

DISS. ETH NO. 27787

Galaxy morphology and cosmological predictions for wide-field surveys analyses

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

Federica Tarsitano

Laurea Magistrale in Fisica Nucleare, Sub-nucleare e Biomedica

Università degli Studi di Torino

born on May 24th, 1990

nationality: Italian

accepted on the recommendation of

Prof. Dr. Alexandre Refregier

Prof. Dr. Adam Amara

Prof. Dr. Lavinia Heisenberg

Galaxy morphology and cosmological predictions for wide-field surveys analyses

Federica Tarsitano

*To my family, who was always there for me.
To my friends and colleagues,
for caring about me and believing in me.
To Sonia and Elisa: the world lost you too early, dear friends.
The memory of you has been giving me strength.
To Leonardo, Antonio, Rainer, Wilfred, Alex, Adam, Will:
you offered me an immense support
and have been a source of inspiration.
I am honored we crossed our paths.*

Contents

List of figures	9
List of tables	13
Nomenclature	14
Summary	19
Sinossi	22
1 Introduction	25
1.1 Cosmology	28
1.1.1 Kinematics and dynamics of the Universe	28
1.1.2 Cosmological distances	31
1.1.3 Structure formation	31
1.1.4 Cosmological probes	36
1.2 Galaxy morphology	45
2 A morphological catalogue for DES Y1	48
2.1 Introduction	48
2.2 Data	51
2.2.1 The Dark Energy Survey	51
2.2.2 Image simulation data	52

2.3	Pre-fitting Pipeline	53
2.3.1	Sample Selection	53
2.3.2	Data processing	56
2.4	Parametric Fits	57
2.4.1	GALFIT Setup	57
2.4.2	Fitting Completeness	58
2.4.3	Validation	66
2.4.4	Calibrations	68
2.5	Non Parametric fits	72
2.5.1	ZEST+ Setup	72
2.5.2	Completeness	75
2.5.3	Validation	76
2.5.4	Calibrations and diagnostics of the corrected results	77
2.6	Science-ready cuts	80
2.7	Conclusions	82
3	Image feature extraction and galaxy classification: a novel and efficient approach with automated machine learning	86
3.1	Introduction	86
3.2	Data	89
3.2.1	The Dark Energy Survey	89
3.2.2	The data set	89
3.2.3	Sample selection	90
3.3	Method	91
3.3.1	Production of stamps	91
3.3.2	Extraction of profiles	92
3.4	AI framework and Modulos	93
3.5	Results and discussion	95
3.5.1	Information extraction	95
3.5.2	Model failures and future perspective	99
3.6	Conclusions	100
4	Predicting Cosmological Observables with PyCosmo	103
4.1	Introduction	103
4.2	Cosmological model	105
4.2.1	Background	106
4.2.2	Linear perturbations	107
4.2.3	Non-linear perturbations	108
4.2.4	Observables	109

4.2.5	Architecture	110
4.2.6	Symbolic calculations	111
4.2.7	Unit tests	112
4.2.8	PyCosmo Hub	112
4.3	Validation and code comparison	113
4.3.1	Cosmological setup and conventions	114
4.3.2	Background	115
4.3.3	Linear Perturbations	116
4.3.4	Non-linear Perturbations	120
4.3.5	Observables	121
4.3.6	Summary	124
4.4	Conclusions	124
4.5	Acknowledgements	126
5	Applications, extensions and further developments	127
5.1	DES Y1 Galaxy Morphology Catalogue and Gravitational Waves	128
5.2	PyCosmo and FFTLog	129
5.3	Kinetic Field Theory in PyCosmo	130
5.4	PyCosmo and the MCCL analysis	132
6	Conclusions	136
7	Contributions	139
	Appendices	142
A	Appendix A	143
A.1	Calibration maps for the g and r filters	143
A.2	Catalog Manual	146
A.2.1	Identification columns	146
A.2.2	SEXTRACTOR parameters for star-galaxy separation and signal-to-noise	146
A.2.3	Columns for Parametric Fits	146
A.2.4	Columns for non-parametric coefficients (ZEST+)	148
B	Appendix B	149
B.1	Code Setup	149
B.2	Heatmaps	151
	Acknowledgments	154

Ringraziamenti	159
Bibliography	164
Curriculum Vitae of Federica Tarsitano	182

List of Figures

1.1	Cosmos flowers in the Chilean Atacama desert and the Blanco Telescope with the Milky Way.	26
1.2	Left panel: representation of the distortions induced by lensing. Right panel: effects of convergence and shear on a circular source.	39
1.3	Geometrical view of gravitational lensing.	40
1.4	Prediction for the CMB temperature angular power spectrum.	45
1.5	The Hubble tuning fork, made with public galaxy images from the DES Collaboration.	47
2.1	Normalised distributions of the variables involved in the sample selection in the i band of the DES Y1 dataset.	55
2.2	Panel A: fitting completeness in g , r and i bands, following the cur in star-galaxy separation. Panels B, C, D: maps of the percentage of converged fits in g , r and i band in each tile (at $mag_auto_i < 23$).	60
2.3	Dependence of fitting completeness at $i < 21.5$ on spatially-dependent survey characteristics, stellar density, PSF FWHM and i -band image depth.	62
2.4	Left panel: fitting completeness calculated in differential bins of magnitude, dividing the sample into sub-populations, according to different ranges of the parameter $\xi = FLUX_RADIUS/PSF_radius$. Right panel: map of the percentage of converged fits per tile with $\xi > 1.25$	63
2.5	Fitting completeness as a function of the magnitude difference between the target galaxy and its closest neighbour.	65

2.6	Difference between the input magnitude (MAG_AUTO) from SExtractor and the output magnitude (MAG_SERSIC) recovered through Single Sérsic fits.	66
2.7	Relation between Sérsic magnitude and effective radius for the i band results.	67
2.8	Discrepancies in recovered Sérsic parameters from running GALFIT on the UFIG-BCC image simulations, as a function of signal to noise (S/N).	69
2.9	Calibration map for the parametric measurements in the i band, obtained from image simulations as described in Section 2.2.2.	71
2.10	Fitting completeness of non-parametric converged fits in the g , r and i bands.	76
2.11	Gini-M20 relation shown as a function of Concentration C , Sérsic Index, Asymmetry A , Clumpiness S , ellipticity ϵ and $g - i$ colour.	78
2.12	Calibration map for the non-parametric measurements in the i band, obtained through the simulation routine described in Section 2.5.4.	79
2.13	Sérsic Index-Concentration relation before (grey) and after (magenta) applying the calibrations.	81
2.14	Healpix map of the ratio between two galaxy samples. We apply to the Y1A1 data the sample-selection cuts to obtain the first sample, and then apply the science-ready cuts to it in order to get the second one. The ratio gives the completeness per pixel of the science-ready sample.	82
3.1	Example of a series extracted from galaxy images (processed through an isophote analysis) and used for classification.	93
3.2	Example of classification between early-type and late-type galaxy, according to their time-series-like profile.	94
3.3	Plot of the sequences corresponding to ETG (in blue) and LTG (in green). Additionally, 10 arbitrary ETG and LTG are drawn. The black, vertical lines indicate the nodes with the largest Gini feature importance.	96
3.4	Summary of the results provided by the Modulos AutoML platform as a function of the F1 score. The solutions are labelled in terms of decision tree models (either XGBoost or Random Forest) and feature engineering methods.	97
3.5	Confusion matrix representing the accuracy achieved in classifying galaxy profiles.	98
3.6	Properties of the classified sample in terms of signal-to-noise ratio, size and ellipticity.	99
3.7	Examples of isophotal fitting for mis-classified galaxies.	100
4.1	Flow-chart summarizing the PyCosmo architecture.	111

4.2	A screenshot of a Jupyter Notebook running on the PyCosmo Hub.	113
4.3	Comparison between PyCosmo, CCL, CLASS and iCosmo in terms of the comoving radius, for the assumed fiducial cosmology.	115
4.4	Heatmaps showing the comparison between PyCosmo and iCosmo, PyCosmo and CCL and PyCosmo and classy in terms of the comoving radius, for a variety of cosmological parameter values.	116
4.5	Code comparison in terms of the growth factor, calculated for the fiducial cosmology.	117
4.6	Comparison of PyCosmo with iCosmo (left panel) and CCL (right panel) in terms of linear power spectrum computed with the EH (solid lines) and the $BBKS$ (dashed lines) fitting functions, for three different redshifts. The y -axis on the left panel is not displayed in logarithmic scale for a better visualization.	118
4.7	Comparison between PyCosmo and iCosmo and between PyCosmo and CCL in terms of linear matter power spectrum computed with the EH fitting function.	118
4.8	Comparison between PyCosmo and iCosmo and between PyCosmo and CCL in terms of the linear matter power spectrum computed with the $BBKS$ fitting function.	119
4.9	Comparison between PyCosmo and classy in terms of the linear power spectrum computed with the Boltzmann solver.	119
4.10	Relative difference in terms of the non-linear matter power spectrum between PyCosmo and iCosmo and PyCosmo and CCL, choosing from a variety of fitting functions.	120
4.11	Heatmaps of the comparison between PyCosmo and iCosmo and between PyCosmo and CCL in terms of the non-linear matter power spectrum, computed with $Halofit+EH$	121
4.12	Heatmaps of the comparison between PyCosmo and iCosmo and between PyCosmo and CCL in terms of the non-linear matter power spectrum, computed with $Halofit+BBKS$	122
4.13	Comparison between the implementations of the <i>Mead Model</i> in PyCosmo and in the original code HMcode for the non-linear power spectrum.	122
4.14	Comparison in terms of $C_\ell^{\gamma\gamma}$ between PyCosmo and iCosmo and PyCosmo and CCL.	123
4.15	Preliminary CMB angular power spectrum computed with the PyCosmo Boltzmann Solver and absolute difference with the same output from classy.	123

5.1	Visualization of the galaxy NGC 4993, the host galaxy of the GW170817 gravitational wave event. The left panel shows the DECam coadded image in the g, r, i filters. In the middle panel the residuals from r-band Sérsic fitting clearly shows its shell-like structure. The HST ACS image (F606W-band) in the right panel emphasized dust lanes crossing the center of the galaxy. Courtesy of Palmese et al., 2017a	128
5.2	Schematic representation of the Kinetic Field Theory (KFT).	131
5.3	Preliminary results from the likelihood analysis using the MCCL output for the DES Y1 dataset.	134
A.1	Map of the corrections for Sérsic parameters in the g (upper panel) and r (lower panel) filters, obtained through the simulation routine described in Section 2.2.2. Symbols and colours have the same meaning as in Figure 2.9, showing the analogue calibration map for the i band.	144
A.2	Map of the corrections for the non-parametric structural measurements in the g (upper panel) and r (lower panel) filters, obtained through the simulation routine described in Section 2.5.4. Symbols and colours have the same meaning as in Figure 2.12, showing the analogue calibration map for the i band.	145
B.1	Heatmaps of the agreement between PyCosmo and iCosmo, PyCosmo and CCL and PyCosmo and CLASS, in terms of the linear growth factor.	151
B.2	Heatmaps of the agreement between PyCosmo and iCosmo, PyCosmo and CCL and PyCosmo and CLASS, in terms of the linear growth factor. In this figure the growth factor in PyCosmo is computed with hypergeometric functions.	151
B.3	Heatmaps of the agreement between PyCosmo and iCosmo and PyCosmo and CCL, in terms of the lensing power spectrum, computed with <i>Halofit+EH</i> in the first comparison and accounting for the revised version of <i>Halofit</i> in the second.	152
B.4	Heatmaps of the agreement between PyCosmo and iCosmo and PyCosmo and CCL, in terms of the lensing power spectrum, computed with the choice of fitting functions <i>Halofit+BBKS</i> in the first comparison and accounting for the revised version of <i>Halofit</i> in the second.	152
B.5	Heatmaps of the agreement between PyCosmo and CCL, in terms of the lensing power spectrum, computed with linear power spectrum, computed with <i>BBKS</i> (left panel) and <i>EH</i> (right panel) linear power spectrum.	153

List of Tables

2.1	Summary of the cuts applied to the overlapping sample between the catalogue provided by the DESDM pipeline and the Y1A1 GOLD catalogue. The selected objects must satisfy the requirements described in section 2.3.1. x identifies the filter ($x = g, r, i$).	54
3.1	Summary of the cuts applied to SExtractor catalogues for the selection of the sample from the DES DR2 dataset.	90
4.1	Summary of the code-comparison between PyCosmo, iCosmo, CCL, CLASS and HMcode.	125

Nomenclature

χ	comoving distance
δ	Statistical matter density of the Universe
δ^3	Dirac Delta function
δ_0	Density perturbations for dark matter at present time
ϵ	Galaxy ellipticity
η	conformal time
$\hat{\rho}$	Mean matter density of the Universe
Ω_b	Fractional baryon density today
Ω_m	Fractional matter density today
Ω_Λ	Fractional dark energy density today
ρ_{crit}	Critical density of the Universe
σ_8	r.m.s. of matter fluctuations in spheres of comoving radius of $8h^{-1}$ Mpc
(k)	Matter transfer function for wave vector k
τ	optical depth

Nomenclature

$\Theta_\ell(k, \eta)$	Photon temperature perturbations field
ξ	Ratio between PSF and galaxy size
$\xi(\phi)$	Two-point correlation function
a	scale factor
c	Speed of light in a vacuum inertial frame
$C_\ell^{\gamma\gamma}$	Lensing power spectrum
C_ℓ^{TT}	CMB angular power spectrum
D	Growth Factor
F_{tot}	Galaxy flux integrated out to $r = \text{inf}$
G	Newton's gravitational constant
$g(\eta)$	Visibility function
H	Hubble parameter
h	Dimensionless Hubble parameter at $z=0$
H_0	Hubble parameter at $z=0$
j_ℓ	Spherical Bessel function of order ℓ
k	Wave vector amplitude
k or b_n	Normalization constant coupled to the Sérsic index
m_ν	Neutrino mass
m_{tot}	Galaxy integrated magnitude
mag_{zpt}	Magnitude zero-point
n	Sérsic index
n_s	Power spectrum scalar spectral index
$P(k)$	Matter power spectrum
PA	Galaxy Position Angle
q	Galaxy axis ratio

r	Galaxy radius
R_e	Galaxy half light radius
$S(k, \eta)$	CMB angular power spectrum Source Function
T_{CMB}	CMB temperature today
t_{exp}	Exposure time for a galaxy image
v	Dark matter velocity
v_b	Baryon velocity
w	Equation of state parameter
z	redshift
z_*	Redshift at Recombination
AI	Artificial Intelligence
BAO	Baryon acoustic oscillations
CANDELS	Cosmic Assembly Near-infrared Deep Extragalactic Legacy
CANVAS	Cut ANd VALIDate Stamps
CAS	Concentration Asymmetry Smoothness system
CDM	Cold dark matter
CMB	Cosmic Microwave Background
CNN	Convolutional Neural Network
DECam	Dark Energy Camera
DES	Dark Energy Survey
DESDM	Dark Energy Survey Data Management
DM	Dark matter
ETG	Early-Type Galaxy
FFTLog	Fast Fourier Transform Log
FLRW	Friedmann-Lemaître-Robertson-Walker metric

Nomenclature

GEMS Galaxy Evolution From Morphology And SEDs

GR General Relativity

HSC Hyper Suprime-Cam

IA Intrinsic Alignment

ISW Integrated Sachs-Wolfe effect

KFT Kinetic Field Theory

LSS Large Scale Structure

LTG Late-Type Galaxy

MCCL Monte Carlo Control Loop

MCMC Monte Carlo Markov Chain

ML Machine Learning

PCA Principal Component Analysis

PSF Point Spread Function

SAND Selection And Neighbours Detection

SDSS Sloan Digital Sky Survey

SLR Stellar Locus Regression

SNe Ia Type Ia supernovae

VRO Vera Rubin Observatory

XGBoost eXtreme Gradient Boosting

Summary

Current and upcoming cosmological wide-field surveys use hundreds of millions of galaxy images to investigate the nature of dark matter, dark energy and large scale gravity. Their measurements lead to thriving research in the fields of Astrophysics and Cosmology. The large number of images they generate is an unprecedented source of information to study the cosmic evolution of galaxies with enhanced statistical power. In particular, they provide information about the shape and the internal structure of galaxies. In order to extract it from such large data sets researchers need to develop and optimize dedicated and automated algorithms. Those surveys have also opened the era of *Precision Cosmology*, where measurements need to be interpreted with theoretical predictions which are equally precise. Thus we need numerical frameworks able to predict cosmological quantities accurately and fast. These two applications, the morphological classification of galaxies, which are tracers of the evolution of the Universe, and the study of the properties of the Universe as a whole, complement each other and are the two main topics of this thesis.

The overall shape of a galaxy (elliptical, spiral, irregular) and its sub-structural properties (presence of a bulge, a disk, a bar or clumpy regions) are related to different scenarios of environmental interactions and star-formation. Combined with information about galaxy luminosity and mass, morphology is of great importance for the understanding the processes governing galaxy formation and evolution history. These key properties can be quantified by fitting the light distribution in galaxy images, thus by extracting structural parameters that describe the size, magnitude, concentration of light and shape

of the galaxy. The *parametric method* relies on an analytical model which is fitted to the galaxy image to estimate these parameters. One can also adopt a non-parametric approach, analyzing the distribution of light collected in the image, targeting potential clumpy regions and asymmetries. These two methods have the advantage of describing the internal structure of a galaxy in great detail and in a quantitative way. However, they can be computationally expensive, especially if applied to the large datasets provided by modern surveys. Moreover, parametric fitting can be prone to numerical degeneracy in the parameter space. Nevertheless, the information provided by these two methods is important to conduct a comprehensive study of the properties of the galaxy population. It is often complemented by a direct classification of galaxy types. This is possible through *citizen science* projects, such as Galaxy Zoo, where the public is invited to assign galaxy images to different classes, mostly choosing between elliptical, spirals and irregulars. Another recent approach applies machine learning algorithms (e.g. Convolutional Neural Networks, boosted decision trees etc.), which have been undergoing rapid development. In this thesis we present two multi-wavelength structural and morphological catalogues of more than 300 millions galaxies observed by the Dark Energy Survey (DES) collaboration. It is the largest structural catalogue to date, more than ten times larger than any catalogue of this type, and contains information obtained with both a parametric and a non-parametric approach. We will show that the combination of these two types of measurements provides a comprehensive description of the main morphological features of the galaxies in this large dataset. The computational framework developed in this work and the acquired knowledge of DES data contributed to a later analysis of the structural properties of the optical counterpart of the gravitational wave source GW170817, detected in August 2017.

The thesis proceeds to expand the study of the morphology of DES galaxies, by proposing a novel method to classify them into early-type and late-type, combining feature engineering and machine learning algorithms. Specifically, we explore the possibility of applying machine learning methods designed for one-dimensional problems to the task of galaxy image classification. The algorithms used for the latter typically rely on multiple costly steps, such as Point Spread Function (PSF) deconvolution and the training and application of complex Convolutional Neural Networks (CNN) with thousands or even millions of parameters. In our approach, we extract features from the galaxy images by analysing the elliptical isophotes in their light distribution and collect the information in a sequence. The sequences obtained with this method present definite features allowing a direct distinction between galaxy types. Then, we train and classify the sequences with machine learning algorithms, designed through the platform Modulos AutoML, and study how they optimize the classification task. We show that our novel method provides accurate galaxy classification and is faster than other approaches. The accuracy achieved is comparable to that reported by CNN-classifications on DES galaxies. Data

dimensionality reduction also implies a significant lowering in computational cost. In the perspective of future data sets obtained with future surveys (e.g. Euclid), this work represents a path towards using a well-tested and widely used ML platform to efficiently tackle galaxy classification problems at the peta-byte scale.

As mentioned above, accurate predictions for cosmological observables are important for cosmological analyses. These can be used to constrain the parameters of the standard cosmological model and its extensions. For this purpose, we further develop PyCosmo, an open-source Python-based framework providing theoretical predictions, and unique in its interactive, user-friendly interface for this new era of collaborative precision Cosmology. We extend PyCosmo features to include predictions for cosmological observables. We show that PyCosmo is competitive with similar codes in computing background quantities, linear and non-linear perturbations and angular power spectra. Also, it has been used within the MCCL (Monte Carlo Control Loop) framework to make cosmic shear analyses for the first release of DES data. We will also describe additional current applications and future extensions. Its architecture provides both the users and the developers a convenient interface to implement and extend new theoretical models. In particular, the Kinetic Field Theory (KFT) has been tested and it will be part of the code in the near future. PyCosmo is publicly available both as a Python package and on the PyCosmo Hub, an online platform where the users can work without the need of installing any software, share their research online and save their results locally. With its demos and tutorial sessions, the hub is conceived to be useful both for educational purposes and for promoting cosmological inferences in the cloud, in a new dynamic way of teaching and doing research.

Together, the advances described in this thesis provide good perspectives for the scientific analysis of future large cosmological surveys.

Sinossi

Come appare l'Universo alle presenti e future wide-field surveys cosmologiche? Agli occhi dei loro telescopi si presenta come una vastissima raccolta di milioni di immagini di galassie che vengono usate per indagare sulla natura della materia e dell'energia oscura, e della struttura a grande scala dell' Universo. Queste misurazioni rendono possibile un'ampia ricerca in Astrofisica e Cosmologia. Il gran numero di immagini messe a disposizione della comunità scientifica costituisce una ricca sorgente di informazioni con un potenziale statistico senza precedenti. Queste immagini vengono manipolate per studiare la forma e la struttura delle galassie e per ricostruirne l'evoluzione cosmica. Poiché sono organizzate in estese raccolte di dati che non sarebbe possibile analizzare manualmente, occorre sviluppare e ottimizzare algoritmi dedicati. Questi esperimenti, inoltre, hanno avviato l'era della *Precision Cosmology*, dove le misure vanno interpretate con predizioni teoriche di ugual precisione. A tal proposito vengono messi a punto algoritmi numerici in grado di fornire tali predizioni in maniera rapida e accurata. Questi due scenari complementari, la classificazione delle galassie che usiamo per tracciare l'evoluzione dell'Universo e lo studio delle proprietà dell'Universo stesso nella sua totalità, sono i due temi principali di questa tesi.

La forma di una galassia, che è ciò che ci fa distinguere tra galassie ellittiche, a spirale o irregolari, e le sue peculiari proprietà strutturali come la presenza di un addensamento centrale, detto *bulge*, di una componente discoidale, una barra centrale o regioni irregolari, dette *clumpy regions*, sono connesse a fenomeni fisici di interazione gravitazionale con oggetti vicini e all'attività di produzione stellare. In combinazione con le infor-

mazioni su luminosità e massa, la morfologia riveste un ruolo importante per comprendere i processi che governano la formazione e l'evoluzione delle galassie. Tali proprietà possono essere quantificate attraverso il fit di un modello analitico all'immagine: questo ci consente di stimare parametri strutturali che descrivono la dimensione, la magnitudine, la concentrazione di luce e la forma di ogni galassia. Questo approccio è conosciuto come *metodo* o *fitting parametrico*. Si può adottare, in alternativa, un approccio *non-parametrico*, analizzando la distribuzione di luce raccolta nell'immagine e cercando la presenza di *clumpy regions* e asimmetrie. Entrambi i metodi hanno il vantaggio di poter descrivere la struttura interna della galassia in grande dettaglio, ma richiedono molto tempo da un punto di vista computazionale. Inoltre, il fitting multi-parametrico è soggetto a soluzioni degeneri, corrispondenti a minimi locali nello spazio dei parametri. Ad ogni modo, le informazioni che possiamo ricavare ci danno la possibilità di condurre uno studio completo delle proprietà strutturali delle galassie. Vi sono poi metodi per una loro classificazione diretta, basati più sulla forma percepita che su quella stimata. Ne è un esempio *Galaxy Zoo*, un popolare progetto di *citizen science* dove il pubblico del web è invitato ad assegnare immagini di galassie a diverse classi cosmologiche, scegliendo tra ellittiche, a spirale e irregolari. Un approccio più recente prevede l'applicazione di algoritmi di Machine Learning (ad esempio Convolutional Neural Network, boosted decision trees ecc.), che stanno attraversando una fase di grande sviluppo. In questa tesi presentiamo due cataloghi strutturali multi-banda di più di 300 milioni di galassie osservate con la Dark Energy Survey (DES). Si tratta del più grande catalogo strutturale mai realizzato ad oggi, e contiene stime strutturali ottenute sia con approccio parametrico sia non-parametrico. Mostreremo come la combinazione di questi due tipi di misure fornisce una completa descrizione delle principali proprietà strutturali delle galassie in questo grande dataset. Gli algoritmi sviluppati per questo progetto e la conoscenza acquisita sul dataset hanno poi dato un importante contributo all'analisi morfologica della galassia ospitante la sorgente di onde gravitazionali identificata come GW170817 nell'Agosto del 2017.

La tesi procede con un'estensione dello studio della morfologia delle galassie di DES e la loro classificazione in ellittiche (di solito anche chiamate *early-type galaxies*) e a spirale (*late-type*). Il progetto propone un modello innovativo che combina *feature engineering* e algoritmi di machine learning. Esploriamo nello specifico la possibilità di applicare alle immagini metodi di machine learning che vengono solitamente adoperati in problemi unidimensionali. Infatti la manipolazione di immagini implica numerosi e lunghi passaggi, come la deconvoluzione con la Point Spread Function (PSF) e il training di complessi Convolutional Neural Network (CNN) con migliaia o milioni di parametri. Nel nostro approccio analizziamo le componenti delle galassie e salviamo le informazioni necessarie in sequenze e osserviamo che esse presentano alcuni dettagli che ben consentono di associarle a diversi tipi di galassie. Dunque procediamo con il training e la classifi-

cazione usando algoritmi di machine learning realizzati con la piattaforma Modulos AutoML. Dall'analisi dei risultati otteniamo indicazioni promettenti: il nostro metodo rivela una precisione comparabile con quella raggiunta per le galassie classificate con CNN, ed offre, grazie all'approccio unidimensionale, una significativa riduzione dei tempi di analisi. Nella prospettiva di futuri progetti osservativi (Euclid, ad esempio), ci proponiamo di migliorare, testare e usare questo metodo per studiare la struttura delle galassie raccolte in dataset di dimensioni maggiori.

In precedenza abbiamo sottolineato l'importanza di come predizioni accurate siano importanti per le analisi cosmologiche. Infatti possiamo usarle per determinare i parametri del modello cosmologico standard e le sue possibili estensioni. A tale scopo abbiamo lavorato allo sviluppo di PyCosmo, un pacchetto scritto in Python che produce predizioni di grandezze cosmologiche, e lo fa proponendo un'interfaccia unica e interattiva. Il nostro progetto si focalizza sullo sviluppo e il test delle osservabili cosmologiche. In questa tesi mostriamo come PyCosmo sia competitivo con altri pacchetti simili nel calcolo del background cosmologico, delle perturbazioni lineari e non lineari di materia e dello spettro di potenza angolare. Inoltre descriveremo il suo utilizzo in MCCL (Monte Carlo Control Loop), una pipeline per svolgere analisi dello shear delle galassie di DES. Illustreremo anche ulteriori applicazioni, e parleremo di come gli utenti possano facilmente includere nuovi modelli ed estensioni. Tra esse, la Kinetic Field Theory (KFT) è stata oggetto di recente sviluppo e test e sarà parte integrante del pacchetto in release future. PyCosmo è disponibile al pubblico sia come pacchetto di Python sia sulla piattaforma PyCosmo Hub, dove gli utenti possono fruire del software senza installarlo, condividere le loro ricerche online e salvare convenientemente i risultati. Con l'inclusione di demo e tutorial, l'hub è concepito come uno strumento didattico e costituisce una piattaforma digitale per analisi cosmologiche, in un modo nuovo e dinamico di condurre ricerca.

Gli sviluppi presentati in questa tesi offrono dunque interessanti prospettive per le analisi scientifiche di future survey cosmologiche.

Introduction



La campana del tempio tace,
ma il suono continua
ad uscire dai fiori.

The temple bell dies away.
The scent of flowers
Is still tolling the bell.

MATSUO BASHO (1644 - 1694)

The word *Universe* speaks for itself. As its etymology suggests (from latin *Univ-ersum*), it includes everything that exists. The orange flowers who dare growing in the middle of the Chilean Atacama desert, sprouting from the dry, dark land around the Blanco telescope, are part of it (see Figure 1.1). Scientists borrowed the big eye of that telescope to look up and try to unveil cosmic properties encoded in photons, the messengers between us and many astronomical objects living as bright flowers in the dark infinite land of the Universe. Some of the galaxies observed with that telescope are part of the analyses presented in this thesis.

Galaxies are dynamically bound systems full of stars, showing a structural and functional complexity which is reflected in the extraordinary diversity of their population. Many aspects characterize them out, starting from morphology and environmental interactions and moving to their color, star formation and nuclear activity. Such properties are often correlated: in general ellipticals are redder than spirals, and galaxies with irregular shapes are often the results of mergers or tidal interactions and might undergo a phase of starburst activity. The size of spiral galaxies, which are supported by rotation,



Figure 1.1: Cosmos flowers in the Chilean Atacama desert (left panel) and the Blanco Telescope with the Milky Way (right panel).

contributes to the measure of their dynamics in terms of angular momentum. The fact that we can observe such bright, well-defined astronomical entities in large numbers over cosmological distances and times scales, not only allows us to study their formation and evolutionary history, but also makes them unique tracers of the evolution of the Universe as a whole. The Hubble law (1929) sees them receding from each other at a speed increasing with their distances, giving the first observational evidence of the cosmic expansion. Their large-scale distribution is one of the probes we use to study the properties of our Universe.

The scientific study of the Universe amounts to observing physical phenomena occurring in a giant laboratory where we have no control. In order to understand them, we need knowledge of the governing physical laws, mainly gravity and particle interactions. If we consider the Universe as a physical system, we need to describe its geometry, material content and how it formed and evolved. The beginning of Cosmology, the study of the evolution and the properties of the Universe, was marked by the theory of General Relativity (GR) that Einstein published in 1915. The most important aspect of GR is that it connects the geometry of space with the matter and energy content of the Universe. The Standard Model of Cosmology starts with the assumption that the Universe is homogeneous and isotropic on large scales and applies it to the space-time geometry. The combination of GR and this Cosmological Principle led to a model that depends on the space-time curvature, K , and the *scale factor*, a . The latter represents the scaling with cosmic time of the distance between two points with fixed comoving coordinates in an expanding background. The time evolution of the scale factor is expressed in terms of a set of *cosmological parameters* that describe the constituents of our Universe. According to the Standard Model of Cosmology, called Λ CDM (Lambda Cold Dark Matter), the cosmic inventory includes baryonic matter, which is described by the

Standard Model of Particle Physics, dark-matter, also referred as *cold dark-matter*, an invisible and non-interacting form of matter, and dark energy (Λ), which is responsible for the recent acceleration of the Universe. Cosmological experiments collect observational evidences for the existence of dark matter and dark energy, and measure multiple observables which can be used to probe the validity of the Λ CDM model. These cosmological probes are the Cosmic Microwave Background (CMB) anisotropies, weak gravitational lensing, galaxy clustering, type Ia supernovae and clusters (e.g. [Wayne Hu and Dodelson, 2002](#); [D. H. Weinberg et al., 2013](#); [Mandelbaum, 2018a](#); [Aghanim et al., 2020](#)). Being able to predict the observables as a function of cosmological parameters, cosmologists can determine the confidence ranges where the predictions are consistent with data. The Λ CDM paradigm, furthermore, includes another component to explain the origin of cosmic structures: the theory of inflation, which gives an explanation for the first seeds of structures. Those seeds are where also galaxies come from.

Current and upcoming cosmological wide-field surveys, such as the Dark Energy Survey (DES)¹, DESI², LSST³, Euclid⁴ and WFIRST⁵ are reshaping the way we conduct research, especially from two main perspectives: an exceptional rise in the amount of data and the precision achievable through the measurements.

In the field of Extragalactic Astronomy, this translates in the availability of images of millions of galaxies whose structural properties can be catalogued and used to study the cosmic evolution of the galaxy population with enhanced statistical power. For example, the latest photometric-release from DES counts 390 million objects ([Sevilla-Noarbe et al., 2021](#)). The morphology catalogue derived from the first release (which counted around 300 million objects) and presented in Chapter 2 includes estimations of the structural properties of 45 million galaxies. A few years earlier, an analogue collection included around 3 million galaxies observed by the Sloan Digital Sky Survey (SDSS) ([Kuminski and Shamir, 2016](#))⁶. In this context, galaxy structure can be measured with different methods. In the *parametric approach*, we can model the shape of a galaxy with two-dimensional parametric functions fit to the images in order to derive an estimate for the size, magnitude, ellipticity and the rate at which lights fades towards the outskirts of the galaxy (named Sérsic index). *Non-parametric* methods study how light is distributed in the image, searching for clumpy regions, studying its concentration and estimating the degree of asymmetry of the distribution. Recent approaches focus on direct automated

¹<http://www.darkenergysurvey.org>

²<http://desi.lbl.gov>

³<http://www.lsst.org>

⁴<http://sci.esa.int/euclid/>

⁵<http://wfirst.gsfc.nasa.gov>

⁶<https://www.sdss.org/>

classifications of galaxy images using machine learning algorithms. In Section 1.2 we present an overview of galaxy morphology and examine these methods. In Chapters 2 and 3 we describe two research works focused on the morphological classification of DES galaxies: the first one using parametric and non-parametric approaches, the second proposing a novel method for automatic feature extraction and classification of galaxy morphology, combining pattern recognition and machine learning.

In the field of Cosmology, the above mentioned surveys offer high constraining power on cosmological parameters. In fact, the precision achieved in recent measurements and achievable in future marks the era of *Precision Cosmology*. In order to interpret these results and test the Λ CDM model, we need an adequate precision in the predictions of the observables. In this context, we develop PyCosmo, a python-based framework computing cosmological predictions. In order to fulfill the precision requirements, we compare it to other similar existing codes, which are implemented in different programming languages. Being written in Python, which is widely used in Astronomy, PyCosmo is the most intuitive and easy to use. Additionally, we propose it in an interactive web form, through a public platform called PyCosmo Hub. In Section 1.1 we describe the Λ CDM model and we give an overview on the quantities computed by the code and in Chapter 4 we present this framework further. Chapter 5 reviews the projects where both the galaxy morphological catalogues and PyCosmo found applications, and describe further extensions. In particular, we show how the DES morphology catalogue, described in Chapter 2, brought substantial contribution to the study of NGC 4993, the host galaxy of the GW170817 gravitational wave event detected in August 2017. We also describe the contribution brought by *PyCosmo* to the cosmic shear analyses of DES data. In Chapter 6, we summarize the main topics of thesis and discuss their implications and possible developments.

1.1 COSMOLOGY

In this section, we provide an overview of the kinematics and dynamics of the Universe and introduce some notation.

1.1.1 KINEMATICS AND DYNAMICS OF THE UNIVERSE

The Cosmological Principle states that the Universe is statistically homogeneous and isotropic. This assumption translates into a model for space-time geometry which uses the concept of *metric* to describe physical distances and introduces a constant value for the curvature of space. To develop these concepts, we can start from something familiar, namely the square distance between two points in a two-dimensional Euclidean plane:

$ds^2 = (dx)^2 + (dy)^2$. A transformation from Cartesian to polar coordinates (r, θ) yields $ds^2 = (dr)^2 + r^2(d\theta)^2$. The distance thus remains the same to an observer, independently of the coordinate system used. The space-time geometry can be expressed in a similar way, as $ds^2 = g_{\mu\nu}dx^\mu dy^\nu$, where $\mu, \nu = 0, 1, 2, 3$ and 0 refers to the time-like coordinates ($dx^0 = dt$) and $g_{\mu\nu}$ is the space-time metric. In an expanding Universe, the metric, which has signature $[1, -1, -1, -1]$, includes the scale factor in order to model the scaling of the distance between two points whose positions are fixed on the background coordinate system, named *comoving coordinates system*. Using the metric and the scale factor, we can introduce the concepts of *peculiar velocity* and *Hubble flow*. In fact, by re-writing the squared distance in this form:

$$ds^2 = dt^2 - a^2(t)\gamma_{ij}dx^i dx^j, \quad (1.1)$$

where $a^2(t)\gamma_{ij}$ is the spatial part of $g_{\mu\nu}$, we can calculate the velocity, after setting the comoving coordinates to $x^i = [x^1, x^2, x^3]$ and using the chain rule:

$$|v^i| = a(t)\frac{dx^i}{dt} + \frac{da}{dt}x^i = v_{pec}^i + Hx^i. \quad (1.2)$$

In the equation above the quantity Hx^i is the Hubble flow and

$$H = \frac{\dot{a}}{a} \quad (1.3)$$

is the *Hubble parameter*. $v_{pec}^i = a(t)\dot{x}^i$ is the peculiar velocity, measured by a comoving observer (i.e. an observer following the Hubble flow).

We can write a specific form of the metric for an homogeneous and isotropic Universe in polar coordinates, as:

$$ds^2 = dt^2 - a^2(t) \left[\frac{dr^2}{1 - Kr^2} + r^2 d\Omega^2 \right], \quad (1.4)$$

where K is the curvature parameter and can be either -1, 0 or 1, referring, respectively, to an hyperbolic, flat or spherical hypersurface. This is called the Friedmann-Robertson-Walker (FRW) metric.

The dynamics of the Universe and the evolution of the metric are described by the Einstein equations:

$$G_{\mu\nu} = 8\pi GT_{\mu\nu}, \quad (1.5)$$

where $G_{\mu\nu}$ is the Einstein tensor, describing the space-time geometry, and $T_{\mu\nu}$ is the stress-energy tensor, measuring the matter content of the Universe. The latter has signature $(-\rho, p, p, p)$, where ρ and p are, respectively, the density and the pressure of a fluid which responds to the assumptions of homogeneity and isotropy. Thus, the com-

1.1. Cosmology

ponents of the tensor are $T_{00} = \rho(t)$, $T_{i0} = 0$ and $T_{ij} = -p\gamma_{ij}$. The Einstein tensor has the following extended form:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}\mathcal{R}g_{\mu\nu}, \quad (1.6)$$

where $R_{\mu\nu}$ and \mathcal{R} are the Ricci tensor and scalar, respectively, and are given by:

$$R_{\mu\nu} = \partial_\lambda \Gamma_{\mu\nu}^\lambda - \partial_\nu \Gamma_{\mu\lambda}^\lambda + \Gamma_{\lambda\rho}^\lambda \Gamma_{\mu\nu}^\rho - \Gamma_{\mu\lambda}^\rho \Gamma_{\nu}^\lambda, \quad (1.7)$$

$$\mathcal{R} = g^{\mu\nu} R_{\mu\nu}, \quad (1.8)$$

with $\Gamma_{\alpha\beta}^\mu = \frac{1}{2}g^{\mu\lambda}(\partial_\alpha g_{\beta\lambda} + \partial_\beta g_{\alpha\lambda} - \partial_\lambda g_{\alpha\beta})$ as the *Christoffel symbols* (see e.g. [Dodelson, 2003](#) and [S. Weinberg, 2008](#)). Due to the isotropy of the metric, the non-vanishing components of the Ricci tensor are:

$$R_{00} = -3\frac{\ddot{a}}{a}, R_{ij} = -\left[\frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2 + 2\frac{k}{a^2}\right], \quad (1.9)$$

and the Ricci scalar becomes:

$$\mathcal{R} = -6\left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2}\right]. \quad (1.10)$$

Replacing these components and the stress-energy tensor in the Einstein equations, we get the *Friedmann Equations*:

$$\frac{\dot{a}}{a} = H^2 = \frac{8\pi G}{3}\rho - \frac{K}{a^2}, \quad (1.11)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p). \quad (1.12)$$

In these equations, with ρ we refer to the total contribution to the energy density of the Universe from different components: ρ_r is the contribution by radiation (ρ_γ by photons and ρ_ν by neutrinos), ρ_m is the contribution by matter (ρ_c for cold-dark matter, ρ_b for baryons) and ρ_Λ is the vacuum energy contribution. From the first Friedmann equation, after setting $K = 0$ (flat Universe), we can define the *critical density* today as:

$$\rho_{crit,0} = \frac{3H_0^2}{8\pi G}, \quad (1.13)$$

and we can use it to define dimensionless density parameters, which are those we aim to constrain with cosmological observations and predictions:

$$\Omega_{I,0} = \frac{\rho_{I,0}}{\rho_{crit,0}}, \quad (1.14)$$

with $I = r, m, c, b, k, \Lambda$, for radiation, matter, dark-matter, baryons, curvature and dark energy, respectively. The left-hand side of the second Friedmann equation (1.12) expresses the acceleration parameter, $q_0 = \frac{\ddot{a}}{a}$, while the right-hand side, including the terms of density and pressure, comes from the contribution of the stress-energy tensor to the non-vanishing Einstein Equations. Assuming that a fluid follows the equation of state $p = w\rho$, by replacing it in Equation 1.14 and solving by ρ we obtain:

$$\rho \propto a^{-3(1+w)}, \quad (1.15)$$

where $w = 1/3, 0, -1$ for photons, matter and dark energy, respectively. Therefore we have $\rho \propto a^{-4}$ for a radiation-dominated Universe, $\rho \propto a^{-3}$ for a matter-dominated Universe and $\rho = \text{const.}$ for the vacuum energy density. From these considerations the Friedmann equation 1.11 can be rewritten as:

$$H^2(a) = H_0^2[\Omega_{r,0}(1+z)^4 + \Omega_{m,0}(1+z)^3 + \Omega_{k,0}(1+z)^2 + \Omega_\Lambda], \quad (1.16)$$

where we used the relation between cosmological redshift and scale factor $(1+z) = a^{-1}$.

1.1.2 COSMOLOGICAL DISTANCES

Given the above framework describing the dynamics of the Universe, we can derive expressions for different cosmological distances. They play an important role in the prediction of cosmological observables, therefore it is important to be able to accurately compute them. To avoid repetitions, we invite the interested reader to consult Chapter 4, Section 4.2.1, for a detailed description of these distances.

1.1.3 STRUCTURE FORMATION

The previous sub-sections gave an overview of the kinematics and the dynamics of an expanding homogeneous and isotropic Universe. Homogeneity and isotropy is a good approximation on large scales, but this is in contrast with the fact that the Universe itself is populated by structures, including galaxies, on small scales. In order to explain structure formation, the smooth Universe treatment above needs an extension, including the evolution of perturbations in the matter density field seeded by inflation (Tsujikawa, 2003). Such fluctuations ultimately lead to the formation of the observed structures.

STATISTICAL DESCRIPTION OF THE MATTER DENSITY FIELD

According to *linear perturbation theory*, we can describe the matter density field as a distribution of over-densities where, due to gravitational interaction, matter progressively accretes. On the other hand, under-dense regions see matter progressively being stripped away. To study such inhomogeneities we can consider over-densities, δ_i , at different points x_i in space. For instance, we can associate the over-density δ_1 to the coordinate x_1 , δ_2 to x_2 and δ_3 to x_3 . Therefore, we can describe the cosmological density field as a realization of a statistical distribution $P(\delta_1, \delta_2, \delta_3, \dots, \delta_N)d\delta_1, d\delta_2, d\delta_3, \dots, d\delta_N$. If we assume a Gaussian random field and we consider just one-point distribution function, then we can write:

$$P(\delta)d\delta = \frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{\delta^2}{2\sigma^2}}d\delta, \quad (1.17)$$

where δ is the density contrast and σ is the variance. The density contrast is defined as:

$$\delta(\bar{x}, t) = \frac{\rho(\bar{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)}, \quad (1.18)$$

where $\bar{\rho}(t)$ is the mean density of the Universe. In Fourier space the density contrast becomes:

$$\delta(\bar{k}, \eta) = \frac{1}{\pi^{3/2}} \int \delta(\bar{x}, \eta) e^{-i\bar{k}\cdot\bar{x}} d\bar{x}, \quad (1.19)$$

where we used the conformal time $d\eta = dt/a(t)$. If we consider N-points, we have a multivariate Gaussian distribution and equation 1.17 becomes:

$$P(\delta_1, \delta_2, \delta_3, \dots, \delta_N) = \frac{e^{-Q}}{[(2\pi^N)\det(\mathcal{C})]^{1/2}}, \quad (1.20)$$

where:

$$Q = \frac{1}{2} \delta_i (C^{-1})_{ij} \delta_j, \quad \mathcal{C} = \langle \delta_i \delta_j \rangle = \xi(r_{ij}), \quad (1.21)$$

where $\xi(r_{ij})$ is the *two-point correlation function* and expresses how two densities correlate as a function of the separation $r_{ij} = |\bar{x}_i - \bar{x}_j|$. The *matter power spectrum*, $P(k)$, is the Fourier Transform of the two-point correlation function:

$$P(\bar{k}) = \int \frac{d^3x}{(2\pi)^3} \xi(\bar{x}) e^{-i\bar{k}\cdot\bar{x}}. \quad (1.22)$$

The statistics of a Gaussian random field is completely determined by either the two-point correlation function or the power spectrum.

LINEAR GROWTH OF STRUCTURES

As expressed in equation 1.19, in Fourier space the perturbations in the matter density field can be considered as the sum of plane waves (Fourier modes) with different wave-numbers, k and associated wavelength, λ , representing a comoving scale. Assuming that the Universe is filled with an ideal fluid of density ρ and pressure p , in linear regime the over-densities evolve as:

$$\frac{\partial^2 \delta}{\partial t^2} + 2 \frac{\dot{a}}{a} \frac{\partial \delta}{\partial t} = \frac{\Delta^2 p}{\bar{\rho} a^2} + 4\pi G \bar{\rho} \delta, \quad (1.23)$$

describing a system on which pressure and gravity act in competition. For collisionless dark-matter we can neglect the pressure in the second term of equation 1.23, which becomes:

$$\frac{\partial^2 \delta}{\partial t^2} + 2 \frac{\dot{a}}{a} \frac{\partial \delta}{\partial t} - 4\pi G \bar{\rho} \delta = 0. \quad (1.24)$$

This equation has the form of a damped harmonic oscillator. This allows us to write a solution decoupling the time and space part, as $\delta(\vec{k}, t) = D(t)\delta_0(\vec{k})$. The time component, D , describing the evolution of the perturbations, is named *growth factor* and follows the following equation:

$$\frac{d^2 D}{dt^2} + 2H \frac{dD}{dt} - \frac{3}{2} H^2 \Omega D = 0, \quad (1.25)$$

whose solution is the sum of *growing* and *decaying* modes, so that we can write:

$$\delta(\vec{x}, t) = D^+(t)\delta^+(\vec{x}, 0) + D^-(t)\delta^-(\vec{x}, 0). \quad (1.26)$$

For an Einstein-de Sitter Universe (dark-matter only) we find $D^+ \propto t^{2/3}$ and $D^- \propto t^{-1/3}$, by solving equation 1.25 for $a(t) \propto t^{2/3}$. This decomposition is convenient because the modes evolve independently and we can describe the evolution of the matter density field by multiplying the initial perturbations by the time-dependent growth factor (see also Chapter 4, Section 4.2.1 for a more detailed definition and discussion of the linear growth rate). We can then describe the linear power spectrum by considering an initial power spectrum, called *primordial power spectrum* and making it evolving in time through the growth factor, and with a function relating the initial amplitudes to the Fourier modes in the post-recombination era, called *transfer function*, $T(k)$. Using the formalism as in [Mo, van den Bosch, and S. White, 2010](#), we can write the linear power spectrum as follows:

$$P(k, t) = \langle |\delta(k, t)|^2 \rangle = P_i(k) T^2(k) D^2(t), \quad (1.27)$$

where $P_i(k)$ is the primordial power spectrum. Following the assumptions of the inflationary models, $P_i(k)$ is described as a power law:

$$P_i(k) \propto k^{n_s}, \quad (1.28)$$

where the power n_s is the *spectral index*. We often consider a dimensionless quantity for the power spectrum:

$$\Delta^2(k) = \frac{1}{2\pi^2} k^3 P(k). \quad (1.29)$$

As we will discuss later in this chapter, precise calculations of the matter power spectrum are needed for accurate predictions of cosmological observables.

NON-LINEAR PERTURBATIONS

In the linear regime, the modes evolve independently following equation 1.26, as a function of the cosmic scale factor. Models for the non-linear regime describe the gravitational collapse and the virialization of structures. The complexity of such descriptions derives from the fact that in this case the modes become coupled and we cannot describe the matter density field as a Gaussian random field any longer. There are different models to describe non-linear perturbations: a simplified, analytic one, called *spherical collapse model*; *high-order perturbation theory*; *numerical simulations*; the *Halo Model*; the *Press-Schechter formalism*. We refer the interested reader to the comprehensive description of these methods in [Mo, van den Bosch, and S. White, 2010](#). Here we briefly describe the main ideas behind the spherical collapse model and the Halo Model, and how they are related to the matter power spectrum. For the spherical collapse model we assume a matter-dominated Universe after recombination, and a collisionless fluid to model the dark matter, interacting only through gravity. Also, we consider a spherical over-density, and describe it using a *top-hat function*. Let us consider concentric mass shells of radius \bar{r}_i making up to the top hat over-density. We assume that the shells expand at the same speed, without crossing each other. The mass inside the top-hat can be modelled as:

$$M(< r) = \frac{4}{3}\pi r^3(t)\bar{\rho}(t)[1 + \delta(t)], \quad (1.30)$$

where $r(t)$ is the expanding radius of the over-density and $\bar{\rho}$ is the background density. According to the top-hat model, the over-density region keeps expanding. When the critical linear density contrast is reached ($\delta_c = 1.686$), the system collapses and the shells contract and cross each other, exchanging energy and exerting mutual gravitational interactions. This non-linear process, which doesn't conserve memory of the initial conditions, results in a virialized dark matter halo. The system is now stable and bound by the gravitational force. The ratio between its average internal density and the mean density of the universe is predicted to be $\Delta_V \approx 178$. Dark matter halos of increas-

ing mass are host of galaxies, groups and clusters of galaxies.

The Halo Model considers dark matter halos as basic units of cosmological structures and uses them to sample the matter density field. The power spectrum can be seen as a superposition of spherical haloes, which contribute to the matter power spectrum as:

$$\Delta_{1H}^2(k) = 4\pi \left(\frac{k}{2\pi}\right)^3 \frac{1}{\bar{\rho}^2} \int M^2 W^2(k, M) F(M) dM, \quad (1.31)$$

where the integral is over the halo masses M , $F(M)$ is the halo-mass function (giving the halo number density per unit mass) and $W(k, M)$ is the Fourier transform of the halo density profile:

$$W(k, M) = \frac{1}{M} \int_0^{r_V} \frac{\sin(kr)}{kr} 4\pi r^2 \rho(r, M) dr, \quad (1.32)$$

where $r_V = \left(\frac{3M}{4\pi\Delta_V\bar{\rho}}\right)^{\frac{1}{3}}$ is the virial radius. Both the halo mass function and the radial internal density structure, described by the function $\rho(r, M)$, can be estimated from cosmological simulations. The power spectrum given in equation 1.31 assumes the halos are randomly distributed, which is not fully correct, since their positions are correlated. In order to account for their spatial correlation, an additional term, proportional to the linear power spectrum, is introduced:

$$\Delta_{2H}^2(k) = \Delta_{\text{lin}}^2(k) \left[\frac{1}{\bar{\rho}} \int dM M F(M) b_h(M) W(k, M) \right], \quad (1.33)$$

where $b_h(M)$ is the halo-bias parameter, introduced to account for the bias of dark matter halos with respect to the mass distribution. The terms expressed in equations 1.31 and 1.33 are called *one-halo* and *two-halo* terms, respectively. In the Halo Model, they form the full non-linear dimensionless matter power spectrum:

$$\Delta^2(k) = \Delta_{1H}^2(k) + \Delta_{2H}^2(k). \quad (1.34)$$

FITTING FUNCTIONS FOR THE LINEAR AND NON-LINEAR POWER SPECTRUM

The matter power spectrum plays an important role in the calculation of cosmological observables. In cosmological analyses, fitting functions were proposed to model linear and non-linear power spectra. They are defined by a set of parameters chosen to match numerical simulations. Widely used fitting functions for the linear-power spectrum are the *Eisenstein & Hu* (Eisenstein and W. Hu, 1998) and *BBKS* (Peacock, 1997) fitting functions, while for the non-linear perturbations they are the *Halofit* fitting functions (Smith et al., 2003; Takahashi et al., 2012) and the more recent *Mead model* (Mead et al., 2015). The latter, in particular, modifies equation 1.34 in order to regulate the transition be-

tween the two *halo terms*, by introducing a smoothing parameter, α :

$$\Delta^2(k) = [(\Delta_{1H}^2(k))^\alpha + (\Delta_{2H}^2(k))^\alpha]^{1/\alpha}. \quad (1.35)$$

Values of $\alpha < 1$ smooth the transition, while values of $\alpha > 1$ sharpen it. Details and references about these models are given in Chapter 4, Section 4.2.1. In the same chapter, we show how to compute and validate the matter power spectrum by choosing from a variety of fitting functions and comparing them to each other.

1.1.4 COSMOLOGICAL PROBES

In the previous sections, we gave an overview on the geometry and the evolution of the matter and energy content in the Universe. We now describe the quantities we can measure with cosmological surveys. Such quantities, called *cosmological observables*, can be predicted with theoretical models involving the background cosmology and perturbations, which in turn are expressed in terms of the cosmological parameters. By varying these, we can find the values that best match our observations. Since the observables have constraining power on the cosmological parameters, we can use them to probe the standard cosmological model and its extensions.

SUPERNOVAE

The term *Supernova* refers to a powerful and luminous stellar explosion, usually associated to the last evolutionary stage of a massive star or to a runaway nuclear fusion in a white dwarf. During this transient phenomenon, the exploding star becomes as bright as an entire galaxy with billions of stars. *Type Ia* supernovae occur in a binary system, where two stars orbit one another, and one of them is a white dwarf. The latter starts accreting mass from its companion and becomes unstable. When it reaches a critical mass (~ 1.4 solar masses) it gets disrupted by a powerful thermonuclear explosion. Type Ia supernovae can be used as *standard candles* because they all reach approximately the same brightness at the peak phase. For this reason, astronomers can use the distance to them to infer properties of the cosmic expansion and acceleration. This relevant distance, called *luminosity distance*, d_L (see also Chapter 4, section 4.2.1), can be expressed in terms of the magnitude of the observed object:

$$d_L = 10^{\frac{(m-M)}{5}+1} \text{ pc}, \quad (1.36)$$

where m and M are the absolute and apparent magnitude, respectively. The absolute magnitude is the magnitude the object would have at 10 pc distance, while the apparent magnitude is the observed one. The latter requires a correction for *reddening*: photons travelling from the emitting source towards us can be absorbed and re-emitted by dust

grains along the line-of-sight, ultimately shifting to larger wavelengths. High energy photons are more prone to such phenomena, because the dust grains are of size comparable to their short wavelength. Therefore the reddening effect on the spectrum of an emitting source is not a global shift towards red (which is the effect of cosmological redshift due to cosmic expansion), but a shift of the bluer part of the spectrum towards lower frequencies. Using the FRW metric (equation 1.16), the luminosity distance can be written in terms of the Hubble parameter, and so in dependence of the cosmological parameters as:

$$d_l = H_0^{-1}(1+z) \int_0^z \frac{dz}{[\Omega_m(1+z)^3 + \Omega_\Lambda(1+z)^{3(1+w)} + \Omega_k(1+z)^2]^{1/2}}. \quad (1.37)$$

For different cosmologies, we can infer the luminosity as a function of redshift and probe the cosmological model. Measurements made in 1998 (Riess et al., 1998) showed that the supernovae appeared 25% fainter than theoretically predicted. This was indication that in the time occurred to light to reach us the Universe had expanded faster than expected, with the consequent evidence for cosmic acceleration.

GALAXY CLUSTER COUNTS

As mentioned in 1.1.3 cosmological simulations can predict the number density of halos per unit mass as a function of redshift. The most massive halos host galaxy clusters, which in turn can be observed. Therefore by matching the observations of cluster counts with the expected number of halos above a certain mass, we can put constraints on cosmological parameters. Considering a volume element in a comoving coordinate system, $d^2V = dAdl$, we need to account for the dependency of the unit area (dA) and the distance element (dl) on redshift. According to the FRW metric, we can write:

$$dA = (rd\theta)(r\sin\theta d\phi) = a^2 r^2 d\Omega = \frac{a_0^2 r^2 d\Omega}{(1+z)^2}, \quad (1.38)$$

$$dl = \frac{da}{\dot{a}} = \frac{dz}{(1+z)} \frac{a}{\dot{a}} = \frac{dz}{H(z)(1+z)}, \quad (1.39)$$

$$d^2V = dAdl = \frac{a_0^2 r^2}{H(z)(1+z)^3} d\Omega dz. \quad (1.40)$$

If we combine equation 1.40 with the predicted number density of halos above a certain mass, $n(> M, z)$, assuming that each halo hosts a cluster, then the predicted number of clusters with mass above M , per unit solid angle and redshift is:

$$\frac{d^2 N(> M, z)}{dzd\Omega} = n(> M, z) \frac{d^2 V}{dzd\Omega} = \frac{n(> M, z) a_0^2 r^2}{H(z)(1+z)^3}. \quad (1.41)$$

Cosmological analyses can further relate the predictions to observables by defining the probability that a halo is found in a certain mass interval. In practice, cosmological surveys can measure the number of clusters above a certain quantity, λ (e.g. mass, richness etc.) and relate it to the models in this way:

$$\frac{d^2 N(> \lambda, z)}{dzd\Omega} = \frac{r^2(z)}{H(z)} \int_0^\infty f(> \lambda, z) d\lambda \int_0^\infty p(\lambda|M, z) \frac{dn(M, z)}{dM} dM, \quad (1.42)$$

where $f(> \lambda, z)$ is a selection function for the observable quantity λ and $p(\lambda|M, z)$ indicates the probability that a halo of mass M will have a particular value of the observable λ . Finally, $dn(M, z)/dM$ expresses the number density of halos per unit mass.

GRAVITATIONAL LENSING AND THE LIMBER APPROXIMATION

An important observable used to probe dark matter is the phenomenon of gravitational lensing. It occurs when the trajectory of photons travelling from a distant source is bent by the gravitational potential of a massive object. This object consists in a distribution of matter (such as a cluster of galaxies) located between the source and the observer and takes the name of *gravitational lens*. We talk about *strong gravitational lensing* when the bending effect is so intense that the source galaxy in the background is imaged multiple times about the lens and/or is heavily distorted, showing arc-shaped patterns. On the other hand, when the images of galaxies are distorted, stretched and magnified only little, then we are in the *weak lensing* regime. Such effect, called *cosmic shear*, is so small (with distributions of typically a few percent) that it cannot be measured with only one galaxy (as it is done for cases of strong lensing): instead it is studied in terms of the coherent alignment of background galaxy images. This is shown in the lensed panels of Figure 1.2 (left side), where the alignment of the galaxies around the central mass generates distortion in their images. We can model such distortions as a coordinate transformation acting on an original distribution of objects. As we will specify below, the main components of this transformation are the *convergence*, κ , and the *shear*, γ : the first dilates the images of the background galaxies, the second shears them. To estimate the shear, we need to account for statistical errors, referenced as *shape noise*. They come from the contamination by the intrinsic ellipticity of galaxy images, including the effects of the Point Spread Function (PSF).

The intrinsic ellipticity of a galaxy is the natural elongation of its shape before lens-

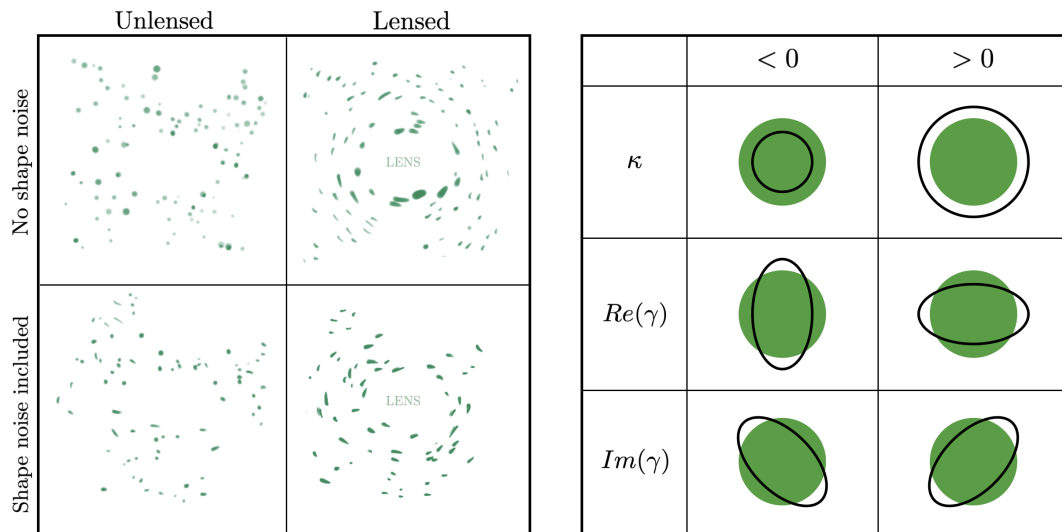


Figure 1.2: Left panel: representation of the distortions induced by lensing, here applied to a random distribution of circular and elliptical objects of various sizes. The lensed objects appear elongated and aligned tangentially to the lens. Right panel: effects of convergence and shear on a circular source (in green). The images have been reproduced taking inspiration from *Wikipedia* (https://en.wikipedia.org/wiki/Weak_gravitational_lensing) and *Wikiwand* (https://www.wikiwand.com/en/Gravitational_lensing_formalism), respectively.

ing. Galaxies are generally not circular, and their ellipticity is much greater than the shear. Models for galaxy morphology (described in section 1.2 and extensively addressed in Chapter 2) approximate the galaxy light distribution with an ellipse of semi-major axis a and semi-minor axis b and define the ellipticity as $\epsilon = 1 - b/a$.

The PSF is a model describing the response of the instrument to a point source. Deviations from a perfect point-like source are due to the properties of the lens of the telescope and to atmospheric effects. They produce smeared images and thus remove information on their shape. To correct for these effects, we measure the PSF of the telescope by observing stars within our own galaxy. Ideally, their images would be perfectly point-like, but in practice they are affected by the instrumental response. Quantifying such effects provides the corrections to be applied to galaxy images across the field of observation. The PSF calibration, being crucial to make accurate structural measurements of galaxies, plays also an important role in galaxy morphology. For additional information and examples we refer the reader to Chapter 2, where we discuss how the PSF affects the structural measurements of small galaxies in the DES dataset.

A schematic view of gravitational lensing is shown in Figure 1.3. The difference between the unlensed angular position, $\bar{\beta}$ and the lensed one, $\bar{\theta}$, is given by the deflection

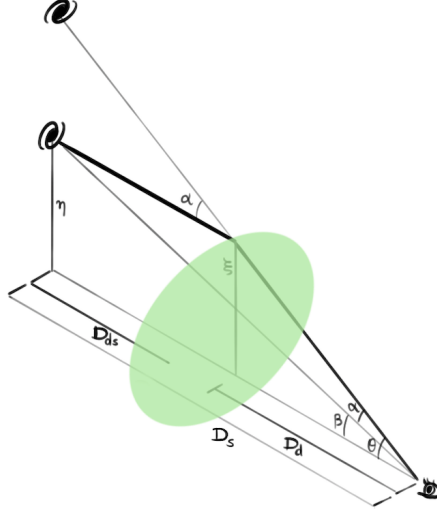


Figure 1.3: Geometrical view of gravitational lensing: the angles α , β and θ give the direction of the real source and the lensed one with respect to the observer. D_s is the radial separation between the source and the lens (in green), D_d between the lens and the observer and D_s between the source and the observer. The image has been reproduced taking inspiration from *Wikiwand* (https://www.wikiwand.com/en/Gravitational_lensing_formalism).

angle $\bar{\alpha}(\bar{\theta})$, expressed by the lens equation:

$$\bar{\beta} = \bar{\theta} - \bar{\alpha}(\bar{\theta}) = \bar{\theta} - \frac{D_{ds}}{D_s} \hat{\alpha}(D_d \bar{\theta}), \quad (1.43)$$

where D_{ds} , D_s and D_d are the angular-diameter distances between the lens and the source, the observer and the source, and the lens and the observer, respectively. The linear transformation between unlensed (x_u, y_u) and lensed (x_l, y_l) coordinates is expressed in terms of convergence κ and shear γ :

$$\begin{pmatrix} x_u \\ y_u \end{pmatrix} = \begin{pmatrix} 1 - \gamma_1 - \kappa & -\gamma_2 \\ -\gamma_2 & 1 + \gamma_1 - \kappa \end{pmatrix} \begin{pmatrix} x_l \\ y_l \end{pmatrix}, \quad (1.44)$$

where the two components γ_1, γ_2 can be combined into a complex number, $\gamma = \gamma_1 + i\gamma_2$. To express the orientation of the shear with respect to the angular separation, ϕ , between two lensed galaxies, it is convenient to define the tangential and cross shear estimators, $\gamma_+ = -\text{Re}(\gamma e^{-2i\phi})$ and $\gamma_\times = -\text{Im}(\gamma e^{-2i\phi})$, respectively. These components can be measured by analyzing the moments of the distribution of light in galaxy images and are used to study the spatial correlation of the shear. As an example, if some image distortion due to weak lensing is measured in a certain direction in the sky, we expect that the distortion measured in a nearby location is similar. Following this, the smaller

the angle chose between two locations, the more similar the distortion is expected to be. On the other hand, when the separation angle becomes larger, then the distortions should become ideally independent. This behavior is modelled by a function called the *angular correlation function*. Let us suppose that a quantity $x(\bar{\theta})$ is measured at different positions θ and directions; its angular correlation function is defined as:

$$\xi_x(\phi) := \langle x(\bar{\theta})x(\bar{\theta} + \bar{\phi}) \rangle, \quad (1.45)$$

where the angular brackets express the average over all positions \bar{x} with separation vector $\bar{\phi}$ on the sky. Using the same formalism presented for the covariance matrix and the matter power spectrum in equations 1.21 and 1.22, it is convenient to calculate the Fourier transform of the angular correlation function, obtaining the *angular power spectrum* for the quantity x :

$$C(l) = \int d^2\phi \xi e^{-i\bar{l}\cdot\bar{\phi}}, \quad (1.46)$$

where \bar{l} is a two-dimensional wave vector and conjugate to the angular separation $\bar{\phi}$. Since the quantity $x(\bar{\theta})$ has a three-dimensional distribution that we want to study on a two-dimensional sky map, we can express it as a function defined in three dimensions, $y(\chi\bar{\theta}, \chi)$, projected onto two dimensions using a weight function, $w(\chi)$:

$$x(\bar{\theta}) = \int_0^{\chi_S} d\chi w(\chi) y(\chi\bar{\theta}, \chi), \quad (1.47)$$

where χ_S is the comoving angular diameter distance from the source to the observer (see equation 4.5 in Chapter 4). The angular power spectrum is often computed using an approximation, instead of computing the full three-dimensional integral (for this calculation, we refer the interested reader to equation 1.58 in the current chapter and to Chapter 4, section 4.2.4, for an application to the CMB angular power spectrum). This method is called the *Limber Approximation* and yields the following power spectrum:

$$C_x(l) \simeq \int_0^{\chi_S} d\chi \frac{w^2(\chi)}{\chi^2} P_y\left(\frac{l}{\chi}\right), \quad (1.48)$$

where P_y is the power spectrum of y taken at the wave-number $k = l/\chi$. We can write the weak lensing correlation functions and power spectra using the formalism specified in equations 1.45, 1.47 and 1.48. The angular correlation functions between the tangential and cross components of the shear, expressing the dependence on angular separation ϕ , are defined as:

$$\begin{aligned}
 \xi_{++}(\phi) &= \langle \gamma_+(\bar{\theta}) \gamma_+(\bar{\theta} + \bar{\phi}) \rangle, \\
 \xi_{\times\times}(\phi) &= \langle \gamma_{\times}(\bar{\theta}) \gamma_{\times}(\bar{\theta} + \bar{\phi}) \rangle, \\
 \xi_{+\times}(\phi) &= \xi_+(\phi) = \langle \gamma_+(\bar{\theta}) \gamma_{\times}(\bar{\theta} + \bar{\phi}) \rangle,
 \end{aligned}
 \tag{1.49}$$

where the last function is expected to be zero.

The convergence can be expressed as the projection of the matter density contrast δ with the weight function:

$$w_{\kappa}(\chi) = \frac{3}{2} \frac{H_0^2}{c^2} \Omega_{m0} \frac{\chi(\chi_S - \chi)}{a\chi_S},
 \tag{1.50}$$

and the convergence angular power spectrum can be calculated using the Limber Approximation:

$$C_{\kappa}(l) = \frac{3}{2} \left(\frac{H_0}{c} \right)^2 \Omega_{m0} \int_0^{\chi_S} d\chi w_{\kappa}(\chi) P_{\delta} \left(\frac{l}{\chi} \right),
 \tag{1.51}$$

where $P_{\delta}(\frac{l}{\chi})$, the matter power spectrum, is important to probe through cosmic shear the cosmic matter distribution. For additional reviews on the formalism and the measures of weak gravitational lensing, we refer the reader to [Bartelmann and Maturi, 2017](#) and [Mandelbaum, 2018b](#).

THE COSMIC MICROWAVE BACKGROUND

Before recombination, the Universe was permeated by a ionized plasma of baryons and photons, made opaque by the highly frequent Thomson scattering of photons by free electrons. This system was in thermal equilibrium, under the competing effect of pressure and gravity: as the latter tried to compress the fluid, radiation pressure resisted, producing Baryon Acoustic Oscillations (BAO). At redshift $z \approx 1000$ (around 300000 years after the Big Bang) and at a temperature of about 3000 K, photons became cool enough to reduce the scattering rate and allow electrons to combine with nuclei to form atoms. At the *surface of last scattering* radiation decoupled from matter and started to travel freely in the Universe. This radiation is observed today as the *relic* Cosmic Microwave Background (CMB). Its energy spectrum is that of a black-body, with a present day temperature $T_{CMB} \simeq 2.726K$. The alternating hot and cold spots generated at earlier times give rise to the anisotropies of the cosmic background radiation we measure today. These anisotropies are a powerful probe of the standard cosmological model because, even filtered through the evolving geometry and dynamics of an expanding universe, they show the imprint of the fluctuations in the early Universe. Modern observations reveal fluctuations in the CMB temperature of the order of $\Delta T/T \approx 10^{-5}$. A

common practice is to decompose the CMB anisotropies on a sphere, by expanding the temperature map in spherical harmonics:

$$\frac{\Delta T}{T} = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell, m}(\theta, \phi), \quad (1.52)$$

where the spherical harmonics, $Y_{\ell, m}(\theta, \phi)$, are normalized so that:

$$\int d\Omega Y_{\ell, m}^*(\theta, \phi) Y_{\ell', m'}(\theta, \phi) = \delta_{\ell, \ell'} \delta_{m, m'}, \quad (1.53)$$

with $d\Omega = \sin\theta d\theta d\phi$. Also, $a_{\ell m}$, the spherical harmonics coefficients, are given by:

$$a_{\ell m} = \int_{\Omega} d\Omega T(\theta, \phi) Y_{\ell, m}^*(\theta, \phi). \quad (1.54)$$

The $a_{\ell m}$ coefficients contain information on the fluctuations from the mean temperature associated to the multipole number ℓ and the azimuthal number, m . In analogy with the matter density field in linear perturbation theory, we can give a statistical interpretation of the anisotropy map: the CMB we observe is a possible realization of a statistical anisotropy distribution. If we assume that the coefficients $a_{\ell m}$ are gaussian distributed, all the information about the field is included in the two-point function:

$$\langle a_{\ell m} a_{\ell' m'}^* \rangle = \delta_{\ell, \ell'} \delta_{m, m'} C_{\ell}, \quad (1.55)$$

where C_{ℓ} is the CMB angular power spectrum and the angle bracket denotes an ensemble average. Given the assumption of homogeneity and isotropy, C_{ℓ} should not depend on the azimuthal number m . However, in formulating the estimator,

$$\hat{C}_{\ell} = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} a_{\ell m} a_{\ell m}^*, \quad (1.56)$$

we are limited by the number of m values for each ℓ . Therefore, in the estimation of C_{ℓ} there is a fundamental uncertainty called *cosmic variance*, given by:

$$\left(\frac{\Delta C_{\ell}}{C_{\ell}} \right) = \left(\frac{2}{2\ell + 1} \right)^{1/2}. \quad (1.57)$$

Using the properties of harmonic functions, the temperature power spectrum can be written as:

$$C_{\ell}^{TT} = \frac{2}{\pi} \int dk k^2 |\Theta_{\ell}(k, \eta_0)|^2 \quad (1.58)$$

where the temperature field, $\Theta_{\ell}(k, \eta_0)$, is given by:

$$\Theta_\ell(k, \eta_0) = \int_0^{\eta_0} d\eta S(k, \eta) j_\ell[k(\eta_0 - \eta)], \quad (1.59)$$

with $S(k, \eta)$ the source function for the temperature field. The presence of the spherical Bessel functions, j_ℓ , makes this integral challenging to evaluate numerically. The temperature field can be expressed in terms of primary anisotropies, arising from the perturbations at the surface of last scattering, and secondary anisotropies, due to the interactions of the photons with hot gas and gravitational potentials, during their journey from the last scattering surface to the observer. Figure 1.4 shows the predicted amplitudes of such fluctuations, that is the CMB angular power spectrum. Its structure is mainly determined by the acoustic oscillations, described earlier, and by *diffusion damping*. The latter phenomenon saw the photons travelling from hot regions in the plasma to cold ones, equalizing the temperature and density of these regions, thus damping anisotropies on small scales. The location of the peaks gives important information on the perturbations in the early Universe (an approximate conversion between the angular scales and multipoles is $\theta \approx 180^\circ/\ell$). Furthermore, we can use the first three peaks to constrain the curvature, the baryon and the dark matter density. Secondary or *late time anisotropies* have smaller effects on the power spectrum. However, they still contain important cosmological information, since they probe the evolution of the Universe from the recombination until today. Important phenomena that affect the observed anisotropies are reionization, gravitational lensing, the Sachs-Wolfe effect and the Sunyaev-Zeldovich effect. We describe them below.

Reionization - this term refers to the process when the first astrophysical objects, e.g. stars, galaxies and Active Galactic Nuclei (AGN), form and start to emit ionizing photons, therefore stripping electrons away of nearby atoms. When CMB photons encounter the resulting free electrons they scatter through Thomson scattering. This produces a suppression of the CMB anisotropies. Reionization happens inhomogeneously in the Universe, since galaxies tend to follow a clumpy distribution. This patchy phenomenon produces secondary anisotropies on small scales (at $\ell \simeq 1000$).

Gravitational lensing - the presence of mass along the line of sight causes deviations in the path of the CMB photons (see Section 1.1.4 above, for a description of gravitational lensing and the distortions produced on galaxy images). The effects on the angular power spectrum result in a smoothing of the peaks and some power on the smallest scales ($\ell \geq 3000$).

Integrated Sachs-Wolfe effect (ISW) - this effect occurs when the CMB photons get redshifted or blueshifted due to a time-varying gravitational potential along their path. In the interaction with a gravitational potential well a photon gains and loses energy when it falls into and escapes from it, respectively. If the amplitude of a potential well decreases before the photon finishes to cross it, then the photon gains energy overall. We

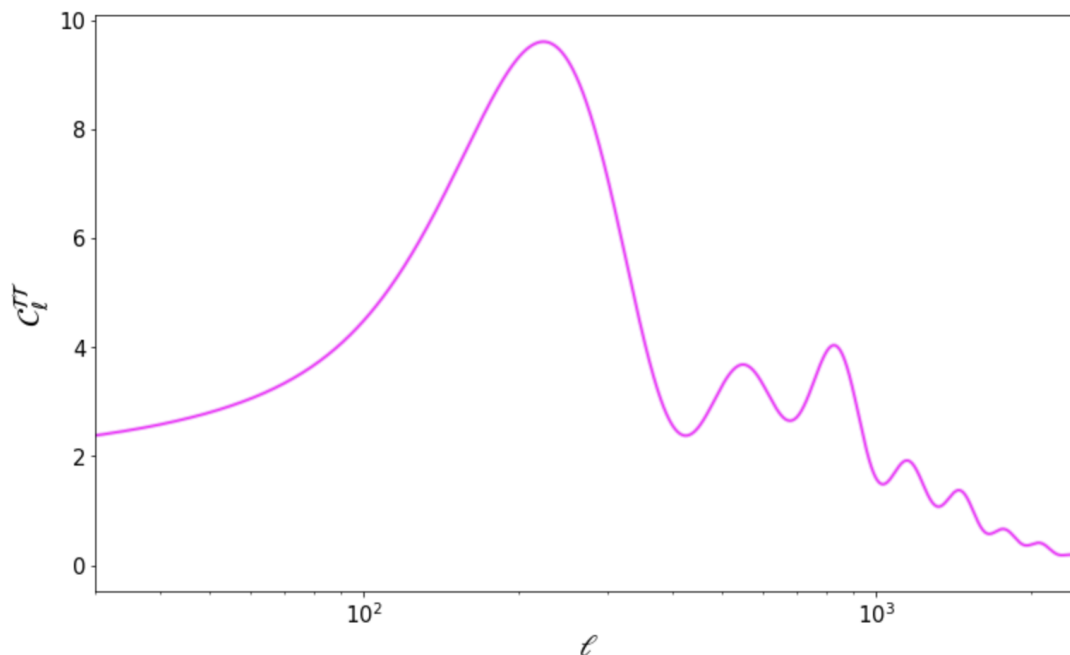


Figure 1.4: Prediction for the CMB temperature angular power spectrum, computed using the fiducial settings in PyCosmo (see Section 4 for more details).

can distinguish between early ISW and late ISW effect, occurring around last scattering and later on, respectively. Such effect is important on large scales.

Sunyaev-Zeldovich effect - a dense structure containing hot gas, such as a galaxy cluster, along the line of sight of the CMB photons, induces interactions between these photons and the free electrons in the hot gas. The involved photons gain energy through inverse Compton scattering.

In addition, anisotropies can occur due to other astronomical sources, such as the Milky Way and other galaxies, which emit signals in the wavelength range of the CMB photons. These contributions are taken into account in CMB analyses.

Details about how we model the temperature field and we calculate the CMB angular power spectrum using the line-of-sight integration are described in Chapter 4, section 4.2.1. We also point to [Wayne Hu and Dodelson, 2002](#) and the [Planck Collaboration, 2020](#) for additional information on the theory underlying the CMB and the latest cosmology results, respectively.

1.2 GALAXY MORPHOLOGY

As mentioned in the first part of this chapter, galaxies are well-defined astronomical objects which can be observed at different cosmic times. A typical bright galaxy, such as our Milky Way, contains $\simeq 10^{10}$ stars, and the stellar number density within a galaxy

is about 10^7 times higher than the mean number density of stars in the Universe. This consideration leads us to consider the first, most important properties of galaxies: they are sources of light. And light is the messenger to our telescopes. Light coming from galaxies is collected in the pixels of an image, which can be used to study the morphology of these objects.

Galaxies constitute a diverse class of objects. This variety is perceived already by looking at their shapes and distribution of light, revealing unique features such as spiral arms, bars and clumps. According to the Hubble sequence, galaxies are divided between ellipticals, spirals and irregulars, as shown in Figure 1.5. Ellipticals have mostly ellipsoidal shapes, and are self-supported by the random motion of their stars. Spirals present a flat disk structure where stars are organized in spiral arms uncoiling from the center. They are rotation-supported and their radial extension determines their angular momentum. However, this distinction is not comprehensive. Most galaxies are neither perfect ellipsoid nor perfect disks, and they can present features belonging to an intermediate state between these two classes. For instance, they can be a combination of both, being formed by two components, a disk and a bulge. And they can be either disk-dominated or bulge-dominated. Irregular galaxies don't enter the aforementioned categories: they can show multiple sub-components and/or filamentary structures that come from mergers or tidal interactions with nearby galaxies. Another type of galaxies, called dwarf galaxies, are neither ellipticals nor spirals. For a complete review of galaxy morphological types we refer the reader to e.g. [Alister W. Graham, 2019](#).

Irregular galaxies provide an example of how morphology can represent the signature of physical processes occurring to them during their evolutionary path. Regarding mergers between galaxies, the final stage of this process can also lead to one single galaxy with morphology depending from the structure and the dimensions of the merging galaxies. In particular, a major merger between two spirals, occurring when the two galaxies have comparable sizes, destroys the galactic disks and dissipates the spiral arms and ultimately forms a giant elliptical. On the other hand, a minor merger, which sees a galaxy colliding with a much smaller one, does not disturb the morphology of the major galaxy. For example, a minor merger between a spiral and a dwarf galaxy typically leads to an enhanced spiral. Morphology is related to other physical properties. In the relation between mass and Star Formation Rate (SFR) ([Schiminovich et al., 2007](#); [Goncalves et al., 2012](#); [Peterken et al., 2021](#)), it plays an important role in distinguishing between three main populations of galaxies. The *main sequence* includes star-forming (blue) galaxies presenting morphological features typical of spirals and irregulars. Objects in this population are also called *Late-Type Galaxies* (LTG). A second population is defined by galaxies with lower SFR and shapes which are mostly elliptical and bulge-dominated. They are also called *Early-Type Galaxies* (ETG). The smooth transition between these two regions in the diagram is called *green valley* and is less populated. Additional correlated

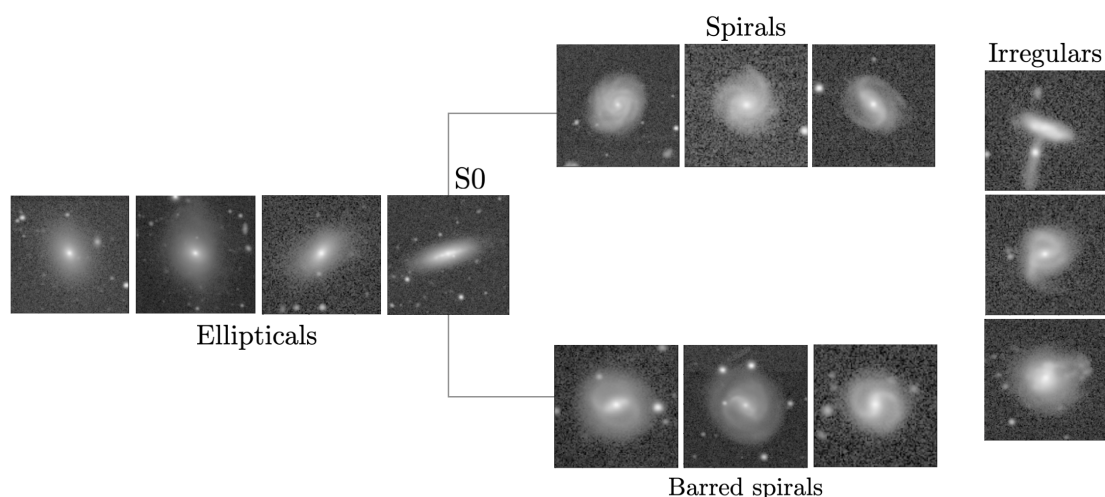


Figure 1.5: The Hubble tuning fork, made with public galaxy images from the DES Collaboration. The galaxies are divided between ellipticals (the first line on the left), spirals (the two following branches on the right) and lenticulars or S0 (the node of the tuning fork), showing intermediate morphological traits. Irregular galaxies, represented in the right end of the diagram, are characterized by disturbed morphology and do not fall into the above categories.

properties of the galaxy population are addressed in the introductory sections of Chapters 2 and 3. Distinguishing between ETG, LTG and irregulars has been historically done through classification by-eye. With the large datasets made available by modern survey this practice, even though fascinating, has become obsolete and unrealistic. Likewise, it is still possible through *citizen science* projects, where the public can access dedicated websites and contribute to the classification of millions of galaxy images (Lintott et al., 2008; Simmons et al., 2017; Willett et al., 2017). On a different approach, the recent development of Machine Learning (ML) algorithms brought great contributions to galaxy classification. Trained Convolutional Neural Networks have been proven effective for direct automated galaxy classification (see Chapter 3 for more details). While this represents an important step forward, it is still complementary to other methods used to extract information from galaxy images. We can indeed quantify the size of a galaxy, its brightness, ellipticity, inclination, concentration, and the rate at which light falls off radially towards the outskirts. We can then study how such quantities evolve with cosmic time, and relate them to other properties such as color, mass and environment (e.g. to study which type of galaxies tend to be isolated or reside in clusters). In order to extract this type of information we need to use more standard approaches, namely *parametric* and *non-parametric* methods. We refer the reader to the introductory section of Chapter 2, and to sections 2.4 and 2.5 for a detailed description of these two methodologies and their main differences.

A morphological catalogue for DES Y1



With great power comes great responsibility!

STAN LEE, *AMAZING FANTASY, SPIDER-MAN*, 1962

The content of this chapter is based on Tarsitano, Hartley, et al., 2018.

2.1 INTRODUCTION

Any explanation of the formation and evolution of galaxies must necessarily include a description of the diverse forms that galaxies take. The morphology of the luminous components of a galaxy, including its classification or decomposition into a bulge and disk (e.g., [Kormendy, 1977](#); [de Jong, 1996a](#)) or identification of features such as bars, rings or lenses (e.g., [Kormendy, 1979](#); [Combes and Sanders, 1981a](#); [B. G. Elmegreen et al., 1996](#)), are a result of its aggregated formation history. Assigning meaningful morphological types or quantifying the distribution of light across the extent of a population of galaxies, is therefore of fundamental importance in understanding the processes that govern their evolution.

A quantitative description of galaxy morphology is typically expressed in terms of structural parameters (brightness, size, shape) and properties of the light distribution (concentration, asymmetry and clumpiness), though human classifications are still used. The development of *citizen science* projects like Galaxy Zoo ([Lintott et al., 2008](#); [Simmons et al., 2017](#); [Willett et al., 2017](#)) and sophisticated machine learning algorithms ([Lahav et al., 1995](#); [Lahav, 1995](#); [Huertas-Company, Rouan, et al., 2008](#); [Huertas-Company, Gravet, et al., 2015b](#); [Banerji et al., 2010](#); [Dieleman, Willett, and Dambre, 2015](#)) have helped to maintain the relevance of these perception-based morphologies in the current literature.

Nevertheless, most recent work on the subject of galaxy morphologies rely on either *parametric* or *non-parametric* approaches to quantify the galaxy's light distribution.

Parametric methods fit two-dimensional analytic functions to galaxy images. The mathematical model of the light fall-off is convolved with the point spread function (PSF) to take into account the seeing. The most general assumed function for this purpose is the Sérsic profile (Sérsic, 1963). The second class, non-parametric methods, perform an analysis of the light distribution within a certain elliptical area, usually defined through the Petrosian radius associated with the galaxy. Common estimates are of the degree to which the light is concentrated, quantifying the asymmetry of the light distribution and searching for clumpy regions: this method is called *CAS system* (Concentration, Asymmetry and Smoothness or Clumpiness) and can be extended with further parameters, Gini and M20 (Conselice, 2003; R. G. Abraham, van den Bergh, and P. Nair, 2003; Lotz, Primack, and Madau, 2004; Law et al., 2007). These parameters together can describe the major features of galaxy structure without resorting to model assumptions about the galaxy's underlying form, as is done with the Sérsic profile. However, they are determined without a PSF deconvolution and need an additional calibration.

Even alone, distributions of morphological quantities represent powerful constraints on possible galaxy formation scenarios. But combined with other physical quantities, they can provide key insights into the processes at play, supporting or even opening new ideas on evolutionary mechanisms (Kauffmann et al., 2004; Weinmann et al., 2006; K. Schawinski, Kaviraj, et al., 2007; van der Wel, 2008; van der Wel, A., 2008; Bamford et al., 2009; K. Schawinski, Urry, et al., 2014). For instance, the relationship between the masses, luminosities and sizes of massive disks and spheroids suggests dissipative formation processes within hierarchical dark matter assembly (S. D. M. White and Rees, 1978; Fall and Efstathiou, 1980) or the occurrence of galaxy-galaxy mergers (A. Toomre and J. Toomre, 1972; A. Toomre, 1977; Barnes, 1988; Naab and Burkert, 2003; Conselice, 2003; Lin et al., 2004; Conselice, 2008; Conselice, Rajgor, and Myers, 2008; Jogee, 2009; Jogee et al., 2009). On the other hand, analysing galaxy sub-structure (e.g. with a bulge + disk decomposition) can open up evidence of further mechanisms: bulges, disks and bars may be formed by secular evolution processes (Kormendy, 1979; Kormendy and Kennicutt, 2004; Bournaud, B. G. Elmegreen, and D. M. Elmegreen, 2007; Genzel et al., 2008; Fisher and Drory, 2008; Sellwood, 2014) or by the interplay between smooth and clumpy cold streams and disk instabilities (Dekel, Birnboim, et al., 2009; Dekel, Sari, and Ceverino, 2009). In this sense bulges may be formed without major galaxy mergers, as is often thought.

Of particular interest in recent years, have been the questions over the degree to which galaxy environment impacts upon morphology (e.g. Dressler, 1980; Postman et al., 2005; Lani et al., 2013; Kuutma, Tamm, and Tempel, 2017), and the connection between morphology and cessation of star formation in galaxies (e.g. Blanton et al., 2003;

[Martig et al., 2009](#); [Bell et al., 2012](#); [Woo et al., 2015](#)). Faced with often subtle correlations or hidden variables within strong correlations, these questions demand far greater statistical power and measurement precision than had been possible from the available data sets in the preceding decades. These demands require efficient pipelines to automate and streamline the analysis of large astronomical data sets. GALAPAGOS ([Gray et al., 2009](#); [Häußler et al., 2011](#); [Barden et al., 2012](#)) is perhaps the most widely used of such pipelines. It offers a routine to simplify the process of source detection, to cut postage stamps, prepare masks for neighbours if needed and estimate a robust sky background and has been used at both low redshift in the GEMS survey ([Häußler et al., 2007](#)), and at higher redshift on the CANDELS ([Wel et al., 2012](#)) data.

At low redshift the state-of-the-art to date are the catalogues constructed from Sloan Digital Sky Survey (SDSS, [York, Adelman, John E. Anderson, et al. 2000](#)) data, in particular the bulge+disk catalogue of [Simard et al., 2011](#) numbering almost 1 million galaxies. Such statistical power has been lacking at higher redshifts, but the advent of large-scale cosmology experiments optimised for weak lensing analyses, such as the Dark Energy Survey (DES) and Hyper Suprime-Cam (HSC) ([Miyazaki et al., 2012](#)), provide a great opportunity to fill in much of this gap. DES is the largest galaxy survey to date, with a narrower PSF and images typically two magnitudes deeper than the SDSS.

In order to create as complete a set of structural measurements for DES as possible we adopt both parametric and non-parametric approaches, using the software GALFIT ([Peng, C. Y. et al., 2002](#); [C. Y. Peng et al., 2010](#)) for parametric Sérsic fitting and ZEST+ for a non-parametric analysis of the structural properties of our galaxy sample. The first provides us with the measurements of the magnitude, effective radius, Sérsic Index, axis ratio and orientation angle of the galaxy; the second one outputs an extended set of parameters, completing the CAS system with Gini and M20, plus the values of magnitude, half light radius and ellipticity, measured within the galaxy Petrosian ellipse.

The scale of the DES data set requires a new dedicated pipeline in order to handle the DES data structure, optimise run-time performance, automate the process of identifying and handling neighbouring sources and prepare tailored postage stamps for input to the two software packages. The resulting dataset is by far the largest catalogue of structural parameters measured to date, numbering 45 million galaxies, which exceeds previous catalogues by more than an order of magnitude in size, and reaches redshift, $z \sim 1$. It includes parametric and non-parametric measurements in three photometric bands, intended to be used in concert and to provide a comprehensive view of the galaxies' morphologies. In this sense, our catalogue constitutes a significant step in our capabilities to study the nature of galaxy morphology in the Universe.

This paper is structured as follows: in Section 2.2 we give an overview of the Dark Energy Survey, describing the data and the image simulation data we used for this work. In Section 2.3 we focus on the details of our sample selection and pre-fitting routine, pre-

presenting the algorithms developed to prepare and process the data. Sections 2.4 and 2.5 are dedicated to the parametric and non-parametric fits, respectively. In each of these two sections, we present a detailed description of the fitting software used for this work, discuss the completeness and validation of the fitted sample from each method, provide an overview of the characteristics of the catalogue and perform a calibration of the output quantities with image simulations. The calibration for the i band are shown in those sections; Section A.1 of Appendix A includes the calibration maps also for the g and r filters. Section 2.5 also introduces a set of basic cuts as a starting point in building a science-ready sample. Finally in Section 2.7 we summarise our work. A manual explaining the catalogue columns is presented in Section A.2 of Appendix A.

2.2 DATA

2.2.1 THE DARK ENERGY SURVEY

The Dark Energy Survey (DES) (DES Collaboration, 2005; The DES Collaboration, 2016) is a wide-field optical imaging survey covering 5000 deg² of the southern equatorial hemisphere in $grizY$ bands¹. Survey observations began in August 2013 and over five years it will provide images of 300 million of galaxies up to redshift ~ 1.4 (Diehl, T. M. C. Abbott, J. Annis, Armstrong, et al., 2014). The survey is designed to have a combination of area, depth and image quality optimized for cosmology, and in particular the measurement of weak gravitational lensing shear. However, its rich data set is well-suited to many areas of astronomy, including galaxy evolution, Milky Way and Local Group science, stellar populations and Solar System science (T. Abbott et al., 2016).

DES uses the Dark Energy Camera (DECam), a mosaic imager with a 2.2° diameter field of view and a pixel scale of 0.263'' per pixel mounted at the prime focus of the Victor M. Blanco 4m Telescope at Cerro Tololo Inter-American Observatory. During the requested 525 observing nights it is expected to reach photometric limits of $g = 24.6$, $r = 24.4$, $i = 23.7$, $z = 22.7$ and $Y = 21.5$ (10σ limits in 1.5'' apertures assuming 0.9'' seeing) following ten single-epoch exposures of 90 seconds each for $griz$ and 45 seconds each for Y (Flaugher, 2005).

The DES data are processed, calibrated and archived through the DES Data Management (DESDM) system (Drlica-Wagner, Sevilla-Noarbe, Rykoff, Gruendl, Yanny, Tucker, Hoyle, Carnero Rosell, et al., 2017; Morganson, Gruendl, Menanteau, Carrasco Kind, Y.-C. Chen, et al., 2018), consisting of an image processing pipeline which performs image de-trending, astrometric calibration, photometric calibration, image co-addition and SExtractor (Bertin and Arnouts, 1996) catalogue creation. The DESDM imaging co-addition combines overlapping single-epoch images in a given filter, which are then

¹<http://www.darkenergysurvey.org>

remapped to artificial tiles in the sky so that one co-add image per band is produced for every tile. These tiles are padded to ensure that each object is entirely contained in at least one tile, but also results in a small fraction of duplicate objects found in different tiles which are removed at a later stage. In order to account for PSF variations caused by object location in the focal plane and the combination of images with different seeing, the catalogue creation process uses PSFex (Bertin, E., 2011; Bertin, 2013) to model the PSF. PSFex produces a basis set of model components on the same pixel scale as the science image that are combined via linear combination into a location-dependent PSF. The final step combines the photometry of each co-add object into a single entry in multi-wavelength SExtractor catalogues. For more details about the DESDM co-addition and PSF modelling we refer the reader to Sevilla et al. 2011, Desai et al. 2012 and Mohr et al. 2012.

In this work we use the DES Y1A1 COADD OBJECTS data release, comprising 139,142,161 unique objects spread over about 1800 deg^2 in 3707 co-add tiles, constructed from the first year of DES survey operations. The tiles are combinations of 1-5 exposures in each of the *grizY* filters and the average coverage depth at each point in the retained footprint is ~ 3.5 exposures. We consider 3690 tiles in total: the catalogue for the remaining tiles, located in the 30 deg^2 of cadenced supernovae fields, will be presented in future work. The data include all the products of the DESDM pipeline and imaging co-addition (the co-add tiles and their respective segmentation maps, the PSF models and the SExtractor catalogues), plus the Y1A1 GOLD catalogue (Drlica-Wagner, Sevilla-Noarbe, Rykoff, Gruendl, Yanny, Tucker, Hoyle, Carnero Rosell, et al. 2017). In the Y1A1 GOLD catalogue, the data collected in DES year-one have been characterised and calibrated in order to form a sample which minimises the occurrence of artefacts and systematic features in the images. It further provides value-added quantities such as the star-galaxy classifier MODEST and photo-z estimates. GOLD magnitudes are corrected for interstellar extinction using stellar locus regression (SLR) (High et al., 2009). We combine the SExtractor DESDM catalogues with the Y1A1 GOLD catalogue to make the sample selection, as described in section 2.3.1, and we also benefit from the application of the MODEST classifier during the analysis of the completeness of our fitting results, reported in more detail in section 2.4.2.

2.2.2 IMAGE SIMULATION DATA

In fitting galaxy light profiles, faint magnitude regimes are well known to present larger systematic errors in the recovered galaxy sizes, fluxes and ellipticities (R. A. Bernstein, Freedman, and Madore, 2002; Häussler et al., 2007; Melchior and Viola, 2012). A larger FWHM of the PSF can also introduce increased uncertainties and systematic errors during morphological estimation. In order to overcome these issues we use sophisticated

image simulations to derive multi-parameter vectors that quantify any biases arising from our analyses, data quality or modelling assumptions. The simulations we use for this purpose are produced by the *Ultra Fast Image Generator* (UFIG) (Bergé et al., 2013) run on the Blind Cosmology Challenge simulation (BCC, Busha et al. 2013) and released for DES Y1 as UFIG-BCC.

UFIG-BCC covers an area of 1750 deg^2 and includes images which are calibrated to match the DES Y1 instrumental effects, galaxy distribution and survey characteristics. Briefly, an input catalogue of galaxies is generated based on the results of an N-body simulation with an algorithm to reproduce the observed luminosity and colour-density relations.

2.3 PRE-FITTING PIPELINE

In this section we describe first the sample selection we apply to the DES Y1A1 COADD OBJECTS, discussing the cuts applied and the initial distributions. Then we describe the process which prepares the data to be fitted both with parametric and non-parametric approach.

2.3.1 SAMPLE SELECTION

For this work we use a tile-by-tile approach, independently for each filter: every step from the sample selection itself to the fitting process is performed separately in each tile and band, with the exception of an overall *i*-band magnitude cut and fiducial star-galaxy separation. We organise the Y1A1 GOLD catalogue into sub-catalogues to include the objects in each co-add tile and match them with the relevant DESDM SExtractor catalogues, extracted from that tile. We apply cuts to specific flags in the catalogues and to the parameters we use as priors for the fits in order to remove the most probable point-like sources, whilst avoiding removing galaxies. In addition we remove a small amount of the survey area in order to work with objects whose SExtractor detection and images are reliable and well-suited for the fitting process. An object is selected if it fulfils the following requirements:

- $\text{FLAGS}_X = 0$;
- $\text{GOLD_MAG_AUTO_I} \leq 23$;
- $\text{FLUX_RADIUS}_X > 0$;
- $\text{KRON_RADIUS}_X > 0$;
- $\text{CLASS_STAR_I} < 0.9$;
- $\text{FLAGS_BADREGION} = 0$,

2.3. Pre-fitting Pipeline

SELECTION TYPE	SELECTION CUT
Gold match	IN_GOLD = True
Image flags	FLAGS_x = 0
S/G	CLASS_STAR_i \leq 0.9
Magnitude	MAG_AUTO_i \leq 23
Size (I)	FLUX_RADIUS > 0 px
Size (II)	KRON_RADIUS > 0 px
Regions	FLAGS_BADREGION = 0

Table 2.1: Summary of the cuts applied to the overlapping sample between the catalogue provided by the DESDM pipeline and the Y1A1 GOLD catalogue. The selected objects must satisfy the requirements described in section 2.3.1. x identifies the filter ($x = g, r, i$).

where $X = g, r, i, z, Y$. The cut in FLAGS removes objects that are either saturated, truncated or have been de-blended. We apply the cuts using the i band as our reference band; indeed the seeing FWHM in this filter is on average the smallest of the five bands. In using the CLASS_STAR classifier at this stage we perform a conservative star-galaxy discrimination (S/G), so that we attempt a fit for any object which could be a galaxy. During the validation analysis we will remove further objects, applying a stricter classifier, named MODEST. We refer to section 2.4.2 for its definition and more details about its impact on this work.

By GOLD_MAG_AUTO we refer to the SExtractor quantity MAG_AUTO, corrected by photometric calibration through SLR as provided by the Y1A1 GOLD catalogue (Drlica-Wagner, Sevilla-Noarbe, Rykoff, Gruendl, Yanny, Tucker, Hoyle, Carnero Rosell, et al. 2017). In the following sections we will simply use the original uncalibrated SExtractor MAG_AUTO. The AUTO photometry is calculated with an elliptical aperture of radius, 2.5 Kron radii. FLUX_RADIUS is the circular radius that encloses half of the light within in the AUTO aperture. Throughout this work, we use KRON_RADIUS to refer to the semi-major axis of the Kron ellipse, i.e. the SExtractor values A_IMAGE and KRON_RADIUS multiplied together.

FLAGS_BADREGION is a flag from the Y1A1 GOLD catalogue tracing the objects that lie in problematic areas, which are close to high-density stellar regions and/or present ghosts and glints. The sample selection cuts described above are summarised in Table 2.1. The normalised distributions of the variables considered during the initial cuts, comparing the selected sample of 45 million objects with the entire dataset (in grey), are shown in Fig. 2.1.

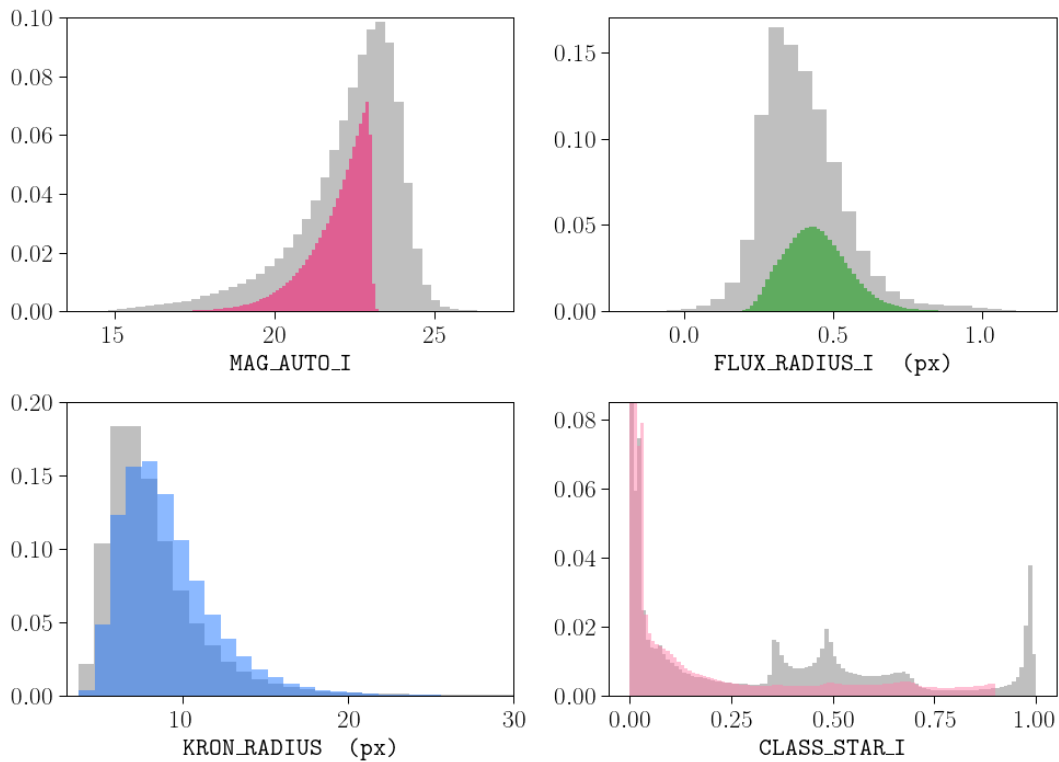


Figure 2.1: Normalised distributions of the variables involved in the sample selection in the *i* band. From upper left to bottom right: `MAG_AUTO`, `CLASS_STAR`, `FLUX_RADIUS` and `KRON_RADIUS`. The cuts applied to each variable are described in more detail in section 2.3.1 and summarised in Table 2.1. In each panel the grey histogram refers to the whole dataset, while the coloured one represents the distribution in that variable for the selected sample.

2.3.2 DATA PROCESSING

The co-add data used in this work are processed in a dedicated pre-fitting pipeline, called Selection And Neighbours Detection (SAND), which has been developed in order to prepare the postage stamps to be fit, their ancillary files in the formats required by GALFIT and ZEST+ and perform essential book-keeping operations. The pipeline performs three steps: *sample selection* (as described in section 2.3.1), *stamp cutting* and *identification of neighbouring sources*. It is important to note that the objects excluded by our initial sample selection (section 2.3.1) are still fit as neighbouring objects where appropriate. For this reason dedicated flags are assigned to each object in the sample, in order to trace their CLASS_STAR classification and possible anomalies in their photometric and structural properties. Collectively, we refer to these flags as STATUS_FLAGS, and document the components and possible values in Appendix A.

For each selected object, an image postage stamp is created, initially with half-width equal to 3 times its Kron radius². Using the relevant segmentation map, the algorithm calculates the percentage of pixels that are not associated with sources (i.e. are background pixels) and approves the stamp if the sky fraction is at least 60%. Otherwise, the image stamp is rejected and is enlarged in size in integer multiples of Kron radius until this requirement is satisfied.

The last step of the pre-fitting routine is dedicated to the identification and cataloguing of neighbours: using the postage segmentation maps it locates the neighbouring objects and, with the above mentioned STATUS_FLAGS, identifies nearby potential stars and/or galaxies with unreliable SExtractor detection. With this last expression we refer to the objects which have unphysical SExtractor parameters (negative sizes, magnitude set to standard error values) and/or are flagged as truncated or saturated objects. In addition to their coordinates and SExtractor properties, the routine catalogues the relative SExtractor magnitude and the presence of overlapping Kron-like isophotes between the central galaxy and its neighbours: these cases are then classified with two dedicated flags, called ELLIPSE_FLAGS and MAX_OVERLAP_PERC, which are fully described in Section A.2 of Appendix A³. This information is now easily accessible during the parametric fitting routine and helps to make decisions on the models to be used to simultaneously fit the objects lying in each stamp (see section 2.4.1); indeed, they are crucial also to the non-parametric approach, since they communicate to ZEST+ all the necessary information to clean the neighbours in the stamps and prepare them for the measuring routine which is described in section 2.5.1.

²i.e. $\text{SETRACTOR_KRON_RADIUS} \times \text{A_IMAGE}$.

³By *Kron-like isophote* we refer to the Kron ellipse enlarged by a factor of 1.5.

2.4 PARAMETRIC FITS

2.4.1 GALFIT SETUP

Image cutouts and PSF models appropriate to each individual object are provided to GALFIT, which is used to find the best-fitting Sérsic models. As reported in (Peng, C. Y. et al., 2002; C. Y. Peng et al., 2010), the adopted Sérsic function has the following form:

$$\Sigma(r) = \Sigma_e \exp \left\{ -k \left[\left(\frac{r}{R_e} \right)^{\frac{1}{n}} - 1 \right] \right\}, \quad (2.1)$$

where Σ_e is the pixel flux at the half-light radius R_e . The Sérsic index n quantifies the profile concentration: if n is large, we have a steep inner profile with a highly extended outer wing; inversely, when n is small, the inner profile is shallow and presents a steep truncation at large radii. In the case of $n = 1$ we have an exponential light profile. We indicate with k the normalization constant coupled to the Sérsic index so that the estimated effective radius always encloses half of the flux (elsewhere, b_n is sometimes used for this quantity). GALFIT produces measurements for the free parameters of the Sérsic function: central position, integrated magnitude (m_{tot}), effective radius (R_e) measured along the major axis, Sérsic index (n), axis ratio (q) and position angle (PA). The integrated magnitude is determined through its definition as a function of the flux (F_{tot}) integrated out to $r = \infty$ for the Sérsic profile:

$$m_{tot} = -2.5 \log \left(\frac{F_{tot}}{t_{exp}} \right) + mag_zpt, \quad (2.2)$$

where t_{exp} is the exposure time and mag_zpt is the zero-point magnitude, both indicated in the image header.

Apart from the central position, which is allowed to vary by only ± 1 pixel by a GALFIT constraints file, all the parameters are left free without constraints: for those, initial guesses are taken from the SExtractor DESDM catalogues (the exception being Sérsic index, which is always started at $n = 2$ and, according to our tests, produces negligible fluctuations in the output if started at other values). Thanks to the large background area available in each stamp (pre-validated with the SAND algorithm), GALFIT is left free to estimate the background level⁴.

For the measurements, GALFIT is left free to build the sigma-image internally. We explored different sizes of the cutout images and convolutions boxes, sequentially enlarging the image until convergence was achieved. Given X and Y the dimensions of the

⁴During initial tests on the fitting routine we randomly selected a sub-sample of objects to be fitted with the background fixed to zero. The outcome of this test was that this choice does not change significantly the results.

cutout image (in pixels), we set the dimensions of the convolution box to $(X + 2, Y + 2)$ pixels. The information provided by the SAND routine is adopted in order to optimise the simultaneous fitting procedure of the central galaxy and its neighbours. Using the ELLIPSE_FLAGS (introduced in section 2.3.2) it is easy to identify most of the neighbours, including faint companions, nearby stars, close objects with overlapping isophotes and neighbours with unreliable priors due to unphysical SExtractor measurements.

Companion objects three magnitudes fainter than the main galaxy are not fit. In the presence of overlapping isophotes, the relevant neighbouring object is fit simultaneously with the target galaxy (even in the cases where it is centred outside the stamp). However, if the overlapping region is 50% or larger than the area within the Kron-like ellipse occupied by the central galaxy, then although a fit is attempted, it is not considered for the analysis discussed in this paper. Given $k1$ and $k2$ as the effective Kron Radii of the central galaxy and its neighbour respectively, they are used to define the isophotes of those objects, intended as enlarged Kron-like ellipses. If the isophotes are not overlapping, but separated by less than the maximum between $k1$ and $k2$, then the neighbour is fit simultaneously. Otherwise it is masked. If the neighbour is a star ($\text{CLASS_STAR} \geq 0.9$), it is simultaneously fit with a PSF model. Finally, if the stamp contains one or more neighbours whose initial guesses from SExtractor contain errors (for example negative magnitudes and radii), no fit is attempted. We adopt a Single Sérsic model with all its parameters free for neighbours also.

2.4.2 FITTING COMPLETENESS

GALFIT uses a non-linear least-squares algorithm which iterates χ^2 minimization in order to find the best solution given a large parameter space. However even when the algorithm outputs a solution, there could be cases where the estimation of one or more parameters is affected by numerical convergence issues, which makes the solution itself an unreliable and non-unique result. These cases include correlated parameters, local minima and mathematically degenerate solutions (C. Y. Peng et al. 2010, Section 6). GALFIT labels the affected parameters enclosing them in between stars (**). In such cases we classify the fit as non-converged and do not trust the set of structural parameters it provides.

We determine the fraction of converged and non-converged fits and investigate their properties and location in the DES field. We present our analysis for all filters taking the i band as reference to discuss the fitting properties and possible causes of failure and incompleteness.

We evaluate the fitting completeness by calculating the percentage of converged fits in differential bins of 0.2 magnitude. The completeness (\mathcal{C}) is calculated by normalising the

number of converged fits in each magnitude bin ($N(c|mag)$) to the number of objects which passed the sample selection (described in section 2.3.1) in that bin, as expressed in the following definition:

$$\mathcal{C}_{|mag} = \frac{N(c|mag)}{N(c|mag) + N(nc|mag) + N(f|mag)}, \quad (2.3)$$

where $N(nc|mag)$ and $N(f|mag)$ refer to the fractions of non-converged and failed fits in each magnitude bin, respectively. We also derive the percentage of converged fits calculated within limiting magnitudes.

The results of this analysis are shown in Fig. 2.2. In the upper left inset (Panel A) the solid lines represent the fitting completeness in magnitude bins and the dashed lines the magnitude limited completeness. They are colour-coded by filter: green and orange lines refer to g and r band, respectively; brown and black to the i band. We start our discussion from the latter.

The dashed black line shows the completeness determined for a sample with a conservative star-galaxy (S/G) cut ($CLASS_STAR > 0.9$): the trend shows that $\sim 90\%$ of the fits are successful at magnitude ~ 17 , after which this value starts to decline and reaches $\sim 80\%$ at magnitude ~ 21 . The completeness decreases more rapidly towards fainter magnitudes.

The brown line shows the completeness after applying a star-galaxy cut based on the $SPREAD_MODEL$ parameter (further details on the star-galaxy classifier and associated analysis are described in the following subsection). In this galaxy sample, a completeness of $\sim 85\%$ is reached at magnitudes $I < 21.5$.

We match the information given by the first panel with the map in Panel B: each point represents a DES tile and is colour-coded by the percentage of converged fits in that tile. The area identified by empty grey circles, where $100 < ra < 60$ and $-70 < dec < -58$, has been excluded from the sample selection because in the GOLD catalogue it is flagged due to its vicinity to the Large Magellanic Cloud (LMC).

We observe that the regions with a higher percentage of non-converged fits are located at the East and West borders of the footprint, towards the Galactic plane. These regions are characterized by high stellar density, as shown in Pieres et al., 2017. One possibility is that many of the unconverged fits at relatively bright magnitudes are stellar contaminants and so there is a poorer completeness where the stellar spatial frequency is higher. Another scenario could be that the edges of the footprints were observed under poorer conditions, for instance with poorer seeing. We now investigate these possibilities through examining correlations between our fitting completeness and maps of survey characteristics (as introduced in Leistedt et al. 2016 and Drlica-Wagner, Sevilla-Noarbe, Rykoff, Gruendl, Yanny, Tucker, Hoyle, Carnero Rosell, et al. 2017), and discuss

2.4. Parametric Fits

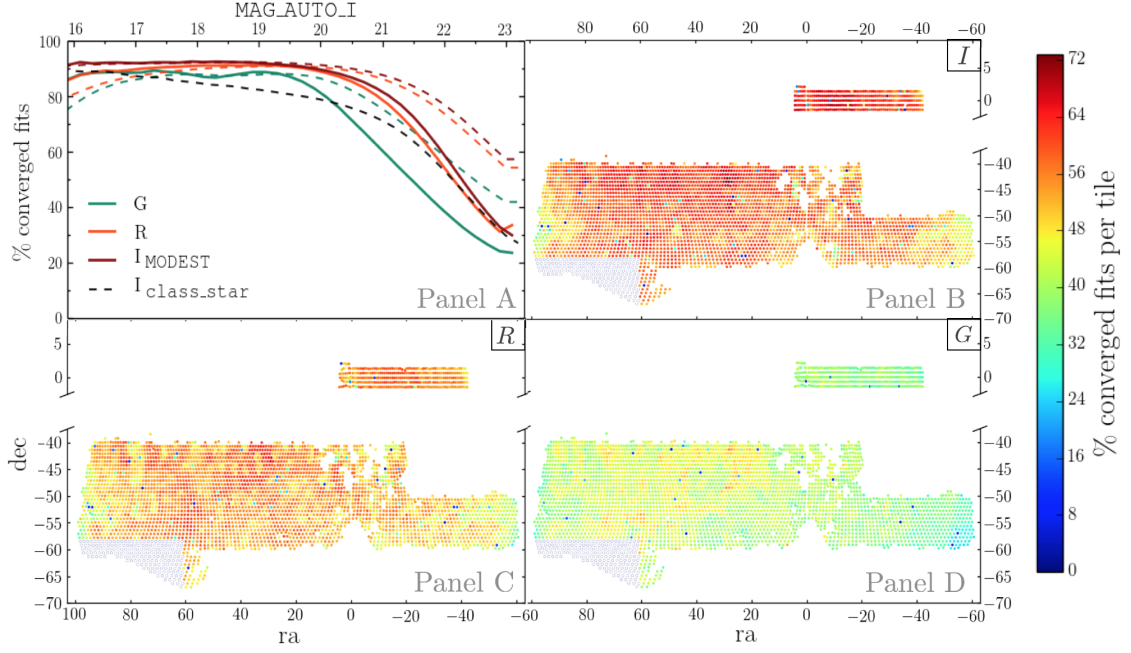


Figure 2.2: **Panel A:** fitting completeness in g , r and i bands (green, orange and brown lines, respectively), following star-galaxy separation using the MODEST classifier (see Section 2.4.2). The completeness, defined in eq. 2.3, is expressed in terms of the percentage of converged fits calculated in bins of 0.2 magnitude. Solid lines show the completeness in differential magnitude bins, while the dashed lines show results for magnitude-limited samples. The dashed black line shows the trend for the i band when using only a conservative S/G cut ($\text{CLASS_STAR} > 0.9$). Using the MODEST classifier we find that the completeness is 90% up to magnitude 21. **Panels B, C, D:** maps of the percentage of converged fits in g , r and i band in each tile (at $\text{mag_auto_i} < 23$). The region in the lower left corner occupied by empty grey circles is entirely flagged as unsuited for extra-galactic work due to its vicinity to the Large Magellanic Cloud (LMC). The regions with a lower fraction of converged fits are found towards the Galactic Plane and close to the LMC. In the g band the percentage of converged fits is poorer, as expected, due to an overall broader PSF.

the likely causes of failures, encompassing stellar contamination, the effect of PSF width, poor signal-to-noise and the effects of neighbouring sources.

STELLAR CONTAMINATION

We used the neural network star-galaxy (S/G) classifier, included as part of SExtractor, for a conservative initial criterion of star-galaxy separation. We apply the cut $\text{CLASS_STAR} < 0.9$, in order to remove only the most obvious stars, and to allow a user to perform their own S/G separation. Point sources will most likely fail to achieve a converged solution in GALFIT and we therefore expect that a substantial fraction of the incompleteness at bright magnitudes seen in the black dotted line in Fig. 2.2 (panel A) is due to contamination by stellar sources. This expectation is supported by the fact that the regions with the lowest percentage of converged fits (Fig. 2.2, panels B-D) are located in regions of known high stellar density. Further, in the upper panel of Fig. 2.3 it can be seen that the converged fraction at $i < 21.5$ depends strongly on the stellar density for the CLASS_STAR S/G separation.

In [Drlica-Wagner, Sevilla-Noarbe, Rykoff, Gruendl, Yanny, Tucker, Hoyle, Carnero Rosell, et al., 2017](#) it is shown that a simple cut in the SExtractor parameters SPREAD_MODEL and SPREADERR_MODEL can achieve a galaxy completeness of $\geq 98\%$, with $\leq 3\%$ stellar contamination at $i < 22$. This cut is known as MODEST classifier. SPREAD_MODEL is a morphological quantity which compares the source to both the local PSF and a PSF-convolved exponential model ([Desai et al., 2012](#); [Soumagnac et al., 2015](#)). In order to optimise the separation of point-like and spatially extended sources, we use the i band as the reference band for object classification due to the depth and superior PSF in this filter. The separation is defined via a linear combination of the SPREAD_MODEL and its uncertainty, the SPREADERR_MODEL:

$$\text{SPREAD_MODEL} + n \times \text{SPREADERR_MODEL} > thr, \quad (2.4)$$

where the coefficients $n = 1.67$ and $thr = 0.005$ are chosen as the optimal compromise between the completeness and purity of the galaxy sample. With the MODEST classifier we recover more than $\sim 90\%$ converged fits at magnitude 20 and $\sim 85\%$ at magnitude 21.5. We apply this additional S/G classification henceforth, and show the converged fraction of galaxies under this additional classification by the coloured lines in Fig. 2.2 and the black points in Fig. 2.3. The dependence of converged fraction on stellar density is vastly reduced with the SPREAD_MODEL classifier (though still present) with a threefold increase in stellar density, from 0.5 to 1.5 stars per sq. arcmin, causing just a 7% point drop in converged fraction. This decrease is almost entirely explained by the expected contamination rate of 3%.

2.4. Parametric Fits

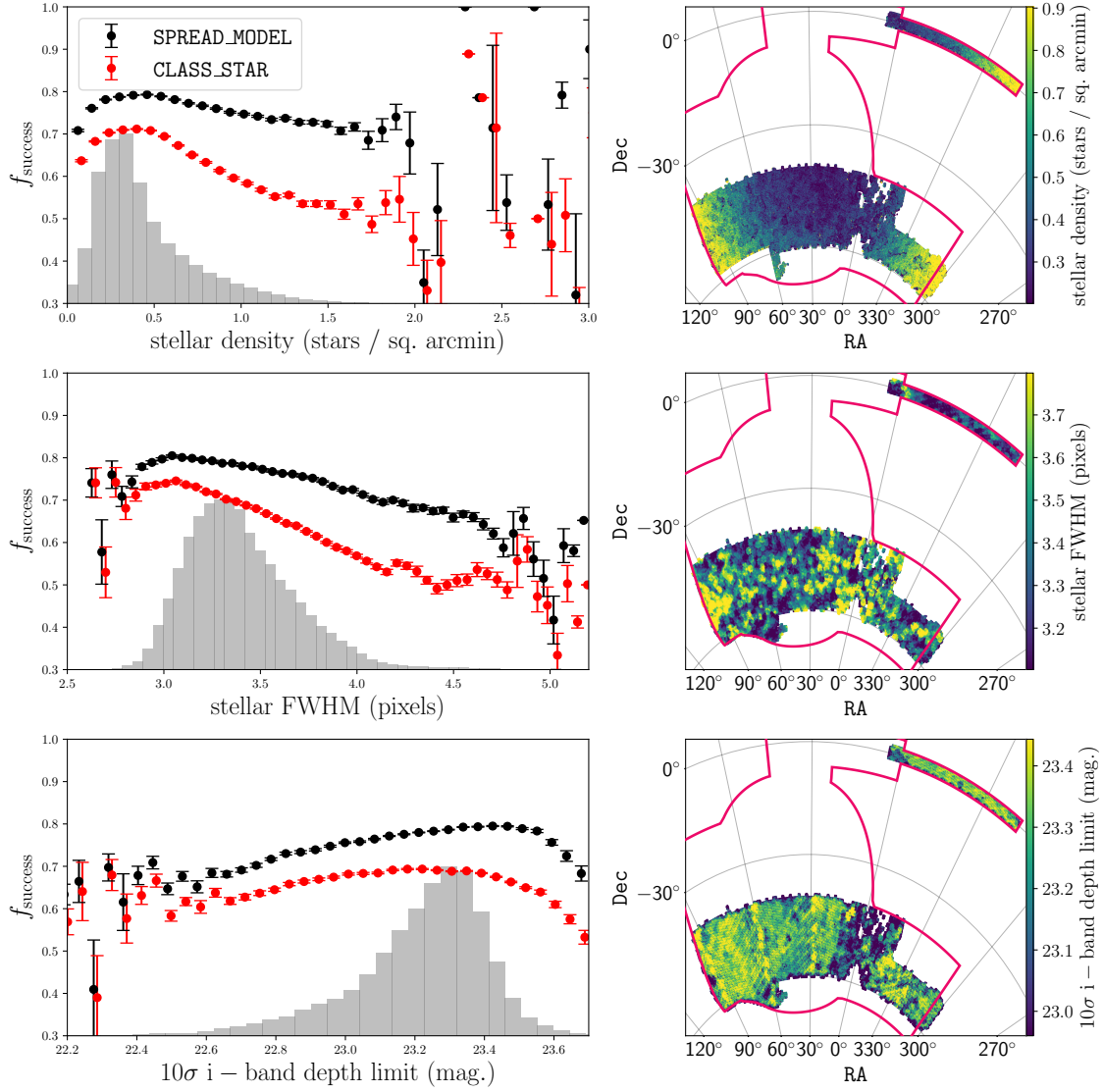


Figure 2.3: Dependence of fitting completeness at $i < 21.5$ on spatially-dependent survey characteristics, stellar density, PSF FWHM and i -band image depth (top, middle and bottom panels respectively). The maps of the nominal DES five-years footprint (outlined in magenta) show the dependences for the DES Y1 area. Grey histograms show the relative distributions of the characteristics in terms of survey area. The results for the galaxy sample are shown, following two star-galaxy classifiers: SExtractor CLASS_STAR (red points) and an additional criterion based on SPREAD_MODEL (black points, see text). Uncertainties are derived by bootstrap resampling. After the improved S/G separation, the fitting completeness is only weakly dependent on survey characteristics, and a high completeness ($> 80\%$) can be maintained with only minimal loss of area. The results at $i < 22$ are very similar in terms of the correlations with survey characteristics, but with overall lower converged fraction.

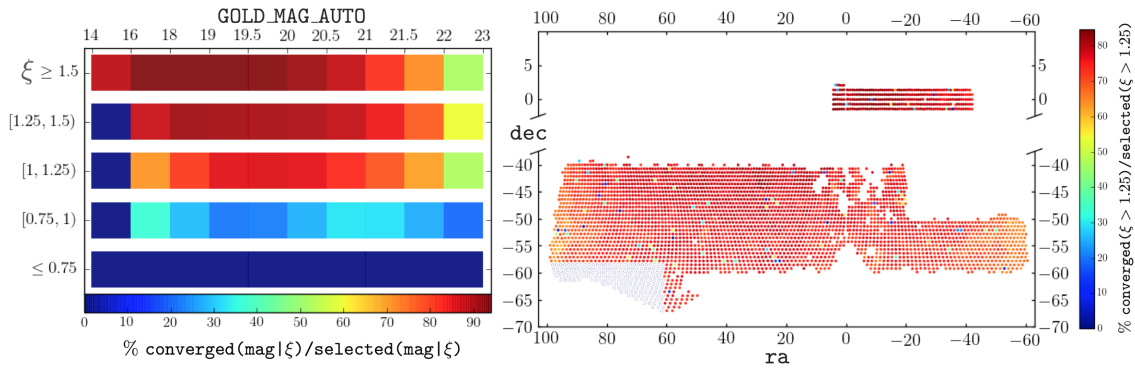


Figure 2.4: **Left panel:** fitting completeness calculated in differential bins of magnitude. The sample is divided into sub-populations, according to different ranges of the parameter $\xi = \text{FLUX_RADIUS}/\text{PSF_radius}$, as reported on the y-axis. Each population is represented by a bar, colour-coded by the percentage of converged fits in each magnitude bin. The figure shows that failed fits are more frequent for the objects with size smaller than the PSF or comparable with it. A critical drop occurs for the population with $\xi < 1.25$. **Right panel:** map of the percentage of converged fits per tile with $\xi > 1.25$. In comparison with the i band map in Fig. 2.2, it is clear that by applying this cut the overall percentage of successful fits increases dramatically, from $\sim 40\%$ to $> 70\%$ at the borders and up to $\sim 90\%$ in the central areas.

PSF WIDTH

In order to take into account the seeing, GALFIT convolves the 2-D model with the PSF, and compares it with the galaxy image. For this reason galaxy fitting requires very accurate knowledge of the PSF. Errors in the PSF model can easily result in attempted fits not converging, or in biased parameters (see section 2.4.4). Here, we assess the fitting incompleteness due to the varying PSF width across the DES survey area. We calculate the completeness for different sub-populations of the sample, delimited by certain values of the ratio between the galaxy half-light radius, estimated by the SExtractor FLUX_RADIUS, and the PSF size; we indicate this parameter with ξ , defined as follows:

$$\xi = \frac{\text{FLUX_RADIUS}}{\text{PSF_radius}}, \quad (2.5)$$

where we calculate the size of the PSF as the radius of the circular aperture enclosing half of the flux of the PSF itself. The typical PSF radius is $\sim 3 \text{ px}$. The left panel in Fig. 2.4 shows the completeness calculated in bins of 1 magnitude for five different populations: $\xi \leq 0.75$, $0.75 < \xi \leq 1$, $1 < \xi \leq 1.25$, $1.25 < \xi \leq 1.5$ and $\xi \geq 1.5$. Values of $\xi < 1$ are unphysical, indicating either noisy photometry, image artefacts or inaccuracies in the PSF model. Each population is represented with a bar coloured by the percentage of converged fits, normalised by the total number of selected objects in each magnitude bin. As expected, we observe lower percentages of converged fits for the objects whose

size is comparable to the size of the PSF used by GALFIT to deconvolve their images. Nevertheless, in the range $1 < \xi \leq 1.25$ the completeness is only around 10% lower than at larger sizes. The right panel in Fig. 2.4 maps the completeness per tile, excluding the galaxy sample whose size is comparable or smaller than the PSF ($\xi < 1.25$). Compared with the i band map in Fig. 2.2, it shows that by applying the cut in ξ the fitting completeness increases dramatically both at the borders (up to $> 70\%$) and in the central areas (up to $\sim 90\%$), and the discrepancy between these two regions is reduced.

In Fig. 2.3, centre panel, we show the dependence of fitting completeness against PSF FWHM ($i < 21.5$). For the SPREAD_MODEL S/G classifier we see that the completeness at $i < 21.5$ only drops below 80% in the extended tail of the distribution of PSF FWHM (grey histogram).

IMAGE DEPTH

There is a clear and expected dependence of the percentage of converged fits on magnitude in both Fig. 2.2 and Fig. 2.4. Although stars are less easily excluded at faint magnitudes and the sizes of galaxies are smaller, much of this dependence is likely to be due simply to the difficulty of GALFIT finding a stable minimum in the χ^2 space at low S/N. In the lower panel of Fig. 2.3 we show how the fitting success rate for $i < 21.5$ galaxies depends on image depth, and hence object S/N. As expected, the completeness falls in shallower regions of the footprint, but the decline is not dramatic for this bright subset and, once again, a high success rate can be maintained by removing only regions corresponding to the tails of the distribution.

IMPACT OF NEIGHBOURING SOURCES

Finally, we assess the impact of neighbouring sources on the fitting success rate. We reduce the complexity of possible arrangements of neighbours to two metric values: the amount of overlapping area⁵ between a galaxy and its neighbours, and the difference in magnitude between the galaxy and its most overlapping neighbour ($(\text{MAG_AUTO_|C}) - (\text{MAG_AUTO_|MON})$). The dependence of the converged percentage as a function of these two quantities is shown in Fig. 2.5, in four intervals of S/N for the target object. Each line in the figure is normalized by the population of objects with attempted fits within the same delta-magnitude range. We observe that even at low S/N the fitting success rate is high if all the neighbours present are sufficiently faint. However, in the range $0 < S/N < 25$ the completeness is a steep function of the magnitude difference between target galaxy and its neighbour. At high S/N neither the degree of overlap nor the relative magnitude of a neighbour are important. Note that, our initial selection removes objects that SExtractor determined to have been blended.

⁵By area, we mean the SExtractor-derived Kron ellipse enlarged by a factor of 1.5

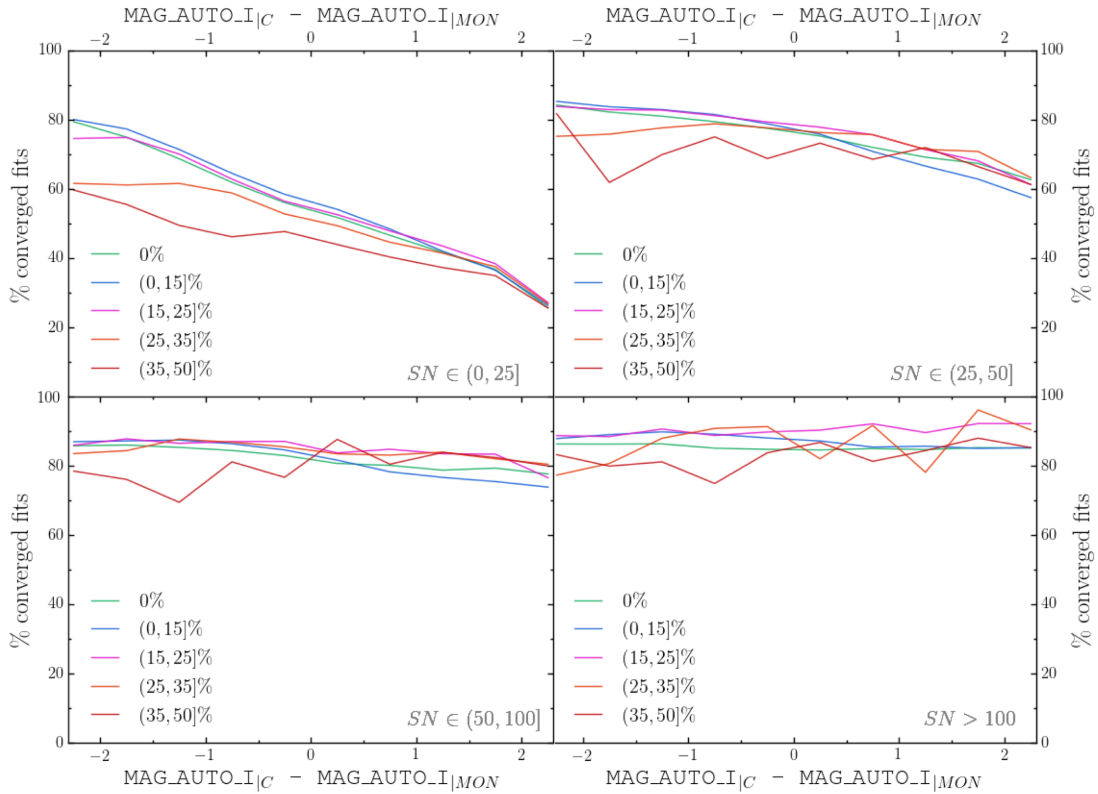


Figure 2.5: Fitting completeness as a function of the magnitude difference between the target galaxy and its closest neighbour. The relation is shown for different percentages of overlap between the two fitted objects, as reported in the legend. Each line is normalized by the population of objects with attempted fits within the same range in magnitude difference. The analysis is repeated in four signal-to-noise intervals. We observe that the fitting completeness decreases when the most overlapping neighbour is much brighter than the central galaxy, with stronger effects in low signal-to-noise regimes. This effect becomes negligible with increasing signal-to-noise.

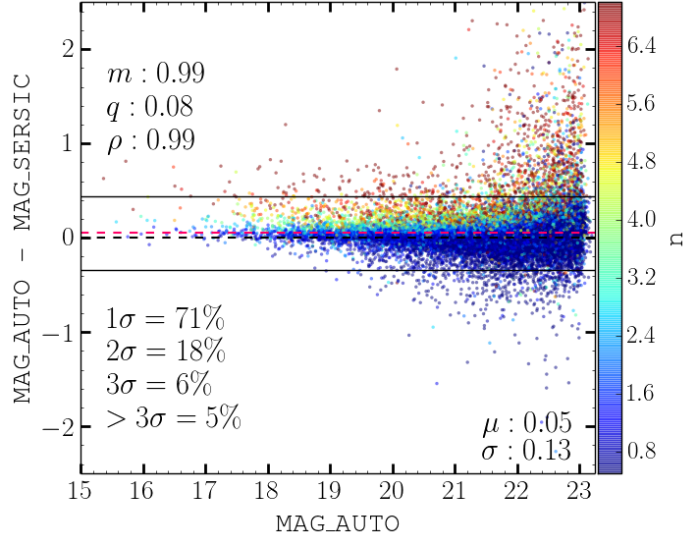


Figure 2.6: Difference between the input magnitude (MAG_AUTO) from SExtractor and the output magnitude (MAG_SERSIC) recovered through Single Sérsic fits. Results are shown as a function of input magnitude and are colour-coded by Sérsic Index. The two solid black lines delimit the population lying within 3 standard deviations from the mean magnitude difference relation, indicated by the dashed red line. The mean and the spread of the relation, printed in the lower right corner of the Figure, are obtained through a 3σ clipping procedure. The banding in Sérsic index is expected (A. W. Graham and Driver, 2005) and the vast majority of outliers (which in total number 5% of the sample) are of low S/N objects.

MULTI-WAVELENGTH COMPLETENESS

As shown by the green and red curves in Panel A in Fig. 2.2, we can recover a relatively high percentage of converged fits for objects brighter than magnitude 21.5 for the *g* and *r* filters also. We notice that the *g* and *r* bands show a drop in the brightest magnitude range ($\text{GOLD_MAG_AUTO_i} \leq 15.5$). Upon inspection we find that the objects responsible are compact objects with size comparable to the PSF and with a MODEST classification which is close to the threshold of 0.005 in the *i*-band. In Panels C and D we can see the spatial completeness for the *r* and *g* band, respectively. In both cases we reconfirm what we observed for the *i* band: a poor fitting completeness at the borders of the field, where stellar density is high, as discussed in the previous sub-sections. The *g* band PSF is typically broader than the *r* and the *i* bands, and the images shallower, which are reflected in an overall poorer recovery of converged fits.

2.4.3 VALIDATION

We now turn to assessing the accuracy of the parameters recovered from those objects that were successfully fit with GALFIT, beginning with simple magnitude and size diag-

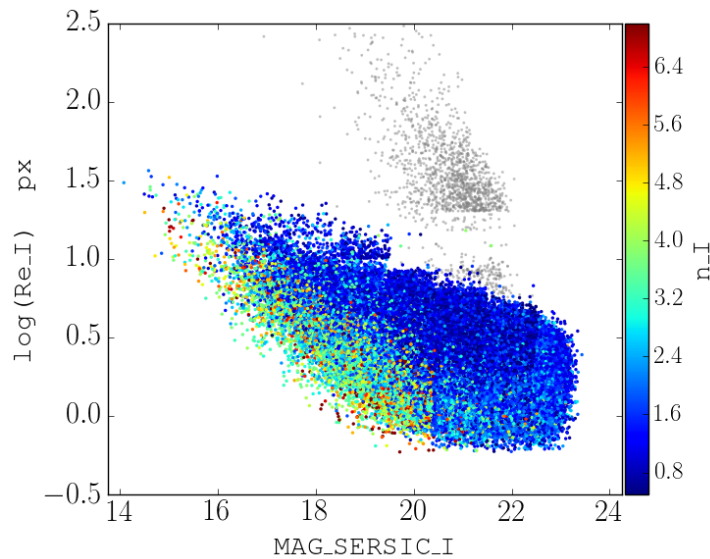


Figure 2.7: Relation between Sérsic magnitude and effective radius for the i band results. Points are colour-coded by Sérsic Index. Outliers are shown in grey.

agnostics of the population. We then investigate whether there are systematic errors from which GALFIT suffers in recovering the structural parameters of the galaxies, depending on their magnitude, size, concentration and shape. We investigate this aspect through image simulations (section 2.2.2) and present the relative calibrations in the next subsection. For this discussion we show the tests performed on the i band, which represents our fiducial filter, starting with a comparison of the total Sérsic magnitude with MAG_AUTO computed by SExtractor. In Fig. 2.6 we show this comparison for 30,000 randomly-selected objects from the full catalogue. We recover the expected behaviour: objects with Sérsic index ~ 1 have magnitudes consistent with MAG_AUTO, while the Sérsic magnitude is brighter at higher n . MAG_AUTO is known to be biased faint for high-Sérsic n objects, losing as much as 50% of the flux in extreme cases (A. W. Graham and Driver, 2005).

The solid black lines in Fig. 2.6 delimit the 3σ outliers in magnitude difference, following an iterative 3-sigma-clipping procedure to find the mean relation and spread (given by the parameters, μ and σ in the figure). The mean relation (red dashed line) is essentially flat in magnitude, suggesting that typically the background computed during catalogue extraction and that estimated by GALFIT are consistent. At faint magnitudes, however, there is a population of outliers with magnitude differences that cannot be explained by simple photometric errors, and that also exhibit very high Sérsic indices. We deem these unreliable fits, possibly caused by an unidentified elevated background. Restricting the sample to objects with $S/N > 30$ removes these objects and entirely removes the group with spurious large radii. We then obtain the relation between mag-

nitude and effective radius from the Sérsic profile fits as shown in Fig. 2.7. Points are colour coded by each object’s Sérsic index. Once again, the data match expectations and similar trends reported in the literature, with high Sérsic n objects forming a steep sequence and galaxies with exponential light profiles dominating at fainter magnitudes. Grey points are sources labelled as outliers during the validation process.

2.4.4 CALIBRATIONS

In this section we illustrate how we calibrate our measurements. As explained in detail in Section 2.2.2, we processed and fit the UFIG-BCC simulated data for DES Y1 in the same way we did for our real galaxy sample. We used ~ 10 million simulated objects. Now we can compare the results from the fits with the true morphological parameters used to generate the UFIG-BCC images. We then calculate the discrepancies between the measured and true parameters and derive appropriate corrections. We show the size of these corrections via a set of calibration maps.

DERIVATION OF THE CORRECTIONS

We derive corrections in a 4-dimensional parameter space, including size, magnitude, Sérsic Index and ellipticity. The ensemble of values assumed by each parameter constitutes a vector in the parameter space. We sample each vector with a list of nodes: the magnitude (mag) in the range [14.5,23.5] in steps of 1 magnitude, the size (r) in the interval [0.5,16.5] px in steps of 2 px, the Sérsic Index (n) in the set [0.2, 2, 4, 10] and the ellipticity (ϵ) in the intervals [0, 0.3, 0.6, 1]. The realization of each combination of these nodes forms an hypervolume which we’ll refer to as a *cell*. In each cell falls a certain number of simulated objects with similar structural properties and the corresponding fitting results: so each parameter is represented by a distribution of simulated values and a distribution of measurements. Each distribution in turn has a median value (m^i) and a standard deviation (σ^i), where $i = mag, r, n, \epsilon$, which represent the central value and the dispersion of the population, respectively. To summarise, in each cell the i -th parameter can be expressed as:

$$\hat{i} = \hat{\mu}^i \pm \hat{\sigma}^i \quad (2.6)$$

for the model and as:

$$i = \mu^i \pm \sigma^i, \quad (2.7)$$

for the fit, where $i(\hat{i}) = mag, r, n, \epsilon$. For all the objects falling in a given cell we calculate the correction (η^i) in each parameter as the discrepancy between the central values of the distributions:

$$\eta^i = \hat{\mu}^i - \mu^i. \quad (2.8)$$

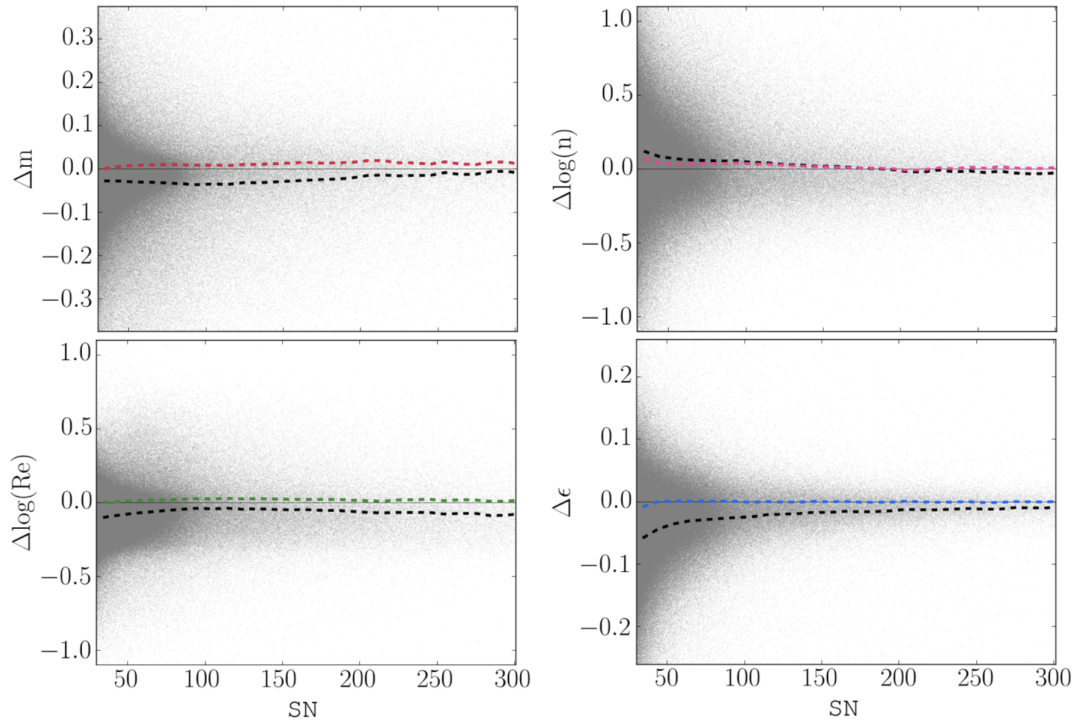


Figure 2.8: Discrepancies in recovered Sérsic parameters from running GALFIT on the UFIG-BCC image simulations, as a function of signal to noise (S/N). From top to bottom the panels display the results for magnitude, half light radius, Sérsic index and ellipticity. The dashed lines show the discrepancy in bins of S/N, calculated before (black line) and after (coloured line) applying calibration corrections (see section 2.4.4). The uncertainties depend to first order on the signal to noise, and the mean deviation is clearly reduced by applying the calibrations. In the calibration map, shown in figure 2.9, we investigate how the parameters and their uncertainties correlate with each other.

We further define a quantity, w , which represents the dispersion of the cell in the 4D parameter space, derived as the quadratic sum of the variances of the model parameters which determine the diagonal of the covariance matrix of the parameter space. It is defined as follows:

$$w = \sqrt{\sum_i \frac{\hat{\sigma}_i^2}{\hat{m}_i^2}}, \quad (2.9)$$

where $i = mag, r, n, \epsilon$ and $\hat{\sigma}_i^2$ and \hat{m}_i are the variance and median values of the model distributions, respectively. For cells with larger dispersion, we expect the correction vector to be less accurate for a given randomly chosen object.

CALIBRATION MAPS

In the validation routine we observed that $\sim 99\%$ of converged fits are well recovered in magnitude (η^{mag} of the order of 0.001), and that cutting objects with $S/N < 30$ we

remove the clear outliers in size and magnitude. In Figure 2.8 we show the discrepancies η^i between the intrinsic values and the parametric measurements as a function of signal to noise for magnitude, half-light radius, ellipticity and Sérsic index. The discrepancies relative to size and Sérsic index are shown in logarithmic space to facilitate visualization. In each panel the dashed lines show the discrepancies in bins of signal to noise. We use the uncalibrated sample to calculate the black line, and the same sample after applying the calibrations for the coloured one. It is clear that the uncertainties on the structural parameters increase in low signal to noise regimes, as one might anticipate, and the scatter clearly reduces when applying the corrections. We observe that GALFIT tends to recover larger sizes and ellipticities, so we pay particular attention to the corrections required for these properties within the multidimensional parameter space.

Figure 2.9 represents a map of the calibrations that we apply to our measurements, derived from our state-of-the-art image simulations. In using this multidimensional calibration map we are able to account for the correlations between parameters and ensure the corrections are appropriate for a true galaxy sample. The arrows represent the strength of the vector corrections, expressed as the distance between the central values of the size and magnitude distributions of the model sample and the relative measured dataset in each cell. The components of the correction vectors are the magnitude discrepancy η^{mag} on the x axis and the size discrepancy η^{size} on the y axis, according to the definitions given in Equations 2.6 and 2.7. If these corrections are small ($\eta^{mag} < 0.1 \wedge \eta^{size} < 10\%$) the length of the arrow is set to zero and only a circle is shown. Apart from the grey circles, which indicate areas with poor statistics, different colours are used to give an indication of the correction applied to ellipticity and Sérsic Index. If $\eta^\epsilon > 0.1$ or $\eta^n > 20\%$, the symbol is coloured in orange and red, respectively. If the correction is large in both cases, then it is coloured in brown. The symbol is empty if the GALFIT recovered value is smaller than the model. The symbols are shaped according to the total scatter (w) in the 4D parameter space of the model parameters, defined in Equation 2.9; we use a pentagon if $w > 1.5$ and a square if $w > 1$, otherwise the symbol is a circle. Figure 2.9 reports the vector corrections for the i band; corrections for the g and r filters are shown in Section A.1 of Appendix A.

We observe that the strength of the corrections and their positions are compatible with the findings we discussed previously in the validation section. In that section we noted that in any range of shape and Sérsic index the uncalibrated measurements of the sub-populations of galaxies at the faintest magnitude range present overestimated half light radii and Sérsic Indices. In the calibration map they are assigned with larger vector corrections in size, which calibrate the measurements towards smaller values. If the correction in size is small, then we observe that a calibration in Sérsic Index is applied, where the recovered value was larger than the model parameter. The same observations are valid also for the other two filters (shown in Appendix A). The fact that

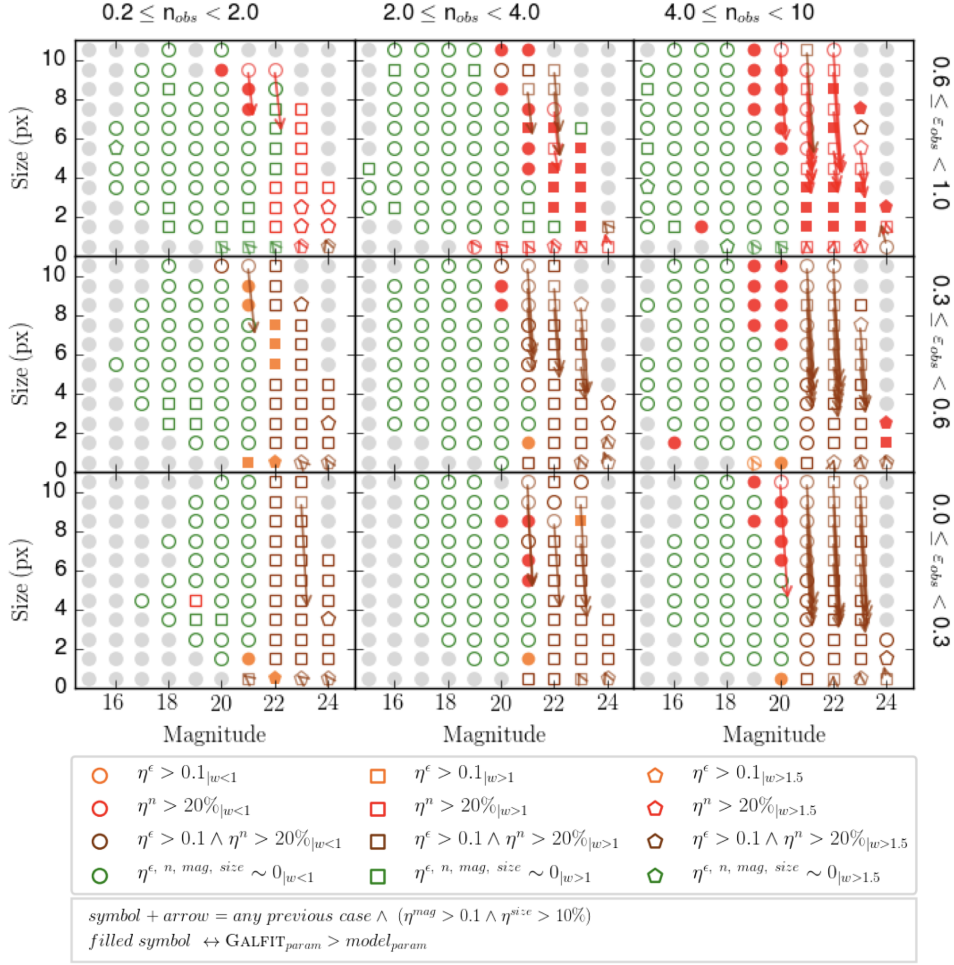


Figure 2.9: Calibration map for the parametric measurements in the i band, obtained from image simulations as described in Section 2.2.2. The calibrations are determined in a 4D parameter space, where the correlation of size, magnitude, ellipticity and Sérsic Index between the simulated galaxy and the model is studied. The information is provided using different marker shapes (circles, squares, pentagons, arrows) and colours, as follows. The calibrations are presented in a size-magnitude plane, divided in different cells according to the shown sub-ranges in ellipticity and Sérsic Index. The components of the correction vectors are the magnitude discrepancy η^{mag} and the size discrepancy η^{size} , according to the definitions given in Equations 2.6 and 2.7. If these corrections are small ($\eta^{mag} < 0.1 \wedge \eta^{size} < 10\%$) the length of the arrow is set to zero and the cell is identified by a symbol only. Points and arrows are coloured according to the scatter in ellipticity (ϵ) and Sérsic Index (n); a scatter in $\eta^\epsilon > 0.1$ or $\eta^n > 20\%$ is expressed in orange and red, respectively, while the cells presenting a large scatter in both parameters are coloured in brown. The symbol is empty if the GALFIT recovered value is smaller than the model. Different shapes are used referring to the total scatter (w) in the 4D parameter space of the model parameters, defined in Equation 2.9; the symbol is a pentagon if $w > 1.5$ and a square if $w > 1$, otherwise it is a circle. The symbols and conventions used in the calibration map are summarised in the legend. In the case of the calibration of non-parametric fits (following in Section 2.5.4) the Sérsic Index is replaced with the concentration parameter.

the measurements and their associated corrections are similar across photometric bands indicates that our final set of calibrated results are robust to the survey characteristics, such as overall PSF size and noise level, that vary between bands. Furthermore, the vast majority of cells across all three calibration maps show little corrections, suggesting that our converged fits are in general reliable and represent the light profiles well.

2.5 NON PARAMETRIC FITS

2.5.1 ZEST+ SETUP

ZEST+ is a C++ software application which uses a non-parametric approach to quantify galaxy structure and perform morphological classification. It is based on the ZEST algorithm by [Scarlata, Carollo, S. Lilly, et al. 2007](#); [Scarlata, Carollo, S. J. Lilly, et al. 2007](#), which saw a first application in [Cameron et al. 2010](#). Compared with its predecessor, ZEST+ has increased execution speed. The software architecture consists of two main modules: *Preprocessing* and *Characterization*. The former performs image cleaning, main object centring and segmentation, the latter calculates structure and substructure morphological coefficients.

PREPROCESSING

In this module the algorithm uses the stamps and the input catalogue provided by the SAND routine. The input catalogue includes the coordinates and the geometrical parameters of the target galaxy and its neighbours in order to remove nearby objects, subtract the background, determine the centre of the galaxy and measure its Petrosian radius. The Petrosian radius is defined as the location where the ratio of flux intensity at that radius, $I(R)$, to the mean intensity within the radius, $\langle I(< R) \rangle$, reaches some value, denoted by $\eta(R)$ ([Petrosian, 1976](#)):

$$\eta = \frac{\mathcal{I}(\mathcal{R})}{\langle \mathcal{I}(\mathcal{R}) \rangle}. \quad (2.10)$$

For this work the Petrosian radius corresponds to the location where $\eta(R) = 0.2$. The Petrosian ellipse associated with the object contains the pixels which are used in the *Characterization* module to calculate the morphological coefficients of the central galaxy.

CHARACTERIZATION

The measurements provided by ZEST+ are galaxy concentration (C), asymmetry (A), clumpiness or smoothness (S) and Gini (G) and M_{20} coefficients. This set of parameters, which we refer to as to the *CASGM system*, quantifies the galaxy light distribution and

is widely used in studies which correlate the galaxy structure to other parameters, such as colour and peculiar features indicating mergers or galaxy interactions (see for example [Conselice, Bershady, and Jangren 2000](#), [Conselice 2003](#), [Lotz, Primack, and Madau 2004](#) and [Zamojski et al. 2007](#)); other similar quantities have been recently introduced by [Freeman et al. \(2013\)](#).

The *concentration* of light, first introduced in [Bershady, Jangren, and Conselice 2000](#) and [Conselice 2003](#), expresses how much light is in the centre of a galaxy as opposed to its outer parts; it is defined as

$$C = 5 \log \left(\frac{r_{80}}{r_{20}} \right), \quad (2.11)$$

where r_{80} and r_{20} are the elliptical radii enclosing, respectively, the 20% and 80% of the flux contained within the Petrosian ellipse of the object. ZEST+ outputs three different values of concentration, C , C_{ext} and C_{circ} . The first parameter is calculated using the total flux measured within the Petrosian ellipse, the second using the flux given as input by the user within the same ellipse and the third one using the Petrosian flux within a circular aperture. For this work we refer to C as the concentration.

The *asymmetry* is an indicator of what fraction of the light in a galaxy is in non-symmetric components. Introduced in [Schade et al. 1995](#) first, and then in [R. G. Abraham, van den Bergh, Glazebrook, et al. 1996](#) and [Conselice 1997](#) independently, asymmetry is determined by rotating individual galaxy images by 180° about their centres and self-subtracting these from the original galaxy images. This procedure is applied after the *Preprocessing* module, where the background is $\kappa\sigma$ -clipped and subtracted. The value of pixel (i, j) in the subtracted image is calculated as:

$$\Delta I(i, j) = I(i, j) - I_{180}(i, j) = I(i, j) - I(2i_c - i, 2j_c - j), \quad (2.12)$$

where I_{180} is the rotated image and (i_c, j_c) are coordinates of the centre of the galaxy. To take into account the asymmetry of the background, ZEST+ works with smoothed images of the galaxies and their rotated version. In this method, proposed in [Zamojski et al. 2007](#), the smoothed image is obtained through a five-point convolution:

$$f_{i,j}^S = \frac{1}{5}(f_{i,j} + f_{i+1,j} + f_{i-1,j} + f_{i,j+1} + f_{i,j-1}), \quad (2.13)$$

where $f_{i,j}$ is the flux at the (i, j) pixel of the image, and $f_{i,j}^S$ is the flux in the same coordinates after the smoothing. The asymmetry of the original image is defined as

$$A_0 = \frac{1}{2} \frac{\sum_{i,j} |I(i, j) - I_{180}(i, j)|}{\sum_{i,j} |I(i, j)|}, \quad (2.14)$$

where $I(i, j)$ and $I_{180}(i, j)$ express the intensity of the flux at the pixel (i, j) in the original and rotated image, respectively. Similarly we define the asymmetry of the smoothed image:

$$A_{0,S} = \frac{1}{2} \frac{\sum_{i,j} |I^S(i, j) - I_{180}^S(i, j)|}{\sum_{i,j} |I^S(i, j)|}. \quad (2.15)$$

Assuming that the intrinsic asymmetry of the light does not change in the smoothed version, we consider that the difference between the two values of asymmetry is due to the background. Smoothing reduces the standard deviation of the background by a factor $\sqrt{5}$ with respect to its un-smoothed version. The combination of A_0 and $A_{0,S}$ then gives the final asymmetry value:

$$A = A_0 - \frac{A_0 - A_{0,S}}{1 - 1/\sqrt{5}}, \quad (2.16)$$

where the subtracted term corresponds to the background correction factor.

The *clumpiness* or *smoothness* parameter, introduced in [Conselice 2003](#), describes the fraction of light which is contained in clumpy distributions. Clumpy galaxies show a large amount of light at high spatial frequencies, and smooth systems at low frequencies. This parameter is therefore useful to catch patches in the galaxy light which reveal star-forming regions and other fine structure. ZEST+ calculates the clumpiness by subtracting a smoothed image, $I_S(i, j)$, from the original, $I(i, j)$, and then quantifying the residual image, $I_\Delta(i, j)$. The smoothed image is obtained by convolving the original image with a Gaussian filter of FWHM equal to 0.25 times the Petrosian radius calculated during the *Preprocessing* module. In $I_\Delta(i, j)$ the clumpy regions are quantified from the pixels with intensity higher than $k = 2.5$ times the background standard deviation in the residual image σ_Δ . These pixels are then used to calculate the clumpiness of the galaxy:

$$S = \frac{\sum_{i,j} I_\Delta(i, j)}{\sum_{i,j} |I(i, j)|_{I_\Delta(i,j) > k\sigma_\Delta}}. \quad (2.17)$$

Similarly, the *Gini* coefficient quantifies how uniformly the flux of an object is distributed among its pixels. A Gini coefficient $G = 1$ indicates that all the light is in one pixel, while $G = 0$ means that every pixel has an equal share. To calculate *Gini* ZEST+ uses the definition by [Lotz, Primack, and Madau \(2004\)](#), [Lotz, Jonsson, et al. \(2008\)](#), and [Lotz, Davis, et al. \(2008\)](#):

$$G = \frac{1}{\hat{I}n(n-1)} \sum_i^n (2i - n - 1) \hat{I}_i, \quad (2.18)$$

where \hat{I} is the mean flux of the galaxy pixels and \hat{I}_i , indicates the flux in the i th pixel, sorted by increasing order.

The M_{20} coefficient is similar to the concentration C in that its value indicates the degree to which light is concentrated in an image; however a high light concentration (denoted by a very negative value of M_{20}) doesn't imply a central light concentration. For this reason it is useful in describing the spatial distribution of bright substructures within the galaxy, such as spiral arms, bars or bright nuclei. The computation of this parameter requires first that the pixels within the Petrosian ellipse of the galaxy are ordered by flux; then the 20% brightest pixels are selected and for each pixel i the second-order moments are calculated:

$$E_i = I_i[(x_i - x_c)^2 + (y_i - y_c)^2], \quad (2.19)$$

where I_i is the flux in the i -th pixel, (x_i, y_i) the coordinates of the pixel and (x_c, y_c) the coordinates of the centre of the Petrosian ellipse. The sum of these moments is $E = \sum_i^{N_{20}} E_i$, where N_{20} is the multiplicity of the 20% brightest selected pixels. Given E_{tot} as the sum of the second order moments of all the pixels in the ellipse, we finally calculate M_{20} as:

$$M_{20} = \log \frac{E}{E_{tot}}. \quad (2.20)$$

2.5.2 COMPLETENESS

The measurements of Gini, M20, Concentration, Asymmetry and Clumpiness are matched with diagnostic flags which inform the user whether errors occurred during the cleaning step of the process or in the calculation of the coefficients. To be more precise, the flag Error (we label it in our catalogue as ERRORFLAG) indicates whether a problem occurred while processing an object: if it is non-zero, it traces an error encountered during the calculation of the structural parameters, and flags the measurements as not reliable. The *contamination flag* informs the user whether the cleaning process was unsuccessful due to the presence of a neighbour covering the centre of the galaxy; in this case the program outputs *contamination flag* = -2. Therefore in this test we considered as converged fits the measurements with $\text{ERRORFLAG} = 0 \wedge \text{contamination flag} \neq -2$. Then we define the fitting completeness as we did for the parametric fits, following Equation 2.3.

The results for the g , r and i bands are shown in Figure 2.10. With the cut in ERRORFLAG and *contamination flag* we discard a total of $\sim 10\%$ of objects. We observe some fluctuations at the brightest end, where we find cases of large bright galaxies whose Petrosian ellipses were underestimated or cases with saturated objects, and at the faintest end,

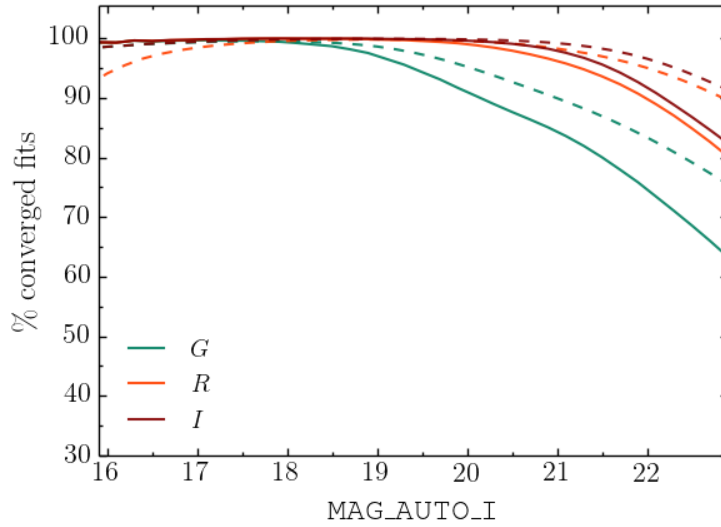


Figure 2.10: Fitting completeness of non-parametric converged fits in the g , r and i bands, expressed in terms of the percentage of converged fits in bins of 0.2 magnitude, normalised on the total number of selected objects in that magnitude bin. By *converged fits* we refer in this case to the objects flagged by ZEST+ as fits without errors, either during the cleaning process or the characterization routine, as described in more detail in Section 2.5.2. Magnitude-limited completeness is represented by the dashed lines. We obtain almost full recovery in the i and r filters up to $i \sim 21$, losing only a few saturated objects.

where it is more common to have higher noise contamination within the Petrosian ellipse. The overall number of successful fits is more than $\sim 90\%$ in the i and r filters and $\sim 80\%$ in the g band. The dashed lines show magnitude-limited, rather than differential, completeness.

2.5.3 VALIDATION

By way of a simple internal validation, we show in Figure 2.11 the uncalibrated measurements from ZEST+ and the relationships between them (only the Concentration parameter is calibrated). In particular we focus on the Gini-M20 relation, studied as a function of other morphological parameters: Concentration (C), Clumpiness (S) and Asymmetry (A), shown in Panels A, C and D, respectively. Since we can benefit from the additional information provided by parametric fitting, we show the same relation as a function of calibrated parametric quantities: Sérsic Index n (Panel B), ellipticity ϵ (Panel E) and $g - i$ colour (Panel F).

In the cross-comparison between non-parametric measurements, we observe that even though those are still un-calibrated, we can easily recover the expected trends with very few outliers. As an example consider the first panel, where the Gini-M20 relation

is colour-coded by the Concentration. The objects with low M20 values present a high concentration of light; from the figure we observe that in the Gini-M20 plane these objects tend to have larger values of Gini, which means that the light is not uniformly distributed. If we now consider the third parameter, we notice that the Concentration (and the Sérsic Indexes) of these objects lies in its highest range: this explains that the light of these galaxies is very concentrated, and located at the centre of the galaxy.

From panels C, D and E we add the expected information that these objects are also symmetric, lack clumpy regions and are mostly rounded. These observations were further confirmed by our visual inspection of image stamps. If the combined analysis of the first five panels helps us to distinguish between two different morphological regions in the Gini-M20 plane, Panel F shows a colour bi-modality which overlaps with the morphological one: disk-like galaxies tend to be bluer and the bulge-dominated ones are redder. Finally, we perform a qualitative comparison with the CAS-GM measurements made by [Zamojski et al., 2007](#) using high-resolution Hubble Space Telescope data (their Figures 3 and 17). The range of values for Gini and M20 are much the same for the bulk of the population, though our far larger sample explores more extreme values of low Gini coefficient and less negative M20. The correlation between M20 and asymmetry, at $M_{20} > -2$, is also clearly present in Figure 2.11, panel C. We expect the PSF to suppress fine substructure, and the trend between clumpiness and Gini coefficient in our sample is not as clear as that found by [Zamojski et al., 2007](#). Nevertheless, redder galaxies do tend to avoid regions of high clumpiness, as expected.

2.5.4 CALIBRATIONS AND DIAGNOSTICS OF THE CORRECTED RESULTS

In order to apply corrections to the non-parametric measurements, which are crucial in accounting for the impact of the PSF, we adopt the same approach used for the parametric fits: we consider the images from the *UFig-BCC* release for DES Y1 and treat them as if they were real data, as explained in detail in Section 2.2.2. We then derive calibration maps exactly as described in Section 2.4.4, determining the correction for each parameter of interest as the discrepancy between the central values of the model and the fitting results distributions in each cell. The equations 2.6, 2.7, 2.8 and 2.9 are valid also in this context, with the exception that the Sérsic Index, n , is now substituted by the Concentration of light, C .

In order to derive correction vectors, we first compute ZEST+ output parameters for the simulated galaxies *before* noise and PSF convolution are applied. We use GALFIT to produce noise and PSF-free image stamps based on the UFIG model parameters and run ZEST+ on them. In this way we construct the truth table of values with which to derive calibration vectors. Figure 2.12 shows the correction map for the i band; the other two filters, g and r , are presented in Appendix A. Also for non-parametric fits we adopt the

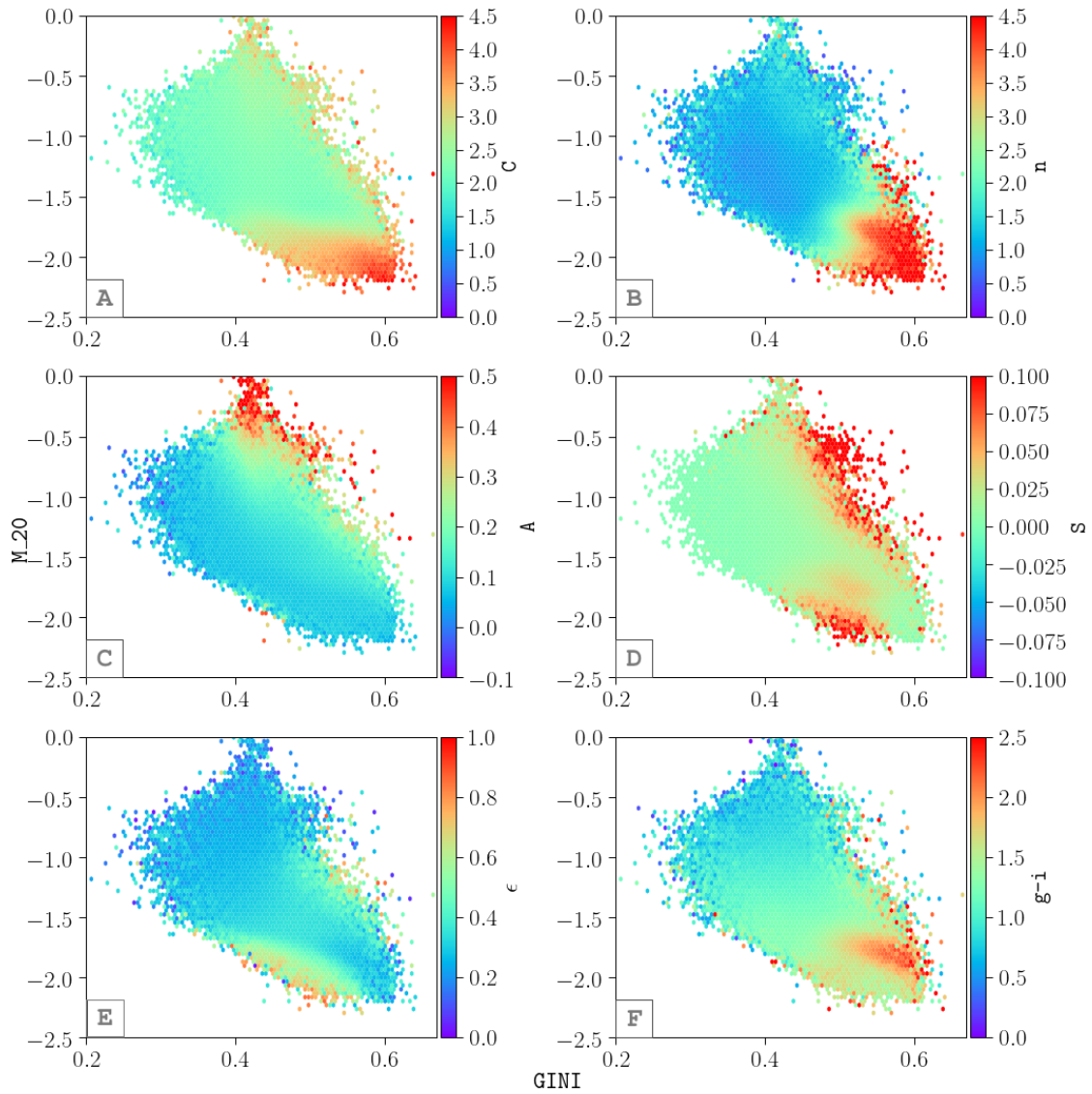


Figure 2.11: Gini-M20 relation shown as a function of Concentration C (Panel A), Sérsic Index (Panel B), Asymmetry A (Panel C), Clumpiness S (Panel D), ellipticity ϵ (Panel E) and $g - i$ colour (Panel F). The expected trends for the relations and their gradients are recovered, as discussed in more detail in Section 2.5.3.

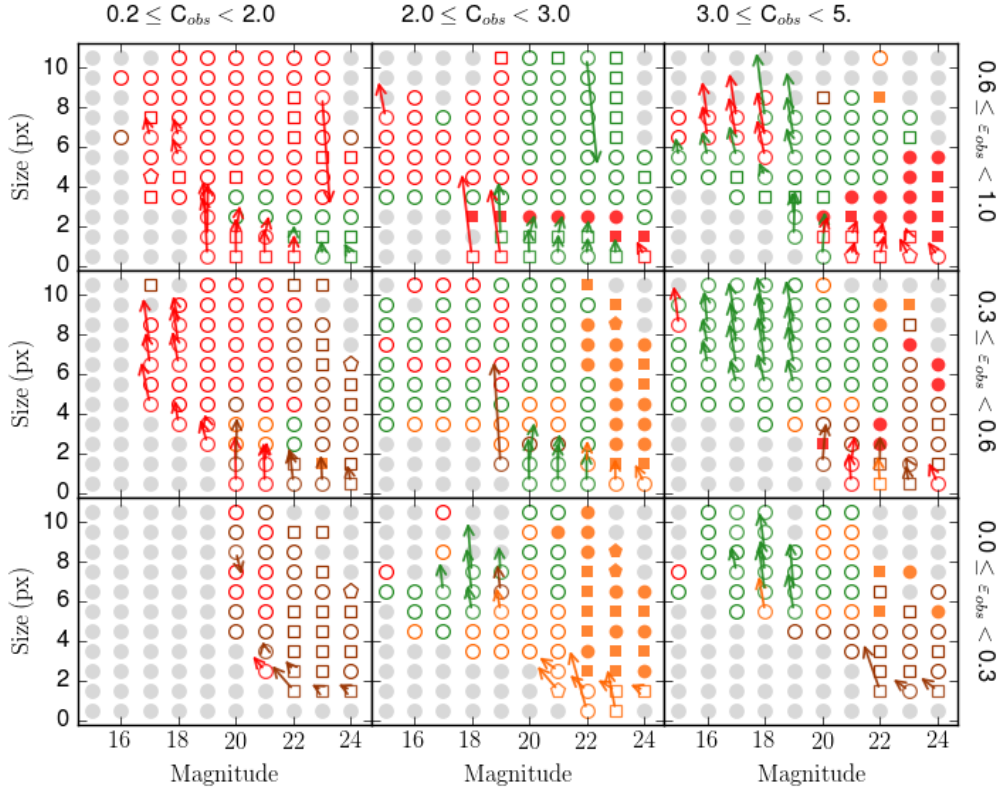


Figure 2.12: Calibration map for the non-parametric measurements in the i band, obtained through the simulation routine described in Section 2.5.4. The calibrations are determined in a 4D parameter space, where the correlation of size, magnitude, ellipticity and Concentration between the measured values and the model parameters is studied. The information in the map is displayed using different symbols and colours with the same GALFIT adopted for the parametric fits. They calibrations are presented in a size-magnitude plane, divided in different cells according to the shown sub-ranges in ellipticity and Concentration. The components of the correction vectors are the magnitude discrepancy η^{mag} on the x axis and the size discrepancy η^{size} on the y axis, according to the definitions given in Equations 2.6 and 2.7. If these corrections are small ($\eta^{mag} < 0.1 \wedge \eta^{size} < 10\%$) the length of the arrow is set to zero and only a symbol identifies them. If the scatter in ellipticity (ϵ) or Concentration (C) is large ($\eta^\epsilon > 0.1$ and $\eta^C > 20\%$, respectively), then the symbol is coloured in orange or red, respectively. If the calibration is large in both parameters, it is coloured in brown. The symbol is empty if the ZEST+ recovered value is smaller than the model. Different shapes are used referring to the total scatter (w) in the 4D parameter space of the model parameters, defined in Equation 2.9; the symbol is a pentagon if $w > 1.5$ and a square if $w > 1$, otherwise it is a circle.

same convention of colours and shapes as in Figure 2.9. The length of the arrows is a visual representation of the strength of the vector correction: their x and y components are the discrepancies between the central values of the model distribution and the fitted dataset in each 4-dimensional cell, projected on the size-magnitude plane. When the correction is small ($\eta^{mag} < 0.1 \wedge \eta^{size} < 10\%$) a symbol in place of the arrow is shown. Apart from the grey circles, which indicate areas with poor statistics, the colour legend reflects the size of the calibration applied to ellipticity and Concentration. If the scatter in ellipticity or Concentration is large ($\eta^\epsilon > 0.1$ or $\eta^C > 20\%$), then the symbol is coloured in orange or red, respectively. If this condition applies to both parameters simultaneously, it is coloured in brown. If the recovered value underestimates the model input, the symbol is empty. Different shapes are used according to the dispersion w of the 4-dimensional parameter space, calculated considering its covariance matrix, as expressed in Equation 2.9. Symbols are pentagons when $w > 1.5$, squares if $w > 1$ and circles otherwise. We observe that the majority of red cells, where a larger correction in Concentration is required, have an empty symbol: this tells us that ZEST+ tends to recover underestimated values of concentration. This behaviour is entirely expected, due to the fact that ZEST+ cannot account the PSF in computing results. We demonstrate this aspect more explicitly in Figure 2.13, which shows the relation between the Sérsic Index and the Concentration before (grey contours) and after (magenta) applying the corrections. For clarity, we have removed objects where the pixel size significantly hampers our ability to measure the concentration (i.e. where $FLUX_RADIUS < 2.5$ px). The solid blue line in this figure is the analytic relationship between Sérsic index and concentration, adapted from [A. W. Graham and Driver, 2005](#) for the case of measurements within the Petrosian radius. The *flattening* effect we observe in the uncalibrated population of Concentration values reflects exactly what we observe in the calibration map and through the corrections we obtain values that are much more consistent with expectations. This test shows that using calibrated values from both parametric and non-parametric approaches to quantifying galaxy structure allows us to use the advantages of both methods and provide a firmer grip on the characteristics of the galaxy population. We will exploit the strength of our dual-method, multi-band morphology catalogue in a series of future papers.

2.6 SCIENCE-READY CUTS

We finish by summarising the overall selection function of the galaxy sample and detail a set of simple cuts that could form the basis of a sample for scientific analysis. We exclude from consideration objects that meet any one of the following criteria:

- `SExtractor FLAGS > 0`

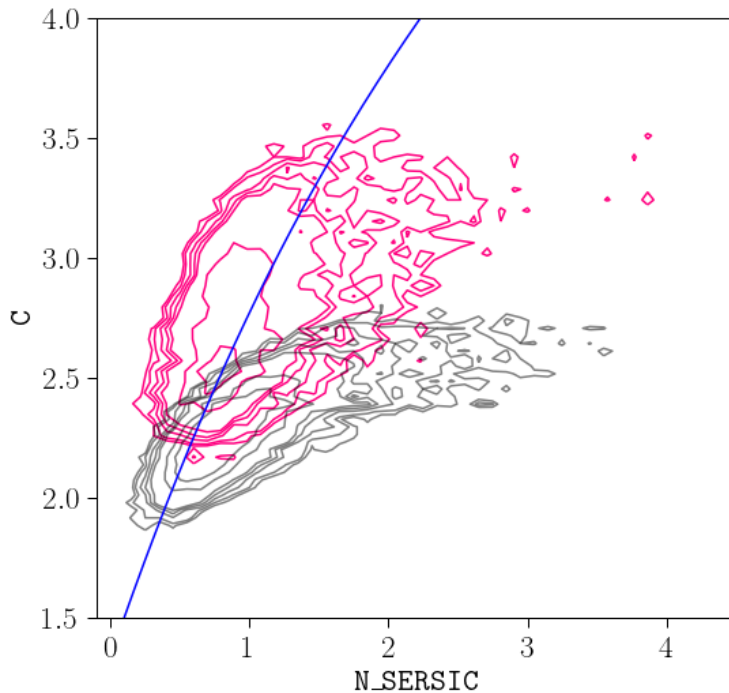


Figure 2.13: Sérsic Index-Concentration relation before (grey) and after (magenta) applying the calibrations. The solid blue line show the analytic relationship between Sérsic index and concentration. The *flattening effect* present in the un-calibrated measurements is due to PSF effects which is corrected by our calibration.

- $\text{CLASS_STAR} > 0.9$
- $\text{MAG_AUTO_I} > 23$
- $\text{FLUX_RADIUS} \leq 0$
- $\text{KRON_RADIUS} \leq 0$
- $\text{FLAGS_BADREGION} > 0$
- Objects with a neighbour that overlaps 50% or more of its expanded Kron ellipse. The relevant column in the catalogue for this criterion is `MAX_OVERLAP_PERC`.
- Objects that have unrecoverable errors in the SExtractor output of their neighbouring objects (if any).

This initial sample comprises 45 million objects over 1800 square degrees that is 80% complete in Sérsic measurements up to magnitude 21.5.

To prepare a high completeness science-ready galaxy sample, we suggest the following initial cuts. Science problems requiring higher completeness and/or greater uniformity

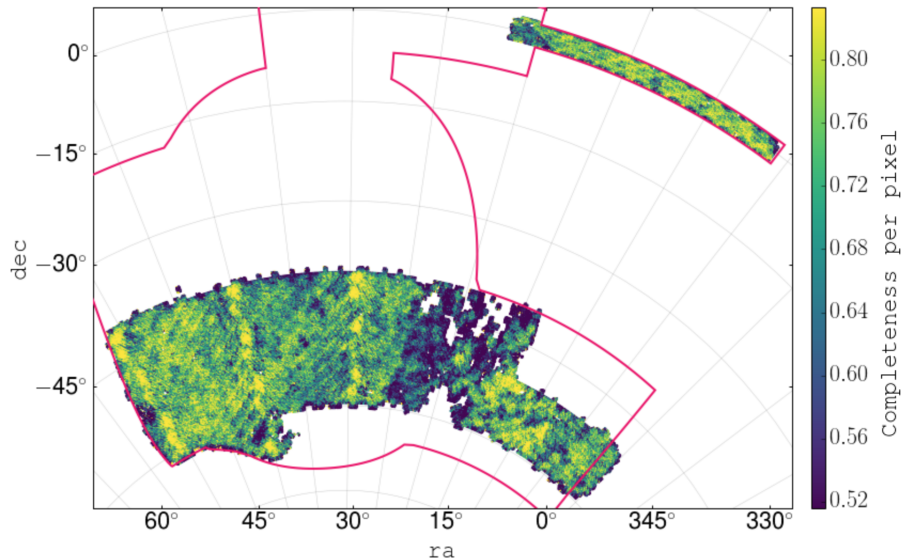


Figure 2.14: Healpix map of the ratio between two galaxy samples. We apply to the Y1A1 data the sample-selection cuts to obtain the first sample, and then apply the science-ready cuts to it in order to get the second one. The ratio gives the completeness per pixel of the science-ready sample.

across the footprint will require additional cuts, dependent on the goals. In some circumstances fainter galaxies could also be included in the sample.

- $\text{MAG_AUTO_I} \leq 21.5$
- $S/N > 30$
- $\text{SPREAD_MODEL} + 1.67 \times \text{SPREADERR_MODEL} > 0.005$

For the i-band catalogue, these cuts produce a sample of 12 million galaxies that is 90% complete in Sérsic measurements and 99% complete in non-parametric measurements. In Fig. 2.14 we show a ratio of two healpix maps realised with two samples. We first applied the cuts used for the sample selection, with an additional cut in $\text{MAG_AUTO} < 21.5$. We chose this threshold according to the analysis of the completeness discussed in Section 2.4.2. Then we select from this sample all the objects with pass the set of science-ready cuts we proposed above. The map shows the completeness per pixel, which is overall uniform. It also guides the catalogue users to possibly select specific areas for future analyses.

2.7 CONCLUSIONS

We have presented the process of preparing, producing and assembling the largest structural and morphological galaxy catalogue to date, comprising 45 million objects over

1800 square degrees, which are taken from the first year of the Dark Energy Survey observations (DES Y1). We adopted both parametric and non-parametric approaches, using GALFIT and ZEST+. In order to optimize their performance according to the characteristics of our sample, in particular in those cases where the galaxy we want to fit has one or more close neighbours, we developed a neighbour-classifier algorithm as part of a pre-fitting pipeline (Section 2.3.2) which automatically prepares the postage stamps and all the settings required to simultaneously fit the objects in the presence of overlapping isophotes. We stress the importance of this step because a precise treatment of the size of the stamps and the neighbouring objects allows the recovery of more accurate measurements.

In Section 2.4.2 we presented the fitting completeness of the parametric fits in the g , r and i filters as a function of object magnitude. Using a tile-by-tile analysis, we show that the highest percentages of non-converged fits are localised at the West and East borders of the footprint, where there is a high stellar density due to the vicinity of the Large Magellanic Cloud. After applying star-galaxy separation based on a linear combination of the parameter SPREAD_MODEL and its uncertainty, we find that the fitting efficiency remains high ($> 80\%$) up to magnitude < 22 for the i and r band, and magnitude < 21 for the g band. We also studied the subsequent fitting completeness in relation to survey data characteristics that are expected to impact the performance of GALFIT: stellar density, PSF FWHM and image depth. We conclude that at relatively bright magnitudes ($i < 21.5$) the completeness has a relatively weak dependence on these quantities, and high completeness can be maintained without much loss of survey area.

In Section 2.4.3 we analysed the properties of the converged fits, isolating a small fraction ($< 5\%$) of outliers in magnitude recovery, and a branch of objects with high Sérsic indices and large radii that we believe to be spurious. Removing low S/N galaxies efficiently cleans the sample of these populations. Following this basic validation, we calibrate the Sérsic measurements using state-of-the-art UFIG image simulations, deriving correction vectors via the comparison of input model parameters and the resulting fits by GALFIT. In Section 2.5 we repeated the above mentioned diagnostics for the non-parametric fits, benefiting from the internal diagnostic flags provided by ZEST+ itself in order to quantify the quality of the image and so the reliability of the measurements. For the non-parametric dataset we adopted the same method to derive the calibrations described in Section 2.2.2, finding that corrections are stronger for low signal to noise galaxies, similar to the parametric case. In particular, we highlight the calibration of galaxy concentration, which is adversely affected due to fact that ZEST+ cannot account for the PSF.

Finally, we summarised the selection function and a recommended set of cuts to form a basic science sample. Our catalogue represents a valuable instrument to explore the properties and the evolutionary paths of galaxies in the DES Y1 survey volume, which

will be used in a series of forthcoming publications.

ACKNOWLEDGEMENTS

FT would like to thank Sandro Tacchella for his useful suggestions about fitting calibration and fruitful discussions.

Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas A&M University, and the OzDES Membership Consortium.

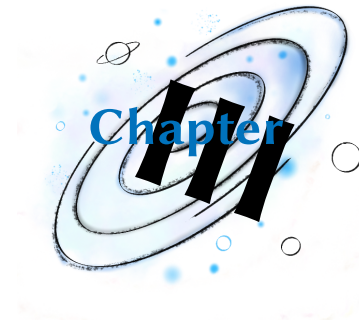
Based in part on observations at Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

The DES data management system is supported by the National Science Foundation under Grant Numbers AST-1138766 and AST-1536171. The DES participants from Spanish institutions are partially supported by MINECO under grants AYA2015-71825,

ESP2015-66861, FPA2015-68048, SEV-2016-0588, SEV-2016-0597, and MDM-2015-0509, some of which include ERDF funds from the European Union. IFAE is partially funded by the CERCA program of the Generalitat de Catalunya. Research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013) including ERC grant agreements 240672, 291329, and 306478. We acknowledge support from the Australian Research Council Centre of Excellence for All-sky Astrophysics (CAASTRO), through project number CE110001020, and the Brazilian Instituto Nacional de Ciência e Tecnologia (INCT) e-Universe (CNPq grant 465376/2014-2).

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

Image feature extraction and galaxy classification: a novel and efficient approach with automated machine learning



La maggior parte degli uomini sono come una foglia secca, che si libra nell'aria e scende ondeggiando al suolo. Ma altri, pochi, sono come le stelle fisse, che vanno per un loro corso preciso, e non c'è vento che li tocchi, hanno in se stessi la loro legge e il loro cammino.

Most people are like a falling leaf that drifts and turns in the air, flutters, and falls to the ground. But a few others are like stars which travel one defined path: no wind reaches them, they have within themselves their guide and path.

HERMANN HESSE, *SIDDHARTA*, 1922

The content of this chapter is based on [Tarsitano, Bruderer, et al., 2021](#).

3.1 INTRODUCTION

Galaxy morphology plays an important role in our studies and understanding of galaxy evolution. Structural components such as bulges, disks, spiral arms and bars formed during galaxies' aggregated formation histories ([Combes and Sanders, 1981b](#); [de Jong, 1996b](#); [Bruce G. Elmegreen et al., 1996](#)). As such, morphology is related to other properties that depend on formation and assembly history, such as colour, stellar-mass and recent Star Formation Rate (SFR) ([Baldry et al., 2004](#); [Noeske et al., 2007](#); [Cano-Díaz et al., 2019](#)). By looking at the relation between mass and SFR ([Schiminovich et al., 2007](#); [Goncalves et al., 2012](#); [Peterken et al., 2021](#)), astronomers have been able to distinguish

between three different populations. Most star-forming galaxies belong to the *main sequence*, and present morphological features typical of spiral or irregular galaxies. Objects in this population are also called *Late-Type Galaxies* (LTG). We can identify another population with much lower SFR and different shapes, mostly elliptical or bulge-dominated morphologies: we refer to these as *Early-Type Galaxies* (ETG). The transition between ETG and the *main sequence* is smoothed by an intermediate and less heavily populated region, called the *green valley* (Salim, 2014; Kevin Schawinski et al., 2014).

Historically, galaxies were classified as early or late type by visual inspection, with a modern example of classification in this way provided by P. B. Nair and Roberto G. Abraham, 2010. Recent *citizen science* projects like *Galaxy Zoo* (Lintott et al., 2008; Simmons et al., 2017; Lingard et al., 2020) use the same approach, while benefiting from a huge network of volunteers who are asked to classify galaxies. Quantitative methods for classifying structural properties includes modeling galaxy light profiles with 2D analytic functions, and fitting them to galaxy images. The most commonly used model is the *Sérsic profile* (Sérsic, 1963), a parametric function with parameters describing structural properties such as size, magnitude, ellipticity, inclination and the rate at which light intensity falls off with radius (*Sérsic index*). The latter quantifies the concentration of light and it is often used to distinguish between ETG and LTG. In fitting galaxy images, the Sérsic model must be convolved with the Point Spread Function (PSF), in order to account for the seeing and any instrumental distortions of the image. This method is proven to be robust when the multi-dimensional fitting delivers non-degenerate solutions. Alternative non-parametric approaches can be used to analyse the light distribution and quantify the galaxy concentration and level of asymmetry and to search for clumpy regions (Conselice, Bershad, and Jangren, 2000). In this case, the PSF is not taken into account. Several large catalogues of galaxy morphologies exist, based on parametric fitting (Simard et al., 2011), a non-parametric approach (Cano-Díaz et al., 2019) or both (Tarsitano, Hartley, et al., 2018).

Going beyond the traditional methods cited above, machine learning algorithms present an attractive way forward in classifying galaxy images. This has been proven to be successful, as in Banerji et al., 2010, Dieleman, Willett, and Dambre, 2015 and Walsley et al., 2019, where data sets from *Galaxy Zoo* were used combining human and machine intelligence. Supervised CNN have been used also on CANDELS images (Grogin et al., 2011; Koekemoer et al., 2011) to provide galaxy visual classifications (Huertas-Company, Gravet, et al., 2015a; Tuccillo et al., 2018), to find specific structural features such as bars (S. Abraham et al., 2018) and to classify galaxies according to their bulge+disk composition Ghosh et al., 2020. In Cheng et al., 2020 unsupervised machine learning algorithms on SDSS images (York, Adelman, Anderson, et al., 2000) were explored, foreseeing the application of these techniques on data with higher resolution and deeper depth from the Dark Energy Survey (DES) (Dark Energy Survey Collabora-

tion et al., 2016) and the Euclid Space Telescope (Amiaux et al., 2012).

The recent increase in the use of machine learning methods has been beneficial for astronomy research, and is of particular interest for extracting information on the evolutionary paths of galaxies from their morphologies. Especially with the exponential rise in the amount of data from modern surveys it has become important to understand and apply intelligent algorithms able to classify galaxies with the same accuracy as human experts, if not even outperforming them. In the works cited above, the CNNs classify galaxy images by processing different levels of information in each layer, aiming at a progressive recognition of complex features. In this approach, image recognition works well if the objects have clear edges. However galaxies' outskirts are smooth: even traditional methods used to measure structural properties, namely the 2D parametric fitting and the non-parametric analyses, are often prone to inaccuracies due to the difficulty of separating galaxy wings from the background. These boundary effects can be mitigated by using model constraints, but cannot completely prevent inaccurate estimations of structural parameters. Machine learning techniques are also subject to mis-classifications for the same reasons, especially with low-resolution images. Another factor to account for when adopting intelligent algorithms is the data management and the speed of the analyses. The increasing volume of available images is difficult to manage and the number of operations processed in CNN models is high. Both training and testing large image data sets requires a lot of time and significant computational costs. These limiting factors led us to search for an alternative method, which performs an isophotal analysis of the galaxy light distribution, stores the information in a more manageable data format and performs classification lowering the total computational costs. In this paper we show that this method, applied to the second public data release, DR2, (T. M. C. Abbott et al., 2021; Morganson, Gruendl, Menanteau, Carrasco Kind, Y. .-. Chen, et al., 2018) of the Dark energy Survey (DES), can lead to competitive results. Recently Vega-Ferrero et al., 2020 applied CNN to an earlier public release (DES DR1), producing a galaxy morphology catalogue which has an accuracy of 87% and 73% for ETG and LTG, respectively, up to magnitude $m_r < 21.5$. In our work we analyse galaxies up to magnitude $m_i < 20$, reaching comparable accuracy: 86% for ETG and 93% for LTG.

Section 3.2 contains more details about the data set. Our method and details on how we perform the isophotal analysis of galaxy images, extract features from their light distribution and collect those into sequences, is described in Section 3.3. The sequences are then processed through a neural network designed and run in the framework of Modulos AutoML¹. More details are given in Section 3.4. We present the results in Section 3.5 and discuss further developments in Section 3.6.

¹<https://www.modulos.ai/>

3.2 DATA

In this work, we use public images from the Dark Energy Survey DR2 release (T. M. C. Abbott et al., 2021; Morganson, Gruendl, Menanteau, Carrasco Kind, Y. -C. Chen, et al., 2018; Flaugher, Diehl, Honscheid, T. M. C. Abbott, Alvarez, Angstadt, J. T. Annis, Antonik, Ballester, Beaufore, G. M. Bernstein, et al., 2015), available through the public *DES Data Management*.² In this section, we provide an overview of the survey, describe the structure of the data set and define the selection function for our sample.

3.2.1 THE DARK ENERGY SURVEY

The Dark Energy Survey (DES) is a project aiming to map hundreds of millions of galaxies to measure the effects of dark energy on the expansion history of the Universe and the growth of cosmic structure. The collected data are analysed through different methods: gravitational lensing, galaxy clustering and Baryonic Acoustic Oscillations (BAO). DES used the Dark Energy Camera (DECam) to detect more than 300 million galaxies between the years 2013 and 2019 (Flaugher, Diehl, Honscheid, T. M. C. Abbott, Alvarez, Angstadt, J. T. Annis, Antonik, Ballester, Beaufore, and al., 2015). Although conceived for cosmological research, the vast data set assembled by DES represents a powerful survey for the fields of galaxy evolution, stellar populations and Solar System Science too (T. Abbott et al., 2016). Moreover, in 2017 DECam provided the optical counterpart of the gravitational wave event GW170817 studied in detail in Palmese et al., 2017a. The camera has a 2.2° diameter field of view and a pixel scale of $0.263''$ (Flaugher, 2005). It is mounted on the Victor M. Blanco 4-meter Telescope at the Cerro Tololo Inter-American Observatory (CTIO) located in the Chilean Andes.

3.2.2 THE DATA SET

The DES survey area is covered by images in five photometric bands, g,r,i,z,Y . The single exposure images have integration time of 90 seconds in the g,r,i,z and 45 seconds in the Y band. Data are later processed through the DESDM (DES Data Management) pipeline, which first applies calibrations and coadds the images, then detects and catalogues all the objects in those images (Drlica-Wagner, Sevilla-Noarbe, Rykoff, Gruendl, Yanny, Tucker, Hoyle, Carnero Rosell, et al., 2017; Morganson, Gruendl, Menanteau, Carrasco Kind, Y.-C. Chen, et al., 2018). In the image co-addition, the pipeline combines overlapping single-epoch images in one filter and remaps them to artificial tiles on the sky as described in Sevilla et al., 2011, Desai et al., 2012 and Mohr et al., 2012. Object detection is made using a specific software, called SExtractor (Bertin, E., 2011), which

²<https://des.ncsa.illinois.edu>

Table 3.1: Summary of the cuts applied to SExtractor catalogues for the sample selection.

SELECTION TYPE	SELECTION CUT
Image flags	FLAGS = 0, 2
Magnitude	MAG_AUTO < 20
S-G	MODEST > 0.005
S/N	FLUX_AUTO/FLUXERR_AUTO > 300

extracts structures from the background and distinguishes between point-like (stars) and galaxies. Then, it performs a photometric analysis, where each object is enumerated and assigned to a set of specific properties, collected in a catalogue. For this analysis, properties of the light distribution are measured, namely the object brightness, quantified in MAG_AUTO, and its size, called FLUX_RADIUS, which includes half of the galaxy light. We use these measures to identify and optimise the sample analysed in this work.

3.2.3 SAMPLE SELECTION

We apply cuts to the SExtractor catalogues (see below) to select a final sample of 6525 galaxies. We choose objects which are neither truncated nor corrupted or blended to other objects by setting FLAGS = 0, 2. Additionally, we choose bright objects by applying a cut in magnitude. In Tarsitano, Hartley, et al., 2018, we observe that robust fits are obtained for objects up to a magnitude of 21.5 in the *i*-band. In order to work with optimal isophotal fitting, in this analysis we make a more conservative cut, setting in the same filter MAG_AUTO \leq 20. For the same reason, we also adopt the cut in signal-to-noise $S/N > 300$. In Tarsitano, Hartley, et al., 2018 we also flagged those galaxies with size smaller than or comparable to the PSF, because in those cases the PSF significantly affects the way the concentration of light is modelled, leading to degeneracies in the estimation of the size and Sérsic index. Therefore our selection function excludes the galaxies with size smaller than 4 px in the *i*-band. We also check that the selected objects have physically meaningful measurements, avoiding galaxies with negative or null radii. In processing the data (see Section 3.3), we will make use of the *Kron radius*, which is the radius within which approximately 90% of the galaxy light is included. According to the definition in SExtractor, we consider as *Kron radius* the product between the KRON_RADIUS and the semi-major axis of the galaxy A_IMAGE. Finally, we exclude from our sample the point-like objects by applying a cut to the MODEST parameter (Drlica-Wagner, Sevilla-Noarbe, Rykoff, Gruendl, Yanny, Tucker, Hoyle, Rosell, et al., 2018). The sample selection is summarized in Table 3.1. Additional information about the MODEST star/galaxy classifier and the SExtractor catalogues can be found here: <https://des.ncsa.illinois.edu/releases/y1a1/gold>.

3.3 METHOD

In this section, we describe our process to transform galaxy images into one-dimensional feature vectors for classification using machine learning. As already mentioned in the introductory sections, this method involves few and fast steps, which is an advantage compared to classification methods that involve several labourious manipulations. More precisely, we refer to two main steps:

1. production of postage stamp images;
2. extraction of profiles.

We describe the steps below, highlighting the main differences with existing methods.

3.3.1 PRODUCTION OF STAMPS

For each of our selected galaxies (see Section 3.2.3), we cut square postage-stamp images from the relevant DES tiles, with dimensions equal to four times the *Kron radius*. This size is chosen to ensure that the image includes the galaxy light distribution entirely and sufficient non-object pixels to be able to determine the background level. For this operation, we use the publicly available CANVAS algorithm³ (Cut ANd VALidate Stamps), presented and optimized in Tarsitano, Hartley, et al., 2018.

BACKGROUND

In standard analyses such as parametric fitting, the background needs to occupy at least 60% of the area of the stamp, in order to obtain a correct fit. The fitting algorithm, in fact, needs to distinguish the light signal from the sky, which becomes challenging towards the outskirts and faint wings of a galaxy. Therefore, a clear separation is only possible if the background occupies a larger area of the stamp than the galaxy (Chien Y. Peng et al., 2010). In works that perform image classification with neural networks, the preparation of the sample usually includes data-augmentation with image simulations to artificially place a galaxy with well-known classification at different redshifts. This process requires the PSF to be deconvolved and the reconstruction of images with an appropriate percentage of background. In our work, we do not perform parametric fitting and we extract information solely from the area inside the *Kron ellipse* of the galaxy. Hence our approach is robust to minor defects or mis-estimations of the background or the image stamp size, and we need not be concerned with multiplying the size of our input images to ensure sufficient background coverage.

³<https://github.com/Federica24/Cosmo>

NEIGHBOURING OBJECTS

A potential source of errors for the galaxy classification is the presence of neighbouring objects. If not taken into account, both standard fitting algorithms and CNNs are prone to imprecise estimations or classification of the galaxy light profile. We consider two cases:

- the neighbour falls outside the *Kron ellipse* of the galaxy;
- the neighbour is placed inside the *Kron ellipse* of the galaxy (partially or fully).

The first scenario is negligible for our analysis, since we only consider the pixels inside of the *Kron ellipse*. However, we cannot ignore the second case. Our aim is to minimize the number of manipulations applied to the images, so we do not apply any algorithm to identify such cases. Moreover, we would need to distinguish cases that are due to chance alignments from more interesting but possibly similar-appearing cases due to, e.g., galaxy-galaxy mergers or star-forming clumps. This latter task is an avenue of future research for our method, but as contaminants do not change the overall trend of the sequences we obtain for elliptical and spiral galaxies, we do not apply corrections for them in the present work.

3.3.2 EXTRACTION OF PROFILES

The extraction of profiles relies on the elliptical isophote analysis of the galaxy in question. Isophotes are curves connecting locations with the same brightness. We use the algorithm of *Elliptical Isophote Analysis* available in the `Photutils` `Astropy` package (Bradley et al., 2020). The algorithm searches for elliptical isophotes iteratively, as described in detail in Jedrzejewski, 1987, up to a user-defined limit, expressed in terms of a maximum value for the semi-major axis of the ellipse. We set this limit to 0.7 times the *Kron radius* in order to exclude the faint wings of a galaxy light distribution from the analysis. Specifically in our fitting routine we observed that such wings result in a noisy tail in the one-dimensional sequences and do not add information useful for distinguishing between different classes of galaxies. Once we measure the isophotes, we proceed by *radially and concentrically* collecting the intensity of pixels falling on the curves, one ellipse at a time. The points collected in this order form our sequence, which we convert to logarithmic scale and normalise such that the brightest pixel has value unity. We can visualize this procedure in Figure 3.1, where the isophotes and their radial intensities collected in the series are matched by colour. The series show a clear and expected pattern of decreasing intensity along the x-axis as the isophotes are read from the centre towards the outskirts of the galaxy. This pattern varies between early and late-type galaxies, as can be seen in 3.2. For early-type, mostly elliptical galaxies (upper panel),

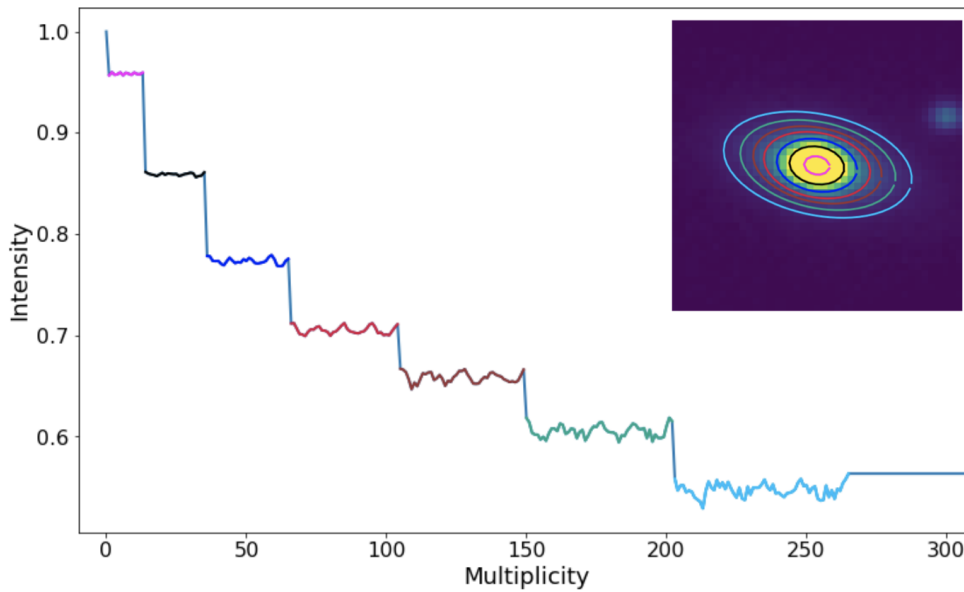


Figure 3.1: Example of a series extracted from galaxy images and used for classification. The galaxy (an early-type in this case) is processed through an isophote analysis, then the logarithm of the radial intensities of pixels lying on each detected ellipse is read and stored in a series, where the multiplicity axis in the main panel is a counter for the pixel values read. The intensity is from the inner to the outer ellipse. Isophotes and their collected intensities are matched by colour.

the trend resembles a step-function, showing regular patterns with decreasing intensity: each step represents an isophote. For late type galaxies, we observe a different trend: in addition to a slower fall off in intensity due to their lower Sérsic index, the presence of spiral arms or clumps adds irregular spikes to the ideal step function. An example is shown in the lower panel of the figure, where we consider a barred-spiral galaxy.

3.4 AI FRAMEWORK AND MODULOS

We run a *Modulos AI* workflow on a data set randomly split between a test (2175 galaxies) and training+validation sample (4350 galaxies), with all objects visually inspected according to their corresponding 1-D sequences and images. As previously mentioned in Section 3.3, we show in Figure 3.2 examples of sequences for ETG (upper panel) and LTG (lower panel). In this section we describe the properties of the workflow in more detail. We use the Modulos⁴ AutoML platform (version 0.3.5) to search for suitable models. The platform is designed to perform automated model selection and training for machine learning tasks, and works in the following way:

⁴<https://www.modulos.ai/>

3.4. AI framework and Modulos

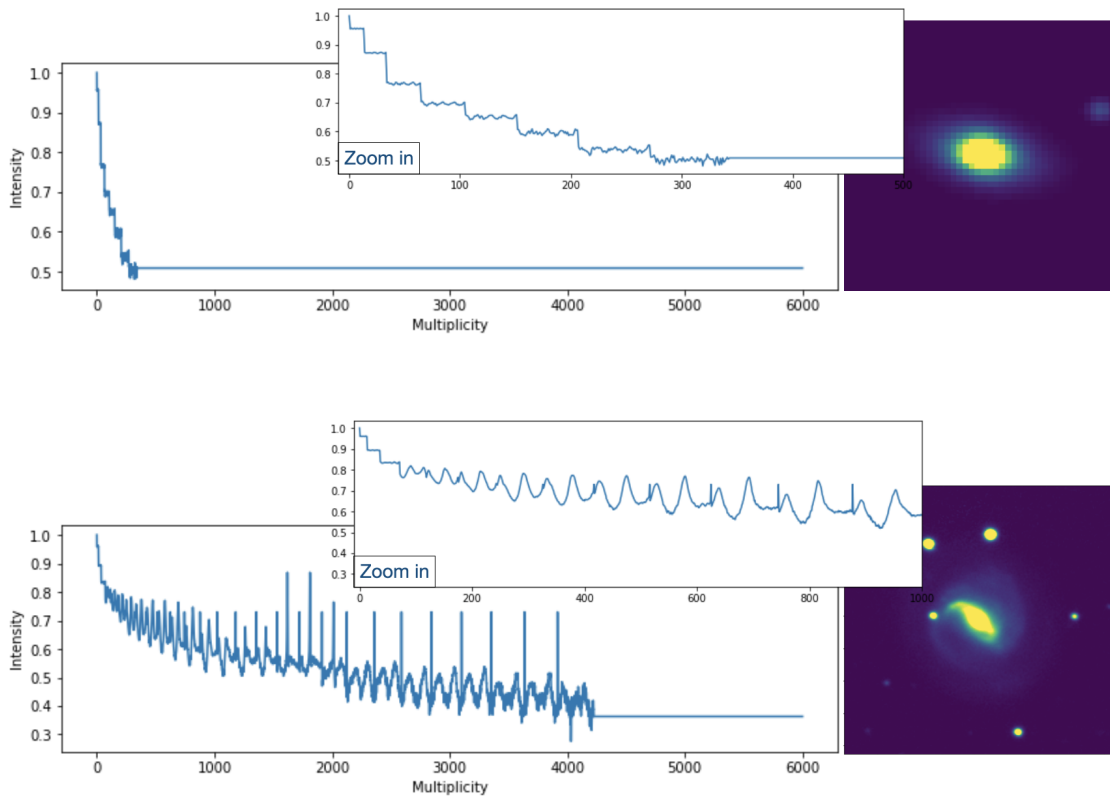


Figure 3.2: Example of classification between early-type (upper panel) and late-type (lower panel) galaxy, according to their time-series-like profile.

1. Workflow configuration (ML task): the user selects the data set to be processed, and sets an objective for which it is optimized.
2. Schema matching: the platform detects the schema of the desired input and output. It then proposes the feature extraction methods and machine learning models applicable to the data set and target objective.
3. Optimization: using a Bayesian optimizer ([Srinivas et al., 2009](#)), the platform tries out various combinations of feature extractors, models and their parameters. At each search step, the platform selects a feature engineering method and a model, chooses its architecture and hyperparameters, and trains it. After completing training, the platform uses a validation set to score this particular choice.
4. End point: There is no clearly defined end point at which the "best model" has been found. However, after a while, the scores for the models begin to converge. As a default, the platform stops if there are no score improvements within 200 steps.
5. Download: Any trained model can be downloaded and used. We choose the best-performing model.

The key advantages of a platform such as AutoML are that the search for models and configuration is principled and not biased by human intervention. It is also significantly more efficient than optimization searches performed "by hand". During this project, the vast majority of the time was spent on the preparation and the analysis of the data set, with only a few hours required for the automated classification. In our case, we found a suitable solution within 4.30 hours of compute time (14 min 19 sec to train the specific solution).

The objective we set for optimisation is the F_1 score. For multi-class classification, the total F_1 score is the unweighted arithmetic mean of the $F_{1,i}$ scores of each class i (macro-averaged). These are the harmonic mean of the precision and recall of the classified samples for each respective class, i.e.

$$F_{1,i} = 2 \cdot \frac{\text{precision}_i \cdot \text{recall}_i}{\text{precision}_i + \text{recall}_i} = \frac{2\text{TP}_i}{2\text{TP}_i + \text{FP}_i + \text{FN}_i}, \quad (3.1)$$

where TP_i are the true positives for the classified samples for the class i and FP_i and FN_i are the false positives and false negatives respectively.

3.5 RESULTS AND DISCUSSION

The automated machine learning framework returns as best solution an XGBoost model with a PCA decomposition as feature engineering method. XGBoost (Extreme Gradient Boosting) is a decision-tree based Machine Learning algorithm using a gradient boosting framework (Friedman, 2001). This model reaches a F_1 macro score (see eq. 3.1) of 90% on training data and 89% on test data, which suggests it is not prone to over or under-fitting. Our best model is publicly available at <https://github.com/Federica24/Cosmo> and can be applied to any DES data processed as described in the previous sections.

3.5.1 INFORMATION EXTRACTION

In order to understand why a combination of an XGBoost model and a PCA feature engineering method is found to be performing best, we review the key information accessed during classification by our model and compare it to the structure of our data. In Figure 3.3 we show the collection of sequences classified as ETG (in blue) and LTG (in green). Additionally, we have highlighted 10 arbitrary sequences from each class to illustrate the individual profiles. The human eye is able to distinguish between the two classes by looking at the slopes of the sequences (greater for ETG) and comparing the abundance of spike-like features, which correspond to spiral arms of LTG, and to the smoothness of the sequences representing ETG. If spikes occur in the latter, they are more sparse and might refer to the presence of a neighbour (see Section 3.3 for reference). The in-

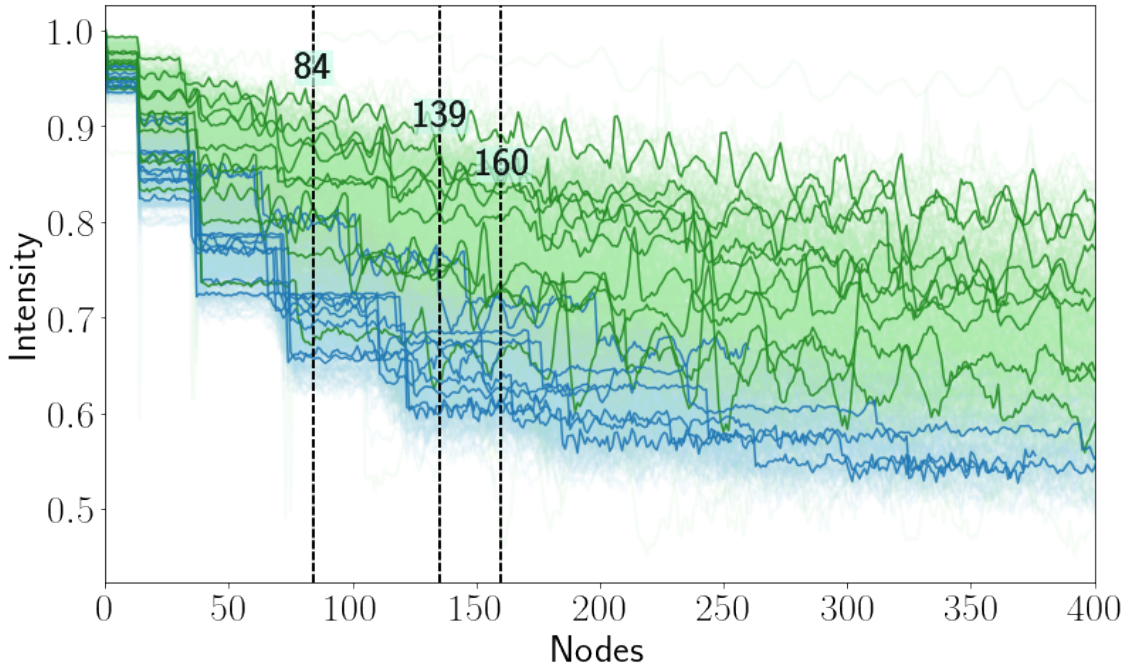


Figure 3.3: Plot of the sequences corresponding to ETG (in blue) and LTG (in green). Additionally, 10 arbitrary ETG and LTG are drawn. The black, vertical lines indicate the nodes with the largest Gini feature importance.

formation contained in the slopes and features is enhanced by the PCA analysis, which encodes it into a set of components. Each component brings a unique contribution to the automatic classification, giving the features different weight. By computing the Gini feature importance of the collection of decision trees, we can then understand which features contribute the most in predicting a class for a new data point in our model. We show the three most important nodes picked up by the Gini feature importance with black, vertical lines on Figure 3.3.

We notice that the three nodes refer to points where the two classes of sequences on average start falling at different rates. In other words, these nodes are the ones with the largest discriminatory power for the collection of decision trees, which perform classifications by sequentially dividing up the sequences. We find that the most important node for our best model (numbered 139) is 15 times more important than the others. This may be because it provides the highest signal-to-noise estimate of the light intensity fall-off. The next few most important nodes provide supporting information to effectively distinguish objects' effective radii. This sequential split becomes especially important to distinguish sequences in the smooth transition region where a few sequences from different classes show similar slopes. This sequential splitting as a qualitative measure of the slopes is physically meaningful: in fact, as commented already in Section 3.3.2, we expect LTG sequences to fall off slower than ETG, due to their lower Sérsic index.

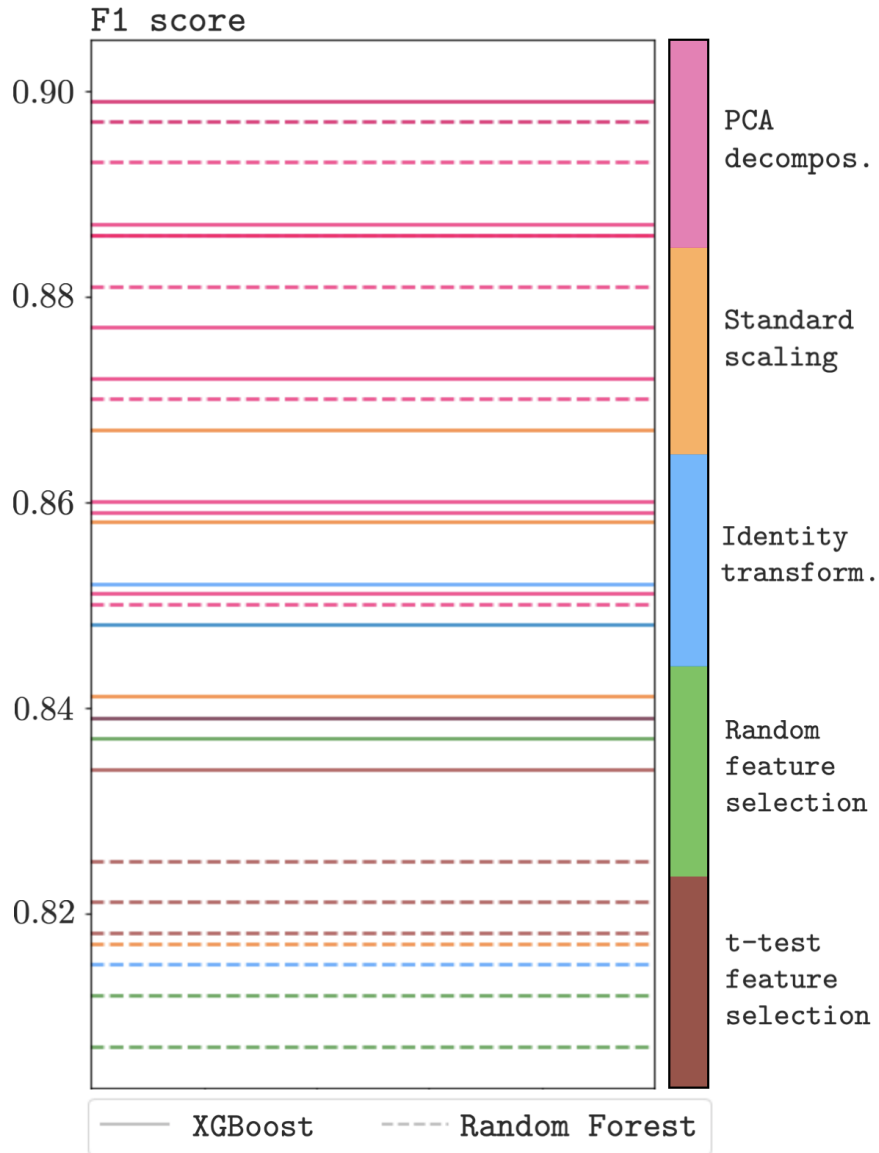


Figure 3.4: Summary of the results provided by the Modulos AutoML platform. The solutions are Decision Tree models, either XGBoost (solid lines) or Random Forest (dashed lines), which are able to sequentially capture and combine information from the sequences in the data set. The lines are color-coded by the feature engineering method. PCA decomposition is associated to the solutions maximizing the F1 score (in the y-axis), given their ability in rotating the feature space so to emphasize the most important nodes in the sequences.

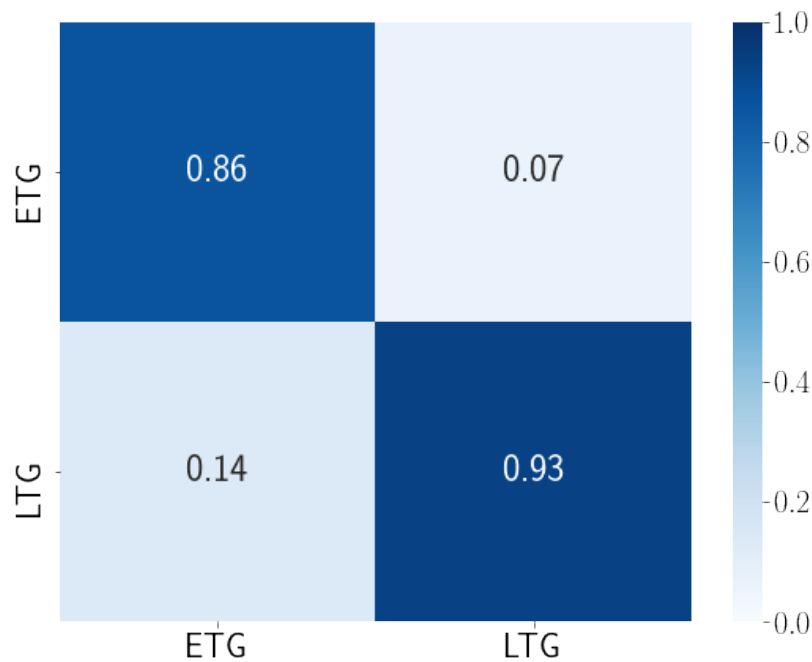


Figure 3.5: Confusion matrix representing the accuracy achieved in classifying galaxy profiles. The x-axis shows the true values, while the y-axis are the predicted categories. The main diagonal shows the correct classifications. The model seems quite robust in classifying the early-type galaxies of the sample.

As mentioned in Section 3.2.3 the PSF can change the rate of fall off from one isophote to the next, leading to a global rescaling of first steps of each sequence. This effect can be particularly significant for galaxies with size smaller than the PSF, but we did not include those in our sample. The scores of all trained solutions provided by the AutoML platform are summarized in Figure 3.4. Looking at these offers insights into which combinations of machine learning models and feature engineering methods are optimal for our task. The models shown are both decision trees-based and are either XGBoost (solid lines) or Random Forest models (dashed lines) and are color-coded by the feature engineering method. We observe that the PCA decomposition performs the best and is more important to the success of the overall model than the choice of XGBoost versus random forest. PCA rotates the feature space in order to successfully emphasize the slopes of the profiles. On the lower end we find the Random and the t-test feature selection methods: since they only select a subset of nodes (either random selection or by applying the Student’s t-test), they seem to be less likely to pick up the most important information to distinguish the profiles and their slopes.

Finally, we use the aforementioned best model to make predictions on our test sample. We quantify the distance between the predictions and the true values by computing the confusion matrix (Figure 3.5), normalized over the number of predictions, for which

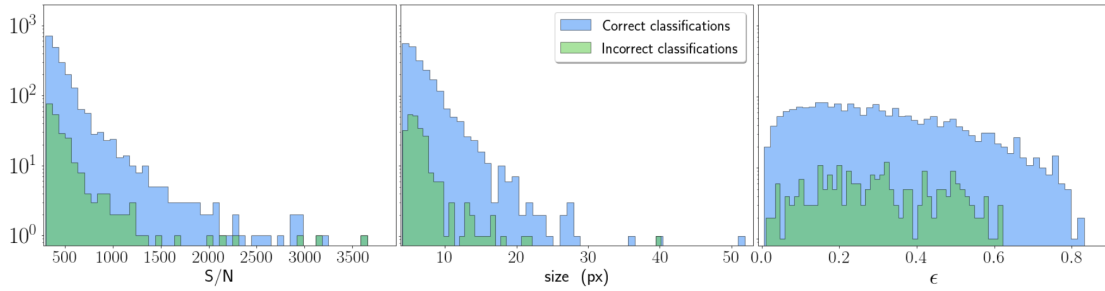


Figure 3.6: Properties of the classified sample in terms of signal-to-noise ratio (left panel), size (central panel) and ellipticity (right panel), distinguishing between objects with and without successful classification. This diagnostic plot alone cannot trace all the mis-classifications. A clearer test is shown in Figure 3.7.

we used the Python `scikit-learn` library. The main diagonal shows the amount of objects correctly classified, while the off-diagonal elements quantify incorrect classifications. The majority of mis-classified galaxies have low S/N ratios and tend to have small sizes and ellipticity, as shown in Figure 3.6.

3.5.2 MODEL FAILURES AND FUTURE PERSPECTIVE

Although there is by no means a simple cut we can perform to identify wrongly classified cases, inspecting examples of the isophotal fittings of both successful and unsuccessful classifications, we notice that objects with bad isophotal fitting tend to be mis-classified more often. This is compatible with the outcome shown in the confusion matrix, where it yields more incorrect classifications for ETG: a poor isophotal fit introduces perturbations into a 1D-sequence which would show a regular pattern typical of such galaxies. A few examples of galaxies with poor isophotal measurements are shown in Figure 3.7. Due to the small apparent size or low image resolution, the fitting does not model the light distribution well, resulting in an incorrect fit of the wings. As can be seen in the middle and right-hand panels, this manifests as a sudden change in the angular orientation of some isophotes with respect to the central regions of the galaxy. This issue can be corrected by applying sigma clipping to the recovered set of isophotes, identifying those that have parameters that are discrepant with the majority of fitted isophotes. However, the appropriate level of clipping varies from object to object and, at present, is not straight forward to determine in an automated way. As our aim is to describe a fully-automated method that can be run efficiently on large survey data, we thus quote our results without this fix. We will return to the issue of mis-aligned isophotes in future work, where we develop a routine to perform flexible isophotal fitting automatically, combining the structural information on the isophotes (e.g. position angle, ellipticity) with new feature engineering solutions, and apply our method to more contextual out-

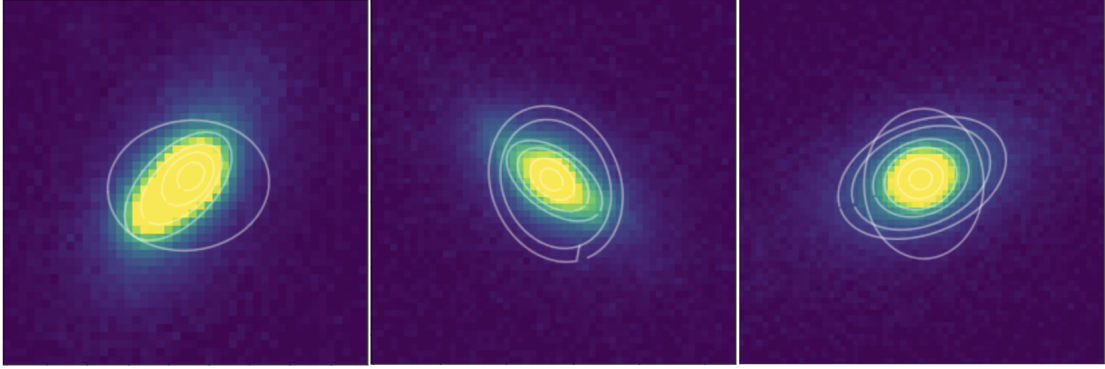


Figure 3.7: Examples of isophotal fitting for mis-classified galaxies. If compared to Fig. 3.2, here we notice that the fitting fails at modelling galaxy wings and introduces rotations in the isophotal ellipses.

puts, such as the presence of clumps or spiral arms.

3.6 CONCLUSIONS

In this work, we describe a novel approach to galaxy morphological classification. It consists of first analysing the main features of the two-dimensional light distribution in a galaxy image with isophotal fitting. This then allows to unravel it to a one-dimensional sequence. The advantage of such an approach is the low complexity of one-dimensional data, which makes both data storage and processing easier and faster compared to classification methods directly analysing images (e.g. parametric fitting). The selection, calibration, and training of classification models is then performed using the Modulos AutoML platform, which allows users to intuitively build and run their workflows and automatizes the search and training of ML solutions. Using this platform also leads to a significant reduction of time spent on building machine learning algorithms. This allowed us to quickly test hypotheses and focus on the scientific analysis. We found ensembles of decision trees (XGBoost and Random Forest models) with a PCA decomposition as a feature engineering method, which transform the feature space to make the profiles more discrepant, to perform well. The resulting best performing model (XGBoost) is physically meaningful as it picks up on the differing slopes of the light profiles of galaxies: LTG profiles are expected to fall off slower due to their lower Sérsic index. We make the best ML solution we have found freely available at <https://github.com/Federica24/Cosmo>. It can be used to predict the galaxy type of other galaxies in the DES DR2 data set. We obtain an overall F_1 score of 90% and 89% on training and test data, respectively, which proves that the dimensionality reduction of the data, even though it implies information loss, still contains enough information to successfully classify galaxies. Our accuracy is

comparable to the results found by (Vega-Ferrero et al., 2020) for image-based classification using DES images. In the future, we will expand upon our promising results by developing a more robust isophotal measurement approach to focus on performance at low S/N, and target higher context features, such as bars, spiral arms and clumps.

ACKNOWLEDGEMENTS

This research made use of Photutils, an Astropy package for detection and photometry of astronomical sources (Bradley et al., 2020). The authors thanks Modulos for the usage of their platform to perform image training and classification.

This project used public archival data from the Dark Energy Survey (DES). Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft, and the Collaborating Institutions in the Dark Energy Survey. The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the OzDES Membership Consortium, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M University. Based in part on observations at Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

DATA AVAILABILITY

The best automated classification model presented in this paper and discussed in Section 3.5 is publicly available at <https://github.com/Federica24/Cosmo> and can be used to classify any DES data from the public release DR2 processed with isophotal fitting, as described in this work.

Predicting Cosmological Observables with PyCosmo



Facilis descensus Averno:
noctes atque dies patet atri ianua Ditis;
sed revocare gradum superasque evadere ad auras,
hoc opus, hic labor est.

The gates of Hell are open night and day;
Smooth the descent, and easy is the way:
But to return, and view the cheerful skies,
In this task and mighty labor lies.

PUBLIUS VERGILIUS MARO, *AENEIS*, VI, 126-129

The content of this chapter is based on Tarsitano, Schmitt, et al., 2020.

4.1 INTRODUCTION

Present research in cosmology investigates the validity of the Λ CDM model and its extensions by testing its parameters through observational probes, such as the Cosmic Microwave Background (CMB), Baryonic Acoustic Oscillations (BAO), weak lensing, cluster counts, supernovae and galaxy surveys. The combination of these observables has high constraining power on the parameters of these cosmological models. Current and upcoming cosmological experiments, such as DES¹, DESI², LSST³, Euclid⁴ and WFIRST⁵

¹<http://www.darkenergysurvey.org>

²<http://desi.lbl.gov>

³<http://www.lsst.org>

⁴<http://sci.esa.int/euclid/>

⁵<http://wfirst.gsfc.nasa.gov>

aim at precise measurements of these observables, thus demanding highly accurate theoretical predictions. Codes fulfilling this task are already available, such as COSMICS (Bertschinger, 1995), CMBFAST (Uros Seljak and Zaldarriaga, 1996b), CMBEASY (Doran, 2005), CAMB (Lewis, Challinor, and Lasenby, 2000), CLASS (Lesgourgues, 2011), iCosmo (A. Refregier, A. Amara, et al., 2011), CosmoLike (Krause and Eifler, 2017a), CosmoSIS (Zuntz et al., 2015), CCL (Chisari et al., 2019). PyCosmo (A. Refregier, Gamper, et al., 2018b) is a recently introduced Python-based framework which provides cosmological model predictions, fitting within the upcoming new era of precision cosmology. As a Boltzmann solver, it computes solutions to the set of Einstein-Boltzmann equations, which govern the linear evolution of perturbations in the Universe. These calculations are at the core of most cosmological analyses. PyCosmo introduces a novel architecture that uses symbolic calculations. As described in a previous work (A. Refregier, Gamper, et al., 2018b) the code, based on the *Python* library *Sympy* (Meurer et al., 2017), uses computer algebra capabilities to produce fast and accurate solutions to the set of Einstein-Boltzmann equations, and provides the user a convenient interface to manipulate the equations and implement new cosmological models.

The PyCosmo version we present in this paper provides accurate predictions without calling the Boltzmann Solver, which is currently under testing and optimization. The computed cosmological quantities are defined in terms of background computations, linear and non-linear perturbations (fitting functions for matter power spectra) and observables (angular power spectra with and beyond the Limber Approximation). The fitting functions for the linear and non-linear power spectrum, which are used to compute predictions for angular power spectra with the Limber Approximation (LoVerde and Afshordi, 2008), have been extensively tested. In particular we refer to the *Halofit* fitting function (Smith et al., 2003; Takahashi et al., 2012) and to a revised version of the Halo Model, presented in Mead et al., 2015 as a more accurate function which also accounts for baryonic feedback; below in this section we will refer to it as the *Mead et al. model*. Both fitting functions within PyCosmo have been used in the Monte Carlo Control Loops (MCCL) analysis of the Dark Energy Survey (DES) data described in Kacprzak et al., 2019. The CMB angular power spectrum is computed using the approach of line-of-sight integration proposed in Uros Seljak and Zaldarriaga, 1996a.

In order to assess the accuracy of such computations it is important to compare PyCosmo to other available codes, with the aim of obtaining the highest possible agreement between algorithms with independent implementations. In PyCosmo such comparisons are constantly monitored through a system of unit tests. Conceived as a user-friendly code, the currently tested and validated version of PyCosmo is currently available on a public hub, called PyCosmo Hub and accessible from <https://cosmology.ethz.ch/research/software-lab/PyCosmo.html>. This server hosts several Jupyter notebooks showing how to use PyCosmo by including tutorial-notebooks and examples.

Registered users can use PyCosmo for their own calculations without the need of local installations. More details about the hub will be provided later in Section 4.2.5.

Future developments focus on a refactoring of the Boltzmann Solver, plus developments of $wCDM$ and massive neutrinos computations, and involve new models for non-linear perturbations as presented in [Bartelmann, Dombrowski, et al., 2020](#). These aspects will be addressed by a set of future works and more advanced code releases. This paper focuses on the implementation of the cosmological observables and the tests made in order to check their accuracy. In this context PyCosmo is compared to the following codes (see also [Chisari et al., 2019](#) for an earlier comparison of some of these codes):

- CLASS ([Lesgourgues, 2011](#)), a C-based Boltzmann solver widely used to compute theoretical predictions, and its python wrapper, `classy`;
- `iCosmo` ([A. Refregier, A. Amara, et al., 2011](#)), an earlier cosmology package written in IDL;
- Core Cosmology Library ([Chisari et al., 2019](#)), developed within the LSST Dark Energy Science Collaboration ([LSST Dark Energy Science Collaboration, 2012](#));
- HMCCode ([Mead et al., 2015](#)), the original implementation of the *Mead et al. model*, coded in Fortran.

The accuracy level goal is tuned according to the results achieved by the other codes. For the linear and non-linear power spectrum, as well as the angular power spectrum, we take as a reference CCL, which is compared to benchmarks targeting the accuracy for Stage III and Stage IV surveys. As we show in this paper, we reach the same level of accuracy as described in [Chisari et al., 2019](#). In Section 4.2 we give an overview of the cosmological observables implemented in PyCosmo. Section 4.2.5 describes how they are implemented, providing details concerning the code architecture. Information about the PyCosmo Hub is also provided. In Section 4.3 we describe the setup and the conventions used for code comparison and we present the main results from those tests.

4.2 COSMOLOGICAL MODEL

In this section we give definitions for the cosmological models implemented in PyCosmo. The current version of the code supports a Λ CDM cosmology, defined in terms of the matter density components Ω_b and Ω_m , the Hubble parameter H_0 , spectral index n_s , normalization of the density fluctuations σ_8 and a dark energy model with equation-of-state $w = -1$. The curvature is defined by $\Omega_k = 1 - \sum_i \Omega_i$, where i refers to matter (Ω_m), radiation (Ω_r) and vacuum (Ω_Λ) density components.

4.2.1 BACKGROUND

Background computations start with the calculation of the Hubble parameter, $H(a)$, and the cosmological distances. The basis of such calculations is the Friedmann equation, obtained by applying the Einstein's equations to the FLRW metric:

$$\left(\frac{1}{a} \frac{da}{dt}\right)^2 = \frac{8\pi G}{3} \rho + \frac{(1 - \Omega)H_0^2}{a^2}. \quad (4.1)$$

In this equation G is the Newton's constant, ρ is the total energy density and Ω is the sum of matter, radiation and vacuum densities expressed in units of critical density, ρ_c , as follows:

$$\Omega \equiv \Omega_m + \Omega_r + \Omega_\Lambda, \text{ where } \Omega_i \equiv \rho_i / \rho_c. \quad (4.2)$$

The critical density is defined as $\rho_c \equiv \frac{3H_0^2}{8\pi G}$, where H_0 is the present value of the Hubble parameter: $H_0 \equiv 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}$. The Hubble parameter, in turn, parametrises the expansion rate of the Universe:

$$\frac{H}{H_0} = [\Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda]^{\frac{1}{2}}. \quad (4.3)$$

Cosmological distances contribute to the computation of observables, so we need accurate predictions for those. A first comoving distance is the *comoving radius*, χ . Out to an object at scale factor a (or, equivalently, at redshift $z = (1/a) - 1$) it is defined as follows:

$$\chi(a) = \int_a^1 \frac{da'}{a^2 H(a')}. \quad (4.4)$$

Using the comoving radius PyCosmo evaluates the *comoving angular diameter distance*, r , as:

$$r(\chi) = \begin{cases} R_0 \sinh(\frac{\chi}{R_0}), & \text{open} \\ \chi, & \text{flat} \\ R_0 \sin(\frac{\chi}{R_0}), & \text{closed,} \end{cases} \quad (4.5)$$

where R_0 is the present value scale radius. The scale radius is defined as $\frac{R}{R_0} = a = (1 + z)^{-1}$ and $R_0 = \frac{c}{\kappa H_0}$, where c is the speed of light and κ is defined as follows:

$$\kappa^2 = \begin{cases} 1 - \Omega, & \text{open} \\ 1, & \text{flat} \\ \Omega - 1, & \text{closed.} \end{cases} \quad (4.6)$$

The *comoving angular diameter distance* is related to the *angular diameter distance*, D_A , and the *luminosity distance*, D_L , according to $D_A = a^2 D_L = ar(\chi)$. The luminosity distance, in turn, is used to compute the distance modulus, $\mu = 5 \log_{10}(D_L/pc) - 5$.

4.2.2 LINEAR PERTURBATIONS

GROWTH OF PERTURBATIONS

PyCosmo computes the linear growth factor of matter perturbations, $D(a)$, observing that for sub-horizon modes ($k \gg \eta^{-1}$) and at late times ($a \gg a_{eq}$), we can derive, from the Einstein-Boltzmann equations:

$$\frac{d^2 \delta_m}{da^2} + \left(\frac{d \ln H}{da} + \frac{3}{a} \right) \frac{d \delta_m}{da} - \frac{3 \Omega_m H_0^2}{2 a^5 H^2} \delta_m = 0. \quad (4.7)$$

Then the growth factor is computed by integrating the differential equation, and normalised so that $D(a) = a$ in the matter dominated case and $D(a) = 1$ when $a = 1$. Another approach to compute the linear growth factor is implemented in PyCosmo and makes use of hypergeometric functions. This formalism is valid for Λ CDM only (Hamilton, 2001; Matsubara, 1995). In Section 4.3 we will show the results of the code comparison using both methods.

LINEAR MATTER POWER SPECTRUM

Theoretical predictions for cosmological observables require knowledge of the matter distribution in the Universe, both at small and large scales. Given the matter density field, ρ , we can write it in terms of its mean matter density, $\bar{\rho}(t)$, and the statistical matter density perturbations:

$$\delta(\mathbf{x}, t) = \frac{\rho(\mathbf{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)}. \quad (4.8)$$

We are interested in the Fourier space overdensity, $\delta(\mathbf{k})$, which is the Fourier transform of the density fluctuations. The power spectrum, $P(k)$, is given by the average of overdensities in Fourier-space:

$$\langle \tilde{\delta}(\mathbf{k}) \tilde{\delta}(\mathbf{k}') \rangle = (2\pi)^3 P(k) \delta^3(\mathbf{k} - \mathbf{k}'), \quad (4.9)$$

where δ^3 is the Dirac delta function.

In addition to the Boltzmann Solver solution for the linear power spectrum, other approaches are used, typically based on numerical simulations. In this context approximate functions have been proposed. The fitting functions implemented in PyCosmo for the linear power spectrum are the *Eisenstein & Hu*, described in [Eisenstein and W. Hu, 1998](#), and a polynomial fitting function, namely *BBKS* ([Peacock, 1997](#)).

4.2.3 NON-LINEAR PERTURBATIONS

As briefly described above, on large scales (small k) the power spectrum can be calculated from linear perturbation theory. On small scales, evolving structures in the Universe become non-linear and perturbation theory breaks down. In analogy to the approximate functions for the linear power spectrum, also the non-linear power spectrum can be computed using fitting functions, following the same approach based on numerical simulations. A recently developed method, described in [Bartelmann, Fabis, Berg, et al., 2016](#) and [Bartelmann, Fabis, Kozlikin, et al., 2017](#), proposes the prediction of the non-linear power spectrum without using N-body simulations, but through non-perturbative analytical computation. Future code developments will also explore this analytical approach. We describe below the two non-linear fitting functions implemented in PyCosmo, *Halofit* ([Smith et al., 2003](#); [Takahashi et al., 2012](#)) and the model proposed in [Mead et al., 2015](#) and originally implemented in the *HMCode*.

NON-LINEAR POWER SPECTRUM

The *Halo Model* describes the dark matter density field as a superposition of spherically symmetric haloes, with mass function and internal density structure derived from cosmological simulations. The power spectrum can be written as:

$$P(k) = P_{1H}(k) + P_{2H}(k), \quad (4.10)$$

where $P_{1H}(k)$ and $P_{2H}(k)$ are denoted the *one-halo* and *two-halo term*, respectively. The first relates to the profile of the spherical haloes, while the second accounts for their spatial distribution, considering that their positions are correlated. For more details concerning the *Halo Model* we refer the reader to [Peacock and Smith, 2000](#); [Uroš Seljak, 2000](#); [Cooray and Sheth, 2002](#). The non-linear fitting functions implemented in PyCosmo are described below.

HALOFIT Predictions for the non-linear matter power spectrum, following the fitting function *Halofit* ([Smith et al., 2003](#)) and its revisions presented in [Takahashi et al., 2012](#), are both implemented in PyCosmo. Both papers propose the formalism described in eq.

4.10, where each term is a parametric function. The revised model provides updated fitting parameters, based on more accurate simulations.

MEAD ET AL. MODEL PyCosmo includes a first *Python* implementation of a revised version of the *Halo Model*, to which we already referred as the *Mead et al. model* (Mead et al., 2015), originally implemented in the HMCcode. In this model physically-motivated new parameters are added to the *Halo Model* formalism, in particular a smoothing parameter between the *one-halo* and the *two-halo* terms, and further parameters used to describe the effects of baryonic feedback on the power spectrum. The latter are found from a set of high-resolution N-body simulations and from OWLS hydrodynamical simulations which investigate the effect of baryons. As in the original HMCcode, three different models accounting for baryons are available: a more general model including prescriptions for gas cooling and heating, star formation and evolution and supernovae feedback, called REF; a model which adds to REF the AGN feedback, called AGN; and a model which is similar to REF, called DBLIM, which includes a more complete treatment of the supernovae feedback, described in van Daalen et al., 2011. For more detailed information about these models and how they are defined in the HMCcode, we refer the reader to Mead et al., 2015. In terms of computational speed, part of the PyCosmo code has been implemented in cython to speed up the computations. PyCosmo and the HMCcode run at comparable speeds.

4.2.4 OBSERVABLES

ANGULAR POWER SPECTRUM WITH THE LIMBER APPROXIMATION

Many observables in cosmology are expressed in terms of angular correlation functions of random fields, or their spherical harmonic transform, the angular power spectrum. Its calculation gives expressions including several integrals which require numerical evaluation. In order to simplify them, we can use approximation methods, such as the *Limber Approximation* (Limber, 1953; N. Kaiser, 1992; Nick Kaiser, 1998; Loverde and Afshordi, 2008). This prescription is implemented in PyCosmo. In particular, the weak lensing shear power spectrum is expressed as:

$$C_\ell = \frac{9}{16} \left(\frac{H_0}{c} \right)^4 \Omega_m^2 \int_0^{\chi_h} d\chi \left[\frac{g(\chi)}{ar(\chi)} \right]^2 P \left(\frac{l}{r}, \chi \right), \quad (4.11)$$

where χ is the comoving distance and χ_h the comoving distance to the horizon. $g(\chi)$ is the lensing radial function, which is defined in terms of $p_\chi(\chi)$, the probability of finding a galaxy at a comoving distance χ :

$$g(\chi) = 2 \int_{\chi}^{\chi_h} d\chi' p_{\chi}(\chi) \frac{r(\chi)r(\chi' - \chi)}{r(\chi')}, \quad (4.12)$$

where $p_{\chi}(\chi)$ is normalised as $\int d\chi p_{\chi}(\chi) = 1$.

In this work we use the lensing power spectrum, $C_{\ell}^{\gamma\gamma}$, as an example of observable.

LINE-OF-SIGHT INTEGRALS

The Boltzmann Solver includes a first python implementation of the Cosmic Microwave Background (CMB) angular power spectrum (C_{ℓ}^{TT}), using the line-of-sight integration. In this method, described in detail in [Uros Seljak and Zaldarriaga, 1996a](#), the temperature field is a time integral over the product of a source term and a spherical Bessel function, therefore splitting between the dynamical and geometrical effects on the anisotropies. The source function, which can be computed semi-analytically, is defined as follows:

$$S(k, \eta) = g \left(\Theta_0 + \Psi + \frac{\dot{u}_b}{k} + \frac{\Pi}{4} + \frac{3\ddot{\Pi}}{4k^2} \right) + \dot{g} \left(\frac{u_b}{k} + \frac{6\dot{\Pi}}{4k^2} \right) + \ddot{g} \left(\frac{3\Pi}{4k^2} \right) + e^{-\tau} (\dot{\Psi} - \dot{\Phi}), \quad (4.13)$$

where $g(\eta)$ is the visibility function, defined in terms of the optical depth as $g(\eta) = -\dot{\tau}e^{-\tau}$. The terms in $\Theta_0 + \Psi$, u_b and Π are the Sachs-Wolfe, Doppler and polarization terms, respectively, while the $\dot{\Psi} - \dot{\Phi}$ term describes the Integrated Sachs-Wolfe effect.

The temperature field is computed along the line of sight as:

$$\Theta_{\ell}(k, \eta) = \int_0^{\eta_0} d\eta S(k, \eta) j_{\ell}[k(\eta_0 - \eta)], \quad (4.14)$$

where $j_{\ell}(\eta)$ is the spherical Bessel function of ℓ order. The temperature field, normalized to the density perturbations for dark matter at present time (δ_0), is integrated over the wave-numbers to get the angular power spectrum:

$$C_{\ell}^{TT} = \frac{2}{\pi} \int dk k^2 P(k) \left| \frac{\Theta_{\ell}(k)}{\delta_0(k)} \right|^2, \quad (4.15)$$

where $P(k)$ is the linear power spectrum computed at present time.

4.2.5 ARCHITECTURE

The flow chart in [Figure 4.1](#) shows the code architecture. After instantiating `PyCosmo`, the user can set the cosmology through a set-function which or, equivalently, an internal configuration file. The latter can be modified also to choose the method to compute the matter power spectra. The *Background* class computes basic background quantities, such as the Hubble parameter and comoving distances. The *Linear Perturbations* class

provides the linear power spectrum either through the Boltzmann Solver or through fitting functions. The output is then used to compute the non-linear power spectrum in the *Non-linear Perturbations* class. In turn, this module offers a choice of different fitting functions. The power spectrum is involved in computing the observables by the class *Observables*. The theoretical models implemented in this routine are described in Section 4.2.

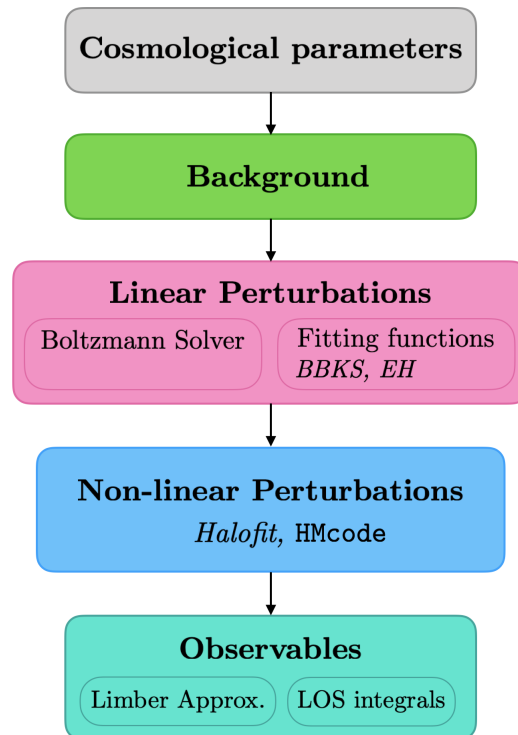


Figure 4.1: Flow-chart summarizing the PyCosmo architecture. From the top: *Cosmological parameters* refers to the initial cosmological setup, which affects all the computations. The *Background* class computes the Hubble parameter and comoving distances. It is followed by the *Linear Perturbations* and *Non-linear Perturbations* modules, which include various methods to compute matter power spectra (described in Section 4.2). The *Observables* module at the end of the chart calls all the other modules before.

4.2.6 SYMBOLIC CALCULATIONS

As shown in the flow-chart in Figure 4.1, one of the classes implemented in PyCosmo provides solutions to the set of Einstein-Boltzmann equations, which govern the linear evolution of perturbations in the Universe. The novelty of this solver is its approach to the equations themselves, which are symbolically represented through the *Python* package *Sympy*. The symbolic representation provides the user a convenient interface to manipulate the equations and implement new cosmological models. The equations are

4.2. Cosmological model

then simplified by a C++ code generator before being evaluated. For more details about how the solver computes a numerical solution for them, we refer the user to a previous work, [A. Refregier, Gamper, et al., 2018b](#), which focusses on the PyCosmo Boltzmann solver.

4.2.7 UNIT TESTS

Each class shown in Figure 4.1 is associated with a unit-test routine. It consists in a series of functions testing the methods implemented in each class. These tests perform code-comparison between PyCosmo and the other codes, and check whether the agreement passes a certain numerical accuracy. Every time the code is updated, the developer can check through unit-tests also the impact the new implementations might have on pre-existing parts of the code. The analysis presented later in Section 4.3 shows the results of code-comparison which is incorporated in the unit-tests. The *coverage* refers to the amount of code tested and validated in each module through unit-tests. With reference to Figure 4.1, currently the PyCosmo modules have the following coverage: 100% for the *Background* class, 97% for *Linear Perturbations*, 96% for *Non Linear Perturbations* and 96% for the *Observables*. Incomplete coverage is due to self-referenced functions, intended as those providing intermediate results called by main functions which are successfully tested.

4.2.8 PYCOSMO HUB

PyCosmo is conceived as a multi-purpose cosmology calculation tool in *Python*, and designed to be interactive and user-friendly. As discussed above, the usage of the *Sympy* package is part of this concept. Indeed, PyCosmo is user-friendly not only in its numerical implementation, but also in terms of its public interface: in order to make its usage immediate to the user, we make PyCosmo publicly available on a hub platform, called PyCosmo Hub (see a screenshot in Fig.4.2). Its current version, accessible from this link, <https://pycosmohub.phys.ethz.ch/hub/login>, includes Jupyter tutorial-notebooks illustrating how to use the code and shows the results of the code-comparison analysis through a series of static notebooks. The hub currently hosts the most recent versions of the codes CLASS and iCosmo, which can be run by the users. The iCosmo code, originally written in IDL language, is interpreted on the hub through GDL, an open source library alternative to IDL. The PyCosmo version installed on the hub can be downloaded via *pip*. Further information about the code release and documentation is available on this web page: (<https://cosmology.ethz.ch/research/software-lab/PyCosmo.html>). The users accessing the hub have space to write their own notebooks, make their own calculations and save the results locally, without the need of installing any software. In this context, the hub is conceived to be useful both for educational purposes and for

promoting cosmological inferences in the cloud, in a new dynamic way of teaching and doing research.

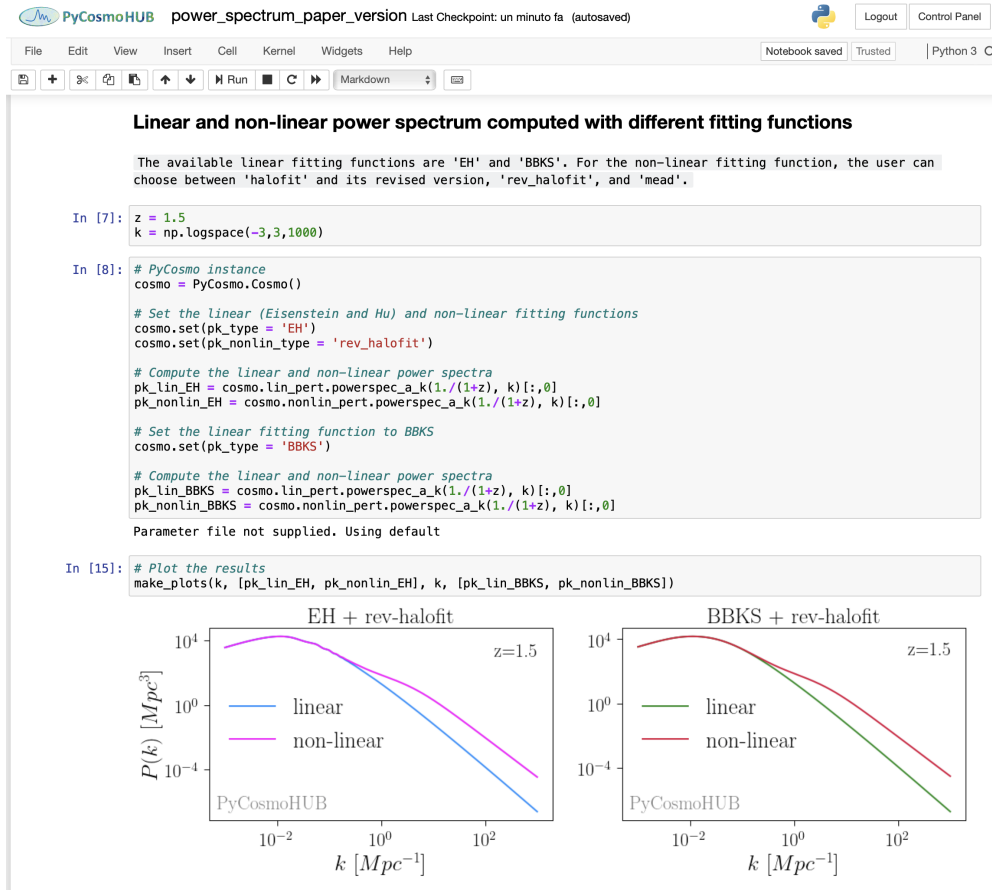


Figure 4.2: A screenshot of a Jupyter Notebook running on the PyCosmo Hub.

4.3 VALIDATION AND CODE COMPARISON

In order to assess the level of accuracy in the computation of cosmological observables, PyCosmo monitors its own predictions internally and making comparisons with other cosmology codes. The reliability of every function in PyCosmo is checked through *unit tests*, described in Section 4.2.7. In this section, we show the main results from those tests: overall we obtain a good agreement between the codes, both using a fiducial cosmology and testing their response by varying the cosmological setup. We compare the algorithms also in terms of execution speed, with the result that PyCosmo runs at a speed comparable with the other codes.

4.3.1 COSMOLOGICAL SETUP AND CONVENTIONS

The tests performed to assess the agreement between the codes are of two kinds, either referring to a fiducial cosmological setup or testing the robustness of the code to changes of cosmological parameters. We assume as our fiducial cosmology: $h = 0.7$, $\Omega_m = 0.3$, $\Omega_b = 0.06$, $\Omega_c = 0.24$, $n_s = 1$, $\sigma_8 = 0.8$, $N_{\text{eff}} = 3$. We vary cosmology in ranges of h and Ω_m : $h \in [0.4, 0.9]$, $\Omega_m \in [0.2, 0.7]$, and we produce heatmaps to show the agreement between the codes across the (h, Ω_m) parameter space. In this section we include the heatmaps only for the background computations and for the linear and non-linear power spectra, showing those for the other classes in Section B.2 of Appendix B.

To illustrate trends as a function of redshift in our fiducial cosmology, for instance in terms of background quantities (cosmological distances, linear growth factor), we consider a redshift range of $z \in [0, 9.5)$ with 5000 grid points. If we vary the cosmological parameters, we consider redshift in the range $[0, 6)$, maintaining the same number of points. When we compare the non-linear power spectrum to the HMcode, we compute it as a function of wavenumbers, k , logarithmically spaced between 10^{-3} and $10^4 Mpc^{-1}$, with a total of 200 points. When we compare the power spectra predicted by different codes we use 200 wavenumbers logarithmically spaced between 10^{-3} and $10^3 Mpc^{-1}$, which is the sampling used by default in iCosmo. Testing the angular power spectrum, we choose a sample of multipoles, ℓ , linearly spaced between 10 and 10^4 , following also in this case the convention adopted in iCosmo.

In each test, the setup described above is matched between the codes, but there are further parameters which need special care in order to make consistent tests. A detailed description of their configuration is given in Section B.1 of Appendix B.

In the next paragraphs, we show the results of the code comparisons. The achieved accuracy is quantified in terms of the relative difference between two compared quantities (i.e. distances, power spectra etc.). Given Q a certain cosmological quantity we consider for comparison between PyCosmo and a code C , the accuracy is defined as follows:

$$A = \frac{|Q_{\text{PyCosmo}} - Q_C|}{Q_{\text{PyCosmo}}}, \quad (4.16)$$

and it is always reported in logarithmic scale. A is a vector including as many points as the two compared quantities. In the heatmaps summarizing the results when varying cosmology, each cell refers to a particular combination of (h, Ω_m) . It is colour-coded by the base-10 logarithm of the maximum accuracy ($\text{Log}[MAX(A)]$) and labelled by the dispersion in accuracy ($\sigma(A)$) obtained for the specific cosmological setup it represents. We structure our analysis as follows: we start with the background quantities, testing

the computation of the cosmological distances. We then proceed with the linear perturbations, discussing the level of agreement reached in terms of the linear growth factor and the linear power spectrum. We move to the non linear perturbations showing the comparisons in terms of the non-linear power spectrum. We conclude with the observables, including the weak lensing and the CMB angular power spectra. We choose this ordering to emphasize the fact that each step, from the background computations to the linear and non-linear perturbations and up to the observables, influences the accuracy reached in the calculation which comes next. We summarize this procedure and the main results later in Table 4.1, which gives an overview of the cosmological quantities which can be computed, the settings used for the comparisons and the level of achieved accuracy.

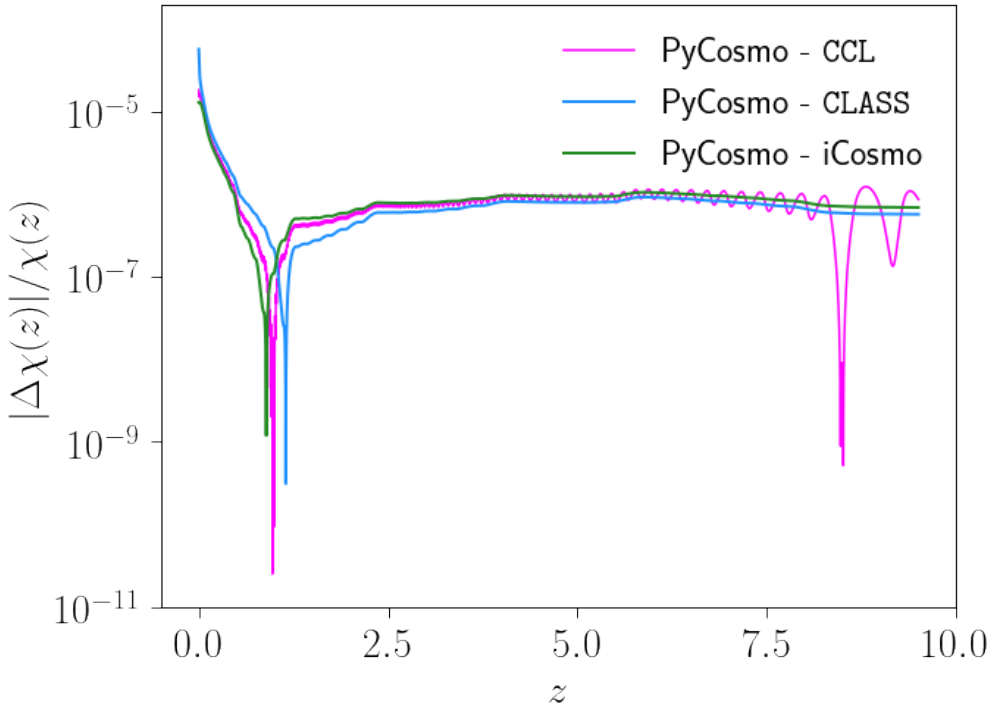


Figure 4.3: Comparison between PyCosmo, CCL, CLASS and iCosmo in terms of comoving radius, $\chi(z)$, for the assumed fiducial cosmology. The test produces an overall accuracy about 10^{-5} .

4.3.2 BACKGROUND

Figure 4.3 summarizes the results of a code comparison made in terms of comoving radius, χ , defined in Eq.4.4. The y -axis shows the relative difference between PyCosmo and the other codes, normalised to PyCosmo (see Eq.4.16), as a function of redshift, z , up to redshift $z = 10$. An overall accuracy around 10^{-6} is observed, with oscillations

4.3. Validation and code comparison

between 10^{-9} and 10^{-5} at lower redshifts. We repeat the same test by varying cosmology, as shown in Figure 4.4. As explained in the paragraph 4.3.1, the heatmaps are colour-coded by the maximum relative difference occurring between PyCosmo and iCosmo (left panel), PyCosmo and CCL (central panel) and PyCosmo and classy (right panel). Each cell, referring to a combination of (h, Ω_m) , is labelled by the value of dispersion in relative difference obtained for that particular cosmological setup. All the results are expressed in logarithmic scale. Overall we can reach an agreement better than about 10^{-4} , with small dispersion (up to $\sim 10^{-6}$) overall.

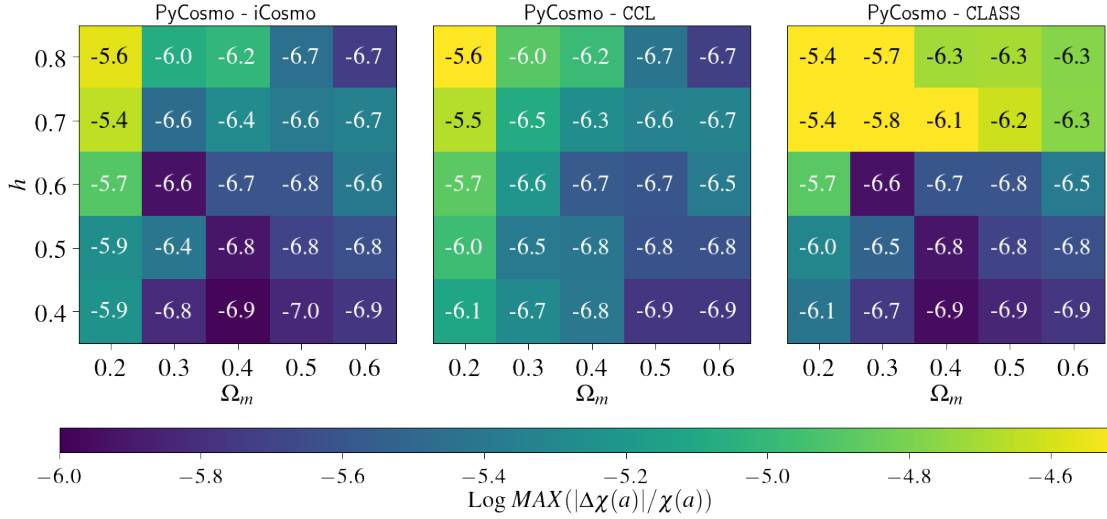


Figure 4.4: Comparison between PyCosmo and iCosmo (left panel), PyCosmo and CCL (central panel) and PyCosmo and classy (right panel) in terms of comoving radius, $\chi(z)$, for a variety of cosmological parameter values. In the heatmaps each cell refers to a specific combination of (h, Ω_m) . As described in paragraph 4.3.1, it is color-coded by the maximum accuracy reached in the comparison, and labelled by the dispersion in accuracy. All the results are expressed in logarithmic scale.

4.3.3 LINEAR PERTURBATIONS

Next, we test the linear perturbations both in terms of the growth factor and the linear power spectrum. In Fig.4.5 we show the results of the code comparison in terms of the linear growth factor, $D(a)$, computed for our fiducial cosmology and with the same settings described in detail in the paragraph 4.3.1 above. Fig.B.1 shows the outcome of the same test, but varying cosmological parameters. All the results are displayed in logarithmic scale. Overall the codes are in agreement, plus we notice a difference between the results obtained by comparing PyCosmo to iCosmo (10^{-7}) and PyCosmo to CCL and CLASS (10^{-3}). This might be due to the different numerical implementations of the algorithm, which have been discussed already in Section 4.1 of Chisari et al., 2019. As

a further test we show the comparison in terms of the hypergeometric growth factor, which offers an analytical reference under the assumption of suppressed radiation. In this test, the dashed lines show the comparison between the hypergeometric growth factor computed in PyCosmo and the integrated growth factor computed with iCosmo, CCL and CLASS. We observe an order of magnitude improvement in the achieved accuracy, as also summarised by the heatmap in Fig.B.2.

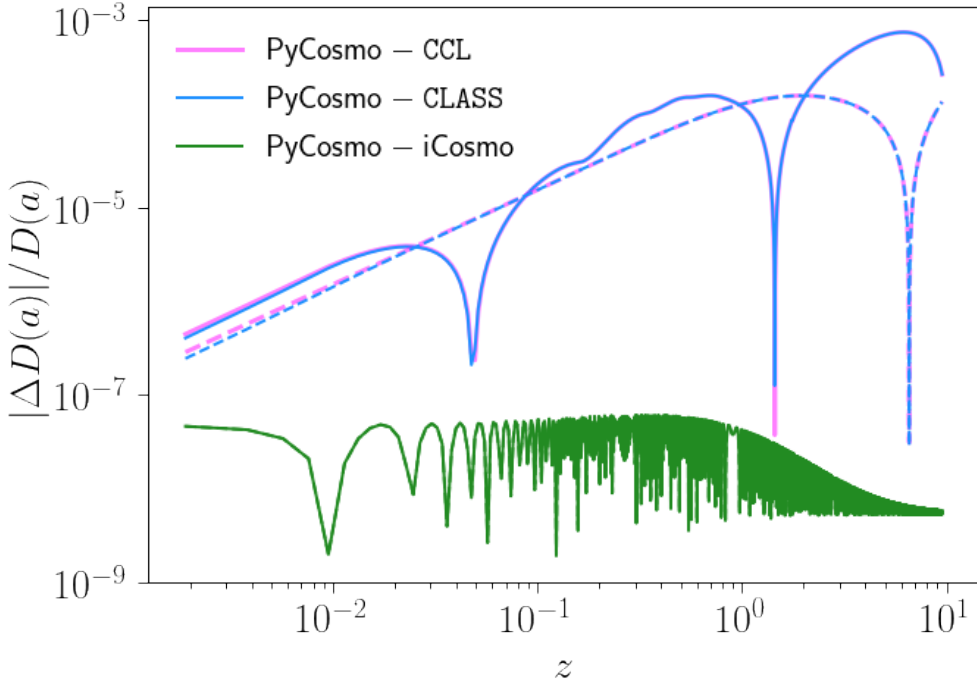


Figure 4.5: Code comparison in terms of the growth factor, for our fiducial cosmology. The values used for integration accuracy on the ODEINT solvers are specified in detail in Section B.2 of Appendix B. PyCosmo, CCL and CLASS agree to better than 10^{-3} . The lines showing the comparison with CCL and CLASS overlap. The dashed lines show the comparison of the hypergeometric growth factor computed in PyCosmo to the integrated growth factor computed with iCosmo, CCL and CLASS. The dashed and the solid lines for the comparison with iCosmo overlap. The agreement between PyCosmo and iCosmo reaches 10^{-7} .

We compute the linear power spectrum both using the *EH* and *BBKS* fitting functions, shown in Fig.4.6 with solid and dashed lines, respectively. We compare PyCosmo to iCosmo on the left panel and to CCL on the right panel. In both cases the linear power spectrum is computed for our fiducial cosmology and at three different values of redshift, using the same settings described in paragraph 4.3.1. Overall we reach a good agreement. The level of accuracy is dominated by the growth factor, whose error propagates into the power spectrum, up to 10^{-7} for iCosmo and 10^{-3} for CCL, as already shown in Fig.4.5,

4.3. Validation and code comparison

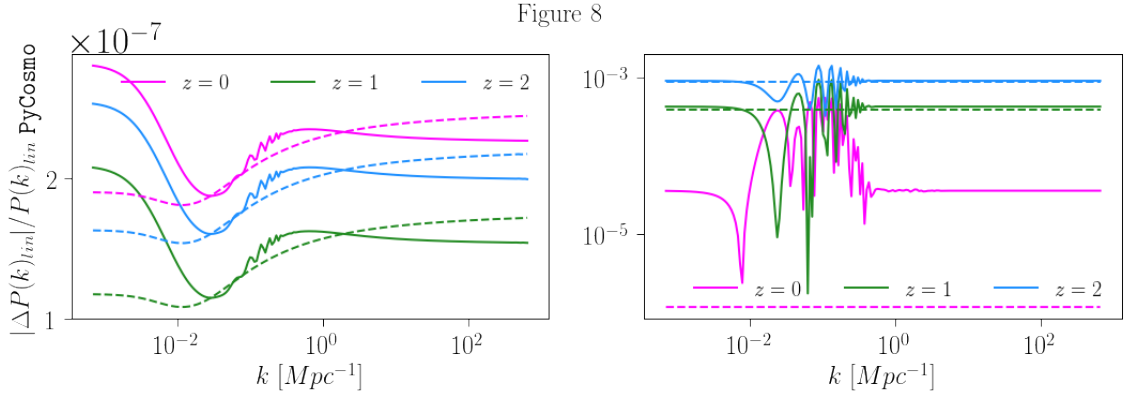


Figure 4.6: Comparison of PyCosmo with iCosmo (left panel) and CCL (right panel) in terms of linear power spectrum computed with the *EH* (solid lines) and the *BBKS* (dashed lines) fitting functions, for three different redshifts. The y -axis on the left panel is not displayed in logarithmic scale for a better visualization.

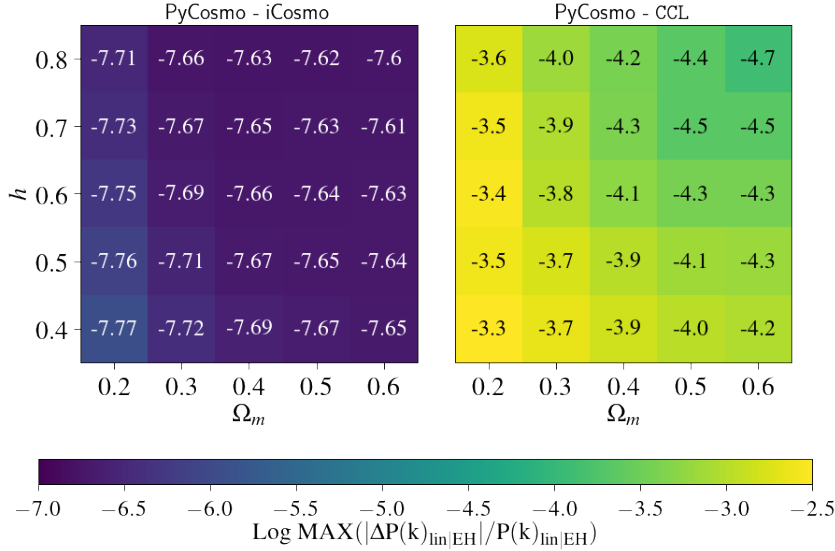


Figure 4.7: Comparison between PyCosmo and iCosmo (left panel) and between PyCosmo and CCL (right panel) in terms of linear matter power spectrum computed with the *EH* fitting function.

and increases with time. As observed in the heatmaps in Figures 4.7 and 4.8, the same level of accuracy is reached when we vary cosmology. The heatmaps are colour-coded and labelled with the same convention used in Fig.4.4 and described in paragraph 4.3.1. A good agreement is also observed between PyCosmo and `classy` when we compare the linear power spectra computed with their respective Boltzmann solvers. Fig.4.9 shows their relative difference at redshift $z=1$ for our fiducial cosmology. We ran `classy` using the same settings listed in the its high-accuracy precision file `pk_ref.pre` (available in the public distribution of CLASS), and PyCosmo with $l_{max} = 100$, $\epsilon = 3 \cdot 10^{-7}$ and

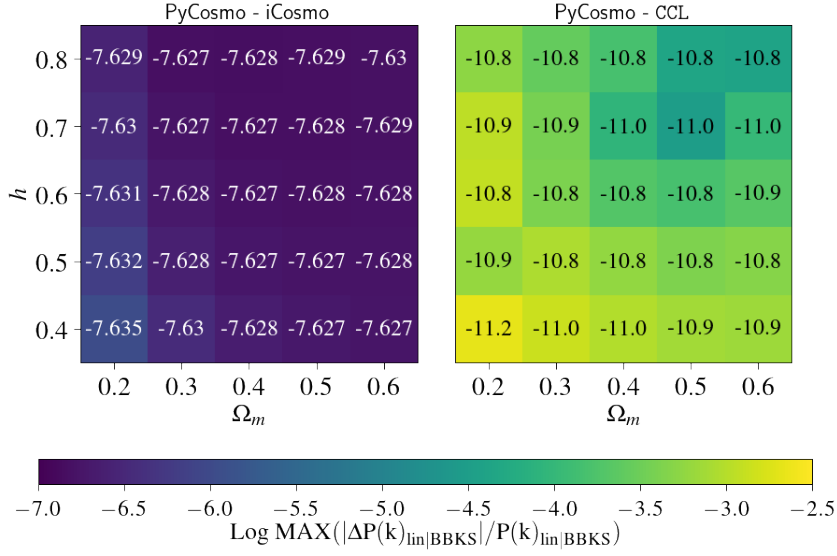


Figure 4.8: Comparison between PyCosmo and iCosmo (left panel) and between PyCosmo and CCL (right panel) in terms of the linear matter power spectrum computed with the *BBKS* fitting function.

$dt = 10^{-5}$. We reach an agreement better than about 10^{-3} .

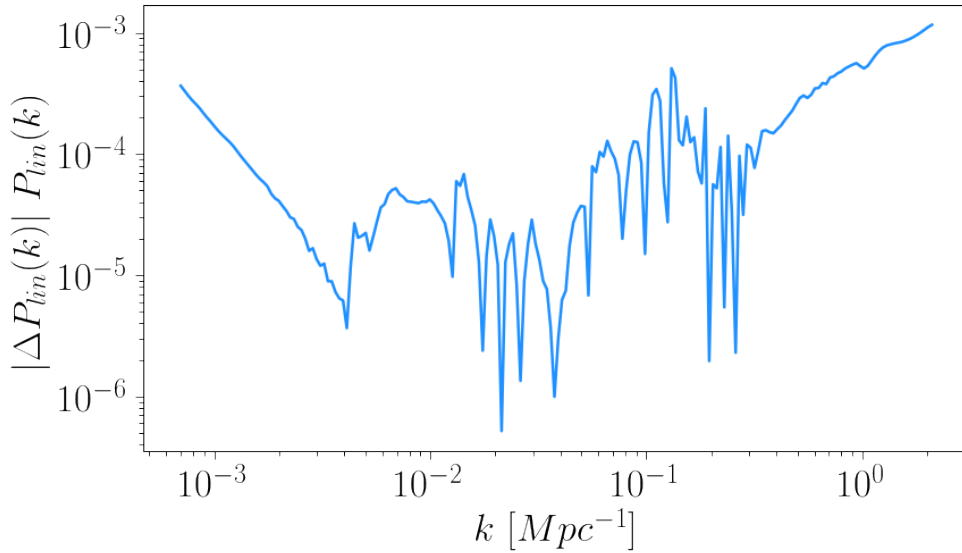


Figure 4.9: Comparison between PyCosmo and classy in terms of the linear power spectrum computed with the Boltzmann solver. The power spectrum is shown at redshift $z = 1$. Both codes were run using high-accuracy settings (described more in detail in Section 4.3.3). A good agreement up to 10^{-3} is reached overall.

4.3.4 NON-LINEAR PERTURBATIONS

The accuracy for non-linear perturbations is assessed in terms of the non-linear matter power spectrum and is reported in Fig.4.10. In comparing PyCosmo to iCosmo (dashed lines) and to CCL (solid lines), we consider the combinations of non-linear and linear fitting functions which are available in the codes. Therefore we show the following tests:

- we compare PyCosmo and iCosmo in terms of non-linear power spectrum computed with the *Halofit* fitting function by [Smith et al., 2003](#). The linear fitting function used is either EH (left panel) or BBKS (right panel).
- PyCosmo and CCL are compared in terms of non-linear power spectrum computed with the *Halofit* fitting function by [Takahashi et al., 2012](#). Also in this case, the linear fitting function used is either EH (left panel) or BBKS (right panel).

We observe that PyCosmo and iCosmo can reach an agreement between 10^{-7} and 10^{-4} . The agreement with CCL, as already observed for the linear power spectrum, is dominated by the growth factor. We obtain analogous results when we vary the cosmological model, as shown in the heatmap of Fig.4.11: overall the codes are in good agreement, and the algorithm is stable across the parameter space. These observations are valid in both choices of linear fitting functions.

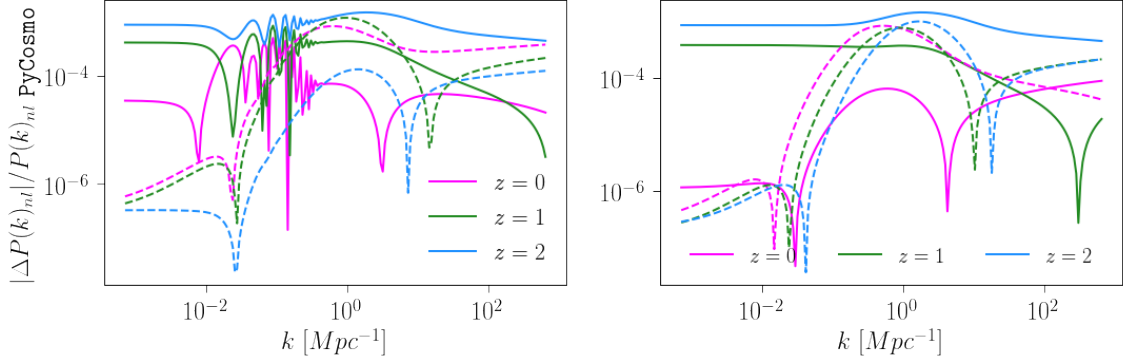


Figure 4.10: Relative difference in terms of the non-linear matter power spectrum between PyCosmo and iCosmo (dashed lines) and CCL (solid lines). In the comparison between PyCosmo and iCosmo we consider the *Halofit* fitting formula by [Smith et al., 2003](#). The test between PyCosmo and CCL accounts for its revision by [Takahashi et al., 2012](#)). The linear fitting formulas used in the computation follow the *EH* and the *BBKS* prescriptions on the left and right panels, respectively.

Moving from *Halofit* to the *HMCode*, Fig.4.13 shows the comparison between its implementation in PyCosmo and the original *HMcode*, for our fiducial cosmology. The

non-linear power spectrum is computed assuming the *EH* linear fitting function. Overall, the computations have been made following the settings described in section 4.3.1. The left panel is dedicated to the dark-matter-only case and the agreement is studied at different redshifts. Concerning the observed difference, the contribution of the growth factor is not much relevant as the one observed in the comparison with CCL (see Figure 4.6), where there vertical shift is clearly due to the degrading accuracy observed for the growth factor as a function of redshift, which is shown in Figure 4.5. Differences in this case are mostly due to numerical discrepancies between the Python and the Fortran implementations. The results on the right panel take into account the baryonic feedback at redshift $z = 1$. In both cases we reach an overall accuracy better than about 10^{-3} . The models for baryonic feedback indicated in the legend are listed in Section 4.2. For more details about those and their differences, we refer the reader to Mead et al., 2015.

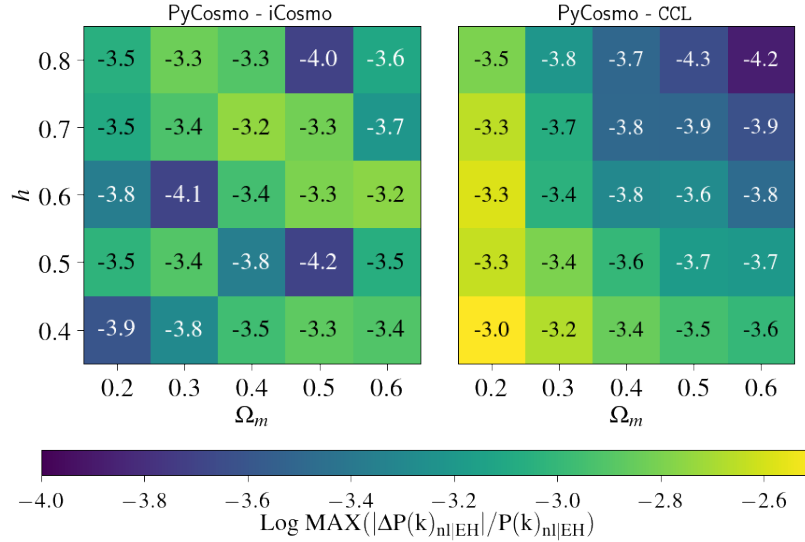


Figure 4.11: Comparison between PyCosmo and iCosmo (left panel) and between PyCosmo and CCL (right panel) in terms of the non-linear matter power spectrum, computed with *Halofit+EH*.

4.3.5 OBSERVABLES

We test the observables computed by PyCosmo in terms of the lensing power spectrum ($C_\ell^{\gamma\gamma}$) and the CMB angular power spectrum (C_ℓ^{TT}). Figure 4.14 shows the comparison to iCosmo (green lines) and CCL (magenta lines) for our fiducial cosmology, in terms of $C_\ell^{\gamma\gamma}$. The non-linear power spectrum involved in the calculation is computed with the *Halofit* fitting formula, combined with both EH (solid lines) and BBKS (dashed lines) fitting functions. The heatmaps in Figures B.3 and B.4 show the same test by varying the cosmological parameters. Overall we recover an accuracy up to $\sim 10^{-3}$ for iCosmo and

4.3. Validation and code comparison

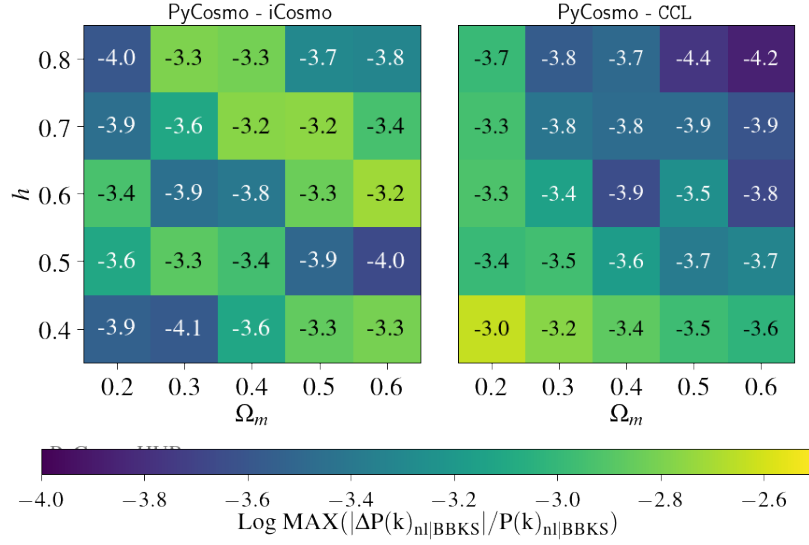


Figure 4.12: Comparison between PyCosmo and iCosmo (left panel) and between PyCosmo and CCL (right panel) in terms of the non-linear matter power spectrum, computed with *Halofit+BBKS*.

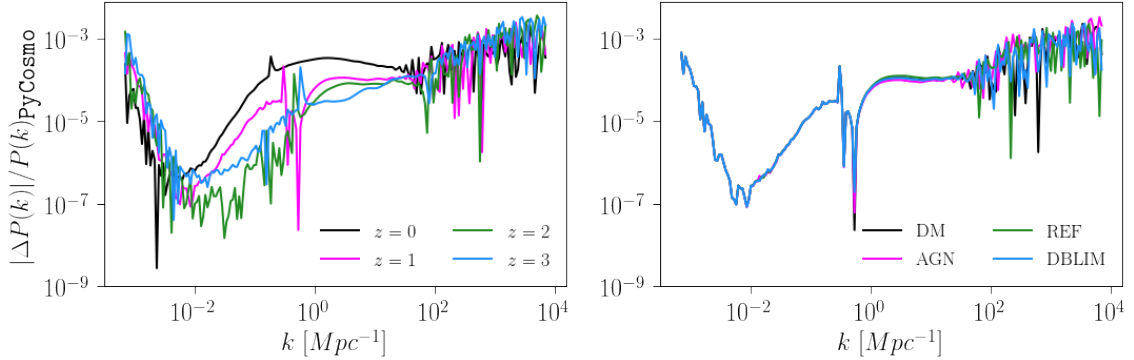


Figure 4.13: Comparison between the implementations of the *Mead Model* in PyCosmo and in the original code HMcode. Our fiducial cosmology is assumed as the cosmological setup and the non-linear power spectrum is computed assuming the *EH* linear fitting function. On the left panel the comparison is shown for the dark-matter-only case at different redshifts. On the right panel, we add the baryonic feedback for redshift $z = 1$.

at the percent level with CCL. The heatmap in Fig. B.5 shows the comparison between PyCosmo and CCL when the C_ℓ^γ are computed with a linear power spectrum, either using the EH or the BBKS fitting function. Also in this case we reach the same level of accuracy as in the previous test.

Fig.4.15 shows preliminary results from our first Python implementation of the C_ℓ^{TT} computed with the line of sight integration. The left panel shows the good agreement between the two Boltzmann Solvers, PyCosmo and classy. More details will be reported in a future paper describing the updates and the performance of the new version

of the PyCosmo Boltzmann Solver.

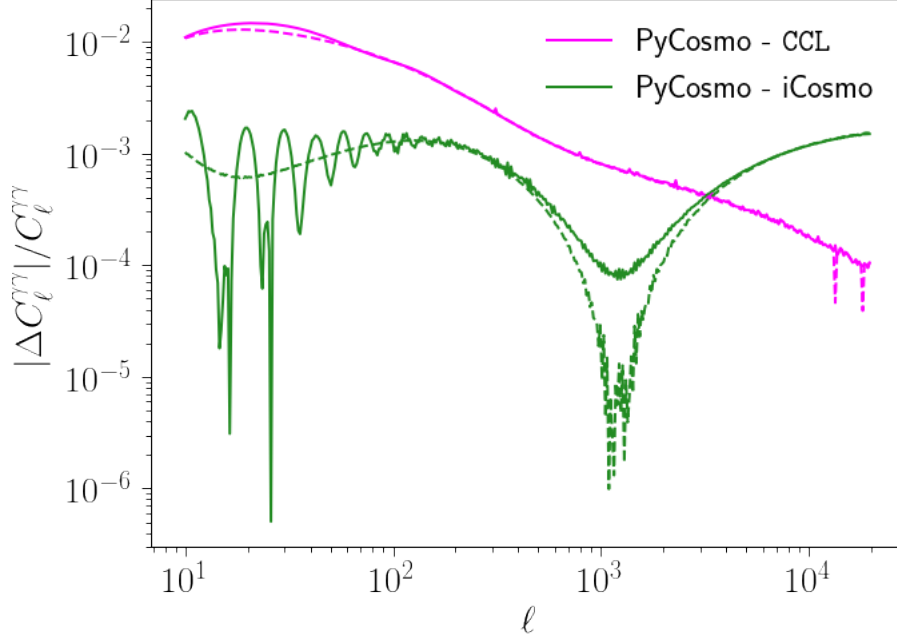


Figure 4.14: Comparison in terms of $C_\ell^{\gamma\gamma}$ between PyCosmo and iCosmo (green lines) and PyCosmo and CLL (magenta lines). *Halofit* and its revised version are used to compute the $C_\ell^{\gamma\gamma}$ in the two respective comparisons. *Halofit* is matched both with EH (solid lines) and BBKS (dashed lines) linear fitting functions.

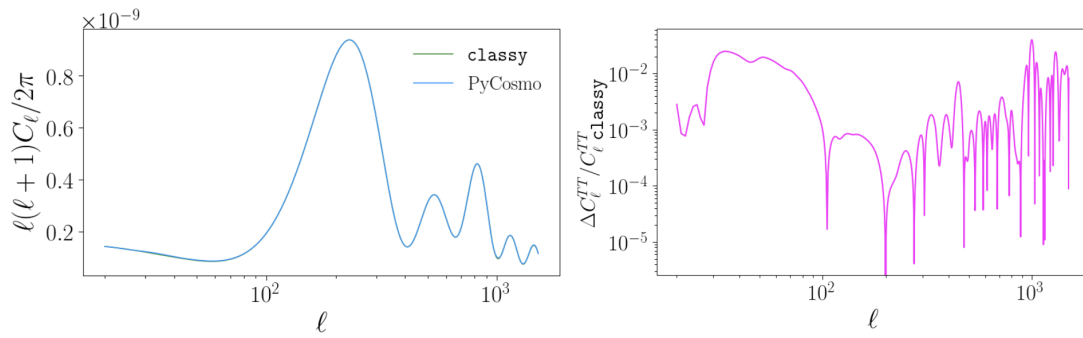


Figure 4.15: Preliminary CMB angular power spectrum computed with the PyCosmo Boltzmann Solver (on the left) and absolute difference with the same output from *classy* (left panel). All the terms entering the source function (Sachs-Wolfe, Doppler, Polarization and Integrated Sachs-Wolfe) are considered for the computation in both codes.

4.3.6 SUMMARY

Table 4.1 represents a summary of the code-comparison described in this paper. It shows the level of agreement between the codes reached in terms of background quantities, power spectra and observables. Each entry quantifies the agreement using the notation $\phi_{fid+\Delta_-}^{-\Delta_+}(\sigma_{fid})$, which is explained as follows.

We consider a certain observable, $Q(\mathbf{x})$, where \mathbf{x} can be, for instance, a collection of values in redshift or wavenumbers. When we run two different codes we get two independent samples of the same observable, $Q(\mathbf{x})$ and $Q'(\mathbf{x})$. For each code we compute their relative difference, expressed as $|Q(\mathbf{x}) - Q'(\mathbf{x})|/Q(\mathbf{x})$, and then extract the maximum relative difference, ϕ , and the dispersion, σ , of this distribution. We repeat the same computations N times, varying cosmological parameters. We get a collection of maximum relative differences, $\Phi = [\phi_1, \phi_2, \dots, \phi_{fid}, \dots, \phi_N]$ and the dispersions of their respective distributions, $\Sigma = [\sigma_1, \sigma_2, \dots, \sigma_{fid}, \dots, \sigma_N]$, where ϕ_{fid} and σ_{fid} refer to the values obtained for our fiducial cosmology. These values are expressed in logarithm base 10. From Φ we extract ϕ_{max} and ϕ_{min} , which represent the worst and the best agreement we could obtain by exploring the parameter space of cosmological parameters. In this context we have $|\phi_{min}| \geq \phi_{fid}$, $|\phi_{max}| \leq \phi_{fid}$.

In the notation used in the table Δ_- and Δ_+ are the distances between ϕ_{fid} and the worst and best agreement, respectively: $\Delta_- = \phi_{max} - \phi_{fid}$, $\Delta_+ = \phi_{min} - \phi_{fid}$. Therefore the notation $\phi_{fid+\Delta_-}^{-\Delta_+}(\sigma_{fid})$ gives the agreement and the dispersion obtained for our fiducial cosmology, plus the maximum and minimum agreement we get by varying cosmological parameters.

All the results reported in the table were obtained following the settings described in section 4.3.1, at redshift $z = 1$. For the comparison with the HMcode we show the maximum and minimum agreement at $z = 0$ for the fiducial cosmology, together with the dispersion accuracy for the fiducial cosmology. The same applies to the comparison with the software `classy` in terms of the CMB angular power spectrum.

The hyphenated entries symbolize cases where a certain computation is not available in one of the codes, so no comparison is currently possible.

4.4 CONCLUSIONS

PyCosmo is a recent python-based framework providing solutions for the Einstein-Boltzmann equations and making theoretical predictions for cosmological observables. In this paper, we first discuss its architecture and the implementation of cosmological observables, computed in terms of background quantities, linear and non-linear matter power spectra and angular power spectra (Section 4.2.5). In order to assess the accuracy of its predictions, PyCosmo is compared to other codes: (CCL), CLASS, HMCode

	iCosmo	CCL	CLASS	HMCode
Background				
$H(a)$	-15.2 (-16.1)	-4.9 (-5.6)	-8.1 (-8.7)	—
$\chi, D_L, D_A(a)$	$-5.5_{+0.9}^{-0.5}$ (-5.7)	$-5.1_{+0.6}^{-0.7}$ (-5.6)	$-4.3_{+0.2}^{-1.6}$ (-5.7)	—
Linear Perturbations				
$D(a)$	$-7.2_{0.0}^{-0.1}$ (-7.9)	$-3.5_{+0.7}^{-0.5}$ (-4.0)	$-3.5_{+0.7}^{-0.5}$ (-3.9)	—
$D(a)$ hyper	$-7.2_{0.0}^{-0.1}$ (-7.9)	$-3.8_{+0.7}^{-0.5}$ (-4.4)	$-3.8_{+0.7}^{-0.5}$ (-4.4)	—
$P(k)_{\text{lin BBKS}}$	$-6.8_{+0.8}^{+0.0}$ (-7.6)	$-3.4_{+0.7}^{-1.1}$ (-10.9)	—	—
$P(k)_{\text{lin EH}}$	$-6.7_{+0.8}^{+0.0}$ (-7.7)	$-3.0_{+0.6}^{-0.9}$ (-3.8)	—	—
$P(k)_{\text{lin boltz}}$	—	—	-2.9 (-4.1)	—
Non-linear Perturbations $P(k)_{\text{nl Halofit}}^{(*)}$				
BBKS + S.	$-3.1_{+0.4}^{-0.6}$ (-3.6)	—	—	—
EH + S.	$-2.9_{+0.2}^{-0.8}$ (-3.4)	—	—	—
BBKS + T.	—	$-3.4_{+0.8}^{-0.5}$ (-3.8)	—	—
EH + T.	—	$-3.1_{+0.7}^{-0.8}$ (-3.7)	—	—
HMCode + EH	—	—	—	[-7.6, -2.6] (-3.3)
Observables $C_\ell^{\gamma\gamma(*)}, C_\ell^{TT}$				
$C_\ell^{\gamma\gamma}$ BBKS	—	$-1.9_{+0.4}^{-0.4}$ (-2.3)	—	—
$C_\ell^{\gamma\gamma}$ EH	—	$-1.8_{+0.3}^{-0.4}$ (-2.3)	—	—
$C_\ell^{\gamma\gamma}$ S. + BBKS	$-2.8_{+0.2}^{-0.1}$ (-3.4)	—	—	—
$C_\ell^{\gamma\gamma}$ S. + EH	$-2.6_{+0.0}^{-0.3}$ (-3.3)	—	—	—
$C_\ell^{\gamma\gamma}$ T. + BBKS	—	$-1.9_{+0.4}^{-0.4}$ (-2.3)	—	—
$C_\ell^{\gamma\gamma}$ T. + EH	—	$-1.8_{+0.3}^{-0.5}$ (-2.3)	—	—
$C_\ell^{\gamma\gamma}$ HMCode + EH	—	—	—	—
C_ℓ^{TT}	—	—	[-5.7, -2.1](-2.7)	—
(*) S. = Halofit Smith, T. = Halofit Takahashi				

Table 4.1: Summary of the code-comparison between PyCosmo, iCosmo, CCL, CLASS and HMCode. The structure, from the background computations to the observables, follows the order schematically shown in Fig.4.1, highlighting the fact that the accuracy reached in each module propagates in the next one. Each cell quantifies this accuracy: we explain in detail the adopted notation in paragraph 4.3.6.

and iCosmo. Details about the codes and the setup used for the comparisons are given in Sections 4.1 and 4.3.1. The tests, performed by comparing the output of different and independent codes, and presented in Section 4.3, show that PyCosmo is in good agreement with the other codes over a range of cosmological models. It also includes a first *Python* implementation of the HMCode, which provides an accurate prediction for the non-linear power spectrum which can take into account baryonic effects. We release the currently tested and validated version of PyCosmo (without the Boltzmann Solver) and we make it available on an online platform called PyCosmo Hub (Section 4.2.5): <https://cosmology.ethz.ch/research/software-lab/PyCosmo.html>. On

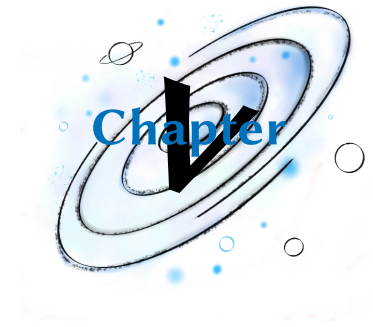
4.5. Acknowledgements

this hub the users can easily access and use PyCosmo without the need of installing the software locally. In this context, PyCosmo presents an easy and user-friendly interface which is accessible to everyone who wants to compute theoretical predictions for precision cosmology.

4.5 ACKNOWLEDGEMENTS

We would like to thank Elisabeth Krause for her constructive discussions about the matter power spectrum and the transfer functions implemented in CCL. We also thank Danielle Leonard, Elisa Chisari and Mustapha Ishak-Boushaki for their useful comments concerning CCL.

Applications, extensions and further developments



[...] Οὐτις ἐμοί γ' ὄνομα· Οὐτὶν δέ με κικλήσκουσι
μήτηρ ἠδὲ πατήρ ἠδ' ἄλλοι πάντες ἐταῖροι.
ὣς ἐφάμην, ὁ δέ μ' αὐτίκ' ἀμείβετο νηλεῖ θυμῶι·
Οὐτὶν ἐγὼ πύματον ἔδομαι μετὰ οἷς ἐτάροισιν,
τοὺς δ' ἄλλους πρόσθεν· τὸ δέ τοι ξεινήιον ἔσται.

'[...] Nobody is my name; Nobody do they call me,
my mother and my father, and all my comrades as well'.
So I spoke, and he answered me with pitiless heart:
'Nobody will I eat last among his comrades,
and the others before him; this shall be my warm gift for you'.

Ὅμηρος, Ὀδύσσεια, IX, 366-370

In this chapter, we will give an overview of the collaborative projects where the works presented above in this thesis found fields of application. More precisely, we will briefly describe how the tools and knowledge fulfilled in the context of the DES Y1 morphology catalogue (described in Chapter 2), as well as the tools developed with PyCosmo, gave substantial contributions to analyses carried on DES data. We will also discuss the work made to implement and test additional theoretical models and computational methods in PyCosmo. Finally, we will show and discuss the results of a likelihood analysis conducted on DES Y1 and Planck data using accurate theoretical predictions provided by PyCosmo.

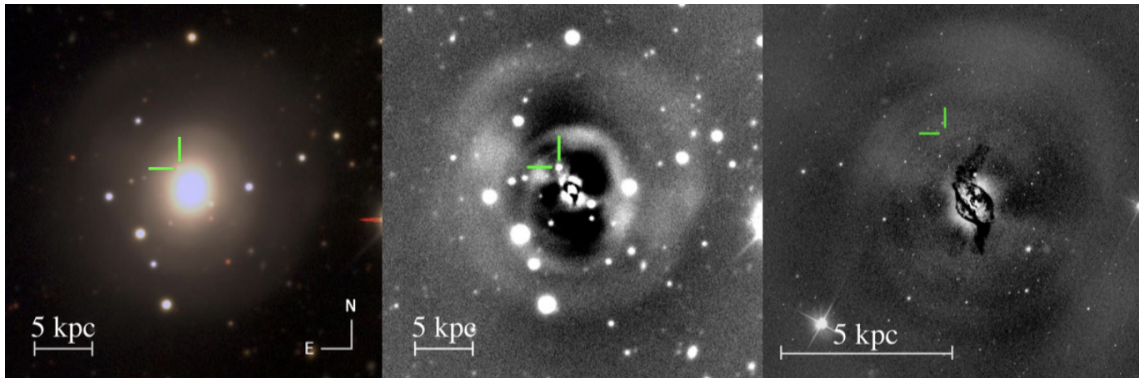


Figure 5.1: Visualization of the galaxy NGC 4993, the host galaxy of the GW170817 gravitational wave event. The left panel shows the DECAM coadded image in the g, r, i filters. In the middle panel the residuals from r -band Sérsic fit clearly shows its shell-like structure. The HST ACS image (F606W-band) in the right panel emphasizes dust lanes crossing the center of the galaxy. Courtesy of [Palmese et al., 2017a](#).

5.1 DES Y1 GALAXY MORPHOLOGY CATALOGUE AND GRAVITATIONAL WAVES

The tools developed for the production of the galaxy morphology catalogue for DES Y1, which is one of the main subjects of this thesis, and the consequent knowledge acquired about DES photometric data, guided the structural analysis of NGC 4993, known also as the host galaxy of the GW170817 gravitational wave event, generated by a neutron star binary system. A complete study of the galaxy star-formation history and morphology is reported in [Palmese et al., 2017a](#). The morphological analysis was made in terms of multi-wavelength parametric fits and non-parametric analyses, whose combined information was able to give important insights on the properties and evolutionary history of this shell-like galaxy. The values found for Concentration, Clumpiness and Asymmetry are in line with an ETG. The results from the parametric fitting reveal aspects of the galaxy sub-structure. More precisely, it reported increasing values for the Sérsic index towards redder bands and different values in the position angle, suggesting a rotation of bluer versus redder wavelengths. This might indicate the presence of two superimposed stellar populations with different orientations. The superposition may have arisen from a galaxy merger, as also suggested by the presence of galaxy shells. The shell structure is clearly visible in the residuals of the parametric fits, as shown in Figure 5.1 (and reported in Figure 1 of the original paper). The authors used the evidences collected from the structural analysis and spectroscopic fitting to sustain the scenario that there is a relation between the dynamical interactions occurred during a galaxy merger and the formation of the neutron star binary that produced the gravitational wave signal.

5.2 PyCosmo and FFTLog

As an alternative to the Line-Of-Sight integration, described in Chapter 4 to calculate the CMB angular power spectrum, we explored the possibility of applying the FFTLog method. The work by [Schöneberg et al., 2018](#) presents an implementation of it for the computation of large-scale structure observables. They show it is convenient since it simplifies the numerical integration over Bessel functions and can assure an accuracy comparable to that achieved with traditional methods. Additionally, it still guarantees a level of precision which is not achievable just with the Limber Approximation. Similarly, in [Fang et al., 2020](#) this method is described and proven to be efficient in the computation of galaxy clustering and weak lensing. Also in this case, the FFTLog is presented as a promising alternative to Limber-Approximation, which is not sufficient to fulfill the precision required by upcoming cosmological surveys.

Formulated in [Talman, 1978](#) and adopted for the first time in cosmology by [Hamilton, 2000](#), the FFTLog method is used to compute an integral containing one spherical Bessel function. Following the formalism in [Fang et al., 2020](#), we briefly summarize such method and discuss its application to the angular power spectra in PyCosmo. Let's suppose we need to evaluate the following integral,

$$F(r) = \int_0^\infty \frac{dk}{k} f(k) j_\ell(kr), \quad (5.1)$$

where the function $f(k)$ is logarithmic-sampled in k . In the FFTLog method we decompose it into a series of power laws. For example, we can write:

$$f(k_q) = \frac{1}{N} \sum_{m=-N/2}^{N/2} c_m k_0^\nu \left(\frac{k_q}{k_0} \right)^{\nu+i\eta_m}, \quad (5.2)$$

where N is the size of the function, ν is a bias index which controls ringing effects and $\eta_m = 2\pi m / (N \Delta_{\ln k})$, where $\Delta_{\ln k}$ encodes the linear spacing in $\ln(k)$ so that, for example, $k_q = k_0 e^{q \Delta_{\ln k}}$, with k_0 representing the smallest value for the k array. After a change of variable, $x = kr$, the integral becomes:

$$\begin{aligned} F(r) &= \frac{1}{Nr^\nu} \sum_{m=-N/2}^{N/2} c_m k_0^{-i\eta_m} r^{-i\eta_m} \int_0^\infty \frac{dx}{x} x^{\nu+i\eta_m} j_\ell(x) = \\ &= \frac{\sqrt{\pi}}{4Nr^\nu} \sum_{m=-N/2}^{N/2} c_m k_0^{-i\eta_m} r^{-i\eta_m} g_\ell(\nu + i\eta_m), \end{aligned} \quad (5.3)$$

where we wrote the integral over the spherical Bessel function in a form which is ana-

lytically solvable through the usage of Gamma functions:

$$g_\ell(z) = 2^z \frac{\Gamma\left(\frac{\ell+z}{2}\right)}{\Gamma\left(\frac{3+\ell-z}{2}\right)}, \quad -\ell < \mathcal{R}(z) < 2, \quad (5.4)$$

with the bias index varying in the interval $(-\ell, 2)$.

In the case of CMB, the equations 4.13, 4.14 and 4.15, reported and commented in Chapter 4 to describe the line-of-sight integration, can be summarized in the following way:

$$C_\ell^{TT} = (4\pi)^2 \int_0^{\eta_0} d\eta \int_0^{\eta_0} d\eta' \int dk k^3 P(k) j_\ell[k(\eta - \eta_0)] j_\ell[k(\eta' - \eta_0)] S(k, \eta) S(k, \eta'), \quad (5.5)$$

where $S(k, \eta)$ is the source function introduced in equation 4.13 and $P(k)$ is the matter power spectrum. In the line-of-sight method first the source function is integrated over the conformal time η , performing a double spherical bessel integral, as shown in equation 4.14. The result is the temperature field, Θ , which is integrated over k together with the power spectrum to get the angular power spectrum (equation 4.15). In the FFTLog approach, we consider $P(k)$ as the logarithmic-spaced function (in analogy to $f(k)$ in equation 5.1 above), and we reduce the bessel integration over k in the general prescription of equation 5.5 to an integral analytically solvable with Gamma functions. In this procedure, we further need to evaluate the source function for each k and make it evolve in time. This approach is demonstrated to be convenient for other observables, where the source function has a less complex form and the angular power spectrum doesn't require the double integration. In the case of the CMB, however, we observe that the FFTLog doesn't bring an improvement in the execution time, which is in general slow for the evaluation of Bessel integrals. For this reason we decided to focus on the standard line of sight integration and optimize the bessel integration so that the execution speed in PyCosmo is comparable to other competitive cosmological codes.

5.3 KINETIC FIELD THEORY IN PYCOSMO

The main idea behind the Kinetic Field Theory (KFT) is to analytically describe the evolution of a system by starting an initial probability distribution on the phase-space of point particles, \mathbf{x} , and make it evolve by mapping through an Hamiltonian flow. The illustration of this idea is sketched in Figure 5.2. Following the full prescription described in Bartelmann, Kozlikin, et al., 2019b, this translates as follows:

$$P(\mathbf{x}) = \int d\mathbf{x}^i P(\mathbf{x}|\mathbf{x}^i) P(i), \quad (5.6)$$

where we split the occupation probability $P(\mathbf{x})$ for the state \mathbf{x} into a probability $P(i)$

that the particle ensemble occupies an initial state \mathbf{x}^i at time $t = 0$, and the conditional probability $P(\mathbf{x}|\mathbf{x}^i)$ expressing how likely the ensemble moves from the initial position to the time-evolved state \mathbf{x} . The conditional probability assumes therefore the meaning of transition probability from state \mathbf{x} to state \mathbf{x}^i , and can be expressed as a function of the classical equation of motion:

$$P(\mathbf{x}|\mathbf{x}^i) = \delta_D[\mathbf{x} - \Phi_{cl}(\mathbf{x}^i)] = \delta_D[E(\mathbf{x}, \mathbf{x}^i)], \quad (5.7)$$

where δ_D is the Dirac delta function and $\Phi_{cl}(\mathbf{x}^i)$ is the classical Hamiltonian flow in the phase space, representing the solutions $E(\mathbf{x}, \mathbf{x}^i)$ to the equation of motion. The Dirac delta function acts in order to single out the trajectories solving the equation of motion. This model is actually represented in Figure 5.2 where it is even clearer an important aspect of this formulation: the trajectories do not cross.

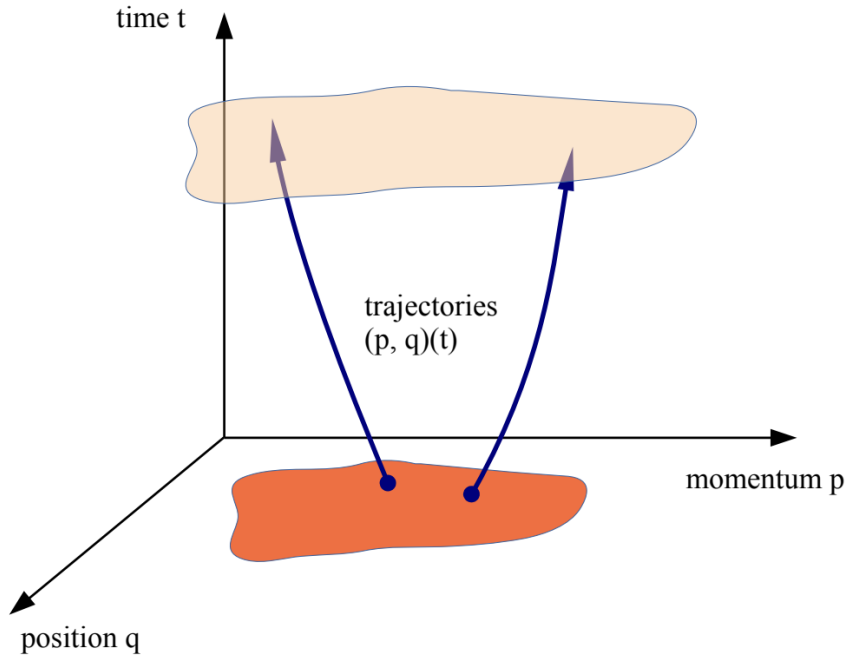


Figure 5.2: Schematic view of the main idea behind KFT, where an initial probability distribution on the phase space is mapped in time. The evolution is modelled by the Hamiltonian of the system. Since trajectories don't cross, this framework can be applied to cosmology and the shell-crossing problem arising in the structure formation theory in non-linear regime. Image: courtesy of [Bartelmann, Kozlikin, et al., 2019b](#).

If applied to the field of cosmological perturbations, this aspect represents an advantage in describing the evolution of the structure formation in the non-linear regime. In fact, in the standard cosmological perturbation theory, as we mentioned in Chapter 1, it is common practice to show that by assuming ideal hydrodynamics the perturbations grow linearly and independently, proportional to the scale factor $D(a)$; however, in non-

linear regimes dark-matter streams meet each other, giving rise to the shell-crossing problem which is observed in numerical simulations. The KFT approach, by avoiding shell crossing, proposes an analytical alternative to cosmological simulations for perturbation theory. In PyCosmo, we explore the possibility of including this theory, starting from the implementation of the non-linear power spectrum, described, at a given scale factor a , as follows:

$$P(a, k) = e^{-\langle S_I \rangle(a)} P_{\text{lin}}(a, k), \quad (5.8)$$

where $P_{\text{lin}}(a, k)$ is the linear power spectrum and S_I is the interaction term encoding the time evolution of the system. Since the linear power spectrum is proportional to the growth factor, it includes the information on the cosmological model and the initial conditions from where the system evolves fully analytically. The current implementation in PyCosmo is under testing and already shows a good agreement with the original code by [Bartelmann, Kozlikin, et al., 2019b](#). Future developments will involve the introduction of symbolic calculations with the Python library *SymPy*.

5.4 PYCOSMO AND THE MCCL ANALYSIS

PyCosmo has been implemented as part of the Monte Carlo Control Loop (MCCL) pipeline ([Alexandre Refregier and Adam Amara, 2014](#); [Kacprzak, 2020](#)). MCCL is a forward-modelling method designed to tackle the issues arising during the estimation of systematic errors in cosmic shear analyses. It relies on the Ultra Fast Image Generator, UFig (already mentioned in Chapter 2 and described in [Bergé et al., 2013](#)), to produce simulated images matching the target data, followed by calibration loops. More precisely, the pipeline starts with performing object detection from co-add DES images, then it models the PSF and proceeds with shear calibration, measurement of the photometric redshift, calculation of the covariance matrix and computation of the power spectra. By measuring the combination of Ω_m, σ_8 and the intrinsic alignment amplitude A_{IA} this work tests the validity of this method for cosmic shear DES analysis and its applicability to future wide-field surveys. The parameter constraints were obtained assuming models and predictions of cosmological observables made with PyCosmo, including both *Halofit* and the *Mead model* for the non-linear power spectrum used to predict the shear power spectrum.

An additional follow-up likelihood analysis has been made to constrain (Ω_m, σ_8) and (S_8, A_{IA}) , but using a different numerical approach to sample the parameter space. Below we give some notation about the likelihood and describe in detail methods and results. The likelihood function, $\mathcal{L}(\mathbf{d}; \boldsymbol{\theta})$, is a statistical quantity allowing to test the validity of a

certain model given a collection of data from observations:

$$\mathcal{L}(\mathbf{d}; \boldsymbol{\theta}) \equiv P(\mathbf{d}|\boldsymbol{\theta}), \quad (5.9)$$

where $\boldsymbol{\theta}$ includes the set of model parameters and \mathbf{d} is the experimental data vector. The scope of a likelihood analysis is to evaluate the range of the model parameters which best reproduce the data. They are given by the probability $P(\boldsymbol{\theta}|\mathbf{d})$, which is related to the likelihood function through the Bayes theorem:

$$P(\boldsymbol{\theta}|\mathbf{d}) = \frac{P(\boldsymbol{\theta})P(\mathbf{d}|\boldsymbol{\theta})}{P(\mathbf{d})} = \frac{P(\boldsymbol{\theta})\mathcal{L}(\mathbf{d}; \boldsymbol{\theta})}{P(\mathbf{d})}, \quad (5.10)$$

where $P(\boldsymbol{\theta})$ is the *prior*, representing the predicted distribution of the model parameters, and $P(\mathbf{d})$ is the analogue information on collected data and reduces to a normalization factor. $P(\boldsymbol{\theta}|\mathbf{d})$ is also referred to as the *posterior* distribution: its maximum value is the most probable set of model parameters matching the data, and its width quantifies the uncertainty on our measurements. If we assume the prior is a flat distribution, where all the parameters are equally probable, then equation 5.10 becomes:

$$P(\boldsymbol{\theta}|\mathbf{d}) \propto \mathcal{L}(\mathbf{d}; \boldsymbol{\theta}), \quad (5.11)$$

where a direct relation between the measurement of the posterior and the likelihood function is established. As we mentioned above, the latter expresses the probability that the experiment measures the data \mathbf{d} given certain model parameters $\boldsymbol{\theta}$. Therefore, maximizing the likelihood provides the most probable set of model parameters, $\boldsymbol{\theta}_{ML}$, where ML stays for Maximum Likelihood. Finding the maximum likelihood can be achieved through maximum likelihood estimators. Alternatively, one can sample the parameter space with a Monte-Carlo algorithm. In such cases, the likelihood function can be written as a multivariate Gaussian distribution with covariance C :

$$\mathcal{L}(\mathbf{d}; \boldsymbol{\theta}) = \frac{1}{(2\pi)^{N/2}|C|^{1/2}} e^{-\frac{1}{2}(\mathbf{d}-\mathbf{m})^T C^{-1}(\mathbf{d}-\mathbf{m})}, \quad (5.12)$$

where \mathbf{d} is the data vector of length N and \mathbf{m} is the theoretical data vector depending on model parameters $\boldsymbol{\theta}$. The argument of the exponential is proportional to the χ^2 distribution, so that minimizing it means maximizing the likelihood. The estimation of the covariance matrix can be challenging and computationally expensive, especially if the parameter space is high dimensional. It can be estimated from the data, from simulations or analytically. In the first two scenarios, the estimate of the covariance matrix takes the form:

$$\hat{C} = \frac{1}{N-1} \sum_{k=1}^N (\mathbf{d}^{(k)} - \boldsymbol{\mu})(\mathbf{d}^{(k)} - \boldsymbol{\mu})^T, \quad (5.13)$$

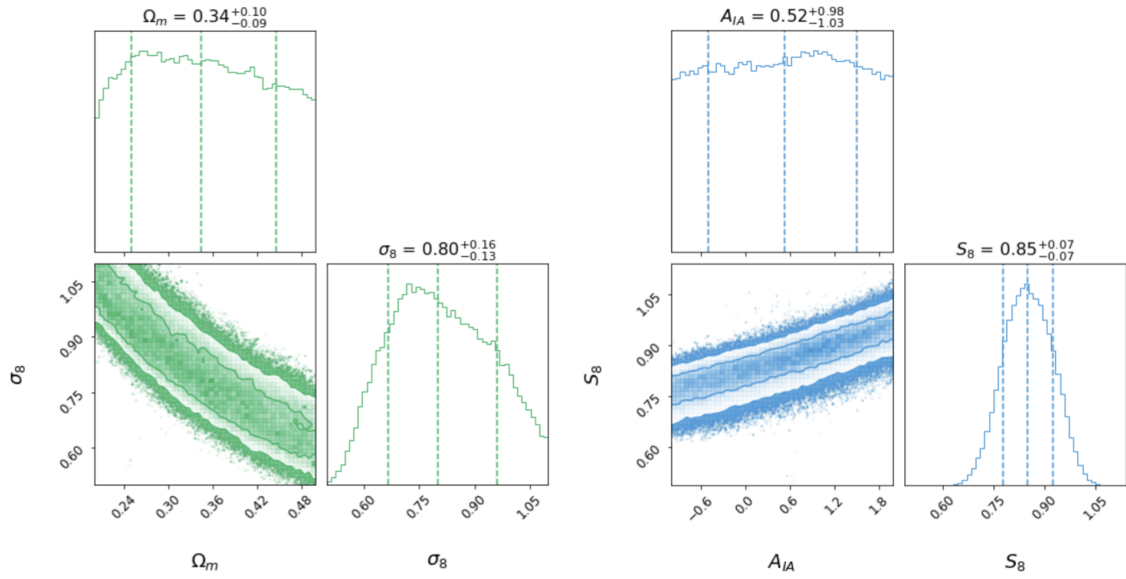
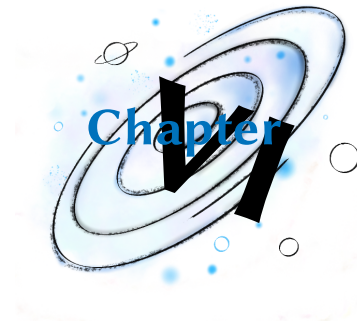


Figure 5.3: Preliminary results from the likelihood analysis using the MCCL output for the DES Y1 dataset. We are able to recover the same results obtained in (Kacprzak, 2020). The weak lensing angular power spectrum was calculated with the *Halofit* prescription for the non-linear power spectrum.

where μ is the average of the data realizations. This term might be challenging to estimate as well because in cosmology we cannot get several realizations of our Universe. We can overcome this limitation using cosmological simulations, which provide both simulated data vectors, $\mathbf{d}_{\text{sim}}^{(k)}$ and the μ vector averaged over several realizations of the data. For the simulation approach, the limit is the computational power required for such realizations. If the covariance matrix is estimated from theory, then the data vector $\mathbf{d}^{(k)}$ follows the model prescriptions and the noise due to the experiment is added. For more information about the different approaches and their properties we refer the reader to e.g. Hartlap, Simon, and P. Schneider, 2007; Dodelson and M. D. Schneider, 2013; Krause and Eifler, 2017b; Lacasa, 2018; A. Schneider et al., 2020.

The MCCL pipeline numerically estimates the covariance matrix using U-FALCON, as described in detail in Sgier et al., 2019. To sample the parameter space we follow a different approach than the one used in the MCCL analysis, using the package *ChaosHammer*, developed in our research group. Our analysis represented also a way to test this code, since it is of recent implementation. The package first calculates the model vector on a grid of user-defined points. The output is collected and passed to an emulator, which is trained over the grid points. Then, each grid point is sampled using an emcee algorithm (Foreman-Mackey et al., 2013). The latter is a python sampler for the Monte Carlo Markov Chain (MCMC) that is crucial to converge to the posterior distribution we

want to determine. Starting from the grid points the algorithm generates an ensemble of chains, which consists in stochastic *walkers*. The walkers follow the direction which maximizes the probability of the parameters. The usage of *ChaosHammer* is more convenient than the direct application of an MCMC, especially when the number of parameters to model is limited, because it allows for much faster computations. In our analysis, with a six-dimensional parameter space $(\Omega_m, \sigma_8, \Omega_m, h, n_s, A_{IA})$, the computations on the grid takes a few hours and training the emulator a few minutes. It represents a great advantage in comparison to the time scales required for an MCMC chain, which can last several days. Also, the emulator reached an accuracy of 99,7%, allowing precise computations of the power spectra during the sampling. In this phase, the likelihood function is computed and the sampler converges towards regions where the likelihood is maximized. Preliminary results probing the planes (Ω_m, σ_8) and (S_8, A_{IA}) with the DES likelihood are in agreement with the constraints recovered in (Kacprzak, 2020), as shown in Figure 5.3.



Conclusions

L'arte di scriver storie sta nel saper tirar fuori da quel nulla che si è capito della vita tutto il resto; ma finita la pagina si riprende la vita e ci s'accorge che quel che si sapeva è proprio un nulla.

The art of writing tales consists in an ability to draw the rest of life from the little one has understood of it; but life begins again at the end of the page, and one realises that one has knew nothing whatsoever.

ITALO CALVINO, *IL CAVALIERE INESISTENTE*

In this thesis we described both the observational and the theoretical work developed and tested mainly within the framework of the Dark Energy Survey (DES). With more than 300 million of observed galaxy images, DES is only one in the list of recent and upcoming surveys delivering *big data* for research in astronomy and cosmology. Such availability of data comes with benefits. First, the large statistical power reachable using these datasets is used to study the properties of our Universe. Second, and not less important, the measurements provided by these surveys are not only unprecedentedly abundant, but also very precise. Therefore, in the context of the standard cosmological model, we need to develop computational tools which are both fast, in order to process a large amount of data points, and precise, in order to match with statistical precision afforded by observations. This motivated us to develop the cosmology code PyCosmo. Galaxies are one of the tracers we use to study the Universe. We highlighted the importance of cataloguing them according to their structural and morphological properties, since they are correlated to other quantities measurable over cosmic time, such as color,

stellar activity and environmental interactions. Such relations help researchers to understand the galaxy formation and evolutionary history.

In Chapter 2 we presented the production and validation of the largest structural and morphological catalogue to date, including measurements for 45 million galaxies detected over 1800 square degrees during the first year of observations from DES. For each galaxy the catalogue comprises a full structural description, thanks to the adopted double approach: parametric fitting, estimating the structural parameters which best match an underlying mathematical model, called Sérsic profile, and a non-parametric approach, studying the moments of the galaxy light distribution and extracting information on its concentration, level of asymmetry and clumpiness. As shown in Figure 2.11 the combination of the parameters obtained with these two approaches leads to a deep understanding of the galaxy populations, and allows the distinction between disk-like and bulge-dominated galaxies, bluer and redder and the identification of objects showing irregular structures and clumpy regions, which are interesting to study environmental effects due to gravitational interactions. The validation of the parametric measurements also contributed to a better understanding of the PSF across the DES field, and to flag those measurements which are mostly affected by it. Both for parametric and non-parametric measurements we applied calibrations using the UFig images simulations. Finally, some of the codes developed for the production of the catalogue were made available to the community to process DES data. The efforts made during the production of such big catalogue and the knowledge acquired about DES data lead to two other analyses.

The first one is a collaborative work described in [Palmese et al., 2017a](#). In this work we studied the properties of NGC 4993, the host galaxy of the GW170817 gravitational wave event. The analysis of this optical counterpart shed some light on an extraordinary event whose messenger are gravitational waves. The galaxy was studied both in terms of its spectral and photometric properties, and with my work I contributed to the latter. The combination of parametric and non-parametric measurements revealed angular shifts across different wavelengths, from bluer to redder, indicating the presence of two overlapping stellar-populations with offset orientations. This configuration triggers the hypothesis of a galaxy merger, which is compatible with the observed shell-like morphology. This particular shape is well visible from the residuals of the parametric fitting. The information collected during the full analysis lead the authors to claim a relation between the merging scenario and the formation of the neutron star binary that sourced the gravitational wave signal. More information are given in Chapter 5.

The second additional work which used the DES Y1 morphological catalogue and DES data is described in Chapter 3 and focuses on a novel approach applicable to galaxy morphology. The method consists in performing isophotal analyses of galaxy images

and collecting the relative information into series. These are not to be confused with Sérsic profiles, which derive from a parametric fitting procedure that accounts for the PSF effects. The advantage of extracting information through isophotes is double: first, it allows the detection of the internal galaxy structure without the need of accounting for the PSF, and provides a data series including features which allow to distinguish between ETG and LTG. Second, the reduction from 2D images into 1D series is of great advantage for the management of big datasets. Also, the application of ML algorithms to such data format is very simple. The results, obtained with a workflow built in collaboration with the Swiss startup Modulos AI, are promising, as they are already comparable with CNN analyses carried on DES data. Future improvements are envisioned not only to optimize the isophotal fitting, but also to apply this routine to the automatic detection of galaxy sub-structural features, e.g. bars and clumps.

In Chapter 4 we presented the development of PyCosmo, a python-based framework providing theoretical predictions for cosmological analyses. Being competitive with similar existing codes, PyCosmo introduces the novelty of being implemented in python and the interactivity of an online platform, named PyCosmo Hub, where the users can make computations without the need of installing any software. The platform can represent also a useful didactic tool. In this thesis, we showed how PyCosmo is fast and accurate in delivering predictions of cosmological observables for the Λ CDM model, assuming different theories for linear and non-linear perturbations. We also explored the possibility of implementing new models, such as the FFTLog for the computation of the CMB angular power spectrum and the KFT for the computation of the matter power spectrum. Finally, we described how PyCosmo has been successfully used in cosmological analyses, such as the MCCL by [Kacprzak, 2020](#) and the likelihood analysis described in Chapter 5. Future developments in PyCosmo will focus on speed optimization in the CMB angular power spectrum and the testing of the KFT theory, as well as the w CDM model. And the inclusion of a module for likelihood functions from different surveys, in order to enable an easy implementation of highly constraining joint cosmological analyses.

The results, methods and advances described above offer together good perspectives for the analysis of present and future cosmological surveys.



Contributions

The scope of this chapter is to describe my work and that of the others who contributed to the projects presented in this thesis.

CHAPTER I In this chapter I give an overview on the scientific topics underlying my research work. The didactic content is based on the literature referenced in the chapter.

CHAPTER II This work focused on the production of the structural and morphological catalogue for DES Y1. I was the main author and contributor to this project, being responsible for all the steps: preliminary analysis and management of the raw data, code development, run on computer clusters, validation and calibration of the results. The validation process benefits from the fruitful discussion with Dr. William Hartley, who is the second author of the paper on which the chapter is based. He is the author of Fig. 2.3 in the same chapter. The calibrations of the measurements were conducted using the simulations produced with the *Ultra Fast Image Generator* (UFIG) (Bergé et al., 2013) and benefited from the valuable inputs by Dr. Claudio Bruderer. The phase of upload of the final catalogue on the DES database was made easier thanks to the support provided by Dr. Matias Carrasco and Dr. Nacho Sevilla from the DES Collaboration. Contributions to the initial steps of the project and further suggestions came from the other authors of the paper and the members of the Galaxy Evolution working group of the DES Collaboration. I also thank Prof. Richard Kron and the DES Management Committee for approving my nomination for DES Builder status, on the basis of the contribution I gave to the experiment with the publication of this catalogue.

CHAPTER III The work presented in this chapter proposed a novel method for both galaxy morphological classification and sub-structure detection, combining isophotal analysis, extracting information from 2D images and collecting it into 1D series, and Machine Learning algorithms. I am the author behind this idea and its implementation. For the automated classification I used the framework provided by the team of Modulos AutoML, and interpreted the results together with Dr. Kevin Schawinski, Dr. Claudio Bruderer and Dr. William Hartley.

CHAPTER IV The PyCosmo code was first presented in (A. Refregier, Gamper, et al., 2018a), focusing on the Boltzmann Solver. In a second phase, part of the background computation and the cosmological observables were implemented by Prof. Adam Amara, Dr. Uwe Schmitt, Dr. Andrina Nicola and Prof. Refregier. I continued the work of development and testing, working on the different models for linear and non-linear matter power spectra and the angular power spectra, and I made the first implementation of the CMB angular power spectra with the line-of-sight integration. I was the lead author of the research described in this chapter. In this project of development and testing, I have been supported by Dr. Uwe Schmitt. Our efforts lead to the first public release of PyCosmo, providing predictions for cosmological observables. Thanks to the collaboration with Janis Fluri, it was possible to make this release available on the public platform PyCosmo Hub. Furthermore, the project benefited from the valuable insights provided by Prof. Lavinia Heisenberg.

CHAPTER V The works described in this chapter refer to several collaborative projects and additional recent analyses. I contributed to the study of the host galaxy of the GW170817 gravitational wave event, reported in full details in Palmese et al., 2017a, focusing on the structural and morphological analysis of the galaxy. The outcome of such analysis, as shown in Section 5.1, provided important insights on the dynamical interaction history of this galaxy and its connection to the neutron star binary which is the source of the gravitational wave event. Regarding the MCCL project described in Section 5.4, I contributed in support of the analysis with my work on PyCosmo, since theoretical predictions are one of the steps of the MCCL pipeline.

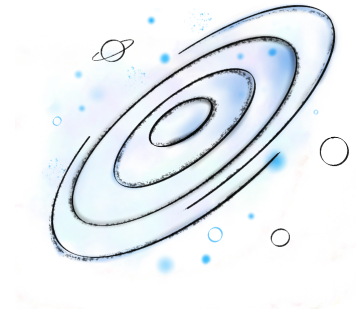
I studied the possible application of the FFTLog method to PyCosmo (Section 5.2) in collaboration with Prof. Alexandre Refregier and Dr. Uwe Schmitt, and I implemented the KFT in collaboration with Stefan Zentarra, supervised by Prof. Lavinia Heisenberg. Finally, I conducted the likelihood analysis described in Section 5.4 in collaboration with Prof. Alexandre Refregier and Dr. Tomasz Kacprzak who is the author of the package *ChaosHammer*.

CHAPTER VI I wrote the conclusions reported in this chapter to summarize the outcome of my research work and discuss its future perspectives.

APPENDICES I reported in Appendix A and B additional details on the analyses described in Chapters [2](#) and [4](#), respectively.

Appendices

Appendix A



A.1 CALIBRATION MAPS FOR THE G AND R FILTERS

In this Appendix we present the calibration maps for both parametric and non-parametric measurements in the g and r bands. They were obtained following the procedure described in Sections 2.2.2 and 2.5.4 for parametric and non-parametric fits, respectively. The maps are displayed following the same conventions adopted for visualising the calibration maps in the i band. Those maps are shown in Figures 2.9 and 2.12.

A.1. Calibration maps for the g and r filters

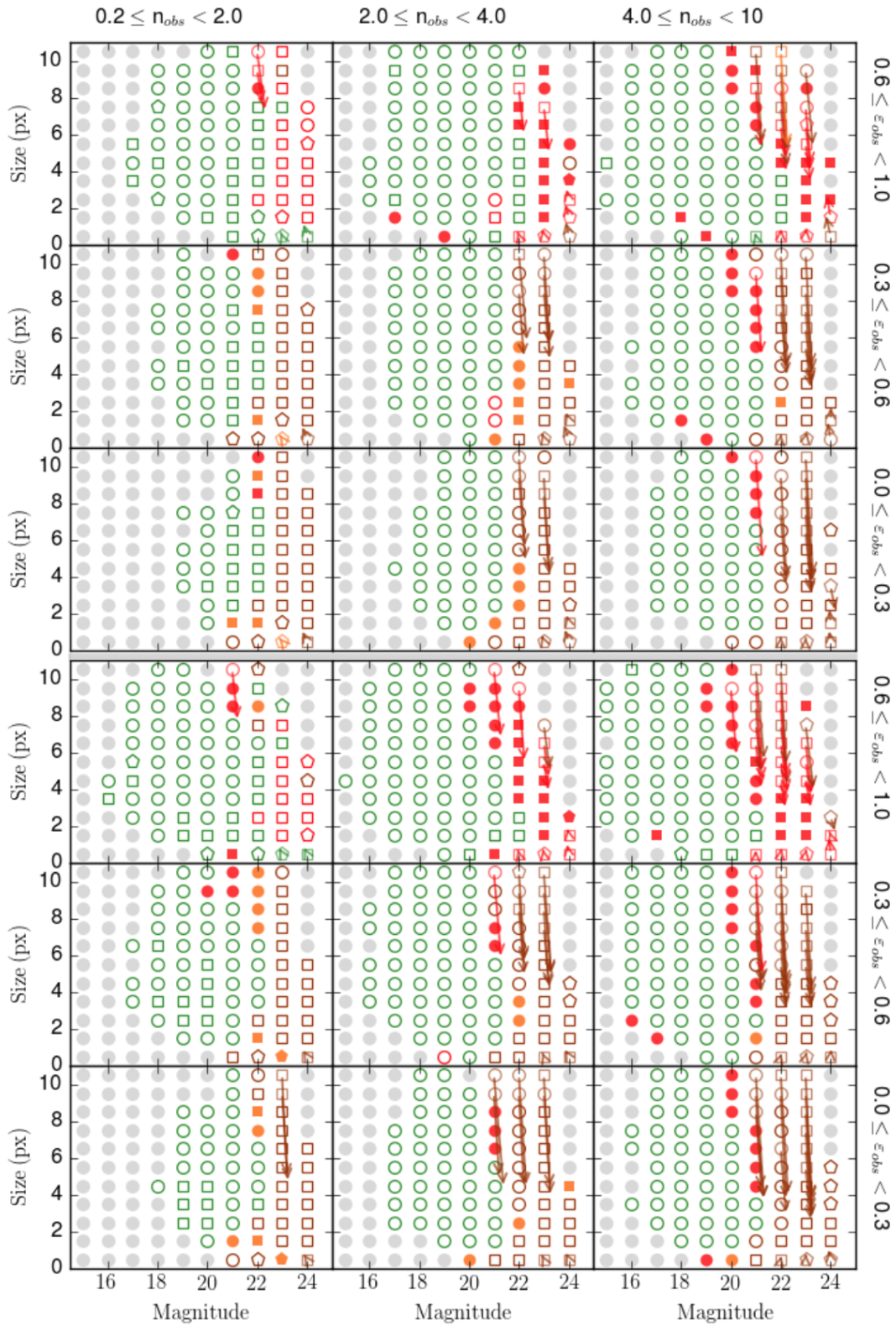


Figure A.1: Map of the corrections for Sérsic parameters in the g (upper panel) and r (lower panel) filters, obtained through the simulation routine described in Section 2.2.2. Symbols and colours have the same meaning as Figure 2.9.

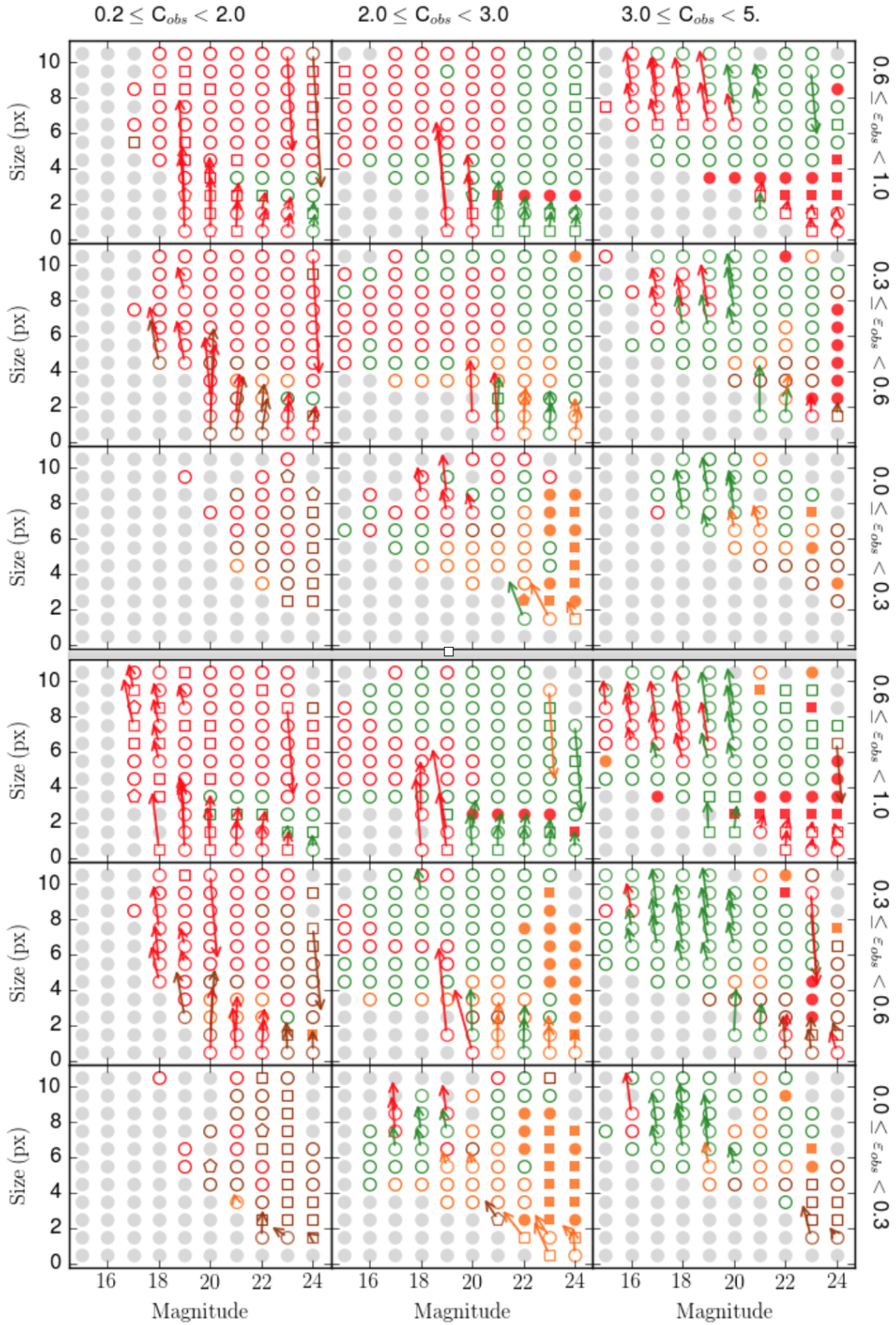


Figure A.2: Map of the corrections for ZEST+ output in the g (upper panel) and r (lower panel) filters, obtained through the simulation routine described in Section 2.5.4. Symbols and colours have the same meaning as Figure 2.12.

A.2 CATALOG MANUAL

A description of the columns of the catalogue follows, both for parametric and non-parametric fits. In order to distinguish between filters, the parameters can be labelled with $_X$, where $X = g, r, i$.

A.2.1 IDENTIFICATION COLUMNS

COADD_OBJECT_ID - Identifier assigned to each object in the co-add DES Y1 dataset, reported here from the *Gold Catalogue*.

TILENAME - Column reporting the name of the tile image where the object lies.

ID - Rows enumerator, running for 1 to the total number of entries in the catalogue.

RA - Right Ascension from the Y1A1 GOLD catalogue.

DEC - Declination from the Y1A1 GOLD catalogue.

A.2.2 SExtractor PARAMETERS FOR STAR-GALAXY SEPARATION AND SIGNAL-TO-NOISE

SG - Linear combination of the star-galaxy classifier SPREAD_MODEL and its uncertainty, SPREADERR_MODEL, according to Equation 2.4. A cut in $SG > 0.005$ is recommended.

SN_X - Signal-to-noise expressed as the ratio between FLUX_AUTO_X and FLUXERR_AUTO_X.

A.2.3 COLUMNS FOR PARAMETRIC FITS

SELECTION AND PRE-FITTING CLASSIFICATION FLAGS

SELECTION_FLAGS_X - If equal to 1, then the relative object has been selected, according to the requirements described in Section 2.3.1. It can assume other numerical values in the following cases:

- if the object passes the selection requirements, but is not included in the intersection between the DESDM catalogues and the Y1A1 GOLD catalogue, then this flag is set to 2;
- if the object passes the selection requirements, but it is fainter than $GOLD_MAG_AUTO_i = 23$, then the flag is set to 3;
- if the object enters in the previous category, but it has no match with the Y1A1 GOLD catalogue, then the flag is set to 4.

If the object is not selected because it doesn't pass any of the selection requirements, then the `SELECTION_FLAGS_X` and all the other flags are set to zero.

The catalogue version made available to the users includes all the objects which have been selected at least in one of the three bands g,r,i .

`C_FLAGS_X` - Number of neighbours in the fitted stamp.

`MAX_OVERLAP_PERC_X` - Percentage of the central galaxy isophotes overlapping with the closest neighbour. If there are no neighbours or no overlapping neighbours, then it is set to 0. A cut in `MAX_OVERLAP_PERC_X < 50` is recommended.

PARAMETRIC MEASUREMENTS (GALFIT)

`MAG_SERSIC_X` - GALFIT value for the magnitude of the galaxy. The value already includes the calibration listed in the column `MAG_CAL_X`.

`RE_SERSIC_X` - GALFIT measure of the half light radius (or Effective radius) of the galaxy. It is expressed in pixels and is already calibrated. The correction is reported in the column `RE_CAL_X`.

`N_SERSIC_X` - GALFIT output for the Sérsic Index. The measure is calibrated, and the can find the relative correction in the column `N_SERSIC_CAL_X`.

`ELLIPTICITY_SERSIC_X` - Ellipticity of the galaxy, calculated by subtracting from unity the GALFIT estimate for the axis-ratio. The value is corrected and the calibration is accessible through the column `ELLIPTICITY_SERSIC_CAL_X`.

`OUTLIERS_X` - If equal to 1, it labels the objects classified as outliers in the catalogue validation process.

`FIT_STATUS_X` - If equal to 1, this flag selects all the objects with a successfully validated and calibrated converged fit.

Important note: by applying the recommended cut `FIT_STATUS_X = 1`, the user is able to collect the sample of validated and calibrated objects in the X filter. This cut is equivalent to applying all together the cuts which are recommended in terms of sample selection, fitting convergence, bad regions masking, exclusion of outliers and significantly overlapping objects, minimization of stellar contamination. A summarising scheme follows:

$$(\text{FIT_STATUS_X}=1) = \left\{ \begin{array}{l} \text{FLAGS_BADREGION=OSG}>0.005 \\ \text{SELECTION_FLAGS_X}=1 \\ \text{FIT_AVAILABLE_X}=1 \wedge \text{WARNING_FLAGS_CENTRAL_X}=0 \\ \text{MAX_OVERLAP_PERC_X}<50 \\ \text{OUTLIERS_X}=0 \\ \text{PARAMETER_CAL_X}<99, \end{array} \right.$$

where the voice `PARAMETER_CAL_X` can be `MAG_CAL_X` etc. In absence of calibration the correction value is set to 99.

For a cleaner sample the user can associate the cut in `FIT_STATUS_X` to the condition `SN_X>30`.

A.2.4 COLUMNS FOR NON-PARAMETRIC COEFFICIENTS (ZEST+)

`SELECTION_NP_X` - If equal to 1, the object is selected in the X filter, otherwise it is 0.

`FIT_STATUS_NP_X` - If equal to 1, this flag selects all the objects with successfully validated and calibrated measurements.

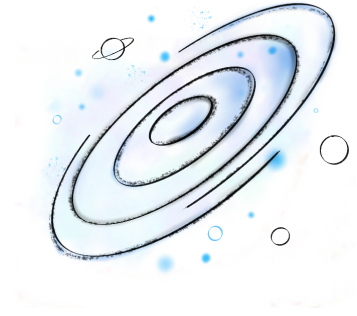
`CONCENTRATION_X` - ZEST+ measurement for the Concentration of light. See Equation 2.11 for its definition. The calibration vector is listed in the column `CONCENTRATION_CAL_X`.

`ASYMMETRY_X` - ZEST+ value for the Asymmetry (see Equation 2.16).

`CLUMPINESS_X` - ZEST+ value for the Clumpiness (see Equation 2.17).

`GINI_X` - Measure of the Gini parameter, defined in Equation 2.18.

`M20_X` - Measure of the M20 parameter, for more details see Equation 2.20.



Appendix B

B.1 CODE SETUP

As mentioned in Section 4.3, the tests performed between the codes require matching those not only in terms of cosmology, but also considering further parameters which change across the codes. They are set as follows:

- iCosmo (version 1.2): the agreement between PyCosmo and iCosmo has been tested by setting to zero the radiation density component ($\Omega_r = 0$), according to the default iCosmo setup. However, even in this configuration the CMB temperature is used in both codes to compute the EH linear fitting function. Therefore we set $T_{CMB} = 2.726K$, assuming for the CMB temperature the same value used in iCosmo. Concerning the growth factor, $D(a)$, both iCosmo and PyCosmo use the ODEINT solver (PyCosmo uses the `scipy.integrate.odeint` solver). We find an agreement up to 10^{-7} if we set the initial condition at $a = 10^{-3}$ and the tolerance parameters as follows:
 - iCosmo configuration: integration accuracy set to 10^{-4} , maximum step size to be attempted by the solver set to 10^{-3} and first attempted step size set to 10^{-3} ;
 - PyCosmo configuration: integration accuracy set to 10^{-9} in terms of relative tolerance and to 10^{-12} as absolute tolerance. First attempted step size set to 10^{-3} .

The computation of non-linear perturbations is tested in terms of the *Halofit* fitting formula proposed in [Smith et al., 2003](#), because this version is the one implemented in iCosmo. *Halofit* is checked both assuming the EH and $BBKS$ linear

- fitting functions. The matter power spectrum is then used to compute the lensing power spectrum. For the latter, we use iCosmo at its slower speed, so that a higher accuracy can be reached.
- **HMcode (Git version):** analogously to the iCosmo setup, the HMcode suppresses the radiation, so we set PyCosmo accordingly. In the HMcode code the CMB temperature enters the computation of the *EH* linear fitting function as a hard-wired value, $T_{CMB} = 2.728K$. We set it to this value also in PyCosmo. We match the codes also in terms of the growth factor: in the HMcode the accuracy of the ODEINT solver is set to 10^{-4} and the initial condition to 10^{-3} . We find the highest agreement if we assume for PyCosmo the same configuration used already in the comparison with iCosmo (see the details above). The comparison between PyCosmo and HMcode consists in testing the computation of the non-linear matter power spectrum as prescribed in the HMcode model, both in terms of dark matter only and exploring the effects of the baryons on the power spectrum. The algorithm has been implemented in Python in PyCosmo, and involves the *EH* linear fitting function, according to original prescription in HMcode.
 - **CCL (developer version 1.0.0):** the comparison between PyCosmo and CCL requires special care in terms of the growth factor. To achieve the best agreement we set PyCosmo so that the initial condition is at $a = 0.1$, the relative and absolute tolerance 10^{-3} and 10^{-12} , respectively, and the first attempted step size 10^{-3} . In addition to the background quantities, we can compare PyCosmo to CCL also in terms of linear and non-linear power spectra. Using the models available in both codes, we are able to compare the linear power spectrum both with the *EH* and *BBKS* fitting functions, and the non-linear power spectrum with the revised *Halofit* fitting function (Takahashi et al., 2012), adopting the two linear fitting functions. The matter power spectra are then involved in the computation of the observables, compared in terms of the lensing power spectrum.
 - **CLASS (version 2.7.1):** the agreement between PyCosmo and CLASS has been tested by using the CLASS python wrapper `classy`. When comparing the linear growth factor, we use for PyCosmo the same setup adopted in the comparison with CCL. Since the linear fitting functions *EH* and *BBKS* are not available in CLASS, we compare the linear power spectra computed with the Boltzmann solver. In this particular test, in order to match the several parameters characterising the two solvers and to achieve the highest possible accuracy, we run the original version of CLASS written in *C* language.

B.2 HEATMAPS

In this section we show the heatmaps summarizing the code comparison performed by varying the fiducial cosmological setup. More details about the cosmology assumed and the results are discussed in Section 4.3. For the description of the quantities shown in the heatmaps, we refer the reader to paragraph 4.3.1 and to Figure 4.4.

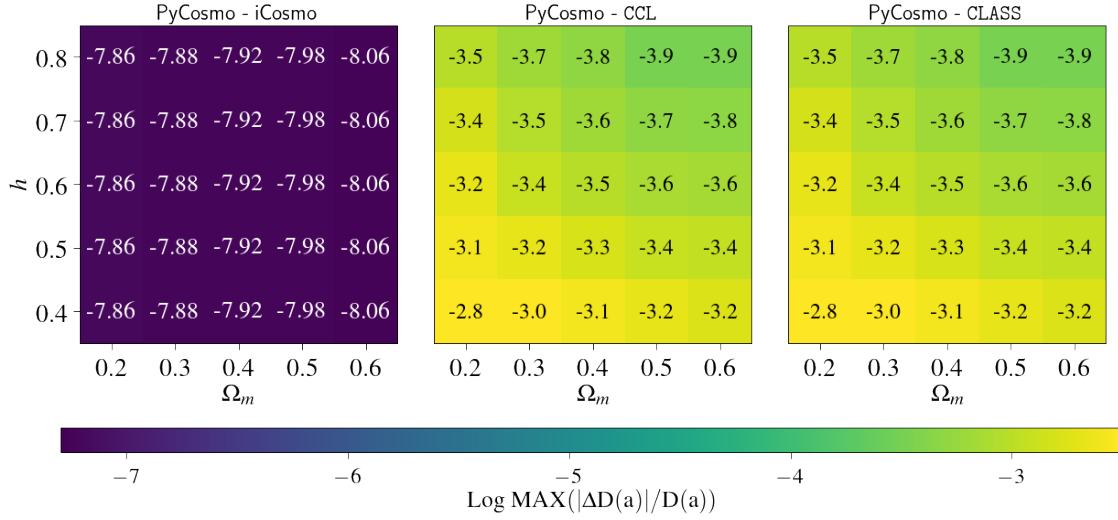


Figure B.1: Comparison between PyCosmo and iCosmo (left panel), PyCosmo and CCL (central panel) and PyCosmo and CLASS (right panel) in terms of the linear growth factor.

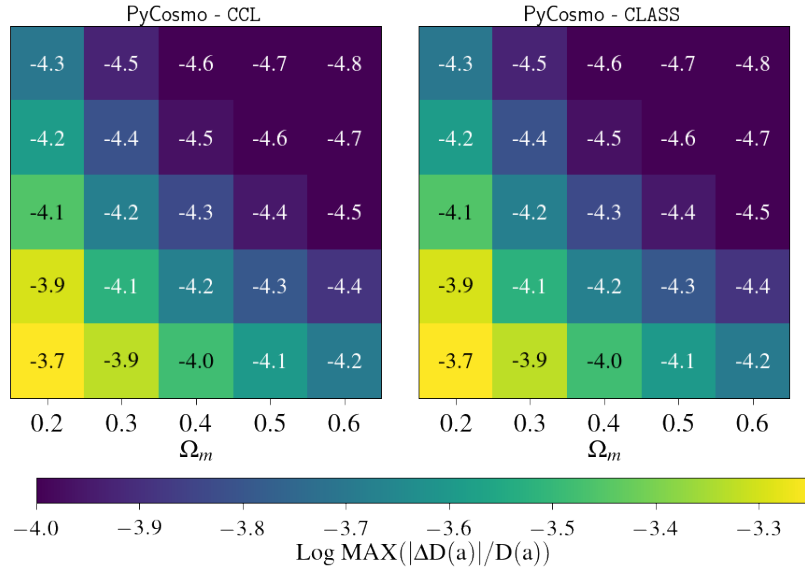


Figure B.2: Comparison between PyCosmo, CCL (left panel) and PyCosmo and CLASS (right panel) in terms of the linear growth factor. In this figure the growth factor in PyCosmo is computed with hypergeometric functions.

B.2. Heatmaps

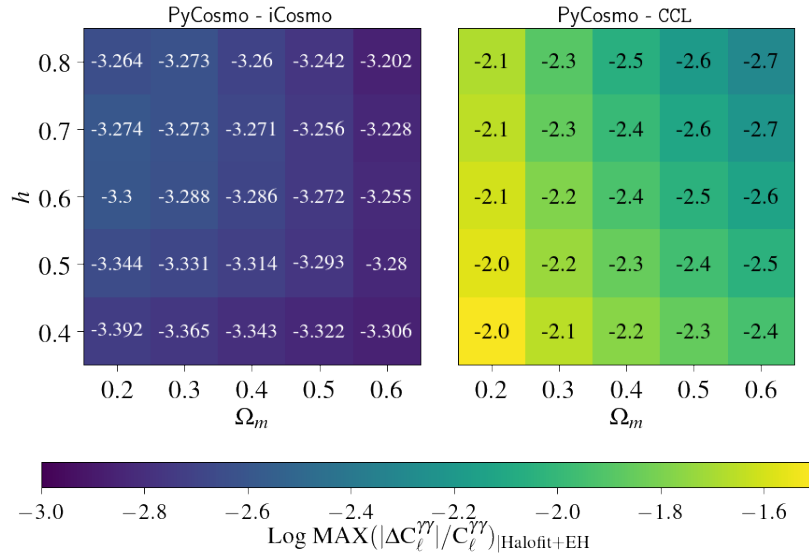


Figure B.3: Comparison between PyCosmo and iCosmo (left panel) and between PyCosmo and CCL (right panel) in terms of the lensing power spectrum, computed with *Halofit+EH* in the first case and accounting for the revised version of *Halofit* in the second case.

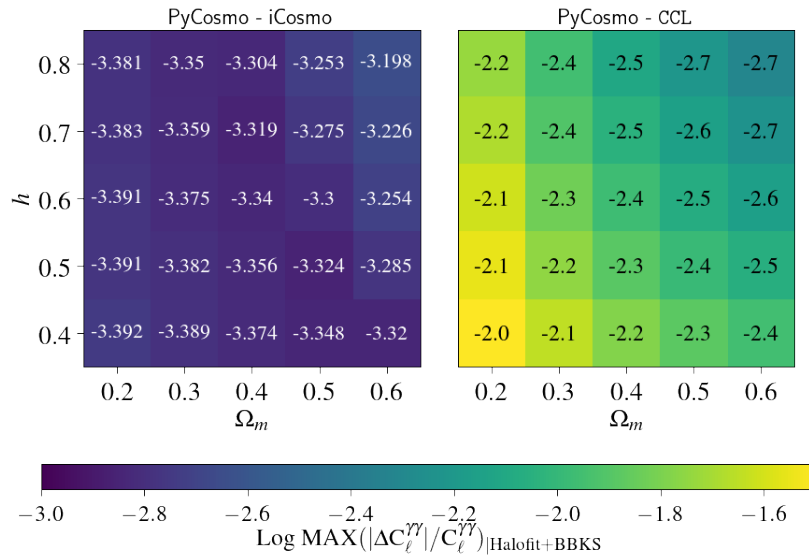


Figure B.4: Comparison between PyCosmo and iCosmo (left panel) and between PyCosmo and CCL (right panel) in terms of the lensing power spectrum, computed with *Halofit+BBKS* in the first case and accounting for the revised version of *Halofit* in the second case.

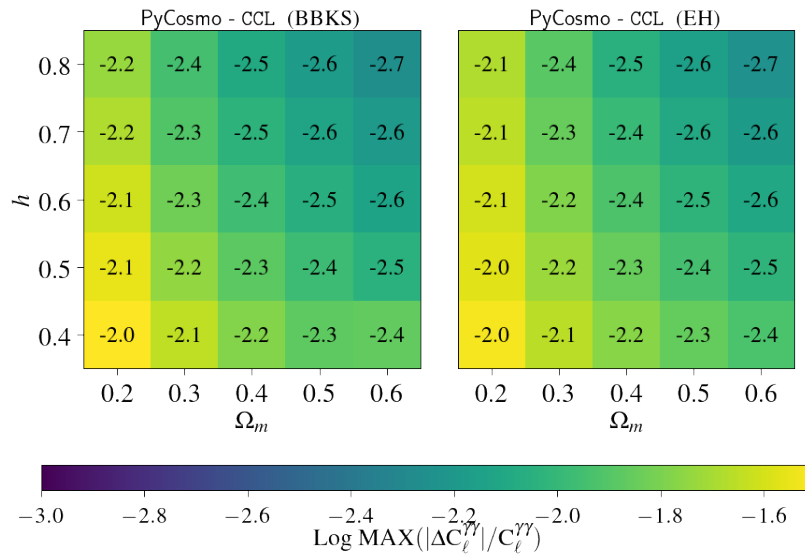


Figure B.5: Comparison between PyCosmo and CCL in terms of lensing power spectrum, computed with *BBKS* (left panel) and *EH* (right panel) linear power spectrum. The heatmaps are color-coded by the maximum relative difference between the two compared codes.

Acknowledgments

Tous pour un, un pour tous!
Tutti per uno, uno per tutti!

ALEXANDRE DUMAS, *LES TROIS MOUSQUETAIRES*, IX,
LE VICOMTE DE BRAGELONNE, CCXLIX

The Italian version follows in the next section *Ringraziamenti*.

Someone once told me that the doctorate is not a sprint, but a marathon. They were right. One should take his time to explore the Universe as well as himself. It is a path to be crossed step by step, without any breaks, sometimes enduring and fighting against opponents. However, when I think about the last few years, I tend to linger over the memories of special people, unique places and important events. And those recollections converge to the image of myself on a ship, in the middle of the ocean, being part of one of those journeys you can read in adventure novels. To think that my native lands do not even overlook the sea! Yet that's exactly what happened: I set sail towards what I thought was a safe destination, and I sailed for a long time instead. During the journey I veered several times and I often confronted storms and even sea monsters. Like Ulysses who almost fell pray to the furious and voracious Scylla and got dragged by Charybdis' vortex. In this myth there is a non negligible detail, though. The valiant Ulysses, despite of his talent and inner strength, would never have escaped that dangerous water without the help of his brave crew. I am surely not an Homeric hero, but similarly I managed to cut through stormy seas with the support of extraordinary people. I have been lucky,

and I will always feel enormous gratitude and esteem towards them. Now that we are back on the mainland, it is time for the acknowledgements.

Rainer Wallny was the main mast, the element which supports the entire propulsion system, and therefore requiring exceptional strength and endurance. Antonio Togni was the sail: even though the winds were adverse, he used their force to keep going forward. Leonardo Degiorgi, the compass, has been a constant source of advice and indicated the right direction. Wilfred van Gunsteren was the ship's banner, proudly raised against the enemy, symbolizing the freedom and the truth.

Will Hartley was the lifeboat: a valuable colleague and a friend endowed with rare empathy. And able to appreciate my nerd talks. Maybe.

And then a heartfelt thanks to my mother. She has been the hull, the part which supports the whole ship. Half submerged, because she shared difficulties with me, and half above the surface, to ensure a valid support.

A special thanks goes to Manfred Sigrist and Roland Munz, and to all those people who chose to join the crew, and rowed together with me, with energy and fearlessness. Among them, I would like to mention William Wester, Tesla Jeltema, Sara Bridle and Rachel Mandelbaum. I would like to express my gratitude to Alexandre Refregier, my supervisor and scientific guidance, and to Adam Amara, my co-supervisor. Adam has this contagious cheerfulness that makes you sing "Everything is Awesooome" all together while queuing at customs at the Chicago Airport. And I thank also Kevin Schawinski for his always prompt suggestions and for being a great scientific collaborator.

During a journey you happen to meet other doctoral students and walk together part of the road, and even sharing the office! In this context, I would like to thank Gabriele Cugno and Tim Lichtenberg for the awesome chats and Lia Sartori for her friendship and support. A big thank you to the mythical Jörg Herbel, a mix of absolute rigor and unpredictable goliardery; to Larry Sin, witty, daring and intelligent; and to Raphael Sgier, a sincere friend and a fantastic travel buddy. I would like to thank Raphael also for helping with the coordination of the PhD Lunch, and Larry for being its emblem, always showing an hilarious ravenous expression in front of fried chicken and potatoes. I am glad I could previously organize the lunch with Anna Weigel and Claudio Bruderer. Special thanks to Claudio for being a good collaborator and for the nice conversations. Since the beginning of the pandemic, the PhD Lunch transitioned into a new format, becoming a PhD virtual coffee. Excellent coffee-break companions have been Polychronis Patapis, gentle giant of Salamis and author of delicious Hellenic biscuits; Larry, Raphael and Jörg; Franziska Menti and her cat, the sweet Ann-Christine Vossberg, Jie Ma, Jean Hayoz; Conrad Schwanitz, often live recording from Davos; Sofia Gallego with her cheerfulness and spontaneity; the amazing and unique Saeed; and Timothy Gebhard,

called Tim, a budding chef and youtuber and mad scientist in constant struggle with the aleatory essence of the cosmos and machine learning algorithms. Thank you so much guys for participating! Despite being far away due to an invisible virus, you remained in touch and spiritually close. Thank you to Andrina Nicola, Beatrice Moser, Dominik Zürcher, Janis Fluri, Pascale Berner and Christiane Lorenz, members of the Cosmology Research Group. Special thanks to Silvan Hunziker, with his wall of Coca-Cola cans and the exclusive astronomical almanac. Thank you to the sporty Christian Tschudi and to Jean-Noël Mettler. I would like to thank my students, Silvan Fischbacher, Beatrice Moser, Paul Moser Röggla, Imelda Romero and Ralf Aeberhard. And thank you also to Masato Onodera, Gabriele Pezzulli, Henning Havenaus, Antonio Garufi, Sebastian Seerhars, Stephen March, Anna Boehle, Lena Bori, Maryam Shirazi, Adrian Glauser, Greta Guidi, Benny Trakhtenbrot, Tomasz Kakprackz, Neven Caplar and Kyuseok Oh. Thank you to Bibiana Prinot, Lorenza Della Bruna, Jocelyne Zofia and Benjamin Stritmatter. To Szymon Hennel e Katrin Kröger, thank you very much for your friendship and support. And I would like to thank Arik Jung and the Students' Association. And a big thank you to the Astronomy Allies, too! Thank you also to Manuel Chinotti for the nice chats, to Matteo Cataneo whom Raphael and I met at the conference in Daejeon, South Korea, and to Kyriakos Flouris for the excellent scientific conversations and the curious anecdotes on the Greek language and Cyprus. Special thanks to Esther Christen, Ina Schultz, Christiane Steuer-Heydkamp, Annie Agius, Milena Ramirez, Caroline Keufer and Marianne Chiesi. I also thank Prof. Lavinia Heisenberg for her scientific collaborations and support, and her student and collaborator Stefan Zentarra. And a big thank you to Antonella Palmese, for asking me to collaborate on her paper about the galaxy NGC 4993 and the gravitational wave signal GW170817. I thank Günther Dissertori and Leonardo Degiorgi for our useful conversations about didactics. Special thanks also to Andreas Faisst, for his welcome when I arrived at ETH and for his good friendship. A special mention goes to the amazing Prof. Richard Kron, whom I met during my observation shift at the Blanco Telescope in Chile. I learned a lot about the telescope and galaxies thanks to his teaching. And thank you also to my other co-observers, Prof. Scott Sheppard and Anthony Kremin.

Finally, a big special thanks to Simon Birrer. I believe in life it is not easy to meet genuine people like you, Simon, and I have been so lucky to receive your friendship. Thank you so much!

And Sandro? A special big thanks also to my wise friend Sandro Tacchella. Intelligence and empathy make you a great colleague and a caring friend. For the conversations and your good advice, thank you very much.

So here I am, close to the epilogue. I thank my father, a voracious connoisseur of

Swiss chocolate and a tireless driver along the Turin-Zürich route. At one point he decided to compete with my brother and me for the number of ankle injuries and broken bones. I still don't know who won, but I certainly stand out for creativity. And I thank my big brother, Egidio. When I was a child and he was already a boy, for me his presence was something taken for granted, a sort of law of nature. It was a constant in my daily life, a safe and warm source of leisure and lightness. Now that we are both adults this dynamic has not changed. Although the pandemic keeps us separate, he is always one call away and every time the phone rings and I answer the air is filled with laughter and happiness. It feels like some magic that takes us in a special dimension where there are neither spatial nor temporal boundaries.

And who am I, after this journey? I don't know. I am the same as before, but also a bit different. What did not change along the way is what I carry in my heart. Somehow the memory of my most precious friend has always been with me. I had the privilege of knowing you, Sonia, and the opportunity to learn and acquire a small part of your immense strength. I like to think that somewhere in the Universe we are still two teenagers with many dreams, sitting in the classroom, near the window, talking about Astrophysics, and that you laugh at my jokes while the sun gilds your hair. In that dimension you are able to explore the world with your amber eyes and describe it to me with your silvery voice. In that world you can continue to live your life. And so with affectionate irony and ill-concealed melancholy, through these verses I say hello to you, wherever you are.

Who knows if at this time on Mars,
on Mercury or Neptune,
someone
in a school desk
is looking for the word
that he misses
to begin the essay
on the blank page.

And certainly in the sky of Orion,
Gemini, Leo,
another one forgets
in the inkwell
punctuation marks...
as I do.

I almost feel it
the creak
of a nib
at the end of the firmament:
in a tiny dot
in the Milky Way
a tiny schoolboy
in his history book
draws a puppet.
He knows it is not good,
and I know too:
but we will laugh together
when I meet him.

Gianni Rodari, *Distrazione Intraplanetaria, Filastrocche in cielo e in terra*,
Einaudi, Torino, 1972, p. 30.

Ringraziamenti

Tous pour un, un pour tous!
Tutti per uno, uno per tutti!

ALEXANDRE DUMAS, *LES TROIS MOUSQUETAIRES*, IX,
LE VICOMTE DE BRAGELONNE, CCXLIX

Qualcuno una volta mi disse che il dottorato non è uno sprint, ma una maratona. Aveva ragione. La conoscenza dell'Universo, così come quella di se stessi, si conquista con calma, passo dopo passo, senza mai fermarsi, vincendo a tratti la fatica e lottando contro gli avversari. Tuttavia quando ripenso a questi ultimi anni e indugio nei ricordi di persone particolari, luoghi speciali ed eventi importanti, allora mi immagino su una nave in mezzo al mare, mentre compio un viaggio di quelli che si leggono nei romanzi di avventura e di formazione. E pensare che le mie sabaude terre natie nemmeno lo vedono il mare! Eppure è andata proprio così: sono salpata verso quella che pensavo una meta sicura, e invece ho navigato a lungo, cambiando rotta molte volte, spesso affrontando tempeste e persino mostri marini. Un po' come Ulisse in preda alla furia e ai voraci attacchi di Scilla e sballottato dai micidiali vortici di Cariddi. Ma c'è un dettaglio niente affatto trascurabile. Il valoroso Ulisse, nonostante il suo ingegno e la sua grande forza d'animo, mai sarebbe sfuggito a quelle acque insidiose senza l'aiuto del suo valoroso equipaggio. Ed io, che certo non sono un eroe omerico, similmente ho solcato burrascosi mari con l'aiuto e il sostegno di persone straordinarie. Sono stata fortunata, e sempre esprimerò la mia gratitudine e la mia stima nei loro confronti. Tornando dunque sulla terraferma,

è arrivato il momento di ringraziarle.

Rainer Wallny è stato un po' come l'albero maestro, che sorregge tutto il sistema di propulsione, un elemento che richiede forza e resistenza eccezionali. Antonio Togni, la vela, ha usato l'energia dei venti, se pur avversi, per assicurare il movimento. Leonardo Degiorgi, la bussola, è stato costante fonte di consigli e di giuste direttive. Wilfred van Gunsteren è stato il vessillo della nave, innalzato fieramente contro il nemico, a simboleggiare libertà e verità.

Will Hartley è stato la scialuppa di salvataggio: un valido collega e un amico dotato di straordinaria empatia. E pure capace di apprezzare i miei discorsi nerd. Forse.

E poi vorrei ringraziare tanto, anzi tantissimo, ma che dico, immensamente, mia madre. Lei è stata lo scafo della nave, la parte che sorregge tutto, che assicura il galleggiamento. In parte sommersa, perché ha condiviso con me le difficoltà, e in parte in superficie, per assicurare sempre un valido supporto.

Un grazie particolare anche a Manfred Sigrist e Roland Munz e a tutte quelle persone speciali che hanno scelto di unirsi all'equipaggio, remando insieme a me e senza temere il rischio di esporsi ai venti e alle tempeste. Tra di loro vorrei menzionare William Wester, Tesla Jeltema, Sara Bridle e Rachel Mandelbaum. Grazie anche al mio supervisor e guida scientifica, Alexandre Refregier, e Adam Amara, il mio co-supervisor. Adam ha questa allegria contagiosa che può portarti a cantare con lui "Everything is Awesooooome" in coda alla dogana dell'Aeroporto di Chicago. E ringrazio anche Kevin Schawinski per i suoi ottimi consigli e per le collaborazioni scientifiche.

Quando si è in viaggio capita di incontrare altri studenti di dottorato e percorrere insieme a loro una parte del percorso, e persino condividere l'ufficio! A tal riguardo, mille grazie a Gabriele Cugno e a Tim Lichtenberg per le fantastiche chiacchierate e a Lia Sartori per la sua amicizia e supporto. Grazie al mitico Jörg Herbel, un misto di rigore assoluto e goliardia imprevedibile, a Larry Sin, spiritoso, audace e intelligente, e a Raphael Sgier, un amico sincero e un fantastico compagno di viaggio. A Raphael va anche un grazie per avermi aiutato a organizzare il PhD Lunch e a Larry per esserne stato l'emblema, con la sua tipica espressione famelica di fronte al suo solito pollo fritto e patatine. Ho avuto il piacere di organizzare il PhD Lunch anche con Anna Weigel and con Claudio Bruderer. Un grazie speciale a Claudio per le collaborazioni scientifiche e per le belle conversazioni. Sin dall'inizio della pandemia, il PhD Lunch si è mutato in un PhD coffee digitale. Miei eccellenti compagni di pause caffè sono stati Polychronis Patapis, gigante buono di Salamina e autore di deliziosi biscotti ellenici, Larry, Raphael, Jörg, Franziska Menti e il suo gatto, la dolce Ann-Christine Vossberg, Jie Ma, Jean Hayoz, Conrad Schwanitz, spesso in collegamento da Davos, Sofia Gallego con la sua allegria e spontaneità, il mitico e unico Saeed, e Timothy Gebhard, detto Tim, cuoco e youtuber

in erba e mad scientist in lotta continua con l'aletorietà del cosmo e degli algoritmi di machine learning. Grazie mille ragazzi perché, pur essendo lontani a causa di un virus invisibile, siete rimasti in contatto e spiritualmente vicini. Grazie ad Andrina Nicola, Beatrice Moser, Dominik Zürcher, Janis Fluri, Pascale Berner e Christiane Lorenz, membri del Cosmology Research Group, a Silvan Hunziker con la sua muraglia di lattine di Coca-Cola e l'esclusivo almanacco astronomico, allo sportivo Christian Tschudi e a Jean-Noël Mettler. Grazie ai miei studenti, Silvan Fischbacher, Beatrice Moser, Paul Moser Rögglä, Imelda Romero e Ralf Aeberhard. E poi grazie a Masato Onodera, Gabriele Pezzulli, Henning Havenaus, Antonio Garufi, Sebastian Seerhars, Stephen March, Anna Boehle, Lena Bori, Maryam Shirazi, Adrian Glauser, Greta Guidi, Benny Trakhtenbrot, Tomasz Kakprackz, Neven Caplar e Kyuseok Oh. Grazie anche a Bibiana Prinot, Lorenza Della Bruna, Jocelyne Zofia e Benjamin Strittmatter. A Szymon Hennel e Katrin Kröger, grazie mille per la vostra amicizia e supporto. E vorrei ringraziare anche Arik Jung e l'associazione studentesca. E invio un grosso grazie ai membri di Astronomy Allies. Grazie a Manuel Chinotti per le simpatiche chiacchierate, a Matteo Cataneo che io e Raphael abbiamo conosciuto in conferenza a Daejeon, in Corea del Sud, e a Kyriakos Flouris per le ottime conversazioni scientifiche e gli aneddoti sulla lingua greca e Cipro. Un grazie speciale va ad Esther Christen, Ina Schultz, Christiane Steuer-Heydkamp, Annie Agius, Milena Ramirez, Caroline Keufer and Marianne Chiesi. Ringrazio anche la Prof.ssa Lavinia Heisenberg per le collaborazioni scientifiche e per il suo supporto, e il suo studente e mio collaboratore Stefan Zentarra. E ringrazio Antonella Palmese, per avermi chiesto di collaborare al paper sulla galassia NGC 4993 e il segnale di onda gravitazionale GW170817. Ringrazio anche Günther Dissertori e Leonardo Degiorgi per le utili conversazioni sulla didattica. Grazie pure ad Andreas Faisst, per la sua accoglienza quando sono arrivata all'ETH e per la sua buona amicizia. Un menzione speciale va al mitico Prof. Richard Kron, che ho conosciuto durante il mio turno di osservazioni presso il Blanco Telescope, in Cile. Ho imparato molte cose sul telescopio e sulle galassie grazie ai suoi insegnamenti. E grazie anche agli altri miei colleghi di osservazioni, il Prof. Scott Sheppard e Anthony Kremin.

Infine, un grazie speciale a Simon Birrer. Credo che nella vita non sia facile incontrare persone genuine come te, Simon, ed io ho avuto la fortuna di ricevere la tua amicizia. Grazie infinitamente.

E Sandro? Un grazie speciale anche al mio saggio amico Sandro Tacchella. Intelligenza ed empatia ti rendono un ottimo collega e un amico premuroso. Grazie mille per le chiacchierate e i tuoi buoni consigli.

Eccomi dunque vicina all'epilogo. Ringrazio mio padre, vorace conoscitore del cioccolato svizzero e instancabile autista lungo la tratta Torino-Zurigo. A un certo punto ha

deciso di competere con me e mio fratello per numero di infortuni alle caviglie e ossa rotte. Non so ancora bene chi ha vinto, ma certamente io primeggio per originalità. E ringrazio il mio fratellone, Egidio. Quando ero una bambina e lui già un ragazzino, per me la sua presenza era qualcosa di scontato, una sorta di legge di natura. Era una costante delle mie giornate, sicura e calorosa fonte di svago e leggerezza. Ora che siamo entrambi adulti questa dinamica non è cambiata. Nonostante la pandemia ci tenga separati, lui è sempre a una chiamata di distanza e ogni volta che il telefono squilla e io rispondo l'aria si riempie di risate e allegria. È un po' come una magia che ci porta in una dimensione speciale che non conosce confini spaziali e temporali.

E chi sono io, dopo questo viaggio? Ancora non lo so. Sono la stessa di prima, ma anche un po' diversa. Ciò che non è mutato durante il percorso è quello che mi porto nel cuore. E il ricordo dell'amica più importante in qualche modo è stato sempre con me. Ho avuto il privilegio di conoscerti, Sonia, e la possibilità di imparare e di acquisire una piccola parte della tua forza immensa. Mi piace pensare che da qualche parte nell'Universo siamo ancora due adolescenti con tanti sogni, sedute in classe vicino alla finestra a parlare di Astrofisica, e che tu ridi ai miei scherzi mentre il sole ti indora i capelli. In una realtà che puoi ancora esplorare con i tuoi occhi color dell'ambra e descrivermi con la tua voce argentina, un mondo dove puoi continuare a vivere la tua vita. E dunque con ironia affettuosa e mal celata malinconia, con questi versi ti mando un saluto, ovunque tu sia.

Chissà se a quest'ora su Marte,
su Mercurio o Nettuno,
qualcuno
in un banco di scuola
sta cercando la parola
che gli manca
per cominciare il tema
sulla pagina bianca.

E certo nel cielo di Orione,
dei Gemelli, del Leone,
un altro dimentica
nel calamaio
i segni d'interpunzione...
come faccio io.

Quasi lo sento

lo scricchiolio
di un pennino
in fondo al firmamento:
in un minuscolo puntino
nella Via Lattea
un minuscolo scolareto
sul suo libro di storia
disegna un pupazzetto.
Lo sa che non sta bene,
e anch'io lo so:
ma rideremo insieme
quando lo incontrerò.

Gianni Rodari, *Distrazione Intraplanetaria, Filastrocche in cielo e in terra*,
Einaudi, Torino, 1972, p. 30.

Bibliography

- Abbott, T. et al. (Aug. 2016). “The Dark Energy Survey: more than dark energy - an overview”. In: *MNRAS* 460, pp. 1270–1299. DOI: [10.1093/mnras/stw641](https://doi.org/10.1093/mnras/stw641). arXiv: [1601.00329](https://arxiv.org/abs/1601.00329).
- Abbott, T. M. C. et al. (2021). *The Dark Energy Survey Data Release 2*. arXiv: [2101.05765](https://arxiv.org/abs/2101.05765) [astro-ph.IM].
- Abraham, R. G., S. van den Bergh, K. Glazebrook, et al. (Nov. 1996). “The Morphologies of Distant Galaxies. II. Classifications from the Hubble Space Telescope Medium Deep Survey”. In: *ApJS* 107, p. 1. DOI: [10.1086/192352](https://doi.org/10.1086/192352).
- Abraham, R. G., S. van den Bergh, and P. Nair (May 2003). “A New Approach to Galaxy Morphology. I. Analysis of the Sloan Digital Sky Survey Early Data Release”. In: *ApJ* 588, pp. 218–229. DOI: [10.1086/373919](https://doi.org/10.1086/373919). eprint: [astro-ph/0301239](https://arxiv.org/abs/astro-ph/0301239).
- Abraham, Sheelu et al. (Mar. 2018). “Detection of bars in galaxies using a deep convolutional neural network”. In: *Monthly Notices of the Royal Astronomical Society* 477.1, pp. 894–903. ISSN: 0035-8711. DOI: [10.1093/mnras/sty627](https://doi.org/10.1093/mnras/sty627). eprint: <https://academic.oup.com/mnras/article-pdf/477/1/894/24654308/sty627.pdf>. URL: <https://doi.org/10.1093/mnras/sty627>.
- Aghanim, N. et al. (Sept. 2020). “Planck 2018 results”. In: *Astronomy Astrophysics* 641, A6. ISSN: 1432-0746. DOI: [10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910). URL: <http://dx.doi.org/10.1051/0004-6361/201833910>.
- Allen, Paul D. et al. (Aug. 2006). “The Millennium Galaxy Catalogue: bulge–disc decomposition of 10095 nearby galaxies”. In: *Monthly Notices of the Royal Astronomical Society* 371.1, pp. 2–18. ISSN: 0035-8711. DOI: [10.1111/j.1365-2966.2006.10586.x](https://doi.org/10.1111/j.1365-2966.2006.10586.x).

- eprint: <https://academic.oup.com/mnras/article-pdf/371/1/2/3017788/mnras0371-0002.pdf>. URL: <https://doi.org/10.1111/j.1365-2966.2006.10586.x>.
- Amiaux, J. et al. (Aug. 2012). “Euclidmission: building of a reference survey”. In: *Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave*. Ed. by Mark C. Clampin et al. DOI: [10.1117/12.926513](https://doi.org/10.1117/12.926513). URL: <http://dx.doi.org/10.1117/12.926513>.
- Baldry, I. K. et al. (Dec. 2004). “Color bimodality: Implications for galaxy evolution”. In: *The New Cosmology: Conference on Strings and Cosmology*. Ed. by Roland E. Allen, Dimitri V. Nanopoulos, and Christopher N. Pope. Vol. 743. American Institute of Physics Conference Series, pp. 106–119. DOI: [10.1063/1.1848322](https://doi.org/10.1063/1.1848322). arXiv: [astro-ph/0410603](https://arxiv.org/abs/astro-ph/0410603) [[astro-ph](#)].
- Bamford, S. P. et al. (Mar. 2009). “Galaxy Zoo: the dependence of morphology and colour on environment”. In: *MNRAS* 393, pp. 1324–1352. DOI: [10.1111/j.1365-2966.2008.14252.x](https://doi.org/10.1111/j.1365-2966.2008.14252.x). arXiv: [0805.2612](https://arxiv.org/abs/0805.2612).
- Banerji, M. et al. (July 2010). “Galaxy Zoo: reproducing galaxy morphologies via machine learning”. In: *MNRAS* 406, pp. 342–353. DOI: [10.1111/j.1365-2966.2010.16713.x](https://doi.org/10.1111/j.1365-2966.2010.16713.x). arXiv: [0908.2033](https://arxiv.org/abs/0908.2033).
- Barden, M. et al. (May 2012). “GALAPAGOS: from pixels to parameters”. In: *MNRAS* 422, pp. 449–468. DOI: [10.1111/j.1365-2966.2012.20619.x](https://doi.org/10.1111/j.1365-2966.2012.20619.x). arXiv: [1203.1831](https://arxiv.org/abs/1203.1831) [[astro-ph.IM](#)].
- Barnes, J. E. (Aug. 1988). “Encounters of disk/halo galaxies”. In: *ApJ* 331, pp. 699–717. DOI: [10.1086/166593](https://doi.org/10.1086/166593).
- Bartelmann, Matthias, Johannes Dombrowski, et al. (2020). *Kinetic field theory: Non-linear cosmic power spectra in the mean-field approximation*. arXiv: [2011.04979](https://arxiv.org/abs/2011.04979) [[astro-ph.CO](#)].
- Bartelmann, Matthias, Felix Fabis, Daniel Berg, et al. (Apr. 2016). “A microscopic, non-equilibrium, statistical field theory for cosmic structure formation”. In: *New Journal of Physics* 18.4, 043020, p. 043020. DOI: [10.1088/1367-2630/18/4/043020](https://doi.org/10.1088/1367-2630/18/4/043020). arXiv: [1411.0806](https://arxiv.org/abs/1411.0806) [[cond-mat.stat-mech](#)].
- Bartelmann, Matthias, Felix Fabis, Elena Kozlikin, et al. (Aug. 2017). “Kinetic field theory: effects of momentum correlations on the cosmic density-fluctuation power spectrum”. In: *New Journal of Physics* 19.8, 083001, p. 083001. DOI: [10.1088/1367-2630/aa7e6f](https://doi.org/10.1088/1367-2630/aa7e6f). arXiv: [1611.09503](https://arxiv.org/abs/1611.09503) [[astro-ph.CO](#)].
- Bartelmann, Matthias, Elena Kozlikin, et al. (2019a). “Cosmic Structure Formation with Kinetic Field Theory”. In: *Annalen der Physik* 531.11, p. 1800446. DOI: <https://doi.org/10.1002/andp.201800446>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.201800446>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.201800446>.

-
- Bartelmann, Matthias, Elena Kozlikin, et al. (Nov. 2019b). “Cosmic Structure Formation with Kinetic Field Theory”. In: *Annalen der Physik* 531.11, p. 1800446. DOI: [10.1002/andp.201800446](https://doi.org/10.1002/andp.201800446). arXiv: [1905.01179](https://arxiv.org/abs/1905.01179) [astro-ph.CO].
- Bartelmann, Matthias and Matteo Maturi (Jan. 2017). “Weak gravitational lensing”. In: *Scholarpedia* 12.1, p. 32440. DOI: [10.4249/scholarpedia.32440](https://doi.org/10.4249/scholarpedia.32440). arXiv: [1612.06535](https://arxiv.org/abs/1612.06535) [astro-ph.CO].
- Bell, E. F. et al. (July 2012). “What Turns Galaxies Off? The Different Morphologies of Star-forming and Quiescent Galaxies since $z \sim 2$ from CANDELS”. In: *ApJ* 753, 167, p. 167. DOI: [10.1088/0004-637X/753/2/167](https://doi.org/10.1088/0004-637X/753/2/167). arXiv: [1110.3786](https://arxiv.org/abs/1110.3786).
- Bergé, J. et al. (Feb. 2013). “An Ultra Fast Image Generator (UFIG) for wide-field astronomy”. In: *Astronomy and Computing* 1, pp. 23–32. DOI: [10.1016/j.ascom.2013.01.001](https://doi.org/10.1016/j.ascom.2013.01.001). arXiv: [1209.1200](https://arxiv.org/abs/1209.1200) [astro-ph.IM].
- Bernstein, R. A., W. L. Freedman, and B. F. Madore (May 2002). “The First Detections of the Extragalactic Background Light at 3000, 5500, and 8000 Å. I. Results”. In: *ApJ* 571, pp. 56–84. DOI: [10.1086/339422](https://doi.org/10.1086/339422). eprint: [astro-ph/0112153](https://arxiv.org/abs/astro-ph/0112153).
- Bershady, M. A., A. Jangren, and C. J. Conselice (June 2000). “Structural and Photometric Classification of Galaxies. I. Calibration Based on a Nearby Galaxy Sample”. In: *AJ* 119, pp. 2645–2663. DOI: [10.1086/301386](https://doi.org/10.1086/301386). eprint: [astro-ph/0002262](https://arxiv.org/abs/astro-ph/0002262).
- Bertin, E. (Jan. 2013). *PSFEx: Point Spread Function Extractor*. Astrophysics Source Code Library. ascl: [1301.001](https://ascl.net/1301.001).
- Bertin, E. and S. Arnouts (June 1996). “SExtractor: Software for source extraction.” In: *A&AS* 117, pp. 393–404. DOI: [10.1051/aas:1996164](https://doi.org/10.1051/aas:1996164).
- 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.
- Bertschinger, Edmund (June 1995). “COSMICS: Cosmological Initial Conditions and Microwave Anisotropy Codes”. In: *arXiv e-prints*, astro-ph/9506070, astro-ph/9506070. arXiv: [astro-ph/9506070](https://arxiv.org/abs/astro-ph/9506070) [astro-ph].
- Blanton, M. R. et al. (Sept. 2003). “The Broadband Optical Properties of Galaxies with Redshifts 0.02z0.22”. In: *ApJ* 594, pp. 186–207. DOI: [10.1086/375528](https://doi.org/10.1086/375528). eprint: [astro-ph/0209479](https://arxiv.org/abs/astro-ph/0209479).
- Bournaud, F., B. G. Elmegreen, and D. M. Elmegreen (Nov. 2007). “Rapid Formation of Exponential Disks and Bulges at High Redshift from the Dynamical Evolution of Clump-Cluster and Chain Galaxies”. In: *ApJ* 670, pp. 237–248. DOI: [10.1086/522077](https://doi.org/10.1086/522077). arXiv: [0708.0306](https://arxiv.org/abs/0708.0306).
- Bradley, Larry et al. (Sept. 2020). *astropy/photutils: 1.0.0*. Version 1.0.0. DOI: [10.5281/zenodo.4044744](https://doi.org/10.5281/zenodo.4044744). URL: <https://doi.org/10.5281/zenodo.4044744>.

- Bruderer, C. et al. (Jan. 2016). “Calibrated Ultra Fast Image Simulations for the Dark Energy Survey”. In: *ApJ* 817, 25, p. 25. DOI: [10.3847/0004-637X/817/1/25](https://doi.org/10.3847/0004-637X/817/1/25). arXiv: [1504.02778](https://arxiv.org/abs/1504.02778).
- Busha, M. T. et al. (Jan. 2013). “Catalog Production for the DES Blind Cosmology Challenge”. In: *American Astronomical Society Meeting Abstracts #221*. Vol. 221. American Astronomical Society Meeting Abstracts, 341.07, p. 341.07.
- Cameron, E. et al. (Nov. 2010). “Bars in early- and late-type discs in COSMOS”. In: *MNRAS* 409, pp. 346–354. DOI: [10.1111/j.1365-2966.2010.17314.x](https://doi.org/10.1111/j.1365-2966.2010.17314.x). arXiv: [1001.1736](https://arxiv.org/abs/1001.1736).
- Cano-Díaz, M et al. (July 2019). “SDSS-IV MaNGA: effects of morphology in the global and local star formation main sequences”. In: *Monthly Notices of the Royal Astronomical Society* 488.3, pp. 3929–3948. ISSN: 0035-8711. DOI: [10.1093/mnras/stz1894](https://doi.org/10.1093/mnras/stz1894). eprint: <https://academic.oup.com/mnras/article-pdf/488/3/3929/29102232/stz1894.pdf>. URL: <https://doi.org/10.1093/mnras/stz1894>.
- Chang, C. et al. (Mar. 2015). “Modeling the Transfer Function for the Dark Energy Survey”. In: *ApJ* 801, 73, p. 73. DOI: [10.1088/0004-637X/801/2/73](https://doi.org/10.1088/0004-637X/801/2/73). arXiv: [1411.0032](https://arxiv.org/abs/1411.0032) [[astro-ph.IM](https://arxiv.org/abs/1411.0032)].
- Cheng, Ting-Yun et al. (Sept. 2020). “Beyond the Hubble Sequence – Exploring Galaxy Morphology with Unsupervised Machine Learning”. In: *arXiv e-prints*, arXiv:2009.11932, arXiv:2009.11932. arXiv: [2009.11932](https://arxiv.org/abs/2009.11932) [[astro-ph.GA](https://arxiv.org/abs/2009.11932)].
- Chisari, Nora Elisa et al. (May 2019). “Core Cosmology Library: Precision Cosmological Predictions for LSST”. In: *ApJS* 242.1, 2, p. 2. DOI: [10.3847/1538-4365/ab1658](https://doi.org/10.3847/1538-4365/ab1658). arXiv: [1812.05995](https://arxiv.org/abs/1812.05995) [[astro-ph.CO](https://arxiv.org/abs/1812.05995)].
- Combes, F. and R. H. Sanders (Mar. 1981a). “Formation and properties of persisting stellar bars”. In: *A&A* 96, pp. 164–173.
- (Mar. 1981b). “Formation and properties of persisting stellar bars.” In: *A&A* 96, pp. 164–173.
- Conselice, C. J. (Nov. 1997). “The Symmetry, Color, and Morphology of Galaxies”. In: *PASP* 109, pp. 1251–1255. DOI: [10.1086/134004](https://doi.org/10.1086/134004). eprint: [astro-ph/9710234](https://arxiv.org/abs/astro-ph/9710234).
- (July 2003). “The Relationship between Stellar Light Distributions of Galaxies and Their Formation Histories”. In: *ApJS* 147, pp. 1–28. DOI: [10.1086/375001](https://doi.org/10.1086/375001). eprint: [astro-ph/0303065](https://arxiv.org/abs/astro-ph/0303065).
- (June 2008). “The Assembly History of Massive Galaxies: What Do We Know?” In: *Pathways Through an Eclectic Universe*. Ed. by J. H. Knapen, T. J. Mahoney, and A. Vazdekis. Vol. 390. Astronomical Society of the Pacific Conference Series, p. 403. arXiv: [0706.3482](https://arxiv.org/abs/0706.3482).
- Conselice, C. J., M. A. Bershadsky, and A. Jangren (Feb. 2000). “The Asymmetry of Galaxies: Physical Morphology for Nearby and High-Redshift Galaxies”. In: *ApJ* 529, pp. 886–910. DOI: [10.1086/308300](https://doi.org/10.1086/308300). eprint: [astro-ph/9907399](https://arxiv.org/abs/astro-ph/9907399).

-
- Conselice, C. J., S. Rajgor, and R. Myers (May 2008). “The structures of distant galaxies - I. Galaxy structures and the merger rate to $z \sim 3$ in the Hubble Ultra-Deep Field”. In: *MNRAS* 386, pp. 909–927. DOI: [10.1111/j.1365-2966.2008.13069.x](https://doi.org/10.1111/j.1365-2966.2008.13069.x). arXiv: [0711.2333](https://arxiv.org/abs/0711.2333).
- Cooray, Asantha and Ravi Sheth (Dec. 2002). “Halo models of large scale structure”. In: *Phys. Rep.* 372.1, pp. 1–129. DOI: [10.1016/S0370-1573\(02\)00276-4](https://doi.org/10.1016/S0370-1573(02)00276-4). arXiv: [astro-ph/0206508](https://arxiv.org/abs/astro-ph/0206508) [[astro-ph](#)].
- Dark Energy Survey Collaboration et al. (Aug. 2016). “The Dark Energy Survey: more than dark energy - an overview”. In: *MNRAS* 460.2, pp. 1270–1299. DOI: [10.1093/mnras/stw641](https://doi.org/10.1093/mnras/stw641). arXiv: [1601.00329](https://arxiv.org/abs/1601.00329) [[astro-ph.CO](#)].
- de Jong, R. S. (Sept. 1996a). “Near-infrared and optical broadband surface photometry of 86 face-on disk dominated galaxies. III. The statistics of the disk and bulge parameters.” In: *A&A* 313, pp. 45–64. eprint: [astro-ph/9601005](https://arxiv.org/abs/astro-ph/9601005).
- (Sept. 1996b). “Near-infrared and optical broadband surface photometry of 86 face-on disk dominated galaxies. III. The statistics of the disk and bulge parameters.” In: *A&A* 313, pp. 45–64. arXiv: [astro-ph/9601005](https://arxiv.org/abs/astro-ph/9601005) [[astro-ph](#)].
- Dekel, A., Y. Birnboim, et al. (Jan. 2009). “Cold streams in early massive hot haloes as the main mode of galaxy formation”. In: *Nature* 457, pp. 451–454. DOI: [10.1038/nature07648](https://doi.org/10.1038/nature07648). arXiv: [0808.0553](https://arxiv.org/abs/0808.0553).
- Dekel, A., R. Sari, and D. Ceverino (Sept. 2009). “Formation of Massive Galaxies at High Redshift: Cold Streams, Clumpy Disks, and Compact Spheroids”. In: *ApJ* 703, pp. 785–801. DOI: [10.1088/0004-637X/703/1/785](https://doi.org/10.1088/0004-637X/703/1/785). arXiv: [0901.2458](https://arxiv.org/abs/0901.2458) [[astro-ph.GA](#)].
- DES Collaboration (Oct. 2005). “The Dark Energy Survey”. In: *ArXiv Astrophysics e-prints*. eprint: [astro-ph/0510346](https://arxiv.org/abs/astro-ph/0510346).
- Desai, S. et al. (Sept. 2012). “The Blanco Cosmology Survey: Data Acquisition, Processing, Calibration, Quality Diagnostics, and Data Release”. In: *ApJ* 757, 83, p. 83. DOI: [10.1088/0004-637X/757/1/83](https://doi.org/10.1088/0004-637X/757/1/83). arXiv: [1204.1210](https://arxiv.org/abs/1204.1210) [[astro-ph.CO](#)].
- Diehl, H. T., T. M. C. Abbott, J. Annis, and R. Armstrong (2012). “The Dark Energy Survey Camera (DECam)”. In: *Physics Procedia* 37, pp. 1332–1340. DOI: [10.1016/j.phpro.2012.02.472](https://doi.org/10.1016/j.phpro.2012.02.472).
- Diehl, H. T., T. M. C. Abbott, J. Annis, R. Armstrong, et al. (Aug. 2014). “The Dark Energy Survey and operations: Year 1”. In: *Observatory Operations: Strategies, Processes, and Systems V*. Vol. 9149. Proc. SPIE, 91490V, p. 91490V. DOI: [10.1117/12.2056982](https://doi.org/10.1117/12.2056982).
- Dieleman, S., K. W. Willett, and J. Dambre (June 2015). “Rotation-invariant convolutional neural networks for galaxy morphology prediction”. In: *MNRAS* 450, pp. 1441–1459. DOI: [10.1093/mnras/stv632](https://doi.org/10.1093/mnras/stv632). arXiv: [1503.07077](https://arxiv.org/abs/1503.07077) [[astro-ph.IM](#)].
- DijkVanvan Dijk, Thomas (1902). “An important paper”. In: *QJRAS* 2, pp. 202–207.
- Dodelson, Scott (2003). *Modern cosmology*.

- Dodelson, Scott and Michael D. Schneider (Sept. 2013). “The effect of covariance estimator error on cosmological parameter constraints”. In: *Phys. Rev. D* 88.6, 063537, p. 063537. DOI: [10.1103/PhysRevD.88.063537](https://doi.org/10.1103/PhysRevD.88.063537). arXiv: [1304.2593](https://arxiv.org/abs/1304.2593) [[astro-ph.CO](#)].
- Doran, Michael (Oct. 2005). “CMBEASY: an object oriented code for the cosmic microwave background”. In: *Journal of Cosmology and Astro-Particle Physics* 2005.10, 011, p. 011. DOI: [10.1088/1475-7516/2005/10/011](https://doi.org/10.1088/1475-7516/2005/10/011). arXiv: [astro-ph/0302138](https://arxiv.org/abs/astro-ph/0302138) [[astro-ph](#)].
- Dressler, A. (Mar. 1980). “Galaxy morphology in rich clusters - Implications for the formation and evolution of galaxies”. In: *ApJ* 236, pp. 351–365. DOI: [10.1086/157753](https://doi.org/10.1086/157753).
- Drlica-Wagner, A., I. Sevilla-Noarbe, E. S. Rykoff, R. A. Gruendl, B. Yanny, D. L. Tucker, B. Hoyle, A. Carnero Rosell, et al. (Aug. 2017). “Dark Energy Survey Year 1 Results: Photometric Data Set for Cosmology”. In: *ArXiv e-prints*. arXiv: [1708.01531](https://arxiv.org/abs/1708.01531).
- Drlica-Wagner, A., I. Sevilla-Noarbe, E. S. Rykoff, R. A. Gruendl, B. Yanny, D. L. Tucker, B. Hoyle, A. Carnero Rosell, et al. (Apr. 2018). “Dark Energy Survey Year 1 Results: The Photometric Data Set for Cosmology”. In: *The Astrophysical Journal Supplement Series* 235.2, p. 33. ISSN: 1538-4365. DOI: [10.3847/1538-4365/aab4f5](https://doi.org/10.3847/1538-4365/aab4f5). URL: <http://dx.doi.org/10.3847/1538-4365/aab4f5>.
- Eisenstein, D. J. and W. Hu (Mar. 1998). “Baryonic Features in the Matter Transfer Function”. In: *ApJ* 496, pp. 605–614. DOI: [10.1086/305424](https://doi.org/10.1086/305424). eprint: [astro-ph/9709112](https://arxiv.org/abs/astro-ph/9709112).
- Elmegreen, B. G. et al. (June 1996). “Light Profiles and Pattern Speeds for Bars in Early- and Late-Type Galaxies”. In: *AJ* 111, p. 2233. DOI: [10.1086/117957](https://doi.org/10.1086/117957).
- Elmegreen, Bruce G. et al. (June 1996). “Light Profiles and Pattern Speeds for Bars in Early- and Late-Type Galaxies”. In: *AJ* 111, p. 2233. DOI: [10.1086/117957](https://doi.org/10.1086/117957).
- Fall, S. M. and G. Efstathiou (Oct. 1980). “Formation and rotation of disc galaxies with haloes”. In: *MNRAS* 193, pp. 189–206. DOI: [10.1093/mnras/193.2.189](https://doi.org/10.1093/mnras/193.2.189).
- Fang, Xiao et al. (May 2020). “Beyond Limber: efficient computation of angular power spectra for galaxy clustering and weak lensing”. In: *J. Cosmology Astropart. Phys.* 2020.5, 010, p. 010. DOI: [10.1088/1475-7516/2020/05/010](https://doi.org/10.1088/1475-7516/2020/05/010). arXiv: [1911.11947](https://arxiv.org/abs/1911.11947) [[astro-ph.CO](#)].
- Fisher, D. B. and N. Drory (Aug. 2008). “The Structure of Classical Bulges and Pseudobulges: the Link Between Pseudobulges and Sérsic Index”. In: *AJ* 136, pp. 773–839. DOI: [10.1088/0004-6256/136/2/773](https://doi.org/10.1088/0004-6256/136/2/773). arXiv: [0805.4206](https://arxiv.org/abs/0805.4206).
- Flaugher, B. (2005). “The Dark Energy Survey”. In: *International Journal of Modern Physics A* 20, pp. 3121–3123. DOI: [10.1142/S0217751X05025917](https://doi.org/10.1142/S0217751X05025917).
- Flaugher, B., H. T. Diehl, K. Honscheid, T. M. C. Abbott, O. Alvarez, R. Angstadt, J. T. Annis, M. Antonik, O. Ballester, L. Beaufore, and et al. (Oct. 2015). “THE DARK ENERGY CAMERA”. In: *The Astronomical Journal* 150.5, p. 150. ISSN: 1538-3881. DOI: [10.1088/0004-6256/150/5/150](https://doi.org/10.1088/0004-6256/150/5/150). URL: <http://dx.doi.org/10.1088/0004-6256/150/5/150>.

-
- Flaugher, B., H. T. Diehl, K. Honscheid, T. M. C. Abbott, O. Alvarez, R. Angstadt, J. T. Annis, M. Antonik, O. Ballester, L. Beaufore, G. M. Bernstein, et al. (Nov. 2015). “The Dark Energy Camera”. In: *AJ* 150.5, 150, p. 150. DOI: [10.1088/0004-6256/150/5/150](https://doi.org/10.1088/0004-6256/150/5/150). arXiv: [1504.02900](https://arxiv.org/abs/1504.02900) [[astro-ph.IM](#)].
- Foreman-Mackey, Daniel et al. (Mar. 2013). “emcee: The MCMC Hammer”. In: *PASP* 125.925, p. 306. DOI: [10.1086/670067](https://doi.org/10.1086/670067). arXiv: [1202.3665](https://arxiv.org/abs/1202.3665) [[astro-ph.IM](#)].
- Förster Schreiber, N. M. et al. (Apr. 2006). “Faint Infrared Extragalactic Survey: Data and Source Catalog of the MS 1054-03 Field”. In: *AJ* 131, pp. 1891–1913. DOI: [10.1086/497293](https://doi.org/10.1086/497293). eprint: [astro-ph/0510186](https://arxiv.org/abs/astro-ph/0510186).
- Fournier, Paul (1901). “Not a Real Paper”. In: *ApJ* 1, pp. 101–111.
- Freeman, P. E. et al. (Sept. 2013). “New image statistics for detecting disturbed galaxy morphologies at high redshift”. In: *MNRAS* 434, pp. 282–295. DOI: [10.1093/mnras/stt1016](https://doi.org/10.1093/mnras/stt1016). arXiv: [1306.1238](https://arxiv.org/abs/1306.1238).
- Friedman, Jerome H. (2001). “Greedy function approximation: A gradient boosting machine.” In: *The Annals of Statistics* 29.5, pp. 1189–1232. DOI: [10.1214/aos/1013203451](https://doi.org/10.1214/aos/1013203451). URL: <https://doi.org/10.1214/aos/1013203451>.
- Genzel, R. et al. (Nov. 2008). “From Rings to Bulges: Evidence for Rapid Secular Galaxy Evolution at $z \sim 2$ from Integral Field Spectroscopy in the SINS Survey”. In: *ApJ* 687, 59–77, pp. 59–77. DOI: [10.1086/591840](https://doi.org/10.1086/591840). arXiv: [0807.1184](https://arxiv.org/abs/0807.1184).
- Ghosh, Aritra et al. (June 2020). “Galaxy Morphology Network: A Convolutional Neural Network Used to Study Morphology and Quenching in $\sim 100,000$ SDSS and $\sim 20,000$ CANDELS Galaxies”. In: *ApJ* 895.2, 112, p. 112. DOI: [10.3847/1538-4357/ab8a47](https://doi.org/10.3847/1538-4357/ab8a47). arXiv: [2006.14639](https://arxiv.org/abs/2006.14639) [[astro-ph.GA](#)].
- Goncalves, Thiago et al. (Aug. 2012). “Quenching star formation at intermediate redshifts: downsizing of the mass flux density in the green valley”. In: *Proceedings of the International Astronomical Union* 8, pp. 163–166. DOI: [10.1017/S1743921313004572](https://doi.org/10.1017/S1743921313004572).
- Graham, A. W. and S. P. Driver (2005). “A Concise Reference to (Projected) Sérsic $R^{1/n}$ Quantities, Including Concentration, Profile Slopes, Petrosian Indices, and Kron Magnitudes”. In: *Publ. Astron. Soc. Australia* 22, pp. 118–127. DOI: [10.1071/AS05001](https://doi.org/10.1071/AS05001). eprint: [astro-ph/0503176](https://arxiv.org/abs/astro-ph/0503176).
- Graham, Alistair W. (Aug. 2019). “A galaxy classification grid that better recognizes early-type galaxy morphology”. In: *MNRAS* 487.4, pp. 4995–5009. DOI: [10.1093/mnras/stz1623](https://doi.org/10.1093/mnras/stz1623). arXiv: [1907.09791](https://arxiv.org/abs/1907.09791) [[astro-ph.GA](#)].
- Gray, M. E. et al. (Mar. 2009). “STAGES: the Space Telescope A901/2 Galaxy Evolution Survey”. In: *MNRAS* 393, pp. 1275–1301. DOI: [10.1111/j.1365-2966.2008.14259.x](https://doi.org/10.1111/j.1365-2966.2008.14259.x). arXiv: [0811.3890](https://arxiv.org/abs/0811.3890).
- Grogin, Norman A. et al. (Dec. 2011). “CANDELS: THE COSMIC ASSEMBLY NEAR-INFRARED DEEP EXTRAGALACTIC LEGACY SURVEY”. In: *The Astrophysical Jour-*

- nal Supplement Series* 197.2, p. 35. ISSN: 1538-4365. DOI: [10.1088/0067-0049/197/2/35](https://doi.org/10.1088/0067-0049/197/2/35). URL: <http://dx.doi.org/10.1088/0067-0049/197/2/35>.
- GuardeDe lade la Garde, Simon (1904). “A Last Unreal Paper”. In: *MNRAS* 4, pp. 404–481.
- Hamilton, A. J. S. (Feb. 2000). “Uncorrelated modes of the non-linear power spectrum”. In: *Monthly Notices of the Royal Astronomical Society* 312.2, pp. 257–284. ISSN: 0035-8711. DOI: [10.1046/j.1365-8711.2000.03071.x](https://doi.org/10.1046/j.1365-8711.2000.03071.x). eprint: <https://academic.oup.com/mnras/article-pdf/312/2/257/2945062/312-2-257.pdf>. URL: <https://doi.org/10.1046/j.1365-8711.2000.03071.x>.
- (Apr. 2001). “Formulae for growth factors in expanding universes containing matter and a cosmological constant”. In: *Monthly Notices of the Royal Astronomical Society* 322.2, pp. 419–425. ISSN: 1365-2966. DOI: [10.1046/j.1365-8711.2001.04137.x](https://doi.org/10.1046/j.1365-8711.2001.04137.x). URL: <http://dx.doi.org/10.1046/j.1365-8711.2001.04137.x>.
- Hartlap, J., P. Simon, and P. Schneider (Mar. 2007). “Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix”. In: *A&A* 464.1, pp. 399–404. DOI: [10.1051/0004-6361:20066170](https://doi.org/10.1051/0004-6361:20066170). arXiv: [astro-ph/0608064](https://arxiv.org/abs/astro-ph/0608064) [[astro-ph](https://arxiv.org/abs/astro-ph)].
- Hartley, W. G. et al. (Dec. 2020). “Dark Energy Survey Year 3 Results: Deep Field Optical + Near-Infrared Images and Catalogue”. In: *arXiv e-prints*, arXiv:2012.12824, arXiv:2012.12824. arXiv: [2012.12824](https://arxiv.org/abs/2012.12824) [[astro-ph.CO](https://arxiv.org/abs/astro-ph)].
- Häußler, B. et al. (July 2011). “Galapagos: A Semi-Automated Tool for Galaxy Profile Fitting”. 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. 155.
- Häussler, B. et al. (Oct. 2007). “GEMS: Galaxy Fitting Catalogs and Testing Parametric Galaxy Fitting Codes: GALFIT and GIM2D”. In: *ApJS* 172, pp. 615–633. DOI: [10.1086/518836](https://doi.org/10.1086/518836). arXiv: [0704.2601](https://arxiv.org/abs/0704.2601).
- High, F. W. et al. (May 2009). “Stellar Locus Regression: Calibrate Photometry Accurately within Minutes of Readout”. In: *American Astronomical Society Meeting Abstracts #214*. Vol. 214. American Astronomical Society Meeting Abstracts, 413.07, p. 680.
- Hu, Wayne and Scott Dodelson (2002). “Cosmic Microwave Background Anisotropies”. In: *Annual Review of Astronomy and Astrophysics* 40.1, pp. 171–216. DOI: [10.1146/annurev.astro.40.060401.093926](https://doi.org/10.1146/annurev.astro.40.060401.093926). eprint: <https://doi.org/10.1146/annurev.astro.40.060401.093926>. URL: <https://doi.org/10.1146/annurev.astro.40.060401.093926>.
- Huertas-Company, M., R. Gravet, et al. (Oct. 2015a). “A CATALOG OF VISUAL-LIKE MORPHOLOGIES IN THE 5 CANDELS FIELDS USING DEEP LEARNING”. In: *The Astrophysical Journal Supplement Series* 221.1, p. 8. DOI: [10.1088/0067-0049/221/1/8](https://doi.org/10.1088/0067-0049/221/1/8). URL: <https://doi.org/10.1088/0067-0049/221/1/8>.

-
- Huertas-Company, M., R. Gravet, et al. (Nov. 2015b). “A Catalog of Visual-like Morphologies in the 5 CANDELS Fields Using Deep Learning”. In: *ApJS* 221, 8, p. 8. DOI: [10.1088/0067-0049/221/1/8](https://doi.org/10.1088/0067-0049/221/1/8). arXiv: [1509.05429](https://arxiv.org/abs/1509.05429).
- Huertas-Company, M., D. Rouan, et al. (Feb. 2008). “A robust morphological classification of high-redshift galaxies using support vector machines on seeing limited images. I. Method description”. In: *A&A* 478, pp. 971–980. DOI: [10.1051/0004-6361:20078625](https://doi.org/10.1051/0004-6361:20078625). arXiv: [0709.1359](https://arxiv.org/abs/0709.1359).
- Jedrzejewski, Robert I. (June 1987). “CCD surface photometry of elliptical galaxies - I. Observations, reduction and results.” In: *MNRAS* 226, pp. 747–768. DOI: [10.1093/mnras/226.4.747](https://doi.org/10.1093/mnras/226.4.747).
- Jogee, S. (Dec. 2009). “Galaxy Mergers and Their Impact over the Last 8 Gyr”. In: *Galaxy Evolution: Emerging Insights and Future Challenges*. Ed. by S. Jogee et al. Vol. 419. Astronomical Society of the Pacific Conference Series, p. 204.
- Jogee, S. et al. (June 2009). “History of Galaxy Interactions and Their Impact on Star Formation Over the Last 7 Gyr from GEMS”. In: *ApJ* 697, pp. 1971–1992. DOI: [10.1088/0004-637X/697/2/1971](https://doi.org/10.1088/0004-637X/697/2/1971). arXiv: [0903.3700](https://arxiv.org/abs/0903.3700).
- Kacprzak, T. (Apr. 2020). “Monte Carlo control loops for cosmic shear cosmology with DES Year 1 data”. In: *Physical Review D* 101.8. ISSN: 2470-0029. DOI: [10.1103/physrevd.101.082003](https://doi.org/10.1103/physrevd.101.082003). URL: <http://dx.doi.org/10.1103/PhysRevD.101.082003>.
- Kacprzak, T. et al. (June 2019). “Monte Carlo Control Loops for cosmic shear cosmology with DES Year 1”. In: *arXiv e-prints*, arXiv:1906.01018, arXiv:1906.01018. arXiv: [1906.01018](https://arxiv.org/abs/1906.01018) [[astro-ph.CO](https://arxiv.org/abs/1906.01018)].
- Kaiser, N. (Apr. 1992). “Weak gravitational lensing of distant galaxies”. In: *ApJ* 388, pp. 272–286. DOI: [10.1086/171151](https://doi.org/10.1086/171151).
- Kaiser, Nick (May 1998). “Weak Lensing and Cosmology”. In: *ApJ* 498.1, pp. 26–42. DOI: [10.1086/305515](https://doi.org/10.1086/305515). arXiv: [astro-ph/9610120](https://arxiv.org/abs/astro-ph/9610120) [[astro-ph](https://arxiv.org/abs/astro-ph)].
- Kauffmann, G. et al. (Sept. 2004). “The environmental dependence of the relations between stellar mass, structure, star formation and nuclear activity in galaxies”. In: *MNRAS* 353, pp. 713–731. DOI: [10.1111/j.1365-2966.2004.08117.x](https://doi.org/10.1111/j.1365-2966.2004.08117.x). eprint: [astro-ph/0402030](https://arxiv.org/abs/astro-ph/0402030).
- Koekemoer, Anton M. et al. (Dec. 2011). “CANDELS: THE COSMIC ASSEMBLY NEAR-INFRARED DEEP EXTRAGALACTIC LEGACY SURVEY—THE HUBBLE SPACE TELESCOPE OBSERVATIONS, IMAGING DATA PRODUCTS, AND MOSAICS”. In: *The Astrophysical Journal Supplement Series* 197.2, p. 36. ISSN: 1538-4365. DOI: [10.1088/0067-0049/197/2/36](https://doi.org/10.1088/0067-0049/197/2/36). URL: <http://dx.doi.org/10.1088/0067-0049/197/2/36>.
- Kormendy, J. (Oct. 1977). “Brightness distributions in compact and normal galaxies. III - Decomposition of observed profiles into spheroid and disk components”. In: *ApJ* 217, pp. 406–419. DOI: [10.1086/155589](https://doi.org/10.1086/155589).

-
- (Feb. 1979). “A morphological survey of bar, lens, and ring components in galaxies
Secular evolution in galaxy structure”. In: *ApJ* 227, pp. 714–728. DOI: [10.1086/156782](https://doi.org/10.1086/156782).
 - Kormendy, J. and R. C. Kennicutt Jr. (Sept. 2004). “Secular Evolution and the Formation of Pseudobulges in Disk Galaxies”. In: *ARA&A* 42, pp. 603–683. DOI: [10.1146/annurev.astro.42.053102.134024](https://doi.org/10.1146/annurev.astro.42.053102.134024). eprint: [astro-ph/0407343](https://arxiv.org/abs/astro-ph/0407343).
 - Krause, Elisabeth and Tim Eifler (Sept. 2017a). “cosmolike - cosmological likelihood analyses for photometric galaxy surveys”. In: *MNRAS* 470.2, pp. 2100–2112. DOI: [10.1093/mnras/stx1261](https://doi.org/10.1093/mnras/stx1261). arXiv: [1601.05779](https://arxiv.org/abs/1601.05779) [[astro-ph.CO](https://arxiv.org/abs/astro-ph)].
 - (May 2017b). “cosmolike – cosmological likelihood analyses for photometric galaxy surveys”. In: *Monthly Notices of the Royal Astronomical Society* 470.2, pp. 2100–2112. ISSN: 0035-8711. DOI: [10.1093/mnras/stx1261](https://doi.org/10.1093/mnras/stx1261). eprint: <https://academic.oup.com/mnras/article-pdf/470/2/2100/18145458/stx1261.pdf>. URL: <https://doi.org/10.1093/mnras/stx1261>.
 - Kuminski, Evan and Lior Shamir (Apr. 2016). “A COMPUTER-GENERATED VISUAL MORPHOLOGY CATALOG OF 3,000,000 SDSS GALAXIES”. In: *The Astrophysical Journal Supplement Series* 223.2, p. 20. ISSN: 1538-4365. DOI: [10.3847/0067-0049/223/2/20](https://doi.org/10.3847/0067-0049/223/2/20). URL: <http://dx.doi.org/10.3847/0067-0049/223/2/20>.
 - Kuutma, T., A. Tamm, and E. Tempel (Apr. 2017). “From voids to filaments: environmental transformations of galaxies in the SDSS”. In: *A&A* 600, L6, p. L6. DOI: [10.1051/0004-6361/201730526](https://doi.org/10.1051/0004-6361/201730526). arXiv: [1703.04338](https://arxiv.org/abs/1703.04338).
 - Lacasa, Fabien (July 2018). “Covariance of the galaxy angular power spectrum with the halo model”. In: *A&A* 615, A1, A1. DOI: [10.1051/0004-6361/201732343](https://doi.org/10.1051/0004-6361/201732343). arXiv: [1711.07372](https://arxiv.org/abs/1711.07372) [[astro-ph.CO](https://arxiv.org/abs/astro-ph)].
 - LagardeDede Laguarde, Anne (1903). “Another Unreal Paper”. In: *Nat* 3, pp. 303–304.
 - Lahav, O. (1995). “Galaxy classification by human eyes and by artificial neural networks”. In: *Astrophysical Letters and Communications* 31, p. 73. eprint: [astro-ph/9505091](https://arxiv.org/abs/astro-ph/9505091).
 - Lahav, O. et al. (Feb. 1995). “Galaxies, Human Eyes, and Artificial Neural Networks”. In: *Science* 267, pp. 859–862. DOI: [10.1126/science.267.5199.859](https://doi.org/10.1126/science.267.5199.859). eprint: [astro-ph/9412027](https://arxiv.org/abs/astro-ph/9412027).
 - Lani, C. et al. (Oct. 2013). “Evidence for a correlation between the sizes of quiescent galaxies and local environment to z 2”. In: *MNRAS* 435, pp. 207–221. DOI: [10.1093/mnras/stt1275](https://doi.org/10.1093/mnras/stt1275). arXiv: [1307.3247](https://arxiv.org/abs/1307.3247).
 - Law, D. R. et al. (Feb. 2007). “The Physical Nature of Rest-UV Galaxy Morphology during the Peak Epoch of Galaxy Formation”. In: *ApJ* 656, pp. 1–26. DOI: [10.1086/510357](https://doi.org/10.1086/510357). eprint: [astro-ph/0610693](https://arxiv.org/abs/astro-ph/0610693).
 - Leistedt, B. et al. (Oct. 2016). “Mapping and Simulating Systematics due to Spatially Varying Observing Conditions in DES Science Verification Data”. In: *ApJS* 226, 24, p. 24. DOI: [10.3847/0067-0049/226/2/24](https://doi.org/10.3847/0067-0049/226/2/24). arXiv: [1507.05647](https://arxiv.org/abs/1507.05647).

-
- Lesgourgues, Julien (Apr. 2011). “The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview”. In: *arXiv e-prints*, arXiv:1104.2932, arXiv:1104.2932. arXiv: [1104.2932 \[astro-ph.IM\]](#).
- Lewis, Antony, Anthony Challinor, and Anthony Lasenby (Aug. 2000). “Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models”. In: *ApJ* 538.2, pp. 473–476. DOI: [10.1086/309179](#). arXiv: [astro-ph/9911177 \[astro-ph\]](#).
- Limber, D. N. (Jan. 1953). “The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field.” In: *ApJ* 117, p. 134. DOI: [10.1086/145672](#).
- Lin, L. et al. (Dec. 2004). “The DEEP2 Galaxy Redshift Survey: Evolution of Close Galaxy Pairs and Major-Merger Rates up to $z \sim 1.2$ ”. In: *ApJ* 617, pp. L9–L12. DOI: [10.1086/427183](#). eprint: [astro-ph/0411104](#).
- Lingard, Timothy K. et al. (Sept. 2020). “Galaxy Zoo Builder: Four-component Photometric Decomposition of Spiral Galaxies Guided by Citizen Science”. In: *ApJ* 900.2, 178, p. 178. DOI: [10.3847/1538-4357/ab9d83](#). arXiv: [2006.10450 \[astro-ph.GA\]](#).
- Lintott, C. J. et al. (Sept. 2008). “Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey”. In: *MNRAS* 389, pp. 1179–1189. DOI: [10.1111/j.1365-2966.2008.13689.x](#). arXiv: [0804.4483](#).
- Lotz, J. M., M. Davis, et al. (Jan. 2008). “The Evolution of Galaxy Mergers and Morphology at $z \sim 1.2$ in the Extended Groth Strip”. In: *ApJ* 672, 177–197, pp. 177–197. DOI: [10.1086/523659](#). eprint: [astro-ph/0602088](#).
- Lotz, J. M., P. Jonsson, et al. (Dec. 2008). “Galaxy merger morphologies and time-scales from simulations of equal-mass gas-rich disc mergers”. In: *MNRAS* 391, pp. 1137–1162. DOI: [10.1111/j.1365-2966.2008.14004.x](#). arXiv: [0805.1246](#).
- Lotz, J. M., J. Primack, and P. Madau (July 2004). “A New Nonparametric Approach to Galaxy Morphological Classification”. In: *AJ* 128, pp. 163–182. DOI: [10.1086/421849](#). eprint: [astro-ph/0311352](#).
- LoVerde, Marilena and Niayesh Afshordi (Dec. 2008). “Extended Limber approximation”. In: *Phys. Rev. D* 78 (12), p. 123506. DOI: [10.1103/PhysRevD.78.123506](#). URL: <https://link.aps.org/doi/10.1103/PhysRevD.78.123506>.
- Loverde, Marilena and Niayesh Afshordi (Dec. 2008). “Extended Limber approximation”. In: *Phys. Rev. D* 78.12, 123506, p. 123506. DOI: [10.1103/PhysRevD.78.123506](#). arXiv: [0809.5112 \[astro-ph\]](#).
- LSST Dark Energy Science Collaboration (Nov. 2012). “Large Synoptic Survey Telescope: Dark Energy Science Collaboration”. In: *arXiv e-prints*, arXiv:1211.0310, arXiv:1211.0310. arXiv: [1211.0310 \[astro-ph.CO\]](#).
- Mandelbaum, Rachel (Sept. 2018a). “Weak Lensing for Precision Cosmology”. In: *Annual Review of Astronomy and Astrophysics* 56.1, pp. 393–433. ISSN: 1545-4282. DOI: [10.](#)

- [1146/annurev-astro-081817-051928](https://doi.org/10.1146/annurev-astro-081817-051928). URL: <http://dx.doi.org/10.1146/annurev-astro-081817-051928>.
- (Sept. 2018b). “Weak Lensing for Precision Cosmology”. In: *ARA&A* 56, pp. 393–433. DOI: [10.1146/annurev-astro-081817-051928](https://doi.org/10.1146/annurev-astro-081817-051928). arXiv: [1710.03235](https://arxiv.org/abs/1710.03235) [[astro-ph.CO](#)].
- Martig, M. et al. (Dec. 2009). “Morphological Quenching of Star Formation: Making Early-Type Galaxies Red”. In: *ApJ* 707, pp. 250–267. DOI: [10.1088/0004-637X/707/1/250](https://doi.org/10.1088/0004-637X/707/1/250). arXiv: [0905.4669](https://arxiv.org/abs/0905.4669) [[astro-ph.CO](#)].
- Matsubara, Takahiko (Dec. 1995). “On Second-Order Perturbation Theories of Gravitational Instability in Friedmann-Lemaitre Models”. In: *Progress of Theoretical Physics* 94.6, pp. 1151–1156. ISSN: 0033-068X. DOI: [10.1143/PTP.94.1151](https://doi.org/10.1143/PTP.94.1151). eprint: <https://academic.oup.com/ptp/article-pdf/94/6/1151/5441959/94-6-1151.pdf>. URL: <https://doi.org/10.1143/PTP.94.1151>.
- Mead, A. J. et al. (Dec. 2015). “An accurate halo model for fitting non-linear cosmological power spectra and baryonic feedback models”. In: *MNRAS* 454.2, pp. 1958–1975. DOI: [10.1093/mnras/stv2036](https://doi.org/10.1093/mnras/stv2036). arXiv: [1505.07833](https://arxiv.org/abs/1505.07833) [[astro-ph.CO](#)].
- Melchior, P. and M. Viola (Aug. 2012). “Means of confusion: how pixel noise affects shear estimates for weak gravitational lensing”. In: *MNRAS* 424, pp. 2757–2769. DOI: [10.1111/j.1365-2966.2012.21381.x](https://doi.org/10.1111/j.1365-2966.2012.21381.x). arXiv: [1204.5147](https://arxiv.org/abs/1204.5147) [[astro-ph.IM](#)].
- Meurer, Aaron et al. (Jan. 2017). “SymPy: symbolic computing in Python”. In: *PeerJ Computer Science* 3, e103. ISSN: 2376-5992. DOI: [10.7717/peerj-cs.103](https://doi.org/10.7717/peerj-cs.103). URL: <https://doi.org/10.7717/peerj-cs.103>.
- Miyazaki, S. et al. (Sept. 2012). “Hyper Suprime-Cam”. In: *Ground-based and Airborne Instrumentation for Astronomy IV*. Vol. 8446. Proc. SPIE, 84460Z, 84460Z. DOI: [10.1117/12.926844](https://doi.org/10.1117/12.926844).
- Mo, Houjun, Frank C. van den Bosch, and Simon White (2010). *Galaxy Formation and Evolution*.
- Mohr, J. J. et al. (Sept. 2012). “The Dark Energy Survey data processing and calibration system”. In: *Software and Cyberinfrastructure for Astronomy II*. Vol. 8451. Proc. SPIE, 84510D, p. 84510D. DOI: [10.1117/12.926785](https://doi.org/10.1117/12.926785). arXiv: [1207.3189](https://arxiv.org/abs/1207.3189) [[astro-ph.IM](#)].
- Morganson, E., R. A. Gruendl, F. Menanteau, M. Carrasco Kind, Y.-C. Chen, et al. (July 2018). “The Dark Energy Survey Image Processing Pipeline”. In: *PASP* 130.989, p. 074501. DOI: [10.1088/1538-3873/aab4ef](https://doi.org/10.1088/1538-3873/aab4ef). arXiv: [1801.03177](https://arxiv.org/abs/1801.03177) [[astro-ph.IM](#)].
- Morganson, E., R. A. Gruendl, F. Menanteau, M. Carrasco Kind, Y.-C. Chen, et al. (Jan. 2018). “The Dark Energy Survey Image Processing Pipeline”. In: *ArXiv e-prints*. arXiv: [1801.03177](https://arxiv.org/abs/1801.03177) [[astro-ph.IM](#)].
- Naab, T. and A. Burkert (Nov. 2003). “Statistical Properties of Collisionless Equal- and Unequal-Mass Merger Remnants of Disk Galaxies”. In: *ApJ* 597, pp. 893–906. DOI: [10.1086/378581](https://doi.org/10.1086/378581). eprint: [astro-ph/0110179](https://arxiv.org/abs/astro-ph/0110179).

-
- Nair, Preethi B. and Roberto G. Abraham (Feb. 2010). “A CATALOG OF DETAILED VISUAL MORPHOLOGICAL CLASSIFICATIONS FOR 14,034 GALAXIES IN THE SLOAN DIGITAL SKY SURVEY”. In: *The Astrophysical Journal Supplement Series* 186.2, pp. 427–456. DOI: [10.1088/0067-0049/186/2/427](https://doi.org/10.1088/0067-0049/186/2/427). URL: <https://doi.org/10.1088/0067-0049/186/2/427>.
- Noeske, K. G. et al. (May 2007). “Star Formation in AEGIS Field Galaxies since $z=1.1$: Staged Galaxy Formation and a Model of Mass-dependent Gas Exhaustion”. In: *ApJ* 660.1, pp. L47–L50. DOI: [10.1086/517927](https://doi.org/10.1086/517927). arXiv: [astro-ph/0703056](https://arxiv.org/abs/astro-ph/0703056) [[astro-ph](#)].
- Palmese, A. et al. (Nov. 2017a). “Evidence for Dynamically Driven Formation of the GW170817 Neutron Star Binary in NGC 4993”. In: *ApJ* 849.2, L34, p. L34. DOI: [10.3847/2041-8213/aa9660](https://doi.org/10.3847/2041-8213/aa9660). arXiv: [1710.06748](https://arxiv.org/abs/1710.06748) [[astro-ph.HE](#)].
- (Nov. 2017b). “Evidence for Dynamically Driven Formation of the GW170817 Neutron Star Binary in NGC 4993”. In: *ApJ* 849.2, L34, p. L34. DOI: [10.3847/2041-8213/aa9660](https://doi.org/10.3847/2041-8213/aa9660). arXiv: [1710.06748](https://arxiv.org/abs/1710.06748) [[astro-ph.HE](#)].
- Peacock, J. A. (Feb. 1997). “The evolution of galaxy clustering”. In: *mnras* 284, pp. 885–898. DOI: [10.1093/mnras/284.4.885](https://doi.org/10.1093/mnras/284.4.885). eprint: [astro-ph/9608151](https://arxiv.org/abs/astro-ph/9608151).
- Peacock, J. A. and R. E. Smith (Nov. 2000). “Halo occupation numbers and galaxy bias”. In: *MNRAS* 318, pp. 1144–1156. DOI: [10.1046/j.1365-8711.2000.03779.x](https://doi.org/10.1046/j.1365-8711.2000.03779.x). eprint: [astro-ph/0005010](https://arxiv.org/abs/astro-ph/0005010).
- Peng, C. Y. et al. (June 2010). “Detailed Decomposition of Galaxy Images. II. Beyond Axisymmetric Models”. In: *AJ* 139, pp. 2097–2129. DOI: [10.1088/0004-6256/139/6/2097](https://doi.org/10.1088/0004-6256/139/6/2097). arXiv: [0912.0731](https://arxiv.org/abs/0912.0731).
- Peng, Chien Y. et al. (June 2010). “Detailed Decomposition of Galaxy Images. II. Beyond Axisymmetric Models”. In: *AJ* 139.6, pp. 2097–2129. DOI: [10.1088/0004-6256/139/6/2097](https://doi.org/10.1088/0004-6256/139/6/2097). arXiv: [0912.0731](https://arxiv.org/abs/0912.0731) [[astro-ph.CO](#)].
- Peng, C. Y. et al. (July 2002). “Detailed Structural Decomposition of Galaxy Images”. In: *AJ* 124, pp. 266–293. DOI: [10.1086/340952](https://doi.org/10.1086/340952). eprint: [astro-ph/0204182](https://arxiv.org/abs/astro-ph/0204182).
- Peterken, Thomas et al. (Jan. 2021). “SDSS-IV MaNGA: when is morphology imprinted on galaxies?” In: *MNRAS* 500.1, pp. L42–L46. DOI: [10.1093/mnras/1/slaa179](https://doi.org/10.1093/mnras/1/slaa179). arXiv: [2010.15213](https://arxiv.org/abs/2010.15213) [[astro-ph.GA](#)].
- Petrosian, V. (Oct. 1976). “Surface brightness and evolution of galaxies”. In: *ApJ* 209, pp. L1–L5. DOI: [10.1086/182253](https://doi.org/10.1086/182253).
- Pieres, A. et al. (June 2017). “A stellar overdensity associated with the Small Magellanic Cloud”. In: *MNRAS* 468, pp. 1349–1360. DOI: [10.1093/mnras/stx507](https://doi.org/10.1093/mnras/stx507). arXiv: [1612.03938](https://arxiv.org/abs/1612.03938).
- Planck Collaboration (Sept. 2020). “Planck 2018 results. V. CMB power spectra and likelihoods”. In: *A&A* 641, A5, A5. DOI: [10.1051/0004-6361/201936386](https://doi.org/10.1051/0004-6361/201936386). arXiv: [1907.12875](https://arxiv.org/abs/1907.12875) [[astro-ph.CO](#)].

- Planck Collaboration et al. (July 2018). “Planck 2018 results. VI. Cosmological parameters”. In: *arXiv e-prints*, arXiv:1807.06209, arXiv:1807.06209. arXiv: [1807.06209 \[astro-ph.CO\]](#).
- Postman, M. et al. (Apr. 2005). “The Morphology-Density Relation in $z \sim 1$ Clusters”. In: *ApJ* 623, pp. 721–741. DOI: [10.1086/428881](#). eprint: [astro-ph/0501224](#).
- Prince, Heather and Jo Dunkley (Oct. 2019). “Data compression in cosmology: A compressed likelihood for Planck data”. In: *Phys. Rev. D* 100.8, 083502, p. 083502. DOI: [10.1103/PhysRevD.100.083502](#). arXiv: [1909.05869 \[astro-ph.CO\]](#).
- Refregier, A., A. Amara, et al. (Apr. 2011). “iCosmo: an interactive cosmology package”. In: *A&A* 528, A33, A33. DOI: [10.1051/0004-6361/200811112](#). arXiv: [0810.1285 \[astro-ph\]](#).
- Refregier, A., L. Gamper, et al. (Oct. 2018a). “PyCosmo: An integrated cosmological Boltzmann solver”. In: *Astronomy and Computing* 25, pp. 38–43. DOI: [10.1016/j.ascom.2018.08.001](#). arXiv: [1708.05177 \[astro-ph.CO\]](#).
- (Oct. 2018b). “PyCosmo: An integrated cosmological Boltzmann solver”. In: *Astronomy and Computing* 25, pp. 38–43. DOI: [10.1016/j.ascom.2018.08.001](#). arXiv: [1708.05177 \[astro-ph.CO\]](#).
- Refregier, Alexandre and Adam Amara (Apr. 2014). “A way forward for Cosmic Shear: Monte-Carlo Control Loops”. In: *Physics of the Dark Universe* 3, pp. 1–3. DOI: [10.1016/j.dark.2014.01.002](#). arXiv: [1303.4739 \[astro-ph.CO\]](#).
- Riess, Adam G. et al. (Sept. 1998). “Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant”. In: *The Astronomical Journal* 116.3, pp. 1009–1038. DOI: [10.1086/300499](#). URL: <https://doi.org/10.1086/300499>.
- Salim, S. (2014). “Green valley galaxies”. In: *Serbian Astronomical Journal* 189, pp. 1–14. ISSN: 1820-9289. DOI: [10.2298/saj1489001s](#). URL: <http://dx.doi.org/10.2298/SAJ1489001S>.
- Scarlata, C., C. M. Carollo, S. Lilly, et al. (Sept. 2007). “COSMOS Morphological Classification with the Zurich Estimator of Structural Types (ZEST) and the Evolution Since $z = 1$ of the Luminosity Function of Early, Disk, and Irregular Galaxies”. In: *ApJS* 172, pp. 406–433. DOI: [10.1086/516582](#).
- Scarlata, C., C. M. Carollo, S. J. Lilly, et al. (Sept. 2007). “The Redshift Evolution of Early-Type Galaxies in COSMOS: Do Massive Early-Type Galaxies Form by Dry Mergers?” In: *ApJS* 172, pp. 494–510. DOI: [10.1086/517972](#). eprint: [astro-ph/0701746](#).
- Schade, D. et al. (Sept. 1995). “Canada-France Redshift Survey: Hubble Space Telescope Imaging of High-Redshift Field Galaxies”. In: *ApJ* 451, p. L1. DOI: [10.1086/309677](#). eprint: [astro-ph/9507028](#).
- Schawinski, K., S. Kaviraj, et al. (Dec. 2007). “The Effect of Environment on the Ultraviolet Color-Magnitude Relation of Early-Type Galaxies”. In: *ApJS* 173, pp. 512–523. DOI: [10.1086/516631](#). eprint: [astro-ph/0601036](#).

-
- Schawinski, K., C. M. Urry, et al. (May 2014). “The green valley is a red herring: Galaxy Zoo reveals two evolutionary pathways towards quenching of star formation in early- and late-type galaxies”. In: *MNRAS* 440, pp. 889–907. DOI: [10.1093/mnras/stu327](https://doi.org/10.1093/mnras/stu327). arXiv: [1402.4814](https://arxiv.org/abs/1402.4814).
- Schawinski, Kevin et al. (Mar. 2014). “The green valley is a red herring: Galaxy Zoo reveals two evolutionary pathways towards quenching of star formation in early- and late-type galaxies”. In: *Monthly Notices of the Royal Astronomical Society* 440.1, pp. 889–907. ISSN: 0035-8711. DOI: [10.1093/mnras/stu327](https://doi.org/10.1093/mnras/stu327). URL: <http://dx.doi.org/10.1093/mnras/stu327>.
- Schiminovich, David et al. (Dec. 2007). “The UV-Optical Color Magnitude Diagram. II. Physical Properties and Morphological Evolution On and Off of a Star-forming Sequence”. In: *ApJS* 173.2, pp. 315–341. DOI: [10.1086/524659](https://doi.org/10.1086/524659). arXiv: [0711.4823](https://arxiv.org/abs/0711.4823) [astro-ph].
- Schneider, Aurel et al. (Apr. 2020). “Baryonic effects for weak lensing. Part I. Power spectrum and covariance matrix”. In: *J. Cosmology Astropart. Phys.* 2020.4, 019, p. 019. DOI: [10.1088/1475-7516/2020/04/019](https://doi.org/10.1088/1475-7516/2020/04/019). arXiv: [1910.11357](https://arxiv.org/abs/1910.11357) [astro-ph.CO].
- Schöneberg, Nils et al. (Oct. 2018). “Beyond the traditional line-of-sight approach of cosmological angular statistics”. In: *Journal of Cosmology and Astroparticle Physics* 2018.10, pp. 047–047. ISSN: 1475-7516. DOI: [10.1088/1475-7516/2018/10/047](https://doi.org/10.1088/1475-7516/2018/10/047). URL: <http://dx.doi.org/10.1088/1475-7516/2018/10/047>.
- Seljak, Uroš (Oct. 2000). “Analytic model for galaxy and dark matter clustering”. In: *Monthly Notices of the Royal Astronomical Society* 318.1, pp. 203–213. ISSN: 0035-8711. DOI: [10.1046/j.1365-8711.2000.03715.x](https://doi.org/10.1046/j.1365-8711.2000.03715.x). eprint: <http://oup.prod.sis.lan/mnras/article-pdf/318/1/203/3943998/318-1-203.pdf>. URL: <https://doi.org/10.1046/j.1365-8711.2000.03715.x>.
- Seljak, Uros and Matias Zaldarriaga (Oct. 1996a). “A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies”. In: *ApJ* 469, p. 437. DOI: [10.1086/177793](https://doi.org/10.1086/177793). arXiv: [astro-ph/9603033](https://arxiv.org/abs/astro-ph/9603033) [astro-ph].
- (Oct. 1996b). “A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies”. In: *ApJ* 469, p. 437. DOI: [10.1086/177793](https://doi.org/10.1086/177793). arXiv: [astro-ph/9603033](https://arxiv.org/abs/astro-ph/9603033) [astro-ph].
- Sellwood, J. A. (Jan. 2014). “Secular evolution in disk galaxies”. In: *Rev. Mod. Phys.* 86 (1), pp. 1–46. DOI: [10.1103/RevModPhys.86.1](https://doi.org/10.1103/RevModPhys.86.1). URL: <http://link.aps.org/doi/10.1103/RevModPhys.86.1>.
- Sérsic, J. L. (1963). “Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy”. In: *Boletín de la Asociación Argentina de Astronomía La Plata Argentina* 6, p. 41.
- Sevilla, I. et al. (2011). “The Dark Energy Survey Data Management System”. In: *Particles and fields. Proceedings, Meeting of the Division of the American Physical Society, DPF*

- 2011, Providence, USA, August 9-13, 2011. arXiv: [1109.6741 \[astro-ph.IM\]](#). URL: http://lss.fnal.gov/cgi-bin/find_paper.pl?conf-11-525.
- Sevilla-Noarbe, I. et al. (May 2021). “Dark Energy Survey Year 3 Results: Photometric Data Set for Cosmology”. In: *The Astrophysical Journal Supplement Series* 254.2, p. 24. ISSN: 1538-4365. DOI: [10.3847/1538-4365/abeb66](#). URL: <http://dx.doi.org/10.3847/1538-4365/abeb66>.
- Sgier, R. J. et al. (Jan. 2019). “Fast generation of covariance matrices for weak lensing”. In: *J. Cosmology Astropart. Phys.* 2019.1, 044, p. 044. DOI: [10.1088/1475-7516/2019/01/044](#). arXiv: [1801.05745 \[astro-ph.CO\]](#).
- Simard, L. et al. (Sept. 2011). “A Catalog of Bulge+disk Decompositions and Updated Photometry for 1.12 Million Galaxies in the Sloan Digital Sky Survey”. In: *ApJS* 196, 11, p. 11. DOI: [10.1088/0067-0049/196/1/11](#). arXiv: [1107.1518](#).
- Simmons, B. D. et al. (Feb. 2017). “Galaxy Zoo: quantitative visual morphological classifications for 48 000 galaxies from CANDELS”. In: *MNRAS* 464, pp. 4420–4447. DOI: [10.1093/mnras/stw2587](#). arXiv: [1610.03070](#).
- Smith, R. E. et al. (June 2003). “Stable clustering, the halo model and non-linear cosmological power spectra”. In: *MNRAS* 341, pp. 1311–1332. DOI: [10.1046/j.1365-8711.2003.06503.x](#). eprint: [astro-ph/0207664](#).
- Soumagnac, M. T. et al. (June 2015). “Star/galaxy separation at faint magnitudes: application to a simulated Dark Energy Survey”. In: *MNRAS* 450, pp. 666–680. DOI: [10.1093/mnras/stu1410](#).
- Srinivas, Niranjana et al. (Dec. 2009). “Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design”. In: *arXiv e-prints*, arXiv:0912.3995, arXiv:0912.3995. arXiv: [0912.3995 \[cs.LG\]](#).
- Tacchella, S. et al. (Apr. 2015). “SINS/zC-SINF Survey of $z \sim 2$ Galaxy Kinematics: Rest-frame Morphology, Structure, and Colors from Near-infrared Hubble Space Telescope Imaging”. In: *ApJ* 802, 101, p. 101. DOI: [10.1088/0004-637X/802/2/101](#). arXiv: [1411.7034](#).
- Takahashi, R. et al. (Dec. 2012). “Revising the Halofit Model for the Nonlinear Matter Power Spectrum”. In: *ApJ* 761, 152, p. 152. DOI: [10.1088/0004-637X/761/2/152](#). arXiv: [1208.2701](#).
- Talman, James D (1978). “Numerical Fourier and Bessel transforms in logarithmic variables”. In: *Journal of Computational Physics* 29.1, pp. 35–48. ISSN: 0021-9991. DOI: [https://doi.org/10.1016/0021-9991\(78\)90107-9](https://doi.org/10.1016/0021-9991(78)90107-9). URL: <https://www.sciencedirect.com/science/article/pii/0021999178901079>.
- Tarsitano, F., C. Bruderer, et al. (2021). *Image feature extraction and galaxy classification: a novel and efficient approach with automated machine learning*. arXiv: [2105.01070 \[astro-ph.GA\]](#).

-
- Tarsitano, F., W. G. Hartley, et al. (Dec. 2018). “A catalogue of structural and morphological measurements for DES Y1”. In: *MNRAS* 481.2, pp. 2018–2040. DOI: [10.1093/mnras/sty1970](https://doi.org/10.1093/mnras/sty1970). arXiv: [1807.10767](https://arxiv.org/abs/1807.10767) [[astro-ph.GA](#)].
- Tarsitano, F., U. Schmitt, et al. (2020). *Predicting Cosmological Observables with PyCosmo*. arXiv: [2005.00543](https://arxiv.org/abs/2005.00543) [[astro-ph.CO](#)].
- The DES Collaboration (Aug. 2016). “The Dark Energy Survey: more than dark energy - an overview”. In: *MNRAS* 460, pp. 1270–1299. DOI: [10.1093/mnras/stw641](https://doi.org/10.1093/mnras/stw641). arXiv: [1601.00329](https://arxiv.org/abs/1601.00329).
- Toomre, A. (1977). “Mergers and Some Consequences”. In: *Evolution of Galaxies and Stellar Populations*. Ed. by B. M. Tinsley and R. B. G. Larson D. Campbell, p. 401.
- Toomre, A. and J. Toomre (Dec. 1972). “Galactic Bridges and Tails”. In: *ApJ* 178, pp. 623–666. DOI: [10.1086/151823](https://doi.org/10.1086/151823).
- Tsujikawa, Shinji (Apr. 2003). “Introductory review of cosmic inflation”. In: *arXiv e-prints*, hep-ph/0304257, hep-ph/0304257. arXiv: [hep-ph/0304257](https://arxiv.org/abs/hep-ph/0304257) [[hep-ph](#)].
- Tuccillo, D. et al. (Mar. 2018). “Deep learning for galaxy surface brightness profile fitting”. In: *MNRAS* 475.1, pp. 894–909. DOI: [10.1093/mnras/stx3186](https://doi.org/10.1093/mnras/stx3186). arXiv: [1711.03108](https://arxiv.org/abs/1711.03108) [[astro-ph.GA](#)].
- van Daalen, M. P. et al. (Aug. 2011). “The effects of galaxy formation on the matter power spectrum: a challenge for precision cosmology”. In: *MNRAS* 415, pp. 3649–3665. DOI: [10.1111/j.1365-2966.2011.18981.x](https://doi.org/10.1111/j.1365-2966.2011.18981.x). arXiv: [1104.1174](https://arxiv.org/abs/1104.1174) [[astro-ph.CO](#)].
- van der Wel, A. (July 2008). “The morphology-density relation: a constant of nature”. In: *Formation and Evolution of Galaxy Bulges*. Ed. by M. Bureau, E. Athanassoula, and B. Barbuy. Vol. 245. IAU Symposium, pp. 59–62. DOI: [10.1017/S1743921308017286](https://doi.org/10.1017/S1743921308017286).
- van der Wel, A. (Mar. 2008). “The Dependence of Galaxy Morphology and Structure on Environment and Stellar Mass”. In: *ApJ* 675, L13, p. L13. DOI: [10.1086/529432](https://doi.org/10.1086/529432). arXiv: [0801.1995](https://arxiv.org/abs/0801.1995).
- Vega-Ferrero, J. et al. (2020). *Pushing automated morphological classifications to their limits with the Dark Energy Survey*. arXiv: [2012.07858](https://arxiv.org/abs/2012.07858) [[astro-ph.GA](#)].
- Walmsley, Mike et al. (Oct. 2019). “Galaxy Zoo: probabilistic morphology through Bayesian CNNs and active learning”. In: *Monthly Notices of the Royal Astronomical Society* 491.2, pp. 1554–1574. ISSN: 0035-8711. DOI: [10.1093/mnras/stz2816](https://doi.org/10.1093/mnras/stz2816). eprint: <https://academic.oup.com/mnras/article-pdf/491/2/1554/31144873/stz2816.pdf>. URL: <https://doi.org/10.1093/mnras/stz2816>.
- Weinberg, David H. et al. (Sept. 2013). “Observational probes of cosmic acceleration”. In: *Physics Reports* 530.2, pp. 87–255. ISSN: 0370-1573. DOI: [10.1016/j.physrep.2013.05.001](https://doi.org/10.1016/j.physrep.2013.05.001). URL: <http://dx.doi.org/10.1016/j.physrep.2013.05.001>.
- Weinberg, Steven (2008). *Cosmology*.
- Weinmann, S. M. et al. (Feb. 2006). “Properties of galaxy groups in the Sloan Digital Sky Survey - I. The dependence of colour, star formation and morphology on halo mass”.

- In: *MNRAS* 366, pp. 2–28. DOI: [10.1111/j.1365-2966.2005.09865.x](https://doi.org/10.1111/j.1365-2966.2005.09865.x). eprint: [astro-ph/0509147](https://arxiv.org/abs/astro-ph/0509147).
- Wel, A. van der et al. (2012). “Structural Parameters of Galaxies in CANDELS”. In: *The Astrophysical Journal Supplement Series* 203.2, p. 24. URL: <http://stacks.iop.org/0067-0049/203/i=2/a=24>.
- White, S. D. M. and M. J. Rees (May 1978). “Core condensation in heavy halos - A two-stage theory for galaxy formation and clustering”. In: *MNRAS* 183, pp. 341–358. DOI: [10.1093/mnras/183.3.341](https://doi.org/10.1093/mnras/183.3.341).
- Willett, K. W. et al. (Feb. 2017). “Galaxy Zoo: morphological classifications for 120 000 galaxies in HST legacy imaging”. In: *MNRAS* 464, pp. 4176–4203. DOI: [10.1093/mnras/stw2568](https://doi.org/10.1093/mnras/stw2568). arXiv: [1610.03068](https://arxiv.org/abs/1610.03068).
- Woo, J. et al. (Mar. 2015). “Two conditions for galaxy quenching: compact centres and massive haloes”. In: *MNRAS* 448, pp. 237–251. DOI: [10.1093/mnras/stu2755](https://doi.org/10.1093/mnras/stu2755). arXiv: [1406.5372](https://arxiv.org/abs/1406.5372).
- York, Donald G., J. Adelman, Jr. Anderson John E., et al. (Sept. 2000). “The Sloan Digital Sky Survey: Technical Summary”. In: *AJ* 120.3, pp. 1579–1587. DOI: [10.1086/301513](https://doi.org/10.1086/301513). arXiv: [astro-ph/0006396](https://arxiv.org/abs/astro-ph/0006396) [[astro-ph](https://arxiv.org/abs/astro-ph)].
- York, Donald G., J. Adelman, Jr. John E. Anderson, et al. (2000). “The Sloan Digital Sky Survey: Technical Summary”. In: *The Astronomical Journal* 120.3, p. 1579. URL: <http://stacks.iop.org/1538-3881/120/i=3/a=1579>.
- Zamojski, M. A. et al. (Sept. 2007). “Deep GALEX Imaging of the COSMOS HST Field: A First Look at the Morphology of $z \sim 0.7$ Star-forming Galaxies”. In: *ApJS* 172, pp. 468–493. DOI: [10.1086/516593](https://doi.org/10.1086/516593). eprint: [astro-ph/0701478](https://arxiv.org/abs/astro-ph/0701478).
- Zuntz, J. et al. (Sept. 2015). “CosmoSIS: Modular cosmological parameter estimation”. In: *Astronomy and Computing* 12, pp. 45–59. DOI: [10.1016/j.ascom.2015.05.005](https://doi.org/10.1016/j.ascom.2015.05.005). arXiv: [1409.3409](https://arxiv.org/abs/1409.3409) [[astro-ph.CO](https://arxiv.org/abs/astro-ph.CO)].

Curriculum Vitae of Federica Tarsitano

PROFILE AND CONTACT INFORMATION

Date of birth: 24/05/1990

Nationality: Italian

Email: federica.tarsitano@phys.ethz.ch

Web: <https://www.federicatarsitano.com/>

EMPLOYMENT 2015-2021: Research Assistant, ETH Zürich

EDUCATION

2015-2021 PhD in Physics, ETH Zürich

- Thesis: *Galaxy morphology and cosmological predictions for wide-field surveys analyses*

- Supervisor: Prof. Dr. Alexandre Refregier

2012-2014 Msc in Nuclear, Subnuclear and Biomedical Physics, 110/110 Summa Cum Laude and Honorable Mention, University of Turin, Italy

2009-2012 Bsc in Physics, University of Turin, Italy

2004-2009 Maturità Classica, Liceo Classico C. Cavour, 100/100, Torino, Italy

TEACHING ASSISTANT, ETH ZURICH

- Physics III, Fall Semester 2020
- Astrophysics II, Spring Semesters 2018, 2019, 2020
- Statistical Methods in Astrophysics, Fall Semester 2018

- Astrophysics I, Fall Semester 2018
- Physics II, Spring Semester 2016
- Physics I, Fall Semester 2015, 2016

MENTORING

2020 Silvan Fishbacher's Semester Project, *Constraining Cosmological Parameters with PyCosmo using the KiDS+VIKING-450 Data Release.*

2020 Paul Moser Röggl's Bachelor Thesis, *The effect of Intrinsic Alignment on cosmological parameter constraints.*

2019 Beatrice Moser's Master Thesis, *Neutrino Cosmology.*

2019 Imelda Romero's Semester Project, *Testing PyCosmo on JupyterHub.*

2018 Ralf Aeberhard's Semester Project, *Semesterproject in Cosmology.*

CONFERENCES AND OBSERVATIONS

- Scipy 2020, online conference, July 2020
- Python in Astronomy, online meeting, April 2020
- Visiting student at Stanford University, December 2019
- Cosmology on Safari, South Africa, March 2019
- Cosmo-18, Daejeon, South Korea, August 2019
- Dark Energy Survey (DES) Collaboration Meeting, Penn State University, USA, June 2019
- DES Collaboration, Observations at Blanco Telescope, Chile, August-September 2016
- DES Collaboration Meeting, Stanford University, USA, May 2016
- DES Collaboration Meeting, University of Michigan, USA, May 2015
- Dark Energy Survey monthly telecons, Galaxy Evolution Group

PUBLICATIONS

- Tarsitano et al., 2021, *Image feature extraction and galaxy classification: a novel and efficient approach with automated machine learning* (Tarsitano, Bruderer, et al., 2021).
- Tarsitano et al, 2020, *Predicting Cosmological Observables with PyCosmo*, accepted by *Astronomy and Computing* (Tarsitano, Schmitt, et al., 2020).
- Tarsitano et al, 2018, A catalogue of structural and morphological measurements for DES Y1 – Monthly Notices of the Royal Astronomical Society (Tarsitano, Hartley, et al., 2018).
- Hartley W. et al., 2021, *Dark Energy Survey Year 3 Results: Deep Field Optical + Near-Infrared Images and Catalogue* (Hartley et al., 2020)
- T. Kacprzak, J. Herbel, A. Nicola, R. Sgier, F. Tarsitano, [...], *Monte Carlo Control Loops for cosmic shear cosmology with DES Year 1*, *Physical Review D* (Kacprzak et al., 2019)
- Palmese, A., Hartley, W., Tarsitano, F., [...], *Evidence for Dynamically Driven Formation of the GW170817 Neutron Star Binary in NGC 4993*, *The Astrophysical Journal Letters*, (Palmese et al., 2017a)

ACADEMIC REVIEWER

I reviewed papers for the journal Monthly Notices of the Royal Astronomical Society (MNRAS) and within the Dark energy Survey (DES) Collaboration.

OUTREACH AND AWARDS

Selected student by ESA and Thales Alenia Space to attend the launch of the shuttle Discovery STS-120, Cape Canaveral, Florida.

Winner of the national scientific competition organized by Tuttoscienze La Stampa in collaboration with the European Space Agency (ESA) and Thales Alenia Space. Prize-trip to witness the launch of the Discovery STS-120 space mission, launched from Cape Canaveral on 23 October 2007.

Guest researcher - Swiss radiotelevision

I took part as a guest in the radio program Mille voci, broadcast on the Swiss RSI Rete Uno radio channel. I took part in the discussion about the topic "From the beginning to the end of the Universe", on March 7th, 2018.

CERN Summer Student

Cern Summer Student to conduct research for my Master thesis in Nuclear, Subnuclear and Biomedical Physics. In my thesis I measured the radiation damage to the CERN-CMS calorimeter and worked on a possible calibration of its crystals. I also used CMS data to study decays of particles containing a beauty quark.

Scientifica outreach event

Scientifica is a biennial scientific outreach event organized by the ETH Zurich and the University of Zurich (UZH), which welcomes more than 25000 visitors every year. I have been contributing to interactive activities for the public.

DES Thought of the Day

Contribution to the outreach activity DES Thought of the Day, run by the Dark Energy Survey collaboration on their social media platforms.

