The Power Spectrum in Cosmology

Emanuele Besana

Abstract

I introduce the notion of the power spectrum as it is used in cosmological applications. The concern here is to convey an intuitive understanding of the concept, while trying to not give up the mathematical formalism. This brief review is intended to be read by the already initiated, as some concepts are only reminded, but not fully explained.

1 The philosophy of cosmological perturbations

In cosmology we usually start with raw observation, namely that the Universe, at least at large scales, seems to respect the properties of homogeneity and isotropy. Naturally, smaller scales break these symmetries, so that for structures (galaxies, clusters of galaxies etc...) to exist, some form of perturbations needs to be introduced into the theory. This is done in the following way. Consider one physical variable f, which for our purposes can be thought of as representing pressure, density, temperature and so on. One then usually starts by describing the background with a Friedmann geometry, a large scale description of our Universe, where the only concern is the *average value* of f, $\bar{f}(x^0)$ (as a consequence of homogeneity, the $\mathbf{x}$ dependence is trivial). Then, perturbations are introduced as fields $\delta f(x)$ that live on this background spacetime, so that the total physical variable $f(x)$ reads:

$$f(x) = \bar{f}(x^0) + \delta f(x) \tag{1}$$

In general, this expression is rewritten more handily:

$$f(x) \equiv \bar{f}(1 + \delta_f) \tag{2}$$

where $\delta_f = \delta f/\bar{f}$ is the *contrast in* f. This last form turns out to be quite helpful, as it hints at an expansion around δ_f, which is how the perturbations are treated in cosmology.

2 The Matter Power Spectrum

In cosmology, one is particularly interested in the spatial distribution of matter, both dark and baryonic, throughout the Universe. If we want to attempt at a mathematical description, the quantity to study is the *matter density contrast*:

$$\delta_\rho(x) = \frac{\rho(x)}{\bar{\rho}} - 1 \tag{3}$$

What does this quantity tell us about the statistics of matter distribution? Since the contrast fields are defined with respect to the mean itself, we trivially have $\langle \delta_\rho \rangle = 0$. In fact, what we are really looking for is the following:

$$\langle \delta_\rho(\mathbf{x}, t)\delta_\rho(\mathbf{x'}, t)\rangle = \xi(t, \mathbf{x}, \mathbf{x'}) \tag{4}$$

where the object appearing on the right is called the *correlation function*. Because of homogeneity and isotropy, we can rewrite this as:

$$\xi(r) = \langle \delta_\rho(\mathbf{x})\delta_\rho(\mathbf{x'})\rangle \tag{5}$$

where $r = |\mathbf{x}-\mathbf{x'}|$ and the time t is implied. This is fundamental, because it asks the question: if I have a certain over(under)-density here $\delta_\rho(\mathbf{x})$, is it correlated to an over(under)-density over there $\delta_\rho(\mathbf{x'})$? This is what we wanted of course, that is $\xi(r)$ measures of the clumpiness of the Universe at a scale r. As with everything in physics, it is easier to work in Fourier space, and this is where the power spectrum comes into play. Let's first transform the δ_ρ:

$$\delta_\rho(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3}\delta_\rho(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}} \tag{6}$$

This, as is known, is just a decomposition of the field into its waves, thus $|\delta(\mathbf{k})|^2$ is the amplitude of a certain wave-number, i.e. how *important* that wave-number is in the decomposition. In fact, we may give the following pseudo-definition that, although not rigorous, is the basis of the idea behind the matter power spectrum $\mathcal{P}_\rho(k)$:

$$|\delta(\mathbf{k})|^2 \sim \mathcal{P}_\rho(k) \tag{7}$$

In other words, the power spectrum just tells us how much "power" resides at a certain scale $\lambda \sim 1/k$. What does this mean? The Fourier transform "divides" 3D space into pieces, which can either be big (small k) or small (large k). Therefore, if, say, $\mathcal{P}_\rho(k)$ has a peak at a certain $k = k_\star$, this means that the perturbations with wavelengths, i.e. scales, $\lambda_\star \sim 1/k_\star$ are very important, which ultimately results in the Universe being highly correlated at $\lambda_\star$. Even though this line of thought gives a rather intuitive explanation, it is only roughly true, leaving behind some details which we later talk about.
Clearly, equations (5) and (7) are very alike, and in fact we can be more precise (although the mathematics will become more involved, the physical intuition given above remains valid). At a glance, we might think of the power spectrum as the Fourier transform of the correlation function. This is exactly what is done traditionally, and may even serve as a definition:

$$\begin{aligned}
\langle \delta_\rho(\mathbf{k})\delta_\rho(\mathbf{k'})\rangle &= \int d^3x d^3x' e^{-i\mathbf{k}\cdot\mathbf{x}}e^{i\mathbf{k'}\cdot\mathbf{x'}}\langle \delta_\rho(\mathbf{x}, t)\delta_\rho(\mathbf{x'}, t)\rangle \\
&= \int d^3r d^3x' e^{-i\mathbf{k}\cdot\mathbf{r}}e^{-i(\mathbf{k}-\mathbf{k'})\cdot\mathbf{x'}}\xi(r) \\
&= (2\pi)^3\delta_D(\mathbf{k}-\mathbf{k'})\int d^3r e^{-i\mathbf{k}\cdot\mathbf{r}}\xi(r) \\
&\equiv (2\pi)^3\delta_D(\mathbf{k}-\mathbf{k'})\mathcal{P}_\rho(k)
\end{aligned} \tag{8}$$

The last line clearly isn't that much different from (7). We have then defined:

$$\mathcal{P}_\rho(k) = \int d^3r e^{-i\mathbf{k}\cdot\mathbf{r}}\xi(r) \tag{9}$$

On the other hand then:

$$\xi(r) = \langle \delta_\rho(\mathbf{x})\delta_\rho(\mathbf{x}+\mathbf{r})\rangle$$
$$= \int \frac{d^3k}{(2\pi)^3}\,\mathcal{P}_\rho(k)e^{i\mathbf{k}\cdot\mathbf{r}} = \frac{1}{2\pi^2}\int_0^\infty dk\; k^2\mathcal{P}_\rho(k)\frac{\sin(kr)}{kr} \tag{10}$$

Practically, even though it is $\xi(r)$ that measures the clumpiness of the Universe in real space, the matter power spectrum is the quantity that's mathematically easier to handle.

Now, one might ask what these averages $\langle\dots\rangle$ mean, and why we are even taking them. The answer lies in one critical observation, namely that the Universe is an intrinsically *stochastic process.* Stochasticity is a property of those systems for which it is impossible to exactly predict the evolution, because they contain a fundamental element of chance. This randomness could be attributed to an incomplete knowledge of the initial conditions, or simply to measurement errors. These systems are described through *random variables.* A random variable is determined by a set of numbers (its values) and a probability measure defined over this set. Then one considers a large (ideally infinite) collection of identical systems (an *ensemble*) and builds *ensemble averages* from there, in order to define the statistics of that random variable (becasue of fundamental stochasticity, statistics are all we have).
A typical example is that of a coin toss. Here the randomness is obviously attributed to our ignorance about initial conditions. A random variable that could be defined is Y, a variable that models a 1\$ payoff for a successful bet on heads. Its values are twofold, 1 and 0. The probability measure defined on this set is clearly 1/2 and 1/2 respectively. Now, we might ask the following question: what is the probability of having 20\$ after 50 throws? The answer is: toss the coin 50 times a large (again, ideally infinite) amount of times, and average over the results, i.e. take the ensemble average.
Our case is fundamentally the same. Clearly, to know the initial conditions of the Universe, we would need an infinite set of values: $\delta_\rho(\mathbf{x}, t_i)$ for every $\mathbf{x}$, which is obviously impossible. Now, we are dealing with *random fields*, like $\delta_\rho(x)$, instead of random variables, but the former are roughly infinite and continuous collections of random variables (one for each point x). If we now ask the probability of correlations in matter densities as a function of scale, i.e. we ask for $\xi(r)$, the line of thinking is clear. We take multiple realizations of the Universe, and then we average over all of these against the probability distribution function which defines the random field; this sequence defines the "observational" two-point function ξ_{obs}:

$$\xi_{\text{obs}}(r) = \int \mathcal{D}\delta_\rho\mathbb{P}[\delta_\rho]\delta_\rho(\mathbf{x}, t)\delta_\rho(\mathbf{x'}, t) \tag{11}$$

Again, since the Universe is a fundamentally stochastic process, statistics are all we can hope to obtain. You might ask at this point what the probability measure of the random variables (in total, a field) is. There is good reason to believe that it is a Gaussian:

$$\mathbb{P}[\delta_\rho]d\delta_\rho = \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left(-\frac{\delta_\rho^2}{2\sigma^2}\right)d\delta_\rho \tag{12}$$

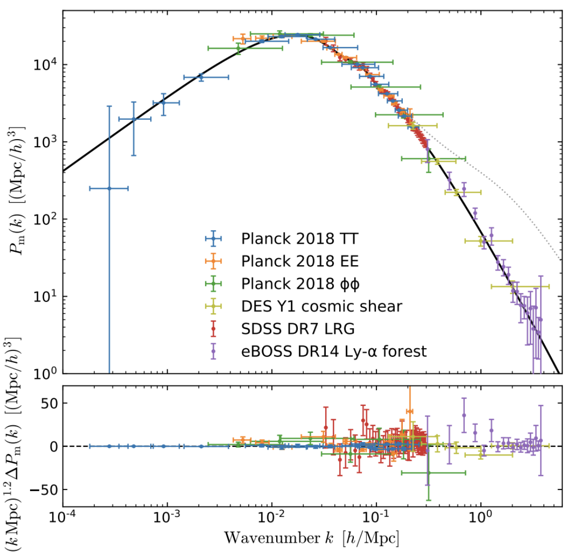
where $\sigma^2 = \xi(0)$. Gaussian random fields are particularly simple and convenient, since, as long as the evolution of the δ_ρ is linear, the statistics remain

Gaussian. Importantly, this means that even *initially* they were Gaussian. In addition, the two point correlation function specifies completely all other N-point functions, as the joint probability reads:

$$\mathbb{P}[\delta_\rho(\mathbf{x}_1), \delta_\rho(\mathbf{x}_2), \dots, \delta_\rho(\mathbf{x}_N)] \propto \frac{1}{\sqrt{\det(\xi_{ij})}} \exp\left(-\frac{1}{2}\delta_\rho(\mathbf{x}_i)\xi_{ij}^{-1}\delta_\rho(\mathbf{x}_j)\right) \quad (13)$$

where $\xi_{ij} \equiv |\mathbf{x}_i - \mathbf{x}_j|$. There is a slight problem with this line of reasoning however. It has to do with the obvious fact that we don't possess any number of copies of Universes to average on, let alone an infinite amount. For this reason, in order to test theoretical predictions with astronomical observations, we have to assume some degree of *ergodicity*, which just says that ensemble averages become spatial averages as the volume becomes infinitely large. This can only hold true if the volume we are averaging over is infinite, which is clearly not true in our case, where we are limited by the speed of light to only see an "observable" Universe. Basically, we are acting as if different parts of the Universe are independent realizations of the underlying random process. For this reason the validity of the ergodic hypothesis depends on the ratio of the scale over which we perform the spatial average to the scale at which spatial correlations become negligibly small. When the surveyed volume contains many statistically independent subsamples, ergodicity is expected to hold. This, however, will introduce statistical errors called *sample variance.*

Let's summarise how we can actually compare theory with experiment. On the observational side of things, we divide the sky into patches, calculate the product $\delta_\rho(\mathbf{x})\delta_\rho(\mathbf{x}')$, then average it over a number of patches given by the ratio of the size of the observable Universe to the patch size. The result is supposed to approximate (11), with an error that's bigger on larger scales (since the patches are bigger, i.e. they are less in number) rather than on small scales. Once we have a $\xi_{\mathrm{obs}}(r)$, we Fourier transform it to the power spectrum (9), which is finally compared to its theoretical prediction.
At present time, we seem to have a rather strong agreement between theory and observations, as is seen in the following plot [2]:



As we have seen, going from a power spectrum to a correlation function requires a Fourier transform. Unfortunately, this step implies that one needs to be careful in trying to build an intuition, so now we will offer more detail. Since the evolution of the δ_ρ is linear, i.e. $\delta_\rho(t) \propto \delta_\rho(t_i)$ for some initial time t_i, we describe the power spectrum at an arbitrary time in the following manner:

$$\mathcal{P}_\rho(k,t) = D^2(t)T^2(k)\mathcal{P}_\rho(k,t_i) \tag{14}$$

Therefore, we only need to know the initial power spectrum, which is given by:

$$\mathcal{P}_\rho(k,t_i) = Ak^n \tag{15}$$

where A is a constant and n is the *spectral index.* It turns out that a value for it that (almost) describes our Universe is $n = 1$, resulting in the *Harrison-Zel'dovich* power spectrum. The question now is the following. Given a power spectrum:

$$\mathcal{P}_\rho(k) = Ak^n \tag{16}$$

what does the two-point correlation function look like? We start out with the easiest value for n, namely $n = 0$. Clearly we have:

$$\langle \delta_\rho(\mathbf{x})\delta_\rho(\mathbf{x}+\mathbf{r})\rangle = \int \frac{d^3k}{(2\pi)^3}\mathcal{P}_\rho(k)e^{-i\mathbf{k}\cdot\mathbf{r}} \sim \delta_D(\mathbf{r}) \tag{17}$$

This power spectrum is known as *white noise*, because there is no correlation between distant point, i.e. the perturbations are distributed randomly in space. For other ns, we note that, given (16), one can solve the integral (10) to find the real space correlation function. Notice that we require that $n > -3$ in order for $\xi(r) \to 0$ as $r \to \infty$, as required by large scale homogeneity. We then get:

$$\xi(r) \sim \frac{1}{r^{n+3}} \tag{18}$$

Finally, let us show why the spectral index $n = 1$ is particularly interesting. In order to do this, while being consistent with cosmological literature, we look at perturbations in another quantity, namely the gravitational/Bardeen potential $\Phi(x)$. This is related to $\delta_\rho(x)$ in a simple way via the Poisson equation:

$$\mathcal{P}_\Phi(k) \propto k^{-4}\mathcal{P}_\rho(k) \tag{19}$$

so that the initial power spectrum for Φ reads:

$$\mathcal{P}_\Phi(k,t_i) \propto k^{n-4} \tag{20}$$

The value $n = 1$, i.e. our Universe, gives rise to what is called a *scale invariant* power spectrum, since, at time t_i:

$$\langle \Phi(\lambda\mathbf{x})\Phi(\lambda\mathbf{x}')\rangle = \int \frac{d^3k}{(2\pi)^3} \underbrace{\frac{1}{k^3}}_{\mathcal{P}_\Phi(k,t_i)} e^{-i\lambda\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')} = \langle \Phi(\mathbf{x})\Phi(\mathbf{x}')\rangle \tag{21}$$

This means that the correlation between any two points is independent of their distance, i.e. there are correlations at all scales. Notice that this is different from white noise, where there is power at all scales, so *no correlation* at any scale.

One final note about the matter power spectrum regards the so-called *dimensionless power spectrum*, frequently mentioned and used in cosmology articles. In the limit $r \to 0$, the two point function reduces to the variance (which is what we used in (12)):

$$\sigma^2 = \langle \delta(\mathbf{x})^2 \rangle \tag{22}$$

Since the power spectrum and the correlation function are Fourier-related, it is expected that a further interpretation of the power spectrum has to do with its contribution to the variance. In fact, we can also use the limit in equation (10):

$$\sigma^2 = \frac{1}{2\pi^2} \int_0^\infty k^2 \mathcal{P}_\rho(k) dk \tag{23}$$

This equation is telling us that the product $\mathcal{P}_\rho(k)/2\pi^2$ gives the power, i.e. the contribution to the variance, per unit k-space volume due to the modes with wave-number k. However, a thin spherical shell in k-space contains many different modes and has a k-space volume $4\pi k^2 dk$, so that the total power contributed by perturbations with wave-number between k and $k + dk$ is proportional to the power spectrum and the number (volume) of contributing modes. In general it is convenient to introduce a new quantity, the dimensionless power spectrum:

$$\Delta^2(k) = \frac{k^3}{2\pi^2} \mathcal{P}_\rho(k) \tag{24}$$

which is the contribution to the variance per bin of $\log k$, instead of unit k-space. Therefore, for instance, $\Delta^2(k) = 1$ is just shorthand for saying that the Fourier modes in a unit logarithmic bin around wave-number k generate fluctuations δ_ρ of order unity.

3 The CMB and the Angular Power Spectrum

A further object of interest in cosmology is the temperature of the CMB, the radiation leftover from the Big Bang that freely streamed after decoupling. The CMB is the most accurate black body we have in nature with temperature $\bar{T}_0 = 2.73\,\mathrm{K}$, but it still presents very small temperature perturbations. Clearly, these perturbations are related to the matter ones we studied above (and to the metric perturbations), since the Einstein equations couple them beyond linear order. In fact, the mathematical treatment in this chapter parallels very closely the one above.

Let's think about what the quantity of interest is in this case. The observations detect the CMB photons at *present time* t_0 and at location $\mathbf{x} = 0$, i.e. where Earth is located. Additionally, and crucially, we observe them on a sphere, so instead of analyzing the problem in 3D space $\mathbf{x}$, we use 2D spherical coordinates (θ, ϕ). Basically, we need to keep track of the following object:

$$\delta_T(\hat{\mathbf{n}}, t_0, \mathbf{x} = 0) = \frac{T(\hat{\mathbf{n}}, t_0, \mathbf{x} = 0)}{\bar{T}_0(t_0)} - 1 \tag{25}$$

Usually, the t_0 and $\mathbf{x} = 0$ are implied. Basically, just like we were interested in $\delta_\rho(\mathbf{x})$ above, now we are interested in $\delta_T(\hat{\mathbf{n}})$. Again, since $\langle \delta_T \rangle = 0$, we instead focus on the two-point function:

$$C(\theta) = \langle \delta_T(\hat{\mathbf{n}})\delta_T(\hat{\mathbf{n}}')\rangle \tag{26}$$

where, like before, isotropy implies that it only depends on the separation of the two points $\theta = \hat{\mathbf{n}}\cdot\hat{\mathbf{n}}'$. This time, instead of going to Fourier space, we go to *harmonic space*, that is we expand $C(\theta)$ into spherical harmonics, a complete basis for scalar functions in the sky:

$$\delta_T(\hat{\mathbf{n}}) = \sum_{\ell,m} a_{\ell m} Y_{\ell m}(\hat{\mathbf{n}}) \tag{27}$$

where $\ell = 1, \ldots, \infty$ and $m = -\ell, \ldots, \ell$. At this point, the parallel should be clear: the $a_{\ell m}$ and $Y_{\ell m}$ now play the role of the $\delta_\rho(\mathbf{k})$ and the plane waves $e^{-i\mathbf{k}\cdot\mathbf{r}}$ respectively. Therefore, along the lines of (7) we now write:

$$|a_{\ell m}|^2 \sim P_{T,lm} \tag{28}$$

The $a_{\ell m}$ tell us the importance of a perturbation of scale $\lambda \sim 1/\ell$, because the spherical harmonics divide the sphere into pieces (just like sines and cosines do for 3D space), either large (small ℓ) or small (large ℓ). In other words, the ks have been exchanged for the ℓs. In fact, the traditional definition is the following:

$$\langle a_{\ell m} a^*_{\ell' m'}\rangle = C_\ell \delta_{\ell\ell'}\delta_{mm'} \tag{29}$$

where, precisely, the C_ℓ is the *angular power spectrum.* The next thing we are going to do should now be clear. The C_ℓ will be the "harmonic transform" of the two-point function $C(\theta)$:

$$\begin{aligned} C(\theta) &= \langle \delta_T(\hat{\mathbf{n}})\delta_T(\hat{\mathbf{n}}')\rangle \\ &= \sum_{\ell m}\sum_{\ell' m'} \langle a_{\ell m} a^*_{\ell' m'}\rangle T_{\ell m}(\hat{\mathbf{n}}) Y^*_{\ell' m'}(\hat{\mathbf{n}}') \\ &= \sum_{\ell} C_\ell \sum_m Y_{\ell m}(\hat{\mathbf{n}}) Y^*_{\ell m}(\hat{\mathbf{n}}') \\ &= \sum_\ell \frac{2\ell+1}{4\pi} C_\ell P_\ell(\cos\theta) \end{aligned} \tag{30}$$

where $P_\ell(\cos\theta)$ are the Legendre polynomials, which are related to the spherical harmonics by:

$$P_\ell(\cos\theta) = \frac{4\pi}{2\ell+1}\sum_m Y_{\ell m}(\hat{\mathbf{n}}) Y^*_{\ell m}(\hat{\mathbf{n}}') \tag{31}$$

Since at the end of the day it is easier to work in harmonic space, the object of focus for theorists is C_ℓ. Even in this case, we are interested in the power per logarithmic bin to the variance:

$$C(0) = \sum_\ell \frac{2\ell+1}{4\pi} C_\ell \approx \int d\log\ell\ \frac{(2\ell+1)\ell C_\ell}{4\pi} \approx \int d\log\ell\ \frac{\ell(\ell+1)C_\ell}{2\pi} \tag{32}$$

where the final equality holds for $\ell \gg 1$. The power per logarithmic interval in ℓ is defined as:

$$\Delta_T^2 \equiv \frac{\ell(\ell+1)}{2\pi} C_\ell \bar{T}_0^2 \tag{33}$$

This quantity is the one that is usually plotted.
We have just seen how this discussion is basically the same as the one for the matter power spectrum, only with the following conceptual replacements:

$$\begin{array}{rcl} \delta_\rho(\mathbf{x}) & \longleftrightarrow & \delta_T(\hat{\mathbf{n}}) \\ k & \longleftrightarrow & \ell \\ \delta_\rho(\mathbf{k}) & \longleftrightarrow & a_{\ell m} \\ \xi(r) & \longleftrightarrow & C(\theta) \\ \mathcal{P}_\rho(k) & \longleftrightarrow & C_\ell \end{array} \tag{34}$$

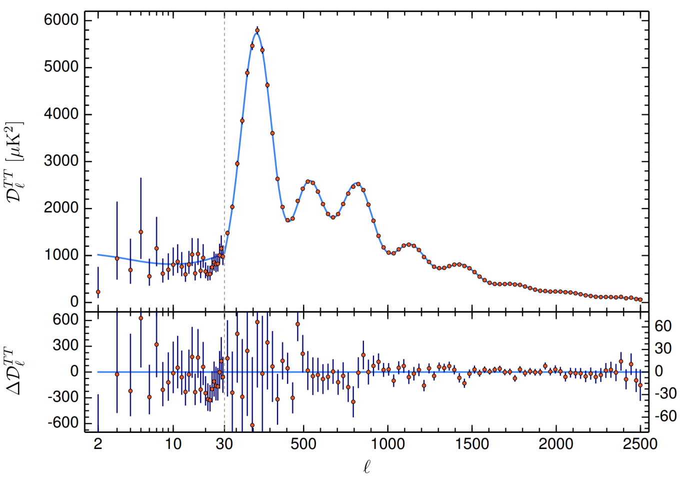
To finish off the discussion, we note that the observational point of view of the two-point function $C_{\rm obs}(\theta)$ should be written as something similar to (11), with a probability function that's still a Gaussian, since all the perturbations δ_T, δ_ρ... are related to each other and to the same object at early times (a *primordial* power spectrum). In much the same way as before, instead of the ensemble average, practically one is forced to use spatial averages, which however induce some error called, in the CMB case, *cosmic variance*. The problem is in fact the following. Suppose we measure the power spectrum in real 2D spherical coordinates, i.e. we measure $C(\theta)$ (let's drop the subscript obs for the time being), then, for fixed ℓ, we would have $2\ell+1$ different $a_{\ell m}$s to make an estimate of the corresponding C_ℓ; clearly, an estimator for it would be:

$$\hat{C}_\ell = \frac{1}{2\ell+1}\sum_m |a_{\ell m}|^2 \tag{35}$$

This estimator has an associated variance, corresponding to the error we incur in trying to determine the real power spectrum:

$$\frac{\Delta C_\ell}{C_\ell} \equiv \frac{\sqrt{\langle (C_\ell - \hat{C}_\ell)^2\rangle}}{C_\ell} = \sqrt{\frac{2}{2\ell+1}} \tag{36}$$

This is what we expect. The error is largest for small ℓ, i.e. large scales, since we have a smaller number of patches in the sky to average on.
Again, there is an outstanding agreement between observation and theory, as can be seen in the following plot [3]:



References

[1] D. Baumann, *Cosmology.* Cambridge University Press, 7 2022.

[2] S. Chabanier, M. Millea, and N. Palanque-Delabrouille, "Matter power spectrum: from Lyα forest to CMB scales," *Mon. Not. Roy. Astron. Soc.*, vol. 489, no. 2, pp. 2247–2253, 2019.

[3] N. Aghanim *et al.*, "Planck 2018 results. VI. Cosmological parameters," *Astron. Astrophys.*, vol. 641, p. A6, 2020.