

The Reality of Baryonic Acoustic Oscillations

J. C. Botke

Ronin Institute, Montclair, New Jersey, USA

Email: [jcbotke\(at\)ronininstitute.org](mailto:jcbotke@ronininstitute.org)

How to cite this paper: Botke, J.C. (2024)
The Reality of Baryonic Acoustic Oscillations.
Journal of Modern Physics, 15, 375-400.
<https://doi.org/10.4236/jmp.2024.153016>

Received: January 24, 2024

Accepted: February 26, 2024

Published: February 29, 2024

Copyright © 2024 by author(s) and
Scientific Research Publishing Inc.
This work is licensed under the Creative
Commons Attribution International
License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The initial idea for baryonic acoustic oscillations (BAO) came about during early efforts to understand the origin of galaxies by studying perturbed versions of the Friedmann-Robertson-Walker (FRW) model. In more recent times, the emphasis has shifted to the idea that 2-point galaxy correlations embedded in the distribution of matter by the BAO could be used as a standard ruler to fix the parameters of cosmological models. In this paper, we first consider the actual business of extracting the correlation length from large data sets of measured galaxy locations. To facilitate this process, we introduce a much-improved method for extracting the correlation peak from the data set. Fundamental to this process in any model is the use of a fiducial cosmological model to transition from redshift space to comoving coordinate space where the correlations actually exist. The belief is that the correlation length so determined can then be reverted to redshift space to fix the parameters of cosmological models. We show, however, that this process is circular and hence of no value whatsoever for fixing model parameters. All one obtains are the parameters of the model used to transition to comoving space in the first place. Finally, we present simple arguments that show that the idea of BAO being responsible for the structure of the universe, *i.e.* the cosmic web, is unworkable.

Keywords

Baryonic Acoustic Oscillations, Two-Point Correlations, CMB,
Anisotropy, Early Universe, Time-Varying Curvature

1. Introduction

The subject of this paper is the baryonic acoustic oscillation (BAO) model. The history of the idea goes back many decades and had its beginnings in efforts to understand galaxy formation by studying models that add perturbations to the standard FRW model. These solutions indicate a possibility of sound waves be-

ing excited in the early photon-baryon fluid that would propagate away from the perturbation at a speed close to the speed of light and would persist until the time of recombination when the photons and protons decoupled. According to the standard Λ CDM model, this latter event occurred at a time of about 10^{13} s. Multiplying by the sound speed gives a characteristic length of about 150 Mpc. The theory is that these sound waves created a pattern of higher-than-average baryon densities in some regions of space over others and that these higher densities induced the formation of galaxies. If that was the case, then the distribution of the so-created galaxies should reflect this dimension. In the few decades that followed, and after a lot of effort, the perturbation model of galaxy formation problem seems to have languished and now recent observations by the James Webb telescope cast doubt on the whole idea. At the end of this paper, we will present arguments to show that the BAO model of structure formation is unworkable.

In the last two decades, interest in the BAO has shifted away from the galaxy formation problem to the idea that this characteristic distance could be used as a standard ruler to better determine the parameters of cosmological models. If we start with a random galaxy at some point in space and calculate the probability that another galaxy exists at some distance from the first one, the existence of such a correlation length would result in a bump in the probability at a distance of about 150 Mpc. Detection of this bump, which is known as the 2-point galaxy correlation distance was accomplished in 2005 [1] which, at the time, was taken as confirmation of the BAO model.

The physical manifestation of this correlation distance is that it is the basic length scale of the cosmic web, both of the size of the superclusters making up the filaments and the inter-filament spacing. The proponents of BAO shouldn't have an issue with that idea since the model is supposed to account for the existence of the cosmic web. The problem for the BAO model is that it is not the only model that can account for the cosmic web so a claim of its conformation is spurious. In particular, our new model proposes a much simpler origin in which the matter in the universe, the over-densities of matter in regions that became galaxies, and the cosmic microwave background (CMB) with its anisotropies all came into existence at the same moment at the beginning of nucleosynthesis under the direction of imprints that were established in the vacuum during an initial Planck era inflation.

Much of this paper will focus on the 2-point correlation issue. In the first several sections, we discuss the problems that arise when trying to recover the correlation peak from the galaxy location data and introduce a new method that goes a long way toward solving those difficulties. This new method illuminates the underlying cosmic structure and allows the correlation peak to be detected without the need for the artificial biases that are used in the standard analysis methods.

Following our discussion of the detection problem, we then consider the standard ruler idea. What we show is that the correlation length so determined

cannot be used to fix the parameters of cosmological models. The essence of the problem is that a fiducial model must be used to transition from redshift space to comoving coordinate space and that coordinate space is a fictitious space whose properties are fixed by the parameters of the fiducial model. It is possible that the “user-created” space is identical to the actual comoving coordinate space but that is both unknowable and irrelevant. The fact is that the extracted correlation peak location and width are properties of the fictitious space so when transitioning back to redshift space, the model parameters that will best fit the observed peak are precisely those that one started with. One can use a fiducial model to obtain an estimate of the actual correlation length or if it were possible to obtain knowledge of the actual comoving coordinates without the use of a fiducial model, one could use a model to extract model parameters from the actual correlation length but using a model in both directions is circular and hence, meaningless.

2. Galaxy 2-Point Correlations—The Data

To determine the 2-point galaxy correlation length, one must begin with a data set containing the redshifts and angular positions of a large number of galaxies. As a basis for our investigation, we chose to use data sets similar to those used in [1] and [2], in part because we wanted to use those results as a check on our methods. Both of these studies are based on data available from the Sloan Digital Sky Survey (SDSS) database [3] which is the repository for an ongoing observational program to identify all the galaxies in a significant angular portion of the sky out to a redshift somewhat larger than 2. At present, the database contains well over a million galaxies.

The SDSS website has an extensive user interface which includes an SQL search engine that users can program as they wish. For our first effort, we utilized the redshift and magnitude selection criteria specified in [1], namely redshifts in the range $0.16 \leq z \leq 0.44$ and “ r ” magnitudes in the range, $r \leq 17.77$. The symbol “ r ” refers to one of the 5 standard frequency passbands used in astronomical observations. Please refer to [1] for a discussion of why these limits were selected. The full SDSS data set divides into two distinct regions and because they are distinct, nothing useful can be obtained by trying to combine the sets. For simplicity, we limited our attention to just the “North Cap” which is much the larger of the two. In Figure 1, we show the SQL query we used. It is the last line of the query that imposes the “North Cap” restriction.

```
SELECT so.z, so.ra, so.dec
FROM SpecObj AS so
INNER JOIN PhotoObjAll AS poa ON so.bestobjID = poa.objID
WHERE Class = 'Galaxy'
AND so.z BETWEEN 0.16 AND 0.47
AND poa.r <= 17.77
AND so.ra BETWEEN 90 AND 285
```

Figure 1. Galaxy data set SQL query.

This query returned a set of 98,240 galaxies at the time of our writing. Because the database is continually being updated, the same query run at a later date is likely to produce a different count. Our query does not restrict the galaxies to any particular type. The authors of [1] on the other hand, restrict their list to just luminous red galaxies (LRG) and they ended up with a total of 46,748 galaxies. The reason for imposing that restriction follows from a determination that those galaxies better reflect the cosmic structure than do the general population of galaxies. By making such a selection, however, the authors are at risk of introducing the bias they are seeking. Our new method which we will introduce shortly, does not make any distinction about galaxy type and so avoids that risk.

With the data set in hand, the next step is to determine the comoving coordinates of each galaxy. The reason this is necessary is that the correlations are a reflection of the comoving coordinate positions of the galaxies instead of their redshifts. (Redshifts are a consequence of the observer, not the galaxies). The observed redshift of a galaxy is based on its apparent velocity which is the sum of the Hubble flow velocity and its peculiar velocity. Our interest, however, is in just its Hubble flow velocity and, while various models have been developed to estimate the peculiar velocities of local galaxies, it is not feasible to determine the peculiar velocities of distant galaxies. Generally, peculiar velocities are small compared to the Hubble flow velocities for redshifts greater than about $z > 0.01$ so one can avoid the whole problem by restricting one's attention to redshifts larger than that value. In any case, discovering the correlation peak is a statistical problem so a moderate repositioning of the galaxies won't change the final results.

Converting from redshift to comoving coordinate is model-dependent. The standard model formula is [4],

$$\chi = \frac{c}{a_0 H_0} \int_0^z dz' \left[\Omega_m (1+z')^3 + \Omega_\Lambda + \Omega_k (1+z')^2 \right]^{-1/2} = \frac{c}{a_0} \int_0^z \frac{dz'}{H(z')} \quad (2-1)$$

with $\chi = \sin(r)$ where r is the original radial coordinate. The FRW radiation contribution has been dropped so $\Omega_k = 1 - \Omega_m - \Omega_\Lambda$. In the standard model, the curvature is constant and in recent years, it has become common to assume a flat spacetime in which case, $\Omega_k = 0$. For the remaining parameters, the values used in [1] were $(\Omega_m, \Omega_\Lambda) = (0.3, 0.7)$. In our new model [5] [6], the curvature varies with time so spacetime is definitely not flat, and at early times, it was quite large, e.g. at $t = 1$ s, $k = 2.1 \times 10^{17}$. Unlike the case above, there does not exist a simple formula relating redshift and comoving coordinate so instead, we use numerical methods to calculate both the radial coordinate, r , and the redshift as functions of look-back time and then combine those results to determine $r = f(z)$ [6].

In **Figure 2**, we show the radial coordinate-redshift relationship for both the new and standard models. For later comparison, we also show the scaling over the same redshift range. According to the FRW model, the scaling is given by

$$a(z) = \frac{a_0}{1+z} \quad (2-2)$$

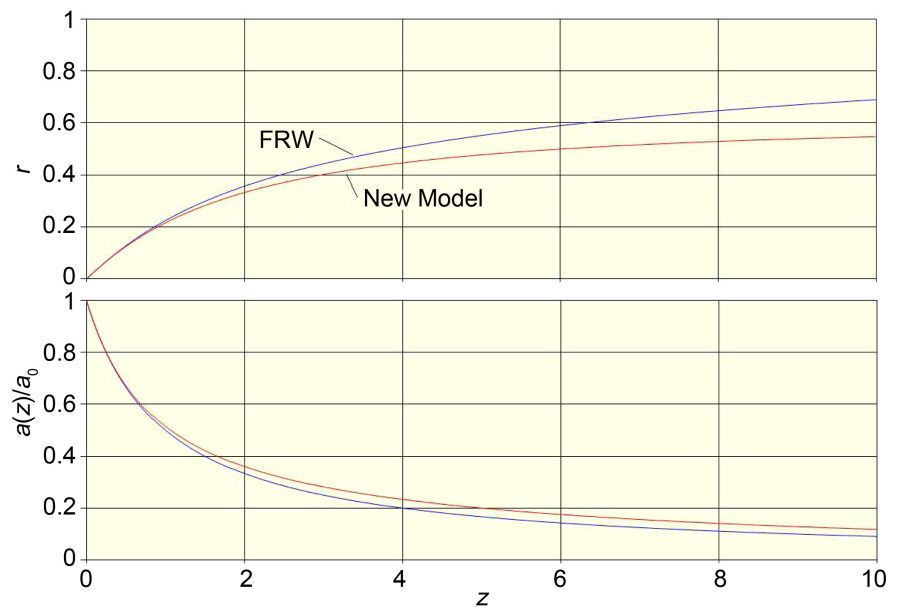


Figure 2. Radial coordinate and scaling versus redshift for both the FRW (blue) and the new model (red).

which, in spite of its widespread use, is not a model independent result. In our new model, the scaling is given by,

$$a(t) = (a_0 e^{-c_1}) \left(\frac{t}{t_0} \right)^{\gamma_*} e^{\frac{t}{t_0} c_1} \quad (2-3)$$

where $\gamma_* = 0.5$ and $c_1 = 0.53$. From this, we obtain the Hubble parameter,

$$H(t) \equiv \frac{\dot{a}(t)}{a(t)} = \frac{\gamma_*}{t} + \frac{c_1}{t_0}. \quad (2-4)$$

Substituting gives $H_0 = 73$ Mpc. We see that the new and FRW model predictions of the radial coordinate are similar for redshift less than about 2.0 and that the two scalings are also initially similar but begin to diverge somewhat earlier than do the coordinates.

The next step is concerned with the coordinate systems used to depict the galaxy locations. The coordinate system used in SDSS is the standard right-ascension system with its origin at the center of the Earth. Instead of using that system, we found it convenient to define two new coordinate systems based on the average location of the galaxies which better represent the sky as viewed by an observer on Earth. In these new systems, the positions of the galaxies are relative to the average position so an observer, for example, would see their locations as being to the right or left of the center point of the collection. Of course, this only amounts to a shift in the axes of the graphs used to display the datasets and results.

We first convert from redshifts to comoving radial coordinates and then determine the average position of the galaxies by summing over the angles and radial coordinates of the full set. (In what follows, to avoid confusion concerning

the variable z , we will use lowercase z to refer to the redshift and uppercase letters to denote cartesian coordinates, e.g. (X, Y, Z) .

To define our new coordinate system, we first rotate about the Z -axis by an amount equal to the right-ascension. After the rotation, the new X -axis lies in both the original XY plane and the plane formed by the average position vector (the line of sight to the average position) and the original Z -axis. Next, we rotate about the new Y -axis to bring the new X -axis into coincidence with the average position vector. In this system, an increase in X corresponds to an increase in redshift. The final Z -axis is now tangent to the original great circle of longitude at the position of the averaged right-ascension. The result is a right-hand coordinate system that presents the galaxies as viewed by an observer on the Earth who is facing the average position.

This coordinate system we call the “Earth” coordinate system (ES). We also make extensive use of a second coordinate system we call the “Galaxy” coordinate system (GS) which we obtain by translating the ES to the average position of the galaxies. The Y and Z coordinates of each galaxy are the same in the two systems but in the GS, the X coordinates of the galaxies will be range over both positive and negative values.

In **Figure 3** we show the distribution of galaxies as viewed by an observer on Earth, (ES). Each frame shows the (Y, Z) positions of all the galaxies lying in an X -coordinate spherical shell with the radius indicated in each frame and with a thickness of, $\Delta X = 0.005$. The angle ϕ is the usual spherical coordinate angle of rotation about the Z -axis. The usual spherical polar angle, θ measures the angular position of the galaxies relative to the Z -axis but in this case, we want the angle relative to the X, Y plane so we define the angle $\psi = \pi/2 - \theta$. As one can see, in our new coordinate system, the galaxies are fairly evenly distributed about $\phi = \psi = 0$. We noted earlier that we are searching for a 2-point correlation at a distance ≈ 150 Mpc. That distance corresponds to a coordinate difference of 0.01 which is shown by the heavy black line at the bottom of each frame.

It is apparent that the number of observed galaxies decreases rapidly with redshift. This, however, is purely an observational issue because, on the length scales we are considering, the universe is homogeneous. The widths of the sample set in both the Z and X directions are about 1/2 the width in the Y direction. The *angular* size of the data set is nearly constant with redshift. At $z = 0.16$, the width in the Y direction is about 10 times the correlation length but that ratio increases with redshift because the correlation length is independent of redshift.

3. Correlations

To discover the correlations, the idea is to calculate the distances from each galaxy to all others and then sort the results into a series of bins. The earliest work considered just the observed galaxies and compared the observed densities in spherical shells surrounding each galaxy with the average density of galaxies [7]. The problem with this method is that it is very sensitive to edge effects. A second

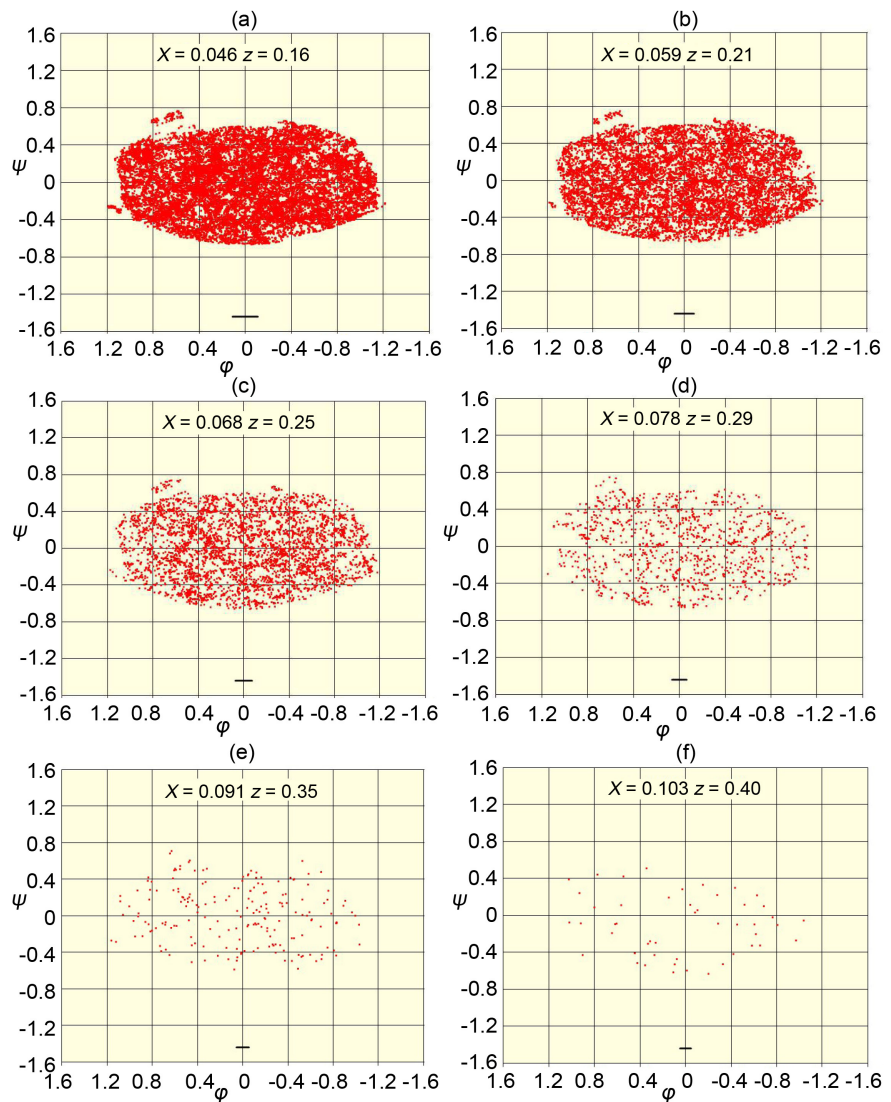


Figure 3. Galaxy distributions for several values of redshift.

problem is that not all galaxies in any particular region of observation are recorded. The latter results in artificially low densities in some regions compared to others.

In more recent work [1] [2] [8], to minimize the edge effects, instead of comparing the sample set with the average expected density, one introduces a population of randomly distributed hypothetical galaxies within the boundaries of the actual galaxies and then compares the distance distribution of the random galaxies with that of the actual galaxies. This mitigates the edge effects since both sets are subject to the same boundaries, but it does not eliminate the problem of artificial underrepresentation of galaxies. We will come back to this point later.

We now need an expression for the probability of finding galaxies with a given separation. Based on the idea of a homogeneous universe, it is customary to assume spherical symmetry in which case the probability of finding a galaxy at a

distance r (in comoving coordinate space) from some other galaxy in a volume dV is given by

$$dP(r) = n(1 + \xi(r))dV \quad (3-1)$$

where n is the average density of galaxies. For a random distribution, $\xi(r) = 0$. The problem is to estimate $\xi(r)$ based on the observed positions of some large set of galaxies. During the development of this field of study, a few different analysis methods have been developed [9]. The most recent and most commonly used is that introduced in [10]. In that model,

$$\begin{aligned} \xi(r) &= \frac{1}{RR(r)} \left(DD(r) \left(\frac{n_R}{n_D} \right)^2 - 2DR(r) \left(\frac{n_R}{n_D} \right) + RR(r) \right) \\ &= \left(\frac{DD(r)}{RR(r)} \left(\frac{n_R}{n_D} \right)^2 - 2 \frac{DR(r)}{RR(r)} \left(\frac{n_R}{n_D} \right) + 1 \right) \end{aligned} \quad (3-2)$$

The symbols D and R refer to the actual galaxies (D) and the hypothetical random galaxies (R) respectively. The n_D and n_R are the average densities of the indicated types. In practice, one specifies a range of r and then divides that range into a number of bins of width Δr . In our case, we used a bin count of 50 and an upper limit of $r \leq 2$ with the result that $\Delta r = 0.04$.

By definition,

$$DD(r) = \sum_{i=1}^N \sum_{j>i}^N \begin{cases} 1 & \text{if } (r - \Delta r/2) \leq |\mathbf{r}_i - \mathbf{r}_j| < (r + \Delta r/2) \\ 0 & \text{otherwise} \end{cases} \quad (3-3)$$

where in this case, both sums run over the actual galaxies. As the calculation proceeds, the calculated distance between each pair is determined and the corresponding bin's count is incremented.

One significant step we introduce in our new method is to sort the galaxies according to their distance from the center of the GS in increasing distance order. Referring to Equation (3-3), the sum over i , which we will call the primary sum, then runs over a sorted list of galaxies from the center outwards. For each of these, we then run over all the galaxies in the list (the secondary sum) with indices greater than i calculating the distance between each pair as we go. Because of our sorting, each of the secondary galaxies is at least as far from the center as the primary galaxy. One advantage of doing it this way is that we can watch the probability distribution as it develops from the center outwards. A second advantage is that we can terminate the secondary sum at the point where the distance from the secondary galaxy to the origin is greater than the distance from the primary galaxy to the origin plus the correlation distance cutoff that we specify. The latter reduces the calculation time by more than a factor of 2.

The measures, $DR(r)$ and $RR(r)$ are computed in the same way with the substitution of the random sets for the actual galaxy sets as appropriate.

We now wish to apply this formalism to the data set described earlier. Before we do, we will first specify the size of a spherical region that will contain our subject list of galaxies. If the region is too small, the number of galaxies will be

too small to achieve reasonable statistical results. We also must require that the diameter of the region is, at a minimum, a few times larger than the expected correlation length. At the other limit, making the region too large adds time to the calculation without adding anything to the results and exacerbates the observational underdensity problem which increases with increasing distance from the center of the collection. Galaxies further and further from the center tend to be randomly located and their inclusion washes out the correlation peak.

We specify our region in terms of two radii. The smaller of the two, the primary radius, corresponds to the primary sum in Equation (3-3). We then define the larger radius by adding to the primary radius, the maximum correlation distance we consider (the maximum r considered in Equation (3-3)). After running several test cases, we found that with this galaxy data set, a primary radius of 0.03 is a good compromise. Choosing a maximum correlation distance of 0.02 then results in a secondary radius of 0.05. The reason for adding this outer region is to make available pair partners for the full range of primary galaxies.

We now need to initialize the random data set. Although they don't explicitly say so, the implication from [2] and [8] is that they create their random galaxies in redshift space and then transpose those into comoving coordinate space. This, however, introduces a bias because the relationship between redshift and comoving coordinate is not linear. What is even worse is that in these same studies, the redshift distribution of the random set is fixed to be the same as for the actual galaxies. The result is a "random" set in comoving coordinate space that isn't random at all.

We instead create our random set directly in comoving coordinate space. We imagine a cubic region in GS using Cartesian coordinates with a side dimension equal to twice the secondary radius which thus encloses the actual galaxy data set. We then create galaxies at random positions in that region and check their distance from the origin. If the random galaxy is within the secondary radius sphere, it is added to the list; if not, it is dropped. The cycle is repeated until the total number of random galaxies equals the required number. At a minimum, the total should equal the number of actual galaxies but after trying a few cases, we found that better results can be obtained by using a multiple of the actual galaxy count.

At one point, we considered sorting the actual galaxies into spherical shells and then adding random galaxies to each shell based on the number of actual galaxies in that shell. This would better reflect the actual distribution of observed galaxies but it would also to some extent impress the correlation length onto the random set so we would be making the same errors that we just finished saying other authors were making.

In the next 2 sections, we will present some results. We emphasize that we are presenting a method for extracting the correlation peak from observational data sets. We do not have any particular interest in the precise parameters of the peaks that result in large part because, as we shown in Section 6, nothing useful

can be done with the result.

4. Low Redshift Results

In **Figure 4**, we show the correlation results obtained using the full data set with a primary radius of 0.03.

There are two curves shown in the figure. The red curve is the result obtained using the full set of 98,240 galaxies. The green curve is the result obtained by removing from that set those galaxies that are primarily responsible for the correlation peak. The procedure for doing this will be explained below. The results show no significant difference between the two curves and no sign of a peak at the expected correlation distance.

The reason for this result is that over a large percentage of the sample area, the density of the observed galaxies is too low to reveal the underlying cosmic structure. There is no reason to doubt that all or almost all galaxies are part of the cosmic web but to expose the structure, the average intergalactic distance of the observed galaxies must be small compared to the characteristic dimension of the web. This sets a lower limit on the required density of observed galaxies in each region of the sky. To get around this problem, the authors of [1] restricted their data set to just LRGs because they had reason to believe that these do reflect the underlying structure. Why any particular type of galaxy should better expose the structure is a question left unanswered. There is no obvious physical reason for such a phenomenon in the grand scheme of the cosmic web so it seems likely that some observational consideration boosts the likelihood of identifying LRG galaxies over other galaxy types in regions where the actual galactic densities are higher than average.

A technique commonly used in an attempt to alleviate the low-density problem is to boost the influence of galaxies in those regions by, in effect, multiplying their number by the ratio of the local density to the total average density. This, however, amounts to data manufacturing. There might be some justification for that step if the low-density galaxies were truly random but that is unlikely because their presence in the SDSS database is a consequence of choices made by

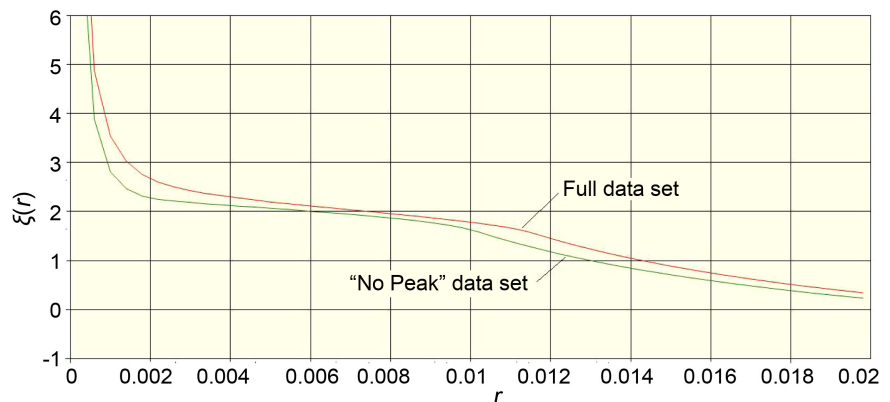


Figure 4. Correlation results for the full galaxy data set with a primary radius of 0.03.

astronomers.

We now introduce a new method for identifying a subset of the observed galaxies that brings to light the underlying cosmic web structure. The idea is simple and is based on the fact that the density of galaxies will be higher in regions defining the backbone of the cosmic web than elsewhere.

We start by defining a cubic grid of cells *in the GS* encompassing the entirety of the observed data set with a cell size considerably smaller than the correlation length. The optimal cell size depends on the particulars of the data set so some experimentation is needed. There is no universal value that works in all cases. We then assign each galaxy in the data set to the cell corresponding to its location in space. The result is that most of the cells will be empty but the remainder will contain from 1 up to some maximum number of galaxies that depends on the chosen cell size. We now come to the essential step. To identify those galaxies that best reflect the backbone of the cosmic web, we simply limit the sample set to those galaxies found in cells with galaxy counts larger than some specified cutoff. This procedure brings into focus the underlying structure without any artificial restructuring being applied and it eliminates from consideration the majority of the galaxies that are in regions where their observational densities are too low to reveal the structure.

Results obtained using a cell size of 0.002 are shown next. With this choice, there were 1,191,016 cells of which 1,169,861 were empty. Of the remainder, the maximum cell galaxy count was 89. Determining the filter cutoff count involved some trial and error. If the cutoff is too low, the peak doesn't show and if it is too high, the filter data set count becomes too small for reasonable statistical results. In this case, we found that 26 is a reasonable compromise. With that value, the filtered data set contained 11,418 galaxies. (In case there is some confusion, the filtered data set contains just the galaxies; the cell structure that was used to generate the filtered set is dropped).

In **Figure 5**, we show the results in the ES system for 4 redshifts. In each case, the figures show the populations of a spherical shell with the redshift indicated in the figure. The frame on the left displays those galaxies from the full data set that lie within the shell and the frame on the right shows the filtered set.

The first observation is that the filtered data set brings out the underlying structure that is hidden in the full data set. The second observation is that the range of redshifts within which the structure is manifested is considerably smaller than the total redshift range of the original data set. By a redshift of $z = 0.21$, even though there are still a considerable number of galaxies in the full data set as shown by the frame to the left, their average spacing has become so large compared to the correlation length that the underlying structure is no longer apparent. Think of the Nyquist frequency. It is not that the structure isn't there or that all the galaxies aren't a part of it, the problem is that we require higher and higher densities of observed galaxies with increasing redshift to see it.

Referring back to **Figure 4**, we can now explain the data set used to generate

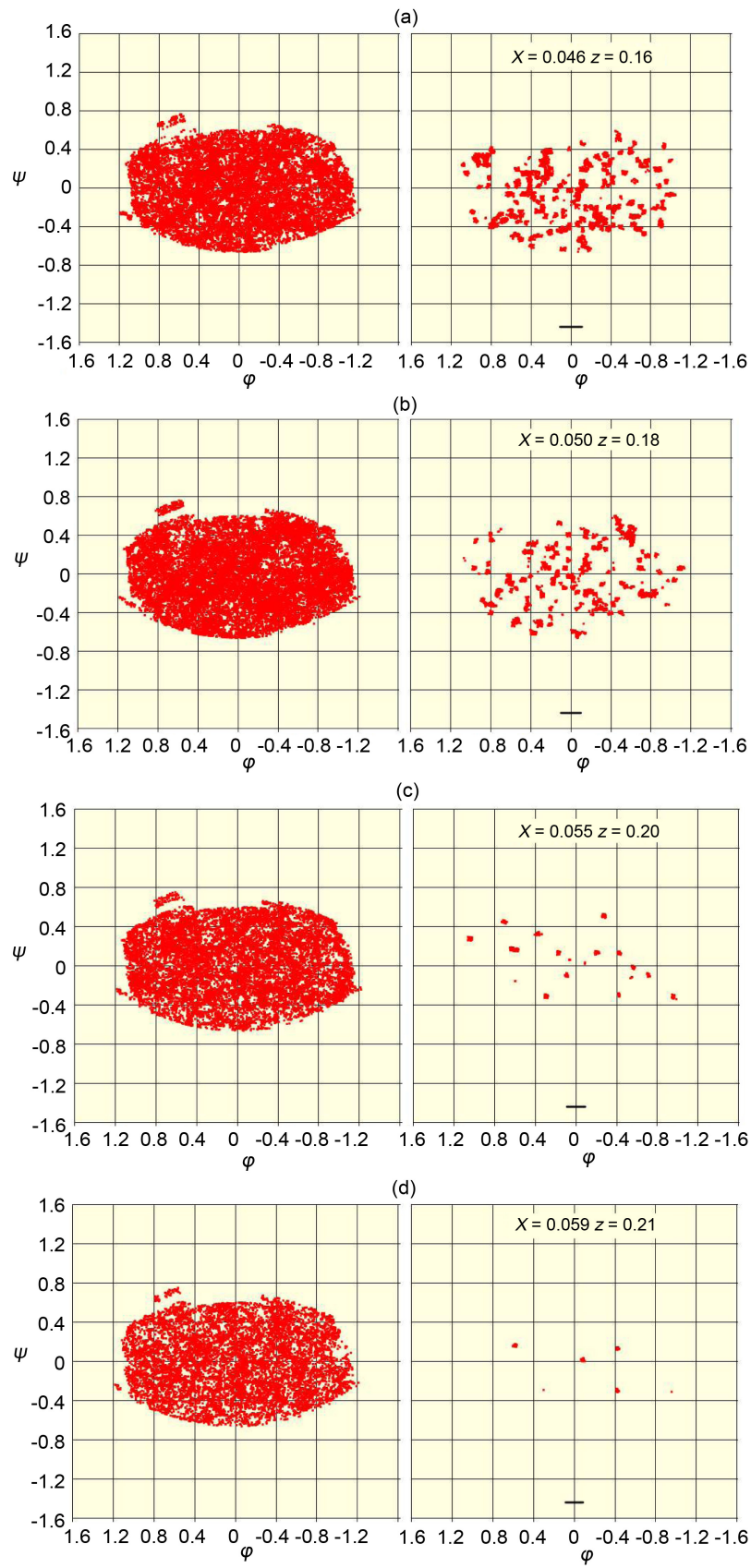


Figure 5. Filtered data set for 4 values of redshift.

the “no peak” curve. To generate the filtered set, we limited the selection to those galaxies in cells with a minimum of, in this case, 26 galaxies. To generate the “no peak” set, we simply did the opposite; we selected only those galaxies that occupy cells that contained at least one galaxy but no more than some upper limit. In the case of **Figure 4**, we set the upper limit to be 20.

In **Figure 6**, we show the distribution of both the actual and random galaxies in the ES for the set of redshifts shown above. In this case, the total number of random galaxies is 4 times the number of actual galaxies. That doesn’t appear to be the case in the figures, but remember that each red dot contains at least 26 galaxies and some contain significantly more.

One can see that the random galaxies lie within a circle containing the outer limit of the actual galaxies. In the last two frames, the actual filtered galaxy distribution becomes sparse so the random galaxies dominate the distribution.

The correlation results with a random galaxy multiplier of 4 are shown in **Figure 7**.

As we noted earlier, by sorting the galaxy list according to distance from the origin in the GS, it is possible to observe the correlation distance result as it develops. The red curve corresponds to the stage at which the primary sum galaxy at index “ I ” of Equation (3-3) was at a distance of 0.0119 from the origin. As can be seen, the 0.01 correlation peak is quite prominent and there are also peaks at about 1/2 and twice the correlation distance. The green curve shows the result at the stage when galaxy “ I ” was at a distance of 0.0138 from the origin. The peak at 0.01 is still visible but it is becoming less prominent and the peak at twice the

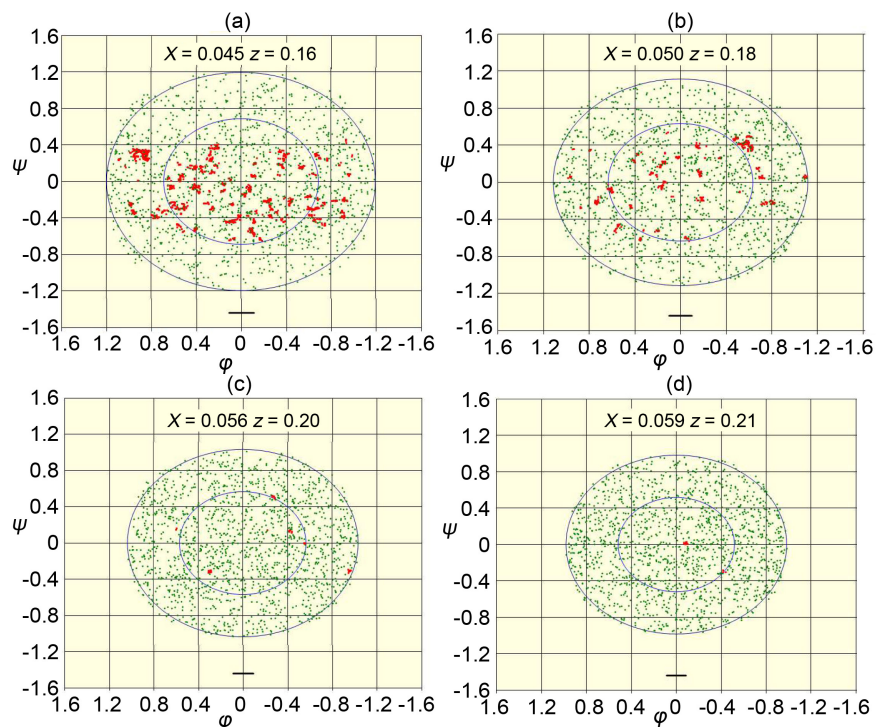


Figure 6. Filtered and random galaxy distributions for 4 redshifts.

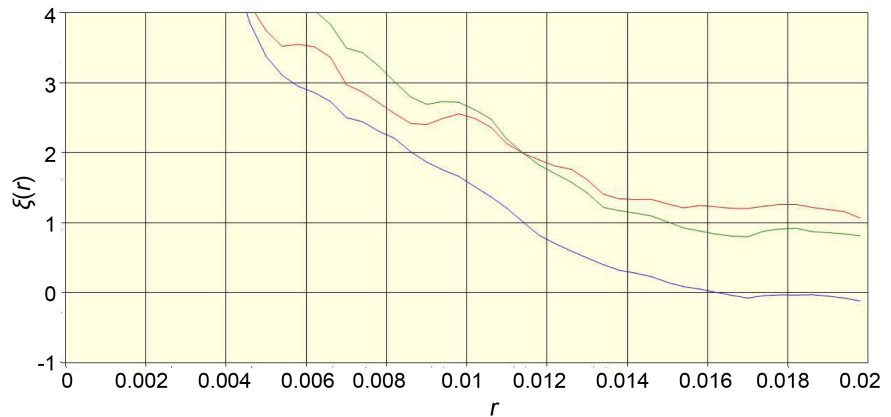


Figure 7. Correlation distance curves for 3 values of the primary galaxy center distance.

distance is still visible. Finally, the blue curve corresponds to a distance of 0.0195 and by that point, the main peak has disappeared but there is still a hint of a peak at twice the distance.

The reason for this result is shown in **Figure 8**. We have suppressed the random galaxies for clarity.

These figures are results corresponding to the first two figures in **Figure 6**. In each case, the current primary radius is indicated by the smaller red circle, and the corresponding secondary radius is indicated by the larger red circle. In Frame (a), we see that both the primary and secondary circles lie largely within the radius of the actual galaxies. The corresponding curve in **Figure 7** is the red curve which shows a prominent correlation peak. In Frame (b), the primary circle is still well within the actual galaxy distribution but the secondary circle is now enclosing a sizeable region containing only random galaxies. The result is the green curve of **Figure 7** which still shows a peak but it is less pronounced. Finally, in Frame (c), the primary circle is now also penetrating regions containing only random galaxies which results in the washing out of the correlation peak. Frames (d)-(f) show a similar pattern but with a reduced density of galaxies. Keep in mind that we are showing slices through the distributions and that the correlation calculation is 3-dimensional with no regard to spherical shells.

We have shown that our new method of analysis reveals the underlying cosmic web when a straightforward application of Equation (3-3) to the full data set does not and that it does so without introducing artificial biases.

5. Higher Redshift Results

We will now look at different data set consisting of quasars instead of normal galaxies. Taking our selection parameters from [2], the redshift range is from 0.8 to 2.2 and, unlike in the previous case, without a magnitude cutoff. Our query is shown in **Figure 9**.

We again restricted our data set to just the “North Cap”. Running this query results in a data set containing 281,572 quasars.

In **Figure 10** we show the distribution at 3 different redshifts.

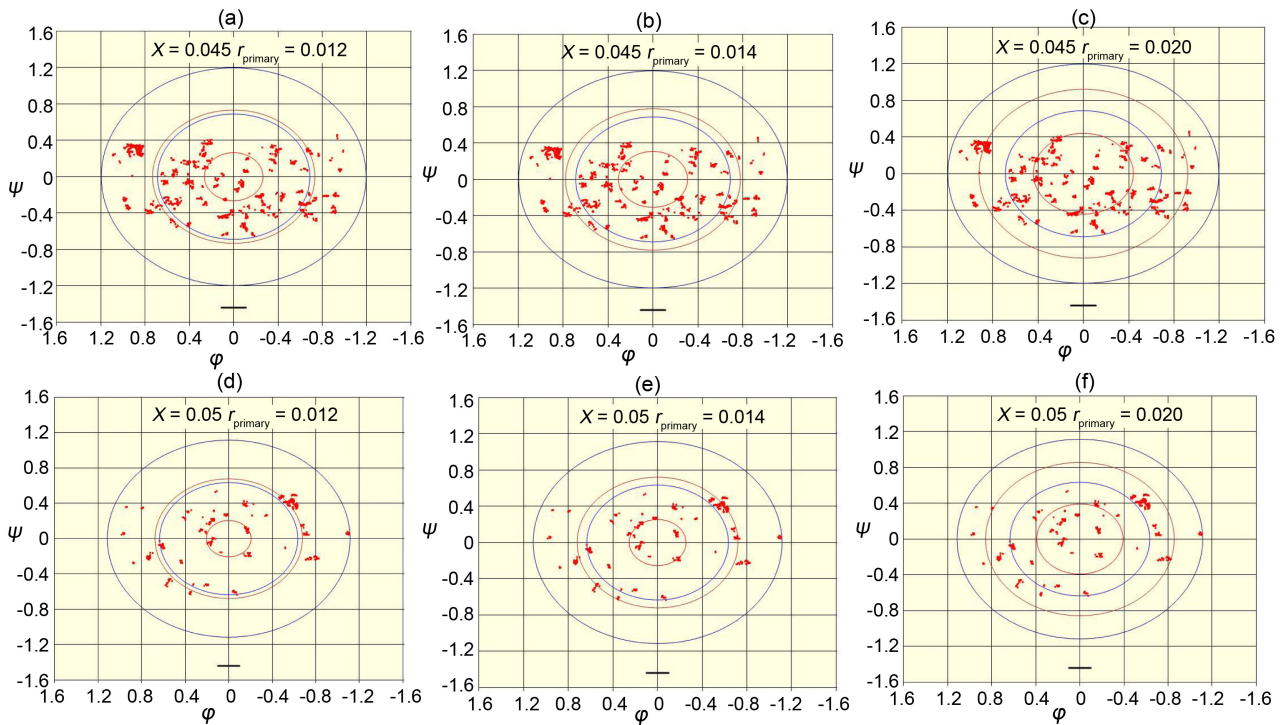


Figure 8. Calculation limits for 3 values of the primary galaxy center distance and 2 values of redshift.

```
SELECT so.z, so.ra, so.dec
FROM SpecObj AS so
WHERE Class = 'QSO'
      AND so.z BETWEEN 0.8 AND 2.2
      AND so.ra BETWEEN 90 AND 285
```

Figure 9. Quasar data set SQL query.

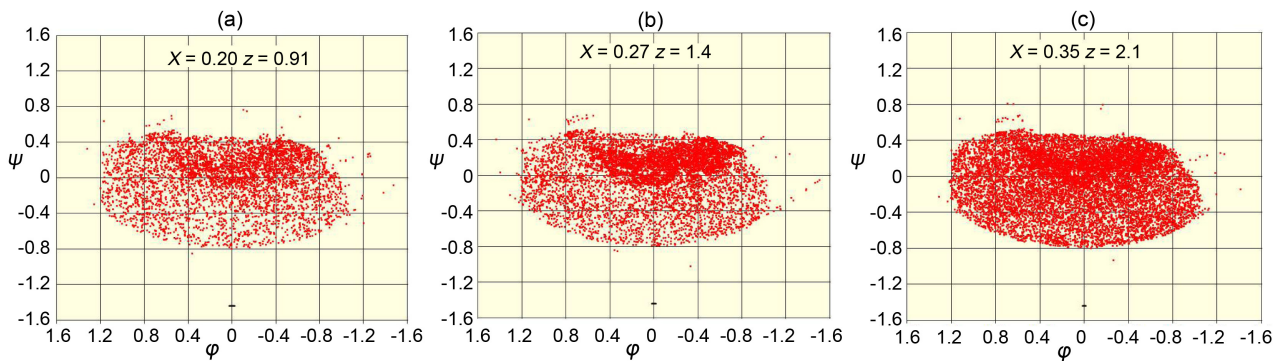


Figure 10. Quasar distributions for 3 values of redshift.

We note that, unlike the case with the galaxies, the count of quasars increases with redshift. In **Figure 11**, we show the correlation results for the complete data set. As can be seen, there is no indication of a correlation peak at 0.01.

We next applied the filtering process we introduced above. Finding a peak was considerably more difficult than it was in the low redshift case but with just the right filtering parameters, a correlation peak can be detected. This result is

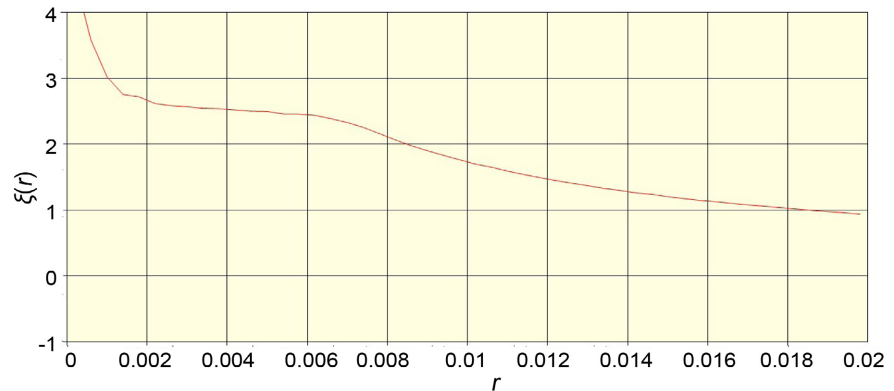


Figure 11. Quasar correlation distribution for the full data set.

consistent with the results of [2] in which they found only a weak indication of a peak. After some experimentation, we settled on a filter cell size of 0.006, which results in a maximal cell quasar count of 28. After more experimentation, we found that the best results were obtained with a filter cutoff of 10, a primary sum radius of 0.1, and a random galaxy multiplier of 6. As before, these parameters are peculiar to the data set.

In **Figure 12**, we show the filtered data set at an intermediate value of redshift.

By comparing with the very small correlation length indicator line at the bottom of the right frame, one can see that the peak is developed within the larger red dots rather than between the dots. The distances between the data points are generally large compared to the correlation length which is an indication that the density of the quasars is too low to give a robust realization of the 0.01 correlation peak.

The resulting correlation curve is shown in **Figure 13**. The curve does show a peak although it is not exactly at a separation of 0.01.

Because the correlations are a function of the comoving separations of the quasars, to reveal the peak at the same level of confidence as with the galaxy data set considered earlier, the density of observed objects in comoving coordinate space must be more or less the same. The density varies as the cube of the distance from the observer and the ratio of the average of the filtered galaxy distances to that of the quasars is about 5.3 so to achieve the same resolution, the quasar data set would need to have a sample count on the order of 150 times greater than the galaxy data set count. The actual ratio is about 2.9.

We can illustrate the same idea in a different way. We create a cubic grid to enclose the entire data set with the cell size this time set to the correlation length, 0.01. We then run through the entirety of both data sets (not the filtered data sets) incrementing the count in each cell by the number of galaxies or quasars that lie in that cell. From these results, we create a distribution in which the index making up the horizontal axis is the number of cells whose galaxy/quasar count equals the index. For example, the value of the bin at index 10 in the total number of cells contains 10 galaxies or quasars. The results are shown in **Figure 14**.

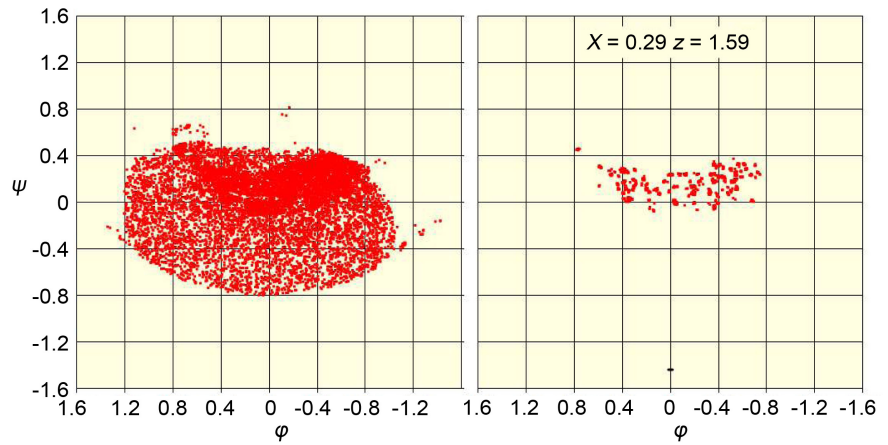


Figure 12. Filtered quasar data set at a redshift of $z = 1.59$.

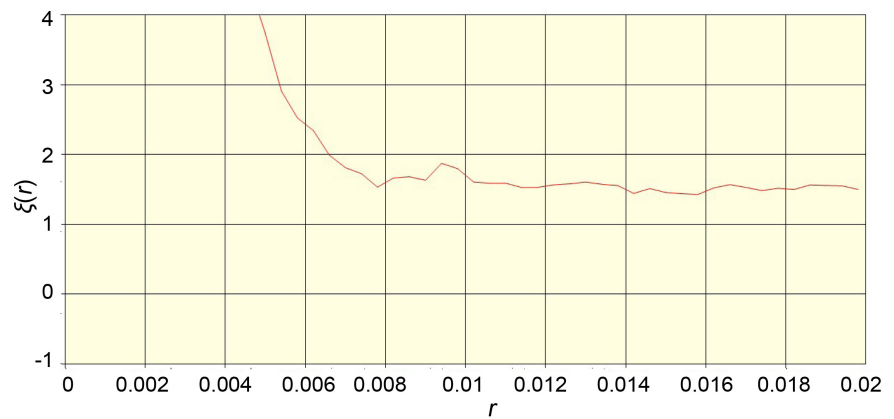


Figure 13. Filtered data set result.

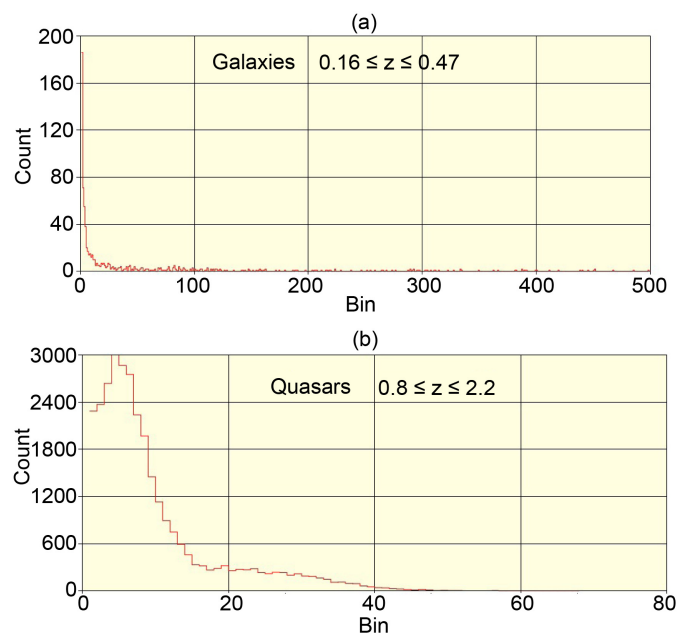


Figure 14. Distributions of galaxies and quasars in a cubic grid with a cell side dimension equal to the correlation length, 0.01.

In the galaxy case, there is a large range of cell populations with the cell containing the largest count having 1506 galaxies. In the quasar case, the range of cell counts is very limited with a maximum of 67 and with the bulk having counts less than 20. Seeing the correlation peak requires a considerable number of cells with large populations. The galaxy data set satisfies this requirement whereas the quasar data does not.

The question now arises as to whether the low density of quasars is a matter of observational limitations or whether it is because there just aren't a lot of identifiable quasars in comparison to the number of ordinary galaxies. If the latter, which we suspect is the case, then quasars will always be of lesser use in determining the correlation distance with the problem growing worse with increasing redshift.

We can gain an appreciation of the quasar distributions by making a plot of the full sky count of quasars versus look-back time. The idea is to specify a circular region of the sky in the ES, count the number of quasars in that circle as a function of redshift, scale up to the full sky count, and plot the result. In **Figure 15**, we show the definition of a typical circle. We have placed the circle in an area of greater density to minimize the effects of limited observations and made it small enough to avoid edge effects over the full range of redshifts.

In **Figure 16**, we show the quasar counts as a function of $\chi(t) = t/t_0$. The upper horizontal line is the total count of galaxies in the observable universe which does not change to any appreciable degree. The starting point of the line is the time of galaxy formation, $t \approx 10^{16}$ s. The quasar counts show a slow decrease with time which is presumably a consequence of the quasars running out of fuel and becoming normal galaxies.

One of the predictions of our new model of cosmology [11] is that all galaxies began life as at least mini-quasars because if they hadn't, they would have undergone free-fall collapse and ceased to exist. We also showed in the same paper

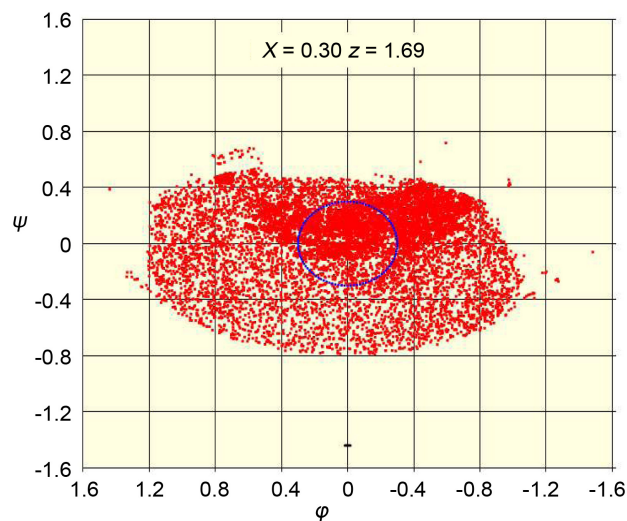


Figure 15. Sampling circle with an angular radius of 0.3 radians.

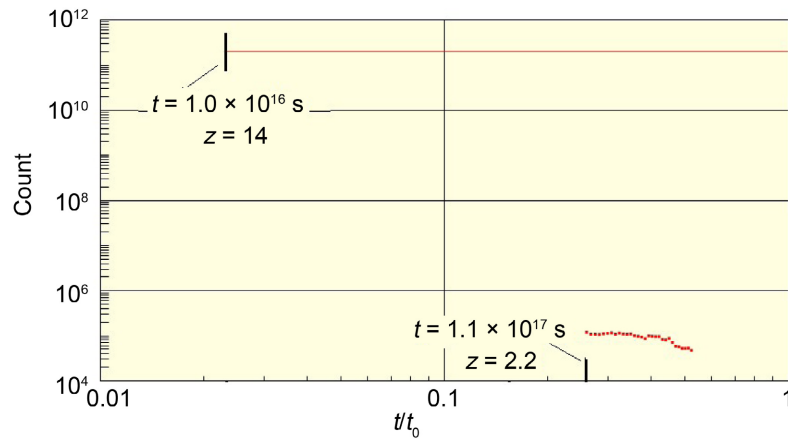


Figure 16. Full sky quasar counts as a function of look-back time. The upper horizontal line is the total (constant) count of galaxies in the observable universe.

that the stability of galaxy clusters demanded that the quasar phase of almost all galaxies must have ended by an extinction time of about $t = 6.8 \times 10^{16}$ s.

We now want to compare that idea with the observed quasar distribution. At this point, we don't know how to model the extinction history of the mini-quasars which would depend in part on their initial supply of quasar fuel. To get some idea of the evolution, we calculate the decay assuming first a Gaussian distribution and second, an exponential, mean lifetime distribution. In both cases, we know that the initial count was about 2×10^{11} and that most were normal galaxies by the extinction time noted earlier. We looked at two cases. In one, we required that 10% of the galaxies were still active at the extinction time, and in the other, that just 1% were still active. The results are shown in **Figure 17**.

The mean lifetime curves indicate a life span that is probably too long because if there were still a sizable number of galaxies in their quasar phase at late times, it seems likely that by now someone would have noticed. The Gaussian results indicate a termination within the redshift range covered by the quasar data set. While these curves are only guesses, they are constrained by the earlier stated conditions. The model does account for the rapid demise by several orders of magnitude of the mini-quasar action by the time range of the observed quasars.

We also see that the decay rate of the observed quasars is not remotely compatible with the calculated curves. This indicates that at the time of galaxy formation, a very small percentage of the newly forming galaxies came into existence with a super-size supply of quasar fuel and these are the ones that are now observed. The remainder underwent rapid decay after the extinction time and even though the Gaussian model indicates that a significant number of mini-quasars were still active at a redshift of 2.2, they would have been running down and not emitting a significant amount of radiation. These would not be recognized as quasars unless observers were specifically looking for that signature. This model suggests that the evolution of quasars, from mini to full

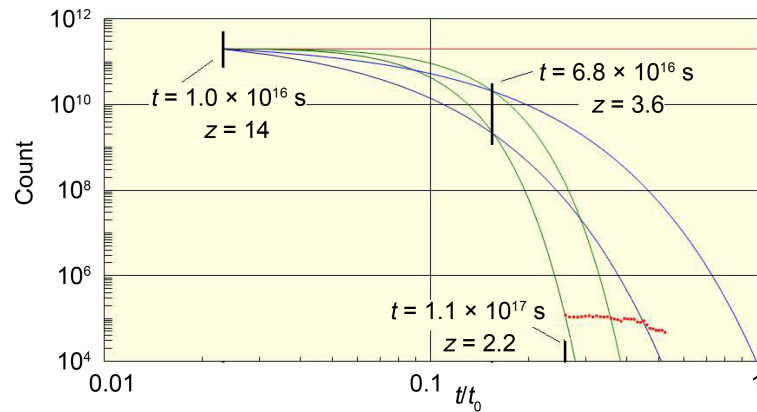


Figure 17. Predicted mini-quasar counts as a function of look-back time. The Gaussian results are shown in green and the exponential results are shown in blue.

strength, during this epoch would be a fruitful area for investigations in the future.

To summarize, we have examined two data sets, one of low redshift galaxies and the other of intermediate redshift quasars. We found in both cases that a computation of the correlation function for the full sets does not reveal a peak. We then applied our new method and discovered peaks in both cases with the effect being stronger in the low redshift galaxy case. That is the good news and now for the bad news. While this observed correlation length may be of interest as a measure of the dimensions of the cosmic web, it has nothing to say about the parameters of cosmological models.

6. Cosmological Model Parameters

A great deal of effort during the past few decades has been expended on the idea of using the observed correlation distance to fix cosmological model parameters such as the Hubble parameter. We will now show that that idea doesn't work. The crux of the problem is that we have no way of measuring the actual correlation distance. What we measure instead is the correlation length that we ourselves create when we pass from redshifts to comoving coordinates using some cosmological model.

The 2-point correlation distance is a comoving coordinate phenomenon. Leaving aside peculiar velocities which are relatively small on the scales we are considering, all galaxies are at rest in comoving coordinate space and since their positions don't change, neither do the actual 2-point correlations. Since on large scales, the universe is homogeneous and isotropic, it follows that measurements of the correlation from any observation point at any epoch will return exactly the same result. This means that nothing bearing on time or location can be extracted from the correlation length itself. The idea that cosmological parameters can be fixed by measuring correlations at different redshifts is rooted in the idea that there are model-independent differences in the *observed* correlation length

when viewed at different redshifts.

All studies begin by using a fiducial model to convert the redshifts into comoving radial coordinates. One then extracts the correlation peak in comoving coordinates as we did in the previous sections and finally, reverts to redshift space using the same model but this time fixing adjustable model parameters along the way. But, *all one is doing is measuring the parameters of the original model.*

We first use a fiducial model to convert from the measured redshifts of these galaxies to comoving coordinate space so

$$r_i = F(z_i; p_1, p_2, \dots) \quad (6-1)$$

where the p_k are the parameters of the model. Unless the chosen model happened to be perfectly correct, the comoving coordinates so determined will not match the actual coordinates of the galaxies and even if they did, *we have no way of knowing it.* Next, we do the usual analysis to determine the correlation length in our “user-created” comoving space with a result that will generally be different from the actual correlation length although probably not by very much.

The standard procedure from then on is to consider the transverse and radial directions separately. In the FRW formulation, any two galaxies with the same angular coordinates will have a separation given by [12]

$$s_{\parallel} = \frac{c}{a_0} \frac{\Delta z}{H(z)} \quad (6-2)$$

where s_{\parallel} is discovered by analyzing a restricted data set that is narrow in angular extent. We then look for the correlation peak and use the result to fix a particular value of the LHS of the equation. Depending on the model, this separation will be larger, smaller, or even the same as the actual correlation length. We now pretend that we don’t know how they were placed and calculate the RHS of the equation for any two galaxies using a model with an unknown set of parameters

$$\Delta z_{i,j} = G(r_i, r_j; p'_1, p'_2, \dots) \quad (6-3)$$

and ask what set of parameters, p'_k , gives a calculated $\Delta z_{i,j}$ that equals the LHS of Equation (6-2). By construction, the LHS is

$$r_i - r_j = F(z_i; p_1, p_2, \dots) - F(z_j; p_1, p_2, \dots) \quad (6-4)$$

so the exact solution is simply

$$G(r_i, r_j; p'_1, p'_2, \dots) = z_i - z_j = F^{-1}(r_i; p_1, p_2, \dots) - F^{-1}(r_j; p_1, p_2, \dots) \quad (6-5)$$

Thus, the whole business boils down to $z_i = F^{-1}(F(z_i; p_1, p_2, \dots); p_1, p_2, \dots)$ with $p'_k = p_k$. This is an identity and hence true for any fiducial model with any set of parameters. The model parameters that best match the measured “user-created” correlation distance are exactly those that were used to create the correlation length in the first place. The actual correlation length never enters into the process and it makes no difference whether or not the “user created”

comoving coordinates match the actual coordinates. What does matter is that a model is used in both directions, redshift-to-comoving-to-redshift.

In the transverse direction, the formula connecting the measured transverse correlation length and the model parameters is [12],

$$s_{\perp} = \frac{a_0}{1+z} r \theta. \quad (6-6)$$

In this case, the correlation is determined from a data set with a restricted range of redshifts. Of course, in reality, the two correlations must be the same. A model error in the initial redshift-to-coordinate conversion that increases the comoving radial coordinates of any two galaxies will increase the average distance between those galaxies so the result will be an increase in the measured correlation length. In real space, the angle cannot be measured directly because one does not know which two galaxies are exactly 1 correlation length apart but the angle can nevertheless be calculated because the redshift-to-comoving transition acts only in radial direction so any errors leave the angle unchanged. The situation is now the same as before. One uses a model to determine the comoving distance which fixes the product $r\theta$ for a given redshift. Reverting to redshift space by calculating the RHS of Equation (6-6) will return the same model parameters that were used in the initial step.

In both cases, there is a possibility of a degeneracy between different parameter sets but the fact remains that using the original parameters will always yield a perfect match.

As noted in [12], researchers in this field are aware that the comoving data set depends on the model used to transform to comoving coordinate space but choose to ignore this issue when doing parameter estimation studies. For some reason, they fail to appreciate that the fiducial model is not a gateway to the actual comoving coordinates but instead, it creates the fictitious comoving coordinate space where the measurements are actually made.

We will now make another point about model dependence. Consider following the previous analysis using first the Λ CDM model with any parameter set one would like and then with our new model. We first convert the observed locations of the galaxies from redshift to comoving coordinate. In both models, there are no off-diagonal components in the metric connecting the angular with the radial coordinate so there would not be any disagreement about the angular coordinates. We see from **Figure 2** that the relationship between coordinate and redshift is essentially the same for redshifts less than 2 so for the galaxy data set considered above, the calculated comoving locations of all the galaxies will be essentially the same in both models and hence, so will the measured value of the correlation distance. We now transition back to redshift coordinates. If the Λ CDM model parameters are treated as adjustable which is the case in the studies, to reproduce the original redshifts, the parameters will have to be the same as were used in the outgoing step. Because the calculated comoving coordinates are the same, and from **Figure 2**, we see that the scaling predicted by the two

models are similar, one might expect that the resulting Hubble parameter would be the same. That, however, is not the case. In **Figure 18**, we show the reduced Hubble parameter, $H(z)/(1+z)$ for the two models.

The red curve is the new model prediction and the 2 blue curves are the Λ CDM predictions for $H_0 = 67.6$ and 73. The former is the fiducial value used in both [2] and [8]. The latter is the new model value. We included the Λ CDM curve with that value to permit a direct comparison with the new model prediction. The curves show that even with identical data sets in comoving coordinate space, different models predict entirely different Hubble parameters.

The orange data points are from [8] and the green data point is from [2]. In [8], for example, they find the same values $(\Omega_m, \Omega_\Lambda, H_0)$ as they started with which is the point we are making. If they had started with a different fiducial model, their results using the same data set would lie on a different Hubble parameter curve.

The crux of the problem is that we have no way of determining the actual comoving locations of the galaxies and hence their spacing. To be useful as a standard ruler, one would need to be able to recover the correlation length without first using a model to create the comoving coordinate space. *The whole process is circular and therefore meaningless.*

7. The Reality of BAO

In this section, we will give arguments that show that the acoustic wave idea is unworkable as the explanation for the cosmic web. To get an idea of the scope of the problem, we will make a few simple estimates. First, we consider the total energy contained in the CMB anisotropies. The energy density of blackbody radiation is given by $\rho_{BB} = a_{SB} T^4$ and, at present, the variance of the CMB spectrum has a peak value of $\Delta T/T_0 = 2 \times 10^{-4}$. Earlier, at the time of recombination, the variance ratio would have been the same but the temperature was then $T_{rec} = 3000$ K. The energy density of the anisotropies would then be

$$\rho_{\Delta T} = a_{SB} \left((T + \Delta T)^4 - T^4 \right) \approx 4a_{SB} T^4 \frac{\Delta T}{T} = 4.9 \times 10^{-5} \text{ J} \cdot \text{m}^{-3} \quad (7-1)$$

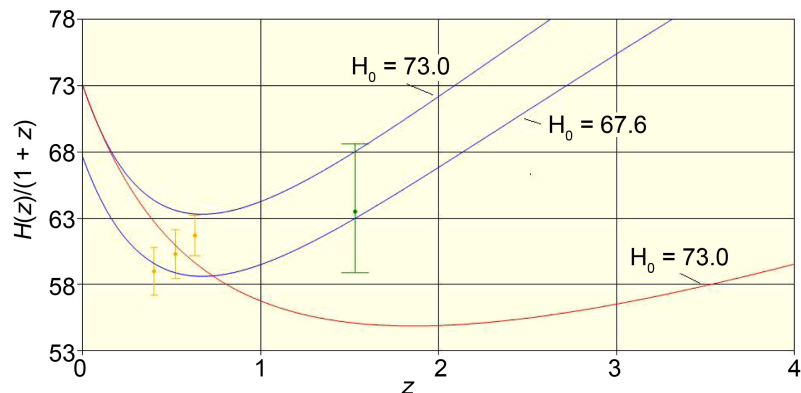


Figure 18. Model predictions for the reduced Hubble parameter.

To get the total energy for a supercluster whose dimensions are the reality of the correlation distance, we multiply by the volume of a supercluster filament. The radius of a filament is about 10% of its length so we find a total of

$E_{total} = 4.9 \times 10^{59}$ j. This energy could not have originated any earlier than the time of nucleosynthesis since there were no baryons in existence before then. If we chose a reference time of 1 second, causality limits the size of a source region to a value of $ct \approx 3 \times 10^8$ m. The radiation energy at that time was $\rho_\gamma = 6.7 \times 10^{22}$ j·m⁻³. The BAO model is based on the idea of perturbations with energy densities small compared to the radiation energy density so the maximum total energy released by a source cell would be considerably less than 7.5×10^{48} j. Comparing with the needed anisotropy total energy shows that the count of sources would have to have been considerably larger than 10^{11} .

We will now approach the source cell count from a different point of view. In comoving coordinates, the average dimension of a supercluster is 0.01 which, for simplicity, we will assume is a reasonable estimate of its size at a time of 1 second. The scaling at that time was $a_1 = 3.9 \times 10^{17}$ m so the comoving dimension of a source cell was 7.7×10^{-10} . The ratio is 1.3×10^7 so within the volume of the perturbation that became the supercluster, there would have been on the order of 10^{21} source cells. Now, going back to the energy-based estimate, the energy density of the outgoing spherical wave would decrease at least as fast as the square of the distance ratio so the estimated total number of cells would now be greater than 10^{25} .

We don't claim great accuracy for these numbers but they do indicate that the required number of source cells was huge and they also indicate that there was not enough space in the assumed initial perturbed regions to account for the total energy needed to explain the anisotropies.

We now come to a more unsurmountable problem for the BAO model; namely that the spherical waves emanating from this huge number of source cells, *which have no causal connection and hence have random phases*, must conspire in exactly the right way across the whole of the universe to account for the structure of the cosmic web with its generally linear filaments. That is just not possible, particularly given the circumstance that all but those waves with the longest wavelengths are suppressed in the BAO perturbation model.

In our new model, because of the different scaling, the radiation temperature at the time of nucleosynthesis was about a factor of 10 smaller than the standard model value, and recombination occurred about a factor of 10 sooner, 10^{12} s versus 10^{13} s. The consequence is that in this model, there would not have been sufficient time for BAO waves to reach the size of superclusters. Instead, the limit of expansion would have been around 15 Mpc indicating that such waves, even if they did actually exist, had nothing to do with the existence of the cosmic web.

8. New Model

We will conclude with a few remarks concerning our new model of cosmology.

First, this new model makes a parameter-independent prediction that the present-day universe must be undergoing an exponential expansion and it does so without any reference to dark energy. In particular, it makes an accurate prediction of the luminosity distance observations over the existing redshift range of observations [13]. The model also proposes a new model of matter creation [14]. In this model, all the matter in the universe came into existence at a time of about 10^{-5} s when a small percentage of the vacuum energy converted into neutron-antineutron pairs with a very small excess of neutrons; a process that was regulated by an imprint established in the vacuum during an initial Planck era inflation. Along with the baryons and leptons, this process created the photons that became the CMB together with its anisotropies. Compared to this model, BAO had a Rube Goldberg air about it.

We have argued above that the BAO model is unworkable so it follows that the BAO fit to the CMB anisotropy spectrum is meaningless. Instead, we show, [5] [14], that the position of the first CMB anisotropy peak was a consequence of the same process that created the cosmic web. Since then, we have accurately calculated the observed maximum temperature of the CMB anisotropy distribution [15] while showing that galaxy clusters are responsible for both that highest temperature and the 3rd peak in the anisotropy spectrum.

9. Conclusion

We have presented a new analysis of the BAO idea. We first performed an analysis of two large data sets and showed that one can extract the 2-point correlation peak as a comoving separation of 0.01 using a new method which avoids shortcomings of the standard methods. We next explain why this measured length cannot be used as a standard ruler to fix cosmological models. We then jump back to the original origin of the BAO idea and finish with simple arguments to show that the idea of BAO being responsible for the cosmic structure of the universe is unworkable.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Eisenstein, D.J., *et al.* (2005) *The Astrophysical Journal*, **633**, 560. <https://iopscience.iop.org/article/10.1086/466512>
- [2] Zarrouk, P., *et al.* (2018) *Monthly Notices of the Royal Astronomical Society*, **477**, 1639-1663. <https://doi.org/10.1093/mnras/sty506>
- [3] Sloan Digital Sky Survey/SkyServer. <https://skyserver.sdss.org/dr7/en/>
- [4] Hobson, M.P., Efstathiou, G.P. and Lasenby, A.N. (2006) *General Relativity: An Introduction for Physicists*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511790904>
- [5] Botke, J.C. (2023) *Cosmology with Time-Varying Curvature: A Summary*.

- <https://www.intechopen.com/online-first/1167416>
- [6] Botke, J.C. (2020) *Journal of High Energy Physics, Gravitation and Cosmology*, **6**, 473-566. <https://doi.org/10.4236/jhepgc.2020.63037>
 - [7] Rivolo, A.R. (1985) *The Astrophysical Journal*, **301**, 70-76.
<https://articles.adsabs.harvard.edu/full/1986ApJ...301...70R>
<https://doi.org/10.1086/163873>
 - [8] Alam, S., *et al.* (2017) *Monthly Notices of the Royal Astronomical Society*, **470**, 2617-2652. <https://academic.oup.com/mnras/article/470/3/2617/3091741>
 - [9] Coil, A.L. (2012) Large Scale Structure of the Universe. ArXiv:1202.6633v2.
<https://arxiv.org/pdf/1202.6633.pdf>
 - [10] Landy, S.D. and Szalay, A.S. (1993) *Astrophysical Journal*, **412**, 64-71.
https://articles.adsabs.harvard.edu/cgi-bin/nph-iarticle_query?1993ApJ...412...64L&defaultprint=YES&page_ind=0&filetype=.pdf
<https://doi.org/10.1086/172900>
 - [11] Botke, J.C. (2022) *Journal of High Energy Physics, Gravitation and Cosmology*, **8**, 345-371. <https://doi.org/10.4236/jhepgc.2022.82028>
 - [12] Bassett, B.A. and Hlozek, R. (2009) Baryon Acoustic Oscillations. ArXiv:0910.5224v1.
<https://arxiv.org/pdf/0910.5224.pdf>
 - [13] Botke, J.C. (2023) *Journal of High Energy Physics, Gravitation and Cosmology*, **9**, 60-82. <https://doi.org/10.4236/jhepgc.2023.91007>
 - [14] Botke, J.C. (2022) *Journal of High Energy Physics, Gravitation and Cosmology*, **8**, 768-799. <https://www.scirp.org/journal/paperinformation.aspx?paperid=118834>
<https://doi.org/10.4236/jhepgc.2022.83053>
 - [15] Botke, J.C. (2024) *Journal of High Energy Physics, Gravitation and Cosmology*, **10**, 257-276. <https://doi.org/10.4236/jhepgc.2024.101020>