Catalog-free modeling of galaxy types in deep survey images : massive dimensional reduction with neural networks

Florian Livet

To cite this version:

Florian Livet. Catalog-free modeling of galaxy types in deep survey images : massive dimensional reduction with neural networks. Astrophysics [astro-ph]. Sorbonne Université, 2021. English. NNT : 2021SORUS173 . tel-03681810

HAL Id: tel-03681810
https://theses.hal.science/tel-03681810v1

Submitted on 30 May 2022

A thesis submitted for the degree of
Doctor of Sorbonne Université

Institut d'Astrophysique de Paris (IAP)

École doctorale d'astronomie et d'astrophysique d'Île-de-France (ED 127)

Discipline : Astrophysique

Catalog-free modeling of galaxy types in deep survey images: Massive dimensional reduction with neural networks

Modélisation sans catalogue des types de galaxies dans les relevés profonds : Compression extrême d'images par réseaux de neurones

---

Author: **Florian LIVET**

Supervisors:
Valérie DE LAPPARENT
Damien LE BORGNE

Members of the jury:

Referees
Laurence TRESSE
Brice MÉNARD

Examiners
Benjamin WANDELT
Micol BOLZONELLA
Marc HUERTAS-COMPANY

Invited member
Tom CHARNOCK

Date of defense: September 28, 2021

Remerciements

Déjà petit j'admirais le ciel et me demandais ce qu'il y avait là-haut. Je me rappelle de ces voyages en voiture pendant lesquels je regardais par la fenêtre et me posais des questions comme "D'où venons-nous ?", "Comment tout ça a commencé ?" ou encore "C'est quoi ces trucs qui brillent là-haut ?". Cette passion pour l'astronomie n'a jamais cessée et enfant je me documentais dans la célèbre revue "Science & Vie", mais ce n'est qu'au lycée que j'ai enfin pu vivre cette passion et c'est grâce à Jean-Jacques Rives, professeur de physique/chimie et astronome amateur à ses heures perdues. Je te suis très reconnaissant d'avoir partagé ta passion avec moi et de m'avoir fait découvrir un domaine extraordinaire. C'est ce projet scientifique anodin qui consistait à faire tourner des pommes de terre sur un tourne-broche qui nous a propulsé de la Vendée à Paris puis à Los Angeles avec mes deux amis de toujours Titouan Coislier et Corentin Pasquier. Merci à tous les deux pour votre présence, votre soutien et pour tous ces moments partagés de rires et de bonheur autour d'un jeu de société de 5 heures. Mais ce projet n'aurait pas été réalisé sans l'aide de ma maman, Nathalie Chevalier, que je remercie plus que tout car elle a toujours été présente pour partager mon bonheur dans les bons moments et pour me soutenir et m'écouter dans les mauvais moments. Toujours positive et optimiste, elle a réussi à me propulser au plus haut toutes ces années et encore aujourd'hui, alors du fond du coeur un grand merci de ton fils qui t'aime et qui est fier de t'appeler maman.

Bien que passionné par l'astronomie et l'astrophysique, je n'étais pas parti pour faire une thèse car mon autre passion pour l'enseignement avait pris le dessus et je voulais prendre mon poste de professeur de mathématiques au lycée. La vie est pleine d'heureuses surprises car cette même année j'ai rencontré Anissa Khamkham, une femme magnifique et extraordinaire, passionnée de civilisation américaine que j'ai suivie aux Etats-Unis et avec qui je partage ma vie depuis. Je ne peux même pas quantifier la chance que j'ai de vivre avec toi chaque instant, d'échanger sur des sujets passionnants, de danser dans tes bras pendant des heures, de délirer sur tout et n'importe quoi. Tu es une personne formidable, drôle, intelligente et passionnante et je te remercie à l'infini de ton écoute pendant des heures sur mes exposés pointues de statistique Bayésienne, d'analyse multi-dimensionnelle et de machine learning. Ton soutien pendant toutes ces années m'a permis de toujours me surpasser et de vivre cette thèse de la meilleure des façons possibles. C'est cette année passée aux USA avec toi, pendant laquelle je faisais un stage sur la détection "éventuelle" de matière noire, qui m'a motivée à faire de la recherche en astrophysique et à commencer une thèse.

Le choix du sujet de ma thèse était bien entendu décisif mais pas autant que le choix des directeurs avec qui j'allais travailler, et je dois dire que sur ce point je n'ai pas été déçu. Ma directrice de thèse, Valérie de Lapparent, est quelqu'un d'extraordinaire dans la recherche et dans la vie de tous les jours. Tu as toujours été présente et tu as su te rendre disponible à chaque instant pour répondre à mes interrogations, toujours prête à faire une visio ou à passer un coup de téléphone pour m'aider sur un point bloquant. Mon autre directeur de thèse, Damien Le Borgne, est une personne passionnante avec qui j'ai eu le plaisir d'échanger sur de nombreux sujets. Je vous remercie infiniment pour ces trois années de thèse, pour tous vos conseils et enseignements qui m'ont rendus plus autonome, pour toutes vos explications qui m'ont rendues plus rigoureux, pour le temps que vous m'avez consacré sans le compter avec ces réunions de quelques minutes qui finalement duraient des heures et pour votre soutien dans chacune de mes initiatives. Merci pour ces trois années épanouissantes et passionnantes. Merci à Tom Charnock pour son expérience, pour le temps qu'il a passé à travailler avec moi, ce fut un plaisir d'échanger de nombreuses heures avec toi pendant ces trois années. Merci Emmanuel Bertin pour ton assistance technique et pour nos nombreux échanges productifs. Je remercie également mes rapporteurs, Laurence Tresse et Brice Ménard, pour leurs précieux conseils et retours ainsi que les autres membres de mon jury de thèse, Benjamin Wandelt, Micol Bolzonella et Marc Huertas-Company pour votre lecture attentive de mon travail. Un grand merci à ma marraine de thèse, Silvia Galli, pour nos échanges et tes conseils toujours justes et encourangeants.

Je voulais bien évidemment remercier mon père, Gérard Livet, pour ton soutien et pour tous ces repas ou nous échangions pendant des heures sur ces questions de l'infiniment grand. Merci aussi à mon beau-père, Jérôme Bacher, ta présence représente énormément pour moi, surtout quand tu m'accompagnais tôt le dimanche matin lorsque j'étais arbitre de foot et qu'il faisait -5 degrés dehors, ou encore pendant ces parties d'échecs infinissables, merci beaucoup. Merci à mes trois petits frères, Bertrand, Mathys et Killian, avoir grandi à vos côtés était un privilège pour moi et je suis fier de vous avoir et de partager tous ces moments avec vous. Merci à Clément Vince et à nos nombreuses sorties pour courir et se vider la tête et les poumons, merci aussi pour toutes ces soirées passées ensemble à discuter et à rire de tout en la compagnie de Luän. Merci à Célia et Loris Sequeira-Paoletti pour tous ces moments passées ensemble, pour ces voyages magnifiques partagés et pour ces soirées avec petit Boubou. Merci à Candice Durandet pour ces soirées passées à partager notre passion commune pour l'astronomie jusqu'au bout de la nuit et même jusqu'au levé du Soleil. Merci aussi à Stéphane Fauvaud pour ces missions astronomies au Pic du midi et à St Véran. Merci aux doctorants de l'IAP, François, Pierre, Lucas, Shweta, Lukas, Quentin, Raphaël, Alexandre, Aline et aux autres pour tous ces repas, cafés et discussions. Et bien entendu, merci à tous les gens que j'ai oublié.

Résumé

Etudier l'évolution des galaxies c'est comprendre comment chaque type morphologique a évolué et qu'elles sont les transformations possibles entre les différentes populations : elliptiques, spirales, irrégulière... Avec l'approche classique de modélisation inverse, des catalogues photométriques de galaxies sont extraits depuis les images afin de modéliser l'évolution avec le redshift des flux, des couleurs et des tailles des galaxies. Les flux apparents sont convertis en flux absolus en utilisant les distances, et la distribution de ces flux est caractérisée par des "fonctions de luminosité". Cependant, ces dernières sont sujettes à de nombreux effets d'incomplétudes qui peuvent biaiser les résultats.

Pour éviter cela, dans cette thèse, nous utilisons une approche de modélisation directe : le modèle crée des images multibandes de champs profonds qui reproduisent les caractéristiques instrumentales et observationnelles du champ D1 d'un degré carré du Canada-France-Hawaii Telescope Legacy Survey (CFHTLS). Les images simulées peuvent donc être comparées directement aux observations car elles contiennent les mêmes effets de sélection.

Ce travail décrit une nouvelle méthode de compression extrême d'images par réseaux de neurones qui permet de se passer de l'extraction de catalogues photométriques ainsi que des effets d'incomplétudes et des corrections à y apporter. L'information contenue dans la fonction de luminosité étant beaucoup trop complexe à décrire analytiquement, nous combinons ici la compression extrême des données avec des techniques de statistiques Bayésiennes pour inférer depuis les observations les valeurs des paramètres du modèle les plus probables.

Abstract

Studying galaxy evolution is understanding how each morphological type has evolved and what are the possible transformations between the different populations: ellipticals, spirals, irregulars... In the classic inverse modeling approach, photometric catalogs of galaxies are extracted from the images and are used to model the joint evolution with redshift of the fluxes, colors and sizes of galaxies. Apparent fluxes are converted into absolute fluxes using estimates of distances, and the distribution of the absolute fluxes of galaxies is characterized using what is called “luminosity functions”. However, they are subject to a variety of incompleteness effects which are likely to bias the results.

In order to circumvent these issues, in this thesis, we use the forward modeling approach: the forward model creates realistic multiband deep field images that reproduce both the instrumental and observational characteristics of the one squared degree Canada-France-Hawaii Telescope Legacy Survey (CFHTLS) D1 deep field. Therefore, the simulated multiband deep field images face the same selection effects as those of the observed survey and this allows for a direct comparison of the two in the “observable space”.

This work describes a novel method of massive data compression using neural networks which completely bypasses the catalog stage, along with correcting for all the biases induced by source extraction and photometry. Because the luminosity function information is so complex to describe analytically, we combine massive compression with approximate Bayesian computation techniques in order to infer from the observation the most probable parameter values of our forward model.

Résumé long

Pendant cette thèse, nous avons développé une nouvelle méthode pour contraindre de manière robuste les paramètres des fonctions de luminosité des galaxies elliptiques et spirales en utilisant une compression massive d'images panchromatiques par réseau de neurones et des techniques d'inférence sans nécessité d'exprimer la fonction de vraisemblance. Cette approche ne nécessite donc aucune extraction de catalogues de flux ou de taille des galaxies et permet une analyse directe des images dans 8 bandes photométriques de l'ultraviolet proche à l'infra-rouge proche : u', g', r', i', z', J, H, Ks. L'utilisation de simulations d'images réalistes obtenues en tenant compte des conditions d'observations des données réelles prend automatiquement en compte les effets de sélections complexes qui affectent ces mêmes données et constitue le point central de cette approche. En effet, les modèles actuellement utilisés pour décrire l'évolution des galaxies sont contraints par des catalogues extraits des images contenant la liste des flux et des tailles des galaxies. Ces catalogues sont affectés par de nombreux effets de sélection (*i.e.* assombrissement cosmologique, correction-K, extinction par la poussière, biais de Malmquist, biais d'Eddington) et des biais d'extraction (*i.e.* estimation du fond de ciel, fausses-détections, séparation étoiles/galaxies) qui peuvent être très corrélés et impossible à exprimer analytiquement.

Des populations synthétiques de galaxies elliptiques et spirales (*i.e.* les deux types dominants de l'Univers) sont simulées en utilisant un modèle de l'évolution des galaxies. Ces dernières sont échantillonnées via des fonctions de luminosité pour chaque population et décomposées en la somme d'un bulbe et d'un disque décrits par un ensemble de distributions d'énergie spectrale (SEDs). Le modèle est rendu encore plus réaliste en introduisant (1) l'extinction interne par la poussière du bulbe et du disque de chaque galaxie séparément, ainsi que (2) le rougissement des objets causé par la Voie Lactée et (3) les étoiles synthétiques du modèle stellaire de Besançon. Le programme "SkyMaker" est ensuite utilisé pour peindre ces galaxies simulées sur des images panchromatiques de taille 1024×1024 pixels dans les bandes optiques u', g', r', i', z' de l'instrument MegaPrime et dans bandes infrarouges J, H, Ks de l'instrument WIRCam. Les effets de sélection reproduits avec réalisme par "SkyMaker" afin que ces simulations soient comparables aux données réelles des images issues du Canada-France-Hawaii Telescope Legacy Survey sur le champ D1. Ce relevé profond d'un degré carré obtenu avec le télescope de 3.6m de diamètre contient plusieurs centaines de milliers de galaxies avec une résolution angulaire remarquable.

Les images simulées par cette technique sont ensuite utilisées pour entraîner un réseau de convolution afin d'extraire l'information contenue dans les simulations concernant les paramètres du modèle pour chaque fonction de luminosité. Cette méthode de compression par réseau de neurones utilisée est l'application de l'algorithme "Information Maximising Neural Network" (IMNN) développé par Charnock et al.(2018). Le réseau de neurones possède une architecture de type "Inception" (voir Szegedy et al., 2015) et est entraîné à maximiser l'information de Fisher des images, permettant ainsi une réduction considérable de la dimension des images avec une taille finale correspondant au nombre de paramètres du modèle. Contrairement aux techniques d'apprentissage profond classiques, ce réseau Inception peut être entraîné avec seulement quelques centaines d'images. Une fois le réseau de neurones entraîné sur des séries de simulations spécifiques, deux procédures d'inférence Bayésienne sont utilisées : (1) Approximate Bayesian Computation (ABC) et (2) Population Monte Carlo (PMC). En partant d'une distribution a priori uniforme de l'espace des paramètres, ces procédures permettent de contraindre les valeurs des paramètres des fonctions de luminosités utilisées pour reproduire au mieux les données réelles. Ces techniques sont appliquées non seulement aux données réelles, mais aussi à des données synthétiques afin de garantir le fonctionnement de la méthode et la fiabilité des résultats. Nous avons également développé une procédure joint-PMC afin de combiner (au sens Bayésien) les résultats d'inférences provenant de plusieurs images de taille 1024×1024 pixels.

Cette méthode prouve son efficacité pour contraindre des paramètres non-triviaux/corrélés des fonctions de luminosité pour des populations de galaxies elliptiques et spirales : l'amplitude et la magnitude caractéristique. En appliquant les deux méthodes Bayésiennes présentées au-dessus et en utilisant le réseau Inception, il est possible d'inférer les valeurs d'amplitude des fonctions de luminosité des populations de galaxies elliptiques et spirales dans les données réelles du CFHTLS. Les fonctions de luminosité dérivées par notre méthode sont globalement en accord avec celles obtenues par la méthode standard utilisant l'extraction de catalogues sur des relevés profonds : López-Sanjuan et al. (2017); Brown et al. (2007); Beare et al. (2015); Salimbeni et al. (2008); Drory et al. (2009); Faber et al. (2007); Zucca et al. (2006); Zucca et al. (2009). L'ensemble de notre travail est expliqué en détails dans l'article publié dans le journal Astronomy

& Astrophysics : Livet, Charnock, Le Borgne, de Lapparent 2021 (A&A 652, A62, 2021).

Si la méthode développée a été appliquée sur deux paramètres des fonctions de luminosité des galaxies elliptiques et spirales, un objectif est de tester son efficacité sur les 5 paramètres possibles : amplitude, magnitude caractéristique, pente des galaxies faibles, évolution d'amplitude et évolution de magnitude caractéristique. De plus, la méthode présentée décompose chaque galaxie en deux composantes (bulbe et disque) et pourrait être complétée par une décomposition plus réaliste développée dans cette thèse car en trois dimensions et avec trois composantes (bulbe, disque épais et disque mince). Cela permettrait de mieux modéliser les galaxies contenant de la poussière et d'obtenir des gradients de couleurs dans les galaxies comparables à ceux des données observées. Ce code est prêt à l'emploi mais n'a pas été mis en place dans le modèle simple actuellement utilisé.

D'autre part, l'architecture du réseau de neurones pourrait éventuellement être améliorée en testant de nouvelles possibilités comme :

- Une architecture "ResNet Inception" en vue d'une amélioration de l'architecture Inception en ajoutant à chaque sortie de bloc l'entrée de ce bloc : ainsi le flot d'informations serait conservé d'un bloc à un autre et l'apprentissage serait plus robuste surtout dans les premières couches de neurones du réseau.
- Augmenter/Réduire le nombre de composantes en sortie de compression par rapport au nombre de paramètres à déduire afin de voir si l'apprentissage peut être plus rapide et/ou plus précis.

Enfin, les méthodes d'inférence Bayésiennes utilisées (ABC/PMC) sont gourmandes en temps de calcul et peuvent devenir inefficaces lorsque le nombre de paramètres du modèle devient grand. Une technique récente développée par Alsing et al. (2019) utilisant des "neural density estimators" (NDE) permettrait de généraliser les méthodes d'inférence développées pendant cette thèse et de gagner du temps de calcul afin de contraindre les paramètres d'un modèle multi-paramétrisé.

Contents

Introduction

The diversity of processes driving galaxy evolution takes astronomers to work on very large scales and leaves us with the challenging study of the most beautiful structures of the Universe. These fantastic objects hide their history within their composition and release information in the form of light that we collect through the state-of-the-art telescopes and that we process with the best computers ever created. Traditionally, the study of galaxy evolution is based on the analysis of large sets of photometric surveys with long exposure times and a wide range of observing filters, sometimes complemented by spectroscopic surveys. These data are used with the goal of answering specific questions about galaxy evolution, such as the star-formation histories, the gas accretion histories, the morphological transformations, the merging rates for galaxies of various types and masses, and the role of environment in these various characteristics. "Inverse modeling" and "forward modeling" are two opposed approaches regarding how these images are used to derive models of galaxy evolution.

In the classic inverse modeling approach, photometric catalogs of galaxies are extracted from the images and are used to model the joint evolution with redshift of the fluxes, colors and sizes of galaxies. Apparent fluxes are converted into absolute fluxes using estimates of distances, and the distribution of the absolute fluxes of galaxies is characterized using what is called "luminosity functions". Their variations with galaxy type and with distance provide us with clues on the transformations that galaxies undergo during their evolution across cosmic times.

Measuring luminosity functions requires to extract complete catalogs of galaxy fluxes. Several well-documented packages are available to perform the extraction of catalogs from survey images such as `FOCAS` (Jarvis and Tyson 1981), `APM` (Maddox et al. 1990) or `SExtractor` (Bertin and Arnouts 1996). Rigorous studies based on flux-limited catalogs can be found in the literature, however they are subject to a variety of incompleteness effects which are likely to bias their results. These selection effects are numerous and can have different impacts on the extracted catalogs: in flux-limited catalogs, the faint galaxies near the sky background are ignored by construction because of the confusion noise (Condon 1974), which inevitably leads to the Malmquist bias (Malmquist 1922, 1925). Poor segmentation, morphological biases and the lack of detection of some galaxies are direct consequences of the cosmological dimming (Tolman 1934; Calvi et al. 2014), that is the decrease of the surface brightness of galaxies by a factor $(1+z)^{-4}$. The Eddington bias (Eddington 1913) is another statistical selection effect that boosts the number of galaxies observed just at the detection threshold and can cause overestimates of the galaxy fluxes. The stellar contamination of deep surveys (Pearson et al. 2014) is another limitation of faint galaxy catalogs because some stars can be misidentified as galaxies and alter the results. Furthermore, redshift changes the fluxes and the colors of galaxies, and K-corrections (Hogg et al. 2002) must be applied to correct for these modifications. Additional biases also arise because of the extraction process, that are the choice of the model point spread function (PSF), the type of magnitude used to perform the extraction (isophotal, aperture, or by profile modeling) and the confusion of sources. Moreover, these various selection effects are correlated in subtle ways that are impossible to express analytically, besides some of them may be spatially variable within the same survey (Bernardi et al. 2017, Appendix B). As a result, the classic inverse modeling approach must correct the extracted catalogs from the consequences of those entangled biases (Marzke 1998; Taghizadeh-Popp et al. 2015). If this is not successful, the models derived from these catalogs tend to be biased in a nontrivial way, in particular the luminosity functions.

In order to avoid solving the hard inverse modeling inference from catalogs, forward modeling coupled with approximate Bayesian computation (ABC) can be used. This technique has been applied in many fields and for various applications, for example, in cosmology (Akeret et al. 2015; Kacprzak et al. 2018), in modeling the initial mass function (Cisewski-Kehe et al. 2019), to study type Ia supernovae (Weyant et al. 2013), and in modeling galaxy evolution (Carassou et al. 2017; Tortorelli et al. 2020). In the specific case of galaxy evolution, a forward model describes several galaxy populations using their luminosity function and is used to measure the evolution of such luminosity functions simultaneously.

The forward model creates realistic multiband deep field images that reproduce both the instrumental and observational characteristics of a given survey such as the exposure time, the filters used, the optics of the telescope, the properties of the detector, the seeing, etc. Therefore, the simulated multiband deep field images face the same selection effects as those of the observed survey and this allows for a direct comparison of the two in the "observable space". In their recent work, Carassou et al. (2017) have developed a method for binning extracted catalogs in fluxes and sizes in order to infer the parameters of the luminosity functions used in their forward model. They successfully measured the evolution parameters of one and two galaxy-type luminosity functions by comparing the distribution of simple observables extracted from the images, such as the flux and size of galaxies. However, the binning of these distributions had to be limited to ten intervals per band (of eight bands), which nevertheless meant that the study took place in a very sparsely populated 10^8 dimension space. Moreover, some of the information from such deep images could have been lost by compressing it into flux and size catalogs: *i.e.* the correlations between the fluxes and sizes of galaxies according to their morphological types, the galaxy profiles, etc.

To circumvent all of these problems simultaneously, this thesis describes a novel method of massive data compression using neural networks in order to enable a direct, hence catalog-free, comparison of the observed and simulated multiband deep field images. This approach completely bypasses the catalog stage, along with all the biases induced by source extraction and photometry. The compression method we use is called the "information maximizing neural network" (IMNN; Charnock et al. 2018) algorithm: it fits a neural network that is only sensitive to the effects of the model parameters in the simulations obtained from our forward model. We implement this method for the first time on deep and large multiband deep field images of galaxies: the $\sim$1 deg^2 common area of the Canada-France-Hawaii Telescope (CFHT) Legacy Survey D1 deep field observed in the ultraviolet/optical, using the MegaPrime instrument in the u', g', r', i', z' filters, and the WIRDS D1 deep field in the near-infrared using the WIRCam instrument in the J, H, Ks filters (also on the CFHT). We use the final releases of the TERAPIX processed data for each survey, which are T0007 for CFHTLS (Hudelot et al. 2012) and T0002 for WIRDS (Bielby et al. 2012). These observed images contain a huge wealth of information for studying the evolution of the various galaxy populations with nearly 400,000 galaxies per square degree with exquisite seeing of ~ 0.6 arcseconds. However, the large range of luminosities of a given galaxy type, as well as the diversity of galaxy types, and their common intervals of luminosities cause the luminosity function information to be so complex that it is impossible to construct a likelihood function which describes the probability of obtaining a given deep field. This is the reason why we combine massive compression with approximate Bayesian computation techniques in order to infer the most probable parameter values of our forward model by comparing observed deep fields with similar simulated images. This thesis is divided into seven thematic chapters as follows.

Chap. 1 briefly describes the cosmological context along with the characteristics, properties and processes which drive galaxy evolution. This first chapter also provides a detailed description of the various populations of galaxies and the common decomposition as a bulge and a disk. Galaxies are often studied using wide and/or deep surveys that are also discussed in this chapter; more specifically, the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS) offers a very large multiband statistical galaxy sample which makes it an ideal choice for this analysis.

The goal of Chap. 2 is to introduce the luminosity functions, that is a commonly used statistical tool to study the evolution with redshift of the number density of galaxies as a function of their brightness. To do so, Chap. 2 describes the various selection effects altering the survey images and presents the technique of extracting catalogs of apparent magnitudes from these surveys. The classic inverse modeling approach aims at modeling these selection and extraction biases and attempts to remove their effects from the derived catalogs of absolute magnitudes. This chapter explains why this approach is however limited because the selection effects can be highly correlated and location-dependent, which inevitably makes the derived galaxy evolution models biased in a non-trivial way. An application of the extraction process with the `SExtractor` program (Bertin and Arnouts 1996) is finally given for a simulated multiband deep field image with three populations of galaxies (*i.e.* elliptical, spiral and irregular) and stars from which the photometric redshift of the sources are estimated with the `ZPEG` program (Le Borgne and Rocca-Volmerange 2002) and the V/V_{max} method (Schmidt 1968) is applied to derive the luminosity functions.

Chap. 3 describes the forward modeling approach and the model used in this thesis. The forward model produces realistic multiband deep field images and is parameterized with the parameters of the luminosity functions used to study the different populations of galaxies. This chapter details the ingredients used in our forward model such as the spectral energy distributions of the bulge and disk components of galaxies, the galactic dust extinctions, the stars and the image generator `SkyMaker`. This chapter also extends the internal dust extinction of galaxies to a 3D version accounting for various color and shape related effects of the distribution of light in the modeled galaxies.

Because each multiband deep field image is a unique realization, the comparison of two images cannot be applied pixel by pixel and summary statistics must be extracted from the images. Chap. 4 therefore aims at proposing a compression scheme of the multiband deep field images through neural networks in order to extract summary statistics while preserving the Fisher information content of the images. This chapter develops the theoretical framework of the information maximising neural network (IMNN; Charnock et al. 2018) compression process and describes the inception network that is trained to optimally compress the data.

Because the likelihood function of our forward model is intractable, as it is too costly to evaluate, we introduce in Chap. 5 the Bayesian likelihood-free framework. This chapter gives an overview of the two likelihood-free techniques that are applied in this thesis, namely approximation Bayesian computation (ABC) and population Monte-Carlo (PMC). Both techniques aim at inferring the most probable parameters for our forward model in order to produce multiband deep field images that are as similar as possible to the cutout images we use from the observed CHFTLS D1 deep field. This chapter also tests the efficiency of the full IMNN, ABC and PMC pipeline designed in this thesis on a first application with only one spiral population and two parameters, $\log_{10}(\phi^*)$ and M^* and synthetic data for which we know the luminosity function parameters. This application demonstrates that the method operates as expected on virtual data and that we are able to accurately retrieve the input parameters as well as the strong correlation between them.

Chap. 6 attempts to fulfill the main objective of this research, that is to infer the values of the model parameters directly from sub-images of the far-ultraviolet/optical/near-infrared multiband CFHTLS D1 deep field. To do so, in this chapter, two populations of elliptical and spiral galaxies are considered in our forward model. In this second application, we only infer the density parameter ϕ_0^* for these two populations simultaneously from five insets of the CFHTLS D1 image. The other luminosity function parameters are fixed to the values that are measured by López-Sanjuan et al. (2017) using the catalog-extraction approach, in particular the forward model allows the luminosity functions to evolve with redshift with distinct parameters for the elliptical and the spiral populations. The results of this application prove that the methodology developed in this thesis can be efficiently applied when two populations are studied at the same time and that the luminosity function results are fully compatible with other inverse modeling analyses. The third application is an ongoing work that aims at reaching the ultimate goal of this research project that is simultaneously inferring six parameters, ϕ_0^*, M_0^* and α for the elliptical and spiral populations.

Chap. 7 gives an overview of improvements that could be made to this thesis to improve the accuracy of the results and to reduce the computing time of the full procedure. This chapter first discusses the various refinements that could be applied to the forward model in order to generate the simulations faster and with more realistic details. We also examine the upgrading methods that could be used to make the training of the inception network faster and more accurate in order to study the various number density and size effects that the luminosity function parameters produce on the images. Finally, we describe the techniques that could be explored to improve the PMC procedure to obtain posterior distributions with a faster convergence, or even to completely bypass this procedure with the use of neural density estimators.

The research work done during this thesis led to the following publication:

F. Livet, T. Charnock, D. Le Borgne, and V. de Lapparent. Catalog-free modeling of galaxy types in deep images. Massive dimensional reduction with neural networks. *A&A*, 652:A62, Aug. 2021. doi: 10.1051/0004-6361/202140383

Chapter 1

Studying galaxy evolution

> *Keep up the good work,*
> *if only for a while,*
> *if only for the twinkling of*
> *a tiny galaxy.*
>
> `Wislawa Szymborska`

Abstract

Galaxies are fascinating objects to admire and to study because of their diversity of colors and shapes. These entities result from the gathering of billions of stars and a colossal amount of gas and dust by gravity. Understanding their properties and reconstructing their evolution is fundamental to describe the Universe itself. This chapter gives an overview of the known properties of galaxies and of their evolution in a cosmological context. We also introduce the main existing deep and wide surveys that are used to study the different populations of galaxies. From this non-exhaustive list of main surveys, we explain why the Canada-France-Hawaii Telescope Legacy Survey is an ideal choice for the analysis carried out in this thesis.

Contents

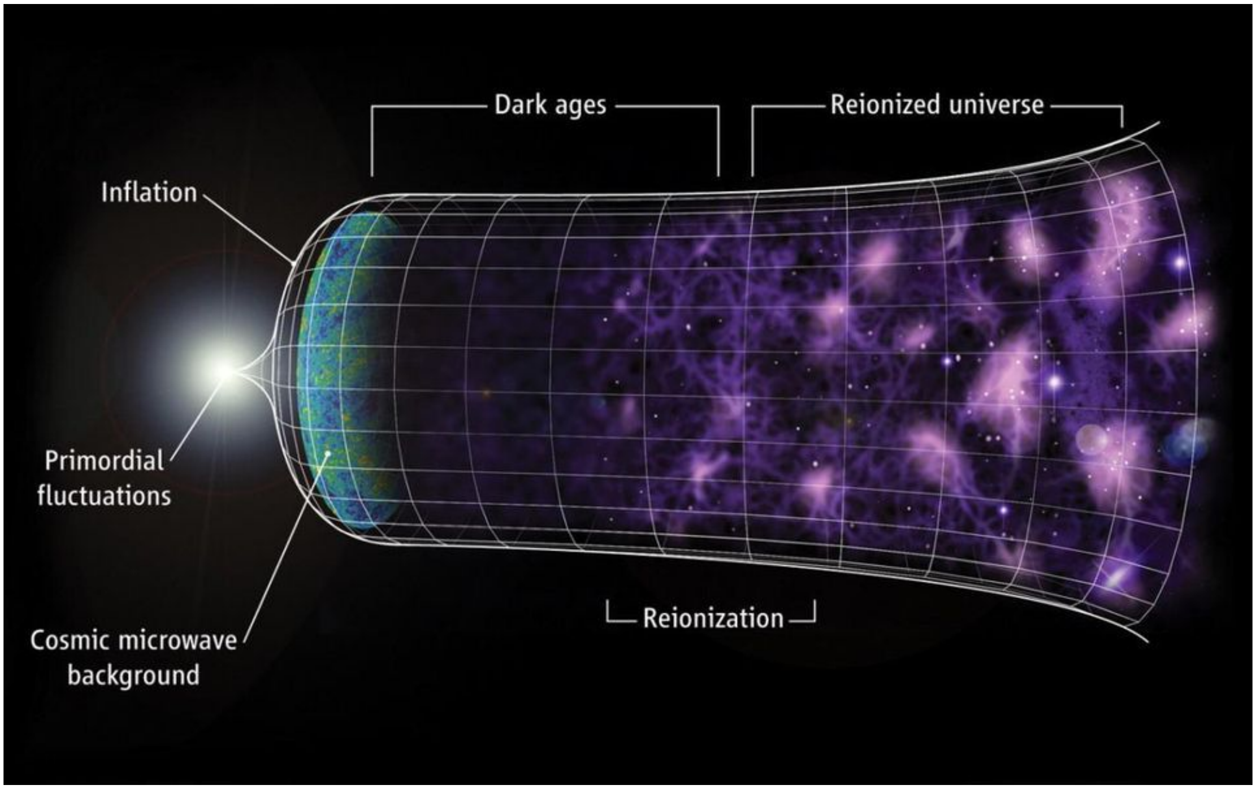


Fig. 1.1: Schematic representation of the expansion of the Universe from Faucher-Giguère et al. (2008) which shows the different periods of galaxy formation and evolution leading to the actual observable Universe.

1. Cosmological context

The Λ**CDM model** (Cold Dark Matter) is the name of the cosmological model that best describes the observable Universe. In particular, this model explains the abundance of light elements (such as hydrogen, helium and lithium), the acceleration of the homogeneous and isotropic expansion of the Universe because of dark energy (Λ cosmological constant, Riess et al. (1998)), the existence and properties of the Cosmic Microwave Background (CMB, Gamow (1948); Penzias and Wilson (1965)) and the distribution of the galaxies at large-scale structure (Tegmark et al. 2004). We use this standard framework to reconstruct the history of galaxy evolution.

Recombination refers to the period when stars and galaxies first began to form 377,000 years after the Big Bang (which corresponds to a redshift $z \approx 1100$). Before that time, the Universe was a hot plasma of mostly electrons and protons and no neutral atom could form. However, at the epoch of recombination, because of its expansion, the temperature of the Universe dropped below 3000K and the first hydrogen atoms were formed. This relatively rapid event freed the photons that were captured beforehand by the electrons, which was the first emission of light called the CMB. Even if the recombination was remarkably homogeneous and isotropic, some overdense regions were created and would grow to form the first galactic structures.

The **Dark Ages** is the name given to the epoch just after the recombination. Due to the expansion of the Universe, the CMB light emitted in the visible was rapidly redshifted to the infrared (IR) and the Universe became dark to the human eye. During this period, the temperature of the Universe progressively dropped and the Universe once hot became cold. This period lasted for several million years before the first luminous objects (massive population III stars, galaxies and quasars) formed and reionized the neutral gas: thus leading to the epoch of **reionization**.

There is no clear boundary between the end of the Dark Ages and the beginning of the reionization as shown by the large-scale polarization study of the CMB of Planck Collaboration et al. (2016), in which they suggest that the epoch of reionization might have started around $z = 14$ (*i.e.* 300 million years after the Big

Bang). Another study of the spectra of high redshift quasars of Fan et al. (2006) indicates that the Universe was fully reionized around $z = 6$ (*i.e.* 940 million years after the Big Bang). The properties of these first million years of existence of the Universe are important to study to better understand the origin of galaxies, which is part of the mission of the James Webb Space Telescope (JWST) scheduled to be launched in the near future. A schematic view of the expansion of the Universe is given in Fig. 1.1.

Today's Universe, after around 13.8 billion years of evolution, is cold and mostly empty and contains galaxies, galaxy clusters, filaments and superclusters that constitute the cosmic web: galaxies form in halos of gas and dark matter that are distributed along the cosmic web. The dynamic of the cosmic web is driven by dark energy at large scales which constitutes 68.3% of the total components of the Universe (other constituents are 26.8% dark matter and 4.9% ordinary matter) according to Planck Collaboration et al. (2014). There are tensions in the values of the cosmological parameters of the ΛCDM model, for example between the Dark Energy Survey and the Planck datasets, see Handley and Lemos (2019). In this dissertation, we use the values: $H_0 = 70$ km s^{-1} Mpc^{-1} (Hubble constant), $\Omega_m = 0.3$ (matter density parameter) and $\Omega_\Lambda = 0.7$ (dark energy density parameter).

2. Galaxy populations and colors

Thanks to Hubble (1926a, 1936), astronomers use a system of galaxy morphological classification to divide galaxies into groups based on their visual appearance: **the Hubble sequence** (see Fig. 1.2). This diagram divides galaxies into ellipticals (E), lenticulars (S0) and spirals (S). The spiral population is also divided into the normal spiral branch and the barred spiral branch. Other galaxies that do not fit into these categories are called Irregulars (Irr). Since then, the original sequence has been improved several times by Morgan (1958); de Vaucouleurs (1959); Sandage (1961); de Vaucouleurs et al. (1991), but is still called the Hubble sequence. The elliptical population together with the lenticular population are often referred to as **early-type** galaxies, while spirals and irregulars are often referred to as **late-type** galaxies. This common nomenclature leads to the logical but mistaken belief that the Hubble sequence was made to reflect an evolution from ellipticals to spirals which is not true.

Short description of each population:

- **Ellipticals (E):** These galaxies generally have smooth and featureless light distributions and they appear as ellipses on images. The integer associated with their name represents the degree of ellipticity. Most of the elliptical galaxies are composed of old low-mass stars with too little gas and dust quantities to form new stars.
- **Spirals (S):** These galaxies are composed of a relatively flat gas-rich disk with spiral arms forming stars and a central bulge of old stars. Roughly two-thirds of the spiral galaxies have a bar-like structure running through their center (and are denoted SB) fueling stars into the bulge. Spirals disks are often rich in dust.
- **Lenticulars (S0):** These galaxies are intermediate between the elliptical and spiral populations and are composed of a bright central bulge and a disk-like structure. The disks of lenticular galaxies do not contain spiral structures and show essentially no star formation. Some rare lenticular disks contain dust.
- **Irregulars (Irr):** These galaxies do not show a distinct regular shape. They generally contain abundant amounts of gas and dust and could have been deformed by external gravitational forces.

The Sloan Digital Sky Survey (see Sect. 5.2) has allowed the determination of large statistical samples of galaxies with visual morphological types from digital images (Lintott et al. 2008; Nair and Abraham 2010; Willett et al. 2013), as well as to revisit the morphological classification of the most nearby, hence well resolved galaxies, and to study their morphological properties along the Hubble sequence (Baillard et al. 2011; de Lapparent et al. 2011).

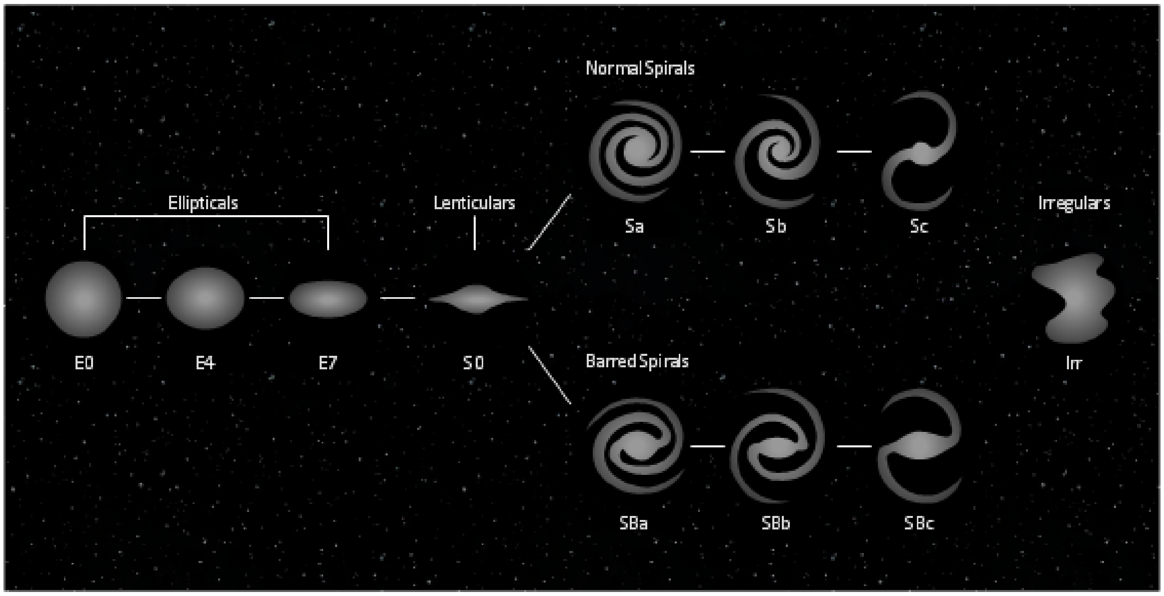


Fig. 1.2: Hubble sequence showing the diversity of different populations of galaxies (e.g. ellipticals, lenticulars, spirals, irregulars). The spiral population is divided into two branches: the "normal" spirals and the "barred" spirals. For the elliptical and linticular populations, the number indicates the degree of ellipticity of each type.

Galaxies are made up from stellar populations that give them their colors which are in general well correlated with their structural type as shown by Holmberg (1958); de Vaucouleurs (1961). This relation is summarized in a **galaxy color-magnitude diagram** with three main features: the **red sequence**, the **blue cloud** and the **green valley**. As indicated by their names, the red sequence brings most of the red early-type galaxies which generally are from the elliptical and lenticular types, the blue cloud includes most late-type galaxies which generally are from the spiral and irregular types. In between these two categories lies the green valley gathering two transitional subgroups of blue late-type galaxies: a small part of spheroids with a quenched star formation and a majority of late-type (*i.e.* disk) galaxies with exhausted gas supplies, see Schawinski et al. (2014). However, the galaxy color-magnitude diagram does not allow for the determination of the galaxy properties because it evolves with time as shown by Willmer et al. (2006); Blanton (2006); Faber et al. (2007); Wyder et al. (2007).

3. Galaxy description

A galaxy is the gathering of stars, gas, dust, dark matter and sometimes with a supermassive black hole at its center. Galaxies contain in general between 10^{10} stars for the dwarf galaxies and 10^{12} stars for the giant galaxies. These objects have a wide range of masses between $10^6 M_\odot$[*] for the dwarf galaxies and $10^{13} M_\odot$ for the giant ones. These entities can have different sizes between 10^2 pc[**] for the most dense galaxies and 10^5 pc for the most extended ones. Following the results of Conselice et al. (2016), they have estimated that there are approximately 2×10^{12} galaxies with mass at least $10^6 M_\odot$ in the observable Universe (up to $z \approx 8$).

Galaxies generally have two main components: a **bulge** and a **disk**, with different properties explained in the next parts. The other structural components of galaxies (such as bars, spiral arms, rings, etc) are not studied in this thesis but the reader is referred to Kormendy (2013) for a more detailed description.

---

[*] $M_\odot$ is the symbol of the solar mass: $1 M_\odot \approx 2 \times 10^{30}$ kg.

[**] pc is the symbol of the parsec, a unit of length used to measure large distances: 1 pc $\approx 3 \times 10^{16}$ m.

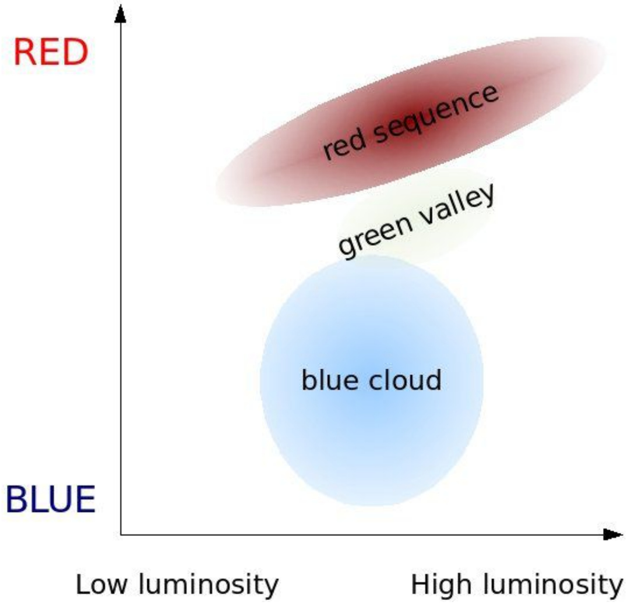


Fig. 1.3: Galaxy color-magnitude diagram with the three main features: the red sequence, the blue cloud and the transitional green valley. This diagram gives the distribution of the galaxies as a function of their luminosity and color. The red sequence gathers most of the elliptical and lenticular galaxies while the blue cloud gathers most of the spiral and irregular galaxies.

3.1 Bulge

Bulges are in the central region of the galaxies and are composed of dense groups of stars. In today's Universe, it has been shown by Driver et al. (2007); Gadotti (2009) that bulges gather approximately one third of the total stellar mass. Furthermore, as shown by Pérez and Sánchez-Blázquez (2011), the stellar populations of bulges present a large range of ages and metallicities. Seeing bulges as small elliptical galaxies is over-simplistic because a bulge of a spiral galaxy can have a very different evolution history (thus different properties) than an elliptical galaxy, see Gadotti (2009). Moreover, most bulges have a supermassive black hole at their center whose mass was shown by Younger et al. (2008) to be related to the mass of the host bulge. Bulges grew in mass during merger events of the early Universe, as shown by Aguerri et al. (2001) but also because of the migration of huge star-forming clouds to the center of galaxies, see Bournaud et al. (2007). Bulges are generally described within two classes with different properties: **classical bulges** and **disk-like bulges**.

Bulges with properties close to those of elliptical galaxies are called classical bulges. These bulges are essentially composed of old stars, hence red, in relatively random orbits giving the bulge its spherical shape. In classical bulges, the quantity of gas and dust is generally very low, therefore they have very low star formation rates. As suggested by Kormendy et al. (2010), around 80% of field galaxies lack a classical bulge compared to only 30% in dense clusters of galaxies. Moreover, the study of Sachdeva and Saha (2016) showed that these proportions were roughly constant for the last 8 billion years. These results indicate that classical bulge formation and evolution are in close relationship with the environment of the galaxy.

Other bulges have properties closer to the inner regions of spiral galaxies than to those of ellipticals and are called pseudo-bulges (or disk-like bulges). The stars of the pseudo-bulges have orbits closer to the orbits of the stars in the outer disk dominated by circular motions. As shown by Kormendy and Kennicutt (2004), the pseudo-bulges generally possess gas and dust boosting their star formation. When no merger event happens for several billion years (*i.e.* secular evolution), because of instabilities, most pseudo-bulges can rearrange their stars, dust and gas and form a bar (Saintonge et al. 2012).

3.2 Disk

Disks are in the outer region of lenticular and spiral galaxies and are composed of stars, gas and dust. The stellar orbits are mostly circular about the galactic center and most of the gas in galaxies lies in their disk. This gas is the fuel for the formation of new stars. The study of both the stars and the gas rotation velocities versus the radial distance from the center of a galaxy allows one to draw the **galaxy rotation curve** which is one of the observational motivations for dark matter. Moreover, it has been shown by Tully and Fisher (1977) that the circular velocity of the gas in the disk is strongly correlated with the intrinsic luminosity of the galaxy (Tully-Fisher relation).

Most stellar disks can be divided into two components: the **thin disk** and the **thick disk**. First studied by Burstein (1979); van der Kruit (1981), and later extended by Yoachim and Dalcanton (2006), thin disks are generally composed of young blue stars with rich metallicities and relatively short scale heights and scale lengths. On the contrary, thick disks are composed of older redder stars with poor metallicities and larger scale heights and scale lengths. Moreover, Bell and de Jong (2000) showed that there is a radial gradient of stellar populations with inner regions of the disk containing older and more metal-rich stars than outer regions. Shen et al. (2003) showed that the massive and brighter galaxies tend to have larger scale lengths of their thin and/or thick disks. Galaxy disks are thoroughly described in van der Kruit and Freeman (2011). In the present thesis, we mainly use the galaxy decomposition as bulge + single disk for computation efficiency (see Chap. 3 for a full description), however as explained in Chap. 3 Sect. 3., a 3D model of dust extinction uses the galaxy decomposition as bulge + thick disk + thin disk.

4. Processes driving galaxy evolution

Galaxy evolution involves multiple processes which can be connected and affect the galaxy components in several ways. We give here a very brief and incomplete overview of the main processes. First, galaxies might form in **dark matter halos** and evolve geometrically and dynamically in various **environments**. They are composed of various **stellar populations** which evolve differently depending on their mass and metallicity. At the end of their evolution, massive stars explode in **supernovae** releasing large amounts of **dust** and **gas** that heat their surrounding interstellar medium and regulate the subsequent **star formation**, see (Mac Low and Klessen 2004). The interstellar medium is progressively enriched in atoms heavier than helium by the series of birth and death of stars richer and richer in heavy elements. The dust of the interstellar medium also absorbs the ultraviolet (UV) photons emitted by stars which contributes to both dim and redden the global galaxy UV light. The corresponding absorbed UV energy is then re-emitted at far-IR wavelengths.

Moreover, galaxies sometimes possess **active galactic nucleus** (AGN) which might be fueled by the accretion of matter onto the **supermassive black hole** at the center. Galaxy evolution depends on their surrounding environment: galaxies can be isolated and develop features such as **bars** and **spiral arms**, on the contrary, in populated environments, they can **merge** more or less violently with other galaxies which highly impact their star formation history and dynamics, see Lin et al. (2008). Because of the complexity and the large range of spatial scales of all these processes, it is very complicated to study and to understand their interactions which are necessary to explain the distribution of galaxies observed in the Universe. Therefore, there are two main approaches reviewed by Somerville and Davé (2015) to model these processes: **cosmological simulations** and semi-analytic models. These two important tools allow the modeling and understanding of the astrophysical processes such as those mentioned above, as well as others: **shock heating, radiative cooling, photoionization**, etc.

Cosmological simulations are a powerful tool to study galaxy evolution on large time scales such as millions even billions of years. These simulations allow for a better understanding of the formation of large scale structures and the growth of galaxy masses through merger events. Cosmological simulations in general consist in studying the evolution of the ΛCDM model in terms of dark matter and gas distributions on large cosmic time scales. To name only a few, state-of-the-art cosmological simulations are Illustris (Vogelsberger et al. 2014), Horizon-AGN (Dubois et al. 2014), EAGLE (Crain et al. 2015), IllustrisTNG (Pillepich et al. 2018; Springel et al. 2018; Nelson et al. 2018), FIRE2 (Hopkins et al. 2018) and SIMBA (Davé et al. 2019).

Semi-analytic models use simplified models to study some of the processes listed above with an accuracy depending on the complexity of each model. In general, semi-analytic models are much less computationally expensive than cosmological simulations and can explore the model's parameters efficiently. In the early

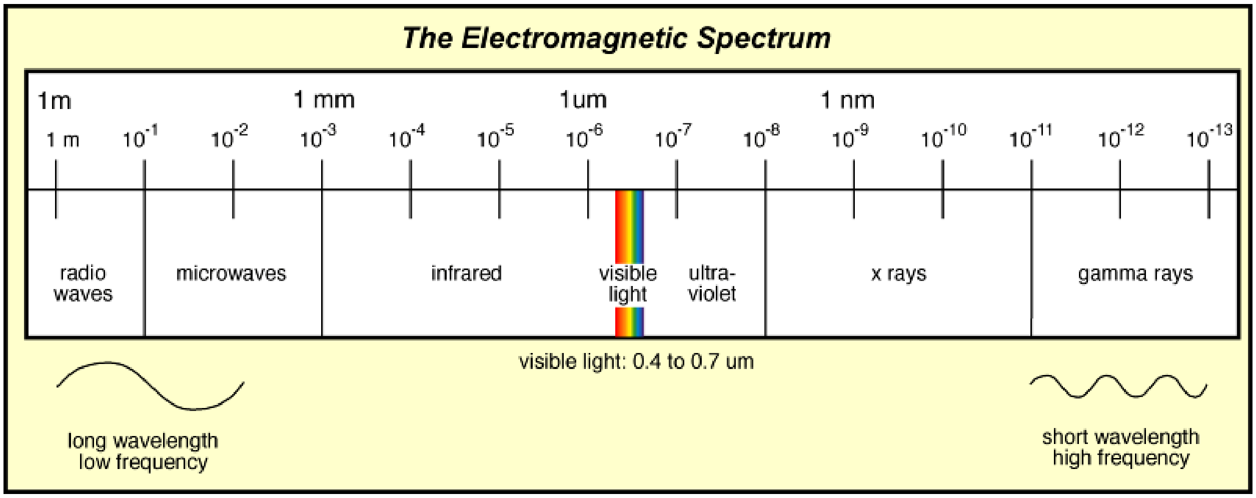


Fig. 1.4: The electromagnetic spectrum from `http://www.columbia.edu/~vjd1/electromag_spectrum.htm` gives the range of frequencies of electromagnetic radiation and their respective wavelengths. In this thesis, we essentially use the ultra-violet, visible and near-infrared wavelengths.

90's, White and Frenk (1991) studied the formation of galaxies by gas evolution and cooling to form stars following the empirical **Schmidt law** of Schmidt (1959) (extended later by Kennicutt 1998, to become the **Kennicutt-Schmidt law**). Then, Kauffmann et al. (1993) used a semi-analytic model of dark matter halo to study galaxy formation through hierarchical clustering. The year after, Cole et al. (1994) described `GALFORM`, a more detailed semi-analytic model of galaxy formation involving most of the processes listed above, further extended by Somerville and Primack (1999), and Cole et al. (2000) with their `Santa Cruz` model. Then, the `GalICS` model of Hatton et al. (2003) which combines N-body and semi-analytic model to study the hierarchical formation of galaxy. Later, Wang et al. (2008) studied the links between galaxy formation and the cosmological parameters. This work was followed by the semi-analytic model of Somerville et al. (2012) that is used to study the dust absorption and re-emission in the interstellar medium. After that, the `Santa Cruz` model (Somerville et al. 2015), the `eGalICS` model (Cousin et al. 2015) and the `GALFORM` model (Lacey et al. 2016) were updated and used to study in depth galaxy formation and evolution. More recently, the semi-analytic models `SHARK` (Lagos et al. 2018), `G.A.S.` (Cousin et al. 2019) and `GalICS 2.1` (Cattaneo et al. 2020) also study the star formation history and galaxy properties.

In this thesis, the physical processes driving galaxy evolution are not modeled, however we study the resulting empirical distribution of galaxies on the sky with deep surveys using a parameterized forward model with several realistic galaxy populations.

5. Main surveys

Galaxies can be observed through multiple techniques such as multiband photometry, long-slit spectroscopy*, fiber spectroscopy or integral field spectroscopy (Ilbert et al. 2006) using Integral Field Units (IFU)**. Galaxy surveys allow us to study galaxies from observations with the advantage to have large statistical samples containing simultaneously multiple galaxy populations, and homogeneous selecting criteria (angular region on the sky, apparent magnitude limit). Over the past decades, large galaxy surveys have been performed and have provided huge samples of panchromatic images containing objects of the past (distant) and the present (nearby) Universe from radio to γ-rays of the electromagnetic spectrum of Fig. 1.4

---

*The long-slit spectroscopy is a technique to observe celestial objects using a spectrograph in which the entrance aperture is an elongated, narrow slit.

**In an IFU the field of view is divided into many segments to obtain a comprehensive overview of the whole. IFUs are used in astronomy to study extended objects, such as nebulae, galaxies or a crowded cluster of stars or galaxies in one shot, using a technique known as integral field spectroscopy.

(where the wavelength λ at the top expressed in meters can also be written as the frequency $\nu = \frac{c}{\lambda}$ where $c = 299792458$ m s^{-1} is the speed of light).

Because the processes driving the evolution of a specific galaxy happen on time-scales of millions to billions of years and cannot be observed on a human time-scale, the properties of galaxies (such as number density, color, shape and size) can be studied with a same survey if deep enough, as large samples of galaxies can be seen at different epochs of their evolution. To make sure that galaxies probed in sufficiently large volumes of the Universe constitute what is called a "fair sample" (see Peebles 1980; Coles and Lucchin 1995; Weinberg 2008), some principles are applied:

- The **Copernican principle** which states that the Earth and more generally our galaxy, the Milky Way, do not occupy a special place in the Universe.
- The **cosmological principle** which assumes a homogeneous and isotropic spatial distribution of matter on large scales in the Universe.
- The **ergodicity principle** which states that the measured statistics (such as the distribution of galaxies) on a large volume is equivalent to the statistics of the whole Universe.

Their are two types of galaxy surveys: (i) **deep surveys** with long exposure times to probe very faint and/or high redshift galaxies – but the down side is that the covered area of the sky is generally smaller ; and (ii) **wide surveys** which cover large areas of the sky to study specific properties of galaxies with more statistics but losing the faintest galaxies. Depending on the scientific purpose, a compromise may be found between the two types of surveys. Moreover, both types of surveys are either or both

- **photometric**: images of sources where the light passes through specific wavelength filters. Thereby, data are available for as many sources as there are in the image, and the light from each source over a filter band is concentrated onto a small area of the detector with a signal-to-noise ratio depending on the apparent flux of the source. However, individual spectral features cannot be seen and the effective wavelength resolution is limited by the width of the filter bands. For instance, this method is particularly suitable to study the morphology, the brightness and the brightness profile of galaxies.
- **spectroscopic**: the light of sources is dispersed through a prism and recorded to obtain the intensity of light as a function of wavelength. The main advantage is that individual spectral features can be seen with excellent discrimination between wavelengths for the same object (for example in the VVDS (Le Fèvre et al. 2005) with one arcsecond slits, the resolution measured at 7500Å is 33Å and this can be compared to the large bandwidth of the Megacam i' filter of 1550 Å[*]). Disadvantages are that the light from the source is spread on the detector as a function of wavelength, therefore faint sources can be lost and long exposure times are generally necessary to obtain spectra with sufficient signal-to-noise for analyses. This method is particularly suitable to study the spectral energy distributions of galaxies, and to estimate their distances by measuring the redshifts of the spectra.

In addition to the area of the sky probed and the wavelength range used, a survey is also characterized by its **sensitivity** (*i.e.* the flux of the faintest detected source above some threshold) and its **completeness limits** (*i.e.* the flux limit at which the ratio of detected sources over the actual number of sources falls below some threshold). A galaxy survey is said to be "complete" when all galaxies brighter than some specified flux limit have been observed. However, it is hard to assess that a survey is actually complete and several methods have been developed. A first test proposed by Hubble (1926b) analyses the variation in galaxy number counts as a function of the adopted limit, but in practice it is difficult to decide whether deviations are indeed an effect of incompleteness, or is due to clustering and/or evolution of the number density of galaxies. Later, the V/V_{max} method developed by Schmidt (1968) makes use of the distance information and considers the ratio of the volume of a sphere of radius equal to the actual distance of observed galaxy, divided by the volume of a sphere of radius equal to the maximum distance at which the galaxy would be observable – but this method suffers from the same problem as the previous test. A third procedure used by Zwaan et al. (2004), Hoppmann et al. (2015), and Xi et al. (2021) makes use of synthetic galaxies and computes the fraction of recovered ones at the limits of the survey to estimate the completeness, but one has to make sure that the synthetic objects are similar enough to the observed galaxies.

---

[*] see `https://www.cfht.hawaii.edu/Instruments/Imaging/Megacam/specsinformation.html`

Using multiple wavelength filters of the same galaxy deep photometric survey allows us to probe processes that drives galaxy evolution such as: morphology, colors, spectral energy distributions. From them, other important quantities can be derived such as: galaxy distance, from which can be derived the properties of the stellar populations, as well as their age and star formation rate; galaxy dust and gas content, and stellar mass can also be inferred. Spectroscopy provides the most accurate measurement of distances, whereas multiband photometric surveys provide less accurate and sometimes degenerate photometric redshifts (Quadri and Williams 2010; Hildebrandt et al. 2012). Moreover, because of optical effects called seeing and pointed out by Hardy (1998) due to atmospheric turbulence, the resolution of ground-based telescope is limited and more or less blurred images – depending on the quality of the observing site – are obtained. This is the reason why space telescopes are used such as the Hubble Space Telescope (HST), and provide a better angular resolution than most of ground-based instruments.

Today, the number of available photometric and/or spectroscopic wide or deep surveys is large with a diversity of characteristics. In order to have a more complete review of these surveys, the reader is invited to read Djorgovski et al. (2013), Madau and Dickinson (2014).

5.1 Wide surveys

Here is a non-exhaustive list of some of the most popular wide-field imaging surveys sorted by wavelength regimes.

- The **Digitized Sky Survey** (DSS; available at `https://archive.stsci.edu/cgi-bin/dss_form`): scan of plates produced by the Space Telescope Science Institute (STScI) between 1983 and 2006 from the Palomar Observatory Sky Survey (POSS) and the European Southern Observatory/Science and Engineering Research Council (ESO/SERC) Southern Sky Atlas. This survey covers the whole sky in several visible wavelength bands and shows around one billion sources.
- The **Sloan Digital Sky Survey** (SDSS; York et al. 2000): both a photometric and spectroscopic survey started in 1998 of 14,500 deg^2 of the sky (*i.e.* approximately one third of the celestial sphere). The last data release of Alam et al. (2015) reveals around 4.7×10^8 objects and 1.8×10^6 spectra.
- The **Baryon Oscillation Spectroscopic Survey** (BOSS; Dawson et al. 2013): spectroscopic survey of approximately 10,000 deg^2 to measure the redshifts of 1.5 million Luminous Red Galaxies (LRGs) between $0 < z < 0.7$, in order to detect the baryon acoustic oscillation (BAO) signature and constrain the acceleration of the expansion rate of the Universe. The project was expanded with the **Extended Baryon Oscillation Spectroscopic Survey** (eBOSS; Abolfathi et al. 2018) to map the distribution of galaxies and quasars from when the Universe was between 3 and 8 billion years old.
- The **Two Micron All-Sky Survey** (2MASS; Jarrett et al. 2000): photometric survey conducted between 1997 and 2001 which covers the whole sky with the large photometric catalog of 1.5 million galaxies in the near-IR bands. Only a sub-sample of 44,000 galaxies has been released by Huchra et al. (2012) with spectroscopic data.
- The **UKIRT Infrared Deep Sky Survey** (UKIDSS; Lucas et al. 2008): photometric survey conducted on the United Kingdom Infrared Telescope on Mauna Kea in Hawaii between 2005 and 2012. It covers 7,500 deg^2 of the Northern sky.
- The **Infrared Astronomical Satellite** (IRAS; Neugebauer et al. 1984): the first-ever space telescope to perform a survey of the whole sky at mid and far IR wavelengths. Launched on 25 January 1983, for ten months, it covered over 250,000 sources.
- The **NRAO VLA Sky Survey** (NVSS; Condon et al. 1998): survey of the Northern Hemisphere performed by the Very Large Array (VLA) of the National Radio Astronomy Observatory (NRAO) of 32,800 deg^2 at radio wavelengths (1.4 GHz) between 1993 and 1996.
- The **Faint Images of the Radio Sky at Twenty-cm** (FIRST; Becker et al. 1995): 10,000 deg^2 survey of the North and South Galactic Caps performed by the NRAO VLA at radio wavelengths (1.4 GHz) between 1993 and 2004. There are approximately 9×10^5 sources and around 35% of them have resolved structures.

- The **Parkes-MIT-NRAO** (PMN; Wright et al. 1994): combination of several surveys of a total of 16,400 deg^2 performed by the Parkes 64m radio telescope at radio wavelengths (4.85 GHz) in 1990.

- The **Galaxy Evolution Explorer** (GALEX; Martin et al. 2005): satellite launched for a NASA mission led by the California Institute of Technology to study star formation in galaxies and its evolution. The survey was performed in the near and far UV between 2003 and 2012, and ultimately observed around 75% of the sky.

- The **ROSAT All Sky Survey** (Voges et al. 1999): satellite mission performed between 1990 and 1999 by ROSAT (short for "Röntgensatellit" because in German X-rays are called "Röntgenstrahlen") at X-ray wavelengths of the full sky.

- The **Fermi Large Area Telescope All Sky Survey** (Fermi LAT; Atwood et al. 2009): satellite survey performed by the Fermi observatory in 2008 of the entire sky at gamma-ray wavelengths.

5.2 Deep surveys

A complement to the wide surveys of the previous section are various deep surveys covering much smaller areas but reaching fainter magnitude limits due to longer exposure times and/or larger telescopes. These surveys aim at studying galaxy formation and evolution; the large depth is required in order to better measure the properties of distant (*i.e.* high redshift) galaxies and compare them to those of nearby or intermediate redshift galaxies.

- The **Hubble Deep Field** (HDF; Williams et al. 1996): survey image obtained by stacking 342 exposures of the same small 2.6 arcmin^2 area of the sky taken with the Hubble Space Telescope for ten consecutive days in 1995 in four filters ranging from near-UV to near-IR. A deeper **Hubble Ultra Deep Field** (UDF; Beckwith et al. 2006) was also observed in 2003/2004 and later in 2009 and covered a total of 11 arcmin^2. Finally, the **Hubble eXtreme Deep Field** (HXDF; Illingworth et al. 2013), a gigantic stack of images corresponding to approximately 23 days of exposure time of 4.6 arcmin^2 contains around 5,500 galaxies with the oldest ones ever observed.

- The **Groth Survey Strip** (Simard et al. 2002): The Groth Survey Strip is a 127 arcmin^2 field observed with the HST and multiple ground-based observatories in only two bands (visible V and near-IR I). The sky area was enlarged in 2004/2005 to the **Extended Groth Strip** to 700 arcmin^2 (Nandra et al. 2005). Finally, it was also extended to multiple wavelengths with the **All-Wavelength Extended Groth Strip International Survey** (AEGIS; Davis et al. 2007)

- The **Chandra Deep Field South** (CDF-S; Giacconi et al. 2001): survey obtained by the Chandra X-ray space telescope from 11 individual day-long exposures in 1999/2000 of a 400 arcmin^2 area of the sky in the X-ray wavelengths.

- The **Great Observatories Origins Deep Survey** (GOODS; Dickinson et al. 2003; Giavalisco et al. 2004): survey covering an area of 324 arcmin^2 in the sky and obtained with a combination of space and ground-based telescopes in the optical, near-IR and X-ray wavelengths (Spitzer, Hubble, Chandra, etc). This survey aims at studying distant galaxies to better understand the history of star formation and galaxy evolution.

- The **Subaru Deep Field** (Kashikawa et al. 2004): survey observed at the 8.2m Subaru telescope with a sky area of around 900 arcmin^2 in the visible and near-IR wavelengths. It was mainly made to observe the Lyman-break of high redshift galaxies and a total of 200,000 galaxies were observed.

- The **Cosmological Evolution Survey** (COSMOS; Scoville et al. 2007): survey initiated with HST observations and extended to other space and ground-based telescopes over a sky area of 2 deg^2 observed in 2004/2005. It was made to study around 2 million galaxies from X-ray to radio wavelengths and was followed by extensive spectroscopic studies.

- The **Cosmic Assembly Near-IR Deep Extragalactic Legacy Survey** (CANDELS; Koekemoer et al. 2011): survey observed by HST of over 250,000 galaxies with a moderate field made with only two orbits of 0.2 deg^2 and a deep field made with twelve orbits of 0.04 deg^2. It aims to study the galactic evolution of high redshift galaxies with $1.5 < z < 8$ at optical and near-IR wavelengths.

- The **Deep Lens Survey** (DLS; Wittman et al. 2002): survey of five fields for a total sky area of approximately 20 arcmin^2 from far-UV to near-IR wavelengths.

- The **NOAO Deep Wide Field Survey** (NDWFS; Jannuzi and Dey 1999): survey of two fields for a total sky area of approximately 18 arcmin^2 from far-UV to near-IR wavelengths, followed by other wavelengths observations.

- The **Canada-France-Hawaï Telescope Legacy Survey** (CFHTLS; available at `http://www.cfht.hawaii.edu/Science/CFHLS`): community-oriented survey of four deep fields for a total sky area of approximately 4 deg^2 (“Deep fields") and a wider and not as deep survey of four fields with a total sky area of 171 deg^2 (“Wide fields”) obtained using the CFHT 3.6 m telescope equipped with the MegaCam imager for visible wavelengths (Hudelot et al. 2012) and the WIRCam instrument for near-IR wavelengths (Bielby et al. 2012) during 450 nights. This survey and more specifically the D1 deep field are further described Sect. 5.3.

- The **Advanced, Large, Homogeneous Area, Medium-Band Redshift Astronomical survey** (ALHAMBRA; Moles et al. 2008): this survey provides a deep photometric data set over 20 contiguous, equal-width (∼300 Å), non-overlapping, medium-band optical filters (3500-9700 Å) plus three standard broad-band near-IR filters (J, H, Ks) over eight different regions of the northern sky and covering a total area of 2.38 deg^2.

- The **Deep Extragalactic Evolution Probe Redshift Surveys** (DEEP and DEEP2; Davis et al. 2003; Newman et al. 2013): community-oriented spectroscopic survey of 3.5 deg^2 using the twin 10 m Keck telescopes in Mauna Kea (Hawaii) to measure the spectra of 50,000 galaxies at redshifts $z \approx 1$.

- The **VIRMOS-VLT Deep Survey** (VVDS; Le Fèvre et al. 2005): spectroscopic redshift survey of 4 fields with a total sky area of 3.5 deg^2 obtained at the 8.2 m VLT to measure the redshifts of 150,000 galaxies in the $0 < z < 5$ redshift range.

5.3 The Canada-France-Hawaï Telescope Legacy Survey

In this thesis, we want to study the evolution of the two main galaxy populations: elliptical and spiral types, which requires a very large statistical galaxy sample of hundred of thousands of objects. Because these two populations can have different color-related properties, we also require multiple wavelength filters in the UV-visible-IR to study them efficiently. Therefore, from the above non-exhaustive lists of galaxy surveys, we need to pick a survey fulfilling these considerations, that is a very deep survey extending from UV to IR; the CFHTLS deep fields fully meets these requirements.

The CFHTLS Deep fields are calibrated stacks of images in both the optical u', g', r', i', z' bands of the MegaCam instrument (Hudelot et al. 2012) and in the near-IR J, H, Ks bands of the WIRCam instrument (Bielby et al. 2012) whose responses curves are shown in Fig. 1.5. These images are of excellent quality, with exquisite (and well sampled) seeing for ground-based observations, ranging from 0.8 to 0.6 from the u' to the Ks band (see Table 3), and high depth: they reach 80% completeness limits at AB magnitudes of $u' = 26.3$, $g' = 26.0$, $r' = 25.6$, $i' = 25.4$ and $z' = 23.9$, thanks to the high sensitivity of both instruments and the 3.6m aperture of the CFHT, and the very long exposure times from 9 hours for the near-IR bands to 72 hours for the visible bands. They cover four fields of ~ 1 square-degree which allows for statistical studies of different galaxy populations with redshift $0 < z < 1$ (for instance there are $\approx 10\,000$ elliptical and $\approx 80\,000$ spiral galaxies in one Deep field to $m_i < 25.4$).

The CFHTLS Deep field are named D1 to D4 according to their location in the sky (see Fig. 1.6) and is a $19,354 \times 19,354$ pixel image covering 1 deg^2 in total and containing several hundred thousand galaxies. The CFHTLS deep fields have been used in a variety of studies such as:

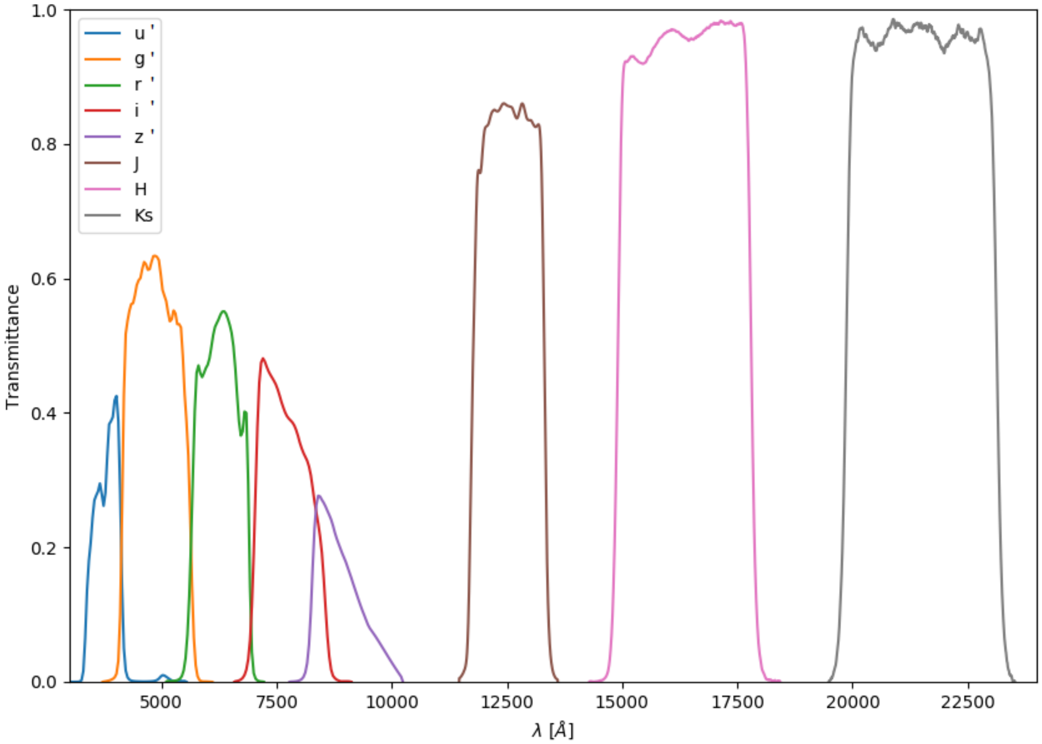


Fig. 1.5: Optical u', g', r', i', z' bands of the MegaCam instrument (Hudelot et al. 2012) and near-IR J, H, Ks bands of the WIRCam instrument (Bielby et al. 2012) used to obtain the CFHTLS D1 deep field images. This plot displays the normalized filter response curves (taking into account the mirror+optics+CCD response) on the y-axis as a function of the wavelength on the x-axis in Angstrom (1 Å $= 10^{-10}$ m).

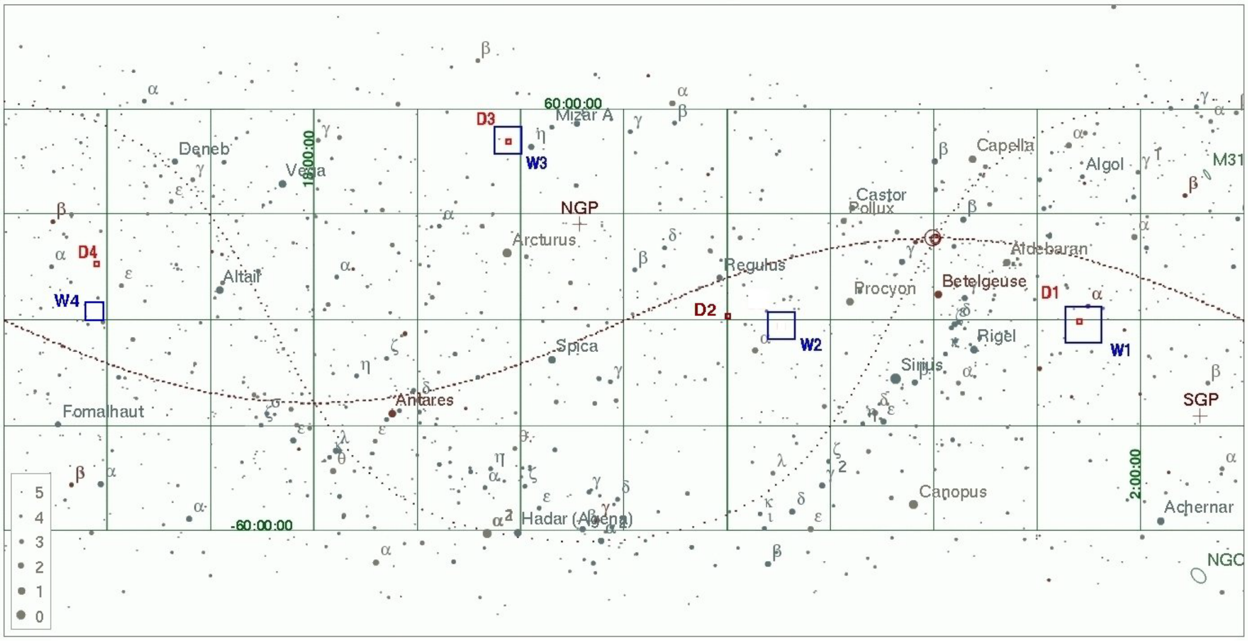


Fig. 1.6: Image from Hudelot et al. (2012). Location of the CFHTLS Deep (denoted D) and Wide (denoted W) fields on the sky.

- The distance measurement of high redshift type Ia supernovae (Astier et al. 2006) from their multi-color light-curves. This was followed by spectroscopic observations at the VLT, Gemini and Keck telescopes to confirm the nature of the supernovae and to measure their redshifts, allowing them to constrain the Ω_M, Ω_Λ and w parameters of the ΛCDM model.
- The photometric redshift measurements of approximately 500,000 objects (Ilbert et al. 2006) based on the VIMOS VLT Deep Survey (VVDS) spectroscopic redshifts of 3241 objects as a calibration for the `Le Phare` photometric redshift code published later by Arnouts and Ilbert (2011).

In this thesis, we use the D1 deep field in the eight photometric bands u', g', r', i', z', J, H, Ks (*i.e.* from far-UV to near-IR wavelengths). Unfortunately, this 1 deg^2 (*i.e.* $19,354 \times 19,354$ pixels) photometric deep field is too large to be processed in its entirety with the technique developed in this thesis, due to computation time constraints. Instead, multiple insets of the D1 deep field of size 1024×1024 pixels in the eight photometric bands are used simultaneously in this study.

Chapter 2

From survey images to catalogs: selection effects

> *Do not go gentle into that good night,*
> *Old age should burn and rave at close of day;*
> *Rage, rage against the dying of the light.*
>
> *Though wise men at their end know dark is right,*
> *Because their words had forked no lightning they*
> *Do not go gentle into that good night. (1-6)*
>
> `Dylan Thomas`

Abstract

Using the deep survey images introduced in the last chapter, the study of galaxy evolution requires the measurement of luminosity functions. Luminosity functions are statistical tools close to the observations used to model the number density and the ratio of each population of galaxies in the Universe and their evolution with redshift. In order to correctly express the luminosity functions, magnitudes are introduced to quantify the brightness of the galaxies. The classic approach aims at extracting catalogs of apparent magnitudes from these survey images and at converting them to absolute magnitudes using estimations of the distance of each galaxy. However, the classic catalog-extraction method inevitably introduces multiple selection effects that lead to highly-correlated biases in the derived luminosity functions. In the next chapters, we will introduce a catalog-free original method which avoids these biases.

Contents

1. Apparent and absolute magnitudes

A **magnitude** is a unitless measure of the brightness of an object (a galaxy, a star, an asteroid, etc). During the last 2000 years, the apparent brightness of stars was thought to be linked to their physical size and several magnitude systems used this assumption to create a size-classification of stars (Miles 2007). However, thanks to parallax calculations, astronomers realized in the mid-nineteenth century that most stars appear as point sources of light because they are far from Earth, and that magnitude is not only linked to the size of the star, but also to its distance. Thus in 1856 Norman Pogson proposed a logarithmic scale of magnitudes in which each increasing step of one magnitude corresponds to a decrease of the brightness of the object by a factor $\sqrt[5]{100} \approx 2.5$, therefore five magnitude steps corresponded precisely to a factor of 100 in brightness (Hoskin 1999). This system was called the **Vega-magnitude system** because the star Vega was used as the fundamental reference star, with an apparent magnitude defined to be zero, whatever the measurement technique or the wavelength filter (this explains why objects brighter than Vega, such as Sirius have a negative Vega-magnitude of -1.5). However, Oke and Gunn (1983) found that Vega varies in brightness making it unsuitable for an absolute reference, so the reference system was modified to the **AB-magnitude system** which does not depend on any particular star's brightness (Vega AB-magnitude in the V (visual) band is 0.03 rather than 0). There are two different definitions of magnitudes commonly used:

- The **apparent magnitude** (m): brightness of an object collected by the observer which depends on the intrinsic brightness emitted by the object, the distance and the possible intergalactic dust extinction between the object and the observer that would dim the brightness of the object. In the AB-magnitude system, considering a filter F with a roughly continuous range of frequencies $[\nu_{\min}, \nu_{\max}]$ (see Fig. 1.4), the apparent magnitude of an object in this filter F is defined as:

$$m_{AB,F} = -2.5 \log_{10} \left(\frac{\int_{\nu_{\min}}^{\nu_{\max}} f(\nu) e(\nu) d\nu}{\int_{\nu_{\min}}^{\nu_{\max}} e(\nu) d\nu} \right) - m_{AB,0} \tag{2.1}$$

  where $f(\lambda)$ is the spectral flux density in W Hz^{-1} m^{-2} (also expressed in erg s^{-1} Hz^{-1} cm^{-2} or in Jy[*]) of the object, $e(\lambda)$ is the system response function (which depends on the quantum efficiency and the transmission curve of the detector, the optics of the telescope and the atmospheric conditions of the observation) and $m_{AB,0} = 48.6$ is the zero point (when inverted this leads to the famous numerical value "3631 Jy": $f_{\nu,0} = 10^{-m_{AB,0}/2.5} \approx 3631 \times 10^{-23}$ erg s^{-1} Hz^{-1} cm^{-2})

- The **absolute magnitude** (M): intrinsic brightness emitted by an object which corresponds to the apparent brightness that the object would have if it were seen from a specific distance of 10 pc. By hypothetically placing all objects at a standard reference distance from the observer, their absolute magnitude can be directly compared among each other and the more luminous an object, the smaller the numerical value of its absolute magnitude. In the AB-magnitude system, the absolute magnitude of an object in a specific filter F is defined as:

$$M_{AB,F} = m_{AB,F} - 5 \log_{10}(d) + 5 - K_{\text{corr}} \tag{2.2}$$

  where $m_{AB,F}$ is the apparent magnitude of the object in filter F, d is the distance between the object and the observer in pc, K_{corr} is the K-correction term due to the fact that we do not observe the light from high-redshift objects in the same wavelengths that it was emitted because of the expansion of the Universe (see Sect. 2.2 for a detailed explanation).

The light of astronomical objects is collected on photometric multiband deep field images (see Chap. 1 Sect. 5.) and catalogs of apparent magnitude for these objects are extracted thanks to extraction packages. The extraction process is explained in Sect. 3. of this chapter along with its inherent limitations. Moreover, using Eq. 2.2, catalogs of absolute magnitudes of objects can be derived directly from the corresponding catalogs of apparent magnitudes using the distance between the objects and the observer. But the determination of the distance is a challenge because at high-redshift, galaxies are faint and their distance cannot be

---

[*] symbol of the Jansky: 1 Jy = 10^{-26} W Hz^{-1} m^{-2} = 10^{-23} erg s^{-1} Hz^{-1} cm^{-2}

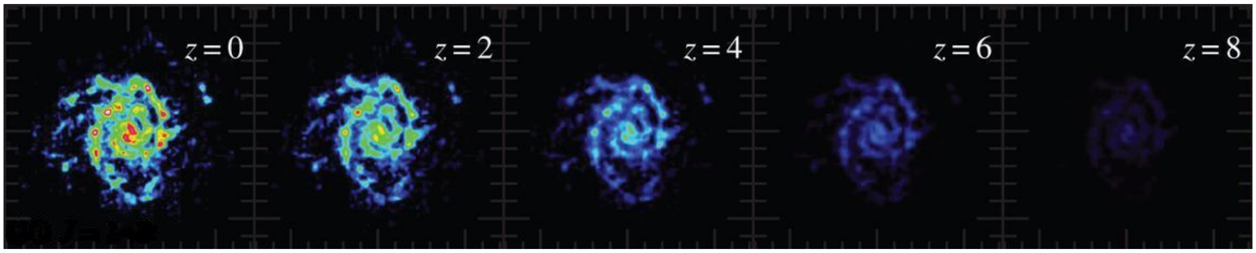


Fig. 2.1: Simulation of the effect of cosmological dimming on the surface brightness of a galaxy. The colors show the intensity scale of the light, more precisely the red regions correspond to very bright star-forming regions. This highlights the fact that the galaxy is progressively obscured because of its redshift/distance and only the very bright regions remain at high redshift.

obtained by spectroscopic redshift which would require prohibitive exposure times. To estimate their distance, a photometric redshift determination is used based only on the apparent magnitudes of these sources in several specific filters. This procedure has a number of biases described by MacDonald and Bernstein (2010) because the precision of photometric redshifts relies on SED templates of different galaxy populations that are assumed to be representative, and are fitted to the apparent fluxes of the galaxies in several photometric bands. Therefore, the number of very faint and very bright galaxies are generally over-estimated as shown by Sheth (2007).

2. Selection effects

2.1 Cosmological dimming

Because of the expansion of the Universe, the bolometric surface brightness (*i.e.* the surface brightness integrated over all wavelengths) of a long-distance source is decreased, an effect called **cosmological dimming**, as explained by Tolman (1930, 1934):

$$I_{\mathrm{obs}} = \frac{I_{\mathrm{e}}}{(1+z)^4} \tag{2.3}$$

where I_{e} (resp. I_{obs}) is the emitted (resp. observed) surface brightness of the source integrated over all wavelengths and z is the redshift of the source.

Consequently, only the very bright galaxies at high-redshift can be seen on a magnitude-limited survey and many sources can be lost which gives a biased view of the early Universe. Moreover, as shown by Calvi et al. (2014), multiple star-forming regions with high surface brightness of the same galaxy can be detected as distinct sources which can have an impact on the galaxy counts. Fig. 2.1 is a simulation that illustrates this effect.

2.2 K-correction

The expansion of the Universe also impacts the wavelengths at which we observe the light coming from high-redshift sources. A photon emitted from a source at redshift z with frequency ν_{e} is redshifted and finally observed with frequency ν_{obs} following the Doppler effect:

$$\nu_{\mathrm{obs}} = \frac{\nu_{\mathrm{e}}}{1+z} \tag{2.4}$$

A classic example of such an effect can be seen in the observation of the very distant galaxies with the HST (Williams et al. 1996) where they observe in UV and visible wavelengths the light that was emitted in far-UV wavelengths. Moreover, as seen in Eq. 2.2, when considering a filter F and converting the apparent magnitude to absolute magnitude, one needs to apply the following K**-correction** (Hogg et al. 2002):

$$K_{\mathrm{corr}} = -2.5 \log_{10} \left[(1+z) \frac{L\left((1+z)\nu_{\mathrm{obs}}\right)}{L\left(\nu_{\mathrm{obs}}\right)} \right] \tag{2.5}$$

where L is the observed luminosity of the source integrated over the wavelengths of the filter F. In practice, the K-correction can be estimated via two distinct approaches:

- By direct measurement of the source spectrum within a spectroscopic survey. But, in most surveys only a small range of the electromagnetic spectrum of the source is available using a set of broadband filters. Therefore, the missing spectrum information can lead to poor estimates of the K-correction, thus the derived absolute magnitude could be biased.
- By fitting the missing spectrum information on a set of empirical or modeled SED templates. But, most of these SED templates are built from observed galaxies of the local Universe which are assumed to be also representative of the SED of distant galaxies. The K-correction is then limited by the accuracy of this fit and the good match between the SED template and the SED of the observed galaxy.

A poor measurement of the K-correction can lead to biases, for instance in the number counts of galaxies in different populations as shown by Gabrielli et al. (2005).

2.3 Dust extinction and inclination

Inside each galaxy, the interstellar dust dims the light coming from the stars mostly in the UV and optical wavelengths depending on the **inclination** angle at which the galaxy is observed. Most of the dust might be contained in the disk of galaxies and almost no dust is seen in the bulge as shown by Graham and Worley (2008). However, the light coming from the bulge region can be extincted by the dust contained in the disk before reaching the observer and as a result, the color of the galaxy can be altered or in the worst case the galaxy itself can be undetectable. This effect is also correlated with the inclination of the galaxy: the attenuation is maximal and the galaxy is redder when seen edge-on (see Fig. 2.2) which has been confirmed in observations as well (Driver et al. 2007). The different populations of galaxies are affected differently by this effect and Nagayama et al. (2004) have shown that spiral galaxies are in general more extincted by dust than elliptical or lenticular galaxies.

Because of this correlated effect between dust and inclination of galaxies, dusty faint galaxies can be missed in magnitude-limited surveys (Devour and Bell 2016) and the galaxy counts biased. Moreover, apparent magnitude and surface brightness profile of galaxies can also be affected (de Jong 1996; Masters et al. 2010) by dust extinction. Bad estimates of galaxy disk scale lengths is another effect of dust extinction because inner regions of disks generally contain more dust than outer regions (Popescu and Tuffs 2005). Finally, Yip et al. (2011) have shown that photometric redshift can be overestimated because of dust-inclination causing mismatches between the SED templates and the SED of the observed galaxy.

The dust content of galaxies and its various effects on the spatial distribution of light and on the colors have been highly investigated and modeled over the last twenty years (Valentijn 1990; de Vaucouleurs et al. 1991; Giovanelli et al. 1995; Tully et al. 1998; Masters et al. 2003; Tuffs et al. 2004; Bailin and Harris 2008; Unterborn and Ryden 2008) but parameters are poorly constrained and models are not in good agreement. Our model of dust extinction for galaxies is inspired from Tuffs et al. (2004) and is explained with more details in Chap. 3 Sect. 3..

2.4 Malmquist bias

In magnitude-limited surveys, only objects brighter than the detectable limit can be seen and faint sources could be undetected. As a consequence, the calculation of the average distance of the galaxies on the same survey is underestimated which directly leads to an overestimate of the average absolute magnitude of the survey, see Fig. 2.3. This effect is known as the **Malmquist bias** (Malmquist 1920, 1922, 1925), first discovered for samples of stars but it equally applies for samples of galaxies. It has to be dealt with to properly measure the average luminosity in bins close to the limiting magnitude.

A first attempt of removing the Malmquist bias is to limit the sample only to the magnitude-range for which all sources can be detected. This subsample is then free from this bias, however many usable data are thrown away and only relatively nearby sources can be analyzed. Another solution to remove this bias was proposed originally by Malmquist (1922) assuming that the luminosity function (see Sect. 4.) is gaussian and that all galaxies have the same spectral type without any dust extinction. This is a very idealized situation which is not met in any photometric survey.

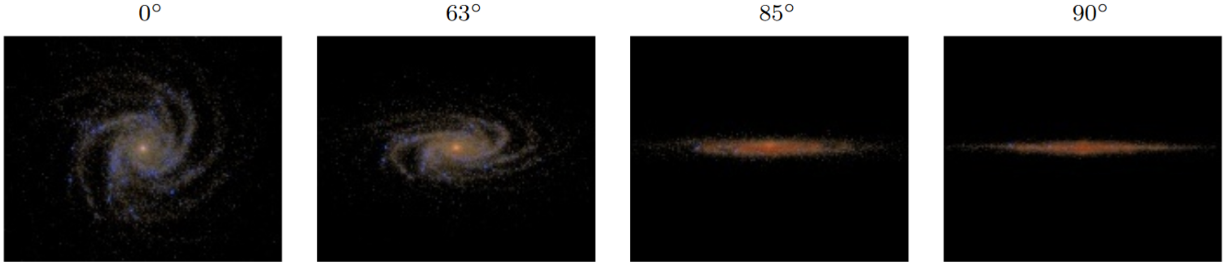


Fig. 2.2: Image from Jonsson et al. (2009). Simulated (Sunrise radiative transfer code; Jonsson 2006) color image (using the u', r', z' filters) of a Sbc galaxy extincted by dust for four different inclinations. The dust reddening effects become more and more important as the inclination increases.

Later, Schmidt (1968) proposed a new method, called $V/V^{\max}$, where each source-magnitude can be weighted by the maximum volume over which the object could have been seen. This method is used in many studies however assuming a constant spatial density distribution of galaxies over all distances which particularly affects the number counts of faint galaxies (Blanton et al. 2005b). Maximum likelihood estimation of magnitude distributions can also be applied (Sandage et al. 1979) accounting for fluctuations in the galaxy number density with redshift. In the literature, many other methods are applied in order to take into account the Malmquist bias.

2.5 Eddington bias

Back in 1913, Eddington described how the rate of bright stars can be overestimated when a fraction of fainter stars is wrongly associated with the brighter class due to the error on the flux measurement (Eddington 1913). The same problem occurs when dealing with galaxies of various magnitudes. This enhances the number of galaxies observed just at the detection threshold and causes overestimates of the galaxy fluxes. Under the assumption that measurement errors are Gaussianly distributed with a constant width in magnitude, Eddington (1940) showed that it is possible to correct the number counts of the observed values. However, as it is the case in most galaxy surveys, measurement errors increase for faint galaxies and no obvious analytical correction can be applied (van der Burg et al. 2010).

3. Extraction process and biases

A large fraction of the scientific work is not done on the deep field images, but rather on catalogs of objects extracted from these images and of course the reliability of the source extraction process is of major importance. Many computer programs have been developed to obtain galaxy catalogs such as `FOCAS` (Jarvis and Tyson 1981), `APM` (Maddox et al. 1990) or `SExtractor` (Bertin and Arnouts 1996), `SourceXtractor++` (Bertin et al. 2020). The analysis and extraction process of an image is generally done in six steps: background estimation, source detection, source separation, source cleaning, photometry and star/galaxy separation.

3.1 Background estimation and confusion noise

Each pixel value of the image is the sum of a residual signal noise (due for instance to the heat of the acquisition system or the readout-noise of a charge-coupled device (CCD) camera) and light coming from the objects we are interested in. Having a good estimate of the background level of the image is essential to measure the magnitude of the faintest objects. Several background estimators are described in the literature (Bijaoui 1980; Almoznino et al. 1993; Bertin and Arnouts 1996) which are more or less biased, noisy and require varying amount of computing time. These techniques of background estimation generally assume that the background does not vary too much over the same image, which is no longer true for crowded (such as star cluster) or bright (such as mid and far-IR surveys) background deep fields. This effect is known as

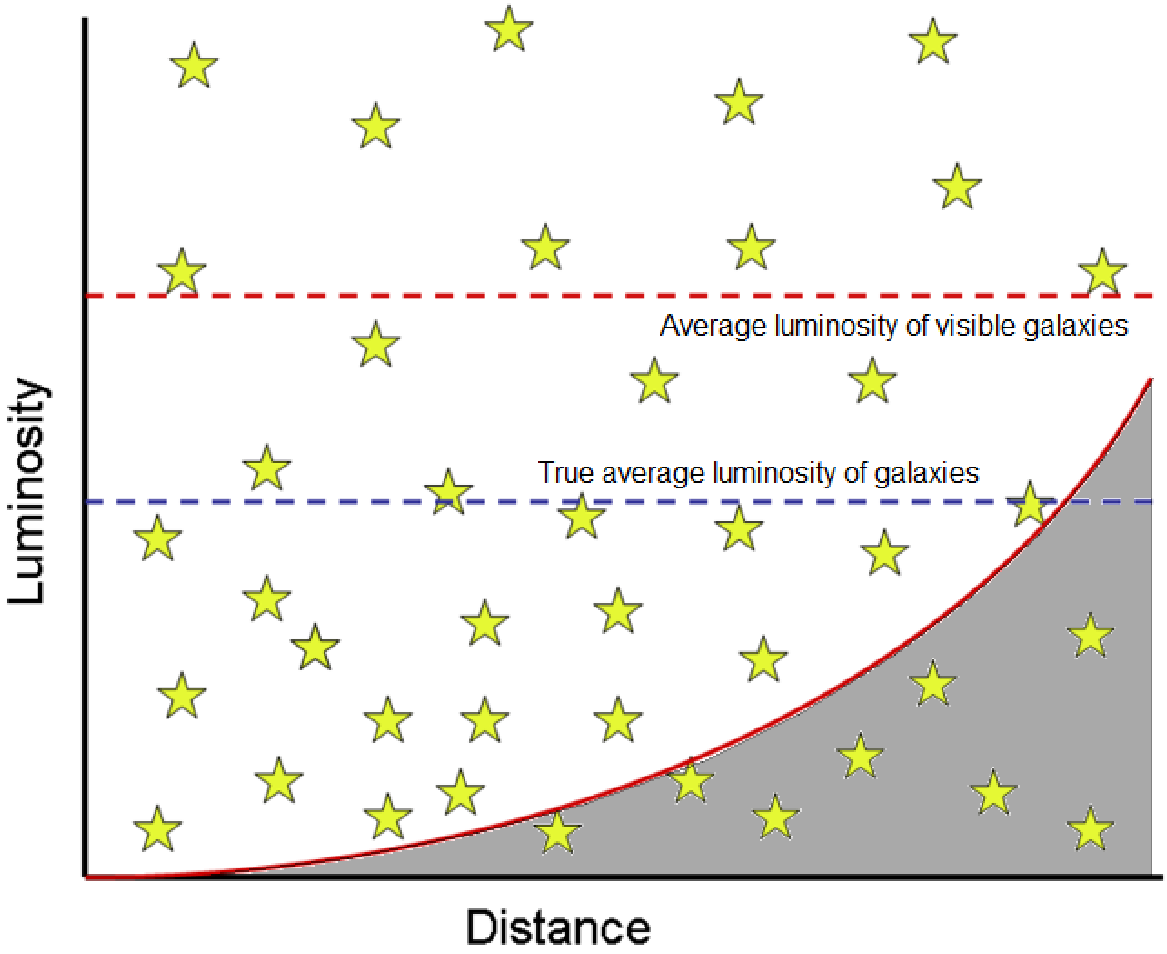


Fig. 2.3: Schematic representation of the effect of the Malmquist bias when calculating the average luminosity of a sample of galaxies. Because galaxies in the gray part are too distant and cannot be detected, the average luminosity of visible galaxies is overestimated compared to the true average luminosity.

confusion noise (Condon 1974) and it can have consequences such as increasing the positional uncertainty of sources as shown by Hogg (2001), boosting the brightness of certain sources or detecting very faint sources that are clustered together as shown by Negrello et al. (2004). This effect is widely described in Dole et al. (2003) and is mainly related to the beam full width at half maximum (beam FWHM), that is the minimal circle diameter at which the brightness of a point source is at least half the maximum.

3.2 Source detection, separation and cleaning

The detection of sources with no particular shape assumption is generally done with two techniques: (*i*) peak finding algorithm which looks for local maxima in brightness and is particularly suited for star detection but not for faint galaxies detection (Yee 1991); and (*ii*) filter convolution algorithm which applies a convolution filter on the original image with user-defined size and values depending on the type of objects to be detected.

With both above techniques, it is necessary to inspect if the detected sources are either single sources or multiple neighboring sources, otherwise the derived catalogs could contain extended galaxies that are split into smaller detected sources or multiple small galaxies that are gathered into a large detected source. This is generally done by applying multiple isophotes* and user-defined thresholds to decide if a detection should be split or not (Bertin and Arnouts 1996).

Depending on the previous user-defined thresholds and on the background estimation, spurious detected sources could contaminate the resulting catalogs and need to be removed. This is a very hard task to perform and this is generally done in studying the mean surface brightness ratio between the detected pixels and the neighboring pixels. More recently, a new technique by Paillassa and Bertin (2019) makes use of a convolutional neural network to detect and separate the sources instead of a multi user-thresholds method.

3.3 Photometry and star/galaxy separation

Photometry is the process of measuring the flux (or intensity of light) of sources on images obtained historically with a photometer and since the 80's, with a CCD camera. The measured value can then be used to compute the apparent magnitude of objects. Photometers generally use specific passband filters across the UV, visible, and IR wavelengths of the electromagnetic spectrum (see Fig. 1.4) such as the classic Ultraviolet Blue Visible (UBV or Johnson) system or near-IR JHK system. Each set of filters with known light transmission capabilities allows the study of particular properties of galaxies.

Because each image is a record of the light from multiple objects at once, various forms of photometric extraction can be performed on the data. The goal is to measure the total flux of galaxies at all signal-to-noise levels, which is a challenge. The observed light signal from each object typically covers several pixels according to the point spread function** (PSF) of the system and the degree of spreading of the light of each source is a measure of the quality of this system. Fixed aperture photometry (Mighell 1999; Laher et al. 2012) is the simplest method to extract the flux of a source: it consists in summing the flux within a circle centered on the object and subtracting the nearby average sky flux. But it misses an increasing fraction of flux with the size of extended objects. Adaptive aperture photometry is an improvement towards estimating total magnitudes, and takes into account the objects isophotal radii (see for example `SExtractor`, Bertin and Arnouts (1996)). However, aperture photometry cannot be applied in case of crowded sources and other techniques such as PSF fitting (Stetson 1987) is required to determine the individual flux of the overlapping sources. Profile modeling is then the best approach (see `SourceXtractor++` (Bertin et al. 2020)); as it allows fluxes to account for the "wings" of the objects hidden below the sky background, which may constitute a significant fraction of the light emitted by faint galaxies. Galaxies often appear as extended objects on images and measuring the spatial distribution of the light within a galaxy is called surface photometry (Palei 1968), which allows for elliptical geometries.

The quality of the extracted apparent magnitudes highly depends on the local background estimation (see Sect. 3.1), the quality of the PSF used, the resolution and the quantity of overlapping sources of the image. Galaxies generally look more extended than stars on images (Colless et al. 1990) and a good dissociation between stars and galaxies is crucial, as faint galaxy catalogs are often contaminated by stars

---

[*] An isophote is a circle or an ellipse connecting separated pixels of an image with roughly the same brightness.

[**] The point spread function (PSF) describes the response of an imaging system to a point source such as a star or a galaxy.

that are misidentified (Pearson et al. 2014). In the literature, many classifiers have been used such as the r^{-2} moment of Kron (1980), magnitude-isophotal area (Reid and Gilmore 1982), magnitude-peak intensity (Jones et al. 1991), magnitude-surface brightness (Harmon and Mamon 1993) or neural network (Bertin and Arnouts 1996). Each classifier has specific properties but they all face the same limitation: to deal with merged/overlapping or neighboring sources.

4. Luminosity function

The **luminosity function (LF)** is an essential tool to describe the properties of galaxy populations and their evolution. The luminosity function gives the number of galaxies per unit comoving volume of the Universe and per magnitude interval, as a function of absolute magnitude. This quasi-probability distribution function can be fit by a common **Schechter profile** (Schechter 1976),

$$\Phi(M) = 0.4\log(10)\phi^* \left[10^{0.4(M^*-M)}\right]^{1+\alpha} \exp\left[-10^{0.4(M^*-M)}\right] \tag{2.6}$$

where ϕ^* is the **normalization density** (in Mpc^{-3} mag^{-1}), α the **faint-end slope**, and M^* a **characteristic absolute magnitude**. Other profiles can also be used to fit specific properties of the data such as the double Schechter profile of Driver and Phillipps (1996) for the luminosity function of galaxy clusters, or Blanton et al. (2005a) and López-Sanjuan et al. (2017) for the upturn in the luminosity function of faint elliptical galaxies. In this thesis, we use the Schechter profile to parameterize both the elliptical and spiral populations of galaxies. For a specific population of galaxies and a fixed redshift, the luminosity function gives both the number of galaxies and the ratio of bright to faint galaxies for this population at this redshift. The α parameter indicates how the number density of faint galaxies varies with magnitude (see Figs. 2.4 and 2.5, left parts of graphs): e.g., for $\alpha < -1$ (as in Figs. 2.4), the number density of faint galaxies increases with magnitude. The normalization density parameter, ϕ^* shifts the entire curve on the y-axis while the characteristic absolute magnitude parameter, M^* determines the change of behavior between the power-law at faint magnitudes and the exponential decrease at bright magnitudes.

Today, the global luminosity function of all types of galaxies combined at redshift $z \approx 0.1$ is relatively well-constrained with the use of the spectroscopic surveys 2dFGRS (Norberg et al. 2002) and SDSS (Blanton et al. 2003) as shown in Fig. 2.4. However, luminosity functions are generally different from one population of galaxies to another as shown by Zucca et al. (2006), and the categorization of galaxies within a given survey is difficult to achieve accurately, and the measured luminosity functions are sensitive to how the galaxy types are split. Moreover, the total absolute magnitude of the sources is required to properly estimate the luminosity function, therefore, a good knowledge of the distance/redshift is necessary and the selection effects described in Sect. 2. apply. Moreover, the total light of the sources is generally estimated using the extraction process of Sect. 3. which also adds biases to the derived absolute magnitudes. Because high-redshift galaxies are generally faint objects unsuitable for spectroscopic redshift determination (which would require excessively high exposure times), it is difficult to properly derive the luminosity function evolution with redshift (Lilly et al. 1995; Gabasch et al. 2004).

In order to describe the redshift evolution of the luminosity function, two other parameters are introduced, ϕ^*_{evol} and M^*_{evol}, defined as

$$\phi^* = \phi^*_0(1+z)^{\phi^*_{\text{evol}}} \tag{2.7}$$

$$M^* = M^*_0 + M^*_{\text{evol}}\log(1+z) \tag{2.8}$$

where z is the redshift at which the luminosity function is measured. Table 1 gives a non-exhaustive list of recent results (López-Sanjuan et al. 2017; Beare et al. 2015; Drory et al. 2009; Salimbeni et al. 2008; Faber et al. 2007; Brown et al. 2007) where the authors attempted to separate quiescent/star-forming (or elliptical/spiral) populations and Fig. 2.5 shows the corresponding luminosity functions that they derived (at redshift $z \approx 0.5$). Their results are in general in good agreement for the spiral population but larger differences appear for the elliptical population, especially at the bright and faint ends. Moreover, as the redshift increases, their samples of higher-redshift galaxies contain less and less objects and their separation methods are less accurate, thus increasing the uncertainties on the evolution of the luminosity function of elliptical/spiral populations.

Table 1: Properties of the studies whose luminosity function are plotted in Fig. 2.5.

Survey	area	Magnitude limit	median redshift	filters	separation elliptic/spiral
Alhambra[1]	2.38 deg^2	$m_I < 24.5$ (∼450k sources)	0.74 (phot.)	20 filters of width 300 Å, from 3500 to 9700 Å + J, H, Ks	model SED fitting + prior probability
NDWFS[2] SDWFS[2] NEWFIRM[2] LBT[2] Subaru[2] AGES[2]	8.26 deg^2	$m_I < 24$ (∼410k galaxies)	∼0.3 (97% phot. 3% spec.)	u, B_W, R, I, z, y, J, H, Ks + 3.6, 4.5, 5.8, 8.0 μm (IR)	model SED fitting + rest-frame $U-B$ color
COSMOS[3]	1.73 deg^2	$m_I < 25$ (∼138k sources)	∼0.7 (phot.)	30 broad and medium bands	$NUV-R$ vs $R-J$ color-color
GOODS[4]	∼0.5 deg^2	$m_z < 26$ (∼14k sources)	∼0.3 (90% phot. 10% spec.	U, B, V, I, z, J, H, Ks + 3.6, 4.5, 5.8, 8.0 μm (IR)	model SED fitting + rest-frame $U-V$ color
DEEP2[5] COMBO-17[5]	1.95 deg^2	$m_R < 24$ (∼70.5k galaxies)	∼0.5 (phot.)	U, B, V, R, I	rest-frame $U-B$ color
NDWFS[6] IRAC[6]	∼7 deg^2	$m_I < 23$ (∼40k elliptical)	∼0.45 (90% phot. 10% spec.)	B_W, R, I, K + 3.6 and 4.5 μm (IR)	rest-frame $U-V$ color

Notes:

"phot." stands for photometric redshifts and "spec." for spectroscopic redshifts.

[1] See López-Sanjuan et al. (2017).

[2] See Beare et al. (2015).

[3] See Drory et al. (2009).

[4] See Salimbeni et al. (2008).

[5] See Faber et al. (2007).

[6] See Brown et al. (2007).

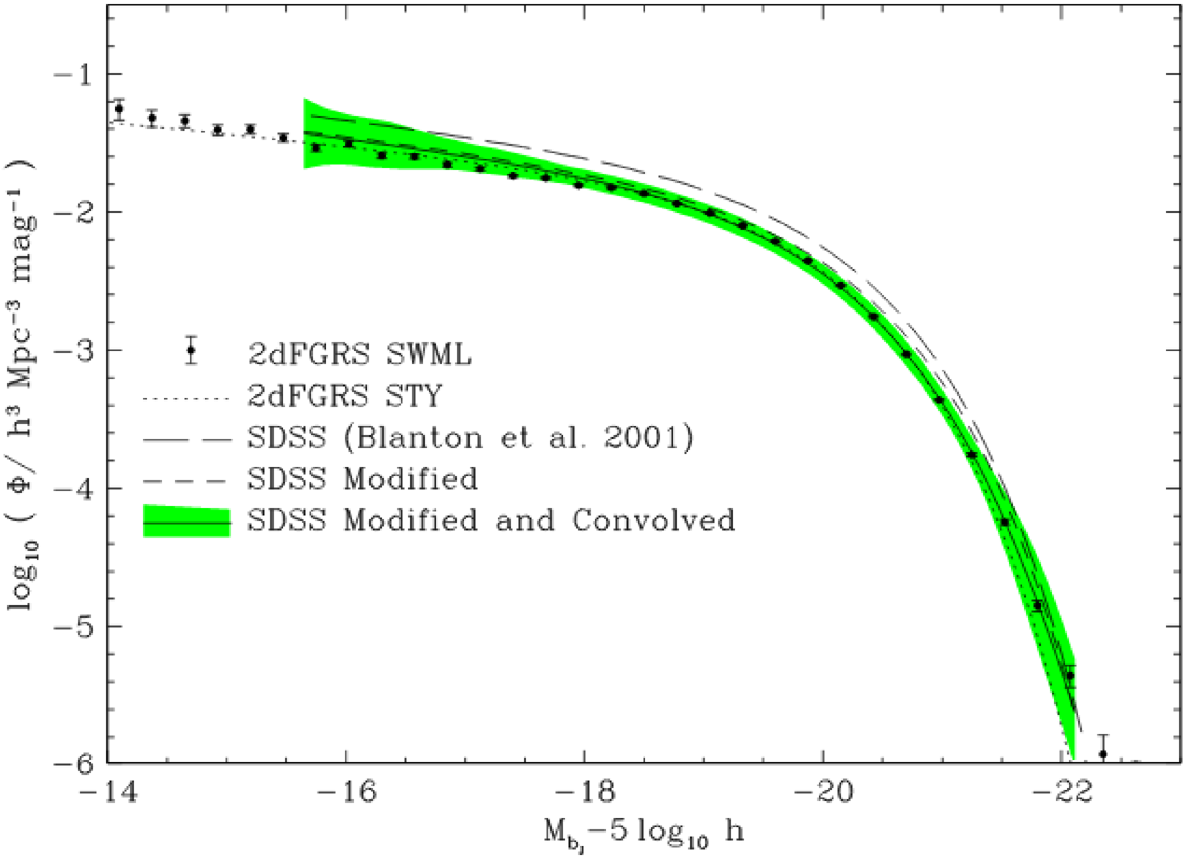


Fig. 2.4: Plot from Norberg et al. (2002). Global luminosity function of all galaxy types combined in the local universe ($z \approx 0.1$). This compares the 2dFGRS luminosity function of Norberg et al. (2002) with the SDSS luminosity function estimates of Blanton et al. (2003) in the Johnson B-band absolute magnitude (results given for $H_0 = 70$ km s^{-1} Mpc^{-1}).

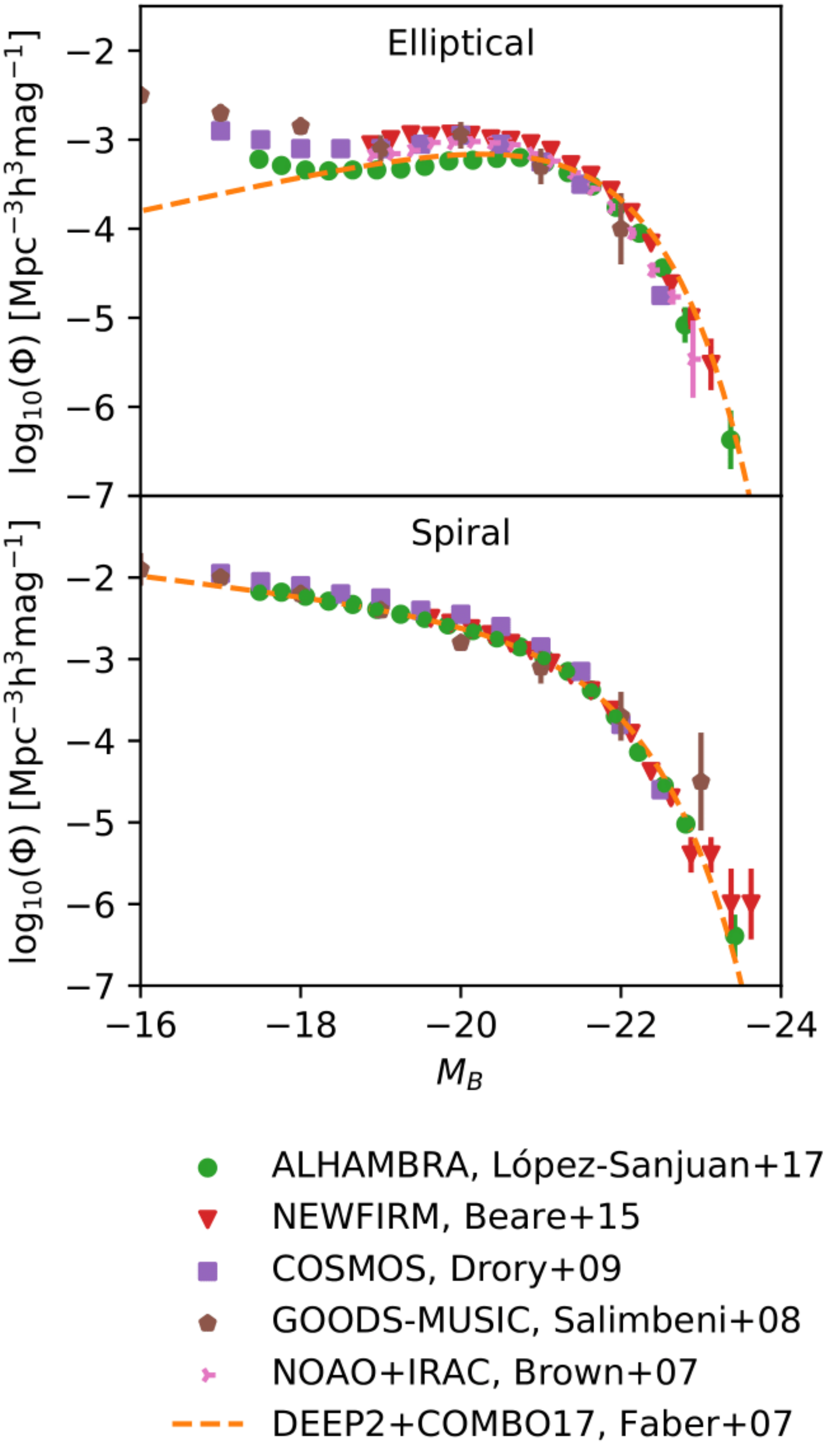


Fig. 2.5: Luminosity function at $z = 0.5$ derived from the results of López-Sanjuan et al. (2017) in green, Brown et al. (2007) in pink, Beare et al. (2015) in red, Salimbeni et al. (2008) in brown, Drory et al. (2009) in purple, and Faber et al. (2007) in orange (results given for $H_0 = 70$ km s^{-1} Mpc^{-1}).

5. Applying the extraction process on a simulated deep field

In this section, we intend to illustrate some biases usually induced in the fluxes and luminosity functions of the various galaxy types by the use of photometric catalogs obtained from source extraction and photometric redshifts. To do so, we apply the forward model that will be described in Chap. 3 in order to obtain a synthetic but realistic multiband deep field and we compare the magnitudes and the luminosity functions of galaxies (and stars) generated by the forward model with the magnitudes and the luminosity functions obtained with the use of the extraction program `SExtractor` (Bertin and Arnouts 1996), as well as the photometric redshift estimates from the `ZPEG` program (Le Borgne and Rocca-Volmerange 2002) and the V/V_{max} method (Schmidt 1968) to derive the luminosity functions.

A large 19354×19354 pixel deep field is generated by the forward model in the 8 photometric bands (*i.e.* u', g', r', i', z', J, H, Ks) with stars and three populations of galaxies: elliptical, spiral and irregular. The elliptical, spiral and irregular galaxies are generated using luminosity functions with the following parameters in the reference g' band:

	Elliptical	Spiral	Irregular
$\log_{10}(\phi_0^*)$	-2.09	-2.04	-2.92
M_0^*	-19.68	-19.71	-19.15
α	-0.53	-1.29	-1.5
ϕ_{evol}^*	-1.37	-0.03	1.9
M_{evol}^*	-1.15	-1.49	-1.5
B/T	1	0.2	0
SED bulge	E	E	$\emptyset$
SED disk	$\emptyset$	Scd	Irr

(2.9)

The above parameters of the luminosity functions of the elliptical and spiral populations are taken from López-Sanjuan et al. (2017, Table 2) (see Chap. 3 Sect. 1.7 for a full description). Those of the irregular population are chosen as a combination of the results of Marzke et al. (1994, 1998), Bromley et al. (1998), Nakamura et al. (2003), and Liu et al. (2008). With these parameters, 2 341 008 galaxies are generated and rendered on realistic images using the `SkyMaker` program (Bertin 2009). Then, the extraction program `SExtractor` is used with the "dual-image" option and the reference band i' (it is in this band that we detect the highest number of objects with the luminosity functions of Eq. (2.9)) and 509 738 sources are detected. With this configuration, 487 detected objects are spurious sources in the i' band. Fig. 2.6 gives in the left panel the number counts of detected sources by `SExtractor` for each population of galaxies and stars and for the sum of these four types. The right panel gives the same number counts for the undetected sources : the fraction of undetected sources (mostly spirals and irregulars) becomes significant at and fainter than $m_i \sim 25$).

Fig. 2.7 shows, in the left panels, the distributions of differences between the true input magnitudes given by the forward model and the "MAG_AUTO" magnitudes (*i.e.* an adaptative aperture magnitude) extracted by `SExtractor` in the i' band for all detected sources (*i.e.* elliptical, spiral, irregular populations and stars) in five different magnitude intervals from bright to faint. The right panels of Fig. 2.7 show the redshift distributions for non detected sources by `SExtractor` compared to those for all sources in each i' magnitude interval. In both columns of Fig. 2.7, the fractions of galaxies over the total number over all types and in each plotted magnitude interval are shown. These graphs show that most bright sources with $m_i < 22$ are detected by `SExtractor`, but several hundreds of very bright stars have under-estimates of extracted magnitudes probably because of saturation issues. For galaxies, the discrepancies between the extracted and the input magnitudes increase with the i' magnitude which is confirmed by the increasing values of the standard deviation for all populations. In the bins $24 < m_i < 25.4$ and $m_i > 25.4$ (*i.e.* around the 80% completeness limit) the discrepancy can be of several magnitudes for significant fractions of spiral and irregular galaxies. Note also that at fainter magnitudes, the `SExtractor` galaxy fluxes are predominantly overestimated over the input values (positive values of the input minus extracted difference dominate), which is counter intuitive: one would expect faint galaxies to become fainter than they are intrinsically, because of the cosmological dimming (a larger fraction of their angular size disappears below the sky background at fainter magnitudes). The observed effect probably results from underestimation of the sky background for these faint sources, a critical aspect of flux estimation from images, as it may not take into account the increased local background due to nearby brighter stars and galaxies. We tested whether this may be partly overcome by using a smaller mesh size of 64 pixels (instead of 256 pixels) in the calculation of the background with `SExtractor`: this has the inconvenient that bright galaxies become fragmented into separate sources ; in any case, this change only adds $\sim 10^4$ galaxies in the extraction catalogue, which is negligible compared to the missing 80% sources at these faint magnitudes.

We emphasize that the right panel of Fig. 2.7 for $m_i > 25.4$ shows the plain (all detected sources) and dotted (undetected sources) curves of the spiral and irregular redshift distributions are close to each other, corresponding to the fact that around 80% of the sources are not detected by `SExtractor` (confirmed in the right panel of Fig. 2.8), probably because these sources are too faint due to the cosmological dimming (see Sect. 2.): they are expected to fall below the 1.5σ detection threshold adopted in the `SExtractor` analysis (see Sect. 3.). Right panels of Fig. 2.7 also show that for all intervals at $m_i < 24$, the spiral population is dominant over the irregular population; moreover, in the magnitude bin $m_i > 25.4$ only $\sim 10\%$ of the irregular galaxies with $0 < z < 1.5$ are detected while $\sim 50\%$ of the spiral galaxies are detected. These are the reason why we only consider the dominant populations of elliptical and spiral galaxies in the applications of Chap. 6 Sects. 1. and 2. Fig. 2.8 gives a full overview of the number (left panel) and the ratio (right panel) of detected and non-detected sources for each population. This figure confirms the 80% completeness limit of 25.4 in the i' band for the D1 deep field.

The `ZPEG` program (Le Borgne and Rocca-Volmerange 2002) is then used to obtain the photometric redshifts from the magnitudes extracted in the 8 bands by `SExtractor`, using the SED templates of Coleman et al. (1980). The derived photometric redshifts are compared to the input hence true galaxy redshifts used by the forward model in Fig. 2.9. The contours of this figure in red, blue, green and black give the number of detected sources per bin above different thresholds. One can see that a large fraction of galaxies have photometric redshift estimates close to their initial values, however several thousands of galaxies have biased estimates of photometric redshifts at $z < 1.5$, that is within the redshift range of a non-negligible fraction of the various galaxy types. In the study of Dahlen et al. (2013), they compared the photometric redshifts obtained by 11 different codes on the same survey (*i.e.* CANDELS) and they have shown that there is no particular code, nor set of template SEDs, that results in significantly improved photometric redshifts compared to others.

The photometric redshifts and extracted magnitudes of the sources are finally used with a $V/V_{\rm max}$ method in order to derive the luminosity functions of the elliptical, spiral and irregular populations of galaxies. This is shown on Fig. 2.10 for 5 bins of redshift between 0 and 1 (dots and best fit dashed curves) and compared to the true input luminosity functions (plain curves) of these populations in the g' band. The fitted luminosity functions are obtained using the `Scipy` function "curve_fit"* which fits a Schechter profile to the $V/V_{\rm max}$ points and error bars. Moreover, Table 2 compares the parameters and the corresponding 1σ errors of the best fits to the $V/V_{\rm max}$ luminosity functions for the three galaxy populations to the true input parameters, in the redshift bins $0.4 < z < 0.6$ and $0.6 < z < 0.8$ (corresponding to the peaks of the redshift distributions for $20 < i' < 25.4$, see Fig. 2.7), as well as the parameter differences between input and measure in terms of σ. This shows that, in the redshift bin $0.4 < z < 0.6$, the $V/V_{\rm max}$ method tends to underestimate the ϕ^* density parameter by a factor of 10 (resp. 100) for the elliptical and spiral (resp. irregular) populations. The underestimated ϕ^* is counterbalanced with a decrease of the M^* parameter by ~ 1 (resp. ~ 2.6) and a decrease of α by 0.3 (resp. 0.5) for the elliptical and spiral (resp. irregular) populations; in all of the table, the offsets have values from 2.3 to 29 times the quoted errors obtained when fitting the $V/V_{\rm max}$ luminosity functions. One can clearly see an offset in the shapes of the luminosity functions with a higher number of faint objects than expected and fainter characteristic magnitudes in these bins. This conclusion is also true in the redshift range $0.6 < z < 0.8$. Fig. 2.10 also shows that in the redshift bins $0.0 < z < 0.2$ and $0.2 < z < 0.4$, the fraction of galaxies considered in the $V/V_{\rm max}$ method is too small to reliably estimate the luminosity functions (*i.e.* large error bars).

This application shows the limits and the biases that one may face when applying the classic approach of extracting catalogs of magnitudes and obtaining the photometric redshifts from these catalogs to derive the luminosity functions of each population of galaxies. It is worth noticing that in this example, the galaxy types initially generated by the forward model are used and there is no discrimination processes (such as

---

*`https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.curve_fit.html`

Table 2: Comparison between the input values of the parameters of the g'-band luminosity functions and the best fit and 1σ errors obtained from the fit of the V/V_{max} luminosity functions in the redshift bins $0.4 < z < 0.6$ and $0.6 < z < 0.8$ for the three populations of galaxies.

		$\log_{10}(\phi^*)$	M^*	α
Elliptical $0.4 < z < 0.6$	Input	-3.10	-20.92	-0.53
	Best fit ($\pm\sigma$)	-3.91 ± 0.05	-22.04 ± 0.13	-0.80 ± 0.04
	Offset	16.2σ	8.6σ	6.8σ
Spiral $0.4 < z < 0.6$	Input	-2.81	-21.09	-1.29
	Best fit ($\pm\sigma$)	-4.04 ± 0.07	-22.01 ± 0.16	-1.59 ± 0.02
	Offset	17.6σ	5.8σ	15σ
Irregular $0.4 < z < 0.6$	Input	-3.35	-20.53	-1.5
	Best fit ($\pm\sigma$)	-5.42 ± 0.50	-23.17 ± 1.14	-2.01 ± 0.02
	Offset	4.1σ	2.3σ	25.5σ
Elliptical $0.6 < z < 0.8$	Input	-3.17	-21.06	-0.53
	Best fit ($\pm\sigma$)	-3.47 ± 0.03	-21.87 ± 0.06	-0.64 ± 0.03
	Offset	10σ	13.5σ	3.7σ
Spiral $0.6 < z < 0.8$	Input	-2.81	-21.28	-1.29
	Best fit ($\pm\sigma$)	-3.24 ± 0.02	-22.43 ± 0.04	-1.26 ± 0.01
	Offset	21.5σ	28.8σ	3σ
Irregular $0.6 < z < 0.8$	Input	-3.25	-20.72	-1.5
	Best fit ($\pm\sigma$)	-4.11 ± 0.07	-21.49 ± 0.16	-1.72 ± 0.01
	Offset	12.3σ	4.8σ	22σ

color thresholds) to determine in which population each galaxy lies, which would add to the biases observed in the luminosity functions of Fig. 2.10. A similar analysis was performed in Carassou et al. (2017) and they also found that the luminosity function of elliptical galaxies can be biased because of extraction and photometric redshift estimates.

In the next chapters, we introduce a novel exploratory method which does not require the extraction of catalog of sources with neither magnitude nor redshift estimates. It is all pixels of the 8 bands deep field images that are used to extract information about the parameters of the luminosity functions. This allows the forward model to simultaneously generate elliptical and spiral populations of galaxies that are affected by the same observational and instrumental biases as those of the observed images and to infer the best parameters of the luminosity functions directly from images.

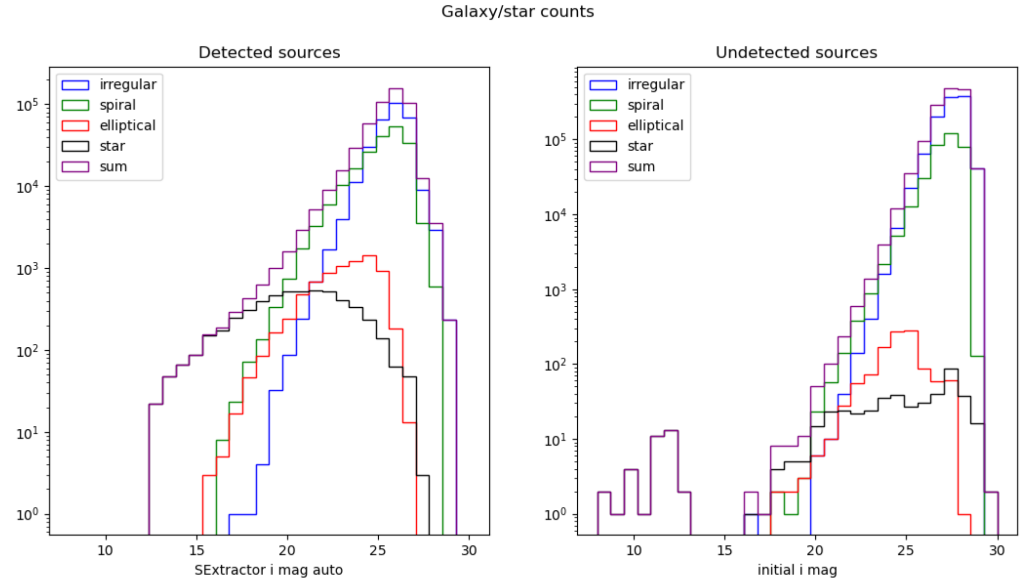
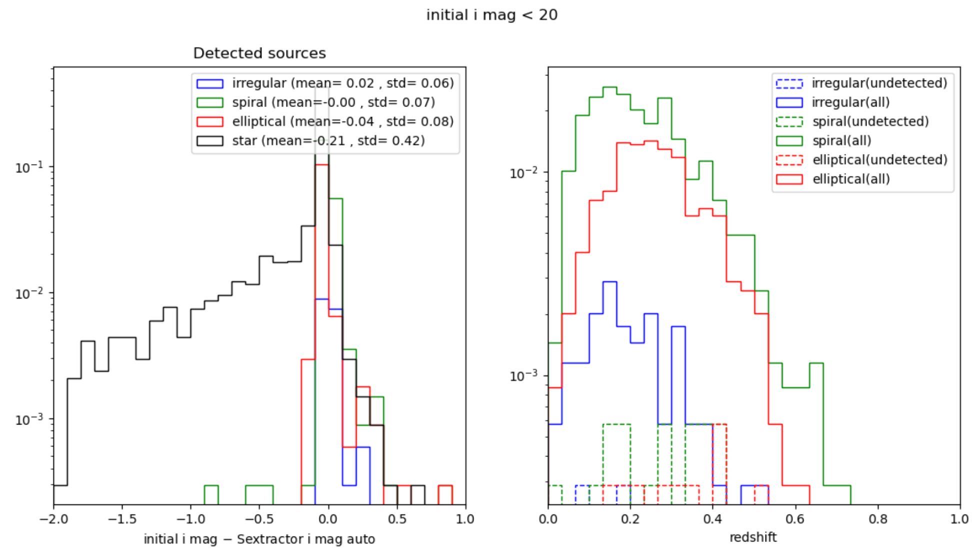
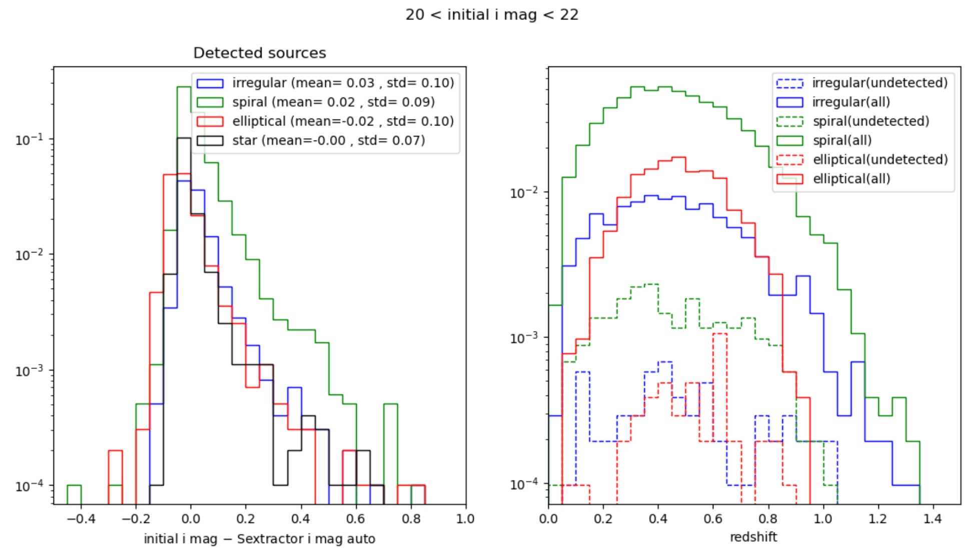
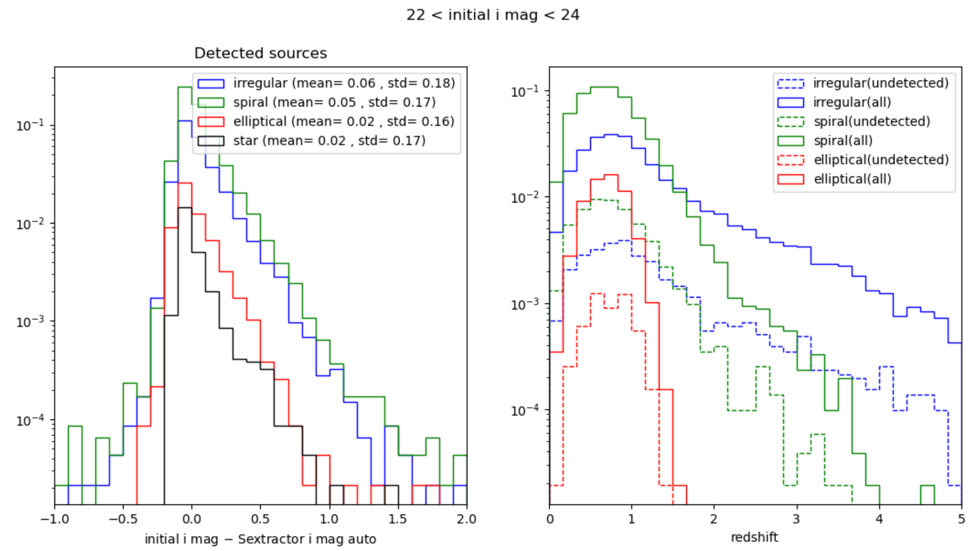


Fig. 2.6: *Left:* Galaxy and star counts for the detected sources by `SExtractor` in the i' band with the elliptical in red, the spiral in green, the irregular in blue, the stars in black and the sum of the four in purple. *Right:* Same as in the left panel, restricted to the undetected sources by `SExtractor`


initial i mag < 20
Detected sources
irregular (mean= 0.02 , std= 0.06)
spiral (mean=-0.00 , std= 0.07)
elliptical (mean=-0.04 , std= 0.08)
star (mean=-0.21 , std= 0.42)
10⁻¹
10⁻²
10⁻³
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
initial i mag − Sextractor i mag auto
irregular(undetected)
irregular(all)
spiral(undetected)
spiral(all)
elliptical(undetected)
elliptical(all)
10⁻²
10⁻³
0.0
0.2
0.4
0.6
0.8
1.0
redshift


20 < initial i mag < 22
Detected sources
irregular (mean= 0.03 , std= 0.10)
spiral (mean= 0.02 , std= 0.09)
elliptical (mean=-0.02 , std= 0.10)
star (mean=-0.00 , std= 0.07)
10⁻¹
10⁻²
10⁻³
10⁻⁴
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
initial i mag − Sextractor i mag auto
irregular(undetected)
irregular(all)
spiral(undetected)
spiral(all)
elliptical(undetected)
elliptical(all)
10⁻²
10⁻³
10⁻⁴
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
redshift


22 < initial i mag < 24
Detected sources
irregular (mean= 0.06 , std= 0.18)
spiral (mean= 0.05 , std= 0.17)
elliptical (mean= 0.02 , std= 0.16)
star (mean= 0.02 , std= 0.17)
10⁻¹
10⁻²
10⁻³
10⁻⁴
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
initial i mag − Sextractor i mag auto
irregular(undetected)
irregular(all)
spiral(undetected)
spiral(all)
elliptical(undetected)
elliptical(all)
10⁻¹
10⁻²
10⁻³
10⁻⁴
0
1
2
3
4
5
redshift

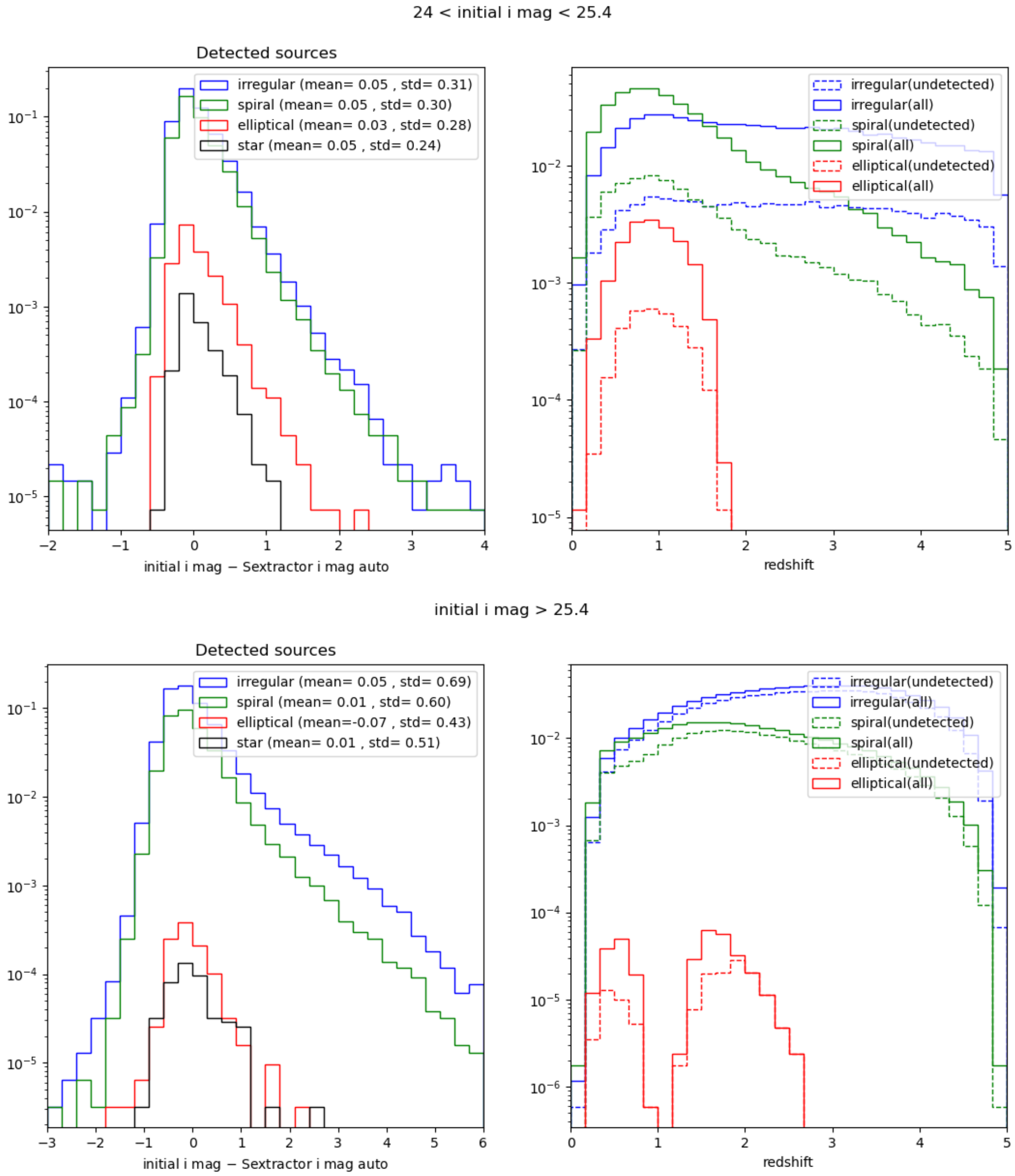


Fig. 2.7: *Left:* Histogram of the difference between the initial magnitude in the i' band generated by the forward model and the extracted magnitude by `SExtractor` for each population of galaxies (*i.e.* elliptical in red, spiral in green and irregular in blue) and stars (in black) for 5 intervals of i' magnitude from bright (top) to faint (bottom). The histograms are normalized over all detected sources by `SExtractor`. The extracted magnitudes are relatively close to the generated magnitudes for $M_i < 24$ but very bright stars with $M_i < 20$ have under-estimated magnitudes. For each population, the mean and the standard deviation are given in the legend of each panel.
Right: Histogram of the redshift of the sources that have not been detected by the `SExtractor` program (dashed) compared to the histogram of the redshift of all sources (plain) for 5 intervals of i' magnitude from bright (top) to faint (bottom). The histograms are normalized over all sources generated by the forward model.

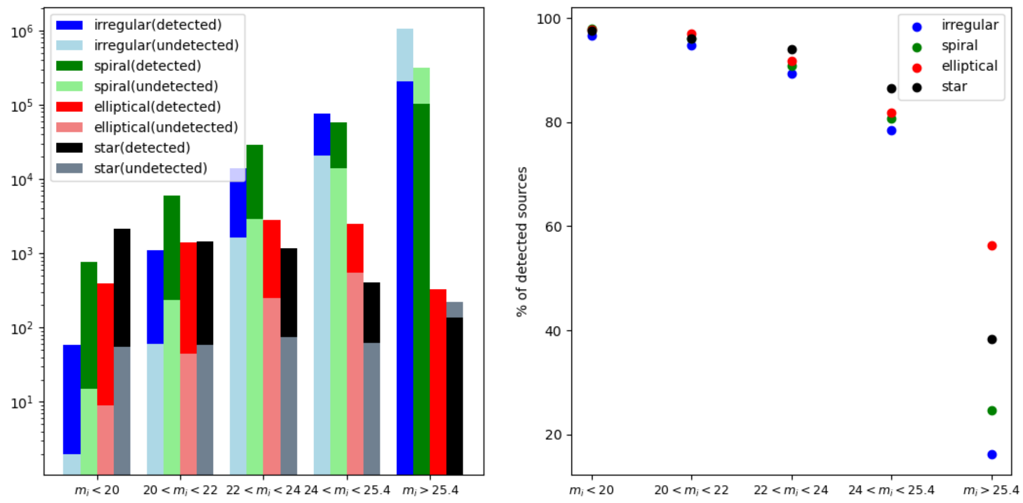


Fig. 2.8: *Left:* Number of detected (dark colors) and undetected (light colors) sources by `SExtractor` for each i' magnitude interval and for each GALAXY population (*i.e.* elliptical in red, spiral in green, irregular in blue and stars in black). Note that the histograms are superimposed with no transparency in their colors (hence the color of the smallest value takes over in the lower part of each bin).
Right: Ratio of detected sources over all sources for each i' magnitude interval and for each population. This confirms the 80% completeness limit in the i' band for the CFHTLS D1 deep field. For fainter objects the detection ratio drops to 55% for the ellipticals, 40% for the stars, 25% for the spirals and only 10% for the irregulars.

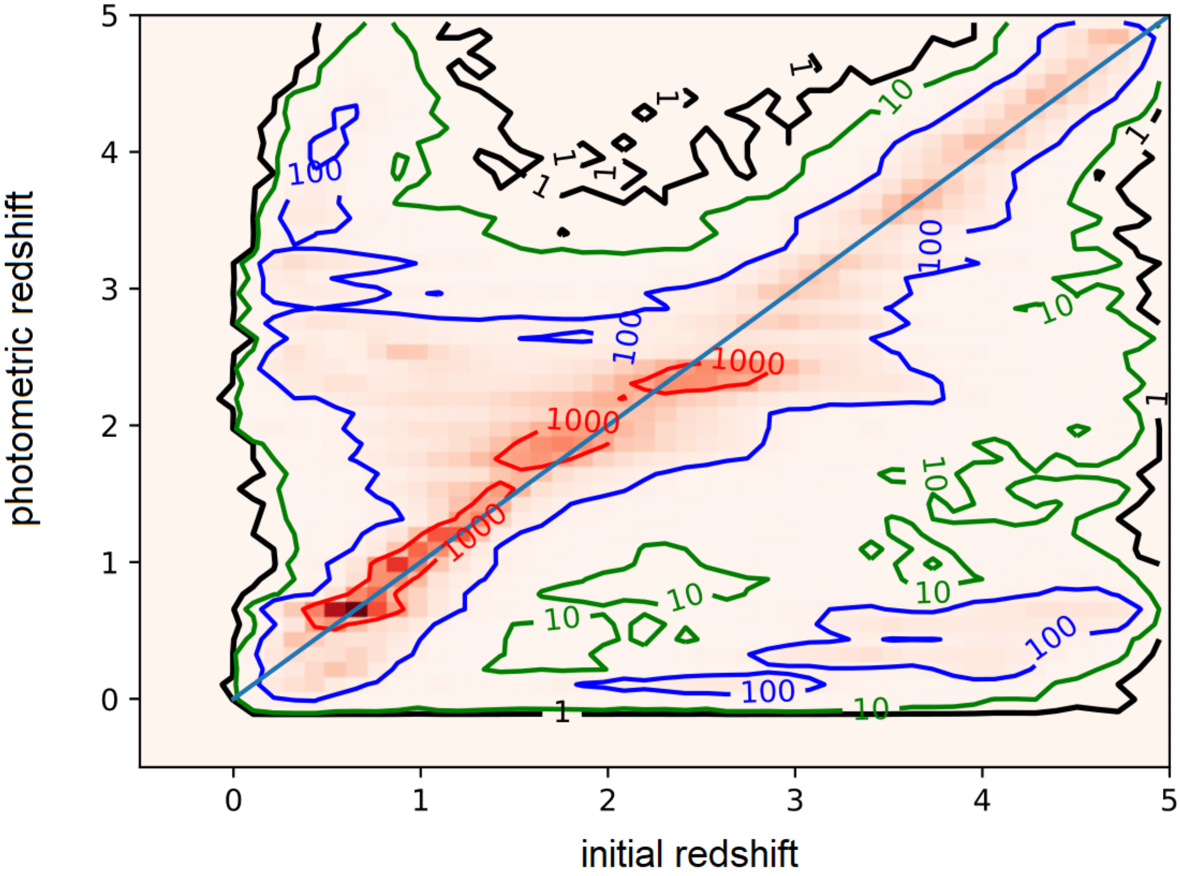
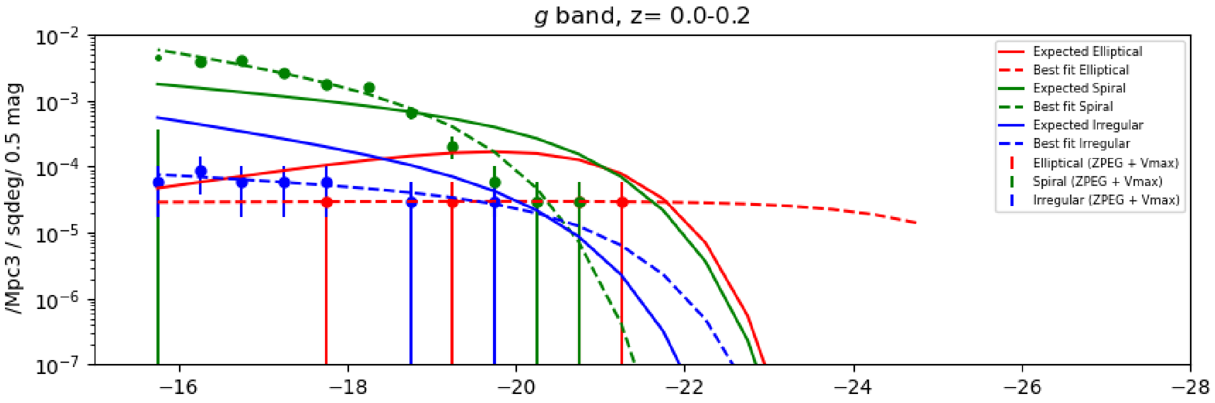
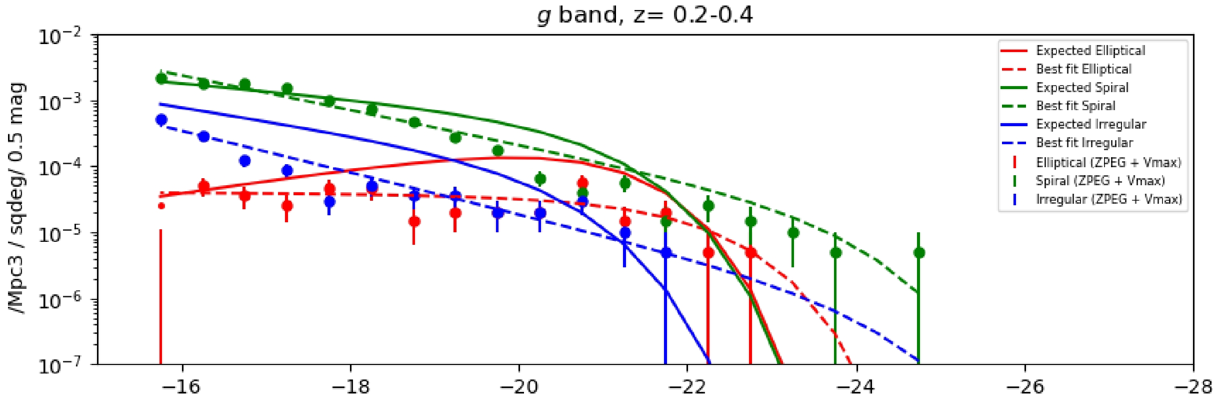
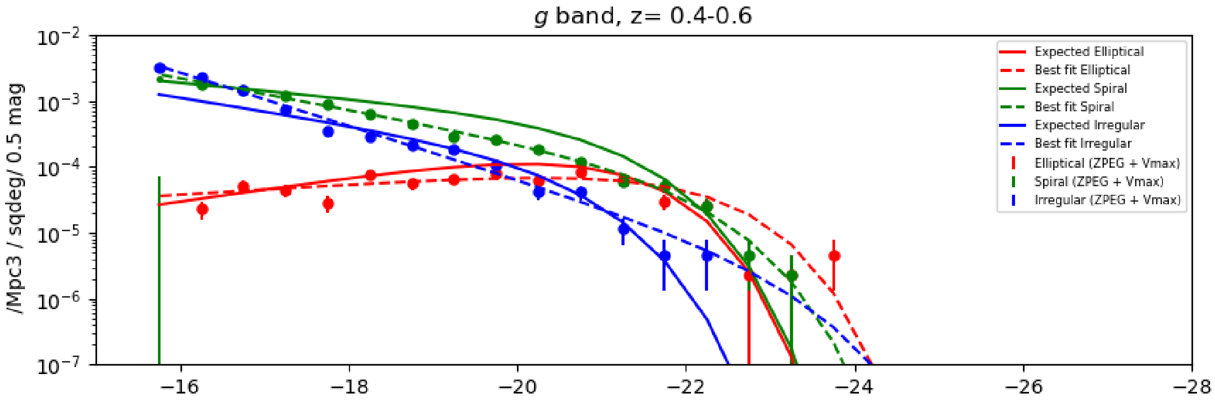


Fig. 2.9: Photometric redshifts vs initial redshifts generated by the forward model. The photometric redshifts are obtained with the `ZPEG` program of Le Borgne and Rocca-Volmerange (2002) using the apparent magnitudes in the 8 photometric bands of the sources detected with the `SExtractor` program and with the SED templates of Coleman et al. (1980). The contours of this figure in red, blue, green and black give the number of detected sources per bin above the indicated thresholds.


g band, z= 0.0-0.2


g band, z= 0.2-0.4


g band, z= 0.4-0.6

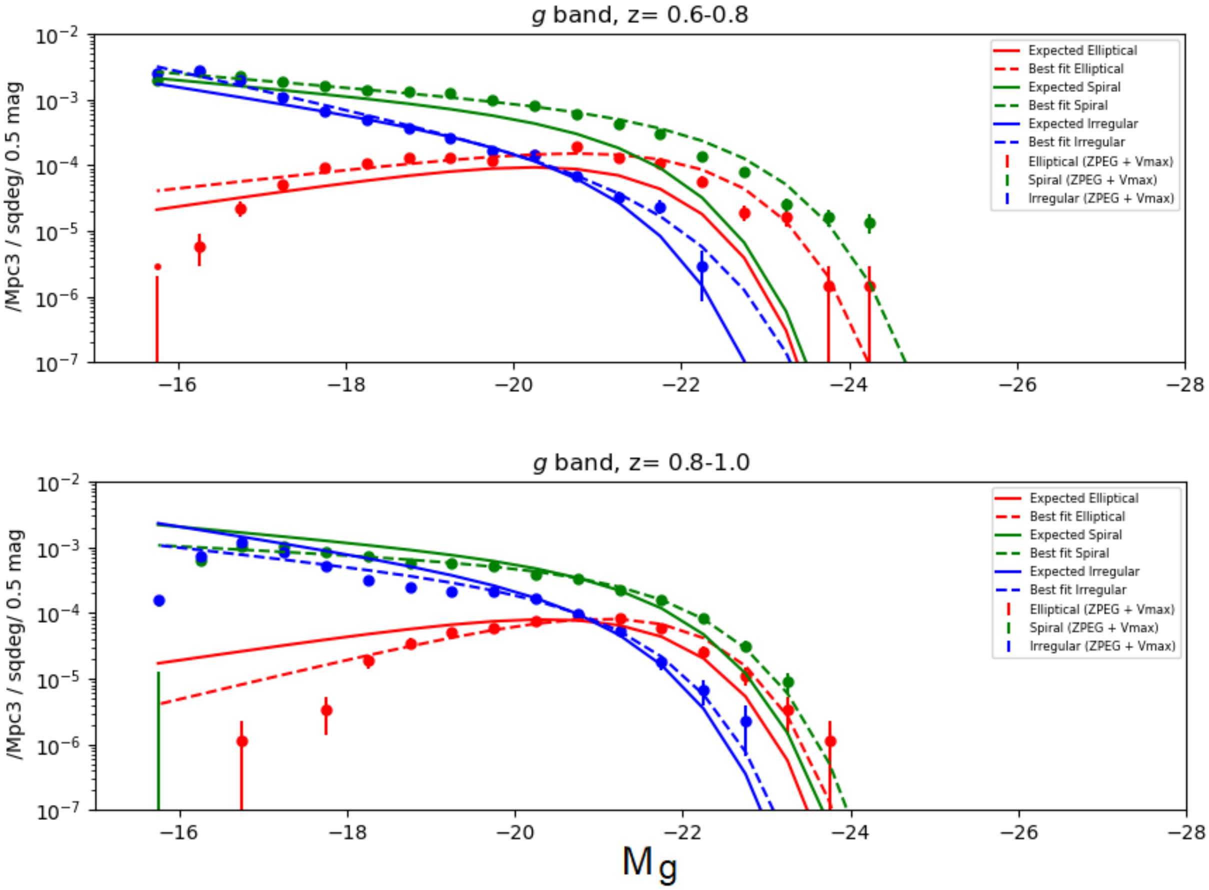


Fig. 2.10: Luminosity functions for the elliptical (red), spiral (green) and irregular (blue) populations of galaxies obtained with a $V/V_{\max}$ method, based on the photometric `ZPEG` redshifts and extracted `SExtractor` magnitudes of the sources for 5 redshift bins of size 0.2 between 0 (top) and 1 (bottom) in the g' band. The points (filled circles) and error bars correspond to the results of the $V/V_{\max}$ method, the dashed curves are the associated best fit luminosity functions (see Table 2) and the plain curves labeled "expected" are the true input luminosity functions from the forward model.

Chapter 3

Forward model

In the womb of the All-spirit
The universe lay ; the will
Blind, an atom, lay still.
The pulse of matter
Obeyed in awe
And strove to flatter
The rhythmic law.
But the will grew ; nature feared,
And cast off the child she reared,
Now her rival, instinct-led,
With her own powers impregnated. (7-17)

`Isaac Rosenberg`

Abstract

This chapter introduces the concepts inherent to the forward modeling approach. The forward model creates mock multiband deep field images that are realistic and generated with the same observational and instrumental characteristics as those of the CFHTLS D1 deep field. Therefore, the same selection effects act on both the simulated and the observed images allowing them for an unbiased comparison. The features and advantages of our forward model used to generate the simulations are fully described in this chapter.

Contents

1. Forward modeling approach

Because the various selection effects described in Chap. 2 can be spatially variable within the same survey (Bernardi et al. 2017) or correlated in subtle ways, it is impossible to express their outcomes analytically. As a result, the classic approach of extracting catalogs is a complex inverse problem of trying to correct for the effects of correlated biases of these catalogs (Marzke et al. 1998). If these effects are not dealt with properly, the models derived from these catalogs tend to be biased in a nontrivial way. In order to avoid solving the hard inverse problem of model inference from catalogs, **forward modeling** can be used (Taghizadeh-Popp et al. 2015). This technique, also called "backward modeling", aims at comparing observed data to simulated data from models. This approach is applied in many fields in astronomy such as cosmology (Akeret et al. 2015), initial mass function study (Cisewski-Kehe et al. 2019), galaxy evolution (Carassou et al. 2017; Tortorelli et al. 2020).

In the study of galaxy evolution, forward models create a distribution of simulated galaxies for various populations and pass them through a virtual telescope with the same instrumental and observational characteristics as the observed data. These are the filters used, the exposure time, the gain, the saturation limits, the detector noise, the seeing properties (see Chap. 1 Sect. 5.) as well as the various cosmological and instrumental biases described in Chap. 2. With this approach, a direct comparison in the observational space between simulated and observed data is made possible and both are affected by the same selection effects. Forward models have been used for various applications in the field of galaxy evolution: to test the reliability of the processing of galaxy catalogs (Yan et al. 2004; McBride et al. 2009), to study galaxy formation (Simard and Somerville 2000; Blaizot et al. 2005) and to construct synthetic images and catalogs (Springel et al. 2005; Bernyk et al. 2016).

2. Description of the model

Our forward model creates realistic mock deep field images in the eight photometric bands u', g', r', i', z', J, H, Ks. These images contain one or several populations of galaxies (e.g. elliptical, spiral, etc) and stars rendered using the same instrumental and observational characteristics as those of the CFHTLS D1 deep field (see Chap. 1 Sect. 5.3). The **input parameters** of our forward model are the **luminosity function parameters** described in Chap. 2 Sect. 4. for each population of galaxies we want to study.

First, the forward model creates mock photometric catalogs of galaxies with a code based on the `Stuff` program (Bertin 2010) that I have rewritten from `C` to `Python` for more flexibility and better modularity in the new developments. This code now accounts for distinct extinction in the bulge and disk components, and for a better control of the sample variance, which allows generating images with variations in the luminosity function parameters by only adding or suppressing specific galaxies with adequate luminosities keeping the other galaxies at their position in the image and with identical bulge and disk parameters (see Chap. 4 Sect. 2.6 and Chap. 5 Sect. 4.2). Each galaxy of these mock catalogs is decomposed as a sum of a bulge and a disk with the following properties: its absolute magnitude in the reference SDSS rest-frame g band, its apparent magnitudes and apparent bulge-to-total light ratio (defining its Hubble type) in each observational band (*i.e.* u', g', r', i', z', J, H, Ks), the characteristic sizes of its bulge and disk, its redshift and its disk inclination. The latter is then used to determine its internal extinction. The code uses a set of galaxy types defined by a specific luminosity function, an intrinsic bulge-to-total luminosity ratio B/T, the SED of the bulge and disk (see Sect. 2.1), the bulge and the disk scale lengths satisfying (with some dispersion) the observed scaling laws (sec Sects. 2.2 and 2.3), and extinction laws (see Sect. 2.4). The galaxies of each type are then generated with Poisson draws in small redshift bins of $5h^{-1}$ Mpc (where $h = H_0/100 = 0.7$ in this thesis and is called the "reduced Hubble constant") from $z = 0.001$ to $z = 20$. Catalogs of stars are also added at the end of this first step to obtain realistic images that can be compared to the CFHTLS D1 deep field. The galaxies and stars generated in this way are then realistically rendered on deep field images using `SkyMaker` (Bertin 2009), see Sect. 2.6. The characteristics and the properties of our forward model are described in the next sections.

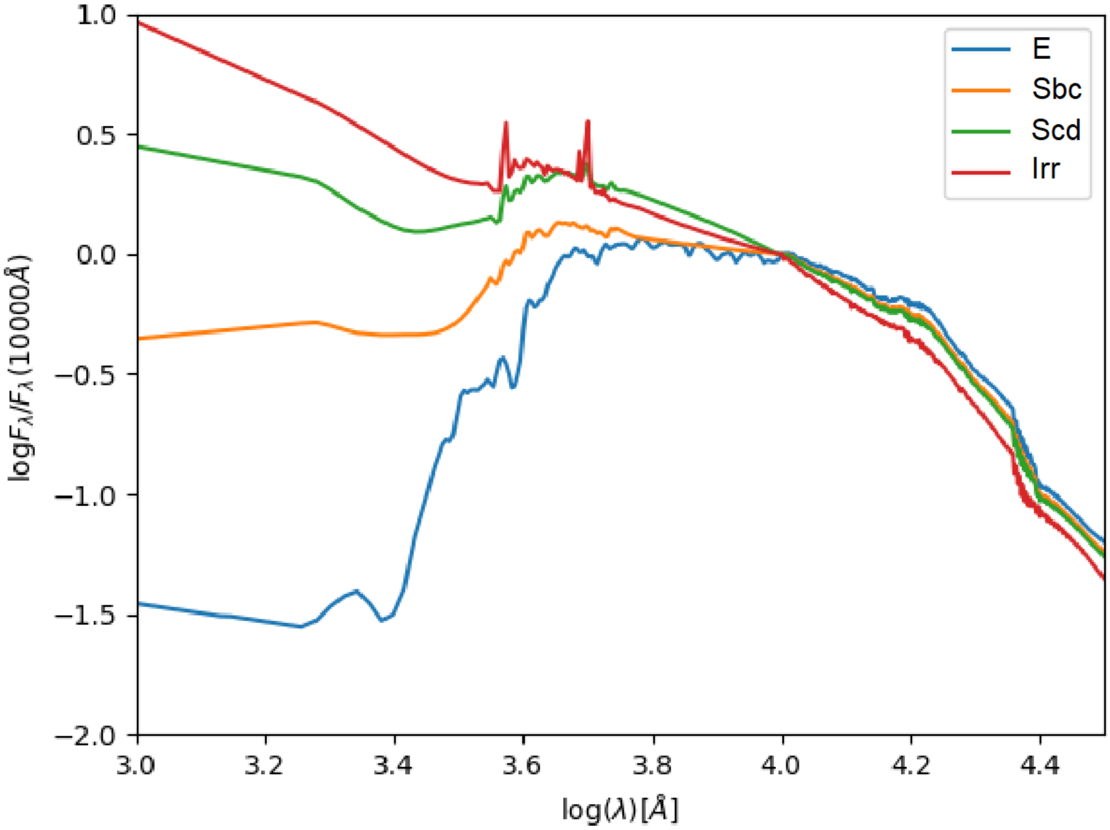


Fig. 3.1: Generic SED templates of Coleman et al. (1980). In this thesis, we use the 'E' template to model the bulge light profile and the 'Irr' template to model the disk light profile in the eight observed photometric bands u', g', r', i', z', J, H, Ks.

2.1 Spectral Energy Distributions (SEDs)

The forward model builds catalogs of apparent magnitudes in eight photometric bands and this relies on our knowledge of the SEDs of galaxies. The SED contains all the information of light emission and light absorption of the stars, gas and dust inside the galaxy. SEDs can be obtained by observation of local bright galaxies such as the templates of Coleman et al. (1980) that is used in this thesis (see Fig. 3.1), or generated from stellar population synthesis models such as `PÉGASE` (Fioc and Rocca-Volmerange 2019). We assign the generic 'E' SED template of Fig. 3.1 to the bulge component and the 'Irr' SED template to the disk component for a given morphological type in variable proportions (using B/T), which leads to a wide range of colors for the individual galaxies. In the current implementation, we keep the SEDs and the B/T ratios fixed with redshift for all galaxy types. These types are therefore defined here by classes of rest-frame colors of galaxies regardless of their redshift. However, the evolution of the SEDs of individual galaxies is implicitly allowed through the evolution of the type-dependent luminosity functions and the different extinction effects (see Sect. 2.4). An example of such an evolution is given in Fig. 3.2.

2.2 Bulge component light profile

Galaxies can have various sizes and shapes and quantifying their photometric properties is fundamental in order to create realistic simulations. The common **de Vaucouleurs profile** (de Vaucouleurs 1953) describes how the surface brightness I_B (in cd m^{-2}) of a bulge or elliptical galaxy varies as a function of the distance R (in kpc) from the center,

$$I_B(R) = I_e \exp\left[-7.669\left(\left(\frac{R}{R_e}\right)^{1/4} - 1\right)\right] \tag{3.1}$$

where R_e (in kpc) is the half-light radius or effective radius (*i.e.* the isophote containing half of the total luminosity of the bulge), and I_e (in cd m^{-2}) is the surface brightness at R_e. The same profile can be written in terms of magnitude as

$$\mu_B(R) = M_B + 8.3268\left(\frac{R}{R_e}\right)^{1/4} + 5\log(R_e) + 16.6337 \tag{3.2}$$

where $\mu_B(R)$ is the bulge surface brightness (in mag kpc^{-2}) at radius R, and M_B is the bulge absolute magnitude. It has been shown that the average effective radius follows an empirical relation with the absolute magnitude of the bulge (Binggeli et al. 1984),

$$\langle R_e \rangle = R_0 10^{-p(M_B - M_0)} \tag{3.3}$$

where $R_0 = 1.58h^{-1}$ kpc, $M_0 = -20.5$, and

$$p = \begin{cases} 0.3, & \text{if } M_B < M_0 \\ 0.1, & \text{otherwise} \end{cases} \tag{3.4}$$

As shown by Trujillo et al. (2006); Damjanov et al. (2009); Williams et al. (2010), the effective bulge radius evolves with redshift with a $(1+z)^{-1}$ factor. Furthermore, it has been shown that the intrinsic flattening*, q, of the bulge follows a normal distribution with mean $\mu = 0.65$ and standard deviation $\sigma = 0.18$ (Sandage et al. 1970). Finally, the absolute magnitude of the bulge M_B can be linked to the total absolute magnitude of the galaxy M through the relation

$$M_B = M - 2.5\log(B/T) \tag{3.5}$$

where B/T is the bulge-to-total light ratio, which is assumed not to evolve for galaxies of a given type. The common de Vaucouleurs profile can be generalized to a Sersic profile (Sérsic 1963) where the power 1/4 of Eqs. (3.1) and (3.2) can be replaced by $1/n$ with n the Sersic index which determines the shape of the light profile. In this thesis, we chose to stick to the de Vaucouleurs profile for the bulge which is well-suited for most elliptical and spiral galaxies as shown in Fig. 3.3.

---

*The flattening corresponds to the compression of a sphere along its diameter to form a spheroid, also called ellipticity. Given a the semi-major axis and b the semi-minor axis of an ellipse, the flattening is defined as $q = \frac{a-b}{a}$.

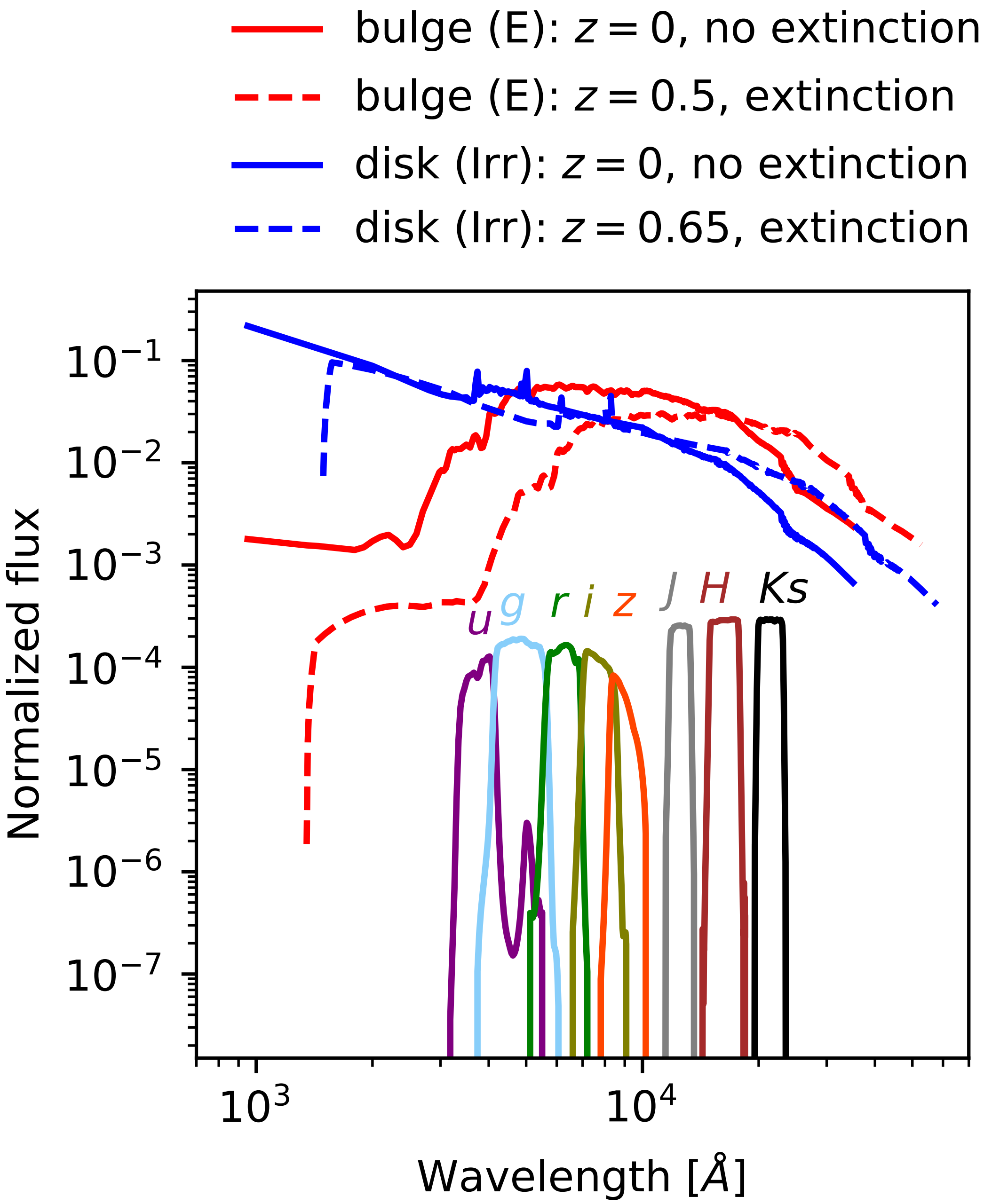


Fig. 3.2: Example of the evolution of the 'E' and 'Irr' SED templates from far-UV to near-IR. The plain red curve shows the generic 'E' template SED of Coleman et al. (1980) used for the bulge modeling at redshift $z = 0$ and without extinction. The dashed red curve shows the same SED at redshift $z = 0.5$ and with extinction (where $\omega = 1.68$ and $i = 45°$, see Sect. 2.4). The plain blue curve shows the generic 'Irr' template SED of Coleman et al. (1980) used for the disk modeling at redshift $z = 0$ and without extinction. The dashed blue curve shows the same SED at redshift $z = 0.65$ and with extinction (where $\omega = 1.68$ and $i = 45°$). The plain curves are normalized so that the integral of the template SED multiplied by the SED of the reference SDSS g' band equals 1. The eight filters used in this thesis (u', g', r', i', z', J, H, Ks) are shown at the bottom.

2.3 Disk component light profile

The **exponential profile** generally describes how the surface brightness I_D (in cd m^{-2}) of a disk varies as a function of the distance R (in kpc) from the center,

$$I_D(R) = I_0 \exp\left(-\frac{R}{h_D}\right) \tag{3.6}$$

where h_D (in kpc) is the disk scale length, and I_0 (in cd m^{-2}) is the surface brightness in the center. As for the bulge, the same profile can be written in terms of magnitude as

$$\mu_D(R) = M_D + 1.8222\frac{R}{h_D} + 5\log(h_D) + 0.8710 \tag{3.7}$$

where $\mu_D(R)$ is the disk surface brightness (in mag kpc^{-2}) at radius R, and $M_D = M - 2.5\log(1 - B/T)$ is the disk absolute magnitude. A log-normal relation linking the disk scale length to its absolute magnitude can be fit (de Jong and Lacey 2000),

$$h_D = h_0 \exp\left[0.921\beta M_D + \mathcal{N}(0,\sigma)\right] \tag{3.8}$$

where $h_0 = 3.85h^{-1}$ kpc, $\beta = -0.214$, and $\mathcal{N}(0,\sigma)$ is a random number following a normal distribution with zero-centered mean and standard deviation $\sigma = 0.36$. As shown by Trujillo et al. (2006); Williams et al. (2010), the disk scale-length evolves with redshift with a $(1+z)^{-0.8}$ factor.

2.4 Dust extinction

2.4.1 Internal extinction

The dust and gas of the interstellar medium (ISM) contained inside each galaxy strongly affect the light we receive. In fact, as shown by Dale et al. (2009), dust grains of the ISM absorb a fraction of the light in the UV/visible wavelengths and re-emit in IR wavelengths. The physical properties of dust can vary across the environment of galaxies but it is generally composed of a mixture of polycyclic aromatic hydrocarbons (PAHs), silicate and graphite (Draine 2009). When we want to derive extinction laws for galaxies, we need to resolve individual bright stars and this can only be done for our closest neighbors (Fitzpatrick and Massa 2007a): the Milky Way (MW), the Small Magellanic Cloud (SMC) and the Large Magellanic Cloud (LMC), see Fig. 3.4. Extinction curves generally present a characteristic bump at 2175 angstrom which shows that carbon is a major component of interstellar dust (Mathis et al. 1977). Quantifying the dust properties of galaxies is a challenge because no dust-free galaxy exists to be used as reference. In this thesis, we use the **Milky Way extinction curve** of Fig. 3.4 as our base SED for internal extinction by dust in distant galaxies following the equation,

$$\mathrm{SED}(\lambda) = \mathrm{SED}_0(\lambda)\exp\left[-\kappa(i,\omega,\lambda)\tau(\lambda)\right], \tag{3.9}$$

where SED(λ) is the extincted SED of the bulge or the disk, SED$_0(\lambda)$ is the face-on non-extincted SED of the bulge or the disk, $\tau(\lambda)$ is the extinction curve, and $\kappa(i,\omega,\lambda)$ is a coefficient depending on the wavelength λ, the inclination i, and the total central opacity ω of the disk. In order to determine the value of $\kappa(i,\omega,\lambda)$, we use the attenuation-inclination relations of Popescu et al. (2011) to compute the differences of extinct and non-extinct magnitudes separately for the bulge and the disk. Then, we interpolate these relations to obtain the values of $\kappa(i,\omega,\lambda)$ for a random inclination i between 0 and $\pi/2$ rad, this procedure is fully described in Appendix A. To determine the total central opacity ω of the galaxy, we use the values obtained from a Markov chain Monte Carlo (MCMC) analysis (Chevallard et al. 2013) of the SDSS Data Release 7 and of Abazajian et al. (2009),

$$\omega = \begin{cases} 1.90^{+0.23}_{-0.10} & \text{for bulgeless galaxies} \\ \\ 1.68^{+0.19}_{-0.10} & \text{for galaxies with bulges.} \end{cases} \tag{3.10}$$

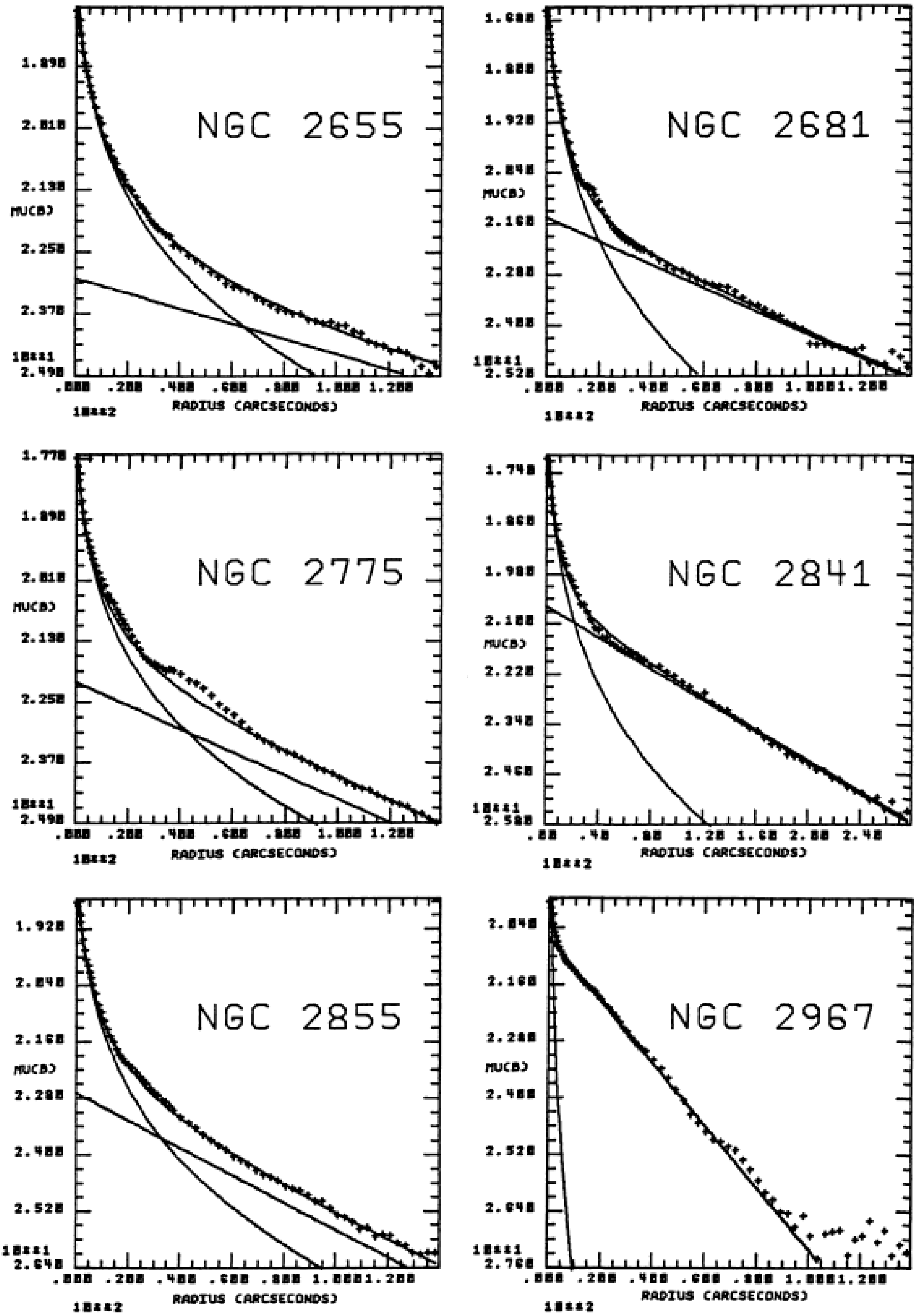


Fig. 3.3: Figure from Boroson (1981) where the surface brightness of spiral galaxies are fitted as a decomposition of a De Vaucouleurs profile for the bulge and an exponential profile for the disk. This figure shows the very good agreement between the modeled surface brightness profiles of the bulge (De Vaucouleurs) + disk (exponential) decomposition compared to the observed surface brightness profile of several galaxies.

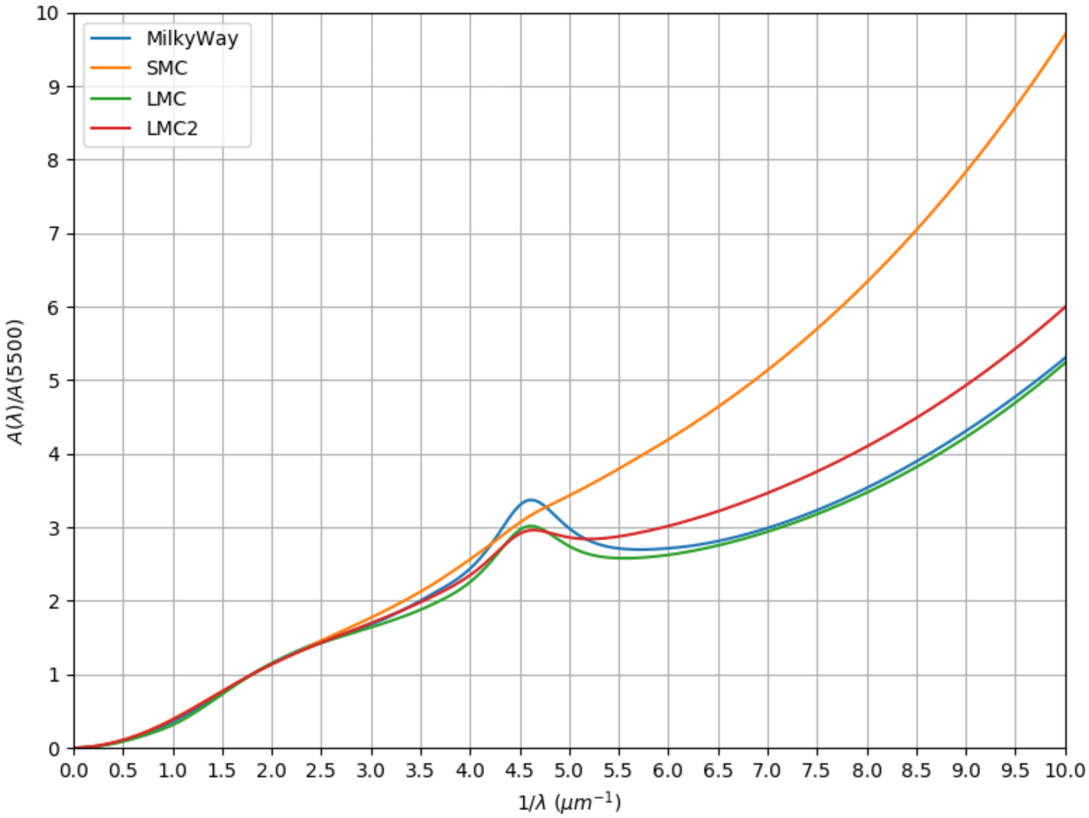


Fig. 3.4: Extinction curves from Fitzpatrick and Massa (2007a) for the Milky Way (MW), the Small Magellanic Cloud (SMC) and the Large Magellanic Cloud (LMC vs LMC2) in the UV to IR wavelengths. In this thesis, we use the Milky Way extinction curve to apply the internal light extinction of the modeled galaxies, see Sect. 2.4

2.4.2 Intergalactic medium absorption

For a specific redshift/distance, the intergalactic medium (e.g. gas clouds between the galaxy and the observer) causes an extinction of both the bulge and the disk magnitude (Madau et al. 1996) that is taken into account on average in the forward model. Fig. 3.2 shows the modifications of the 'E' and 'Irr' SED templates of Coleman et al. (1980) caused by the internal dust extinction coupled with the intergalactic medium absorption with redshift of the galaxy source.

2.4.3 Milky Way attenuation

Because we observe extra-galactic sources from inside our own galaxy, the light received has also been extincted by the dust and the gas of the Milky Way. We take this effect into account in our forward model, and each galaxy apparent magnitude is corrected following the parameterization of Fitzpatrick and Massa (1990) in the UV and spline fitting in the optical and IR with the R_V factor of 3.1 and an average $E(B-V)$ of 0.02 using the Python `extinction` package*. These corrections taken at the specific location of the CFHTLS D1 deep field are wavelength-dependent and are given in Table 3.

---

*https://github.com/kbarbary/extinction

Table 3: Parameters for the `SkyMaker` program

Passband	u'	g'	r'	i'	z'	J	H	Ks
Effective wavelength (μm)	0.355	0.475	0.640	0.776	0.925	1.253	1.631	2.146
Effective gain (e^-/ADU)	1026	7095	6919	6807	2148	1116	2379	2134
Effective exposure time (sec)	1	1	1	1	1	1	1	1
Saturation level (ADU)	6466	2710	3088	4231	12331	68068	114324	110884
Read-out noise (e^-)	5	5	5	5	5	30	30	30
Pixel size (arcsec)	0.186	0.186	0.186	0.186	0.186	0.186	0.186	0.186
Seeing FWHM (arcsec)	0.833	0.817	0.771	0.744	0.718	0.65	0.65	0.65
Sky level (AB mag/arcsec^2)	22.2	21.7	20.8	20.0	19.1	16.7	15.4	15.4
Zero-point (AB mag)	30	30	30	30	30	30	30	30
Milky Way attenuation (AB mag)	0.097	0.075	0.049	0.036	0.027	0.016	0.010	0.007

Note: The long exposure times of the CFHTLS deep fields are reflected in the high effective gains with a 1-second exposure time.

2.5 Stars

The surveys also contain stars among the galaxies we are studying. Therefore, in order to obtain realistic deep field images, our forward model adds stars using the Besançon model (Robin et al. 2003) web service, which includes their recent improvements (Robin et al. 2012, 2014; Bienaymé et al. 2015; Amôres et al. 2017). This model was run at the coordinates of the CFHTLS D1 and D2 deep fields in the eight photometric bands. When a smaller image than the 1 deg^2 deep field is generated, a Poisson random draw is performed and stars are positioned randomly (uniform law) on the image with the same spatial number density. The stars represent only a few percent of the objects below the completeness limit of $g \simeq 26.5$ mag.

2.6 Image generation with `SkyMaker`

`SkyMaker` (Bertin 2009) can generate simulated deep field images as if passed through a virtual telescope and can reproduce observed data with accuracy, see Fig. 3.5. It mimics both the instrumental and observational properties used for the observation such as the exposure time, the filters used, the background level, the diameter of the primary and secondary mirror for simulating the "aureole" effect (artefacts in the apparent "wings" of stars), and the detector characteristics (e.g. gain, image size, full-well capacity of pixels, zero-point, saturation level). Each galaxy generated in the previous steps is rendered on a high-resolution pixel grid and is convolved with an auto-computed realistic PSF* (Bertin 2011). The galaxy is also decomposed as a bulge and a disk where the bulge luminosity distribution follows the De Vaucouleurs profile of Eq. (3.1) and the disk luminosity distribution follows the exponential profile of Eq. (3.6).

In addition, for each simulation, we use the weight maps provided by Hudelot et al. (2012) and Bielby et al. (2012) to realistically model the dead pixels and the various behaviors of the CCD detectors when used on the CFHTLS D1 deep field. The image is then subsampled at the final resolution and centered at some specified coordinates. `SkyMaker` finally adds the sky background, saturation, photon (Poisson) noise and read-out (Gaussian) noise. The various parameters used for each passband are shown in Table 3.

2.7 Data conditioning

Fluxes from galaxies and/or stars can have a wide dynamic range that reaches the saturation level in the observed and simulated deep field images. This can be problematic when a neural network is fit on these fluxes because it will have more effect on very high input values than it does on the low pixel values. Therefore in the forward model we apply the following inverse hyperbolic sine transform to each pixel x of

---

*The PSF is not suitable for very bright stars or galaxies that are above the saturation level, which explains why diffraction patterns cannot be seen in our simulations.

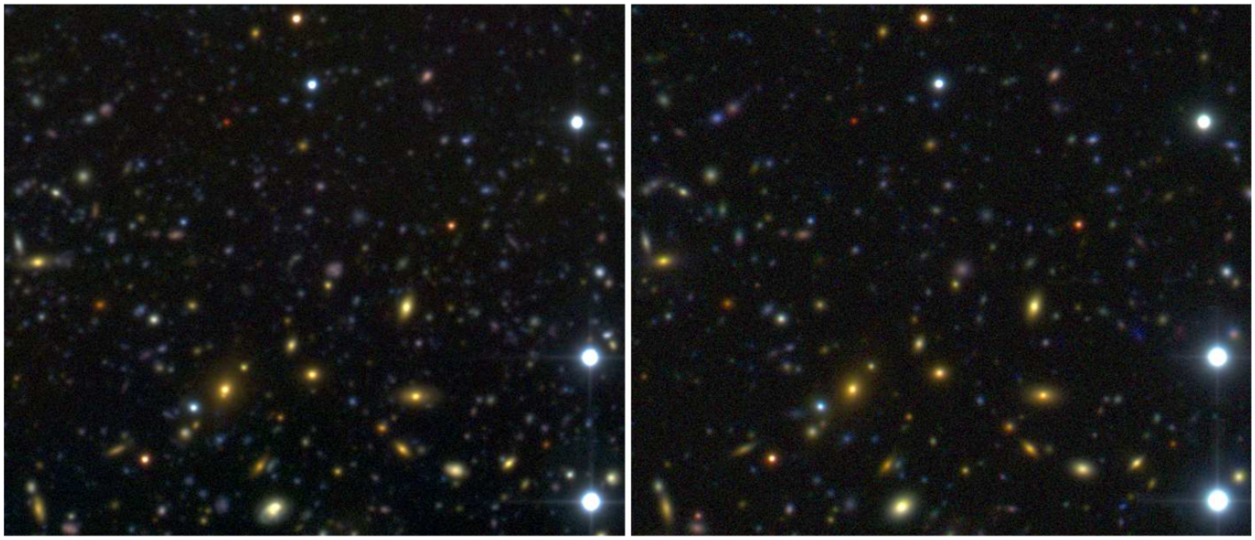


Fig. 3.5: Figure from Bertin (2009). *Left:* cutout of the CFHTLS D1 deep field (Cuillandre and Bertin 2006) in g',r',i' bands. *Right:* Using the CFHTLS D1 extracted catalogs, this simulated image is a reconstruction of the observed left image with `SkyMaker` with the same filters, exposure times, telescope properties as the data and the same color coding. This figure shows a good agreement for the visual aspects and colors of the galaxies generated by our forward model with the galaxies observed in the CFHTLS D1 deep field.

the images (see Fig. 3.6):

$$g(x) = \sinh^{-1}(x) = \log\left(x + \sqrt{x^2+1}\right). \tag{3.11}$$

This function reduces the dynamic range of the fluxes, handles negative values and allows successful fitting of the neural network with the observed and simulated deep field images, see Chap. 4.

3. 3D modeling of the dust extinction

Star-forming galaxies are fundamentally heterogeneous, with possibly obscured star-forming regions but also large zones of isolated stars and diffuse dust. This spatial diversity of dust plays a major role in moderating the propagation of photons. This can have an important effect on the shape of the observed galaxy components and on the color gradient observed (Xilouris et al. 1999). Moreover, the bulge can be partially obscured or hidden by the dust disk when a galaxy has a high inclination, as shown by Driver et al. (2007). Furthermore, the dust properties are related to the age of the surrounding stellar populations (Silva et al. 1998; Charlot and Fall 2000). Stellar winds and supernovae can cause the migration of stars and consequently form, on billion years timescales, a thick disk with on average an old red stellar population and a thin disk with a younger blue stellar population (Wielen 1977).

The work of Popescu et al. (2011) provides a three-dimensional model of dust extinction dealing with these considerations which can lead to even more realistic simulations of galaxies extincted by dust, see a schematic view on Fig. 3.7. This model extends the 2D light profiles of the bulge and the disk to equivalent deprojected 3D distributions:

- The bulge light profile of Eq. (3.1) becomes

$$I_B(R,z) = I_{B,0} \exp\left[-7.669 B^{1/4}\right] B^{-\gamma} \tag{3.12}$$

This equation is known as the "deprojected De Vaucouleurs profile" (Young 1976; Mellier and Mathez 1987).

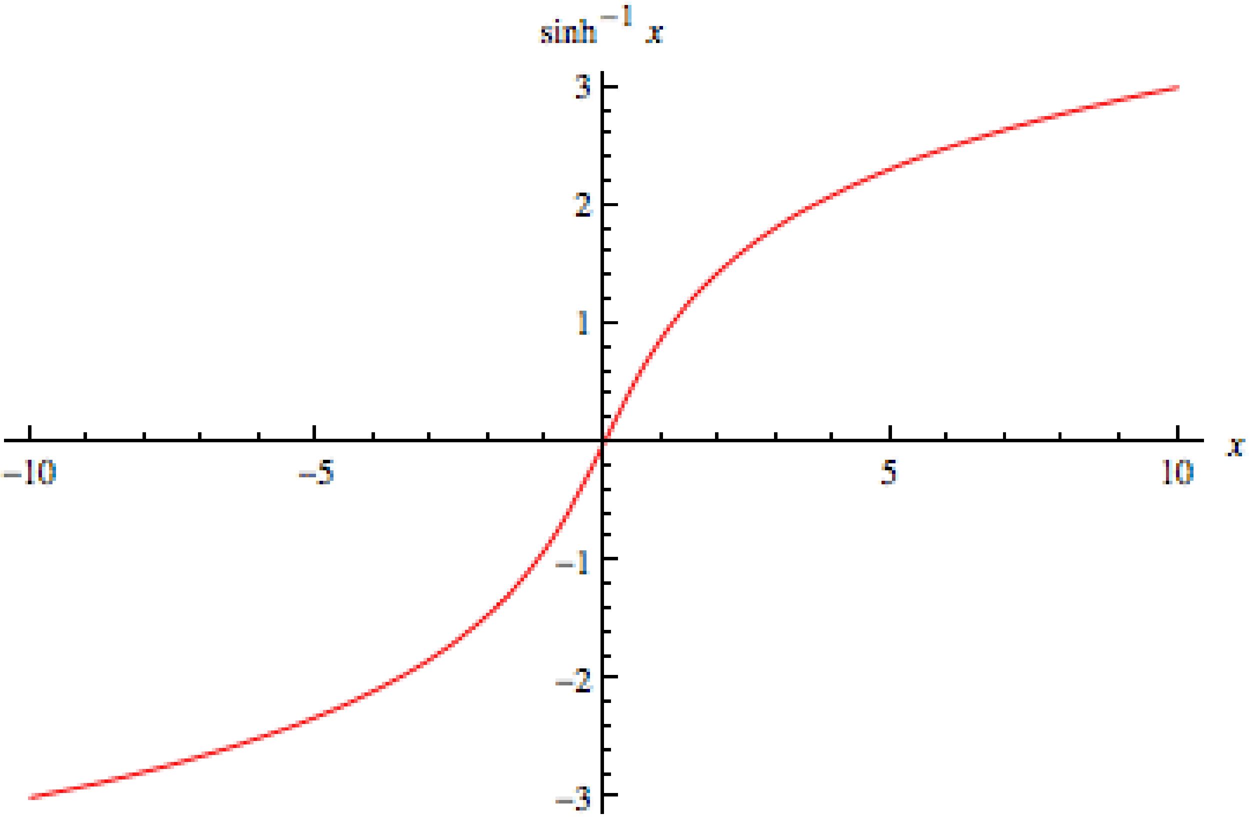


Fig. 3.6: Graph of the inverse hyperbolic sine function used to reduce the wide dynamic range of the fluxes of the galaxies and/or stars of our deep field images.

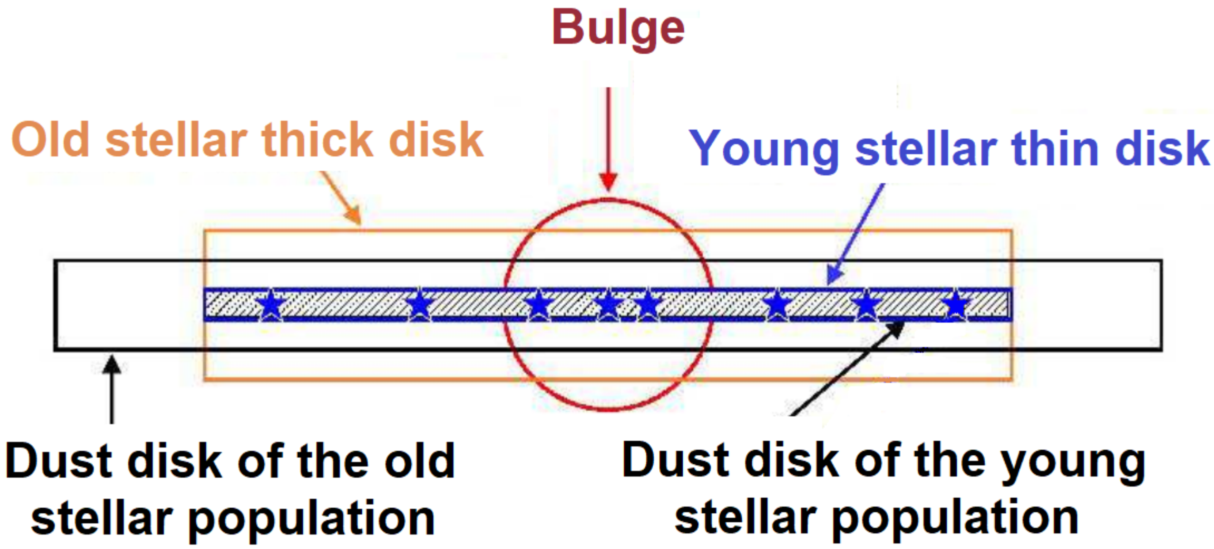


Fig. 3.7: Schematic representation of the galaxy decomposition in the 3D model of dust extinction with bulge + thick disk + thin disk + two dust disks. Image from Popescu et al. (2011).

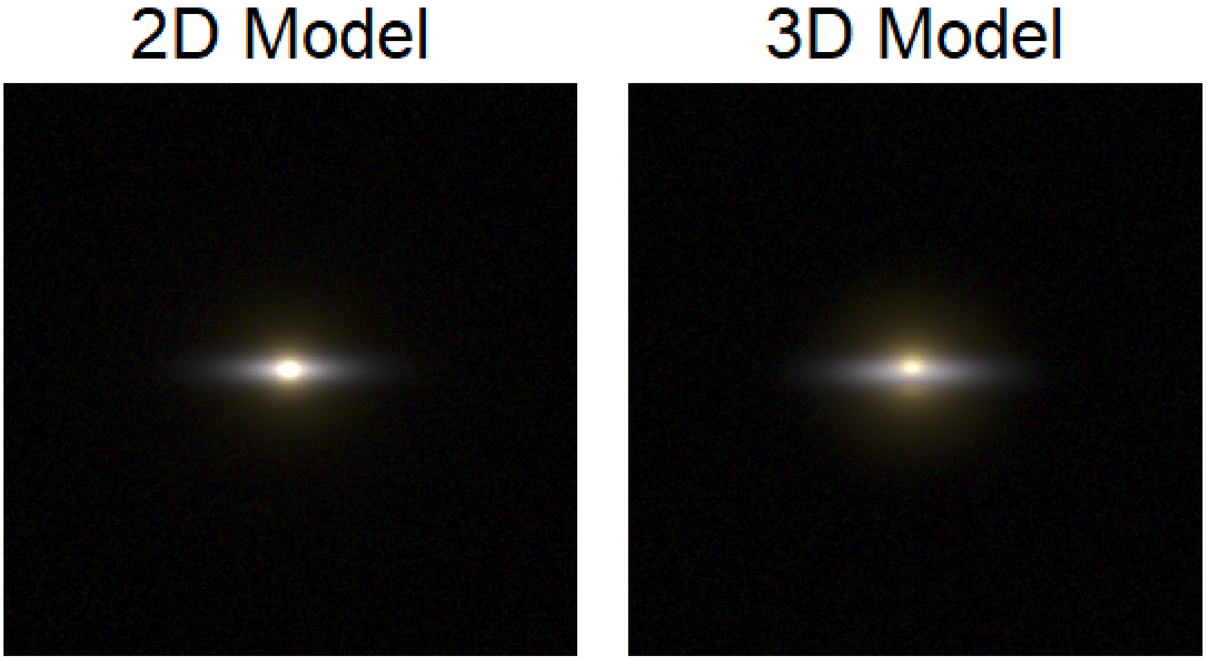


Fig. 3.8: Comparison between the current 2D model (left) and the future 3D model (right) described in Sect. 3. for the internal dust extinction. The galaxy is modeled with the following properties: total absolute magnitude $M_g = -20$, redshift $z = 0.1$, bulge-to-total ratio $B/T = 0.4$, thick disk-to-total ratio 0.4, bulge SED 'E' (Coleman et al. 1980), thick disk SED 'E', thin disk SED 'Irr', bulge effective radius $R_e = 1.34$ kpc, bulge flattening $q = 0.91$, disk scale length $h = 1.72$ kpc, inclination angle $i = 78.5°$, total central opacity $\omega = 1.75$. The two RGB color images were obtained using the g', r' and i' bands and the `STIFF` software (Bertin 2012). This comparison shows that the 3D model of internal dust extinction effectively creates a color-gradient through the three components of the galaxy and also takes into account the fact that the bulge is now partially obscured by the disks as compared to the 2D model.

- The disk light profile of Eq. (3.6) becomes for the thick disk

$$I_{\text{thick}}(R,z) = I_{\text{thick},0} \exp\left[-\frac{R}{h_D} - \frac{|z|}{z_{\text{thick}}}\right] \tag{3.13}$$

  and for the thin disk

$$I_{\text{thin}}(R,z) = I_{\text{thin},0} \exp\left[-\frac{R}{h_D} - \frac{|z|}{z_{\text{thin}}}\right] \tag{3.14}$$

where $I_{B,0}$ (resp. I_{thick} and I_{thin}) is the bulge (resp. thick disk and thin disk) central surface brightness, $B = \frac{\sqrt{R^2+z^2q^2}}{R_e}$, R and z are the cylindrical coordinates, R_e is the bulge effective radius, h_D the disk scale length which is the same for the thick and the thin disk, q (resp. $z_{\text{thick}} = 0.074h_D$ and $z_{\text{thin}} = 0.016h_D$) the bulge flattening (resp. the thick disk scale height and the thin disk scale height), and $\gamma = \frac{3}{4}$ if $B < 1$ and $\frac{7}{8}$ otherwise. This model also includes two dust disks with 3D exponential distributions of dust, as those of Eqs. 3.13 and 3.14. A first dust disk is associated with the old stellar population (with scale length $h_D^{\text{old}} = 1.406h_D$ and scale height $z_D^{\text{old}} = 0.048h_D$) and another dust disk is associated with the young stellar population (with scale length $h_D^{\text{young}} = h_D$ and scale height $z_D^{\text{young}} = 0.016h_D$). These five distributions (*i.e.* those of the bulge, thick disk, thin disk, old stellar dust disk and young stellar dust disk) are combined in a 3D-grid in a `Python` code that I have written. The bulge, thick disk and thin disk are individually extincted by the two dust disks as a function of the inclination angle i and the total central opacity ω of the galaxy and then added to account for the various effects described above. Fig. 3.8 shows an example of the differences between the current 2D model of uniform extinction and this 3D model for a galaxy simulated with the same characteristics. We can see on the right panel the attenuation of the bulge light by the dust disks now taken into account as well as the clear color gradient of the three components.

The effect of this new model on the luminosity functions and on the redshift distributions will probably be low on our current images of 1024 × 1024 pixel,s but this 3D model will have two major advantages in the future:

- The training of the network (see Chap.4) will be improved because images of galaxies with a variety of colors will be seen: for example spiral galaxies with orange-like disk due to internal dust, or with partially extincted bulge by the dust disk, will make the images more realistic.
- In the near future, we would like to apply the forward model to also measure the evolution of the scale lengths of the bulge and the disk, and this more realistic 3D model of bulges and disks will be important with this task.

Unfortunately, the `Python` CPU version of the 3D model is not suitable for the work presented in this thesis because it is too expensive in computational time and for the following results we therefore stick to the 2D model with a bulge + one disk decomposition (see Sects 2.2 to 2.3 above). However, I am currently writing a `Tensorflow` GPU version of the 3D model which should drastically decrease the time needed (by at least a factor 50), allowing me to fully use this work in future applications.

4. Comparison CFHTLS / forward model

Given that our forward model is potentially able to produce realistic multiband deep field images, we check in this section the consistency between the mock galaxies generated with the forward model and the observed galaxies of the CFHTLS Deep fields, using the magnitude number counts and the redshift distributions. To this end, we generate two populations of galaxies with specific values of the parameters of the luminosity function extracted from López-Sanjuan et al. (2017, Table 2):

$$\begin{array}{c|c}
\text{Elliptical} & \text{Spiral} \\
\hline
\log_{10}\left(\phi^*_{0,E}\right) = -2.09 + 3\log_{10} h & \log_{10}\left(\phi^*_{0,Sp}\right) = -2.04 + 3\log_{10} h \\
M^*_{0,E} = -19.68 + 5\log_{10} h & M^*_{0,Sp} = -19.71 + 5\log_{10} h \\
\alpha_E = -0.53 & \alpha_{Sp} = -1.29 \\
\phi^*_{\text{evol},E} = -1.37 & \phi^*_{\text{evol},Sp} = -0.03 \\
M^*_{\text{evol},E} = -1.15 & M^*_{\text{evol},Sp} = -1.49
\end{array} \tag{3.15}$$

The various parameters are defined in Eqs. (2.6), (2.7) and (2.8) of Chap. 2 Sect. 4., and the above parameters ϕ_0^* and M_0^* are given for a Hubble constant $H_0 = 100\ h$ km s^{-1} Mpc^{-1}. With this choice of luminosity function parameters, an eight-band deep field image of size 1024×1024 pixels is generated in 1.5s on average by our forward model. This very fast computation allows us to generate in parallel hundreds of thousands of simulated images in the PMC procedure described in Chaps. 5 and 6.

As for the evolution parameters of the luminosity function, we use the following equations to transform the values of López-Sanjuan et al. (2017, Table 2) to the values of ϕ^* and M^* at $z = 0$ and with $h = 1$:

$$\begin{aligned} \phi_0^*(z=0,h=1) &= \frac{\phi_0^*(z,h) \times 10^{-zP}}{h^3} \\ M_0^*(z=0,h=1) &= M_0^*(z,h) - zQ - 5\log_{10}(h) \\ \phi_{\mathrm{evol}}^* &= \frac{P}{\log_{10}(2)} \\ M_{\mathrm{evol}}^* &= \frac{Q}{\log(10) \times \log_{10}(2)} \end{aligned} \tag{3.16}$$

López-Sanjuan et al. (2017, Table 2) measure the values of $P = 0.41$, $Q = -0.80$ (for their quiescent population) and $P = -0.01$, $Q = -1.03$ (for their star-forming population) for $h = 0.7$ and $z = 0.5$.

Fig. 3.9 shows the total differential counts for the D1+D2 deep fields compared to those of our forward model including stars in all bands (u', g', r', i', z', J, H, Ks) and we see a good agreement for $i < 26$ which corresponds to the completeness limit of the CFHTLS. Fig. 3.10 shows the decomposition of our forward model into stars, elliptical and spiral galaxies in the i' band alone and we also see that the model counts match the observation up to the completeness limit. Moreover, this apparent "excess" of faint galaxies are visually lost when `SkyMaker` paints them onto the images because they can be drowned into the sky background.

Fig. 3.11 shows the redshift distribution of the galaxies generated from our forward model with the values in Eq. (3.15) compared to the photometric redshift distribution derived by Ilbert et al. (2006). This figure is decomposed into four intervals of apparent i' magnitude from bright $[17.5, 22]$ to faint $[24, 25]$ for redshift bins $\Delta z = 0.1$ between $z = 0$ and $z = 4$. We see a fairly good consistency between the photometric and the modeled redshift distributions. It is worth noting that in the fourth panel $24 < i' < 25$, the high redshift tail of the distribution obtained from our forward model is above those for the CFHTLS Deep fields, and this is again because these magnitudes are beyond the completeness limit of the four CFHTLS deep surveys.

We conclude from these three figures to the similarity of the model images and the observations in the total number density of objects, in their counts in each observed photometric band, and in the redshift distributions. This confirms that our forward model generates realistic multiband deep field images which can be properly compared to the CFHTLS Deep field images.

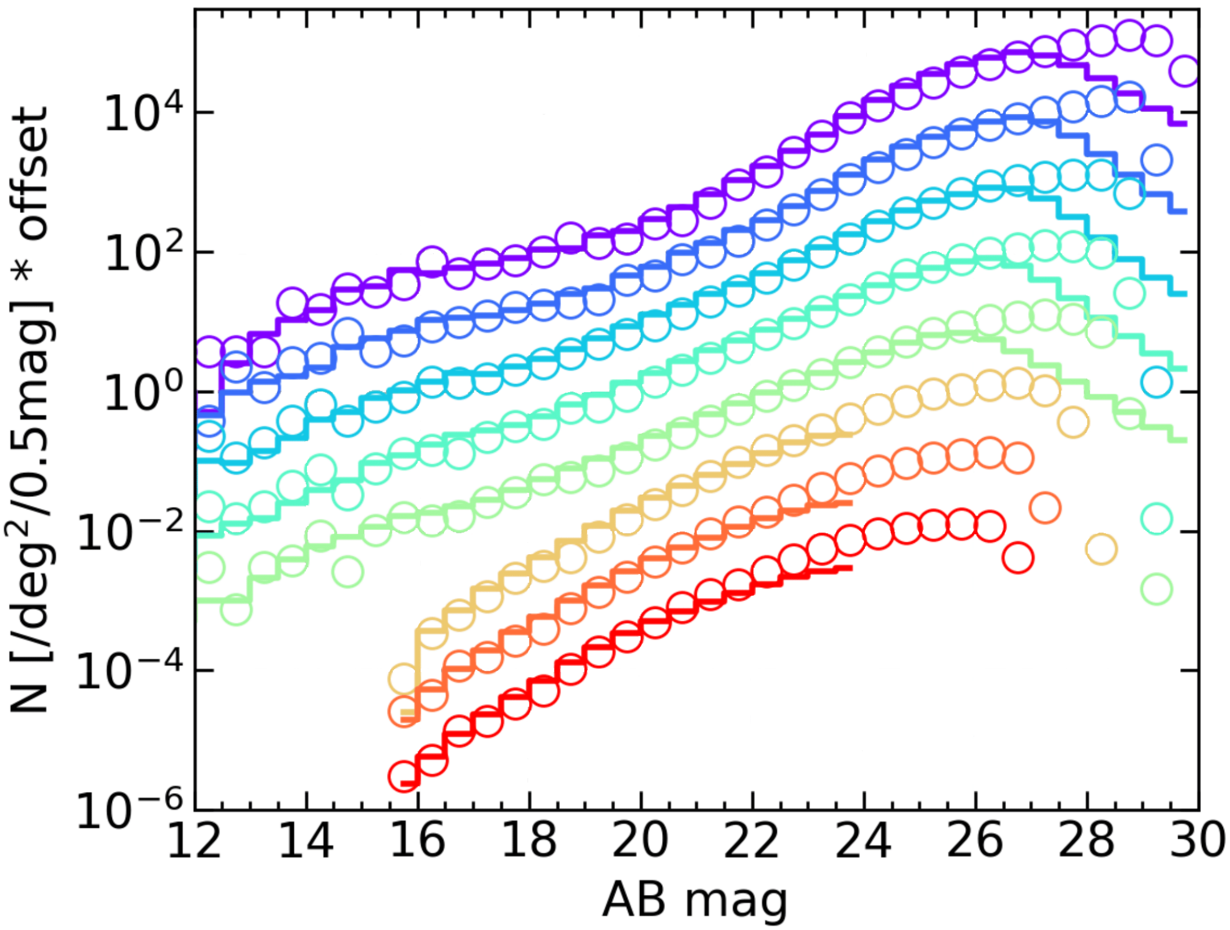


Fig. 3.9: Differential source counts in u', g', r', i', z', J, H, Ks (from top to bottom). Histograms show CFHTLS observations: star+galaxy counts built from u', g', r', i', z' source catalogs from MegaPrime D1+D2 (Hudelot et al. 2012), and J, H, Ks galaxy counts taken from WIRCam D1+D2+D3+D4 (Bielby et al. 2012). Our forward model is shown as empty circles (without stars for near-IR bands, like the observation data). For clarity, the counts in each band are regularly offset vertically downward by 1 dex from u' to Ks. This graph shows that the magnitudes of the galaxies in the forward model agree well with the observations down to their completeness limits in all eight photometric bands.

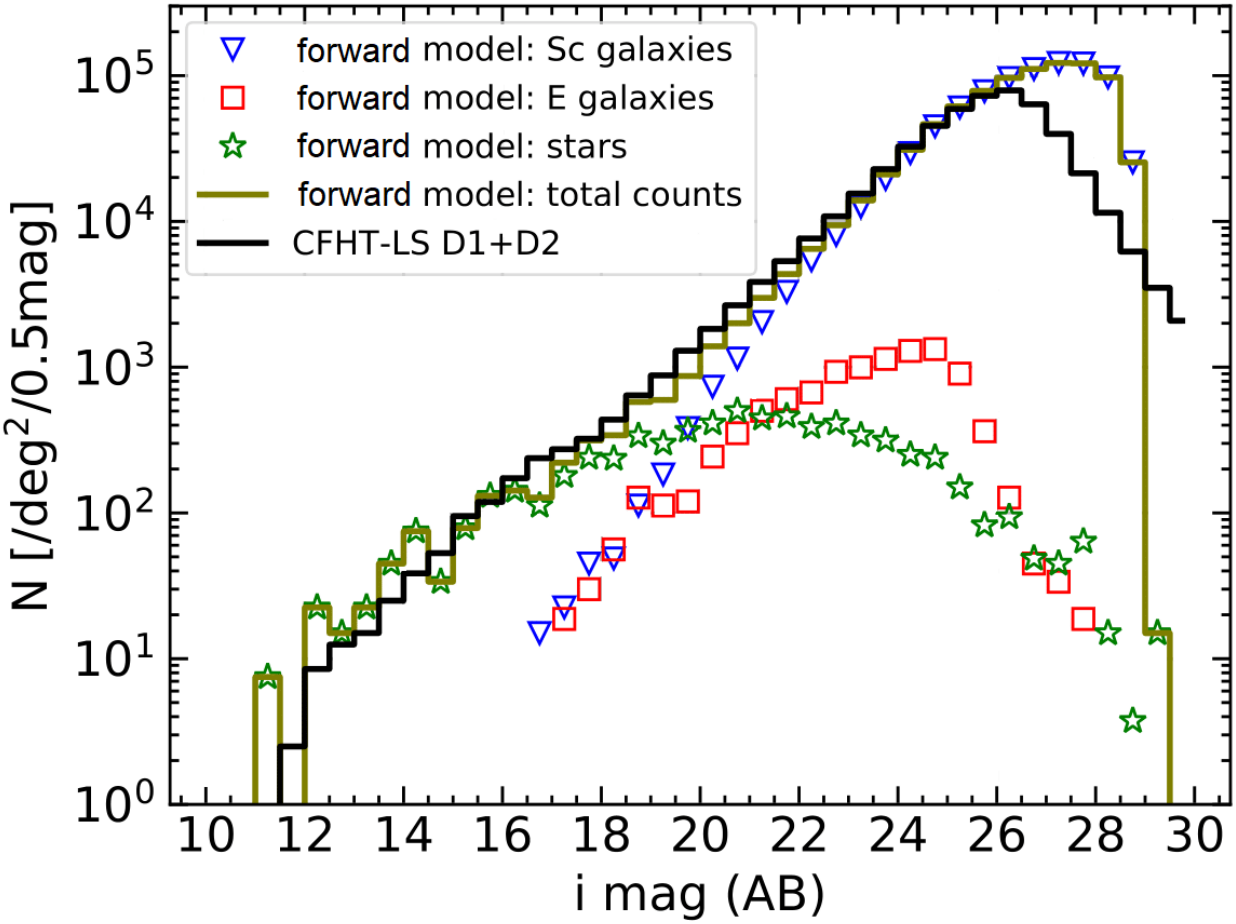


Fig. 3.10: Differential counts (stars and galaxies) in the i' band for the CFHTLS D1+D2 fields matching our forward model decomposed into stars (Besançon model), elliptical galaxies, and spiral galaxies down to the completeness limit. This graph shows the dominance of stars and spiral galaxies at the bright and faint ends, respectively, and the small number of elliptical galaxies. At the faint end, the completeness of the CFHTLS source extractions is limited to $i \simeq 26$, whereas the forward model shows the fainter input source distribution.

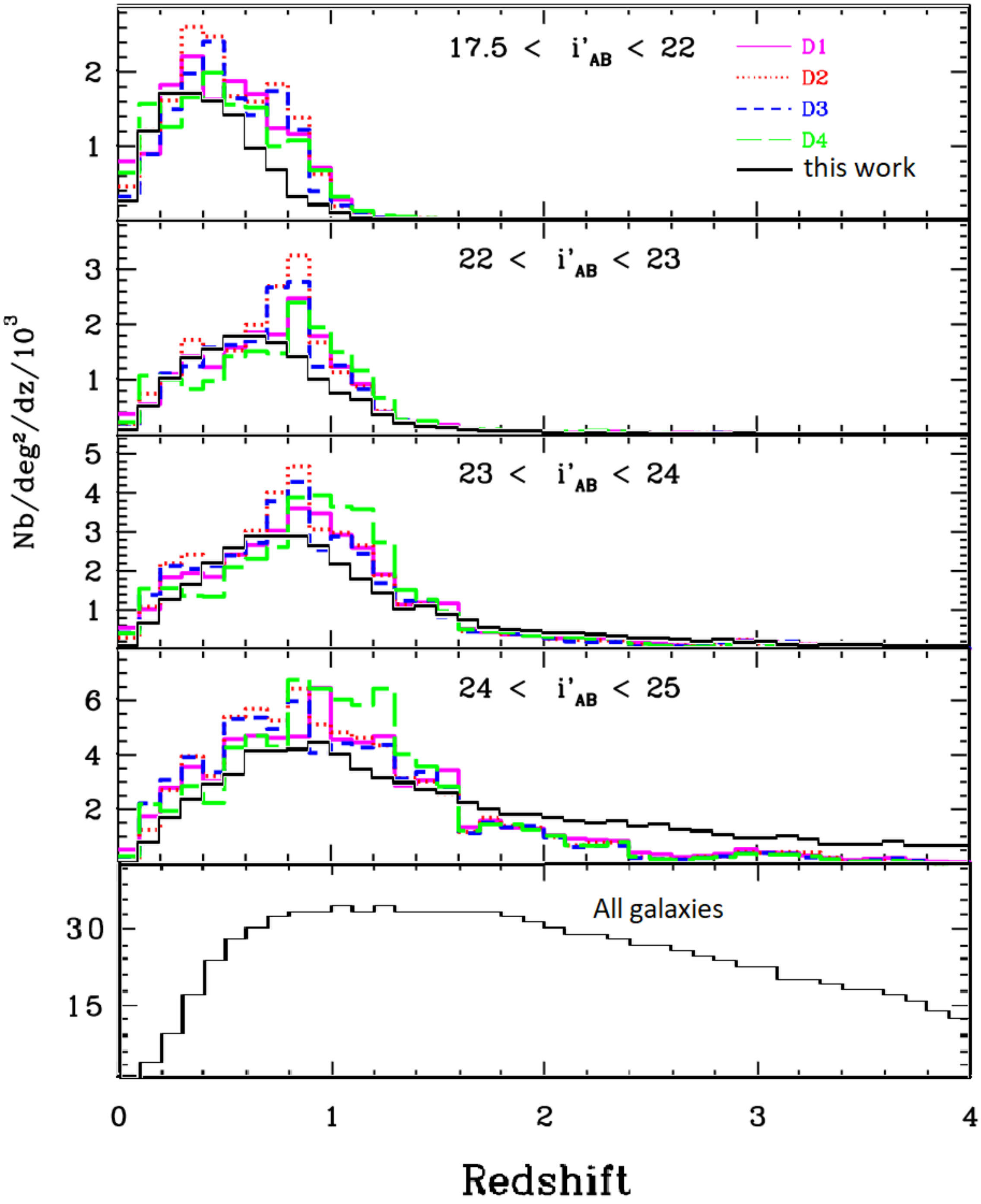


Fig. 3.11: Comparison of the redshift distributions from Ilbert et al. (2006) in the four CFHTLS Deep fields (in pink, red, blue and green) with the redshift distributions for our forward model (in black) composed of elliptical galaxies and spiral galaxies, for a total of 522 286 galaxies. This graph shows the relatively good agreement between the photometric redshifts derived from the "T0003" release of the CFHTLS (Ilbert et al. 2006) and the redshifts of the galaxies generated by our forward model. The last panel gives the redshift distribution of all galaxies generated by our forward model (without any apparent magnitude cut).

Chapter 4

Compression of multiband deep field images

> *While heading to office I hit against the door,*
> *And my forehead started to soar!*
> *Seeing in the mirror I came to know,*
> *I got a big bump with purple hue!*
> *Ice was rubbed, compression was applied,*
> *To help the purple bump to subside*
> *Pain was lessened; I got relieved,*
> *And the Purple bump was gone indeed!*
>
> `Neeru Bhatt`

Abstract

Because a straightforward comparison of multiband deep field images is not possible, in this chapter we describe the compression procedure we use. The massive compression is performed with a fully convolutional inception network whose architecture is outlined here along with its training. This compression procedure allows a catalog-free comparison of our multiband deep field images while preserving the Fisher information of our model parameters.

Contents

1. Compression maximizing the Fisher information

Using the forward model described in Chap. 3, we are now able to generate realistic deep field images and our goal is to minimize the discrepancy between the simulated and observed data. Therefore, a distance* is requested in order to quantify the similarities between two deep field images. In the case of multiband deep field images, a straightforward comparison cannot be applied, first, because of the huge amount of information contained in such images. Moreover, two multiband deep field images are two different realizations of the same luminosity functions which therefore display a limited number of galaxies with different shapes, sizes, colors and locations within the images making the comparison pixel by pixel impossible. Carassou et al. (2017) have successfully bypassed this issue with the binning of the catalogs of fluxes and sizes extracted from simulations to infer the parameters of the luminosity function, and by measuring the distance between these histograms in the observed and simulated images. However, they limited their binning process to ten per wavelength band which nevertheless leads to a 10^8-dimension space, therefore their approach is too expensive in computational time. Another way to deal with these considerations, and that we use in this thesis, is the information maximising neural networks (IMNN) method developed by Charnock et al. (2018)** which allows the compression of a number of realizations of the multiband deep field images as well as their "derivatives", using a neural network, while maximizing the Fisher information extracted from the images.

1.1 Score and Fisher information

The parameters of our forward model are the luminosity function parameters described in Chap. 2 Sect. 4. for each population of galaxies we want to study. Given a parameter vector θ of $p \in \mathbb{N}$ components $\theta = (\theta_1, \ldots, \theta_p)$, our forward model of Chap. 3 creates a data set of $n \in \mathbb{N}$ mock multiband deep field images $D = \{d_1, \ldots, d_n\}$. The likelihood function $\mathcal{L}(d_k|\theta)$ gives the probability that the multiband deep field image d_k was generated from our forward model using a specific parameter vector θ. For instance, given a specific parameter vector θ, if we compare the likelihood function of two multiband deep field images, d_1 and d_2, and find that $\mathcal{L}(d_1|\theta) > \mathcal{L}(d_2|\theta)$ then it is more likely to generate the deep field d_1 than d_2 from our forward model using the parameter vector θ.

Using the above likelihood function, we define the **score** of a piece of data d_k given the parameter vector θ:

$$S\left(d_k, \theta\right) = \nabla_\theta \log \mathcal{L}\left(d_k|\theta\right) \tag{4.1}$$

That is the gradient of the log-likelihood function with respect to the model parameters, thus the score is a vector with p components (*i.e.* like the parameter vector). In other words, the score indicates the sensitivity of the likelihood function to small changes in each component of the parameter vector of the model. The variance of the score is what is called the **Fisher information**:

$$F(D, \theta) = -\frac{1}{n} \sum_{k=1}^{n} S\left(d_k, \theta\right) S\left(d_k, \theta\right)^T \tag{4.2}$$

that is a matrix with dimension p. Formally, the Fisher information is the amount of information on the unknown model parameters that the data set, D, contains when considered from a particular position θ of the parameters space.

1.2 Unknown likelihood and summary statistics

Each piece of data d_k can be very large (e.g., in our applications of Chap. 5 Sect. 4. and Chap. 6 Sects. 1. and 2., each simulation is a 3D image with size $1024 \times 1024 \times 8$). Our goal is therefore to compress them down to summary statistics with very few components, whilst retaining as much Fisher information about the parameters as possible. A way of summarizing data have been developed in the massively optimized parameter estimation and data (MOPED) algorithm of Heavens et al. (2000). MOPED uses linear combinations to compress the original data down to summary statistics; moreover, this compression is lossless in

---

*In mathematics, a distance, also called metric, is a way of describing what it means for elements of some space to be "close to", or "far away from" each other.

**https://bitbucket.org/tomcharnock/imnn terms of Fisher information when the likelihood function is Gaussian, at least to first order. However, in our problem of studying luminosity function parameters in multiband deep field images, the likelihood is not Gaussian, and worse, is not known. In addition, compressing the data with only linear combinations may not be optimal to extract the information available.

We therefore combine the work of massive data compression from Alsing and Wandelt (2018) with the conceptual scheme of the MOPED algorithm in order to introduce the "information maximising neural networks" (IMNN) method of Charnock et al. (2018). IMNN finds the most informative nonlinear summary statistics from the multiband deep field images by training a neural network using the Fisher information matrix. This compression is in practise without any loss in terms of Fisher information, and down to a vector of p components which corresponds to the number of model parameters ($p = 2$ in the applications of Chap. 5 Sect. 4. and Chap. 6 Sect. 1. and $p = 6$ in the application of Chap. 6 Sect. 2.). Assuming a "perfect" compression scheme of the data, the work of Charnock et al. (2018) proves that p-components summary statistics are sufficient for p parameters, *i.e.* no more than p summaries are needed and logically fewer than p summaries would lose information, unless the model parameters are degenerate. Of course, the quality of the compression depends on the architecture of the neural network used, see Sect. 2.

In the following, let us consider a nonlinear application, $f : D \to T$, which compresses each piece of data, d_k ($k = 1, \ldots, n$), to a summary statistics vector, t_k with p components, under the constraint of maximizing the Fisher information of the data set. Moreover, we suppose that the application f transforms the **original unknown likelihood function** $\mathcal{L}(d_k|\theta)$ of the data into an **asymptotically Gaussian likelihood function** of the summary statistics, $\mathcal{L}(t_k|\theta)$, which can therefore be written as:

$$-2\log\mathcal{L}(t_k|\theta) = (t_k-\mu)^T\, C^{-1}\,(t_k-\mu) + \log(2\pi\det C) \tag{4.3}$$

where

- $\mathcal{L}(t_k|\theta)$ is the likelihood function of the summary statistic $t_k = f\,(d_k)$ given the parameter vector θ.
- $\mu = \frac{1}{n}\sum\limits_{k=1}^{n} t_k$ is the **mean vector** (with p components) over all the summary statistics.
- $C = \frac{1}{n-1}\sum\limits_{k=1}^{n}(t_k-\mu)\,(t_k-\mu)^T$ is the **covariance matrix** (of dimension p) of the summary statistics.

The summary statistics t_k are considered mathematically as random vectors following the multivariate normal distribution $\mathcal{N}(\mu(\theta), C(\theta))$. Because the likelihood function of the summary statistics is expressed given the parameter vector θ, we can consider that mathematically only the mean vector μ and the covariance matrix C depend on θ.

Moreover, we make the assumption that the covariance matrix C is very close to the identity matrix and therefore does not depend on the parameter vector θ (*i.e.* $\nabla_\theta C = 0$ and $\nabla_\theta C^{-1} = 0$); the expression of the loss function Λ_C of Eq. (4.27) in Sect. 2.5 gives an explanation for this hypothesis. With this assumption, and by derivation of Eq. (4.3), the score $S\,(t_k, \theta)$ of the summary statistics t_k becomes:

$$S\,(t_k,\theta) \underset{\text{def}}{=} \nabla_\theta \log\mathcal{L}\,(t_k|\theta) = -\,(\nabla_\theta\mu)^T\, C^{-1}\,(t_k-\mu) \tag{4.4}$$

In this equation, only the mean vector μ is derived with respect to the vector parameter θ. Consequently, the Fisher information of Eq. (4.26) becomes:

$$F(T,\theta) \underset{\text{def}}{=} \frac{1}{n}\sum_{k=1}^{n} S\,(t_k,\theta)\,S\,(t_k,\theta)^T = -\frac{1}{n}\sum_{k=1}^{n}(\nabla_\theta\mu)^T\,C^{-1}\,(t_k-\mu)\,(t_k-\mu)^T\,C^{-1}\,(\nabla_\theta\mu) \tag{4.5}$$

where the only term that depends on the sum index is $\frac{1}{n}\sum\limits_{k=1}^{n}(t_k-\mu)\,(t_k-\mu)^T = C$ by definition of the covariance matrix. Therefore, the Fisher information matrix simply becomes:

$$F = (\nabla_\theta\mu)^T\,C^{-1}\,(\nabla_\theta\mu) \tag{4.6}$$

where $\nabla_\theta \mu$ is a matrix of dimension p, that is the derivative of the mean vector with respect to the vector parameter θ.

In the following, we consider a **fiducial** parameter vector θ_{fid} and a new multiband deep field image d generated from our forward model with unknown parameter vector θ that we compress using the application f into a summary statistics $t = f(d)$. We want to find the most likely parameter vector θ that was used to generate this new data image d (*i.e.* we want to obtain the maximum likelihood estimate $\widetilde{\theta}(t)$ for this summary statistics t). To do so, we follow the idea of Alsing and Wandelt (2018) and we write the order 2 Taylor expansion of the log-likelihood function of Eq. (4.3) for θ close to θ_{fid}:

$$l(t,\theta) \approx l(t,\theta_{\text{fid}}) + (\theta-\theta_{\text{fid}})^T \nabla_\theta l(t,\theta_{\text{fid}}) + (\theta-\theta_{\text{fid}})^T \nabla_\theta \nabla_\theta^T l(t,\theta_{\text{fid}})\,(\theta-\theta_{\text{fid}}) \tag{4.7}$$

where $l(t,\theta) = \log \mathcal{L}(t,\theta)$ is the log-likelihood of the summary statistics t for parameter θ in Eq. (4.3). Therefore, the quasi-maximum likelihood estimator of θ about θ_{fid} is obtained by derivation of Eq. (4.7), that is:

$$\nabla_\theta l(t,\theta) = 0 \iff \nabla_\theta l(t,\theta_{\text{fid}}) + \nabla_\theta \nabla_\theta^T l(t,\theta_{\text{fid}})\,(\theta-\theta_{\text{fid}}) = 0 \tag{4.8}$$

where

$$\nabla_\theta \nabla_\theta^T l(t,\theta_{\text{fid}}) \underset{\text{def}}{=} S\left(t,\theta_{\text{fid}}\right) S\left(t,\theta_{\text{fid}}\right)^T \tag{4.9}$$

In practice, it is useful to replace the term of Eq. (4.9) with its expectation value:

$$\nabla_\theta \nabla_\theta^T l(t,\theta_{\text{fid}}) \approx \frac{1}{n}\sum_{k=1}^{n} \nabla_\theta \nabla_\theta^T l(t_k,\theta_{\text{fid}}) = \frac{1}{n}\sum_{k=1}^{n} S\left(t_k,\theta_{\text{fid}}\right) S\left(t_k,\theta_{\text{fid}}\right)^T \underset{\text{def}}{=} -F\left(T,\theta_{\text{fid}}\right) \tag{4.10}$$

Replacing the first term of Eq. (4.8) with Eq. (4.4) and the second term of Eq. (4.8) with Eq. (4.10) gives:

$$\left[\nabla_\theta \mu\left(\theta_{\text{fid}}\right)\right]^T C^{-1}\left[t-\mu\left(\theta_{\text{fid}}\right)\right] - F\left(T,\theta_{\text{fid}}\right)\left(\theta-\theta_{\text{fid}}\right) \approx 0 \tag{4.11}$$

which simplifies to the formula for the quasi-maximum likelihood estimator around the fiducial value θ_{fid}:

$$\widetilde{\theta}(t) \approx \theta_{\text{fid}} + F^{-1}\left(T,\theta_{\text{fid}}\right)\left[\nabla_\theta \mu\left(\theta_{\text{fid}}\right)\right]^T C^{-1}\left[t-\mu\left(\theta_{\text{fid}}\right)\right] \tag{4.12}$$

and finally, this $\widetilde{\theta}(t)$ is the most likely parameter vector that was used to generate the multiband deep field image d that was compressed to the summary statistics $t = f(d)$. This quasi-maximum likelihood estimator $\widetilde{\theta}$ is unbiased around the fiducial parameter (*i.e.* its expectation value is equal to the fiducial parameter): for instance, if we consider the summary statistics $t_1, \ldots, t_m$ of $m \in \mathbb{N}$ compressed multiband deep field images that are generated with m unknown parameter vectors all relatively close to θ_{fid}, then Eq. (4.12) gives:

$$\frac{1}{m}\sum_{k=1}^{m}\widetilde{\theta}\left(t_k\right) \approx \theta_{\text{fid}} + F^{-1}\left(T,\theta_{\text{fid}}\right)\left[\nabla_\theta \mu\left(\theta_{\text{fid}}\right)\right]^T C^{-1} \underbrace{\left[\frac{1}{m}\sum_{k=1}^{m} t_k - \mu\left(\theta_{\text{fid}}\right)\right]}_{\approx \mu(\theta_{\text{fid}})-\mu(\theta_{\text{fid}})=0} \approx \theta_{\text{fid}} \tag{4.13}$$

As a consequence, Cramér-Rao bound expresses the fact that, for parameters around the fiducial value, the variance of the quasi-maximum likelihood estimator of Eq. (4.12) is at least as high as the inverse of the Fisher information of the data, even if this estimator operates on the summary statistics (Alsing and Wandelt 2018), $\forall \gamma \in \{1,\ldots,p\}$:

$$\operatorname{Var}\left[\widetilde{\theta}\left(t_1\right)_\gamma, \ldots, \widetilde{\theta}\left(t_m\right)_\gamma\right] \geq F\left(D,\theta_{\text{fid}}\right)_\gamma^{-1} \tag{4.14}$$

This equation ensures that the IMNN compression scheme is lossless in terms of Fisher information, therefore, the summary statistics are as informative as possible about the model parameters (around the fiducial value).

In the following, the application f which compresses the data d to the summary statistics t will be an inception neural network that is introduced in the next sections. In particular, the significant result of Eq. (4.14) shows that, by fitting a neural network to maximize the Fisher information of Eq. (4.6), we can massively compress our large multiband deep field images down to sufficient statistic vectors.

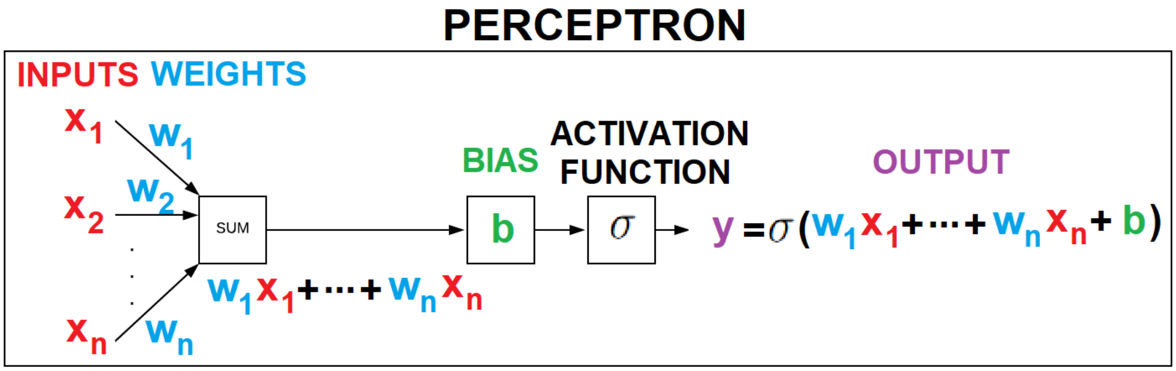


Fig. 4.1: Schematic representation of a perceptron.

2. Inception neural network

An **artificial neural network** is an application, generally nonlinear with parameters* called "weights" and "biases" to be fitted on the data. Neural networks can be used for regression or classification tasks and their success comes from the **universal approximation theorem** of Cybenko (1989) and Hornik (1991). In the past two decades, progress in the field of pattern recognition or classification on images has been tremendous. It was shown in 2004 that standard neural networks can be greatly accelerated by using graphics processing units (GPUs), whose implementation is 20 times faster than the same implementation on central processing units (CPUs; Oh and Jung 2004). Convolutional neural networks were shown to outperform more traditional machine-learning methods for images (Cireşan et al. 2010): a deep convolutional neural network won the 2012 ImageNet Large Scale Visual Recognition Challenge (Krizhevsky et al. 2012). In this thesis, given the multiscale properties and the various shapes of the galaxies in the deep field images, we therefore use a **fully convolutional inception architecture** described in the next sections.

2.1 Neural network basics: perceptron and activation functions

The idea of an "artificial neuron" also called **perceptron** was developed by Rosenblatt (1957), inspired by the earlier work of McCulloch and Pitts (1943), which basically mimics how neurons pass information among themselves. A perceptron takes several inputs, $x_1, x_2, \ldots, x_n$, and produces a single output, y. Rosenblatt introduced real numbers called "weights", $w_1, w_2, \ldots, w_n$, expressing the strength of each input x_i to the output y. A bias b is then added to introduce a variable shift and an activation function σ is applied to map the inputs to the required value, see Fig. 4.1. The output is then given by:

$$y = \sigma\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{4.15}$$

A perceptron without an activation function would only perform a linear transformation on the inputs using the weights and bias which makes it simpler but less powerful to learn complex patterns from the data. The activation functions can be of various types and are generally characterized by properties such as non-linearity, differentiability, support, monotonicity and zero-behavior. The simplest activation function is called the "binary step function" which tells whether or not the neuron should be activated based on the value from the linear transformation and is defined as:

$$\sigma(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.16}$$

---

[*] In this thesis, we always call "weights" the parameters of a neural network to make the difference with the parameters of our forward model. Therefore, "weights" are referring to the weights and biases of the neural network to be fitted on the data.

The very limitation of the binary step function is its inefficiency when more than two outputs are required. Moreover, the gradient of this function is always zero which can be an obstacle when fitting the weights/biases to a model with a gradient-based optimization. The sigmoid function is another choice as one of the most widely used non-linear activation function. The sigmoid function transforms the input real number to an output between zero and one following the equation:

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{4.17}$$

Unlike the binary step function, the sigmoid function is continuously differentiable but is very sensitive to high inputs and the output is always positive. This can be addressed by scaling the function which gives the hyperbolic tangent activation function:

$$\sigma(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{4.18}$$

The hyperbolic tangent function is also continuously differentiable and its gradient is steeper as compared to the sigmoid function. The rectified linear unit (ReLU) function is another popular non-linear activation function with the advantage of not activating all the neurons at the same time. The output of the ReLU function is only activated when the input value is greater than zero:

$$\sigma(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.19}$$

Because only a fraction of the neurons is activated, the ReLU function is far more computationally efficient when compared to the sigmoid and hyperbolic tangent functions. However, for negative x values, the gradient of the ReLU function is always zero and can lead to “dead” neurons which never get activated. The leaky rectified linear unit (leaky ReLU) function is an improved version of the ReLU function which deals with this issue:

$$\sigma(x) = \begin{cases} x & \text{if } x \geq 0 \\ \zeta x & \text{otherwise} \end{cases} \tag{4.20}$$

where ζ is a very small constant ($\zeta = 0.01$ is often considered) allowing the gradient of the negative values to be nonzero, hence no longer creating “dead” neurons. Other variants of the leaky ReLU exist such as the parameterized ReLU (for which ζ becomes a trainable parameter of the neural network) and exponential ReLU functions. The choice of the activation function largely depends upon the characteristics of the problem and will impact the convergence efficiency of the network. In this thesis, the **leaky ReLU activation function** with parameter $\zeta = 0.01$ is applied because it combines the ReLU advantages for the convergence of many neural networks (e.g. efficient computation because only comparison, addition and multiplication operations are applied; scale-invariant property; Nair and Hinton 2010; Glorot et al. 2011; Maas 2013) with the dealing of the vanishing gradient which no longer creates dead neurons.

2.2 Convolutional and pooling layer

Perceptrons are generally assembled together to form a fully-connected layer where each neuron of the input is connected to several neurons. These neurons can then be stacked such that the output of each layer is used as the input of subsequent layers. This configuration is limited in terms of computational time because the weights add up quickly, particularly when the network is composed of many fully connected layers with large inputs. When dealing with high-dimensional inputs such as multiband deep field images it is impractical to connect all inputs to all neurons in any particular layer due to both the high dimensionality of the input, but also the distribution of informative features not being location dependent as galaxies can be located anywhere in the images without any change in the luminosity function. To circumvent this issue, the **convolutional layers** calculate the signal of each output neuron as the sum of the convolved images taken in each filter: this is in effect only a local region of the input volume, as shown in Fig. 4.3. Convolutional layers are characterized by:

- a set of tunable discreet convolutional kernels often known as “receptive fields”. Normally, each receptive field is small spatially along width and height, but extends through the full depth of the input volume, *i.e.* along the different filters of our deep field images.

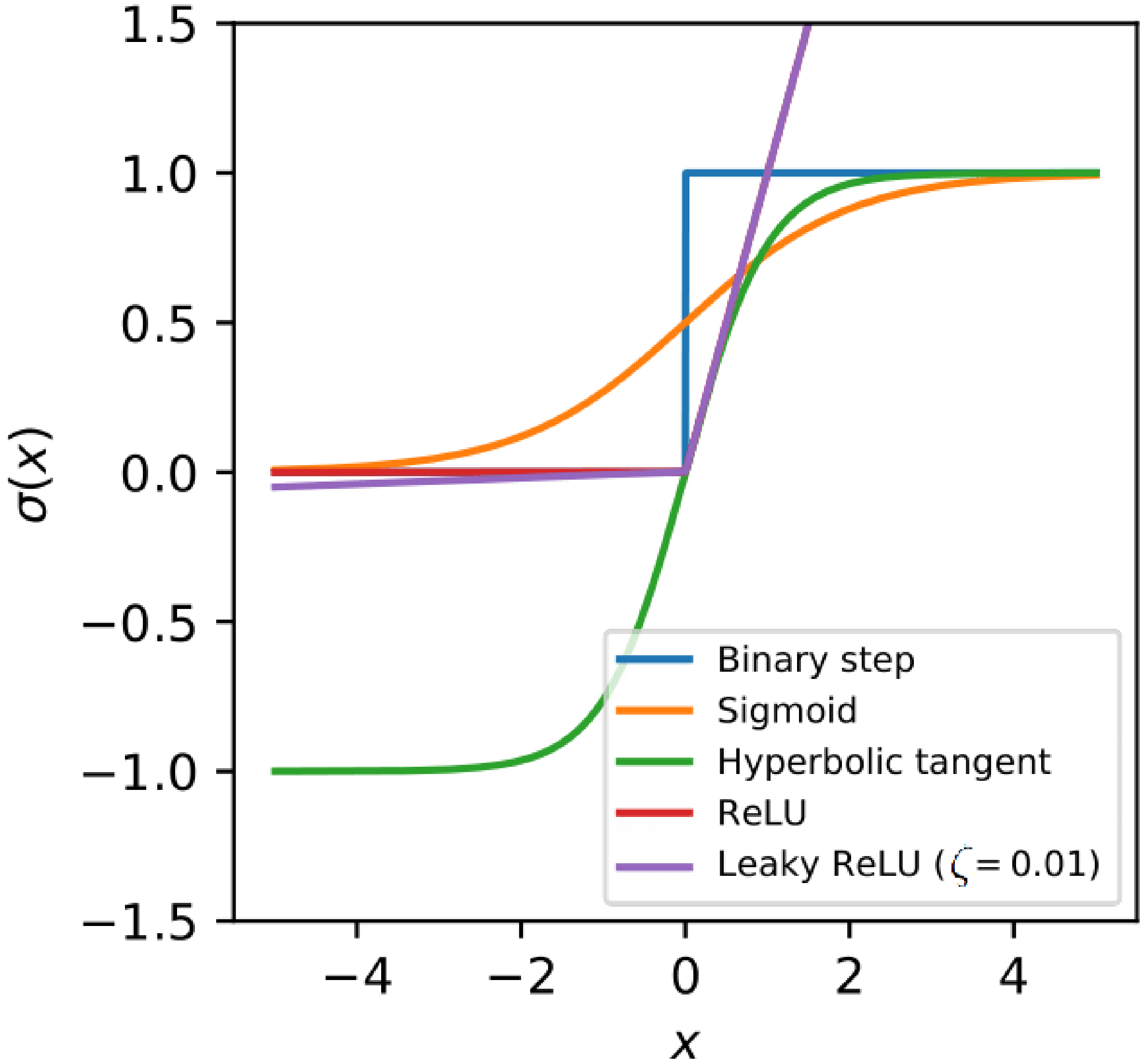


Fig. 4.2: Graphs of the activation functions described in Chap. 4 Sect. 2.1. In this thesis, we use the Leaky ReLU activation function with parameter $\zeta = 0.01$.

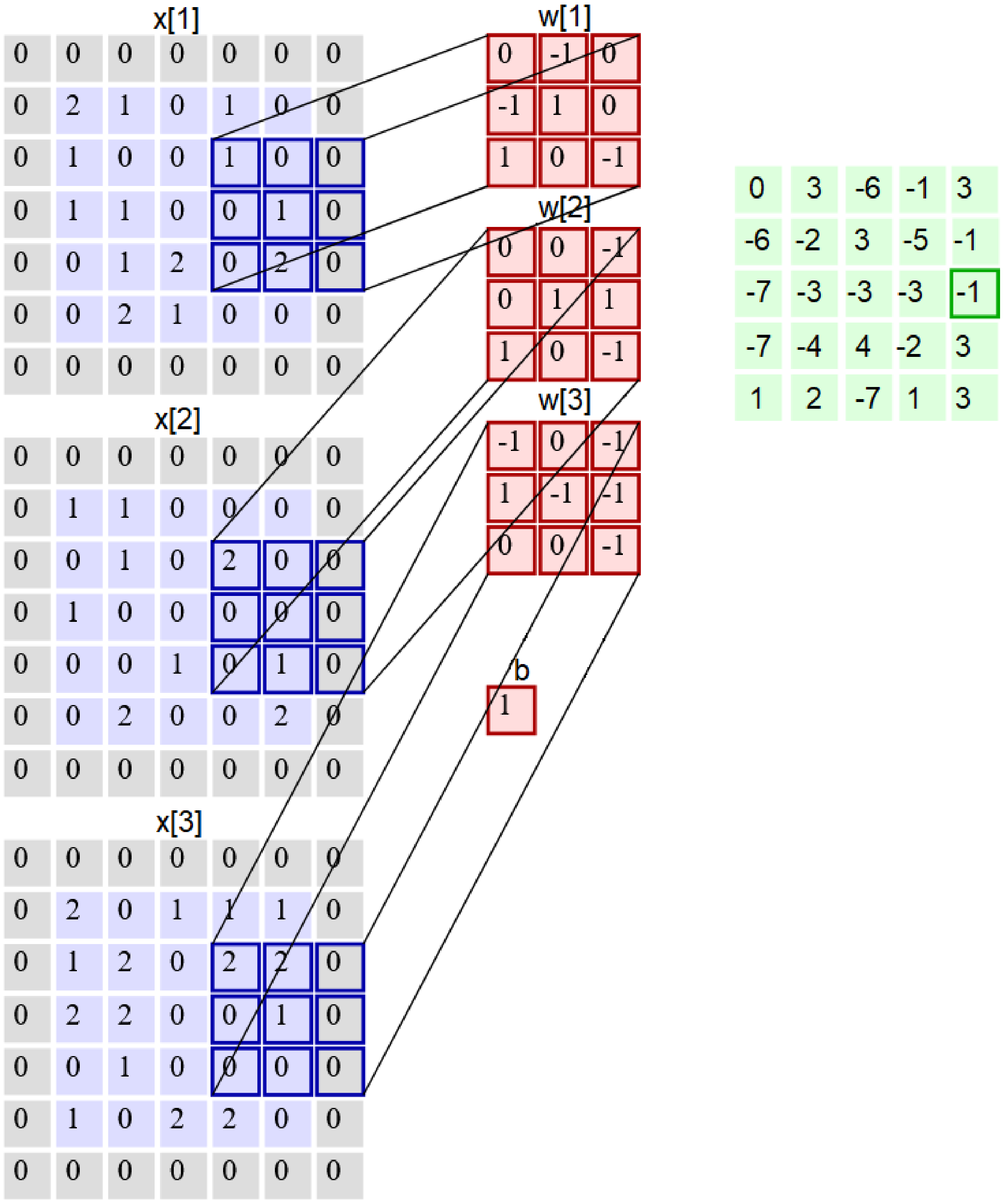


Fig. 4.3: Schematic representation of a convolutional layer with the following properties: the squares with the blue background depicts the input x of width 5, height 5 and depth 3, the squares with the red background depicts the receptive field w of width 3, height 3, depth 3 and stride 1, the squares with the green background depicts the output y of width 5, height 5 and depth 1. The green boxed square corresponds to the sum of the product of each blue boxed square by the corresponding red boxed square, plus the bias b.

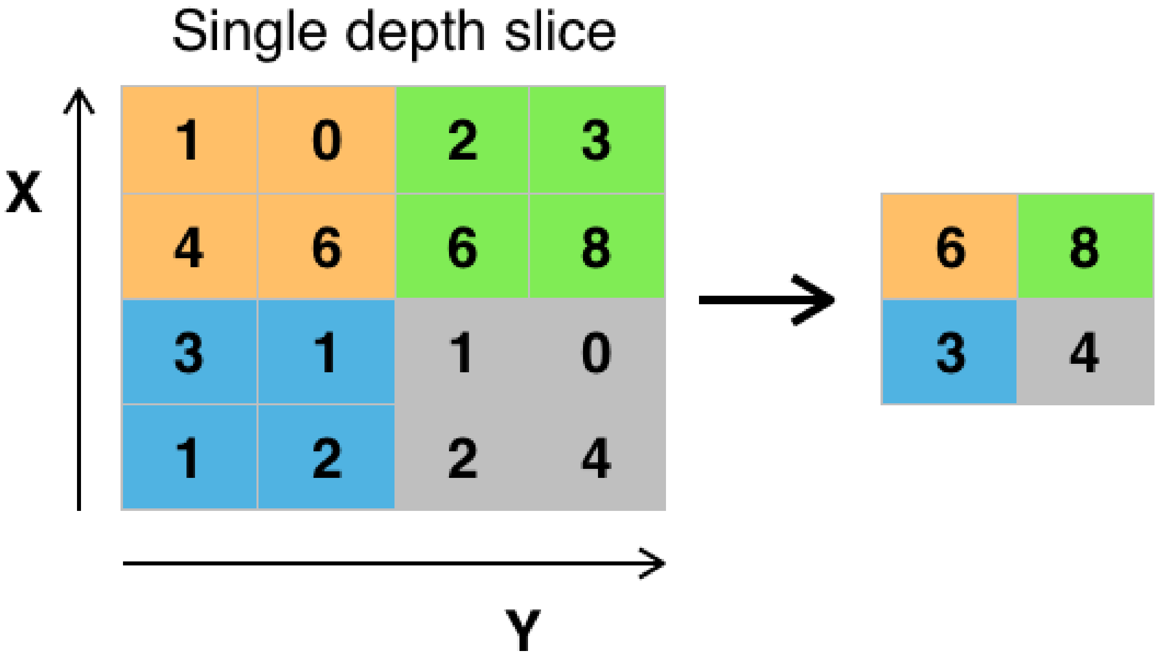


Fig. 4.4: Schematic representation of a pooling layer with the following properties: input of width 4 and height 4, receptive field of width 2, height 2, stride 2 and the "MAX" operation.

- a stride, that is the number of pixel shifts over the input matrix when applying the convolutional kernel (e.g. when the stride is 1 then the receptive field is moved by 1 pixel at a time).
- the padding, that is the amount of pixels added to the processed input (e.g. if the padding is set to one, there will be a one pixel border added to the image with a pixel value of zero) in order to adapt the width and length of the output.

Intuitively, the convolutional layer fits receptive fields that activate under the input of some types of visual feature, such as a pattern with some orientation or a patch with some particular color.

It is common to periodically insert a **pooling layer** in-between successive convolutional layers in a network. The function of a pooling layer is to progressively reduce the spatial size in order to work at different scales, the first layer works on the smallest scales then after pooling, the convolution kernel works on a larger scale, etc. The pooling layer operates independently on every slice in depth of the input and re-sizes it spatially by taking the maximum or the average of a running local patch of the output of the convolutional layer.

2.3 Inception architecture

Galaxies can be of different sizes and shapes, hence salient parts in the image can have large variation in size. Because of this variation in the distribution in scale of the information, choosing the correct kernel size for the convolution operation is delicate but critical. Moreover, very deep networks with several large convolutional layers and with millions of weights are prone to overfitting because there are not enough data to correctly adjust the weights; and naively stacking large convolution operations is computationally expensive. A way to circumvent these problems was developed (Szegedy et al. 2015) using a so-called **inception block**: a parallel mixture of convolutions and pooling layers, see Fig. 4.5. Inception blocks allow studying the input data at different scales in parallel and extracting features of different sizes from the same input. A succession of inception blocks can be used to create a full network architecture, which is called an **inception network**. The network architecture used in this thesis is shown in Fig. 4.6.

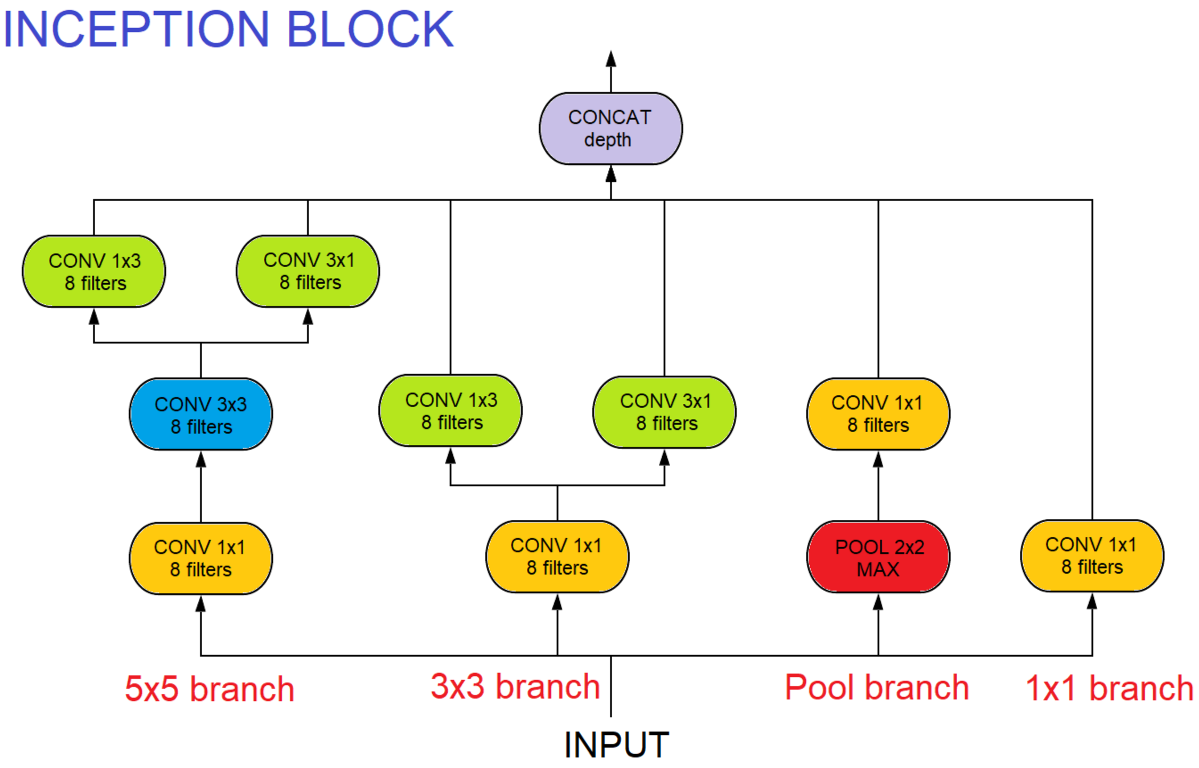


Fig. 4.5: Schematic representation of an inception block used in this thesis with 4 branches of parallel convolutional and/or pooling layers. For each layer, the size of the receptive field and the number of output filters are given. A depth concatenation is finally applied to gather the results of the different branches.

Each inception block that we use is decomposed as four branches:

- The "1×1 branch" only contains a convolutional layer with receptive field width 1, height 1 in order to apply a pixel by pixel linear combination along the full depth (of size 8) of the 3D input.
- The "Pool branch" contains a pooling layer which focuses only on the very bright pixels through a receptive field of width 2, height 2, stride 1. Then, a convolutional layer with receptive field width 1, height 1, depth 8 applies a linear combination of the bright pixels extracted by the previous pooling layer.
- The "3×3 branch" contains, first, a convolutional layer strictly equivalent to the one in the 1×1 branch and then followed by two parallel convolutional layers with receptive field of width 1, height 3 and width 3, height 1 respectively. These two parallel convolutional layers have been shown by Szegedy et al. (2015) to be equivalent to one convolutional layer 3×3 (hence the name of the branch) but with the advantage of reducing the number of weights (from $3 \times 3 = 9$ to $2 \times 3 \times 1 = 6$ weights per filter). This 33% gain can seem relatively derisory for one convolutional layer but it is very precious when the convolutional layers add up. This 3×3 branch allows for linear combinations of neighboring pixels inside a 3×3 grid which is relevant for our analysis of galaxies that extend over few pixels.
- The "5×5 branch" contains, first, a convolutional layer strictly equivalent to the one in the 1×1 branch and then followed by two consecutive convolutional layers 3×3 which replace the classic convolutional layer 5×5 (hence the name of the branch). Szegedy et al. (2015) have exploited the fact that the computation graph of the 5×5 convolution is translational invariant and can be replaced by two consecutive convolutional layers 3×3. In our architecture, we also decompose the second convolutional layer 3×3 into two parallel convolutional layers 3×1 and 1×3 as in the previous branch in order to again decrease the number of weights. This 5×5 branch allows for linear combinations of neighboring pixels inside a 5×5 grid which is also relevant for our analysis of more extended galaxies.

The number "8 filters" given in each layer of the inception block in Fig. 4.5 is not necessary in general inception architecture but is used in this thesis to keep the total number of weights relatively low (*i.e.* around 12 800 parameters for our architecture) compared to traditional inception networks with millions of parameters and because exactly eight wavelength filters are used in the multiband deep field images (*i.e.* u', g', r', i', z', J, H, Ks). At the end of each inception block, the outputs coming from all four branches are concatenated in depth to gather the results of each branch. Finally, an average pooling is applied over a 4×4 receptive field of the output to decrease the resolution of the image by 4. This process is applied exactly five times until the output is of the expected dimension, that is, the number of parameters of the model. Table 4 gives a complete description of each component of our inception network. Specifically, the full $1024 \times 1024 \times 8$ input data used in the applications described in Chap. 5 Sect. 4. and Chap. 6 Sect. 1. are massively compressed by the network down to the number of parameters in the corresponding luminosity function model. These compressed summaries are then used to compute the quantities defined in Sect. 1.2: the mean vector μ, the covariance matrix C, and finally, the Fisher information matrix F. The inception network used is fully convolutional, without any fully connected layer so that complete translational invariance is incorporated.

2.4 Backpropagation and the chain rule

A neural network can be seen as an application, f, which transforms given input data into given output data. To estimate the match between the output data and the transformed input, one needs a **loss function** (also called "cost function" or "error function" or "reward function"), Λ, that maps the output values of a neural network (transformed from the input data) onto a real number intuitively representing some "cost" associated with the input data. The neural network f contains weights and biases, $\mathbb{W}$, that are therefore fitted on the input data by minimization of the loss function in order to find a configuration for the set of weights $\mathbb{W}$ such as the transformation f is optimal. This requires the computation of the gradient of the loss function $\frac{\partial \Lambda}{\partial w}$ with respect to the weights $w \in \mathbb{W}$, which will be used to update the weights (gradients are only necessary in gradient-based optimization as in the method developed in this thesis, but other non-gradients method also exist), so that the loss function decreases. This process is called **backpropagation**.

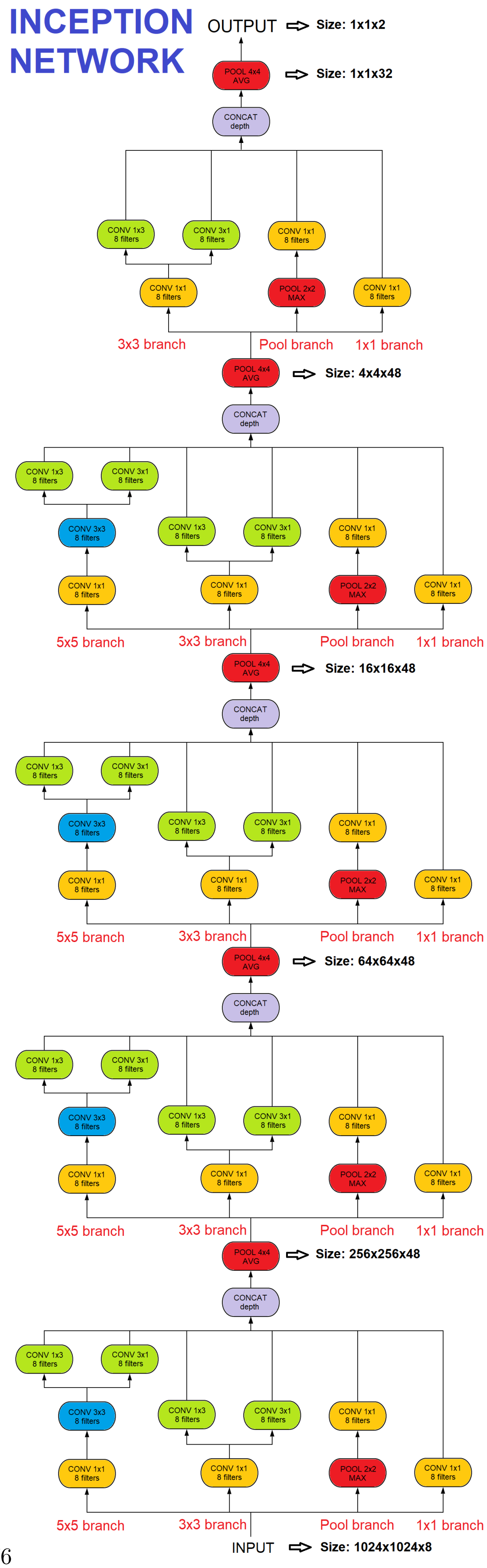


Figure 4.6: Fully convolutional inception network used in this thesis to perform the compression, see Table 4 for a full description. Each inception block is composed of parallelized convolutions that simultaneously process the same input at different scales to extract different features, then concatenates the full output. After each inception block, the input is compressed with a 4 × 4 average pooling layer to decrease the resolution by a factor 4, the resulting size is given in black. Finally, the output layer allows a compression down to the number of parameters of the model and is the summary statistics vector of Sect. 1.2

Table 4: Description of the inception network used in this thesis (see Figs. 4.5 and 4.6).

Label	Width	Height	Channels	Stride	Padding	Activation
Input	1024	1024	8	∅	∅	∅
1 × 1 convolution	1	1	8	1	SAME	Leaky ReLU 0.01
3 × 1 convolution	3	1	8	1	SAME	Leaky ReLU 0.01
1 × 3 convolution	1	3	8	1	SAME	Leaky ReLU 0.01
3 × 3 convolution	3	3	8	1	SAME	Leaky ReLU 0.01
2 × 2 max pooling	2	2	∅	2	∅	∅
4 × 4 avg pooling	4	4	∅	4	∅	∅
Output	1	1	2*	∅	∅	Linear

Note:

Each line of this table describes the colored boxes of Figs. 4.5 and 4.6.

The number of weights of a convolutional layer is given by: Number of inputs × Width × Height × Channels.

* The number of channels for the output of the network is indeed 2 in the applications of Chap. 5 Sect. 4. and Chap. 6 Sect. 1., but it is 6 in the application of Chap. 6 Sect. 2..

Because the weights are located in successive layers of the neural network, the gradient of the loss function cannot be immediately computed with respect to the weights and **chain rule** must be applied. The chain rule is similar for a fully-connected network (*i.e.* a network where all the perceptrons/neurons of a specific layer are connected to all neurons of the next layer) and a fully-convolutional network (*i.e.* as the inception network used in this thesis) but the concepts of the chain rule are easier to understand and to mathematically write down for a fully-connected network. This is the reason why, in the following, we give an example of the chain rule applied to a fully-connected network with L layers. We note, for $k = 1, \dots, n$, $y_{k,(L)}$ the n outputs of the L^{th} layer (that are also the n outputs of the neural network) which are used to calculate the value of the loss $\Lambda = \Lambda\left(y_{1,(L)}, \dots, y_{n,(L)}\right)$. We also note, for $j = 1, \dots, m$, $y_{j,(L-1)}$ the m outputs of the $(L-1)^{\text{th}}$ layer (that are also the m inputs of the L^{th} layer). Finally, we note, for $k = 1, \dots, n$ and $j = 1, \dots, m$, $w_{j,k}$ the weights, b_k the biases and σ the activation function that are used to compute the outputs of the L^{th} layer from the outputs of the $(L-1)^{\text{th}}$ as in Fig. 4.1:

$$y_{k,(L)} = \sigma\left(\sum_{j=1}^{m} w_{j,k} y_{j,(L-1)} + b_k\right) \tag{4.21}$$

To compute the gradient of the loss with respect to the weight $w_{j,i}$ (for some fixed indices $j \in \{1, \dots, m\}$ and $i \in \{1, \dots, n\}$), we therefore apply the chain rule:

$$\frac{\partial \Lambda}{\partial w_{j,i}} = \sum_{k=1}^{n} \frac{\partial \Lambda}{\partial y_{k,(L)}} \times \frac{\partial y_{k,(L)}}{\partial w_{j,i}} \tag{4.22}$$

where we can replace the last term with the partial derivative of Eq. (4.21) and the chain rule of Eq. (4.22) finally gives:

$$\frac{\partial \Lambda}{\partial w_{j,i}} = \sum_{k=1}^{n} \left[\frac{\partial \Lambda}{\partial y_{k,(L)}} \times \sigma'\left(\sum_{l=1}^{m} w_{l,k} y_{l,(L-1)} + b_k\right) \times y_{j,(L-1)}\right] \tag{4.23}$$

This backpropagation chain rule is successively applied for the previous layers to compute the gradient of the loss with respect to all weights of the neural network. Moreover, the backpropagation is automatically optimized to avoid redundant intermediate calculations, for instance in Eq. (4.23), we see that the term

$\sum_{j=1}^{m} w_{j,k} y_{j,(L-1)} + b_k$ is already calculated in Eq. (4.21). Finally, it is worth noting that usually in machine learning, the backpropagation is performed after passing only a batch (*i.e.* a subset of the full dataset) through the network. However, with the IMNN, the full dataset must be passed through the network to accurately compute the loss function, see Sect. 2.5 for the expression of the loss we use. Moreover, we demonstrate in Appendix B. that the chain rule can be applied to calculate the partial derivative of our loss function with respect to the outputs of the network, and also be applied to the other layers of our network, therefore demonstrating that the backpropagation can be properly applied.

Afterwards, using the above gradients of the loss function, the weights of the network are updated via a **gradient method**. The **stochastic gradient descent** algorithm is a simple iterative method to update the weights. Let us suppose that we are at the i^{th} iteration of the stochastic gradient descent algorithm. As in the previous paragraph, we use the n outputs of the network at the i^{th} iteration, $y_{k,(L)}^{(i)}$ for $k = 1, \ldots, n$ and we consider a weight $w \in \mathbb{W}$ (anywhere in one of the L layers of the network) with value at the i^{th}-iteration that we call $w^{(i)}$. Then, applying the stochastic gradient descent, the weight w is updated following the rule:

$$w^{(i+1)} = w^{(i)} - \frac{\eta}{n} \sum_{k=1}^{n} \frac{\partial \Lambda}{\partial w} \left(y_{k,(L)}^{(i)} \right) \tag{4.24}$$

where η is the **learning rate**, that is a user-defined hyper-parameter. This method ensures an almost sure convergence to a global minimum when the loss function is convex/pseudoconvex or to a local minimum otherwise, as shown by Bottou (1998); Kiwiel (2001). In this thesis, we use a variant of the stochastic gradient descent called the **momentum** method (Rumelhart D. E. and J. 1986) which linearly combines the gradients of the previous iteration with the current gradient and that minimizes the loss function faster. With the same notations as in Eq. (4.24), the weight w is updated following the rule:

$$w^{(i+1)} = w^{(i)} - \frac{\eta}{n} \sum_{k=1}^{n} \frac{\partial \Lambda}{\partial w} \left(y_{k,(L)}^{(i)} \right) + \frac{\delta}{n} \sum_{k=1}^{n} \frac{\partial \Lambda}{\partial w} \left(y_{k,(L)}^{(i-1)} \right) \tag{4.25}$$

where δ is the **exponential decay**, that is another user-defined hyper-parameter, set in this thesis to the value $\delta = 0.9$ (Goodfellow et al. 2016). When needed, we also use the **Adam** method (short for Adaptive Moment Estimation; Kingma and Ba 2014) which takes advantage of the averages for both the gradients and the second moments of the gradients. The Adam method is particularly useful at the beginning of the training because it stabilizes the updates faster than the momentum stochastic gradient descent algorithm and it avoids the oscillations in the gradient computations, see the applications of Chap. 5 Sect. 4. and Chap. 6 Sects. 1. and 2.

2.5 Multi-objective loss function

In many neural network optimization scheme, the loss function is a measure of the fit of the transformed input data by the neural network to the output data. However, with the IMNN, the output data is the set of summary statistics extracted from the input data. Therefore, in IMNN, the loss function must be related to the amount of information that is extracted from the data. The IMNN approach is to fit the weights of our inception network to maximize the determinant of the Fisher information matrix of the output summary statistics obtained from the network.

Moreover, we want to ensure that the covariance matrix of the summary statistics is as close as possible to the identity matrix for two reasons:

1. a covariance matrix of the summary statistics close to the identity matrix is relatively independent of the model parameters θ; *i.e.* $\nabla_\theta C = 0$ and $\nabla_\theta C^{-1} = 0$. This assumption made in Sect. 1.2 allows the use of Eq. (4.6) for the Fisher information matrix which gives the amount of information extracted from the data.

2. to limit the scaling of the summary statistics and therefore to avoid the determinant of the Fisher information matrix to reach infinite values because Lehmann and Casella (1983, Eqs. (2.5.11) and (2.6.16)) have shown that the Fisher information is invariant under linear scaling of the summaries.

One simple way to achieve this is to constrain the norm of the covariance matrix minus the identity matrix to be as small as possible and two loss functions are therefore defined:

$$\Lambda_F = -\log\left[\det F\right] \tag{4.26}$$

$$\Lambda_C = 0.5\left(||C - I||_{\mathrm{F}}^2 + ||C^{-1} - I||_{\mathrm{F}}^2\right) \tag{4.27}$$

where $||.||_{\mathrm{F}}$ is the Frobenius norm. The first loss function Λ_F measures the Fisher information, and the second loss function Λ_C measures the difference of the covariance from the identity matrix. The second loss simultaneously constrains both the covariance matrix and its inverse to be close to the identity matrix to limit the asymmetry of the convergence, thus to speed up the convergence, an example is given in Fig. 5.3 of our first application in Chap. 5 Sect. 4.2.

Our objective is to find a configuration of the weights such that both loss functions are minimal. Minimizing Λ_F maximizes the information, while minimizing Λ_C provides a covariance of the summaries that is close to the identity matrix. In the neural network framework, the minimization of two loss functions simultaneously is called a **multiobjective problem** and has to be done carefully to efficiently train the network. Several techniques (Hwang et al. 1979) exist to achieve the minimization process of such a multi-objective problem such as the linear scalarization, that is the linear combination of the two loss functions, or the ϵ-constraint method, that is an alternative use of the two loss functions functions depending of their respective value compared to an ϵ threshold. In this thesis, I have found an efficient way to combine both concepts into what is called a **continuously penalized optimization** in which the loss function becomes:

$$\Lambda = \Lambda_F + r_{\Lambda_C}\Lambda_C, \tag{4.28}$$

where $r_{\Lambda_C} = \frac{\kappa\Lambda_C}{\Lambda_C + \exp(-a\Lambda_C)}$ is a sigmoid function with user-defined parameters κ and a. The κ parameter is the strength given to the second loss function and in our applications of Chap. 5 Sect. 4. and Chap. 6 Sects. 1. and 2., we use $\kappa = 10$. At the beginning of the training, due to the random initialization of the weights following the technique of Glorot and Bengio (2010), the covariance matrix is generally far from the identity matrix; with the value $\kappa = 10$, we force the covariance matrix to quickly reach the identity matrix. The a parameter gives the degree of closeness expected between the covariance matrix and the identity matrix. In our applications, a is determined by solving the equation: $r_{0.01} = 1 - 0.01 \iff a \approx 9200$ which ensures that the r_{Λ_C} function returns $1 - 0.01$ when $\Lambda_C \approx 0.01$ in order to decrease the effect of r_{Λ_C} as soon as the covariance matrix C is close to the identity matrix.

As a result, when the covariance is far from identity, the r_{Λ_C} function is large and the optimization concentrates on quickly bringing the covariance and its inverse close to identity; when the covariance is close to identity, the r_{Λ_C} function is small and the optimization concentrates on maximizing the Fisher information.

2.6 Training the inception network

Following the description of the IMNN code (Charnock et al. 2018) to train the network when exact gradients of the forward model simulator are not available, we need a vector of **fiducial parameter values**, θ_{fid}, and a vector of **deviation values** about the fiducial, $\Delta\theta$. From these values, we generate a **training set** of multiband deep field images generated by the forward model of Chap. 3 consisting of:

- A first collection of n multiband deep field images generated with values at the fiducial parameters θ_{fid}.
- A second collection of m multiband deep field images generated with parameter values $\theta_{\mathrm{fid}} + \Delta\theta_\gamma{}^*$, for each $\gamma \in \{1, \ldots, p\}$.
- A third collection of m multiband deep field images generated with parameter values $\theta_{\mathrm{fid}} - \Delta\theta_\gamma$, for each $\gamma \in \{1, \ldots, p\}$.

As illustrated in Fig. 4.7, the training set of multiband deep field images is passed through the network as a whole to obtain the summary statistics t_k $(k = 1, \ldots, n)$ and t_k^+, t_k^- $(k = 1, \ldots, m)$, and we compute the

---

[*] $\theta_{\mathrm{fid}} + \Delta\theta_\gamma = \left(\theta_{\mathrm{fid},1}, \ldots, \theta_{\mathrm{fid},\gamma} + \Delta\theta_\gamma, \ldots, \theta_{\mathrm{fid},p}\right)$

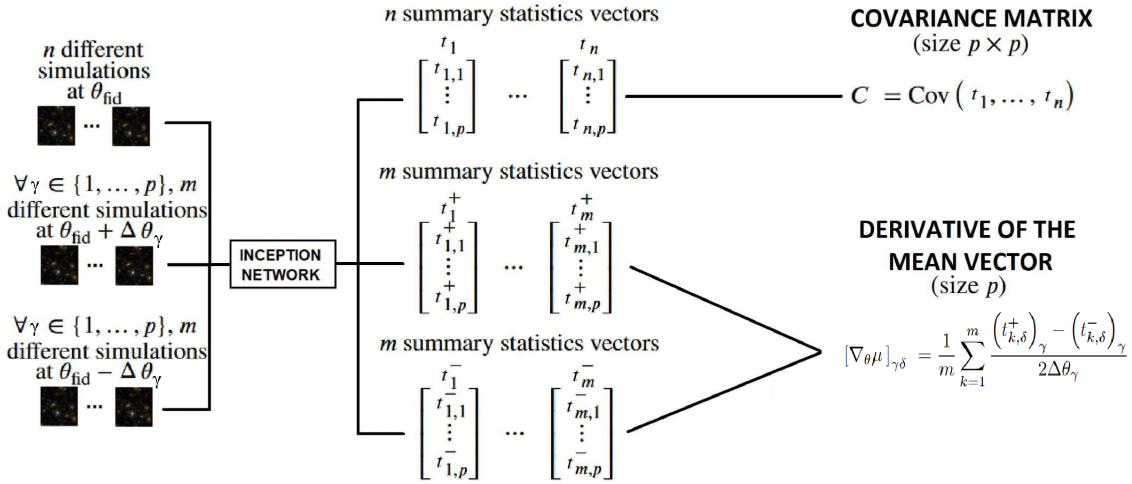


Fig. 4.7: Schematic description of the training of the network to obtain the covariance matrix and the derivative of the mean vector that are used in Eq. (4.6) to obtain the Fisher information matrix. The top, middle, and bottom lists of images form what we call the training set and are generated with fiducial parameters θ_{fid} (respectively with $\theta_{\text{fid}} + \Delta\theta_\gamma$ and $\theta_{\text{fid}} - \Delta\theta_\gamma$, for each $\gamma \in \{1, \ldots, p\}$). From the first collection of images, we compute the covariance matrix, and from the second and third collections of images, we compute the derivative of the mean vector for each $\gamma \in \{1, \ldots, p\}$.

quantities of interest: the covariance matrix C from the fiducial set of images, the derivative of the mean $\nabla_\theta \mu$ of the summary statistics vector for each parameter from the perturbed parameter sets ($\theta_{\text{fid}} \pm \Delta\theta$), and the resulting Fisher information matrix F of Eq. (4.6) in Sect. 1.2. For the simulations for the numerical derivatives, each parameter is varied independently, keeping the other variable parameters fixed at their fiducial values. Furthermore, for the second and third collections of images, the initial seeds are fixed to be the same for each individual image, so that the galaxies are generated at the same location in the paired-images and with identical flux and shape. This ensures that the only changes from one image of the second collection to the corresponding image of the third collection are galaxies that appear or disappear according to the luminosity function. Moreover, this process also guarantees that the stars are generated exactly at the same location in the paired-images and with identical flux and shape. The suppression of this variance allows us to make fewer simulations (*i.e.* $m < n$) for the derivatives than for the fiducial simulations used to calculate the covariance matrix (e.g. $n = 200$ and $m = 50$ in the applications of Chap. 5 Sect. 4. and Chap. 6 Sect. 1. ; and $n = 350$ and $m = 100$ in the application of Chap. 6 Sect. 2.).

If the covariance matrix of the summary statistics is easy to obtain, it is not the case for the derivative of the mean vector of the summary statistics, because the parameters of our forward model are only accessible in the simulated images. We therefore perform first order **numerical differentiation** to obtain the derivatives of the mean vector of the summary statistics, that is a matrix of dimension p defined as, $\forall \gamma, \delta \in \{1, \ldots, p\}$

$$[\nabla_\theta \mu]_{\gamma\delta} = \frac{1}{m}\sum_{k=1}^{m} \frac{\left(t^+_{k,\delta}\right)_\gamma - \left(t^-_{k,\delta}\right)_\gamma}{2\Delta\theta_\gamma} \tag{4.29}$$

where $\left(t^+_{k,\delta}\right)_\gamma$ (resp. $\left(t^-_{k,\delta}\right)_\gamma$) is the δ^{th}-component of the summary statistics vector associated with a simulated multiband deep field image of the second (resp. third) collection of images described above with parameter values $\theta_{\text{fid}} + \Delta\theta_\gamma$ (resp. $\theta_{\text{fid}} - \Delta\theta_\gamma$) for each $\gamma \in \{1, \ldots, p\}$.

At each iteration of the training phase, the training set of multiband deep field images is passed as a whole (this corresponds to one **epoch** of the training process) through the network before the backpropagation is run (see Sect. 2.5 of this chapter and Appendix B. for a full description). Consequently, as the training advances, the network summarizes the data while retaining increasingly more information about the model parameters.

Furthermore, as shown in the work of Charnock et al. (2018, part VI), the inception network becomes increasingly insensitive to parameter-independent noise, providing a function that calculates extremely robust summary statistics. Another independent set of different multiband deep field images that the network did not train on is also used to validate the performance of the training, it is called **the validation set**.

For a given training (or validation) set of multiband deep field images, and given a specific network architecture, the expected Fisher information is generally unknown. We therefore train the network until the loss reaches a plateau (or starts to decrease) when evaluated on the validation set, indicating that the network has compressed the data with the maximum Fisher information available. At this point, the summaries are **Gaussianly distributed** in regions close to the fiducial choice of parameter values, even if the likelihood of the actual data is non-Gaussian (Charnock et al. 2018).

2.7 Iterative procedure

Once trained, the IMNN will be used to derive for each input image its optimal summary statistics, to eventually produce the posterior distribution of the parameters (see Chap. 5). To this end, the IMNN generally performs well when evaluated on new data with parameters close to the fiducial choice of training parameters, but if applied in parameters regions that are "far" from the fiducial choice of training parameters, then there is a risk that the summaries are no longer Gaussianly distributed. This can lead to very wide or flat posterior distributions when the Bayesian likelihood-free inference techniques of Chap. 5 Sects. 2. and 3. are applied. To fix this problem, the following iterative method can be applied to choose a closer set of fiducial training values:

1. Train the IMNN with any guess (or random choice) for the fiducial values following the procedure of the previous section.
2. Compute the quasi-maximum likelihood estimate of the data using Eq. (4.12) in Sect. 1.2.
3. Use this estimate as the new fiducial values and train a new IMNN with these values.

This iterative process can be applied until the quasi-maximum likelihood estimate is relatively constant between iterations. It generally takes very few iterations for convergence. By iterating toward the quasi-maximum likelihood, more information about the model parameters can potentially be extracted. In our applications of Chap. 5 Sect. 4. and Chap. 6 Sect. 1., we do not need to apply this iterative method because our choice of fiducial values for the training of the network is close (or identical) to the inferred parameter values of the data.

Chapter 5

Likelihood-free Bayesian inference and validation of the method

You have the prior,
Make it known,
But with the likelihood,
It's still not set in stone.
But maybe there is a chance,
For your team to win!
And so the posterior may show you,
If and a celebration will begin.

`Irene Helenowski`

Abstract

We want to express the probability of the parameters of our forward model for one or two galaxy populations given some multiband deep field images. To do so, this chapter introduces the Bayesian framework with the well-known Bayes theorem to express the posterior distribution of our parameters given data images. Because the likelihood function of our forward model is not accessible, we therefore describe two accept-reject iterative procedures (namely approximate Bayesian computation, and population Monte Carlo) to derive such posterior distributions. This chapter also contains a first application where the whole procedure is applied to synthetic nonrealistic deep field images with only one spiral population and two parameters M^* and ϕ^*. This application demonstrates the succes of the methodology and that, by fitting our inception network, we are able to exhibit the posterior distribution of the two correlated parameters.

Contents

1. Bayes' theorem

The goal of this thesis is to obtain the probability of a set of luminosity functions' parameters given data images. To do so, we apply the Bayes theorem which relates the probability $\mathbb{P}(\theta|d)$ (called **the posterior** distribution) of a particular set of parameter values, θ, given data, d to the probability $\mathcal{L}(d|\theta)$ (that is **the likelihood** function of Chap. 4 Sect. 1.1) of d given θ:

$$\mathbb{P}(\theta|d) \propto \mathcal{L}(d|\theta)\mathbb{P}(\theta) \tag{5.1}$$

where $\mathbb{P}(\theta)$ is called **the prior** distribution of parameters θ. The prior represents the current knowledge about the parameters θ before the data d is available. For convenience the prior is often chosen from well-known and tractable distributions (e.g; uniform, gaussian, etc), allowing an easy evaluation of the prior probability of a given set of parameters as well as a fast computation of the generation of random parameters following such a distribution. In this thesis, we use a **uniform prior** with specific bounds that are chosen per parameter of our forward model using the range of luminosity function parameters measured in other redshift surveys.

2. Approximate Bayesian Computation (ABC) procedure

Because the likelihood function of our forward model is unknown, we can use **ABC** to bypass its evaluation by considering the density distribution of model parameters for which the generated forward simulations are "close" to some observed data; *i.e.* the distance between the two is small. The first ABC-related ideas date back to the 1980s (Rubin 1984) with a sampling method that asymptotically yields the posterior distribution. The **ABC-rejection sampling** in its most basic form generates N simulations, $\{\hat{d_i}(\theta_i)|i \in [1, N]\}$, from model parameters θ_i following the prior distribution $\mathbb{P}(\theta)$, and compares them to the observed data, d. Simulations that are distinctly different from the observed data d are considered unlikely to have been generated from the same parameter values describing d, and the associated simulation parameter values are rejected. In more precise terms, the sample $\hat{d_i}$ is accepted with tolerance $\epsilon \geq 0$ if

$$\rho(\hat{d_i}, d) \leq \epsilon, \tag{5.2}$$

where ρ is a distance that measures the discrepancy between $\hat{d_i}$ and d. The probability of generating a data set $\hat{d_i}$ with a small distance to d typically decreases as the dimensionality of the problem increases. A common approach to somewhat alleviate this issue is to replace $\hat{d_i}$ and d with a set of lower-dimensional summary statistics $\hat{t_i} = f\left(\hat{d_i}\right)$ and $t_{\text{obs}} = f(d)$, which are selected to capture the relevant information in $\hat{d_i}$ and d. The full relevance of Chap. 4 for obtaining the summary statistics thus becomes clear in this context, where f is the trained IMNN inception architecture. The acceptance criterion in the ABC rejection algorithm then becomes

$$\rho\left(\hat{t_i}, t_{\text{obs}}\right) = \rho\left(f\left(\hat{d_i}\right), f(d)\right) \leq \epsilon, \tag{5.3}$$

where the summary statistics are derived using the compression described in Chap. 4. In general, the form of the distance measure, ρ, can be difficult to choose for ABC (amounting to a similar interpretation as a choice of likelihood function). However, Charnock et al. (2018) show that in the IMNN approach, the distance used between a summary statistic of a simulation $\hat{t_i} = f\left(\hat{d_i}\right)$ and that of the observed data, $t_{\text{obs}} = f(d)$, is conveniently defined by the Fisher information of the previously trained network (Chap. 4) by

$$\rho\left(\hat{t_i}, t_{\text{obs}}\right) = \left(\widetilde{\theta}\left(\hat{t_i}\right) - \widetilde{\theta}\left(t_{\text{obs}}\right)\right)^T F(T, \theta_{\text{fid}}) \left(\widetilde{\theta}\left(\hat{t_i}\right) - \widetilde{\theta}\left(t_{\text{obs}}\right)\right) \tag{5.4}$$

This distance measure is justified with the quasi-maximum likelihood estimator $\widetilde{\theta}$ described in Eq. (4.12) of Chap. 4 Sect. 1.2 because that space should be asymptotically Euclidean (with the Fisher information as the metric) once the network has converged and the summary statistics become Gaussianly distributed (Charnock et al. 2018). Even when using summary statistics, the acceptance rate of the ABC is low for a small ϵ because we sample from the whole multidimensional prior distribution. This means that finding a new set of parameters whose simulations are more similar to the data is unlikely, which makes the sampling from the prior distribution inefficient and hence slow.

3. Population Monte Carlo (PMC) procedure

The **PMC** framework introduced in Cappé et al. (2004) falls within the family of the **adaptive importance sampling** methods and uses resampling and weighting steps to adapt the location parameters of the iterative proposal distributions. PMC is a submethod of ABC (often called "PMC-ABC") and its goal is to have a higher density of draws in the more probable regions of the posterior distribution. The different steps of the PMC algorithm we use in this thesis are as follows:

- **Step 1:** From the prior distribution, we draw N (user-defined) sets of parameters. For each set of parameters in this sample, we follow the steps listed below.
    - ⋆ **Step 1a:** We simulate the multiband images.
    - ⋆ **Step 1b:** We compress the simulated images with the trained IMNN to reduce the dimensional complexity.
    - ⋆ **Step 1c:** We compute the distance between the summarized simulation and the summary of the observed data using Eq. (5.4).
    - ⋆ **Step 1d:** We define a weighting that corresponds to the probability of this set of parameters under the prior distribution.
- **Step 2:** We compute the mean vector and the weighted-covariance matrix of the N sets of parameters. We set the R counter to 0.
- **Step 3:** From the N sets of parameters, we identify the $Q\%$ (user-defined) of the compressed simulations that have the largest distance from the summary of the observed data. For each set of parameters in this subsample, we follow the steps listed below.
    - ⋆ **Step 3a:** We resample it from a proposal distribution that is a Gaussian using the current parameter values as the mean and the weighted-covariance matrix from Step 2. Each time we go through this step, we increment the R counter.
    - ⋆ **Step 3b:** We simulate the multiband images at this new proposed set of parameters.
    - ⋆ **Step 3c:** We compress the simulated image with IMNN to reduce the dimensional complexity.
    - ⋆ **Step 3d:** We compute the distance between the summarized simulation and the summary of the observed data using Eq. (5.4).
    - ⋆ **Step 3e:** If this distance is smaller than the previous distance, we keep this set of parameters. Otherwise, we return to **Step 3a**. After 100 subsequent times through **Steps 3a** to **3e**, we go back to **Step 2** to avoid trying to redraw very few parameters sets for an unreasonable long time.
- **Step 4:** Each weight is updated using the probability of the corresponding resampled set of calculated parameters and the weighted-covariance matrix of Step 2. Each initial weight (under the prior distribution) is divided by the normalized Gaussian using the difference between the new and initial parameter set as the mean and the weighted-covariance matrix from Step 2.
- **Step 5:** As long as the counter R is smaller than the user-defined threshold M, we return to **Step 2** and re-identify the $Q\%$ simulations with the largest distance from the observed data. Otherwise, the PMC concludes.

The $Q\%$ in Step 3 is a user choice, and the chosen value allows us to efficiently parallelize the procedure. Moreover, taking $Q = 100\%$ is not very efficient because it means that at each iteration the whole N parameter sets will be redrawn, including the parameter sets which give simulations that are very close to the observed data (i.e. those parameters will be very hard to redraw in order to find even closer simulations). On the other side, taking Q too small is not very efficient because the procedure will be slow to converge. In this thesis, we use the values $Q = 25\%$ or $Q = 75\%$ as intermediate values. The value of M is another user choice which gives the number of draws to achieve during the same iteration for the PMC procedure to conclude. If M is large enough, the distribution of the N sets of parameters becomes approximately stationary and the procedure stops, and this distribution can then be considered as a good approximation of the posterior distribution (Del Moral et al. 2012). We take a value of M inferior to $100 \times N \times Q\%$ and we choose $\frac{2}{3}$ of this value: this means that when the PMC stops, on average, at least $\frac{2}{3}$ of the parameter sets need more than 100 redraws each.

4. First application: 1 population and 2 parameters

Here we validate the methodology developed above by applying it to a nonrealistic but simple set of synthetic multiband deep field images containing only one population of spiral galaxies described by only two parameters for the luminosity function, M^* and ϕ^*.

4.1 Description

Only one spiral population of galaxies is modeled with two "free" parameters of the luminosity function, M^* and ϕ^*, while the slope is fixed to $\alpha = -1.3$ and there is no redshift evolution in the luminosity function, $M^*_{\mathrm{evol}} = 0$ and $\phi^*_{\mathrm{evol}} = 0$. However, the full description of the forward model of Chap. 3 still applies, particularly the bulge and disk redshift evolutions described in Chap. 3 Sects. 2.2 and 2.3 are used. Our spiral population is defined with a bulge-to-total ratio $B/T = 0.2$, with the 'E' SED for the bulge and the 'Irr' SED for the disk, see Fig. 3.1. As explained in Chap. 4 Sect. 2.6, the following fiducial parameter values and their offsets for the numerical derivatives are used to fit the network:

$$\begin{array}{lclcl} \theta_{\mathrm{fid},1} & = & M^* & = & -20 \\ \Delta\theta_1 & = & \Delta M^* & = & 0.5 \\ \theta_{\mathrm{fid},2} & = & \log_{10}(\phi^*) & = & -2.01 \\ \Delta\theta_2 & = & \Delta\log_{10}(\phi^*) & = & 0.2, \end{array} \tag{5.5}$$

where ϕ^* and M^* are given for $H_0 = 100$ km s^{-1} Mpc^{-1}. Table 9 gives an overview of the values that are used to generate the simulations according to the description of the forward model of Chap. 3.

Fig. 5.1 shows the five theoretical luminosity functions with fiducial/perturbed parameters of Eq. (5.5) and fixed values $\alpha = -1.3$, $M^*_{\mathrm{evol}} = 0$ and $\phi^*_{\mathrm{evol}} = 0$. Fig. 5.2 shows the effect on the RGB image (using the i', r', g' filters) of a simulation when only one of the two parameters M^* or ϕ^* is changed. The number density of spirals is boosted for the $\Delta\log_{10}(\phi^*)$ increase in $\log_{10}(\phi^*)$ and for the ΔM^* decrease in M^*. Consequently, a strong correlation between these two parameters is expected in this application.

4.2 Training of the network

To fit the weights and biases of the inception network, we follow the prescription described in Chap. 4 Sect. 2.6 and $n = 200$ deep field 1024×1024 pixel images are generated (in the eight photometric bands u', g', r', i', z', J, H, Ks) with parameters at fiducial values of Eq. (5.5). For the multiband deep field images used for the numerical derivatives of Eq. (4.29), $m = 50$ simulations are generated (in each of the eight bands) for each parameter, that is, M^* and $\log_{10}(\phi^*)$, at values $\theta_{\mathrm{fid},\gamma} + \Delta\theta_\gamma$, $\gamma = 1, 2$, and another $m = 50$ simulations per parameter at values $\theta_{\mathrm{fid},\gamma} - \Delta\theta_\gamma$, $\gamma = 1, 2$, see Eq. (5.5). This yields a total of $200 + 2 \times 2 \times 50 = 400$ deep field images (in the eight bands) for the training set, and the same number is used for the validation set. The optimizer techniques and the corresponding learning rates used during the training are listed in Table 5. The Adam optimizer is applied for the first 100 epochs to quickly find a transformation that forces the covariance matrix of the summaries to be approximately the identity matrix, then we switch to the SGD momentum technique. For epochs 100-200 the learning rate was set to 1×10^{-6} and because the training was relatively stable, we decided to increase the learning rate to 3×10^{-6} for epochs 200-600 to boost the procedure. For the remaining epochs, the learning rate is decreased from time to time in order to explore the more complex landscape of the parameter-dependent Fisher information.

The architecture of the network is that of Fig. 4.6, which yields a total of 12,170 network parameters to be fitted. Because of the number and size of the images, within this configuration, one epoch of training (*i.e.* when the whole training and validation sets are passed through the network and the network weights are updated) takes about 40 seconds to run (on a NVIDIA QUADRO RTX 8000 45GB GPU) with a total of 6.5 days of training.

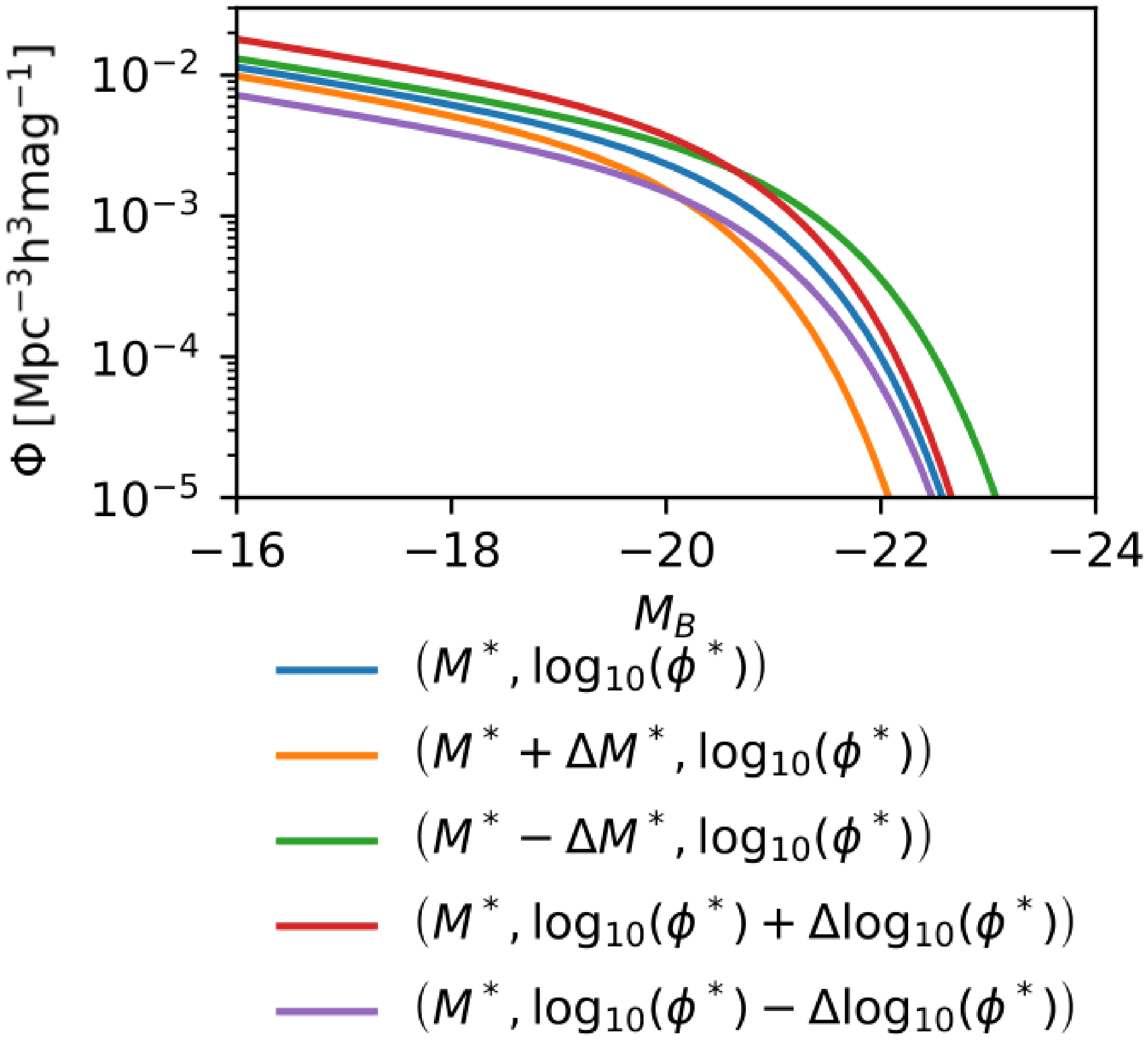


Fig. 5.1: Theoretical luminosity functions of the spiral population with the perturbed values of the parameters used in the first application of Sect. 4. The parameters used for each curve are listed in Eq. (5.5) and in Table 9. The explicit features in the data, generated from these functions, provide the differences that the IMNN will become sensitive to when training, therefore mapping the effect of $\log_{10}(\phi^*)$ and M^* to informative summaries.

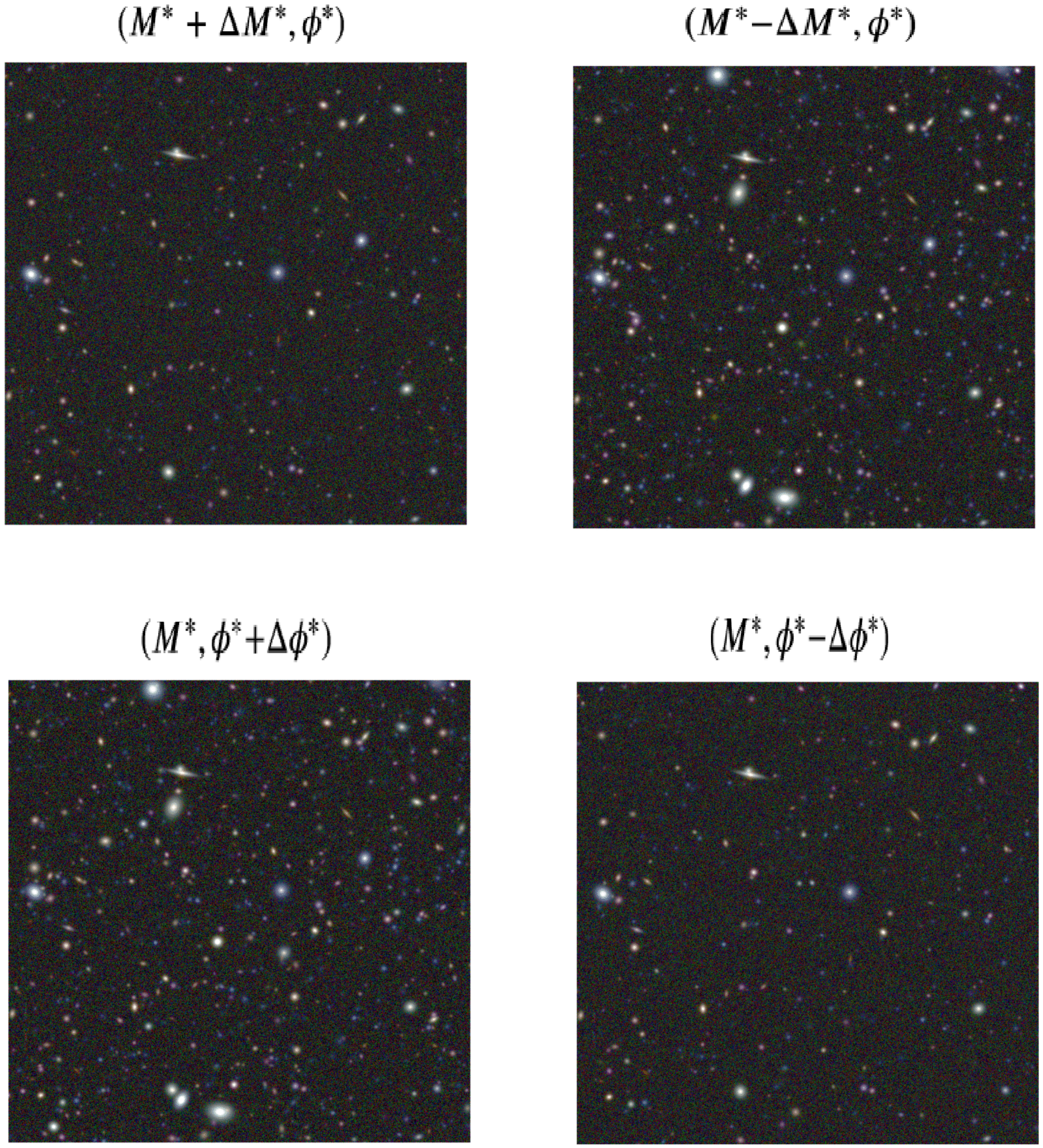


Fig. 5.2: Color images (g', r', i' filters) showing the effect of the perturbed values from fiducial for each parameter, as listed in Eq. (5.5). For each subplot we only added or removed the offset for one parameter listed in Eq. (5.5) and kept the other at its fiducial value. Decreasing the value of M^* (top right) increases the number of galaxies and decreasing the value of $\log_{10}(\phi^*)$ (bottom right) decreases the number of galaxies. This shows that these two parameters are highly correlated.

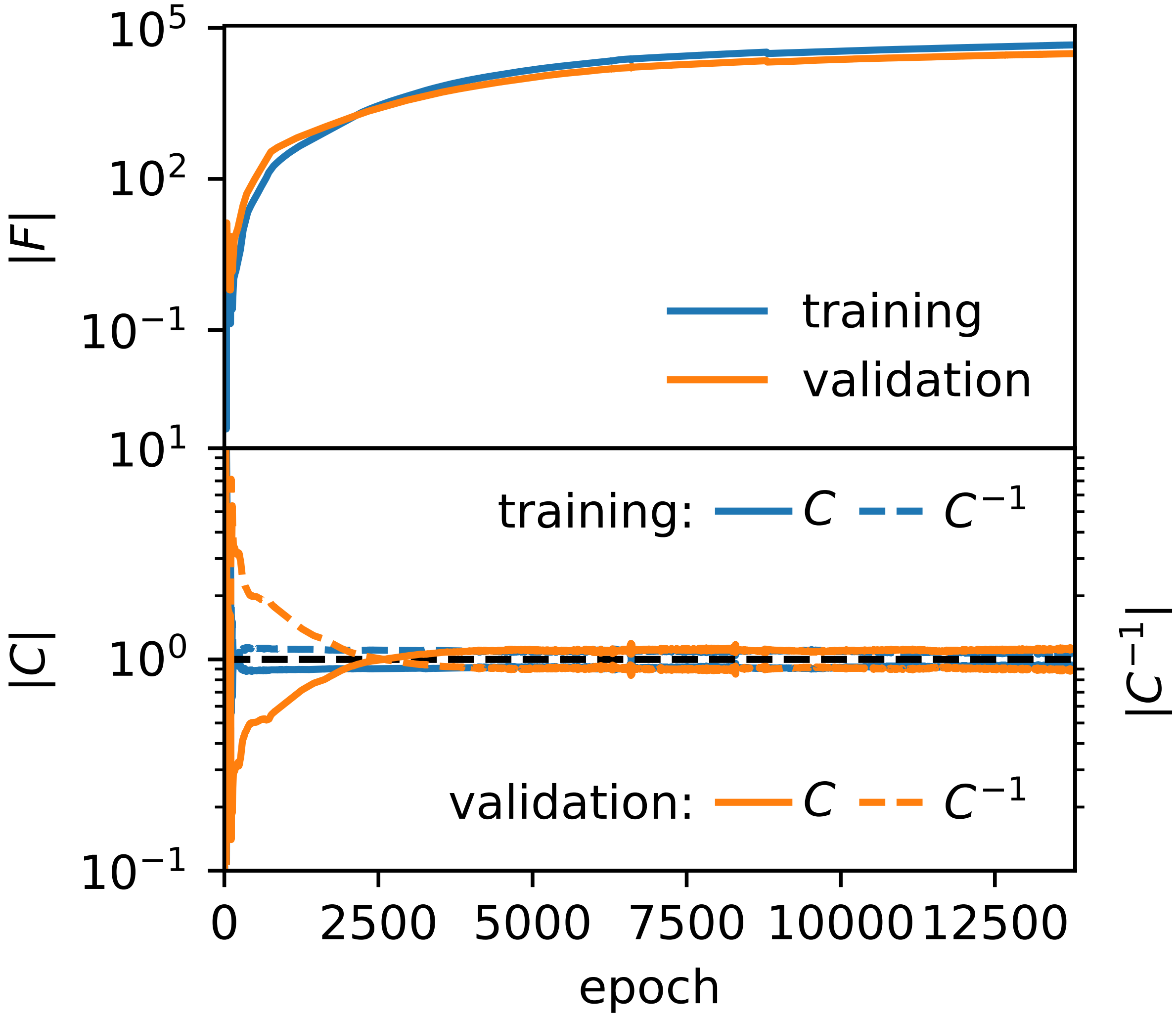


Fig. 5.3: *Top:* Evolution of the determinant of the Fisher information matrix during almost 14000 epochs. The blue curve represents the information extracted from the training set, and the orange curve the information from the validation set. The determinant of the Fisher information is maximized on the training set and is comparable to the determinant of the Fisher information calculated with the validation set. This confirms that the two sets are representative samples for our inference problem of the ϕ^* and M^* parameters of the spiral population. The training is stopped when the Fisher information of the validation set becomes relatively constant. *Bottom:* Evolution of the determinants of the covariance matrix (solid line) and of the inverse of the covariance matrix (dashed line) for the training set (blue) and for the validation set (orange). The values quickly become constant, which shows that the network suppresses the parameter dependence of the covariance. Because the correlation between the parameters is strong, the covariance matrix cannot exactly diagonalize to the identity matrix, but the fact that it is constant shows that the parameter dependence is weak, which is the only requirement for the IMNN.

Table 5: Optimizer and learning rate used during the training of Chap. 5 Sect. 4..

epoch	optimizer	learning rate
0	Adam	0.02
100	SGD	1×10^{-6}
200	SGD	3×10^{-6}
600	SGD	2×10^{-6}
6300	SGD	1.5×10^{-6}
8000	SGD	1×10^{-6}
8500	SGD	7.5×10^{-7}
13300	SGD	5.5×10^{-7}

Note: SGD stands for stochastic gradient descent.

The network was trained for almost 14000 epochs and Fig. 5.3 shows the evolution of the determinant of the Fisher information matrix (top) and the determinants of the covariance and inverse covariance matrices (bottom). The determinant of the Fisher information matrix increases during the training, and the determinant of the covariance matrix and its inverse become constant after a few hundred epochs. This shows that the covariance has very little parameter dependence, and the sensitivity in the network instead is due to the derivative of the mean of the summaries with respect to the parameters. The determinant of the covariance matrix (and its inverse) are not precisely unity because the M^* and ϕ^* parameters are extremely correlated (a larger number of galaxies can be obtained either by increasing the value of $\log_{10}(\phi^*)$ or by decreasing the value of M^*), which is clearly seen in Fig. 5.2. However, the only requirement for the IMNN is that the covariance matrix does not depend on the parameter values (*i.e.* $\nabla_\theta C = 0$ and $\nabla_\theta C^{-1} = 0$), at least for parameters close to the fiducial choice, which is the case as seen by the relatively constant values for the determinants of C and C^{-1} after 2500 epochs in Fig. 5.3. After 5000 epochs, the determinant of the Fisher information begins to flatten (in the training and validation data), and the training is stopped at epoch around 14000. The Fisher matrix at the end of the training is the amount of information we are willing to extract from the data given the training time, the data sets and the network architecture.

4.3 ABC posterior

After the network is trained, an ABC procedure is run to test the reliability of the method, see Sect. 2. for more details. Here, we replaced what is referred to as "observed data" in Sect. 2. with a simulation called "virtual data", generated with parameters set to the fiducial parameter values given in Eq. (5.5), and used to train the network:

$$\begin{aligned} \theta_{\mathrm{vir},1} &= M^* = -20 \\ \theta_{\mathrm{vir},2} &= \log_{10}(\phi^*) = -2.01 \end{aligned} \tag{5.6}$$

where ϕ^* and M^* are given for $H_0 = 100$ km s^{-1} Mpc^{-1} and $z = 0$. The formalism of Sect. 2. is applied and a uniform prior is chosen for the two parameters,

$$\begin{aligned} -24 \leq & \; M^* \leq -16 \\ -4 \leq & \; \log_{10}(\phi^*) \leq -0.5 \end{aligned} \tag{5.7}$$

From these prior intervals, an initial sample of 5000 pairs of parameters is drawn and a simulated multiband deep field (using the forward model of Chap. 3) is generated for each pair. Then, both the virtual data and the 5000 simulated fields are passed through the network in order to compress them into the most informative summary statistics. The bottom left plot of Fig. 5.4 shows the values of these parameters for the entire 5000 prior simulations in orange, and in blue, for only the 50 simulations for which the distance to the virtual data is minimal. The dashed black lines are the parameter values used to generate the virtual data. The ABC procedure retrieves the expected strong correlation between the two parameters M^* and ϕ^*. The upper and right pannels show the marginal 1D distributions for each individual parameter. However, because we do not know which value of ϵ should be used, the posterior is not properly approximated here and

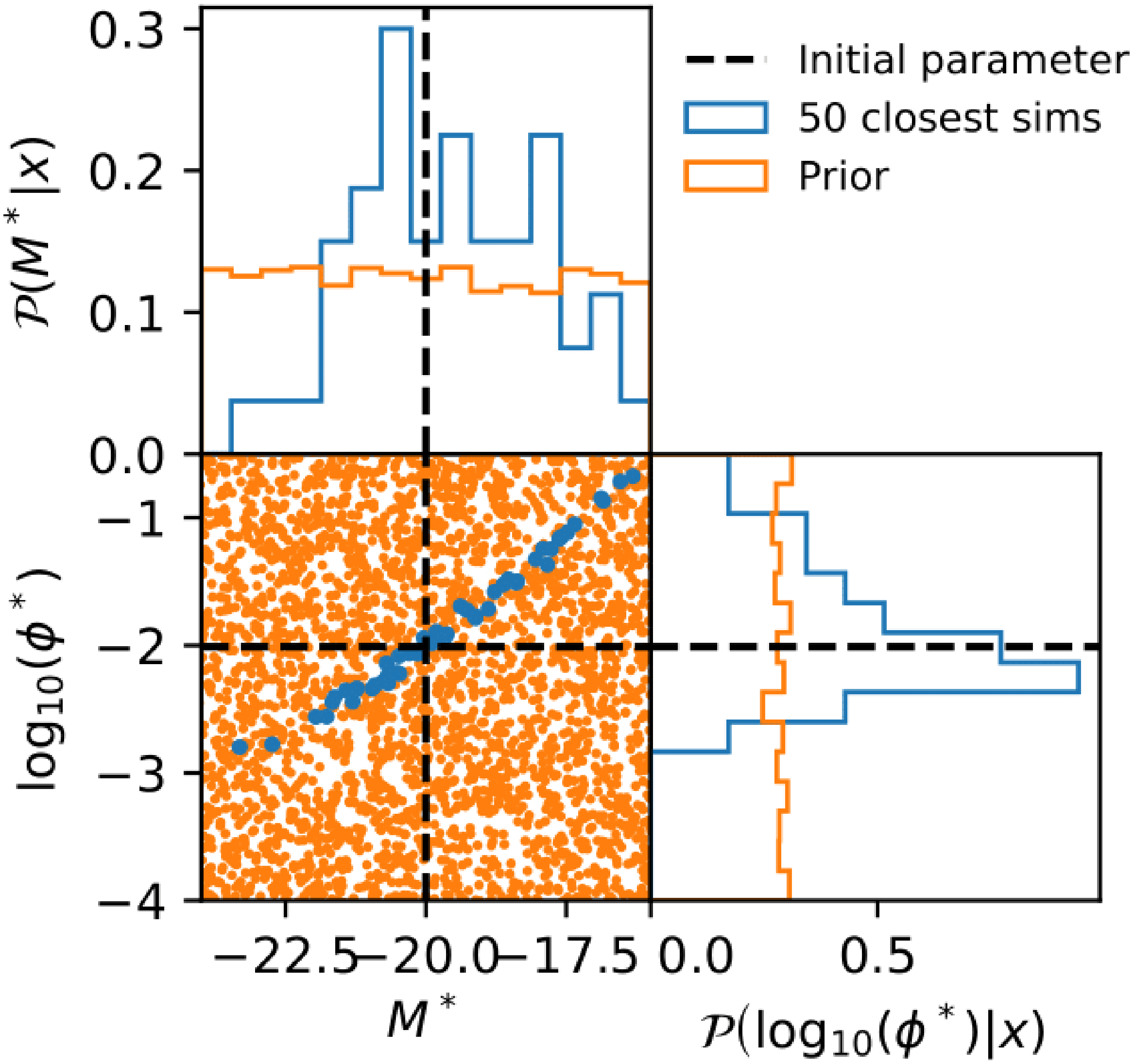


Fig. 5.4: Results of the ABC procedure. *Bottom left:* Values of the parameters for the 5000 prior simulations (orange) and for the 50 simulations with minimum distance (blue) to the virtual data. The dotted black lines are the parameter values used to generate the virtual data. We retrieve the strong correlation between the two parameters. *Top left and bottom right:* 1D marginal distributions of the parameter values.

it would take a prohibitive amount of time to obtain a correct posterior distribution, therefore the population Monte Carlo (PMC) procedure of Chap. 5 Sect. 3. is applied to in the next Section to improve the posterior determination.

4.4 PMC posterior

For the PMC, the same prior is used for the two parameters as in Eq. (5.7), and the PMC procedure of Chap. 5 Sect. 3. is applied with $N = 1000$, $M = 50000$, and $Q = 75\%$ for resampling the N sets of parameters. At each iteration, the PMC procedure re-draws 750 samples and concludes when the number of attempts in the same iteration is higher than 50000. This yields 232577 draws and generated simulations in total and the PMC procedure lasted $\approx$ 2 days. The results are shown in Fig. 5.5. The 2D posterior distribution (bottom left) and the 1D marginal distributions are concentrated around the values of the parameters used to generate the virtual data (*i.e.* $M^* = -20$ and $\log_{10}(\phi^*) = -2.01$). The procedure also identifies the strong correlation regarding the number density of objects. The red dashed lines show the values of the parameters used to generate the virtual data. The PMC is therefore able to infer that the parameter values used to generate the virtual data lie in the 2D region of the posterior distribution with the highest density. This first application illustrates that the network is able to summarize the effects in the images of the two correlated parameters, and that it is able to robustly constrain the possible values of the parameters used to generate the virtual data.

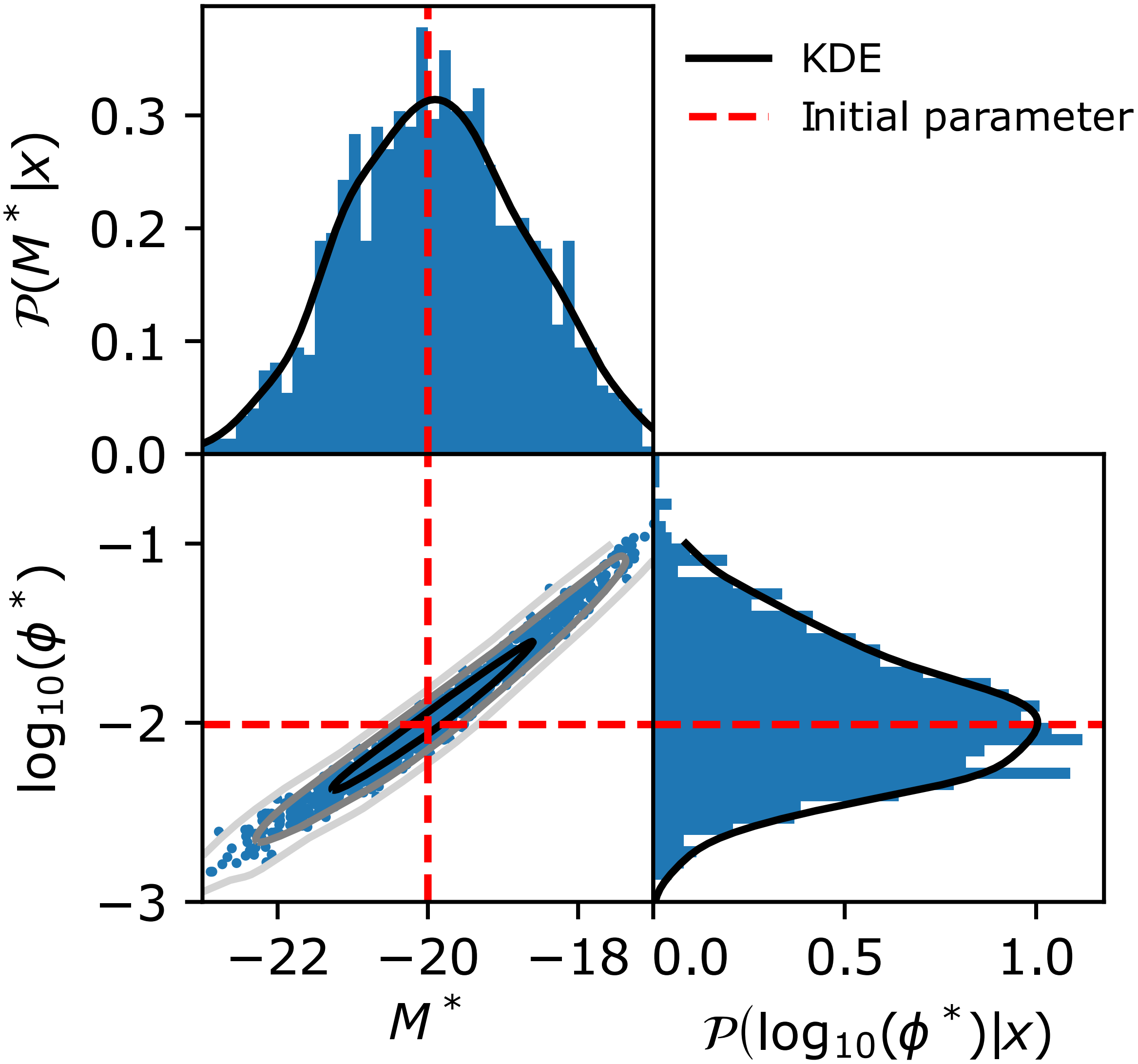


Fig. 5.5: Results of the PMC procedure. *Bottom left:* Distribution of the parameter values for the final 1000 points obtained by the PMC. The black, dark gray and light gray show the 68%, 95%, and 99% contours, and the dotted red line shows the values of the parameters used to generate the virtual data. *Top left and bottom right:* 1D marginal distributions of the parameters and their kernel density estimates (KDE) in black. The procedure converges around the parameter values used to generate the virtual data. There is evidence here, in 2D, that the uncertainty due to the correlation is large between the parameters whose effect is clear in the data, see Fig. 5.2.

Chapter 6

Inference on the CFHTLS D1 deep field

Thought of result is very scary,
It might bring tears or shower cherry !!

Whatever result may reveal,
Failure or Success,
But feel yourself blessed !!

No result is an ending,
Just a new beginning !! (8-14)

`Mansi Kotecha`

Abstract

This chapter aims at inferring the parameters of our forward model for the elliptical and spiral populations simultaneously from the eight photometric bands of the CFHTLS D1 deep field images. We describe first the procedure and the results when only the density parameter ϕ_0^* is considered for these two populations. We demonstrate the accuracy of the method and we show that our results are fully compatible with other studies which have applied the classic catalog-extraction method. The third application shows preliminary results when the procedure is applied with a total of six parameters: ϕ_0^*, M_0^* and α for the elliptical and spiral populations simultaneously.

Contents

1. Second application: 2 populations and 2 parameters

The main objective of this thesis is to infer the parameters of the luminosity function for several populations of galaxies simultaneously on real observed multiband deep field images, and specifically here for the CFHTLS D1 deep field. Therefore, a more realistic model than the previous one (see Chap. 5 Sect. 4.) is considered, with, here, two populations of galaxies, ellipticals and spirals and only the density parameter ϕ_0^* of the luminosity functions is studied simultaneously for these two populations. The other parameters of the luminosity functions are fixed following the recent results of López-Sanjuan et al. (2017) in the Alhambra survey where they provided a deep statistical analysis of the luminosity function parameters for these two populations of galaxies.

1.1 Description

In this application, using the results of López-Sanjuan et al. (2017), the following fiducial parameter values and offsets for the numerical derivatives are used to train the network:

$$\begin{array}{ccccc|ccccc} \multicolumn{5}{c|}{\text{Elliptical}} & \multicolumn{5}{c}{\text{Spiral}} \\ \hline \theta_{\mathrm{fid},1} & = & \log_{10}\left(\phi^*_{0,E}\right) & = & -2.09 & \theta_{\mathrm{fid},2} & = & \log_{10}\left(\phi^*_{0,Sp}\right) & = & -2.04 \\ \Delta\theta_1 & = & \Delta\log_{10}\left(\phi^*_{0,E}\right) & = & 0.5 & \Delta\theta_2 & = & \Delta\log_{10}\left(\phi^*_{0,Sp}\right) & = & 0.05 \end{array} \tag{6.1}$$

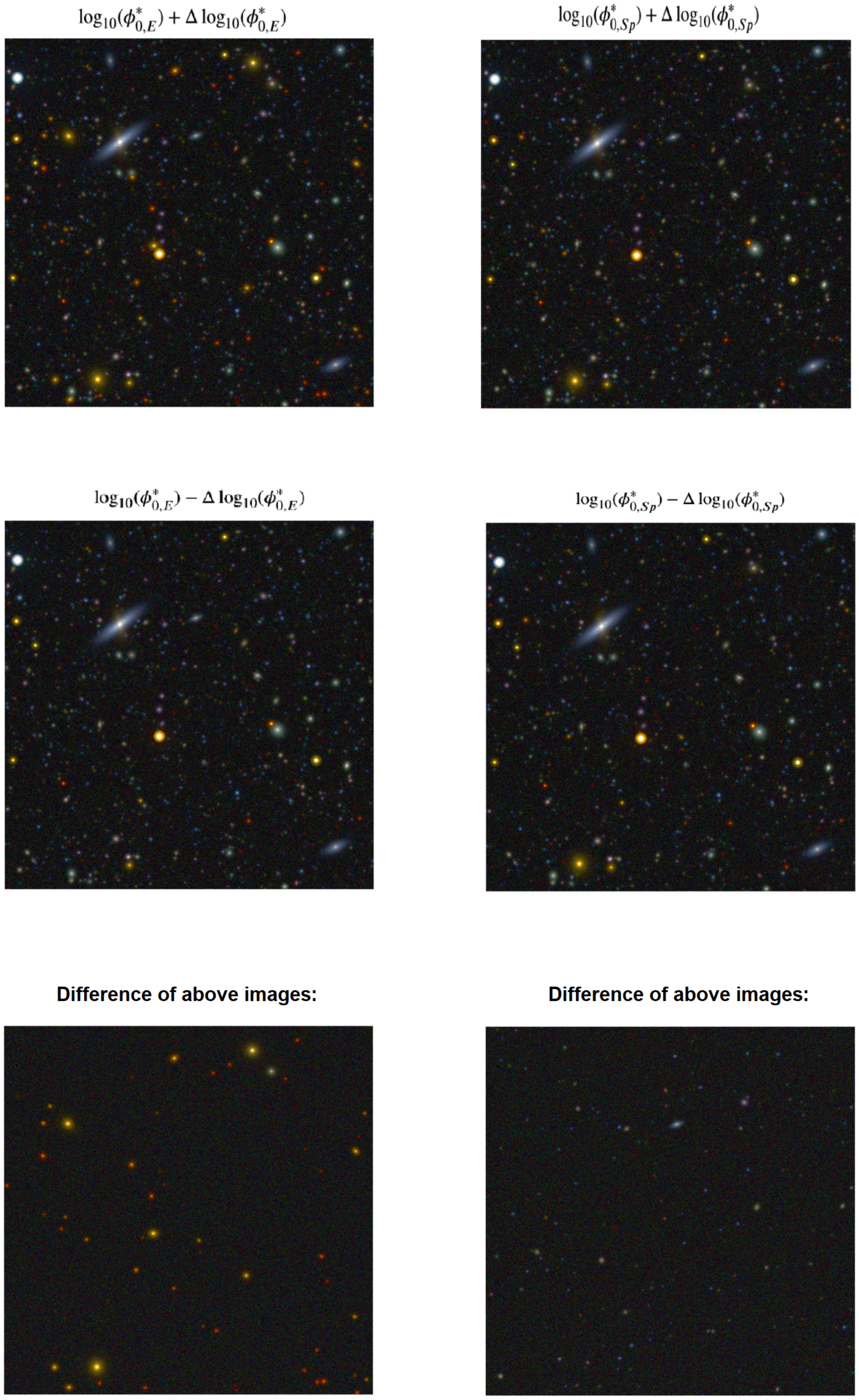
where ϕ^* is given for $H_0 = 100$ km s^{-1} Mpc^{-1} and $z = 0$. The other luminosity function parameters are fixed with the values from López-Sanjuan et al. (2017), see Table 9. The offset value of $\phi^*_{0,\mathrm{Sp}}$ of Eq. (6.1) is a factor of 10 smaller for the spiral galaxies than $\phi^*_{0,\mathrm{E}}$ for the ellipticals when the simulations for the numerical derivatives are generated because the number density of spirals is higher than that of ellipticals, as seen in Fig. 6.2. Because, in this thesis, the reference SDSS g' band for the absolute magnitude is used, we therefore apply a correction of $B - g' = 0.78$ for the elliptical population and $B - g' = 0.40$ for the spiral population, as suggested by Table 7 of Fukugita et al. (1995), to express the magnitudes in the reference B band as in López-Sanjuan et al. (2017).

For computational efficiency, only 3.17 arcmin^2 images (corresponding to 1024×1024 pixels of 0.186 arcsec) of the full 1 deg^2 D1 deep field are used in the eight photometric bands u', g', r', i', z', J, H, Ks in order to infer the values of ϕ_0^* for the elliptical and spiral populations. Fig. 6.1 shows the effects on a RGB image of the g', r', i' simulations for the elliptical population (left column) and the spiral population (right column) of changing the fiducial parameters by their respective positive (top images) and negative (middle images) offset from the fiducials that are used to calculate the numerical derivatives. The bottom image of each column is the difference of the above images and shows the galaxies that appear in the images used to calculate the numerical derivatives of the summaries using the IMMN. The bottom left panel shows that only elliptical galaxies remain when only ϕ_0^* for the elliptical is changed (these galaxies appear as yellow or red because of the SED template 'E' used to model them), as opposed to the blue galaxies (modeled as a bulge with the 'E' SED and a disk with the 'Irr' SED) of the spiral population remaining in the bottom right panel. Table 9 gives an overview of the values that are used to generate the simulations according to the description of the forward model of Chap. 3.

Fig. 6.2 shows the six theoretical luminosity functions, at $z = 0$ and $z = 2$, used to train the network. The number density of bright red ellipticals is lower at $z = 2$ than at $z = 0$: fewer galaxies with early-type SEDs existed in the past than today because they were still forming their stars and therefore had bluer SEDs. Moreover, a relatively higher number density of faint spirals is expected compared to ellipticals at both $z = 0$ and $z = 2$. The curves for the perturbed luminosity functions of spirals are closer together because the offset $\Delta\theta$ is a factor 10 lower than for the ellipticals.

1.2 Training of the network

As in the previous application of Chap. 5 Sect. 4.2, simulations are generated to train the network in the same way, with $n = 200$ fiducial images of 3.17 arcmin^2 and 1024×1024 pixels in the eight photometric bands, and $2 \times 2 \times 50 = 200$ ($m = 50$) seed-matched images for the numerical derivatives. The same inception architecture is used with the Adam optimizer (learning rate of 0.04) for the first 4000 epochs of the training, then the stochastic gradient descent optimizer (learning rate of 5×10^{-6}) for the remaining epochs. The



$\log_{10}(\phi^*_{0,E}) + \Delta\log_{10}(\phi^*_{0,E})$

$\log_{10}(\phi^*_{0,Sp}) + \Delta\log_{10}(\phi^*_{0,Sp})$

$\log_{10}(\phi^*_{0,E}) - \Delta\log_{10}(\phi^*_{0,E})$

$\log_{10}(\phi^*_{0,Sp}) - \Delta\log_{10}(\phi^*_{0,Sp})$

Difference of above images:

Difference of above images:

Fig. 6.1: *Left:* Difference of the two upper images is shown in the bottom panel. The two upper images are simulated by only changing the fiducial parameter of the elliptical population to the perturbed parameter values by $+\Delta\theta_1$ (top) and $-\Delta\theta_1$ (middle). Only yellow or red elliptical galaxies remain, which confirms that $\phi^*_{0,E}$ affects the number density of ellipticals in the simulations. *Right:* Difference of the two upper images is shown in the bottom panel. The two upper images are simulated by only changing the fiducial parameter of the spiral population to the perturbed parameter values by $+\Delta\theta_2$ (top) and $-\Delta\theta_2$ (middle). Only blue spiral galaxies remain, which confirms that $\phi^*_{0,Sp}$ affects the number density of spirals in the simulations. These RGB color images are obtained from the i', r', g' filters.

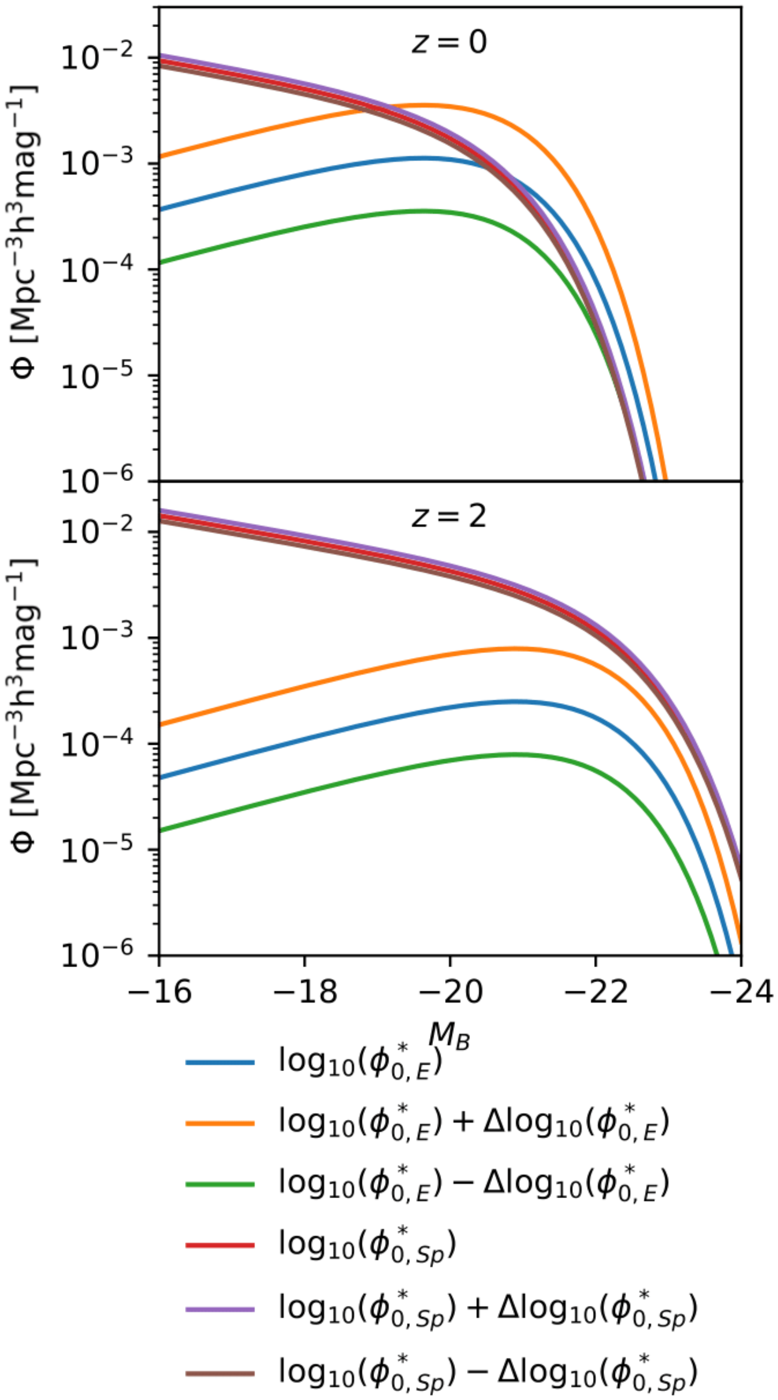


Fig. 6.2: Theoretical luminosity functions of the elliptical (blue, orange, and green) and spiral (red, purple, and brown) populations at redshift $z = 0$ (top) and $z = 2$ (bottom). The legend applies to both panels. The parameters used for each luminosity function are given in Table 9. These curves show a higher integrated number density of spiral galaxies than of elliptical galaxies. The steep faint-end slope of the spiral population implies that the images contain many faint spiral galaxies, which is not the case for the elliptical population. There is also a redshift effect that decreases the proportion of ellipticals or spirals when looking at distant objects.

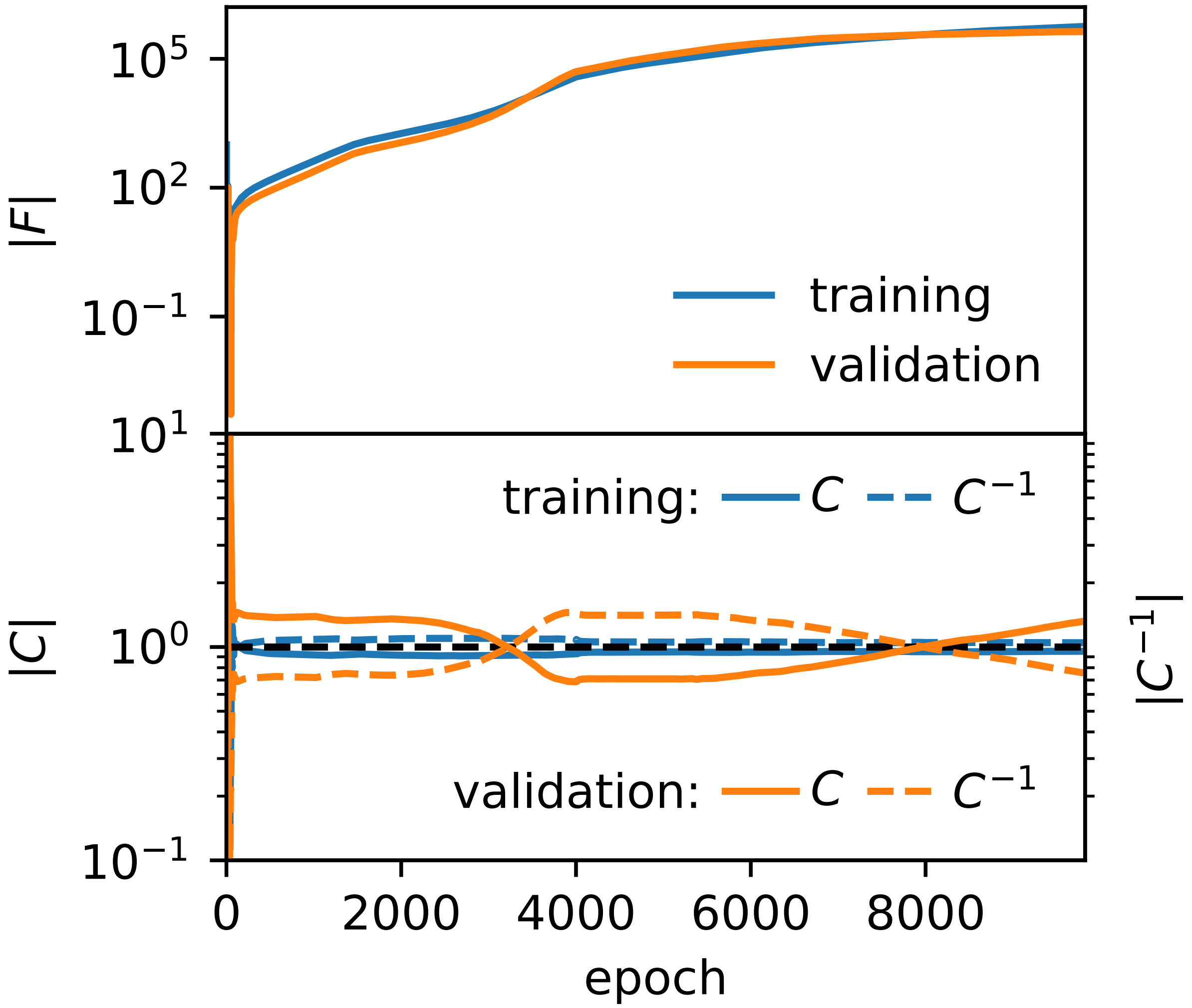


Fig. 6.3: *Top:* Evolution of the determinant of the Fisher information matrix during almost 10 000 epochs. The blue curve represents the information extracted from the training set, and the orange curve shows the same from the validation set. The curves increase, which means that the network learns more about the two parameters as the training continues. The training was stopped when the validation curve flattened, suggesting that the network is close to convergence. *Bottom:* Evolution of the determinant of the covariance matrix (solid line) and the inverse of the covariance matrix (dashed line). The blue curves show the training set, and the orange curves show the validation set. The training curves reach 1 very fast, which shows that the loss is stabilized and that the magnitude of the summaries is under control. The validation curve oscillates while being still very close to identity, which is a sign that there is some weak parameter dependence on the covariance.

network was trained on the GPU NVIDIA TITAN X 12Go for almost 10 000 epochs ($\approx$ 8 days), and Fig. 6.3 shows the convergence of the determinant of the Fisher matrix (top) and the determinant of the covariance and inverse covariance matrices (bottom) for the training set (blue) and the validation set (orange). During the fitting, the training and validation curves remain close to each other, which indicates that the training and validation sets contain approximately the same amount of information about the two parameters, $\phi^*_{0,E}$ and $\phi^*_{0,Sp}$. Consequently, both sets seem representative enough of the diversity of realizations of a deep field of spiral and elliptical galaxies with these fiducial parameters.

1.3 Observed and virtual data

Our ultimate goal is to infer the parameters on the 1 deg^2 D1 deep field. However, the analysis is limited to deep fields of size 3.17 arcmin^2 (*i.e.* 1024×1024 pixels) for computational efficiency (memory limitation), and five such regions of the D1 are randomly chosen, see Fig. 6.5. In order to confirm that the network is trained properly and performs well, five simulations, called “virtual data”, are also generated with the fiducial parameters of Table 9 and the same angular size as the CFHTLS data (see Fig. 6.4). The goal is to obtain the individual posterior distributions of the five observed data images and to compute their joint posterior distribution to constrain the values of the parameters and extract the confidence intervals as explained in Sect. 1.6. The same procedure is applied to the five virtual data with the only difference we know in advance the values of the parameters that are used to generate the virtual data. Five images in both cases correspond to a relatively good compromise between the computational time and the accuracy of the joint posterior, see Sect. 1.6.

1.4 ABC posterior

The ABC procedure of Chap. 5 Sect. 2. is applied and a uniform prior distribution is used with the same lower bound for the amplitudes of the luminosity functions of the elliptical and spirals galaxies, but different upper bounds:

$$\begin{array}{ccccc|ccccc} \multicolumn{5}{c|}{\text{Elliptical}} & \multicolumn{5}{c}{\text{Spiral}} \\ \hline -4 & \leq & \log_{10}(\phi^*_{0,\mathrm{E}}) & \leq & 0.5 & -4 & \leq & \log_{10}(\phi^*_{0,\mathrm{Sp}}) & \leq & -1 \end{array} \tag{6.2}$$

The choice of a smaller upper bound of the prior for the spirals is made because the luminosity function for this population has a steep faint end slope (see Table 9 and Fig. 6.2), and it can lead to very long generation times of the simulations with high values of the amplitude $\phi^*_{0,\mathrm{Sp}}$. Moreover, previous studies such as López-Sanjuan et al. (2017) have shown that $\log_{10}(\phi^*_{0,\mathrm{Sp}}) > -1$ is very unlikely, and indeed, this is what we also find from Fig. 6.8. We wish to note that cutting possible regions of parameter space due to computational resources is not a statistically rigorous process, but the information from López-Sanjuan et al. (2017) gives us reason to believe that it is acceptable to truncate here.

Fig. 6.6 shows the results of the ABC procedure with 5000 draws of the parameters following the uniform prior of Eq. (6.2). A distance threshold ϵ can be chosen from the data to accept or reject some of the parameters, as shown by the green, blue, and red points in the bottom left plot. The number of accepted points decreases when this threshold is low, but their 2D region also becomes more concentrated. As a consequence, the corresponding 1D marginalized distribution of $\log_{10}(\phi^*_{0,E})$ (top left) varies significantly depending on this choice for a distance threshold. This is caused by the difficulty to densely sample from a stationary distribution, a problem that can occur when sparsely drawing from the prior and not choosing a small enough $\epsilon-$ball in which to accept simulations. In order to properly approximate the posterior, the distance threshold must approach $\epsilon = 0$ and more simulations need to be done for the ABC until a steady distribution is reached. This highlights the need for a PMC procedure that automatically approaches the posterior.

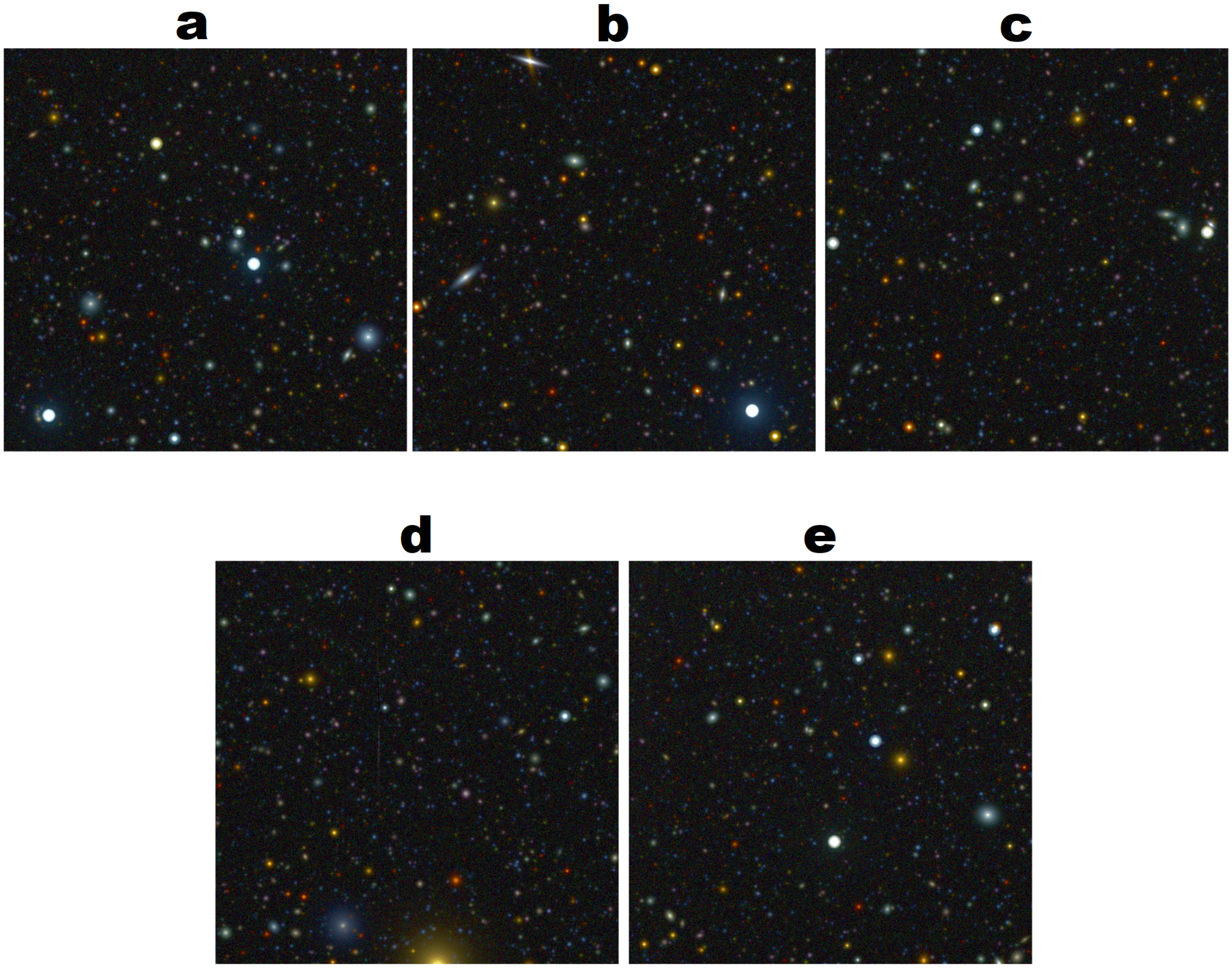


Fig. 6.4: RGB images using the g', r', i' filters for the five virtual data of 3.17 arcmin^2 generated with fiducial values of the luminosity function parameters (see Table 9), used to validate the method.

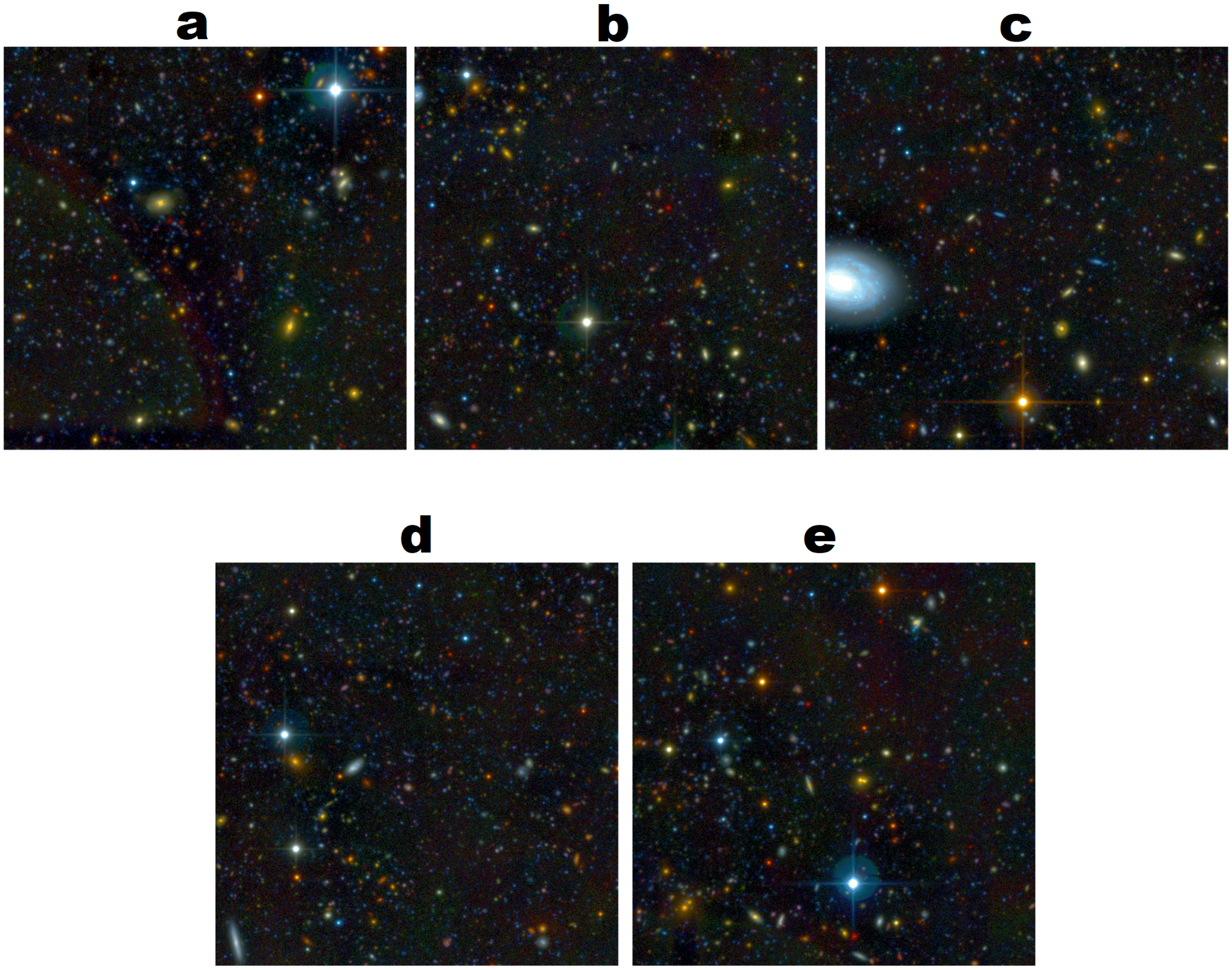


Fig. 6.5: RGB images using the g', r', i' filters for five random regions of 3.17 arcmin^2 within the 1 deg^2 CFHTLS D1 deep field that are used to infer the luminosity function parameters of the elliptical and spiral galaxies, namely the logarithm of their amplitudes $\log_{10}(\phi^*_{0,\mathrm{E}})$ and $\log_{10}(\phi^*_{0,\mathrm{Sp}})$.

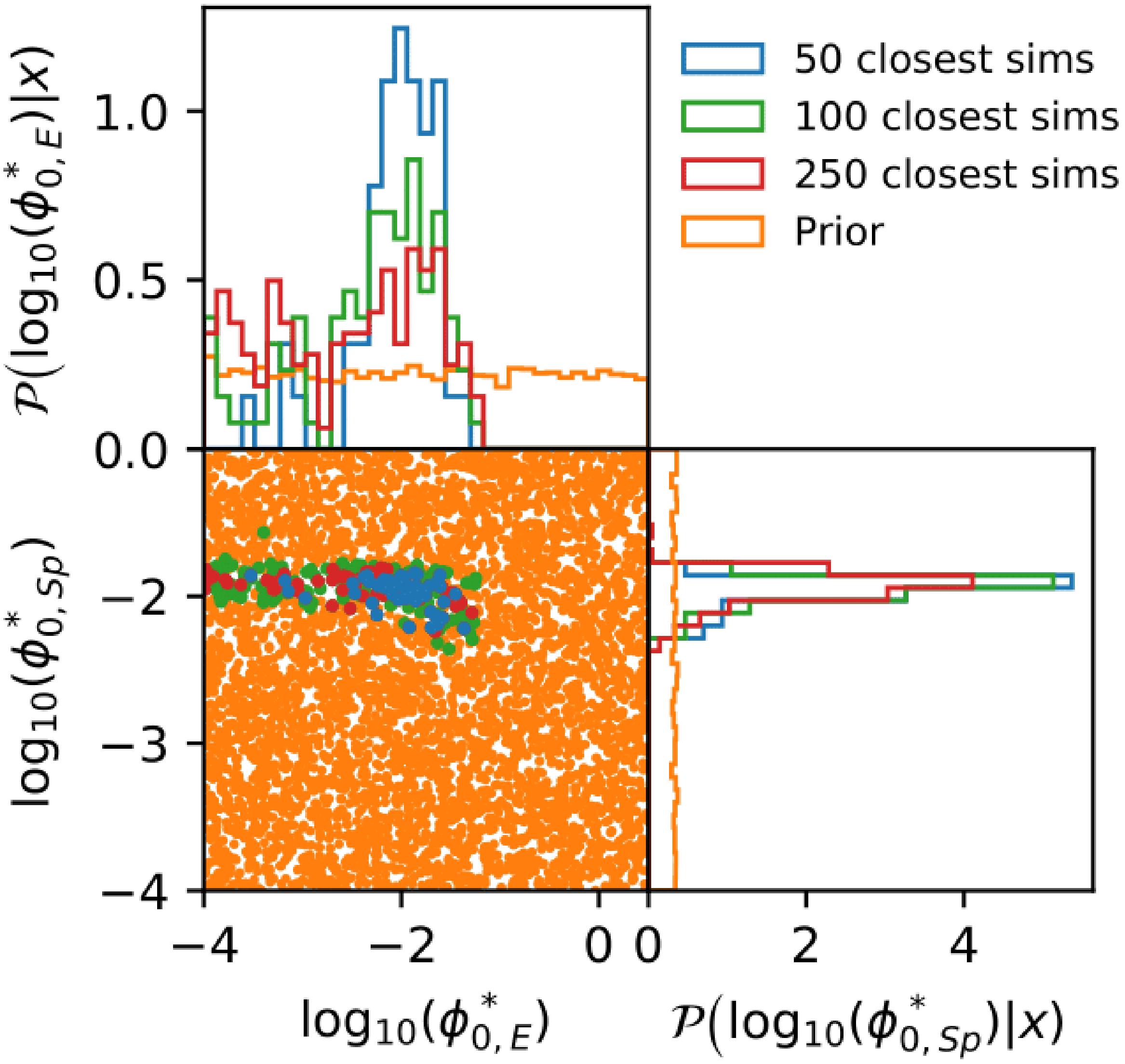


Fig. 6.6: Results of the ABC procedure. *Bottom left:* Parameter values corresponding to 5000 simulations (dots) drawn from our random uniform prior (orange) of Eq. (6.2). The colored points are those with a small distance ρ of Eq. (5.4) to the observed data (frame 'b' of Fig. 6.5 from the CFHTLS D1 Deep field): the 50 closest points are shown in blue, the 100 closest points are shown in green, and the 250 closest points are shown in red. *Top left and bottom right:* Marginalized distributions of the distance selections in the bottom left panel.

Table 6: 1D confidence intervals for the inference from the five images of the virtual data and initial values at which the virtual data are generated.

	Initial	68%	95%	99%
virtual 1				
$\log_{10}(\phi^*_{0,\mathrm{E}})$	-2.09	[-3.20 , -1.63]	[-3.86 , -1.23]	[-3.99 , -0.93]
$\log_{10}(\phi^*_{0,\mathrm{Sp}})$	-2.04	[-2.11 , -1.93]	[-2.23 , -1.86]	[-2.45 , -1.86]
virtual 2				
$\log_{10}(\phi^*_{0,\mathrm{E}})$	-2.09	[-2.86 , -1.94]	[-3.67 , -1.76]	[-4.00 , -1.59]
$\log_{10}(\phi^*_{0,\mathrm{Sp}})$	-2.04	[-2.09 , -1.98]	[-2.16 , -1.93]	[-2.31 , -1.86]
virtual 3				
$\log_{10}(\phi^*_{0,\mathrm{E}})$	-2.09	[-2.78 , -1.97]	[-3.56 , -1.76]	[-3.93 , -1.75]
$\log_{10}(\phi^*_{0,\mathrm{Sp}})$	-2.04	[-2.10 , -2.00]	[-2.15 , -1.96]	[-2.21 , -1.91]
virtual 4				
$\log_{10}(\phi^*_{0,\mathrm{E}})$	-2.09	[-3.15 , -1.60]	[-3.94 , -1.27]	[-3.99 , -0.85]
$\log_{10}(\phi^*_{0,\mathrm{Sp}})$	-2.04	[-2.11 , -1.92]	[-2.26 , -1.85]	[-2.92 , -1.85]
virtual 5				
$\log_{10}(\phi^*_{0,\mathrm{E}})$	-2.09	[-2.60 , -1.76]	[-3.50 , -1.58]	[-3.95 , -1.58]
$\log_{10}(\phi^*_{0,\mathrm{Sp}})$	-2.04	[-2.09 , -1.99]	[-2.17 , -1.95]	[-2.23 , -1.88]
chain-rule joint				
$\log_{10}(\phi^*_{0,\mathrm{E}})$	-2.09	[-2.55 , -1.76]	[-3.40 , -1.54]	[-3.90 , -1.51]
$\log_{10}(\phi^*_{0,\mathrm{Sp}})$	-2.04	[-2.08 , -1.99]	[-2.14 , -1.94]	[-2.20 , -1.91]

Note: where ϕ^* is given for $H_0 = 100$ km s^{-1} Mpc^{-1} and for $z = 0$ (ϕ^* is in units of Mpc^{-3} mag^{-1}).

1.5 PMC posterior

Because, the inference is simultaneously made from 5 pieces of data (observed or virtual), a parallelized version of the PMC procedure* is applied for the five images of the virtual and observed data. The same priors as in Eq. (6.2) are used and the PMC procedure of Chap. 5 Sect. 3. is applied with $N = 500$, $M = 5000$ and a $Q = 25\%$ threshold for resampling the N sets of parameters. At each iteration, the procedure therefore tries to re-draw 125 samples and concludes when the number of attempts in the same iteration is higher than 5000. This yields about 60000 draws of parameters and 60000 generated simulations and the PMC procedure lasted ≈ 1 days per observed and virtual data. The summaries provided by the trained network encode information about the relation of the number of ellipticals and spirals in the images to the model parameters, and therefore can be used for the inference of this property using the PMC procedure.

The results of the PMC procedures are shown in Fig. 6.7 for the virtual data and Fig. 6.8 for the observed CFHTLS D1 data. As shown by the 2D distribution of points (bottom left) and the 1D projected Gaussian kernel density estimates (KDE), the posteriors are concentrated around similar regions of parameter values for the five insets of each sample for the observed and virtual data. This is because the data come from the same generative process, while the differences in the posteriors are due to the independent realizations containing more or less constraining power. The 68%, 95%, and 99% confidence intervals for the 1D marginal distributions are given in Tables 6 and 7 for the five individual images. Because the distribution is not Gaussian in 2D, neither its covariance nor the 1D marginal standard deviations are sufficient to wholly encapsulate the information in this posterior distribution. This gives us further good reason to consider such likelihood-free inference methods. The parameter values used to generate the virtual data images are shown with dotted red lines in Fig. 6.7 and can be compared to the posterior distributions. All results, namely the parameter values used to generate the data (for the virtual data alone) and the 68%, 95%, and 99% 1D confidence intervals for $\Phi^*_{0,\mathrm{E}}$ and $\Phi^*_{0,\mathrm{Sp}}$ are listed in Tables 6 and 7 for the five individual images.

It is worth noticing that for the observed and virtual data, Figs. 6.7 and 6.8 show that the results are

---

[*] parallelized version that I wrote using a Docker environment and Tensorflow Serving, see `https://www.tensorflow.org/tfx/serving/docker` for a more detailed description

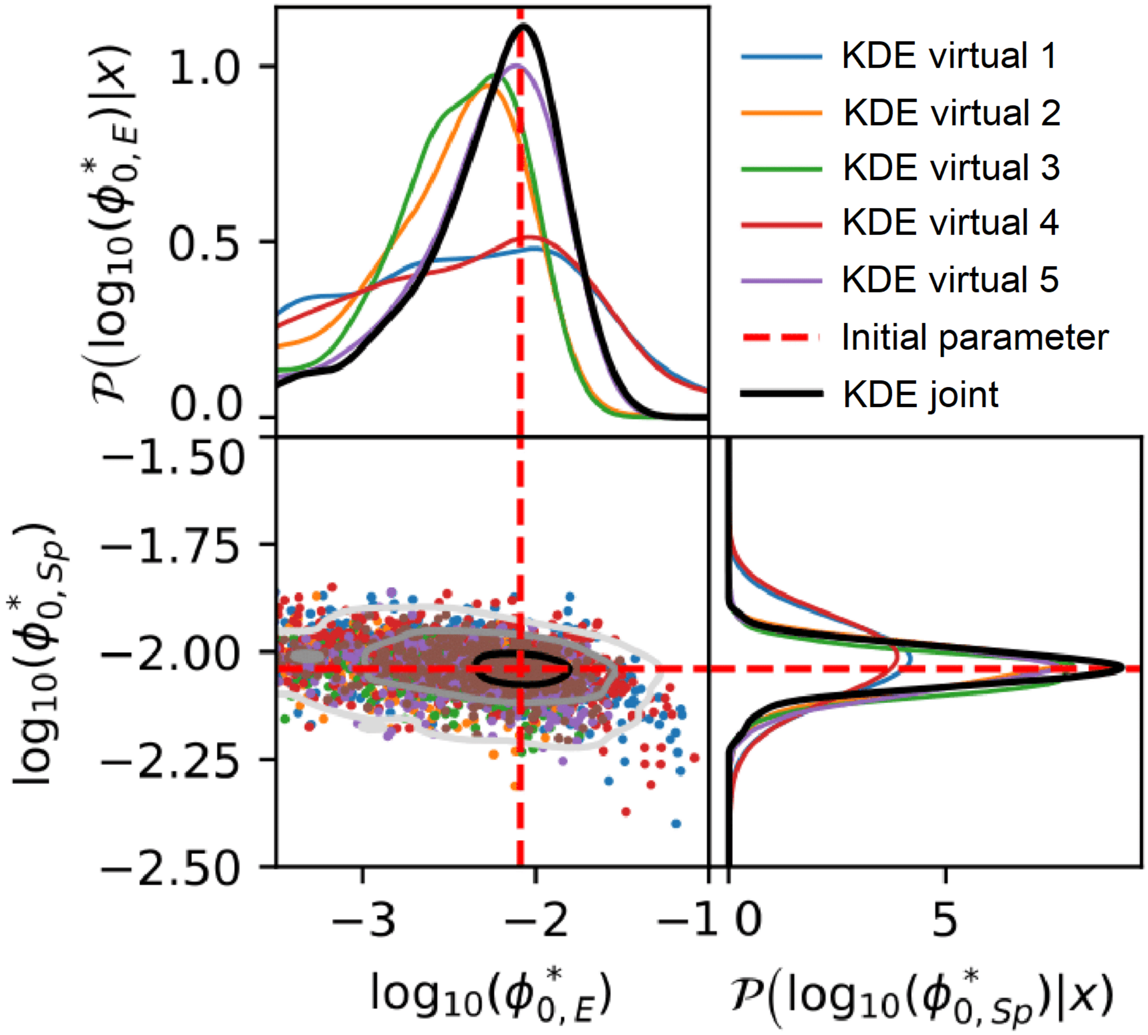


Fig. 6.7: Posterior distributions of the two parameters $\log_{10}(\phi^*_{0,\mathrm{E}})$ and $\log_{10}(\phi^*_{0,\mathrm{Sp}})$ for the five images of virtual data (with different colors from blue to purple) and for the joint-PMC (brown) described in Sect. 1.6. The 68%, 95%, and 99% contours of the joint-PMC are plotted in black, gray, and light gray, respectively, in the bottom left panel. The initial parameters used to generate the virtual data are indicated with dashed red lines. The 2D region of the posterior distribution with the highest density is at the same location for the five individual 1D posteriors for each image, which indicates that the most likely parameters are consistent among the five images. The deviation between the different posteriors arises from the fact that these fields are stochastically sampled from a random process and so statistical differences exist in the virtual data. The joint posterior is tighter and shows how likely the parameters would be if we considered the five images simultaneously.

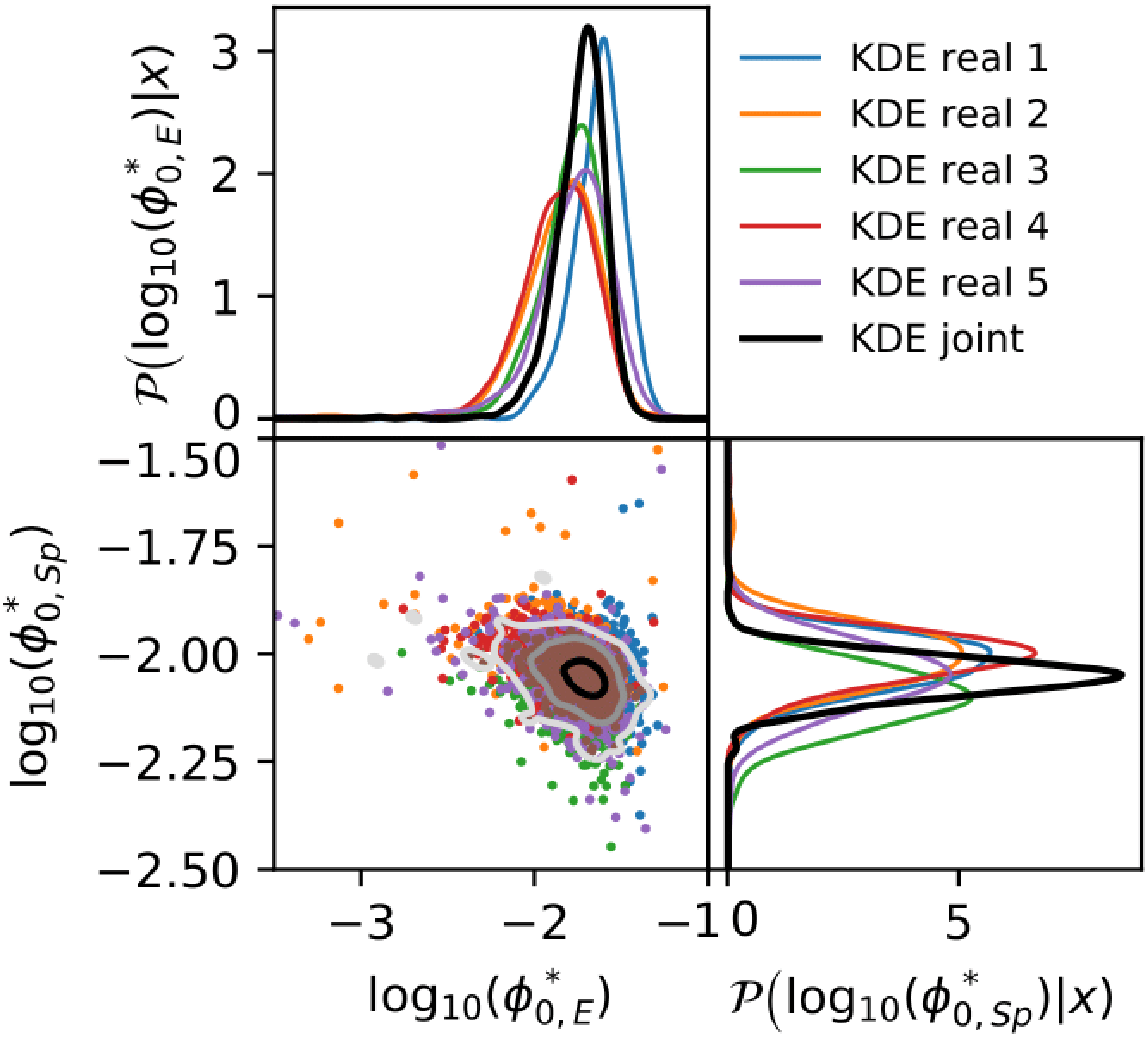


Fig. 6.8: Posterior distributions of the two parameters $\log_{10}(\phi^*_{0,\mathrm{E}})$ and $\log_{10}(\phi^*_{0,\mathrm{Sp}})$ for the five insets of observed CFHTLS D1 data (blue to purple) and for the joint-PMC (brown) described in Sect. 1.6. The 68%, 95%, and 99% contours of the joint-PMC are plotted in black, gray, and light gray, respectively, in the bottom left panel. The 2D region of the posterior distribution with the highest density is at the same location for the five individual 1D posteriors for each image, which indicates that the most likely parameters are consistent among the five insets of the CFHTLS D1 deep field. The differences in the posteriors obtained from the different images come from the fact that the observed data come from different patches of the sky with statistically different amounts of information in the patches due to their independent environments. The joint posterior is tighter and shows how likely the parameters would be if we considered the five images simultaneously.

Table 7: 1D confidence intervals for the five insets of the D1 observed data and the joint data.

	68%	95%	99%
real 1			
$\log_{10}(\phi^*_{0,\mathrm{E}})$	[-1.75 , -1.47]	[-1.94 , -1.35]	[-2.12 , -1.29]
$\log_{10}(\phi^*_{0,\mathrm{Sp}})$	[-2.08 , -1.95]	[-2.17 , -1.89]	[-2.31 , -1.83]
real 2			
$\log_{10}(\phi^*_{0,\mathrm{E}})$	[-2.01 , -1.60]	[-2.31 , -1.43]	[-2.89 , -1.26]
$\log_{10}(\phi^*_{0,\mathrm{Sp}})$	[-2.08 , -1.93]	[-2.17 , -1.85]	[-2.27 , -1.53]
real 3			
$\log_{10}(\phi^*_{0,\mathrm{E}})$	[-1.92 , -1.58]	[-2.16 , -1.47]	[-2.51 , -1.33]
$\log_{10}(\phi^*_{0,\mathrm{Sp}})$	[-2.17 , -2.02]	[-2.24 , -1.96]	[-2.46 , -1.96]
real 4			
$\log_{10}(\phi^*_{0,\mathrm{E}})$	[-2.03 , -1.63]	[-2.30 , -1.47]	[-2.76 , -1.31]
$\log_{10}(\phi^*_{0,\mathrm{Sp}})$	[-2.06 , -1.94]	[-2.15 , -1.89]	[-2.22 , -1.84]
real 5			
$\log_{10}(\phi^*_{0,\mathrm{E}})$	[-1.93 , -1.54]	[-2.28 , -1.34]	[-2.90 , -1.24]
$\log_{10}(\phi^*_{0,\mathrm{Sp}})$	[-2.13 , -1.96]	[-2.23 , -1.88]	[-2.44 , -1.73]
chain-rule joint			
$\log_{10}(\phi^*_{0,\mathrm{E}})$	[-1.84 , -1.58]	[-2.05 , -1.48]	[-2.39 , -1.38]
$\log_{10}(\phi^*_{0,\mathrm{Sp}})$	[-2.10 , -2.00]	[-2.15 , -1.96]	[-2.23 , -1.92]

Note: where ϕ^* is given for $H_0 = 100$ km s^{-1} Mpc^{-1} and for $z = 0$ (ϕ^* is in units of Mpc^{-3} mag^{-1}).

narrower 1D marginalized distributions for the spiral population than for the elliptical population. This comes from the steep faint-end slope of the luminosity function for the spirals: a small change in $\log_{10}(\phi^*_{0,\mathrm{Sp}})$ has a strong effect on the number of faint spirals, which can be more easily distinguished. This is not exactly the case for the other parameter: a strong change in $\log_{10}(\phi^*_{0,\mathrm{E}})$ is needed to considerably alter the number of elliptical galaxies, which equates to greater uncertainty on this parameter.

Comparison of Figs. 6.7 and 6.8 show that if the posterior distributions of the spirals have a similar dispersion for the virtual and observed data, the posterior distributions of the elliptical population is narrower for the observed than for the virtual data. This is an effect of the small number of ellipticals in an image: with a low value for $\log_{10}(\phi^*_{0,E})$, there are very few elliptical galaxies in the image, and this lack of a statistically significant sample increases the width of the posterior due to the lack of available information. In comparison, the posterior is narrower for the observed data because more elliptical galaxies appear to be present: the 2D region of the joint posterior distribution with the highest density gives a value of $\log_{10}(\phi^*_{0,E})$ around -1.7 for the observed data, compared to a lower value of $\log_{10}(\phi^*_{0,E})$ around -2.1 for the virtual data (see Figs. 6.7 and 6.8 and Tables 6 and 7). This effect is discussed with more details in Sect. 1.7.

For the observed data frame labeled "real 3" in Fig. 6.8, the 1D confidence interval of the parameter $\log_{10}(\phi^*_{0,Sp})$ of Table 7 is slightly displaced compared to the other observed images, and moreover toward lower densities of spirals. This can be explained by the presence of a large spiral galaxy in the image of frame "c" in Fig. 6.5. This galaxy covers $\sim 100 \times 100$ pixels, an area of the image (*i.e.* $\sim 1\%$ of the full 1024×1024 image) that reduces the amount of detected spirals, which therefore appear to be fewer because the area in which they are visible is smaller. This illustrates the fact that the inference is correct even if the PMC procedure happens to be biased because of statistically anomalous data. This is the reason why five insets of the D1 deep field are used, with the goal to improve the robustness of the results.

1.6 Joint posterior and confidence intervals

Because only individual posteriors for each image are available from the PMC procedures, we need to combine them to derive a unique posterior for multiple data. Unfortunately, posterior chains cannot be combined in a simple way (*i.e.* by a simple product of the posterior distributions). A rigorous way to achieve such a combined-posterior distribution is to use the Bayesian chain rule,

$$p(\theta|X_1,\dots,X_n) \propto p(\theta)p(X_1|\theta)p(X_2|X_1,\theta)\cdots p(X_n|X_1,\dots,X_{n-1},\theta), \tag{6.3}$$

where θ is a set of parameters, and $X_1,\dots,X_n$ are the n individual pieces of data. The chain rule allows us to obtain the posteriors sequentially: if $\forall i \in \{1,\dots,n-1\}$ the posterior distribution of $p(\theta|X_1,\dots,X_i)$ is already available, then

1. Consider $p(\theta|X_1,\dots,X_i)$, derived from the pieces of data $X_1,\dots,X_i$, as an approximate proposal distribution for θ.
2. Use that proposal distribution of θ for the PMC prior, given the new piece of data, X_{i+1}.
3. Derive a new posterior distribution from $p(\theta|X_1,\dots,X_{i+1})$ that can be used in turn as the proposal distribution for the next piece of data.

I wrote a parallelized version of such a sequential-PMC for the five images for the virtual and observed data. The resulting joint posteriors are shown as black lines (called kernel density estimate “KDE joint”) in Figs. 6.7 and 6.8 and can be compared to the posterior distributions obtained for the individual images. In these figures, the 68%, 95% and 99% contours are drawn using the joint posterior in black, gray, and light gray, respectively. The 1D 68%, 95%, and 99% confidence intervals for the joint posteriors are listed at the end of Table 6 for the virtual data and in Table 7 for the observed data. The confidence intervals are narrower for the joint posteriors because more information is available, which allows us to draw tighter constraints on the parameter values.

1.7 Comparison with other studies

Our results are compared with other measurements of the luminosity functions for the elliptical (or red, or quiescent) galaxy population and for the spiral (or blue, or star-forming) population obtained from deep multiband galaxy surveys: López-Sanjuan et al. (2017), Brown et al. (2007), Beare et al. (2015), Salimbeni et al. (2008), Drory et al. (2009), Faber et al. (2007), Zucca et al. (2009), and Zucca et al. (2006). In these studies, the two populations are identified either by comparing the observed magnitudes in different bands or by finding the best-fit SEDs to each galaxy. These methods are affected by several biases because of the catalog extraction process (see Chap. 2 Sect. 3.) and because of the choice of threshold used to split galaxies into red or blue.

We have shown in Sect. 1.6 that the 2D region of the posterior distribution with the highest density for the observed joint data agrees well with the one for the virtual data; in fact, the values used to generate our virtual data are chosen to be based on the catalogs-method of the Alhambra survey (López-Sanjuan et al. 2017). However, there are differences regarding the elliptical population. In order to properly compare the results, we focus on the global shape of the luminosity functions of elliptical and spiral populations at $z = 0.5$ which includes the correlations among the inferred parameter ϕ^*_0 and the other fixed luminosity function parameters M^*_0, α, ϕ^*_{evol}, M^*_{evol}.

Fig. 6.9 shows the luminosity function using the 68% confidence intervals of the joint posterior over the five insets of the CFHTLS+WIRDS D1 field in blue for the elliptical (top) and spiral populations (bottom) compared to the results from the other articles. This shows a good agreement between our results and the other surveys for the spiral population (bottom panel) for which the network was able to precisely constrain the luminosity function (narrow 68% confidence interval). Despite a different sample and no catalog extraction, when our results are compared to those of López-Sanjuan et al. (2017), we obtain for the dominant spiral population a 1σ confidence interval on ϕ^* comparable to theirs. The top panel of Fig. 6.9 shows that for the elliptical population, there are larger differences between all analyses, including ours. This is likely partly due to Poisson noise: the number density of galaxies in the elliptical population is typically a factor of 10 to 100 lower than for the spiral population in a given magnitude bin, introducing more dispersion. Moreover, as pointed out in several of the analyses we quoted, the chosen samples of red or blue galaxies contain misidentified galaxies, which affect the smaller elliptical population more.

The top panel of Fig. 6.9 also indicates evidence for an excess of red faint galaxies: López-Sanjuan et al. (2017), Drory et al. (2009), and Salimbeni et al. (2008) have used the sum of two Schechter functions

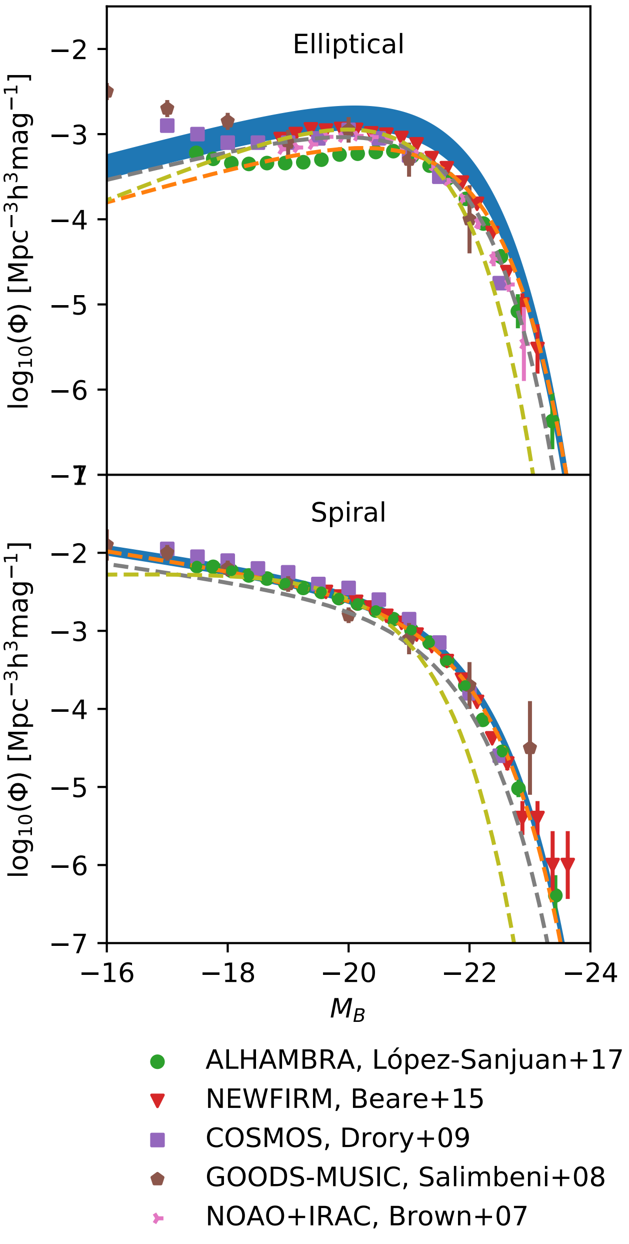


Fig. 6.9: Inferred luminosity functions derived for the elliptical (top) and the spiral (bottom) populations using the 68% confidence intervals (blue) for the joint posterior of Table 7. Our results are compared with the results of López-Sanjuan et al. (2017) in green, Brown et al. (2007) in pink, Beare et al. (2015) in red, Salimbeni et al. (2008) in brown, Drory et al. (2009) in purple, Faber et al. (2007) in orange, Zucca et al. (2009) in gray, and Zucca et al. (2006) in light green. Our g' magnitudes are converted into the Johnson B-band absolute magnitude, at redshift $z = 0.5$ and for $H_0 = 100$ km s^{-1} Mpc^{-1} using $B - g' = 0.78$ for ellipticals and $B - g' = 0.40$ for spirals, listed in Table 7 of Fukugita et al. (1995).

to better model the luminosity function of the elliptical population and its “rising” faint end. We suspect that this apparently observed excess of faint elliptical galaxies to which our method is in principle quite sensitive to tends to pull our single-Schechter luminosity function upward for the elliptical population and is therefore systematically higher than the luminosity functions derived by Brown et al. (2007), Beare et al. (2015), Faber et al. (2007), Zucca et al. (2009), and Zucca et al. (2006). In this case, our limited model of two galaxy populations with single evolving Schechter luminosity functions for each population is too coarse. Nevertheless, the rising of the elliptical luminosity function at the faint end could be caused by some of the faint spiral galaxies, which are numerous, being categorized as red galaxies by some source extraction method.

Conversely, some of the elliptical galaxies in the other analyses may have been misclassified by some source extraction method as blue galaxies and might lead to a systematically too low amplitude for the elliptical luminosity function. Moreover, dusty star-forming galaxies with red colors may be modeled as ellipticals in our analysis (due to the resemblance of the strong starburst to a bulge-dominated object in the image), whereas they may be classified as spirals in the other analyses. Both effects would affect only the luminosity function of elliptical galaxies because spiral types vastly dominate the elliptical types in number.

Finally, Fig. 6.10 the photometric redshift distribution derived by Ilbert et al. (2006) compared to the redshift distribution of the galaxies generated from our forward model with the best fit values measured by the PMC procedure:

$$\begin{array}{c|c}
\text{Elliptical} & \text{Spiral} \\
\hline
\log_{10}\left(\phi^*_{0,E}\right) = -1.76 \quad \text{(best fit)} & \log_{10}\left(\phi^*_{0,Sp}\right) = -2.05 \quad \text{(best fit)} \\
M^*_{0,E} = -19.68 & M^*_{0,Sp} = -19.71 \\
\alpha_E = -0.53 & \alpha_{Sp} = -1.29 \\
\phi^*_{\text{evol},E} = -1.37 & \phi^*_{\text{evol},Sp} = -0.03 \\
M^*_{\text{evol},E} = -1.15 & M^*_{\text{evol},Sp} = -1.49
\end{array} \tag{6.4}$$

Apart from the two best fit values $\log_{10}\left(\phi^*_{0,E}\right)$ and $\log_{10}\left(\phi^*_{0,Sp}\right)$ (that are chosen as the maximum of the Gaussian kernel density estimate of the 2D posterior distribution of Fig. 6.8), the other parameter values are fixed to the values measured in López-Sanjuan et al. (2017, Table 2) which correspond to the fiducial values of our forward model. We see an even better consistency between the photometric and the modeled redshifts than what is seen in Fig. 3.11. It is worth noting that in the fourth panel $24 < i' < 25$, the high redshift tail of the distribution obtained from our forward model is above those of the CFHTLS and this is again because we reach the completeness limit of this survey.

2. Third application: 2 populations and 6 parameters (Preliminary results)

In the previous application, the shape of the luminosity functions for the two populations, elliptical and spiral, was fixed because only two free parameters (*i.e.* the density parameter ϕ^*_0) were considered. In this third application, we try to extend these results to a model with now six parameters and two populations (*i.e.* ϕ^*_0, M^*_0 and α for the elliptical and spiral populations). Therefore, the shape of the luminosity functions can now be properly inferred, thus the number density and the ratio faint/bright galaxies can be studied simultaneously for these two major populations.

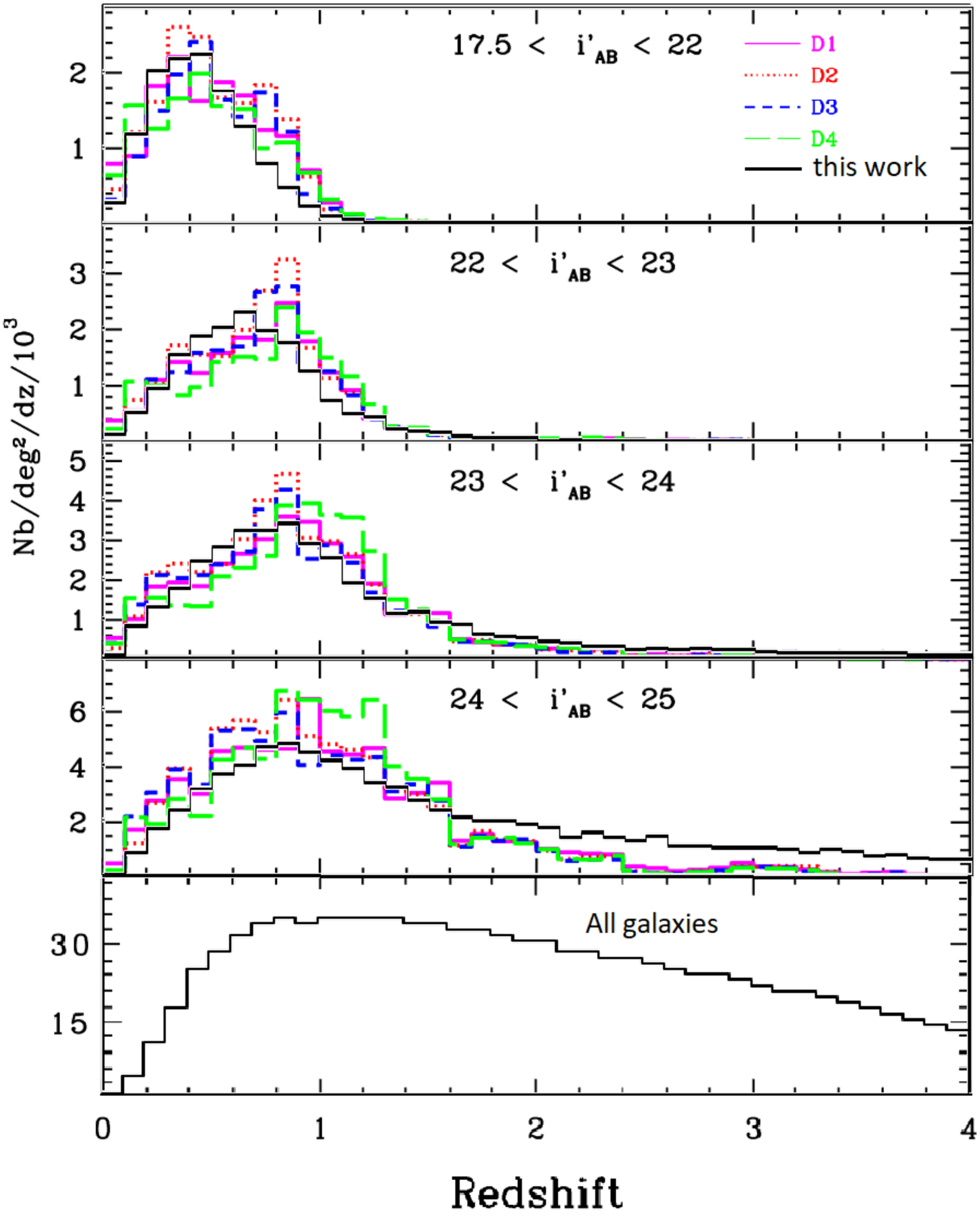


Fig. 6.10: Comparison between the redshift distribution from Ilbert et al. (2006) in the four deep fields (in pink, red, blue and green) with the redshift distribution for our forward model (in black) decomposed into elliptical galaxies and spiral galaxies, for best fit values of $\log_{10}\left(\phi^*_{0,E}\right)$ and $\log_{10}\left(\phi^*_{0,Sp}\right)$ inferred from the 5 insets of the CFHTLS D1 deep field (see Fig. 6.5), and the other luminosity function parameters fixed to their fiducial values. This graph shows the very good agreement between the photometric redshifts derived from the 'T0003' release of the CFHTLS (Ilbert et al. 2006) and the redshift of the galaxies generated by our forward model. The last panel gives the redshift distribution of all galaxies generated by our forward model.

2.1 Description

In this application, we still use the values of López-Sanjuan et al. (2017) to choose the fiducial parameter parameters and offsets for the numerical derivatives that are used to fit the network:

$$\begin{array}{cccccc|cccccc}
 & & \text{Elliptical} & & & & & & \text{Spiral} & & \\
\hline
\theta_{\text{fid},1} & = & \log_{10}\left(\phi^*_{0,E}\right) & = & -2.09 & & \theta_{\text{fid},4} & = & \log_{10}\left(\phi^*_{0,Sp}\right) & = & -2.04 \\
\Delta\theta_1 & = & \Delta\log_{10}\left(\phi^*_{0,E}\right) & = & 0.5 & & \Delta\theta_4 & = & \Delta\log_{10}\left(\phi^*_{0,Sp}\right) & = & 0.1 \\
\theta_{\text{fid},2} & = & M^*_{0,E} & = & -19.68 & & \theta_{\text{fid},5} & = & M^*_{0,Sp} & = & -19.71 \\
\Delta\theta_2 & = & \Delta M^*_{0,E} & = & 1 & & \Delta\theta_5 & = & \Delta M^*_{0,Sp} & = & 0.25 \\
\theta_{\text{fid},3} & = & \alpha_E & = & -0.53 & & \theta_{\text{fid},6} & = & \alpha_{Sp} & = & -1.29 \\
\Delta\theta_3 & = & \Delta\alpha_E & = & 0.25 & & \Delta\theta_6 & = & \Delta\alpha_{Sp} & = & 0.1
\end{array} \tag{6.5}$$

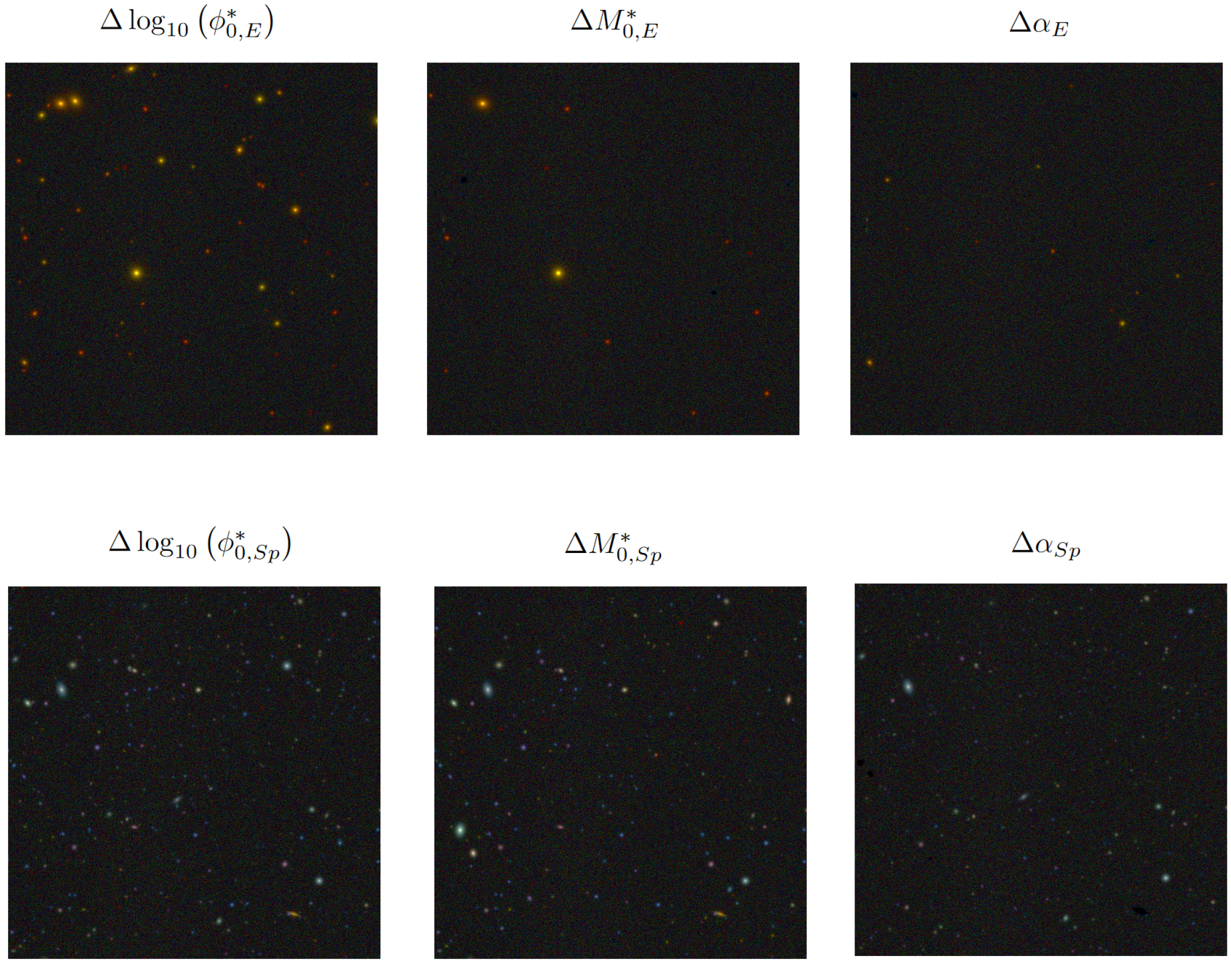
where ϕ^* and M^* are given for $H_0 = 100$ km s^{-1} Mpc^{-1} and $z = 0$. The other luminosity function parameters $\phi^*_{\text{evol},E}$, $M_{\text{evol},E}$, $\phi^*_{\text{evol},Sp}$ and $M_{\text{evol},Sp}$ are fixed with the values from López-Sanjuan et al. (2017), see Table 9. As in the previous application, the offset values of the parameters for the elliptical population are higher than those for the spiral population when the simulations for the numerical derivatives are generated because the number density of spirals is higher than that of ellipticals.

The simulated and observed images still are of size 1024 × 1024 pixels in the eight photometric bands u', g', r', i', z', J, H, Ks. Fig. 6.11 shows the effects on a RGB image of the g', r', i' filters for the six parameters when subtracting the simulation generated with positive offset to the simulation generated with the corresponding negative offset. This shows the galaxies that appear in the images used to calculate the numerical derivatives of the summaries using the IMMN. The top panels show that only elliptical galaxies remain (these galaxies appear as yellow or red because of the SED template 'E' used to model them), as opposed to the blue galaxies (modeled as a bulge with the 'E' SED and a disk with the 'Irr' SED) of the spiral population remaining in the bottom panels. Table 9 gives an overview of the values that are used to generate the simulations according to the description of the forward model of Chap. 3.

2.2 Training of the network

In this application, more simulations are generated to efficiently train the network using the forward model with six parameters. We use $n = 350$ fiducial images of 3.17 arcmin^2 and 1024 × 1024 pixels in the eight photometric bands, and $m = 100$ seed-matched images for the numerical derivatives for the six parameters at values $\theta_{\text{fid},\gamma} + \Delta\theta_\gamma$, $\gamma = 1, \ldots, 6$, and another $m = 100$ simulations per parameter at values $\theta_{\text{fid},\gamma} - \Delta\theta_\gamma$, $\gamma = 1, \ldots, 6$, see Eq. (6.5). This yields a total of $350 + 6 \times 2 \times 100 = 1550$ deep field images (in the eight bands) for the training set, and the same number is used for the validation set. The optimizer techniques and the corresponding learning rates used during the training are listed in Table 8. The Adam optimizer is applied for the first 1200 epochs to quickly find a transformation that forces the covariance matrix of the summaries to be approximately the identity matrix, then we switch to the stochastic gradient descent momentum algorithm with a decreasing learning rate to refine the training process. The architecture of the network is that of Fig. 4.6 (but with output summary statistics of size 6 instead of 2), which yields a total of 12,302 network parameters to be fitted. Because of the number and size of the images, within this configuration, one epoch of training (*i.e.* when the whole training and validation sets are passed through the network and the network weights are updated) takes about 90 seconds to run (on a NVIDIA QUADRO RTX 8000 45GB GPU). The network is trained for ∼24k epochs (∼20 days) and the procedure is not fully converged but is stopped because the determinant of the Fisher matrix does not increase too much for the last thousand epochs.

At the end of the training the determinant of the Fisher information matrix (top panel of Fig. 6.12) is higher (*i.e.* $\det F \approx 10^{12}$) for the training set than for the validation set (*i.e.* $\det F \approx 10^{10}$) as compared to the previous applications. This discrepancy of two order of magnitude between the training curve and the validation curve indicates that the training is not optimal and a slightly lower amount of Fisher information is extracted from the validation set. This difference is also seen for the determinant of the covariance between the training set and the validation set (bottom panel of Fig. 6.12) which indicates that only 350 simulations generated with fiducial parameters are not enough to properly estimate the covariance matrix of the summary statistics.



$\Delta \log_{10}\left(\phi^*_{0,E}\right)$ $\Delta M^*_{0,E}$ $\Delta \alpha_E$

$\Delta \log_{10}\left(\phi^*_{0,Sp}\right)$ $\Delta M^*_{0,Sp}$ $\Delta \alpha_{Sp}$

Fig. 6.11: Each panel corresponds to the subtraction of a simulation generated with positive offset to the simulation generated with the corresponding negative offset for the six parameters used in the third application. These RGB color images are obtained from the i', r', g' filters.

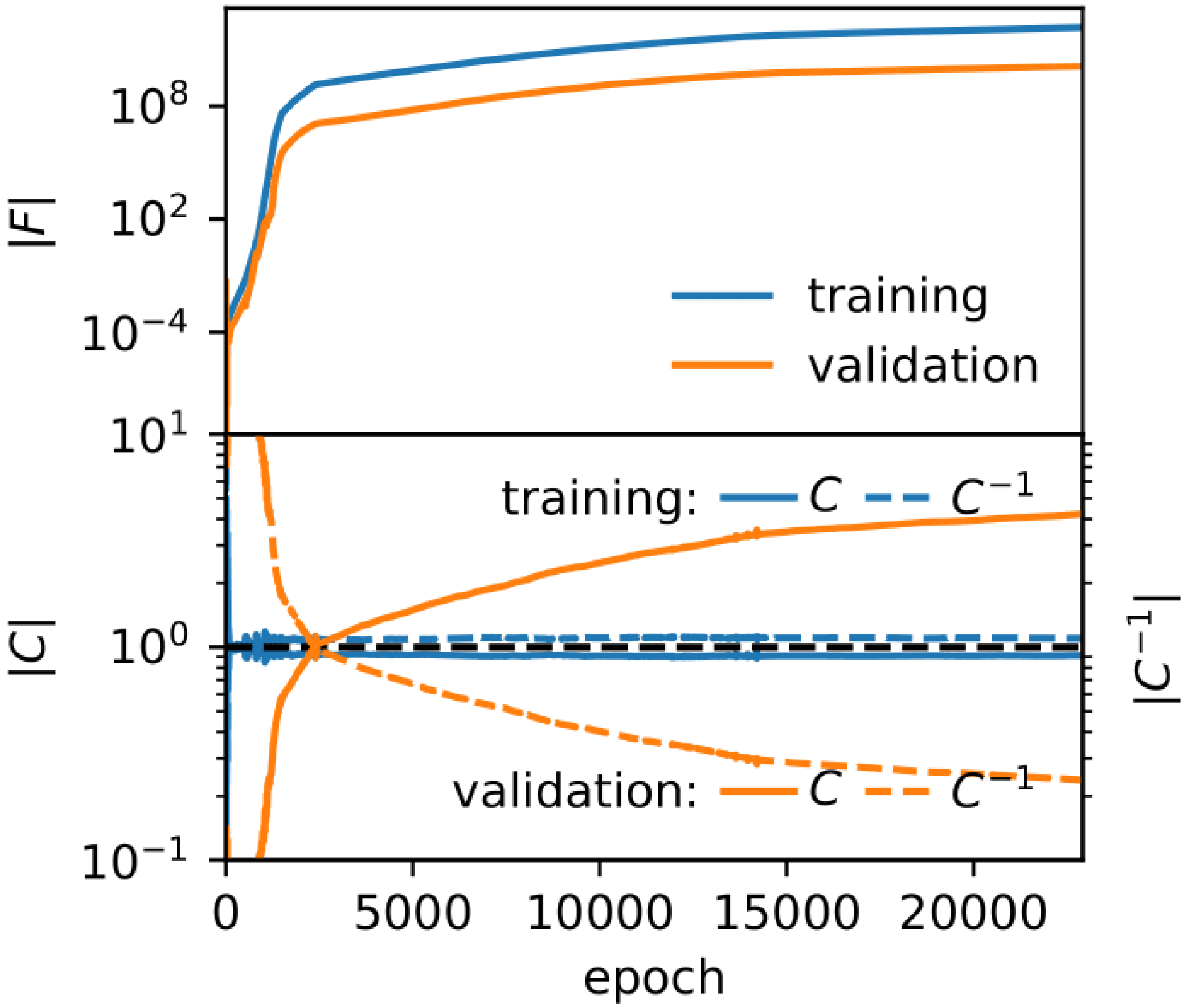


Fig. 6.12: *Top:* Evolution of the determinant of the Fisher information matrix during almost 24 000 epochs. The blue curve represents the information extracted from the training set, and the orange curve shows the same from the validation set. The curves increase, which means that the network learns more about the six parameters as the training continues. The training was stopped when the validation curve flattened, suggesting that the network is close to convergence. *Bottom:* Evolution of the determinant of the covariance matrix (solid line) and the inverse of the covariance matrix (dashed line). The blue curves show the training set, and the orange curves show the validation set. The training curves reach 1 very fast, which shows that the loss is stabilized and that the magnitude of the summaries is under control. The validation curve stays close to identity, which is a sign that there is some weak parameter dependence on the covariance.

Table 8: Optimizer and learning rate used during the training of Chap. 6 Sect. 2..

epoch	optimizer	learning rate
0	Adam	0.01
1200	SGD	5×10^{-5}
1500	SGD	2×10^{-5}
2400	SGD	5×10^{-6}
14200	SGD	2×10^{-6}

Note: SGD stands for stochastic gradient descent.

The discrepancy between the two extracted Fisher information can also be explained by the fact that the values of the deviations around the fiducial parameters produce different effects on the number of galaxies that appear/disappear from one parameter to the other. For instance, the $\Delta\alpha_E$ deviation of Fig. 6.11 causes less differences than for the $\Delta \log_{10}\left(\phi^*_{0,E}\right)$ deviation and these differences in terms of number of galaxies can disturb the training process. The same argument applies for the $\Delta\alpha_{Sp}$ deviation parameter. These considerations probably explain why the marginalized posterior 1D distributions in Fig. 6.14 look flatter for the two parameters α_E and α_{Sp}.

For this application, performing an ABC procedure to derive the posterior distribution is unfeasible because it would take a prohibitive amount of computing time to populate the 6D prior region of parameters; for instance, if the prior interval for each parameter is binned with only 10 bins it would mean that we need to generate 10^6 simulations which is not possible. This is the reason why we only perform the PMC procedure in the following section.

2.3 PMC posterior

In this section, we apply the PMC procedure of Chap. 5 Sect. 3. on Fig. 6.13, that is an observed image of size 1024×1024 pixels of the CFHTLS. A uniform prior is used for the six parameters as follows:

Prior distribution (uniform)									
Elliptical					Spiral				
-4	$\leq$	$\log_{10}(\phi^*_{0,\mathrm{E}})$	$\leq$	-0.5	-4	$\leq$	$\log_{10}(\phi^*_{0,\mathrm{Sp}})$	$\leq$	-0.5
-23	$\leq$	$M^*_{0,\mathrm{E}}$	$\leq$	-16	-23	$\leq$	$M^*_{0,\mathrm{Sp}}$	$\leq$	-16
-1.6	$\leq$	α_{E}	$\leq$	1	-1.6	$\leq$	α_{Sp}	$\leq$	1

(6.6)

The PMC is performed with $N = 1\,000$ sets of parameters, $M = 50\,000$ (stopping criterion of the procedure), and $Q = 75\%$ for resampling the N sets of parameters. At each iteration, the PMC procedure therefore re-draws 750 samples and concludes when the number of attempts in the same iteration is higher than 50 000. The procedure yielded 311 057 draws and generated simulations in total and the PMC procedure has run for ≈ 4 weeks. The results are shown in Fig. 6.14 with the 1D and 2D marginalized posterior distribution. The PMC procedure gives the following 68% confidence intervals for the 1D marginal distribution:

68% confidence intervals (1D marginal)									
Elliptical					Spiral				
-1.34	$\leq$	$\log_{10}(\phi^*_{0,\mathrm{E}})$	$\leq$	-0.60	-2.32	$\leq$	$\log_{10}(\phi^*_{0,\mathrm{Sp}})$	$\leq$	-1.14
-18.72	$\leq$	$M^*_{0,\mathrm{E}}$	$\leq$	-16.69	-20.11	$\leq$	$M^*_{0,\mathrm{Sp}}$	$\leq$	-16.87
-1.46	$\leq$	α_{E}	$\leq$	-0.03	-1.31	$\leq$	α_{Sp}	$\leq$	0.38

(6.7)

These preliminary results for a single piece of data image of 1024×1024 pixels from the CFHTLS D1 deep field gives relatively wide marginalized 1D posteriors as seen in Fig. 6.14. The procedure clearly identifies the strong correlation between the two pairs of parameters $\left[\log_{10}\left(\phi^*_{0,E}\right), M^*_{0,E}\right]$ and $\left[\log_{10}\left(\phi^*_{0,Sp}\right), M^*_{0,Sp}\right]$ and this effect is extensively discussed in the application of Chap. 5 Sect. 4.. The procedure also indicates that the more probable region for the parameter $\log_{10}\left(\phi^*_{0,Sp}\right)$ is around -2 and this corresponds to the value

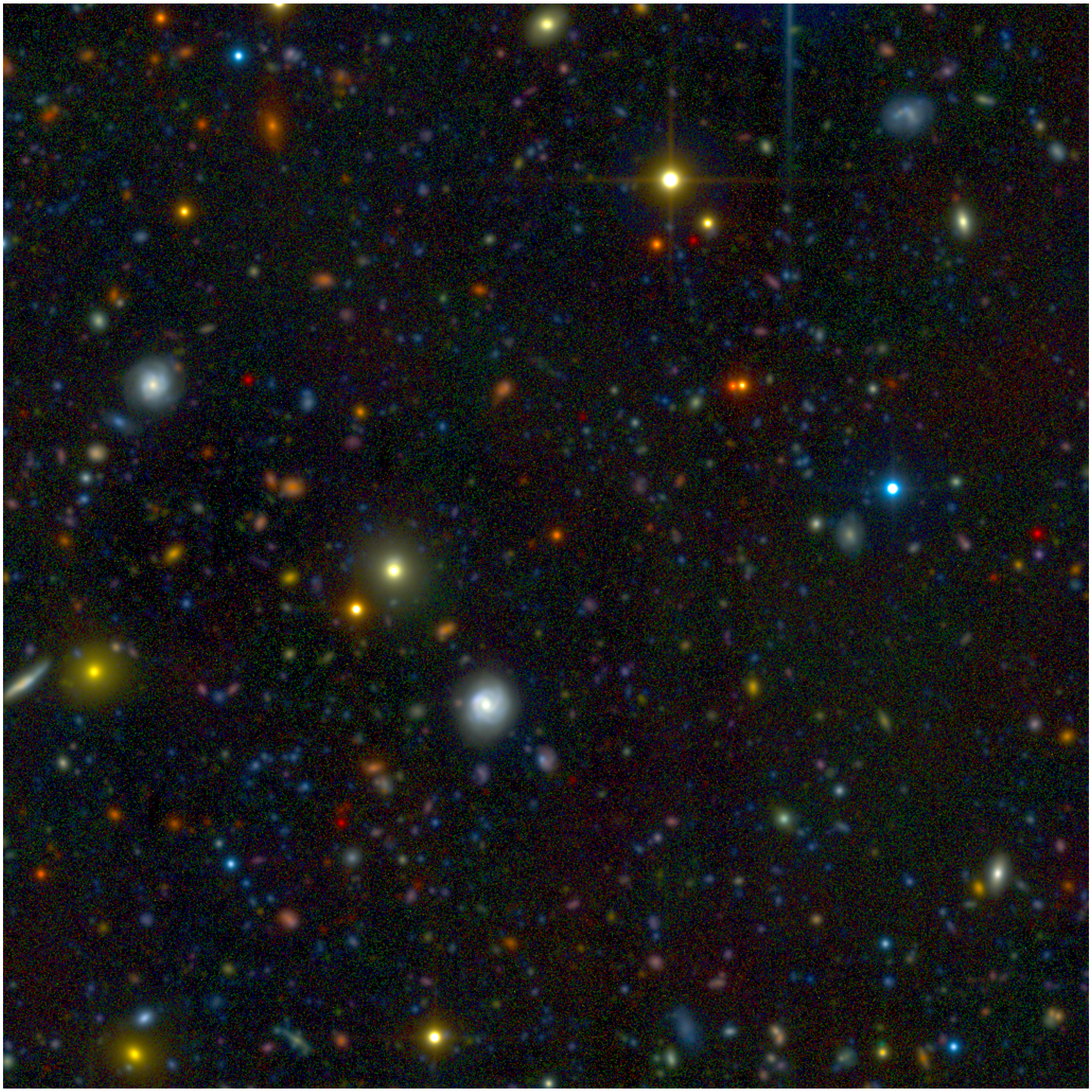


Fig. 6.13: RGB image using the g', r', i' filters for a random region of 3.17 arcmin^2 (1024×1024 pixels) within the 1 deg^2 CFHTLS D1 deep field that is used to infer the luminosity function parameters of the elliptical and spiral galaxies, namely $\log_{10}(\phi^*_{0,\mathrm{E}}), M^*_{0,\mathrm{E}}, \alpha_\mathrm{E}, \log_{10}(\phi^*_{0,\mathrm{Sp}}), M^*_{0,\mathrm{Sp}}, \alpha_\mathrm{Sp}$.

measured by López-Sanjuan et al. (2017). Moreover, the general trend for the parameter $M^*_{0,Sp}$ seems to be located around the value -19.5 which is also a value measured in López-Sanjuan et al. (2017).

However, the more probable region for the three parameters of the elliptical populations ($i.e.$ $\log_{10}\left(\phi^*_{0,E}\right) \approx -0.8$, $M^*_{0,E} \approx -17.5$ and $\alpha_E \approx -1.2$) do not agree with the values measured in the other studies: López-Sanjuan et al. (2017), Brown et al. (2007), Beare et al. (2015), Salimbeni et al. (2008), Drory et al. (2009), Faber et al. (2007), Zucca et al. (2009), and Zucca et al. (2006). In fact, the PMC procedure seems to detect a large number of faint and a very few number of bright elliptical galaxies with respect to what is found in the other studies above. These discrepancies can be explained by the two following limitations:

- The training process of the inception network is not optimal as suggested by the difference between the Fisher information extracted from the training set and the Fisher information extracted from the validation set in Fig. 6.12. The training scheme can be improved by slightly increasing the amplitude of the deviations around the fiducial parameters for the numerical derivative of the mean vector in order to have on average the same number of galaxies that appear/disappear because of these deviations. This would therefore help the inception network to better train on distinguishing the statistical effect of each parameter. In addition, slightly increasing the number of fiducial simulations (from 350 to 400 or even 500) would allow the network to better evaluate the covariance matrices and therefore extract a closer amount of Fisher information from the training set and the validation set.

- The very small number of elliptical galaxies on the single cutout image of 1024×1024 pixels from the CFHTLS D1 deep field can lead to very poor estimates of the three luminosity function parameters for this population. This can be improved by considering simultaneously several such 1024×1024 pixel images and combine the PMC results as in the joint-PMC procedure of Sect. 1.6. In addition, the size of the images can be increased, at least by a factor 8 ($i.e.$ from 1024×1024 to 4096×4096 pixels), and at best considering the full 1 $\deg^2$ (19354×19354 pixels) D1 deep field, in order to increase the number of elliptical galaxies from which half of the luminosity function parameters are inferred.

It is worth noting that the results of the PMC procedure of Fig. 6.14 are not necessarily wrong and this can be a good approximation of the true posterior distribution given this piece of data alone. For instance, the number of elliptical galaxies can be unusually higher on this particular small portion of the full CFHTLS D1 deep field and can mislead our judgment. This is the reason why we cannot trust a PMC procedure performed on a single piece of data and we need to consider either multiple pieces of data images or a larger data image in order to derive robust constraints.

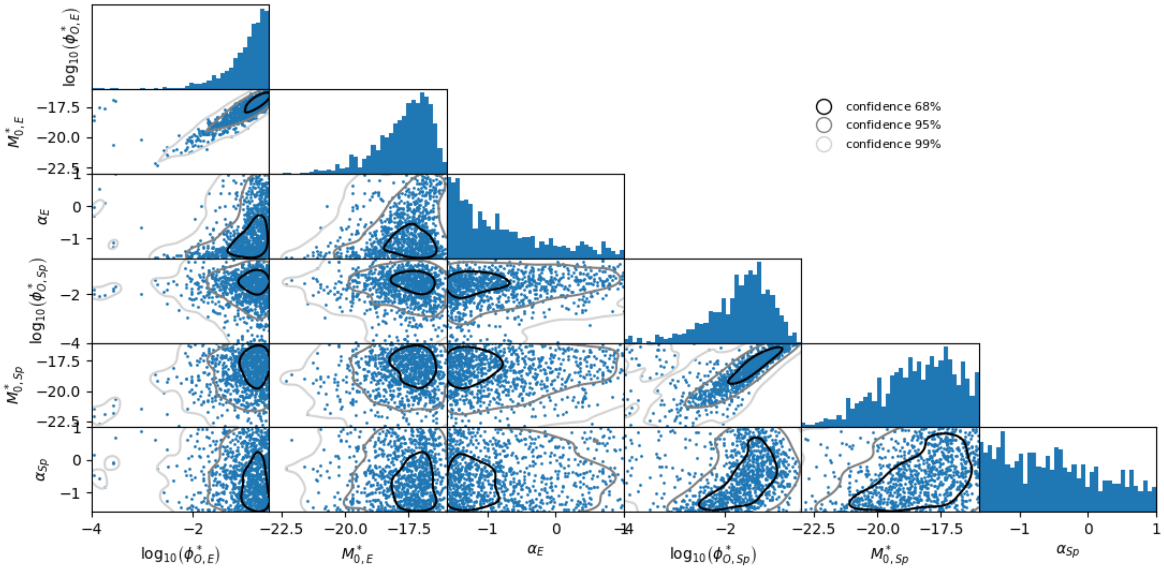


Fig. 6.14: 1D and 2D marginalized posterior distributions of the PMC procedure after 311 051 draws for the initial 6D prior for parameters $\log_{10}(\phi^*_{0,\mathrm{E}}), M^*_{0,\mathrm{E}}, \alpha_{\mathrm{E}}, \log_{10}(\phi^*_{0,\mathrm{Sp}}), M^*_{0,\mathrm{Sp}}, \alpha_{\mathrm{Sp}}$. The 68%, 95%, and 99% contours of the 2D marginalized posterior distributions are given in each panel in black, gray, and light gray, respectively.

Table 9: Overview of the parameters of the forward model for the applications of Chap. 5 Sect. 4. and Chap. 6 Sects. 1. and 2.

APPLICATION	Chap. 5 Sect. 4.	Chap. 6 Sect. 1.	Chap. 6 Sect. 2.
COSMOLOGY			
Hubble constant	$H_0 = 70$ km s^{-1} Mpc^{-1}		
Matter density	$\Omega_M = 0.3$		
Cosmological constant	$\Omega_\Lambda = 0.7$		
Redshift integration step	5 h^{-1} Mpc		
IMAGE			
Pixel size	0.186 arcsec		
Field size	1024×1024 px		
Apparent magnitude limits	$[17, 29]$ for SDSS g' filter		
ZERO POINTS / BACKGROUND			
Collecting area	10.17		
Gain	1.67		
Background SED	puxley		
PASSBANDS			
Reference	SDSS g'		
Observed	MegaPrime u', g', r', i', z', WIRCam J, H, Ks		
Detection type	Photons		
Calibration	AB		
ELLIPTICAL POPULATION			
Hubble type		-6	-6
Bulge to total		1	1
Bulge SED[1]		E	E
Disk SED[1]		$\emptyset$	$\emptyset$
$log_{10}(\phi^*_0)$		-2.09 Mpc^{-3} mag^{-1} (fid*)	-2.09 Mpc^{-3} mag^{-1} (fid*)
M^*_0		-19.68	-19.68 (fid*)
α		-0.53	-0.53 (fid*)
ϕ^*_{evol}		-1.37	-1.37
M^*_{evol}		-1.15	-1.15
SPIRAL POPULATION			
Hubble type	4	4	4
Bulge to total	0.2	0.2	0.2
Bulge SED[1]	E	E	E
Disk SED[1]	Irr	Irr	Irr
$log_{10}(\phi^*_0)$	-2.01 Mpc^{-3} mag^{-1} (fid*)	-2.04 Mpc^{-3} mag^{-1} (fid*)	-2.04 Mpc^{-3} mag^{-1} (fid*)
M^*_0	-20.0 (fid*)	-19.71	-19.71 (fid*)
α	-1.3	-1.29	-1.29 (fid*)
ϕ^*_{evol}	0	-0.03	-0.03
M^*_{evol}	0	-1.49	-1.49
EXTINCTION			
Extinction curve	Milky Way[2]		
Reddening	0.097 (u'), 0.0745 (g'), 0.049 (r'), 0.036 (i'), 0.027 (z'), 0.016 (J), 0.010 (H), 0.007 (Ks)		
Inter-galactic medium	Madau model[3]		
STARS	Besançon model[4]		
BULGE PARAMETERS[5]			
R_0	1.58 h^{-1} kpc		
M_0	-20.5		
γ_B	-1		
DISK PARAMETERS[6]			
h_0	3.85 h^{-1} kpc		
β	-0.214		
σ	0.36		
γ_D	-0.8		
COMPUTATION TIME			
Training	$\approx$ 6.5 days[7]	$\approx$ 8 days[8]	$\approx$ 20 days[7]
PMC	$\approx$ 2 days	$\approx$ 7 days	$\approx$ 30 days
TRAINING/VALIDATION SET			
Fiducial (n)	200	200	350
Deviation (m)	50	50	100
Total ($n + 2 \times n_{\mathrm{param}} \times m$)	400	400	1550

Notes for Table 9:

ϕ^* is given for $H_0 = 100$ km s^{-1} Mpc^{-1} and $z = 0$.

[*] "fid" indicates that these values are used as fiducial values of the parameters to train the inception network.

[1] The SED templates are taken from Coleman et al. (1980) for the far UV and visible and extended to the near-IR.

[2] The extinction curve is taken from Fitzpatrick and Massa (2007b) and extended to the near-IR.

[3] The extinction curve is corrected for the intergalactic medium effects as in Madau et al. (1996).

[4] Stars are added following the Besançon model of Robin et al. (2003) and their recent improvements (Robin et al. (2012, 2014); Bienaymé et al. (2015); Amôres et al. (2017)).

[5] For a full description of these parameters, see Chap. 3 Sect. 2.2.

[6] For a full description of these parameters, see Chap. 3 Sect. 2.3.

[7] Trained on NVIDIA QUADRO RTX 8000 45GB GPU.

[8] Trained on NVIDIA TITAN X 12Go GPU.

Chapter 7

Improvements and perspectives

Criticism is nice
But what surprises?
If it is not aimed at an improvement
But works as an impediment

So steer out
From fall and talk about
How best you can achieve?
And improve (13-20)

Hasmukh Amathalal

Abstract

This last chapter gives an overview of the possible improvements to make the whole approach faster and more accurate. We discuss the current forward model and we propose a list a refinements in order to generate more realistic multiband deep field images. We also present the improvements that could be applied to the inception network to speed up the training and improve its accuracy. Finally, we discuss the two limitations that may occur with the PMC procedure and we briefly introduce the neural density estimators as a promising tool to efficiently derive likelihood-free posterior distributions.

Contents

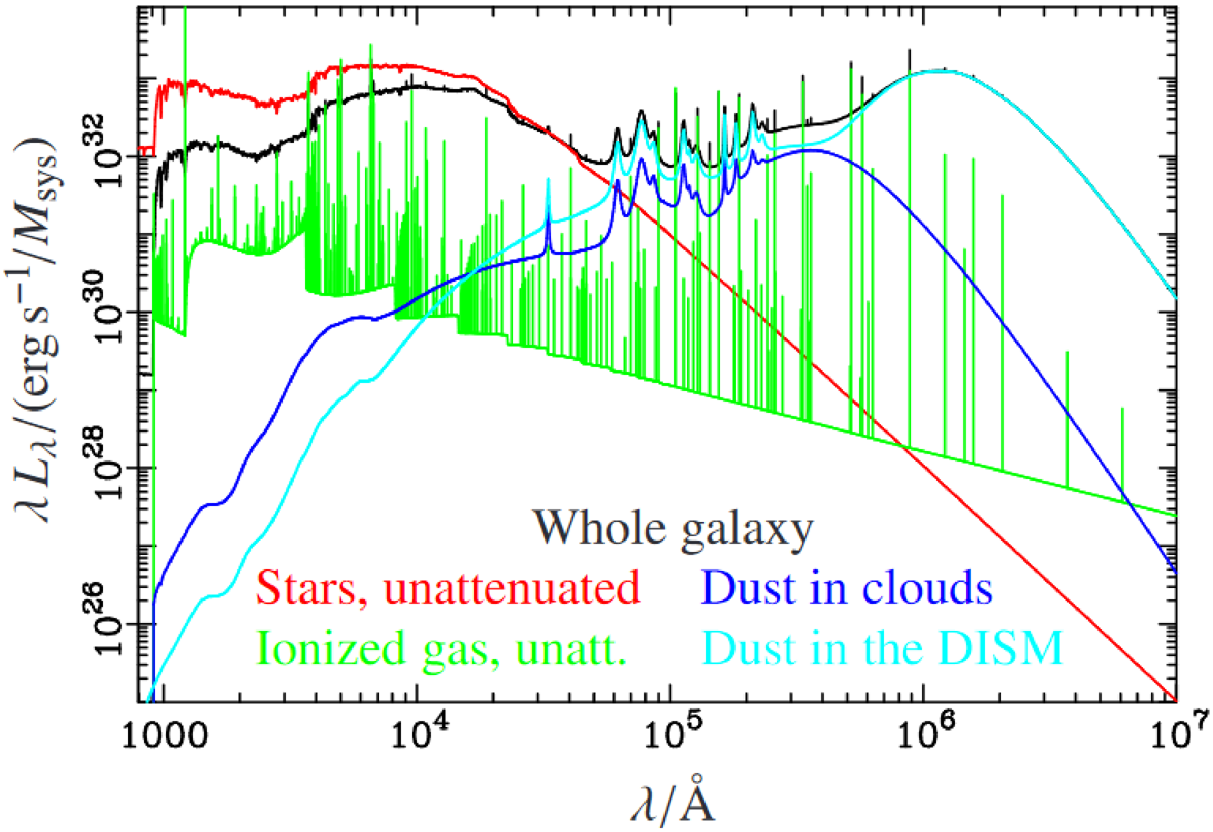


Fig. 7.1: Image from Fioc and Rocca-Volmerange (2019). Spectral energy distribution of their 13 Gyr-old Milky Way model (black line) and of its components: unattenuated nebular continuum and lines produced by the ionized gas (green line); dust emission from star-forming clouds (dark blue line); dust emission from the diffuse Interstellar medium (light blue line). The unattenuated stellar continuum is plotted in red.

1. An even more realistic forward model

In this thesis, we prove that the proposed method is accurate for a forward model with two galaxy populations (elliptical and spiral), but further realism in the simulations is still required, in particular concerning the continuum of galaxy types observed in data images. In a first step, one could include more galaxy types among those that significantly contribute to the observed deep fields, such as lenticular galaxies, and dividing the spirals galaxies into early and late type spirals (with a high bulge-to-total ratio for Sa-Sb types and a low ratio for Sc-Sd types) ; it is not clear whether including irregular galaxies would be useful, as their low surface brightness makes them undetectable except at the closest distances. A larger variety of galaxy types may also bring essential information if these types evolve differently from redshift $z = 1$.

The current forward model uses the non-evolving SEDs of Coleman et al. (1980) to model the bulges and disks of galaxies, see Chap. 3 Sect. 2.1. Even if the evolution of the SEDs of individual galaxies is allowed through the evolution of the type-dependent luminosity functions and the different extinction effects, it can still be improved with evolving scenarios of galaxies such as those of the PÉGASE model of Fioc and Rocca-Volmerange (2019), for the entire galaxies, or separately for their bulges and disks. This model takes into account the ionized gas, the star-forming clouds and the diffuse interstellar medium to model the extincted SEDs of galaxies, see Fig. 7.1. Moreover, with their scenario-based SEDs, the PÉGASE model (Fioc and Rocca-Volmerange 2019) could provide a simple way to model the diversity of galaxy types as described in the previous paragraph.

In addition and as explained in Chap. 3 Sect. 3., the 3D galaxy model that I have developed will be used in the near future to properly model the light of the three component "bulge + thick disk + thin disk" decomposition, which can be attenuated by a two component dust disk associated with old and young stellar populations (see Fig. 3.8). This model takes into account the galaxy properties (such as the inclination angle of the galaxy, the total central opacity, the ratio of the scale length to the scale height of each component, etc) and realistically renders visual effects such as the bulge being partially obscured by the dust disks. This 3D model can also include a bar component that may be needed to realistically render the barred lenticular and spiral galaxies (10-40% of such galaxies have a bar; de Lapparent et al. 2011). Moreover, E. Bertin has recently updated the `SkyMaker` program and it can now be run with raster images of galaxies (or of each individual component: bulge and one/two disks) instead of only painting the galaxies according to previously chosen light distributions. This update could be used extensively with the 3D model described in Chap. 3 Sect. 3. The 3D model can indeed be used to produce small raster images of the bulge, the thick disk and the thin disk, and the updated version of `SkyMaker` can realistically render them (with the PSF, convolution filters, sky background, etc). The right panel of Fig. 3.8 in Chap. 3 Sect. 3. shows such a combination of the 3D model with the updated `SkyMaker` accounting for more realistic visual effect such as the bulge partially obscured by the dust disks.

Although this thesis focuses only on the parameters of the luminosity functions but in principle, the parameters of the forward model could be any parameter that we fixed in Chap. 3. For instance, we could also infer the bulge and disk scale lengths R_e (given by the Binggeli et al. (1984) law of Eq. (3.3) parameterized with R_0 and M_0 in Chap. 3 Sect. 2.2) and h_D (given by the de Jong and Lacey (2000) law of Eq. (3.8) parameterized with h_0, β and σ in Chap. 3 Sect. 2.3); along with their redshift evolution parameters that we fixed in the present thesis to the value $(1+z)^{-1}$ for the bulge and $(1+z)^{-0.8}$ for the disk, but which can be also be varied and inferred from the data. Such an approach is probably necessary to lift the discrepancies in the size evolution of elliptical galaxies with redshift. For instance, Fan et al. (2010) find that the size evolution of massive early-type galaxies occurs rapidly at $z \gtrsim 1$ and becomes negligible at $z < 1$. Damjanov et al. (2011), however, find a continuous size evolution from $z = 2.5$ to $z = 0$, whereas Saracco et al. (2010) find no strong evidence for size growth from $z = 2$ to $z = 0$.

Moreover, the bulge-to-total ratio B/T that is fixed in this thesis to 0.2 for the spiral population could also be inferred from observed data (and a larger number of spiral populations – or additional lenticular populations – could be considered and their B/T ratio also inferred). The dust extinction parameter called "total central opacity" ω of Eq. (3.10) in Chap. 3 Sect. 2.4.1 could also be a variable for which the mean and 1σ standard deviation values could be inferred from the CFHTLS data. At last, some of the parameters defining the PÉGASE scenarios (Fioc and Rocca-Volmerange 2019) may be inferred on. Of course, inferring on more and more parameters requires increasing amounts of computing time for generating the simulations, training the network, and calculating the posterior distributions of the parameters using the PMC. The challenge here will be to discriminate between the parameters that one should in priority infer on, and those that could be fixed to already measured values.

In addition, our current forward model applies the same intergalactic medium attenuation for all galaxies in each redshift bin (Madau et al. 1996). However, considering the improvements of Haardt and Madau (2012), a more realistic stochastic model of the intergalactic medium attenuation could be considered to determine the absorption of light by gas/dust for each individual galaxy depending on the number of gas clouds crossed along their line of sight and at their redshift.

Finally, the forward model uses both the `Stuff` program (Bertin 2010) – that I have entirely rewritten in `Python` programming language and updated as described in Chap. 3 Sect. 2. – and the `SkyMaker` (Bertin 2009) software written in `C` programming language. In order to generate the simulations even faster, both of these codes could be improved: for instance, some of the loops in the `Python` code could be changed into `Numpy` arrays for faster calculations; and `SkyMaker` could be rewritten in `TensorFlow` programming language in order to be executed on GPU to quickly render on the images the galaxies generated by `Stuff`, thus drastically decreasing the computing time.

2. Improving the inception network

Our specific inception architecture shown in Fig. 4.6 and described in Chap. 4 Sect. 2. efficiently compresses the data images into sufficient summary statistics that enable a fast comparison between simulated and observed mutliband deep field images. The main advantage of our inception architecture is the fact that the number of weights is relatively low (*i.e.* ≈ 13000) as compared to classic inception networks with millions of parameters, avoiding too much overfitting by the network. However, the architecture could be slightly refined by the implementation of the residual technique (Szegedy et al. 2016). This consists in adding the input of each inception block to its output in order to stabilize the flow of information and the gradient calculations. Szegedy et al. (2016) give empirical evidence that training with residual connections improves both the training time and the training accuracy of inception networks.

In addition, each image is preconditioned before being passed through our inception network using the inverse hyperbolic sine function $\sinh^{-1}$ described in Chap. 3 Sect. 2.7, in order to reduce the dynamical range of fluxes, and to ease the training procedure. More specifically, this function is applied in the same way for the eight photometric bands of each image, even though the flux intensity of a given galaxy is generally higher (by a factor ≈ 10) in the Ks band than in the u' band. Instead of considering the same function for all photometric bands, we could insert some parameters in the preconditioning function $g(x)$, as follows:

$$g(x) = a \times \sinh^{-1}(bx + c) + d \tag{7.1}$$

where a, b, c, d are parameters that vary for each photometric band (and x is a pixel). Moreover, because we do not know a priori the best values for these a, b, c, d parameters, we can make them weights of the inception network that are updated at each iteration during the training process. By doing so, the network will be able to learn by itself the best way to perform the data conditioning for the various photometric bands.

It is worth noting that the classical techniques to improve the training accuracy and to decrease the computing time such as dropout, batch-training and batch-normalization are not possible here. Dropout (Hinton et al. 2013) is a regularization method which randomly deactivates a fraction of neurons of the network during the training, with the hope that overfitting, hence the discrepancy between the training and the validation sets, are reduced. But the Fisher information needs to be computed exactly in the same way from one iteration to another, in order to maximize its evaluation during the training: this makes the use of dropout irrelevant because the neurons that are activated would change from one iteration to the other. Batch-training is a technique of calculating the loss function and updating the weights only on a portion of the training set in order to multiply the weight updates and therefore to train faster. Batch normalization (Ioffe and Szegedy 2015) is a method of normalizing the inputs (*i.e.* re-centering and re-scaling) of some layers of the network in order to make the training faster and more stable. However, the Fisher information is calculated on the whole training set (or the whole validation set) at each iteration, which makes the batch-training and batch-normalization techniques unfeasible.

Finally, as explained in Chap. 4 Sect. 2.6, because in the current version of the code, the gradients of the forward model with respect to the parameters θ are not available, we perform numerical differentiation to obtain the derivatives of the mean vector of the summary statistics as in Eq. (4.29). However, numerical differentiation, as it is currently performed, is a first order approximation of the true gradients of the forward model. In the next developments of this approach, we could consider to write a new version of the code for the forward model which may then be differentiable so that the gradients may be calculated by exact differentiation (*i.e.* the code automatically computes the exact derivatives of all functions with respect to the parameters of the forward model). This could be done with the use of `TensorFlow`, `JAX` or `Autograd` Python-based programming languages. As a result, the simulations generated with deviations around the fiducial values would not be necessary anymore, only those at fiducial values would be used to train the inception network, making the training much more stable. Moreover, the possibility to make the simulations on the fly for both the training and validation sets would reduce the differences of the covariance and Fisher information matrices between both sets.

3. Alternatives to the PMC sampling procedure

The PMC procedure gives a gigantic gain in computing time as compared to the classic ABC approach, especially when the number of parameters to infer is high. The second to last block of Table 9 gives an idea of the computing time required by the PMC procedures of our three applications. If the PMC has the advantage to efficiently adapt the location of the resampled parameters in order to obtain a higher density of draws in the more probable regions of the posterior distribution, this procedure suffers from several limitations:

- Density estimation becomes dramatically harder as the dimensionality of the model increases often referred to as the **curse of dimensionality** (Lee et al. 2011).
- Proposed new samples with low associated weights are likely to be removed in the next resampling steps, thus yielding a reduction of diversity, this effect is known as **path degeneracy** (Elvira et al. 2017).
- the PMC only adapts the location of the proposed samples, while the scale (*i.e.* the variance of the proposed samples) remains fixed from the beginning of the procedure. This could have a negative effect on the efficiency of the procedure because the resampled parameters could be drawn far from the expected values. Adapting the scale can be a challenge because it requires prior knowledge of the scale of the posterior distribution: this is even more complicated when the posterior parameters are correlated (Elvira et al. 2015).

Several approaches and studies have been developed in order to circumvent these limitations and an extensive review of such attempts is given in Bugallo et al. (2017). In their recent work, Elvira and Chouzenoux (2021) have developed a method which exploits the geometric information of the target distribution, more specifically by adapting the location and scale of a set of temporary density distributions. This method could be adapted to a certain extent to this work to improve the sampling procedure of the PMC.

In addition, we aim at obtaining the highest fidelity posterior distribution with the fewest simulations possible from our forward model. We therefore prefer to generate simulations in the most "interesting" regions of the parameter space, which are not known a priori. To fulfill this objective, the study of Alsing et al. (2019) uses the trained IMNN in order to fit neural density estimators (NDEs) to derive the posterior distribution given the data: NDEs are parameterized neural networks with weights w, which can be trained on a set of simulated summary-parameter pairs (t, θ) to learn the non-linear relations between the summary statistics and the associated parameters. NDEs automatically determine on-the-fly where the best parameters are, based on their current state of knowledge of the target posterior distribution. The IMNN algorithm (Charnock et al. 2018) is particularly relevant in this approach because: (1) by construction IMMN provides Gaussianly distributed compressed summary statistics and a pseudo-maximum likelihood estimator allowing for relatively simple NDEs architecture and requiring fewer simulated data to converge. (2) IMNN also provides the Fisher information matrix of the summary statistics, that can be used directly to initialize the NDEs. In order to train such networks, Alsing et al. (2019) developed a `Python` algorithm called `pydelfi`[*], that could be used in future applications of the method developed in this thesis, therefore avoiding very long PMC searches across the parameter space.

---

[*] `https://github.com/justinalsing/pydelfi`

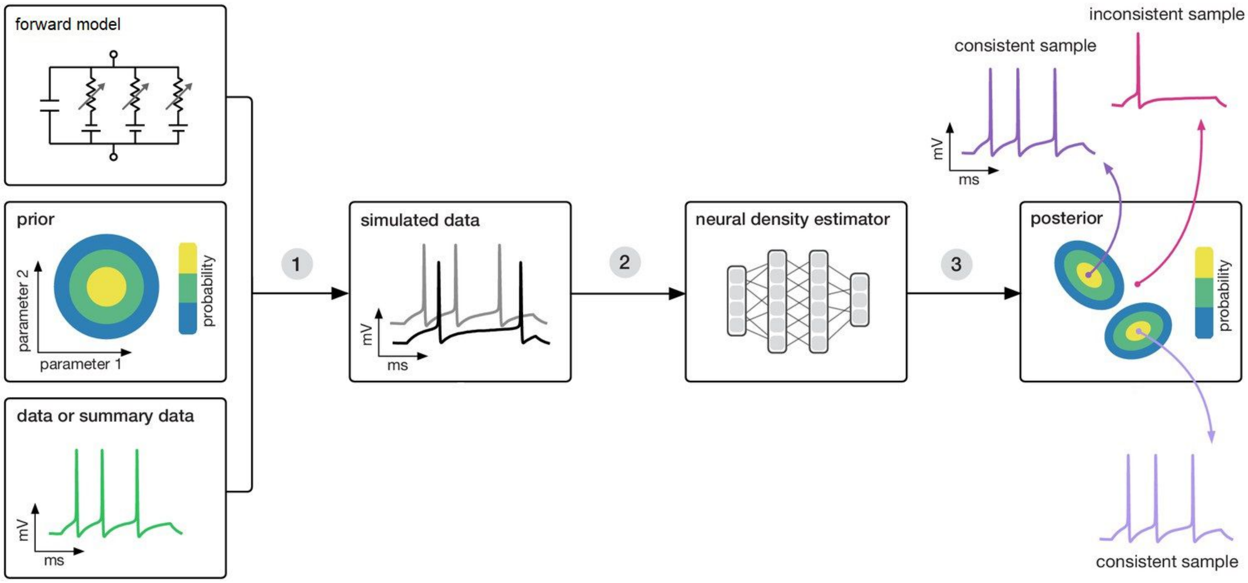


Fig. 7.2: Schematic representation of the neural density estimators (NDEs) from Gonçalves et al. (2019) for which the general concepts are as follow: (1) sampling parameters from the prior, simulating synthetic multiband deep field images from these parameters, and compressing them with the trained IMNN. (2) Using NDEs to learn the non-linear relations between the summary statistics and the model parameters. (3) Applying the trained NDEs to derive the posterior distribution. High posterior probability is assigned to parameters which are consistent with both the data and the prior, low probability to inconsistent parameters.

Conclusion

In the present thesis, we have introduced a novel method to infer robust constraints on the luminosity functions of elliptical and spiral galaxies using a massive compression of panchromatic images through a neural network and likelihood-free (simulation-based) inference. This approach directly analyzes multiband deep field images using an inception network and performs a likelihood-free inference without the need for any source catalog or any explicit galaxy count. The key elements of our approach are: (i) the use of simulated images in similar "observing" conditions as the data, which takes into account the complex selection biases that affect the survey images (see Chap. 2), as a central part of the inference process; (ii) the direct comparison of the simulated images to the observed CFHTLS deep fields.

Synthetic populations of elliptical and spiral galaxies are simulated using a forward model of galaxy evolution, sampled from luminosity functions for each population and decomposed into their bulge and disk through a set of SEDs. The forward model is made even more realistic by adding the internal extinction by dust of each component of the galaxies, the reddening caused by the Milky Way, and the stars from the Besançon stellar model. The `SkyMaker` software is then used to paint these simulated galaxies and stars in realistic panchromatic images in the optical u', g', r', i', z' MegaPrime filters and the near-IR J, H, Ks WIRCam filters, making sure to reproduce the instrumental selection effects of the CFHTLS deep field images.

The images simulated through this process are then used to train a fully convolutional inception network to extract the most informative summary statistics about the parameters of each luminosity function. The network is trained to maximize the Fisher information of the images and enables the drastic reduction in the dimensionality of the images down to the number of parameters of the model. In contrast to deep-learning approaches requiring hundred of thousands to millions of images, training the proposed inception network with relatively few images (only 400 images, that is 200 for estimating the covariance matrix and 200 for estimating the derivative of the mean vector in the applications of Chap. 5 Sect. 4. and Chap. 6 Sect. 1.) is sufficient to obtain very good results. After the network is trained on simulated deep field images, ABC/PMC procedures are run, starting from a uniform prior, to infer the parameters of the luminosity functions that best describe the data images.

This method proves its efficiency in constraining the nontrivial and correlated parameters of two luminosity functions of elliptical and spiral populations of galaxies: the amplitude ϕ^* and the characteristic magnitude M^*. Using the likelihood-free inference and the compressed network summary statistics, we are able to recover the input values of the parameters for virtual data, as well as the values of the model parameters best describing observed CFHTLS D1 data; in both cases, we only use 5 insets of size 1024×1024 pixel of the CFHTLS D1 one square-degree field. We also developed a joint-PMC procedure in order to infer the parameters using multiple 1024×1024 pixel images at once. For the observed data, the derived luminosity functions agree well with those for the elliptical (or red, or quiescent) and the spiral (or blue, or star-forming) galaxy populations obtained from deep multiband galaxy surveys by López-Sanjuan et al. (2017), Brown et al. (2007), Beare et al. (2015), Salimbeni et al. (2008), Drory et al. (2009), Faber et al. (2007), Zucca et al. (2009), and Zucca et al. (2006); the agreement is particularly good for the spiral population. For the elliptical population, information about the excess of faint galaxies that some authors have tried to model as the sum of two Schechter functions is encoded via the IMNN and yields a higher amplitude for this galaxy type.

The results of the inference obtained with our method are limited for the moment to two parameters of the luminosity functions (*i.e.* the density parameters $\phi^*_{0,E}$ and $\phi^*_{0,Sp}$), but could be extended in the near future to six parameters (*i.e.* $\phi^*_{0,E}$, $M^*_{0,E}$, α_E, $\phi^*_{0,Sp}$, $M^*_{0,Sp}$ and α_{Sp}, see Chap. 6 Sect. 2.), and eventually to all the luminosity function parameters (*i.e.* including the evolution parameters $\phi^*_{\mathrm{evol},E}$, $M^*_{\mathrm{evol},E}$, $\phi^*_{\mathrm{evol},Sp}$ and $M^*_{\mathrm{evol},Sp}$). The important advantage of our approach is that we completely bypass the need for galaxy redshifts, whether spectroscopic or the less precise photometric redshifts, as are used in all the other studies mentioned (López-Sanjuan et al. 2017; Brown et al. 2007; Beare et al. 2015; Salimbeni et al. 2008; Drory et al. 2009; Faber et al. 2007; Zucca et al. 2009, 2006). In these studies, the need for redshifts is a major effort beyond the mere multiband photometry of sky fields to characterize galaxy evolution. Photometric redshifts are improved by calibration on a spectroscopic redshift sub-sample (Shuntov et al. 2020), but spectroscopic redshifts requires large amounts of telescope time. Photometric redshifts may also be flawed because of the mismatches between the galaxies and the SED templates, and photometric redshifts suffer from redshift-template degeneracies, which may affect predominantly faint sources near the survey limit (Cunha et al. 2014). We actually show in this thesis how the photometric redshifts and the catalog extraction of magnitudes can bias the derived luminosity functions of elliptical, spiral and irregular populations of galaxies (see Chap. 2 Sect. 5.).

We proved in this thesis that our method allows the likelihood-free inference of luminosity function parameters simultaneously and directly from multiband deep field images, without any knowledge of observed galaxy distances or even galaxy positions in the image, thus circumventing the various and complex selection biases affecting astronomical images. This approach could be applied to analyze the future scheduled photometric surveys with very large samples of galaxies, such as the Euclid space mission (2022), the Legacy Survey of Space and Time (LSST) ground mission (2024) and the James Webb Space Telescope (JWST) space mission (2021). We emphasize that a reliable study of galaxy evolution based on several surveys with complementary redshift intervals, hence covering a wide redshift range (for example the SDSS (York et al. 2000) at $z \sim 0 - 0.1$, the CFHTLS (Hudelot et al. 2012; Bielby et al. 2012) at $z \sim 0.1 - 0.5$ and the Hubble UDF (Beckwith et al. 2006) at $z \sim 0.5 - 2$) would put an even stronger constraint on the need to overcome the selection effects: surveys are often markedly different due to different telescopes, observing conditions, instruments, filters, angular resolutions, etc. Such a multi-survey study would express the full power of our approach.

List of Figures

List of Tables

Bibliography

K. N. Abazajian, J. K. Adelman-McCarthy, M. A. Agüeros, et al. The Seventh Data Release of the Sloan Digital Sky Survey. *ApJS*, 182(2):543–558, June 2009. doi: 10.1088/0067-0049/182/2/543.

B. Abolfathi, D. S. Aguado, G. Aguilar, et al. The Fourteenth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the Extended Baryon Oscillation Spectroscopic Survey and from the Second Phase of the Apache Point Observatory Galactic Evolution Experiment. *ApJS*, 235(2):42, Apr. 2018. doi: 10.3847/1538-4365/aa9e8a.

J. A. L. Aguerri, M. Balcells, and R. F. Peletier. Growth of galactic bulges by mergers. I. Dense satellites. *A&A*, 367:428–442, Feb. 2001. doi: 10.1051/0004-6361:20000441.

J. Akeret, A. Refregier, A. Amara, et al. Approximate Bayesian computation for forward modeling in cosmology. *J. Cosmology Astropart. Phys.*, 2015(8):043, Aug. 2015. doi: 10.1088/1475-7516/2015/08/043.

S. Alam, F. D. Albareti, C. Allende Prieto, et al. The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III. *ApJS*, 219(1):12, July 2015. doi: 10.1088/0067-0049/219/1/12.

E. Almoznino, F. Loinger, and N. Brosch. A Procedure for the Calculation of Background in Images. *MNRAS*, 265:641, Dec. 1993. doi: 10.1093/mnras/265.3.641.

J. Alsing and B. Wandelt. Generalized massive optimal data compression. *MNRAS*, 476(1):L60–L64, May 2018. doi: 10.1093/mnrasl/sly029.

J. Alsing, T. Charnock, S. Feeney, and B. Wandelt. Fast likelihood-free cosmology with neural density estimators and active learning. *MNRAS*, 488(3):4440–4458, Sept. 2019. doi: 10.1093/mnras/stz1960.

E. B. Amôres, A. C. Robin, and C. Reylé. Evolution over time of the Milky Way's disc shape. *A&A*, 602: A67, June 2017. doi: 10.1051/0004-6361/201628461.

S. Arnouts and O. Ilbert. LePHARE: Photometric Analysis for Redshift Estimate, Aug. 2011.

P. Astier, J. Guy, N. Regnault, et al. The Supernova Legacy Survey: measurement of Ω_M, Ω_Λ and w from the first year data set. *A&A*, 447(1):31–48, Feb. 2006. doi: 10.1051/0004-6361:20054185.

W. B. Atwood, A. A. Abdo, M. Ackermann, et al. The Large Area Telescope on the Fermi Gamma-Ray Space Telescope Mission. *ApJ*, 697(2):1071–1102, June 2009. doi: 10.1088/0004-637X/697/2/1071.

J. Bailin and W. E. Harris. Inclination-Independent Galaxy Classification. *ApJ*, 681(1):225–231, July 2008. doi: 10.1086/588091.

A. Baillard, E. Bertin, V. de Lapparent, et al. The EFIGI catalogue of 4458 nearby galaxies with detailed morphology. *A&A*, 532:A74, Aug. 2011. doi: 10.1051/0004-6361/201016423.

R. Beare, M. J. I. Brown, K. Pimbblet, et al. The $z > 1.2$ Optical Luminosity Function from a Sample of 410,000 Galaxies in Bootes. *ApJ*, 815(2):94, Dec. 2015. doi: 10.1088/0004-637X/815/2/94.

R. H. Becker, R. L. White, and D. J. Helfand. The FIRST Survey: Faint Images of the Radio Sky at Twenty Centimeters. *ApJ*, 450:559, Sept. 1995. doi: 10.1086/176166.

S. V. W. Beckwith, M. Stiavelli, A. M. Koekemoer, et al. The Hubble Ultra Deep Field. *AJ*, 132(5): 1729–1755, Nov. 2006. doi: 10.1086/507302.

E. F. Bell and R. S. de Jong. The stellar populations of spiral galaxies. *MNRAS*, 312(3):497–520, Mar. 2000. doi: 10.1046/j.1365-8711.2000.03138.x.

M. Bernardi, J. L. Fischer, R. K. Sheth, et al. Comparing pymorph and SDSS photometry - II. The differences are more than semantics and are not dominated by intracluster light. *MNRAS*, 468(3):2569–2581, July 2017. doi: 10.1093/mnras/stx677.

M. Bernyk, D. J. Croton, C. Tonini, et al. The Theoretical Astrophysical Observatory: Cloud-based Mock Galaxy Catalogs. *ApJS*, 223(1):9, Mar. 2016. doi: 10.3847/0067-0049/223/1/9.

E. Bertin. SkyMaker: astronomical image simulations made easy. *Mem. Soc. Astron. Italiana*, 80:422, Jan. 2009.

E. Bertin. Stuff: Simulating "Perfect" Astronomical Catalogues, Oct. 2010.

E. Bertin. Automated Morphometry with SExtractor and PSFEx. In I. N. Evans, A. Accomazzi, D. J. Mink, and A. H. Rots, editors, *Astronomical Data Analysis Software and Systems XX*, volume 442 of *Astronomical Society of the Pacific Conference Series*, page 435, July 2011.

E. Bertin. Displaying Digital Deep Sky Images. In P. Ballester, D. Egret, and N. P. F. Lorente, editors, *Astronomical Data Analysis Software and Systems XXI*, volume 461 of *Astronomical Society of the Pacific Conference Series*, page 263, Sept. 2012.

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

E. Bertin, M. Schefer, N. Apostolakos, et al. The SourceXtractor++ Software. In R. Pizzo, E. R. Deul, J. D. Mol, et al., editors, *Astronomical Society of the Pacific Conference Series*, volume 527 of *Astronomical Society of the Pacific Conference Series*, page 461, Jan. 2020.

R. Bielby, P. Hudelot, H. J. McCracken, et al. The WIRCam Deep Survey. I. Counts, colours, and mass-functions derived from near-infrared imaging in the CFHTLS deep fields. *A&A*, 545:A23, Sept. 2012. doi: 10.1051/0004-6361/201118547.

O. Bienaymé, A. C. Robin, and B. Famaey. Quasi integral of motion for axisymmetric potentials. *A&A*, 581:A123, Sept. 2015. doi: 10.1051/0004-6361/201526516.

A. Bijaoui. Sky background estimation and application. *A&A*, 84(1-2):81–84, Apr. 1980.

B. Binggeli, A. Sandage, and M. Tarenghi. Studies of the Virgo Cluster. I. Photometry of 109 galaxies near the cluster center to serve as standards. *AJ*, 89:64–82, Jan. 1984. doi: 10.1086/113484.

J. Blaizot, Y. Wadadekar, B. Guiderdoni, et al. MoMaF: the Mock Map Facility. *MNRAS*, 360(1):159–175, June 2005. doi: 10.1111/j.1365-2966.2005.09019.x.

M. R. Blanton. Galaxies in SDSS and DEEP2: A Quiet Life on the Blue Sequence? *ApJ*, 648(1):268–280, Sept. 2006. doi: 10.1086/505628.

M. R. Blanton, D. W. Hogg, N. A. Bahcall, et al. The Galaxy Luminosity Function and Luminosity Density at Redshift z = 0.1. *ApJ*, 592(2):819–838, Aug. 2003. doi: 10.1086/375776.

M. R. Blanton, R. H. Lupton, D. J. Schlegel, et al. The Properties and Luminosity Function of Extremely Low Luminosity Galaxies. *ApJ*, 631(1):208–230, Sept. 2005a. doi: 10.1086/431416.

M. R. Blanton, D. J. Schlegel, M. A. Strauss, et al. New York University Value-Added Galaxy Catalog: A Galaxy Catalog Based on New Public Surveys. *AJ*, 129(6):2562–2578, June 2005b. doi: 10.1086/429803.

T. Boroson. The distribution of luminosity in spiral galaxies. *ApJS*, 46:177–209, June 1981. doi: 10.1086/190742.

L. Bottou. Online algorithms and stochastic approximations. In D. Saad, editor, *Online Learning and Neural Networks*. Cambridge University Press, Cambridge, UK, 1998. URL `http://leon.bottou.org/papers/bottou-98x`. revised, oct 2012.

F. Bournaud, B. G. Elmegreen, and D. M. Elmegreen. Rapid Formation of Exponential Disks and Bulges at High Redshift from the Dynamical Evolution of Clump-Cluster and Chain Galaxies. *ApJ*, 670(1):237–248, Nov. 2007. doi: 10.1086/522077.

B. C. Bromley, W. H. Press, H. Lin, and R. P. Kirshner. Spectral Classification and Luminosity Function of Galaxies in the Las Campanas Redshift Survey. *ApJ*, 505(1):25–36, Sept. 1998. doi: 10.1086/306144.

M. J. I. Brown, A. Dey, B. T. Jannuzi, et al. The Evolving Luminosity Function of Red Galaxies. *ApJ*, 654 (2):858–877, Jan. 2007. doi: 10.1086/509652.

M. F. Bugallo, V. Elvira, L. Martino, et al. Adaptive Importance Sampling: The past, the present, and the future. *IEEE Signal Processing Magazine*, 34(4):60–79, July 2017. doi: 10.1109/MSP.2017.2699226.

D. Burstein. Structure and origin of S0 galaxies. II. Disk-to-bulge ratios. *ApJ*, 234:435–447, Dec. 1979. doi: 10.1086/157512.

V. Calvi, M. Stiavelli, L. Bradley, et al. The Effect of Surface Brightness Dimming in the Selection of High-z Galaxies. *ApJ*, 796(2):102, Dec. 2014. doi: 10.1088/0004-637X/796/2/102.

O. Cappé, A. Guillin, J. M. Marin, and C. P. Robert. Population monte carlo. *Journal of Computational and Graphical Statistics*, 13(4):907–929, 2004. doi: 10.1198/106186004X12803. URL `https://doi.org/10.1198/106186004X12803`.

S. Carassou, V. de Lapparent, E. Bertin, and D. Le Borgne. Inferring the photometric and size evolution of galaxies from image simulations. I. Method. *A&A*, 605:A9, Sept. 2017. doi: 10.1051/0004-6361/201730587.

A. Cattaneo, I. Koutsouridou, E. Tollet, et al. GalICS 2.1: a new semianalytic model for cold accretion, cooling, feedback, and their roles in galaxy formation. *MNRAS*, 497(1):279–301, Sept. 2020. doi: 10.1093/mnras/staa1832.

S. Charlot and S. M. Fall. A Simple Model for the Absorption of Starlight by Dust in Galaxies. *ApJ*, 539 (2):718–731, Aug. 2000. doi: 10.1086/309250.

T. Charnock, G. Lavaux, and B. D. Wandelt. Automatic physical inference with information maximizing neural networks. *Phys. Rev. D*, 97(8):083004, Apr. 2018. doi: 10.1103/PhysRevD.97.083004.

J. Chevallard, S. Charlot, B. Wandelt, and V. Wild. Insights into the content and spatial distribution of dust from the integrated spectral properties of galaxies. *MNRAS*, 432(3):2061–2091, July 2013. doi: 10.1093/mnras/stt523.

D. C. Cireşan, U. Meier, L. M. Gambardella, and J. Schmidhuber. Deep, big, simple neural nets for handwritten digit recognition. *Neural Computation*, 22(12):3207–3220, 2010. doi: 10.1162/NECO_a_00052. URL `https://doi.org/10.1162/NECO_a_00052`. PMID: 20858131.

J. Cisewski-Kehe, G. Weller, and C. Schafer. A preferential attachment model for the stellar initial mass function. *Electronic Journal of Statistics*, 13(1):1580 – 1607, 2019. doi: 10.1214/19-EJS1556. URL `https://doi.org/10.1214/19-EJS1556`.

J. Cisewski-Kehe, G. Weller, and C. Schafer. A Preferential Attachment Model for the Stellar Initial Mass Function. *arXiv e-prints*, art. arXiv:1904.11306, Apr. 2019.

S. Cole, A. Aragon-Salamanca, C. S. Frenk, et al. A recipe for galaxy formation. *MNRAS*, 271:781–806, Dec. 1994. doi: 10.1093/mnras/271.4.781.

S. Cole, C. G. Lacey, C. M. Baugh, and C. S. Frenk. Hierarchical galaxy formation. *MNRAS*, 319(1): 168–204, Nov. 2000. doi: 10.1046/j.1365-8711.2000.03879.x.

G. D. Coleman, C. C. Wu, and D. W. Weedman. Colors and magnitudes predicted for high redshift galaxies. *ApJS*, 43:393–416, July 1980. doi: 10.1086/190674.

P. Coles and F. Lucchin. *Cosmology. The origin and evolution of cosmic structure.* 1995.

M. Colless, R. S. Ellis, K. Taylor, and R. N. Hook. The LDSS deep redshift survey. *MNRAS*, 244:408, June 1990.

J. J. Condon. Confusion and Flux-Density Error Distributions. *ApJ*, 188:279–286, Mar. 1974. doi: 10.1086/152714.

J. J. Condon, W. D. Cotton, E. W. Greisen, et al. The NRAO VLA Sky Survey. *AJ*, 115(5):1693–1716, May 1998. doi: 10.1086/300337.

C. J. Conselice, A. Wilkinson, K. Duncan, and A. Mortlock. The Evolution of Galaxy Number Density at z $<$ 8 and Its Implications. *ApJ*, 830(2):83, Oct. 2016. doi: 10.3847/0004-637X/830/2/83.

M. Cousin, G. Lagache, M. Bethermin, and B. Guiderdoni. Towards a new modelling of gas flows in a semi-analytical model of galaxy formation and evolution. *A&A*, 575:A33, Mar. 2015. doi: 10.1051/0004-6361/201424462.

M. Cousin, P. Guillard, and M. D. Lehnert. G.A.S.. I. A prescription for turbulence-regulated star formation and its impact on galaxy properties. *A&A*, 627:A131, July 2019. doi: 10.1051/0004-6361/201834673.

R. A. Crain, J. Schaye, R. G. Bower, et al. The EAGLE simulations of galaxy formation: calibration of subgrid physics and model variations. *MNRAS*, 450(2):1937–1961, June 2015. doi: 10.1093/mnras/stv725.

J. C. Cuillandre and E. Bertin. CFHT Legacy Survey (CFHTLS) : a rich data set. In D. Barret, F. Casoli, G. Lagache, et al., editors, *SF2A-2006: Semaine de l'Astrophysique Francaise*, page 265, June 2006.

C. E. Cunha, D. Huterer, H. Lin, et al. Spectroscopic failures in photometric redshift calibration: cosmological biases and survey requirements. *MNRAS*, 444(1):129–146, Oct. 2014. doi: 10.1093/mnras/stu1424.

G. Cybenko. Approximation by superpositions of a sigmoidal function. *Math. Control Signal Systems*, 2: 303–314, 1989. doi: 10.1007/BF02551274. URL `https://doi.org/10.1007/BF025512742`.

T. Dahlen, B. Mobasher, S. M. Faber, et al. A Critical Assessment of Photometric Redshift Methods: A CANDELS Investigation. *ApJ*, 775(2):93, Oct. 2013. doi: 10.1088/0004-637X/775/2/93.

D. A. Dale, S. A. Cohen, L. C. Johnson, et al. The Spitzer Local Volume Legacy: Survey Description and Infrared Photometry. *ApJ*, 703(1):517–556, Sept. 2009. doi: 10.1088/0004-637X/703/1/517.

I. Damjanov, P. J. McCarthy, R. G. Abraham, et al. Red Nuggets at z ~1.5: Compact Passive Galaxies and the Formation of the Kormendy Relation. *ApJ*, 695(1):101–115, Apr. 2009. doi: 10.1088/0004-637X/695/1/101.

I. Damjanov, R. G. Abraham, K. Glazebrook, et al. Red Nuggets at High Redshift: Structural Evolution of Quiescent Galaxies Over 10 Gyr of Cosmic History. *ApJ*, 739(2):L44, Oct. 2011. doi: 10.1088/2041-8205/739/2/L44.

R. Davé, D. Anglés-Alcázar, D. Narayanan, et al. SIMBA: Cosmological simulations with black hole growth and feedback. *MNRAS*, 486(2):2827–2849, June 2019. doi: 10.1093/mnras/stz937.

M. Davis, S. M. Faber, J. Newman, et al. Science Objectives and Early Results of the DEEP2 Redshift Survey. In P. Guhathakurta, editor, *Discoveries and Research Prospects from 6- to 10-Meter-Class Telescopes II*, volume 4834 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 161–172, Feb. 2003. doi: 10.1117/12.457897.

M. Davis, P. Guhathakurta, N. P. Konidaris, et al. The All-Wavelength Extended Groth Strip International Survey (AEGIS) Data Sets. *ApJ*, 660(1):L1–L6, May 2007. doi: 10.1086/517931.

K. S. Dawson, D. J. Schlegel, C. P. Ahn, et al. The Baryon Oscillation Spectroscopic Survey of SDSS-III. *AJ*, 145(1):10, Jan. 2013. doi: 10.1088/0004-6256/145/1/10.

R. S. de Jong. Near-infrared and optical broadband surface photometry of 86 face-on disk dominated galaxies. IV. Using color profiles to study stellar and dust content of galaxies. *A&A*, 313:377–395, Sept. 1996.

R. S. de Jong and C. Lacey. The Local Space Density of SB-SDM Galaxies as Function of Their Scale Size, Surface Brightness, and Luminosity. *ApJ*, 545(2):781–797, Dec. 2000. doi: 10.1086/317840.

V. de Lapparent, A. Baillard, and E. Bertin. The EFIGI catalogue of 4458 nearby galaxies with morphology. II. Statistical properties along the Hubble sequence. *A&A*, 532:A75, Aug. 2011. doi: 10.1051/0004-6361/201016424.

G. de Vaucouleurs. On the distribution of mass and luminosity in elliptical galaxies. *MNRAS*, 113:134, Jan. 1953. doi: 10.1093/mnras/113.2.134.

G. de Vaucouleurs. Classification and Morphology of External Galaxies. *Handbuch der Physik*, 53:275, Jan. 1959. doi: 10.1007/978-3-642-45932-0_7.

G. de Vaucouleurs. Structure of the Virgo Cluster of Galaxies. *ApJS*, 6:213, May 1961. doi: 10.1086/190064.

G. de Vaucouleurs, A. de Vaucouleurs, J. Corwin, Herold G., et al. *Third Reference Catalogue of Bright Galaxies*. 1991.

P. Del Moral, A. Doucet, and A. Jasra. An adaptive sequential monte carlo method for approximate bayesian computation. *Statistics and Computing*, 22(5):1009–1020, Sep 2012. ISSN 1573-1375. doi: 10.1007/s11222-011-9271-y. URL `https://doi.org/10.1007/s11222-011-9271-y`.

B. M. Devour and E. F. Bell. Global dust attenuation in disc galaxies: strong variation with specific star formation and stellar mass, and the importance of sample selection. *MNRAS*, 459(2):2054–2077, June 2016. doi: 10.1093/mnras/stw754.

M. Dickinson, M. Giavalisco, and GOODS Team. The Great Observatories Origins Deep Survey. In R. Bender and A. Renzini, editors, *The Mass of Galaxies at Low and High Redshift*, page 324, Jan. 2003. doi: 10.1007/10899892_78.

S. G. Djorgovski, A. Mahabal, A. Drake, et al. *Sky Surveys*, page 223. 2013. doi: 10.1007/978-94-007-5618-2_5.

H. Dole, G. Lagache, and J. L. Puget. Predictions for Cosmological Infrared Surveys from Space with the Multiband Imaging Photometer for SIRTF. *ApJ*, 585(2):617–629, Mar. 2003. doi: 10.1086/346130.

B. T. Draine. Interstellar Dust Models and Evolutionary Implications. In T. Henning, E. Grün, and J. Steinacker, editors, *Cosmic Dust - Near and Far*, volume 414 of *Astronomical Society of the Pacific Conference Series*, page 453, Dec. 2009.

S. P. Driver and S. Phillipps. Is the Luminosity Distribution of Field Galaxies Really Flat? *ApJ*, 469:529, Oct. 1996. doi: 10.1086/177801.

S. P. Driver, C. C. Popescu, R. J. Tuffs, et al. The Millennium Galaxy Catalogue: the B-band attenuation of bulge and disc light and the implied cosmic dust and stellar mass densities. *MNRAS*, 379(3):1022–1036, Aug. 2007. doi: 10.1111/j.1365-2966.2007.11862.x.

N. Drory, K. Bundy, A. Leauthaud, et al. The Bimodal Galaxy Stellar Mass Function in the COSMOS Survey to z ~1: A Steep Faint End and a New Galaxy Dichotomy. *ApJ*, 707(2):1595–1609, Dec. 2009. doi: 10.1088/0004-637X/707/2/1595.

Y. Dubois, C. Pichon, C. Welker, et al. Dancing in the dark: galactic properties trace spin swings along the cosmic web. *MNRAS*, 444(2):1453–1468, Oct. 2014. doi: 10.1093/mnras/stu1227.

A. S. Eddington. On a formula for correcting statistics for the effects of a known error of observation. *MNRAS*, 73:359–360, Mar. 1913. doi: 10.1093/mnras/73.5.359.

S. Eddington, A. S. The correction of statistics for accidental error. *MNRAS*, 100:354, Mar. 1940. doi: 10.1093/mnras/100.5.354.

V. Elvira and E. Chouzenoux. Optimized Population Monte Carlo. working paper or preprint, Mar. 2021. URL `https://hal.archives-ouvertes.fr/hal-03136318`.

V. Elvira, L. Martino, D. Luengo, and J. Corander. A gradient adaptive population importance sampler. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4075–4079, 2015. doi: 10.1109/ICASSP.2015.7178737.

V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo. Improving population monte carlo: Alternative weighting and resampling schemes. *Signal Processing*, 131:77–91, 2017. ISSN 0165-1684. doi: https://doi.org/10.1016/j.sigpro.2016.07.012. URL `https://www.sciencedirect.com/science/article/pii/S0165168416301633`.

S. M. Faber, C. N. A. Willmer, C. Wolf, et al. Galaxy Luminosity Functions to z~1 from DEEP2 and COMBO-17: Implications for Red Galaxy Formation. *ApJ*, 665(1):265–294, Aug. 2007. doi: 10.1086/519294.

L. Fan, A. Lapi, A. Bressan, et al. Cosmic Evolution of Size and Velocity Dispersion for Early-type Galaxies. *ApJ*, 718(2):1460–1475, Aug. 2010. doi: 10.1088/0004-637X/718/2/1460.

X. Fan, C. L. Carilli, and B. Keating. Observational Constraints on Cosmic Reionization. *ARA&A*, 44(1): 415–462, Sept. 2006. doi: 10.1146/annurev.astro.44.051905.092514.

C.-A. Faucher-Giguère, A. Lidz, and L. Hernquist. Numerical simulations unravel the cosmic web. *Science*, 319(5859):52–55, 2008. ISSN 0036-8075. doi: 10.1126/science.1151476. URL `https://science.sciencemag.org/content/319/5859/52`.

M. Fioc and B. Rocca-Volmerange. PÉGASE.3: A code for modeling the UV-to-IR/submm spectral and chemical evolution of galaxies with dust. *A&A*, 623:A143, Mar. 2019. doi: 10.1051/0004-6361/201833556.

E. L. Fitzpatrick and D. Massa. An Analysis of the Shapes of Ultraviolet Extinction Curves. III. an Atlas of Ultraviolet Extinction Curves. *ApJS*, 72:163, Jan. 1990. doi: 10.1086/191413.

E. L. Fitzpatrick and D. Massa. An Analysis of the Shapes of Interstellar Extinction Curves. V. The IR-through-UV Curve Morphology. *ApJ*, 663(1):320–341, July 2007a. doi: 10.1086/518158.

E. L. Fitzpatrick and D. Massa. An Analysis of the Shapes of Interstellar Extinction Curves. V. The IR-through-UV Curve Morphology. *ApJ*, 663(1):320–341, July 2007b. doi: 10.1086/518158.

M. Fukugita, K. Shimasaku, and T. Ichikawa. Galaxy Colors in Various Photometric Band Systems. *PASP*, 107:945, Oct. 1995. doi: 10.1086/133643.

A. Gabasch, R. Bender, S. Seitz, et al. The evolution of the luminosity functions in the FORS Deep Field from low to high redshift. I. The blue bands. *A&A*, 421:41–58, July 2004. doi: 10.1051/0004-6361:20035909.

A. Gabrielli, F. Sylos Labini, M. Joyce, and L. Pietronero. *Statistical Physics for Cosmic Structures*. 2005. doi: 10.1007/b138455.

D. A. Gadotti. Structural properties of pseudo-bulges, classical bulges and elliptical galaxies: a Sloan Digital Sky Survey perspective. *MNRAS*, 393(4):1531–1552, Mar. 2009. doi: 10.1111/j.1365-2966.2008.14257.x.

G. Gamow. The Origin of Elements and the Separation of Galaxies. *Physical Review*, 74(4):505–506, Aug. 1948. doi: 10.1103/PhysRev.74.505.2.

R. Giacconi, P. Rosati, P. Tozzi, et al. First Results from the X-Ray and Optical Survey of the Chandra Deep Field South. *ApJ*, 551(2):624–634, Apr. 2001. doi: 10.1086/320222.

M. Giavalisco, H. C. Ferguson, A. M. Koekemoer, et al. The Great Observatories Origins Deep Survey: Initial Results from Optical and Near-Infrared Imaging. *ApJ*, 600(2):L93–L98, Jan. 2004. doi: 10.1086/379232.

R. Giovanelli, M. P. Haynes, J. J. Salzer, et al. Dependence on Luminosity of Photometric Properties of Disk Galaxies: Surface Brightness, Size, and Internal Extinction. *AJ*, 110:1059, Sept. 1995. doi: 10.1086/117586.

X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In Y. W. Teh and M. Titterington, editors, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *Proceedings of Machine Learning Research*, pages 249–256, Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010. PMLR. URL `http://proceedings.mlr.press/v9/glorot10a.html`.

X. Glorot, A. Bordes, and Y. Bengio. Deep sparse rectifier neural networks. In G. Gordon, D. Dunson, and M. Dudík, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 315–323, Fort Lauderdale, FL, USA, 11–13 Apr 2011. PMLR. URL `http://proceedings.mlr.press/v15/glorot11a.html`.

P. J. Gonçalves, J.-M. Lueckmann, M. Deistler, et al. Training deep neural density estimators to identify mechanistic models of neural dynamics. *bioRxiv*, 2019. doi: 10.1101/838383. URL `https://www.biorxiv.org/content/early/2019/11/12/838383`.

I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

A. W. Graham and C. C. Worley. Inclination- and dust-corrected galaxy parameters: bulge-to-disc ratios and size-luminosity relations. *MNRAS*, 388(4):1708–1728, Aug. 2008. doi: 10.1111/j.1365-2966.2008.13506.x.

F. Haardt and P. Madau. Radiative Transfer in a Clumpy Universe. IV. New Synthesis Models of the Cosmic UV/X-Ray Background. *ApJ*, 746(2):125, Feb. 2012. doi: 10.1088/0004-637X/746/2/125.

W. Handley and P. Lemos. Quantifying tensions in cosmological parameters: Interpreting the DES evidence ratio. *Phys. Rev. D*, 100(4):043504, Aug. 2019. doi: 10.1103/PhysRevD.100.043504.

J. W. Hardy. *Adaptive Optics for Astronomical Telescopes*. 1998.

R. Harmon and G. Mamon. The Detection of Galaxies in Infrared Surveys. In B. T. Soifer, editor, *Sky Surveys. Protostars to Protogalaxies*, volume 43 of *Astronomical Society of the Pacific Conference Series*, page 15, Jan. 1993.

S. Hatton, J. E. G. Devriendt, S. Ninin, et al. GALICS- I. A hybrid N-body/semi-analytic model of hierarchical galaxy formation. *MNRAS*, 343(1):75–106, July 2003. doi: 10.1046/j.1365-8711.2003.05589.x.

A. F. Heavens, R. Jimenez, and O. Lahav. Massive lossless data compression and multiple parameter estimation from galaxy spectra. *MNRAS*, 317(4):965–972, Oct. 2000. doi: 10.1046/j.1365-8711.2000.03692.x.

H. Hildebrandt, T. Erben, K. Kuijken, et al. CFHTLenS: improving the quality of photometric redshifts with precision photometry. *MNRAS*, 421(3):2355–2367, Apr. 2012. doi: 10.1111/j.1365-2966.2012.20468.x.

G. E. Hinton, K. A., S. I., and S. I. System and method for addressing overfitting in a neural network. art. USS PATENT: US9406017B2, 2013.

D. W. Hogg. Confusion Errors in Astrometry and Counterpart Association. *AJ*, 121(2):1207–1213, Feb. 2001. doi: 10.1086/318736.

D. W. Hogg, I. K. Baldry, M. R. Blanton, and D. J. Eisenstein. The K correction. *arXiv e-prints*, art. astro-ph/0210394, Oct. 2002.

E. Holmberg. A photographic photometry of extragalactic nebulae. *Meddelanden fran Lunds Astronomiska Observatorium Serie II*, 136:1, Jan. 1958.

P. F. Hopkins, A. Wetzel, D. Kereš, et al. FIRE-2 simulations: physics versus numerics in galaxy formation. *MNRAS*, 480(1):800–863, Oct. 2018. doi: 10.1093/mnras/sty1690.

L. Hoppmann, L. Staveley-Smith, W. Freudling, et al. A blind H I mass function from the Arecibo Ultra-Deep Survey (AUDS). *MNRAS*, 452(4):3726–3741, Oct. 2015. doi: 10.1093/mnras/stv1084.

K. Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991. ISSN 0893-6080. doi: https://doi.org/10.1016/0893-6080(91)90009-T. URL `https://www.sciencedirect.com/science/article/pii/089360809190009T`.

M. Hoskin. *The Cambridge Concise History of Astronomy*. 1999.

E. Hubble. No. 324. Extra-galactic nebulae. *Contributions from the Mount Wilson Observatory / Carnegie Institution of Washington*, 324:1–49, Jan. 1926a.

E. P. Hubble. Extragalactic nebulae. *ApJ*, 64:321–369, Dec. 1926b. doi: 10.1086/143018.

E. P. Hubble. *Realm of the Nebulae*. 1936.

J. P. Huchra, L. M. Macri, K. L. Masters, et al. The 2MASS Redshift Survey—Description and Data Release. *ApJS*, 199(2):26, Apr. 2012. doi: 10.1088/0067-0049/199/2/26.

P. Hudelot, J. C. Cuillandre, K. Withington, et al. VizieR Online Data Catalog: The CFHTLS Survey (T0007 release) (Hudelot+ 2012). *VizieR Online Data Catalog*, art. II/317, Sept. 2012.

C.-L. Hwang, A. S. M. M. Masud, and S. R. Paidy. *Multiple objective decision making : methods and applications : a state-of-the-art survey*. Berlin : Springer, 1979. ISBN 3540091114. URL `http://lib.ugent.be/catalog/rug01:000011811`.

O. Ilbert, S. Arnouts, H. J. McCracken, et al. Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey. *A&A*, 457(3):841–856, Oct. 2006. doi: 10.1051/0004-6361:20065138.

G. D. Illingworth, D. Magee, P. A. Oesch, et al. The HST eXtreme Deep Field (XDF): Combining All ACS and WFC3/IR Data on the HUDF Region into the Deepest Field Ever. *ApJS*, 209(1):6, Nov. 2013. doi: 10.1088/0067-0049/209/1/6.

S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, page 448–456. JMLR.org, 2015.

B. T. Jannuzi and A. Dey. The NOAO Deep Wide-Field Survey. In R. Weymann, L. Storrie-Lombardi, M. Sawicki, and R. Brunner, editors, *Photometric Redshifts and the Detection of High Redshift Galaxies*, volume 191 of *Astronomical Society of the Pacific Conference Series*, page 111, Jan. 1999.

T. H. Jarrett, T. Chester, R. Cutri, et al. 2MASS Extended Source Catalog: Overview and Algorithms. *AJ*, 119(5):2498–2531, May 2000. doi: 10.1086/301330.

J. F. Jarvis and J. A. Tyson. Focas: faint object classification and analysis system. *AJ*, 86:476–495, Mar. 1981. doi: 10.1086/112907.

L. R. Jones, R. Fong, T. Shanks, et al. Galaxy number counts. - I. Photographic observations to B=23.5 mag. *MNRAS*, 249:481–497, Apr. 1991. doi: 10.1093/mnras/249.3.481.

P. Jonsson. SUNRISE: polychromatic dust radiative transfer in arbitrary geometries. *MNRAS*, 372(1):2–20, Oct. 2006. doi: 10.1111/j.1365-2966.2006.10884.x.

P. Jonsson, B. Groves, and T. J. Cox. High-Resolution Panchromatic Spectral Models of Galaxies including Photoionisation and Dust. *arXiv e-prints*, art. arXiv:0906.2156, June 2009.

T. Kacprzak, J. Herbel, A. Amara, and A. Réfrégier. Accelerating Approximate Bayesian Computation with Quantile Regression: application to cosmological redshift distributions. *J. Cosmology Astropart. Phys.*, 2018(2):042, Feb. 2018. doi: 10.1088/1475-7516/2018/02/042.

N. Kashikawa, K. Shimasaku, N. Yasuda, et al. The Subaru Deep Field: The Optical Imaging Data. *PASJ*, 56:1011–1023, Dec. 2004. doi: 10.1093/pasj/56.6.1011.

G. Kauffmann, S. D. M. White, and B. Guiderdoni. The formation and evolution of galaxies within merging dark matter haloes. *MNRAS*, 264:201–218, Sept. 1993. doi: 10.1093/mnras/264.1.201.

J. Kennicutt, Robert C. The Global Schmidt Law in Star-forming Galaxies. *ApJ*, 498(2):541–552, May 1998. doi: 10.1086/305588.

D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. *arXiv e-prints*, art. arXiv:1412.6980, Dec. 2014.

K. Kiwiel. Convergence and efficiency of subgradient methods for quasiconvex minimization. *Math. Program. 90, 1–25*, 2001. doi: 10.1007/PL00011414. URL `https://doi.org/10.1007/PL00011414`.

A. M. Koekemoer, S. M. Faber, H. C. Ferguson, et al. CANDELS: The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey—The Hubble Space Telescope Observations, Imaging Data Products, and Mosaics. *ApJS*, 197(2):36, Dec. 2011. doi: 10.1088/0067-0049/197/2/36.

J. Kormendy. *Secular Evolution in Disk Galaxies*, page 1. 2013.

J. Kormendy and J. Kennicutt, Robert C. Secular Evolution and the Formation of Pseudobulges in Disk Galaxies. *ARA&A*, 42(1):603–683, Sept. 2004. doi: 10.1146/annurev.astro.42.053102.134024.

J. Kormendy, N. Drory, R. Bender, and M. E. Cornell. Bulgeless Giant Galaxies Challenge Our Picture of Galaxy Formation by Hierarchical Clustering. *ApJ*, 723(1):54–80, Nov. 2010. doi: 10.1088/0004-637X/723/1/54.

A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012. URL `http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf`.

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

C. G. Lacey, C. M. Baugh, C. S. Frenk, et al. A unified multiwavelength model of galaxy formation. *MNRAS*, 462(4):3854–3911, Nov. 2016. doi: 10.1093/mnras/stw1888.

C. d. P. Lagos, R. J. Tobar, A. S. G. Robotham, et al. Shark: introducing an open source, free, and flexible semi-analytic model of galaxy formation. *MNRAS*, 481(3):3573–3603, Dec. 2018. doi: 10.1093/mnras/sty2440.

R. R. Laher, V. Gorjian, L. M. Rebull, et al. Aperture Photometry Tool. *PASP*, 124(917):737, July 2012. doi: 10.1086/666883.

D. Le Borgne and B. Rocca-Volmerange. Photometric redshifts from evolutionary synthesis with PÉGASE: The code Z-PEG and the z=0 age constraint. *A&A*, 386:446–455, May 2002. doi: 10.1051/0004-6361:20020259.

O. Le Fèvre, G. Vettolani, B. Garilli, et al. The VIMOS VLT deep survey. First epoch VVDS-deep survey: 11 564 spectra with 17.5 ≤ IAB ≤ 24, and the redshift distribution over 0 ≤ z ≤ 5. *A&A*, 439(3):845–862, Sept. 2005. doi: 10.1051/0004-6361:20041960.

J. Lee, R. McVinish, and K. Mengersen. Population Monte Carlo Algorithm in High Dimensions. *Springer*, Methodol Comput Appl Probab 13:369–389, June 2011. doi: 10.1007/s11009-009-9154-2. URL `https://doi.org/10.1007/s11009-009-9154-2`.

E. L. Lehmann and G. Casella. *Theory of Point Estimation.* Springer, 1983. doi: 10.1007/b98854. URL `https://doi.org/10.1007/b98854`.

S. J. Lilly, L. Tresse, F. Hammer, et al. The Canada-France Redshift Survey. VI. Evolution of the Galaxy Luminosity Function to Z approximately 1. *ApJ*, 455:108, Dec. 1995. doi: 10.1086/176560.

L. Lin, D. R. Patton, D. C. Koo, et al. The Redshift Evolution of Wet, Dry, and Mixed Galaxy Mergers from Close Galaxy Pairs in the DEEP2 Galaxy Redshift Survey. *ApJ*, 681(1):232–243, July 2008. doi: 10.1086/587928.

C. J. Lintott, K. Schawinski, A. Slosar, et al. Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *MNRAS*, 389(3):1179–1189, Sept. 2008. doi: 10.1111/j.1365-2966.2008.13689.x.

C. T. Liu, P. Capak, B. Mobasher, et al. The Faint-End Slopes of Galaxy Luminosity Functions in the COSMOS Field. *ApJ*, 672(1):198–206, Jan. 2008. doi: 10.1086/522361.

F. Livet, T. Charnock, D. Le Borgne, and V. de Lapparent. Catalog-free modeling of galaxy types in deep images. Massive dimensional reduction with neural networks. *A&A*, 652:A62, Aug. 2021. doi: 10.1051/0004-6361/202140383.

C. López-Sanjuan, E. Tempel, N. Benítez, et al. The ALHAMBRA survey: B-band luminosity function of quiescent and star-forming galaxies at 0.2 ≤ z < 1 by PDF analysis. *A&A*, 599:A62, Mar. 2017. doi: 10.1051/0004-6361/201629517.

P. W. Lucas, M. G. Hoare, A. Longmore, et al. The UKIDSS Galactic Plane Survey. *MNRAS*, 391(1):136–163, Nov. 2008. doi: 10.1111/j.1365-2966.2008.13924.x.

A. L. Maas. Rectifier nonlinearities improve neural network acoustic models. 2013.

M.-M. Mac Low and R. S. Klessen. Control of star formation by supersonic turbulence. *Reviews of Modern Physics*, 76(1):125–194, Jan. 2004. doi: 10.1103/RevModPhys.76.125.

C. J. MacDonald and G. Bernstein. Photometric Redshift Biases from Galaxy Evolution. *PASP*, 122(890):485, Apr. 2010. doi: 10.1086/651702.

P. Madau and M. Dickinson. Cosmic Star-Formation History. *ARA&A*, 52:415–486, Aug. 2014. doi: 10.1146/annurev-astro-081811-125615.

P. Madau, H. C. Ferguson, M. E. Dickinson, et al. High-redshift galaxies in the Hubble Deep Field: colour selection and star formation history to z~4. *MNRAS*, 283(4):1388–1404, Dec. 1996. doi: 10.1093/mnras/283.4.1388.

S. J. Maddox, W. J. Sutherland, G. Efstathiou, and J. Loveday. The APM galaxy survey - I. APM measurements and star-galaxy separation. *MNRAS*, 243:692–712, Apr. 1990.

G. K. Malmquist. A study of the stars of spectral type A. *Meddelanden fran Lunds Astronomiska Observatorium Serie II*, 22:3–69, Mar. 1920.

K. G. Malmquist. On some relations in stellar statistics. *Meddelanden fran Lunds Astronomiska Observatorium Serie I*, 100:1–52, Mar. 1922.

K. G. Malmquist. A contribution to the problem of determining the distribution in space of the stars. *Meddelanden fran Lunds Astronomiska Observatorium Serie I*, 106:1–12, Feb. 1925.

D. C. Martin, J. Fanson, D. Schiminovich, et al. The Galaxy Evolution Explorer: A Space Ultraviolet Survey Mission. *ApJ*, 619(1):L1–L6, Jan. 2005. doi: 10.1086/426387.

R. O. Marzke. *The Galaxy Luminosity Function at Zero Redshift: Constraints on Galaxy Formation*, volume 231 of *Astrophysics and Space Science Library*, page 23. 1998. doi: 10.1007/978-94-011-4960-0_3.

R. O. Marzke, M. J. Geller, J. P. Huchra, and J. Corwin, Harold G. The Luminosity Function for Different Morphological Types in the CFA Redshift Survey. *AJ*, 108:437, Aug. 1994. doi: 10.1086/117081.

R. O. Marzke, L. N. da Costa, P. S. Pellegrini, et al. The Galaxy Luminosity Function at Z <= 0.05: Dependence on Morphology. *ApJ*, 503(2):617–631, Aug. 1998. doi: 10.1086/306011.

K. L. Masters, R. Giovanelli, and M. P. Haynes. Internal Extinction in Spiral Galaxies in the Near-Infrared. *AJ*, 126(1):158–174, July 2003. doi: 10.1086/375758.

K. L. Masters, R. Nichol, S. Bamford, et al. Galaxy Zoo: dust in spiral galaxies. *MNRAS*, 404(2):792–810, May 2010. doi: 10.1111/j.1365-2966.2010.16335.x.

J. S. Mathis, W. Rumpl, and K. H. Nordsieck. The size distribution of interstellar grains. *ApJ*, 217:425–433, Oct. 1977. doi: 10.1086/155591.

C. McBride, A. Berlind, R. Scoccimarro, et al. LasDamas Mock Galaxy Catalogs for SDSS. In *American Astronomical Society Meeting Abstracts #213*, volume 213 of *American Astronomical Society Meeting Abstracts*, page 425.06, Jan. 2009.

W. S. McCulloch and W. Pitts. *A logical calculus of the ideas immanent in nervous activity.* 1943.

Y. Mellier and G. Mathez. Deprojection of the de Vaucouleurs R exp 1/4 brightness profile. *A&A*, 175(1-2): 1–3, Mar. 1987.

K. J. Mighell. Algorithms for CCD Stellar Photometry. In D. M. Mehringer, R. L. Plante, and D. A. Roberts, editors, *Astronomical Data Analysis Software and Systems VIII*, volume 172 of *Astronomical Society of the Pacific Conference Series*, page 317, Jan. 1999.

R. Miles. A light history of photometry: from Hipparchus to the Hubble Space Telescope. *Journal of the British Astronomical Association*, 117:172–186, Aug. 2007.

M. Moles, N. Benítez, J. A. L. Aguerri, et al. The Alhambra Survey: a Large Area Multimedium-Band Optical and Near-Infrared Photometric Survey. *AJ*, 136(3):1325–1339, Sept. 2008. doi: 10.1088/0004-6256/136/3/1325.

W. W. Morgan. A Preliminary Classification of the Forms of Galaxies According to Their Stellar Population. *PASP*, 70(415):364, Aug. 1958. doi: 10.1086/127243.

T. Nagayama, P. A. Woudt, C. Nagashima, et al. A deep near-infrared survey around the giant radio galaxy PKS 1343-601. *MNRAS*, 354(4):980–990, Nov. 2004. doi: 10.1111/j.1365-2966.2004.08277.x.

P. B. Nair and R. G. Abraham. A Catalog of Detailed Visual Morphological Classifications for 14,034 Galaxies in the Sloan Digital Sky Survey. *ApJS*, 186(2):427–456, Feb. 2010. doi: 10.1088/0067-0049/186/2/427.

V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, page 807–814, Madison, WI, USA, 2010. Omnipress. ISBN 9781605589077.

O. Nakamura, M. Fukugita, N. Yasuda, et al. The Luminosity Function of Morphologically Classified Galaxies in the Sloan Digital Sky Survey. *AJ*, 125(4):1682–1688, Apr. 2003. doi: 10.1086/368135.

K. Nandra, E. S. Laird, K. Adelberger, et al. A deep Chandra survey of the Groth Strip - I. The X-ray data. *MNRAS*, 356(2):568–586, Jan. 2005. doi: 10.1111/j.1365-2966.2004.08475.x.

M. Negrello, M. Magliocchetti, L. Moscardini, et al. Confusion noise at far-infrared to millimetre wavelengths. *MNRAS*, 352(2):493–500, Aug. 2004. doi: 10.1111/j.1365-2966.2004.07938.x.

D. Nelson, A. Pillepich, V. Springel, et al. First results from the IllustrisTNG simulations: the galaxy colour bimodality. *MNRAS*, 475(1):624–647, Mar. 2018. doi: 10.1093/mnras/stx3040.

G. Neugebauer, H. J. Habing, R. van Duinen, et al. The Infrared Astronomical Satellite (IRAS) mission. *ApJ*, 278:L1–L6, Mar. 1984. doi: 10.1086/184209.

J. A. Newman, M. C. Cooper, M. Davis, et al. The DEEP2 Galaxy Redshift Survey: Design, Observations, Data Reduction, and Redshifts. *ApJS*, 208(1):5, Sept. 2013. doi: 10.1088/0067-0049/208/1/5.

P. Norberg, S. Cole, C. M. Baugh, et al. The 2dF Galaxy Redshift Survey: the b_J-band galaxy luminosity function and survey selection function. *MNRAS*, 336(3):907–931, Nov. 2002. doi: 10.1046/j.1365-8711.2002.05831.x.

K.-S. Oh and K. Jung. Gpu implementation of neural networks. *Pattern Recognition*, 37(6):1311 – 1314, 2004. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2004.01.013. URL `http://www.sciencedirect.com/science/article/pii/S0031320304000524`.

J. B. Oke and J. E. Gunn. Secondary standard stars for absolute spectrophotometry. *ApJ*, 266:713–717, Mar. 1983. doi: 10.1086/160817.

M. Paillassa and E. Bertin. Deblending in Crowded Star Fields Using Convolutional Neural Networks. In M. Molinaro, K. Shortridge, and F. Pasian, editors, *Astronomical Data Analysis Software and Systems XXVI*, volume 521 of *Astronomical Society of the Pacific Conference Series*, page 382, Oct. 2019.

A. B. Palei. Integrating Photometers. *AZh*, 45:207, Jan. 1968.

C. P. Pearson, S. Serjeant, S. Oyabu, et al. The first source counts at 18 μm from the AKARI NEP Survey. *MNRAS*, 444(1):846–859, Oct. 2014. doi: 10.1093/mnras/stu1472.

P. J. E. Peebles. *The large-scale structure of the universe.* 1980.

A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *ApJ*, 142: 419–421, July 1965. doi: 10.1086/148307.

I. Pérez and P. Sánchez-Blázquez. Study of stellar populations in the bulges of barred galaxies. *A&A*, 529: A64, May 2011. doi: 10.1051/0004-6361/201015002.

A. Pillepich, V. Springel, D. Nelson, et al. Simulating galaxy formation with the IllustrisTNG model. *MNRAS*, 473(3):4077–4106, Jan. 2018. doi: 10.1093/mnras/stx2656.

Planck Collaboration, P. A. R. Ade, N. Aghanim, et al. Planck 2013 results. XVI. Cosmological parameters. *A&A*, 571:A16, Nov. 2014. doi: 10.1051/0004-6361/201321591.

Planck Collaboration, P. A. R. Ade, N. Aghanim, et al. Planck 2015 results. XIII. Cosmological parameters. *A&A*, 594:A13, Sept. 2016. doi: 10.1051/0004-6361/201525830.

C. C. Popescu and R. J. Tuffs. Modelling the ultraviolet/submillimeter spectral energy distributions of normal galaxies. In C. C. Popescu and R. J. Tuffs, editors, *The Spectral Energy Distributions of Gas-Rich Galaxies: Confronting Models with Data*, volume 761 of *American Institute of Physics Conference Series*, pages 155–177, Apr. 2005. doi: 10.1063/1.1913927.

C. C. Popescu, R. J. Tuffs, M. A. Dopita, et al. Modelling the spectral energy distribution of galaxies. V. The dust and PAH emission SEDs of disk galaxies. *A&A*, 527:A109, Mar. 2011. doi: 10.1051/0004-6361/201015217.

R. F. Quadri and R. J. Williams. Quantifying Photometric Redshift Errors in the Absence of Spectroscopic Redshifts. *ApJ*, 725(1):794–802, Dec. 2010. doi: 10.1088/0004-637X/725/1/794.

N. Reid and G. Gilmore. New light on faint stars. II. A photometric study of the low luminosity main sequence. *MNRAS*, 201:73–94, Oct. 1982. doi: 10.1093/mnras/201.1.73.

A. G. Riess, A. V. Filippenko, P. Challis, et al. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *AJ*, 116(3):1009–1038, Sept. 1998. doi: 10.1086/300499.

A. C. Robin, C. Reylé, S. Derrière, and S. Picaud. A synthetic view on structure and evolution of the Milky Way. *A&A*, 409:523–540, Oct. 2003. doi: 10.1051/0004-6361:20031117.

A. C. Robin, D. J. Marshall, M. Schultheis, and C. Reylé. Stellar populations in the Milky Way bulge region: towards solving the Galactic bulge and bar shapes using 2MASS data. *A&A*, 538:A106, Feb. 2012. doi: 10.1051/0004-6361/201116512.

A. C. Robin, C. Reylé, J. Fliri, et al. Constraining the thick disc formation scenario of the Milky Way. *A&A*, 569:A13, Sept. 2014. doi: 10.1051/0004-6361/201423415.

F. Rosenblatt. *The Perceptron—a perceiving and recognizing automaton.* 1957.

D. B. Rubin. Bayesianly justifiable and relevant frequency calculations for the applies statistician. *The Annals of Statistics*, 12(4):1151–1172, 1984. ISSN 00905364. URL `http://www.jstor.org/stable/2240995`.

H. G. E. Rumelhart D. E. and W. R. J. Learning representations by back-propagating errors. *Nature, 323(6088), 533–536*, 1986. doi: 10.1038/323533a0. URL `https://doi.org/10.1038/323533a0`.

S. Sachdeva and K. Saha. Survival of Pure Disk Galaxies over the Last 8 Billion Years. *ApJ*, 820(1):L4, Mar. 2016. doi: 10.3847/2041-8205/820/1/L4.

A. Saintonge, L. J. Tacconi, S. Fabello, et al. The Impact of Interactions, Bars, Bulges, and Active Galactic Nuclei on Star Formation Efficiency in Local Massive Galaxies. *ApJ*, 758(2):73, Oct. 2012. doi: 10.1088/0004-637X/758/2/73.

S. Salimbeni, E. Giallongo, N. Menci, et al. The red and blue galaxy populations in the GOODS field: evidence for an excess of red dwarfs. *A&A*, 477(3):763–773, Jan. 2008. doi: 10.1051/0004-6361:20077959.

A. Sandage. *The Hubble Atlas of Galaxies.* 1961.

A. Sandage, K. C. Freeman, and N. R. Stokes. The Intrinsic Flattening of e, so, and Spiral Galaxies as Related to Galaxy Formation and Evolution. *ApJ*, 160:831, June 1970. doi: 10.1086/150475.

A. Sandage, G. A. Tammann, and A. Yahil. The velocity field of bright nearby galaxies. I. The variation of mean absolute magnitude with redshift for galaxies in a magnitude-limited sample. *ApJ*, 232:352–364, Sept. 1979. doi: 10.1086/157295.

P. Saracco, M. Longhetti, and A. Gargiulo. The number density of superdense early-type galaxies at $1 < z < 2$ and the local cluster galaxies. *MNRAS*, 408(1):L21–L25, Oct. 2010. doi: 10.1111/j.1745-3933.2010.00920.x.

K. Schawinski, C. M. Urry, B. D. Simmons, et al. The green valley is a red herring: Galaxy Zoo reveals two evolutionary pathways towards quenching of star formation in early- and late-type galaxies. *MNRAS*, 440(1):889–907, May 2014. doi: 10.1093/mnras/stu327.

P. Schechter. An analytic expression for the luminosity function for galaxies. *ApJ*, 203:297–306, Jan. 1976. doi: 10.1086/154079.

M. Schmidt. The Rate of Star Formation. *ApJ*, 129:243, Mar. 1959. doi: 10.1086/146614.

M. Schmidt. Space Distribution and Luminosity Functions of Quasi-Stellar Radio Sources. *ApJ*, 151:393, Feb. 1968. doi: 10.1086/149446.

N. Scoville, H. Aussel, M. Brusa, et al. The Cosmic Evolution Survey (COSMOS): Overview. *ApJS*, 172(1): 1–8, Sept. 2007. doi: 10.1086/516585.

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

S. Shen, H. J. Mo, S. D. M. White, et al. The size distribution of galaxies in the Sloan Digital Sky Survey. *MNRAS*, 343(3):978–994, Aug. 2003. doi: 10.1046/j.1365-8711.2003.06740.x.

R. K. Sheth. On estimating redshift and luminosity distributions in photometric redshift surveys. *MNRAS*, 378(2):709–715, June 2007. doi: 10.1111/j.1365-2966.2007.11812.x.

M. Shuntov, J. Pasquet, S. Arnouts, et al. PhotoWeb redshift: boosting photometric redshift accuracy with large spectroscopic surveys. *A&A*, 636:A90, Apr. 2020. doi: 10.1051/0004-6361/201937382.

L. Silva, G. L. Granato, A. Bressan, and L. Danese. Modeling the Effects of Dust on Galactic Spectral Energy Distributions from the Ultraviolet to the Millimeter Band. *ApJ*, 509(1):103–117, Dec. 1998. doi: 10.1086/306476.

L. Simard and R. S. Somerville. Testing Hierarchical Galaxy Evolution Models with the Luminosity-Size Relation of Distant Field Galaxies. In F. Combes, G. A. Mamon, and V. Charmandaris, editors, *Dynamics of Galaxies: from the Early Universe to the Present*, volume 197 of *Astronomical Society of the Pacific Conference Series*, page 439, Jan. 2000.

L. Simard, C. N. A. Willmer, N. P. Vogt, et al. The DEEP Groth Strip Survey. II. Hubble Space Telescope Structural Parameters of Galaxies in the Groth Strip. *ApJS*, 142(1):1–33, Sept. 2002. doi: 10.1086/341399.

R. S. Somerville and R. Davé. Physical Models of Galaxy Formation in a Cosmological Framework. *ARA&A*, 53:51–113, Aug. 2015. doi: 10.1146/annurev-astro-082812-140951.

R. S. Somerville and J. R. Primack. Semi-analytic modelling of galaxy formation: the local Universe. *MNRAS*, 310(4):1087–1110, Dec. 1999. doi: 10.1046/j.1365-8711.1999.03032.x.

R. S. Somerville, R. C. Gilmore, J. R. Primack, and A. Domínguez. Galaxy properties from the ultraviolet to the far-infrared: Λ cold dark matter models confront observations. *MNRAS*, 423(3):1992–2015, July 2012. doi: 10.1111/j.1365-2966.2012.20490.x.

R. S. Somerville, G. Popping, and S. C. Trager. Star formation in semi-analytic galaxy formation models with multiphase gas. *MNRAS*, 453(4):4337–4367, Nov. 2015. doi: 10.1093/mnras/stv1877.

V. Springel, S. D. M. White, A. Jenkins, et al. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435(7042):629–636, June 2005. doi: 10.1038/nature03597.

V. Springel, R. Pakmor, A. Pillepich, et al. First results from the IllustrisTNG simulations: matter and galaxy clustering. *MNRAS*, 475(1):676–698, Mar. 2018. doi: 10.1093/mnras/stx3304.

P. B. Stetson. DAOPHOT: A Computer Program for Crowded-Field Stellar Photometry. *PASP*, 99:191, Mar. 1987. doi: 10.1086/131977.

C. Szegedy, W. Liu, Y. Jia, et al. Going deeper with convolutions. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

C. Szegedy, V. Vanhoucke, S. Ioffe, et al. Rethinking the Inception Architecture for Computer Vision. *arXiv e-prints*, art. arXiv:1512.00567, Dec. 2015.

C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *arXiv e-prints*, art. arXiv:1602.07261, Feb. 2016.

M. Taghizadeh-Popp, S. M. Fall, R. L. White, and A. S. Szalay. Simulating Deep Hubble Images with Semi-empirical Models of Galaxy Formation. *ApJ*, 801(1):14, Mar. 2015. doi: 10.1088/0004-637X/801/1/14.

M. Tegmark, M. R. Blanton, M. A. Strauss, et al. The Three-Dimensional Power Spectrum of Galaxies from the Sloan Digital Sky Survey. *ApJ*, 606(2):702–740, May 2004. doi: 10.1086/382125.

R. C. Tolman. On the Estimation of Distances in a Curved Universe with a Non-Static Line Element. *Proceedings of the National Academy of Science*, 16(7):511–520, July 1930. doi: 10.1073/pnas.16.7.511.

R. C. Tolman. *Relativity, Thermodynamics, and Cosmology.* 1934.

L. Tortorelli, M. Fagioli, J. Herbel, et al. Measurement of the B-band galaxy Luminosity Function with Approximate Bayesian Computation. *J. Cosmology Astropart. Phys.*, 2020(9):048, Sept. 2020. doi: 10.1088/1475-7516/2020/09/048.

I. Trujillo, N. M. Förster Schreiber, G. Rudnick, et al. The Size Evolution of Galaxies since z~3: Combining SDSS, GEMS, and FIRES. *ApJ*, 650(1):18–41, Oct. 2006. doi: 10.1086/506464.

R. J. Tuffs, C. C. Popescu, H. J. Völk, et al. Modelling the spectral energy distribution of galaxies. III. Attenuation of stellar light in spiral galaxies. *A&A*, 419:821–835, June 2004. doi: 10.1051/0004-6361:20035689.

R. B. Tully and J. R. Fisher. Reprint of 1977A&A....54..661T. A new method of determining distance to galaxies. *A&A*, 500:105–117, Feb. 1977.

R. B. Tully, M. J. Pierce, J.-S. Huang, et al. Global Extinction in Spiral Galaxies. *AJ*, 115(6):2264–2272, June 1998. doi: 10.1086/300379.

C. T. Unterborn and B. S. Ryden. Inclination-Dependent Extinction Effects in Disk Galaxies in the Sloan Digital Sky Survey. *ApJ*, 687(2):976–985, Nov. 2008. doi: 10.1086/591898.

E. A. Valentijn. Opaque spiral galaxies. *Nature*, 346(6280):153–155, July 1990. doi: 10.1038/346153a0.

R. F. J. van der Burg, H. Hildebrandt, and T. Erben. The UV galaxy luminosity function at z = 3-5 from the CFHT Legacy Survey Deep fields. *A&A*, 523:A74, Nov. 2010. doi: 10.1051/0004-6361/200913812.

P. C. van der Kruit. The thickness of the hydrogen layer and the three-dimensional mass distribution in NGC 891. *A&A*, 99(2):298–304, June 1981.

P. C. van der Kruit and K. C. Freeman. Galaxy Disks. *ARA&A*, 49(1):301–371, Sept. 2011. doi: 10.1146/annurev-astro-083109-153241.

M. Vogelsberger, S. Genel, V. Springel, et al. Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe. *MNRAS*, 444(2):1518–1547, Oct. 2014. doi: 10.1093/mnras/stu1536.

W. Voges, B. Aschenbach, T. Boller, et al. The ROSAT all-sky survey bright source catalogue. *A&A*, 349:389–405, Sept. 1999.

J. Wang, G. De Lucia, M. G. Kitzbichler, and S. D. M. White. The dependence of galaxy formation on cosmological parameters: can we distinguish between the WMAP1 and WMAP3 parameter sets? *MNRAS*, 384(4):1301–1315, Mar. 2008. doi: 10.1111/j.1365-2966.2007.12797.x.

S. Weinberg. *Cosmology.* 2008.

A. Weyant, C. Schafer, and W. M. Wood-Vasey. Likelihood-free Cosmological Inference with Type Ia Supernovae: Approximate Bayesian Computation for a Complete Treatment of Uncertainty. *ApJ*, 764(2):116, Feb. 2013. doi: 10.1088/0004-637X/764/2/116.

S. D. M. White and C. S. Frenk. Galaxy Formation through Hierarchical Clustering. *ApJ*, 379:52, Sept. 1991. doi: 10.1086/170483.

R. Wielen. The Diffusion of Stellar Orbits Derived from the Observed Age-Dependence of the Velocity Dispersion. *A&A*, 60(2):263–275, Sept. 1977.

K. W. Willett, C. J. Lintott, S. P. Bamford, et al. Galaxy Zoo 2: detailed morphological classifications for 304 122 galaxies from the Sloan Digital Sky Survey. *MNRAS*, 435(4):2835–2860, Nov. 2013. doi: 10.1093/mnras/stt1458.

R. E. Williams, B. Blacker, M. Dickinson, et al. The Hubble Deep Field: Observations, Data Reduction, and Galaxy Photometry. *AJ*, 112:1335, Oct. 1996. doi: 10.1086/118105.

R. J. Williams, R. F. Quadri, M. Franx, et al. The Evolving Relations Between Size, Mass, Surface Density, and Star Formation in 3×10^4 Galaxies Since z = 2. *ApJ*, 713(2):738–750, Apr. 2010. doi: 10.1088/0004-637X/713/2/738.

C. N. A. Willmer, S. M. Faber, D. C. Koo, et al. The Deep Evolutionary Exploratory Probe 2 Galaxy Redshift Survey: The Galaxy Luminosity Function to z~1. *ApJ*, 647(2):853–873, Aug. 2006. doi: 10.1086/505455.

D. M. Wittman, J. A. Tyson, I. P. Dell'Antonio, et al. Deep lens survey. In J. A. Tyson and S. Wolff, editors, *Survey and Other Telescope Technologies and Discoveries*, volume 4836 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 73–82, Dec. 2002. doi: 10.1117/12.457348.

A. E. Wright, M. R. Griffith, B. F. Burke, and R. D. Ekers. The Parkes-MIT-NRAO (PMN) Surveys. II. Source Catalog for the Southern Survey (-87 degrees -4pt.5 < delta < -37 degrees). *ApJS*, 91:111, Mar. 1994. doi: 10.1086/191939.

T. K. Wyder, D. C. Martin, D. Schiminovich, et al. The UV-Optical Galaxy Color-Magnitude Diagram. I. Basic Properties. *ApJS*, 173(2):293–314, Dec. 2007. doi: 10.1086/521402.

H. Xi, L. Staveley-Smith, B.-Q. For, et al. The Arecibo Ultra-Deep Survey. *MNRAS*, 501(3):4550–4564, Mar. 2021. doi: 10.1093/mnras/staa3931.

E. M. Xilouris, Y. I. Byun, N. D. Kylafis, et al. Are spiral galaxies optically thin or thick? *A&A*, 344: 868–878, Apr. 1999.

R. Yan, M. White, and A. L. Coil. Mock Catalogs for the DEEP2 Redshift Survey. *ApJ*, 607(2):739–750, June 2004. doi: 10.1086/383588.

H. K. C. Yee. A Faint-Galaxy Photometry and Image-Analysis System. *PASP*, 103:396, Apr. 1991. doi: 10.1086/132834.

C.-W. Yip, A. S. Szalay, S. Carliles, and T. Budavári. Effect of Inclination of Galaxies on Photometric Redshift. *ApJ*, 730(1):54, Mar. 2011. doi: 10.1088/0004-637X/730/1/54.

P. Yoachim and J. J. Dalcanton. Structural Parameters of Thin and Thick Disks in Edge-on Disk Galaxies. *AJ*, 131(1):226–249, Jan. 2006. doi: 10.1086/497970.

D. G. York, J. Adelman, J. Anderson, John E., et al. The Sloan Digital Sky Survey: Technical Summary. *AJ*, 120(3):1579–1587, Sept. 2000. doi: 10.1086/301513.

P. J. Young. Tables of functions for a spherical galaxy obeying the 7-1/1 law in projection. *AJ*, 81:807, Oct. 1976. doi: 10.1086/111959.

J. D. Younger, P. F. Hopkins, T. J. Cox, and L. Hernquist. The Self-Regulated Growth of Supermassive Black Holes. *ApJ*, 686(2):815–828, Oct. 2008. doi: 10.1086/591639.

E. Zucca, O. Ilbert, S. Bardelli, et al. The VIMOS VLT Deep Survey. Evolution of the luminosity functions by galaxy type up to z = 1.5 from first epoch data. *A&A*, 455(3):879–890, Sept. 2006. doi: 10.1051/0004-6361:20053645.

E. Zucca, S. Bardelli, M. Bolzonella, et al. The zCOSMOS survey: the role of the environment in the evolution of the luminosity function of different galaxy types. *A&A*, 508(3):1217–1234, Dec. 2009. doi: 10.1051/0004-6361/200912665.

M. A. Zwaan, M. J. Meyer, R. L. Webster, et al. The HIPASS catalogue - II. Completeness, reliability and parameter accuracy. *MNRAS*, 350(4):1210–1219, June 2004. doi: 10.1111/j.1365-2966.2004.07782.x.

Appendix

Contents

A. Comparison of new model for internal dust extinction with the model of Popescu et al. (2011)

The attenuation of the magnitude of galaxy bulge and disk by Popescu et al. (2011) follows the same procedure as in the previous work of Tuffs et al. (2004), where they fit the attenuation curves Δm versus the inclination angle i (*i.e.* the difference between extincted and non-extincted magnitudes as a function of i):

$$\Delta m = \begin{cases} \sum\limits_{j=0}^{n} a_j (1-\cos i)^j & \text{for } 1-\cos i \leq 0.9 \\ b_0 & \text{for } 0.9 \leq 1-\cos i \leq 0.95 \\ b_1 & \text{for } 0.95 \leq 1-\cos i \leq 1 \end{cases} \tag{8.1}$$

where $n = 5$ for the disk and $n = 4$ for the bulge; each a_j, b_0 and b_1 are coefficients depending on the total central opacity ω in the B-band and on the wavelength of the considered filter given in Tables 4-6 of Tuffs et al. (2004) and updated in Popescu et al. (2011). In their work, the whole SED of each component is attenuated by a modeled SED of dust that takes into account the composition of the dust, the metallicity of the galaxy, the fraction of old/young stars and the re-emission of the light in the IR wavelength. Such a level of details is not required in our forward model: for computing time efficiency, we choose in our forward model to use only a global attenuation of light by dust for each component with fixed values of these various parameters. The internal dust extinction of the galaxy is therefore applied as follows:

$$L^c = \frac{B}{T}\int \mathrm{SED_{bulge}} \exp\left(-\mathrm{coeff_{bulge}} \times \mathrm{SED_{extinct}}\right) + \left(1-\frac{B}{T}\right)\int \mathrm{SED_{disk}} \exp\left(-\mathrm{coeff_{disk}} \times \mathrm{SED_{extinct}}\right) \tag{8.2}$$

where L^c is the flux of the galaxy with dust extinction, $\frac{B}{T}$ is the bulge-to-total ratio, $\mathrm{SED_{bulge}}$ is the SED of the bulge component (see Fig. 3.1), $\mathrm{SED_{disk}}$ is the SED of the disk component (see Fig. 3.1) and $\mathrm{SED_{extinct}}$ is the SED of the dust extinction for the Milky Way (see Fig. 3.4). The coefficients $\mathrm{coeff_{bulge}}$ and $\mathrm{coeff_{disk}}$ are found by equating Eq. (8.1) (that is the blue line in Fig. 8.1) with Eq. (8.2) (that are the dotted horizontal lines in Fig. 8.1 which depend on the inclination i of the galaxy). We only calculate these extinction coefficients in the SDSS g reference band in which the galaxies are generated with our forward model, and we verify in Fig. 8.2 that, the differences between extincted and non-extincted magnitudes of the bulge and the disk in all eight bands of our forward model are comparable to those calculated in Eq. (8.1) by Popescu et al. (2011).

This comparison between our attenuation and the one of Popescu et al. (2011) is done in Fig. 8.2 where the difference is < 0.2 mag for $1-\cos i < 0.9$ (*i.e.* an inclination angle $< 77°$) in all bands. Moreover, the difference is < 0.4 mag for $1-\cos i > 0.9$ which can be explained by the limiting choice of fitting the high inclination angles with constant slopes in Popescu et al. (2011). Finally, the discrepancies between the attenuation curves for the disk are slightly better than those for the bulge because the degree of the polynomial fit of Popescu et al. (2011) is higher for the disk component. Of course, for the g'-band, the discrepancy is almost zero because this band is very close to the SDSS g' reference band on which the extinction coefficients are calculated. Overall, there is a very good agreement between our attenuation curves and those of Popescu et al. (2011) which confirms the reliability of our forward model for the internal dust extinction. We have also verified that, in Eq. (8.2), the $\frac{B}{T}$ term has indeed no significant impact on these results (see Fig. 8.3 where the comparison is done for $0 \leq \frac{B}{T} \leq 1$ with step 0.1).

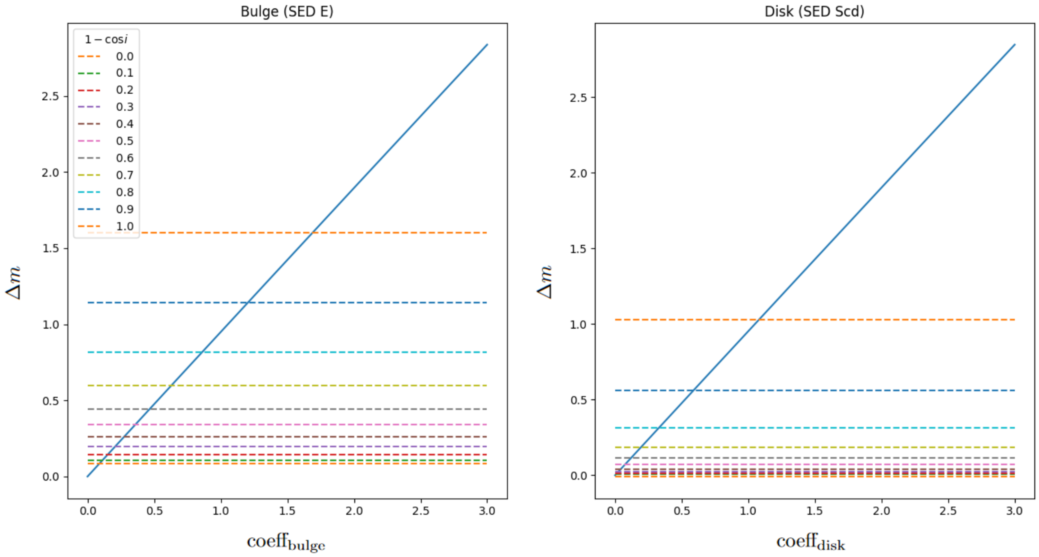


Fig. 8.1: Interpolating process on the model of Tuffs et al. (2004) to retrieve the extinction coefficients separately for the bulge and the disk as a function of the inclination angle i in the g' reference band (obtained for a total central opacity $\omega = 1$).

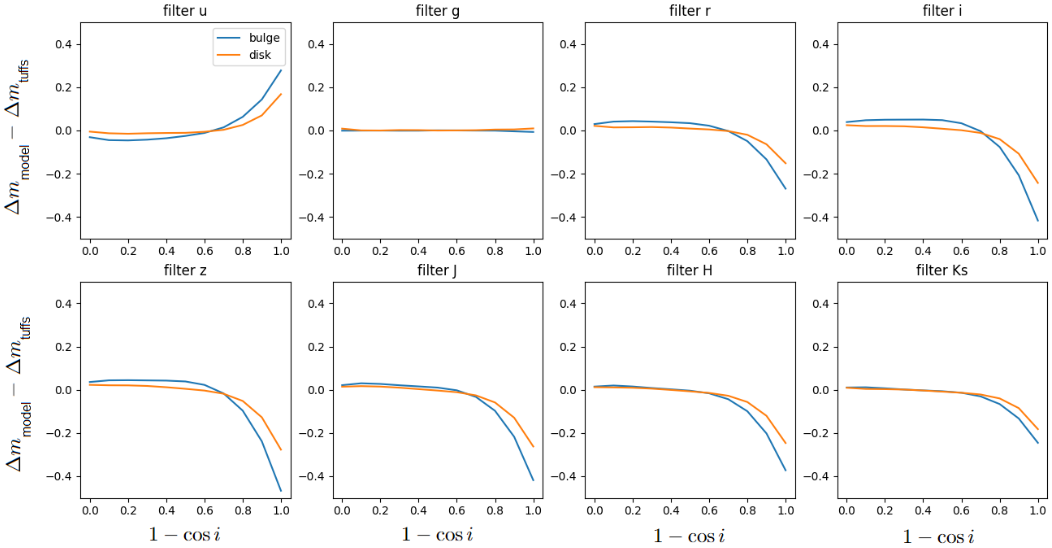


Fig. 8.2: Comparison between the attenuation curves generated in our forward model and those of Popescu et al. (2011) in the eight MegaPrime and WIRCam observed bands for the bulge and the disk (obtained for a total central opacity $\omega = 1$).

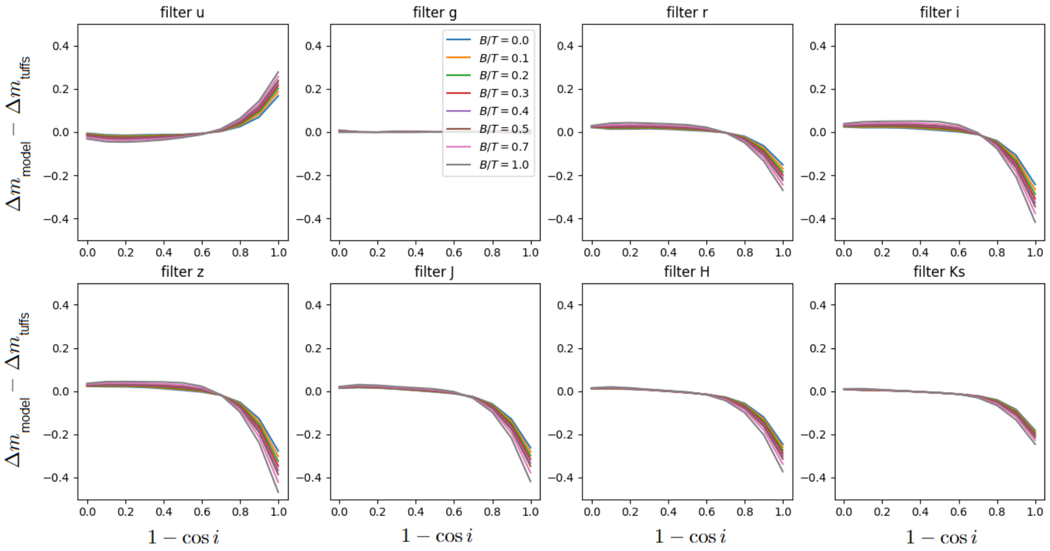


Fig. 8.3: Comparison between the attenuation curves generated in our forward model and those of Popescu et al. (2011) in the eight MegaPrime and WIRCam observed bands in Eq. (8.2) for varying values of $0 \leq \frac{B}{T} \leq 1$ with a 0.1 step (obtained for a total central opacity $\omega = 1$).

B. Backpropagation and the chain rule

Following the description of Chap. 4 Sect. 2.4, the gradients of the loss function, $\frac{\partial \Lambda}{\partial w}$, are needed with respect to the weights, $\forall w \in \mathbb{W}$. To do so, the chain rule can be applied to calculate the partial derivatives of the loss with respect to the outputs of the network, $\frac{\partial \Lambda}{\partial t}$, and to calculate the partial derivative of the outputs of the network with respect to the weights, $\frac{\partial t}{\partial w}$. Then, $\forall w \in \mathbb{W}$,

$$\frac{\partial \Lambda}{\partial w} = \sum_{k=1}^{n} \frac{\partial \Lambda}{\partial t_k} \frac{\partial t_k}{\partial w} + \sum_{\gamma=1}^{p} \sum_{k=1}^{m} \left(\frac{\partial \Lambda}{\partial \left(t_k^+\right)_\gamma} \frac{\partial \left(t_k^+\right)_\gamma}{\partial w} + \frac{\partial \Lambda}{\partial \left(t_k^-\right)_\gamma} \frac{\partial \left(t_k^-\right)_\gamma}{\partial w} \right) \tag{8.3}$$

It is worth noting that these sums are calculated once all the n fiducial images and the m pairs of deviation images have been passed through the network. To do so, we introduce in the following sections some definitions and we use algebra operations to express $\frac{\partial \Lambda}{\partial t_k}$ (resp. $\frac{\partial \Lambda}{\partial \left(t_k^+\right)_\gamma}$ or $\frac{\partial \Lambda}{\partial \left(t_k^-\right)_\gamma}$) which is a 4D-tensor (resp. a 5D-tensor).

B.1 Partial derivatives of the covariance matrix with respect to the summary outputs the network: $\frac{\partial C}{\partial t}$

Because the covariance matrix C is obtained only from the summary statistics associated with the n simulations generated with the fiducial parameter values, we know that $\frac{\partial C}{\partial t^+} = \frac{\partial C}{\partial t^-} = 0$. Let α and β be the indices of the output summary statistics vector t of the network. We can express the terms of the 2D covariance matrix of the summary statistics, $\forall \alpha, \beta \in \{1, \ldots, p\}$, as

$$C_{\alpha\beta} = \frac{1}{n-1} \sum_{k=1}^{n} \left(t_{k,\alpha} - \mu_\alpha\right) \left(t_{k,\beta} - \mu_\beta\right) \tag{8.4}$$

where $\forall \alpha \in \{1, \ldots, p\}$, $\mu_\alpha = \frac{1}{n} \sum\limits_{l=1}^{n} t_{l,\alpha}$ is the term α of the mean vector of the summary statistics. Then, $\frac{\partial C}{\partial t}$ is a 4D-tensor defined, $\forall \alpha, \beta, \gamma \in \{1, \ldots, p\}$, $\forall i \in \{1, \ldots, n\}$, as

$$\begin{aligned}
\frac{\partial C_{\alpha\beta}}{\partial t_{i,\gamma}} &= \frac{\partial}{\partial t_{i,\gamma}} \left[\frac{1}{n-1} \sum_{k=1}^{n} \left(t_{k,\alpha} - \mu_\alpha\right) \left(t_{k,\beta} - \mu_\beta\right) \right] \\
&= \frac{1}{n-1} \sum_{k=1}^{n} \left(\frac{\partial \left(t_{k,\alpha} - \mu_\alpha\right)}{\partial t_{i,\gamma}} \left(t_{k,\beta} - \mu_\beta\right) + \left(t_{k,\alpha} - \mu_\alpha\right) \frac{\partial \left(t_{k,\beta} - \mu_\beta\right)}{\partial t_{i,\gamma}} \right) \\
&= \frac{1}{n-1} \sum_{k=1}^{n} \left[\left(\frac{\partial t_{k,\alpha}}{\partial t_{i,\gamma}} - \frac{1}{n} \sum_{l=1}^{n} \frac{\partial t_{l,\alpha}}{\partial t_{i,\gamma}} \right) \left(t_{k,\beta} - \mu_\beta\right) + \left(t_{k,\alpha} - \mu_\alpha\right) \left(\frac{\partial t_{k,\beta}}{\partial t_{i,\gamma}} - \frac{1}{n} \sum_{l=1}^{n} \frac{\partial t_{l,\beta}}{\partial t_{i,\gamma}} \right) \right]
\end{aligned} \tag{8.5}$$

Because $\forall \alpha \in \{1, \ldots, p\}$ and $\forall k \in \{1, \ldots, n\}$, $\frac{\partial t_{k,\alpha}}{\partial t_{i,\gamma}} = 0$ if $k \neq i$ or $\alpha \neq \gamma$, we can simplify the last equation, $\forall \alpha, \beta, \gamma \in \{1, \ldots, p\}$, $\forall i \in \{1, \ldots, n\}$, as

$$\frac{\partial C_{\alpha\beta}}{\partial t_{i,\gamma}} = \begin{cases} 2\frac{\left(t_{i,\gamma} - \mu_\gamma\right)}{n-1} & \text{if} \quad \alpha = \beta = \gamma \\ \frac{t_{i,\beta} - \mu_\beta}{n-1} & \text{if} \quad \alpha = \gamma \neq \beta \\ \frac{t_{i,\alpha} - \mu_\alpha}{n-1} & \text{if} \quad \beta = \gamma \neq \alpha \\ 0 & \text{if} \quad \alpha \neq \gamma \quad \text{and} \quad \beta \neq \gamma \end{cases} \tag{8.6}$$

B.2 Partial derivatives of the derivatives of the mean vector with respect to the deviation outputs of the network: $\frac{\partial \nabla_\theta \mu}{\partial t^+}$, $\frac{\partial \nabla_\theta \mu}{\partial t^-}$

Because the mean vector is obtained only from the summary statistics associated with simulations generated with deviations around the fiducial parameter values, we know that $\frac{\partial \nabla_\theta \mu}{\partial t} = 0$. Now, given the numerical derivative of the mean vector of the summary statistics of Eq. (4.29), we have, $\forall \alpha, \beta, \gamma, \delta \in \{1, \ldots, p\}$,

$\forall i \in \{1, \ldots, m\}$,

$$\frac{\partial \left[\nabla_\theta \mu\right]_{\alpha\beta}}{\partial \left(t^+_{i,\delta}\right)_\gamma} = \frac{\partial}{\partial \left(t^+_{i,\delta}\right)_\gamma} \left(\frac{1}{m} \sum_{k=1}^{m} \frac{\left(t^+_{k,\beta}\right)_\alpha - \left(t^-_{k,\beta}\right)_\alpha}{2\Delta\theta_\alpha} \right) \tag{8.7}$$

Knowing that $\forall \alpha, \beta \in \{1, \ldots, p\}$ and $\forall i \in \{1, \ldots, m\}$, $\frac{\partial \left(t^+_{k,\beta}\right)_\alpha}{\partial \left(t^+_{i,\delta}\right)_\gamma} = 0$, if $\alpha \neq \gamma$ or $\beta \neq \delta$, and that $\frac{\partial \left(t^-_{k,\beta}\right)_\alpha}{\partial \left(t^+_{i,\delta}\right)_\gamma}$ is always 0, we can simplify the last equation, $\forall \alpha, \beta, \gamma, \delta \in \{1, \ldots, p\}$, $\forall i, \in \{1, \ldots, m\}$, into

$$\frac{\partial \left[\nabla_\theta \mu\right]_{\alpha\beta}}{\partial \left(t^+_{i,\delta}\right)_\gamma} = \begin{cases} \frac{1}{2m\Delta\theta_\gamma} & \text{if} \quad \alpha = \gamma \quad \text{and} \quad \beta = \gamma \\ 0 & \text{otherwise} \end{cases} \tag{8.8}$$

Similarly, we have, $\forall \alpha, \beta, \gamma, \delta \in \{1, \ldots, p\}$, $\forall i \in \{1, \ldots, m\}$,

$$\frac{\partial \left[\nabla_\theta \mu\right]_{\alpha\beta}}{\partial \left(t^-_{i,\delta}\right)_\gamma} = \begin{cases} -\frac{1}{2m\Delta\theta_\gamma} & \text{if} \quad \alpha = \gamma \quad \text{and} \quad \beta = \delta \\ 0 & \text{otherwise} \end{cases} \tag{8.9}$$

B.3 Partial derivatives of the Fisher information matrix with respect to the outputs of the network: $\frac{\partial F}{\partial t}$, $\frac{\partial F}{\partial t^+}$, $\frac{\partial F}{\partial t^-}$

Firstly, we can express each component of the Fisher information matrix using Eqs. (4.6) and (4.29), $\forall \alpha, \beta \in \{1, \ldots, p\}$,

$$F_{\alpha\beta} = \frac{1}{2} \left(\left[\nabla_\theta \mu\right]_{\cdot\alpha}^T C^{-1} \left[\nabla_\theta \mu\right]_{\cdot\beta} + \left[\nabla_\theta \mu\right]_{\cdot\beta}^T C^{-1} \left[\nabla_\theta \mu\right]_{\cdot\alpha} \right) \tag{8.10}$$

where, $\left[\nabla_\theta \mu\right]_{\cdot\alpha}$is a vector with p components given by $\left[\nabla_\theta \mu\right]_{\gamma\alpha}$, $\forall \gamma \in \{1, \ldots, p\}$. Then, using the matrix algebra formula $\partial M^{-1} = -M^{-1}(\partial M)M^{-1}$, we can express the partial derivative of each component of the Fisher information matrix with respect to the summary statistics at fiducial parameter values, $\forall \alpha, \beta, \gamma \in \{1, \ldots, p\}$, $\forall i \in \{1, \ldots, n\}$, as

$$\frac{\partial F_{\alpha\beta}}{\partial t_{i,\gamma}} = -\frac{1}{2} \left(\left[\nabla_\theta \mu\right]_{\cdot\alpha}^T C^{-1} \frac{\partial C}{\partial t_{i,\gamma}} C^{-1} \left[\nabla_\theta \mu\right]_{\cdot\beta} + \left[\nabla_\theta \mu\right]_{\cdot\beta}^T C^{-1} \frac{\partial C}{\partial t_{i,\gamma}} C^{-1} \left[\nabla_\theta \mu\right]_{\cdot\alpha} \right) \tag{8.11}$$

Similarly, $\forall \alpha, \beta, \gamma, \delta \in \{1, \ldots, p\}$, $\forall i \in \{1, \ldots, m\}$, we have

$$\begin{aligned} \frac{\partial F_{\alpha\beta}}{\partial \left(t^+_{i,\delta}\right)_\gamma} =& \frac{1}{2} \left(\frac{\partial \left[\nabla_\theta \mu\right]_{\cdot\alpha}}{\partial \left(t^+_{i,\delta}\right)_\gamma}^T C^{-1} \left[\nabla_\theta \mu\right]_{\cdot\beta} + \left[\nabla_\theta \mu\right]_{\cdot\alpha}^T C^{-1} \frac{\partial \left[\nabla_\theta \mu\right]_{\cdot\beta}}{\partial \left(t^+_{i,\delta}\right)_\gamma} \right. \\ & \left. + \frac{\partial \left[\nabla_\theta \mu\right]_{\cdot\beta}}{\partial \left(t^+_{i,\delta}\right)_\gamma}^T C^{-1} \left[\nabla_\theta \mu\right]_{\cdot\alpha} + \left[\nabla_\theta \mu\right]_{\cdot\beta}^T C^{-1} \frac{\partial \left[\nabla_\theta \mu\right]_{\cdot\alpha}}{\partial \left(t^+_{i,\delta}\right)_\gamma} \right) \end{aligned} \tag{8.12}$$

and

$$\begin{aligned} \frac{\partial F_{\alpha\beta}}{\partial \left(t^-_{i,\delta}\right)_\gamma} =& \frac{1}{2} \left(\frac{\partial \left[\nabla_\theta \mu\right]_{\cdot\alpha}}{\partial \left(t^-_{i,\delta}\right)_\gamma}^T C^{-1} \left[\nabla_\theta \mu\right]_{\cdot\beta} + \left[\nabla_\theta \mu\right]_{\cdot\alpha}^T C^{-1} \frac{\partial \left[\nabla_\theta \mu\right]_{\cdot\beta}}{\partial \left(t^-_{i,\delta}\right)_\gamma} \right. \\ & \left. + \frac{\partial \left[\nabla_\theta \mu\right]_{\cdot\beta}}{\partial \left(t^-_{i,\delta}\right)_\gamma}^T C^{-1} \left[\nabla_\theta \mu\right]_{\cdot\alpha} + \left[\nabla_\theta \mu\right]_{\cdot\beta}^T C^{-1} \frac{\partial \left[\nabla_\theta \mu\right]_{\cdot\alpha}}{\partial \left(t^-_{i,\delta}\right)_\gamma} \right) \end{aligned} \tag{8.13}$$

B.4 Partial derivatives of the loss with respect to the outputs of the network: $\frac{\partial\Lambda}{\partial t}$, $\frac{\partial\Lambda}{\partial t^+}$, $\frac{\partial\Lambda}{\partial t^-}$

Here, we use the definition of our two loss functions of Eqs. (4.26) and (4.27), and we use the two following matrix algebra formula:

- $\partial||M||_F^2 = 2\mathrm{Tr}[M\partial M]$
- $\partial\log(\det M) = \mathrm{Tr}\left[\left(M^{-1}\right)^T\partial M\right]$

Therefore, $\forall\gamma\in\{1,\ldots,p\}$, $\forall i\in\{1,\ldots,n\}$, we have

$$\frac{\partial\Lambda_F}{\partial t_{i,\gamma}} = -\mathrm{Tr}\left[F^{-1}\frac{\partial F}{\partial t_{i,\gamma}}\right] \tag{8.14}$$

$$\frac{\partial\Lambda_C}{\partial t_{i,\gamma}} = 0.5\left(\mathrm{Tr}\left[(C-I)\frac{\partial C}{\partial t_{i,\gamma}}\right] + \mathrm{Tr}\left[\left(C^{-1}-I\right)C^{-1}\frac{\partial C}{\partial t_{i,\gamma}}C^{-1}\right]\right) \tag{8.15}$$

Similarly, $\forall\gamma,\delta\in\{1,\ldots,p\}$, $\forall i\in\{1,\ldots,m\}$, we can write

$$\frac{\partial\Lambda_F}{\partial\left(t^+_{i,\delta}\right)_\gamma} = -\mathrm{Tr}\left[F^{-1}\frac{\partial F}{\partial\left(t^+_{i,\delta}\right)_\gamma}\right] \quad\text{and}\quad \frac{\partial\Lambda_F}{\partial\left(t^-_{i,\delta}\right)_\gamma} = -\mathrm{Tr}\left[F^{-1}\frac{\partial F}{\partial\left(t^-_{i,\delta}\right)_\gamma}\right] \tag{8.16}$$

$$\frac{\partial\Lambda_C}{\partial\left(t^+_{i,\delta}\right)_\gamma} = 0 \quad\text{and}\quad \frac{\partial\Lambda_C}{\partial\left(t^-_{i,\delta}\right)_\gamma} = 0 \tag{8.17}$$

because Λ_C does not depend on the summary statistics associated with simulations generated with deviations around the fiducial parameter values.

Here, we have therefore proven that the chain rule can be applied to calculate the partial derivatives of the loss with respect to the outputs of the network, $\frac{\partial\Lambda}{\partial t}$. This chain rule process can also be applied to the other layers of our inception network as an iterative procedure from the outputs of the network, therefore demonstrating that the backpropagation can be properly performed.

C. Choosing the number of simulations (n, m) to train the inception network

Training the inception network requires three collections of multiband deep field images generated by our forward model: the first collection gathers n multiband deep field images generated by our forward model with fiducial values, and the second and third collections gather m such images for each component of the parameter vector θ, but generated with deviations around the fiducial values (see Chap. 4 Sect. 2.6). We discussed in Chap. 4 Sect. 2.6 the fact that $m < n$ because the sample variance is removed by placing the galaxies at the same locations for each pair of simulations of the second and third collections. However, we did not discuss how to choose the values of m and n that are given in Table 9. This choice is not easy to make and depends on the architecture of the network (*i.e.* number of weights to fit), the number of parameters of the forward model and their possible correlations. Moreover, choosing large values for n and m (*i.e.* > 1000) is prohibited because it would require a very long time for the network to converge. In our applications of Chap. 5 Sect. 4. and Chap. 6 Sects. 1. and 2., this choice was a combination between "trial and error" of the full training combined with the following procedure for attempting to estimate the n parameter suggested by T. Charnock:

- Initialize an IMNN inception network with random weights.
- Pass $n = 2$ simulations inside the network and compute only the covariance matrix of the compressed summaries.

- Keep increasing the number n of simulations until the determinant of the covariance is relatively stable.

We have applied this technique for the application of Chap. 6 Sect. 2. Fig. 8.4 shows the evolution of the determinant of the covariance matrix when more and more simulations are considered. This procedure cannot be justified mathematically because the weights of the inception network are randomly initialized but empirically after many repetitions we reach a plateau of the similar level of $\log \det C$. The more simulations we use, the better the estimate of the covariance matrix, we must however find a compomise between a very large n that would require too much time for the training and a too small n that would no be sufficient to train on the luminosity function parameters. Note that Fig. 8.4 seems to indicate that only $n = 50$ simulations could be enough to estimate the covariance (which corresponds to the first plateau). We nevertheless considered that $n = 50$ was not sufficient in this 6D model parameters application and we finally decided to choose $n = 350$. This larger value was probably nevertheless not enough, and may be the cause for the difference between the determinant of the covariance matrix, which is higher for the training set ($\approx 10^{12}$) than for the validation set ($\approx 10^{10}$), see Fig. 6.12 in Chap. 6 Sect. 2. (this discrepancy could also be caused by a non-optimal training of the network); in that case, ideally, a larger value of n should be chosen and the full training should be restarted. Once the choice of n is made, we must choose m, which can also depend on the architecture of the network and the complexity of the forward model: ideally, m should be as large as possible, so that the network succeed in approximating the derivatives of the mean vector of the summary statistics, but in practice, we limit ourselves to $m \approx 25\% n$ given that we have to generate $2 \times n_{\text{param}} \times m$ simulations.

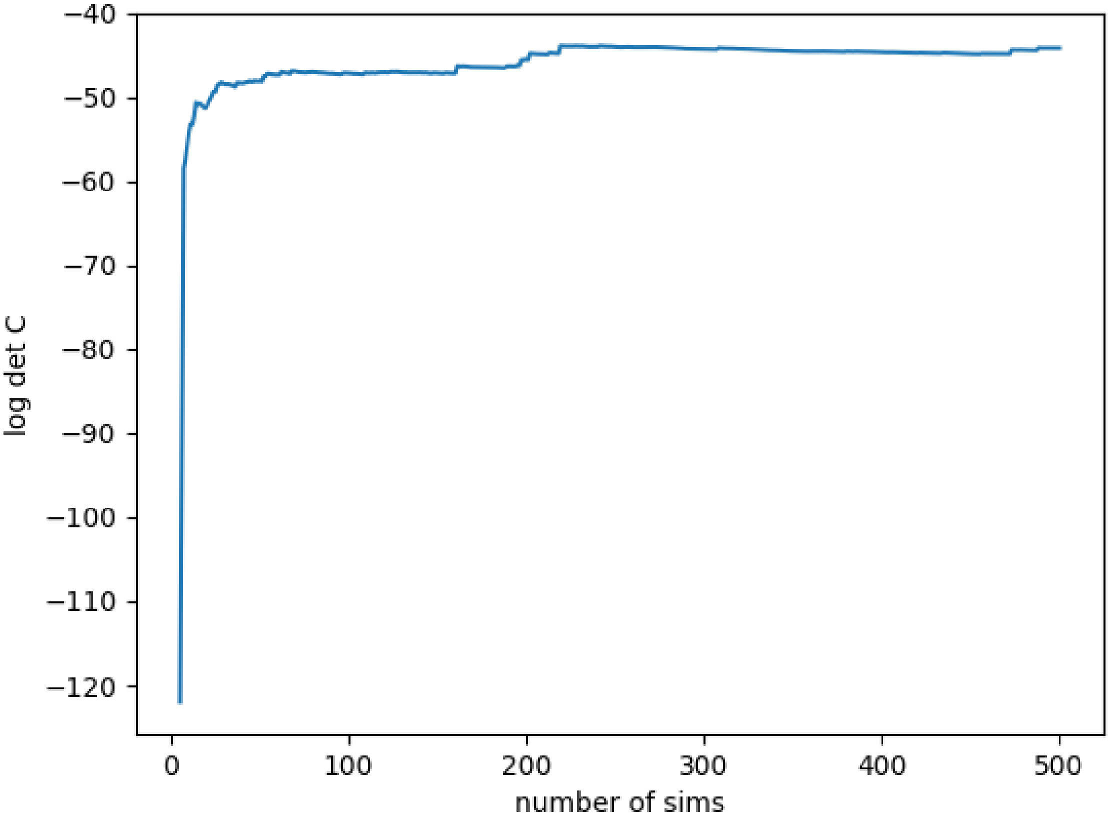


Fig. 8.4: Results of the procedure suggested by T. Charnock to estimate the number of simulations required to properly estimate the covariance matrix of the summary statistics. This plot is given for the third application of Chap. 6 Sect. 2., where $n = 350$ was finally chosen.