

# The Impact of Cosmic Variance and Satellites on JWST Clustering Measurements at redshift $\approx 6$

Jiamu Huang<sup>1\*</sup>, Elia Pizzati<sup>2†</sup>, Joseph Hennawi<sup>1,3</sup>, Joop Schaye<sup>3</sup>, Matthieu Schaller<sup>3,4</sup>, Benjamin Snyder<sup>1</sup>, Yi Kang<sup>1</sup>

<sup>1</sup>*Department of Physics, University of California, Santa Barbara, CA 93106, USA*

<sup>2</sup>*Center for Astrophysics | Harvard & Smithsonian, 60 Garden St., Cambridge, MA 02138, USA*

<sup>3</sup>*Leiden Observatory, Leiden University, NL-2300 RA Leiden, The Netherlands*

<sup>4</sup>*Lorentz Institute for Theoretical Physics, Leiden University, PO Box 9506, NL2300 RA Leiden, The Netherlands*

Accepted XXX. Received YYY; in original form ZZZ

## ABSTRACT

We present a framework for inferring the dark matter halo masses of quasars and [O III]-emitting galaxies from JWST/NIRCam Wide Field Slitless Spectroscopy (WFSS) clustering measurements at  $z \approx 6$ . Using the FLAMINGO-10k N-body simulation, we construct mock realizations of quasar and galaxy catalogs that incorporate realistic selection functions, spatial coverage, and sensitivity limits matched to the ASPIRE survey. These mocks enable accurate measurements of the quasar–galaxy cross-correlation and galaxy auto-correlation functions, with covariance matrices derived from 1000 realizations that capture both cosmic variance and bin-to-bin correlations. We employ Bayesian inference to fit the correlation functions and infer the minimum halo masses for quasars and galaxies. Our results demonstrate that Poisson pair-count uncertainties, commonly adopted in high-redshift clustering studies, significantly underestimate the true measurement errors. The dominant missing component is cosmic variance: even the diagonal of the full covariance matrix exceeds the Poisson expectation, with off-diagonal bin-to-bin correlations contributing a smaller additional correction. In particular, 1) the commonly used Poisson error on the correlation functions underestimates the true uncertainty by a factor of  $\approx 3$ ; 2) the uncertainties on the inferred minimum halo masses are underestimated by a factor of  $\approx 1.5$ –3 when adopting Poisson errors instead of the full covariance matrix; 3) the inferred QSO halo mass is robust to whether central and satellite [O III]-emitters share a common mass threshold. Our framework provides a more complete error budget for JWST/WFSS clustering analyses, enabling robust constraints on the host halo masses and duty cycles of high-redshift quasars and emission-line galaxies.

**Key words:** galaxies: high-redshift - quasars: supermassive black holes - galaxies haloes - large-scale structure of Universe

## 1 INTRODUCTION

The detection of quasars at high redshift hosting supermassive black holes (SMBHs) with masses of  $\sim 10^9 M_\odot$  (Mortlock et al. 2011; Matsuoka et al. 2019; Yang et al. 2020; Wang et al. 2021; Fan et al. 2023) poses significant challenges to our understanding of black hole formation and growth in the early Universe. To observationally constrain the growth scenarios of early SMBHs, it is essential to determine not only how massive these black holes are when active, but also what fraction of their growth occurs during the UV-bright phase that we observe as quasars. Quasar clustering provides a direct way to quantify this through measurements of the quasar duty cycle  $f_{\text{duty}} \equiv n_{\text{quasar}}/n_{\text{halo}}$ , the ratio of the observed quasar number density to the number density of their host dark matter halos (Efstathiou & Rees 1988; Cole & Kaiser 1989; Haiman & Hui 2001; Martini & Weinberg 2001). Since more massive dark matter halos exhibit stronger clustering due to their higher bias (Kaiser 1984; Mo & White

2002), we can statistically infer the typical host halo mass of quasars from the observed clustering strength, and thereby determine the duty cycle from the abundance of halos above that mass threshold. Under the assumption that each halo undergoes one quasar phase, the duty cycle is equivalent to the quasar lifetime as a fraction of the Hubble time,  $f_{\text{duty}} \approx t_{\text{Q}}/t_{\text{H}}$  (Efstathiou & Rees 1988; Cole & Kaiser 1989; Haiman & Hui 2001; Martini & Weinberg 2001).

Clustering analysis through the two-point correlation function (2PCF) has become a powerful tool to probe the environments and host halo masses of quasars (Osmer 1981; Shen et al. 2007). At redshifts  $1 \lesssim z \lesssim 3$ , quasar clustering measurements indicate host halo masses of order of  $10^{12} M_\odot$ , increasing to  $\sim 10^{13} M_\odot$  at  $z \approx 4$  (Porciani et al. 2004; Croom et al. 2005; Coil et al. 2007; Pizzati et al. 2024a; Giner Mascarell et al. 2025).

However, high-redshift measurements of quasar auto-correlation functions become challenging due to the low spatial density of luminous quasars at  $z \gtrsim 6$  (Wang et al. 2019; Schindler et al. 2023). Recently, Arita et al. (2023) measured the quasar auto-correlation at  $z \approx 6$  by targeting the fainter, more abundant quasar population ( $M_{1450} \approx -25$ ) from the SHELLQs survey, although the resulting

\* E-mail: jiamu\_huang@ucsb.edu

† NHFP Einstein Fellow

constraints on the host halo mass carry large statistical uncertainties due to the limited sample size, motivating the use of cross-correlation measurements with more abundant tracers. For the rarer, more luminous quasars, cross-correlation measurements between quasars and galaxies offer a complementary approach for determining host halo masses at high redshift (García-Vergara et al. 2017, 2019, 2022; Eilers et al. 2024; Pizzati et al. 2024b; Schindler et al. 2026).

Before *JWST*, empirical studies of quasar environments at  $z > 5$  yielded mixed results: narrow band searches for Ly $\alpha$  emitters (LAEs) and Lyman break galaxies (LBGs) sometimes reported galaxy excesses around high- $z$  quasars (e.g., Kim et al. 2009; Morselli et al. 2014), whereas others found no significant overdensities (Bañados et al. 2013; Simpson et al. 2014). These mixed conclusions likely reflected observational limitations, including small fields of view, heterogeneous depth, redshift offsets between narrow band filters and quasar systemic redshifts, and the heterogeneity of the LAE/LBG tracer populations themselves, whose host halo masses are poorly constrained but directly affect the interpretation of the cross-correlation signal (Overzier 2016). More recently, Arita et al. (2026) studied four low-luminosity quasars ( $-24 < M_{1450} < -22$ ) at  $z \approx 6.2$  using Subaru/HSC narrowband imaging of Ly $\alpha$  emitters, finding that only one of four fields shows an overdensity ( $\delta_{\text{LAE}} = 3.77 \pm 0.97$ ) while the remaining three are consistent with the field, further illustrating the diversity of quasar environments inferred from narrowband studies.

Recent observations with *JWST*/NIRCam using the Wide Field Slitless Spectroscopy (WFSS) mode have opened new opportunities to probe quasar environments with higher sensitivity and larger spatial coverage. Several ongoing programs are now using WFSS to search for line-emitting galaxies, including [O III] and H $\alpha$  emitters, in quasar fields. These studies have revealed an abundance of [O III]-emitting galaxies around some high-redshift quasars with  $z \gtrsim 6$  (Wang et al. 2023; Kashino et al. 2023; Eilers et al. 2024; Champagne et al. 2025), suggesting that luminous quasars may reside in significant overdensities. However, Schindler et al. (2026) measured the quasar–[O III] emitter cross-correlation for two quasars at  $z \approx 7.3$  and found a weaker clustering signal, indicating that quasar environments may vary from object to object. The growing sample of emission-line galaxy tracers has enabled more precise clustering analyses: the combination of [O III]-emitter auto-correlation and quasar–[O III]-emitter cross-correlation measurements has yielded the first joint constraints on the dark matter halo masses hosting galaxies and quasars at  $z > 6$  (Eilers et al. 2024; Pizzati et al. 2024b; Huang et al. 2026), with achievable measurement precision broadly consistent with pre-launch forecasts of *JWST*/WFSS clustering uncertainties (Endsley et al. 2020).

Beyond quasar–galaxy cross-correlations, *JWST*/NIRCam WFSS has enabled galaxy auto-correlation measurements at high redshift. Shuntov et al. (2025) measured the angular auto-correlation of H $\alpha$  emitters at  $3.7 < z < 5.1$  (CONGRESS, F356W) and  $4.9 < z < 6.7$  (FRESCO, F444W; Oesch et al. 2023), and [O III] emitters at  $6.8 < z < 9.0$  (FRESCO, F444W) in the GOODS fields. Through HOD modeling of the clustering and UV luminosity function, they inferred characteristic halo masses decreasing from  $\log(M_h/M_\odot) = 11.46$  at  $\langle z \rangle = 4.3$  to  $10.99$  at  $\langle z \rangle = 7.3$ . Lin et al. (2025a) measured the auto-correlation of H $\alpha$  emitters at  $z = 4$ – $6$  using the same CONGRESS+FRESCO data, finding correlation lengths of  $r_0 = 4.61^{+1.00}_{-0.68}$  cMpc  $h^{-1}$  at  $z = 4$ – $5$  and  $r_0 = 6.23^{+1.68}_{-1.13}$  cMpc  $h^{-1}$  at  $z = 5$ – $6$ , corresponding to host halo masses of  $\log(M_h/M_\odot) \approx 11.0$ – $11.2$ . Lin et al. (2025b) further measured the AGN–H $\alpha$  emitter cross-correlation at  $3.9 < z < 6$  and found a comparable clustering amplitude to the HAE auto-correlation, indicating that low-luminosity AGNs share similar bias and host halo masses ( $\log(M_h/M_\odot) \approx 11.0$ –

$11.2$ ) with HAEs. This places them in a population distinct from the more biased, more massive luminous quasars.

From the theoretical perspective, galaxy clustering at high redshift breaks a key degeneracy in the UV luminosity function: the same observed number counts can be produced either by bright galaxies in rare, massive halos (high duty cycle) or by occasional bright phases of more common, lower-mass halos (low duty cycle). Since these two scenarios predict different clustering amplitudes, measuring the galaxy auto-correlation function constrains the typical host halo mass and duty cycle of the bright UVLF tracers (Muñoz et al. 2023). Weibel et al. (2025) recently applied this approach at  $z \approx 10$ , using 34 independent NIRCam sightlines from PANORAMIC to measure the cosmic variance of Lyman-break galaxies and, combined with the UV luminosity function, disfavor models with globally enhanced star-formation efficiency.

A critical but underappreciated issue in all existing high- $z$  clustering analyses is the treatment of measurement uncertainties. It is common practice in high-redshift quasar clustering studies to estimate the uncertainty of the correlation function using Poisson errors on the pair counts (García-Vergara et al. 2017, 2019; Eilers et al. 2024; Schindler et al. 2026). This approach assumes that the covariance matrix is diagonal and ignores two key sources of correlated uncertainty. First, different galaxy pairs are not statistically independent because the same galaxies contribute to multiple pairs, so any large-scale fluctuation in the galaxy density (for example, a locally overdense region) boosts the number of pairs across a wide range of separations, producing positive correlations between radial bins. Second, cosmic variance, i.e. the field-to-field variation in the large-scale density field, introduces coherent fluctuations that affect all bins simultaneously. The Poisson approximation therefore underestimates not only the diagonal error bars but also completely neglects the off-diagonal covariances between radial bins, so that the true error budget is substantially larger and more correlated than implied by Poisson statistics alone. This effect has recently been quantified using the FLAMINGO suite by Lim et al. (2025), who showed that for *JWST*-like (100 cMpc)<sup>3</sup> survey volumes at  $z \approx 6$ , the field-to-field variance in massive-halo number counts is 2–3 times the Poisson expectation, with a comparable enhancement in the variance of the most-massive halo per field.

With only a few independent quasar fields per program (e.g., four in EIGER, Eilers et al. 2024), internal error estimators such as jackknife or bootstrap resampling cannot provide stable covariance estimates: the resulting covariance matrices are typically noisy and can become poorly conditioned or even singular (Norberg et al. 2009). This effectively forces high- $z$  studies (Eilers et al. 2024; Schindler et al. 2026) to adopt Poisson pair-count uncertainties, leaving the reported error bars as underestimates of the true uncertainty (Yang et al. 2005; Robertson 2010). In contrast, at low-to-intermediate redshifts ( $z \lesssim 3$ ), the methodology for robust clustering inference is well established. Large galaxy surveys such as the Sloan Digital Sky Survey (SDSS; Zehavi et al. 2011; Zheng et al. 2007; Guo et al. 2015), the Baryon Oscillation Spectroscopic Survey (BOSS; Manera et al. 2013; White et al. 2014; Kitaura et al. 2016), the Dark Energy Survey (DES; Friedrich et al. 2021), and most recently the Dark Energy Spectroscopic Instrument (DESI; Forero-Sánchez et al. 2025) have invested heavily in constructing hundreds to thousands of mock galaxy catalogs from N-body simulations or fast approximate methods (see Lippich et al. 2019, for a systematic comparison) to estimate covariance matrices for correlation function measurements. These covariance matrices are then used directly in HOD fitting of the volume-averaged correlation function  $w_p(r_p)$  to infer halo masses

and the galaxy–halo connection (Sinha et al. 2018; van den Bosch et al. 2013).

Empirical covariance estimators such as jackknife or bootstrap resampling require many independent sub-volumes to converge (Norberg et al. 2009), and the small number of independent ASPIRE fields makes such observation-based estimators unreliable in our regime. We therefore compute the covariance from forward-modeled mock catalogs, which capture both cosmic variance and correlated bin-to-bin uncertainties. These mock ensembles provide two key ingredients: (i) mock covariance matrices that capture cosmic variance and correlated bin-to-bin uncertainties, and (ii) simulation-based clustering models that correctly account for non-linear structure growth and scale-dependent bias at small separations. The importance of full covariance modeling has been demonstrated repeatedly. Norberg et al. (2009) showed that internal error estimators such as jackknife and bootstrap are biased relative to ensemble-based covariance from independent mock realizations, with neither method reliably reproducing the off-diagonal covariance structure on scales of  $1\text{--}25 h^{-1} \text{Mpc}$  (Mohammad & Percival 2022). The practical requirements for the number of mock realizations needed to obtain an unbiased inverse covariance matrix have been quantified by Hartlap et al. (2007), Dodelson & Schneider (2013), and Percival et al. (2014), who showed that using too few mocks inflates parameter confidence intervals and that  $O(10^3)$  realizations are typically required for stable estimates (see also Sellentin & Heavens 2016). Moreover, within the framework of the halo occupation distribution (HOD; Berlind & Weinberg 2002; Zheng et al. 2005), the covariance matrix itself depends on the assumed halo mass parameters: more massive, more biased tracers produce stronger cosmic variance and more prominent off-diagonal correlations. As we show in Section 5.3, ignoring this mass dependence of the covariance, or replacing the full covariance with diagonal Poisson errors, underestimates the true uncertainty on inferred halo masses by factors of  $\approx 1.5\text{--}3\times$ .

The need for proper covariance modeling is even more acute at high redshift. The narrow pencil-beam fields of *JWST*/WFSS ( $\approx 2' \times 2'$  each) make cosmic variance the dominant source of uncertainty (Yang et al. 2005; Trenti & Stiavelli 2008; Moster et al. 2011), and the highly biased tracers at  $z > 5$  further amplify the effect. In our companion paper (Huang et al. 2026), we apply these mass-dependent covariance matrices to the ASPIRE observations and show that adopting Poisson errors underestimates the  $1\sigma$  uncertainty on the inferred halo masses by about 0.7 dex for quasar minimum halo mass ( $M_{h,\min}^{\text{QSO}}$ ) and 0.5 dex for galaxy host minimum mass ( $M_{h,\min}^{\text{[O III]}}$ ).

In addition to getting the error budget correct, simulations are also required to physically interpret the measured clustering amplitude in terms of halo masses. In previous high-redshift studies, halo mass estimates are commonly obtained by fitting a power-law model to the large-scale clustering signal, converting the fitted amplitude into an effective bias, and then associating this bias with a characteristic halo mass using analytical prescriptions calibrated on linear theory (e.g., Arita et al. 2023). This approach is not appropriate for *JWST* WFSS measurements of small pencil-beam fields, which probe small scales ( $r_p \sim 0.1\text{--}5 \text{cMpc } h^{-1}$ ) and highly biased tracers such as quasars at high redshift. At these redshifts and separations, the halo bias becomes strongly non-linear and scale-dependent, and standard analytical bias models do not capture these effects. In the quasi-linear regime near the one-halo to two-halo transition, analytical models can significantly underpredict the clustering amplitude if these non-linearities are not treated correctly (see discussions in Jose et al. 2016; Shuntov et al. 2025). Recent work has also shown that fitting a rigid power-law correlation function (as in e.g. Arita et al. 2023; Eilers

et al. 2024) can lead to biased halo mass estimates that reflect the assumed power-law shape rather than the underlying halo population (demonstrated on large scales in Appendix Fig. C1 of Pizzati et al. 2024b). In contrast, our simulation-based clustering model links the observed signal directly to the spatial distribution of halos in the simulation, providing a physically grounded inference of the masses of quasar-host halos.

This work addresses the gap between large-area clustering surveys, where mock covariance methodology is well established, and the small-area, pencil-beam *JWST*/WFSS regime where cosmic variance dominates the error budget, by constructing mock covariance matrices from the FLAMINGO-10k (F10k) cosmological simulation for *JWST*-style clustering inference. We employ the F10k dark-matter-only simulation (Schaller et al., in prep.; see also Pizzati et al. 2024b), a  $(2.8 \text{cGpc})^3$  volume containing  $10080^3$  CDM particles from the FLAMINGO suite (Schaye et al. 2023; Kugel et al. 2023), to generate realistic mock catalogs of [O III]-emitters around  $z \sim 6.6$  quasars. Each mock realization replicates the observational properties of the *JWST* ASPIRE survey, incorporating the field-by-field spatial coverage, spectral sensitivity, and flux limits through the same selection function applied to the real data. From 1000 such realizations, we construct the full covariance matrices for the [O III]-emitter auto-correlation function  $\chi_{\text{GG}}(r_p)$  and the quasar-[O III]-emitter cross-correlation function  $\chi_{\text{QG}}(r_p)$ . Crucially, these covariance matrices are computed on a two-dimensional grid of minimum halo masses ( $M_{h,\min}^{\text{QSO}}, M_{h,\min}^{\text{[O III]}}$ ), capturing the mass dependence of the covariance structure that arises from the bias dependence of cosmic variance.

Our approach differs from prior high- $z$  clustering studies in three specific aspects. First, we use mock covariance matrices based on F10k that capture cosmic variance and off-diagonal bin-to-bin correlations, rather than Poisson pair-count errors. Second, we employ simulation-based clustering models that accurately reproduce the non-linear, scale-dependent halo bias at small separations, rather than analytical power-law prescriptions. Third, we incorporate the full mass dependence of the covariance matrix into the likelihood, so that the error budget is self-consistently tied to the halo mass model. Together, these elements bring the robust statistical methodology of low- $z$  clustering analyses to the high- $z$  regime for the first time.

The paper is structured as follows. In Section 2 we define the volume-averaged correlation function estimators. Section 3 describes the construction of the selection function, including coverage and sensitivity maps. Section 4 presents the mock catalog generation and covariance matrix construction. In Section 5 we describe the Bayesian inference framework and compare constraints from the full covariance to those from Poisson errors. We summarize our conclusions in Section 6.

## 2 CORRELATION FUNCTION ESTIMATORS

We quantify galaxy clustering using the volume-averaged correlation function  $\chi$  (Hennawi et al. 2006; García-Vergara et al. 2017), which measures the mean overdensity within cylindrical shells in redshift space. For a shell bounded by projected separations  $r_p^{\min}$  and  $r_p^{\max}$ , this is defined as:

$$\chi(r_p^{\min}, r_p^{\max}) = \frac{2}{V} \int_{r_p^{\min}}^{r_p^{\max}} \int_0^{\pi_{\max}} \xi(r_p, \pi) 2\pi r_p dr_p d\pi, \quad (1)$$

where  $\xi(r_p, \pi)$  is the two-dimensional correlation function,  $\pi$  is the comoving line-of-sight separation, and  $V$  is the cylindrical shell volume bounded by  $r_p^{\min}, r_p^{\max}$ , and  $\pi_{\max}$ . Following Huang et al.

(2026), we adopt  $\pi_{\max} = 7 \text{ cMpc } h^{-1}$ , corresponding to a line-of-sight velocity of  $dv = 1037 \text{ km s}^{-1}$  at  $z = 6.5$ . The pair counts entering these estimators are computed with `Corrfunc` (Sinha & Garrison 2020), a suite of highly optimized, OPENMP parallel clustering codes.

For the [O III]-emitter auto-correlation, we use the Landy–Szalay estimator (Landy & Szalay 1993):

$$\chi_{GG}(r_p^{\min}, r_p^{\max}, \pi_{\max}) = \frac{\langle D_G D_G \rangle - 2\langle D_G R_G \rangle + \langle R_G R_G \rangle}{\langle R_G R_G \rangle}, \quad (2)$$

where  $\langle D_G D_G \rangle$ ,  $\langle D_G R_G \rangle$ , and  $\langle R_G R_G \rangle$  are the normalized data–data, data–random, and random–random pair counts, respectively. For the quasar–[O III]-emitter cross-correlation, each ASPIRE pointing contains only a single quasar at a fixed location, so no random quasar catalog is generated. We therefore adopt the Davis & Peebles (1983) estimator:

$$\chi_{QG}(r_p^{\min}, r_p^{\max}, \pi_{\max}) = \frac{\langle D_Q D_G \rangle}{\langle D_Q R_G \rangle} - 1, \quad (3)$$

where  $\langle D_Q D_G \rangle$  and  $\langle D_Q R_G \rangle$  are the normalized quasar–galaxy and quasar–random pair counts. The normalization of all pair counts follows the convention described in Huang et al. (2026).

### 3 COMPLETENESS AND SELECTION FUNCTION

Accurate random catalogs for the Landy–Szalay and Davis–Peebles estimators require two products that characterize the ASPIRE selection function: a spectral coverage map and a sensitivity map. We briefly summarize their construction here, with a more detailed discussion provided in Huang et al. (2026).

#### 3.1 Coverage Map

The spectral coverage map determines whether an [O III]-emitter at a given sky position and redshift has both [O III]  $\lambda\lambda 4960, 5008$  lines falling within the WFSS detector boundaries. For each ASPIRE pointing, we generate a grid of test positions from the F356W direct imaging, map each position through the grism WCS and the wavelength-dependent dispersion solution (Sun et al. 2024) to all dithered WFSS frames, and flag the source as covered if the dispersed lines fall on the detector in at least one exposure. The coverage map is evaluated over  $z = 5.3\text{--}7.0$ , corresponding to a comoving line-of-sight depth of  $\approx 485 \text{ cMpc } h^{-1}$ , to capture the redshift-dependent spatial coverage.

#### 3.2 Sensitivity Map

The sensitivity map accounts for spatial variations in the detection threshold arising from differences in exposure depth, background level, and detector sensitivity between NIRCcam modules A and B. A grid of mock [O III]-emitters with fixed flux ( $f_{\text{mock}} = 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$ ) and line width (FWHM =  $200 \text{ km s}^{-1}$ ) is injected into the observed WFSS data and extracted with `unfold_JWST` (Wang et al. 2023). At each grid position, the  $5\sigma$  flux limit is defined as  $F_{\text{lim}} \equiv 5 f_{\text{mock}} / (S/N)_{\text{mock}}$ .

#### 3.3 Random Catalog Construction

To construct the random catalog, we employ a Monte Carlo approach. For each ASPIRE pointing,  $5 \times 10^6$  mock sources are placed uniformly in RA and Dec within the F356W imaging footprint. Each source

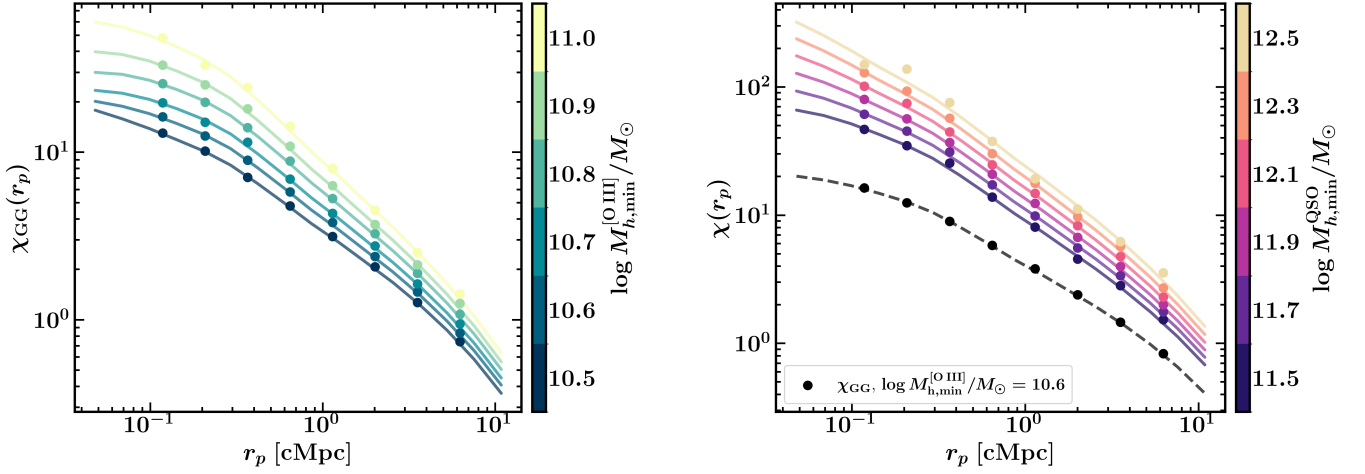
is assigned a redshift and [O III] luminosity drawn from the [O III] luminosity function measured by the EIGER program (Matthee et al. 2023), which covers the same redshift range ( $5.3 < z < 7.0$ ) and WFSS mode as ASPIRE. Sources are then filtered through the coverage map (§3.1) to verify geometric observability and through the sensitivity map (§3.2) to verify detectability above the local flux limit. The resulting effective completeness, defined as the fraction of injected sources that survive both selection cuts, is  $\langle S \rangle \approx 0.54$ . We note that this completeness depends on both the flux limit of the survey and the area of the Monte Carlo injection region relative to the redshift-dependent WFSS spectral coverage. The WFSS coverage varies with redshift because the dispersed location of the [O III] lines shifts across the detector, changing the effective area observed by the two NIRCcam modules. Since the injection region encompasses the full rectangular F356W direct imaging footprint, which is larger than the spectral coverage at any given redshift, a fraction of injected sources fall outside the observable area. The effective completeness restricted to the WFSS-covered area, which is the relevant quantity for clustering measurements, is substantially higher; we refer the reader to Huang et al. (2026) for the explicit definition and value. The Landy–Szalay estimator normalizes the pair counts using  $N_{R_G}$ , the expected number of [O III]-emitters in a blank field after applying the ASPIRE selection function, which is obtained by downscaling the oversampled random catalog by the effective completeness  $\langle S \rangle$  (see Huang et al. 2026 for details).

## 4 COVARIANCE MATRIX BASED ON MOCK REALIZATIONS

### 4.1 Simulation Setup

Previous high-redshift clustering studies commonly estimate uncertainties using Poisson errors on the pair counts. However, because galaxies contribute to multiple pair-count bins and because large-scale density fluctuations (cosmic variance) coherently affect all bins, the true covariance matrix has significant off-diagonal structure that Poisson errors neglect entirely. The small number of independent ASPIRE fields makes the empirical or data-driven error estimators such as jackknife unreliable (Norberg et al. 2009). We therefore construct mock covariance matrices from the FLAMINGO-10k (F10k) simulation, a dark-matter-only (DMO) run from the FLAMINGO suite of cosmological simulations (Schaye et al. 2023; Kugel et al. 2023) in a  $(2.8 \text{ cGpc})^3$  volume (Schaller et al., in prep.; see also Pizzati et al. 2024b). The simulation adopts the “ $3 \times 2\text{pt} + \text{all}$ ” cosmology from Abbott et al. (2022):  $\Omega_m = 0.306$ ,  $\Omega_b = 0.0486$ ,  $\sigma_8 = 0.807$ ,  $H_0 = 68.1 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ,  $n_s = 0.967$ , with a summed neutrino mass of  $0.06 \text{ eV}$ . The simulation contains  $10080^3$  CDM particles and  $5600^3$  neutrino particles, with a CDM particle mass of  $M_{\text{dm}} = 8.40 \times 10^8 M_\odot$ .

Halos are identified using the HBT-HERONS (Hierarchical Bound-Tracing) halo finder (Han et al. 2012, 2018; Forouhar Moreno et al. 2025), which tracks the formation and evolution of subhalos by following their bound particles across cosmic time. We define the subhalo mass as the peak bound mass ( $M_{\text{peak}}$ ), i.e. the maximum mass a subhalo has attained over its history, and adopt the most bound particle as the subhalo position. The HBT-HERONS catalog includes both central and satellite subhalos. We use the simulation snapshot at  $z_{\text{snap}} = 6.14$ , the closest match to the median redshift of [O III]-emitters observed in ASPIRE ( $\langle z_{[\text{O III}]} \rangle = 6.13$ ). We only select subhalos with more than 40 dark matter particles, corresponding to a minimum resolved mass of  $\log(M_{\text{min}}^{\text{res}}/M_\odot) = 10.5$  (Pizzati et al. 2024b). See Pizzati et al. (2024b) for a detailed description of the halo catalog.



**Figure 1.** *Left:* Modeled [O III]-emitter auto-correlation function. The solid curves represent the results from the full simulated catalog, while scatter points show the mean of 1,000 ASPIRE-like realizations. The colors correspond to step-function halo occupation distributions with  $\log M_{h,\min}^{[\text{O III}]} / M_{\odot} = 10.5\text{--}11.1$ . *Right:* QSO–galaxy cross-correlation function with  $\log M_{h,\min}^{\text{QSO}} / M_{\odot} = 11.5\text{--}12.5$  and a fixed  $\log M_{h,\min}^{[\text{O III}]} / M_{\odot} = 10.7$ . The dashed line shows the reference [O III] auto-correlation function  $\chi_{\text{GG}}$  at fixed  $\log M_{h,\min}^{[\text{O III}]} / M_{\odot} = 10.7$  for comparison.

We adopt a step-function halo occupation distribution (Zheng et al. 2005) in which both quasars and [O III]-emitters populate all subhalos (centrals and satellites) above their respective minimum mass thresholds,  $M_{h,\min}^{\text{QSO}}$  and  $M_{h,\min}^{[\text{O III}]}$ . To match the observed abundance, we down-sample the galaxy catalog to reproduce the number density measured in Matthee et al. (2023). Each selected halo is assigned an [O III] luminosity drawn from the EIGER luminosity function at  $z \sim 5.3\text{--}7$  (Matthee et al. 2023), which is necessary because the mock sources are subsequently passed through the sensitivity map (§3.2) to determine detectability. The full mock construction procedure is described in §4.2.

#### 4.2 Construction of Mock [O III]-emitter Catalog

Each ASPIRE pointing covers approximately  $2.2' \times 5.1'$  on the sky, corresponding to a transverse comoving size of  $\approx 3.8 \times 8.8 \text{ cMpc } h^{-1}$  at  $z \approx 6.6$ , with a line-of-sight depth of  $\approx 485 \text{ cMpc } h^{-1}$  over  $z = 5.3\text{--}7.0$ . All three dimensions are much smaller than the  $\approx 1900 \text{ cMpc } h^{-1}$  FLAMINGO-10k box side, so the simulation hosts many independent ASPIRE-like realizations from non-overlapping regions.

For each pair of minimum peak subhalo masses ( $M_{h,\min}^{\text{QSO}}, M_{h,\min}^{[\text{O III}]}$ ), we construct mock ASPIRE-like observations as follows:

(i) Select quasars and [O III]-emitters from the simulation using a step-function HOD: all subhalos (both centrals and satellites) with  $M_{\text{peak}} > M_{h,\min}^{\text{QSO}}$  are assigned as quasar hosts, and all subhalos with  $M_{\text{peak}} > M_{h,\min}^{[\text{O III}]}$  as galaxy hosts. The positions of both quasars and galaxies are taken to be the most bound particle of their host subhalo. We randomly down-sample the galaxy catalog to match the observed [O III]-emitter number density from Matthee et al. (2023); the down-sampling factor corresponds to the [O III]-emitter duty cycle,  $f_{\text{duty}}^{[\text{O III}]} = n_{[\text{O III}]} / n_{M_h > M_{h,\min}^{[\text{O III}]}}$ . This assumes a common mass threshold for centrals and satellites; the impact of relaxing this assumption is explored in §5.5. We adopt  $\log(M_{h,\min}^{\text{QSO}} / M_{\odot}) = 11.0\text{--}12.5$  and  $\log(M_{h,\min}^{[\text{O III}]} / M_{\odot}) = 10.5\text{--}11.0$ , bracketing the values inferred from the EIGER survey (Pizzati et al. 2024b).

(ii) For each realization, randomly select 25 quasars from the catalog. Using the periodic boundary conditions of the simulation box, re-center the galaxy catalog around each selected quasar to avoid edge effects. Assign observed sky coordinates and redshifts by mapping the simulation-frame positions relative to each quasar onto the true ASPIRE quasar locations. Galaxy redshifts include the line-of-sight peculiar velocity from the simulation, encoding redshift-space distortions (see Huang et al. 2026, for the explicit redshift-assignment procedure). Each galaxy is assigned an [O III] luminosity drawn randomly from the EIGER luminosity function at  $z \sim 5.3\text{--}7$  (Matthee et al. 2023), independently of its host halo mass and assuming no redshift evolution over this range.

(iii) Apply the redshift-dependent coverage map (§3.1) and sensitivity map (§3.2) field by field to determine which mock [O III]-emitters would be detected in the ASPIRE survey.

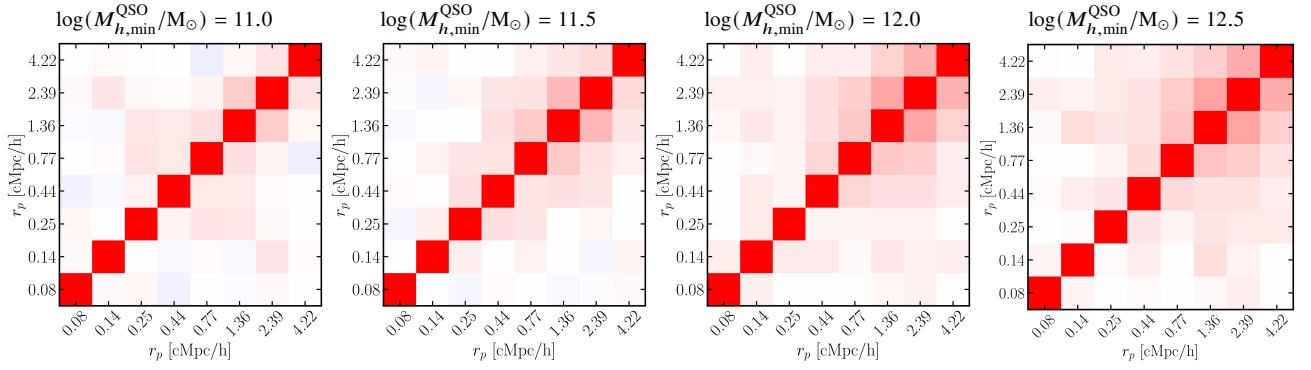
(iv) Compute the [O III]-emitter auto-correlation function  $\chi_{\text{GG}}$  (Eq. 2) and the quasar–[O III]-emitter cross-correlation function  $\chi_{\text{QG}}$  (Eq. 3), based on the total pair counts across all 25 mock ASPIRE fields.

(v) Repeat steps (ii)–(iv) for  $N_{\text{real}} = 1000$  realizations and compute the covariance matrix using Eq. 4.

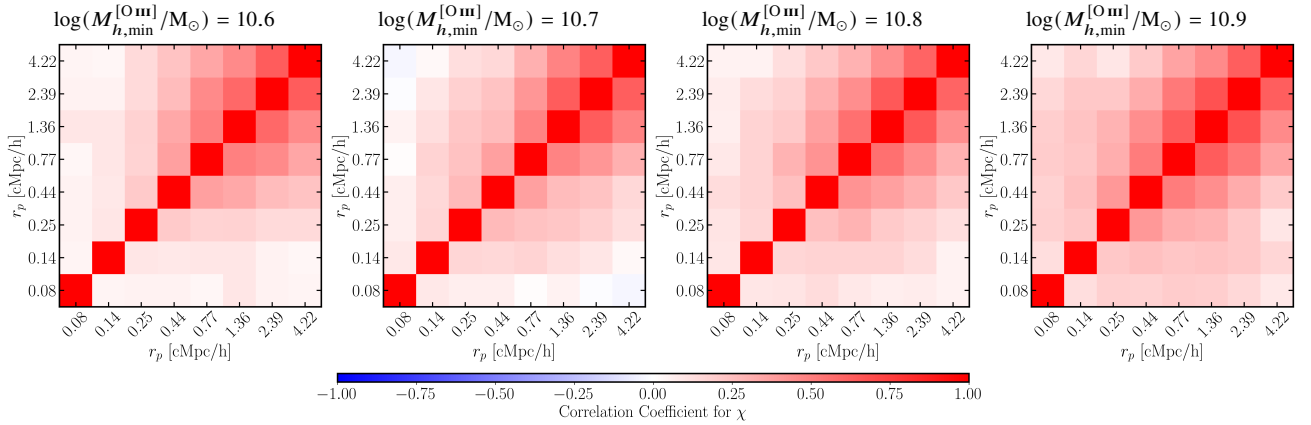
Figure 1 compares the mean of the 1000 mock correlation functions (scatter points) with the full-box halo model predictions (solid curves) for both  $\chi_{\text{GG}}$  and  $\chi_{\text{QG}}$  across a range of minimum halo masses, confirming that the ASPIRE-like mocks faithfully recover the input clustering signal. From these  $N_{\text{real}} = 1000$  realizations, we compute the covariance matrix elements as:

$$C_{ij}(\chi_a) = \frac{1}{N_{\text{real}}} \sum_{k=1}^{N_{\text{real}}} (\chi_{a,i}^k - \langle \chi_{a,i} \rangle) (\chi_{a,j}^k - \langle \chi_{a,j} \rangle), \quad (4)$$

where indices  $i, j = 1, \dots, N_{\text{bin}}$  label radial bins,  $k = 1, \dots, N_{\text{real}}$  labels realizations,  $\chi_{a,i}^k$  is the  $i$ -th bin for the  $k$ -th mock realization of the volume-averaged correlation function,  $a \in \{\text{GG}, \text{QG}\}$ , and  $N_{\text{real}} = 1000$  is the total number of realizations. To visualize the



**Figure 2.** Example correlation structure of the volume-averaged cross correlation function,  $\chi_{QG}$ , computed using Eq. 5 based on 1,000 realizations. From left to right, the minimum quasar host halo mass,  $M_{h,min}^{QSO}/M_{\odot}$ , increases from 11.0 to 12.5, while the minimum galaxy host halo mass is fixed at  $\log(M_{h,min}^{[OIII]})/M_{\odot} = 10.6$  for better comparison. The off-diagonal correlation becomes slightly stronger as  $M_{h,min}^{QSO}/M_{\odot}$  increases, indicating stronger correlations for more massive quasar host halos.



**Figure 3.** Correlation structure of the volume-averaged auto-correlation function,  $\chi_{GG}$ , computed using Eq. 5 based on 1,000 realizations. From left to right, the minimum galaxy host halo mass,  $M_{h,min}^{[OIII]}/M_{\odot}$  increases from 10.6 to 10.9. The off-diagonal correlation becomes slightly stronger as  $M_{h,min}^{[OIII]}/M_{\odot}$  increases.

bin-to-bin correlations, we define the correlation matrix:

$$\rho_{ij}(\chi_a) = \frac{C_{ij}(\chi_a)}{\sqrt{C_{ii}(\chi_a) C_{jj}(\chi_a)}}. \quad (5)$$

Figs. 2 and 3 show the resulting correlation matrices (Eq. 5) for the cross-correlation and auto-correlation, respectively. The cross-correlation correlation matrix (Fig. 2) shows weak off-diagonal terms, since  $\chi_{QG}$  is measured around a single quasar per field and each galaxy contributes to only one separation bin relative to the quasar. The off-diagonal elements increase with  $M_{h,min}^{QSO}$ : for more massive, more strongly biased quasar host halos, large-scale density fluctuations contribute a larger fraction of the total variance relative to shot noise, increasing the correlation between radial bins (see, e.g., Bernstein 1994). The auto-correlation correlation matrix (Fig. 3) exhibits stronger off-diagonal correlations, which increase with  $M_{h,min}^{[OIII]}$ . This arises from two effects: (1) the same galaxy contributes to pair counts at multiple separations, so uncertainties in different bins are inherently coupled; and (2) large-scale density fluctuations coherently boost or suppress galaxy counts across all bins simultaneously.

## 5 STATISTICAL INFERENCE WITH COVARIANCE MATRICES

### 5.1 Likelihood functions

By construction,  $\chi_{QG}$  is computed using only [O III]-emitters within  $\pm 7$  cMpc  $h^{-1}$  of the quasar along the line of sight (i.e., the same  $\pi_{\max}$  window used in the cross-correlation estimator), while  $\chi_{GG}$  uses only galaxies outside that region. The two estimators therefore use disjoint pair sets, and we have verified directly from the 1000 mock realizations that the cross-covariance between  $\chi_{GG}$  and  $\chi_{QG}$  is negligible. We therefore treat the two correlation functions as statistically independent. The total likelihood function is:

$$\mathcal{L}_{\text{tot}}(X | \Theta) = \mathcal{L}_{GG} \cdot \mathcal{L}_{QG}, \quad (6)$$

where  $X$  is a vector of observed correlation functions,  $X = (\chi_{GG}, \chi_{QG})$  and  $\Theta$  is the modeled parameters for the minimum masses of subhalos hosting quasars and galaxies,  $\Theta = (M_{h,min}^{QSO}, M_{h,min}^{[OIII]})$ . The individual likelihoods  $\mathcal{L}_{GG}$  and  $\mathcal{L}_{QG}$  are:

$$\mathcal{L}_{GG} = \mathcal{N}\left(\chi_{GG}^{\text{obs}} \middle| \chi_{GG}^{\text{model}}(M_{h,min}^{[OIII]}), \mathbf{C}_{GG}(M_{h,min}^{[OIII]})\right) \quad (7)$$

and

$$\mathcal{L}_{QG} = \mathcal{N}\left(\chi_{QG}^{\text{obs}} \middle| \chi_{QG}^{\text{model}}(M_{h,min}^{QSO}, M_{h,min}^{[OIII]}), \mathbf{C}_{QG}(M_{h,min}^{QSO}, M_{h,min}^{[OIII]})\right).$$

(8)

Here,  $\mathcal{N}(\chi^{\text{obs}} | \chi^{\text{model}}, \mathbf{C})$  represents a Gaussian distribution with mean  $\chi^{\text{model}}$  and covariance  $\mathbf{C}$ , evaluated at the observed data  $\chi^{\text{obs}}$ . The model prediction  $\chi^{\text{model}}$  is taken from the FLAMINGO-10k clustering fit of Pizzati et al. (2024b), which is consistent with the mean of our 1000 ASPIRE-like mock realizations for  $\log(M_h/M_\odot) \geq 10.5$  (Fig. 1). We use this fit rather than the mock mean directly because the inference grid extends below the FLAMINGO-10k resolution limit ( $\log(M_h/M_\odot) = 10.5$ ); the fit can be extrapolated to lower masses, although the resulting prediction may not be exact in that regime.

For comparison, we also evaluate the likelihood using Poisson pair-count uncertainties:

$$\log \mathcal{L}_a^{\text{Pois}} = -\frac{1}{2} \sum_{i=1}^{N_{\text{bin}}} \frac{(\chi_{a,i}^{\text{obs}} - \chi_{a,i}^{\text{model}})^2}{\sigma_{a,i}^2}, \quad (9)$$

where  $a \in [\text{GG}, \text{QG}]$  and  $\sigma_{a,i}$  is the Poisson error in bin  $i$ , obtained by propagating the  $1\sigma$  Poisson confidence interval on the raw data–data pair counts through the correlation function estimator. For the [O III]-emitter auto-correlation, we propagate the Poisson interval on  $D_G D_G$  through the Landy–Szalay estimator (Eq. 2); for the cross-correlation, we propagate the interval on  $D_Q D_G$  through the Davis–Peebles estimator (Eq. 3). In both cases, the data–random and random–random pair counts are treated as noiseless, since the random catalog is oversampled to  $5 \times 10^6$  sources per field (see Huang et al. 2026 for details on the pair count normalization). When fitting a single mock realization, we use the Poisson error computed from the pair counts of that realization, as one would with real observational data. When the “observed” data vector is instead set to the mean of all realizations, we correspondingly use the mean Poisson error across all 1000 realizations (see detailed discussion in §5.3). This allows us to directly quantify the impact of adopting the full covariance matrix on the inferred parameters.

## 5.2 Minimum Halo Masses

We now apply the framework described above to infer the minimum halo masses ( $M_{h,\text{min}}^{\text{QSO}}, M_{h,\text{min}}^{\text{[O III]}}$ ), using the simulation-based clustering models and mass-dependent covariance matrices constructed in §4. Unlike the power-law correlation length analysis presented in Appendix B, where the covariance must be fixed at a fiducial mass, here both the model correlation function and the covariance matrix depend on the fitted parameters ( $M_{h,\text{min}}^{\text{QSO}}, M_{h,\text{min}}^{\text{[O III]}}$ ).

### 5.2.1 Mass Dependent Covariance Matrices

Since the correlation functions  $\chi_{\text{QG}}$  and  $\chi_{\text{GG}}$  depend on the minimum halo mass threshold, so does their variance in each radial bin. As shown in Fig. 2 and 3, the off-diagonal covariance structure also varies with the minimum halo mass for both  $\chi_{\text{QG}}$  and  $\chi_{\text{GG}}$ . To account for this mass dependence, we compute the covariance matrix on a grid of minimum halo masses: for  $\chi_{\text{GG}}$ , six values of  $\log M_{h,\text{min}}^{\text{[O III]}}/M_\odot \in [10.5, 11.0]$  in 0.1-dex steps; for  $\chi_{\text{QG}}$ , a 2D grid of 30 mass pairs spanning  $\log M_{h,\text{min}}^{\text{[O III]}}/M_\odot \in [10.5, 10.9]$  in 0.1-dex steps and  $\log M_{h,\text{min}}^{\text{QSO}}/M_\odot \in [11.5, 12.5]$  in 0.2-dex steps.

### 5.2.2 Example fit to a single mock realization

To further illustrate the impact of the covariance matrix on parameter inference, we perform Bayesian inference and sample the posterior with MCMC on individual mock realizations. We select one realization

from our mock catalog generated with the fiducial input masses  $\log(M_{h,\text{min}}^{\text{QSO}}/M_\odot) = 12.2$  and  $\log(M_{h,\text{min}}^{\text{[O III]}}/M_\odot) = 10.6$ , and treat its correlation functions  $\chi_{\text{QG}}(r_p)$  and  $\chi_{\text{GG}}(r_p)$  as “observed” data. We then run two independent MCMC inferences using the affine-invariant ensemble sampler emcee (Foreman-Mackey et al. 2013): one using the full mock covariance matrix in the likelihood (Equation 6), and one replacing it with Poisson errors derived from the pair counts of that realization. Each chain uses 36 walkers run for 5000 steps, with the first 1000 steps discarded as burn-in.

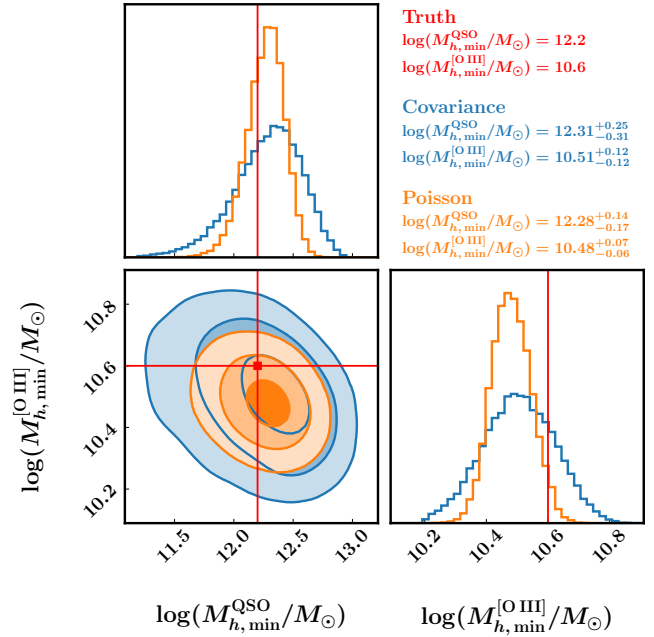
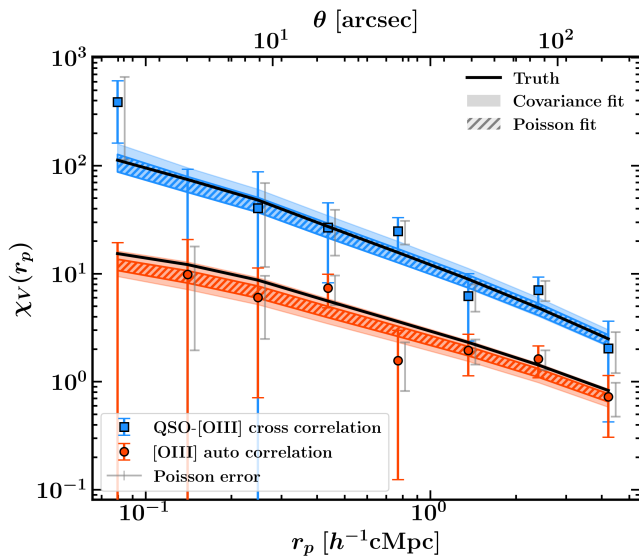
The left panel of Fig. 4 shows the resulting fits. The data points (blue squares for the quasar–[O III] cross-correlation; red circles for the [O III] auto-correlation) are plotted with colored error bars from the diagonal of the full covariance matrix, which represent the realistic  $1\sigma$  uncertainties including cosmic variance. Grey offset error bars show the Poisson errors derived from pair counts for comparison, which are visibly smaller. The black solid curves show the true halo model evaluated at the input masses. The solid shaded bands show the 16th–84th percentile model range from 200 random draws of the covariance-based posterior, while the hatched bands show the same for the Poisson-based posterior.

The Poisson fit bands (hatched) are visibly narrower than the covariance fit bands (solid shaded), directly demonstrating that the Poisson likelihood produces artificially tight constraints. The two fits also yield slightly different median parameters: the Poisson fit gives  $\log(M_{h,\text{min}}^{\text{QSO}}/M_\odot) = 12.28^{+0.14}_{-0.17}$  versus  $12.31^{+0.25}_{-0.31}$  for the covariance fit, and  $\log(M_{h,\text{min}}^{\text{[O III]}}/M_\odot) = 10.48^{+0.07}_{-0.06}$  versus  $10.51^{+0.12}_{-0.12}$ . This shift arises because the two likelihoods weight the radial bins differently.

The right panel of Fig. 4 shows the corresponding MCMC posterior distributions. We note that the posterior contours extend below the FLAMINGO-10k resolution limit of  $\log M_{h,\text{min}}^{\text{[O III]}}/M_\odot = 10.5$ ; in this regime  $\chi^{\text{model}}$  is the extrapolated model from Pizzati et al. (2024b) (see §5.1) and the inferred values should be interpreted with appropriate caution. The covariance-based posterior (blue contours) is approximately 1.8× broader in both  $M_{h,\text{min}}^{\text{QSO}}$  and  $M_{h,\text{min}}^{\text{[O III]}}$  compared to the Poisson posterior (orange contours). For this realization, the true input values (red crosshairs) fall within the  $1\sigma$  contour of the covariance-based posterior but are excluded by the Poisson posterior for  $M_{h,\text{min}}^{\text{[O III]}}$ , illustrating the overconfidence of Poisson error estimates. We quantify this in §5.4: across 1000 mock realizations, the truth lies more than  $1\sigma$  from the Poisson estimate in  $\approx 80\%$  of cases, far above the  $\approx 32\%$  expected for well-calibrated intervals. The single-realization example above demonstrates the effect of the full covariance at one fiducial mass pair. However, since the covariance matrix itself depends on the assumed halo masses, the degree to which Poisson errors underestimate the true uncertainty is expected to vary across the parameter space. In §5.3, we extend this comparison across a grid of minimum halo masses to quantify this mass dependence. A complementary noise-free version of this single-realization fit, using the mean of all 1000 mock realizations as the observed data, is presented in Appendix A.

## 5.3 Impact of covariance on halo mass constraints

To quantify the importance of the full covariance matrix in constraining halo minimum masses, we compare parameter uncertainties derived using the full mock covariance  $\mathbf{C}$  (Eq. 6) against those obtained using the Poisson likelihood (Eq. 9). To isolate the effect of the error model from noise in any particular realization, we set the “observed” data vector to the mean of 1000 mock realizations, i.e., the noiseless true model. This ensures that both likelihoods peak at the same input value by construction, so the difference between the two is purely in



**Figure 4.** *Left:* Correlation function fit for a single mock realization with input masses  $\log(M_{h,\min}^{\text{QSO}}/M_{\odot}) = 12.2$  and  $\log(M_{h,\min}^{[\text{O III}]}/M_{\odot}) = 10.6$ . Blue squares show the quasar–[O III] cross-correlation and red circles show the [O III] auto-correlation. Colored error bars show the  $1\sigma$  uncertainty from the diagonal of the full covariance matrix; grey offset error bars show the Poisson errors from pair counts for comparison. Black solid curves show the true model at the input masses. Solid shaded bands show the 16th–84th percentile model range from 200 posterior draws using the full covariance likelihood; hatched bands show the same for the Poisson likelihood. The top axis shows the corresponding angular scale at  $z = 6.1$ . *Right:* Joint posterior distribution for  $M_{h,\min}^{\text{QSO}}$  and  $M_{h,\min}^{[\text{O III}]}$ . Blue filled contours show the 1, 2, and  $3\sigma$  posterior using the full mock covariance matrix; orange filled contours show the posterior using diagonal Poisson errors. Red crosshairs mark the true input values. The covariance-based posterior is  $\approx 1.8\times$  broader in both parameters; the Poisson posterior excludes the true  $M_{h,\min}^{[\text{O III}]}$  at  $1\sigma$ , illustrating the overconfidence of Poisson error estimates.

the width of the constraints. As described in §5.1, we use the mean Poisson error across all realizations as the representative Poisson uncertainty, consistent with using the noise-free mean as the observed data. Unlike the full covariance, this Poisson error captures only shot noise and ignores both cosmic variance and correlations between  $r_p$  bins.

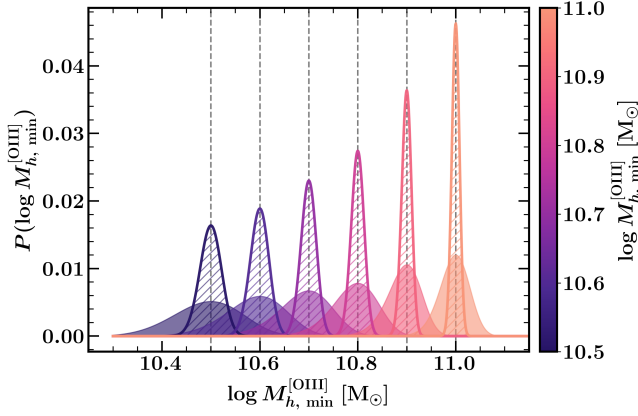
We generate the covariance matrices  $\mathbf{C}_{\text{QG}}$  and  $\mathbf{C}_{\text{GG}}$  from 1000 mock realizations of the ASPIRE survey. For the galaxy auto-correlation, we compute  $\mathbf{C}_{\text{GG}}$  independently at six values of the galaxy halo mass threshold,  $\log(M_{h,\min}^{[\text{O III}]}/M_{\odot}) \in [10.5, 11.0]$  in steps of 0.1 dex. For the cross-correlation, mock covariances are computed on a 2d grid of 30 mass pairs spanning  $M_{h,\min}^{[\text{O III}]} \in [10.5, 10.9]$  in steps of 0.1 dex and  $M_{h,\min}^{\text{QSO}} \in [11.5, 12.5]$  in steps of 0.2 dex ( $5 \times 6 = 30$  mass pairs).

For each  $(M_{h,\min}^{[\text{O III}]}, M_{h,\min}^{\text{QSO}})$  pair, we perform two parallel inferences: one using the full covariance matrix (Equation 6) and one replacing  $\mathbf{C}$  with the diagonal Poisson errors. From the resulting likelihood curves we compute the  $1\sigma$  uncertainty (half the 16th–84th percentile interval of the normalized likelihood) and compute the ratio  $\sigma_{\text{cov}}^{M_{h,\min}^{\text{QSO}}}/\sigma_{\text{Poisson}}^{M_{h,\min}^{\text{QSO}}}$ .

Fig. 5 shows the normalized likelihood  $P(M_{h,\min}^{[\text{O III}]}) \propto \mathcal{L}_{\text{GG}}(M_{h,\min}^{[\text{O III}]})$  from the GG auto-correlation alone, evaluated on a grid of  $M_{h,\min}^{[\text{O III}]}$  values for six input mass thresholds (dashed vertical lines). Because  $\chi_{\text{GG}}$  depends only on  $M_{h,\min}^{[\text{O III}]}$  and is independent of  $M_{h,\min}^{\text{QSO}}$ , no marginalization over the quasar halo mass is required. As described at the beginning of §5.3, the “observed” data vector is set to the mean mock  $\chi_{\text{GG}}$  at each true input mass, so both likelihoods peak at the same value by construction. For each input value, the darker filled curve shows the likelihood obtained using the full covariance

$\mathbf{C}_{\text{GG}}$ , while the lighter filled curve of the same color uses Poisson errors. The difference between the two is therefore entirely due to the error budget: the covariance-based constraints are systematically and significantly broader than their Poisson counterparts, reflecting the combined effect of cosmic variance and off-diagonal bin-to-bin correlations that Poisson errors neglect entirely.

Fig. 6 presents the ratio  $\sigma_{\text{cov}}^{M_{h,\min}^{\text{QSO}}}/\sigma_{\text{Poisson}}^{M_{h,\min}^{\text{QSO}}}$  for the marginalized  $M_{h,\min}^{\text{QSO}}$  constraint from the full joint QG+GG fit across the grid of 30 mass pairs. The general trend is clear: the Poisson approximation underestimates the true error more severely at higher halo masses, for both  $M_{h,\min}^{\text{QSO}}$  and  $M_{h,\min}^{[\text{O III}]}$ . At low masses ( $M_{h,\min} = 11.5$ ,  $M_{h,\min} = 10.5$ ), the ratio is modest ( $\sigma_{\text{cov}}^{M_{h,\min}^{\text{QSO}}}/\sigma_{\text{Poisson}}^{M_{h,\min}^{\text{QSO}}} \approx 1.4\text{--}1.5$ ), while at high masses ( $M_{h,\min}^{\text{QSO}} \geq 12.1$ ,  $M_{h,\min}^{[\text{O III}]} = 10.9$ ) it reaches  $\approx 2.4\text{--}2.7$ . However, at the highest quasar halo masses ( $\log M_{h,\min}^{\text{QSO}}/M_{\odot} \geq 12.3$ , top two rows of Fig. 6), the ratio does not grow monotonically. This is driven by two boundary effects of the precomputed grids: the mock covariance was generated only on  $\log M_{h,\min}^{\text{QSO}}/M_{\odot} \in [11.5, 12.5]$ , so inferences at higher input masses fall back to the  $M_{h,\min}^{\text{QSO}} = 12.5$  covariance rather than the true mass-dependent covariance; and the model  $\chi(r_p)$  grid extends only to  $\log M_{h,\min}^{\text{QSO}}/M_{\odot} = 13.0$ , truncating the posterior at this upper boundary. Because the covariance posterior is broader than the Poisson posterior, it is clipped more aggressively by this upper-boundary truncation, so  $\sigma_{\text{cov}}^{M_{h,\min}^{\text{QSO}}}$  shrinks faster than  $\sigma_{\text{Poisson}}^{M_{h,\min}^{\text{QSO}}}$  as  $M_{h,\min}^{\text{QSO}}$  approaches the grid edge. The ratio therefore stops growing — and in some cells decreases — at the top of the heatmap,



**Figure 5.** Normalized likelihood  $P(M_{h,\min}^{[\text{OIII}]}) \propto \mathcal{L}_{\text{GG}}(M_{h,\min}^{[\text{OIII}]})$  evaluated on a grid of  $M_{h,\min}^{[\text{OIII}]}$  values, using only the GG auto-correlation function  $\chi_{\text{GG}}(r_p)$ . Because  $\chi_{\text{GG}}$  depends only on  $M_{h,\min}^{[\text{OIII}]}$  and is independent of  $M_{h,\min}^{\text{QSO}}$ , no marginalization is required. For each of six input mass thresholds (dashed vertical lines), the “observed” data vector is set equal to the mean mock  $\chi_{\text{GG}}$  at the true  $M_{h,\min}^{[\text{OIII}]}$ , so that both likelihoods peak at the correct input value by construction. The darker filled curve evaluates  $\mathcal{L}_{\text{GG}}$  using the full mock covariance  $\mathbf{C}_{\text{GG}}$ , and the lighter filled curve uses diagonal Poisson errors. The difference between the two is therefore purely in the error budget: the covariance-based constraints are systematically broader, reflecting cosmic variance and off-diagonal bin-to-bin correlations that Poisson errors neglect.

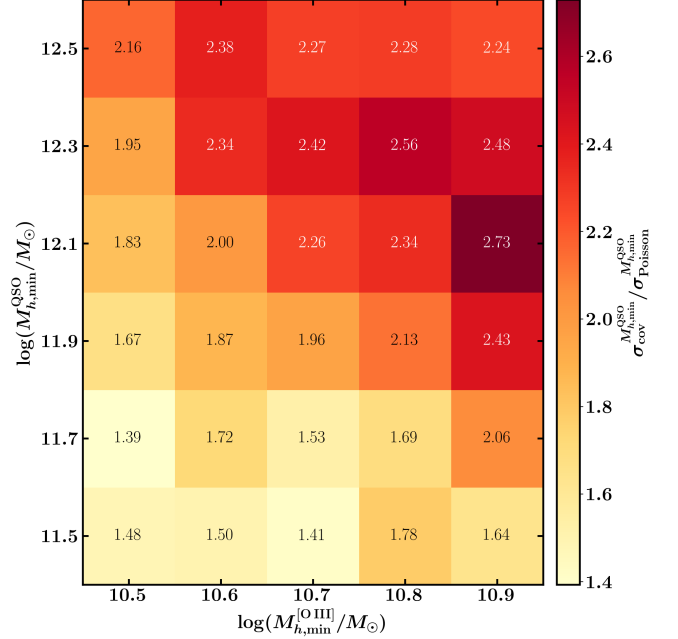
instead of continuing to increase with  $M_{h,\min}^{\text{QSO}}$  as it does at lower masses.

This mass-dependent trend can be understood from the off-diagonal elements of the correlation matrix becoming more prominent at higher halo masses. As illustrated in Figs. 2 and 3, the positive off-diagonal correlations arise because large-scale density fluctuations boost galaxy pair counts at all separations simultaneously: an overdense region increases the number of pairs in every  $r_p$  bin, causing the correlation function values to rise and fall together across bins. For more massive (i.e., more highly biased) halos, these correlated fluctuations are amplified because the clustering response to the underlying density field is stronger (Valageas et al. 2011; Mo & White 1996). This is consistent with the trend seen in Fig. 6, where the ratio increases with both  $M_{h,\min}^{\text{QSO}}$  and  $M_{h,\min}^{[\text{OIII}]}$ .

#### 5.4 Inference test

The mass-dependent comparison in Figs. 5 and 6 shows that Poisson errors underestimate the width of the halo mass constraints by a factor of  $\approx 1.5$ – $2.7$ . We now verify that the narrower Poisson credible intervals fail to contain the true parameters at the expected rate, using a coverage test (see appendix A of Hennawi et al. 2025), which checks whether the resulting credible intervals contain the true parameters at the expected rate. We perform a coverage test using 1000 mock realizations generated at the fiducial input masses  $(\log M_{h,\min}^{[\text{OIII}]} / M_\odot, \log M_{h,\min}^{\text{QSO}} / M_\odot) = (10.6, 12.2)$ . For each realization, we run two independent MCMC inferences, one with the full mock covariance matrix and one with diagonal Poisson errors, and record the log-posterior probability evaluated at the MCMC samples,  $\ln P(\theta_i)$ , and at the true input parameters,  $\ln P(\theta_{\text{true}})$ .

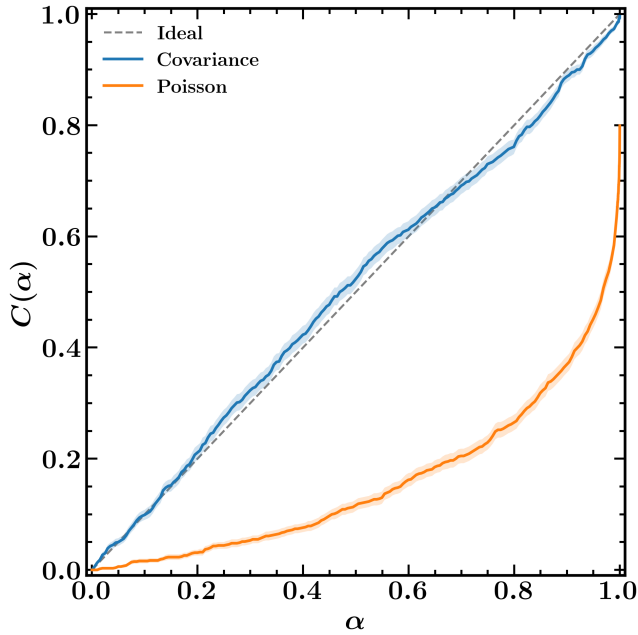
For a given credible level  $\alpha$ , we define the  $\alpha$ -contour as the iso-probability surface enclosing fraction  $\alpha$  of the posterior mass. We then measure the coverage  $C(\alpha)$ : the fraction of mock realizations for which the true parameters fall within the  $\alpha$ -contour. A posterior



**Figure 6.** Ratio  $\sigma_{\text{cov}}^{M_{h,\min}^{\text{QSO}}} / \sigma_{\text{Poisson}}^{M_{h,\min}^{\text{QSO}}}$  for the marginalized  $M_{h,\min}^{\text{QSO}}$  constraint from the joint QG+GG fit, plotted across a grid of minimum halo masses for quasars ( $M_{h,\min}^{\text{QSO}}$ , y-axis) and galaxies ( $M_{h,\min}^{[\text{OIII}]}$ , x-axis). The ratio increases with both masses, reaching  $\approx 2.4$ – $2.7$  at the highest masses, reflecting the stronger cosmic variance for more biased tracers. At the highest quasar halo masses ( $\log M_{h,\min}^{\text{QSO}} / M_\odot \geq 12.3$ , top two rows), the ratio does not grow monotonically. This is because the mock covariance was generated only on  $\log M_{h,\min}^{\text{QSO}} / M_\odot \in [11.5, 12.5]$ , so inferences at higher input masses fall back to the  $M_{h,\min}^{\text{QSO}} = 12.5$  covariance; and the model  $\chi(r_p)$  grid extends only to  $\log M_{h,\min}^{\text{QSO}} / M_\odot = 13.0$ , truncating the posterior at this upper boundary. Because the covariance posterior is broader than the Poisson posterior, it is clipped more by this upper-boundary truncation, so  $\sigma_{\text{cov}}^{M_{h,\min}^{\text{QSO}}}$  shrinks faster than  $\sigma_{\text{Poisson}}^{M_{h,\min}^{\text{QSO}}}$ . The ratio therefore stops growing — and in some cells decreases — at the top of the grid, instead of continuing to increase with  $M_{h,\min}^{\text{QSO}}$  as it does at lower masses.

with perfect coverage satisfies  $C(\alpha) = \alpha$ , meaning that the  $\alpha$ -credible region contains the true parameters in exactly fraction  $\alpha$  of realizations. If  $C(\alpha) < \alpha$ , the credible regions are too narrow (overconfident); if  $C(\alpha) > \alpha$ , they are too wide (underconfident).

Fig. 7 shows the resulting coverage curves  $C(\alpha)$  for both likelihoods, with  $1\sigma$  binomial confidence bands from the 1000 realizations. The covariance-based posterior (blue) closely follows the ideal diagonal  $C(\alpha) = \alpha$ , confirming that the full mock covariance matrix, which encodes cosmic variance, inter-bin correlations, and the non-independent nature of galaxy pairs, produces credible intervals with the correct coverage. This also validates the Gaussian likelihood assumption: despite the discrete nature of pair counts at small separations, the Gaussian likelihood combined with the full covariance matrix produces well-calibrated credible intervals. In contrast, the Poisson posterior (orange) falls dramatically below the diagonal: at  $\alpha = 0.68$  (the nominal  $1\sigma$  level), the true parameters are contained in only  $\approx 20\%$  of the Poisson credible regions. This is a direct consequence of the Poisson approximation ignoring cosmic variance and treating all radial bins as statistically independent, which artificially shrinks the posterior and leads to severe overconfidence in the inferred halo masses.



**Figure 7.** Coverage test comparing the covariance-based and Poisson likelihoods across 1000 mock realizations at the fiducial input masses ( $\log M_{h,\min}^{[\text{O III}]} / M_{\odot}, \log M_{h,\min}^{\text{QSO}} / M_{\odot}$ ) = (10.6, 12.2). The coverage  $C(\alpha)$  gives the fraction of realizations for which the true parameters fall within the  $\alpha$ -credible region of the posterior. Shaded bands show  $1\sigma$  binomial uncertainties. The covariance likelihood (blue) follows the ideal diagonal, confirming that the full mock covariance correctly captures the scatter across realizations. The Poisson likelihood (orange) is severely overconfident: at  $\alpha = 0.68$ , the true parameters are contained in only  $\approx 20\%$  of realizations instead of the expected 68%, because the Poisson approximation ignores cosmic variance and inter-bin correlations.

We note that a fully rigorous inference test (Sellentin & Heavens 2016; Talts et al. 2020; Hennawi et al. 2025) would draw the true parameters from the prior for each mock realization, requiring a separate set of mock catalogs for every  $(M_{h,\min}^{[\text{O III}]}, M_{h,\min}^{\text{QSO}})$  draw. Since our 1000 realizations share a single set of input masses, our test measures the *conditional* coverage at one point in parameter space. Nevertheless, this is sufficient to demonstrate that the Poisson likelihood systematically underestimates the true uncertainty, whereas the covariance-based likelihood does not.

## 5.5 Effect of Separate Central and Satellite Mass Thresholds

The minimum halo mass inference framework presented in §§5.1–5.4 adopts a single galaxy mass threshold  $M_{h,\min}^{[\text{O III}]}$  shared by both central and satellite [O III]-emitters. We now relax this assumption and investigate how the cross- and auto-correlation functions, and the inference of the underlying halo masses, change when centrals and satellites are allowed to populate halos above different mass thresholds.

### 5.5.1 Physical motivation

The single-threshold HOD assumes that the probability of a galaxy having a given luminosity depends only on its (sub)halo peak mass  $M_{\text{peak}}$ , regardless of whether the galaxy is a central or a satellite. In practice, however, the luminosity distribution  $L(M_{\text{peak}})$  is unlikely to be the same for both populations. Several environmental processes

systematically alter the properties of satellite galaxies after infall, including strangulation of the hot gas supply (Larson et al. 1980; Balogh et al. 2000), ram pressure stripping of the cold gas reservoir (Gunn & Gott 1972; Abadi et al. 1999), and tidal stripping and harassment from the host potential and neighboring subhalos (Moore et al. 1996; Read et al. 2006). These effects could either suppress or, in some cases, temporarily boost (e.g., via ram-pressure-induced starbursts) the [O III] luminosity of satellites relative to centrals of the same  $M_{\text{peak}}$ .

Since our galaxy sample is selected via an [O III] flux limit, any systematic difference in  $L(M_{\text{peak}})$  between centrals and satellites translates directly into different effective mass thresholds for the two populations, which in turn changes the satellite fraction of the observed sample. For instance, if satellite luminosities are systematically suppressed at fixed  $M_{\text{peak}}$ , the flux-limited sample preferentially selects satellites from more massive subhalos, yielding an effective  $M_{\text{min,sat}} > M_{\text{min,cen}}$  and a smaller satellite fraction. Conversely, if environmental interactions temporarily boost satellite [O III] emission, the effective satellite threshold could be lower than the central threshold and the satellite fraction correspondingly larger.

For QSOs, the integrated black hole mass is relatively robust to the central–satellite distinction, since much of the black hole growth likely occurred while the galaxy was still a central. The instantaneous accretion rate that powers an observed QSO phase is more vulnerable: gas stripping in satellites depletes the cold-gas reservoir that fuels the central black hole, so a single  $M_{h,\min}^{\text{QSO}}$  would over-count QSOs from the satellite population. We therefore restrict QSOs to centrals throughout this section (§5.5.2); a fully general treatment would also split  $M_{h,\min}^{\text{QSO}}$ . For emission-line luminosities, which respond to instantaneous star formation rates, the central–satellite distinction is similarly consequential, motivating the split  $M_{h,\min}^{\text{cen}, [\text{O III}]} / M_{h,\min}^{\text{sat}, [\text{O III}]}$  treatment that follows.

### 5.5.2 Model setup and mock generation

We generalize the step-function HOD adopted throughout this work by introducing separate halo-mass thresholds for central and satellite [O III]-emitters,  $M_{h,\min}^{\text{cen}, [\text{O III}]}$  and  $M_{h,\min}^{\text{sat}, [\text{O III}]}$ , while retaining a single threshold  $M_{h,\min}^{\text{QSO}}$  for QSOs. Each subhalo is classified using its rank in the halo catalog: rank = 0 identifies a central, while rank > 0 identifies a satellite. A halo or subhalo with  $M_{\text{peak}}$  above its corresponding threshold hosts exactly one tracer, with no scatter. QSOs are assumed to populate only central subhalos (rank = 0), with their position taken as the most bound particle of the host. To remove numerically unresolved objects we additionally require that every subhalo contains at least 20 bound dark-matter particles ( $N_{\text{bound}} \geq 20$ ), consistent with the resolution cut described in §4.

The remainder of the mock construction follows the procedure in §4.2. We use the same FLAMINGO-10k snapshot at  $z_{\text{snap}} = 6.14$ , the same 25 ASPIRE pointings with their selection functions, the same line-of-sight integration scale  $\pi_{\text{max}} = 7 \text{ cMpc } h^{-1}$ , and the same set of  $r_p$  bins. The only change is in step (i) in §4.2: instead of a single galaxy mass cut, we now select centrals from subhalos with rank = 0 and  $M_{\text{peak}} > M_{h,\min}^{\text{cen}, [\text{O III}]} M_{\odot}$ , and satellites from subhalos with rank > 0 and  $M_{\text{peak}} > M_{h,\min}^{\text{sat}, [\text{O III}]} M_{\odot}$ . Selection function and down-sampling to the Matthee et al. (2023) number density are then applied as in §4.2.

For the inference described in §5.5.3 we precompute  $\chi_{\text{GG}}(r_p)$  and  $\chi_{\text{QG}}(r_p)$  on a grid of  $(M_{h,\min}^{\text{QSO}}, M_{h,\min}^{\text{cen}, [\text{O III}]}, M_{h,\min}^{\text{sat}, [\text{O III}]})$ . The grid spans  $\log M_{h,\min}^{\text{QSO}} / M_{\odot} \in [11.25, 12.75]$  in steps of 0.25 dex,

$\log M_{h,\min}^{\text{cen}, [\text{O III}]} / M_\odot \in [10.5, 11.5]$  in steps of 0.1 dex, and  $\log M_{h,\min}^{\text{sat}, [\text{O III}]} / M_\odot \in [10.5, 12.5]$  in steps of 0.1 dex. Likelihood evaluations during MCMC use linear interpolation on this grid.

Figs. 8 and 9 illustrate how the two galaxy thresholds shape the volume-averaged auto- and cross-correlation functions. Each figure shows the [O III] auto-correlation  $\chi_{\text{GG}}$  on the left and the QSO–[O III] cross-correlation  $\chi_{\text{QG}}$  on the right, with the swept threshold encoded by the colormap; in each case we also decompose the signal into 1-halo and 2-halo contributions. We note that the 1-halo terms (in Figs. 8 and 9) remain finite at small  $r_p$  rather than dropping to zero as halo exclusion would imply for the 3D correlation function  $\xi(r)$ ; this is a projection effect of the volume-averaged  $\chi(r_p)$  definition (Eq. 1), which integrates over the cylindrical shell with  $|\pi| \leq \pi_{\text{max}}$  and therefore picks up contributions from pairs with larger 3D separations even at small  $r_p$ .

**Varying the central mass threshold.** When  $M_{h,\min}^{\text{cen}, [\text{O III}]}$  is varied with  $M_{h,\min}^{\text{sat}, [\text{O III}]}$  held fixed (Fig. 8):

(i) The one-halo term (single black dashed line) is essentially unchanged across all models in both  $\chi_{\text{GG}}$  and  $\chi_{\text{QG}}$ , since the satellite population in galaxy host halos and in quasar host halos depends only on  $M_{h,\min}^{\text{sat}, [\text{O III}]}$ , which is held fixed.

(ii) The two-halo term (colored dotted lines) increases with  $M_{h,\min}^{\text{cen}, [\text{O III}]}$  in both  $\chi_{\text{GG}}$  and  $\chi_{\text{QG}}$ , as expected: selecting centrals from more massive halos increases the mean galaxy bias.

(iii) The total auto- and cross-correlations (solid colored lines) increase with  $M_{h,\min}^{\text{cen}, [\text{O III}]}$  beyond the two-halo contribution alone. At fixed total galaxy number density, increasing  $M_{h,\min}^{\text{cen}, [\text{O III}]}$  reduces the number of centrals while the satellite count remains fixed, raising the satellite fraction from  $f_{\text{sat}} \approx 5\%$  at  $\log M_{h,\min}^{\text{cen}, [\text{O III}]} / M_\odot = 10.5$  to  $\approx 87\%$  at  $\log M_{h,\min}^{\text{cen}, [\text{O III}]} / M_\odot = 11.5$ . The higher satellite fraction increases the relative weight of the one-halo term in the total signal of both  $\chi_{\text{GG}}$  and  $\chi_{\text{QG}}$ .

**Varying the satellite mass threshold.** When  $M_{h,\min}^{\text{sat}, [\text{O III}]}$  is varied with  $M_{h,\min}^{\text{cen}, [\text{O III}]}$  held fixed (Fig. 9):

(i) The two-halo term is insensitive to  $M_{h,\min}^{\text{sat}, [\text{O III}]}$  in both the auto- and cross-correlation. At  $r_p \gtrsim 1 \text{ cMpc } h^{-1}$ , all colored curves converge to the single black dashed 2-halo line in both  $\chi_{\text{GG}}$  (left) and  $\chi_{\text{QG}}$  (right). The two-halo clustering amplitude depends on the total galaxy bias, which is dominated by the more numerous centrals; since the central population is identical across all models, the large-scale signal is nearly unchanged.

(ii) The one-halo term depends strongly on  $M_{h,\min}^{\text{sat}, [\text{O III}]}$  in both  $\chi_{\text{GG}}$  and  $\chi_{\text{QG}}$ . At  $r_p \lesssim 0.5 \text{ cMpc } h^{-1}$ , the one-halo signal arises from satellites sharing a host halo with another tracer: another galaxy in the case of  $\chi_{\text{GG}}$ , and a quasar in the case of  $\chi_{\text{QG}}$ . Lowering  $M_{h,\min}^{\text{sat}, [\text{O III}]}$  adds more satellites to these halos and boosts the one-halo pair count. The effect is more pronounced in  $\chi_{\text{QG}}$  because quasar halos are massive ( $M_h \geq 10^{12.2} M_\odot$ ) and host many satellites: the satellite fraction within quasar halos grows faster than the global mean galaxy density, so  $\chi_{\text{QG}}^{\text{1h}}$  has a larger dynamic range in  $M_{h,\min}^{\text{sat}, [\text{O III}]}$  than  $\chi_{\text{GG}}^{\text{1h}}$ .

The complementary roles of  $M_{h,\min}^{\text{cen}, [\text{O III}]}$  (large-scale, 2-halo) and  $M_{h,\min}^{\text{sat}, [\text{O III}]}$  (small-scale, 1-halo) are the physical reason the two thresholds are independently constrainable when both auto- and cross-correlation functions are fit jointly.

### 5.5.3 Inference with the full mock covariance matrix

The likelihood structure follows §5.1 (Eqs. 6–7), with the model means  $\chi_{\text{GG}}^{\text{model}}$  and  $\chi_{\text{QG}}^{\text{model}}$  evaluated as functions of the three free parameters ( $M_{h,\min}^{\text{QSO}}, M_{h,\min}^{\text{cen}, [\text{O III}]}, M_{h,\min}^{\text{sat}, [\text{O III}]}$ ) via linear interpolation on the precomputed grid above. In principle the covariance matrices should also be made mass-dependent, following the construction in §4 where  $C_{\text{GG}}(M_{h,\min}^{\text{[O III]}})$  and  $C_{\text{QG}}(M_{h,\min}^{\text{QSO}}, M_{h,\min}^{\text{[O III]}})$  are computed on the two-parameter grid, with the natural extension here being  $C_{\text{GG}}(M_{h,\min}^{\text{cen}, [\text{O III}]}, M_{h,\min}^{\text{sat}, [\text{O III}]})$  and  $C_{\text{QG}}(M_{h,\min}^{\text{QSO}}, M_{h,\min}^{\text{cen}, [\text{O III}]}, M_{h,\min}^{\text{sat}, [\text{O III}]})$  on the three-parameter grid. However, since our goal in this section is to demonstrate the effect of separating the central and satellite thresholds rather than to fully propagate their impact on the error budget, we therefore adopt a single covariance matrix for each of  $\chi_{\text{GG}}$  and  $\chi_{\text{QG}}$ , computed once at the fiducial truth  $\log[(M_{h,\min}^{\text{QSO}}, M_{h,\min}^{\text{cen}, [\text{O III}]}, M_{h,\min}^{\text{sat}, [\text{O III}]}) / M_\odot] = (12.20, 10.70, 10.60)$ , and hold it fixed during MCMC sampling. The Poisson comparison case follows §5.1 exactly: the per-bin Poisson confidence interval is computed from the raw  $D_{\text{QD}}D_{\text{G}}$  and  $D_{\text{G}}D_{\text{G}}$  pair counts of the realization being fit, and propagated through the Landy–Szalay and Davis–Peebles estimators (Eqs. 2–3) to give per-bin errors on  $\chi_{\text{GG}}$  and  $\chi_{\text{QG}}$ .

The ( $M_{h,\min}^{\text{cen}, [\text{O III}]}, M_{h,\min}^{\text{sat}, [\text{O III}]}$ ) samples are mapped to derived satellite and central fractions  $f_{\text{sat}}$  and  $f_{\text{cen}} = 1 - f_{\text{sat}}$ , computed by counting the centrals (rank = 0) above  $M_{h,\min}^{\text{cen}, [\text{O III}]}$  and the satellites (rank > 0) above  $M_{h,\min}^{\text{sat}, [\text{O III}]}$  in the FLAMINGO-10k subhalo catalog with the same  $N_{\text{bound}} \geq 20$  resolution cut.

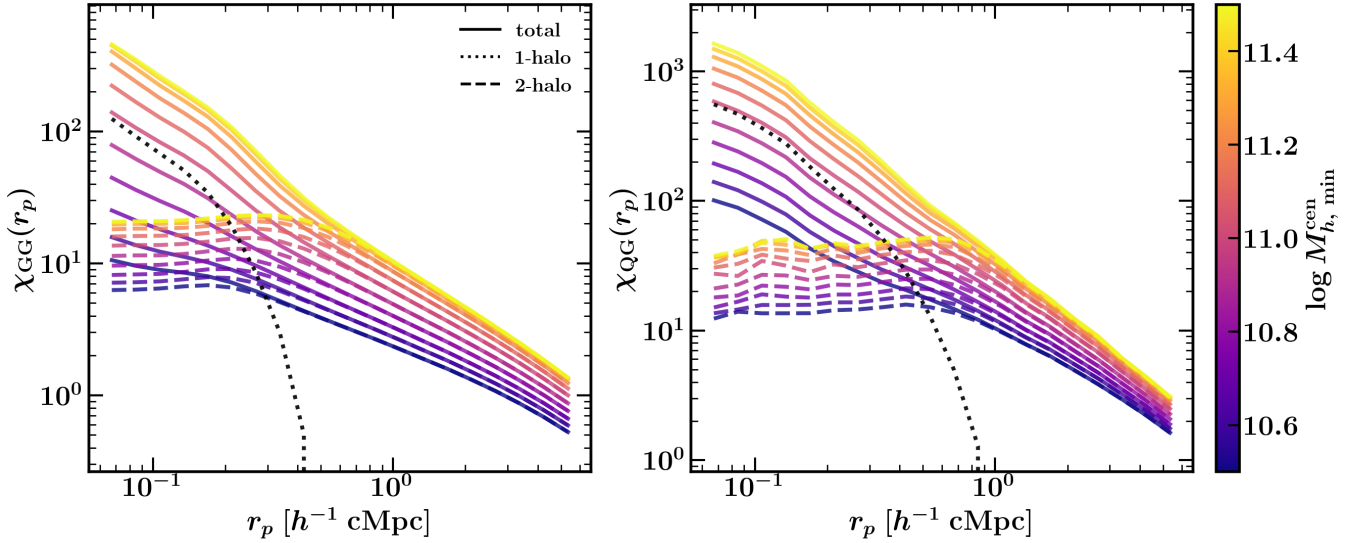
Fig. 10 shows the inference result for one random realization, with the covariance-based posterior recovering the truth model within the bands. Fig. 11 shows the corresponding 5-parameter posterior; the covariance contours enclose the truth at the  $2\sigma$  level in all three free parameters.

Unlike the mass thresholds,  $f_{\text{cen}}$  and  $f_{\text{sat}}$  are derived rather than free parameters, and a flat prior in  $\log M_{h,\min}^{\text{cen}, [\text{O III}]}$  and  $\log M_{h,\min}^{\text{sat}, [\text{O III}]}$  does *not* translate to a flat prior on  $f_{\text{cen}}$  and  $f_{\text{sat}}$ . The mapping ( $M_{h,\min}^{\text{cen}, [\text{O III}]}, M_{h,\min}^{\text{sat}, [\text{O III}]}$ )  $\rightarrow$  ( $f_{\text{cen}}, f_{\text{sat}}$ ) has a non-trivial Jacobian: because the satellite halo-mass function is steep,  $n_{\text{sat}}$  falls rapidly with  $M_{h,\min}^{\text{sat}, [\text{O III}]}$ , so that for  $\log M_{h,\min}^{\text{sat}, [\text{O III}]} / M_\odot \gtrsim 11$  it becomes negligible compared to  $n_{\text{cen}}$  and drives  $f_{\text{sat}} \rightarrow 0$ ,  $f_{\text{cen}} \rightarrow 1$ . Most of the flat prior on  $M_{h,\min}^{\text{sat}, [\text{O III}]} \in U(10.5, 12.5)$  lies in this regime. The induced prior on  $f_{\text{cen}}$  is therefore strongly skewed toward 1 (and the prior on  $f_{\text{sat}}$  toward 0), as shown by the dashed black curves on the  $f_{\text{cen}}$  and  $f_{\text{sat}}$  panels of Fig. 11. We keep the priors flat in log-mass and accept this Jacobian, because the experiment probes mass thresholds rather than fractions; flattening the  $f_{\text{cen}}$  prior would impose a non-trivial prior on the masses we are trying to infer. The Jacobian also explains why the recovered posteriors on  $f_{\text{cen}}$  sit close to 1: the fiducial truth  $\log M_{h,\min}^{\text{sat}, [\text{O III}]} / M_\odot = 10.60$  lies near the lower edge of the  $M_{h,\min}^{\text{sat}, [\text{O III}]}$  prior, in a regime where the prior already pulls  $f_{\text{cen}}$  toward 1.

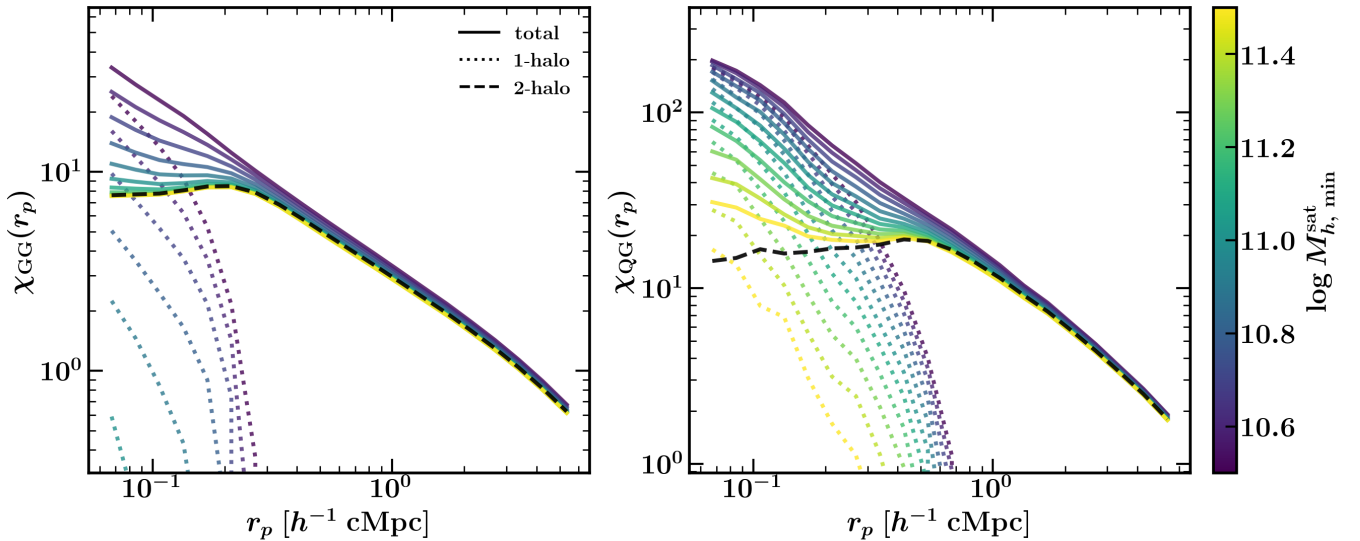
The marginalized posterior on  $M_{h,\min}^{\text{QSO}}$  is essentially unchanged between the single-threshold analysis of §5.2 and the analysis with separate central and satellite thresholds here, indicating that the QSO halo-mass inference is robust to the central–satellite separation in the galaxy sample.

### 5.5.4 Implications for the single-threshold HOD

These results demonstrate that  $M_{h,\min}^{\text{cen}, [\text{O III}]}$  and  $M_{h,\min}^{\text{sat}, [\text{O III}]}$  affect the volume-averaged correlation functions in distinct ways. Varying



**Figure 8.** Sensitivity of the volume-averaged correlation functions to the minimum mass of the central host halos for [O III]-emitters  $\log M_{h,\min}^{\text{cen}, [\text{O III}]} / M_{\odot}$ , at fixed  $\log M_{h,\min}^{\text{QSO}} / M_{\odot} = 12.20$  and  $\log M_{h,\min}^{\text{sat}, [\text{O III}]} / M_{\odot} = 10.60$ . **Left:** [O III] auto-correlation  $\chi_{\text{GG}}(r_p)$ . **Right:** quasar–[O III] cross-correlation  $\chi_{\text{QG}}(r_p)$ . In both panels, solid colored curves show the total correlation function, dotted colored curves show the 2-halo term, and the single black dashed curve shows the 1-halo term, which is essentially independent of  $M_{h,\min}^{\text{cen}, [\text{O III}]}$  at fixed  $M_{h,\min}^{\text{sat}, [\text{O III}]}$ . Increasing  $M_{h,\min}^{\text{cen}, [\text{O III}]}$  selects centrals from more massive halos, raising the mean galaxy bias and boosting the 2-halo and total amplitudes; the 1-halo curves remain unchanged because they are set by  $M_{h,\min}^{\text{sat}, [\text{O III}]}$ , which is held fixed.



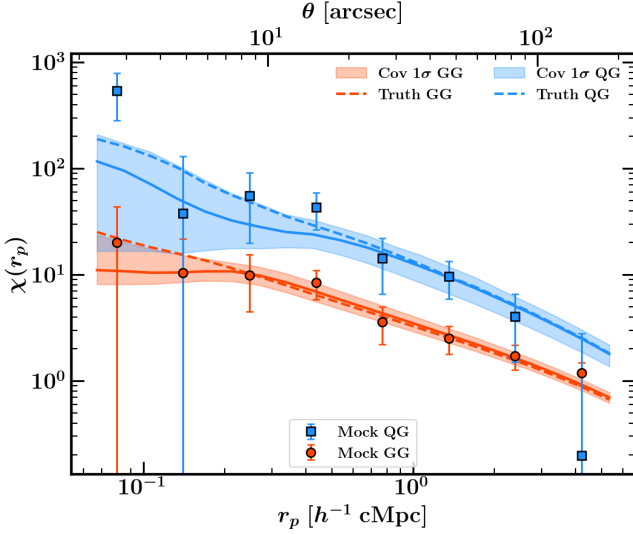
**Figure 9.** Same as Fig. 8, but varying the satellite halo-mass threshold  $\log M_{h,\min}^{\text{sat}, [\text{O III}]} / M_{\odot}$  at fixed  $\log M_{h,\min}^{\text{QSO}} / M_{\odot} = 12.20$  and  $\log M_{h,\min}^{\text{cen}, [\text{O III}]} / M_{\odot} = 10.70$ . **Left:**  $\chi_{\text{GG}}(r_p)$ . **Right:**  $\chi_{\text{QG}}(r_p)$ . In both panels, solid colored curves show the total correlation function, dotted colored curves show the 1-halo term, and the single black dashed curve shows the 2-halo term, which is set by the (fixed) central population. The 1-halo curves drop steeply at small  $r_p$  as  $M_{h,\min}^{\text{sat}, [\text{O III}]}$  increases and strips satellites out of QSO and galaxy host halos, while the large- $r_p$  2-halo signal is essentially unchanged. Comparing to Fig. 8, the small-scale 1-halo regime carries essentially all of the information about  $M_{h,\min}^{\text{sat}, [\text{O III}]}$  in this model, while the large-scale 2-halo amplitude is controlled by  $M_{h,\min}^{\text{cen}, [\text{O III}]}$ .

$M_{h,\min}^{\text{cen}, [\text{O III}]}$  changes the mean galaxy bias, which controls the two-halo amplitude at large separations, and at fixed total galaxy number density also increases the satellite fraction, boosting the relative weight of the one-halo term. Varying  $M_{h,\min}^{\text{sat}, [\text{O III}]}$  primarily affects the one-halo term at small projected separations ( $r_p \lesssim 0.5 \text{ cMpc } h^{-1}$ ), while leaving the two-halo regime essentially unchanged.

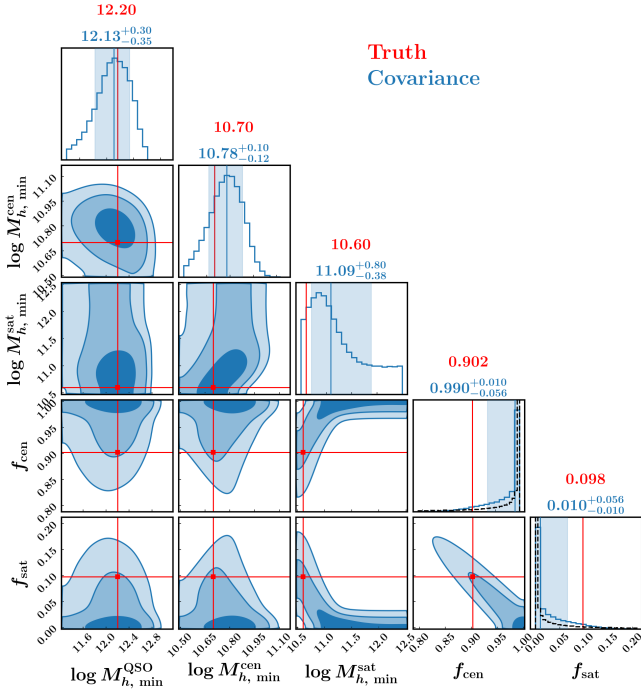
This has two practical consequences for the analysis presented in this work:

(i) The standard single-threshold HOD is a good approximation in the two-halo regime that dominates the total signal and drives the halo mass inference. The two-halo term is controlled by the central population, which is the same in both the single-threshold and split models when  $M_{h,\min}^{\text{cen}, [\text{O III}]} = M_{h,\min}^{[\text{O III}]}$ .

(ii) At small scales where one-halo contributions become significant, the single-threshold model could introduce a systematic bias if the true central and satellite mass thresholds differ substantially. Several mock data points fall within the one-halo radial range



**Figure 10.** Best-fit  $\chi_{GG}(r_p)$  (red) and  $\chi_{QG}(r_p)$  (blue) for a single mock realization at the fiducial truth  $\log[(M_{h,\min}^{\text{QSO}}, M_{h,\min}^{\text{cen}, [\text{O III}]}, M_{h,\min}^{\text{sat}, [\text{O III}]})/M_{\odot}] = (12.20, 10.70, 10.60)$ . Errorbars are computed from the diagonal of the full mock covariance at the fiducial model. The dashed red curve is the truth at the input parameters. Solid lines and shaded bands show the posterior median and the 16–84% range of the model.



**Figure 11.** Joint posterior in the five parameters  $\log[(M_{h,\min}^{\text{QSO}}, M_{h,\min}^{\text{cen}, [\text{O III}]}, M_{h,\min}^{\text{sat}, [\text{O III}]})/M_{\odot}]$ ,  $f_{\text{cen}}$ ,  $f_{\text{sat}}$  for the same realization as Fig. 10. Contours show the posterior using the full mock covariance, enclosing 1, 2, and 3 $\sigma$  Gaussian-equivalent posterior mass. Red lines mark the truth. The lower two parameters ( $f_{\text{cen}}$ ,  $f_{\text{sat}}$ ) are derived deterministically from  $\log[(M_{h,\min}^{\text{cen}, [\text{O III}]}, M_{h,\min}^{\text{sat}, [\text{O III}]})/M_{\odot}]$  via the FLAMINGO subhalo halo-mass functions and carry no independent information. The dashed black curves on the  $f_{\text{cen}}$  and  $f_{\text{sat}}$  1D diagonals show the induced prior on these fractions, obtained by mapping the flat  $\log[(M_{h,\min}^{\text{cen}, [\text{O III}]}, M_{h,\min}^{\text{sat}, [\text{O III}]})/M_{\odot}]$  priors.

( $r_p \lesssim 0.5 \text{ cMpc } h^{-1}$ ), and even though these data points do not dominate the fit, they do contribute to the total likelihood. If environmental processes have preferentially suppressed (or boosted) satellite [O III] luminosities, the effective satellite mass threshold would differ from the central threshold, introducing a model misspecification that is not captured by our current framework.

We flag this as a systematic uncertainty that future analyses with larger datasets and deeper spectroscopy should address. In principle, the one-halo regime offers a direct handle on the satellite fraction and could be used to jointly constrain  $M_{h,\min}^{\text{cen}, [\text{O III}]}$  and  $M_{h,\min}^{\text{sat}, [\text{O III}]}$  (and hence also the satellite fraction), provided sufficient signal-to-noise at small separations. At  $z \approx 6$ , the relatively short time since satellite infall may limit the magnitude of environmental effects, but they cannot be excluded *a priori*, particularly for emission-line selected samples where even modest changes in the instantaneous star formation rate translate directly into changes in [O III] detectability.

## 6 DISCUSSION AND CONCLUSIONS

We have developed a framework for robust high-redshift clustering inference by constructing realistic mock observations from the FLAMINGO-10k (F10k) dark-matter-only simulation. By generating 1000 mock realizations matched to the ASPIRE setup of 25 independent *JWST/NIRCam* WFSS pointings around  $z \approx 6.6$  quasars (including the field-by-field spatial coverage and sensitivity maps), we construct mock covariance matrices for the [O III]-emitter auto-correlation function  $\chi_{GG}(r_p)$  and the quasar-[O III]-emitter cross-correlation function  $\chi_{QG}(r_p)$ . These covariance matrices are computed on a grid of minimum halo masses ( $M_{h,\min}^{\text{QSO}}, M_{h,\min}^{\text{[O III]}}$ ), capturing the mass dependence of cosmic variance. We validate the inference framework through coverage tests, demonstrating that the covariance-based credible intervals have the correct statistical properties, and quantify the impact of adopting Poisson pair-count errors versus the full mock covariance matrices on inferred halo masses and correlation lengths. Our main results are as follows:

(i) Poisson errors underestimate the true uncertainty by factors of  $\approx 1.5$ – $2.7$ . By comparing the  $1\sigma$  constraints on  $M_{h,\min}^{\text{QSO}}$  from the joint QG+GG fit using the full covariance versus diagonal Poisson errors across a grid of 30 mass configurations, we find a systematic ratio  $\sigma_{\text{cov}}^{M_{h,\min}^{\text{QSO}}}/\sigma_{\text{Poisson}}^{M_{h,\min}^{\text{QSO}}}$  that increases with both  $M_{h,\min}^{\text{QSO}}$  and  $M_{h,\min}^{\text{[O III]}}$  (Fig. 6). At fiducial ASPIRE masses ( $\log(M_{h,\min}^{\text{QSO}}/M_{\odot}) = 12.2$ ,  $\log(M_{h,\min}^{\text{[O III]}}/M_{\odot}) = 10.6$ ), the Poisson posterior on the minimum halo masses is  $\approx 1.8\times$  narrower than the covariance-based posterior (Fig. 4).

(ii) The covariance structure is mass dependent. The off-diagonal elements of the covariance matrix become more prominent at higher halo masses (Figs. 2 and 3), because more massive, more highly biased tracers cluster more strongly, producing larger correlations and correspondingly larger covariance. This means that using a fixed fiducial covariance matrix, while preferable to Poisson errors, may still not capture the full mass dependence of the error budget. We therefore advocate for mass-dependent covariance matrices in the likelihood, as implemented in this work.

(iii) The inference test confirms that Poisson credible intervals are severely overconfident. Across 1000 mock realizations at the fiducial masses, the covariance-based  $1\sigma$  credible region contains the true parameters in  $\approx 65\%$  of realizations, consistent with the expected 68% within statistical uncertainty, while the Poisson  $1\sigma$  region contains the true parameters in only  $\approx 20\%$  of realizations (Fig. 7).

(iv) The QSO halo-mass inference is robust to the central–satellite treatment of the galaxy sample. The marginalized posterior on  $M_{h,\min}^{\text{QSO}}$  is essentially unchanged between the single-threshold analysis (§5.2) and the analysis with separate central and satellite thresholds (§5.5), indicating that the inferred  $M_{h,\min}^{\text{QSO}}$  does not depend on the detailed treatment of the 1-halo term.

Our results have important implications for existing and future high-redshift clustering studies. The commonly adopted Poisson error bars on the two-point correlation function at  $z \gtrsim 6$  should be regarded as lower bounds on the true uncertainty. Studies that have reported halo mass constraints without full simulation-based covariance matrices (e.g., Lee et al. 2006; Barone-Nugent et al. 2014; Harikane et al. 2016, 2022; García-Vergara et al. 2017, 2019; Arita et al. 2023; Eilers et al. 2024; Schindler et al. 2026; Shuntov et al. 2025; Lin et al. 2025a) likely underestimate the true error bars on inferred halo masses by factors of  $\approx 1.5$ –3, depending on the halo mass and survey geometry. This does not necessarily invalidate the central values of previous measurements, but it does mean that reported confidence intervals are too narrow, and that apparent tensions between studies may be less significant than they appear. Robust covariance-based methodology is well established at low-to-intermediate redshifts (see §1 and references therein), but has not previously been applied to clustering at  $z \gtrsim 6$ .

The framework presented here is readily applicable to other *JWST* WFSS programs targeting emission-line galaxies. In particular, the COSMOS-3D survey (Kakiichi et al. 2024) covers  $\approx 0.3 \text{ deg}^2$  with NIRCcam/WFSS, providing a large sample of emission-line galaxies (Meyer et al. 2025). The galaxy auto-correlation function measured from this blank-field survey can constrain the host halo mass and duty cycle of emission-line galaxies, and the mock covariance framework developed here can be adapted to the COSMOS-3D survey geometry to ensure robust error estimates. Other WFSS programs, including pure-parallel surveys such as SAPPHIRES (Sun et al. 2025), will further expand the sample of emission-line galaxies available for clustering analyses across multiple redshifts and tracers. Even in larger contiguous-mosaic surveys such as COSMOS-3D ( $\approx 0.3 \text{ deg}^2$ ,  $\approx 4.5\times$  the total ASPIRE area and sampled in a single connected volume rather than 25 independent pointings), cosmic variance does not become negligible: applying our framework to mock COSMOS-3D [O III]-emitter auto-correlation measurements, we find that Poisson errors still underestimate the true uncertainty of the halo mass by a factor of  $\approx 2$  (Huang et al., in prep.). Proper covariance modeling therefore remains essential across the range of current and near-future *JWST*/WFSS surveys to avoid overconfident constraints on the galaxy–halo connection.

## ACKNOWLEDGEMENTS

We acknowledge helpful conversations with the ENIGMA group at UC Santa Barbara and Leiden University. JH is grateful to the discussions with Feige Wang, Shane Bechtel, Daming Yang, Ben Wang, Bingcheng Jin, Xiaojing Lin, and Romain Meyer. JH and JFH acknowledges support from the National Science Foundation under Grant No. 2307180. JFH also acknowledges support for program XXX provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127. EP acknowledges support provided by NASA through the NASA Hubble Fellowship grant HST-HF2-51586.001-A awarded by the Space Telescope Science Institute, which is operated by the

Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555.

## DATA AVAILABILITY

The data underlying the findings of this study, together with the analysis code used to produce the results, are available from the corresponding author upon request.

## REFERENCES

- Abadi M. G., Moore B., Bower R. G., 1999, *MNRAS*, **308**, 947
- Abbott T. M. C., et al., 2022, *Phys. Rev. D*, **105**, 023520
- Arita J., et al., 2023, *ApJ*, **954**, 210
- Arita J., et al., 2026, *arXiv e-prints*, p. [arXiv:2602.11736](#)
- Bañados E., Venemans B., Walter F., Kurk J., Overzier R., Ouchi M., 2013, *ApJ*, **773**, 178
- Balogh M. L., Navarro J. F., Morris S. L., 2000, *ApJ*, **540**, 113
- Barone-Nugent R. L., et al., 2014, *ApJ*, **793**, 17
- Berlind A. A., Weinberg D. H., 2002, *ApJ*, **575**, 587
- Bernstein G. M., 1994, *ApJ*, **424**, 569
- Champagne J. B., et al., 2025, *ApJ*, 981
- Coil A. L., Hennawi J. F., Newman J. A., Cooper M. C., Davis M., 2007, *ApJ*, **654**, 115
- Cole S., Kaiser N., 1989, *MNRAS*, **237**, 1127
- Croom S. M., Shanks T., Boyle B. J., Smith R. J., Miller L., Loaring N. S., Hoyle F., 2001, *MNRAS*, **325**, 483
- Croom S. M., et al., 2005, *MNRAS*, **356**, 415
- Davis M., Peebles P. J. E., 1983, *ApJ*, **267**, 465
- Dodelson S., Schneider M. D., 2013, *Phys. Rev. D*, **88**, 063537
- Efstathiou G., Rees M. J., 1988, *MNRAS*, **230**, 5p
- Eilers A.-C., et al., 2024, *arXiv e-prints*, p. [arXiv:2403.07986](#)
- Endsley R., Behroozi P., Stark D. P., Williams C. C., Robertson B. E., Rieke M., Gottlöber S., Yepes G., 2020, *MNRAS*, **493**, 1178
- Fan X., Bañados E., Simcoe R. A., 2023, *ARA&A*, **61**, 373
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, **125**, 306
- Forero-Sánchez D., Rashkovetskiy M., Alves O., et al., 2025, *JCAP*, **2025**, 055
- Forouhar Moreno V. J., Helly J., McGibbon R., Schaye J., Schaller M., Han J., Kugel R., Bahé Y. M., 2025, *MNRAS*, **543**, 1339
- Friedrich O., Andrade-Oliveira F., Camacho H., et al., 2021, *MNRAS*, **508**, 3125
- García-Vergara C., Hennawi J. F., Barrientos L. F., Rix H.-W., 2017, *ApJ*, **848**, 7
- García-Vergara C., Hennawi J. F., Barrientos L. F., Rix H.-W., 2019, *ApJ*, **886**, 79
- García-Vergara C., Rybak M., Hodge J., Hennawi J. F., Decarli R., 2022, *ApJ*, **927**, 65
- Giner Mascarell M., Eilers A.-C., Storey-Fisher K., 2025, *arXiv e-prints*, p. [arXiv:2511.17413](#)
- Gunn J. E., Gott III J. R., 1972, *ApJ*, **176**, 1
- Guo H., Zheng Z., Zehavi I., Behroozi P. S., Chuang C.-H., Comparat J., et al., 2015, *MNRAS*, **453**, 4368
- Haiman Z., Hui L., 2001, *ApJ*, **547**, 27
- Han J., Jing Y. P., Wang H., Wang W., 2012, *MNRAS*, **427**, 2437
- Han J., Cole S., Frenk C. S., Benítez-Llambay A., Helly J., 2018, *MNRAS*, **474**, 604
- Harikane Y., et al., 2016, *ApJ*, **821**, 123
- Harikane Y., et al., 2022, *ApJS*, **259**, 20
- Hartlap J., Simon P., Schneider P., 2007, *A&A*, **464**, 399
- Hennawi J. F., et al., 2006, *AJ*, **131**, 1
- Hennawi J. F., Kist T., Davies F. B., Tamanas J., 2025, *MNRAS*, **539**, 2621
- Huang J., et al., 2026, *arXiv e-prints*, p. [arXiv:2602.04974](#)
- Jose C., Lacey C. G., Baugh C. M., 2016, *MNRAS*, **463**, 270
- Kaiser N., 1984, *ApJ*, **284**, L9
- Kakiichi K., Egami E., Fan X., Lyu J., Wang F., Yang J., et al., 2024, COSMOS-3D: A Legacy Spectroscopic/Imaging Survey of the Early Universe, JWST Proposal. Cycle 3, ID. #5893
- Kashino D., et al., 2023, *ApJ*, **950**, 66
- Kim S., et al., 2009, *ApJ*, **695**, 809
- Kitaura F.-S., Rodríguez-Torres S., Prada F., et al., 2016, *MNRAS*, **456**, 4156
- Kugel R., et al., 2023, *MNRAS*, **526**, 6103
- Landy S. D., Szalay A. S., 1993, *ApJ*, **412**, 64
- Larson R. B., Tinsley B. M., Caldwell C. N., 1980, *ApJ*, **237**, 692
- Lee K.-S., Gialalisco M., Gnedin O. Y., Somerville R. S., Ferguson H. C., Dickinson M., Ouchi M., 2006, *ApJ*, **642**, 63
- Lim S., Tacchella S., Maiolino R., Lovell C. C., Schaye J., 2025, *arXiv e-prints*, p. [arXiv:2511.09618](#)
- Lin X., et al., 2025a, *arXiv e-prints*, p. [arXiv:2504.08028](#)
- Lin X., et al., 2025b, *arXiv e-prints*, p. [arXiv:2505.02896](#)
- Lippich M., Sánchez A. G., Colavincenzo M., Sefusatti E., Monaco P., Blot L., Crocce M., et al., 2019, *MNRAS*, **482**, 1786
- Manera M., Samushia L., Tojeiro R., et al., 2013, *MNRAS*, **428**, 1036
- Martini P., Weinberg D. H., 2001, *ApJ*, **547**, 12
- Matsuoka Y., et al., 2019, *ApJ*, **872**, L2
- Matthee J., Mackenzie R., Simcoe R. A., Kashino D., Lilly S. J., Bordoloi R., Eilers A.-C., 2023, *ApJ*, **950**, 67
- Meyer R. A., Wang F., Kakiichi K., Giovinazzo E., Weibel A., Brammer G., Oesch P. A., et al., 2025, *arXiv e-prints*, p. [arXiv:2510.11373](#)
- Mo H. J., White S. D. M., 1996, *MNRAS*, **282**, 347
- Mo H. J., White S. D. M., 2002, *MNRAS*, **336**, 112
- Mohammad F. G., Percival W. J., 2022, *MNRAS*, **514**, 1289
- Moore B., Katz N., Lake G., Dressler A., Oemler A., 1996, *Nature*, **379**, 613
- Morselli L., et al., 2014, *A&A*, **568**, A1
- Mortlock D. J., et al., 2011, *Nature*, **474**, 616
- Moster B. P., Somerville R. S., Newman J. A., Rix H.-W., 2011, *ApJ*, **731**, 113
- Muñoz J. B., Mirocha J., Furlanetto S., Sabti N., 2023, *MNRAS*, **526**, L47
- Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, *MNRAS*, **396**, 19
- Oesch P. A., et al., 2023, *MNRAS*, **525**, 2864
- Osmer P. S., 1981, *ApJ*, **247**, 762
- Overzier R. A., 2016, *A&ARv*, **24**, 14
- Percival W. J., Ross A. J., Sánchez A. G., et al., 2014, *MNRAS*, **439**, 2531
- Pizzati E., Hennawi J. F., Schaller M., 2024a, *MNRAS*, **528**, 4466
- Pizzati E., Hennawi J. F., Eilers A.-C., et al., 2024b, *MNRAS*, **534**, 3155
- Porciani C., Magliocchetti M., Norberg P., 2004, *MNRAS*, **355**, 1010
- Read J. I., Wilkinson M. I., Evans N. W., Gilmore G., Kleya J. T., 2006, *MNRAS*, **366**, 429
- Robertson B. E., 2010, *ApJ*, **713**, 1266
- Schaye J., et al., 2023, *MNRAS*, **526**, 4978
- Schindler J.-T., et al., 2023, *ApJ*, **943**, 67
- Schindler J.-T., et al., 2026, *A&A*, **708**, A160
- Sellentin E., Heavens A. F., 2016, *MNRAS*, **456**, L132
- Shen Y., Strauss M. A., Oguri M., et al., 2007, *AJ*, **133**, 2222
- Shuntov M., et al., 2025, *arXiv e-prints*, p. [arXiv:2503.14280](#)
- Simpson C., Mortlock D., Warren S., Cantalupo S., Hewett P., McLure R., McMahon R., Venemans B., 2014, *MNRAS*, **442**, 3454
- Sinha M., Garrison L. H., 2020, *MNRAS*, **491**, 3022
- Sinha M., Berlind A. A., McBride C. K., Scoccimarro R., Piscionere J. A., Wibking B. D., 2018, *MNRAS*, **478**, 1042
- Sun F., Boyer M. L., Egami E., Pirzkal N., Rieke M. J., Sun J., 2024, A Novel Wavelength Calibration of NIRCcam WFSS with a Nearby Star-Forming Galaxy, JWST Proposal. Cycle 3, ID. #4924
- Sun F., et al., 2025, *arXiv e-prints*, p. [arXiv:2503.15587](#)
- Talts S., Betancourt M., Simpson D., Vehtari A., Gelman A., 2020, *arXiv e-prints*
- Trenti M., Stiavelli M., 2008, *ApJ*, **676**, 767
- Valageas P., Clerc N., Pacaud F., Pierre M., 2011, *A&A*, **536**, A95
- Wang F., et al., 2019, *ApJ*, **884**, 30
- Wang F., et al., 2021, *ApJ*, **907**, L1
- Wang F., et al., 2023, *ApJ*, **951**, L4
- Weibel A., et al., 2025, *arXiv e-prints*, p. [arXiv:2512.14212](#)
- White M., Tinker J. L., McBride C. K., 2014, *MNRAS*, **437**, 2594
- Yang X., Mo H. J., van den Bosch F. C., Jing Y. P., 2005, *MNRAS*, **357**, 608
- Yang J., et al., 2020, *ApJ*, **897**, L14
- Zehavi I., Zheng Z., Weinberg D. H., et al., 2011, *ApJ*, **736**, 59
- Zheng Z., Berlind A. A., Weinberg D. H., et al., 2005, *ApJ*, **633**, 791
- Zheng Z., Coil A. L., Zehavi I., 2007, *ApJ*, **667**, 760
- van den Bosch F. C., More S., Cacciato M., Mo H., Yang X., 2013, *MNRAS*, **430**, 725

## APPENDIX A: INFERENCE WITH THE MEAN MOCK CORRELATION FUNCTION

As a complement to the single-realization inference shown in Fig. 4, we repeat the MCMC analysis using the mean of all 1000 mock realizations as the “observed” data vector. Because the mean mock correlation function converges to the true halo model prediction, this isolates the effect of the error model (covariance vs. Poisson) on the posterior width without the confounding effect of noise fluctuations in any particular realization.

Fig. A1 shows the resulting correlation function fit and MCMC posterior. By construction, the data points coincide with the true model (black curves), and both the covariance and Poisson fits recover the input masses. However, the covariance-based posterior remains substantially broader than the Poisson posterior, confirming that the difference in uncertainty estimates is a systematic property of the error model rather than a consequence of noise in any individual realization.

While Fig. 2 shows how the correlation structure of  $\chi_{\text{QG}}$  varies with the quasar halo mass  $M_{h,\text{min}}^{\text{QSO}}$  at fixed  $M_{h,\text{min}}^{\text{[O III]}}$ , Fig. A2 shows the complementary dependence on  $M_{h,\text{min}}^{\text{[O III]}}$  at fixed  $M_{h,\text{min}}^{\text{QSO}}$ .

## APPENDIX B: CORRELATION LENGTH

Many high-redshift clustering studies infer halo masses by fitting a power-law correlation function to extract the correlation length  $r_0$ , converting  $r_0$  into an effective large-scale bias, and then mapping this bias onto a characteristic halo mass using analytical prescriptions (e.g., Shen et al. 2007; Arita et al. 2023; Eilers et al. 2024). To demonstrate that the Poisson error underestimation is not limited to the minimum halo mass inference presented in the main text, we repeat the covariance versus Poisson comparison for the power-law correlation length, following the same approach as in Huang et al. (2026). We model the real-space correlation function as a power law  $\xi(r) = (r/r_0)^{-\gamma}$ , where  $r_0$  is the correlation length and  $\gamma$  is the slope, fixed at  $\gamma = 2$ . The model prediction for each radial bin is obtained by substituting  $\xi(r_p, \pi) = (\sqrt{r_p^2 + \pi^2}/r_0)^{-\gamma}$  into Eq. 1, giving  $\chi_{V,\text{model}}(r_{p,\text{min}}, r_{p,\text{max}}; r_0, \gamma)$ .

We note that, unlike the minimum halo mass inference in §5.2 where the covariance matrix varies with the model parameters ( $M_{h,\text{min}}^{\text{[O III]}}$ ,  $M_{h,\text{min}}^{\text{QSO}}$ ) as written in Eq. 7, the power-law correlation length  $r_0$  is not a parameter of the mock catalog generation. Since our mock realizations are constructed from step-function HOD models parameterized by ( $M_{h,\text{min}}^{\text{[O III]}}$ ,  $M_{h,\text{min}}^{\text{QSO}}$ ), the covariance matrix cannot easily be made  $r_0$ -dependent. We therefore adopt a *fixed* covariance matrix evaluated at the fiducial model ( $\log M_{h,\text{min}}^{\text{[O III]}}/M_\odot$ ,  $\log M_{h,\text{min}}^{\text{QSO}}/M_\odot$ ) = (10.6, 12.2) for the  $r_0$  fit, as is common when the covariance cannot be directly linked to the fitted parameters (see, e.g., Norberg et al. 2009).

Assuming deterministic bias, where galaxies and quasars trace the same underlying dark matter distribution, the quasar auto-correlation function can be expressed as  $\xi_{\text{QQ}} = \xi_{\text{QG}}^2/\xi_{\text{GG}}$  (Croom et al. 2001; García-Vergara et al. 2017; Eilers et al. 2024). Following Eilers et al. (2024) and Huang et al. (2026), we jointly fit the [O III]-emitter auto-correlation and the quasar–[O III]-emitter cross-correlation to simultaneously constrain the quasar auto-correlation length  $r_0^{\text{QQ}}$  and

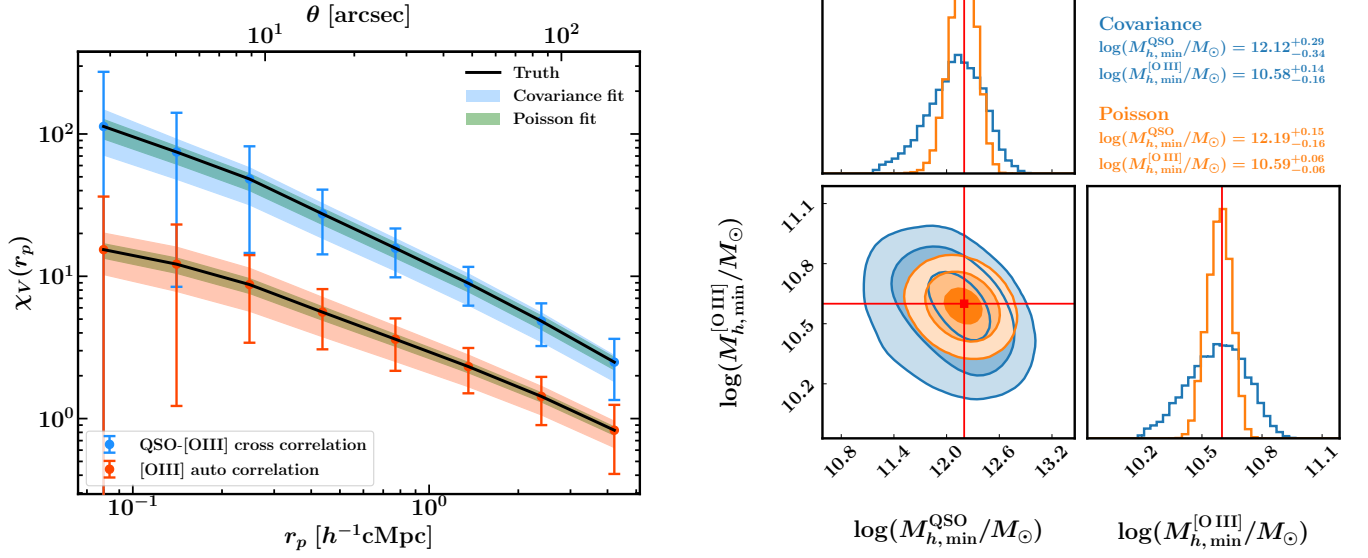
the galaxy auto-correlation length  $r_0^{\text{GG}}$ . The total likelihood is:

$$\mathcal{L}_{\text{tot}}(r_0^{\text{QQ}}, r_0^{\text{GG}}) = \mathcal{N}\left(\chi_{\text{GG}}^{\text{obs}} \middle| \chi_{\text{GG}}^{\text{model}}(r_0^{\text{GG}}, \mathbf{C}_{\text{GG}})\right) \cdot \mathcal{N}\left(\chi_{\text{QG}}^{\text{obs}} \middle| \chi_{\text{QG}}^{\text{model}}(r_0^{\text{QQ}}, r_0^{\text{GG}}, \mathbf{C}_{\text{QG}})\right), \quad (\text{B1})$$

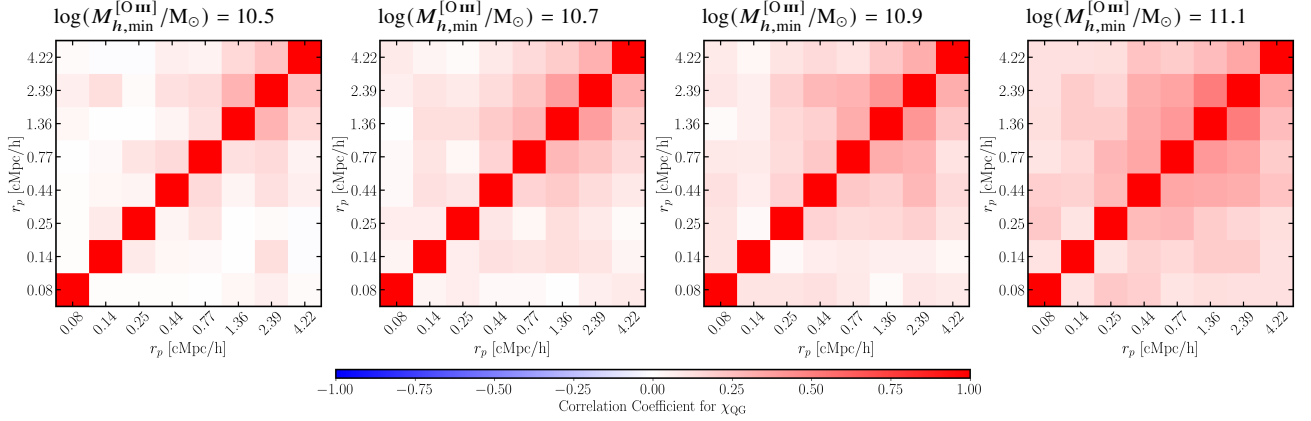
where  $\mathbf{C}_{\text{GG}}$  and  $\mathbf{C}_{\text{QG}}$  are evaluated at the fiducial masses.

Fig. B1 shows the joint posterior for ( $r_0^{\text{QQ}}$ ,  $r_0^{\text{GG}}$ ) when the true power-law model is used as the observed data (i.e., noiseless). The anti-correlation between the two parameters is evident: the joint fit couples  $r_0^{\text{QQ}}$  and  $r_0^{\text{GG}}$  through the cross-correlation, so increasing  $r_0^{\text{QQ}}$  while decreasing  $r_0^{\text{GG}}$  can leave  $\chi_{\text{QG}}$  nearly unchanged. As with the minimum halo mass inference, the covariance-based posterior (blue) is broader than the Poisson posterior (orange), demonstrating that Poisson errors underestimate the uncertainty on the correlation lengths.

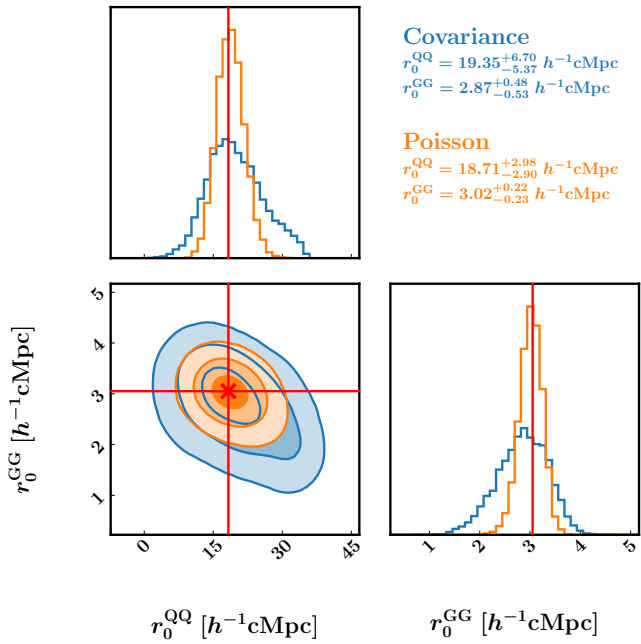
This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.



**Figure A1.** Same as Fig. 4, but using the mean of 1000 mock realizations as the observed data. *Left:* Since the mean converges to the true model, the data points lie on the black truth curves by construction. Only the  $1\sigma$  posterior bands are shown (no median curves). The covariance fit bands (blue/red shaded) remain broader than the Poisson fit bands (green shaded), confirming that the broadening is a systematic effect of the error model rather than noise in any single realization. *Right:* Both posteriors peak at the true input masses, but the covariance-based posterior (blue) is broader than the Poisson posterior (orange), consistent with the single-realization result.



**Figure A2.** Similar to Fig. 2, but showing how the correlation structure of  $\chi_{\text{QG}}$  depends on the galaxy halo mass. From left to right,  $\log(M_{h,\min}^{[\text{OIII}]} / M_{\odot})$  increases from 10.5 to 11.1, while the quasar halo mass is fixed at  $\log(M_{h,\min}^{\text{QSO}} / M_{\odot}) = 12.5$ . The off-diagonal correlation becomes slightly stronger as  $M_{h,\min}^{[\text{OIII}]}$  increases. Compared with Fig. 2, the correlation structure of  $\chi_{\text{QG}}$  depends more strongly on the tracer galaxy halo mass than on the quasar halo mass.



**Figure B1.** Joint posterior for the quasar auto-correlation length  $r_0^{\text{QQ}}$  and galaxy auto-correlation length  $r_0^{\text{GG}}$ , from the joint power-law fit ( $\gamma = 2$ ) to  $\chi_{\text{GG}}$  and  $\chi_{\text{QG}}$  assuming deterministic bias ( $\xi_{\text{QQ}} = \xi_{\text{QG}}^2 / \xi_{\text{GG}}$ ). The observed data is set to the true (noiseless) model. Blue filled contours use the full mock covariance; orange filled contours use Poisson errors. Red cross mark the true input values. The anti-correlation between  $r_0^{\text{QQ}}$  and  $r_0^{\text{GG}}$  arises from the coupling through the cross-correlation constraint. The Poisson posterior is systematically narrower, underestimating the true uncertainty.