ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Weak lensing map inference: a physics informed Gaussian process approach

Master Thesis

Nicolò Massari

April 02, 2024

Advisors: Prof. Dr. Alexandre Réfrégier, Dr. Tilman Tröster. Veronika Oehl

Department of Physics, ETH Zürich

Abstract

In this work we propose the use of physically informed Gaussian processes (GP) to analyse cosmological fields at the map level. We will show that GPs can capture the statistical behaviour of cosmological fields, providing us with a likelihood as a function of the cosmological parameters conditioned to the map. In practice, we set the Gaussian process kernel to be the 2-point autocorrelation function associated to a 2D discrete flat-sky convergence map. We find that a GP in this setup is not only able to generate maps with the wanted 2-point statistics, but also to reconstruct masked data with an associated uncertainty. Additionally, we perform a Bayesian inference analysis in order to test the ability of Gaussian processes to recover cosmological parameters. We find that we are able to consistently recover the σ_8 and Ω_m degeneracy, recovering S_8 within two sigma uncertainty. The data is simulated by a Gaussian random field realisation of a convergence map of size $(10°, 10°)$, 64×64 grid, mask $\sim 10\%$ and noise given by a galaxy density of $n_g = 10$ galaxies/arcmin^2.

keywords: Gaussian processes, weak lensing.

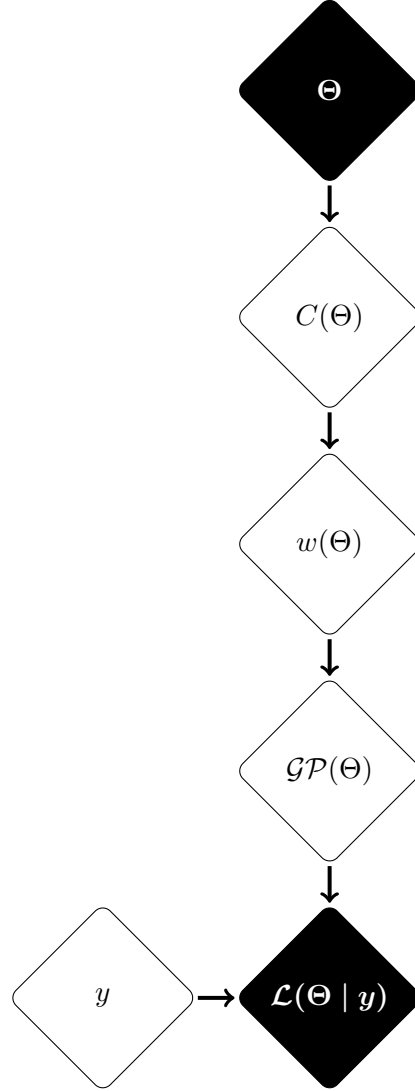# Contents

Chapter 1

Introduction

Matter bends light. The theory of general relativity predicts that the presence of matter or energy changes the geometry of space and time, which in turn can cause what would otherwise be the straight path of a beam of light to curve. Take a distant source of light. If we assume the universe not to be empty, then between us and said source there exists a non trivial matter disposition. As light travels through everything that is in between us and the source, it gets blocked, bent and distorted. We call this phenomenon weak lensing. In practice what we measure is the shape of distant galaxies. These images do not show a distinct lensing feature individually, as such tiny changes can only be seen with a large number of sources. For example, we observe that galaxies have a tendency of aligning along a preferred axis, causing a statistical discrepancy in an otherwise seemingly isotropic universe. For further details on lensing, please refer to [1] [2]; for weak lensing [3]. The image of a distant galaxy can change shape or size. Changes in shape are fully characterised by the shear distortion $\vec{\gamma}$ vector field, where the change in size is given by its magnitude, the convergence field κ. We lay out a theoretical framework for weak lensing in *Sec.* 2.1 as well as a details on the cosmology used in this thesis in *Sec.* 3.1.1.

A lot of work goes into translating a measurement of distant galaxies to its mathematically friendly counterpart $\vec{\gamma}$. This is one of the reasons why we will not be dealing with it in this thesis, as it is outside of our scope. Instead we will be simulating our own convergence fields. We use a *Gaussian random field (GRF)* algorithm for the creation of Gaussianly distributed data. As well as lognormal transformations to create fields with a distribution that resembles more closely that in our universe. These transformations are listed in *Sec.* 2.2; we then use them to simulate our data as explained in *Sec.* 3.1. We also verify that the generated fields recover the fiducial power spectrum in *Sec.* 4.1.1.

A *Gaussian process (GP)* usually assumes little prior knowledge about the data it is applied to. Current research in the field of cosmology views GPs as a machine learning tool to be trained. It is used to accelerate and optimise models [4] [5] [6], as well as for its interpolation qualities applied to the reconstruction of functions determining the evolution of the universe [7] [8] [9]. Our work, however, is based on a different approach. We apply our prior knowledge about 2-point statistics in cosmology to create a fully informed GP. Restricting ourselves to 2D flat-sky weak lensing convergence fields, as shown in *Fig.* 1.1, we can:

- compute the angular power spectrum $C(\Theta)$ from a set of cosmological parameters Θ,

- transform it in the convergence angular autocorrelation function $w(\Theta)$,

- create a zero mean GP with kernel given by said correlation function,

- evaluate the likelihood $\mathcal{L}$ of Θ given a set of data points y.

With a Bayesian approach we make use of this pipeline to infer the values of the cosmological parameters. Running a *Markov chain Monte Carlo (MCMC)* we can sample the posterior distribution of the cosmological parameters, in particular we will get contours for Ω_m, σ_8 and S_8. Other than that, GPs have several other interesting properties at the field level. They are not only able to generate fields that recover the fiducial 2-point statistics, but are also able to reconstruct

Figure 1.1: Simplified steps taken by the model to go from cosmological parameters Θ to likelihood $\mathcal{L}$ using GP. Here C and w stand for the power spectrum and correlation function respectively, y is the data.

masked fields, a task that usually brings many challenges to C_ℓ estimation [10] [11]. In the field of weak lensing in particular, foreground objects like bright stars or galaxies can contaminate measurements, leading to the need of masking such a region, essentially removing the signal.

Here we list the advantages of our method:

- minimal information loss, as it is a map based method we use all available data points,

- it can be used as a likelihood for cosmological parameters inference,

- easily deals with masked fields, providing estimates with an associated uncertainty for the masked points.

Whilst some of the disadvantages:

- the conditioning process depends on the inverse of a correlation matrix, with a computational effort that grows as $\sim \mathcal{O}(n^3)$ for $n \propto$ data points, indicating possible scaling issues;

- due to the intrinsic Gaussianity of GPs, the distribution of the samples will be Gaussian, making it hard to apply them to other fields, say for example lognormal fields.

GPs in this thesis are presented in a general introduction in *Sec.* 2.3, followed by a detailed account of the computational methods used to recover a working kernel for GPs in *Sec.* 3.2. In our results we show their ability to create maps that follow the desired statistic *Sec.* 4.1.2 and reconstruct data *Sec.* 4.2. We also present our attempt at cosmological parameters inference with GPs in *Sec.* 4.3.

It is important to note that throughout the thesis we follow the extremely useful guidelines set by the `Miko` pipeline [12] on how to deal with discrete maps in weak lensing.

Theoretical Framework

2.1 Weak lensing

2.1.1 Convergence field

Let's introduce what can only be defined as the protagonist cosmological field of this thesis: the convergence field [13] [3] [1] [2]. As explained in the introduction weak lensing happens due to the presence of matter between us and a distant source. In mathematical terms we write this as an integral over the line of sight of the matter overdensity field, weighted accordingly.

$$\kappa(\theta) = \int_0^{\chi*} d\chi W(\chi) \delta_m(\chi\theta, \chi), \tag{2.1}$$

$$W(\chi) = \tfrac{3}{2} H_0^2 \Omega_m (1 + z(\chi))\chi \int_\chi^{\chi*} d\chi' n(z(\chi')) \left(1 - \frac{\chi}{\chi'}\right). \tag{2.2}$$

With $W(\chi)$ being a measure of the lensing weights; $W(\chi)$ incorporates all of the relevant cosmological parameters, as well as knowledge of the redshift distribution of the source galaxies $n(z)$.

2.1.2 2-point statistics

2-point statistics such as correlation functions and power spectra are arguably the most powerful tool of analysis in cosmology. They work as a way to summarise raw data into something simpler, while still retaining most of the relevant information; for examples, allowing us to extract constraints on cosmological parameters.

Full sky

Let us start by introducing the statistics of a 3D spherically symmetric field, also known as a full sky field. The 2-point autocorrelation function is defined

as the expectation value $\mathbb{E}$ of the product of a random field with its complex conjugate

$$w(\boldsymbol{x}, \boldsymbol{y}) \equiv \mathbb{E}[\phi(\boldsymbol{x})\phi^*(\boldsymbol{y})]. \tag{2.3}$$

Where $\phi(\boldsymbol{x})$ is the 2D projection of a spherically symmetric 3D field. We can now perform a decomposition in spherical harmonics $Y_{\ell m}$. Such a decomposition gives rise to the correlation function associated to the harmonic coefficients $\phi_{\ell m}$, the 2D power spectrum C_ℓ,

$$\phi(\boldsymbol{x}) = \sum_{\ell m} \phi_{\ell m} Y_{\ell m}(\hat{\boldsymbol{x}}), \tag{2.4}$$

$$C_\ell \equiv \delta_{\ell m} \delta_{\ell' m'} \mathbb{E}[\phi_{\ell m} \phi^*_{\ell' m'}]. \tag{2.5}$$

In this setup correlation function and power spectrum are therefore related by,

$$\begin{aligned}
&\mathbb{E}[\phi(\boldsymbol{x})\phi^*(\boldsymbol{y})] \\
&= \sum_{\ell m} \sum_{\ell' m'} Y_{\ell m}(\hat{\boldsymbol{x}}) Y^*_{\ell' m'}(\hat{\boldsymbol{y}}) \mathbb{E}[\phi_{\ell m} \phi^*_{\ell' m'}] \\
&= \sum_{\ell m} Y_{\ell m}(\hat{\boldsymbol{x}}) Y^*_{\ell m}(\hat{\boldsymbol{y}}) C_\ell \\
&= \sum_\ell \frac{2\ell+1}{4\pi} C_\ell P_\ell(\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}}),
\end{aligned}$$

where we have used the identity relating Legendre polynomials to spherical harmonics $P_\ell(\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}}) = \frac{4\pi}{2\ell+1} \sum_m Y_{\ell m}(\hat{\boldsymbol{x}}) Y^*_{\ell m}(\hat{\boldsymbol{y}})$. Cleaning up the equations we are just left with the well known full sky equation

$$w(\theta) = \sum_\ell \frac{2\ell+1}{4\pi} C_\ell P_\ell(\cos(\theta)), \tag{2.6}$$

which relates the angular correlation function to the angular power spectrum.

Flat-sky

Most analysis of weak lensing data actually use the flat-sky approximation [14] [15] [16] [17] [18] [19], our work will not be an exception. Such an approximation essentially changes our setup to a flat-sky 2D field, instead of the 2D projection of a 3D one, allowing for simpler analytical expressions. In this setup, harmonic decomposition will not work, we will have to Fourier transform our space instead. To do so, we can assume the existence of a function $C(l)$ which is the continuous extension of C_ℓ. Then we say that $C(l)$ is the result of a 2D inverse Fourier transformation $\mathcal{F}$,

$$\{\mathcal{F}^{-1}C\}(\theta) = \int \frac{d^2l}{4\pi^2} e^{i\boldsymbol{l} \cdot \boldsymbol{\theta}} C(l). \tag{2.7}$$

To prove that $\{\mathcal{F}^{-1}C\}(\theta)$ is none other than the angular correlation function, we make use of the radial symmetry of the cosmological field. Which simplifies the equation into a 1D integral. Consider the polar coordinate substitution from (l_x, l_y) to (l, ϕ), the integral becomes

$$\int \frac{dl}{2\pi} l C(l) \int \frac{d\phi}{2\pi} e^{il\theta \cos \phi}.$$

Lastly, we make use of the identity $\int d\phi e^{il\theta \cos \phi} = 2\pi J_0(l\theta)$, where J_0 is the zeroth order Bessel function. Which leaves us with

$$\{\mathcal{F}^{-1}C\}(\theta) = \int \frac{dl}{2\pi} l C(l) J_0(l\theta). \tag{2.8}$$

In this form it is clear that *Eq.* (2.6) and *Eq.* (2.8) are asymptotically equivalent, since it is known that $J_0(\ell\theta) \longrightarrow P_\ell(\cos(\theta))$ as $\ell \to \infty$. This concludes our heuristic proof that the angular correlation function is recovered from an inverse Fourier transformation of the angular power spectrum $w(\theta) \sim \{\mathcal{F}^{-1}C\}(\theta)$ in the flat-sky approximation. Physically speaking this approximation makes sense when we consider a patch of sky of size L. The wavenumber of the flat-sky angular power spectrum is related to its dimensionless counterpart p by

$$l = \frac{2\pi}{L} p, \tag{2.9}$$

in other words, it is inversely proportional to the size of the map. Just as we would expect the geometry of a sphere to become flat when zooming in, the flat-sky approximation holds as $L \to 0$.

Limber approximation

The way we computationally obtain κ's 2-point statistics is with the Limber approximation. We use the code `jaxcosmo` [20] to compute the 2D power spectrum $C(l)$ from the 3D matter power spectrum $P_\delta(k)$ with the efficient Limber approximation,

$$C(l) = C_{\kappa\kappa}(l) = \int_0^{\chi*} \frac{d\chi}{\chi} W(\chi) W(\chi) P_\delta(k = \frac{l}{\chi}, \chi). \tag{2.10}$$

The assumptions of the Limber approximation are:

- flat-sky, as described previously and

- the matter power spectrum depends only on modes on the field $\boldsymbol{k}_\perp$, essentially setting the modes parallel to the line of sight to zero, $k_\parallel = 0$.

Such assumptions allow to simplify the relation between $C(l)$ and $P_\delta(k)$ to a one line integral, *Eq.* (2.10). It is noteworthy to mention that the Limber approximation introduces significant errors only for modes $l < 10$, as explained in detail in [17]. As far as this work is concerned, we only deal with patches of size $10°$, which means we have $l > 36$ according to *Eq.* (2.9), well within the range for a good approximation.

2.2   Field generation

2.2.1   Gaussian random field

In order to simulate cosmological fields with a specific power spectrum, we make use of Gaussian random fields. They are fast to generate and only need information about the 2-point power spectrum of the field. The algorithm we use for the generation of GRFs is common to most packages and is explained in detail in [21] [22]. In 1D, assume a pair of functions functions ξ and P, related by a Fourier transformation

$$\xi(x, y) = \int \frac{dk}{2\pi} e^{ik(x-y)} P(k).$$

Let $W(x)$ be a Gaussian white noise field and let $\mathcal{F}$ denote a Fourier transformation, we then define

$$\phi(x) \equiv (\mathcal{F}^{-1} P^{1/2} \mathcal{F} W)(x) \tag{2.11}$$

to be a Gaussian random field. Such a procedure ensures that the covariance $\mathbb{E}\left[\phi(x)\phi(y)\right]$ of the GRF recovers the correlation function $\xi(x, y)$,

$$
\begin{aligned}
&\mathbb{E}\left[\phi(x)\phi(y)\right] \\
&= \iiiint dx' dy' \frac{dk}{2\pi} \frac{dl}{2\pi} \, e^{i(kx+ly)} P(k)^{1/2} P(l)^{1/2} e^{-i(kx'+ly')} \mathbb{E}[W(x')W(y')] \\
&= \iint \frac{dk}{2\pi} \frac{dl}{2\pi} \, e^{i(kx+ly)} P(k)^{1/2} P(l)^{1/2} \int dx' \, e^{-i(k+l)x'} \\
&= \int \frac{dk}{2\pi} \, e^{ik(x-y)} P(k) \\
&= \xi(x, y).
\end{aligned}
$$

Where we have used $\mathbb{E}[W(x')W(y')] = \delta(x' - y')$, as white noise is defined by a constant power spectrum.
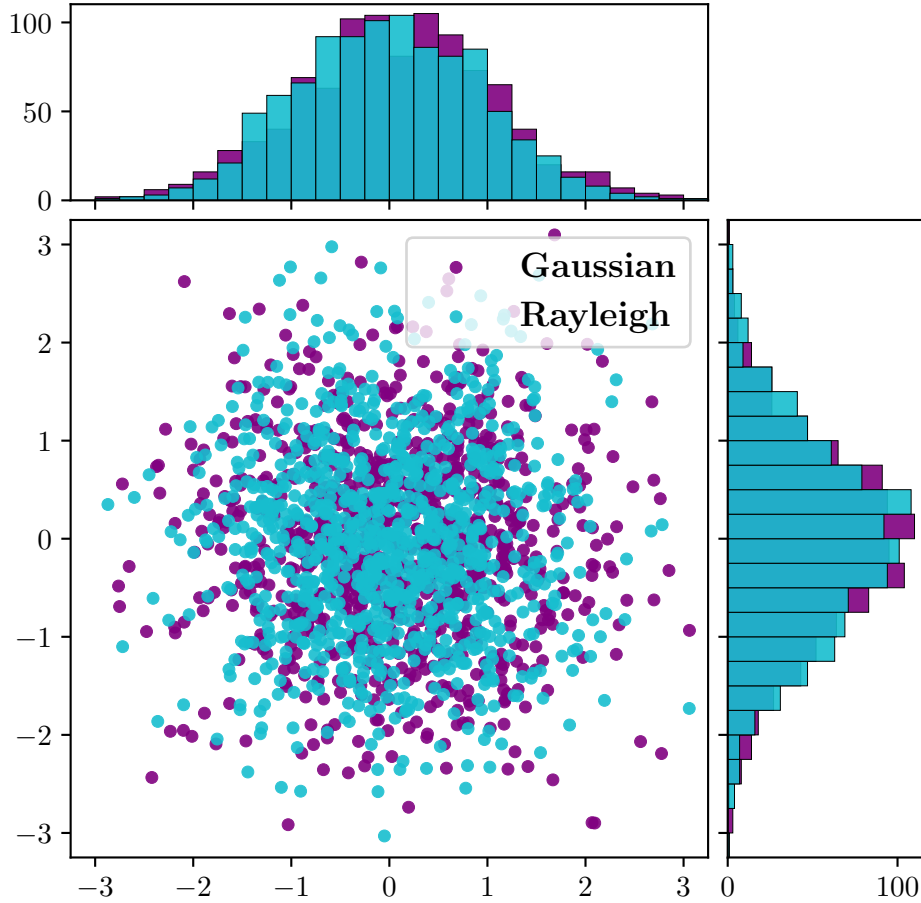
Rayleigh distribution

Since we are dealing with 2D maps, our algorithm will have to be implemented in two dimensions. To do so, we decide to implement the algorithm in *Eq. (2.11)* with the use of the Rayleigh distribution $\mathcal{R}(\sigma)$. Given two independent Gaussian random variables X and Y, the random variable R given by

$$R = \sqrt{X^2 + Y^2},$$

is said to be Rayleigh distributed. If we then multiply R by the complex exponential $e^{i\theta}$ of a uniformly distributed random variable $\theta \sim \mathcal{U}(0, 2\pi)$, we obtain a map of Gaussianly distributed complex numbers, which will substitute the $\mathcal{F}W$ term in *Eq. (2.11)*. We showcase this equivalency in *Fig. 2.1*

for distributions of $\mu = 0$ and $\sigma = 1$. The sequence of transformations in 2D therefore becomes,

$$\phi(x) \equiv (\mathcal{F}^{-1} P^{1/2} R e^{i\theta})(x). \tag{2.12}$$

Figure 2.1: Sampling a 2D Gaussian against a Rayleigh distributed amplitude with uniform complex phase. In purple is the complex number $X + iY$ with $X, Y \sim \mathcal{N}(0,1)$, in cyan is $Re^{i\theta}$ with $R \sim \mathcal{R}(1)$ and $\theta \sim \mathcal{U}(0, 2\pi)$.

2.2.2 Lognormal field

The universe today is not Gaussian. GRFs are able to capture the 2-point statistics of cosmological fields, but they cannot capture their skewed distribution. A possible way to get closer to the true distribution of the convergence field is with a lognormal transformation [23] [24] [25] [12]. In this work we will denote lognormal transformations as $\mathcal{L}$ and adopt the following convention for

both the field and the correlation function,

$$\mathcal{L}_w^{-1}(w^L, a) \equiv \log\left(\frac{w^L}{a^2} + 1\right) = w^G, \tag{2.13}$$

$$\mathcal{L}_\kappa(\kappa^G, a) \equiv a\left(\exp(\kappa^G - \tfrac{1}{2}\mathrm{Var}(\kappa^G)) - 1\right) = \kappa^L. \tag{2.14}$$

Where the superscripts L and G respectively stand for lognormal and Gaussian, $\mathrm{Var}(\cdot)$ stands for the variance of the field and a is the so-called shift parameter, which is an indicator of the non-Gaussianity of the resulting field.

2.3 Gaussian process

One way to think of Gaussian processes is as an extension of random vectors to infinite dimensions. Following this train of thought, let's begin with the concept of a random variable following a normal distribution. We say,

$$X \sim \mathcal{N}\left(\mu, \sigma^2\right),$$

to mean that X is a sample of a Gaussian of mean μ and variance σ^2. If we were to get enough samples X, we would eventually recover its distribution. The generalisation of this concept to n-dimensions is a collection of random variables, described by a so-called multivariate normal distribution,

$$\boldsymbol{X} \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{K}\right).$$

Where $\boldsymbol{X} = (X_0, X_1, ...)$ is a vector of random variables, $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{K}$ the covariance matrix. For a zero mean field, the covariance matrix is formed by the variance of each of the random variables on its diagonal, while the cross correlation terms populate the rest of the matrix.

2.3.1 Definition

Adopting the philosophy of Rasmussen [26], functions can be thought of as very long vectors of the form $(f(x_1), f(x_2), ...)$. Such a view allows us to extend the definition of multivariate Gaussians, to functions. Defining $\mathcal{GP}$ a Gaussian process, a sample function f will be given by:

$$f(\boldsymbol{x}) \sim \mathcal{GP}\left(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')\right) \tag{2.15}$$

with $m(\boldsymbol{x})$ and $k(\boldsymbol{x}, \boldsymbol{x}')$ defined as,

$$m(\boldsymbol{x}) = \mathbb{E}[f(\boldsymbol{x})], \tag{2.16}$$

$$k(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}[(f(\boldsymbol{x}) - m(\boldsymbol{x}))(f(\boldsymbol{x}') - m(\boldsymbol{x}'))], \tag{2.17}$$

Mathematically a GP is defined for a continuous function. Computationally this is not possible and we must treat space as a discrete grid.

2.3.2 Prior and posterior samples

Given some $m(\cdot)$ and $k(\cdot, \cdot)$ which define a GP, a random sample from said GP would be a function $\boldsymbol{f}_*$ defined on a domain D_*, which is our grid.

$$\boldsymbol{f}_* \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{K}_{**}). \tag{2.18}$$

Here we adopt the convention $\boldsymbol{K} = k(D_*, D_*)$. When drawing a sample function from *Eq.* (2.18), computationally the operation is equivalent to drawing a vector from a multivariate Gaussian. What we obtain is a so-called prior, or priors, see *Fig.* 2.2.
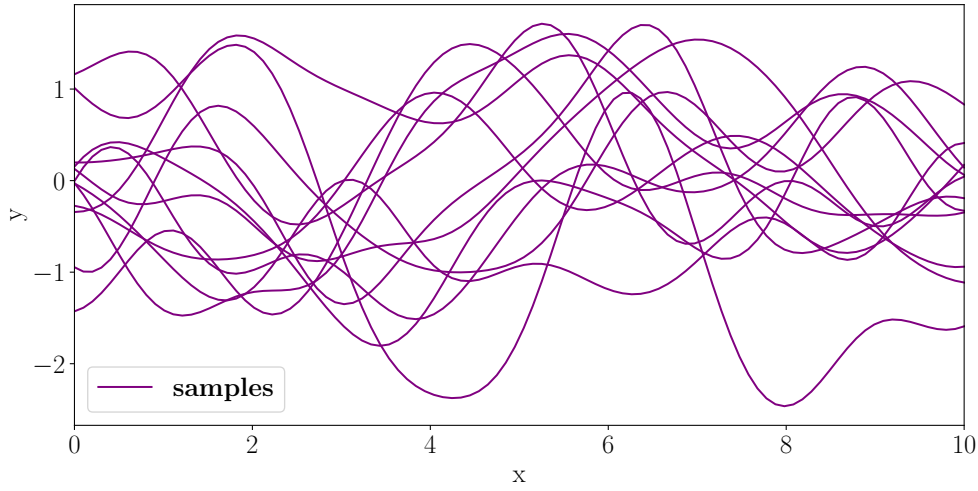
Figure 2.2: Prior samples of a GP with mean $m(\boldsymbol{x}) = 0$ and squared exponential kernel $k(\boldsymbol{x}, \boldsymbol{x'}) = e^{-(\boldsymbol{x} - \boldsymbol{x'})^2}$.

Let's now see how we can introduce knowledge of data points in this system. We divide the grid in training points D and test points D_*. To each training point is associated a known value $\boldsymbol{y}$ and variance σ_n^2, whereas the values of the function at the test points $\boldsymbol{f}_*$ are unknown. We can summarise this as,

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{f}_* \end{bmatrix} \sim \mathcal{N} \left(\boldsymbol{m}, \begin{bmatrix} \boldsymbol{K} + \sigma_n^2 I & \boldsymbol{K}_* \\ \boldsymbol{K}_*^T & \boldsymbol{K}_{**} \end{bmatrix} \right) \tag{2.19}$$

where we have once again adopted the notation $\boldsymbol{K} = k(D, D)$, $\boldsymbol{K}_* = k(D, D*)$, $\boldsymbol{K}_{**} = k(D_*, D_*)$. At this point, one way to find samples that follow the data would be to blindly draw priors until we get something that goes through all data points. This would be inefficient and computationally wasteful. Instead, we make a better guess for the test function values. This operation is called conditioning, because we condition the joint Gaussians on the training points, this gives

$$\boldsymbol{f}_* \mid D_*, D, \boldsymbol{y} \sim \mathcal{N} \left(\boldsymbol{K}_*^T [\boldsymbol{K} + \sigma_n^2 I]^{-1} \boldsymbol{y}, \boldsymbol{K}_{**} - \boldsymbol{K}_*^T [\boldsymbol{K} + \sigma_n^2 I]^{-1} \boldsymbol{K}_* \right). \tag{2.20}$$

11

Conditioning can therefore give rise to what is called a posterior sample, *Fig. 2.3*. The result is still a multivariate Gaussian, but the mean and variance given by *Eq.* (2.20) generate samples that are a better guess of the behaviour of the function outside of the training points.
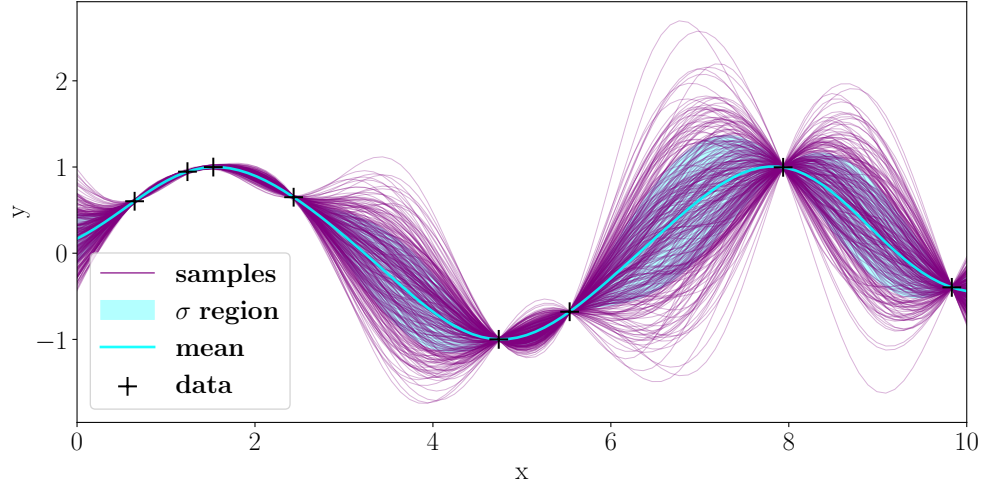
Figure 2.3: Summary plot of a GP conditioned to some data. The cyan line is the mean of the GP and the filled region corresponds to 1σ. The purple lines are posterior samples, which are distributed Gaussianly around the mean. The data points are clearly marked in black, they are also the points where all samples converge to.

Chapter 3

Methods

3.1 Data simulation

3.1.1 Cosmology

This work follows in the general footsteps of data analysis of weak lensing surveys like *HSC* [27] and *KiDS* [28][29], as we aim to replicate their methodologies. Throughout the work, we will assume a fiducial cosmology of fixed parameter values as shown in *Tab.* 3.1. In particular, all data maps will be generated from a power spectrum following this cosmology. We will refer to this power spectrum as the fiducial power spectrum $C(l)$. A leading modelling choice comes with the redshift distribution $n(z)$. We model it as a Smail-type distribution [30][31],

$$n(z) = z^\alpha \exp\left[-\left(\frac{z}{z_0}\right)^\beta\right]. \tag{3.1}$$

The choice of parameters has been made to emulate bin 5 of the *KiDS1000* survey [32], with parameters $\alpha = 3.5$, $\beta = 4.5$, $z_0 = 1$.

Table 3.1: Fiducial cosmology values given to `jaxcosmo`.

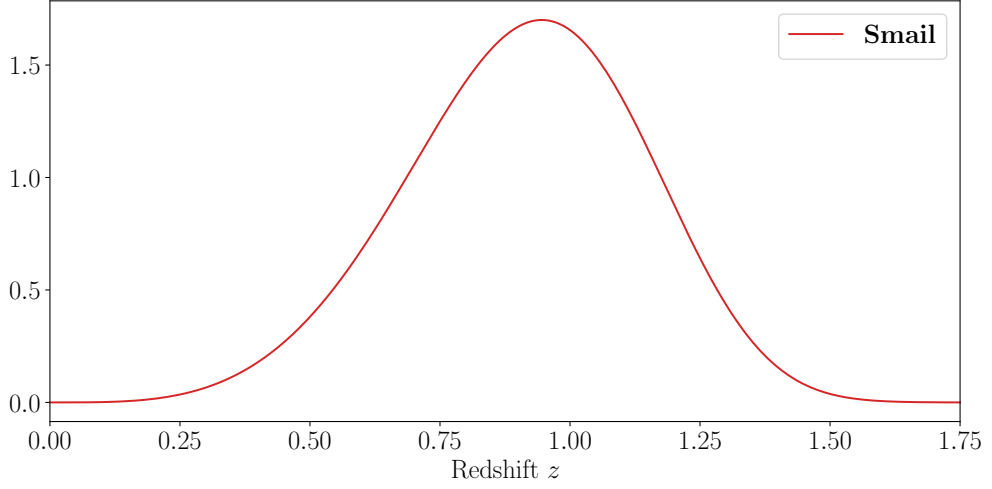|	Fiducial
Ω_m	0.3
Ω_b	0.05
Ω_k	0
h	0.7
n_s	0.97
σ_8	0.8
w_0	-1
w_a	0

3.1.2 Map making pipeline

When dealing with maps of finite size (L, L) and pixel resolution (N, N), Fourier space is going to have boundaries, just as real space does. These limits are given by

$$l_{\min} = \frac{2\pi}{L},$$
$$l_{\max} = \frac{2\pi}{L}N.$$

13

Figure 3.1: Smail-type redshift distribution for the chosen parameters: $\alpha = 3.5$, $\beta = 4.5$, $z_0 = 1$.

A map of size $(10°, 10°)$ and grid 64×64 will therefore have limits $(l_{\min}, l_{\max}) = (36, 2304)\mathrm{rad}^{-1}$. Now that we have the physical range for the power spectrum, we can generate data.
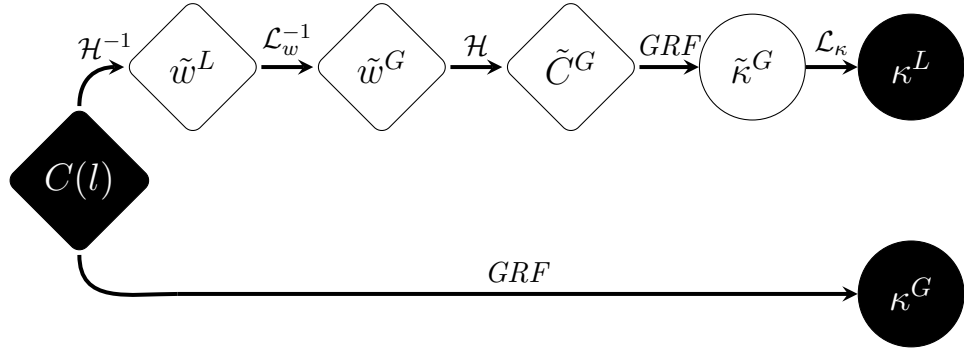

Figure 3.2: Sequence of transformations used to generate Gaussian and lognormal maps starting from the fiducial angular power spectrum. The top branch shows how to generate a GRF which when transformed to a lognormal field, follows the fiducial $C(l)$. The bottom branch is a standard GRF realisation starting from the fiducial $C(l)$. L and G stand for lognormal and Gaussian respectively. $\mathcal{H}$ is the Hankel transformation, and tilde refers to intermediate results.

In order to generate a GRF, we employ the algorithm mentioned in *Eq.* (2.12). Whereas to make a lognormal field, the matter is a bit more complicated. Any GRF on which we apply the lognormal transformation $\mathcal{L}_\kappa$, from *Eq.* (2.14), becomes a lognormal field. The reason we don't just lognormal transform any field is given by the fact that they wouldn't recover the fiducial power spectrum $C(l)$. Our goal is therefore to find a transfer GRF $\tilde{\kappa}^G$ which, when

lognormally transformed, gives rise to a lognormal field κ^L that does recover $C(l)$. To do so, we follow the sequence of transformations illustrated in the top branch of *Fig.* 3.2.

3.1.3 Noise and mask
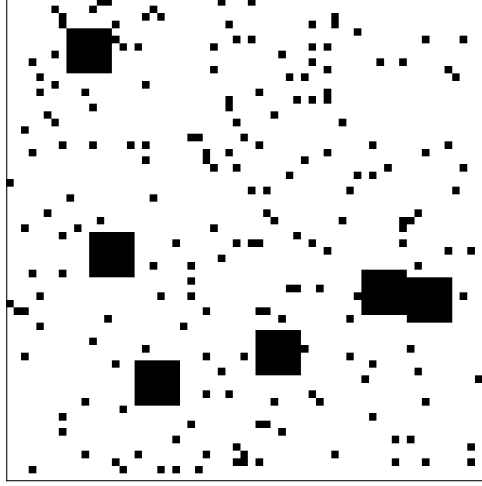
Figure 3.3: Approximately 10% mask applied to the data. Size $(10°, 10°)$ and grid 64×64.

As all data is being simulated, we only take into account the so-called shape noise, which is due to the intrinsic distribution of ellipticities and angle formed with respect to us. GPs will treat each pixel of the map as a random variable Gaussianly distributed with standard deviation given by [33],

$$\sigma_{\text{noise}} = \frac{\sigma_e}{\sqrt{n_g A_{px}}}.$$ (3.2)

We use values of $\sigma_e = 0.26$, $n_g = 4, 10, 30, 100$ galaxies/arcmin^2, and a pixel area given by the pixel resolution squared, $A_{px} = (L/N)^2$. For our highest resolution run, we have $N = 64$ and use $n_g = 10$ galaxies/arcmin^2, which results in a noise standard deviation of $\sigma_{\text{noise}} \sim 0.0088$. The mask being used as seen in *Fig.* 3.3, covers approximately 10% of the patch [34]. We keep the noise and mask random seeds fixed throughout the work.

3.2 Kernel

The kernel of a Gaussian process is given by a function of the form $k(x, y)$. It takes in two points, x and y, and returns the value of their correlation. In our case specifically, x and y will be two points of a grid of shape (N, N), and the kernel function will be the convergence angular autocorrelation function. As all code used for this paper is written in JAX [35] we opted for the use of the

library `tinygp` [36] for all Gaussian process computations. `tinygp` allows for the use of a custom kernel with a custom `evaluate` method, which takes two points on grid and returns the correlation value. We follow with two `Python` pseudocodes of our kernel implementations.

Listing 3.1: `kernel_Hankel` uses the helper function `Hankel` which returns a 1D callable correlation function `w`. Then finds the euclidean distance between x and y and evaluates the correlation function at that point. We will discuss how we perform the Hankel transformation in the following section.

```python
class kernel_Hankel:
    def __init__(self, cl, N, L):
        self.w   = Hankel(cl, N, L)
        self.r   = L / N

    def evaluate(self, x, y):
        theta    = self.r * sqrt(sum(x - y))
        return self.w(theta)
```

Listing 3.2: `kernel_FFT` performs a 2D Fourier transform on the power spectrum `cl2D`. This returns a 2D array that is the correlation function. Then to evaluate `w2D` we need two indeces. These are given by the difference of x and y component wise. Furthermore we have to be careful about possible mismatches between the shape of `cl2D` and the grid of our map. For this we have the renormalisation factor `r`. We will discuss further how we get `cl2D` in the following section.

```python
class kernel_FFT:
    def __init__(self, cl2D, N, L):
        M        = sqrt(cl2D.size)
        self.w2D = abs(ifft2(cl2D)) * M**2 / L**2
        self.r   = M / N

    def evaluate(self, x, y):
        d0       = self.r * abs(x[0] - y[0])
        d1       = self.r * abs(x[1] - y[1])
        return self.w2D[int(d0)][int(d1)]
```

3.2.1 Hankel transform

In order to build a Gaussian process kernel, we need to find the correlation function that best describes the data. The way we do this is by computing the angular power spectrum and transforming it to the corresponding angular correlation function. Let's explore the previously discussed flat-sky relation between angular power spectrum and correlation function given by *Eq.* (2.8). This particular integral of a Bessel function is also known as a zeroth order Hankel transformation,

$$w(\theta) = \int \frac{dl}{2\pi} l C(l) J_0(l\theta). \tag{3.3}$$

We will explore two methods for computing this integral: the *integration* method *Sec.* 3.2.1 and the *FFTlog* method *Sec.* 3.2.1.

Integration

Integration is the most straight forward way to evaluate the integral, but it requires to implement an algorithm for the approximation of the Bessel function J_0. The advantages of this method:

- it is easy to integrate over the correct l-range, from $l_{\min}$ and $l_{\max}$, a freedom that we do not have with the FFTlog.

The disadvantages:

- integration is computationally slow, especially when dealing with highly oscillatory behaviour introduced by the Bessel function, which requires fine sampling.

FFTlog

The FFTlog method [37] is a fast implementation of the Hankel transformation. In fact it simplifies *Eq.* (3.3) by assuming a power decomposition of $C(l)$. Such a decomposition is achievable by taking the *fast Fourier transformation (FFT)* in $\log k$, hence the name FFTlog. The power spectrum becomes,

$$C(l) = \sum_\alpha c_\alpha l^{\upsilon + i\eta_\alpha}. \tag{3.4}$$

Substituting in *Eq.* (3.3),

$$w(\theta) = \sum_\alpha c_\alpha \int_0^\infty \frac{dl}{2\pi} l^{\upsilon + i\eta_\alpha + 1} J_0(l\theta).$$

Take $x = l\theta$ and $s_\alpha - 1 = \upsilon + i\eta_\alpha + 1$,

$$w(\theta) = \sum_\alpha c_\alpha \theta^{-s_\alpha} \int_0^\infty \frac{dx}{2\pi} x^{s_\alpha - 1} J_0(x).$$

Lastly we recognise $\int_0^\infty dx x^{s-1} J_0(x) = \frac{2^{s-1}}{\pi} \sin(\pi s/2)[\Gamma(s/2)]^2$, a Mellin transform. Using this tabulated result we conclude that the correlation is given by the sum,

$$w(\theta) = \frac{1}{2\pi^2} \sum_\alpha c_\alpha \theta^{-s_\alpha} 2^{s_\alpha - 1} \sin(\pi s_\alpha/2)[\Gamma(s_\alpha/2)]^2. \tag{3.5}$$

For the Mellin transform identity to hold analytically, the integral bounds have to go from 0 to ∞. Although computationally we don't need to consider such a wide range, we still have to broaden the integration limits to something larger than our l-range; so as to avoid ringing effects. Experimentally we have found that extending the l-range between $l_{\min}/4$ and $l_{\max}$, is enough to compensate for such effects. The advantages of this method:

17

- the FFTlog method is ultimately much faster than the integration method.

The disadvantages:

- the widening of the *l*-range needed to avoid ringing effects inevitably adds power to the correlation function resulting in a higher variance and overall amplitude.

3.2.2   Fourier transform

The second transformation that we can use to go from power spectrum to correlation function, is the Fourier transformation discussed in *Eq.* (2.7). It is a more fundamental relation than the Hankel transformation as it does not assume that *l* is radially symmetric. In our case, low resolution, coupled with a square map, means that the radial assumption might not apply. Note that, the lowest resolution grid we use is 32×32, with $L/N \sim 19$ arcmin; which is one order of magnitude worse than today's weak lensing data, around ~ 3 arcmin according to [12]. In the following *Sec.* 3.2 we will find experimentally that the Hankel methods described above do not work well with our GP setup. For this reason, we explore this method of conversion between angular power spectrum and correlation function,

$$w(\theta) = \int \frac{d^2l}{4\pi^2} e^{-i\boldsymbol{l} \cdot \boldsymbol{\theta}} C(l).$$ (3.6)

As computers can only deal with discrete functions, it is important to note that we will be performing a *discrete Fourier transformation (DFT)*. We will use the widely known FFT algorithm to compute DFTs. In particular we use JAX's implementation of the 2D FFT algorithm, `jax.numpy.fft.ifft2`.

Let us find the relation between continuous Fourier transformation and DFT. The definitions of continuous and discrete Fourier transformations in 1D, are respectively:

$$\mathcal{F}^{-1} = \int \frac{dk}{2\pi} e^{ikx},$$ (3.7)

$$\mathrm{F}^{-1} = \frac{1}{N} \sum_p e^{ik_p x}.$$ (3.8)

First of all, a DFT is dimensionless. Secondly, it is discrete and bounded. We can therefore rewrite *Eq.* (3.7) using the substitution $k(p) = 2\pi \frac{p}{L}$ to discretise k-space,

$$\int \frac{dk}{2\pi} e^{ikx} = \int \frac{dp}{L} e^{ik(p)x} = \frac{1}{L} \sum_p e^{ik_p x}.$$

Applying the definition of DFT as seen in *Eq.* (3.8), it follows that in 1D

$$\mathcal{F}^{-1} = \frac{N}{L} \mathrm{F}^{-1}.$$ (3.9)

Which means that the correlation function is given by a *backwards* normalised inverse DFT to be scaled by a factor $(N/L)^d$, where d is the dimension of the considered space. In our case, $d = 2$.

In order to perform a FFT in 2D, we will need a two dimensional extension of the angular power spectrum. We make use of its radial symmetry with respect to $\boldsymbol{l} = (l_x, l_y)$ and create a 2D grid of shape (M, M) as shown in *Fig.* 3.4. Now
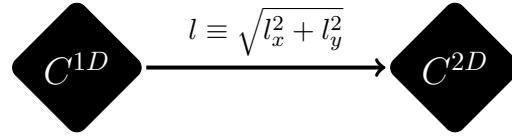

$$l \equiv \sqrt{l_x^2 + l_y^2}$$

Figure 3.4: 1D to 2D extension of the power spectrum $C(l)$

that we have a 2D power spectrum we can take the inverse two dimensional fast Fourier transformation to obtain a 2D correlation function. In practice, we will test two grids, which we name *full-range FFT* method and *half-range FFT* method.

Full-range FFT

The full-range FFT is defined by a grid of shape $(2N, 2N)$. Since the grid is radial, it will be centered. This implies that if we want to keep information of all N modes,

$$\frac{2\pi}{L}, \frac{4\pi}{L}, \cdots, \frac{2\pi}{L}N, \tag{3.10}$$

the grid will have to be at least of shape $M = 2N$. The advantages of this method:

- it keeps information on the full range of modes.

The disadvantages:

- it introduces rounding errors, as the field has shape (N, N) and not all of the possible distance combinations of such a grid are covered bya grid of shape $(2N, 2N)$.

Half-range FFT

The half-range FFT is defined by a grid of shape (N, N). The advantages of this method:

- has the perk of having no shape mismatch between field and correlation function.

The disadvantages:

- it loses half of the l-range, missing information on small scales.

Chapter 4
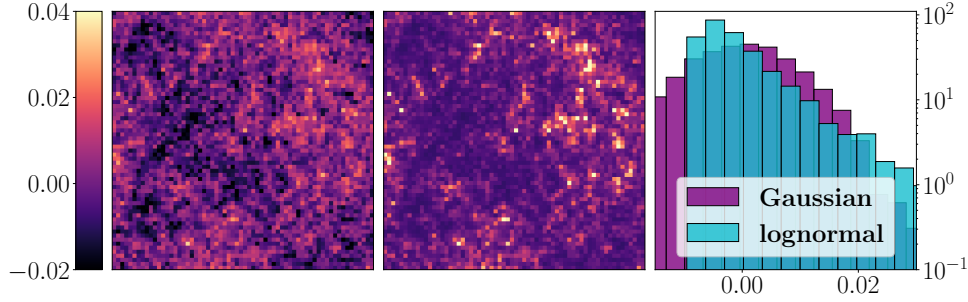
Results

4.1 Power spectrum recovery

4.1.1 Gaussian and lognormal fields

We begin by testing the consistency of our Gaussian and lognormal maps generation pipeline.

Figure 4.1: Comparison of a Gaussian map, on the left, with a lognormal map on the right. Both maps arise from the same random seed. The colorbar has been adjusted to enhance the differences between the two. The histogram plot shows the clear difference in the map distributions.

We show example realisations of the two fields in *Fig.* 4.1, Gaussian on left and lognormal on the right. There's a visible difference between the two, as it can be seen clearly from the distribution plot. The main check to perform is for testing whether the generated fields recover the theoretical power spectrum. *Fig.* 4.2 shows that this is the case for the Gaussian fields. They recover the fiducial $C(l)$ within a few percent error, with larger deviations $\sim 5\%$ at the low and high ends of the l-range. Instead, lognormal fields present deviations $\gtrsim 10\%$. As the lognormal transformations we use have been reported by different sources [12][23], the issue must lie with our `JAX` implementation of the Hankel

transformation. Resolving such issues could be achieved by future iterations of this work. In this work, we restrict ourselves to the use of Gaussian fields, as it is enough to prove our thesis and show that Gaussian processes can be applied to cosmological fields.
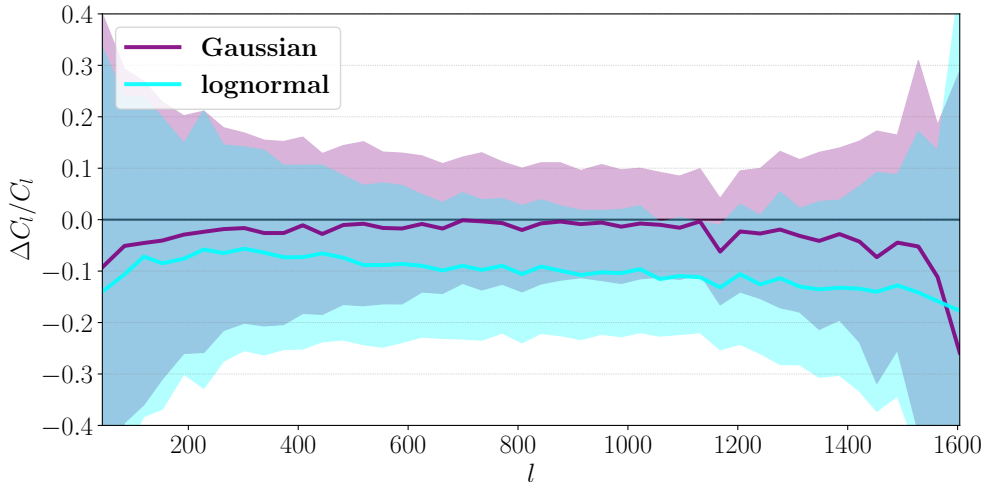

Figure 4.2:   Power spectrum estimation from Gaussian and lognormal maps. Mean and standard deviation are calculated with 500 realisation of both fields.

4.1.2   Gaussian process priors

First, we test the ability of the kernels we have built in *Sec.* 3.2 to recover the power spectrum of our cosmology. We test five models in *Fig.* 4.3: integration, FFTlog, full-range FFT, half-range FFT and sinc FFTlog. The first four methods are described in the Gaussian process kernels *Sec.* 3.2, whereas the sinc FFTlog referes to a FFTlog model on which we applied smoothing, by multiplying the power spectrum by a factor of $sinc^4(l\frac{L}{2\pi N})$. The recovered power spectra are plotted against the fiducial power spectrum, or smoothed power spectrum for the sinc FFTlog. Mean and standard deviation associated to the plots are calculated from 500 samples. As expected the integration, FFTlog and full-range FFT perform similarly, as they all contain the same ammount of information. As these models deviate so strongly from the fiducial power spectrum we tried applying smoothing, which helps to recover half of the *l*-range at large scales. The only method that seems to be consistently recovering the fiducial power spectrum is the half-range FFT. One could argue that due to the inherent discreteness and boundedness of the fields we are working with, using FFTs is the most natural choice; also, half-range FFT uses the only grid that recovers a correlation function of the same shape as the field without having to perform binning.
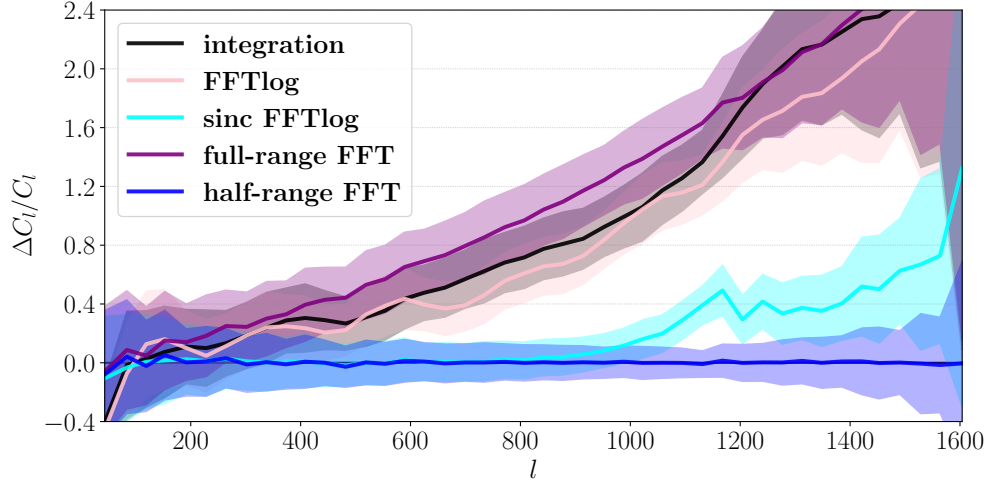
Figure 4.3:   Reconstructed power spectrum from prior sample of GP with the four proposed kernels: integration, FFTlog, full-range FFT, half-range FFT and sinc FFTlog. Mean and standard deviation are calculated with 500 prior samples from each GP.
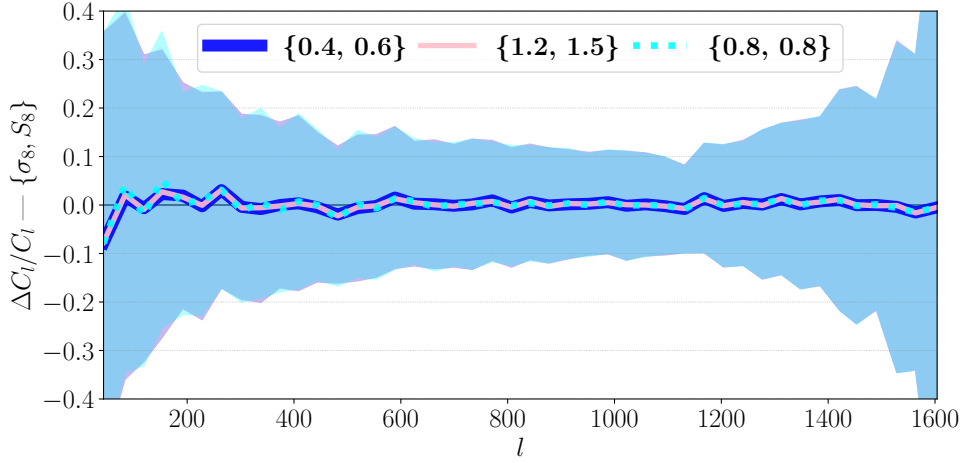

Figure 4.4:   Reconstructed power spectrum from prior samples of a GP, as a function of $\{\sigma_8, S_8\}$. Mean and standard deviation are calculated with 500 prior samples for each different cosmology.

We have also tested the efficacy of the half-range FFT model for different cosmologies of values $\{\sigma_8, S_8\}$ equal to $\{0.4, 0.2\}$, $\{1.2, 1.5\}$ and, our fiducial cosmology, $\{0.8, 0.8\}$. As *Fig.* 4.4 shows, the model is independent of the choice of cosmology. From here on the results will be presented assuming a kernel built with the half-range FFT model.

4.2 Gaussian process map reconstruction

Armed with a reliable kernel, let's embark upon the journey of reconstructing a heavily masked cosmological field. What we will do is: create a noiseless GRF in the fiducial cosmology *Tab.* 3.1, *True* map; apply a mask to obtain the *Data* map; condition a Gaussian process which assumes the fiducial cosmology. *Fig.* 4.5 lists the result of this operation, showing the resulting mean μ and standard deviation σ of the conditioned GP. We also plot the ratio between residuals $\Delta = \mu - True$ and standard deviation squared, to test the goodness of fit of our model, the values of the map sum up to $\chi^2 \sim 2495$. With the mask covering $\nu = 2353$ pixels, we obtain $\chi^2/\nu = 1.06$. Of course, this is just a noiseless application, which is unreasonable for a real application.
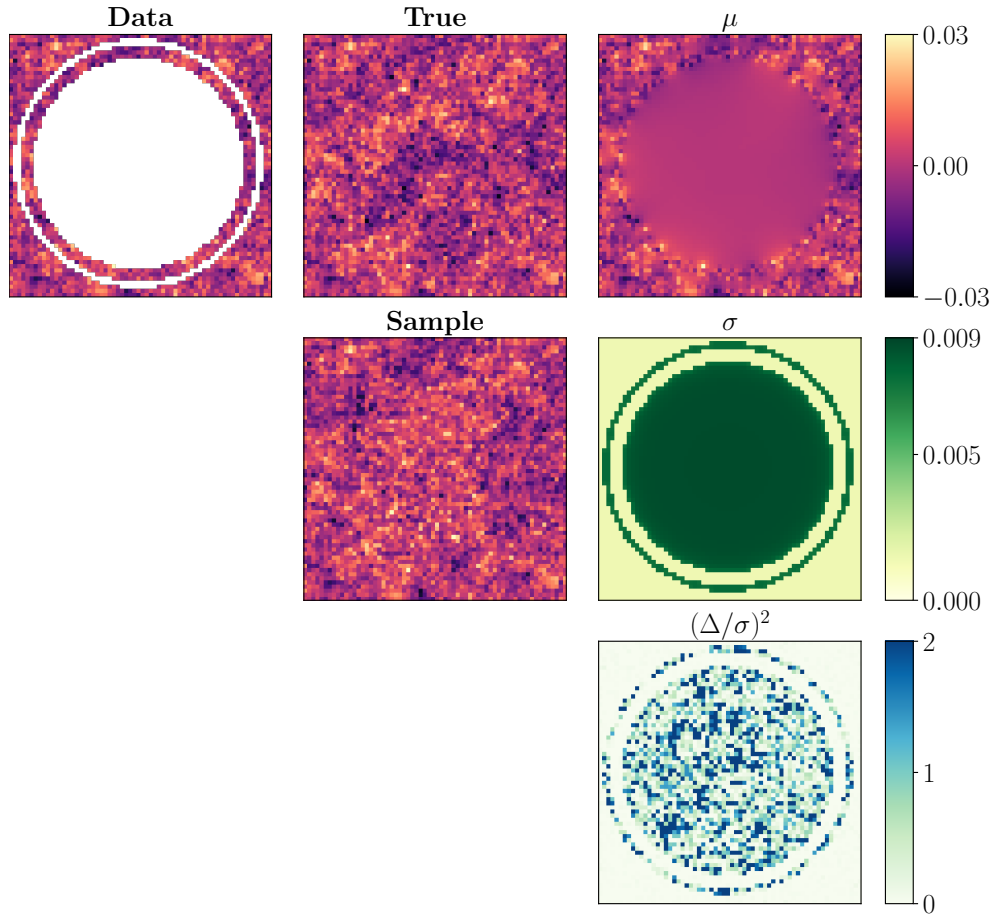
Figure 4.5: Summary of field reconstruction abilities of a Gaussian process conditioned on data. The left column shows the masked GRF, which is our data. The middle column shows the true GRF without masks and a posterior sample drawn from the conditioned GP. The right column shows maps of the mean, standard deviation and residuals over standard deviation squared of the conditioned GP. Regions of higher uncertainty correspond to the masked regions. The residuals over standard deviation map also shows how regions with low mask recover the data.

4.3 Inference of cosmological parameters

To test the ability of Gaussian processes to recover cosmological parameters without any prior knowledge except a noisy and masked map, we perform a MCMC simulation to infer the posterior distributions of σ_8 and S_8. We use the convention

$$S_8 = \sigma_8 \sqrt{\frac{\Omega_m}{0.3}}, \tag{4.1}$$

Table 4.1: Priors used by our `numpyro` model for cosmological parameter inference

	Prior
S_8	$\mathcal{U}[0.565, 1.78]$
σ_8	$\mathcal{U}[0.4, 1]$
Ω_m	$\mathcal{U}[0.15, 0.95]$

to infer deterministically a posterior for Ω_m. Such a reparametrisation is needed due to the strong degeneracy between σ_8 and Ω_m. *Eq.* (4.1) breaks this degeneracy, changing the geometry of the sampling space and making the sampling more consistent. The model assumes uninformed flat priors for the cosmological parameters, as shown in *Tab.* 4.1, such prior bounds are also in accordance with the `jaxcosmo` release [20]. The likelihood of the model is given by a Gaussian process distribution conditioned on *Data*, with a standard deviation equal to the noise applied to the map. The analysis is coded with `numpyro` [38] [39], using a the *No-U-Turn Sampler (NUTS)* method with `max_tree_depth=16`, `target_accept _prob=0.8`. We simulate 8 chains for the 32×32 grid and 4 chains for the 64×64. Each chain performs 1000 warmup steps and 3000 samples.

4.3.1 One parameter

As a first step and for a consistency check, we run the inference model for one cosmological parameter, keeping all others fixed. Using a 64×64 grid with $n_g = 10$ galaxies/arcmin^2. In *Fig.* 4.6 we show the inferred distribution for both σ_8 and Ω_m. We find that we are able to recover the true value for both parameters within two sigmas, $\sigma_8 = 0.776 \pm 0.015$ and $\Omega_m = 0.284 \pm 0.010$. We notice a slight tendency of the inferred distribution to be biased low; a tendency we also observe next for both sampled parameters, S_8 and σ_8.

4.3.2 Two parameters

Effect of noise

We perform some tests on low resolution 32×32 grids to see the effect that the noise level has on the recovered parameters, see *Tab.* 4.2. Here we report the inferred cosmological parameters for one data realisation and different noise levels, corresponding respectively to $n_g = 4$, 10, 30 and 100 galaxies/arcmin^2, see *Eq.* (3.2). The inferred value of S_8 can vary as much as a full σ between high and low noise runs. Keeping in mind that σ_8 and Ω_m are extremely unreliable due to relative uncertainties of $\sim 25-30\%$ caused by the degeneracy:
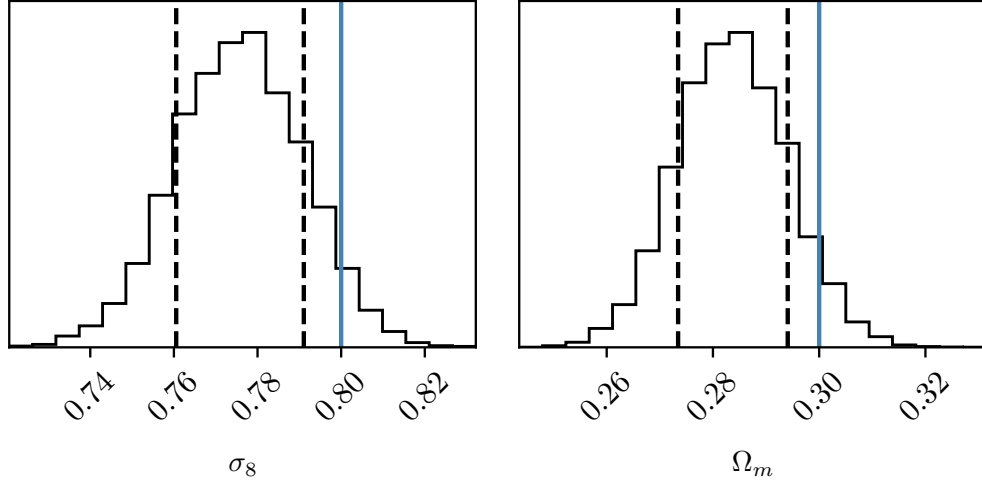
Figure 4.6: Inferred posterior distribution of σ_8 on the left and Ω_m on the right. Dotted lines indicate the 1σ level. Truth values corresponding to the fiducial cosmology are indicated in blue.

Table 4.2: List of inferred cosmological parameters inferred by the model with a small 32×32 grid and for a fixed true GRF realisation. We present the cosmological parameters inferred as we increase the noise level, corresponding to $n_g = 4$, 10, 30 and 100 galaxies/arcmin^2.

σ_{noise}	0.0069	0.0044	0.0025	0.0014
S_8	0.716 ± 0.043	0.733 ± 0.040	0.747 ± 0.041	0.752 ± 0.042
σ_8	0.645 ± 0.157	0.628 ± 0.158	0.632 ± 0.166	0.631 ± 0.170
Ω_m	0.423 ± 0.164	0.469 ± 0.175	0.485 ± 0.186	0.497 ± 0.193

as a general trend we notice Ω_m gets bigger when σ_8 gets smaller with less noise.

Inferred cosmological parameters

Running the model for a larger 64×64 grid with $n_g = 10$ galaxies/arcmin^2, gives much better constraints on the cosmological parameters. We present the values recovered by the posterior distributions, listed as follows in *Tab.* 4.3.

Table 4.3: Mean and sigma values recovered from the inferred distributions of the cosmological parameters.

S_8	σ_8	Ω_m
0.762 ± 0.028	0.745 ± 0.151	0.353 ± 0.143

Fig. 4.7 shows the inferred posterior distributions and contours for the three cosmological parameters σ_8, Ω_m and S_8. Looking at the contours, we obtain the well known banana-shaped degeneracy between σ_8 and Ω_m. The S_8 and

Ω_m contour presents sharp cuts for high and low Ω_m, indicating an issue with the bounds of the uniform priors imposed. Unfortunately the `jaxcosmo` package does not allow for the choice of priors to be wider than what shown in *Tab.* 4.1, as the model then starts to have divergent samples.
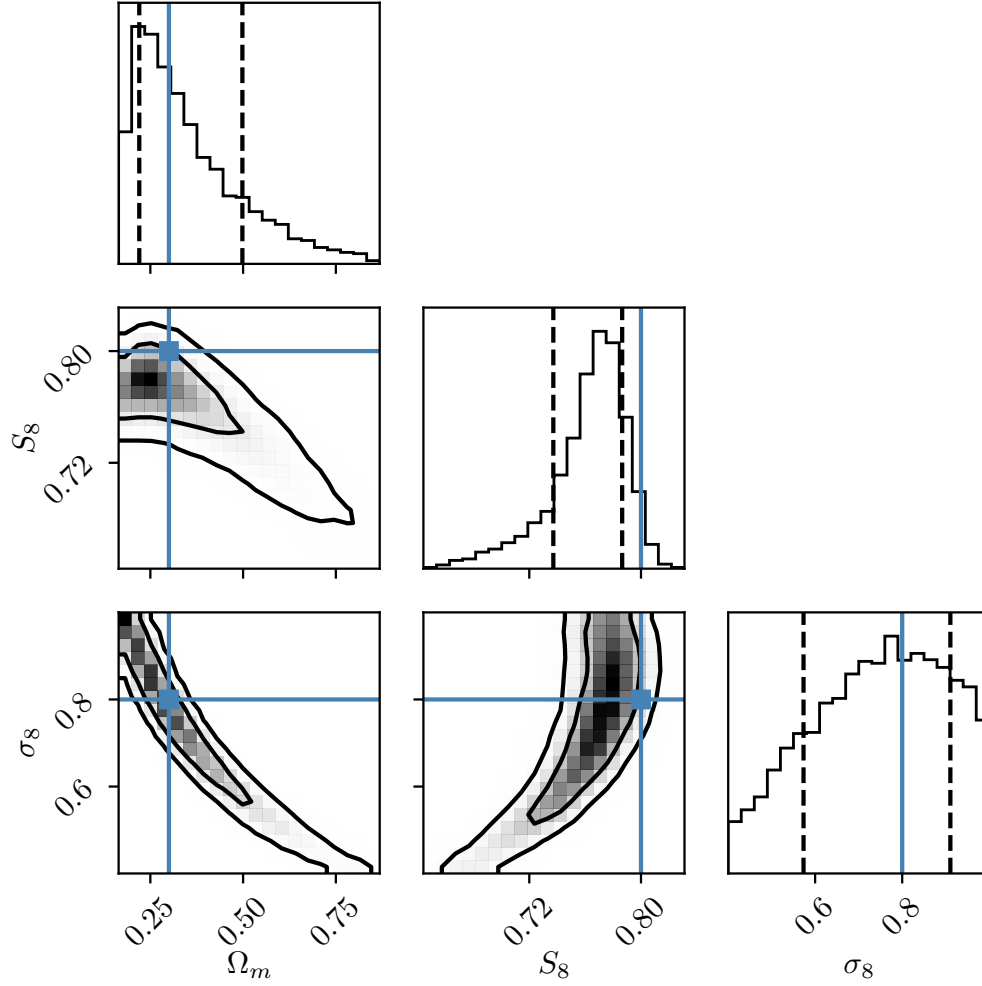
Figure 4.7: Inferred posterior distributions of S_8, σ_8 and Ω_m. For noise level $\sigma_{\text{noise}} \sim 0.0088$. Contours indicate the 1σ and 2σ credible interval respectively. Dotted lines indicate the 1σ level. Truth values corresponding to the fiducial cosmology are indicated in blue.

4.3.3 Posterior checks

Following the two parameter inference model, we perform some posterior checks at the map level [40]. *Fig.* 4.9 sums up the ability of the model to recover the true map, noiseless and unmasked. Here we present the run with noise level $\sigma_{\text{noise}} \sim 0.0088$ and a 64×64 grid. We show the mean and standard

deviation for the sample with highest likelihood out of the 12000. The mean field μ is visibly different to the true field in the masked regions and it seems to be of overall lower amplitude. The sample map is comparable to the noisy data; which is to be expected, as the internal noise given to the Gaussian process is the same as the noise level of the data. The standard deviation map σ presents an overall amplitude comparable to the noise level ~ 0.010, with higher values for the masked regions. Summing up the map values of the residuals divided by standard deviation squared, we obtain a $\chi^2 \sim 1297.4$. Compared to the number of free parameters ν in our inference model, which for a 10% mask and a 64×64 grid, is $\nu = 3689$. The value of χ^2 therefore seems to be low, indicating that the noise level assumed by the GP is overestimated. This is supported by the fact that the sample map looks just as noisy as the data, according to *Fig.* 4.8, its distribution is in fact just as wide as the noise.
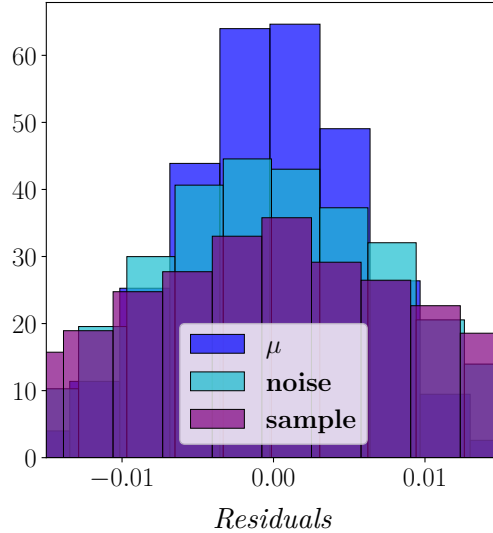
Figure 4.8: Residual distributions of the mean and sample compared to noise. The mean is less spread, whereas the sample is wider.
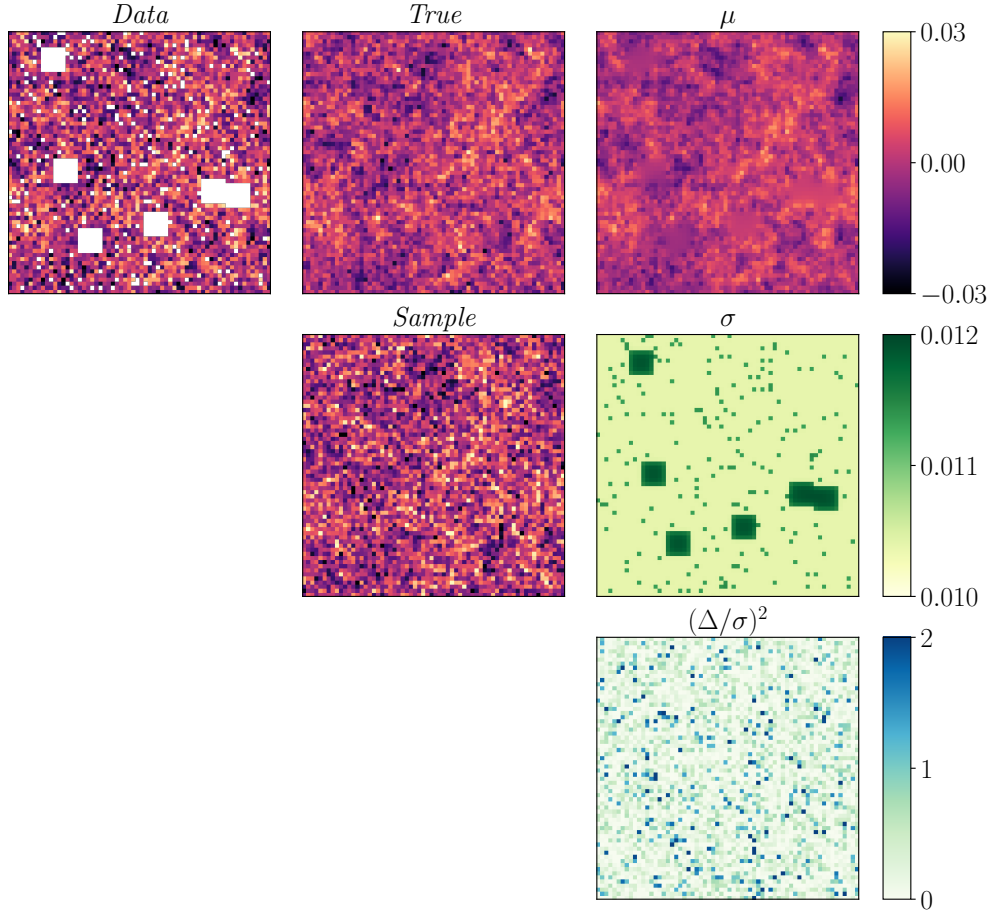
Figure 4.9: Summary of the two parameter inference at the map level. The left column shows the masked and noisy GRF realisation used, which is our data. The middle column shows the true GRF and a sample from the conditioned GP. The right column shows maps of the mean, standard deviation and residuals over standard deviation squared resulting from the numpyro model sample with highest likelihood. Regions of higher uncertainty correspond to the masked regions.

Chapter 5

Conclusion

We have hereby introduced the tool of Gaussian processes to the landscape of map inference of cosmological fields, in particular weak lensing convergence. We considered how the 2-point statistics of cosmological fields changes when we are dealing with bounded and discrete maps in *Sec.* 2.1. We discussed the realisation of Gaussian and lognormal fields in *Sec.* 2.2, and showed their ability to recover the 2-point statistics that they encode, in *Sec.* 4.1.1. We have included masking and noise to the data to simulate realistic maps in *Sec.* 3.1. We have shown that it is possible to apply physical knowledge about the 2-point correlation function of a cosmological field in order to set up a Gaussian process able to produce a Gaussian realisation of such a field. We considered different set ups for the Gaussian process kernel in *Sec.* 3.2 and showed how they fare against one another in *Sec.* 4.1.2, ultimately proving empirically that the half-range FFT model is the best. In *Sec.* 4.2 we present an application of Gaussian processes to a noiseless masked convergence map, in order to showcase its ability to reconstruct a heavily masked map. Finally we present our results for the cosmological parameters inference with GPs, conditioning on noisy and masked data. When running the inference model on one cosmological parameter we recover both parameters within two sigmas, $\sigma_8 = 0.776 \pm 0.015$ and $\Omega_m = 0.284 \pm 0.010$. For the two parameter inference we observe the well known banana-shaped degeneracy between σ_8 and Ω_m, as well as recovering 0.762 ± 0.028 within two sigmas.

In future studies GPs could be tested on maps with larger grids. In order to achieve a resolution of ~ 3 arcmin with a map of size $(10°, 10°)$, a 200×200 grid is needed. Too big for a GP. The bottleneck is given by the inversion of the kernel matrix, see (2.20). Approximations of this operation could enable the use of GPs on larger grids. This can lead to the possibility of applying this method on current weak lensing catalogues and perhaps even full sky catalogues. Another pathway to explore are lognormal fields, as they do a much better job at simulating data than GRFs. Due to GPs being Gaussian,

their associated likelihood is not suitable to treat lognormal fields. A modified likelihood could therefore unlock a correct application of GPs to lognormal fields.

Acknowledgments

Thank you, reader.

Bibliography

[1] M. Kilbinger, "Cosmology with cosmic shear observations: a review", Reports on Progress in Physics **78**, 086901 (2015) (pages 1, 5).

[2] M. Bartelmann and M. Maturi, "Weak gravitational lensing", Scholarpedia **12**, revision #186553, 32440 (2017) (pages 1, 5).

[3] R. D. Blandford, A. B. Saust, T. G. Brainerd, and J. V. Villumsen, "The distortion of distant galaxy images by large scale structure", AIP Conference Proceedings **222**, 455–458 (1991) (pages 1, 5).

[4] A. Mootoovaloo, A. F. Heavens, A. H. Jaffe, and F. Leclercq, "Parameter inference for weak lensing using Gaussian Processes and MOPED", Monthly Notices of the Royal Astronomical Society **497**, 2213–2226 (2020) (page 2).

[5] S. S. Boruah, T. Eifler, V. Miranda, and P. M. S. Krishanth, "Accelerating cosmological inference with Gaussian processes and neural networks – an application to LSST Y1 weak lensing and galaxy clustering", Monthly Notices of the Royal Astronomical Society **518**, 4818–4831 (2022) (page 2).

[6] K. Karchev, A. Coogan, and C. Weniger, "Strong-lensing source reconstruction with variationally optimized Gaussian processes", Monthly Notices of the Royal Astronomical Society **512**, 661–685 (2022) (page 2).

[7] A. Shafieloo, A. G. Kim, and E. V. Linder, "Gaussian process cosmography", Phys. Rev. D **85**, 123530 (2012) (page 2).

[8] M. Seikel, C. Clarkson, and M. Smith, "Reconstruction of dark energy and expansion dynamics using gaussian processes", Journal of Cosmology and Astroparticle Physics **2012**, 036 (2012) (page 2).

[9] T. Holsclaw, U. Alam, B. Sansó, H. Lee, K. Heitmann, S. Habib, and D. Higdon, "Nonparametric reconstruction of the dark energy equation of state", Phys. Rev. D **82**, 103502 (2010) (page 2).

[10] G. Chon, A. Challinor, S. Prunet, E. Hivon, and I. Szapudi, "Fast estimation of polarization power spectra using correlation functions", Monthly Notices of the Royal Astronomical Society **350**, 914–926 (2004) (page 2).

[11] M. L. Brown, P. G. Castro, and A. N. Taylor, "Cosmic microwave background temperature and polarization pseudo-C_ℓ estimators and covariances", Monthly Notices of the Royal Astronomical Society **360**, 1262–1280 (2005) (page 2).

[12] A. J. Zhou, X. Li, S. Dodelson, and R. Mandelbaum, "Accurate field-level weak lensing inference for precision cosmology", (2023) (pages 3, 9, 18, 21).

[13] S. Dodelson and F. Schmidt, "13 - probes of structure: lensing", in Modern cosmology (second edition), edited by S. Dodelson and F. Schmidt, Second Edition (Academic Press, 2021), pp. 373–399 (page 5).

[14] A. Nicola, C. García-García, D. Alonso, J. Dunkley, P. G. Ferreira, A. Slosar, and D. N. Spergel, "Cosmic shear power spectra in practice", Journal of Cosmology and Astroparticle Physics **2021**, 067 (2021) (page 6).

[15] Z. Gao, A. Raccanelli, and Z. Vlah, "Asymptotic connection between full- and flat-sky angular correlators", Phys. Rev. D **108**, 043503 (2023) (page 6).

[16] W. L. Matthewson and R. Durrer, "The flat sky approximation to galaxy number counts", Journal of Cosmology and Astroparticle Physics **2021**, 027 (2021) (page 6).

[17] Z. Gao, Z. Vlah, and A. Challinor, "Flat-sky angular power spectra revisited", (2023) (pages 6, 7).

[18] C. García-García, D. Alonso, and E. Bellini, "Disconnected pseudo-C_ℓ covariances for projected large-scale structure data", Journal of Cosmology and Astroparticle Physics **2019**, 043 (2019) (page 6).

[19] Schneider, P., van Waerbeke, L., Kilbinger, M., and Mellier, Y., "Analysis of two-point statistics of cosmic shear - i. estimators and covariances", A&A **396**, 1–19 (2002) (page 6).

[20] J.-E. Campagne, F. Lanusse, J. Zuntz, A. Boucaud, S. Casas, M. Karamanis, D. Kirkby, D. Lanzieri, A. Peel, and Y. Li, "Jax-cosmo: an end-to-end differentiable and gpu accelerated cosmology library", The Open Journal of Astrophysics **6**, 10.21105/astro.2302.05163 (2023) (pages 7, 25).

[21] E. Bertschinger, "Multiscale gaussian random fields and their application to cosmological simulations", The Astrophysical Journal Supplement Series **137**, 1 (2001) (page 8).

[22] A. Lang and J. Potthoff, Monte Carlo Methods and Applications **17**, 195–214 (2011) (page 8).

[23] S. S. Boruah, E. Rozo, and P. Fiedorowicz, "Map-based cosmology inference with lognormal cosmic shear maps", (2022) (pages 9, 21).

[24] F. Leclercq and A. Heavens, "On the accuracy and precision of correlation functions and field-level inference in cosmology", Monthly Notices of the Royal Astronomical Society: Letters **506**, L85–L90 (2021) (page 9).

[25] Hilbert, S., Hartlap, J., and Schneider, P., "Cosmic shear covariance: the log-normal approximation", A&A **536**, A85 (2011) (page 9).

[26] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (The MIT Press, Nov. 2005) (page 10).

[27] C. Hikage, M. Oguri, T. Hamana, S. More, R. Mandelbaum, M. Takada, F. Köhlinger, H. Miyatake, A. J. Nishizawa, H. Aihara, R. Armstrong, J. Bosch, J. Coupon, A. Ducout, P. Ho, B.-C. Hsieh, Y. Komiyama, F. Lanusse, A. Leauthaud, R. H. Lupton, E. Medezinski, S. Mineo, S. Miyama, S. Miyazaki, R. Murata, H. Murayama, M. Shirasaki, C. Sifón, M. Simet, J. Speagle, D. N. Spergel, M. A. Strauss, N. Sugiyama, M. Tanaka, Y. Utsumi, S.-Y. Wang, and Y. Yamada, "Cosmology from cosmic shear power spectra with Subaru Hyper Suprime-Cam first-year data", Publications of the Astronomical Society of Japan **71**, 43 (2019) (page 13).

[28] F. Köhlinger, M. Viola, B. Joachimi, H. Hoekstra, E. van Uitert, H. Hildebrandt, A. Choi, T. Erben, C. Heymans, S. Joudaki, D. Klaes, K. Kuijken, J. Merten, L. Miller, P. Schneider, and E. A. Valentijn, "KiDS-450: the tomographic weak lensing power spectrum and constraints on cosmological parameters", Monthly Notices of the Royal Astronomical Society **471**, 4412–4435 (2017) (page 13).

[29] Asgari, Marika, Lin, Chieh-An, Joachimi, Benjamin, Giblin, Benjamin, Heymans, Catherine, Hildebrandt, Hendrik, Kannawadi, Arun, Stölzner, Benjamin, Tröster, Tilman, van den Busch, Jan Luca, Wright, Angus H., Bilicki, Maciej, Blake, Chris, de Jong, Jelte, Dvornik, Andrej, Erben, Thomas, Getman, Fedor, Hoekstra, Henk, Köhlinger, Fabian, Kuijken, Konrad, Miller, Lance, Radovich, Mario, Schneider, Peter, Shan, HuanYuan, and Valentijn, Edwin, "Kids-1000 cosmology: cosmic shear constraints and comparison between two point statistics", A&A **645**, A104 (2021) (page 13).

[30] I. Smail, D. W. Hogg, L. Yan, and J. G. Cohen, "Deep optical galaxy counts with the keck telescope*", The Astrophysical Journal **449**, L105 (1995) (page 13).

[31] D. Kirk, A. Rassat, O. Host, and S. Bridle, "The cosmological impact of intrinsic alignment model choice for cosmic shear", Monthly Notices of the Royal Astronomical Society **424**, 1647–1657 (2012) (page 13).

[32] Hildebrandt, H., van den Busch, J. L., Wright, A. H., Blake, C., Joachimi, B., Kuijken, K., Tröster, T., Asgari, M., Bilicki, M., de Jong, J. T. A., Dvornik, A., Erben, T., Getman, F., Giblin, B., Heymans, C., Kannawadi, A., Lin, C.-A., and Shan, H.-Y., "Kids-1000 catalogue: redshift distributions and their calibration", A&A **647**, A124 (2021) (page 13).

[33] R. A. C. Croft, P. E. Freeman, T. S. Schuster, and C. M. Schafer, "Prediction of galaxy ellipticities and reduction of shape noise in cosmic shear measurements", Monthly Notices of the Royal Astronomical Society **469**, 4422–4427 (2017) (page 15).

[34] N. Grewal, J. Zuntz, and T. Tröster, "Comparing mass mapping reconstruction methods with minkowski functionals", (2024) (page 15).

[35] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, *JAX: composable transformations of Python+NumPy programs*, version 0.3.13, 2018 (page 15).

[36] D. Foreman-Mackey, W. Yu, S. Yadav, M. R. Becker, N. Caplar, D. Huppenkothen, T. Killestein, R. Tronsgaard, T. Rashid, and S. Schmerler, *dfm/tinygp: The tiniest of Gaussian Process libraries*, version v0.3.0, Jan. 2024 (page 16).

[37] M. Simonović, T. Baldauf, M. Zaldarriaga, J. J. Carrasco, and J. A. Kollmeier, "Cosmological perturbation theory using the fftlog: formalism and connection to qft loop integrals", Journal of Cosmology and Astroparticle Physics **2018**, 030 (2018) (page 17).

[38] D. Phan, N. Pradhan, and M. Jankowiak, "Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro", arXiv e-prints, arXiv:1912.11554, arXiv:1912.11554 (2019) (page 25).

[39] E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos, R. Singh, P. A. Szerlip, P. Horsfall, and N. D. Goodman, "Pyro: deep universal probabilistic programming", J. Mach. Learn. Res. **20**, 28:1–28:6 (2019) (page 25).

[40] N. Porqueres, A. Heavens, D. Mortlock, and G. Lavaux, "Bayesian forward modelling of cosmic shear data", Monthly Notices of the Royal Astronomical Society **502**, 3035–3044 (2021) (page 27).

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every written paper or thesis authored during the course of studies. In consultation with the supervisor, one of the following three options must be selected:

- (●) I confirm that I authored the work in question independently and in my own words, i.e. that no one helped me to author it. Suggestions from the supervisor regarding language and content are excepted. I used no generative artificial intelligence technologies[1].

- ( ) I confirm that I authored the work in question independently and in my own words, i.e. that no one helped me to author it. Suggestions from the supervisor regarding language and content are excepted. I used and cited generative artificial intelligence technologies[2].

- ( ) I confirm that I authored the work in question independently and in my own words, i.e. that no one helped me to author it. Suggestions from the supervisor regarding language and content are excepted. I used generative artificial intelligence technologies[3]. In consultation with the supervisor, I did not cite them.

Title of paper or thesis:

Weak lensing map inference: a physics informed Gaussian processes approach

Authored by:
If the work was compiled in a group, the names of all authors are required.

Last name(s):	**First name(s):**
Massari	Nicolo'

With my signature I confirm the following:
- – I have adhered to the rules set out in the Citation Guide.
- – I have documented all methods, data and processes truthfully and fully.
- – I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for originality.

Place, date	**Signature(s)**
Zürich, 02.04.2024	Nicolò Massari

If the work was compiled in a group, the names of all authors are required. Through their signatures they vouch jointly for the entire content of the written work.

---

[1] E.g. ChatGPT, DALL E 2, Google Bard
[2] E.g. ChatGPT, DALL E 2, Google Bard
[3] E.g. ChatGPT, DALL E 2, Google Bard