

# Clustering Measurements of broad-line AGNs: Review and Future

Mirko Krumpel<sup>1,3</sup>, Takamitsu Miyaji<sup>2,3</sup>, and Alison L. Coil<sup>3</sup>

<sup>1</sup>*European Southern Observatory, Karl-Schwarzschild-Straße 2, 85748 Garching bei München, Germany*

<sup>2</sup>*Instituto de Astronomía, UNAM, Apdo. Postal 106, Ensenada, BC, México*

<sup>3</sup>*University of California San Diego, CASS, 9500 Gilman Drive, La Jolla, CA 92093-0424, USA*

Corresponding author: [mkrumpe@eso.org](mailto:mkrumpe@eso.org)

## Abstract

Despite substantial effort, the precise physical processes that lead to the growth of super-massive black holes in the centers of galaxies are still not well understood. These phases of black hole growth are thought to be of key importance in understanding galaxy evolution. Forthcoming missions such as eROSITA, HETDEX, eBOSS, BigBOSS, LSST, and Pan-STARRS will compile by far the largest ever Active Galactic Nuclei (AGNs) catalogs which will allow us to measure the spatial distribution of AGNs in the universe with unprecedented accuracy. For the first time, AGN clustering measurements will reach a level of precision that will not only allow for an alternative approach to answering open questions in AGN and galaxy co-evolution but will open a new frontier, allowing us to precisely determine cosmological parameters. This paper reviews large-scale clustering measurements of broad line AGNs. We summarize how clustering is measured and which constraints can be derived from AGN clustering measurements, we discuss recent developments, and we briefly describe future projects that will deliver extremely large AGN samples which will enable AGN clustering measurements of unprecedented accuracy. In order to maximize the scientific return on the research fields of AGN and galaxy evolution and cosmology, we advise that the community develops a full understanding of the systematic uncertainties which will, in contrast to today’s measurement, be the dominant source of uncertainty.

**Keywords:** large-scale structure of universe - galaxies: active.

## 1 Introduction

Large area surveys such as the Two Degree Field Galaxy Redshift Survey (2dFGRS; Colless et al. 2001) and the Sloan Digital Sky Survey (SDSS; Abazajian et al. 2009) have measured positions and redshifts of millions of galaxies. These measurements allow us to map the 3D structure of the nearby universe<sup>1</sup>.

Galaxies are not randomly distributed in space. They form a complex cosmic network of galaxy clusters, groups, filaments, isolated field galaxies, and voids, which are large regions of space that are almost devoid of galaxies. The current understanding of the distribution of galaxies and structure formation in the universe is based on the theory of gravitational instability. Very early density fluctuations became the “seeds” of cosmic structure. These have been observed as small temperature fluctuations ( $\delta T/T \sim 5 \times 10^{-5}$ ) in the cosmic microwave background with the Cosmic Background Explorer (Smoot et al. 1992). The small primordial matter density enhancements have progressively grown through gravitational collapse and created the complex network seen in the distribution of matter in the later universe.

During a galaxy’s lifetime different physical processes, which are still not well understood, can trigger a mass flow onto the central super-massive black hole (SMBH). In this phase of galaxy evolution, the galaxy is observed as an Active Galactic Nucleus (AGN). After several million years, when the SMBH has consumed its accretion reservoir, the central engine shuts down, and the object is again observed as a normal galaxy. The AGN phase is thought to be a repeating special epoch in the process of galaxy evolution. In recent years it has become evident that both fundamental galaxy and AGN parameters change significantly between low ( $z < 0.3$ ) and intermediate redshifts ( $z \sim 1 - 2$ ), e.g., global star formation density (Hopkins & Beacom 2006) and accretion rate onto SMBHs. For example, the contribution to black hole growth has shifted from high luminosity objects at high redshifts to low luminosity objects at low redshifts (AGN “downsizing”; e.g., Hasinger et al. 2005). It has also become clear that SMBH masses follow a tight relation with the mass or velocity dispersion of the stars in galactic bulges (Magorrian et al. 1998; Gebhardt et al. 2000; Ferrarese & Merritt 2000). These observational correlations moti-

<sup>1</sup>A visual impression is given in this video: <http://vimeo.com/4169279>

vate a co-evolution scenario for galaxies and AGNs and provide evidence of a possible interaction or feedback mechanism between the SMBH and the host galaxy. The interpretation of this correlation, i.e., whether and to what extent the AGN influences its host galaxy, remains controversial (e.g., Jahnke & Macció 2011).

Since AGNs are generally much brighter than (inactive) galaxies, one major advantage of AGN large-scale (i.e., larger than the size of a galaxy) clustering measurements over galaxy clustering measurements is that they allow the study of the matter distribution in the universe out to higher redshifts. At these redshifts, it becomes challenging and observationally expensive to detect galaxies in sufficient numbers. Furthermore, as the distribution of AGNs and galaxies in the universe depends on galaxy evolution physics, large-scale clustering measurements are an independent method to identify and constrain the physical processes that turn an inactive galaxy into an AGN and are responsible for AGN and galaxy co-evolution.

In the last decade the scientific interest in AGN large-scale clustering measurements has increased significantly. As only a very small fraction of galaxies contain an AGN ( $\sim 1\%$ ), the remaining and dominating challenge in deriving physical constraints based on AGN clustering measurements is the relative small sample size compared to galaxy clustering measurements. However, this situation will change entirely in the next decade when several different surveys come online that are expected to identify millions of AGN over  $\sim 80\%$  of cosmic time.

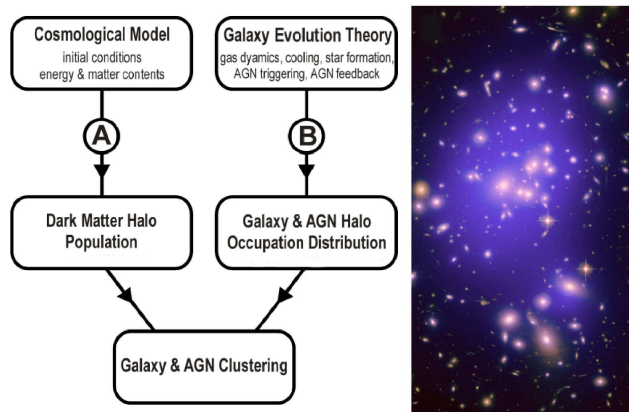
We therefore review broad-line AGN clustering measurements. A general introduction to clustering measurements is given in Sections 2 & 3. In Section 4 we briefly summarize how AGN clustering measurements have evolved and discuss recent developments. In Section 5 we discuss the outlook for AGN clustering measurements in future upcoming projects.

## 2 Understanding Observed Clustering Properties

In our current understanding, the observed galaxy and AGN spatial distribution in the universe – i.e., large-scale clustering – is caused by the interplay between cosmology and the physics of galaxy evolution.

In the commonly assumed standard cosmological model, Lambda-CDM, the universe is currently composed of  $\sim 70\%$  dark energy,  $\sim 25\%$  dark matter (DM), and  $\sim 5\%$  baryonic matter (Larsen et al. 2011). Dark matter plays a key role in structure formation as it is the dominant form of matter in the universe. Baryonic matter settles in the deep gravitational potentials created by dark matter, the so-called dark matter halos (DMHs). The term “halo” commonly refers to a

bound, gravitationally collapsed dark matter structure which is approximately in dynamical equilibrium. The parameters of the cosmological model determine how the DMHs are distributed in space (Fig. 1, left panel, A-branch) as a function of the DMH mass and cosmic time. Different cosmological models lead to different properties of the DMH population.



**Figure 1:** Current conceptual model of the physical processes involved in large-scale galaxy and AGN clustering. *Left:* The two branches (A and B) in the diagram show the primary causes of clustering: (A) the properties of the dark matter halo population, which are based on the cosmological model, and (B) the physics of complex processes in galaxy formation and evolution, which lead to a distinct baryonic population within collapsed dark matter halos. Figure adapted from Weinberg 2002. *Right:* Illustration of the spatial distribution of galaxies within a dark matter halo. The picture maps the galaxy cluster Abell 1689, where an optical image showing the galaxy cluster members is superimposed with the distribution of dark matter shown in purple. Credit: NASA, ESA, E. Jullo, P. Natarajan, and J-P. Kneib.

Inside DMHs, or within halos inside another DMH, called sub-halos, the baryonic gas will radiatively cool. If the gas reservoir is large enough, star and galaxy formation will be initiated. The gas can also be accreted onto the SMBH in the center of the galaxy. On scales comparable to the size of the galaxy, the AGN can heat and/or eject the surrounding gas, preventing star formation, and eventually removing the gas fueling the AGN itself. All the galaxy evolution processes described here determine how galaxies and AGNs are distributed within DMHs (Fig. 1, left panel, B-branch). This distribution of AGN and galaxies within DMHs (Fig. 1, right panel) is described by the halo occupation distribution (HOD; Peacock & Smith 2000). In addition to the spatial distribution of AGN and galaxies in DMHs, the HOD describes the probability distributions

of the number of AGNs and galaxies per DMH of a certain mass and the velocity distribution of AGNs and galaxies within a DMH.

The interplay between cosmology and galaxy evolution causes the observed large-scale clustering of galaxies and AGNs. The goal of AGN and galaxy clustering measurements is to reverse the causal arrows in the Fig. 1 (left panel), working backwards from the data to the galaxy & AGN halo occupation distribution and DMH population properties, in order to finally draw conclusions about galaxy and AGN physics, as well as to constrain fundamental cosmological parameters.

### 3 Clustering Measurements

The most common statistical estimator for large-scale clustering is the two-point correlation function (2PCF; Peebles 1980)  $\xi(r)$ . This quantity measures the spatial clustering of a class of object *in excess* of a Poisson distribution. In practice,  $\xi(r)$  is obtained by counting pairs of objects with a given separation and comparing them to the number of pairs in a random sample with the same separation. Different correlation estimators are described in the literature (e.g., Davis & Peebles 1983; Landy & Szalay 1993).

The large-scale clustering of a given class of objects can be quantified by computing the angular (2D) correlation function, which is the projection onto the plane of the sky, or with the spatial (3D) correlation function, which requires redshift information for each object. Obtaining spectra to measure the 3D correlation function is observationally expensive, which is the main reason why some studies have had to rely on angular correlation functions. However, 3D correlation function measurements are by far preferable, since the deprojection (Limber 1954) of the angular correlation function introduces large systematic uncertainties. Despite these large caveats and the already moderately low uncertainties of current 3D correlation measurements, the use of angular correlation functions might still be justified when exploring a new parameter space. However, the next generation multi-object spectrographs (e.g., 4MOST (de Jong et al. 2012), BigBOSS (Schlegel et al. 2011), and WEAVE (Dalton et al. 2012)), will make it far easier to simultaneously obtain thousands of spectra over wide fields. Hence, measurements of the 3D correlation function will soon become ubiquitous.

As one measures line-of-sight distances for 3D correlation functions from redshifts, measurements of  $\xi(r)$  are affected by redshift-space distortions due to peculiar velocities of the objects within DMHs. To remove this effect,  $\xi(r)$  is commonly extracted by counting pairs on a 2D grid of separations where  $r_p$  is perpendicular to the line of sight and  $\pi$  is along the line of sight. Then, integrating along the  $\pi$ -direction leads to the projected

correlation function,  $w_p(r_p)$ , which is free of redshift distortions. The 3D correlation function  $\xi(r)$  can be recovered from the projected correlation function (Davis & Peebles 1983).

The resulting signal can be approximated by a power law where the largest clustering strength is found at small scales. At large separations of  $>50 \text{ Mpc } h^{-1}$  the distribution of objects in the universe becomes nearly indistinguishable from a randomly-distributed sample. Only on comoving scales of  $\sim 100 \text{ Mpc } h^{-1}$  can a weak positive signal be detected (e.g., Eisenstein et al. 2005; Cole et al. 2005) which is caused by baryonic acoustic oscillations (BAO) in the early universe.

The spatial clustering of observable objects does not precisely mirror the clustering of matter in the universe. In general, the large-scale density distribution of an object class is a function of the underlying dark matter density. This relation of how an object class traces the underlying dark matter density is quantified using the linear bias parameter  $b$ . This contrast enhancement factor is the ratio of the mean overdensity of the observable object class, the so-called tracer set, to the mean overdensity of the dark matter field, defined as  $b = (\delta\rho/\langle\rho\rangle)_{\text{tracer}}/(\delta\rho/\langle\rho\rangle)_{\text{DM}}$ , where  $\delta\rho = \rho - \langle\rho\rangle$ ,  $\rho$  is the local mass density, and  $\langle\rho\rangle$  is the mean mass density on that scale. In terms of the correlation function, the bias parameter is defined as the square root of the 2PCF ratio of the tracer set to the dark matter field:  $b = \sqrt{\xi_{\text{tracer}}/\xi_{\text{DM}}}$ . Rare objects which form only in the highest density peaks of the mass distribution have a large bias parameter and consequently a large clustering strength.

Theoretical studies of DMHs (e.g., Mo & White 1996; Sheth et al. 2001) have established a solid understanding of the bias parameter of DMHs with respect to various parameters. Comparing the bias parameter of an object class with that of DMHs in a certain mass range at the same cosmological epoch allows one to determine the DMH mass which hosts the object class of interest. A halo may contain substructures, but the DMH mass inferred from the linear bias parameter refers to the single, largest (parent) halo in the context of HOD models.

#### 3.1 Why are we interested in AGN clustering?

AGN clustering measurements explore different physics on different scales. At scales up to the typical size of a DMH ( $\sim 1 - 2 \text{ Mpc}$ ), clustering measurements are sensitive to the physics of galaxy and AGN formation and evolution. Constraints on the galaxy and AGN merger rate and the radial distribution of these objects within DMHs can be derived. On scales larger than the size of DMHs, the large-scale clustering is sensi-

tive to the underlying DM density field, which essentially depends only on cosmological parameters. Consequently, with only one measurement both galaxy and AGN co-evolution as well as cosmological parameters can be studied.

Future high precision AGN clustering measurements have the potential to accurately establish missing fundamental parameters that describe when AGN activity and feedback occur as a function of luminosity and redshift. Since they will precisely determine how DMHs are populated by AGN host galaxies, these measurements will also improve our theoretical understanding of galaxy and AGN evolution by enabling comparisons to galaxy measurements and cosmological simulations. Here, we elaborate on some (though not all) of the critical observational constraints which are provided by AGN clustering measurements:

- *AGN host galaxy* – AGN cannot be more clustered than the type of galaxies they reside in. Thus, AGN clustering measurements determine the host galaxy type in a statistical sense for the entire AGN sample, regardless of the AGN’s luminosity. Comparing the observed AGN clustering to very accurate galaxy clustering measurements, which depend on different galaxy subclasses (morphological, spectral type, luminosity), constrains the AGN host galaxy type.
- *External (mergers) vs. internal triggering* – Different theoretical models (e.g., Fry 1996; Sheth et al. 2001; Shen 2009) of how AGNs are triggered predict very different large-scale clustering properties with AGN parameters such as luminosity and redshift. Moderately precise AGN clustering measurements allow us to distinguish between these different models (Allevato et al. 2011). Furthermore, the validity of different models can be tested for different luminosities and cosmological epochs.
- *Fundamental galaxy/AGN physics* – AGN large-scale clustering dependences with various AGN properties could potentially be a key source of independent constraints on galaxy/AGN physics. Comparing the observed AGN clustering properties with results from simulations with different inputs for galaxy/AGN physics could identify the physics that links the evolution of AGNs and galaxies.
- *AGN lifetimes* – AGN clustering measurements allow us to estimate the AGN lifetime at different cosmological epochs (Martini & Weinberg 2001). The underlying idea is that rare, massive DMHs are highly biased tracers of the underlying mass distribution, while more common objects are less strongly biased (Kaiser 1984). Therefore, if AGNs

are heavily biased they must be in rare, massive DMHs. The ratio of the AGN number density to the host halo number density is a measure of the “duty cycle”, i.e., the fraction of the time that the object spends in the AGN phase.

- *Cosmological parameters* – As AGN clustering measurements extend to much higher redshifts than galaxy clustering measurements, they can be used to derive constraints on cosmological parameters (e.g., Basilakos & Plionis 2009) back to the time of the formation of the first AGNs. Currently, the detection of the BAO imprint on clustering measurements at different cosmological epochs is of great interest to constrain the equation of state of dark energy. AGN large-scale clustering measurements with very large AGN samples can detect the BAO signal in a redshift range that is not accessible with galaxy clustering measurements.

## 4 AGN Clustering Measurements: Past and Present

Until the 1980s, studies had to primarily rely on small, optically-selected, very luminous AGN samples for clustering measurements. Then the main question was whether AGNs are randomly distributed in the universe (e.g., Bolton et al. 1976; Setti & Woltjer 1977). The extremely small sample sizes did not allow clustering measurements at scales below  $\sim 50$  Mpc, where a significant deviation from a random distribution is present. Thanks to the launch of major X-ray missions in the 1980s and 1990s such as Einstein (Giacconi et al. 1979) and ROSAT (Truemper 1993), much larger AGN samples enabled successful detections of the AGN large-scale clustering signal. A detailed review on the history of X-ray AGN clustering measurements is given in Cappelluti et al. (2012).

Although AGN clustering measurements are far from being as precise as galaxy clustering measurements, some general findings have emerged in recent years. Interestingly, over all of studied cosmic time ( $z \sim 0 - 3$ ), broad-line AGNs occupy DMH masses of  $\log(M_{\text{DMH}}/[h^{-1}M_{\odot}]) \sim 12.0 - 13.5$  and therefore cluster like groups of galaxies. More detailed information about the current picture of broad-line AGN clustering is presented in Section 6.6 of Krumpke et al. (2012).

Some puzzling questions remain. For example, at  $z < 0.5$  a weak X-ray luminosity dependence on the clustering strength is found (in that luminous X-ray AGNs cluster more strongly than their low luminosity counterparts, e.g., Krumpke et al. 2010; Cappelluti et al. 2010; Shen et al. 2013). However, at high redshift it seems that high luminosity, optically-selected AGNs cluster less strongly than moderately-luminous

X-ray selected AGNs. Whether this finding is due to differences in the AGN populations, an intrinsic luminosity dependence to the clustering amplitude, or an observational bias is yet not understood.

We note that different studies have used different relations to translate the measured linear bias parameter to DMH mass, as well as different  $\sigma_8$  values. Therefore, instead of blindly comparing the derived DMH mass, re-calculating the masses based on the same linear bias to DMH mass relation and the same  $\sigma_8$  is essential when comparing measurements in the literature.

#### 4.1 Recent Developments

In the last few years several new approaches have been used to improve the precision of AGN clustering measurements or their interpretation. We summarize these developments below.

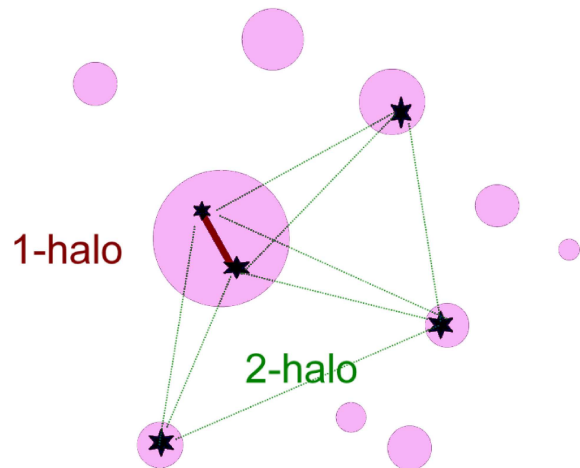
##### *Cross-correlation measurements:*

Auto-correlation function (ACF) measurements of broad-line AGNs often have large uncertainties due to the low number of objects. Especially at low redshifts, large galaxy samples with spectroscopic redshifts are frequently available. In such cases, the statistical uncertainties of AGN clustering measurements can be reduced significantly by computing the cross-correlation function (CCF). The CCF measures the clustering of objects between two different object classes (e.g., broad-line AGNs and galaxies), while the ACF measures the spatial clustering of objects in the same sample (e.g., galaxies or AGNs). CCFs have been used before to study the dependence of the AGN clustering signal with different AGN parameters. However, these values could not be compared to other studies as the CCFs also depend on the galaxy populations used and their clustering strength. Only recently has an alternative approach (Coil et al. 2009) allowed the comparison of the results from different studies by inferring the AGN ACF from the measured CCF and ACF of the galaxy tracer set. The basic idea of this method, which is now frequently used (e.g., Krumpel et al. 2010, 2012; Mountrichas & Georgakakis 2012; Shen et al. 2013), is that both populations trace the same underlying DM density field.

##### *Photometric redshift samples:*

Large galaxy tracer sets with spectroscopic redshifts are not available at all redshifts. Some studies therefore rely on photometric redshifts. The impact of the large uncertainties and catastrophic outliers when using photometric redshifts is commonly not considered but it can be essential. The use of the full probability density function (PDF) of the photometric redshift fit, instead of a single value for the photometric redshift, has been used in some studies (e.g., Mountrichas et al. 2013). Here, photometric galaxies samples are used as tracer sets to derive the CCF between AGN and galaxies. Each ob-

ject is given a weight for the probability that it is actually located at a certain redshift based on the fit to the photometric data.



**Figure 2:** In the conceptual model of the HOD approach, there are two contributions to the pairs that account for the measured correlation function. Pairs of objects (black stars) can either be located within the same DMH (pink filled circles), such that their measured separation contributes to the 1-halo term (red solid line in the large DMH), or can reside in different DMHs, such that their separations (green dotted line) contribute to the 2-halo term.

##### *AGN Halo Occupation Distribution Modeling:*

Instead of deriving only mean DMH masses from the linear bias parameter, HOD modeling of the correlation function allows the determination of the full distribution of AGN as a function of DMH mass. The derived distribution also connects observations and simulations as it provides recipes for how to populate DMHs with observable objects.

In the HOD approach, the measured 2PCF is modeled as the sum of contributions from pairs within individual DMHs (Fig. 2; 1-halo term) and in different DMHs (2-halo term). The superposition of both components describes the shape of the observed 2PCF better than a simple power law. In the HOD description, a DMH can be populated by one central AGN or galaxy and by additional objects in the same DMH, so-called satellite AGN and galaxies. Applying the HOD approach to the 2PCF allows one to determine, e.g., the minimum DMH needed to host the object class of interest, the fraction of objects in satellites, and the number of satellites as a function of DMH mass. Instead of using the derived AGN ACF from CCF measurements, Miyaji et al. (2011) utilize the HOD model directly on high precision AGN vs. galaxy CCF and achieve additional constraints on the AGN and galaxy co-evolution and AGN physics.



### Theoretical predictions:

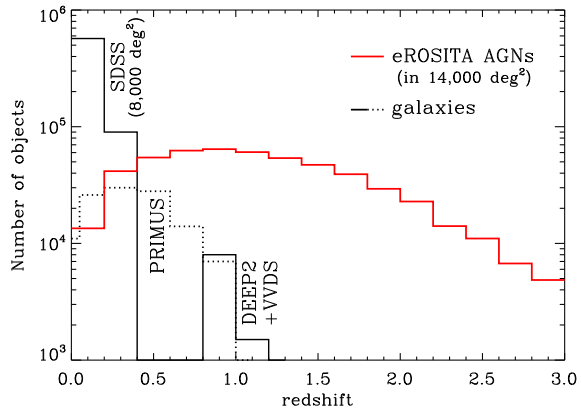
Only recently have several different theoretical models been published which try to explain the observed AGN clustering with different physical approaches (e.g. Fanidakis et al. 2013; Hütsi et al. 2014). The key to observationally distinguish between these models are their different predictions for the clustering dependences of different AGN parameters. In addition to theoretical models of the observed clustering, other very recently developed models predict the halo occupation distribution of AGNs at different redshifts, e.g., Chatterjee et al. (2012). The major challenge presented by all of these models is the urgent need for observational constraints with higher precision than can be provided with current AGN samples. In the future, progress in AGN physics and AGN and galaxy evolution will be achieved through a close interaction between state-of-the-art cosmological simulations and observational constraints from high precision clustering measurements. Simulations which incorporate different physical processes will lead to different predictions of the AGN and galaxy large-scale clustering trends and their halo occupation distributions. Observational studies will then identify the correct model and consequently the actual underlying physical processes.

## 5 The future of AGN clustering measurements

AGN clustering measurements from several upcoming projects will significantly extend our knowledge of the growth of cosmic structure and will also provide a promising avenue towards new discoveries in the fields of galaxy and AGN co-evolution, AGN triggering, and cosmology. For example, eROSITA (Predehl et al. 2010; launch 2015/2016) will perform several all-sky X-ray surveys. After four years the combined survey is expected to contain approximately three million AGNs. HETDEX (Hill et al. 2008; start 2015) will use an array of integral-field spectrographs to provide a total sample of  $\sim 20,000$  AGNs without any pre-selection over an area of  $\sim 450 \text{ deg}^2$ . The SDSS-IV/eBOSS and BigBOSS builds upon the SDSS-III/BOSS project and will use a fiber-fed spectrograph. Over an area of  $14,000 \text{ deg}^2$ , it will observe roughly one million QSOs at  $1.8 < z < 3.5$ . In addition to these projects, there will be other major enterprises such as LSST (LSST collaboration 2009) and Pan-STARRS (Kaiser et al. 2002) which will detect several million AGNs but currently lack dedicated spectroscopic follow-up programs.

In the following we will focus on eROSITA, as this mission will compile the largest AGN sample ever observed. Figure 3 shows that eROSITA AGN detections will outnumber current galaxy samples with spectroscopic redshifts at  $z > 0.4$ . Using a large number

of AGNs that continuously cover the redshift space, will allow us (in contrast to galaxy samples) to measure the distribution of matter with high precision in the last  $\sim 11$  Gyr of cosmic time. To fully exploit the eROSITA potential for AGN clustering measurements, a massive spectroscopic follow-up program is needed. Several spectroscopic multi-object programs and instruments are currently planned or are in an early construction phase (e.g., SDSS-IV/SPIDERS and 4MOST).



**Figure 3:** Number of expected eROSITA AGNs (red) and currently available galaxies with spectroscopic redshifts (black solid line at  $z < 0.4$  – SDSS data release 7; black dotted line – PRIMUS (Coil et al. 2011); black solid line at  $z \sim 1$  – DEEP2 (Davis et al. 2003) and VVDS (Le Fèvre et al. 2005)). Instead of the full sky area, we consider only the expected number of eROSITA AGNs with spectroscopic redshifts from 4MOST over  $14,000 \text{ deg}^2$ .

eROSITA AGN clustering measurements at  $z \sim 0.8 - 1$  will even allow for the detection of the BAO signal. The feasibility of such a measurement can be estimated using the BAO detection found with  $\sim 46,000$  SDSS LRGs ( $\langle z \rangle = 0.35$ ) over  $3,816$  square degrees of sky ( $0.72 h^{-3} \text{ Gpc}^3$ ) as a standard for comparison (Eisenstein et al. 2005). The observed AGN X-ray luminosity function (Gilli et al. 2007) and the eROSITA sensitivity determine the number density of eROSITA AGNs. In the abovementioned redshift range, the eROSITA AGN area density will be comparable to that of SDSS LRGs at lower redshifts. Therefore, the comoving volume number density of eROSITA AGNs will be five times lower than that of SDSS LRGs. Since eROSITA will conduct an all-sky survey, the increased sky area will counterbalance the lower volume density. Given the signal-to-noise ratio (S/N) of the BAO detection of Eisenstein et al. (2005) and an assumed spectroscopic area of  $14,000 \text{ deg}^2$ , we expect a  $\sim 3\sigma$  BAO detection using eROSITA AGNs only in the redshift range of  $z \sim 0.8 - 1$ . This is consistent with

Kolodzig et al. (2013), who use a different approach based on the angular power spectrum for estimating the significance of a BAO detection with eROSITA AGNs.

With the much larger AGN datasets that will exist in the future, the statistical uncertainties in clustering measurements will be significantly decreased. Systematic uncertainties will then be the dominant source of uncertainty. The impact and level of different systematic uncertainties can only be carefully explored and quantified through simulations. Thus far, there has not been a need for such studies because the AGN samples to date are i) drawn from surveys that (with exceptions) cover a rather moderate sky area and are therefore likely to suffer from the problem of cosmic variance<sup>2</sup> and/or ii) comprised of up to several thousand objects and are consequently Poisson noise dominated. Both limitations will be removed in future AGN clustering measurements with the upcoming extensive AGN samples covering extremely large sky areas. However, to derive reliable constraints on AGN physics and cosmology, as well as to avoid any possible misinterpretations of future unprecedented high precision AGN clustering measurements, we have to fully understand and be able to correctly model the impact of the systematic uncertainties. Only then can we maximize the scientific return of future AGN clustering measurements and have a major impact in the field of cosmology and galaxy and AGN evolution.

## Acknowledgement

MK received funding from the European Community's Seventh Framework Programme (/FP7/2007-2013/) under grant agreement No 229517. TM is supported by UNAM/PAPIIT IN104113 and CONACyT 179662. ALC acknowledges support from NSF CAREER award AST-1055081.

## References

- [1] Abazajian et al. 2009, ApJS, 182, 543.  
[doi:10.1088/0067-0049/182/2/543](https://doi.org/10.1088/0067-0049/182/2/543)
- [2] Alleinato et al. 2011, ApJ, 736, 99.  
[doi:10.1088/0004-637X/736/2/99](https://doi.org/10.1088/0004-637X/736/2/99)
- [3] Bolton et al. 1976, ApJ, 210, L1.
- [4] Cappelluti et al. 2010, ApJ, 716, L209.  
[doi:10.1088/2041-8205/716/2/L209](https://doi.org/10.1088/2041-8205/716/2/L209)
- [5] Cappelluti et al. 2012, AdAst, 2012, 25.
- [6] Chatterjee et al. 2012, MNRAS, 419, 2657.  
[doi:10.1111/j.1365-2966.2011.19917.x](https://doi.org/10.1111/j.1365-2966.2011.19917.x)
- [7] Coil et al. 2009, ApJ, 701, 1484.  
[doi:10.1088/0004-637X/701/2/1484](https://doi.org/10.1088/0004-637X/701/2/1484)
- [8] Coil et al. 2011, ApJ, 741, 8.  
[doi:10.1088/0004-637X/741/1/8](https://doi.org/10.1088/0004-637X/741/1/8)
- [9] Cole et al. 2005, MNRAS, 362, 505.  
[doi:10.1111/j.1365-2966.2005.09318.x](https://doi.org/10.1111/j.1365-2966.2005.09318.x)
- [10] Colless et al. 2001, MNRAS, 328, 1039.  
[doi:10.1046/j.1365-8711.2001.04902.x](https://doi.org/10.1046/j.1365-8711.2001.04902.x)
- [11] Dalton et al. 2012, SPIE, 8446.
- [12] Davis & Peebles 1983, ApJ, 267, 465.
- [13] Davis et al. 2003, SPIE, 4834, 161.
- [14] de Jong et al. 2012, SPIE, 8446.
- [15] Eisenstein et al. 2005, ApJ, 633, 560.  
[doi:10.1086/466512](https://doi.org/10.1086/466512)
- [16] Fanidakis et al. 2013, MNRAS, 435, 679.
- [17] Ferrarese & Merritt 2000, ApJ, 539, L9.  
[doi:10.1086/312838](https://doi.org/10.1086/312838)
- [18] Fry 1996, ApJ, 461L, 65.
- [19] Gebhardt et al. 2000, ApJ, 539, 13.  
[doi:10.1086/312840](https://doi.org/10.1086/312840)
- [20] Giacomoni et al. 1979, ApJ, 230, 540.
- [21] Gilli et al. 2007, A&A, 463, 79.
- [22] Hasinger et al. 2005, A&A, 441, 417.
- [23] Hill et al. 2008, ASPC, 399, 115.
- [24] Hopkins & Beacom 2006, ApJ, 651, 142.  
[doi:10.1086/506610](https://doi.org/10.1086/506610)
- [25] Hütsi et al. 2014, A&A, 561, 58.
- [26] Jahnke & Macció 2011, ApJ, 734, 92.  
[doi:10.1088/0004-637X/734/2/92](https://doi.org/10.1088/0004-637X/734/2/92)
- [27] Kaiser 1984, ApJ, 284, L9.
- [28] Kaiser et al. 2002, SPIE, 4836, 154.
- [29] Kolodzig et al. 2013, A&A, 558, 90.
- [30] Krumpe et al. 2010, ApJ, 713, 558.  
[doi:10.1088/0004-637X/713/1/558](https://doi.org/10.1088/0004-637X/713/1/558)
- [31] Krumpe et al. 2012, ApJ, 746, 1.  
[doi:10.1088/0004-637X/746/1/1](https://doi.org/10.1088/0004-637X/746/1/1)

<sup>2</sup>Surveys with a small sky area will not sample a representative part of the universe due to cosmic variance, i.e., if a survey incidentally aims at an underdensity/overdensity in the universe a lower/higher clustering amplitude will be measured than when aiming at a representative part of the universe.

- [32] Landy & Szalay 1993, ApJ, 412, 64.
- [33] Larsen et al. 2011, ApJS, 192, 16.  
[doi:10.1088/0067-0049/192/2/16](https://doi.org/10.1088/0067-0049/192/2/16)
- [34] Le Fèvre et al. 2005, A&A, 439, 845.
- [35] Limber 1954, ApJ, 119, 655.
- [36] LSST Collaboration 2009, arXiv0912.0201.
- [37] Magorrian et al. 1998, AJ, 115, 2285.
- [38] Martini & Weinberg 2001, ApJ, 547, 12.  
[doi:10.1086/318331](https://doi.org/10.1086/318331)
- [39] Miyaji et al. 2011, ApJ, 726, 83.  
[doi:10.1088/0004-637X/726/2/83](https://doi.org/10.1088/0004-637X/726/2/83)
- [40] Mo & White 1996, MNRAS, 282, 347.
- [41] Mountrichas & Georgakakis 2012, MNRAS, 420, 514. [doi:10.1111/j.1365-2966.2011.20059.x](https://doi.org/10.1111/j.1365-2966.2011.20059.x)
- [42] Mountrichas et al. 2013, MNRAS, 430, 661.  
[doi:10.1093/mnras/sts666](https://doi.org/10.1093/mnras/sts666)
- [43] Peacock & Smith 2000, MNRAS, 318, 1144.  
[doi:10.1046/j.1365-8711.2000.03779.x](https://doi.org/10.1046/j.1365-8711.2000.03779.x)
- [44] Peebles 1980, Princeton University Press.
- [45] Predehl et al. 2010, SPIE, 7732, 23.
- [46] Schlegel et al. 2011, arXiv1106.1706.
- [47] Setti & Woltjer 1977, ApJ, 218, L33.
- [48] Shen 2009, ApJ, 704, 89.
- [49] Shen et al. 2013, ApJ, 778, 98.
- [50] Sheth et al. 2001, MNRAS, 323, 1.  
[doi:10.1046/j.1365-8711.2001.04006.x](https://doi.org/10.1046/j.1365-8711.2001.04006.x)
- [51] Smoot et al. 1992, ApJ, 396, 1.
- [52] Basilakos & Plionis 2009, MNRAS, 400, L57.
- [53] Truemper 1993, Sci., 260, 1769.
- [54] Weinberg 2002, ASPC, 283, 3.