Weak lensing analysis of the Canada-France Imaging Survey : from pixels to cosmology, preparation for the Euclid mission

Axel Guinot

Université de Paris

Ecole doctorale: Astronomie et d'Astrophysique d'Ile-de-France, 127
Laboratoire: Commissariat à l'Energie Atomique et aux Energies Alternatives
Thèse de doctorat en Astronomie et Astrophysique

Weak Lensing analysis of the Canada-France Imaging Survey: from pixels to cosmology, preparation for the Euclid mission

Axel Guinot

Présentée et soutenue publiquement le 10 Décembre 2020

Directeur: Dr. Martin Kilbinger
Co-Directeur: Pr. Yannick Mellier

Composition du jury:
Président:
Dr. Ken Ganga, Professor, APC

Rapporteurs:
Pr. Gary Bernstein, UPenn, USA
Pr. Jean-Paul Kneib, EPFL, Switzerland

Examinateurs:
Dr. Mathilde Jauzac, Assistant Professor, Durham University, UK
Dr. Eric Jullo, Astronome adjoint, LAM
Dr. Cécile Roucelle, Maître de conférence, APC

Invité:
Pr. Nicholas Kaiser, Professor, ENS

To my grandfathers...

Abstract

Among the big questions cosmology faces today, the nature of dark matter and dark energy are at the center of upcoming surveys. The future stage IV missions Euclid and LSST will cover a surface on the sky never reached before to unveil structures at very large scales and different epochs. Weak gravitational lensing will be one of the cosmological probes used to trace dark matter. Gravitational lensing is a physical phenomenon which uses the distortion of light to trace the presence of mass in the Universe. The interesting point of weak lensing is its sensitivity to total mass, *i.e.* baryonic and non-baryonic. Due to gravitational lensing, distant galaxies will appear distorted on the observed images.

The measurement of the distortions induced by gravitational shear requires a very accurate estimation of the shape of galaxies. This thesis will present the data reduction pipeline built for weak lensing studies, from the telescope to cosmological parameter inference. The work focuses on the analysis of the Canada-France Imaging Survey (CFIS), a u- and r-band survey covering 5,000 $\deg^2$ in the Northern hemisphere. The high resolution and depth of those data make it one the best survey candidates for weak lensing science to date. Among other things, accurate measurement of the shape galaxies requires a very good knowledge of the PSF for which a suite of validation tests have been developed. Due to the noise, and approximations used in the shape measurement, the results can be biased. The residual multiplicative and additive biases have been reduced to $\Delta m < 0.1\%$ and $c < 0.001\%$ respectively by using state-of-art techniques such as Metacalibration. This thesis will also present the work that is required for the development of a weak lensing pipeline, such as the elaboration of highly accurate and data-representative image simulations. We will show validation tests performed to ensure systematic-free measurements.

Finally, preliminary science results will be presented demonstrating the viability of the pipeline. We have constructed maps of dark matter over a surface of 2,000 $\deg^2$. We have measured tangential shear around 50 clusters and compared to theoretical predictions. To conclude, we will present a first 3x2 points analysis combining the weak lensing study performed on CFIS and the redshift measurement from eBOSS observations on the 50 $\deg^2$ chosen for science verification purposes.

Keywords: Cosmology, Observation, Weak gravitational lensing, Data processing

Résumé

Parmi les grandes questions auxquelles la cosmologie fait face aujourd'hui, la nature de la matière noire et de l'énergie noire sont au centre des relevés à venir. Les futures missions de stade IV Euclid et LSST vont couvrir une surface du ciel jamais atteinte auparavant dans le but de révéler les structures aux très grandes échelles et de différentes époques. Le lentillage gravitationnel faible sera une des sondes cosmologiques utilisées pour tracer la matière noire. Le lentillage gravitationnel est un phénomène physique qui utilise la distorsion de la lumière pour tracer la présence de masses dans l'Univers. Ce qui est intéressant avec le lentillage gravitationnel faible est sa sensitivité à la masse totale, *i.e.* baryonique et non-baryonique. Dû au lentillage gravitationnel, les galaxies distantes apparaissent distordues sur les images observées.

La mesure des distorsions provoquées par le cisaillement gravitationnel requiert une très précise estimation des formes des galaxies. Cette thèse présentera la chaîne de réduction de données construites pour l'étude du lentillage gravitationnel faible, depuis le télescope jusqu'à l'inférence des paramètres cosmologiques. Le travail se focalise sur l'analyse du relevé Canada-France Imaging Survey (CFIS), un relevé couvrant 5,000 $\deg^2$ de l'hémisphère Nord dans les bandes u et r. La haute résolution et la profondeur de ces données en font à ce jour un des meilleurs candidats pour l'étude de la science du lentillage gravitationnel faible. Entre autres choses, la mesure précise de la forme des galaxies nécessite une très bonne connaissance de la PSF pour laquelle une suite de tests ont été développés pour la validation. Dû au bruit, et l'utilisation d'approximations pour la mesure de formes, les résultats peuvent être biaisés. Par l'utilisation de techniques de pointe comme la Metacalibration, les biais multiplicatifs et additifs résiduels ont été réduits à $\Delta m < 0.1\%$ et $c < 0.001\%$ respectivement. Cette thèse présentera aussi le travail qui est demandé pour le développement d'une chaîne de traitements pour du lentillage gravitationnel faible, comme l'élaboration de simulations d'images très précises et représentatives des données. Nous présenterons les tests de validations réalisés pour assurer une mesure dénuée d'erreurs systématiques.

Enfin, des résultats scientifiques préliminaires seront présentés pour démontrer la viabilité de la chaîne de traitements. Nous avons construit des cartes de matière noire sur une surface de 2,000 $\deg^2$. Nous avons mesuré et comparé aux prédictions théoriques le cisaillement gravitationnel tangentiel autour d'environ 50 amas. Pour conclure, nous présenterons une première analyse 3x2 points combinant le lentillage gravitationnel faible de l'étude réalisée sur CFIS et la mesure du décalage vers le rouge des galaxies des observations de eBOSS sur les 50 $\deg^2$ choisis dans le but de vérifications scientifiques.

Mots clés: Cosmologie, Observation, Lentillage gravitationnel faible, Traitement de données

Acknowledgements

I would like first to thank K. Ganga, G. Bernstein, J.-P. Kneib, M. Jauzac, E. Jullo, C. Roucelle and N. Kaiser for accepting to be members of my Jury. A particular thanks to my reviewers G. Bernstein and J.-P. Kneib for taking the time to thoroughly read my manuscript and provide comments to improve it.

I would like to thanks the DIM-ACAV labex, the CEA and the CNES for putting their trust in me by founding this project.

Now, I would like to thanks Martin Kilbinger, who first took me as an intern before being my PhD supervisor. I got the opportunity to work on real observations, be deeply involved in two large international collaborations, learn more than I would never imagine, participate to several meetings, and more. I want to thank you for taking the time to answer all my questions (almost everyday) and all the discussion we had. I also thank the entire CosmoStat team. J.-L. Starck who lead the team and helped me a lot with my work. S. Farrens, M. Schmitz, F. Nammour and K. Themelis with who I shared an office and all the others who were always there to support me.

I thank J.-C. Cuillandre and R. Gavazzi for being my "comité de suivi" during those three years. The weak-lensing group of UNIONS, M. Hudson, L. Van Waerbeke, S. Fabbro, L. Miller and I. Spitzer. They followed my work and gave me feedback.

I have to thank the all my friends, the "Cosmo team", who support me and with who I spent great time at the Euclid summer school and after-work parties. I could not go through this PhD without them.

Finally, I thank all my family and particularly my parents. They always were behind me and bring the support I needed during bad times and share my joy during the best time.

Contents

List of Figures **xiii**

List of Tables **xv**

List of Acronyms **xvii**

1 Introduction **1**
1.1 Cosmology . . . 2
1.1.1 General Relativity . . . 2
Metric . . . 2
Geodesics . . . 3
Einstein equation . . . 3
1.1.2 **Λ**CDM . . . 4
the "standard model" . . . 4
The Cosmological Principle . . . 5
A word on redshift . . . 6
1.1.3 Gravitational lensing . . . 7
Basics of gravitationel lensing . . . 7
Weak Lensing formalism . . . 7
1.2 Lensing observables . . . 9
1.2.1 Basic principle . . . 10
1.2.2 Tangential alignment and correlation functions . . . 11
Tangential alignment . . . 11
Correlation functions . . . 11
1.2.3 PSF introduction . . . 11
1.2.4 Quadrupole moments . . . 12
General method . . . 12
Adaptive moments . . . 13
The KSB method . . . 13
1.2.5 Model fitting . . . 14
Methodology . . . 14
1.2.6 Shear bias . . . 15
Model bias . . . 15
Noise bias . . . 15
Detection bias . . . 15
Selection bias . . . 16
Additive bias . . . 16

2 Data and Simulations **17**
2.1 Weak-Lensing surveys . . . 17
2.1.1 The CFIS survey . . . 17
The survey strategy . . . 18

Data pre-processing . . . 19
2.2 Simulations . . . 20
2.2.1 The Point Spread Function (PSF) . . . 22
2.2.2 The galaxies . . . 23

3 A weak lensing pipeline **25**
3.1 Masking . . . 25
3.1.1 Why? What? . . . 25
3.1.2 How? . . . 26
3.2 Source extraction and classification . . . 27
3.2.1 Object identification . . . 27
Blends . . . 28
3.2.2 Classification . . . 29
Star selection . . . 29
Galaxy selection . . . 29
3.3 PSF modeling . . . 34
3.3.1 PSFEx model . . . 34
3.3.2 Stacked PSF model . . . 34
3.4 Shape measurements . . . 36
3.4.1 Metacalibration . . . 36
3.4.2 Multi-epoch shape measurement . . . 37
3.4.3 Shape measurement on the stack . . . 38

4 Validation tests **39**
4.1 PSF Validation . . . 39
4.1.1 Residuals . . . 39
4.1.2 ρ-statistics . . . 40
4.1.3 Validation of the stacked PSF . . . 43
4.2 Shear systematics . . . 44
4.2.1 Multiplicative bias . . . 44
4.2.2 Additive bias . . . 45
4.3 PSF/Shapes correlations . . . 46
4.3.1 PSF leakage . . . 47
4.3.2 PSF systematics . . . 51

5 Cosmological results **55**
5.1 Redshift distribution of lensed galaxies . . . 58
5.2 Cluster lensing . . . 59
5.2.1 Theoretical model . . . 59
5.2.2 Mass estimation . . . 59
5.3 Mass mapping . . . 60
5.4 Parameter inference . . . 62
5.4.1 Test on a simulation . . . 68
5.4.2 Setup . . . 69
Lensing data . . . 69
LRG sample . . . 69
Covariance matrix . . . 69
Correlation functions and theoretical model . . . 70
5.4.3 Results . . . 70

6 Conclusion **75**

Bibliography **77**

A PSF leakage **83**

B Metacalibration response **99**
Patch P1 99
Patch P2 99
Patch P4 99

List of Figures

1.1 Representation of ds. 3
1.2 Strong lensing system observed by Hubble Space Telescope (HST). Credit: NASA/ESA 8
1.3 This figure shows the propagation of two light beams emitted by a galaxy in the source plan through a cluster located in the lens plane. 9

2.1 Map showing the light distribution of the Milky Way. Credit: Euclid consortium 18
2.2 Map showing the dust distribution mainly coming from the Milky Way. Credit: Euclid consortium 19
2.3 Map showing the zodiacal light on the ecliptic plane. Credit: Euclid consortium 19
2.4 Ultra-violet Near-Infrared Optical Northern Survey (UNIONS) footprint. Credit: UNIONS collaboration 20
2.5 Astrometric residuals with respect to Gaia DR2. Credit: Stephen Gwyn . . 21
2.6 Magnitude zero-point residuals accros MegaCam field of view after calibration. Credit: Stephen Gwyn 21
2.7 Examples of simulated galaxies. 24

3.1 ShapePipe flowchart diagram. 26
3.2 Example of a mask for a chip. Green: border mask. Red: spikes around bright stars. Light blue: halos around bright stars. Light green: dead pixels. 27
3.3 Mode of the Full Width at Half Maximum (FWHM) of the stars for each Charge-Coupled Device (CCD). 30
3.4 Example of a size-magnitude plot. The selected stars are show in orange while the other objects are ploted in blue. 30
3.5 Examples of simulated objects. *Top row*: Stacked stars. *Bottom row*: Stacked galaxies 31
3.6 Spread-Model classification. In *Blue*: the real stars. In *Orange*: the real galaxies. In *Green*: the objects selected as galaxies. The *Threshold* represent a cut at 0.0035. 32
3.7 Spread-Model classification. The *orange* delimited area corresponds to the objects that have been selected for the galaxy sample. 33
3.8 Average number of stars selected to construct the PSF model for entire processed area. I set the lower limit to 20 stars to ensure a good model. . . 35

4.1 Average number of stars selected for testing the model. 40
4.2 PSF ellipticity residuals in the focal plane. 41
4.3 PSF size residuals in the focal plane. 41
4.4 PSF residuals ρ-statistics for the multi-epoch PSF model. The *grey* area represents the requirement described in 4.1.2. Here $T_{\mathrm{PSF}}/T_{\mathrm{gal}} = 1$ and $\alpha = 0.03$. For the computation of ξ_+ Core Cosmology Library (CCL) is used. 43

4.5 PSF residuals ρ-statistics for the stacked PSF model. The *grey* area represents the requirement described in 4.1.2. Here $T_{\rm PSF}/T_{\rm gal} = 1$ and $\alpha = 0.03$. For the computation of ξ_+ CCL is used. 43
4.6 Residual multiplicative bias without weighting and calibration. (For clarity, only one fifth of the points are shown) 45
4.7 Residual multiplicative bias after calibration without weighting. (For clarity, only one fifth of the points are shown) 45
4.8 Residual multiplicative bias after calibration and weighting. (For clarity, only one fifth of the points are shown) 46
4.9 PSF leakage using the average galaxy shape in bins of PSF ellipticity (top two panels) and size (bottom panel) on the field P3 for the ngmix catalogue. 48
4.10 PSF leakage using the average galaxy shape in bins of PSF ellipticity (top two panels) and size (bottom panel) on the field P3 for the Galsim catalogue. 49
4.11 PSF leakage using the average galaxy shape in bins of true PSF ellipticity (top two panels) and size (bottom panel) on the simulated data. 50
4.12 $\xi_\pm^{\rm sys}(\theta)$ measured for ngmix (*Top*) and Galsim (*Bottom*). 52
4.13 Ratio of $\xi_+^{\rm sys}(\theta)/\xi_+^{\rm ss}(\theta)$ for ngmix (*Top*) and Galsim (*Bottom*). For Galsim the systematics are larger at smaller scales compared to ngmix. 53
4.14 Leakage α shown as a function of scale θ for ngmix (*Top*) and Galsim (*Bottom*). 54

5.1 Process patches of the Canada-France Imaging Survey (CFIS) dataset. . . 56
5.2 Diagonal terms of the response martrix for the patch P3. The *solid line* for the galaxies and the *dashed line* for the stars. 57
5.3 Magnitude distribution for the patch P3. 57
5.4 Galaxy distribution of matched galaxies derived from the redshifts measured in Canada France Hawaii Telescope Lensing Survey (CFHTLenS). 58
5.5 Number of clusters as function of mass (SZ) and redshift. 60
5.6 Stacked shear profiles using CFIS weak-lensing data (blue curves with statistical error bars), and best-fit Navarro-Frenk-White (NFW) model. The three panels show the three bins, see Table 5.1. 61
5.7 Mass map for the patch P1. The *black crosses* represent the positions of Planck clusters. The value on *top* of each cross is the cluster redshift, and the *bottom* value indicates the cluster mass ($10^{14}{\rm M}_\odot$). The *top* (*bottom*) panel shows the E-mode (B-mode). 63
5.8 Mass map for the patch P2. The *black crosses* represent the positions of Planck clusters. The value on *top* of each cross is the cluster redshift, and the *bottom* value indicates the cluster mass ($10^{14}{\rm M}_\odot$). The *top* (*bottom*) panel shows the E-mode (B-mode). 64
5.9 Mass map for the patch P3. The *black crosses* represent the positions of Planck clusters. The value on *top* of each cross is the cluster redshift, and the *bottom* value indicates the cluster mass ($10^{14}{\rm M}_\odot$). The *top* (*bottom*) panel shows the E-mode (B-mode). 65
5.10 Mass map for the patch P4. The *black crosses* represent the positions of Planck clusters. The value on *top* of each cross is the cluster redshift, and the *bottom* value indicates the cluster mass ($10^{14}{\rm M}_\odot$). The *top* (*bottom*) panel shows the E-mode (B-mode). 66

5.11 Mass maps stacked at the Planck cluster positions for P1, P2, P3 and P4 from *top* to *bottom*. Galaxies for the tangential shear stacks are selected in a radius of 5 Mpc around each cluster. This distance is computed at the cluster redshift. 67
5.12 Posterior for the 3×2-point analysis on MICE simulation. The contours represent 1-, 2- and 3-σ errors. *Dashed* lines show the input values of the simulation, $\Omega_m = 0.25$ and $\sigma_8 = 0.8$. 68
5.13 Redshift distribution of the LRG sample. 69
5.14 Comparison of the diagonal of the covariance from CosmoCov and the shot/shape noise measured on data. 70
5.15 Posterior for the 3×2-point analysis. The contours represent 1-, 2- and 3-σ errors. 71
5.16 Best fit of the correlation functions (*orange*). In *green* I show the theoretical value for Planck cosmology. 72
5.17 Posterior for the 3×2-point analysis in comparison with Planck and DES Y1. The contours represent 1-, 2- and 3-σ errors. 72
5.18 Posterior for the shear-shear only compared to the 3×2-point analysis. The contours represent 1-, 2- and 3-σ errors. 73
5.19 Shear-shear correlation function computed from the entire processed area of 1,695 deg^2. In *green* I show the theoretical value for the Planck cosmology. The (*orange*) curve represents the DES Y1 best-fit cosmology. 74

A.1 PSF leakage using the average galaxy shape in bins of PSF ellipticity (top two panels) and size (bottom panel) on the field P1 for the ngmix catalog. 84
A.2 PSF leakage using the average galaxy shape in bins of PSF ellipticity (top two panels) and size (bottom panel) on the field P1 for the Galsim catalog. 85
A.3 PSF leakage using the average galaxy shape in bins of PSF ellipticity (top two panels) and size (bottom panel) on the field P2 for the ngmix catalog. 86
A.4 PSF leakage using the average galaxy shape in bins of PSF ellipticity (top two panels) and size (bottom panel) on the field P2 for the Galsim catalog. 87
A.5 PSF leakage using the average galaxy shape in bins of PSF ellipticity (top two panels) and size (bottom panel) on the field P4 for the ngmix catalog. 88
A.6 PSF leakage using the average galaxy shape in bins of PSF ellipticity (top two panels) and size (bottom panel) on the field P4 for the Galsim catalog. 89
A.7 $\xi_{\pm}^{\rm sys}(\theta)$ measured for ngmix (*Top*) and Galsim (*Bottom*) on patch P1. . . . 90
A.8 Ratio of $\xi_{+}^{\rm sys}(\theta)/\xi_{+}^{\rm ss}(\theta)$ for ngmix (*Top*) and Galsim (*Bottom*) on patch P1. 91
A.9 Leakage α shown as a function of scale θ for ngmix (*Top*) and Galsim (*Bottom*) on patch P1. 92
A.10 $\xi_{\pm}^{\rm sys}(\theta)$ measured for ngmix (*Top*) and Galsim (*Bottom*) on patch P2. . . . 93
A.11 Ratio of $\xi_{+}^{\rm sys}(\theta)/\xi_{+}^{\rm ss}(\theta)$ for ngmix (*Top*) and Galsim (*Bottom*) on patch P2. 94
A.12 Leakage α shown as a function of scale θ for ngmix (*Top*) and Galsim (*Bottom*) on patch P2. 95
A.13 $\xi_{\pm}^{\rm sys}(\theta)$ measured for ngmix (*Top*) and Galsim (*Bottom*) on patch P4. . . . 96
A.14 Ratio of $\xi_{+}^{\rm sys}(\theta)/\xi_{+}^{\rm ss}(\theta)$ for ngmix (*Top*) and Galsim (*Bottom*) on patch P4. 97
A.15 Leakage α shown as a function of scale θ for ngmix (*Top*) and Galsim (*Bottom*) on patch P4. 98

B.1 Diagonal terms of the response martrix for the patch P1. The *solid line* for the galaxies and the *dashed line* for the stars. 100

B.2 Diagonal terms of the response martrix for the patch P2. The *solid line* for the galaxies and the *dashed line* for the stars. 100
B.3 Diagonal terms of the response martrix for the patch P4. The *solid line* for the galaxies and the *dashed line* for the stars. 101

List of Tables

1.1 Cosmological parameters for ΛCDM. The value are from Planck Collaboration XVI (Aghanim et al. 2020) 6

3.1 SExtraction parametrisation. 28
3.2 PSFEx parametrisation. 34

4.1 Additive bias for both components of the ellipticity with ngmix 46
4.2 Additive bias for both components of the ellipticity with Galsim. 47

5.1 Cluster selection bins. 60
5.2 Average mass of the clusters for each bin. 62
5.3 Number of clusters in each patch. 62
5.4 Priors used for the Markov Chain Monte Carlo (MCMC). 71

List of Acronyms

BFD	Bayesian Fourier Domain.
BOSS	Baryon Oscillation Spectroscopic Survey.
CCD	Charge-Coupled Device.
CCL	Core Cosmology Library.
CEFCA	Centro de Estudios de Física del Cosmo de Aragón.
CFHT	Canada France Hawaii Telescope.
CFHTLenS	Canada France Hawaii Telescope Lensing Survey.
CFHTLS	Canada-France Telescope Legacy Survey.
CFIS	Canada-France Imaging Survey.
CMB	Cosmic Microwave Background.
COSMOS	Cosmic Evolution Survey.
DES	Dark Energy Survey.
eBOSS	Extended Baryon Oscillation Spectroscopic Survey.
FLRW	Friedmann-Lemaître-Robertson-Walker.
FWHM	Full Width at Half Maximum.
GR	General Relativity.
GSC	Guide Star Catalog.
HSC	Hyper Suprime-Cam.
HST	Hubble Space Telescope.
JWST	James Webb Space Telescope.
KSB	Kaiser-Squires-Broadhurst.
LRG	Luminous Red Galaxy.
MCMC	Markov Chain Monte Carlo.
MPI	Message Passing Interface.
NFW	Navarro-Frenk-White.
Pan-STARRS	Panoramic Survey Telescope And Rapid Response System.

PSF	Point Spread Function.
PSFEx	PSF Extraction.
RA	Right Ascension.
RMS	Root Mean Square.
SDSS	Sloan Digital Sky Survey.
SMP	Symmetric Multi-Processing.
SNR	Signal to Noise Ratio.
SR	Special Relativity.
SZ	Sunyaev-Zeldovich.
UNIONS	Ultra-violet Near-Infrared Optical Northern Survey.
WCS	World Coordinate System.

Chapter 1

Introduction

It is at the beginning of the year 1610 that Galileo Galilei observed for the first time the satellites of Jupiter. From this observation he deduced the Earth was orbiting the Sun. This experience started what we call today modern science and, of course, modern astronomy. After Galileo's work, more people started to look at the sky using instruments that were perfected over the years. Today, we do not observe the sky by eye anymore, we rather take images of it. Even if the human eye is one of the best optical tools at our disposal, using a camera allows one to increase the exposure time and also to observe the sky at different wave-lenghts.

From a theoretical point-of-view, cosmology started when Albert Einstein described Gravitation as a deformation of space and time in his publication in 1915 (Einstein 1915). This theory was first confirmed during the solar eclipse of 1919 by Sir Arthur Eddington by observing the deflected light of stars located behind the sun.

By multiplying our observations and always looking deeper and deeper at the sky, Edwin Hubble discovered that galaxies were moving away from us. The deeper you look, the faster the galaxies seem to move. This escape is not due to the proper motion of galaxies, but it is space between them that is expanding. That was the first observation of the expansion of the Universe.

After this discovery, people started to think that if the Universe was expanding, it had to be smaller before. Then, they tried to reverse the process using General Relativity (GR) equations. One of the people who was working on this, Georges Lemaître, proposed a theory that explains the apparition of space and time. His work led to The Big Bang Theory. This theory predicts the evolution of the Universe and its content. The first image we have of the Universe, 380 000 years after the Big Bang, called the Cosmic Microwave Background (CMB), shows the distribution of matter. This signal was detected for the first time in 1964 by Arno Penzias and Robert Wilson.

Most recently, the Nobel committee rewarded in 2011 the discovery of the acceleration of the cosmic expansion. These results confirm the actual model used in cosmology: Λ-CDM. This model remains the best we have today, and can explain the evolution of the Universe and all its content from the Big Bang to the current time. This model can be fully described using only 7 parameters. To have constrains on those parameters we use different probes. The CMB being one of them, allows us to constrain the early Universe. The study of the distribution of galaxies, also referred to as clustering, constrains the evolution of structure. The clustering can be combined with weak lensing which is one of the few direct measurements of the total matter distribution.

My work has been mainly focused on the study of weak lensing. This science uses the same principle of GR as the one used to prove it. Indeed, the light follows a curved trajectory when propagating through space-time, deformed by the presence of matter. Measurements of this light distortions allow one to deduce the matter distribution between the source of

the light and the observer. The resulting observations consist of slightly sheared objects. From the observer's point of view, the main goal is to measure this shear.

The main difficulty of weak lensing when it comes to measuring the shapes of objects, is to precisely take care of all the other effects that could distort a galaxy. First, the light will go through an optical system, a telescope, and in the case of ground-based observation, through the atmosphere. We can encompass those effects in what we call the Point Spread Function (PSF). In addition to the PSF, the photons will be captured by a camera which will add noise to the final image. Finally, the effect we want to measure is particularly faint which can be overcome by using very large surveys.

In this introduction, I have sketched the importance of observations in astrophysics. I first give a short introduction to the current cosmological model used in the community. Then I will do the transition to data by showing how theoretical equations are derived to be applied to real data. Finally, I will introduce the basic method I am using in observational weak lensing to measure the signal.

1.1 Cosmology

1.1.1 General Relativity

Metric

The most basic principal underlying both Special and General Relativity is that time and space are reference-frame dependent. Time and space are linked through the metric. The metric allows one to describe how to compute geometrical distances in our Universe. For that we define an invariant distance, ds. If one maps the entire Universe by a grid, ds will be the side length of one cell (see figure 1.1 for an illustration). It can be defined as:

$$ds^2 = \sum_{\mu,\nu=0}^{3} g_{\mu\nu} dx^\mu dx^\nu, \tag{1.1}$$

where μ, ν are in the range [0,3] for a 4-dimensions space-time, 0 being time by convention, and $g_{\mu\nu}$ is a tensor that translates how the metric evolves. In Special Relativity (SR) the Universe is static, so it can define as:

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.2}$$

But in the case of GR the Universe is not necessarily static. For this case, I will consider a flat Universe in expansion with time. In that case the metric will depend on the scale factor, $a(t)$ (which will be described later), and $g_{\mu\nu}$ becomes:

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & a^2(t) & 0 & 0 \\ 0 & 0 & a^2(t) & 0 \\ 0 & 0 & 0 & a^2(t) \end{pmatrix}, \tag{1.3}$$

called the Friedmann-Lemaître-Robertson-Walker (FLRW) metric.

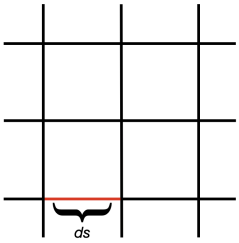


Figure 1.1: Representation of ds.

Geodesics

Once I have defined the metric that describe our space, one can see how objects move in this space. Let's take light as an example. In a Newtonian space, without any force, the trajectory of light beams are straight lines. In the context of relativity those straights lines are called geodesics. The equation of motion of the photons traveling along these geodesics can be written as:

$$\frac{d^2x^\mu}{d\lambda^2} = -\Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{d\lambda}\frac{dx^\beta}{d\lambda}, \tag{1.4}$$

with $\Gamma^\mu_{\alpha\beta}$, the Christoffel symbol, defined as:

$$\Gamma^\mu_{\alpha\beta} = \frac{g^{\mu\nu}}{2}\left[\frac{\partial g_{\alpha\nu}}{\partial x^\beta} + \frac{\partial g_{\beta\nu}}{\partial x^\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x^\nu}\right]. \tag{1.5}$$

One can see here that the equation of motion depends only on the metric used. It is important to note that objects, or light in the example, would be dependent of the space geometry. If space is curved, light will follow a curved trajectory even in the absence of applied forces.

Einstein equation

In the framework of GR, gravitation plays a key role. Indeed, gravitation can not be characterized as a force, in the sense of the other three fundamental forces, but more as a geometrical effect due to the presence of matter. Gravitation can be encoded in the metric. However, to describe the interaction of the Universe's content and the metric we use the Einstein equation:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi\mathcal{G}}{c^4}T_{\mu\nu}. \tag{1.6}$$

In this equation:

- $G_{\mu\nu}$ is the Einstein tensor, also written as: $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}\mathcal{R}$.
  Here we have $R_{\mu\nu}$, the Ricci tensor, $\mathcal{R}$, the Ricci scalar, and $g_{\mu\nu}$, the metric.
- $T_{\mu\nu}$ is the energy-momentum tensor.
  That describes the energy content of the Universe.
- $\mathcal{G}$ is Newton's gravitational constant.
- c is the speed of light in vacuum.
- Λ is the cosmological constant. This constant was set to 0 by Einstein because we first thought the Universe was static in space (no expansion) and in time (it has always

been there and will always be there). We will discuss more in detail this term in section 1.1.2.

One can see the Einstein equation as equivalence of $Metric \Leftrightarrow Energy$.

1.1.2 ΛCDM

the "standard model"

In the previous section, the most elementary brick of a cosmological model has been defined. Here, I will present the preferred one to date which describes best our observations (Aghanim et al. 2020). First things first, the name of this model already gives a lot of information:

- **CDM**: Cold Dark Matter. Dark Matter was introduced in 1933 by Fritz Zwicky after observations of the Coma Cluster. He measured the velocity of galaxies orbiting the cluster and found that the "luminous mass" was not enough to explain the observed speed. "Dark matter" was then introduced to refer to the missing mass. Among the different properties of dark matter, the two most important ones are that it is non-baryonic and does not interact with light. The term "Cold" refers to the fact that it is not relativistic (small speed compared to the speed of light). This kind of Dark Matter model which is the most used in the community today.
- **Λ**: The cosmological constant. This constant was first introduced by Albert Einstein when he developed the equations of GR. At this time it was thought that the Universe was static in time and space, but the GR equations gave the possibility to have a non-static Universe. To fix this problem Einstein introduced a constant, Λ. Later, in 1927 Alexander Friedmann and Georges Lemaitre proposed an expanding Universe which was confirmed in 1929 by observations from Edwin Hubble. Finally, in 1998 it was discovered based on type Ia supernovae that the expansion was accelerating. Λ was then reintroduced to Einstein's equations, but to explain a non-static Universe this time.

Now we can make the inventory of the Universe's content (Aghanim et al. 2020):

- **Matter (≈ 5%):** This includes all baryonic matter. From the largest object that one can find to the smallest dust particles, which compose the interstellar medium.
- **Dark Matter (≈ 20%):** The non-baryonic component of the Universe.
- **Dark Energy (≈ 75%):** This last component which composes the larger part of the content of the Universe is also the one we least understand. Dark energy could be Λ, which fits the observations very well, or a more generic substance. Today there are a lot of ongoing experiments which aim to discover the nature of dark energy.

Before going further, we have to mention the Cosmological Principle. This principle lays on two observed properties: The Universe is homogeneous and isotropic on large scales. We will develop below the implication of this principal on the GR equations.

The Cosmological Principle

Using the Cosmological Principle, we can simplify Einstein's equations. First, we can rewrite equation 1.1 and define the FLRW metric named after their authors:

$$ds^2 = -c^2dt^2 + a(t)^2\left(d\chi^2 + f_K(\chi)^2 d\Omega^2\right), \tag{1.7}$$

with χ being the comoving distance, $a(t)$ the scale factor and $f_K(\chi)$ the comoving angular distance, which describes the curvature of space, defined as:

$$f_K(\chi) = \begin{cases} K^{-1/2}\sin(K^{-1/2}\chi) & \text{for } K > 0 \text{ (sperical);} \\ \chi & \text{for } K = 0 \text{ (flat);} \\ |K|^{-1/2}\sinh(|K|^{-1/2}\chi) & \text{for } K < 0 \text{ (hyperbolic).} \end{cases} \tag{1.8}$$

The Cosmological Principle implies that the content of the Universe can be described as a perfect fluid. Then the energy-momentum tensor becomes:

$$T^{\mu\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}, \tag{1.9}$$

and the Einstein equations are:

$$\frac{\dot{a}}{a} = \frac{8\pi G}{3}\rho - \frac{Kc^2}{a^2}, \tag{1.10}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right), \tag{1.11}$$

which are called the Friedmann equations. Here, ρ is the energy density, p the pressure of the fluid, K describes the curvature, G the gravitational constant, and c the speed of light in vacuum.

From equation 1.10 one can define a critical density $\rho_{\text{crit}}(a)$ which vanished the curvature as:

$$\rho_{\text{crit}}(a) = \frac{3H^2(a)}{8\pi G}. \tag{1.12}$$

Using this definition we can link to density parameters of the model with:

$$\Omega_x = \frac{\rho_x(a)}{\rho_{\text{crit}}(a)}, \tag{1.13}$$

with $x \in \{m, \Lambda, c, b, k, rad\}$.

Finally, one can link together the density parameters to define $H(a)$ as:

$$H(a) \equiv \frac{\dot{a}}{a} = H_0\sqrt{\Omega_m a^{-3} + \Omega_{rad}a^{-4} + \Omega_\Lambda}, \tag{1.14}$$

for a flat λCDM Universe where $\Omega_k = 0$ and $\omega = -1$. Ω_m is the matter density parameter and Ω_Λ is the density parameter of dark energy. Here H_0 is the Hubble constant which is usually defined using its reduced form h:

$$H_0 = 100h \text{ km.s}^{-1}.\text{Mpc}^{-1}. \tag{1.15}$$

Cosmological parameters		
Parameter	Value	Short description
Ω_m	0.3153	Density of Matter
Ω_Λ	0.6847	Density of Dark Energy
Ω_b	0.0493	Density of baryon
h	0.6736	Hubble parameter
σ_8	0.8111	Growth of structure
n_s	0.9649	Scalar spectral index
τ	0.0544	Reionisation optical depth

Table 1.1: Cosmological parameters for ΛCDM. The value are from Planck Collaboration XVI (Aghanim et al. 2020)

To fully describe the model we need to introduce other parameters. Two of them are used to define the primordial power spectrum which describes the initial matter fluctuation in the Universe. The primordial power spectrum is given as:

$$P(k) = A_s \left(\frac{k}{k_p}\right)^{n_s} \tag{1.16}$$

with A_s the amplitude of the power spectrum, n_s the scalar spectral index and k_p a pivot scale. It is more common to normalize the power spectrum using σ_8, which represents the Root Mean Square (RMS) of the density contrast at a scale of 8 Mpc.h^{-1}.

The last parameter we need is τ, the reionisation optical depth. It can be defined as:

$$\tau(z_{\text{ri}}) = \sigma_T \int_{t(z_{\text{ri}})}^{t_0} a(t) n_e(t) dt \tag{1.17}$$

with σ_T the Thomson scattering amplitude, n_e the number density of free electrons and z_{ri} the reionisation redshift ($z_{\text{ri}} \geq 6$). This phenomenon is not fully understand today due to the lack of data at high redshift. The future James Webb Space Telescope (JWST) will improve our understanding of this period.

We can then describe the ΛCDM model using only 7 parameters presented in table 1.1. We can add to those parameters, w, describing the equation of state of dark energy. For this model we have $w = -1$ but there are more complex models which use a parametric equation of the form:

$$w(a) = w_0 + (1 + a) w_a \tag{1.18}$$

which describe the evolution through time. Such models will not be discussed in this work.

A word on redshift

Redshifts are, in most cases, used to refer to a distance. But since the speed of light is constant, when we observe an object far away from us we also see it in the past. So redshift can also be used to define specific epoch. By convention, today, we are at redshift $z = 0$. We can write the redshift as function of the scale factor $a(t)$:

$$z = \frac{1 - a(t)}{a(t)}. \tag{1.19}$$

Another way to talk about distances in the Universe is to use the comoving distance χ, which can be expressed as a function of the scale factor:

$$\chi(a) = \int_a^1 \frac{c}{a' H(a')} da', \tag{1.20}$$

or as a function of redshift:

$$\chi(z) = \int_0^z \frac{c}{H(z')} dz'. \tag{1.21}$$

The problem to use the comoving distance is that you have to assume a cosmology due to the term $H(z)$ and fix the parameters. This is not the case with redshift, which is defined independently of either the cosmological model or its parameters.

1.1.3 Gravitational lensing

In the previous section I presented a general non-exhaustive description of the ΛCDM model. Here, we will give more details on the matter distribution in the Universe and how the light propagates through it.

Basics of gravitationel lensing

This physical phenomenon of gravitational lensing takes its name from what we can see with a simple optical system composed by a glass lens. When light passes through it, the trajectories of the beams are distorted by the shape of the glass. Gravitational lensing is the exact same thing except that the lens does not have a physical reality. Indeed, as we mentioned in section 1.1.1, light follows the curvature of space which can be affected by the presence of matter. This distortion of space-time will act as a "lens" and will create images as we can see in figure 1.2, this is what we call strong lensing. To see this phenomenon, we have to observe galaxies in the background for which the light will go through a high-density region, which can be a massive galaxy or a galaxy cluster for example. What makes gravitational lensing interesting for science is that it is one of very few methods that have access to the total mass of an object, including the mass we can not see (*i.e.* dark matter). Finding the configuration I described above is however extremely rare. Today we know only a few thousand strong-lensing systems (this number will significantly increase with the upcoming stage IV surveys). In the case of weak gravitational lensing we do not see a large distortion, but we can measure an alignment of the background galaxies around high-density regions. This is what we will describe in the following section.

Weak Lensing formalism

As shown in section 1.1.2, one bases their development on the Cosmological Principle, implying, among other things, that the Universe is homogeneous on large scales. Here I will have to introduce inhomogeneity on a local scale to describe the weak lensing formalism. Those inhomogeneities will be described as fluctuations of the density field δ around an average value $\bar{\rho}$ as:

$$\delta = \frac{\rho - \bar{\rho}}{\bar{\rho}}. \tag{1.22}$$

One then use the Poisson equation to describe the evolution of such fluctuations:

$$\nabla^2 \Phi = 4\pi G a^2 \bar{\rho} \delta, \tag{1.23}$$

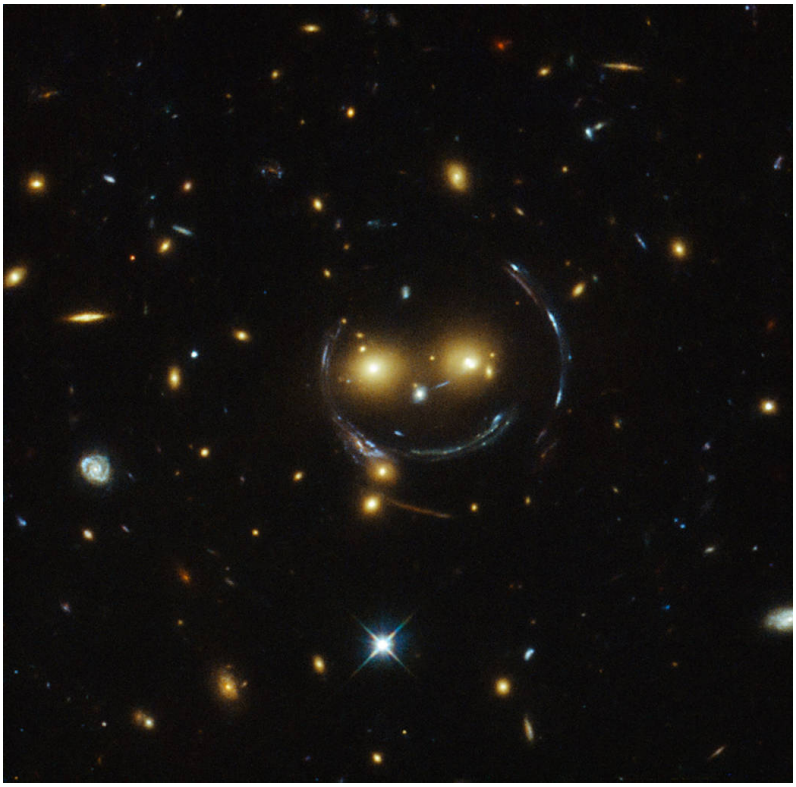


Figure 1.2: Strong lensing system observed by Hubble Space Telescope (HST).
Credit: NASA/ESA

with Φ the Newtonian gravitational potential and G Newton's constant.

Now, I will introduce the Newtonian gravitational potential in the FLRW metric (eq. 1.7) to get:

$$ds^2 = -\left(1+\frac{2\Phi}{c^2}\right)c^2dt^2 + \left(1-\frac{2\Phi}{c^2}\right)a(t)^2\left(d\chi^2 + f_K(\chi)^2 d\Omega^2\right), \tag{1.24}$$

which assumes weak gravitational fields ($\Phi \ll c^2$). In my study such condition will always be fulfilled since I am not looking at compact objects. Photons travel following null geodesics. One can derive the travel time from this newly defined metric:

$$t \approx \frac{1}{c}\int\left(1-\frac{2\Phi}{c^2}\right). \tag{1.25}$$

Applying Fermat's principle of minimal travel time to this equation will give us the Euler-Lagrange equations. By integrating these equations along the light path, one finds the deflection angle α which represents the angle between the observed and the emitted light paths:

$$\vec{\alpha} = -\frac{2}{c^2}\int \vec{\nabla}_\perp \Phi dr, \tag{1.26}$$

$\nabla_\perp$ represents the gradient of the Newtonian potential taken perpendicularly to the light path.

Under the assumption of a homogeneous Universe, one can write the comoving separation between two light rays as:

$$\vec{x}_0(\chi) = f_K(\chi)\vec{\theta}, \tag{1.27}$$

with $\vec{\theta}$ being the angle under which the separation is seen. To this, one adds the deflection due to presence of matter along the line of sight and obtains:

$$\vec{x}(\chi) = f_K(\chi)\vec{\theta} - \frac{2}{c^2}\int_0^{\chi} f_K(\chi - \chi')\left[\vec{\nabla}_\perp\Phi(x(\chi'),\chi') - \vec{\nabla}_\perp\Phi(0,\chi')\right]d\chi'. \tag{1.28}$$

Using equation 1.27 the angle $\vec{\beta}$ can be defined under which the separation vector $\vec{x}$ would have been seen without the lens as:

$$\vec{\beta} = \frac{\vec{x}}{f_K(\chi)}. \tag{1.29}$$

Finally, one defines the lensing equation as the difference between the separation angle with and without the lens according to the deflection angle as:

$$\vec{\beta} = \vec{\theta} - \vec{\alpha}, \tag{1.30}$$

where:

$$\vec{\alpha} = \frac{2}{c^2}\int_0^{\chi} \frac{f_K(\chi - \chi')}{f_K(\chi)}\left[\vec{\nabla}_\perp\Phi(x(\chi'),\chi') - \vec{\nabla}_\perp\Phi(0,\chi')\right]d\chi'. \tag{1.31}$$

The reader can report to figure 1.3 for a sketch representation of the different quantities introduced in the above equations.

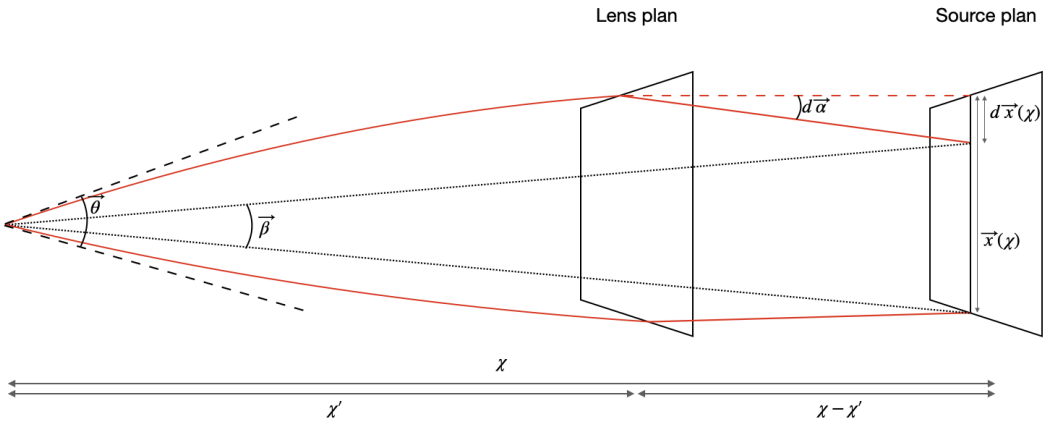


Figure 1.3: This figure shows the propagation of two light beams emitted by a galaxy in the source plan through a cluster located in the lens plane.

1.2 Lensing observables

In the previous section, I presented the weak lensing formalism. Now I will show how it can link the theory to actual observations. Since galaxies are distorted due to the lensing effect, the goal of observations is to measure these distortions. To achieve that, we measure

the shapes of galaxies. Of course, galaxies can have very complex shapes, and I will show that there are effects other than the one we are interested in that can alter the shapes of galaxies.

1.2.1 Basic principle

We start by simplifying equation 1.31 by using the Born approximation in combination with the introduction of the Jacobian $\vec{A}$ defined as:

$$\vec{A} = \frac{\partial \vec{\beta}}{\partial \vec{\theta}}. \tag{1.32}$$

Going further one has:

$$\begin{aligned} A_{ij} &= \frac{\partial \beta_i}{\partial \theta_j} \\ &= \delta_{ij} - \frac{\partial \alpha_i}{\partial \theta_j} \\ &= \delta_{ij} - \frac{2}{c^2}\int_0^\chi \frac{f_K(\chi-\chi')f_K(\chi')}{f_K(\chi)}\frac{\partial^2}{\partial x_i \partial x_j}\Phi(f_K(\chi')\vec{\theta},\chi')d\chi'. \end{aligned} \tag{1.33}$$

The 2D lensing potential ψ can now be defined as:

$$\psi(\vec{\theta},\chi) = \frac{2}{c^2}\int_0^\chi \frac{f_K(\chi-\chi')}{f_K(\chi)f_K(\chi')}\Phi(f_K(\chi')\vec{\theta},\chi')d\chi', \tag{1.34}$$

which finally gives:

$$A_{ij} = \delta_{ij} - \frac{\partial^2}{\partial\theta_i\partial\theta_j}\psi = \delta_{ij} - \partial_i\partial_j\psi. \tag{1.35}$$

The shear γ_i for $i \in 1,2$ and the convergence κ can be defined as:

$$\gamma_1 = \frac{1}{2}(\partial_1^2 - \partial_2^2)\psi, \tag{1.36}$$

$$\gamma_2 = \partial_1\partial_2\psi, \tag{1.37}$$

$$\kappa = \frac{1}{2}(\partial_1^2 + \partial_2^2)\psi, \tag{1.38}$$

which relates to A such that:

$$A = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}. \tag{1.39}$$

The convergence affects the size of galaxies, while the shear affects the shape. As mentioned in the introduction to this section, we measure the shapes of objects but we do not have a direct access to the shear. What we are measuring is the reduced shear g which is defined as:

$$g_i = \frac{\gamma_i}{1-\kappa} \quad \text{for } i \in 1,2. \tag{1.40}$$

We now have all the required framework for weak lensing studies. In the sections 1.2.4 and 1.2.5 I will detail the two main techniques used to measure the shapes of objects.

1.2.2 Tangential alignment and correlation functions

Tangential alignment

When the light of a galaxy is affected by a gravitational potential, the galaxy will appear tangentially aligned around the potential. To have a better visualization of this phenomenon, two quantities can be defined, γ_t and γ_x as:

$$\gamma_t = -\mathfrak{Re}\left(\gamma e^{-2i\phi}\right); \tag{1.41}$$

$$\gamma_x = -\mathfrak{Im}\left(\gamma e^{-2i\phi}\right), \tag{1.42}$$

where $\gamma = \gamma_1 + i\gamma_2$, and ϕ designates the orientation angle of the direction vector. The negative sign is there by convention to have a tangential alignment around overdensities and radial alignment for underdense regions.

Correlation functions

From γ_t and γ_x we can write two non-zero 2-point correlation functions:

$$\xi_+(\theta) = \langle\gamma_t\gamma_t\rangle(\theta) + \langle\gamma_x\gamma_x\rangle(\theta); \tag{1.43}$$

$$\xi_-(\theta) = \langle\gamma_t\gamma_t\rangle(\theta) - \langle\gamma_x\gamma_x\rangle(\theta). \tag{1.44}$$

These equations can be linked to cosmology through the convergence power spectrum P_κ:

$$P_\kappa(l) = \frac{9H_0^4\Omega_m^2}{4c^4}\int_0^{\chi_{lim}} \frac{g^2(\chi)}{a^2(\chi)} P_\delta\left(k = \frac{l}{f_K(\chi)}, \chi\right) d\chi, \tag{1.45}$$

where $P_\delta(k, \chi)$ is the matter power spectrum.
Using the Hankel transform we get the 2-point correlation functions ():

$$\xi_+(\theta) = \frac{1}{2\pi}\int_0^{+\infty} l J_0(l\theta) P_\kappa(l) dl; \tag{1.46}$$

$$\xi_-(\theta) = \frac{1}{2\pi}\int_0^{+\infty} l J_4(l\theta) P_\kappa(l) dl, \tag{1.47}$$

with J_n the Bessel function of the n-th order.

1.2.3 PSF introduction

As shown, the shapes of galaxies are affected by gravitational lensing. However, gravitational lensing is not the only effect modifying the observed shapes. When images of the sky are taken for a photometric survey, one uses a complex optical system, the main piece being the telescope. Most of the time this system will also have additional mirrors and lenses. All of those optical elements have their own optical response, and this response is not perfect. All the elements the light will go through before reaching the camera will alter the shape of the objects in a non-negligible way. Finally, the telescope can be in space or on the ground. In the case of ground-based surveys, the atmosphere is included in the optical system.

Since weak lensing requires high-precision measurements, one can not ignore these effects. In order to measure the response of the system one uses a similar technique as the

one used in electronics when facing a "black box". One sends an impulse to the system and measures the PSF. We proceed similarly with our optical system. There are two main techniques to estimate the PSF:

- Space-based surveys: In this case, the telescope comprises the entire system. What people can do, is to study in a laboratory how the telescope behaves, and build a model before the telescope is sent to space. This model can be adjusted using observed stars.
- Ground-based surveys: Here, the atmosphere is part of the optical system, which adds randomness and makes the elaboration of a model almost impossible in a laboratory. Fortunately, the stars can be used to measure the PSF. Indeed, stars can not be resolved by most telescopes [1] which makes then the impulsion we are looking for. One can use them to derive a model for each observation and correct for it.

Later, I will explain in more detail how the PSF model is constructed in the case of ground-based surveys. In the next section I will present two generically used methods to measure the shapes of objects that can take care of the PSF.

1.2.4 Quadrupole moments

Here I present one familly of methods used to estimate the shape of objects. This is based on the quadrupole moments of the surface brightness. This method is also used to give a definition when one talks about the ellipticity of galaxies. Indeed, for elliptical galaxies can can do a direct analogy with ellipses and the use of ellipticity might be direct. When it comes to more complex objects like spiral barred galaxies or irregular galaxies the analogy with an ellipse is more complex. The quadrupole moments allow us to define an ellipticity which will always exist.

General method

For a light profile $I(\vec{X})$ with $\vec{X} = \begin{pmatrix} x \\ y \end{pmatrix}$, the quadrupole moments can be defined as:

$$M_{ij} = \int_{\vec{X}\in\mathbb{R}} (x - x_0)^i (y - y_0)^j I(\vec{X}) d\vec{X}, \tag{1.48}$$

with $\vec{X}_0$ being the center of the light profile which can be defined using the first order moments:

$$x_0 = \frac{M_{1,0}}{M_{0,0}}, \tag{1.49}$$

$$y_0 = \frac{M_{0,1}}{M_{0,0}}. \tag{1.50}$$

On real data, the presence of noise makes this measurement very unstable. It is preferable to use weighted moments defined by:

$$Q_{ij} = \frac{\int_{\vec{X}\in\mathbb{R}} (x - x_0)^i (y - y_0)^j I(\vec{X}) H(\vec{X}) d\vec{X}}{\int_{\vec{X}\in\mathbb{R}} I(\vec{X}) H(\vec{X}) d\vec{X}}, \tag{1.51}$$

[1] This statement stands for the most generic cases. Some observations ha been done of resolved stars (other then the Sun) thanks to the progress of adaptive optics (Gilliland and Dupree 1996).

where $H(\vec{X})$ represents the weighting function. In most cases this function is a 2D-Gaussian. We can now define the complex ellipticity e_i using the second-order moments[2]:

$$e = \frac{Q_{11} - Q_{22} + 2iQ_{12}}{Q_{11} + Q_{22}}, \tag{1.52}$$

and the size:

$$R^2 = Q_{11} + Q_{22} \tag{1.53}$$

The ellipticity defined here does not directly provide the reduced shear g. In the quantity measured here there is a contribution from the intrinsic ellipticity of the galaxy. To get the value of the shear one has to average measurements on several galaxies:

$$g \approx \frac{< e >}{2} \tag{1.54}$$

Adaptive moments

When weighted moments are computed, one can fix the weight function to an isotropic Gaussian with a given size, based on the size of the object for example. But one can also fit the weight function iteratively as the moments are being estimated. In that case one can chose to have an elliptical Gaussian function. The aim of this process is to maximize the Signal to Noise Ratio (SNR) of the object. The convergence criteria of this method can be based on the variation of the Gaussian parameters. One can chose to only use the center positions, or the center positions and the size of the Gaussian. Once we get to the convergence, a last measure of the weighted moments is performed from which the ellipticity is deduced. This technique is similar to model fitting, except that the final result is not the parameters of the model.

The KSB method

As seen previously, to get a correct estimation of the shear one haw to correct for PSF effects. Over the years people have developed several methods based on quadrupole moments to account for PSF effects. One typically does not perform a proper deconvolution of the PSF but rather apply corrections on the moments due to the PSF. The most well-known method is called Kaiser-Squires-Broadhurst (KSB) after the name of their inventors (Kaiser, Squires, and Broadhurst 1995). But there are other implementations such as Bernstein and Jarvis 2002, or re-Gaussianization (Hirata and Seljak 2003). Other works also perform deconvolution in the moments space with a method called DEIMOS (Melchior et al. 2011).

The main advantage of these methods is that they are very fast since all the computation is analytical. The main issue is that one cannot use them with a multi-epoch approach. Today however, a lot of surveys attempt to take full advantage of multi-epoch data without creating stacked images, to avoid a PSF that is altered by the stacking procedure. I will later present a method for stacked images, for which we have implemented a specific stacking method for the PSF.

[2] This quantity is usually called χ but to avoid any confusion with the comoving distance definition we will call it e. Not to be confused with the other existing convention ϵ.

1.2.5 Model fitting

The other family of methods to estimate the ellipticity of a galaxy is to approximate the light profile of the object with an analytic function. Several models can be found in the literature, the most simple one being a 2D-Gaussian. To have a better description one can use an Exponential or a Sersic model (Sérsic 1963). One can also use two distinct models to describe the disk and the bulge of galaxies. The point of an analytical model is to have the ellipticity as one of the parameters of the model to get an approximation of the object shape.

Methodology

Model fitting methods are based on the minimization of a cost function. The cost function can be either in real or Fourier space, and estimates how good the model is to describe the data. Here I present the method I will use later with the ngmix software package[3]. The method used for the minimization is least squares. In that case the loss function is defined as:

$$L(p) = \sum_x \frac{(\mathrm{I}(x) - \mathrm{M}(x;p))^2}{\sigma^2(x)}, \tag{1.55}$$

where I is the image vector as a function of the pixel coordinates x, M is the model for a set of parameters p, and σ^2 describes the noise variance on the real image for each pixel. I then run a minimization algorithm to find the best set of parameters for a given model.

This method can account for PSF corrections. One can define a model M′ as:

$$M'(x;p) = M(x;p) * \mathrm{H}(x), \tag{1.56}$$

where H represents the PSF model and $*$ denotes a convolution operation. In our case the main interest of this method is the ability to fit several representations of the same galaxy with one model. This will allow us to have better constraints on the parameters with an increased signal to noise ratio. In that case we can rewrite the loss function as:

$$L(p) = \sum_{n=1}^{n_e} \sum_x \frac{(\mathrm{I}_n(x) - [\mathrm{M}(x;p) * \mathrm{H}_n(x)])^2}{\sigma_n^2(x)}, \tag{1.57}$$

where n represents one epoch of the image I and the PSF model H over the total number of epochs n_e. It is important to note that the model and parameters are the same for all epochs. We can imagine more complex situations where for example one has several observations of the object but in different photometric bands. In that case one could have some parameters common to each band, and other parameters varying across the different filters.

The minimization can include priors for the different parameters. This is not always appreciated since some prior can influence the results. In the present case, I use uninformative priors for most of the parameters. They take the form of a uniform distribution within a given range chosen to be wide enough to not constrain the fit. The only informative prior is on the ellipticity, where one assumes a Gaussian distribution from Bernstein and Armstrong 2014a. The purpose of these priors is only to regularize the fit.

[3] https://github.com/esheldon/ngmix

1.2.6 Shear bias

Independently of the method one uses, the measured shear will be biased[4]. The different origins of the multiplicative bias are going to be detailed below. The problem of this bias is that it is of the same order (or larger) than the precision required on the shear signal. The shear bias is usually modeled as:

$$e^{\text{obs}} = (1 + m)\gamma + c + e^{\text{int}} \tag{1.58}$$

where e^{obs} is the observed ellipticity, e^{int} the intrinsic ellipticity, m the shear bias and c the additive bias. I will discuss again the bias in the section dedicated to the calibration in 3.4. It is often assumed that only the shear altered the shape of galaxies. However, one has observed that the clustering can also have an impact on the orientation of galaxies. Indeed, galaxies in filaments will tend to have the same orientation. This effect is called intrinsic alignment and its contribution is counted as a systematic effect which will contribute to the bias the shear measurements.

Model bias

The model bias comes from the approximation used to describe the object from which one tries to extract the ellipticity. This become clear in the case of model fitting where the model used is very simplistic and may not take into account all features on the surface brightness of a galaxy. It is important to notice that this also impacts methods which are moments based since there is the intrinsic implication of the object being an ellipsoid. The window function used will also have a contribution in terms of model bias. This bias can be captured using simulations with realistic images of galaxies.

Noise bias

This bias is due to the presence of noise on the images. Since the ellipticity varies non-linearly with the pixel values, the noise will bias the estimated shear. The noise will also make it hard to define the right morphology of the object. This bias can also be captured with image simulations.

Detection bias

The object detection can induce a bias on the shear. This means that some objects will be preferably detected depending of their shear. This has a strong impact on blended objects or close-by objects. This bias can also be related to the PSF. Indeed, depending on the orientation of the object with respect to the orientation of the PSF it might be easier (or harder) to detect it. This will also create detection bias on the shear. The calibration of this bias can also be handled with simulations but it requires very complex ones. The simulation has to reflect a realistic spatial distribution of galaxies, meaning the implementation of clustering effects. A new approach for the calibration of this bias has been proposed by Sheldon et al. 2020.

[4]We can make an exception for the Bayesian Fourier Domain (BFD) method which claim to give an unbiased measurement of the shear (Bernstein and Armstrong 2014b).

Selection bias

The selection bias is created by the different cuts that are applied on the detected sample to obtain a confident sample of galaxies. The measured properties on which one applies the cuts can be correlated to the shear and then induce a bias. For example properties like the SNR or the size are very likely to create a selection bias. This bias is also very complicated to capture with simulations for the same reason as the detection bias. In this work I will use the metacalibration method (Huff and Mandelbaum 2017) which provides the tools to make the calibration directly with the real data. This will be detailed in the section 3.4.

Additive bias

The additive bias has to be handled separately since it is directly correlated to the shear. This bias will be mainly due to a bad correction of the PSF effects in the shape of galaxies. This bias depends only on the method used for the shape measurement (a perfectly modeled PSF is assumed). This bias is the sign of non-optimal shape measurement algorithm consequently there is not a way to calibrate for it. However, it is possible to estimate the impact of it on the data. This will be presented in the section 4.3.

Chapter 2

Data and Simulations

2.1 Weak-Lensing surveys

Weak-lensing analyses are mostly based on optical imaging surveys, but it is also possible to conduct weak lensing studies using CMB data or other radio data. The first weak-lensing analyses have been carried out in the optical on very small regions of the sky, and centered on galaxy clusters. Most of those were done using the HST, which provides high resolution data. These studies provided very accurate maps of the mass distribution of clusters. Later, the focus shifted to mapping the matter distribution on larger regions in the sky. Surveys like the Canada France Hawaii Telescope Lensing Survey (CFHTLenS) from the Canada-France Telescope Legacy Survey (CFHTLS) provided around 200 Deg^2 of images, resulting in shape measurements of millions of galaxies. Today, we are entering the *golden era* of lensing with surveys like Dark Energy Survey (DES), Canada-France Imaging Survey (CFIS), or the upcoming Euclid survey and the Vera Rubin Observatory's Legacy Survey of Space and Time. Those surveys will provide around twenty thousands of square degrees of data with over one billion objects.

For this study, I used CFIS, a component of the Ultra-violet Near-Infrared Optical Northern Survey (UNIONS) collaboration which aims at collecting photometric data for Euclid in particular. In this chapter I will present the data used and the pre-processing required for the analysis.

2.1.1 The CFIS survey

CFIS started in 2017 and has reached a coverage of 2500 Deg^2 of the 5000 Deg^2 planned for the final coverage (completion before 2025). CFIS is a photometric survey taking deep images in the r- (640 nm) and u-bands (355 nm). It takes advantage of the excellent sky quality of Mauna Kea, with an average seeing of 0.65 arcsec in the r-band. It is a perfect data set for weak lensing studies. In 2018, the UNIONS collaboration has been created to gather in a single scientific group the various multi-band surveys covering the planned Euclid footprint in the Northern hemisphere. CFIS will provide the r- and u-band while the Panoramic Survey Telescope And Rapid Response System (Pan-STARRS) will observe the i-band, and Subaru Hyper Suprime-Cam (HSC) the z-band, with the aim of seeing the Centro de Estudios de Física del Cosmo de Aragón (CEFCA) soon join UNIONS through their Euclid effort on the g-band. A key element in the Northern hemisphere is the availability of the previous studies from the Sloan Digital Sky Survey (SDSS), the Baryon Oscillation Spectroscopic Survey (BOSS), and Extended Baryon Oscillation Spectroscopic Survey (eBOSS) to conduct science today. Thus, one has all the tools available to make a cosmological study from the combination of probes like clustering and cosmic shear for example.

The survey strategy

The objectives of the CFIS cosmic shear effort were both the photometric quality and cosmological inference. These types of analyses require a clear and dark sky. This means, from the ground, taking images during new moons. But there are other sources of light in the sky: a first offender is the light coming from the stars of the Milky Way. To avoid this, the observations are done away from the galactic plane (figure 2.1). Related to being within a galaxy, the presence of dust in the Milky Way creates absorption which needs to be avoided as well (figure 2.2). Another source is the zodiacal light coming from the Solar System, a concern only for space missions really since it is fairly faint compared to the night sky brightness even on the best sites, such as Maunakea. This light follows the ecliptic plane and is due to light scattering on small particles which reflect the light from the Sun (figure 2.3). Figure 2.4 shows the goal footprint of UNIONS, driven by Euclid, which takes into account all the effects mentioned before. Finally, there is also a few extremely bright stars whose halos from optical reflections would make the images unusable, and thus have to be avoided.

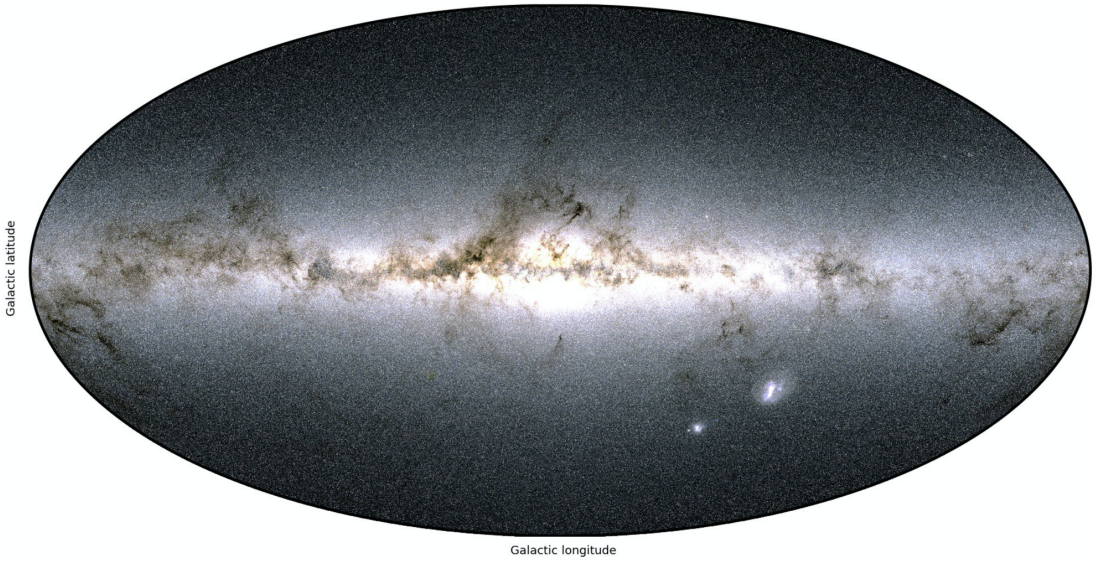


Figure 2.1: Map showing the light distribution of the Milky Way.
Credit: Euclid consortium

When observations are taken, it is very important to find the best compromise between quality, depth, and coverage for the allocated time. I will take the example of CFIS to illustrate this process. To increase the total covered area, CFIS has a large dither between each of the three exposures (of ∼200 seconds each) necessary to reach the depth goal (SNR=10 for 24.1 mag on the r-band for extended objects). The main drawback of this strategy is to bring more variability into the stacked images. This has been overcome by adopting a multi-epoch processing which will be discussed later. CFIS benefits from the observatory location: thanks to the high altitude of Mauna Kea, around 4200 m above the sea level, the atmosphere is thinner which means less turbulence. With perfect sky conditions, the seeing can be as low as 0.4 arcsec but is on average around 0.6 arcesec in the r-band. Finally, one of the ingredients for a photometric survey and weak lensing study is depth homogeneity. For CFIS the exposure time varies according to observing conditions (sky brightness, image quality, sensitivity of the system telescope+camera) to

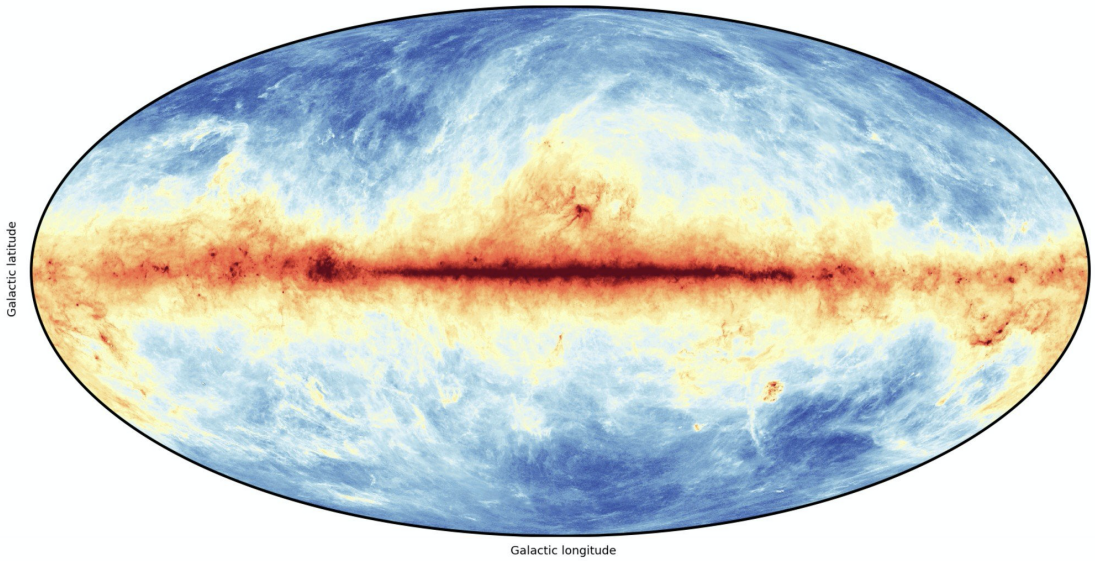


Figure 2.2: Map showing the dust distribution mainly coming from the Milky Way.
Credit: Euclid consortium

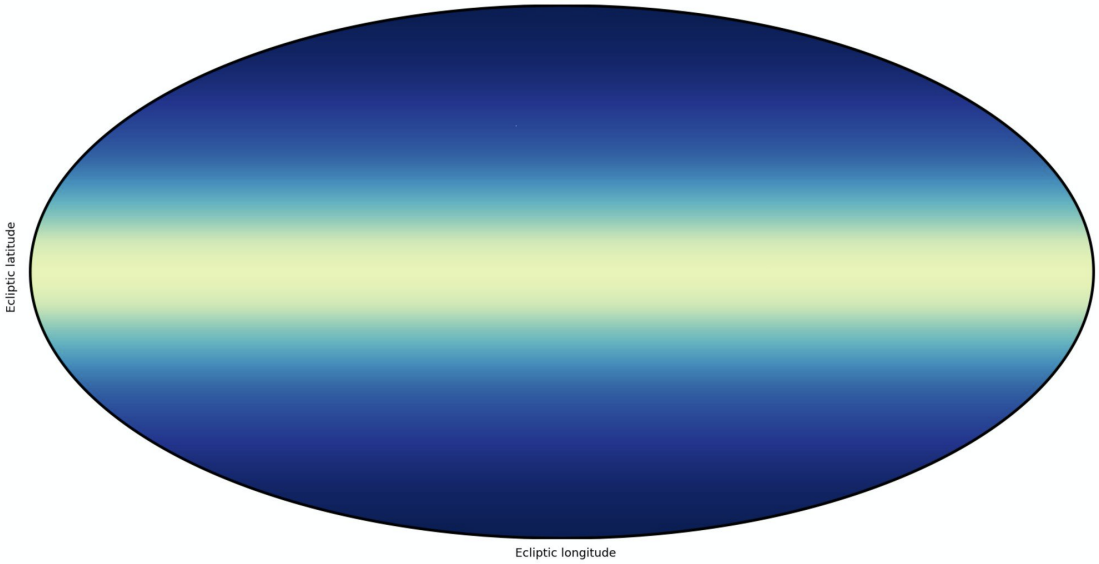


Figure 2.3: Map showing the zodiacal light on the ecliptic plane.
Credit: Euclid consortium

achieve a constant magnitude depth across the survey footprint (Cuillandre, Mahoney, and Withington 2014).

Data pre-processing

The pre-processing of data is a key step in the processing chain. This consists in positioning the images on the sky (astrometry) and calibrating the flux (photometry). For CFIS, the Canada France Hawaii Telescope (CFHT) images are calibrated using the pipeline

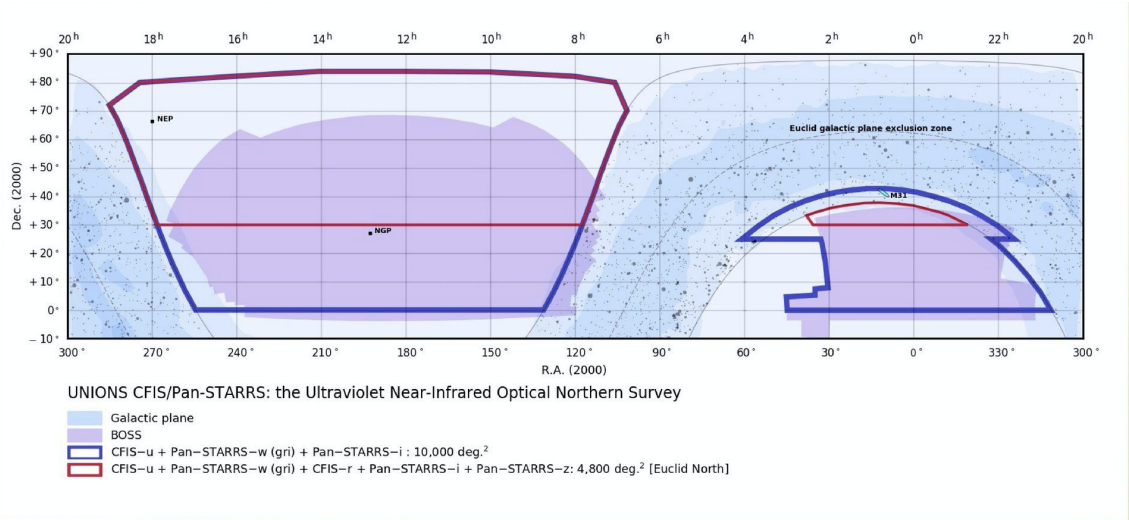


Figure 2.4: UNIONS footprint.
Credit: UNIONS collaboration

MegaPipe (Gwyn 2008). With the Gaia observations (Brown et al. 2018) of 1.7 billion stars we can achieve an astrometric calibration within 20 milli-arcsec as shown in figure 2.5. To calibrate the flux, we can rely on the observations from the Pan-STARRS PS1 survey (Chambers et al. 2016) to get a photometric solution as good as 1 milli-magnitude in the r-band (figure 2.6) internally (camera field-of-view) and 4 milli-magnitude in absolute with respect to an all-sky reference. Both of those steps are important for shape measurement since the gravitational lensing signal is extremely sensitive: achieving a calibration at such levels reduces systematic effects.

Another part of the processing is the creation of stacked images. To improve the signal to noise ratio, single exposures are combined together. There are two ways to handle this process. The "median-like" method will reduce the signal of time-dependent outliers like satellite or asteroid tracks, and cosmic rays. The main problem with such technique is that it alters the PSF significantly, and as I have shown previously, a stable and smooth PSF is very important for weak lensing. The second method is "average-like": this method conserves better the PSF but is less efficient at removing outliers although these can be reasonably easily flagged and rejected when stacking. In CFIS we have adopted the second technique: the stacks are created using a weighted average so the stacks can be used for weak lensing. The stacks are created with the SWARP software package[1] (Bertin et al. 2002).

2.2 Simulations

To develop a new technique one needs a controlled environment. This is only possible by using simulations. Simulations are an important part of weak lensing surveys. They can intervene at different steps of the development. They can be used for testing only, or to calibrate results. The calibration requires very precise simulations which reflect the data perfectly. Different effects which influence measurements have to be taken into account for the calibration. For testing, one can use simpler simulations that focus on one part of the processing pipeline. In my case, the simulations have been used to test the implementation

[1] https://github.com/astromatic/swarp

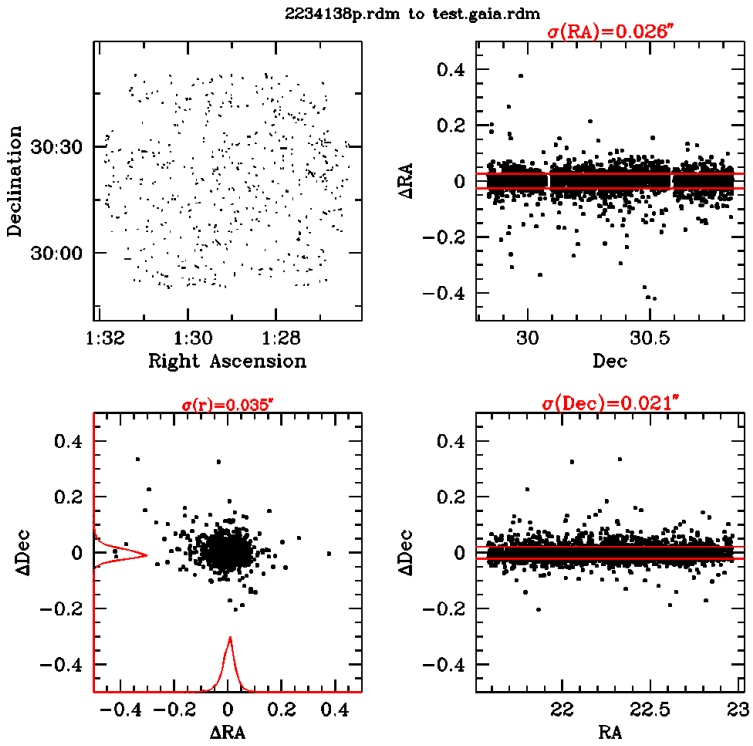


Figure 2.5: Astrometric residuals with respect to Gaia DR2.
Credit: Stephen Gwyn

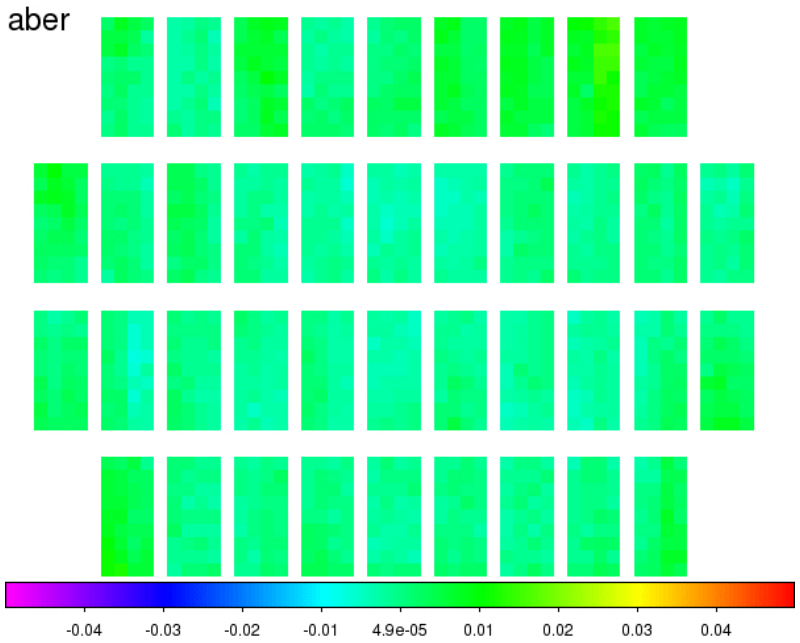


Figure 2.6: Magnitude zero-point residuals accros MegaCam field of view after calibration.
Credit: Stephen Gwyn

of the shape measurement method. The calibration is done using another technique which

will be developed later.

Since only shape measurements are tested, the following assumptions have been made. First I consider a perfect modeling of the PSF, meaning that I use the simulated PSF model as input to the galaxy shape measurement methods. I also consider a perfect extraction of sources with no detection biases. Finally, only postage stamps of objects are simulated, independently from one another. The simulation is composed of two steps: the PSF and the galaxies, which are described below.

The simulations have been created using the GalSim sofware package. To reflect the CFIS survey, all objects have three epochs with a distinct uncorrelated PSF and noise. I also used the real, observed optical distortions from the World Coordinate System (WCS) fitted on the data. To minimize the shape noise each galaxy is simulated in pairs with a $\pi/2$ rotation as the only difference. To have access to shear estimation, the simulations contain 200 batches of 10 000 galaxies (5 000 + the rotated pairs) with the same constant shear applied on each batch.

2.2.1 The PSF

To simulate a realistic CFIS PSF, I have created a two-component PSF model, optical and atmospheric. The optical part is approximated by a Moffat profile described by the equation:

$$I(x,y) = \frac{\beta - 1}{\pi\alpha^2}\left(1 + \frac{(x-x_0)^2 + (y-y_0)^2}{\alpha^2}\right)^{-\beta}, \tag{2.1}$$

with

$$\alpha = \frac{\mathrm{FWHM}}{2\sqrt{2^{1/\beta} - 1)}}, \tag{2.2}$$

where FWHM represents the Full Width at Half-Maximum, x and y are the image coordinates in pixels, x_0 and y_0 are the center of the profile and $\beta = 4.765$ (Trujillo et al. 2001)[2]. This model is then sheared to reproduce the real optical ellipticity variations. To access the optical part of the real PSF, I average the ellipticities of 500,000 stars in bins over the entire focal plane. For observations done during different nights over the years, one can estimate the atmospheric variations to be random, and is left with only the optical part. It is possible to create more complex models for an optical PSF but I wanted to have a perfect control of the ellipticity.

For the atmospheric part of the simulated PSF, I use a Kolmogorov model (Buscher et al. 1995), which can be described in Fourier space by the following transfer function:

$$T(k) \approx e^{-\frac{D(k)}{2}}, \tag{2.3}$$

with

$$D(k) = 6.8839\left(\frac{\lambda k}{2\pi r_0}\right)^{\frac{5}{3}}, \tag{2.4}$$

where λ is the wavelength of the light, and r_0 the Fried parameter. The average value of this parameter for most observing sites is 0.1 m and can go up to 0.2 m under excellent conditions. For our case it was easier to define the FWHM of the PSF instead, which can

[2]This value should be used for the atmospheric component and not the optical one. For our test this does not make a huge difference since we do not fit the PSF. However, this will be fixed in a future analysis.

be expressed as (Racine 1996):

$$\text{FWHM} \approx 0.976\frac{\lambda}{r_0}. \tag{2.5}$$

For the atmospheric component, the ellipticities are randomly set. In this case we neglect the correlations due to the atmosphere which is not a bad approximation given the exposure time used for the real images is around 200 s and the turbulence time scale is around 30 s.

Finally, both models are convolved together, and the size is set in order to replicate the distribution measured on the real images. This method allows me to have a simple PSF model, which at the same time is close to the real one. The most important advantage of this model is the varying ellipticity, which make it possible to measure residual correlations between the PSF and the PSF-corrected galaxies. This will be developed in Chapter 4.

2.2.2 The galaxies

The simulated galaxies are constructed from a sub-sample of the Cosmic Evolution Survey (COSMOS) (Scoville et al. 2007). This survey provides high-resolution and high-quality images taken by HST. The images in this sample have been carefully selected to remove blended objects. They have also been pre-processed to provide artifact-free galaxies. This sample is accessible through the GalSim software package[3] (Mandelbaum et al. 2012). Those images can then be used to simulate galaxies using different survey properties. In this case I rescaled the flux to emulate 300 s observations taken from the CFHT 3.6 m telescope. Images have been re-sampled on larger pixels with a size of ≈ 0.187 arcsec. Finally, galaxies are convolved by the previously described PSF. The advantage of this sample is that galaxies are real images including all their complexity, some examples are presented in figure 2.7.

The simulated images are created as multi-epoch objects. This means that each galaxy has three representations with different noise realizations, independent PSF and intra-pixel shifts. The noise is assumed to be Gaussian with a standard deviation of $\sigma = 14.5$ ADU.pixel^{-2} to replicate the CFIS observations. As mentioned before, I have created 200 batches with a constant shear for each batch. Shear values have been drawn from a normal distribution with $\sigma = 0.03$.

Since the PSF has full range of variations in terms of ellipticity and size, one can make an estimation of the PSF leakage (see 4.3). This test is interesting to do on simulated data given that we do not have errors coming from the creation or the interpolation of the model.

One of the crucial concerns today regarding shape measurement is handling of blended objects. To try to answer this challenge the simulations have been created in a way that allows one to create blends. At the moment, the blends are simple. I assume the galaxies are at the same redshift, which means they are sheared similarly. I also use the same PSF since the galaxies are simulated on a small postage stamp (≈ 9.5 arcsec^2). However, I have chosen not to use blended galaxies for our testing at the moment.

[3] https://github.com/GalSim-developers/GalSim

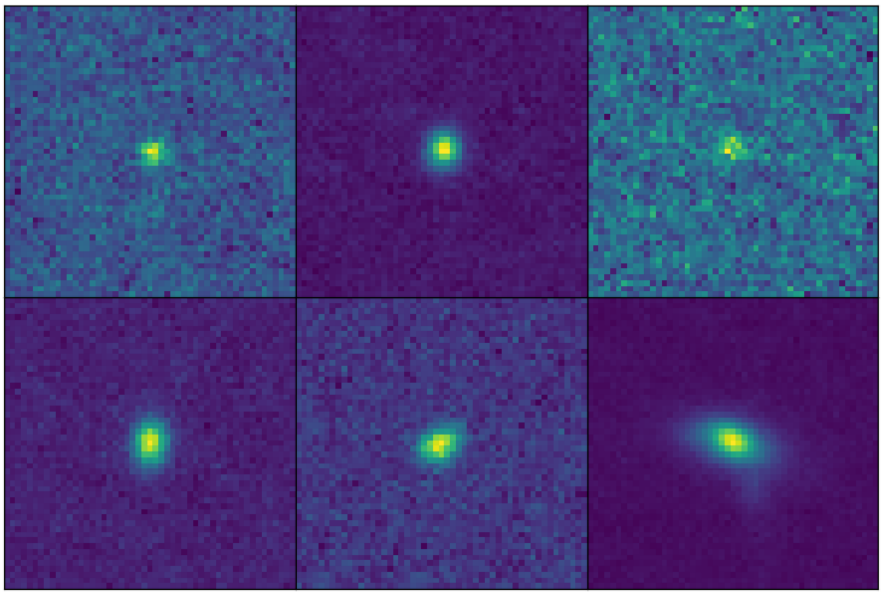


Figure 2.7: Examples of simulated galaxies.

Chapter 3

A weak lensing pipeline

In this chapter I will present the principal steps of a weak lensing pipeline. In most cases, such a pipeline starts once the images have been calibrated for both the astrometry and the photometry. A weak lensing pipeline aims to precisely measure the shapes of a pure sample of galaxies. Achieving this goal is extremely challenging since all the steps are important. I will start by describing the masking of outliers and the areas to avoid. Then, I will present the extraction of the sources and their classification. The two classes of interest are stars, to model the PSF, and galaxies, to measure the shear signal. As explained in section 1.2.3, it is primordial to have a good description of the PSF distortion. To reach the precision required for the PSF, stars are selected on single exposures and the model is constructed from them. Finally, the lensing signal to be measured only affects galaxies, which is why we need a pure galaxy sample. For the fitting of the model of the light distribution of galaxies I use a multi-epoch technique, which consists in jointly fitting all the images of a galaxy from each exposure where the galaxy appears. I also have a secondary method based on quadrupole moments applied on the stacked images with a stacked PSF model derived from single exposures.

Here I present our pipeline, ShapePipe. This pipeline has been developed in Python and makes use of external software packages, such as SExtractor, written in C. The pipeline is implemented as a modular architecture. Each module has a specific task focused on one type of images or catalogues. The core of the pipeline handles the distribution of files across the available resources. The pipeline can use different multiprocessing techniques such as Symmetric Multi-Processing (SMP) or Message Passing Interface (MPI). An inbuilt logging system allows us to track any error that is encountered during the processing. The handling of files and communication between processes and modules has been implemented to process thousands of images of several Tb size in a reasonable computing time. The pipeline is able to process the $1,695$ deg^2 of CFIS data, as we will demonstrate in chapter 5. An overview of the different modules is shown in the workflow of the pipeline presented in figure 3.1.

Most of the ShapePipe code has been developed as part of this PhD work. This has consisted in creating wrappers for external software packages (SExtractor, PSFEx, WeightWatcher, CDSClient), created the interface between the several file formats, adapted the parametrization to our specific data set. All the pipeline modules (around 10, with a total of more than 6,000 lines) have been written from scratch. We also had to modify some already existing code to adapt them to our specific problem and our data.

3.1 Masking

3.1.1 Why? What?

The masking consists on removing outliers. Artifacts can affect both the estimation of the PSF and the shape measurements of galaxies. These artifacts can have different

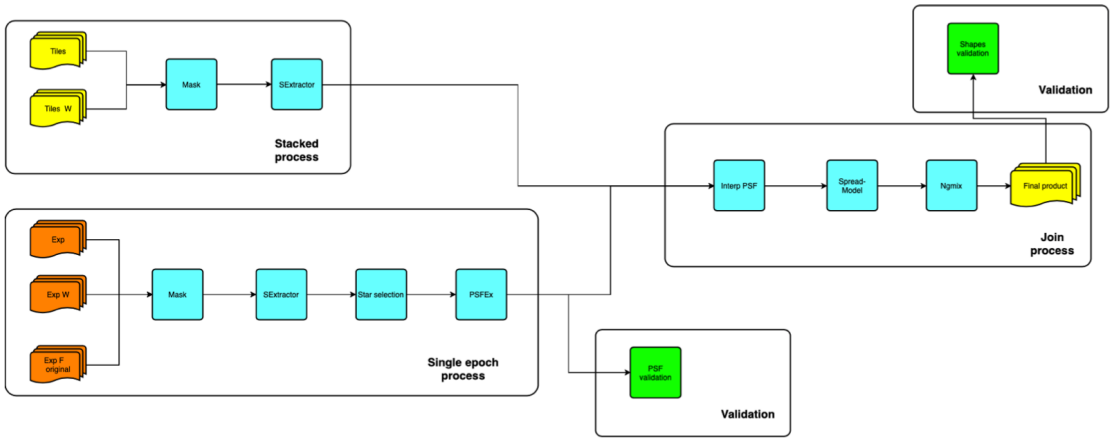


Figure 3.1: ShapePipe flowchart diagram.

origins, they can be due to electronic defects on the camera under the form of dead pixels for example. There are also areas on each Charge-Coupled Device (CCD) which we want to avoid, specifically at the edge where the behavior of pixels can be affected by their support. The sky itself can also be a source of outliers, with cosmic rays as an example. The optics of the telescope can also create halos and spikes for the brightest stars in the field-of-view.

3.1.2 How?

To mask artifacts created by the electronics, one can use the flat field images. Such images are usually used for the calibration of the illumination of the focal plane but they can also be transformed into weight maps. Indeed, flat field images will show the defective pixels but they can also show the presence of small dust particles on the optics which can be accounted for with weighting.

To mask spikes and halos around bright stars, we first need to select the objects which might represent a problem. For this I use the Guide Star Catalog (GSC) version 2.2 which provides the position of 500 million stars. Then I use templates from the THELI pipeline (Schirmer 2013; Erben et al. 2005) that are fitted using an empirical formula on the stars depending on their magnitude. For this step we have ported in python a bash script from the THELI pipeline. Masks for halos (spikes) are created for stars up to magnitude 13 (18). I also create masks for Messier objects, which are very large and bright and can disturb the object selection. An example of this kind of mask is shown in figure 3.2. The masking was validated via visual inspection of some images and compared to THELI masks.

The masking of cosmic rays is performed using a combination of the EyE package[1] with SExtractor. EyE, a machine-learning based algorithm, is used to create a special kernel filter to emphasize the signal of cosmic rays. This kernel is fed to SExtractor which will detect cosmic rays and create a segmentation image. Finally, the segmentation image is transformed to binary flags with WeightWatcher[2]. This step is integrated to the the MegaPipe pipeline used for the pre-processing detailed in section 2.1.1.

[1] https://www.astromatic.net/software/eye
[2] https://www.astromatic.net/software/weightwatcher

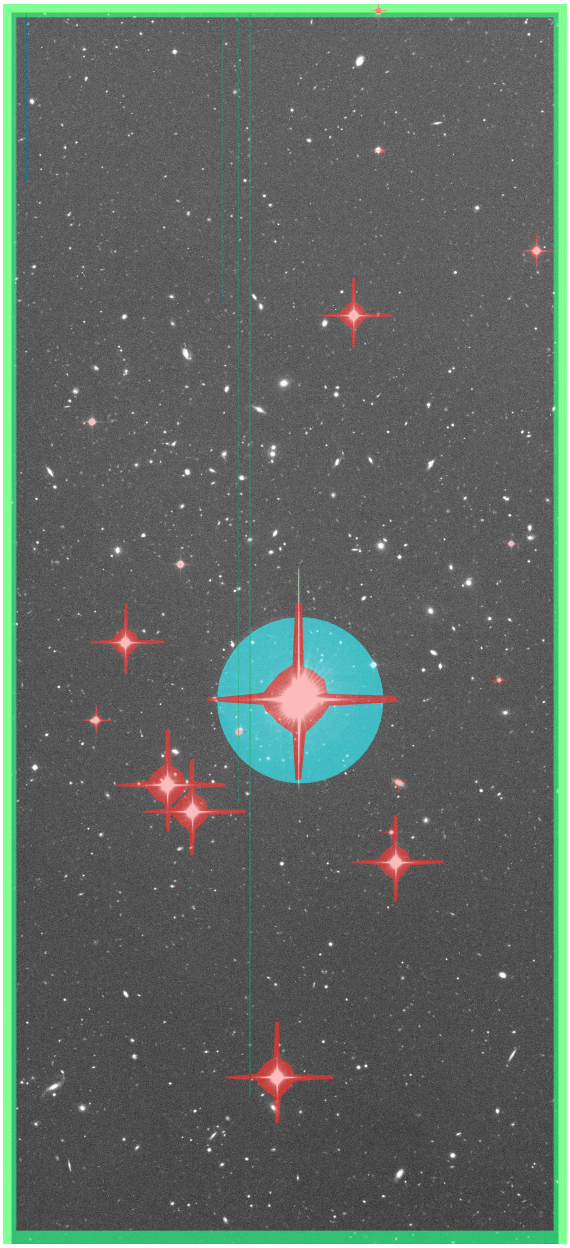


Figure 3.2: Example of a mask for a chip. Green: border mask. Red: spikes around bright stars. Light blue: halos around bright stars. Light green: dead pixels.

3.2 Source extraction and classification

3.2.1 Object identification

The source extraction consists in carefully extracting the scientific signal from the images. To achieve this selection I use the software package SExtractor (Bertin and Arnouts 1996). The idea of this algorithm is: if a contiguous number of pixels have a signal above a threshold (see table 3.1), the object is selected. Here, the threshold is a multiple of the standard deviation of the noise. In order to avoid too many spurious detections, the image can be filtered by a specific kernel (3x3 Gaussian kernel in this case). I also use in-built

Parameter	Value
THRESH_TYPE	RELATIVE
DETECT_THRESH	1.5
DETECT_MINAREA	10
FILTER	Y
FILTER_NAME	kernel_3x3.conv (default)
DEBLEND_NTHRESH	32
DEBLEND_MINCONT	0.001
Specific to single exposures	
BACK_TYPE	AUTO
BACK_SIZE	64
BACK_FILTER	3

Table 3.1: SExtraction parametrisation.

SExtractor functions to estimate the background sky level on single exposures.

SExtractor also performs photometry and morphology measurements. In my case, I make use of the MAG_AUTO estimate which is based on the Kron technique (Kron 1980). To have an estimate of the size, I use FWHM_IMAGE which assumes a Gaussian morphology, and returns the Full Width at Half-Maximum (FWHM) of the light profile. These two parameters can be combined for the star selection. Finally, the last parameters I mention here are the ones regarding the position. I use the windowed positions since they are a more reliable estimation: one first fits a Gaussian window function to the light profile to reduce the influence of the noise. Then, the center of the object is determined.

For this work, I carry out two separate runs of SExtractor. One is done on the single exposures, used to select stars for the PSF modeling. This run is also used to estimate the background which is subtracted before performing the shape measurements. The second run is performed on the stacked images where all sources for which shapes are measured are being extracted. For these two runs the main parameters have been set as presented in table 3.1. It is important to notice that DETECT_MINAREA is set to a higher value than what is generally chosen. In this case, I require at least 10 contiguous pixels to select an object. This is a conservative choice: 10 pixels roughly represent the area covered by a circular aperture with a radius of 2 pixels. This avoids the selection of too many outliers. However, in the future I plan to reduce this criterium in addition to a more robust artifact rejection.

Blends

A blend is a situation where at least two objects have an overlap of their light profile. It is very important to identify them because they are one of the main sources of bias for shape measurements (Samuroff et al. 2017), which will be even more important for upcoming, deeper surveys. In my work, I rely only on the identification performed by SExtractor based on multi-thresholding of the light distribution (see figure 2, Bertin and Arnouts 1996). In the case of the CFIS survey, the identification of blends is particularly challenging due to the fact that I am using only one photometric band. I have an on-going project on this which uses a machine-learning based algorithm.

However, in the case of weak lensing, if the objects involved in a blend are at the same redshift, they will be affected by the same gravitational distortions. Even if the blend affects the observed morphologies of those objects, an appropriate calibration can reduce the effect

on the selected sample for science studies. If the objects are at two different redshifts, the calibration can be more complex. To avoid most of the problem, the objects classified as blended are removed from our final sample, with the aim to improve the identification and handling of such objects in the future.

3.2.2 Classification

In this section I will describe how stars and galaxies are selected. The main constraint I have for this step is the fact that, for CFIS, only the r-band used for the weak-lensing science. Despite the emergence of machine learning, state-of-the-art algorithms are based on multi-band information (Machado et al. 2016). In my case, I chose to use established methods with conservative criteria to ensure the purity of my star and galaxy samples.

Star selection

To select stars, I use the fact that stars are not resolved. This property implies that the observed size of stars will be independent of their luminosity. The size of stars is only driven by the optics of the telescope and the atmospheric conditions (for ground-based surveys). Then I can plot the size, the Full Width at Half Maximum (FWHM), versus the luminosity, the magnitude in that case. Figure 3.4 shows this plot for one chip of the camera.

The selection is performed automatically. I first estimate the mode of the size distribution. This is performed by iteratively binning the distribution and selecting the bin that contains the maximum number of objects[3]. Once the mode is found, I select objects with a size of ± 0.2 pix around the mode. Figure 3.3 shows the distribution of the mode for all CCD used in our analysis of 280,000 CCDs. Finally, a cut in magnitude is applied. Even if MegaCam has a negligible brighter-fatter effect (Guyonnet et al. 2015), I prefer to remove the brightest objects with mag < 18. At low magnitudes, it becomes harder to separate stars from galaxies since we can have galaxies as small as the PSF. To avoid contamination of the star sample, I limit stars to magnitude 22. Other selections are applied when I construct the PSF model, see 3.3.1.

From the star candidates selected on each chip, I keep 20% to perform the validation tests presented in chapter 4.

Galaxy selection

Since galaxies are extracted on the stacks, the selection is also performed on the stacks for consistency. Galaxies are selected using the Spread-Model. This consists in quantifying the extensiveness of an object. To do this, each extracted object is compared to a Gaussian fit of the PSF, and an exponential profile convolved with the PSF (Mohr et al. 2012; Desai et al. 2012). The problem encountered with CFIS data is the handling of the PSF. As mentioned, due to the survey strategy, the PSF on the stack is not reliable, just as a star selection on the stacks would be a problem. The Spread-Model does not require a perfect PSF model, since the only needed information is the size to create the Gaussian profile. I have implemented my own version of the Spread-Model using the cited formalism, but applied on stacked images. The Spread-Model is written as:

$$\mathrm{SM} = \frac{\vec{G}^{\mathsf{T}}\tilde{\mathrm{W}}\vec{I}}{\vec{P}^{\mathsf{T}}\tilde{\mathrm{W}}\vec{I}} - \frac{\vec{G}^{\mathsf{T}}\tilde{\mathrm{W}}\vec{P}}{\vec{I}^{\mathsf{T}}\tilde{\mathrm{W}}\vec{P}}, \tag{3.1}$$

[3]This method does not guarantee to find the mode for every distribution. Nevertheless, our experience shows that is good enough for our purpose.

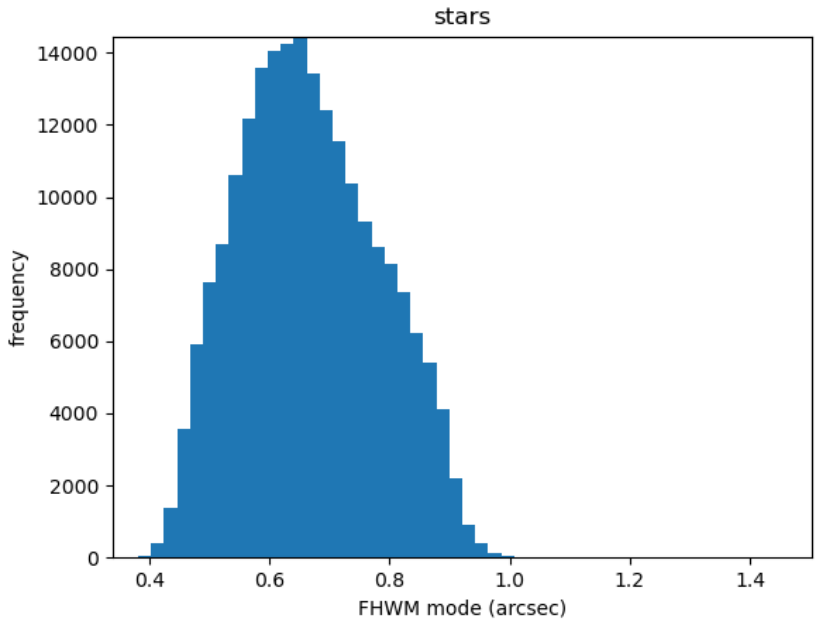


Figure 3.3: Mode of the FWHM of the stars for each CCD.

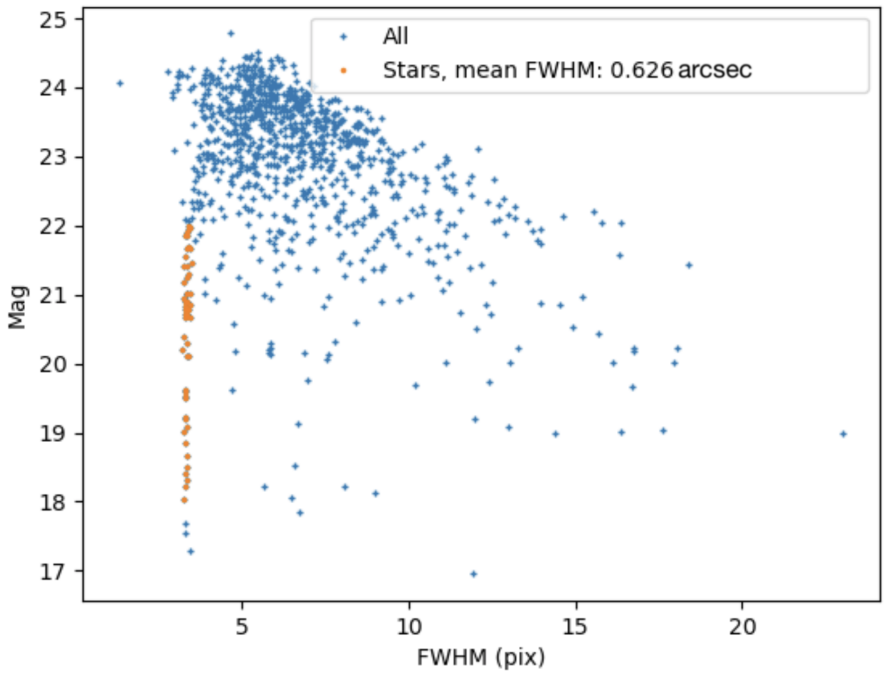


Figure 3.4: Example of a size-magnitude plot. The selected stars are show in orange while the other objects are ploted in blue.

its associated error is:

$$\sigma_{SM} = \frac{1}{(P^{\mathrm{T}}\mathrm{W}I)^2}\Big((G^{\mathrm{T}}\mathrm{Cov}G)(P^{\mathrm{T}}\mathrm{W}I)^2(P^{\mathrm{T}}\mathrm{W}I)^2 + (P^{\mathrm{T}}\mathrm{Cov}P)(G^{\mathrm{T}}\mathrm{W}I)^2 - 2(G^{\mathrm{T}}\mathrm{Cov}P)(G^{\mathrm{T}}\mathrm{W}I)(P^{\mathrm{T}}\mathrm{W}I)\Big)^{1/2}, \tag{3.2}$$

where:

- P : Isotropic Gaussian with sigma equal to the mean sigma of the PSF model of the single epoch images interpolated to the position of the object detected on the stack.
- G : Exponential profile with a scale radius of $1/16$ $\mathrm{PSF}_{\mathrm{FWHM}}$ convolved by the Gaussian P.
- I : Postage stamp of the object.
- W : Postage stamp of the weight.
- Cov : Covariance matrix of the noise, assumed to be diagonal. $\mathrm{Cov} = \mathrm{Diag}(1/\mathrm{W})$.

To test our implementation of the Spread-Model, I create image simulations of stars and galaxies. This set of simulated images has been created for the only purpose of testing the Spread-Model, and therefore are relatively simplistic. Images are created as single epochs with three epochs per objects. Galaxies are simulated from the COSMOS catalogue described in 2.2.2. Noise was added to sample the range of SNR within $[1, 50]$. The PSF is represented by a double Gaussian. One peaked Gaussian to represent the inner core of stars with 80% of the total flux. The second Gaussian, wider, represents the extension of the PSF with 20% of the total flux. Those two Gaussian profiles are then added together. A random shear is applied for each epoch of the PSF, randomly drawn from a normal distribution with a standard deviation of 0.01 and zero mean for each shear components. The size of each Gaussian profile also varies for each of the three epochs. For the inner core sizes are $[0.5, 0.6, 0.7]$ arcsec, and $[2., 2.3, 2.6]$ arcsec for the outer core. The star sample is created from the PSF by adding the same noise as the one used for the galaxies. Examples of galaxies and stars are presented in figure 3.5. To create stacked images, I use a weighted average. For this test, astrometric and mis-centering errors have been ignored. All objects are perfectly centered in each postage stamp.

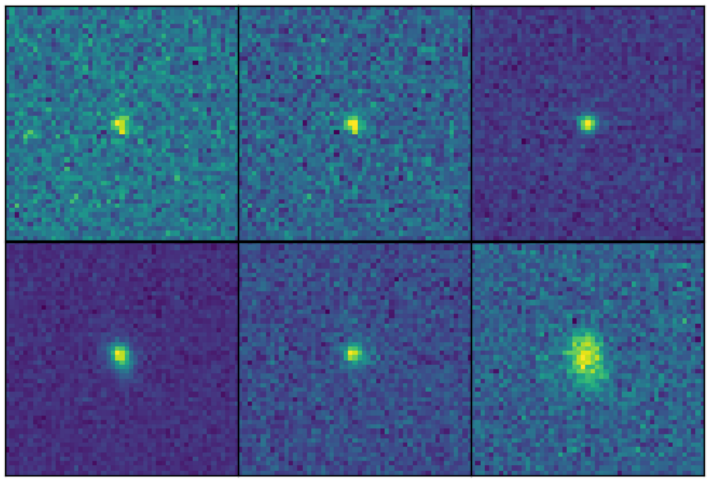


Figure 3.5: Examples of simulated objects. *Top row*: Stacked stars. *Bottom row*: Stacked galaxies

This simulation allowed me to test the behavior of the Spread-Model on stacked images of galaxies and stars in combination with PSF information coming from single epoch images. Here, I am only interested in selecting galaxies. To separate the samples I use the Spread-Model as follows:

$$\mathrm{SM} + 2\mathrm{SM}_{\mathrm{err}} > 0.0035 \tag{3.3}$$

Figure 3.6 represents the value of the two samples. It is clear that I have disjoint distributions as expected. Stars have a Spread-Model around 0 since they perfectly represent the PSF, and galaxies have a larger value since they are extended objects. An object with a $SM < 0$ would mean that it is smaller than the PSF, which are artifacts like cosmic rays, but it can also happen for objects at very low SNR. In addition to the separation presented above, a cut is applied directly on the Spread-Model values. First, I want to make sure the Spread-Model is positive. Second, I remove very large objects for which shape measurements might fail. Finally, a cut is applied on the SNR to remove the noisiest objects. The galaxy sample follows the cuts:

- $\mathrm{SM} + 2\mathrm{SM}_{\mathrm{err}} > 0.0035$
- $0 < \mathrm{SM} < 0.03$
- $\mathrm{SNR} > 10$

Applying those cuts to the simulation with 10,000 stars and 10,000 galaxies, 2.46% of stars are being mis-selected as galaxies. 74.78% of the objects are correctly classified as galaxies. In figure 3.6, the green crosses represent the selected sample.

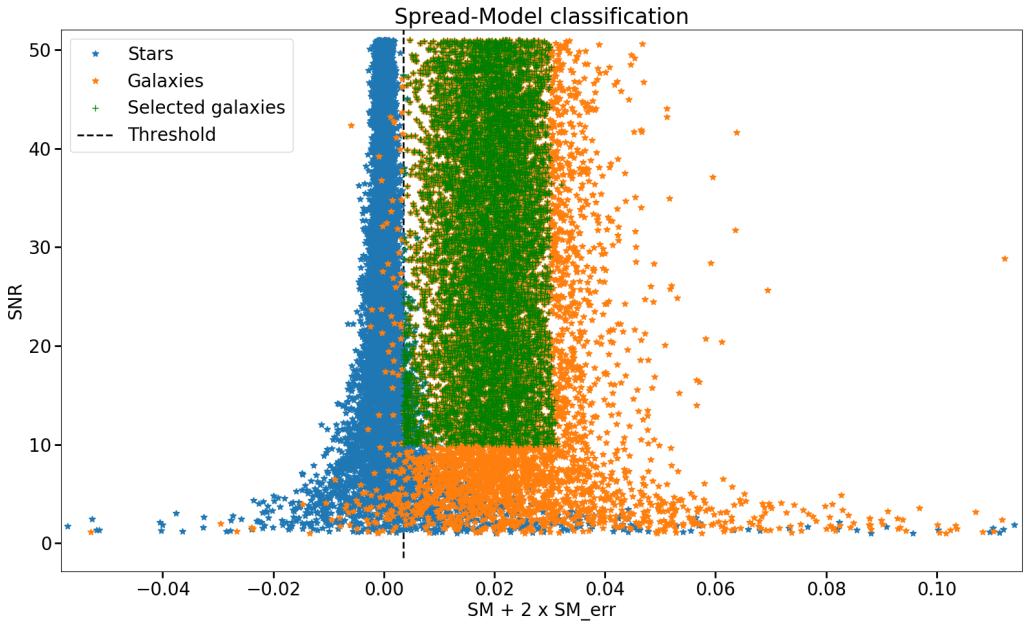


Figure 3.6: Spread-Model classification. In *Blue*: the real stars. In *Orange*: the real galaxies. In *Green*: the objects selected as galaxies. The *Threshold* represent a cut at 0.0035.

When applied to real data, the classification is shown as a function of magnitude in figure 3.7. I tested the star contamination of the galaxy sample by matching with stars

selected on single exposures for the PSF model. I find less than 10^{-2}% of stars. This contamination is a likely underestimation since I do not identify stars below magnitude 22. However, I do not expect a contamination of more than a few percent. With the fusion of CFIS with Pan-STARRS and Subaru-HSC, the purity of our galaxy sample will be improved thanks to more photometric bands. Given that very bright objects can be hard to use for shape measurements, objects brighter then magnitude 20 are removed. Regarding the science to be performed, one can extrapolate that the brightest objects are probably at low redshift, not the most interesting for weak lensing analyses. To summarize, here are the cuts performed on the real sample:

- $\mathrm{SM} + 2\mathrm{SM}_{\mathrm{err}} > 0.0035$
- $0 < SM < 0.03$
- $26 < M_r < 20$

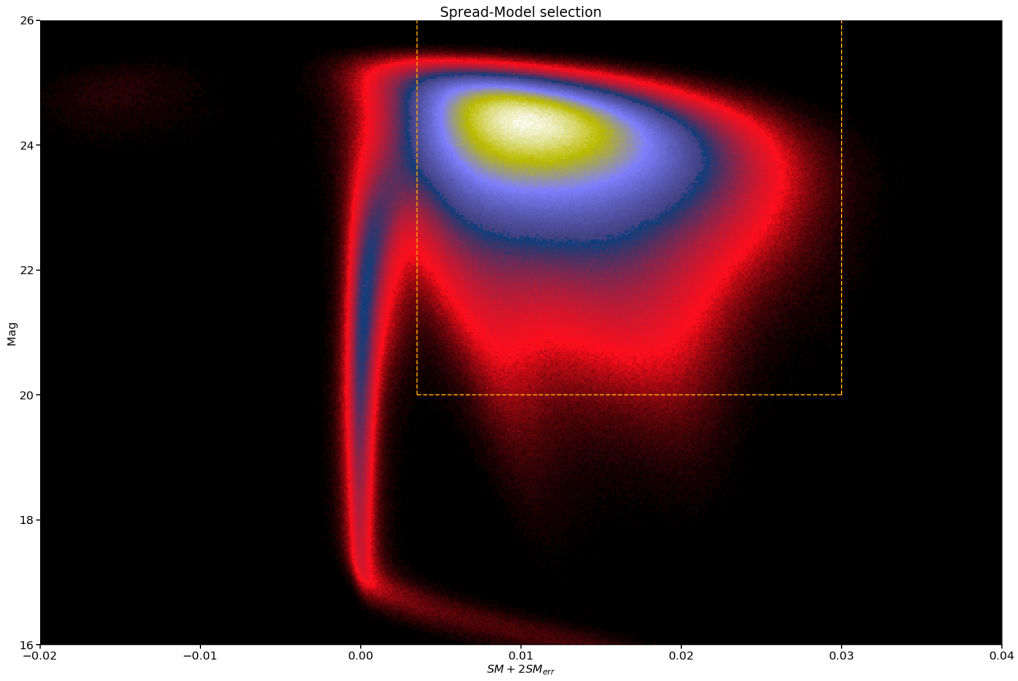


Figure 3.7: Spread-Model classification. The *orange* delimited area corresponds to the objects that have been selected for the galaxy sample.

In addition to the Spread-Model classification, more cuts are applied after shape measurements. Shape measurements are done in the framework of metacalibration which will be detailed in 3.4. To reduce selection effects, it is prefered to perform the selection at the same step as metacalibration, which allows one to account for those effects and include them in the calibration. We use the output of the model fitting procedure to define a resolution cut defined as T/T_{PSF}, where T is the size of an object deconvolved by the PSF, and T_{PSF} is the size of the PSF at the position of the object. We also apply a cut on the SNR defined as flux/σ_{flux} with σ_{flux} being the error of the flux estimation. The additional cuts are:

Parameter	Value
BASIS_TYPE	PIXEL
PSF_SAMPLING	1
PSF_PIXELSIZE	1
PSF_ACCURACY	0.01
CENTER_KEYS	XWIN_IMAGE, YWIN_IMAGE
PSFVAR_KEYS	XWIN_IMAGE, YWIN_IMAGE
PSFVAR_DEGREES	2

Table 3.2: PSFEx parametrisation.

- $\frac{T}{T_{\mathrm{PSF}}} > 0.5$
- $SNR > 10$
- $SNR < 500$

The cut of very high SNR values removes objects that are most probably a bad fit of our model.

3.3 PSF modeling

3.3.1 PSFEx model

To create the PSF model I use the PSF Extraction (PSFEx) software package[4] (Bertin 2011). PSFEx does not assume a specific model to fit the data, or any priors regarding the optical system. The PSF is solely based on the image pixels. The software builds a polynomial interpolation as a function of the position or any other parameter. The model was built to be only position-dependent with a degree-two polynomial. It is constructed on each CCD of the focal plane, independently from one another. I am currently working on a new approach for PSF modeling, which creates a model with smooth variation for all CCDs of the focal plane (Liaudat et al., *submitted*). PSFEx includes its own star selection, which I chose to bypass to force the model on the pre-selected sample. Yet, I kept the iterative model with the chi square rejection of outliers. The main parameters used are presented in table 3.2.

To construct the model, I request at least 20 stars per CCD. Their distribution over 280,000 CCDs is shown in figure 3.8. To validate the model, I keep 20% of the stars for testing. The different tests performed on the PSF will be presented in chapter 4. Those stars are not used to construct the PSF model.

3.3.2 Stacked PSF model

As mentioned earlier, it is not possible to construct a proper PSF model on the CFIS stacks due to the survey strategy. However, I want to perform a moments-based shape measurement algorithm on the stacks. To go around this problem, I adopt a strategy similar to the one used by the HSC collaboration (Bosch et al. 2017). The idea is to stack the PSF

[4] https://github.com/astromatic/psfex

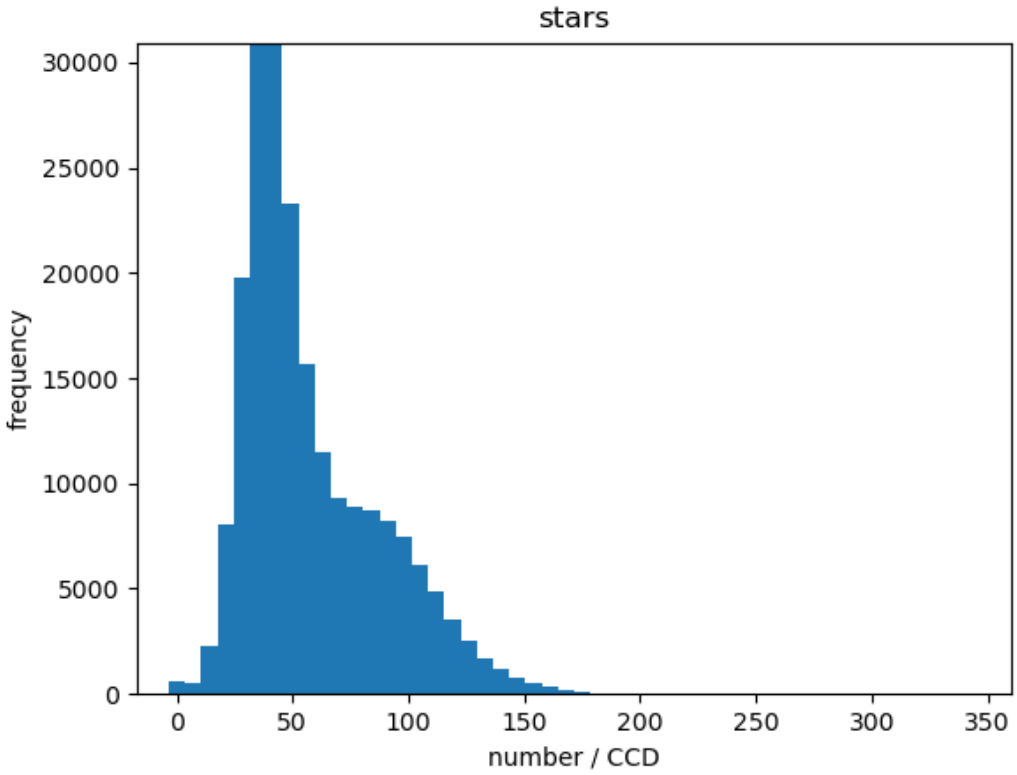


Figure 3.8: Average number of stars selected to construct the PSF model for entire processed area. I set the lower limit to 20 stars to ensure a good model.

model in the same way as the stacked images. To make the model consistent with the actual PSF on the stacks, images have to be combined in a linear way. This oriented the choice of having a weighted averaged stacking. Also, since the stacking process implies performing interpolation on single epoch images, the PSF model should be interpolated identically. In the present case, this was not possible so I used an independent interpolation method using the reproject software package[5].

Even though I used the weighted average technique, I had to use approximations for the PSF model. Indeed, usually the weights are different for each pixel. If I had applied these weights to the PSF model pixels I would have lost the convolution property of the model. I therefore homogenized the weights. The weights are averaged to a single value using a Gaussian weighting to maximize the value of the weights at the center of the images where the maximum information for the PSF is located. With this new effective weight, which I construct over the postage stamp, I perform a weighted average stacking of the model. If the weight of the PSF appears to be 0, the object is rejected to avoid inconsistency between the true stack and the stacked model. This procedure is performed on the interpolated PSF postage stamps at the position of the object from which we measure the shape.

[5]https://github.com/astropy/reproject

3.4 Shape measurements

For shape measurements, I implemented two independent methods in the pipeline. This allows for internal consistency checks. Yet, the multi-epoch method is the main measurement which will thus be used for all science analyses. Furthermore, I chose to perform shape measurements in the framework of metacalibration. This new method of shear calibration avoids the need for creating highly realistic and computer intensive simulations.

Despite the fact that the stacking method is mostly used for testing, it is important to develop such a method at the dawn of new large surveys. Indeed, for a survey like the Vera Rubin Observatory's Legacy Survey of Space and Time, it will be almost impossible to perform a multi-epoch method given the very large number of epochs, around 1,000 after 10 years of observations. We intend to keep these future applications in mind and to experiment with different techniques to make our pipeline as flexible as possible.

3.4.1 Metacalibration

Metacalibration has already been mentioned a few times, it has been introduced by Huff and Mandelbaum 2017 and used for the DES survey (Jarvis et al. 2016). Here I present the formalism and how it is applied in practice. First, I start with the usual equations which introduce how the ellipticity of galaxies is linked to the gravitational shear:

$$e_i^{\rm obs} = e_i^{\rm int} + (1 + {\rm m})\gamma_i + {\rm c} \quad i \in \{1, 2\}. \tag{3.4}$$

Here, $e_i^{\rm obs}$ is the observed ellipticity, *i.e.* the ellipticity given by shape measurements of the i-th component. $e_i^{\rm int}$ is the intrinsic ellipticity of the galaxy (before any shear is applied). γ_i is the gravitational shear, the scientific quantity we want to measure. The bias of shape measurements is supposed to be linear with the shear. ${\rm m}_i$ represents the multiplicative part, and c the additive part (they will be referred to as multiplicative and additive biases, respectively). The additive bias can also be written as a function of the PSF ellipticity. I will come back to that in the following chapter on validation tests.

The problem with equation 3.4 is the presence of $e_i^{\rm int}$ which cannot be measured. To avoid this term the shear is measured by averaging observed ellipticities. One makes the assumption that galaxies have random shapes without a preferred orientation in the absence of gravitational shear[6]. If these hypotheses are applied to the previous equation we obtain:

$$< e_i^{\rm obs} >= (1 + {\rm m}) < \gamma_i > + {\rm c}. \tag{3.5}$$

The idea of metacalibration is to measure the response of shape measurements to the shear. For this, we generalize the previous equation and introduce the response matrix as:

$$< \boldsymbol{e_i^{\rm obs}} >= \boldsymbol{R_{ij}} < \boldsymbol{\gamma_i} > + {\rm c}, \tag{3.6}$$

with

$$\boldsymbol{R_{ij}} = \frac{\boldsymbol{\partial e_i}}{\boldsymbol{\partial \gamma_j}}. \tag{3.7}$$

To calibrate the data we can then invert this matrix to get the shear:

$$< \boldsymbol{\gamma_i} >= \boldsymbol{R_{ij}^{-1}}(< \boldsymbol{e_i^{\rm obs}} > - \boldsymbol{c}). \tag{3.8}$$

[6] Observations tend to prove that this approximation is actually wrong. We call that the cosmic variance.

To estimate the response defined in equations 3.7, the derivative is approximated by finite differences. For that, one needs to apply a known, small shear to the object. This means the object has to be deconvolved from the PSF first. One can then apply the shear. Finally the object is reconvolved by an, in principle, arbitrary PSF, see below. Such operations would be trivial if images were noiseless, but of course with real observations, images are noisy and the point of the procedure used here is to handle the noise properly.

The deconvolution from the real PSF is performed in Fourier space. This step creates correlated noise and artifacts in high frequencies. When the shear is applied, additional correlated noise is created. To reduce the impact at high frequencies, the PSF used for the reconvolution is taken slightly larger than the original one. Finally, to handle the correlations due to the shear, a noise map is added on top of our images. This noise image went through the same process as the real image, except that the shear is added with a rotation by $\pi/2$. This procedure slightly reduces the SNR of the objects.

To estimate the response matrix, a positive and negative shear are applied to both components of the ellipticity with an absolute value of $|\gamma| = 0.01$. The response matrix is then approximated using finite differences:

$$\boldsymbol{R_{ij}} \approx \frac{\boldsymbol{e_i^+} - \boldsymbol{e_i^-}}{2|\boldsymbol{\gamma_j}|}, \tag{3.9}$$

with $e_i^{\pm}$ being the i-th ellipticity component with a shear $\pm|\gamma|$ applied.

For the reconvolution, one can use a simpler PSF than the original one since any anisotropy has been taken care of with the deconvolution. In this case, I chose an isotropic Gaussian profile. During this procedure an image is created without the application of shear. This will be the reference image used for the scientific measurement, while the other sheared images are used for the calibration. It is necessary to proceed this way to make sure all effects introduced during the calibration also appear on the science image. Otherwise the correction will not be adapted. The entire procedure is performed by the software package ngmix[7] [8] (Sheldon and Huff 2017).

Another interesting point of metacalibration is the ability to correct for selection effects. This is particularly interesting because the quantities on which we performed cuts, like the size or the SNR, can be correlated with the shear and lead to systematics if not corrected. To quantify selections biases we can define another response as:

$$< \boldsymbol{R_{ij}^S} > \approx \frac{< \boldsymbol{e_i^{0,M+}} > - < \boldsymbol{e_i^{0,M-}} >}{2|\boldsymbol{\gamma_j}|}, \tag{3.10}$$

were $< e_i^{0,M\pm} >$ is the averaged no-shear ellipticity of the i-th component with the mask computed on the image with a shear $\pm|\gamma|$ applied. The masking includes all the cuts on the different measured properties (SNR, size, ...).

3.4.2 Multi-epoch shape measurement

The multi-epoch measurement consists on performing a joint fit of several images of the same object. The fitting is performed using ngmix. The model used here is a Gaussian profile. This choice is mainly driven by the computing time. Also, this reduces the number of parameters to be minimized during the fit. Here, there are only 6 free parameters:

[7] https://github.com/esheldon/ngmix

[8] Link to my branch including my modifications to the code: https://github.com/aguinot/ngmix/tree/modif_observation

- $\Delta x, \Delta y$: the centering shifts,
- e_1, e_2: the 2 ellipticity components,
- r_{50}: the half-light radius,
- F: the flux.

The fit is done in the WCS referential. This has the advantage of having a fixed referential in which one can define ellipticities. That way, the distortion are also handled properly.

At the end, a covariance matrix is obtained. It can be used in different ways. One can construct weights based on the errors on the ellipticity as follows (Jarvis et al. 2016):

$$w = \frac{1}{2\sigma_{sn}^2 + \sigma_{e_1}^2 + \sigma_{e_2}^2}, \tag{3.11}$$

where σ_{sn}^2 is the shape noise fixed at 0.31 (measured on data). $\sigma_{e_i}^2$ for $i \in \{1, 2\}$ is the variance of the measured ellipticity. The error on the flux, σ_{flux}, is used to define the SNR as:

$$\text{SNR} = \frac{\text{flux}}{\sigma_{\text{flux}}}. \tag{3.12}$$

The model is created with the software package Galsim. This software package can draw the model on the sphere using WCS. Using Galsim makes this method flexible for future applications. We could for example create a more complex model for other purposes such as precise photometry or studies of morphology.

3.4.3 Shape measurement on the stack

To measure the ellipticity on the stack, an independent method was chosen. My choice went to quadrupole moments for efficiency reasons once again. First thoughts pointed me to the KSB method (Kaiser, Squires, and Broadhurst 1995), but given the framework of metacalibration, shape measurements are actually performed on the object with an isotropic PSF. The tests show that weighted moments give better results than KSB in terms of multiplicative bias where the response distribution did not have a well define peak. In the future, I plan to move to a partial KSB to keep the correction from the window function and the isotropic PSF. By using the weighted moments and neglecting the effects of the PSF, the amplitude of the shear is under-estimated. This effect is included in the multiplicative bias defined in equation 3.5, and calibrated.

Chapter 4

Validation tests

In this chapter, I present the tests I performed on the data to validate the different steps of the pipeline. The first step is to validate the PSF model. As I will demonstrate, this step can be done using data only. Contrariwise, the validation of shape measurements have to be performed on simulations. Here I used the simulations presented in section 2.2. One can estimate the residual additive and multiplicative biases after calibration. Finally, I look at the residual PSF information in galaxies shapes. This is commonly called PSF leakage.

4.1 PSF Validation

First I focus on the validation of the PSF model. To validate the model I have kept 20% of the star sample, which was not used to create the model. This allows us to test the accuracy of the model, the interpolation of the model, and also to make sure there is no over-fitting problem.

4.1.1 Residuals

A simple test of the PSF model is to look at the residuals between the ellipticity/size of the model compared to measured stars. The residuals are defined as $\delta e_{\rm PSF} = e_{\rm star} - e_{\rm PSF}$ and $\delta T_{\rm PSF} = T_{\rm star} - T_{\rm PSF}$ for ellipticity and size, respectively. This provides indications on the fit of the model but it is difficult to link those values to an impact on science results. Yet, it is interesting to look for particular patterns in the focal plane. The residuals are computed at the single-exposure level. Since there are overlaps between exposures, the same object may appear more than once. This is not an issue given the dithering used to construct the survey, the object will be at different positions in the focal plane. I decided to keep the repeated objects because they improve the statistics. I used all of the processed area for those tests, which represents around 1,700 deg^2 with 2.5 million stars. More details on the processed area are given in chapter 5.

Figure 4.1 shows the number of stars in small spatial bins. Each CCD has been split into 15 bins along the x-axis and 30 bins along the y-axis. This represents squares with a size of 150 pixels x 150 pixels, or 28 arcsec x 28 arcsec. The first feature to notice on this plot is the lack of stars in the corners. The bottom right corner is highly impacted. At the moment the reason why such a small number of stars is detected in this region is not known. This problem was also observed at the zero-point calibration level with larger deviation in this CCD compared to the others.

The residuals on the ellipticity are presented in the second row of figure 4.2. Their amplitude is around 10 times smaller than the amplitude of the ellipticity, shown in the first row of that figure. Perfect residuals would have a random pattern. Here, one can see circular patterns. This can be caused by a too small polynomial order for the interpolation of the

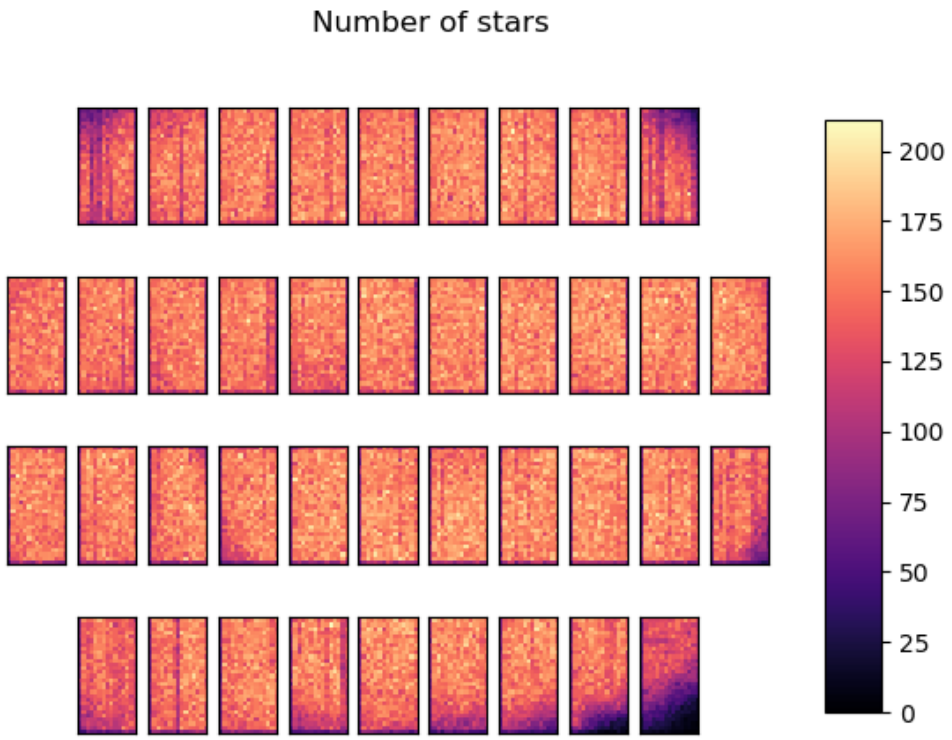


Figure 4.1: Average number of stars selected for testing the model.

model. In that case, those patterns are residuals of higher order PSF variations. However, increasing the degree of the polynomial comes with a cost. If there are not enough stars, the model will be noisier and precision on smaller scales will be lost. To be conservative, a degree 2 polynomial is kept and those residuals are neglected now. Another idea to reduce the impact on large scales would be to have a model that fits the entire focal plane at once. As mentioned in 3.3.1, we are currently working on such a method which shows promising results. We also look at the residual on the size, see figure 4.3, where the large-scale patterns are less significant.

4.1.2 ρ-statistics

To know the actual impact of the residuals on science results, I consider the so-called ρ-statistics. These two-point correlation functions have been introduced by Rowe 2010 and

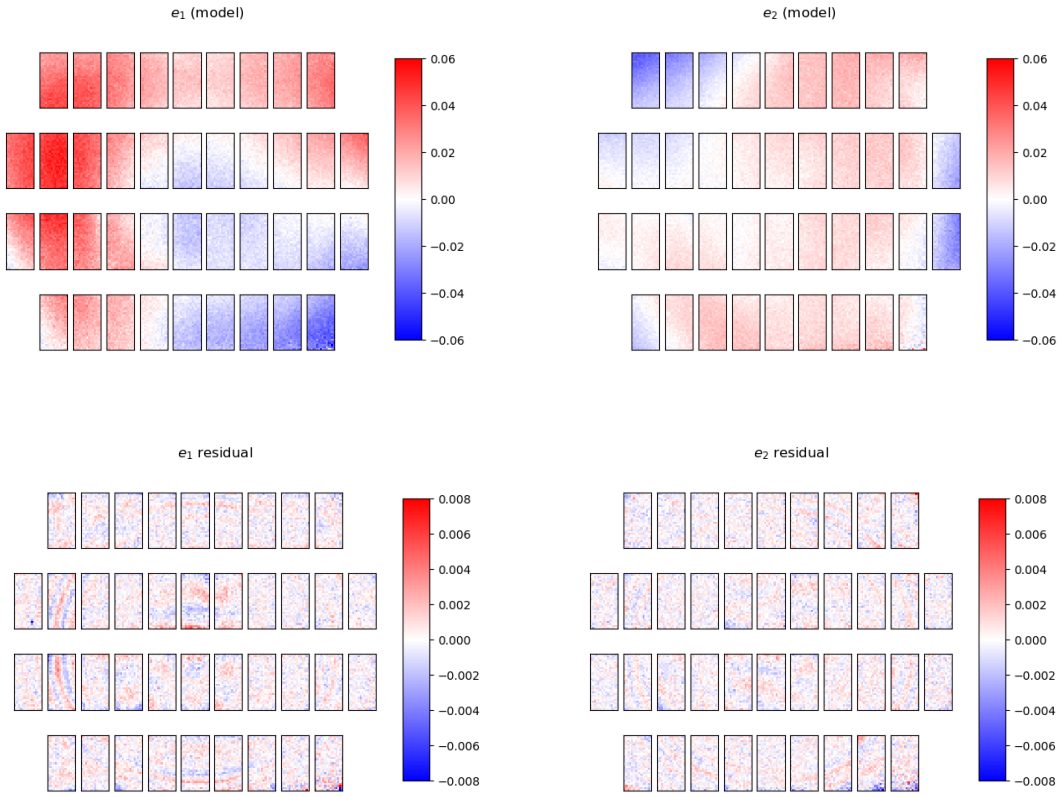


Figure 4.2: PSF ellipticity residuals in the focal plane.

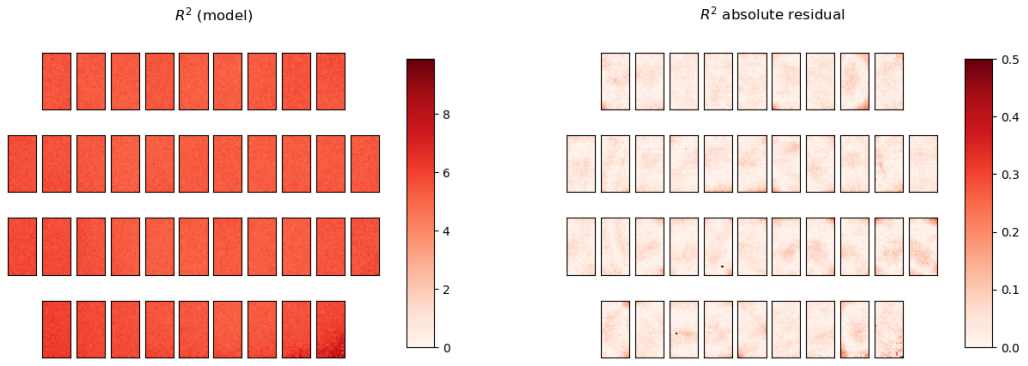


Figure 4.3: PSF size residuals in the focal plane.

then extended by Jarvis et al. 2016. They are defined as follows:

$$\rho_1(\theta) = \langle \delta e^*_{\rm PSF}(\vec{X}) \delta e_{\rm PSF}(\vec{X}+\vec{\theta}) \rangle; \tag{4.1}$$

$$\rho_2(\theta) = \langle e^*_{\rm PSF}(\vec{X}) \delta e_{\rm PSF}(\vec{X}+\vec{\theta}) \rangle; \tag{4.2}$$

$$\rho_3(\theta) = \left\langle \left(e^*_{\rm PSF} \frac{\delta T_{\rm PSF}}{T_{\rm PSF}} \right)(\vec{X}) \left(e_{\rm PSF} \frac{\delta T_{\rm PSF}}{T_{\rm PSF}} \right)(\vec{X}+\vec{\theta}) \right\rangle; \tag{4.3}$$

$$\rho_4(\theta) = \left\langle \delta e^*_{\rm PSF}(\vec{X}) \left(e_{\rm PSF} \frac{\delta T_{\rm PSF}}{T_{\rm PSF}} \right)(\vec{X}+\vec{\theta}) \right\rangle; \tag{4.4}$$

$$\rho_5(\theta) = \left\langle e^*_{\rm PSF}(\vec{X}) \left(e_{\rm PSF} \frac{\delta T_{\rm PSF}}{T_{\rm PSF}} \right)(\vec{X}+\vec{\theta}) \right\rangle, \tag{4.5}$$

with $e_{\rm PSF}$ being the PSF ellipticity, $T_{\rm PSF}$ the size of the PSF defined as $T_{\rm PSF} = 2\sigma^2$, $\delta e_{\rm PSF}$ the ellipticity residuals, and $\delta T_{\rm PSF}$ the size residuals. This metric can relate the residuals of the PSF model to cosmology based on the shear 2-points correlation function. It can be shown in Jarvis et al. 2016 that the ρ_i contribute as additive systematic terms to the shear correlation function. One can then set an upper value to the ρ-statistics to be within defined requirements. The constraints are set as follows:

$$|\rho_{1,2,3}(\theta)| < \left\langle \frac{T_{\rm PSF}}{T_{\rm gal}} \right\rangle^{-2} \delta\xi_+^{max}(\theta); \tag{4.6}$$

$$|\rho_{2,5}(\theta)| < |\alpha|^{-1} \left\langle \frac{T_{\rm PSF}}{T_{\rm gal}} \right\rangle^{-1} \delta\xi_+^{max}(\theta), \tag{4.7}$$

where $\delta\xi_+^{\rm max}$ is the sensitivity of ξ_+ with respect to the cosmology, and α represents the PSF leakage (to be detailed in 4.3). In my case, I have considered only one parameter for the requirements, σ_8, from which I can define $\delta\xi_+^{\rm max}$ as:

$$\delta\xi_+^{max}(\theta) = \frac{\partial\xi_+(\theta)}{\partial\sigma_8}\delta\sigma_8. \tag{4.8}$$

In CFIS, no specific requirements this or other parameters have been specified, since cosmology is not the primary focus of the survey. Yet, I am interested to know if this processing can reach a similar precision compared to recent cosmological surveys such as DES Y1. Therefore, I have considered an error of 3% on the estimation of σ_8, which can be writen as $\delta\sigma_8/\sigma_8 < 0.03$. For the purpose of testing, I have considered a PSF leakage $|\alpha| < 0.03$, and I have set the size ratio $T_{\rm PSF}/T_{\rm gal} = 1$. Results are presented in figure 4.4 for the multi-epoch PSF model. The region in *grey* represents the requirements, we can see that on large scales we tend to go out of the requirements. I assume this is due to the large-scale patterns observed in the residuals. It could be improved by a model fitted on the entire focal plane. On scales smaller than around 100 arcmin, the ρ-statistics are within our requirements, and I conclude that our PSF modeling is of comparable quality than DES Y1.

The requirements presented here have been measured assuming a Planck cosmology (Aghanim et al. 2020) and the 2-point correlation functions have been computed using the Core Cosmology Library (CCL)[1]. For the redshift distribution of source galaxies, I have matched our data in the W3 region to the CFHTLenS study (Erben et al. 2013), and constructed our $N(z)$ from their photo-z's.

[1] https://github.com/LSSTDESC/CCL

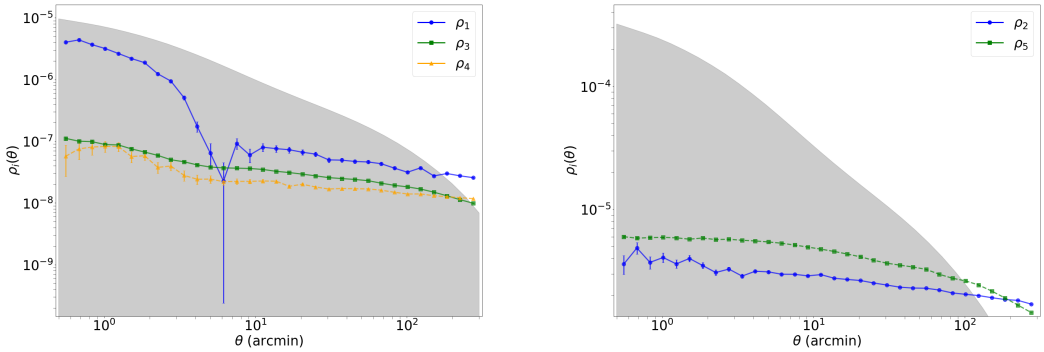


Figure 4.4: PSF residuals ρ-statistics for the multi-epoch PSF model. The *grey* area represents the requirement described in 4.1.2. Here $T_{\mathrm{PSF}}/T_{\mathrm{gal}} = 1$ and $\alpha = 0.03$. For the computation of ξ_+ CCL is used.

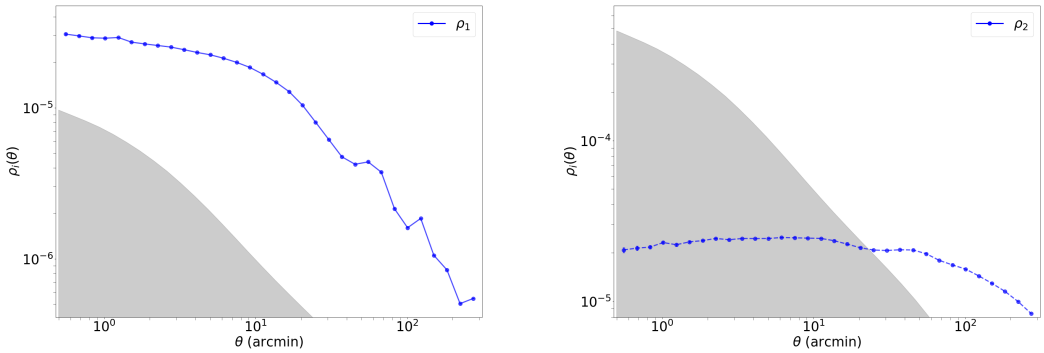


Figure 4.5: PSF residuals ρ-statistics for the stacked PSF model. The *grey* area represents the requirement described in 4.1.2. Here $T_{\mathrm{PSF}}/T_{\mathrm{gal}} = 1$ and $\alpha = 0.03$. For the computation of ξ_+ CCL is used.

4.1.3 Validation of the stacked PSF

Given the transformation applied to the model constructed at the single exposure level, one cannot guarantee that the stacked model is precise enough. With the survey strategy used for CFIS, I cannot reproduce the residual plots with the stacked model. However, it is possible to look at the ρ-statistics.

To compute the ρ-statistics, I need the ellipticity and size information for both the stars and the PSF. At the moment, I use different estimators for both[2]. The star size and ellipticity are measured using SExtractor weighted moments, while the corresponding values of the PSF are estimated using the adaptive moments of Galsim.

As one can see in figure 4.5, this model does not fulfill the requirements. Due to the differences of the two estimators, especially on the size, I only present ρ_1 and ρ_2. This is not the only explanation for the differences observed. Here, the stacking also has a negative effect. This can be explained by the fact that the interpolation process applied on the stars

[2]This will be fixed in future processing of the data. Unfortunately it was not possible to re-run the pipeline on the entire dataset before the submission of this manuscript.

and the PSF are not identical. I plan to improve this part in the future to make the PSF model more consistent with stacked images.

4.2 Shear systematics

To estimate the residual shear bias for the shape measurements, I have to use image simulations. To have a meaningful estimation of the errors, simulations have to reflect as much as possible the real images. In my case, the purpose was to test the implementation of ngmix, and not to validate the method itself since this has already been done by other groups (Sheldon and Huff 2017; Jarvis et al. 2016; Zuntz et al. 2018).

Those tests are only focused on shape measurements. I have excluded all other potential effects due to object selection or PSF modeling. The simulations used here are described in 2.2. I am not validating my measurement on the stack with the quadrupole moments, because I already know from 4.1.3 that our PSF model is a limiting factor.

4.2.1 Multiplicative bias

To estimate the multiplicative bias, I run the shape measurement algorithm including the metacalibration step on 10,000 galaxies with the same constant shear. I repeat this operation 200 times with a different shear applied each time. Once all measurements have been calibrated for each field of constant shear, independently, I look at the residual between measured and true shear.

The selection done on the simulations is similar to the one performed on the real data. Only the selection post-metacalibration is applied. Here one need to be careful when applying the different cuts. To reduce the shape noise all galaxies are simulated with a rotated version. When an object is removed, I remove the entire pair to ensure that the average shear of the sample is consistent with the true shear.

The residuals are defined as: $\Delta g_i = g_i^{\text{obs}} - g_i^{\text{true}}$ with $g_i^{\text{obs}} =< e_i^{\text{obs}} >$. I perform a linear fit of Δg_i as a function of g_i^{true} to obtain the multiplicative and additive bias. To see the impact of the calibration and the weighting, I test different configurations. First, figure 4.6 shows the residual bias without applying the metacalibration correction and without weighting. One can see a residual multiplicative bias of around $25\% \pm 0.25\%$. Shape measurement methods that do not use metacalibration are typically at around 5%. The large bias in our case is due to operations of deconvolution/convolution applied to the image to perform metacalibration. It is interesting to see that the additive bias is not affected, it is in the 10^{-4} range.

Next, figure 4.7 shows the impact of weighting. When calibration is applied we reach a residual multiplicative bias of $4\% \pm 0.34\%$. If compared to figure 4.8 where the weighting is applied, one can see that weights have a non-negligible impact. Once again, I explain this by the transformation applied for the purpose of metacalibration. For galaxies at low SNR, metacalibration can have a major impact and disturb shape measurements drastically.

Despite the impact of metacalibration on the images, overall this method seams to be able to calibrate the shear bias. My results on the simulations, presented in figure 4.8, show a residual multiplicative bias less than $1\% \pm 0.3\%$, and an additive bias of around 10^{-4}. The implementation used in my pipeline shows results in the expected range for this method (Sheldon and Huff 2017).

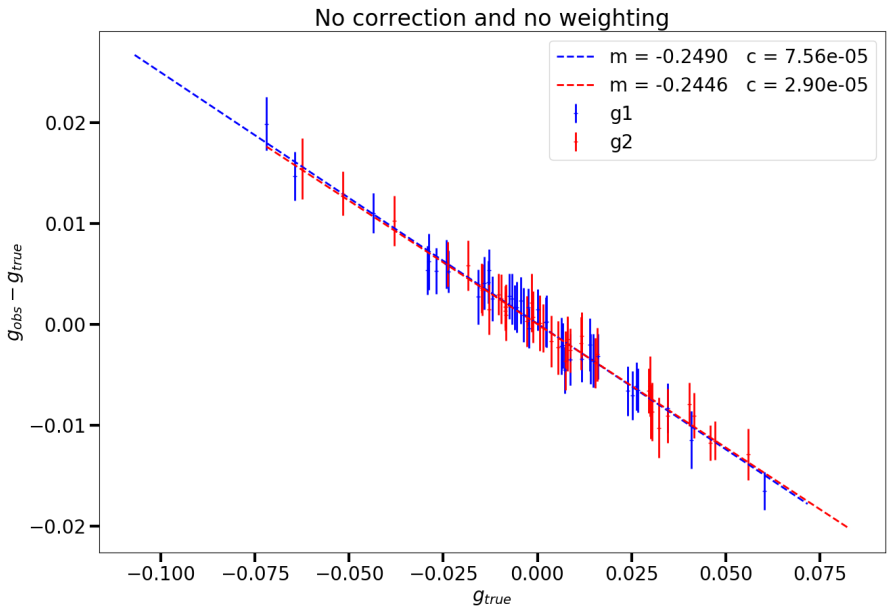


Figure 4.6: Residual multiplicative bias without weighting and calibration. (For clarity, only one fifth of the points are shown)

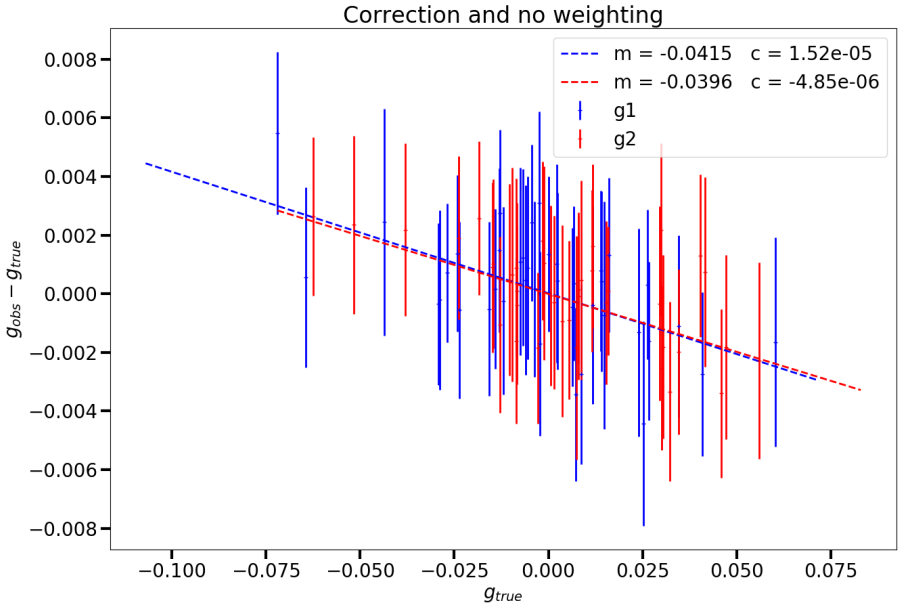


Figure 4.7: Residual multiplicative bias after calibration without weighting. (For clarity, only one fifth of the points are shown)

4.2.2 Additive bias

The additive bias has been tested on the simulation as well, see previous section. Yet, one can also estimate it on real data. To measure this bias I compute the weighted average of the calibrated shear measurements. This weighted average is done using a jackknife to

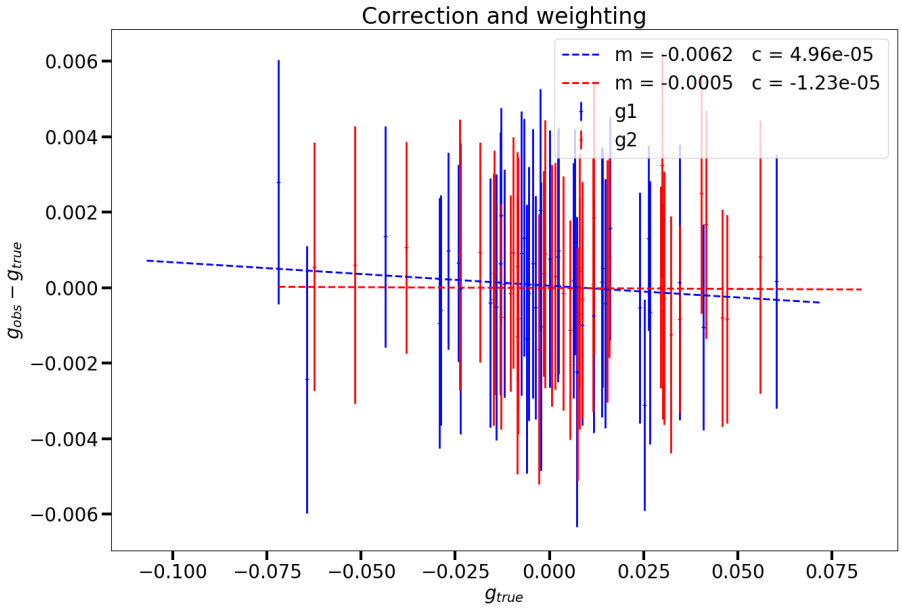


Figure 4.8: Residual multiplicative bias after calibration and weighting.
(For clarity, only one fifth of the points are shown)

Field	additive bias	
	c_1	c_2
P1	$-8.0\text{x}10^{-4} \pm 7.9\text{x}10^{-5}$	$4.8\text{x}10^{-4} \pm 8.0\text{x}10^{-5}$
P2	$2.8\text{x}10^{-4} \pm 1.5\text{x}10^{-4}$	$5.6\text{x}10^{-4} \pm 1.5\text{x}10^{-4}$
P3	$-7.6\text{x}10^{-4} \pm 1.4\text{x}10^{-4}$	$7\text{x}10^{-4} \pm 1.5\text{x}10^{-4}$
P4	$-1.7\text{x}10^{-4} \pm 1.2\text{x}10^{-4}$	$2.1\text{x}10^{-4} \pm 1.3\text{x}10^{-4}$

Table 4.1: Additive bias for both components of the ellipticity with ngmix

be sure the measurement is more robust against outliers. I compute this average 500 times and randomly remove 5% of the data each time. Results obtained for the different fields studied are presented in table 4.1 for ngmix, and in table 4.2 for Galsim. This technique is sensitive to the cosmic variance which could bias this estimation. However, I used an area large enough to assume that the average shear is zero.

For both methods, the measured additive bias is below 1% of the shear of order 0.1 we aim to measure. Yet, ngmix yields better results, which strengthen my preference for this method to perform the science analysis. One hypothesis to explain the larger values for shape measurements with Galsim is the imperfection in the PSF model.

4.3 PSF/Shapes correlations

In the previous two sections I validated the PSF model and shape measurements. To finish the validation process, I have to make sure that the PSF contribution to galaxy shapes is negligible. Usually, if there is a correlation between galaxy ellipticities and the PSF it will

Field	additive bias	
	c_1	c_2
P1	$-3.3\text{x}10^{-4} \pm 6.9\text{x}10^{-5}$	$-1.3\text{x}10^{-4} \pm 6.9\text{x}10^{-5}$
P2	$-1.8\text{x}10^{-4} \pm 1.2\text{x}10^{-4}$	$-1.0\text{x}10^{-4} \pm 1.1\text{x}10^{-4}$
P3	$-3.1\text{x}10^{-3} \pm 1.2\text{x}10^{-4}$	$8.1\text{x}10^{-4} \pm 1.2\text{x}10^{-4}$
P4	$-2.4\text{x}10^{-4} \pm 1.1\text{x}10^{-4}$	$-2.7\text{x}10^{-4} \pm 1.1\text{x}10^{-4}$

Table 4.2: Additive bias for both components of the ellipticity with Galsim.

appear as an additive bias. One can include that in the equation 3.4:

$$e_i^{\text{obs}} = e_i^{\text{int}} + (1 + \text{m})\gamma_i + \alpha e_i^{\text{psf}}, \tag{4.9}$$

where α represents the PSF leakage.

I will use two different statistics to estimate this leakage. The first one measures the shape of the galaxy as a function of the shape of the PSF at the galaxy position (4.3.1). The second method is based on the cross-correlation between galaxy ellipticities and the PSF at the position of the stars (4.3.2). The first method is a local estimate of the leakage, whereas the second method accounts for the spatial variations of the PSF leakage, measuring its impact on different scales of the shear correlation function.

4.3.1 PSF leakage

First, I measure the average ellipticity of galaxies in bins of PSF ellipticity. For this test, I take the ellipticity of the PSF at the position of the galaxies. The bins are constructed to have the same number of galaxies in each of them. I compute the weighted average in those bins using a jackknife to have an error estimation. For the fit, I use all data points, not the averaged value of each bin. Figure 4.9 shows the PSF leakage for ngmix and figure 4.10 for Galsim in the field P3 (see 5 for a definition) (the results for the other patches are presented in appendix A). Ngmix does not show a cross leakage, which is a correlation between different ellipticity components ($e_1^{\text{gal}}\text{x}e_2^{\text{psf}}$ for example). This is not the case for Galsim for e_2^{psf}. Overall, both show very small leakages around 1%. One can also see that there is no correlation with the size of the PSF, which could be due to a bad behavior of the metacalibration operations.

It is important to note that this leakage could come from a bad deconvolution in the metacal process, or a bad modeling of the PSF. In this plot, it is not possible to distinguish between the two cases, but as presented in Zuntz et al. 2018, they observed that this leakage is in fact coming from errors in the model. I will improve this test in the future to see if I am facing a similar scenario. To give a first answer, I had a look at the PSF leakage in the simulated data. In that case, I do not have errors due to a bad PSF model. Figure 4.11 shows the leakage with the true value of PSF ellipticities. I do not find a significant leakage in the simulation. At the moment the measurement is dominated by statistical errors because I do not have enough objects in the simulation. In the future, I will reproduce the test with a larger batch of simulated data.

Ngmix

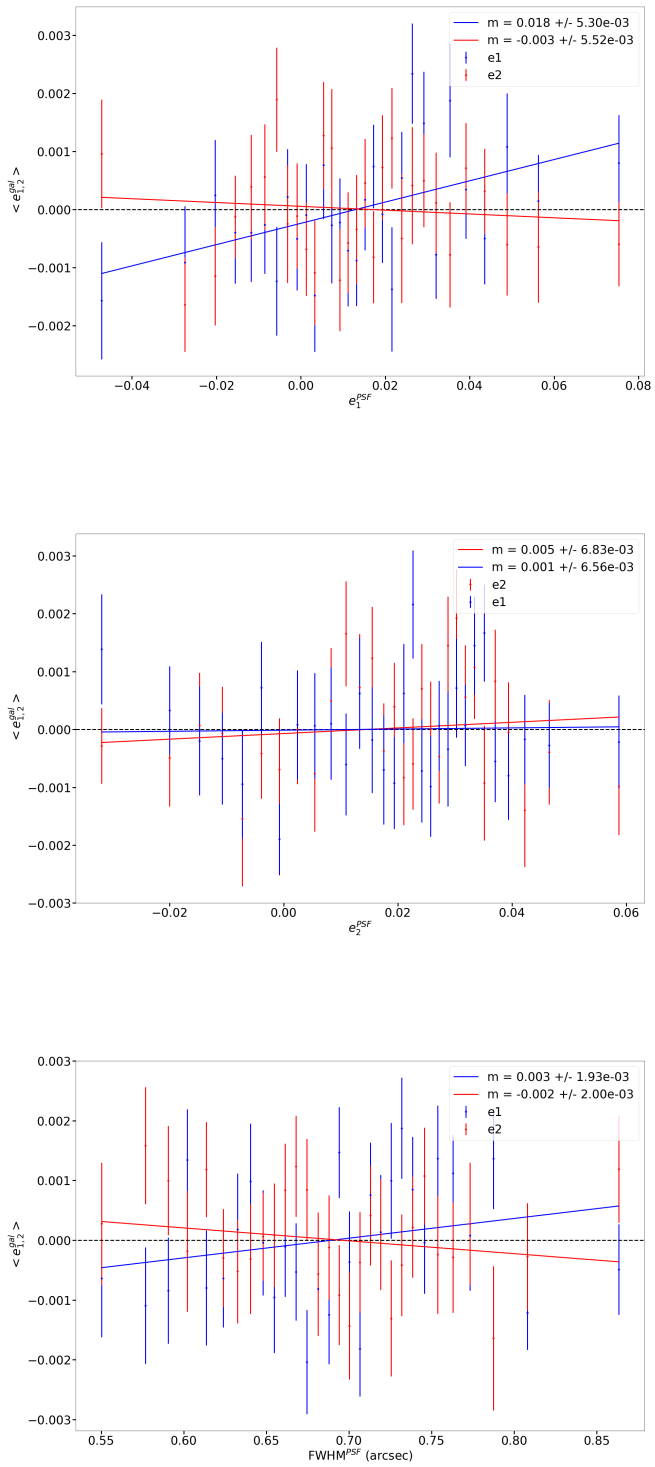


Figure 4.9: PSF leakage using the average galaxy shape in bins of PSF ellipticity (top two panels) and size (bottom panel) on the field P3 for the ngmix catalogue.

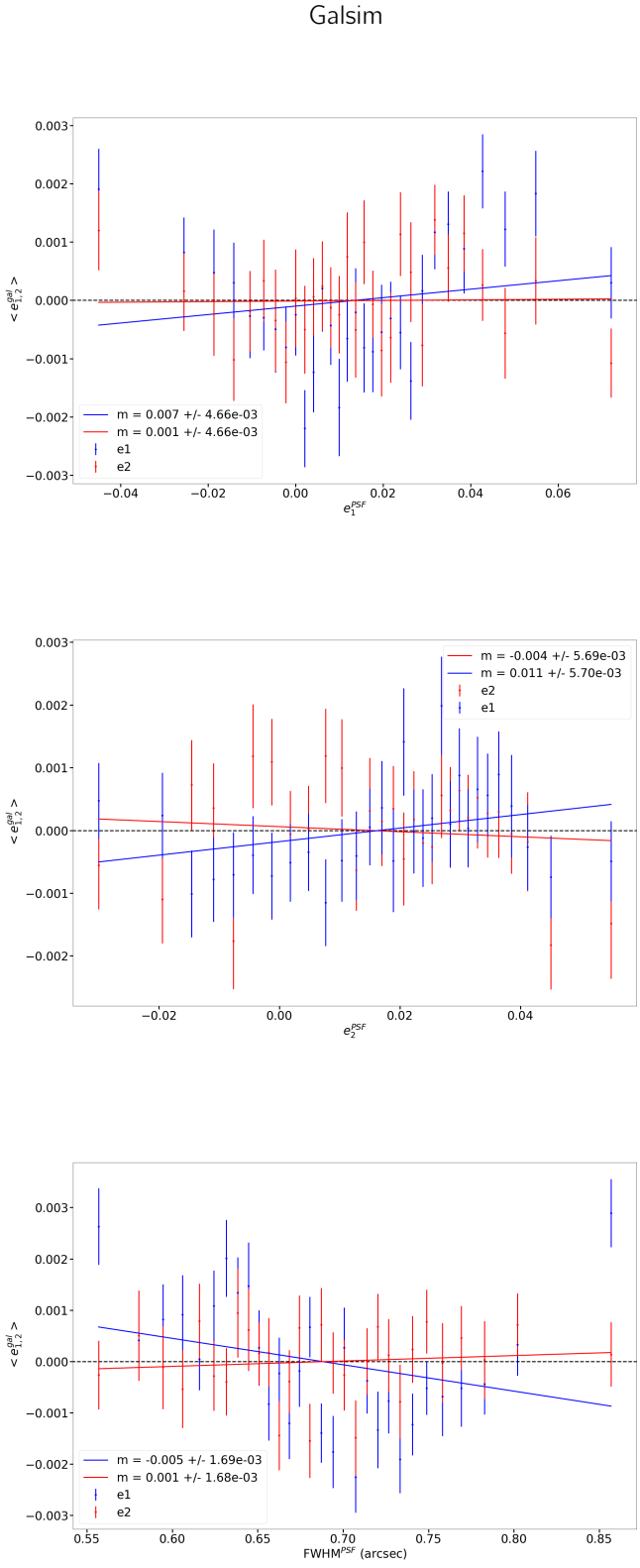


Figure 4.10: PSF leakage using the average galaxy shape in bins of PSF ellipticity (top two panels) and size (bottom panel) on the field P3 for the Galsim catalogue.

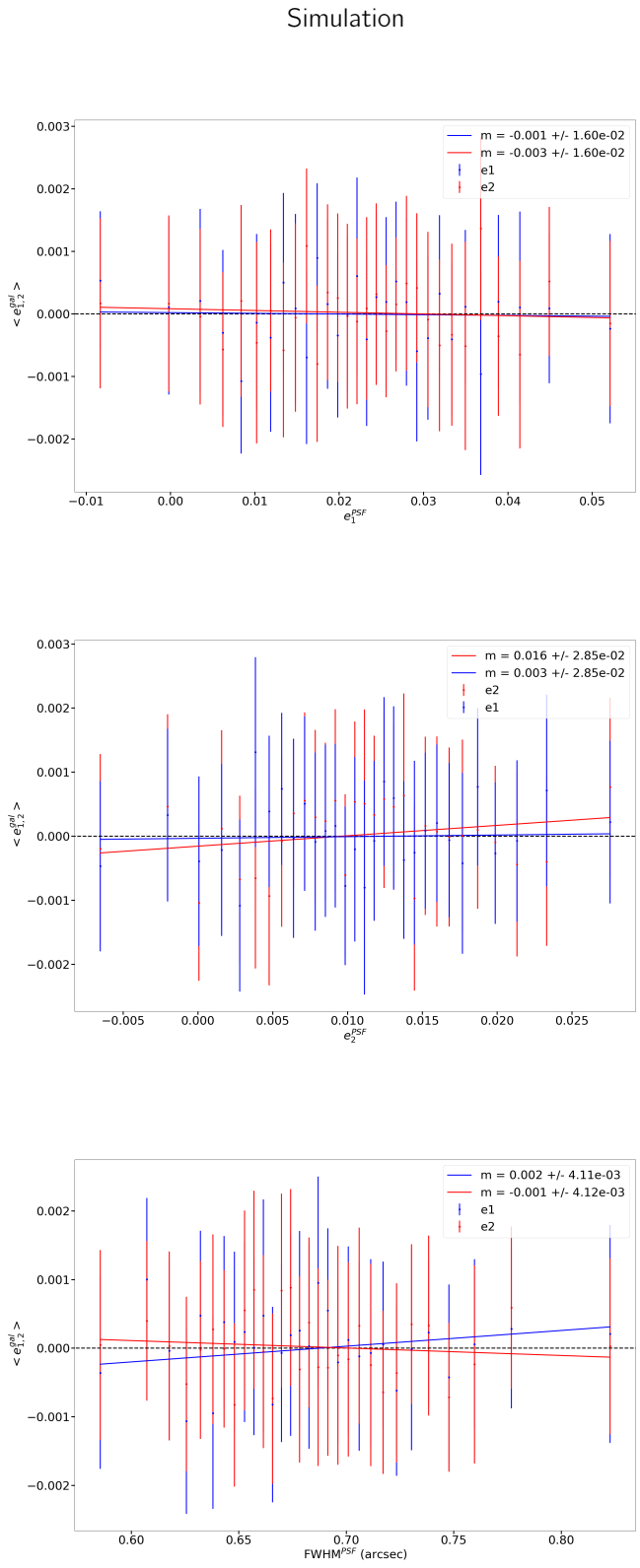


Figure 4.11: PSF leakage using the average galaxy shape in bins of true PSF ellipticity (top two panels) and size (bottom panel) on the simulated data.

4.3.2 PSF systematics

For the second test, I use the ellipticity of the PSF at the positions of test stars. The idea of this test is to compare the amplitude of this cross-correlation to the shear-shear correlation, the actual signal I want to use for the science analysis. I use the estimator $\xi_{\pm}^{\rm sys}(\theta)$ (Bacon et al. 2003) defined as:

$$\xi_{\pm}^{sys}(\theta) = \frac{\xi_{\pm}^{\rm gp^2}(\theta)}{\xi_{\pm}^{\rm pp}(\theta)}, \tag{4.10}$$

where $\xi_{\pm}^{\rm gp}(\theta)$ is the cross-correlation between star and the galaxy ellipticities, and $\xi_{\pm}^{\rm pp}(\theta)$ is the auto-correlation of star ellipticities.

In figure 4.12 I compare $\xi_{\pm}^{\rm sys}(\theta)$ to the theoretical value of the shear-shear correlation $\xi_{\pm}^{\rm ss}(\theta)$ for a flat ΛCDM model using a Planck cosmology. To have a better view of the impact on cosmology, figure 4.13 shows the ratio $\xi_{\pm}^{\rm sys}(\theta)/\xi_{\pm}^{\rm ss}(\theta)$. One sees that on small scales the contribution is less than a percent, and increases towards larger scales. These plots are presented in appendix A for the patches P1, P2 and P4.

I also estimate the leakage α using the cross-correlation. This has been introduced in Jarvis et al. 2016 as:

$$\alpha(\theta) = \frac{\xi_{+}^{\rm gp}(\theta) - \langle e_{\rm gal}\rangle^{*}\langle e_{\rm PSF}\rangle}{\xi_{+}^{\rm pp}(\theta) - |\langle e_{\rm PSF}\rangle|^2}. \tag{4.11}$$

In figure 4.14, I show the evolution of α as a function of distance. Similar to $\xi_{\pm}^{\rm sys}(\theta)$, the leakage degrades as the scale increases. The average values are around 2% for ngmix and 3% for Galsim, which is in good agreement with the results from 4.3.1 (the other patches are shown in the appendix A).

If the leakage is actually due to PSF modeling errors, one can explain the degradation on large scales by the large-scale patterns observed on the residuals in figure 4.2. I will explore this hypothesis further in my future analyses.

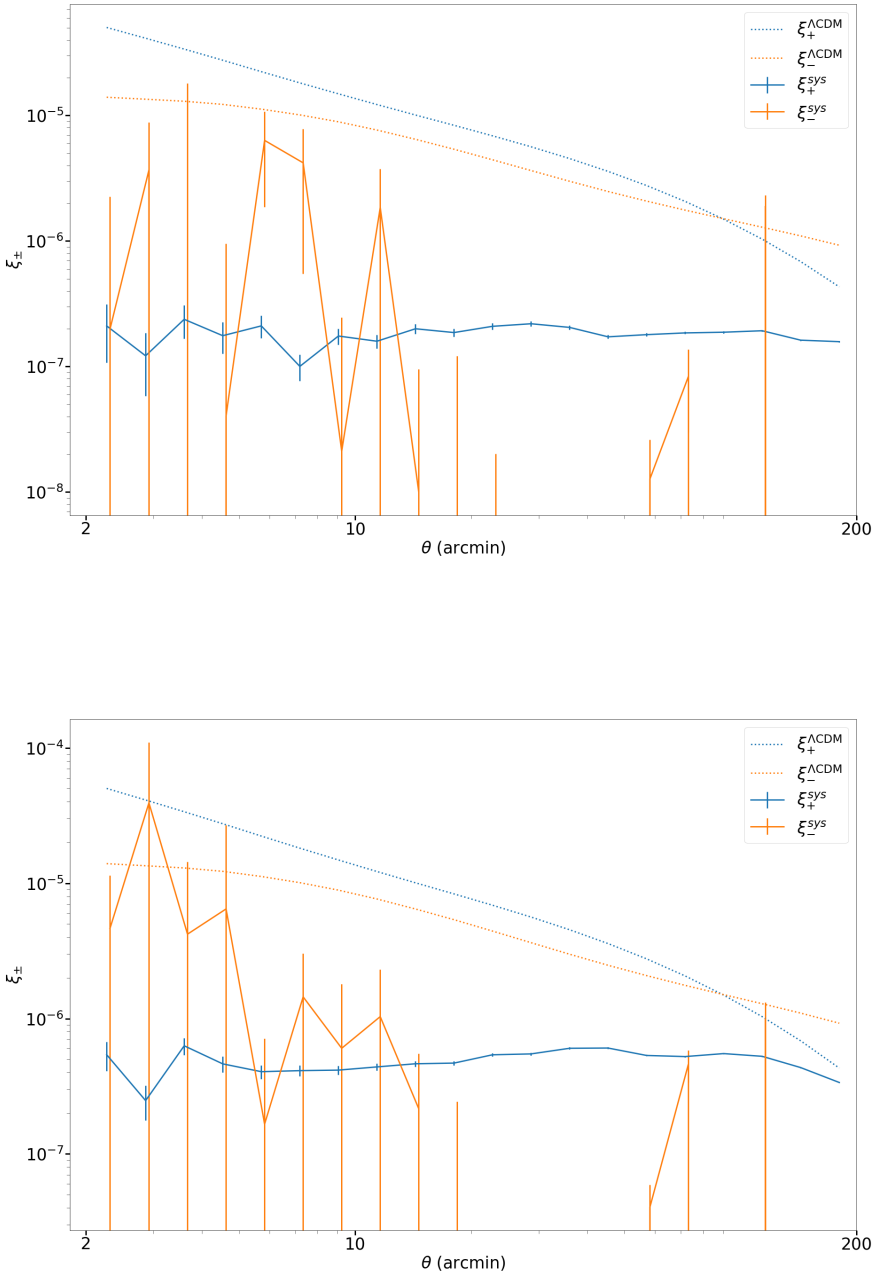


Figure 4.12: $\xi_{\pm}^{\rm sys}(\theta)$ measured for ngmix (*Top*) and Galsim (*Bottom*).

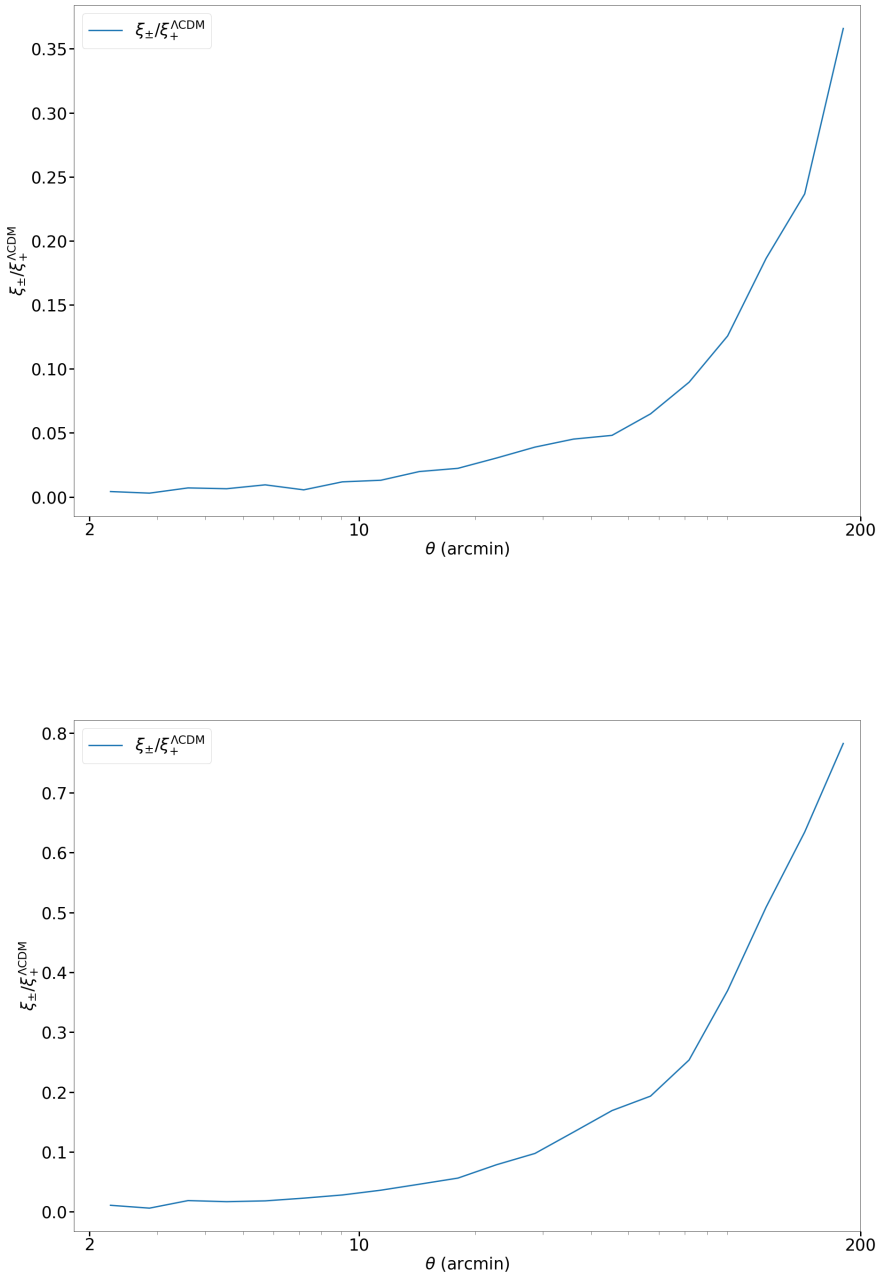


Figure 4.13: Ratio of $\xi_+^{\rm sys}(\theta)/\xi_+^{\rm ss}(\theta)$ for ngmix (*Top*) and Galsim (*Bottom*). For Galsim the systematics are larger at smaller scales compared to ngmix.

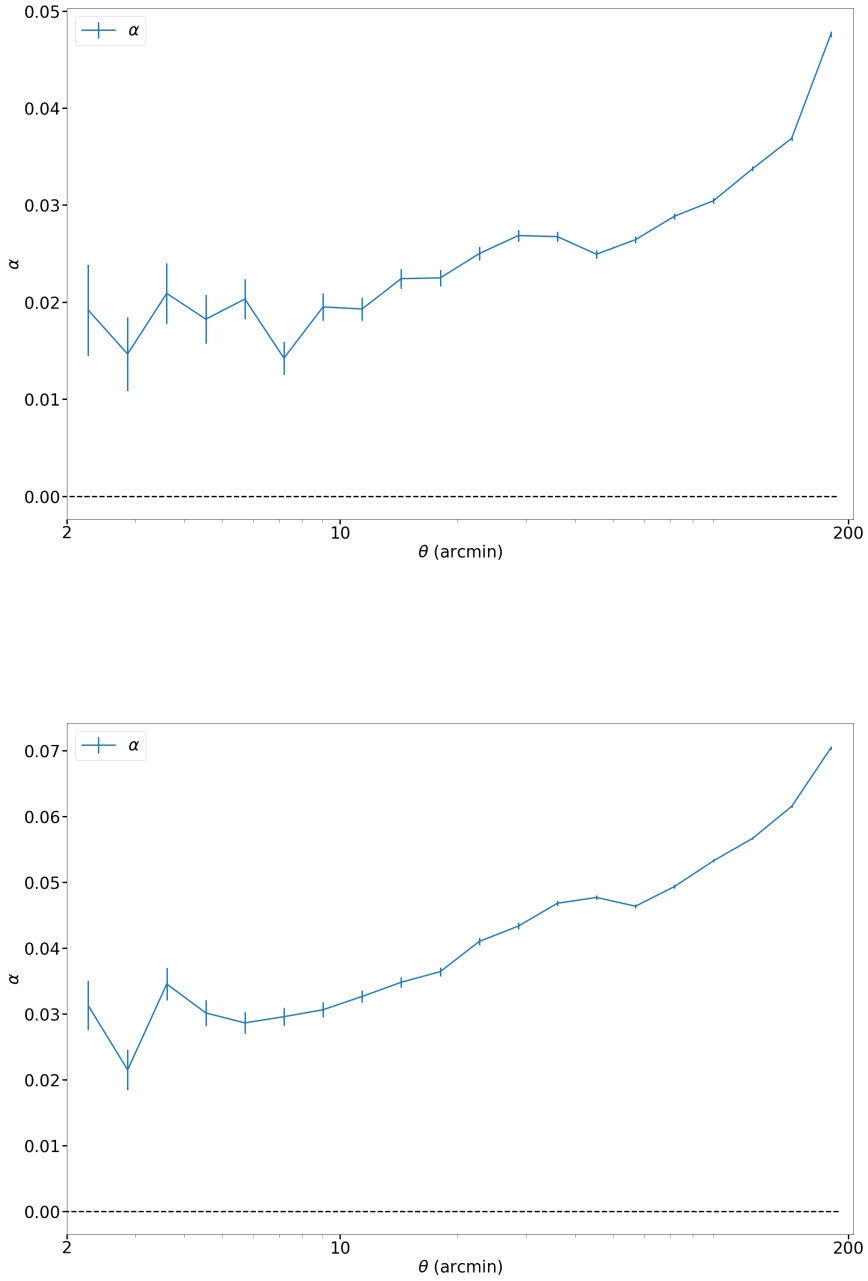


Figure 4.14: Leakage α shown as a function of scale θ for ngmix (*Top*) and Galsim (*Bottom*).

Chapter 5

Cosmological results

For my science results, I have studied four patches of 806, 282, 249 and 358 square degrees, corresponding to P1, P2, P3, and P4, respectively. These patches are presented in figure 5.1. They represent around 5,200 single exposures and 6,800 tiles, with a total area of $1,695$ square degrees. The processing has been carried out on the high-performance computing clusters CANDIDE[1] and CANFAR[2].

The amount of data processed is considerable. All required input images for the nearly $1,700$ square degrees (tile images and weights, single-exposure images, weight, and flag maps) add to several terabytes of disk space. To this, one has to add the intermediate data and images created by the pipeline, which are for example PSF files, star catalogues for masking, or metacalibration postage stamps. These data would take up several hundreds of terabytes in total, which makes the processing and storage very challenging, in particular on multi-core processors.

The processing time for a single tile from masking to shape measurements takes around 30-40 hours on a typical computing cluster node. This time is dominated by shape measurements and metacalibration, which uses 90% of the total time. Even with multi-core processing, the total run time can easily be several months, which makes testing extremely cumbersome. In the following, I describe how I solved these technical computing challenges.

CANFAR allowed me to perform the processing in parallel, on 200 virtual machines at a time. Each virtual machine behaves as a stand-alone computer, which processes one tile at a time, from masking to shape measurements, including PSF modeling on single exposures. Shape measurements were performed in parallel using all 8 CPUs per virtual machine, by splitting up the galaxy sample into 8 subsamples. Note that each galaxy can be processed independently including added shear versions for metacalibration. The calibration of the entire patch is done in a post-processing step, see below. This reduced the end-to-end processing time to around five hours per tile, adding up to a total wall-clock time of approximately two weeks with overheads.

The independent processing of individual tiles has further advantages. First, it keeps the memory per computing node relatively low. If a tile fails, I can easily identify it and reprocess it without having to redo the entire patch. Due to the survey strategy, a single exposure is typically used for several tiles, therefore I have some repeated processing of CCDs up to the PSF modeling. This is however not the bottleneck in terms of computing time compared to shape measurements.

Once a virtual machine is launched, the required images are downloaded via virtual observatory software (VOSpace[3]). The download speed is very fast, since both images and processing are hosted on the CANFAR platform. Once the processing is succeful, the resulting catalogues and meta-data (for example log files, PSF, statistics) are uploaded to

[1] https://candideusers.calet.org/
[2] https://www.canfar.net/en/
[3] http://www.ivoa.net/documents/VOSpace/

VOSpace, from where I retrieve it for post-processing. The virtual machine is deleted, and no temporary data is stored after processing has finished.

For each patch, I download the resulting catalogues and create a merged main catalogue. The file size for the largest patch is 50 Gb, which are at the limit of being read using standard python libraries such as numpy.

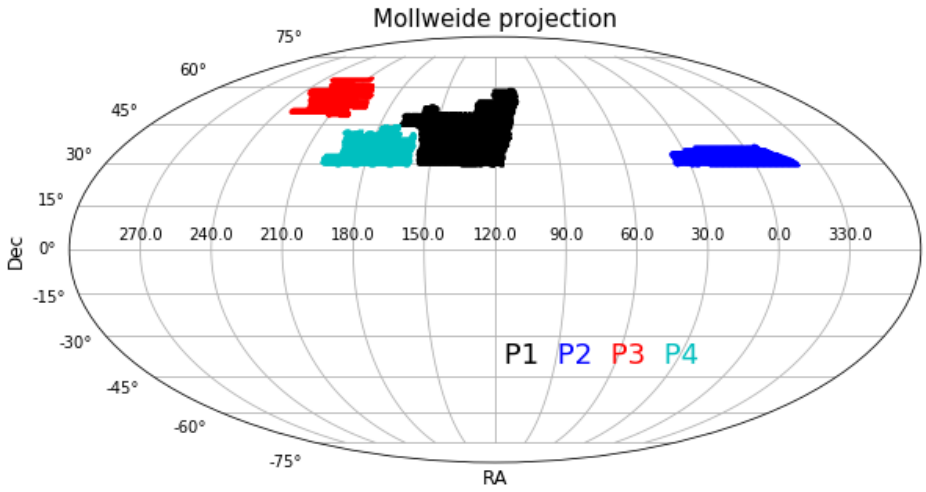


Figure 5.1: Process patches of the CFIS dataset.

The data I used here have been calibrated with metacalibration for each patch independently. In figure 5.2 I show the diagonal terms of the response matrix for both galaxies and stars in patch P3 (I find similar results in the other patches). The stars are expected to have a zero response, here one sees a scatter which could be due to a bad deconvolution or the presence of other objects in the sample. Similar results have been found in DES Y1 (Zuntz et al. 2018). The averaged total response for galaxies in this patch is (the other patches are given in appendix B):

$$R^{\text{total}} = \begin{pmatrix} 0.66 & 5.11 \times 10^{-5} \\ 7.3 \times 10^{-4} & 0.64 \end{pmatrix}, \tag{5.1}$$

while the response for selection effects is:

$$R^{\text{S}} = \begin{pmatrix} -0.148 & 0. \\ 0. & -0.153 \end{pmatrix}. \tag{5.2}$$

I present the magnitude distribution for P3 in figure 5.3. As mentioned in the previous chapter, bright galaxies with $r < 20$ are cut from the final sample. The mode of the distribution peak around 23.7, which is consistent with expectations corresponding to the survey strategy. I found an average galaxy density between 6 and 7 galaxies.arcmin^{-2}.

In the following, I present first science results from the shear catalogue derived with my pipeline. I will show tangential shear measured around known galaxy clusters together with mass estimates of stacked cluster samples binned in mass and redshift. I also plot convergence maps that relate to the projected dark matter distribution. Finally, I have inferred parameter constraints on the field W3. This field of 55 square degree is a sub-field of the patch P3, and corresponds to the CFHTLS-W3 (Erben et al. 2013). I compute

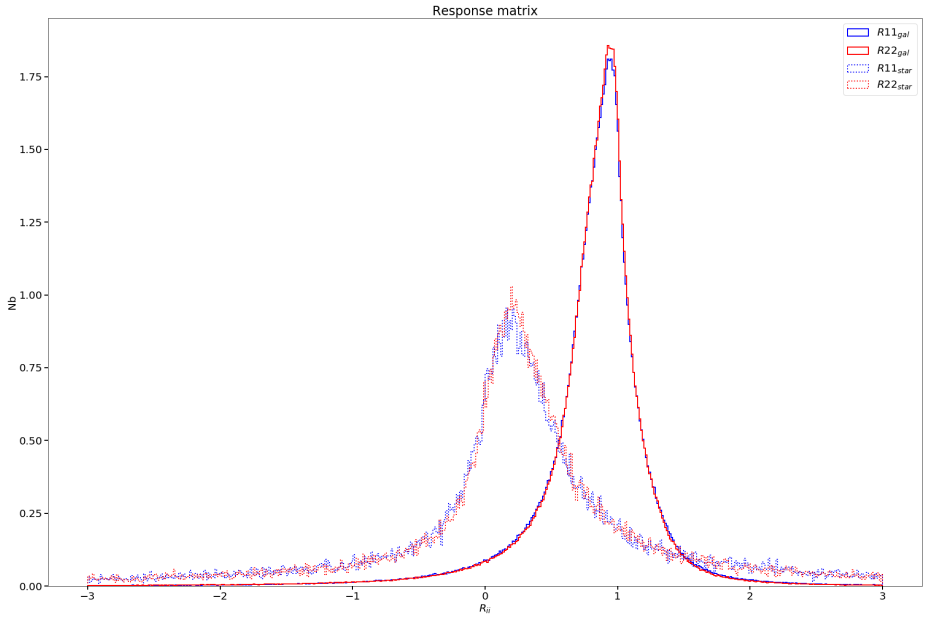


Figure 5.2: Diagonal terms of the response martrix for the patch P3. The *solid line* for the galaxies and the *dashed line* for the stars.

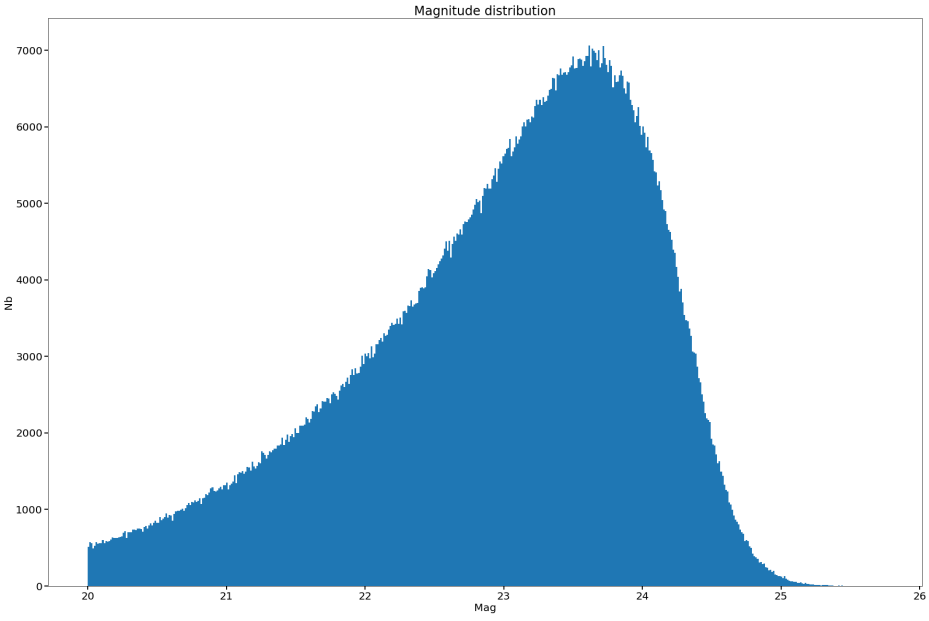


Figure 5.3: Magnitude distribution for the patch P3.

the "3×2"-point correlations functions, using the correlations of our CFIS lensing sample and a galaxy population from BOSS. The three correlation functions correspond to shear-shear, shear-position, and position-position, respectively. I ran this analysis using shear-shear alone, and using the three correlation functions including the covariance, and estimate the parameters σ_8 and Ω_m for a flat ΛCDM model.

5.1 Redshift distribution of lensed galaxies

To estimate cluster masses from weak lensing and to perform the 3×2-point analysis, I need a redshift estimate for the source galaxies, or at least the number density as a function of redshift, $n(z)$, for the lensing sample. At the moment, I do not have such information from CFIS data only. There is an ongoing effort to obtain redshifts with the so-called cluster-z technique as presented in Ménard et al. 2013. In this work, I have derived the $n(z)$ distribution from the CFHTLenS study. I match our galaxies to the ones detected in CFHTLenS in the W3 region. This matching is done at 0.3 arcsec, roughly 2 pixels. I matched around 700,000 galaxies with CFHTLenS. I then reconstruct the $n(z)$ from the photometric redshift measurements in CFHTLenS, using the best-fit value (z_B). This method is far from being perfect, but it gives a first estimate of the galaxy distribution. The derived $n(z)$ is presented in figure 5.4.

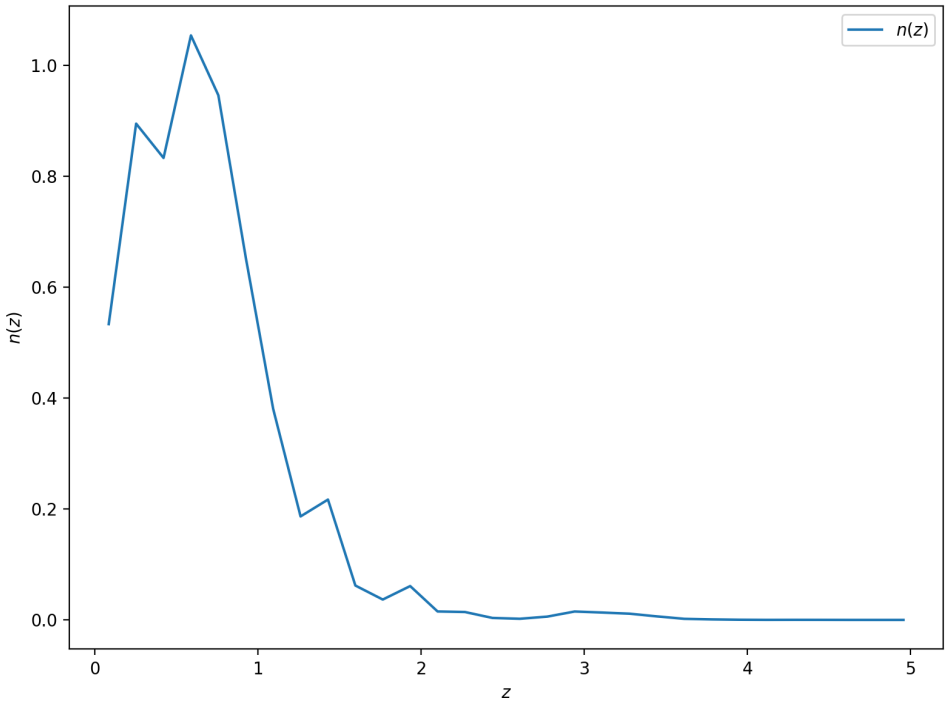


Figure 5.4: Galaxy distribution of matched galaxies derived from the redshifts measured in CFHTLenS.

5.2 Cluster lensing

As described in the introduction, when the light of a galaxy goes through a high-density region, due to the gravitational lensing effect the galaxy image is on average tangentially aligned around the dense region. In this section, I present results from tangential shear measurements around known clusters from the Planck experiment (Ade et al. 2016) detected via the Sunyaev-Zeldovich (SZ) effect (Sunyaev and Zeldovich 1970, 1980).

5.2.1 Theoretical model

To estimate the masses of clusters, one has to assume a theoretical model of their mass distribution. Here, I use the Navarro-Frenk-White (NFW) profile (Navarro, Frenk, and White 1996). This profile can be written as:

$$\rho(r) = \frac{\rho_s}{\frac{r}{R_s}\left(1+\frac{r}{R_s}\right)^2}, \tag{5.3}$$

where $\rho(r)$ gives the density of the dark matter halo, R_s is the scale radius and ρ_s the density at that scale. Usually, people define the concentration of a halo c_Δ as: where $\rho(r)$ indicates the density of the dark-matter halo, R_s is the scale radius, and ρ_s the density at that scale. Usually, one defines the concentration of a halo c_Δ as:

$$c_\Delta = \frac{R_\Delta}{R_s}, \tag{5.4}$$

where Δ is the virial overdensity. For this study I used $\Delta = 500$. Integrating the density profile over the volume of the cluster out to some cut-off size provides the cluster mass.

One can also define the gravitational potential by solving the Poisson equation 1.23, which leads to:

$$\Phi(r) = -\frac{4\pi G \rho_s R_s^2}{r} \ln\left(1+\frac{r}{R_s}\right). \tag{5.5}$$

Then, using equations 1.34 and 1.41, provides us with the tangential shear expected for a given cluster, see Takada and Jain 2003 for the explicit expression.

5.2.2 Mass estimation

To estimate the mean mass of the clusters I fit an NFW profile to the measured stacked γ_t, with the cluster mass as the only free parameter. The cosmology has been fixed to a flat ΛCDM model with Planck parameters. The theoretical model is computed through the CLMM software package[4]. The mass-concentration relation is given by Groener, Goldberg, and Sereno 2015, from which I used the X-ray best-fit values from their table 2.

Clusters are selected in three different bins in mass and redshift. I base this selection on figure 5.5, which shows the number of clusters as a function of mass and redshift. The cuts applied are presented in table 5.1. I choose the three bins with the highest number of clusters. I have not selected clusters above redshift 0.4 because the number of background sources would be too low to get a significant signal.

For each bin, I measure the stacked shear profile of the clusters to improve the SNR and fit the mass. The results are presented in table 5.2, in comparison with measurements from the SZ signal. One can see that for lower-mass clusters I get a good agreement between

[4] https://github.com/LSSTDESC/CLMM

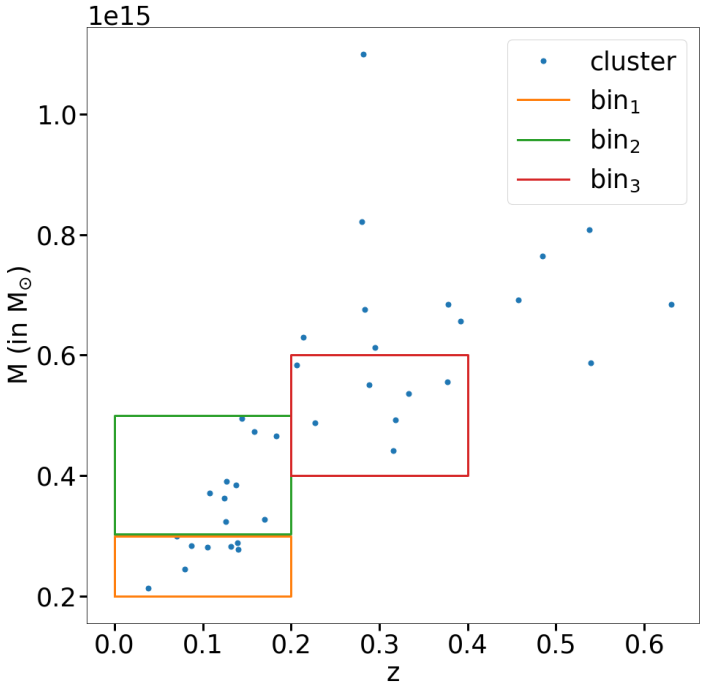


Figure 5.5: Number of clusters as function of mass (SZ) and redshift.

Bin name	z^{clust}	$M_{500}^{\text{clust}}(10^{14}M_{\odot})$	n^{clust}
bin_1	[0; 0.2]	[2; 3]	8
bin_2	[0; 0.2]	[3; 5]	9
bin_3	[0.2; 0.4]	[4; 6]	7

Table 5.1: Cluster selection bins.

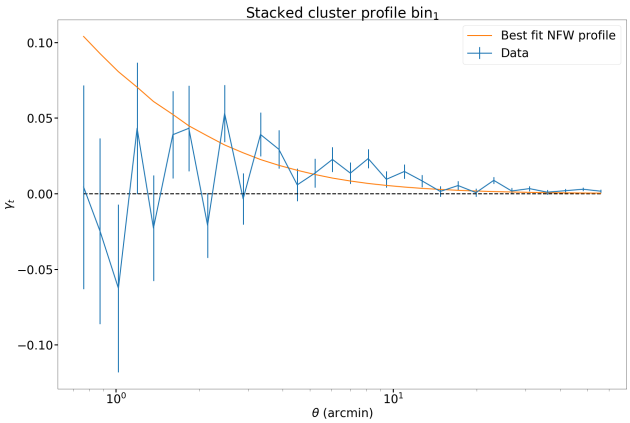
the two estimates. It is important to note that error bars only account for statistical errors due to the fit. These errors are under-estimated, since I did not include the photometric redshift errors or the residual shear bias. I also suspect that the differences observed with M_{500}^{SZ} are mainly driven by the fact that I do not have individual redshifts for source galaxies and then I cannot cut galaxies that are in front of the cluster. The best fit of the NFW profile for each bin is presented in figure 5.6.

5.3 Mass mapping

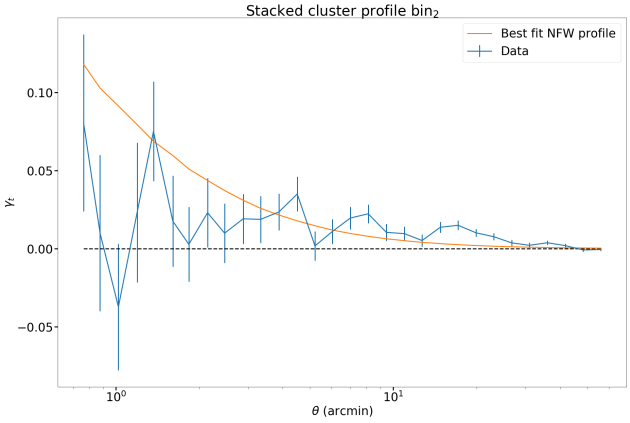
As mentioned in the introduction, weak-lensing is a tracer of the dark matter distribution in the Universe. To visualize this spatial distribution in 2D, one can draw convergence maps, usually called mass maps. Those maps are directly proportional to the density of matter along the line-of-sight. To create these map, I perform an inversion of a shear map in Fourier space. I used the method presented in Kaiser and Squires 1993 and implemented in the LensPack software package[5].

[5]https://github.com/CosmoStat/lenspack

bin$_1$



bin$_2$



bin$_3$

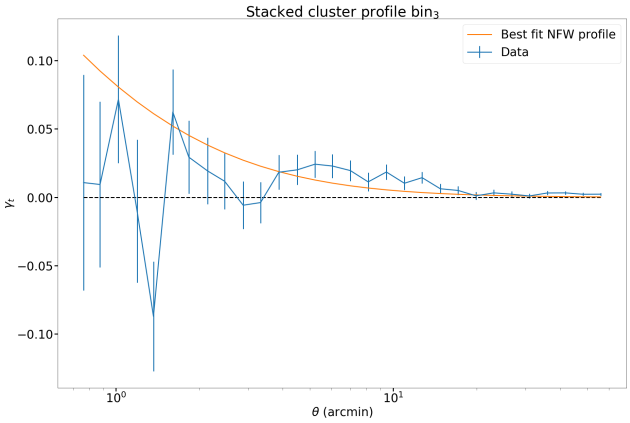


Figure 5.6: Stacked shear profiles using CFIS weak-lensing data (blue curves with statistical error bars), and best-fit NFW model. The three panels show the three bins, see Table 5.1.

Bin name	$M_{500}^{WL}(10^{14}M_{\odot})$	$M_{500}^{SZ}(10^{14}M_{\odot})$
bin_1	2.56 ± 0.06	$2.7^{+0.22}_{-0.22}$
bin_2	3.11 ± 0.12	$3.99^{+0.3}_{-0.32}$
bin_3	2.57 ± 0.11	$5.21^{+0.46}_{-0.49}$

Table 5.2: Average mass of the clusters for each bin.

Patch name	$n_{cluster}$
P1	12
P2	3
P3	12
P4	5

Table 5.3: Number of clusters in each patch.

Since mass maps are proportional to the projected matter density, one expects to see an excess of signal at the position of over-densities. In the maps presented here, I show the positions of all overlapping clusters from Ade et al. 2016, which were also used in the previous section. The maps are created independently for each of the 4 patches to limit projection effects and discontinuities.

Mass maps are presented in figure 5.7 for P1, figure 5.8 for P2, figure 5.9 for P3, and figure 5.10 for P4. Figure 5.11 presents the stacked signal at the position of each cluster for the four patches. The number of clusters for each patch is given in table 5.3. One can see that there is a clear signal in the E-mode in most cases, which correlates with the presence of an over-density. In comparison, the B-mode looks homogeneous which indicates that I do not have dominant systematics in the data. For the patch P2, I do not see any signal at the cluster positions, even on the stacked mass map. One can speculate that this might be a projection problem due the large extend of this patch in Right Ascension (RA) of over 60 deg. To resolve this issue, I will need to perform the shear-to-convergence transformation on the sphere.

5.4 Parameter inference

In this section, I present a first cosmological analysis done with the CFIS data. One of the major concerns for this kind of work is blinding, which is now routinely done in weak-lensing surveys to reduce observer biases. Blinding is however not done for the analysis presented here. Yet, I have limited my analysis to a small part of the survey to not compromise future work. The data used here are in the W3 region defined in CFHTLenS covering an effective area of 50.5 deg^2. I have measured the shear-shear correlation function from the processed CFIS data with the pipeline described in this manuscript. For galaxy clustering, I use a Luminous Red Galaxy (LRG) sample from SDSS DR12 LOWZ (Reid et al. 2016). This sample has been chosen because it provides massive galaxies at low redshift, which will maximize the galaxy-lensing signal. Those three probes (shear-shear, shear-position, and position-position correlations) are combined together for a 3×2-point analysis.

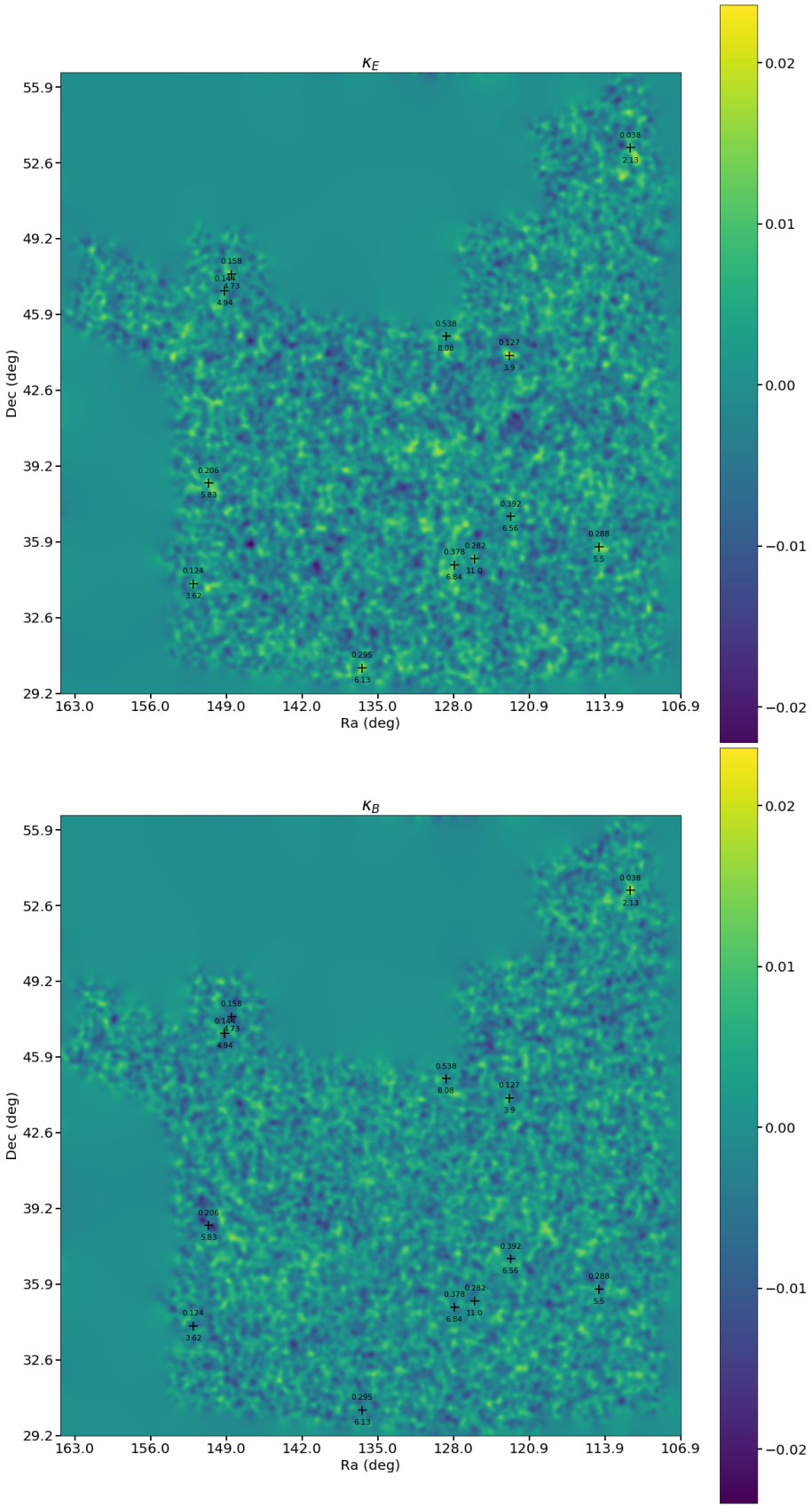


Figure 5.7: Mass map for the patch P1. The *black crosses* represent the positions of Planck clusters. The value on *top* of each cross is the cluster redshift, and the *bottom* value indicates the cluster mass ($10^{14}M_{\odot}$). The *top* (*bottom*) panel shows the E-mode (B-mode).

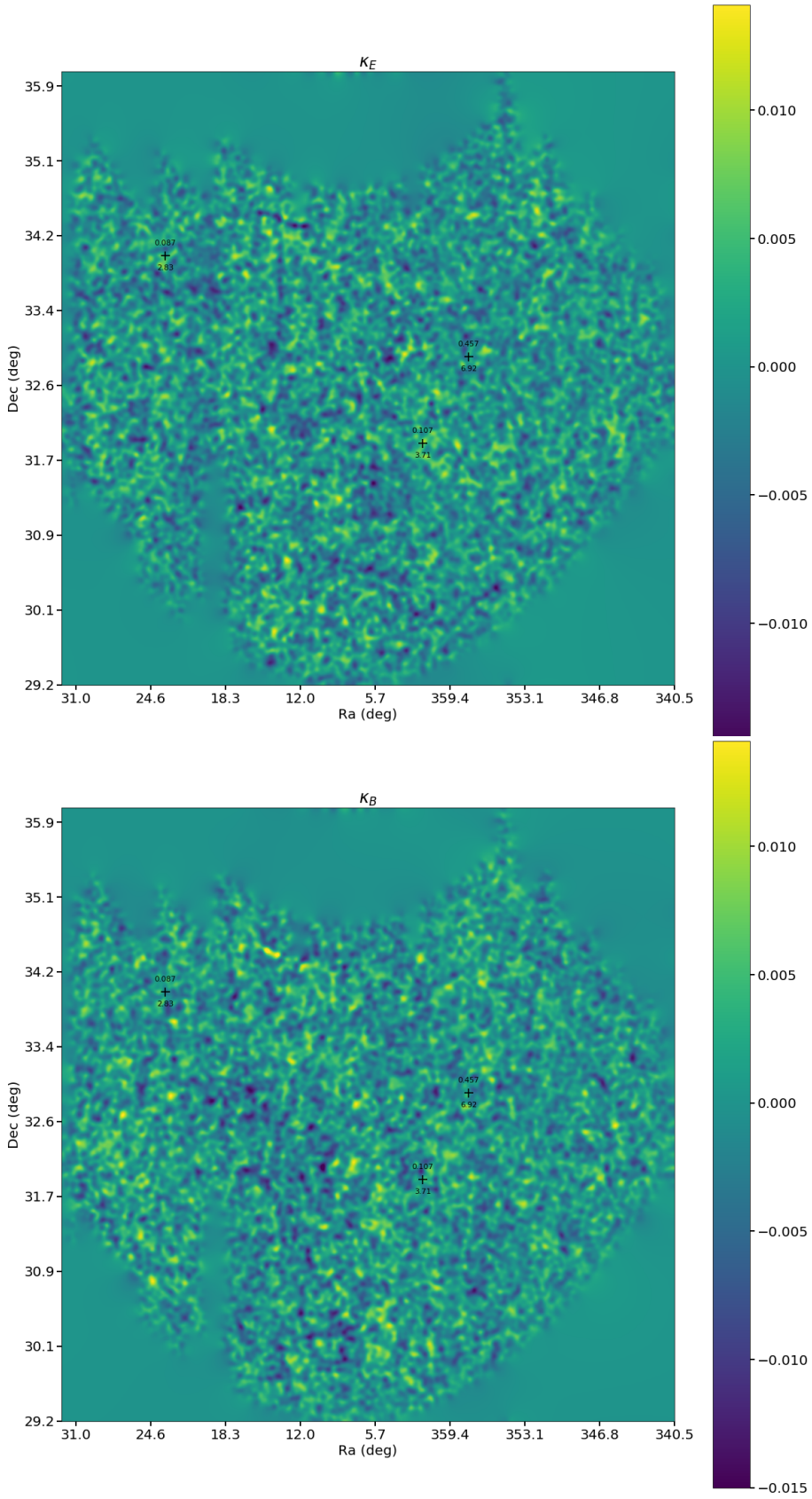


Figure 5.8: Mass map for the patch P2. The *black crosses* represent the positions of Planck clusters. The value on *top* of each cross is the cluster redshift, and the *bottom* value indicates the cluster mass ($10^{14}M_{\odot}$). The *top* (*bottom*) panel shows the E-mode (B-mode).

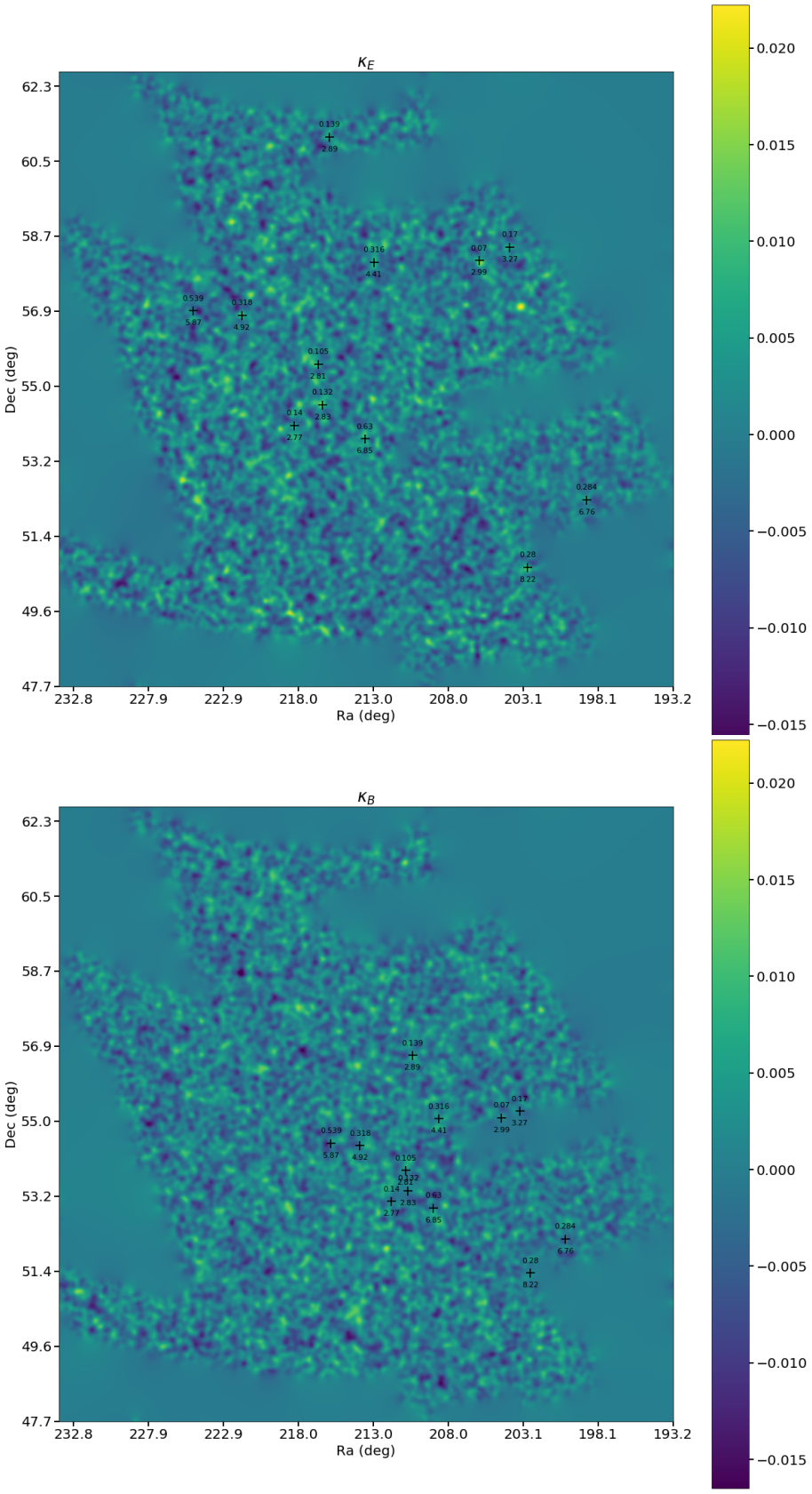


Figure 5.9: Mass map for the patch P3. The *black crosses* represent the positions of Planck clusters. The value on *top* of each cross is the cluster redshift, and the *bottom* value indicates the cluster mass ($10^{14}M_{\odot}$). The *top* (*bottom*) panel shows the E-mode (B-mode).

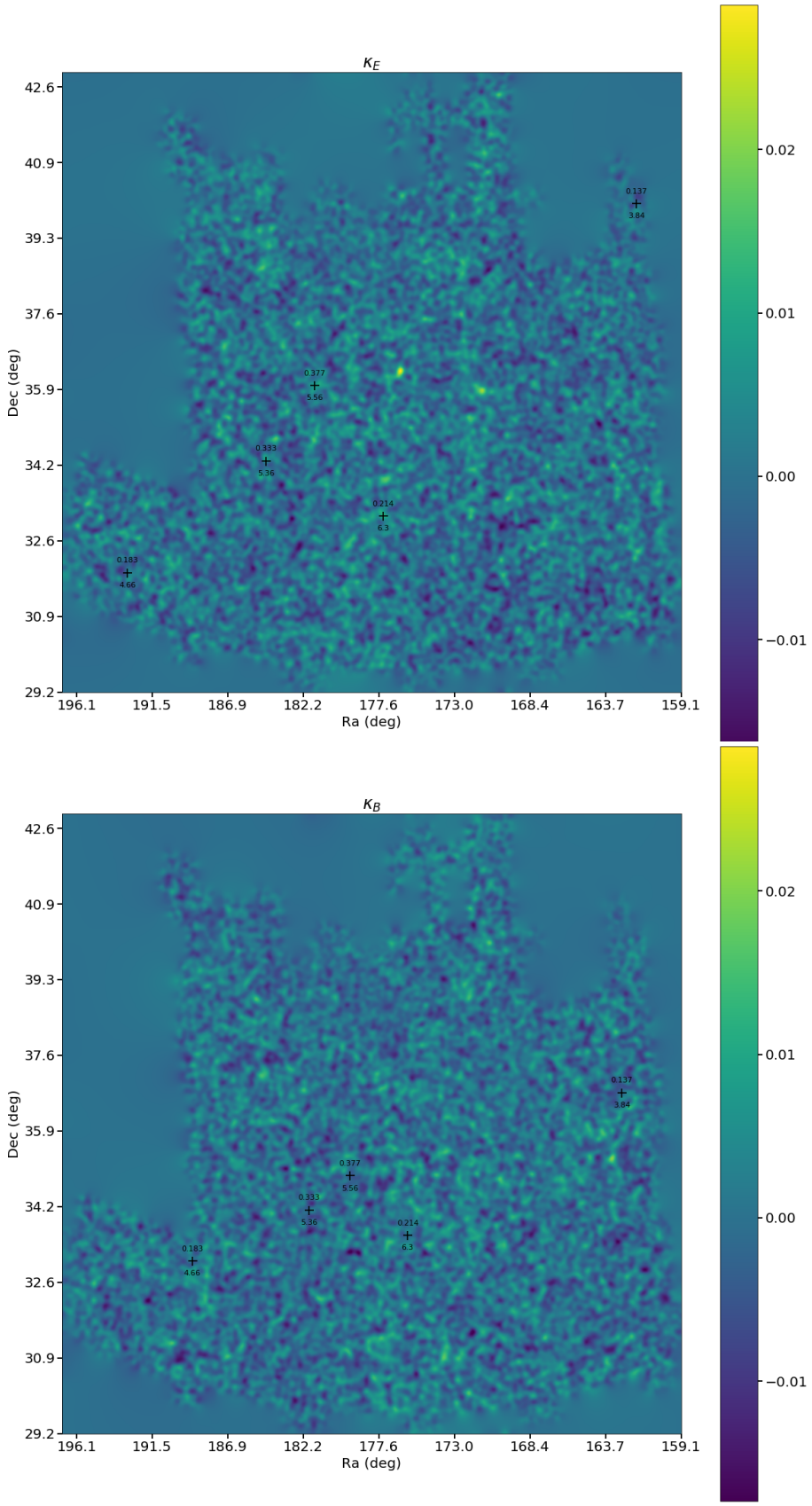


Figure 5.10: Mass map for the patch P4. The *black crosses* represent the positions of Planck clusters. The value on *top* of each cross is the cluster redshift, and the *bottom* value indicates the cluster mass ($10^{14}M_{\odot}$). The *top* (*bottom*) panel shows the E-mode (B-mode).

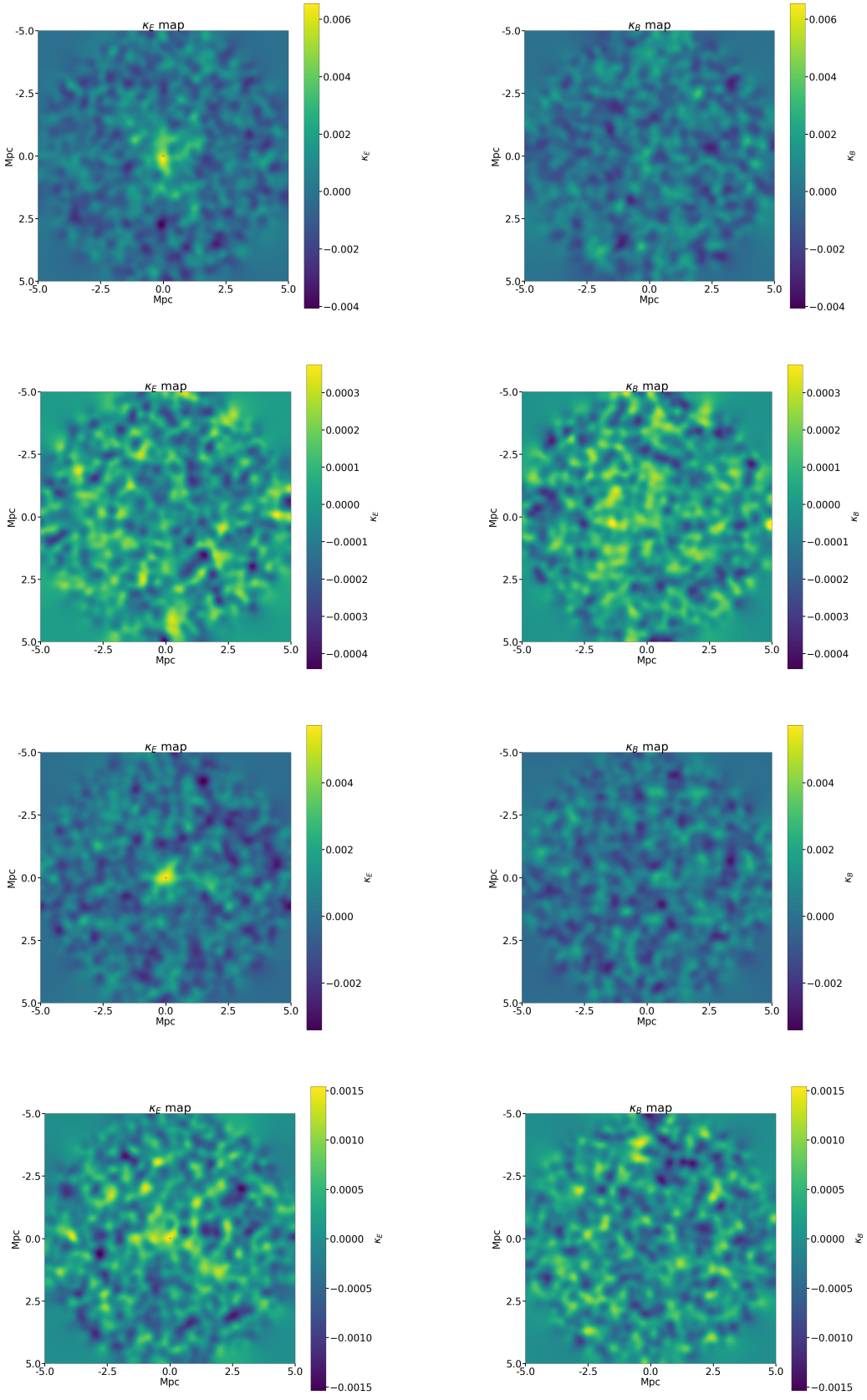


Figure 5.11: Mass maps stacked at the Planck cluster positions for P1, P2, P3 and P4 from *top* to *bottom*. Galaxies for the tangential shear stacks are selected in a radius of 5 Mpc around each cluster. This distance is computed at the cluster redshift.

5.4.1 Test on a simulation

Before running the analysis on real data, I have performed a test on a simulation. For this test, I used MICE v2 (Fosalba et al. 2015; Crocce et al. 2015; Fosalba et al. 2014; Carretero et al. 2014; Hoffmann et al. 2015). I created a field of 49 deg^2 with one million galaxies which gives 5.9 galaxies.arcmin^{-2}. I select an LRG-like sample using the following criteria:

- $M_r < -21$. With M_r the absolute magnitude in the r-band;
- $(g-r) > 0.8$. With g and r the observed magnitude in the g- and r-band respectively;
- $0 < z < 0.45$. A similar redshift range as the one used by our real LRG sample.

The LRG sample has a density of 0.027 galaxies.arcmin^{-2}. To reflect the real data, I added an artificial shape noise to shear value of the lensing sample. The values have been drawn from a normal distribution with zero mean and $\sigma = 0.3$.

The analysis is identical to the one performed wit the real data, and described in more details below. Results are presented in figure 5.12. One can see that the true value (input of the simulation) are within 1-sigma. The errors are thought to be mainly due to the small chosen area. Yet, this test probes that the algorithm works and is ready to move to real observations.

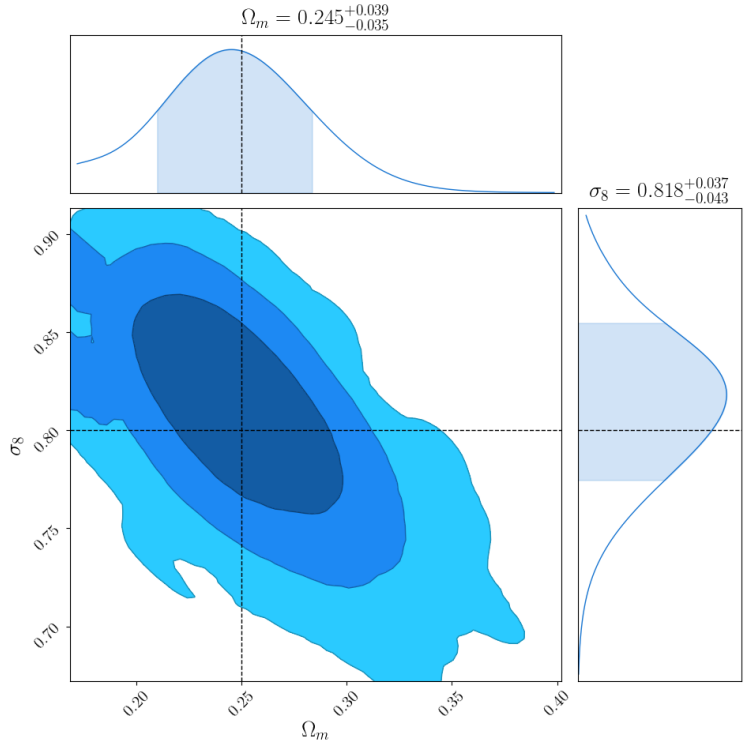


Figure 5.12: Posterior for the 3×2-point analysis on MICE simulation. The contours represent 1-, 2- and 3-σ errors. *Dashed* lines show the input values of the simulation, $\Omega_m = 0.25$ and $\sigma_8 = 0.8$.

5.4.2 Setup

Lensing data

There are 1.1 million source galaxies in W3, corresponding to a density of 6.06 galaxies.arcmin^{-2}. The redshift distribution is the one presented in figure 5.4. Shear values have been calibrated using the metacalibration technique presented in 3.4. The shape noise measured in the data is $\sigma_\epsilon = 0.31$ per ellipticity component.

LRG sample

There are 2031 LRG galaxies in W3, corresponding to a density of 0.0112 galaxies.arcmin^{-2}. The redshift distribution of this sample is presented in figure 5.13. For the measurement of the clustering, I used the provided random catalogues appropriately masked to fit the geometry of the lensing sample. I used a linear galaxy bias fixed to 2 (Reid et al. 2016).

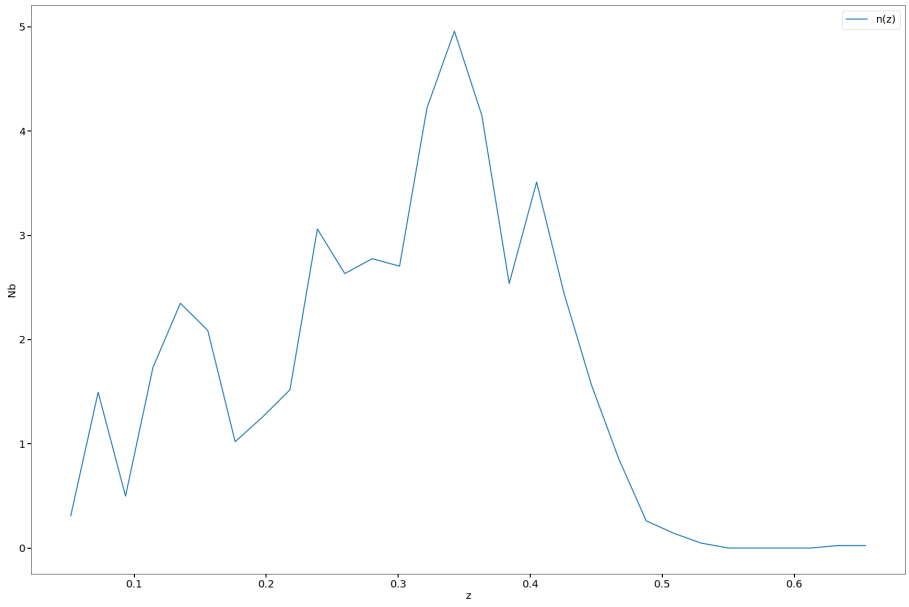


Figure 5.13: Redshift distribution of the LRG sample.

Covariance matrix

For this analysis, I use a covariance matrix derived from theory. I employ the software package CosmoCov[6]. The covariance does not include non-Gaussianities and intrinsic alignment. To compute the covariance, I assume a flat ΛCDM model with Planck parameters. In figure 5.14 I compare the diagonal of the computed covariance with the shot noise measured on data. One can see a good agreement.

[6]https://github.com/CosmoLike/CosmoCov

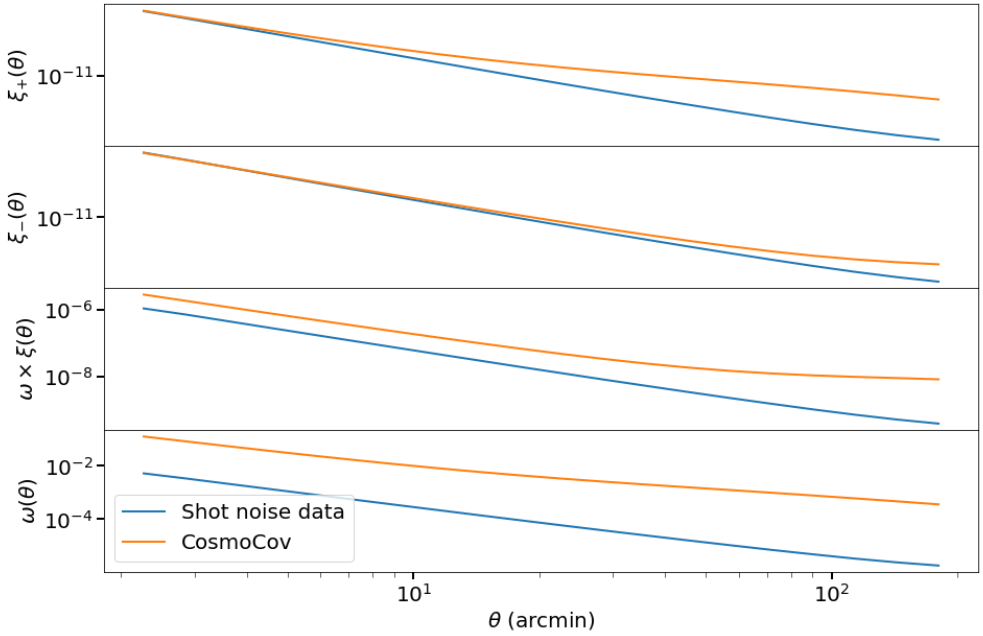


Figure 5.14: Comparison of the diagonal of the covariance from CosmoCov and the shot/shape noise measured on data.

Correlation functions and theoretical model

Correlation functions have been measured with the software package TreeCorr[7]. Error bars on the plots in figure 5.16 are only statistical errors due to shot noise and shape noise measured with TreeCorr.

Theoretical correlation functions are obtained by the CCL library. The transfer function comes from Eisenstein and Hu 1998, and the matter power spectrum is given by HaloFit (Smith et al. 2003). For efficiency reasons, I did not use a Boltzmann code, but restrict myself to a transfer function.

5.4.3 Results

The minimization is carried out by the emcee library[8] (Foreman-Mackey et al. 2013). I only vary the two parameters Ω_m and σ_8, and for now do not include nuisance parameters. The priors used are presented in table 5.4. I ran 16 walkers on 5,000 steps for the Markov Chain Monte Carlo (MCMC). I removed the first 500 steps to avoid problems due to initialization.

The purpose of this analysis is to give an idea of what is possible with this shape catalogue. I do not claim to provide trustworthy parameter constraints with a thorough error and systematics analyses. I compare our likelihood to other surveys to provide a reference to the reader.

[7] https://github.com/rmjarvis/TreeCorr
[8] https://github.com/dfm/emcee

Parameter	Prior
Ω_m	flat(0.06, 1)
σ_8	flat(0.4, 1.2)

Table 5.4: Priors used for the MCMC.

Figure 5.15 shows the results for the 3×2-point analysis. We found the following mean values for the parameters:

$$\Omega_m = 0.295^{+0.113}_{-0.094}; \tag{5.6}$$

$$\sigma_8 = 0.81^{+0.11}_{-0.13}, \tag{5.7}$$

errors are given at 1-sigma. The best-fit model of the correlation functions is shown in figure 5.16. Figure 5.17 shows the same posterior, with previous results from DES Y1 (Abbott et al. 2018) and Planck (Aghanim et al. 2020). One can see that the results form CFIS are consistant with Planck and DES Y1 which is encouraging for future analysis. Figure 5.18 shows the posterior for the shear-shear only compared with the 3×2-point.

To conclude this section, I present in figure 5.19 the shear-shear correlation function for the full processed area ($\approx$ 1,700 deg^2), to show the potential of the survey for future cosmological works.

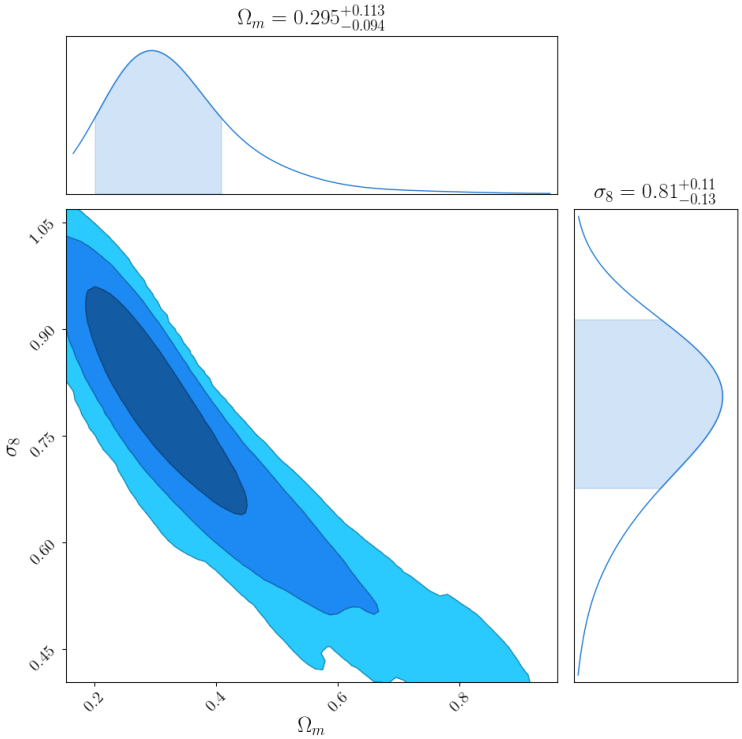


Figure 5.15: Posterior for the 3×2-point analysis. The contours represent 1-, 2- and 3-σ errors.

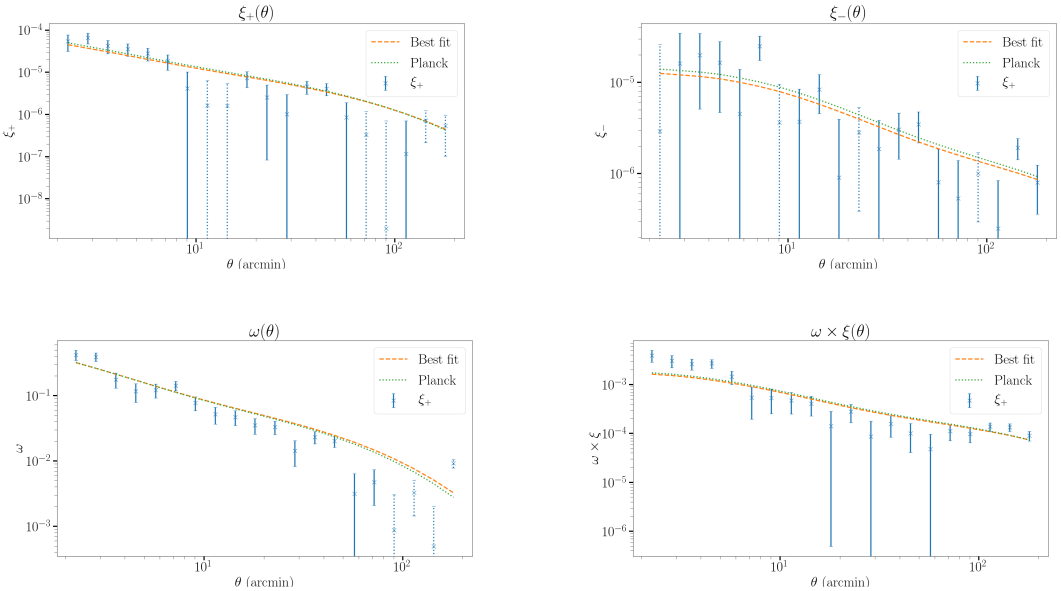


Figure 5.16: Best fit of the correlation functions (*orange*). In *green* I show the theoretical value for Planck cosmology.

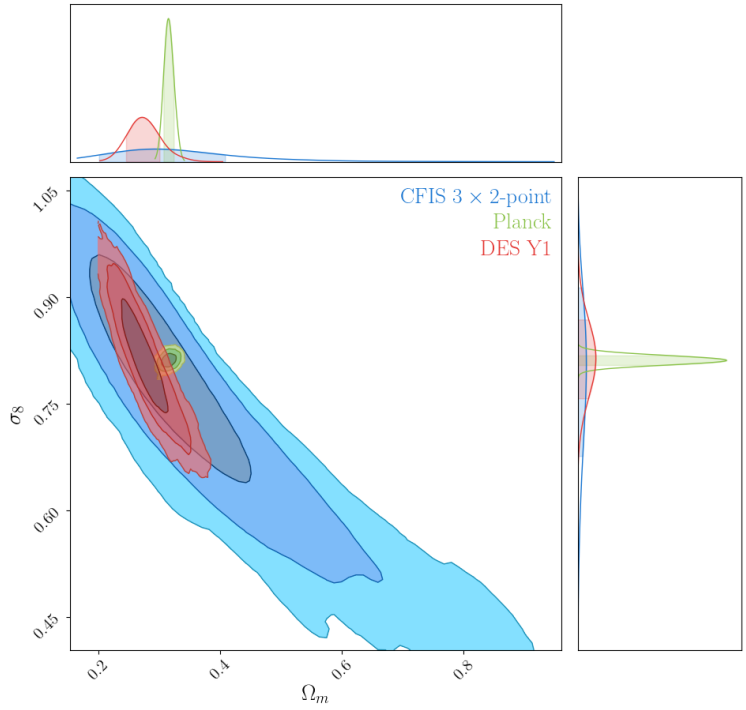


Figure 5.17: Posterior for the 3×2-point analysis in comparison with Planck and DES Y1. The contours represent 1-, 2- and 3-σ errors.

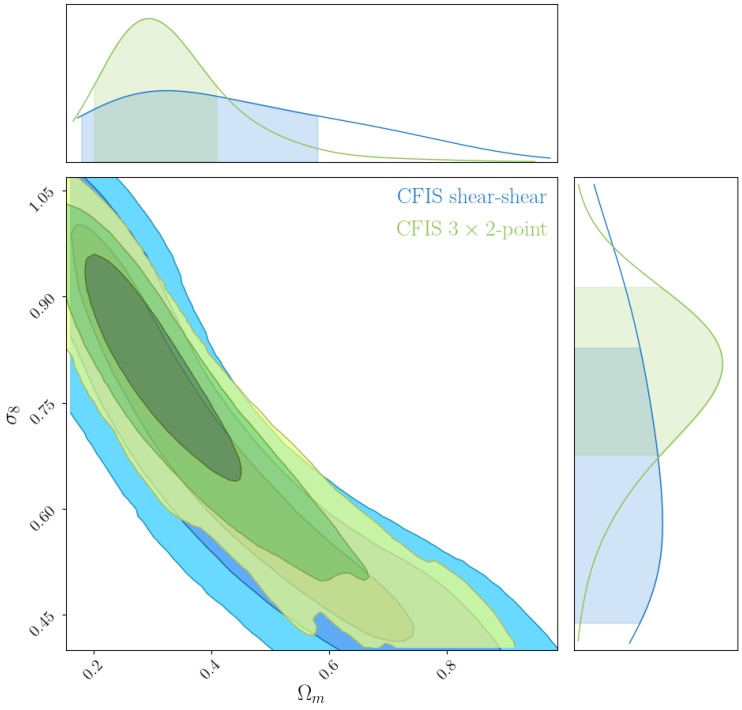


Figure 5.18: Posterior for the shear-shear only compared to the 3×2-point analysis. The contours represent 1-, 2- and 3-σ errors.

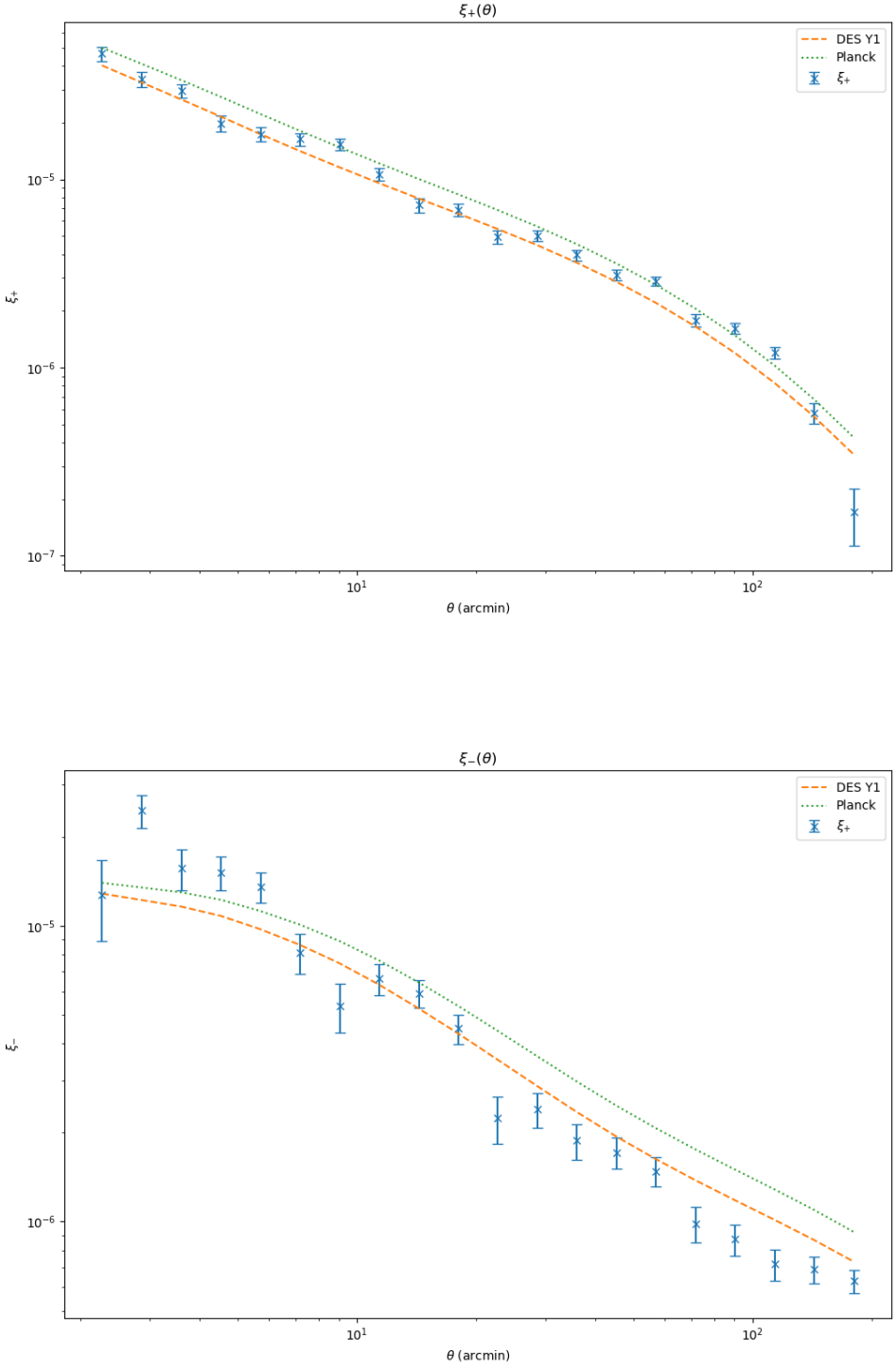


Figure 5.19: Shear-shear correlation function computed from the entire processed area of 1,695 deg^2. In *green* I show the theoretical value for the Planck cosmology. The (*orange*) curve represents the DES Y1 best-fit cosmology.

Chapter 6

Conclusion

In this manuscript I have presented a new competitive pipeline for weak lensing analysis. I gave an overview of the entire process required to go from the "pixels" to an estimation of cosmological parameters. I combined both well-established and state-of-the-art techniques in the processing. As our Universe, the pipeline I built will evolve in time.

First, I have presented how a survey like CFIS is planned to maximize the allocated time. We have to consider our surrounding, the Milky Way, which can be at the same time a source of scientific information and a source of contamination for extra-galactic work. Even the smallest objects of our Solar System can disturb such analysis. Once all those effects have been taken care of and the observations have started, one can focus on other sources of systematics. The pre-processing is a crucial steps of the processing chain. The photometry and astrometry calibrations have to be precisely controlled to reach the level of precision that is required in weak lensing analyses.

The next step is the selection of objects of interest. For either the star or the galaxy samples, the selection has to be as outlier-free as possible. The star sample used for the PSF plays a key role. Indeed, the PSF is one of the principal sources of systematics when it comes to shape measurement. As I have shown, the model has to be reliable at all scales. In this manuscript, I make use of widely used tools and techniques to reach the requirements.

The shape measurement methods presented here are state-of-the-art techniques. Despite the fact that they were used in previous works, they have to be adapted to each specific data set. This step has been one of the most challenging part of this work. I elaborated realistic simulations to replicate all sources of potential problems that can be caused by real data. I presented the importance of calibration of the shear bias. Linked to this issue, I contributed to a parallel work presented in Pujol et al. 2020, which uses a machine-learning approach to solve this problem.

The control of systematics is critical, and guided me in my work to improve the final results. This control and check of systematics have to be done at each major step of the processing chain. The validation of the PSF model was the first step to achieve before going further. The validation of the shape measurement was the final phase from the pipeline's point of view. The different tests that I implemented and carried out allowed me to identify where the pipeline needed to be improved.

The last part, on the scientific analysis, was the most engaging. Even if my results are preliminary, it is the final proof to show that the pipeline can be used for science. The mass maps presented here show the distribution of dark matter over a total area of $1,695$ deg^2, the largest mass maps from galaxy weak-lensing published to date. The work on parameter inference gives a first idea of what will be possible with CFIS.

A paper will be submitted (Guinot et al. *in prep*) describing the study I have carried out on the CFIS data. It will go through the pipeline we have described in this manuscript and present a suite of validation tests and scientific results on cluster profiles and lensing mass maps. This will prepare the more cosmology focused papers that will follow.

As I mentioned several times in the previous chapters there is a lot of place for improvement. Machine learning has demonstrated to be a very powerful tool for astronomy and I intend to make use of it. The star/galaxy separation could be widely improved with such techniques, in combination with the new photometric bands that will be added to CFIS.

During my PhD, a new method to calibrate shear measurements was published, called metadetection (Sheldon et al. 2020), with very promising results. This method would be particularly interesting for measurements based on stacked images. The stacked PSF I presented appears to be the bottleneck for the secondary shape measurements. I will develop this method further since it has the potential to be very useful for future surveys. I plan to achieve this goal by experimenting new stacking methods, such as the one presented in Bosch et al. 2017. I also contributed to a new PSF modeling algorithm (Liaudat et al., 2020, submitted to A&A), which I did not use for this work due to time constraints. Yet, this method will be implemented as the primary technique to estimate and interpolate the PSF for future releases.

I briefly discussed the blending problem in this manuscript. This will require a particular attention in the future given that it will be as problematic as the PSF in coming surveys. The work I have started using machine learning algorithms will be pursued. We can imagine to implement a technique similar to what the DES collaboration introduced with Multi-Object Fitting. This problem is linked to metadection, which will provide a well suited framework to solve it.

Finally, I have neglected a lot of systematics in the science analysis. As crucial the careful treatment of systematics is in data processing, as important this will be for the scientific analysis. I will spend more time on this part to make fair comparison with other studies in the future.

Bibliography

Abbott, T. M. C. et al. (2018). "Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing". In: *Physical Review D* 98.4. issn: 2470-0029. doi: 10.1103/physrevd.98.043526. url: http://dx.doi.org/10.1103/PhysRevD.98.043526.

Ade, P. A. R. et al. (2016). "Planck2015 results". In: *Astronomy & Astrophysics* 594, A27. issn: 1432-0746. doi: 10.1051/0004-6361/201525823. url: http://dx.doi.org/10.1051/0004-6361/201525823.

Aghanim, N. et al. (2020). "Planck 2018 results". In: *Astronomy & Astrophysics* 641, A6. issn: 1432-0746. doi: 10.1051/0004-6361/201833910. url: http://dx.doi.org/10.1051/0004-6361/201833910.

Bacon, David J. et al. (Sept. 2003). "Joint cosmic shear measurements with the Keck and William Herschel Telescopes". In: MNRAS 344.3, pp. 673–685. doi: 10.1046/j.1365-8711.2003.06877.x. arXiv: astro-ph/0203134 [astro-ph].

Bernstein, G. M. and M. Jarvis (Feb. 2002). "Shapes and Shears, Stars and Smears: Optimal Measurements for Weak Lensing". In: AJ 123.2, pp. 583–618. doi: 10.1086/338085. arXiv: astro-ph/0107431 [astro-ph].

Bernstein, Gary M. and Robert Armstrong (2014a). "Bayesian lensing shear measurement". In: *Monthly Notices of the Royal Astronomical Society* 438.2, 1880–1893. issn: 1365-2966. doi: 10.1093/mnras/stt2326. url: http://dx.doi.org/10.1093/mnras/stt2326.

— (2014b). "Bayesian lensing shear measurement". In: *Monthly Notices of the Royal Astronomical Society* 438.2, 1880–1893. issn: 1365-2966. doi: 10.1093/mnras/stt2326. url: http://dx.doi.org/10.1093/mnras/stt2326.

Bertin, E. (July 2011). "Automated Morphometry with SExtractor and PSFEx". In: *Astronomical Data Analysis Software and Systems XX*. Ed. by I. N. Evans et al. Vol. 442. Astronomical Society of the Pacific Conference Series, p. 435.

Bertin, E. and S. Arnouts (June 1996). "SExtractor: Software for source extraction." In: A&AS 117, pp. 393–404. doi: 10.1051/aas:1996164.

Bertin, Emmanuel et al. (Jan. 2002). "The TERAPIX Pipeline". In: *Astronomical Data Analysis Software and Systems XI*. Ed. by David A. Bohlender, Daniel Durand, and Thomas H. Handley. Vol. 281. Astronomical Society of the Pacific Conference Series, p. 228.

Bosch, James et al. (2017). "The Hyper Suprime-Cam software pipeline". In: *Publications of the Astronomical Society of Japan* 70.SP1. issn: 2053-051X. doi: 10.1093/pasj/psx080. url: http://dx.doi.org/10.1093/pasj/psx080.

Brown, A. G. A. et al. (2018). "Gaia Data Release 2". In: *Astronomy & Astrophysics* 616, A1. issn: 1432-0746. doi: 10.1051/0004-6361/201833051. url: http://dx.doi.org/10.1051/0004-6361/201833051.

Buscher, D. F. et al. (Feb. 1995). "Interferometric seeing measurements on Mt. Wilson: power spectra and outer scales". In: Appl. Opt. 34.6, p. 1081. doi: 10.1364/AO.34.001081.

Carretero, J. et al. (Dec. 2014). "An algorithm to build mock galaxy catalogues using MICE simulations". In: *Monthly Notices of the Royal Astronomical Society* 447.1, pp. 646–670. issn: 0035-8711. doi: 10.1093/mnras/stu2402. eprint: https://academic.oup.com/mnras/article-pdf/447/1/646/4911622/stu2402.pdf. url: https://doi.org/10.1093/mnras/stu2402.

Chambers, K. C. et al. (2016). *The Pan-STARRS1 Surveys*. arXiv: 1612.05560 [astro-ph.IM].

Crittenden, R. G. et al. In: ().

Crocce, M. et al. (2015). "The MICE Grand Challenge lightcone simulation – II. Halo and galaxy catalogues". In: *Monthly Notices of the Royal Astronomical Society* 453.2, 1513–1530. issn: 1365-2966. doi: 10.1093/mnras/stv1708. url: http://dx.doi.org/10.1093/mnras/stv1708.

Cuillandre, J., B. Mahoney, and K. Withington (May 2014). "A Signal-to-noise Approach to Queue Observing with Ground-based Telescopes". In: *Astronomical Data Analysis Software and Systems XXIII*. Ed. by N. Manset and P. Forshay. Vol. 485. Astronomical Society of the Pacific Conference Series, p. 81.

Desai, S. et al. (2012). "THE BLANCO COSMOLOGY SURVEY: DATA ACQUISITION, PROCESSING, CALIBRATION, QUALITY DIAGNOSTICS, AND DATA RELEASE". In: *The Astrophysical Journal* 757.1, p. 83. issn: 1538-4357. doi: 10.1088/0004-637x/757/1/83. url: http://dx.doi.org/10.1088/0004-637X/757/1/83.

Dodelson, Scott (2003). *Modern cosmology*.

Einstein, Albert (Jan. 1915). "Die Feldgleichungen der Gravitation". In: *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin*, pp. 844–847.

Eisenstein, Daniel J. and Wayne Hu (1998). "Baryonic Features in the Matter Transfer Function". In: *The Astrophysical Journal* 496.2, 605–614. issn: 1538-4357. doi: 10.1086/305424. url: http://dx.doi.org/10.1086/305424.

Erben, T. et al. (July 2005). "GaBoDS: The Garching-Bonn Deep Survey. IV. Methods for the image reduction of multi-chip cameras demonstrated on data from the ESO Wide-Field Imager". In: *Astronomische Nachrichten* 326.6, pp. 432–464. doi: 10.1002/asna.200510396. arXiv: astro-ph/0501144 [astro-ph].

Erben, T. et al. (2013). "CFHTLenS: the Canada–France–Hawaii Telescope Lensing Survey – imaging data and catalogue products". In: *Monthly Notices of the Royal Astronomical Society* 433.3, 2545–2563. issn: 0035-8711. doi: 10.1093/mnras/stt928. url: http://dx.doi.org/10.1093/mnras/stt928.

Fang, Xiao, Tim Eifler, and Elisabeth Krause (2020). "2D-FFTLog: efficient computation of real-space covariance matrices for galaxy clustering and weak lensing". In: *Monthly Notices of the Royal Astronomical Society* 497.3, 2699–2714. issn: 1365-2966. doi: 10.1093/mnras/staa1726. url: http://dx.doi.org/10.1093/mnras/staa1726.

Fang, Xiao et al. (2020). "Beyond Limber: efficient computation of angular power spectra for galaxy clustering and weak lensing". In: *Journal of Cosmology and Astroparticle Physics* 2020.05, 010–010. issn: 1475-7516. doi: 10.1088/1475-7516/2020/05/010. url: http://dx.doi.org/10.1088/1475-7516/2020/05/010.

Foreman-Mackey, Daniel et al. (2013). "emcee: The MCMC Hammer". In: *Publications of the Astronomical Society of the Pacific* 125.925, 306–312. issn: 1538-3873. doi: 10.1086/670067. url: http://dx.doi.org/10.1086/670067.

Fosalba, P. et al. (Dec. 2014). "The MICE Grand Challenge light-cone simulation – III. Galaxy lensing mocks from all-sky lensing maps". In: *Monthly Notices of the Royal Astronomical Society* 447.2, pp. 1319–1332. issn: 0035-8711. doi: 10.1093/mnras/stu2464. eprint: https://academic.oup.com/mnras/article-pdf/447/2/1319/8093572/stu2464.pdf. url: https://doi.org/10.1093/mnras/stu2464.

Fosalba, P. et al. (2015). "The MICE grand challenge lightcone simulation – I. Dark matter clustering". In: *Monthly Notices of the Royal Astronomical Society* 448.4, 2987–3000. issn: 0035-8711. doi: 10.1093/mnras/stv138. url: http://dx.doi.org/10.1093/mnras/stv138.

Gilliland, Ronald L. and A. K. Dupree (1996). "First Image of the Surface of a Star with the [ITAL]Hubble Space Telescope[/ITAL]". In: *The Astrophysical Journal* 463.1, pp. L29–L32. doi: 10.1086/310043. url: https://doi.org/10.1086%2F310043.

Groener, A. M., D. M. Goldberg, and M. Sereno (2015). "The galaxy cluster concentration–mass scaling relation". In: *Monthly Notices of the Royal Astronomical Society* 455.1, 892–919. issn: 0035-8711. doi: 10.1093/mnras/stv2341. url: http://dx.doi.org/10.1093/mnras/stv2341.

Guyonnet, A. et al. (2015). "Evidence for self-interaction of charge distribution in charge-coupled devices". In: *Astronomy & Astrophysics* 575, A41. issn: 1432-0746. doi: 10.1051/0004-6361/201424897. url: http://dx.doi.org/10.1051/0004-6361/201424897.

Gwyn, Stephen D. J. (Feb. 2008). "MegaPipe: The MegaCam Image Stacking Pipeline at the Canadian Astronomical Data Centre". In: PASP 120.864, p. 212. doi: 10.1086/526794. arXiv: 0710.0370 [astro-ph].

Hirata, Christopher and Uroš Seljak (Aug. 2003). "Shear calibration biases in weak-lensing surveys". In: MNRAS 343.2, pp. 459–480. doi: 10.1046/j.1365-8711.2003.06683.x. arXiv: astro-ph/0301054 [astro-ph].

Hoffmann, K. et al. (Feb. 2015). "Measuring the growth of matter fluctuations with third-order galaxy correlations". In: MNRAS 447.2, pp. 1724–1745. doi: 10.1093/mnras/stu2492. arXiv: 1403.1259 [astro-ph.CO].

Huff, Eric and Rachel Mandelbaum (2017). *Metacalibration: Direct Self-Calibration of Biases in Shear Measurement*. arXiv: 1702.02600 [astro-ph.CO].

Jarvis, M. et al. (May 2016). "The DES Science Verification weak lensing shear catalogues". In: *Monthly Notices of the Royal Astronomical Society* 460.2, pp. 2245–2281. issn: 0035-8711. doi: 10.1093/mnras/stw990. eprint: https://academic.oup.com/mnras/article-pdf/460/2/2245/8118034/stw990.pdf. url: https://doi.org/10.1093/mnras/stw990.

Jarvis, M. et al. (Aug. 2016). "The DES Science Verification weak lensing shear catalogues". In: MNRAS 460, pp. 2245–2281. doi: 10.1093/mnras/stw990. arXiv: 1507.05603 [astro-ph.IM].

Kaiser, Nick and Gordon Squires (Feb. 1993). "Mapping the Dark Matter with Weak Gravitational Lensing". In: ApJ 404, p. 441. doi: 10.1086/172297.

Kaiser, Nick, Gordon Squires, and Tom Broadhurst (1995). "A Method for Weak Lensing Observations". In: *The Astrophysical Journal* 449, p. 460. issn: 1538-4357. doi: 10.1086/176071. url: http://dx.doi.org/10.1086/176071.

Kilbinger, Martin (2015). "Cosmology with cosmic shear observations: a review". In: *Reports on Progress in Physics* 78.8, p. 086901. issn: 1361-6633. doi: 10.1088/0034-4885/78/8/086901. url: http://dx.doi.org/10.1088/0034-4885/78/8/086901.

Krause, Elisabeth and Tim Eifler (2017). "cosmolike – cosmological likelihood analyses for photometric galaxy surveys". In: *Monthly Notices of the Royal Astronomical Society* 470.2, 2100–2112. issn: 1365-2966. doi: 10.1093/mnras/stx1261. url: http://dx.doi.org/10.1093/mnras/stx1261.

Kron, R. G. (June 1980). "Photometry of a complete sample of faint galaxies." In: ApJS 43, pp. 305–325. doi: 10.1086/190669.

Machado, Eduardo et al. (2016). "Exploring machine learning methods for the star/galaxy separation problem". In: *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE, pp. 123–130.

Mandelbaum, Rachel et al. (2012). "Precision simulation of ground-based lensing data using observations from space". In: *Monthly Notices of the Royal Astronomical Society* 420.2, pp. 1518–1540.

Melchior, P. et al. (Apr. 2011). "Weak gravitational lensing with DEIMOS". In: MNRAS 412.3, pp. 1552–1558. doi: 10.1111/j.1365-2966.2010.17875.x. arXiv: 1008.1076 [astro-ph.IM].

Mohr, Joseph J. et al. (2012). "The Dark Energy Survey data processing and calibration system". In: *Software and Cyberinfrastructure for Astronomy II*. Ed. by Nicole M. Radziwill and GianlucaEditors Chiozzi. doi: 10.1117/12.926785. url: http://dx.doi.org/10.1117/12.926785.

Ménard, Brice et al. (2013). *Clustering-based redshift estimation: method and application to data*. arXiv: 1303.4722 [astro-ph.CO].

Navarro, Julio F., Carlos S. Frenk, and Simon D. M. White (May 1996). "The Structure of Cold Dark Matter Halos". In: ApJ 462, p. 563. doi: 10.1086/177173. arXiv: astro-ph/9508025 [astro-ph].

Pujol, Arnau et al. (June 2020). "Shear measurement bias II: a fast machine learning calibration method". In: *arXiv e-prints*, arXiv:2006.07011, arXiv:2006.07011. arXiv: 2006.07011 [astro-ph.CO].

Racine, Rene (Aug. 1996). "The Telescope Point Spread Function". In: PASP 108, p. 699. doi: 10.1086/133788.

Reid, Beth et al. (Jan. 2016). "SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: galaxy target selection and large-scale structure catalogues". In: MNRAS 455.2, pp. 1553–1573. doi: 10.1093/mnras/stv2382. arXiv: 1509.06529 [astro-ph.CO].

Rowe, B. (2010). "Improving PSF modelling for weak gravitational lensing using new methods in model selection". In: MNRAS 404, pp. 350–366. doi: 10.1111/j.1365-2966.2010.16277.x. arXiv: 0904.3056 [astro-ph.CO].

Samuroff, S et al. (2017). "Dark Energy Survey Year 1 results: the impact of galaxy neighbours on weak lensing cosmology with im3shape". In: *Monthly Notices of the Royal Astronomical Society* 475.4, 4524–4543. issn: 1365-2966. doi: 10.1093/mnras/stx3282. url: http://dx.doi.org/10.1093/mnras/stx3282.

Schirmer, M. (Dec. 2013). "THELI: Convenient Reduction of Optical, Near-infrared, and Mid-infrared Imaging Data". In: ApJS 209.2, 21, p. 21. doi: 10.1088/0067-0049/209/2/21. arXiv: 1308.4989 [astro-ph.IM].

Schneider, P. (2006). "Weak Gravitational Lensing". In: *Gravitational Lensing: Strong, Weak and Micro*, 269–451. issn: 1861-7980. doi: 10.1007/978-3-540-30310-7_3. url: http://dx.doi.org/10.1007/978-3-540-30310-7_3.

Scoville, N. et al. (Sept. 2007). "The Cosmic Evolution Survey (COSMOS): Overview". In: ApJS 172, pp. 1–8. doi: 10.1086/516585. eprint: astro-ph/0612305.

Sérsic, J. L. (Feb. 1963). "Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy". In: *Boletin de la Asociacion Argentina de Astronomia La Plata Argentina* 6, pp. 41–43.

Sheldon, Erin S. and Eric M. Huff (2017). "Practical Weak Lensing Shear Measurement with Metacalibration". In: *Astrophys. J.* 841.1, p. 24. doi: 10.3847/1538-4357/aa704b. arXiv: 1702.02601 [astro-ph.CO].

Sheldon, Erin S. et al. (2020). *Mitigating Shear-dependent Object Detection Biases with Metacalibration*. arXiv: 1911.02505 [astro-ph.CO].

Smith, R. E. et al. (June 2003). "Stable clustering, the halo model and non-linear cosmological power spectra". In: MNRAS 341.4, pp. 1311–1332. doi: 10.1046/j.1365-8711.2003.06503.x. arXiv: astro-ph/0207664 [astro-ph].

Sunyaev, R. A. and Ia. B. Zeldovich (Jan. 1980). "Microwave background radiation as a probe of the contemporary structure and history of the universe". In: ARA&A 18, pp. 537–560. doi: 10.1146/annurev.aa.18.090180.002541.

Sunyaev, R. A. and Ya. B. Zeldovich (Apr. 1970). "Small-Scale Fluctuations of Relic Radiation". In: Ap&SS 7.1, pp. 3–19. doi: 10.1007/BF00653471.

Takada, Masahiro and Bhuvnesh Jain (Sept. 2003). "Three-point correlations in weak lensing surveys: model predictions and applications". In: MNRAS 344.3, pp. 857–886. doi: 10.1046/j.1365-8711.2003.06868.x. arXiv: astro-ph/0304034 [astro-ph].

Trujillo, I. et al. (Dec. 2001). "The effects of seeing on Sérsic profiles - II. The Moffat PSF". In: MNRAS 328.3, pp. 977–985. doi: 10.1046/j.1365-8711.2001.04937.x. arXiv: astro-ph/0109067 [astro-ph].

Zuntz, J et al. (2018). "Dark Energy Survey Year 1 results: weak lensing shape catalogues". In: *Monthly Notices of the Royal Astronomical Society* 481.1, 1149–1182. issn: 1365-2966. doi: 10.1093/mnras/sty2219. url: http://dx.doi.org/10.1093/mnras/sty2219.

Appendix A

PSF leakage

In this appendix we show the PSF leakage for the 3 other patches (P1, P2 and P4) not mention in chapter 4. We can see that the leakage vary between the different fields. This due to the variation of the PSF. The images are taken in various atmospheric conditions and sky location. Figures A.1 to A.6 present the PSF leakage as defined in section 4.3.1 for the ngmix and Galsim methods.

Figures A.7 to A.15 show the leakage α as well as ξ^{sys} for the same three patches. We do not notice large variations between the different patches.

Ngmix

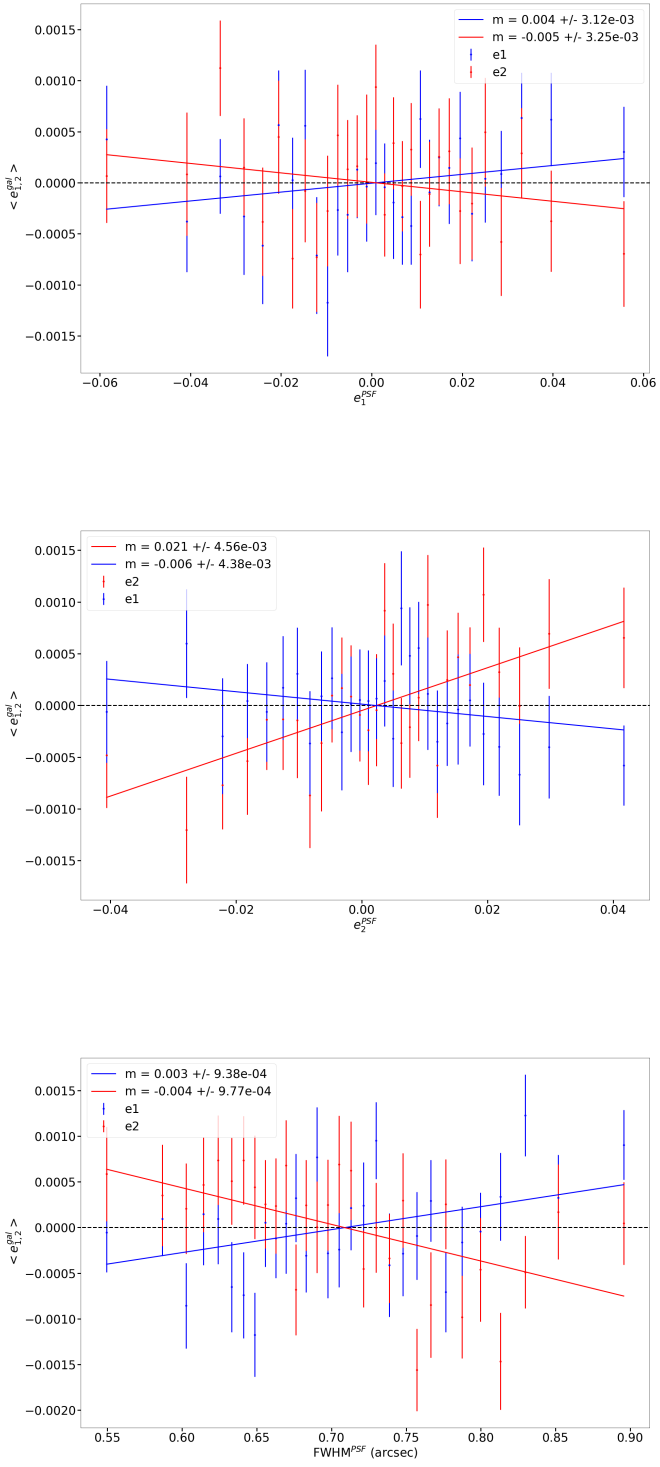


Figure A.1: PSF leakage using the average galaxy shape in bins of PSF ellipticity (top two panels) and size (bottom panel) on the field P1 for the ngmix catalog.

Galsim

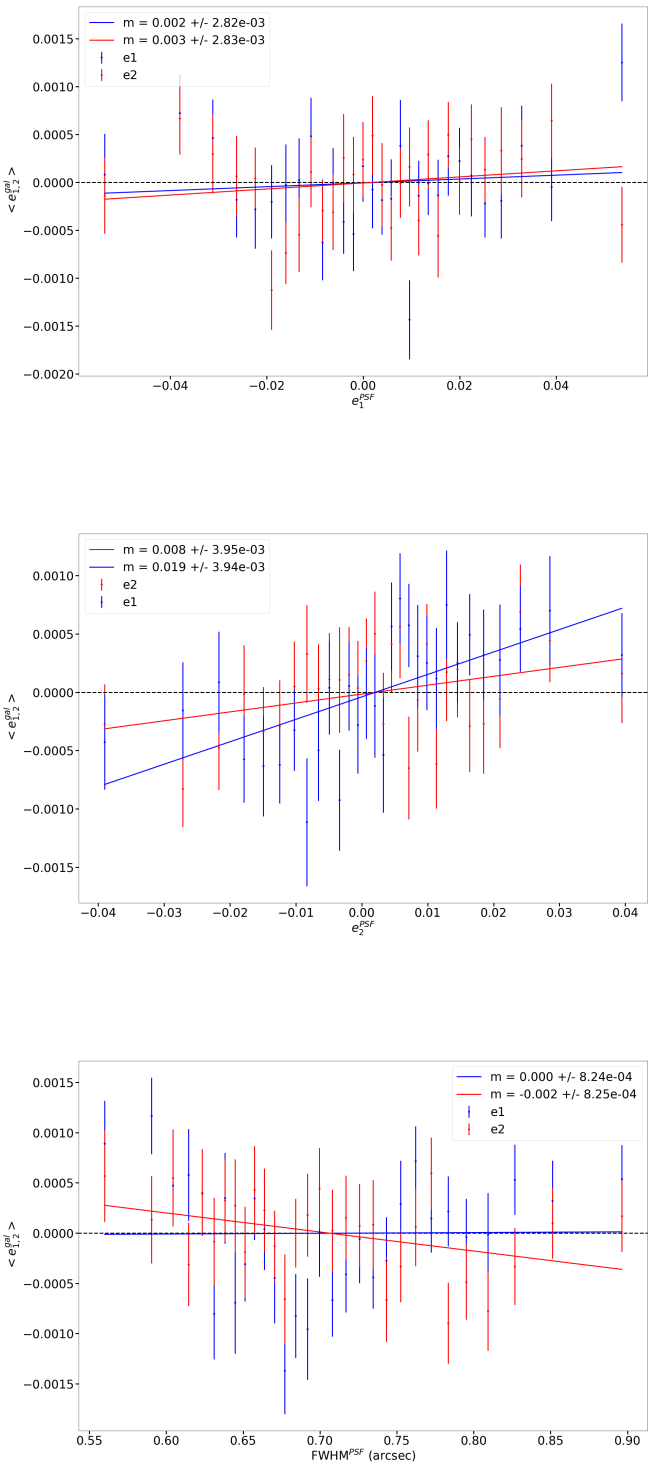


Figure A.2: PSF leakage using the average galaxy shape in bins of PSF ellipticity (top two panels) and size (bottom panel) on the field P1 for the Galsim catalog.

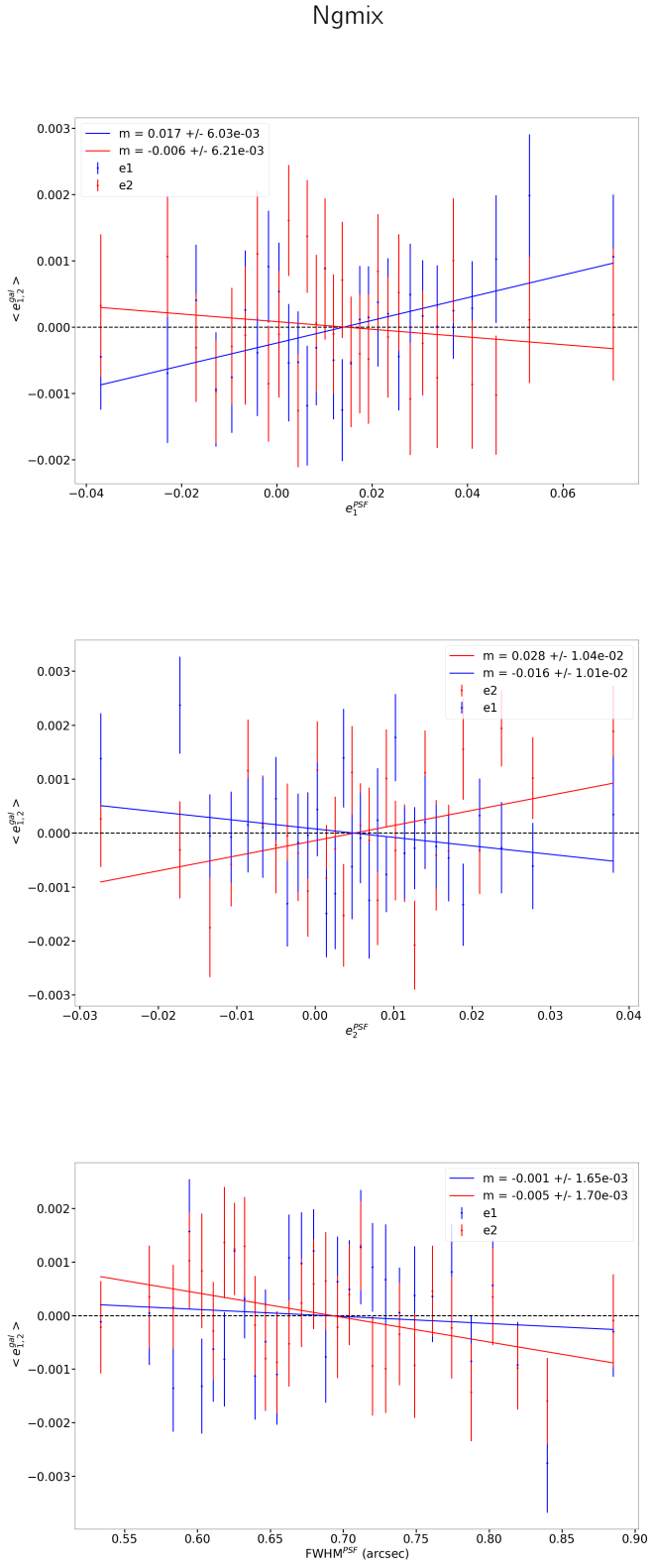


Figure A.3: PSF leakage using the average galaxy shape in bins of PSF ellipticity (top two panels) and size (bottom panel) on the field P2 for the ngmix catalog.

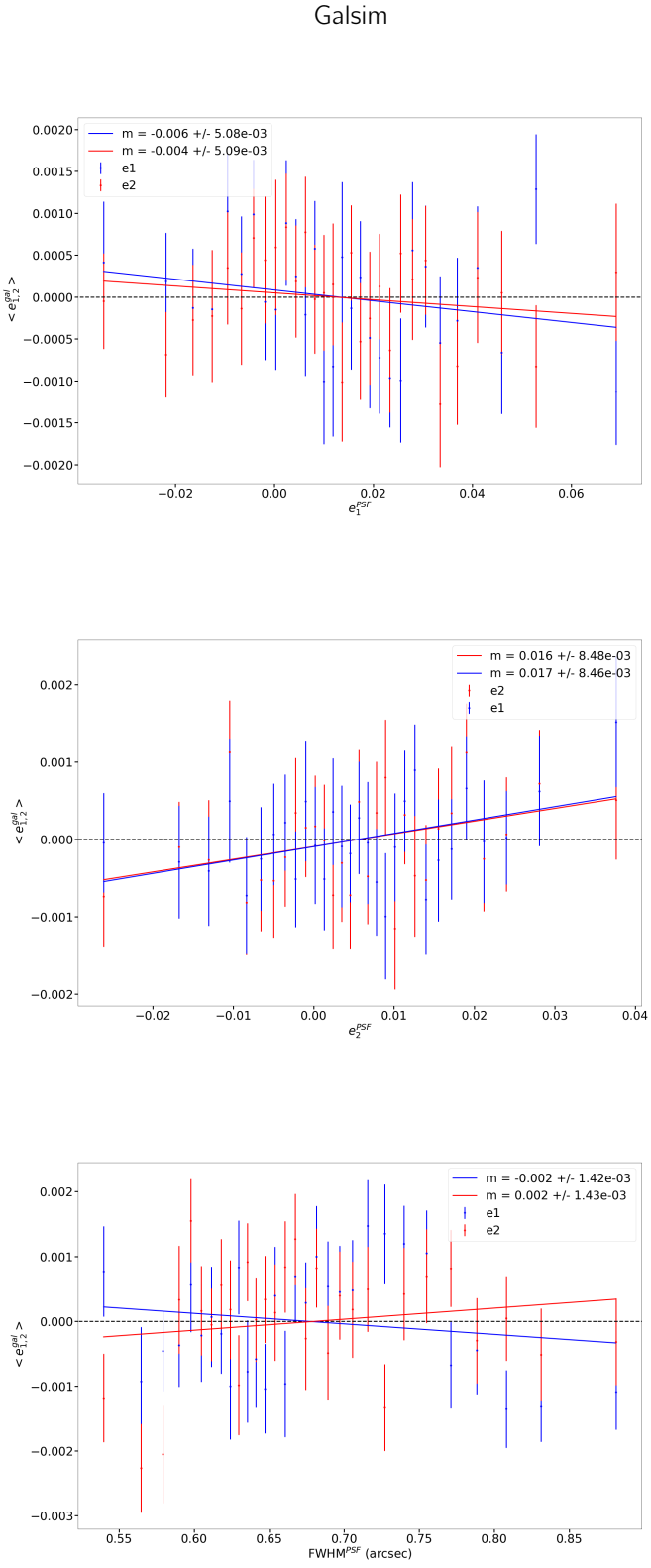


Figure A.4: PSF leakage using the average galaxy shape in bins of PSF ellipticity (top two panels) and size (bottom panel) on the field P2 for the Galsim catalog.

Ngmix

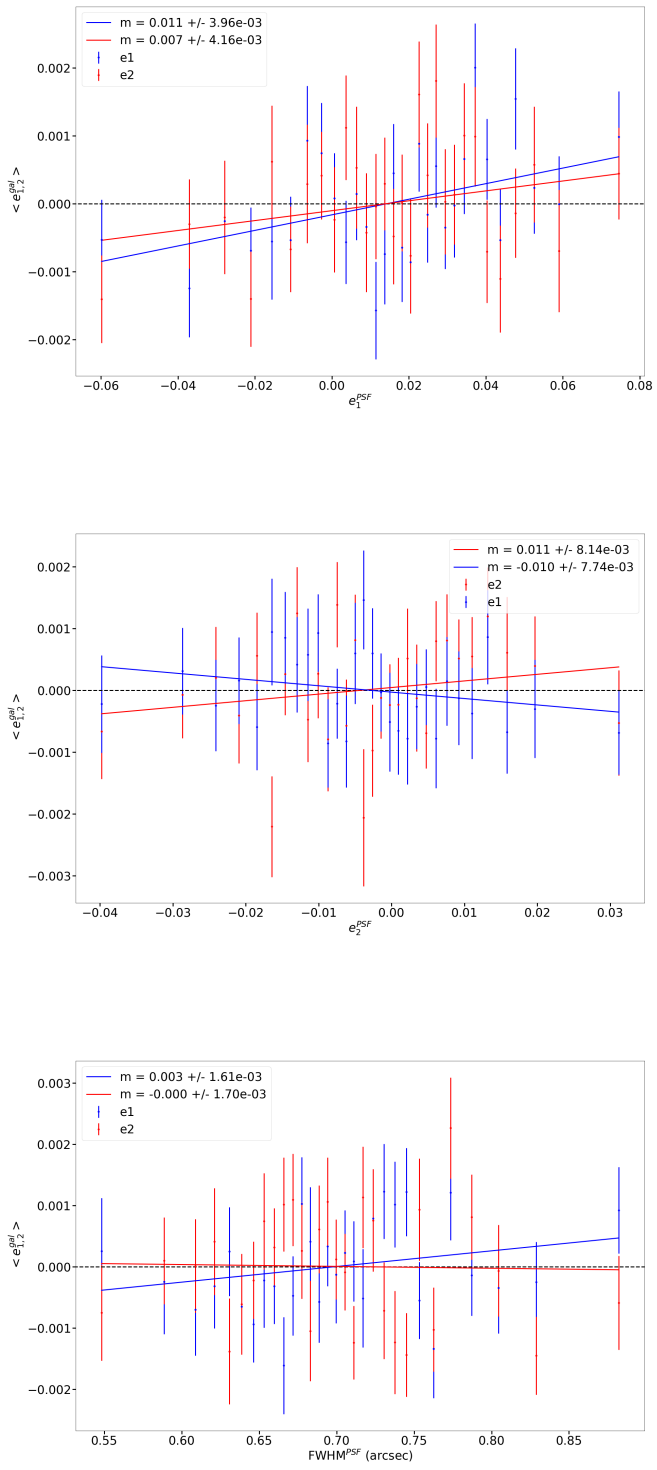


Figure A.5: PSF leakage using the average galaxy shape in bins of PSF ellipticity (top two panels) and size (bottom panel) on the field P4 for the ngmix catalog.

Galsim

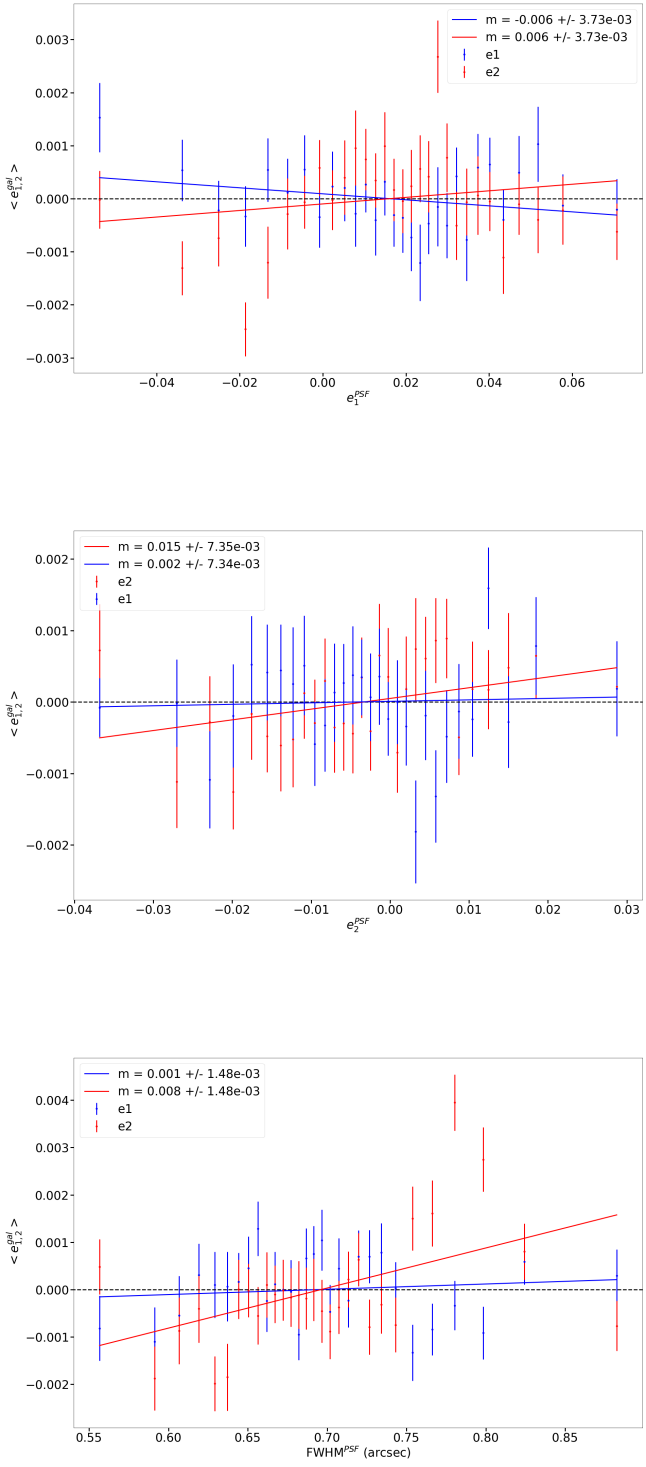


Figure A.6: PSF leakage using the average galaxy shape in bins of PSF ellipticity (top two panels) and size (bottom panel) on the field P4 for the Galsim catalog.

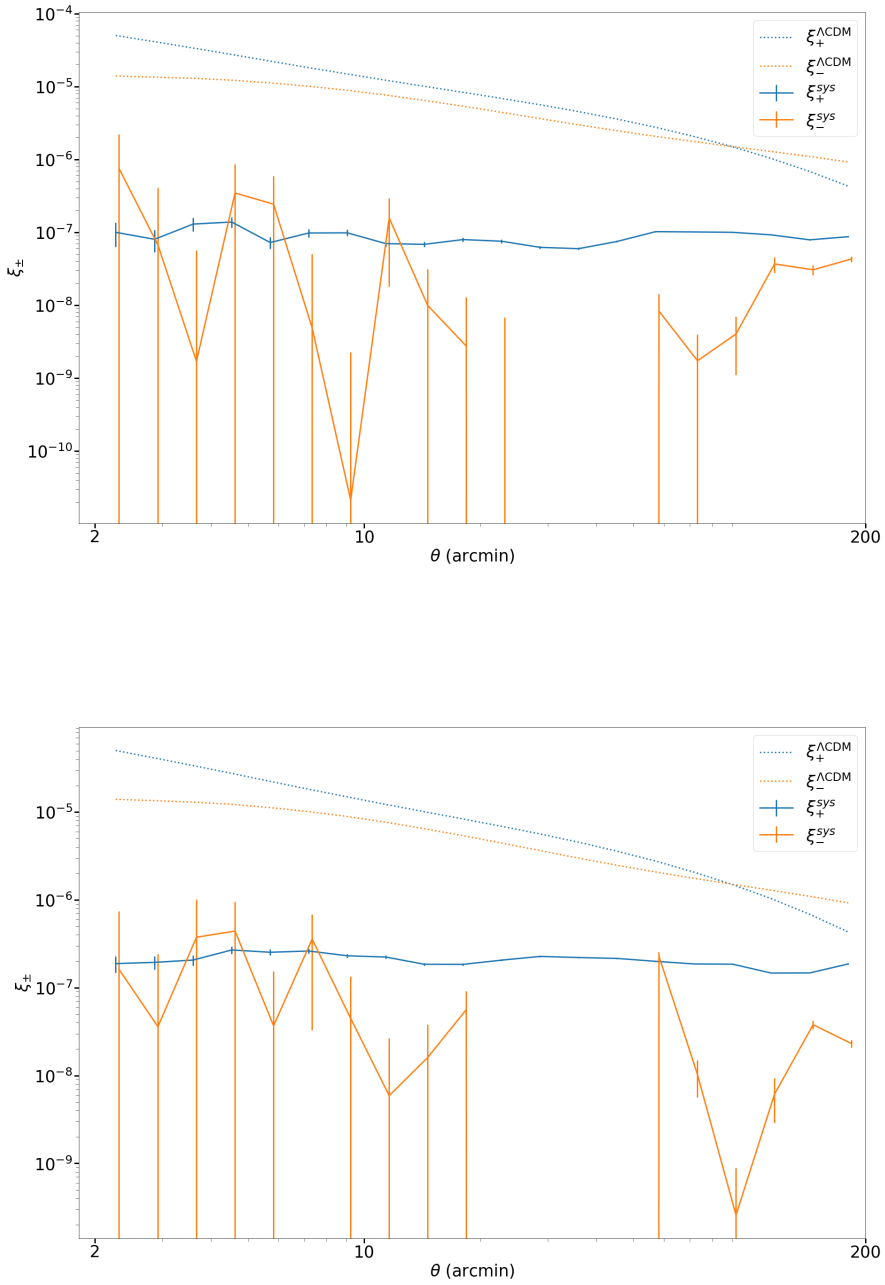


Figure A.7: $\xi_{\pm}^{\rm sys}(\theta)$ measured for ngmix (*Top*) and Galsim (*Bottom*) on patch P1.

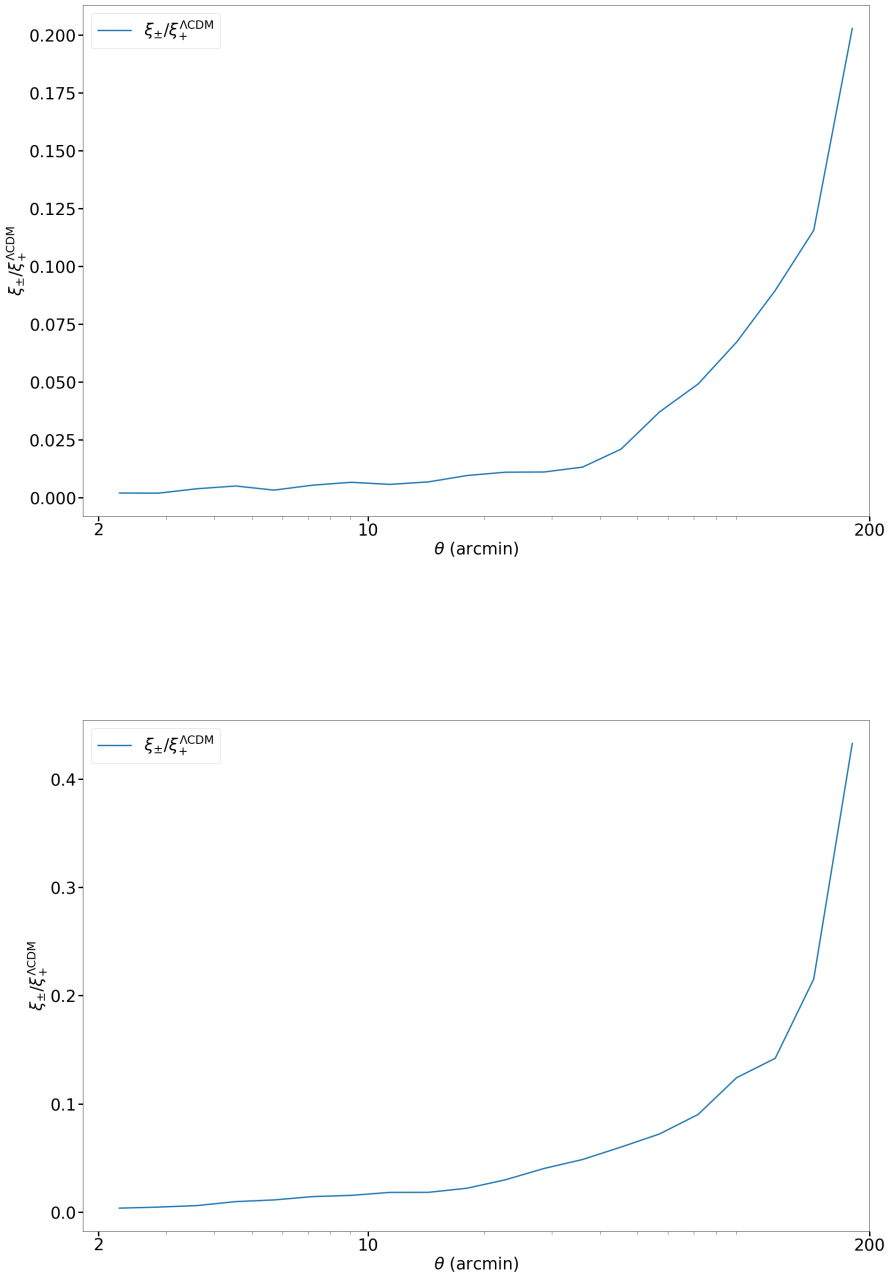


Figure A.8: Ratio of $\xi_+^{sys}(\theta)/\xi_+^{ss}(\theta)$ for ngmix (*Top*) and Galsim (*Bottom*) on patch P1.

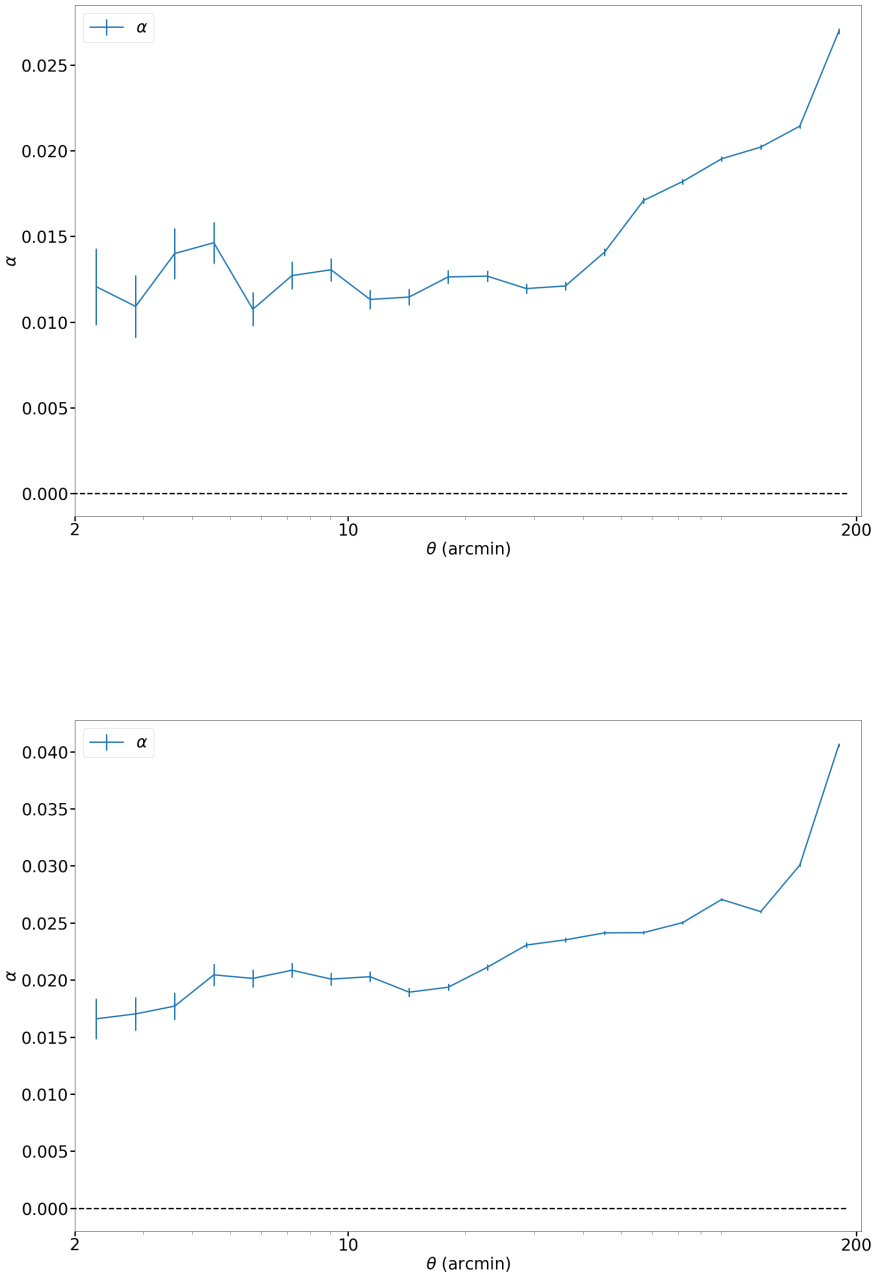


Figure A.9: Leakage α shown as a function of scale θ for ngmix (*Top*) and Galsim (*Bottom*) on patch P1.

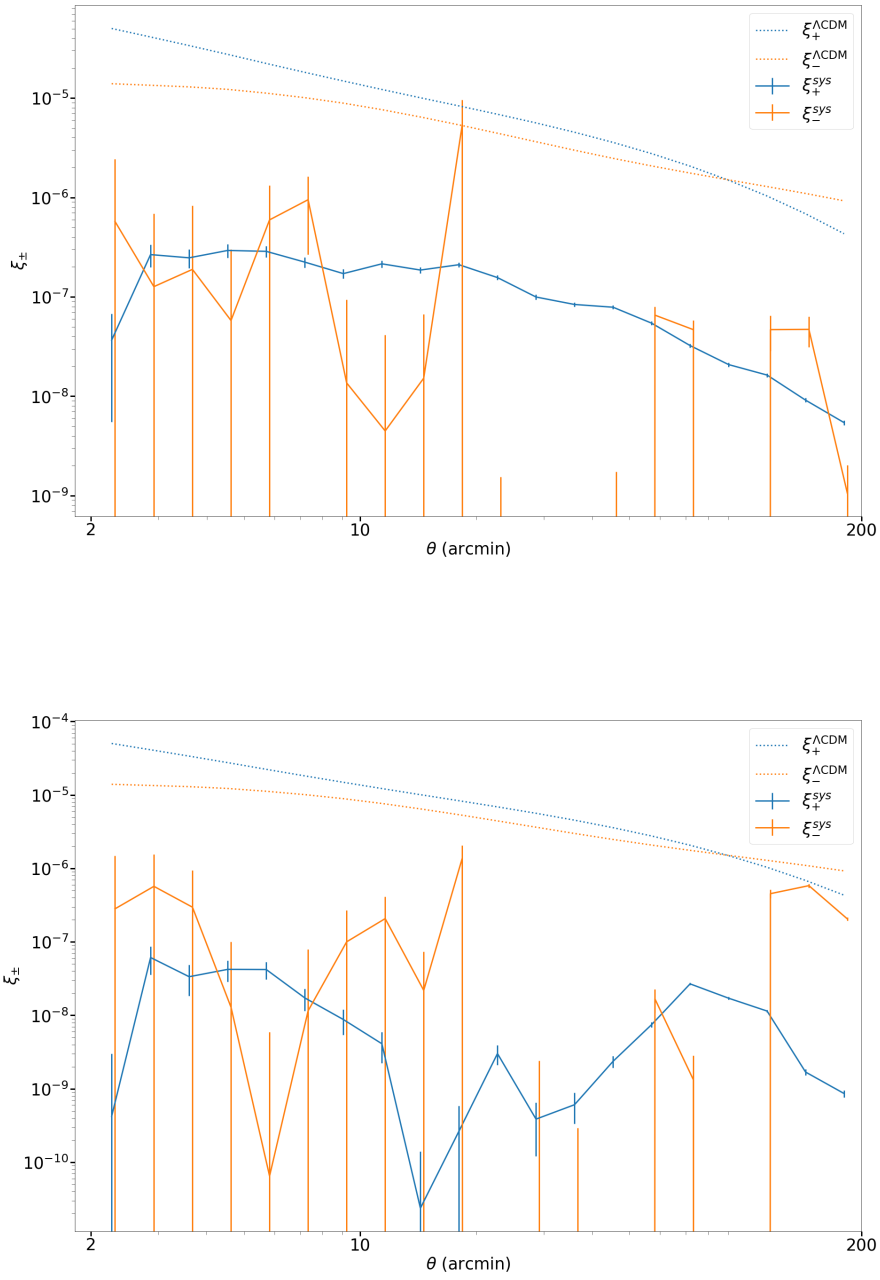


Figure A.10: $\xi_\pm^{sys}(\theta)$ measured for ngmix (*Top*) and Galsim (*Bottom*) on patch P2.

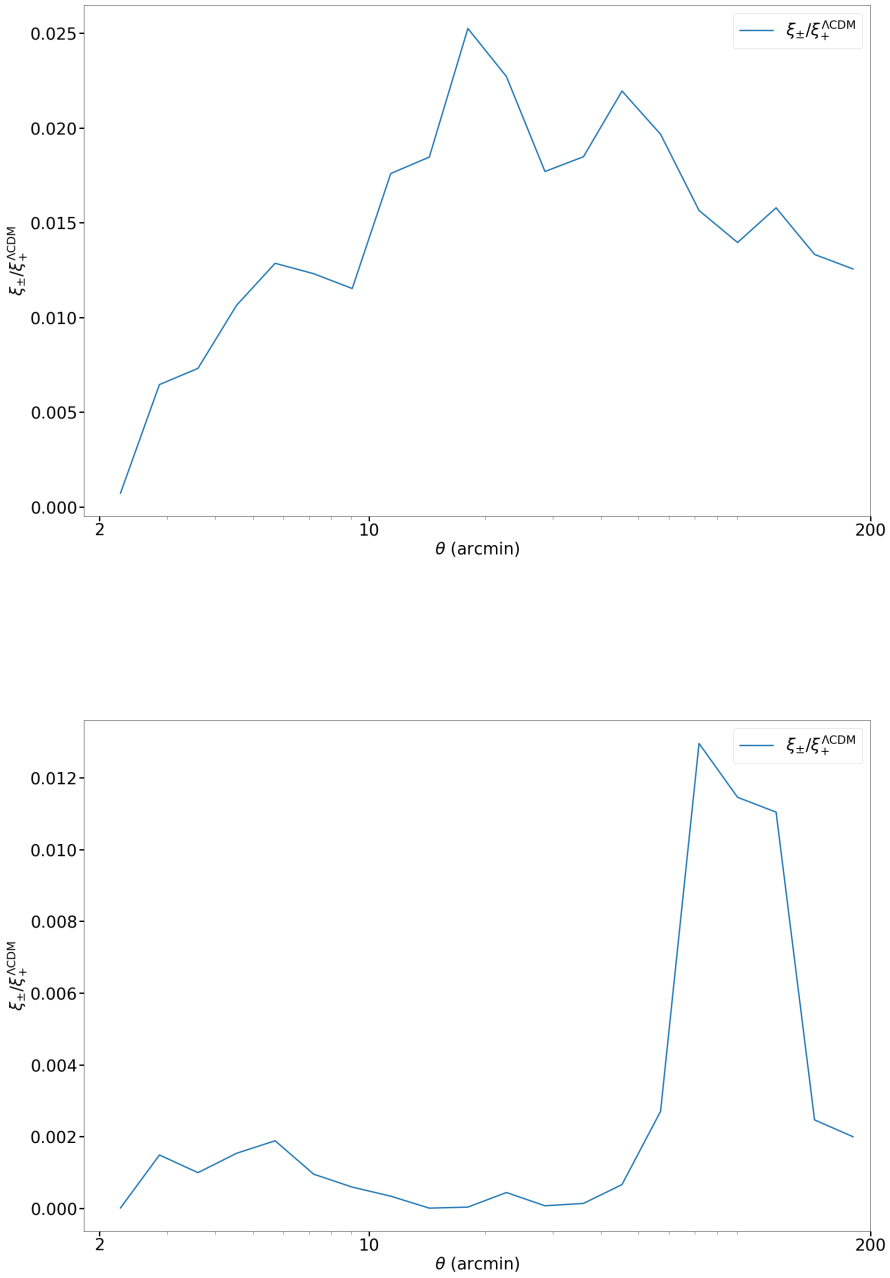


Figure A.11: Ratio of $\xi_+^{\rm sys}(\theta)/\xi_+^{\rm ss}(\theta)$ for ngmix (*Top*) and Galsim (*Bottom*) on patch P2.

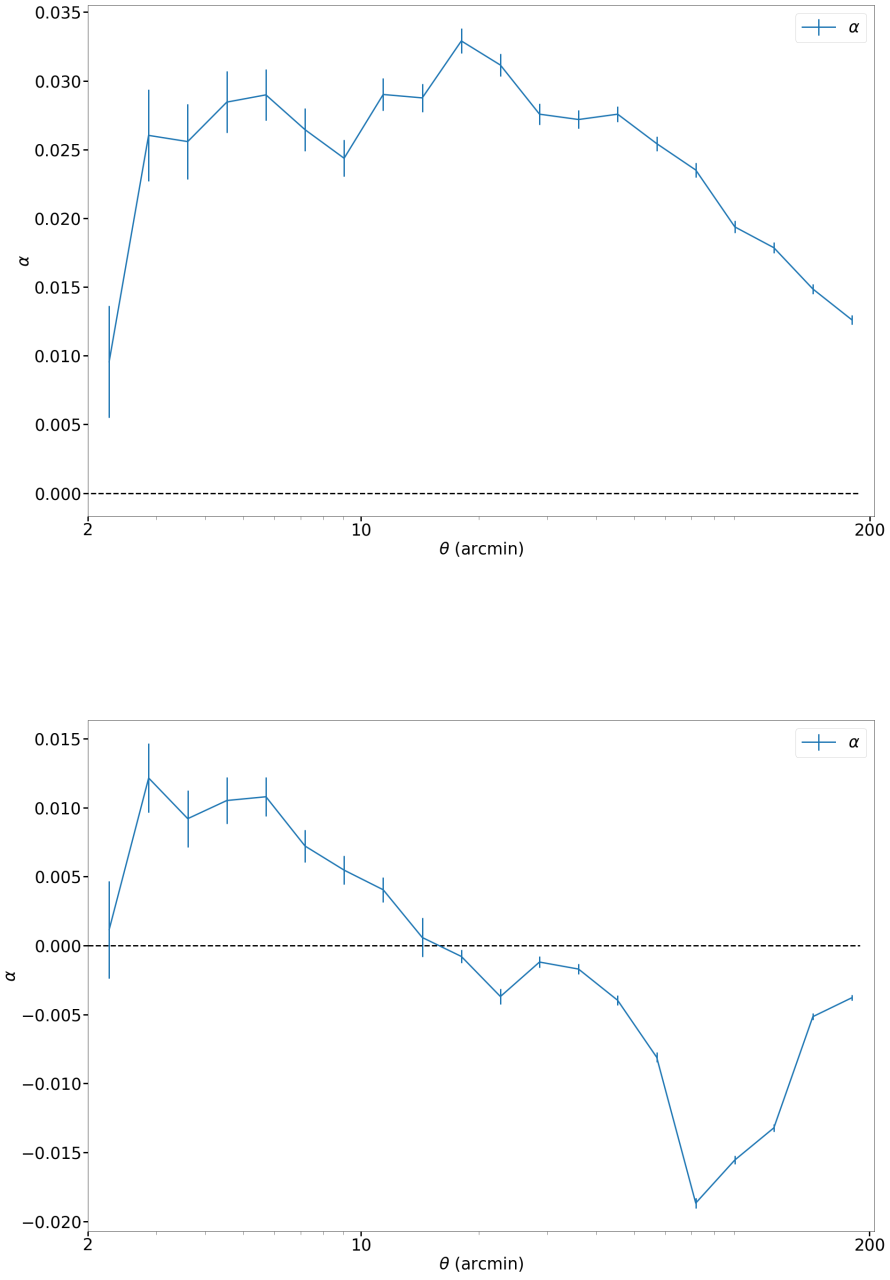


Figure A.12: Leakage α shown as a function of scale θ for ngmix (*Top*) and Galsim (*Bottom*) on patch P2.

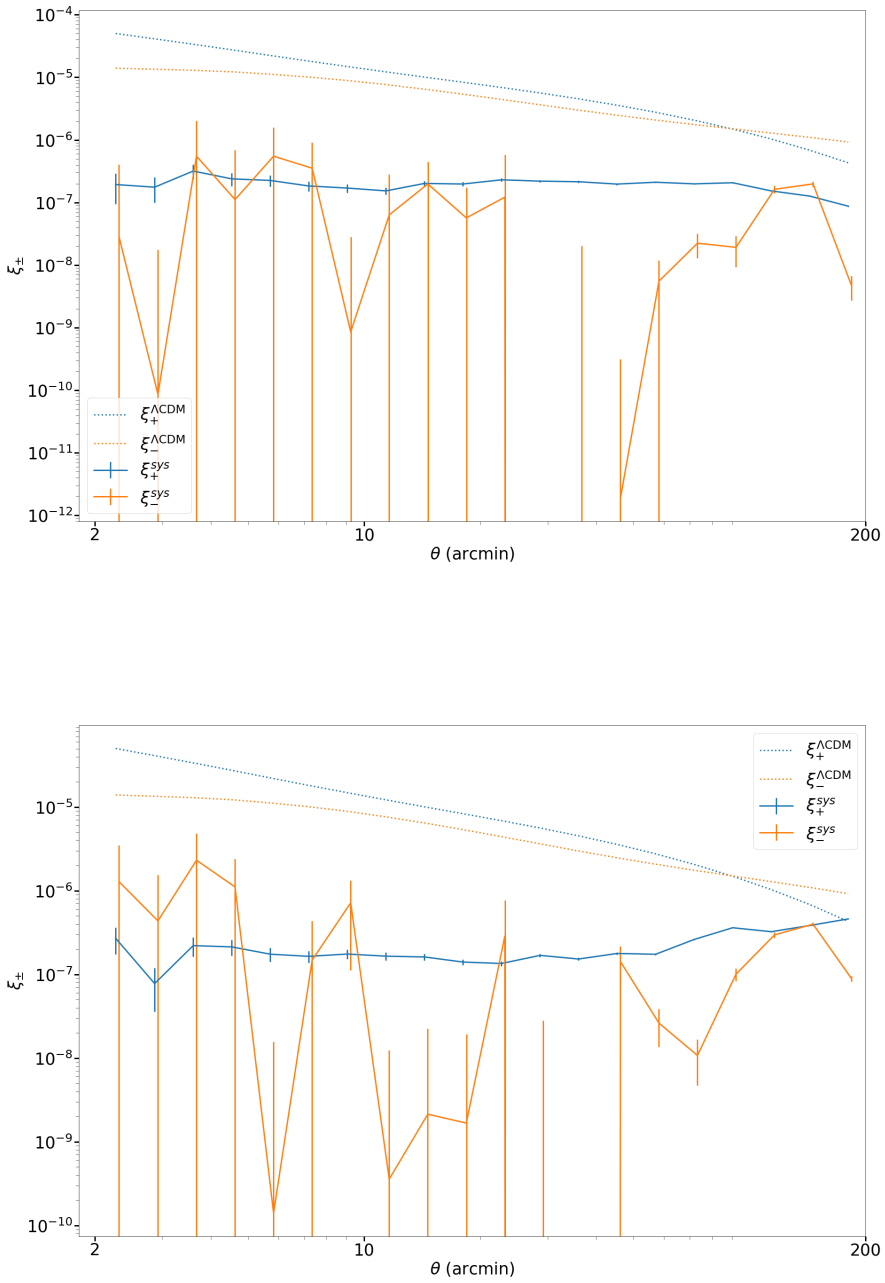
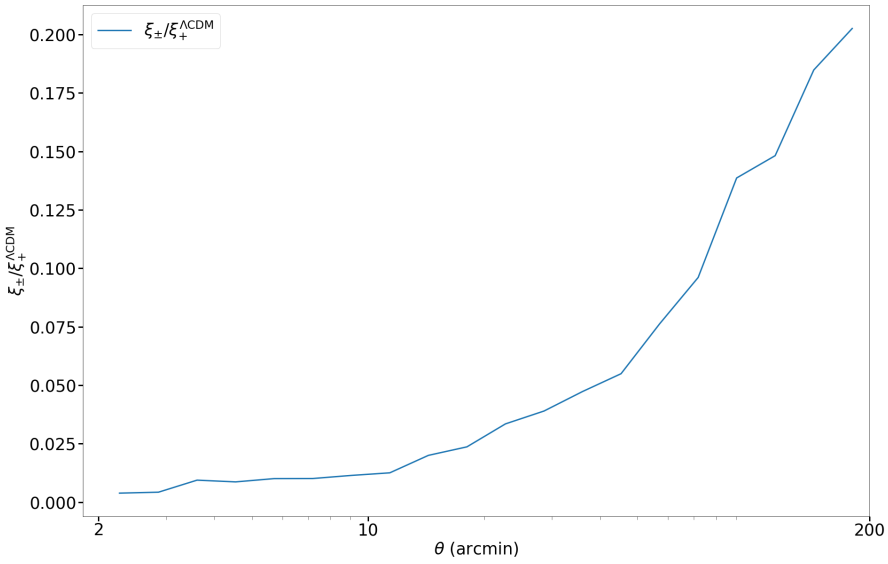


Figure A.13: $\xi_{\pm}^{sys}(\theta)$ measured for ngmix (*Top*) and Galsim (*Bottom*) on patch P4.



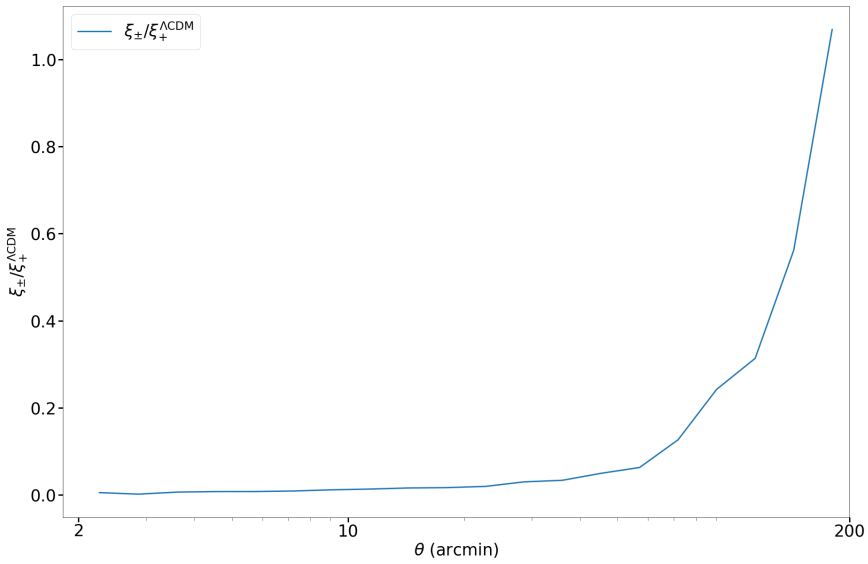


Figure A.14: Ratio of $\xi_{+}^{\rm sys}(\theta)/\xi_{+}^{\rm ss}(\theta)$ for ngmix (*Top*) and Galsim (*Bottom*) on patch P4.

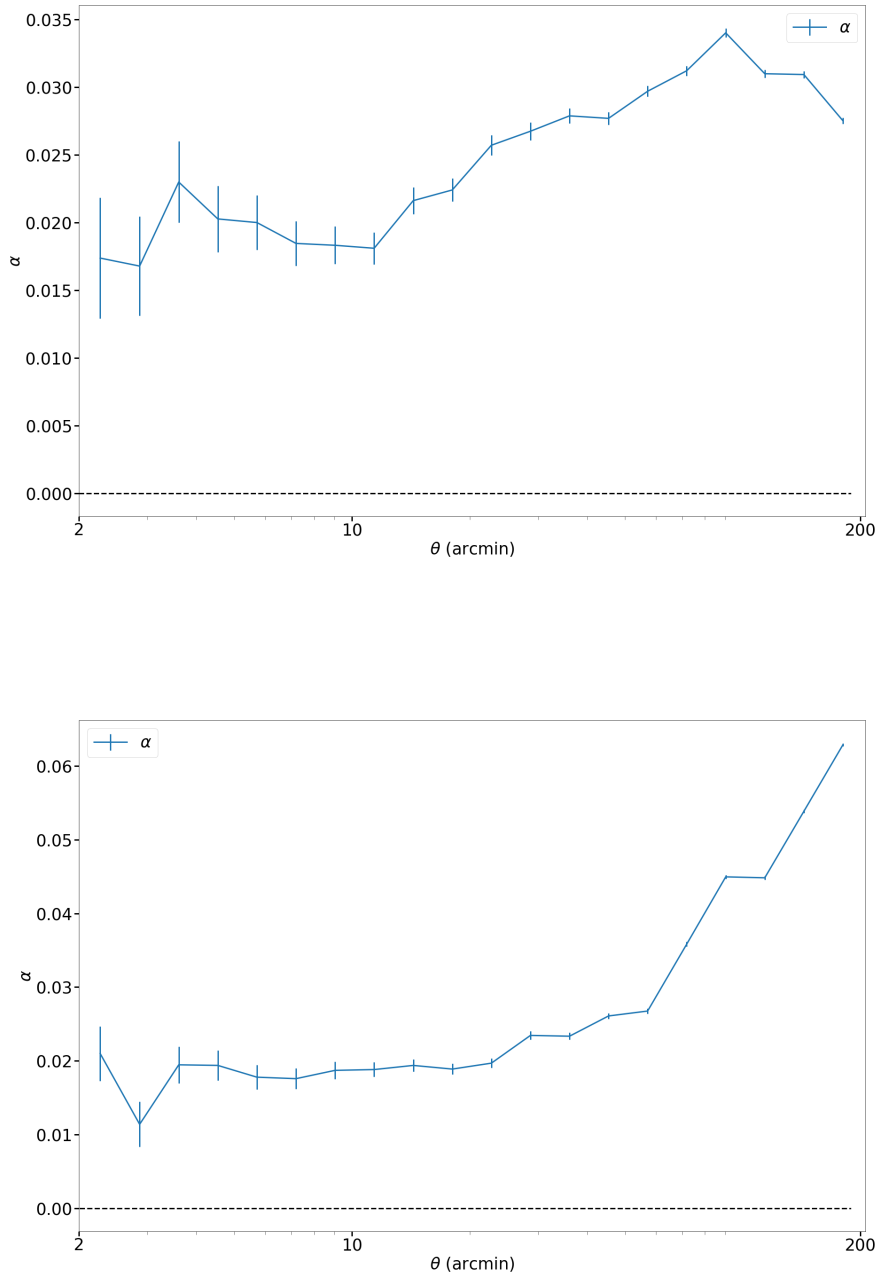


Figure A.15: Leakage α shown as a function of scale θ for ngmix (*Top*) and Galsim (*Bottom*) on patch P4.

Appendix B

Metacalibration response

In this appendix we present the response matrices from metacalibration for the field P1, P2 and P3 with the ngmix method. The values between are very closed since the survey is homogeneous.

Patch P1

Here is the total response:

$$R^{\text{total}} = \begin{pmatrix} 0.66 & 3.65 \times 10^{-4} \\ -1.36 \times 10^{-4} & 0.64 \end{pmatrix}, \tag{B.1}$$

while the response for selection effects is:

$$R^{\text{S}} = \begin{pmatrix} -0.137 & 0. \\ 0. & -0.144 \end{pmatrix}. \tag{B.2}$$

Figure B.1 show the diagonal terms for the galaxies and the stars.

Patch P2

Here is the total response:

$$R^{\text{total}} = \begin{pmatrix} 0.67 & -7.69 \times 10^{-4} \\ -1.76 \times 10^{-4} & 0.66 \end{pmatrix}, \tag{B.3}$$

while the response for selection effects is:

$$R^{\text{S}} = \begin{pmatrix} -0.143 & 0. \\ 0. & -0.149 \end{pmatrix}. \tag{B.4}$$

Figure B.2 show the diagonal terms for the galaxies and the stars.

Patch P4

Here is the total response:

$$R^{\text{total}} = \begin{pmatrix} 0.64 & -1.25 \times 10^{-3} \\ 1.38 \times 10^{-4} & 0.62 \end{pmatrix}, \tag{B.5}$$

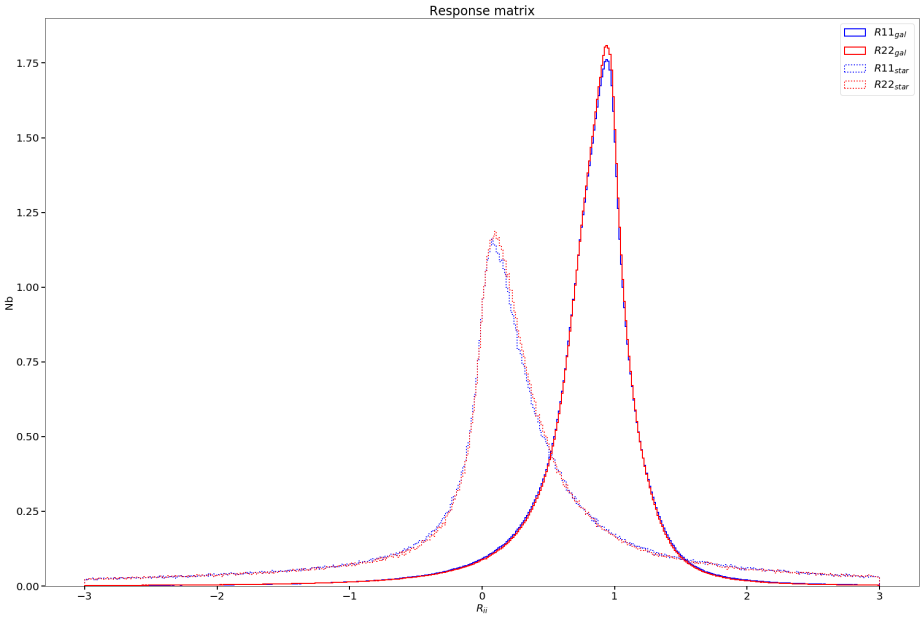


Figure B.1: Diagonal terms of the response martrix for the patch P1. The *solid line* for the galaxies and the *dashed line* for the stars.

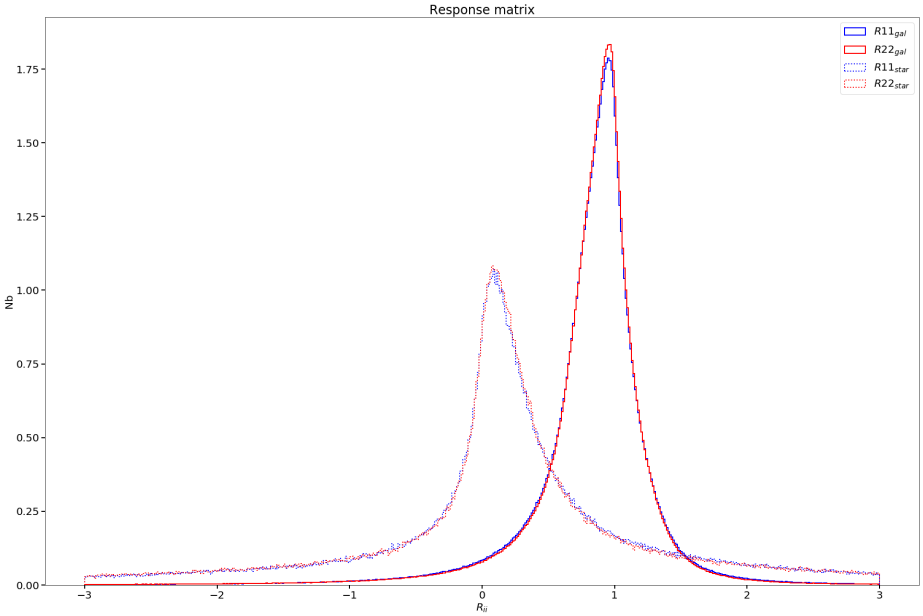


Figure B.2: Diagonal terms of the response martrix for the patch P2. The *solid line* for the galaxies and the *dashed line* for the stars.

while the response for selection effects is:

$$R^{\mathrm{S}} = \begin{pmatrix} -0.143 & 0. \\ 0. & -0.149 \end{pmatrix}. \tag{B.6}$$

Figure B.3 show the diagonal terms for the galaxies and the stars.

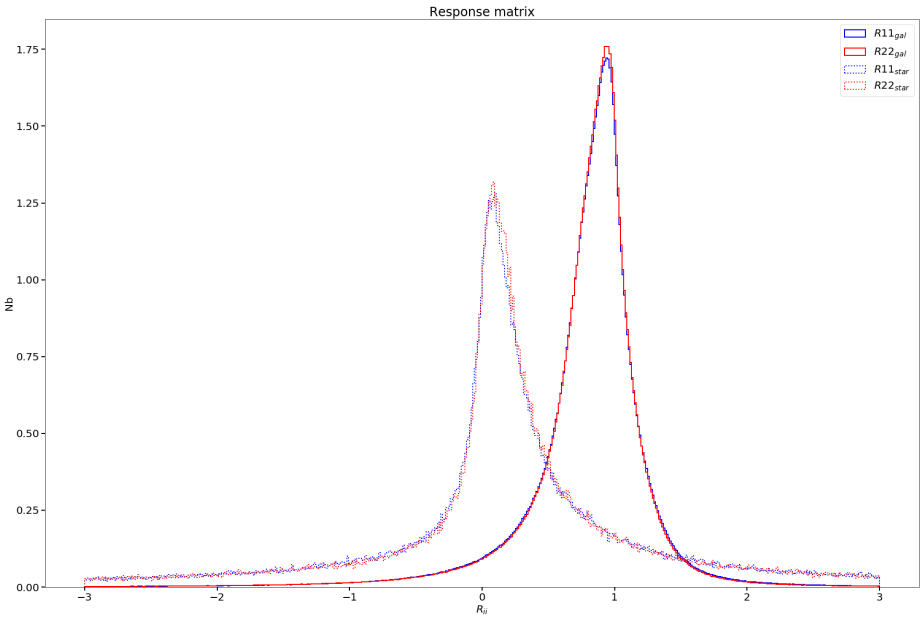


Figure B.3: Diagonal terms of the response martrix for the patch P4. The *solid line* for the galaxies and the *dashed line* for the stars.