

Angular clustering and bias of photometric quasars in the Kilo-Degree Survey Data Release 4

Anjitha John William, Maciej Bilicki, Wojciech A. Hellwing, Szymon J. Nakoneczny, and Priyanka Jalan

Center for Theoretical Physics, Polish Academy of Sciences, al. Lotników 32/46, 02-668 Warsaw, Poland *

ABSTRACT

We investigate the angular clustering and effective bias of photometrically selected quasars in the Kilo-Degree Survey Data Release 4 (KiDS DR4). We update the previous photometric redshifts (photo- z s) of the KiDS quasars using Hybrid- z , a deep learning framework combining four-band KiDS images and nine-band KiDS+VIKING magnitudes. Hybrid- z is trained on the latest Dark Energy Spectroscopic Instrument (DESI) DR1 and Sloan Digital Sky Survey (SDSS) DR17 quasars matching with KiDS, and achieves average bias $\langle \delta z \rangle < 0.01$ and scatter $\sim 0.04(1+z)$ on a test sample. The updated catalog of $\sim 157k$ quasars over 777 deg^2 is divided into four tomographic bins spanning $0.1 \leq z_{\text{phot}} \leq 2.7$. In each bin, we measure the angular two-point correlation function and compare it with theoretical predictions for dark matter clustering. We estimate the best-fit scale-independent quasar bias, which increases from $b \approx 1.6$ at $z \approx 0.6$ to $b \approx 4.0$ at $z \approx 2.2$, and is well matched by a quadratic relation in redshift. Our clustering analysis indicates that KiDS quasars reside in dark matter halos of mass $\log_{10}(M_{\text{eff}}/h^{-1}M_{\odot})$ in the range $\sim 12.7\text{--}12.9$ and effective peak heights ν_{eff} rising from ~ 1.5 to 2.9 over our redshift span. We study two systematics that could affect the bias derivation: stellar contamination and the redshift distribution assumed in the theoretical modeling. The former has a negligible effect, whereas the latter significantly impacts the derived $b(z)$, emphasizing the importance of redshift calibration. Our work is the first cosmological application of quasars selected from KiDS and paves the way for future extensions in the final KiDS DR5, the Legacy Survey of Space and Time, or the 4-metre Multi-Object Spectroscopic Telescope.

Key words. quasars: photometric redshifts - angular clustering - techniques: machine learning - catalogs: photometric & spectroscopic

1. Introduction

Quasars, also known as quasi-stellar objects (QSO, Schmidt 1963), are the brightest of the active galactic nuclei (AGN, Rees 1984). Their high luminosity ($\sim 10^{10} - 10^{14} L_{\odot}$) arises from the matter accretion onto a supermassive black hole (SMBH) situated at the centre of their host galaxy (Salpeter 1964; Zel'dovich & Novikov 1964). This allows us to observe quasars up to very high redshifts ($z \gtrsim 7$), which makes them valuable astrophysical and cosmological probes of the late and early-time Universe (Fan et al. 2023). Quasar emission originates from regions that lie well within the gravitational influence of the supermassive black hole (SMBH), providing information on the physics of black hole accretion and AGN activity (Marziani et al. 2006; Spilker et al. 2025). At galactic scales ($\sim 1\text{--}10 \text{ kpc}$), quasars help us to understand galaxy formation, growth, and quenching processes (Prochaska & Hennawi 2009). On cosmological scales ($\gtrsim 1 \text{ Mpc}$), they trace the large-scale structure (LSS) of the Universe (Song et al. 2016). For instance, measuring quasar clustering allows us to constrain the redshift evolution of their bias with respect to the underlying dark matter (DM) and to reveal their host halo masses, and they provide access to cosmological information such as the detection of baryon acoustic oscillations (e.g. Myers et al. 2007; Ata et al. 2018; Adame et al. 2025). Their spatial distribution can be cross-correlated with galaxy and cosmic microwave background weak gravitational lensing maps to probe the growth of structure and constrain cosmological parameters (e.g. Ménard & Bartelmann 2002; DiPompeo et al. 2016;

Alonso et al. 2023; Luo et al. 2024; Piccirilli et al. 2024). Additionally, the large-scale clustering of quasars is sensitive to primordial non-Gaussianity through scale-dependent bias effects (Slosar et al. 2008).

Even though quasars are among the brightest extragalactic objects, it remains challenging to select them in surveys. Their compact emission region outshines the host galaxy, and their small angular size, combined with the limited resolution of telescopes, causes them to appear point-like at high redshifts, making them morphologically indistinguishable from stars in imaging data. As a result, the quasar selection must rely on other properties, such as colors, variability, or spectroscopy, to separate them from the vastly more numerous stellar population. However, the spectroscopic quasar identification method (Richards et al. 2002; Lyke et al. 2020; Storey-Fisher et al. 2024) is time-consuming and observationally expensive. It restricts surveys to relatively bright sources and limits the sky coverage and depth of quasar samples. To overcome these limitations, we can apply photometric selection techniques to produce larger and more complete QSO catalogues (e.g. Carrasco et al. 2015; Nakoneczny et al. 2019; Chaussidon et al. 2023; Feng et al. 2025). Photometrically classified quasars offer a higher completeness but at the cost of increased contamination compared to their spectroscopic counterparts. However, they provide substantial cosmological information through their large numbers and wide sky coverage.

A photometric target selection technique for quasar identification was applied within the Kilo-Degree Survey (KiDS, Kuijken et al. 2019) by Nakoneczny et al. 2021, hereafter N21. KiDS covers about 1000 deg^2 of the southern sky in its fourth

* e-mail: (anjithajm,bilicki,hellwing)@cft.edu.pl

data release. This paper presents the first study of angular clustering and effective bias for KiDS DR4 quasars. Our results demonstrate that the quasar bias evolution is consistent with a quadratic dependence on redshift, in agreement with prior findings from other surveys (Myers et al. 2007; Shen et al. 2009; Laurent et al. 2017; Elvedt et al. 2024). We investigate potential systematic effects, finding that uncertainties in the redshift distribution significantly affect the quasar bias estimation. This highlights the necessity of a precise redshift calibration. In contrast, stellar contamination within the sample has a negligible effect on the clustering measurements and bias inference. Together, these findings establish a framework for quasar clustering studies with KiDS, and they contribute insight into quasar bias evolution across cosmic time.

A prerequisite to studying the evolution of quasar properties such as their bias is the availability of redshift measurements. Similar to non-active galaxies, the redshifts of quasars can be determined through spectroscopic and photometric methods. Measuring the accurate and precise spectroscopic redshifts (hereafter spec- z s) is expensive and time-consuming, however. The less precise photometric redshift (photo- z) estimation is based on broadband photometry (Baum 1957; Koo 1985; Salvato et al. 2019; Newman & Gruen 2022). This alternative approach can be relatively quick compared to spectroscopy, and may give a redshift estimate for all sources within an imaging survey. We used the empirical approach of machine-learning (ML) to derive photo- z s for KiDS DR4 quasars. These algorithms typically belong to the supervised ML category and find the relation between multi-band photometric quantities and the redshift by being trained on datasets containing photometry and the corresponding spec- z s. Examples of their applications to quasars include Brescia et al. (2013); D’Isanto & Polsterer (2018); Nakoneczny et al. (2021); Nakazono et al. (2024).

As compared to other galaxies, quasars exhibit prominent broad emission lines, such as Ly α , C IV, Mg II (Peterson 1997), and they move across different photometric bands at different redshifts. An accurate photometric redshift estimation for quasars therefore benefits from a broad wavelength coverage that extends from the blue to the infrared (IR) to help break degeneracies caused by different emission lines entering similar filter combinations at different redshifts, as well as ambiguities from their smooth power-law continua (Salvato et al. 2019). Still, these effects often produce characteristic peaks and dips in the quasar photometric redshift distribution. Furthermore, QSO spectra are dominated by non-thermal emission, giving them colors distinct from most other extragalactic sources and necessitating training sets specifically tailored for quasar photo- z measurements.

Within the broader ML framework, deep-learning (DL), often using images among the inputs, has continued to find increasingly numerous uses for photo- z estimation, in particular for quasars (e.g. Pasquet-Itam & Pasquet 2018; Yao et al. 2023; Roster et al. 2024). The term “deep” indicates artificial neural networks (ANN) with multiple layers (often tens or hundreds of them concatenated) and numerous neurons in each layer. We integrated two types of neural networks: dense (McCulloch 1943; Rosenblatt 1958) and convolutional neural networks (CNN, Lecun & Bengio 1995). Our model is called Hybrid- z ¹ since it uses magnitudes and multi-channel images, and it was originally developed for the KiDS bright galaxy sample in John William et al. 2025, hereafter A25. In its previous application, Hybrid- z improved the earlier ANN-based photo- z s

for the KiDS DR4 bright sample (Bilicki et al. 2021), reducing the scatter of photo- z residuals by 20%. We applied a similar DL model to update the photo- z s of the photometrically classified KiDS DR4 QSO sample, which was previously derived with ANNs employing quasar magnitudes and colors by N21.

In addition to developing a new ML model for photo- z s, a key improvement of this work over N21 is the use of the recent Dark Energy Spectroscopic Instrument Data Release 1 (DESI DR1, DESI Collaboration et al. 2026) and the Sloan Digital Sky Survey Data Release 17 (SDSS DR17, Abdurro’uf et al. 2022) to build the spectroscopic training set for our photo- z model. As compared to the SDSS DR14 (Abolfathi et al. 2018) spec- z data used previously by N21, the addition of DESI DR1 considerably improved the quality of our training set and the resulting photo- z estimation. DESI not only includes many more quasars than SDSS in the KiDS fields, it also extends the QSO spectroscopic coverage across the color-redshift space.

We used the KiDS DR4 quasar sample with the updated photo- z s to study the connection between the QSO distribution and the underlying DM field via angular clustering, which is a common observational statistic for studying the properties of the LSS projected on the sky (Peebles 1973, 1980). Following previous studies (e.g. Myers et al. 2007), we considered quasar auto-correlations in photo- z bins. We used these measurements to constrain the evolution with redshift of the effective quasar bias and of their effective host halo mass.

In the standard cosmological framework, the evolution of density perturbations yields predictions for the power spectrum of the total matter density contrast field, including contributions from both dark and baryonic matter. Observationally, we did not directly probe the clustering of the total matter density field. Instead, we measured the spatial distribution of luminous tracers, in our case, quasars. On sufficiently large scales (megaparsecs), the DM and quasar overdensities (δ_q) are related via the bias parameter, such that $\delta_q = b\delta_m$. This bias is expected to evolve with redshift (Shen et al. 2009; Laurent et al. 2017) and its evolution can be traced from angular clustering measured tomographically, that is, in redshift bins. If additional information on quasar redshift distribution per bin is available, for instance via photo- z s, the angular two-point correlation function (2PCF) can be related to the theoretical three-dimensional 2PCF or its Fourier counterpart, power spectrum. The factor linking the observed quasar correlations and those theoretically expected for DM is the bias b . In summary, incorporating the redshift distribution enables the projection of theoretical three-dimensional dark matter clustering into angular space, allowing us to infer the quasar bias from the observed 2PCF.

This paper is structured as follows: In Sec.2, we describe the photometric and spectroscopic datasets used in our analysis. Sec.3 presents our Hybrid- z model and the photo- z estimation for the quasar sample. The clustering analysis of this sample is detailed in Sec.4. Finally, we summarise our findings and discuss future prospects in Sec.5.

2. Data

In this section, we describe the data we used to select quasars and estimate their photo- z s. The quasar sample was derived from KiDS Data Release 4 (DR4, Kuijken et al. 2019), and the spectroscopic data for the training sample were obtained from DESI DR1 and SDSS DR17.

KiDS is a multi-band imaging survey designed for weak lensing studies and it was carried out by the European Southern Observatory (ESO) at the VLT Survey Telescope (VST,

¹ <https://hybrid-z.readthedocs.io/en/latest/>

Capaccioli & Schipani 2011). The images were taken in four optical broad bands (ugri) and supplemented with five near-infrared bands (ZYJHK_s) from the VISTA Kilo-degree INfrared Galaxy survey (VIKING, Edge et al. 2013) overlapping with KiDS on the sky. Here we employ KiDS DR4, covering about 1000 deg² before masking. The nine-band images were processed by the KiDS team, using in particular the Gaussian aperture and point spread function method (GAAP, Kuijken 2008) to obtain uniformly measured u to K_s magnitudes, which we employ throughout. In addition to the magnitudes, similarly as in A25, for photo- z derivations we also use four-band ugri images, leaving the extension to VIKING images for future work. The optical images in KiDS DR4 are organised into 4×1006 survey tiles² with a uniform pixel scale of 0.2 arcsec (de Jong et al. 2015). For the Hybrid- z model described below, we made cutouts centred at quasar positions with a size of $5'' \times 5''$ (25×25 pixels), as most of our quasars are smaller than this.

Our quasar sample is selected from KiDS DR4 following the previous work by N21³. That selection was obtained by applying an ANN-based classification model using 9-band detections, trained and tested on SDSS DR14 spectroscopy. In the KiDS DR4 QSO sample, the objects are further categorised into safe and extrapolation regimes based on the feature space coverage with respect to the training set. For the clustering analysis presented in Sec. 4, we only used the safe subset, limited to $r < 22$ mag, whose feature space lies well within the domain of the N21 training data. For this subset, the N21 classification model achieves test-data purity of 97% and completeness of 94%. As we do not update the quasar selection over that previous study, limiting our analysis to only the safe QSO set mitigates potential contamination from stars and non-active galaxies, which could affect photo- z quality and subsequent clustering analyses. There are $\sim 157k$ KiDS DR4 objects in this safe QSO category. Unless otherwise specified, the term KiDS DR4 quasar sample hereafter refers to this subset of safe quasars. We note, however, that the general training set does not have the safe selection applied (see below). We leave for future study a possible analysis of quasars derived from KiDS DR5 (Wright et al. 2024; Feng et al. 2025).

While we adopt the same KiDS DR4 QSO sample as selected by N21, we update and improve the photo- z s derived there. This was achieved by changing the method (ML photo- z model) and by adding considerably more training data with spectroscopic labels. The former is provided thanks to using the Hybrid- z approach, employing the nine-band magnitudes and four-band optical images, previously developed and validated on the KiDS-bright galaxy sample (Bilicki et al. 2021) by A25. We discuss the particular Hybrid- z implementation and its performance for KiDS DR4 quasars in the next Sec. 3.

As for the new training data, we replace the SDSS DR14 QSOs previously used by N21 by a joint sample of SDSS DR17 and DESI DR1 quasars. These two datasets cover the northern (equatorial) KiDS area and provide spectroscopically confirmed quasars with spectroscopic redshifts (that may be considered exact) to train and validate our photo- z model. The joint SDSS+DESI sample, with only unique objects kept⁴, is dominated by DESI and has about 115k counterparts in the full KiDS DR4 catalogue before we apply any further cuts. Of these, about 104k have all nine ugriZYJHK_s bands measured in KiDS DR4

– a selection we apply as our neural network model requires a numerical value to be provided for each of the magnitudes. This general cross-match of DESI+SDSS with KiDS DR4 will be our training and test set for the Hybrid- z model. Note that this training sample goes deeper in magnitude and redshift than the final safe-QSO sample we use for the clustering analysis, which is a conscious choice to provide photo- z estimates also for quasars fainter than eventually employed. The median r -band magnitude of the quasar sample with spectroscopic redshift labels is $r \sim 21.4$ mag, while the depth expressed as the 99th percentile is $r \sim 23.7$. The median spectroscopic redshift of this training set is $z = 1.66$ and 99% of the quasars are contained within the redshift range $0.22 \leq z \leq 3.32$.

3. Deriving the photometric redshift

The architecture and configuration of our photo- z model, Hybrid- z , are detailed in A25. In brief, Hybrid- z comprises two parallel branches: a CNN branch that processes ugri KiDS galaxy cutouts through initial convolutional layers and four Inception modules to extract multi-scale features, and an ordinary (fully-connected) neural network (ONN) branch that processes standardised nine-band KiDS+VIKING magnitudes through a sequence of dense layers. The CNN and ONN outputs are concatenated and passed through fully connected layers, with the final layer providing a photo- z estimate. The primary modification between the current model configuration and that detailed in A25 is the exclusion of the sigmoid activation function in the output layer. This adjustment was made since the redshift distribution of quasars within our dataset extends well above $z > 1$, while previously we were limiting the predictions to the $0 < z_{\text{phot}} < 1$ range, appropriate for the KiDS-bright sample. Therefore, we applied a linear activation function in the final layer. The rest of the Hybrid- z architecture remains identical to that of A25. The hyperparameters differ from those in A25 and were selected through empirical testing across various configurations and evaluating the corresponding model performance on the quasar dataset we used. We employed a batch size of 32 and an initial learning rate of 0.001.

As far as data preprocessing is concerned, we adapt the same techniques as in A25; normalisation to the image pixel values and standardisation of the nine-band magnitude features. Furthermore, we employed data augmentation, including 90° rotation, flipping, and translations (height and width shifts), to expand our training set (Yang et al. 2022). We used $\zeta = \ln(1 + z_{\text{spec}})$ as labels (Baldry 2018) to achieve a more balanced distribution in redshift. This transformation of true values helps to maintain stable gradients in the descent-based optimisation of the model by ensuring numerical stability. Although the model is trained in this log-transformed space, all evaluations and error metrics are reported after inverting the transformation.

The dataset is divided into training, validation, and testing disjoint sets using the SCIKIT-LEARN library function `train_test_split`, with approximately two-thirds of the data assigned to training and the remainder evenly split between validation and testing subsets. In our case, this corresponds to $\sim 104k$ quasars in total, with $\sim 73k$ allocated for training and $\sim 18k$ each for validation and testing.

The performance of the Hybrid- z model is primarily evaluated using the Huber loss function (Huber 1992). The loss curves for the training and validation sets show a steady decline over successive epochs, indicating efficient learning and convergence. The close agreement between the two curves suggests that the model generalises well to unseen data without signs of overfit-

² <https://kids.strw.leidenuniv.nl/DR4>

³ <https://kids.strw.leidenuniv.nl/DR4/quasarcatalog.php>

⁴ Note that majority of SDSS quasars are also present in the DESI catalog. In the cross-match with KiDS DR4, only about 1900 out of 17,000 SDSS QSOs are not in DESI.

ting. We used early stopping criteria (Prechelt 2012) to determine the optimal number of epochs. Training ceased automatically after 10 consecutive epochs without improvement, with any reduction in validation loss considered as an indication of improvement. The model was trained over 70 epochs in our model setup. Finally, the estimated QSO photo- z s for the test sample are evaluated by standard statistical metrics. To quantify the statistical accuracy, we calculated the bias, $\delta z = z_{\text{phot}} - z_{\text{spec}}$, and normalised bias, $\Delta z = \frac{\delta z}{1+z_{\text{spec}}}$. Standard deviation of normalised bias $\sigma_{\Delta z}$ and scaled median absolute deviation (SMAD) of Δz are the measures of scatter, that is, statistical precision. Here, $\text{SMAD}(\Delta z) = 1.4826 \times \text{median}(|\Delta z - \text{median}(\Delta z)|)$.

3.1. Results for the test data

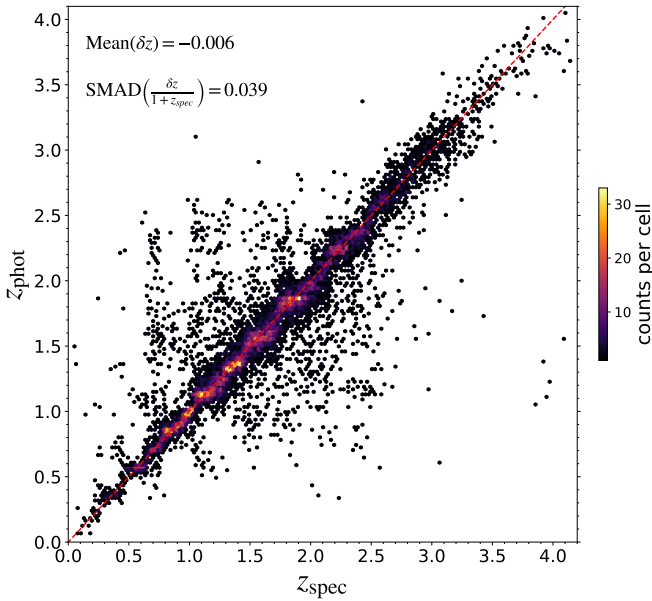


Fig. 1. Density plot comparing spectroscopic and photometric redshifts of KiDS DR4 quasars derived with our Hybrid- z method. The color bar indicates the counts per hex bin for a test sample not seen in the training. The red dashed line corresponds to the identity relation $z_{\text{spec}} = z_{\text{phot}}$.

In this section, we analyze the KiDS DR4 QSO photo- z s derived using the Hybrid- z model for the test sample, with true redshift known, but not seen by the model in training nor validation. Starting with the statistics averaged over the entire test set of about 18,200 quasars drawn from a general cross-match of KiDS with DESI+SDSS spectroscopy, the mean bias is below 0.01 in absolute terms (see first row of Table 1), while the scatter of $\Delta z \equiv (z_{\text{phot}} - z_{\text{spec}})/(1 + z_{\text{spec}})$ is $\sigma_{\Delta z} = 0.133$ (standard deviation) or $\text{SMAD}(\Delta z) = 0.039$ (scaled median absolute deviation). Therefore, our photo- z s are on average unbiased, that is, the overall scatter is considerably larger than the mean bias. On the other hand, the fact that SMAD is much smaller than SD indicates their non-Gaussian nature with extended tails (outliers), which is also visible in the direct $z_{\text{spec}} - z_{\text{phot}}$ comparison, presented in Fig. 1, as estimates far apart from the diagonal. This is typical behaviour for quasar photo- z s (e.g. Curran 2022; Yao et al. 2023).

In order to directly compare the Hybrid- z performance for KiDS DR4 QSOs with the previous derivations by N21, we cross-matched the above-mentioned test sample with the safe set from that earlier work. This reduces the number of common objects to about 10k mostly because of the $r < 22$ mag limit in that sample. The results are presented in rows 2-3 of Table 1 and we notice that the Hybrid- z model exhibits a notable improvement over N21 in all the statistics. As we use the N21 safe quasar sample for clustering measurements below, the statistics provided in row 3 of Table 1 us general quantification of the photo- z performance for our dataset.

We illustrate a direct comparison of our photo- z predictions for the test sample with the true ones in Fig. 1. While the data generally follow the diagonal (identity line $z_{\text{spec}} = z_{\text{phot}}$), as expected for photo- z s unbiased on average, characteristic step-like behaviour is observed in detail. This specific redshift focusing, where a range of spec- z s have similar photo- z s predicted, or vice versa, is commonly observed in quasar photo- z derivations (e.g. Nakoneczny et al. 2021; Kunsági-Máté et al. 2022; Yao et al. 2023; Moss et al. 2025). We interpret these steps as artefacts driven by emission line transitions across filters. For instance, At lower redshifts (e.g., $z \sim 0.4$), the $H\alpha$ emission line (restframe $\lambda = 6562.8\text{\AA}$) lies within the Z-band (central $\lambda = 8800\text{\AA}$), contributing significantly to redshift estimation. As redshift increases, $H\alpha$ shifts out of that filter coverage, and although $H\beta$ remains visible, this transitional zone introduces increased uncertainty. Around $z \sim 1.1$, $H\beta$ ($4861\text{\AA}$) also exits the optical range, and the model increasingly relies on redshifted MgII ($2800\text{\AA}$) and CIII] ($1908.73\text{\AA}$) emission lines, which can lead to degeneracies due to similar observed colors across different redshifts. These transitions result in redshift confusion, where the model tends to assign many objects the same redshift, producing the observed step-like artefacts in the distribution. Each “step” reflects a regime where the model’s redshift estimation is dominated by a different combination of emission lines.

The KiDS quasar clustering analysis presented below in Sec. 4 is done tomographically in four photo- z bins, it is therefore relevant to analyze the redshift performance for each of these bins. This is presented in rows 4-7 of Table 1, where the numbers refer to a cross-match of our test set with the N21 safe quasar sample, binned in Hybrid- z redshift estimates. This time we observe more considerable biases in our photo- z s, with mean $\delta z = z_{\text{phot}} - z_{\text{spec}}$ surpassing 0.01 in absolute terms, with some improvement if the $1 + z$ scaling is applied (i.e. for Δz). On the other hand, the scatter is better behaved, staying at a similar level for the particular bins as for the full test sample in terms of the standard deviation and SMAD of the residuals. This indicates that, once binned in photo- z , our redshift estimates retain their statistical precision, but lose on accuracy.

3.2. Updated KiDS DR4 quasar photometric redshifts

Having checked the Hybrid- z performance for quasars with known redshifts from spectroscopic data, we trained the model on the KiDS DR4 $\times$ DESI+SDSS cross-match and applied it to the entire safe quasar sample as selected by N21. As already clarified, our feature space are 4-band ugr images and 9-band KiDS+VIKING magnitudes, with a requirement of having detections in each. This dataset, flux-limited to $r < 22$ mag, includes 157,419 quasar candidates on $\sim 777 \text{ deg}^2$ (effective sky coverage of KiDS DR4 after masking). In Fig. 2 we compare the redshift distribution of the spectroscopic training quasars present

Table 1. Photometric redshift statistics of the Hybrid-z model for quasars in a test sample derived from a cross-match of KiDS DR4 with DESI+SDSS spectroscopy.

Selection	Size ¹	$\langle z_{\text{spec}} \rangle$	$\langle z_{\text{phot}} \rangle$	$\langle \delta z \rangle$	$\langle \Delta z \rangle$	$\sigma_{\Delta z}$	SMAD(Δz)
full test set ²	18,274	1.662	1.656	-0.006	0.008	0.133	0.039
test set cross-matched with the N21 sample: ³							
previous N21 derivations ⁴	10,226	1.672	1.738	0.067	0.038	0.131	0.054
this work	10,226	1.672	1.664	-0.007	0.005	0.116	0.033
$0.1 \leq z_{\text{phot}} \leq 0.8$	718	0.656	0.596	-0.061	-0.019	0.104	0.029
$0.8 < z_{\text{phot}} \leq 1.2$	1996	1.079	1.018	-0.061	-0.017	0.095	0.030
$1.2 < z_{\text{phot}} \leq 1.8$	3459	1.510	1.489	-0.021	0.002	0.110	0.038
$1.8 < z_{\text{phot}} \leq 2.7$	3290	2.110	2.158	0.047	0.027	0.139	0.031

¹ Number of quasars in the subsample.

² Entire spectroscopic test set without cross-matching to N21 selection nor any cuts.

³ Spectroscopic test set cross-matched with the N21 safe quasar sample ($r < 22$ mag).

⁴ N21 derived photo-z statistics in the spectroscopic test set cross-matched with the N21 safe quasar sample ($r < 22$ mag).

⁵ Row 1 applies to the entire test sample (not seen by the model in training). Rows 2-3 are for a cross-match with the earlier N21 derivations: original results in row 2, and our updates in row 3. Finally, rows 4-7 present a break-down into photometric redshift bins which are employed for tomographic angular clustering in this paper.

also in our photometric set (filled histogram) with two photometric redshift distributions: previous from N21 (green dashed) and the one derived here with Hybrid-z (blue solid). Each of the histograms was normalised to unity under the curve. The comparison indicates that the original relatively smooth dN/dz_{spec} is mapped to dN/dz_{phot} with visible peaks and dips for N21 and Hybrid-z derivations. This is another manifestation of the effects already discussed previously in the context of Fig. 1 that quasar photo-zs tend to be focusing around particular redshift values due to redshifted line transitions. This is a caveat one should bear in mind when analysing the overall photo-z performance for quasars. Indeed, as shown above in Table 1, once broken down into particular redshift bins, the statistics may deteriorate. On the other hand, for the subsequent clustering analysis we chose sufficiently broad photo-z ranges to include neighboring ‘peaks’ and ‘dips’ in the redshift distribution and hence partly mitigate the related issues.

4. Angular clustering and bias evolution of KiDS quasars

In this section, we study the clustering of KiDS quasars along with their effective bias, peak height and host halo mass evolution. Our observable is the angular 2PCF measured in photo-z bins and compared to theoretical predictions for the underlying DM field. We also examine two possible systematics – stellar contamination and redshift distribution modeling.

4.1. Measurements: angular correlation function

We use the angular two-point correlation function (2PCF), $\omega(\theta)$, as our observable. The 2PCF quantifies the excess probability of finding a pair of quasars separated by a given distance, here expressed as an angle θ on the sky (Peebles 1973), as we are dealing with photometric data without exact redshifts. To compute the 2PCF, we use the standard Landy & Szalay (1993) estima-

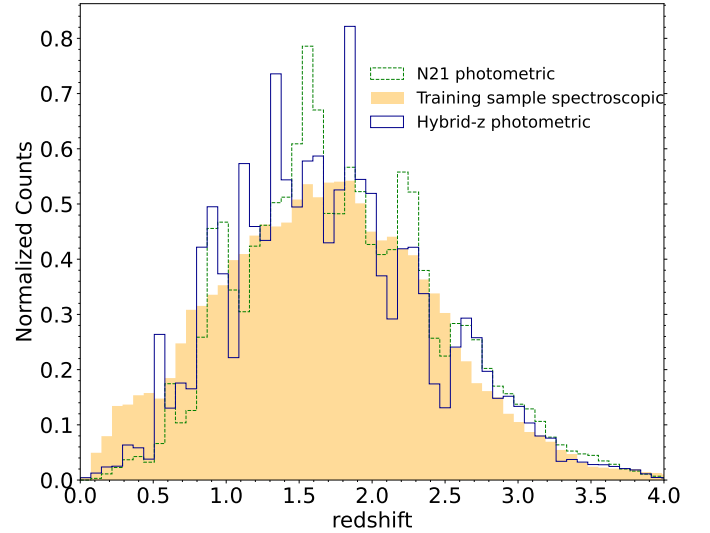


Fig. 2. Comparison of the redshift distributions for the KiDS DR4 safe quasar sample. The filled orange histogram shows the spectroscopic training sample used in our Hybrid-z model. The green dotted line represents the photo-zs of KiDS DR4 safe quasars from N21, and the blue line corresponds to the photo-zs derived in this work using the Hybrid-z framework.

tor.

$$\omega_{\text{obs}}(\theta) = \frac{DD(\theta) - 2DR(\theta) + RR(\theta)}{RR(\theta)}. \quad (1)$$

Here, $DD(\theta)$, $DR(\theta)$, and $RR(\theta)$ represent the pair counts normalised by number density within an angular bin centred at θ , for the data-data, data-random, and random-random pairs, respectively. The random catalogue covers the same footprint as the real data and includes about 100x more points than the quasar sample, for each of the redshift bins used. It is generated by

first uniformly distributing points within the range of the minimum and maximum values of right ascension and declination of the data and then applying the KiDS DR4 binary HEALPix mask with $N_{\text{side}} = 4096$ to restrict the random dataset to the survey coverage. We did not employ any corrections to the randoms that would compensate for survey systematics. While appropriate methodology has been developed for KiDS (Johnston et al. 2021; Yan et al. 2025), our quasar sample is too sparse for such organised randoms approach to be applicable. Position-dependent variations of survey depth could introduce systematic effects in clustering measurements, which usually lead to the amplitudes being overestimated (e.g. Johnston et al. 2021; Vakili et al. 2023; Yan et al. 2025). However, our sample is limited to $r < 22$ mag, which is well below the KiDS limiting magnitude of $r \sim 25$ mag. In addition, we work with point sources, which should be less sensitive to effects such as varying PSF. Therefore, we do not expect the effective depth to be a dominant source of uncertainty for our analysis.

Our measurements are performed in nine logarithmically-spaced angular bins in the range $0.05^\circ < \theta < 1^\circ$. The lower value is driven by shot noise effects in our relatively sparse dataset, while the upper one corresponds to very large physical scales once deprojected from angles to distances, and additionally is chosen to mitigate the main KiDS systematic, which is number density variations between tiles of 1 degree at a side. We have checked that indeed the measured $\omega(\theta)$ displays ‘jumps’ when passing to larger scales, hence we discard them in the analysis. Restricting measurements to scales below 1° , while necessary, will of course not remove the systematics arising from tile-to-tile variations, as correlations between quasars near tile edges will still be present. However, similarly as with the general effect of variable depth, we assess this to be of less importance here as our sample covers the bright end of KiDS. As for small scales, while we do measure the correlations down to arcminutes, the first angular bins are not used to derive the best-fit quasar bias when comparing with theoretical predictions. This is because the corresponding physical scales fall within the non-linear regime where our assumed simple relation between matter and quasar clustering is no longer adequate.

In our analysis we employ four photo- z bins with edges $z_{\text{phot}} = \{0.1, 0.8, 1.2, 1.8, 2.7\}$ (Table.1). The minimum and maximum values were chosen to cover the bulk of the redshift distribution (Fig. 2), while the bin widths and divisions are selected to account for photo- z scatter, quasar number density and the artefacts in the redshift distribution, as discussed in Sec. 3.2. The angular 2PCF is computed separately in each photo- z bin using the TreeCorr package (Jarvis et al. 2004), and the measurements are shown as points in Fig. 3. The number of sources in each redshift bin is indicated in the legend of each panel in Fig. 3. The error bars attached to the measurements in Fig. 3 are the square root of the diagonal elements of the covariance matrix (Sec. 4.3). Generally, the larger error bars in the first redshift bin than those in the other three bins are primarily driven by the smaller number of sources and the increased relative impact of shot noise in that bin.

We only used auto-correlations of quasars in the redshift bins. However, due to photo- z errors, there is also clustering signal in bin cross-correlation (Blake et al. 2006; Agarwal et al. 2014; Ho et al. 2015), notwithstanding the magnification lensing effect correlating the bins even for exact redshift tomography (Breton et al. 2022). What is more, photo- z bin cross-correlations could be used as a further diagnostic for systematics, for instance, in case of redshift outliers (Balaguera-Antolínez et al. 2018). However, we chose to focus

on auto-correlations only for two main reasons. First, proper modeling of cross-correlations would require deeper knowledge of the underlying true redshift distributions for the bins, which we do not have. Second, such modeling would need to properly account for lensing magnification, which is usually quantified by assuming magnitude-limited sampling (Maartens et al. 2021). While our quasar selection does have a flux cut applied, it is far from magnitude-limited due to complexities of ML-based catalogue construction. We are therefore not able to properly model the cross-correlation measurements to either use them as extra information or to evaluate residual systematics.

4.2. Theoretical modelling

We compare the measured angular 2PCF with the theoretical prediction derived for matter clustering to estimate quasar bias. The theoretical angular 2PCF is obtained from the angular power spectrum C_ℓ via the Legendre transformation:

$$\omega_m(\theta) = \frac{1}{4\pi} \sum_{\ell=\ell_{\min}}^{\ell_{\max}} (2\ell + 1) C_\ell P_\ell(\cos \theta), \quad (2)$$

where $P_\ell(\cos \theta)$ is the Legendre polynomial of degree ℓ , and the multipole moment ℓ relates to the angular scale θ via $\ell = 180^\circ/\theta$. The sum in the above equation should formally run from $\ell = 1, \dots, +\infty$, but we restrict the summation to $\ell_{\min} = 30$ and $\ell_{\max} = 10^4$ since our angular binning ranges from 0.05° to 1° .

The angular power spectrum (PS) is derived by projecting the three-dimensional PS, $P(k, z)$, along the line of sight, typically using the Limber approximation:

$$C_\ell = \int \frac{dz}{H(z)\chi^2(z)} \left(\frac{dN}{dz} \right)^2 P_m \left(\frac{\ell + 0.5}{\chi(z)}, z \right), \quad (3)$$

where $H(z)$ is the Hubble parameter and $\chi(z)$ is the comoving radial distance to redshift z . The function dN/dz represents the normalised redshift distribution of the quasar population for a given redshift bin; ideally this should be the true redshift distribution, which we do not know exactly not having spec- z s for the full sample. As discussed below, we will compare two approaches to approximate dN/dz , one by using photo- z s directly and the other by cross-matching with spec- z samples.

The term $P_m(k, z)$ is the three-dimensional non-linear matter PS evaluated at the wavenumber $k = \frac{\ell+0.5}{\chi(z)}$ and redshift z . To obtain it, for the matter-only case, we use the Core Cosmology Library (CCL; Chisari et al. (2019)) which uses the CLASS algorithm (Blas et al. 2011) for the linear PS and non-linear corrections to the matter PS are applied using the Halofit prescription (Smith et al. 2003; Takahashi et al. 2012). This requires cosmological parameters and here we adopt the flat Λ CDM cosmology with $\Omega_c = 0.2642$ (cold dark matter fraction), $\Omega_b = 0.0493$ (baryonic matter fraction), $n_s = 0.965$ (spectral index), $H_0 = 67.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Hubble constant), $\sigma_8 = 0.811$ (amplitude of matter density fluctuations on a scale of $8 h^{-1} \text{ Mpc}$) and a cosmological constant (Planck Collaboration et al. 2020). In principle, the cosmological model could be varied in the analysis, but our measurements are not sufficiently sensitive to constrain the quasar bias and cosmological parameters. We therefore fixed the latter to best-fit Planck values. The resulting $\omega_m(\theta)$ curves, computed for each redshift bin using dN/dz_{phot} as the assumed redshift distribution, are shown as black dashed lines in Fig. 3. In Sec. 4.4 we detail how we use these theoretical predictions to derive quasar bias by comparing with the measured 2PCF. First, we discuss in Sec. 4.3 the details of the covariance matrix (CM) needed for these calculations.

4.3. Covariance estimation

In order to derive quasar bias by comparing the observed 2PCF with its theoretical counterpart, we need the CM quantifying the correlations of $\omega(\theta)$ at different scales, which in our case are particular bins in θ . We estimate the CM directly from the data via the Bootstrap method (Mohammad & Percival 2022). It is a non-parametric internal error estimation method for a measured quantity, where multiple realisations of the dataset are generated by randomly sampling with replacement. In our analysis, we divided the KiDS DR4 survey footprint into $N_{\text{patch}} = 200$ spatial patches, and each bootstrap sample was constructed by randomly selecting N_{patch} patches with replacement. This allowed us to repeat some patched and omit others.

For each bootstrap realisation, $\omega(\theta)$ is recalculated using the selected patches. This procedure is repeated $N_{\text{boot}} = 15,000$ times, resulting in N_{boot} bootstrap realisations of $\omega(\theta)$. N_{boot} is determined by checking that the eigenvalues of the covariance matrix have converged. The bootstrap estimate of the covariance matrix is then computed as:

$$C_{ij} = \frac{1}{N_{\text{boot}} - 1} \sum_{k=1}^{N_{\text{boot}}} [\omega_k(\theta_i) - \bar{\omega}(\theta_i)] [\omega_k(\theta_j) - \bar{\omega}(\theta_j)], \quad (4)$$

where $\omega_k(\theta_i)$ and $\omega_k(\theta_j)$ are the angular 2PCF measured in the k -th bootstrap sample at angular bins θ_i and θ_j , respectively, and $\bar{\omega}(\theta_i)$ is the mean value across all bootstrap samples:

$$\bar{\omega}(\theta_i) = \frac{1}{N_{\text{boot}}} \sum_{k=1}^{N_{\text{boot}}} \omega_k(\theta_i). \quad (5)$$

The inverse of the CM, $C_{ij}^{-1, \text{boot}}$, estimated using Bootstrap method is biased because of the finite number of subsamples; that is, $C_{ij}^{-1, \text{boot}}$ does not exactly correspond to the true inverse CM. We applied the Anderson-Hartlap-Kaufman (AHK) debiasing factor (Hartlap et al. 2007; Kaufman et al. 2008; Vakili et al. 2023) to correct for this bias in the inverse CM calculation. AHK debiasing factor is defined as:

$$f_{\text{AHK}} = \frac{N_{\text{patch}} - N_d - 2}{N_{\text{patch}} - 1}, \quad (6)$$

where $N_d = 9$ is the number of angular bins used. The corrected inverse CM thus becomes:

$$C_{ij}^{-1, \text{AHK}} = f_{\text{AHK}} \times C_{ij}^{-1, \text{boot}}. \quad (7)$$

This corrected inverse covariance is then used for uncertainty-weighted model fitting to derive the quasar bias. We use the TreeCorr package to compute the Bootstrap estimate of the CM. TreeCorr employs k-means clustering using the angular positions (right ascension, declination) of quasars to generate the patches for covariance estimation.

4.4. Quasar bias and host halo mass estimation

Here we detail how we derive the quasar effective bias of KiDS quasars, $b(z)$, from the angular correlation measurements. Under the assumption of scale-independent biasing, the observed angular 2PCF of quasars is modelled as a scaled version of the matter $\omega(\theta)$, with the scaling governed by the square of the effective bias parameter:

$$\omega_q(\theta) = b^2(z) \omega_m(\theta). \quad (8)$$

In each redshift bin, square of quasar bias $b^2(z)$ is inferred through a chi-squared minimisation:

$$\chi^2 = \sum_{i,j} [\omega_q(\theta_i) - b^2(z) \omega_m(\theta_i)]^T C_{ij}^{-1, \text{AHK}} [\omega_q(\theta_j) - b^2(z) \omega_m(\theta_j)], \quad (9)$$

In this analysis, we fit the observed quasar clustering over the angular range $0.05 \leq \theta(\text{deg}) \leq 1.0$. We employed a Monte Carlo approach to robustly estimate the goodness-of-fit (Fumagalli et al. 2022). We used the χ^2 calibration method to account for covariance matrix estimation effects and correlated data points. This is done through moment-matching to a scaled χ^2 distribution. The result is a calibrated reduced χ^2 , denoted as χ_{red}^2 in Fig. 3. We determine b^2 in each redshift bin by minimising the χ^2 statistic, and derive b and its uncertainties from the resulting $\chi^2(b)$ curve. Owing to the asymmetry of $\chi^2(b)$, the inferred bias values have asymmetric 1σ uncertainties. These are presented in Figs. 3 and 5, as well as in Table. A.1.

Given a normalised redshift distribution dN/dz , the effective redshift in each bin is computed as:

$$z_{\text{eff}} = \frac{\int_{z_{\text{min}}}^{z_{\text{max}}} z \left(\frac{dN}{dz} \right) dz}{\int_{z_{\text{min}}}^{z_{\text{max}}} \left(\frac{dN}{dz} \right) dz} \quad (10)$$

In Fig. 3, the z_{eff} was computed using photometric dN/dz , which we take as our fiducial approach.

The fitted quasar bias increases steadily from ~ 1.6 in the first redshift bin to ~ 4.0 in the last z -bin (Fig. 3 and Table A.1). This trend agrees with the expectations from halo bias models and previous studies (Myers et al. 2007; Ross et al. 2009; Eftekharzadeh et al. 2019). We calculated the effective peak height ν_{eff} directly from our bias measurements using the Tinker et al. (2010) and Comparat et al. (2017) (hereafter T10 and C17 respectively) bias- ν_{eff} relations, and derived the effective host halo masses M_{eff} for each redshift bin by inverting their halo mass-bias relations. Table 2 presents the ν_{eff} and $\log_{10} M_{\text{eff}}(h^{-1} M_{\odot})$ and their lower and upper bounds at each effective redshift. For both models, ν_{eff} increases with redshift, and the corresponding halo masses lie in the range $\log_{10} M_{\text{eff}}(h^{-1} M_{\odot}) \simeq 12.7-12.9$. These trends are consistent with the widely accepted picture of quasar host halo mass (Shen et al. 2009; Eftekharzadeh et al. 2019; Petter et al. 2023; Pizzati et al. 2024; Giner Mascarell et al. 2025). We note that the selection effects of our quasar sample, particularly at high redshifts, are not accounted for in this analysis. Table 3 presents a comparison of quasar host halo mass estimates from the literature alongside our results based on C17. A direct comparison across studies requires caution for several reasons, including differences in the adopted halo mass statistic, cosmological parameters, redshift ranges, and sample selection criteria such as luminosity thresholds, photometric versus spectroscopic identification, and survey depth. Giner Mascarell et al. (2025) and Eftekharzadeh et al. (2019) report minimum halo masses; Pizzati et al. (2024) reports an average halo mass; while Petter et al. (2023), Shen et al. (2009), and this work report effective halo masses. Nevertheless, we find that our estimates remain consistent with the literature, and the typical values are $\sim 10^{12.5} - 10^{13} h^{-1} M_{\odot}$.

4.5. Correcting for stellar contamination

Stellar contamination of the quasar sample is a possibly significant factor affecting the clustering measurements and quasar bias

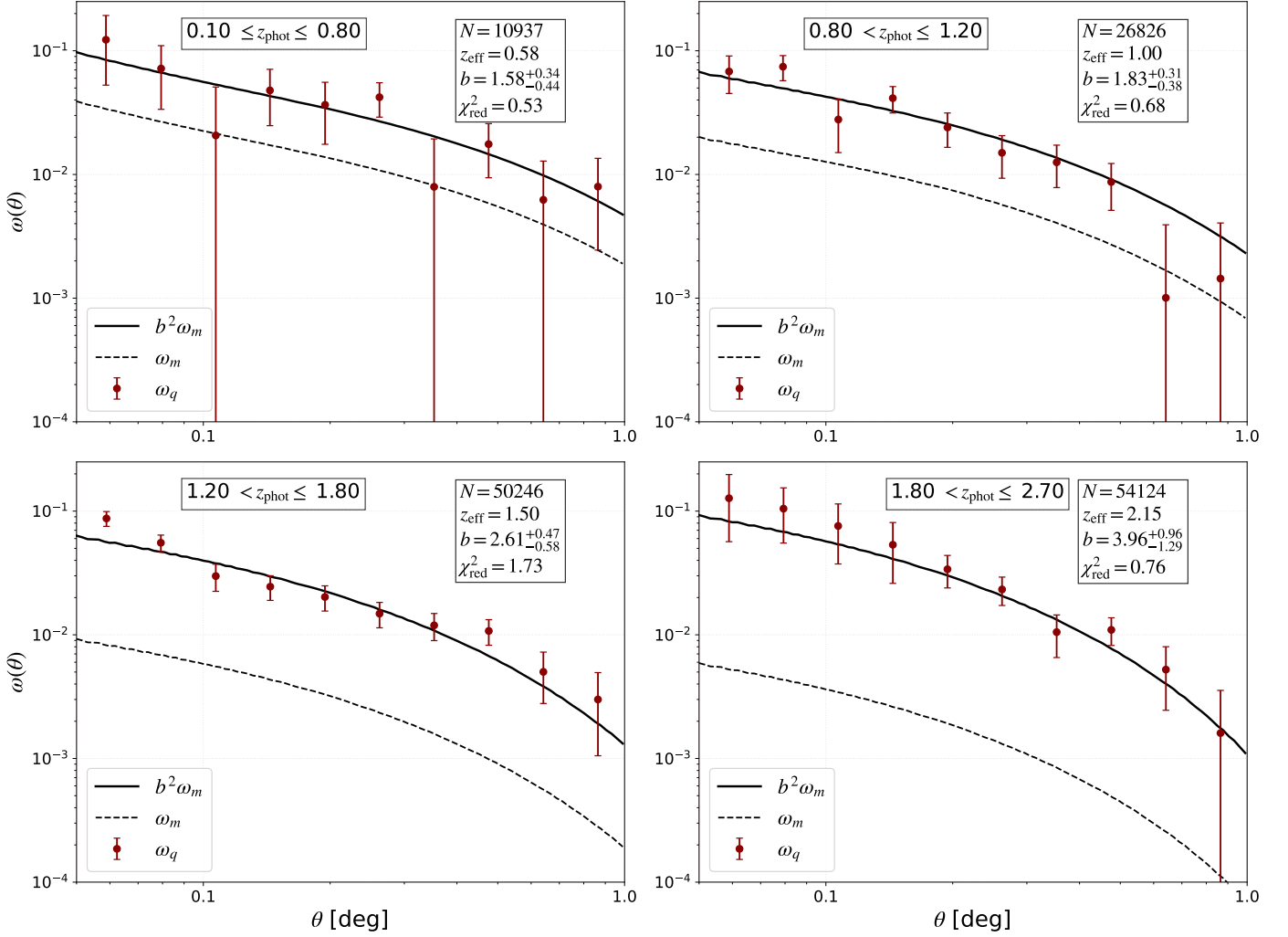


Fig. 3. Angular auto-correlation function of KiDS DR4 quasars measured in four photometric redshift bins. The red data points with error bars are the quasar 2PCF measurements, while the dashed lines show the predicted matter 2PCF $\omega_m(\theta)$, obtained for Planck Λ CDM cosmology and quasar redshift distribution taken as dN/dz_{phot} . The solid black line shows $b^2\omega_m$, with the best-fit bias value indicated in the legend. N and z_{eff} are the number of quasar objects and the effective redshift, respectively. The measurements shown here were corrected for stellar contamination using $p_{\text{star}} = 0.99$ as discussed in Sec. 4.5.

derivation. Stars are expected to have a nearly flat angular 2PCF, in contrast to steep power-law for extragalactic sources such as quasars. Stellar contamination would therefore affect the slope of the measured quasar 2PCF by making it less steep.

In order to inspect and mitigate the stellar contamination effect, we take a similar approach as in Myers et al. (2006). Namely, we model the observed angular 2PCF of our sample as being composed of a true quasar clustering component and that from stars, appropriately weighted by the purity of the quasar sample.⁵ As discussed in Myers et al. (2006), it can be shown that this leads to the following expression for true quasar clustering (neglecting cross-terms from star-quasar correlations):

$$\omega_q(\theta) = \frac{\omega_{\text{obs}}(\theta) - (1 - a^2)\omega_s(\theta)}{a^2}, \quad (11)$$

where $\omega_s(\theta)$ is angular 2PCF of stars, while a is quasar sample purity, for which we take $a = 0.98$ from the KiDS DR4 QSO classification model of N21.

⁵ In what follows we assume that only stars can contaminate the quasar sample, that is, we neglect possible contamination from non-QSO galaxies.

We measure the star 2PCF ω_s from samples of star candidates provided by the N21 classification model⁶. For that, we select objects based on different stellar classification probability thresholds, p_{star} . A threshold of $p_{\text{star}} \geq 0.99$ yields a star catalogue of higher purity, containing fewer quasars misclassified as stars compared to catalogues with $p_{\text{star}} \geq 0.90$ and $p_{\text{star}} \geq 0.75$. The autocorrelation function of these stellar samples, $\omega_s(\theta)$, was measured over the angular range $0.05^\circ \leq \theta \leq 1.0^\circ$ following the procedure described in Sec. 4.1, and the results are presented in Fig. B.1. Notably, the clustering amplitude of the $p_{\text{star}} \geq 0.99$ sample remains approximately flat across the considered angular scales, indicating a low residual clustering signal from extragalactic sources, while the angular 2PCF for the $p_{\text{star}} \geq 0.90$ and $p_{\text{star}} \geq 0.75$ samples exhibit a clear decline with increasing angular separation.

We used the above-discussed measurements of $\omega_s(\theta)$ for the various p_{star} thresholds in Eq. (11) to correct the quasar clustering measurements for stellar contamination. These corrected measurements were further employed to estimate quasar bias as

⁶ <https://kids.strw.leidenuniv.nl/DR4/quasarcatalog.php>

Table 2. Effective peak height ν_{eff} and halo mass, M_{eff} of KiDS DR4 quasars obtained by using the T10 and C17 models for each redshift z_{eff} .

Model	z_{eff}	ν_{eff}	$\log_{10} M_{\text{eff}} (h^{-1} M_{\odot})$
T10	0.58	$1.54^{+0.25}_{-0.41}$	$12.86^{+0.18}_{-0.22}$
	1.00	$1.73^{+0.21}_{-0.29}$	$12.65^{+0.11}_{-0.13}$
	1.50	$2.21^{+0.24}_{-0.34}$	$12.70^{+0.07}_{-0.07}$
	2.15	$2.85^{+0.38}_{-0.61}$	$12.73^{+0.13}_{-0.15}$
C17	0.58	$1.54^{+0.25}_{-0.41}$	$12.93^{+0.17}_{-0.21}$
	1.00	$1.73^{+0.21}_{-0.29}$	$12.68^{+0.12}_{-0.13}$
	1.50	$2.21^{+0.24}_{-0.34}$	$12.72^{+0.07}_{-0.07}$
	2.15	$2.85^{+0.38}_{-0.61}$	$12.76^{+0.13}_{-0.15}$

Table 3. Comparison of quasar host halo mass estimates in this work and in the literature.

Work	z_{eff}	$\log_{10} M (h^{-1} M_{\odot})$
This work (using C17 model) ³	0.58	$12.93^{+0.17}_{-0.21}$
	1.00	$12.68^{+0.12}_{-0.13}$
	1.50	$12.72^{+0.07}_{-0.07}$
	2.15	$12.76^{+0.13}_{-0.15}$
Giner Mascarell et al. (2025) ¹	0.5	13.03
	1.5	12.91
	2.5	12.96
	3.5	12.99
Pizzati et al. (2024) ²	6.0	12.33
Petter et al. (2023) ³		
Unobscured QSO	1.4	12.58
Obscured QSO	1.4	12.93
Eftekharzadeh et al. (2019) ¹	1.5	12.36
Shen et al. (2009) ³		
Radio-loud QSO	1.5	~ 13.0
Radio-quiet QSO	1.5	~ 12.30

¹ Minimum halo mass

² Average halo mass

³ Effective halo mass

discussed in Sec. 4.4. We found no significant variation in the bias values across the different p_{star} thresholds. This suggests that our bias estimates are robust to moderate levels of residual stellar contamination. For the final analysis and interpretation, we adopted the stellar catalogue with $p_{\text{star}} \geq 0.99$ for the stellar contamination correction in the quasar clustering measurements (Fig. 3).

4.6. Effect of the assumed redshift distribution

Bias inference from the angular 2PCF depends on the assumed redshift distribution dN/dz , as this quantifies the projection of the 3D quasar distribution onto the sky, and enters the model quadratically (see Eq. 3). To investigate the impact of the assumed redshift distribution on quasar bias measurements, we compared two approaches. The fiducial one, for which the results are presented in Fig. 3, uses the photometric estimates, dN/dz_{phot} , directly. The second method follows Myers et al. (2007): KiDS DR4 quasars in each of the photo- z bins are cross-matched with the DESI+SDSS spec- z quasar sample, which

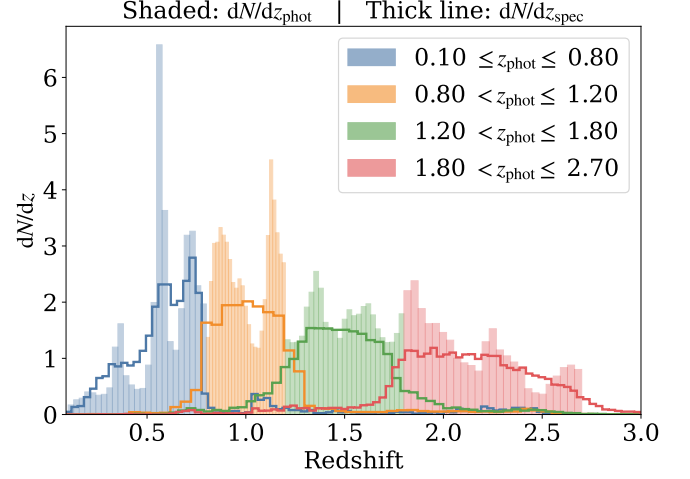


Fig. 4. Quasar redshift distributions for the four tomographic bins used in this study. Shaded histograms are the photo- z distributions of the KiDS DR4 QSO sample, while the thick lines correspond to the spec- z distributions obtained by direct cross-match of the quasars from a given photo- z bin with overlapping DESI and SDSS spectroscopy. Each distribution is normalised to unity under the curve.

gives dN/dz_{spec} for a given bin. The derived dN/dz are compared in Fig. 4, where shaded bars present the photo- z distributions, while the thick lines illustrate dN/dz_{spec} . In that plot, each of the tomographic redshift distributions is normalised to unity under the curve to appreciate better the differences in the redshift support for each of the bins.

By design, the dN/dz_{phot} are truncated at bin edges and do not reflect the photo- z uncertainties scattering the objects between the bins. The true redshift distributions will therefore always be broader than the photo- z s suggest. What is more, at higher redshifts, photo- z errors generally increase, leading to a larger fraction of outliers. Consequently, dN/dz_{phot} will deviate more substantially from the true underlying redshift distribution as the redshifts increase.

On the other hand, the spec- z distributions derived via the cross-match are broader than the photometric ones and indicate some outliers at values far apart from bin centres. We emphasise, however, emphasize that the spec- z distribution derived from the cross-matched sample is not necessarily representative of the true redshift distribution of our entire quasar sample, as the N21 selection may contain quasar candidates not represented by DESI or SDSS in the color space. Therefore, bias values inferred using dN/dz_{spec} may not be inherently more accurate than those obtained from the photo- z distribution of the entire sample.

The shape and width of the redshift distribution significantly affect the theoretical predictions of the angular 2PCF. If dN/dz is narrower, the integrand in Eq. (3) results in a higher angular clustering amplitude compared to that of a broader distribution. This in turn will lead to lower best-fit quasar bias for the narrower dN/dz for a given redshift bin. This is indeed what we find, as illustrated in Fig. 5 and quantified in Table A.1. The two tables compare the best-fit effective bias derivations for the KiDS DR4 quasars when using dN/dz_{phot} (blue points in Fig. 5) or dN/dz_{spec} (green points) in the modeling. In Fig. 5 we additionally show simple phenomenological fits for the redshift dependence of quasar bias: $b(z) = b_0 + b_1z + b_2z^2$ for dN/dz_{phot} and dN/dz_{spec} . The best-fit values for b_0, b_1, b_2 are provided in the plot while their errors are listed in Table C.1 and illustrated as shaded bands in the Figure. We note that such quadratic re-

lations fit the datapoints very well. On the other hand, as we do not account for redshift-bin correlations (which would require building a relevant covariance matrix), the fit uncertainties will be somewhat underestimated. We can firmly conclude, however, that using spectroscopic redshifts to model dN/dz gives bias values larger than the photometric case for the considered redshift range.

These results show that the choice of the underlying redshift distribution has a non-negligible impact on the inferred quasar bias estimation for photometric samples. This emphasises the importance of reliable dN/dz calibration for clustering-based cosmological analyses with such quasar datasets. Here, we adopt the quasar bias derived from the dN/dz_{phot} of the full sample as our fiducial result, because this redshift distribution corresponds directly to the dataset used in the observed angular 2PCF measurements, unlike the spec- z distribution of the cross-match, which may not be fully representative of the entire sample.

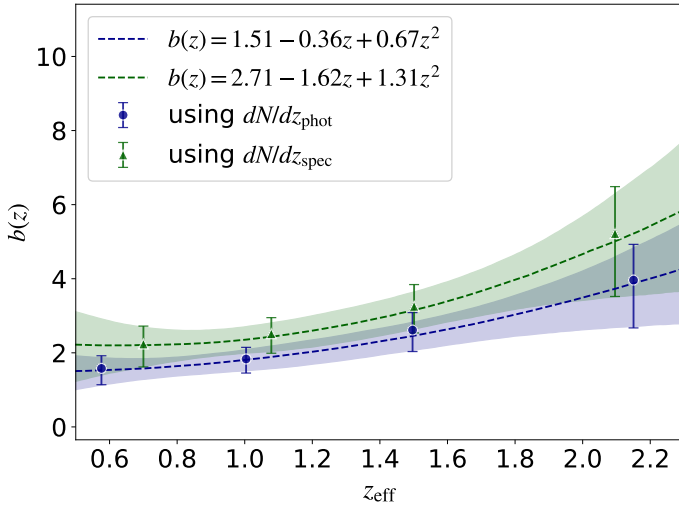


Fig. 5. Effective bias of KiDS DR4 quasars as a function of redshift, estimated with the angular clustering analysis. We present results from two modelling choices: blue circles were obtained by assuming dN/dz_{phot} directly as the underlying redshift distribution in the photo- z bins, while for green triangles we used per-bin cross-matches of KiDS photometric quasars with DESI and SDSS spectroscopy. Other model assumptions are the same in both cases. The dashed lines indicate the best-fit models of the form $b(z) = b_0 + b_1z + b_2z^2$. The corresponding best-fit parameters are shown in the legend, while their uncertainties are reported in Table C.1. The shaded regions represent the $\pm 1\sigma$ uncertainty bands of the fitted models.

4.7. Comparison with previous derivations of the quasar bias

The effective bias of quasars has been measured in a number of past studies and here we compare our results with the literature. For this comparison, we chose those where the large-scale bias was estimated in tomographic bins using photometric and spectroscopic redshifts. These are in particular: Croom et al. (2005), who employed the spectroscopic sample from the 2dF QSO Redshift Survey (2QZ, Croom et al. 2004) covering the range $0.3 < z < 2.2$; already discussed Myers et al. (2007), who used photometrically classified quasars drawn from SDSS DR4 spanning $0.4 < z < 2.8$; Ross et al. (2009), who presented bias measurements for spectroscopic quasars from SDSS DR5 within the redshift range $0.3 \leq z \leq 2.2$; and Laurent et al. (2017), who studied eBOSS spectroscopic quasar sample for redshifts $0.9 < z < 2.2$. We note that these past works used different

cosmological parameters than us, care should be therefore taken when comparing the bias derivations. The main factor influencing the bias estimate will be different adopted σ_8 , which rescales the amplitude of the theoretical power spectrum, and hence directly affects the best-fit effective bias. Moreover, the previous analyses reported $b(z)$ values at the mean redshifts of the adopted bins, which depends on the assumed redshift distributions per bin.

We illustrate the comparison of our results from this paper with the past quasar bias measurements in Fig. 6. We observe very good consistency of our estimates with the 2QZ and SDSS derivations respectively by Croom et al. (2005) and Myers et al. (2007); Ross et al. (2009), while the more recent eBOSS constraints by Laurent et al. (2017) lie systematically below ours, especially for the highest- z bins. This difference is enhanced especially due to considerably smaller errorbars on $b(z)$ in this latter paper as compared to other works, including ours. However, the discrepancy hardly exceeds 2σ even for the most diverging datapoint at $z \sim 2$, which is appreciated especially when compared to our best-fit quadratic model of the form $b(z) = 1.51 - 0.36z + 0.67z^2$ (Sec. 4.4 & Fig. 5). This overall consistency between our quasar bias fits and those from the previous analyses is especially remarkable taking into account various surveys and sample selections, as well as cosmologies used. We conclude that within the explored redshift ranges and scales the bias measurement for quasars is relatively insensitive to such factors.

As an additional diagnostic, we also examined a full-redshift angular correlation measurement and a fixed-slope power-law Limber inversion in order to compare more directly with broad-redshift analyses such as Petter et al. (2023) and Eltvedt et al. (2024). These checks confirm that such comparisons are difficult to interpret quantitatively for our sample: the redshift distributions, selections, effective redshifts, and available covariance information differ between studies, and the full-redshift angular correlation function is not well described by a single power law. We therefore retain the tomographic bias measurements as our fiducial result.

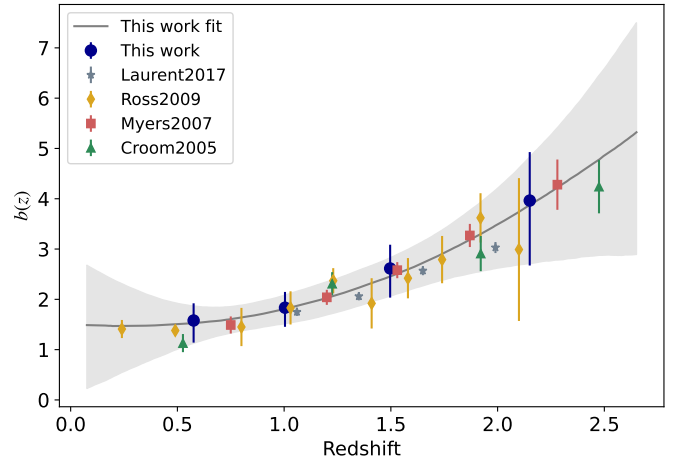


Fig. 6. Comparison of the the effective quasar bias as a function of redshift in this work and previous analyses. The dark blue circles represent the KiDS DR4 measurements from this paper, while the others are respectively: 2QZ from Croom et al. (2005) (upward green triangles); SDSS DR4 from Myers et al. (2007) (red squares); SDSS DR5 from Ross et al. (2009) (orange diamonds); and eBOSS from Laurent et al. (2017) (downward grey triangles). The background grey line with error band is our best-fit quadratic $b(z)$.

5. Conclusions

This study provides the first comprehensive measurement of angular clustering and the scale-independent effective bias for photometrically classified quasars from the Kilo-Degree Survey Data Release 4 (Kuijken et al. 2019). Using the original quasar selection from N21, we restricted our analysis to the safe sample, where the object features lie within the range of the training data of the QSO classification model. Although we used the previous photometrically classified quasar catalog, we did not adopt the photometric redshifts originally derived by N21. Instead, we applied our new DL model, Hybrid-z, which was developed earlier by John William et al. (2025). Hybrid-z combines KiDS ugri imaging with KiDS+VIKING nine-band photometry (ugriZYJHK_s) and was trained on DESI DR1 and SDSS DR17 spectroscopic quasars. In this way we updated the approach of N21, who relied solely on SDSS DR14 and only used colours and magnitudes without incorporating imaging information.

Our new model significantly improves the previous KiDS QSO photo-z derivation in practically all the statistics, including the mean bias and residual scatter. On the other hand, despite the better statistical accuracy and precision of our photo-zs, we still observe the characteristic artefacts in quasar redshift estimates as in previous studies: the redshift focusing arises from the emission line shifts across photometric filters and results narrow peaks in the photometric redshift distributions.

After testing the Hybrid-z model on spectroscopic data with known redshifts, we trained it on the overlap between KiDS DR4 and DESI+SDSS quasars and applied it to the entire photometric sample of safe quasars from N21. The final product was used for the clustering analysis and is an updated KiDS DR4 QSO catalogue containing about 157k objects in an effective area of $\sim 777 \text{ deg}^2$.

We divided our quasar sample into four photometric redshift bins in the range $0.1 \leq z \leq 2.7$, where most of the objects are located. For each of the bins, we measured the quasar angular two-point correlation function (2PCF) over scales from 0.05° to 1° . These measurements were then compared with theoretical predictions for dark-matter 2PCF assuming Planck Collaboration et al. (2020) Λ CDM cosmological parameters. Using a simple bias relation of the form $\omega_Q(\theta) = b^2(z)\omega_m(\theta)$, we obtained best-fit bias values per bin. Our analysis of the quasar clustering exhibits a clear increase in bias with redshift, from $b \sim 1.6$ at $z \sim 0.6$ to $b \sim 4.0$ at $z \sim 2.2$. We calculated the effective peak height ν_{eff} and host halo mass M_{eff} of our quasars. ν_{eff} increases from ~ 1.5 at $z \sim 0.6$ to ~ 2.85 at $z \sim 2.15$. This trend is consistent with halo bias models and previous observational studies, confirming that quasars observed at higher redshifts reside in progressively more massive dark matter haloes.

We assessed the effect of stellar contamination on our quasar clustering measurements by applying the correction formula from Myers et al. (2006) using angular scale-dependent stellar clustering based on stellar samples with varying classification probabilities. Our analysis shows that the inferred quasar bias remains largely unchanged across different stellar purity thresholds, indicating that the results are robust to moderate levels of residual stellar contamination. For the final analysis, we adopted the highest-purity stellar sample ($p_{\text{star}} \geq 0.99$) to correct the quasar angular correlation function, ensuring reliable bias estimates.

We also investigated how our quasar bias derivations depended on the per-bin redshift distributions employed in the the-

oretical model. We used two approaches: taking dN/dz_{phot} for each bin directly, or employing the cross-matches of the bins with spectroscopic quasars from DESI+SDSS as the redshift distributions. These latter dN/dz are broader than the photometric ones due to photo-z errors, and they display tails far away from the bin centres. This is related to catastrophic outliers. Consequently, using dN/dz_{spec} yields systematically higher bias values than dN/dz_{phot} as the amplitudes of the theoretical predictions for dark matter 2PCF are lower in the former case than in the latter. Nevertheless, the spectroscopic distribution is not fully representative of the entire sample, especially because the original quasar selection from N21 that we employed was trained on SDSS DR14. We therefore adopted the bias derived from the photo-z distribution of the full KiDS DR4 quasar sample as our fiducial result. For the future robust clustering-based cosmological inference from photometric quasar samples, such as from the final KiDS DR5 or the forthcoming LSST, an accurate characterisation of the underlying redshift distribution will be crucial.

Our effective quasar bias measurements for the four effective redshifts of the bins can be modelled as $b(z) = b_0 + b_1 z + b_2 z^2$ with best-fit parameters $(b_0, b_1, b_2) = (1.51, -0.36, 0.67)$. A comparison with the previous derivations of quasar $b_Q(z)$, using samples such as 2QZ (Croom et al. 2005), SDSS: photometric (Myers et al. 2007) and spectroscopic (Ross et al. 2009), and eBOSS (Laurent et al. 2017), shows consistency with our results overall, except for the latter work, where the highest- z bias estimates depart from our best-fit by more than 2σ . A caveat regarding these comparison is that these previous works used different cosmologies than we did, which is especially important when σ_8 do not match because they directly rescale the matter power spectrum.

We estimated updated photo-zs for quasars selected from the previous KiDS DR4, and performed the first measurements of KiDS quasar clustering and of their scale-independent effective bias. In the near future, we plan to extend these studies to the final KiDS DR5 (Wright et al. 2024). A quasar selection from that release was already presented in Feng et al. (2025), where, the SDSS data were used to train the classifier. Updating this to DESI is expected to allow us to further improve the previous classification, especially at the faint end. Other avenues to explore it to use all nine KiDS+VIKING passbands in the convolutional part of the Hybrid-z model, that is, they incorporated both imaging and magnitudes from the whole u to K_s range.

Angular 2PCF measurements give high signal-to-noise ratio down to 0.1° (at least), which might be further explored with more sophisticated theoretical frameworks, such as the halo occupation distribution (Mitra et al. 2018; Eftekharzadeh et al. 2019; Petter et al. 2023; Chowdhary & Chatterjee 2025). Furthermore, the quasar bias and other cosmological parameters might also be derived by cross-correlating the quasar distribution with CMB lensing (Hirata et al. 2008; Sherwin et al. 2012; Eltvedt et al. 2024; de Belsunce et al. 2025). A crucial ingredient for the robustness of the cosmological constraints when using photometric quasars will be to properly calibrate their redshift distributions, and techniques such as self-organising maps might be employed for this purpose (Jalan et al. 2024; Wright et al. 2025).

In the longer term, forthcoming large data from the Vera Rubin Observatory Legacy Survey of Space and Time (LSST, Ivezić et al. 2019), as well as from the 4-metre Multi-Object Spectroscopic Telescope (4MOST, de Jong et al. 2019) will allow us to build and explore new quasar samples over most of the sky. This gives great perspectives not only for better under-

standing their large-scale clustering properties, but also to connect these to smaller-scale environments of the cosmic web.

Code availability

The Deep learning framework for the photometric redshift estimation, named Hybrid-z, will be publicly available at <https://hybrid-z.readthedocs.io/en/latest/>.

Acknowledgements. We thank Angus Wright for his valuable comments and suggestions on the manuscript, and Andrej Dvornik and Ziang Yan for their helpful feedback and discussions during the early stages of this project.

This work is supported by the Polish National Science Center through grants no. 2020/38/E/ST9/00395, and 2020/39/B/ST9/03494. We thank Polish Ministry of Science and Higher Education under the agreement no. 2026/WK/06 for the infrastructure support.

Based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under program IDs 177.A-3016, 177.A-3017 and 177.A-3018, and on data products produced by Target/OmegaCEN, INAF-OACN, INAF-OAPD and the KiDS production team, on behalf of the KiDS consortium. OmegaCEN and the KiDS production team acknowledge support by NOVA and NWO-M grants. Members of INAF-OAPD and INAF-OACN also acknowledge the support from the Department of Physics & Astronomy of the University of Padova, and of the Department of Physics of Univ. Federico II (Naples).

We have made use of TOPCAT (Taylor 2005) software, as well as of PYTHON (www.python.org), including the packages NUMPY (Harris et al. 2020), SciPy (Virtanen et al. 2020), COLOSSUS (Diemer 2018), and MATPLOTLIB (Hunter 2007).

References

- Abdurro'uf et al., 2022, *ApJS*, **259**, 35
- Abolfathi B., et al., 2018, *ApJS*, **235**, 42
- Adame A. G., et al., 2025, *J. Cosmology Astropart. Phys.*, **2025**, 012
- Agarwal N., et al., 2014, *J. Cosmology Astropart. Phys.*, **2014**, 007
- Alonso D., Fabbian G., Storey-Fisher K., Eilers A.-C., García-García C., Hogg D. W., Rix H.-W., 2023, *J. Cosmology Astropart. Phys.*, **2023**, 043
- Ata M., et al., 2018, *MNRAS*, **473**, 4773
- Balaguera-Antolínez A., Bilicki M., Branchini E., Postiglione A., 2018, *MNRAS*, **476**, 1050
- Baldry I. K., 2018, *arXiv e-prints*, p. [arXiv:1812.05135](https://arxiv.org/abs/1812.05135)
- Baum W. A., 1957, *AJ*, **62**, 6
- Bilicki M., et al., 2021, *A&A*, **653**, A82
- Blake C., Pope A., Scott D., Mobasher B., 2006, *MNRAS*, **368**, 732
- Blas D., Lesgourgues J., Tram T., 2011, CLASS: Cosmic Linear Anisotropy Solving System, Astrophysics Source Code Library, record ascl:1106.020
- Brescia M., Cavuoti S., D'Abrusco R., Longo G., Mercurio A., 2013, *ApJ*, **772**, 140
- Breton M.-A., de la Torre S., Piat J., 2022, *A&A*, **661**, A154
- Capaccioli M., Schipani P., 2011, *The Messenger*, **146**, 2
- Carrasco D., et al., 2015, *A&A*, **584**, A44
- Chaussidon E., et al., 2023, *ApJ*, **944**, 107
- Chisari N. E., et al., 2019, *ApJS*, **242**, 2
- Chowdhary A., Chatterjee S., 2025, *ApJ*, **992**, 21
- Comparat J., Prada F., Yepes G., Klypin A., 2017, *MNRAS*, **469**, 4157
- Croom S. M., Smith R. J., Boyle B. J., Shanks T., Miller L., Outram P. J., Loaring N. S., 2004, *MNRAS*, **349**, 1397
- Croom S. M., et al., 2005, *MNRAS*, **356**, 415
- Curran S. J., 2022, *MNRAS*, **512**, 2099
- DESI Collaboration et al., 2026, *AJ*, **171**, 285
- D'Isanto A., Polsterer K. L., 2018, *A&A*, **609**, A111
- DiPompeo M. A., Hickox R. C., Myers A. D., 2016, *MNRAS*, **456**, 924
- Diemer B., 2018, *ApJS*, **239**, 35
- Edge A., Sutherland W., Kuijken K., Driver S., McMahon R., Eales S., Emerson J. P., 2013, *The Messenger*, **154**, 32
- Eftekharzadeh S., Myers A. D., Kourkchi E., 2019, *MNRAS*, **486**, 274
- Eltvedt A. M., Shanks T., Metcalfe N., Ansarinejad B., Barrientos L. F., Murphy D. N. A., Alexander D. M., 2024, *MNRAS*, **535**, 2105
- Fan X., Bañados E., Simcoe R. A., 2023, *ARA&A*, **61**, 373
- Feng H.-C., et al., 2025, *ApJS*, **279**, 26
- Fumagalli A., Biagetti M., Saro A., Sefusatti E., Slosar A., Monaco P., Veropalumbo A., 2022, *J. Cosmology Astropart. Phys.*, **2022**, 022
- Giner Mascarell M., Eilers A.-C., Storey-Fisher K., 2025, *arXiv e-prints*, p. [arXiv:2511.17413](https://arxiv.org/abs/2511.17413)
- Harris C. R., et al., 2020, *Nature*, **585**, 357
- Hartlap J., Simon P., Schneider P., 2007, *A&A*, **464**, 399
- Hirata C. M., Ho S., Padmanabhan N., Seljak U., Bahcall N. A., 2008, *Phys. Rev. D*, **78**, 043520
- Ho S., et al., 2015, *J. Cosmology Astropart. Phys.*, **2015**, 040
- Huber P. J., 1992, Robust Estimation of a Location Parameter. Springer New York, New York, NY, pp 492–518, doi:10.1007/978-1-4612-4380-9_35, https://doi.org/10.1007/978-1-4612-4380-9_35
- Hunter J. D., 2007, *Computing in Science and Engineering*, **9**, 90
- Ivezić Ž., et al., 2019, *ApJ*, **873**, 111
- Jalan P., et al., 2024, *A&A*, **692**, A177
- Jarvis M., Bernstein G., Jain B., 2004, *MNRAS*, **352**, 338
- John William A., Jalan P., Bilicki M., Hellwing W. A., Thuruthipilly H., Nakoneczny S. J., 2025, *A&A*, **698**, A276
- Johnston H., et al., 2021, *A&A*, **648**, A98
- Kaufman C., Schervish M., Nychka D., 2008, *Journal of the American Statistical Association*, **103**, 1545
- Koo D. C., 1985, *AJ*, **90**, 418
- Kuijken K., 2008, *A&A*, **482**, 1053
- Kuijken K., et al., 2019, *A&A*, **625**, A2
- Kunsági-Máté S., Beck R., Szapudi I., Csabai I., 2022, *MNRAS*, **516**, 2662
- Landy S. D., Szalay A. S., 1993, *ApJ*, **412**, 64
- Laurent P., et al., 2017, *J. Cosmology Astropart. Phys.*, **2017**, 017
- Lecun Y., Bengio Y., 1995, Convolutional networks for images, speech, and time-series. MIT Press
- Luo W., et al., 2024, *ApJ*, **977**, 59
- Lyke B. W., et al., 2020, *ApJS*, **250**, 8
- Maartens R., Fonseca J., Camera S., Jolicoeur S., Viljoen J.-A., Clarkson C., 2021, *J. Cosmology Astropart. Phys.*, **2021**, 009
- Marziani P., Dultzin-Hacyan D., Sulentic J. W., 2006, in Kreitler P. V., ed., , New Developments in Black Hole Research. p. 123, doi:10.48550/arXiv:astro-ph/0606678
- McCulloch W. S., 1943, Bulletin of mathematical biophysics, **5**, 115
- Ménard B., Bartelmann M., 2002, *A&A*, **386**, 784
- Mitra K., Chatterjee S., DiPompeo M. A., Myers A. D., Zheng Z., 2018, *MNRAS*, **477**, 45
- Mohammad F. G., Percival W. J., 2022, *MNRAS*, **514**, 1289
- Moss J., Curran S., Perrott Y., 2025, *PASA*, **42**, e113
- Myers A. D., et al., 2006, *ApJ*, **638**, 622
- Myers A. D., Brunner R. J., Nichol R. C., Richards G. T., Schneider D. P., Bahcall N. A., 2007, *ApJ*, **658**, 85
- Nakazono L., et al., 2024, *MNRAS*, **531**, 327
- Nakoneczny S., Bilicki M., Solarz A., Pollo A., Maddox N., Spiniello C., Brescia M., Napolitano N. R., 2019, *A&A*, **624**, A13
- Nakoneczny S. J., et al., 2021, *A&A*, **649**, A81
- Newman J. A., Gruen D., 2022, *ARA&A*, **60**, 363
- Pasquet-Itam J., Pasquet J., 2018, *A&A*, **611**, A97
- Peebles P. J. E., 1973, *ApJ*, **185**, 413
- Peebles P. J. E., 1980, The large-scale structure of the universe
- Peterson B. M., 1997, An Introduction to Active Galactic Nuclei
- Petter G. C., Hickox R. C., Alexander D. M., Myers A. D., Geach J. E., Whalen K. E., Andonie C. P., 2023, *ApJ*, **946**, 27
- Piccirilli G., Fabbian G., Alonso D., Storey-Fisher K., Carron J., Lewis A., García-García C., 2024, *J. Cosmology Astropart. Phys.*, **2024**, 012
- Pizzati E., et al., 2024, *MNRAS*, **534**, 3155
- Planck Collaboration et al., 2020, *A&A*, **641**, A6
- Prechelt L., 2012, Early Stopping — But When?. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 53–67, doi:10.1007/978-3-642-35289-8_5, https://doi.org/10.1007/978-3-642-35289-8_5
- Prochaska J. X., Hennawi J. F., 2009, *ApJ*, **690**, 1558
- Rees M. J., 1984, *ARA&A*, **22**, 471
- Richards G. T., et al., 2002, *AJ*, **123**, 2945
- Rosenblatt F., 1958, *Psychological Review*, **65**, 386
- Ross N. P., et al., 2009, *ApJ*, **697**, 1634
- Roster W., et al., 2024, *A&A*, **692**, A260
- Salpeter E. E., 1964, *ApJ*, **140**, 796
- Salvato M., Ilbert O., Hoyle B., 2019, *Nature Astronomy*, **3**, 212
- Schmidt M., 1963, *Nature*, **197**, 1040
- Shen Y., et al., 2009, *ApJ*, **697**, 1656
- Sherwin B. D., et al., 2012, *Phys. Rev. D*, **86**, 083006
- Slosar A., Hirata C., Seljak U., Ho S., Padmanabhan N., 2008, *J. Cosmology Astropart. Phys.*, **2008**, 031
- Smith R. E., et al., 2003, *MNRAS*, **341**, 1311
- Song H., Park C., Lietzen H., Einasto M., 2016, *ApJ*, **827**, 104
- Spilker J. S., Champagne J. B., Fan X., Fujimoto S., van der Werf P. P., Yang J., Yue M., 2025, *ApJ*, **982**, 72
- Storey-Fisher K., Hogg D. W., Rix H.-W., Eilers A.-C., Fabbian G., Blanton M. R., Alonso D., 2024, *ApJ*, **964**, 69
- Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, *ApJ*, **761**, 152
- Taylor M. B., 2005, in Shopbell P., Britton M., Ebert R., eds, Astronomical Society of the Pacific Conference Series Vol. 347, Astronomical Data Analysis Software and Systems XIV. p. 29

- Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, [ApJ](#), 724, 878
- Vakili M., et al., 2023, [A&A](#), 675, A202
- Virtanen P., et al., 2020, [Nature Methods](#), 17, 261
- Wright A. H., et al., 2024, [A&A](#), 686, A170
- Wright A. H., et al., 2025, [A&A](#), 703, A144
- Yan Z., et al., 2025, [A&A](#), 694, A259
- Yang S., Xiao W., Zhang M., Guo S., Zhao J., Shen F., 2022, arXiv preprint arXiv:2204.08610
- Yao L., et al., 2023, [MNRAS](#), 523, 5799
- Zel'dovich Y. B., Novikov I. D., 1964, [Soviet Physics Doklady](#), 9, 246
- de Belsunce R., et al., 2025, [J. Cosmology Astropart. Phys.](#), 2025, 077
- de Jong J. T. A., et al., 2015, [A&A](#), 582, A62
- de Jong R. S., et al., 2019, [The Messenger](#), 175, 3

Appendix A: Measured bias values

Table A.1. Comparison of effective redshifts (z_{eff}) and bias values obtained using photo- z distributions of KiDS DR4 quasar sample (dN/dz_{phot}) and spec- z distributions of KiDS DR4 $\times$ (DESI DR1 + SDSS DR17) quasar data (dN/dz_{spec}) for the theoretical prediction of angular clustering.

photo- z bin	using dN/dz_{phot}		using dN/dz_{spec}	
	z_{eff}	bias	z_{eff}	bias
$0.1 \leq z_{\text{phot}} \leq 0.8$	0.58	$1.58^{+0.34}_{-0.44}$	0.70	$2.24^{+0.48}_{-0.62}$
$0.8 < z_{\text{phot}} \leq 1.2$	1.00	$1.83^{+0.31}_{-0.38}$	1.08	$2.52^{+0.43}_{-0.53}$
$1.2 < z_{\text{phot}} \leq 1.8$	1.50	$2.61^{+0.47}_{-0.58}$	1.50	$3.26^{+0.59}_{-0.72}$
$1.8 < z_{\text{phot}} \leq 2.7$	2.15	$3.96^{+0.96}_{-1.29}$	2.10	$5.22^{+1.27}_{-1.69}$

Appendix B: Angular clustering of stars

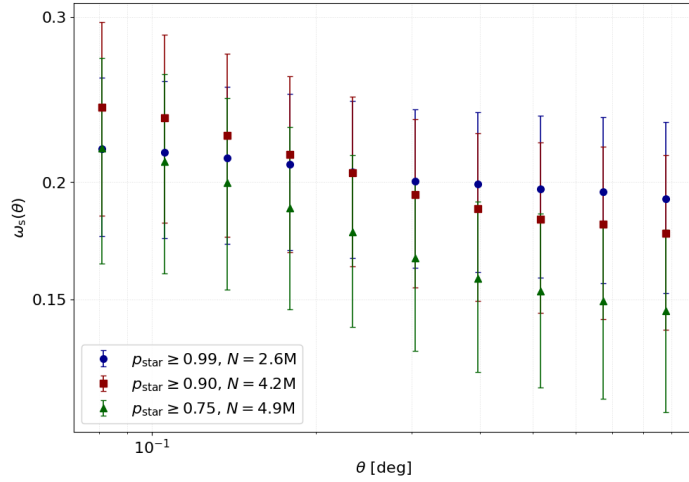


Fig. B.1. Angular two-point correlation function for stellar sources, $\omega_s(\theta)$, measured from 0.05 to 1.0 deg selected by different stellar classification probability (p_{star}) threshold. The corresponding number of sources N is indicated in the legend.

Appendix C: Best-fit parameters

Best-fit parameters and corresponding covariance matrix for the quadratic fit, $b_0 + b_1z + b_2z^2$, describing the evolution of quasar bias with redshift as follows. Covariance matrix of best-fit parameters when using dN/dz_{phot} is:

$$C_{ij} = \begin{bmatrix} 2.06 & -3.65 & 1.41 \\ -3.65 & 6.84 & -2.75 \\ 1.41 & -2.75 & 1.16 \end{bmatrix}$$

and using dN/dz_{spec} is:

$$C_{ij} = \begin{bmatrix} 7.38 & -12.16 & 4.55 \\ -12.16 & 20.9 & -8.1 \\ 4.55 & -8.1 & 3.26 \end{bmatrix}$$

Appendix D: Comparison of clustering measurements with previous studies

Table C.1. The parameter uncertainties, corresponding to 1σ deviation, are derived from the square root of the diagonal elements of their covariance matrix.

Parameters	Using dN/dz_{phot}	Using dN/dz_{spec}
b_0	1.51 ± 1.44	2.71 ± 2.72
b_1	-0.36 ± 2.62	-1.62 ± 4.57
b_2	0.67 ± 1.08	1.31 ± 1.81

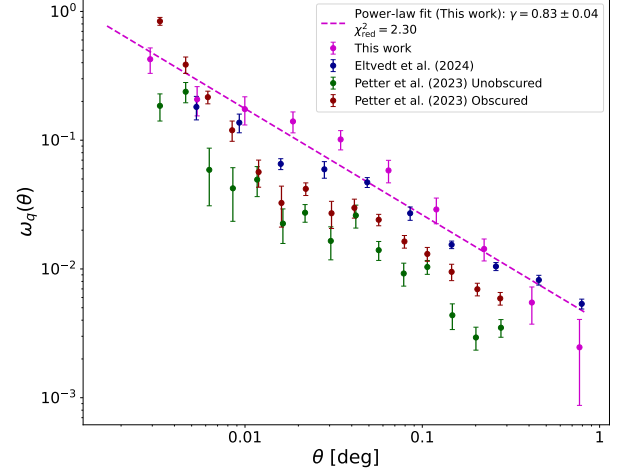


Fig. D.1. Comparison of the quasar angular auto correlation function, $\omega_q(\theta)$, we measured in this work (magenta circles) with the [Eltvedt et al. \(2024\)](#) measurements (blue circles), and the obscured (dark red circles) and unobscured (green circles) quasar samples from [Petter et al. \(2023\)](#). The solid magenta line represents the best-fit power-law model, $\omega_q(\theta) = A\theta^{-\delta}$, over the range $0.002 < \theta < 1.0$ deg. The fit yields a slope $\delta = 0.83 \pm 0.04$ with a reduced χ^2 of 2.30.

We compare our measurements of the angular correlation function with those of [Petter et al. \(2023\)](#) and [Eltvedt et al. \(2024\)](#), who used a similar angular correlation technique. Since our main analysis is performed in tomographic redshift bins and over a different angular range (see Sec.4), we construct an additional quasar sample spanning the full redshift range $0.1 \leq z \leq 2.7$ to enable a direct comparison with these works. For this purpose, we measure ω_q over angular scales $[0.002, 1.0]$ deg, matching as closely as possible the setup of previous studies. The comparison is shown in Fig. D.1. The differences in amplitude may arise from differences in quasar sample selection, redshift distributions, and the treatment of systematic effects.

In addition, we fit a power-law model,

$$\omega_q(\theta) = A\theta^{-\gamma} \quad (\text{D.1})$$

Over the angular range $0.002 < \theta < 1.0$ deg, we obtain a best-fitting slope of $\gamma = 0.83 \pm 0.04$ with a reduced chi-square of $\chi^2_{\text{red}} = 2.3$. The measured slope is fully consistent with the canonical value of $\gamma \approx 0.8$ commonly found for galaxy angular correlation functions ([Peebles 1980](#)). The large reduced chi-square value indicates that a single power law is not a fully adequate description of our angular clustering measurements.