Validation of the weak-lensing power spectrum and point-spread function residuals for Euclid

Author: Jingwei WANG

September 21, 2023

Keywords:

Weak Lensing, Galaxy Shear, Two-Point Correlation Function, Point-Spread Function, the Euclid mission

Conducted under the supervision of:

Martin Kilbinger - martin.kilbinger@cea.fr

Acknowledgements

I want to express my gratitude to my tutor, Dr Martin Kilbinger, for providing me with the precious chance to work with him and with this amazing CosmoStat Lab. I will always be grateful for all his kindness and patient guidance throughout the internship. He gave me a lot of strength and encouragement to move forward. I would also thank my dear colleagues, including but not limited to Alex Szapiro, Benjamin Remy, Clara Bonini, Denise Lanzieri, Fabian Hervas Peters, Jennifer Pollack, Jonnyy Torres, Lisa Goh, Samuel Farrens, Sergio La Hera, Tobías Liaudat, Vilasini Tinnaneri Sreekanth, Ziwen Zhang, who gave me strong support in several aspects, both in work (in theoretical physics and in computer/coding techniques) and mentally. They have helped me to integrate into the lab and enjoy the time at CEA. Thank them for letting me be a part of their fantastic legendary cool life. I will always remember the sincerity, humility, idealism, vitality, and openness on each of them. Thank my ancient tutor and friend Alain Coulais for his visiting and for the coffee chats. Additionally, I'd also like to thank some other kind professors and colleagues in CEA. The chats with them are like surprising candies in daily work, they are Bertrand Cordier, Margaret, Marius, René Gastaud, Mathieu Esclingand, etc.

Last but not least, I want to thank my dear friends, Fernando Gutierrez Canales, Clara Grometto, Mila Veliz, Yue Kong, Anne Lorenz, Huichen Zhang, Nabiha Ikram, Mathilde Ferrer, Mostafa Elraies, Zheng Li, etc, who are my support and my source of happiness. I also want to thanks some schoolmates, Juncen Ye, Etienne Bargel, Louis Blazejczak, Pierre Samaha, Joshua Tuckey, Elie Louis, Shuyu Ding, etc, for their help during the semester.

Abstract

This report details my internship in CosmoStat for Euclid from March to August, 2023. Euclid is a space mission that aims to explore the evolution of the dark Universe. It primarily employs two probes: weak gravitational lensing and galaxy clustering. My internship is intricately connected to the work of the Euclid Science Ground Segment (SGS), with a primary focus on the weak gravitational lensing probe.

My project can be divided into two main components:

1) Testing the angular power spectrum calculation. I utilised several programs in parallel with the SGS LE3 pipeline, which generates science-ready data products, to assess the angular power spectrum. After simulating a weak lensing galaxy data set, the angular power spectrum can be achieved through different methods. The outcomes showed a good agreement among them.

2) Utilising diagnostic function, called ρ-statistics, to quantify PSF-model-induced biases on the two-point correlation function. I firstly generated PSF simulations and the counterpart PSF prediction from a data-driven model, including low resolution and super-resolution, with different spectral energy distributions. Then I computed the ρ-statistics from those data sets. Finally, I compared the ρ-statistics with Euclid requirements.

Contents

1. Introduction

This internship was carried out in the CosmoStat Laboratory in the astrophysics division of CEA (Commissariat à l'énergie atomique et aux énergies alternatives), where my main work was on the recently launched international space mission Euclid.

This report contains four main parts. This chapter provides an overview of the host organization of this internship and the Euclid mission; chapter 2 introduces the terms and notions used in the report; the third and fourth parts will respectively discuss the two parts of my work, where chapter 3 is about the calculation and validation of angular power spectrum and chapter 4 is about the analysis of PSF-model-induced bias in the Two Point Correlation Function (2PCF).

1.1. The host organization

CEA is a government-funded research organization established in 1945 and operating under the authority of the French Ministry of the Armed Forces and the Ministry of Higher Education, Research and Innovation. Even though CEA's primary focus is on research and development in fields related to nuclear energy, defense, and various alternative energy sources, it has expanded and now covers a wide range of scientific and technological domains, including information technologies, health and life sciences, climate change and environment protection, matter and universe, etc. To carry out its missions, the CEA relies on four so-called "directions": Direction des énergies, Direction des applications militaires, Direction de la recherche technologique, and Direction de la recherche fondamentale.

The Direction de la Recherche Fondamentale (DRF), composed of 9 institutes in Saclay, Grenoble, and Marseille, focuses on the fundamental study in biology, chemistry, physics, and calculation. At the time of April 2021, there were more than 5000 persons working for DRF, 566 active patents and 179 licences achieved and 38 start-up were created since 2000 [5]. Among all the institutes, Irfu (Institut de recherche sur les lois fondamentales de l'Univers) is the one in the domain of astrophysics, nuclear physics and particle physics [7].

The Département d'Astrophysique (DAp) within Irfu is one of the major laboratories in space astronomy in France, at European and international level. It comprises nearly 200 people, including 130 permanent staff. DAp brings together research engineers and technicians from the CEA's Astrophysics Department within DRF on the Laws of the Universe as well as research engineers from DEDIP (Département d'Electronique, des Détecteurs et d'Informatique pour la Physique, another Irfu department). DAp has contributed to many international missions, including the James Webb Space Telescope, the Kepler space telescope, LISA (Laser Interferometer Space Antenna), etc [3]. DAp, is also an important partner of the Euclid mission, which is the base of this internship.

The CosmoStat Lab, where I did my internship, is a young, active, and international team in DAp, Irfu. As a laboratory, it gathers both cosmologists and computer scientists to work together on developing new methods of statistics, signal processing, and apply them to cosmological data sets. The research topics of CosmoStat include Weak Lensing, Cosmic Microwave Background, Machine Learning, PSF estimation, radio interferometry, etc. It is involved in the various international astronomical projects, including Euclid, UNIONS (the Ultraviolet Near-

Infrared Optical Northern Survey), SKA (Square Kilometre Array), and LSST DESC (Legacy Survey of Space and Time Dark Energy Science Collaboration). As an education unit, CosmoStat offers a very equal, free, open minded, and mutual-helping atmosphere for the younger researchers. The Journal Club, the Learning Slot and the dev meeting happen every week, they help the members to keep up with the newest happening things in the domain, and provide an occasion for the juniors to ask questions of any level. About 20 professors, researchers, engineers, Ph.D. candidates, and interns from the world work, discuss, and develop together.

1.2. The Euclid mission

Euclid is a survey mission of the European Space Agency (ESA) to research the nature of dark energy and dark matter, collaborated by 300 institutes from thirteen European countries, the U.S., Canada, and Japan. As the density of matter and energy is linked to the geometry of the universe, the mission is named after the Greek mathematician Euclid, who founded the subject of geometry.

Euclid is expected to reveal how the Universe expanded and how structures formed over cosmic history, by observing the Universe evolving over the past 10 billion years (to a redshift $z \sim 2$). For that, further insights on the fundamental physical laws and the originate of the Universe will be possible. To achieve its goal, Euclid will map the spatial distribution of matter in the Universe, whose properties are influenced by the expansion of the Universe and the nature of the dark matter. Simplified, a stronger expansion causes less "clumping" of matter and thus a smoother large-scale structure (LSS). Smoother structures may also be expected, to a lesser extent, depending on the properties of dark matter itself, uniquely connecting cosmological scales with micro-physics. By comparing the statistical properties and morphology of the large-scale structure with simulations and theoretical models, we can infer how the Universe expanded, and study how matter structures formed [2]. Euclid will make the 3-D map of the Universe by observing billions of galaxies out to 10 billion light-years, across 15 000 $\deg^2$, more than a third of the sky. The payload module of Euclid comprises two scientific instruments: a visible-wavelength camera (VIS) for visible light with wavelength from 0.55 μm to 0.90 μm, and the Near-Infrared Spectrometer and Photometer (NISP) for near infrared light from 0.95 μm to 2.0 μm.

Euclid was successfully launched on a SpaceX Falcon 9 launch vehicle from Cape Canaveral, Florida, in the U.S., on July 1st, 2023, and arrived one month later at an orbit around the Sun-Earth Lagrange point 2 (L2), 1.5 million kilometers away from Earth in the direction opposite to the sun. This special location offers relatively stable gravitational conditions and relatively minor light influence from the Sun [4].

1.2.1. The probes of Euclid

In the Euclid mission, the primary probes employed to create the map are galaxy clustering and weak gravitational lensing.

Galaxy clustering

Galaxies are luminous tracers of the matter distribution. In addition to a 2-D map on the sky, by detecting the redshift of galaxies, the distances to galaxies can be inferred, thus a 3-D galaxy distribution map can be reconstructed. The inhomogeneity on different scales in those 3-D maps will be quantified through specific statistics of the density fluctuations and

used to constrain model predictions. These predictions depend on key ingredients of the theory describing the Universe, such as the density of dark and standard matter. Euclid will measure the redshift distribution of 35 million galaxies over an enormous volume in the distant Universe from their Hα emission line. This will enable the scientists to get the value of these ingredients to an unprecedented precision, and by comparison between two probes, to check those values.

Weak gravitational lensing

Weak gravitational lensing (short as weak lensing) uses the specific phenomenon that mass bends light rays similarly to optical lenses. The images of far-away galaxies observed are already deflected by the matter structure. Therefore, the distortions images encode the quantity of matter that the light encountered on their way to us, and thus can be used to map the matter distribution and get other statistical parameters. To better evaluate and remove the intrinsic shapes, as many objects as possible are needed. Galaxies, with their relatively high density, high-redshift visibility, and as they are always distorted by gravitational lensing, are usually the objects researched for weak lensing. Euclid will obtain shape measurements of 1.5 billion galaxies in the survey area, with a redshift accuracy better than $0.05(1+z)$ [2].

In conclusion, by observing a huge amount of galaxies from $z \sim 2$ until now, the two complementary probes help construct the distribution of both dark and luminous matter, thus help to better measure the accelerated expansion of the Universe at different cosmological times, and be a strong tool of constraining cosmological models and understanding dark energy.

1.2.2. The Euclid Science Ground Segment

The Science Ground Segment (SGS) refers to an essential part of a space mission's infrastructure that handles the reception, processing, and distribution of the scientific data collected by the satellite or telescope. The Euclid SGS assumes the following data processing levels to define input and output data products, and to follow a global processing scheme:

- Data Level 1. Unpacked and checked telemetry and housekeeping data;
- Data Level 2. Calibrated data and intermediate data products produced during the calibrations, all instrumental fingerprints removed, dithered images co-added, spectra extracted;
- Data Level 3. Science-ready data products, mostly catalogues fulfilling the science objectives of the mission. Level 3 includes pre-defined science data products (3-D galaxy power spectra, dark matter power spectra, etc) and some data analysis (basic two-point statistics and cosmological interpretation);
- External Data. Quality-controlled data from existing missions and ground-based surveys which are used for calibrations, photometric redshift derivations, and simulation validations;
- Data Level S (Simulation). Pre-launch simulations and modelling impacting on calibrations and observing strategies. Post-launch massive Monte Carlo simulations are likely required to assess systematic effects and derive uncertainties on the final cosmological parameters;
- Data Level Q (Quick-release), which the Science Operation Center generates from the raw telemetry. Level Q data form a subset of level 1 (LE1) data which are suitable for distribution to the general scientific community during the proprietary period [20].

1.2.3. Goals of the internship

My host lab CosmoStat is highly involved in the Euclid project at various levels, including management, data analysis and cosmological interpretation. Dr Martin Kilbinger is in charge of validating shear data and processing tools, especially the two-point correlation function (2PCF) method of weak lensing probe in Data Level 3 (LE3). This method is supposed to obtain statistics on base of 2PCF, which is generated from the observed galaxies.

The two goals of my internship are, respectively, to test and validate the 2PCF LE3 calculation and to analyze the bias in the PSF model in 2PCF.

2. Weak gravitational lensing in cosmology

In this chapter, I will provide a general introduction to gravitational lensing, with a specific focus on weak gravitational lensing, as well as statistical methods for analyzing weak lensing data.

Gravitational lensing is a phenomenon predicted by Einstein's theory of general relativity, which describes how the presence of massive objects can bend and distort the path of light as it travels through spacetime. Due to its sensitivity to all massive components in the universe, gravitational lensing can serve as an ideal complement to other probes, particularly in the study of dark energy and dark matter, which are not luminous.

Gravitational lensing distorts the images both in size and shape. When gravitational lensing is strong, the images can be heavily distorted to the extent that multiple images of the same source may appear, or the source images might be transformed into arcs. These phenomena are called strong lensing. In the case of weak lensing, both the distortions and magnifications are subtle, typically on the order of a few percent or less, making them difficult to identify in individual sources, but in a statistical sense. Nevertheless, in the field of cosmology, it is weak lensing, particularly the weak lensing of galaxies as explained in subsection 1.2.1, that draws more interest [11].

2.1. Weak Gravitational Lensing

2.1.1. Deflection angle

To make things easier, the schema in Fig. 2.1 shows the process that a beam of light travels from the source to the observer through one single lensing at distance χ', with the light paths approximated as straight lines. Angle $\boldsymbol{\beta}$ represents the original angle, $\boldsymbol{\theta}$ refers to the observed angle, and $\boldsymbol{\alpha} = \boldsymbol{\theta} - \boldsymbol{\beta}$ refers to the scaled deflection angle of this beam of light.

For general metric that describes an expanding universe including first-order perturbations, the line element $\mathrm{d}s$ is given as below, with $\mathrm{d}l$ represents proper coordinates:

$$\mathrm{d}s^2 = (1 + 2\Phi/c^2)c^2\mathrm{d}t^2 - (1 - 2\Phi/c^2)\mathrm{d}l^2. \tag{2.1}$$

Since light is relativistic, with the propagation condition, $\mathrm{d}s^2 = 0$, so that Eq. (2.1) can be written as $(1 + 2\Phi/c^2)c^2\mathrm{d}t^2 = (1 - 2\Phi/c^2)\mathrm{d}l^2$. In the case of large scale weak lensing, the field

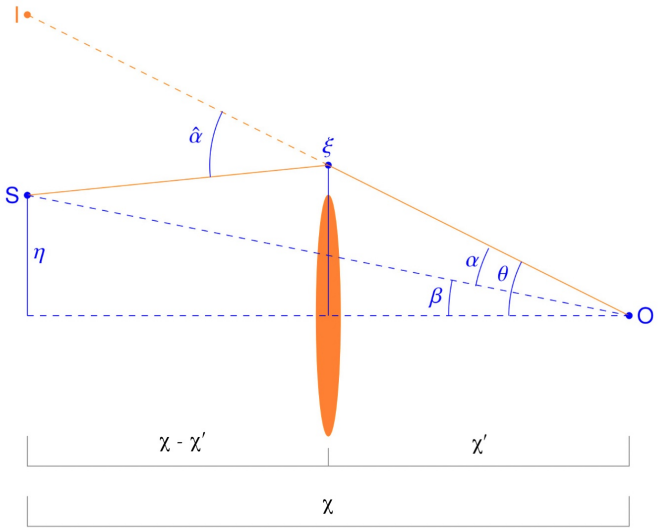


Figure 2.1: Sketch of a gravitational lensing system [9], adapted from [18]

Φ is supposed to be tiny: $\Phi \ll c^2$. Using Taylor expansion, we can get the equation

$$t = \frac{1}{c}\int_{\text{path}} (1 - 2\Phi/c^2)\mathrm{d}l. \tag{2.2}$$

Based on Fermat's principle minimal light travel time $\delta t = 0$,

$$\hat{\boldsymbol{\alpha}} = -\frac{2}{c^2}\int_{\text{Source}}^{\text{Observer}} \nabla_\perp \Phi \mathrm{d}l. \tag{2.3}$$

While taking into consideration of the potential along the light path

$$\mathrm{d}\hat{\boldsymbol{\alpha}} = -\frac{2}{c^2}\nabla_\perp \Phi(\boldsymbol{\eta}, \chi')\mathrm{d}\chi', \tag{2.4}$$

so the distortion of the whole η can be calculated by integral the changes along the path

$$\mathrm{d}\boldsymbol{\eta}(\chi) = \frac{2}{c^2}\int_0^\chi f_K(\chi - \chi')[\nabla_\perp \Phi(\boldsymbol{\eta}, \chi') - \nabla_\perp \Phi^0(\chi')]\mathrm{d}\chi'. \tag{2.5}$$

So that the deflection angle α can be deduced as

$$\boldsymbol{\alpha} = \frac{2}{c^2}\int_0^\chi \frac{f_K(\chi - \chi')}{f_K(\chi)}[\nabla_\perp \Phi(\boldsymbol{\eta}, \chi') - \nabla_\perp \Phi^0(\chi')]\mathrm{d}\chi'. \tag{2.6}$$

2.1.2. Jacobian matrix and shear of galaxy

Provided that the angular size of a source image is much smaller than the characteristic angular scale on which the gravitational field varies, we approximate the equation by substituting the separation vector $\boldsymbol{\eta}$ in the integral by the 0th-order solution, $\boldsymbol{\eta} = f_K(\chi)\boldsymbol{\theta}$. This corresponds to integrating the potential gradient along the unperturbed ray, which is called Born approximation. Thus, we can linearise the lens equation $\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}$, and define the inverse amplification matrix as the Jacobian matrix $\mathcal{A} = \partial\boldsymbol{\beta}/\partial\boldsymbol{\theta}$,

$$\mathcal{A}_{ij} = \frac{\partial \beta_i}{\partial \theta_j} = \delta_{ij} - \frac{\partial \alpha_i}{\partial \theta_j}. \tag{2.7}$$

$\mathcal{A}$ will describe a linear mapping from the observed image to the source, and the inverse Jacobian $\mathcal{A}^{-1}$ describes the local mapping of the source to image. With Eq. (2.6), given that the second term does not depend on the angle $\boldsymbol{\theta}$, and because of

$$\frac{\partial \nabla_\perp \Phi}{\partial \theta_i} = \frac{\partial(\frac{\partial \Phi}{\partial \eta_{j,\chi'}})}{\partial \theta_i} = \frac{1}{f_K(\chi')} \frac{\partial^2}{\partial \theta_i \partial \theta_j} \Phi, \tag{2.8}$$

the derivative of $\boldsymbol{\alpha}$ with respect to $\boldsymbol{\theta}$ can be written as

$$\frac{\partial \alpha_i}{\partial \theta_j} = \frac{2}{c^2} \int_0^\chi \mathrm{d}\chi' \frac{f_K(\chi - \chi')}{f_K(\chi) f_K(\chi')} \frac{\partial^2}{\partial \theta_i \partial \theta_j} \Phi(f_K(\chi)\boldsymbol{\theta}, \chi'). \tag{2.9}$$

In this case, the deflection angle can be written as the gradients of lensing potential, which is defined as:

$$\Psi(\boldsymbol{\theta}, \chi) = \frac{2}{c^2} \int_0^\chi \mathrm{d}\chi' \frac{f_K(\chi - \chi')}{f_K(\chi) f_K(\chi')} \Phi(f_K(\chi)\boldsymbol{\theta}, \chi') \tag{2.10}$$

The Jacobian matrix can also be described by a scalar convergence κ, which presents the difference of size from observed image to the source, and a spin-two[1] value shear $\boldsymbol{\gamma}$, which quantifies the anisotropic stretching, having the influence of turning a circular into an elliptic, as shown in the Fig. 2.2:

$$\mathcal{A} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}. \tag{2.11}$$

Here shows the relation between κ, γ_i and the lensing potential:

$$\boldsymbol{\kappa} = \frac{1}{2}(\partial_1\partial_1 + \partial_2\partial_2)\Psi;\ \gamma_1 = \frac{1}{2}(\partial_1\partial_1 - \partial_2\partial_2)\Psi;\ \gamma_2 = \partial_1\partial_2\Psi. \tag{2.12}$$

If we define the reduced shear $g = g_1 + ig_2 = |g|^{2i\varphi}$, where $g_i = \frac{\gamma_i}{1-\kappa}$, then the Jacobian matrix can be written as this, and the (reduced) shear is also a spin-2 value:

$$\mathcal{A} = (1 - \kappa) \begin{pmatrix} 1 - g_1 & -g_2 \\ -g_2 & 1 + g_1 \end{pmatrix}. \tag{2.13}$$

If there is a circular source with radius R, the observed image after distortion of $\mathcal{A}^{-1}$, the semi-major axis and semi-minor axis will respectively be

$$a = \frac{R}{(1-\kappa)(1-|g|)};\ b = \frac{R}{(1-\kappa)(1+|g|)}.$$

In the way reverse, the shear can be deduced from semi-axis:

$$|g| = \frac{1 - b/a}{1 + b/a}.$$

However, the galaxies are not intrinsically circular, so that the ellipticity of observed image is combined of intrinsic ellipticity and shear. The strategy to estimate the shear is to assume that the intrinsic ellipticities are randomly oriented, and average over many galaxies in closed location. Moreover, the intrinsic shapes of galaxies are not usually simply elliptical, and the

[1] because an ellipse transforms into itself after a rotation by 180°.

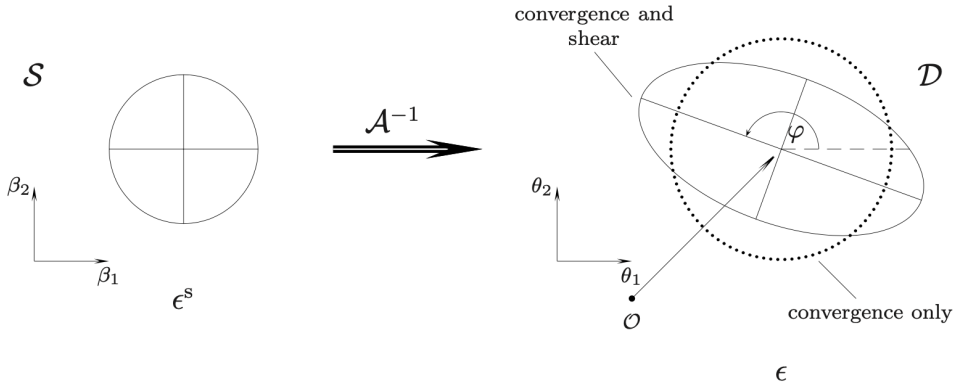


Figure 2.2: The convergence and the shear of a source after distorted by gravitational lensing. Adapted from [22], source: M. Bradac.

accuracy of our observation is always limited, which means the pixel size[2] can not be ignored. For this problem, the scientists defined image ellipticities [14]

$$e^{obs} = \frac{I_{xx} - I_{yy} + 2iI_{xy}}{I_{xx} + I_{yy} + 2\sqrt{I_{xx}I_{yy} - I_{xy}^2}}, \tag{2.14}$$

where I_{ij} is second moments of the intensity profile $I(x, y)$, defined as

$$I_{\mu\nu} = \frac{\int \mathrm{d}x\mathrm{d}y I(x,y)(\mu - \bar{\mu})(\nu - \bar{\nu})}{\int \mathrm{d}x\mathrm{d}y I(x,y)}. \tag{2.15}$$

2.2. Statistical methods of cosmic shear research

From Eq. (2.14), we can obtain the ellipticity e of galaxies from the images observed. The shape e is supposed to include both cosmic shear and their intrinsic shape $\boldsymbol{e}^{\mathbf{obs}} \approx \boldsymbol{e}^{\mathbf{S}} + \boldsymbol{\gamma}$. However, in case that the intrinsic galaxy ellipticities are randomly distributed, the local average ellipticities can be used to estimate the shear $\langle \boldsymbol{e}^{\mathbf{obs}} \rangle = \boldsymbol{\gamma}$. So a galaxy shape survey can give us a general map of the distribution of cosmic shear, and ultimately a map of dark matter. Whereas, in cosmology weak lensing research, we are more interested in statistics such as the correlation function or its the power spectrum [11]. In this subsection, we will introduce the two-point correlation function (2PCF) and the angular power spectrum.

2.2.1. Two-point correlation function

The two components of shear 2PCF are defined as

$$\begin{aligned} \xi_+(\theta) &= \langle \boldsymbol{\gamma}\boldsymbol{\gamma}^* \rangle(\theta) \\ \xi_-(\theta) &= \Re[\langle \boldsymbol{\gamma}\boldsymbol{\gamma} \rangle(\theta) e^{-4i\phi}], \end{aligned} \tag{2.16}$$

where ϕ is the polar angle of a given direction vector θ, and $\xi_\pm$ turns out to be independent of ϕ.

To be more convenient, shapes of galaxies can be decomposed into tangential component and cross-component:

$$\gamma_t = -\Re(\boldsymbol{\gamma} e^{-2i\phi}),\ \gamma_\times = -\Im(\boldsymbol{\gamma} e^{-2i\phi}),$$

[2]For VIS Euclid, pixel size is $0.1''$.

where ϕ refers to the angle between the position vector from point of reference to the galaxy.

Parity conservation implies that $\langle\gamma_t\gamma_\times\rangle = \langle\gamma_\times\gamma_t\rangle = 0$, so Eq. (2.16) can be written with tangential component and cross-component:

$$\begin{aligned}\xi_+(\theta) &= \langle\gamma_t\gamma_t\rangle(\theta) + \langle\gamma_\times\gamma_\times\rangle(\theta),\\ \xi_-(\theta) &= \langle\gamma_t\gamma_t\rangle(\theta) - \langle\gamma_\times\gamma_\times\rangle(\theta).\end{aligned} \tag{2.17}$$

It is a statistical tool that provides insights into the distribution of matter in the universe, particularly the large-scale structure of dark matter.

2.2.2. Angular power spectrum $C(\ell)$

The angular convergence power spectrum is always denoted as $C_\kappa(\ell)$ or $P_\kappa(\ell)$, and quantifies the correlation of the cosmic convergence at different angular separations on the sky. The convergence power spectrum is defined as a Fourier-space ($\tilde{\boldsymbol{\kappa}}(\boldsymbol{\ell})$) two-point correlation function [15]

$$(2\pi)^2\delta_D(\boldsymbol{\ell}-\boldsymbol{\ell}')C_\kappa(\ell) = \langle\tilde{\boldsymbol{\kappa}}(\boldsymbol{\ell})\tilde{\boldsymbol{\kappa}}^*(\boldsymbol{\ell}')\rangle, \tag{2.18}$$

where ℓ refers to modulus, and $\boldsymbol{\ell}$ is 2-D wave vector, the Fourier conjugate of the angular separation $\boldsymbol{\theta}$. Similarly, the angular shear power spectrum can be defined as

$$(2\pi)^2\delta_D(\boldsymbol{\ell}-\boldsymbol{\ell}')C_\gamma(\ell) = \langle\tilde{\boldsymbol{\gamma}}(\boldsymbol{\ell})\tilde{\boldsymbol{\gamma}}^*(\boldsymbol{\ell}')\rangle. \tag{2.19}$$

In many applications, it is convenient to Fourier-transform the correlation function $\xi(\theta)$ to obtain the angular power spectrum $C(\ell) = \int \mathrm{d}^2\theta\boldsymbol{\xi}(\theta)e^{-i\ell\boldsymbol{\theta}}$. After carrying out the integration over the polar angle ϕ, we find:

$$C_\gamma(\ell) = 2\pi\int_0^\infty \mathrm{d}\theta\,\theta\,\xi_+(\theta)J_0(\ell\theta) = 2\pi\int_0^\infty \mathrm{d}\theta\,\theta\,\xi_-(\theta)J_4(\ell\theta)). \tag{2.20}$$

From Eq. (2.12), we notice that the shear angular power spectrum and convergence angular power spectrum are equal ($C_\gamma(\ell) = C_\kappa(\ell)$), so we will not differentiate them later. Since we can never observe and get unlimited 2PCF value as expected, an estimator[3] for the power spectrum will be used instead [24]:

$$\hat{C}(\ell) = 2\pi\int_{\theta_{min}}^{\theta_{max}} \mathrm{d}\theta\,\theta\,[K_+\xi_+(\theta)J_0(\ell\theta) + (1-K_+)\xi_-(\theta)J_4(\ell\theta))], \tag{2.21}$$

where K_+ is a value in the range $[0, 1]$.

2.2.3. Shear tomography

The methods discussed above focused on the method themselves, but the redshift distribution of source galaxies, which determines the source distance distribution, also encodes a lot of information. To extract the third dimension information besides the projected 2-D density contrast, we can separate the source galaxies into several shells according to their redshift, by which we also slide the lensing fields with different redshift.

[3] J_0, J_4 are Bessel functions

If we split the redshift distribution into N bins, we can generate $N(N+1)/2$ power spectra, including N auto-spectra, and $N(N-1)/2$ cross spectra. This allows us to discover the time-varying universe in history, and a 3-D tomographic view of the large-scale structures.

3. Validation of LE3 $C(\ell)$

As introduced in the first chapter, weak lensing is one of the two main probes of Euclid. In the weak lensing part of SGS LE3 (Data Level 3), several different methods generate 2PCF, angular power band, COSEBIs[1], etc, and evaluate cosmological parameters such as $\hat{\Omega}_m$, $\hat{H}_0$ through the IST (Inter-SWG[2] Taskforce). During the internship, I validated the power band output from LE3 2PCF-WL branch, where people firstly summarize the 2PCF $\xi_\pm$ from the observed database, and then generate other functions on the base of $\xi_\pm$. The general idea is to calculate the angular power spectrum or power band for a simulated universe using different programs, and compare their results at the end.

3.1. The simulation of the cosmology

Cosmological parameters

As the creator of this simulated "universe", the first thing to do is to settle the needed basic cosmological parameters, including H_0, Ω_c (dark matter density parameter), Ω_b (baryon density parameter), and the perturbation of the universe. We used the Python library "Cosmology" [26] to accomplish this.

Generating matter

The second step is to distribute matter into the universe which will be used to simulate the cosmic lensing. To cover the field that Euclid observes, we set the redshift range from 0 to 3. This 3-D field with a depth of $z = 3$ is then comoving-distance uniformly divided into shells with a thickness of $\mathrm{d}x = 90$ Mpc. With assuming matter weights to be uniformly distributed in redshift, we compute angular matter power spectra $C(\ell)$ of the shells with the package `glass.camb` from the Python library `glass`[3]. Then, we compute Gaussian $C(\ell)$ for a lognormal field with 3 correlated shells which takes into consideration the correlation angular matter power spectrum between different shells. Then, with the Gaussian angular matter power spectra, we can generate a 3-D tomographic matter distribution.

Generating lensed galaxies

The real galaxy redshift distribution follows a Smail distribution [25], so we use the corresponding function from the package `glass.observations` to generate a sample of Smail-distributed galaxies, giving the density and redshift limit. The Python function used is constructed on base of A. Amara *et al.* [8]:

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

Then, to make further calculation convenient, we divide the sky into ten redshift bins, assuming that the observed redshift are measured with a Gaussian error. Now, we have both matter

[1] COSEBIs: Complete Orthogonal Sets of E-/B-mode Integrals, a method of extracting the full E-/B-mode information from cosmic shear correlation functions [23]

[2] SWG: Science Working Group

[3] version 2023.1: https://github.com/glass-dev/glass/releases/tag/v2023.1.

distribution and galaxy distribution. For every shell of matter, we calculate and add their lensing to all the galaxies effected (which means the galaxies with a further distance to the observer than the shell of matter). As a result, we get the simulation of a large number of galaxies, with their shape $e^{\rm sim} = e^{\rm intrinsic} + \gamma$, their (RA, DEC)[4] position, and a label indicating in which redshift bin they are.

Mapping cosmic shear

The last step is to map the "observed" cosmic shear from the galaxies. The redshift bins compose a new series of shells, from Shell 0 to Shell 9, with redshift from $z = 0$ to $z = 3$. We divide the region of each shell into smaller pixels[5], and put every galaxy into the right pixel and redshift bin. The average of all the galaxy shapes in each pixel is calculated and assumed as the estimated local cosmic shear. After this, we successfully transferred a galaxy data set into ten shell maps of cosmic shear.

3.2. The prediction of $C(\ell)$

On base of the simulated galaxies, four programs were executed including Euclid SGS LE3 2PCF-WL. Their results show a good agreement among them.

Pallas

Pallas (Power-spectrum for kAppa of eLL bAnd eStimator) is a package written by Martin Kilbinger to calculate power spectrum from 2PCF data. It is also the method embedded in Euclid Pipeline LE3. I edited a little bit of the Pallas that it can give both the evaluated power band $\hat{\mathcal{P}}$ and the prediction of $\hat{C}(\ell)$.

The power in band i here is based on the definition in [24], as following:

$$\begin{aligned}\hat{\mathcal{P}}_i :&= \frac{1}{\Delta_i}\int_{\ell_{il}}^{\ell_{iu}} \mathrm{d}\ell\,\ell\,\hat{C}(\ell) \\ &= \frac{2\pi}{\Delta_i}\int_{\theta_{\min}}^{\theta_{\max}} \frac{\mathrm{d}\theta}{\theta}\{[K_+\xi_+(\theta)[g_+(\ell_{iu}\theta) - g_+(\ell_{il}\theta)] + (1-K_+)\xi_-(\theta)[g_-(\ell_{iu}\theta) - g_-(\ell_{il}\theta)]\},\end{aligned} \tag{3.2}$$

where $\Delta_i = \ln \ell_{iu}/\ell_{il}$, is the logarithmic width of the band; and $g_+(x) = xJ_1(x)$, $g_-(x) = (x - 8/x)J_1(x) - 8J_2(x)$. Assuming that all the bands are narrow that $\Delta_i \ll \ell_i$, so the power in band i can be estimated as:

$$\hat{\mathcal{P}}_i \cong \frac{\ell_{il} - \ell_{iu}}{\ln \ell_{iu}/\ell_{il}}\ell_i\,\hat{C}(\ell_i) \cong \frac{\mathrm{d}\ell_i}{\mathrm{d}\ln \ell_i}\ell_i\,\hat{C}(\ell_i) = \ell_i^2\,\hat{C}(\ell_i). \tag{3.3}$$

Then, we can get the power-band-deduced $C(\ell)$:

$$\hat{C}(\ell) = \hat{\mathcal{P}}\,/\,\ell^2. \tag{3.4}$$

The $C(\ell)$ function directly derived from 2PCF in Pallas is based on Eq. (2.21).

In terms of the input ξ_{pm} values, we utilise the 2PCF output of Euclid LE3 2PCF-WL pipeline.

[4]RA: Right ascension [0°, 360°); DEC: Declination (-90°, 90°]

[5]The pixels are generated by HEALPix (Hierarchical Equal Area isoLatitude Pixelation of a sphere), which produces a subdivision of a spherical surface in which each pixel covers the same surface area as every other pixel.

Euclid LE3 2PCF-WL

Euclid LE3 2PCF-WL is the branch that analyses the two-point galaxy correlation functions in the Euclid Data Level 3 weak-lensing pipeline. The input data are one or two shells of galaxies. In the case of one shell, it gives out the auto-correlated calculations (including power band, 2PCF, COSEBIs), otherwise it will calculate cross-correlated parameters. To compare the results, Shell 2 and Shell 5 are used as input data, so that each output file includes two auto-correlation results and one cross-correlation result.

The SGS LE3 2PCF-WL utilises Pallas as the main method but with an adjustment parameter different from that in Pallas, so that the power band output from LE3 actually represents $P_{\rm LE3} = \hat{\mathcal{P}}/2\pi$. Fig. 3.1a is a comparison between Pallas results and LE3 2PCF-WL output power band. It shows very good consistency between the modified Pallas package and Euclid LE3 weak-lensing pipeline.

Using Healpix function Anafast

The other method is to use the function `healpix.sphtfunc.anafast()`. Given shear maps of some shells, `anafast()` can directly get the auto angular power spectrum of one shell or the cross angular power spectrum of two shells. The output is composed as (TT, EE, BB, TE, EB, TB), where the second term is the C_ℓ^{EE} expected, and third term is the C_ℓ^{BB} expected [6].

For this test, we also used Shell 2 and Shell 5, the same as used in LE3 2PCF-WL, and got the cross angular power spectrum C_ℓ.

Using CAMB

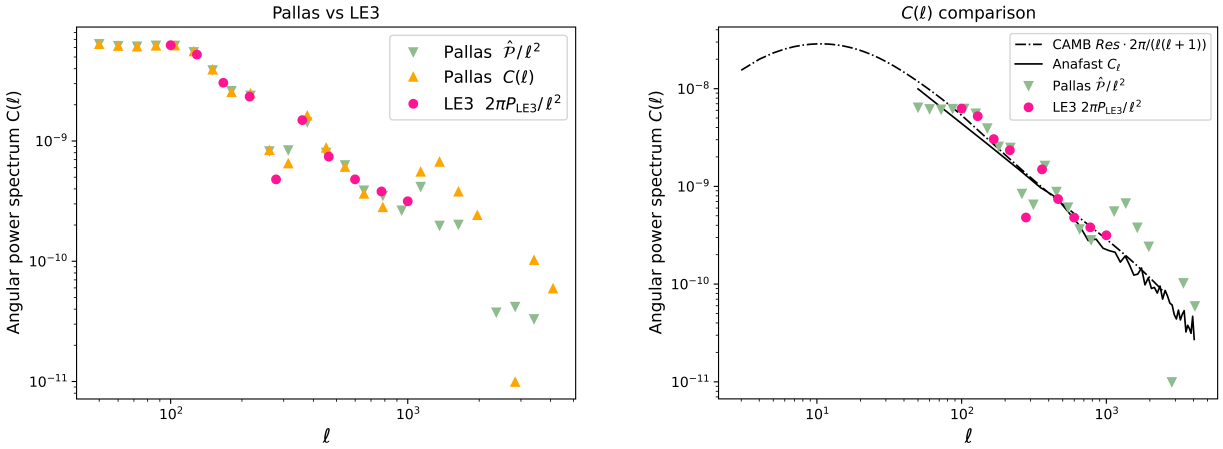
This method is actually utilising CAMB[6] to extract the C_ℓ directly from the basic parameters, but not the simulated galaxies. CAMB is a cosmology code for calculating CMB, lensing, galaxy count, etc. It is available both in Python and Fortran, where we used the Python library. CAMB is widely used in CMB, and we also used it in the simulation.

This process is done by CAMB package, without setting the keyword `raw_cl` to be `True` [1], the default output is $\ell(\ell+1)C_\ell/2\pi$. So the output of this step is the "not-raw C_ℓ" among all ten shells.

The C_ℓ of CAMB represents the true input power spectra, which were used to create the simulation, so that it is supposed to be the "expected" result of the output of other methods.

Fig. 3.1b shows the result comparison among the four methods. The $C(\ell)$ deduced from the output of Pallas and LE3 are consistent with the underlying CAMB power spectrum; the output of Anafast, after a prefactor, also matches the input C_ℓ. The difference is that the estimators are computed from a noisy, pixellated realisation of the simulation; LE3 output is calculated with ℓ ranged in [100,1000].

[6]Code for Anisotropies in the Microwave Background, https://camb.info



(a) Pallas vs. LE3 output (b) $C(\ell)$ Comparison

Figure 3.1: (a) Comparison between Pallas outputs and Euclid outputs; (b) Comparison among the evaluations from simulated galaxy data set and the input C_ℓ.

4. ρ-statistics analysis in Euclid

The optical system and the sensor of a telescope will distort the image observed by diffraction, aberrations, and other optical effects. As a result, instead of being recorded as a perfect point on the sensor, the light of point source is spread out over a small area, creating a blurred image of the point source. The point-spread function (PSF) is a fundamental concept of this blurring effect. The observed image is convoluted by PSF from the original image $I^{\rm obs} = I^{\rm ori} * \text{PSF}$. PSF depends on the design of the instrument and also the position and wavelength distribution of the light source. If a PSF model is well constructed, it can be used for deconvolution techniques to reverse the blurring effect and to estimate the original image, as well as to assess and correct for the limitations of an imaging system.

In the internship, I used the PSF model of Tobías Liaudat [17], which is called WaveDiff. WaveDiff is a data driven modeling framework with a differentiable optical forward model. It takes in consideration both the position and the spectral energy distribution (SED) of the source, and can give out its special predicted PSF. The Figure 4.3a shows the image of observed PSF, predicted PSF, and the residual in one case.

In weak gravitational lensing, the faint galaxies are small, and therefore their observed shape is strongly affected by the point-spread function. The mission Euclid poses stringent requirements on the quality of the PSF model, since the correlation analysis of galaxy shapes is important. To quantify the influence of the PSF model error on the 2PCF analysis, we imported the method of ρ-statistics from M. Jarvis *et al.* (2016) [14].

4.1. WaveDiff PSF model

So far, there are two main approaches of building a PSF model: parametric PSF models, which consist in building a parametric model of the optical system; and data-driven PSF models, which take use of the observed stars and assume some special or chromatic regularity to deduce PSF over the field-of-view [17]. Since stars far away are of small angular size, we can assume the observed bright star images to be the local PSF for the color.

The WaveDiff model of Tobías Liaudat *et al.* is a combination of both the two approaches, keeping the merits of both, and avoiding the problem of the correction for possible errors or disturbances of parametric models, and the problem of under-sample for a space mission of data-driven models. WaveDiff can solve the cases with special variation and chromatic variations of the source light. Moreover, the WaveDiff model offers two resolutions options.

My mission was to analyse the ρ-statistics for one trained Euclid PSF model generated by WaveDiff. Therefore, the PSF prediction and the weak-lensing related measurement were given by WaveDiff. WaveDiff is available to public in GitHub named as "wf-psf"[1]. I used a version modified by myself[2], based on the branch `dummy_main`, which is maintained by Jennifer Pollack.

4.2. The theory of ρ-statistics

ρ-statistics is a diagnostic method to evaluate PSF models. As described in Chapter 2, cosmic shear is an important goal of data processing, and for a large number of galaxies, the average of shear can represent the local cosmic shear ($\langle e\rangle = g \equiv \frac{\gamma}{1-\kappa}$). In this section, we take systematical biases into consideration, $\langle e\rangle = (1+m)g_{\rm true} + \alpha e_{\rm PSF} + c$, where m and c quantify respectively multiplicative error and additive error, and α represents the leakage of the PSF shape into the galaxy shapes. M. Jarvis et al. used the approximation of Paulin-Henriksson *et al.* [19], which quantifies the bias of PSF model error on the measurement of galaxy ellipticity:

$$\delta e_{sys} = (e - e_{\rm PSF})\left(\frac{T_{\rm PSF}}{T_{\rm gal}}\right)\frac{\delta T_{\rm PSF}}{T_{\rm PSF}} - \left(\frac{T_{\rm PSF}}{T_{\rm gal}}\right)\delta T_{\rm PSF}\,, \tag{4.1}$$

where $T_{\rm gal}$ is the intrinsic galaxy size, and T is the intensity-weighted second moment of the light distribution:

$$T \equiv I_{xx} + I_{yy},$$

where $I_{\mu\nu}$ is defined in Eq. (2.15).

With Eq. (4.1), the systematic error in shear 2PCF ξ_+ in terms of PSF model error is:

$$\begin{aligned}\delta\xi_+(\theta) =& 2\Big\langle\frac{T_{\rm PSF}}{T_{\rm gal}}\frac{\delta T_{\rm PSF}}{T_{\rm PSF}}\Big\rangle\xi_+(\theta) + \Big\langle\frac{T_{\rm PSF}}{T_{\rm gal}}\Big\rangle^2\rho_1(\theta) - \alpha\Big\langle\frac{T_{\rm PSF}}{T_{\rm gal}}\Big\rangle\rho_2(\theta)\\ &+ \Big\langle\frac{T_{\rm PSF}}{T_{\rm gal}}\Big\rangle^2\rho_3(\theta) + \Big\langle\frac{T_{\rm PSF}}{T_{\rm gal}}\Big\rangle^2\rho_4(\theta) - \alpha\Big\langle\frac{T_{\rm PSF}}{T_{\rm gal}}\Big\rangle\rho_5(\theta),\end{aligned} \tag{4.2}$$

where M. Jarvis *et al.* defined[3] $\rho_i(\theta)$ as:

$$\rho_1(\theta) \equiv \langle\delta e^*_{\rm PSF}(\mathbf{x})\delta e_{\rm PSF}(\mathbf{x}+\theta)\rangle \tag{4.3}$$

$$\rho_2(\theta) \equiv \langle e^*_{\rm PSF}(\mathbf{x})\delta e_{\rm PSF}(\mathbf{x}+\theta)\rangle \tag{4.4}$$

$$\rho_3(\theta) \equiv \Big\langle\Big(e^*_{\rm PSF}\frac{\delta T_{\rm PSF}}{T_{\rm PSF}}\Big)(\mathbf{x})\Big(e_{\rm PSF}\frac{\delta T_{\rm PSF}}{T_{\rm PSF}}\Big)(\mathbf{x}+\theta)\Big\rangle \tag{4.5}$$

$$\rho_4(\theta) \equiv \Big\langle\delta e^*_{\rm PSF}(\mathbf{x})\Big(e_{\rm PSF}\frac{\delta T_{\rm PSF}}{T_{\rm PSF}}\Big)(\mathbf{x}+\theta)\Big\rangle \tag{4.6}$$

$$\rho_5(\theta) \equiv \Big\langle e^*_{\rm PSF}(\mathbf{x})\Big(e_{\rm PSF}\frac{\delta T_{\rm PSF}}{T_{PSF}}\Big)(\mathbf{x}+\theta)\Big\rangle. \tag{4.7}$$

[1] https://github.com/CosmoStat/wf-psf
[2] https://github.com/Jeanne-jw/wf-psf
[3] $\rho_{1,2}$ are cited from Rowe [21]

The ρ-statistics quantify the correlation between PSF prediction residuals ($\delta e_{\rm PSF}$ and $\delta T_{\rm PSF}$), or including the PSF leakage (parameterised with the coefficient α) in the case of $\rho_{2,5}$. If we set the limit of PSF-model-reduced ξ_+ bias as $\delta\xi_+^{\rm PSF}(\theta)$, the ρ_is are part of additive terms that should be set into less than $\frac{1}{2}\delta\xi_+^{\rm PSF}(\theta)$[4], so that the integral bias can meet the requirement. Based on the equations (4.4-4.7), the requirements for ρ_i can be written as below:

$$|\rho_{1,3,4}(\theta)| \;<\; \frac{1}{2}\left\langle\frac{T_{\rm PSF}}{T_{\rm gal}}\right\rangle^{-2}\delta\xi_+^{\rm PSF}(\theta), \tag{4.8}$$

$$|\rho_{2,5}(\theta)| \;<\; \frac{1}{2}|\alpha|^{-1}\left\langle\frac{T_{\rm PSF}}{T_{\rm gal}}\right\rangle^{-1}\delta\xi_+^{\rm PSF}(\theta). \tag{4.9}$$

4.3. The PSF model and data

Since the real data of Euclid are not released yet, I used simulated data as "observed" data to generate the prediction and measure PSF shapes through a WaveDiff PSF model. Alex Szapiro, a colleague in CosmoStat Lab, trained the WaveDiff model using 52 000 simulated star images and prepared an additional 20 000 simulated star images for testing. I employed these 20 000 test data as "observed" stars, thus "observed" PSFs, for the ρ-statistics analysis to assess the accuracy of his trained model.

In this subsection, I will introduce the important parameters of the simulated data and the WaveDiff output data.

4.3.1. Simulated PSF data

For each simulated star, its observed position, the images of two resolutions, the SED (spectral energy distribution), etc are available.

The observed **position** for the simulated stars are not in the coordinates on the sky, as RA and DEC that correspond to longitude and latitude on sky. Instead, they refer to the star's coordinate on the focal plane, in relative scales ranged in $[0, 1000]$. The test data are randomly distributed on the entire focal plane.

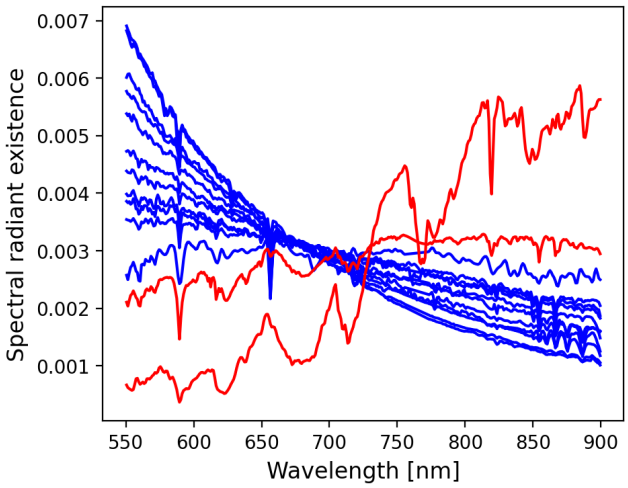


Figure 4.1: All the possible SEDs in the samples. Red curves are assumed to be relatively red, others are relatively blue.

[4]In the reference [14], there is no prefactor $\frac{1}{2}$.

The **SEDs** of the stars are also included in the simulation data set. As shown in Fig. 4.1, the wavelengths are limited between 550 nm and 900 nm, as Euclid is sensitive to this light band. The SEDs of this simulation sample are set among 13 types, where generally speaking, two of them seem "redder" than the others.

There are two **resolutions** of images available, including low resolution, with 32×32 pixels, which is the original resolution of Euclid, and super-resolution, composed with 64×64 pixels. The images of two resolutions of one star are shown in Fig. 4.2. A super-resolution image from real data can be obtained by combining multiple exposures close in time and sky position.

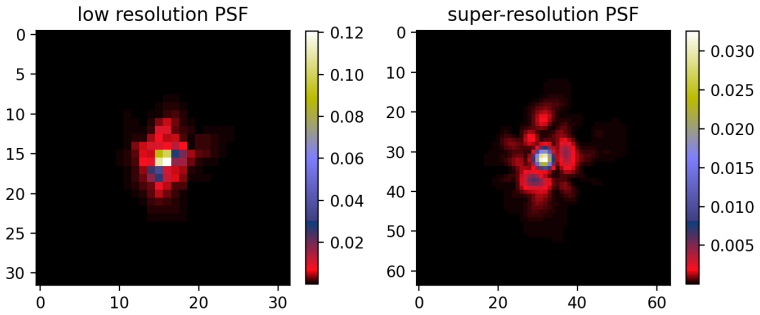


Figure 4.2: Low resolution and super-resolution images of one star. The axes show image position in pixel; colors represent the intensity of light in each pixel, where the summation upon one whole image is 1.

4.3.2. The output of WaveDiff model

As the equations in the subsection 4.2 show, the essential information of the PSF model are the ellipticities (e_{PSF}) and second moments (T_{PSF}) of observed PSF and predicted PSF. Therefore, I modified the pipeline and added some metrics functions [16], so that the output data can include: predicted PSF image, the ellipticities (e_1, e_2) of both simulated PSF image (which is the star image) and predicted PSF image, the second moment of the two images, and the flag indicating the existence of the values mentioned.

The parameters are calculated by the HSM module of the Python library GalSim [13]. It needs to be aware that e_1 and e_2 used in ρ-statistics actually refer to g_1 and g_2 in GalSim; the size T_{PSF} is given by the dispersion of the light distribution as $2\sigma_{\mathrm{I}}^2$. For some samples, the star image or the predicted PSF pattern is too small for the HSM module to compute the expected parameters. So an output parameter "Flag" is added to note if the values are calculated successfully. For the samples with "Flag" equal to 0, es and Ts are automatically set as 0 by the HSM module, and are meaningless for further analysis. In this case, they will be discarded.

4.4. PSF data collection and ρ calculation

As mentioned earlier, we can categorize the data into low resolution and super-resolution based on resolution, and we can classify the simulated PSFs into "blue" and "red" based on different SEDs. Furthermore, by reconstructing monochromatic PSF images, we can even test the idealistic cases where the sources are monochromatic with different wavelengths.

In this subsection, I will explain the data collection in different cases to compare the results, and the calculation of ρ-statistics.

4.4.1. Low resolution and super-resolution images

To cover the possibility of both two resolutions in Euclid research, the PSF and shear-related parameters for both of low and super-resolutions are predicted and calculated.

Without distinguishing chromatic diversity of the simulated stars, we can predict the PSF images of all the samples through the WaveDiff model, either in low resolution, or in super-resolution.

Low resolution PSF

Using low resolution, more than half of the test data have to be discarded, because of the images are too small to be successfully calculated by the HSM module. As for our test data set, only 7429 out of 20 000 given samples work as expected and provide needed values. All the calculation for low resolution correspond to the 7429 effective samples. Fig. 4.3a shows a simulated PSF, the predicted PSF and the residual of prediction, in low resolution.

As for shape parameters, an ellipticity map can be generated. Every bar is decided by a position angle, which refers to the angle between the arrow and x-axis, and a bar length. Here, the position angle is $\arctan(e_2/e_1)/2$, and the length of bar is proportional to $\sqrt{e_1^2 + e_2^2}$. Fig. A.1a shows the ellipticities of observed and predicted PSF, and the residual between them. They match well each other, both in position angular and the length. With the ellipticity comparison, we suppose that the PSF prediction and the calculation of shapes are reliable.

Super-resolution PSF

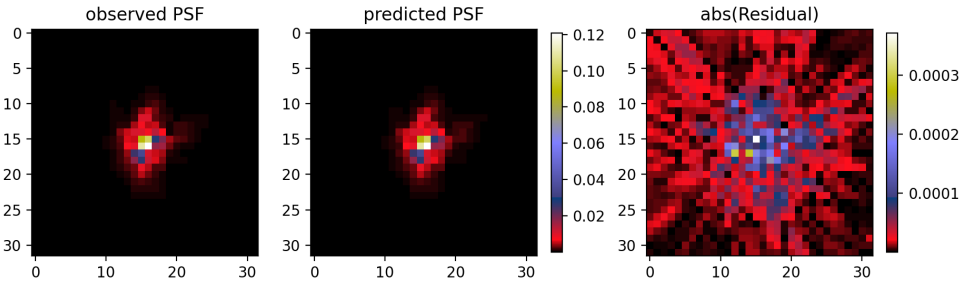
Super-resolution images are composed of 64×64 pixels, which is more accurate than the original Euclid resolution, so the shapes are clearer then low resolution (as shown in Fig. 4.3b), so more stars are calculable with HSM module to get shape parameters. In our case, the ellipticity of every star is available and usable. As a result, population of the sample for super-resolution is 2.7 times larger than that for low resolution calculation. Fig. A.1b shows the ellipticity map of super-resolution images. Obviously, the samples spread in the whole field of view, other than cluster in some region as the case in low resolution.

In addition, I also wanted to compare the ellipticity results between low resolution and super-resolution PSFs. Fig. A.1c shows the difference in ellipticities between low resolution and super resolution PSFs on the overlapped samples. This map indicates the need for further discussion regarding the methods of calculating ellipticities and the choice of resolution

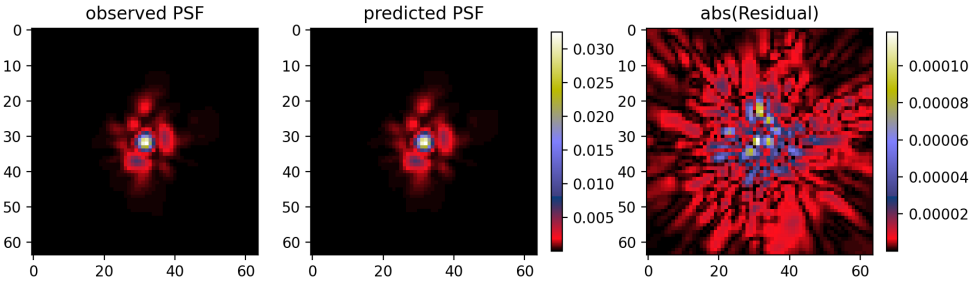
ρ calculation

Based on the output data, we can calculate ρ-statistics. The script used to do the calculation is modified from the Python library "Stile"[5] by Tobías Liaudat, which makes the functions match Jarvis' definitions. What is needed is: 1) to add one more type choice as "Euclid". So that the script can identify the source of data, and correctly read the data; 2) to generate a data files in format of .fits and matching input data model, from the output WaveDiff. After a successful calculation, a .npy file including ρ-statistics results will be saved. This step is similar for different PSF data sets, so will not be repeated in the following parts.

[5]Stile: the Systematics Tests In LEnsing pipeline. https://github.com/msimet/Stile



(a) low resolution



(b) super-resolution

Figure 4.3: Observed (simulated) PSF, predicted PSF, and residual of the prediction.

4.4.2. PSF of different light wavelengths

Divide the data set into red and blue

As introduced above, the PSF varies as a function of wavelength; and in the test samples, there are 13 modes of SEDs among which two SEDs are redder than the others (shown in Fig. 4.1). Therefore, in aim to test the difference in ρ-statistics on varying the color of source, we can divide the dataset into "red" and "blue" two groups.

The "red" group includes 3111 simulated PSF images with the two reddest SEDs. The "blue" group covers all the rest, with 16 889 samples in total. Using super-resolution image data, all the samples will be useable, so the calculation of ρ-statistics will be based on one sample with 3111 PSF and another sample with 16 889 PSF images. In case of low resolution, there are only 1287 "red" samples and 6142 "blue" samples available.

Generate monochromatic PSFs

To better compare the influence of light wavelength, I modified the WaveDiff pipeline and made it possible to reconstruct monochromatic PSF simulation. Since VIS Euclid is sensitive between $0.55\,\mu$m and $0.90\,\mu$m, I generated the simulated monochromatic PSF images and predicted monochromatic PSF by the trained WaveDiff model with wavelength equal to $0.55\,\mu$m and $0.90\,\mu$m, as shown in Fig. 4.4.

4.4.3. ρ-statistics and Euclid requirement

Now, with the ρ-statistics results covering many cases, we can move on to analyse the PSF model effectiveness by comparing our result with the Euclid requirements.

From Table 1 and Table 11 in Euclid Consortium Document [10], the pessimistic requirement for σ_8 prediction from weak lensing is $\delta\sigma_8 = 0.019$, and fiducial value σ_8^{fid} is 0.816, which is the

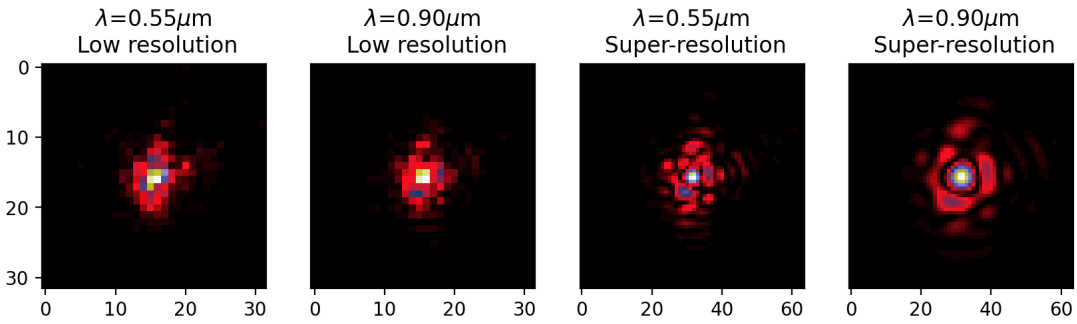


Figure 4.4: Monochromatic PSF images, with $\lambda = 0.55\,\mu$m or $0.90\,\mu$m, in low resolution or super-resolution

best-fit value measured by the Planck Collaboration.

We can propagate σ_8 requirement to the shear correlation functions:

$$\delta\xi_+^{\max}(\theta) \sim \left|\frac{\partial\xi_+(\theta)}{\partial\sigma_8}\delta\sigma_8\right|, \tag{4.10}$$

which can be approximated by calculating the variation of $\xi_+(\theta)$ given $\sigma_8 = \sigma_8^{\rm fid} \pm \frac{\delta\sigma_8}{2}$. For the limit of PSF-model-reduced error:

$$\delta\xi_+^{\rm PSF}(\theta) \sim x\delta\xi_+^{\max}(\theta), \tag{4.11}$$

where we chose two values for x, $x = 0.3$ or $x = 0.05$, which correspond to a maximum contribution to the statistical error on σ_8 from PSF model uncertainties of 30% and 5% respectively.

With Eqs. (4.8-4.9), we can get the requirement of $\rho_i(\theta)$,

$$|\rho_{1,3,4}(\theta)| < \frac{1}{2}\left\langle\frac{T_{\rm PSF}}{T_{\rm gal}}\right\rangle^{-2} x\delta\xi_+^{\max}(\theta); \tag{4.12}$$

$$|\rho_{2,5}(\theta)| < \frac{1}{2}x|\alpha|^{-1}\left\langle\frac{T_{\rm PSF}}{T_{\rm gal}}\right\rangle^{-1} x\delta\xi_+^{\max}(\theta), \tag{4.13}$$

where the Euclid PSF leakage is not known yet, but $\alpha = 0.03$ is a reasonable value measured in previous surveys. For the partition of PSF size and galaxy size, the galaxy effective radius is fixed to $0.7''$, following [12] see their Fig. 6; in the simulated PSF images, 6-7 central pixels can cover 50% flux, where the length of one pixel is $0.1''$, so that $\left\langle\frac{T_{\rm PSF}}{T_{\rm gal}}\right\rangle \sim \frac{7\times0.1^2}{\pi(0.7)^2} \sim 0.05$.

To make the reference of the ρ limits more convenient, I define

$$\rho_{1,3,4}^{\max}(\theta) \sim \frac{1}{2}\left\langle\frac{T_{\rm PSF}}{T_{\rm gal}}\right\rangle^{-2}\delta\xi_+^{\max}(\theta); \tag{4.14}$$

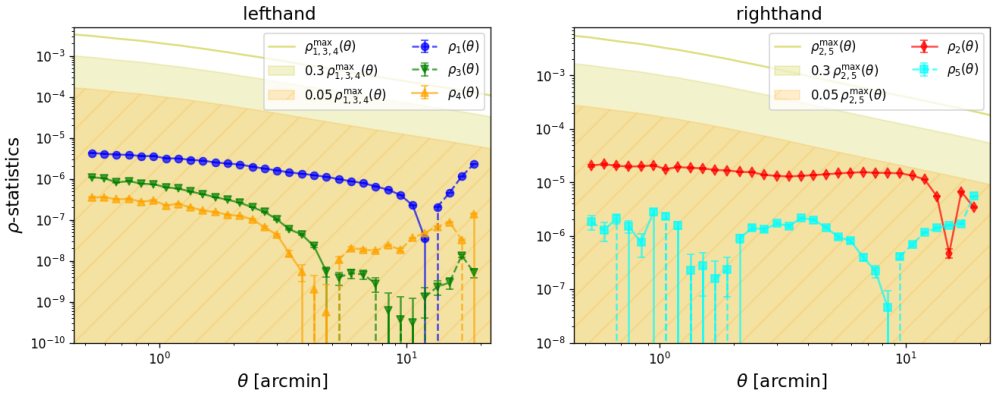
$$\rho_{2,5}^{\max}(\theta) \sim \frac{1}{2}|\alpha|^{-1}\left\langle\frac{T_{\rm PSF}}{T_{\rm gal}}\right\rangle^{-1}\delta\xi_+^{\max}(\theta), \tag{4.15}$$

such that the ρ requirement in Eqs. (4.12-4.13) can be simplified as

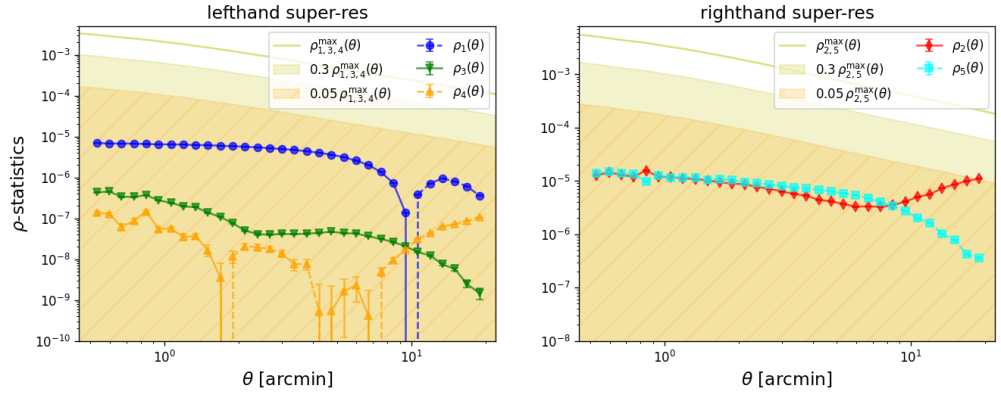
$$|\rho_{1,3,4}(\theta)| < x\rho_{1,3,4}^{\max}(\theta);\ |\rho_{2,5}(\theta)| < x\rho_{2,5}^{\max}(\theta).$$

Fig. 4.5 plots the comparison between ρ requirements and the simulated ρ-statistics, in the cases of low resolution and super-resolution. The error of PSF model fits well the requirements up

to scales of $\theta \sim 20$ arcminute. The results of other cases are available in Appendix A. Fig. A.2 compares the ρ-statistics of relatively blue sources and relatively red sources. Both colors and both resolutions see a good match with the requirements. Fig. A.3 compares idealistic situations of monochromatic source with wavelength $\lambda = 0.55\,\mu$m and $\lambda = 0.90\,\mu$m.



(a) low resolution



(b) super-resolution

Figure 4.5: ρ-statistics v.s. the requirements for low resolution and super-resolution PSF.

The $\rho^{\max}(\theta)$ curves fall off towards larger scales, because the cosmic-shear correlation decreases with distance, the large-scale structure becomes less coherent to larger scales. The $\rho_i(\theta)$ curves do not obviously decrease to larger scales. In some cases, it looks like some of them will go beyond the requirements (for example, ρ_2). However, that behavior will most likely change when we combine many images of the focal plane, not just one. Then we will have correlations across images, where the effect of position on the focal plane will decrease, so do the value of ρ_i statistics. The low resolution PSF prediction of $\lambda = 0.55\,\mu$m sources (Fig A.3a) shows a big gap to fit the requirements. If monochromatic analysis is needed (since it's for idealistic cases, it should not happen in real data) and if the combination of images does not help, there should be further study on the low-resolute monochromatic PSF models.

According to all the figures, super-resolution model shows a sightly better performance in ρ-statistics than low resolution model, but in some cases ρ_1 of super-resolution model is not better than the other. From the super-resolution PSF, we get more information and therefore might obtain a more accurate model. However, it also shows more complex features, which makes fitting more difficult.

According to chromatic-related comparisons (Figs. A.2 and A.3), red sources see a better performance than blue sources. This might can be explained by their smoother PSF patterns.

5. Internship conclusion

The objectives of my training revolved around validating the outputs of the 2PCF branch of the Euclid Data Level 3 weak-lensing pipeline and analysing the ρ-statistics of a WaveDiff PSF model.

For the first goal, the outcomes exhibited a notable level of concurrence across various methods. My work contributed to an important step towards the science validation of the Euclid LE3 weak-lensing pipeline, which helped to get ready for the first Euclid science data arriving in around 6 months. However, there is still a concern: the Anafast approach requires a multiplication factor of 360 to agree with the other methods. Despite reviewing many Anafast documents, this discrepancy remains unresolved. That could potentially be attributed to a lack of normalization or a unit disparity that has yet to be identified.

With regard to the second subject, it is evident that the ρ-statistics methodology can be applied to the Euclid model. The diagnostics of the ρ-statistics will be an important tool to compare different PSF models and to assess their accuracy in terms of requirements on cosmological parameter measurements from Euclid weak lensing. In the future, with actual Euclid data, we can directly employ the existing programs and procedures for evaluation.

In term of my personal development, I learned both theoretical and technical knowledge. I studied weak lensing concepts, point-spread function model methods to move better understand my work. I also used various Python libraries related to cosmology, gained experience in creating installable Python packages, and learned how to run programs on the calculation center, including GPU and HPC in a large and complex pipeline environment. I am grateful to all my colleagues for their support, which greatly contributed to the progress of my work. While I encountered challenges, my tutor, as well as other professors and colleagues, provided invaluable assistance.

Bibliography

[1] Camb calculation results online document, 2023. https://camb.readthedocs.io/en/latest/results.html#camb.results.CAMBdata.get_source_cls_dict.

[2] Core science: cosmology, 2023. https://www.euclid-ec.org/public/core-science/.

[3] Département d'astrophysique cea saclay, 2023. https://irfu.cea.fr/dap/index.ph.

[4] Euclid overview, 2023. https://www.esa.int/Science_Exploration/Space_Science/Euclid_overview.

[5] Fabrique de savoirs, 2023. https://www.cea.fr/drf/Pages/La-DRF.aspx.

[6] healpy.sphtfunc.anafast document, 2023. https://healpy.readthedocs.io/en/latest/generated/healpy.sphtfunc.anafast.html.

[7] Institut de recherche sur les lois fondamentales de l'univers, 2023. https://irfu.cea.fr.

[8] Adam Amara and Alexandre Réfrégier. Optimal surveys for weak-lensing tomography. *Monthly Notices of the Royal Astronomical Society*, 381(3):1018–1026, oct 2007.

[9] Matthias Bartelmann and Matteo Maturi. Weak gravitational lensing, 2016.

[10] Euclid Collaboration. Euclid preparation: Vii. forecast validation for euclid cosmological probes. *A&A*, 642:A191, oct 2020.

[11] Scott Dodelson and Fabian Schmidt. 13 - probes of structure: lensing. In Scott Dodelson and Fabian Schmidt, editors, *Modern Cosmology (Second Edition)*, pages 373–399. Academic Press, second edition edition, 2021.

[12] Euclid Collaboration. Euclid preparation - xiii. forecasts for galaxy morphology with the euclid survey using deep generative models. *A&A*, 657:A90, 2022.

[13] GalSim-developers. The hsm module, 2019. https://galsim-developers.github.io/GalSim/_build/html/hsm.html.

[14] M. Jarvis, E. Sheldon, J. Zuntz, and T. et al. Kacprzak. The DES Science Verification weak lensing shear catalogues. *Monthly Notices of the Royal Astronomical Society*, 460(2):2245–2281, may 2016.

[15] Martin Kilbinger. Cosmology with cosmic shear observations: a review. *Reports on Progress in Physics*, 78(8):086901, jul 2015.

[16] Tobías Liaudat and Ezequiel Centofanti. wf-psf, 2022. https://github.com/Jeanne-jw/wf-psf/tree/dummy_main.

[17] Tobías Liaudat, Jean-Luc Starck, Martin Kilbinger, and Pierre-Antoine Frugier. Rethinking data-driven point spread function modeling with a differentiable optical model. *Inverse Problems*, 39(3):035008, feb 2023.

[18] Ramesh Narayan and Matthias Bartelmann. Lectures on gravitational lensing, 1997.

[19] S. Paulin-Henriksson, Asim Amara, L. Voigt, A. Refregier, and Sarah Bridle. Point spread function calibration requirements for dark energy from cosmic shear. *http://dx.doi.org/10.1051/0004-6361:20079150*, 484, 06 2008.

[20] Giuseppe D. Racca and et al. René Laureijs. The euclid mission design. In *SPIE Proceedings*. SPIE, Jul 2016.

[21] Barnaby Rowe. Improving PSF modelling for weak gravitational lensing using new methods in model selection. *Monthly Notices of the Royal Astronomical Society*, feb 2010.

[22] P. Schneider. Weak gravitational lensing. In *Saas-Fee Advanced Courses*, pages 269–451. Springer Berlin Heidelberg, 2006.

[23] P. Schneider, T. Eifler, and E. Krause. COSEBIs: Extracting the full e-/b-mode information from cosmic shear correlation functions. *Astronomy and Astrophysics*, 520:A116, sep 2010.

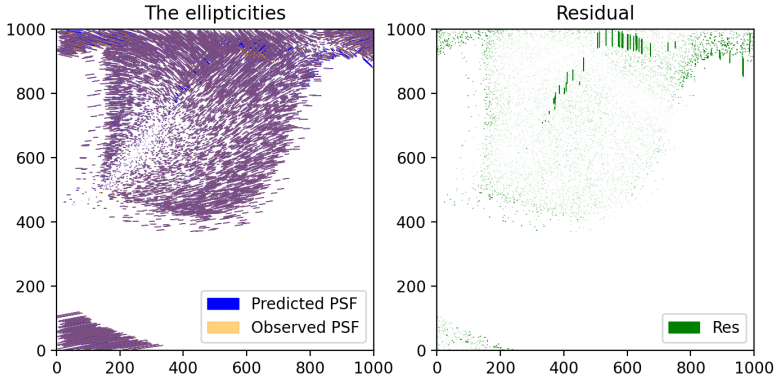
[24] Schneider, P., van Waerbeke, L., Kilbinger, M., and Mellier, Y. Analysis of two-point statistics of cosmic shear - i. estimators and covariances. *A&A*, 396(1):1–19, 2002.

[25] Ian Smail, Richard S. Ellis, and Mike J. Fitchett. Gravitational lensing of distant field galaxies by rich clusters – I. Faint galaxy redshift distributions. *Monthly Notices of the Royal Astronomical Society*, 270(2):245–270, 09 1994.

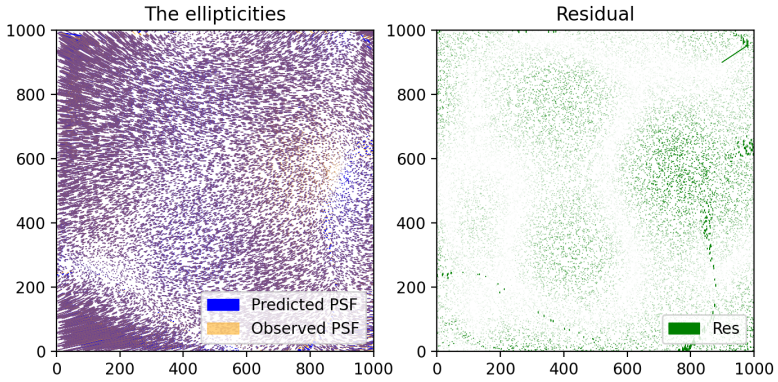
[26] Nicolas Tessore. cosmology 2022.10.9, 2022. `https://pypi.org/project/cosmology/`.

A. Plots

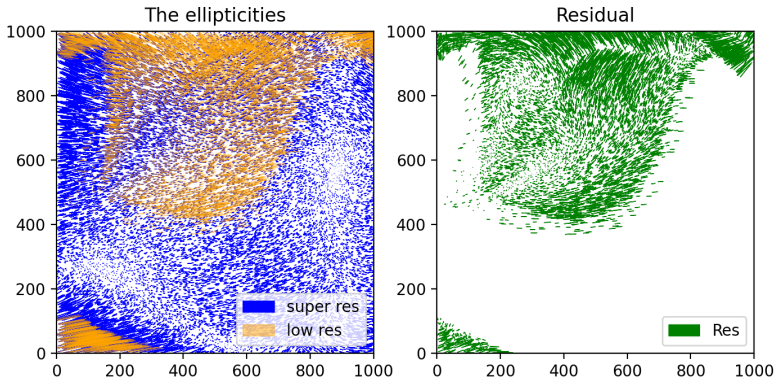
Figure A.1: Maps of ellipticities, x, y axes represent relative positions in the field of view. (a),(b) show the comparison between simulated PSF and predicted PSF ellipticities; (c) compare low resolution and super-resolution.



(a) Low resolution: (Left) maps of ellipticities; (right) ellipticity residual



(b) Super-resolution: (Left) maps of ellipticities; (right) ellipticity residual



(c) Galaxy ellipticities from low resolution and from super-resolution data

Figure A.2: ρs statistics for blue source and red source: (a) blue, low resolution;(b)red, low resolution; (c)blue, super-resolution; (b)red, super-resolution. ρ statistics in all 4 cases fit the requirements.

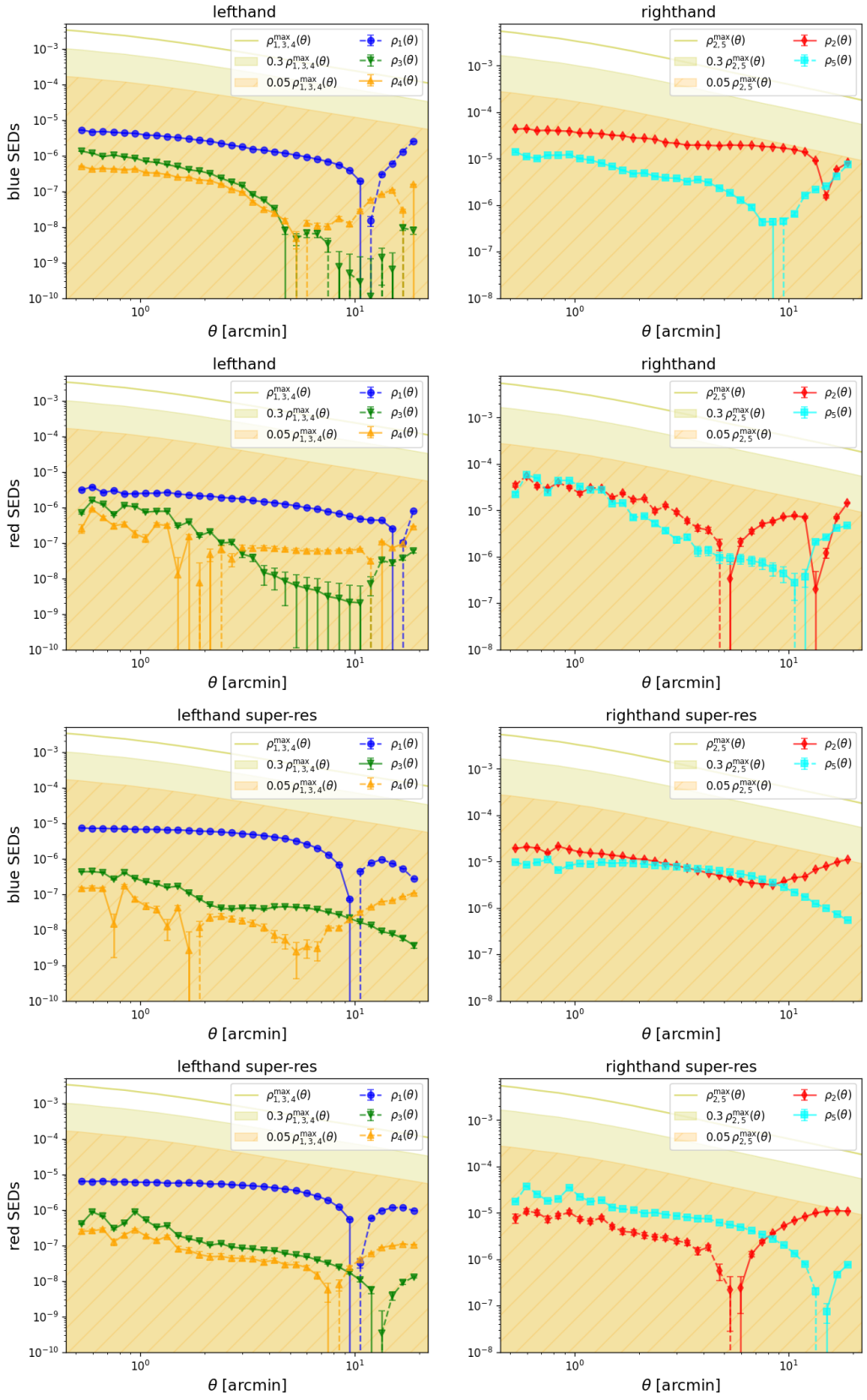
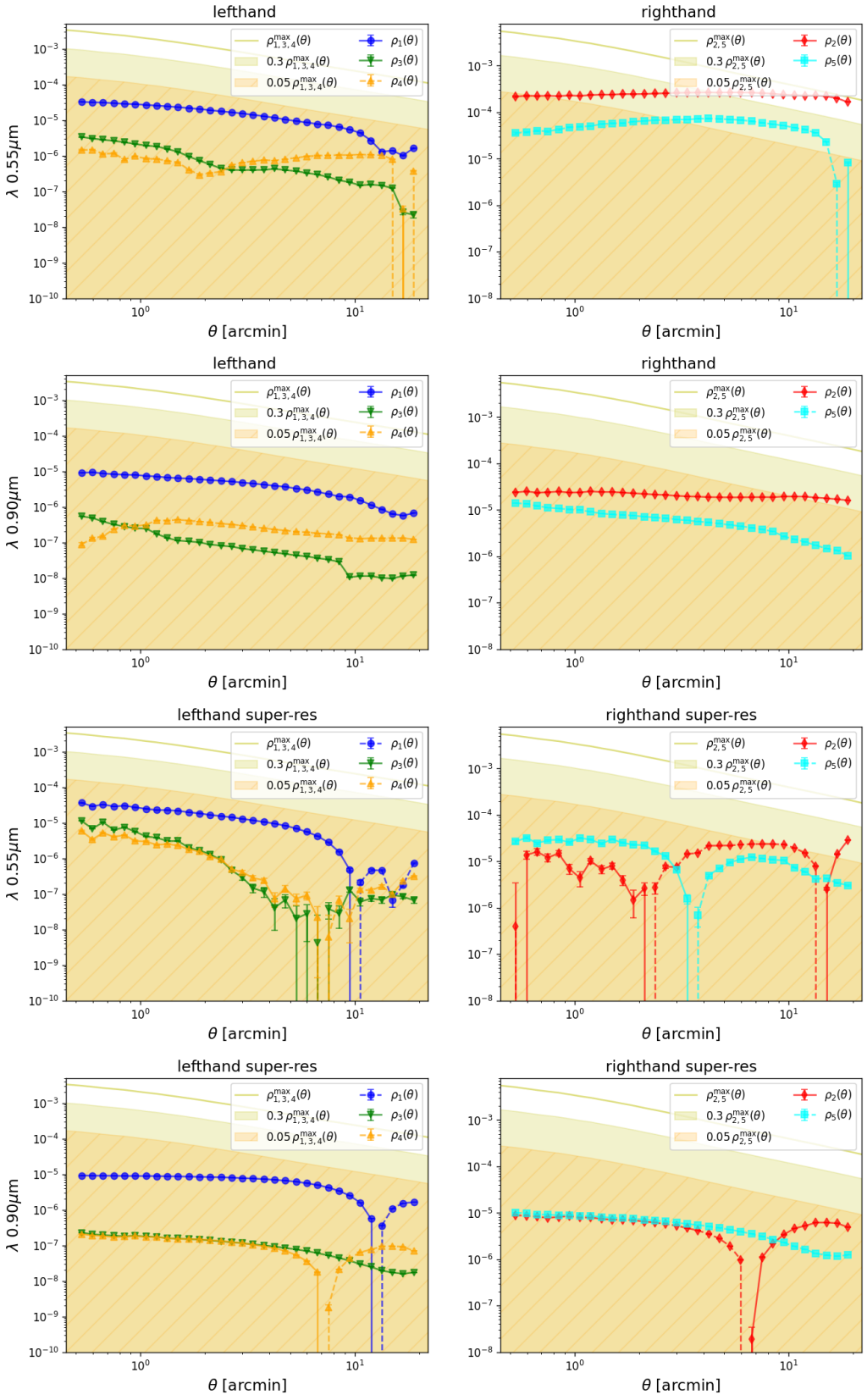


Figure A.3: ρs statistics for monochromatic source: (a)$\lambda = 0.55\mu$m, low resolution; (b)$\lambda = 0.90\mu$m, low resolution; (c)$\lambda = 0.55\mu$m, super-resolution; (b)$\lambda = 0.90\mu$m, super-resolution. The monochromatic PSF model for low resolution still needs improvement or further study.



B. Bibliographie commentée

[**1**] Martin Kilbinger. Cosmology with cosmic shear observations: a review. Reports on Progress in Physics, 78(8):086901, 2015, https://arxiv.org/abs/1411.0115.

This review provides an overview of the field of cosmology using cosmic shear observations. I found it very comprehensive and clear, that it can be a perfect introduction and reference for students. It derives important terms (from basic notions to statistical methods); it discusses the role of cosmic shear in constraining fundamental cosmological parameters, understanding dark matter and dark energy, and testing gravitational theories on large scales.

[**2**] Tobías Liaudat, et al. Rethinking data-driven point spread function modeling with a differentiable optical model. Inverse Problems, 39(3):035008, 2023, https://arxiv.org/abs/2203.04908.

The paper introduces a perspective on data-driven point-spread function (PSF) modeling using a differentiable optical model. It proposes an approach that enhances PSF estimation accuracy by incorporating the optical model's differentiable aspects. The part of introduction can help beginners learn an overview on the current PSF models.

[**3**] CosmoStat Tutorial pages, https://github.com/CosmoStat/Tutorials.

It comprises 4 parts: Career Development, Cosmology, Data Science, and Software Carpentries. I found the second and third parts very helpful for starting some specific science-related calculations. The youtube channel is also very useful:

https://www.youtube.com/channel/UCuTjur87q56Rw3V26fHgDHA.