

MASTER THESIS

Exploring the impact of baryons on cosmological observables

Institute for Theoretical Physics

Utrecht University
Utrecht

Author
Úna Parkinson



**Utrecht
University**

First Reviewer
Dr. Elisa Chisari
Institute for Theoretical Physics
Utrecht University
Utrecht

Second Reviewer
Dr. Tomislav Prokopec
Institute for Theoretical Physics
Utrecht University
Utrecht

Utrecht, January 2023

Exploring the impact of baryons on cosmological observables

Úna Parkinson

January 2023

Abstract

Most predictions for matter distribution in the universe only consider dark matter. However, there have been several recent improvements in the quality of observational data. This means that we can no longer ignore the effects of luminous matter, that is to say baryons, on models of large-scale structure. We investigate a halo model of the universe which includes baryons, and see what their effects are on the matter power spectrum. Including baryons adds many free parameters to the model. We vary all the parameters to see which have the largest effect, and look for ways to constrain them. We do a Fisher forecast to explore the uncertainty on the most important of these. We show that adding priors from observational data helps mitigate the impact of the baryonic parameters.

Contents

1	Introduction	1
2	The model	4
2.1	Halo models	4
2.2	Baryonic halo model	4
3	Changes made to baryonic halo model	10
3.1	Gas density profile	10
3.2	Other changes	10

4	Free parameters	11
4.1	Stellar parameters	11
4.2	Gas parameters	13
5	Observational constraints	17
5.1	Observational constraints on $\bar{\rho}_*(z)$	17
5.2	Simulation inspiration	18
5.3	Schechter method for $n(m)$	18
6	Angular power spectrum	20
7	Fisher forecasting	22
7.1	Initial analysis	23
7.2	Investigating differentiation methods	24
7.3	Uncertainty on $\bar{\rho}_*$	26
7.4	Setting a prior	27
7.5	Fixing the third parameter	28
7.6	A different parameter	30
8	Discussion and Conclusions	33
A	Matter power spectrum code	38

1 Introduction

The distribution of all matter in the universe is usually modelled by what are called ‘halo models’ [1]. These models describe the non-linear gravitational clustering of matter, by assuming all matter to be contained within hard, spherical ‘halos’. Most of these predictions only consider dark matter. However, recent and upcoming improvements in the quality of observational data, for example LSST (Legacy Survey of Space and Time) [2], mean that we can no longer ignore the effects of baryons on these models. We will have measurements to 1% uncertainty, and simulations have shown that baryonic feedback can alter the matter distribution by a much larger percentage [3], so it is important they be included. Luminous matter is known to redistribute matter [4]. For example, supermassive black holes tend to heat and push gas around overdensities of dark matter. This causes a change in the density map.

This project will aim to further develop efficient models which account for baryonic impact on cosmological observables, based on previous developments [5, 6]. We start with a halo model [1] which has been modified to include baryons, that is to say, stars and hot gas [6]. We use a Python package called `pyccl` to reconstruct this model for our own purposes. It is necessary to make some small changes to the model as presented in [6], so that we can reproduce the results practically. This mostly consists of making it compatible with the way we code the model.

A halo model relies on parametrisations of the ‘halo profile’ of each component, which is the radial profile of its density within the halo. Here we have three different profiles. Using these, the model is constructed by considering the correlation between two particles, (i) within the same halo, and (ii) within different halos. It is further complicated by the fact that gas also has a diffuse component which exists outside the halos altogether [6]. This quantity must also be considered in constructing contributions to the model. Consequently, there are a lot of free parameters, most of which are not observationally constrained. We find out which of these parameters are most important for the model, and attempt to constrain them.

We investigate this by looking at their effect on the matter power spectrum. This is a quantity which measures the clustering, or amount of structure, of matter [7]. It is therefore useful to look at when considering matter distri-

bution. It is noticeably different when baryons are included in the model compared to when they are not [6]. The baryons have a small contribution, not appreciably changing the shape of the matter power spectrum, but they do make it overall larger, and this is significant. It turns out that some free parameters affect the model more than others, so we want to constrain these to make the model as accurate as it can be. The goal is to find ways to constrain this model using observations of baryonic matter.

Once we have found the parameters with the largest effect on the matter power spectrum, we look for observational data with which to constrain them. This is crucial if we want to know specifically how the baryons affect the model. Currently, in these kinds of models, the parameters are primarily just fitted to match simulations of the universe’s large-scale structure [6]. Unfortunately, this means we do not know exactly how the models reflect reality, only having the results of simulations to compare with what we can observe. As dark matter is not observable except where it gravitationally affects visible matter, we have to rely on measurements of baryonic observables. Therefore, we want to use observations of baryonic quantities to give these parameters their own values, so they do not have to be left free anymore.

For the analysis of the parameter uncertainty, we will also need a quantity called the *angular power spectrum*. This is a Fourier transform of the angular correlations between points on the sky [7, 8]. It contains all the information in one function and can be calculated in `pyccl` using the matter power spectrum. We then do a *Fisher forecast* on the model. This is a way of investigating the uncertainties on the parameters. It involves differentiating the angular power spectrum with respect to the parameters, and encodes the uncertainty on and correlation between parameters. We explore the uncertainties on two of the most significant parameters, and compare with cosmological parameters which are better constrained. Using a visualisation technique known as *confidence ellipses*, we show how pairs of parameters are correlated. The Fisher formalism also allows us to add priors based on the observations, so that we can lower those uncertainties. We find that adding a prior on a parameter can reduce uncertainty, on that parameter but also on the others. However, it can also make parameters less strongly correlated.

We first introduce the model in section 2. Then we discuss any changes made to the pre-existing halo model in section 3. Following this, section 4

investigates the effect of each free parameter of the baryonic halo model. In section 5, we explore possibilities for constraining the model with observational data. Next, in section 6, we introduce the concept of an angular power spectrum, as this will be needed for the uncertainty analysis. We then perform a Fisher analysis on the most significant parameters in section 7. Finally, section 8 discusses and summarises our findings.

2 The model

2.1 Halo models

Halo models are a way of describing the non-linear gravitational clustering of matter. Linear and perturbation theories cannot describe the clustering of matter at scales smaller than a few megaparsecs (Mpc), so we must use another method to investigate the highly non-linear range of gravitational clustering. All matter is considered to be contained within hard, spherical, virialised ‘halos’. A halo is a region which was sufficiently overdense that it collapsed [1]. With an initially smooth universe, matter over time clusters into structures including large filaments, sheets, and knots. These dense knots are often referred to as dark matter (DM) halos. Most halo models just concern themselves with DM.

In a DM halo model, there are two contributions to the model. The “1-halo” term concerns the correlation between two particles within the same halo, while the “2-halo” term concerns the correlation between two particles in two separate halos. These are added together to obtain the full DM halo model.

Halo models do have some limitations [1]. We assume they are all spherically symmetric, with smooth parameterised profiles, but even at a fixed mass there is variation in halo profile shape. Also, in simulations the halos do not have smooth profiles and are rarely symmetric. We just have to recognise that this assumption may be inaccurate and take all the predictions of the model with this as a caveat. Halo models are still a very useful way of describing the distribution of all matter in the universe. Until recently, halo models have only considered DM, as it makes up the majority of the matter in the universe. However, with current and near-future observational capabilities, we can no longer ignore the baryonic contribution. Therefore we want to include the contribution of baryonic components within the halo model.

2.2 Baryonic halo model

We start with a (semi-analytic) baryonic halo model as presented by Fedeli [6]. In this model, about 80% of all matter is DM, while the remaining 20%

is a mixture of gas and stars. These baryonic components can have an effect on the distribution of matter through several processes. For example, the gas can exchange energy with a radiation background, leading to radiative cooling. The stars, for their part, can inject energy and also metals into the gas, which alters the entropy in the system and affects the radiative properties. The energy feedback from Active Galactic Nuclei (AGNs) is also expected to affect baryonic matter.

The matter power spectrum of each component can be related to its abundance and distribution within the gravitational potential wells of DM halos. It is useful to consider a model like this to study the effect of baryons on cosmological parameter estimation, because hydrodynamical simulations cannot do this. We use the cosmological parameter values in [6].

As a measure for the average energy density of matter in the universe, we use the density of an *equivalent mass*, m . This is calculated as if in a universe with only DM,

$$\bar{\rho}_m = \int_0^\infty dm m n(m, z). \quad (1)$$

We will consider the equivalent average energy density as being equal to the sum of the average densities of all the components,

$$\bar{\rho}_m = \bar{\rho}_{\text{DM}} + \bar{\rho}_g(z) + \bar{\rho}_*(z). \quad (2)$$

The average density of each component is then defined according to their respective mass fractions. These are defined with respect to the equivalent mass as follows:

$$f_i(m, z) = \frac{m_i(m, z)}{m}, \quad (3)$$

where $m_i(m, z)$ is the mass within a bound structure whose equivalent mass is m . The specific mass fractions are parametrised as:

$$f_{\text{DM}}(m, z) = 1 - \frac{\Omega_{b,0}}{\Omega_{m,0}}, \quad (4)$$

$$f_*(m, z) = A \exp \left[-\frac{\log^2(m/m_{0s})}{2\sigma_s} \right], \quad (5)$$

$$f_g(m, z) = \frac{\Omega_{b,0}}{\Omega_{m,0}} \text{erf} \left[\frac{\log(m/m_{0g})}{\sigma_g} \right]. \quad (6)$$

For DM and stars, using this to define the average density is straightforward, as all of the DM and stellar mass is contained within halos,

$$\bar{\rho}_{\text{DM}} = \int_0^\infty dm m f_{\text{DM}}(m, z) n(m, z), \quad (7)$$

$$\bar{\rho}_*(z) = \int_0^\infty dm m f_*(m, z) n(m, z). \quad (8)$$

However, for the gas component, there is also a certain amount which is diffused in the space between bound structures, so the average gas density must be multiplied by the fraction of gaseous matter which does reside within halos to be defined in the same way,

$$F_g(z) \bar{\rho}_g(z) = \int_0^\infty dm m f_g(m, z) n(m, z). \quad (9)$$

For the actual total gas energy density, we can define it by using the total baryon fraction and subtracting the stellar part,

$$\bar{\rho}_g = \frac{3H_0^2}{8\pi G} \Omega_{b,0} - \bar{\rho}_*(z). \quad (10)$$

The density profiles of each component within a halo are then defined. We use the standard Navarro-Frenk-White (NFW) profile for the DM component [9],

$$\rho_{\text{DM}}(x|m) = \frac{\rho_s}{x(1+x)^2}, \quad (11)$$

where $x = r/r_s$. The two parameters ρ_s and r_s can be related by requiring that the mass fraction within one halo must match that in the universe overall, thus eliminating a free parameter.

The stellar profile is an exponential profile as noted in [6],

$$\rho_*(x|m) = \frac{\rho_t}{x} \exp(-x^\alpha), \quad (12)$$

where α is set to 1 for simplicity, and $x = \frac{r}{R_\Delta} x_\Delta$. The two parameters ρ_t and x_Δ can be related in the same way as for the DM profile.

The gas density profile as defined in [6] did not actually work in our computations, so we substituted a different profile of the same type. This is shown

in subsection 3.1.

The power spectrum of each component was calculated using the integrals defined in [6]. These take the form of combinations of

$$I_{2p,ij} = \int_0^\infty dm n(m) m^2 y_i(k|m) y_j(k|m) \quad (13)$$

and

$$I_{bp,i} = \int_0^\infty dm n(m) b(m) m y_i(k|m), \quad (14)$$

where i and j refer to which component is being calculated, and y is the profile divided by m . Each component (DM, stars, gas) has different contributions to the matter power spectrum. The DM and stellar contributions are simply the 1-halo and 2-halo terms added together for each component,

$$P_{\text{DM}}^{(1)}(k) = \frac{1}{\bar{\rho}_{\text{DM}}^2} I_{2p,\text{DMDM}}, \quad P_{\text{DM}}^{(2)}(k) = \frac{P_L(k)}{\bar{\rho}_{\text{DM}}^2} [I_{bp,\text{DM}}]^2, \quad (15)$$

$$P_*^{(1)}(k) = \frac{1}{\bar{\rho}_*^2} I_{2p,**}, \quad P_*^{(2)}(k) = \frac{P_L(k)}{\bar{\rho}_*^2} [I_{bp,*}]^2, \quad (16)$$

while the gas is more complicated. It has the 1- and 2-halo terms,

$$P_{g,h}^{(1)}(k) = \frac{1}{F_g^2 \bar{\rho}_g^2} I_{2p,gg}, \quad P_{g,h}^{(2)}(k) = \frac{P_L(k)}{F_g^2 \bar{\rho}_g^2} [I_{bp,g}]^2, \quad (17)$$

but also a term for the diffuse component and the cross-correlation between the diffuse component and the halo,

$$P_{g,d}(k) = b_d^2 P_L(k), \quad P_{g,dh}(k) = b_d \frac{P_L(k)}{F_g^2 \bar{\rho}_g^2} I_{bp,g}. \quad (18)$$

The diffuse component is simply considered to have a linear power spectrum, modified by a bias constant b_d . The gas components are also not simply added, but are combined linearly as follows:

$$P_g(k) = (1-F_g)^2 P_{g,d}(k) + 2F_g(1-F_g) P_{g,dh}(k) + F_g^2 \left[P_{g,h}^{(1)}(k) + P_{g,h}^{(2)}(k) \right]. \quad (19)$$

There are also cross contributions from all three components. The DM-stars cross-correlation just has a 1-halo and 2-halo contribution,

$$P_{\text{DM}*}^{(1)}(k) = \frac{1}{\bar{\rho}_{\text{DM}}\bar{\rho}_*} I_{2p,\text{DM}*} \quad P_{\text{DM}*}^{(2)}(k) = \frac{P_L(k)}{\bar{\rho}_{\text{DM}}\bar{\rho}_*} I_{bp,\text{DM}} I_{bp,*}. \quad (20)$$

The DM-gas one has those and also a diffuse-halo contribution,

$$P_{\text{DM}g,h}^{(1)}(k) = \frac{1}{\bar{\rho}_{\text{DM}}\bar{\rho}_g} I_{2p,\text{DM}g} \quad P_{\text{DM}g,h}^{(2)}(k) = \frac{P_L(k)}{\bar{\rho}_{\text{DM}}\bar{\rho}_g} I_{bp,\text{DM}} I_{bp,g}, \quad (21)$$

$$P_{\text{DM}g,dh}^{(1)}(k) = b_d \frac{P_L(k)}{\bar{\rho}_{\text{DM}}} I_{bp,\text{DM}}, \quad (22)$$

and is combined as follows:

$$P_{\text{DM}g}(k) = (1 - F_g) P_{\text{DM}g,dh}(k) + F_g \left[P_{\text{DM}g,h}^{(1)}(k) + P_{\text{DM}g,h}^{(2)}(k) \right]. \quad (23)$$

The stars-gas cross-correlation is similar,

$$P_{*g,h}^{(1)}(k) = \frac{1}{\bar{\rho}_*\bar{\rho}_g} I_{2p,*g}, \quad P_{*g,h}^{(2)}(k) = \frac{P_L(k)}{\bar{\rho}_*\bar{\rho}_g} I_{bp,*} I_{bp,g}, \quad P_{*g,sh}^{(1)}(k) = b_d \frac{P_L(k)}{\bar{\rho}_*} I_{bp,*}, \quad (24)$$

$$P_{*g}(k) = (1 - F_g) P_{*g,sh}(k) + F_g \left[P_{*g,h}^{(1)}(k) + P_{*g,h}^{(2)}(k) \right]. \quad (25)$$

All the contributions are combined by adding them together, weighting each component based on their average density, to get the total matter power spectrum [6],

$$P(k) = \frac{1}{\bar{\rho}_m^2} [\bar{\rho}_{\text{DM}}^2 P_{\text{DM}}(k) + \bar{\rho}_*^2 P_*(k) + \bar{\rho}_g^2 P_g(k) + 2\bar{\rho}_{\text{DM}}\bar{\rho}_* P_{\text{DM}*}(k) + 2\bar{\rho}_{\text{DM}}\bar{\rho}_g P_{\text{DM}g}(k) + 2\bar{\rho}_*\bar{\rho}_g P_{*g}(k)]. \quad (26)$$

This gives us the total matter power spectrum which, when plotted relative to the (DM-only) non-linear matter power spectrum (as coded by Code for Anisotropies in the Microwave Background (CAMB)), shows that baryons can have an impact (Figure 1). Soon we will have new observations which will measure the matter distribution to within 1% uncertainty [2]. The effect of the baryons is larger than this [3], as can be seen in Figure 1, so is necessary to be included in models going forward.

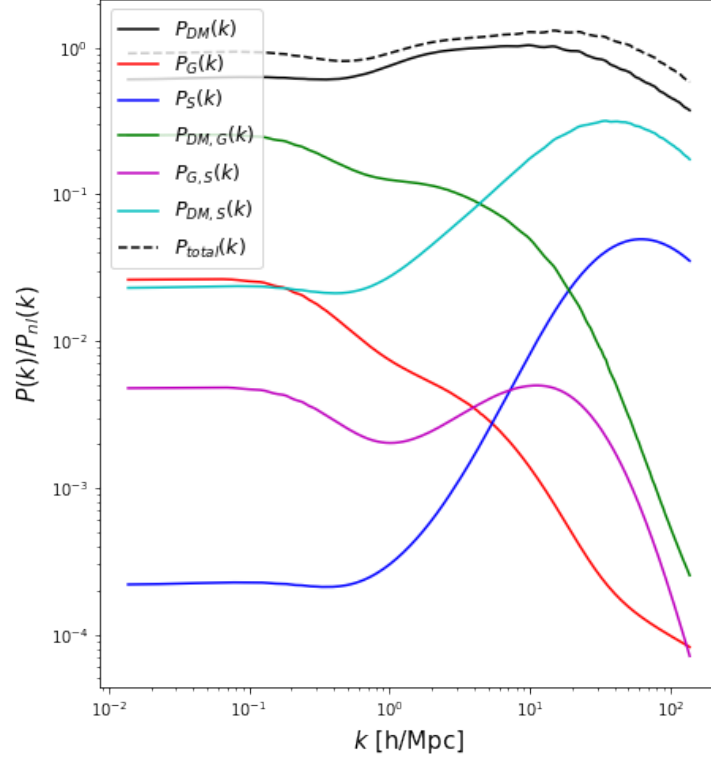


Figure 1: Total matter power spectrum relative to the non-linear matter power spectrum, with components.

The model was coded in Python, using a package called `pycc1` (Python Core Cosmology Library). This package has some built-in cosmology functions and integrals, so it was very useful. See Appendix A for the notebook which calculates the matter power spectra.

3 Changes made to baryonic halo model

There were some parts of the model in [6] which we changed in the course of this work. Not everything suited our purposes exactly, so we used different formulations of things here and there.

3.1 Gas density profile

The most significant change made was using a different β -profile for the gas than used in [6]. This is because with our code it did not work as a profile. Instead, we used a β -profile from Schneider [5], which produces a very similar result to the Fedeli profile, but could actually be implemented in our Python code. It is also a more complex profile, with not only a β parameter and core radius, but also an ejection radius:

$$\rho_g(r) = \frac{\rho_{g,0}}{(1+u)^\beta(1+v^2)^{(7-\beta)/2}}, \quad (27)$$

where $u = r/r_{co}$ and $v = r/r_{ej}$. We then do the mass fraction matching as in [6], constraining the $\rho_{g,0}$ parameter.

3.2 Other changes

Some definitions such as the virial mass and the concentration-mass relation, as well as the halo mass function and halo bias, were defined using inbuilt functions in `pyccl`, rather than taking them from the sources used in [6]. The NFW profile for dark matter is also an inbuilt halo profile in `pyccl`, so this was used instead of the explicit parametrisation in [6].

We also interpolated over each of the profiles to allow for better numerical integration. This involves selecting an amount of values of the function and then creating a smooth function based on those points, so that the numerical integration is easier. It made the code run faster, and smoothing out the functions left less room for numerical instabilities.

Everything else in the calculation of the matter power spectra follows Fedeli [6].

4 Free parameters

The introduction of the baryons to the halo model adds many free parameters. In this section, we explore the effects of varying each free parameter on the matter power spectrum. The actual values chosen for upper and lower bounds are purely illustrative, and are not intended to be specific percentages of difference. The fiducial values are taken from [6] and [5], unless otherwise stated.

The parameters have been grouped based on whether they affect the stellar or gas component, though the average density parameters of both are directly linked (Equation 2) and only effectively count as one free parameter. The aim here is to discover which parameters change the total matter power spectrum the most, so that we know which ones are the most important to constrain.

4.1 Stellar parameters

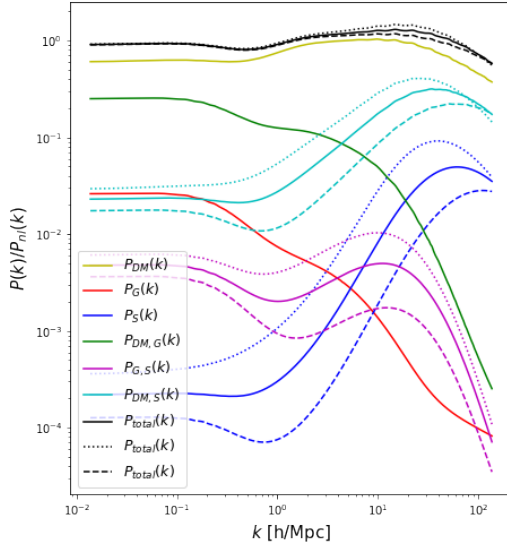


Figure 2: m_{0s}

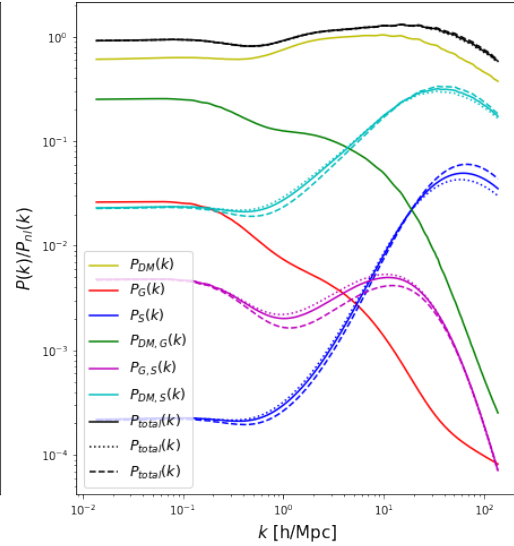


Figure 3: σ_s

m_{0s} : The mass at which star formation peaks. Varying it affects the stellar power spectrum, so it mostly only changes the total power spectrum at

higher k (smaller scales). Above a k -value of ~ 10 , we do not expect to have high-quality observational data, so this part of the spectrum is not our main focus. Varied by: an order of magnitude either side.

σ_s : This parameter determines how wide or narrow the stellar mass fraction is. It therefore controls the sharpness of the star formation peak. Varying this has a negligible effect on the total matter power spectrum, and what small effect it does have is restricted to high k . Varied by: ± 0.4 .

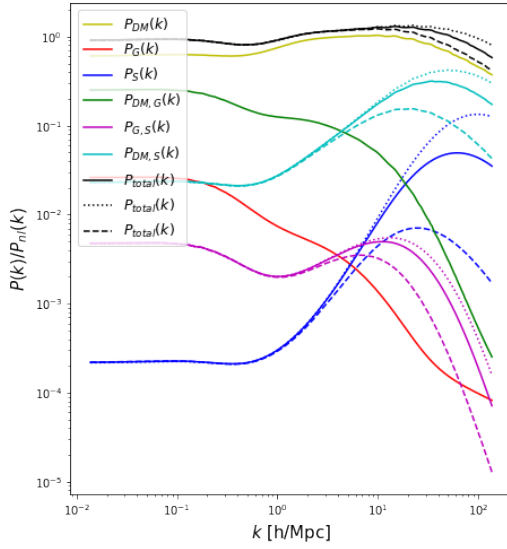


Figure 4: x_{Δ_s}

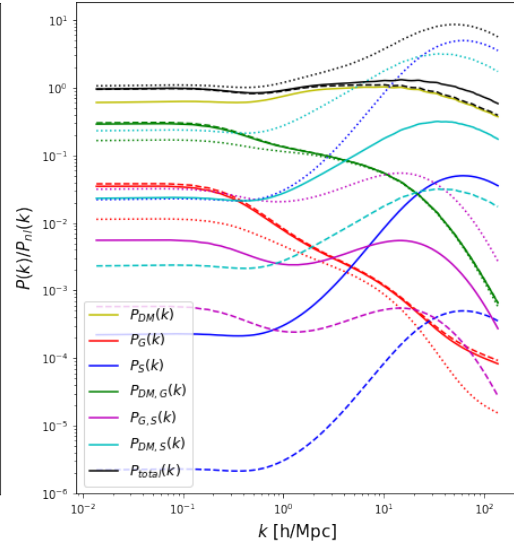


Figure 5: $\bar{\rho}_*$ ($\bar{\rho}_g$)

x_{Δ_s} : Depends (inversely) on the radius scale which lies between the inner and outer parts of the exponential disk profile. Only has an effect at large k , so this is not the most important parameter to constrain. If the transition between inner to outer happens over a larger distance, this parameter is smaller, which means the power spectrum (at large k) is also smaller. Varied by: ± 20 .

$\bar{\rho}_*$: The average stellar density across the universe. Changing this also affects $\bar{\rho}_g$, as they are defined (with the constant $\bar{\rho}_{DM}$) as part of an unchanging total average matter density (see Equation 2). There is a significant effect on the total matter power spectrum, so it will be important to constrain this quantity. Most of the effect is at large k , but it has an effect across all scales,

so is significant overall. This parameter was varied while keeping the total baryon fraction Ω_b and mass fraction Ω_m constant. It would also be possible to vary it by changing the cosmological parameters, but this would then have to propagate through the whole model and have a much more complex effect. Varied by: an order of magnitude either side.

4.2 Gas parameters

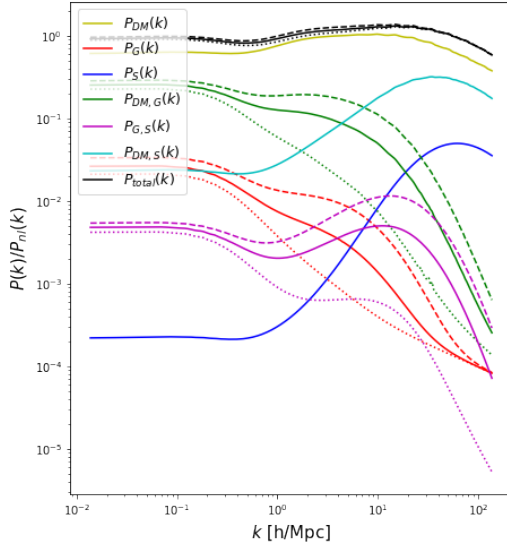


Figure 6: m_{0g}

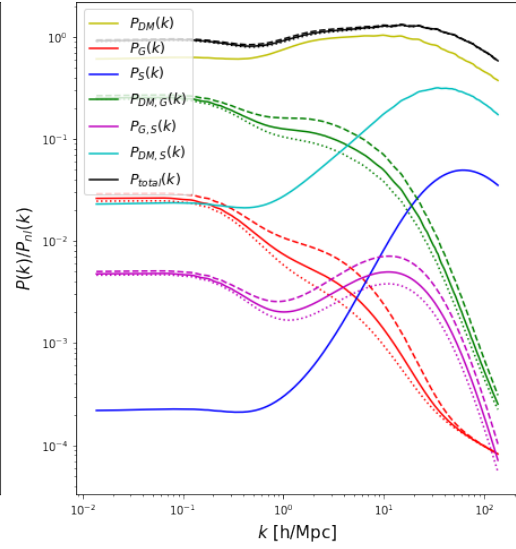


Figure 7: σ_g

m_{0g} : Mass at which gas fraction drops. If gas exists at lower masses, this means the power spectrum gets larger, as there is more range of mass which includes gas particles. Has a noticeable effect on the total power spectrum, so it should be important to constrain this, especially because of the impact at low and mid-range k . It should be possible to measure this quantity, all that would be required would be the ranges of mass density which have a gaseous component. Varied by: an order of magnitude either side.

σ_g : This parameter defines how sharp the drop is in the mass fraction at lower masses. This also has an effect at the low and middle ranges of k . If the drop is steeper, then the fraction is higher at the lower masses above the

drop, and this makes the power spectrum smaller overall, since the larger masses have a smaller contribution. Could possibly be measured through observation of the gas profile. Varied by: ± 1 .

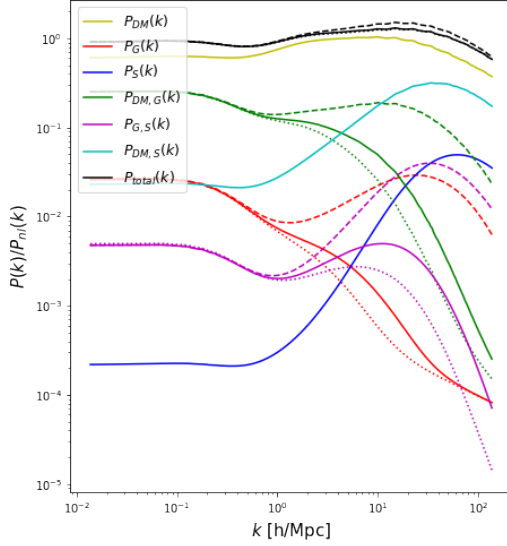


Figure 8: r_{co}

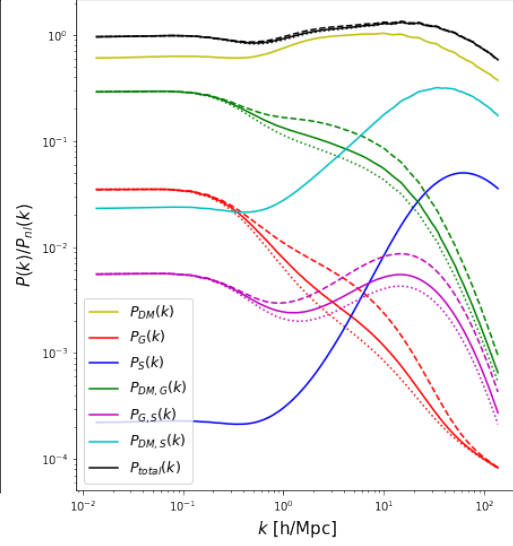


Figure 9: r_{ej}

r_{co} : This is the radius of the gas profile's core. It only affects the power spectrum for large k , which makes sense as it only relates to the smaller scales of the profile. This means it is not particularly important to constrain, since our measurements will not be that precise up to this small of scales. If the gas profile is in reality a cored profile, then we can definitely measure the core radius. Its dependence on the size of the halo, however, through its definition with respect to R_Δ , may make it complicated to constrain. Varied by: ± 0.09 .

r_{ej} : Radius of gas ejection from the profile. There is no effect at linear scales, but has the most effect in the mid-range, so it may be important to constrain. However, the effect on the total matter power spectrum is small. There is a constraint in [5] that sets prior assumptions of values between [2, 8]. Problems are caused with the computation of the profile if $M = m_{0g}$ and/or if $r_{ej} < 2$, so not all values of this parameter are possible in our construction of the model. Fiducial value: $4.5R_\Delta$. Varied by: $+3.5R_\Delta, -2.5R_\Delta$.

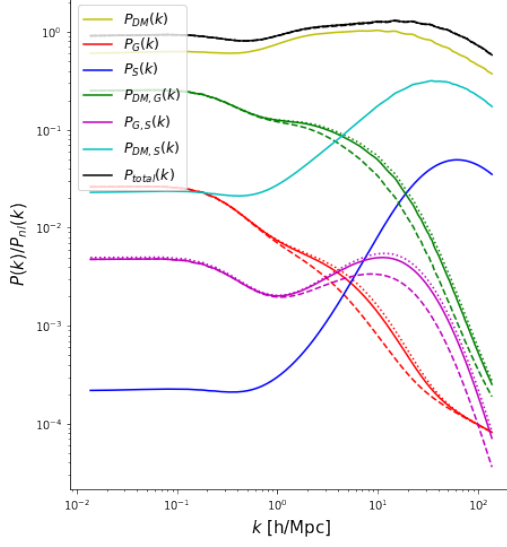


Figure 10: β

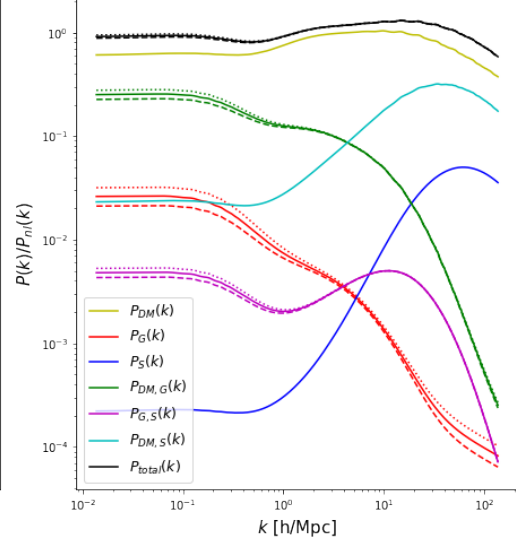


Figure 11: b_d

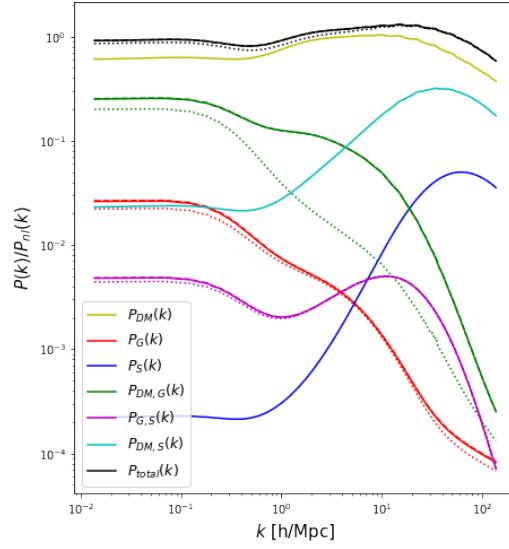


Figure 12: F_g

β : Measures the slope of the decrease of the power-law gas profile. It depends on halo mass, so is larger (steeper) for higher masses. This has a very slight

effect on the mid-to-larger ranges of k in the power spectrum. It already has a constraint of $\beta \leq 3$ [5]. Varied by: $+0.1, -0.4$.

b_d : This is a bias parameter which deals with the diffuse component of the gas. The diffuse component just has a linear power spectrum, scaled by b_d , so this parameter only affects the power spectrum at linear scales, simply scaling that part up or down. There seem to be constraints on b_d which say it is approximately equal to or less than 1 [6]. Varied by: ± 0.1 .

F_g : The fraction of gas which exists within halos. This means that it affects how much of the gas contributes to the non-linear part of the matter power spectrum. If it is larger, there is less gas in the diffuse component, so the power spectrum is smaller at linear scales. It has an effect at small and middle ranges of k . Varied by: an order of magnitude either side.

5 Observational constraints

We now start to look for observational data with which to constrain the most important free parameters.

5.1 Observational constraints on $\bar{\rho}_*(z)$

The most useful constraint from observations which exists is for the average stellar density. It is a parameter for which changing it has a large effect on the matter power spectrum, so it is important that we have a way to constrain it. This parameter is exhaustively measured at a range of redshifts, and a large quantity of these measurements (data from 2008-2013) are compiled, including error bars, in what is known as the Madau plot [10]. This is a plot of $\log \bar{\rho}_*$ at different redshifts (see [10] for the figure).

We can use this data to set $\bar{\rho}_*$ at any specific redshift and calculate the matter power spectrum. Some of the source data within the plot is extracted here.

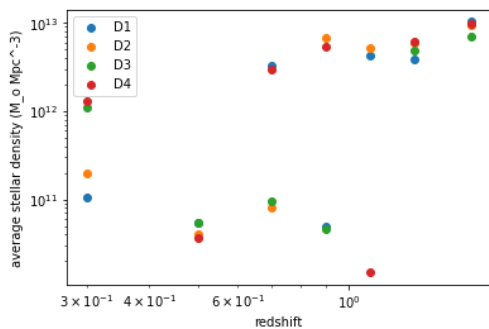


Figure 13: Data from Bielby et al [11]. Consists of four fields of view, which were averaged over to include this data in the Madau plot.

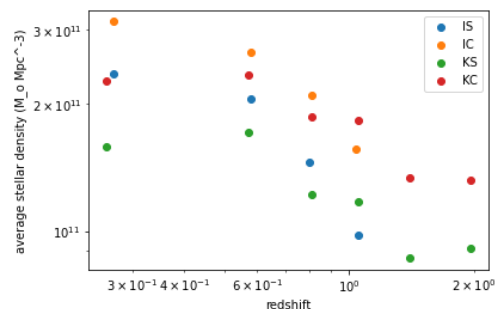


Figure 14: Data from Pozzetti et al [12]. Each colour of data points is from a different sampling method.

Figure 13 shows some of the data used to create the Madau plot. This data is from four separate fields of view in the sky, and appears to have some anomalous points which do not coincide. To avoid bias due to this dependence on field of view, they were averaged before being included in the compilation in [10]. Figure 14 illustrates another set of points which was

included by Madau. These points are from different sampling methods, but they appear to be self-consistent so an average would hopefully remove any bias.

5.2 Simulation inspiration

In an effort to find more sources for observational constraints, we looked at some simulation papers to find out where they got the data for their initial conditions [13][14][15].

Schaye et al, in their paper on the EAGLE project [13], used the galaxy stellar mass function (GSMF) to calibrate the efficiency of their feedback model. The Planck 2013 data was used to constrain the cosmological parameters. They also compare the simulation results to the observed GSMF, the stellar-to-halo mass relation, the sizes of galaxies, and the relation between black hole (BH) mass and stellar mass. They used data from the Sloan Digital Sky Survey (SDSS) and the GAMA survey for these.

McCarthy et al, in their paper on the BAHAMAS project [14], also used data from SDSS and GAMA to calibrate their simulations. They also used X-ray measurements of hot gas to calibrate the hot gas mass fraction-halo mass relation.

We also looked at a paper by Dubois et al [15], but they did not appear to use any observational data to calibrate their simulation.

5.3 Schechter method for $n(m)$

The halo mass function is usually determined by simulations. Because we wanted to make this model more observationally constrained, we looked at luminosity measurements of galaxies. These can be used to calculate values for the GSMF, the stellar counterpart to $n(m)$, as luminosity is directly related to mass by the equation [16]

$$\phi(M) = \phi(L) 0.4 \ln 10 L. \quad (28)$$

The GSMF (or $n_*(m_*, z)$) can be parameterised by a Schechter function [16, 17], with n (usually two or three) terms to capture the multiple phases,

$$\phi(M) dM = \sum_{i=1}^n \exp\left(-\frac{M}{M_i^*}\right) \left[\phi_i^* \left(\frac{M}{M_i^*}\right)^{\alpha_i}\right] \frac{dM}{M_i^*}. \quad (29)$$

The free parameters M_i^* , ϕ_i^* , and α_i^* have been set by fitting to various sky surveys, for example the Galaxy And Mass Assembly (GAMA) survey [18, 19].

$n_*(m_*, z)$ does not actually appear in our equations for the model, as everything is calculated in relation to $n(m, z)$, but these are related, meaning we can obtain one from the other:

$$\int_0^\infty dm_* n_*(m_*, z) = \int_0^\infty dm n(m, z). \quad (30)$$

Because of this, it can also be used to constrain the average stellar mass density, if we follow

$$\implies n(m, z) = n_*(m_*, z) \frac{dm_*}{dm}, \quad (31)$$

and substituting in Equation 8,

$$\implies \bar{\rho}_*(z) = \int_0^\infty dm_* m f_*(m, z) n_*(m_*, z) = \int_0^\infty dm_* m_* n_*(m_*, z). \quad (32)$$

Then, combined with the fact that m_* can be defined through a parametrisation of the stellar-to-halo mass relation [20],

$$m_*(m) = \frac{(m/m_{h,1})^{\beta_1}}{1 + (m/m_{h,1})^{\beta_1 - \beta_2}}, \quad (33)$$

where $m_{h,1}$, β_1 , and β_2 are free parameters fixed by observation, we can define $f_*(m, z)$ by dividing by m . This therefore specifies $\bar{\rho}_*(m, z)$ in another way, relying only on observational data. We did not actually end up using this version of $f_*(m, z)$, because within our model it was not normalised in a way that made it work right, but further work in this direction could be interesting.

6 Angular power spectrum

We will now need to introduce the angular power spectrum, or as it is also known, the C_ℓ s. This can be formulated as the Fourier transform of the angular correlation function [7], analogous to how the matter power spectrum is the Fourier transform of the spatial density correlation function, and is a function of the spherical harmonic index ℓ . It can be calculated from the matter power spectrum, which is three-dimensional, as a kind of projection to two angular dimensions [8],

$$C(\ell) \sim \int_0^{\chi_{\text{lim}}} d\chi \frac{g^2(\chi)}{a^2(\chi)} P(k, \chi), \quad (34)$$

where we integrate over comoving distance χ . We do this using `pyccl`, where there is a function which automatically calculates the angular power spectrum. Also necessary for this is a distribution of galaxies with redshift, so that the projection onto the sky is correct. We use a redshift distribution equal to that detectable by LSST [2].

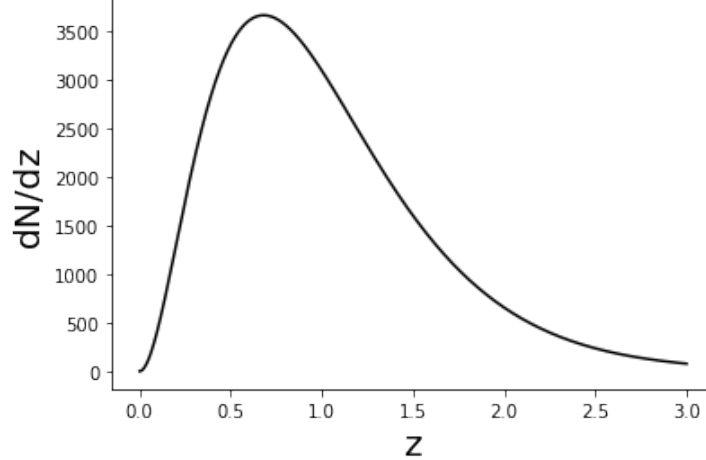


Figure 15: Redshift distribution used, comparable to the LSST redshift distribution [2].

The parameter with seemingly the most possible effect on the matter power spectrum is $\bar{\rho}_*$, so that is the first one to be investigated. When passed onto the angular power spectrum, variation in this parameter has a noticeable

effect, as seen in Figure 16. Here $\bar{\rho}_*$ has been varied an order of magnitude either side of its fiducial value.

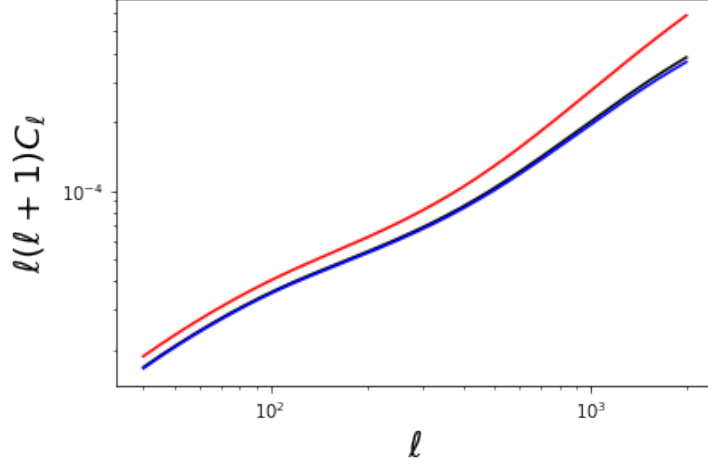


Figure 16: The change in C_ℓ s when $\bar{\rho}_*$ is varied by an order of magnitude either side.

We have calculated our matter power spectrum at a redshift of zero, as we were following Fedeli in [6], so when we calculate the C_ℓ s we need to add a dependence on redshift. Since this project is mostly a qualitative exploration, we do this roughly in `pycc1` by just multiplying the matter power spectrum by the background growth factor squared, which is a function of the scale factor, in our cosmology. This is strictly speaking only applicable for the linear power spectrum, but we extrapolate to give us an idea of what the angular power spectrum should be like, making the assumption that the variation of the free parameters with redshift is negligible within $0 \leq z \leq 3$. The angular power spectrum is necessary for the next section, which is concerned with analysis of the uncertainty on parameters.

7 Fisher forecasting

To get a quantitative idea of how much each free parameter can be constrained, we do a Fisher forecast for some of the most important ones. This involves constructing a *Fisher matrix*, which begins with calculating the angular power spectrum for this model. This is then differentiated numerically with respect to each parameter to form the entries of the matrix [21, 22],

$$F_{\alpha\beta} = \sum_{\ell} \frac{1}{[\text{Cov}]_{\alpha\beta}} \frac{\partial C^{\text{theory}}(\ell, \{\bar{\lambda}_{\gamma}\})}{\partial \bar{\lambda}_{\alpha}} \frac{\partial C^{\text{theory}}(\ell, \{\bar{\lambda}_{\gamma}\})}{\partial \bar{\lambda}_{\beta}}. \quad (35)$$

Here $[\text{Cov}]_{\alpha\beta}$ is an entry in what is called the *covariance matrix*, a measure of the error on the C_{ℓ} s. It is a diagonal matrix, with entries formulated as follows:

$$[\text{Cov}]_{ii} = \frac{2}{(2\ell + 1)f_{sky}} \left(C_{\ell}^i + \frac{\sigma_e^2}{n_{gal}} \right)^2, \quad (36)$$

where f_{sky} is the fraction of sky being observed, σ_e is the intrinsic ellipticity variance of source galaxies, n_{gal} is the number of galaxies per steradian. The σ_e^2/n_{gal} term corresponds to noise.

Each row and column of the Fisher matrix corresponds to a parameter, for example:

$$F = \begin{matrix} & \begin{matrix} \Omega_m & \sigma_8 & \bar{\rho}_* \end{matrix} \\ \begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{12} & F_{22} & F_{23} \\ F_{13} & F_{23} & F_{33} \end{bmatrix} & \end{matrix}. \quad (37)$$

The Fisher matrix reveals the possible uncertainty on each parameter through its inverse. The uncertainty on a parameter x is

$$\sigma_x = \sqrt{[F^{-1}]_{xx}}, \quad (38)$$

where the indices on F^{-1} refer to the diagonal entry corresponding to parameter x [22]. We look at the amount of uncertainty on each parameter, and then add priors based on our constraints to improve this.

Parameters can be fixed or marginalised over simply by removing the parameter in question from the Fisher matrix or its inverse respectively [22]. In

this way, all parameters excluded from the matrix are considered perfectly known. Priors from constraints can be added by just adding a value to the entry corresponding to the necessary parameter. By marginalising over all but two parameters, we can make a visual representation (called a confidence ellipse) of how those two parameters' uncertainty is correlated.

7.1 Initial analysis

We start with an analysis involving three parameters, Ω_m , σ_8 , and $\bar{\rho}_*$. Using numerical differentiation, with $\Delta\Omega_m$ and $\Delta\sigma_8$ as 0.01, and $\Delta\bar{\rho}_*$ as 7×10^6 , we get the following uncertainty values from our Fisher matrix:

- Ω_m uncertainty = 2.68×10^{-5}
- σ_8 uncertainty = 9.24×10^{-3}
- $\bar{\rho}_*$ uncertainty = 1.80×10^8 .

The derivatives were calculated using the formula

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h}, \quad (39)$$

where h is Δx , for each parameter x . As an example of this type of analysis, Figure 17 shows the 1σ uncertainty in the two cosmological parameters purely based on this model.

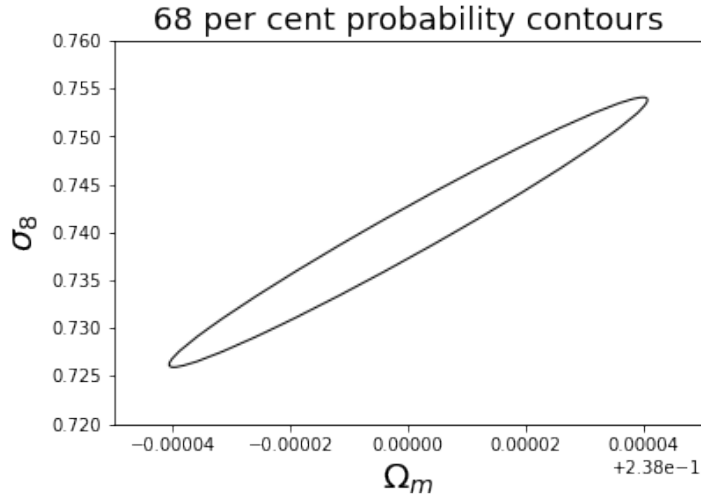


Figure 17: 68% uncertainty confidence ellipse for σ_8 and Ω_m .

The narrow ellipse indicates that they are strongly correlated parameters. Here, $\bar{\rho}_*$ has been marginalised over, that is to say its corresponding row and column have been removed from the inverse of the Fisher matrix [22].

7.2 Investigating differentiation methods

To check the results, we compare with a different numerical differentiation formula,

$$f'(x) = \frac{-f(x+2h) + 8f(x+h) - 8f(x-h) + f(x-2h)}{12h}, \quad (40)$$

which is more precise. This method converged to the same quantity as the previous one for all the parameters except for Ω_m .

The numerical differentiation on Ω_m was inconsistent, with different results depending on the formula used (see Figure 18). Therefore, we tried decreasing the value of h until the methods converged.

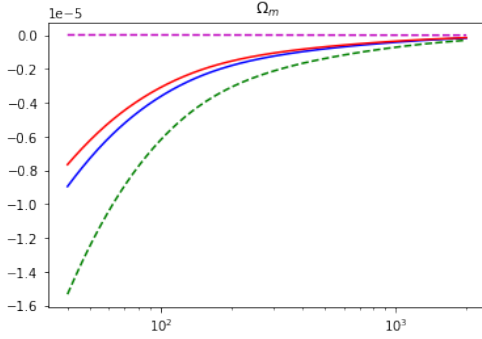


Figure 18: Comparing different numerical differentiation methods for a step size of $h=0.01$. Plot is of $\partial C_\ell / \partial \Omega_m(\ell)$ as a function of ℓ .

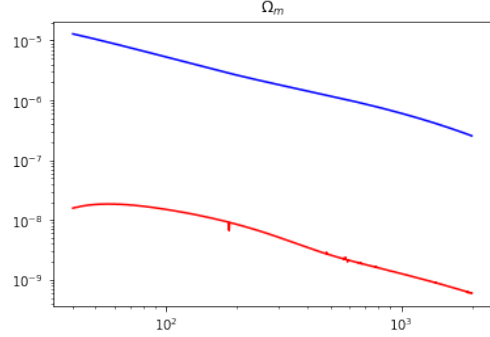


Figure 19: The methods are still separate at a step size of $h=0.001$.

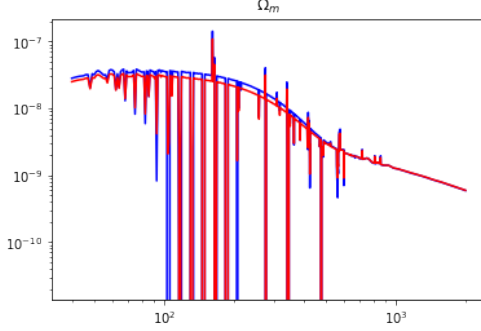


Figure 20: At a step size of $h=0.0001$, the methods almost converge, but there is a lot of numerical instability.

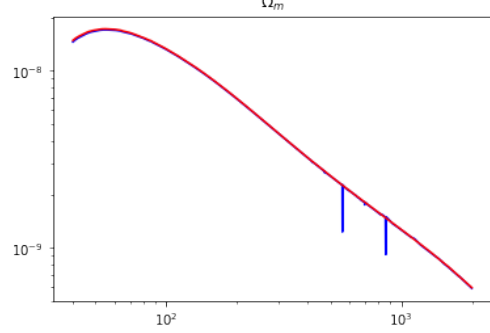


Figure 21: The methods finally converge at a step size of $h=0.00001$.

The first attempt also uses two extra differentiation formulae: left and right as opposed to central methods. These were henceforth discarded as they were much less convergent than the two central methods. With a step size of $h = 0.00001$, the derivatives converge, as can be seen in Figure 21.

We could then re-form our Fisher matrix with this more reliable derivative. It can be seen that the uncertainties changed:

- Ω_m uncertainty = 9.35×10^{-3}
- σ_8 uncertainty = 0.0118
- $\bar{\rho}_*$ uncertainty = 2.00×10^8 .

If we again marginalise over $\bar{\rho}_*$, we can investigate the two cosmological parameters. The ellipse now shows that they are in fact inversely correlated, which matches with what is known [23].

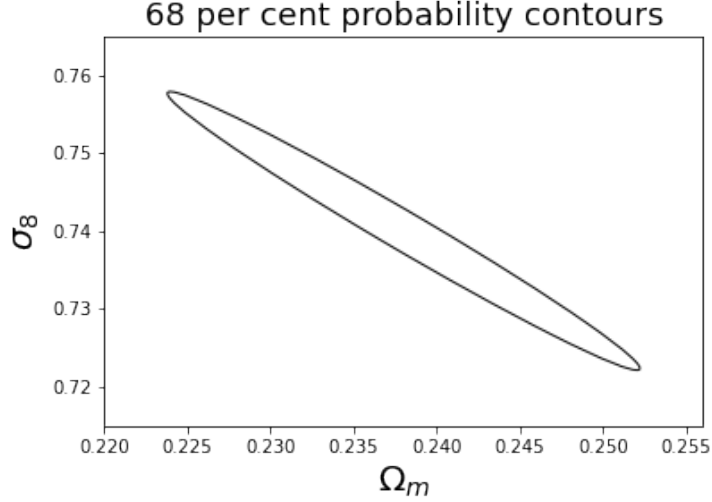


Figure 22: 68% uncertainty confidence ellipse for σ_8 and Ω_m .

7.3 Uncertainty on $\bar{\rho}_*$

Parameters are selected as important based on how much varying them affects first the matter power spectrum, and then the angular power spectrum. Based on this, we first investigate $\bar{\rho}_*$.

A transformation of variables was necessary, as the orders of magnitude involved were very different. Therefore, $\bar{\rho}_*$ is multiplied by a constant, $\xi = 1/(7 \times 10^8)$, so as to make it more comparable. The transformation of variables can be implemented by differentiating the old parameter set, $(\Omega_m, \sigma_8, \bar{\rho}_*)$, with respect to the new one, $(\Omega_m, \sigma_8, \xi\bar{\rho}_*)$, to create a kind of Jacobian matrix [22]. This is used to transform to a new Fisher matrix.

Marginalising over the new Fisher matrix to remove σ_8 dependence, we can look at the confidence ellipse for $\xi\bar{\rho}_*$ and Ω_m . If we instead marginalise over Ω_m , we get the dependence between $\xi\bar{\rho}_*$ and σ_8 instead. We see that these ellipses show a correlation, but not one as tight as that of the two cosmological parameters.

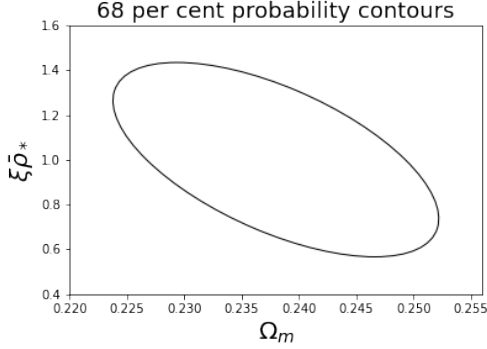


Figure 23: 68% uncertainty confidence ellipse for $\xi\bar{\rho}_*$ and Ω_m .

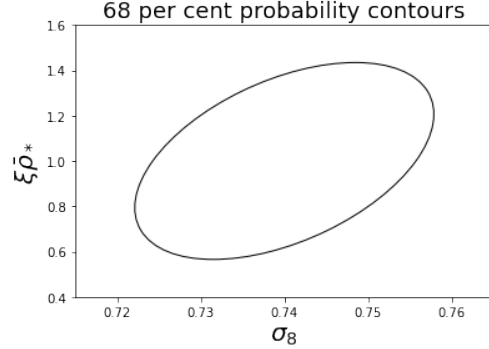


Figure 24: 68% uncertainty confidence ellipse for $\xi\bar{\rho}_*$ and σ_8 .

7.4 Setting a prior

Using the data from [10], we can add a prior on $\bar{\rho}_*$, constraining its uncertainty. We find the tightest constraint on a measurement of $\bar{\rho}_*$ at $z \sim 0$ is $\sim 9 \times 10^6$. This means we add a value of $\frac{1}{(9 \times 10^6)^2} = 1.23 \times 10^{-14}$ onto the diagonal entry corresponding to $\bar{\rho}_*$ in the original (non-transformed) Fisher matrix [22]. This leads to a lower uncertainty on $\bar{\rho}_*$ than without this prior, and actually improves the uncertainty on the cosmological parameters too. This makes sense, as we have shown them to be correlated quantities.

- Ω_m uncertainty = 7.43×10^{-3}
- σ_8 uncertainty = 0.0104
- $\bar{\rho}_*$ uncertainty = 9.01×10^6

Transforming this Fisher matrix to include $\xi\bar{\rho}_*$ instead of $\bar{\rho}_*$, and marginalising over the third parameter once again, we see the effect on the confidence ellipses.

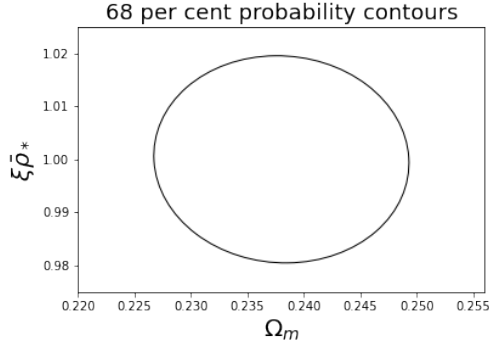


Figure 25: 68% uncertainty confidence ellipse for $\xi\bar{\rho}_*$ and Ω_m , with prior on $\bar{\rho}_*$.

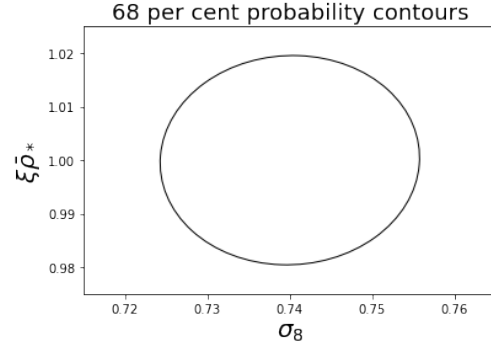


Figure 26: 68% uncertainty confidence ellipse for $\xi\bar{\rho}_*$ and σ_8 , with prior on $\bar{\rho}_*$.

These ellipses almost look horizontal, but there is a slight tilt on each, showing that there is a correlation between the parameters. However, it is small, as evinced by the broadness of the almost circular ellipses. Adding the prior has decreased the uncertainty on $\bar{\rho}_*$, but it also seems to have decreased the dependence between it and the cosmological parameters.

7.5 Fixing the third parameter

In the interests of completeness, I also check what happens when the third parameter is considered perfectly known rather than marginalised over. This is done by removing the corresponding row and column from the Fisher matrix before inverting it to get the uncertainties.

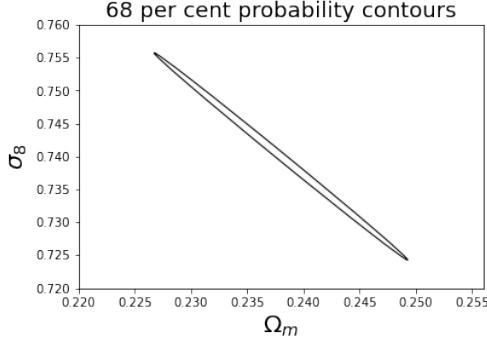


Figure 27: 68% uncertainty confidence ellipse for σ_8 and Ω_m , where $\bar{\rho}_*$ has been fixed.

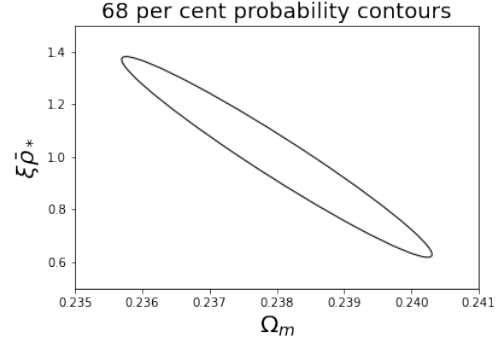


Figure 28: 68% uncertainty confidence ellipse for $\xi\bar{\rho}_*$ and Ω_m , where σ_8 has been fixed.

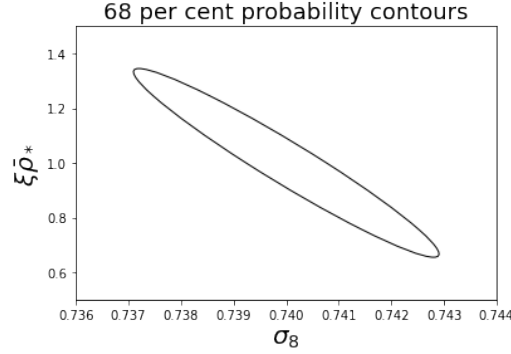


Figure 29: 68% uncertainty confidence ellipse for $\xi\bar{\rho}_*$ and σ_8 , where Ω_m has been fixed.

The ellipses are all narrower, showing that two parameters become more correlated when the third one is fixed. The uncertainty in all of the parameters also decreases, but the correlation for the last one has flipped. Further exploration will be necessary to find out why this switch of direction happens.

Considering a parameter as fixed essentially means we have a very strong prior on it. This is not something we currently have, but if very precise and accurate measurements did become available, we might hope to attain a constraint similar to this.

7.6 A different parameter

We now do a similar analysis, but with a different parameter, m_{0g} , replacing $\bar{\rho}_*$. As shown in subsection 4.2, this parameter may be important to constrain as it has an effect at a wide range of k . We construct a 3x3 Fisher matrix much as before. However, the third row/column now corresponds to m_{0g} .

- Ω_m uncertainty = 7.76×10^{-3}
- σ_8 uncertainty = 0.0131
- m_{0g} uncertainty = 1.44×10^{12}

The uncertainty on m_{0g} is actually similar to the value itself, which is very high. To look at how these parameters are correlated, we do the same as with $\bar{\rho}_*$ and multiply m_{0g} by a constant $\eta = 1/(10^{12}/h)$, and do a transformation of variables. The results appear similar to those for $\bar{\rho}_*$.

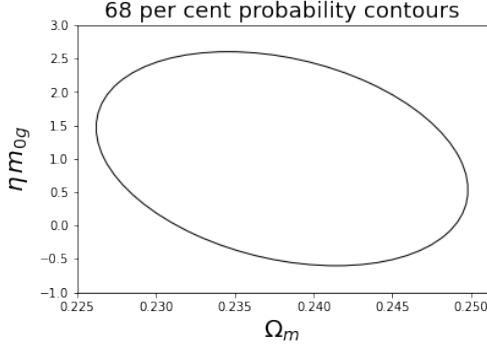


Figure 30: 68% uncertainty confidence ellipse for ηm_{0g} and Ω_m .

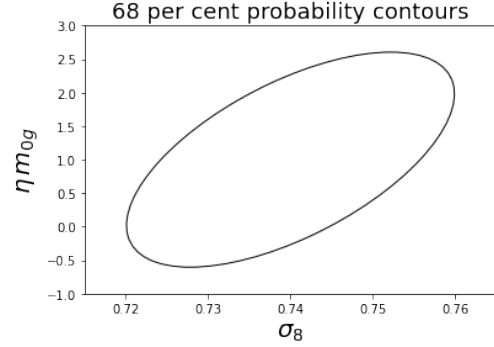


Figure 31: 68% uncertainty confidence ellipse for ηm_{0g} and σ_8 .

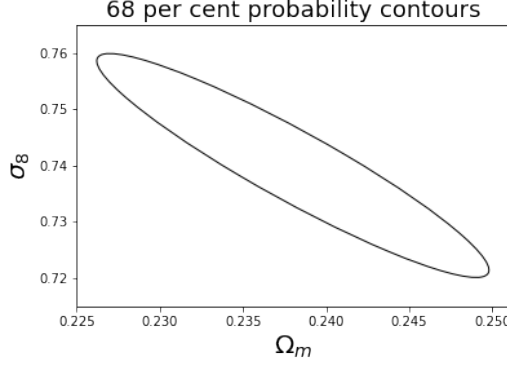


Figure 32: 68% uncertainty confidence ellipse for Ω_m and σ_8 , when the third parameter (marginalised over) is m_{0g} instead of $\bar{\rho}_*$.

We could also put a prior on m_{0g} . Say we knew its value to within $\pm 10\%$, then we would add $\frac{1}{(10^{11})^2}$ to the diagonal entry corresponding to m_{0g} . This would have the effect of improving the uncertainty on m_{0g} , but it would also slightly improve the uncertainty on the cosmological parameters.

- Ω_m uncertainty = 7.43×10^{-3}
- σ_8 uncertainty = 0.0104
- m_{0g} uncertainty = 1.36×10^{11}

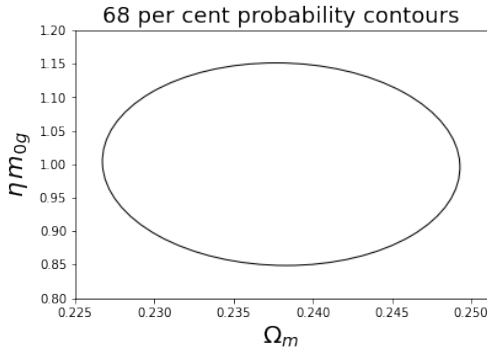


Figure 33: 68% uncertainty confidence ellipse for ηm_{0g} and Ω_m , with a prior on m_{0g} .

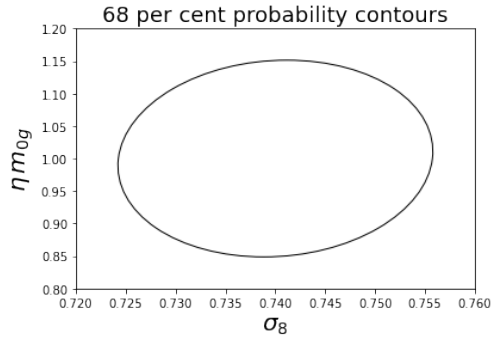


Figure 34: 68% uncertainty confidence ellipse for ηm_{0g} and σ_8 , with a prior on m_{0g} .

These are again almost horizontal, but there is a slight angle so they are not circular but elliptical. This effect is useful because constraining any parameter will help constrain the model as a whole. Therefore, a relevant observation of any parameter in this model will help improve the uncertainty, at least somewhat, on all the related parameters.

And, if we fix the third parameter instead of marginalising, the ellipses once again become narrower. This time the Ω_m dependence flips instead of the σ_8 one.

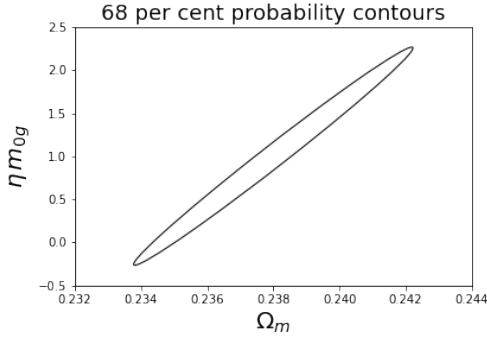


Figure 35: 68% uncertainty confidence ellipse for ηm_{0g} and Ω_m , with σ_8 fixed.

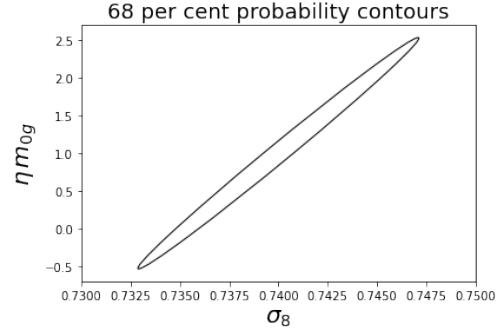


Figure 36: 68% uncertainty confidence ellipse for ηm_{0g} and σ_8 , with Ω_m fixed.

It is clear that these parameters $(\bar{\rho}_*, m_{0g})$ have a large uncertainty. This means that it is of the utmost importance to constrain them, so that we can be sure of their impact on the model. We have shown that adding a prior from observational data on $\bar{\rho}_*$ improves the uncertainty in the model. Finding more data for the other parameters will help this cause further, and it may be possible to have a fully observationally-motivated model of matter distribution.

8 Discussion and Conclusions

Modelling the distribution of matter in the universe is an ongoing process. Including baryons will be necessary for greater accuracy of these matter distribution models, as we will have more precise measurements of the large-scale structure of the universe in the near future. We have shown the possible impact of baryons on the matter power spectrum by considering a halo model which includes stars and gas in addition to DM. They can cause a noticeable change in the model when compared to a DM-only model.

However, the inclusion of luminous matter does add many free parameters to the model. These parameters are usually either left free and fit to the most observationally correct results, or fixed using data from simulations, which may not be accurate to reality. The observations they are fit to will soon be out of date and can have large uncertainty, and it is improbable that simulations perfectly reproduce reality, so there are not many constraints on these models. This is why it is important to use observational data to constrain these free parameters.

We investigated which parameters have the largest effect on the model by varying them and looking at the impact on the total matter power spectrum. Some parameters had very little impact, or only had an effect at very small scales, so were not considered as very important to constrain. On the other hand, a few of the parameters were very impactful. They could possibly change the model to a large degree, depending on their values.

Different sources of possible observational constraints were then explored. A detailed compilation of data on one of the parameters, $\bar{\rho}_*$, was found in the form of the Madau plot [10]. Also investigated was the possibility of finding constraints through exploring how simulations set their initial conditions. This revealed that measurements of the GSMF exist, and might be useful to us. This was examined to find out if it could constrain any of our free parameters. We did not end up finding anything directly relevant, but this might be an interesting avenue to study in future.

The two most effective of the parameters were selected and were investigated with a more detailed analysis. This involved doing a Fisher forecast on each one in turn. We first created a Fisher matrix with $\bar{\rho}_*$ and two cosmolog-

ical parameters, Ω_m and σ_8 . The uncertainties on these parameters inherent in the model were then obtainable from this. There was a very large uncertainty on $\bar{\rho}_*$, comparable to its own value. This meant it was very desirable to constrain it somewhat. It was necessary to do a transformation of variables to compare the parameters visually, so that the orders of magnitude involved were more similar. This allowed us to look at the confidence ellipses for each pair of parameters, to see how they were correlated. Following this, we added a prior from observational data on $\bar{\rho}_*$ and this reduced the uncertainty. However, it interestingly made the correlations with the cosmological parameters much less. There was not time to investigate why this might be, but it is nevertheless an intriguing result.

After that, we did a similar process for m_{0g} . There was also a very large uncertainty on this parameter. We did not find any observations with relation to this parameter, but we made a prediction for what effect having even a nominal prior to constrain the uncertainty would have. It was shown that if we knew the value of this parameter to within 10%, it would decrease the uncertainty by an order of magnitude, and improve the uncertainty in the other parameters as well. This would be very useful in the effort to constrain the model.

It is evident that the inclusion of baryons in this model and others like it will be necessary in order to have models precise enough for comparison to upcoming surveys. A considerable amount of uncertainty comes with this inclusion, so it is important to find ways to constrain the baryonic parameters. Current observational data can provide us with the means to add priors on some parameters, hence lowering the uncertainty. This must be further explored, as there is still a lot of freedom in this model. However, it is reassuring that most of the free parameters do not have a huge effect on the model if changed, so there are only a few which urgently need constraining. To continue with this research, it would be interesting to search for more sets of observational data which can constrain some of the other parameters. Depending on how everything is correlated, a measurement on any one parameter may improve the uncertainty in the whole model. There is still much to learn in terms of these models, and they will not be perfect immediately, but we hope to have added some small improvement.

Acknowledgements

I never would have been able to do this alone, so first I would like to thank most of all my supervisor Elisa Chisari, for being such a helpful and patient guide throughout this project. Also big thanks go to Aart and my friends and family for all the necessary moral support.

References

- [1] Asantha Cooray and Ravi Sheth. “Halo models of large scale structure”. In: *Physics reports* 372.1 (2002), pp. 1–129.
- [2] Rachel Mandelbaum et al. “The LSST Dark Energy Science Collaboration (DESC) Science Requirements Document”. In: *arXiv preprint arXiv:1809.01669* (2018).
- [3] S Kaviraj et al. “The Horizon-AGN simulation: evolution of galaxy properties over cosmic time”. In: *Monthly Notices of the Royal Astronomical Society* 467.4 (2017), pp. 4739–4752.
- [4] Nora Elisa Chisari et al. “Modelling baryonic feedback for survey cosmology”. In: *arXiv preprint, arXiv:1905.06082* (2019).
- [5] Aurel Schneider et al. “Quantifying baryon effects on the matter power spectrum and the weak lensing shear correlation”. In: *Journal of Cosmology and Astroparticle Physics* 2019.03 (2019), p. 020.
- [6] Cosimo Fedeli. “The clustering of baryonic matter. I: a halo-model approach”. In: *arXiv preprint, arXiv:1401.2997* (2014).
- [7] Phillip James Edwin Peebles. *The large-scale structure of the universe*. Vol. 98. Princeton University Press, 1980.
- [8] Martin Kilbinger. “Cosmology with cosmic shear observations: a review”. In: *Reports on Progress in Physics* 78.8 (2015), p. 086901.
- [9] Julio F Navarro, Carlos S Frenk, and Simon DM White. “The assembly of galaxies in a hierarchically clustering universe”. In: *Monthly Notices of the Royal Astronomical Society* 275.1 (1995), pp. 56–66.
- [10] Piero Madau and Mark Dickinson. “Cosmic star formation history”. In: *arXiv preprint, arXiv:1403.0007* (2014).

- [11] R Bielby et al. “The WIRCam Deep Survey-I. Counts, colours, and mass-functions derived from near-infrared imaging in the CFHTLS deep fields”. In: *Astronomy & Astrophysics* 545 (2012), A23.
- [12] L Pozzetti et al. “zCOSMOS–10k-bright spectroscopic sample-the bimodality in the galaxy stellar mass function: exploring its evolution with redshift”. In: *Astronomy & Astrophysics* 523 (2010), A13.
- [13] Joop Schaye et al. “The EAGLE project: simulating the evolution and assembly of galaxies and their environments”. In: *Monthly Notices of the Royal Astronomical Society* 446.1 (2015), pp. 521–554.
- [14] Ian G McCarthy et al. “The BAHAMAS project: calibrated hydrodynamical simulations for large-scale structure cosmology”. In: *Monthly Notices of the Royal Astronomical Society* (2016).
- [15] Yohan Dubois et al. “Dancing in the dark: galactic properties trace spin swings along the cosmic web”. In: *Monthly Notices of the Royal Astronomical Society* 444.2 (2014), pp. 1453–1468.
- [16] Peter Schneider. *Extragalactic astronomy and cosmology: an introduction*. Vol. 146. Springer, 2006.
- [17] Paul Schechter. “An analytic expression for the luminosity function for galaxies.” In: *The Astrophysical Journal* 203 (1976), pp. 297–306.
- [18] Ivan K Baldry et al. “Galaxy And Mass Assembly (GAMA): the galaxy stellar mass function at $z < 0.06$ ”. In: *Monthly Notices of the Royal Astronomical Society* 421.1 (2012), pp. 621–634.
- [19] Cheng Li and Simon DM White. “The distribution of stellar mass in the low-redshift Universe”. In: *Monthly Notices of the Royal Astronomical Society* 398.4 (2009), pp. 2177–2187.
- [20] Edo Van Uitert et al. “The stellar-to-halo mass relation of GAMA galaxies from 100 deg² of KiDS weak lensing data”. In: *Monthly Notices of the Royal Astronomical Society* 459.3 (2016), pp. 3251–3270.
- [21] Scott Dodelson and Fabian Schmidt. *Modern cosmology*. Academic Press, 2020.
- [22] Dan Coe. “Fisher matrices and confidence ellipses: a quick-start guide and software”. In: *arXiv preprint, arXiv:0906.4123* (2009).

- [23] Bhuvnesh Jain and Uroš Seljak. “Cosmological model predictions for weak lensing: Linear and nonlinear regimes”. In: *The Astrophysical Journal* 484.2 (1997), p. 560.

A Matter power spectrum code

I include my Python notebook used to calculate the matter power spectrum in the following pages. This can also be found at https://github.com/LSSTDESC/CCLX/blob/master/Baryons_halo_model_power_spectrum.ipynb as an example of how to use the `pyccl` package to construct a baryonic halo model. In future this kind of model may be implemented into `pyccl` as an inbuilt function, as it currently does not have a model of this type.

▼ Baryonic halo model

In this example, we illustrate how to build a baryonic halo model in CCL. We follow mostly the equations in Fedeli (2014), arXiv:1401.2997, with some modifications for speed or numerical simplicity.

```
import numpy as np
import pycc1 as ccl
import pylab as plt
import matplotlib.cm as cm
from scipy.special import erf
from scipy import integrate
from scipy import interpolate
from functools import partial
%matplotlib inline
```

▼ Preliminaries

Generate a cosmology and a vector with wavenumbers. We will work only at $z = 0$ for this example.

```
# Cosmology
cosmo = ccl.Cosmology(Omega_c=0.196, Omega_b=0.042, h=0.73, sigma8=0.74, n_s=0.951)

# Wavenumbers
k_arr = np.geomspace(1E-2, 1E2, 256)

# Scale factor
a_sf=1
```

Let us also initialize a series of halo model quantities. We will need:

- A halo mass definition.
- A concentration-mass relation.
- A halo mass function parametrization.
- A halo bias parametrization.
- Halo density profiles (for the dark matter part of the model).

```
# We will use a mass definition with Delta = 200 times the matter density
hmd_200m = ccl.halos.MassDef200m()

# The Duffy 2008 concentration-mass relation
cM = ccl.halos.ConcentrationDuffy08(hmd_200m)

# halo mass function
hmf_200m = ccl.halos.MassFuncTinker10(cosmo, mass_def=hmd_200m)

# The Tinker 2010 halo bias
hbf = ccl.halos.HaloBiasTinker10(cosmo, mass_def=hmd_200m)

# The NFW profile - this is for dark matter only
pM = ccl.halos.profiles.HaloProfileNFW(cM)
```

The baryonic halo model consists of three ingredients: gas, stars and dark matter. The contribution of each ingredients is weighted by their energy density at $z = 0$, which are given by

```
# Define some useful densities
rho_star = 7E8*cosmo['h']**2
rho_m = cosmo['Omega_m']*ccl.physical_constants.RHO_CRITICAL*cosmo['h']**2
rho_DM = cosmo['Omega_c']*ccl.physical_constants.RHO_CRITICAL*cosmo['h']**2
rho_g = rho_m - rho_star - rho_DM
```

We are now going to define some useful constants that we will need later.

```
#For the stellar mass fraction within halos
m_0s = 5E12/cosmo['h']
sigmas = 1.2

#For the stellar profile
x_deltas = 1/0.03

#For the gas mass fraction within halos
```

```

m_0g = 1E12/cosmo['h']
sigmag = 3

#For the gas profile
beta = 2.9

#Parameters that inform the power spectrum of the gas,
#specifically the extended component.
Fg = 0.05 #this is an example, it should be <<1
bd = 0.85 #from Fedeli

# A reference halo mass that will be used later
mhalo = 1E14/cosmo['h']

```

To compute halo model ingredients of the power spectrum, we will need to integrate over the masses of the contributing halos. These will be slightly different for stars, DM and gas. Here we define them:

```

# Array of masses
mmin_DM=1E6
mmax_DM=1E16
mmin_gas=m_0g
mmax_gas=mmax_DM
mmin_stars=1E10
mmax_stars=1E15
m_arr = np.geomspace(mmin_DM,mmax_DM,128)

```

▼ Stellar component

First we will compute the stellar mass fraction.

$$f_*(m) = A \exp\left(-\frac{\log_{10}^2(m/m_{0s})}{2\sigma_s^2}\right)$$

We fix A by requiring that $\bar{\rho}_*$ have a particular value.

$$\bar{\rho}_* = \int_0^\infty dm m f_*(m) n(m)$$

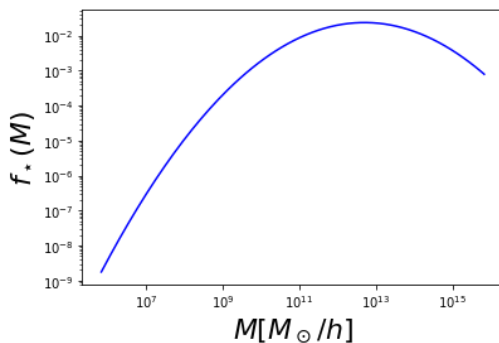
```

#This computes the normalization of the stellar mass fraction
def f_star_integrand(m):
    mf = hmf_200m.get_mass_function(cosmo,m,a_sf)/(m*np.log(10))
    return m*np.exp(-(np.log10(m/m_0s))**2/(2*sigmas**2))*mf
def f_star_norm():
    return integrate.quad(f_star_integrand,mmin_stars,mmax_stars,epsabs=0,epsrel=1E-3,limit=5000)[0]

#Stellar mass fraction
A = rho_star/f_star_norm()
def f_star(m):
    return A*np.exp(-(np.log10(m/m_0s))**2/(2*sigmas**2))

plt.figure()
plt.plot(m_arr*cosmo['h'], f_star(m_arr), 'b-')
plt.xlabel(r'$M$ [M_\odot/h]$', fontsize=22)
plt.ylabel(r'$f_\star(M)$', fontsize=22)
plt.xscale('log')
plt.yscale('log')
plt.show()

```



Next we define the shape of the stellar profiles and we make a class out of them, so we can then pass them on to the power spectrum method.

$$\rho_*(x|m) = \frac{\rho_t}{x} \exp(-x^\alpha), \text{ with } x := r/r_t.$$

An additional constraint on ρ_t comes from requiring that the mass fraction overall is equal to the mass fraction within a halo.

```
#Stellar profile
class HaloProfileStars(ccl.halos.HaloProfile):
    def __init__(self):
        super(HaloProfileStars, self).__init__()

    #scale radius
    def _rs(self, cosmo, M, a, mdef):
        return mdef.get_radius(cosmo, M, a) / a

    def _real(self, cosmo, r, M, a, hmd_200m):
        # Generate 2D array by default
        r_use = np.atleast_1d(r)
        M_use = np.atleast_1d(M)

        # Compute profile
        r_delta = self._rs(cosmo, M_use, a, hmd_200m)
        r_t = r_delta/x_deltas
        x = r_use/r_t[:, None]
        rhot = M_use*f_star(M_use)/(4* np.pi * r_t**3)
        prof = rhot[:,None] * np.exp(-x)/x

        # Make sure the output has the right shape
        if np.ndim(r) == 0:
            prof = np.squeeze(prof, axis=-1)
        if np.ndim(M) == 0:
            prof = np.squeeze(prof, axis=0)
        return prof

ps = HaloProfileStars()
ps.update_precision_ffftlog(padding_hi_ffftlog=1E3,
                             padding_lo_ffftlog=1E-3,
                             n_per_decade=1000,
                             plaw_fourier=-2.)
```

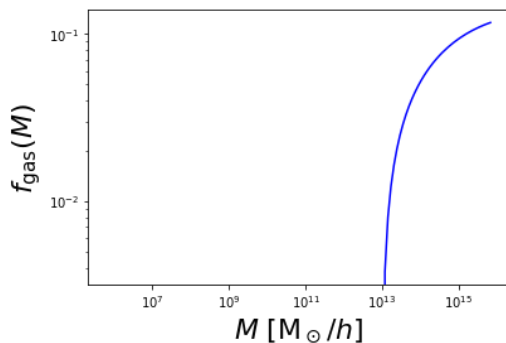
▼ Gas component

We now turn to the gas, again defining the gas fraction in each halo.

$$f_g(m) = \frac{\Omega_b}{\Omega_m} \text{erf}\left(\frac{\log_{10}(m/m_{0g})}{\sigma_g}\right)$$

```
def f_gas(m):
    fgas = []
    for i in m:
        if i < m_0g:
            fgas.append(0.)
        else:
            fgas.append((cosmo['Omega_b']/cosmo['Omega_m'])*erf(np.log10(i/m_0g)/sigmag))
    return fgas

plt.figure()
plt.plot(m_arr*cosmo['h'], f_gas(m_arr), 'b-')
plt.xscale('log')
plt.yscale('log')
plt.xlabel(r'$M$ [M$_\odot$/h]', fontsize=22)
plt.ylabel(r'$f_{\rm gas}(M)$', fontsize=22)
plt.show()
```



We now define a class that describes the profiles for the bound gas. We use a different gas profile to Fedeli (2014) because it is more compatible with the code.

$$\rho_g(r) = \frac{\rho_{g,0}}{(1+u)^\beta (1+v^2)^{(\gamma-\beta)/2}}, \text{ with } u := r/r_{co} \text{ and } v := r/r_{ej}.$$

```

class HaloProfileGas(ccl.halos.HaloProfile):
    def __init__(self):
        super(HaloProfileGas, self).__init__()

    #scale radius
    def _rs(self, cosmo, M, a, mdef):
        return mdef.get_radius(cosmo, M, a) / a

    def _rint(self, M, Rd, r):
        r_co = 0.1*Rd
        r_ej = 4.5*Rd #just placeholder value
        return r**2/((1+r/r_co)**beta*(1+(r/r_ej)**2)**((7-beta)/2))

    def _norm(self,M,Rd):
        nn = f_gas(M)*M/(4*np.pi*np.array([integrate.romberg(partial(self._rint, mm, rd),
                                                                1E-3,50,tol=0,rtol=1E-1,divmax=5000)
                                                                for mm, rd in zip(M, Rd)]))

        return nn

    def _real(self, cosmo, r, M, a, mdef):
        # Generate 2D array by default
        r_use = np.atleast_1d(r)
        M_use = np.atleast_1d(M)

        # Compute profile
        r_delta = self._rs(cosmo, M_use, a, mdef)
        r_co = 0.1*r_delta
        r_ej = 4.5*r_delta

        prof = 1/((1+r_use[None,:]/r_co[:,None])**beta*(1+(r_use[None,:]/r_ej[:,None])**2)**((7-beta)/2))
        norm = self._norm(M_use,r_delta)
        prof = prof[:, :]*norm[:,None]

        # Make sure the output has the right shape
        if np.ndim(r) == 0:
            prof = np.squeeze(prof, axis=-1)
        if np.ndim(M) == 0:
            prof = np.squeeze(prof, axis=0)
        return prof

pg = HaloProfileGas()
pg.update_precision_ffftlog(padding_hi_ffftlog=1E3,padding_lo_ffftlog=1E-3,
                             n_per_decade=1000,plaw_fourier=-2.)

```

▼ Interpolation of profiles

It takes a while to evaluate the profiles. For this reason, we are going to interpolate them.

```

#Interpolate the profiles for speed

M = np.logspace(np.log10(mmin_DM),np.log10(mmax_DM),num=20)
yDM=np.zeros((20,500))
M2D = np.zeros((20,500))
for i in range(0,500):
    M2D[:,i]=M
K = np.linspace(1E-2,1E2,num=500)
for j in range(0,20):
    yDM[j,:]=pg.fourier(cosmo, K,M[j], a_sf, hmd_200m)/M[j]
fDM = interpolate.interp2d(K, M, yDM)

print("DM profiles interpolated...\n")

M = np.logspace(np.log10(mmin_gas),np.log10(mmax_gas),num=10)
yg=np.zeros((10,500))
M2D = np.zeros((10,500))
for i in range(0,500):
    M2D[:,i]=M
K = np.linspace(1E-2,1E2,num=500)
for j in range(0,10):
    yg[j,:]=pg.fourier(cosmo, K,M[j], a_sf, hmd_200m)/M[j]
f = interpolate.interp2d(K, M, yg)

print("Gas profiles interpolated...\n")

M = np.logspace(np.log10(mmin_stars),np.log10(mmax_stars),num=10)
M2D = np.zeros((len(M),50))
for i in range(0,50):
    M2D[:,i]=M
K = np.linspace(1E-2,1E2,num=50)

```

```
ys=ps.fourier(cosmo, K,M, a_sf, hmd_200m)/M2D
fs = interpolate.interp2d(K,M,ys)

print("Star profiles interpolated...\n")
    DM profiles interpolated...

    Gas profiles interpolated...

    Star profiles interpolated...
```

▼ Calculation of power spectra

We define some generic integrals:

$$I_{2p} = \int_0^\infty dm n(m) m^2 y_1(k) y_2(k)$$

$$I_{bp} = \int_0^\infty dm n(m) b(m) m y(k)$$

```
# Halo Mass Calculator (for DM)
hmc = ccl.halos.HMCalculator(cosmo, hmf_200m, hbf, hmd_200m)

# Non-linear matter power spectrum, for comparison
pk_nl = ccl.nonlin_matter_power(cosmo, k_arr, a_sf)

#Generic integrals we need
#(of the form 5.x in Fedeli (2014))
def integrandGEN2p(m,k,prof1,prof2):
    dndlog10m = hmf_200m.get_mass_function(cosmo, m, a_sf)
    dndm=dndlog10m/(m*np.log(10.))
    y1=prof1(k,m)
    y2 =prof2(k,m)
    return dndm*m**2*y1*y2
def integratedGEN2p(mmin,mmax,k,p1,p2):
    return integrate.quad(integrandGEN2p, mmin, mmax, args=(k,p1,p2),epsabs=0,epsrel=1E-3,limit=5000)[0]
def integrandGENbp(m,k,prof1):
    bm = hbf.get_halo_bias(cosmo,m,a_sf)
    dndlog10m = hmf_200m.get_mass_function(cosmo, m, a_sf)
    dndm=dndlog10m/(m*np.log(10.))
    y=prof1(k,m)
    return dndm*m*bm*y
def integratedGENbp(mmin,mmax,k,p1):
    return integrate.quad(integrandGENbp, mmin, mmax, args=(k,p1),epsabs=0,epsrel=1E-3,limit=5000)[0]
```

We start with the dark matter:

```
#Dark matter auto-spectra

def PDM2h(k):
    return ccl.linear_matter_power(cosmo, k, a_sf)/rho_DM**2*integratedGENbp(mmin_DM,mmax_DM,k,fDM)**2

def PDM1h(k):
    return 1/rho_DM**2*integratedGEN2p(mmin_DM,mmax_DM,k,fDM,fDM)

PDM1h_arr=np.zeros(len(k_arr))
PDM2h_arr=np.zeros(len(k_arr))
for i in range(0,len(k_arr)):
    PDM1h_arr[i]=PDM1h(k_arr[i])
    PDM2h_arr[i]=PDM2h(k_arr[i])

#Gas auto-spectra

def PG2h(k):
    return ccl.linear_matter_power(cosmo, k, a_sf)/(Fg*rho_g)**2*integratedGENbp(mmin_gas,mmax_gas,k,f)**2

def PG1h(k):
    return 1/(Fg*rho_g)**2*integratedGEN2p(mmin_gas,mmax_gas,k,f,f)

def PGd(k):
    return bd**2*ccl.linear_matter_power(cosmo, k, a_sf)

def PGdh(k):
    return bd*ccl.linear_matter_power(cosmo, k, a_sf)/(Fg*rho_g)*integratedGENbp(mmin_gas,mmax_gas,k,f)

PG2h_arr=np.zeros(len(k_arr))
PG1h_arr=np.zeros(len(k_arr))
PGd_arr=np.zeros(len(k_arr))
```

```

PGdh_arr=np.zeros(len(k_arr))
for i in range(0,len(k_arr)):
    PG2h_arr[i]=PG2h(k_arr[i])
    PGd_arr[i]=PGd(k_arr[i])
    PG1h_arr[i]=PG1h(k_arr[i])
    PGdh_arr[i]=PGdh(k_arr[i])

PG = Fg**2*(PG2h_arr + PG1h_arr) + (1-Fg)**2*PGd_arr + 2*Fg*(1-Fg)*PGdh_arr

#DM x gas

def PDMgdh(k):
    return (bd*ccl.linear_matter_power(cosmo, k, a_sf)/rho_DM)*integratedGENbp(mmin_DM,mmax_DM,k,fDM)
def PDMg1h(k):
    return 1/(rho_DM*Fg*rho_g)*integratedGEN2p(mmin_gas,mmax_gas,k,f,fDM)
def PDMg2h(k):
    return ccl.linear_matter_power(cosmo, k, a_sf)/(rho_DM*Fg*rho_g)*integratedGENbp(mmin_DM,mmax_DM,k,fDM)*integratedGENbp(mmin_gas,mmax_gas,k,f,fDM)

PDMgdh_arr=np.zeros(len(k_arr))
PDMg2h_arr=np.zeros(len(k_arr))
PDMg1h_arr=np.zeros(len(k_arr))
for i in range(0,len(k_arr)):
    PDMgdh_arr[i]=PDMgdh(k_arr[i])
    PDMg1h_arr[i]=PDMg1h(k_arr[i])
    PDMg2h_arr[i]=PDMg2h(k_arr[i])

PDMg = (1-Fg)*PDMgdh_arr + Fg*(PDMg1h_arr+PDMg2h_arr)

```

We will now compute all spectra that involve a stellar contribution:

```

#Stars auto spectrum
def PS2h(k):
    return ccl.linear_matter_power(cosmo, k, a_sf)/rho_star**2*integratedGENbp(mmin_stars,mmax_stars,k,fs)**2

def PS1h(k):
    return 1/rho_star**2*integratedGEN2p(mmin_stars,mmax_stars,k,fs,fs)

PS2h_arr=np.zeros(len(k_arr))
PS1h_arr=np.zeros(len(k_arr))
for i in range(0,len(k_arr)):
    PS2h_arr[i]=PS2h(k_arr[i])
    PS1h_arr[i]=PS1h(k_arr[i])

#DM x stars

def PDMs1h(k):
    return 1/(rho_DM*rho_star)*integratedGEN2p(mmin_stars,mmax_stars,k,fDM,fs)

def PDMs2h(k):
    ii=integratedGENbp(mmin_stars,mmax_stars,k,fs)*integratedGENbp(mmin_DM,mmax_DM,k,fDM)
    return ccl.linear_matter_power(cosmo, k, a_sf)/(rho_DM*rho_star)*ii

#Eval and add DM x stars
PDMs1h_arr=np.zeros(len(k_arr))
PDMs2h_arr=np.zeros(len(k_arr))
for i in range(0,len(k_arr)):
    PDMs1h_arr[i]=PDMs1h(k_arr[i])
    PDMs2h_arr[i]=PDMs2h(k_arr[i])
PDMs = PDMs1h_arr+PDMs2h_arr

#Stars x gas
def Psg1h(k):
    return bd*ccl.linear_matter_power(cosmo, k, a_sf)/rho_star*integratedGENbp(mmin_stars,mmax_stars,k,fs)

def Psg1h(k):
    return 1/(rho_star*Fg*rho_g)*integratedGEN2p(mmin_gas,mmax_stars,k,f,fs)

def Psg2h(k):
    ii=integratedGENbp(mmin_stars,mmax_stars,k,fs)*integratedGENbp(mmin_gas,mmax_gas,k,f)
    return ccl.linear_matter_power(cosmo, k, a_sf)/(rho_star*Fg*rho_g)*ii

#Eval and add star x gas
Psg1h_arr=np.zeros(len(k_arr))
Psg2h_arr=np.zeros(len(k_arr))
Psg1h_arr=np.zeros(len(k_arr))
for i in range(0,len(k_arr)):
    Psg1h_arr[i]=Psg1h(k_arr[i])
    Psg2h_arr[i]=Psg2h(k_arr[i])

```

```
Psg2h_arr[i]=Psg2h(k_arr[i])
Psg = (1-Fg)*Psg_h_arr + Fg*(Psg1h_arr+Psg2h_arr)
```

Now plot them all and add them to form the total:

```
#Auto-spectra
Pauto = (rho_DM/rho_m)**2*(PDM2h_arr+PDM1h_arr)+(rho_g/rho_m)**2*PG+(rho_star/rho_m)**2*(PS1h_arr+PS2h_arr)
#Cross-spectra
Pcross = (2*rho_DM*rho_star*PDMs + 2*rho_g*rho_star*Psg+2*rho_DM*rho_g*PDMg)/(rho_m**2)
#Add all
Ptotal = Pauto + Pcross

#Relative power spectra
plt.figure(figsize=(7,8))
#All auto correl
plt.plot(k_arr/cosmo['h'], (rho_DM/rho_m)**2*(PDM2h_arr+PDM1h_arr)/pk_n1, 'k-', label='$P_{DM}(k)$')
plt.plot(k_arr/cosmo['h'], (rho_g/rho_m)**2*PG/pk_n1, 'r-', label='$P_G(k)$')
plt.plot(k_arr/cosmo['h'], (rho_star/rho_m)**2*(PS2h_arr+PS1h_arr)/pk_n1, 'b-', label='$P_S(k)$')
#All cross correl
plt.plot(k_arr/cosmo['h'], 2*rho_DM*rho_g/(rho_m**2)*PDMg/pk_n1, 'g-', label='$P_{DM,G}(k)$')
plt.plot(k_arr/cosmo['h'], 2*rho_g*rho_star/(rho_m**2)*Psg/pk_n1, 'm-', label='$P_{G,S}(k)$')
plt.plot(k_arr/cosmo['h'], 2*rho_DM*rho_star/(rho_m**2)*PDMs/pk_n1, 'c-', label='$P_{DM,S}(k)$')
#And the total with respect to CAMBs nonlinear power
plt.plot(k_arr/cosmo['h'], Ptotal/pk_n1, 'k--', label='$P_{total}(k)$')
plt.xscale('log')
plt.yscale('log')
plt.legend(loc='lower left', fontsize=12.5)
plt.ylabel(r'$P(k)/P_{nl}(k)$', fontsize=15)
plt.xlabel(r'$k\,h\,Mpc^{-1}$', fontsize=15)
plt.show()
```

