THE UNIVERSITY
of EDINBURGH

Machine Cosmology: Investigating The Dark Sector Through Novel Inference Methods

Benjamin Möws

Doctor of Philosophy
The University of Edinburgh
June 2021

Lay summary

As one of the oldest sciences, astronomy has a rich history of discoveries and attempts to explain the Universe around us. Over time, and with the emergence of natural philosophy and subsequently physics as a dedicated discipline, these efforts grew into a scientific field of study that is now the cornerstone of our understanding of the evolution and future of our Universe. At the time Albert Einstein put forward his theory of gravity, slightly over a century ago, one of the often-favored hypotheses was that of a static Universe, neither contracting nor expanding, which Einstein himself believed in at first. In the same decade, research by Vesto Slipher and Carl Wilhelm Wirtz, later solidified by Edwin Hubble's observations, also determined that the distance between us and most galaxies is increasing, although they were not understood as such at that point. Shortly thereafter, in 1920, what is known as the Great Debate took place, in which the dominant view of 'spiral nebulae' in the outskirts of our own galaxy was confronted with the notion of these observations really being far-away galaxies just like our own. Similarly, later observations led to the establishment of the Big Bang theory proposed by Georges Lemaître as the standard model of cosmology, with later refinements such as an initial period of exponential expansion called 'inflation' to explain the emergence of large-scale structure in the Universe.

As can be seen from these pioneering steps in what is now the field of cosmology, the discipline in the modern sense is only around one hundred years old, which pales in comparison to many other fields of research in the natural sciences. Just like observations at that time enabled us to further our understanding of the larger world around us, recent years have pushed these efforts to new heights through advancing technologies for cosmological surveys. These include both increasingly large ground-based and space-based telescopes, allowing us to look farther into the Universe than ever before. Luckily for us, from improved observations naturally follows the potential to test theoretical advances.

This thesis is, after the general introduction laying the groundwork, split into four chapters, which contribute to this close relationship between cosmology, statistics, and now machine learning by targeting challenges at different levels of granularity. In practice, this means that we first tackle the largest-scale problem of constraining the fundamental parameters of our Universe such as the mass density of matter and the Hubble constant measuring the rate of expansion. As the number of parameters to estimate translates to the dimensionality in which we need to operate, and calculations involved in estimating these parameters tend to be computationally costly, methods for solving this problem in a reasonable time is a challenge in modern cosmology. For this, we make use of and expand upon recent advances in statistics and machine learning to develop a novel approach that is naturally parallel, and demonstrate the scalability and speed advantage by using supercomputing facilities. Following that, we address the estimation of the dark energy equation of state, the ratio of pressure to energy density in a cosmological model, through the use of type Ia supernovae as luminous explosions at the end of certain stars' lives. In developing and applying a constrainable generator of random alternative cosmologies in this context, we show that larger deviations from the generally accepted model do not necessarily lead to easier-to-detect discrepancies when relying solely on these types of measurements, showing that physics beyond the standard model could hide in plain sight.

Next, we look at the large-scale structure of our Universe, meaning the filamentary construct emerging from the agglomeration of galaxies on very large scales. Measuring the latter is often a noisy process, which makes the denoising of those measurements an important step in finding what are known as 'cosmic voids', large and nearly empty regions of space in-between said filamentary construct. We make use of and extend recent advances in statistical methodology to extract the underlying structure from these noisy measurements of matter density as a way to unravel finer-grained empty regions, and compare it to alternative approaches. Lastly, in zooming into the last level of granularity on our journey, we delve into the area of galaxy evolution. While simulations that include baryonic properties, meaning properties stemming from 'normal' matter as opposed to dark matter, are substantially costlier in terms of computational requirements than those only simulating dark matter, they are a crucial tool in cosmology. We develop and apply a novel way to complete galactic dark matter halos with baryonic properties by combining an analytic approach of modeling the evolutionary relationship between baryonic and dark matter with machine learning, boosting the latter with the larger amount of information available in this framework.

Abstract

Cosmology during the last few decades has experienced an influx of new theory and observations, pushed forward by ever-increasing capabilities of current and upcoming large-scale surveys, computational and methodological capabilities, and new theoretical work being fueled by these latter factors. Observational measurements often carry uncertainties from noise or random processes, with inference methods being concerned with inverse probability as the quest to explore underlying distributions of data. Over the same time frame, Bayesian statistics has thus quickly found itself in a central role in cosmological analysis, as the field is rife with inverse problems such as hypothesis testing, model selection, and parameter estimation. More recently, inference models from the field of machine learning have also experienced a surge in applications to cosmology. We delve into the utility of such inference methods for challenges in cosmology in different degrees of granularity and focusing on the dark sector of our Universe, traveling from the largest scale to more local problems in the process.

Starting in the area of cosmological parameter estimation, we develop a novel parallel-iterative parameter estimation method rooted in Bayesian nonparametrics and recent developments in variational inference from the field of machine learning in Chapter 2. In doing so, we propose, implement, and test a new approach to fast high-dimensional parameter estimation in an embarrassingly parallel manner. For this work, we make use of large-scale supercomputing facilities to speed up the functional extraction of cosmological parameter posteriors based on data from the Dark Energy Survey. Next, we concentrate on the dark energy equation of state in Chapter 3, stress-testing its imprint on type Ia supernovae measurements through an introduced random curve generator for smooth function perturbation. We then investigate the robustness of standard model analyses based on such data with regard to deviations from a cosmological constant in the form of a redshift-dependent equation of state.

With regard to large-scale structure, we show the advantages of density ridges as curvilinear principal curves from Dark Energy Survey weak lensing data for cosmic trough identification in Chapter 4. Denoising large-scale structure in this way allows for the more fine-grained identification of structural components in the cosmic web. We also compare the results of our extended version of the subspace-constrained mean shift algorithm to curvelet denoising as an alternative method, as well as trough structure from measurements of the foreground matter density field. Lastly, in the area of galaxy formation and evolution, we combine analytic formalisms and machine learning methods in a hybrid prediction framework in Chapter 5. We use a two-step process to populate dark matter haloes taken from the SIMBA cosmological simulation with baryonic galaxy properties of interest. For this purpose, we use the equilibrium model of galaxy evolution as a precursory module to enable an improved prediction of remaining baryonic properties as a way to quickly complete cosmological simulations.

Declaration

I declare that no part of this thesis has been submitted for any other degree or professional qualification. This thesis was composed by myself and the work contained herein is my own except where explicitly stated otherwise in the text.

Parts of this thesis are based on existing and published papers by the author, such as parts of Chapter 1 being based on the publications listed in the remainder of this declaration. Chapter 2 is based on Moews, B. and Zuntz, J. (2020), "Gaussbock: Fast parallel-iterative cosmological parameter estimation with Bayesian nonparametrics", *The Astrophysical Journal*, 896(2), 98, Chapter 3 on Moews, B. et al. (2019), "Stress testing the dark energy equation of state imprint on supernova data", *Physical Review D*, 99, 123529, Chapter 4 on Moews, B. et al. (2020), "Ridges in the Dark Energy Survey for cosmic trough identification", *Monthly Notices of the Royal Astronomical Society*, 500(1), 859, and Chaper 5 on Moews, B. et al. (2021), "Hybrid analytic and machine-learned baryonic property insertion into galactic dark matter haloes", *Monthly Notices of the Royal Astronomical Society*, 504(3), 4024.

(Benjamin Möws, June 2021)

Acknowledgements

I would, of course, like to thank my primary PhD advisor, Joe Zuntz, who was an indispensable well of knowledge and quippy wisdom. His support has been instrumental, and his willingness to take me and my ominous ideas about machine learning-driven cosmology on as part of a multi-year commitment baffles me to this day. I would like to apologize for my tendency to run off to spearhead collaborations all over the place, sometimes not even in my field, which must have been a source of frequent headaches. His support in terms of not only accepting but encouraging the freedom to explore various avenues was a driving force behind my continuing efforts and the completion of this thesis.

One of these collaborations, which ended up as a chapter, took place with Romeel Davé, who has taken me in as what felt like one of his own students. His uncanny ability to humorously cut through nonsense and unearth the root of a problem was vital for parts of this thesis. I would also like to thank Andy Taylor, my second PhD advisor, for our down-to-earth chats about my research, as well as Catherine Heymans, now the Astronomer Royal for Scotland, for both her insights into personal happiness as a vital factor for good research and kidnapping me as the 'statistics person' for a project. Further people at the Institute for Astronomy helped me along my journey, and listing them all would fill an entire pamphlet, so be assured that I have you in mind, and that I am thankful to you.

I would like to express my gratitude to the University of Edinburgh and its Institute for Astronomy for funding me through a Principal's Career Development Scholarship, as well as the Cosmostatistics Initiative led by Rafael de Souza, Emille Ishida, and Alberto Krone-Martins, which was instrumental for two chapters of this thesis. And, as each family can be viewed as a small organization itself, I want to thank my parents, Karin and Stefan Möws, whose consistent support of me chasing my dreams is what brought me here in the first place.

The most important entry should be placed at the end, and at least partially circles back to research due to having actually published together in another of those side projects that happened to be in neither of our actual fields. I want to thank Antonia Gieschen, my partner in crime, who was the foundation of my sanity during this journey. Her strength and unending support are what kept me going, and despite both of us thinking that the other one is the better researcher, I use this last line to insist that my hypothesis in that regard is correct.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

While astronomy is one of the oldest sciences in existence, cosmology in its current form is a comparatively new phenomenon triggered by a series of breakthroughs in both theory and observation in the early twentieth century, including milestones such as an improved understanding of gravity and the confirmation of other galaxies. Since then, a flurry of new advances took place, which include the discovery of the cosmic microwave background, the use of supernova surveys, and the emergence of dark matter and dark energy, with the latter sparking from the observation of an accelerating expansion of our Universe, to form what is now called the 'standard model' of cosmology. The wide variety of objects and structures of interest in cosmology, as well as their corresponding observations, means that cosmology as a field lives at different levels of granularity, from fundamental measurements like the dark energy equation of state and observations of the cosmic microwave background to large-scale weak lensing surveys and, at the lower end, single observations such as gravitational waves and inquiries into the inner working of galaxies. Due to a row of important observations and theoretical developments, combined with the availability of cheap and effective computing resources, the last two to three decades are also often referred to as the 'golden age' of cosmology.

Over roughly the same time frame, Bayesian statistics, which part of this thesis makes heavy use of, also experienced a sharp rise in popularity due to modern challenges. These include, but are not limited to, higher-dimensional parameter estimation and an interest in the distributions of such estimates. Apart from methodological advances, the same argument regarding computing resources also

applies to these developments and further boosted the field of computational statistics as a powerful tool for science. The field of astrostatistics emerged as the application of statistical methodology to astronomical data, including the development of new methods. This disposition to the 'in-house' creation of new approaches has a rich history in physics more generally, with the Markov chain Monte Carlo method being one of the primary examples.

More recently, machine learning has taken up a spotlight in the already statistics-heavy field of cosmology. Over the course of the last ~5 years, this former niche topic is now featured in dedicated sessions at many conferences and has sparked research groups across the globe. Indeed, during the development of the work featured in this thesis, we have seen machine learning develop from a topic viewed with some caution to a vibrant area of research. Debates on what fraction of modern machine learning is knowledge previously used under the umbrella of statistics are possible, and we dedicate some text to discussing the often close interplay between those areas. Some of the mentioned caution when approaching machine learning is, of course, not without justification, which we will discuss.

This thesis tackles a set of challenges in modern cosmology by going from the largest to more fine-grained scales, leveraging statistics and machine learning in the process. We start with cosmological parameter estimation in Chaper 2 and the dark energy equation of state via supernova surveys in Chapter 3. In Chapter 4, we then investigate large-scale structure and the identification of separate cosmic voids and troughs, and conclude with the insertion of baryonic properties into single simulated galaxies based only on dark matter information in Chapter 5. Lastly, Chapter 6 summarizes our findings and conclusions for the presented work.

The below introduction is split into two parts. In Section 1.1, we provide an introduction to modern cosmology, starting with the basics as well as the types and relevance of different observables. We cover the standard model and its alternatives, dark energy and its equation of state, large-scale structure in general and cosmic voids in particular, and end with the different types of cosmological simulations. Section 1.2, in contrast, covers the methodological background relevant to this thesis. This includes a short introduction to Bayesian analysis, an overview of parameter estimation and sampling methods with a discussion of recent advances, relevant developments in variational inference and Dirichlet processes, and the recent rise of machine learning with a focus on the methods and field relevant to our work on baryonic property prediction.

1.1 Modern cosmology and the dark sector

1.1.1 An introduction of the cosmological basics

Although the saying is traced back to Bernard of Chartres in the twelfth century, it usually is Isaac Newton who popularized the statement that one "stands on the shoulders of giants" when pursuing scientific endeavors. The concept emphasizes the progressive nature of science, building on the work and results of those that came before us and contributed to leaps in our understanding of the world around us. While he is, of course, now considered to be one of these giants himself, it certainly holds true that the history of scientific disciplines is built from the efforts of a countless number of humans trying to unravel the inner workings of the Universe, from the smallest to the largest scales.

Over the last few centuries, science also further diversified and specialized to an unprecedented degree. While this development inevitably heralded the end of the era of polymaths due to the amount of information accumulating in each field, it also allowed for an exponential growth of in-depth knowledge in a vast array of disciplines and including physics. Over the last few decades, scientific endeavors have also grown increasingly reliant on the pooling of both resources and researchers, with project teams of sometimes startling size in comparison to earlier times. As a result, while the saying about giants still rings true, we now collectively stand, so to speak, on each other's shoulders in a way that would most likely seem structurally unsound if the metaphor was taken too far. The following parts of this thesis will explain how some of the 'gigantic' developments relate to each other and form our current understanding of cosmology. As many of the described concepts are common throughout the literature, and citing the same text books after each equation does not seem like a sensible approach, we refer to Peacock (1999) and Dodelson (2020) at this point.

At present times, the evidence available to us strongly suggests that our Universe is expanding, which means distances between objects such as our own and other galaxies were smaller at previous times (Riess et al., 2016; Aghanim et al., 2018). This is encoded in the *scale factor* a, which thus was smaller at these previous times. By convention, its current value is set to $a = 1$, and it directly leads us to one of the most important observations leading up to the current era of cosmological research. The *redshift* of given electromagnetic radiation describes

an increase in its wavelength (which also means a decrease in photon energy and frequency) due to a number of causes. For the purpose of this introduction and in relation to the scale factor this pertains to *cosmological redshift*. Due to electromagnetic raditation such as light from a distant source being bound by the speed of light in a given medium as a constant, it experiences redshift as the wavelength is stretched when travelling through expanding space. Redshift is usually denoted as z and can, in terms of the expansion of the Universe, be defined as

$$1 + z \equiv \frac{\lambda_o}{\lambda_e} = \frac{a_o}{a_e} = \frac{1}{a_e}, \text{ with } a \equiv \frac{R(t)}{R_0}, \tag{1.1}$$

where λ describes a wavelength, and o and e denote observation and emission, respectively. R_0 is the scale factor evaluated at the current time and $R(t)$ is the time-dependent scale factor to quantify the distance between points in a universe relying on the *cosmological principle* at a given time t. The latter specifies that, at large-enough scales, we can consider (or at least approximate) the Universe as homogeneous and isotropic, meaning the same everywhere and in all directions (Liddle, 2003). The two other types of redshift in physics are relativistic, due to objects that are moving apart, and gravitational, due to electromagnetic radiation traveling toward an object in flatter spacetime, meaning a weaker gravitational potential.

The introduction of general relativity by Einstein (1915), enabled the development of cosmology as a modern scientific discipline driven by testable hypotheses (Einstein, 1917). The relationship between the curvature of spacetime and the contained energy density is expressed in Einstein's Gravitational Field Equations,

$$R_{ab} - \frac{1}{2} R g_{ab} = -8\pi G T_{ab}. \tag{1.2}$$

Here, R_{ab} and R denote the Ricci tensor and scalar, respectively, which are derivatives of g_{ab} as the metric tensor with respect to coordinates. While G is the universal gravitational constant, T_{ab} denotes the energy-momentum tensor that encodes information about the energy density distribution (Misner et al., 1973). In an iostropic smooth universe, said tensor takes, for density ρ and pressure p,

the form of

$$T_{ab} = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}. \tag{1.3}$$

Due to the equations being tensorial, we are presented with invariance under coordinate transformations. The issue with the equations is that they are highly non-linear and, while some analytic solutions exist for a number of symmetric cases, there are no known general solutions. In a scenario featuring the cosmological principle, the assumed symmetry can be used to constrain the metric tensor in the *Friedmann-Lemaître-Robertson-Walker* (FLRW) metric,

$$ds^2 = g_{ab}dx^a dx^b = dt^2 - R(t)^2[dr^2 + S_k^2(r)(d\theta^2 + \sin^2\theta d\phi^2)], \tag{1.4}$$

where ds denotes a universally agreed-upon spacetime interval between two given events (Friedmann, 1922; Lemaître, 1927; Robertson, 1935; Walker, 1937). In addition, r, θ, and ϕ represent spherical polar coordinates, with the radial coordinate r as a dimensionless comoving coordinate. The FLRW metric's symmetry leads to the simple form of T_{ab} described above. To make use of $S(k)$, we require the *Friedmann equations* to describe the relationship between the curvature constant k and R,

$$\left(\frac{\dot{R}}{R}\right)^2 \equiv H^2 = \frac{8\pi G}{3}\sum_i \rho_i(R) - \frac{k}{R^2}, \tag{1.5}$$

with H as the *Hubble parameter* (see Hubble, 1929), c as the speed of light in vacuum, and ρ as the total energy density for which we sum over all energy-density components. $S(k)$ then is a geometry-dependent function of a given universe's curvature,

$$S_k = \begin{cases} \sin r, & k = 1 \\ r, & k = 0 \\ \sinh r, & k = -1 \end{cases}. \tag{1.6}$$

Here, the first, second, and third cases correspond to the curvature of a given universe and are called 'closed', 'flat', and 'open' geometries. This can easily be imagined by subtracting one dimension for parallel lines on a spherical, flat, and saddle-shaped sheet, respectively. The geometry of a universe is dependent on the total energy density, with a critical value of $\rho_{\text{crit}} \approx 1.9 \cdot 10^{-29}\text{g cm}^{-3}$. For a lower and higher value, we face a closed and open universe, respectively, as a higher mass density leads to the universal Hubble expansion eventually halting and reversing, presenting us with a closed Universe. A total energy density equal to said critical value, on the other hand, presents us with a flat (or 'Euclidean') universe, which is the scenario generally supported by currently available data.

We can view the FLRW metric as an energy equation, with expansion-driven kinetic energy on the left side, while the first and second term on the right side describe the potential and constant total energy in terms of the curvature. In doing so, the matter present in a given universe, as well as its expansion rate, determine its energy content. This becomes apparent if we write down the curvature constant in relation to the present-time scale factor,

$$k = R_0^2 \left(\frac{8\pi G \rho_0}{3} - H_0^2 \right), \tag{1.7}$$

for the present-time values of the *Hubble constant* and the total energy density, H_0 and ρ_0, where the former is the current value describing the expansion of our Universe. Here, we see the reason for the conventional definition of R_0 in Eq. 1.1, as R_0 becomes undefined for a flat universe with $k = 0$. Each component of the energy density of a given universe affects the expansion or contraction of the latter. The *continuity equation* offers a description of the conservation of the stress-energy tensor,

$$\dot{\rho}_i + 3H(\rho_i + p_i) = 0. \tag{1.8}$$

Solutions for one of the components involved in the above equation require an *equation of state* that describes the relationship between those two factors, which will be discussed in Section 1.1.4. For now, we note that the reason of each component affecting a given universe in a different manner is due to the relationship between pressure and scale factors as follows. The matter density changes according to $\rho_m \propto a^{-3}$ due to the conservation of particles, the radiation

density according to $\rho_r \propto a^{-4}$ because of relativistic effects, and dark energy as one of the primary constituents of the standard model of cosmology remains constant, with its pressure denoted as ρ_Λ. This leads us to a rewriting of the Friedmann equation as

$$\left(\frac{\dot{a}}{a}\right)^2 \equiv H^2 = \frac{8\pi G}{3}\left(\frac{\rho_m}{a^3} + \frac{\rho_r}{a^4} + \rho_\Lambda\right) - \frac{kc^2}{R_0^2 a^2}. \tag{1.9}$$

We can bring the above in a more convenient form by declaring dimensionless parameters so that, for Ω as the sum of said dimensionless parameters,

$$\left(\frac{\dot{a}}{a}\right)^2 = H_0^2\left(\frac{\Omega_m}{a^3} + \frac{\Omega_r}{a^4} + \Omega_\Lambda + \frac{1-\Omega}{a^2}\right), \text{ with } \Omega_i(a) = \frac{8\pi G}{3H^2}\rho_i(a). \tag{1.10}$$

The family of *Big Bang cosmologies* emerges when we have a beginning at $a = t = 0$. The latter holds, for example, true for a universe with flat geometry that is comprised exclusively of matter, with $a \propto t^{2/3}$, or exclusively of radiation, with $a \propto t^{1/2}$. In the case of such a universe being comprised of only *vacuum energy*, however, meaning an underlying background energy as a property of space, we get $a \propto \exp(H_0 t)$ and are left with no starting point. From Eq. 1.9, we arrive at the realization that a universe in such a framework needs to either expand or contract. In the case of parameters leading to a scaling parameter of zero in the future, the resulting cosmologies are known *Big Crunch cosmologies*, mirroring the 'Big Bang' nomenclature (see, for example, Hertog & Horowitz, 2004).

1.1.2 Cosmological observations and their utility

In order to discuss cosmological observations, we will first have a look at distance calculations. Returning to redshift as defined in Eq 1.1, given a radial null geodesic and noting that the source's comoving coordinate does not depend on the expansion, we can write the metric down as

$$ds^2 = c^2 dt^2 - R^2(t) dr^2 = 0 \Rightarrow r = \int_{t_e}^{t_o} \frac{dt}{R(t)} = \int_{t_e+\Delta t}^{t_o+\Delta t} \frac{dt}{R(t)}, \tag{1.11}$$

by solving for the radial distance to arrive at the comoving radial distance between emission and observation. As the comoving coordinates of galaxies are not subject to change, save for peculiar velocities, shifting the time frame via Δt does not affect the validity of the above expression due to the comoving distance remaining constant. Given our understanding of redshift from Eq. 1.1 further above, we can now rewrite the comoving distance, with the integral limits depending on the target quantity at hand, as

$$r = \frac{1}{R_0} \int \frac{da}{a^2 H(a)} = \frac{1}{R_0} \int \frac{dz}{H(z)}. \tag{1.12}$$

In the case of Big Bang cosmologies, which include the standard model of cosmology, we are faced with a finite time and, accordingly, a distance limit that light could have traveled in said time. Given the latter, the *particle horizon* is that maximum distance available for light particles to travel since $R = 0$,

$$R_0 r_p(z) = R_p(z) = \int_0^t \frac{dt}{R(t)} = \int_z^\infty \frac{dz}{H(z)}, \tag{1.13}$$

with the caveat that $R(t)$ needs to be monotonic for the right-most part of the above (Liddle & Lyth, 2000). The limit for the comoving distance for travel at the speed of light, the *event horizon*, can be calculated and includes the same caveat as above,

$$R_0 r_p(z) = R_p(z) = \int_{t_0}^\infty \frac{dt}{R(t)} = \int_{-1}^0 \frac{dz}{H(z)}, \tag{1.14}$$

and the question of whether the above is finite or not depends on the universe in question. This is, for example, the case for 'Big Crunch' cosmologies in which $t = \infty$ is not an option.

While $R_0 r_p$ is generally referred to as the *physical distance*, two other metrics are of primary relevance in cosmology. The first is the *luminosity distance*, conventionally denoted as D_L, which is subject to satisfying the below flat-space relation between bolometric Flux (F) and luminosity (L),

$$D_{\mathrm{L}} = (1+z) R_0 S_k(r) \text{ so that } F = \frac{L}{4\pi D_L^2}. \tag{1.15}$$

The second metric is the *angular diameter distance*, commonly written as D_{A}, which defines a distance in terms of an object of size ℓ subtending an angle $\delta\theta$ on the sky so that

$$D_{\mathrm{A}} = \frac{\ell}{\delta\theta} = \frac{R_0 S_k(r)}{1+z}, \tag{1.16}$$

with the right-most part of the equation relying on the metric defined in Eq. 1.4. In an expanding universe, objects are bound to appear larger than they would otherwise, while the positive redshift also makes them appear dimmer. The luminosity distance is greater than the angular diameter distance, with $D_{\mathrm{L}}/D_{\mathrm{A}} = (1+z)^2$. The observation that we live in an expanding Universe, which is a cornerstone of our modern understanding of cosmology, was initially made by Slipher (1915), measuring the recession speed of galaxies (v). This observation was then cemented by work by Hubble (1929), describing a linear relation between v and x as the distance,

$$v = H_0 x, \tag{1.17}$$

with H_0 being the current value of the Hubble parameter as introduced in Section 1.1.1. As mentioned before, H_0 is generally referred to as a constant, although that is not technically the case due to its dependence on time. For objects in our neighborhood, meaning local objects for which distance measures converge, H_0 can also be calculated through measurements due to the relation to redshift on this scale, $z = H_0 x$. The known drawback of measuring the redshift in this context is that it can be affected by physical factors such as peculiar and internal velocities of galaxies, which requires us to measure large-enough numbers of galaxies to address and correct these issues. For Hubble's work, so-called *Cepheid Variables*, stars with a narrowly defined relationship between their luminosity and pulsation periods, played an important role (Leavitt, 1908). The luminosity of these stars can then be calculated using said periods, inferring the distance through the received flux.

Now that we are aware of distances as commonly used in cosmology, we can also move on to one of the primary observables used in the field, which extends our reach in terms of distance when compared to Cepheid Variables and played an important role in establishing the standard model of cosmology. Named

after what appears on the sky like a very luminous ('super-') new ('nova') star, *supernovae*, often abbreviated to 'SNe' in the plural, are powerful stellar explosions. As such, they qualify as transient events and signify the end of a star's life as either white dwarfs[1] or massive stars, and collapse into black holes or neutron stars, or are completely destroyed. Their luminosity sets them apart from the more common novae, being able to temporarily outshine their host galaxy. In the case of massive stars, they are the result of the core's sudden gravitational collapse, while white dwarfs transitioning into a supernova are the result of a sudden reignition of nuclear fusion. They are generally classified into type I and type II, with the latter spectrum containing hydrogen lines (Filippenko, 1997; Weiler & Sramek, 1988).

For the purpose of cosmology, and especially in the context of this thesis, *type Ia supernovae* (SNe Ia) are the primary subject of interest (Branch, 1998). They exhibit a strong ionized silicon absorption line, specifically a singly ionized silicon line at 615 nm near peak luminosity. If a massive-enough carbon-oxygen white dwarf reaches the Chandrasekhar limit, the maximum mass of a (non-rotating) stable white dwarf at $\sim 1.44\ M_\odot$ (solar masses), the pressure from electron degeneracy is unable to counteract the gravitational force and triggers a collapse. Further factors, however, paint a slightly more nuanced picture, with metallicity and spin being able to influence the collapse. Due to their characteristic light curve, we can use SNe Ia for distance measurements, although a correction is required. The latter is the so-called Phillips relationship between SN Ia peak luminosity and the luminosity evolution speed after maximum light (Phillips, 1993). Specifically, the decline in B-band magnitude light curve from maximum light within 15 days, denoted as Δm_{15}, is used to express the relation with the intrinsic B-band magnitude,

$$M_{\max}(B) = -21.726 + 2.698\Delta m_{15}(B). \tag{1.18}$$

As such, SNe Ia are not technically *standard candles*, meaning an astronomical object with known magnitude, but they are 'standardizable' candles. Other, more accurate methods were developed afterwards, for example the multi-color light curve shape method (MLCS) by Riess et al. (1996), parameterizing SN light curve shape as a function of their maximum absolute magnitude, and the 'stretch

[1] In a shocking deviation from astronomers' usual fondness of naming things after elements of J. R. R. Tolkien's works, including asteroids, geographical features of moons and other celestial objects, and software, the field proves resistant to the plural form 'dwarves'.

method', which works based on the time stretching of a canonical light curve to represent the entire range of SN Ia light curves in the B-band and V-band, providing a parameterized set of light curve shapes (Perlmutter et al., 1997, 1999). In terms of distance measurements, for an SN Ia observed at low redshift of $z \ll 1$, we can approximate the luminosity distance as

$$D_{\mathrm{L}} = \frac{z}{H_0}\left(1 - \frac{z}{4}(\Omega_m - 2\Omega_\Lambda - 2)\right). \tag{1.19}$$

Here, we encounter Ω_Λ from Eq. 1.10 again, which denotes the ratio between the energy density from the cosmological constant and the Universe's critical density. The latter constant can be introduced into the Einstein Field Equations in Eq. 1.2 to produce universes with an accelerating expansion, which will be discussed in more detail in Section 1.1.3 when we arrive at the standard model of cosmology,

$$R_{ab} - \frac{1}{2}Rg_{ab} + \Lambda_{ab} = -8\pi G T_{ab}. \tag{1.20}$$

In terms of cosmological models, one of the features that Big Bang cosmologies have in common is a comparatively hot early stage, as well as light from that time we should be able to observe. These remnants, popularized by Dicke et al. (1965), are known as the *cosmic microwave background* (CMB), the discovery of which happened rather by fortuitous accident. When working with a microwave horn antenna originally constructed for passive communications satellites, Penzias & Wilson (1965) detected what was later found[2] to be remnant electromagnetic radiation from the early universe.

In the standard model, about 378,000 years after the beginning and due to cooling via expansion, charged protons and electrons bound to neutral hydrogen, which is generally referred to as *recombination*. Resulting in the decoupling of baryon plasma and photons, as well as no quick reabsorption, the latter were free to travel throughout the universe, leading to the CMB (Ydri, 2017). Through expansion, the Universe experiences a decrease in temperature due to the resulting decrease in density. The temperature observation as an ideal black body of around 2.73 K can be expressed in terms of a total energy density in radiation by using the

[2]On a humorous note, they at first thought bird droppings inside of the horn antenna were to blame for the perceived interference, which they tactfully described as 'white dielectric matter'.

Steffan-Boltzmann constant with $\sigma = 5.67 \cdot 10^{-8}$ W m^2 K^{-4},

$$\rho_r = \frac{4\sigma T_0^4}{c^3}. \tag{1.21}$$

The estimated CMB temperature evolution with redshift is corroborated by rotational excitation of molecules and the Sunyaev-Zel'dovich effect (Noterdaeme et al., 2011; Luzzi et al., 2015). With only matter and radiation being relevant in the early Universe, the Friedmann equation can be written as

$$\frac{\dot{a}}{a} = H_0^2 \left(\frac{\Omega_r}{a^4} + \frac{\Omega_m}{a^3} \right), \tag{1.22}$$

and even earlier, when the energy density of radiation dominates, the growth rate is $a \propto t^{1/2}$. That relation changes with the growing relevance of matter,

$$H_0 t = \frac{2\Omega_r^{3/2}}{3\Omega_m^2} \left(\left(\frac{\Omega_m}{\Omega_r} a - 2 \right) \sqrt{1 + \frac{\Omega_m}{\Omega_r} a} + 2 \right). \tag{1.23}$$

The time of matter-radiation equality (t_{eq}) can then be calculated as

$$t_{\mathrm{eq}} = 13.04 \frac{\Omega_r^{3/2}}{\Omega_m^2} h^{-1} \text{ Gyr due to } a_{\mathrm{eq}} = \frac{\Omega_r}{\Omega_m} \text{ with } \Omega_m a_{\mathrm{eq}}^{-3} = \Omega_r a_{\mathrm{eq}}^{-4}. \tag{1.24}$$

The observation that the CMB is isotropic brings us back to the cosmological principle, which extends this observation to homogeneity and results (on large-enough scales) in a smooth Universe. The CMB also features regions that do seem to share a causal relationship, although that should not be the case due to their distance given the Universe's age when calculating the size of the particle horizon, which led to the concept of *inflation*, a period of exponential expansion roughly 10^{-34} s after the Big Bang, which will be described further in Section 1.1.3. Together with the Hubble–Lemaître law on the distance-redshift relation, the CMB constitutes a landmark in establishing the Big Bang family of cosmologies, and we have access to increasingly more fine-grained maps through surveys such as COBE, WMAP and Planck (Bennett et al., 1996, 2013; Planck Collaboration et al., 2020a). Figure 1.1 shows CMB maps from the `Planck` 2018 results taken

from Planck Collaboration et al. (2020b) and using spectral matching independent component analysis (SMICA) (Cardoso et al., 2008).

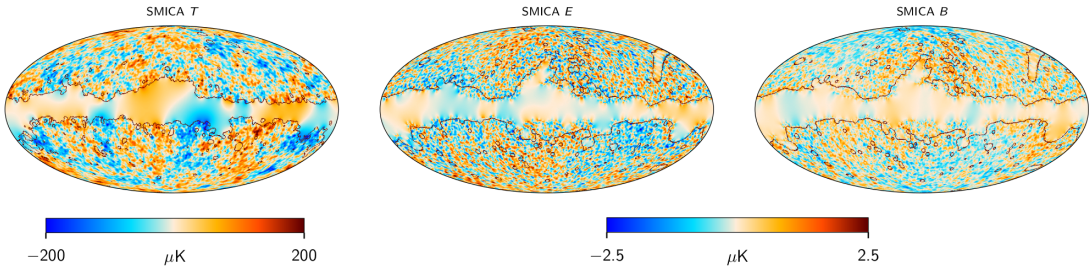

Figure 1.1 *Component-separated CMB maps at 80' resolution, with columns depicting, from left to right, temperature, E-mode and B-mode maps, as shown in Planck Collaboration et al. (2020b). The left-hand temperature map is inpainted within the common mask, and monopoles and dipoles are subtracted from the temperature maps, with parameters fitted to the data after masking.*

Another large-scale measurement in cosmology, which is of special relevance for this thesis, as both Chapter 2 and Chapter 4 make use of it, is *weak gravitational lensing* (Kaiser, 1998). On a more general note, gravitational lensing is the perturbation of the path of photons from a source by the gravitational field of massive objects, effectively deviating from the null geodesics along which light otherwise travels. The light from a source is thus displaced, with a deflection angle $\hat{\alpha}$, the mass of the lensing object M, and an impact parameter ξ, as

$$\hat{\alpha} = \frac{4GM}{c^2\xi} \text{ with } \xi \gg R_s \equiv 2GMc^{-2}, \tag{1.25}$$

with the right-hand part describing the condition that the lensing object's Schwarzschild radius, R_s, and the impact parameter need to satisfy. Let the angular diameter distances between observer and lensing object, observer and source, and lensing object and source be denoted as D_A^{ol}, D_A^{os}, and D_A^{ls}, respectively, as well as the observed angle if there were no lensing taking place as β and the observed angle with lensing as θ, then we can rewrite Eq. 1.25 through the small angle approximation to retrieve the lensing equation as

$$\theta D_A^{\mathrm{os}} = \beta D_A^{\mathrm{os}} - \hat{\alpha} D_A^{\mathrm{ls}} \Rightarrow \theta = \beta - \alpha \text{ with } \alpha = \hat{\alpha}\frac{D_A^{\mathrm{ls}}}{D_A^{\mathrm{os}}}. \tag{1.26}$$

The source as described above is sometimes also referred to as the 'backlight',

as the object of interest in lensing is often the lensing object between said source, or backlight, and the observer, and confusion can arise when referring to the lensing object as the source of lensing in addition. Gravitational lensing is another observable that played an important role in establishing our current understanding of cosmology, specifically due to its role in experimentally validating general relativity through the observed alteration of star positions by the Sun that is larger than what Newtonian physics would predict (Dyson et al., 1920). It can be split into strong lensing, microlensing, and weak lensing, with the first describing the arc-like structures or multiple images of the same source galaxy caused by massive foreground objects. Microlensing, on the other hand, is caused by massive and sufficently compact objects such as stars and its subsequent evolutions passing in front of source objects. As a transient event, this leads to a brief spike in the source object's light curve.

In contrast to these localized observations, weak lensing pertains to larger-scale coherent distortions of observed galaxy shapes in surveys. While not as visually appealing due to measurements relying on a large number of galaxies with minuscule changes to orientation and shape, or *shear*, due to large-scale structure, it is especially relevant to trace dark matter and estimate cosmological parameters (Munshi et al., 2008). The estimation of the two-point shear correlation function, denoted here with $\xi_{\pm}(\theta)$ for angular separations θ, can be written by summing over galaxy pairs $\{\mathrm{a}, \mathrm{b}\}$,

$$\hat{\xi}_{\pm}(\theta) = \frac{\sum_{\mathrm{ab}} w_{\mathrm{a}} w_{\mathrm{b}} (\epsilon_t(\theta_{\mathrm{g,a}})\epsilon_t(\theta_{\mathrm{g,b}}) \pm \epsilon_{\times}(\theta_{\mathrm{g,a}})\epsilon_{\times}(\theta_{\mathrm{g,b}}))}{\sum_{\mathrm{ab}} w_{\mathrm{a}} w_{\mathrm{b}}}, \tag{1.27}$$

where w_i is used to weight the ellipticity measurements, and with ϵ_t and $\epsilon_{\pm}$ denoting tangential and cross-components of ellipticities relative to $\theta_{\mathrm{g,a}} - \theta_{\mathrm{g,b}}$ as the connecting vector. We can relate measurements of the shear correlation function, for given tomographic bins i and j, to the angular convergence power spectrum ($P_{\kappa}^{ij}(\ell)$), through the following integrals,

$$\hat{\xi}_{+}^{ij}(\theta) = \frac{1}{2\pi} \int d\ell \ell J_0(\theta\ell) P_{\kappa}^{ij}(\ell), \tag{1.28}$$

$$\hat{\xi}_{-}^{ij}(\theta) = \frac{1}{2\pi} \int d\ell \ell J_4(\theta\ell) P_{\kappa}^{ij}(\ell), \tag{1.29}$$

with J_n as the n^{th}-order first-kind Bessel function, and assuming that the Universe

is flat. $P_\kappa^{ij}(\ell)$ can be related to the matter power spectrum (P_δ) (see below),

$$P_\kappa^{ij}(\ell) = \int_0^{rH} dr \frac{q^i(r) q^j(r)}{D_A(r^2)} P_\delta \left(\frac{\ell + 0.5}{D_A(r)}, r \right), \tag{1.30}$$

by using the Limber approximation in its harmonic-space version. To see the relationship of shear measurements and cosmological parameters, we can take a look at the lensing efficiency, $q(r)$, with r_H as the comoving radial distance with regard to the horizon,

$$q^i(r) = \frac{3}{2}\Omega_m \left(\frac{H_0^2}{c^2} \right) \frac{D_A(r)}{a(r)} \int_r^{r_H} \mathrm{d}r' \; n^i(r') \frac{D_A(r' - r)}{D_A(r')}, \tag{1.31}$$

with $n^i(r')$ being the effective number density of galaxies such that $\int d\mathcal{X} n^i(r) = 1$. Shear measurements are particularly useful for constraints on σ_8 and Ω_m as cosmological parameters. In contrast to shear maps, *convergence maps*, or weak lensing mass maps, show the integrated total matter density along the line of sight, using a lensing kernel peaking approximately half-way between source and observer. Convergence is a dimensionless surface mass density, denoted here as κ, and can be written in its general form as

$$\kappa(\theta) = \frac{\Sigma(\theta)}{\Sigma_{\mathrm{crit}}} \text{ with } \Sigma_{\mathrm{crit}} = \frac{c^2}{4\pi G} \frac{D_{os}}{D_{ol} D_{ls}}, \tag{1.32}$$

where Σ_{crit} is the critical surface mass density of the lens. Up to a constant, one can convert between convergence and shear due to both being second derivatives of the lensing potential. In summary, shear influences a 'stretch' of observed galaxy shapes, with shear components being responsible for the orientation, while convergence results in enlarged galaxy images.

While weak lensing is accessed through photometric surveys, another observable is explored through spectroscopic surveys. As another component featured in the matter power spectrum, and stemming from the early stages of the Universe, *baryon acoustic oscillations* (BAOs) are caused by the pre-recombination photon-baryon plasma (Eisenstein, 2005). With the connection between the latter, radiation pressure, and gravitational interaction, this leads to perturbations trying to collapse increasing in pressure, resulting in oscillating sound waves and

thus their name. As dark matter interacts gravitationally with these baryonic perturbations, this effect can be found in the matter power spectrum. Similar to SN Ia observations as standard(izable) candles, measurements of BAO peaks help us to constrain cosmology, specifically via $\Omega_m h^2$ and $\Omega_b h^2$, providing us with constraints on the Hubble constant under the standard model of cosmology. With precise measurements of the sound horizon, BAOs serve as a 'standard ruler' for length scale by comparing today's sound horizon via galaxy clustering to that during recombination derived from the CMB.

In recent years, another type of observable received a lot of attention. *Gravitational waves* are disturbances in the curvature of spacetime, which were initially hypothesized by Poincaré (1906) and later predicted based on general relativity (Einstein, 1916, 1918). While a separate type of radiant energy, parallels can be drawn (to a degree) to electromagnetic radiation, and they offer yet another way to confirm general relativity when compared to Newtonian gravity, which involves instantaneous propagation of gravitational effects. The first verified gravitational wave signal was named GW150914 due to its discovery on 14 September 2015, and subsequently reported by the LIGO Scientific Collaboration (LSC) on 11 February 2016 (Abbott et al., 2016b). For the purpose of this thesis, and as opposed to type Ia supernovae, gravitational waves are not a source used in later chapters, but a mention should be made in response to an increased interest in multi-messenger astronomy as the effort to combine multiple astronomical source types (Bartos & Kowalski, 2017).

1.1.3 The standard model of cosmology and beyond

The observational milestones we covered in the previous section contributed to what is now referred to as the 'standard model of cosmology' (or the 'concordance model') due to being generally accepted as the best-fitting model of our Universe that we currently have. More formally, the model in question is called the ΛCDM *model*, which we will now have a closer look at. This thesis has, so far, mentioned *dark matter* in passing, but the latter is a central concept in modern cosmology. Early indications include the observation that galaxy clusters show a velocity dispersion that should be lower given their stellar mass and hot gas, as well as rotation speeds of stars and gas in galaxies in excess of what their baryonic content would suggest (Zwicky, 1937; Oort, 1940; Rubin et al., 1980). The sum of these observations strongly pointed toward there being more gravity than mass

observed from electromagnetic interactions, which is where the 'dark' in dark matter stems from. More specifically, dark matter refers to an as of yet unknown form of matter that seems to interact only gravitationally with itself and baryonic matter. One intial explanation in the literature deals with *massive compact halo objects* (MACHOs), high concentrations of matter like, for example, black holes and sufficiently dark stars (Alcock et al., 2000).

Aside from events such as recombination, another epoch of interest in the evolution of the Universe is known as *big bang nucleosynthesis* (BBN), which occured within the first 20 minutes after the Big Bang and deals with the rate of nuclear reactions and their relationship to the baryon density in the early Universe (Gamow, 1948). As the latter is made up of fundamental particles at very early stages, and temperatures are high enough to trigger nuclear reactions due to the density, neutrons and protons convert to deuterium, helium, and a number of heavier nuclides. We can make use of this by measuring the abundances of said nuclides in gas reservoirs that are left over from that time, and constrain the nuclear reaction rate. This provides us with estimates of the baryon density in the early Universe and leads to a baryon density of $\mathbf{\Omega}_b \sim 0.05$, allowing us to constrain the standard model (e.g., Cyburt et al., 2016; De Souza et al., 2019a,b). As observational evidence in cosmology points toward a total matter density of $\mathbf{\Omega}_m \sim 0.3$, BBN is one of the pieces of a puzzle that led to the general acceptance of dark matter as a part of our Universe, as this difference in densities for total and baryonic matter conflicts with an explanation relying on MACHOs, and subsequent efforts to detect the latter through microlensing have done little to strengthen them as the explaining factor.

Candidates for this (or these) fundamental particle(s) that constitute dark matter are plenty, with one class being referred to as *weakly interacting massive particles* (WIMPs), featuring weak interaction, but lacking electromagnetic interaction, and demonstrating the continued excellence of the field when it comes to conjuring up sensible acronyms. Other candidates include the *axion*, which offers a solution to the strong charge-parity problem in quantum chromodynamics and forms a Bose-Einstein condensate belonging to the *cold dark matter* (CDM) family (Peccei & Quinn, 1977). The latter describes dark matter proposals that exhibit a free streaming length (FSL) much smaller than a protogalaxy. Here, the FSL is the distance which objects move as a result of random motion in the early Universe before being slowed by the latter's expansion.

Conversely, *warm dark matter* (WDM) and *hot dark matter* (HDM) describe

particle proposals that feature an FSL comparable to and much larger than that of a protogalaxy, respectively (Bode et al., 2001; Primack & Gross, 2001). Another way to look at them is to view CDM, WDM and HDM proposals as decoupling while non-relativistic (with a number density akin to photons), semi-relativistic, and relativistic, in that order. This 'temperature' of dark matter proposals can be constrained in terms of their velocity at decoupling due to 'colder' and 'hotter' proposals resulting in less and more smooth structure in the Universe, respectively. Current observations suggest CDM as the most likely proposal, thus making it part of the standard model.

The statement about $\Omega_m \sim 0.3$ above does, of course, leave us with the question what the remaining energy density of the Universe is composed of. At the same time, we have specified that the standard model is referred to as 'ΛCDM', while not addressing the Λ in question so far. The answer lies in the late-time acceleration of the Universe (see, for example, Riess et al., 1998; Perlmutter et al., 1999; Peebles & Ratra, 2003), for which supernova surveys delivered the first evidence due to the observation that SNe Ia that are further away exhibit higher redshifts than we would expect without an acceleration. Other observables that cemented late-time acceleration as part of the standard model include BAOs and galaxy clustering, and, more recently, gravitational waves as described in Section 1.1.2 are expected to further aid in these efforts (Ur Rahman, 2018). As demonstrated earlier in Eq. 1.20, the Einstein Field Equations can be modified to produce universes with an accelerating expansion. We can then rewrite the Friedmann Equation in Eq. 1.5 as

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho(a) - \frac{kR_0^2}{a^2} + \frac{\Lambda}{3}. \tag{1.33}$$

While such an addition does, technically, affect gravitation in general, including at the galaxy level, the value of Λ necessary to satisfy our observation of late-time acceleration is small enough to not have a significant effect at that scale. While this affects the G_{ab} part of Eq. 1.5, we can also change the stress-energy tensor, T_{ab}, in the same equation by introducing a homogeneous vacuum energy that conforms to $p = -\rho$ as well as $\dot{\rho}_\Lambda = 0$. The benefit of this approach is that it conforms to predictions of quantum field theory (QFT). One issue with that, however, is that the corresponding calculation only takes on finite values if a cut-off value for some energy scale is used, as QFT holds only up to this value. This gives rise to the so-called cosmological constant problem, as the theoretical zero-

point energy value of QFT is larger than the observed value of vacuum energy density by a comically large factor, which is one of the reasons why a combination of gravity and quantum mechanics remains a topic of high interest.

In this context, *dark energy*, which will be covered in terms of its equation of state in Section 1.1.4, is a low-density form of energy that constitutes around 70% of the total energy in the observable Universe, meaning $\Omega_\Lambda \approx 0.7$, of which the cosmological constant as described above is one explanatory variety. The question whether a more detailed model of the Universe in terms of general relativity, as well as a successful merging with quantum mechanics, will do away with the need for dark energy remains a topic of debate, and currently investigated alternatives will be touched upon further below and in the next section. As a visual aid, Figure 1.2 shows the rough timeline of the Universe with the events previously described, as well as galaxy formation, with their respective redshift values.

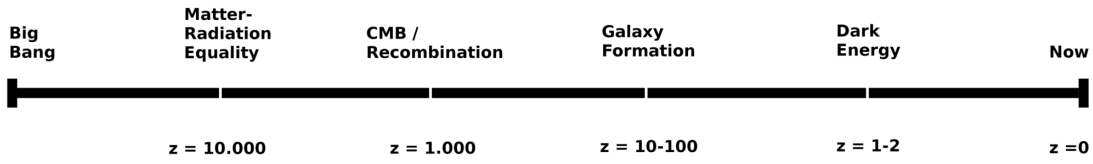

Figure 1.2 *Schematic representation of important events in the history of the Universe along the cosmological timeline from left to right, with the approximate corresponding redshift values.*

ΛCDM universes belong to the family of Big Bang cosmologies, meaning they feature a starting point at $a = t = 0$ and no 'Big Crunch' scenario due to $a, t \to \infty$. The Friedmann equation for such universes, when dealing with flat universes based on matter and with a cosmological constant, is

$$H^2 = H_0^2 \left(\frac{\Omega_m}{a^3} + 1 - \Omega_m \right), \text{ so } a(t) = \left(\left(\frac{\Omega_m}{1 - \Omega_m} \right) \sinh^2 \left(\frac{3}{2} \sqrt{1 - \Omega_m} H_0 t \right) \right)^{1/3} \quad (1.34)$$

when solving said equation for the time-dependent scale factor as described in Section 1.1.1. Consequently, we can use this information to calculate the age of the given universe as

$$t_0 \approx 6.52 \frac{1}{1 - \Omega_m} \operatorname{asinh} \left(\sqrt{\frac{1 - \Omega_m}{\Omega_m}} \right) h^{-1} \text{Gyr}. \quad (1.35)$$

As previously indicated in Section 1.1.2, the CMB features regions that seem to share a causal relationship despite the Universe's age and the particle horizon indicating otherwise, leading to the concept of inflation as a period of rapid and exponential expansion around 10^{-34} s after the Big Bang. In this context, if we let a be a scale factor with $a \propto t^{\alpha}$ so that $\alpha > 1$, we can use Eq. 1.8 and differentiate Eq. 1.5 to arrive at the *Friedmann acceleration equation*,

$$\frac{\ddot{a}}{\dot{a}} = \frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right), \text{ with } pc^2 + 3p < 0, \tag{1.36}$$

as a negative pressure to satisfy an acceleration as implied by $\alpha > 1$ above. The *de Sitter solution* offers a solution to Eq 1.5 for a flat universe dominated by dark energy, introducing an exponential expansion through

$$a \propto e^{Ht}, \text{ where } H = \sqrt{\frac{8\pi H \rho_{\Lambda}}{3}}. \tag{1.37}$$

One notable effect of inflation is that it promotes a flat geometry, meaning that curved universes would have their geometries 'flattened' in the process, making it difficult to differentiate them from a truly flat universe. To demonstrate this, we can have a look at the total density parameter's evolution for a homogeneous universe,

$$1 - \Omega(a) = \frac{(1 - \Omega_{\Lambda})}{a^2 \Omega_{\Lambda}}, \tag{1.38}$$

with $\Omega(a) \rightarrow 1$ for increasing values for a, leading to an ever-more-distant curvature scale. Importantly, inflation also provides an explanation for structure observed in the Universe, especially with regard to seemingly causal connections between different parts of the sky. Quantum-mechanical fluctuations at the beginning of this period would rapidly expand beyond their limits of causal contact, imprinting these fluctuations into the density pattern of the Universe. While inflation is not reflected in the name of the ΛCDM model, it usually is a part of the standard-model view of our Universe due to its explanatory power.

As the name 'standard model' suggests, there are non-standard models of cosmology as well. Throughout history, cosmological models have occurred as

part of the progressive nature of science, with early examples including the so-called 'Copernican Revolution' in favour of the heliocentric model and the 'Great Debate' mentioned before, which established the existence of galaxies other than our own. More recently, the general acceptance of the Big Bang point of view provides another dominant example. While the ΛCDM model is generally accepted to provide the best fit to available observations, alternative theories exist and are the subject of ongoing research.

With regard to the 'CDM' part of the standard model, we have already covered warm and hot dark matter as alternatives. Another dark matter alternative that is the subject of current research is *modified Newtonian dynamics* (MOND), which covers a modification of Newton's laws for low-acceleration environments and belongs to the larger family of *modified gravity* proposals (Milgrom, 1983). For the cosmological constant, similar alternatives exist, for example *quintessence* as a scalar field to explain dark energy, with 'phantom energy' as a special case, which is linked to the dark energy equation of state covered in Section 1.1.4 Peebles & Ratra (1988). In this context, *holographic dark energy* refers to the notion that the origin of dark energy lies within quantum fluctuations that are limited by our Universe's event horizon (see Wang et al., 2017, for a review).

In addition, as a non-random component of galaxy clusters' peculiar velocity, *dark flow* describes a debated 600-1000 km/s flow toward a 20-degree patch of the sky based on WMAP data (Kashlinsky et al., 2008). Alternative cosmologies such as *scalar-tensor theories* and *massive gravity* employ screening mechanisms for compatibility, decoupling a proposed fifth force from matter living in high-density regions like galactic interiors and making the fifth force a function of environment (Desmond et al., 2019). These screening mechanisms would have the least effect in highly underdense regions, making voids an ideal test bed for investigations, which will be important in Chaper 4. For a more general overview of alternative cosmologies, including further proposals such as $f(R)$ gravity and exotic dark matter, which is a too-broad topic to cover exhaustively in this introduction, we refer the interested reader to suitable reviews (Narlikar & Padmanabhan, 2001; Pardo & Spergel, 2020).

1.1.4 Dark energy and its equation of state

As described before, the standard model of cosmology, in which the Universe is composed primarily of cold dark matter and a cosmological constant, is mainly

supported by three observational pillars: BBN, CMB, and the discovery of late-time accelerating cosmic expansion, while the discovery of accelerated cosmic expansion relies on evidence such as the observation that type Ia supernovae (SN Ia) appear fainter than it would be expected in a decelerating universe. The late-time acceleration of the expansion of our Universe is especially relevant with regard to Chapter 3, in which we explore deviations from the standard model in this context. For the purpose of this introduction, we pick up where we left with the continuity equation in Eq. 1.8 and the observation that solutions for one of its components requires an equation of state describing the relationship between energy density and pressure, as well as the rewriting of the Friedmann equation in Eq. 1.9 and Eq. 1.10.

As described in Section 1.1.1, an equation of state is useful when talking about dark energy to relate energy density to pressure, and the resulting equation of state parameter is commonly denoted as w in the form

$$w \equiv \frac{p}{\rho}, \text{ with } \begin{cases} w = 0 & \text{for matter,} \\ w = \frac{1}{3} & \text{for radiation, and} \\ w = -1 & \text{for a cosmological constant.} \end{cases} . \tag{1.39}$$

Depending on the cosmological model used, w can have different values, including redshift-dependent ones that change as a function of time. The postulate of a cosmological constant corresponding to $w = -1$ has been consistently supported by observational evidence (see, for example, Riess et al., 2007; Wood-Vasey et al., 2007; Amanullah et al., 2010; Komatsu et al., 2011; Sullivan et al., 2011; Suzuki et al., 2012; Anderson et al., 2012, and references therein). This constant value is commonly interpreted as a form of vacuum energy in the context of said equation of state of dark energy, the nature of which has garnered the interest of cosmologists for the last two decades (Riess et al., 1998; Frieman et al., 2008; O'Raifeartaigh et al., 2018). The general notion of a cosmological constant itself predates the discovery of the accelerating expansion of the Universe (e.g., (Einstein, 1917; Friedmann, 1922; Lemaître, 1927)). The concept of dark energy, however, is much broader and has long served as a generic placeholder for the physical cause of an accelerating expansion, which is not necessarily restricted to a constant w (see, for example, Frieman et al., 2008, for a review). We can retrieve

a Friedmann equation corresponding to a late-time universe with flat geometry,

$$H^2 = \left(\frac{\Omega_m}{a^3} + \Omega_w a^{-3(1+w)} \right), \tag{1.40}$$

by considering how the energy density relates to the dimensionless scale factor, a, that we covered in Section 1.1.1,

$$\rho(a) = \rho_0 a^{-3(1+w)}. \tag{1.41}$$

For universes dominated by dark energy, as is the case in the concordance model used in cosmology, values of $w < -1/3$ will feature an accelerated expansion of a given universe, while cosmological models generally operate in the range of $-1 < w \leq 0$, and a smaller value for w will result in a faster acceleration. Typical attempts to probe deviations from the ΛCDM model assume modifications at the background level, which can be described as a relativistic fluid with an effective time-dependent equation of state. The form of the variable equation of state depends on the theory involved, subject to underlying kinetic and potential terms, which can result in considerable variations of w as a function of z. This also leads to proposals like the Chevallier-Polarski-Linder (CPL) parameterization (Chevallier & Polarski, 2001; Linder, 2003).

Examples of other non-constant models of dark energy include quintessence and, more generally, scalar-tensor theories, which we already mentioned above (Gannouji et al., 2006; Copeland et al., 2006). Theories relying on non-constant parameterizations of w have been tested on real datasets, with no evidence of statistically significant deviations from ΛCDM being reported (Garnavich et al., 1998; Hannestad & Mortsell, 2004; Chávez et al., 2016; Tripathi et al., 2017). The same inability to rule out competing theories of dark energy is reported when using SN Ia data under a specialized hypothesis test for ranges of w, though future survey data could provide stronger constraints (Genovese et al., 2009). This competition between a constant and a variable, often redshift-dependent, equation of state is a matter of continuing debate (Huterer & Shafer, 2018). A recent example of efforts in testing the CPL parameterization is carried out using the Pan-STARRS[3] Medium Deep Survey SN Ia data in combination with CMB measurements (Jones et al., 2018; Jones et al., 2019).

[3] `https://panstarrs.stsci.edu/`

Apart from common parameterizations of $w(z)$ as seen in Jassal et al. (2005) and De Felice et al. (2012), non-parametric approaches make use of linear or cubic spline interpolation as well as Gaussian processes (GPs) (Zhao et al., 2008; Serra et al., 2009; Vázquez et al., 2012; Hee et al., 2017). The latter replace the need for placing a limited number of nodes for an interpolation with the choice of a suitable covariance function $K(z, z')$ (Holsclaw et al., 2010a,b). Related research also makes use of non-parametric Bayesian methods based on correlated priors (Crittenden et al., 2012). Regardless of the preferred representation for the equation of state, the standard analysis consists of including the chosen $w(z)$ model in the supernova likelihood and evaluating the results with the ΛCDM model as the null hypothesis. In this scenario, the goal is to determine which type of behavior is allowed by the data in the context of a given dark energy model, with the prevailing conclusion that currently allowed behaviors are indistinguishable from the ΛCDM model (Abbott et al., 2019b).

In light of these results, we address the contrapositive question in Chapter 3 and investigate the robustness of a standard SN Ia analysis pipeline to deviations from the standard model of cosmology in the data used for such analyses. For this purpose, we introduce a novel random curve generator that can be subjected to customizable constraints to create redshift-dependent deviations from a cosmological constant as mock observations that purposefully deviate from the ΛCDM model. In doing so, we stress-test currently used methods to differentiate between dark energy equations of state based on SN Ia data.

1.1.5 Large-scale structure and cosmic voids

As we have mentioned before, the cosmological principle holds only on large-enough scales, and the FLRW metric does not seem to apply to the local Universe. *Large-scale structure* (LSS) describes, as the name suggests, the overarching structure observed in our Universe, which is also known as the *cosmic web* due to its filamentary nature. This latticework structure of enormous proportions represents one of the largest physical patterns in the Universe (Bond et al., 1996). Its structural composition is a byproduct of the hierarchical growth of large-scale structure and gives rise to four main classes of substructures: Galaxy groups, clusters, and superclusters, as well as filaments, sheets, and the large regions of near-emptiness known as *cosmic voids* (Zeldovich et al., 1982).

We can define a dimensionless matter density parameter that is defined in terms

of comoving coordinates, $\delta(\mathbf{r})$, and describes the density field perturbations as

$$\delta(\mathbf{r}, t) \equiv \frac{\rho(\mathbf{r}, t) - \bar{\rho}(t)}{\bar{\rho}(t)}, \tag{1.42}$$

with $\mathbf{r}$ as the 3D comoving scale size and $\bar{\rho(t)}$ as the homogeneous mass density operating as a function of time. Due to the random distribution describing the spatial dependence of perturbations at a given time, statistical investigations of the distribution of galaxies in our Universe are of great interest for the study of cosmology. Theories provide us with expectations about galaxy clustering and their size distribution, for which scales are commonly denoted as $\mathbf{k}$, which relate to the Fourier transform of comoving coordinates $\mathbf{r}$ (meaning the comoving wave number), are used. We can then relate Fourier and real space, with V as a given volume over which the transform is applied, as

$$\delta_{\mathbf{k}} = \frac{1}{V} \int \delta(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} d^3r, \text{ with } \delta(\mathbf{r}) = \sum_{\mathbf{k}} \delta_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{1.43}$$

As mentioned in Section 1.1.2, we are often more interested in statistical properties than spatial coordinates when investigating LSS, enabling us to average over all Fourier nodes (δ_k in Eq. 1.43) of a certain amplitude and establishing the matter power spectrum as

$$P(\mathbf{k}) \equiv |\delta_{\mathbf{k}}|^2, \text{ and } \langle \delta_{\mathbf{k}} \delta^*_{\mathbf{k}'} \rangle = \delta^{\mathrm{K}}_{\mathbf{k}\mathbf{k}'} P(|\mathbf{k}|) \tag{1.44}$$

for the Kronecker delta ($\delta^{\mathrm{K}}_{\mathbf{k}\mathbf{k}'}$) if we assume the cosmological principle to hold, meaning that isotropy and homogeneity are satisfied. For sufficiently small sub-horizon density perturbations ($\delta \ll 1$), that are subject only to gravitational interactions, the linear evolution can, with $\bar{\rho}_m$ as the homogeneous matter density, be written as

$$\ddot{\delta} + 2H\dot{\delta} = 4\pi G \bar{\rho}_m \delta = \frac{3}{2} H^2 \Omega_m(t) \delta, \tag{1.45}$$

with Peacock (1999) offering a more detailed coverage of this topic. Due to the dependence on time, we can also apply this to Fourier space for $\delta \rightarrow \delta_k$. The

acceleration, in this case, can be viewed as the result of gravitational interactions stemming from perturbations in the Universe, as a Newton's law of universal gravitation for density perturbations. In the left-hand and right-hand parts of Eq. 1.45, $2H$ refers to the result of using comoving coordinates without inertia, which is commonly referred to as 'Hubble drag'. This is, of course, a simplified view that excludes relevant perturbations outside of the energy density of the modeled universe in question, and multiple perturbations with corresponding equations of state (denoted as δ_i and w_i below) complicate matters to

$$\begin{aligned} \ddot{\delta}_i + 2H\dot{\delta}_i &= 4\pi G(1+w_i)\sum_j (1+3w_j)\bar{\rho}_j\delta_j \\ &= \frac{3}{2}H^2(1+w_i)\sum_j (1+3w_j)\Omega_j(a)\delta_j. \end{aligned} \tag{1.46}$$

Here, we can see that a dark matter equation of state equating to minus one, as defined in Eq 1.39, bars dark energy perturbations from growing due to setting the right-hand term to zero. When considering a full set of species (meaning baryons, dark matter, radiation, and neutrinos), we are faced with a suite of coupled second-order differential equations (Ma & Bertschinger, 1995). These are tackled with specialized software such as `CAMB` and `CLASS` (Lewis et al., 2000; Blas et al., 2011). Here, we should also note that at early times and for large scales, the evolution stays linear, but at late times and for small scales, the latter becomes non-linear, and density fluctuations eventually grow to no longer be linear. Calculations of the non-linear power spectrum rely on fitting functions based on simulations, with `halofit` being a prominent example (Smith et al., 2003). Figure 1.3 shows such a computation of (pseudo) non-linear matter power spectra for extensions of the standard model.

These approaches are, of course, refined on a constant basis, for example by Takahashi et al. (2012) for `CAMB`, and modern emulation approaches based on Gaussian processes that use the natural abbreviation of 'emulator' have emerged with software such as `CosmicEmu` by Lawrence et al. (2010) and `FrankenEmu` by Heitmann et al. (2014). Some opt for different routes, for example `PkANN` via artificial neural networks, while other emulators aim to improve on the more traditional approach (see, for example, Giblin et al., 2019; Winther et al., 2019).

Cosmic voids are characterized by underdensities in the dark matter distribution, presenting much simpler dynamics than their non-linear and high-density

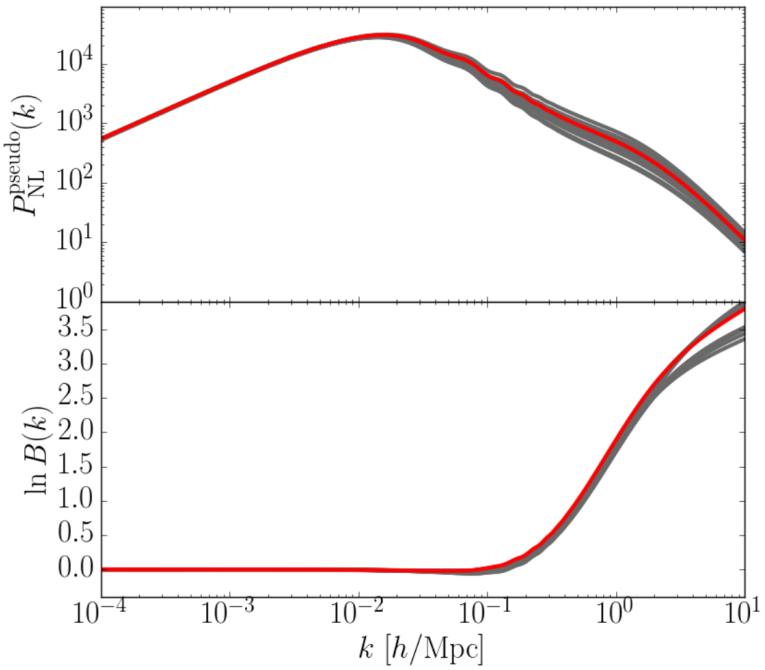

Figure 1.3 *Example of non-linear matter power spectra (Giblin et al., 2019). The upper panel shows (pseudo) non-linear matter power spectra at $z = 0$ computed with* `halofit`*. The lower panel shows the natural logarithm of the boost factor defined by the ratio of non-linear and linear power spectra. In both panels the base ΛCDM cosmology is drawn in red.*

counterparts (Hamaus et al., 2016). They are valuable cosmological probes, since they can encode relevant physical information and suffer from fewer sources of systematical error (Peebles, 2001; Lavaux & Wandelt, 2012). Their population statistics may be predicted from a given cosmological theory (see Fry, 1986; White et al., 1987; Li, 2011), which can provide observational constraints to current models (Hoyle & Vogeley, 2004; Gruen et al., 2018). They are especially useful for testing alternative cosmological models and probing screening mechanisms that are predicted to have a significantly reduced influence in such low-density regions, which we cover in more detail in Section 4.3.2. However, in spite of their usefulness, the detection of voids is no trivial task. The challenge stems from the lack of a dominant definition thereof, and their detection remains a focal topic of interest in cosmology (Cai et al., 2015; Gruen et al., 2016; Sánchez et al., 2017; Nadathur et al., 2017; Adermann et al., 2018; Brouwer et al., 2018; Xu et al., 2019; Davies et al., 2021).

Several established approaches to void detection employ Voronoi tessellation as described by El-Ad & Piran (1997) and Gaite (2005). Treated as particles, galaxies can, for example, be enclosed in distance-based Voronoi cells and grouped into

larger zones, where a watershed simulation naturally leads to the identification of large basins (low-density regions) corresponding to voids (Neyrinck, 2008). Markedly different approaches to void identification have also been proposed. For instance, Aragón-Calvo et al. (2007) use computer vision techniques to classify void morphology, applying scale-independent morphology filters to identify primary cosmological structures such as walls, filaments, and voids.

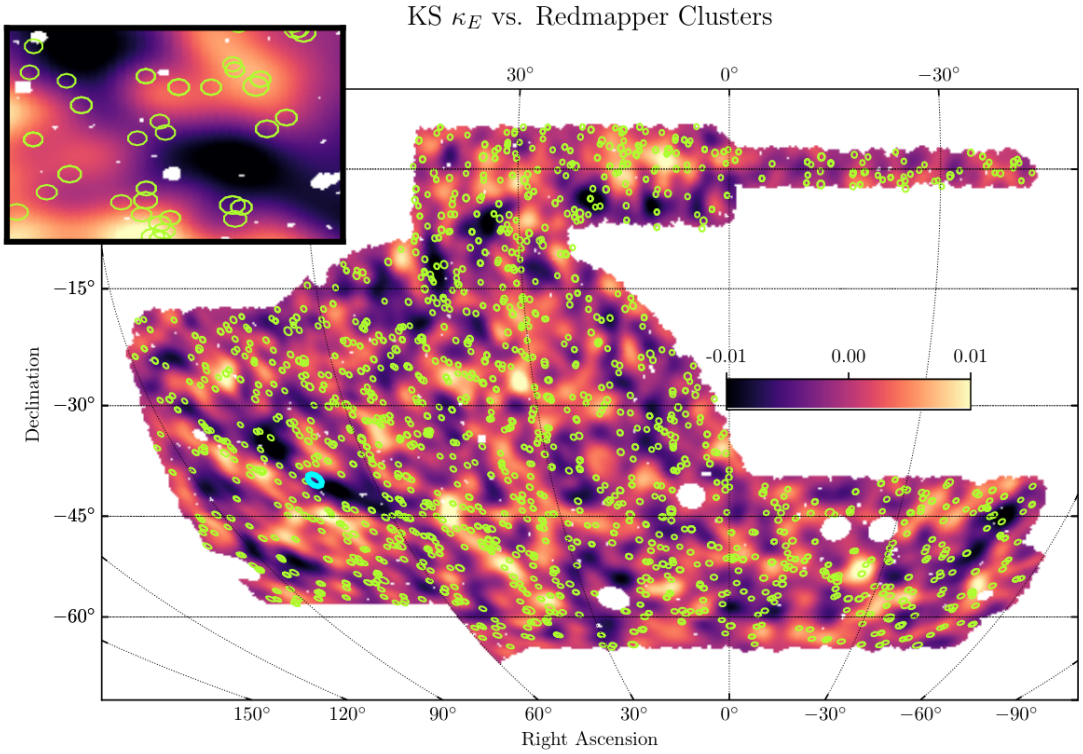

Figure 1.4 *DES Year 3 weak lensing mass maps based on galaxies in the underlying publication's third redshift bin (Jeffrey et al., 2021). These were obtained using the Kaiser-Squires method (KS, see Kaiser et al., 1995, for details), with clusters in the range of* $0.3 < z < 0.5$ *as identified via* `redMaPPer` *by Rykoff et al. (2014) superimposed as green circles. The location of the small inset on the upper left in the wide-field map is indicated with a cyan marker.*

Alternatively, a different family of techniques exploits notions of data topology for void identification. Aragón-Calvo et al. (2010), for example, apply topological segmentation, while Xu et al. (2019) apply notions of topological data analysis to find dimensional holes via persistent homology, in which zero, one, and two-dimensional holes are identified as clusters, loops of filaments, and voids, respectively. In the latter case, voids are considered statistically significant if their structure persists as long as data neighborhoods increase in size. Classifications yielded by these techniques exhibit differences that impact the science case for

which each method was developed (Libeskind et al., 2018). Figure 1.4 shows weak lensing mass maps based on DES Year 3 data as an illustration of the filamentary structure and empty regions on the sky.

Despite the difficulties posed by automated void detection, the rich variety of current methods and techniques have led to important advances in cosmology. As an example, the accurate location and modeling of voids can be exploited to derive clustering and abundance statistics such as void mass, and to constrain dark energy (Pisani et al., 2015). The dynamics of matter flowing away from the center of a void are instrumental to gain more insight on cosmological parameters, as described by Dekel & Rees (1994), while other applications include probing alternative dark matter models and tests of general relativity (Yang et al., 2015; Barreira et al., 2015b).

Regarding the latter, Li (2011) analyze void statistics in the context of scalar field theories in terms of their sizes, demonstrating through simulations that the fifth force produced by scalar field coupling increases the fraction of large underdense regions and leads to sharper transitions between voids and filaments. They observe that this information can be used not only to establish constraints under ΛCDM, but also to distinguish between different coupled scalar field models. Cai et al. (2015) show that $f(R)$ gravity results in weaker gravitational lensing of voids due to their lower dark matter content, inducing differentiation from general relativity via the lensing tangential shear signal around voids. Similarly, void statistics such as abundances, ellipticities, radial density profiles, and radial velocity profiles have been used to study voids in simulations, leading to an increase in average void size and the elimination of small voids, as well as emptier voids (Zivick et al., 2015). The sharper transitions may be detectable in the projected 2D ridges that Chapter 4 investigates.

Another difficulty of void detection is the need for accurate redshift measurements. Reliable redshifts demand larger, and ideally complete, galaxy spectroscopic surveys, which tend to cover small areas of the sky. Photometric galaxy surveys such as the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS; Chambers et al., 2016), the Dark Energy Survey (DES; Flaugher et al., 2015), and the upcoming Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST; Ivezic et al., 2008), on the other hand, can provide observations covering larger areas. Beyond these studies of voids, or the more common study of the galaxy clusters that occupy the nodes of the cosmic web, the characterization of connecting filaments can also be of great interest.

The filament between clusters Abell 0399 and 0401, for example, has been shown to host quiescent galaxies and hot gas, and to emit in radio (Bonjean et al., 2018; Govoni et al., 2019). Automated detection of filaments in a large volume has recently been carried out by Malavasi et al. (2020), using galaxy samples from the Sloan Digital Sky Survey (SDSS; York et al., 2000), while Galárraga-Espinosa et al. (2020) apply a similar approach to hydrodynamical simulations to derive the expected statistical properties of filaments.

In these studies, the distributions of galaxies from spectroscopic surveys (or simulations thereof) are used to detect filaments. Alternatively, weak gravitational lensing can be used as the observable to extract filaments (Mead et al., 2010; Maturi & Merten, 2013). While a few individual detections have been reported between specific clusters, as described by Dietrich et al. (2012) and Jauzac et al. (2012), the very low signal-to-noise ratio of the filament signal typically requires the stacking of large numbers of pairs of clusters to make detection from weak lensing possible (see, for example, Xia et al., 2020, and references therein). The very low amplitude of the lensing signal of filaments poses a significant challenge to detection efforts from wide photometric surveys and using lensing observables. Even if this were not the case, any such attempt to characterize many filaments over a large 3D volume would suffer from the same dependency on redshift accuracy as that discussed in the case of studies focussed on voids.

A way to avoid relying on accurate redshift measurements, while still readily taking advantage of photometric surveys, is to consider the 2D-projected counterparts to the elements that make up large-scale structure instead (though see also Sánchez et al., 2017, for a void finder built to work on photometric surveys, using redshift slices instead of a full 2D projection). In the case of voids, such projections are known as *troughs*, which represent the most underdense regions in the plane of the sky. Since troughs comprise regions of lower density across the line of sight in the projected space only, it is sufficient to use photometric measurements, obviating spectroscopic redshifts (Clampitt et al., 2013; Gruen et al., 2016). In related research, Brouwer et al. (2018) make use of the photometric Kilo-Degree Survey (KiDS; De Jong, Jelte T. A. et al., 2017) and the spectroscopic Galaxy And Mass Assembly Survey (GAMA; Driver et al., 2011) to identify troughs from foreground galaxies; their simulations ultimately forecast that upcoming surveys such as LSST and Euclid will enable us to constrain the redshift evolution of cosmic troughs. This finding is further supported by research showing 2D-projected underdensities, meaning troughs, to be a powerful probe

to distinguish between ΛCDM and modified gravity models in future lensing surveys, potentially even more so than 3D cosmic voids (Higuchi & Shirasaki, 2016; Barreira et al., 2017; Cautun et al., 2018).

1.1.6 Types and use of cosmological simulations

Despite massive efforts in investigating both the Milky Way as our home galaxy and the myriad of others we know about so far, the physical processes involved in their formation and evolution remain an active topic of research, and we still lack a lot of the pieces of that puzzle. While some of these uncertainties pertain to the nature of dark matter, the dynamics of baryonic components still leave a lot of questions unanswered.

Generally, cosmological simulations can be split into three distinct approaches, which will be outlined in the following parts. As the simplest method for simulating the Universe, N-body simulations, which are also known as 'pure-gravity' or 'dark matter-only' simulations, employ dynamical systems of particles to calculate gravitational forces acting on them. This can be done either directly via numerical integration, or with the inclusion of general-relativistic effects to establish a scale factor a necessary for modeling the expansion of the Universe (Efstathiou et al., 1985). The particles in these simulations are only subject to gravitational interaction and do not represent physical objects; instead, they are a discretization of dark matter mass.

Simulations of this kind played an essential role in establishing the ΛCDM model as the 'standard model' of cosmology. Influential N-body simulations include the Millennium Simulation by Springel et al. (2005) and its Millennium-II successor as described by Boylan-Kolchin et al. (2009a), the Bolshoi simulation by Klypin et al. (2011), the subsequent MultiDark simulation by Riebe et al. (2013), the MICE Grand Challenge Lightcone Simulation as described by Fosalba et al. (2015b,a) and Crocce et al. (2015), and the EUCLID Flagship Simulation using PKDGRAV3 (Hopkins, 2014; Potter et al., 2017).

Since the baryonic content of the Universe can be described by treating gas in a continuous manner as an ideal fluid, hydrodynamic cosmological simulations using particle-based methods with discrete masses and grid-based methods with discrete spaces have been developed (Dolag et al., 2008). Influential hydrodynamic simulations include MareNostrum Universe as described by Hoeft et al. (2008),

the Illustris Simulation and its successor IllustrisTNG introduced by Genel et al. (2014) and Pillepich et al. (2018), respectively, MassiveBlack-II by Khandai et al. (2015), EAGLE by Schaye et al. (2015), BlueTides as described by Feng et al. (2016), and Horizon-AGN by Dubois et al. (2016), as well as MUFASA and, more recently, its successor simulation SIMBA, with the latter being used in this work (Davé et al., 2016, 2019).

These kinds of computational models are regularly used to further our understanding of galaxy formation and evolution. Cosmological simulations are an invaluable part of theoretical research in cosmology, allowing for the implementation of new ideas and their testing against existing observational data, as well as the detailed study of cosmological phenomena with large amounts of simulated data. The realistic modeling of processes when compared to observational data is a primary concern, as is the trade-off between simulation size and resolution (Dolag et al., 2008).

While constraints on cosmological parameters are mostly directly tied to the overall matter distribution, observations can only directly probe the luminous baryonic component. Modeling the latter entails complex additional physical processes beyond gravity that result in much higher computational costs, which precludes parameter space explorations within $\sim$Gpc^3 volumes as needed for cosmological applications. It is, therefore, important to develop accurate frameworks to tie observable galaxy properties to the dark matter halo distribution.

Several approaches to solving this issue exist in the literature. One is based on abundance matching, in which the baryonic properties are tied to the stellar mass, which, in turn, is tied to the halo mass by assuming that rank ordering in mass is preserved. Here, satellites are extracted directly from the simulation, and rank ordering is implemented for all galaxies instead of modeling central and satellite galaxy relations separately. Conversely, halo occupancy distribution (HOD) modeling treats halo mass functions, meaning the distribution of the prevalence of binned dark matter halo masses, as an input. It assumes a satellite distribution describing the distribution of satellite galaxies around the central galaxy, and models the latter to match clustering constraints (Berlind & Weinberg, 2002). The assumption of rank ordering, however, is not true in detail, and since there is no underlying physical model, it is not obvious that the often locally-calibrated relations apply at all the redshifts considered. That being said, with appropriate choices, it is possible to populate galaxies into dark matter haloes roughly in accord with observations.

Semi-analytic models (SAMs) are another approach (see, e.g., White & Frenk, 1991; Kauffmann et al., 1993; Cole et al., 1994), which provides a full physical framework with increased computational cost, albeit still far cheaper than full hydrodynamic models. With appropriate parameter tuning, these can be calibrated to local relationships, and they are built on models of baryonic physics to relate the hierarchical growth of dark matter haloes to galaxy population properties, as summarized by Mitchell et al. (2018), with an introduction to the field provided by Baugh (2006). Such models do, however, typically have a large number of free parameters that are difficult to constrain simultaneously, and so either tune parameters by hand, or else constrain only a subset of parameters to observations via an MCMC approach. As such, it is difficult to formally constrain the uncertainties in the physical parameters, as required for precision cosmology.

As SAMs combine (often simplified) physically motivated prescriptions with estimates of dark matter halo distributions and merger trees to calculate physical galaxy properties, and since the calculation of baryonic components with hydrodynamic simulations is computationally costly, large-volume investigations benefit from the SAM approach. An interface to N-body simulations exists through the use of dark matter halo merger trees extracted from such simulations as SAM inputs, as direct simulations are better suited for capturing non-linear structure formation than analytic methods like the Press-Schechter formalism (Press & Schechter, 1974; Knebe et al., 2015).

Influential models include, while not an exhaustive list, work by Somerville & Primack (1999) and Somerville et al. (2008), GALFORM as described by Cole et al. (2000) and later recalibrated by Baugh et al. (2018), research by Monaco (2004) and Kang et al. (2005), GalICS and GalICS 2.0 by Hatton et al. (2003) and Cattaneo et al. (2017), respectively, the Munich galaxy formation model by Henriques et al. (2015), the GAEA model by Hirschmann et al. (2016), and SAGE by Croton et al. (2016), as well as a broad review of galaxy formation theory by Benson (2010). Notably, Neistein et al. (2012) present a method for turning hydrodynamic simulations into SAMs by transforming efficiencies in physical processes of galaxies into functions of z and M_{halo}. While larger deviations for instantaneous properties like star formation rates are reported, this success increases the feasibility of machine learning methods that are trained on hydrodynamic simulations.

1.2 Inference methods and simulations

1.2.1 A short primer on Bayesian analysis

Bayesian methods are now a standard approach to data analysis and inference in astrophysics. In this approach, probabilities are regarded as a means of quantifying information, and in particular the information contained in an experimental dataset about a specific model. The field of Bayesian analysis is vast and occupied by a multitude of application areas as well as different philosophical schools of thought, so any coverage that only forms part of an introduction to a specialized body of work is, of course, limited to a basic overview relevant to further parts.

In terms of its foundations, there are two dominant schools of thoughts on probability, especially in the application of statistics in science and including cosmology, to which we shall limit ourselves in this context. The *Bayesian interpretation of probability* treats the latter as a degree of belief given specified evidence, which can be translated to a reasonable expectation provided an available dataset. In contrast, the *frequentist interpretation of probability* treats a probability as the true underlying frequency relative to other possible outcomes in the limit of an infinite number of repetitions of the same experiment.

While both interpretations can be viewed as somewhat at odds, especially from a philosophical point of view, methods stemming from both are frequently used in science and have their place in the statistical analysis of data. Frequentist approaches to testing hypotheses are prevalent due to additional requirements of Bayesian hypothesis testing such as prior probabilities for approaches like the Bayes factor. Bayesian statistics, on the other hand, experienced a sharp rise in usage in cosmology during the past two decades due to its ability to work with singular phenomena and the fact that, at the largest scale, our sample is $N = 1$ universe, something that the frequentist foundations formally struggle with (Trotta, 2008).

The core of Bayesian statistics is formed by *Bayes' theorem*, named, like the underlying interpretation, after Thomas Bayes and the posthumous publication of his results (Bayes, 1763). In addition, Laplace (1812) independently developed and further refined the Bayesian approach, with much of the early Bayesian interpretation of probability being thanks to his efforts. Given an event A, the

probability of said event occurring is denoted as $P(A)$. The *conditional probability* $P(A|B)$ expresses the probability of event A given that event B takes place. Similarly, the probability of both events taking place is encoded in the *joint probability* $P(A \cap B)$. The latter can be split into a conditional probability and the probability of the conditional event, meaning

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A). \tag{1.47}$$

Bayes' theorem, through the lens of Bayesian statistics, formulates the *posterior probability* $P(H|D)$ of a hypothesis H given available data D. This calculation relies on the *likelihood* $P(D|H)$, meaning the probability of obtaining the given data assuming that H is true, as well as the *prior probability* (often just called the 'prior') $P(H)$ of the hypothesis and the probability of the *evidence* $P(D)$. It can, in this case, be written as

$$P(H|D) = \frac{P(H \cap D)}{P(D)} = \frac{P(D|H)P(H)}{P(D)} \quad \text{s.t. } P(D) \neq 0, \tag{1.48}$$

with P(D) calculated via the *law of total probability*, which states said calculation through a partition of the sample space into possible explanations. This means that, for a set of N explanations $\{H_1, H_2, \ldots, H_N\}$, the evidence probability is

$$P(D) = \sum_{i=1}^{N} P(D \cap H_i) = \sum_{i=1}^{N} P(D|H_i)P(H_i). \tag{1.49}$$

For textbooks providing an introduction and overview of Bayesian methods, we refer interested readers to Bernardo & Smith (1994), MacKay (2003) , and Gelman et al. (2013), as well as Murphy (2012) and Hobson et al. (2009) for an overview centered on machine learning and cosmology, respectively.

In most realistic cases, the analytic or direct numerical evaluation of posterior probability distributions is impossible or infeasible, especially in cases that feature many parameters, due to the large volume of high-dimensional spaces. The widespread use of Bayesian methods has largely been driven by the availability of *sampling* algorithms, which can generate samples from a posterior distribution without exploring the full space. These samples can then be used to generate

summary statistics like means and limits on individual parameters, or correlations between them. For a shorter overview of the application of Bayesian inference and especially sampling in cosmology, see Trotta (2008).

1.2.2 Parameter estimation and sampling methods

Parameter estimation is a valuable and widely-used tool in cosmology and other sciences, as we are often interested in the values of certain variables. The most obvious example in our case is cosmological parameter estimation, where we wish to infer certain fundamental values of our model of the Universe based on empirical data. Before this millennium, these parameters were commonly approached with maximum likelihood estimation (MLE), which attempts to optimize the likelihood function. One prevalent shortcoming of this basic approach is that the parameter(s) of interest are treated as a point estimate, as opposed to a random variable as is the case in Bayesian approaches. There are options to estimate the variances in this case, but they generally require assumptions about the likelihood.

As a result, we lack further information about the parameter distribution, such as the variance, for example to compare the fits between different estimates from various surveys. From a frequentist point of view, the general case encompassing MLE are extremum estimators for parametric models, in which case the likelihood serves as the objective function. This takes the form of

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \, \hat{\mathcal{L}}_n(\theta; \mathbf{y}), \tag{1.50}$$

for a parameter vector θ, its associated parameter space Θ, and a likelihood function $\hat{\mathcal{L}}_n$ with observed data $\mathbf{y} = y_1, y_2, \ldots, y_n$. Analytic solutions are a possibility in some cases, for example for CMB calibration and beam uncertainties in power spectrum measurements as described in Bridle et al. (2002), as well as for galaxy clustering and dark energy parameters from 3D cosmic shear (Taylor & Kitching, 2010). These approaches do, however, assume that the distribution of observations can be accurately described by Gaussians. If not solvable analytically, this is usually done with methods such a *gradient ascent*, which

updates parameter values to slowly approach to a good-fit value as

$$\theta_i^{\text{new}} = \theta_i^{\text{old}} + \eta \cdot \frac{\partial \mathcal{L}\mathcal{L}\theta^{\text{old}}}{\partial \theta_i^{\text{old}}}, \tag{1.51}$$

with a step size η for updates. We can, of course, see how this leads to issues in multimodal parameter distributions, as only one peak is found. In addition, the step size requires fine-tuning or adaptive approaches to not get stuck in 'valleys' within the distribution. In a demonstration of the intimate relationship between statistics and machine learning, which we will discuss further below in Section 1.2.5, gradient ascent and related methods are commonly used in machine learning, especially in artificial neural network architectures, to update parameters of artificial neurons.

Within the cosmology literature, Christensen et al. (2001) proposed initial arguments for the use of Bayesian methods for the purpose of cosmological parameter estimation. They argued in favor of MCMC approaches due to their superiority in terms of sampling from, and converging to, the true posterior distribution in the limit of an infinite sample size. The application of MCMC approaches in these early efforts was centered on the *Metropolis-Hastings algorithm*, which was named after work done by Metropolis et al. (1953) and, for the more general case, Hastings (1970). As the result of statistical research efforts in the realm of physics, the algorithm is also a prime example of the close connection between the two fields, with statistics-driven work not being a new phenomena of the age of data science.

In a *Markov chain*, the distribution of subsequent steps x_{t+1} at a time t depends only on the prior position x_t. The distinguishing feature of the algorithm is the acceptance of new points in the Markov chain if the acceptance ratio of the proposed point and the last point is larger than one, and the probabilistic acceptance of points with a lower ratio if the latter is larger than a random number $n \in [0, 1]$. The Metropolis-Hasting algorithm requires a *transition kernel* K that satisfies $\pi(x)k(x, \hat{x}) = \pi(\hat{x})k(\hat{x}, x)$ for a target distribution π and is thus called 'π-reversible', meaning that the probability of transferring from x to $\hat{x}$ is the same as transferring from $\hat{x}$ to x. After initializing x_0, for example at a random point in the parameter space, a new proposal location $\hat{x} \sim k(\hat{k}|x_t)$ is then sampled and

accepted with the probability

$$x_{t+1} = \begin{cases} \hat{x}, & \text{if } \min\left(1, \frac{k(x_t|\hat{x})\pi(\hat{x})}{k(\hat{x}|x_t)\pi(x_t)}\right) > u \sim \mathcal{U}(u; 0, 1) \\ x_t, & \text{else} \end{cases} . \tag{1.52}$$

In practice, the symmetric special case originally devised by Metropolis et al. (1953) is often used, for example by using a Gaussian distribution with a specified variance centered around the current location as the distribution to sample proposals from. This acceptance of less likely points dependent on the likelihood leads to the sampling from the posterior distribution and, notably, does not require marginalization via the evidence. While many commonly used proposal distributions are symmetric, choosing a suitable proposal can still be difficult. Even for the simple case of using a Gaussian distribution, the variance of the latter has to be selected carefully, as too small a variance leads to a slow exploration of the parameter space, while too large a variance can overlook high-density areas of interest. This challenge is similar to step sizes in gradient ascent methods, which have to deal with the risk of a small step size getting stuck in valleys, while large step sizes can lead to jumping over target peaks. Attempts to dynamically tune the proposal parameters on the fly exist, for example in the family of adaptive MCMC algorithms (Roberts & Rosenthal, 2009).

Knox et al. (2001) then followed the proposal of Christensen et al. (2001) to constrain the age of the universe to $t_0 = 14.0 \pm 0.5$ Gyr. Earlier work includes Saha & Williams (1994), who made use of the Metropolis-Hastings algorithm for galaxy kinematics, Jaffe (1996), who was among the pioneers of Bayesian model selection in cosmology, and Christensen & Meyer (1998), who employed the related Gibbs sampler for gravitational wave analysis (Geman & Geman, 1984). For more in-depth information covering the wide array of contributions from both the astrophysical and statistical literature, we recommend Trotta (2008) as a more complete overview of the development of Bayesian inference in cosmology in particular, and Robert & Casella (2011) and Brooks et al. (2011) for a history of MCMC methods and their development in general.

Up to, and into, the new millennium, the Metropolis-Hastings algorithm remained the standard approach to cosmological parameter estimation, which was further supported by the development of a dedicated implementation in `CosmoMC` (Lewis & Bridle, 2002). A variety of algorithms and codes are, however, available for

different types of problems. The optimal choice depends on multiple factors, including the dimensionality of the problem, meaning the number of parameters to estimate, the evaluation speed, the need for Bayesian evidences, the availability of analytic derivatives, the ability to sample from marginal distributions, and the possibility and degree of using parallelization. The latter is a problem commonly encountered in basic MCMC approaches from the perspective of computational costs, as the Markov property of new states being dependent on the last state makes them sequential by design. One can, of course, use multiple 'walkers' in the parameter space, but in the limit of a number of steps equaling one and as many walkers as samples are required, this approaches random sampling, which means that there is a trade-off between the number of walkers and the number of steps necessary to converge to the underlying distribution.

1.2.3 Recent developments in Bayesian sampling

It should be noted that the statistical literature on sampling methods is rich and vast, and a complete review of both their history and all current developments would exceed the scope of this thesis. The methods covered in more detail here are those likely to be more familiar to the astrophysical community, due to being widespread or featuring field-specific implementations. While we aim to cover relevant comparisons, this should, of course, not be misunderstood as a judgment about these methods being superior in the wider context of all statistical developments, but to place this work in the context of astrostatistics.

In more recent years, new MCMC sampling techniques were proposed and subsequently applied to cosmological parameter estimation. Examples include *Population Monte Carlo* (PMC) techniques introduced by Cappé et al. (2004) and Wraith et al. (2009), and used by Kilbinger et al. (2010) to develop `CosmoPMC`; *affine-invariant MCMC ensembles* by Goodman & Weare (2010), which led to the publication of `emcee` by Foreman-Mackey et al. (2013) and `CosmoHammer` by Akeret et al. (2013); and renewed interest in *Approximate Bayesian Computation* (ABC) for likelihood-free inference based on simulations to introduce `CosmoABC` and `abcpmc` (Ishida et al., 2015; Akeret et al., 2015).

Density estimation likelihood-free inference (DELFI) is another recently developed technique that trains a flexible density estimator to approximate the target posterior, circumventing the large number of simulations that traditional ABC approaches can require (Bonassi et al., 2011; Fan et al., 2013; Papamakarios

& Murray, 2016). Using the JLA sample of 740 type Ia supernovae as described in Betoule et al. (2014), Alsing et al. (2018) subsequently deploy this method to estimate cosmological parameters. Their approach, however, makes a few simplifying assumptions, for example normally distributed priors and likelihoods. Other advanced methods, like the *Hamiltonian Monte Carlo* approach developed by Duane et al. (1987), have also been applied, for example by Hajian (2007). These developments are driven by the computationally costly likelihood calculations involved in most MCMC algorithms, trying to alleviate this issue with a certain degree of parallelization due to the increased availability of cheap computing resources, faster convergence or, in the case of ABC, the circumvention of direct likelihood computations altogether.

As such methods either fail to reduce the runtime enough for modern problems or have their own pitfalls, for example through an increased risk of introducing biases, the quest for highly parallelized and fast alternatives for cosmological parameter estimation continues. This need is further exacerbated by upcoming missions like LSST and Euclid requiring high-dimensional posterior approximations with a large number of required nuisance parameters of no interest for a given application, but which have to be accounted for, predicted to vastly exceed previous missions (Amendola et al., 2018).

Lastly, *nested sampling* is a Bayesian take on numerical Lebesgue integration for model selection introduced by Skilling (2006). While targeting the calculation of Bayesian evidence, posterior samples are generated as a by-product, and the algorithm was quickly shown to require considerably fewer posterior evaluations (Mukherjee et al., 2006). Due to denser and sparser sampling from high-posterior and low-posterior regions, respectively, nested sampling provides increased efficiency when compared to previous MCMC methods. This has led to extensions and implementations for applications in cosmology, notably `CosmoNest` by Liddle et al. (2006), `MultiNest` as described in Hobson & Feroz (2008) and Feroz et al. (2009), and `PolyChord` (Handley et al., 2015). In cosmology, such implementations have been used in areas as diverse as cosmic ray propagation models, cosmoparticle physics, and gravitational wave astronomy (Trotta et al., 2011; Del Pozzo, 2012; Verde et al., 2013; Del Pozzo et al., 2017; Wang et al., 2018b). A comparison between nested sampling and state-of-the-art MCMC methods can be found in Allison & Dunkley (2014), while an investigation of statistical uncertainties in nested sampling is provided by Keeton (2011). Nested sampling has also been adopted by other fields of research, including GPU-

accelerated implementations, for example in systems biology (Aitken & Akman, 2013; Stumpf et al., 2014).

The statistical literature, however, points out various issues of nested sampling methods that have prevented wide-spread adoption in statistics. Among these are the assumption that perfect and independent samples from a constrained version of the prior are drawn in each iteration, the underestimation of sampling errors due to the simulated-weights method it employs, and an asymptotic approximation variance that scales linearly with the dimensionality of a given parameter space (Chopin & Robert, 2010; Higson et al., 2018). In recent years, the rise of machine learning in cosmology has also led to the inclusion of associated techniques into parameter estimation approaches, for example work by Alsing et al. (2019) on learning the likelihood function in the context of the DELFI approach described above, and to parameterize efficient MCMC proposals via deep learning (Moss, 2020). Similarly, the same applies to nested sampling, as `MultiNest` has been combined with artificial neural networks to approximate the likelihood function (Graff et al., 2012). Nested sampling has been combined with importance sampling to soften the hard likelihood constraint as described in Chopin & Robert (2010), which has been shown to counteract some of the sampling issues in `MultiNest` (Buchner, 2016). Further recent extensions include, but are not limited to, the introduction of embarrassing parallelism to nested sampling by Griffiths & Wales (2019) as well as variations in the number of live points, for example in dynamic nested sampling (Higson et al., 2019).

In this thesis, we use example likelihoods from the Dark Energy Survey (DES) collaboration's analysis of lensing and clustering data for Chapter 2, as presented in Abbott et al. (2018a). These calculations make use of the `CosmoSIS` and `CosmoLike` pipelines, which contain implementations of both `MultiNest` and `emcee` (Zuntz et al., 2015; Krause & Eifler, 2017).

For a comparison of approaches designed for the acceleration of MCMC methods in particular, including additional parallelization methods, see Robert et al. (2018), who cover methods targeting both the exploration stage of the algorithms and the exploitation level. The second approach includes Rao-Blackwellization and scalability, with the latter encompassing parallelization under this nomenclature. Other examples of methods trying to optimize the performance of established algorithms include the *no-U-turn sampler* (NUTS) by Hoffmann & Gelman (2014), which alleviates the need by the previously mentioned HMC algorithm for tuning by computing the trajectory length via recursively built

candidate proposals, as well as work by Neiswanger et al. (2014) on asymptotically exact and embarrassingly parallel MCMC sampling. The latter solves the slowing-down of parallel MCMC methods by reducing the amount of required communication in a divide-and-conquer tactic that splits up the dataset and which the authors justify with prohibitively long runtimes of many serial methods.

The need for sped-up posterior estimation approaches is further elaborated on by Bardenet et al. (2014) and Wilkinson (2005), with the latter pointing out the need for parallelized methods due to: "[...] weeks of CPU time on powerful computers" for serial MCMC methods on high-dimensional problems of interest. This need for parallelization approaches stems mostly from cases in which parts of the computations are very expensive, but which can be transformed into an, ideally, embarrassingly parallel problem that allows the respective steps to take full advantage of a greater number of cores, thus cutting otherwise infeasible runtimes to a fraction. For a more general overview of the history of Monte Carlo methods, such as multi-stage Gibbs samplers, see Robert & Casella (2004).

1.2.4 Variational inference and Dirichlet processes

Variational Bayesian methods were originally developed and explored in the context of artificial neural networks, and gained initial interest from research on inference in graphical models (Peterson & Anderson, 1987; Peterson & Hartman, 1989; Jordan et al., 1999). They provide a way of approximation for intractable integrals prevalent in both Bayesian inference and machine learning. The use of variational Bayesian methods for inference is commonly known as *variational inference* (VI) and provides a faster and more scalable alternative to Markov chain Monte Carlo (MCMC) methods in many contexts; the main difference between them is that VI treats parameter estimation not as a sampling problem, but instead as an optimization problem. From a research point of view, these methods also garnered the interest of the statistics community because they are currently not as well understood as MCMC methods and form the basis of much of Chapter 2 (Blei et al., 2017). VI can be used directly to analytically approximate posterior probabilities of unobserved variables, but its use for model selection by establishing a lower bound for marginal likelihoods is of special interest; a model's higher marginal likelihood indicates a better fit to given data, which shares similarities to the Bayes factor as a way to determine models with a higher probability of generating said data.

The *Kullback-Leibler divergence* D_{KL} is a central concept in VI and defines by how much a distribution diverges from another, or how similar it is. For a reference distribution $p(\mathbf{x})$ and a proposal distribution $q(\mathbf{x})$, the D_{KL} can be expressed as

$$D_{\mathrm{KL}}(p(\mathbf{x})||q(\mathbf{x})) = \int_{-\infty}^{\infty} p(\mathbf{x}) \log \frac{p(\mathbf{x})}{q(\mathbf{x})} \, d\mathbf{x}. \tag{1.53}$$

The fact that the D_{KL} is an asymmetric difference measure means that $D_{\mathrm{KL}}(p(\mathbf{x})||q(\mathbf{x})) \neq D_{\mathrm{KL}}(q(\mathbf{x})||p(\mathbf{x}))$, which is due to its calculation as a directional loss of information. The former formulation in the last sentence is called the 'forward' D_{KL}, while the latter formulation is known as the 'reverse' D_{KL} due to swapping reference and proposal distributions in the equation.

In VI, the D_{KL} is used to find a best-fitting distribution to a set of samples. Let $\mathscr{Q}$ be a selected family of distributions, $\mathbf{x}$ and $\mathbf{z}$ observations and parameters, respectively, and $p(\mathbf{z})$ a prior density that can be related to observations via the likelihood $p(\mathbf{x}|\mathbf{z})$ to calculate the posterior $p(\mathbf{z}|\mathbf{x})$. The family member $\hat{q}(\mathbf{z})$ that best matches the posterior can be found in the framework of an optimization problem, finding with some specified tolerance the value of

$$\hat{q}(\mathbf{z}) = \underset{q(\mathbf{z}) \in \mathscr{Q}}{\operatorname{argmin}} \, D_{\mathrm{KL}} \left(q(\mathbf{z}) || p(\mathbf{z}|\mathbf{x}) \right). \tag{1.54}$$

Calculating this quantity directly is often infeasible, since it is equivalent to measuring the Bayesian evidence. Instead, VI methods (equivalently) maximize an alternative quantity, the *evidence lower bound* (ELBO),

$$\begin{aligned} \mathrm{ELBO}(q) &= \mathbb{E}[\log \, p(\mathbf{z}, \mathbf{x})] - \mathbb{E}[\log \, q(\mathbf{z})] \\ &= \mathbb{E}[\log \, p(\mathbf{x}|\mathbf{z})] - D_{\mathrm{KL}} \left(q(\mathbf{z}) || p(\mathbf{z}) \right), \end{aligned} \tag{1.55}$$

which is numerically easier to calculate than the D_{KL}. Here, it should be noted that the above uses the reverse D_{KL}. The difference between the two formulations of the divergence measure is that the forward D_{KL} exhibits mean-seeking behavior, as the proposal distribution is forced to cover high-probability regions of the reference distribution without being penalized for covering low-probability regions of the reference distribution with high probabilities. In contrast, the reverse D_{KL} exhibits mode-seeking behavior, meaning that it forces the proposal distribution to concentrate samples on modes at the cost of omitting the pressure to place high probabilities on all modes of the reference distribution. This can be easily imagined with the simple example of trying to fit a single

Gaussian to a bimodal distribution; where the forward D_{KL} would place the a very broad distribution between the modes, covering both modes while placing the mean between them, the reverse D_{KL} would focus on on one of the modes given the limitation of using just one Gaussian. For this reason, determining the right amount of proposal distributions to use with VI when using the reverse D_{KL} is crucial to have enough distributions to cover all areas of interest appropriately.

The ELBO also delivers a lower bound for the evidence, which is the reason for the utility of VI for model selection, as covered in Blei et al. (2017). The ELBO serves as an alternative to using MLE on model parameters, as it functions as an approximation of the marginal likelihood. While the use of a bound for model selection lacked theoretical justifications despite practical applications, strong theoretical guarantees regarding the consistency for model selection have recently been provided for both mixture models and the general case (Chérief-Abdellatif & Alquier, 2018; Chérief-Abdellatif, 2019). A more extensive introduction to VI for the interested reader can be found in Murphy (2012). In the traditional approach of mixture models, the number of separate distributions used to model the posterior is either set manually as an input variable or has to be optimized in computationally costly ways. VI, on the other hand, can be used to determine the number of distributions directly from the available data, employing a *Dirichlet Process* (DP) as a prior on the number of parameters.

Developed by Ferguson (1973), DPs are distributions of distributions and serve as a measure of measures, featuring a base distribution G_0 and a scaling parameter $\alpha \in \mathbb{R}_+$, and with realizations denoted as $G \sim \mathrm{DP}(\alpha, G_0)$. Given said base distribution as well as a measurable set on which to apply a probability distribution, a measurable (and finite) partition of that set with n elements, which we can write as $\{B_i\}_{i=1}^n$, the requirement $(G(B_1), \ldots, G(B_n)) \sim \mathrm{Dir}(\alpha G_0(B_1), \ldots, \alpha G_0(B_n))$ holds. This area has important applications as the prior in infinite mixture models, and gained new traction in both statistics and machine learning in recent years (Gershman & Blei, 2012). The DP mixture model presented originally by Antoniak (1974) takes θ_i as the distribution parameter of observation i and uses the discrete nature of the base distribution G_0 to view the DP mixture as an infinite mixture model. For values s from such a DP mixture, the predictive density with available data $\{s_1, \ldots, s_N\}$ is

$$p(s|s_1, \ldots, s_N, \alpha, G_0) = \int p(s|\theta) p(\theta|s_1, \ldots, s_N, \alpha, G_0)\, d\theta. \tag{1.56}$$

Here, the predictive density is the distribution of unobserved data points given the available observations. As the computation of that density is, again, infeasible, Blei & Jordan (2006) introduce the use of VI for DP mixtures. Bayesian takes on mixture models employ a prior over the mixing distribution as well as over the cluster parameters, with the former commonly being a Dirichlet and the latter being a Gaussian distribution in our case. Given the discrete nature of random measures drawn from a DP, a mixture of the latter can be viewed as a mixture model with an unbounded number of components (Blei & Jordan, 2006).

The Bayesian nonparametrics approach employs the *stick-breaking process* by Sethuraman (1994), which exploits the discrete nature of DPs to calculate the probability mass function, and can be used for Bayesian Gaussian mixtures with an undetermined number of Gaussians. Let β_k be the random variables drawn from a beta distribution according to $\beta_k \sim \text{beta}(1, \alpha)$, $\{\theta_k\}_{k=1}^{\infty}$ samples from G_0, δ_{θ_k} a Dirac measure, and $\{\beta_k\}_{k=1}^{\infty}$ the corresponding probabilities. The probability mass function of the discrete random distribution is then

$$f(\theta) = \sum_{k=1}^{\infty} \beta_k \delta_{\theta_k}(\theta) \ , \ \text{with } \beta_k = \beta'_k \cdot \prod_{i=1}^{k-1}(1 - \beta'_i). \tag{1.57}$$

The name is based on the analogy of breaking a stick of unit length into infinite segments by consecutively breaking off pieces as per β'_k from the stick and assigning them to our stick of length β_k until the remainder is truncated to recover a finite-dimensional representation. The truncated variational distribution is then used to approximate the posterior of an infinite DP mixture. As a mathematical description of the subsequent application of VI to DPs with stick-breaking would go beyond the scope of this overview, we refer the reader to Blei & Jordan (2006). A less concise introduction to DPs and Bayesian nonparametrics in general, as well as its applications, is provided in Hjort et al. (2010).

As the posterior distribution, given a DP mixture prior, cannot be directly calculated, VI offers a deterministic approach to approximate them. In this chapter, we employ the mean-field family within VI to optimize the D_{KL}, using this approach to approximate the joint posterior for parameters of an infinite Gaussian mixture, made finite to a maximum number of components through stick-breaking.

1.2.5 Machine learning as part of cosmology

Around half a decade ago, in a paper on variational inference for generative models for astronomical images, Regier et al. (2015) described the field of astrophysics as receiving comparably little attention from, and being underserved by, the latest advances in machine learning. At the time, with emerging overviews of things happening in the field by works such as Ball & Brunner (2010), this sentiment rang true, although the latter niche has experienced an explosive growth in the short amount of time that has since passed (Fluke & Jacobs, 2020; Feigelson et al., 2021). Despite this, however, the mentioned statement refers to the collaboration with, and attention from, machine learning as a discipline, as opposed to astronomers applying and, at times, developing methodology from that area. Efforts to pool researchers from these fields together exist, though, for example in the interdisciplinary McWilliams Center for Cosmology[4] and the Cosmostatistics Initiative[5] (COIN), the latter of which was a driving factor for the work presented in Chapter 3 and Chapter 4.

While some subareas of machine learning clearly belong to the field, the differentiation between statistics and machine learning can be a contentious topic due to the close interconnection between the two areas of research. With the more recent wave of 'data science' as the new in-vogue term, both fields, as well as the field of operations research at times, can be seen trying to claim this new word for their own. In reality, these areas have long shared methodology and seeded each other, with operations research stemming from mathematical optimization, and both the latter and statistics more generally are heavily used in machine learning. Optimization is crucial to learning from data, while statistics delivers much of the underlying methodology of many machine learning models and is, especially in terms of inference methods, often linked to optimization itself. In fact, one can observe the occasional debate on social media or over a coffee at academic institutions, pondering the question whether simple linear regression is, technically, machine learning, while statistical learning theory underpins much of the former area as a whole (Hastie et al., 2001).

An often-cited difference between the two fields lies in their purpose; while statistical inference aims to uncover relationships between variables of interest and test hypotheses, machine learning is preoccupied with the prediction of

[4] https://www.cmu.edu/cosmology/
[5] https://cosmostatistics-initiative.org/

unobserved events based on existing datasets. This definition does, of course, only account for *supervised machine learning*, meaning exactly said prediction based on training data, and does not include *unsupervised machine learning* methods such as cluster analysis, which tries to find patterns in unlabeled data. While a generalization, it serves as a useful differentiation between the two fields, although Bzdok et al. (2018) correctly point out that many methods from both fields can be used for both inference and prediction, and are used that way due to both being of interest in scientific applications. In practice, the overlap is significant, and a lot of the perceived differentiation comes down to historical reasons, meaning that machine learning methods and terminology stem primarily from computer science departments, while statistical methods are generally pursued at mathematics departments.

In this thesis, we make no use of the particular set of architectures that led much of the recently renewed interest in machine learning in a variety of fields, namely *artificial neural networks*. Although the author's published work includes papers on this topic, the latter form no part of the work covered here (Moews et al., 2019b; Fussell & Moews, 2019; Boucaud et al., 2019; Moews & Ibikunle, 2020). For this reason, we spare readers an in-depth coverage of this part of the field, as it would go beyond the scope of this thesis. We should also note that machine learning, while a powerful tool for a variety of tasks that has started to attract a lot of interest during the past few years, is subject to the same limitations that often apply to tools, meaning that they are made for specific purposes. As such, machine learning, too, should be viewed through the lens of utility and not be used for its glamor alone, as there are often established methods more suitable for a given problem. In this context, interpretability is a limitation that is generally encountered when trying to link the performance of models to underlying physics. We do, therefore, stress that the advantages and disadvantages of these methods should be weighed and contrasted with both the problem that is attempted to be solved and the answers one hopes to get from doing so.

Machine learning techniques employed in the context of optimization can be useful when trying to constrain parameters of cosmological simulations themselves, both in terms of observational data and formalisms like the equilibrium model. For parameter optimization, MCMC methods are commonly used in astronomy, but can face certain limits in terms of expensive likelihood calculations and high numbers of dimensions, as described in Section 1.2.3. Swerving away from the more traditional statistical approaches to uncertainty estimation, Monte Carlo

dropout, as introduced by Gal & Ghahramani (2016), provides an approximation of Bayesian inference, specifically to a Gaussian process. Generally speaking, dropout is a regularization technique in neural network models to avoid overfitting by randomly dropping nodes during the training process, forcing the model to spread it weights more evenly across inputs. In Monte Carlo dropout, where dropout is applied at the testing instead of training step, the model architecture in each dropout realization is interpreted as being different, with the outcome being an averaging ensemble of different models acting on a subset of the data. While this method circumvents the often prohibitive computational costs associated with Bayesian methods, the dropout rate has to be fine-tuned to a given problem, and a convergence to concentrated distributions is not guaranteed, showcasing the trade-off often made in machine learning (Osband, 2016; Gal et al., 2017).

In contrast to the more limited body of work available when the work presented in this thesis started, an overview of the entirety of machine learning in cosmology would exceed the scope of this introduction, which demonstrates the growth this area experienced in the last few years. We will, for this reason and because the more statistically oriented machine learning relevant to other chapters was covered in previous sections, focus on work relevant to Chapter 5.

Recently, machine learning has been employed to directly learn the relationship between dark matter haloes and baryonic properties within hydrodynamic simulations, which then allows populating those galaxy properties into dark matter haloes (see, e.g., Kamdar et al., 2016a,b; Agarwal et al., 2018; Moster et al., 2020). While stellar mass is quite accurately predicted, other properties such as star formation rates and gas contents have substantially poorer accuracy, which limits its usefulness in the age of increasingly precise surveys. The type of machine learning model associated with our work in Chapter 5 is of interest in terms of preparing the reader for the latter, especially in connection with related research on baryonic property prediction.

The equilibrium model covered in Chapter 5 only directly predicts stellar mass ($\boldsymbol{M}_*$), star formation rate (SFR), and metallicity ($\boldsymbol{Z}$), whereas cosmological applications often require a wider suite of baryonic properties such as neutral hydrogen content for HI intensity mapping. While the directly-predicted star formation and metallicity histories straightforwardly yield galaxy luminosities, accessing a wider suite of galaxy properties requires employing machine learning on hydrodynamic simulations that directly predict such properties. Although it is possible to use machine learning directly on the dark matter and thus

bypassing the equilibrium model, Agarwal et al. (2018) report that their approach to predicting HI properties for neutral hydrogen results in a large scatter relative to the true values. Factors like feature selection can, due to their respective relevance, play a role in these predictions, and a pure machine learning approach is not ruled out by this exploratory study.

When, however, additionally provided with true SFR and metallicity information, their machine learning model performs much better, even qualitatively recovering the second-parameter correlation between stellar mass, metallicity, and SFR (the so-called fundamental metallicity relation) (Agarwal et al., 2018). Thus, we can presume that by first applying the equilibrium model to dark matter-only simulations to predict key baryonic properties, and then feeding that information into the machine learning model along with dark matter properties, it becomes possible to substantially improve the accuracy, or ease the process of reaching improved accuracies, of inserting galaxies into dark matter haloes.

In Chapter 5, we extend and merge the equilibrium model into a machine learning framework to predict galaxy stellar and gas evolution within a merger tree. We demonstrate the effectiveness of this approach using the recently completed Simba cosmological hydrodynamic simulation (Davé et al., 2019). Along the way, we implement extensions of the equilibrium model to account for the free-fall time within haloes, and enable the model to process largest-progenitor merger trees instead of just initial halo masses. By fusing this extended equilibrium model into an extremely randomized tree ensemble, a machine learning technique previously identified to perform well on the problem of baryonic property prediction, we advance the fields of both analytic galaxy evolution models for cosmological applications and machine learning for specialized tasks in the investigation of galaxy evolution (Kamdar et al., 2016a,b; Agarwal et al., 2018; Jo & Kim, 2019; Hearin et al., 2020; Moster et al., 2020; Fluke & Jacobs, 2020).

1.2.6 Decision trees and ensemble methods

Due to their simplicity and interpretability in terms of their decision-making process, *decision trees* remain one of the most widely used machine learning algorithms (for a review, see Wu et al., 2008). As the name suggests, a decision tree is a hierarchical structure, starting with the input at the 'root' and being subsequently split at nodes, usually in a binary manner, with final nodes corresponding to predicted values or classes being known as 'leaves'. As

such, these models can be viewed as generative models to create induction rules, making them an example of 'white-box' models with clearly interpretable decision paths, the counterpart to black-box models like many types of neural networks. Notably, the use of multiple decision trees makes the resulting 'forest' emerge as a black-box model, providing a trade-off of often better performance against a decrease in transparency (Guidotti et al., 2018).

Introduced by Breiman et al. (1984), `CART` (short for 'Classification And Regression Trees') is an early decision tree learning algorithm suitable for regression problems. `CART` traditionally makes use of the Gini impurity as a way to quantify the probability of incorrect classifications of data points, assuming that the latter's classification is drawn randomly from the label distribution in the available data. This property is used to choose variables to split at the tree's nodes in a way that maximizes homogeneity to build efficient hierarchical structures. It takes the form

$$\mathrm{I}_G(p) = \sum_{i=1}^{N} p_i \sum_{k \neq i} p_k = \sum_{i=1}^{N} p_i(1 - p_i) = 1 - \sum_{i=1}^{N} p_i^2, \tag{1.58}$$

summing the probability of a given label, p_i, multiplied with the probability of a labelling error, and corresponds to Tsallis entropy, a generalized version of Boltzmann-Gibbs entropy in statistical mechanics. Conversely, the `ID3` algorithm by Quinlan (1986) and its successor generally use the information gain,

$$\mathrm{IG}(T, a) = -\sum_{i=1}^{N} \; p_i \; \log_2 \; p_i - \sum_{i=1}^{N} -\mathrm{p}(i|a) \; \log_2 \; \mathrm{p}(i|a), \tag{1.59}$$

for N classes, a given label i, and p_i as the percentages of each of these labels present in a splitting node. The expected information gain is then the mutual information, or reduction in entropy given by an optimal split. Given the iterative locally optimal splitting of datasets, building decision trees is a type of greedy algorithm (Quinlan, 1986). Algorithm 1 shows the pseudocode for a two-class ID3 algorithm to illustrate how decision trees are built (see also Quinlan, 1986).

Data: S := Training dataset,
a_t := Target attribute,
a := Total set of attributes
Result: r := Root to be returned
Require: $S \neq \emptyset$, length$(a) > 0$
$r \longleftarrow$ newly created root node
If all labels are one class, return a single-node root
if $\forall a_i \in a : a_i == 0$ **then**
  return r as a single-node tree root with label 0
if $\forall a_i \in a : a_i == 1$ **then**
  return r as a single-node tree root with label 1
Specify the label with which to proceed
$l \longleftarrow$ most common value of a_t
Determine the splitting attribute via Eq. 1.59
$\hat{a} \longleftarrow a_i \in a$ that best classifies examples
Set the decision tree attribute for r to $\hat{a}$
for all possibles values v_i of $\hat{a}$ **do**
  Add a tree branch to r that corresponds to $\hat{a} = v_i$
  $S_{v_i} \longleftarrow \forall s \in S : s == v_i$ for $\hat{a}$
  if $S_{v_i} == \emptyset$ **then**
    Add a lead node with label l to the tree branch
  else
    Iteratively continue building the tree
    Add a sub-tree ID3$(S_{v_i}, a_t, \forall a_i \in a : a_i \neq \hat{a})$
  end
end
return r

Algorithm 1: Clasification tree building in the ID3 algorithm.

While decision trees have a number of advantages, such as little preprocessing when compared to one-hot encoding like in neural network approaches, training speed, and being non-parametric models that lack a requirement for assumptions about a dataset's shape, they are not without fault. Important hyperparameters to tune the building of decision trees include the maximum depth to avoid overfitting, as well as the minimum amounts of samples in a split or in a leaf node, and further hyperparameters such as a class balance requirement in leaf nodes. Another disadvantage is that, through the discrete nature of a finite number of leaf nodes, predictions in regression problems tend to exhibit step-like demarcations. For regression trees, the commonly used criterion is the mean

squared error (MSE), which can be calculated, for a node n, associated data D_n, and samples M_n, as

$$\text{MSE}(D_n) = \frac{1}{M_n} \sum_{y \in D_n} (y - \bar{y}_n)^2, \text{ with } \bar{y}_n = \frac{1}{M_n} \sum_{y \in D_n} y. \tag{1.60}$$

Another common splitting criterion in regression is the mean absolute error (MAE), which replaces the squaring above with the modulus to obtain the non-squared difference. Using the MAE has the disadvantage of not being very punishing toward gross mispredictions with a linearly scaling error, often limiting its suitability to datasets in which outliers can be safely disregarded.

In machine learning, an *ensemble* refers to a finite set of models, the output of which is used to produce the final output, for example by averaging or weighting the individual outputs. This is primarily done to combine a multitude of 'weak learners' into a stronger predictive model, and to realise better generalization. Two of the primary methods are *boosting* and *bootstrap aggregation* (also known as *bagging*). The former incrementally refines an ensemble by sequentially training models on data points previously determined to be 'hard', while the latter involves random draws with replacement from the dataset to create artificial training sets for the separate trees in an ensemble, averaging their output for predictions (Bishop, 2006).

In the case of decision trees, the earliest and most wide-spread type of ensemble is the *random forest*, an ensemble learning method first introduced by Ho (1995) and further expanded by Breiman (2001). Making use of the random subspace method, trees are trained on bagging-style random samples of data points with replacement. The analogy to bagging as explained above comes into play as each model in the ensemble is trained on a dataset with a randomly sampled subsets of features, as opposed to randomly sampled subsets of observations, to increase accuracy and prevent overfitting to the training dataset (Ho, 2002). For multivariate regression problems, while one could build separate models for each output, training time concerns and correlations between output values for a given input usually lead to implementations using decision trees that predict all outputs (Dumont et al., 2009; Segal & Xiao, 2011).

Feature importances are commonly used to calculate the contribution of given inputs variables to the growth of tree-type machine learning models. These

importances do, however, relate to which features are used most heavily in the construction of trees and do not necessarily connect with relationships in the underlying data, in this case physical importance. Correlated features can further bias the results, which is why these approaches should be treated with caution when it comes to their interpretation. For this reason, the development of statistical methods suitable to explore the underlying importances of interest is an active area of research (Strobl et al., 2007; Altmann et al., 2010; Fisher et al., 2019; Zhou & Hooker, 2020). Correlation also plays a role in terms of errors, as already work already demonstrates that a reduction in correlated errors feature a linear relationship with error reduction in tree ensembles (Quinlan, 1986).

Adding an additional layer of randomness, *extremely randomized trees* (usually shortened to *extra trees*), while retaining the random subspace method targeting a random subset of features, discard the bootstrap sample, training each tree on the complete training dataset and choosing node splits based on a randomized selection instead of computing information-theoretically optimal splits (Geurts, 2006). While complete randomness can be used to generate the splitting choices, split points for a given feature are usually randomly sampled from a uniform distribution of the feature's value range in the training dataset, followed by using the optimal choice among those randomly generated split proposals. Using the mean squared error as the splitting criterion translates to variance reduction for the feature selection. They also feature less correlated errors than random forests, although the increased variance reduction comes at the cost of a slightly increased bias. The splitting process of extra trees in particular is further discussed in Chapter 5. Notably, the randomization aspect of extra trees means that another of the advantages over random forest models is a reduction in runtime.

Chapter 2

Gaussbock: Fast parallel-iterative cosmological parameter estimation with Bayesian nonparametrics

In this chapter, we propose a parallel-iterative algorithm to address current challenges in high-dimensional parameter estimation with expensive posterior calculations, making use of recent advances in the fields of statistics and machine learning. Our method starts with a preliminary approximation of the target distribution, either through a built-in affine-invariant MCMC ensemble or a user-provided initial sample guess. We then fit a non-parametric model to the sample and employ a variation of sampling-importance-resampling to iteratively move the samples toward the true distribution, repeating these steps until the process converges. In doing so, we also offer a user-friendly Python package to both the cosmology and the wider astronomy community, as well as a general parameter estimation tool for other disciplines dealing with the same issues. We test our implementation on the DES Year 1 (Y1) posterior, and on a fast approximation to the latter for extended tests.

This chapter is organized as follows. We cover the relevant methodology, which includes an overview of variational inference for Bayesian mixture models and truncated importance sampling, in Section 2.1, and describe the mathematical architecture of the proposed approach in Section 2.2. In Section 2.3, we introduce an open-source implementation based on our method, explain computational considerations and parallelization, and provide a quickstart tutorial. Experiments

for both toy examples and an approximation of the DES Y1 likelihood are covered in Section 2.4, together with cosmological parameter estimation runs on supercomputing facilities for the full DES Y1 likelihood. We present and discuss the results of these experiments in Section 2.5, and summarize our findings in Section 2.6. This work has been peer-reviewed and published in *The Astrophysical Journal* (Moews & Zuntz, 2020).

2.1 Mathematical background

While Bayesian inference has earned its place as a powerful instrument for cosmological research, complex problems often suffer from the need to approximate probability densities that are difficult or outright infeasible to compute. Since Bayesian methods rely on the posterior density, approximations are a necessary evil. In the algorithm presented in this chapter, we fit a mixture model to sample from the posterior using variational inference methods, while avoiding fixing the number of mixture components by using a Dirichlet process. We iteratively improve these samples using truncated importance sampling until a convergence criterion is fulfilled.

In this section, we provide an overview of sampling-importance-resampling in Section 2.1.1. After covering truncated importance sampling as a recent development in Section 2.1.2, we introduce a novel method for parallel-iterative parameter estimation in Section 2.2.

2.1.1 Importance Sampling

Importance sampling was described early by Kahn & Marshall (1953) in the context of sample size reduction in Monte Carlo methods and continues to inspire a wide array of extensions. This includes physics-specific techniques like umbrella sampling for difficult energy landscapes by Torrie & Valleau (1977) and, more recently, methods to alleviate issues with poorly approximated proposal distributions (Ionides, 2008). It is also a staple in cosmological parameter estimation, for example in Wraith et al. (2009) and Kilbinger et al. (2010). Generally, the basic method is a way to estimate distribution properties if only samples from a different, often approximated, distribution are given. Let $p(\mathbf{z})$ be the target distribution, $q(\mathbf{z})$ an approximate (or proposed) distribution, and $f(\mathbf{z})$

some function. The expectation of $f(\mathbf{z})$ can then be computed as

$$\begin{aligned}\mathbb{E}[f] &= \int f(\mathbf{z})p(\mathbf{z})\ d\mathbf{z} \\ &= \int f(\mathbf{z})\frac{p(\mathbf{z})}{q(\mathbf{z})}q(\mathbf{z})\ d\mathbf{z} \\ &\simeq \frac{1}{N}\sum_{i=1}^{N}\frac{p(\mathbf{z}_i)}{q(\mathbf{z}_i)}f(\mathbf{z}_i),\end{aligned} \tag{2.1}$$

with N as the number of drawn samples. The ratios in this equation, which are called the importance weights (or importance ratios) and are a central concept of the method, are given as

$$r_l \equiv \frac{p(\mathbf{z}_l)}{q(\mathbf{z}_l)}. \tag{2.2}$$

Sampling-importance-resampling (SIR) is a two-step approach in which the importance weights for a set of samples are calculated, after which an equally-sized subset of these samples is generated by drawing from them with probabilities per sample indicated by the normalized importance weights. For a more in-depth introduction to importance sampling and other related sampling methods, see Bishop (2006).

2.1.2 Counteracting high-weight samples

One issue with this approach is the possibility of overly dominant samples, meaning points with disproportionately high posterior values in comparison to the rest of a set of model samples. During the importance resampling step, this dominance leads to copies of these samples being overrepresented, resulting in sets that are too narrow in their densities. We address this issue with *truncated* importance sampling, an extension of importance sampling that truncates weights of high-value samples based on the total number of drawn samples, with guarantees for finite variance and mean-square consistency under weak conditions (Ionides, 2008). For a set of N_i samples, proposal distribution posteriors $q(\theta_i)$, actual posteriors $p(\theta_i)$ and a set truncation value α with justifications to be set at $\alpha = 2$, the weight w_i of a single sample is updated according to

$$w_i = \min\left(r_i, \bar{r}N_i^{\frac{1}{\alpha}}\right), \text{with } r_i = \frac{p(\theta_i)}{q(\theta_i)}, \tag{2.3}$$

where $\bar{r}$ is the mean of all importance weights for the sample. With this extension applied to SIR, the weighted drawing of samples is limited by the truncation value. This change improves the behavior of importance sampling during the early part of the algorithm described below, when the estimated distribution q is a poor approximation to the desired posterior p, and alleviates the issue of working with relatively small sample sizes for high-dimensional parameter spaces.

2.2 The Gaussbock Algorithm

Based on Bayesian nonparametrics and machine learning, we introduce an algorithm that uses variational inference on an infinite Dirichlet process approximated via stick-breaking to fit variational Bayesian Gaussian mixture models (GMMs) in an iterative manner. This algorithm offers a highly adaptive and embarrassingly parallel way to approximate high-dimensional posteriors with computationally expensive likelihoods.

The idea behind our approach is to start from an initial sample guess, either from existing data or a short run of another sampler such as `emcee`. Based on the work on nonparametric VI by Gershman & Blei (2012), our algorithm uses a variational Bayesian GMM due to its ability to automatically determine the number of Gaussians required to produce a good fit by stick-breaking an infinite Dirichlet process mixture. For this reason, only the provision of a maximal number of Gaussians is required. The algorithm then determines means and variances for the optimal number of Gaussians given a sample and fitting tolerance. This is followed by drawing a new sample from the fitted mixture model, and a truncated SIR step to move the sample distribution further toward the true the posterior density. These steps are then repeated in an iterative manner until convergence, which is assessed from the change in the variance of importance weights at the end of each iteration:

1. Fit a variational Bayesian GMM to the sample,
2. draw a new sample from the newly fitted model,
3. perform an SIR step for a weighted sample, and
4. check inter-iteration variances for convergence.

Data: Initial posterior-space samples $\boldsymbol{\theta}_{\text{start}}$,
number of required output samples n,
array of allowed ranges per parameter $\mathbf{r}$,
number of samples drawn per iteration m,
safety margin multiplier for sampling c,
maximum number of mixture components g,
dynamically shrinking fitting tolerance d,
value for importance weight truncation α,
log-posterior function for $p(\boldsymbol{\theta}|\mathcal{D})$
Result: Approximated posterior samples $\boldsymbol{\theta}_{\text{final}}$
$\boldsymbol{\theta}_{\text{new}} \leftarrow \boldsymbol{\theta}_{\text{start}}$;
for $i \leftarrow 1$ **to** N **do**
  Calculate the (shrinking) model fitting tolerance
  $d \leftarrow a_1 - i \cdot \Delta a \cdot (N-1)^{-1}$
  Fit a variational Bayesian GMM to the samples
  $\mathcal{M}_i \leftarrow \text{VBGMM}(\boldsymbol{\theta}_{\text{new}}, d, g)$
  Sample a set of parameters from the fitted model
  $\boldsymbol{\theta}_i \leftarrow \boldsymbol{\theta} \sim \mathcal{M}_i$ s.t. $\text{length}(\boldsymbol{\theta}) = m \cdot c$
  Cut samples straying beyond the allowed ranges
  $R = \text{r}_1 \times \text{r}_2 \times \ldots \times \text{r}_{\dim(\boldsymbol{\theta}_i)}$
  $\boldsymbol{\theta}_i \leftarrow \boldsymbol{\theta}_i \cap R$
  Keep the required number of parameter samples
  $\boldsymbol{\theta}_i \leftarrow \boldsymbol{\theta}_i^{(1:n)}$
  Parallel calculation of true log-posterior values
  $\mathbf{p} \leftarrow p(\boldsymbol{\theta}_i|\mathcal{D})$
  Compute importance probabilities in linear space
  $\mathbf{w}_i \leftarrow \exp(\mathbf{p} - p(\boldsymbol{\theta}_i|\mathcal{M}))$;
  Compute the truncated importance probabilities
  $\mathbf{w}_i \leftarrow \min(\mathbf{w}_i, \bar{\mathbf{w}}_i \cdot \text{length}(\boldsymbol{\theta}_i^{\frac{1}{\alpha}}))$
  Renormalize the updated importance probabilities
  $\mathbf{w}_i \leftarrow \mathbf{w}_i \times (\sum \mathbf{w}_i)^{-1}$;
  Weighted sampling from the parameter samples
  $L \leftarrow \text{length}(\boldsymbol{\theta}_i)$
  $\boldsymbol{\theta}_{\text{new}} \leftarrow \text{sample}(\boldsymbol{\theta}_i, \mathbf{w}_i)$ s.t. $\text{length}(\boldsymbol{\theta}_{\text{new}}) = L$
  Terminate if convergence criterion is reached
  if $|\Delta\sigma_i^2 < t$ **then**
    break
  end
end
Return the user-specified number of final samples
return $\boldsymbol{\theta}_{\text{final}} \leftarrow \boldsymbol{\theta} \sim \mathcal{M}_i$ s.t. $\text{length}(\boldsymbol{\theta}) = n$

Algorithm 2: Pseudo-code for `Gaussbock`.

The way in which a mixture model is fit to a given set of data points deserves a short overview to bolster the previous material covered in the introduction. Belonging to the family of hierarchical models, mixture models fit a number of distributions of the same parametric family (in this case normal), with mixture weights as probabilities that, accordingly, sum up to one. For each mixture component, meaning for each distribution that forms part of the model, there are parameters such as, in the Gaussian case, mean and variance. In the Bayesian case, which we make use of in this chapter, both weights and parameters rely on prior distributions, as they are viewed as random variables drawn from a Dirichlet distribution themselves.

Standard mixture models employ the expectation-maximization algorithm (see Dempster et al., 1977), which is used to iteratively calculate the probability of each data point to be generated by each mixture component. These parameters are then updated to maximize the likelihood. In the variational case, VI is used as an extension of said algorithm to maximize a lower bound on the model evidence, as detailed in Section 1.2.4, adding regularization by integrating information of prior distributions. The additional level of hierarchy can be found in the treatment of parameters as random variables with their own (pseudo-)posterior distribution. Section 1.2.4 provides a more detailed explanation of variational inference and Dirichlet processes as the underpinning methodology. While a more in-depth overview of variational Bayesian inference and corresponding mixture models would inevitably go beyond the scope of this chapter, we recommend the reasonably short overview by Fox & Roberts (2012) and the more extensive works of Blei & Jordan (2006) and McAuliffe et al. (2006).

We use a dynamically shrinking tolerance d for the model-fitting step. Let a be the tuple denoting the initial and final model-fitting tolerances, with $a_1 > a_2$, and let N be the maximum number of iterations, then the tolerance d_i for a given iteration $i \in \{1, 2, \ldots, N\}$ is

$$d_i = a_1 - i \cdot \Delta a \cdot (N-1)^{-1}, \text{ with } \Delta a = a_1 - a_2. \tag{2.4}$$

This approach is related to the previously mentioned PMC algorithms initially introduced by Cappé et al. (2004), and applied to cosmological inference in Kilbinger et al. (2010). It differs, though, by the nonparametric nature of the model, which eliminates the bias present in the predetermined number of distributions in classical GMMs. It also adds the weight truncation to reduce the influence of overly dominant samples with high posterior values in relatively

small samples. Our method bears motivational similarity, although considerable methodological differences, to `CosmoABC`, while not being subject to the potential pitfalls of forward-simulation inference in ABC (Ishida et al., 2015).

In Algorithm 2, we provide a more complete pseudo-code representation of the most relevant parts of the approach described in this chapter, which we name `Gaussbock`. For this algorithm, we let $\mathbf{r}$ be the array of tuples representing the allowed ranges (min, max) per dimension, that is, per parameter. Furthermore, let N be the maximum number of iterations, m the number of samples to be drawn from each iteration's model, n the number of samples returned after termination, g the maximum number of Gaussians available for approximating the posterior distribution, and c a safety margin parameter greater than one to draw additional GMM samples in case some fall outside the parameter bounds. Finally, let $\mathcal{D}$ be the empirical data used for calculating the true likelihood. The specifics of the variational Bayesian GMM (VBGMM) with reasonable default settings, like the prior of the covariance distribution and the parameter initialization for the VBGMM, are omitted in order to keep the pseudo-code concise.

As (mostly adaptive) defaults are used for the settings of `Gaussbock`, only the initial approximative sample set $\boldsymbol{\theta}_{\text{start}}$, the number of iterations n, and the handle of a function to compute $p(\boldsymbol{\theta}_i|\mathcal{D})$ have to be provided with regard to the above pseudo-code. In addition, if no $\boldsymbol{\theta}_{\text{start}}$ is provided, the implementation described in Section 2.3 will automatically run an affine-invariant MCMC ensemble to procure that initial set of posterior-space samples. Since the determination of convergence is a common issue in MCMC methods, `Gaussbock` uses a convergence threshold t that terminates the iterative fitting-resampling procedure if reached before the maximum number of iterations. For this purpose, we measure the difference in inter-iteration weight variances $\Delta\sigma_i^2$, which takes the form

$$\begin{aligned} &\Delta\sigma_i^2 = |\bar{\sigma}_i^2 - \bar{\sigma}_{i-1}^2|, \\ &\text{with } \bar{\sigma}_i^2 = \dim(D)^{-1} \sum_{d=1}^{\dim(D)} \sigma(\log(w_{i_d}))^2. \end{aligned} \tag{2.5}$$

Here, the average logarithmic importance weight variance is denoted as $\bar{\sigma}_i^2$, providing the arithmetic mean over the dimensionality $\dim(D)$, meaning the number of parameters.

2.3 Software implementation

In order to make this algorithm readily available, we have released a Python 3 package incorporating the complete `Gaussbock` algorithm. The package is installable via `pip` from the Python Package Index[1], while documentation and source code are available in a public repository[2].

Figure 2.1 shows the schematic workflow of `Gaussbock`, with a choice between an automated initial posterior approximation and a user-provided sample guess, as well as the option to return importance weights and the final fitted model. The automated initial approximation makes use of an affine-invariant MCMC ensemble, as introduced by Goodman & Weare (2010), through the package `emcee` by Foreman-Mackey et al. (2013) and with parameters like the number of walkers being automatically determined based on the required function inputs. The only required inputs to the tool's main function are the lower and upper limits for each parameter ('parameter_ranges'), the handle of a function that accepts a point in the problem's parameter space and returns its log-posterior value ('posterior_evaluation'), and the desired number of posterior samples to be returned ('output_samples').

An overview of settable inputs is shown in Table 2.1. We strongly encourage users to provide parameter ranges that are scaled to the interval $[0, 1]$ when setting a threshold for the optional convergence determination ('convergence_threshold') due to its mean variance-based functionality. When setting a convergence threshold, we recommend a value of $\sim 0.01 \cdot \dim(\boldsymbol{D})$ as a choice that, based on the tests performed in the course of this work, takes increased dimensionalities into account when using the built-in convergence criterion. The implementation uses `schwimmbad`, a library for parallel processing tools, to provide MPI parallelization on parallel computing architectures (Price-Whelan & Foreman-Mackey, 2017). The use of MPI can be activated with the optional boolean input ('mpi_parallelization') being set to 'True'. Alternatively, for running the algorithms across multiple cores locally, the optional input 'processes' can be set to the number of desired cores to be used. The initial sample to start from can be provided by the user, for example through sampling a best-guess approximation or using the posterior from previous research ('initial_samples').

[1] `https://pypi.org`
[2] `https://github.com/moews/gaussbock`

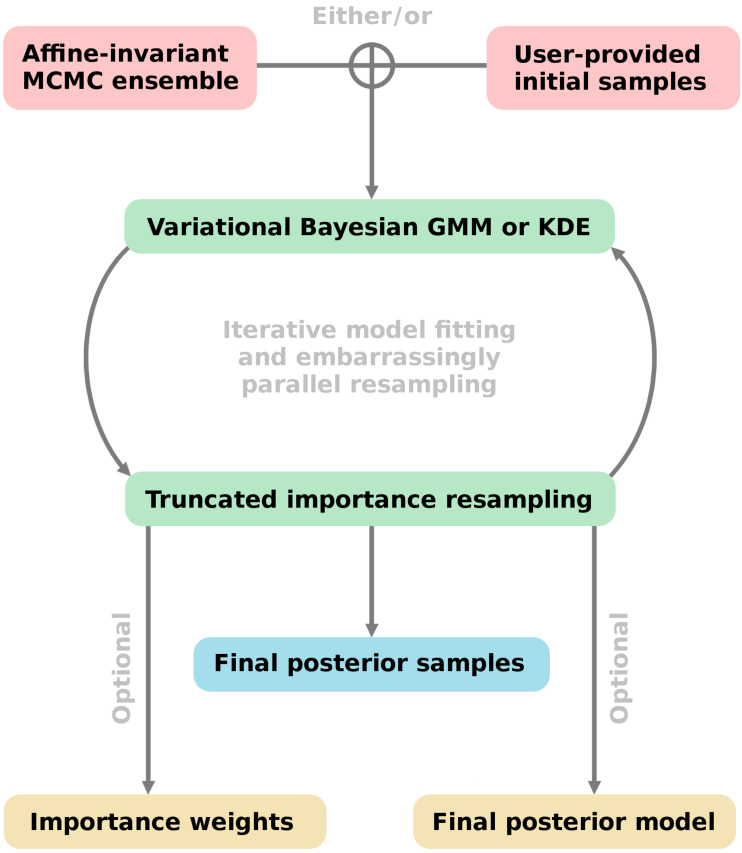

Figure 2.1 *Schematic workflow of* `Gaussbock`*. Inputs are colored in red, iterative steps in green, primary outputs in blue, and optional outputs in yellow. Starting with an initial set of samples that roughly appproximates the posterior distribution, the method uses an iterative model-fitting and parallelized sampling-importance-resampling step using importance ratio truncation to evolve toward tighter fits for the true posterior. Depending on the dimensionality of the problem, a variational Bayesian Gaussian mixture model (GMM) or kernel density estimation (KDE) can be used. This iterative step is repeated until convergence or a maximum number of iterations is reached. As indicated by the exclusive OR connection, the initial sample set can be user-provided or automatically inferred through a short-chained affine-invariant Markov chain Monte Carlo (MCMC) ensemble.*

Table 2.1 *`Gaussbock` inputs. The table lists all 19 possible inputs that can be set by the user, as well as a short explanation for each input, with the first three being required. The remaining 16 optional inputs are marked with an asterisk before their name and are default values are based on the tests presented in this chapter and should, as a result, generally achieve desirable performance for a wide array of problems reasonably similar to those described here.*

Input	Explanation	Default
1. `parameter_ranges`	The lower and upper limit for each parameter	
2. `posterior_evaluation`	Evaluation function handle for the posterior	
3. `output_samples`	Number of posterior samples that are required	
4. `*initial_samples`	Choice of `emcee` or a provided start sample	['automatic', $2 \cdot \dim(\boldsymbol{D}) + 2$, 10^3]
5. `*gaussbock_iterations`	Maximum number of Gaussbock iterations	10
6. `*convergence_threshold`	Threshold for inter-iteration convergence checks	None
7. `*mixture_samples`	Number of samples drawn for importance sampling	10^4
8. `*em_iterations`	Maximum number of EM iterations for the mixture	10^3
9. `*tolerance_range`	The range for the shrinking convergence threshold	$[10^{-2}, 10^{-7}]$
10. `*model_components`	Maximum number of Gaussians fitted to samples	$\mathrm{ceil}((2/3) \cdot \dim(\boldsymbol{D}))$
11. `*model_covariance`	Type of covariance for the GMM fitting process	'full'
12. `*parameter_init`	How to intialize model weights, means and covariances	'random'
13. `*model_verbosity`	The amount of information printed during runtime	1
14. `*mpi_parallelization`	Whether to parallelize Gaussbock using an MPI pool	False
15. `*processes`	Number of processes Gaussbock should parallelize over	1
16. `*weights_and_model`	Whether to return importance weights and the model	False
17. `*truncation_alpha`	Truncation value for importance ratio reweighting	2.0
18. `*model_selection`	Type of model used for the fitting process	'gmm' if $\dim(\boldsymbol{D}) > 2$, else 'kde'
19. `*kde_bandwidth`	Kernel bandwidth used when fitting via KDE	0.5

An input of special importance is the ability to set the variable parameter for truncated importance sampling ('truncation_alpha'), the ideal value of which can change based on the difficulty of the posterior approximation problem. By default, the recommended value of 2.0 is used (Ionides, 2008). When dealing with, for example, high-dimensional truncated Gaussians or similarly hard-to-approximate shapes, a value of up to 3.0 can enforce a stronger truncation to combat high-weight samples. Similarly, the truncation value can be set down to a minimum of 1.0 for weaker importance weight truncation. Interlinked with this input are the dimensionality of the problem and number of samples drawn from a fitted model in each iteration ('mixture_samples'), as a lower number of samples in a higher-dimensional parameter space increases the odds of importance weights with comparatively high values due to sparse samples. Time requirements and the number of available cores are the limiting factors for such considerations, which is discussed in the experiments in Section 2.4.

The algorithm's runtime can be further influenced by limiting the maximum number of Gaussians to be used for fitting a VBGMM during each iteration ('model_components'). By default, this input is determined based on the number of parameters to be estimated, but user knowledge about the complexity of the target distribution can inform the requirement for lower or higher maximums. Low-dimensional problems with $\dim(\boldsymbol{D}) < 3$ trigger the use of kernel density estimation (KDE) instead of a VBGMM by default, as this density estimation approach is quite powerful in such scenarios, but faces issues in higher-dimensional problems (O'Brien et al., 2016). The use of KDE or a VBGMM can, however, be forced by the user by setting the respective optional input ('model_selection') to either 'kde' or 'gmm'. The bandwidth used for the KDE functionality can be customized with an optional input ('kde_bandwidth'). We advise the use of KDE for low-dimensional problems due to the ability to catch hard-to-approximate posteriors in combination with our iterative method, which we demonstrate in Section 2.4.4.

2.4 Experiments

DES is an imaging survey that covers 5000 square degrees of the southern celestial hemisphere, operating a wide-field camera on the 4-meter Víctor M. Blanco Telescope located at the Cerro Tololo Inter-American Observatory (Abbott et al., 2016c). The survey probes cosmology using multiple different sources, including

galaxy clustering and lensing, cluster counts, and supernova measurements. Preliminary constraints from DES Science Verification (SV) data are presented in Abbott et al. (2016a) and Kacprzak et al. (2016) while, more recently, results and data for DES Y1 observations are described by Abbott et al. (2018a) and have been made public[3].

In this work, we use the Y1 weak lensing and galaxy clustering measurements as a test of `Gaussbock`. These measurements consist of a set of 2D two-point correlation functions of galaxy shape and position ("3x2pt") in tomographic bins by redshift. These functions can be predicted from the cosmological matter power spectrum and redshift-distance relation, both of which are sensitive to the underlying cosmological parameters, and especially to the matter density fraction Ω_m and the variance of cosmic structure σ_8. DES analyses yield constraints on these parameters comparable to those obtained from the CMB with Planck (Aghanim et al., 2018). For our experiments, we use the baseline ΛCDM model with varied neutrino density as our test likelihood. The sampling methods used in the main DES analysis are discussed in Krause et al. (2017); they use both the `emcee` affine-invariant sampler and the `MultiNest` nested sampling method, and found close agreement between the two methods.

In Section 2.4.1, we describe a fast-likelihood approximation of the DES Y1 posterior, followed by a performance test for `Gaussbock`. We explore scaling behavior of our implementation on the same approximation with experiments in Section 2.4.2. In Section 2.4.3, we run `Gaussbock` on the full DES Y1 posterior to test both the performance in real scenarios and the the ability to run fully parallelized via MPI on supercomputing facilities. Lastly, in Section 2.4.4 we test the behavior of the method on distributions with specific challenges and determine what types of failure modes it experiences.

2.4.1 Approximating the Dark Energy Survey posterior

The real DES Y1 likelihood is slow to evaluate, with durations per likelihood that make serial algorithms non-viable, as in Wilkinson (2005). In order to enable experiments that target controlled assessment and scaling behavior, we use an approximation to the DES Y1 posterior with a multivariate truncated Gaussian distribution, for which we employ the mean and covariance values

[3] `https://des.ncsa.illinois.edu/releases/y1a1`

for 26 cosmological and nuisance parameters, as well as their limits from the respective DES data release. This approach results in an extremely fast parameter set evaluation based on a DES Y1 approximation suitable for our purposes. A perfectly Gaussian approximation to the posterior would be an artificially easy test of a model that fits Gaussians; our posterior is truncated within a few sigma of the peak in many of its parameters, and thus provides a reasonable challenge.

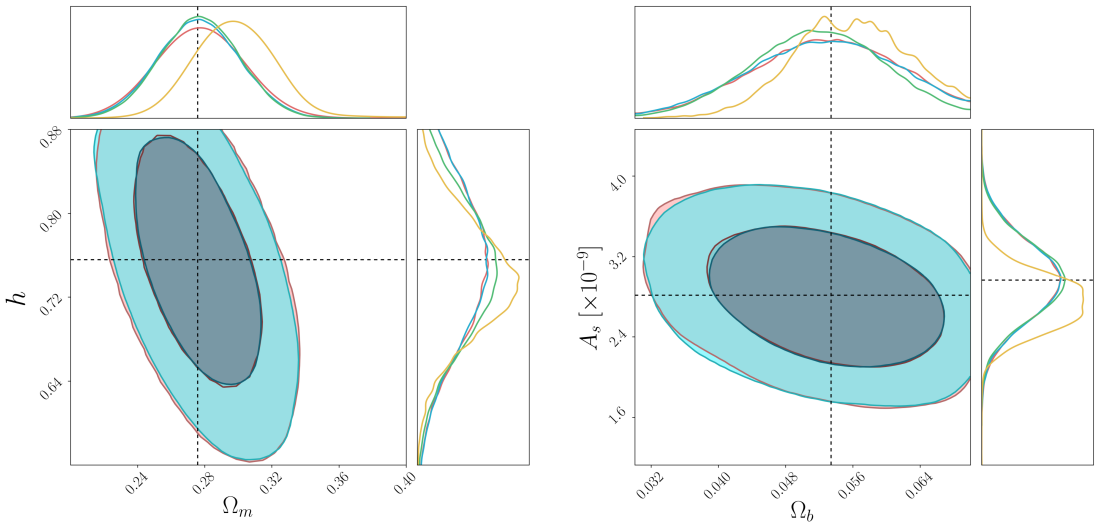

Figure 2.2 *DES Y1 posterior approximation with* `Gaussbock`*. The left figure depicts the matter density parameter (Ω_m) versus the Hubble constant (H_0), whereas the right figure shows the baryon density parameter (Ω_b) versus the scalar amplitude of density fluctuations (A_s). Contours for the importance-weighted samples generated with* `Gaussbock` *are drawn in blue, with contours for an* `emcee` *chain with* 5.4 *million samples across* 54 *walkers drawn in red. Darker and lighter shaded contour areas depict the 68% and 95% credible intervals, respectively. In addition to the same color coding as used in the contour plots, one-dimensional subplots for each parameter also show the unweighted distribution of* `Gaussbock` *samples in green, and the initial guess from which* `Gaussbock` *starts, obtained through a short-chained* `emcee` *run with* 1000 *steps per walker, in yellow. True means for DES Y1 data are indicated with dashed black lines to demonstrate the correct centering of both the fast approximation we employ in the experiment and the* `Gaussbock` *outputs.*

As discussed in Section 2.3, we use an increased truncation value for the SIR step of `Gaussbock`, which we set to 3.0, and a convergence threshold of $0.01 \cdot 26 = 0.26$ that follows the previously outlined best-practice guidelines and triggers the use of the built-in convergence determination. The number of samples per iteration is set to 15000, with the reasoning behind this choice further outlined in Section 2.4.2. As we want to weight the returned posterior samples with their importance weight, we activate the additional return of the final model and importance

weights. Apart from these settings, we use the default behavior of `Gaussbock` by not providing other optional inputs. Table 2.2 shows the lower and upper limits for cosmological and nuisance parameters employed in our approximation.

Table 2.2 *Cosmological and nuisance parameter limits for a fast approximation of the DES Y1 posterior. The lower and upper limits shown as open intervals closely follow prior distribution features previously used by DES for data from the first year of observations (Abbott et al., 2018a).*

Category	Parameter	Interval
Cosmology	Ω_m	$[0.1, 0.9]$
	H_0	$[0.55, 0.9]$
	Ω_b	$[3 \cdot 10^{-2}, 7 \cdot 10^{-2}]$
	n_s	$[0.87, 1.07]$
	A_s	$[5 \cdot 10^{-10}, 5 \cdot 10^{-9}]$
	ω_ν	$[6 \cdot 10^{-4}, 10^{-2}]$
Lens galaxy bias	$b_1, \ldots, b_5$	$[0.8, 3.0]$
Shear calibration	$m_1, \ldots, m_4$	$[-0.1, 0.1]$
Intr. alignment	A_{IA}	$[-5.0, 5.0]$
	μ_{IA}	$[-5.0, 5.0]$
Source photo-z	$\Delta z_s^1, \ldots, \Delta z_s^4$	$[-0.1, 0.1]$
Lens photo-z	$\Delta z_l^1, \ldots, \Delta z_l^5$	$[-5 \cdot 10^{-2}, 5 \cdot 10^{-2}]$

The results of this experiment are shown in Figure 2.2 and demonstrate the abilily of `Gaussbock` to recover correct constraints. Starting from a short and unconverged `emcee` chain, for which distributions are shown in yellow, the importance-weighted posterior samples marked in blue closely match the long-run `emcee` samples highlighted in red. The achieved level of agreement is good enough to make posterior contours and distributions for the target distribution and the importance-weighted samples hard to separate by eye. While the distributions for unweighted posterior samples in green show a good agreement with the long-run samples, weighting the output samples with the optionally returned importance weights pushes the sample distributions further toward to target posterior, thus validating the additionally provided functionality related to KDE for low-dimensional parameter estimation. While this experiment is based on an approximation of the full DES Y1 posterior, it offers a suitable testbed to prepare for the full-scale run described in Section 2.4.3.

When testing for considerably worse coverage of the true posterior by the initial sample, reaching convergence takes longer due to the increased amount of shifting that is required to reach a reasonable fit, so an as-good-as-possible coverage in the initial sample needs to be weighed against the time required to achieve

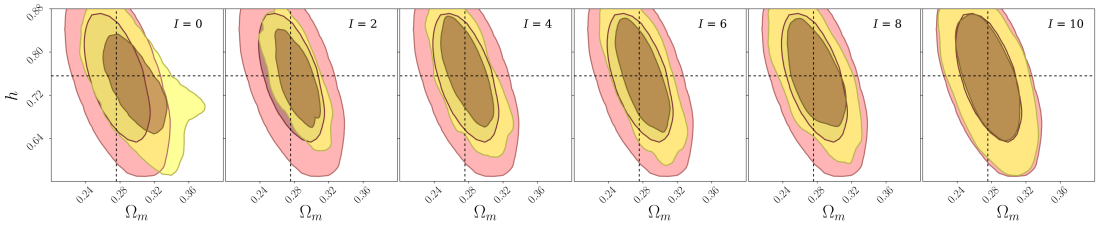

Figure 2.3 *Gradual improvement of contours across* `Gaussbock` *iterations. The figure depicts, in yellow, the importance-weighted posterior approximations for the matter density parameter (Ω_m) versus the Hubble constant (H_0). Each panel indicates the respective number of iterations I in the upper right corner, for iteration numbers from the the set $\{0, 2, \ldots 10\}$ to cover easily visible morphing behavior before fine-tuning takes place. Contours for an* `emcee` *chain with* 5.4 *million samples across* 54 *walkers are drawn in red to serve as a target distribution and orientation point across panels. Darker and lighter shaded contour areas depict the 68% and 95% credible intervals, respectively. On the far left, at $I = 0$, the posterior approximation corresponds to the initial sample guess. True means for DES Y1 data are indicated with dashed black lines.*

sufficient coverage. For this reason, the necessity of realizing a reasonable initial sample in the same approximate region as the true posterior is one of the major drawbacks. Due to the reliance on importance sampling, an essentially complete lack of coverage by the initial sample will also block the algorithm from iteratively shifting the approximation.

Another factor of interest is the iterative behavior of our algorithm, as `Gaussbock` is supposed to continuously improve the agreement of its internally generated samples with the true posterior distribution. In Figure 2.3, we illustrate this behavior, showing the gradual improvement of the constraints. The plots depict the morphing and shifting behavior of `Gaussbock` samples for the number of iterations as even integers in the interval [0, 10]. The cosmological parameters chosen for this experiment are the same as in the left-hand panel of Figure 2.2.

The evolution across the different panels showcase the algorithm's ability to start from a very rough sample guess and gradually move toward the target distribution. The latter is closely approximated by an extremely long `emcee` chain as an ideal sample. As demonstrated through this visualization, the algorithm first shifts generated samples toward the true mean with a lower-variance distribution, followed by incrementally spreading out to create a close fit to the target distribution.

2.4.2 Exploration of scaling behavior

In algorithms designed for the use with highly parallelized architectures, as well as in approaches for high-dimensional estimation problems, the question of how the algorithm in questions scales for different factors is important. For this reason, we now explore the scaling behavior of our algorithm. We quantify the time to convergence using the criterion introduced in Section 2.2, measured on the fast DES Y1 approximation covered in Section 2.4.1.

Higher-dimensional problems can, in general, be assumed to lead to a greater complexity of the estimation procedure, forcing `Gaussbock` to morph and shift the distribution in each iteration across more dimensions. We test our implementation for dimensionalities $3 \leq \dim(\boldsymbol{D}) \leq 26$, up to the full set of cosmological and nuisance parameters in our DES Y1 approximation, as a heuristic that proved to be robust for the various tests performed in the development of our approach. Other ways of convergence checks were tested, especially checks not taking dimensionality into account, but these tests resulted in very fragile convergence checks that required extreme fine-tuning due to higher-dimensional estimation problems leading to a larger variances between iterations. We perform this parameter estimation 50 times for each number of dimensions to create confidence intervals, with the respective subset of parameters being randomly selected. In each case, we use the convergence threshold $0.01 \cdot \dim(\boldsymbol{D})$.

The left panel of Figure 2.4 plots the number of iterations required to reach convergence versus the number of estimated parameters, showing the rise with problems of increased dimensionality. The 95% confidence intervals around the average number of iterations to convergence highlight the larger variance with increasing numbers of parameters. The average number of 26.6 iterations for estimating the full set of 26 parameters provides an indicator for the full DES Y1 posterior computation in Section 2.4.3.

The second question in terms of scaling behavior targets the embarrassingly parallel part of our algorithm, as we can vary the number of samples drawn at each iteration. Although the ability to parallelize across large numbers of cores is one of the strengths of `Gaussbock`, and while access to parallel computing architectures is wide-spread in modern cosmology, the number of available cores for a given task still faces upper limits. As described in Section 2.3, a higher number of samples drawn from a given iteration's fitted model is generally preferable, which

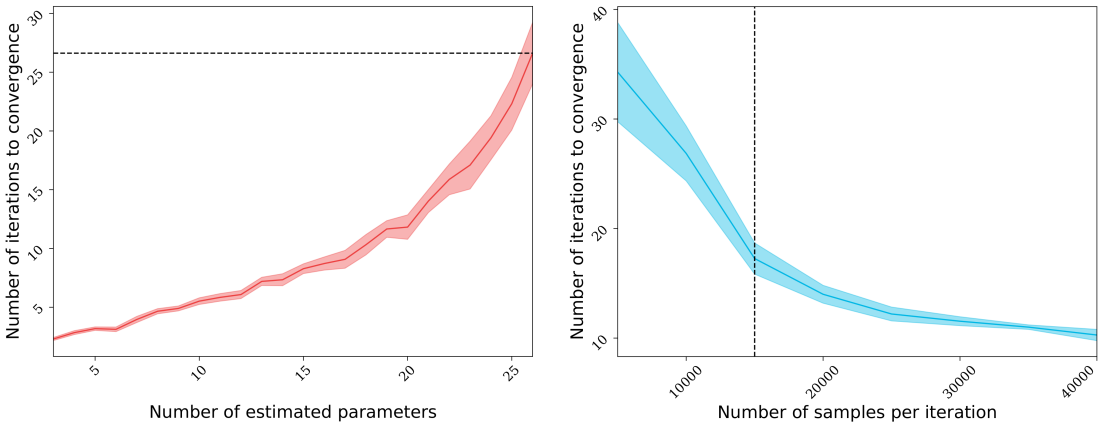

Figure 2.4 *Relationship between time to convergence, dimensionality, and the number of samples per iteration for* `Gaussbock`*. The left panel shows the number of iterations needed to achieve convergence, as a function of the dimensionality of the problem. The dashed black line indicates the mean number of iterations (26.6) needed for the full 26D DES Y1 parameter set. The right panel shows the number of iterations before convergence, as a function of the number of importance samples taken at each iteration, in steps of 5000. The dashed line marks the 'elbow criterion' for the trade-off in terms of time requirements from iterations and sample size, at 15000 samples. In both panels, the central line shows the mean and the shaded band the 95% confidence intervals over 50 simulations per point.*

translates to a preference for a higher number of cores due to the subsequent parallelization of the truncated SIR step. This poses the question of the scaling behavior of this benefit, as the required number of iterations to convergence is expected to decrease with a higher number of samples per iteration. The right panel of Figure 2.4 shows the scaling behavior of the required number of iterations to convergence versus the number of samples drawn from the fitted model during each iteration. We perform 50 `Gaussbock` runs per number of samples to create confidence intervals, in the interval [5000, 40000] and in steps of 5000.

Let I be the number of required iterations to convergence, C the number of available cores, and S the number of used samples per iteration. Then the total number of posterior value calculations per core over the course of a `Gaussbock` run is $I \cdot S \cdot C^{-1}$. Increasing numbers of samples constrain the variance of required iterations, and the dashed black line in the right panel of Figure 2.4 indicates an optimal trade-off (in terms of total core time) between the two variables as $\min(I \cdot S)$ at $S = 15000$ for the number of samples, which informs our input choices in Section 2.4.1. This visualization also bears resemblance to the 'elbow criterion' in cluster analysis, which determines the optimal number of clusters by plotting

that number against the explained variance (Thorndike, 1953).

In terms of a comparison to alternatives, as pointed out by Blei & Jordan (2006), comparing stochastic MCMC methods and deterministic variational approaches in a standardized way presents a challenge. The less constricted parallelizability of methods such as `Gaussbock`, though, is clear when noting that even multi-walker MCMC methods have a sequential component to them due to the Markov property of new states (see Section 1.2.2 for details). Assuming that the number of live points in nested sampling scales linearly with the dimensionality $\dim(D)$ of a given problem, Skilling (2009) provides a computational complexity of $O(\dim(D)^2)$, although Handley et al. (2015) challenge that view on scaling for higher dimensions. With the VBGMM employed by our approach scaling linearly with $\dim(D)$, the number of samples, and the number of mixture components used, respectively, the latter is bounded internally, and the number of points is fixed. It should also be noted that `MultiNest`, for example, is not fully embarrassingly parallel. Importantly, as visualized in Figure 2.4, positive scaling with the tackled dimensionality can be offset by the negative scaling with the number of samples per iteration, making this approach especially suitable for missions with access to allotments on large-scale supercomputing facilities.

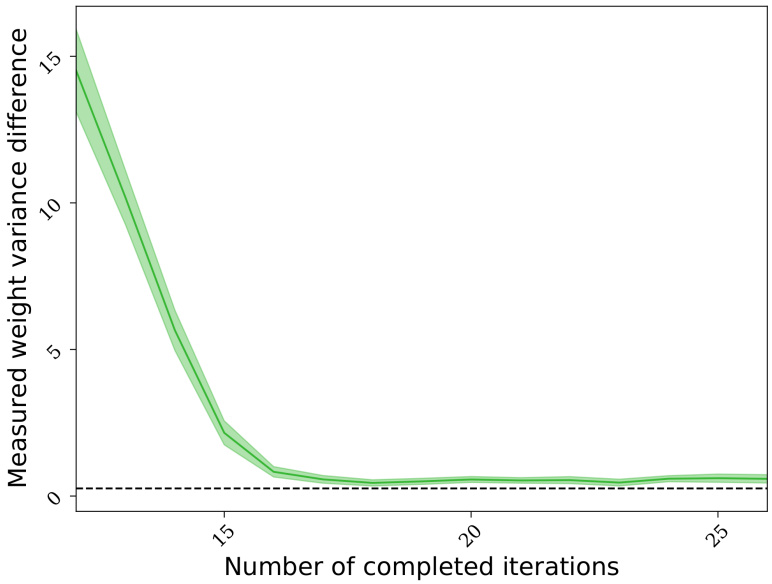

Figure 2.5 *Convergence behavior of `Gaussbock` for the number of completed iterations in approximated 26D DES Y1 analyses. The figure shows the inter-iteration change in variances of the logarithmic weights, used as a convergence criterion, with the dashed line marking the default convergence threshold for this problem. The mean value over 50 runs is shown as the central line, and the shaded band shows the 95% confidence interval.*

Lastly, we investigate the convergence behavior of `Gaussbock` as a follow-up to Figure 2.3, to ensure that both the convergence check itself and the recommended calculation of a convergence threshold behave as intended. The algorithm is run on the same parameter estimation problem as in Section 2.4.1, for a total of 27 iterations to cover the previously computed mean number of iterations to convergence of 26.6. As for previous tests, we run this experiment 50 separate times to generate 95% confidence intervals. The results are shown in Figure 2.5, starting after the first 10 iterations to cover fine-tuning behavior after the initial morphing and shifting explored in Figure 2.3, and with the dashed black line indicating the convergence threshold set to $0.01 \cdot \dim(\boldsymbol{D}) = 0.26$. The figure, showing a remarkably consistent and well-constrained behavior, demonstrates both convergence behavior for the threshold calculation and narrow confidence intervals for multiple experiments.

2.4.3 The full Dark Energy Survey posterior

In order to expose our method to a fully realistic experiment without approximations, we apply `Gaussbock` to the full DES posterior from the DES Y1 experiments and data release (Abbott et al., 2018a,b; Krause et al., 2017). We use the public `CosmoSIS` implementation of the public Y1 likelihood, which includes `CAMB` as described by Lewis & Bridle (2002) and Howlett et al. (2012), and `Halofit` as introduced in Smith et al. (2003) and Takahashi et al. (2012) to compute distances and matter power spectra, `CosmoSIS`-specific modules for the Limber integral and other intermediate steps, and `Nicaea`[4] for the computation of real-space correlations from Fourier space values (Kilbinger et al., 2009). Since the public implementation of the Y1 likelihood differs very slightly from the released chains, we rerun the model referred to as `d_l3` in the public DES Y1 chains using `MultiNest` for an identical comparison. The experiment starts with the same initial sample guess via a short-chained `emcee` run that we use for our fast approximation of the DES Y1 posterior in Section 2.4.1, demonstrating that our approach is able to start from approximative guesses that only partially fall within the vicinity of the target posterior and are not necessarily based on calculations using the actual target in question.

Making use of `Gaussbock`'s innate embarrassing parallelism, we run this experiment on supercomputing facilities of the National Energy and Scientific Research

[4] `http://www.cosmostat.org/software/nicaea`

Computing Center (NERSC)[5] (He et al., 2018). We run on 32 nodes of the Cori computer, for a total of 1024 cores and 2048 threads. The results below were generated in approximately two hours in this configuration, showcasing the total runtime advantage of our approach. With the runtime scaling being inversely linear with the number of cores, up to the number of samples used per iteration due to the model-fitting process not requiring a lot of time, up to 15000 cores can be used in an idealized scenario for our experimental setup to gain a further order-of-magnitude reduction. In order to make use of existing posterior implementations, we employ `CosmoSIS` to use `Gaussbock` with the DES Y1 posterior (Zuntz et al., 2015).

Table 2.3 lists the cosmological parameters as estimated by both `Gaussbock` and its comparison baseline, meaning the fiducial `MultiNest` run, demonstrating a satisfactory level of agreement for both means and credible intervals. In addition to the cosmological parameters shown in the experiments for Figures 2.2 and 2.6, the table also includes the scalar spectral index n_s and the massive neutrino density ω_ν, covering the full set of cosmological parameters previously listed in Table 2.2.

Table 2.3 *Cosmological parameters for DES Y1 data. The table shows figures of merit for common cosmological parameters used in the original DES Y1 experiments, with the latter's implementation of* `MultiNest` *and, for comparison, the results for a highly parallel* `Gaussbock` *run.*

Parameter	`MultiNest`	`Gaussbock`
Ω_m	$0.276^{+0.031}_{-0.031}$	$0.275^{+0.029}_{-0.026}$
H_0	$0.787^{+0.080}_{-0.106}$	$0.781^{+0.078}_{-0.080}$
Ω_b	$0.056^{+0.010}_{-0.012}$	$0.057^{+0.009}_{-0.013}$
n_s	$1.020^{+0.043}_{-0.064}$	$1.013^{+0.043}_{-0.065}$
A_s	$2.470^{+0.510}_{-0.440} \times 10^{-9}$	$2.430^{+0.420}_{-0.400} \times 10^{-9}$
ω_ν	$5.100^{+2.900}_{-2.800} \times 10^{-3}$	$5.000^{+3.000}_{-2.800} \times 10^{-3}$

Figure 2.6 shows the posterior contours for both the `d_l3` rerun with `MultiNest` and the `Gaussbock` result in red and blue, respectively. Both matter and baryon density parameters, Ω_m and Ω_b, are shown to match the baseline computation well, whereas the Hubble parameter H_0 and scalar amplitude of density fluctuations A_s are in reasonable agreement, but do not correctly recover the tails of the posterior distribution. An exploration of the 26-dimensional

[5] `https://www.nersc.gov`

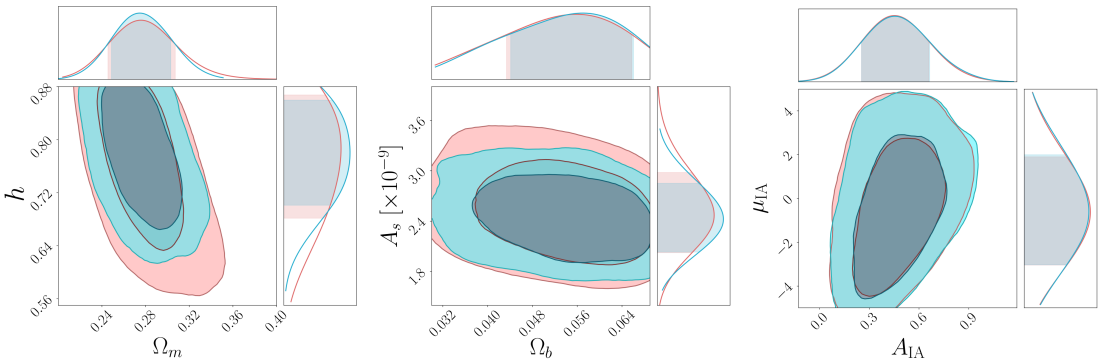

Figure 2.6 *DES Y1 posteriors with* `Gaussbock`*. The left panel depicts the matter density parameter (Ω_m) versus the Hubble constant (H_0), the middle figure shows the baryon density parameter (Ω_b) versus the scalar amplitude of density fluctuations (A_s), and the right figure shows the two intrinsic alignment parameters ($A_{\rm IA}$, $\mu_{\rm IA}$). Contours for the importance-weighted samples generated with* `Gaussbock` *are drawn in blue, with contours for the original nested sampling implementation as used by DES drawn in red. Darker and lighter shaded contour areas depict the 68% and 95% credible intervals, respectively, with the same levels shaded in the histograms.*

approximation shows that `Gaussbock` accurately models the parameters which are well-constrained, but fails to recover the tails on unconstrained parameters like H_0 and A_a that have very broad intervals, as listed in Table 2.2. Where possible, it might help to provide narrower constraints for such parameters. In terms of general difficulties when approximating higher-dimensional posteriors with Gaussbock as possible reasons for these results, the fact that Gaussians are used offers another explanation. If we imagine a simplified example of a single Gaussian being used to approximate a broad truncated Gaussian, this can lead to an underestimation of the tails, as the approximation will naturally overestimate the concentration due to a non-truncated distribution used for the fitting process. Another possibility is that the thresholding parameter used for the truncated importance sampling needs to be further fine-tuned for a given problem. While this is one side of the coin, another possibility is that `MultiNest` overestimates tails for these parameters due to the increasing inaccuracies in higher dimensions discussed by Chopin & Robert (2010) and Higson et al. (2018). In addition, Figure 2.6 shows the joint posterior of the two intrinsic alignment parameters, $A_{\rm IA}$ and $\mu_{\rm IA}$ in the right panel.

The results demonstrate the ability of `Gaussbock` to recover non-Gaussian shapes of correlated parameters to a high degree of accuracy, as can be seen in the 2D posterior shapes for the fiducial `MultiNest` and `Gaussbock` runs, as well as in the

agreement between histograms in the figure. The effective sample size for this `Gaussbock` run is $N_{\text{eff}} = 2104$, compared to $N_{\text{eff}} = 4316$ for the original `MultiNest` chain, although with a smaller overall runtime for our algorithm.

While the results are not in near-perfect agreement, as is the case for the fast truncated Gaussian approximation in Section 2.4.1, a trade-off between considerably reduced runtime and accuracy is to be expected analogous to the No Free Lunch Theorem in optimization (Wolpert & Macready, 1997). The described experiment on the full DES Y1 posterior makes use of `Gaussbock`'s adaptive default behavior and, for the number of samples per iteration, is based on our fast approximation, so fine-tuning to a specific application case can be expected to further improve the performance of the algorithm. Other reasons for the results not showing the same goodness of fit for all parameters, as observed in Section 2.4.1, are a diminished smoothness of posteriors and less Gaussian tails, which we discuss in Section 2.5.

2.4.4 Stress tests on additional distributions

In this subsection we run `Gaussbock` on distributions with more challenging features to determine when it starts to fail. As outlined in Section 2.3, KDE is a powerful density estimation technique, but faces issues in higher-dimensional problems (O'Brien et al., 2016). In this experiment, we exemplify the built-in default to use KDE for problems in which $\dim(D) \leq 2$, allowing `Gaussbock` to make use of the method most suitable to a given problem. For this purpose, we construct a posterior of three approximately equilateral triangles with a flat posterior surface, meaning that posterior values are uniform across the triangle shapes. Due to the convergence criterion of `Gaussbock`, which we discuss in Section 2.2, being geared toward the use of a VBGMM as its primary application in high-dimensional setting, we set the number of iterations to 20. We let the initial sample guess be generated automatically with the same number as for previous experiments in Section 2.4.1, and let `Gaussbock` use its default behavior for optional inputs.

The results of this low-dimensional parameter estimation experiment is shown in Figure 2.7, with 95% credible intervals for the flat-surface posterior demonstrating the ability of `Gaussbock` to approximate complex shapes with pronounced edges and corners. The three separate triangles are clearly reconstructed through the importance-weighted samples generated by the algorithm, validating its

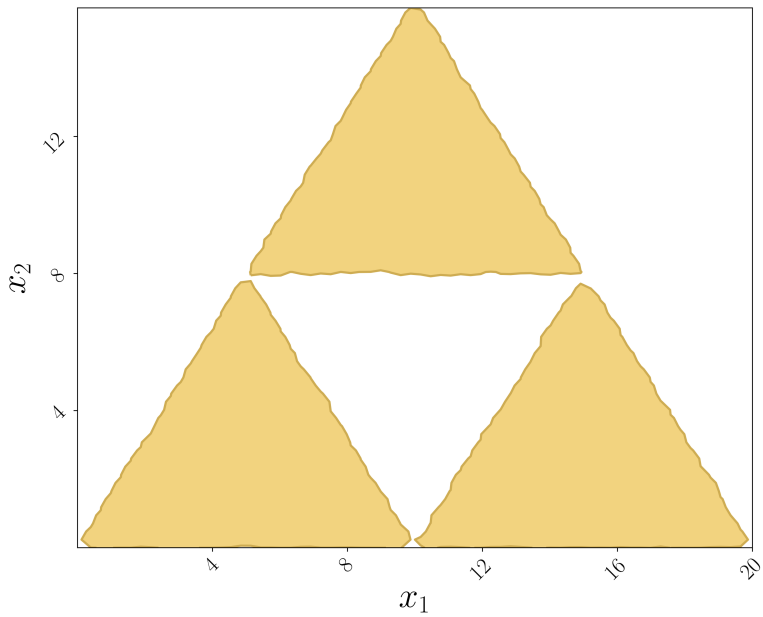

Figure 2.7 *Approximation of a hard-to-estimate posterior with* `Gaussbock`*. The two-dimensional posterior distribution features uniform values across the surface of three triangles. With a completely flat distribution of the posterior shape, the importance-weighted sample contours in the plot show the 95% credible interval for the generated samples.*

integrated KDE functionality for low-dimensional estimation problems.

Next, we consider similar stress tests based on those described in Hobson & Feroz (2008) and Feroz et al. (2009). First, we test with a posterior in the form of a double Gaussian shell, as described in Allanach & Lester (2008),

$$\mathcal{L}(\theta) = C(\theta; \mathbf{c_1}, w, r) + C(\theta; \mathbf{c_2}, w, r), \tag{2.6}$$

where

$$C(\theta; \mathbf{c}, w, r) = \mathcal{N}(|\theta - \mathbf{c}|; r, w^2). \tag{2.7}$$

At low dimensions, Gaussbock can sample effectively from such a distribution; the results from a 2D example with $w = 0.1$ and $r = 2$ are shown in Figure 2.8. The samples correctly trace the distribution, with a close-to-ideal match between the brute-force percentiles and the fraction of samples inside them. At moderate dimensions, from around 5D, Gaussbock fails on the sharp edges in this distribution, as the required number of Gaussians to capture the full shape becomes too high.

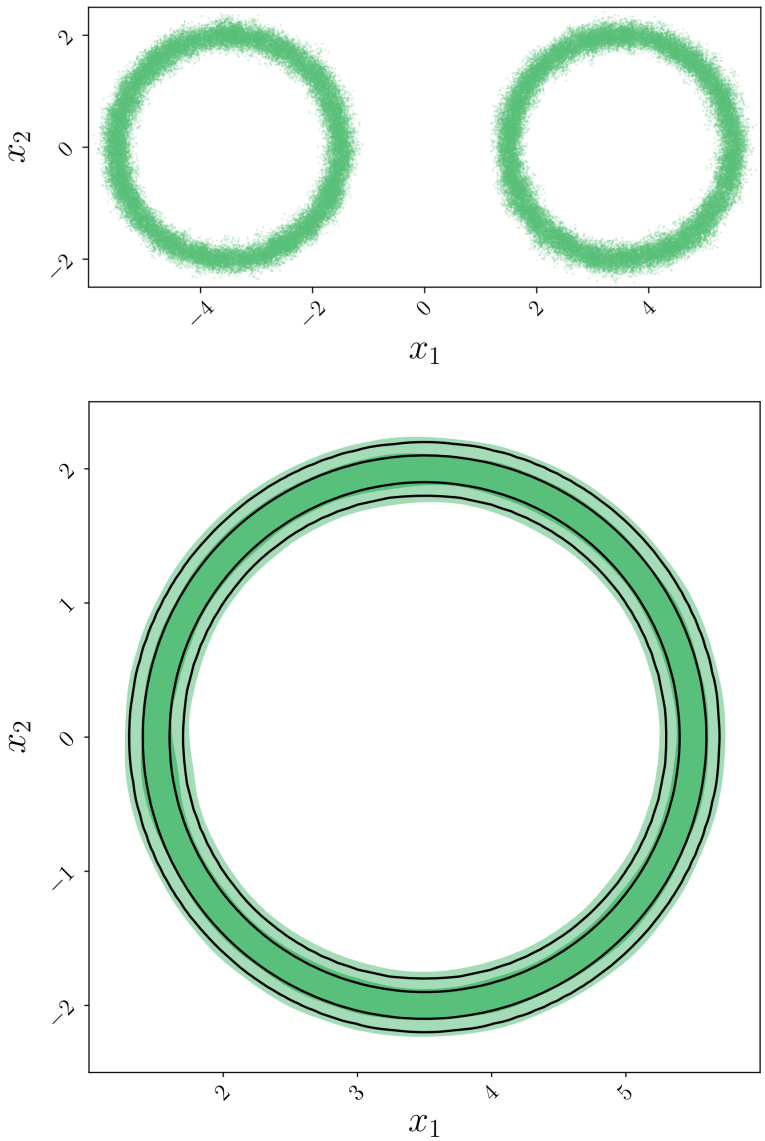

Figure 2.8 *Samples from a 2D Gaussian shell distribution. The upper panel shows a scatter plots of the resulting* `Gaussbock` *samples, while the lower panel zooms in on one of the two shells. For the latter, we show inner 68% and outer 95% contours from a brute-force grid evaluation in black, and KDE on* `Gaussbock` *samples as blue-shaded regions, with darker and lighter shaded contour areas depicting the 68% and 95% credible intervals, respectively. At higher dimensions,* `Gaussbock` *fails on such distributions.*

Next, we consider sharp edges that are poorly fit by Gaussian mixtures as another possible failure cases. We sample from

$$\mathcal{L} = \begin{cases} \exp -2 \cdot |\mathbf{x}|, & \text{if } \forall x \in \mathbf{x} : x > 0 \\ 0, & \text{else} \end{cases} \tag{2.8}$$

using a 4D example. This form has a sharp edge at $\mathbf{x}_i = 0$ in each dimension. Figure 2.9 shows the 1D distribution of one of the four parameters, as sampled using `Gaussbock`, `emcee`, and a brute-force evaluation. Both samplers undersample at this boundary[6], and this effect will worsen for `Gaussbock` at higher dimension.

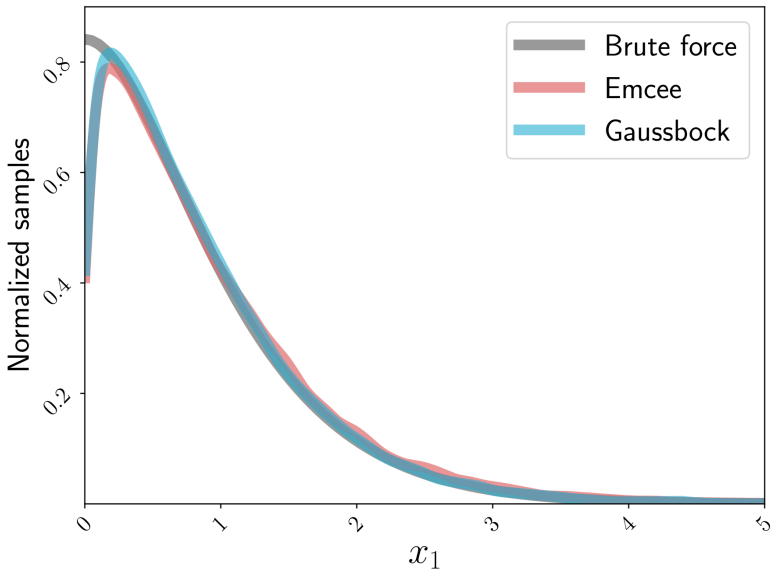

Figure 2.9 *Sampling behavior of* ***Gaussbock*** *on the distribution in Equation 2.8, with a sharp boundary in 4D, compared to a long-chained* `emcee` *run and a brute-force evaluation. Both samplers underestimate the PDF near the edge, although* ***Gaussbock*** *maintains a slightly smoother adherence to the true distribution otherwise.*

Finally, as a multimodal example, we consider sets of identical Gaussians, with centers arranged in a Latin hypercube formation so that they do not overlap in any dimension. The algorithm, starting from a random scattering throughout the space, finds all the modes for dimensions up to about six, as shown in Figure 2.10. At higher dimensions, the algorithm often misses some of the modes; this is an important failure case that is based in the reliance on an initial sample provided

[6]Many sampling methods based on Markov chains can suffer from repulsive effects at sharp edges of distributions, since proposals to points near the boundary can only happen from one direction; a variety of methods have been used to correct for this behavior (Ahmadian et al., 2011).

by either the built-in affine-invariant MCMC sampler or a method of choice. If the latter fails to catch at least part of some modes, the algorithm is unlikely to recover them.

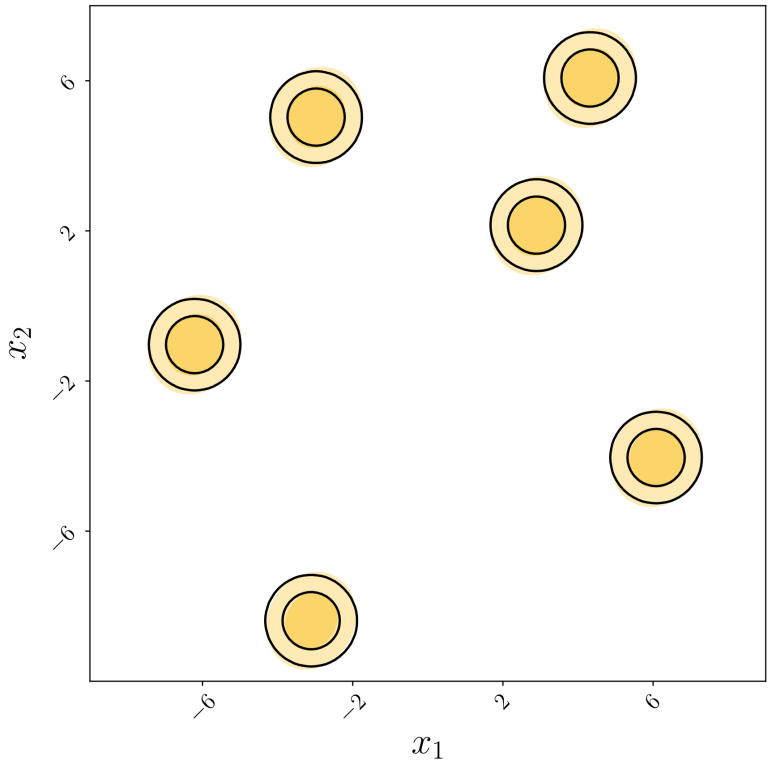

Figure 2.10 *A 2D projection of a six-dimensional distribution with six Latin hypercube-located Gaussian modes. We show a KDE on* `Gaussbock` *samples as yellow-shaded regions, with darker and lighter shaded contour areas depicting the 68% and 95% credible intervals, respectively. The algorithm typically finds all the modes up to about 6D, and then begins to miss them at higher dimensions due to the difficulty of catching them in the initial sample generation..*

It should be noted that most distributions found in the additional tests of this section are not usually found in the intended field of application, cosmological parameter estimation, but serve as a demonstration of the method's capabilities for classical tests found in the statistical literature, and could be of use in other application areas. The high-dimensional experiment performed in Section 2.4.1, as an approximation to the subsequently used DES posterior with a truncated Gaussian over 26 variables, bears closer resemblance to practical applications in cosmology. For new challenges such as the upcoming LSST and Euclid missions, however, we recommend additional stress tests specific to these, as is common in the preparation for new data sources.

2.5 Discussion

The primary advantage of our approach is the considerable reduction in the required runtime, given a large-enough number of cores available for parallelization. This strength offers a way to tackle increasing complexities in cosmological parameter estimation for current and upcoming surveys such as LSST and Euclid (Amendola et al., 2018). Since cosmological parameter estimation efforts rely on computationally costly posterior evaluations, the embarrassing parallelization of their calculation allows for an immense speed-up in comparison to standard MCMC approaches. This reduction in total runtime comes, however, at the cost of an increase in the required core time, meaning the number of computing hours necessary to achieve suitable results. For this reason, and assuming a sufficiently costly posterior evaluation, making use of `Gaussbock`'s parallelization capabilities is a requirement rather than an optional feature, as demonstrated in Section 2.4.3.

A direct comparison to MCMC methods is a double-edged sword in that such methods, run for a very large number of steps, provide a close fit to the true posterior. The downside of MCMC approaches is that they tend to not scale well with the number of dimensions, and that they are only parallelizable over the number of walkers. This means that computationally expensive likelihoods provide an obstacle to implementations such as `emcee` (Foreman-Mackey et al., 2013). While nested sampling methods circumvent this restriction by requiring fewer posterior evaluations, they rely on assumptions about perfect and independent samples and can sometimes underestimate an asymptotically Gaussian sampling error. In many cases, though, they can be highly effective, for both posterior and evidence estimation, depending on the problem at hand (Chopin & Robert, 2010).

As mentioned in Section 2.4.3, posteriors based on real-world survey data may have a less smooth posterior surface, which can hamper the effectiveness of the truncated SIR step used in our approach. Adjusting the 'truncation_alpha' input can alleviate this issue for isolated samples with higher posterior values, although a more effective solution is to increase the number of samples drawn from the posterior approximation of a given iteration of the algorithm. This approach does, in turn, require either a correspondingly larger number of cores or additional runtime. Alternatively, the initial sample guess to which the first-iteration model is fitted can be based on a longer-chain `emcee` run. As a result, this approach offers a better approximation of the posterior to start from, as it more closely resembles

the target distribution and leads to broader coverage of relevant areas. We hope that the presented work will lead to further investigations of this and related parallelized iterative approaches to parameter estimation, alleviating the issues arising from increased computational demands in inference based on modern surveys.

Apart from cases with sufficiently smooth posteriors and well-constrained parameters, `Gaussbock` also offers a way to quickly approximate a posterior to reasonable degrees. For this purpose, we recommend using either uniform-random samples from an n-sphere scaled to the admissible ranges or, if feasible, samples from a better-suited distribution like an n-dimensional Gaussian to provide an initial sample guess covering the posterior area. The reason for such approaches is the elimination of the need for computationally more expensive sample guess generators such as short-chained `emcee` runs, which require costly evaluations of the posterior. While short chains are fast in comparison to exhaustive runs of MCMC methods, runtimes should be kept to a minimum for fast approximations in order to provide an edge in speed over alternative approaches.

An additional use case pertains to lower-dimensional problems, or scenarios with posterior evaluations that are sufficiently cheap to compute, and offers a way to achieve very tight fits to posteriors that are hard to approximate and feature clean cuts, with an example given in Section 2.4.4 and one commonly-encountered example of such posterior shapes being truncated Gaussians. The suitability for the latter type also extends to higher dimensions, as we demonstrate with the truncated Gaussian approximation of the 26-dimensional DES Y1 posterior in Section 2.4.1. For the latter, as described in Section 2.4.3, an important finding is that `Gaussbock` accurately models well-constrained parameters, but can have trouble to recover the tails on unconstrained parameters perfectly. For that reason, setting sensible parameter constraints as one of the three required inputs to the implementation is strongly advised.

Unlike in most MCMC methods, the final mixture model is an optional output of our implementation, which can be saved and used again at a later point. It can act as an approximate but analytic description of the posterior, allowing, for example, the subsequent drawing of an arbitrary number of samples for which importance weights can be calculated and which can be easily disseminated. In this context, our approach offers a way to easily exchange and compare posterior approximations based on different datasets, with mixture models whose components can be combined.

For common problems faced by contemporary research in cosmology, `Gaussbock` offers a considerable speed-up. This is especially relevant for upcoming missions with larger numbers of parameters, for which our approach provides a way to quickly compute high-fidelity posterior approximations and the underlying mixture model. While, in this work, we use a wrapper to run `Gaussbock` through `CosmoSIS` on NERSC facilities, a complete integration into `CosmoSIS` will further enhance the ease of access to our methodology. Regarding the scaling behavior tested in Section 2.4.2, a higher number of dimensions leads to a higher number of iterations to reach convergence, as demonstrated in Figure 2.4. `Gaussbock` also benefits from an as-close-as-possible fit to the true posterior for the initial sample to start from. In cases in which such a sample guess is available, it lends an advantage to the method's performance when compared to using the built-in affine-invariant MCMC sampler. Notably, the ability to feed an arbitrary set of initial samples into the tool also means that `Gaussbock` can be combined with any sampling algorithm to create such an initial sample, allowing users to employ cutting-edge methods of their choice to make full use of the current statistical literature and personal preferences.

In terms of its internal functionality, our approach inherently lends itself to combating issues with defaulting cores, as the failure or a subset of processes to return importance values can be safely ignored. The respective parameter sets can simply be omitted from the set of samples used to approximate the posterior in a given iteration, using the large-enough amount of remaining parameter sets to fit the model in a given iteration. While the capability to do so is not part of our implementation and is primarily of interest for large-scale cloud computing, our code easily lends itself to being extended toward this safety redundancy.

2.6 Summary

In this chapter, we introduce and apply `Gaussbock`, a novel approach to cosmological parameter estimation that makes use of recent advances in machine learning and statistics. By coupling variational Bayesian GMMs with a truncation-based extension of importance sampling in an iterative approach with a convergence criterion, our method offers an embarrassingly parallel way to achieve high-speed parameter estimation for problems with computationally expensive likelihood calculations.

We initially test `Gaussbock` on a fast approximation of the DES Y1 posterior to demonstrate its capabilities on a high-dimensional realistic example, and to investigate scaling relations and the effectiveness of the convergence criterion, both of which prove to be well-behaved. We then apply our method to the full DES Y1 posterior, making use of `Gaussbock`'s built-in MPI capabilties to run it on NERSC supercomputing facilities. The results showcase the immense speed-up that constitutes the primary strength of our method, achieving a good fit to the original DES approach of using `MultiNest`.

While achieving excellent fits in most cases across our experiments, we observe that less Gaussian posteriors of unconstrained parameters result in a slightly worse fit to the tails of the distribution and discuss the potential issues arising from less smooth posterior surfaces. In addition, we stress-test the algorithm using more complex distributions. We also demonstrate that `Gaussbock` achieves tight fits to hard-to-approximate posteriors such as double Gaussian shells, scattered multivariate Gaussians, and exponential distributions in lower dimensions. The reliance on an initial sample guess roughly covering the areas of interest, however, means that it will break down if the latter is not the case, for example if modes of a multivariate distribution are not caught in that initial sample. In addition, we verify that our method, like other parameter estimation techniques based on Gaussian mixture models, is limited by the degree to which distributions can be formalized as a weighted mixture of Gaussians, which becomes problematic if, for example, facing Gaussian shells of moderate to high dimensionality.

We implement `Gaussbock` as a pure-Python package to conduct our experiments described in this chapter. In doing so, we also provide the astronomy community with a user-friendly and readily installable implementation of `Gaussbock`, bearing the same name. While our method is developed specifically with contemporary parameter estimation problems in cosmology in mind, it represents a general-purpose inference tool applicable to many problems dealing with high-dimensional parameter estimation with computationally costly posteriors. As a result, our work contributes to the wider field of estimation theory in addition to current and upcoming astronomical surveys.

Chapter 3

Stress testing the dark energy equation of state imprint on supernova data

In this chapter, given the literature on alternative parameterizations of the dark energy equation of state and their testing against the standard model described in Section 1.1.4, we aim to address the contrapositive question: How robust is a standard SN Ia analysis pipeline to deviations from ΛCDM in the data? We thus investigate whether the traditional ΛCDM analysis framework is, in this context, a meaningful process to begin with. By creating arbitrary realizations of $w(z)$, we stress-test the viability of currently wide-spread methods to measure w via SN Ia data for the assessment of dark energy models. To accomplish this goal, we explore current capabilities to discriminate between different models beyond a cosmological constant by running a standard cosmological inference pipeline on random fluctuations of the dark energy parameter w that adhere to physically motivated constraints.

This chapter is organized as follows. The SN Ia mock samples generated for subsequent experiments are described in Section 3.1, along with our procedure for generating data perturbations and the theoretical considerations that have to be taken into account when constraining $w(z)$. The analysis is performed according to the procedure outlined in Section 3.2, which provides an overview of the cosmological inference pipeline, the choice of priors, and the measure of posterior differences. We present and discuss the results of both the primary

investigation and additional experiments for relaxed constraints in Section 3.3 and provide our summary in Section 3.4. This work has been peer-reviewed and published in *Physical Review D* (Moews et al., 2019a).

3.1 Data

In order to test the limits of a standard SN Ia cosmological pipeline, we generate a series of mock catalogs, each one corresponding to a universe with a different underlying behavior for the dark energy equation of state parameter. The individual $w(z)$ curves are obtained using a smooth random curve generator described in Section 3.1.1, coupled with physically motivated constraints explained in Section 3.1.2. The generated curves are subsequently fed into into a SN Ia simulation pipeline, based on the statistical properties and redshift distribution of the PANTHEON SN Ia sample (Scolnic et al., 2018). Details on our simulation, the process for which is shown in Figure 3.1, are given in Section 3.1.3.

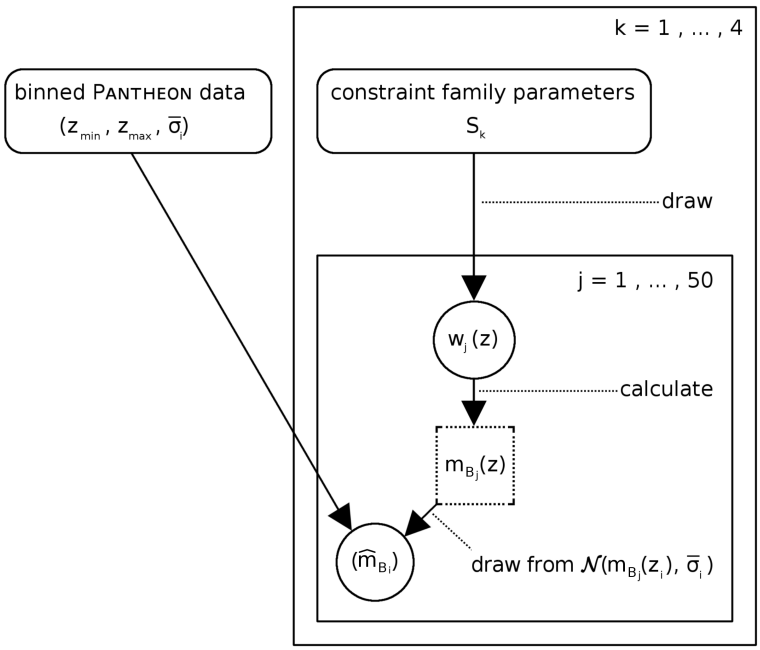

Figure 3.1 *Schematic flowchart of the generation for* PANTHEON*-based SN Ia simulations. Dotted rectangles denote calculated values, whereas rounded rectangles and circles indicate known values and random variables, respectively. Dotted lines mark operations performed at a given point during the process.*

3.1.1 Generating perturbations of ΛCDM

The construction of mock type SN Ia datasets that can mimic universes with varying dark energy equations of state requires the ability to create $w(z)$ realizations under arbitrarily flexible sets of constraints, for example to define vertical intervals and regulate the maximum number of gradient sign changes. To this end, we introduce a general-purpose smooth random curve generator that satisfies the need for extensive constraints, together with an easy-to-handle implementation for the wider research community. While we use this generator to create realizations of $w(z)$, our method is applicable to a wide array of problems in which generic curves are needed. In this context, curve realizations can also be used for function perturbations of arbitrary measurement detail, treating the value at each measurement point as a multiplier for the respective value in a function that is to be smoothly perturbed.

Both node-dependent interpolation approaches and GPs present some significant drawbacks. Linear splines lead to sharp changes in the generated functions, while cubic splines are prone to introducing spurious features. Similarly, GPs require setting a covariance function and, depending on the kernel, may lack smoothness (Rasmussen & Williams, 2005). In addition, the aforementioned methods hamper the ability to easily subject the generated curves to customized sets of constraints.

To overcome such limitations, we introduce and employ `Smurves`, a random smooth curve generator that allows for highly customizable and physically motivated constraints to be placed on the curve-generating process. The source code of the curve generator, as well as a tutorial and examples, can be found in a public code repository[1]. Based on the concept of changes in gravitational direction and magnitude along projectile paths, the generator employs Newtonian projectile motion, adapted to allow for negative values, as the basis for generating curves as its outputs.

Given a set of user-specified constraints, `Smurves` generates smooth curves through uniform-random sampling of the number of changes in gravitational direction and the locations of such changes, while adhering to the specified constraints. The path is segmented at the sampled change points, and uniform-random samples of the gravitational acceleration are drawn within the bounds of possible curve paths, while respecting the set of interval constraints.

[1] `https://github.com/moews/smurves`

Data: v := velocity
α := step size
β := direction
s := partial steps
p_0 := start point
f := vertical force
θ := launch angle
Result: Path p, impact angle θ_{imp}, velocity v
Set the initial horizontal displacement to zero
$\Delta x \longleftarrow 0$
Calculate the horizontal and vertical velocities
$v_x \longleftarrow v\cos(\theta)$
$v_y \longleftarrow v\sin(\theta)$
Initialize start velocity and path measurements
$v_0 \longleftarrow v$
$p \longleftarrow p_0$
Loop over the given x-axis measurement points
for $i \leftarrow 1$ **to** $\underline{\text{length}(s)}$ **do**
    Horizontal distance, displacement and time
    $d \longleftarrow s[i]$
    $\Delta x \longleftarrow \Delta x + \alpha$
    $t \longleftarrow \frac{\Delta x}{v_x}$
    Calculate vertical velocity and displacement
    $v_y \longleftarrow v_0\sin(\theta) - ft$
    $\Delta y \longleftarrow -\left(v_0\sin(\theta)t - \frac{1}{2}ft^2\right)$
    Total velocity and directional displacement
    $v \longleftarrow \sqrt{v_x^2 + v_y^2}$
    $D \longleftarrow \beta\Delta x$
    Append the projectile location at that point
    $p \longleftarrow \text{append}(p, (d, p_0 + D))$
end
Calculate the impact angle for the partial path
$\theta_{\text{imp}} \longleftarrow \arctan(-\frac{v_y}{v_x})$
return $p, \theta_{\text{imp}}, v$

Algorithm 3: Partial trajectory calculation

The segmented path calculation of `Smurves` follows, in its broadest terms, the classical Newtonian calculation of a projectile path: Given a velocity, an acceleration magnitude as a force acting on the projectile, and a launch angle, a flight path can be easily computed as vertical axis values along a set of measurement points on the horizontal axis. At the end of the partial path computation, the function returns the path measurements, the impact angle, and the final velocity of the projectile. Depending on the number of sampled change points, and on whether parts of the full path are not yet calculated, a new force acting in the opposite direction of the previous one is sampled, and previously returned values are re-used as inputs to the same function. This lets the projectile continue its flight with the same characteristics, but with changed gravitational magnitude and direction, to ensure a smooth curve evolution that easily lends itself to subsequent splining.

The corresponding method for curve segment calculations is specified, as pseudocode, in Algorithm 3 to allow for an easier replication and easier understanding both of our approach and the accompanying open-source code implementation for smooth random curve generation.

While we primarily make use of the ability to set intervals and the number of maximum gradient sign changes for this chapter, `Smurves` features a variety of additional options that make it applicable to a wider array of problems. Examples of other capabilities include the use of logarithmic scales and the capacity for perfect convergence in a specified point along the generated curves' paths.

The next section describes the use of `Smurves` to create 50 $w(z)$ curves per constraint family, which imposes boundaries in both dimensions, z and w, on each curve sampled at 500 equally-spaced redshift bins on a linear scale. For brevity, we call each such constraint family generated with `Smurves` a "SmurF".

3.1.2 Constraints on w(z)

We explore families of $w(z)$ curves that evolve within the redshift range covered by the binned Pantheon data, $0.0140 < z < 1.6123$, and that are constrained to regions of allowed constant-w models, with a broadest envelope of $-5/3 < w < -1/3$. The upper bound of $w = -1/3$ is obtained by requiring an accelerated expansion of the Universe at the present time driven by dark energy. For each component i of the Universe, this limit corresponds to $\sum_i (\rho_i + 3p_i) < 0$, defining

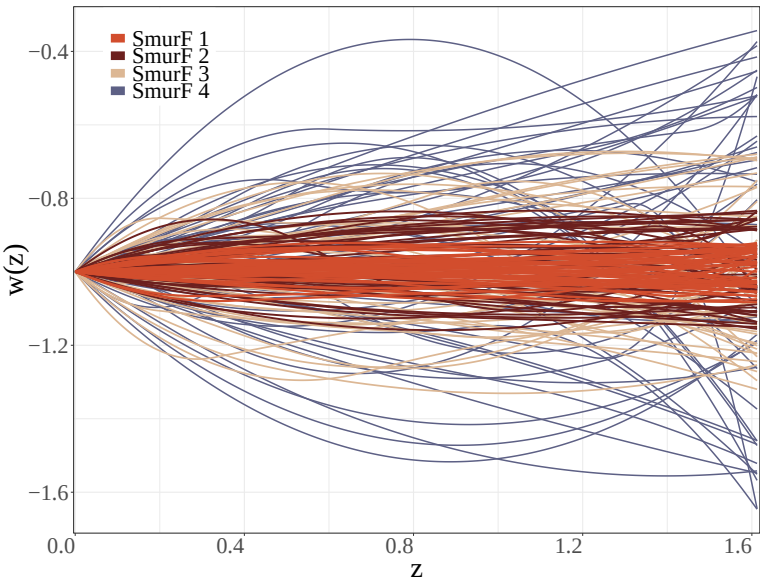

Figure 3.2 *Smooth random $w(z)$ curves generated with* `Smurves` *to create SN Ia mock observations. The figure shows curves from four different constraint families ("SmurFs"), with 50 curves per family, while adhering to a maximum of one gradient sign change for a given curve. The varying parameters are the upper and lower boundaries of $w(z)$ for each family.*

the strong energy condition, with equation of state $w_i \equiv p_i/\rho_i$, pressure p_i, and energy density ρ_i of energy component i (Dodelson, 2003). The limit of $w < -1/3$ corresponds to a cosmological constant that dominates over other constituents. The lower bound on w results from the requirement that a so-called Big Rip scenario cannot have occurred within the age of the Universe of roughly one Hubble time H_0^{-1}. The previous term implies that phantom energy, with $w < -1$, becomes infinite in finite time and overcomes all other forms of energy, ripping apart everything, from cosmic structure to atoms, with the Universe ending in a "Big Rip" (Caldwell et al., 2003). We also note that phantom dark energy violates the null energy condition (Carroll et al., 2003).

While the lowest redshift for the PANTHEON data is $z = 0.0140$, we set another constraint to let all curves start at $z = 0$ so that $w(0) = -1$. This is to agree with near-z cosmological probes bearing small scatter at the lowest redshift bin. The resulting set of constrained $w(z)$ curves, shown in Figure 3.2, exhibits behaviors that can be found, among others, in effective fluid descriptions of $f(R)$ models as described by Arjona et al. (2019), scaling, or interacting, dark matter as in Chevallier & Polarski (2001), and bimetric theories of gravity (Koennig et al., 2014). In terms of what types of $w(z)$ realizations can be modeled to fit with

theoretical models, our approach allows for the coverage of a wide variety of $w(z)$ realizations. For linear realizations, this can be done through large-enough acceleration or small-enough gravitational force to approximate linear behavior, or, alternatively, by constraining the allowed y-axis interval, fixing the number of gravitational sign shifts to zero, or not using left-hand convergence and setting the gravitational pull start point as the right-hand limit, thus disabling curve-like behavior in general. For arbitrarily complex functions, by setting a large limit on the number of sign shifts, any curve can be reasonably approximated by the presented method as a sequence of piece-wise quadratic steps. The one drawback that is shared with related approaches to $w(z)$ curve simulation is that sharp interruptions in a curve that, for example, abruptly change the $w(z)$ value along the redshift evolution, can not be modeled.

In practice, this approach means that we evolve the Friedmann equation while including both matter and dark energy as energy components. For a flat Universe, this implies

$$H(z) = H_0 \left[\Omega_\mathrm{m}(1+z)^3 + \Omega_\Lambda (1+z)^{3(1+w)}\right]^{1/2}, \tag{3.1}$$

where Ω_m and Ω_Λ represent the dark matter and dark energy density parameters, respectively. For a flat Universe, we note that $\Omega_\Lambda = 1 - \Omega_\mathrm{m}$. The current age $t > H_0^{-1}$ of the Universe sets a lower limit on w for a given Ω_m. The more negative a phantom component ($w < -1$) is, the faster we reach a Big Rip scenario. A lower boundary of $w \gtrsim -2$ corresponds to $\Omega_\mathrm{m} = 0.6$, while, for example, $\Omega_\mathrm{m} = 0.8$ leads to the requirement $w \gtrsim -2.2$, and $\Omega_\mathrm{m} = 0.01$ yields $w \gtrsim -5/3$. Therefore, we constrain our broadest envelope of $w(z)$ curves to a lower limit of $w = -5/3$, conservatively corresponding to a very low matter density and yielding symmetric intervals for the curve limits.

For the three remaining SmurFs, we halve the preceding symmetric interval around $w = -1$ for each new family, shrinking the allowed envelopes each time to let curves generated from the corresponding families stay closer to the value of the ΛCDM model. As a result, we generate four curve families with increasing maximum and average deviations from the ΛCDM model to investigate the degree of compliance for different degrees of compliance with $w(z) = -1$.

We put a final constraint on the curve generator, specifying a maximum number of one for gradient sign changes in the created curves to keep our $w(z)$ curves in

line with shapes in related research, but explore an increased maximum number of gradient sign changes, as well as the effect of an omission of the $w(z) = 0$ constraint, later in Section 3.3.2.

3.1.3 SN Ia data simulation

Observations sensitive to the background expansion such as SN Ia data can be employed to measure the luminosity distance,

$$d_{\mathrm{L}}(z) = (1+z)\, d_{\mathrm{H}} \int_0^z \frac{\mathrm{d}z'}{E(z')}, \tag{3.2}$$

where the Hubble distance is $d_{\mathrm{H}} = c/H_0$ and the Hubble parameter is $E(z) = H(z)/H_0$, with $H(z)$ given by Equation 3.1. This is related to the peak B-band magnitude,

$$m_{\mathrm{B}i} = 5 \log_{10} d_{\mathrm{L}}(z_i) + M, \tag{3.3}$$

of a given supernova i at redshift z_i, with absolute magnitude M. We generate SN Ia peak B-band magnitude catalogs by inserting each $w(z)$ curve seen in Figure 3.2 into Equation 3.1 and following the process shown in Figure 3.1.

Our mock data are constructed to mimic the statistical properties and redshift distribution of the PANTHEON SN Ia sample[2], which consists of a total of 1048 SN Ia at redshifts $0.03 < z < 2.3$, representing the largest combined sample of SN Ia observations to date (Scolnic et al., 2018). We use the publicly available catalog, which is summarized by 40 redshift bins from $z_1 = 0.0140$ to $z_{40} = 1.6123$. We note that differences in w between the binned and unbinned versions are smaller than $(1/16)\,\sigma$ for statistical measurements Scolnic et al. (see 2018), which makes this an adequate and easy-to-handle data representation for a large number of analysis pipeline runs.

We propagate the curves through a simulation pipeline using `CosmoSIS`, as described in Section 3.2.1. The simulation pipeline also takes into account the full covariance matrix, which includes effects due to photometric error, the uncertainty in the mass step correction, uncertainty from peculiar velocity and

[2] `https://archive.stsci.edu/prepds/ps1cosmo/index.html`

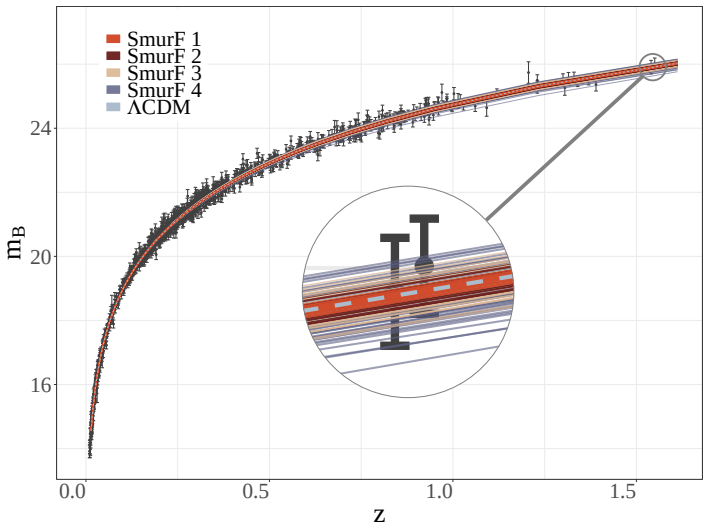

Figure 3.3 *Peak B-band magnitudes m_B as a function of redshift z for different dark energy equation of state ($w(z)$) realizations. The figure shows the diagrams for the ΛCDM model (dashed line), as well as 50 random $w(z)$ curves for each of the four constraint families, which represent increasing deviations from ΛCDM. Black points depict the* Pantheon *dataset and respective uncertainties, and the insets highlight $w(z)$ models regarding ΛCDM as mostly falling within the data uncertainty, even at redshifts as high as $z \gtrsim 1.5$.*

redshift measurement, distance bias correction, and uncertainty from stochastic lensing and intrinsic scatter. Peak B-band magnitudes for $w(z)$ curves are shown in Figure 3.3 to demonstrate the similarity of results even at high redshifts.

3.2 Methods

We run a full analysis pipeline that assumes a constant-w dark energy model, hereafter called $\Psi_{w_{\rm const}}$, to infer the posterior probability distribution of w, $\Omega_{\rm m}$, and M as described in Section 3.2.1. In Section 3.2.2, we list and justify our choice of priors for parameters. Finally, in Section 3.2.3, we introduce the metric by which we compare simulation-based posteriors and those from real SN Ia Pantheon data.

3.2.1 Pipeline with `CosmoSIS`

`CosmoSIS` is a cosmological parameter estimation code by Zuntz et al. (2015), which models cosmological likelihoods and calculations as a sequence of independent modules that read and write their inputs and outputs to a central data storage block. The package has been used extensively for parameter estimation by the Dark Energy Survey (DES) (see, for example, Abbott et al., 2018a; Abbott et al., 2019a; Abbott et al., 2018c; Elvin-Poole et al., 2018; Troxel et al., 2018), among others (Barreira et al., 2015a; Harrison et al., 2016; Krause & Eifler, 2017; Lin & Ishak, 2017).

We utilize two `CosmoSIS` pipelines; the first simulates data using the $w(z)$ realizations described above, and the second analyzes the simulated data using the `emcee` sampler, as described by Goodman & Weare (2010) and Foreman-Mackey et al. (2013), under a standard cosmological model. This approach extends the classic Metropolis-Hastings algorithm with a parallel "stretch move".

A number K of walkers explore the parameter space, with their respective steps drawn from a proposal distribution that depends on other walkers' positions. A walker at position Y is drawn by chance to propose a new position X' for the walker that is to be updated and currently at position X, meaning that

$$X \rightarrow X' = Y + Z[X - Y]. \tag{3.4}$$

Here, Z acts as a random variable with $S \coloneqq [0.5, 2]$ and $Z \sim g(z) \propto \mathbf{1}_S(z) \cdot \sqrt{z}^{-1}$, with the indicator function $\mathbf{1}_S(z)$ taking a value of one for all $z \in S$ and a value of zero for all $z \notin S$. Alternatively, this can be written as

$$g(z) \propto \begin{cases} \frac{1}{\sqrt{z}} & \text{if } z \in \left[\frac{1}{2}, 2\right] \\ 0 & \text{otherwise} \end{cases}. \tag{3.5}$$

The "parallel stretch" mentioned above splits the K walkers into two equal-sized subsets and updates all walkers of one subset using the other, followed by the corresponding opposite procedure, which allows for the parallelization of this computationally expensive update step.

An affine-invariant MCMC algorithm satisfies $X_a(t) = AX_b(t) + b$ for different

starting points X_a and X_b, and two probability densities π and $\pi_{A,b}$, for any affine transformation $Ax+b$. The independence of the aspect ratio in highly anisotropic distributions offers a speed advantage in highly skewed distributions.

We connect these two pipelines in a script to iterate the process over the curves from each SmurF using four standard library modules: `consistency`, which computes the complete set of cosmological parameters, `CAMB` as described by Lewis et al. (2000), which, in our case, calculates cosmological background functions, and `pantheon`, which computes the PANTHEON likelihood. A custom module is used to read in tabulated $w(z)$ functions and cast them to the form used in `CAMB`.

For Gaussian likelihoods, `CosmoSIS` automatically generates simulated outputs incorporating both the signal based on the used model and noise. The tool creates simulations of peak B-band magnitudes as $m_{\mathrm{B}}(z_i)$ double arrays based on binned PANTHEON SN Ia data. From the PANTHEON noise covariance $C \equiv \langle nn^T \rangle$, we can generate this simulation using its (unique) Cholesky decomposition $C = LL^T$ and a random vector r, where each element is a random normal value with $r_i \sim N(0,1)$. We can then form $n = L \cdot r$ as our noise simulation, as the noise covariance is then

$$\langle nn^T \rangle = \langle Lrr^T L^T \rangle = \langle LL^T \rangle = C. \tag{3.6}$$

Accordingly, the total simulated values $m_{\mathrm{B_{sim}}}$ obtained through `CosmoSIS` are

$$m_{\mathrm{B_{sim}}} = m_{\mathrm{B_{truth}}}(z_i) + L \cdot r, \tag{3.7}$$

for true values $m_{\mathrm{B_{truth}}}$. Initial experiments to compare the original Pantheon data with SN Ia data generated using flat $w(z)$ curves as a null test uncovered a bug in `CosmoSIS`. After this was reported and subsequently fixed, the flat-curve simulations of SN Ia peak B-band magnitudes returned to expected values of m_{B}. Employing the reported uncertainties on m_{B} and the full covariance matrix, we use this process to simulate peak B-band magnitudes at the same redshift values as reported for the real data in the binned PANTHEON sample. The distributions of these mock peak B-band magnitudes are provided in Figure 3.4.

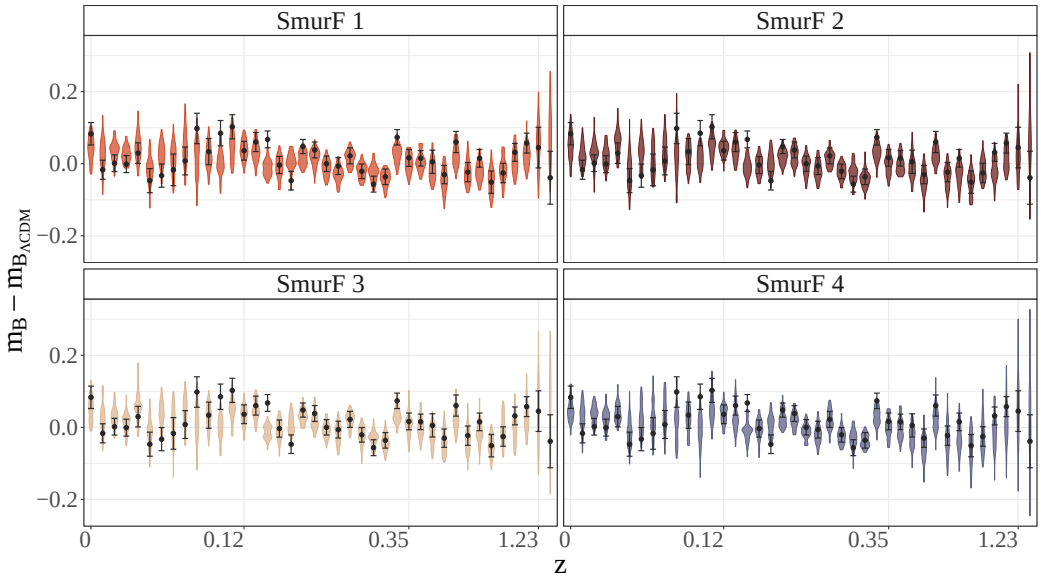

Figure 3.4 *Visualization of peak B-band magnitude (m_B) residuals between our simulated data and ΛCDM, as well as between observed* PANTHEON *data and the ΛCDM model. In both cases, ΛCDM corresponds to* $\Omega_m = 0.307$ *and* $M = -19.255$. *The violin plots for each of the 40 redshift (z) bins show a rotated kernel density plot of the distributions of values for each of 50 different realizations for one SmurF per panel. Black dots indicate binned* PANTHEON *data, with vertical black lines representing the error bars of one standard deviation. The comparison is plotted as the difference between the respective peak B-band magnitudes and expected ΛCDM values,* $m_B - m_{B\Lambda CDM}$, *to show both the deviation from theoretical values and the distributions of simulated SN Ia data around observed values.*

3.2.2 Choice of priors

We vary our cosmology via the present-day matter density Ω_m and the dark energy equation of state w. We assume a flat Universe with $\Omega_k = 0$ and, therefore, a dark energy density of $\Omega_\Lambda = 1 - \Omega_m$. We keep the present-day Hubble parameter fixed to $h_0 = 0.7324$ Riess et al. (see 2016), and the cosmic baryon density to $\Omega_b = 0.04$ (Cooke et al., 2014). An additional nuisance parameter is the absolute magnitude of SN Ia M, which is degenerate with the Hubble parameter. This chapter addresses the question of how sensitive commonly used λCDM analysis pipelines are to non-standard $w(z)$ deviations found in alternative models such as the Chevallier-Polarski-Linder (CPL) parameterization (Copeland et al., 2018). For this reason, the focus here lies on an investigation of such a standard analysis pipeline. We do, however, recommend experiments that use non-λCDM analyses

for follow-up research, in order to check the sensitibity of cosmological analyses in a more general framework.

Table 3.1 *Priors for the estimation of cosmological and nuisance parameters. $U(\cdot)$ denotes a uniform distribution, whereas we use "fixed" to indicate a Dirac delta function with $\delta(x) = \infty$ for an x from the column of initial values.*

Parameter	Prior	Initial value
Ω_m	U(0.01, 0.6)	0.307
M	U(-20.0, -18.0)	-19.255
w	U(-2.0, -0.3333)	-1.026
Ω_k	fixed	0
Ω_b	fixed	0.04
h_0	fixed	0.7324

Our set of estimated parameters from the `emcee` sampler is $\{\Omega_\mathrm{m}, w, M\}$. We choose uniform priors for all parameters, with bounds given in Table 3.1. The range for the absolute magnitude M encompasses previous constraints given, for example, by the SDSS-II/SNLS3 Joint Light-Curve Analysis (JLA) (Betoule et al., 2014). The central starting value of $M = -19.255$ is chosen from a preliminary maximum likelihood run with Pantheon data. The prior over Ω_m covers allowed parameter ranges as estimated by present-day SN Ia samples like JLA and Pantheon. The starting point for the dark matter parameter is $\Omega_\mathrm{m} = 0.307$, which corresponds to the Pantheon wCDM best-fit value. Analogously, the central value for w is set to $w = -1.026$ (Scolnic et al., 2018).

The prior range on w coincides with the allowed values for the families of $w(z)$ curves considering the prior upper bound of $\Omega_\mathrm{m} = 0.6$ (see Section 3.1.2 for a detailed description of the allowed w-interval). For our parameter estimation, we loosen the symmetric lower-bound requirement, with $w = -2$ as our lower limit to cover the allowed upper boundary of Ω_m from SN Ia at $3\,\sigma$.

3.2.3 Comparison criteria

Conventional error contours, used ubiquitously in cosmology, are estimated from samples from posterior probability distributions $p(\theta|D, \Psi)$ of parameters of interest, in our case $\theta = \{\Omega_\mathrm{m}, w, M\}$, conditioned on the cosmological model Ψ and data $D = \{d_i\}_N$, where i runs over the number N of observations. For Pantheon, the data is presented as $D_\mathrm{Pantheon} = \{z_i, m_{\mathrm{B}i}, \sigma_{m_\mathrm{B},i}\}_{40}$ for bins i.

We consider each individual $w(z)$ curve separately, but group them by constraint family S_k, as depicted in Figure 3.1, for interpretability. Given the way in which posteriors from $w(z)$ curve realizations from the same constraint family are used in this chapter, one might ask why posterior samples obtained from instances of the same SmurF are not simply combined to arrive at a posterior for the constraint family. Considering error contours as being comprised of samples from $p(\theta|D, \Psi)$, as introduced in Section 3.2.3, neglects the role of the initial conditions C_0 that have been implicitly marginalized out as

$$p(\theta|D, \Psi) = \int p(C_0, \theta|D, \Psi) dC_0. \tag{3.8}$$

Since we generally cannot constrain the initial conditions as such, an obvious question to ask is why they matter.

When combining constraints on cosmological parameters from different probes D and D', we are really asking for $p(\theta|D, D', \Psi)$ when we have $p(\theta|D, \Psi)$ and $p(\theta|D', \Psi)$. To make use of the independence of the datasets, we would expand this in terms of Bayes' Rule as

$$\begin{aligned} p(\theta|D, D', \Psi) &= \int p(C_0, \theta|D, D', \Psi) dC_0 & (3.9) \\ &= \int p(D, D'|C_0, \theta, \Psi) \frac{p(C_0, \theta|\Psi)}{p(D, D'|\Psi)} dC_0. & (3.10) \end{aligned}$$

If D and D' are our standard independent probes, every term in Equation 3.9 is well-defined. This means that the integral is separable and we can recover the intuitive way to combine the posteriors.

The situation investigated in this chapter, however, is different. In our case, D and D' correspond to different SmurF instances j and j'. These two datasets are inherently contradictory; they could never be observed in the same instantiation of the universe, even under the same physical model and values of the cosmological parameters θ. In other words, $p(D, D'|C_0) = 0$ for any pair of mock-Pantheon data we consider. What distinguishes one SmurF from another is rolled into the initial conditions C_0, leading to well-defined $p(\theta|D, \Psi)$ and $p(\theta|D', \Psi)$, but to an internally inconsistent $p(\theta|D, D', \Psi)$. Thus, it would be inappropriate to combine samples of the cosmological parameters obtained through a Markov chain Monte Carlo (MCMC) method from any collection of SmurF instances with different

$w(z)$ curves, divided by constraint family or not.

For $j \in \{1, 2, \ldots, 50\}$, each of 50 simulated data sets D_j is generated with the curve $w_j(z)$, and our experimental design yields samples from the posteriors $p_j \equiv p(\theta|D_j, \Psi_{w_{\rm const}})$. Each posterior corresponds to the probability of parameters from a cosmological model $\Psi_{w_{\rm const}}$ conditioned on the data generated from $w_j(z)$. We also apply the same pipeline to 50 realizations of the data under the ΛCDM model, producing $p_{\Lambda_j} \equiv p(\theta|D_{\Lambda{\rm CDM}_j}, \Psi_{w_{\rm const}})$, and to the real Pantheon data, producing $p_{\rm Pantheon} \equiv p(\theta|D_{\rm Pantheon}, \Psi_{w_{\rm const}})$.

To compare the samples from each mock universe to their ΛCDM counterparts, we adopt a measure suited to quantifying the difference between probability distributions. The Kullback–Leibler divergence ($D_{\rm KL}$) introduced by Kullback & Leibler (1951),

$$D_{\rm KL} = \int_{-\infty}^{\infty} p(x) \ln \left[\frac{p(x)}{\hat{p}(x)} \right] {\rm d}x, \tag{3.11}$$

is the directional difference between a reference probability distribution $p(x)$ and a proposed approximating probability distribution $\hat{p}(x)$. The $D_{\rm KL}$ has been applied within astronomy only to a limited extent, but is gaining popularity (Kilbinger et al., 2010; Ben-David et al., 2015; De Souza et al., 2017; Hee et al., 2017; Malz et al., 2018; Nicola et al., 2019).

Unlike symmetric measures of the distance between two probability distributions, such as the familiar root-mean-square-error, the $D_{\rm KL}$ is defined as the directional loss of information due to using an approximation in place of the truth; we must designate one distribution as a reference from which the proposal distribution diverges. A generic example of a pair of reference and proposal distributions can be defined by posterior samples derived from a large set of observations, as opposed to posterior samples derived from a small subset thereof. There is, therefore, an implicit assumption that the former is closer to the truth than the latter, which may be an approximation when the rest of the observations are unavailable.

In our case, the samples from $p_{\rm Pantheon}$ always serve as the reference distribution, and the samples from p_j and p_{Λ_j} always act as the proposal distribution.

3.3 Results and Discussion

In the previous sections, we describe both the data and our methodology. In Section 3.3.1, we present the results of primary experiments, together with a discussion of the underlying causes and implications for SN Ia investigations. In addition, we relax the different constraints for two of the constraints families in Section 3.3.2 to explore the impact such changes have on the resulting D_{KL} distributions. In the first of these two additional experiments, we generate $w(z)$ curves with an increased maximum number of gradient sign shifts, whereas the second experiment eliminates the requirement that $w(z) = -1$.

3.3.1 Primary experiments

For each SmurF, as described in Section 3.1.1, we generate 50 $w(z)$ curves that are fed into the `CosmoSIS` simulation and analysis pipeline described in Section 3.2.1. This results in four sets of 50 posterior distributions for parameters $\{\Omega_{\mathrm{m}}, w, M\}$, or $p_{S_k,j}$, where $k \in \{1, 2, 3, 4\}$ identifies the SmurF and $j \in \{1, 2, \ldots, 50\}$ denotes its realizations (see Section 3.2.3 for details on notation). In addition, 50 datasets from a ΛCDM model are generated to illustrate the impact allowed by current statistics and systematic uncertainties. We feed these simulations, as well as the original binned Pantheon dataset, into the same analysis pipeline. Posteriors derived from all simulated data are then compared to the Pantheon results using the Kullback-Leibler divergence D_{KL}, described in Section 3.2.3.

Figure 3.5 shows histograms of D_{KL} values for each SmurF along with those from ΛCDM simulations. In accordance with our expectations, the distributions of D_{KL} values for constraint families with increasingly wider w-intervals, from SmurF 1 through 4, show a systematic shift toward higher means, larger variances, and multimodality. These differences are, however, small enough that the bulk of D_{KL} values for each SmurF coincides with the D_{KL} range covered by the ΛCDM case, presenting a serious obstacle for the detection of deviations from a cosmological constant. For the goodness of fit of posteriors, further challenges arise from a statistical degeneracy, which we describe below and in Fig. 3.6.

This effect is better visualized by a representative $w(z)$ function for each SmurF and the respective posteriors, shown in each column of Figure 3.6. The top row shows $w(z)$ curve associated with the median D_{KL} value for each SmurF,

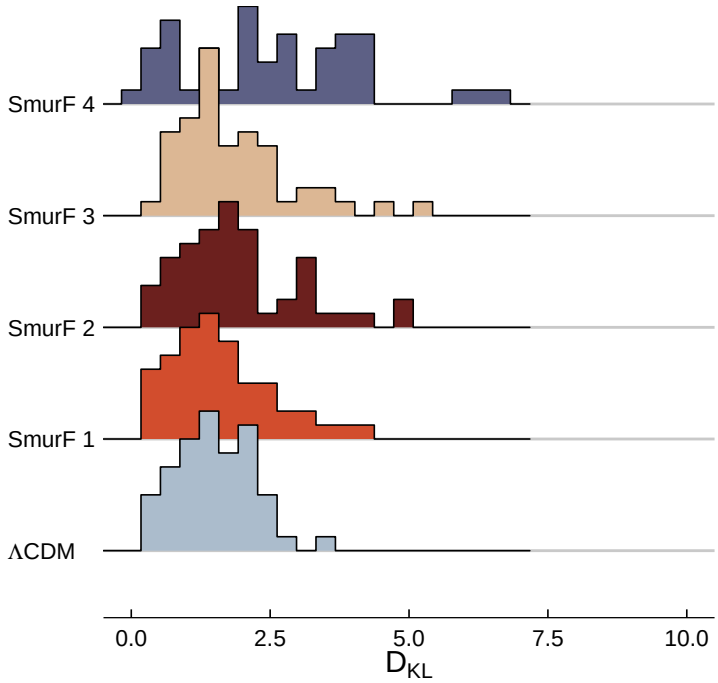

Figure 3.5 *Histograms of the Kullback-Leibler divergence (D_{KL}) for different sets of constraints. The shown histograms depict the distribution of D_{KL} values for the ΛCDM case and each SmurF used to generate simulated SN Ia peak B-band magnitudes. D_{KL} values are calculated for the posterior distributions of parameters obtained through a standard ΛCDM analysis pipeline that considers only constant w models.*

as well as the constant $w = -1$ line. In doing so, we enable the comparison of single representative curves, which we also visualize to ensure that deviations from $w(z) = -1$ in representatives follow the same progression toward larger deviations as the increasing deviations distinguishing different SmurFs.

Each curve approximately covers the allowed w intervals of its respective constraint family, thus confirming the applicability of a median-D_{KL} approach for choosing a representative SmurF instance. The bottom three rows show two-dimensional posterior distributions, for parameters $\{\Omega_{\mathrm{m}}, w, M\}$, for each SmurF and the ΛCDM case (colored contours) superimposed on the posteriors from Pantheon data (black contours). Similarly, posterior distributions from the ΛCDM model, together with SmurFs 1, 3, and 4, go from agreement to disagreement with Pantheon. Posteriors from SmurF 2, on the other hand, show an unexpected visual match with both real Pantheon data results and the ΛCDM case, despite its associated $w(z)$ exhibiting larger deviations from $w = -1$ than the one associated with SmurF 1. Notably, the representative

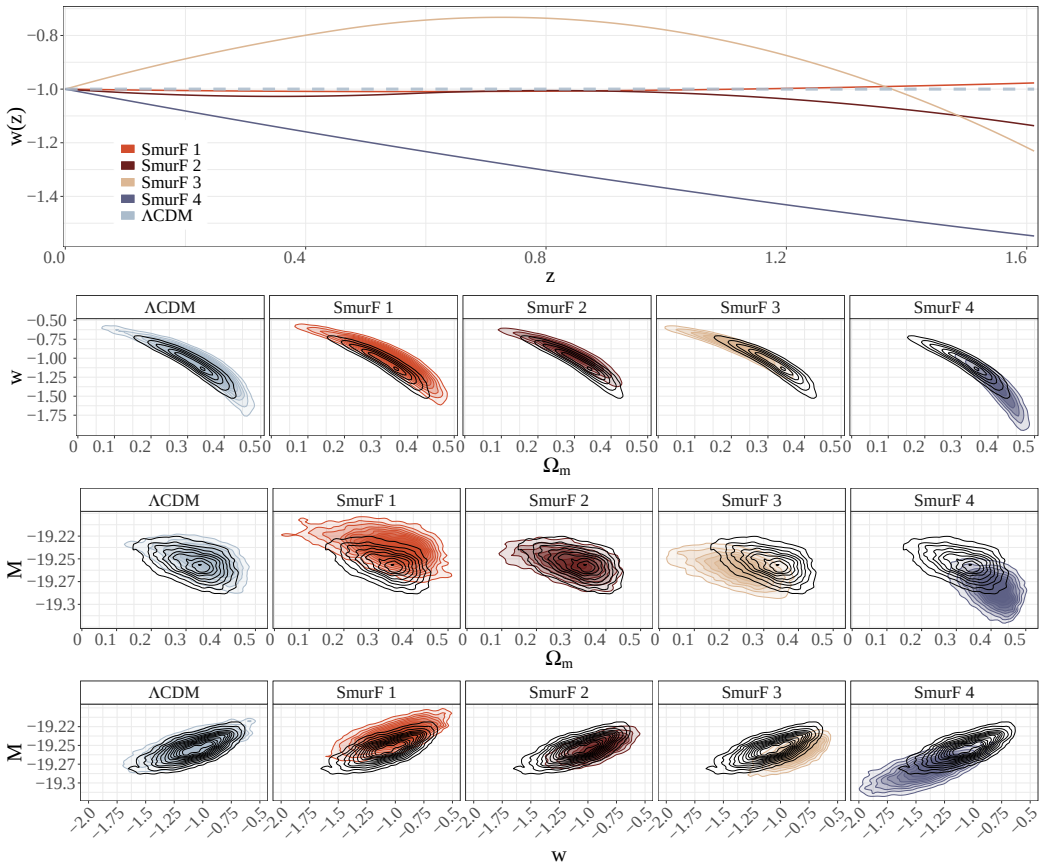

Figure 3.6 *First row: Representative redshift-dependent dark energy equation of state ($w(z)$) curves associated with the median D_{KL} per constraint family (full lines) and the ΛCDM case (dashed line). Second row: Posteriors for w and dark matter density Ω_{m} per constraint family. The four plots depict the posterior distributions for the above-mentioned curves (colored contours), as well as the posteriors for the* Pantheon *analysis case (black contours). Third and fourth row: With M as the absolute magnitude, the plots show two-dimensional posteriors for $M \times \Omega_{\mathrm{m}}$ and $M \times w$, respectively. The cause of the comparatively good posterior fit of SmurF 2 is discussed in the text.*

curve from SmurF 2 features larger deviations from the ΛCDM case than the representative curve from SmurF 1 in both low-z and high-z regimens, meaning that larger deviations from the ΛCDM case do not necessarily result in posteriors considerably different from the ones produces by $w(z) = -1$. This is an important finding, which requires a discussion in the following part.

The apparent discrepancy between notable inconsistencies in $w(z)$ and compliant posterior estimates derives from the fact that, while $w(z)$ can change widely, the observable signature of $w(z)$ relies on the peak B-band magnitude m_{B}. The dependence of m_{B} on the integral of the Hubble parameter leads to a statistical

degeneracy that makes such posteriors indistinguishable from ΛCDM within the current magnitude precision level and probed redshift range. Coupled with the large D_{KL} overlap between SmurF instances and ΛCDM results seen in Figure 3.5, this directly extends to a considerable chance of mistaking an equation of state varying significantly with redshift for one in reasonable agreement with a cosmological constant. Put simply, this means that even comparatively compliant $w(z)$ curves can lead to higher differences in posteriors as seen in Fig. 3.6. Follow-up research on such effects for alternatives to the ΛCDM model such as the CPL parameterization should also check for similar discrepancies.

A more detailed view of all posteriors over w is shown in the ridgeline plots of Figure 3.7, in which the means, as well as the bulk of the probability, fall within the 95% credible intervals of the PANTHEON results under a constant-w hypothesis. SmurF 2, in particular, shows more constrained posteriors, which offers an explanation for the agreement of the median-D_{KL} representative's posterior with the ΛCDM case. It does, however, also feature four obvious outliers reaching far beyond the left boundary of the credible interval, which demonstrates the variability in the agreement of w-posteriors within the same constraint family.

Naturally, all of the the aforementioned results are bounded by the PANTHEON-like quality of our simulations. Current surveys such as DES continue to contribute to the number of SN Ia observations (Abbott et al., 2018a). Though the DES SN Ia samples used in combination with additional external samples amount to less than a third of PANTHEON's sample size, DES results indicate smaller intrinsic scatter in the Hubble diagram, taking one step further in the attainment of higher-quality SN Ia samples (Brout et al., 2019). These new and future datasets will certainly increase our ability to discriminate between different models for the dark energy equation of state parameter.

It is, however, important to highlight the non-intuitive and unavoidable behavior derived from the nature of distance measurements as an integral over the Hubble parameter. Given a dataset with sufficiently low measurement and systematic uncertainties, especially at high redshifts, discrimination between phenomenologically close models is possible, but we cannot rely on the assumption that substantial redshift-dependent changes in $w(z)$ will necessarily result in detectable biases under a constant-w analysis. This is especially the case for SN Ia-only analyses (Miranda & Dvorkin, 2018; Zhao et al., 2017; Abbott et al., 2019b; L'Huillier et al., 2019).

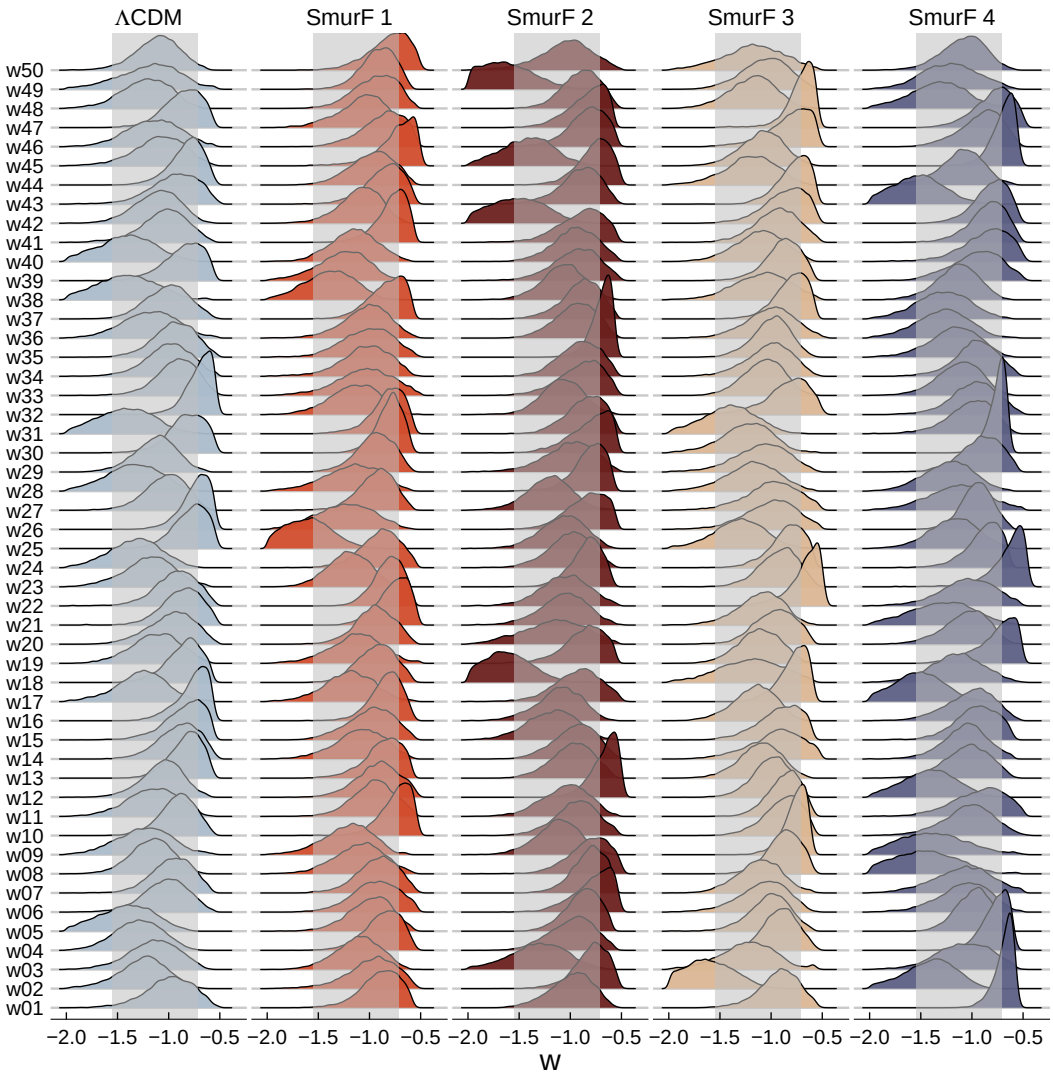

Figure 3.7 *Ridgeline plots for the dark energy equation of state parameter w. Each row depicts the posterior densities of w for all 50 curves, for each of the four constraint families as well as the simulations for the ΛCDM case. The transparent bands covering the middle section of each column show the 95% credible interval for the* Pantheon *sample, analyzed under a constant-w model.*

Caution should be exercised in using other cosmological observables to break the degeneracy via constraining additional parameters. This strategy is wide-spread in the literature, to the point that recent research questions the use of SN Ia data without such additional observables (Solà Peracaula et al., 2019). It is, however, important to keep in mind that supernovae are the primary dynamical observable that probe the line of sight directly, and consequently impose boundaries in the behavior of w. The use of additional probes such as weak lensing can, with insufficient information on the baryonic physics involved, introduce new biases, for example in the CPL parameterization (Copeland et al., 2018).

In summary, we recognize the need to combine complementary observables, for example baryon acoustic oscillations and CMB data, while making use of careful statistical analyses capable of probing more subtle behaviors of the dynamical evolution of dark energy. Although paramount for a more general discussion of this topic, the addition of extra observables exceeds the scope of this thesis.

3.3.2 Relaxed constraints on w(z)

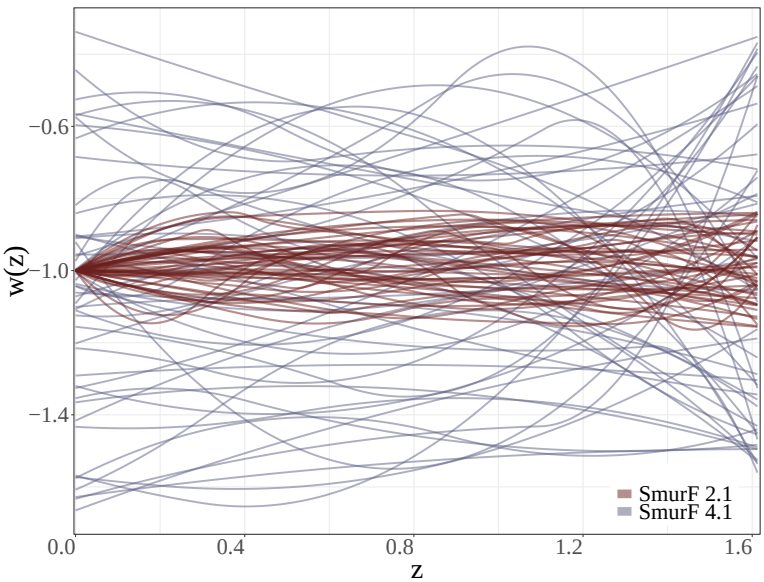

Figure 3.8 *Smooth random dark energy equation of state ($w(z)$) curves generated with* `Smurves` *to create mock SN Ia observations for additional experiments. The figure shows curves from two different constraint families, SmurF 2.1 and SmurF 4.1, with 50 curve realizations per family.*

In a bid to push our analysis a bit further, we relax the constraints put on the

curve generator for SmurFs 2 and 4 for illustrative purposes. For SmurF 2, we increase the maximum number of gradient sign changes from one to 10, allowing for more complicated functions to be realized. In contrast, for SmurF 4, we omit the requirement that $w(0) = -1$ to allow curves to start at arbitrary values within the allowed $w(z)$ interval. The respective curves used in these additional experiments are depicted in Figure 3.8.

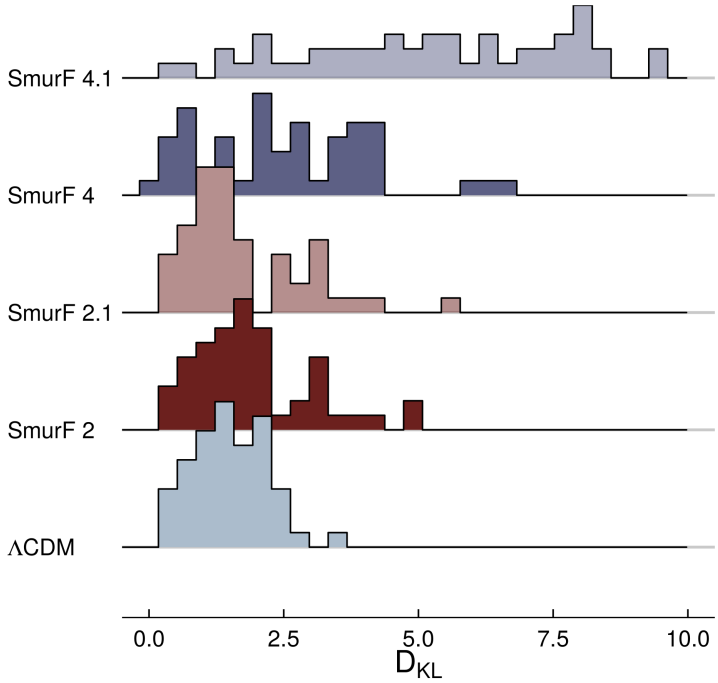

Figure 3.9 *Histograms of the Kullback-Leibler divergence (D_{KL}) for different constraint families. The histograms show the distributions of D_{KL} values, with a total of 50 redshift-dependent dark energy of state curves $w(z)$ per family. In doing so, this figure facilitates the comparison of two previous constraint families, SmurF 2 and SmurF 4, with further relaxed constraint families, namely SmurF 2.1 and SmurF 4.1, as well as with the ΛCDM case.*

To assess the impact of these further constraint relaxations, their D_{KL} distributions are shown in Figure 3.9, along with those from SmurF 2, SmurF 4, and the ΛCDM case. The D_{KL} distribution of SmurF 2.1 still holds the same overall shape of SmurF 2 and occupies a range of D_{KL} values between those covered by SmurF 2 and 4. This demonstrates that the use of more complicated functions, for example the larger maximum number of gradient sign changes in SmurF 2.1, has a lesser impact than simpler functions allowed to vary in a larger interval, as is the case for SmurF 4.1, when constrained to the same $w(z)$ intervals and initial conditions. The complexity of $w(z)$ curves does, as a result, seem to have less of

an effect on distinguishability than the intervals in which they live. This is, again, a consequence of the dependence of m_B on the integral over the Hubble parameter, meaning that faster variations in $w(z)$ tend to be smoothed out observationally. Residual additional variations, which are still present, lead to the slightly higher spread in the corresponding D_{KL} distribution.

When we omit the $w(0) = -1$ constraint, which restricts generated $w(z)$ curves to exhibit stark variations from the ΛCDM case at very low redshifts, we find ourselves confronted with a very different result. Relative to SmurF 4, SmurF 4.1 exhibits larger D_{KL} values with a considerably wider spread. We also note that the distribution of D_{KL} values is much flatter than for distributions constrained to $w(0) = -1$, without a peak at low D_{KL} values. This wider spread and flattened distribution can be attributed to introducing an offset in our observable m_B, since m_B averages over $w(z)$ via the Hubble parameter. Curves like those in SmurF 4.1 can, for example, always lie above or below -1, with an additional offset of varying magnitude depending on its $w(0)$ value, leading to a posterior very different from the ΛCDM case. Intuitively, choosing random $w(0)$ anchoring points leads to a roughly flat distribution of D_{KL} values until reaching a maximal possible deviation from ΛCDM that depends on our allowed $w(0)$ prior range.

3.4 Summary

Searching for new physics beyond the standard ΛCDM model inherently requires the capability to discriminate between competing models for the dark energy equation of state. This work scrutinizes the pitfalls of standard cosmological analysis pipelines in their ability to detect signals of ΛCDM deviations. For this task, we introduce a novel smooth random curve generator, `Smurves`, which uses random sampling and modified Newtonian projectile motion as the means for its generative process. This method is highly customizable and facilitates the use of physically motivated constraints into the curve-generating process. While applied to a specific cosmological case in this chapter, `Smurves` represents a general multi-purpose methodology for constrained curve generation and function perturbation. We also provide a user-friendly implementation of the code for the sake of reproducible science.

We employ `Smurves` to generate mock SN Ia observations representing four constraint families, or SmurFs, each one representing increasing degrees of

deviation from the ΛCDM model. Making use of 50 random $w(z)$ curves per SmurF, we run a Bayesian cosmological inference pipeline for each curve to subsequently produce 200 joint posteriors of Ω_m, w, and M. We then compare these posteriors to those from an analysis of the PANTHEON sample derived under the assumption of a constant-w model.

We show that SN Ia cosmology observables under extensive redshift dependencies of the dark energy equation of state are virtually indistinguishable from those of ΛCDM models using current state-of-the-art analysis pipelines. Notably, $w(z)$ realizations that exhibit a stronger deviation from $w = -1$ can lead to posterior samples of Ω_m, w, and M exhibiting a slightly better agreement with ΛCDM than realizations with lesser levels of deviation. This result highlights a fundamental and generally unstated caveat underpinning the current methodology used to estimate w from SN Ia observations: If ΛCDM is assumed as the null hypothesis in a test for compatibility with observational SN Ia data, the inability to rule out the standard model could, in a given case, be based on such similarities in posteriors with potentially large underlying deviations due to statistical degeneracies.

In addition, we test the effect of both an increased number of gradient sign changes, leading to more complex curves, and of larger deviations from $w(z) = -1$ with the omission of an anchor point of $w(0) = -1$ for generated curves. While the complexity of curves has little impact on the compliance with the standard model, we find that this omission of an anchor constraint at $z = 0$ reduces ΛCDM compliance considerably. We recommend further research on the topic, specifically in terms of an investigation focused on different curve characteristics to reduce the set of viable candidate hypotheses. In doing so, further insights into the specific features of redshift-dependent dark energy equations of state by identifying regions of $w(z)$ parameterizations that favor certain cosmologies.

The upcoming arrival of larger and higher-quality data sets, especially at high redshifts, will certainly improve our capability to distinguish between dark energy models. There are, however, intrinsic characteristics of distance-based observables that can render the identification of strong deviations unattainable. The application of redshift-dependent analyses, parametric or non-parametric, alongside the constant-w scenario and the careful use of additional cosmological observables, are crucial steps in providing a realistic picture of our current knowledge regarding properties of dark energy. Due to these caveats, and given the significant loss in precision when redshift-dependence is taken into account, physics beyond the standard model may be hidden in plain sight.

Chapter 4

Ridges in the Dark Energy Survey for cosmic trough identification

In this chapter, we propose an algorithm to detect 2D density ridges as a way to denoise cosmic structure in mass density maps. For this purpose, we extend the subspace-constrained mean shift (SCMS) algorithm introduced by Ozertem & Erdogmus (2011) to fit our application case, and apply our method to the DES Year 1 data release (Flaugher et al., 2015). The general methodology works by defining a denoised and sparse representation of the filamentary structure, and, as a consequence, the locations of regions of emptiness emerge naturally. Considering the limitations of spectroscopic surveys discussed in Section 1.1.5, trough finders can be applied as an alternative to void finders to recover and study underdense regions from weak lensing studies, giving particular relevance to the present application to DES Year 1 data. We incorporate the haversine distance, a more suitable approach for spherical surfaces, and also customise the mesh size for ridge estimation and optimization of the bandwidth. The density ridges we recover are extracted from the projected matter distribution from weak lensing data. We compare our ridges to a search based on curvelets, an extension of the wavelet transform for filamentary structures (see Candès et al., 2006), as well as to foreground matter density fields derived from luminous red galaxies.

This chapter is organized as follows. Section 4.1 explains the SCMS algorithm and implemented extensions; it includes our kernel density estimation technique, experimental data, and simulations. Section 4.2 describes our experimental results and compares them to a curvelet-based denoising technique. Section 4.3

comments on our approach and future directions, with an emphasis on the advantages and drawbacks of the proposed methodology. Finally, Section 4.4 provides our summary. This work has been peer-reviewed and published in *Monthly Notices of the Royal Astronomical Society* (Moews et al., 2020). It extends and applies prior methodological developments published in *Decision Support Systems* (Moews et al., 2021b).

4.1 Methodology and data

This section provides background information on the SCMS algorithm in Section 4.1.1 and describes past applications, as well as extensions from both this and prior work, in Section 4.1.2. Lastly, we introduce the Dark Energy Survey and its mass maps, together with our approach to sample generation and the creation of noisy and noiseless simulations for verification purposes, in Section 4.1.3, and provide a summary in Section 4.4.

4.1.1 Subspace-constrained mean shift

Introduced by Ozertem & Erdogmus (2011), the SCMS algorithm is a recent addition to statistical methods dealing with the estimation of density ridges. Starting with a mesh of points placed in equidistant steps across the parameter space, the algorithm seeks to establish local principal curves in iterative steps. This can be visualized as a cloud of points shifting closer toward the nearest underlying structure at each iteration, akin to the process in which mass in our universe converges toward better-defined cosmic filaments over time. The latter can be observed in N-body simulations such as the Millennium Simulation by Springel et al. (2005) and its Millennium-II successor by Boylan-Kolchin et al. (2009b), as well as the Bolshoi simulation by Klypin et al. (2011) and the MultiDark simulation (Riebe et al., 2013).

In more formal terms, a ridge is a maximizer of the local density in the normal direction as given by the Hessian matrix. Let $\nabla p(x)$ be the gradient of a probability density function p on a space of dimension d, $H(x)$ its Hessian matrix of second derivatives, and v the eigenvectors of $H(x)$ corresponding to

a descending sorting of eigenvalues λ. We can diagonalise $H(x)$ according to

$$H(x) = U(x)\lambda(x)U(x)^\top. \tag{4.1}$$

We then take the $d-1$ eigenvectors (columns of $U(x)$) corresponding to the $d-1$ smallest eigenvalues λ as v', thereby omitting the column for the largest eigenvalue, which corresponds to the direction parallel to the ridge. Taking these eigenvectors and their linear projection operator,

$$L(x) \propto L(H(x)) = v'v'^\top, \tag{4.2}$$

we can then project the gradient of p onto the eigenvectors as

$$G(x) = L(x)\nabla p(x). \tag{4.3}$$

Following the deeper investigation of nonparametric ridge estimation methods of Genovese et al. (2014), a ridge R can thus be expressed as

$$R = \{x : ||G(x)|| = 0, \lambda_{d+1}(x) < 0\}, \tag{4.4}$$

with x as the locations in which G is zero everywhere and the omitted largest eigenvalue is positive[1].

Pseudocode describing the SCMS algorithm as it appears in Moews et al. (2021b) can be found in Algorithm 4, including thresholding, with the notation of $\theta_{*,1}$ and $\theta_{*,2}$ defining all rows of the first and second column of an N-by-2 matrix, respectively. Line 4 shows a kernel density estimation with a radial basis function (RBF) kernel, $\mathcal{K}(x) = (1/\sqrt{2\pi})\exp(-0.5x^2)$, meaning

$$\mathrm{KDE}_{\mathrm{RBF}}(x, \beta) = \frac{1}{\dim(\theta)(2\pi\beta^2)^{\frac{d}{2}}} \sum_{i=1}^{\dim(\theta)} e^{\left(\frac{||x-\theta_i||^2}{2\beta^2}\right)}. \tag{4.5}$$

[1]Without this second condition, we also locate valleys.

Data: Coordinates θ,
bandwidth β,
threshold τ,
iterations N
Result: Density ridge point coordinates ψ
$\kappa(x) \longleftarrow \mathrm{KDE}_{\mathrm{RBF}}(\theta, \beta)$, using Eqn. 4.5
$x_{\min} \longleftarrow (\min(\theta_{*,1}), \max(\theta_{*,1}))$
$y_{\min} \longleftarrow (\min(\theta_{*,2}), \max(\theta_{*,2}))$
$\psi \longleftarrow \psi \sim U(x_{\min}, y_{\min})_{\dim(\theta)}$
$\psi \longleftarrow \forall y \in \psi : \kappa(y) > \tau$
for $n \longleftarrow 1, 2, \ldots, N$ **do**
  for $i \longleftarrow 1, 2, \ldots, \dim(\psi)$ **do**
    for $j \longleftarrow 1, 2, \ldots, \dim(\theta)$ **do**
      $a_j = \frac{\psi_i - \theta_j}{\beta^2}$
      $b_j = \mathcal{K}\left(\frac{\psi_i - \theta_j}{\beta}\right)$
    end
    $H(x) = \frac{1}{\dim(\theta)} \sum_{j=1}^{\dim(\theta)} b_j \left(a_j a_j^\top - \frac{1}{\beta^2}\mathbb{I}\right)$
    $v, \lambda \longleftarrow v, \lambda$ from diagonalization $\mathrm{eig}(H(x))$
    $v' \longleftarrow$ entries in v corresp. to $\mathrm{sort}_{\mathrm{asc}}(\lambda)_{1,2,\ldots,d-1}$
    $\psi_i \longleftarrow v' v'^\top \frac{\sum_{j=1}^{\dim(\psi)} b_j \theta_j}{\sum_{j=1}^{\dim(\psi)} b_j}$
  end
end
return ψ

Algorithm 4: SCMS with thresholding as in Moews et al. (2021b).

4.1.2 Previous applications and extensions

In the few years since its introduction, the SCMS algorithm has proven valuable in a variety of fields, from the analysis of 3D neuron structures in tissue images by Bas & Erdogmus (2011) to the identification of road networks in satellite images by combining the SCMS algorithm with the geodesic method and tensor voting (Miao et al., 2014).

The first applications of the SCMS algorithm in astronomy are also the ones that are most closely related to our work. For the purpose of investigating galaxy evolution, and after initially using SDSS data as a test dataset for a methodological paper by Chen et al. (2014), Chen et al. (2015a) employ the algorithm to constrain matter distributions at different redshifts by applying *thresholding*, the practice of discarding low-density areas of the initial grid of unconverged ridge points according to a kernel density estimate of the data, precluding their identification as filaments. More formally, the threshold τ is determined by computing the root mean square of the differences between the average density estimate and the grid points' density estimates,

$$\tau = \sqrt{\frac{\sum(\phi - \bar{\phi})^2}{G}}, \tag{4.6}$$

for an element-wise subtraction of an array of grid point density estimates ϕ with average $\bar{\phi}$ over G grid points. In a related paper, they also explore the algorithm's suitability to show that dark matter is traced by baryonic matter across large-scale structure (Chen et al., 2015b). Additionally, Chen et al. (2016) provide a filament catalogue for SDSS, and show the influence of nearest-filament distances on galaxy properties like size, color, and stellar mass.

Another application, this time in cosmology, investigates non-Gaussianities of the matter density field to provide lensing effects based on filaments using the SCMS algorithm (He et al., 2017). Hendel et al. (2019) gain a better understanding of galactic mergers through left-over collision disruptions by using the algorithm to classify stellar debris and identify morphological substructures therein.

The most recent use of the algorithm pertains to the field of quantitative criminology, where multiple extensions are introduced (Moews et al., 2021b). The study investigates the optimization of police patrols via density ridges, using

publicly available data from the City of Chicago over multiple years to assess the validity and stability of predictive ridges. Apart from thresholding as described above, the study makes use of the haversine formula as described by Inman (1835), which calculates the great-circle (or orthodromic) distance as a way to prevent distorted measures that would result from, for example, using the Euclidean distance. As a more specialized case of the law of haversines, it computes the distance between two points along the surface of a sphere. The formula makes use of the haversine function for a given angle α,

$$\mathrm{hav}(\alpha) = \frac{1 - \cos(\alpha)}{2} = \sin^2\left(\frac{\alpha}{2}\right), \tag{4.7}$$

and can be used to calculate the relative haversine distance between two such points, θ_1 and θ_2, as

$$\begin{aligned}\delta_{\mathrm{hav}}(\theta_1, \theta_2) = \mathrm{hav}(&\theta_{2,1} - \theta_{1,1} \\ &+ \cos\theta_{1,1}\cos\theta_{2,1}\mathrm{hav}(\theta_{2,2} - \theta_{1,2})).\end{aligned} \tag{4.8}$$

The resulting ridge estimation tool by Moews et al. (2021b) is publicly available as a python package called DREDGE[2], with a corresponding open-source repository[3].

In this chapter, we adapt and extend DREDGE in order to apply it to DES Year 1 mass density maps. In addition to the existing thresholding and integrated use of the haversine formula, we parallelise time-consuming parts of the code, enable the manual setting of the mesh size for the ridge estimation, and implement a mathematically grounded optimization of the required bandwidth. The latter is due to Moews et al. (2021b) using a simple automatic bandwidth calculation suited to the initial application in criminology. Some of the features in the original implementation of DREDGE are redundant for this thesis as well, most notably the ability to set a top-percentage threshold to extract only ridges falling within the highest-density areas, as researchers and practitioners in criminal justice are often interested in focussing on 'hot spots' (Braga, 2005).

For the purpose of parallelising DREDGE, we make use of the embarrassing parallelism inherent in the updating function of the ridge points at each iteration.

[2] `https://pypi.org/project/dredge`
[3] `https://github.com/moews/dredge`

As this update is not reliant on other ridge points during the respective iteration, multiprocessing offers an easily accessible option to speed up the algorithm's runtime.

Kernel density estimation is a well-developed non-parametric data smoothing technique that has seen wide use in cosmological research applications. Park et al. (2007), for example, use an adaptive smoothing bandwidth with a spline kernel, and Mateus et al. (2007) apply a k-nearest neighbours density estimator to estimate the local number density of galaxies in an SDSS sample.

Similar methods have been employed in other surveys, for example by Scoville et al. (2007b), who identify large-scale structure and estimate dimensions, number of galaxies, and mass using an adaptive smoothing technique over a sample in the Cosmic Evolution Survey (COSMOS; Scoville et al., 2007a). Relatedly, Jang (2006) uses a multivariate kernel density estimator with a cross-validated smoothing parameter to estimate galaxy cluster density over a sample of the Edinburgh-Durham Cluster Catalogue (EDCC).

The SCMS algorithm's bandwidth β, plays a crucial role in determining the bias-variance relationship of the resulting distribution. A larger bandwidth results in a smoother distribution with less variance and more bias, whereas a smaller bandwidth results in a less smooth distribution with more variance and less bias. Finding an optimal bandwidth for the kernel estimator in the context of DES mass maps is crucial to ensure that dense regions will not be oversmoothed and that higher-density areas of the projected large-scale structure will not be blurred into troughs. Conversely, optimising the bandwidth allows us to preserve the properties of low-density and extended structures.

In this application, we use a likelihood cross-validation approach to find the optimal bandwidth parameter. This method provides a density estimate that is close to the actual density in terms of the Kullback-Leibler divergence (KLD; Kullback & Leibler, 1951, a measure of relative entropy). Cross-validation is performed using a maximum likelihood estimation of the leave-one-out kernel estimator of f_{-i}, which is given by

$$f_{-i}(X_i) = \frac{1}{(n-1)h} \sum_{j=1, j\neq i} K_h(X_i, X_j), \tag{4.9}$$

where h is the bandwidth parameter and K_h represents the generalized product

kernel estimator. Put more simply, we use, due to operating in latitude and longitude coordinates, a bivariate KDE to optimize the bandwidth in an error-robust way. By training the KDE in a leave-one-out way, we can evaluate using the left-out data point, repeating that step while alternating through the points. For the latter reason, this process is also known as 'rotation estimation'. The idea is to use the available data twice; first for estimation and then to evaluate. Leaving the evaluation data out in each step avoids dependence on the sample. The f_{-i} estimator thus excludes the basis function centered on that left-out point, and a high leave-one-out likelihood corresponds to a low cross-validation error. In doing so, outliers relative to other data points will be assigned lower likelihoods and contribute less to the optimal bandwidth estimate. We use the generalized product kernel estimator of Li & Racine (2006) on the latitude and longitude coordinates of the DES Y1 data as

$$K_h(X_i, X_j) = \prod_{s=1}^{q} h_s^{-1} k\left(\frac{X_{is} - X_{js}}{h_s}\right), \tag{4.10}$$

where q is the dimension of X_i, X_{is} is the s^{th} component of X_i $(s = 1, \ldots, q)$, h_s is the smoothing parameter for the given component of X_i, and $k(\cdot)$ is a univariate kernel function. This nonparametric kernel estimator does not assume any functional form of the data, only that it satisfies regularity conditions such as smoothness and differentiability. In our case, the β value we use in SCMS is the best h value obtained by cross-validation, while the variables X_i, X_j correspond to the sky positions θ.

4.1.3 Data and simulations

The Dark Energy Survey (DES) is a six-year photometric survey project to image 5000 square degrees of the sky in grizY filters using the DECam camera on the Blanco telescope, Cerro Tololo, Chile (Flaugher et al., 2015). The primary purpose of the survey is to generate a dataset for cosmology, and in particular one suitable for weak gravitational lensing measurements. DES observations completed in January 2019, and data analysis for the project is ongoing.

Weak lensing measurements use galaxies as a backlight to determine the projected gravitational fields along the paths of their light rays. Gravity bends the light paths, resulting in a shearing of galaxy images that can be measured from the

mean ellipticity of a large-enough sample of objects. One application of weak lensing is to generate mass maps. Assuming the general relativity relationship between gravitational convergence κ and mass, we can use ellipticity catalogues to map the projected, weighted overdensity in a given pixel of the survey. The weighting depends on the redshift distribution of the observed galaxies and the redshift-distance relationship that, in turn, depends on the underlying cosmology.

The DES Year 1 (Y1) data release[4], as described by Drlica-Wagner et al. (2018), includes 2D-projected mass maps (see Chang et al., 2018) estimated from cosmic shear measurements from the survey's first year (Zuntz et al., 2018). The creation of these maps is based on the redshifts estimated in Hoyle et al. (2018) and the shear catalogues made using the METACALIBRATION method (see Huff & Mandelbaum, 2017; Sheldon & Huff, 2017) which inverts shears to convergences by applying the spherical Kaiser-Squires method to the galaxy shear catalogues (Kaiser & Squires, 1993; Schneider, 1996).

Various mass maps were made using different selections of source galaxies, thus having different redshift weight functions. For this initial project, we use only the maps made with the widest range of galaxies, from $z = 0.2$ to $z = 1.3$. Here, we use the E-mode maps and their corresponding masks[5]. Although the mass maps are already in the form of a field on which Hessians and gradients may be calculated, in practice, the DES Y1 mask, which has a large number of small excised regions at the locations of bright stars, makes calculating derivatives a very noisy process, even with aggressive masking. It is considerably simpler to instead generate samples from the map and apply the SCMS algorithm described above. In order to generate these samples, we compute a mean value μ_i per pixel,

$$\mu_i = \max(1 + \omega\kappa_i, 0), \tag{4.11}$$

where κ_i is the projected overdensity in pixel i, and ω is a parameter that we can tune as desired. If ω is too low, the ridges will not be detectable in the map, and if it is too high, the lower ridges will disappear because the highest density peaks dominate. We find that a value of $\omega = 50$ works well for DES data to suppress the $\sim 2\%$ map fluctuations to within $o(1)$ point density variation. We then generate a number n_i of samples per pixel, using a Poisson distribution $n_i \sim \text{Poi}(\mu_i)$, and

[4] `https://des.ncsa.illinois.edu/releases/y1a1`

[5] `http://desdr-server.ncsa.illinois.edu/despublic/y1a1_files/mass_maps`, files `y1a1_spt_mcal_0.2_1.3_kE.fits` and `y1a1_spt_mcal_0.2_1.3_mask.fits`

place these samples uniformly within the pixel.

Note that no interpolation or reconstruction is performed in masked regions; masked pixels are omitted, and no samples are generated within them. This method is simple to apply and causes no numerical difficulties, but it does mean that ridges at the edge of the mask regions may not be detected. This occurs only in regions where interpolation would also fail, as the SCMS algorithm, like reconstructive methods, is able to identify structure across missing pixels. Methods using these ridge catalogs should account for this, for example in simulations.

To provide a testbed for our method before employing real data, we also build a suite of simulated maps using the FLASK[6] software (see Xavier et al., 2016a,b), which generates tomographic log-normal random fields that approximate large-scale structure distributions. We use the Planck best-fit ΛCDM cosmological parameters in the code, $\sigma_e = 0.25$, and the same redshift distribution as estimated for the DES source galaxies (see Hoyle et al., 2018), normalized to the correct overall density. We then generate true κ maps, which we treat as idealized noiseless simulations, and galaxy ellipticity catalogues, which we use with a spherical Kaiser-Squires map-making method to generate a noisy κ map. Lastly, we apply the DES masks to both noisy and noiseless simulated maps.

While this approach works well for the experiments performed in this work, it is insufficient for pseudo-3D extensions of our method discussed in Section 4.3.2, as FLASK cannot generate features like clusters and filaments.

4.2 Experimental results

In this section, we discuss and implement a distance-based statistical test to verify our results, and we use noisy and noiseless simulated dark matter density maps in Section 4.2.1 to test the degree of robustness to noise. We explore the general and specific properties of our method's tracing of the large-scale structure through a quantifiable comparison to the curvelet transforms in Section 4.2.2. Lastly, we present the extracted ridges, together with a comparison to previous research on trough identification from DES Y1 data, in Section 4.2.3.

[6] `https://github.com/hsxavier/flask`

4.2.1 Statistical functionality verification

As is common among studies dealing with cosmic voids, validating our approach is not an easy task, even when we can apply it to simulations. Indeed, while simulations allow us to apply our method to noiseless data, they do not provide us with forward-modeled ridges or troughs, since these need to be estimated and defined, even in the absence of noise. For ground-truth experiments on the SCMS algorithm's accuracy, see a variety of examples in the original work by Ozertem & Erdogmus (2011) and follow-up studies by Aliyari Ghassabeh et al. (2012) and Ghassabeh & Rudzicz (2021), as well as Chen et al. (2015a) for astronomical tests. A viable analysis is the comparison of ridges and troughs recovered when running the proposed approach in the diverse settings of a noiseless simulation and one that contains realistic levels of noise, as described in Section 4.1.3. In other words, we can test for robustness to observational noise.

Such a test, however, requires the choice of a similarity criterion, or distance metric, between sets of ridge points, in this case those extracted from noisy and noiseless maps. If such a metric could be defined on the ridges themselves, it could be applied directly to the ridges recovered from the proposed approach. This would allow us to directly probe the approach's sensitivity to observational noise, without requiring any specification of how to relate the ridges to a scientifically meaningful quantity like cosmological parameters. While such a test is not sufficient to guarantee that the ridges contain the information required to compute such quantities, it can confirm that the ridges are not dominated by observational noise, a necessary if not sufficient condition for any science case, provided that an appropriate metric is chosen.

Optimal transport theory (see Villani, 2008), and specifically the Wasserstein distance, provides us with a natural, principled means of computing such distances. Per the Monge-Kantorovich interpretation of optimal transport, the Wasserstein distance can be understood as the minimal possible cost incurred to move a certain amount of mass from one distribution to another. This is precisely what a set of ridges and troughs represent; a certain distribution of mass, projected in two dimensions across (a part of) the sky.

Consider two distributions of mass, $p_1, p_2 \in \mathbb{R}^N$, sampled on some discrete space of dimension N. Let $C \in \mathbb{R}^{N \times N}$ be the cost matrix whose entries C_{ij} contain the cost of moving mass from position i to position j. We define the Wasserstein

distance as

$$W(p_1, p_2) = \text{argmin}_{T \in \Pi(p_1, p_2)} < T, C > . \tag{4.12}$$

Here, $\Pi(p_1, p_2)$ is the set of transport plans between p_1 and p_2. For any such matrix T, each of its entries T_{ij} contains the amount of mass of p_1 that is transported from position i within p_1 (denoted $p_{1,i}$) to position j within p_2. By construction, for any row i, we then have

$$\sum_j T_{ij} = p_{1,i}. \tag{4.13}$$

Similarly, summing over the columns of T yields the entries of p_2. We can then simply express the set of acceptable transport plans as

$$\Pi(p_1, p_2) = \left\{ T \in \mathbb{R}^{N \times N}, \forall i, j, \sum_j T_{ij} = p_{1,i}, \sum_i T_{ij} = p_{2,j} \right\}. \tag{4.14}$$

In order to find the optimal T, the solution to Eqn. (4.12), we use the entropic regularization scheme proposed by Cuturi (2013), which allows for the Wasserstein distance between discrete measures to be computed by using an iterative scheme (see the textbook by Peyré & Cuturi, 2019, for a recent overview of computational schemes for the practical computation of optimal transport quantities).

The raw output of SCMS is a set of shifted 'mesh points', each with a 2D position vector, that make up the ridges. While we could compute the Wasserstein distance directly between two sets of mesh points, the objects of interest in our case are the ridges they comprise, and the troughs thus delimited, rather than the points themselves. For this reason, we convert each set of ridge points into a binary 2D image with each pixel's value set to zero if no mesh point fell within it, and one otherwise. We then compute the Wasserstein distance between the two resulting images. Here, N is equal to the total number of pixels, in our case 10952, and the cost matrix C is the Euclidean distance between each pair of pixel positions. The use of the latter distance metric is motivated by comparing binarized 2D images, as opposed to within-image calculations on the curved sky.

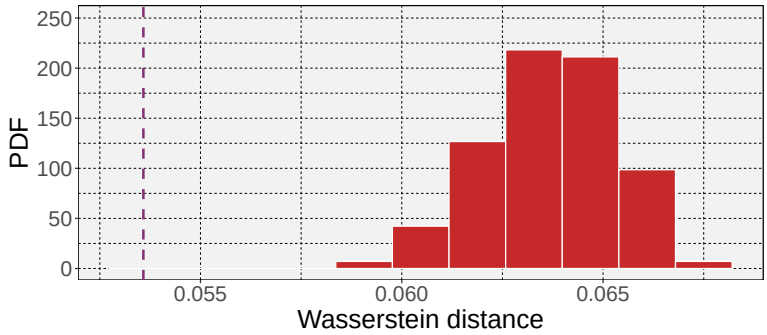

Figure 4.1 *Wasserstein distance between the ridges obtained on the noisy simulation and either its noiseless counterpart, as a vertical dashed line in purple, or a set of 101 random distributions of mass in red.*

In order to provide a basis for comparison, we generate random maps by projecting the DES mask onto the image plane and uniformly sampling the same number of non-zero pixels as those present in the images that contain our ridges. We then compute the Wasserstein distance between those ridges and the ones obtained from the noisy simulation. The distribution of these distances is represented in Figure 4.1, along with the distance computed between the two sets of ridge points based on noisy and noiseless maps. As can be seen, our ridge-finding method shows robustness to realistic amounts of noise. This test allows for a general confirmation that the ridges we obtain contain physical information about the distribution of matter, obviating any specification of the problem under study, that is, a precise definition of troughs or voids. As mentioned above, further testing tailored to the application of interest is recommended. These tests could likely use the same settings and simulations, where we compare the output in both noiseless and noisy cases, for example the cosmological constraints derived from the troughs delimited by those two sets of ridges.

4.2.2 DES ridges and curvelet comparison

Sparse signal processing (see Starck et al., 2015) provides solutions to many signal retrieval problems including, but not exclusive to, image denoising. The approach relies on finding a representation space in which the signal can be sparsely represented, a typical example being a sinusoidal signal, which can be fully expressed with very few non-zero coefficients in Fourier space. While natural signals are rarely sinusoidal, wavelets are commonly used as a sparse basis of representation (Mallat, 1999). They can, however, perform poorly when the features to be recovered are rectilinear or elongated, as is the case for estimating

ridges. Because of this shortcoming, analogous transforms have been designed specifically for these cases, namely ridgelets and curvelets (Candès & Donoho, 1999; Candès et al., 2006).

These transforms have led to a wide range of applications in astrophysics and cosmology (Starck et al., 2003). Starck et al. (2004) use wavelets, ridgelets and curvelets to detect and characterise, on simulations, various sources of CMB anisotropies, which include imprints of inflation, the Sunyaev-Zel'dovich effect, and cosmic strings. The latter have also been studied in a CMB framework by Vafaei Sadr et al. (2017) and Hergt et al. (2017), using the curvelet transform, while Laliberte et al. (2018) use ridgelets to that end within N-body simulations. Gallagher et al. (2011) apply both to solar astrophysics, and Jiang et al. (2019) use curvelets for radio transient detection.

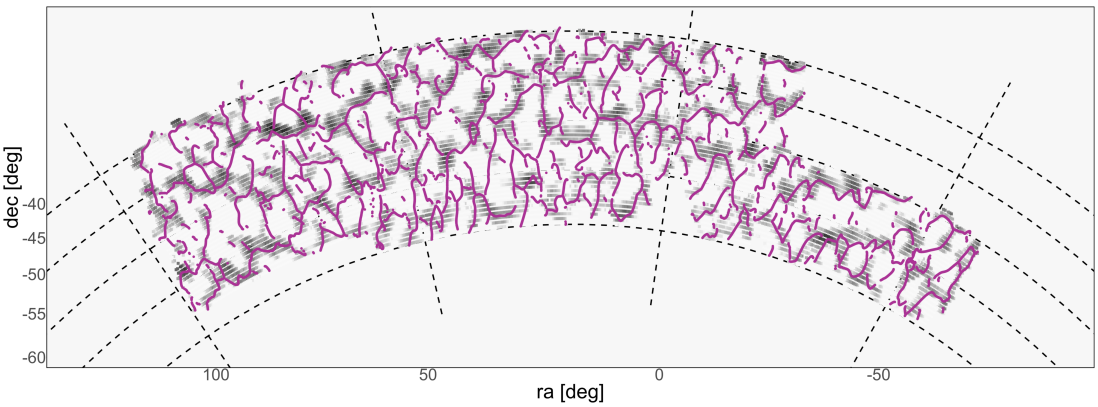

Figure 4.2 *Comparison of density ridges and a curvelet reconstruction. Ridges in purple are superimposed on structural constraints obtained via curvelet denoising in shades of grey, with higher densities shifting from lighter to darker. Both results are based on DES Y1 weak lensing mass density maps.*

In our case, the application of curvelets allows for a straightforward and entirely independent means of recovering the ridges. We perform a simple denoising, that is, a thresholding of the input mass in curvelet space. Our SCMS-recovered ridges are obtained from the samples described in Section 4.1.3, which themselves rely on the choice of the ω parameter. We use these samples to generate a two-dimensional, discrete image by counting their number in each pixel bin. We then apply curvelet denoising to the resulting image, using the freely available `Sparse2D` package[7]. The thresholds are selected using the False Discovery Rate approach described by Benjamini & Hochberg (1995), and the coarse scale is

[7] `https://github.com/CosmoStat/Sparse2D`

discarded. The resulting denoised map, converted back to sky coordinates, is shown in Figure 4.2. Upon visual inspection, we observe good agreement between the uncovered structured overdensities and the ridges obtained by SCMS, overplotted in purple.

To get a more quantitative view of the differences in the structures recovered by both approaches, we project the ridges yielded by the SCMS algorithm into the same pixel grid as that used by the curvelet denoising step. For every pixel that then contains part of one of the ridges, we check the corresponding value in the curvelet reconstruction. We consider all pixels where that value is 0 to be a mismatch. Figure 4.3 shows all such mismatches in orange, while parts of the ridges that match with the curvelet reconstruction are shown in purple. About 31% of the SCMS-derived ridges mismatch the curvelet reconstruction.

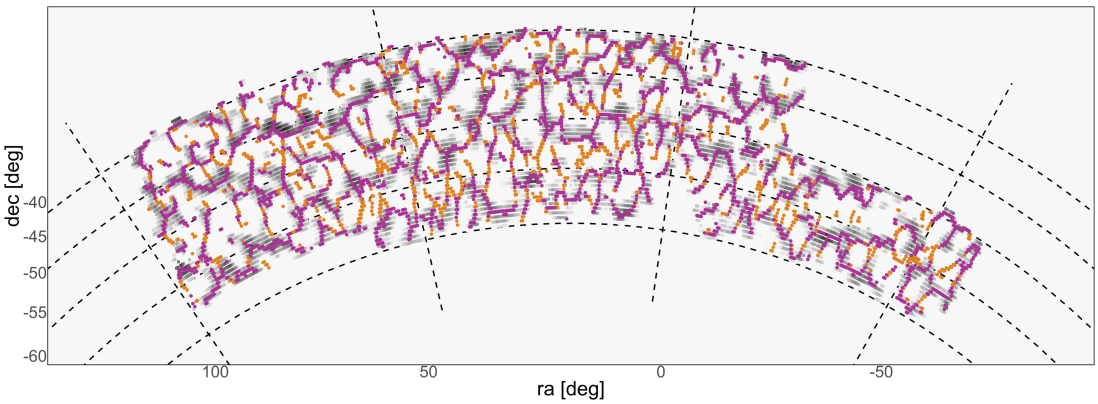

Figure 4.3 *Similar to Figure 4.2, but with ridges shown in purple where they match the curvelet reconstruction, and orange otherwise.*

As can be seen from Figure 4.3, the mismatches mostly correspond to areas where structure is present in two nearby areas of the sky, and where the curvelet reconstruction keeps those two areas disjoint, while SCMS-derived ridges link them together. Although very different in their heuristics, both approaches ultimately perform some form of denoising of the input mass maps. Those differences between the two denoising approaches could indicate that one of the two fails to properly separate noise from signal; the areas of disagreement could either be due to the curvelets considering signal to be noise, or to SCMS algorithm turning noise into ridges.

In order to root out the cause of these discrepancies, we reprocess the DES data while tuning the parameters of each method. In the case of SCMS, we impose a stronger denoising by multiplying the bandwidth, β, obtained as described in

Section 4.1.2, by a factor $\zeta \in \{1.25, 1.50, 2.00\}$, or considering a higher threshold $\Upsilon > \tau$ (see Section 4.1.2 as well). The percentage of mismatching ridges for both cases is shown in Figure 4.4.

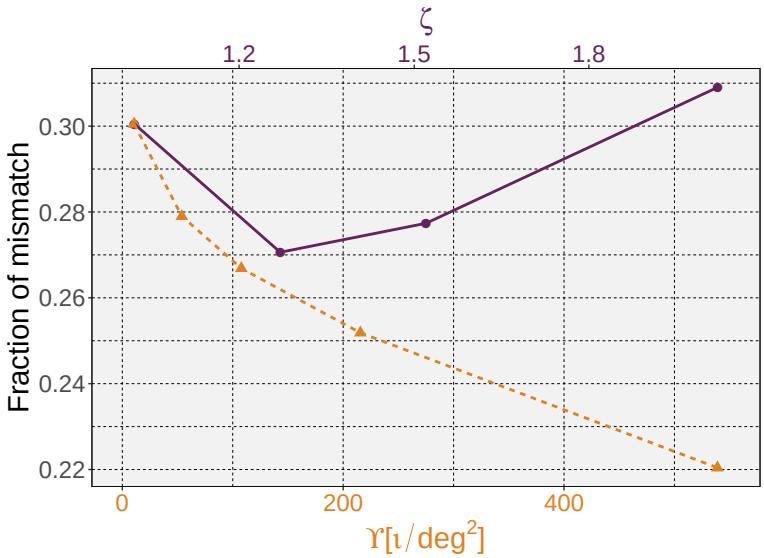

Figure 4.4 *Fraction of 'mismatch' between SCMS-derived ridges and the curvelet reconstruction, when varying the τ and β hyperparameters of the SCMS algorithm. The bottom axis shows the threshold on the ridges (Υ) in units of meshpoints (ι) per square degree (deg^2). The top axis shows the multiplication factor (ζ) of the bandwidth. The further to the right, the stronger the implicit denoising performed by the algorithm. The absence of a clear decreasing trend shows that the differences between both methods, as illustrated by Figure 4.3, are not due to hyperparameter choices.*

Artificially increasing the strength of the denoising performed by the SCMS algorithm does not lead to ridges that match the curvelet reconstruction more closely. In fact, the percentage of mismatch between the two approaches is not even monotonic with respect to the bandwidth. In all cases, we still observe ridges that tend to be more connected in the SCMS case. Similarly, we try imposing weaker denoising in the curvelet approach, by using increasingly lower values of k in $k\hat{\sigma}$ thresholds instead of the False Detection Rate approach, where $\hat{\sigma}$ is estimated from the data using the Median Absolute Deviation estimator. Once again, this yields no clear increase in the match between the two outputs.

This shows that the support of each method, in their respective (hyper)parameter spaces, are disjoint. In other words, the differences between curvelet reconstruction and SCMS-derived ridges seen in Figure 4.3 are due to intrinsic differences between the approaches, as opposed to a failure of either at the denoising task.

Implications of those differences for potential applications are discussed further in Section 4.3.2.

4.2.3 Ridges in the Dark Energy Survey

After the above comparison with curvelet transforms, we perform a second comparison of our results with independent methodology, as void and trough detection remain a current focus of interest within the cosmology community. Specifically, Gruen et al. (2018) derive cosmological constraints via density split statistics, counting tracers in cells to split lines of sight and measuring counts-in-cells and gravitational shear in regions of varying density. Using REDMAGIC luminous red galaxies at $0.2 < z < 0.45$ to trace the foreground matter density field, they count the number of galaxies that fall within circular top-hat apertures within radii $\theta_T = [10', 20', 30', 60']$.

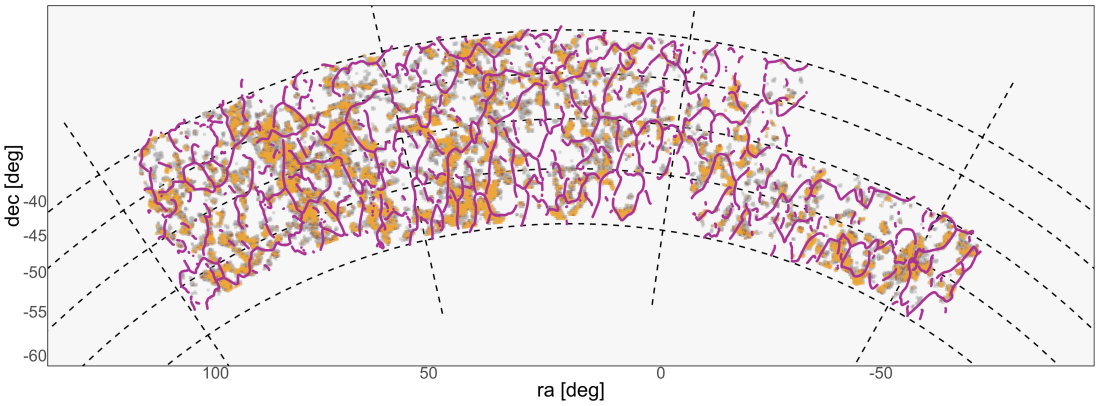

Figure 4.5 *Comparison of density ridges and previous results from Gruen et al. (2018). Ridges from this work are shown in purple, and are superimposed on mass density probabilities that were obtained by measuring counts-in-cells along lines of sight of the foreground luminous red galaxies* REDMAGIC *sample.*

Each line of sight is then assigned to a density quintile based on those counts, the data for which is publicly available[8], with the highest quintile being of interest to us in terms of high-density ridge analogues. Due to the DES Y1 data being more inhomogeneous in depth than the SDSS DR8 data used for the REDMAGIC catalogue (see Aihara et al., 2011), tracer galaxies are removed and the area is defined as fully masked if the sample from the catalogue is not complete to $z = 0.45$. We compare our findings to Gruen et al. (2018), which uses the same DES

[8] `https://des.ncsa.illinois.edu/releases/y1a1/density`, file `trough_maps.tar.gz`

Y1 data we make use of in this work in combination with SDSS DR8 data, without sharing the same underlying mass density map, as in the curvelet comparison. Figure 4.5 shows the same DES Y1-extracted ridges as in Figure 4.2, underlaid with the highest-quintile density measurements from Gruen et al. (2018).

While these quintile-based comparisons are more 'spotty' due to the masking based on depth completeness described above, and qualitatively different when compared to curvilinear structures extracted by the SCMS algorithm, the lines generally trace the same high-density regions shown in the figure. The shown percentiles also use foreground galaxies as tracers of the matter field, as opposed to the mass density maps based on weak lensing that are used as the input for our modified version of `DREDGE`. In contrast to the previous comparison with curvelets as an alternative method in Section 4.2.2, the goal of this experiment is to perform a comparison across both methods and underlying data to reach a consensus in terms of high-density regions on the sky.

4.3 Discussion

This section provides an in-depth discussion of our findings, their implications, and future direction. In Section 4.3.1, we recapitulate the implemented extensions and performed experiments, discuss the results of the latter, provide an overview of advantages and disadvantages, and compare the SCMS algorithm to ridge-like structures extracted by other means. In Section 4.3.2, we describe planned follow-up research for developing and testing cosmological probes and models, and to validate cosmological analysis pipelines.

4.3.1 Overview

Ridges derived from our methodology have several properties that make them interesting for lensing trough studies, compared to simply using local minima. First, since the method generates a point cloud as a starting point, rather than working with the map directly, it does not require convergence map smoothing. This means that smaller-scale troughs can potentially be probed. Secondly, since the trough points maximise the distance from local ridge structures, they probe inter-cluster and inter-filament regions directly, rather than by proxy. Finally, masks from bright stars on small scales (below the typical ridge segment length)

do not significantly affect the operation of the algorithm.

We incorporate several extensions from previous research into our implementation of the SCMS algorithm, both for accuracy and performance , and describe these in Section 4.1.2. Thresholding the initial mesh of points based on density estimates, as introduced by Chen et al. (2015a), ensures that the curvilinear structure is constrained to higher-density areas, which solves potential issues associated with identifying ridges in sparsely populated regions in the data. We also make use of the haversine formula, as implemented by Moews et al. (2021b) for geospatial analysis, to prevent the distortion of measurements on the curved sky.

We further extend the algorithm by exploiting the potential for embarrassing parallelism inherent in the updating function of ridge points at each iteration. This allows users to reduce the runtime significantly by using a multiprocessing setup, thus removing a major obstacle when applying the SCMS algorithm to large datasets. The algorithm itself is reliant on the choice of a bandwidth, which plays a crucial role in determining the bias-variance relationship of the distribution, with larger bandwidths resulting in smoother distributions with less variance and more bias, and with smaller bandwidths resulting in a less-smooth distribution with more variance and less bias.

For our current application, this means that large-scale structure in dark matter density maps could be blurred into cosmic troughs through bandwidths that are too large, while recovered ridges could show spurious fine-grained structure through bandwidths that are too small. We solve this by introducing a likelihood cross-validation to the SCMS algorithm to automatically find the optimal bandwidth in a data-driven way, providing a density estimate close to the actual density in terms of the Kullback-Leibler divergence.

The experiments performed in this work are designed to test for both noise robustness and the correct tracing of high-density regions. We generate noisy and noiseless simulations that correspond to the DES Y1 mass density maps in Section 4.2.1, and calculate the Wasserstein distance between binarized maps of the resulting curvilinear structures. In addition, we calculate the same metric for randomly generated maps with uniform sampling to place the results into context, showing robustness to realistic levels of noise. While this test provides some measure of the robustness of our method to noise in real data, as discussed in Section 4.2.1, additional testing is recommended to avoid introducing biases specific to any application. However, as our goal is to propose a general approach

to recover ridges, our tests were chosen to be as general as possible.

The overall agreement we find between our ridges and the structure recovered by curvelet denoising in Section 4.2.2, despite the independence and vast difference between the two approaches, is a good indication that our ridges successfully capture information contained in the matter distribution. The differences between the two methods lie mostly in the curvelet reconstruction leading to more disconnected patches, while our ridges tend to connect these areas.

Our experiments show that this is the result of intrinsic differences between both approaches rather than a poor choice of hyperparameters; while clearly different, neither of the two resulting estimates are 'wrong'. This illustrates an important point: For a given matter distribution, there is no such thing as a set of true ridges. Choosing a proper definition is a non-trivial task in itself, which is precisely one of the motivations for the present work. In this chapter, where we aim to present both methods in as general a framework as possible, we will simply point out the difference in the structures identified by each, highlighting once again that neither is more correct than the other. Tests tailored to a specific application, however, could be used to determine which of the two is more appropriate or useful.

Another key difference between the two approaches is that the SCMS algorithm does indeed produce ridges, that is, curvilinear structures, whereas the output of the curvelet denoising is still a full two-dimensional map. The curvelet reconstruction also still contains information about the amplitude of the signal at each position, while our ridges are binary, meaning that every position either contains a ridge or not. If such a truly curvilinear format is required for a given application, and the patchier nature of the curvelet-recovered 'ridges' is better suited for the respective task at hand, it would be straightforward to combine both approaches (see Figure 4.3).

To further our analysis, we compare ridges extracted from DES Y1 mass density maps, which are based on weak lensing, to trough mass probabilities by Gruen et al. (2018) in Section 4.2.3, using the highest quintile corresponding to large-scale structure. Since this test uses redMaGiC luminous red galaxies at $0.2 < z < 0.45$ rather than weak lensing to trace the foreground matter density field and is also limited to the depth coverage of the combined DES Y1 and SDSS DR8 data, it serves as a comparison across both methods and sources. Given this combination in Gruen et al. (2018), a stronger masking is present, leading to a more 'spotty' nature of the visible structure when compared to the curvelet

representation in Section 4.2.2. Despite these differences, we see both ridges and quintile probabilities falling into the same areas, with DES Y1-extracted ridges tracing this complementary dataset.

Our cross-validated bandwidth optimization approach to kernel density estimation provides a data-driven and, via the Kullback-Leibler divergence, mathematically motivated way to perform ridge estimation. As with any approach that is data-driven, sanity checks should be performed to ensure physically plausible ridges. These derived ridges are, therefore, quite dependent on both the data quality of the DES Y1 survey and the shape assumptions in the density estimation, as well as on the distribution space being approximately symmetric and unimodal. In our approach, care should be taken when considering samples derived from heavy-tailed distributions, as the bandwidth optimization can be prone to overestimation (Silverman, 1986). To encourage both verification and further analysis of this denoising approach to dark matter density maps, we release a catalogue[9] of the ridges extracted from DES Y1 data.

4.3.2 Applications

In terms of future directions and follow-ups, we intend to extend our method to the three-dimensional case to identify full voids instead of trough projections. One way to reach this goal is to use tomographic 2D information, effectively identifying voids within a layered pseudo-3D structure. This can be performed using tomographic reconstruction techniques (see Herman, 2009), which have been highly enhanced using deep neural networks (Wang et al., 2018a).

Another interesting avenue we plan to pursue is the inclusion of lensing information along identified ridges to further bolster the viability of this approach, as well as the matching of a cosmic web reconstructed from DES data to Planck LSS (Bouchet, 2016). This follow-up study will also continue the work of Gruen et al. (2016) on weak lensing shear around troughs to further probe the connection between convergence fields and matter density, and extend research on optimal trough finders for weak lensing maps by Davies et al. (2021). For the latter, we propose an iterative expanding-circle approach to trough identification, qualitatively similar to the spherical void finder by (Padilla et al., 2005), as the preparatory step of finding troughs in our catalogues. Ridges such as the ones

[9]The data table is available at CDS via anonymous ftp to `cdsarc.u-strasbg.fr` (130.79.128.5) or via http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/MNRAS.

presented in this work are especially useful for this, as the lack of a smoothing requirement of the convergence map allows for the potential identification of smaller-scale structure, and thus troughs. By inducing a sparse, hence denoised, representation of ridges structures, our approach is also particularly useful for the topological analysis of ridge structures via, for example, persistent diagrams and Betti numbers (Xu et al., 2019).

As pointed out in Section 1.1.3, an interesting avenue of follow-up research is the differentiation between the standard model and alternative cosmologies through screening mechanisms, as both scalar-tensor theories and massive gravity include the latter to ensure compatibility with observations. The approach for tomographic pseudo-3D voids outlined above could, thus, be easily adapted to test alternative cosmologies. Davies et al. (2019) find that troughs, through abundances and weak lensing profiles, can also act as sensitive probes of gravity, opening up a direct extension of the presented work to tests of cosmological models. Insofar as the ridges we locate here are projections of filaments (which remains to be shown) and other structures intermediate between voids and clusters, lensing along and around them also offers a potential probe of screening mechanisms: We expect screening to switch off at some point as we move away from the ridge, and this may be detectable.

A stacking approach akin to that presented in Xia et al. (2020), but around ridges, should prove informative on these poorly understood structures, along with other tracers of ridge topology and density. Any prediction code used to obtain likelihoods needs to account for the strong noise biases due to identifying ridges via gravitational lensing, and then measuring lensing around those same ridges. The most likely method for theory predictions of such quantities is interpolating between simulations, as is done in most peak and trough analyses. Consequently, this means that simulations must have very accurate noise properties, as opposed to being homogeneous.

In addition to the testing potential of voids for cosmological models outlined in Section 1.1.5, Barreira et al. (2015b) investigate cubic Galileon and nonlocal gravity theories as modified gravity pathways that change the lensing potential. In terms of screening, the former theory does not screen such modifications inside of voids, featuring lensing effects roughly twice as strong as predicted by general relativity, offering an avenue for lensing-based cosmological tests. For the Vainshtein mechanism, a specific screening mechanism native to a variety of gravity models, Falck et al. (2018) also note that dark matter in large-scale

structure is not screened by the latter, and demonstrate that voids are completely unscreened and can be used to show deviations from the ΛCDM model through tangential velocity and velocity dispersion of stacked voids. Regarding void stacking, Cautun et al. (2016) also offer a non-spherical stacking approach based on a boundary profile, which could further enhance such tests. These previous results offer convincing evidence that a number of void and trough statistics can be used effectively to test for alternative cosmological models, especially those making use of screening mechanisms.

This planned future application would yield a three-dimensional void catalogue, the statistical properties of which could constitute powerful cosmological probes. To start, the higher-order statistics of the spatial distribution of voids could be compared to results from Hamaus et al. (2014) and Pycke & Russell (2016). Additional insights could also be gained from the spatial correlation of voids as a function of their sizes and shapes. These void statistics also open up possibilities as a sensitive probe of the dark energy equation of state, as demonstrated by Pisani et al. (2015) for the number of observed voids and Demchenko et al. (2016) for their profiles in the context of the spherical evolution model, possibly to be combined with complementary constraints on the dark energy equation of state (Moews et al., 2019a). Conversely, Bos et al. (2012) find that morphological properties of voids are too strongly affected by sparsity and spatial bias in their given sample to differentiate cosmologies at a statistically significant level.

Just as the statistical characterization of the triaxiality of galaxy clusters has proven fruitful for cosmological analysis (see Simet et al., 2017; Melchior et al., 2017; Chiu et al., 2018), so too might the triaxiality distribution of voids. One could further imagine a hierarchical inference procedure that uses the void size and shape distributions as a prior to iteratively update ambiguous photometric redshift probability density functions, just as those same probability density functions are used to hierarchically infer the void size and shape distributions from the tomographic projections used to derive the triaxial void catalogue. A supplemental application of these distributions could serve as a validation test for the sophisticated simulated catalogues being developed for future LSS surveys (Korytov et al., 2019).

In short, this method brings within reach a number of promising avenues for testing cosmological models, developing novel cosmological probes, and validating cosmological analysis pipelines. In this chapter, we demonstrate the utility of curvilinear structures for the denoising of large-scale structure based on weak

lensing, extending the available methodology for this purpose. We plan to explore these applications in future papers, but also welcome potential users of the ridge catalogue to experiment with the dataset we provide alongside the results.

4.4 Summary

This work presents a new ridge estimation approach based on an extension of the subspace-constrained mean shift algorithm, a filamentary search method, and releases the corresponding results as a catalogue of curvilinear structure. As an application case of current relevance, we apply the method to dark matter mass density maps from the DES Y1 data release to extract high-density ridges between cosmic troughs. Our results demonstrate the viability of ridge estimation as a precursory step for denoising cosmic filaments, leading to a versatile and effective identification of cosmic troughs.

We extend the SCMS algorithm by including the haversine distance, a customization of the mesh size for ridge estimation, automatic optimization of the bandwidth used in the process, and the parallelization of the updating function for mesh points to scale down the algorithm's runtime. We also include the thresholding extension of the SCMS algorithm from a previous application to astronomical data. In order to test the robustness of our method, we recover the ridges from simulations under different noise levels, and use the Wasserstein distance as a comparison metric.

We further compare our extracted ridges, which are based on DES Y1 weak lensing data, with curvelet denoising of the same data, and with high-density quintiles derived from both DES Y1 and SDSS DR8 foreground galaxies limited by inhomogeneous depth coverage. This allows us to compare results across both methods and data sources, leading to highly reasonable agreement. Lastly, we discuss the utility of our approach in the context of further investigations, with a focus on ridge lensing, testing alternative cosmological models through troughs and pseudo-3D voids, and screening mechanisms.

Chapter 5

Hybrid analytic and machine-learned baryonic property insertion into galactic dark matter haloes

In this chapter, we introduce a novel framework that aims to marry the benefits of multiple approaches. Large dynamic ranges, large Reynolds numbers, and highly supersonic flows make the modeling of baryonic physics in cosmological simulation numerically demanding when compared to the collisionless dynamics of dark matter in N-body simulations. Such efforts are, however, necessary to investigate theories of galaxy formation and evolution, as well as alternative cosmological models and their impact on galaxy populations (Vogelsberger et al., 2020). Our approach is based on the *equilibrium model* introduced by (Davé et al., 2012), a simple galaxy evolution framework whose free parameters correspond to baryon cycling, meaning the flows of material in and out of galaxies, which modern hydrodynamic simulations indicate is the main modulator of galaxy growth. For many cosmological applications, our approach to hybrid analytic and machine-learned baryonic property prediction based on dark matter halo merger trees offers a way to considerably reduce the computing resources required to predict comprehensive baryonic property sets, while leveraging information from both state-of-the-art Gpc-scale dark matter-only simulations and the latest hydrodynamic simulations.

This chapter is organized as follows. In Section 5.1, we provide the background on the equilibrium model of galaxy evolution, an explanation of the type of machine

learning model used in this work, and related recent research on machine learning for baryonic galaxy property prediction. Our methodology and data are described in Section 5.2, covering our newly proposed and subsequently implemented extensions of the equilibrium model, the creation of a hybrid prediction framework based on both the equilibrium model and machine learning, and the simulation from which we draw our dataset. We present our experimental setup and results, both for preliminary experiments with partial enhancements and a full experimental suite, in Section 5.3. Lastly, we discuss the results and implications of our work in Section 5.4 and offer a summary in Section 5.5. This work has been peer-reviewed and published in *Monthly Notices of the Royal Astronomical Society* (Moews et al., 2021a).

5.1 Background

In this section, we provide the necessary background upon which this work builds, covering previous research as well as both the analytic and machine-learned part of our hybrid approach to inserting baryonic properties into N-body simulations. The equilibrium model, an analytic formalism of galaxy evolution, is described in Section 5.1.1, while Section 5.1.2 offers an overview of decision tree learning and ensembles, and Section 5.1.3 describes related research on machine learning for baryonic property prediction.

5.1.1 The equilibrium model of galaxy evolution

Galaxies in hydrodynamic simulations have been observed to fluctuate around a self-regulatory relation on short timescales (Dutton et al., 2010). In this context, the equilibrium model of galaxy evolution is an analytic formalism inspired by such simulations and based on the premise that galaxies are situated in slowly-evolving equilibria between inflow and outflow through accretion and feedback, as well as star formation, aiming to capture the evolution of galaxies in simulations (Davé et al., 2012). The equilibrium condition in the vicinity of which star-forming galaxies are seen to fall in such simulations is, with mass inflow rate $\dot{M}_{\text{in}}$, mass outflow rate $\dot{M}_{\text{out}}$ and star formation rate (SFR) $\dot{M}_{*}$,

$$\dot{M}_{\text{in}} = \dot{M}_{\text{out}} + \dot{M}_{*}. \tag{5.1}$$

The reason for omitting a term for gas reservoirs, meaning the prevalence of molecular gas that is related to the star formation rate, is the finding by Finlator & Davé (2008) that the rate of change for such reservoirs is negligible in relation to the other terms in Equation 5.1. While this constitutes a simplification, and despite changes in the gas reservoir having effects over short time frames, omitting the term still results in realistic galaxy growth when averaged over cosmological time frames. The interplay of mass inflow, outflow, and SFR in Equation 5.1 also bears resemblance to the reservoir model and the bath tub model (Bouché et al., 2010; Krumholz & Dekel, 2012). The notable difference is that the SFR is expressed as $(1 - R)\dot{M}_*$ in these models, with R as the (constant) gas recycling factor; in the equilibrium model, a time-dependent fitting formula is used for R instead (Leitner & Kravtsov, 2011). The gas regulator model by Lilly et al. (2013), in the simplest form of which the specific star formation rate is set to the galaxy's specific accretion rate, is more accurate on shorter timescales, but increases the complexity of equations in return.

It also resembles SAMs in that both are analytic, but differs in the omission of merger trees (until now) and angular momentum conservation to cool gas. Using a Bayesian MCMC approach, Mitra et al. (2015) show that the model is well-suited to describe observations of scaling relations in galaxies from $z = 0$ to $z = 2$, with more recent investigations of $z = 0.5$ to $z = 3$ and including gas and dust observations (Saintonge et al., 2013; Mitra et al., 2017). The equilibrium model rests on three central equations that describe the behavior of the stellar, gas, and metal content over cosmological time frames. The SFR takes the form

$$\dot{M}_* = \frac{\zeta \dot{M}_{\mathrm{grav}} + \dot{M}_{\mathrm{recyc}}}{(1 + \eta)}, \tag{5.2}$$

with $\eta \equiv \dot{M}_{\mathrm{out}}/\dot{M}_*$ as the mass loading factor acting as an ejective feedback parameter, and ζ as the preventive feedback parameter describing the rate of growth of halo gas as the amount of halo-entering gas that does not reach the ISM, which can be defined by rearranging equation (5.2). $\dot{M}_{\mathrm{recyc}}$ is parameterized as a function of t_{recyc}, which is the time frame it takes for ejected gas to be recycled. $\dot{M}_{\mathrm{grav}}$ denotes the baryonic inflow into the dark matter halo as the infall into the halo that is assumed to be metal-free and derived from dark matter simulations, and $\dot{M}_{\mathrm{recyc}}$ is the wind recycling parameter, meaning the re-accretion of previously ejected gas. The above equation is central to the equilibrium model's description of accretion and baryon cycling feedback steering the SFR, and can be derived as a

reformulation of equation 5.1. For a full derivation starting with the equilibrium condition, see Davé et al. (2012). The metallicity in the ambient interstellar medium (ISM) gas Z_{ISM} in the equilibrium model is, with y as the survey-derived metal yield,

$$Z_{\mathrm{ISM}} = y \frac{\mathrm{SFR}}{\zeta \dot{M}_{\mathrm{grav}}}. \tag{5.3}$$

The first part can be understood in the context of the metal enrichment rate, which is the yield times the SFR. Lastly, with the depletion time scaling as $t_{\mathrm{dep}} \propto t_H M_*^{-0.3}$ for Hubble time t_H, and with the specific star formation rate sSFR, the dependence of a galaxy's gas fraction f_{gas} on both t_{dep} and sSFR is described as

$$f_{\mathrm{gas}} = \frac{M_{\mathrm{gas}}}{M_{\mathrm{gas}} + M_*} = \frac{1}{1 + (t_{\mathrm{dep}}\ \mathrm{sSFR})^{-1}}. \tag{5.4}$$

Specifically, t_{dep} represents the time frame in which gas in the ISM is converted into stars. Related work includes the modeling of the regulation of galactic star formation rates in disk galaxies by Ostriker et al. (2010), research on HI content of galaxies in hydrodynamic simulations by Davé et al. (2013), and the investigation of galactic outflows in cosmological zoom simulations (Anglés-Alcázar et al., 2014). For a more in-depth introduction, the reader is referred to the original paper by Davé et al. (2012) or, for a broader overview of physical models of galaxy formation, Somerville & Davé (2015).

5.1.2 Extremely randomized tree ensembles

A more in-depth overview of decision trees and ensembles can be found earlier in Section 1.2.5 as part of the methodological introduction, but we assume that a reminder with suitable visualizations is beneficial for this chapter. As hierarchically built models with path criteria at each note, a single decision tree is usually an interpretable white-box model, but the use of an ensemble of trees often boosts performance considerably at the price of losing much of that interpretability. Figure 5.1 shows the splitting process in building said hierarchical structure, while Algorithm 5 describes this splitting process in extra trees as commented pseudocode.

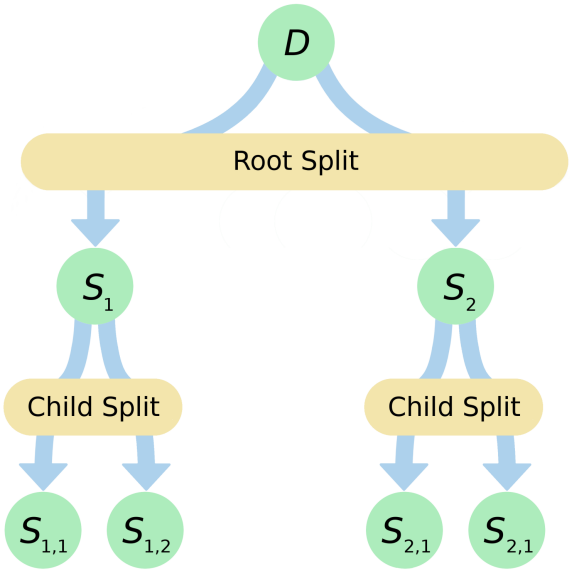

Figure 5.1 *Splitting process in extremely randomized trees. For a dataset D, which acts as the 'root', the tree is built by generating binary splits to produce a deterministic flowchart, further splitting along the subsequent 'child' nodes representing subsets until terminating in the end nodes as 'leaves'.*

Data: S := Local learning subset for node-splitting,
k := Number of attributes selected at each node,
n := Minimum sample size for splitting a node
Result: $\hat{t}$:= Optimal splitting choice for the given node
Check whether the subset S fulfils splitting criteria
if $|S| < n$ **or** $\forall c \in S : c$ const. **or** Tree(S) const. **then**
  If not, return no splitting choice
  return None
end
else
  Select k non-constant attributes in S
  $\{a_1, \ldots, a_k\} \sim \forall c \in S : c$ const.
  Create an empty set to store random splits
  $T \longleftarrow \{\}$
  Create and store k random splits for attributes
  for $i \in \mathbb{N}_{\leq k}$ **do**
    $T \longleftarrow T \cup \{a_{\text{cut}} \sim \mathcal{U}(\min(S_{a_i}), \max(S_{a_i}))\}$
  end
  Calculate and return the optimal splitting choice
  return $\hat{t} \in T$ s.t. $\text{score}(t, S) = \max_{i \in \mathbb{N}_{\leq k}} (\text{score}(T_i, S))$
end

Algorithm 5: Splitting in extremely randomized trees.

The general way in which a machine learning ensemble, for example using extra trees, works is that multiple models are trained and then used to make a prediction given the same input. In order for this to work, these models do, of course, have to differ from each other to make use of the added predictive power the development of such ensembles is motivated by. In the case of extra trees, this is not done by training each model on a subset of the available training data. Instead, each model is built not by using optimal splits for each node of the tree, but by choosing node splits based on a randomized selection. The flow of information in building an ensemble of extra trees can be seen in Figure 5.2.

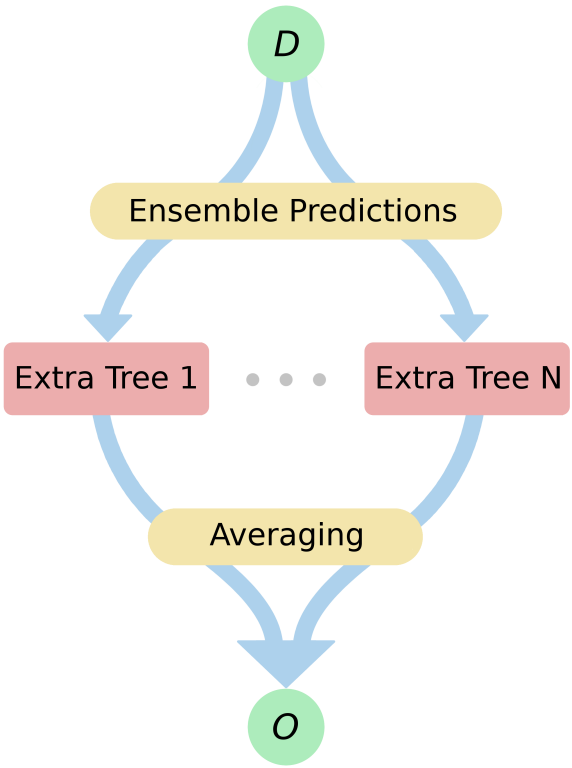

Figure 5.2 *Layout of predictions by extremely randomized trees as a previously trained ensemble of N decision trees. For a given input dataset D to produce predictions, the data is fed, in full, into each separate tree to generate predictions, which are then averaged to produce a final output O.*

Previous related research has explored the viability of different types of models for baryonic property prediction based on N-body simulations, allowing us to draw on existing research to determine the model of choice, as described below in Section 5.1.3 (Kamdar et al., 2016b; Agarwal et al., 2018; Jo & Kim, 2019).

5.1.3 Machine learning and baryonic properties

While machine learning to paint dark matter haloes with galactic properties is sparse in the literature, several works have explored this topic so far. Kamdar et al. (2016a) introduced the application of machine learning techniques to semi-analytic cosmological simulations, training various algorithms to predict the total stellar mass M_*, the stellar mass in the bulge approximated via $M_{*,\mathrm{half}}$, and the central black hole mass M_{BH}, as well as hot and cold gas masses, for each dark matter halo in the Millennium simulation at $z = 0$ (Springel et al., 2005). Their research targets the prediction of these baryonic constituents based solely on dark matter halo merger trees as the training input, using the GADGET-2 algorithm described by Springel (2005), as well as the Tree-PM method by Xu (1995) to simulate gravitational interactions. In doing so, they extract both partial dark matter halo merger trees, with only the largest-mass progenitors, and the baryonic components. While hot gas masses, black hole bulge masses and stellar masses, both total and within the bulge, are predicted well with a slight overprediction of the bulge mass for lower-mass haloes, their approach suffers from poor predictive accuracy for cold gas masses.

In a follow-up project, Kamdar et al. (2016b) extend their previous machine learning framework for hydrodynamic simulations by using the public data release of the Illustris Simulation (Vogelsberger et al., 2014; Genel et al., 2014; Nelson et al., 2015). Due to the previous success with extremely randomized trees in Kamdar et al. (2016a), which investigates decision trees, random forests, and the k-nearest neighbours algorithms for this problem, the same technique is employed again to predict M_*, M_{BH}, gas mass M_g, SFR, and $g - r$ color based on an input of dark matter halo properties without merger trees and a cosmology consistent with WMAP9 (Hinshaw et al., 2013). While recovering a similar population of galaxies via the used algorithm, a noticeable underprediction of the scatter is observed, with the possible explanation that the dark matter properties used as inputs do not contain enough information to learn the underlying physical processes. Notably, Kamdar et al. (2016b) make use of considerably more information in their inputs, namely the total dark matter subhalo mass, velocity dispersion, maximum circular velocity in the subhalo, the number of dark matter particles bound to the subhalo, and the three different spin components.

Similarly, Agarwal et al. (2018) perform experiments with decision trees, gradient-boosted trees, random forests, feed-forward neural networks, support vector

regressors, the k-nearest neighbours algorithm, and extra trees, again finding that extra trees perform best for the prediction of baryonic properties when testing models on the hydrodynamic MUFASA simulation (Davé et al., 2016). Agarwal et al. (2018) follow the aforementioned research by populating dark matter-only simulations with baryonic galaxy parameters via predicting M_*, SFR, metallicity Z, and both neutral ($M_{\rm HI}$) and molecular ($M_{\rm H2}$) hydrogen masses based on dark matter halo properties. The latter are, in this case, the dark matter halo mass, velocity dispersion, spin parameter, and nearby halo mass densities within radii at 200, 500, and 1000 kpc.

In applying this approach to the hydrodynamic MUFASA simulation introduced by Davé et al. (2016), they observe the same underprediction of scatter around the mean relations as Kamdar et al. (2016b) for the Illustris Simulation, and report that ensemble methods do not improve this result because none of the employed machine learning techniques reproduce the necessary scatter. They test a 'meta-ensemble' by averaging the outputs of various machine learning algorithms, with methods like random forests and extremely randomized trees already being, as a combination of regression trees and bootstrap aggregation, an ensemble. Stacking or boosting would be more suitable for such an attempt to leverage the strength of different algorithms. Agarwal et al. (2018) also find that adding key baryonic parameters to the inputs, for example the SFR to predict $M_{\rm HI}$, improves the obtained results, which will become an important motivation for our present work in Section 5.2.2.

Similarly, Moster et al. (2020) include the halo mass and peak halo mass, growth rate for both halo mass and peak halo mass, and the scale factor for halo mass, peak halo mass, and half-peak mass, as well as the virial radius, concentration parameter and spin parameter in their inputs to predict the stellar mass and SFR with a deep neural network using reinforcement learning. Jo & Kim (2019) make use of the MultiDark-Planck (see Klypin et al., 2016) and IllustrisTNG (see Pillepich et al., 2018) simulations to estimate baryonic galaxy properties based on dark matter haloes and showing that results are largely compatible with SAMs. Extra trees are chosen by Jo & Kim (2019), too, further strengthening the evidence for this specific type of ensemble model in the literature. These results do, of course, pose the question of why extra trees outperform random forests for this problem. While random forests introduce randomness with bootstrapping the input data, extra trees omit this randomization, but instead introduce it through randomly selecting a feature subset for a given split. This generally

leads to more diversified trees, resulting in smoother decision boundaries as the algorithm asymptotically creates continuous piece-wise multilinear functions and features less correlated errors and a lower variance (Geurts, 2006). This can be especially helpful when dealing with continuous variables. For this reason, given the previous determination of the model most suitable for the problem at hand, we focus on the use of extra trees in this work.

In related research, Lucie-Smith et al. (2018) show that random forests can be used to classify whether dark matter particles will be part of haloes in a given mass range, matching the predictions of common spherical collapse approximations. Interestingly, the application of modern machine learning to cosmological simulations is still relatively sparse in the literature, although the number of contributions is growing, with close demonstrations with regard to large-scale structures tackling the prediction of cosmological parameters with 3D simulations of the cosmic web by Ravanbakhsh et al. (2016), as well as the creation of cosmic web simulations and synthetic galaxies (Ravanbakhsh et al., 2017; Rodríguez et al., 2018; Fussell & Moews, 2019).

5.2 Methodology and data

In this section, we provide details on the methodological considerations and data sources contributing to this work. We propose extensions to the equilibrium model of galaxy evolution, covering the inclusion of free-fall time and merger trees, and describe them in Section 5.2.1. Following this, Section 5.2.2 outlines the hybrid prediction approach we create from merging our extended equilibrium model into a machine learning framework. In Section 5.2.3, we describe the cosmological simulation we make use of in this work, as well as the extraction of the data used in the presented experiments.

5.2.1 Extension of the equilibrium model

We include some minor improvements to the equilibrium model presented in Davé et al. (2012) and Mitra et al. (2015). First, we introduce a time delay between the accretion onto the halo and the accretion onto the ISM, given by the free-fall

time of the halo at the given redshift,

$$t_{\rm ff} = \sqrt{\frac{3\pi}{32G\rho}} \ , \ \text{where } \rho = 200\rho_{\rm crit,z=0}(1+z)^3. \tag{5.5}$$

More significantly, another novel addition to the equilibrium model is that the halo growth rate is now computed based on largest-progenitor merger trees. The original equilibrium model of Davé et al. (2012) employed the fitting formula from Dekel et al. (2009) for the average growth rate of M_h as

$$\dot{M}_h = 6.6 \left(\frac{M_h}{10^{12}}\right)^{1.15} (1+z)^{2.25} f_{0.165} M_\odot {\rm yr}^{-1}, \tag{5.6}$$

which was based on the assumption of a flat Universe with a fluctuation normalization parameter of $\sigma_8 = 0.8$, mass-dominated by cold dark matter and with 72% dark energy content. Mitra et al. (2017) extended this to include inflow fluctuations based on a parameterization from the Millenium simulation.

In our new version, we employ halo merger trees to compute the halo growth rate. The dark matter particle mass resolution is $\approx 10^8 M_\odot$, which is not ideal for probing the very earliest phases of galaxy growth, but for this initial proof of concept it suffices. For each halo, between each of the 114 time steps, we compute the average growth rate during that step. If the growth rate is negative, we employ 'backwards capping' and wait until the halo increases in mass at a later step, and take the average growth rate over all steps until it becomes positive.

Owing to the finite number of merger tree outputs resulting in up to hundreds of Myr between time steps, we augment this growth using the same formalism as Mitra et al. (2017) to account for short-timescale inflow fluctuations, with the limit that the fluctuations cannot grow the halo more than the amount for the entire time step. In this way, we account for both the individual long-term growth history of haloes, as well as (statistically) the fluctuations that may drive the scatter in galaxy scaling relations. This provides a more realistic description of haloes taken directly from an N-body simulations, which we later explore in Section 5.3.2.

5.2.2 Creating a hybrid prediction framework

In order to create a hybrid framework making use of both analytic formalisms and machine learning, we implement two modules, the first of which is an extra trees ensemble as introduced in Section 5.1.2. Figure 5.3 shows the training workflow of this model, starting with input values from a hydrodynamic simulation such as SIMBA (Davé et al., 2019). The depicted setup is related to the previously introduced conjecture that the inclusion of additional baryonic properties improves the accuracy of predictions for remaining properties (Agarwal et al., 2018). As N-body simulations, while relatively fast to compute, provide only dark matter properties, the equilibrium model offers a way to estimate a subset of baryonic properties as additional machine learning inputs.

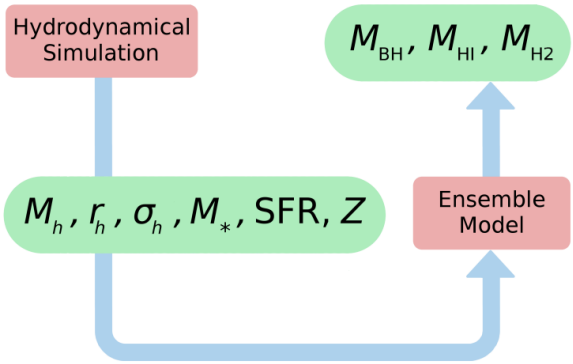

Figure 5.3 *Training process of the machine learning component of the hybrid framework presented here. The depicted workflow shows the training of an ensemble model based on the dark matter halo mass (M_h), dark matter half-mass radius (r_h), dark matter halo velocity dispersion (σ_h), stellar mass (M_*), star formation rate (SFR), and metallicity (Z) of a galaxy within a hydrodynamic simulation to predict the corresponding black hole mass (M_{BH}), neutral hydrogen (M_{HI}) mass, and molecular hydrogen mass (M_{H2}).*

In our case, the dark matter halo mass (M_h), dark matter half-mass radius (r_h), dark matter halo velocity dispersion (σ_h), stellar mass (M_*), star formation rate (SFR), and metallicity (Z) of a galaxy are used as inputs to predict the corresponding black hole mass (M_{BH}), neutral hydrogen (M_{HI}), and molecular hydrogen (M_{H2}). Apart from the hybrid approach for additional baryonic inputs, this presents an additional difference to previous research, specifically to (Kamdar et al., 2016b) not predicting neutral and molecular hydrogen but the total gas mass. Due to working on the basis of a different hydrodynamic simulation, the latter also use additional dark matter inputs, namely the three spin components, the maximum circular velocity in the subhalo, and the number of dark matter

particles bound to the subhalo, but omit r_h as an input.

Our set of inputs more closely resembles work by Agarwal et al. (2018), but the latter make use of the halo spin instead of r_h. As r_h tightly connects with halo concentration, and thus the halo formation time which dominates the scatter in the $M_* - M_h$ relation, we expect that r_h provides an orthogonal dimension to the other properties. Additionally, they include nearby halo mass densities within radii at 200, 500, and 1000 kpc, which translates to environmental data, and do not predict $M_{\rm BH}$. Notably, though, they use arrays of M_h for both the present and the five immediately preceding snapshot in their inputs, thus including merger tree information directly into the machine learning model. This consideration is later adapted into the full merger tree experiments in Section 5.3.2.

Importantly, this means that our approach runs on a reduced amount of information when compared to previous research on methods using only machine learning, with only three basic dark matter inputs fed into the framework. This is partly counterbalanced by the inclusion of merger tree data in the equilibrium model introduced in Section 5.1.1, featuring the extensions described in Section 5.2.1. As shown in Figure 5.4, the merger tree-based initial halo mass (M_{h_0}) of a galaxy, as well as initial and final redshifts (z_0, z) for a given merger tree, are fed into our modified version of the equilibirum model to produce the corresponding stellar mass (M_*), star formation rate (SFR), and metallicity (Z). These outputs, together with the previously described dark matter inputs, are then used as inputs to the trained ensemble model, predicting the full set of baryonic properties of a given galaxy. The baryonic properties generated by the equilibrium model are used both as an input and as target variables of the extra trees ensemble to further refine the results based on the combined inputs. Stepwise, our model thus works in the following way:

- Train the machine learning model on Simba data
  - Inputs: M_*, SFR, Z, M_h, r_h, σ_h
  - Outputs: M_*, SFR, Z, $M_{\rm BH}$, $M_{\rm HI}$, $M_{\rm H2}$
- Derive baryonic outputs of the equilibrium model
  - Inputs: M_{h_0}, z_0, z
  - Outputs: M_*, SFR, Z
- Predict baryonic quantities with the trained model

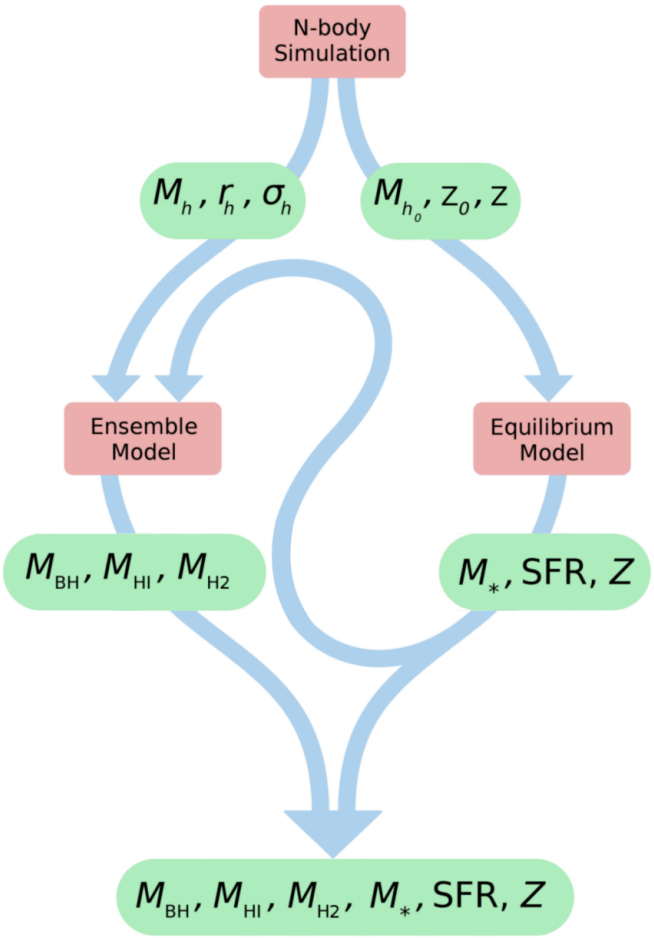

Figure 5.4 *Prediction with the full hybrid analytic and machine learning framework presented in this work. In the shown workflow, along the right path, the merger tree-based initial halo mass (M_{h_0}) of a given galaxy, as well as initial and final redshifts (z_0, z) for the same merger trees, are fed into our modified version of the equilibirum model to produce the corresponding stellar mass (M_*), star formation rate (SFR), and metallicity (Z). These values, along the left path and together with the dark matter halo mass (M_h), half-mass radius (r_h), and dark matter halo velocity dispersion (σ_h), are then used by the previously trained ensemble model to predict the black hole mass ($M_{\rm BH}$), neutral hydrogen ($M_{\rm HI}$), and molecular hydrogen ($M_{\rm H2}$), as well as updated outputs of the values predicted by the equilibrium model.*

In doing so, we make use of the equilibrium model to create the additional baryonic inputs necessary to run a machine learning model relying on them, while only requiring dark matter properties commonly found in comparatively fast-running N-body simulation as inputs for the completed framework.

5.2.3 Data from SIMBA

The dataset is extracted from the m100n1024 run of the SIMBA simulation (Davé et al., 2019), which has a volume of 100 h^{-1}Mpc with 1024^3 dark matter particles and 1024^3 gas elements. It assumes a Planck Collaboration XIII (2016) concordant cosmology of $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $\Omega_b = 0.048$, $H_0 = 68\ km\ s^{-1}\ Mpc^{-1}$, $\sigma_8 = 0.82$, and $n_s = 0.97$. We refer to Davé et al. (2019) for its detailed physical models for baryon processes. This simulation starts at $z = 249$, with ~ 150 outputs spanning from $z = 30$ to zero.

Haloes are identified on the fly by a 3D friends-of-friends algorithm within GIZMO, with a linking length set to 0.2 times the mean inter-particle spacing and without the consideration of subhaloes. We identify galaxies with a YT-based package, CAESAR[1], which uses a 3D friends-of-friends galaxy finder that assumes a spatial linking length of 0.0056 times the mean inter-particle spacing (equivalent to twice the minimum softening length). Black holes that are gravitationally bound, as well as gas elements with a minimum SF threshold density of $n_H > 0.13$ H atoms cm^{-3}, are attached to the host galaxy with the same linking length. We treat the most massive black hole in the galaxy as the central one, with mass $M_{\rm BH}$. The neutral and molecular hydrogen of the galaxy are calculated based on its gas properties; these are computed self-consistently in SIMBA assuming self-shielding from Rahmati et al. (2013) and atomic/molecular separation based on the subgrid model of Krumholz & Dekel (2012).

Galaxies and haloes are cross-matched in post-processing within CAESAR, and the most massive galaxy close to the halo minimum potential center is assigned as the central galaxy. The merger tree of a galaxy, instead of a halo, is built by tracking the unique star particle IDs, while the most massive progenitor is treated as the main progenitor of the descent, which we use for tracking galaxy evolution here. As there is a one-to-one connection between the central galaxy and its host halo, it is simple to have the host halo information attached to its merger tree. We constrict our data to final halo masses of $\log_{10}(M_h) \in [11, 14]$ and make sure logarithms of values do not lead to unsuitable infinities, but do not make use of any ways of further restricting the dataset that could lead to better fits, either for the data extracted from the SIMBA simulation or our framework's predictions, in order to present generalisable results. In doing so, we generate a dataset of $30,555$ haloes with corresponding merger trees and baryonic properties to create

[1] `https://github.com/dnarayanan/caesar`

training and test sets from, as well as a separate dataset for the baryon cycling parameter optimization via MCMC, as described in Section 5.3.1.

5.3 Experiments and results

In this section, we describe the experiments performed to evaluate the performance of our approach to baryonic property insertion, both in a half-way setup and a full-scale experiment the extensions proposed and described in this work. A preliminary experimental run is covered in Section 5.3.1, without merger trees and dark matter halo mass variability, but including a free-fall time modification of the equilibrium mode and relying on splining outputs of the equilibrium model to build a relation between initial and final halo mass in lieu of using largest-progenitor merger trees. After including the use of merger trees in the equilibrium model, we present the final results in Section 5.3.2.

5.3.1 Preliminary splining and free-fall time

In the first step of our experiments, we include the free-fall time, as described in Section 5.2.1, into the equilibrium model, but omit merger trees for now. We represent the baryon cycling parameters by nine free variables to assess their behavior with halo masses and redshifts (see also Mitra et al., 2015, 2017),

$$\eta = \left(\frac{M_h}{10^{\eta_1+\eta_2(1+\min(z,2))}}\right)^{\eta_3}, \tag{5.7}$$

$$t_{\rm rec} = \tau_1 t^{\tau_2}\left(\frac{M_h}{10^{12}}\right)^{\tau_3}, \tag{5.8}$$

$$\zeta_{\rm quench} = \begin{cases} \left(\frac{M_h}{M_q}\right)^{\zeta_1(1+z)}, & \text{if } M_h > M_q \\ 1, & \text{else} \end{cases}, \tag{5.9}$$

where η is the ejective feedback parameter, $t_{\rm rec}$ is the wind recycling time at a given time t, and $\zeta_{\rm quench}$ is the quenching feedback parameter, with a corresponding quenching mass $M_q = 10^{12}\ \mathrm{M}_\odot(\zeta_2 + \zeta_3 z)$. We note that the parameterizations have changed slightly from Mitra et al. (2015), as these were found to give more reasonable extrapolations to higher redshifts.

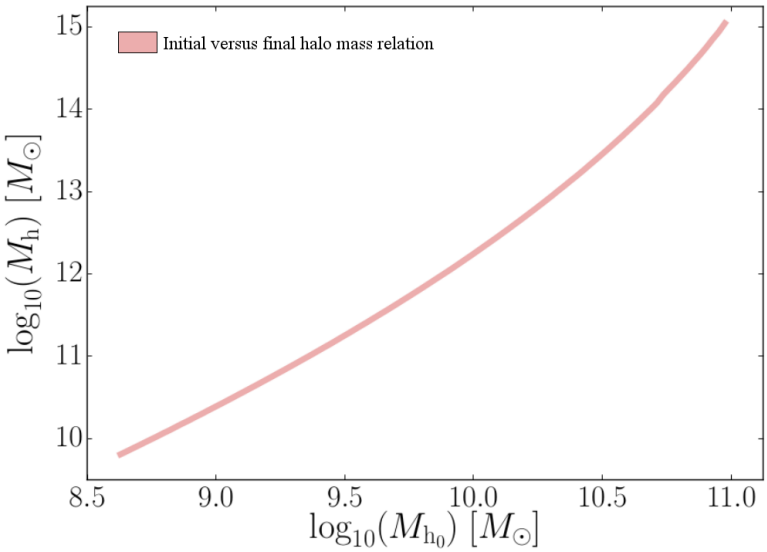

Figure 5.5 *Splining of the relation between initial halo mass (M_{h_0}) and final halo mass (M_h) of galaxies in the equilibrium model. Based on the halo masses extracted from the* Simba *simulation and a redshift of $z = 10$, 100 equidistantly spread initial halo mass values are fed into the model to cover the corresponding final halo mass range, with the result being splined to approximate a continuous look-up function.*

As in Mitra et al. (2015), these free parameters $\{\eta_1, \eta_2, \eta_3, \tau_1, \tau_2, \tau_3, \zeta_1, \zeta_2, \zeta_3\}$ are then constrained using a Bayesian MCMC analysis with $N = 500$ walkers against recent observations on the $M_* - M_{\rm h}$ relation by Behroozi et al. (2019), $M_* - Z$ relation (combined from Andrews & Martini, 2013; Zahid et al., 2014; Ly et al., 2016; Sanders et al., 2018), and M_*–SFR relation by Speagle et al. (2014) at redshifts $z \in 0, 1, 2$. The best-fit values we obtain from the resulting analysis are listed in Table 5.1, and are subsequently used for the equilibrium model component of our framework. The match between our model predictions and observed datasets are quite similar to earlier results on this MCMC fitting approach by Mitra et al. (2015) and Mitra et al. (2017), and we refer the reader to those papers for a more detailed description.

Table 5.1 *Baryon cycling parameters for the equilibrium model used in the analytic formalism part of our framework. The best-fit values are achieved through a Bayesian MCMC estimation for ejective feedback parameters (η), wind recycling parameters (τ) and quenching feedback parameters (ζ).*

η_1	η_2	η_3	τ_1	τ_2	τ_3	ζ_1	ζ_2	ζ_3
10.822	0.405	−1.517	3.184	−2.161	−1.381	−0.229	1.122	0.007

We choose a redshift of $z = 10$ as a baseline value and feed initial halo masses

M_{h_0} into the equilibrium model using the Dekel et al. (2009) mean accretion rate (not including fluctuations) to generate final halo masses covering values on a closed interval $\log_{10}(M_h) \in [11, 14]$ for later evaluation against SIMBA data. Using the mean accretion rate allows a one-to-one mapping of initial and final halo masses that is on average correct, and enables us to identify the starting halo mass for any halo in the merger tree. Figure 5.5 shows the resulting spline as a continuous look-up function for initial versus final halo mass. For training and test sets, we split our dataset in a four-to-one ratio via random subsampling without replacement, resulting in a training set of 24,444 and a test set of 6,111 examples.

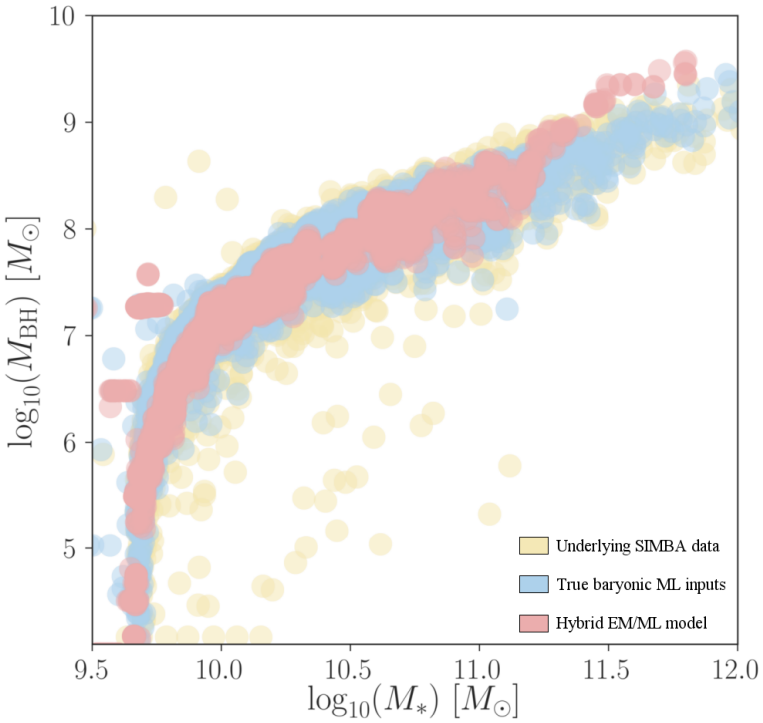

Figure 5.6 *Scatter plot for the restricted experimental run, with reduced scatter typical of machine learning approaches. The figure shows results for stellar mass (M_*) versus black hole mass ($M_{\rm BH}$), with true* SIMBA *data plotted in yellow, results of an extremely randomized tree ensemble with additional baryonic inputs shown in blue, and results for the preliminary test of the hybrid analytic and machine learning framework without some of the extensions introduced in this work shown in red.*

We train an extra trees ensemble as described in Section 5.1.2 and Section 5.2.2 and shown in Figure 5.3, feeding both dark matter properties and baryonic properties predicted via the equilibrium model into it. Specifically, this means that an array of $\{M_{h_0}, z_0, z\}$ values are used by the equilibrium model to predict

the corresponding array of $\{M_*, \text{SFR}, Z\}$ values as shown in Figure 5.4, with $z_0 = 10$, $z = 0$, and M_{h_0} predicted from the aforementioned splining of SIMBA halo masses. These output values, together with SIMBA values $\{M_h, r_h, \sigma_h\}$, are then used by the ensemble to predict $\{M_{\text{BH}}, M_{\text{HI}}, M_{\text{H2}}\}$. In addition, we also train two additional extra trees ensembles for the purpose of comparison. First, we train one model that is allowed to use true underlying target values from SIMBA instead of equilibrium model outputs in order to gauge the effect the inclusion of these partial baryonic properties has on the quality of the results. Effectively, this machine learning-only setup emulates the assumption of perfect predictions by the equilibrium model. Secondly, we also train one model that disregards any baryonic input information, predicting solely based on SIMBA values $\{M_h, r_h, \sigma_h\}$ as a pure machine learning baseline.

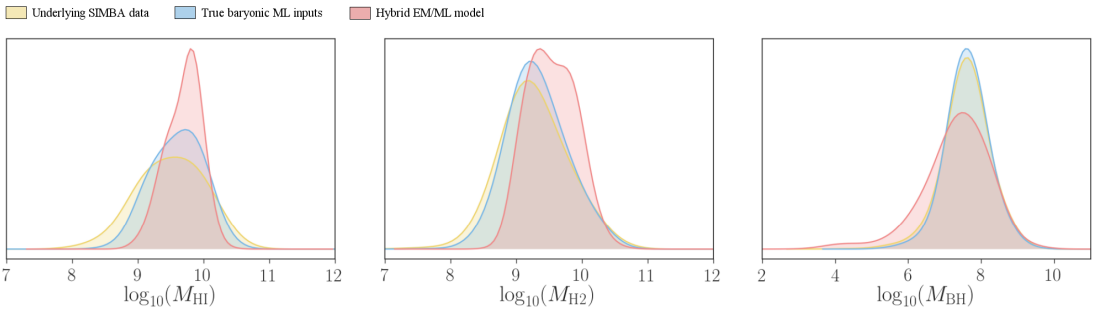

Figure 5.7 *Density plots for the restricted experimental run. The panel show results for neutral hydrogen (M_{HI}), molecular hydrogen (M_{H2}), and black hole mass (M_{BH}). In all three panels, the true underlying* SIMBA *target data is plotted in yellow. Results of an extremely randomized tree ensemble with additional true baryonic inputs from* SIMBA*, mimicking a hypothetical 'perfect' equilibrium model by receiving the actual target values for these inputs, are shown in blue and lead to a green tint when fitting the underlying data. Lastly, results for the preliminary test of the hybrid analytic and machine learning framework without some of the extensions introduced in this work are shown in red.*

Due to the limited information and fixed redshift value for looking up initial halo masses from a spline, we can expect results to feature some irregularities. Figure 5.6, depicting a two-dimensional plot for M_* and M_{BH} of this restricted setup as an easy-to-eyeball combination with a noticeable bend, demonstrates this by showing a lack of scatter in M_{BH} for larger values of M_*. For an closer visual analysis of the results, taking a look at separate variables can be useful to determine the level of overconstrained or underconstrained distributions and eventual missed or superfluous multimodal features. Density plots of all output values are shown in Figure 5.7, and are further discussed in Section 5.4. The data

shown in Figure 5.6 can be summarized as follows, and follows the same color scheme for subsequent Figures 5.7 and 5.8 further below:

- Yellow: True underlying SIMBA data, meaning the target values taken directly from the cosmological simulation instead of the output of a predictive model, as a comparison baseline
- Blue: Machine learning-only results when, as an idealized scenario, feeding true underlying SIMBA data for $\{M_*, \text{SFR}, Z\}$ into the extra trees ensemble instead of using equilibrium model estimates, thus mimicking a hypothetical 'perfect' equilibrium model
- Red: Hybrid model results, predicting $\{M_*, \text{SFR}, Z\}$ via the equilibrium model and using the extra trees machine learning model to estimate the full set of baryonic target properties

While visual inspections provide a reasonable overview, stricter statistical validation is necessary to assess the results. For this purpose, we calculate the coefficient of determination,

$$R^2(x, y) = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{5.10}$$

with x and y as the true target values and their predictions generated by our framework, respectively, representing the proportion of the dependent variable's variance that can be explained through the independent variable. As such, it offers a way to quantify the performance of a model's replication of observed outcomes and, while having an upper limit of one, features no lower limits for models that perform arbitrarily badly. Specifically, negative values for non-linear functions, as is the case in our work, mean that the data's mean provides a better fit than the predictions in question. In addition, we also calculate Pearson's correlation coefficient,

$$\rho(x, y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2(y_i - \bar{y})^2}}, \tag{5.11}$$

as a measure that can be employed together with R^2 (see, for example, Agarwal et al., 2018, for related research) to provide additional insight. It measures

the linear correlation two variables, in our case the true target values and the framework's predictions, and is limited to $\rho(x, y) \in [-1, 1]$.

Table 5.2 *Statistical validation for the restricted experimental run. The table lists the coefficient of determination (R^2) and Pearson's correlation coefficient (ρ) for different setups. The column denoted as 'True' shows results for the prediction of neutral hydrogen (M_{HI}), molecular hydrogen (M_{H2}), and black hole mass (M_{BH}) when feeding true underlying* Simba *target values for stellar mass (M_*), and star formation rate (SFR) into the machine learning model, while the column under 'ML' shows results for excluding M_*, SFR, Z from the inputs, predicting only based on dark matter halo information. The column under 'Hybrid' shows the results when using the equilibrium model without merger tree information for these baryonic inputs, and predicting these as well, for an invariant initial redshift of $z = 0$ and initial halo masses predicted from splined equilibrium model results.*

	True		**ML**		**Hybrid**	
Variable	R^2	ρ	R^2	ρ	R^2	ρ
M_{HI}	0.5560	0.7497	0.4125	0.6538	0.2563	0.5343
M_{H2}	0.7743	0.8800	0.7456	0.8635	0.5457	0.7631
M_{BH}	0.7207	0.8832	0.6354	0.8765	0.5980	0.8517
M_*	–	–	–	–	0.7286	0.9621
SFR	–	–	–	–	0.7266	0.8663
Z	–	–	–	–	−7.4543	0.4847

The results for these calculations are listed in Table 5.2, with the column under 'True' corresponding to a machine learning model receiving true underlying M_*, SFR, and Z values from Simba (colored yellow in Figures 5.6 and 5.7 further above), and the column under 'ML' to predictions based on feeding only dark matter halo features into the machine learning model while disregarding the equilibrium model (colored blue in Figures 5.6 and 5.7 further above). The column under 'Hybrid' refers to predictions using the equilibrium model for these baryonic inputs, albeit with a fixed redshift and initial halo masses M_{h_0} produced by the continuous look-up function via a spline described in this section and shown in Figure 5.5 (colored red in Figures 5.6 and 5.7 further above).

5.3.2 Inclusion of merger tree information

In the next step, we include the remaining extension of the equilibrium model to run a full test suite for our approach. This means that merger tree information described in Section 5.2.1 is incorporated into the hybrid model, both in terms of

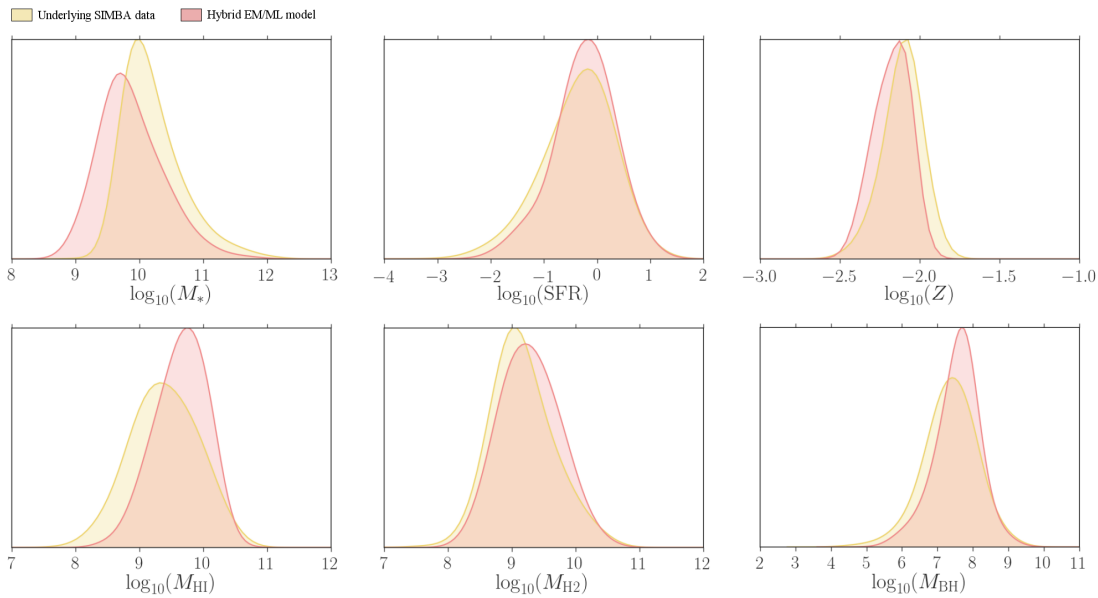

Figure 5.8 *Density plots for the full experimental run including merger trees. The panel show results for stellar mass (M_*), star formation rate (SFR), metallicity (Z), neutral hydrogen (M_{HI}), molecular hydrogen (M_{H2}), and black hole mass (M_{BH}). In all six panels, the true underlying* Simba *data is plotted in yellow,*

and results for the test of the hybrid analytic and machine learning framework with all extensions introduced in this work are shown in red.

the internal use of merger trees by the equilibrium model and the five most recent halo masses as part of the inputs as in Agarwal et al. (2018). For this experiment, we make use of the same dataset as previously, with training and test set examples numbering $24,444$ and $6,111$, respectively, to enable an as-accurate-as-possible measurement of the impact that the inclusion of merger tree data, specifically the associated halo masses, have on the hybrid framework.

As previously in Section 5.3.1, we use an extra trees ensemble trained on the Simba-extracted data described in Section 3.1, as shown in Figure 5.3, and use the already MCMC-fitted baryon cycling parameters. Unlike in these preliminary experiments, however, we feed full largest-progenitor merger trees instead of just the respective initial halo mass estimates into the equilibrium model, allowing the model to steer more closely along each dark mater halo's true evolutionary history. As the prediction of Z stays too constrained when making use of equilibrium model outputs, we instead predict the property directly from the other outputs. Fitting tailored parameters for the given larger dataset could possibly provide a more accurate fit, but this would incur a significantly higher computational cost due to the increased wall time of including merger trees. In addition, using pre-

Table 5.3 *Statistical validation for the full experimental run including merger trees. The table lists the coefficient of determination (R^2) and Pearson's correlation coefficient (ρ) for different setups in alphabetically indicated columns. The column under 'True' shows results for the prediction of neutral hydrogen ($M_{\rm HI}$), molecular hydrogen ($M_{\rm H2}$), and black hole mass ($M_{\rm BH}$) when feeding true underlying* Simba *target values for stellar mass (M_*), star formation rate (SFR), and metallicity (Z) into the machine learning model. The column under 'Hybrid' shows results for the prediction of the same properties as well as M_*, SFR, Z when using the updated equilibrium model that includes merger tree information.*

	True		**Hybrid**	
Variable	R^2	ρ	R^2	ρ
$M_{\rm HI}$	0.4873	0.7110	0.3650	0.6533
$M_{\rm H2}$	0.7826	0.8863	0.5470	0.7513
$M_{\rm BH}$	0.7774	0.8907	0.7143	0.8815
M_*	–	–	0.8229	0.9478
SFR	–	–	0.7395	0.8728
Z	–	–	0.2167	0.6641

fitted parameters relying on a previous and smaller dataset also provides a use case for further applications, as future research is planned to apply our framework to N-body simulations that are unable to provide the target values for these fitting procedures.

Density plots of all output values are shown in Figure 5.8, and are further discussed in Section 5.4. Again, we compare a hypothetical 'perfect' equilibrium model by using true underlying Simba values instead of mode-outputted values for M_*, SFR, and Z, and find that the respective statistical key performance indices remain at a very similar level to the restricted experimental run, with only a slight decrease and increase in accuracy for $M_{\rm HI}$ and $M_{\rm BH}$, respectively. The results for our statistical validation are listed in Table 5.3, with the column under 'True' corresponding to a machine learning model receiving true underlying M_*, SFR, and Z values from Simba (colored yellow in Figure 5.8 further above) and the column under 'Hybrid' to predictions using the equilibrium model for these baryonic inputs while making use of merger tree information (colored blue in Figure 5.8 further above).

5.4 Discussion

The hybrid nature of our approach, making use of both an established analytic formalism for galaxy evolution and machine learning, provides a number of advantages by combining, in an adage to timeless music, 'the best of both worlds'. We derive a subset of baryonic parameters corresponding to dark matter haloes from the equilibrium model, an established formalism that we improve upon with a number of extensions. We include the gravitational free-fall time for mass accretion, as described in Section 5.2.1, to further refine the model's capabilities of accurately tracing the evolution of haloes and their properties. In addition, we enable the equilibrium model to be fed complete largest-progenitor merger trees with corresponding redshift values, thus obviating the need to estimate halo masses at each time step to let the model follow the underlying mass evolution more closely. In adding these extension, one contribution of our work is the improvement of an established analytic formalism, making said formalism more suitable for fine-grained estimations and its application to N-body simulations and their merger trees.

Similarly, for the second half of our work, we merge the extended equilibrium model into a machine learning framework comprised of an ensemble of extra trees, as discussed in Section 5.2.2, as the latter has been shown to deliver the best results for the problem at hand (Ravanbakhsh et al., 2017; Agarwal et al., 2018). With increased dataset sizes in future research, which allows for larger training sets to fit machine learning models, we expect neural network models to catch up to, and surpass, the performance of tree-based ensembles. This can be realized with both larger-scale simulations and the combination of existing simulations, which we discuss further below. We contribute to the expanding literature on machine learning methods for cosmological simulations in general, and its application to, and analysis of, N-body and hydrodynamic simulations for baryonic property prediction in particular.

As these simulations are of crucial importance for several research areas in cosmology such as AGN feedback, survey analysis, covariance estimation, large-scale structure, and small-scale matter power behavior, the resulting speed-up of going from N-body simulations to full hydrodynamic property sets per dark matter halo is of importance in the context of ever-growing simulation sizes and the analysis of upcoming surveys like Euclid and LSST (Racca et al., 2016; Ivezić et al., 2019).

In a first step, we omit the inclusion of merger tree information and compare our hybrid approach, using the equilibrium model, to a hypothetical 'perfect model', for which we simply use the true underlying properties normally derived from the equilibrium model as inputs. Given that the purpose of this part of the preliminary experimental run was to confirm the beneficial effect of including baryonic values in the input when predicting M_{HI}, M_{H2}, and M_{BH}, the values for these properties fed into the machine learning model are the same as the target values, meaning that the model is given an unrealistic 'ideal world' advantage. Even with perfect inputs for these three parameters, the model still performs slightly worse for M_{HI} when compared to M_{H2} and M_{BH}, as shown in Table 5.2.

When compared to this hypothetical perfect scenario, a machine learning-only approach still performs reasonably well, confirming previous research on this topic. While we observe decreased accuracies for all assessed properties, the largest drop happens, again, for M_{HI}, with the remaining parameters being less affected and M_{BH} experiencing the smallest decrease in accuracy, pointing toward a diminished reliance on baryonic inputs. Figure 5.6 also successfully recovers the $M_* - M_{\mathrm{BH}}$ relation, including the sharp drop at lower stellar masses, as a litmus test of the usability of our approach. The slight overprediction at higher stellar masses can be explained by more sparse data in the SIMBA dataset.

In the same preliminary experiment, we also establish that our hybrid approach underperforms, albeit still at a level useful for the completion of N-body simulations, when not making use of any baryonic inputs. This does, of course, not come as a surprise, as the preliminary experiment gauges initial halo masses based on a simple spline function fitted to the reduced equilibrium model itself, as shown in Figure 5.5, and makes use of an invariant assumed initial redshift for this purpose. We also observe that the prediction of metallicity proves especially difficult in this setup.

We then incorporate the fully extended equilibrium model into our framework to test for the impact of including merger tree information. The results, depicted in Figure 5.8, show that our approach recovers property distributions to a very reasonable degree, although we observe a slight underprediction of M_* and overpredictions, especially of M_{BH} and M_{HI}, which can be explained by the resolution challenge induced through a lower limit of $M_* \approx 10^{9.5} M_\odot$ in SIMBA. While distribution recoveries are a reliable way to ballpark the overall reproduction of values, it is important to statistically verify the results for a complete overview. The results in Table 5.3 confirm that, for hypothetical perfect

inputs of M_*, SFR, and Z, the results stay virtually the same for the assessed properties. Notably, the inclusion of the five last halo masses in the machine learning inputs does not improve the outcome, which further confirms that these properties do not rely on halo masses as much as the other inputs when predicted through a machine learning model.

For both $M_{\rm HI}$ and $M_{\rm H2}$, the results are slightly worse when compared to a machine learning-only approach as listed in Table 5.2, while the hybrid model outperforms on the prediction of $M_{\rm BH}$. Compared to the reduced equilibrium model without merger tree information, the predictions also outperform on both $M_{\rm HI}$ and $M_{\rm BH}$, while the accuracy for $M_{\rm H2}$ stays at the same level. For the baryonic properties, M_*, SFR, and Z, the extended setup clearly outperforms the reduced approach, with only the SFR remaining at a very similar, but satisfactory, accuracy. These results are especially useful as they not only demonstrate the viability of our hybrid approach, but also show that the use of baryon cycling parameters fitted on a reduced equilibrium model are viable on the extended version, confirming the robustness of the formalism.

While there is some degradation in the coefficient of determination, R^2, and Pearson's correlation coefficient, ρ, the degradation caused by using the equilibrium model as an intermediary is modest, for instance for HI with $\rho \approx 0.71$ in the ideal scenario with true baryonic inputs and $\rho = 0.65$ when using the equilibrium model. HI is, in fact, the most difficult quantity to predict, while $M_{\rm BH}$ ($\rho \approx 0.88$) and M_{H2} ($\rho \approx 0.75$) are predicted with substantially higher fidelity. The equilibrium model also does a good job of reproducing the original SIMBA values of the input parameters M_* and SFR, with $\rho \approx 0.95$ and $\rho \approx 0.87$, respectively, although the metallicity Z is reproduced less effectively at $\rho \approx 0.66$. As the latter is predicted from the other inputs, as opposed to using the equilibrium model outputs, the accuracy is expectedly lower when compared to the other parameters, but still follows the intended distribution closely, as shown in the histograms in Figure 5.8. Overall, this shows that the approach of using the equilibrium model to 'pre-predict' a subset of baryonic quantities, which can be employed to improve training, is an effective approach toward more accurately bridging hydrodynamic simulations and large-volume N-body simulations.

For planned follow-up research, and aside from further investigations of the equilibrium model to improve the calculation of metallicities in the context of merger tree information, we propose the use of current developments in machine learning, especially in terms of boosting methods, active learning, and

meta-ensembles, to alleviate this remaining issue. More generally, our work demonstrates that the hybrid model using analytic formalisms for baryonic properties outperforms in some areas, but underperforms in others. The use of such tailored meta-ensemble methods to base the weights placed on pure machine learning and hybrid approaches on the individual inputs could further improve the results, but would go beyond the scope of this work and is planned for future research. For an initial test, to evaluate the use of the equilibrium model against a pure machine learning approach, we provide the same type of extra trees ensemble with the entire set of redshift values and the evolutionary history of masses for each halo and find that such a replacement leads to little difference for the stellar mass ($\rho \approx 0.97$) and significantly underperforms for the star formation rate ($\rho \approx 0.50$), but outperforms on the metallicity ($rho \approx 0.79$) when compared to the results in Table 5.3. For this reason, we plan on taking pure machine learning approaches into consideration specifically with regard to the problematic prediction of the metallicity, as well as for further investigations into correlations and informativeness in the context of machine learning models.

As with all research targeting specific datasets, both previous work on baryonic property prediction and this chapter are limited to the recreation of specific simulations they are working with. Moreover, while machine learning excels at generalization in terms of interpolation, meaning successful prediction within the value ranges presented by training datasets, extrapolation beyond these ranges is considerably more difficult and an active topic of research (Webb et al., 2020). These issues could, for example, take the form of attempting to predict the described hybrid emulator to populate haloes from a larger box that happen to be more massive than the ones present in the SIMBA dataset used in our work, which presents a limitation of common machine learning approaches in general as well as our model in particular.

Planned follow-up research will, for this reason, target the combination of data sources by assessing the compatibility of a variety of existing hydrodynamic simulations, allowing the framework to train on different pathways taken to model our Universe. Such a combination of simulations will also allow for larger training sets to fit more complex machine learning models, thus enabling the research community to revisit, for example, deep-layered neural network architectures for this purpose. Future improvements in cosmological simulations will further enhance the realism, which means that this framework can be reapplied and, thus, updated to develop into an increasingly robust predictive model.

In addition, sourcing simulations covering a larger range of values for the relevant features will help to alleviate the challenge of extrapolation mentioned above. For further follow-up research, we also suggest to assess the performance of hybrid approaches on haloes purposefully beyond the dataset ranges. Due to the modular nature of our two-part approach, both the machine learning component and the equilibrium model can also be replaced with novel developments in either of these areas in case they prove to be more suitable for the problem at hand. In this context, highly parallelized parameter estimation methods as described in Section 1.2.3, which draw from recent developments in both statistics and machine learning, will prove useful in potential follow-up research to constrain models relying on large datasets and likelihood calculations that require the running of computationally costly code. For this reason, we propose the use of these methods to investigate both the fast tuning of our model to given datasets and the application of such methods to simultaneous parameter optimization for cosmological simulations.

An additional limitation is the narrower dynamic range in simulations, which means that the population of large-column simulations requires extrapolations beyond the dataset. As previously pointed out by Agarwal et al. (2018), this shortcoming could be tackled by using zoom simulations for dwarfs and galaxy clusters to retrieve anchor points for small and large halo masses (Cui et al., 2016). Similarly, the focus on entire dark matter haloes using a largest-progenitor merger tree represents another limitation. The extended equilibrium model does, at this stage, not include satellite galaxies due to both data-side and model-side challenges. The former relies on the detection of subhaloes and numerical resolution in simulations, which often poses an issue, while the latter requires the inclusion of complete merger trees with all relevant progenitors in a suitable format, as well as the extension of the latter in the internal equilibrium model calculations.

5.5 Summary

In this chapter, we introduce a hybrid framework using both machine learning and analytic components to predict baryonic galaxy properties based on dark matter halo information. In doing so, we lay the groundwork for a new class of merged approaches between analytic formalisms and machine learning. For this purpose, we extend the equilibrium model, a feedback-based description of the evolution of

the stellar, gas, and metal content of galaxies, by including the ability to process largest-progenitor merger trees of dark matter haloes, as well as the free-fall time within haloes themselves. We then feed the partial baryonic outputs, together with dark matter properties, into a machine learning model that connects these properties to a full set of baryonic properties with stellar and black hole mass, neutral and molecular hydrogen, star formation rate, and metallicity, trained on the Simba cosmological hydrodynamic simulation. This framework is then able to predict with reasonable, though far from perfect, accuracy when compared to the true values taken from Simba.

We first introduce several modifications to the equilibrium model as described by Mitra et al. (2017), including a slightly updated parameterization of the baryon cycling parameters and the introduction of a delay time between halo and galaxy accretion given by the free-fall time. These minor updates improve the physical realism, but do not substantially change the goodness of fit versus observations. Next, we modify the equilibrium model to accept halo growth rates taken from merger trees, and use the equilibrium model to predict the baryonic properties M_*, SFR, and Z.

We feed this information, in addition to dark matter halo information, into an extremely randomized trees machine learning algorithm. The outputs of this process are various physical parameters that are not predicted directly by the equilibrium model. Here, we examine $M_{\rm BH}$, M_{HI}, and M_{H2}, and train the extra trees model on Simba data. This now extends the predictive power of our framework to these additional quantities that are not directly predicted by the equilibrium model, using information from full hydrodynamic simulations. It is trivial to extend this to predicting other desired quantities, so long as they are outputs of a hydrodynamic simulation such as Simba for training.

We test this approach by comparing two cases: In the first case, we input the true values for $\{M_*$, SFR, $Z\}$ values from Simba, and then predict the remaining quantities; this is effectively the ideal case for machine learning predictions, since the extra baryonic inputs are taken directly from the hydrodynamic simulation itself. In the second case, we use the equilibrium model to obtain these properties from merger trees and associated redshifts, and then use those to predict the remaining quantities, namely $M_{\rm BH}$, M_{HI}, and M_{H2}. Generally, we find that the second case has correlation coefficients that are not greatly degraded from the ideal case, indicating that the equilibrium model can provide a useful intermediary to improve baryonic property predictions within haloes at minimal computational

cost, with the latter being a concern in modern cosmology for both temporal and environmental reasons. As with simulations in general, one thing that applies to this approach is that the framework learns to predict based on such a simulation, meaning that it learns not 'true' physics but rather how properties are related in that simulation. Parameter inference thus also relates to how the simulation in question evolves galaxies, which should be taken into consideration when using our framework for such investigations.

In the future, we aim to extend this work in a three-pronged approach targeting all components of our framework: The equilibrium model is planned to include full merger trees with smaller progenitors, as well as satellite galaxies and black holes, to further push the model's accuracy and its predictive power for metallicities. On the machine learning side of our framework, we intend to make use of meta-learning approaches to weight input variables or the analytic and machine learning modules themselves. Lastly, advances in observational data and cosmological simulations allow for the equilibrium model to be fitted more accurately, and for the machine learning model to be trained on a wide variety of simulation approaches. As the equilibrium model provides reliable Bayesian posteriors, planned follow-up research will also investigate constraints on dark matter.

In addition, we plan to investigate modeling the correlated scatter in galaxy quantities more accurately. In particular, the division between quenched and star-forming galaxies, as well as the associated trends in gas content and other properties, has often been challenging to recover using machine learning. These issues suggest that perhaps a combination of methodologies including both classification and regression may be more optimal. Alternatively, different machine learning approaches such as generative adversarial networks may be more effective at picking out the more subtle trends in the galaxy population.

The resulting framework will, in principle, have a wide applicability for both cosmology and galaxy evolution studies, including populating dark matter-only simulations, examining the physical constraints on baryon cycling parameters, and investigating environmental trends in the galaxy population such as assembly bias. Machine learning applied to galaxy evolution is still a developing field, but offers great promise for delivering the most accurate mock universes incorporating information from both high-resolution hydrodynamic simulations and large-volume N-body simulations.

Chapter 6

Conclusion

The development of this thesis was, at least in our and thus very biased opinion, an exciting journey across various areas, both in terms of cosmology and methodology, and with machine learning and statistical inference at its heart. As such, we hope that it provides a holistic slice through the beginning era of 'machine cosmology' as the rapidly expanding use of novel inference methods in the field. We focus on the dark sector of our Universe, and proceed from the largest scale to more granular challenges, starting with cosmological parameter estimation and the dark energy equation of state, and subsequently visit the detection of cosmic voids and troughs in the large-scale structure, and, finally, the prediction of baryonic properties in modern cosmological simulations.

At the turn of the millennium, the field of cosmology experienced a stark shift toward Bayesian methodology, which quickly became a staple of cosmological inference. Chapter 2 heavily relies on Bayesian nonparametrics in its methodological approach, and we present a new iterative cosmological parameter estimator with embarrassing parallelism as one of its core features. We test our approach on a modern supercomputing architecture to constrain cosmological parameters based on DES Year 1 data, and show that our method scales well into a sufficient number of dimensions, and benefits greatly from large numbers of cores to parallelize over in terms of its total runtime. The development of this new approach is primarily motivated by the need for fast and robust parameter estimators in the context of upcoming surveys such as ESA's Euclid and the Vera C. Rubin Observatory's LSST, which will extend the current requirements in terms of the number of dimensions and the data provided.

More broadly, modern cosmology is an inherently statistical field due to measurements that rely on a large number of observations, and both Chapter 3 and Chapter 4 reflect this importance. In the former chapter, we introduce a novel randomized algorithm to create smooth random curves following customizable constraints to mimic redshift-dependent deviations from the dark energy equation of state, and run a full cosmological analysis pipeline with SN Ia data to assess the detectability against the standard model. One of the most interesting findings of that chapter is that larger deviations from the standard model do not automatically translate to an improved statistical detectability. This shows that, for analyses based on SN Ia data, physics beyond the standard model might hide in plain sight, and the ruling out of deviations from the cosmological constant through statistical methodology has to be treated with the necessary caution. Chapter 4 extends the subspace-constrained mean shift algorithm, a method for density ridge estimation, to create a new tool for cosmic trough detection in current surveys, followed by the application of our approach to DES Year 1 mass density maps. Like Chapter 2, this work is part of the preparation for upcoming missions, as future lensing surveys will provide a better access to investigations of alternative gravity theories through empty regions on the sky.

During the course of the development of the presented work, cosmology experienced yet another shift in, or addition of, methodology. Machine learning rose to become a full-fledged part of the cosmological tool kit, with conference sessions, discussion groups in departments, and special editions in journals dedicated to its application within the field. Chapter 2 and Chapter 5 most clearly fall into this category. The former chapter's methodology is based on variational inference stemming from developments in the field of machine learning to create a novel cosmological parameter estimation approach as described above.

In contrast, Chapter 5 makes use of supervised machine learning ensembles to create a hybrid analytic and machine-learned framework predicting baryonic properties of galactic dark matter haloes based only on dark matter information. Our findings demonstrate the viability of our approach and open the path for follow-up research on hybrid methods in galaxy evolution. Just like statistics and other broad fields, though, machine learning is not a monolithic approach, but a conglomeration of methods for different purposes. We hope that our development of various methods for different areas of application highlight the utility that novel inference methods can provide across a variety of cosmological challenges, while demonstrating that 'the right tool for the right problem' still applies.

Bibliography

Abbott T., et al., 2016a, Phys. Rev. D, 94, 022001

Abbott B. P., et al., 2016b, Phys. Rev. Lett., 116, 061102

Abbott T., et al., 2016c, MNRAS, 460, 1270

Abbott T. M. C., et al., 2018a, Phys. Rev. D, 98, 043526

Abbott T. M. C., et al., 2018b, ApJS, 239, 18

Abbott T. M. C., et al., 2018c, MNRAS, 480, 3879

Abbott T. M. C., et al., 2019a, Phys. Rev. Lett., 122, 171301

Abbott T. M. C., et al., 2019b, ApJ, 872, L30

Adermann E., Elahi P. J., Lewis G. F., Power C., 2018, MNRAS, 479, 4861

Agarwal S., Davé R., Bassett B. A., 2018, MNRAS, 478, 3410

Aghanim N., et al., 2018, preprint (`arXiv:1807.06209`)

Ahmadian Y., Pillow J. W., Paninski L., 2011, Neural Comput., 23, 46

Aihara H., et al., 2011, ApJS, 193, 29

Aitken S., Akman O. E., 2013, BMC Syst. Biol., 7, 72

Akeret J., Seehars S., Amara A., Refregier A., Csillaghy A., 2013, Astron. Comput., 2, 27

Akeret J., Refregier A., Amara A., Seehars S., Hasner C., 2015, J. Cosmology Astropart. Phys., 2015, 043

Alcock C., et al., 2000, ApJ, 542, 281

Aliyari Ghassabeh Y., Linder T., Takahara G., 2012, in 25th IEEE Canadian Conference on Electrical and Computer Engineering. pp 1–5, doi:10.1109/CCECE.2012.6334859

Allanach B., Lester C., 2008, Comput. Phys. Commun., 179, 256

Allison R., Dunkley J., 2014, MNRAS, 437, 3918

Alsing J., Wandelt B., Feeney S., 2018, MNRAS, 477, 2874

Alsing J., Charnock T., Feeney S., Wandelt B., 2019, MNRAS, 488, 4440

Altmann A., Toloşi L., Sander O., Lengauer T., 2010, Bioinformatics, 26, 1340

Amanullah R., et al., 2010, ApJ, 716, 712

Amendola L., et al., 2018, Living Rev. Relativ., 21, 2

Anderson L., et al., 2012, MNRAS, 427, 3435

Andrews B. H., Martini P., 2013, ApJ, 765, 140

Anglés-Alcázar D., Davé R., Özel F., Oppenheimer B. D., 2014, ApJ, 782, 84

Antoniak C. E., 1974, Ann. Statist., 2, 1152

Aragón-Calvo M. A., Jones B. J. T., van de Weygaert R., van der Hulst J. M., 2007, A&A, 474, 315

Aragón-Calvo M. A., Platen E., van de Weygaert R., Szalay A. S., 2010, ApJ, 723, 364

Arjona R., Cardona W., Nesseris S., 2019, Phys. Rev. D, 99, 043516

Ball N. M., Brunner R. J., 2010, Int. J. Mod. Phys. D, 19, 1049

Bardenet R., Doucet A., Holmes C., 2014, in Proceedings of the 31st International Conference on Machine Learning (ICML 2014). pp 405–413

Barreira A., Li B., Jennings E., Merten J., King L., Baugh C. M., Pascoli S., 2015a, MNRAS, 454, 4085

Barreira A., Cautun M., Li B., Baugh C. M., Pascoli S., 2015b, J. Cosmology Astropart. Phys., 2015, 028

Barreira A., Bose S., Li B., Llinares C., 2017, J. Cosmology Astropart. Phys., 2017, 031

Bartos I., Kowalski M., 2017, Multimessenger astronomy. 2399-2891, Bristol, UK: IOP Publishing

Bas E., Erdogmus D., 2011, Neuroinformatics, 9, 181

Baugh C. M., 2006, Rep. Prog. Phys., 69, 3101

Baugh C. M., et al., 2018, MNRAS, 483, 4922

Bayes T., 1763, Philos. Trans. R. Soc., 53, 370

Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143

Ben-David A., Liu H., Jackson A. D., 2015, J. Cosmology Astropart. Phys., 2015, 051

Benjamini Y., Hochberg Y., 1995, J. Royal Stat. Soc. B, 57, 289

Bennett C. L., et al., 1996, ApJ, 464, L1

Bennett C. L., et al., 2013, ApJS, 208, 20

Benson A. J., 2010, Phys. Rep., 495, 33

Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587

Bernardo J., Smith A. F. M., 1994, Bayesian theory. Hoboken, USA: John Wiley & Sons, Inc.

Betoule M., et al., 2014, A&A, 568, A22

Bishop C. M., 2006, Pattern recognition and machine learning. Heidelberg, Germany: Springer-Verlag

Blas D., Lesgourgues J., Tram T., 2011, J. Cosmology Astropart. Phys., 2011, 034

Blei D. M., Jordan M. I., 2006, Bayesian Anal., 1, 121

Blei D. M., Kucukelbir A., McAuliffe J. D., 2017, J. Am. Stat. Assoc., 112, 859

Bode P., Ostriker J. P., Turok N., 2001, ApJ, 556, 93

Bonassi F. V., Lingchong Y., West M., 2011, Stat. Appl. Genet. Mol. Biol., 10, 49

Bond J. R., Kofman L., Pogosyan D., 1996, Nature, 380, 603

Bonjean V., Aghanim N., Salomé P., Douspis M., Beelen A., 2018, A&A, 609, A49

Bos E. G. P., van de Weygaert R., Dolag K., Pettorino V., 2012, MNRAS, 426, 440

Boucaud A., et al., 2019, MNRAS, 491, 2481

Bouché N., et al., 2010, ApJ, 718, 1001

Bouchet F., 2016, in Frontiers of Fundamental Physics 14. Berlin, Germany: Springer, doi:10.22323/1.224.0002

Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009a, MNRAS, 398, 1150

Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009b, MNRAS, 398, 1150

Braga A. A., 2005, J. Exp. Criminol., 1, 317

Branch D., 1998, ARA&A, 36, 17

Breiman L., 2001, Mach. Learn., 45, 5

Breiman L., Friedman J. H., Olshen R. A., Stone C. J., 1984, Classification and regression trees. Monterey, USA: Wadsworth & Brooks

Bridle S. L., Crittenden R., Melchiorri A., Hobson M. P., Kneissl R., Lasenby A. N., 2002, MNRAS, 335, 1193

Brooks S., Gelman A., Jones G., Xiao-Li m., 2011, Handbook of Markov chain Monte Carlo. New York, USA: Chapman & Hall (CRC Press)

Brout D., et al., 2019, ApJ, 874, 106

Brouwer M. M., et al., 2018, MNRAS, 481, 5189

Buchner J., 2016, Stat. Comput., 26, 383

Bzdok D., Altman N., Krzywinski M., 2018, Nat. Methods, 15, 233

Cai Y.-C., Padilla N., Li B., 2015, MNRAS, 451, 1036

Caldwell R. R., Kamionkowski M., Weinberg N. N., 2003, Phys. Rev. Lett., 91, 071301

Candès E. J., Donoho D. L., 1999, Philos. Trans. R. Soc. A, 357, 2495

Candès E., Demanet L., Donoho D., Ying L., 2006, Mult. Mod. Sim., 5, 861

Cappé O., Guillin A., Marin J.-M., Robert C. P., 2004, J. Comput. Graph. Stat., 13, 907

Cardoso J.-F., Le Jeune M., Delabrouille J., Betoule M., Patanchon G., 2008, IEEE J. Sel. Top. Signal Process., 2, 735

Carroll S. M., Hoffman M., Trodden M., 2003, Phys. Rev. D, 68, 023509

Cattaneo A., et al., 2017, MNRAS, 471, 1401

Cautun M., Cai Y.-C., Frenk C. S., 2016, MNRAS, 457, 2540

Cautun M., Paillas E., Cai Y.-C., Bose S., Armijo J., Li B., Padilla N., 2018, MNRAS, 476, 3195

Chambers K. C., et al., 2016, preprint (`arXiv:1612.05560`)

Chang C., et al., 2018, MNRAS, 475, 3165

Chávez R., Plionis M., Basilakos S., Terlevich R., Terlevich E., Melnick J., Bresolin F., González-Morán A. L., 2016, MNRAS, 462, 2431

Chen Y.-C., Genovese C. R., Wasserman L., 2014, preprint (arXiv:1406.1803)

Chen Y.-C., Ho S., Freeman P. E., Genovese C. R., Wasserman L., 2015a, MNRAS, 454, 1140

Chen Y.-C., et al., 2015b, MNRAS, 454, 3341

Chen Y.-C., Ho S., Brinkmann J., Freeman P. E. P., Wasserman L., 2016, MNRAS, 461, 3896

Chérief-Abdellatif B.-E., 2019, in Ruiz F., Zhang C., Liang D., Bui T., eds, Proceedings of Machine Learning Research Vol. 96, Proceedings of The 1st Symposium on Advances in Approximate Bayesian Inference. pp 11–31

Chérief-Abdellatif B.-E., Alquier P., 2018, Electron. J. Stat., 12, 2995

Chevallier M., Polarski D., 2001, Int. J. Mod. Phys. D, 10, 213

Chiu I.-N., Umetsu K., Sereno M., Ettori S., Meneghetti M., Merten J., Sayers J., Zitrin A., 2018, ApJ, 860, 126

Chopin N., Robert C., 2010, Biometrika, 97, 741

Christensen N., Meyer R., 1998, Phys. Rev. D, 58, 082001

Christensen N., Meyer R., Knox L., Luey B., 2001, Class. Quantum Grav., 18, 2677

Clampitt J., Cai Y.-C., Li B., 2013, MNRAS, 431, 749

Cole S., Aragón-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, MNRAS, 271, 781

Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168

Cooke R. J., Pettini M., Jorgenson R. A., Murphy M. T., Steidel C. C., 2014, ApJ, 781, 31

Copeland E. J., Sami M., Tsujikawa S., 2006, Int. J. Mod. Phys. D, 15, 1753

Copeland D., Taylor A., Hall A., 2018, MNRAS, 480, 2247

Crittenden R. G., Zhao G.-B., Pogosian L., Samushia L., Zhang X., 2012, J. Cosmology Astropart. Phys., 2, 048

Crocce M., Castander F. J., Gaztañaga E., Fosalba P., Carretero J., 2015, MNRAS, 453, 1513

Croton D. J., et al., 2016, ApJS, 222, 22

Cui W., et al., 2016, MNRAS, 458, 4052

Cuturi M., 2013, in Advances in Neural Information Processing Systems. Red Hook, USA: Curran Associates Inc., p. 2292

Cyburt R. H., Fields B. D., Olive K. A., Yeh T. H., 2016, Rev. Mod. Phys., 88, 015004

Davé R., Finlator K., Oppenheimer B. D., 2012, MNRAS, 421, 98

Davé R., Katz N., Oppenheimer B. D., Kollmeier J. A., Weinberg D. H., 2013, MNRAS, 434, 2645

Davé R., Thompson R., Hopkins P. F., 2016, MNRAS, 462, 3265

Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827

Davies C. T., Cautun M., Li B., 2019, MNRAS, 490, 4907

Davies C. T., Paillas E., Cautun M., Li B., 2021, MNRAS, 500, 2417

De Felice A., Nesseris S., Tsujikawa S., 2012, J. Cosmology Astropart. Phys., 2012, 029

De Jong, Jelte T. A. et al., 2017, A&A, 604, A134

De Souza R. S., et al., 2017, MNRAS, 472, 2808

De Souza R. S., Boston S. R., Coc A., Iliadis C., 2019a, Phys. Rev. C, 99

De Souza R. S., Iliadis C., Coc A., 2019b, ApJ, 872

Dekel A., Rees M. J., 1994, ApJ, 422, L1

Dekel A., et al., 2009, Nature, 457, 451

Del Pozzo W., 2012, Phys. Rev. D, 86, 043011

Del Pozzo W., Li T. G. F., Messenger C., 2017, Phys. Rev. D, 95, 043502

Demchenko V., Cai Y.-C., Heymans C., Peacock J. A., 2016, MNRAS, 463, 512

Dempster A. P., Laird N. M., Rubin D. B., 1977, J. R. Stat. Soc. Series B Stat. Methodol., 39, 1

Desmond H., Ferreira P. G., Lavaux G., Jasche J., 2019, MNRAS, 483, L64

Dicke R. H., Peebles P. J. E., Roll P. G., Wilkinson D. T., 1965, ApJ, 142, 414

Dietrich J. P., Werner N., Clowe D., Finoguenov A., Kitching T., Miller L., Simionescu A., 2012, Nature, 487, 202

Dodelson S., 2003, Modern cosmology, 2 edn. San Diego, USA: Academic Press

Dodelson S., 2020, Modern cosmology, 2 edn. San Diego, USA: Academic Press

Dolag K., Borgani S., Schindler S., Diaferio A., Bykov A. M., 2008, Space Sci. Rev., 134, 229

Driver S. P., et al., 2011, MNRAS, 413, 971

Drlica-Wagner A., et al., 2018, ApJS, 235, 33

Duane S., Kennedy A. D., Pendleton B. J., Roweth D., 1987, Phys. Lett. B, 195, 216

Dubois Y., Peirani S., Pichon C., Devriendt J., Gavazzi R., Welker C., Volonteri M., 2016, MNRAS, 463, 3948

Dumont M., Marée R., Wehenkel L., Geurts P., 2009, in Proceedings of the International Conference on Computer Vision Theory and Applications (VISAPP). pp 196–203

Dutton A. A., van den Bosch F. C., Dekel A., 2010, MNRAS, 405, 1690

Dyson F. W., Eddington A. S., Davidson C., 1920, Philos. Trans. R. Soc. A, 220, 291

Efstathiou G., Davis M., White S. D. M., Frenk C. S., 1985, ApJS, 57, 241

Einstein A., 1915, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), pp 844–847

Einstein A., 1916, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), pp 688–.696

Einstein A., 1917, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), pp 142–152

Einstein A., 1918, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin, pp 154–167

Eisenstein D. J., 2005, New Astron. Rev., 49, 360

El-Ad H., Piran T., 1997, ApJ, 491, 421

Elvin-Poole J., et al., 2018, Phys. Rev. D, 98, 042006

Falck B., Koyama K., Zhao G.-B., Cautun M., 2018, MNRAS, 475, 3262

Fan Y., Nott D. J., Sisson S. A., 2013, Stat, 2, 34

Feigelson E. D., de Souza R. S., Ishida E. E., Babu G. J., 2021, Annu. Rev. Stat. Appl., 8, 493

Feng Y., Di-Matteo T., Croft R. A., Bird S., Battaglia N., Wilkins S., 2016, MNRAS, 455, 2778

Ferguson T. S., 1973, Ann. Statist., 1, 209

Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601

Filippenko A. V., 1997, ARA&A, 35, 309

Finlator K., Davé R., 2008, MNRAS, 385, 2181

Fisher A., Rudin C., Dominici F., 2019, J. Mach. Learn. Res., 20, 1

Flaugher B., et al., 2015, AJ, 150, 150

Fluke C. J., Jacobs C., 2020, Data Min. Knowl. Disc., 10, e1349

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306

Fosalba P., Gaztañaga E., Castander F. J., Crocce M., 2015a, MNRAS, 447, 1319

Fosalba P., Crocce M., Gaztañaga E., Castander F. J., 2015b, MNRAS, 448, 2987

Fox C. W., Roberts S. J., 2012, Artif. Intell. Rev., 38, 85

Friedmann A., 1922, Z. Phys., 10, 377

Frieman J. A., Turner M. S., Huterer D., 2008, ARA&A, 46, 385

Fry J. N., 1986, ApJ, 306, 358

Fussell L., Moews B., 2019, MNRAS, 485, 3203

Gaite J., 2005, Eur. Phys. J. B, 47, 93

Gal Y., Ghahramani Z., 2016, in Balcan M. F., Weinberger K. Q., eds, Proceedings of Machine Learning Research Vol. 48, Proceedings of The 33rd International Conference on Machine Learning. pp 1050–1059

Gal Y., Hron J., Kendall A., 2017, in Guyon I., Luxburg U. V., Bengio S., Wallach H., Fergus R., Vishwanathan S., Garnett R., eds, Proceedings of Machine Learning Research Vol. 30, Proceedings of the 31st International Conference on Neural Information Processing Systems. pp 3584–3593

Galárraga-Espinosa D., Aghanim N., Langer M., Gouin C., Malavasi N., 2020, A&A, 641, A173

Gallagher P. T., Young C. A., Byrne J. P., McAteer R. T. J., 2011, Adv. Space Res., 47, 2118

Gamow G., 1948, Nature, 162, 680

Gannouji R., Polarski D., Ranquet A., Starobinsky A. A., 2006, J. Cosmology Astropart. Phys., 9, 016

Garnavich P. M., et al., 1998, ApJ, 509, 74

Gelman A., Carlin J. B., Stern H. S., Rubin D. B., 2013, Bayesian data analysis. New York, USA: Chapman & Hall (CRC Press)

Geman S., Geman D., 1984, IEEE Trans. Pattern Anal. Mach. Intell., 6, 721

Genel S., et al., 2014, MNRAS, 445, 175

Genovese C., Freeman P., Wasserman L., Nichol R., Miller C., 2009, Ann. Appl. Stat., 3, 144

Genovese C. R., Perone-Pacifico M., Verdinelli I., Wasserman L., 2014, Ann. Stat., 42, 1511

Gershman S. J., Blei D. M., 2012, J. Math. Psychol., 56, 1

Geurts P., 2006, Mach. Learn., pp 3–42

Ghassabeh Y. A., Rudzicz F., 2021, J. Classif., 38, 27

Giblin B., Cataneo M., Moews B., Heymans C., 2019, MNRAS, 490, 4826

Goodman J., Weare J., 2010, Commun. Appl. Math. Comput. Sci., 5, 65

Govoni F., et al., 2019, Science, 364, 981

Graff P., Feroz F., Hobson M. P., Lasenby A., 2012, MNRAS, 421, 169

Griffiths M., Wales D. J., 2019, J. Chem. Theory Comput., 15, 6865

Gruen D., et al., 2016, MNRAS, 455, 3367

Gruen D., et al., 2018, Phys. Rev. D, 98, 023507

Guidotti R., Monreale A., Ruggieri S., Turini F., Giannotti F., Pedreschi D., 2018, ACM Comput. Surv., 51, 93

Hajian A., 2007, Phys. Rev. D, 75, 083525

Hamaus N., Sutter P. M., Wandelt B. D., 2014, Proc. Int. Astron. Union, 11, 538

Hamaus N., Pisani A., Sutter P. M., Lavaux G., Escoffier S., Wandelt B. D., Weller J., 2016, Phys. Rev. Lett., 117, 091302

Handley W. J., Hobson M. P., Lasenby A. N., 2015, MNRAS, 453, 4384

Hannestad S., Mortsell E., 2004, J. Cosmology Astropart. Phys., 2004, 001

Harrison I., Camera S., Zuntz J., Brown M. L., 2016, MNRAS, 463, 3674

Hastie T., Tibshirani R., Friedman J., 2001, The elements of statistical learning. New York, USA: Springer

Hastings W. K., 1970, Biometrika, 57, 97

Hatton S., Devriendt J. E. G., Ninin S., Bouchet F. R., Guiderdoni B., Vibert D., 2003, MNRAS, 343, 75

He S., Alam S., Ferraro S., Chen Y.-C., Ho S., 2017, Nat. Astron., 2, 401

He Y., et al., 2018, Concur. and Computat.: Practice and Experience, 30, e4291

Hearin A., Korytov D., Kovacs E., Benson A., Aung H., Bradshaw C., Campbell D., LSST Dark Energy Science Collaboration 2020, MNRAS, 495, 5040

Hee S., Vázquez J. A., Handley W. J., Hobson M. P., Lasenby A. N., 2017, MNRAS, 466, 369

Heitmann K., Lawrence E., Kwan J., Habib S., Higdon D., 2014, ApJ, 780, 111

Hendel D., Johnston K. V., Patra R. K., Sen B., 2019, MNRAS, 486, 3604

Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, MNRAS, 451, 2663

Hergt L., Amara A., Brandenberger R., Kacprzak T., Refregier A., 2017, J. Cosmology Astropart. Phys., 2017, 004

Herman G. T., 2009, Fundamentals of computerized tomography: Image reconstruction from projections. Berlin, Germany: Springer

Hertog T., Horowitz G. T., 2004, J. High Energy Phys., 2004, 073

Higson E., Handley W., Hobson M., Lasenby A., 2018, Bayesian Anal., 13, 873

Higson E., Handley W., Hobson M., Lasenby A., 2019, Stat. Comput., 29, 891

Higuchi Y., Shirasaki M., 2016, MNRAS, 459, 2762

Hinshaw G., et al., 2013, ApJS, 208, 19

Hirschmann M., De Lucia G., Fontanot F., 2016, MNRAS, 461, 1760

Hjort N. L., Holmes C., Mueller P., Walker S. G., 2010, Bayesian nonparametrics: Principles and practice. Cambridge, UK: Cambridge University Press

Ho T. K., 1995, in Proceedings of the 3rd International Conference on Document Analysis and Recognition. pp 278–282, doi:10.1109/ICDAR.1995.598994

Ho T. K., 2002, Pattern Anal. Appl., 5, 102

Hobson M. P., Feroz F., 2008, MNRAS, 384, 449

Hobson M. P., Jaffe A. H., Liddle A. R., Mukherjee P., Parkinson D., 2009, Bayesian methods in cosmology. Cambridge, UK: Cambridge University Press

Hoeft M., Brüggen M., Yepes G., Gottlöber S., Schwope A., 2008, MNRAS, 391, 1511

Hoffmann M. D., Gelman A., 2014, J. Mach. Learn. Res., 15, 1593

Holsclaw T., Alam U., Sansó B., Lee H., Heitmann K., Habib S., Higdon D., 2010a, Phys. Rev. D, 82, 103502

Holsclaw T., Alam U., Sansó B., Lee H., Heitmann K., Habib S., Higdon D., 2010b, Phys. Rev. Lett., 105, 241302

Hopkins P. F., 2014, GIZMO: Multi-method magneto-hydrodynamics + gravity code (ascl:1410.003)

Howlett C., Lewis A., Hall A., Challinor A., 2012, J. Cosmology Astropart. Phys., 2012, 027

Hoyle F., Vogeley M. S., 2004, ApJ, 607, 751

Hoyle B., et al., 2018, MNRAS, 478, 592

Hubble E., 1929, Contrib. Mt. Wilson Obs., 3, 23

Huff E., Mandelbaum R., 2017, preprint (arXiv:1702.02600)

Huterer D., Shafer D. L., 2018, Rep. Prog. Phys., 81, 016901

Inman J. W., 1835, Navigation and nautical astronomy for the use of British seamen, 3 edn. London, UK: W. Woodward, C. & J. Rivington

Ionides E. L., 2008, J. Comput. Graph. Stat., 17, 295

Ishida E. E. O., et al., 2015, Astron. Comput., 13, 1

Ivezic Z., et al., 2008, preprint (arXiv:0805.2366)

Ivezić Ž., et al., 2019, ApJ, 873, 111

Jaffe A., 1996, ApJ, 471, 24

Jang W., 2006, Comput. Stat. Data Anal., 50, 760

Jassal H. K., Bagla J. S., Padmanabhan T., 2005, MNRAS, 356, L11

Jauzac M., et al., 2012, MNRAS, 426, 3369

Jeffrey N., et al., 2021, MNRAS, p. in print

Jiang M., Cui B., Yu Y.-F., Cao Z., 2019, IEEE Access, 7, 107389

Jo Y., Kim J.-h., 2019, MNRAS, 489, 3565

Jones D. O., et al., 2018, ApJ, 857, 51

Jones D. O., et al., 2019, ApJ, 881, 19

Jordan M. I., Ghahramani Z., Jaakkola T. S., Saul L. K., 1999, Mach. Learn., 37, 183

Kacprzak T., et al., 2016, MNRAS, 463, 3653

Kahn H., Marshall A. W., 1953, Oper. Res., 1, 263

Kaiser N., 1998, ApJ, 498, 26

Kaiser N., Squires G., 1993, ApJ, 404, 441

Kaiser N., Squires G., Broadhurst T., 1995, ApJ, 449, 460

Kamdar H. M., Turk M. J., Brunner R. J., 2016a, MNRAS, 455, 642

Kamdar H. M., Turk M. J., Brunner R. J., 2016b, MNRAS, 457, 1162

Kang X., Jing Y. P., Mo H. J., Borner G., 2005, ApJ, 631, 21

Kashlinsky A., Atrio-Barandela F., Kocevski D., Ebeling H., 2008, ApJ, 686, L49

Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201

Keeton C. R., 2011, MNRAS, 414, 1418

Khandai N., Di Matteo T., Croft R., Wilkins S., Feng Y., Tucker E., DeGraf C., Liu M.-S., 2015, MNRAS, 450, 1349

Kilbinger M., et al., 2009, A&A, 497, 677

Kilbinger M., et al., 2010, MNRAS, 405, 2381

Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102

Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340

Knebe A., et al., 2015, MNRAS, 451, 4029

Knox L., Christensen N., Skordis C., 2001, ApJ, 563, L95

Koennig F., Akrami Y., Amendola L., Motta M., Solomon A. R., 2014, Phys. Rev. D, 790, 124014

Komatsu E., et al., 2011, ApJS, 192, 18

Korytov D., et al., 2019, ApJS, 245, 26

Krause E., Eifler T., 2017, MNRAS, 470, 2100

Krause E., et al., 2017, preprint (`arXiv:1706.09359`)

Krumholz M. R., Dekel A., 2012, ApJ, 753, 16

Kullback S., Leibler R. A., 1951, Ann. Math. Statist., 22, 79

L'Huillier B., Shafieloo A., Linder E. V., Kim A. G., 2019, MNRAS, 485, 2783

Laliberte S., Brandenberger R., Camargo Neves da Cunha D., 2018, preprint (`arXiv:1807.09820`)

Laplace P., 1812, Théorie analytique des probabilités. Paris, France: Courcier

Lavaux G., Wandelt B. D., 2012, ApJ, 754, 109

Lawrence E., Heitmann K., White M., Higdon D., Wagner C., Habib S., Williams B., 2010, ApJ, 713, 1322

Leavitt H. S., 1908, "AnHar", 60, 87

Leitner S. N., Kravtsov A. V., 2011, ApJ, 734, 48

Lemaître G. H. J. E., 1927, PhD thesis, MIT

Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 103511

Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473

Li B., 2011, MNRAS, 411, 2615

Li Q., Racine J. S., 2006, Nonparametric econometrics: Theory and practice. Princeton, USA: Princeton University Press

Libeskind N. I., et al., 2018, MNRAS, 473, 1195

Liddle A., 2003, An introduction to modern cosmology, 2 edn. Hoboken, USA: John Wiley & Sons, Inc.

Liddle A. R., Lyth D. H., 2000, Cosmological inflation and large-scale structure. Cambridge, UK: Cambridge University Press

Liddle A., Parkinson D., Mukherjee P., 2006, Astron. Geophys., 47, 4.30

Lilly S. J., Carollo C. M., Pipino A., Renzini A., Peng Y., 2013, ApJ, 772, 119

Lin W., Ishak M., 2017, Phys. Rev. D, 96, 083532

Linder E. V., 2003, Phys. Rev. Lett., 90, 091301

Lucie-Smith L., Peiris H. V., Pontzen A., Lochner M., 2018, MNRAS, 479, 3405

Luzzi G., Génova-Santos R. T., Martins C. J. A. P., De Petris M., Lamagna L., 2015, J. Cosmology Astropart. Phys., 9, 011

Ly C., Malkan M. A., Rigby J. R., Nagao T., 2016, ApJ, 828, 67

Ma C.-P., Bertschinger E., 1995, ApJ, 455, 7

MacKay D. J. C., 2003, Information theory, inference and learning algorithms. Cambridge, UK: Cambridge University Press

Malavasi N., Aghanim N., Douspis M., Tanimura H., Bonjean V., 2020, A&A, 642, A19

Mallat S., 1999, A wavelet tour of signal processing: The sparse way. Amsterdam, Netherlands: Elsevier

Malz A. I., Marshall P. J., DeRose J. Graham M. L., Schmidt S. J., Wechsler R., (LSST Dark Energy Science Collaboration) 2018, AJ, 156, 35

Mateus A., Sodré L., Cid Fernandes R., Stasińska G., 2007, MNRAS, 374, 1457

Maturi M., Merten J., 2013, A&A, 559, A112

McAuliffe J. D., Blei D. M., Jordan M., 2006, Stat. Comput., 16, 5–

Mead J. M., King L. J., McCarthy I. G., 2010, MNRAS, 401, 2257

Melchior P., et al., 2017, MNRAS, 469, 4899

Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E., 1953, J. Chem. Phys., 21, 1087

Miao Z., Wang B., Shi W., Wu H., 2014, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 7, 4762

Milgrom M., 1983, ApJ, 270, 365

Miranda V., Dvorkin C., 2018, Phys. Rev. D, D98, 043537

Misner C. W., Thorne K. S., Wheeler J. A., 1973, Gravitation. Princeton, USA: Princeton University Press

Mitchell P. D., et al., 2018, MNRAS, 474, 492

Mitra S., Davé R., Finlator K., 2015, MNRAS, 452, 1184

Mitra S., Davé R., Simha V., Finlator K., 2017, MNRAS, 464, 2766

Moews B., Ibikunle G., 2020, Physica A Stat. Mech. Appl., 547, 124392

Moews B., Zuntz J., 2020, ApJ, 896, 98

Moews B., de Souza R. S., Ishida E. E. O., Malz A. I., Heneka C., Vilalta R., Zuntz J., COIN Collaboration 2019a, Phys. Rev. D, 99, 123529

Moews B., Herrmann J. M., Ibikunle G., 2019b, Expert Syst. Appl., 120, 197

Moews B., et al., 2020, MNRAS, 500, 859

Moews B., Davé R., Mitra S., Hassan S., Cui W., 2021a, MNRAS (forthcoming)

Moews B., Argueta J. R., Gieschen A., 2021b, Decis. Support Syst., 144, 113518

Monaco P., 2004, MNRAS, 352, 181

Moss A., 2020, MNRAS, 496, 328

Moster B. P., Naab T., Lindström M., O'Leary J. A., 2020, preprint (`arXiv:2005.12276`)

Mukherjee P., Parkinson D., Liddle A. R., 2006, ApJ, 638, L51

Munshi D., Valageas P., van Waerbeke L., Heavens A., 2008, Phys. Rep., 462, 67

Murphy K. P., 2012, Machine learning: A probabilistic perspective. Cambridge, USA: The MIT Press

Nadathur S., Hotchkiss S., Crittenden R., 2017, MNRAS, 467, 4067

Narlikar J. V., Padmanabhan T., 2001, ARA&A, 39, 211

Neistein E., Khochfar S., Dalla Vecchia C., Schaye J., 2012, MNRAS, 421, 3579

Neiswanger W., Wang C., Xing E. P., 2014, in Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence (UAI'14). pp 623–632

Nelson D., et al., 2015, Astron. Comput., 13, 12

Neyrinck M. C., 2008, MNRAS, 386, 2101

Nicola A., Amara A., Refregier A., 2019, Journal of Cosmology and Astroparticle Physics, 2019, 011

Noterdaeme P., Petitjean P., Srianand R., Ledoux C., López S., 2011, A&A, 526, L7

O'Brien T. A., Kashinath K., Cavanaugh N. R., Collins W. D., O'Brien J. P., 2016, Comput. Stat. Data Anal., 101, 148

O'Raifeartaigh C., O'Keeffe M., Nahm W., Mitton S., 2018, Eur. Phys. J. H, 43, 73

Oort J. H., 1940, ApJ, 91, 273

Osband I., 2016, in Gal Y., Louizos C., Gharamani Z., Murphy K., Welling M., eds, NeurIPS Workshops Vol. 192, Neural Information Processing Systems Workshop on Bayesian Deep Learning 2016.

Ostriker E. C., McKee C. F., Leroy A. K., 2010, ApJ, 721, 975

Ozertem U., Erdogmus D., 2011, J. Mach. Learn. Res., 12, 1249

Padilla N. D., Ceccarelli L., Lambas D. G., 2005, MNRAS, 363, 977

Papamakarios G., Murray I., 2016, in Lee D. D., Sugiyama M., Luxburg U. V., Guyon I., Garnett R., eds, , Advances in Neural Information Processing Systems 29. Red Hook, USA: Curran Associates, pp 1028–1036

Pardo K., Spergel D. N., 2020, Phys. Rev. Lett., 125, 211101

Park C., Choi Y.-Y., Vogeley M. S., Gott J. Richard I., Blanton M. R., SDSS Collaboration 2007, ApJ, 658, 898

Peacock J. A., 1999, Cosmological physics. Cambridge, UK: Cambridge University Press

Peccei R. D., Quinn H. R., 1977, Phys. Rev. Lett., 38, 1440

Peebles P. J. E., 2001, ApJ, 557, 495

Peebles P. J. E., Ratra B., 1988, ApJ, 325, L17

Peebles P. J., Ratra B., 2003, Rev. Mod. Phys., 75, 559

Penzias A. A., Wilson R. W., 1965, ApJ, 142, 419

Perlmutter S., et al., 1997, ApJ, 483, 565

Perlmutter S., et al., 1999, ApJ, 517, 565

Peterson C., Anderson J. R., 1987, Complex Syst., 1, 995

Peterson C., Hartman E., 1989, Neural Netw., 2, 475

Peyré G., Cuturi M., 2019, Found. Trends Mach. Learn., 11, 355

Phillips M. M., 1993, ApJ, 413, L105

Pillepich A., et al., 2018, MNRAS, 475, 648

Pisani A., Sutter P. M., Hamaus N., Alizadeh E., Biswas R., Wandelt B. D., Hirata C. M., 2015, Phys. Rev. D, 92, 083531

Planck Collaboration et al., 2020a, A&A, 641, A1

Planck Collaboration et al., 2020b, A&A, 641, A7

Poincaré M. H., 1906, Annu. Rev. Stat. Appl., 21, 129

Potter D., Stadel J., Teyssier R., 2017, CompAC, 4, 2

Press W. H., Schechter P., 1974, ApJ, 187, 425

Price-Whelan A. M., Foreman-Mackey D., 2017, J. Open Source Softw., 2, 357

Primack J. R., Gross M. A. K., 2001, in Caldwell D. O., ed., , Current aspects of neutrino physics. Berlin, Germany: Springer, pp 287–308

Pycke J. R., Russell E., 2016, ApJ, 821, 110

Quinlan J. R., 1986, Mach. Learn., 1, 81

Racca G. D., et al., 2016, in Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave. p. 99040O, doi:10.1117/12.2230762

Rahmati A., Pawlik A. H., Raičevič M., Schaye J., 2013, MNRAS, 430, 2427

Rasmussen C. E., Williams C. K. I., 2005, Gaussian processes for machine learning (adaptive computation and machine learning). Cambridge, USA: The MIT Press

Ravanbakhsh S., Oliva J., Fromenteau S., Price L. C., Ho S., Schneider J., Poczos B., 2016, in Proceedings of the 33rd International Conference on Machine Learning. pp 2407–2416

Ravanbakhsh S., Lanusse F., Mandelbaum R., Schneider J., Poczos B., 2017, in Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI-17). pp 1488–1494

Regier J., Miller A., McAuliffe J., Adams R., Hoffman M., Lang D., Schlegel D., Prabhat 2015, in Bach F., Blei D., eds, Proceedings of Machine Learning Research Vol. 37, Proceedings of the 32nd International Conference on Machine Learning. pp 2095–2103

Riebe K., et al., 2013, Astron. Nachr., 334, 691

Riess A. G., Press W. H., Kirshner R. P., 1996, ApJ, 473, 88

Riess A. G., et al., 1998, AJ, 116, 1009

Riess A. G., et al., 2007, ApJ, 659, 98

Riess A. G., et al., 2016, ApJ, 826, 56

Robert C. P., Casella G., 2004, Monte Carlo statistical methods. Heidelberg, Germany: Springer-Verlag

Robert C., Casella G., 2011, Stat. Sci., 26, 102

Robert C. P., Elvira V., Tawn N., Wu C., 2018, WIREs Comput. Stat., 10, e1435

Roberts G. O., Rosenthal J. S., 2009, J. Comput. Graph. Stat., 18, 349

Robertson H. P., 1935, ApJ, 82, 284

Rodríguez A. C., Kacprzak T., Lucchi A., Amara A., Sgier R., Fluri J., Hofmann T., Réfrégier A., 2018, "CompAC", 5, 4

Rubin V. C., Ford W. K. J., Thonnard N., 1980, ApJ, 238, 471

Rykoff E. S., et al., 2014, ApJ, 785, 104

Saha P., Williams T. B., 1994, AJ, 107, 1295

Saintonge A., et al., 2013, ApJ, 778, 2

Sánchez C., et al., 2017, MNRAS, 465, 746

Sanders R. L., et al., 2018, ApJ, 858, 99

Schaye J., et al., 2015, MNRAS, 446, 521

Schneider P., 1996, MNRAS, 283, 837

Scolnic D. M., et al., 2018, ApJ, 859, 101

Scoville N., et al., 2007a, ApJS, 172, 1

Scoville N., et al., 2007b, ApJS, 172, 150

Segal M., Xiao Y., 2011, WIREs Data Min. Knowl. Discovery, 1, 80

Serra P., Cooray A., Holz D. E., Melchiorri A., Pandolfi S., Sarkar D., 2009, Phys. Rev. D, 80, 121302

Sethuraman J., 1994, Stat. Sin., 4, 639

Sheldon E. S., Huff E. M., 2017, ApJ, 841, 24

Silverman B. W., 1986, Density estimation for statistics and data analysis. Vol. 26, Cleveland, USA: CRC press

Simet M., McClintock T., Mandelbaum R., Rozo E., Rykoff E., Sheldon E., Wechsler R. H., 2017, MNRAS, 466, 3103

Skilling J., 2006, Bayesian Anal., 1, 833

Skilling J., 2009, in Goggans P. M., Chan C.-Y., eds, American Institute of Physics Conference Series Vol. 1193, 29th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering. pp 277–291

Slipher V. M., 1915, Pop. Astron., 23, 21

Smith R. E., et al., 2003, MNRAS, 341, 1311

Solà Peracaula J., Gómez-Valent A., de Cruz Pérez J., 2019, Phys. Dark Universe, 25

Somerville R. S., Davé R., 2015, ARA&A, 53, 51

Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087

Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, MNRAS, 391, 481

Speagle J. S., Steinhardt C. L., Capak P. L., Silverman J. D., 2014, ApJS, 214, 15

Springel V., 2005, MNRAS, 364, 1105

Springel V., et al., 2005, Nature, 435, 629

Starck J.-L., Donoho D. L., Candès E. J., 2003, A&A, 398, 785

Starck J.-L., Aghanim N., Forni O., 2004, A&A, 416, 9

Starck J.-L., Murtagh F., Fadili J., 2015, Sparse image and signal processing: Wavelets and related geometric multiscale analysis. Cambridge, UK: Cambridge University Press

Strobl C., Boulesteix A.-L., Zeileis A., Hothorn T., 2007, BMC Bioinform., 8, 25

Stumpf M. P. H., Kirk P., Johnson R., 2014, Bioinformatics, 31, 604

Sullivan M., et al., 2011, ApJ, 737, 102

Suzuki N., et al., 2012, ApJ, 746, 85

Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, ApJ, 761, 152

Taylor A. N., Kitching T. D., 2010, MNRAS, 408, 865

Thorndike R. L., 1953, Psychometrika, 18, 267

Torrie G. M., Valleau J. P., 1977, J. Comput. Phys., 23, 187

Tripathi A., Sangwan A., Jassal H. K., 2017, J. Cosmology Astropart. Phys., 2017, 012

Trotta R., 2008, Contemp. Phys., 49, 71

Trotta R., Jóhannesson G., Moskalenko I. V., Porter T. A., Ruiz de Austri R., Strong A. W., 2011, ApJ, 729, 106

Troxel M. A., et al., 2018, Phys. Rev. D, 98, 043528

Ur Rahman S. F., 2018, Astron. Geophys., 59, 2.39

Vafaei Sadr A., Movahed S., Farhang M., Ringeval C., Bouchet F., 2017, MNRAS, 475, 1010

Vázquez J. A., Bridges M., Hobson M. P., Lasenby A. N., 2012, J. Cosmology Astropart. Phys., 2012, 020

Verde L., Feeney S. M., Mortlock D. J., Peiris H. V., 2013, J. Cosmology Astropart. Phys., 2013, 013

Villani C., 2008, Optimal transport: Old and new. Vol. 338, Berlin, Germany: Springer

Vogelsberger M., et al., 2014, MNRAS, 444, 1518

Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020, Nat. Rev. Phys, 2, 42

Walker A. G., 1937, Proc. London Math. Soc., 42, 90

Wang S., Wang Y., Li M., 2017, Phys. Rep., 696, 1

Wang G., Ye J. C., Mueller K., Fessler J. A., 2018a, IEEE Trans. Med. Imaging, 37, 1289

Wang S., Wang Y.-F., Xia D.-M., 2018b, Chin. Phys. C, 42, 065103

Webb T., Dulberg Z., Frankland S., Petrov A., O'Reilly R., Cohen J., 2020, in III H. D., Singh A., eds, Proceedings of Machine Learning Research Vol. 119, Proceedings of the 37th International Conference on Machine Learning. pp 10136–10146

Weiler K. W., Sramek R. A., 1988, ARA&A, 26, 295

White S. D. M., Frenk C. S., 1991, ApJ, 379, 52

White S. D. M., Frenk C. S., Davis M., Efstathiou G., 1987, ApJ, 313, 505

Wilkinson D. J., 2005, in , Handbook of parallel computing and statistics. New York: Chapman & Hall (CRC Press), pp 477–513

Winther H. A., Casas S., Baldi M., Koyama K., Li B., Lombriser L., Zhao G.-B., 2019, Phys. Rev. D, 100, 123540

Wolpert D. H., Macready W. G., 1997, IEEE Trans. Evol. Comput., 1, 67

Wood-Vasey W. M., et al., 2007, ApJ, 666, 694

Wraith D., Kilbinger M., Benabed K., Cappé O., Cardoso J.-F., Fort G., Prunet S., Robert C. P., 2009, Phys. Rev. D, 80, 023507

Wu X., et al., 2008, KAIS, 14, 1

Xavier H. S., Abdalla F. B., Joachimi B., 2016a, FLASK: Full-sky Lognormal Astro-fields Simulation Kit (ascl:1606.015)

Xavier H. S., Abdalla F. B., Joachimi B., 2016b, MNRAS, 459, 3693

Xia Q., et al., 2020, A&A, 633, A89

Xu G., 1995, ApJS, 98, 355

Xu X., Cisewski-Kehe J., Green S. B., Nagai D., 2019, Astron. Comput., 27, 34

Yang L. F., Neyrinck M. C., Aragón-Calvo M. A., Falck B., Silk J., 2015, MNRAS, 451, 3606

Ydri B., 2017, in 2053–2563, Lectures on General Relativity, Cosmology and Quantum Black Holes. IOP Publishing, pp 3–1 to 3–45, doi:10.1088/978-0-7503-1478-7ch3

York D. G., et al., 2000, AJ, 120, 1579

Zahid H. J., Dima G. I., Kudritzki R.-P., Kewley L. J., Geller M. J., Hwang H. S., Silverman J. D., Kashino D., 2014, ApJ, 791, 130

Zeldovich I. B., Einasto J., Shandarin S. F., 1982, Nature, 300, 407

Zhao G.-B., Huterer D., Zhang X., 2008, Phys. Rev. D, 77, 121302

Zhao G.-B., et al., 2017, Nat. Astron., 1, 627

Zhou Z., Hooker G., 2020, preprint (arXiv:1903.05179)

Zivick P., Sutter P. M., Wandelt B. D., Li B., Lam T. Y., 2015, MNRAS, 451, 4215

Zuntz J., et al., 2015, Astron. Comput., 12, 45

Zuntz J., et al., 2018, MNRAS, 481, 1149

Zwicky F., 1937, ApJ, 86, 217