Hi everyone, I’m Jiaxi. My master project is the SubHalo abundance matching for eboss galaxies. My supervisors are Prof. Jean-Paul Kneib and Dr. Cheng Zhao.

I’ll start with two questions: why we need the Subhalo Abundance matching and what it is.

We know that the DM theory describes the gravitational growth of structures in the universe. DM particles interact through the gravitational force, their behaviour could be modeled by the N-body simulation. Simulation results validate the DM theories to ensure we work in a self-consistent framework.

The observations of large scale structures can constrain the parameters under the LCDM framework, such as H0. However, the structure formation theory of LCDM described the DM, not the baryonic matter. So there is something missing in between the DM and galaxy theories. Studies in the 1990s has also found the correlation functions of the N-body simulation were higher than those of the observed galaxies.

The missing link is the galaxy bias theory. The theory describes how galaxies reside in the DM halos.

Once we confirm it is a reliable link under the LCDM framework, then the galaxy theory is also reliable and the constraints on the parameter are meaningful.

The SHAM is one method to calibrate our galaxy bias hypothesis.

The bias study includes the halo and subhalo catalogue from the simulation and the galaxy catalogue from the massive spectroscopic survey.

There are two common ways do the bias model calibration, that is to say, assigning galaxies to halos. The first one is the halo occupation distribution. The basic assumption is for a dark matter halo with a given mass Mh, there is a certain probability to host N galaxies, one central and (N-1) satellites. While for the Subhalo Abundance Matching method, the principle is that halos and subhalos have possibility to host one central galaxy. Those reside in the subhalos are equivalent to the satellite galaxies. The possibility is a function of the halo or subhalo mass.The next question is, how we can achieve the SHAM galaxy distribution in halos.

The core of this method is the selection of DM halos. Those selected halos can host a galaxy in its centre and have the same peculiar velocity as the halo.

The galaxies in the simulation box should have the same number density as the observations. It determines the total number of selected halos or SHAM galaxies.

They follow the empirical distribution function in halos. The probability distribution function here is the probability for a halo with a given mass to host a galaxy inside. So it equals to the number of selected halos divided by the total number of halos in a specific mass bin. The basic assumption is massive halos have higher probability of having a galaxy inside.

For massive halos, their gravitational potential is strong enough to host any LRG. However, the star-forming processes will be quenched in such dense environment, so the PDF is very close to 0.

For low mass halos, their gravitational potentials are too weak to host any galaxy.

Due to some physical processes, the transition from the most massive halos and the light halos is not sharp. The stochasticity makes it smooth.

Finally, their two-point correlation functions should match the observations on 5-25 Mega pc per h. This is also the step to calibrate the galaxy probability distribution. The calibration is conducted in the redshift space.

The 2PCF describes the excess probability of finding a galaxy pair compared to the random uniform distribution. For AM galaxies from the simulation box, we employ the PH estimator. DD is the normalised galaxy-galaxy counts, and RR is the analytical expression of the normalised Random-random counts.

For observations, the LS estimator is used. DR is the normalised galaxy-random counts and RR is the normalised random-random counts. They both need the help of a random catalogue to complete the counting. But why do we choose the correlation function to calibrate the galaxy bias model?

These are expressions of the density contrast. Affected by the Redshift space distortion, the density contrast is a function of the peculiar velocity.

With the linear bias theory, the multipoles of the correlation function is shown to be affected by the galaxy bias. The influence of galaxy bias (that is to say, the change of the SHAM model) is stronger on the monopole than on the quadrupole.

There is a small conclusion to the first part: the galaxy bias enables the precise description of galaxy clustering with the dark matter theory. The Subhalo abundance matching method select halos according to the empirical bias models and assign one galaxy in those halos. The bias model is calibrated by the 2PCF of the observation. The 2PCF multipoles are functions of the bias and the peculiar velocity.

The second section is to describe the implementation of SHAM. I’ll present the data used in this study, the process of SHAM method and the bias model calibration.

The halo-subhalo catalogues are from the Universe N-body simulations for the Investigation of Theoretical models from galaxy survey, i.e., the UNIT simulation. The simulation models a LCDM universe from z=99 to z=0 and provides 128 snapshots. The selected snapshot one is for ELG and the second one is for LRG. Vpeak is employed to represents the halo mass because it’s more robust for subhalos.

The observation data for the SHAM model calibration is the pairwise-inverse-probability and the angular up-weighted galaxy pairs. Those two method is employed to correct the fibre collision effect in the correlation function on small scales. Due to the finite physical size of fibres, some galaxies in between their covering area cannot be observed. The PIP effect is for those pairs that have one galaxy missing there and the ANG is for the missing pairs there.

The useful properties of galaxies for SHAM are presented here. The effective redshift determined the snapshots of UNIT and the effective number density determines the number of SHAM galaxies.

There are several methods to estimate the statistical errors of the 2PCF. We choose the EZmocks to help us. It employs the effective Zel’dovich approximation and a compound bias model to quickly generate galaxy catalogues at a give redshift. In my project, the covariance matrices of one tracer in NGC or SGC are calculated with 1000 realisations of EZmocks.

As I introduced before, the SHAM galaxy catalogue should follow empirical distributions in which the most massive halos are sure to have an LRG inside. However, to avoid the BOSS samples, the i-band luminosity of the eBOSS LRG has a lower limit. It means that some massive galaxies might be removed.

Taking this effect into account, the modified PDF of LRGs also has a peak as the ELG PDF. That is to say, the SHAM models for eBOSS LRG and ELG are the same. Then I’ll explain how to achieve such distribution.

If we have a halo catalogue, truncate its massive end after Vcut and select the most massive ones from the remaining catalogue, the PDF will be a rectangle.

But if we scattering the Vpeak before the truncation, the acute turings become smooth and form a peak as we expect in the empirical distribution.

To conclude, the SHAM is implemented with Vpeak. By scattering it with a Guassian distribution and then cut its massive ends, the remaining Nth largest Vpeak,scat is halos and subhalos that can host a central galaxy. So how does those parameters affects the correlation functions and Vpeak probability distribution?

We change the scattering parameter sigma and the massive-truncation parameter Vcut by 50% respectively, their correlation function monopoles and quadrupoles varies like this. From the figure, the smaller sigma have a similar effects on the correlation functions as a larger Vcut. Because both of them mean the catalogue will include more massive halos that have a larger bias. This also implies that those two parameters are possibly degenerate.

The effect of smaller sigma and larger Vcut can be illustrate more clearly in the Vpeak PDF. We can see that their most-probable Vpeak both shift to the massive end. I’ll explain why there are error bars in those distributions in the next two slides.

I’ve introduced the method to realise one SHAM galaxy catalogue in the implementation part. Its 2PCF will be calculated with the PH estimator in the redshift space.

However, due to the statistical fluctuations brought by different random seeds, the correlation function and Vpeak PDF has a scattered range. The fluctuation makes it difficult to distinguish the effect caused by the fluctuation and a small change of the parameter set. It will destroy the precision of our SHAM model calibration.

To reduce the effect of statistical fluctuation, we take the averaged correlation function over 20 realisations as the SHAM model for this parameter set.

And use the Monte-Carlo Nested Sampling technique to find the maximum-likelihood parameters and their posteriors. Its results are our eBOSS SHAM galaxy models. The results of the chi2 minimiser -- iminuit are also presented as a contrast.

Here comes the exciting part. I’m going to introduce the models for ELG and LRG. Additionally, a new effect -- the redshift uncertainty is considered to optimise our present model.

The small table on the top shows the maximum-likelihood parameters, the corresponding chi2 and reduced chi2. The data vector is the combination of monopole and quadrupole, and it has 40 elements in total. This figure is the correlation function of SHAM models compared with observations with errors. Monopole is on the left hand side and quadrupole is on the right hand side. The standard is the cyan line produced by the multinest. For ELGs in NGC, the SHAM model fits well even though the observation is not smooth.

The Vpeak PDF on the left hand side and DF are on the right hand side. The most probable Vpeak or the typical Vpeak of th ELG halos is 230 km/s.

This is a triangle plot of the posterior distribution of ELGs in NGC. The subplot on top of the 2D posterior is the posterior of sigma. The Vcut posterior is on its right. The narrow inclined confidence interval tells us that sigma and Vcut are highly degenerate.

For ELG in SGC, the SHAM fitting of correlation function is also good.

The Vpeak PDF also shows a typical value at 230km/s.

Unlike two big peaks, The posterior in SGC has a significant main peak, but the degeneracy remains.

Then let’s see the result for LRGs in NGC. Although the monopole matches well, the quadrupole of SHAM galaxies on small scales has a significant deviation from the observation. This discrepancy also embodies in the reduced chi2 value.

With the present SHAM LRG model, the Vpeak has typical value at around 700 km/s.

The posterior doesn’t present a strong sign of degeneracy. But it does point out that the prior of LRG should be extended.

For LRG in SGC, the discrepancy for quadrupole on small scales remains, even though the reduced chi2 is not a large as that in NGC. The monopole still matches the observation well.

With the maximum-likelihood results, the typical Vpeak for LRG in SGC is around 660 km/s.

The posterior distribution of Vcut is quite irregular, but the degeneracy feature remains.

Now we have a problem: how to solve the quadrupole discrepancy without affecting the monopole. As implied in the bias--multipole relations and the parameter impact plot, the attempts to play with bias models will strongly affect the monopole instead of the quadrupole. So it’s not a favourable solution. What else?

We know that the measured redshift is the combination of true redshift, the line-of-sight peculiar velocity effect, and the measurement uncertainty. The last one has been omitted in the study. To reduce this fluctuation, multiple observations for one target are necessary. The eBOSS working group measures the redshift difference for more than 8000 of LRG pairs. They found a Gaussian function can well represent their distribution. The large dispersion indicates that the uncertainty can have an effect on small scales.

To add this effect to the SHAM processes, we smear the peculiar velocity with a Gaussian function. The aim is to model the measurement uncertainty by introducing uncertainty in the peculiar velocity along the line of sight. Since the change of peculiar velocity has a stronger effect on the quadrupole than monopole, we expect it can solve our problem.

By simply adding a fixed Gaussian smearing, the chi2 is significantly reduced. The new maximum-likelihood results are within the 1-sigma confidence interval of the previous result.

To summarise, in this study, a SHAM method is applied to the UNIT halo catalogue. Considering the target selection effect on eBOSS LRGs and the quenching process for ELGs, they have the same Vpeak model. Calibrated by the 2PCF of eBOSS observations, the SHAM Vpeak models for ELGs and LRGs are presented blow. The redshift uncertainty effect greatly improved the fitting of LRG models in the quadrupole.

So until now, I’ve had a reliable SHAM pipeline for eBOSS galaxies. The next step is to make the pipeline more robust. The first point is to add the peculiar smearing parameter sigma\_pec to the present model. Meanwhile, since sigma and Vcut are highly degenerate, so we might have to find another two independent parameters. The second point is to test the new model in different redshift bins.

After I get the reliable LRG and ELG SHAM models, I’ll continue to multi-tracer SHAM. It will

Enable us to have different kinds of galaxies within one SHAM process. The difficulty here would be assign galaxies in the mass range where there are more than one kind of galaxies. The following step will be the cross-correlation study and error analysis. That’s all for my work.