Universidad de Oviedo

MACHINE LEARNING EN ASTROFÍSICA Y COSMOLOGÍA

MACHINE LEARNING IN ASTROPHYSICS AND COSMOLOGY

TESIS DOCTORAL

PROGRAMA DE DOCTORADO EN CIENCIAS Y TECNOLOGÍAS DEL ESPACIO Y LA MATERIA

José Manuel Casas González

Universidad de Oviedo

MACHINE LEARNING EN ASTROFÍSICA Y COSMOLOGÍA

MACHINE LEARNING IN ASTROPHYSICS AND COSMOLOGY

TESIS DOCTORAL

PROGRAMA DE DOCTORADO EN CIENCIAS Y TECNOLOGÍAS DEL ESPACIO Y LA MATERIA

José Manuel Casas González

Directores de tesis

Dra. Dña. Laura Bonavera

Dr. D. Joaquín González-Nuevo

Universidad de Oviedo

RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
Español/Otro Idioma: Machine Learning en Astrofísica y Cosmología	Inglés: Machine Learning in Astrophysics and Cosmology

2.- Autor	
Nombre: José Manuel Casas González	
Programa de Doctorado: Programa en Ciencias y Tecnologías del Espacio y la Materia	
Órgano responsable: Centro Internacional de Postgrado	

RESUMEN (en español)

Aproximadamente 380.000 años después del Big Bang, los fotones se desacoplaron de los bariones y viajaron libres por el Universo. Hoy en día, aún pueden observarse en la banda de microondas. A este efecto se le llama el fondo cósmico de microondas y es una sonda clave para los cosmólogos y cosmólogas para conocer la naturaleza y evolución del Universo.

No obstante, en las frecuencias en las que puede observarse el fondo cósmico de microondas existen una serie de emisiones provenientes de nuestra Galaxia y de fuentes extragaláticas llamadas foregrounds que contaminan los mapas de temperatura y polarización. La caracterización de estas emisiones y, por tanto, la recuperación del fondo cósmico de microondas depende principalemente de la calidad de la metodología escogida.

Debido al aumento de la calidad de los intrumentos utilizados en astrofísica y cosmología, la cantidad de datos disponibles en futuros experimentos del fondo cósmico de microondas también aumentará, requiriendo métodos más sofisticados a la par que automáticos. Debido a la mejora de la capacidad computacional, los modelos de machine learning o aprendizaje automático, que tienen la habilidad de aprender de los datos una tarea particular, pero requieren una gran cantidad de memoria y datos, han aumentado su impacto en diversas areas de la vida cotidiana. Más aún, las redes neuronales, que son modelos de aprendizaje automático inspirados en la neurociencia, son perfectos para la recuperación del fondo cósmico de microondas y la caracterización de foregrounds, ya que están diseñados para lidiar con comportamientos no lineales en los datos, que son precisamente aquellos que caracterizan dichas emisiones.

Esta tesis de doctorado presenta nuevas metodologías basadas en redes neuronales artificiales para varios análisis del fondo cósmico de microondas. Más precisamente, mediante recorte de parches cuadrados del cielo de microondas visto por el satélite Planck, se han entrenado con simulaciones realistas varias redes neuronales convolucionales para detección de radio galaxias, para el ajuste de sus propiedades de polarización y para la recuperación del fondo cósmico de microondas en temperatura y polarización. Por último, se describirán los usos futuros y los desarrollos de estas redes neuronales.

Universidad de Oviedo

RESUMEN (en Inglés)

Nearly 380.000 years after the Big Bang, photons decoupled from baryons and freely traveled along the Universe. Today, they still can be observed in the microwave regime. This effect is called the cosmic microwave background, and it is a key probe for cosmologists to understand the nature and evolution of the Universe.

However, at the frequencies where the cosmic microwave background can be observed, there are several emissions from our Galaxy and extragalactic sources called foregrounds, which contaminate both temperature and polarization maps. The characterization of these emissions and therefore the recovery of the cosmic microwave background depend mainly on the quality of the chosen methodology.

Due to the increasing in the quality of the instruments used in Astrophysics and Cosmology, the quantity of available data in future cosmic microwave background experiments will also increase, requiring more sophisticated and automatic methods. Due to the increasing in the computational capability, machine learning models, which have the ability of learning from data a particular task but require high amounts of data and memory, have been increased their impact in many areas of human live. Furthermore, artificial neural networks, which are machine learning models inspired in neuroscience, are perfect for cosmic microwave background recovery and foreground characterization since they are designed to deal with non-linear behaviors from data, which are precisely the ones that characterize these emissions.

This PhD thesis presents new methodologies based on artificial neural networks for several cosmic microwave background analyses. More precisely, by cutting squared patches of the microwave sky as seen by the Planck satellite, several convolutional neural networks have been trained with realistic simulations for radio galaxies detection, for the constraining of their polarization properties and for the recovery of the cosmic microwave background in both temperature and polarization. Lastly, future uses and developments of these neural networks will be described.

SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO
EN Programa en Ciencias y Tecnologías del Espacio y la Materia

A mis padres

Abstract

Nearly 380.000 years after the Big Bang, photons decoupled from baryons and freely traveled along the Universe. Today, they still can be observed in the microwave regime. This effect is called the cosmic microwave background, and it is a key probe for cosmologists to understand the nature and evolution of the Universe.

However, at the frequencies where the cosmic microwave background can be observed, there are several emissions from our Galaxy and extragalactic sources called foregrounds, which contaminate both temperature and polarization maps. The characterization of these emissions and therefore the recovery of the cosmic microwave background depend mainly on the quality of the chosen methodology.

Due to the increasing in the quality of the instruments used in Astrophysics and Cosmology, the quantity of available data in future cosmic microwave background experiments will also increase, requiring more sophisticated and automatic methods. Due to the increasing in the computational capability, machine learning models, which have the ability of learning from data a particular task but require high amounts of data and memory, have been increased their impact in many areas of human live. Furthermore, artificial neural networks, which are machine learning models inspired in neuroscience, are perfect for cosmic microwave background recovery and foreground characterization since they are designed to deal with non-linear behaviors from data, which are precisely the ones that characterize these emissions.

This PhD thesis, based mainly in several high-impact papers led by the author (Casas et al. (2022a), Casas et al. (2022b), Casas et al. (2023a), Casas et al. (2023b)), presents new methodologies based on artificial neural networks for several cosmic microwave background analyses. More precisely, by cutting squared patches of the microwave sky as seen by the Planck satellite, several convolutional neural networks have been trained with realistic simulations for radio galaxies detection, for the constraining of their polarization properties and for the recovery of the cosmic microwave background in both temperature and polarization. Lastly, future uses and developments of these neural networks will be described.

Resumen

Aproximadamente 380.000 años después del Big Bang, los fotones se desacoplaron de los bariones y viajaron libres por el Universo. Hoy en día, aún pueden observarse en la banda de microondas. A este efecto se le llama el fondo cósmico de microondas y es una sonda clave para los cosmólogos y cosmólogas para conocer la naturaleza y evolución del Universo.

No obstante, en las frecuencias en las que puede observarse el fondo cósmico de microondas existen una serie de emisiones provenientes de nuestra Galaxia y de fuentes extragaláticas llamadas foregrounds que contaminan los mapas de temperatura y polarización. La caracterización de estas emisiones y, por tanto, la recuperación del fondo cósmico de microondas depende principalemente de la calidad de la metodología escogida.

Debido al aumento de la calidad de los intrumentos utilizados en astrofísica y cosmología, la cantidad de datos disponibles en futuros experimentos del fondo cósmico de microondas también aumentará, requiriendo métodos más sofisticados a la par que automáticos. Debido a la mejora de la capacidad computacional, los modelos de machine learning o aprendizaje automático, que tienen la habilidad de aprender de los datos una tarea particular, pero requieren una gran cantidad de memoria y datos, han aumentado su impacto en diversas areas de la vida cotidiana. Más aún, las redes neuronales, que son modelos de aprendizaje automático inspirados en la neurociencia, son perfectos para la recuperación del fondo cósmico de microondas y la caracterización de foregrounds, ya que están diseñados para lidiar con comportamientos no lineales en los datos, que son precisamente aquellos que caracterizan dichas emisiones.

Esta tesis de doctorado, basada principalmente en varios artículos de alto impacto liderados por el autor (Casas et al. (2022a), Casas et al. (2022b), Casas et al. (2023a), Casas et al. (2023b)), presenta nuevas metodologías basadas en redes neuronales artificiales para varios análisis del fondo cósmico de microondas. Más precisamente, mediante recorte de parches cuadrados del cielo de microondas visto por el satélite Planck, se han entrenado con simulaciones realistas varias redes neuronales convolucionales para detección de radio galaxias, para el ajuste de sus propiedades de polarización y para la recuperación del fondo cósmico de microondas en temperatura y polarización. Por último, se describirán los usos futuros y los desarrollos de estas redes neuronales.

Contents

CHAPTER 1

Introduction

1.1 Background and motivation

Nearly 13.800 billion years from today, all the known things in the Universe, that is, planets, satellites, stars and galaxies, were constrained in a state of extremely high energy. At that *instant*, there was not space nor time. After a spontaneous change of the initial conditions, time started and space become expanded, a phenomenon known as Big Bang (Weinberg (2008), Baumann (2018), Schneider (2015)). Physical laws, as we know today, at least started at the smallest time interval that can be measured, which is 10^{-44} seconds after the Big Bang, a time called Planck time. The initial change of phase produced several quantum fluctuations at 10^{-36} seconds after the Big Bang, during the exponentially expansion of the Universe known as Cosmic Inflation (Guth (1981), Linde (1982)). Those primordial fluctuations formed structures in the early Universe that evolve into stars, galaxies and groups of galaxies by following the gravitational instability paradigm. Once expanded, the mean temperature of the Universe started to decrease.

When the Universe reached $T \approx 10^{12}K$, radiation density was composed by all the protons and neutrons (both baryons) today forming the visible baryonic matter, as so as electrons, photons, positrons, neutrinos, and also dark matter particles. All those particles were in equilibrium. Then, about one second after the Big Bang, temperature reached $T \approx 10^{10}K$, while particles maintained the equilibrium due to the shorter mean time between reaction with respect to the expansion of the Universe. However, while temperature decreased bellow $10^{10}K$, neutrinos could not be longer in equilibrium, and then they decoupled from the other particles, a process called *freeze-out*. Although neutrinos decoupled, the Universe continued its expansion. Then, temperature continued to decrease. When it reached $\approx 5 \times 10^{9}K$, electron and positron pairs could no longer be produced efficiently, a process called pair annihilation. After that, nearly all electrons were converted into photons, remaining the Universe electrically neutral.

After one second and before 3 minutes after the Big Bang, at about $\sim 10^9 K$, protons and neutrons formed lights elements such as Hidrogen, Helium and Litium nuclei and also their isotopes by nuclear reactions such as the ones in the stars. That process is called Big Bang Nucleosynthesis (BBN). Then, about 3 minutes after the Big Bang, when temperature reached $\sim 3000K$, BBN was over. At those stages, the Universe mainly consisted in photons, protons, helium nuclei, other light elements and electrons. Following the optical depth for Thomson scattering parameter (Planck Collaboration, 2020d), photons could propagate freely from a redshift $z \sim 1000$, which is inside the range of the corresponding redshift for the Hubble expansion during recombination. This is called the *last scattering surface*. Those photons from recombination are predicted, by following the Wien law, to be propagated from the recombination epoch until today without further interaction with other particles. They should follow a Planck spectrum, with only changing their temperature after recombination. Moreover, they could be observed today, redshifted into the microwave regime of the electromagnetic spectrum.

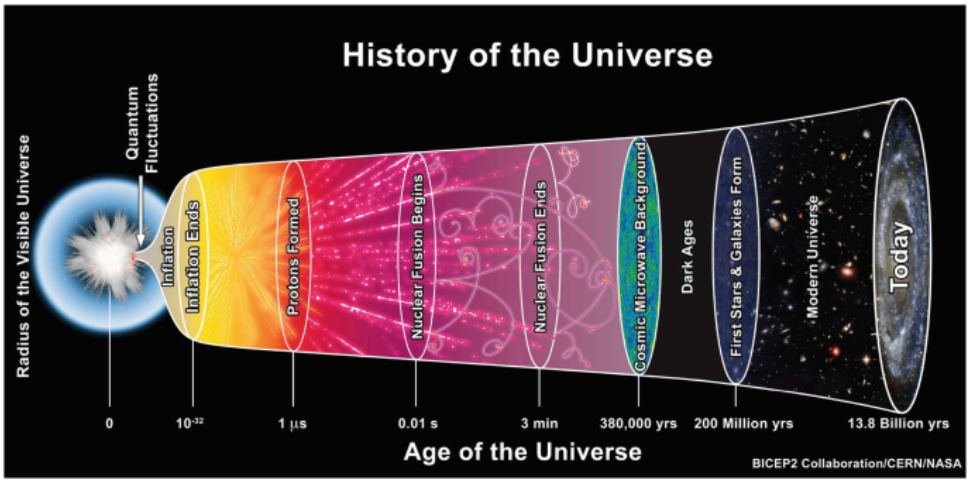

Figure 1.1: Scheme of the thermal history of the Universe (Mambrini, 2021).

These arguments firstly showed up in 1948, when George Gamow and collaborators predicted a thermal radiation of between 5 and 50 K due to the current abundance of nuclei. Years after, in 1964, Arno Penzias and Robert Wilson detected a fixed radiation of 3.5 K at 7.35 cm in the spectrum (Penzias and Wilson, 1965). This excess in the antenna was precisely the first detection of the photons from recombination epoch, a radiation called the cosmic microwave background (CMB). This observation, considered to be isotropically emitted in all directions of the sky, is correlated with the theoretical predictions by Lemaître of a Universe adiabatically expanding and cooling from its first stages. Then Penzias and Wilson won the Nobel Prize for

that astonishing discovery.

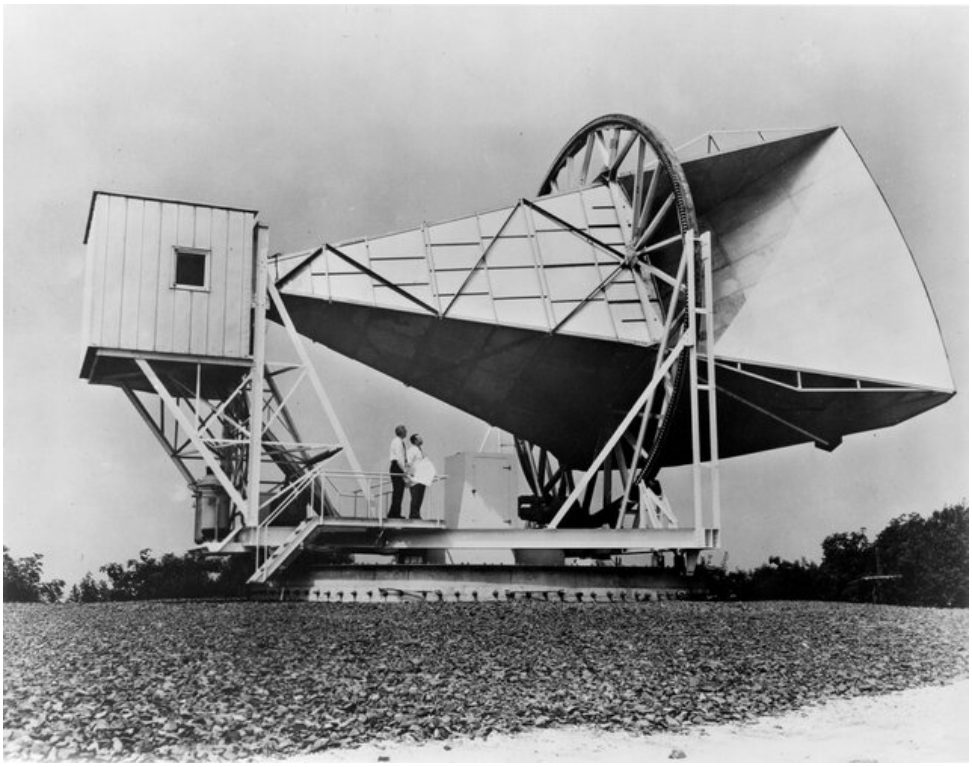

Figure 1.2: Horn antenna used in 1964 by Penzias and Wilson to discover the CMB (*NASA image*).

Immediately after, the main idea was that, if the structures of the Universe evolved from small initial fluctuations by gravitational instability, they would be present in the CMB, since the photons carrying out the radiation, they could not interact with such matter, as explained before. Therefore, the CMB, although isotropic as a first approximation, has to be an inhomogeneus radiation at $10^{-5}K$ levels. Cosmologists then started to study these CMB anisotropies.

The dipole, firstly observed in 1969 and 1971 by Conklin (Conklin, 1969) and Henry (Henry, 1971), respectively, was the first detected CMB anisotropy. It is interpreted as our proper motion with respect to the surface of *the last scattering*. After that, it was clear that more anisotropies would be present in the CMB, and their study could shade light about the structure and evolution of the Universe. For that purpose, NASA launched in 1989 the COsmic Background Explorer (COBE, Smoot et al. (1992)). It saw the sky 3 years before Smoot and collaborators published the first detection of the CMB anisotropies on angular scales $\theta \gtrsim 7^{\circ}$, corresponding to the first 20 harmonics [1].

[1] See Section 2.1.1 for understanding the comparison between angular scales and harmonics.

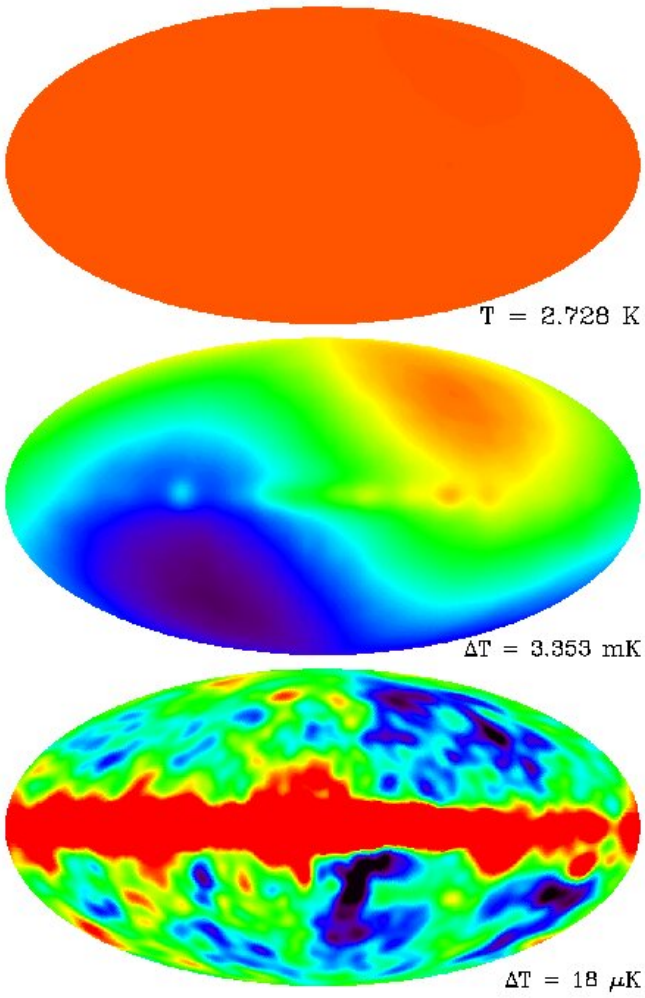

Figure 1.3: COBE CMB anisotropy results. Top figure: the isotropic CMB emission. Middle figure: the CMB dipole. Bottom figure: $l \sim 20$ multipole anisotropies, being the mean temperature fluctuation represented on each case (*NASA image*).

A Nobel Prize for George Smoot and John Mather, the principal investigators of DMR and FIRAS detectors aboard COBE, respectively, was the reward for COBE results.

Moreover, since the CMB is theoretically generated by an isotropic Gaussian stochastic process, all its information can be extracted from the power spectrum of its map, which could be expanded in spherical harmonics as described in Section 2.1.1, and it is usually plotted as $l(l+1)C_l/(2\pi)$ because of historical reasons and also since it would be constant for a scale invariant pattern on the sky.

Following this expression, the *acoustic* oscillations of the photon/baryon fluid before pho-

tons finally decoupled would left an imprint in the CMB power spectrum in form of peaks on angular scales $\theta \gtrsim 2^{\circ}$, corresponding to multipoles $l > 100$. For this purpose, several experiments were built, standing out BOOMERANG (de Bernardis et al., 2001), Maxima (Balbi et al., 2001) and DASI (Halverson et al., 2002), which observed the first *acoustic* peak at $l_1 \sim 200$. They also predicted more peaks at $l \sim nl_1$, being n even. However, their sensitivity was not enough for a reliable detection.

Furthermore, two main more aspects motivate the necessity of a precise measurement of the CMB anisotropies: firstly, as explained in Bucher (2015), several codes bases in Markov Chain Monte Carlo (MCMC) routines were developed in the 90ties with the aim of estimating cosmological parameters from CMB anisotropies [2]. Therefore, with a precise measure of the CMB power spectra, scientific community could determine the nature and evolution of the Universe by using the imprint of the photons when climbed out the gravitational potentials of the structures in the primordial Universe. Secondly, following Hu and White (1997) and Zaldarriaga and Seljak (1997), polarization is a key probe for Cosmic Inflation, described in Starobinsky (1982), Guth (1981) and Linde (1982), then a precise measurement of the polarization, in particular, the B-mode polarization at tensor-to-scalar levels of $r = EE/BB \leq 0.001$ would become an unprecedented discovery in science.

Therefore, a new precise CMB experiment should be prepared for these two purposes. This was the Wilkinson Microwave Anisotropy Probe (WMAP, Bennett et al. (2013)), which was launched in 2001 and took data for 9 years. It measured the CMB anisotropy with unprecedented precision, as can be seen by comparing its measure with COBE's one, as shown in Figure 1.4. With the new CMB power spectra results, cosmological parameters become fitted with high accuracy, cimenting the ΛCDM model as the one explaining our Universe, that is, thanks to WMAP we know that we live in a flat, expanding Universe represented by a cosmological constant, which started from a high-dense energy state called Big Bang, dominated by dark matter with respect to baryonic one. After WMAP, new polarization measures came out from BOOMERANG and the high-resolution ground-based experiments Atacama Cosmology Telescope (ACT) and South-Pole Telescope (SPT).

In May 2009, ESA launched the Planck Satellite (Planck Collaboration, 2020a). It was formed by the HFI and LFI instruments, being the first one dedicated to observed the sky at six frequencies: 100, 143, 217, 353, 545 and 857 GHz and the second one dedicated to observed the

[2]The reader is encouraged to read references in its concise review.

sky at 30, 44 and 70 GHz. The precision vary from channels but the maximum one achieved was around 5 arcminutes, allowing to constrain the CMB power spectrum to $l \sim 2500$. Furthermore, the much larger range of covering with respect previous experiments allowed *Planck* to unveil other aspects in Cosmology, such as a proper characterization (at least in temperature) of the microwave emissions from our Galaxy, the behavior of the different populations of extragalactic radio and dusty sources, the behavior of clusters of galaxies presenting Sunyaev-Zeldovich effects [3].

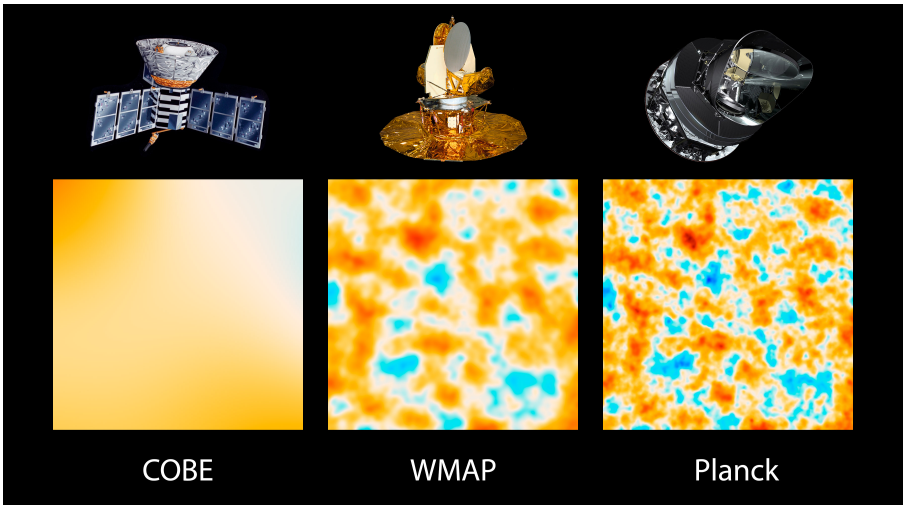

Figure 1.4: Three generations of satellites dedicated to the study of the CMB. From left to right: COBE, WMAP and *Planck* (*NASA image*).

In any case, two main goals from *Planck* were achieved: the release of an unprecedented detailed all sky CMB maps with its temperature anisotropies constrained to multipoles $l \sim 2500$, as the one shown in Figure 1.5, and the estimation of the cosmological parameters with higher accuracy than WMAP. However, the third main goal of constraining the polarization was not covered: the latests results from *Planck* allowed to estimate the E-mode with high accuracy up to multipoles $l \sim 1500$, as seen in the power spectrum of Figure 2.6, and its correlation with temperature. However, the B-mode polarization could not be constrained with *Planck* data.

Therefore, after *Planck*, it was clear that B-mode polarization and the evidence of the Cosmic Inflation period would be the main goal of the CMB scientific community. Actually, the B-mode is characterized at low and middle scales by present experiments such as BICEP2, POLARBEAR, ACTPol and SPT. Furthermore, the tensor-to-scalar ratio is constrained to

[3]The reader is encouraged to read *Planck* results https://www.cosmos.esa.int/web/planck/publications for more information.

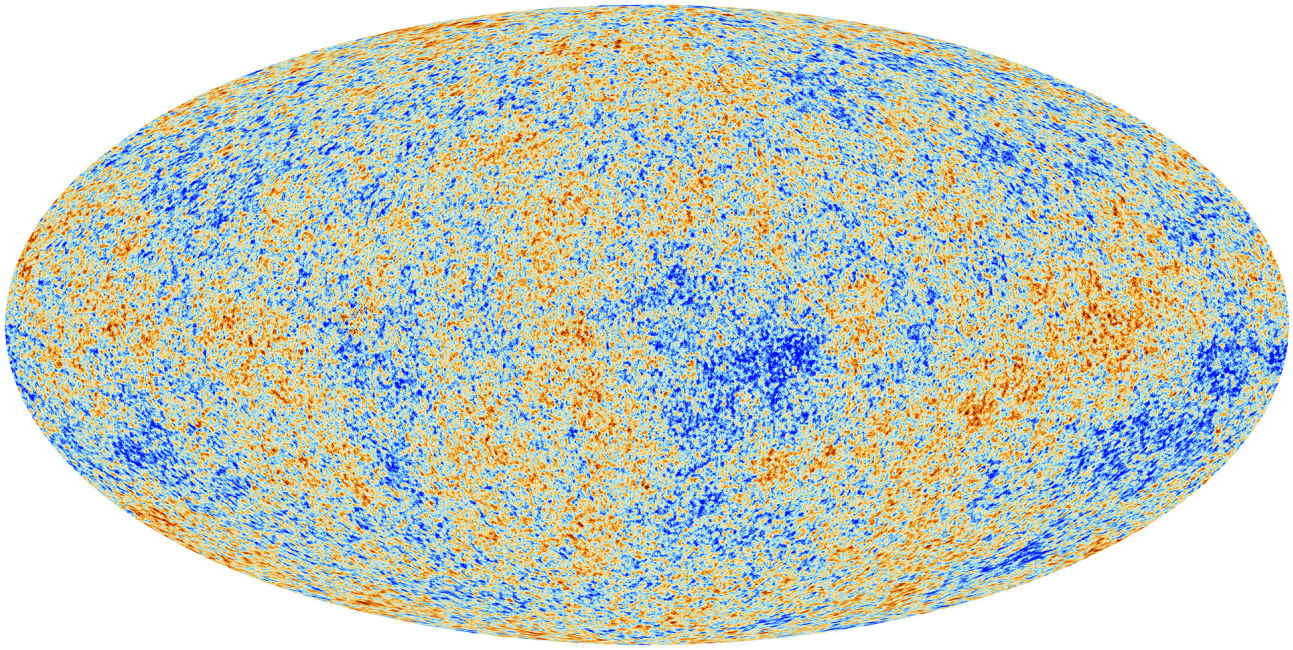

Figure 1.5: The cosmic microwave background as seen by *Planck* (*ESA image*).

$r < 0.032$ at 95% confidence level using *Planck* PR4 data and BICEP/Keck 2018 data Tristram et al. (2022). Upcoming experiments such as LiteBIRD and CMB-S4 are expected to improve this limit significantly, achieving sensitivities of $\sigma(r) = 0.001$ LiteBIRD Collaboration (2023a) and $\sigma(r) = 0.0005$ CMB-S4 Collaboration (2016) respectively.

While CMB community studied the anisotropies of the early Universe, mathematicians and engineers dreamed of machines doing tasks and thinking like humans. That field of research is called Machine Learning (ML), which is the ability of a machine to learn some kind of experience from data in order to aboard a particular task. ML models started to be popular in the last decades of the XX century, while increasing our computational power (see Hastie et al. (2001) for a review of the most popular ML approachs and their applications to real data). In the 40ties, it started the idea of mixing ML with the human brain. Today, this is called Deep Learning (DL), which is basically the ability of machines inspired on neuroscience to learn from data. In the past, it was called *cybernetics* in its early times between the 40ties and the 60ties, and *connectionism* in the 80ties-90ties. The actual name of DL mainly arose in 2006 until today (Goodfellow et al., 2016).

The first wave of neural network research started in the 40ties-60ties with the development of theories of biological learning and the implementation of the Perceptron model Sanger and Baljekar (1958) with the aim of training a single neuron.

The second wave started in the 80ties, with the movement called *connectionism* or par-

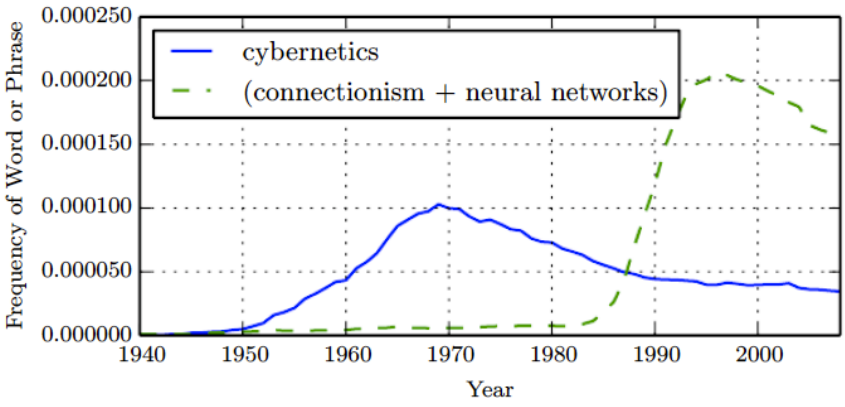

Figure 1.6: Deep learning research waves representation based on the frequency of its names (cybernetics, connectionism, neural networks) with respect the years (Goodfellow et al., 2016).

allel distributed processing. The main idea of this wave was that a large number of simple computational units can process information with intelligence when networked together, which is the basis of the hidden layers of actual neural networks. Another key ideas arose during this movement. The first one was the distributed representation, explained in Hinton et al. (1986). This is that many features should represent each input to a system, and each feature should represent many inputs. The second one was the use of back-propagation to train deep neural networks (Rumelhart et al., 1986). However, neural networks popularity decrease while artificial intelligence research cannot deal with the unreasonable expectations community had. Also the computational cost was too high for training and testing the neural networks.

In 2006, neural networks training took advantage from Hinton et al. (2006), Bengio et al. (2006) and other works, allowing to train deeper structures and improve generalization on test examples. After that, with the increasing of the datasets sizes due to the improvement in the computation and data storage, deeper structures were developed and tasks such image recognition, sequence learning, reinforcement learning, speech recognition and language processing can now be automatically solved with neural networks. The future is that, with the actual computational capability, neural networks will have the same number of neurons than a human in 2050.

Machine learning and neural networks have also been increasingly used in the last years in Astronomy, Astrophysics and Cosmology. For example, there are models based on machine learning for large-scale structure formation and identification, reionization, weak and strong gravitational lensing identification, redshift prediction or cosmological parameter estimation.

The reader is encouraged to follow several reviews with comprehensive lists of published machine learning applications to Astronomy (Baron, 2019), Astrophysics (Rodríguez et al., 2022) and Cosmology (Dvorkin et al., 2022). With respect to the CMB, principal machine and deep learning applications are lensing reconstruction, foreground model inference, inpainting and component separation [4].

1.2 Objectives and published contributions

The main goal of this PhD thesis is to develop several methodologies based on artificial neural networks for analyzing cosmic microwave background data. After preprocessing datasets used for training the models and formed by realistic simulations of the microwave sky as seen by the Planck satellite, different convolutional neural networks architectures are developed to investigate which one is better performing and can be adopted for different experiments to produce reliable results. In particular, two parallel (but related) goals can be defined:

- **Detection and characterization of point sources in *Planck* maps**
  Although reliable catalogues of radio and infrared galaxies polluting the CMB signal were published in *Planck* releases, future experiments with higher resolution will be able to detect a higher number of them. Neural networks seem to be suitable models for that purpose, following previous works such as Bonavera et al. (2021). Once detected, polarization properties about the AGNs behavior of the radio sources, which are not well known today, could also be constrained with neural networks by using a non-blind detection methodology.

- **Recovering the cosmic microwave background in temperature and polarization *Planck* maps**
  Although component separation (the process of disentangle one signal in a map from the others) models have recovered the CMB temperature and polarization maps with high accuracy during the last years, neural networks seem to be an alternative and accurate models for performing that task, by following previous works such as Petroff et al. (2020), since they can learn non-linear behaviors from data, which are precisely the ones of the foregrounds which contaminate the CMB signal.

[4]The most relevant works are listed in this comprehensive Github: https://github.com/georgestein/ml-in-cosmology.

This PhD thesis resume the main results from the following published contributions, all of them led by the author:

- **J. M. Casas**, J. Gonzalez-Nuevo, L. Bonavera, et al. (2022). Multi-frequency point source detection with fully-convolutional neural networks: Performance in realistic microwave sky simulations. *Astronomy and Astrophysics*, 658:A110, cited throughout the thesis as Casas et al. (2022b).

- **J. M. Casas**, L. Bonavera, J. Gonzalez-Nuevo, et al. (2022). CENN: A fully convolutional neural network for CMB recovery in microwave sky simulations. *Astronomy and Astrophysics*, 666:A89, cited throughout the thesis as Casas et al. (2022a).

- **J. M. Casas**, L. Bonavera, J. Gonzalez-Nuevo, et al. (2023). Constraining polarisation flux density and angle of point sources by training a convolutional neural network. *Astronomy and Astrophysics*, 670:A76, cited throughout the thesis as Casas et al. (2023a).

- **J. M. Casas**, L. Bonavera, J. Gonzalez-Nuevo, et al. Recovering the E and B-mode CMB polarization at sub-degree scales with neural networks, cited throughout the thesis as Casas et al. (2023b).

1.3 Outline of the thesis

This PhD thesis has been structured as follows. Chapter 2 describes the different components forming the microwave sky, all of them simulated in order to form the datasets used for training the neural networks. Chapter 3 explains the concepts of machine learning and deep learning, describing the elements forming the architectures of the neural networks used along this PhD thesis. Chapters 4 and 5 detail the results obtained when using those neural networks for detecting point sources in *Planck*-like realistic simulated patches of the sky and for constraining their polarization flux density and angle. Both kind of results are published in Casas et al. (2022b) and Casas et al. (2023a), respectively. Chapters 6 and 7 show the results obtained when using neural networks for recovering the CMB signal in both temperature and polarization *Planck*-like realistic simulated patches of the sky, which are published in Casas et al. (2022a) and Casas et al. (2023b), respectively. Chapter 8 concludes this PhD thesis by summarizing the main results and ideas, and also describing ongoing work and future prospects beyond this PhD thesis.

CHAPTER 2

The microwave sky

This chapter will cover the physics of the components of the microwave sky, which are simulated for training the neural networks developed in this PhD thesis. Figures 2.2 and 2.3 show the microwave sky at the nine frequencies covered by *Planck* in temperature and the seven ones observing the polarization maps, respectively. As shown, CMB anisotropies are visible at nearly all frequencies covered by *Planck*, especially up to 217 GHz, and always outside the Galactic plane (latitudes above/bellow $\pm 30^{\circ}$). The reader is encouraged to compare these all sky maps with Figure 2.1, where the intensity of Galactic microwave emissions is represented with respect the frequency.

Therefore, on the one hand, as seen in the bottom colorbars representing the intensity of the maps in μK_{CMB} units, bellow 100 GHz, Galactic emission dominates most of the map at 30 and 44 GHz in temperature and at 30, 44, 70 and 100 GHz in polarization. This is mainly due to synchrotron emission from ultra-relativistic electrons spiralling the magnetic field of our Galaxy. As shown, the highest emission is produced at 30 GHz. In temperature, also free-free and spinning dust (also known as anomalous microwave emission, AME) are contaminating the signal at these frequencies. As explained before, these were the channels observed by the LFI instrument aboard the *Planck* satellite.

On the other hand, a similar situation can be seen at higher frequencies, which are the ranges which were covered by the HFI instrument of *Planck*. In both temperature and polarization, starting at 217 GHz and increasing with frequency, dust from our Galaxy pollute the signal at this frequency and 353 GHz, especially in polarization. At small scales, thousands of radio sources detected by *Planck* contaminate the signal in temperature and, at least hundreds, in polarization. Also early galaxies in the process of forming their stellar masses emit a huge amount of dust, contaminating the signal at high frequencies, an emission called the cosmic infrared background.

As seen, the CMB is not the only emission of the microwave sky, and its recovery and characterization depend on the quality of the methodologies used for separating these Galactic and extragalactic contaminants called foregrounds, and also on the knowledge of their behavior

at these frequencies. The reader is encouraged to see the works by Leach et al. (2008) and Delabrouille et al. (2013), where pre-*Planck* methods for separating the signals from the CMB and the modelling of the different foreground emissions are extensively explained, respectively. Furthermore, Planck Collaboration (2020b) shows the final four methods used in the Collaboration and their resulted CMB and foreground maps. More actual, oriented for polarization, methods are described in LiteBIRD Collaboration (2023b). Following subsections will describe briefly each of these components in the microwave sky, emphasizing the ones in the simulations of this PhD thesis.

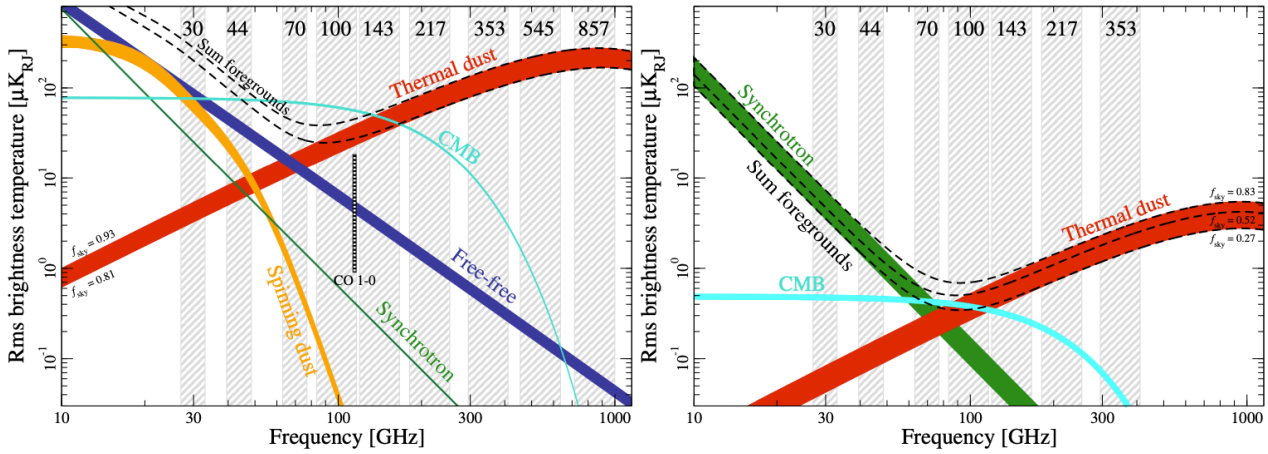

Figure 2.1: Frequency dependence of the main Galactic emissions in the microwave sky in temperature (left panel) and polarization (right panel). Grey vertical lines show each one of the Planck channels. Black dashed lines show the total sum of these emissions. The bands for each emission represent the percentage of sky coverage by *Planck* (Planck Collaboration, 2020a).

2.1 The cosmic microwave background

The CMB is the principal signal along this PhD thesis. In order to understand why we need a methodology based on neural networks for its recovery, we firsly should take into account its nature. This section will describe the main characteristics of the signal that change the cosmology since its discovery by Penzias and Wilson.

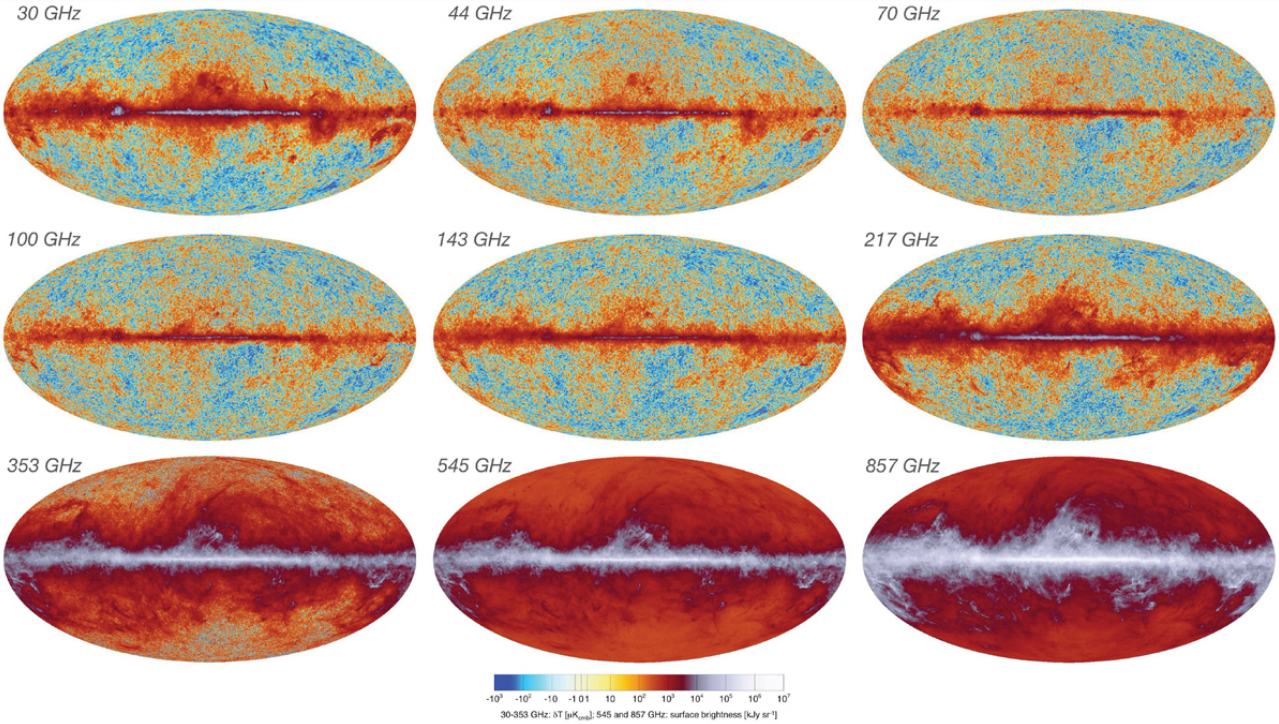

Figure 2.2: Temperature fluctuations of the microwave sky as seen by *Planck*. The colorbar in the bottom is the scale in μK_{CMB} units, representing the intensity signal for $30 - 353$ GHz channels and the surface brightness in k Jy sr^{-1} for 545 and 857 GHz channels (Planck Collaboration, 2020a).

2.1.1 The cosmic microwave background anisotropy

For a given model, pretended to be compared with observations, the most useful quantity to use is the power spectrum, which is the harmonic transform of the 2-point correlation function (Durrer, 2008), which contains all the statistical information of the model. In Fourier space it is defined as

$$\langle X(\mathbf{k}, t_0) X^*(\mathbf{k}', t_0) \rangle = (2\pi)^3 \delta(\mathbf{k} - \mathbf{k}') P_X(k), \tag{2.1}$$

where, in flat space, the function $X(\mathbf{k}, t_0)$ is the ordinary Fourier transform of the function in position space $X(\mathbf{x}, t_0)$, $\delta(\mathbf{k}-\mathbf{k}')$ is the Dirac δ-function and $\langle\rangle$ indicates the statistical average.

The CMB anisotropies can also be statistically studied with the power spectrum. Given $\Delta T/T$, a function of position $\mathbf{x} = \mathbf{x_0}$, time $\mathbf{t} = \mathbf{t_0}$ and photon direction on the sphere $\mathbf{n}$, the CMB anisotropies can be described in terms of spherical harmonics, Y_{lm}'s. Assuming that the distribution of $\Delta T/T(\mathbf{n})$ is the same for all directions, the CMB anisotropies become described

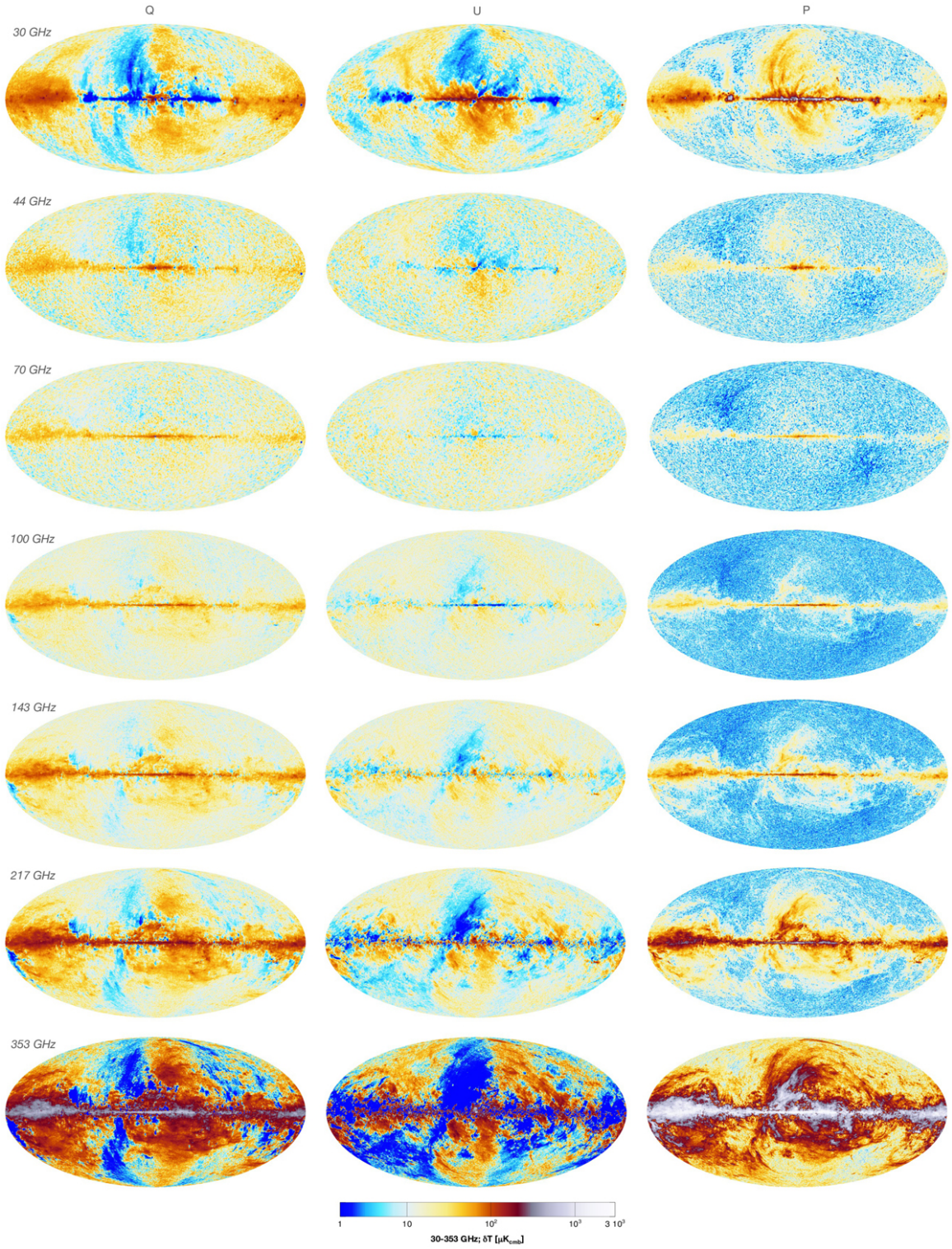

Figure 2.3: Polarized microwave sky as seen by *Planck*. The first two columns show the Q and U Stokes parameters while the last column represents the polarized intensity, $P = \sqrt{Q^2 + U^2}$. The colorbar in the bottom is the scale in μK_{CMB} units, representing the intensity signal for $30 - 353$ GHz channels (Planck Collaboration, 2020a).

as

$$\frac{\Delta T}{T}(\mathbf{x_0}, \mathbf{t_0}, \mathbf{n}) = \sum_{l,m} a_{lm}(\mathbf{x_0}) Y_{l,m}, \qquad \langle a_{lm} \cdot {a_{l'm'}}^* \rangle = \delta_{ll'}\delta_{mm'}C_l, \tag{2.2}$$

where a_{lm} are the expansion coefficients and the C_ls are the CMB power spectrum. Then, the CMB power spectrum can be divided into three regions, as shown in Figure 2.4:

- **The Sachs-Wolfe Term**

  On $l \lesssim 30$, the CMB photons become blueshifted as they fall into Newtonian potential wells and redshifted when they climb out them. If the depth of the potential well does not vary with time, both effects cancel, implying that there will not be an integrated Sachs-Wolfe (ISW) contribution. But if the depth of the potential well changes with time, in particular if the overall scale of the potential is decaying, the two effects no longer cancel and an integrated Sachs-Wolfe contribution is imprinted into the power spectrum. Actually, it is difficult to determine the nature of the Sachs-Wolfe Term because of the cosmic variance effect from the noisy primary CMB anisotropies emanating from the *last scattering surface*. *Planck* established a ISW effect consistent with a ΛCDM model, and its 3σ detection allowed to constrain the Ω_Λ cosmological parameter.

- **The Acoustic Peaks**

  On $l \gtrsim 100$, the contribution to $\Delta T/T(\mathbf{x_0}, \mathbf{t_0}, \mathbf{n})$ is related to *acoustic* oscillations due to the photon/baryon fluid experimenting graviational variations during recombination, a process commonly known as "Acoustic Peaks". Expanding the temperature anisotropies into spherical harmonics, the angular scale θ_n relatively corresponds to the harmonic number

  $$l_n \simeq \pi/\theta_n = \pi d_A(t_{dec})/\lambda_n = d_A(t_{dec})k_n = n\sqrt{3}\pi d_A(t_{dec})/t_{dec}, \tag{2.3}$$

  where $d_A(t_{dec})$ is the angular distance to the last scattering surface. For a flat, matter-dominated Universe, $d_A(t_{dec}) \simeq t_0$, then $l_n \simeq 180n$ [1]. Futhermore, these Acoustic Peaks depend on the sound speed of the radiation-baryon plasma, leading to the nth peak

  $$l_n \simeq k_n t_0 \cong n\pi\sqrt{3}\frac{t_0}{t_{dec}}. \tag{2.4}$$

  Then, for the first peak, $l_1 \sim 220$. Subsequent peaks are then given by $l_n = nl_1$.

[1] For a more concise explanation, the reader is encouraged to consult Durrer (2008) and Schneider (2015) books.

- **The Silk Damping**

  On $l \gtrsim 1000$, the CMB power spectrum become damped when the coupling between photons and the baryon/electron gas is still present but no longer perfect, being imprinted smaller acoustic peaks into the power spectrum.

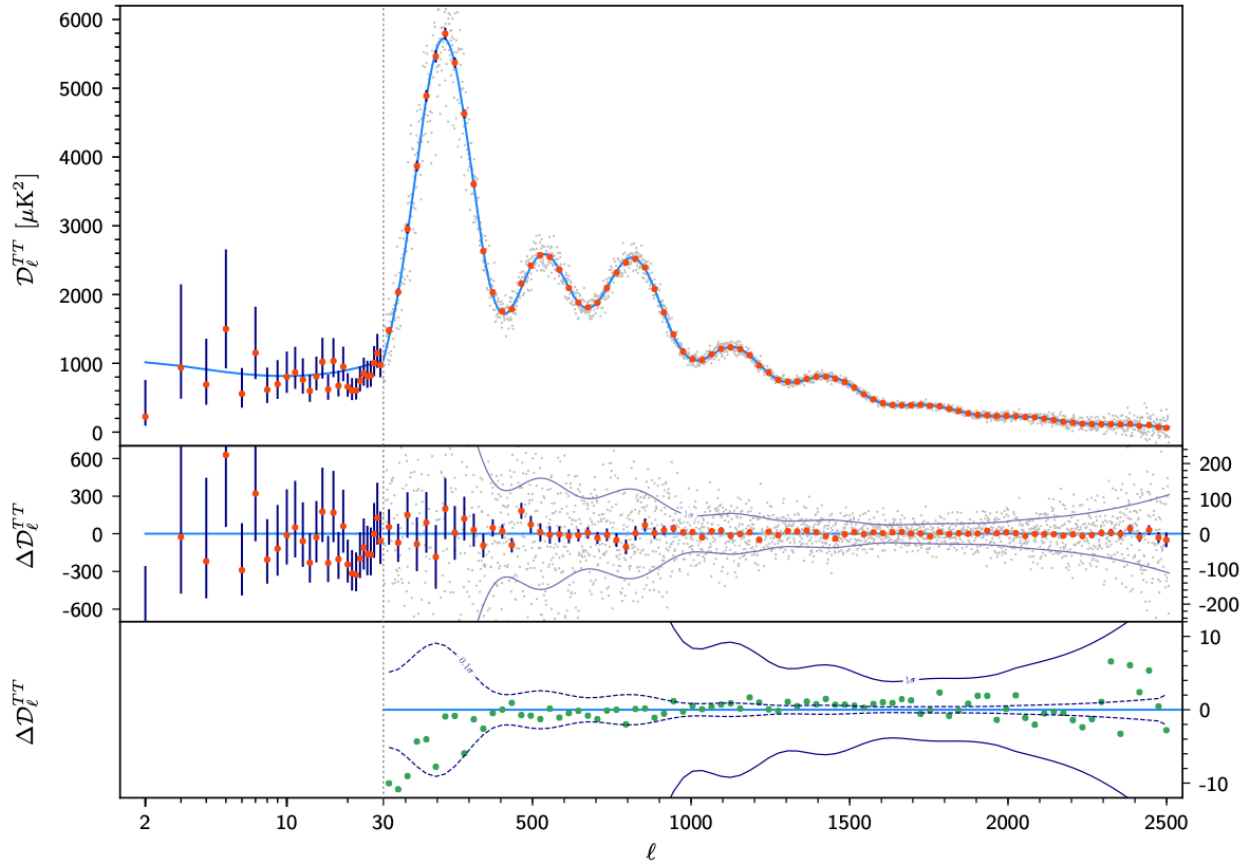

Figure 2.4: *Planck* 2018 temperature power spectrum in μK^2. Red dots show the binned results while grey ones represent the unbinned results. Blue line shows the ΛCDM theoretical spectrum. The $\pm 1\sigma$ uncertainty is represented as vertical lines for each bin. Residuals with respect to the ΛCDM model are shown in the middle panel. Unbinned data point uncertainties are represented in the $\pm 1\sigma$ blue region. The difference between 2015 and 2018 *Planck* results is represented as green dots in the bottom panel (Planck Collaboration, 2020c).

2.1.2 Polarization in the cosmic microwave background

In the early 80's, Starobinsky proposed the first semi-realistic model of inflation (Starobinsky, 1982), based on conformal anomaly in quantum gravity. However, it did not solve the homogeneity and isotropy problems, although it was the first model predicting gravitational waves

with a flat spectrum. A more physically clear model, now called "old inflation", was proposed by Alan Guth in 1981 (Guth, 1981), considering inflation as an exponential expansion of the Universe in a supercooled false vacuum state, that is, the Universe is considered in his theory as a metastable space without any fields or particles although with a large energy density. Then, when expanded, the energy density did not change so it should experiment an exponential expansion, making it very big and flat. While the false vacuum decays, the bubbles of the new phase collide and the Universe becomes hot. However, as concluded by Linde in 1982 (Linde, 1982), that explanation is misleading because if the new phase bubbles were formed near each other, Guth's inflation was too short to solve horizon, flatness and isotropy problems. If they were formed far away from each other, each of them represents a separate open Universe. Furthermore, Linde described the "new inflationary theory", under the assumption that inflation is a state either in the false vacuum or in an unstable state at the top of the effective potential. Then, the inflation field slowly rolled down to the minimum of its effective potential, producing density perturbations which are responsible for the homogeneity of our Universe. These density perturbation are classified into two categories: scalar and tensor ones. In particular, tensor perturbations produced gravitational waves in the primordial Universe. 380.000 years after that moment, they become imprinted into the polarization of the CMB.

Following Zaldarriaga and Seljak (1997), the all-sky CMB polarization could be described as an expansion in spin-weighted harmonics, an approach more realistic than small scale limit ones (see references therein). More particularly, the CMB radiation field is a 2×2 intensity tensor I_{ij}. The Stokes parameters Q and U, that characterized the linear polarization of the photons, are defined as $Q = (I_{11} - I_{22})/4$ and $U = I_{12}/2$. The temperature anisotropy is described as $T = (I_{11} + I_{22})/4$ and the Stokes parameter V, defining circular polarization, is neglected since in Cosmology it cannot be generated through Thomson scattering. Temperature is then invariant under rotation in the plane perpendicular to photon direction $\hat{\mathbf{n}}$, while Q and U transform under rotation by an angle ψ as

$$\begin{aligned} Q' &= Q cos(2\psi) + U sin(2\psi) \\ U' &= -Q sin(2\psi) + U cos(2\psi), \end{aligned} \tag{2.5}$$

where $\hat{\mathbf{e_1}}' = cos(\psi \mathbf{e_1}) + sin(\psi \mathbf{e_2})$ and $\hat{\mathbf{e_2}}' = -sin(\psi \mathbf{e_1}) + cos(\psi \mathbf{e_2})$. Then

$$(Q \pm U)'(\hat{\mathbf{n}}) = e^{\mp 2i\psi}(Q \pm iU)\hat{\mathbf{n}}, \tag{2.6}$$

which can be expanded in terms of a spin-weighted basis

$$\begin{aligned} T(\hat{\mathbf{n}}) &= \sum_{lm} a_{T,lm}\, Y_{lm}(\hat{\mathbf{n}}) \\ (Q+iU)(\hat{\mathbf{n}}) &= \sum_{lm} a_{2,lm}\, {}_2Y_{lm}(\hat{\mathbf{n}}) \\ (Q-iU)(\hat{\mathbf{n}}) &= \sum_{lm} a_{-2,lm}\, {}_{-2}Y_{lm}(\hat{\mathbf{n}}). \end{aligned} \tag{2.7}$$

While temperature power spectrum, explained above, is easily computed since its respective Stokes parameter is invariant under rotation in the plane perpendicular to $\hat{\mathbf{n}}$, both Q and U not. They have to be rotated for each wavevector $\mathbf{k}$ and direction of the sky ($\hat{\mathbf{n}}$). Then, the estimation of their power spectrum is only valid for small-scale limit. For all sky maps, there is mandatory to define spin zero rotationally invariant quantities. Using sping raising and lowering operators $\eth$ into $Q \pm iU$, the expansion coefficients $a_{T,lm}, a_{2,lm}, a_{-2,lm}$ [2] and their linear combinations, these quantities are

$$\begin{aligned} a_{E,lm} &= -(a_{2,lm} + a_{-2,lm})/2 \\ a_{B,lm} &= i(a_{2,lm} - a_{-2,lm})/2, \end{aligned} \tag{2.8}$$

then, one can define the rotationally invariant quantities

$$\begin{aligned} C_{Tl} &= \frac{1}{2l+1}\sum_m \langle a^*_{T,lm} a_{T,lm}\rangle \\ C_{El} &= \frac{1}{2l+1}\sum_m \langle a^*_{E,lm} a_{E,lm}\rangle \\ C_{Bl} &= \frac{1}{2l+1}\sum_m \langle a^*_{B,lm} a_{B,lm}\rangle \\ C_{Cl} &= \frac{1}{2l+1}\sum_m \langle a^*_{T,lm} a_{E,lm}\rangle, \end{aligned} \tag{2.9}$$

in order to estimate the final temperature and polarization power spectra. They imply that, when E remains unchanged, B changes the sign. Then, TT, TE, EE and BB power spectra characterize both temperature and polarization CMB signals. *Planck* TT power spectra is represented in Figure 2.4, while EE is shown in Figure 2.5. Their correlation, since does not have relation with this PhD thesis is not shown, but the reader is encouraged to see Planck Collaboration (2020c) for the latests results. BB power spectra was not recovered by *Planck*, but Figure 2.6 shows its actual constrain by modern experiments. Moreover, 2.7 shows the comparison between the estimates from past and present experiments with the predictions by LiteBIRD.

[2] The reader is encouraged to follow the complete derivation in Zaldarriaga and Seljak (1997).

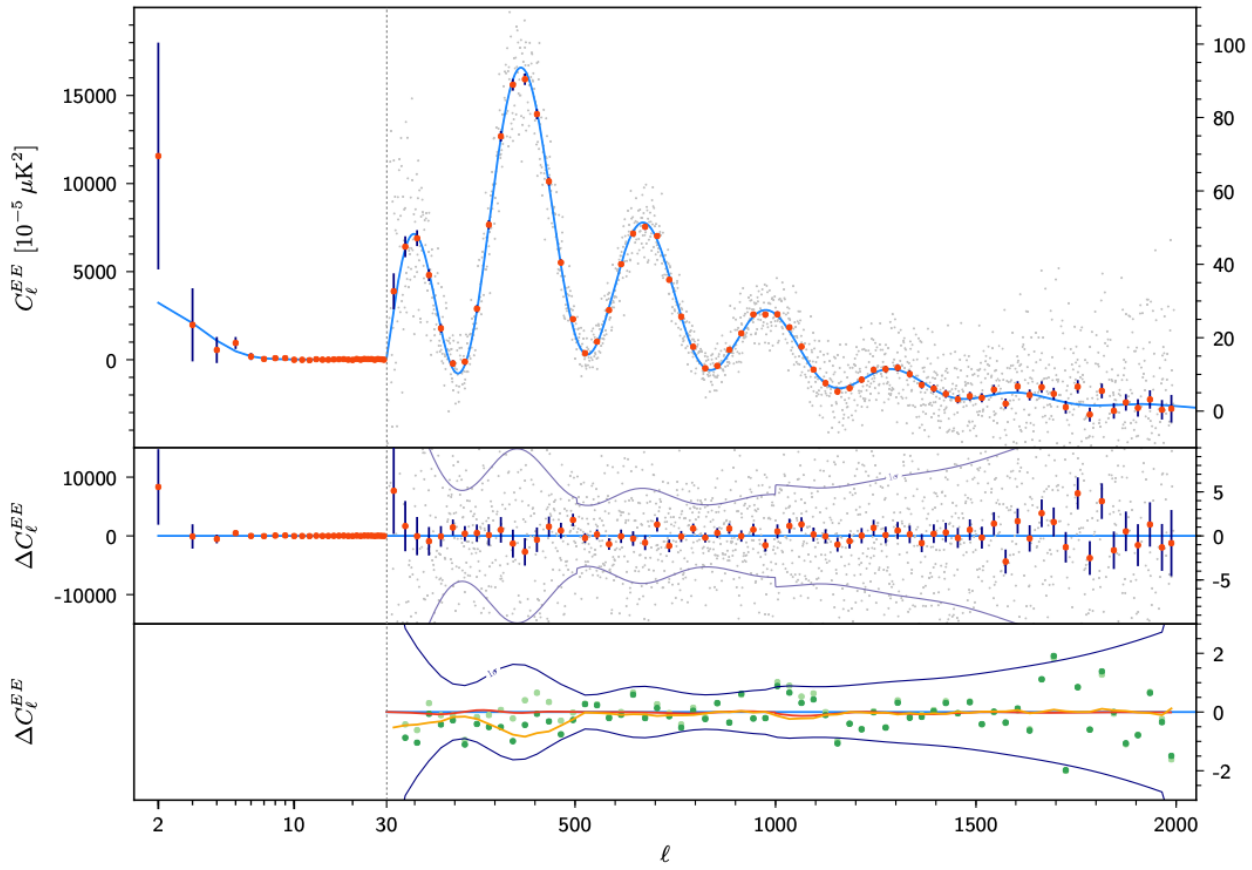

Figure 2.5: Planck 2018 EE polarization power spectrum in $10^{-5}\mu K^2$. Red dots show the binned results while grey ones represent the unbinned results. Blue line show the ΛCDM theoretical spectrum. The $\pm 1\sigma$ uncertainty is represented as vertical lines for each bin. Residuals with respect to the ΛCDM model are shown in the middle panel. Unbinned data point uncertainties are represented in the $\pm 1\sigma$ blue region. The difference between 2015 and 2018 Planck results is represented as green dots in the bottom panel (Planck Collaboration, 2020c).

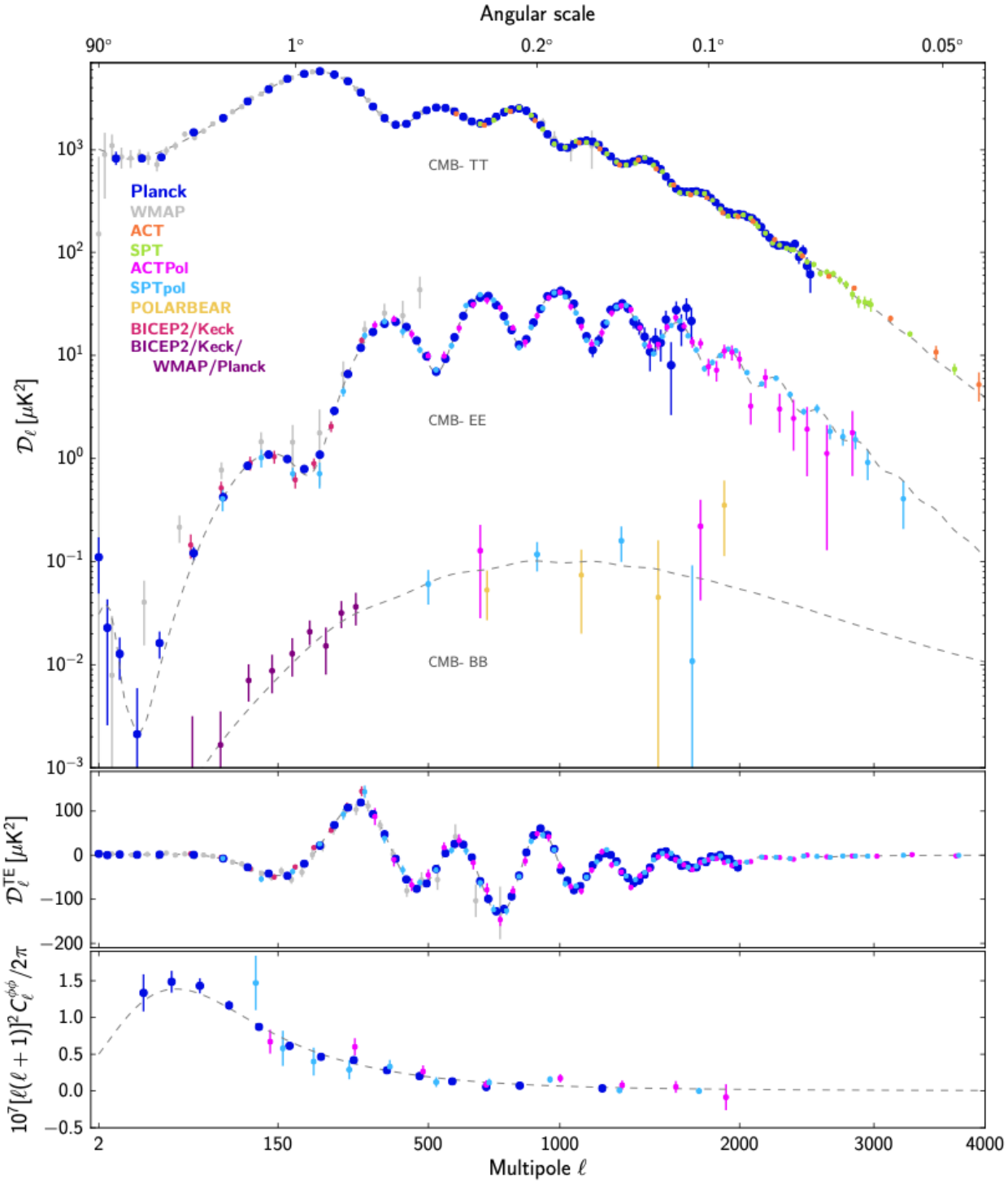

Figure 2.6: CMB angular power spectrum measurements for recent experiments. Upper panel: temperature and E-mode and B-mode polarization signals. Middle panel: cross-correlation spectrum between TT and EE. Lower panel: lensing deflection power spectrum. The dashed line shows the best-fit ΛCDM model to the *Planck* temperature, polarization, and lensing data (Planck Collaboration, 2020a).

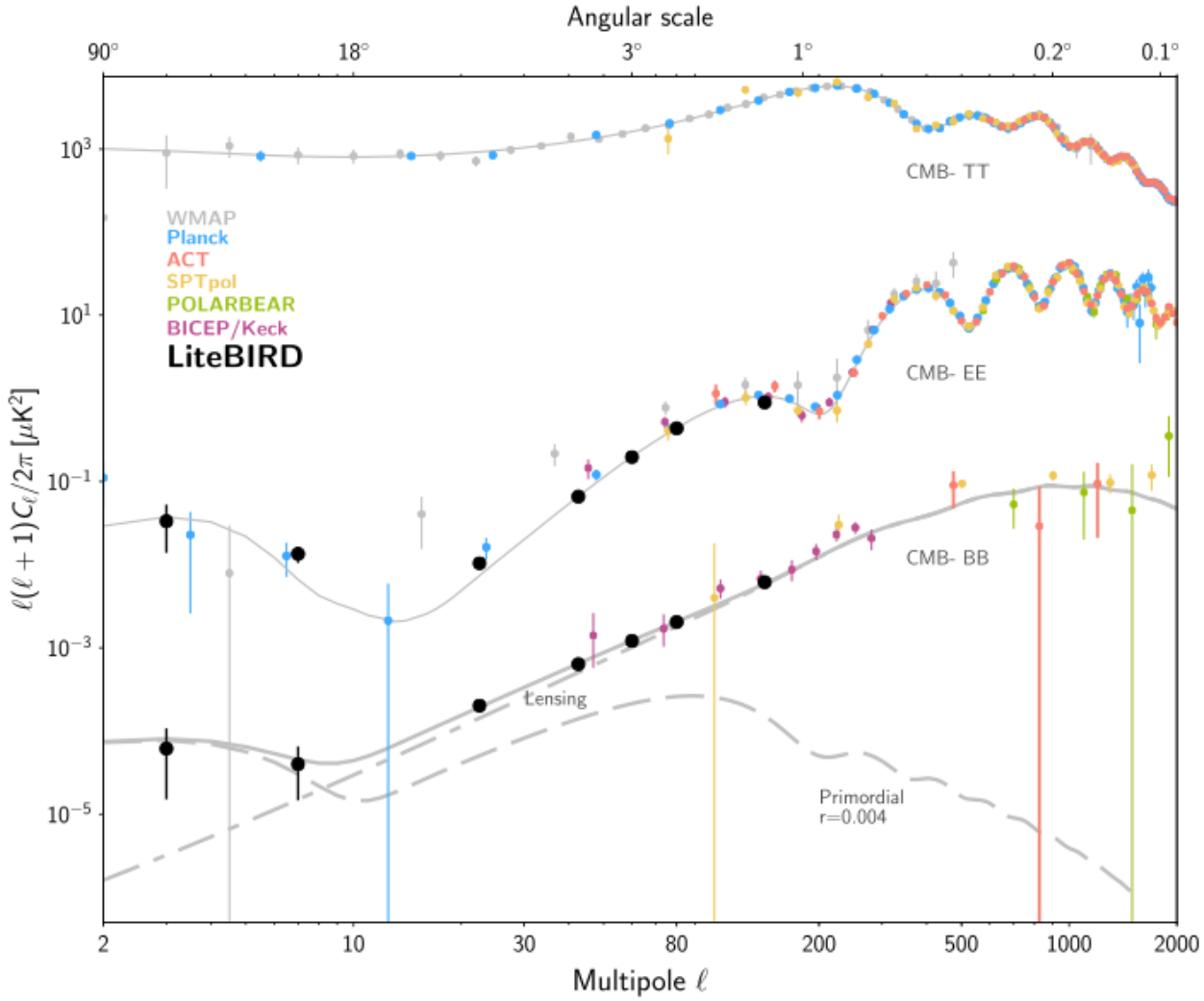

Figure 2.7: CMB angular power spectrum predictions for LiteBIRD in comparison with recent and past experiments. Upper panel: temperature and E-mode and B-mode polarization signals. Middle panel: cross-correlation spectrum between TT and EE. Lower panel: lensing deflection power spectrum. The grey lines show the theoretical predictions by the ΛCDM model, being the primordial B-modes at $r = 0.004$ represented as a dashed grey line (LiteBIRD Collaboration, 2023a).

2.2 Galactic and extragalactic emissions

As shown in Figure 2.1, several components from our Galaxy and another galaxies emit in the microwave sky. In fact, at determined frequencies, they dominate the total emission. Galactic components are mainly thermal dust from interstellar grains and synchrotron emission from ultra-relativistic electrons spiralling the Galactic magnetic field. Extragalactic emissions are mainly synchrotron and dust radiation from radio galaxies behaving such active galactic nuclei and dusty star-forming galaxies, respectively, and Sunyaev-Zeldovich effects from galaxy clusters. This section will describe each emission forming the data used along the PhD thesis for training the neural networks.

2.2.1 Interstellar dust

Interstellar dust is present in almost all wavelengths of astronomical insterest. As a first approximation it is mainly formed by amorphous silicate and carbonaceous materials (Draine, 2003). However, after the high-sensitivity emission measures by *Planck*, providing the scientific community with unprecedented detailed dust maps up to 5 arcmin resolution, several dust models from the 90s used in the pre-*Planck* simulations become unrealistic, especially in polarization.

In a recent concise review, Hensley and Draine (2021) described the constraints of interstellar dust in the post-*Planck* era. With respect to the dust grains morphology, they show that the grains in the interstellar medium (ISM) emitting in the microwave regime are mainly formed by heavy elements ejected from the bulk of stars to the ISM. Then, while some of the atoms remain in the gas, a fraction of them get locked in the grains. In particular: C, O, Mg, Si and Fe are mainly the elements forming the interstellar dust mass, while elements such as Al, S, Ca and Ni compose a minor fraction of the total mass. Their actual abundances in the solar mass are corrected at the time of the Sun's formation by using several chemical evolution models or by using observations of young stars in order to estimate their abundances in the ISM and therefore in the dust grains emitting in the microwave regime [3].

With respect to their emission, the bulk of the dust grains are heated to $\sim 20K$, which imply that they emit thermally in the far infrared (FIR). Both DIRBE and FIRAS instruments aboard COBE constrained the dust spectrum between 3.5 and 1000 μm. After the unprecedented detailed measurements of the FIR-submillimeter dust emission over the full sky, *Planck*

[3]The reader is encouraged to see the references therein for detailed information about these observations.

constrained the dust spectrum from 80 to 10000 μm (Planck Collaboration (2014c), Planck Collaboration (2015), Planck Collaboration (2016d)).

Furthermore, in the beginnings of the 00ties, it was found that both aligned and aspherical grains are responsible of polarizing the starlight they absorb, being published the first full-sky polarized dust by WMAP, and found that the dust spectral energy distribution (SED) is consistent with a modified blackbody with a power-law opacity $\kappa_\nu \propto \nu^\beta$. After several releases, *Planck* found agreement with a modified blackbody with a temperature $T_d = 19.6K$ and $\beta = 1.53 \pm 0.02$ on the dust SED. However, this is a first order approximation. It has been found that, even on the degree scale, both parameters show variations in the sky (Planck Collaboration (2016b), Planck Collaboration (2016c)). This is therefore a challenge for component separation algorithms, which have to known the spectral variation of dust along the sky.

Moreover, *Planck* observation has confirmed in Coulton and Spergel (2019) that dust power spectrum, although can be approximated at first order with a power law, $A_d(l) \propto l^{\alpha_d}$, with $\alpha_d \approx -2.4$, has high level of non-Gaussianity with non-zero polarized bispectra. In fact, this mislead characterization of polarized dust emission has led to a false detection of primordial B-modes from BICEP experiment. In order to reach those levels of precision, dust simulations must have enough information about the smaller scale structures from dust and their variation along the sky, which is actually a limitation due to the higher levels of contamination from systematics and instrumental noise.

2.2.2 Synchrotron emission

At lower frequencies ($\lesssim$ 30 GHz in temperature and $\lesssim$ 100 GHz in polarization), as shown in Fig 2.1, the sky signal is dominated by synchrotron emission.

Synchrotron radiation is emitted by high energy electrons gyrating in a magnetic field. Ginzburg and Syrovatskii (1965) reviewed the relation between non-thermal emission and synchrotron radiation. His model from the early 50ties was improved in Ginzburg and Syrovatskii (1969) and supported by maps of Galactic and extragalactic sources and by the discovery of the polarized nature of the radiation. Following that theoretical model, the observed non-thermal emission comes from high-energy relativistic electrons spiralling the weak Galactic magnetic fields.

As previously shown, dust constraints from *Planck* allowed a reliable characterization for future models in both temperature and polarization. This is not the same for synchrotron

emission, due to the sensitivity of the LFI *Planck* instrument. In fact, models used in recent collaborations for simulating the microwave sky are based on the power law scaling with a spatially varying spectral index from the Haslam 408 MHz emission maps (Haslam et al. (1981), Haslam et al. (1982)). Its spectral index was constrained using WMAP and *Planck* data as $\beta_s \sim -3$. However, post-*Planck* observations using S-PASS experiment (Krachmalnicoff et al. (2018), Fuskeland et al. (2021)) show that the spectral index vary from $\beta_s \approx -2.8$ at low Galactic latitudes to $\beta_s \approx -3.3$ at higher ones. Also using BICEP2 and SPIDER fields, it is found a mean synchrotron spectral index of $\beta_s \approx -3.22 \pm 0.06$ and $\beta_s \approx -3.21 \pm 0.03$ (Fuskeland et al., 2021). In any case, characterizing synchrotron radiation is crucial for constraining the tensor-to-scalar ratio at values lower than $r = 0.001$.

2.2.3 Cosmic infrared background

The cosmic infrared background (CIB) is formed by accumulated emissions over the entire history of the Universe, from nucleosynthesis sources to sources present in the first-stars-era and actual new stellar populations (Kashlinsky et al., 2018). After dividing CIB observations into 2-5 μm and 1-2 μm wavelength ranges, there is general agreement between models for the nature of CIB fluctuations for the first range, although the second range is actually not well constrained. *Euclid* mission will observe with unprecedented detail the CIB anisotropies from high redshifhts, allowing the scientific community to study its properties.

2.2.4 Thermal Sunyaev-Zeldovich effect

Thermal Sunyaev-Zeldovich is caused by CMB photons scattering off hot electrons in galaxy clusters (Sunyaev and Zeldovich, 1972), being the effective spectrum no longer a perfect black-body. Its signature observation is that, bellow 217 GHz, the effective SZ spectrum is negative and above that frequency it is positive. Around 217 GHz, the effective power spectrum is around zero. Moreover, Planck Collaboration analyzed the SZ effect in several clusters such as the Coma and Virgo ones, and published a reliable catalogue of SZ clusters in the sky [4].

[4]The reader is encouraged to see *Planck* releases for more information.

2.2.5 Other emissions

As shown in Figure 2.1, there are several other emissions in the microwave sky. Since they are mainly secondary contaminants for the CMB, they will be superficially described in this Phd thesis.

- **Free-free**

  Free-free emission is produced by electron-ion interactions in the ionised phased of the ISM. Dickinson et al. (2003) model was used in pre-*Planck* simulations, describing the variation of the spectral index of free-free emission from -2.12 to -2.2 considering both the frequency and the electronic temperature. After *Planck*, free-free emission is more reliable modeled by using the two-component model by Draine (2011). The spectral index then varies from -2.13 at 500 K to -2.15 for 20000 K.

- **Spinning dust**

  Spinning dust, also called Anomalous Microwave Emission, is produced by rotational dust grains due to non-zero electric dipole moments, as explained in Draine and Lazarian (1998). Although it was firstly observed in the late 90s, *Planck* gave unprecedented detailed observations of this effect (Planck Collaboration, 2014b). Actually, the Ali-Haïmoud et al. (2009) accurately model this effect. The method is based on a sum of two spinning dust populations, one having a spatially varying peak frequency, and the other one having a spatially constant peak frequency.

- **Molecular lines**

  Molecular line emission from dense molecular clouds in our Galaxy and external ones also pollutes the CMB maps. In particular, ^{12}CO lines strongly contaminate the signal at 115.27 GHz, 230.54 GHz and 345.80 GHz for $J = 1 \longrightarrow 0$, $J = 2 \longrightarrow 1$ and $J = 3 \longrightarrow 2$, respectively, which are near the HFI *Planck* channels 100, 217 and 353 GHz. Other lines such as CN, HNC, C_2H, HCO+, CS and many others also contaminate the signal at 5-10% levels.

2.3 Point sources

A Point Source (PS) is a single identifiable localized source of an emission. In astronomy, a PS usually adopt the shape of the instrument's beam through the Point Spread Function (PSF),

as is shown in figure 2.8.

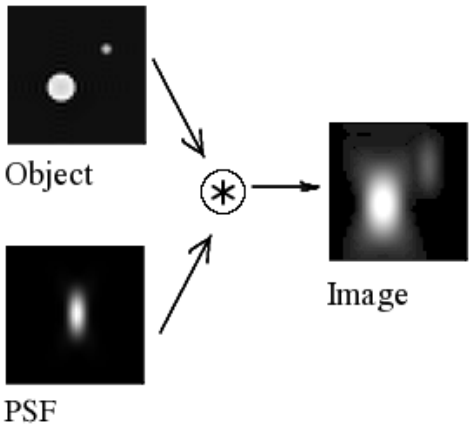

Figure 2.8: Illustration of a PS arising from the convolution of an object with the PSF.

A compact source in an astronomical map is similar to a point-like one with a certain flux density. While luminosity is the radiant flux from a source over the power of signal, being an absolute measure of radiated electromagnetic power, and quantified by Watts (W), flux density can be defined as the luminosity received by a surface per unit area. Its units are W/m^2. Another way to define this quantity is as the irradiance of a surface per unit frequency or wavelength. In this case, its units are $W/m^2/Hz$. In Astrophysics, the flux density unit is usually the Jansky (Jy), which are related to previous units as $1\ Jy = 10^{-26}\ W m^{-2} Hz^{-1}$.

PS are relevant in CMB science. At low frequencies (< 10 GHz), early evolutionary models of radio sources (Toffolatti et al. (1998), de Zotti et al. (2005)) provided accurate fits for them. In particular, the first one remarkably constrained the number counts of WMAP sources and was used by González-Nuevo et al. (2005) to give predictions for *Planck* counts. Moreover, González-Nuevo et al. (2008) used the second evolutionary model to also estimate statistical properties of WMAP sources. Both models also give good constraints for the number counts of radio sources at high frequencies (≥ 100 GHz), after a statistical extrapolation of flux densities by adopting a simple power-law with a fixed spectral index. Furthermore, as explained by Tucci, M. et al. (2011), that is a first approximation of the behavior of the active galactic nuclei (AGN) core at high frequencies, due to a steepening of the spectral index at $\nu > 30$ GHz [5], which could be more relevant in blazars at frequencies in the range $10 - 1000$ GHz. In fact, a steepening of the spectral index from 0.5 to 1.2 was observed in *Planck*, as shown in Figure 2.10. Furthermore, at those frequencies, the model by Tucci, M. et al. (2011) is

[5] See references in Tucci, M. et al. (2011) for detailed information.

actually the state-of-the art in order to estimate the number counts of radio sources. Indeed, it was used in pre-*Planck* simulations Delabrouille et al. (2013), giving remarkable fits compared with observations from *Planck*, as shown in their two published catalogues of compact sources (Planck Collaboration (2014a), Planck Collaboration (2016a)).

The model by Tucci, M. et al. (2011) is basically an extrapolation of the 5GHz number counts to higher frequencies by following the next assumptions:

1. Three different populations of radio sources are considered, based on their spectral index α_1^5 in the $1-5$ GHz frequency range: steep-spectrum sources if $\alpha_1^5 < -0.5$, flat-spectrum ones if $-0.5 \leq \alpha_1^5 < 0.3$ and inverted-spectrum ones if $\alpha_1^5 \geq 0.3$.

2. Sources are simulated at 5 GHz using the models by Toffolatti and de Zotti. More particularly, for flat and inverted sources, the number counts are given by fitting observational data using
$$\bar{n}(S) = n_0 \frac{(S/S_0)^k}{1-e^{-1}}(1-e^{-(S/S_0)^{l-k}}), \tag{2.10}$$
where $\bar{n}(S)$ is the differential number counts normalized to $S^{5/2}$, $n_0 = 47.4\ Jy^{-1}sr^{-1}$, $S_0 = 1.67Jy$, $k = 0.5$ and $l = -0.66$, as found in observations. On the other hand, for steep-spectrum sources the number counts are estimated as the difference between the previous fit and the total number counts predicted by Toffolatti model. The fraction of *Blazar Lacertae* (BL Lacs) sources from flat-spectrum populations are taken at 5 GHz by using the evolutionary model by de Zotti.

3. Then, the extrapolation is done by using several spectral models. For steep-spectrum sources, it is used
$$S(\nu) = S_{5GHz}(\nu/5)^{\alpha_{hi}} \tag{2.11}$$
where $\alpha_{hi} = \alpha_1^5 - \Delta\alpha$, being the spectral steepening $\Delta\alpha$ extracted from a Gaussian distribution with average 0.3 and dispersion 0.2, as founded in observations. For inverted-spectrum sources it is used
$$S(\nu) = S_0(\nu/\nu_0)^k(1-e^{-(S/S_0)^{l-k}}), \tag{2.12}$$
where ν_0 is related to the peak frequency in the spectrum and $S_0 = (1-e^{-1})S(\nu_0)$, $k = \alpha_1^5$ and l value us fitted by using observations. For flat-spectrum sources there are used four different spectral models. As seen in Tucci, M. et al. (2011), the C2Ex one is the one

better fitting the data. In that model, spectra is described by

$$S(\nu) = \begin{cases} S_{5GHz}(\nu/5)^{\alpha_{flat}} & 5 \leq \nu \leq \nu_M \\ S(\nu_M)(\nu/\nu_M)^{\alpha_{steep}} & \nu \geq \nu_M, \end{cases}$$

where ν_M is the break frequency, which is the one related to the transition from the optically thick to the optically thin regime of the synchrotron radiation from the central black hole of the active galactic nuclei, $\alpha_{flat} = \alpha_1^5 - \Delta\alpha$ is the spectral index in the optically thick regime and $\alpha_{spectral} = -0.8 \pm 0.2$ is the spectral index in the optically thin regime. More particularly, for this model $0.01 \leq r_M \leq 0.3$ parsecs, which is the smallest radius from which optically-thin synchrotron emission can be observed with $\alpha < -0.5$.

The other kind of point sources are called far-IR sources. They are mainly early-type galaxies, formed by old stellar populations with mass-weighted ages of $\gtrsim 8 - 9$ Gyr, corresponding to redshifts $z \gtrsim 1 - 1.5$, and spiral and irregular galaxies formed by significantly younger stellar populations (see Lapi et al. (2011), Cai et al. (2013) for more detailed information) of about $\lesssim 7$ Gyr, corresponding to redshifts $z \lesssim 1$. Therefore, their emission in the microwave regime is mainly due to high amounts of dust while forming their stellar populations.

2.3.1 Point sources are also polarized

Due to their synchrotron radiation from relativistic electrons in their jets and lobes, point sources can be highly polarized, with maximum values of $\sim 70 - 75\%$ of linear polarization, as explained in Tucci and Toffolatti (2012). However, with actual instruments, only a few sources are observed to be polarized, with a total fractional polarization value, $\Pi = P/S$, as high as $\sim$10% of the total flux density, S. Although improvements have been made in the total number of detected sources last years after *Planck* by using more sophisticated ground-based instruments with much higher resolution such as the Atacama Cosmology Telescope (ACT) and the South-Pole Telescope (SPT), the properties of polarized point sources are not well constrained yet, which is not only a lack of knowledge in the AGN research but also relevant for CMB studies since they constitute a major contaminant for the primordial B-mode detection when $r < 10^{-4}$, as predicted by Puglisi et al. (2018).

Since the estimated polarization is only a few percentage of the total flux density of the source, it is theoretically believed that the magnetic fields formed in the AGN of radio sources are highly inhomogeneous, or almost without ordering, as explained in concise reviews

(Ginzburg and Syrovatskii (1969), Saikia and Salter (1988), Mesa, D. et al. (2002)). However, observations based on measurements of the orientations of the linear polarization from the jets and lobes in the sources show coherent structures across the images. This change of ordering can be related not only to the degree of polarization but also to the change in the position angle of the linearly polarized radiation passing through a magneto-ionic medium, an effect known as Faraday rotation. Therefore, the position angle, also known as polarization angle, is one of the most relevant characteristics of extragalactic radio sources. Actually, its nature is observationally unconstrained and believed to be random.

With respect to the polarization, it is generally used the fraction with respect the total flux density in order to characterize the degree of polarization. Due to the challenging of the detection in faint signals as in Q and U polarization maps, completeness levels are hard to achieve, and it is more usual to work with the probability function $\mathcal{P}(\Pi)$ of the polarization fraction. Although it is only an approximation, as shown in Puglisi et al. (2018), after assuming that polarization fraction and flux density are uncorrelated and statistically independent, one can model $\mathcal{P}(\Pi)$ by means of a log-normal distribution

$$\mathcal{P}(\Pi) = \frac{A}{\sqrt{2\pi\sigma^2}\Pi} exp\left[-\frac{(ln(\Pi)/\mu)^2}{2\sigma^2}\right], \tag{2.13}$$

where μ and σ are the median and the standard deviation, respectively. Furthermore, the fractional polarization can be then approximated by the log-normal parameters

$$\begin{aligned} \langle \Pi \rangle &\approx \mu e^{\frac{1}{2}\sigma^2} \\ \langle \Pi^2 \rangle &\approx \mu^2 e^{2\sigma^2} \\ \Pi_{med} &\approx \mu \end{aligned} \tag{2.14}$$

In the last years, different values of the mean and median fractional polarization have been published at all microwave regimes, all of them bellow 10% of the total flux density, as predicted in Tucci and Toffolatti (2012). Table 2.9 from Datta et al. (2018) collect the number of sources and statistical degree of polarization in different catalogues published in the last years, being the analysis by Bonavera et al. (2017a), Bonavera et al. (2017b) and Trombetti et al. (2018) the most relevants when dealing with *Planck* sources as in this PhD thesis. On the one hand, Bonavera et al. (2017a) and Bonavera et al. (2017b) used the stacking technique for constraining the mean polarization fraction of radio and dusty sources respectively in Planck data, reaching values of about 3%, as seen in Table 2.9. On the other hand, Trombetti et al. (2018) reanalysed the median fractional polarization at the same *Planck* channels by using the

intensity distribution analysis method, obtaining values between 2 and 3%. In those all recent works, the degree of polarization seems to be independent of both frequency and total flux density.

Reference	N_{srcs}	Freq (GHz)	Polarization fraction	Remarks
Condon et al. (1998)	30 000	1.4	SS*: 1.1% median, 2% mean SS: 1.8% median, 2.7% mean FS*: 1.3% median, 2% mean	$S_{1.4} > 800$ mJy $200 > S_{1.4} > 100$ mJy
Mesa et al. (2002)	8032	1.4	SS NVSS sources: 1.82% median	$200 > S_{1.4} > 100$ mJy
	3700		SS NVSS sources: 1.45% median	$400 > S_{1.4} > 200$ mJy
	1438		SS NVSS sources: 1.37% median	$800 > S_{1.4} > 400$ mJy
	660		SS NVSS sources: 0.74% median	$S_{1.4} > 800$ mJy
	6198		FS/IS NVSS sources: 1.84% median	$200 > S_{1.4} > 100$ mJy
	2859		FS/IS NVSS sources: 1.50% median	$400 > S_{1.4} > 200$ mJy
	1150		FS/IS NVSS sources: 1.32% median	$800 > S_{1.4} > 400$ mJy
	496		FS/IS NVSS sources: 1.05% median	$S_{1.4} > 800$ mJy
			2.2% median combined	$S_{1.4} > 800$ mJy Anti-correlated with flux density
Klein et al. (2003)	106	2.7, 4.85, 10.5	SS: 2 – 6% median, FS: 2.5% median	1.4–10.5 GHz
Ricci et al. (2004)	197	18.5	FS: 2.7% median, SS: 4.8% median	Weakly correlated with $\alpha_{5-18.5}$
Sadler et al. (2006)	108	20	2.3% median	Anti-correlated with flux density
López-Caniego et al. (2009)	138, 122, 93, 81	23, 33, 41, 61	1.7%, 0.91%, 0.68%, 1.3% mean	Statistical estimates for bright WMAP sources
Murphy et al. (2010)	768	20	2.6% median, 2.7% mean	FS: 2.9%, SS: 3.8%
Agudo et al. (2010)	149	86	1.5%	Flat-radio-spectrum AGNs
Battye et al. (2011)	105	8.4, 22, 43	2% median, 3.5% mean	Independent of frequency, flux density
Sajina et al. (2011)	159	4.86 – 43.34 (four bands)	typically 2–5%, tail extending to ~15%	Slight increase with frequency, trend is stronger for SS and dim sources
Bonavera et al. (2017b)	881	30	3.05% mean	Fractional polarization estimated by applying the stacking technique on 881 sources detected in the 30 GHz Planck map that are outside the Planck Galactic mask
		44	3.27% mean	
		70	2.51% mean	
		100	3.26% mean	
		143	3.06% mean	
		217	3.07% mean	
		353	3.52% mean	
Puglisi et al. (2017)	32	95	FS: 2.07% median	Independent of frequency
Trombetti et al. (2018)	35	30	3.3% median	Independent of frequency, flux density
	9	44	2.2% median	Fractional polarization estimated using the Intensity Distribution Analysis (IDA) method
	4	70	2.8% median	
	14	100	1.9% median	
	15	143	2.9% median	
	8	217	3.1% median	
	1	353	3.0% median	
This work	169	148	$2.8 \pm 0.5\%$	Independent of flux density

Note. *SS, FS, and IS denote source populations with steep, flat, and inverted spectrum, respectively.

Figure 2.9: Actual fractional polarization constrains for radio and infrared sources Datta et al. (2018).

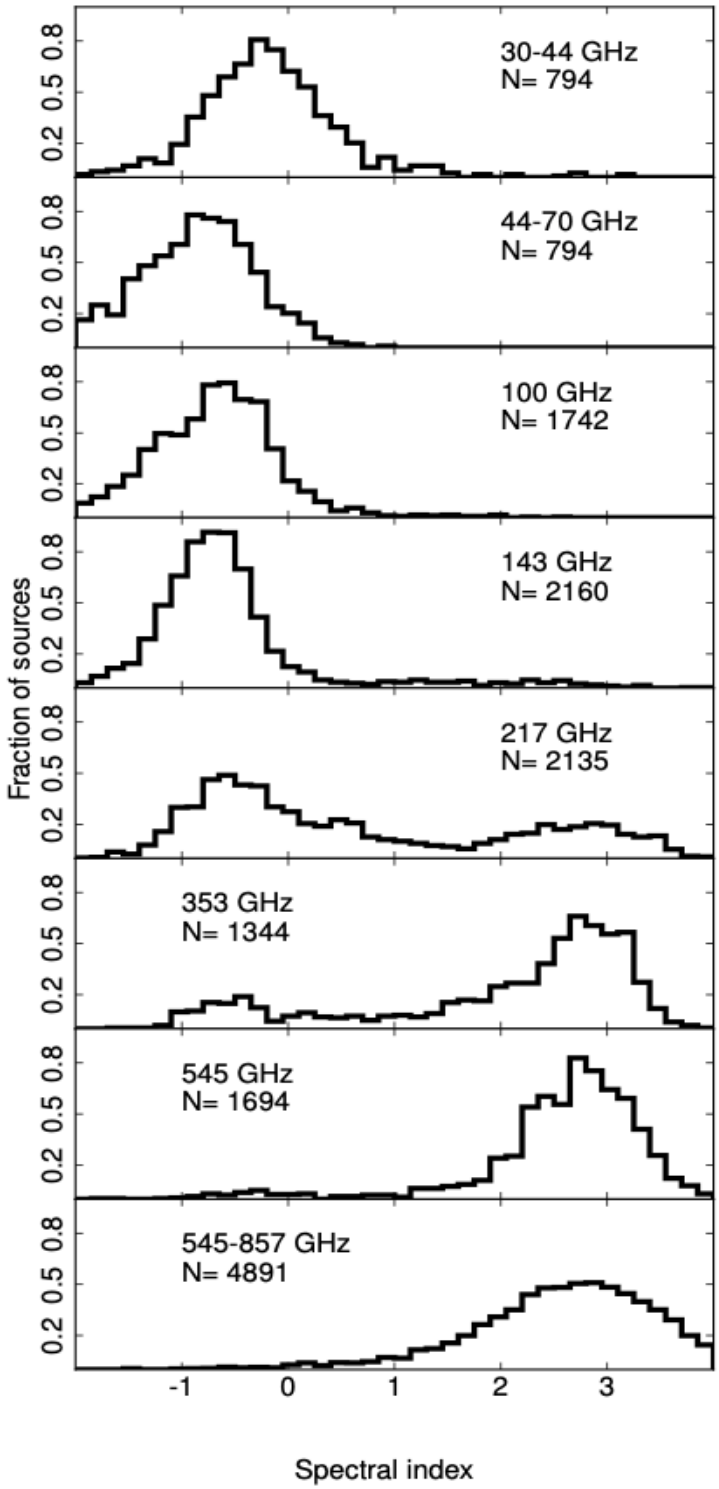

Figure 2.10: Fraction and number of radio and infrared sources with respect to their spectral index distribution for each *Planck* channel Planck Collaboration (2016a).

2.4 Simulations

This section will resume the process for simulating the microwave sky described above. Along this PhD thesis, it has been used 2-dimensional patches of the total simulated sky for training and testing the neural networks. The frequencies used along the thesis are the 100, 143, 217 and 353 GHz HFI channels of the Planck satellite. The sky is simulated in both temperature and polarization. Depending on each study of this PhD thesis, the sky is divided into latitude regions, as can be seen in Figure 2.11, because of the strong contamination of our Galaxy. Black lines delimite the extragalactic region (ER), with simulations above $|b| > 30^\circ$ of latitude, black and white lines contain the Galactic region (GR, $10^\circ < |b| < 30^\circ$) and between white lines is represented the Galactic plane (GP, $|b| < 10^\circ$).

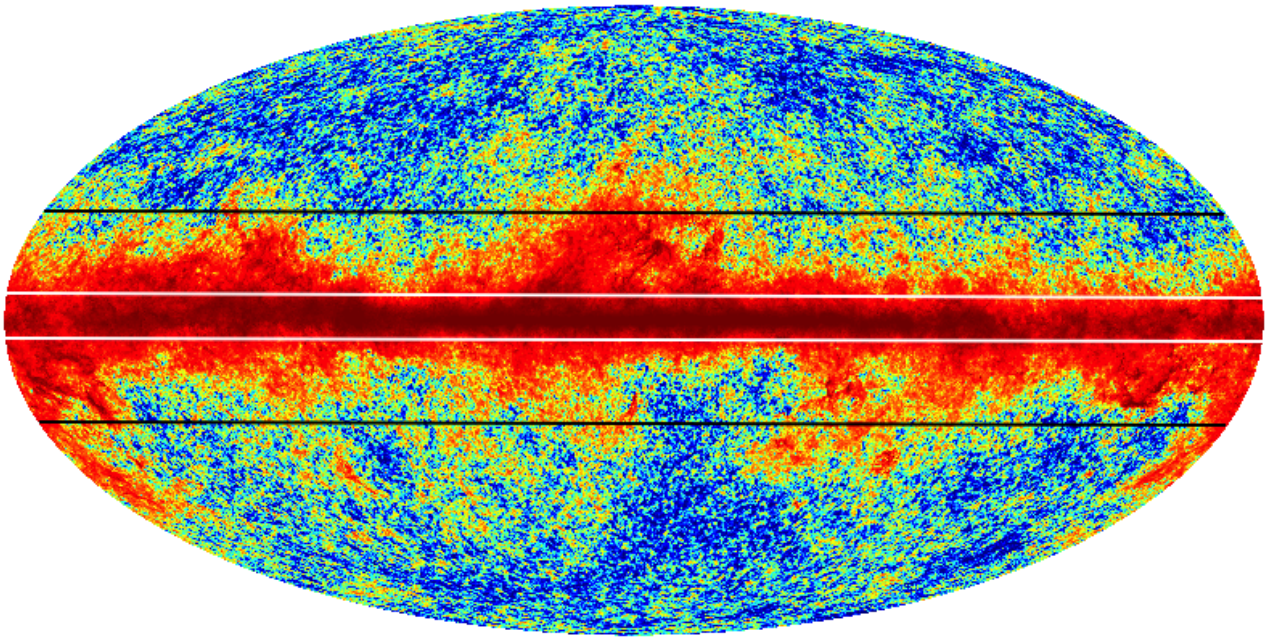

Figure 2.11: The microwave sky as seen by *Planck* at 217 GHz. Horizontal lines delimite the different regions analyzed along this PhD thesis: black lines delimite the extragalactic region, black and white lines contain the Galactic region and between white lines is represented the Galactic plane (*ESA image*).

2.4.1 Temperature simulations

Temperature sky is used in both section 4 and section 6 for the detection of point sources and for recovering the CMB, respectively. In both cases, the used frequencies are 143, 217 and 353 GHz. Then, in temperature, the three maps are simulated before cutting the patches. Foreground and CMB maps are downloaded from the Planck Legacy Archive (PLA[6]) database. It is a reliable website containing all the releases from *Planck* and also the simulated sky from the models best fitting the published data.

Firstly, a position in the sky is defined by means of a longitude and latitude parameters. They are the centered position of each patch of the sky. Then, both the CMB and foreground maps are cutted by taking this position, the patch area and the pixel size. In both cases, the pixel size is 90 arcsec, and the patch area is 128×128 pixels for the detection of point sources, and 256×256 pixels for recovering the CMB. For cutting the patches, it is used the gnomonic projection of the HEALPix framework. The CMB is a simulation for a ΛCDM model Universe, with the cosmological parameters published in *Planck*. At these frequencies, as shown in Figure 2.1, only thermal dust pollutes the signal. Then, it is also downloaded from the PLA and, following its documentations, it is based on the GNILC map applied to the *Planck* second release, extrapolated to the other channels assuming a modified blackbody emission law with the spectral index map from *Planck* 2013 first release.

Thermal Sunyaez-Zeldovich effect is also included in the simulations after downloaded from the PLA. Also following its documentation, clusters are generated by following the number counts method of Delabrouille et al. (2002).

Finally, random white instrumental noise is added to the patch. In particular, the *Planck* sensitivity for each channel is: 0.55, 0.78 and 2.56 $\mu K_{CMB} deg$ (Planck Collaboration, 2020a).

After forming the patch with these emissions (which is called Input Total along this PhD thesis), point sources are injected on each patch by simulating their number counts at 217 GHz with the C2Ex model by Tucci, M. et al. (2011). Their flux density is assigned by using the CORRSKY method by González-Nuevo et al. (2005), which was used in *Planck* for simulating point source catalogues. Their spectral behavior is considered by extrapolating their flux density below/above the 217 GHz channel. The spectral index for radio and infrared sources is estimated by assuming a Gaussian distribution of the published spectral index for each channel (Planck Collaboration, 2016a), as shown in Figure 2.10.

[6]https://pla.esac.esa.int/#home.

The CIB is simulated by assuming that it is entirely formed by proto-spheroidal galaxies in the process of forming their stellar masses. For simulate their number counts, the model by Cai et al. (2013) is used as so as the angular power spectrum by Lapi et al. (2011). Their flux density is also assigned by using CORRSKY. Their spectral behavior is considered by using the same spectral index estimated for the infrared point sources.

An example of a patch for each simulated emission of the microwave sky cited above is represented in Figures 2.12 and 2.13.

2.4.2 Polarization simulations

Polarization simulations, used in sections 5 and 7, are similar to the temperature ones. In this case, the frequencies used are 100, 143 and 217 HFI *Planck* channels, due to the low sensitivity of the 353 GHz channel. For constraining the polarization flux density and angle of point sources, the patches are cutted by also using the gnomonic projection of the HEALPix framework, with a patch area of 32×32 pixels. For recovering the CMB, the methodology for cutting the patches is the projection by Krachmalnicoff and Puglisi (2021), since it was found during the development of the work by Casas et al. (2023b) that, when using the HEALPix projection, a strong E-to-B leakage (E-mode polarization contaminating the B-mode spectrum) dominates the power spectrum.

Then, as for the temperature simulations, several emissions formed the Input Total patch. As shown in Figure 2.1, Galactic emission at these frequencies is due to synchrotron and dust. Both of them, as so as the CMB, are downloaded from the PLA in Q and U stoke parameter maps. Since the power spectra of both CIB and thermal Sunyaev-Zeldovich effect are several orders of magnitude bellow the CMB signal, these emissions are not included in the polarized sky of this PhD thesis.

The CMB is also a simulation for a ΛCDM Universe with the published cosmological parameters by *Planck*, and both scalar and tensor modes.

Thermal dust, following the PLA documentation, is simulated with a realization of the Vansyngel et al. (2017) model at 353 GHz and extrapolated to the lower frequencies used in this work by using *Planck* second release dust maps and *Planck* first release dust spectral index maps. On the other hand, synchrotron emission is simulated by following a power law scaling with a spatially varying spectral index for such emission.

Point sources are differently injected into the patch, depending on each Chapter of the

PhD thesis: in Chapter 5, a single, central PS is injected into the patch. In Chapter 7, PS are distributed along the patch similarly than in temperature. In both cases, their polarization is taking into account by assuming that they follow a log-normal distribution with the parameters of Bonavera et al. (2017a).

Random instrumental noise is added to the patch similarly to the temperature case. In polarization, the sensitivity of *Planck* is 1.96, 1.17 and 1.75 $\mu K_{CMB} deg$.

An example of a patch with each simulated emission of the microwave sky cited above is represented in Figures 2.14 and 2.15 for Q, and in Figures 2.16 and 2.17 for U.

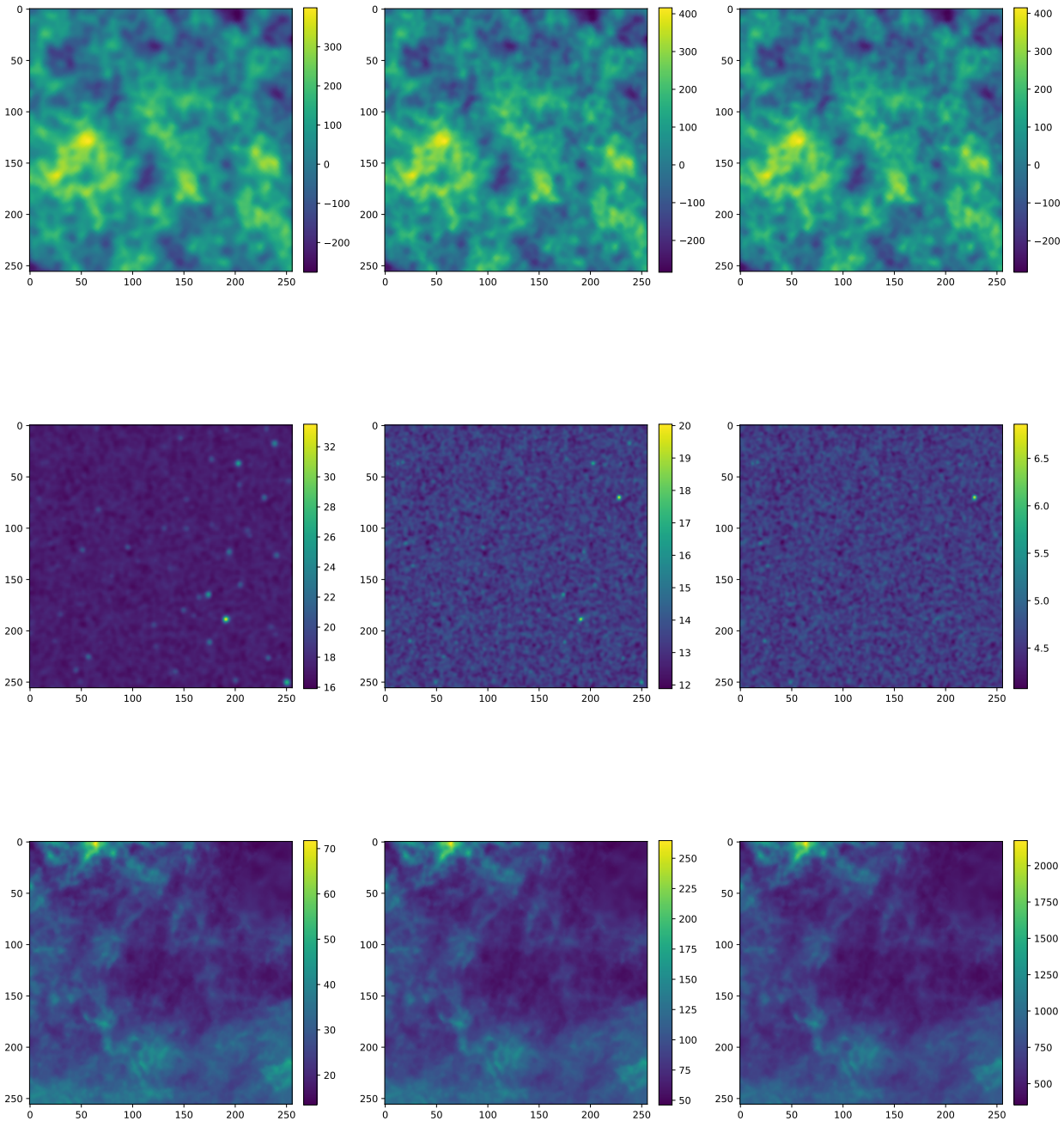

Figure 2.12: Example of each emission simulated to train the networks in temperature at 143, 217 and 353 GHz (from left to right). The emissions are the CMB, radio and dusty PS, and dust from top to bottom. The units, represented in the colorbars, are μK_{CMB}.

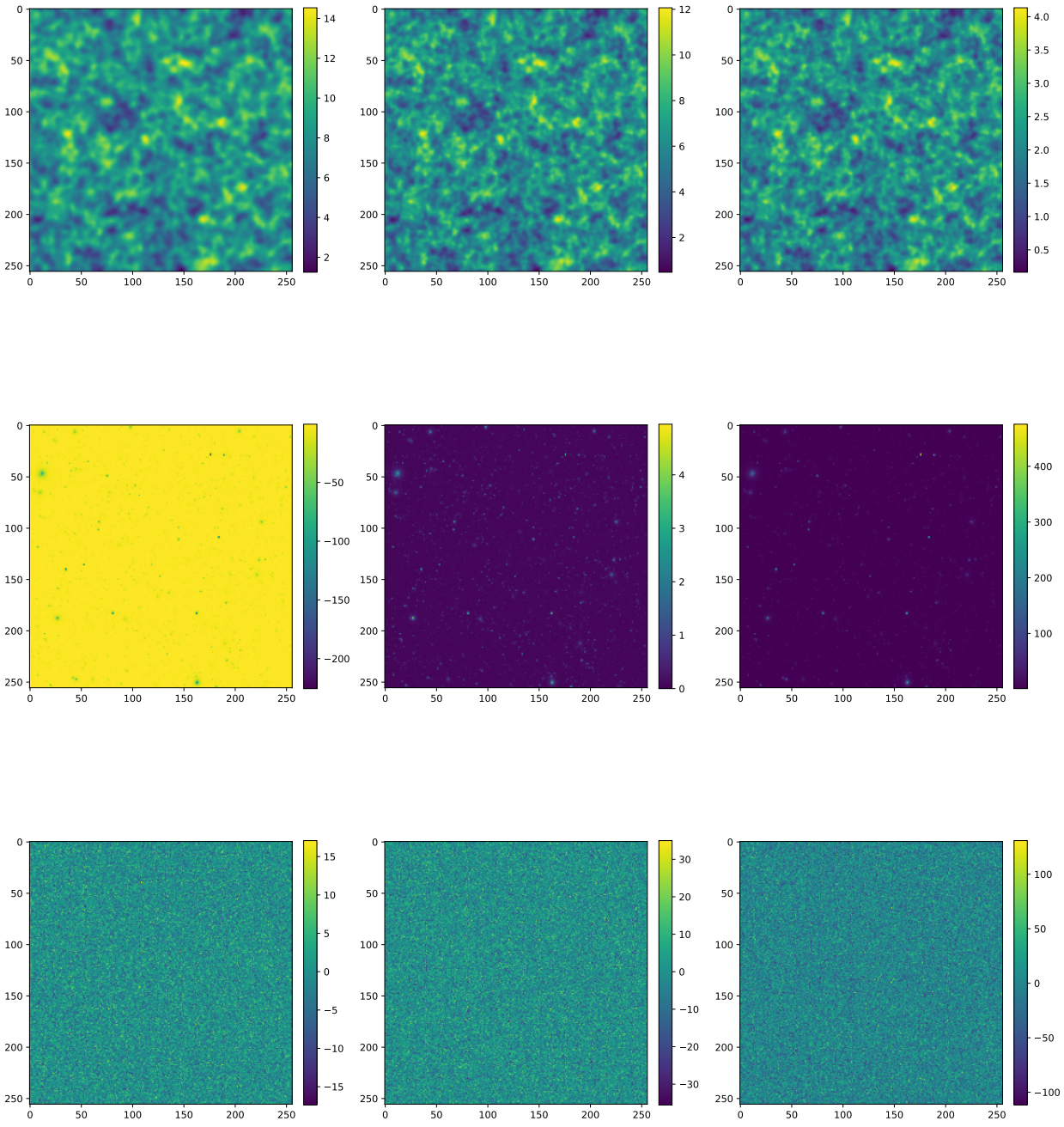

Figure 2.13: Example of each emission simulated to train the networks in temperature at 143, 217 and 353 GHz (from left to right). The emissions are the CIB, thermal SZ effect and instrumental noise from top to bottom. The units, represented in the colorbars, are μK_{CMB}.

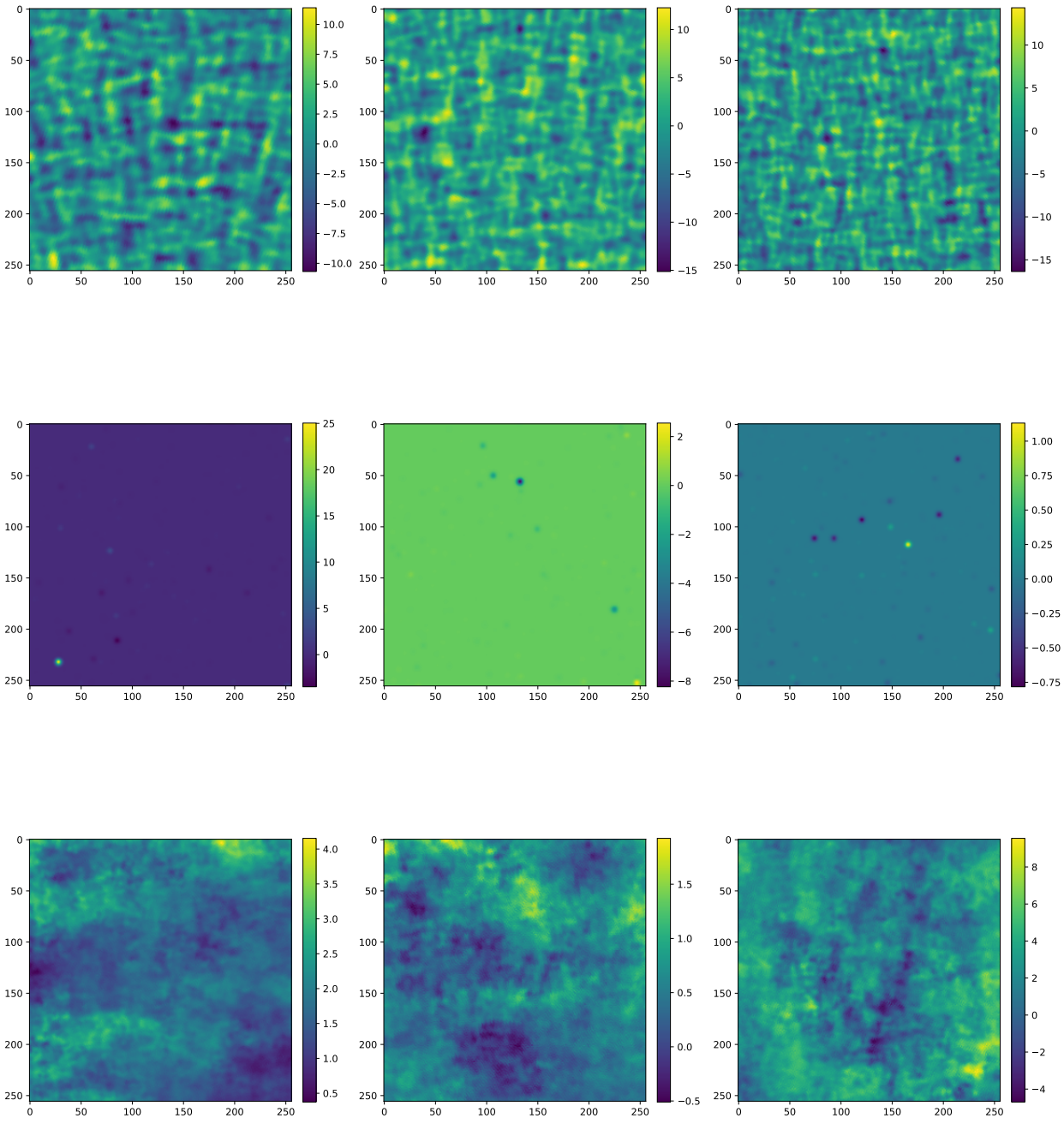

Figure 2.14: Example of each emission simulated to train the networks in Q polarization at 100, 143 and 217 GHz (from left to right). The emissions are the CMB, radio PS and dust from top to bottom. The units, represented in the colorbars, are μK_{CMB}.

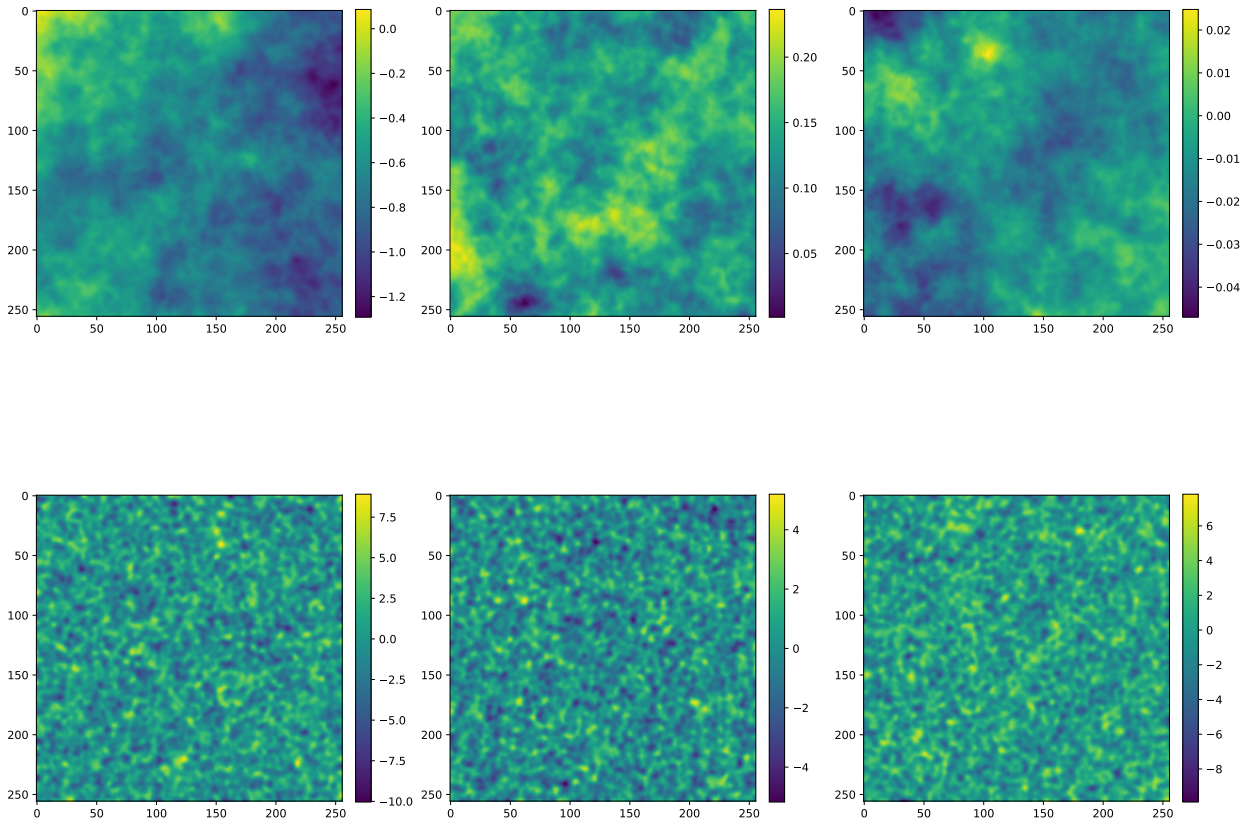

Figure 2.15: Example of each emission simulated to train the networks in Q polarization at 100, 143 and 217 GHz (from left to right). The emissions are synchrotron and instrumental noise from top to bottom. The units, represented in the colorbars, are μK_{CMB}.

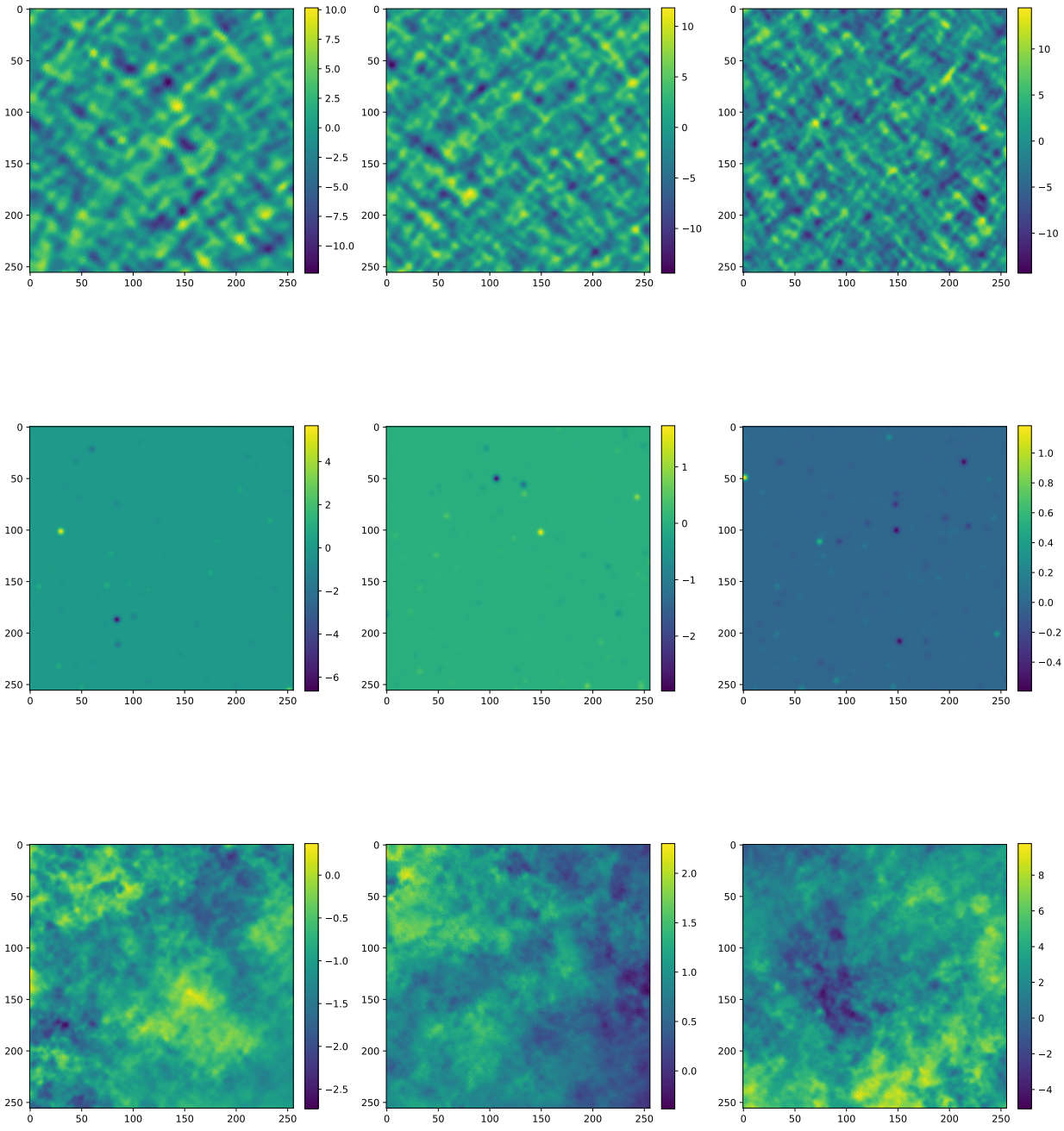

Figure 2.16: Example of each emission simulated to train the networks in U polarization at 100, 143 and 217 GHz (from left to right). The emissions are the CMB, radio PS and dust from top to bottom. The units, represented in the colorbars, are μK_{CMB}.

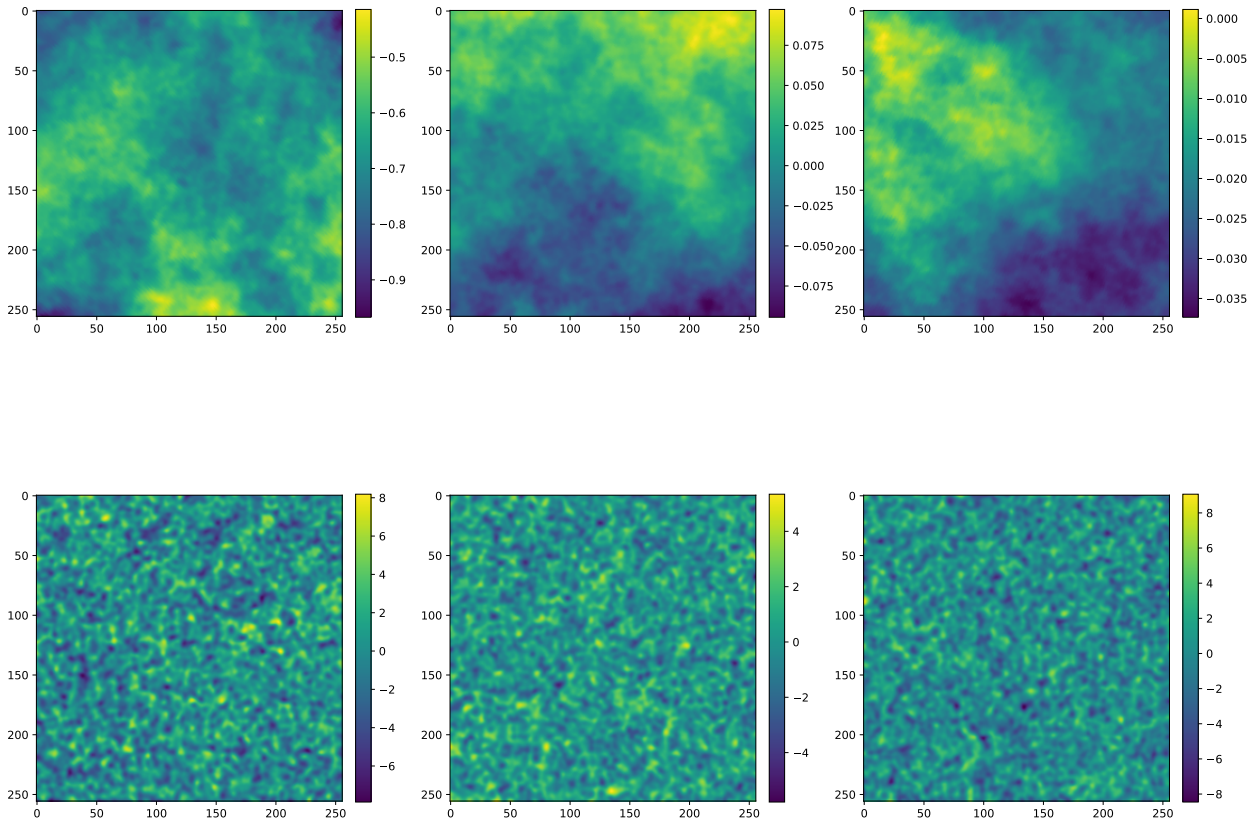

Figure 2.17: Example of each emission simulated to train the networks in U polarization at 100, 143 and 217 GHz (from left to right). The emissions are synchrotron and instrumental noise from top to bottom. The units, represented in the colorbars, are μK_{CMB}.

CHAPTER 3

Machine Learning

3.1 Machine Learning Basics

3.1.1 The Concept of Learning

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E (Goodfellow et al., 2016).

- **The Task, *T***

  Learning is the ability to perform a task. A ML algorithm can have tasks like classification, regression, anomaly detection and others. For example, a classification task or problem may consist to find a category for a input in the data. In the case of regression, the main objective will be to predict a numerical value using input information. Therefore, a classification ML algorithm can be used for example to asign authomatically a category to a galaxy based on input data, and then classify it. On the other hand, a regression ML algorithm can be used for example to predict the numerical value of a function, based on input data.

- **The Performance measure, *P***

  In order to evaluate how good a ML algorithm is, one must design a quantitative value to determine its performance. This measure is the performance measure P, and it is related to the task T. For tasks such as classification, the performance measure is often the accuracy of the model. This is the proportion of examples for which the model produces the correct output. In tasks such as regression, the performance measure is the error rate or loss, that is, the proportion of examples for which the model produces an incorrect output.

- **The Experience, *E***

  ML algorithms are generally categorized as unsupervised or supervised according to the

experience they are allowed to have during their learning. These types of learning can be seen schematically in figure 3.1. Unsupervised learning algorithms learns properties of a dataset from its structure. On the other hand, supervised learning algorithms learns behaviors from a dataset by using a label or target associated to each input data.

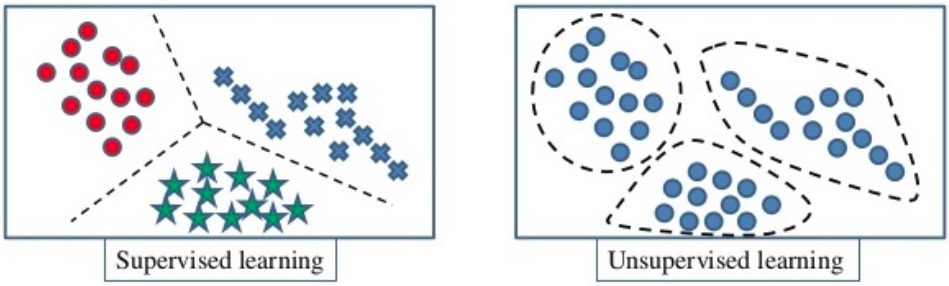

Figure 3.1: Supervised (left pannel) and unsupervised (right pannel) learning algorithms, in this case with the task of classifying samples into different categories. For Supervised learning, the data classifies between red dots, green stars and blue crosses according to the label imposed by the agent. For unsupervised learning, the algorithm classify the data into blue dot clusters according to the similarity of their intrinsic parameters (Goodfellow et al., 2016).

3.1.2 Overfitting and Underfitting

The main challenge in a ML problem is to perform well on inputs, but since in fields like Astrophysics or Cosmology the quantity of data increases every day, the idea is not to reduce the error in the input data but to reduce the expected value of the error on new input data. That is called the test error. To deal with that usually while training a ML algorithm is to have minimum two datasets: one for train the algorithm and another for testing it. Furthermore, the data on test dataset must not be contained on train dataset. This is to prevent a phenomenon called overfitting.

Overfitting occurs when there is a gap between train and test error, on practice, when the model works extremely well with present data, but it fails while evaluating with new data. On the other hand, underfitting occurs when the model is not able to obtain a sufficiently low error value on train dataset. On practice, underfitting can be interpreted as a model which cannot understand well the task for which it is trained. In figure 3.2 is represented a schematic view of this elements. As shown, the goal is to adjust one trained model with data. In the left image is presented an underfitting case: the model cannot fit the data because it has not sufficient previous knowledge or training. In the right image is shown an overfitting case: the model

works unrealistic well through fitting all points. This is because it works well with this data, but not necessarily well with other data. In the center image the model have a good fitting.

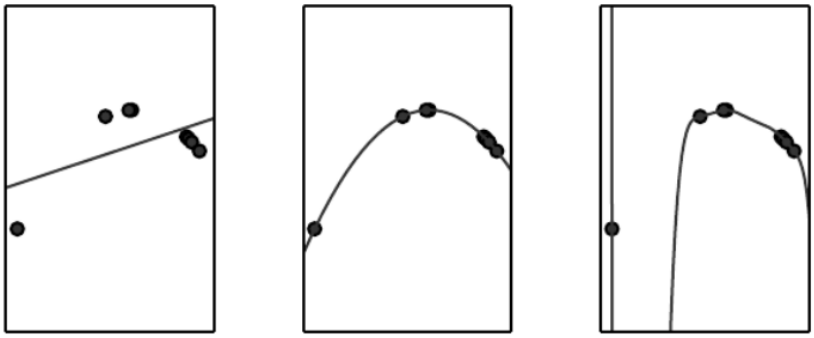

Figure 3.2: An schematic view of points and its fitting showing overfitting (right) and underfitting (left). In the center image is represented a good fitting (Goodfellow et al., 2016).

To deal with underfitting, one must use a great amount of input data to train the model. On the other hand, to deal with overfitting, the model performance should be evaluated with new data not contained on the train dataset.

3.2 Deep learning

Deep Learning is a field of ML that works with data to learn non-linear parameters. Its principal models are Neural Networks (NN), a type of models inspired on human brain with the mission of learning non-linear behaviors from the data. Moreover, the goal of a NN is to approximate some function f^*. For example, in a classification problem, the function $y = f^*(x)$ looks to map an input x to a category y. Then, in general, a NN define a mapping $y = f(x;\theta)$ in order to learn the value of the parameters θ to approximate a function. The simplest type of NNs are the Feedforward Neural Networks (FNN) or Multilayer perceptrons:

- They are called feedforward because the train data flows through the function evaluated at x, passing the intermediate computations used to define f, and finally reaching the output y.

- They are called networks because they are represented by composing functions, i.e $f(x) = f^{(3)}(f^{(2)}(f^{(1)}(x)))$. Each of these functions are called layers. Therefore, $f^{(1)}$ is the first layer of the neural network, $f^{(2)}$ is the second layer, and so on. The length of all the

composing functions is called the depth of the model, and that is the reason why these models are usually called deep models.

The first layer is called the input layer. The final layer is then the output layer. Since they are ML models, they are need to be trained. During this process, the objective is to drive $f(x)$ to obtain its approximate function $f^*(x)$. Each input data x is accompanied by a label $y \approx f^*(x)$. Then the output layer produce a value theoretically close to y.

Since the learning algorithm decide how to use the NN to produce a desired output by implementing the best approximation of f^*, the layers between the input and the output layers are called hidden layers. These layers are the main problem of NNs: if the number of hidden layers is high, then the number of parameters increase. In this case, it is difficult to follow the optimization on each parameter and compute that analytically implies a waste of time. Therefore, it is computed numerically and the researcher generally do not follow what happens inside the NNs. This is why NNs are usually called *black boxes*.

- They are called neural because they are inspired by neuroscience: each element of a layer is similar to a neuron in the brain, and each hidden layer is vector-valued. Thus, the functions $f^{(i)}(x)$ represent mathematically the behavior of biological neurons.

In figure 3.3 it is presented the general structure of a FNN. At the top panel is represented a neuron, formed by inputs x_i and weights w_i, which will adjust their value through a cost function f on each epoch of training. An epoch is the training step when backpropagation algorithm updates the weights by applying the learning algorithm through the entire training dataset. On the other hand, at the bottom panel it is represented the general architecture of a feedforward neural network.

3.2.1 Gradient-Based Optimization

Generally ML and DL models use optimization to learn behaviors from data. Optimization is the task of minimazing or maximizing some function $f(x)$ by altering x. The function is called the cost or loss function.

There are multiple loss functions to minimize a model. For example, for tasks such as classification, the Binary Crossentropy function is a well-suited loss function. For tasks like

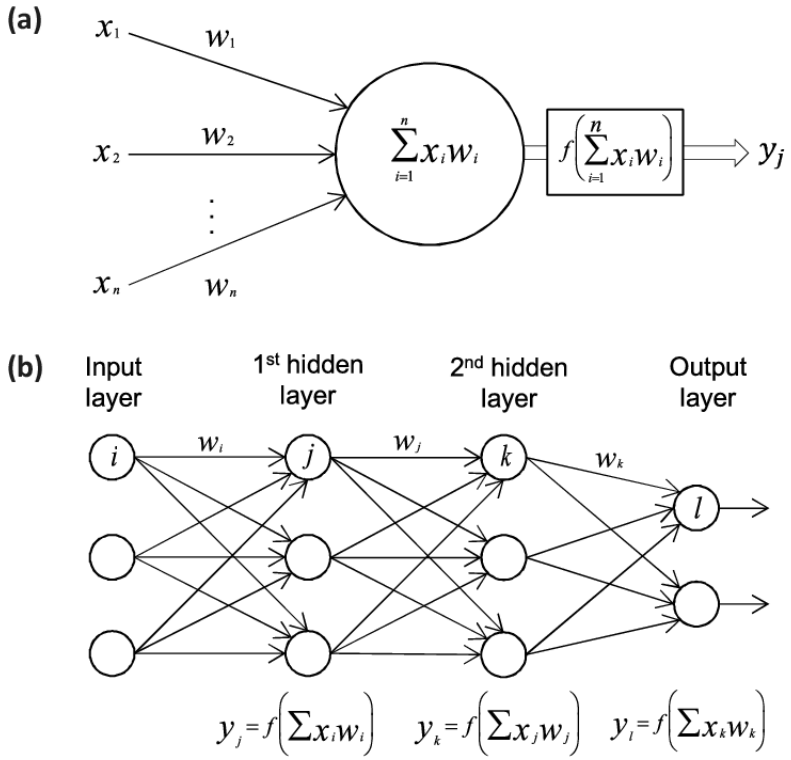

Figure 3.3: General architecture of a feedforward neural network model. In the figure, x_i are the inputs, w_i are the weights of the nodes or neurons and $y_i = f(\sum x_i w_i)$ are the cost functions (Goodfellow et al., 2016).

regression, the Mean Squared Error (MSE) works generally well. It has the form

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2 \tag{3.1}$$

where n is the number of points in the dataset, y_i are the data points, and $\tilde{y}_i$ are the data points fitted by the model. Therefore, the model would respond to certain inputs compared with their corresponding labels.

NNs are usually trained by using iterative, gradient-based optimizers that looks to obtain a low value for the cost function. This method is called gradient descent: a model is determined by a function $y = f(x)$, where x and y are real numbers. Its derivative is $f'(x) = dy/dx$ and it gives the change of $f(x)$ at the point x, in other words

$$f(x + \epsilon) \approx f(x) + \epsilon f'(x) \tag{3.2}$$

The derivative is useful for minimizing a function. As $f(x - \epsilon sign(f'(x)))$ have a lower value than $f(x)$ for a given ϵ, $f(x)$ can be reduced by displacing x in small steps with opposite sign of the derivative. This technique is how the gradient descent method works. An illustration of this method is shown in figure 3.4.

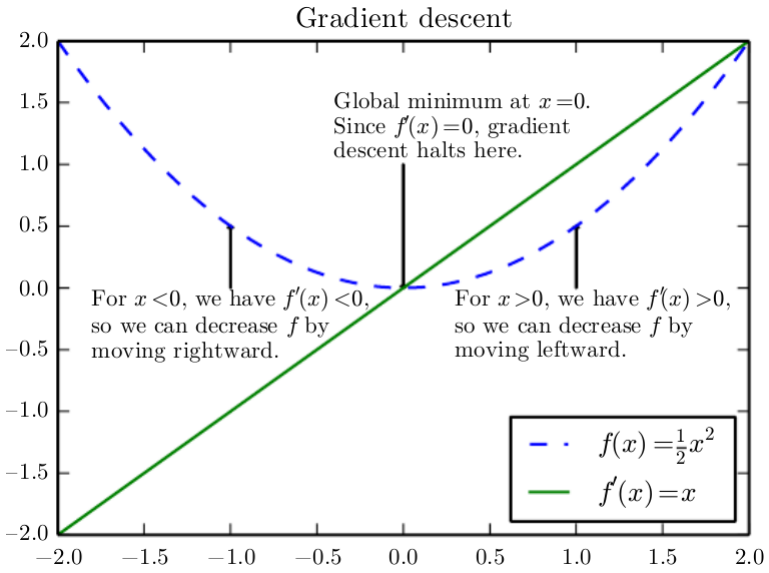

Figure 3.4: Scheme of how the derivatives of a function can be used to lead its way to a minimum (Goodfellow et al., 2016).

When $f'(x) = 0$, the model have no information about which direction should move. These points are known as critical or stationary points. A local minimum is a point where $f(x)$ is lower than all its neighboring points, therefore the model cannot decrease more. A point with the absolute lowest value of $f(x)$ is called a global minimum.

The inputs have the initial information and propagates through the net to the output layer. This phenomenon is called forward propagation. When the model compute the loss value from the cost function, the information goes back from the cost function to the input layers of the network, in order to compute the gradient. This is called the back-propagation algorithm (Rumelhart et al., 1986).

When the input and output of the function to minimize are both vectors, optimization is based on computing the jacobian matrix, defined as

$$J_{i,j} = \frac{\partial}{\partial x_j} f(x)_i \tag{3.3}$$

When the function has multiple input dimensions, the task is to compute the hessian matrix, defined as

$$H(f)(x)_{i,j} = \frac{\partial^2}{\partial x_i \partial x_j} f(x) \tag{3.4}$$

which is the jacobian of the gradient. In a NN, a critical point $\nabla_x f(x) = 0$ is examined in order to determine if a point is a local minimum. When the hessian is positive definite, that is, when the directional second derivative in any direction is positive, the point will be a local

minimum. However, in multiple dimensions, a single point can have different second derivatives for each direction. In this case, gradient descent fails to determine the nature of a point using the hessian matrix, implying a slow learning.

3.2.2 Stochastic Gradient Descent

The stochastic gradient descent (SGD) is an evolution of the gradient descent method for taking into account the misleading with multiple dimension data. It is used to obtain an unbiased value for the gradient by taking the average gradient for a subset of samples $\mathbb{B} = \{x, ..., x^m\}$ called minibatch uniformly from the training set on each step of the algorithm. The minibatch size m is chosen to be a relatively small number of examples between 1 and a few hundred, depending on the size of the training dataset. This hyperparameter is one of the most important in a NN model, and can be tuned to search for an improve in the model.

The algorithm takes the initial parameter θ from the model, and compute the gradient estimate

$$g \leftarrow \frac{1}{m}\nabla_\theta \sum_{i=1}^{m} L(f(x^{(i)}; \theta), y^{(i)}) \tag{3.5}$$

where ∇_θ is the gradient of the parameter θ, L is the per-example loss, $x^{(i)}$ are the m examples in a minibatch from the training dataset and $y^{(i)}$ are their corresponding targets. After that, it applies an update in the parameters

$$\theta \leftarrow \theta - \epsilon g \tag{3.6}$$

and then it starts a new epoch. On the other hand, ϵ is an hyperparameter called learning rate, which determines the step size at each iteration while moving toward a minimum of the loss function. The negative side of SGD is that, in general, it is a slow algorithm since it may not be guaranteed to arrive at a local minimum in a reasonable amount of time. Because of that, several faster algorithms called optimizers have been developed the last years, which are extensively described in Goodfellow et al. (2016) and the reader is encouraged to consult it for more information.

3.2.3 Convolutional Neural Networks

Convolutional Neural Networks (CNN) (Cun et al., 1990) are NNs generally used to extract behaviors from image data, which can be thought as a 2D grid of pixels, but it is usual to work

with this type of NNs to process data with tensor-like topology, for example, a coloured image have which have more than one channel (RGB). In this case, each tensor consist in a 2D pixel image for each channel. The difference of these NNs with respect to the ones seen before is that CNNs use convolution in place of general matrix multiplication in at least one of their layers.

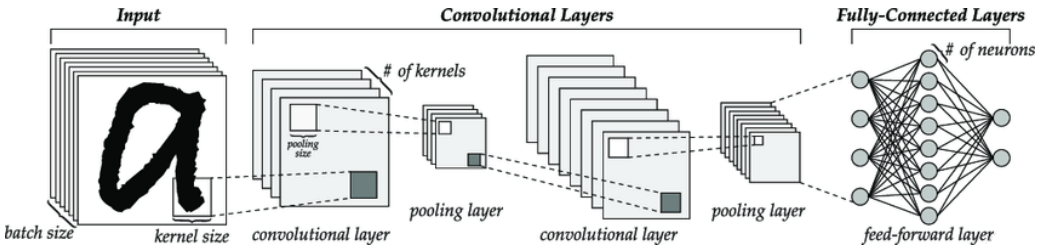

Figure 3.5: General architecture of a CNN model (Goodfellow et al., 2016).

Figure 3.5 shows the general architecture of a CNN model. The layers are known as convolutional blocks, and they consist of three stages: in the first one, the layer uses parallel convolution operations in order to produce a set of linear activations. In the second stage, a nonlinear activation function changes the values from the previous step. In the third step, the output of the layer changes when using a pooling function, which aggregates information by grouping neighboring pixels using either their maximum or average values. A FNN can be used after the output block, as in the figure 3.5, to categorize the outputs of the net but this is not a general case. Since the networks used along this PhD thesis do not have pooling stages, both convolution and detector ones will be described in more detail [1].

3.2.4 Convolution Stage

In the convolution stage, the NN applies a convolution operation on each pixel. It is defined as

$$s(t) = \int x(a)\, w(t-a)\, da = (x * w)(t) \tag{3.7}$$

where x is the input, and w is the kernel. The output s is called the feature map. The input is usually a multidimensional array of data or tensor, and the kernel is usually a tensor of parameters related to the learning algorithm. If the network use convolutions over more than one axis at a time, i.e a two-dimensional kernel K, the feature map is defined by a discrete convolution

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m, j+n) K(m,n) \tag{3.8}$$

[1] The reader is encouraged to consult Goodfellow et al. (2016) for a concise review of the pooling stage.

This type of function, more useful to implement in a ML model thanks to the less variation on m and n, is called the cross-correlation function. Figure 3.6 shows an example of convolution applied to a $2-D$ tensor.

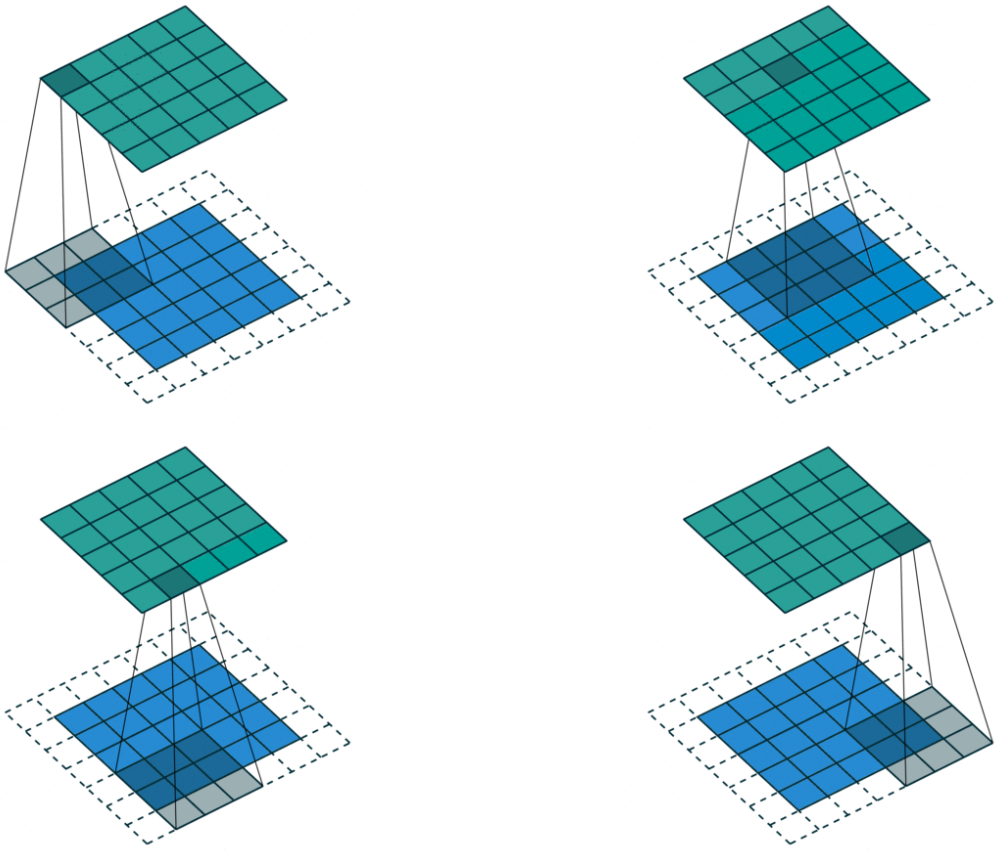

Figure 3.6: Convolution operation on a $2-D$ image in the convolution stage of a CNN. After a convolution in the green grid, a feature map in the blue grid with high dimensionality is created on each step (Goodfellow et al., 2016).

3.2.5 Detector Stage

In the detector stage, the main objective is to take the set of linear activations produced in the convolution stage, and run it through a nonlinear activation function. Generally, the activation functions $g(z)$ have a left derivative defined by the values of the function to the left of z, and a right derivative defined by the values to the right of z, and they are not differentiable at $z = 0$ In general, this stage has a transformation such as

$$z = W^T x + b \tag{3.9}$$

where W is a matrix of kernels, and b are the bias. On the other hand, x are the inputs. After that, the NN applies the activation function $g(z)$. The most popular activation functions are the sigmoid, the hyperbolic tangent, the linear units and the rectified linear units (ReLU) functions.

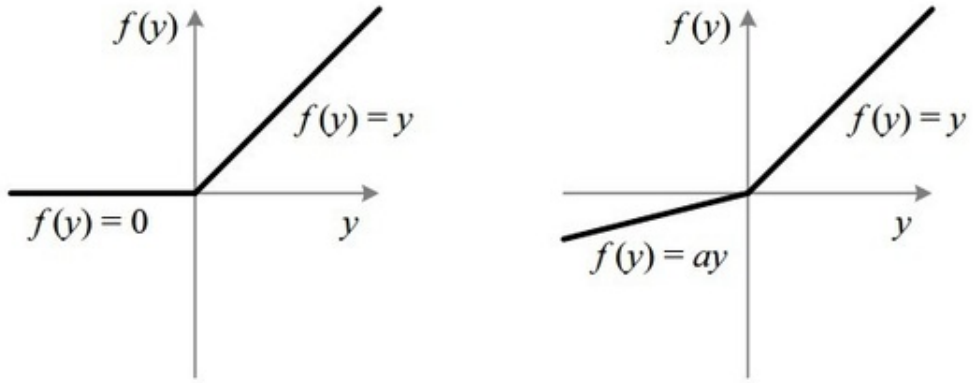

Figure 3.7: Illustration of the activation function $f(y)$ ReLU (left) and its generalization leaky ReLU (right), which depends on a parameter a (Goodfellow et al., 2016).

ReLU is popular because it is relatively easy to optimize. As it is shown in Figure 3.7 (left panel), it outputs a non-zero value at half of its domain like the linear units and a zero value at the other half of its domain. This implies that the derivatives through ReLU are high when the unit is active. On the other hand, the second derivative is 0 in all its domain. This means that the gradient direction is more useful for learning than using activation functions with second-order effects. ReLU follows the transformation

$$h = g(W^T x + b) \tag{3.10}$$

where the parameters are the same as in equation (3.9).

Another possibility is to use a generalization of ReLU called the Leaky Rectified Linear Units (leaky ReLU), which is the one used in the neural networks of this PhD thesis. This function uses a non-zero slope α_i when $z_i < 0$, that is

$$h_i = g(z, \alpha)_i = max(0, z_i) + \alpha_i min(0, z_i) \tag{3.11}$$

as is shown at right side in figure 3.7. After that, it fixes α_i to a small value.

Tunning the detector stage through changing the activation function is another procedure to search for an improvement in a NN model. This tunning consists generally on a genetic grid

seach, which tests model for evenly spaced out values of hyperparametrs and determines which combination gives best results on validation set. Because of that, understand the detector stage is an extremely active area of research in the world of DL.

3.2.6 Fully Convolutional Neural Networks

Object segmentation is another recently and important goal in the field of DL (Long et al., 2014). In this case, the DL model looks for classify each pixel instead of the whole image. If the image is accompanied by a label, a well-trained model theoretically is able to classify this label as one class. There are two types of image segmentation: Instance Segmentation and Semantic Segmentation. The last one is the type of image segmentation which looks to classify the objects belonging to the same class in the image with a single label. An example of Semantic Segmentation is shown in figure 3.8.

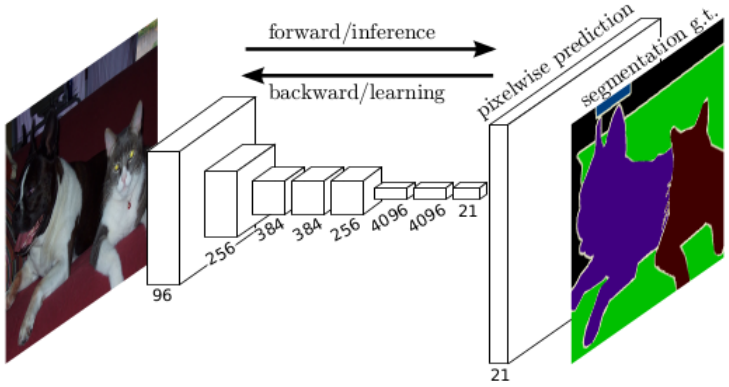

Figure 3.8: Illustration of Semantic Segmentation studied with a NN (Long et al., 2014).

Fully-Convolutional (FCNN) are NNs trained to perform image segmentation. They make both learning and inference on the whole image at-a-time by feedforward computation and backpropagation. In their upsampling steps they make a prediction at each pixel and in their subsampling steps they make a learning at each pixel.

Each layer is a three-dimensional array of size $h \times w \times d$, where h and w are height and width respectively, and d is the feature or channel dimension.

Each FCNN block have three steps (convolution stage, detector stage and pooling stage) which depend only on relative spatial coordinates. The activation functions of each layer compute outputs through

$$y_{ij} = f_{ks}(\{x_{si+\delta i, sj+\delta j}\}_{0 \leq \delta i, \delta j \leq k}) \tag{3.12}$$

where k is the kernel size, s is a parameter called the stride or subsampling factor, (i, j) is the location and f_{ks} determines the layer type (convolution, detector or pooling stage). Therefore, while a CNN computed a nonlinear function, a FCNN computes a nonlinear filter.

A FCNN have two general steps: first, it convolves the image as a CNN with a factor f. The second step is the upsampling of the image. Upsampling with factor f consist on take a fractional input stride of $1/f$. Since f is integer, upsampling can be seen as backwards convolution or deconvolution with an output stride of f.

Their learning process is also based on backpropagation and optimization. In this case, backpropagation algorithm updates both filters and kernels. Since they extract a signal from others contaminats in an image, they are specially designed for the tasks of this PhD thesis: both point source detection and CMB recovery are then image segmentation problems [2].

3.2.7 Learning multi-dimensional kernels

As shown in previous sections, the microwave sky has a particular spectral behavior. In this PhD thesis, the learning is supervised. Therefore, the similar the input image is to the label, the better the network is performing. However, in some occasions, the input patch could have complex and different structures with respect to the label, for example, thermal dust cirrus or strong radio sources. For this purpose, multi-frequency information could help the neural network to learn. In particular, the networks developed in this PhD thesis read multi-frequency data, then, their learning process vary with respect the one described in Goodfellow et al. (2016). In this case, the network have to learn multi-dimensional kernels. This process is extensively described in Goodfellow (2010), but next pages will describe it briefly.

Firstly, let's define new notation for simplicity: $@_d$ and $@_d^T$ represent the convolution and deconvolution between both visible and hidden channels, respectively. On the other hand, $\#_d$ is the convolution between both the gradient and the visible channel. Thus, the convolutional blocks make a convolution operation

$$H_{cij} = \sum_{k,m} W_{kmi} V_{c \circ d + k, m, j} = W @_d V, \tag{3.13}$$

while the deconvolutional blocks, connected to them, make a transpose convolution operation

$$R_{qmj} = \sum_{c,k \,|\, d \circ c + k = q} \sum_{i} W_{kmi} H_{c,i,j} = W @_d{}^T H, \tag{3.14}$$

[2] The mathematical updating of uni-dimensional FCN kernels are extensively explained in Goodfellow et al. (2016). The reader is encouraged to consult it for more information.

where d is a vector of strides, c is a pixel position, $\circ$ is the elementwise product, i is the visible channel and j is the hidden channel, that is, the patches of the sky read on each block and the resulting feature maps. Then, the information flows forward the network. After the last block, the mean squared error function is computed outputing a loss value. Then, the basic gradient computed at the end of the architecture is

$$\frac{\partial L(H = W@_d V)}{\partial W_{c,i,j}} = \sum_{k,m} \frac{\partial L}{\partial H_{kjm}} V_{d\circ k+c,i,m} = \nabla_W L(H = W@_d V) = (\nabla_H L)\#_d V. \tag{3.15}$$

However, there are three more gradients to compute for updating the parameters of the network: the gradient of the loss with respect to the weights W of the final block, the gradient of the loss with respect to the outputs H of the final block and the gradient of the outputs with respect to the inputs V. The first one is used for updating the weights of the final block, the second one for updating the outputs of the final block and the third one for propagating the gradient backwards.

- **Gradient with respect to the weights**

$$\begin{aligned}
R_{q,m,j} &= \sum_{c,k\,|\,d\circ+k=q} \sum_i W_{kmi} H_{cij} \Rightarrow \frac{\partial L}{\partial W_{xyz}} = \sum_{qmj} \frac{\partial L}{\partial R_{qmj}} \frac{\partial R_{qmj}}{\partial W_{xyz}} = \\
&= \sum_{qmj} \frac{\partial L}{\partial R_{qmj}} \frac{\partial \sum_{c,k\,|\,d\circ+k=q} \sum_i W_{kmi} H_{cij}}{\partial W_{xyz}} = \\
&= \sum_{qj} \frac{\partial L}{\partial R_{qyj}} \frac{\partial \sum_{c,k\,|\,d\circ+k=q} W_{xyz} H_{czj}}{\partial W_{xyz}} = \\
&= \sum_{qj} \frac{\partial L}{\partial R_{qyj}} \sum_{c,k\,|\,d\circ+k=q} H_{czj} = \sum_{cj} \frac{\partial L}{\partial R_{d\circ c+x,y,j}} H_{czj} \Rightarrow \frac{\partial L}{\partial W_{cij}} = \\
&= \sum_{km} H_{kjm} \frac{\partial L}{\partial R_{d\circ k+c,i,m}} \Rightarrow \frac{\partial L}{\partial W_{cij}} = \sum_{km} \frac{\partial L}{\partial H_{kjm}} V_{d\circ c,i,m} \Rightarrow \\
&\Rightarrow \nabla_W L(R + W@_d{}^T H) = H\#_d \nabla_R L
\end{aligned} \tag{3.16}$$

- **Gradient with respect to the outputs**

$$
\begin{aligned}
R_{q,m,j} = \sum_{c,k|d\circ+k=q} \sum_{i} W_{kmi} H_{cij} \Rightarrow \frac{\partial L}{\partial H_{xyz}} = \sum_{qmj} \frac{\partial L}{\partial R_{qmj}} \frac{\partial R_{qmj}}{\partial H_{xyz}} = \\
= \sum_{qmj} \frac{\partial L}{\partial R_{qmj}} \frac{\partial \sum_{c,k|d\circ+k=q} \sum_i W_{kmi} H_{cij}}{\partial H_{xyz}} \\
= \sum_{qm} \frac{\partial L}{\partial R_{qmz}} \frac{\partial \sum_{k|d\circ x+k=q} W_{kmy} H_{xyz}}{\partial H_{xyz}} = \\
= \sum_{qm} \frac{\partial L}{\partial R_{qmz}} \sum_{k|d\circ x+k=q} W_{kmy} = \sum_{km} \frac{\partial L}{\partial R_{d\circ x+k,m,z}} W_{kmy} \Rightarrow \frac{\partial L}{\partial H_{cij}} = \\
= \sum_{km} W_{kmi} \frac{\partial L}{\partial R_{d\circ c+k,m,j}} \Rightarrow H_{cij} = \sum_{km} W_{kmi} V_{d\circ c+k,m,j} \Rightarrow \\
\Rightarrow \nabla_H L(R + W@_d{}^T H) = W@_d \nabla_R L
\end{aligned}
\tag{3.17}
$$

- **Propagating the gradient backwards**

$$
\begin{aligned}
\frac{\partial L(H = W@_d V)}{\partial V_{xyz}} = \sum_{cij} \frac{\partial L}{\partial H_{cij}} \frac{\partial H_{cij}}{\partial V_{xyz}} = \\
= \sum_{cij} \frac{\partial L}{\partial H_{cij}} \frac{\partial \sum_{km} W_{kmi} V_{d\circ c+k,m,j}}{\partial V_{xyz}} = \\
= \sum_{ci} \frac{\partial L}{\partial H_{ciz}} \frac{\partial \sum_{k|d\circ c+k=x} W_{kyi} V_{xyz}}{\partial V_{xyz}} = \\
= \sum_{ci} \frac{\partial L}{\partial H_{ciz}} \sum_{k|d\circ c+k=x} W_{kyi} = \\
= \sum_{k|d\circ c+k=x} \sum_{i} \frac{\partial L}{\partial H_{ciz}} W_{kyi} \Rightarrow \frac{\partial L(H = W@_d V)}{\partial V_{qmj}} = \\
= \sum_{c,k|d\circ c+k=q} \sum i W_{kmi} \frac{\partial L}{\partial H_{cij}} \Rightarrow \nabla_V L(H = W@_d V) = \\
= W@_d{}^T \nabla_H L
\end{aligned}
\tag{3.18}
$$

Once each gradient is computed, the parameters become updated based on an optimizer. Along this PhD thesis, the optimizer used is AdaGrad. However, deep learning research shaded light to many more, which are extensively described in Goodfellow et al. (2016). Then, each parameter θ is updated by following

$$
\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\epsilon I + diag(G_t)}} g_t, \tag{3.19}
$$

where g_t is any of the above gradients at the time-step t, η is the initial learning rate, ϵ is a small quantity to avoid the division by zero, I is the identity matrix and $G_t = \sum_{\tau=1}^{t} g_\tau g_\tau^T$ is the sum of the outer product of the gradients until time-step t.

CHAPTER 4

Detecting point sources

4.1 Introduction

As explained before, radio and dusty galaxies are a major contaminant for recovering the CMB at small scales ($l > 1000$) in all the range of frequencies where the signal is present in the sky. Moreover, the first ones pollute the signal mostly bellow 220 GHz, as shown in Figure 2.10, while the second ones are more relevant at higher frequencies. This is because the emission from radio galaxies come mainly from the high-energy relativistic jets from their AGN, being generally non-thermal sources, more relevant at lower frequencies. On the other hand, the grains from star-forming regions of the dusty galaxies behave similarly to the dust emission from our Galaxy, which is more relevant in the FIR spectrum.

Planck detected several thousands of PS in temperature and several hundreds in polarization, as published in both Planck Catalogues of Compact Sources (PCCS Planck Collaboration (2014a) and PCCS2 Planck Collaboration (2016a)). However, the Mexican Hat Wavelet 2 (MHW2, González-Nuevo et al. (2006)), which was the method used for PS detection in temperature maps in *Planck*, has several conditionants that could affect a reliable detection of PS: firstly, it assumes a gaussian distribution for the PS, which implies it can only properly characterize this kind of shape. Secondly, its convolution affects the total image, implying that the patch changes its properties once the method is used. Thirdly, it always produce artifacts near the borders that should be taken into account in a post-processing step in order to avoid false detections. Because of that, for future experiments, new models should be developed. Neural networks seem to be good candidates for detecting PS in background noisy maps, as concluded by Bonavera et al. (2021), who developed a FCNN called the Point Source Image Detection Network (PoSeIDoN) to detect PS in single-frequency temperature patches of the sky by performing image segmentation.

Furthermore, PoSeIDoN was trained with realistic simulations of the 217 GHz microwave sky as seen by *Planck*, and tested against realistic simulations not only at the channel used for

training, but also at 143 and 353 GHz channels. Its behavior was also compared against the performance of the MHW2 on the same patches. The network successfully recovered PS at 90% completeness corresponding to 253, 126 and 250 mJy for 143, 217 and 353 GHz respectively. On the other hand, with a 3σ flux density detection limit, the MHW2 recovered PS up to 181, 102 and 153 mJy. However, while PoSeIDoN reached a maximum of 50% of spurious detections (false positives), the MHW2 recovered sources with much more than 50% at all flux density intervals. After applying a safer 4σ flux density detection limit to reduce the number of spurious detections, it recovered PS up to 235, 137 and 192 mJy, but still the number of spurious detections doubled the one from PoSeIDoN.

These methods are designed to detect PS in single-frequency conditions, but each PS has its own spectral behavior and single-frequency detection imply a huge wasting of physical information. Because of that, the use of multi-frequency information could allow to reliable detect more sources with respect to single-frequency methods.

As shown in Planck Collaboration (2018), *Planck* detected thousands of multi-frequency sources with the Matrix Filters approach (Herranz et al., 2009). This chapter describes the first published method of multi-frequency PS detection based on a neural network model. The methodology is an evolution of PoSeIDoN. Instead of learning how to segmentate PS in background noisy patches of the sky in one frequency channel, it uses three channels to take into account the spectral behavior of the emissions presented in the sky for recovering the PS contribution. It is called the Multi-frequency Point Source Image Detection Network (MultiPoSeIDoN), and its performance was published in the *Astronomy and Astrophysics* journal Casas et al. (2022b), in a work led by the author of this PhD thesis.

Moreover, MultiPoSeIDoN is trained with 50000 realistic simulations of the microwave sky as seen by *Planck*. Each simulation has three 128×128 pixel patches of the 143, 217 and 353 GHz channels, with a pixel size of 90 arcseconds. The patches are cut by using the gnomonic projection of the HEALPix framework (Górski et al., 2005) outside the Galactic mask ($|b| > 30°$), in the region above/bellow the black horizontal lines of Figure 2.11.

The simulations are formed by thermal dust, the CMB, the CIB, thermal SZ efect, radio and dusty galaxy emissions and instrumental noise. In particular:

- The **CMB** is the map published in *Planck* after using the SEVEM method Martínez-González et al. (2003).
- **Thermal dust** is a simulation from the PLA website, consisting in applying of the

GNILC method Remazeilles et al. (2011) to the PR2-2015 *Planck* data at 353 GHz. Their extrapolation to lower channels has been made by using the PR1-2013 *Planck* dust spectral indeces.

- The **CIB** is simulated by assuming that it is formed by proto-spheroidal galaxies in the process of forming their stellar masses. To simulate them, the source number counts given by Cai et al. (2013), their angular power spectrum given by Lapi et al. (2011) and the software CORRSKY are used.

- The **thermal SZ** effect is also a simulation from the PLA, based on the number counts given by the method of Delabrouille et al. (2002).

- **Radio and IR late-type PS** are simulated by taking into account the C2Ex model by Tucci, M. et al. (2011) for the number of counts of the first ones, and the source number counts by Cai et al. (2013) for the second ones. They are injected into the patches by using CORRSKY. Their spectral behavior is simulated by assuming that vary as

$$S = S_0 \left(\frac{\nu}{\nu_0}\right)^{\alpha}, \tag{4.1}$$

  were α is the spectral index for each population, and assumed to be a gaussian value over the data published in the PCCS2, and shown in Fig. 2.10.

- **Instrumental noise** is simulated by assuming a random white noise along the patch, with the sensitivity of *Planck*, given in Planck Collaboration (2020a): 0.55, 0.78 and 2.56 $\mu K_{CMB} deg$ for 143, 217 and 353 GHz, respectively.

Figure 4.1 shows one example of the simulations used in this chapter. In the first column it is shown the three patches *seen* by MultiPoSeIDoN in both train and test datasets (143, 217 and 353 GHz respectively). The colorbars show the scale of the temperature emission in Jy. As shown, the 353 GHz channel presents higher contamination from the dust of our Galaxy, as so as CIB pollution at small scales. It could also be seen the different resolution between the 143 GHz channel and the other two, due to the full width at half maximum (FWHM) of the Planck instrument. A brighter PS appears in the patch, which could also be seen in the second column patches, used as a label for the supervised learning of MultiPoSeIDoN. The spectral behavior of the PS is easier to see: there are sources visible at 353 GHz that start to be apparent at 217 GHz, but they are not in the 143 GHz patch. That is, following the spectral index distributions

of Figure 2.10, these sources are probably dusty ones, while the brighter PS is probably a radio galaxy, since it also appears at 143 GHz.

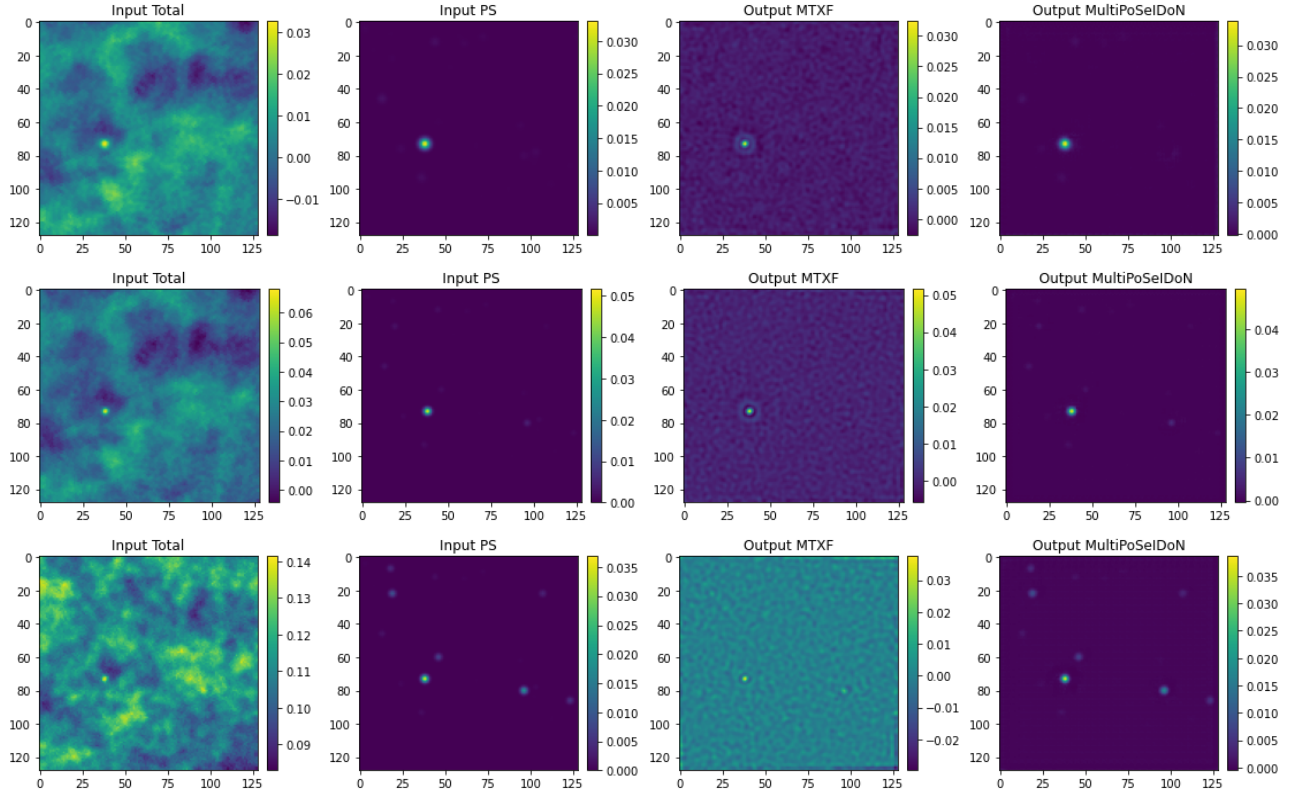

Figure 4.1: From left to right, an example of one simulation used for training and testing MultiPoSeIDoN and the MTXFs, with the Input total and PS-only input patches, and the MTXFs and MultiPoSeIDoN PS outputs at 143, 217, and 353 GHz from top to bottom at $b > 30°$ Galactic latitudes. The flux density values (in Jy) for each panel are shown in the colorbars.

Thus, the choose of these channels for studying the performance of MultiPoSeIDoN is not trivial: the main channel is the 217 GHz one, because the fraction of sources is higher than in the other two, as shown in Figure 2.10. On the other hand, 353 GHz channel helps MultiPoSeIDoN to separate the thermal dust contribution when comparing against the 143 GHz channel. Moreover, 217 GHz is the channel with the higher resolution of *Planck*, as shown in Planck Collaboration (2020a), and also the one when SZ effects are negligible, as explained in Sect. 2.2.4, which helps the network to detect small structures such PS.

As said before, the detection of PS is considered an image segmentation problem: MultiPoSeIDoN reads patches of the sky and separate the PS signal with respect the other ones. Then, it is a FCNN formed by several convolutional blocks connected to deconvolutional ones.

At the end of its architecture, a loss function takes the output and the PS label.

The network is formed by six convolutional blocks of 8, 2, 4, 2, 2, and 2 kernels of sizes of 9, 9, 7, 7, 5, and 3, respectively, which make inference over the input patches. They are connected to six deconvolutional blocks of 2, 2, 2, 4, 2, and 8 kernels of sizes of 3, 5, 7, 7, 9, and 9, which make prediction about the location and flux density of every PS in the patch. Every layer has a leaky ReLU activation function, which allows the information to pass. At the end of the architecture, the mean squared error loss function compares the output with the label, estimating a loss value. After that, the gradient is computed and the parameters of the network are updated through the backpropagation algorithm. This architecture is called Flat MultiPoSeIDoN because both label and output are a patch formed by PS at 217 GHz, with a flat behavior, that is, PS have the same spectral behavior in the three channels. An extension of this architecture to three outputs, and taking into account the spectral behavior of the PS for input and output, is also developed for comparison. Its name is Spectral MultiPoSeIDoN. In that case, the architecture is similar but having 9 kernels in the first convolutional block and another 9 in the last deconvolutional one instead of 8. Both flat and spectral architectures are shown in Figures 4.2 and 4.3, respectively.

4.2 Results

After being trained, both networks are tested with similar data, not used for training. In particular, they read a test set formed by 5000 more patches of the sky, then predicting 5000 patches with the PS. Since creating a catalogue from pixel positions in a patch is not an optimal process, the detection algorithm from Bonavera et al. (2021) is used, which is based on two parameters: the minimum flux density limit for considering a pixel value as a detection, and the number of pixels between detection, for taking into account a PS instead of an artifact produced by the network. This algorithm produces a series of both position and flux density for each detection, in order to compare them with the true catalogues, which are the PS in the label patches. After that, three statistical quantities are used for analyzing the detections. In particular:

- **Photometry**

  It is simply the comparison between the flux density of a PS in the input catalogue with respect the flux density of its corresponding estimate. It is usually represented as a scatter

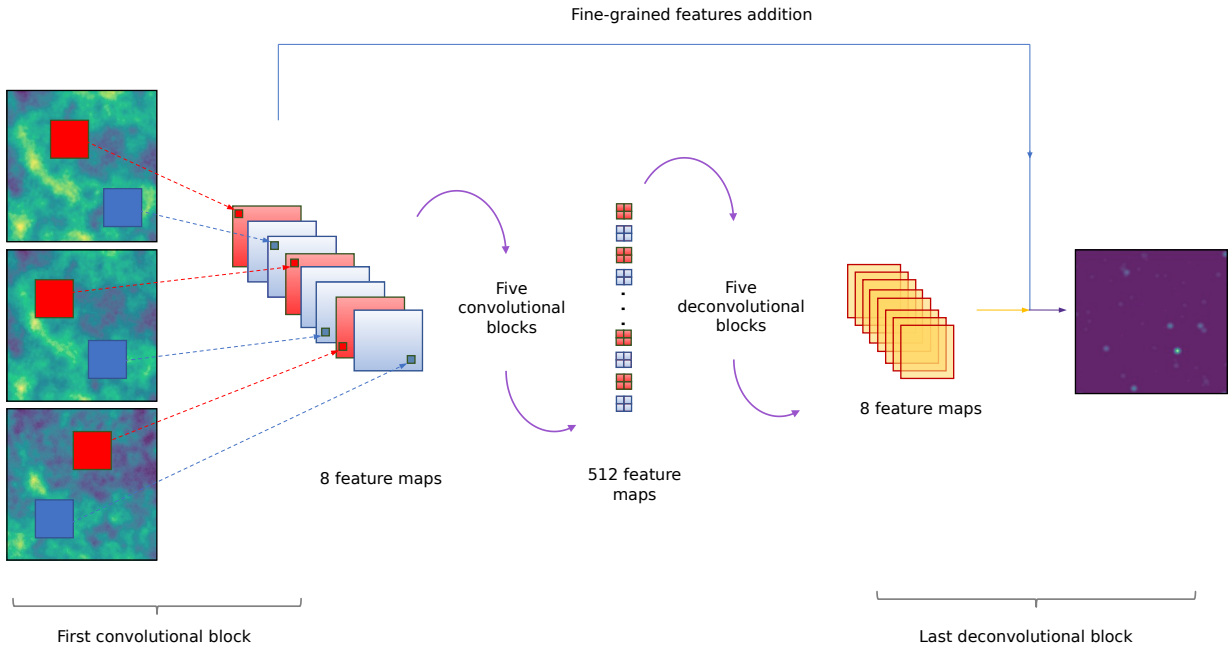

Figure 4.2: Architecture of Flat MultiPoSeIDoN. Firstly, it has a convolutional block, which produces eight feature maps. After that, the space dimensionality increases to 512 feature maps through five more convolutional blocks. These layers are connected to deconvolutional ones, which decreases the space dimensionality to eight feature maps in the last deconvolutional block. Fine-grained features are added from each convolution to its corresponding deconvolution.

plot with a line representing the ideal 1:1 case.

- **Completeness**

  It is the ratio between the number of true detections and the total number of input sources over a given flux limit, that is

  $$C(>S_0) = \frac{N_{true\,detected}(>S_0)}{N_{input}(>S_0)} \tag{4.2}$$

  It is usually represented as a cumulative plot with respect to the flux density limit.

- **Reliability**

  It is the number of detected sources that do not have a counterpart in the input catalogue (called spurious sources or false positives), that is

  $$R(S_0) = \frac{N_{spurious}(>S_0)}{N_{input}(>S_0)} = \frac{N_{predicted}(>S_0) - N_{true\,detected}(>S_0)}{N_{input}(>S_0)} \tag{4.3}$$

  It is usually represented as a percentage plot with respect to the flux density limit, and in combination with the completeness plot.

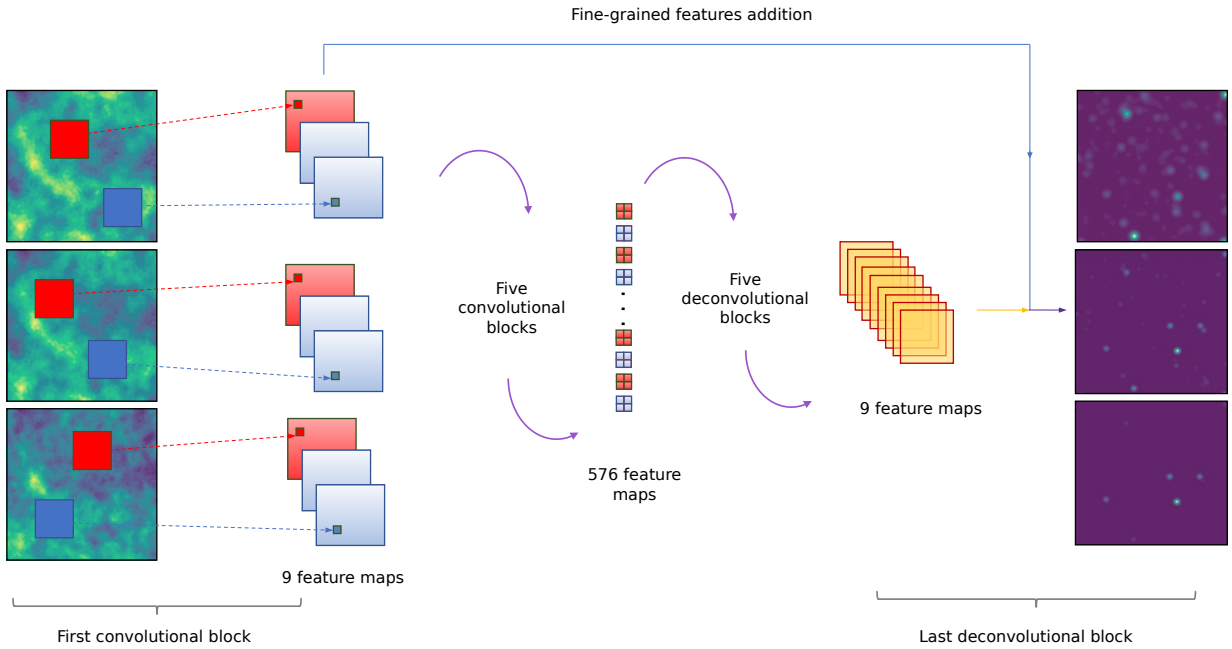

Figure 4.3: Architecture of spectral MultiPoSeIDoN. Firstly, it has a convolutional block, which produces nine feature maps. After that, the space dimensionality increases to 576 feature maps through five more convolutional blocks. These layers are connected to deconvolutional ones, which decreases the space dimensionality to nine feature maps in the last deconvolutional block. Fine-grained features are added from each convolution to its corresponding deconvolution.

4.2.1 Photometry

Figure 4.4 shows the comparison between input and recovered flux density in mJy for each detected source in the test dataset at 143, 217 and 353 GHz channels. Blue, red, cyan and orange circles represent the estimates from Flat MultiPoSeIDoN, Spectral MultiPoSeIDoN, the MTXF at and the MTFX after applying a safer 4σ flux density detection limit, respectively.

In general, both methodologies present a similar behavior along the three channels. Neural networks recover the flux density without presenting the Eddington Bias below the 100 mJy Planck limit of reliability. At these flux density low ranges, the uncertainty is at least lower than 10 mJy. Flat approach is slightly more accurate than the Spectral one, especially at 217 GHz, where Spectral MultiPoSeIDoN estimates have an uncertainty of 10 mJy up to 10^3 mJy flux density range.

On the other hand, the MTXF also shows accurate estimates of the flux density of the detected sources for the three channels. At the same flux density limit bellow the 100 mJy

Planck's one, there is still not Eddington Bias at 143 and 217 GHz but the uncertainty when recovering is about 30 mJy, slightly worse than both MultiPoSeIDoN's. At 353 GHz, due to dust and CIB contamination, the MTXF start to present an Eddington Bias at 120 mJy, while MultiPoSeIDoN are still reliable in their estimates at those flux density ranges.

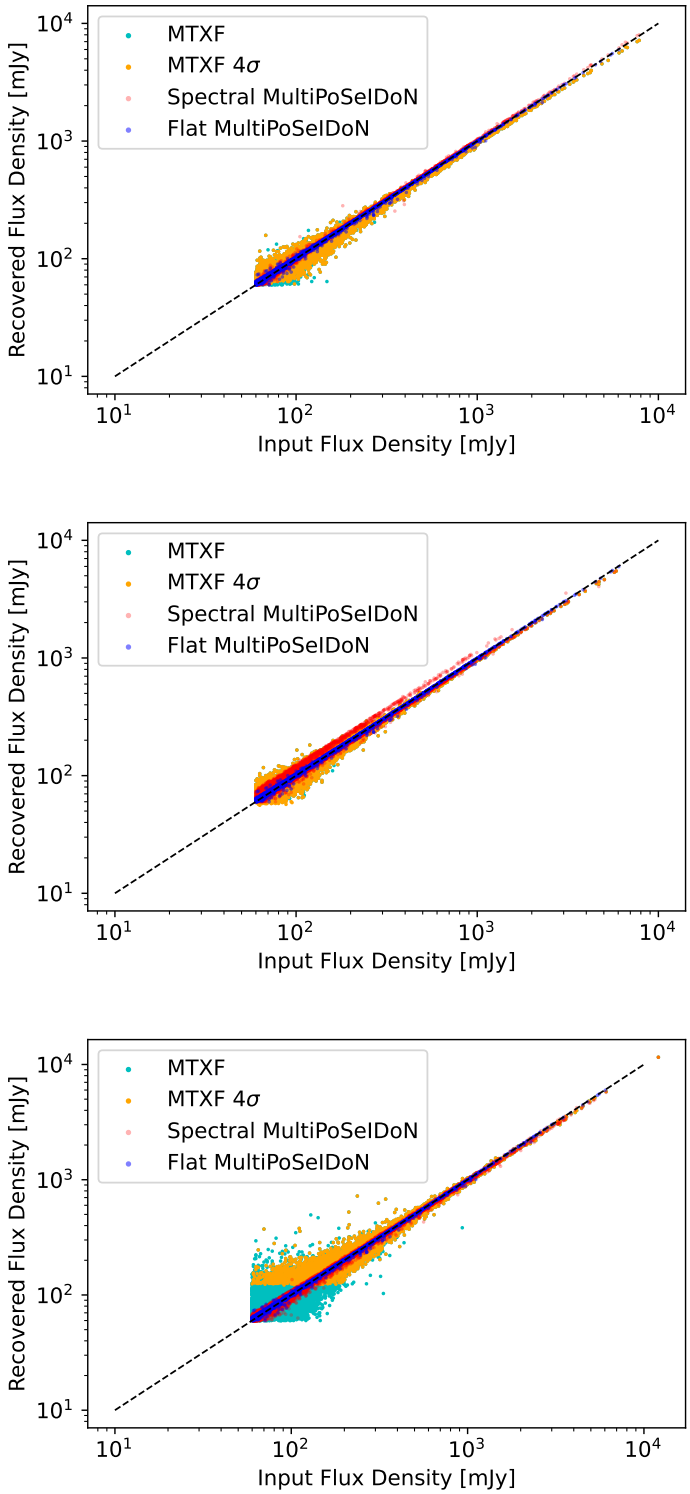

Figure 4.4: Flux density comparison between the input sources and the recovered ones by flat (blue dots) and spectral (red dots) MultiPoSeIDoNs, the MTXF (cyan dots), and the MTXF with a 4σ flux detection limit (orange dots) at 143, 217 and 353 GHz from top to bottom. Black dashed lines represent the ideal 1:1 case.

4.2.2 Completeness and reliability

Figure 4.5 shows the comparison between input and recovered completeness and reliability for each detected source in the test dataset at 143, 217 and 353 GHz channels. Blue, red, cyan and orange lines represent the estimates from Flat MultiPoSeIDoN, Spectral MultiPoSeIDoN, the MTXF and the MTXF after applying a safer 4σ flux density detection limit, respectively.

Completeness performance is similar for both methodologies at 143 and 217 GHz: they reach the 90% of completeness before Planck levels (represented as a grey dotted line). In general, MultiPoSeIDoN has a better performance than the MTXF, being the flat case more accurate than the spectral one. The respective percentage of spurious sources shows similar results than in Bonavera et al. (2021), when compared PoSeIDoN with the MHW2: the percentage of spurious sources from the neural network is clearly lower than from the filters at similar completeness levels, especially for fainter sources, at low flux density ranges. While for the MTXF, the percentage of spurious sources slowly decreases when increasing the frequency channel, it remains relatively stable for MultiPoSeIDoN at all flux density intervals.

In fact, at 143 GHz, the percentage of spurious detections varies from 100% to above 0% for fainter sources between 60 and 200 mJy, while MultiPoSeIDoN spurious detections are bellow 30% levels. At 217 GHz, the MTXF spurous detections vary from 100% to 0% for fluxes between 60 and 1000 mJy, while MultiPoSeIDoN only presents a relatively high percentage of spurious detections for fainter sources bellow 100 mJy.

At 353 GHz, the performance is clearly better for MultiPoSeIDoN: while completeness and reliability are similar than the other frequencies, the MTXF show a worse performance due to dust contamination. Moreover, although reaching the 90% of completeness levels at the same flux density interval than *Planck*, the percentage of spurious detections is above 50% even for brighter sources between 1000 and 2000 mJy.

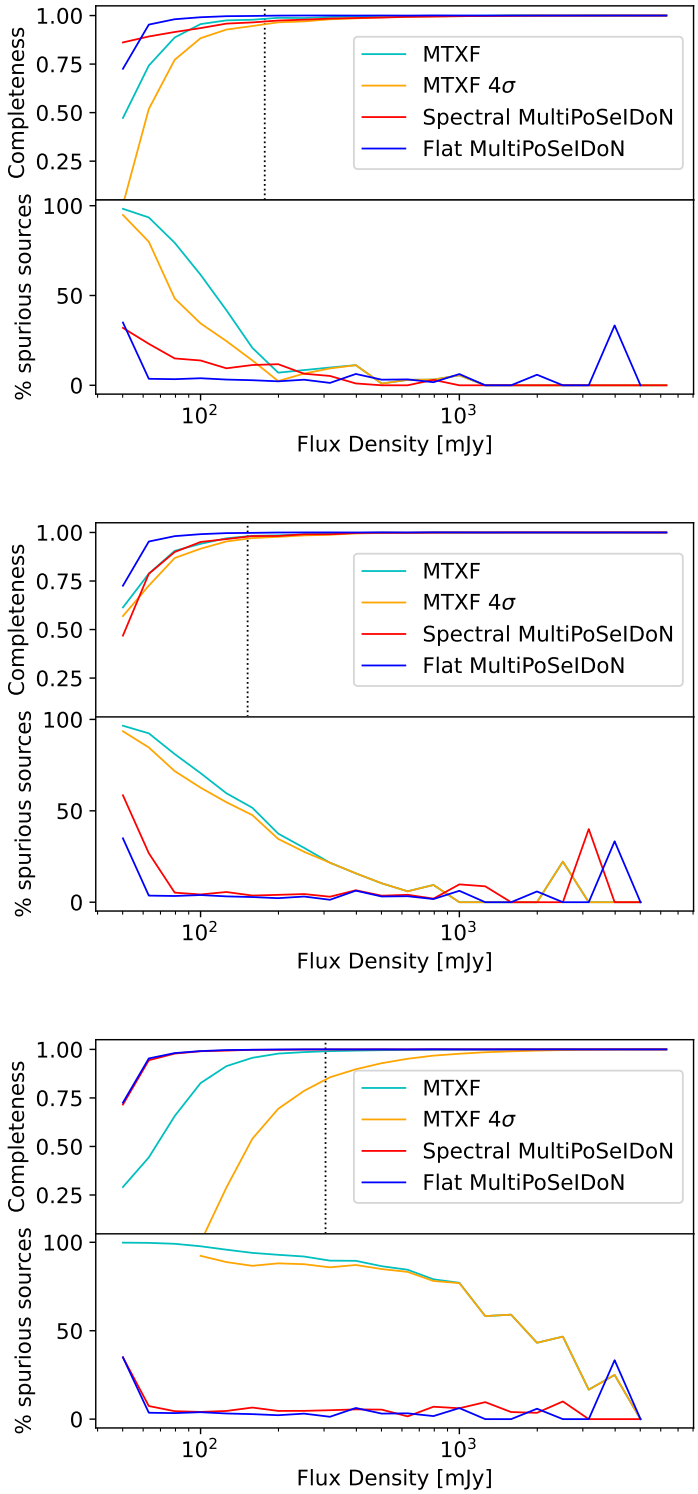

Figure 4.5: Completeness (top panels) and reliability (bottom panels) of flat (blue lines) and spectral (red lines) MultiPoSeIDoNs, the MTXF (cyan lines), and the MTXF with a 4σ flux detection limit (orange lines) at 143, 217 and 353 GHz from top to bottom. Grey dotted lines represent the 177, 152 and 304 mJy 90% of completeness level reached by *Planck* for each channel.

4.2.3 MultiPoSeIDoN vs PoSeIDoN

Finally, Figure 4.6 shows the performance comparison between PoSeIDoN (in green) and MultiPoSeIDoN (in blue and red for flat and spectral cases, respectively) at 217 GHz in photometry and completeness and reliability plots from top to bottom, respectively.

As shown, using multi-frequency information allows the neural network to have good performance: while MultiPoSeIDoN constraints accurately the flux density of the recovered sources, following the 1:1 case represented with the black dashed line, PoSeIDoN tends to underestimate the flux of the detected sources below 300 mJy. On the other hand, while MultiPoSeIDoN reaches the 90% of completeness level at lower flux density intervals than *Planck* nearly without spurious detections, PoSeIDoN reaches the 90% of completeness at higher flux density intervals, with more than 50% of spurious detections bellow 100 mJy.

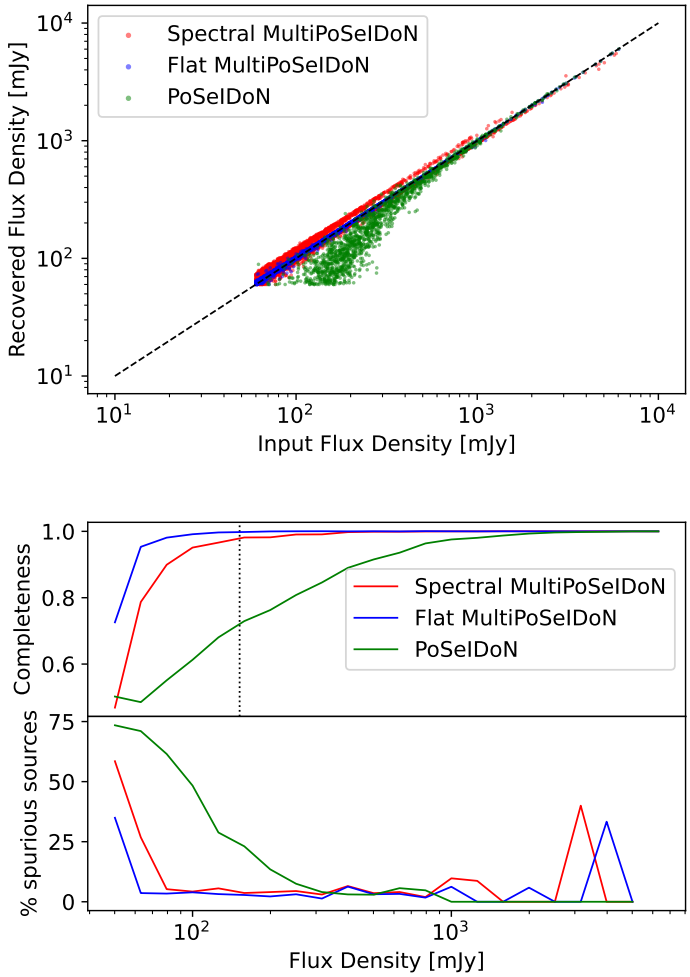

Figure 4.6: Comparison between PoSeIDoN and MultiPoSeIDoN performances. Top panel: Flux density recovering from PoSeIDoN (green dots) and from flat and spectral MultiPoSeIDoNs (blue and red dots, respectively). The black dashed line represents the ideal 1:1 case. Bottom panel: Completeness and reliability from PoSeIDoN (green lines) and from flat and spectral MultiPoSeIDoNs (blue and red lines, respectively). The grey dotted line represents the 90% of completeness level by *Planck*.

CHAPTER 5

Characterizing the polarization of point sources

5.1 Introduction

After the publishing the Second Planck Catalogue of Compact Sources, 234 PS were detected in polarization, which corresponds to only 2.715% of the sources detected in total intensity. Moreover, as shown in Section 2.3.1, relevant properties to understand the nature of the polarized PS could not be constrained with those poor statistics. In fact, post-*Planck* studies gave relatively better statistics than *Planck* but still these physical properties could not been accurately estimated.

Today, the method generally used for detecting PS in polarization is the Filtered Fusion (FF), proposed in Argueso et al. (2009). However, this method is reliable only to 300 mJy, as shown in Herranz et al. (2021), where an additional method based on a bayesian estimator is presented. However, also this method could not give accurate estimates of the polarization flux density bellow the 300 mJy limit. Moreover, the FF method is biased by the errors from Q and U maps, and its estimates must be processed, as shown in Planck Collaboration (2016a). Therefore, the question is that if only a few percent of the PS present polarization or if the methods used for their detection could not give reliable estimations for fainter sources.

Based on the accurate results presented in the last chapter, neural networks could be also reliable methods for detecting PS in polarization. For this task, there could be two possibilities: the first one is a *blind* detection as in the last chapter, where PS are detected in noisy background maps while extracting their flux density. The second one is a non-*blind* detection, where the PS polarization properties are estimated after knowing where they are located in temperature. Since *Planck* presents accurate estimations for the PS in temperature and it has a worse polarization sensitivity, it is easier to perform in a non-*blind* way. Thus, this Chapter presents a method based on a convolutional neural network for estimating both polarization flux density and angle of a point source centered in a small patch of the sky, a method published in the *Astronomy and Astrophysics* journal Casas et al. (2023a), in a work led by the author

of this PhD thesis.

The network is called the POint Source Polarisation Estimation Network (POSPEN), and it is simultaneously trained with Q and U *Planck* 10000 realistic simulations. Each one has a patch of the sky at 217 GHz of 32×32 pixels of area, with a pixel size of 90 arcseconds. The patches were cutted by using the gnomonic projection of the HEALPix framework outside the Galactic mask ($|b| > 30°$). Firstly, the microwave sky is simulated by taking into account all the elements emitting in polarization at this frequency (CMB, dust and synchrotron). Then, the central PS is injected into the patch. Finally, random white noise is added to the patch. In particular:

- The **CMB** is the map released by *Planck* using the SEVEM method.
- **Thermal dust** is a simulation from the PLA website, based on the model by Vansyngel et al. (2017) with spectral indeces from *Planck* PR1 release and large scale corrections with *Planck* PR2 dust release map.
- **Synchrotron** emission is also a simulation from the PLA, and it is based on a power law scale per pixel, being the spectral index derived from WMAP data.
- **Radio** PS are injected in the center of the polarized patches by assuming a log-normal distribution with $\mu = 0.7$ and $\sigma = 0.9$ parameters from Bonavera et al. (2017a) for the brighter source selected in a catalogue produced by using the source number counts from the C2Ex model by Tucci, M. et al. (2011) in temperature.
- **Instrumental noise** is added as random white noise at *Planck* levels (1.75 $\mu K_{CMB} deg$ at 217 GHz).

Figure 5.1 shows one example of the simulations used in this Chapter. In the top row, it is shown the patch with all emissions that POSPEN reads in $P = \sqrt{Q^2 + U^2}$ total polarization, Q and U from left to right. Bottom row shows the PS only map, used as a label for the training. In all cases, the units are Jy.

Furthermore, Figure 5.2 shows the architecture of POSPEN. After reading the Input total patch in its first convolutional block, five more are connected, with 8, 32, 64, 128, and 256 kernels, with sizes of 2, 2, 3, 5, and 5 and strides of 2, 2, 3, 5, and 5. After the convolutional blocks, there are two layers of 128 and 1 neurons, respectively, which convert the image information into numerical values. In all layers, the leaky ReLU activation function is used. At the end of

the architecture, the MSE loss function compares the output with the label for estimating the loss and the gradient.

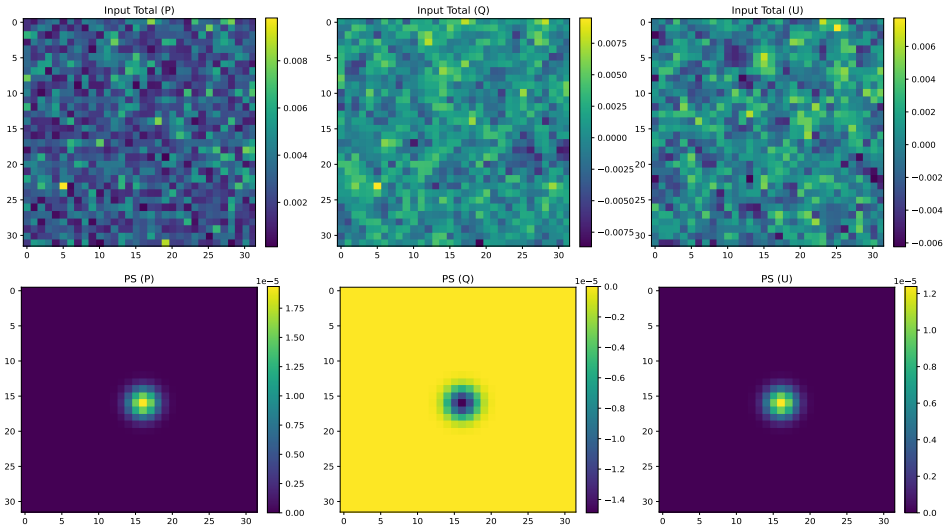

Figure 5.1: Example of one simulation used to train or validate POSPEN for P, Q, and U polarisation maps (left, middle, and right columns, respectively). Top panels show the input patch formed by all the emissions, where the network learns how to extract the polarization of the PS. Bottom figures show the simulated PS, where their integrated intensity are used as a label for the training. The units are in Jy.

5.2 Results

5.2.1 Constraining the polarization flux density

Once trained, POSPEN is tested with 1000 P more patches not used for training. It estimates a value for the total polarization flux density of the central PS on each patch. Its performance is then evaluated by comparing the estimates of the flux density for each PS against their known value.

This is shown in blue in Figure 5.3, as so as the relative error of the estimates against the input flux density levels, from top to bottom, respectively. Both dashed black lines represent the ideal case. Therefore, POSPEN reliably recovers the flux density of sources above 80 mJy, where Eddington bias starts to appear. In comparison with the bottom panel, for faint sources bellow 120 mJy, polarization is recovered with an error lower than $30 \pm 50\%$. Above that limit, although with poor statistics, POSPEN has fluctuating errors but, in general, they are bellow

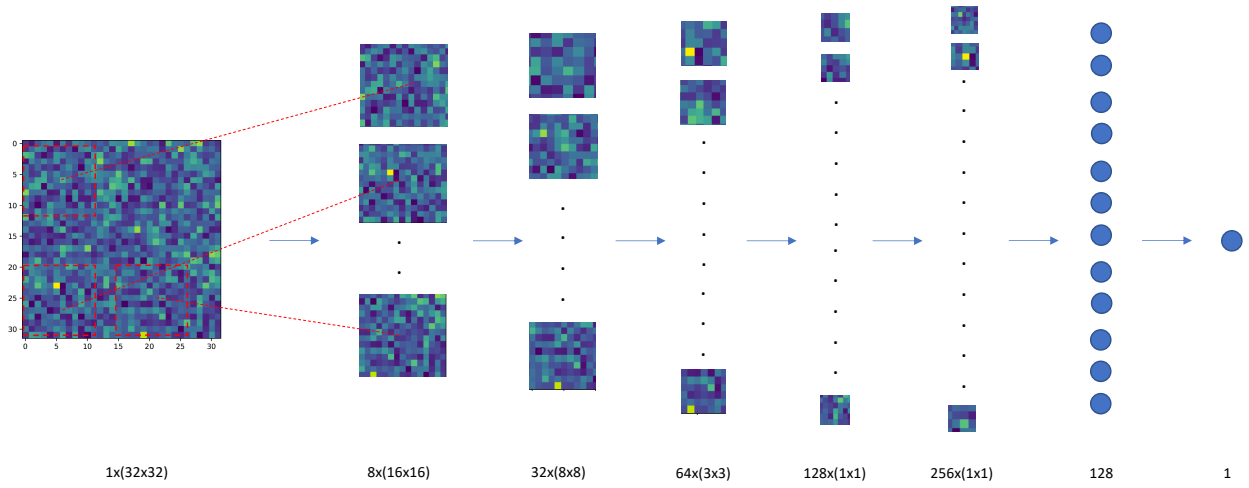

Figure 5.2: Architecture of POSPEN. The information in the 217 GHz patch is convolved into 8 filters in the first convolutional block. After that, several convolutions are made through four more layers of 32, 64, 128, and 256 filters. The information is then processed through a layer of 128 neurons. Finally, one last neuron gives the output value.

30%.

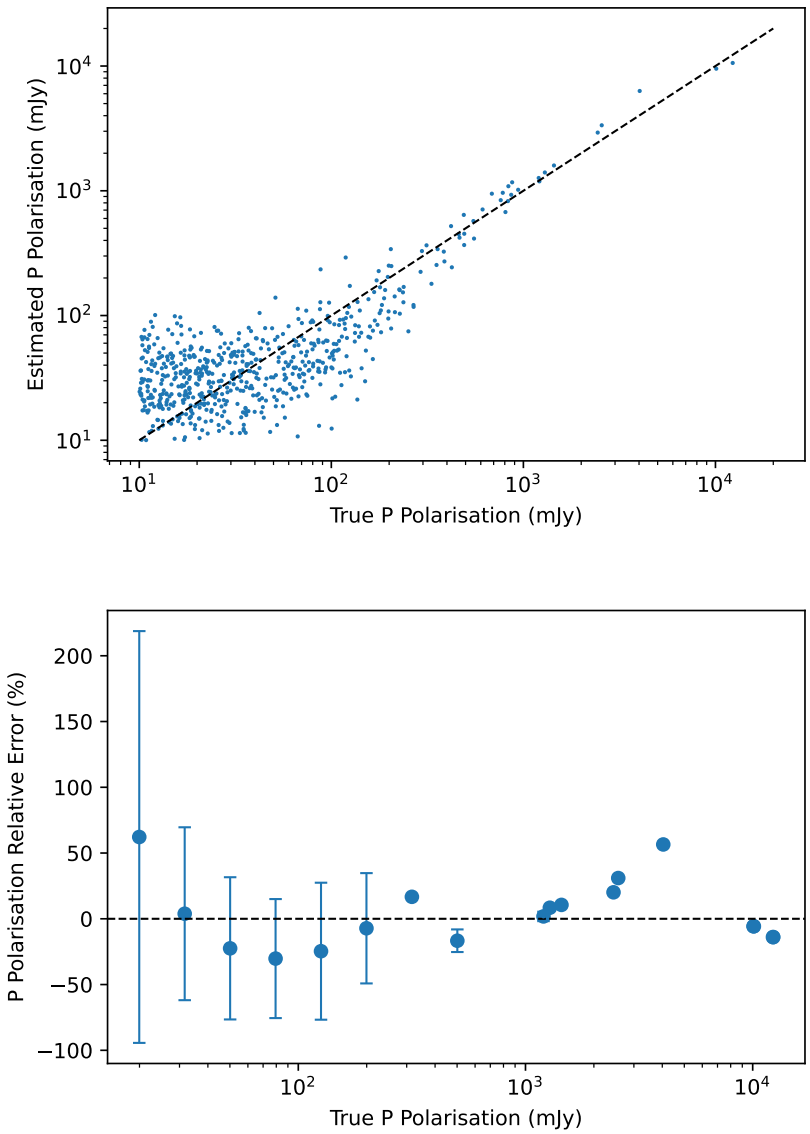

Figure 5.3: Estimation of total polarization flux density of the test catalogue. Top panel: correlation between true and estimated polarization flux density. Bottom panel: polarization relative error against the input polarization flux density. The error bars represent the uncertainty of each bin, which is considered to be equal to the standard deviation. In both cases, the dashed black lines show the ideal case.

5.2.2 Constraining the polarization angle

Although the total polarization P is the most natural quantity for characterizing the value of the polarized emission from an extragalactic source, due to the fact that it is more related than Q and U to the real processes occurring along the path of photons from the source to the Earth (Herranz et al., 2021). Thus, an estimation of both Q and U allows to estimate a polarization angle value

$$\psi = tan^{-1}\left(\frac{Q}{U}\right), \tag{5.1}$$

where $0 \leq \psi \leq 2\pi$ is the polarization angle. However, this expression implies having reliable detections in both polarization channels. Then, for flexibility, polarization angle are estimated by using

$$\begin{aligned} Q &= Pcos(\psi) \\ U &= Psin(\psi). \end{aligned} \tag{5.2}$$

Then, 10000 Q and U simulations each are used for training POSPEN simultaneously. After being trained, 1000 more simulations are used for testing. Firstly, Q and U photometry is analyzed in the same way as for the total polarization, in order to give reliable estimations for the polarization angle.

Figures 5.4 and 5.5 compare the estimates for the polarization flux densities with respect the input ones for Q and U sources, respectively, being both of them discriminated into $Q > 0$ sources (in orange), $Q < 0$ ones (in blue), $U > 0$ sources (in green) and finally $U < 0$ ones (in brown). Also the relative error against the input polarization is shown in the bottom panels of each figure, being the errorbars the standard deviation of each bin, representing the uncertainty of the model on each flux density range.

As shown, although separately analyzing both positive and negative Q and U sources, respectively, results show the same as for the total polarization: POSPEN is reliable above the 80 mJy limit. Below this value, the Eddington Bias appears in both cases. In this case, uncertainty increases due to the use of a different sign instead of the positive definite values of the total polarization.

Moreover, although Q and U could be accurately estimated and therefore the sign of the polarization angle, there are cases where one of them could not be correctly estimated. Because of that, the associated sing of the polarization angle could not be in the correct quadrant. To take into account this issue, Q and U values should be compared for every source: when $Q > 0$

and $U < 0$, the final angle is obtained by adding $\pi/2$ to the estimation, while when $Q < 0$ and $U > 0$, $\pi/2$ are substracted to the estimation. In order to give reliable estimations of the polarization angle, Figure 5.6 represents the root mean squared error of the polarization angle with respect the cut in polarization flux density for Q and U detections, that is, with a 250 mJy cut in both channels, POSPEN can estimate the polarization angle with an uncertainty of 30°.

The performance in the estimation of the polarization angle is shown in Figure 5.7 for Q and U sources in orange and green dots, respectively, being the coloured areas their 30° uncertainties. As shown, the estimation is generally accurate for bright sources above the 250 mJy cut in flux density since they have to follow the ideal 1:1 case, represented as a dashed black line in both panels.

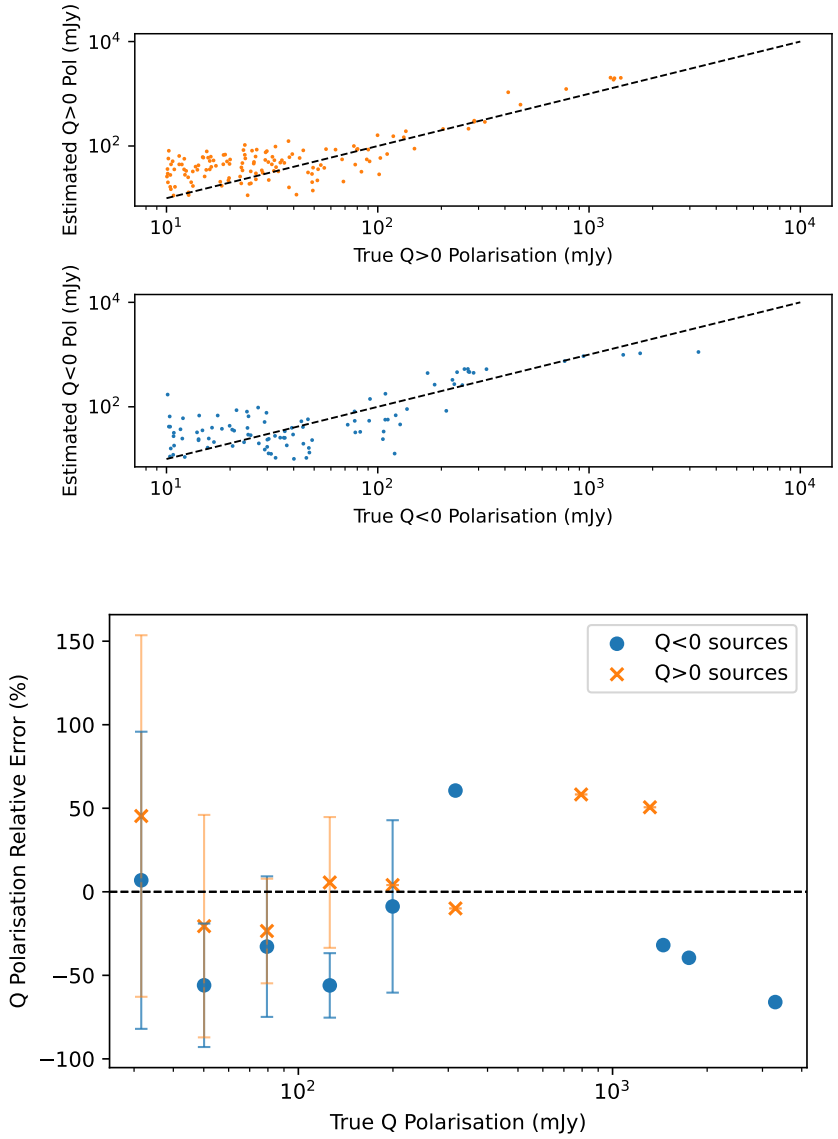

Figure 5.4: Estimation of Q polarization flux density of the positive and negative sources (in orange and blue, respectively) in the test dataset. Top panel: correlation between true and estimated polarization flux density. Bottom panel: polarization relative error against the input polarization flux density. The error bars represent the uncertainty of each bin, which is considered to be equal to the standard deviation. In both cases, the dashed black lines show the ideal 1:1 case.

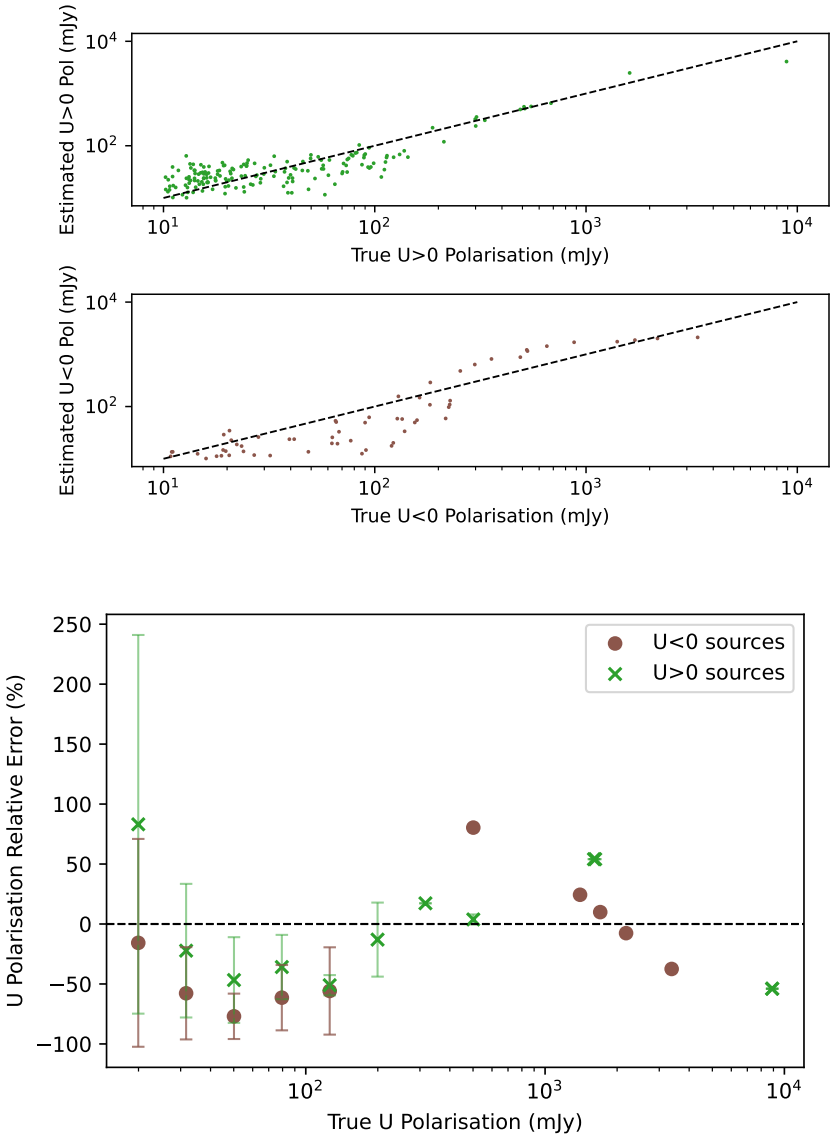

Figure 5.5: Estimation of U polarization flux density of the positive and negative sources (in green and brown, respectively) in the test dataset. Top panel: correlation between true and estimated polarization flux density. Right panel: polarization relative error against the input polarization flux density. The error bars represent the uncertainty of each bin, which is considered to be equal to the standard deviation. In both cases, the dashed black lines show the ideal 1:1 case.

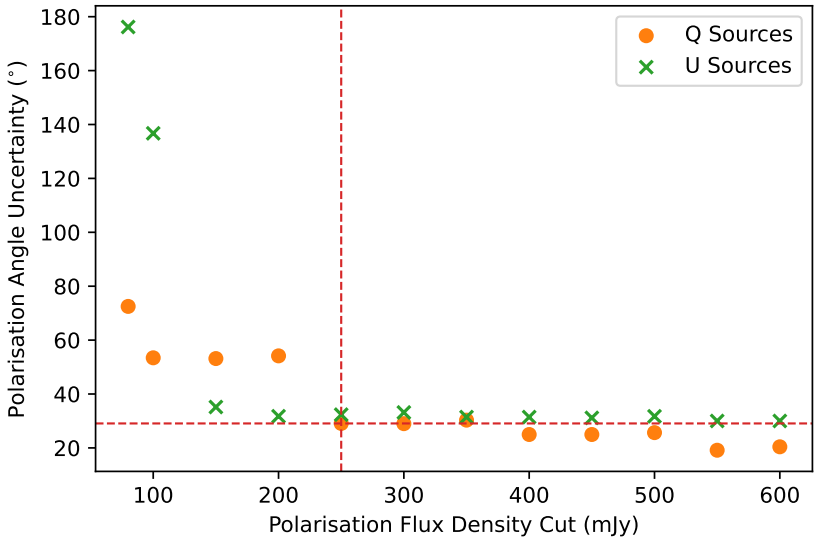

Figure 5.6: Relation between polarization angle uncertainty for the estimations of POSPEN and the cut in polarization flux density in both Q and U sources (in orange and green, respectively). The dashed red line shows the minimum uncertainty considering enough number of sources to have good statistics.

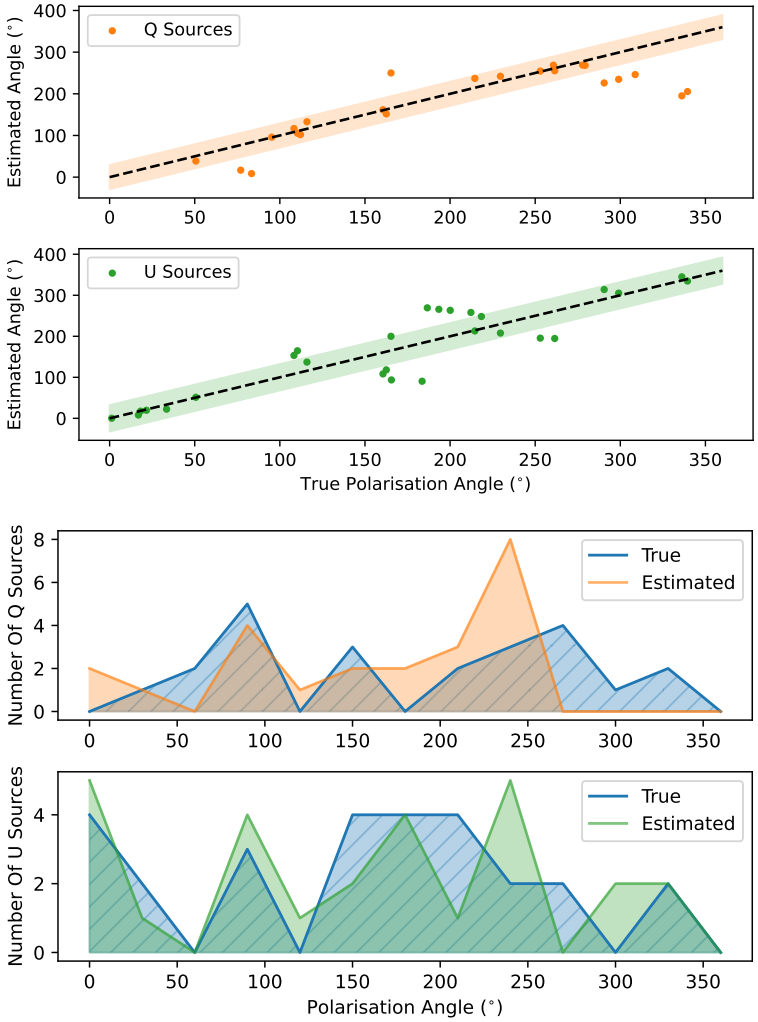

Figure 5.7: Estimation of the polarization angle. Top panel: Correlation between true and estimated polarization angles for Q and U sources (in orange and green, respectively). The coloured areas represent a confidence interval of $\pm 29°$ for Q sources and $\pm 32°$ U ones, which are the standard deviation of the relative error of each population, respectively. Dashed black lines represent the 1:1 case. Right panel: Distribution of polarization angles for both true and estimated Q and U angles (in orange and green, respectively).

CHAPTER 6

Recovering the cosmic microwave background in temperature

6.1 Introduction

As described in Leach et al. (2008), several pre-*Planck* methods were developed and optimized to separate the CMB from the other signals in the microwave sky. This process, called component separation, was one of the main objectives in the Collaboration, which finally was carried out by four methods, which published unprecedented detailed CMB maps in both temperature and polarization, as the one seen in Figure 1.5. They will be briefly described, but the reader is encouraged to see more information in Planck Collaboration (2020b), and the following references for more information about each method.

- **Commander** (Eriksen et al. (2008), Galloway et al. (2023)) is a Bayesian approach employing a Monte Carlo method called Gibbs sampling that fits the data as a parametric model with standard posterior sampling or maximization techniques, where each parameter represent an emission of the microwave sky.

- **NILC** (Basak and Delabrouille, 2012) extracts the CMB by minimizing the contamination from Galactic and extragalactic foregrounds and noise by computing the linear combination of the input maps that minimize the variance in a basis spanned by spherical wavelets called needlets.

- **SEVEM** (Martínez-González et al., 2003) is an internal template-cleaning approach based on difference maps between two neighboring Planck channels that trace the foreground contaminants at the corresponding frequency ranges. After that, a linear combination of these templates is subtracted from CMB-dominated frequency maps by minimizing the variance of the clean map outside a given mask. Finally, a combination of individually-cleaned frequency maps in harmonic space produces a cleaned, co-added, CMB map.

- **SMICA** (Cardoso et al., 2008) separates the CMB by synthetizing the maps from spherical harmonic coefficients obtained from the combination of a set of frequency maps with a multipole vector of weights.

However, although giving reliable and unprecedented detailed CMB maps, the implementation of these methods required years of research and developing in order to reach the final precise CMB measurement, with the introduction of notable improvements from the early to latest releases. Moreover, only 78% of the sky were confidently masked by these methods, due to strong contamination in the Galactic plane and brighter PS, which were masked once detected.

Moreover, as shown, for an accurate CMB measurement, methods with different characteristics should be developed and optimized. As explained before, the microwave sky is formed by several emissions, some of them with non-linear behaviors. Thus, neural networks seem to be the perfect candidates for separating the CMB from these emissions, especially after analyzing their performance when detecting PS by following an image segmentation problem in Chapter 4. Then, this Chapter will describe a similar approach but for recovering the CMB in temperature realistic simulations of *Planck*. The results shown in this Chapter are published in the *Astronomy and Astrophysics* journal (Casas et al., 2022a), in a work led by the author of this PhD thesis.

The Cosmic microwave background Extraction Neural Network (CENN) is the FCN used along this Chapter. It is trained with 60000 realistic simulations of the microwave sky as seen by *Planck*. Simulations, frequencies used and architecture are the same as in the Flat MultiPoSeIDoN case. However, there are slightly variations with respect that case:

- Patches have 256×256 pixels of area in order to have enough statistics at the large scale multipoles and clearly observe the first recombination peak at $l \sim 200$.

- Both train and test simulations comprises the 100% of the sky instead of avoiding the Galactic plane, in order to study the performance of the network at different conditions. In fact, the sky is divided into different latitude regions for testing the network. The divisions are explained in Figure 2.11.

- The label is now the CMB patch.

Figure 6.1 shows an example of simulation used for training and testing CENN at three latitude ranges: the Galactic plane (hereafter GP) at $0° < |b| < 5°$, being b the latitude

coordinate, the Galactic region (hereafter GR) at $5° < |b| < 30°$ and the extragalactic region (hereafter ER) at $30° < |b| < 90°$.

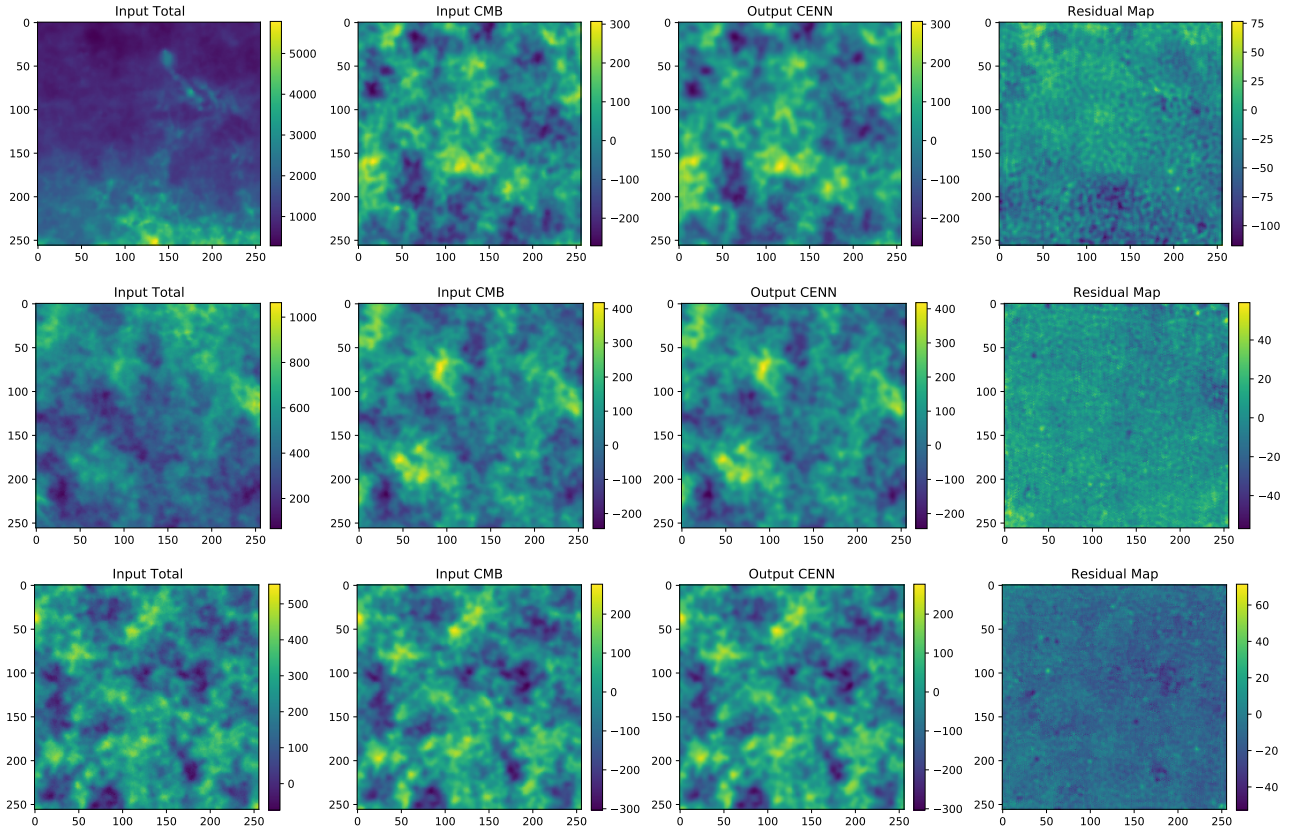

Figure 6.1: Patches at $0° < |b| < 5°$ (top row), $5° < |b| < 30°$ (middle row), and $30° < |b| < 90°$ (bottom row) latitude intervals. Columns from left to right represent the maps that CENN reads with all the emissions, the CMB on each simulation, used as a label during the training, CENN estimation, and the residual map computed as the difference between the second and the third column patches from left to right. The frequency is 217 GHz, and their temperature (in μK_{CMB}) is shown in the right vertical colourbars. The x and y labels are the pixel coordinates for each 256×256 patch.

Furthermore, Figure 6.2 shows the CENN's architecture, which is the same as for Flat MultiPoSeIDoN and the reader is encouraged to follow Section 4.1 for more information.

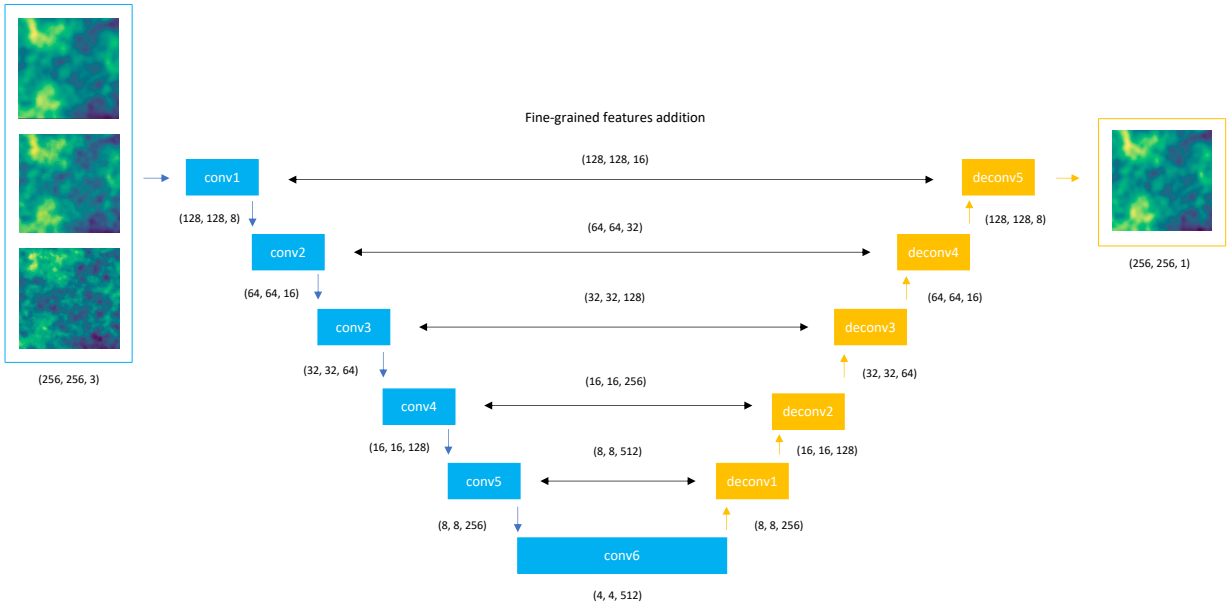

Figure 6.2: Architecture of CENN. The first one has a convolutional block, which produces eight feature maps. After that, the space dimensionality increases to 512 feature maps through five more convolutional blocks. These layers are connected to deconvolutional ones, which decreases the space dimensionality to eight feature maps in the last deconvolutional block. Fine-grained features are added from each convolution to its corresponding deconvolution.

6.2 Results

6.2.1 Power spectrum

Once trained, CENN is tested with similar, although not used for training, data. Four datasets are formed: one with 6000 realistic simulations of *Planck* at random positions at all sky, and three more, 2000 realistic simulations each, one for each latitude interval (GP, GR and ER).

The main statistical quantity used for evaluate the performance of the network is the power spectrum, described in Section 2.1. In this first analysis, CENN is only tested using the first dataset. After obtaining a set of CMB predictions, an algorithm for estimating the TT power spectrum is run for each of them, as so as for the input CMB signal contained on each simulation. After estimating the power spectrum of each patch, the average data is rebinned and then both mean and standard deviation are saved in order to have an uncertainty for each multipole range.

Figure 6.3 shows, in the top panel, the comparison between both $l(l+1)C_l/2\pi$ average power spectrum of the CMB in the input simulations (in blue) and the power spectrum estimates by CENN (in red). It is also represented in the bottom panel the difference between both signals (in black). In all cases the units are μK^2.

As shown, CENN recovers the CMB with high accuracy at all multipole ranges. In general, the mean absolute error is lower than $50 \pm 100\mu K^2$, especially at small scales ($l \sim 1500$) and large scales ($l < 500$), although the second ones could be an effect of poor information due to the patch smallness. It is remarkable however that in general the power spectrum difference between input and estimation is around $0 \pm 100\mu K^2$ at the rest of the multipole intervals. In fact, at smaller scales than $l \sim 1500$, *Planck* component separation methods generally overestimate the CMB power spectra due to instrumental noise effects.

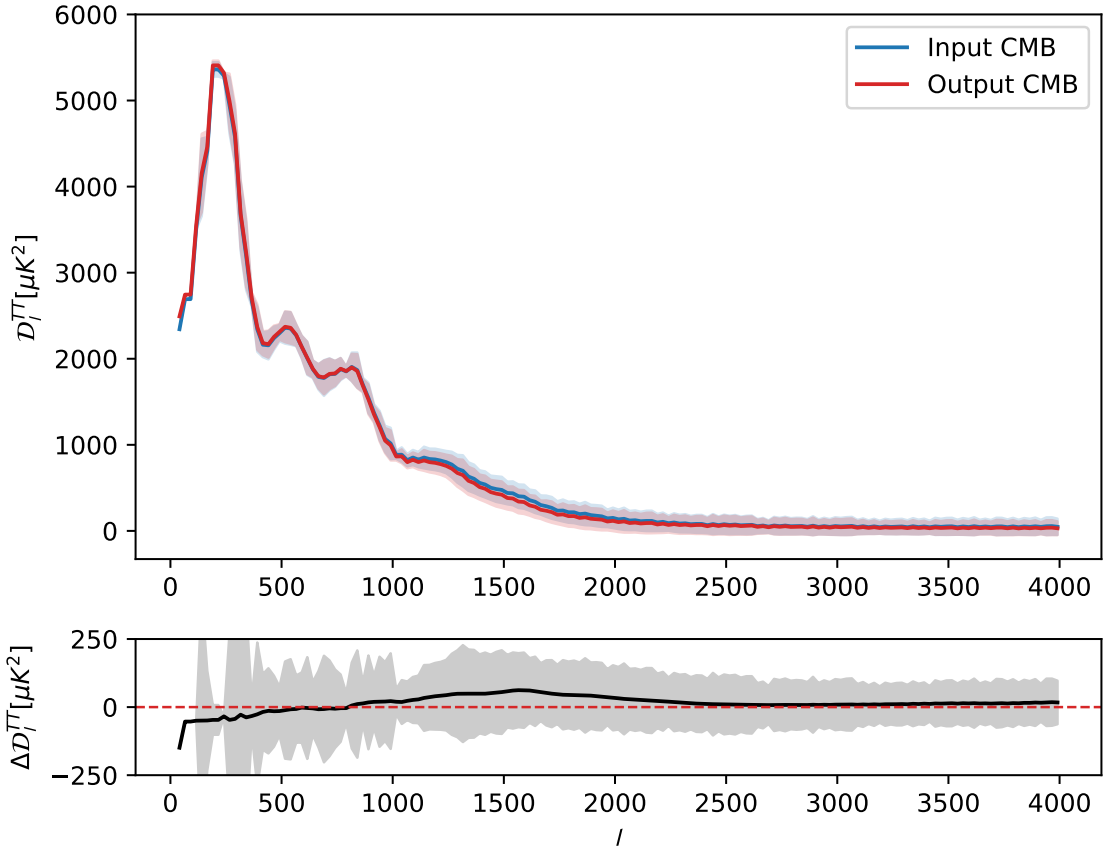

Figure 6.3: Average CMB power spectrum comparison computed over the entire test dataset. The input CMB is represented as a blue line, and the estimates from CENN are shown as a red line. The corresponding uncertainties, computed as the standard deviation of each bin, are the blue and red areas, respectively. The difference between input and output is shown in the bottom panel as a black line, with its uncertainty being the grey area. Red dashed line represents the ideal case.

6.2.2 Residuals

Although powerful information can be obtained from the power spectrum analysis, another useful quantity to evaluate is the power spectrum of the residual patches, that is, firstly, the residual patch is estimated as the difference between input and output patches (an example of that could be seen in the right column of Fig. 6.1). Then, the TT power spectrum is estimated for each patch and the mean and standard deviation are taken in the same as in the previous section.

Figure 6.4 shows the power spectrum of the residual patches (in black), in comparison with both input and output power spectrum (blue and red lines) and the power spectrum of each element in the simulations: dust, CIB, PS and Noise in orange, green, pink and brown respectively. All the power spectra are estimated not only for the test dataset with simulations at all sky (bottom right panel), but also at the three latitude test datasets (GP, GR and ER in the top left panel, top right panel and bottom left panel, respectively)

As shown, residuals are generally below the input signal in the four cases, even at small scales ($l > 2000$) when PS, CIB and instrumental noise starts to dominate the signal. Even more, although dust is the dominant emission polluting the signal with scales from 1 to 3 orders of magnitude above the input, residuals are generally between 1 and 2 orders of magnitude below it. Moreover, as seen by CENN's uncertainties, the CMB is accurateryly recovered at all *Planck* multipole range, since uncertainty results are compatible with zero until $l \sim 2000$.

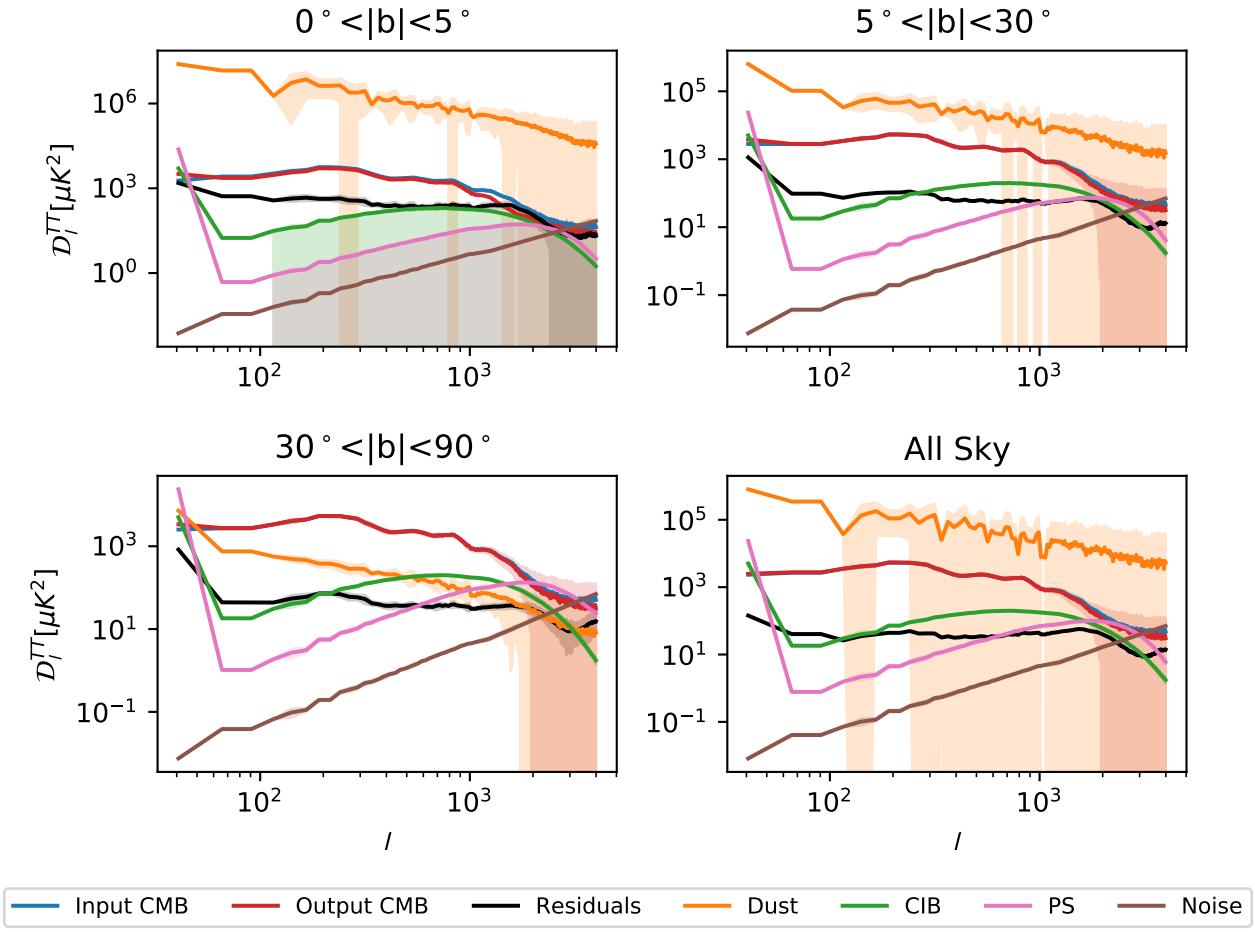

Figure 6.4: Average power spectrum of the residuals (black) and its uncertainty (grey area) for GP (top left panel), GR (top right panel) and ER latitude regions (bottom left panel) and for the whole sky (bottom right panel) against the input and output CMB (blue and red lines, respectively) and their uncertainties (blue and red areas, respectively). In all cases, these uncertainties are the standard deviation of each bin. The power spectrum of each foreground in the simulations is also plotted. The thermal dust is represented by the orange line, the PS and CIB by the pink and green lines, respectively, and the instrumental noise is shown by the brown line.

6.2.3 Looking inside CENN

Finally, Figure 6.5 shows how convolutional and deconvolutional blocks in CENN make inference and prediction, respectively over the input images about where the CMB is located and which is its value on each pixel. As shown, by using convolution operations, all regions in the patch are analyzed by CENN pixel-at-pixel. Then, at the deepest layers of the network, CENN learns by comparing the CMB values with the other foregrounds, in order to separate the signals. This is then used in the deconvolutional blocks to infer the shape of the CMB signal.

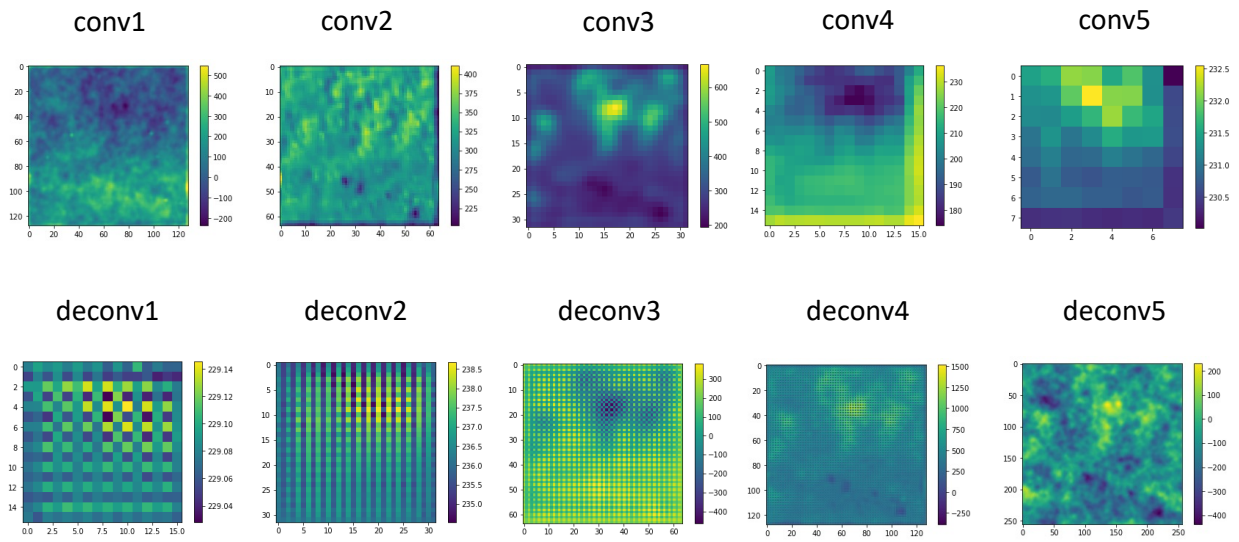

Figure 6.5: Example of the feature maps of each convolutional block of CENN (top panel) and each deconvolutional one (bottom panel).

CHAPTER 7

Recovering the cosmic microwave background in polarization

7.1 Introduction

As shown if Figures 2.6 and 2.7, after the third *Planck* release, the E-mode, as well as its correlation with temperature, were unprecedentedly constrained to multipoles $l \sim 1700$. However, the B-mode could not be constrained due to the low sensitivity of the instrument aboard the satellite. Post-*Planck* experiments such as BICEP/Keck estimated the lensed B-modes ($50 < l < 300$) with high accuracy, while ACT and SPTPol gave upper limits for the lensed B-mode at small scales ($l > 1000$). Moreover, future CMB experiments such LiteBIRD are designed to unveil the primordial B-mode signal at larger scales than BICEP/Keck ($l < 200$).

Following LiteBIRD Collaboration (2023b), in addition to Commander and NILC algorithms described in Chapter 6, three more component separation algorithms are already been used with LiteBIRD simulations to constrain the tensor-to-scalar ratio in future data releases:

- **FGBuster** (Puglisi et al., 2022) is a parametric method which uses nonlinear optimization to explore the likelihood function by firstly estimating all nonlinear spectral foreground parameters, secondly by estimating amplitudes conditionally with respect to the spectral parameters and, finally, by computing the angular power spectra from the amplitude maps.

- **Moment expansion** (Chluba et al. (2017), Vacher et al. (2022)) is a parametric method which allows Taylor-expansion based distortions in the SED models for each foreground, taking into account their nonlinear averaging effects.

- **cMILC** (Remazeilles et al., 2021) combines the moment-expansion and `NILC` methods by imposing additional constraints on the `ILC` weights that cancels out individual foreground components through their Taylor-expanded SEDs. Then each moment corresponds to one

additional linear constraint in the `ILC`, solving the combination of foreground-specific and variance constraints by single Lagrange multiplier system.

As shown, all these methods are different between them although they share the same goal. However, none of those methods are based on artificial intelligence, therefore it could be useful to explore some ML approach as in temperature. In this Chapter, a new methodology for recovering both E and B-mode CMB polarization based on neural networks is presented, work submitted to the *Astronomy and Astrophysics* journal (Casas et al., 2023b) and led by the author of this PhD thesis. The network is a simplified version of CENN, trained with less information and, because of that, with a simple and shallower architecture, as shown in Figure 7.2.

The network is trained with the same Q and U sky than POSPEN from Chapter 5, although with several modifications for this work:

- In order to take into account multi-frequency information, three HFI channels are used for helping CENN to discriminate between foreground contamination and the CMB. Due to the sensitivity given in Planck Collaboration (2020a), the channels used are 100, 143 and 217 GHz, which are the ones with the lower noise levels.

- The ~ 5 arcmin *Planck* sky is smoothed with several Gaussian filters of 30, 25 and 20 arcmin. This has to be done for increasing the signal-to-noise ratio of the patches. With this approach, it is easier to give a prediction, although losing part of the signal.

- Patches are cutted from all-sky maps by using the Krachmalnicoff and Puglisi (2021) approach, because the gnomonic projection from HEALPix causes a strong E-to-B leakage, as explained in Casas et al. (2023b).

Thus, CENN is trained with 10.000 simulations for Q and U simultaneously. Figure 7.1 shows an example of the sky used for training and testing the network.

Furthermore, CENN is formed in this study by four convolutional blocks with 8, 2, 4 and 2 kernels of sizes of 9, 9, 7 and 7, respectively. The number of filters is 8, 16, 64 and 128, respectively. Each layer has a subsampling factor of 2 and a padding type "Same" for adding space around the input data or the feature map in order to deal with possible loss in width and/or height dimension in the feature maps after having applied the filters. The convolutional blocks are connected to four deconvolutional ones to help the network to predict low-level

features by taking high-level features into account, previously inferred by the convolutional blocks, as explained in Casas et al. (2022a). The deconvolutional blocks have 2, 2, 2 and 4 kernels of sizes of 3, 5, 7 and 7, respectively. The number of filters is 64, 16, 8, and 1 with the same subsampling and padding type than the convolutional ones. In all layers, the activation function is leaky ReLU. This architecture is represented in Figure 7.2.

Once trained, the network is tested with 1.000 simulations. In order to analyze the performance of CENN, both EE and BB power spectra is estimated from Q and U input and recovered patches by using the NaMaster framework (Alonso et al., 2019).

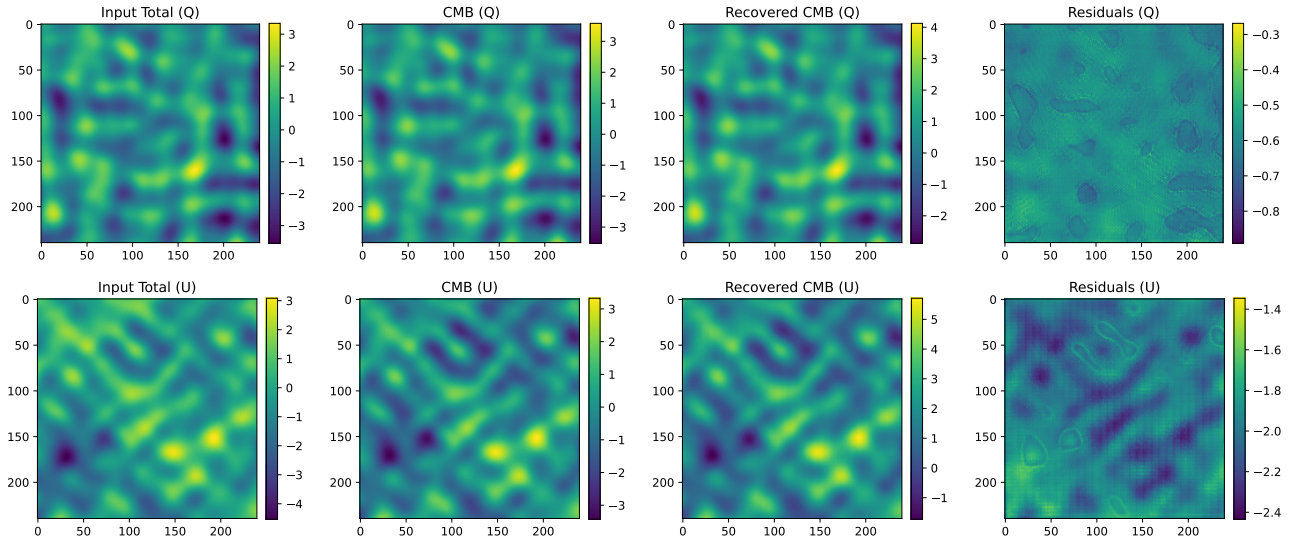

Figure 7.1: Example of one simulation used for training CENN in Q (top row) and U (bottom row) polarization at 30 arcmin resolution. The left column shows the input patch formed by all the emissions, the second one represents only the CMB signal, used for minimizing the loss function during training. The third column shows the output CMB after validating CENN and the fourth one shows the residual patch, which is computed as the difference between input and output CMB. The units in all the patches are μK_{CMB}.

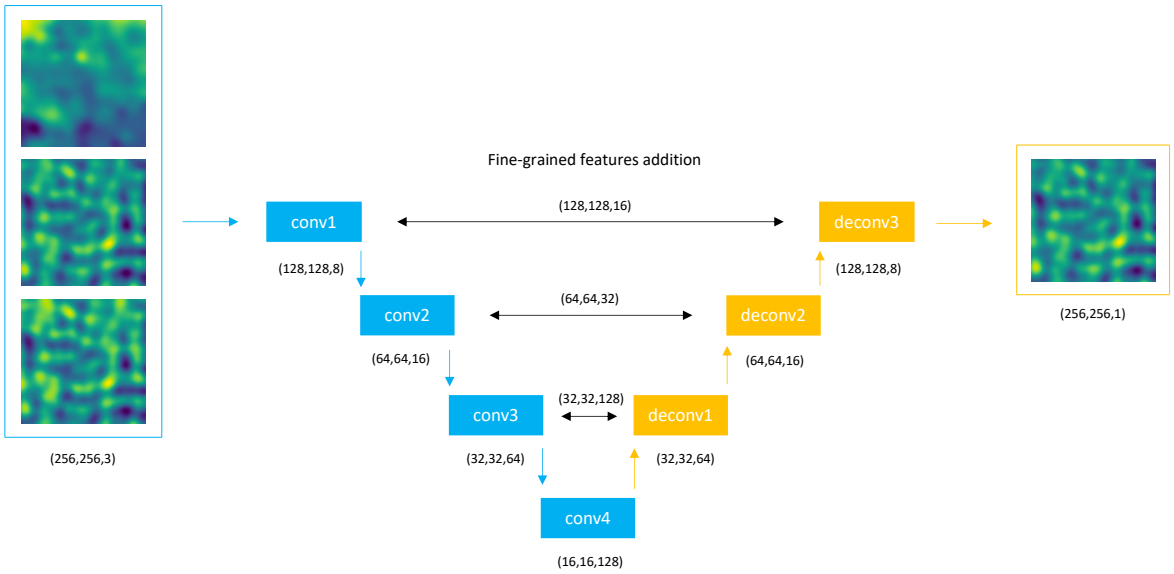

Figure 7.2: Architecture of CENN. It has four convolutional blocks connected to other four deconvolutional ones, which are trained to read 3 input images of the microwave sky and recovery the CMB signal of the central frequency channel (in this case corresponding to 143 GHz).

7.2 Results

7.2.1 Power spectrum

In order to give reliable results from CENN, the sky is smoothed with a Gaussian filter of 30 arcmin for increasing the signal-to-noise ratio. At this situation, the CMB is the dominant signal of the patch, making easier to recover it, although only multipoles up to $l \sim 600$ can be recovered. Figure 7.3 shows the comparison between input (in blue) and recovered power spectra (in red) for the E-mode (top panel) and the B-mode (bottom panel). Black line shows the difference between them, while the coloured areas are the uncertainty of each signal.

As seen, the E-mode is recovered with a mean difference of $0.1 \pm 0.3\ \mu K^2$, while the B-mode is estimated with $0.01 \pm 0.05\ \mu K^2$. At large scales ($l < 200$), CENN has a worse performance, probably due to the smallness of the patch and the poor statistics for learning the anisotropies.

In order to compare this performance against high resolution data, Figure 7.4 shows the network trained at 30 arcmin and tested against a new dataset of 25 arcmin resolution (orange line), formed by patches at the same position in the sky. As a comparison, the same architecture but trained at 25 arcmin (red line) is shown. Both differences with respect to the input signal (in blue) are shown as brown and black lines, respectively. The coloured areas represent the uncertainty of each signal. In this case, CENN can recover the signal for multipoles up to $l \sim 700$.

As shown, results are similar to the last case: the E-mode is recovered with a mean difference of $0.1 \pm 0.5\ \mu K^2$ for the 25 arcmin network and $0.3 \pm 0.7\ \mu K^2$ for the 30 arcmin one. On the other hand, the B-mode is estimated with a difference of $0.005 \pm 0.04\ \mu K^2$ in the 25 arcmin case and $0.01 \pm 0.05\ \mu K^2$ in the 30 arcmin one. Therefore, results show that it is better to train the network at the same resolution than testing, and that it is reliable at 25 arcmin resolution.

Then, Figure 7.5 shows the same analysis but testing three networks at 20 arcmin resolution, where instrumental noise starts to dominate the signal: both trained at 30 and 25 arcmin from the last cases (in magenta and orange, respectively), and a third one trained at 20 arcmin (in red). Their differences with respect the input signal (in blue) are shown in yellow, brown and black, respectively. The coloured areas represent the uncertainty of each signal. In this case, CENN can recover the signal for multipoles up to $l \sim 800$.

This case shows that CENN recovers the E-mode with similar performance to that of the previous cases when trained at 20 arcmin. However, both networks trained at 25 and 30 arcmin show slightly worse performance than in the previous cases. On the other hand, the B-mode show completely opposite performance: similar results are obtained when training at lower resolution and much worse performance is obtained when training at 20 arcmin. This implies that, firstly, the network can infer small scales from high resolution data, although trained at lower resolution and, secondly, noise levels are crucial during the weights optimization, that is, it seems better to train the network in unrealistic conditions although testing with realistic ones.

Finally, CENN is tested in Figure 7.6 with 5 arcmin resolution data without noise, in order to give an approximation for future CMB experiments with higher resolution than *Planck*. In this case the recovered signal can reach multipoles $l \sim 1500$. The performance is compared against a different foreground model (d4s2 from PySM) at the same position in the sky. The input signal is represented in blue, both train and d4s2 foreground models in red and orange, respectively. Their difference with respect to the input signal are shown in black and brown, respectively.

As shown, CENN accurately recovers both modes with similar results than the 30 arcmin case, especially at middle and smaller scales ($l > 200$), showing the d4s2 foreground model case worse results with respect the train foreground model.

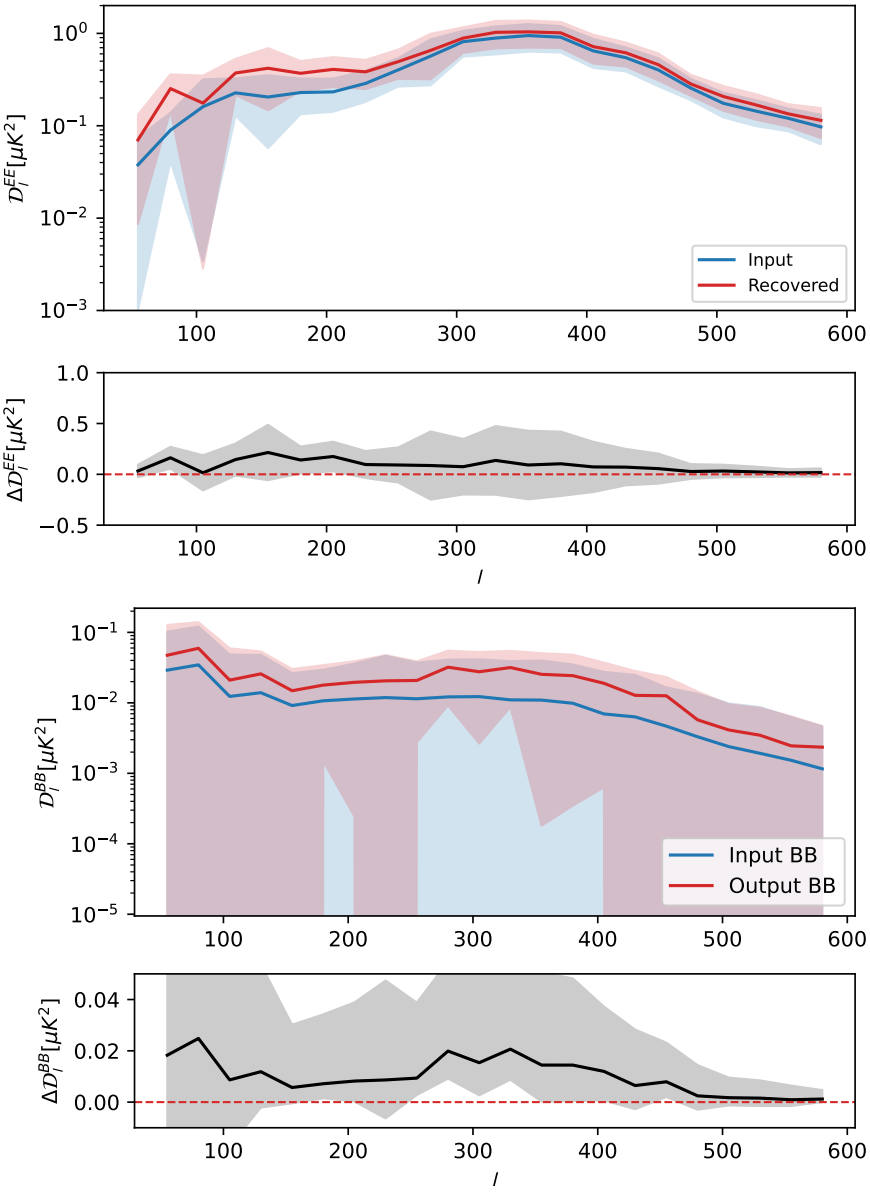

Figure 7.3: EE and BB power spectra recovery after training and testing CENN at 30 arcmin resolution. Top panel: EE power spectra in the input simulations (blue line) and patch average output power spectra from CENN at 30 arcmin red line. Top bottom panel: Difference between input and recovered EE power spectra at 30 arcmin (black line). Bottom panel: the same but for the B-mode. In all cases, coloured areas shown the standard deviation of each bin, considered as the uncertainty of the model. Red dashed lines in bottom sub-panels represent the ideal case.

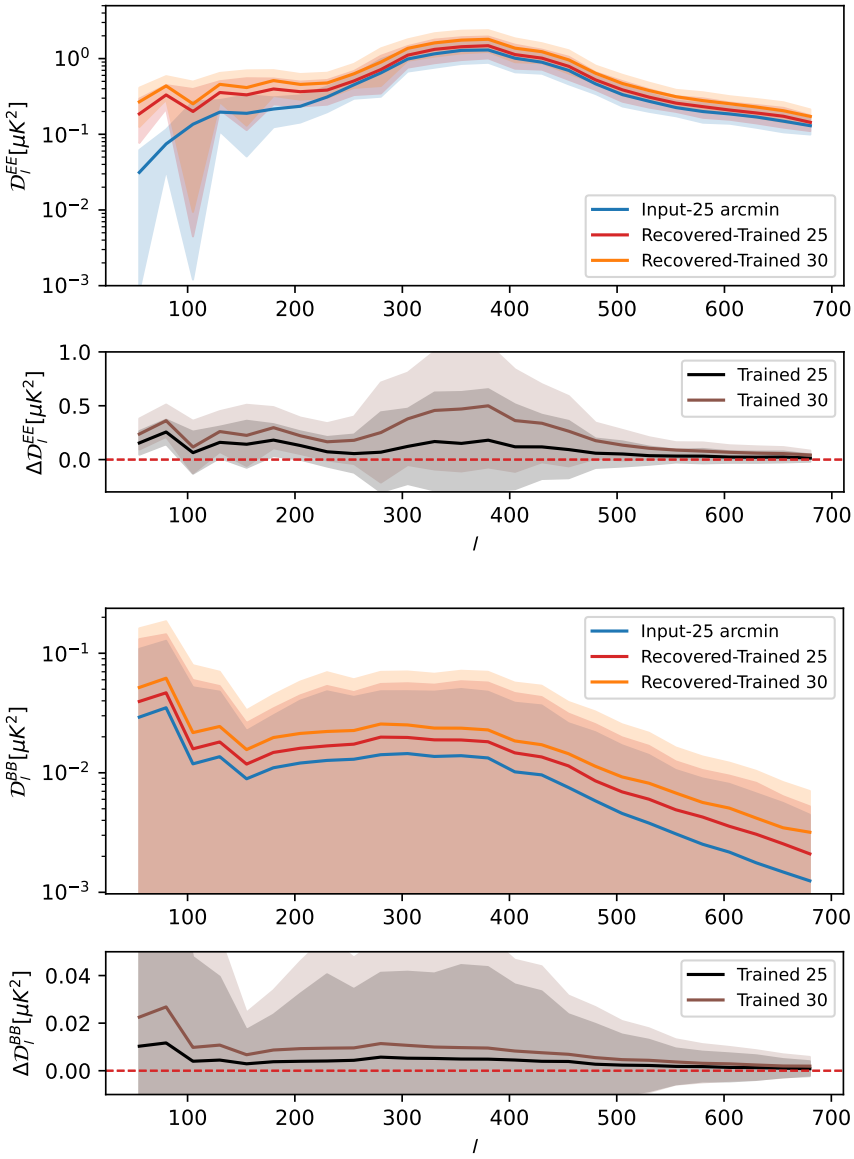

Figure 7.4: EE and BB power spectra recovery after training CENN at 25 and 30 arcmin resolution and tested at 25 arcmin. Top panel: EE power spectra in the input simulations (blue line), patch average output power spectra from CENN trained at 25 (red line) and 30 arcmin (orange line). Top bottom panel: Difference between input and recovered EE power spectra from CENN trained at 25 arcmin (black line) and 30 arcmin (brown line). Bottom panel: the same but for the B-mode. In all cases, coloured areas shown the standard deviation of each bin, considered as the uncertainty of the model. Red dashed lines in bottom sub-panels represent the ideal case.

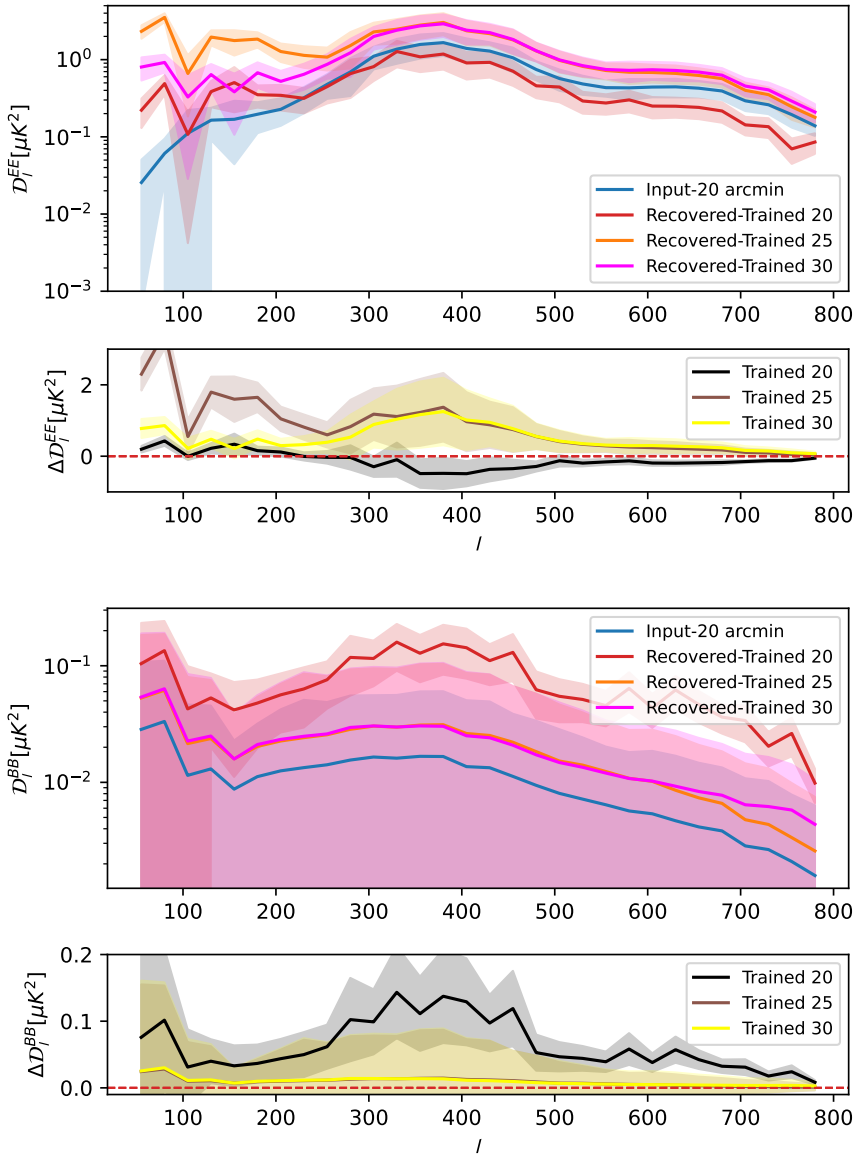

Figure 7.5: EE and BB power spectra recovery after training CENN at 20, 25 and 30 arcmin resolution and tested at 20 arcmin. Top panel: EE power spectra in the input simulations (blue line), patch average output power spectra from CENN trained at 20 (red line), 25 (orange line) and 30 arcmin (magenta line). Top bottom panel: Difference between input and recovered EE power spectra from CENN trained at 20 (black line), 25 (brown line) and 30 arcmin (yellow line). Bottom panel: the same but for the B-mode. In all cases, coloured areas shown the standard deviation of each bin, considered as the uncertainty of the model. Red dashed lines in bottom sub-panels represent the ideal case.

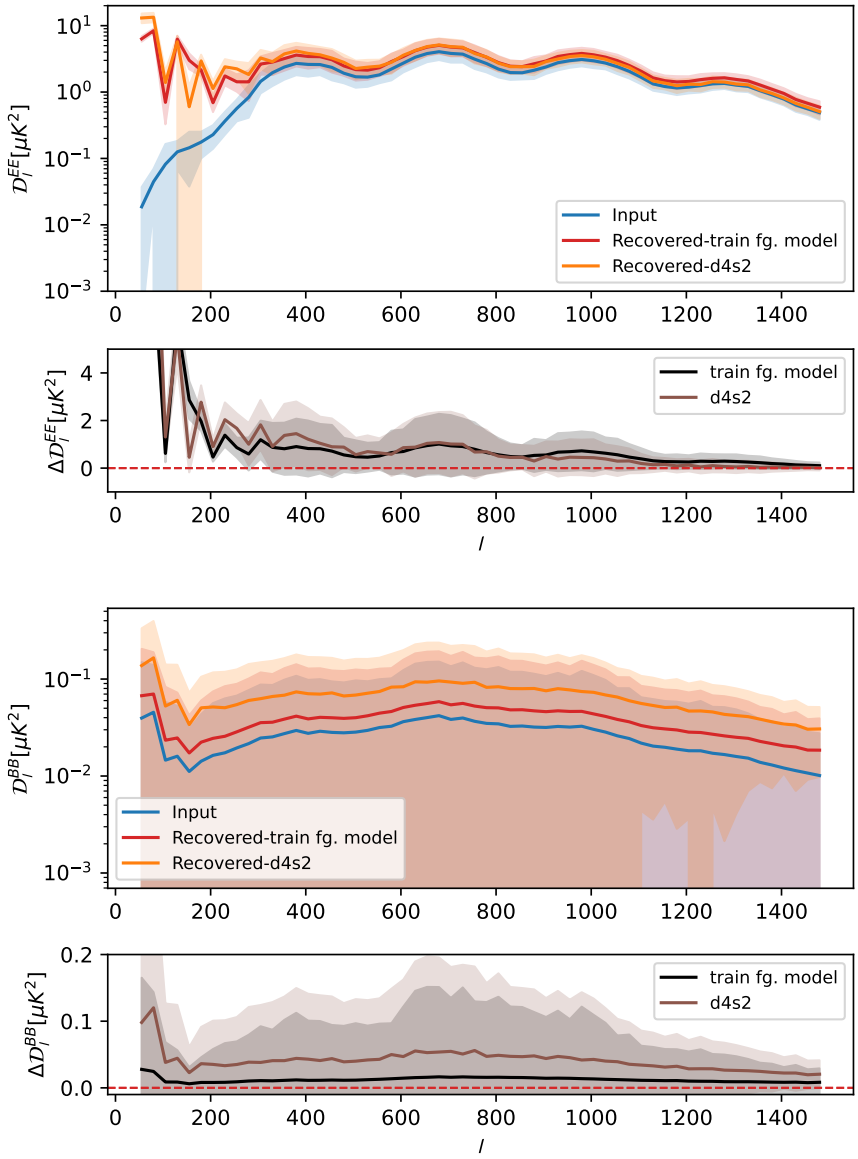

Figure 7.6: EE and BB power spectra recovery after training CENN at 5 arcmin resolution without noise. Left top panel: EE power spectra in the input simulations (blue line), patch average output power spectra from CENN training foreground model (red line) and d4s2 (orange line) foreground models. Left bottom panel: Difference between input and recovered EE power spectra from CENN for the training (black line) and d4s2 (brown line) foreground models. Right panel: the same but for the B-mode. In all cases, coloured areas shown the standard deviation of each bin, considered as the uncertainty of the model. Red dashed lines in bottom sub-panels represent the ideal case.

7.2.2 Residuals

Although the last section shows a reliable analysis to understand the performance of the network when recovering the CMB in polarization, it is more relevant to work with the power spectra of the residual patches, which are defined as the difference between input and recovered patches. In fact, the last comparison between input and recovered signals could not show the possible introduction of artifacts created along the patch by CENN.

Then, following the same resolution analysis than the last section, Figure 7.7 shows the power spectrum of the 30 arcmin residuals patches (in black), in comparison with the input signal (in blue) and the highest noise *seen* by CENN in the test simulations (represented as a dashed grey line), corresponding to the one for the 100 GHz channel.

As shown, for the E-mode, most of the residuals appear at large scales ($l < 200$), where they are of the order of the input signal. Above that scales, residuals decrease from 10^{-1} μK^2 to 10^{-2} μK^2. The B-mode shows accurate results, being the power spectrum of the residuals bellow the input signal at all scales, although the noise levels are above residual ones.

As shown in Figure 7.8, similar results are obtained when training at 25 arcmin and testing both 25 and 30 arcmin networks. In this case, the E-mode is accurately recovered above $l > 200$, since residuals are of two orders of magnitude lower, being the 25 arcmin (in black) ones below the 30 arcmin ones (in brown). Furthermore, at the smaller scales ($l > 500$), residuals are 10^{-2} μK^2, although the noise levels are above them. Similar results are obtained when recovering the B-mode. In fact, noise levels dominate the signal above $l > 500$.

Figure 7.9 shows the performance when training and tested at 20 arcmin resolution (in black). Both networks trained at 30 and 25 arcmin (in yellow and brown) are also tested at this resolution. The E-mode is similarly recovered with respect the previous cases, being the residuals one order of magnitude below the input one (in blue) above $l > 200$. The 20 arcmin case has higher residuals than the other cases above $l > 400$. On the other hand, the B-mode is accurately recovered for both 25 and 30 arcmin cases, although noise levels are higher than those train cases and indeed dominates the signal at all scales. However, the 20 arcmin case (in black) recovered the signal with higher residuals than the input.

Finally, when training and testing with 5 arcmin data without noise, CENN recovers the CMB with much better results than in the previous cases at $l > 200$, being the E-mode recovered with residuals between 1 and 2 orders of magnitude below the input signal (in blue) for the same foreground model as in the training (in black) and for the d4s2 foreground model (in brown).

On the other hand, the B-mode is accurately recovered when testing with the same foreground model, obtaining residuals of about 5×10^{-3} μK^2 for $l > 100$, although when changing the foreground model, it is obtained a worse recovering because residuals are at the order of the input signal. CENN seems to be sensitive to the sky used for training. Therefore, in order to a properly recovery of the B-mode, the foregrounds should be correctly characterized and the network should be optimized for each kind of sky.

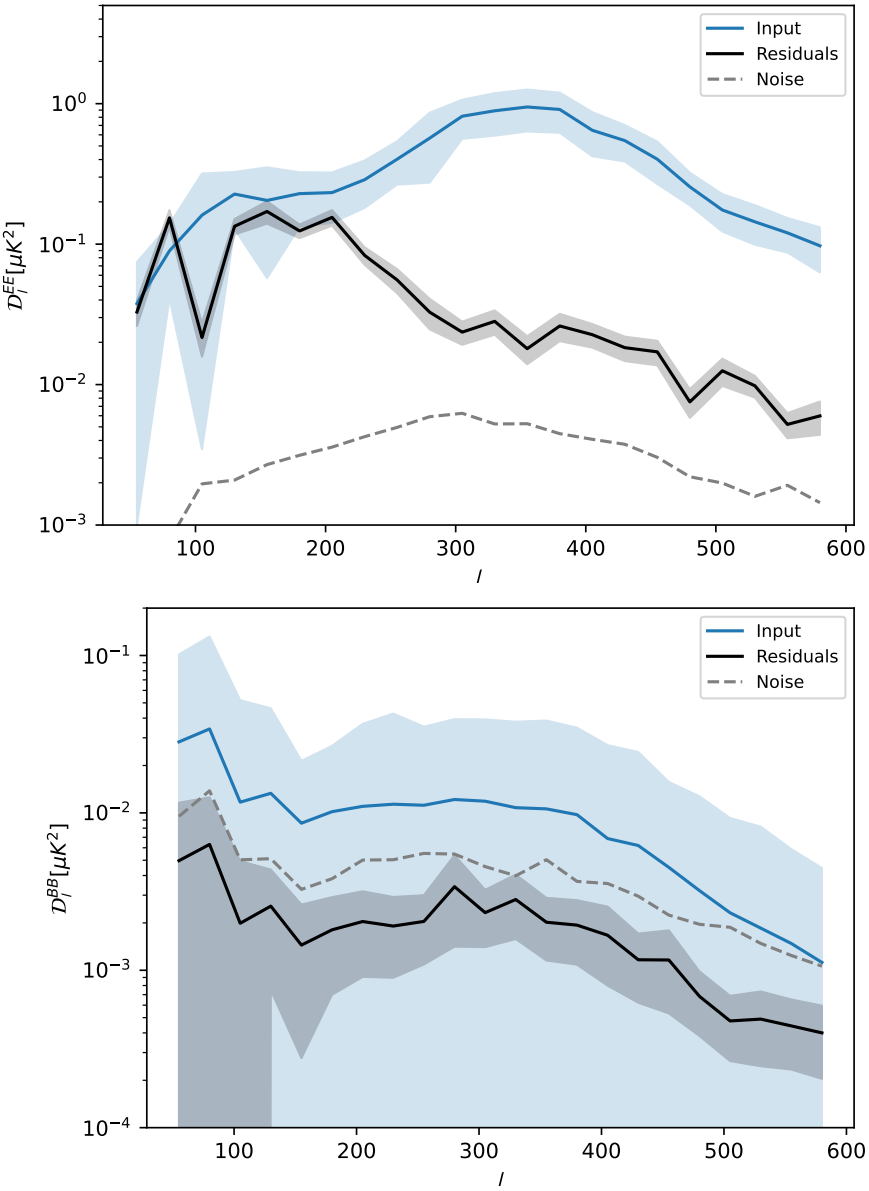

Figure 7.7: EE and BB residuals after training CENN at 30 arcmin resolution and tested at the same resolution. Top panel: EE power spectra in the input simulations (blue line) and residual patch average power spectra from CENN at 30 arcmin (black line). Bottom panel: the same but for the B-mode. In all cases, coloured areas shown the standard deviation of each bin, considered as the uncertainty of the model. Grey dashed line shows the higher instrumental noise levels *seen* by CENN.

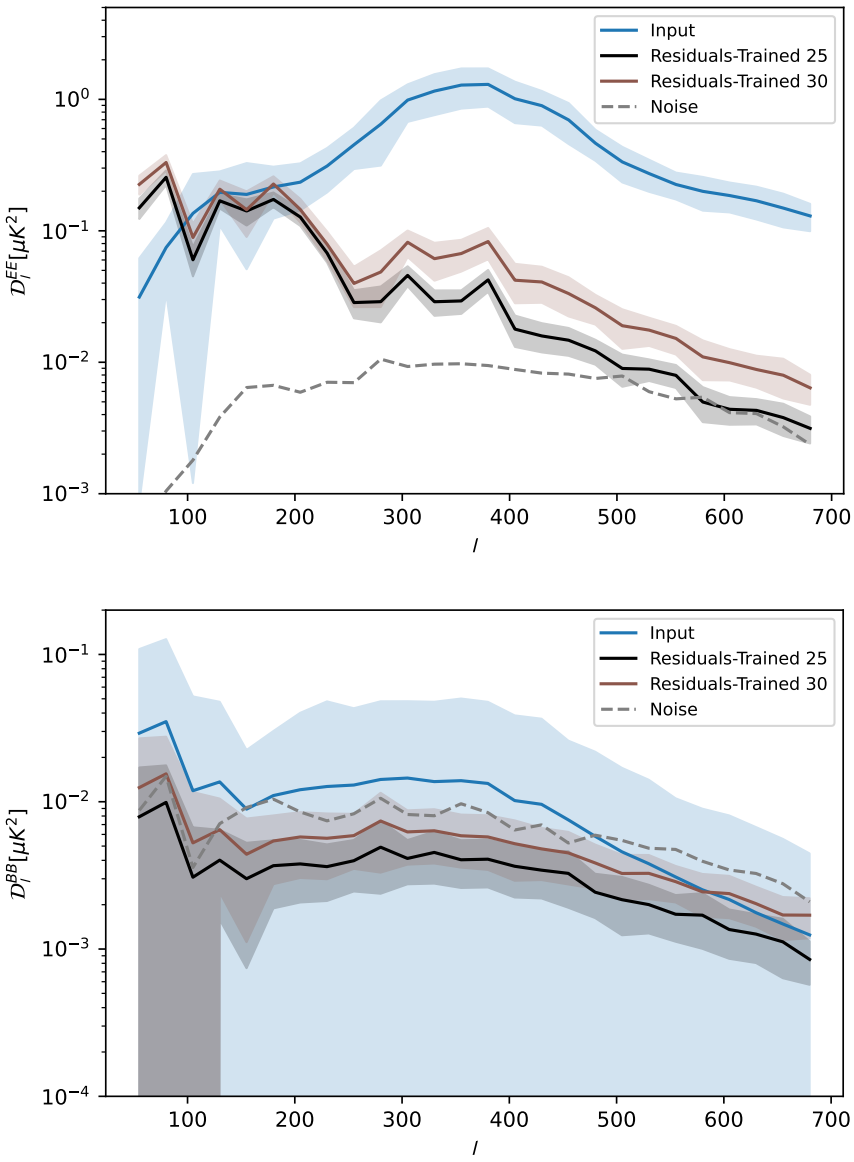

Figure 7.8: EE and BB residuals after training CENN at 25 and 30 arcmin resolution and tested at 25 arcmin. Top panel: EE power spectra in the input simulations (blue line), residual patch average power spectra from CENN at 25 (black line) and 30 arcmin (brown line). Bottom panel: the same but for the B-mode. In all cases, coloured areas shown the standard deviation of each bin, considered as the uncertainty of the model. Grey dashed line shows the higher instrumental noise levels *seen* by CENN.

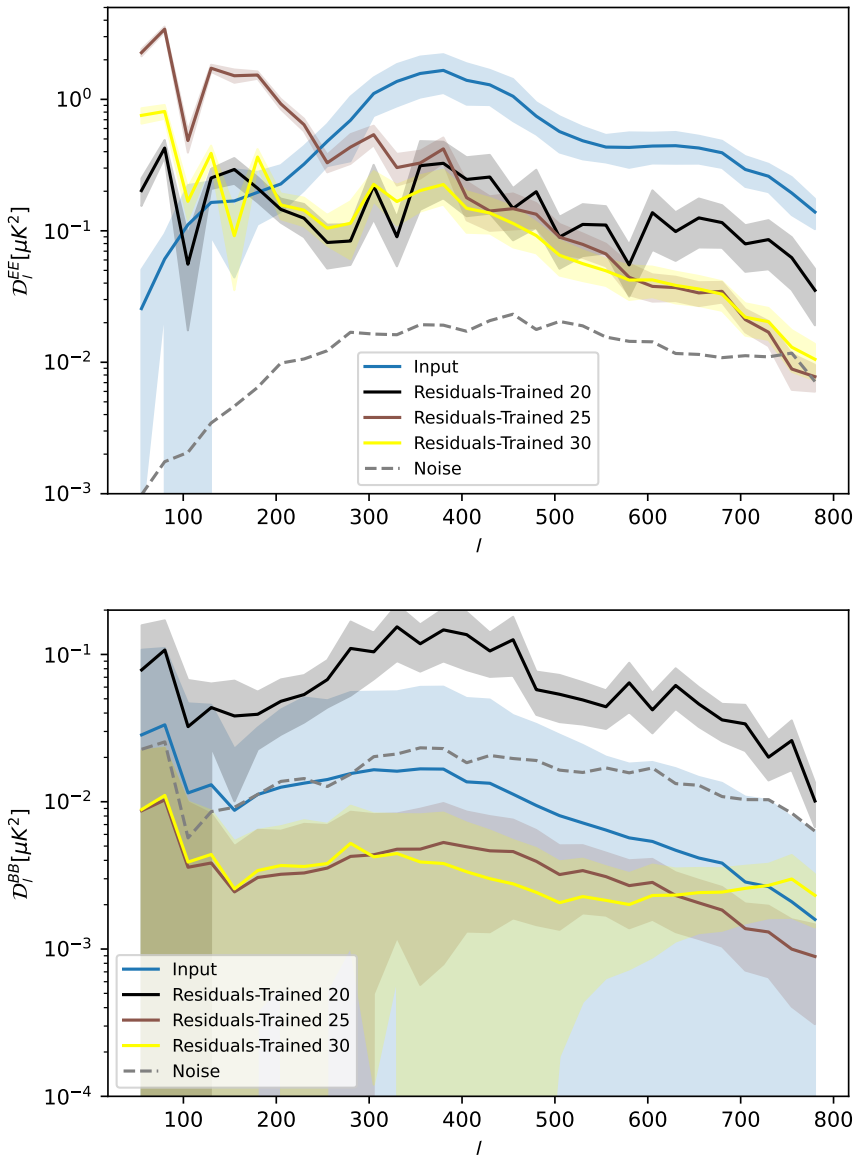

Figure 7.9: EE and BB residuals after training CENN at 20, 25 and 30 arcmin resolution and tested at 20 arcmin. Top panel: EE power spectra in the input simulations (blue line), residual patch average power spectra from CENN at 20 (black line), 25 (brown line) and 30 arcmin (yellow line). Bottom panel: the same but for the B-mode. In all cases, coloured areas shown the standard deviation of each bin, considered as the uncertainty of the model. Grey dashed line shows the higher instrumental noise levels *seen* by CENN.

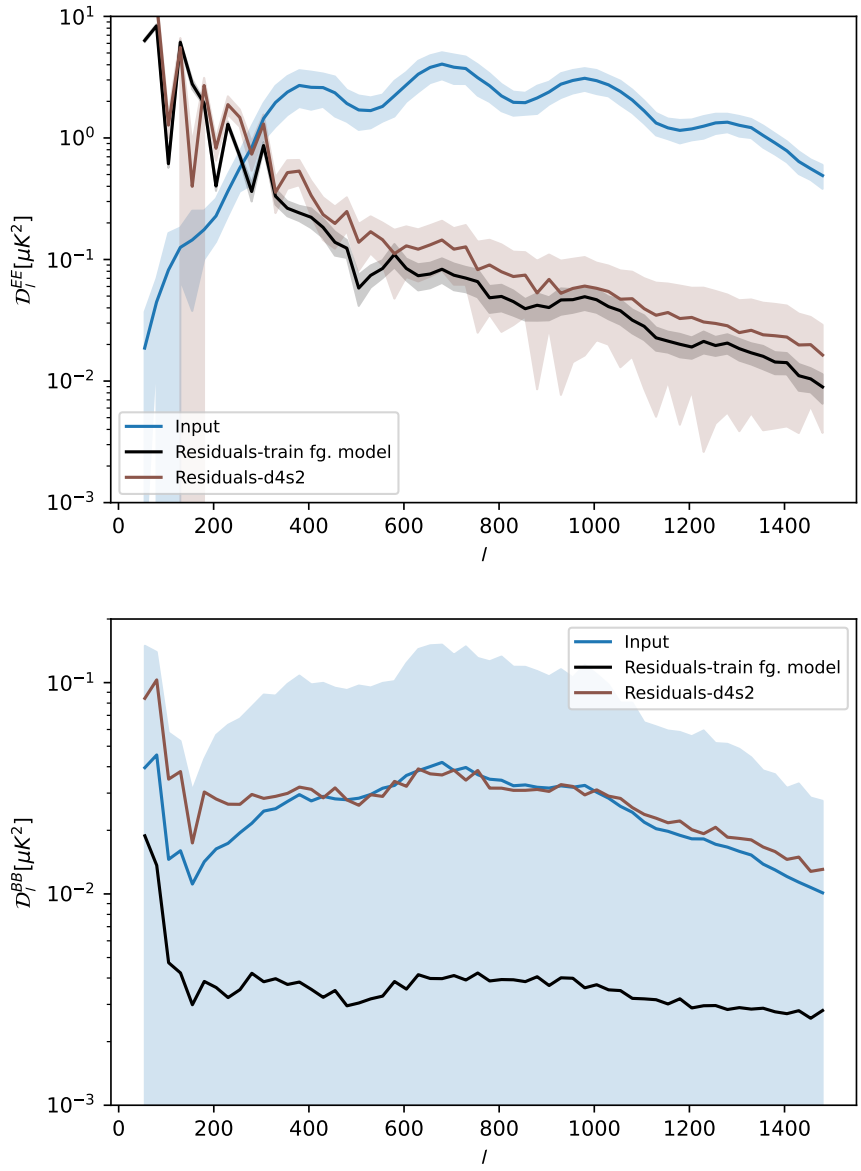

Figure 7.10: EE and BB residuals after training CENN at 5 arcmin resolution without noise. Left panel: EE power spectra in the input simulations (blue line), residual patch average power spectra from CENN for the training (black line) and d4s2 (brown line) foreground models. Right panel: the same but for the B-mode. In all cases, coloured areas shown the standard deviation of each bin, considered as the uncertainty of the model.

7.2.3 Looking inside CENN

Finally, Figures 7.11 and 7.12 show how convolutional and deconvolutional blocks in CENN make inference and prediction over the input Q and U patches, respectively, about where the CMB is located and which is its value on each pixel. As shown, by using convolution operations, all regions in the patch are analyzed by CENN pixel-at-pixel. Then, at the deepest layers of the network, CENN learns by comparing the CMB values with the other foregrounds, in order to separate the signals. This is then used in the deconvolutional blocks to infer the shape of the CMB signal.

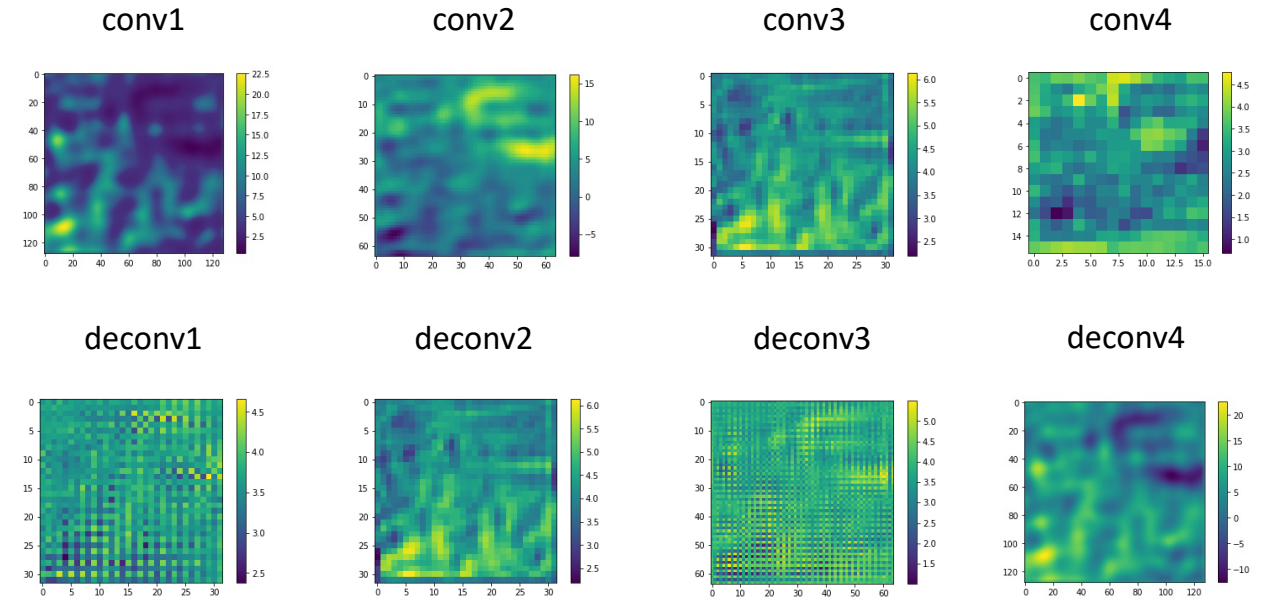

Figure 7.11: Example of the feature maps of each convolutional block of CENN (top panel) and each deconvolutional one (bottom panel) when testing with a Q patch.

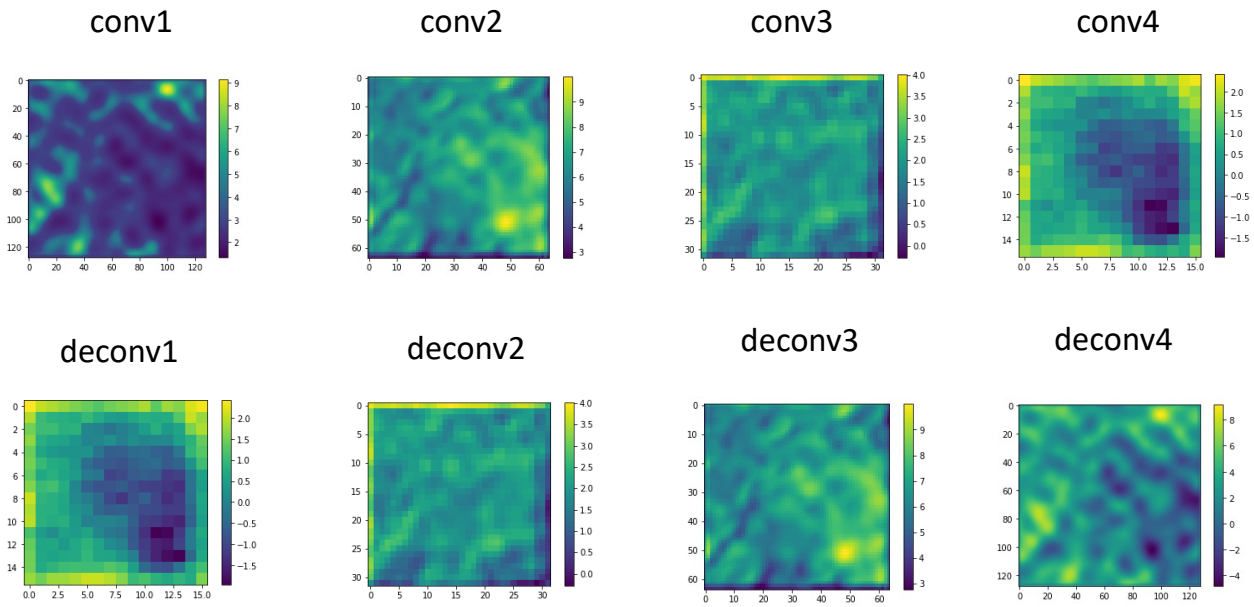

Figure 7.12: Example of the feature maps of each convolutional block of CENN (top panel) and each deconvolutional one (bottom panel) when testing with a U patch.

CHAPTER 8

Summary and conclusions

The cosmic microwave background, although it has been extensively studied these last decades, is still a profitable source of information about the first stages of the Universe and their relation with its actual nature. Future experiments will allow to unveil most details about the first instants of the Universe. However, with the improvement in the quality and quantity of data along with the challenging of the future CMB experiments, powerful automatic and flexible methods should be develop and optimize.

Artificial neural networks seem to be the perfect candidates for this objective since, once trained, they are capable to automatically and instantaneously perform a task. Moreover, they are specially designed to deal with non-linear behaviors from data, which are precisely the ones that have the CMB foregrounds, that is, the Galactic and extragalactic emissions that contaminate the CMB maps.

This PhD thesis wants to cement several methodologies based on neural networks, applied to different studies: radio source detection in temperature maps, the constraining of their polarization properties and the recovery of the CMB in both temperature and polarization maps. The that have been obtained with these neural networks are presented in four chapters (4, 5, 6 and 7, respectively), and published in high-impact papers (Casas et al. (2022b), Casas et al. (2023a), Casas et al. (2022a) and Casas et al. (2023b)) led by the author of this PhD thesis.

Despite some variations in the methodological details of each analysis, the main idea along the four works is to train and test convolutional neural networks with realistic simulations of the Planck satellite. The sky *seen* by the neural networks is not only formed by the CMB, but it also contains dust and synchrotron emission from our Galaxy, thermal SZ from Galaxy clusters, the CIB, radio and dusty sources, and instrumental noise. In all train datasets, the microwave sky is cut into squared patches with different area size, depending on each analysis.

MultiPoSeIDoN, CENN and POSPEN are the neural networks developed, trained and analyzed in this PhD thesis. First and second ones are fully-convolutional neural networks, with architectures designed to perform image segmentation. The third one is a convolutional

neural network, designed to extract a numerical value from an image. In all cases the learning is supervised, that is, an emission in the simulations is kept and used for minimizing a loss function at the end of each architecture. This is called a label. In Chapter 4, PS detection is performed by using the PS patch as a label. In Chapters 6 and 7, the CMB is used instead. In Chapter 5, the polarization flux density of the PS is the quantity of interest.

MultiPoSeIDoN is trained in Chapter 4 to detect PS in realistic simulations of the HFI *Planck* channels. Detection is performed by using multi-frequency information, and results are compared with the Matrix Filters, a method used in *Planck*. Based on the analysis of the photometry, completeness and reliability, the neural network is more accurate than the filter, especially at the highest frequencies, where Galactic dust dominates the signal. More on, it is more reliable than the filter at the same completeness, since it detects a much lower number of spurious sources or false positives. The comparison against the same neural network trained with single-frequency information show that accurate results are obtained when training with multi-frequency data.

POSPEN is trained in Chapter 5 for characterizing the polarization properties of PS in realistic simulations of the 217 GHz *Planck* channel. It is a convolutional neural network that reads a patch in the sky and outputs a numerical value for the PS located in the center of the patch. Once obtained the polarization flux density in both Q and U channels, it can estimate the polarization angle. Results shown a better performance with respect to the actual methods, because it could give reliable results at 80 mJy, instead of the 200 mJy limit for the traditional methods. Moreover, it could estimate accurately the polarization angle for PS above 200 mJy.

CENN is trained in Chapters 6 and 7 to recover the CMB signal in temperature and polarization, respectively. Although varying the size of the train dataset and the deepness of the network, accurate results are obtained in both cases in terms of the power spectra. In particular, in temperature, CENN shows reliable results by comparing both true and recovered CMB signals, and also when estimating the power spectra of the residual patches, especially at small scales, where noise starts to dominate the signal. In fact, the neural network behaves differently with respect to traditional component separation methods, giving a precise estimation of the power spectra instead of overestimate the signal.

In polarization, although the sensitivity of the *Planck* instrument did not allow an estimation for the B-mode, CENN is able to give, at least in the actual conditions, an upper limit in *Planck* realistic simulations. Even more, CENN accurately estimates the E-mode. Due to

the low signal-to-noise ratio, patches should be smoothed, therefore the network only can recover the polarization up to multipoles $l \sim 800$. When testing without noise, the network can accurately recover the CMB at smaller scales, an approximation for future CMB experiments with higher sensitivity than *Planck*. By comparing the performance when training at higher resolution, it is found that the network is more accurate when trained with lower levels of noise, although tested in worse conditions. By testing with a different foreground model with respect to the one in the training dataset, it is found that CENN is sensitive to the sky used for training, and therefore the network should be optimize for each kind of sky, and also foregrounds should be correctly characterize, especially for the B-mode recovery.

Therefore, this PhD thesis is not only a compendium of methodologies based on neural networks for CMB studies, but also wants to study how neural networks *behave* with this kind of data. Understanding these methods would allow to use them in future CMB experiments. For example, the CENN extension to all sky should be the next objective, in order to give not only estimations for the B-mode at large scales, where recombination peak represents the primordial signal, but also the use of information of larger areas for recover the entire signal in both modes. Furthermore, although *Planck* estimated with high reliability the polarization of point sources, poor statistics were obtained. Probably, a higher number of sources lies bellow the reliability limit of the Filtered Fusion method used in *Planck*. POSPEN can probably characterize their flux density, and also the polarization angle for the brighter sources in each *Planck* channel. As commented above, better statistics would allow a better understanding of their polarized behavior.

Resumen y concusiones

El fondo cósmico de microondas, a pesar de haber sido detenidamente estudiado estas últimas décadas, es aún una gran fuente de información sobre los primeros instantes del Universo y su relación con su actual naturaleza. Experimentos futuros serán capaces de descubrir más detalles aún sobre los primeros instantes del Universo. Sin embargo, el aumento de la cantidad y calidad de los datos, junto con los desafíos a los que deberán enfrentarse los futuros experimentos del fondo cósmico de microondas, nuevos métodos automáticos y flexibles deben desarrollarse y optimizarse.

Las redes neuronales artificiales parecen ser las perfectas candidatas para este objetivo dado que, una vez entrenadas, son capaces de realizar una tarea automática e instantáneamente. Además, están especialmente diseñadas para lidiar con comportamientos no lineales en los datos, que son precisamente los que tienen los foregrounds del fondo cósmico de microondas, es decir, las emisiones Galácticas y extragalácticas que contaminan la señal.

Esta tesis de doctorado pretende cimentar varias metodologías basadas en redes neuronales, aplicadas a diferentes estudios: la detección de fuentes radio en mapas de temperatura, el ajuste de sus propiedades de polarización y la recuperación del fondo cósmico de microondas en mapas de temperatura y polarización, resultados divididos en cuatro capítulos de esta tesis (4, 5, 6 and 7, respectivamente) y publicados en artículos de alto impacto (Casas et al. (2022b), Casas et al. (2023a), Casas et al. (2022a) y Casas et al. (2023b), respectivamente), liderados por el autor de esta tesis de doctorado.

A pesar de variar detalles de la metodología en cada análisis, la idea principal en los cuatro trabajos es entrenar y testear redes neuronales convolucionales con simulaciones realistas del satélite Planck. El cielo *visto* por las redes neuronales no sólo está formado por el CMB, sino que además contiene polvo y emisión sincrotrón de nuestra galaxia, efectos termales de SZ de grupos de galaxias, el CIB, fuentes radio y dusty, y ruido instrumental. En todos los sets de entrenamiento, el cielo de microondas es cortado en parches cuadrados con diferente area, dependiendo de cada análisis.

MultiPoSeIDoN, CENN y POSPEN son las redes neuronales desarrolladas, entrenadas y analizadas en esta tesis doctoral. La primera y la segunda son redes neuronales totalmente convolucionadas, con arquitecturas diseñadas para realizar segmentación de imagenes. La tercera es una red neuronal convolucional, diseñada para extraer un valor numérico de una imagen.

En todos los casos, el aprendizaje es supervisado, es decir, se guarda una emisión de las simulaciones para minimizar la función de coste al final de cada arquitectura. A esta emisión se la llama etiqueta. En el Capítulo 4, para la detección de fuentes se utiliza el parche con las fuentes puntuales como etiqueta. En los Capítulos 6 y 7, se utiliza el CMB en su lugar. En el Capítulo 5 la densidad de flujo de polarización de la fuente puntual es la cantidad de interés.

MultiPoSeIDoN se entrena en el Capítulo 4 para detectar fuentes puntuales en simulaciones realistas de los canales HFI de *Planck*. La detección se realiza utilizando información multi-frecuencia, y se comparan los resultados con los Matrix Filters, un método usado en *Planck*. En base al análisis de la fotometría, completitud y fiabilidad, la red neuronal es más precisa que el filtro, especialmente a las frecuencias más altas, cuando el polvo Galáctico domina la señal. Más aún, es más fiable que el filtro a la misma completitud, ya que detecta un número mucho más bajo de fuentes espúreas or false positives. La comparación con la misma red neuronal entrenada con información uni-frecuencia muestra resultados más precisos cuando la red se entrena con datos multi-frecuencia.

POSPEN se entrena en el Capítulo 5 para caracterizar las propiedades de polarización de fuentes puntuales en simulaciones realistas del canal 217 GHz de *Planck*. Es una red neuronal convolucional que lee un parche del cielo y devuelve un valor numérico de la densidad de flujo de polarización de la fuente puntual localizada en el centro del parche. Una vez obtenida la polarización de los canales Q y U, puede estimarse el ángulo de polarización. Los resultados muestran un rendimiento mejor que los métodos utilizados actualmente, ya que es capaz de dar resultados fiables hasta 80 mJy, en lugar de los 200 mJy que suelen tener los métodos tradicionales como límite. Además es capaz de estimar precisamente el ángulo de polarización de fuentes por encima de los 200 mJy.

CENN se entrena en los Capítulos 6 y 7 para recuperar la señal del CMB en temperatura y polarización, respectivamente. A pesar de variar el tamaño del set de entrenamiento y la profundidad de la red, se obtienen resultados precisos en ambos casos en términos del power spectrum. En particular, en temperatura, CENN obtiene resultados fiables al comparar las señales del CMB real con el recuperado, y además cuando se estima el power spectrum de los parches residuo, especialmente a pequeña escala, donde el ruido domina la señal. De hecho, la red neuronal se comporta de manera diferente con respecto a los métodos tradicionales de separación de componentes, ya que recupera con precisión la señal a pequeña escala en lugar de sobreestimarla.

En polarización, a pesar de que la baja sensibilidad del instrumento de *Planck* no permitió a la Colaboración obtener una estimación del modo B, CENN es capaz de dar, al menos en las condiciones actuales, un upper limit en simulaciones realistas de *Planck*. Aún más, CENN estima precisamente el modo E. Debido a la baja señal con respecto al ruido, los parches deben filtrarse, de modo que la red sólo puede recuperar la polarización hasta multipolos $l \sim 800$. Cuando se testea sin ruido, una aproximación de futuros experimentos del CMB con mayor sensibilidad que *Planck*, la red recupera precisamente escalas más pequeñas. Mediante la comparación cuando se entrena con mayor resolución, se encuentra que la red es más precisa cuando se entrena con niveles más bajos de ruido, aún siendo testeada en peores condiciones. Cuando se testea con un modelo de foregrounds distinto con respecto al del set de entrenamiento, se encuentra que CENN es sensible al cielo usado para su entrenamiento y, por tanto, la red debería optimizarse para cada tipo de cielo, así como los foregrounds deberían caracterizarse correctamente, especialmente para la recuperación del modo B.

Por lo tanto, esta tesis de doctorado no es sólo un compendio de metodologías basadas en redes neuronales para estudios del CMB, sino que pretende estudiar como las redes neuronales *se comportan* con esta clase de datos. La comprensión de estos métodos permitiría su uso en futuros experimentos del CMB. Por ejemplo, la extensión de CENN a todo cielo debería ser el siguiente objetivo, para dar no sólo estimaciones del modo B a gran escala, donde el pico de la recombinación representa la señal primordial, sino que permitiría recuperar toda la señal de ambos modos utilizando información de áreas más grandes. Más aún, aunque *Planck* estimo con gran fiabilidad las fuentes puntuales polarizadas, se obtuvo una estadística pobre. Probablemente, un mayor número de fuentes presentan polarización por debajo del límite de fiabilidad del método Filtered Fusion usado en *Planck*. Probablemente POSPEN pueda caracterizar su densidad de flujo, y además el ángulo de polarización de las fuentes más brillantes en cada uno de los canales de *Planck*. Como se comentó durante la tesis, una mejora en la estadística de la detección de las fuentes permita entender mejor su comportamiento en polarización.

Bibliography

Ali-Haïmoud, Y., Hirata, C. M., and Dickinson, C. (2009). A refined model for spinning dust radiation. *Monthly Notices of the Royal Astronomical Society*, 395(2):1055–1078.

Alonso, D., Sanchez, J., Slosar, A., and LSST Dark Energy Science Collaboration (2019). A unified pseudo-C_ℓ framework. *Monthly Notices of the Royal Astronomical Society*, 484(3):4127–4151.

Argueso, F., Sanz, J. L., Herranz, D., Lopez-Caniego, M., and Gonzalez-Nuevo, J. (2009). Detection/estimation of the modulus of a vector. Application to point source detection in polarization data. *Mon. Not. Roy. Astron. Soc.*, 395:649–656.

Balbi, A., Ade, P., Bock, J., Borrill, J., Boscaleri, A., de Bernardis, P., Ferreira, P. G., Hanany, S., Hristov, V., Jaffe, A. H., Johnson, B., Lange, A. E., Lee, A. T., Mauskopf, P. D., Netterfield, C. B., Oh, S., Pascale, E., Rabii, B., Richards, P. L., Smoot, G. F., Stompor, R., Winant, C. D., and Wu, J. H. P. (2001). Maps of the CMB from the MAXIMA experiment. *Mem. Societa Astronomica Italiana*, 72:849–852.

Baron, D. (2019). Machine Learning in Astronomy: a practical overview. *arXiv e-prints*, page arXiv:1904.07248.

Basak, S. and Delabrouille, J. (2012). A needlet internal linear combination analysis of WMAP 7-year data: estimation of CMB temperature map and power spectrum. *Monthly Notices of the Royal Astronomical Society*, 419(2):1163–1175.

Baumann, D. (2018). TASI Lectures on Primordial Cosmology. *arXiv e-prints*, page arXiv:1807.03098.

Bengio, Y., Lamblin, P., Popovici, D., and Larochelle, H. (2006). Greedy layer-wise training of deep networks. In Schölkopf, B., Platt, J., and Hoffman, T., editors, *Advances in Neural Information Processing Systems*, volume 19. MIT Press.

Bennett, C. L., Larson, D., Weiland, J. L., Jarosik, N., Hinshaw, G., Odegard, N., Smith, K. M., Hill, R. S., Gold, B., Halpern, M., Komatsu, E., Nolta, M. R., Page, L., Spergel, D. N., Wollack, E., Dunkley, J., Kogut, A., Limon, M., Meyer, S. S., Tucker, G. S., and Wright,

E. L. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. *Astrophysical Journal, Supplement*, 208(2):20.

Bonavera, L., Gonzá lez-Nuevo, J., Argüeso, F., and Toffolatti, L. (2017a). Statistics of the fractional polarization of compact radio sources in planck maps. *Monthly Notices of the Royal Astronomical Society*, 469(2):2401–2411.

Bonavera, L., Gonzá lez-Nuevo, J., Marco, B. D., Argüeso, F., and Toffolatti, L. (2017b). Statistics of the fractional polarization of extragalactic dusty sources in planck HFI maps. *Monthly Notices of the Royal Astronomical Society*, 472(1):628–635.

Bonavera, L., Suarez Gomez, S. L., González-Nuevo, J., Cueli, M. M., Santos, J. D., Sanchez, M. L., Muñiz, R., and de Cos, F. J. (2021). Point source detection with fully convolutional networks. Performance in realistic microwave sky simulations. *Astronomy & Astrophysics*, 648:A50.

Bucher, M. (2015). Physics of the cosmic microwave background anisotropy. *International Journal of Modern Physics D*, 24(2):1530004–303.

Cai, Z.-Y., Lapi, A., Xia, J.-Q., De Zotti, G., Negrello, M., Gruppioni, C., Rigby, E., Castex, G., Delabrouille, J., and Danese, L. (2013). A Hybrid Model for the Evolution of Galaxies and Active Galactic Nuclei in the Infrared. *The Astrophysical Journal*, 768(1):21.

Cardoso, J.-F., Martin, M., Delabrouille, J., Betoule, M., and Patanchon, G. (2008). Component separation with flexible models. Application to the separation of astrophysical emissions. *arXiv e-prints*, page arXiv:0803.1814.

Casas, J. M., Bonavera, L., González-Nuevo, J., Baccigalupi, C., Cueli, M. M., Crespo, D., Goitia, E., Santos, J. D., Sánchez, M. L., and de Cos, F. J. (2022a). CENN: A fully convolutional neural network for CMB recovery in realistic microwave sky simulations. *Astronomy & Astrophysics*, 666:A89.

Casas, J. M., Bonavera, L., González-Nuevo, J., Cueli, M. M., Crespo, D., Goitia, E., González-Gutiérrez, C., Santos, J. D., Sánchez, M. L., and de Cos, F. J. (2023a). Constraining the polarisation flux density and angle of point sources by training a convolutional neural network. *Astronomy & Astrophysics*, 670:A76.

Casas, J. M., Bonavera, L., González-Nuevo, J., Puglisi, G., Baccigalupi, C., Cueli, M. M., Crespo, D., González-Gutiérrez, C., and de Cos, F. J. (2023b). Recovering the E and B-mode CMB polarization at sub-degree scales with neural networks. *arXiv e-prints*, page arXiv:2310.07590.

Casas, J. M., González-Nuevo, J., Bonavera, L., Herranz, D., Suarez Gomez, S. L., Cueli, M. M., Crespo, D., Santos, J. D., Sánchez, M. L., Sánchez-Lasheras, F., and de Cos, F. J. (2022b). Multi-frequency point source detection with fully convolutional networks: Performance in realistic microwave sky simulations. *Astronomy & Astrophysics*, 658:A110.

Chluba, J., Hill, J. C., and Abitbol, M. H. (2017). Rethinking CMB foregrounds: systematic extension of foreground parametrizations. *Monthly Notices of the Royal Astronomical Society*, 472(1):1195–1213.

CMB-S4 Collaboration (2016). CMB-S4 Science Book, First Edition. *arXiv e-prints*, page arXiv:1610.02743.

Conklin, E. K. (1969). Velocity of the Earth with Respect to the Cosmic Background Radiation. *Nature*, 222(5197):971–972.

Coulton, W. R. and Spergel, D. N. (2019). The bispectrum of polarized galactic foregrounds. *Journal of Cosmology and Astroparticle Physics*, 2019(10):056.

Cun, Y. L., Boser, B., Denker, J. S., Howard, R. E., Habbard, W., Jackel, L. D., and Henderson, D. (1990). *Handwritten Digit Recognition with a Back-Propagation Network*, page 396–404. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Datta, R., Aiola, S., Choi, S. K., Devlin, M., Dunkley, J., Dünner, R., Gallardo, P. A., Gralla, M., Halpern, M., Hasselfield, M., Hilton, M., Hincks, A. D., Ho, S.-P. P., Hubmayr, J., Huffenberger, K. M., Hughes, J. P., Kosowsky, A., López-Caraballo, C. H., Louis, T., Lungu, M., Marriage, T., Maurin, L., McMahon, J., Moodley, K., Naess, S. K., Nati, F., Niemack, M. D., Page, L. A., Partridge, B., Prince, H., Staggs, S. T., Switzer, E. R., Wollack, E. J., and Farren, G. (2018). The Atacama Cosmology Telescope: two-season ACTPol extragalactic point sources and their polarization properties. *Monthly Notices of the Royal Astronomical Society*, 486(4):5239–5262.

de Bernardis, P., Ade, P. A. R., Bock, J. J., Bond, J. R., Borrill, J., Boscaleri, A., Coble, K., Crill, B. P., de Gasperis, G., de Troia, G., Farese, P. C., Ferreira, P. G., Ganga, K., Giacometti, M., Hivon, E., Hristov, V. V., Iacoangeli, A., Jaffe, A. H., Lange, A. E., Martinis, L., Masi, S., Mason, P., Mauskopf, P. D., Melchiorri, A., Miglio, L., Montroy, T., Netterfield, C. B., Pascale, E., Piacentini, F., Pogosyan, D., Pongetti, F., Prunet, S., Rao, S., Romeo, G., Ruhl, J. E., Scaramuzzi, F., Sforna, D., and Vittorio, N. (2001). First results from the BOOMERanG experiment. In Durrer, R., Garcia-Bellido, J., and Shaposhnikov, M., editors, *Cosmology and Particle Physics*, volume 555 of *American Institute of Physics Conference Series*, pages 85–94.

de Zotti, G., Ricci, R., Mesa, D., Silva, L., Mazzotta, P., Toffolatti, L., and González-Nuevo, J. (2005). Predictions for high-frequency radio surveys of extragalactic sources. *Astronomy & Astrophysics*, 431(3):893–903.

Delabrouille, J., Betoule, M., Melin, J. B., Miville-Deschênes, M. A., Gonzalez-Nuevo, J., Le Jeune, M., Castex, G., de Zotti, G., Basak, S., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A. J., Bernard, J. P., Bouchet, F. R., Clements, D. L., da Silva, A., Dickinson, C., Dodu, F., Dolag, K., Elsner, F., Fauvet, L., Faÿ, G., Giardino, G., Leach, S., Lesgourgues, J., Liguori, M., Macías-Pérez, J. F., Massardi, M., Matarrese, S., Mazzotta, P., Montier, L., Mottet, S., Paladini, R., Partridge, B., Piffaretti, R., Prezeau, G., Prunet, S., Ricciardi, S., Roman, M., Schaefer, B., and Toffolatti, L. (2013). The pre-launch Planck Sky Model: a model of sky emission at submillimetre to centimetre wavelengths. *Astronomy & Astrophysics*, 553:A96.

Delabrouille, J., Melin, J. B., and Bartlett, J. G. (2002). Simulations of Sunyaev-Zel'dovich Maps and Their Applications. In Chen, L.-W., Ma, C.-P., Ng, K.-W., and Pen, U.-L., editors, *AMiBA 2001: High-Z Clusters, Missing Baryons, and CMB Polarization*, volume 257 of *Astronomical Society of the Pacific Conference Series*, page 81.

Dickinson, C., Davies, R. D., and Davis, R. J. (2003). Towards a free-free template for CMB foregrounds. *Monthly Notices of the Royal Astronomical Society*, 341(2):369–384.

Draine, B. T. (2003). Interstellar Dust Grains. *Annual Review of Astronomy and Astrophysics*, 41:241–289.

Draine, B. T. (2011). *Physics of the Interstellar and Intergalactic Medium*.

Draine, B. T. and Lazarian, A. (1998). Electric Dipole Radiation from Spinning Dust Grains. *The Astrophysical Journal*, 508(1):157–179.

Durrer, R. (2008). *The Cosmic Microwave Background*.

Dvorkin, C., Mishra-Sharma, S., Nord, B., Villar, V. A., Avestruz, C., Bechtol, K., Ćiprijanović, A., Connolly, A. J., Garrison, L. H., Narayan, G., and Villaescusa-Navarro, F. (2022). Machine Learning and Cosmology. *arXiv e-prints*, page arXiv:2203.08056.

Eriksen, H. K., Jewell, J. B., Dickinson, C., Banday, A. J., Górski, K. M., and Lawrence, C. R. (2008). Joint Bayesian Component Separation and CMB Power Spectrum Estimation. *The Astrophysical Journal*, 676(1):10–32.

Fuskeland, U., Andersen, K. J., Aurlien, R., Banerji, R., Brilenkov, M., Eriksen, H. K., Galloway, M., Gjerløw, E., Næss, S. K., Svalheim, T. L., and Wehus, I. K. (2021). Constraints on the spectral index of polarized synchrotron emission from WMAP and Faraday-corrected S-PASS data. *Astronomy & Astrophysics*, 646:A69.

Galloway, M., Andersen, K. J., Aurlien, R., Banerji, R., Bersanelli, M., Bertocco, S., Brilenkov, M., Carbone, M., Colombo, L. P. L., Eriksen, H. K., Eskilt, J. R., Foss, M. K., Franceschet, C., Fuskeland, U., Galeotta, S., Gerakakis, S., Gjerløw, E., Hensley, B., Herman, D., Iacobellis, M., Ieronymaki, M., Ihle, H. T., Jewell, J. B., Karakci, A., Keihänen, E., Keskitalo, R., Maggio, G., Maino, D., Maris, M., Mennella, A., Paradiso, S., Partridge, B., Reinecke, M., San, M., Suur-Uski, A. S., Svalheim, T. L., Tavagnacco, D., Thommesen, H., Watts, D. J., Wehus, I. K., and Zacchei, A. (2023). BEYONDPLANCK. III. Commander3. *Astronomy & Astrophysics*, 675:A3.

Ginzburg, V. L. and Syrovatskii, S. I. (1965). Cosmic Magnetobremsstrahlung (synchrotron Radiation). *Annual Review of Astronomy and Astrophysics*, 3:297.

Ginzburg, V. L. and Syrovatskii, S. I. (1969). Developments in the Theory of Synchrotron Radiation and its Reabsorption. *Annual Review of Astronomy and Astrophysics*, 7:375.

González-Nuevo, J., Argüeso, F., López-Caniego, M., Toffolatti, L., Sanz, J. L., Vielva, P., and Herranz, D. (2006). The Mexican hat wavelet family: application to point-source detection in cosmic microwave background maps. *Monthly Notices of the Royal Astronomical Society*, 369(4):1603–1610.

González-Nuevo, J., Massardi, M., Argüeso, F., Herranz, D., Toffolatti, L., Sanz, J. L., López-Caniego, M., and de Zotti, G. (2008). Statistical properties of extragalactic sources in the New Extragalactic WMAP Point Source (NEWPS) catalogue. *Monthly Notices of the Royal Astronomical Society*, 384(2):711–718.

González-Nuevo, J., Toffolatti, L., and Argüeso, F. (2005). Predictions of the Angular Power Spectrum of Clustered Extragalactic Point Sources at Cosmic Microwave Background Frequencies from Flat and All-Sky Two-dimensional Simulations. *The Astrophysical Journal*, 621(1):1–14.

Goodfellow, I. J. (2010). Technical report: Multidimensional, downsampled convolution for autoencoders. Technical report, Université de Montréal.

Goodfellow, I. J., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press, Cambridge, MA, USA. `http://www.deeplearningbook.org`.

Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., and Bartelmann, M. (2005). HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *The Astrophysical Journal*, 622(2):759–771.

Guth, A. H. (1981). Inflationary universe: A possible solution to the horizon and flatness problems. *Pyisical Review D*, 23(2):347–356.

Halverson, N. W., Leitch, E. M., Pryke, C., Kovac, J., Carlstrom, J. E., Holzapfel, W. L., Dragovan, M., Cartwright, J. K., Mason, B. S., Padin, S., Pearson, T. J., Readhead, A. C. S., and Shepherd, M. C. (2002). Degree Angular Scale Interferometer First Results: A Measurement of the Cosmic Microwave Background Angular Power Spectrum. *The Astrophysical Journal*, 568(1):38–45.

Haslam, C. G. T., Klein, U., Salter, C. J., Stoffel, H., Wilson, W. E., Cleary, M. N., Cooke, D. J., and Thomasson, P. (1981). A 408 MHz all-sky continuum survey. I - Observations at southern declinations and for the North Polar region. *Astronomy & Astrophysics*, 100:209–219.

Haslam, C. G. T., Salter, C. J., Stoffel, H., and Wilson, W. E. (1982). A 408-MHZ All-Sky Continuum Survey. II. The Atlas of Contour Maps. *American Association of Pharmaceutical Scientists*, 47:1.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA.

Henry, P. S. (1971). Isotropy of the 3 K Background. *Nature*, 231(5304):516–518.

Hensley, B. S. and Draine, B. T. (2021). Observational Constraints on the Physical Properties of Interstellar Dust in the Post-Planck Era. *The Astrophysical Journal*, 906(2):73.

Herranz, D., Argüeso, F., Toffolatti, L., Manjón-García, A., and López-Caniego, M. (2021). A Bayesian method for point source polarisation estimation. *Astronomy & Astrophysics*, 651:A24.

Herranz, D., López-Caniego, M., Sanz, J. L., and González-Nuevo, J. (2009). A novel multifrequency technique for the detection of point sources in cosmic microwave background maps. *Monthly Notices of the Royal Astronomical Society*, 394(1):510–520.

Hinton, G. E. et al. (1986). Learning distributed representations of concepts. In *Proceedings of the eighth annual conference of the cognitive science society*, volume 1, page 12. Amherst, MA.

Hinton, G. E., Osindero, S., and Teh, Y.-W. (2006). A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, 18(7):1527–1554.

Hu, W. and White, M. (1997). A CMB polarization primer. *New Astronomy*, 2(4):323–344.

Kashlinsky, A., Arendt, R. G., Atrio-Barandela, F., Cappelluti, N., Ferrara, A., and Hasinger, G. (2018). Looking at cosmic near-infrared background radiation anisotropies. *Reviews of Modern Physics*, 90(2):025006.

Krachmalnicoff, N., Carretti, E., Baccigalupi, C., Bernardi, G., Brown, S., Gaensler, B. M., Haverkorn, M., Kesteven, M., Perrotta, F., Poppi, S., and Staveley-Smith, L. (2018). S-PASS view of polarized Galactic synchrotron at 2.3 GHz as a contaminant to CMB observations. *Astronomy & Astrophysics*, 618:A166.

Krachmalnicoff, N. and Puglisi, G. (2021). ForSE: A GAN-based Algorithm for Extending CMB Foreground Models to Subdegree Angular Scales. *The Astrophysical Journal*, 911(1):42.

Lapi, A., González-Nuevo, J., Fan, L., Bressan, A., De Zotti, G., Danese, L., Negrello, M., Dunne, L., Eales, S., Maddox, S., Auld, R., Baes, M., Bonfield, D. G., Buttiglione, S., Cava,

A., Clements, D. L., Cooray, A., Dariush, A., Dye, S., Fritz, J., Herranz, D., Hopwood, R., Ibar, E., Ivison, R., Jarvis, M. J., Kaviraj, S., López-Caniego, M., Massardi, M., Michałowski, M. J., Pascale, E., Pohlen, M., Rigby, E., Rodighiero, G., Serjeant, S., Smith, D. J. B., Temi, P., Wardlow, J., and van der Werf, P. (2011). Herschel-ATLAS Galaxy Counts and High-redshift Luminosity Functions: The Formation of Massive Early-type Galaxies. *The Astrophysical Journal*, 742(1):24.

Leach, S. M., Cardoso, J. F., Baccigalupi, C., Barreiro, R. B., Betoule, M., Bobin, J., Bonaldi, A., Delabrouille, J., de Zotti, G., Dickinson, C., Eriksen, H. K., González-Nuevo, J., Hansen, F. K., Herranz, D., Le Jeune, M., López-Caniego, M., Martínez-González, E., Massardi, M., Melin, J. B., Miville-Deschênes, M. A., Patanchon, G., Prunet, S., Ricciardi, S., Salerno, E., Sanz, J. L., Starck, J. L., Stivoli, F., Stolyarov, V., Stompor, R., and Vielva, P. (2008). Component separation methods for the PLANCK mission. *Astronomy & Astrophysics*, 491(2):597–615.

Linde, A. D. (1982). A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems. *Physics Letters B*, 108(6):389–393.

LiteBIRD Collaboration (2023a). Probing cosmic inflation with the LiteBIRD cosmic microwave background polarization survey. *Progress of Theoretical and Experimental Physics*, 2023(4):042F01.

LiteBIRD Collaboration (2023b). Tensor-to-scalar ratio forecasts for extended LiteBIRD frequency configurations. *Astronomy & Astrophysics*, 676:A42.

Long, J., Shelhamer, E., and Darrell, T. (2014). Fully Convolutional Networks for Semantic Segmentation. *arXiv e-prints*, page arXiv:1411.4038.

Mambrini, Y. (2021). *Particles in the Dark Universe; A Student's Guide to Particle Physics and Cosmology.*

Martínez-González, E., Diego, J. M., Vielva, P., and Silk, J. (2003). Cosmic microwave background power spectrum estimation and map reconstruction with the expectation-maximization algorithm. *Monthly Notices of the Royal Astronomical Society*, 345(4):1101–1109.

Mesa, D., Baccigalupi, C., De Zotti, G., Gregorini, L., Mack, K.-H., Vigotti, M., and Klein, U. (2002). Polarization properties of extragalactic radio sources and their contribution to microwave polarization fluctuations. *A&A*, 396(2):463–471.

Penzias, A. A. and Wilson, R. W. (1965). A Measurement of Excess Antenna Temperature at 4080 Mc/s. *The Astrophysical Journal*, 142:419–421.

Petroff, M. A., Addison, G. E., Bennett, C. L., and Weiland, J. L. (2020). Full-sky Cosmic Microwave Background Foreground Cleaning Using Machine Learning. *The Astrophysical Journal*, 903(2):104.

Planck Collaboration (2014a). Planck 2013 results. xxviii. the planck catalogue of compact sources. *A&A*, 571:A28.

Planck Collaboration (2014b). Planck intermediate results. XV. A study of anomalous microwave emission in Galactic clouds. *Astronomy & Astrophysics*, 565:A103.

Planck Collaboration (2014c). Planck intermediate results. XVII. Emission of dust in the diffuse interstellar medium from the far-infrared to microwave frequencies. *Astronomy & Astrophysics*, 566:A55.

Planck Collaboration (2015). Planck intermediate results. XXII. Frequency dependence of thermal emission from Galactic dust in intensity and polarization. *Astronomy & Astrophysics*, 576:A107.

Planck Collaboration (2016a). Planck 2015 results - xxvi. the second planck catalogue of compact sources. *A&A*, 594:A26.

Planck Collaboration (2016b). Planck 2015 results. X. Diffuse component separation: Foreground maps. *Astronomy & Astrophysics*, 594:A10.

Planck Collaboration (2016c). Planck intermediate results. XLVIII. Disentangling Galactic dust emission and cosmic infrared background anisotropies. *Astronomy & Astrophysics*, 596:A109.

Planck Collaboration (2016d). Planck intermediate results. XXIX. All-sky dust modelling with Planck, IRAS, and WISE observations. *Astronomy & Astrophysics*, 586:A132.

Planck Collaboration (2018). Planck intermediate results. LIV. The Planck multi-frequency catalogue of non-thermal sources. *Astronomy & Astrophysics*, 619:A94.

Planck Collaboration (2020a). Planck 2018 results. I. Overview and the cosmological legacy of Planck. *Astronomy & Astrophysics*, 641:A1.

Planck Collaboration (2020b). Planck 2018 results. IV. Diffuse component separation. *Astronomy & Astrophysics*, 641:A4.

Planck Collaboration (2020c). Planck 2018 results. V. CMB power spectra and likelihoods. *Astronomy & Astrophysics*, 641:A5.

Planck Collaboration (2020d). Planck 2018 results. VI. Cosmological parameters. *Astronomy & Astrophysics*, 641:A6.

Puglisi, G., Galluzzi, V., Bonavera, L., Gonzalez-Nuevo, J., Lapi, A., Massardi, M., Perrotta, F., Baccigalupi, C., Celotti, A., and Danese, L. (2018). Forecasting the contribution of polarized extragalactic radio sources in CMB observations. *The Astrophysical Journal*, 858(2):85.

Puglisi, G., Mihaylov, G., Panopoulou, G. V., Poletti, D., Errard, J., Puglisi, P. A., and Vianello, G. (2022). Improved galactic foreground removal for B-mode detection with clustering methods. *Monthly Notices of the Royal Astronomical Society*, 511(2):2052–2074.

Remazeilles, M., Delabrouille, J., and Cardoso, J.-F. (2011). Foreground component separation with generalized Internal Linear Combination. *Monthly Notices of the Royal Astronomical Society*, 418(1):467–476.

Remazeilles, M., Rotti, A., and Chluba, J. (2021). Peeling off foregrounds with the constrained moment ILC method to unveil primordial CMB B modes. *Monthly Notices of the Royal Astronomical Society*, 503(2):2478–2498.

Rodríguez, J.-V., Rodríguez-Rodríguez, I., and Woo, W. L. (2022). On the application of machine learning in astronomy and astrophysics: A text-mining-based scientometric analysis. *WIREs Data Mining and Knowledge Discovery*, 12(5):e1476.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323:533–536.

Saikia, D. J. and Salter, C. J. (1988). Polarization properties of extragalactic radio sources. *Annual Review of Astronomy and Astrophysics*, 26(1):93–144.

Sanger, T. and Baljekar, P. N. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65 6:386–408.

Schneider, P. (2015). *Extragalactic Astronomy and Cosmology: An Introduction*.

Smoot, G. F., Bennett, C. L., Kogut, A., Wright, E. L., Aymon, J., Boggess, N. W., Cheng, E. S., de Amici, G., Gulkis, S., Hauser, M. G., Hinshaw, G., Jackson, P. D., Janssen, M., Kaita, E., Kelsall, T., Keegstra, P., Lineweaver, C., Loewenstein, K., Lubin, P., Mather, J., Meyer, S. S., Moseley, S. H., Murdock, T., Rokke, L., Silverberg, R. F., Tenorio, L., Weiss, R., and Wilkinson, D. T. (1992). Structure in the COBE Differential Microwave Radiometer First-Year Maps. *The Astrophysical Journal Letters*, 396:L1.

Starobinsky, A. A. (1982). Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations. *Physics Letters B*, 117(3-4):175–178.

Sunyaev, R. A. and Zeldovich, Y. B. (1972). The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies. *Comments on Astrophysics and Space Physics*, 4:173.

Toffolatti, L., Argueso Gomez, F., de Zotti, G., Mazzei, P., Franceschini, A., Danese, L., and Burigana, C. (1998). Extragalactic source counts and contributions to the anisotropies of the cosmic microwave background: predictions for the Planck Surveyor mission. *Monthly Notices of the Royal Astronomical Society*, 297(1):117–127.

Tristram, M., Banday, A. J., Górski, K. M., Keskitalo, R., Lawrence, C. R., Andersen, K. J., Barreiro, R. B., Borrill, J., Colombo, L. P. L., Eriksen, H. K., Fernandez-Cobos, R., Kisner, T. S., Martínez-González, E., Partridge, B., Scott, D., Svalheim, T. L., and Wehus, I. K. (2022). Improved limits on the tensor-to-scalar ratio using BICEP and Planck data. *Pyisical Review D*, 105(8):083524.

Trombetti, T., Burigana, C., Zotti, G. D., Galluzzi, V., and Massardi, M. (2018). Average fractional polarization of extragalactic sources at planck frequencies. *Astronomy & Astrophysics*, 618:A29.

Tucci, M. and Toffolatti, L. (2012). The impact of polarized extragalactic radio sources on the detection of CMB anisotropies in polarization. *Advances in Astronomy*, 2012:1–17.

Tucci, M., Toffolatti, L., De Zotti, G., and Martínez-González, E. (2011). High-frequency predictions for number counts and spectral properties of extragalactic radio sources. new evidence of a break at mm wavelengths in spectra of bright blazar sources. *A&A*, 533:A57.

Vacher, L., Aumont, J., Montier, L., Azzoni, S., Boulanger, F., and Remazeilles, M. (2022). Moment expansion of polarized dust SED: A new path towards capturing the CMB B-modes with LiteBIRD. *Astronomy & Astrophysics*, 660:A111.

Vansyngel, F., Boulanger, F., Ghosh, T., Wandelt, B., Aumont, J., Bracco, A., Levrier, F., Martin, P. G., and Montier, L. (2017). Statistical simulations of the dust foreground to cosmic microwave background polarization. *Astronomy & Astrophysics*, 603:A62.

Weinberg, S. (2008). *Cosmology*.

Zaldarriaga, M. and Seljak, U. (1997). All-sky analysis of polarization in the microwave background. *Pyisical Review D*, 55(4):1830–1840.