

Cosmology and neutrino mass with the minimum spanning tree

Krishna Naidoo^{1,2★}, Elena Massara^{3,4} and Ofer Lahav²

¹Center for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, PL-02-668 Warsaw, Poland

²Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK

³Waterloo Centre for Astrophysics, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada

⁴Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada

Accepted 2022 April 22. Received 2022 March 17; in original form 2021 November 24

ABSTRACT

The information content of the minimum spanning tree (MST), used to capture higher order statistics and information from the cosmic web, is compared to that of the power spectrum for a $\nu\Lambda$ CDM model. The measurements are made in redshift space using haloes from the Quijote simulation of mass $\geq 3.2 \times 10^{13} h^{-1} M_{\odot}$ in a box of length $L_{\text{box}} = 1 h^{-1}$ Gpc. The power spectrum multipoles (monopole and quadrupole) are computed for Fourier modes in the range $0.006 h\text{Mpc}^{-1} < k < 0.5 h\text{Mpc}^{-1}$. For comparison the MST is measured with a minimum length-scale of $l_{\text{min}} \simeq 13 h^{-1}$ Mpc. Combining the MST and power spectrum allows for many of the individual degeneracies to be broken; on its own the MST provides tighter constraints on the sum of neutrino masses M_{ν} and cosmological parameters h , n_s , and Ω_b but the power spectrum alone provides tighter constraints on Ω_m and σ_8 . Combined we find constraints that are a factor of two (or greater) on all parameters with respect to the power spectrum (for M_{ν} there is a factor of four improvement). These improvements appear to be driven by the MST's sensitivity to small scale clustering, where the effect of neutrino free-streaming becomes relevant, and high-order statistical information in the cosmic web. The MST is shown to be a powerful tool for cosmology and neutrino mass studies, and therefore could play a pivotal role in ongoing and future galaxy redshift surveys (such as DES, DESI, *Euclid*, and Rubin-LSST).

Key words: neutrinos – cosmological parameters – large-scale structure of Universe.

1 INTRODUCTION

The Lambda cold dark matter (Λ CDM) paradigm has remained at the forefront of cosmology for over 20 years, cementing it as the standard cosmological model. Observations and simulations over that time have largely strengthened the case for Λ CDM despite the model consisting overwhelmingly of things we still do not understand – principally the nature of dark matter and dark energy. While understanding the former will most likely need significant contributions from particle physics experiments, understanding the latter is a key goal for future experiments in cosmology. The next generation of galaxy redshift surveys (such as the Dark Energy Spectroscopic Instrument (DESI),¹ *Euclid*,² the Nancy Grace Roman Space Telescope,³ the Prime Focus Spectrograph,⁴ the Rubin Observatory Legacy Survey of Space and Time,⁵ and the 4-m Multi-Object Spectroscopic Telescope⁶) will map the positions of hundreds of millions of galaxies. Determining the nature of dark energy is a major scientific mission for these surveys. In Λ CDM dark energy is simply Einstein's cosmological constant (resulting from a scalar field), but

better data may reveal that dark energy is actually changing with time (dynamical dark energy) or that the requirement for dark energy is actually a symptom that General Relativity, on which these models are based, is incomplete and requires modification (the premise of Modified Gravity theories).

The discovery of neutrino oscillation (Fukuda et al. 1998; Ahmad et al. 2001) provided evidence that neutrinos are not massless (as had been predicted by the standard model of particle physics). This discovery makes neutrinos of keen interest to particle physicists, as the origins of their mass and hierarchy could provide hints to new physics. Particle physics experiments currently place a lower bound on the sum of neutrino masses (denoted in this work by M_{ν}) of $M_{\nu} \gtrsim 0.06$ eV. However, it is cosmological experiments that provide the tightest upper bound: currently $M_{\nu} \lesssim 0.11$ eV (95 per cent confidence level, CL; Planck Collaboration VI 2020; eBOSS Collaboration 2021); by contrast the upper bound from particle physics experiment KATRIN is currently $M_{\nu} \leq 1.1$ eV (90 per cent CL; Aker et al. 2019; although unlike cosmological measurements this is model independent). The sensitivity to neutrino mass in cosmology comes from the role neutrinos play in the growth of large-scale structure (LSS). The neutrino's characteristic free-streaming length, a quantity dependent on its mass, will wash out small-scale structure. This effect can be quantified as a suppression of small-scale modes in the power spectrum. Over the next 5 yr, surveys such as DESI expect to be sensitive to $M_{\nu} \lesssim 0.06$ eV (95 per cent CL; Font-Ribera et al. 2014) and therefore anticipate a first cosmological detection of a non-zero mass for neutrinos.

* E-mail: krishna.naidoo.11@ucl.ac.uk

¹<http://desi.lbl.gov/>

²<http://www.euclid-ec.org/>

³<https://roman.gsfc.nasa.gov>

⁴<https://pfs.ipmu.jp/index.html>

⁵<https://www.lsst.org/>

⁶<https://www.4most.eu/cms/>

Cosmological surveys have largely focused on two-point statistics, whether in real or Fourier space. While these methods are tried-and-tested, they fail to fully explore and capture all the information content of galaxy surveys. This is particularly relevant at low redshifts where the highly non-linear structure of the cosmic web is very pronounced; here the distribution of matter cannot be fully characterized by two-point statistics and further statistical methods are required if we are to fully extract all the information present. Such analyses are particularly timely as future data sets will probe the Universe with tracers (galaxies, quasars, etc.) at higher number densities providing a greater sensitivity to the cosmic web. Relevant statistical methods include the three-point correlation function (the Bispectrum in Fourier space; e.g. Gil-Marín et al. 2017; Gualdi et al. 2019), Minkowski functionals (e.g. Petri et al. 2013), the 1D probability distribution function (Uhlemann et al. 2020), marked power spectra (Massara et al. 2021), machine learning (ML; e.g. Fluri et al. 2018), and the minimum spanning tree (MST; Naidoo et al. 2020) – the focus of this paper.

The MST was first introduced to astronomy by Barrow, Bhavsar & Sonoda (1985) and has successfully been used as a filament finder for cosmic web studies (Bhavsar & Ling 1988; van de Weygaert, Jones & Martínez 1992; Bhavsar & Splinter 1996; Krzewina & Saslaw 1996; Ueda & Itoh 1997; Coles et al. 1998; Adami & Mazure 1999; Colberg 2007; Alpaslan et al. 2014; Beuret et al. 2017; Libeskind et al. 2018). The MST is the minimum weighted graph that connects a set of points without forming loops. More recently, Naidoo et al. (2020) investigated how the MST could be used to incorporate the cosmic web when constraining cosmological parameters. However, unlike the conventional two-point analysis as performed by most galaxy redshift surveys, reference MST values cannot be calculated analytically and instead need to be calculated from simulations. Fortunately, this problem is not unique to the MST – conventional statistics such as the power spectrum and bispectrum cannot be computed analytically in the non-linear regime (for Fourier modes $k \gtrsim 0.3 h\text{Mpc}^{-1}$) and hence require simulations, as do artificial intelligence (AI) and ML algorithms (as well as other algorithms used to measure non-linear features in the cosmic web). This has created a growing demand for large suites of cosmological simulations and the development of accurate emulators as cosmologists push to extract more information from the distribution of galaxies.

The Quijote simulations (Villaescusa-Navarro et al. 2020) were designed precisely for this use (i.e. to test new summary statistics such as the MST and AI/ML and to push conventional statistics to smaller scales). In this paper, we will use these simulations to measure the information content of the MST. The simulations have previously been used to conduct Fisher matrix analysis for the power spectrum (Villaescusa-Navarro et al. 2020), bispectrum (Hahn et al. 2020), 1D probability distribution function (Uhlemann et al. 2020), and marked power spectrum (Massara et al. 2021). In this paper, we extend this analysis to the MST; in Naidoo et al. (2020) the MST was tested against measurements of the power spectrum and bispectrum for a few parameters (matter density Ω_m , amplitude of scalar fluctuations A_s , and neutrino mass M_ν) to test whether the MST adds new information. Fisher matrix analysis is useful in cosmology as it places a lower bound (the Cramer–Rao bound; Rao 1945; Cramer 1946) on the uncertainty of cosmological parameters inferred from a given statistic. If the posterior distribution is Gaussian then the Fisher matrix constraints will be realized; otherwise the constraints on any given parameter will be weaker. This analysis is explored for parameters of the $\nu\Lambda\text{CDM}$ model (i.e. the standard model of cosmology ΛCDM + massive neutrinos M_ν). This will help

determine the role that the MST can play in constraining parameters from the current and next generation of galaxy surveys.

The paper is organized as follows. In Section 2, we discuss the methodology and data used. In Section 3, we present the constraints from components of the MST and demonstrate how including the MST together with measurements of the power spectrum improves parameter constraints in a $\nu\Lambda\text{CDM}$ model. Finally, in Section 4 we discuss the main results and their implications for cosmology and future surveys.

2 METHOD

In this section, we explain the Fisher matrix formalism used to measure the information content of several summary statistics, we explain how we measure the power spectrum multipoles and the MST statistics in redshift space, and we describe properties of the Quijote simulations used in this analysis.

2.1 Fisher formalism

The Fisher matrix (Tegmark, Taylor & Heavens 1997) F is defined to have elements

$$F_{ij} = \sum_{\alpha,\beta} \frac{\partial S_\alpha}{\partial \theta_i} C_{\alpha\beta}^{-1} \frac{\partial S_\beta}{\partial \theta_j}, \quad (1)$$

where S_α and S_β are the elements α and β of the data vector \mathbf{S} , \mathbf{C} is the sample covariance matrix defined to have elements

$$C_{\alpha\beta} = \langle (S_\alpha - \langle S_\alpha \rangle) (S_\beta - \langle S_\beta \rangle) \rangle, \quad (2)$$

and θ_i and θ_j are parameters i and j of the model. We multiply the inverse of the covariance matrix by the Kaufman–Hartlap factor (Kaufman 1967; Hartlap, Simon & Schneider 2007) $(N_{\text{sim}} - 2 - N_S)/(N_{\text{sim}} - 1)$, where N_S is the length of the data vector \mathbf{S} and N_{sim} is the number of simulations used to estimate the covariance matrix; this compensates for the error in the sample covariance estimation. An implicit assumption of this formalism is that the covariance matrix has no parameter dependence and can be accurately defined by one fiducial point in parameter space.

When reference summary statistics are available analytically, the partial derivatives in the Fisher matrix are straightforward to estimate. However, for some summary statistics (such as the MST) where reference values must be obtained via simulations, the partial derivatives must be estimated numerically; typically we use

$$\frac{\partial \mathbf{S}}{\partial \theta} \simeq \frac{\mathbf{S}(\theta + d\theta) - \mathbf{S}(\theta - d\theta)}{2d\theta} + \mathcal{O}(d\theta^2), \quad (3)$$

where $d\theta$ is a small deviation from a fiducial θ . We cannot use this when estimating the partial derivative with respect to neutrino mass when this mass is zero, as this would require simulations with negative neutrino mass; instead here we use one of the estimators

$$\left[\frac{\partial \mathbf{S}}{\partial M_\nu} \right]_1 \simeq \frac{\mathbf{S}(dM_\nu) - \mathbf{S}(M_\nu = 0)}{dM_\nu} + \mathcal{O}(dM_\nu), \quad (4)$$

$$\left[\frac{\partial \mathbf{S}}{\partial M_\nu} \right]_2 \simeq \frac{-\mathbf{S}(2dM_\nu) + 4\mathbf{S}(dM_\nu) - 3\mathbf{S}(M_\nu = 0)}{2dM_\nu} + \mathcal{O}(dM_\nu^2), \quad (5)$$

$$\left[\frac{\partial \mathbf{S}}{\partial M_\nu} \right]_3 \simeq \frac{\mathbf{S}(4dM_\nu) - 12\mathbf{S}(2dM_\nu) + 32\mathbf{S}(dM_\nu) - 21\mathbf{S}(M_\nu = 0)}{12dM_\nu} + \mathcal{O}(dM_\nu^3). \quad (6)$$

These non-symmetric estimators are designed to use simulations with $M_\nu = 0.1, 0.2, \text{ or } 0.4 \text{ eV}$ (in addition to $M_\nu = 0$). In this case, the

Table 1. A summary of the Quijote simulations used in this study, highlighting the deviations from the fiducial cosmological parameters, the type of initial conditions (IC; either first-order perturbation theory – ZA (Zel’dovich approximation) or second-order perturbation theory – 2LPT) and the number of realizations.

Name	Deviation from fiducial	IC	Realizations
Fiducial	n/a	2LPT	15 000
Fiducial_ZA	n/a	ZA	500
Ω_m^+	$\Delta\Omega_m = +0.01$	2LPT	500
Ω_m^-	$\Delta\Omega_m = -0.01$	2LPT	500
Ω_b^{++}	$\Delta\Omega_b = +0.002$	2LPT	500
Ω_b^{--}	$\Delta\Omega_b = -0.002$	2LPT	500
h^+	$\Delta h = +0.02$	2LPT	500
h^-	$\Delta h = -0.02$	2LPT	500
n_s^+	$\Delta n_s = +0.02$	2LPT	500
n_s^-	$\Delta n_s = -0.02$	2LPT	500
σ_8^+	$\Delta\sigma_8 = +0.015$	2LPT	500
σ_8^-	$\Delta\sigma_8 = -0.015$	2LPT	500
M_v^+	$\Delta M_v = +0.1 \text{ eV}$	ZA	500
M_v^{++}	$\Delta M_v = +0.2 \text{ eV}$	ZA	500
M_v^{+++}	$\Delta M_v = +0.4 \text{ eV}$	ZA	500

increment dM_v can in equation (4) be any of these three values, in equation (5) can be $M_v = 0.1$ or 0.2 eV , and in equation (6) must be $M_v = 0.1 \text{ eV}$.

The likelihood is assumed to follow a multivariate Gaussian (e.g. Heavens 2009, i.e. with Gaussian errors for each parameter) defined by

$$\mathcal{L}(\boldsymbol{\theta}) = \sqrt{\frac{\det \mathbf{F}}{(2\pi)^M}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_{\text{Fid}})^\top \cdot \mathbf{F} \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}_{\text{Fid}})\right), \quad (7)$$

where $\boldsymbol{\theta}$ (of length M) are the parameters of a $\nu\Lambda\text{CDM}$ model and $\boldsymbol{\theta}_{\text{Fid}}$ are fiducial parameters.

2.2 Summary statistics in redshift space

In redshift space, redshift space distortions (RSD; Kaiser 1987) caused by peculiar velocities alter the observed redshifts of galaxies. This causes a line-of-sight (LOS) shift given by

$$\mathbf{x}_{\text{RSD}} = \mathbf{x} + \frac{1+z}{H(z)}(\mathbf{v} \cdot \mathbf{e}), \quad (8)$$

where \mathbf{x} is the real space coordinate, \mathbf{x}_{RSD} is the redshift space coordinate, \mathbf{v} is the peculiar velocity, z is the redshift, $H(z)$ is the Hubble expansion rate at redshift z , and \mathbf{e} is the unit vector defining the LOS. In this paper, the LOS is taken to be the z -axis ($\mathbf{e} = (0, 0, 1)$); while the accuracy and convergence for the partial derivatives estimates is improved by additionally using both the x -axis ($\mathbf{e} = (1, 0, 0)$), and y -axis ($\mathbf{e} = (0, 1, 0)$) as the LOS (Hahn et al. 2020).

2.2.1 Power spectrum multipoles

The density field is reexpressed in Fourier space; let $\mathbf{k} = (k_x, k_y, k_z)$ be a Fourier mode vector. We bin the density field by $k = |\mathbf{k}|$ and $\mu = k_z/k$ (the cosine of the angle between \mathbf{k} and the LOS \mathbf{e}).

The power spectrum multipoles are

$$P_\ell(k) = (2\ell + 1) \int_0^1 P(k, \mu) \mathcal{L}_\ell(\mu) d\mu, \quad (9)$$

where \mathcal{L}_ℓ is a Legendre polynomial. The monopole and quadrupole are P_0 and P_2 , respectively.

2.2.2 Masking small scales by grouping

Unlike the power spectrum, small-scale information cannot be removed by simply cutting the distribution of edge lengths in the MST statistics, instead these scales need to be removed from the input catalogues. To do this, small scales are masked by grouping together points with small separations. This is carried out by grouping two haloes if their separation is less than $l_{\text{min}} = 2\pi/k_{\text{max}} = 4\pi h^{-1} \text{ Mpc} \simeq 13 h^{-1} \text{ Mpc}$ (equivalent to the maximum Fourier mode $k_{\text{max}} = 0.5 h \text{ Mpc}^{-1}$ measured for the power spectrum). The grouping is transitive (if A and B are close, and B and C are close, then all three are grouped together regardless of the distance between A and C); as a result, l_{min} needs to be set well below the mean separation of points (to avoid the entire catalogue collapsing to one point). A group of haloes is given coordinates equal to the mean coordinates of its constituent haloes. This process yields a catalogue of nodes (a collection of grouped and ungrouped haloes); on average this node catalogue is about one quarter the size of the original halo catalogue. See Appendix A for evidence that this masking technique is effective.

2.2.3 Minimum spanning tree statistics

The MST is constructed, in 3D comoving coordinates in redshift space, from the node catalogue. The distribution function $N(x)$ of the MST statistics is measured, where x is the degree d , edge length l , branch length b , or branch shape s (Naidoo et al. 2020). An edge is a line in the MST graph, the degree is the number of edges attached to each node, and a branch is a chain of edges connected continuously by nodes of degree $d = 2$. For the branches we measure their length b (i.e. the sum of the lengths of member edges) and their shape s (i.e. the square root of one minus the ratio between the straight line distance between branch ends and the branch length – with this definition, straighter branches have $s \simeq 0$ while larger values indicate more curved branches). Furthermore, to ensure $N(x)$ can be described by a Gaussian distribution we remove the tails of the distribution function. We measure the mean of the cumulative distribution function (CDF) of $N(x)$ for the fiducial simulations and then measure the MST $N(x)$ only in the region where $0.05 < \text{CDF} < 0.95$ (with the exception of d where we include $N(d)$ in the range $1 \leq d \leq 4$). The publicly available PYTHON package MISTREE (Naidoo 2019) was used to construct and measure the statistics of the MST.

2.3 Quijote simulations

The Quijote simulations (Villaescusa-Navarro et al. 2020) are a large set of N -body simulations designed for quantifying the information content of cosmological observables and for training ML algorithms. The simulations are constructed in boxes of length $L_{\text{box}} = 1 h^{-1} \text{ Gpc}$, using 512^3 dark matter particles and 512^3 neutrino particles (for simulations with massive neutrinos). A detailed table of the parameters used for the Quijote simulations can be found in table 1 of Villaescusa-Navarro et al. (2020). The simulations are based on a fiducial ΛCDM cosmology (based on Planck Collaboration VI 2020) with matter density $\Omega_m = 0.3175$, baryon density $\Omega_b = 0.049$, Hubble constant $h = 0.6711$, primordial spectral tilt $n_s = 0.9624$, the root mean square of the linear power spectrum at spheres of radius $8 h^{-1} \text{ Mpc}$ $\sigma_8 = 0.834$, sum of neutrino masses $M_\nu = 0 \text{ eV}$, and dark energy equation of state $w = -1$. The power spectrum multipoles and MST are computed on haloes with masses larger than $3.2 \times 10^{13} h^{-1} M_\odot$. For each parameter, we determine the dependence with respect to that parameter using 500 simulations

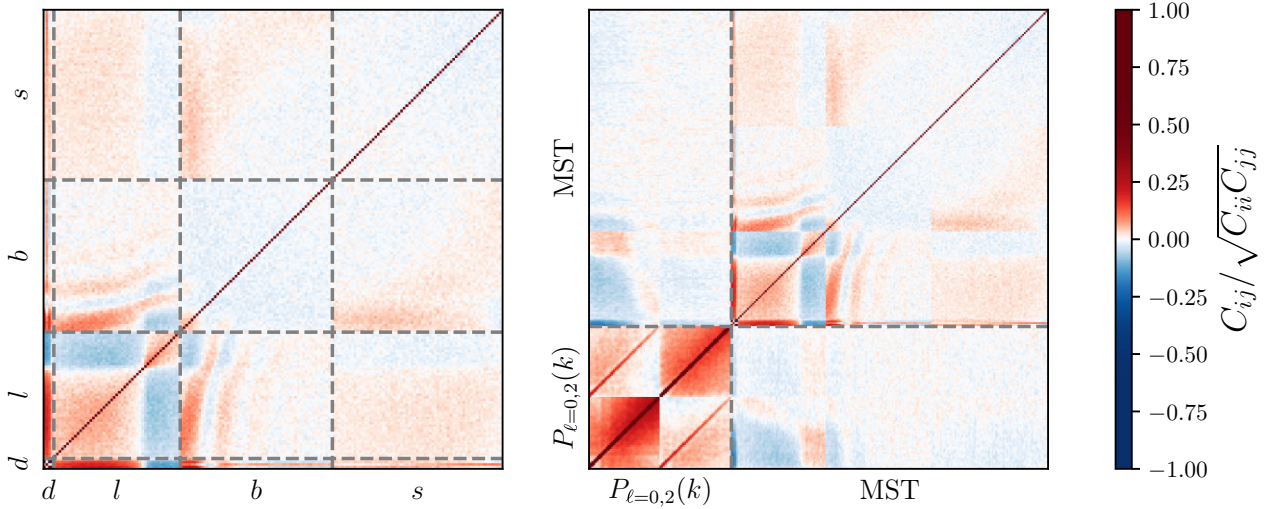


Figure 1. Correlation matrices (*left*) of the MST statistics and (*right*) and power spectrum multipoles. The components of the MST statistics (the degree d , edge length l , branch length b , and branch shape s) are labelled and divided by dashed grey lines on the left. The matrix has significant non-diagonal features. The most striking is the correlation between the edge length l and branch length b where there are several fading lines of positive correlation. These originate from correlations between branches formed from two edges, three edges and so on, becoming fainter for branches formed from more edges. For the degree d and l there is virtually no correlation while there are correlations between d and the two branch statistics length b and shape s . These correlations appear to be strongly tied to $d = 2$ presumably due to the definition for branches requiring intermediate nodes with degree $d = 2$. Lastly there is a weak correlation between b and s and negligible correlations between other MST statistics (i.e. between l and d and between l and s). The power spectrum multipoles and components of the MST statistic are divided by a dashed grey lines on the right. The correlation matrix for the power spectrum multipoles is shown to be strongest diagonally, with the most striking feature being the positive correlations between multipoles. Correlations between the power spectrum multipoles and the MST are negligible with only a small inverse correlation seen. This is consistent with expectations since smaller edge lengths should correspond to larger Fourier modes.

in which only that parameter deviates from its fiducial value (while maintaining zero curvature, i.e. a shift in $\Omega_m \Rightarrow \Omega_\Lambda = 1 - \Omega_m$). To construct the covariance matrix we use 15 000 *fiducial simulations* constructed with the fiducial cosmology. See Table 1 for a summary. Simulations with massive neutrinos (M_ν^+ , M_ν^{++} , and M_ν^{+++}) are produced from simulations with initial conditions following first-order perturbation theory (i.e. the Zel’dovich approximation – ZA) instead of second-order Lagrangian perturbation theory (2LPT) since the 2LPT approach is not implemented for massive neutrino cosmologies (for further details, see Villaescusa-Navarro et al. 2020). Therefore, to remove any potential systematic bias this discrepancy may present we will be using the Fiducial_ZA simulations for the neutrino partial derivative estimates, rather than the Fiducial simulations, since these simulations are also produced with ZA initial conditions.

3 RESULTS

This section discusses the following results: (1) the covariance matrix for the MST statistics and the internal correlations and correlations with the power spectrum, (2) the partial derivatives of the power spectrum and MST statistics, and (3) parameter constraints for a $\nu\Lambda$ CDM model obtained from individual and combined measurements of the MST and power spectrum.

3.1 Covariance matrix

The covariance matrix is constructed from equation (2) using data vectors measured from 15 000 fiducial simulations. Fig. 1 shows the correlation matrix for the MST on the left and the correlation matrix for the combined data vector of the power spectrum multipoles and MST statistics on the right. Unlike the correlation matrix for

the power spectrum, the correlation matrix for the MST contains several non-diagonal features. One of the most striking features is the correlation between the edge length l and branch length b which show ‘waves’ of positive correlations between short edges and short branches followed by negative correlations and then positive correlation for longer edges and branches. These positive correlations stem from the correlations between branches formed from two edges, three edges, and so on. For branches formed from more edges, these correlations become weaker as branches formed from more than three edges are rare. Other correlations in the MST statistics appear to stem from branches, which by definition have intermediate nodes with degree $d = 2$. As a result we see strong correlations between the degree and branch length. The correlation between branch length and shape is weak but indicates that longer branches are more curved than short ones. The correlations between the power spectrum multipoles and MST are weak so adding the MST to $P(k)$ is beneficial. This is clearest for the monopole and edge lengths which show an inverse correlation between edges and Fourier modes; this is completely consistent with the inverse relation between Fourier space and real space. Furthermore the large-scale modes of the monopole (the first half of the data vector) show positive correlations with longer edges. This indicates that most of the large-scale clustering information is stored in the large edges of the MST.

3.2 Fisher matrix and partial derivatives

The Fisher matrix is calculated from equation (1) from the data vector \mathbf{S} . In this section, the data vectors are either the power spectrum monopole and quadrupole $P_{\ell=0,2}(k)$ or combinations of the MST statistics: degree d , edge length l , branch length b , and branch shape s . To calculate the Fisher matrix, we require measurements of the derivatives $\partial\mathbf{S}/\partial\theta$; these are estimated using simulations

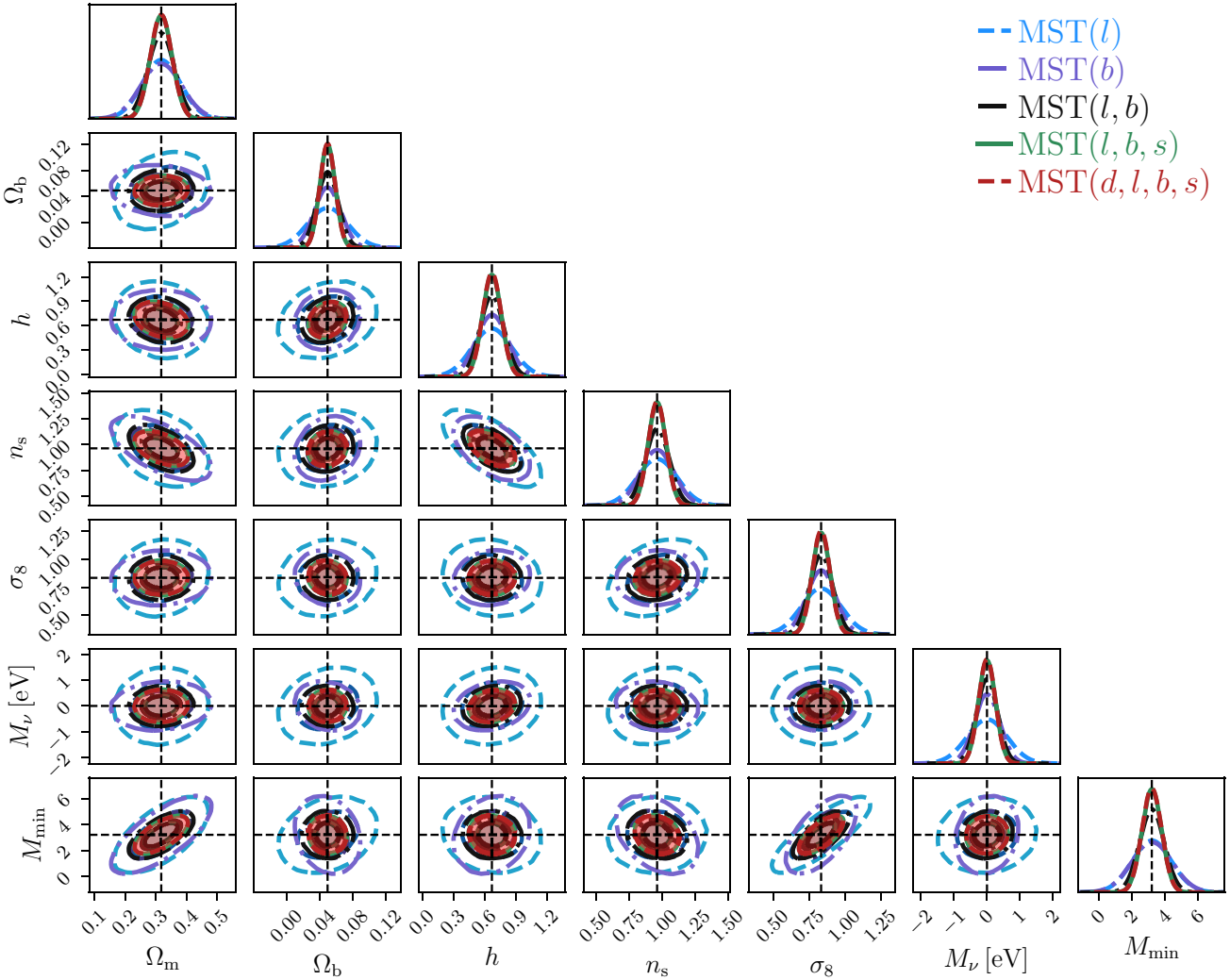


Figure 2. Fisher matrix constraints for $\nu\Lambda\text{CDM}$ parameters from the MST edge and branch length distributions in combination with the distribution of degree and branch shape in redshift space at $z = 0.5$. The individual constraints from the edge MST(l) and branch length distributions MST(b) are shown in blue and purple, respectively. The combined constraint from the edge and branch lengths MST(l, b) is shown with black dashed lines, from the edge, branch length and degree distributions MST(l, b, s) with green lines, and from the edge, branch length, degree and branch shape distribution MST(d, l, b, s) with red lines and contours. The contours show that adding the degree does very little to improve the constraints from the edge and branch length distribution while the branch shape tightens constraints overall.

summarized in Table 1. For each set of simulations (consisting of 500 individual simulations and 15 000 simulations for the fiducial set) the mean and standard deviation of the summary statistics are obtained. The derivatives for the parameters are then obtained using equation (3), with the exception of neutrino mass where three estimators are used (equations 4, 5, and 6). In real data we would not be able to place a clean cut on the galaxy masses; instead, these limits would be imposed by survey designs and magnitude limits. To account for this in our analysis, we follow Hahn et al. (2020) by adding the nuisance parameter M_{\min} which characterizes the dependence on the minimum halo mass which is calculated by running the statistics on the fiducial suite with $M_{\min} = 3.1$ and $M_{\min} = 3.3 \times 10^{13} h^{-1} M_{\odot}$. However, unlike in the study by Hahn et al. (2020) we do not include a linear bias nuisance parameter (as it is not clear how a linear bias would affect the MST statistics and including it solely in the power spectrum measurements would weaken the constraints in the power spectrum but leave the MST unaffected, in

effect biasing our results in favour of the MST). To avoid this, we use only the minimum halo mass as a nuisance parameter in this study.

The more accurate estimators for the partial derivatives are those for which the errors are given to higher orders of $d\theta$. For most parameters, the derivative $\partial S/\partial\theta$ is determined by the symmetric derivative estimator equation (3) which has errors of order $\mathcal{O}(d\theta^2)$, while the most accurate estimator for $\partial S/\partial M_{\nu}$ is given by equation (6) which has errors of order $\mathcal{O}(d\theta^3)$. However, in order for the estimators to be consistent for all parameters the appropriate estimator to use for M_{ν} is equation (5) which has errors of order $\mathcal{O}(d\theta^2)$. Previous studies (Hahn et al. 2020; Uhlemann et al. 2020; Villaescusa-Navarro et al. 2020) have used equation (6) (while Massara et al. 2021 use equation 5), so to facilitate comparisons to these studies the results obtained using this estimator are additionally provided in Table C2. However, throughout this paper we will refer to results obtained using equation (5) with $dM_{\nu} = 0.2 \text{ eV}$. The partial derivatives for the data

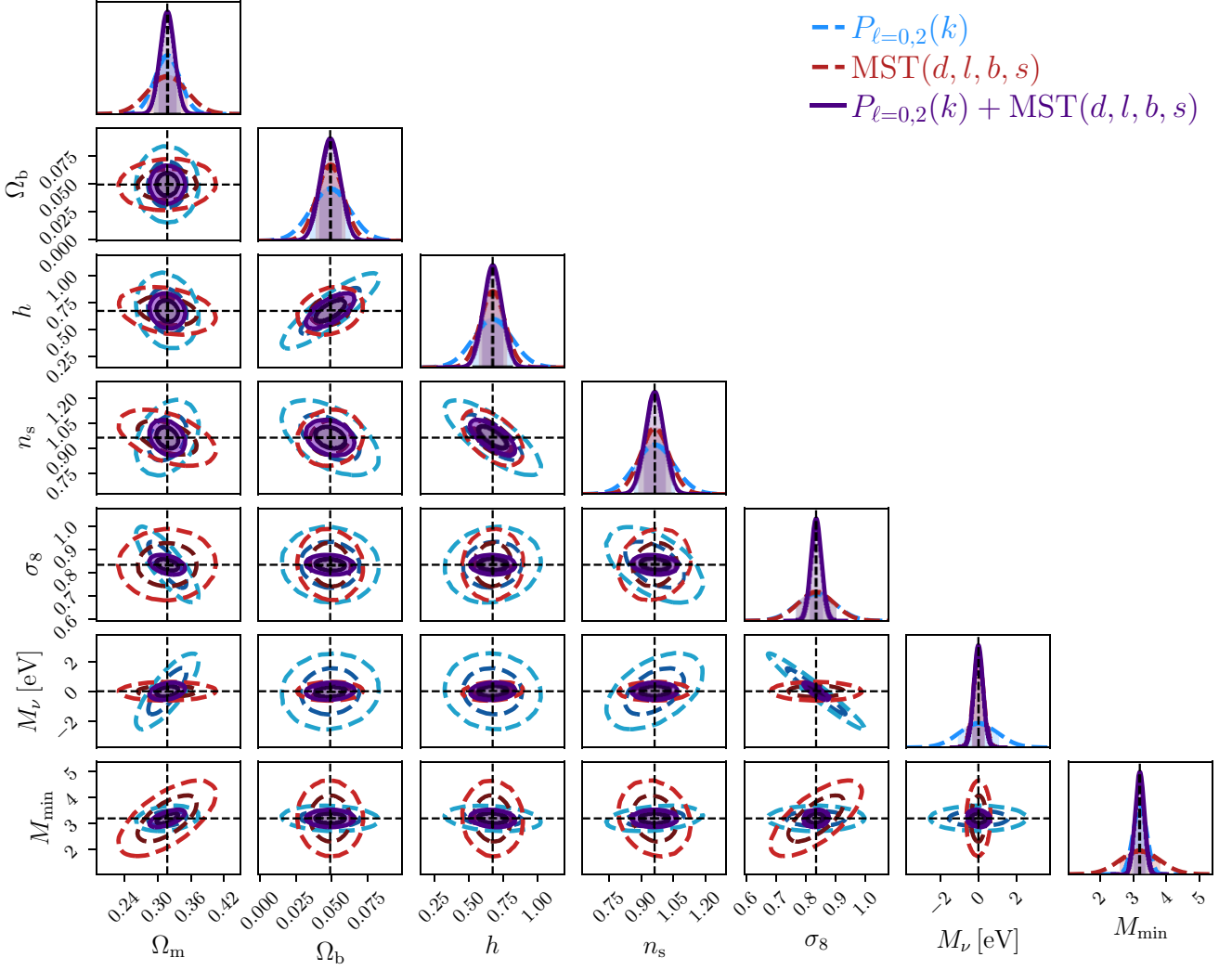


Figure 3. Fisher matrix constraints on $\nu\Lambda$ CDM parameters from the power spectrum and MST separately and combined in redshift space at $z = 0.5$. The constraints from the power spectrum multipoles $P_{\ell=0,2}(k)$ are shown with blue dashed lines, from the four MST statistics $\text{MST}(d, l, b, s)$ with red dotted lines, and from the combination of the power spectrum multipoles and MST $P_{\ell=0,2}(k) + \text{MST}(d, l, b, s)$ with purple lines and contours. Constraints for σ_8 and Ω_m are dominated by the power spectrum multipoles while h , n_s , Ω_b , and M_ν are dominated by the MST. Significant degeneracies are broken when combined, leading to much tighter constraints (in comparison to the individual constraints from the power spectrum multipoles) on h , n_s , Ω_m , and Ω_b (which improve by a factor of ~ 2); σ_8 (which improves by a factor of ~ 3.89) and M_ν (which improves by a factor of ~ 4.35).

vectors (as a function of the $\nu\Lambda$ CDM parameters Ω_m , Ω_b , h , n_s , σ_8 , and M_ν) are shown in Appendix B.

We test for convergence of the partial derivatives by estimating the Fisher matrix from a fraction of the total simulations available. For the power spectrum, we find that partial derivatives are converged for all M_ν derivative estimates and for all parameters while for the MST we find that equation (5) with $dM_\nu = 0.2$ eV provides the most reliable converged estimates at all redshifts (see Fig. B5) and quote results from this estimator unless stated otherwise.

3.3 MST bin size dependence

We test the sensitivity of the MST distribution functions to the bin size. In real data, the size of bins would be dictated by numerical constraints, i.e. the number of mocks used to obtain the covariance matrices and the sample size. Since the suite used to estimate the covariance matrix is large, we do not need to keep the data vector

short and therefore can test the sensitivity to the MST distribution function's bin size. To this end, we use three binning strategies for the distribution functions of l , b , and s , with $N_{\text{bins}} = 200$, 100, and 40 bins. The constraints from the different binning strategies are described and discussed in the following sections (in particular, see Fig. 4).

3.4 Sensitivity to neutrinos and Λ CDM

In this section, the forecast constraints (derived from Fisher matrices) for parameters of a $\nu\Lambda$ CDM universe are obtained and discussed.

3.4.1 Constraints from the minimum spanning tree

In Fig. 2, the individual and combined constraints from the edge length l , branch length b , branch shape s , and degree d distributions are shown using l , b , and s with $N_{\text{bins}} = 200$. The edge and branch

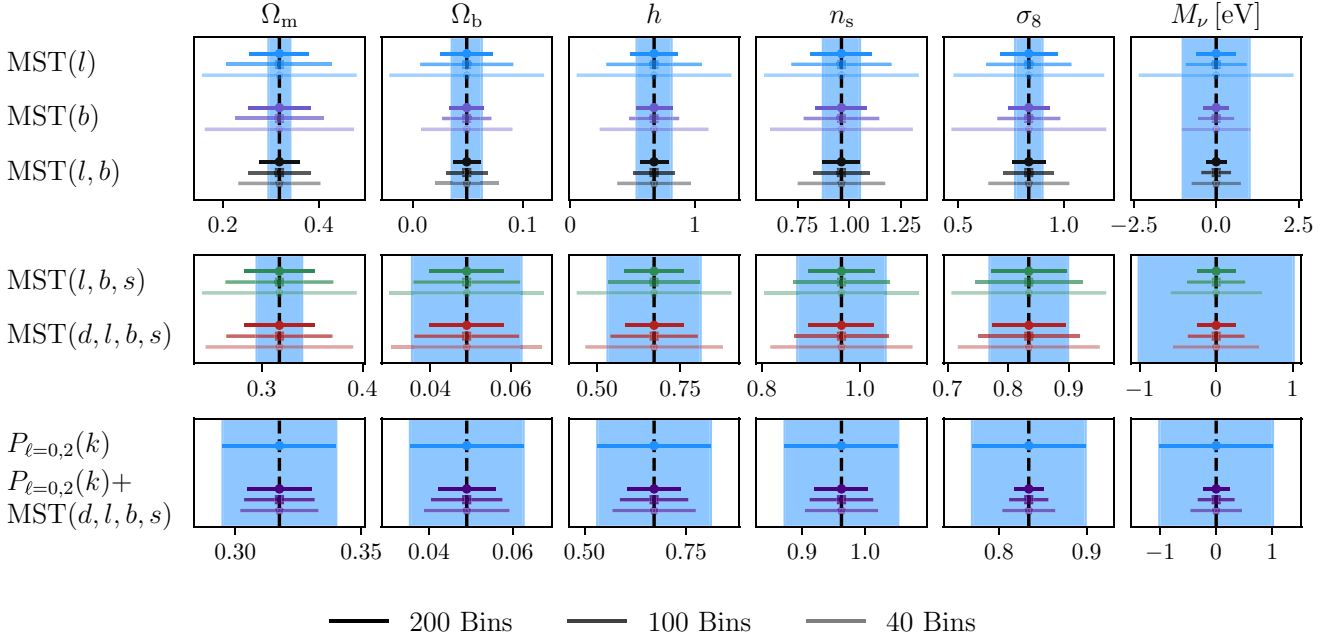


Figure 4. The 1D parameter constraints from the MST statistics individually and in combination with the power spectrum multipoles (measured at redshift $z = 0.5$). From left to right, we show the constraints on the cosmological parameters of the $\nu\Lambda$ CDM model: Ω_m , Ω_b , h , n_s , σ_8 , and M_ν . From top to bottom, we show the constraints for the individual MST statistics: edge length $MST(l)$ and branch length $MST(b)$, then constraints from the combined MST statistics: edge and branch length $MST(l, b)$, with the addition of branch shape $MST(l, b, s)$ and with the addition of the degree $MST(d, l, b, s)$, and finally in combination with the power spectrum: first showing the power spectrum multipoles alone $P_{\ell=0,2}(k)$ (these are marked with blue envelopes for comparison to the other statistics) and in combination $P_{\ell=0,2}(k) + MST(d, l, b, s)$. These are further subdivided for constraints involving the MST with different binning schemes N_{bins} : 200 bins are shown with a dark coloured error bar at the top, 100 bins are shown with slightly paler coloured error bars in the middle, and 40 bins are shown with the palest coloured error bars at the bottom. These results indicate the significance that binning will have on the individual MST constraints; they also show that, regardless of binning, combining MST statistics with the power spectrum significantly improves constraints on all parameters.

length distributions have constraints that are competitive with each other, with branch length providing but that in combination are significantly tighter. In augmenting the constraints first with the branch shape and then with the degree, we see that much of the constraining power for the MST is contained in the edge and branch length distribution; the addition of the branch shape adds a modest improvement to the overall constraints (~ 25 per cent for cosmological parameters at $z = 0.5$) but the addition of the degree appears to provide a negligible improvement (~ 1 per cent for cosmological parameters at $z = 0.5$).

3.4.2 Combined constraints from the minimum spanning tree and power spectrum

The separate and combined constraints from the MST (using $l, b,$ and s with $N_{\text{bins}} = 200$) and power spectrum measured at redshift $z = 0.5$ are shown in Fig. 3 (for the constraints obtained from different redshifts refer to Table C2). The constraints obtained for $\Omega_b, h, n_s,$ and neutrino mass M_ν are significantly tighter for the MST, while the power spectrum yields better constraints on Ω_m and σ_8 . When the two are combined, significant degeneracies are broken leading to much tighter constraints than those obtained using either statistic alone. For neutrino mass M_ν , the combination provides a 1σ constraint that is $4.35\times$ tighter, for Ω_m that is $1.77\times$ tighter, for Ω_b that is $2.06\times$ tighter, for h that $2.12\times$ tighter, for n_s that is $2.09\times$ tighter, and for σ_8 that is $3.82\times$ tighter.

The impact of the MST binning is explored in Fig. 4, where we display the marginalized constraints for the MST statistics (individually and for different combinations) at redshift $z = 0.5$

in comparison to the power spectrum multipoles with different levels of binning. Larger bins are associated with poorer constraints but in each case combining with the power spectrum still leads to significant improvements in the constraints of all parameters. This highlights the importance of including the MST in future galaxy redshift surveys to test and constrain Λ CDM parameters and to determine the sum of neutrino mass M_ν – a key scientific goal of many future surveys.

3.4.3 Sensitivity to the neutrino partial derivative estimator

The partial derivative of the summary statistics as a function of neutrino mass can be estimated using equations (4), (5), and (6). Fig. 5 compares the neutrino mass constraints $\Delta M_\nu = \sqrt{F_{\nu\nu}^{-1}}$ using the different estimators for the power spectrum and MST measured at redshift $z = 0.5$ (see Table C1 for a full summary of the different estimators for different components and combinations of the MST statistics at different redshifts). The intrinsic accuracy of the estimators appear to show that less accurate estimators are associated with poorer constraints. The most accurate estimators, i.e. estimator 2 and 3 with $M_\nu = 0.1$ eV, are shown to be fairly consistent, with estimator 3 generally providing tighter constraints. Without prior knowledge of the true partial derivative it is difficult to know which estimator better captures the true behaviour. For consistency, we focus on the results from estimator 2 (with $M_\nu = 0.2$ eV) which has the same absolute errors as the estimator used on the other parameters (equations 3) and has been found to have the best convergence properties at all redshifts (see Appendix B). Nevertheless as is shown in Fig. 5 the relative strengths of the constraints from the power spectrum and MST individually and combined remains, meaning our

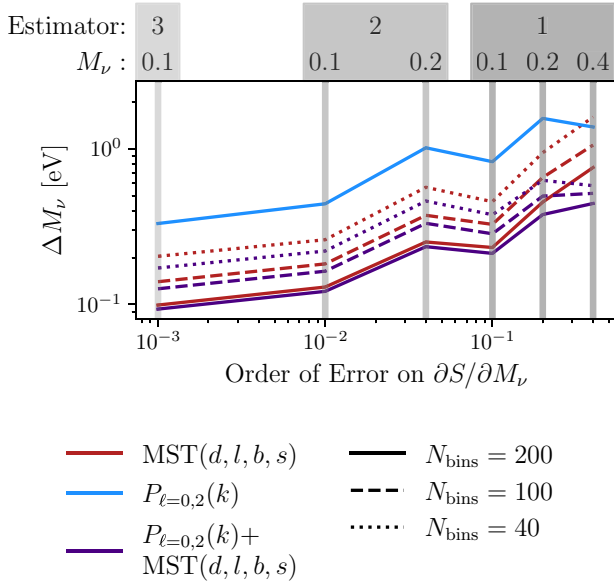


Figure 5. The constraints on neutrino mass from different partial derivative estimates for the MST and power spectrum measured at redshift $z = 0.5$ (individually and in combination). The MST statistics are shown with three binning schemes ($N_{\text{bins}} = 200, 100,$ and 40). The constraints are given as a function of the order of the absolute error on $\partial S/\partial M_\nu$. Estimator 1 is given by equation (4), Estimator 2 by equation (5), and Estimator 3 by equation (6). The estimators are indicated by shaded grey lines with the value of M_ν used for each estimator labelled. With the exception of the 40 bins data, the MST consistently outperforms the power spectrum constraints. However, irrespective of the bin size, significant improvements in the constraints of neutrino mass are obtained when combining the MST with the power spectrum.

overall conclusions are insensitive to the neutrino derivative estimator used.

In all cases, the MST constraints on M_ν are stronger than ones obtained from the power spectrum. If more bins are used for the MST then the combined constraints with the power spectrum are significantly improved in comparison to the power spectrum alone.

4 CONCLUSION

Using haloes from the Quijote simulations, we calculate the information content (from the Fisher matrix) of the MST statistics in a $\nu\Lambda\text{CDM}$ model. The Quijote simulations are a large suite of N -body simulations designed to test the information content of summary statistics and to test AI/ML algorithms. In this paper, we use a subset of the full suite of simulations (summarized in Table 1), measuring the power spectrum multipoles and MST statistics in redshift space on haloes with mass $\geq 3.1 \times 10^{13} h^{-1} M_\odot$. The analysis presented in this paper is based on the Fisher matrix formalism which considers only second derivatives of the data vector, rather than a full exploration of the parameter space via for e.g. MCMC. Fisher matrices are limited in the sense that they place only lower bounds (the Cramer–Rao bound) on the uncertainties of cosmological parameters and assume the covariance matrix has no cosmological parameter dependence; as such analysis on real data is likely to depart significantly due to non-Gaussianities (see Hawken et al. 2012; Foroozan, Krolewski & Percival 2021).

From the MST, we measure the distribution of degree d , edge length l , branch length b , and branch shape s . In Naidoo et al.

(2020), it was shown that l is the most constraining MST statistic. However, this previous analysis was based on COLA simulations of length $L_{\text{box}} = 250 h^{-1} \text{Mpc}$ and limited to the most massive 5000 haloes. Due to the size of the sample, the study saw only modest improvements of 17 per cent on the 1σ constraints on Ω_m and 12 per cent on the 1σ constraints on A_s . By using the Quijote simulations, we are able to expand this analysis to a wider set of cosmological parameters ($\Omega_m, \Omega_b, h, n_s, \sigma_8,$ and M_ν compared to only $A_s, \Omega_m,$ and M_ν), over a larger volume ($L_{\text{box}} = 1 h^{-1} \text{Gpc}$), and using a larger catalogue of haloes (of the order of 10^5). In Fig. 2, the constraints from the MST are shown to be dominated by the distribution of edge and branch lengths, with the degree adding very little information and the branch shape providing a moderate improvement overall.

In Fig. 3 and Table C2, the MST and power spectrum constraints are compared and combined. The power spectrum multipoles provide much stronger constraints on Ω_m and σ_8 . On the other hand, the MST dominates the constraints on M_ν : at $z = 0.5$ the power spectrum multipoles gives $\Delta M_\nu = 1 \text{ eV}$ (1σ), the MST gives $\Delta M_\nu = 0.25 \text{ eV}$ (1σ), and the power spectrum multipoles and the MST combined give $\Delta M_\nu = 0.23 \text{ eV}$ – an improvement of a factor of ~ 4.35 . For the other parameters we find that combining the two sets of statistics breaks several degeneracies and thereby improves constraints on $h, n_s, \Omega_m,$ and Ω_b by a factor of ~ 2 and σ_8 by a factor of ~ 3.82 in comparison to the individual power spectrum constraints. We measure the dependence on the number of bins N_{bins} , showing that fewer bins will decrease the constraining power but that the inclusion of the MST will, irrespective of the bin number, significantly improve constraints on all parameters.

In this paper, we show that adding the MST to the power spectrum greatly improves constraints for parameters of the $\nu\Lambda\text{CDM}$ model. In particular, constraints on M_ν are dominated by the MST since it is significantly more sensitive to the effects of neutrinos on the distribution of haloes. This appears to come from the sensitivity of the MST to extra information contained in the higher order statistics of the cosmic web. Bonnaire et al. (2021) have shown that filaments in the cosmic web are the most sensitive environment to neutrino mass. Given the MST propensity for filament detection, we believe the sensitivity to the neutrino mass is related to the detection of filamentary structures by the MST. However, only future work comparing both cosmic web environments from the density field and MST structures from haloes will be able to determine this for certain.

Future galaxy surveys such as DESI, *Euclid*, and LSST are projected to bring the upper limit on neutrino mass below or close to 0.06 eV (Font-Ribera et al. 2014), the lower limit from neutrino oscillations (assuming normal hierarchy). As we have illustrated using halo simulations, the MST can improve by a factor of four the constraints on neutrino mass with respect to just the power spectrum. Therefore, we can expect the MST to provide greater constraints from current and future surveys, possibly enabling this interesting regime to be probed sooner. These constraints could be further improved by combining the MST and the power spectrum with other probes such as the CMB. This demonstrates the importance of measuring more than just the power spectrum (or two-point statistics) and provides a powerful argument for making measurements of the MST on current and future galaxy surveys such as BOSS, eBOSS, and DESI. In future work, we will look to develop techniques for addressing the challenges associated with real galaxy survey data; these challenges include the impact of the survey’s selection function across the sky (completeness and depth) and as a function of redshift, the incorporation of galaxy weights, the mitigation of small-scale effects including the effect of fibre collisions in spectroscopic surveys, the modelling

of galaxy bias and halo occupation distribution parameters, and the development of MST emulators or likelihood free techniques for computing the posterior probability for real data.

ACKNOWLEDGEMENTS

We thank Lorne Whiteway for his generous contributions to the development of this paper and for providing insightful suggestions and comments. We thank Peter Coles and Andrew Pontzen for providing useful comments and discussions. KN acknowledges support from the Science and Technology Facilities Council grant ST/N50449X and from the (Polish) National Science Centre grant #2018/31/G/ST9/03388. OL acknowledges support from an STFC Consolidated Grant ST/R000476/1.

DATA AVAILABILITY

In this paper, we use halo simulation catalogues from the Quijote simulation. Information on how to access these data products can be found on the Quijote documentation page.⁷

REFERENCES

- Adami C., Mazure A., 1999, *A&AS*, 134, 393
 Ahmad Q. R. et al., 2001, *Phys. Rev. Lett.*, 87, 071301
 Aker M. et al., 2019, *Phys. Rev. Lett.*, 123, 221802
 Alpaslan M. et al., 2014, *MNRAS*, 438, 177
 Barrow J. D., Bhavsar S. P., Sonoda D. H., 1985, *MNRAS*, 216, 17
 Beuret M., Billot N., Cambrésy L., Eden D. J., Elia D., Molinari S., Pezzuto S., Schisano E., 2017, *A&A*, 597, A114
 Bhavsar S. P., Ling E. N., 1988, *PASP*, 100, 1314
 Bhavsar S. P., Splinter R. J., 1996, *MNRAS*, 282, 1461
 Bonnaire T., Aghanim N., Kuruvilla J., Decelle A., 2021, preprint (arXiv:2112.03926)
 Colberg J. M., 2007, *MNRAS*, 375, 337
 Coles P., Pearson R. C., Borgani S., Plionis M., Moscardini L., 1998, *MNRAS*, 294, 245
 Cramer H., 1946, *Mathematical Methods of Statistics*. Princeton University Press, Princeton
 eBOSS Collaboration, 2021, *Phys. Rev. D*, 103, 083533
 Fluri J., Kacprzak T., Refregier A., Amara A., Lucchi A., Hofmann T., 2018, *Phys. Rev. D*, 98, 123518
 Font-Ribera A., McDonald P., Mostek N., Reid B. A., Seo H.-J., Slosar A., 2014, *J. Cosmol. Astropart. Phys.*, 5, 023
 Forozan S., Krolewski A., Percival W. J., 2021, *J. Cosmol. Astropart. Phys.*, 10, 044
 Fukuda Y. et al., 1998, *Phys. Rev. Lett.*, 81, 1562
 Gil-Marín H., Percival W. J., Verde L., Brownstein J. R., Chuang C.-H., Kitaura F.-S., Rodríguez-Torres S. A., Olmstead M. D., 2017, *MNRAS*, 465, 1757
 Galdi D., Gil-Marín H., Schuhmann R. L., Manera M., Joachimi B., Lahav O., 2019, *MNRAS*, 484, 3713
 Hahn C., Francisco V.-N., Emanuele C., Roman S., 2020, *J. Cosmol. Astropart. Phys.*, 03, 040
 Hartlap J., Simon P., Schneider P., 2007, *A&A*, 464, 399
 Hawken A. J., Abdalla F. B., Hütsi G., Lahav O., 2012, *MNRAS*, 424, 2
 Heavens A., 2009, preprint (arXiv:0906.0664)
 Kaiser N., 1987, *MNRAS*, 227, 1
 Kaufman G., 1967
 Krzewina L. G., Saslaw W. C., 1996, *MNRAS*, 278, 869
 Libeskind N. I. et al., 2018, *MNRAS*, 473, 1195
 Massara E., Villaescusa-Navarro F., Ho S., Dalal N., Spergel D. N., 2021, *Phys. Rev. Lett.*, 126, 011301

- Naidoo K., 2019, *J. Open Source Softw.*, 4, 1721
 Naidoo K., Whiteway L., Massara E., Galdi D., Lahav O., Viel M., Gil-Marín H., Font-Ribera A., 2020, *MNRAS*, 491, 1709
 Petri A., Haiman Z., Hui L., May M., Kratochvil J. M., 2013, *Phys. Rev. D*, 88, 123002
 Planck Collaboration VI, 2020, *A&A*, 641, A6
 Rao C. R., 1945, *Reson. J. Sci. Educ.*, 20, 78
 Tegmark M., Taylor A. N., Heavens A. F., 1997, *ApJ*, 480, 22
 Ueda H., Itoh M., 1997, *PASJ*, 49, 131
 Uhlemann C., Friedrich O., Villaescusa-Navarro F., Banerjee A., Codis S. r., 2020, *MNRAS*, 495, 4006
 van de Weygaert R., Jones B. J. T., Martínez V. J., 1992, *Phys. Lett. A*, 169, 145
 Villaescusa-Navarro F. et al., 2020, *ApJS*, 250, 2

APPENDIX A: MASKING SMALL SCALES USING GROUPS

Cosmological simulations tend to suffer from inaccuracies at small scales due to limitations in resolution and complex baryonic effects which require expensive hydrodynamics simulations. For these reasons, it is important to be able to remove these troublesome scales from our analysis. For traditional N -point statistics this is relatively straightforward as this simply requires placing scale cuts on the data vector. However, the MST is different as there is no way to mask small scales once the MST has been constructed; instead, the input data vector needs to be prepared such that scales are not present when the MST is constructed. To remove these scales from our analysis, we implement a grouping scheme where points separated by less than l_{\min} are combined.

In this section, we investigate the effectiveness of masking small scales by grouping using two Levy–Flight simulations as a proxy for N -body simulations with accurate and inaccurate small scale effects (such as baryonic effects). These simulations were chosen as their clustering properties are relatively simple to control and are fast to generate. We generate 50 realizations of two Levy–Flight simulation models: (1) the standard Levy Flight (LF) and (2) the Adjusted Levy Flight (ALF; see Naidoo et al. 2020, for more details). The simulations are generated using MISTREE (Naidoo 2019). For each simulation 50 000 points are generated in a box of length 75 with parameters $t_0 = 0.24$ and $\alpha = 1.6$ for the LF simulations and parameters $t_0 = 0.325$, $t_s = 0.015$, $\alpha = 1.5$, $\beta = 0.45$, and $\gamma = 1.3$ for the ALF.

The two-point correlation function (2PCF) and MST statistics are measured for both simulations and compared in Fig. A1. The simulations are designed to exhibit very different small-scale properties but identical large-scale correlation functions.

All points in the simulations with separations less than 0.4 are grouped and replaced by a single point with the average position of the group members. A new catalogue of points is now constructed with these grouped points and any remaining ungrouped points. For both simulations this results in a catalogue of roughly ~ 13 000 points. To ensure differences between the statistics are not due to slight differences in density, we sample at random 12 000 points from each catalogue and measure the 2PCF and MST statistics. In Fig. A1, the statistics before and after grouping are shown. Prior to grouping the simulations exhibit different small-scale clustering properties, which is seen by differences in the 2PCF and MST statistics, while after grouping all the statistics are consistent with each other. This shows that grouping effectively masks the small-scale differences between the simulations.

⁷<https://quijote-simulations.readthedocs.io/en/latest/index.html>

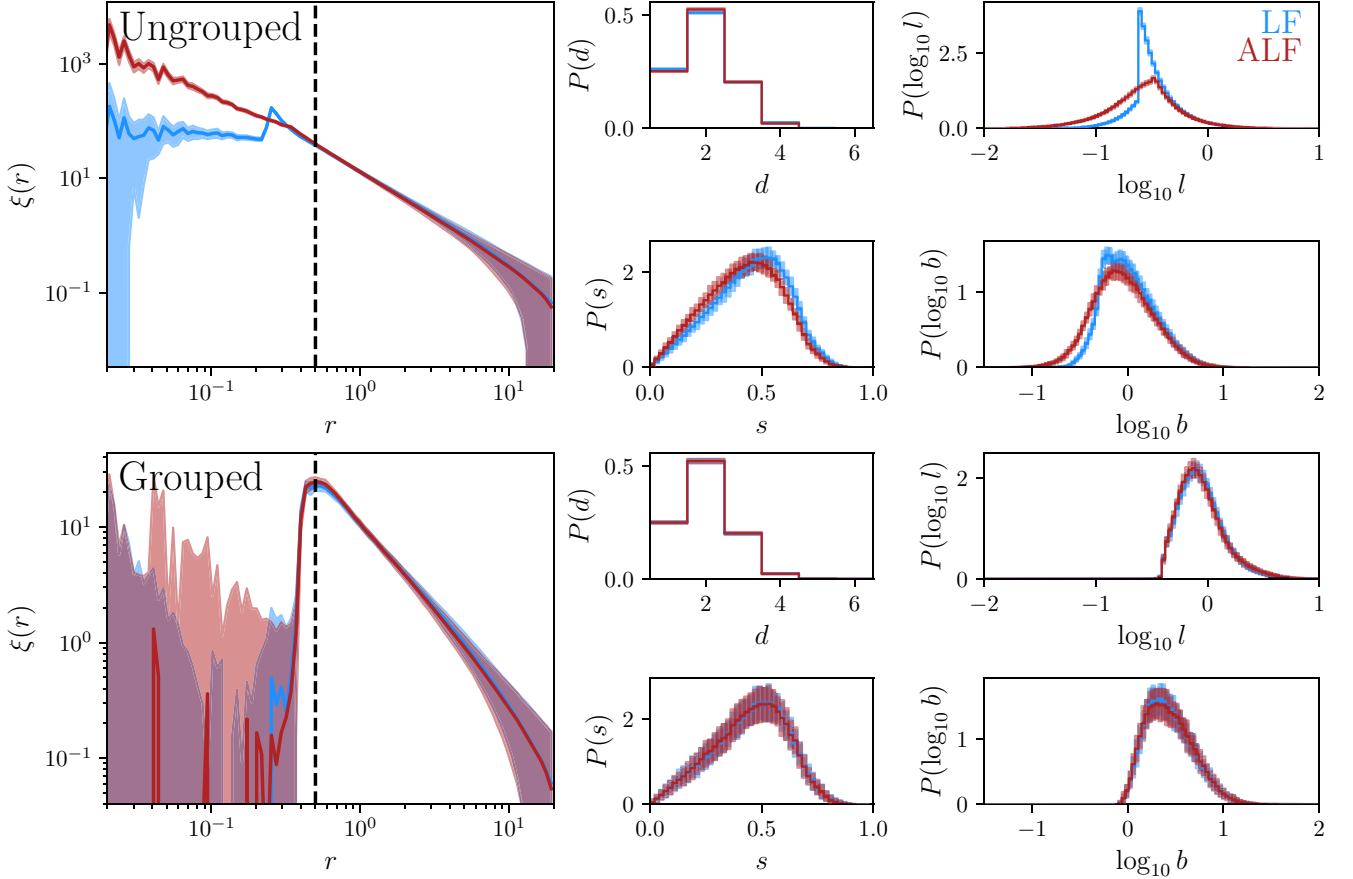


Figure A1. The effect of grouping Levy Flight simulations are shown for the two-point correlation function (2PCF) and MST statistics. Two sets of Levy Flight simulations are used: the standard Levy Flight (LF) simulations are shown in blue and the adjusted Levy Flight (ALF) simulations are shown in red. They exhibit almost identical large scale 2PCF $\xi(r)$ by design but have very different small-scale correlation properties. The differences between the ungrouped simulations are shown in the top subpanels. On the left is the 2PCF and on the right are the four MST statistics degree (d), edge length (l), and branch length (b). The differences between the grouped simulations are shown in the bottom subpanels. On the left is the 2PCF and on the right are the four MST statistics d , l , s , and b . These plots show that grouping provides an effective solution for masking small scales as the small-scale differences between the simulations do not appear once grouping has been performed.

APPENDIX B: MINIMUM SPANNING TREE PARTIAL DERIVATIVES

The partial derivative estimates for the MST statistics are shown for the distribution of degree d in Fig. B1, for the distribution of edges l in Fig. B2, for the distribution of branches b in Fig. B3, and for the distribution of branch shapes s in Fig. B4. The derivatives for neutrino mass are shown for estimator 2 (equation 5) with $M_\nu = 0.2$ eV.

For a fixed MST statistic, the partial derivatives of that statistic (with respect to the various cosmological parameters) all have similar shapes. The similarity can be explained if we think about the construction of the MST as an optimization problem: differences in parameters will lead to trees with longer (poorly optimized) or shorter edges (highly optimized). Depending on the outcome, the MST statistics will to first order be pulled to smaller or larger values but since there are roughly the same number of points this change has to be counter balanced with a reduction in the opposite direction. Although the profiles are similar in their general shape if we instead look to the relative peaks and troughs with respect to zero we can see that each parameter behaves slightly differently; take for example the partial derivatives for edge length with respect to Ω_m and M_{\min} where the troughs for M_{\min} is much deeper. The

similarities suggest that the MST statistics data vectors could be compressed into a few values (this possibility will be investigated in future work). In any case, the similarities are not a cause for concern as the interdependencies of these parameters are taken into account in the Fisher matrix calculation and would appear as covariances.

To test whether the partial derivatives have converged we compute the Fisher matrix from a subset of the total simulations available. In Fig. B5, we show the convergence of the MST partial derivatives by showing whether the individual components of the Fisher matrix are converging to their final values. Convergence is loosely defined to be Fisher matrix components which have settled to within 5 per cent of their final value when N_{deriv} (the number of data vectors used to compute partial derivatives) is greater than 1300. We find that estimator 2 (equation 5) with $M_\nu = 0.2$ eV provides the best convergence properties at all redshifts. Estimator 2 (equation 5) with $M_\nu = 0.1$ eV and Estimator 3 (equation 6) while being more accurate estimators tend to show poorer convergence at redshift $z = 0$ where the constraining power for M_ν is generally worse. This is likely the combination of two factors: (1) at low redshifts the halo catalogues are larger meaning more groupings take place resulting in some loss of information due to percolation and (2) gravity infall scrambles

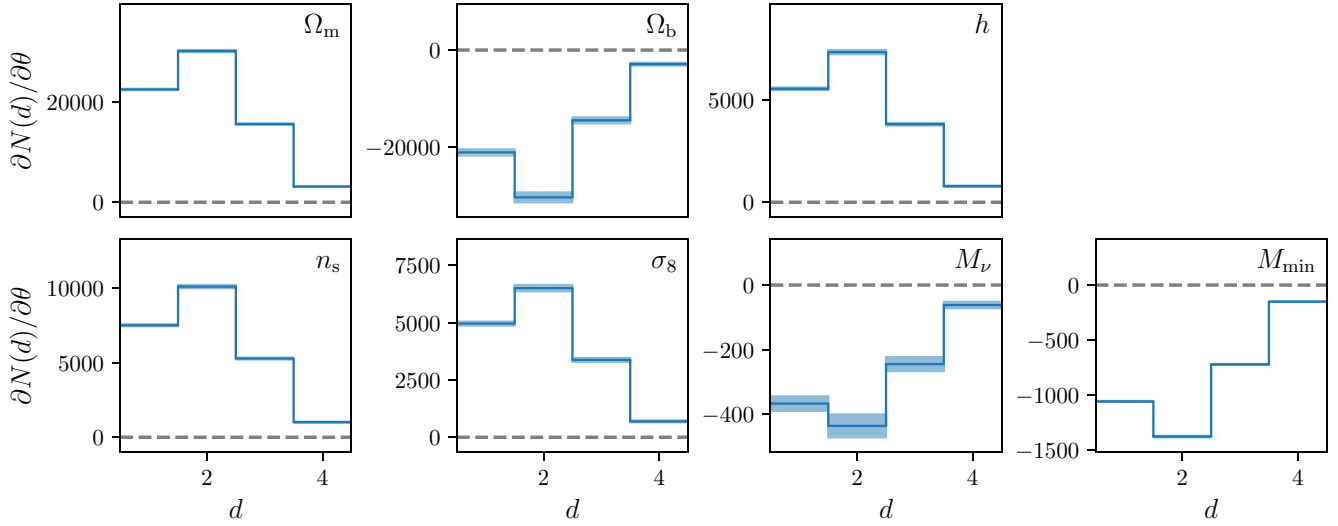


Figure B1. Derivatives for the MST degree d are shown for the six $\nu\Lambda$ CDM parameters (h , n_s , Ω_b , Ω_m , σ_8 , and M_ν) in redshift space and an additional nuisance parameter M_{\min} . The sensitivity to each parameter can be assessed by the significance of deviations away from $\partial N(d)/\partial\theta = 0$ (dotted black line).

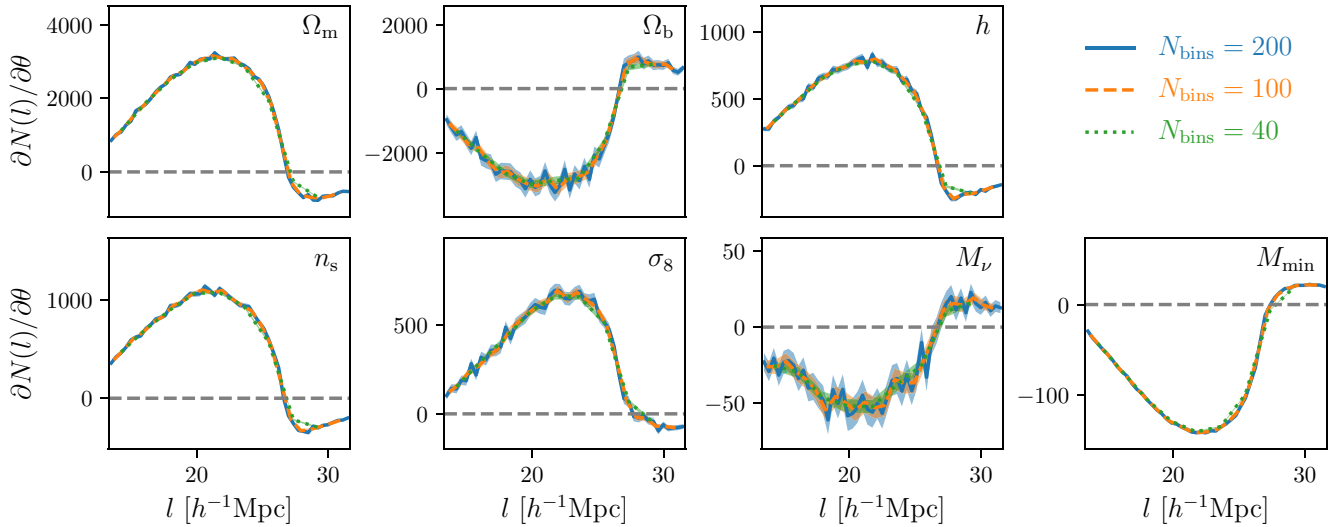


Figure B2. Derivatives for the MST edge length l are shown for the six $\nu\Lambda$ CDM parameters (h , n_s , Ω_b , Ω_m , σ_8 , and M_ν) in redshift space and an additional nuisance parameter M_{\min} . The derivatives are shown for three binning schemes (N_{bins}): 200 (blue), 100 (orange), and 40 (green).

the effect of neutrinos free-streaming making their effect harder to detect. We believe the former is the more important feature but this will need to be investigated in later work to understand how to limit

the effects of percolation on the MST, especially for dealing with scenarios where l_{\min} is larger than the mean separation between haloes.

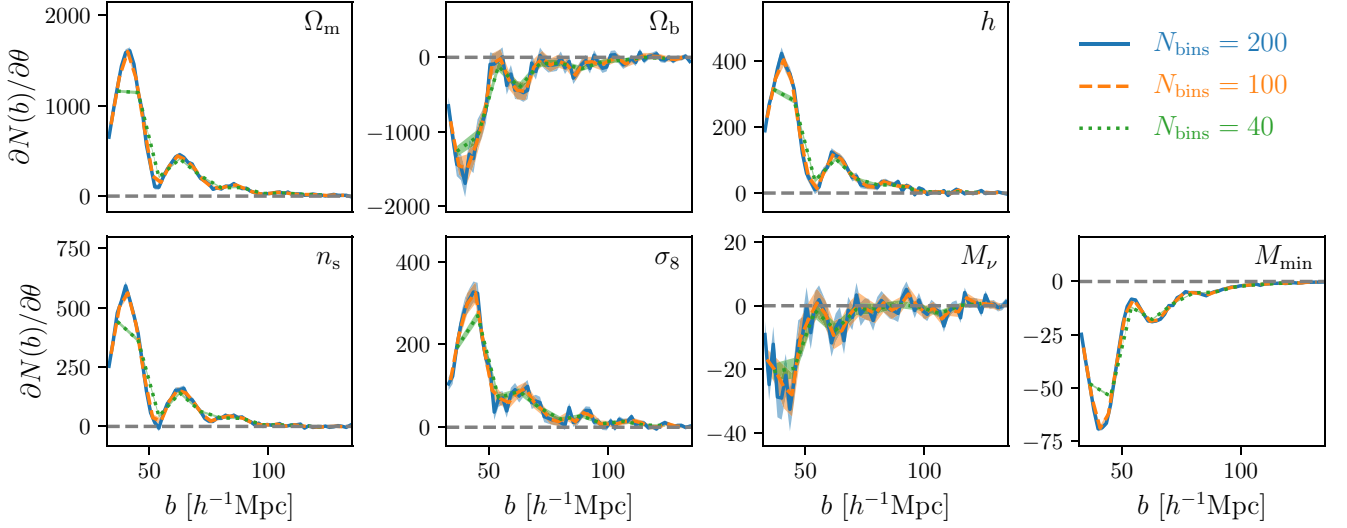


Figure B3. Derivatives for the MST branch length b are shown for the six $\nu\Lambda$ CDM parameters (h , n_s , Ω_b , Ω_m , σ_8 , and M_ν) in redshift space and an additional nuisance parameter M_{\min} . The sensitivity to each parameter can be assessed by the significance of deviations away from $\partial N(b)/\partial\theta = 0$ (dotted black line). The derivatives are shown for three binning schemes (N_{bins}): 200 (blue), 100 (orange), and 40 (green).

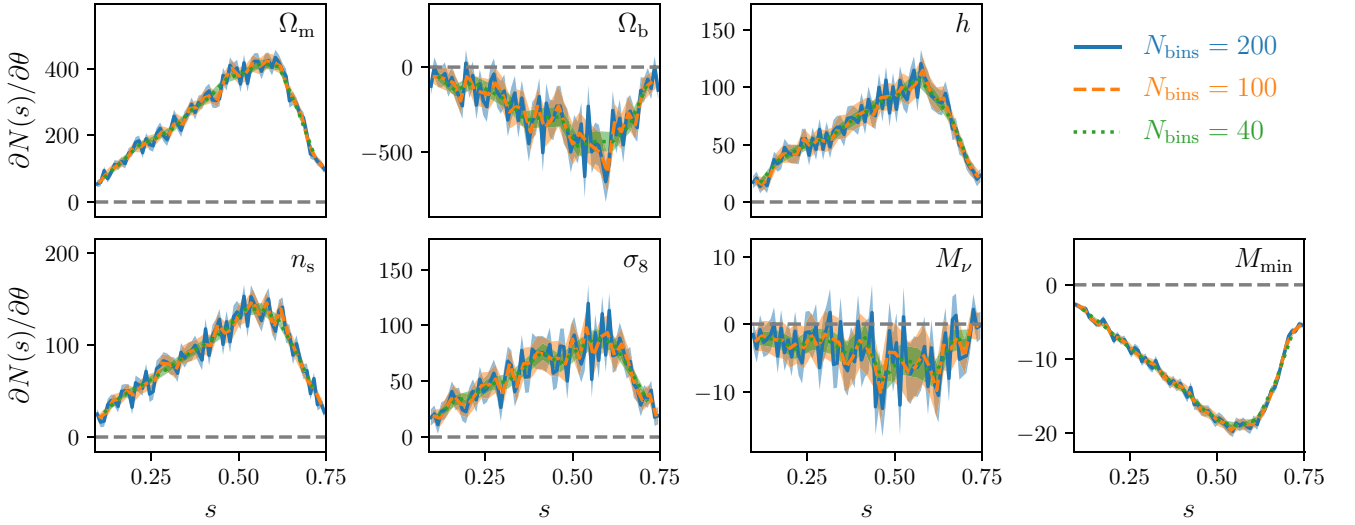


Figure B4. Derivatives for the MST branch shape s are shown for the six $\nu\Lambda$ CDM parameters (h , n_s , Ω_b , Ω_m , σ_8 , and M_ν) in redshift space and an additional nuisance parameter M_{\min} . The sensitivity to each parameter can be assessed by the significance of deviations away from $\partial N(s)/\partial\theta = 0$ (dotted black line). The derivatives are shown for three binning schemes (N_{bins}): 200 (blue), 100 (orange), and 40 (green).

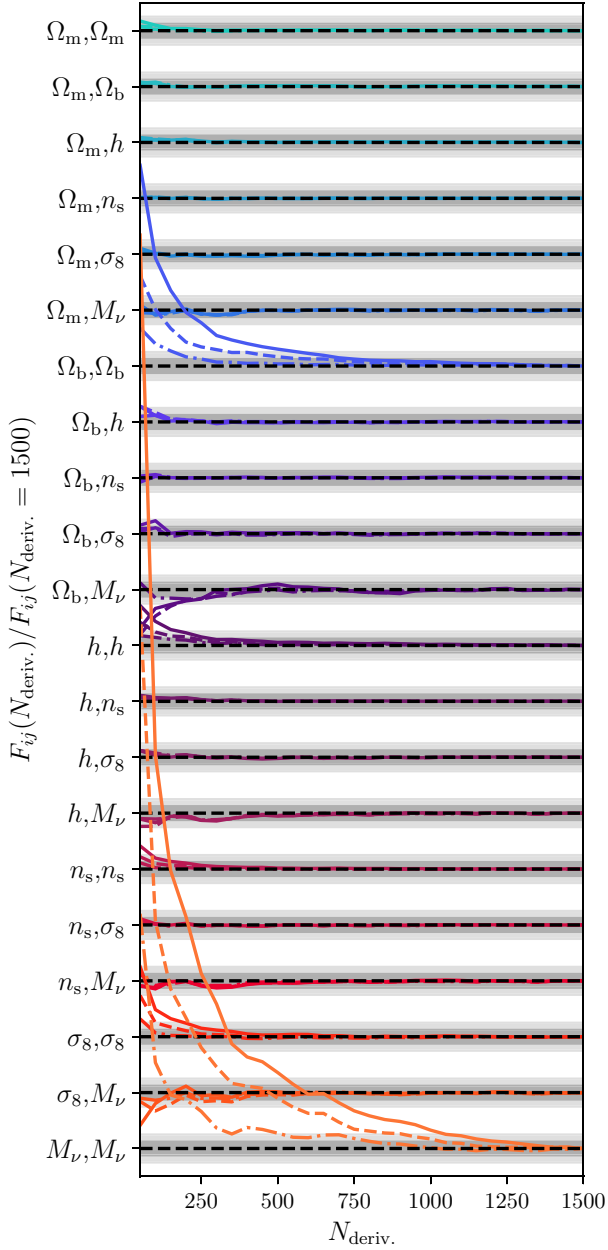


Figure B5. Convergence of the MST Fisher matrix elements as a function of the number of data vectors used for the derivative estimates. The x -axis shows the number of data vectors used to compute the partial derivatives ($N_{\text{deriv.}}$) of the MST statistics at redshift $z = 0.5$ (similar results are obtained for $z = 0$ and $z = 1$). On the y -axis the components of the Fisher matrix are shown. The grey bands show 10 per cent (light) and 5 per cent (dark) convergence from the Fisher matrix values after all 1500 derivative estimates are used. The full lines show the convergence for MST statistics with $N_{\text{bins}} = 200$, dashed lines with $N_{\text{bins}} = 100$, and dash-dotted lines with $N_{\text{bins}} = 40$. For most of these statistics convergence is reached relatively quickly with the exception of Ω_b and M_ν whose effects are generally more subtle. In both cases, we show convergence within 5 per cent is achieved for $N_{\text{deriv.}} > 1300$.

APPENDIX C: FISHER MATRIX CONSTRAINTS FROM THE MST

Additional constraints from the MST and power spectrum multipoles are presented in this section. In addition to the measurements made in the paper at redshift $z = 0.5$, we also make measurements at redshift $z = 0$ and $z = 1$, and finally combine the analysis over the three redshifts.

In Table C1, we compare the constraints of the MST and power spectrum multipoles for the different neutrino mass derivative estimates presented in this paper.

In Table C2, we compare the constraints of the MST and power spectrum multipoles for the six $\nu\Lambda\text{CDM}$ parameters and the nuisance minimum halo mass parameter M_{min} .

Table C1. Constraints on M_ν using different estimators for $\partial S/\partial M_\nu$ at $z = 0$: estimator 1 (equation 4) using $M_\nu = 0.1, 0.2$, and 0.4 eV, estimator 2 (equation 5) using $M_\nu = 0.1$ and 0.2 eV, and estimator 3 (equation 6) using $M_\nu = 0.1$ eV. Typically estimator 3 has been used in previous studies but in this paper we use estimator 2 with $M_\nu = 0.2$ eV since the accuracy of this estimator is consistent with equation (4) used for the other parameters. All the estimators in redshift space show tighter constraints for the MST than the power spectrum. Furthermore, when they are combined the constraints appear to be dominated by the MST.

Statistics	Redshift	Est. 1 (equation 4)			Est. 2 (equation 5)		Est. 3 (equation 6)
		0.1 eV	0.2 eV	0.4 eV	0.1 eV	0.2 eV	0.1 eV
MST(l)	0.5	0.5	1.0	1.8	0.29	0.6	0.22
MST(b)	0.5	0.45	0.7	1.4	0.26	0.38	0.2
MST(l, b)	0.5	0.33	0.56	1.1	0.19	0.31	0.15
MST(l, b, s)	0.5	0.23	0.46	0.79	0.13	0.25	0.099
MST(d, l, b, s)	0.5	0.23	0.45	0.76	0.13	0.25	0.099
$P_{\ell=0,2}(k)$	0.5	0.82	1.6	1.4	0.44	1.0	0.33
$P_{\ell=0,2}(k) + \text{MST}(d, l, b, s)$	0.5	0.21	0.38	0.44	0.12	0.23	0.093

Table C2. Separate and combined constraints for parameters from the $\nu\Lambda$ CDM model determined from measurements of the power spectrum (multipoles in redshift space) and MST at redshift $z = 0, 0.5$, and 1 . The constraints are obtained using equation (3) for all of the parameters except M_ν , which are obtained using equation (5) with $M_\nu = 0.2$ eV. For the standard Λ CDM parameters we obtain competitive constraints from the MST at all redshifts, with the exception of Ω_m and σ_8 where the power spectrum dominates; however, for M_ν the MST dominates. When measurements from the different redshifts are combined we find the MST is competitive for all parameters, including Ω_m and σ_8 , but still dominates constraints on M_ν .

Statistics	Redshift	N_{bins}	$\Delta\Omega_m$	$\Delta\Omega_b$	Δh	Δn_s	$\Delta\sigma_8$	ΔM_ν [eV]	ΔM_{min}
MST(l)	0	1	0.076	0.018	0.13	0.12	0.14	0.53	1.2
MST(b)	0	1	0.062	0.015	0.14	0.14	0.098	0.45	1.4
MST(l, b)	0	1	0.048	0.011	0.089	0.091	0.079	0.33	0.75
MST(l, b, s)	0	1	0.036	0.0083	0.073	0.067	0.067	0.24	0.59
MST(d, l, b, s)	0	1	0.036	0.0083	0.073	0.065	0.067	0.23	0.55
$P_{\ell=0,2}(k)$	0		0.029	0.013	0.13	0.079	0.061	1.1	0.22
$P_{\ell=0,2}(k) + \text{MST}(d, l, b, s)$	0	1	0.015	0.0059	0.054	0.036	0.016	0.22	0.14
$P_{\ell=0,2}(k) + \text{MST}(d, l, b, s)$	0	2	0.016	0.0074	0.069	0.041	0.019	0.29	0.14
$P_{\ell=0,2}(k) + \text{MST}(d, l, b, s)$	0	5	0.017	0.0099	0.093	0.046	0.023	0.37	0.15
MST(l)	0.5	1	0.062	0.024	0.19	0.15	0.14	0.6	1.2
MST(b)	0.5	1	0.065	0.016	0.15	0.12	0.099	0.38	1.2
MST(l, b)	0.5	1	0.042	0.013	0.11	0.091	0.079	0.31	0.73
MST(l, b, s)	0.5	1	0.035	0.0092	0.089	0.068	0.062	0.25	0.59
MST(d, l, b, s)	0.5	1	0.035	0.0092	0.088	0.067	0.061	0.25	0.58
$P_{\ell=0,2}(k)$	0.5		0.023	0.014	0.14	0.09	0.065	1.0	0.2
$P_{\ell=0,2}(k) + \text{MST}(d, l, b, s)$	0.5	1	0.013	0.0068	0.066	0.043	0.017	0.23	0.13
$P_{\ell=0,2}(k) + \text{MST}(d, l, b, s)$	0.5	2	0.014	0.0086	0.084	0.051	0.022	0.33	0.14
$P_{\ell=0,2}(k) + \text{MST}(d, l, b, s)$	0.5	5	0.016	0.01	0.1	0.058	0.03	0.46	0.14
MST(l)	1	1	0.055	0.018	0.17	0.12	0.097	0.52	1.2
MST(b)	1	1	0.066	0.016	0.16	0.11	0.089	0.38	1.1
MST(l, b)	1	1	0.04	0.011	0.11	0.079	0.063	0.3	0.77
MST(l, b, s)	1	1	0.034	0.0088	0.086	0.064	0.051	0.24	0.62
MST(d, l, b, s)	1	1	0.034	0.0087	0.08	0.063	0.051	0.24	0.58
$P_{\ell=0,2}(k)$	1		0.024	0.016	0.18	0.14	0.065	0.88	0.25
$P_{\ell=0,2}(k) + \text{MST}(d, l, b, s)$	1	1	0.014	0.0063	0.057	0.046	0.02	0.23	0.12
$P_{\ell=0,2}(k) + \text{MST}(d, l, b, s)$	1	2	0.016	0.0078	0.073	0.054	0.025	0.32	0.13
$P_{\ell=0,2}(k) + \text{MST}(d, l, b, s)$	1	5	0.017	0.01	0.098	0.069	0.033	0.43	0.14
MST(l)	All	1	0.034	0.01	0.083	0.064	0.035	0.29	0.46
MST(b)	All	1	0.031	0.0083	0.075	0.061	0.033	0.21	0.45
MST(l, b)	All	1	0.023	0.0065	0.055	0.045	0.022	0.17	0.32
MST(l, b, s)	All	1	0.019	0.0049	0.043	0.035	0.019	0.14	0.26
MST(d, l, b, s)	All	1	0.019	0.0049	0.043	0.035	0.019	0.14	0.26
$P_{\ell=0,2}(k)$	All		0.014	0.0083	0.07	0.048	0.026	0.37	0.15
$P_{\ell=0,2}(k) + \text{MST}(d, l, b, s)$	All	1	0.0081	0.0039	0.033	0.025	0.01	0.13	0.071
$P_{\ell=0,2}(k) + \text{MST}(d, l, b, s)$	All	2	0.0091	0.0049	0.041	0.029	0.013	0.17	0.073
$P_{\ell=0,2}(k) + \text{MST}(d, l, b, s)$	All	5	0.01	0.0062	0.052	0.033	0.015	0.21	0.075