field in steps, progressing from 5 Mpc/h smoothing to no smoothing. This is inspired by work done in Modi et al. (2018). The smoothing range we use was found through empirical testing to give the most accurate reconstruction.

4. Changes in the reconstruction implementation have significantly reduced reconstruction time, allowing us to reconstruct larger volumes. In this work we analyze $L = 128 \ h^{-1} \text{ Mpc}$ cubes as opposed to $L = 64 \ h^{-1} \text{ Mpc}$ in Horowitz et al. (2019b), and each full reconstruction takes approximately 10 minutes on a single NVIDIA Tesla V-100 GPU[3].

3. MOCK DATASETS

In this paper, we focus our reconstruction efforts on mock survey data that simulate the planned Galaxy Evolution Survey to be carried out as part of a planned Subaru Strategic Program (SSP) on the 8.2m Subaru Telescope's upcoming 2400-fiber Prime Focus Spectrograph (Sugai et al. 2015). The overall PFS Galaxy Evolution Survey is planned[4] to

---

[3] Memory considerations limit larger volumes at the same particle resolution, but in practice large volumes can be reconstructed by concatenating smaller subvolumes with overlap to reduce edge effects.

[4] Note that while Takada et al. (2014b) outlines an early version of this survey, it is no longer up-to-date and pre-dates the PFS IGM tomography program.

observe three separate extragalactic fields of $\sim 4-5 \deg^2$ each, with multiple visits to build up $\sim 350,000$ medium-resolution spectra (covering 380nm-1.25μm) over multiple target classes ranging from $0.7 \lesssim z \lesssim 6$. Of pertinence to us are the IGM tomography background and foreground components designed to cover the $2.2 < z < 2.7$ redshift range[5], which we will aim to simulate in the subsequent mock reconstructions.

For the density field reconstructions, we run high resolution nbody simulations using a 10-step FastPM simulation with a particle resolution of 256^3 over a $(128 \text{ Mpc/h})^3$ volume. FastPM has quickly become a standard tool for large mock catalog generation (Modi et al. 2019a) and has been demonstrated to reconstruct the statistical properties of other simulation methods with sufficient stage number (Feng et al. 2016). As we will not be studying small scale ($\lesssim 500$ kpc) features within individual galaxy halos, fastPM provides a straightforward tool which can be easily interfaced to nbodykit to analyze the properties of group- and cluster-sized halos.

To populate our simulation with galaxies, we use the nbodykit friends-of-friends algorithm with linking length 0.2 and minimum particle number of 5 in order to find halos. We then use the Zheng et al. (2007) halo occupancy density model with $\alpha = 0.5$ and $\sigma_{\log M} = 0.40$ to populate the halos with galaxies using halotools (Hearin et al. 2017). From this catalog we can then model a realistic survey by uniformly down-sampling those galaxies, combining centrals and satellites with equal weight, to match the expected number density of our desired survey. We use the expected number density from the 'IGM foreground' target class of the Subaru PFS Galaxy Evolution Survey as our fiducial model, which corresponds to a projected density of 1000 galaxies per square degree coeval with the IGM tomography redshift of $2.1 < z < 2.6$. This is equivalent to a comoving number density of $n_{\text{gal}} \approx 4.5 \times 10^{-4} \ h^3 \text{ Mpc}^{-3}$ within the redshift range. Within our $V = (128 \ h^{-1} \text{ Mpc})^3$ simulation volume, we therefore assume a sample of 1000 foreground galaxies.

We then apply a redshift-space distortion along the line of sight based on the velocity of the underlying halos plus velocity dispersion. Note that at high redshift the relative impact of these distortions is small as most structure is still in the early stage of collapse and has relatively small line of sight velocity.

During our reconstruction we do not explicitly model individual galaxies but instead use them as a field. Computing the power spectra of the observed galaxy field, we approximately match the shape and amplitude of the matter power spectra with the choice of $b_1 = 0.301$ and $b_2 = 1.05$ (match-

ing up to $k \approx 3.0$ (h / Mpc)) for the matter field. We show this comparison in Figure 1.

We infer the hydrogen Lyα optical depth for the tomography using the Fluctuating Gunn Peterson Approximation (FGPA), with $T = T_0(\rho/\bar{\rho})^{(\gamma-1)}$ with slope $\gamma = 1.6$ (Horowitz et al. 2019a; Lee et al. 2015). This is then redshift space distorted and exponentiated to generate the underlying flux field. This approximation, as opposed to a fully hydrodynamical simulation, is justified in Appendix A where the same reconstruction is performed on a hydrodynamical simulation. We do not use this field in this main work as we have found that the reconstruction quality does not appreciably suffer when using the fully hydrodynamical simulations, while the dark matter only n-body simulations allow direct comparison to previous works.

We then compute the flux field from the optical depth, $F = \exp(-\tau)$. The flux field varies from 0 to 1, with 0 being complete absorption and 1 being complete transmission of the background spectrum. We select random X-Y coordinate pairs, and take the flux along the Z-direction as one mock skewer. We select 2700 such skewers within our $L = 128 \ h^{-1} \text{ Mpc}$ volume, in order to achieve the average sightline separation of the proposed PFS IGM tomography survey of $\sim 2.4 \ h^{-1} \text{ Mpc}$. Note that it may seem counterintuitive that the number of Lyα forest skewers outnumbers the number of coeval galaxies (1000) within the same assumed survey, but this is due to the fact that each Lyα forest sightline probes a redshift range of $\Delta z \sim 0.4 - 0.5$. Therefore, within our simulation box which corresponds to $\Delta z = 0.14$ along the line-of-sight, Lyα forest sampling comes from a relatively wide range of background source redshifts (see Lee et al. 2014b).

We follow the procedure detailed in Horowitz et al. (2019a) to add pixel noise to each mock skewer. In summary, each skewer is assigned a Gaussian noise level, which is constant across the skewer. To determine each skewer's noise level, we follow Stark et al. (2015a) and Krolewski et al. (2018)), by drawing an S/N ratio from a power-law distribution. Between a minimum and maximum S/N (S/N_{min} and S/N_{max}), the distribution follows a power-law: $dn_{\text{skewer}}/dS/N = S/N^\alpha$, with $\alpha = 2.7$. To replicate the noise properties of the Prime Focus Spectrograph, we then assign a minimum/maximum S/N of $S/N_{\text{min}} = 2$, and $S/N_{\text{max}} = 10$. Finally, we draw values for each pixel in the skewer from a Gaussian distribution with a standard deviation equaling the noise level, and add those values to the flux.

In order to model the error resulting from misfitting a background quasar or galaxy's spectral flux continuum, we apply the continuum error model of Krolewski et al. (2018) to the mock skewers. The flux values within each skewer is offset

---

[5] See the Appendix in Nagamine et al. (2020) for more details on the PFS IGM tomography program.
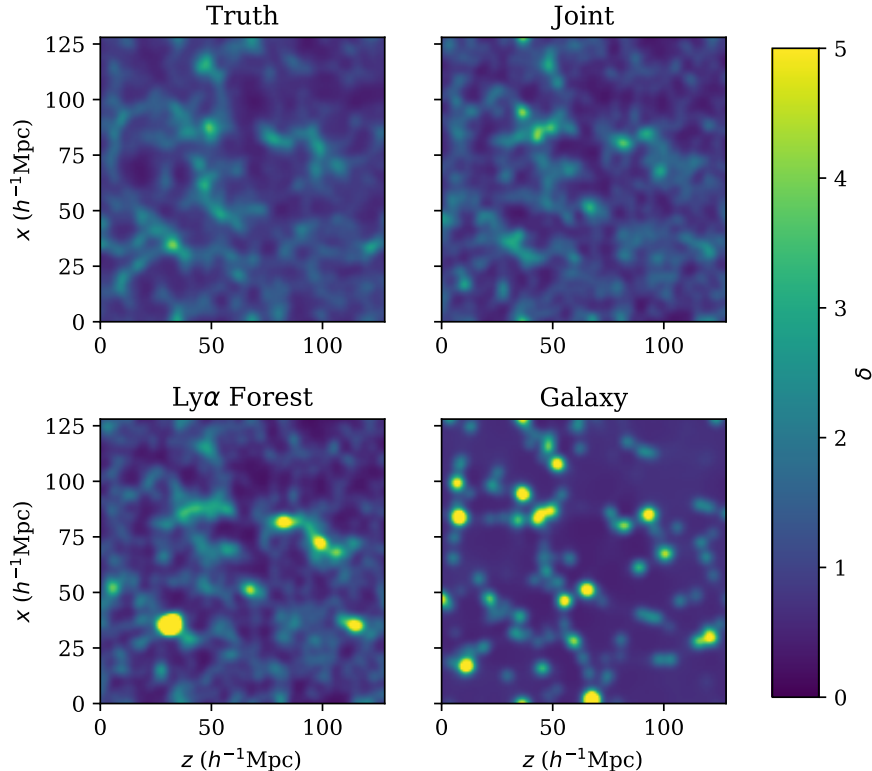
Figure 2. Average $z = 2.5$ density field across a $10\ h^{-1}\,\mathrm{Mpc}$ slab projections through the y axis for the three reconstructions, with the true density field for comparison. The joint reconstruction clearly yields a better match to the true density field than using the Lyα forest and galaxy data individually. These density fields were smoothed at $2.0\,h^{-1}\,\mathrm{Mpc}$.

such that the final observed flux is

$$F_{obs} = \frac{F}{1 + \delta_c},\qquad(4)$$

with δ_c being a value drawn from a Gaussian distribution with width

$$\sigma_c = \frac{0.205}{S/N} + 0.015.\qquad(5)$$

We emphasize that this is a separate source of error from the pixel noise; σ_c is added to the likelihood noise array σ_{obs}, but is not included in the pixel noise level used to draw noise for each pixel in a skewer.

4. RESULTS

We apply our updated TARDIS algorithm (hereafter "TARDIS-II"), as described in Section 2 to the mock catalogs described in Section 3. We explore a number of different properties of the reconstruction, both in terms of the mode reconstruction properties as well as in terms of $z = 2.4$ cosmic web classification. Since we are solving for the initial density fields that give rise to those structures, our reconstruction also will give us the velocity field at $z = 2.4$ and allow us to further evolve our particles to $z = 0.0$.

4.1. *Density Reconstruction*

We qualitatively show the matter density fields of our different reconstructions in Figure 2. Visually, the Lyα forest reconstructions does a good job of recovering the filamentary structures at close to mean density, but at overdensities the amplitude of the recovered structures is inaccurate. This is due to the known effect that the Lyα forest saturates at relatively mild overdensities, therefore making overdensities ill-constrained. The galaxy-only reconstructions, in contrast, appear as peaks in the matter density field where galaxies trace strong overdensities, but does not recover more average densities and underdensities. The joint reconstructions using both Lyα forest and galaxy information, on the other hand, yields both a good recovery of the overdensities as well the lower-density structures.

To assess the quality of our reconstructions, we first directly compare, on a cell-by-cell basis, the real space matter density field between the simulated truth and our reconstructions for each data combination. This is shown in Figure 3. The joint reconstruction has significantly less bias and variance in the recovered matter field than either individual reconstructions. We can quantify the variance properties with the Pearson correlation coefficients, finding [0.72, 0.56, 0.54] for the joint, Lyα forest, and galaxy reconstructions respec-

tively — a coefficient of 1 would indicate a perfect density reconstruction. The density fields in the joint reconstructions clearly have a more linear relationship with the true density field.

To measure the quality of our reconstruction with the variety of datasets as a function of scale we use two-point statistics to quantify the fields. We use the reconstructed auto power spectra, P_{RR}, and cross-correlation coefficient defined in terms of the crosspower (P_{RT}) and auto-power (P_{TT} for the truth) as;

$$r_c = \frac{P_{RT}}{\sqrt{P_{RR}P_{TT}}}. \tag{6}$$

We show the auto-power in Figure 4 and for the cross-correlation coefficient in Figure 5. We find a comparable agreement across a range of scales, with the joint dataset having the best reconstruction across all scales, outperforming either probe individually. This suggests that there is a synergistic reconstruction, where our dynamic forward model is able to use the gravitational phase coupling information to inform the final optimized field.

Because TARDIS-II chooses a random Gaussian random field as its starting point, we expect to find some variance in a reconstruction for a particular dataset. In order to quantify the level of this variance, we run TARDIS-II 30 times on the mock catalogs with a random starting point, and calculate the cross-correlation coefficient for each run. We then consider the maximum difference in r_c across runs for each scale, $\Delta r_c(k)$. For each individual probe, we find that Δr_c remains at the $\sim 10^{-2}$ level across all scales. The joint dataset has a Δr_c which is smaller than each individual probe, at the $\sim 5 \cdot 10^{-3}$ level. This suggests that the joint reconstruction has a synergistic effect for both reconstruction quality, and reconstruction variance.

4.2. *Cosmic Web Classification*

For a quantitative comparison of the cosmic web recovery in TARDIS-II, we use the eigenvalues and vectors of the pseudo-deformation tensor as described in Lee & White (2016); Krolewski et al. (2017) and inspired by work in Bond et al. (1996); Hahn et al. (2007); Forero-Romero et al. (2009). This tensor has a strong physical interpretation within the Zel'dovich approximation (Zel'dovich 1970) and, while there exist other cosmic web classification algorithms (see summary in Cautun et al. 2014), we choose to use the eigenvectors/values of the deformation tensor since it provides a continuous field which can be compared easily point-wise. Unlike past work, such as Lee & White (2016); Krolewski et al. (2017), we will be working directly with the reconstructed density field rather than in flux. All the density fields used in our comparisons will be smoothed with a $R = 2.0 \ h^{-1}$ Mpc Gaussian kernel.

The (pseudo-)deformation tensor is the Hessian of the underlying gravitational potential,

$$D_{ij} = \frac{\partial^2 \Phi}{\partial x_i \partial x_j}, \tag{7}$$

or equivalently in Fourier space in terms of the density field, δ_k, as

$$\tilde{D}_{ij} = \frac{k_i k_j}{k^2} \delta_k. \tag{8}$$

The eigenvectors of the deformation tensor relate to the principle curvature axes of the density field at each point, corresponding in the Zel'dovich approximation with the principle inflow/outflow directions. The corresponding eigenvalues determine if the net flow is inward or outward. Points with three eigenvalues above some nonzero threshold value λ_{th} (as in Forero-Romero et al. 2009) are classified as nodes (i.e. (proto)clusters), two values above λ_{th} are filaments, one value above λ_{th} are sheets, and zero values above λ_{th} are voids. The use of this thresh-hold value allows for the rescaling of the relative number of each cosmic web type and allowing comparisons without an additional complication of overall normalization. Note that we could use the reconstructed velocity field to determine net inflow and outflow in a more physically accurate sense (i.e. beyond the Zel'dovich approximation), but keep the deformation tensor approach to maintain ease of comparison with other works (such as the Wiener filter). We follow past work (Lee & White 2016; Krolewski et al. 2017; Horowitz et al. 2019a) and define our threshhold value λ_{th} for each reconstruction such that the voids occupy 21% of the total volume at $z = 2.5$ (see also Cautun et al. (2014)).

We calculate the deformation tensor on each reconstructed density field as well as the true density field, smoothed at 2.0 h^{-1} Mpc with a Gaussian kernel. We then choose a threshold λ_{th} for each reconstruction in order to keep the void fraction to 21%. We show the qualitative results of this analysis in Figure 6 for a given slice through the reconstruction, where one can see visually the improved reconstruction of cosmic structure with the joint analysis. A quantitative volume overlap comparison is shown in the confusion matrix in Figure 7, as well as common misclassifications. When weighted by volume type, we find agreement of [68%, 63%, 58%] for exact match in classification for Joint, Lyman-α forest, and Galaxy reconstructions, respectively. Significant misclassification, defined as classifications off by more than one eigenvalue sign (i.e. nodes mistaken as a sheet or void), are rare and comprise only [1.1%, 2.0%, 3.3%] fraction of the volume for the joint, Lyα forest and galaxy reconstructions respectively. This high-quality cosmic web reconstruction should allow PFS to confidently study the variation of galaxy properties with respect to their cosmic web environment at $z \sim 2.5$ (or lack thereof, e.g. Martizzi et al. 2019).
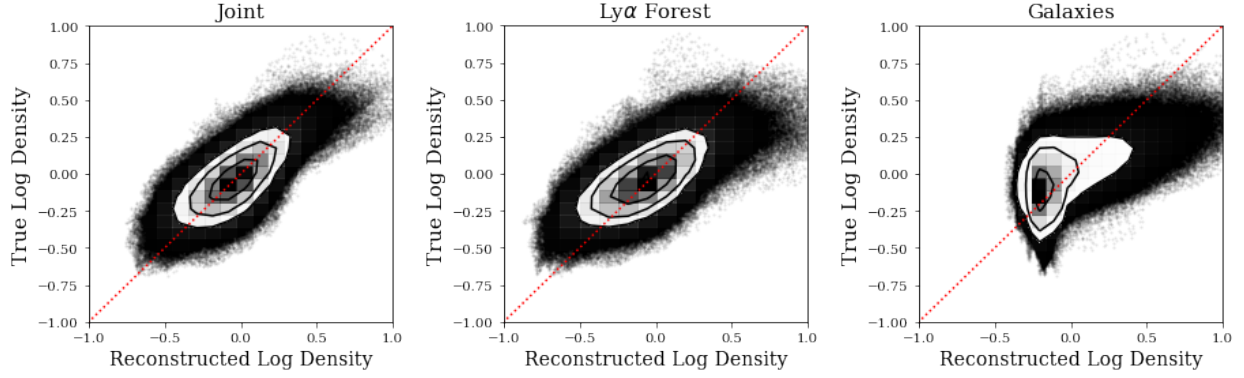
Figure 3. Scatter plot showing the reconstructed densities in the simulation volume in comparison with the the true density in each grid cell. The joint reconstruction clearly gives a more accurate and unbiased reconstruction of the density field compared with the Lyα and galaxy reconstructions alone. All the densities have been smoothed on a 2 h^{-1} Mpc scale, which will be used for the later cosmic web analysis.

Table 1. Cosmic Web Recovery at $z = 2.5$ (Eulerian Comparison)

Mock Data	Pearson Coefficients			Volume Overlap (%)			
	λ_1	λ_2	λ_3	Node	Filament	Sheet	Void
Joint	0.853	0.796	0.708	51	73	70	62
Lyα	0.771	0.718	0.613	46	70	65	58
Galaxy	0.746	0.635	0.518	37	58	58	58
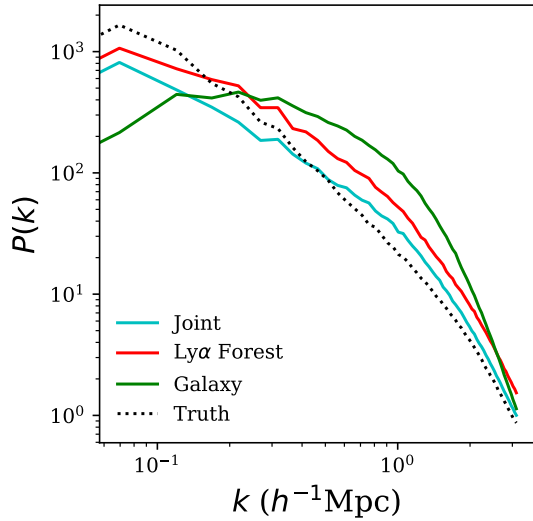
Figure 4. Power spectra of the reconstructed fields in comparison to the true power spectra at $z = 2.5$. Note that while Lyα forest and galaxies-only reconstructions have a significant excess of power at mid/small scales, the joint analysis has only a mild excess.

Figure 5. Correlation coefficient defined in Equation 6, showing agreement between the two fields as a function of scale at $z = 2.5$. Complete agreement would correspond to an $r_c = 1$. Each individual probe has a regime of scale with better reconstruction quality; galaxies appear to reconstruct small scale features while Lyα forest is able to reconstruct the largest modes.

In Horowitz et al. (2019a), we had studied the possibility of tracking the trajectories of coeval $z \sim 2.5$ galaxies to their $z = 0$ environments. While we do not explicitly perform the same analysis here, on the basis of the superior cosmic web recovery of the joint reconstruction method we estimate
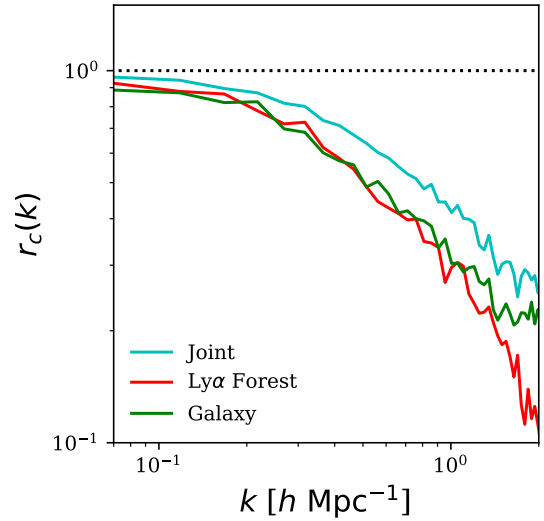
that $\sim 50\%$ of the $z \sim 2.5$ galaxies can be predicted to their correct $z = 0$ environment.
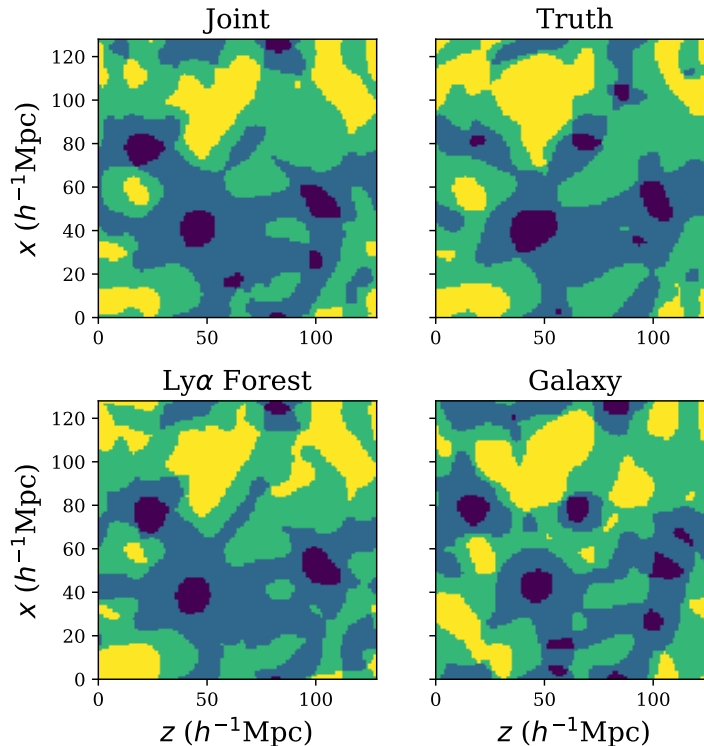
Figure 6. Cosmic Web Structures from a single 1 h^{-1} Mpc slice through our volume. Dark blue indicates node, light blue indicates filament, green indicates sheet, and yellow indicates void regions. Qualitatively, the joint reconstruction better reconstructs the topology of the web.
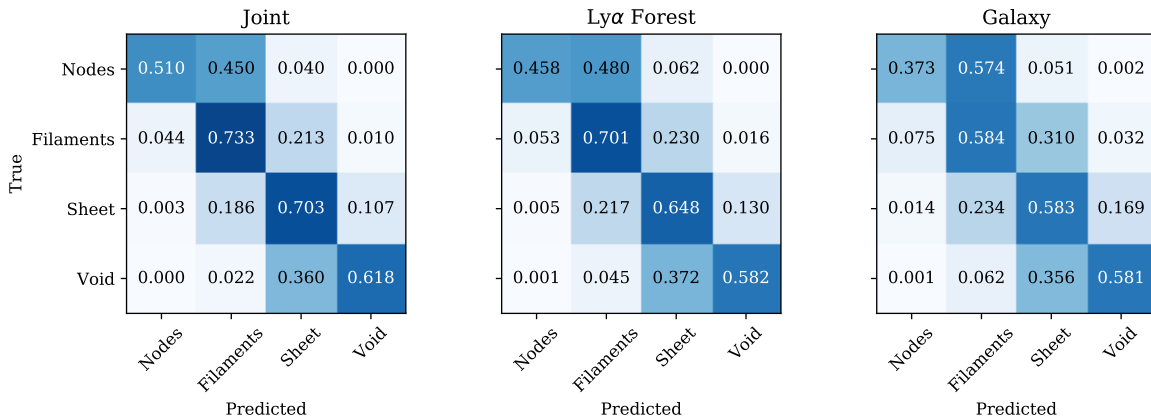

Figure 7. Confusion matrix for cosmic structures; numbers on the diagonal the fraction of each true cosmic structure correctly identified by the reconstruction, while those on the diagonal reflect the faction of each structure misidentified. The joint reconstruction appears to outperform either field individually across all cosmic structure types.

We can also study the alignment of the reconstructed cosmic web with the true cosmic web via cell-by-cell comparison of the characteristic eigenvectors of the pseudo-deformation tensor. Matching the direction of the eigenvectors is particularly important for intrinsic alignment studies, where the direction of the filament is oriented along $\hat{\mathbf{e}}_3$ (e.g., Forero-Romero et al. 2014). Figure 8 shows the dot product between the true and reconstructed eigenvectors, finding that the joint analysis provides substantial improvement in the

recovery of the eigenvectors. In $\sim 80\%$ of the grid cells, the recovered $\hat{\mathbf{e}}_1$ and $\hat{\mathbf{e}}_3$ are aligned to within $\cos\theta > 0.9$ of the true eigenvectors. This is comparable to the accuracy forecasted by Krolewski et al. (2017) for a TMT-like Lyα forest-only survey using Wiener filtering, although note that in this paper we actually adopt smaller smoothing scales (2 h^{-1} Mpc vs 4 h^{-1} Mpc). Note that $\hat{e}_1$ and $\hat{e}_3$ has similar recovery, which has been attributed to the primary orientation for filaments and sheets respectively, while $\hat{e}_2$ has a
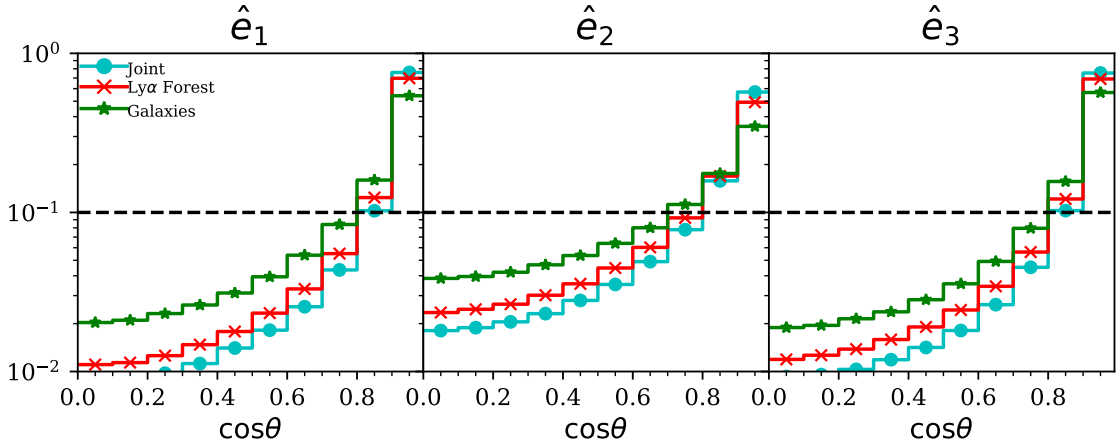
Figure 8. Cosine of the angle between the reconstructed eigenvectors and true eigenvectors of the density pseudo-deformation tensor. Random orientations would correspond to the black dotted line, while complete agreement would correspond to all density in the far right bin (i.e. $\cos\theta = 1$). The galaxy reconstructions perform considerably worse than the Lyα forest since it doesn't properly recover filamentary structures, but the joint reconstruction significantly outperforms either probe individually.

more ambiguous geometric interpretation and worse recovery. This therefore bodes well for the ability of Subaru PFS to constrain intrinsic alignments between the galaxies and the underlying cosmic web during the 'Cosmic Noon' epoch of $z \sim 2 - 3$ (e.g., Codis et al. 2018), especially since the survey will deliver both the Lyα forest sightlines and foreground galaxy sample over a large volume.

4.3. *Cluster and Protocluster Reconstruction*

A common application of Lyα tomography is the study of massive overdensities, such as galaxy (proto)clusters, at high redshift (as highlighted in works such as Lee et al. 2018; Mukae et al. 2020; Newman et al. 2020b; Ravoux et al. 2020). In this section we examine various statistics related to the reconstruction characteristics of the TARDIS-II technique to our mock catalogs.

4.3.1. *Cluster Reconstruction*

Since we have no prior on the underlying matter power spectrum, the lowest order statistic to check is the distributional qualities of the reconstructed massive halos in the form of a halo mass function. We use a friends-of-friends halo finder on our resulting particle positions at $z = 0$ and $z = 2.5$, with linking length of 0.2. In Figure 9, we show the halo mass function for the three reconstructions in comparison to the truth. For the joint reconstruction we find the mass function is consistent with the truth at $M > 10^{12.5} M_\odot$, indicating very good agreement across a wide range of mass scales of astrophysical and cosmological interest, while for either galaxies or Lyα alone there are significant residual biases.

4.3.2. *Protocluster identification and reconstruction*

Although we are reconstructing the density and velocity field using $z = 2.5$ mock observables, we can further evolve the resulting fields to $z = 0$ to provide a rigorous definition of "proto-cluster" as the progenitor of modern-day galaxy clusters. In this case we will take any halo identified as a "cluster" at $z = 0.0$ by our friends-of-friends halo-finder and identify the corresponding particles at $z = 2.5$ to define our proto-cluster region. We show projection plots of the most massive halos in our reconstructions in Figure 11. These show the $z = 2.5$ real-space dark matter density and redshift-space flux fields, as well as the evolved $z = 0.0$ late time structure of these clusters. To compare masses of individual identified clusters at $z \sim 2.5$ we need to match analogous structures from our reconstructions and the simulated truth. While in abstract it is possible to perform this analysis with the friends of friends halo catalog used above, in practice there is some uncertainty due to the variation in the sub-structure within the halo causing slightly different halos to be identified (i.e. one large halo in the mock could be identified as two adjacent halos in the reconstruction, or vice-versa). To correct for this effect, we calculate protocluster mass by summing over the $z = 2.5$ particle number within ten times the halo radius of each $z = 0$ halo (i.e. an adjusted spherical over-density mass) at the $z = 0$ Eulerian position. The choice of this radii was chosen to allow slight shifts in real-space position between the reconstructed and true halo positions. We show the protocluster mass recovery of the reconstruction in Figure 10. This indicates good agreement across a wide range of masses with little residual bias except at the very high mass end, which suffers from small-number statistics. This plot could be roughly compared to Figure 9 in Lee et al. (2016), which included additional calibrations to determine a "tomographic mass." They found roughly a factor 2 wider spread in
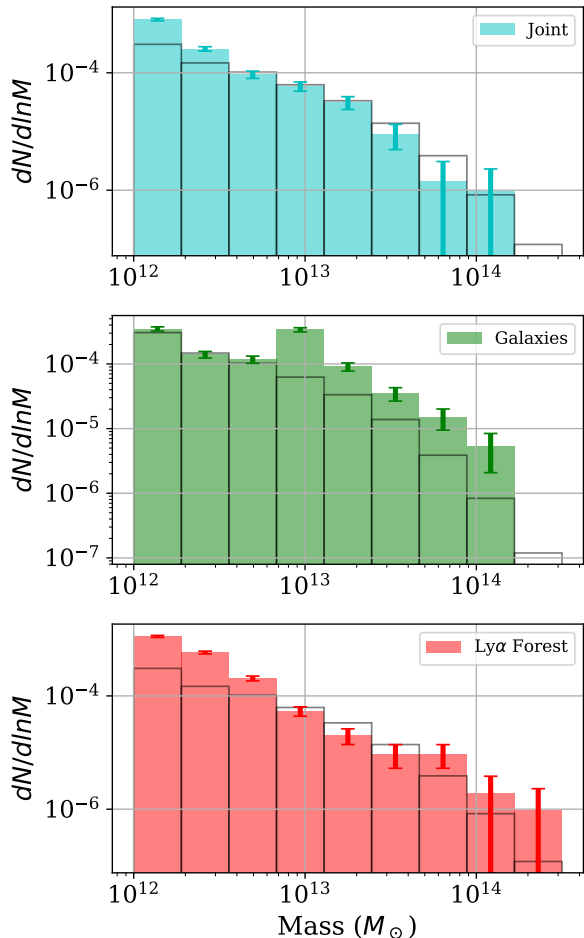
Figure 9. Corresponding $z = 2.5$ halo mass functions for the three reconstructions found via friends of friends halo finding. Black box represents the true halo mass function, while error bars are Poissonian. The joint reconstruction yields a halo mass function that is consistent with the true distribution at $M > 10^{12.5}\, M_\odot$.

calculated cluster mass compared to our approach, although this did not include the galaxy field information.

5. CONCLUSION

In this work we have demonstrated the possibility of jointly reconstructing a density field from multiple disjoint tracers using a maximum likelihood framework. By combining a $z \sim 2.5$ Lyα forest survey and an overlapping galaxy catalog that mock up the upcoming Subaru PFS Galaxy Evolution Survey, we have shown that it is possible to reconstruct a more accurate 3D map than either one individually. We attribute this reconstruction quality to the multi-scale properties of our different unique tracers. Throughout all metrics we analyzed (power-spectra, cross-power, cosmic structure classification, and halo mass function) we found the joint optimization provides higher fidelity and less biased reconstructions. In light of upcoming surveys with overlapping
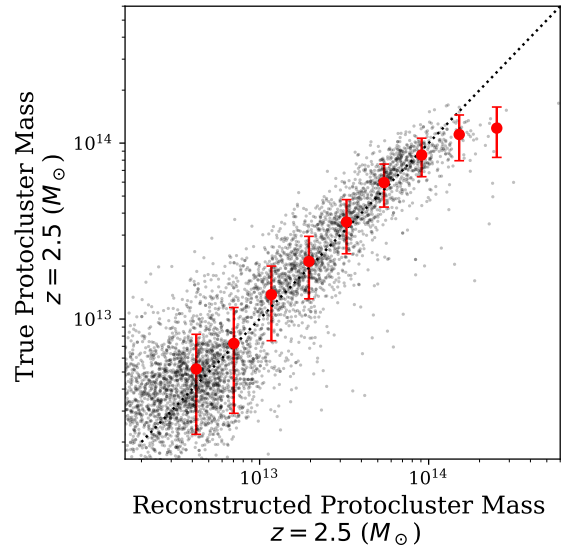

Figure 10. Comparison of the 5000 most massive identified $z = 2.5$ protoclusters in the joint reconstructed volume, as identified by the predicted $z = 0.0$ cluster positions, with the corresponding volume in the true field. Note that this mass is spherical overdensity mass, as opposed to FOF mass in Figure 9.

Lyα forest information and galaxy spectra, these joint analyses seem like a promising way to extract the optimal information.

In this work we use a simple biasing scheme to assign galaxies to the density field in our forward model. A more physical approach may be possible by using a more nuanced differential model to identify halos, such as a neural network as done in Modi et al. (2019b), and populate them according to a halo occupancy density model. This latter step is non trivial as it requires a differentiable stochastic model, but work in this direction is ongoing. However, it is useful to note that even with this extremely simple model we are able to achieve surprisingly accurate reconstructions of the cosmic web despite the sparse number densities used. In addition, more accurate object-by-object bias estimates could be extracted from spectra and photometry to better inform our reconstructions within the current framework.

Alternatively, instead of foreword modeling the HOD model, one could use a more nuanced biasing schemes such as that done in Kitaura et al. (2019); Ata et al. (2020). In these works a non-linear polynomial bias is used in Lagrangian space as well as a non-local tidal field bias encoded in the displacement fields to map the galaxy tracers from the Eulerian to the Lagrangian frame. As long as the bias model is a differentiable function of the initial density field, it could be included in TARDIS. While not implemented in this work, it is also possible to jointly optimize for these bias parameters while solving for the initial density field. Note
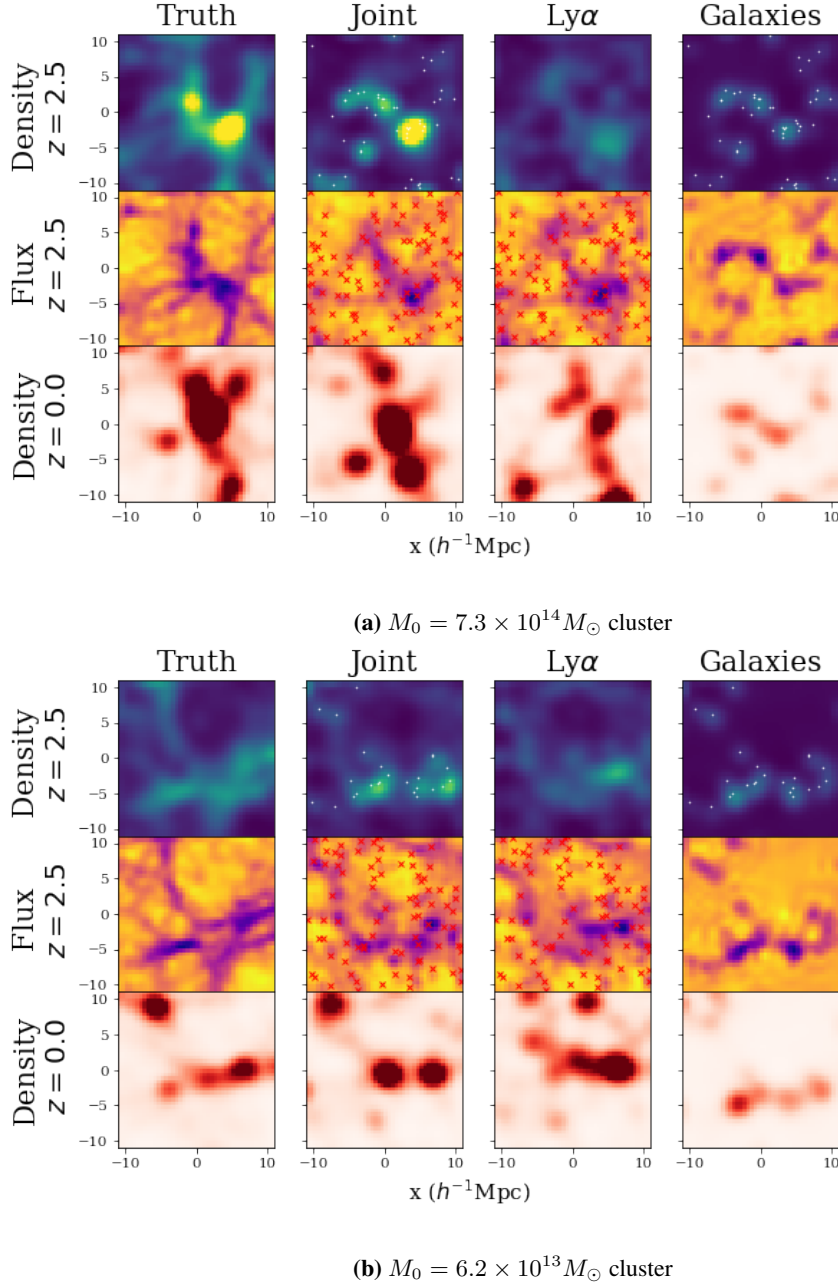
(a) $M_0 = 7.3 \times 10^{14} M_\odot$ cluster

(b) $M_0 = 6.2 \times 10^{13} M_\odot$ cluster

Figure 11. Reconstructions of protocluster regions as identified by their predicted $z = 0$ particle locations, projected over a 20 h^{-1} Mpc box in the line of sight direction. Clusters are categorized by their friends-of-friends halo mass at $z = 0$, M_0, and the same Eulerian region of space is displayed at z=2.5 in flux and density. In the top row, the positions of observed galaxies are shown as white dots and Lyα lines of sight are shown as red crosses. **(a)**; shows the most massive cluster in our mock volume, while **(b)** shows a more typical halo.

that this will likely introduce degeneracies which would require additional consideration, possibly including using a sampling-based optimization method.

Similarly, we assume the Fluctuating Gunn Peterson Approximation to model the Lyα forest in our simulations, which is known to break down in massive environments. (Sorini et al. 2018) While we explore the impact of these deviations briefly in Appendix A, finding the effect on cosmic structure reconstruction small, for more futuristic po-

tential surveys (i.e. TMT/GMT/EELT) it could be a significant source of error. In addition, a more detailed study of reconstructed cluster interiors would likely show significant systematic deviations when assuming FGPA. Depending on the application, a more nuanced model could be implemented to map from dark matter to hydrodynamical quantities; higher order analytical approximations, gradient based methods (Dai et al. 2018), or deep learning methods (i.e. Thiele et al. (2020)).

It is interesting to note that we are able to get high fidelity reconstruction quality through using an "off the shelf" optimization scheme within SciPy's optimizer framework (Virtanen et al. 2020). While in this work we did include an annealing scheme which improved reconstruction of large scale modes, this did not require any fine tuning to get satisfactory reconstruction quality, and wasn't used in TARDIS-I, the earlier iteration of our algorithm. Related techniques using Hamiltonian Monte Carlo for Lyα flux reconstruction have also been developed (Porqueres et al. 2019, 2020) but this approaches have significantly increased computational costs and it is currently unclear what benefits there are compared to an optimization approach for cosmic web recovery.

These reconstructions also provide a possible powerful cosmological tool as we are reconstructing the initial density field corresponding the observed region which should contain all cosmological information. This aspect was a driving motivator for earlier interest in such reconstructions (Seljak et al. 2017), and would be possible to do with our reconstructions as well. Residual biases in Figure 4 could in theory be removed and associated covariance calculated within a response formalism (see Seljak et al. (2017); Horowitz et al. (2019b)). We leave this potentially promising avenue to future paper(s).

It is important to stress that the late time reconstructions are constrained to follow statistical properties of ΛCDM by nature of the forward model. For example, void density/velocity profiles would still be forced to follow the standard cosmological model. Deviations from this model (massive neutrinos, modified dark energy, etc.) would require alteration of the forward model. Either this deviation could be parametrized and fit jointly, or the reconstructions could be run separately with each model and the resulting likelihood values compared.

An additional natural extension to this work would be the inclusion of additional tracers beyond Lyα forest and spectroscopic galaxy surveys, such as photometric redshift catalogs and cosmic microwave background lensing. The latter field could be of significant benefit, as the CMB lensing kernel peaks at $z \sim 2.0$ and it provides a direct probe of the matter density field on large scales. Not only could this provide a field complementary to both galaxies and Lyα but it could be used to directly calibrate galaxy biases within the optimization.

APPENDIX

A. HYDRODYNAMICAL EFFECTS

In the main body of the text we used the Flucuating Gunn Peterson Approximation for fast generation of mock catalogs and in our reconstruction. However, as the name suggests, this is only an approximation which is known to break down in various cosmic environments. In this section we briefly explore how well TARDIS performs on reconstructing density fields from Lyman Alpha forests including observational effects. We focus on the z=2.5 density field reconstructions on the map level. To perform this comparison, we will compare reconstruction quality using the full hydrodynamical flux field versus an FGPA transformed dark matter field for two sets of hydrodynamical simulations; NyX and Illustrus. How hydrodynamical effects propagate to our reconstruction quality can then be quantified by the degradation of each statistic.

In order to isolate the hydrodynamical effects arising from feedback processes vs those arising from only baryon pressure, we use two different hydrodynamical simulations for the comparisons. The Nyx code (Almgren et al. 2013), which relies on solving the evolution of a system of Lagrangian fluid elements coupled gravitationally to inviscid ideal fluid. The fluid evolution is modeled using a finite volume representation in an Eulerian framework using adaptive mesh refinement. While this technique allows accurate modeling of gas properties of diffuse structure (such as those which dominate the volume of Lyα studies), it notably does not include feedback processes, such as AGN formation which could dominate at the centers of massive clusters. We use a 100 Mpc/h sidelength box with particle resolution 4096^3 and choose our sightline and noise characteristics to match the PFS-like survey in the main text.

Unlike Nyx, the Illustris simulation does include feedback from AGN and star formation. For this purpose, we adopt the hydrodynamical Illustris simulation (Vogelsberger et al. 2014a,b; Genel et al. 2013; Sijacki et al. 2015), which employs the

Table 2. Cosmic Web Recovery at $z = 2.5$ (Variation of Forward Model)

Mock Data	Pearson Coefficients			Volume Overlap (%)			
	λ_1	λ_2	λ_3	Node	Filament	Sheet	Void
Nyx	0.754	0.692	0.573	43.4	63.5	64.7	53.1
FGPA	0.786	0.709	0.538	47.7	68.2	64.6	44.6

Table 3. Cosmic Web Recovery at $z = 2.5$ (Variation of Forward Model)

Mock Data	Pearson Coefficients			Volume Overlap (%)			
	λ_1	λ_2	λ_3	Node	Filament	Sheet	Void
Illustris	0.739	0.689	0.595	42.4	62.3	63.9	47.7
FGPA	0.794	0.744	0.629	47.5	66.7	67.2	50.6

adeptive mesh refinement code $AREPO$ (Springel 2010). The AGN feedback prescription in Illustris produces too much heating, which is able to remove small-scale structure in the gas (e.g., Genel et al. 2014; Park et al. 2018). Therefore, it should provide a good test case here, since the difference from a non-feedback simulation such a Nyx would be relatively larger compared to other simulations with more realistic feedback prescriptions (see e.g., Sorini et al. 2018, for a detailed comparison between the two).

The dark matter density field from Illustris was deposited onto a grid using the same cloud-in-cell algorithm adopted in Martizzi et al. (2019) with 150 cells on a side and corresponding spatial resolution of 0.5 Mpc/h. We extracted skewers from the simulation along the z-direction with a resolution of 1 km/s on a regular grid separated 0.5 Mpc/h in the x- and y-directions. The skewers were generated using the $fake\_spectra$ package (Bird 2017), which is designed to work on moving mesh simulations like Illustris to use the HI fraction directly from the simulation and takes into account the voronoi kernel adopted in $AREPO$. The Illustris-1 simulation volume has a sidelength of 75 Mpc/h and initially 1820^3 particles, resulting in a gas mass resolution of $1.3 \times 10^6 M_\odot$ (Nelson et al. 2015).

To address how our reconstruction from quality degrades, we compare a reconstruction of using the hydrodynamical modelled optical depth for mock generation to a reconstruction using the Fluctuating Gunn-Peterson Approximation on the corresponding dark matter field for mock generation. This should allow us to compare in what sort of environments our approximate forward model fails.

We show results for our reconstruction in terms of cosmic structure classifications for the nyx simulation mock in Table 2 and for the Illustris simulation mock in Table 3, using the identical procedure as in the main text (i.e. comparable to Table 1). The performance of the two models in this context are quite comparable, with only a slight reduction in cosmic web characterization accuracy, justifying our use of the FGPA approximation in this context. In Figure 12 we show a comparison of the reconstructed density error for both simulation mock catalog reconstructions. In the absence of the feedback processes (i.e. for the nyx simulation), the reconstructed density error is nearly identical whether or not hydrodynamical effects are included. With feedback, there is a slight skew in the reconstructed density, but still not sufficient to appreciably effect reconstructed cosmic structure.

REFERENCES

Almgren, A. S., Bell, J. B., Lijewski, M. J., Lukić, Z., & Van Andel, E. 2013, ApJ, 765, 39

Ata, M., Kitaura, F.-S., Lee, K.-G., et al. 2020, arXiv e-prints, arXiv:2004.11027

Becker, G. D., Bolton, J. S., Madau, P., et al. 2015, MNRAS, 447, 3402

Bielby, R. M., Shanks, T., Crighton, N. H. M., et al. 2017, MNRAS, 471, 2174

Bird, S. 2017, FSFE: Fake Spectra Flux Extractor, ascl:1710.012

Bond, J. R., Kofman, L., & Pogosyan, D. 1996, Nature, 380, 603

Cai, Z., Fan, X., Peirani, S., et al. 2016, ApJ, 833, 135

Cai, Z., Fan, X., Yang, Y., et al. 2017a, ApJ, 837, 71

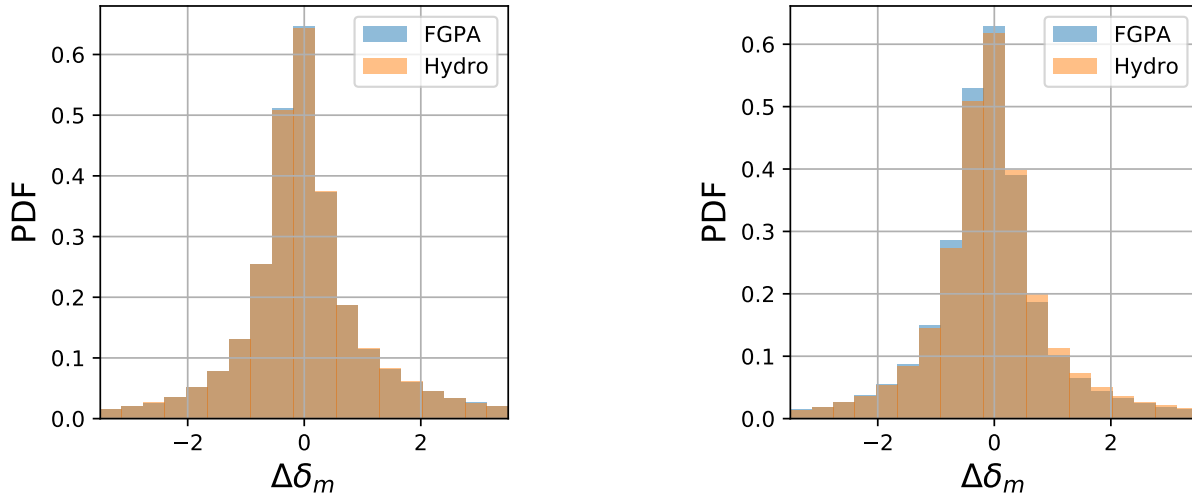Cai, Z., Fan, X., Bian, F., et al. 2017b, ApJ, 839, 131

Figure 12. Histogram comparing pixel error in reconstructed mass using mock catalogs generated off of Nyx (left) and Illustris (right) hydrodynamical models. The presence of feedback within Illustris slightly skews the reconstruction quality vs. those generated from dark matter alone.

Caucci, S., Colombi, S., Pichon, C., et al. 2008, MNRAS, 386, 211

Cautun, M., van de Weygaert, R., Jones, B. J. T., & Frenk, C. S. 2014, MNRAS, 441, 2923

Codis, S., Jindal, A., Chisari, N. E., et al. 2018, MNRAS, 481, 4753

Crocce, M., & Scoccimarro, R. 2006, PhRvD, 73, 063519

Dai, B., Feng, Y., & Seljak, U. 2018, JCAP, 11, 009

de Sainte Agathe, V., Balland, C., du Mas des Bourboux, H., et al. 2019, A&A, 629, A85

Desjacques, V., Jeong, D., & Schmidt, F. 2018, PhR, 733, 1

D'Odorico, V., Viel, M., Saitta, F., et al. 2006, MNRAS, 372, 1333

Evans, C. J., Puech, M., Barbuy, B., et al. 2014, in Proc. SPIE, Vol. 9147, Ground-based and Airborne Instrumentation for Astronomy V, 914796

Fan, X., Strauss, M. A., Becker, R. H., et al. 2006, AJ, 132, 117

Feng, Y., Chu, M.-Y., Seljak, U., & McDonald, P. 2016, MNRAS, 463, 2273

Font-Ribera, A., Arnau, E., Miralda-Escudé, J., et al. 2013, JCAP, 5, 18

Forero-Romero, J. E., Contreras, S., & Padilla, N. 2014, MNRAS, 443, 1090

Forero-Romero, J. E., Hoffman, Y., Gottlöber, S., Klypin, A., & Yepes, G. 2009, MNRAS, 396, 1815

Genel, S., Vogelsberger, M., Nelson, D., et al. 2013, MNRAS, 435, 1426

Genel, S., Vogelsberger, M., Springel, V., et al. 2014, MNRAS, 445, 175

Gunn, J. E., & Peterson, B. A. 1965, The Astrophysical Journal, 142, 1633

Hahn, O., Carollo, C. M., Porciani, C., & Dekel, A. 2007, MNRAS, 381, 41

Hearin, A. P., Campbell, D., Tollerud, E., et al. 2017, AJ, 154, 190

Hernquist, L., Katz, N., Weinberg, D. H., & Miralda-Escudé, J. 1996, ApJL, 457, L51

Horowitz, B., Lee, K.-G., White, M., Krolewski, A., & Ata, M. 2019a, ApJ, 887, 61

Horowitz, B., Seljak, U., & Aslanyan, G. 2019b, JCAP, 2019, 035

Jain, B., & Bertschinger, E. 1994, ApJ, 431, 495

Japelj, J., Laigle, C., Puech, M., et al. 2019, A&A, 632, A94

Johns, M., McCarthy, P., Raybould, K., et al. 2012, in Proc. SPIE, Vol. 8444, Ground-based and Airborne Telescopes IV, 84441H

Kitaura, F.-S., Ata, M., Rodriguez-Torres, S. A., et al. 2019, arXiv e-prints, arXiv:1911.00284

Krolewski, A., Lee, K.-G., Lukić, Z., & White, M. 2017, ApJ, 837, 31

Krolewski, A., Lee, K.-G., White, M., et al. 2018, ApJ, 861, 60

Lee, K.-G., Hennawi, J. F., White, M., Croft, R. A. C., & Ozbek, M. 2014a, ApJ, 788, 49

—. 2014b, ApJ, 788, 49

Lee, K.-G., & White, M. 2016, ApJ, 831, 181

Lee, K.-G., Hennawi, J. F., Spergel, D. N., et al. 2015, ApJ, 799, 196

Lee, K.-G., Hennawi, J. F., White, M., et al. 2016, ApJ, 817, 160

Lee, K.-G., Krolewski, A., White, M., et al. 2018, ApJS, 237, 31

Lindsay, B. G. 1988, Contemporary mathematics, 80, 221

Martizzi, D., Vogelsberger, M., Artale, M. C., et al. 2019, MNRAS, 486, 3766

McConnachie, A., Babusiaux, C., Balogh, M., et al. 2016, arXiv e-prints, arXiv:1606.00043

Miller, J. S. A., Bolton, J. S., & Hatch, N. 2019, MNRAS, 489, 5381

Modi, C., Castorina, E., Feng, Y., & White, M. 2019a, JCAP, 2019, 024

Modi, C., Castorina, E., & Seljak, U. 2017, MNRAS, 472, 3959

Modi, C., Feng, Y., & Seljak, U. 2018, JCAP, 10, 028

Modi, C., Lanusse, F., & Seljak, U. 2020, arXiv e-prints, arXiv:2010.11847

Modi, C., White, M., Slosar, A., & Castorina, E. 2019b, arXiv e-prints, arXiv:1907.02330

Momose, R., Shimizu, I., Nagamine, K., et al. 2020a, arXiv e-prints, arXiv:2002.07334

Momose, R., Shimasaku, K., Kashikawa, N., et al. 2020b, arXiv e-prints, arXiv:2002.07335

Mukae, S., Ouchi, M., Kakiichi, K., et al. 2017, ApJ, 835, 281

Mukae, S., Ouchi, M., Cai, Z., et al. 2020, ApJ, 896, 45

Nagamine, K., Shimizu, I., Fujita, K., et al. 2020, arXiv:2007.14253

Nelson, D., Pillepich, A., Genel, S., et al. 2015, Astronomy and Computing, 13, 12

Newman, A. B., Rudie, G. C., Blanc, G. A., et al. 2020a, ApJ, 891, 147

—. 2020b, ApJ, 891, 147

Park, H., Alvarez, M. A., & Bond, J. R. 2018, ApJ, 853, 121

Pichon, C., Vergely, J. L., Rollinde, E., Colombi, S., & Petitjean, P. 2001, MNRAS, 326, 597

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2018, arXiv e-prints, arXiv:1807.06209

Porqueres, N., Hahn, O., Jasche, J., & Lavaux, G. 2020, arXiv e-prints, arXiv:2005.12928

Porqueres, N., Jasche, J., Lavaux, G., & Enßlin, T. 2019, A&A, 630, A151

Ravoux, C., Armengaud, E., Walther, M., et al. 2020, arXiv e-prints, arXiv:2004.01448

Rollinde, E., Petitjean, P., Pichon, C., et al. 2003, MNRAS, 341, 1279

Schlegel, D., Kollmeier, J. A., & Ferraro, S. 2019, in BAAS, Vol. 51, 229

Schmidt, T. M., Hennawi, J. F., Lee, K.-G., et al. 2019, ApJ, 882, 165

Seljak, U., Aslanyan, G., Feng, Y., & Modi, C. 2017, ArXiv e-prints, arXiv:1706.06645

Seljak, U., Slosar, A., & McDonald, P. 2006, JCAP, 2006, 014

Sheth, R. K., & Tormen, G. 1999, MNRAS, 308, 119

Sijacki, D., Vogelsberger, M., Genel, S., et al. 2015, MNRAS, 452, 575

Skidmore, W., TMT International Science Development Teams, & Science Advisory Committee, T. 2015, Research in Astronomy and Astrophysics, 15, 1945

Sorini, D., Oñorbe, J., Hennawi, J. F., & Lukić, Z. 2018, ApJ, 859, 125

Springel, V. 2010, MNRAS, 401, 791

Stark, C. W., White, M., Lee, K.-G., & Hennawi, J. F. 2015a, MNRAS, 453, 311

—. 2015b, MNRAS, 453, 311

Sugai, H., Tamura, N., Karoji, H., et al. 2015, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 035001

Takada, M., Ellis, R. S., Chiba, M., et al. 2014a, ApJ, 66, R1

—. 2014b, PASJ, 66, 1

Thiele, L., Villaescusa-Navarro, F., Spergel, D. N., Nelson, D., & Pillepich, A. 2020, arXiv e-prints, arXiv:2007.07267

Varin, C., Reid, N., & Firth, D. 2011, Statistica Sinica, 5

Viel, M., Becker, G. D., Bolton, J. S., & Haehnelt, M. G. 2013, PhRvD, 88, 043502

Viel, M., Lesgourgues, J., Haehnelt, M. G., Matarrese, S., & Riotto, A. 2005, PhRvD, 71, 063534

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261

Vogelsberger, M., Genel, S., Springel, V., et al. 2014a, MNRAS, 444, 1518

—. 2014b, Nature, 509, 177

Wakker, B. P., Hernandez, A. K., French, D. M., et al. 2015, ApJ, 814, 40

Zel'dovich, Y. B. 1970, A&A, 5, 84

Zheng, Z., Coil, A. L., & Zehavi, I. 2007, ApJ, 667, 760