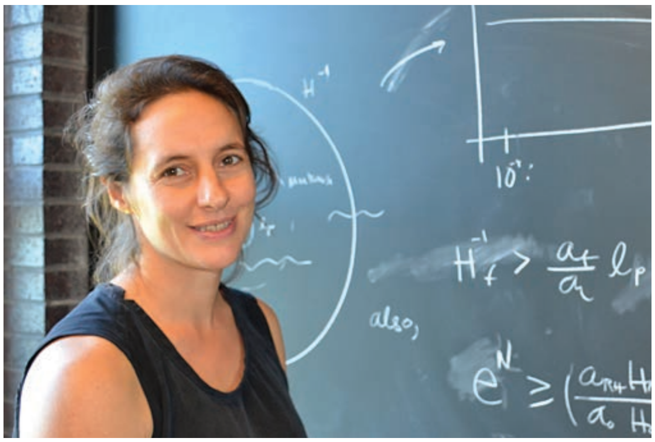
Mapping the Universe

By Marilena Loverde

Assistant Professor of Physics
C.N. Yang Institute for Theoretical Physics
Stony Brook University



Marilena Loverde
Photo: Lorraine Walsh

Cosmologists are largely statisticians. We cannot directly observe the evolution of the Universe, but we can make maps of the Universe at different epochs in time and compare how the statistical properties have changed. Similarly, we cannot set up and perform cosmological experiments with different variables held fixed, but we can search for correlations between the things we observe and quantities that vary throughout space. In fact, much of observational and theoretical cosmology involves devising clever statistics to isolate physical relationships that cannot be measured directly.

Cosmologists make maps of the Universe through a variety of probes: the temperature and polarization of the cosmic microwave background radiation (CMB), the abundance of galaxies or the magnitude of their line-of-sight velocities, the ellipticities and brightnesses of galaxies, which can be used to infer the distribution of matter by gravitational lensing, and the strength of emission and absorption lines from atomic and molecular spectra in intergalactic gas. From these quantities we can extract information about both the initial distribution of different types of matter in our Universe and dynamical processes occurring through the history of our Universe. In this article, I give a brief description of some examples of both types of inferences.

To set the stage for this discussion, an image showing a map of the anisotropy in the temperature of the CMB as mapped by the Planck satellite [1], along with a map of the distribution of galaxies mapped by the Sloan Digital Sky Survey (SDSS) [2] is shown in Figure 1. The temperature and polarization anisotropies in the CMB provide a snapshot of the Universe at a fixed time about $380,000$ years after the big bang. The galaxy surveys (and other methods) provide images of the Universe at a range of times in cosmic history, beginning with the faintest observable galaxies formed about a billion years after the big bang until today. In the coming years, new cosmological surveys will map out more and more of the observable Universe.

The most familiar examples of cosmological statistics are *power spectra*, the Fourier transforms of two-point correlation functions. For instance, if $\rho_m(\mathbf{x})$ de-

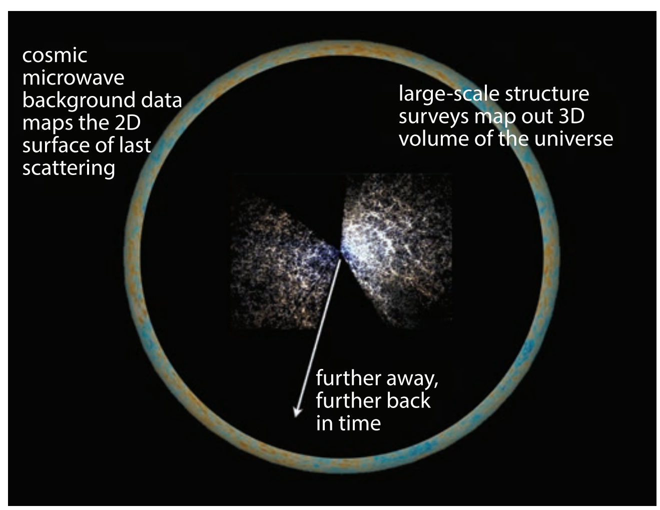


Figure 1: The large colorful sphere is a map of the temperature anisotropies in the CMB from the Planck satellite [1]. The CMB anisotropies primarily show inhomogeneities in the Universe at the time when these photons are emitted, the map they provide is therefore a 2-dimensional sphere of the so-called *last-scattering surface*. Also shown are a subset of galaxies in our universe mapped by the Sloan Digital Sky Survey. A galaxy survey can detect galaxies at a range of distances from Earth, and therefore at a range of epochs in cosmic history. The maps of galaxies show that there is *large-scale* structure in the distribution of matter in the Universe.

scribes the matter density at position $\mathbf{x}$ in a volume V, then the fluctuations in the matter density are given by,

$$\delta_m(\mathbf{x}) \equiv \rho(\mathbf{x})/\bar{\rho}_m - 1\,, \tag{1}$$

with $\bar{\rho}_m$ the mean matter density in V. The power spectrum of matter fluctuations is estimated by

$$P_{mm}(k_b) = \frac{1}{VN_k}\sum_{\mathbf{k}\in k_b}|\delta_m(\mathbf{k})|^2 \tag{2}$$

where k_b is a bin in wavenumber and $N_k = \sum_{\mathbf{k}\in k_b}$ and I have adopted the cosmologist's convention of distinguishing between real and Fourier space quantities only by their argument. Similarly, if $n_g(\mathbf{x})$ is the number density of galaxies, and then the matter-galaxy cross-power spectrum is given by

$$P_{gm}(k_b) = \frac{1}{VN_k}\sum_{\mathbf{k}\in k_b}\frac{1}{2}\left(\delta_g^*(\mathbf{k})\delta_m(\mathbf{k}) + \delta_m^*(\mathbf{k})\delta_g(\mathbf{k})\right)\,. \tag{3}$$

where $\delta_g(\mathbf{x}) = n_g(\mathbf{x})/\bar{n}_g - 1$, where $\bar{n}_g$ is the mean number of galaxies in V. Cosmologists measure the auto- and cross-power spectra of nearly any quantity they can measure! These are used for a variety of purposes. For instance, the matter power spectrum in Eq. (2) characterizes the typical amplitudes of fluctuations in the matter density on wavelength $2\pi/k$. More precisely, the variance of matter fluctuations on scale k is given by $\Delta_{mm}(k) = 4\pi^2k^3P_{mm}(k)/(2\pi)^3$. In our Universe, $\Delta_{mm}(k)$ is an increasing function of k so that the typical amplitude of density perturbations is larger on smaller scales. Equivalently, the Universe appears most inhomogeneous on small scales, and on large scales typical fluctuations in the density are very tiny. The variance of matter fluctuations in our Universe is shown in Figure 2. For comparison, Figure 3 shows a realization of the matter distribution taken from a snapshot of a cosmological simulation. By construction, the power spectrum of the matter distribution in the simulation will be consistent with Figure 2.

A working assumption in cosmology is that the particular realization of the distribution of different types of matter in our Universe is a random draw from some underlying probability distribution functional I'll call $\mathcal{P}[\delta_c, \delta_b, \delta_\gamma, \delta_\nu, \ldots]$. In this expression δ_c indicates fluctuations in the cold dark matter density, δ_b indicates fluctuations in baryonic matter density[1], δ_γ indicates fluctuations in the photon energy density, δ_ν indicates fluctuations in neutrino energy density,

[1]Because the masses of the proton and neutron are so much larger than the mass of the electron, cosmologists typically use the term "baryonic matter" to include the energy density in nuclei, atoms, and all standard model particles other than neutrinos, even though electrons are leptons and not baryons.

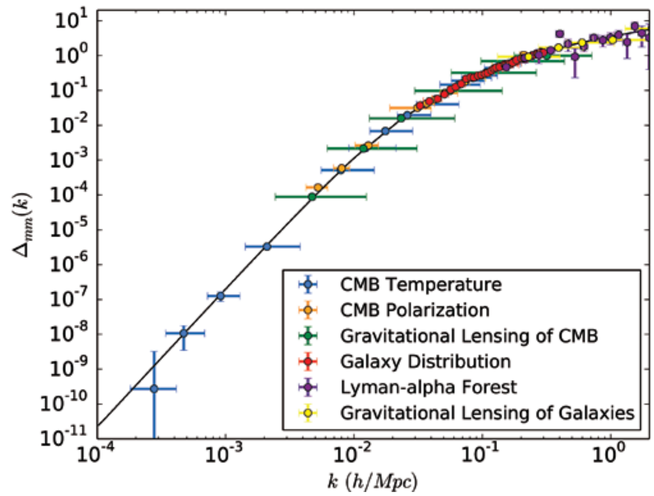


Figure 2: The power spectrum of fluctuations in the matter density throughout our Universe compiled from a variety of datasets, adapted from [1]. Data from the Planck satellite's measurements of CMB temperature and polarization anisotropies determines the power spectrum on the largest scales (smallest k). The distribution of galaxies from SDSS is used on intermediate scales, while the smallest scale measurements come from maps made using the Lyman-α absorption line in intergalactic Hydrogen from BOSS, and correlated distortions to the shapes of galaxies caused by gravitational lensing, as measured by the Dark Energy Survey (DES). This plot uses cosmologist length units of Mpc ($\approx 3\times10^{22}$ meters) with the dimensionless Hubble parameter $h = H_0/(100km/s) \approx 0.7$ scaled out.

and the ... allow for any other type of matter that may exist.

In this framework, if the "experiment" of our Universe were run many times we would expect to see different realizations of these fields $\delta_c(\mathbf{x}), \delta_b(\mathbf{x}), \delta_\gamma(\mathbf{x}), \delta_\nu(\mathbf{x}), \ldots$. Yet, if one measured the power spectra and cross-power spectra of these fields in each realization via Eq. (2) and then computed the averages, we would recover the "true" power spectra, e.g.

$$P_{cc}^{true}(k) = \langle P_{cc}(k)\rangle\,, \qquad P_{cb}^{true}(k) = \langle P_{cb}(k)\rangle \ldots \tag{4}$$

The angular brackets, $\langle\rangle$ above, indicate ensemble averages over realizations of the matter fields. The "true" power spectra are the functions that characterize the variances and co-variances of fluctuations in the probability distribution functional $\mathcal{P}[\delta_c, \delta_b, \delta_\gamma, \delta_\nu, \ldots]$. A theory that provides a mechanism for the origin of the structure, or inhomogeneities, in our Universe should provide an explanation for both the form of the probability distribution functional $\mathcal{P}[\delta_c, \delta_b, \delta_\gamma, \delta_\nu, \ldots]$ and for any quantities, e.g. power spectra or bispectra, needed to characterize it. A completely general probability distribution functional will depend on an infinite number of correlation functions, or higher-order polyspectra, of each independent quantity! Amazingly, our

Universe appears much simpler. All data can be described if the initial values of each field are determined by a single random field $\mathcal{R}(\mathbf{x})$, sometimes called the *primordial curvature perturbation*. That is, the initial spatial distributions of all quantities we have observed appear to determined by $\mathcal{R}$ through

simple proportionalities,

$$\delta_c(\mathbf{x}) \propto \delta_b(\mathbf{x}) \propto \delta_\gamma(\mathbf{x}) \propto \delta_\nu(\mathbf{x}) \propto \mathcal{R}(\mathbf{x})\,. \qquad (5)$$

This is a special class of initial conditions called *adiabatic*. Relative fluctuations between different components are called *isocurvature* modes, and at present there is no evidence for any primordial isocurvature modes in our Universe.

The properties of the field $\mathcal{R}$ are also very special: $\mathcal{R}$ is a Gaussian random field. This means that each Fourier mode $\mathcal{R}(\mathbf{k})$ is statistically independent and the phases of each mode are drawn from a flat probability distribution. The two-point function completely characterizes the statistics,

$$\langle \mathcal{R}(\mathbf{k})\mathcal{R}(\mathbf{k}')\rangle = (2\pi)^3 P_{\mathcal{R}\mathcal{R}}(k)\delta_{Dirac}(\mathbf{k}+\mathbf{k}') \qquad (6)$$

where δ_{Dirac} is the Dirac-delta function and

$$\langle \mathcal{R}(\mathbf{k_1})\mathcal{R}(\mathbf{k_2})\ldots\mathcal{R}(\mathbf{k_n})\rangle_c = 0 \qquad \text{for} \qquad n > 2\,. \qquad (7)$$

The subscript $_c$ in Eq. (7) indicates the connected part of the correlation function (i.e. subtracting all pairwise contractions, which are just determined by Eq. (6)). The amplitudes of the individual Fourier modes of $\mathcal{R}(\mathbf{k})$ are drawn from a Gaussian distribution with a variance given by [3],

$$\Delta_{\mathcal{R}\mathcal{R}}(k) \equiv \frac{4\pi^2}{(2\pi)^3} P_{\mathcal{R}\mathcal{R}}(k) \qquad (8)$$

$$\approx 2.1 \times 10^{-9} \left(\frac{0.05\mathrm{Mpc}^{-1}}{k}\right)^{0.035} \qquad (9)$$

From Eq. (9) we learn that the initial perturbations in the spatial curvature are very small and that the power spectrum is nearly scale-invariant but with slightly larger-amplitude perturbations on large scales (small k). At present, there is no evidence for any deviation from the functional form in Eq. (8), for instance any change in the power law index with k. Furthermore, there is no evidence for any non-trivial higher order correlation functions of $\mathcal{R}$.

The initial conditions described above (adiabatic, Gaussian, and nearly scale-invariant) can be explained by one of the most popular theories for the origin of structure, cosmological inflation. Within inflation, the primordial curvature perturbations are generated by quantum mechanical fluctuations during a phase of exponential expansion in the very early Universe. The near scale-invariance of the primordial curvature power spectrum is directly related to the near-constancy of the expansion rate. While we don't know the precise time or energy scale during inflation, it could be as early as 10^{-36} seconds after the Big Bang when the typical energy scale was 10^{15}GeV. Measurements of the statistics of $\mathcal{R}$, therefore, provide a window into this era. Near-term experiments such as SPHEREx aim to detect higher-order correlation functions in $\mathcal{R}$, which can provide information about the types of matter and interactions that were important during inflation – all at energy scales vastly beyond what is accessible by terrestrial particle colliders (see, e.g. [4], for a recent review).

Now, how do cosmologists actually determine the statistics of matter fluctuations? Traditionally, the simplest approach has been to use observations of the anisotropies in the CMB. As shown in Figure 1, these primarily give a map of our universe at a single snapshot in time. Dramatic improvements in our understanding of many aspects of our Universe will require data from more epochs in cosmic history. One classic way of mapping the large-scale distribution of matter is via galaxy surveys – maps of the positions of galaxies across our Universe. To interpret data from galaxy surveys, however, requires understanding how fluctuations in the distribution of galaxies relate to fluctuations in other types of matter, e.g. cold dark matter and baryons. Fortunately, these quantities are closely related. In Figure 3 two snapshots of simulations of structure in the Universe are shown. The image on the left shows the distribution of dark matter, while the image on the right shows just the dark matter halos (an excellent proxy for the positions of galaxies, since galaxies reside in halos). The galaxy distribution clearly reflects some of the underlying structure in the matter distribution, but provide an incomplete picture where the matter is, particularly on smaller scales.

The correlation between large-scale fluctuations in the abundance of galaxies $\delta n_g(\mathbf{x})$ and large-scale fluctuations in the matter density $\delta\rho_m(\mathbf{x})$ quantifies how closely galaxies trace matter. The cross-correlation coefficient between these two quantities is referred to as *galaxy bias* b_g,

$$b_g \equiv \frac{P_{gm}(k)}{P_{mm}(k)}\,. \qquad (10)$$

This quantity is important for interpreting data from galaxy surveys because galaxies are much easier to observe than dark matter, which dominates the mat-

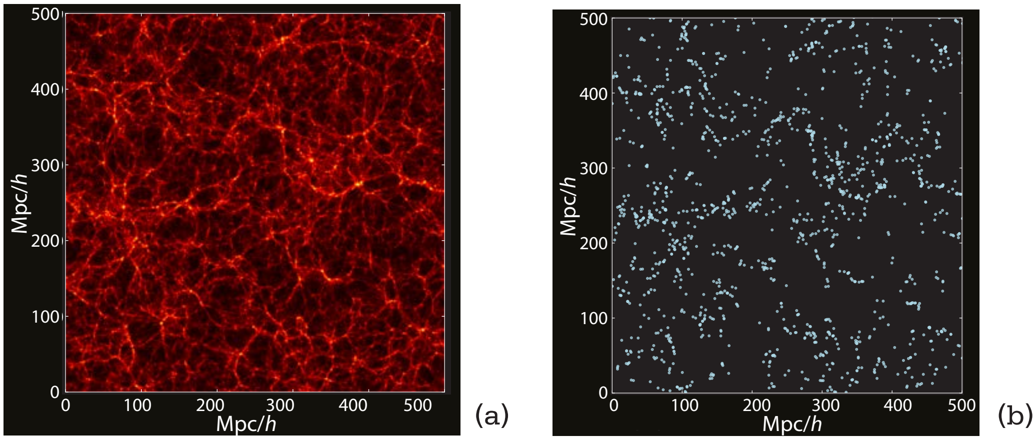


Figure 3: (a) A map of the distribution of dark matter in a slice of an N-body simulation of our Universe (b) A map of the position dark matter halos, the hosts of galaxies, in the same simulation. The image on the right resembles what we can observe directly in galaxy surveys (e.g. the wedges from the Sloan Digital Sky Survey shown in Figure 1), while the image on the left shows the full large-scale structure of the Universe. The large-scale fluctuations in the halo distribution are strongly correlated with large-scale fluctuations in the matter distribution on the left. On the other hand, from the halo distribution alone, it is difficult to infer the small-scale finer features in the matter distribution. Both Figures are by Drew Jamieson, adapted from [5].

ter density ρ_m of our Universe, but the distribution of dark matter is a better diagnostic for addressing fundamental questions about cosmology. The galaxy bias in Eq. (10) is also interesting in its own right, it quantifies how the presence of long-wavelength matter fluctuations helps or hinders galaxy formation. This can be understood simply if we consider the number of galaxies in some region of space to be a functional of the matter densities in the same region of space,

$$n_g(\mathbf{x}) = n_g[\rho_c(\mathbf{x}), \rho_b(\mathbf{x}), \ldots] \,. \tag{11}$$

In Eq. (11), $n_g(\mathbf{x})$ should be understood to be the number of galaxies in some region around $\mathbf{x}$ that is larger than the typical size of a galaxy, say R_{gal}, and where the local values of the cold dark matter and baryon densities are $\rho_c(\mathbf{x})$ and $\rho_b(\mathbf{x})$ (as before we allow ... for any other quantity that may affect galaxy formation). Because galaxy formation is a process that occurs over a long period of time, the number of galaxies in some region will depend on the *history* of the local density fields, not just their values at a single instant in time. By cross-correlating Eq. (11) with the total fluctuations in the matter density $\delta_m = (\rho_c\delta_c + \rho_b\delta_b)/(\rho_c + \rho_b)$ we can extract the linear galaxy bias w.r.t. the total matter fluctuation defined in Eq. (10) via,

$$b_g = \frac{1}{\langle n_g \rangle} \frac{\langle \delta_m n_g[\rho_m] \rangle}{\langle \delta_m \delta_m \rangle} = \frac{\delta \log n_g}{\delta \delta_m} \,. \tag{12}$$

Traditionally, the galaxy bias is studied under the assumption that only the total matter density affects galaxy formation and that all long-wavelength perturbations in the matter density $\delta_m(k \ll 1/R_{gal})$ evolve in the exact same way. While this is true in simple cases, for instance a Universe with inhomogeneities in only cold dark matter, it is violated in many interesting examples, including a Universe like our own! At Stony Brook, we have been pioneering the study of how the galaxy bias depends on the full evolutionary history of the local matter density. To do this, we have developed techniques to study dark matter halo formation in radically different environments. This, in essence, allows us to compute the functional derivative in Eq. (12) by computing the response of n_g to different evolutionary histories for δ_m. By doing this we have gained insight into the formation of structure in our Universe. Because the true evolutionary history of δ_m will depend on the types of matter present in our Universe, as well as the initial conditions (e.g. Eq. (5) or violations of that form), these studies have produced new tests of properties of our Universe. In the coming years, these tools will be increasingly important for interpreting data from surveys of the galaxies and matter distributed throughout the Universe.♦

References

[1] **Planck** Collaboration, Y. Akrami *et al.*, "Planck 2018 results. I. Overview and the cosmological legacy of Planck," arXiv:1807.06205 [astro-ph.CO].

[2] https://www.sdss.org/. https://www.sdss.org/.

[3] **Planck** Collaboration, N. Aghanim *et al.*, "Planck 2018 results. VI. Cosmological parameters," arXiv:1807.06209 [astro-ph.CO].

[4] P. D. Meerburg *et al.*, "Primordial Non-Gaussianity," arXiv:1903.04409 [astro-ph.CO].

[5] D. Jamieson and M. Loverde, "Separate Universe Void Bias," arXiv:1909.05313 [astro-ph.CO].