Universidade de São Paulo
Instituto de Física

Estudos sobre a luz extragaláctica de fundo com raios gama por meio de métodos de Monte Carlo via cadeias de Markov

Matheus Genaro Dantas Xavier

Orientador: Prof. Dr. Edivaldo Moura Santos

Dissertação de mestrado apresentada ao Instituto de Física da Universidade de São Paulo, como requisito parcial para a obtenção do título de Mestre em Ciências.

Banca Examinadora:
Prof. Dr. Edivaldo Moura Santos - Orientador (IFUSP)
Prof. Dr. Luiz Vitor de Souza Filho (IFSC/USP)
Prof. Dr. Manuel Meyer (University of Southern Denmark)

São Paulo
2023

FICHA CATALOGRÁFICA
Preparada pelo Serviço de Biblioteca e Informação
do Instituto de Física da Universidade de São Paulo

Xavier, Matheus Genaro Dantas

Estudos sobre a luz extragaláctica de fundo com raios gama por meio de métodos de Monte Carlo via cadeias de Markov. São Paulo, 2023.

Dissertação (Mestrado) - Universidade de São Paulo. Instituto de Física. Depto. de Física Experimental.

Orientador: Prof. Dr. Edivaldo Moura Santos

Área de Concentração: Física.

Unitermos: 1. Radiação gama; 2. Núcleos ativos galácticos; 3. Astrofísica.

USP/IF/SBI-082/2023

University of São Paulo
Physics Institute

Extragalactic background light studies with gamma rays via Markov chain Monte Carlo methods

Matheus Genaro Dantas Xavier

Supervisor: Prof. Dr. Edivaldo Moura Santos

Dissertation submitted to the Physics Institute of the University of São Paulo in partial fulfillment of the requirements for the degree of Master of Science.

Examining Committee:
Prof. Dr. Edivaldo Moura Santos - Supervisor (IFUSP)
Prof. Dr. Luiz Vitor de Souza Filho (IFSC/USP)
Prof. Dr. Manuel Meyer (University of Southern Denmark)

São Paulo
2023

Aos meus pais.

Agradecimentos

Os primeiros agradecimentos vão ao meu orientador Edivaldo, não por simples decoro ou cortesia, mas pela certeza de seu papel fundamental em tornar a experiência do Mestrado muito recompensadora e abrir grandes oportunidades para minha carreira e vida. Quatro anos foram pouco para aprender tudo o que poderia sob sua orientação e convívio. É indescritível sua dedicação e comprometimento com seus alunos e sinto-me extremamente sortudo por fazer parte disso.

O trabalho também não teria sido possível sem as contribuições e apoio cruciais do Douglas Pimentel e Luiz Stuani, cujas discussões e participações nas várias reuniões semanais proporcionaram o contínuo aperfeiçoamento e andamento do projeto.

Agradeço à banca a atenção dada ao meu trabalho. Importantes conselhos do Prof. Vitor de Souza para minha carreira científica, além do apoio estrutural através dos projetos temáticos e por fomentar o contato com uma grande rede de cientistas brasileiros nesta área. Also, many thanks to Prof. Manuel Meyer for participating in my defense and making many interesting questions.

Aos meus colegas do Journal Club — minhas irmãs, irmãos (e primos) de orientador — agradeço o companheirismo, os desabafos, os encontros e risadas essenciais na trajetória de um estudante. Mesmo sendo breve o nosso convívio, sei que ainda nos encontraremos em futuros eventos profissionais (e não profissionais) daqui pra frente, sempre unidos em nossa identidade de grupo. Lembrarei desse período com vocês com muito carinho.

Finally, while I would love to express myself in German at this moment, I must keep it in English for the sake of clarity. I want to make a very special thanks to Prof. Stefan Funk for accepting me as his student during my stay at ECAP and allowing all this to be just the beginning. I also want to thank Rodrigo's guidance throughout the development of my project there, providing a good momentum to my research. Não posso deixar de registrar também meu obrigado pela recepção amigável e por facilitar minha adaptação ao exótico cotidiano alemão. My stay at ECAP was also made very special by the wonderful people I met there. It was a joy to spend some time with all of you and I am very excited to "see you later" (soon).

O presente trabalho teve apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), nº do processo 2021/01473-6 e apoio do CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil, processo 130534/2021-8. Também agradeço o vínculo com os Projetos Temáticos FAPESP 2015/15897-1 e 2021/01089-1. Por fim, parte da dissertação resulta do projeto desenvolvido com Bolsa Estágio de Pesquisa no Exterior FAPESP, nº 2022/07692-4. As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade do autor e não necessariamente refletem a visão da FAPESP.

O trabalho foi desenvolvido com recursos computacionais HPC da Superintedência

de Tecnologia da Informação da Universidade de São Paulo.

This research made use of ctools, a community-developed gamma-ray astronomy science analysis software. ctools is based on GammaLib, a community-developed toolbox for the scientific analysis of astronomical gamma-ray data.

NOËLLE, *watching the TV with rapt interest*: “GRANDMA, what is a Blazar?”

GRANDMA, *Busy with eating*: “A piece of clothing, honey.”

— “Santa’s Guide to the Galaxy” (Christmas Lecture 2022, ECAP)

Abstract

The Extragalactic Background Light (EBL) is the second most intense background radiation field in the universe, being the product of the integrated stellar emission and light reprocessed by dust throughout the history of structure formation and cosmic evolution. The precise spectral shape of the EBL is not completely known, as direct measurements are difficult to make due to dominant foregrounds. However, it is possible to probe the EBL indirectly using gamma rays, since, during their propagation over cosmological distances, very high energy (VHE) photons can interact with the EBL producing electron-positron pairs. This dissertation explores the use of Markov Chain Monte Carlo methods to obtain simultaneous constraints on the EBL and intrinsic spectral parameters of gamma-ray sources. Thus, the fundamental goal is to reconstruct the posterior probability density of parameters characterising the EBL and the intrinsic flux emission of active galactic nuclei (AGNs), in a Bayesian approach. The first part of the work is mainly concerned with validating the methodology with a sample of synthetic BL Lacs observed with the instrument configuration of the future Cherenkov Telescope Array (CTA). In this controlled scenario, we investigate the impacts on EBL constraints by progressively including more spectra in the likelihood function. We identify a consistent improvement in uncertainties by combining different sources in the analysis, while also being capable of recovering the spectral indices of all intrinsic spectra. We further explore the impacts of increasing the observation time of the sources and possible systematic effects associated to the choice of EBL modelling. In the second part, we analyse a sample of 65 real spectra from 36 AGN observed by various Imaging Atmospheric Cherenkov Telescopes, obtaining constraints on the EBL and intrinsic parameters that are consistent to other results found in the literature. We identify the Markarian 501 flare data as essential for constraining the far-Infrared part of the EBL, while the combination of all other sources provided robust constraints on the mid-Infrared. Such analysis was possible through the use of the Hamiltonian Monte Carlo method, which is very efficient in a parameter space with a high number of dimensions. Two other extensions are also explored. With the synthetic sample, we discuss the possibility of constraining the Hubble constant while probing the EBL. We also present a Bayesian method for signal estimation in On/Off measurements and perform a preliminary analysis based on H.E.S.S. data. Such method allows improved signal estimation without performing selection cuts on data, which could be useful for improving VHE measurements and detection of faint sources.

Keywords: Extragalactic Background Light; gamma rays; Hamiltonian Monte Carlo; Bayesian analysis; signal estimation.

Resumo

A Luz Extragaláctica de Fundo (EBL) é o segundo campo de radiação de fundo mais intenso no universo, sendo o resultado da emissão estelar e luz reprocessada por poeira integradas ao longo da história de evolução e formação de estruturas. A forma espectral exata da EBL não é completamente conhecida, dado que medidas diretas são difíceis de realizar devido a *foregrounds* dominantes. No entanto, é possível vincular a EBL indiretamente utilizando raios gama, pois, durante sua propagação por distâncias cosmológicas, fótons de energias muito altas (VHE) podem interagir com a EBL, produzindo pares elétron-pósitron. Esta dissertação explora o uso de métodos de Monte Carlo via Cadeias de Markov (MCMC) para obter simultaneamente vínculos sobre a EBL e parâmetros espectrais intrínsecos de fontes de raios gama. Assim, o objetivo fundamental consiste em reconstruir a distribuição posterior de probabilidade de parâmetros que caracterizam a EBL e o fluxo intrínseco de emissão de núcleos ativos de galáxias (AGNs), em uma abordagem Bayesiana. A primeira parte do trabalho está focada em validar a metodologia com uma amostra de BL Lacs sintéticos observada com a configuração instrumental do futuro Cherenkov Telescope Array (CTA). Neste cenário controlado, nós investigamos os impactos sobre vínculos da EBL através da inclusão progressiva de mais espectros na função de verossimilhança. Identificamos melhorias consistentes nas incertezas ao combinar fontes distintas na análise, concomitantemente recuperando os índices espectrais de todos os espectros intrínsecos. Além disso, exploramos os impactos do aumento do tempo de observação das fontes e possíveis erros sistemáticos associados à escolha de modelagem da EBL. Na segunda parte, analisamos uma amostra de 65 espectros reais de 36 AGNs observados por diversos Telescópios Cherenkov de Imageamento Atmosférico, obtendo vínculos sobre a EBL e parâmetros intrínsecos que são consistentes com outros resultados encontrados na literatura. Nós identificamos os dados de *flare* de Markarian 501 como essenciais para vincular a região infravermelha distante da EBL, enquanto a combinação de todas as fontes restantes gerou vínculos robustos sobre o infravermelho médio. Esta análise foi possível pelo uso do método de Monte Carlo Hamiltoniano, que é muito eficiente para espaços de parâmetros com um número grande de dimensões. Duas outras extensões também foram exploradas. Com as amostras sintéticas, discutimos a possibilidade de vincular a constante de Hubble junto da EBL. Também apresentamos um método Bayesiano para a estimativa de sinal em medidas On/Off e realizamos uma análise preliminar baseada em dados do H.E.S.S. Este método permite aperfeiçoar a estimativa de sinal sem realizar cortes de seleção nos dados, o que pode ser útil para aprimorar medidas de VHE e a detecção de fontes fracas.

Palavras-chave: Luz Extragaláctica de Fundo; raios gama; Monte Carlo Hamiltoniano; análise Bayesiana; estimação de sinal.

Acronyms

AGN Active Galactic Nucleus.

CGRH Cosmic Gamma ray Horizon.

CMB Cosmic Microwave Background.

CTA Cherenkov Telescope Array.

EBL Extragalactic Background Light.

ESS Effective Sample Size.

HMC Hamiltonian Monte Carlo.

IACT Imaging Atmospheric Cherenkov Telescope.

IMF Initial Mass Function.

IR Infrared.

KSP Key Science Project.

LG Large Grains.

LIV Lorentz Invariance Violation.

LP Log-parabola.

MCMC Markov Chain Monte Carlo.

PAH Polycyclic Aromatic Hydrocarbons.

PL Power Law.

PLC Power Law with exponential Cut-off.

QED Quantum Electrodynamics.

SED Spectral Energy Distribution.

SFR Star Formation Rate.

SG Small Grains.

SNR Signal-to-Noise Ratio.

SSC Synchrotron Self-Compton.

UV Ultraviolet.

VHE Very High Energy.

Contents

Chapter 1

Introduction

The background electromagnetic content of the Universe — from radio waves to gamma rays — encapsulates a profound history of structure formation at various scales. It emerges from a large range of different physical processes, such as thermal emission from stars and dust inside galaxies; radiation produced in the vicinity of supermassive black holes; or even the relics of a much denser and much hotter Universe of the past. Furthermore, the wide wavelength interval requires a variety of unique instruments and techniques to produce measurements of its intensity and spectral energy distribution. The main challenge comes essentially from its definition: as background, its origin is extragalactic and approximately isotropic, being usually sub-dominant in relation to other astrophysical sources. Therefore, any successful measurement needs the correct identification (and subtraction) of foregrounds.

The knowledge on the background radiation is not only key for describing how the universe evolves, but also how particles propagate through it. Energetic particles, such as charged cosmic rays and gamma rays, may interact with low energy photons during propagation, transferring their energy or producing new particles in the process. These effects need to be modelled and accounted when interpreting observations of distant sources. For instance, very high energy gamma rays can interact with optical and infrared light via pair production, resulting in an effective absorption of this radiation in the intergalactic medium. Therefore, the suppression of the emitted gamma-ray flux from propagation must be considered when distinguishing between observed and intrinsic properties of extragalactic sources.

At the same time, it is possible to use some assumptions about the intrinsic emission to set limits on the extragalactic absorption. This indirect method of probing the background radiation content of the universe is the main subject of this dissertation. Here, we are concerned with the Extralagactic Background Light (EBL), which is the byproduct of star formation, active galactic emission and radiation reprocessed by dust, in an expanding universe. Its light, from the near ultraviolet to the far infrared, is mainly responsible for the attenuation of very high energy photons, at the TeV energy range.

In turn, these gamma rays are generally produced in active galactic nuclei (AGNs) at hundreds of megaparsec away and can be detected by ground telescopes of the Cherenkov type. From the flux data measured in these observatories, our goal is to simultaneously recover EBL and intrinsic spectral parameters of AGNs, through a Bayesian approach and the use of Markov Chain Monte Carlo (MCMC) methods. We do this by combining multiple sources in the likelihood, so we can investigate how this aggregate information improves the constraints on the EBL. As the precise shape of the EBL and the intrinsic emission of extragalactic sources is not known a priori, by sampling the posterior probability distribution of all unknown variables and marginalising it over a subset of parameters, we can easily incorporate the statistical uncertainties of the EBL into the intrinsic spectral parameters of the sources and vice versa.

The structure of the dissertation is as follows: in Chapter 2, we briefly present general aspects of the background radiation, before discussing in more details specifically the EBL (direct measurements and models). We do a short review on works that constrain the EBL through gamma-ray observations and also discuss possible routes of probing cosmological parameters in a similar way. In Chapter 3, we describe the methodology and tools used throughout the work. Starting with the adopted EBL model, we proceed by discussing the kinematics of the pair production interaction (the Breit-Wheeler process) and arrive at the optical depth expression, which quantifies the attenuation of gamma rays of a given energy and redshift of emission. Then, we construct the likelihood and posterior distribution of the target parameters, concluding with the presentation of the MCMC tools.

In Chapter 4, we start the analysis by creating a sample of synthetic sources and spectra simulated according to the expected response of the CTA ideal configuration. With this sample, we study EBL constraints by progressively adding more spectra in the likelihood, as well as identifying the impact of increasing the observation time of each source. We also study how changes in the "true" EBL opacity affect the inference on all parameters. Continuing with the analysis of the synthetic sources, in Chapter 5, we investigate the possibility of constraining the Hubble constant, by exploring the dependence of the optical depth with this cosmological parameter. In Chapter 6, we move to analysing a set of real flux measurements collected from the literature. With a sample of 65 spectra from 36 different AGN, we simultaneously constrain EBL parameters and the intrinsic gamma ray emission of all of these sources. These results are the culmination of all that we have learned from the previous chapters and they consist on the main contributions of this dissertation. Finally, in Chapter 7, we present a complementary Bayesian approach for signal estimation in gamma-ray astronomy. Although not yet used for probing the EBL, it has the potential of improving measurements at the highest energies or for weak sources, which can be beneficial for constraining the extragalactic opacity. This last chapter is the result of a five months internship in the Erlangen Centre

of Astroparticle Physics (ECAP), in Germany, an opportunity that directly and indirectly contributed to the overall quality of this work. In Chapter 8 we revise the main results and draw the final conclusions.

Chapter 2

Background Radiation and Gamma Rays

In this chapter, we start in section 2.1 with a brief introduction on the different components of the extragalactic background radiation, followed by a detailed discussion on the Extragalactic Background Light in section 2.2. Then, we present the connection between this radiation field and gamma rays in section 2.3. This chapter concludes mentioning the role of ground-based Cherenkov telescopes for studying these themes.

2.1 Why is the night sky dark?

A curious observer may look at the night sky and wonder: why is there so much darkness among a sea of virtually infinite stars that are constantly shining? Such poetic question was more precisely formulated by past astronomers and it is modernly referred as the Olbers' paradox, after the 18th–19th century German astronomer Heinrich Wilhelm Olbers. The argument goes as follows: imagine a static infinite flat universe, uniformly populated with stars. Under this assumption, no matter in what direction you look, your line of sight will eventually intercept a star. That star with a given luminosity L produces light that reaches us, at a distance r, with a flux $f = L/(4\pi r^2)$. Furthermore, the angular size of the same star with radius R will be $\Omega = \pi R^2/r^2$, so the observed surface brightness $f/\Omega = L/(4\pi^2 R^2)$ is independent of the distance. To have a crude estimation of its value, we can adopt a blackbody approximation for the stellar emission, meaning that $f/\Omega = \sigma T^4/\pi$, where $\sigma = 5.6703 \times 10^{-8}\,\mathrm{W\,m^{-2}\,K^{-4}}$ is the Stefan-Boltzmann constant and T the effective temperature of the star. At the order of magnitude of $T = 5 \times 10^3\,\mathrm{K}$, similar to the Sun, we find $f/\Omega \approx 9.025 \times 10^4\,\mathrm{W\,m^{-2}\,sr^{-1}}$. Such brightness is a factor $\sim 10^{13}$ larger than the observed optical and infrared (IR) background (e.g. Hauser & Dwek 2001), hence the paradox.

The description above is pedagogical and not historical, but it captures the essence of this discussion: if it is not bright, what makes the night sky dark? Naturally, the way this question is formulated changes as our perception of the universe is also transformed,

and so changes its answer. During Olbers' epoch, stars, planets and "nebulae" were the only observed constituents of the cosmos. It is in the 20th century that the extragalactic nature of many of these "nebulae" was realised, as our Galaxy became only one of many others. With Edwin Hubble, there was observational evidence that the universe was expanding, while General Relativity provided a powerful explanation and tools for modelling this dynamic. In this context, many of the assumptions to formulate the paradox were put in question, while during the last century actual measurements of the background radiation (not only in the visible light) were performed with the advance of various techniques and telescopes. This lead to other important questions, such as the origins of the various types of background radiation; the role of the expansion of the universe in explaining the observed intensities; or how the structure formation history and opacity, inside and outside the galaxies, relate to the brightness of the night sky.

The more modern discussions on the Olbers' paradox started around 1948, when the steady-state cosmological model — as an alternative to the Big Bang — was proposed by Bondi & Gold (1948). As Wesson (1989) presents, H. Bondi had a particular interest in investigating Olbers' paradox, with a discussion on his book on Cosmology in 1952. From that, most of the solutions to the Olbers' paradox accepted the movement of the galaxies and the expansion of the Universe as the main cause of the low levels of background light. The opposed view was notably represented by Harrison (1964), where it is argued that the finite age of galaxies is the dominant reason. The debate and some confusion persisted for more than two decades, as Wesson et al. (1987) and Wesson (1991) identified many textbooks published during the 1970s and 1980s with opposite views, heavily criticising those where the expansion of the universe is appointed as the solely factor. The presentation by Wesson (1991) puts in more straightforward and quantitative terms the relative importance of both factors. It is shown (Wesson et al. 1987) that bolometric computations (i.e. integrated over wavelengths) of the background light are reduced only by a factor of ~ 2 in a expanding universe compared to a static one, while the age of galaxies determines the order of magnitude of the intensity. For spectral computations (wavelength dependent), Wesson (1991) identifies a slightly increased importance of the expansion, but the conclusion remains the same.

Another factor contributing to the low background radiation, which is generally neglected in previous discussions, is the role of extinction due to interstellar dust. This decreases the brightness of galaxies and produces obscuration of more distant galaxies from nearby ones (Vavryčuk 2016). Although some authors, such as Knutsen (1997), argue that it has a marginal relevance, Vavryčuk (2016) shows that even in a static infinite universe, it is possible to achieve an intensity in the optical band around only 5 times larger than the observed value. A result obtained by considering observed mean values of the number density of galaxies, the mean free path between them, opacity of galaxies and the intergalactic absorption.

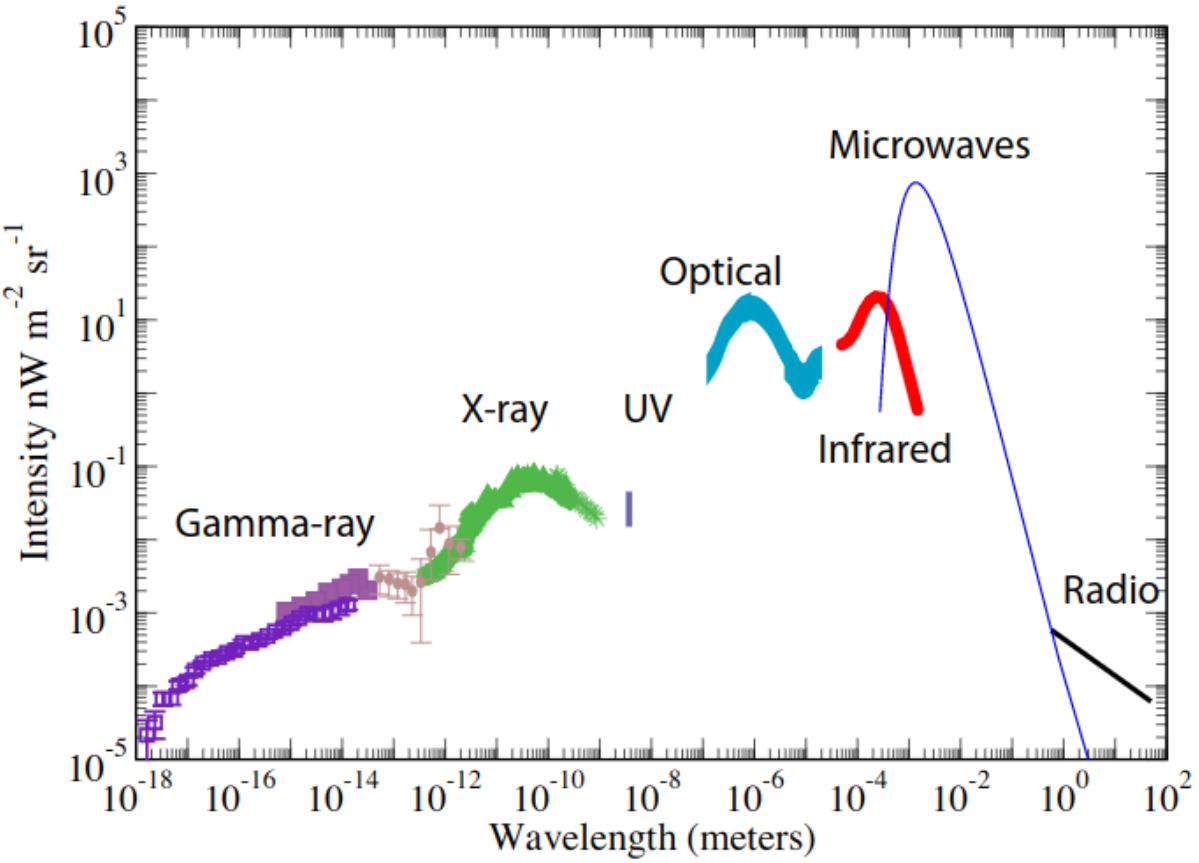

Figure 2.1: Measured intensity of the extragalactic background radiation from gamma rays to radio. Extracted from figure 1 of Cooray (2016), where the exact references of the data points can be found.

Today, we have a much deeper picture of the background radiation content of the universe, and while the reasons for the low intensity of the visible light are essentially understood, a dark night sky does not mean an empty one. Direct and indirect measurements of the diffuse extragalactic radiation span a wide range of wavelengths, from gamma rays to the radio. Figure 2.1 compiles some of these results, revealing a complex environment of distinct physical origins. The most intense radiation field is the Cosmic Microwave Background (CMB), represented as the thin blue line. The CMB is a relic of the early universe ($z \sim 1090$, around 380 thousand years after the Big Bang), being the radiation emerging from the surface of last scattering, after the decoupling of matter and radiation, both previously at thermal equilibrium. Due to cosmological redshift, its peak is observed today in the microwave band. Measurements from the Wilkinson Microwave Anisotropy Probe (WMAP) in combination with the Far-InfraRed Absolute Spectrophotometer (FIRAS) were analysed by Fixsen (2009) to reveal a spectrum very accurately described by Plank's law for a blackbody, with an average temperature of $T_0 = 2.7260(13)\,\mathrm{K}$. More recent and precise measurements from the Planck satellite (Planck Collaboration et al. 2020a) show temperature fluctuations of the order of only

3×10^{-4} K, revealing an impressive isotropy and homogeneity. Such small anisotropies, however, carry a lot of information, as they are associated with perturbations in matter and energy distribution from which all structures of the universe (galaxies, clusters) eventually formed. Therefore, the study on CMB anisotropies allows constraints on many cosmological parameters, such as the Huble constant, density parameters, the acoustic scale, fluctuation parameters and various others (see Planck Collaboration et al. 2020b).

At the lowest energies, the cosmic radio background is formed by the integrated emission of star-forming and radio galaxies (Protheroe & Biermann 1996). In the former, electrons in the ionised interstellar medium undergo free-free emission and high energy electrons produce synchrotron radiation in magnetic fields, while in the latter radio jets are produced in Active Galactic Nuclei (AGNs), as well as synchrotron radiation created in the magnetised relativistic plasma at the end of these jets (lobes), which interact with the extragalactic medium (Niţu et al. 2021). The observational challenge of measuring the radio background comes from the opacity and distortions caused by the ionosphere and the necessity of subtracting the foreground emission from our Galaxy (Niţu et al. 2021). Direct measurements, however, are possible by balloon-borne telescopes, such as the ARCADE-2 (Absolute Radiometer for Cosmology, Astrophysics and Diffuse Emission), constraining the background in the range of 3–90 GHz (Fixsen et al. 2011). Furthermore, the knowledge on the cosmic radio background intensity is essential for studying the propagation of ultra-high-energy photons above $\sim 10^{19}$ eV, as this radiation field becomes the dominant source of opacity due to pair production from photon-photon interaction (De Angelis et al. 2013).

At the highest energies, we find the Extragalactic Gamma-ray Background, largely probed by the Large Area Telescope (LAT) on the Fermi Gamma-ray Space Telescope (*Fermi*). *Fermi*–LAT (Atwood et al. 2009) is one of the most important instruments for gamma-ray astronomy, as its large field of view, with good angular resolution, operating in sky-survey mode for many years resulted in a detailed mapping of the gamma-ray sky, particularly in the range 10–300 GeV, where it has the highest sensitivity. Besides discovering many galactic and extragalactic sources, *Fermi*–LAT measured the diffuse gamma-ray emission between 100 MeV ($\lambda \approx 1.24 \times 10^{-14}$ m) and 820 GeV ($\lambda \approx 1.5 \times 10^{-18}$ m), describing it as a power law of spectral index $\Gamma = 2.32(2)$ with an exponential cutoff at a break energy of 279(52) GeV (Ackermann et al. 2015). This diffuse background is the cumulative result of unresolved sources, such as AGNs, star-forming galaxies and gamma-ray bursts. At energies above 100 GeV, another diffuse contribution comes from ultra-high energy cosmic rays (UHECR) and gamma rays interacting with low energy photons. The former can produce neutral pions, which decay into photons, while the latter produce electron-positron pairs which generate secondary electromagnetic cascades in magnetic fields. As discussed by Ackermann et al. (2015), some of the challenges of performing these measurements come from modelling the diffuse Galactic emission, which is the main foreground, resulting from the interaction of cosmic rays with the interstellar

gas and radiation fields.

The high energy background emission, as seen in Figure 2.1 continue in the X-Ray band, but encounters a significant gap in the ultraviolet (UV). The main difficult for observations below 100 nm comes from the absorption of these extreme UV photons by neutral hydrogen in the Galaxy and in the intergalactic medium. Below the ionisation energy of hydrogen ($\sim$13.6 eV), corresponding to a wavelength of ~ 0.1 µm, direct measurements contain large statistical and systematic uncertainties, according to Cooray (2016). However, the rough shape of the diffuse UV emission from that point on through the visible wavelengths is approximately understood, due to a combination of distinct measurement techniques (see section 2.2). A near-IR peak around 1 µm can be identified (the "optical" blue band in Figure 2.1) and also a secondary peak in the far-IR. Following Hauser & Dwek (2001) and Dwek & Krennrich (2013), we define this isotropic component between 0.1–1000 µm, excluding the CMB, as the Extragalactic Background Light (EBL).

The EBL constitutes the second most intense background radiation field and it will be the main topic of discussion throughout this work. As it will become clear in section 2.3, the knowledge on EBL' spectrum is fundamentally important for extragalactic gamma-ray astronomy, while the study of very energetic gamma rays can also contribute to indirectly probe the EBL intensity and its evolution. Before tackling into these questions, we will take a closer look at EBL's origin, measurements and current models.

2.2 The Extragalactic Background Light

The EBL is a residual of structure formation in an expanding universe (Hauser & Dwek 2001). Since the first stars were born and the first galaxies were formed, light — mostly from thermal processes — permeates the intergalactic medium and is redshifted by cosmological expansion. The first peak of the EBL was essentially formed by stellar activity, but it also has a contribution from AGN emission (Dwek & Krennrich 2013) impacting the mid-IR (here defined as the range of 5–50 µm). Since AGNs are strong X-ray emitters and believed to be surrounded by optically thick dust, their environment is heated by this energetic radiation and consequently produces significant IR thermal emission. According to Treister et al. (2006), this process can contribute from 3% to 8% of the EBL emission between 3–8 µm, while reaching 15% of the intensity at 24 µm. The most relevant contribution to the mid- to far-IR (> 50 µm), however, comes from various types of dust present in the interstellar medium of galaxies, as they absorb part of the stellar radiation and re-emit at longer wavelengths. This is the main mechanism for forming the second EBL peak around 100–200 µm.

2.2.1 EBL measurements

Direct measurements of the EBL are difficult to make due to its subdominant intensity in comparison to foregrounds. Besides atmospheric IR emission, there are sources in the Solar system and in the Galaxy that must be subtracted in order to correctly evaluate the extragalactic background. One of these sources is the zodiacal light, which results from the scattering and emission of light by interplanetary dust, and is the largest contribution of foreground in the range of 1.25–140 µm (Hauser & Dwek 2001). Other sources come from the integrated starlight from unresolved stars, obtained from star count models, which can be several times larger than the EBL (Mattila 2006) and are more substantial in the range 1.25–3.5 µm (Hauser & Dwek 2001). Finally, the diffuse galactic light, resulting from interstellar dust, is relevant above 60 µm. Naturally, the CMB must also be distinguished, specially above 400 µm, where it becomes the dominant background field. The incorrect subtraction of foregrounds may lead to wrong conclusions about the EBL intensity and consequently on the overall luminosity density of the universe. As discussed by Mattila (2006), an inadequate modelling of the zodiacal light may have lead some authors (e.g. Bernstein et al. 2005) to measure a discontinuity of the EBL around 1 µm, which could have indicated a signature of Population III stars (the first generation of stars, with the lowest metallicity), never before detected.

The first reliable direct measurements came from the operation of the Cosmic Background Explorer (COBE) satellite (Boggess et al. 1992), carrying the Diffuse Infrared Background Experiment (DIRBE) and Far Infrared Absolute Spectrophotometer (FIRAS) instruments. Initial measurements by DIRBE had uncertainties dominated by systematic effects associated with the foreground determination (Hauser & Dwek 2001), however, it was still possible to affirm the detection of an isotropic component exceeding the foregrounds in the 140 µm and 240 µm wavelengths, while obtaining upper limits from 1.25–100 µm (see Hauser et al. 1998 and Dwek et al. 1998). In the near-IR, the analysis performed by Matsumoto et al. (2005) on data from the Infrared Telescope in Space (IRTS) produced additional limits on the EBL between 1.4 to 4 µm, complementing previous results from COBE.

In the UV and optical bands, the same foregrounds from zodiacal light and starlight impact the measurements. Upper limits were reported since the 1970s (Hauser & Dwek 2001), but with the Hubble Space Telescope (HST), newer attempts of EBL constraints were made (e.g. Bernstein et al. 2002, later revised in Bernstein 2007). For instance, Bernstein et al. (2002) take photometric and low resolution measurements from the HST to measure the total mean flux of night sky, subtracting an estimate of the zodiacal light obtained from a surface spectrophotometer and modelling the diffuse galactic light. However, as HST was not designed for absolute photometry, the measurements were subject to large uncertainties (Cooray 2016). Other experiments that have also contributed

in providing measurements of the EBL in the UV were the Galaxy Evolution Explorer (GALEX), which resulted in a measurement of the UV background around $\sim 0.150\,\mu m$ (Murthy et al. 2010), and data from the Voyager mission, resulting in constraints at $\sim 0.110\,\mu m$ (Murthy et al. 1999) and $0.100\,\mu m$ (Edelstein et al. 2000).

Dwek & Krennrich (2013) provide a comprehensive table of many of these measurements and limits, listing the references within. To give some examples of more recent results, we mention Matsuura et al. (2017), who analysed data from the rocket experiment CIBER (Cosmic Infrared Background Experiment) to put constraints on the EBL range between 0.8–1.70 µm, using models to derive the foreground. Another interesting approach comes from analysing data of the cruising phase of the New Horizons probe, which in 2015 had its closest approach to Pluto. Zemcov et al. (2017) were able to obtain 2σ upper limits on the optical range between 0.4–0.8 µm. A few years later, Lauer et al. (2021) refined these results by using extended data from New Horizons, as the probe travels through the Kuiper Belt (~ 40 au). In fact, at distances greater than ~ 10 au from the Sun, the emission from interplanetary dust is negligible (Lauer et al. 2021), suppressing one of the most important foreground components.

Complementing direct measurements, indirect methods can also be used to set limits to the EBL. One of these approaches is to use direct background data to analyse anisotropy fluctuations in the diffuse radiation, as performed by Kashlinsky et al. (1996) with DIRBE data. Since the EBL originates from discrete sources, correlations in the fluctuation distribution can give information about the number of sources and help to constrain the EBL level from models (Hauser & Dwek 2001). It is also possible to probe lower limits on the EBL from measurements of galaxy counts (e.g. Madau & Pozzetti 2000). As not all existing galaxies have been observed, the integrated emission from identified objects should give the minimum intensity of the background radiation compatible to the data. Such analysis was made possible by deep field data coming from space instruments (such as GALEX and HST), or large ground-based observatories (Driver et al. 2016). In the mid- and far-IR, number counts measured by Spitzer (Werner et al. 2004) and Herschel (Pilbratt et al. 2010) space telescopes can also be used to set limits on the EBL at these wavelength bands. One example comes from Béthermin et al. (2012), who were able to estimate the intensity at 250 µm, 350 µm and 500 µm using Herschel data, as also deriving the integrated intensity between 8–1000 µm. Another example is the work by Driver et al. (2016), who used optical and UV data also to extended the limits throughout the whole EBL wavelength range.

2.2.2 EBL models

The interpretation of EBL data can be done with a more detailed modelling of the EBL's spectral energy distribution (SED) and its evolution in redshift. In the literature

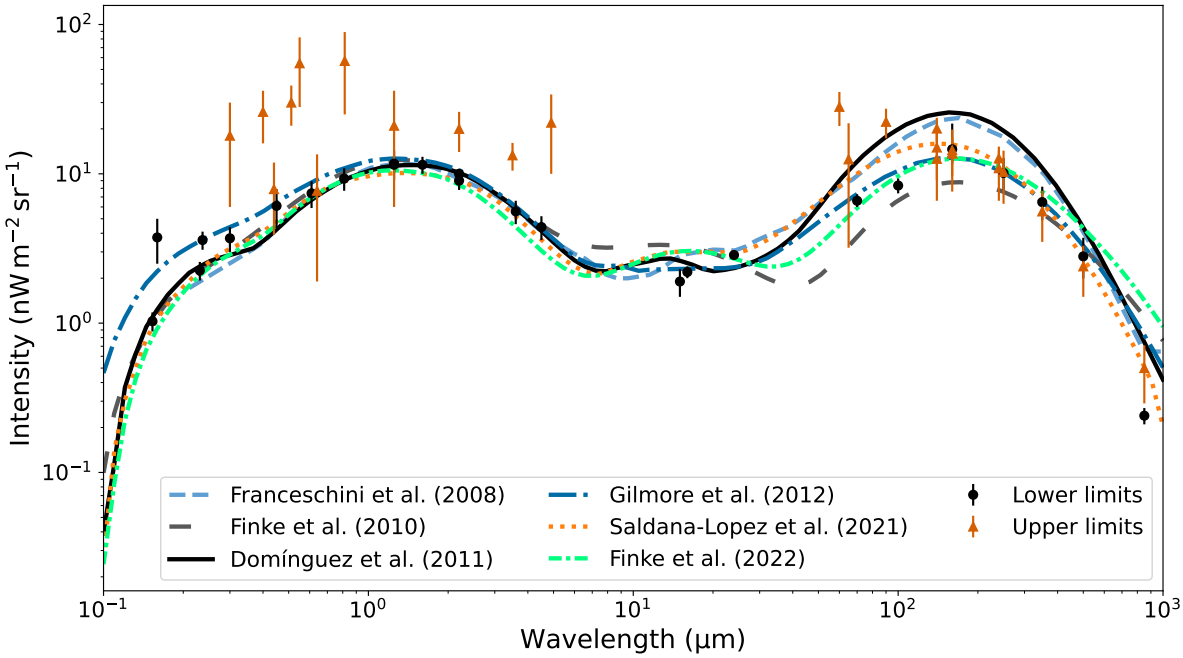

Figure 2.2: Local ($z = 0$) EBL intensity as described by different models. Curves were created from the `ebltable` package (Meyer 2022). Lower and upper limits on the spectrum, extracted from Biteau & Williams (2015), are also presented.

(e.g. Hauser & Dwek 2001, Kashlinsky 2005, Dwek & Krennrich 2013) it is common to classify the existing models in a few categories. Here we express three of the most common ones: backward evolution, forward evolution and semi-analytical models. For illustration purposes, we show in Figure 2.2 some of the more recent and frequently used ones. All the curves correspond to the local ($z = 0$) EBL intensity. Upper and lower limits were taken from Biteau & Williams (2015), which is a selection largely based on Dwek & Krennrich (2013).

Backward evolution models start with observed local characteristics of galaxy populations and then compute their evolution backwards in time, which can be compared to observed galaxy counts (Dwek & Krennrich 2013). Therefore, the key ingredient in this approach is to get an appropriate functional form of the galaxy luminosity function (number density of galaxies in a given luminosity interval) and then propose a behaviour for its redshift evolution. Then, one needs to integrate the spectral luminosity density over the redshifts (up to the start of galaxy formation) to obtain the resulting EBL. The simplest case is to assume that the comoving number density and the spectral energy distribution (SED) of galaxies do not evolve in time (Hauser & Dwek 2001). However, as mentioned by Dwek & Krennrich (2013), Spitzer and Herschel measurements revealed a significant evolution in the number of counts. One example of backward evolution model is that of Franceschini et al. (2008). In their work, they use new UV, optical and IR measurements, recently obtained at that time, to improve the characterisation of the luminosity evolu-

tion. Furthermore, they use independent models to describe the UV to near-IR evolution and the far-IR region, as the astrophysical processes that produced these emissions have different origins (direct stellar light versus dust emission) and distinct redshift evolution.

Similar to the backward evolution approach (and classified as such by Dwek & Krennrich 2013), the model by Domínguez et al. (2011) uses observations of galaxy counts and a luminosity function, but takes a more empirical approach, as the galaxy evolution represented in the luminosity function was directly observed up to $z = 4$ in the K band (near-IR). This is combined with a large sample of galaxies from AEGIS (All-Wavelength Extended Groth Strip International Survey), statistically attaching an SED type to each galaxy, from which the luminosity density can be computed.

Forward evolution models begin at the star formation origins and evolve the stellar populations to the present day, computing the properties of galaxies in the process. The parameters of the models can then be adjusted, for instance, to observed metallicities, galaxy number counts and SEDs of galaxies (Hauser & Dwek 2001). While backward evolution approaches usually arrive at the luminosity density from integrating the galaxy luminosity function, forward evolution models are more focused on the cosmic star formation rate and its redshift evolution to arrive at luminosity density of the EBL (Dwek & Krennrich 2013). One example of such models is Finke et al. (2010) — which we are going to discuss in detail in Chapter 3. An extended and updated model was recently published (Finke et al. 2022) and a quick discussion and comparison is presented in Chapter 6.

This classification, however, is not absolute, as Domínguez et al. (2011) uses a different third category to describe Finke et al. (2010)'s approach, since their inference on the galaxy evolution is anchored in the star formation rate, based on observations, leaving the forward evolution label to more semi-analytical methods. The semi-analytical approach takes cosmological parameters, usually in the ΛCDM framework, to set initial conditions and follow the formation of galaxies from dark matter halos (Dwek & Krennrich 2013), incorporating the physical processes of the gas cooling, formation of stars and feedback mechanisms (Hauser & Dwek 2001). One of the models come from Gilmore et al. (2012), which takes cosmological parameters from the 5 years results of WMAP and constructs semi-analytical models based on Monte Carlo simulations of dark matter halo merges. The physical processes taken in consideration are the accretion of gas and subsequent cooling to initiate star formation, while feedback from supernovae heats the environment and can eject cold gas out of the galaxies; besides, mergers can initiate AGNs by driving gas to the central regions, which can also interrupt star formation. Parallel to this, the chemical evolution from stellar populations and the history of star formation rate are used to compute the resulting emission spectrum. The absorption and emission from dust is also modelled. Two general components are considered (one being dense dust in giant molecular clouds and the other the diffuse distribution in the interstellar medium) and the reemission in the IR is computed from templates describing the spectra

of galaxies.

As seen in Figure 2.2, there is a general agreement between the different models, specially in the UV to near-IR peak, and they are close to the lower limits coming from galaxy counts. At the mid- and far-IR we see larger differences and even some apparent incompatibility to direct measurements (upper limits), but there is also a significant uncertainty in all models and possible systematic effects in the data, due to the observational challenges previously discussed. Nonetheless, the overall scenario for the local EBL spectrum is of convergence. However, as pointed by Saldana-Lopez et al. (2020), the general agreement holds only up to $z < 1$, as order of magnitude divergences in the UV start to appear at higher redshifts and a factor of 5 in difference in the IR. The evolution of the EBL spectrum is still very uncertain, but efforts have been made to improve constraints at higher redshifts. The work from Saldana-Lopez et al. (2020) aims at contributing to this by constructing the SEDs of a very large sample of galaxies from the CANDELS survey (a HST observation program), combining with other IR measurements, to stack the data with corrections and extract the evolving EBL. Their result is compatible with luminosity density determinations, galaxy counts and lower bounds, while giving more insight into the redshift evolution of the EBL.

2.3 EBL and Astroparticle Physics

The radiation content of the universe is an important tracer of all physical processes during formation and evolution of structures. Besides helping researchers to unveil this history, the knowledge on the extragalactic background radiation is a key ingredient for interpreting signals from high energy messengers. Low energy photons from the EBL can interact with energetic electrons through inverse Compton scattering, or with protons causing photo-pion production and with very high energy (VHE) gamma rays, above $\sim$100 GeV, by electron-positron pair production. These mechanisms represent sources of opacity for such particles when traversing the intergalactic medium and should be accounted when one wishes to disentangle intrinsic emission phenomena from propagation effects.

The pair production from photon-photon scattering, or Breit-Wheeler, phenomenon (Breit & Wheeler 1934) is the dominant process that affects the propagation of VHE photons. As will be shown in Chapter 3, the cross section of this interaction peaks at an observed gamma ray energy of $E_\gamma(\mathrm{TeV}) \sim 0.83\lambda\,(\mu\mathrm{m})$, for an EBL wavelength λ, meaning that for $\sim$TeV energies, emission is attenuated by near-IR photons ($\sim$1 µm), while gamma rays of dozens of TeV are particularly sensitive to the mid-IR diffuse background. This opacity implies that observed extragalactic fluxes are a result of the intrinsic emission exponentially attenuated by a factor $\exp(-\tau)$, where τ — known as the optical depth — involves an integration over the path, cross section and EBL photon density, described in

section 3.2.2 of Chapter 3.

Since the attenuation due to pair production is energy and redshift dependent, the field of extragalactic gamma-ray astronomy requires the knowledge on the EBL spectrum and its evolution to analyse observations. The limited range of direct or indirect EBL measurements means the interpretation of VHE fluxes requires the use of EBL models to extract the expected absorption during propagation and reconstruct the intrinsic emission of the source. Conversely, it is possible to use gamma-ray observations to indirectly constrain the EBL, if some assumptions are made about the intrinsic emission. This approach complements very well the direct and indirect methods previously discussed. In particular, the range between 5–100 µm can be fully explored with VHE measurements of distant sources, whereas the excessive foreground in this interval turns direct measurements very difficult to perform (Hauser & Dwek 2001).

The gamma-ray sources generally used for EBL constraints are AGNs, as the blazar object type constitutes the majority of extragalactic detections in VHE (Biteau & Meyer 2022). To better understand the physical processes that can explain the gamma-ray emission of these sources and, consequently, the assumptions made on their intrinsic spectrum to constrain EBL levels, we present a brief review about blazars and other AGN classes. This is followed by a more in-depth discussion on different approaches to probe the EBL using gamma rays.

2.3.1 What is a blazar?

From the observational point of view, active galaxies are those with intense emission in all wavelengths, which cannot be explained simply by the aggregate thermal emission from stars. Furthermore, this non-thermal emission is located in a very compact region in the centre — hence the active galactic nuclei name — suggesting that it is powered by a supermassive black hole. In the unified AGN scheme, a structure with some key components is proposed (a black hole, a gaseous disk, a dusty torus, clouds and a relativistic jet), while the orientation relative to the observer — besides intrinsic properties — determines the observed characteristics. Here we are only interested in the broad classification scheme, but a review and more detailed discussion on the unification model can be found in Netzer (2015).

We can classify AGNs between radio-loud and radio-quiet galaxies, depending on the strength of their radio emission. This difference is related to the presence of a relativistic jet (see figures in Dermer & Giebels 2016). Considering radio-loud objects, this jet can be more or less aligned with our line of sight. In case of a misalignment, the object is observed as a radio galaxy, while aligned jets correspond to what we call blazars. Blazars can be of two types: BL Lac or Flat Spectrum Radio Quasars (FSRQ). These subclasses can be distinguished by strong and broad optical emission lines (FSRQ), or weaker/absent

(BL Lac). Physically, the lines should be formed from a intense radiation in the accretion disk or by a dense material in the broad-line region (above the disk plane) (Dermer 2015). The alignment of the jet also implies that the radiation is relativistically beamed, so they become bright gamma-ray emitters even at cosmological distances. In fact, as discussed by Dermer (2015), if the emission region is at rest, the gamma rays produced would be heavily attenuated due to pair production by ambient photons, while a bulk relativistic motion strongly suppress this opacity and boost the luminosity.

When measuring a blazar SED from radio to gamma rays, two distinct bumps are recognised, which should be directly associated to the physical processes that produce this radiation. The first bump — from radio to UV or X-rays — is believed to be due to synchrotron radiation from relativistic electrons (Boettcher 2010, Dwek & Krennrich 2013, Dermer 2015). The position of this peak motivates a further classification of blazars: low-frequency peaked, if the synchrotron peak is in the infrared, for $\nu \leq 1 \times 10^4$ Hz; intermediate-frequency for a peak in the UV–optical between $1{\times}10^{14}$ Hz $< \nu \leq 1{\times}10^{15}$ Hz; and high-frequency ones for $\nu > 1 \times 10^{15}$ Hz. The second bump, related to the highest energies, is usually modelled as a result of inverse Compton scattering of the low energy photons from the synchrotron peak, a mechanism named as Synchrotron Self-Compton (SSC). It can also originate from photons created outside the jet intercepting the relativistic electrons, known as the External Compton model (Dermer 2015). A scenario different from the leptonic one is the case in which protons are accelerated to sufficiently high energies such that the production of pions are possible. In these hadronic models, intense magnetic fields may generate synchrotron radiation from the protons and secondary muons and mesons (Boettcher 2010), or gamma rays produced by photons from π^0 decay, as well as electromagnetic cascades involving charged particles produced in the process.

2.3.2 Probing the EBL and Cosmology using gamma rays

There are a few approaches to constrain EBL levels. Upper limits can be derived from general assumptions about the steepness of the gamma ray spectrum of the source. For a sufficiently narrow energy range, the tail end of the spectrum can be usually described by a power law $\propto E^{-\Gamma}$. If this emission originates from relativistic electrons in the diffusive acceleration scenario, then $\Gamma = 1.5$ should be a reasonable lower bound, while in the hadronic scenario of proton acceleration, $\Gamma = 2$ is expected from the decay of neutral pions (Mazin & Raue 2007). By limiting the hardness of the intrinsic spectrum, the observed fluxes give the maximum attenuation and consequently upper limits on the EBL intensity. The analysis by Aharonian et al. (2006b) applied this approach to the blazars H 2356–309 and 1ES 1101–232, both observed by H.E.S.S.[1], scaling down a phenomenological EBL model to achieve $\Gamma > 1.5$ for both spectra. This resulted in EBL levels close to

[1]The High Energy Stereoscopic System is a ground based system of five Cherenkov telescopes, located in Namibia, for observing VHE gamma rays.

lower limits and theoretical calculations of that time. An EBL level consistent to galaxy counts was also found by MAGIC Collaboration et al. (2008) analysing MAGIC[2] data from one of the most distant VHE sources, the quasar 3C 279 (at $z = 0.536$). Fine tuning an EBL model such that the intrinsic spectral index is no smaller than 1.5, they were able to get compatible results with an EBL evolution that reaches a maximum star formation rate at $z \geq 1$, as expected.

The limiting hardness of gamma-ray spectra was also used as an exclusion criterion for EBL configurations by Mazin & Raue (2007). They have built a generic grid of EBL intensity, from which a continuous optical depth shape could be obtained through spline, constraining the spectral form by analysing 13 gamma-ray spectra from various sources. The lower limit on the photon index was chosen to be 3/2 or 2/3, defining a "realistic" and an "extreme" case, in which a more peculiar acceleration scenario would allow for a harder spectral index. Besides, by testing different forms of the intrinsic spectrum parametrization, they exclude EBL configurations that would result in an exponential pile-up in the gamma-ray emission. Such a strong rise in flux for higher energies can arise from an overestimation of the EBL density, while also being expected in exotic scenarios of Lorentz Invariance Violation (LIV) (Lang et al. 2019) or coupling with axion-like particles (Abdalla et al. 2021). For instance, LIV introduces a change in the energy-dispersion relation for particles (including photons), increasing the energy threshold of pair-production, which decreases the optical depth and consequently increases the distance VHE gamma rays can travel. Depending on the the energy scale of LIV, we can expect a recovery in the observed flux at the highest energies. With a large sample of VHE spectra and adopting Franceschini et al. (2008) EBL model, Lang et al. (2019) were able to put limits on the LIV scale that improved over previous results.

The limits on the EBL by Mazin & Raue (2007) were produced from individual spectra and also with a combined analysis. The combination of different spectra has the advantage of improving the statistics and possible constraints, as the resulting EBL should be able to simultaneously and uniquely describe the attenuation in all spectra. In fact, Mazin & Raue (2007) found that the large majority of possible EBL shapes are excluded by more than one spectrum simultaneously. Consequently, with combined sources, they have found stronger constraints in the interval between 4–60 µm, compatible to the lower limits from galaxy counts by a factor of 2 to 2.5 higher, given the "realistic" case for the photon index limit.

A similar and more recent approach with a generic EBL grid is discussed by Abeysekara et al. (2019), in which 14 spectra from VERITAS[3] and the lower limit in the intrinsic spectral index was set to $\Gamma = 1$. As this method does not incorporate the redshift

[2]The Major Atmospheric Gamma Imaging Cherenkov is a system of two Cherenkov telescopes, located in La Palma, for observing VHE gamma rays.

[3]The Very Energetic Radiation Imaging Telescope Array System is a set of four Cherenkov telescopes, located in Arizona, for observing VHE gamma rays.

evolution of the EBL — by only constraining the local EBL —, the authors estimated the resulting uncertainty from the spread in the optical depth when comparing Franceschini et al. (2008) and Gilmore et al. (2012) models. They identified a maximum impact of 12% in the EBL intensity. Another source of systematic uncertainty comes from the absolute energy scale of the observations, alongside the energy resolution of the instrument, which propagate to the spectral index of the source and consequently to the EBL level. They have estimated an impact of the order of 10% for this VERITAS sample.

Other approach consists of physically modelling the gamma-ray emission, fitting a spectral model to observations in a larger range of wavelengths and consequently inferring the EBL attenuation to match the VHE data. One example can be seen in Guy et al. (2000), where the connection between EBL constraints and model parameters in the SSC framework is investigated. They explore measurements of Markarian 501 in gamma rays during a flare state, while also taking into account sub-TeV and X-ray data to limit parameters of the SSC model. From this work it was possible to set EBL constraints in the range 0.4–3 µm and show that the limits given by variations of SSC parameters are compatible and more stringent than the upper limit which results from simply setting a lower bound in the photon index. While Guy et al. (2000) achieved EBL constraints by scaling existing models, it is also possible to probe directly the optical depth with more general assumptions. Using a samṕle of 15 blazars, Domínguez et al. (2013) were able to determine the cosmic gamma ray horizon (CGRH) — the values of energy and redshift such that $\tau = 1$ — without specifying an EBL model. This was done by gathering data from radio to VHE for all selected sources, such that the observation dates are as close as possible. Then, an SSC model is fitted to the data to recover the intrinsic emission of gamma rays, which can be compared to the observed VHE data points to compute the optical depth. Finally, a third-order polynomial in the logarithm of energy is fitted to the logarithm of the optical depth with some additional constraints, from which the CGRH can be extracted. Their results reveal a good agreement with the prediction from Domínguez et al. (2011) EBL model, while also indicating a possible contamination by zodiacal light in some direct measurements.

A crucial development to gamma-ray astronomy happened with the launch and operation of *Fermi*-LAT. For probing the EBL, this instrument provided valuable measurements of a vast number of extragalactic sources in the range of 100 MeV–100 GeV, which can serve as an estimate of the intrinsic spectra, since the attenuation at these high energies is much less intense than at the TeV region. The work by Orr et al. (2011) explored this fact through the application of two methods to constrain the EBL. First, they assume that the power law that describes the GeV range can be extrapolated to TeV. For this method, they choose three hard spectrum sources, as it is believed that below the inverse Compton peak — which for harder sources occurs at higher TeV energies — the spectrum can be well represented by a power law. Consequently, they were able

to set constraints on the EBL wavelength of 1.6 µm, while not being able to determine the intensity at 15 µm, as the spectra did not extend to higher TeV values. The second method consists of analysing the spectral break between GeV and TeV, which is a result from a change of slope in the EBL opacity around ∼1 TeV. Using 12 sources and plotting the change in spectral index as a function of the redshift, it can be compared with different EBL scenarios to set constraints on the relative intensity in 1.6 µm and 15 µm, complementing the previous approach.

Other works have used a larger sample of VHE measurements to improve EBL constraints alongside the use of *Fermi*-LAT data, notably Meyer et al. (2012), Biteau & Williams (2015) and Desai et al. (2019). In Meyer et al. (2012), high energy measurements from the 2-year Fermi catalogue are used to set an upper limit on the inclination of the VHE spectra of 22 sources detected by ground observatories. The assumption is that the spectral index becomes softer with energy, as the Klein-Nishina effect[4] takes place in the leptonic scenario, or there is a suppression in the accelerated particle spectrum. With this assumption, they were able to eliminate incompatible EBL configurations created through the spline of a generic grid. Other criterion utilised was the exclusion of shapes that produce exponential pile-ups in VHE. An even larger data sample is used by Biteau & Williams (2015), comprised by 106 VHE spectra from 38 sources, associating the corresponding *Fermi*-LAT data as close to the epoch of observation. They also impose a softer spectral index in VHE compared to HE and use an iterative process to obtain the intrinsic spectrum model, starting with a power law and then testing if there is an improvement in the fit with a model with intrinsic curvature (such as log-parabola or exponential cut-off). They were able to probe the EBL in 8 wavelength bins between 0.26–105 µm in good agreement with the estimates from galaxy counts. Desai et al. (2019) uses the same sample as Biteau & Williams (2015), but they derive constraints on the optical depth by scaling existing EBL models and fitting the best configuration to the sample divided in energy and redshift bins. Their optical depth values were then combined with the result from Fermi-LAT Collaboration et al. (2018), in which 739 blazars and one gamma-ray burst detected by *Fermi*–LAT from 9 years of operation were used, generating EBL constraints between redshifts 0.03 and 3.1. While the results from Fermi-LAT Collaboration et al. (2018) can only probe the UV and optical, due to the energy range of *Fermi*–LAT observations, the work by Desai et al. (2019) can complement — at least in lower redshifts — the optical to far-IR due to VHE data. Yet, while consistent constraints are obtained in the mid-IR, current data is not capable of probing beyond $\gtrsim 100$ µm.

This discussion reveals some key points. In most of the works, EBL constraints from gamma rays tend to be consistent with lower limits from galaxy counts. However, while these analyses can tell a great deal on the UV, optical and near-IR intensities of the

[4]The suppression of the photon-electron scattering cross section at high energies, which can affect the SSC mechanism. See e.g. Cerruti (2020).

EBL, current data from ground observatories reaching VHE are insufficient to adequately probe the EBL in longer wavelengths, specially in the far-IR. Finally, it is understood that the simultaneous use of different sources at distinct redshifts can improve the constraints on the EBL intensity and its evolution.

The study on the propagation of VHE photons goes beyond EBL investigations, as gamma rays can also probe the strength of the intergalactic magnetic field, the possibility of coupling between photon and axion-like particles, perform tests to Lorentz Invariance Violation and also constrain cosmological parameters. For an overview on these topics, see e.g. Abdalla et al. (2021), Biteau & Meyer (2022) and Martínez-Huerta et al. (2020). Here, we focus on the latter topic, which will be further discussed on Chapter 5.

Since the EBL is a product of the integrated light emitted throughout the history of star and galaxy formation, it is influenced by the large scale dynamics and, in particular, by the expansion factor of the universe. Besides, extragalactic gamma rays travel cosmological distances and thus are also subject to how the universe expands. Therefore, when modelling the EBL and the extragalactic medium opacity to gamma rays, some cosmological assumptions are needed. In the framework of the ΛCDM model, at least three cosmological parameters explicitly appear when computing the optical depth (see chapters 3 and 5): the matter and cosmological constant density parameters; and the Hubble constant, H_0. Thus, it is possible to vary these parameters and study their impact on the propagation of gamma rays. One of the earliest proposals to constrain H_0 from gamma-ray data is from Salamon et al. (1994), in which the required observation time of a blazar (or number of gamma ray photons) is estimated such that the Hubble constant can be determined with a given precision. At that time, however, only Markarian 421 had been measured at TeV energies and there were very few instruments capable of such measurements. With the advent of new experiments in early 2000's, such as the MAGIC Cherenkov Telescope, new prospects of probing cosmological parameters were discussed by Blanch & Martinez (2005). They extrapolated spectra measured by EGRET to MAGIC energies, given some assumptions about the modelled emission, EBL attenuation and instrument response. Then, they reconstructed the gamma-ray horizon, which can be compared to calculations using different cosmologies. The best fit produced confidence intervals for the density parameters, which are comparable to Supernova Ia determinations. One advantage of this approach is to provide an independent method that does not rely on luminosity-distance relations and the cosmic distance ladder (Riess et al. 2022). However, there are other systematic uncertainties in place, such as the knowledge on the EBL and the flux measurements of the sources. Actual estimates were possible with the growth of VHE data in the past decades. The work by Biteau & Williams (2015), with their large blazar sample, estimated H_0 by comparing the EBL constraints from their gamma-ray data analysis to limits from galaxy counts, in a likelihood marginalised over EBL parameters. Introducing a dependence on the Hubble constant, they constructed its likelihood

distribution and obtained the estimate of $H_0 = 88 \pm 8(\text{stat}) \pm 13(\text{syst})\,\text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1}$, considering the systematic uncertainties from the determination of the optical depth.

One of the most competitive constraint on H_0, however, comes from Domínguez & Prada (2013). In this work, the cosmological dependence on the EBL is emphasized — not only on the distance-redshift relation —, explicitly modifying the luminosity functions used to construct the Domínguez et al. (2011) EBL model with different Hubble constant values. Taking the CGRH for these different cosmologies, they fit the horizon data from Domínguez et al. (2013) and obtain $H_0 = 71.6^{+4.6}_{-5.6}(\text{stat})^{+7.2}_{-13.8}(\text{syst})\,\text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1}$. This approach was expanded in Domínguez et al. (2019) with the new optical depth data from *Fermi*–LAT (Fermi-LAT Collaboration et al. 2018). This allowed for a simultaneous determination of the Hubble constant and matter density parameter, with respective values of $H_0 = 76.4^{+6.0}_{-6.2}\,\text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1}$ and $\Omega_\text{m} = 0.14^{+0.06}_{-0.07}$. In particular, the high redshift optical depth measurements from Fermi-LAT Collaboration et al. (2018) were more relevant to constraining the matter density parameter, while low redshift data can better probe H_0. More robust results are also achieved when a joint likelihood analysis is performed, incorporating independent estimates of these cosmological parameters, such as from baryon acoustic oscillation and type Ia supernovae.

2.4 IACTs and prospects with CTA

A key element for the continuous improvement of EBL constraints from gamma-ray observations — and even the possibility of probing cosmological parameters — came from the expressive growth in data. In the past couple of decades, there was a substantial increase in the discovery of new extragalactic gamma-ray sources, as new instruments provided a leap in sensitivity and coverage. In the high energy range (above 100 MeV), ∼100 extragalactic sources prior to the year 2010 were known, but this number jumped to more than 1000 after the release of the third *Fermi*-LAT catalogue in 2015. For VHE measurements, since 2005, there was an order of magnitude increase primarily due to ground-based instruments such as VERITAS, MAGIC and H.E.S.S., from ∼10 known sources to nearly one hundred today (Biteau & Meyer 2022).

These instruments belong to a category called Imaging Atmospheric Cherenkov Telescopes (IACTs) and are most sensitive to energies of hundreds of GeV to few TeV, as their large effective area — compared to space satellites — allows measurements of the comparatively low photon fluxes at VHEs. They operate by taking into their advantage Earth's opacity to gamma rays. The interaction of energetic photons with atoms in the atmosphere is primarily through pair-production. Due to the high energies involved, the electron-positron pair is emitted close to the direction of the incoming gamma ray. Their subsequent interaction with other nuclei result in bremsstrahlung emission of secondary gamma rays, which in turn — if sufficiently energetic — produce more electron-positron

pairs. This results in an electromagnetic cascade of particles and photons that develops until there is not enough energy to produce new particles (i.e. ionisation energy losses prevail over the bremsstrahlung channel). In the process, the relativistic charged particles can emit a flash of Cherenkov light: the radiation produced by the polarisation of the medium when a charged particle moves along it faster than the speed of light in the material. The coherent light can then be collected by large mirrors and focused to be registered by a light detector — usually a camera comprised by photomultiplier tubes.

The next generation of IACTs, the Cherenkov Telescope Array (CTA) will bring important advances to the field of Astroparticle physics and Gamma-ray Astronomy, as the new instruments — the two largest arrays of Cherenkov telescopes — may reach an order of magnitude improved sensitivity in the VHE range over current IACTs (see e.g. figure 1.1 in Cherenkov Telescope Array Consortium et al. 2019). CTA will have three types of telescopes, of small, medium and large size. The small-sized ones will be distributed in a large area in the southern observatory site, in Chile, to be most sensitive to the highest energies (from ∼TeV to ∼300 TeV), which will be particularly valuable for observing Galactic sources and the Galactic centre. The northern observatory, in La Palma, will contain the large-sized telescopes, most sensitive to the lowest energies, while the medium-sized telescopes will be distributed across both observatories[5]. The improved sensitivity and angular resolution will allow the discovery of even more gamma-ray sources, as well as measurements that reach higher energies, expanding the current description of known objects.

The CTA Consortium elaborated (Cherenkov Telescope Array Consortium et al. 2019) a series of Key Science Projects (KSP), establishing many of the scientific goals to be developed. Some of the topics involve the origin of cosmic particles, probing the physical process of extreme environments and investigating Physics frontier topics, such as dark matter and LIV. In particular, improved measurements at VHEs will contribute to investigate any mechanism affecting photon propagation over extragalactic distances, as presented in Abdalla et al. (2021) . This involves attenuation due to the EBL, but also the development of electromagnetic cascades in the intergalactic magnetic field (allowing constraints on the strength of these fields), possible coupling of photons to axion-like particles and LIV phenomena, both of which could alter the transparency of the universe to VHE gamma rays. The KSP related to AGN observations, with its long-term monitoring strategy, search for flares and systematic coverage for high quality spectra (Cherenkov Telescope Array Consortium et al. 2019) will therefore be very important to cover these subjects.

[5]This is a general description of the "alpha" configuration programmed to be initially constructed. In the ideal, "omega" configuration, the large telescopes would also be installed in the southern array. See e.g. Acharyya et al. (2019) for the Consortium' study on the performance of the optimal array distribution.

Chapter 3

Methodology

This chapter contains a detailed explanation on all the general ingredients and theoretical background to comprehend the analysis of subsequent chapters. In section 3.1, we present the main EBL model used throughout this work. Then, we discuss the resulting opacity to gamma rays in section 3.2. In section 3.3, we introduce the Bayesian approach that guides the inference on the model parameters. This chapter concludes with the description of the statistical tools for performing such inference, in section 3.4.

3.1 Modelling the EBL energy density

The base EBL model used throughout this work is from Finke et al. (2010), which we are going to abbreviate as F10 from now on. It considers the EBL as a sum of two contributions: direct stellar emission (dominating the UV and optical) and reprocessed emission from dust contained in the interstellar medium of galaxies (dominating the IR band). While the stellar emission depends on properties of stellar populations and their evolution, the dust is comprised by three independent components assumed to emit as blackbodies (at different temperatures) which absorb a fraction of the stellar radiation and re-emit it at longer wavelengths. The choice of this model involves some considerations. Firstly, the parameters used in F10 have a clear physical interpretation, as we can directly relate the EBL intensity to distinct astronomical origins. Secondly, as we are mostly interested in investigating EBL constraints in the IR region, it is advantageous that this model has free parameters which enable wavelength-dependent changes in the EBL intensity, through modifying the abundance and temperature of the dust components — as will be shown later. Finally, this work is expanding the research from de Matos Pimentel & Moura-Santos (2019), in which the F10 model was implemented, so we share some of their tools. In what follows, we are going to explain how this EBL model is constructed.

As the EBL has a cosmological dependency, we start in the comoving frame — i.e.,

the frame in which we factor out the expansion of the universe — and compute the stellar emissivity. In this context, emissivity is the comoving luminosity (energy per time) per unit of comoving volume, or comoving luminosity density, expressed in general as

$$J^{\mathrm{star}}(E,z) = E^2 \frac{dN}{dtdEdV}, \tag{3.1}$$

where E is the EBL photon energy in the comoving frame and $dN/(dtdEdV)$ is the number of photons dN emitted in a time interval dt, energy range dE and comoving volume dV. Following F10, we define the dimensionless quantity

$$\epsilon = \frac{E}{m_e c^2} \tag{3.2}$$

using the rest energy of an electron (mass m_e). Making a change of variable in eq. 3.1, we arrive at

$$J^{\mathrm{star}}(\epsilon,z) = m_e c^2 \epsilon^2 \frac{dN}{dtd\epsilon dV}. \tag{3.3}$$

The differential emission $dN/(dtdEdV)$ at a given redshift depends on the integration of the photon emission of all existing stars up to that point. Let's first consider the emission of a single star. For simplicity, F10 assumes each star to emit as a blackbody, which was verified to approximately reproduce the spectrum of clusters of stars in simple stellar populations[1] of different ages (see figure 1 in F10). Naturally, this assumption does not reproduce high resolution characteristics of these spectra (e.g. absorption lines), but it reasonably describes the essence of the thermal process.

To obtain the differential emission, we remember that the radiance of a blackbody (energy emitted per unit of time, area and frequency) is related to its energy density $\rho_T(\nu)$ by

$$\frac{dE}{dtd\nu dA} = \frac{c}{4}\rho_T(\nu), \tag{3.4}$$

for an equilibrium temperature T, which follows from the homogeneity and isotropy of the radiation field. Moreover,

$$\rho_T(\nu)\mathrm{d}\nu = \frac{8\pi}{c^3}\frac{h\nu^3}{[\exp(h\nu/k_B T)-1]}\mathrm{d}\nu, \tag{3.5}$$

which can be expressed as a photon number density

$$n_T(\nu)\mathrm{d}\nu = \frac{8\pi}{c^3}\frac{\nu^2}{[\exp(h\nu/k_B T)-1]}\mathrm{d}\nu, \tag{3.6}$$

dividing eq. 3.5 by $h\nu$. Now changing variables from ν to ϵ and realising that $\mathrm{d}\nu =$

[1] A simple stellar population, SSP, is a population of stars born at the same time.

$(m_e c^2/h)\,\mathrm{d}\epsilon$, we obtain

$$n_\Theta(\epsilon)\,\mathrm{d}\epsilon = \frac{8\pi}{\lambda_C^3}\frac{\epsilon^2}{[\exp(\epsilon/\Theta)-1]}\,\mathrm{d}\epsilon\,, \tag{3.7}$$

defining $\Theta = k_B T/(m_e c^2)$ as the dimensionless effective temperature and using the Compton wavelength $\lambda_C = h/(m_e c)$. The numerical radiance is therefore

$$\frac{dN}{dtd\epsilon dA} = \frac{c}{4} n_\Theta(\epsilon). \tag{3.8}$$

Integrating eq. 3.8 over the area surface of a spherical star of radius R, i.e. $4\pi R^2$, its photon emission rate turns out to be

$$\dot{N}_*(\epsilon; m, t_*) \equiv \frac{dN_*}{dtd\epsilon} = \pi R(m,t_*)^2 c n_\Theta(\epsilon). \tag{3.9}$$

The stellar radius is actually a function of the star's mass m and age t_*, as it changes during different evolutionary phases. Ultimately, the temperature is also a function of these properties, $\Theta = \Theta(m, t_*)$, as it is tied to the star's radius and luminosity $L(m, t_*)$. This can be easily seen in the Stefan-Boltzmann law, as the radiance of a blackbody is proportional to T^4, so

$$\frac{L(m,t_*)}{4\pi R(m,t_*)^2} = \sigma T^4, \tag{3.10}$$

where $\sigma = 5.6703 \times 10^{-8}\,\mathrm{W\,m^{-2}\,K^{-4}}$ is the Stefan-Boltzmann constant. Both radius and luminosity can be obtained by the position of the star in the Hertzsprung-Russel (HR) diagram in combination with stellar evolution models. This information is encapsulated in Eggleton et al. (1989) stellar formulae, used with some corrections by F10. These formulae relate the time a star stays in the main sequence depending on its mass, as it also describes luminosity and radius evolution beyond the main sequence, from giant to white dwarf phases.

Having the emission due to one star at each moment in time (eq. 3.9), to obtain the total stellar emissivity, we need to integrate it over the populations of stars of different ages and masses. This is firstly done by convolving the star's emission with the comoving star formation rate (SFR) density, $\psi(t)$, then integrating it with the initial mass function (IMF), $\xi(m)$. The SFR describes the rate at which the mass in the interstellar medium is converted into stars, so the comoving SFR density is the SFR per comoving volume. The IMF, on the other hand, is the distribution of initial masses in a stellar population. Therefore, mathematically one can write:

$$J^{\mathrm{star}}(\epsilon, z) = m_e c^2 \epsilon^2 f_{\mathrm{esc}}(\epsilon) \int_{m_{\min}}^{m_{\max}} \mathrm{d}m\, \xi(m) \int_z^{z_{\max}} \mathrm{d}z_1 \left|\frac{\mathrm{d}t_*}{\mathrm{d}z_1}\right| \psi(z_1) \dot{N}_*(\epsilon; m, t_*(z, z_1)). \tag{3.11}$$

Conceptually, by fixing the redshift z and photon energy ϵ, we need to sum the blackbody

emission at this energy from all stars born at $z_1 \in [z, z_{\max}]$ (weighted by the SFR) in the amount accordingly to the distribution of masses in this population, given by the IMF. F10 investigated different SFR and IMF model combinations, but found that Hopkins & Beacom (2006) SFR and Baldry & Glazebrook (2003) IMF parametrizations had the best agreement with luminosity density data, specially at lower redshifts. Equation 3.11 has also two other terms: first, $f_{\rm esc}(\epsilon)$ represents the fraction of the radiation that effectively escapes the galaxies and enters the intergalactic medium (i.e., is not absorbed by dust). This was first determined by Driver et al. (2008) using dust absorption models and later parameterised by Razzaque et al. (2009) using power law fits. F10 adopt a further assumption of full absorption of photons with energy $E > 13.6\,\mathrm{eV}$, due to the ionisation of neutral hydrogen in the interstellar and intergalactic medium, so the UV radiation beyond this energy is not reprocessed in their model. The second term in eq. 3.11 is the cosmological factor $|\,\mathrm{d}t_*/\mathrm{d}z_1\,|$, which emerges from the change of variables from time to redshift. In the ΛCDM cosmology, we consider a flat universe comprised essentially by non-relativistic matter and a dark energy (cosmological constant) component, resulting in

$$\left|\frac{\mathrm{d}t}{\mathrm{d}z}\right| = \frac{1}{H_0(1+z)}\frac{1}{\sqrt{\Omega_m(1+z)^3+\Omega_\Lambda}}, \tag{3.12}$$

where H_0 is the Hubble constant, Ω_m and Ω_Λ are the density parameters for matter and dark energy, respectively. Here, we actually neglect the radiation density parameter, as it has a negligible effect in the cosmology for low redshifts. For the results that are discussed in Chapters 4 and 6, we set $\Omega_m = 0.3 = 1 - \Omega_\Lambda$ and $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, but a change in these cosmological parameters is discussed in Chapter 5.

The stellar radiation which did not escape the galaxies is then absorbed by dust in their interstellar medium. According to Desert et al. (1990), a minimal model for the dust composition capable of explaining the extinction and emission of the diffuse interstellar medium requires three components: below 15 µm, Polycyclic Aromatic Carbon molecules (PAHs) could explain emission features in the IR; between 15–60 µm, very small grains (of several nanometers) are required to describe an absorption bump in 220 nm; while for longer wavelengths in the far-IR, above 100 µm, one component of big grains can be utilised to describe the IR and visible extinction curves. In F10's model, the emission of these three components is simplified as blackbodies of distinct temperatures. In fact, such modelling was previously presented by Kneiske et al. (2002a). They argue that, although PAHs are known to not emit in thermal equilibrium, their broad emission lines can be modelled as a very hot blackbody continuum of low flux. Since the EBL is comprised by a sum of these emissions at various redshifts, the detailed spectral shapes would be smoothed out by the integration. Small grains (SG) and large grains (LG), on the other hand, are in thermal equilibrium and represent the hot and warm dust components, respectively.

Each dust component is responsible for absorbing a fraction f_n ($n = 1, 2, 3$) of the

stellar radiation retained in the galaxies. Since they emit as blackbodies, their respective emissivities are given by

$$J_n(\epsilon, z) \equiv m_e c^2 \epsilon^2 \left(\frac{dN}{dt d\epsilon dV} \right)_{\text{dust}} = m_e c^2 \epsilon J_n^0 \frac{\epsilon^3}{[\exp(\epsilon/\Theta_n) - 1]}, \tag{3.13}$$

with dimensionless temperature Θ_n and a normalisation constant J_n^0. This constant can be determined by realising that, because all of the absorbed energy is re-emitted, there is an equality between the integrated energies

$$\int_0^\infty \mathrm{d}E \left(E \frac{dN^{(n)}}{dt dE dV} \right)_{\text{dust}} = f_n \int_0^\infty \mathrm{d}E \, (1 - f_{\text{esc}}) \frac{1}{f_{\text{esc}}} \left(E \frac{dN}{dt dE dV} \right)_{\text{star}}, \tag{3.14}$$

observing that the stellar part is the complementary amount to eq. 3.11 multiplied by the dust fraction f_n. This can be further written as

$$\int_0^\infty \mathrm{d}\epsilon \, J_n^0 \frac{\epsilon^3}{[\exp(\epsilon/\Theta_n) - 1]} = f_n \int_0^\infty \mathrm{d}\epsilon \, (1 - f_{\text{esc}}(\epsilon)) \frac{J^{\text{star}}(\epsilon, z)}{m_e c^2 \epsilon f_{\text{esc}}(\epsilon)}. \tag{3.15}$$

The integration on the left-hand side of eq. 3.15 results in $J_n^0 \Theta_n^4 \Gamma(4) \zeta(4) = J_n^0 \Theta_n^4 \pi^4 / 15$, so the normalisation constant is

$$J_n^0 = \frac{15}{\pi^4} \frac{f_n}{\Theta_n^4} \int_0^\infty \mathrm{d}\epsilon \, (1 - f_{\text{esc}}(\epsilon)) \frac{J^{\text{star}}(\epsilon, z)}{m_e c^2 \epsilon f_{\text{esc}}(\epsilon)}. \tag{3.16}$$

Substituting it back in eq. 3.13 and summing over n, the total emissivity coming from dust becomes

$$J^{\text{dust}}(\epsilon, z) = \frac{15}{\pi^4} \left[\int_0^\infty \mathrm{d}\epsilon' \left(\frac{1}{f_{\text{esc}}(\epsilon')} - 1 \right) \frac{J^{\text{star}}}{\epsilon'}(\epsilon', z) \right] \sum_{n=1}^{3} \frac{f_n}{\Theta_n^4} \frac{\epsilon^4}{[\exp(\epsilon/\Theta_n) - 1]}. \tag{3.17}$$

Naturally, the total comoving luminosity density is simply the sum

$$J^{\text{total}}(\epsilon, z) = J^{\text{star}}(\epsilon, z) + J^{\text{dust}}(\epsilon, z). \tag{3.18}$$

This quantity can then be used to compute the total comoving EBL energy density U by integrating over the past emissivity up to a given redshift, as

$$U(\epsilon, z) \equiv m_e c^2 \epsilon^2 \frac{dN}{d\epsilon dV} = \int_z^{z_{\max}} \mathrm{d}z_1 \, \frac{J^{\text{total}}(\epsilon', z_1)}{1 + z_1} \left| \frac{\mathrm{d}t}{\mathrm{d}z_1} \right|, \tag{3.19}$$

where $z_{\max} = 6$ is a maximum redshift for which we consider star formation and $\epsilon' =$

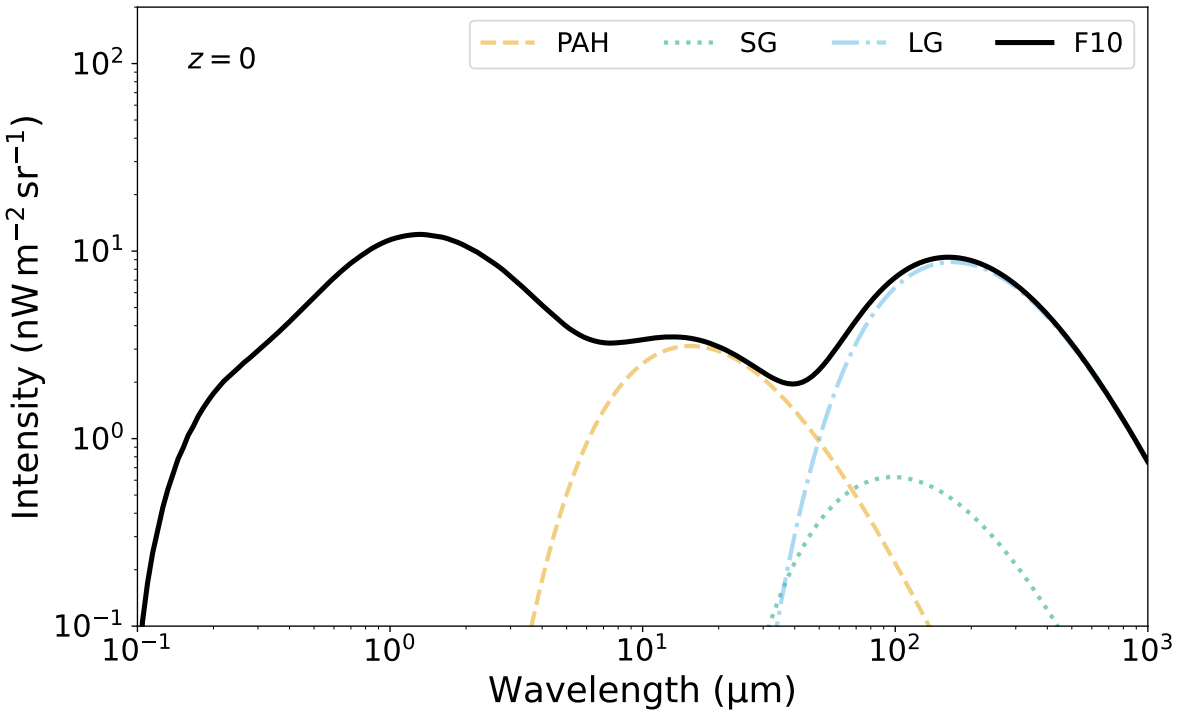

Figure 3.1: Local ($z = 0$) EBL intensity according to F10 model. The individual contributions of the PAH, small grains and large grains are also shown.

$\epsilon(1 + z_1)$. To understand the redshift factors, first we bring attention to the fact that comoving quantities are those that remain constant with the expansion of the universe, so they are equal, by definition, to their values in the present time ($z = 0$), if the scale factor is normalised to $a(t_0) = 1$ today. Then, the contributions to the comoving EBL energy density at a redshift z and photon energy ϵ are from the emission at $z_1 \geq z$ of photons with energy $1 + z_1$ times greater than observed today, due to the cosmological redshift, so the emissivity must be computed at $\epsilon' = \epsilon(1 + z_1)$. Besides, the luminosity density requires another transformation, as it is a temporal rate of a comoving density. This is because photons separated by a time interval δt_e during emission at z_1 will have a comoving distance of $c\delta t_e(1 + z_1)$ (Ryden 2017), so the effective time interval observed today would be enlarged by $\delta t_0 = \delta t_e(1 + z_1)$. This means the luminosity at z_1 will decrease by the same factor $1 + z_1$ as it is summed for its contribution to the comoving EBL energy density.

The local ($z = 0$) EBL density spectrum from F10's model, using eq. 3.19, can be visualised in figure 3.1, showing the contribution of each dust component to the total energy content. However, instead of plotting the energy density, we transform to intensity (energy crossing an area element per unit of time and solid angle) by simply multiplying it by $c/(4\pi)$ (an isotropic radiation field in vacuum). This figure was elaborated using an EBL grid previously computed by de Matos Pimentel & Moura-Santos (2019). The details of this grid and a discussion on the resulting opacity to gamma rays are presented

next, in section 3.2.

3.2 Modelling the EBL opacity

As discussed in section 2.3 of Chapter 2, the interaction of gamma rays with the EBL results in a decrease of the VHE photon flux during propagation. The attenuation of the specific intensity, I_ν, of gamma rays along the propagation path l can be described by the radiation transfer equation in the absorption-only case

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}l} = -\alpha_\nu I_\nu, \tag{3.20}$$

where α_ν is the absorption coefficient. We neglect source terms that could contribute to an increased intensity, like a secondary emission from pair-produced electrons and positrons deflected in intergalactic magnetic fields, which can interact with low energy background photons through inverse Compton scattering (Ichiki et al. 2008).

Equation 3.20 can be formally solved by introducing the optical depth τ_ν, such that $\mathrm{d}\tau_\nu = \alpha_\nu \,\mathrm{d}l$, resulting in

$$I_\nu(\tau_\nu) = I_\nu(0)e^{-\tau_\nu} \tag{3.21}$$

with

$$\tau_\nu(l) = \int_{l_0}^{l} \mathrm{d}l' \, \alpha_\nu(l'), \tag{3.22}$$

considering the coordinate position l_0 such that $\tau_\nu(l_0) = 0$. The optical depth measures the average length that the radiation has travelled, in units of mean free path, while the mean free path can be thought of as the typical scale of propagation before the particle interacts once. Therefore, the optical depth is the fundamental quantity we need to compute in order to quantify the attenuation to the intensity (and consequently to the flux) of extragalactic gamma rays. For this task, we must first obtain the absorption coefficient (which also corresponds to the reciprocal of the mean free path). It depends on the numerical density of the interacting particles (the EBL photons) and the cross section of the interaction. The first of these ingredients we have obtained in section 3.1, so in the following subsection the cross section is presented.

3.2.1 The Breit-Wheeler interaction

The EBL photon density is the dominant source of opacity for VHE gamma rays in the intergalactic medium. As mentioned in Chapter 2, the Breit-Wheeler process — the electron-positron pair production by two photons in vacuum — is the physical interaction

responsible for the effective absorption of freely propagating gamma rays. This process is best described by Quantum Electrodynamics (QED), since, with the Feynman diagrams formalism, it is possible to analytically compute the invariant amplitude at tree level and the correspondent cross section. In Appendix A, we provide a guide for this computation. Using the notation from Gould & Schréder (1967), the Breit-Wheeler cross section is

$$\sigma_{\mathrm{BW}}(\beta) = \frac{\pi}{2}\left(\frac{\alpha\hbar c}{m_e c^2}\right)^2 (1-\beta^2)\left[(3-\beta^4)\ln\left(\frac{1+\beta}{1-\beta}\right) - 2\beta(2-\beta^2)\right], \tag{3.23}$$

where α is the fine-structure constant and $\beta = v/c$ is the electron and positron speed in the centre-of-momentum system.

Such interaction has some kinematic restrictions. Considering a gamma ray photon with 4-momentum $p'_\gamma = (E'_\gamma/c, \mathbf{p}'_\gamma)$ and an EBL photon with $p' = (E'/c, \mathbf{p}')$ in the centre-of-momentum frame (primed quantities), the squared magnitude of the total 4-momentum of the system

$$P'^2 = \frac{(E'_\gamma + E')^2}{c^2} - \left|\mathbf{p}'_\gamma + \mathbf{p}'\right|^2 \tag{3.24}$$

is an invariant quantity. In the centre-of-momentum frame, $\mathbf{p}'_\gamma + \mathbf{p}' = 0$ and the conservation of energy imposes

$$E'_\gamma + E' = 2E'_e, \tag{3.25}$$

where E'_e is the energy of the electron/positron in this frame (using the fact that they have the same mass and the total momentum is zero, so their energies must be the same). Then,

$$P'^2 = \frac{4E'^2_e}{c^2}. \tag{3.26}$$

In the "laboratory" frame, on the other hand, we orient the gamma ray in the z-axis and consider a collision on the xz-plane where the EBL photon makes an angle θ with the z-axis. Therefore,

$$\mathbf{p}_\gamma + \mathbf{p} = \left(\frac{E}{c}\sin\theta, 0, \frac{E_\gamma}{c} + \frac{E}{c}\cos\theta\right), \tag{3.27}$$

using the energy-dispersion relation $E = pc$. So computing P^2 and using eq. 3.26, we arrive at the relation

$$2E_\gamma E(1-\cos\theta) = 4E'^2_e. \tag{3.28}$$

For a given gamma ray, the energy threshold of this interaction occurs when the electron-positron pair is produced at rest in the centre-of-momentum frame ($E'_e = m_e c^2$),

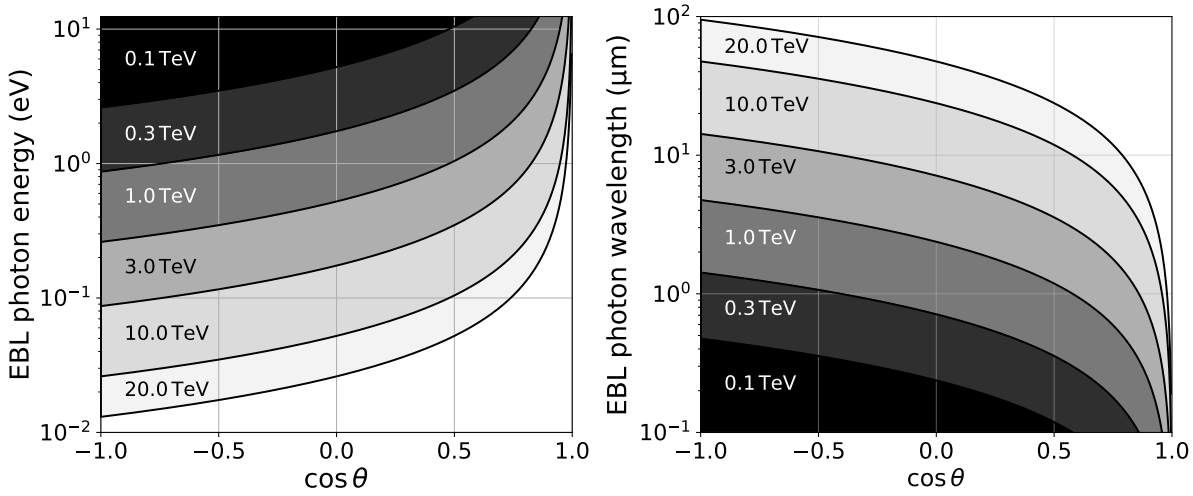

Figure 3.2: The shaded regions delimited by the solid lines are the allowed kinematic space for pair production, according to eq. 3.29. The different shades correspond to fixed energies of the gamma-ray photon (written next to the line threshold).

so the allowed EBL energies for pair production obey

$$E \geq \frac{2m_e^2c^4}{E_\gamma(1-\cos\theta)}, \tag{3.29}$$

for $0 < \theta \leq \pi$. This restriction is illustrated in figure 3.2, where the (E, θ) possible values for the interaction are plotted for different gamma-ray energies.

Finally, we can relate β to other kinematic quantities, observing that

$$\beta^2 = \frac{v^2}{c^2} = \frac{p^2c^2}{E_e'^2} = \frac{E_e'^2 - m_e^2c^4}{E_e'^2} = 1 - \frac{m_e^2c^4}{E_e'^2} = 1 - \frac{2m_e^2c^4}{E_\gamma E}\frac{1}{(1-\cos\theta)}, \tag{3.30}$$

using eq. 3.28. Figures 3.3 and 3.4 present different slices of the cross section, as a function of β and $\cos\theta$, E_γ and $\lambda_{\text{EBL}} = hc/E$. Numerically, the cross section peaks at $\beta_{\max} \approx 0.7013$, which means we can obtain a relation between the EBL wavelength and gamma ray energy in the peak of the interaction to be

$$\lambda_{\text{EBL}} = (1-\cos\theta)(1-\beta_{\max}^2)\frac{hc}{2m_e^2c^4}E_\gamma. \tag{3.31}$$

Substituting the values and considering the average of the interaction angles, $\langle\cos\theta\rangle = 0$,

$$\left(\frac{\lambda_{\text{EBL}}}{\mu\text{m}}\right) \sim 1.206\left(\frac{E_\gamma}{\text{TeV}}\right) \tag{3.32}$$

is an order of magnitude estimate of the typical wavelengths and energies of the interaction.

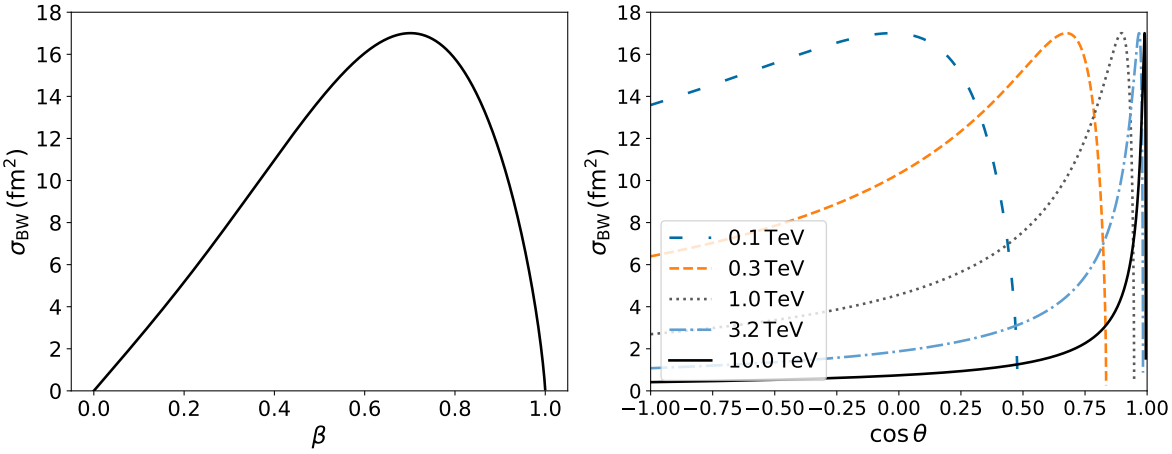

Figure 3.3: Cross section of the Breit-Wheeler process (eq. 3.23) as a function of β (left) and of $\cos\theta$ (right) for different energies of the gamma ray photon. The EBL photon energy was fixed to 10 eV ($\lambda \approx 0.124\,\mu$m).

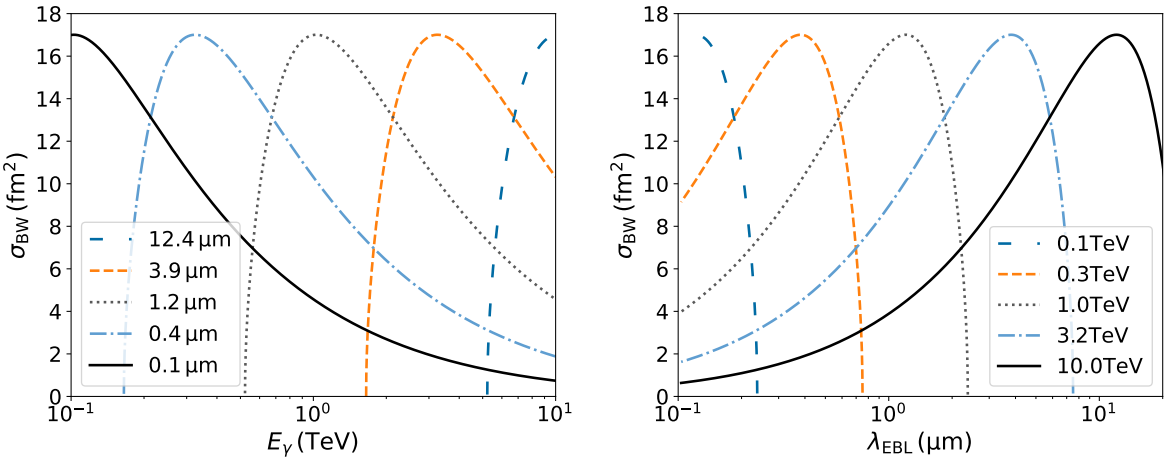

Figure 3.4: Cross section of the Breit-Wheeler process (eq. 3.23) as a function of the gamma ray energy (left) and of the EBL photon energy (right). Each curve corresponds to a different wavelength or energy of the respective other photon. The interaction angle was fixed to $\cos\theta = 0$.

3.2.2 Attenuation of gamma rays

We want to obtain the differential optical depth with respect to the path, or the reciprocal mean free path λ^{-1},

$$\alpha_\nu \equiv \frac{\mathrm{d}\tau}{\mathrm{d}l} = \lambda^{-1}(E_\gamma, z) \tag{3.33}$$

as a function of the observed gamma ray energy and redshift of emission. Since the gamma ray may interact with photons coming from different directions and in a range of different energies, to compute the absorption coefficient we need to integrate the cross section and the EBL numerical density over the allowed kinematic phase space $\mathcal{P}$, as

$$\frac{\mathrm{d}\tau}{\mathrm{d}l'} = \int_{\mathcal{P}} \sigma \, \mathrm{d}n' \,. \tag{3.34}$$

Such expression can be more easily computed in the case of a massive particle interacting with a photon — such as the Compton scattering (Blumenthal & Gould 1970) — by going to the particle's rest frame. Then, we will take the limit of massless particles to arrive at the correct form for the Breit-Wheeler process.

Consider the same geometry of the γ-EBL collision, but with an electron travelling at speed $v \equiv \beta c$, instead of the gamma ray. When transforming to the electron's rest frame (primed quantities), the path $\mathrm{d}l'$ is contracted as $\mathrm{d}l' = \mathrm{d}l \,/\gamma$, where $\gamma = (1-\beta^2)^{-1/2}$ is the Lorentz factor. Therefore, the differential optical depth, or reciprocal mean free path, in observed "laboratory" coordinates, is given by

$$\frac{\mathrm{d}\tau}{\mathrm{d}l} = \frac{1}{\gamma}\frac{\mathrm{d}\tau}{\mathrm{d}l'} = \frac{1}{\gamma}\int_{\mathcal{P}} \sigma \, \mathrm{d}n' \,, \tag{3.35}$$

where $\mathrm{d}n'$ is the differential photon density with which the electron interacts in this rest frame. To transform to the laboratory frame the quantity $\mathrm{d}n$, we may realise that while the particle number is an invariant quantity, the spatial volume is not. Taking the differential spacetime volume $\mathrm{d}X = \mathrm{d}x^0 \, \mathrm{d}V$, where x^0 is the time component, we can write

$$\mathrm{d}n = \frac{\mathrm{d}N}{\mathrm{d}V} = \frac{\mathrm{d}N}{\mathrm{d}X}\,\mathrm{d}x^0 \tag{3.36}$$

and, since $\mathrm{d}N/\mathrm{d}X$ is invariant, $\mathrm{d}n$ should transform as the "time" component of a 4-vector. This is useful because we can construct the invariant quantity (Blumenthal & Gould 1970)

$$\frac{\mathrm{d}n}{E} = \frac{\mathrm{d}n'}{E'}, \tag{3.37}$$

where E is the photon energy. This is because the energy is the zeroth-component of the

4-momentum. Therefore, we arrive at

$$\frac{\mathrm{d}\tau}{\mathrm{d}l} = \frac{1}{\gamma}\int_{\mathcal{P}} \sigma \frac{E'}{E}\,\mathrm{d}n\,. \tag{3.38}$$

Explicitly, the boost in the direction of the electron's velocity produces the usual Lorentz transformation matrix representation

$$\Lambda = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{3.39}$$

while the photon's 4-momentum is

$$p^{\mu} = (E/c, p\cos\theta, -p\sin\theta, 0), \tag{3.40}$$

implying that

$$E' = \gamma E(1 - \beta\cos\theta). \tag{3.41}$$

This results in

$$\frac{\mathrm{d}\tau}{\mathrm{d}l} = \int_{\mathcal{P}} \sigma(1 - \beta\cos\theta)\,\mathrm{d}n\,, \tag{3.42}$$

where now the angle and the photon density are in laboratory quantities.

For two photons, there is no particle rest frame, but the correct result (Gould & Schréder 1967) can be achieved by taking the limit $\beta \to 1$. The intuition behind it is that the mean free path expression (eq. 3.34) is valid for an incident beam on a stationary target or, conversely, to a target moving in the direction of a particle at rest. Therefore, by going to a reference frame where both the beam and target are moving in relation to the observer, it is necessary to take into account the relative velocity in this new frame. The relativistic transformations of this change of reference then impose a modification on the target photon density that the incident particle encounters. In the case of photons, the relativistic effect implies that there is no contribution to the reciprocal mean path coming from $\theta = 0$ photons, as parallel photons cannot intercept.

To obtain the final expression for the differential optical depth, it is important to realise that the interaction also happens in the "proper" cosmological reference frame (index p), meaning that all quantities need to be evaluated at redshift z (e.g. $E_{\gamma,\mathrm{p}} = E_{\gamma}(1+z)$). That is,

$$\frac{\mathrm{d}\tau}{\mathrm{d}l} = \lambda^{-1}(E_{\gamma}, z) = \int_{\mathcal{P}} \mathrm{d}\Omega\,\mathrm{d}E_{\mathrm{p}}\,(1 - \cos\theta)\sigma_{\mathrm{BW}}(\theta, E_{\mathrm{p}}, E_{\gamma,p})\tilde{n}(E_{\mathrm{p}}, z), \tag{3.43}$$

where we write $\mathrm{d}n = \mathrm{d}\Omega\,\mathrm{d}E_\mathrm{p}\,\tilde{n}(E_\mathrm{p}, z)$. The solid angle integration then can be further simplified as there is no dependence in the cross section to the azimuth angle ϕ, which can be integrated resulting in a factor of 2π. Besides, as the EBL photons have an isotropic distribution, for a fixed solid angle, only a fraction 4π of the total density will contribute to the absorption probability. Calling $\tilde{n}(E_p, z) = n(E_p, z)/(4\pi)$ and changing the integration variable $\cos\theta \to \mu$, we arrive at

$$\lambda^{-1}(E_\gamma, z) = \int_{-1}^{1} \mathrm{d}\mu\,\frac{1-\mu}{2} \int_{E_\mathrm{min,p}}^{\infty} \mathrm{d}E_p\,\sigma_\mathrm{BW}(\mu, E_p, E_{\gamma,p}) n(E_p, z), \tag{3.44}$$

showing explicitly the integration limits.

The numerical density $n(E_\mathrm{p}, z)$ is the proper EBL numerical density in the frame of the interaction. While the comoving numerical density can be readily extracted from eq. 3.19, as

$$n_c(\epsilon, z) \equiv \frac{1}{m_e c^2 \epsilon^2} U(\epsilon, z), \tag{3.45}$$

the relation between $U(\epsilon, z)$ and $U_\mathrm{p}(\epsilon_\mathrm{p}, z)$ — the proper EBL energy density — is given by

$$U_\mathrm{p}(\epsilon_\mathrm{p}, z) = (1+z)^4 U(\epsilon, z). \tag{3.46}$$

This is because the volume was $(1+z)^3$ times smaller in the past (so the density was bigger by the same factor) and the radiation energy is $(1+z)$ times smaller today, due to the cosmological redshift. Thus,

$$n(\epsilon_\mathrm{p}, z) = \frac{1}{m_e c^2 \epsilon_\mathrm{p}^2} (1+z)^4 U(\epsilon, z), \tag{3.47}$$

where the energies are related by $\epsilon_\mathrm{p} = \epsilon(1+z)$.

The integration limits of eq. 3.44 are also consistent with the phase space in figure 3.2, as we can first integrate over the EBL photon energies, from the lower bound of eq. 3.29,

$$E_\mathrm{min,p} = \frac{2 m_e^2 c^4}{E_{\gamma,p}(1-\mu)}, \tag{3.48}$$

to infinity, and then integrate over the angles. Bear in mind that, if we change the integration order, the angle integration limits become $\mu = -1$ to $\mu_\mathrm{max} = \mu(E_\mathrm{min,p})$ and, in energy, from $E_\mathrm{min,p}(\mu = -1) = m_e^2 c^4 / E_{\gamma,\mathrm{p}}$ to infinity. This order is useful because, as presented by Gould & Schréder (1967), it is possible to perform an analytical integration over θ (using an appropriate change of variables). For a different parameterization, Biteau & Williams (2015) presents another formula. This in turn reduces the computational effort of obtaining the optical depth, as one less integral needs to be made.

Finally, the optical depth can then be computed by integrating the reciprocal mean

free path (eq. 3.44) through the cosmological distance, which is done by transforming the distance integration to redshift, $\mathrm{d}l = c|\,\mathrm{d}t/\mathrm{d}z\,|\,\mathrm{d}z$, as

$$\tau(E_\gamma, z) = c \int_0^z \mathrm{d}z' \left|\frac{\mathrm{d}t}{\mathrm{d}z'}\right| \int_{-1}^{1} \mathrm{d}\mu\, \frac{1-\mu}{2} \int_{E_{\mathrm{min,p}}}^{\infty} \mathrm{d}E_p\, \sigma_{\mathrm{BW}}(\mu, E_p, E_{\gamma,p}) n(E_p, z) \tag{3.49}$$

and applying the correct redshift factors to the proper energies.

Previous work from de Matos Pimentel & Moura-Santos (2019) (hereinafter P19) computed from the ground up an optical depth grid, based on eq. 3.49 and the implementation of F10's model of section 3.1. Since the total emissivity is the sum of the stellar and dust components (eq. 3.18), the linearity of all integration steps implies the optical depth can be separated as

$$\tau(E_\gamma, z) = \tau^{\mathrm{star}}(E_\gamma, z) + \sum_n f_n \tau^{(\mathrm{n})}(E_\gamma, z, \Theta_n), \tag{3.50}$$

where $n = \{\mathrm{PAH}, \mathrm{SG}, \mathrm{LG}\}$. So, for fixed dust temperature values, four grids were computed as a function of the redshift of emission and observed gamma-ray energy. Another grid in which Θ_n is a free parameter was also created by Dr. Douglas de Matos Pimentel and we use it in Chapter 4 to investigate gamma ray constraints of the PAH effective temperature. A last grid was elaborated by the author of this dissertation in which the Hubble constant H_0 is added as free parameter. The cosmological constraints by studying gamma-ray absorption are presented in Chapter 5. Table 3.1, then, informs the technical specifications of all of these grids. They are an important asset to quickly compute the optical depth a gamma ray with arbitrary energy and redshift is subject to, which is exactly what is needed when investigating EBL constraints from gamma-ray observations. For illustration, Figure 3.5 presents the optical depth curves as a function of the energy for all redshift values, fixing $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Next, we present the statistical analysis that allows the inference of EBL and spectral properties of gamma-ray sources.

3.3 Bayesian Inference

Our goal is to recover information about parameters, $\mathbf{\Omega}$, describing the EBL and the spectral fluxes of extragalactic gamma-ray sources, by comparing the observed fluxes with modelled ones, which take into account the attenuation due to the EBL. The Bayesian approach allows us to write a joint probability distribution of all parameters (the posterior distribution) in terms of the likelihood of the data, D, and prior information we may have. This can be expressed by the Bayes' theorem

$$p(\mathbf{\Omega}|D, I) = \frac{p(D|\mathbf{\Omega}, I)p(\mathbf{\Omega}|I)}{p(D|I)}, \tag{3.51}$$

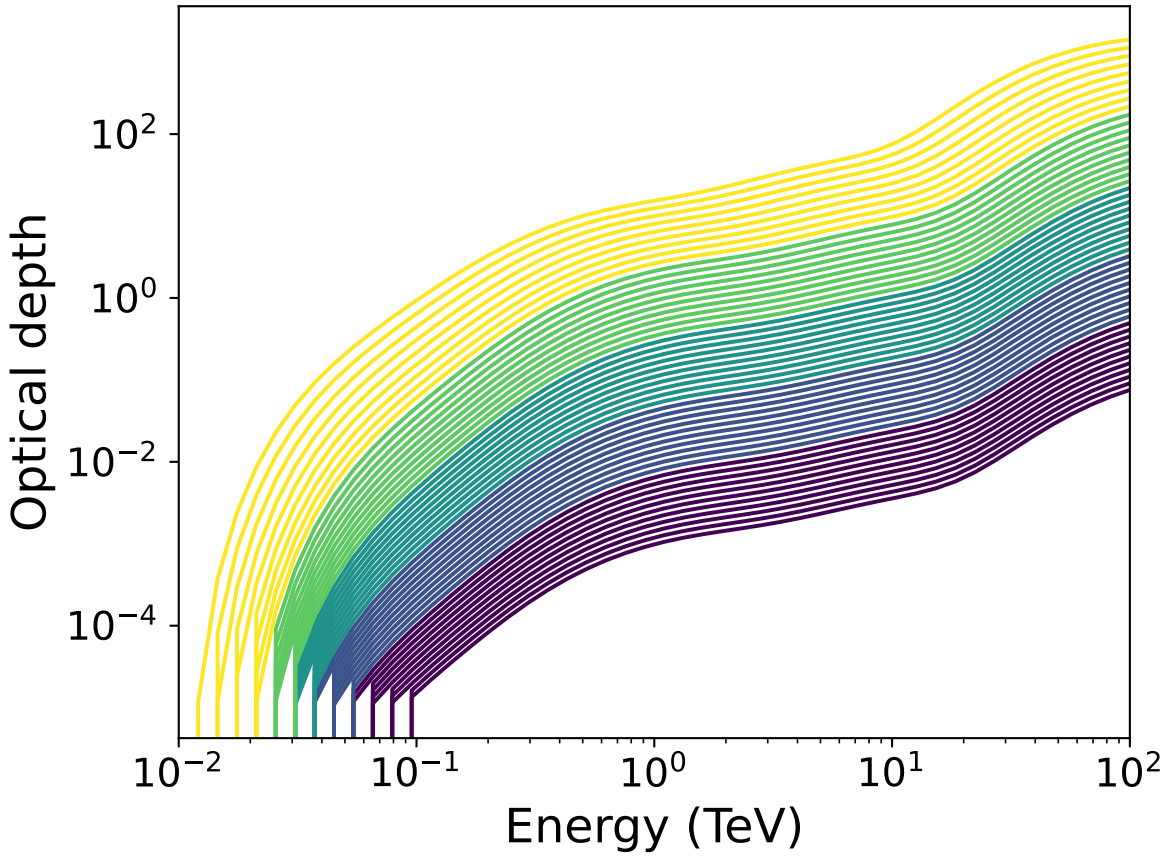

Figure 3.5: The optical depth for 50 logarithmically-spaced redshift values between $z = 10^{-4}$ and $z = 1$, as a function of the gamma ray energy, according to eq. 3.50. The dust fraction values were fixed to $f_{\mathrm{PAH}} = 0.25$, $f_{\mathrm{SG}} = 0.05$ and $f_{\mathrm{LG}} = 0.70$. All curves correspond to the grid with $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. The colours are for better visual inspection, they do not carry meaning.

which is simply derived from basic probability rules. Here, $\mathbf{\Omega} = \{\omega_{\mathrm{EBL}}, \omega_{\mathrm{S}}\}$ is a vector of EBL and source parameters, while I are other information and hypotheses (e.g. the models utilised). The term $p(D|I)$, sometimes called "evidence", is a constant normalisation factor and its value is not important for our goals, as the MCMC method does not require its computation (see section 3.4.1). That is, we are interested in computing

$$p(\mathbf{\Omega}|D, I) \propto p(D|\mathbf{\Omega}, I)p(\mathbf{\Omega}|I). \tag{3.52}$$

The likelihood $p(D|\Omega, I)$ can be constructed from the hypothesis that the measured spectral fluxes of the sources have Gaussian errors, which is a valid assumption for high photon counts. A more low-level analysis which starts from the event list of observations — as discussed in Chapter 7 — would require the use of Poisson statistics. For independent

Grid	Variable	Range	No. of points	Spacing
P19	Energy	0.01–100 TeV	100	log
	Redshift	0.01–6.00	600	linear
P19+Temperature	Energy	0.01–100 TeV	50	log
	Redshift	10^{-4}–1	50	linear
	Dust Temperature	5.93–5930 K	50	log
Cosmology	Energy	0.01–100 TeV	50	log
	Redshift	10^{-4}–1	50	log
	Hubble constant	50–90(*)	41	linear

Table 3.1: Description of the EBL optical depth grids, based on the F10 model. P19 refers to the grid computed by de Matos Pimentel & Moura-Santos (2019), P19+Temp. is a new grid with the addition of dust temperature as a third dimension and "Cosmology" is the grid with the variable Hubble constant. (*) in $\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

spectra, the likelihood is simply the product of Gaussian distributions, as

$$p(D|\mathbf{\Omega}, I) = \frac{1}{Z}\exp\left\{-\frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{n_j}\left[\frac{\Phi_{\rm obs}^{(j)}\left(E_i^{(j)}\right) - \Phi_{\rm mod}^{(j)}\left(E_i^{(j)};\mathbf{\Omega}\right)}{\sigma\left(E_i^{(j)}\right)}\right]^2\right\}, \tag{3.53}$$

where $\Phi_{\rm obs}^{(j)}(E_i^{(j)})$ is the observed flux of the j-th source at the energy bin E_i — with uncertainty $\sigma(E_i^{(j)})$ — and $\Phi_{\rm mod}^{(j)}$ is the respective flux model. Naturally, the sums are over the n_j energy bins of each source and over all N spectra. The factor Z is the normalisation of the probability, which we too will not need to compute.

The model flux

$$\Phi_{\rm mod}(E;\mathbf{\Omega}) \equiv E^2\phi_{\rm mod}(E;\mathbf{\Omega}) = e^{-\tau(E;\omega_{\rm EBL})}E^2\phi_{\rm intr}(E;\omega_{\rm S}), \tag{3.54}$$

is expressed in terms of the differential flux $\phi \equiv dN/(dEdtdA)$ and the optical depth $\tau(E;\omega_{\rm EBL})$, omitting the redshift dependence. The *intrinsic* differential flux, $\phi_{\rm intr}$, refers to the observed flux at Earth after removing the EBL attenuation, so it has no redshift corrections due to the source's distance. Throughout this work, we use up to three different parametrizations for the intrinsic flux: the power law (PL), log-parabola (LP) and power law with exponential cut-off (PLC) models, given by

$$\phi_{\rm intr}(E) = \begin{cases} N_0\left(\frac{E}{E_0}\right)^{-\Gamma} & \text{(PL)} \\ N_0\left(\frac{E}{E_0}\right)^{-a-b\log(E/E_0)} & \text{(LP)} \\ N_0\left(\frac{E}{E_0}\right)^{-\Gamma}\exp(-E/E_{\rm cut}) & \text{(PLC).} \end{cases} \tag{3.55}$$

These are common models used in the literature (e.g. Biteau & Williams 2015). The power law flux usually emerges in acceleration models, while the LP and PLC models are different

ways of parameterizing intrinsic curvature of the spectra if there is some mechanism for local absorption or power limitations. In our analysis, we have set $E_0 = 1\,\mathrm{TeV}$, so the source parameters $\omega_S^{(j)}$ are comprised by two (PL) or three (LP, PLC) variables for each spectrum.

The EBL parameters ω_{EBL} are constituted by the dust fractions (f_{PAH}, f_{SG}, f_{LG}) presented in section 3.1 and, eventually, the PAH temperature Θ_{PAH}, as will be discussed in Chapter 4. In Chapter 5 we also explore the inference of H_0. The dust fractions have a normalisation condition that imposes the constraint

$$f_{\mathrm{PAH}} + f_{\mathrm{SG}} + f_{\mathrm{LG}} = 1, \tag{3.56}$$

which we use as a prior condition to define $f_{\mathrm{LG}} = 1 - f_{\mathrm{PAH}} - f_{\mathrm{SG}}$. In general, we set an uniform prior $p(\omega_{\mathrm{EBL}}|I) = \text{constant}$ in the domain $0 \leq f_{\mathrm{PAH}} + f_{\mathrm{SG}} \leq 1$, with zero probability outside to enforce the constraint of eq. 3.56. The source parameters also have analogous uniform priors which enforce zero probability for negative values and define an upper limits for them. This is mainly done for the practical implementation of the Markov Chain Monte Carlo (MCMC), as an unbounded parameter space may compromise the stability and convergence of the algorithm.

3.4 Markov Chain Monte Carlo

MCMC tools are made for sampling complicated probability distribution functions (Hogg & Foreman-Mackey 2018), which is often the case when we have a parameter space of high dimensionality. Furthermore, in the Bayesian approach, we are interested in the whole shape of the posterior distribution — often marginalised over specific parts of the parameter space, in which we can compute expectation values and probability intervals. MCMC methods allow precisely this, as the marginalisation becomes simply a projection of the sample into the subspaces of interest and a fair sample means we can replace integral computations by sums (Hogg & Foreman-Mackey 2018).

The MCMC technique combines the random sampling (Monte Carlo) with a stochastic process called Markov chain. A Markov chain is a sequence of random variables in which each transition depends only on the current state, so the sampling of the subsequent step does not depend on the chain's past history (Sokal 1997). When a stationary distribution exists to which the Markov chain can converge, the steps taken in this converged state serve as a fair sample of the target distribution. This, however, does not mean that each sample is independent of each other, but the chain is often autocorrelated, meaning the variance of estimated quantities will be larger than the independent case (Sokal 1997). Therefore, some of the fundamental challenges for the practical implementation of MCMC methods are to determine if the chain has converged to the target distribution

and what is the degree of correlation between samples. Another problem it may occur is an initialisation bias if the chain starts far from the target distribution, as this transient period must be identified and eliminated (the burn-in phase).

To deal with these questions, we follow the recommendations from Sokal (1997) and Foreman-Mackey et al. (2013) and use the integrated autocorrelation time, τ_I, as the indicator of statistical errors in the Monte Carlo estimation, given by

$$\tau_I \equiv \sum_{\tau=-\infty}^{\infty} \rho(\tau), \tag{3.57}$$

where $\rho(\tau)$ is the autocorrelation function with lag τ. The integrated quantity essentially determines the number of steps required in a chain to produce each independent sample and can be estimated (Foreman-Mackey et al. 2013) as

$$\hat{\tau}_I = 1 + 2\sum_{\tau=1}^{M} \hat{\rho}(\tau), \tag{3.58}$$

where $\hat{\rho}(\tau)$ is the autocorrelation estimator and the sum is up to $M \ll N$ for a chain with size N. This cut avoids summing too much noise due to the estimated autocorrelation for very large lags. In this work, we use the routines from the package `emcee`[2] (Foreman-Mackey et al. 2013) to estimate τ_I.

Once the integrated autocorrelation time is computed, we also discard the initial burn-in phase of a few τ_I's length to mitigate the initialisation bias. This allows us to obtain the effective sample size (ESS): the total number of steps in the chain divided by the autocorrelation time. Then, we follow Gong & Flegal (2016) in quantifying the Monte Carlo standard error (difference between an estimator and the parameter's true value). For some estimate computed from the MCMC sample (e.g. mean), the idea is to stop the simulation when the width w_i of the confidence interval corresponding to each parameter i is smaller than some threshold. Instead of fixing an absolute value, which is not optimal for multivariate problems, Gong & Flegal (2016) chooses the condition that all confidence intervals are shorter than a fraction ϵ of posterior standard deviations λ_i, that is, $w_i < \epsilon\hat{\lambda}_i$.

By controlling the Monte Carlo error, this approach allows for a practical criterion of terminating an MCMC simulation and can be applied in a multivariate scenario, as discussed by Vats et al. (2015). Asymptotically ($N \to \infty$), the ESS is related to the posterior variance and the variance of the independent sample case σ^2 (Gong & Flegal 2016) as

$$\mathrm{ESS}_i = \frac{N\lambda_i^2}{\sigma_i^2}. \tag{3.59}$$

[2]Available online under the MIT License: `https://github.com/dfm/emcee`

Then, we can apply the condition $w_i < \epsilon\hat{\lambda}_i$ to define the minimum ESS such that, for a given confidence interval of $(1-\alpha)\%$, we achieve a precision ϵ. The `R` package `mcmcse` (Flegal et al. 2021) presents a formula for this computation for multivariate problems, where the minimum ESS is

$$\min \text{ESS} = \frac{2^{2/p}\pi}{[p\Gamma(p/2)]^{p/2}} \frac{\chi^2_{1-\alpha,p}}{\epsilon^2} \tag{3.60}$$

for an MCMC sampling in p dimensions and $\chi^2_{1-\alpha,p}$ is the percent point function of the $\chi^2(p)$ distribution. Conversely, given the effective sample size of ESS $= N/\tau_I$ and α, we can obtain the corresponding relative precision ($\epsilon = 0.02$ is recommended by Flegal & Gong (2013)).

In this work, we use two distinct MCMC methods. In Chapters 4 and 5, we implement the `emcee` code, while in Chapter 6 we apply a Hamiltonian Monte Carlo (HMC) algorithm. The following subsections discuss the details and differences regarding these methods.

3.4.1 emcee

The MCMC Python package `emcee` implements the affine invariant ensemble sampler from Goodman & Weare (2010), which is an alternative approach to the Metropolis-Hastings (M-H) algorithm (Metropolis et al. 1953; Hastings 1970). The M-H method consists of a proposal function — from which we sample the proposed next step in the Markov chain — and a criterion of acceptance. One of the simplest implementations uses a Gaussian proposal $\mathcal{N}(\theta'|\theta, \Sigma)$, which is symmetric over the exchange of the current parameters θ to the proposed ones θ'. Then, the proposed step is accepted with probability

$$\min\left(1, \frac{\pi(\theta')}{\pi(\theta)}\right), \tag{3.61}$$

where π is the distribution we are sampling. This criterion more clearly shows that only the ratio of the posterior probability distributions is relevant for our MCMC operation, explaining why we have neglected the constant factors in section 3.3. What Goodman & Weare (2010) does to improve the M-H prescription is to create a number of chains such that a subset of them (an ensemble) is used to evolve the position of the complementary ensemble[3]. This is described as the "stretch move", as it randomly selects a chain from the complementary ensemble and updates the current ensemble chain based on the parameter distance stretched by a number draw from another distribution. This algorithm was implemented in `emcee` to allow for the simultaneous update of the chains in parallel

[3]It also has the property of being an algorithm which makes the ensemble invariant to affine transformations, which preserves the performance in anisotropic distributions.

(Foreman-Mackey et al. 2013), which speeds up the process and it also benefits from the smaller integrated autocorrelation time compared to the standard M-H, as discussed by Goodman & Weare (2010).

As of the writing of this text, the current version of `emcee` (v.3.0.0 and above) allows for different choices of the "move" proposal, beyond the stretch move. In our investigation, the Kernel Density Estimator (KDE) produced the lowest integrated autocorrelation times, so we have used it in all MCMC runs of Chapters 4 and 5. The KDE uses a non-parametric estimation to obtain the proposal distribution and evolve the simulation, but it usually requires a large number of chains — also called *walkers* — to achieve a good convergence.

3.4.2 Hamiltonian Monte Carlo

According to Huijser et al. (2015), affine invariant ensemble methods may have problems when dealing with a large number of parameters (more than 50). In our work, it was also becoming increasingly costly, in computational time, to increase the number of parameters by adding more gamma-ray sources. Under this motivation, we utilised a different algorithm that allowed us to use a larger data set: the Hybrid or Hamiltonian Monte Carlo.

HMC incorporates information about the geometry of the target distribution to more efficiently explore its typical set (Betancourt 2017). The idea behind it is that the goal of MCMC methods is to be able to approximate the computation of expectation values (integrals) with discrete sums from a fair sample. For such computations, it would be interesting to prioritise regions that numerically contribute the most to the expectation value. Naturally, our intuition says such regions must be closer to the mode of the distribution, while low probability regions contribute less. However, in spaces with high dimensionality, we must also be aware of the comparatively small volume in the neighbourhood of the mode to the volume of the whole sampling space. More precisely, the integration of a probability density depends on the product $\pi(\theta_n)\mathrm{d}^n\theta$, where $\mathrm{d}^n\theta$ is the volume element in an n-dimensional parameter space. In spherical coordinates of arbitrary dimension, it is easy to see that $\mathrm{d}^n\theta$ depends on r^{n-1}, where r is the distance coordinate. By centring the coordinate system in the mode, we see that the volume grows indefinitely far away from it, so the most relevant contribution to the integral may come from adjacent regions where the growth in volume can compensate the lower probability density. This middle ground between the density peak and distant large volume contributions is what constitutes the typical set, which an MCMC sampler should focus to more efficiently compute expectation values.

We have already seen a strategy to quantify the typical set: the M-H algorithm. Being purely stochastic, together with the fact that the volume exterior to the typical set is

overwhelmingly larger in high dimensions (Betancourt 2017), means that often a proposed step will be outside the typical set and rejected, so the chain will not move or will need to move very slowly to prevent it to be stuck (worsening the overall efficiency). HMC presents a solution to this problem, by finding a vector field that allows a guided exploration of the typical set. The answer to this comes from differential geometry (Betancourt 2017), or — to give a physical analogy — from Hamiltonian dynamics. Hamilton's equations have three important properties for MCMC algorithms (Neal 2011): reversibility (which leaves the target distribution invariant); conservation of the Hamiltonian, for explicit time-independent Hamiltonians (which would result in an acceptance probability of 1); and conservation of volume, a consequence of the more general property of symplecticness (so no change in volume needs to be accounted in the acceptance criterion).

The foundation of HMC is to map our original problem of sampling an N-dimensional space to a problem in $2N$ dimensions, by adding conjugate momenta to our original set of parameters (generalised positions). This means our target distribution can be written as

$$\Pi(\mathbf{q}, \mathbf{p}) = \tilde{\pi}(\mathbf{p}|\mathbf{q})\pi(\mathbf{q}), \tag{3.62}$$

where now $\mathbf{q}$ contains our parameters of interest and $\mathbf{p}$ is their conjugate momenta. This probability can be written in terms of the canonical distribution

$$\Pi(\mathbf{q}, \mathbf{p}) = \frac{1}{Z} e^{-H(\mathbf{q},\mathbf{p})}, \tag{3.63}$$

so the Hamiltonian is defined as

$$H(\mathbf{q}, \mathbf{p}) = -\ln(\Pi(\mathbf{q}, \mathbf{p})Z) = -\ln(\tilde{\pi}(\mathbf{p}|\mathbf{q})) - \ln(\pi(\mathbf{q})) - \ln(Z). \tag{3.64}$$

Since the acceptance criterion, as M-H, depends on the ratio of the probabilities (differences of Hamiltonians), we can neglect the constant term from the normalization Z. This means we can associate kinetic and potential energies

$$\begin{aligned} K(\mathbf{p}|\mathbf{q}) &= -\ln(\tilde{\pi}(\mathbf{p}|\mathbf{q})) \\ U(\mathbf{q}) &= -\ln(\pi(\mathbf{q})). \end{aligned} \tag{3.65}$$

Naturally, in our case, $\pi(\mathbf{q}) = p(\mathbf{\Omega}|D, I)$, so the potential energy is the negative of the logarithm of the posterior distribution. Furthermore, we adopt the Euclidean-Gaussian kinetic energy (Betancourt 2017)

$$K(\mathbf{p}|\mathbf{q}) \equiv K(\mathbf{p}) = \frac{1}{2}\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p} \tag{3.66}$$

which is the simplest implementation. Here, $\mathbf{M}$ is the "mass" matrix, that will be discussed

in Chapter 6. This leads to Hamilton's equations

$$\frac{\mathrm{d}\Omega_i}{\mathrm{d}t} = \frac{\partial H}{\partial p_i} = (\mathbf{M}^{-1}\mathbf{p})_i \tag{3.67}$$

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = -\frac{\partial H}{\partial \Omega_i} = \frac{\partial}{\partial \Omega_i} \ln(p(\mathbf{\Omega}|D, I)) \tag{3.68}$$

for each pair of parameters (Ω_i, p_i). The explicit expressions for the posterior derivatives are presented in section E.1 of Appendix E.

The algorithm of HMC proceeds like this (see e.g. Neal 2011): given the initial coordinates, momentum values are randomly generated, specifying the "energy" level of the current state. This is done from the definition of the kinetic energy, as eq. 3.66 implies $\tilde{\pi}(\mathbf{p}|\mathbf{q})$ is a Gaussian distribution for $\mathbf{p}$ with mean zero and covariance matrix $\mathbf{M}$. Then, Hamilton's equations are solved to evolve coordinates $(\mathbf{q}, \mathbf{p}) \to (\mathbf{q}', \mathbf{p}')$ for a specified integration time. Finally, the momentum variables are changed to the opposite sign and the proposed step is accepted with probability

$$\min\left[1, \exp(-H(\mathbf{q}', -\mathbf{p}') - H(\mathbf{q}, \mathbf{p}))\right]. \tag{3.69}$$

The negation of the momenta is just a formal operation to ensure the Hamiltonian transition to be reversible, but it does not have practical consequences in this context, since $K(-\mathbf{p}) = K(\mathbf{p})$. Besides, regardless if the proposal is accepted or rejected, new momentum coordinates are sampled at each iteration. Therefore, HMC combines a stochastic exploration of the parameter space (entering random energy levels of the canonical distribution) with deterministic trajectories that allow moving the chain close to the typical set.

In practice, the integration of Hamilton's equations requires numerical methods that can — as best as possible — conserve the energy along the trajectory, such as symplectic integrators. We use the fourth-order Forest-Ruth integrator (Forest & Ruth 1990) for this task, as it is a robust method with simple implementation. The explicit algorithm can also be found in Das et al. (2015). It consists of a sequence of updates in the coordinates and momenta, by discretizing time in steps ϵ, which is chosen to achieve good numerical precision. HMC requires other tuning, such as the choice of the mass matrix and the number of discrete time steps to reach a new proposal. A more detail discussion on these topics is presented in Chapter 6.

Chapter 4

Accuracy and precision of EBL constraints through a multi-source analysis

In this chapter, we investigate simultaneous EBL and intrinsic spectral parameters constraints through a multi-source combined analysis, using a sample of synthetic blazars. Our focus is to understand the accuracy and precision of the inferred parameters. We use two data sets: one with a "short" observation time (total exposure of less than 5h per source) and a "long" one (more than 5h), firstly probing the dust fractions in section 4.3 and then the PAH temperature in section 4.3.1. Secondly, we repeat the analysis in section 4.4 for new data sets generated with a different EBL attenuation model to study systematic errors. We finalise in section 4.5 with a discussion on the constraints on the intrinsic spectral indices.

4.1 Synthetic Sources

The distribution of extragalactic gamma-ray sources in the universe, such as blazars, can be described by a luminosity function — the number density of objects per luminosity, redshift and spectral index (in the GeV range). To start with an approximately realistic distribution of sources that will be observed in the future by CTA, we consider the parametrization from Ajello et al. (2014). This luminosity function was built from the first year data catalogue of *Fermi*–LAT. Figure 4.1 shows the resulting distribution of redshift, luminosities, and spectral indices, sampled from the luminosity function. As *Fermi*–LAT is only sensible up to high energies (hundreds of GeV), the VHE intrinsic emission of these sources was assumed to be a power law extrapolation of their spectral index from GeV to the TeV range. This hypothesis is not necessarily valid, as some softening of the spectrum can be expected, above the SSC peak. However, our goal is not to perform a precise forecast of future observations, but to understand how EBL and intrinsic parameter constraints evolve in a combined fit.

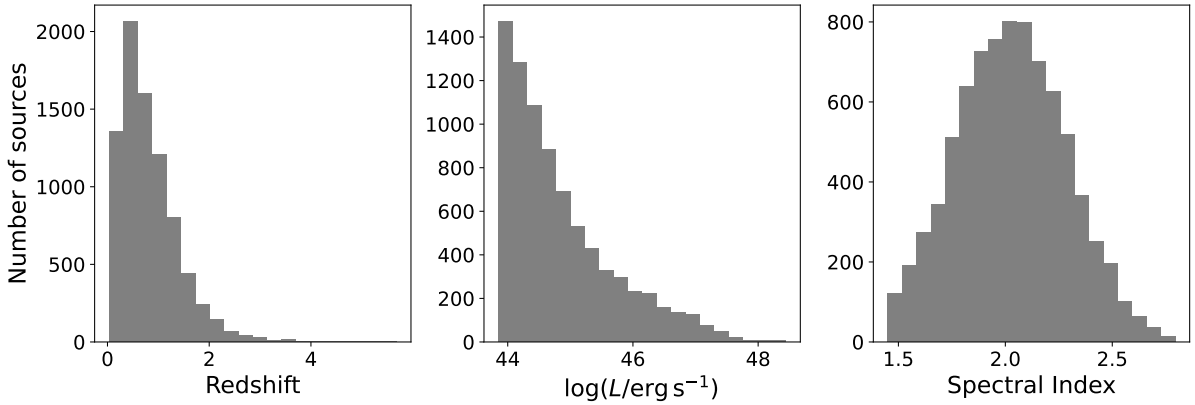

Figure 4.1: Distribution of redshift, luminosities (L) and spectral indices from a sample of 8033 BL Lacs according to the luminosity function from Ajello et al. (2014).

The observation strategy for these sources was defined as follows. A large sample of BL Lacs from the luminosity function was uniformly distributed in the celestial sphere. Then, a uniform pointing grid in an area corresponding to 25% of the sky, 5° above the Galactic plane, was determined to be observed by the South and North observatories of CTA. This was inspired by the operation of the Extragalactic Survey, one of the Key Science Projects of the CTA collaboration (Cherenkov Telescope Array Consortium et al. 2019). Having an estimate 1000 h of total observation time divided between both hemispheres, this period was distributed to each pointing of the South and North arrays, proportional to the area visible by each observatory. This resulted typically in observations no longer than 5 hours. This approximately realistic pointing strategy was important to achieve a plausible distribution of offsets (angular distance between the position of the source and the camera centre), which is a variable that affects the significance of detection. Then, the CTA observation was simulated[1] using the software `ctools` (Knödlseder et al. 2016) and the instrument response function `prod3b-v2` Observatory & Consortium (2016). The significance of detection was computed according to the likelihood ratio test, as presented by Li & Ma (1983). In general terms, we can compute the test-statistics TS

$$\mathrm{TS} \equiv -2(\ln \mathcal{L}_0 - \ln \mathcal{L}_1), \tag{4.1}$$

where $\mathcal{L}_0$ is the likelihood under the null hypothesis (pure background) and $\mathcal{L}_1$ is the maximum likelihood when both diffuse background and a point-like source signal are present. The square-root of the TS can be used to quantify the "number of sigmas" of the rejection of the null hypothesis (no source detection). In Chapter 7 we provide a detailed explanation on the subject in the context of On/Off measurements. For sources with TS > 25, we have created another data set by increasing the total observation time (to

[1]The simulation effort was mainly coordinated by Dr. Luiz Augusto Stuani, so we greatly appreciate his contribution.

Variable	$f_{\rm PAH}$	$f_{\rm SG}$	$f_{\rm LG}$	$T_{\rm PAH}$	$T_{\rm SG}$	$T_{\rm LG}$
Benchmark value	0.25	0.05	0.70	450 K	70 K	40 K

Table 4.1: Benchmark values of the parameters describing the EBL intensity of the F10 EBL model, as implemented by de Matos Pimentel & Moura-Santos (2019).

5h per pointing). We refer to these sets as the "short" (less than 5h in total) and "long" (more than 5h) observation time data.

The gamma-ray photon flux was also attenuated by two different EBL models: F10 and the model from Domínguez et al. (2011) (D11). For the F10 implementation, we have considered the nominal dust fraction and temperature values from de Matos Pimentel & Moura-Santos 2019 (P19), reproduced in Table 4.1, which will be our benchmark model. This set of parameters will be sometimes referred as F10+P19, since there are small differences in the actual dust fraction values when compared to the original F10 publication. In other words, F10+P19 should be read as "the F10 model computationally reproduced by P19".

The characteristics of the most significant sources in our sample (redshift, spectral index, TS, total observation time and maximum energy bin) can be found in Table 4.2 for sources attenuated by the F10 model. When the sources were attenuated by D11, there are some fluctuations in the TS values and a small change in ordering, but this information is not relevant for our studies.

4.2 Marginalised posterior distribution

In this chapter, we make a modification to the posterior probability presented in section 3.3 by performing an analytical marginalisation over the flux normalisation parameters, N_0. Since all sources are described by the PL model, only the spectral indices remain to be determined by the MCMC sampling, besides the EBL parameters. This approach was a solution to the increased difficulty of sampling in a space of high dimensionality, as chains needed to be progressively longer to produce a reasonable number of effective samples. After determining the EBL and the spectral indices from the posterior, it is possible to perform a one parameter fit to the spectrum and obtain the unknown normalisation. This procedure means we lose the information about the probability distribution of this parameter and its immediate correlations in the posterior, but it can be useful if the focus is on EBL properties. It also significantly speeds up the MCMC simulations, as it reduces the dimensionality of the parameter space by almost half (or one third if the LP and PLC models are used).

To perform the marginalisation of the posterior, we assume the independence of

ID	z	$\Gamma_{\rm true}$	Obs. T. (h)	$E_{\rm max}$(*)	TS	Obs. T. (h)	$E_{\rm max}$(*)	TS
A	0.051	1.578	0.98	21.13	15528	5	29.85	85362
B	0.038	1.873	1.96	21.13	15499	5	21.13	39116
C	0.058	1.587	3.92	14.96	12095	20	21.13	48921
D	0.061	1.714	3.92	14.96	10250	20	14.96	44412
E	0.079	1.485	0.98	14.96	9691	5	21.13	51800
F	0.076	1.477	0.98	12.59	4081	20	21.13	19381
G	0.083	1.899	2.21	12.59	2887	15	10.59	9236
H	0.125	1.503	1.96	5.31	2054	10	5.31	11936
I	0.192	1.705	3.92	2.66	1949	20	10.59	9483
J	0.077	1.623	3.92	14.96	1701	20	14.96	7070
K	0.092	1.514	3.92	10.59	1316	20	10.59	5811
L	0.076	1.800	3.92	7.50	1130	20	14.96	5193

Table 4.2: Source parameters (redshift z and spectral index $\Gamma_{\rm true}$) of the synthetic sample described by a PL intrinsic spectrum and EBL attenuation according to the F10 model as implemented by de Matos Pimentel & Moura-Santos (2019). Each source has a letter label for identification. For the "short" (left side) and "long" (right side) observation times (Obs. T., in hours), we show their respective TS values and the maximum energy bin of their spectra. (*) in TeV.

the normalisation parameters. Explicitly, we want to perform the computation of

$$p(\mathbf{\Omega}_r|D,I) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\prod_{j=1}^{N}\mathrm{d}N_0^{(j)}\right)\frac{p(D|\mathbf{\Omega},I)p(\mathbf{\Omega}|I)}{p(D|I)}, \tag{4.2}$$

where $\mathbf{\Omega}_r$ are all the parameters except the flux normalisation ones. We also adopt an improper prior $p(N_0^{(j)}|I) = 0$ for $N_0^{(j)} < 0$, and constant otherwise. Therefore, we simply need to compute

$$p(D|\mathbf{\Omega}_r,I) = \int_{0}^{\infty}\cdots\int_{0}^{\infty}\left(\prod_{j=1}^{N}\mathrm{d}N_0^{(j)}\right)\frac{1}{Z}\exp\left(-\frac{1}{2}\chi^2\right), \tag{4.3}$$

where

$$\chi^2 = \sum_{j=1}^{N}\sum_{i=1}^{n_j}\left[\frac{\Phi_{\mathrm{obs}}^{(j)}\left(E_i^{(j)}\right) - \Phi_{\mathrm{mod}}^{(j)}\left(E_i^{(j)};\mathbf{\Omega}\right)}{\sigma\left(E_i^{(j)}\right)}\right]^2. \tag{4.4}$$

The explicit computation and the final expression can be found in Appendix B.

4.3 EBL constraints using synthetic sources

To investigate the potential of constraining EBL properties while simultaneously inferring spectral parameters of gamma-ray sources, we perform four sets of MCMC simulations using the 12 highest TS sources attenuated by F10+P19 model. Firstly, we consider the data set of "short" total observational time, sampling the posterior probability of EBL's dust fractions and spectral indices of the sources. Secondly, we use the spectra taken with the longer observational period. Then, both cases are repeated while also letting the PAH's blackbody temperature, T_{PAH}, to be a free parameter.

For this analysis, we use the ensemble sampler `emcee` described in section 3.4.1 and perform MCMC runs starting with the two sources with highest TS. Then, we sequentially add the next highest TS source, up to 12 spectra. Thus, we can investigate the evolution of EBL constraints from combining progressively more data. We also adopt the posterior distribution marginalised over the normalisation parameters. The statistical properties of all results can be found in Table C.1 of Appendix C. We were able to maintain an ESS sufficiently large to achieve a precision ϵ of less than 2% for a confidence interval of 95%. We also have tried to keep the acceptance fraction (fraction of proposed steps accepted during the chain) between 0.2–0.5, following the recommendation from Foreman-Mackey et al. (2013). However, for more than 10 sources, we have usually found higher autocorrelations — in general associated with low acceptance fractions. Even increasing the number of parallel chains (*walkers*), we did not quite reach the optimal range for the acceptance fraction. Nevertheless, it was still possible to obtain a fair amount of effective

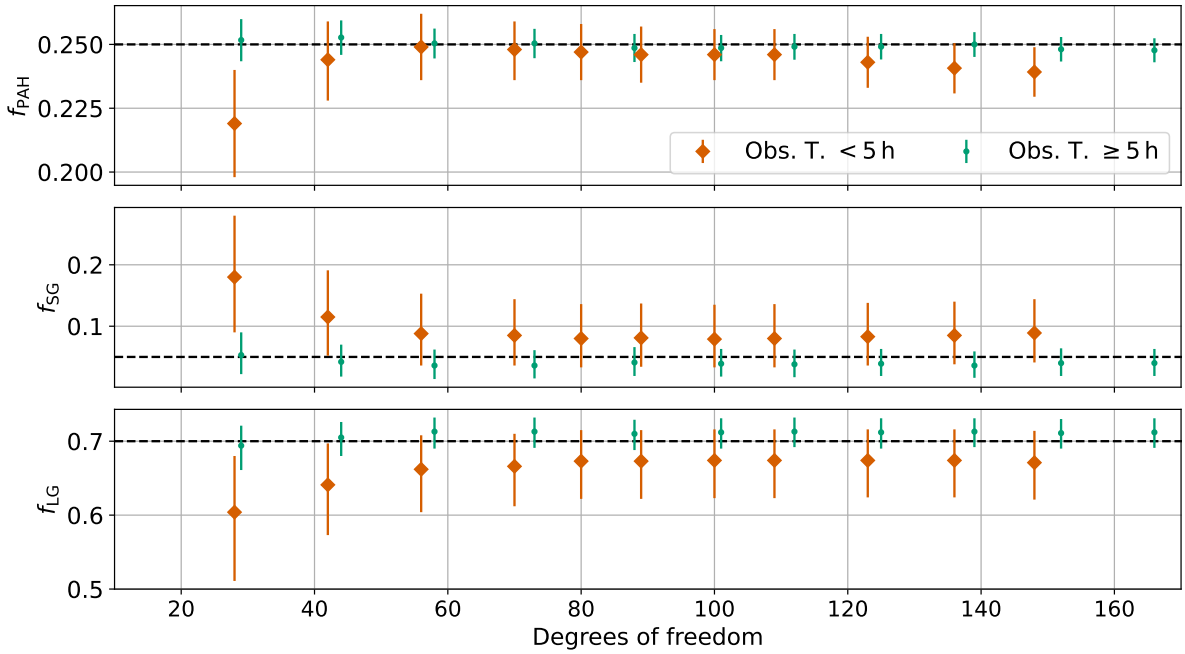

Figure 4.2: Median values of the dust fractions as a function of the number of degrees of freedom in each simulation. The leftmost point correspond to the case with the 2 highest TS sources and the rightmost with all 12 spectra. The horizontal dashed lines are the true values of the F10+P19 attenuation.

samples, which was the priority.

Once the MCMC samples were obtained, we took the one dimensional marginal projections of the posterior distribution and have computed the median values for all parameters. We also associate the uncertainty to be the range between the 16th and 84th percentiles (a 68% interval). This is suggested by Hogg & Foreman-Mackey (2018), since it ensures the central value to be contained in the uncertainty range (as opposed to the mean in highly skewed distributions). Besides, in these asymmetric distributions, the median is insensitive to outliers, unlike the mean. However, it is important to make a distinction between these statistics and the "best fit" values. In spaces of higher dimensions, it is possible that the median (or mean) values of all parameters, with respect to their marginal distributions, are not a good fit to the overall data (Hogg & Foreman-Mackey 2018). That is why it is also important to compare the result to a random selection of parameters from the posterior sample, as this can better informs us if the high probability regions of the joint posterior are a good representation of the data.

Regarding the EBL constraints, the evolution of the dust fractions with each successive simulation (from 2 to 12 sources) can be seen in Figure 4.2, comparing the short and long observational time data sets. The median values with respective uncertainties are plotted as a function of the number of degrees of freedom (ndof). Naturally the data with more observation time — and higher statistical significance — result in more flux points, each point with higher photon counts on average, increasing the ndof for the same

number of sources.

Two main effects of adding new sources in the likelihood can be identified: a more accurate inference of EBL parameters (in the absence of systematic errors), approaching the true values marked as the horizontal dashed lines; and it also results in a more precise estimation, since the uncertainties get progressively smaller. Looking at the low observational time results, after 4 sources with the highest TS, all three parameters get closer to their respective true values (less than 1σ). The exact values and uncertainties can be found in Table 4.3. Notably, the improvement in precision is best seen in the PAH component, as the uncertainty diminishes until 12 sources, while for the other components the uncertainties are somewhat stable beyond 7 sources. In fact, the PAH is the best constrained parameter, reaching the smallest uncertainty relative to the central value. Another remark is that the inferred fractions are systematically below or above the true values. We should stress, however, that the points are not independent from each other, as one result contains the data of the previous. Besides, the values are dominated by the few highest TS sources, generating the trend. Finally, with the data set of higher observational time, an expressive reduction of the uncertainties in all parameters is achieved, even with a lower amount of sources. In particular, the result with the two highest TS sources already produced an estimate more accurate and precise than the case with lower observational time with more sources.

Another way of discussing these results is through the corner plot in Figure 4.3. In it, we show the 1- and 2-dimensional marginal distributions of EBL parameters. We see few differences between the results of 5 and 12 sources, specially compared to the one with 2 spectra, as the distributions become more well defined. For the 2-dimensional distributions, they reveal a strong negative correlation between the SG and LG components ($\rho_{\text{SG-LG}} = -0.98$ for 12 sources). This strong correlation may be related to the difficulty in disentangling the small and large grain components, since their contribution to the EBL density essentially overlaps in wavelength (e.g. Figure 3.1). Only with the longer observational time data it was possible to obtain more accurate estimates of the SG and LG parameters, probing the far-IR portion of the EBL. We identify a much softer correlation between PAH and SG, of $\rho_{\text{PAH-SG}} = -0.52$, and a positive correlation between PAH and LG of $\rho_{\text{PAH-LG}} = +0.36$. Complementary to this discussion, Figure 4.4 presents the marginal distributions of the higher observational time case.

Therefore, we have found that the addition of more spectra, with progressively lower TS, improves the EBL constraints by converging the model to a configuration — in this case where no systematic errors are introduced — closer to the true EBL model. Such improvements, however, become less pronounced as new sources are added, suggesting a limit on how well one can constrain the EBL. This limit can be surpassed if higher TS observations are made, producing better measured spectra (more flux points and lower uncertainties). We then found that the group of sources observed for more than 5 h per

Short Observation Time: F10+P19 (dust fractions)				
N.S.	ndof	$f_{\rm PAH}$	$f_{\rm SG}$	$f_{\rm LG}$
2	28	$0.219^{+0.021}_{-0.021}$	$0.18^{+0.10}_{-0.09}$	$0.604^{+0.076}_{-0.093}$
3	42	$0.244^{+0.015}_{-0.016}$	$0.115^{+0.076}_{-0.063}$	$0.641^{+0.056}_{-0.068}$
4	56	$0.249^{+0.013}_{-0.013}$	$0.088^{+0.065}_{-0.052}$	$0.662^{+0.046}_{-0.058}$
5	70	$0.248^{+0.011}_{-0.012}$	$0.085^{+0.059}_{-0.049}$	$0.666^{+0.044}_{-0.054}$
6	80	$0.247^{+0.011}_{-0.011}$	$0.080^{+0.056}_{-0.047}$	$0.673^{+0.042}_{-0.051}$
7	89	$0.246^{+0.011}_{-0.011}$	$0.081^{+0.056}_{-0.047}$	$0.673^{+0.042}_{-0.051}$
8	100	$0.246^{+0.010}_{-0.010}$	$0.079^{+0.056}_{-0.046}$	$0.674^{+0.042}_{-0.051}$
9	109	$0.246^{+0.010}_{-0.010}$	$0.080^{+0.056}_{-0.047}$	$0.674^{+0.042}_{-0.051}$
10	123	$0.243^{+0.010}_{-0.010}$	$0.083^{+0.055}_{-0.047}$	$0.674^{+0.042}_{-0.050}$
11	136	$0.2407^{+0.0098}_{-0.0099}$	$0.085^{+0.055}_{-0.047}$	$0.674^{+0.042}_{-0.050}$
12	148	$0.2392^{+0.0097}_{-0.0097}$	$0.089^{+0.055}_{-0.048}$	$0.671^{+0.043}_{-0.050}$
Long Observation Time: F10+P19 (dust fractions)				
2	29	$0.2517^{+0.0082}_{-0.0083}$	$0.053^{+0.037}_{-0.031}$	$0.694^{+0.027}_{-0.033}$
3	44	$0.2527^{+0.0067}_{-0.0068}$	$0.042^{+0.028}_{-0.024}$	$0.705^{+0.021}_{-0.025}$
4	58	$0.2504^{+0.0058}_{-0.0059}$	$0.036^{+0.026}_{-0.022}$	$0.713^{+0.019}_{-0.023}$
5	73	$0.2504^{+0.0057}_{-0.0058}$	$0.036^{+0.025}_{-0.021}$	$0.713^{+0.019}_{-0.022}$
6	88	$0.2486^{+0.0055}_{-0.0055}$	$0.041^{+0.025}_{-0.022}$	$0.710^{+0.019}_{-0.022}$
7	101	$0.2486^{+0.0051}_{-0.0052}$	$0.039^{+0.024}_{-0.021}$	$0.712^{+0.019}_{-0.022}$
8	112	$0.2491^{+0.0050}_{-0.0051}$	$0.038^{+0.024}_{-0.021}$	$0.713^{+0.019}_{-0.021}$
9	125	$0.2491^{+0.0050}_{-0.0051}$	$0.039^{+0.024}_{-0.020}$	$0.712^{+0.019}_{-0.022}$
10	139	$0.2500^{+0.0048}_{-0.0049}$	$0.036^{+0.023}_{-0.020}$	$0.713^{+0.018}_{-0.021}$
11	152	$0.2481^{+0.0048}_{-0.0048}$	$0.040^{+0.024}_{-0.021}$	$0.711^{+0.019}_{-0.021}$
12	166	$0.2477^{+0.0047}_{-0.0047}$	$0.040^{+0.023}_{-0.021}$	$0.712^{+0.019}_{-0.021}$

Table 4.3: EBL parameters from the combined inference of the highest TS simulated sources (described by the PL intrinsic model and attenuated by F10+P19 EBL). Upper half: simulations using spectra with total observed time lower than 5 h. Lower half: spectra with total observed time equal or greater than 5 h. N.S. refers to the number of sources. Values are the median from their marginal distribution, while uncertainties represent the 68% interval around it.

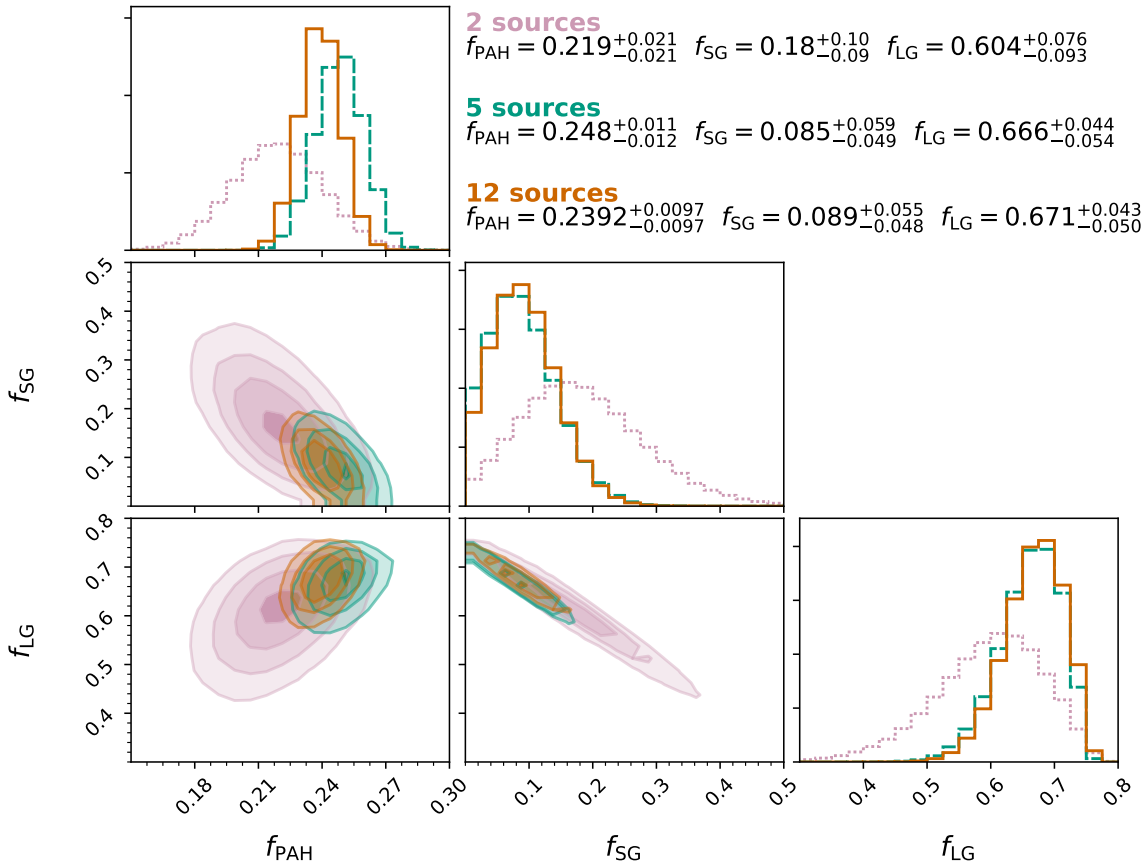

Figure 4.3: 1- and 2-dimensional marginal distributions of EBL parameters using 2, 5 and 12 highest TS sources attenuated by F10+P19's EBL model. The distribution of the LG fraction is simply obtained by the constraint $f_{LG} = 1 - f_{PAH} - f_{SG}$. The inferred values correspond to the median and the 68% interval around it.

pointing is essential for improving constraints in the SG and LG fractions, as the higher energy bins in their spectra may interact more strongly with the IR portion of the EBL. Nevertheless, the VHE data can greatly constrain the PAH fraction — corresponding to the mid-IR band of the EBL.

4.3.1 Constraints on the PAH temperature

As previously shown, the PAH fraction is the best constrained parameter in the EBL, so we investigate how the results differ if the blackbody temperature, T_{PAH}, controlling the emissivity coming from the PAH component, is also included as a free parameter. Figure 4.5 shows the evolution of the median values and Table 4.4 presents the specific values. Differently from the situation of fixed temperatures, considering the short observation time sample, the 68% interval around the median mostly does not contain the true values, even in the simulation with 12 sources. Since the PAH temperature is systematically above the true value, this is compensated by a higher SG median fraction (above 0.15

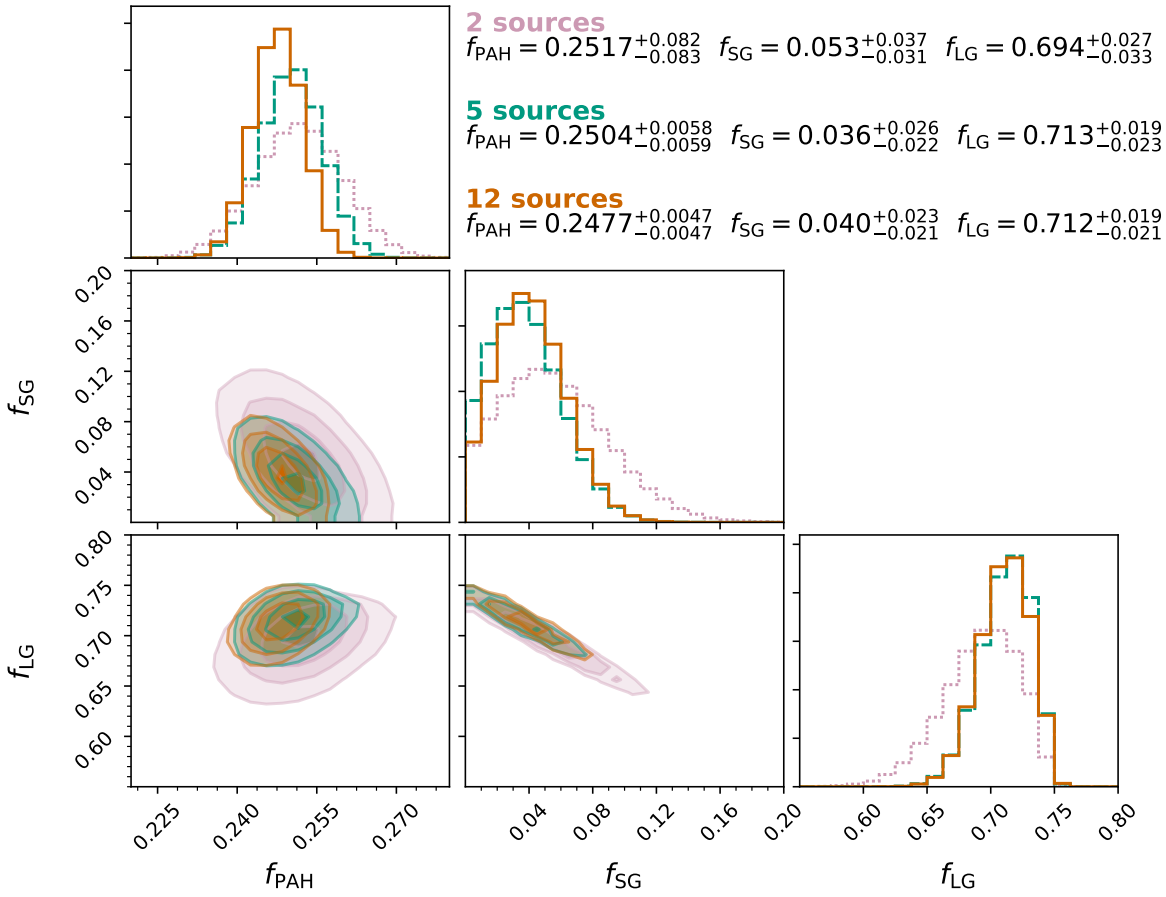

Figure 4.4: Same as Figure 4.3, but for the longer observation time data.

in all cases), as a hotter PAH component shift the mid-IR EBL density toward shorter wavelengths, leaving the EBL density at longer wavelengths to be replaced by the emission from small grains. This behaviour essentially disappears when we use observations above 5 hours, revealing that it is possible to have a good constraint on the position of the EBL's mid-IR peak through the temperature of the PAH component, determining it with a resolution of a few dozen Kelvins. Still, some saturation in the resolution of the median values are observed, with the results using 5 sources being compatible to the 12 sources case.

We have found a similar overlap of the marginal distributions, in Figure 4.6, for the SG and LG fraction components, besides the PAH temperature, showing that the EBL constraints are dominated by few of the highest TS sources. The correlations between the dust fractions are also similar to the case with fixed temperatures, but we also identify that $T_{\rm PAH}$ correlates weakly with $f_{\rm PAH}$, as $\rho_{\text{T-PAH}} = -0.067$, while it has a stronger correlation with the SG component ($\rho_{\text{T-SG}} = +0.71$) and the LG ($\rho_{\text{T-LG}} = -0.74$). For the longer observational time, we identify similar correlations and we find that the PAH fraction, in particular, is already well constrained with only 2 sources. The marginal distributions

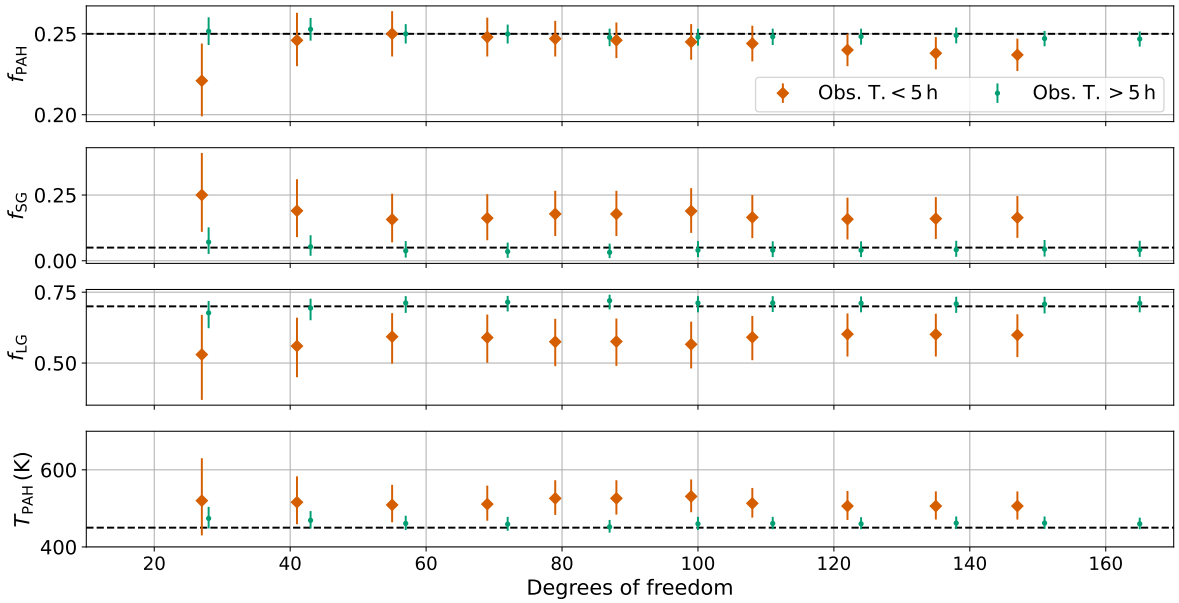

Figure 4.5: Same as Figure 4.2, but with the PAH dust temperature as a free parameter.

can be found in Figure C.1 of Appendix C. The addition of other sources also contribute for reducing uncertainties in all parameters and the estimate of the PAH temperature is more accurate than the short observational time case, with the median less than 1σ apart from the true value.

4.4 Systematic effects associated to the modelling of the EBL density

The previous analyses assumed that the true EBL corresponds exactly to the F10 model. In such a scenario where there is no systematic effect associated with the modelling of the EBL density, we were able to successfully recover the EBL dust fractions and PAH temperature. Now we turn our attention to a situation where the same sources were attenuated by Domínguez et al. (2011) (D11) EBL model and we apply the same methodology to infer the dust fractions and PAH temperature, discussing how adequately they can describe this EBL spectrum and the intrinsic parameters of the sources. Table C.2 in Appendix C shows the statistical properties of the MCMC simulations.

Starting with the case of fixed temperatures, Figure 4.7 presents the evolution of the dust fractions and Table 4.5 shows the exact median values and uncertainties. We can see that the addition of sources in the likelihood still has the effect of progressively reducing the uncertainties — in particular for the PAH fraction —, but now the median values converge to different levels compared to the F10+P19 attenuation. While the PAH fraction is slightly lower, there is a large increase in the SG contribution accompanied by a decrease in the LG fraction. This can be better understood by looking at the local

Short Observation Time: F10+P19 (dust fractions + PAH temperature)					
N.S.	ndof	$f_{\rm PAH}$	$f_{\rm SG}$	$f_{\rm LG}$	$T_{\rm PAH}$ (K)
2	27	$0.221^{+0.023}_{-0.022}$	$0.25^{+0.16}_{-0.14}$	$0.53^{+0.14}_{-0.16}$	520^{+110}_{-90}
3	41	$0.246^{+0.017}_{-0.016}$	$0.19^{+0.12}_{-0.10}$	$0.56^{+0.10}_{-0.11}$	516^{+67}_{-57}
4	55	$0.250^{+0.014}_{-0.014}$	$0.157^{+0.098}_{-0.087}$	$0.593^{+0.083}_{-0.095}$	509^{+52}_{-45}
5	69	$0.248^{+0.012}_{-0.012}$	$0.162^{+0.091}_{-0.084}$	$0.590^{+0.081}_{-0.089}$	511^{+48}_{-43}
6	79	$0.247^{+0.011}_{-0.011}$	$0.178^{+0.088}_{-0.084}$	$0.575^{+0.081}_{-0.086}$	526^{+47}_{-43}
7	88	$0.246^{+0.011}_{-0.011}$	$0.178^{+0.088}_{-0.084}$	$0.576^{+0.081}_{-0.086}$	526^{+47}_{-42}
8	99	$0.245^{+0.011}_{-0.011}$	$0.189^{+0.087}_{-0.083}$	$0.566^{+0.080}_{-0.085}$	531^{+44}_{-41}
9	108	$0.244^{+0.011}_{-0.011}$	$0.165^{+0.085}_{-0.079}$	$0.591^{+0.075}_{-0.081}$	513^{+40}_{-37}
10	122	$0.240^{+0.010}_{-0.010}$	$0.158^{+0.082}_{-0.077}$	$0.602^{+0.073}_{-0.079}$	506^{+39}_{-36}
11	135	$0.238^{+0.010}_{-0.010}$	$0.160^{+0.082}_{-0.077}$	$0.601^{+0.073}_{-0.078}$	506^{+38}_{-35}
12	147	$0.237^{+0.010}_{-0.010}$	$0.164^{+0.082}_{-0.077}$	$0.599^{+0.073}_{-0.078}$	506^{+38}_{-35}
Long Observation Time: F10+P19 (dust fractions + PAH temperature)					
2	28	$0.2516^{+0.0086}_{-0.0085}$	$0.071^{+0.056}_{-0.045}$	$0.677^{+0.042}_{-0.054}$	474^{+30}_{-26}
3	43	$0.2528^{+0.0070}_{-0.0070}$	$0.053^{+0.044}_{-0.034}$	$0.694^{+0.033}_{-0.043}$	469^{+24}_{-21}
4	57	$0.2500^{+0.0060}_{-0.0060}$	$0.038^{+0.037}_{-0.026}$	$0.712^{+0.024}_{-0.035}$	461^{+20}_{-17}
5	72	$0.2499^{+0.0058}_{-0.0059}$	$0.035^{+0.034}_{-0.024}$	$0.715^{+0.022}_{-0.033}$	459^{+19}_{-17}
6	87	$0.2478^{+0.0054}_{-0.0054}$	$0.032^{+0.033}_{-0.022}$	$0.720^{+0.021}_{-0.031}$	452^{+18}_{-15}
7	100	$0.2479^{+0.0052}_{-0.0052}$	$0.040^{+0.035}_{-0.026}$	$0.712^{+0.024}_{-0.033}$	460^{+18}_{-16}
8	111	$0.2482^{+0.0050}_{-0.0051}$	$0.040^{+0.034}_{-0.026}$	$0.712^{+0.024}_{-0.032}$	461^{+17}_{-15}
9	124	$0.2483^{+0.0049}_{-0.0050}$	$0.040^{+0.034}_{-0.026}$	$0.711^{+0.024}_{-0.032}$	460^{+17}_{-14}
10	138	$0.2490^{+0.0049}_{-0.0049}$	$0.042^{+0.034}_{-0.027}$	$0.709^{+0.025}_{-0.032}$	462^{+17}_{-15}
11	151	$0.2471^{+0.0047}_{-0.0048}$	$0.044^{+0.035}_{-0.028}$	$0.708^{+0.026}_{-0.033}$	462^{+17}_{-15}
12	165	$0.2468^{+0.0047}_{-0.0047}$	$0.042^{+0.034}_{-0.027}$	$0.711^{+0.025}_{-0.032}$	460^{+16}_{-14}

Table 4.4: Same as Table 4.3, but for the case with PAH dust temperature as a free parameter.

Short Observation Time: D11 (dust fractions)				
N.S.	ndof	$f_{\rm PAH}$	$f_{\rm SG}$	$f_{\rm LG}$
2	20	$0.188^{+0.021}_{-0.021}$	$0.53^{+0.12}_{-0.11}$	$0.28^{+0.10}_{-0.11}$
3	34	$0.192^{+0.016}_{-0.016}$	$0.61^{+0.10}_{-0.10}$	$0.199^{+0.088}_{-0.093}$
4	48	$0.183^{+0.013}_{-0.013}$	$0.670^{+0.086}_{-0.088}$	$0.147^{+0.081}_{-0.079}$
5	62	$0.187^{+0.012}_{-0.011}$	$0.677^{+0.082}_{-0.085}$	$0.136^{+0.078}_{-0.076}$
6	75	$0.185^{+0.011}_{-0.011}$	$0.684^{+0.080}_{-0.084}$	$0.131^{+0.078}_{-0.074}$
7	85	$0.181^{+0.011}_{-0.010}$	$0.695^{+0.077}_{-0.081}$	$0.124^{+0.075}_{-0.071}$
8	95	$0.183^{+0.010}_{-0.010}$	$0.689^{+0.077}_{-0.080}$	$0.129^{+0.075}_{-0.072}$
9	107	$0.1836^{+0.0099}_{-0.0097}$	$0.686^{+0.077}_{-0.080}$	$0.130^{+0.074}_{-0.072}$
10	120	$0.1838^{+0.0097}_{-0.0094}$	$0.680^{+0.078}_{-0.079}$	$0.136^{+0.074}_{-0.073}$
11	133	$0.1824^{+0.0094}_{-0.0093}$	$0.680^{+0.078}_{-0.078}$	$0.138^{+0.073}_{-0.073}$
12	146	$0.1811^{+0.0093}_{-0.0091}$	$0.687^{+0.076}_{-0.078}$	$0.133^{+0.073}_{-0.071}$
Long Observation Time: D11 (dust fractions)				
2	27	$0.2464^{+0.0078}_{-0.0077}$	$0.485^{+0.051}_{-0.047}$	$0.268^{+0.043}_{-0.047}$
3	41	$0.2311^{+0.0066}_{-0.0066}$	$0.517^{+0.044}_{-0.041}$	$0.252^{+0.038}_{-0.040}$
4	55	$0.2212^{+0.0059}_{-0.0060}$	$0.546^{+0.040}_{-0.038}$	$0.233^{+0.035}_{-0.037}$
5	69	$0.2232^{+0.0058}_{-0.0058}$	$0.565^{+0.040}_{-0.038}$	$0.212^{+0.035}_{-0.037}$
6	83	$0.2175^{+0.0055}_{-0.0055}$	$0.578^{+0.039}_{-0.038}$	$0.204^{+0.035}_{-0.036}$
7	96	$0.2130^{+0.0052}_{-0.0052}$	$0.593^{+0.039}_{-0.037}$	$0.195^{+0.034}_{-0.036}$
8	109	$0.2131^{+0.0051}_{-0.0051}$	$0.591^{+0.039}_{-0.037}$	$0.195^{+0.034}_{-0.036}$
9	121	$0.2128^{+0.0050}_{-0.0050}$	$0.592^{+0.038}_{-0.037}$	$0.195^{+0.034}_{-0.036}$
10	134	$0.2117^{+0.0049}_{-0.0048}$	$0.597^{+0.038}_{-0.036}$	$0.191^{+0.034}_{-0.036}$
11	147	$0.2088^{+0.0047}_{-0.0047}$	$0.608^{+0.038}_{-0.036}$	$0.183^{+0.034}_{-0.036}$
12	160	$0.2090^{+0.0046}_{-0.0046}$	$0.610^{+0.038}_{-0.036}$	$0.181^{+0.034}_{-0.036}$

Table 4.5: EBL parameters from the combined inference of the highest TS simulated sources (described by the PL intrinsic model and attenuated by D11 EBL). Upper half: simulations using spectra with total observed time shorter than 5 h. Lower half: spectra with total observed time equal or greater than 5 h. N.S. refers to the number of sources. Values are the median from their marginal distribution, while uncertainties represent the 68% interval around it

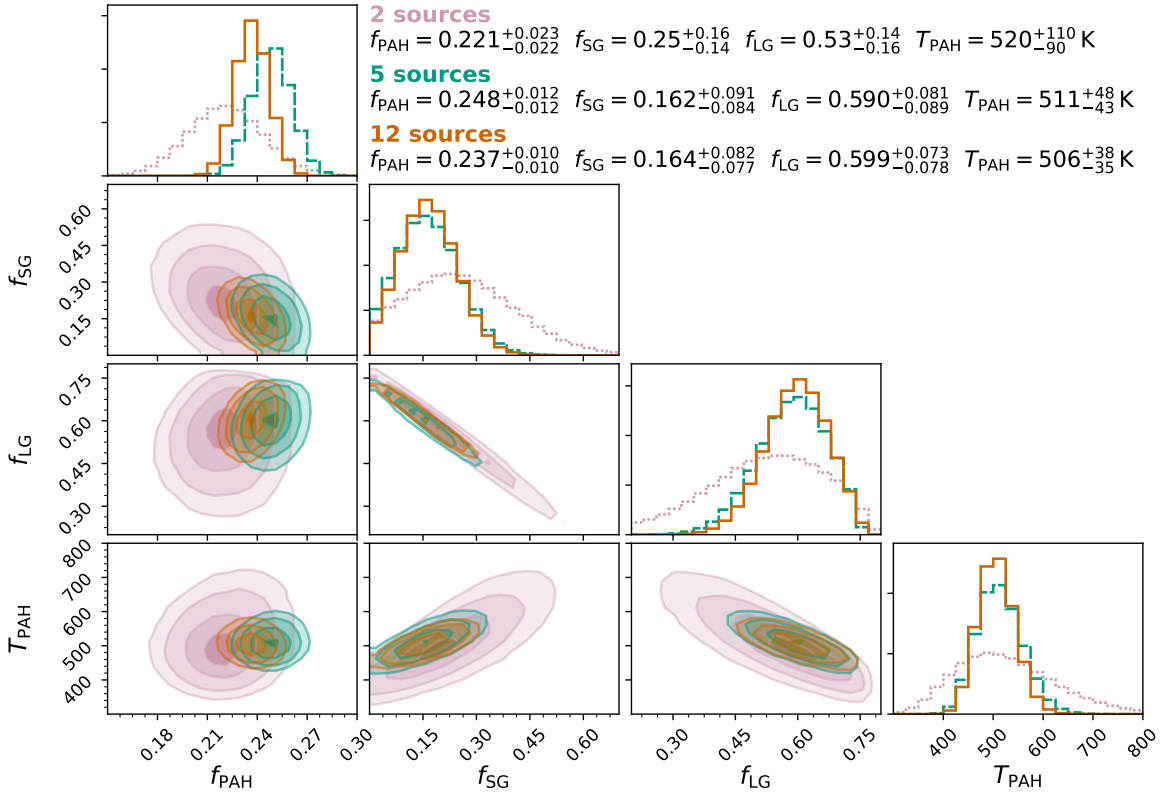

Figure 4.6: One and two dimensional marginal distributions of EBL parameters, including the temperature of the PAH component, T_{PAH}, using the set of short observation time attenuated by F10+P19's model.

EBL intensity inferred from the 12-source case and comparing to D11's actual EBL, in Figure 4.8. In this figure, we show the individual contributions of the dust components and consider the lower and upper uncertainties in the dust fractions to construct the uncertainty in the EBL intensity, shown as the shaded region. Clearly, D11 and F10 EBL models have very similar intensities in the UV and optical parts (at $z = 0$), but differ in the mid- and far-IR, where D11's model predicts lower and higher EBL levels, respectively. The median values of the dust fractions were able to approximately reproduce the EBL's spectrum in the range of ∼10–60 µm, which was possible by lowering the PAH fraction and increasing the warm SG component, as the rise of the third EBL peak occurs at a smaller wavelength in D11's model compared to F10. However, the model with free dust fractions (normalised to 1) cannot fully reproduce D11's EBL intensity in the far-IR (beyond 100 µm), since the value of its peak is at least a factor of 2 larger than F10+P19.

By increasing the observational time of the sources, the limitations of inferring D11 EBL attenuation through the F10 model with free dust fractions become even more apparent. Figure 4.7 shows that, although the inferred SG and LG fractions continue to agree under 1σ in both data sets, there is a more expressive shift in the PAH component, as its contribution is increased. This could be a result of the higher energy bins and more

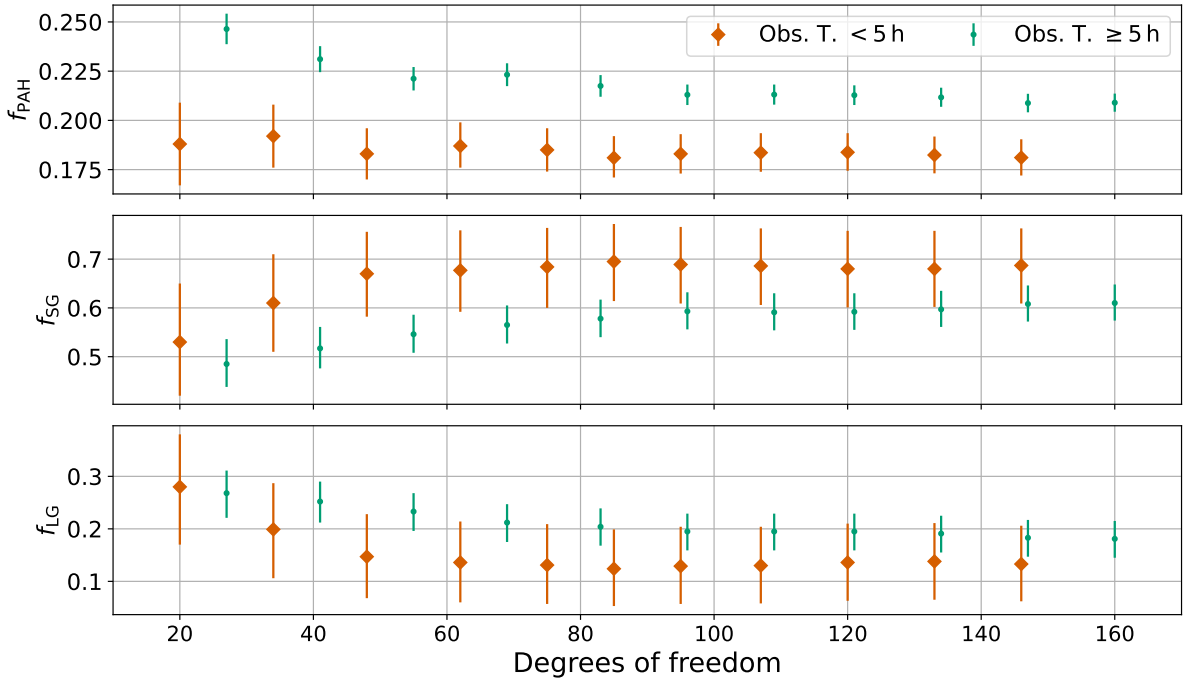

Figure 4.7: Median values of the dust fractions as a function of the number of degrees of freedom in each simulation, considering the attenuation from D11 model. The leftmost point correspond to the case with the 2 highest TS sources and the rightmost with all 12 spectra.

precise fluxes being able to probe part of the far-IR region, while D11 model requires a much higher EBL intensity that cannot be reached by simply changing the normalised fractions. This causes an increase of the PAH fraction to compensate the required attenuation level of the highest energy bins. In terms of the EBL spectrum, presented in Figure 4.9, we see that the mid-IR intensity seems to be overestimated, while above ∼30 µm the result is very similar to what is seen on Figure 4.8, besides the smaller uncertainties.

Compared to the situation with fixed dust temperatures, when the PAH temperature becomes a free parameter, only small differences in the median dust fractions are apparent, as Table 4.6 and Figure 4.10 show. In fact, the inferred PAH temperature values are very similar to the median values of the F10+P19 case, although with a slightly higher uncertainty. For instance, with 12 sources in the short observation time set and F10+P19 attenuation, we found $T_{\mathrm{PAH}} = 506^{+38}_{-35}\,\mathrm{K}$, while for D11 sources, the value turns out to be $T_{\mathrm{PAH}} = 500^{+41}_{-42}\,\mathrm{K}$. For the longer observation time data, they become respectively $T_{\mathrm{PAH}} = 460^{+16}_{-14}\,\mathrm{K}$ and $T_{\mathrm{PAH}} = 476^{+20}_{-19}\,\mathrm{K}$. As a consequence, the EBL spectra for the short and longer observation time are also similar to what was already discussed in the fixed temperatures case, so we do not see any clear improvements in mimicking D11's EBL intensity by having T_{PAH} as a free parameter. The corresponding spectra can be found in Figures C.2 and C.3 in Appendix C.

Short Observation Time: D11 (dust fractions + PAH temperature)					
N.S.	ndof	$f_{\rm PAH}$	$f_{\rm SG}$	$f_{\rm LG}$	$T_{\rm PAH}$ (K)
2	19	$0.190^{+0.023}_{-0.022}$	$0.60^{+0.13}_{-0.15}$	$0.22^{+0.15}_{-0.13}$	530^{+110}_{-100}
3	33	$0.195^{+0.017}_{-0.017}$	$0.70^{+0.07}_{-0.11}$	$0.10^{+0.11}_{-0.07}$	553^{+73}_{-73}
4	47	$0.183^{+0.014}_{-0.014}$	$0.72^{+0.07}_{-0.11}$	$0.10^{+0.10}_{-0.07}$	516^{+64}_{-63}
5	61	$0.185^{+0.012}_{-0.012}$	$0.72^{+0.07}_{-0.10}$	$0.11^{+0.11}_{-0.07}$	502^{+54}_{-55}
6	74	$0.183^{+0.011}_{-0.011}$	$0.71^{+0.08}_{-0.11}$	$0.11^{+0.11}_{-0.07}$	484^{+52}_{-52}
7	84	$0.179^{+0.011}_{-0.010}$	$0.73^{+0.06}_{-0.10}$	$0.089^{+0.096}_{-0.062}$	494^{+50}_{-51}
8	94	$0.180^{+0.010}_{-0.010}$	$0.74^{+0.06}_{-0.10}$	$0.085^{+0.092}_{-0.060}$	499^{+46}_{-46}
9	106	$0.1805^{+0.0098}_{-0.0095}$	$0.743^{+0.056}_{-0.090}$	$0.077^{+0.086}_{-0.054}$	507^{+43}_{-44}
10	119	$0.1805^{+0.0096}_{-0.0094}$	$0.737^{+0.060}_{-0.091}$	$0.083^{+0.087}_{-0.058}$	505^{+43}_{-43}
11	132	$0.1790^{+0.0093}_{-0.0091}$	$0.742^{+0.057}_{-0.089}$	$0.079^{+0.084}_{-0.055}$	509^{+42}_{-42}
12	145	$0.1775^{+0.0092}_{-0.0089}$	$0.738^{+0.061}_{-0.092}$	$0.085^{+0.087}_{-0.059}$	500^{+41}_{-42}
Long Observation Time: D11 (dust fractions + PAH temperature)					
2	26	$0.2469^{+0.0082}_{-0.0081}$	$0.635^{+0.064}_{-0.066}$	$0.118^{+0.065}_{-0.063}$	552^{+30}_{-28}
3	40	$0.2290^{+0.0067}_{-0.0066}$	$0.608^{+0.062}_{-0.061}$	$0.163^{+0.059}_{-0.061}$	515^{+27}_{-27}
4	54	$0.2186^{+0.0060}_{-0.0060}$	$0.621^{+0.059}_{-0.057}$	$0.161^{+0.055}_{-0.057}$	505^{+27}_{-26}
5	68	$0.2204^{+0.0059}_{-0.0058}$	$0.656^{+0.056}_{-0.056}$	$0.124^{+0.054}_{-0.054}$	514^{+24}_{-25}
6	82	$0.2142^{+0.0056}_{-0.0055}$	$0.648^{+0.056}_{-0.055}$	$0.138^{+0.053}_{-0.054}$	501^{+25}_{-24}
7	95	$0.2091^{+0.0052}_{-0.0053}$	$0.645^{+0.055}_{-0.054}$	$0.146^{+0.051}_{-0.053}$	489^{+23}_{-22}
8	108	$0.2091^{+0.0052}_{-0.0051}$	$0.647^{+0.055}_{-0.054}$	$0.144^{+0.052}_{-0.053}$	491^{+23}_{-22}
9	120	$0.2080^{+0.0051}_{-0.0051}$	$0.654^{+0.054}_{-0.053}$	$0.138^{+0.051}_{-0.052}$	493^{+22}_{-21}
10	133	$0.2070^{+0.0050}_{-0.0050}$	$0.651^{+0.054}_{-0.053}$	$0.142^{+0.050}_{-0.051}$	488^{+21}_{-20}
11	146	$0.2042^{+0.0049}_{-0.0048}$	$0.657^{+0.054}_{-0.053}$	$0.139^{+0.051}_{-0.051}$	485^{+21}_{-20}
12	159	$0.2048^{+0.0049}_{-0.0048}$	$0.643^{+0.054}_{-0.053}$	$0.152^{+0.050}_{-0.051}$	476^{+20}_{-19}

Table 4.6: Same as Table 4.5, but for the case with PAH dust temperature as a free parameter.

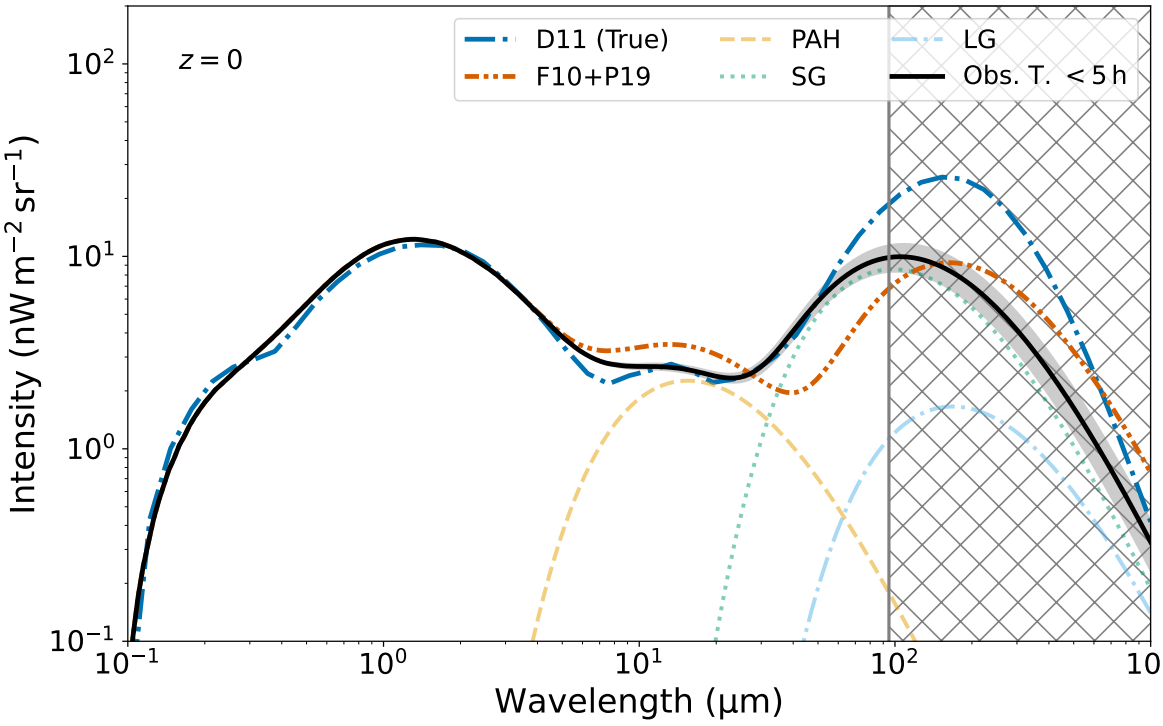

Figure 4.8: EBL's local spectral intensity as inferred with the data set of short observational time and sources attenuated by D11. The individual PAH, SG and LG contributions are shown, alongside the F10+P19 and D11 nominal curves. The hatched area is the region of the EBL that does not interact with the observed gamma ray flux data.

4.5 Inferring spectral parameters of sources

We compare the median values of the spectral indices to the true values, considering the simulations with 12 sources, listing the values in Table 4.7. Figures 4.11 and 4.12 present the cases without and with $T_{\rm PAH}$ as a free parameter, respectively, and the χ^2 between the median and true values are written in each box. Firstly, the results for F10+P19 sources reveal low χ^2 values, not larger than 3 times the number of degrees of freedom (12)[2], reaffirming that no biases are found and, in our case, the median is a good estimate of the regions where the posterior assumes its most significant values. Furthermore, this shows the robustness of Monte Carlo marginalisation, as these spectral index estimates incorporate the uncertainties in the EBL parameters. Comparing both figures, we notice that the inclusion of another EBL parameter (PAH temperature) also does not significantly change the median value of the marginal distribution of spectral indices, regardless if the sources were attenuated by F10+P19 or D11. Therefore, it reinforces that inferred $T_{\rm PAH}$ values are very consistent to the F10+P19 benchmark case.

The larger differences, however, arise when we compare the different EBL models. On one hand, for the low observation time data, there is a good agreement between

[2]Since no fit has been performed to computed this χ^2, the ndof corresponds to the number of data points.

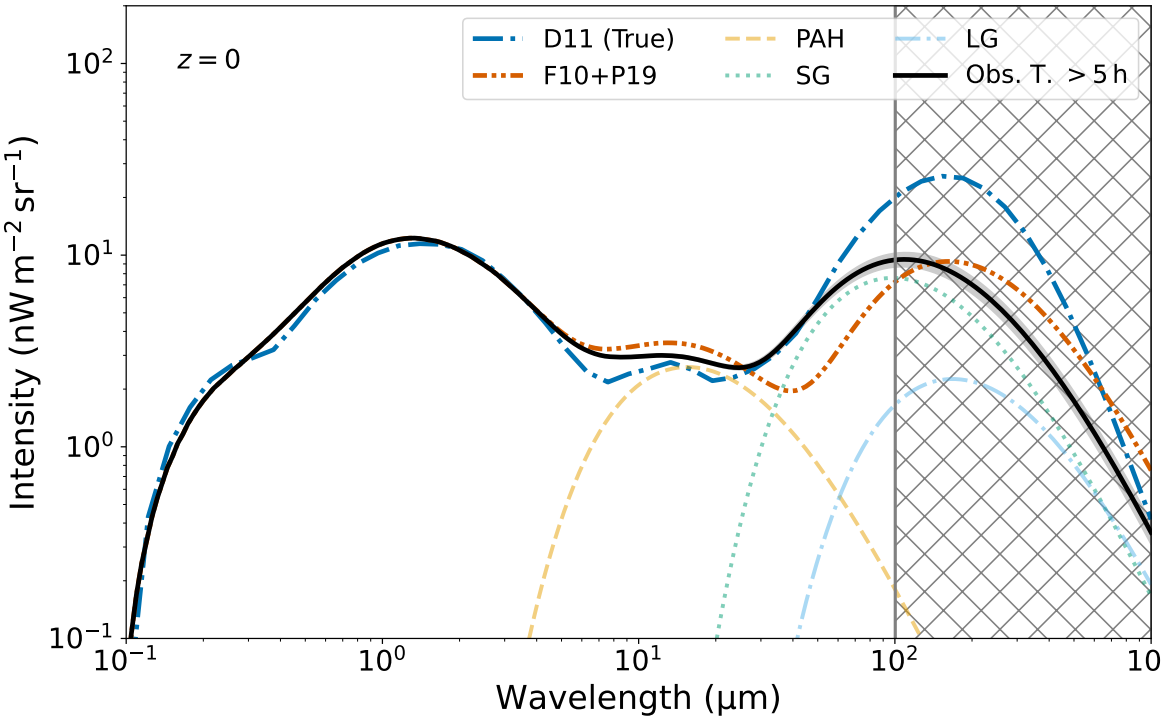

Figure 4.9: Same as Figure 4.8, but for the longer observational time data.

median and true values considering D11 attenuation. On the other hand, increasing the observation time reveals a noticeable bias in the spectral indices (more than 3–5σ), in particular by underestimating their values. As previously mentioned, the long observation time case resulted in an overestimation of the mid-IR compared to the expected D11 spectrum. Therefore, the increased overall extragalactic opacity, tends to underestimate the spectral index, as it gives more weight to the EBL in explaining the fall in the differential flux with energy. Interestingly, the bias is particularly larger for harder sources (low spectral index). Since they have a stronger emission at VHEs compared to softer spectra, their gamma rays can interact with a larger range of the EBL, so their inferred intrinsic spectra are subject to more systematic errors emerging from the wrong EBL modelling. This suggests that improved gamma-ray measurements that reach higher energies can possibly more finely distinguish different EBL opacities, or signal eventual limitations of the adopted model, reflected in the overall trend of the spectral indices.

Since we are working with the marginalised likelihood, the posterior sampling only provides information about the spectral indices of the sources. But we can recover the flux normalisation values by doing a one parameter fit to each gamma-ray spectrum, given the estimated EBL parameters and spectral indices. To gather more statistics, we can repeat this procedure using 100 samples from the posterior distribution constructed using 12 sources. This random sample help us to identify if the posterior distribution represents a reasonable global fit of the data. Having done these fits, we can explore the general behaviour of the residuals in all cases studied. Figure 4.13 shows the pull distribution

Short Obs. T.		F10+P19		D11	
ID	True	Frac.	Frac.+Temp.	Frac.	Frac.+Temp.
A	1.5783	$1.587^{+0.017}_{-0.017}$	$1.577^{+0.018}_{-0.018}$	$1.579^{+0.017}_{-0.017}$	$1.573^{+0.018}_{-0.018}$
B	1.8728	$1.882^{+0.015}_{-0.015}$	$1.876^{+0.015}_{-0.015}$	$1.867^{+0.014}_{-0.014}$	$1.863^{+0.015}_{-0.015}$
C	1.5870	$1.609^{+0.017}_{-0.017}$	$1.598^{+0.018}_{-0.019}$	$1.601^{+0.017}_{-0.017}$	$1.594^{+0.018}_{-0.018}$
D	1.7135	$1.739^{+0.018}_{-0.018}$	$1.728^{+0.019}_{-0.019}$	$1.682^{+0.017}_{-0.017}$	$1.675^{+0.018}_{-0.018}$
E	1.4850	$1.498^{+0.021}_{-0.021}$	$1.484^{+0.022}_{-0.023}$	$1.461^{+0.020}_{-0.020}$	$1.452^{+0.021}_{-0.021}$
F	1.4772	$1.476^{+0.030}_{-0.031}$	$1.461^{+0.032}_{-0.033}$	$1.468^{+0.031}_{-0.031}$	$1.458^{+0.032}_{-0.032}$
G	1.8992	$1.906^{+0.035}_{-0.034}$	$1.897^{+0.036}_{-0.035}$	$1.891^{+0.026}_{-0.025}$	$1.884^{+0.026}_{-0.026}$
H	1.5030	$1.454^{+0.038}_{-0.038}$	$1.428^{+0.042}_{-0.042}$	$1.511^{+0.036}_{-0.036}$	$1.496^{+0.039}_{-0.039}$
I	1.7051	$1.762^{+0.047}_{-0.046}$	$1.748^{+0.049}_{-0.048}$	$1.681^{+0.039}_{-0.038}$	$1.667^{+0.042}_{-0.041}$
J	1.6232	$1.607^{+0.041}_{-0.040}$	$1.591^{+0.042}_{-0.042}$	$1.654^{+0.039}_{-0.039}$	$1.644^{+0.040}_{-0.040}$
K	1.5137	$1.487^{+0.045}_{-0.045}$	$1.465^{+0.048}_{-0.048}$	$1.540^{+0.042}_{-0.042}$	$1.527^{+0.044}_{-0.044}$
L	1.7996	$1.865^{+0.045}_{-0.045}$	$1.851^{+0.047}_{-0.046}$	$1.792^{+0.043}_{-0.043}$	$1.783^{+0.044}_{-0.043}$
Long Obs. T.		F10+P19		D11	
A	1.5783	$1.5777^{+0.0083}_{-0.0084}$	$1.5744^{+0.0087}_{-0.0088}$	$1.6128^{+0.0074}_{-0.0075}$	$1.6084^{+0.0079}_{-0.0080}$
B	1.8728	$1.854^{+0.011}_{-0.011}$	$1.852^{+0.011}_{-0.011}$	$1.8976^{+0.0087}_{-0.0088}$	$1.8950^{+0.0089}_{-0.0089}$
C	1.5870	$1.5997^{+0.0094}_{-0.0095}$	$1.597^{+0.010}_{-0.010}$	$1.5491^{+0.0094}_{-0.0096}$	$1.545^{+0.010}_{-0.010}$
D	1.7135	$1.7184^{+0.0098}_{-0.0098}$	$1.715^{+0.010}_{-0.010}$	$1.6981^{+0.0095}_{-0.0094}$	$1.6937^{+0.0099}_{-0.0099}$
E	1.4850	$1.480^{+0.011}_{-0.011}$	$1.476^{+0.012}_{-0.012}$	$1.4858^{+0.0091}_{-0.0091}$	$1.480^{+0.010}_{-0.010}$
F	1.4772	$1.481^{+0.015}_{-0.015}$	$1.476^{+0.016}_{-0.016}$	$1.451^{+0.015}_{-0.015}$	$1.445^{+0.016}_{-0.016}$
G	1.8992	$1.899^{+0.020}_{-0.020}$	$1.896^{+0.020}_{-0.020}$	$1.877^{+0.019}_{-0.019}$	$1.872^{+0.020}_{-0.019}$
H	1.5030	$1.459^{+0.018}_{-0.018}$	$1.452^{+0.020}_{-0.020}$	$1.412^{+0.018}_{-0.018}$	$1.403^{+0.019}_{-0.019}$
I	1.7051	$1.679^{+0.022}_{-0.021}$	$1.673^{+0.022}_{-0.022}$	$1.675^{+0.021}_{-0.020}$	$1.666^{+0.022}_{-0.022}$
J	1.6232	$1.644^{+0.021}_{-0.021}$	$1.639^{+0.021}_{-0.021}$	$1.566^{+0.020}_{-0.020}$	$1.559^{+0.021}_{-0.021}$
K	1.5137	$1.522^{+0.024}_{-0.024}$	$1.516^{+0.025}_{-0.025}$	$1.400^{+0.023}_{-0.023}$	$1.392^{+0.024}_{-0.024}$
L	1.7996	$1.841^{+0.024}_{-0.023}$	$1.837^{+0.024}_{-0.024}$	$1.807^{+0.022}_{-0.022}$	$1.801^{+0.023}_{-0.022}$

Table 4.7: Spectral indices inferred from the 12 highest TS sources using the short and long observation time data sets, for both F10+P19 and D11 EBL attenuation, compared to their true values. "Frac.+Temp." refers to the case in which the PAH temperature was a free parameter.

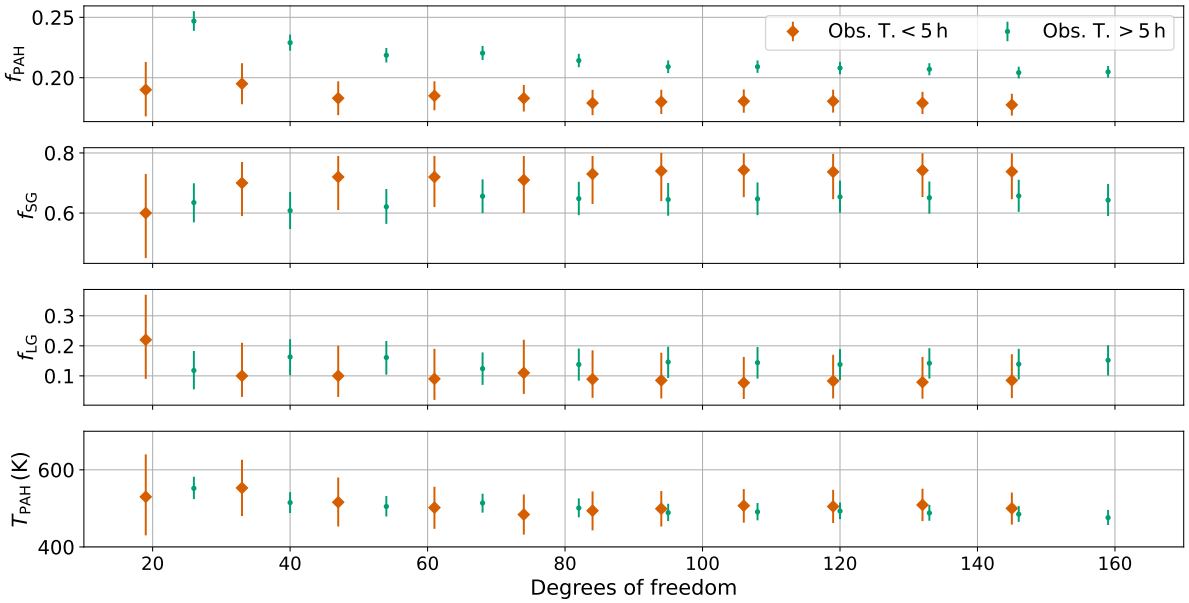

Figure 4.10: Same as Figure 4.7, but with the PAH dust temperature as a free parameter.

(residual over uncertainty) for the whole sample. Firstly we notice that the upper and bottom panels are very similar, as the inclusion of the PAH temperature did not change the fit significantly. Comparing the left and right panels, however, we can see that there is a noticeable difference when we increase the observation time if D11 is the true EBL. The residual distribution is much more widespread with long tails, revealing a worse global fit due to the systematic error in the EBL modelling.

4.6 Summary and discussion

The simultaneous inference on EBL and intrinsic spectral parameters is advantageous due to the fact that both emission and propagation characteristics are intertwined. Furthermore, the MCMC sampling of the posterior distribution allows a simple way of marginalising over a subset of variables, meaning that we can treat the EBL or the intrinsic spectra as nuisance parameters and incorporate their uncertainty into the another when projecting the posterior distribution into the dimensions of interest. In this chapter, we have also presented an analytical marginalisation of the posterior over the flux normalisation of all sources participating in the likelihood. This can be used for a wide variety of intrinsic (and EBL) models and it is best suited for optimisation purposes when dealing with a large number of sources, especially if the focus is on EBL constraints.

We have seen that the combination of a relatively low number of highly significant VHE sources at low redshift is capable of improving EBL constraints on the IR (through fits of the dust fractions or PAH temperature). However, the biggest impacts come from increasing the observational time of the same sources to gather even higher energy flux measurements and smaller flux uncertainties. With such data set, it was possible to probe

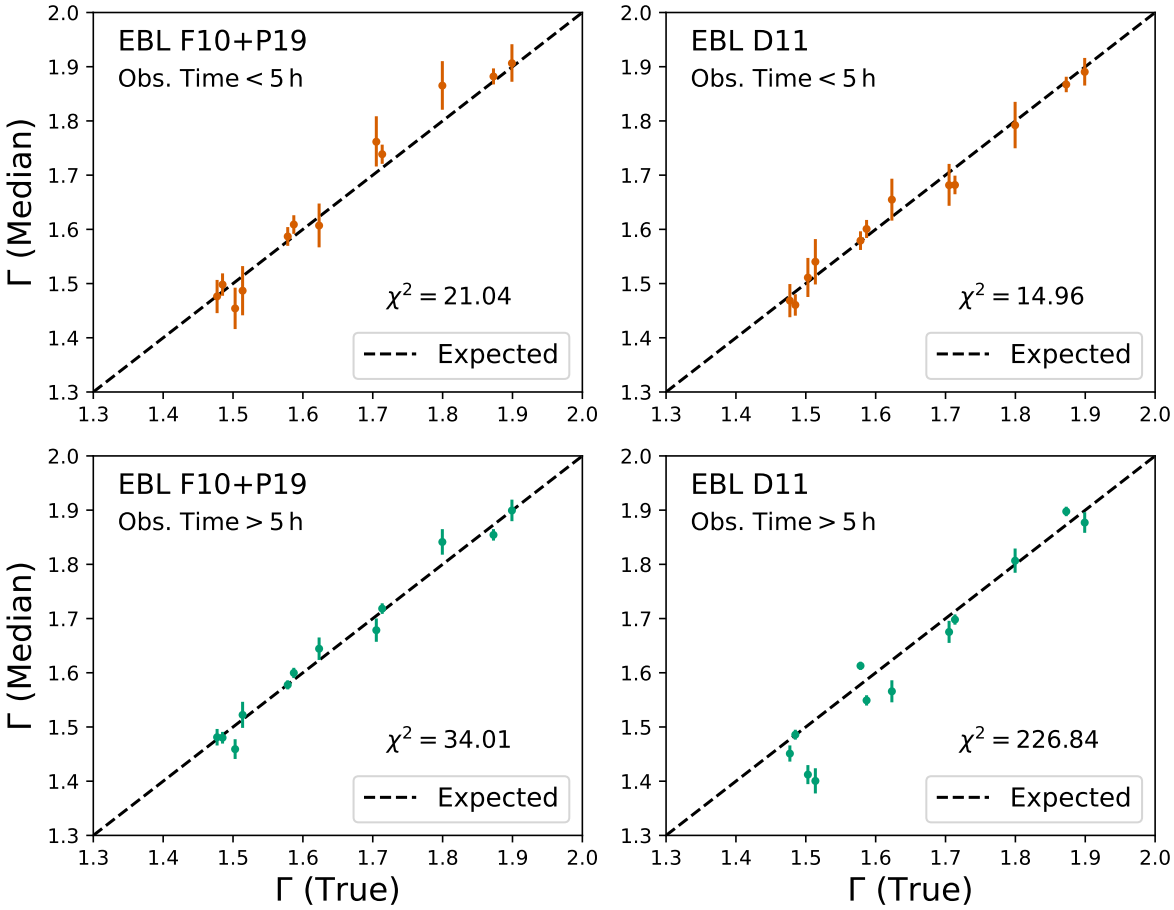

Figure 4.11: Comparison between the median spectral indices and their respective true values for the 12 highest TS sources, considering the different observation times and EBL opacities. The dashed diagonal line is the expected equality and the χ^2 value measures the sum of the quadratic difference in units of standard deviation.

systematic differences on EBL models, by identifying systematic errors on the inferred spectral indices, if the sources were attenuated by D11 model. This reflects the fact that the F10 model with free dust fractions cannot completely reproduce the EBL opacity from D11, due to differences mainly in the far-IR. In this situation, we have also found a worse global fit of all spectra. Nevertheless, the mid-IR range of the D11 model can be approximately well described by a lower PAH and a higher SG fraction value, revealing that the F10 model with free dust fractions admits some flexibility to describe the EBL opacity, if the available spectra is limited in energy and precision — as was the case with the short observation time data.

In this aspect, CTA has the potential of greatly contributing to improve measurements of current known AGN and even discovering new extragalactic VHE emission. With these data, it is likely that we will achieve an increased capability of distinguishing EBL models, while also improving constraints on current ones.

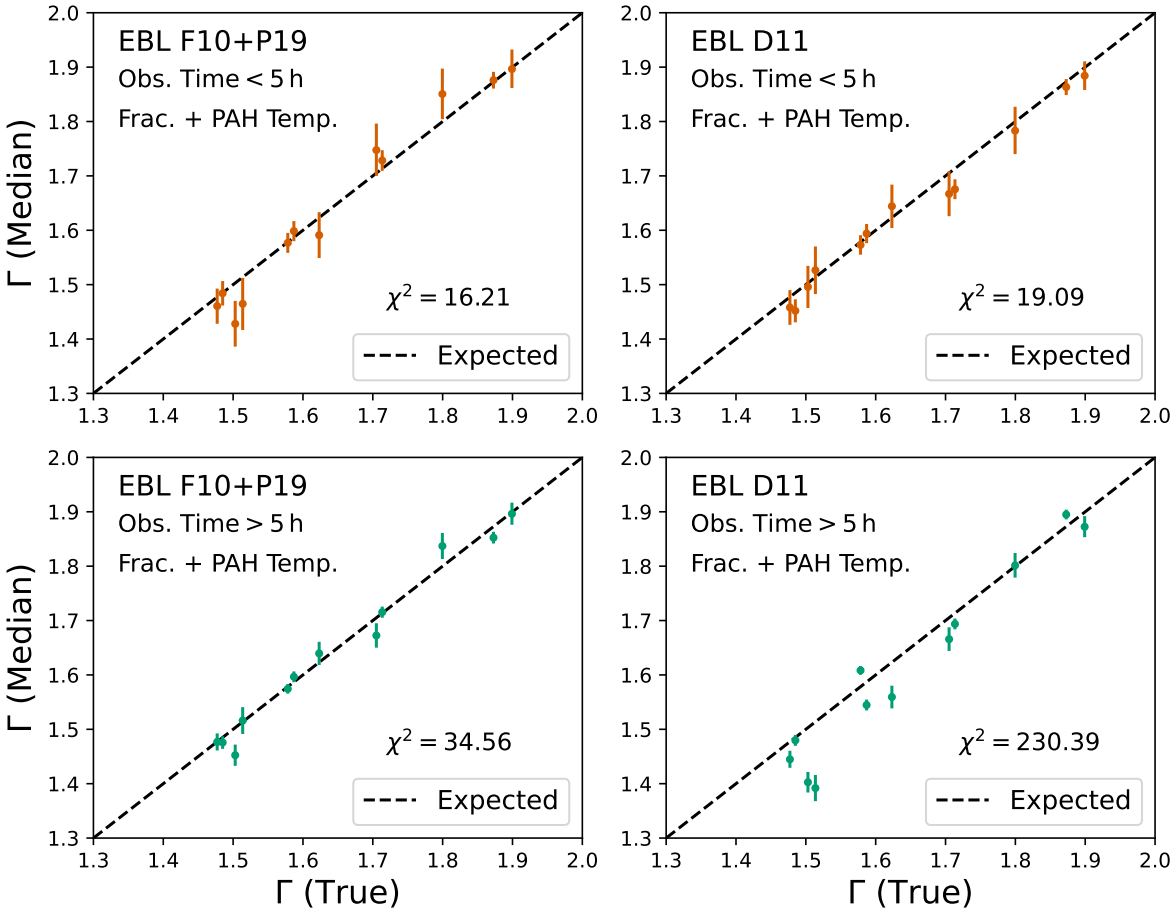

Figure 4.12: Same as Figure 4.11, but for the simulations where the PAH temperature is a free parameter.

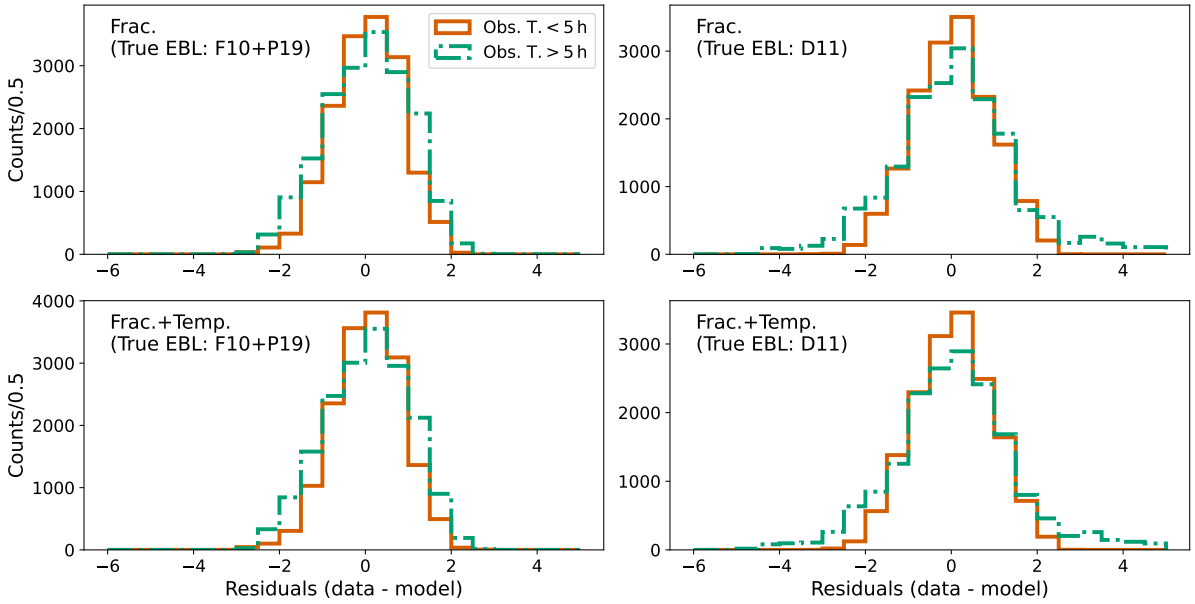

Figure 4.13: Pull distribution (residuals in units of standard deviation) from the global fit of 100 random models (combinations of dust fractions and spectral indices) sampled from the marginal posterior probability. The flux normalisation was obtained from a maximum likelihood fit for each source, given the other parameters.

Chapter 5

Probing Cosmology with gamma rays

In Chapter 3, section 3.1, we have seen how F10 model is constructed from the comoving luminosity density of stars and dust. Our goal is to understand the impact that changes on the cosmology — in particular the value of the Hubble constant $H_0 = 100h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ — has on the optical depth due to EBL's opacity, so section 5.1 describes the changes on the emissivity and EBL energy density that one needs to make to construct the optical depth for an arbitrary cosmology. Section 5.2 follows this discussion by presenting the constraints on H_0 when analysing the synthetic sources, as done in Chapter 4.

5.1 Cosmology and the EBL

From eq. 3.49, there are two ways that the assumed cosmology can impact the optical depth. The first is explicitly through the distance element $c|\,\mathrm{d}t/\mathrm{d}z\,|$. Qualitatively, $c|\,\mathrm{d}t/\mathrm{d}z\,|$ measures the correspondence between a redshift measurement and the comoving distance. For instance, in a universe with faster expansion rate (greater H_0), a given redshift would represent an object closer to the observer than in our current universe. This is because we can associate a recession velocity ($\mathrm{km\,s^{-1}}$) to a measured redshift, so a higher H_0 ($\mathrm{km\,s^{-1}\,Mpc^{-1}}$) with a fixed velocity means that the object must be closer. For smaller distances, given the mean free path, the probability of interaction would also be smaller, so the optical depth would be inversely proportional to H_0.

The second way the cosmological model impacts the optical depth is through the EBL density. The cosmological dependency in the EBL density comes from the integration of the total emissivity over the redshifts (eq. 3.19) and the computation itself of the stellar emissivity (eq. 3.11), which also reflects in the dust component (eq. 3.17). In both cases, the $|\,\mathrm{d}t/\mathrm{d}z\,|$ factor appears inside the integrals, but in the stellar emissivity there are two other channels. On one hand, the stellar emission $\dot{N}_*(\epsilon; m, t_*)$ depends on the star's age $t_* = t_*(z, z_1)$, expressed in terms of its birth redshift, z_1, and the redshift

of observation z, which is computed by integrating $|\mathrm{d}t_*/\mathrm{d}z|$ in this interval. Thus, a change in cosmology modifies the age of stars at each redshift and therefore can alter the evolutionary stage of them (characterised by a radius and a luminosity), impacting the overall stellar emission. On the other hand, the stellar emissivity also depends on the star formation rate $\psi(z)$ which, in turn, depends on the cosmology. In F10's model, the SFR from Hopkins & Beacom (2006) assumes a standard ΛCDM cosmology with $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $\Omega_m = 0.30$. Following Domínguez et al. (2019) and Ascasibar et al. (2002), we can approximate the SFR for arbitrary cosmological parameters (but still in the ΛCDM framework) by the relation

$$\psi(z) = \psi_{70}(z)\frac{H_0\sqrt{\Omega_m(1+z)^3+\Omega_\Lambda}}{H_0^{(70)}\sqrt{0.3(1+z)^3+0.7}}, \tag{5.1}$$

where now $H_0^{(70)} = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $\psi_{70}(z)$ is the SFR of Hopkins & Beacom (2006). Here, we are only interested in possible changes in H_0 and Ω_m. This approximation is valid because the computation of the comoving SFR density, in relation to different cosmologies, only involves transformations of the luminosity densities (Ascasibar et al. 2002). Therefore, the ratio of SFR for different cosmological parameters is essentially a ratio between luminosity densities. Consider for instance a comoving volume contained in a given solid angle, between redshifts $z + \Delta z$ and $z - \Delta z$, then

$$V(z, \Delta z) \propto D^3(z + \Delta z) - D^3(z - \Delta z), \tag{5.2}$$

where $D(z)$ is the comoving distance. For a given observed flux, the luminosity scales with D_L^2 where $D_L = (1 + z)D$ is the luminosity distance, so the luminosity density can be written as

$$\dot{\rho} = \frac{L(z)}{V(z, \Delta z)} \propto \frac{(1+z)^2 D^2(z)}{D^3(z + \Delta z) - D^3(z - \Delta z)}. \tag{5.3}$$

Approximating to first order in Δz,

$$D^3(z + \Delta z) - D^3(z - \Delta z) \approx 6\frac{D^2(z)}{H(z)}\Delta z, \tag{5.4}$$

where $H(z) = H_0\sqrt{\Omega_m(1+z)^3+\Omega_\Lambda}$, so we can write the ratio

$$\frac{\dot{\rho}}{\dot{\rho}_{70}} = \frac{\psi(z)}{\psi_{70}(z)} = \frac{H(z)}{H^{(70)}(z)}, \tag{5.5}$$

as expected.

Introducing only H_0 as a free parameter (keeping $\Omega_m = 0.3 = 1 - \Omega_\Lambda$) and implementing the SFR modification, we have built a new grid of optical depth (now in the E_γ, z, H_0 space) by recomputing the emissivities and EBL density. The parameter's range,

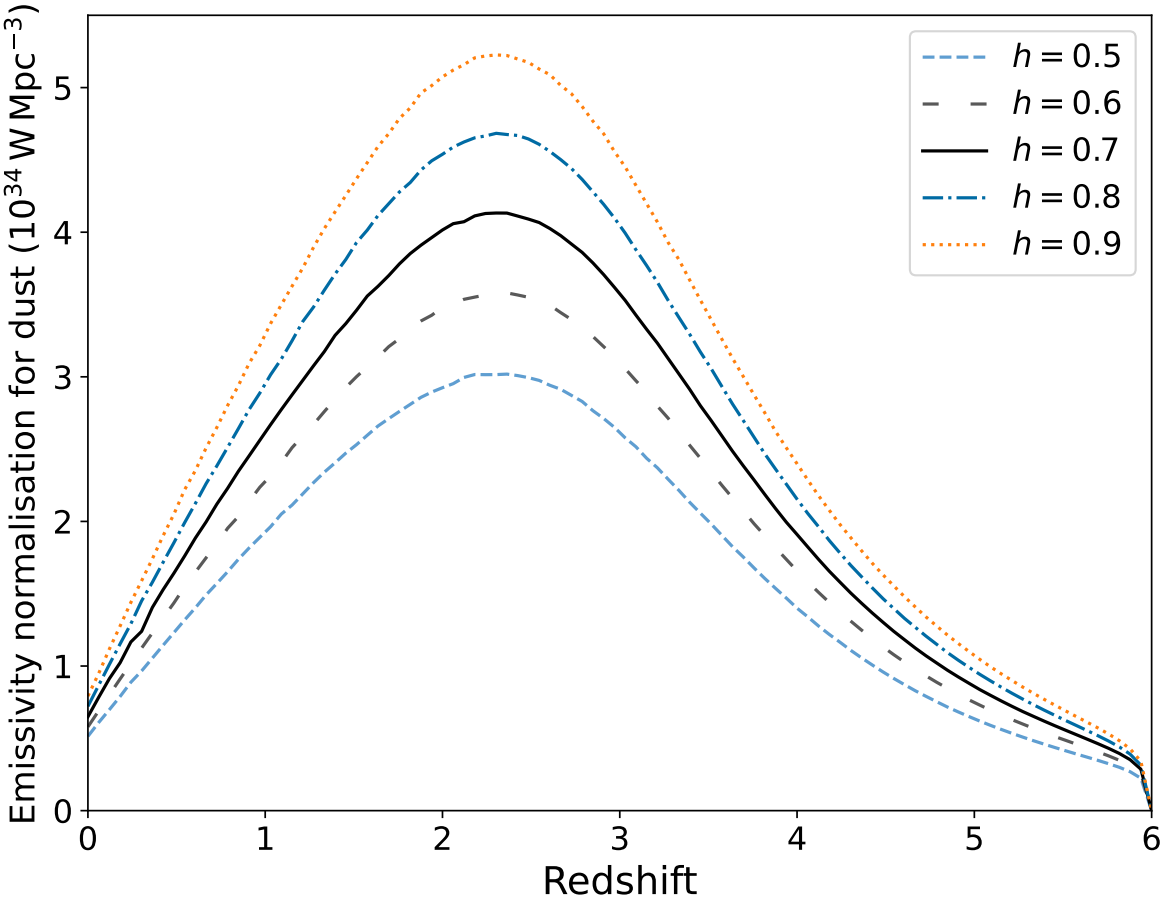

Figure 5.1: Normalisation of the emissivity coming from dust as a function of redshift for different values of h. The peak around $z \sim 2$ is explained by the maximum star formation rate around that epoch.

number of points and spacing are found in Table 3.1. For such computations, we have kept the dust temperatures fixed in their respective nominal values.

Starting with the emissivity, we can realise that the correction in the SFR, as $\psi(z) = \psi_{70}(z)(H_0/H_0^{(70)})$, actually cancels the new Hubble constant in $|\mathrm{d}t/\mathrm{d}z|$, which appears in the denominator. Therefore, all changes with respect to the nominal $H_0^{(70)}$ case comes from $\dot{N}_*(\epsilon; m, t_*)$, due to the difference in the star's age for a given redshift interval. Furthermore, since $t_*(z, z_1)$ is obtained by integrating $|\mathrm{d}t_*/\mathrm{d}z|$, the age is simply inversely proportional to the Hubble constant. So higher values for the Hubble constant result in younger stars, which probably mean a larger fraction of them remains in the main sequence. The lower amount of post-main sequence stars could then result in an overall higher luminosity density. That is because, although evolved stars can reach high luminosities, their photosphere is much colder, so their lower effective temperature imply less energy output from the blackbody emission.

Figure 5.1 presents the normalisation of the emissivity coming from dust for various H_0 values, where we can see the effect of increased luminosity density with larger cosmic

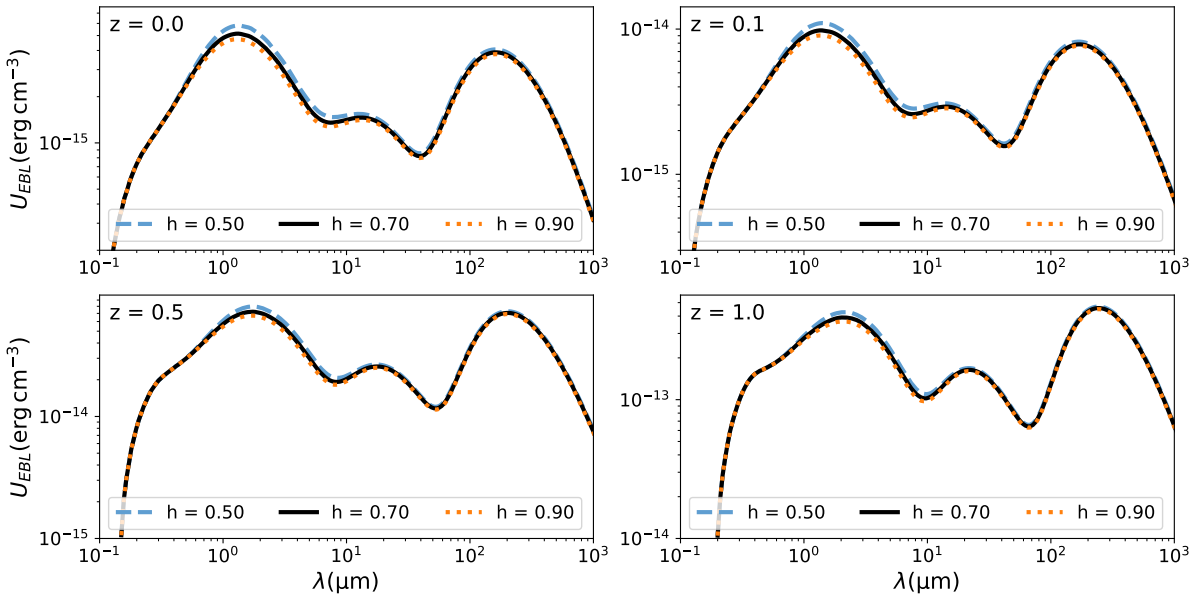

Figure 5.2: Comoving EBL density in erg cm^{-3} as a function of comoving wavelength (μm) for four redshift values ($z = \{0, 0.1, 0.5, 1.0\}$). Each box presents the EBL computed for three values of the dimensionless Hubble constant h.

expansion rates. The normalisation factor is simply the integral over the energies of the absorbed part of the stellar emissivity, as in eq. 3.17, which is redshift dependent. This figure also shows a consistent growth of emissivity up to $2 < z < 3$. Such phenomena is a result from the behaviour of the SFR. This is because, starting at very high redshifts ($z \sim 6$), we find low star formation rates, which progressively increase until a peak around $z \sim 2$ (see e.g. Behroozi et al. 2013 for a compilation of data). This represents a rise in stellar (and consequently dust) emissivity until such epoch. However, after that point, the star formation rates decreases. Stars are also evolving and the more massive ones are dying, which contribute to the decrease of photon flux and lower emissivity levels today. Even varying Hubble's constant, the normalisation peak appears to remain in the same position. This is of course in agreement with the correction implemented in the SFR, as it did not introduce any redshift dependent terms.

In the computation of the comoving EBL energy density (eq. 3.19), there are two competing factors, as increasing the Hubble constant implies higher emissivity, but at the same time, it decreases $|\mathrm{d}t/\mathrm{d}z|$. Figure 5.2 reveals that the second factor dominates, so higher H_0 values produce lower EBL comoving densities. Alongside the $|\mathrm{d}t/\mathrm{d}z|$ in the optical depth formula (eq. 3.2), inversely proportional to H_0, we should expect stronger gamma-ray attenuation for lower H_0 values. The differences in EBL density levels are more pronounced in the UV and optical parts, while few differences are seen in the dust-dominated region. This means we can also expect that most of the constraints in H_0 should come from low redshift sources emitting at the low-end of the VHEs, where they are most sensible to the stellar component of the EBL. To see this more clearly, we can look at

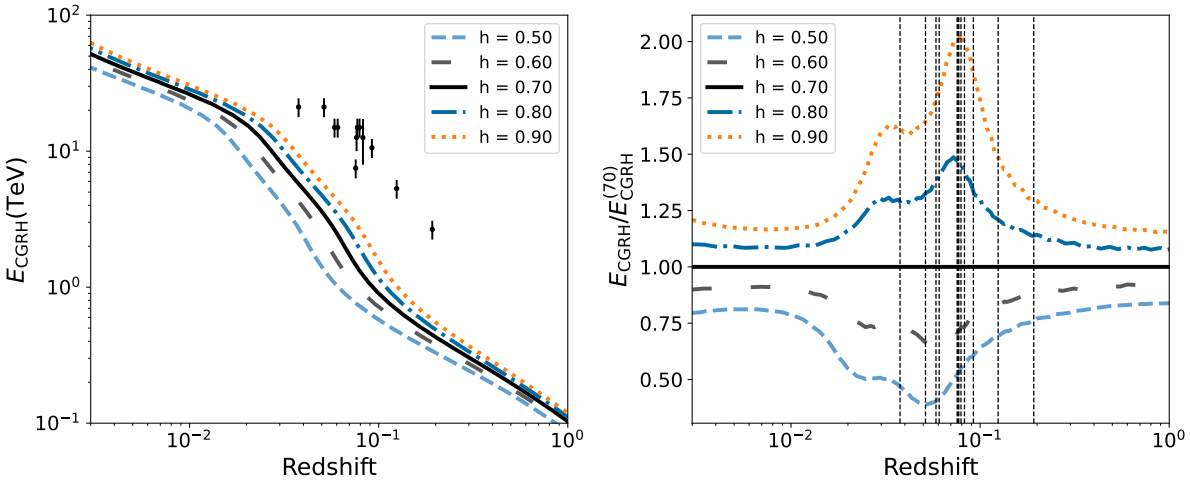

Figure 5.3: On the left, the cosmic gamma-ray horizon (CGRH) for different dimensionless Hubble constants h. On the right, the same lines are presented, but normalised by the CGRH line corresponding to the fiducial cosmology with $h = 0.70$. The points on the left are the last energy bin of the synthetic sources used for this analysis and the vertical dashed lines on the right correspond to their redshift values.

the CGRH constructed with the nominal dust fractions of the P19+F10 implementation, in Figure 5.3. In this figure, we see that the biggest differences in the CGRH appear for $z \lesssim 0.1$, which is more noticeable by taking the ratio of the CGRHs in relation to the $H_0^{(70)}$ case. Therefore, local VHE emitters at a few TeV could probe the CGRH and, in principle, be sensitive to changes in the Hubble constant. This observation is consistent with Domínguez & Prada (2013), in which the CGRH based on D11's model is plotted.

We investigate this proposition by working once again with the synthetic sources discussed in Chapter 4. Their redshift values fall exactly on the desired range to investigate H_0 constraints, represented as the vertical lines on the right side of Figure 5.3, so we analyse and discuss the results in section 5.2.

5.2 Inferring the Hubble Constant with synthetic spectra

We take the same sources as presented in Table 4.2 of Chapter 4, considering both data sets of short and long observation time, as defined in section 4.1. Since the spectra of these sources were generated considering the EBL model F10 with our benchmark values for the dust fractions (Table 4.1) and the fixed Hubble constant of $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, we can investigate this controlled scenario without introducing systematic errors in the EBL modelling. We use the same methodology as Chapter 4, adopting the marginal posterior distribution (i.e., after integrating out the flux normalisation parameters) and the sampler `emcee`. Table D.1 in Appendix D gives the detailed statistical information of each MCMC run.

Figure 5.4 shows the median values of the dust fractions and H_0 — normalised

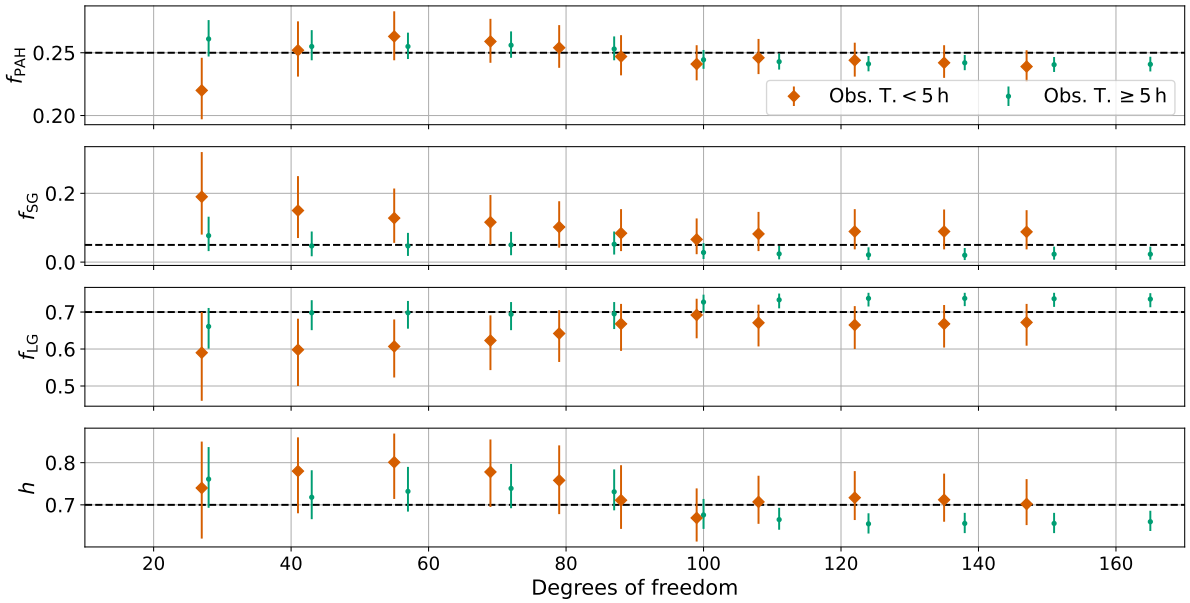

Figure 5.4: Median values of the dust fractions and $h = H_0/(100\,\mathrm{km\,s^{-1}\,Mpc^{-1}})$ as a function of the number of degrees of freedom in each MCMC simulation. The leftmost point correspond to the case with the 2 highest TS sources and the rightmost with all 12 spectra. The horizontal dashed lines are the true values of the F10+P19 attenuation.

as the variable $h = H_0/(100\,\mathrm{km\,s^{-1}\,Mpc^{-1}})$ —, compared to their respective true values (horizontal dashed lines), from the samplings using the highest TS sources (from 2 to 12 spectra). Table 5.1 presents the specific values. In the short observation time (total less than 5h), we see a very similar trend for the EBL parameters compared to the results of Chapter 4 with fixed H_0. There is a gradual convergence to the true values and progressively smaller uncertainties when we add other spectra, but the results are still dominated by the highest TS sources. A similar pattern happens with H_0, which is surprisingly very well constrained with more than 7 sources. In fact, beyond this number of sources, the inferred values of all parameters become very stable, as probably only higher TS sources with spectra extending to higher energies would improve the results.

By increasing the observation time, we find a visible bias in the parameters, as now the Hubble constant appears to be underestimate when combining all the sources, while the EBL parameters are systematically shifted in relation to their true values. One possibility is this simply being an effect of the low uncertainties, as the marginal distributions are still compatible to the true values. It is also possible that the specific selection of sources does not produce the most accurate inference on H_0. Also, we cannot discard numerical errors in our optical depth grid, as it was not possible to achieve the desired precision for its computation, given the time constraints and scope of this work. Some numerical imprecision was identified in the SG and LG components of the grid, so only sufficiently high VHE measurements with low uncertainties will be sensible to these errors — explaining the possible systematic shifts in the long observation time data

Short Observation Time: F10+P19 (dust fractions+ H_0)					
N.S.	ndof	$f_{\rm PAH}$	$f_{\rm SG}$	$f_{\rm LG}$	H_0 (*)
2	27	$0.220^{+0.026}_{-0.023}$	$0.19^{+0.13}_{-0.11}$	$0.59^{+0.11}_{-0.13}$	74^{+11}_{-12}
3	41	$0.252^{+0.023}_{-0.021}$	$0.15^{+0.10}_{-0.08}$	$0.598^{+0.084}_{-0.098}$	78^{+8}_{-10}
4	55	$0.263^{+0.020}_{-0.019}$	$0.128^{+0.086}_{-0.072}$	$0.607^{+0.073}_{-0.084}$	$80.1^{+6.8}_{-8.7}$
5	69	$0.259^{+0.018}_{-0.017}$	$0.116^{+0.079}_{-0.065}$	$0.623^{+0.068}_{-0.080}$	$77.8^{+7.7}_{-8.3}$
6	79	$0.254^{+0.018}_{-0.016}$	$0.102^{+0.075}_{-0.060}$	$0.642^{+0.063}_{-0.077}$	$75.8^{+8.3}_{-8.0}$
7	88	$0.247^{+0.017}_{-0.015}$	$0.084^{+0.070}_{-0.052}$	$0.668^{+0.054}_{-0.073}$	$71.1^{+8.3}_{-6.8}$
8	99	$0.241^{+0.015}_{-0.013}$	$0.066^{+0.061}_{-0.043}$	$0.692^{+0.044}_{-0.063}$	$66.9^{+7.0}_{-5.6}$
9	108	$0.246^{+0.015}_{-0.013}$	$0.082^{+0.064}_{-0.050}$	$0.671^{+0.049}_{-0.064}$	$70.7^{+6.2}_{-5.2}$
10	122	$0.244^{+0.014}_{-0.013}$	$0.089^{+0.065}_{-0.052}$	$0.665^{+0.051}_{-0.065}$	$71.7^{+6.3}_{-5.3}$
11	135	$0.242^{+0.014}_{-0.012}$	$0.089^{+0.064}_{-0.052}$	$0.668^{+0.051}_{-0.064}$	$71.2^{+6.2}_{-5.2}$
12	147	$0.239^{+0.013}_{-0.012}$	$0.088^{+0.063}_{-0.051}$	$0.672^{+0.050}_{-0.063}$	$70.2^{+5.9}_{-5.0}$
Long Observation Time: F10+P19 (dust fractions + H_0)					
2	28	$0.261^{+0.015}_{-0.014}$	$0.077^{+0.055}_{-0.045}$	$0.661^{+0.050}_{-0.060}$	$76.1^{+7.6}_{-6.8}$
3	43	$0.255^{+0.013}_{-0.011}$	$0.047^{+0.042}_{-0.030}$	$0.698^{+0.034}_{-0.047}$	$71.8^{+6.4}_{-5.2}$
4	57	$0.255^{+0.011}_{-0.010}$	$0.047^{+0.038}_{-0.029}$	$0.698^{+0.032}_{-0.043}$	$73.2^{+5.8}_{-4.8}$
5	72	$0.256^{+0.011}_{-0.010}$	$0.050^{+0.038}_{-0.030}$	$0.694^{+0.033}_{-0.043}$	$73.9^{+5.8}_{-4.7}$
6	87	$0.253^{+0.010}_{-0.009}$	$0.052^{+0.037}_{-0.030}$	$0.695^{+0.032}_{-0.041}$	$73.1^{+5.3}_{-4.4}$
7	100	$0.2444^{+0.0078}_{-0.0072}$	$0.028^{+0.027}_{-0.019}$	$0.727^{+0.020}_{-0.028}$	$67.6^{+3.8}_{-3.3}$
8	111	$0.2429^{+0.0065}_{-0.0063}$	$0.024^{+0.023}_{-0.016}$	$0.733^{+0.017}_{-0.023}$	$66.5^{+2.8}_{-2.4}$
9	124	$0.2411^{+0.0063}_{-0.0059}$	$0.021^{+0.022}_{-0.015}$	$0.737^{+0.015}_{-0.022}$	$65.5^{+2.5}_{-2.3}$
10	138	$0.2420^{+0.0062}_{-0.0059}$	$0.020^{+0.021}_{-0.014}$	$0.737^{+0.015}_{-0.021}$	$65.6^{+2.5}_{-2.3}$
11	151	$0.2405^{+0.0061}_{-0.0058}$	$0.023^{+0.022}_{-0.016}$	$0.736^{+0.016}_{-0.022}$	$65.6^{+2.5}_{-2.3}$
12	165	$0.2408^{+0.0060}_{-0.0057}$	$0.023^{+0.022}_{-0.016}$	$0.735^{+0.016}_{-0.022}$	$66.0^{+2.6}_{-2.2}$

Table 5.1: EBL parameters and H_0 from the combined inference of the highest TS simulated sources (described by the PL intrinsic model and attenuated by F10+P19 EBL). Upper half: simulations using spectra with total observed time lower than 5 h. Lower half: spectra with total observed time equal or greater than 5 h. N.S. refers to the number of sources. Values are the median from their marginal distribution, while uncertainties represent the 68% interval around it. (*) In $\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

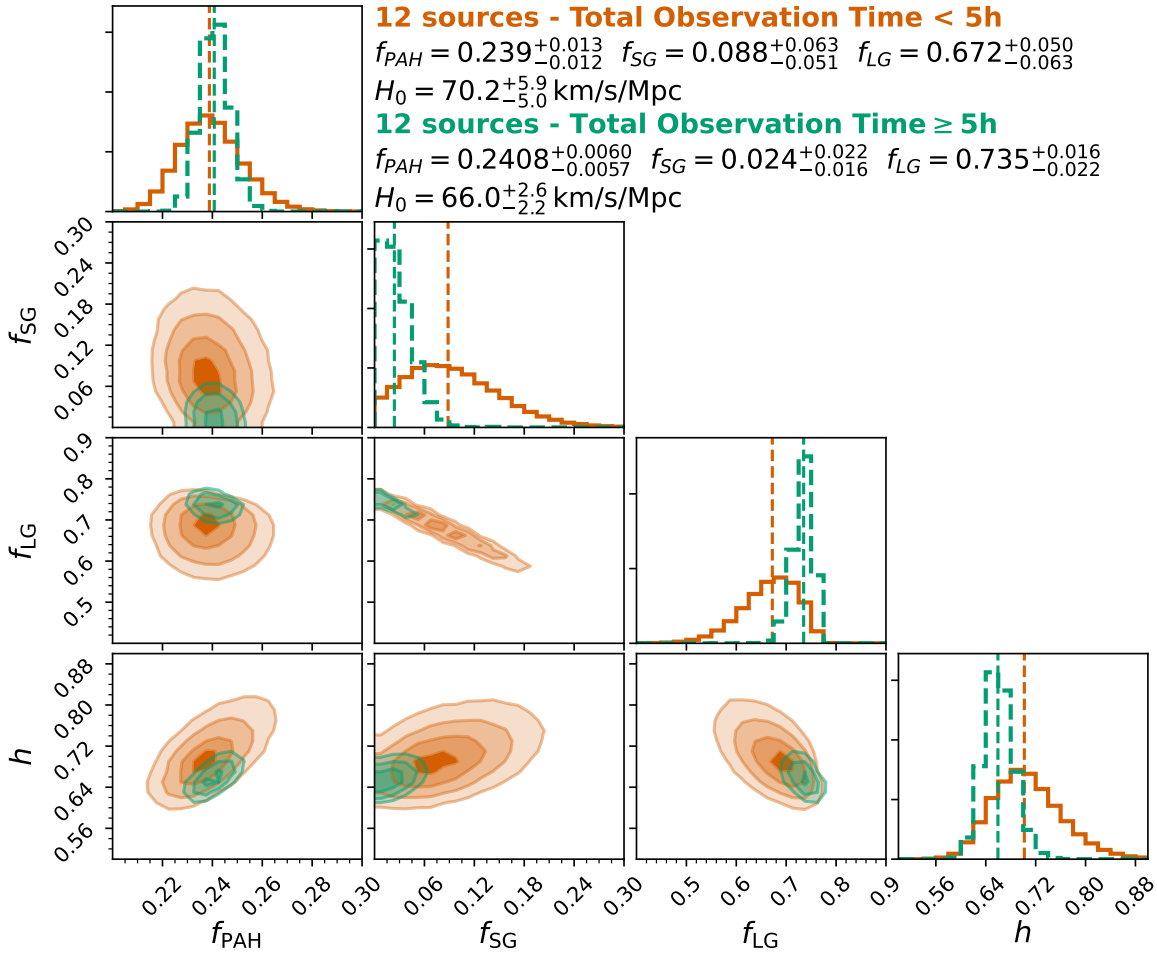

Figure 5.5: Marginal distributions of EBL parameters and $h = H_0/(100\,\mathrm{km\,s^{-1}\,Mpc^{-1}})$ using 12 highest TS sources attenuated by F10 EBL model with benchmark parameter values from P19. Both observation time data sets are represented. The distribution of the LG fraction is simply obtained by the constraint $f_{\mathrm{LG}} = 1 - f_{\mathrm{PAH}} - f_{\mathrm{SG}}$. The inferred values correspond to the median (represented by the vertical dashed line in the histograms) and the uncertainty is the 68% interval around it.

set analysis. Nevertheless, we will treat the results qualitatively, as it is still possible to identify a substantial improvement in uncertainty and a gradual trend of stabilisation of all parameters due to the increment of more spectra in the likelihood.

Figure 5.5 shows the marginal distribution of EBL parameters and h for the 12 sources case of both observation time data sets. We can see the notable improvement in uncertainty (of the order of 2), as the marginal distributions become more localised. Considering only the dust fractions, there are a few differences in the correlation between the variables, compared to the case in which the Hubble constant was fixed (see section 4.3). For the short observation time, we still identify a strong negative correlation between the SG and LG, of $\rho_{\text{SG-LG}} = -0.97$, but now we find a much weaker correlation between PAH and SG ($\rho_{\text{PAH-SG}} = -0.12$, compared to previously -0.52) and also a negative correlation between PAH and LG ($\rho_{\text{PAH-LG}} = -0.10$), instead of a positive one. By

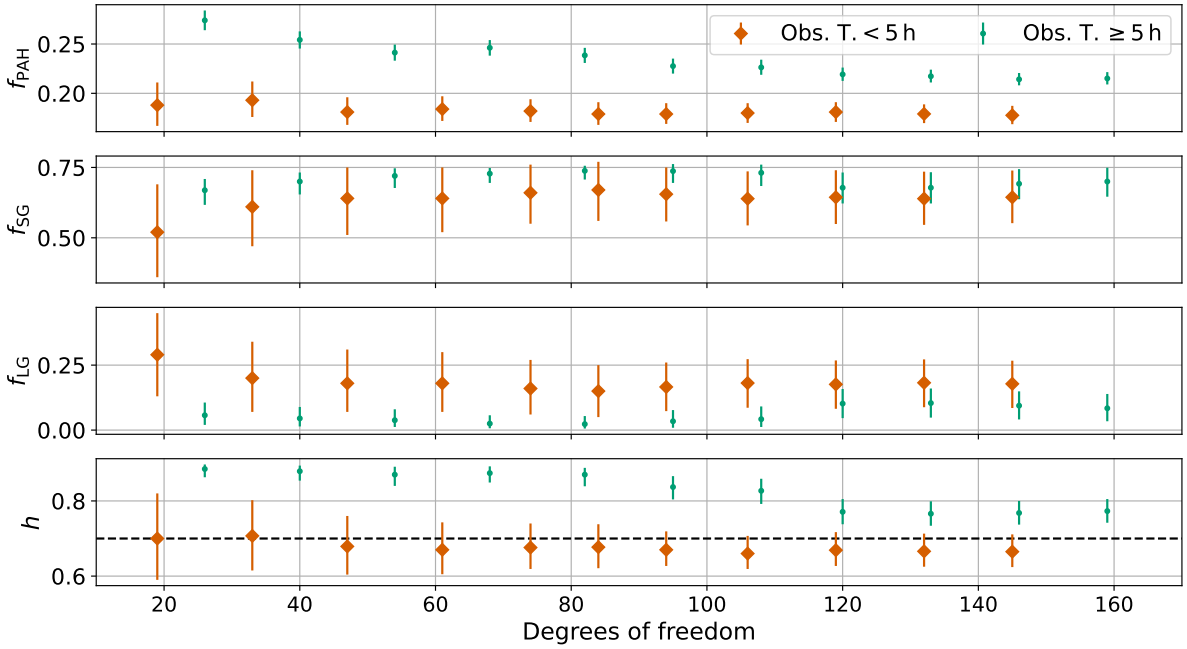

Figure 5.6: Median values of the dust fractions and $h = H_0/(100\,\mathrm{km\,s^{-1}\,Mpc^{-1}})$ as a function of the number of degrees of freedom in each MCMC simulation, considering sources attenuated by D11 EBL model. The leftmost point correspond to the case with the 2 highest TS sources and the rightmost with all 12 spectra. The horizontal dashed lines corresponds to D11's choice for the Hubble constant.

increasing the observation time, only small fluctuations in the correlations appear. In relation to the Hubble constant, the PAH component has the strongest correlation, of $\rho_{\text{h-PAH}} = 0.60$, indicating that, given our sample, a higher mid-IR EBL intensity would be associated to larger H_0 values. A positive correlation is also found for the SG fraction, as $\rho_{\text{h-SG}} = 0.40$, but a negative one is present for the LG ($\rho_{\text{h-LG}} = -0.54$).

Given these considerations, we proceed by investigating H_0 constraints by changing the true EBL attenuation to the D11 model. The Hubble constant in D11 work was also fixed at $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. The median values for the combined inference using the highest TS sources are presented in Figure 5.6, with the associated parameter values in Table 5.2. From the short observation time data set using 12 sources, we recover similar dust fraction values compared to the case presented in Chapter 4, in which H_0 was fixed (see Figure 4.7 and Table 4.5). Specifically, the inferred PAH fraction of $f_{\text{PAH}} = 0.1777^{+0.0096}_{-0.0090}$ matches the previous result of $f_{\text{PAH}} = 0.1811^{+0.0093}_{-0.0091}$ and the new SG and LG fractions of $f_{\text{SG}} = 0.644^{+0.095}_{-0.092}$ and $f_{\text{LG}} = 0.178^{+0.089}_{-0.093}$ are compatible to the Chapter 4 results of $f_{\text{SG}} = 0.687^{+0.076}_{-0.078}$ and $f_{\text{LG}} = 0.133^{+0.073}_{-0.071}$. Although the inferred Hubble constant value of $H_0 = 66.5^{+4.6}_{-4.1}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, given the uncertainties, is also compatible to the reference value from D11 model, there is a $\sim$4 $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ shift in the median that was not observed when using sources attenuated by F10+P19 EBL. Therefore, we attribute a systematic error of around 4 $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ in the Hubble constant due to the

incorrect EBL modelling.

By increasing the observation time, larger systematic shifts in the parameters appear, as the inference with few sources resulted in a higher amount of the PAH component (a fraction value close to 0.25) and also an exceedingly high value for the Hubble constant (above $80\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$). With 12 sources, however, the PAH fraction value $f_{\mathrm{PAH}} = 0.2151^{+0.0064}_{-0.0061}$ is closer to the result with fixed H_0, of $f_{\mathrm{PAH}} = 0.2090^{+0.0046}_{-0.0046}$, although the SG median is higher ($f_{\mathrm{SG}} = 0.700^{+0.049}_{-0.054}$) than before ($f_{\mathrm{SG}} = 0.610^{+0.038}_{-0.036}$) and the LG fraction is comparatively lower ($f_{\mathrm{LG}} = 0.084^{+0.055}_{-0.050}$, while previously it was $f_{\mathrm{LG}} = 0.181^{+0.034}_{-0.036}$). The resulting impact on the Hubble constant is an overestimation of its value, to $H_0 = 77.3^{+3.2}_{-3.1}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. This could be related to the previously mentioned positive correlation between the PAH component and H_0. As the long observation time data provides a deeper probe to longer wavelengths of the EBL, systematic effects from describing D11 opacity using F10 model start to appear, which result in an increase of the PAH component (as discussed in section 4.4 of Chapter 4) and its propagation to a higher H_0 value. Taking into consideration that numerical errors in the grid may have caused an underestimation of the Hubble constant, as shown in Figure 5.5, it is not clear the exact magnitude of these effects, but this analysis has shown that any H_0 constraints from gamma-ray observations are definitively sensitive to the EBL modelling and the availability of VHE data. Figure 5.7 summarises the EBL spectral shape for the short and long observation time data, compared to the true D11 spectrum. The shaded area for each curve represents the uncertainty in the dust fractions, as the uncertainty in H_0 is actually very negligible to the overall intensity.

Finally, we can analyse the impact on the spectral indices from the four scenarios studied. Figure 5.8 shows the median and respective uncertainties compared to the true photon indices of the 12 sources, alongside the corresponding χ^2 values. Notably, there is no significant bias in the short observation time case with F10+P19 EBL attenuation. Compared to the fixed H_0 case presented in Chapter 4, the uncertainties are a bit larger, which may be a result of the marginalisation over the Hubble constant and EBL parameters, since, even though they all agree with the benchmark values, the free H_0 variable broadens the posterior distribution, augmenting the uncertainty of the other parameters. The increased observation time results in a larger χ^2, which is partially due to the smaller uncertainties, but also from a systematic underestimation of the spectral indices, shifting the distribution below the dashed line — probably an effect of the intrinsic numerical resolution of the optical depth grid used. A similar shift is observed for the low observation time case with D11 attenuation, but we attribute this effect to the systematic error introduced by the EBL modelling. Curiously, in both situations, the inferred Hubble constant is approximately $\sim 66\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, so the underestimation of the photon indices may be related to the underestimation in H_0. The opposite effect is observed in the long observation time case for D11 attenuation, in which the spectral indices are overestimated

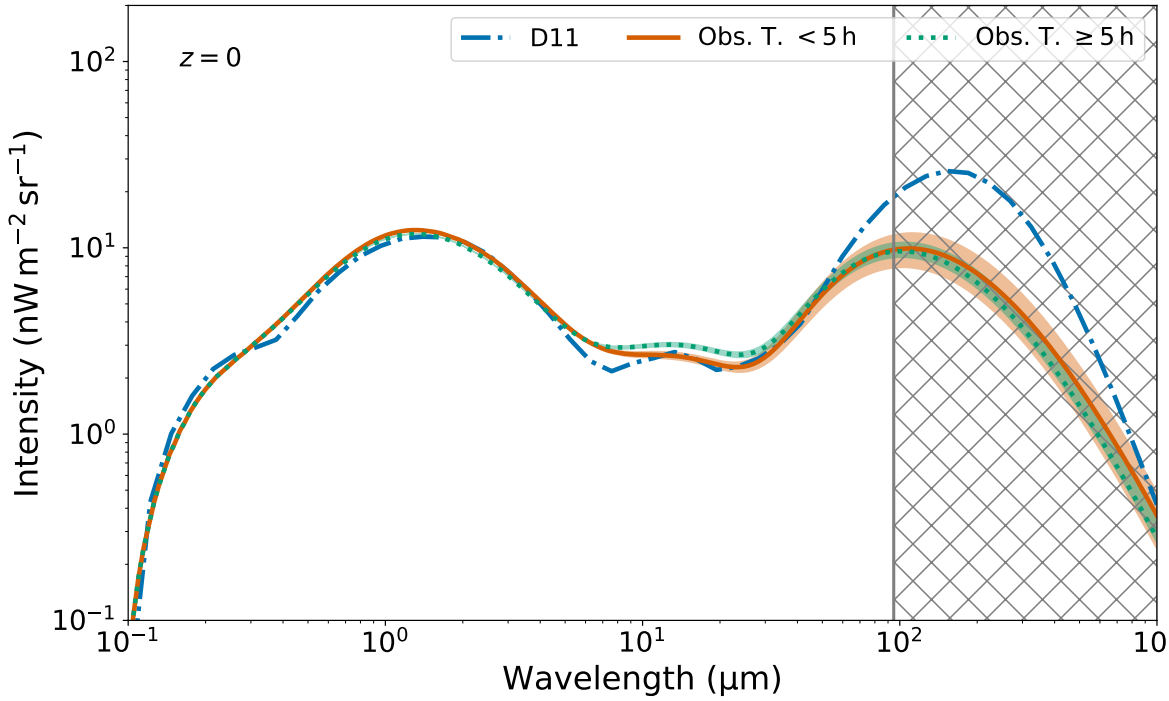

Figure 5.7: Local EBL spectrum ($z = 0$) inferred using the 12 highest TS sources, subject to the D11 opacity, considering the short ($< 5\,\mathrm{h}$) and long ($> 5\,\mathrm{h}$) observation times. The shaded area around each curve represents the uncertainty in the dust fractions. The hatched area is the region of the EBL that does not interact with the observed gamma ray flux data.

in general, as is the Hubble constant. Since, as seen in Figure 5.2, lower H_0 are associated with a higher EBL density in the UV to near-IR peak, this may imply an overestimation of the EBL opacity and consequently an underestimation of the spectral indices — or vice versa.

5.3 Discussion

In general, there is a good perspective of using VHE measurements of low redshift sources to constrain the Hubble constant. Given the limitations of the adopted models, the MCMC marginalisation is a simple way of propagating the statistical uncertainties of the EBL modelling and the parameters of individual sources to H_0. Ideally, different EBL models should also be used to compare the results and improve the estimation of the systematic uncertainties associated to the EBL modelling. As we have seen, systematic effects may dominate the analysis, producing significant shifts in H_0 if the wrong EBL model is used. As our knowledge on the EBL advances, alongside new data from CTA, there will be new perspectives of probing cosmological parameters independently from traditional methods.

Another possible extension is to investigate constraints on the matter density pa-

Short Observation Time: D11 (dust fractions+ H_0)					
N.S.	ndof	$f_{\rm PAH}$	$f_{\rm SG}$	$f_{\rm LG}$	H_0 (*)
2	19	$0.188^{+0.023}_{-0.021}$	$0.52^{+0.17}_{-0.16}$	$0.29^{+0.16}_{-0.16}$	70^{+12}_{-11}
3	33	$0.193^{+0.019}_{-0.017}$	$0.61^{+0.13}_{-0.14}$	$0.20^{+0.14}_{-0.13}$	$70.7^{+9.5}_{-9.2}$
4	47	$0.181^{+0.015}_{-0.013}$	$0.64^{+0.11}_{-0.13}$	$0.18^{+0.13}_{-0.11}$	$67.9^{+8.1}_{-7.5}$
5	61	$0.184^{+0.013}_{-0.012}$	$0.64^{+0.11}_{-0.12}$	$0.18^{+0.12}_{-0.11}$	$67.0^{+7.3}_{-6.5}$
6	74	$0.182^{+0.012}_{-0.011}$	$0.66^{+0.10}_{-0.11}$	$0.16^{+0.11}_{-0.10}$	$67.6^{+6.4}_{-5.7}$
7	84	$0.179^{+0.012}_{-0.011}$	$0.67^{+0.10}_{-0.11}$	$0.15^{+0.10}_{-0.10}$	$67.7^{+6.1}_{-5.6}$
8	94	$0.179^{+0.011}_{-0.010}$	$0.655^{+0.095}_{-0.097}$	$0.166^{+0.094}_{-0.093}$	$67.0^{+4.9}_{-4.3}$
9	106	$0.180^{+0.010}_{-0.010}$	$0.639^{+0.097}_{-0.095}$	$0.181^{+0.092}_{-0.095}$	$66.0^{+4.7}_{-4.1}$
10	119	$0.181^{+0.010}_{-0.010}$	$0.644^{+0.096}_{-0.095}$	$0.176^{+0.092}_{-0.094}$	$66.9^{+4.8}_{-4.2}$
11	132	$0.1791^{+0.0097}_{-0.0093}$	$0.639^{+0.096}_{-0.093}$	$0.182^{+0.090}_{-0.094}$	$66.6^{+4.7}_{-4.1}$
12	145	$0.1777^{+0.0096}_{-0.0090}$	$0.644^{+0.095}_{-0.092}$	$0.178^{+0.089}_{-0.093}$	$66.5^{+4.6}_{-4.1}$
Long Observation Time: D11 (dust fractions + H_0)					
2	26	$0.274^{+0.010}_{-0.010}$	$0.669^{+0.040}_{-0.052}$	$0.057^{+0.049}_{-0.037}$	$88.5^{+1.2}_{-2.2}$
3	40	$0.2542^{+0.0086}_{-0.0088}$	$0.700^{+0.032}_{-0.046}$	$0.045^{+0.044}_{-0.031}$	$87.9^{+1.5}_{-2.5}$
4	54	$0.2413^{+0.0081}_{-0.0082}$	$0.720^{+0.027}_{-0.043}$	$0.038^{+0.042}_{-0.026}$	$87.0^{+2.1}_{-3.0}$
5	68	$0.2462^{+0.0078}_{-0.0080}$	$0.728^{+0.020}_{-0.033}$	$0.025^{+0.032}_{-0.018}$	$87.4^{+1.8}_{-2.5}$
6	82	$0.2385^{+0.0076}_{-0.0077}$	$0.738^{+0.018}_{-0.031}$	$0.023^{+0.031}_{-0.017}$	$87.0^{+1.8}_{-3.1}$
7	95	$0.2276^{+0.0076}_{-0.0076}$	$0.737^{+0.025}_{-0.042}$	$0.034^{+0.043}_{-0.025}$	$83.7^{+2.9}_{-3.3}$
8	108	$0.2263^{+0.0078}_{-0.0075}$	$0.731^{+0.029}_{-0.047}$	$0.042^{+0.049}_{-0.030}$	$82.7^{+3.2}_{-3.5}$
9	120	$0.2192^{+0.0069}_{-0.0066}$	$0.678^{+0.054}_{-0.056}$	$0.102^{+0.056}_{-0.056}$	$77.1^{+3.4}_{-3.3}$
10	133	$0.2173^{+0.0067}_{-0.0063}$	$0.678^{+0.055}_{-0.056}$	$0.104^{+0.056}_{-0.056}$	$76.6^{+3.3}_{-3.2}$
11	146	$0.2142^{+0.0064}_{-0.0061}$	$0.692^{+0.052}_{-0.055}$	$0.094^{+0.055}_{-0.053}$	$76.8^{+3.2}_{-3.1}$
12	159	$0.2151^{+0.0064}_{-0.0061}$	$0.700^{+0.049}_{-0.054}$	$0.084^{+0.055}_{-0.050}$	$77.3^{+3.2}_{-3.1}$

Table 5.2: EBL parameters and H_0 from the combined inference of the highest TS simulated sources (described by the PL intrinsic model and attenuated by D11 EBL). Upper half: simulations using spectra with total observed time lower than 5 h. Lower half: spectra with total observed time equal or greater than 5 h. N.S. refers to the number of sources. Values are the median from their marginal distribution, while uncertainties represent the 68% interval around it. (*) In $\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

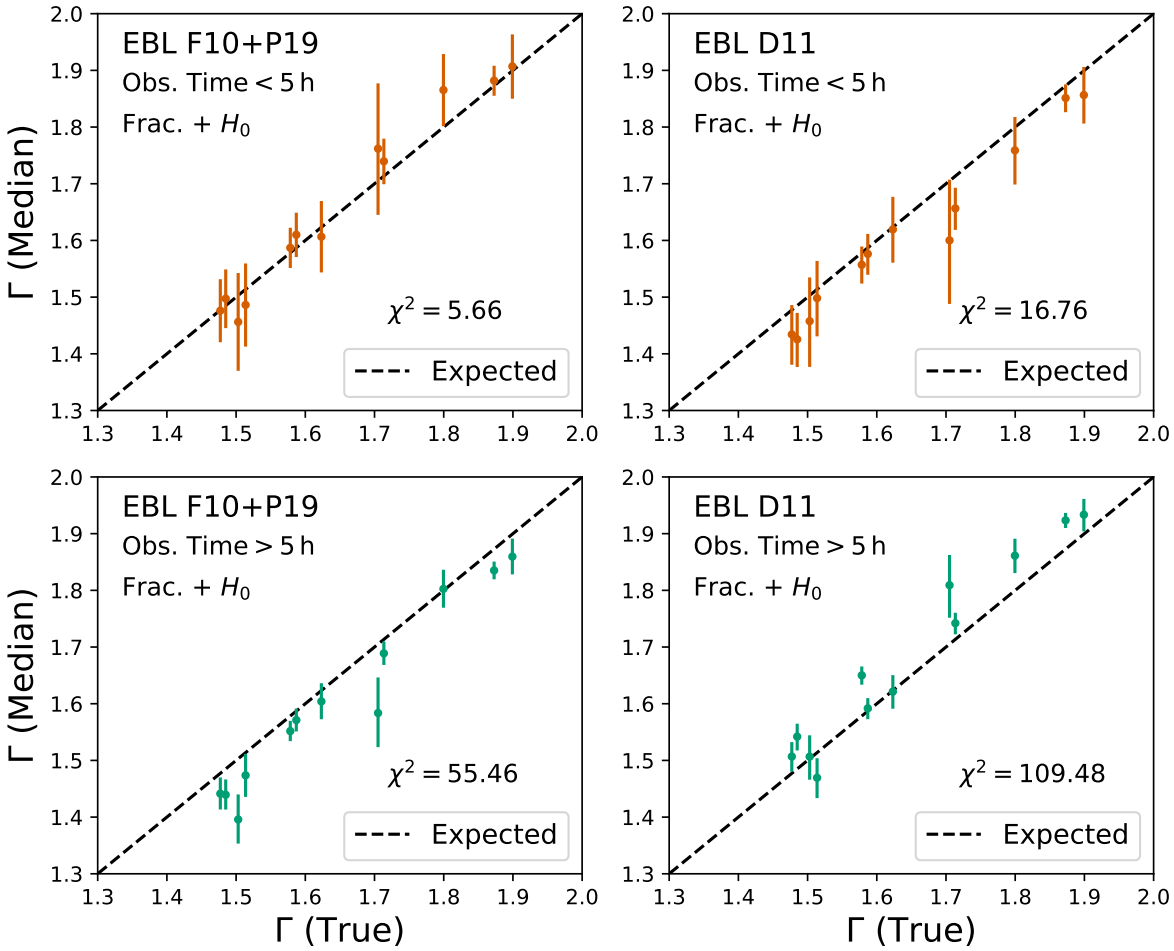

Figure 5.8: Inferred spectral indices of the 12 highest TS sources, compared to their true values, considering the different observation times and EBL models. Each point corresponds to the median and the 68% confidence interval around it. The dashed diagonal line is the expected equality between axis and the χ^2 value measures the sum of the quadratic difference in units of standard deviation.

rameter Ω_m. In our methodology, this would imply the addition of an extra dimension to the optical depth grid (computationally expensive, but a doable task). Analogously to changes in H_0, modifying Ω_m would affect all $|\mathrm{d}t/\mathrm{d}z|$ factors and the SFR, but these changes cannot be factored out, introducing redshift dependent corrections. Nevertheless, we do not expect strong constraints from the same data set of low redshift sources. This is because changes in the matter density parameter mostly impact higher redshifts ($z \gtrsim 1$), as we can see from the behaviour of the $|\mathrm{d}t/\mathrm{d}z|$ curves for different Ω_m, on the right side of Figure 5.9. Conversely, changes in H_0 affect $|\mathrm{d}t/\mathrm{d}z|$ primarily at low z.

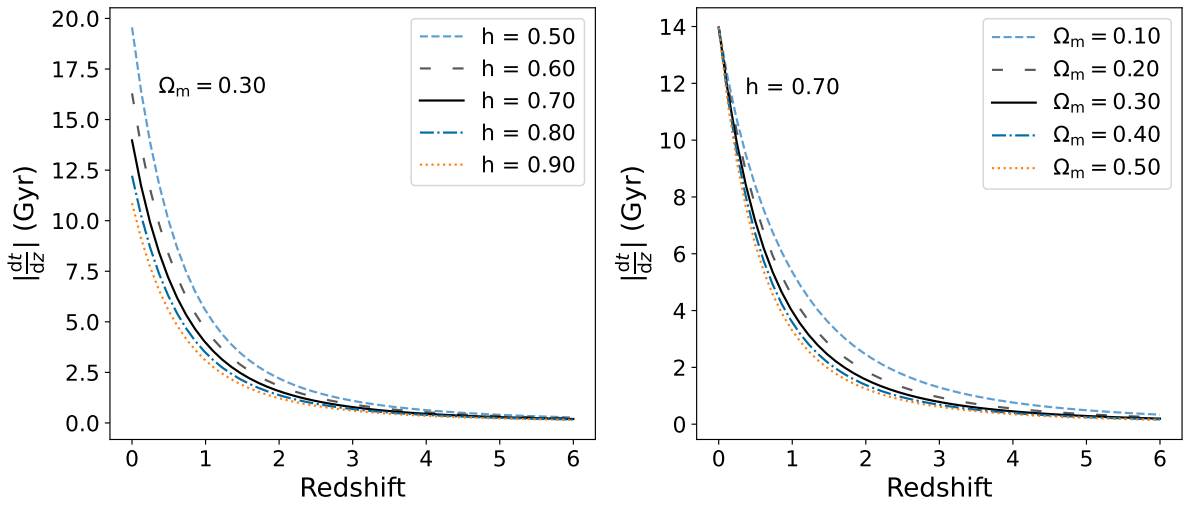

Figure 5.9: Behaviour of the cosmological $|\mathrm{d}t/\mathrm{d}z|$ factor for different Hubble constant values $h = H_0/(100\,\mathrm{km\,s^{-1}\,Mpc^{-1}})$, but fixed $\Omega_\mathrm{m} = 0.30$ (left) and different matter density parameter values, but fixed $h = 0.70$ (right).

Chapter 6

Probing the EBL with IACT data

Previously, de Matos Pimentel & Moura-Santos (2019) (P19) investigated EBL constraints from Markarian 501 and applied the best fit dust fractions to a sample of other spectra from known sources identified in the TeVCat catalog (Wakely & Horan 2008)[1], which were observed by various IACTs. Collecting these data — based on the methodology described in Chapter 3 and the results of Chapter 4 —, we take the next step and use the gathered spectra to perform a simultaneous inference of EBL and intrinsic spectral parameters of all sources. Section 6.1 describes the characteristics of the data sample, while in section 6.2 we discuss the choice of the intrinsic models and in section 6.3 the tuning of the HMC algorithm. Finally, in sections 6.4, 6.5 and 6.6, the results of the MCMC simulations are presented.

6.1 Data selection

From the initial sample gathered by P19, we have eliminated the spectra with less than 4 flux points and some sources with imprecise or unknown redshift. This resulted in a set of 65 gamma-ray spectra from 36 distinct AGNs. The unique sources are described in Table 6.1. The sample comprises blazars and radio galaxies, which are radio-loud AGNs. Most of the blazars are BL Lac objects, with the majority (25) being BL Lacs of the high-frequency peaked type (HBL), while 2 are intermediate-frequency (IBL) and one low-frequency peaked type (LBL). Besides blazars, four sources are Flat Spectrum Radio Quasars (FSRQ). The Radio Galaxies are all of the Fanaroff–Riley I (FR-I) type, a subclass for which the radio surface brightness decreases away from the galactic core and is typically less luminous than Fanaroff–Riley II objects (Schneider 2014). The only source with unclear classification is IC 310, as it was initially classified as a radio galaxy, but latter a blazar-like nucleus was found (a detailed and more recent discussion can be

[1] An online catalogue of TeV gamma-ray sources, which can be accessed in `http://tevcat.uchicago.edu/`.

Name	z	Type	Name	z	Type
1ES 0229+200	0.14	HBL	Centaurus A	0.00183	FR-I
1ES 0347-121	0.188	HBL	H 1426+428	0.129	HBL
1ES 0414+009	0.287	HBL	H 2356-309	0.165	HBL
1ES 0806+524	0.138	HBL	IC 310	0.0189	(Unknown)
1ES 1011+496	0.212	HBL	M87	0.0044	FR-I
1ES 1101-232	0.186	HBL	Markarian 180	0.045	HBL
1ES 1215+303	0.13	HBL	Markarian 421	0.031	HBL
1ES 1218+304	0.182	HBL	Markarian 501	0.034	HBL
1ES 1312-423	0.105	HBL	NGC 1275	0.017559	FR-I
1ES 1727+502	0.055	HBL	PKS 0447-439	0.343	HBL
1ES 1959+650	0.048	HBL	PKS 1441+25	0.939	FSRQ
1ES 2344+514	0.044	HBL	PKS 1510-089	0.361	FSRQ
1RXS J101015.9	0.142639	HBL	PKS 2005-489	0.071	HBL
3C 279	0.5362	FSRQ	PKS 2155-304	0.116	HBL
3C 66A	0.34	IBL	RBS 0413	0.19	HBL
4C +21.35	0.432	FSRQ	RGB J0152+017	0.08	HBL
AP Librae	0.049	LBL	RGB J0710+591	0.125	HBL
BL Lacertae	0.069	IBL	RX J0648.7+1516	0.179	HBL

Table 6.1: List of gamma-ray sources used to infer EBL parameters. The redshift and type information were extracted from TeVCat.

read in Gendron-Marsolais et al. 2020).

6.2 Choice of the intrinsic models

Most of the selected data in our work is also utilised by Biteau & Williams (2015) and they found that the majority of the intrinsic flux can be described by a power law, as intrinsic curvature models did not improve their estimation in such cases. Therefore, we start from the hypothesis that the PL model is a good description for all spectra and we proceed by doing MCMC simulations with each spectrum, simultaneously inferring their intrinsic (power law) parameters and the EBL dust fractions. The idea is to identify a priori unrealistic constraints in the EBL that could emerge from the wrong choice of the intrinsic model. This was apparent when some sources produced distributions for the PAH component such that $f_{\mathrm{PAH}} \gtrsim 0.70$, which results in an unreasonable bump in the mid-IR portion of the EBL compared to other EBL descriptions found in the literature. For these spectra, we repeated the MCMC simulation with the PLC and LP models. In some cases, we find a large degeneracy in the PLC model, due to the exponential cut-off parameter being poorly constrained. This means we need to impose an arbitrary upper limit on this parameter, as we find an essentially uniform distribution for it. Thus, we opt for the LP model in these situations, avoiding potential problems in the MCMC sampling. The choice of intrinsic curvature for some spectra was also supported by a search in the

literature. For instance, Markarian 421 spectrum is described with a source-inherent curvature, such as Albert et al. (2007b) using the PLC model for the MAGIC 2004–2005 data, or Acciari et al. (2011b) performing a fit to the VERITAS 2008 data with the same model, although not taking into account the EBL attenuation. Another example is the source PKS 2155–304, with the HESS 2006 data, as it is analysed with a log-parabola model in the context of axion-like particles constraints.

Finally, we pay particular attention to one source. As discussed by P19, the majority of the spectra in our work's sample are dominated by the attenuation coming from the stellar component of the EBL. However, there is an important flare data of Markarian 501, observed by the HEGRA telescope in 1997 (Aharonian et al. 2001), which is subject to a strong opacity from the dust components, since it is a low redshift source ($z = 0.034$) observed up to $\sim$20 TeV. This fact, as we shall see, means the choice of the intrinsic model for this particular spectrum can strongly affect the EBL inference. Therefore, we study three cases: we do the MCMC simulation without Mkn 501's flare data, then proceed to include it, firstly considering the PL model and secondly the LP one. The references for all spectra used in our work and the respective choice of model are presented in Tables E.1 to E.4 of section E.2 in Appendix E. We start in section 6.3 with a brief discussion on the adjustments made to the statistical method, followed by the results in section 6.4.

6.3 Tuning HMC

In our simple implementation of the HMC, it is necessary to manually tune some input parameters: the number of integration steps, their size, and the "mass" matrix. They heavily influence the efficiency of the algorithm, so we have tried to adjust them (by trial and error) until a high acceptance fraction was achieved with a sufficiently low integrated autocorrelation time. We have found that an integration step size of $\Delta t = 8 \times 10^{-3}$ and $L = 80$ steps was enough for our purposes and it maintained the total "energy" approximately conserved. To choose the mass matrix, we emphasise that the inverse mass matrix resembles the posterior covariance matrix (Betancourt 2017), so we did an iterative process by starting with a diagonal matrix where its entries are of the order of the typical variance of each parameter. Then, doing some preliminary runs, we refined the estimate by computing the covariance matrix and took only the main diagonal. By working with a diagonal matrix, we wanted to minimise possible numerical errors during matrix multiplication or inversion operations and reduce the number of free variables in our tuning. Besides, for the non-diagonal elements, we expect vanishing correlations between most of the parameters, as the spectral parameters are independent. Once we achieved a mass matrix that stabilised the simulation, we did some fine-tuning on the "masses" of specific parameters when we had observed high autocorrelation values. For three sources in particular it was more challenging to lower their autocorrelation than the

rest: PKS 1510-089, 1ES 1011+496 and 3C 279, so we first did a simulation without these sources and then progressively included them by finding compatible mass values. During these tests we realised that the distribution of EBL parameters (and the other intrinsic parameters) appeared to be unchanged with or without their inclusion.

Once the integrated autocorrelation time was computed, after eliminating correlated samples, we have obtained 6195 effective samples for the simulation without Markarian 501 flare data, 11075 samples with Mkn 501 modelled as PL and 11539 samples with the LP model. Fixing the confidence in 95%, these ESS translate to efficiencies of 5.64%, 4.22% and 4.13% respectively.

6.4 Overall Results

A comparison between the marginal distributions of EBL parameters from the three MCMC runs can be seen in Figure 6.1, alongside the median values and respective uncertainties. One clear effect of the addition of Markarian 501's flare data is apparent: it is essential to constrain the SG and LG components, as its absence leaves essentially flat distributions for them. On the other hand, the PAH fraction is still very much constrained and its distribution is consistent between the three cases. This reveals to us that the combination of the other 64 spectra was very important to probe the mid-IR portion of the EBL. Comparing the results using the two different intrinsic models for Markarian 501, we see that the PL choice gives more stringent constraints to the SG and LG fractions, while the LP model results in broader distributions and consequently in larger uncertainties. This is consistent with the fact that the use of a PL model for the intrinsic spectrum implies that any curvature in the measured spectrum should be accounted by the EBL attenuation, whereas the LP one allows for an intrinsic component to also explain the decreasing flux. Therefore, the marginal distribution of the dust fractions is widened by the introduction of an intrinsic curvature parameter, as it enables more EBL configurations, otherwise unfeasible with the PL description.

The median values are consistent with the best fit results published by P19, but we find a noticeable improvement in the PAH constraint, reaching a smaller uncertainty. We attribute this to the combination of many spectra in our likelihood — as observed in the synthetic sample on Chapter 4 —, since P19's results come solely from the fit of Markarian 501 flare data. Such improvement is irrespective of the model chosen for Markarian 501, as P19 found $f_{\mathrm{PAH}} = 0.32 \pm 0.15$ for the PL case, while we found $f_{\mathrm{PAH}} = 0.219^{+0.047}_{-0.045}$: an uncertainty three times smaller in absolute terms. For the LP model, P19 obtained $f_{\mathrm{PAH}} = 0.27 \pm 0.25$, an uncertainty five times higher than our result of $f_{\mathrm{PAH}} = 0.214^{+0.047}_{-0.044}$. Interestingly, there is a bigger shift in the PAH value when P19 changes the intrinsic model for Markarian 501, compared to the near stable behaviour of our inference. This reinforces that the combination of distinct sources increases the robustness of the results. For the

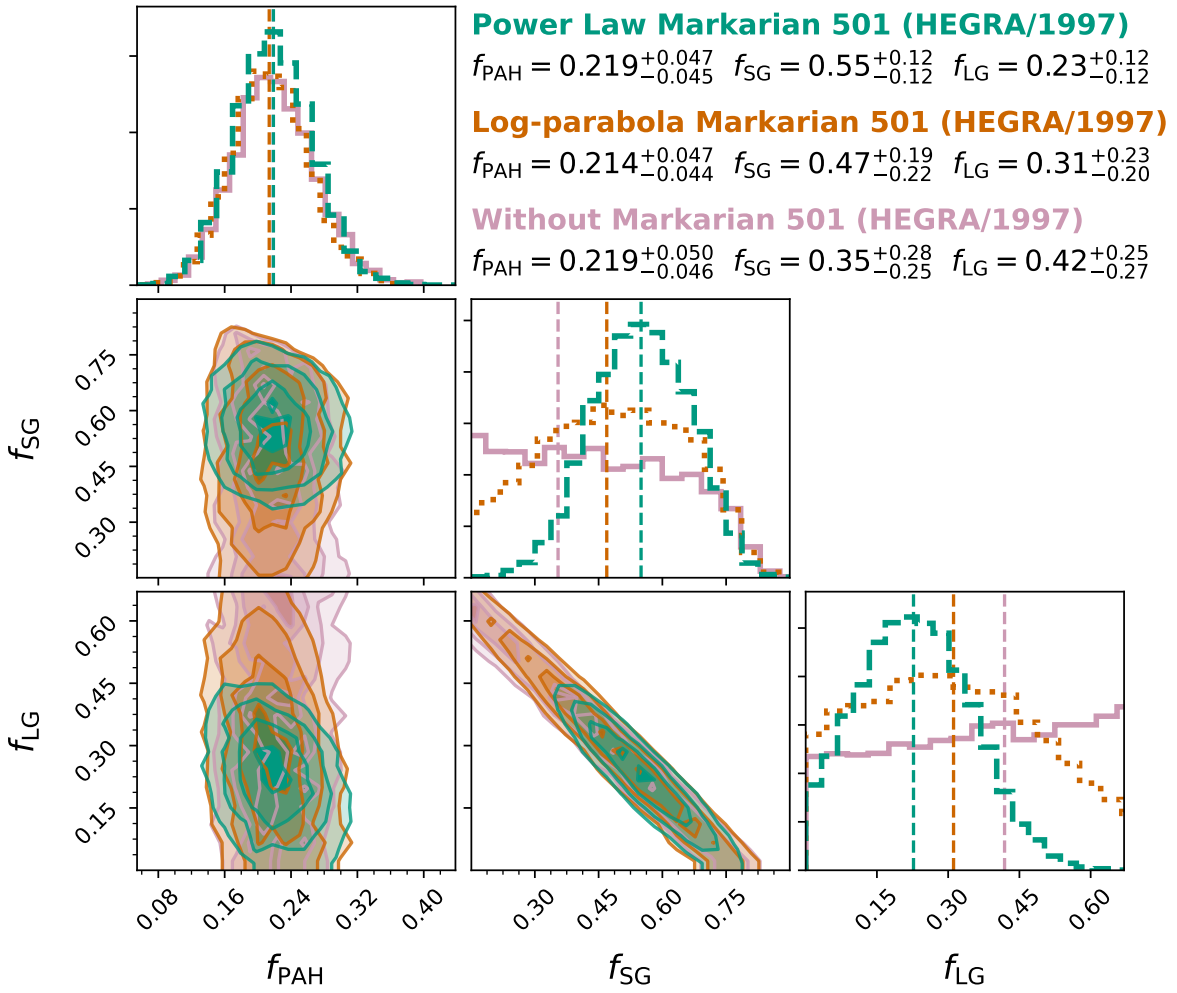

Figure 6.1: Marginal distributions of EBL dust fractions from the three HMC samples. The green dashed line corresponds to the case with Markarian 501 flare data modelled as power-law, the dotted orange one as log-parabola and the purple solid line corresponds to the case without this spectrum. The median of the histograms are represented as the vertical lines, while the respective values with their uncertainties are also shown.

SG fraction, however, we find very similar results, as P19 obtained $f_{SG} = 0.56 \pm 0.12$ using the PL model (equivalent to our result of $f_{SG} = 0.55^{+0.12}_{-0.12}$) and, with the LP model, their result of $f_{SG} = 0.49 \pm 0.28$ also agrees with ours, of $f_{SG} = 0.47^{+0.19}_{-0.22}$, but our uncertainty is a bit smaller. Once again, the majority of the data in the sample has little to none information regarding the far-IR of the EBL.

Figure 6.2 presents the local EBL intensity of the LP case, compared to other models and some upper and lower limits found in the literature for reference. Our result is consistent to the F10+P19 nominal case and D11, having an EBL approximately in between both models, up to $\sim$100 µm. We also observe a particular agreement with the D11 curve around $\sim$ 30 µm, which indicates an early rise of the second IR peak compared to F10 model. Consequently, we find a higher value for the SG fraction. Figure 6.2 also

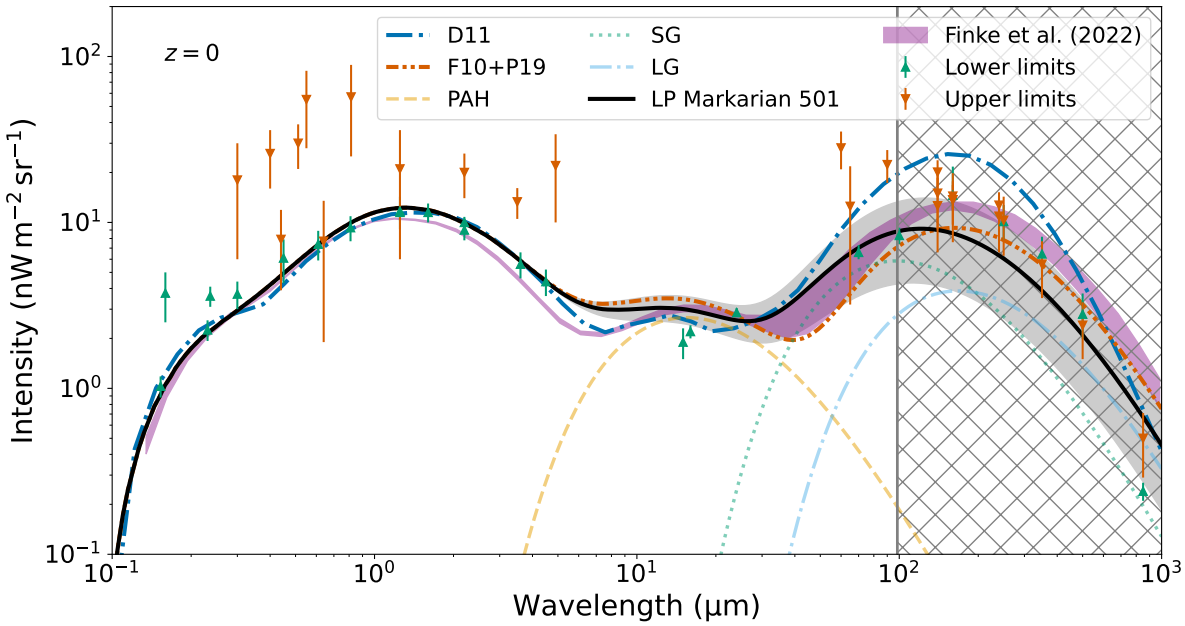

Figure 6.2: Local EBL spectrum from the combined analysis modelling Markarian 501 flare emission as a log-parabola. The shaded grey region indicates the 1σ interval due to the uncertainty in the dust fractions. The individual contributions of the PAH, SG and LG components are also presented, alongside the F10, D11 and Finke et al. (2022) models. Lower and upper limits of the EBL intensity were taken from Biteau & Williams (2015).

shows an updated EBL model based on F10, that was recently published (Finke et al. 2022). The authors worked with improved stellar models, a different parameterization of the star formation rate and they have also incorporated metalicity and dust extinction evolution with redshift. Using a variety of data (EBL opacity from LAT and IACTs; luminosity, stellar density and dust extinction from galaxy surveys), they have fitted the model parameters, including the dust fractions. This updated result lowers a bit the stellar component of the EBL and it gives new values for the SG and LG fractions, of $f_{\mathrm{SG}} = 0.26^{+0.18}_{-0.17}$ and $f_{\mathrm{LG}} = 0.56^{+0.17}_{-0.18}$. However, the LG dust is now warmer, at a temperature of $T_{\mathrm{LG}} = 60.5^{+2.3}_{-3.5}\,\mathrm{K}$, higher than F10, where $T_{\mathrm{LG}} = 40\,\mathrm{K}$. This turns this component closer to the SG temperature of $T_{\mathrm{LG}} = 70\,\mathrm{K}$. Nevertheless, we find a reasonable agreement with our result in the range 10–100 μm.

The EBL results can also be presented in the form of a contour plot, showing the changes in optical depth for different observed gamma ray energies and redshifts of emission, as in Figure 6.3. This plot also emphasises the Cosmic Gamma-ray Horizon (CGRH), corresponding to combinations of energy and redshit for which the optical depth $\tau = 1$. Since there are uncertainties in the dust fraction, we also plot the the 1σ limits in dotted lines. Therefore, we can see that the dust component of the EBL is mainly responsible for the low redshift ($z < 0.1$) and very high energy ($E_\gamma > 1\,\mathrm{TeV}$) parts of the CGRH. The last energy bins of all the spectra used in the likelihood are presented as the

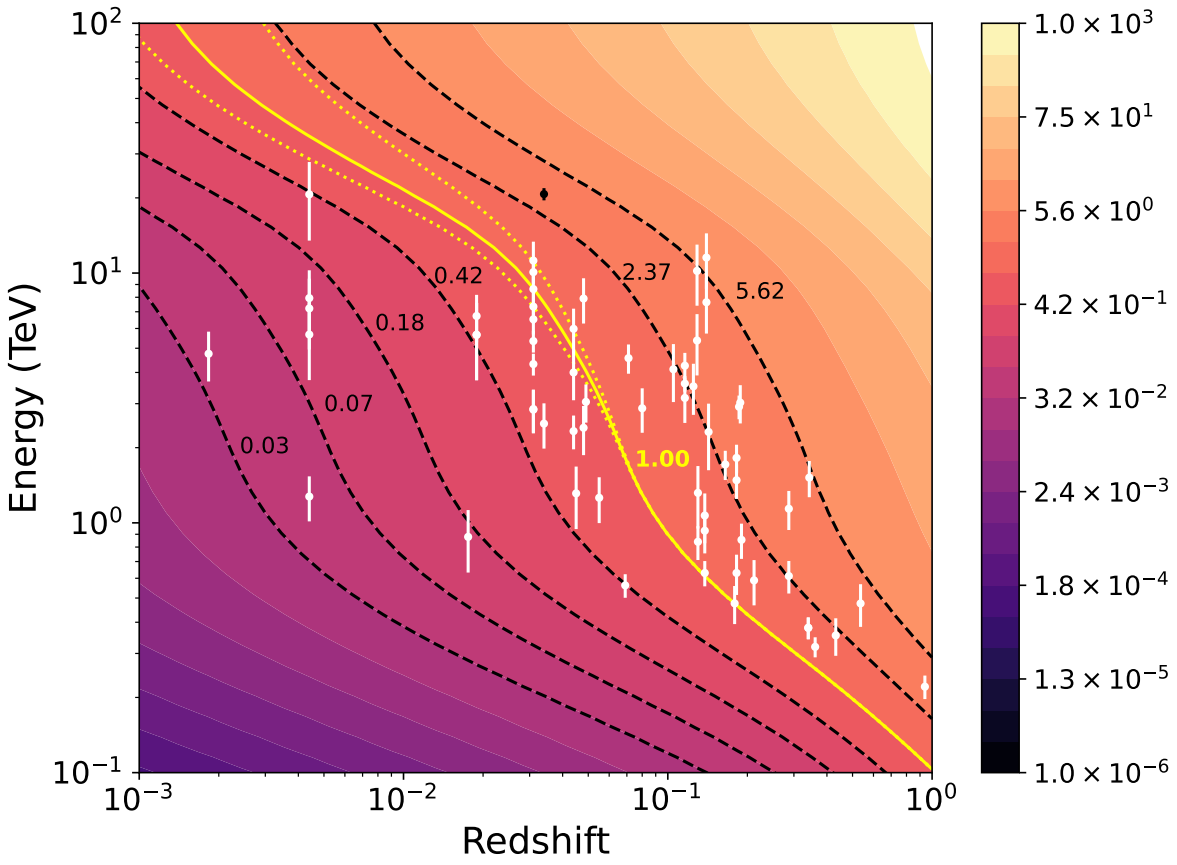

Figure 6.3: Optical depth (τ) as a function of the observed gamma ray energy and the redshift of emission, considering the median dust fractions of the HMC result with Mkn 501 as LP. The yellow line corresponds to the cosmic gamma-ray horizon ($\tau = 1$), with the dotted lines indicating the uncertainty in the dust fractions. Other values of the optical depth are marked for reference. The white dots are the last energy bin of all the spectra used in the likelihood, with Mkn 501 flare data highlighted in black.

white dots, with uncertainty corresponding to the half distance in logarithm scale to the previous energy bin. The point representing Markarian 501 flare data is highlighted in black. Due to its low redshift and having the highest energy in our sample, we can see that it is a very important measurement to probe the CGRH in the IR region.

A complementary information is the mean free path computed from the local EBL spectrum, tuned to the dust fractions obtained in the LP case for Mkn 501. Figure 6.4 presents the corresponding curve, according to the reciprocal of eq. 3.44, also showing the contributions from each individual EBL component. For gamma ray energies above 100 TeV, we need to take into consideration the CMB, as this is the dominant background radiation field. For energies above $\sim 10^{19}$ eV we should also consider the radio background (see e.g. De Angelis et al. 2013), but we have omitted this aspect in the figure. At the typical energies of our data sample, of a few TeV, the mean free path is mostly determined by the stellar and PAH components, being of the order of a few hundreds of Mpc. Only observations above ~20 TeV can be more substantially affected by the IR emission from

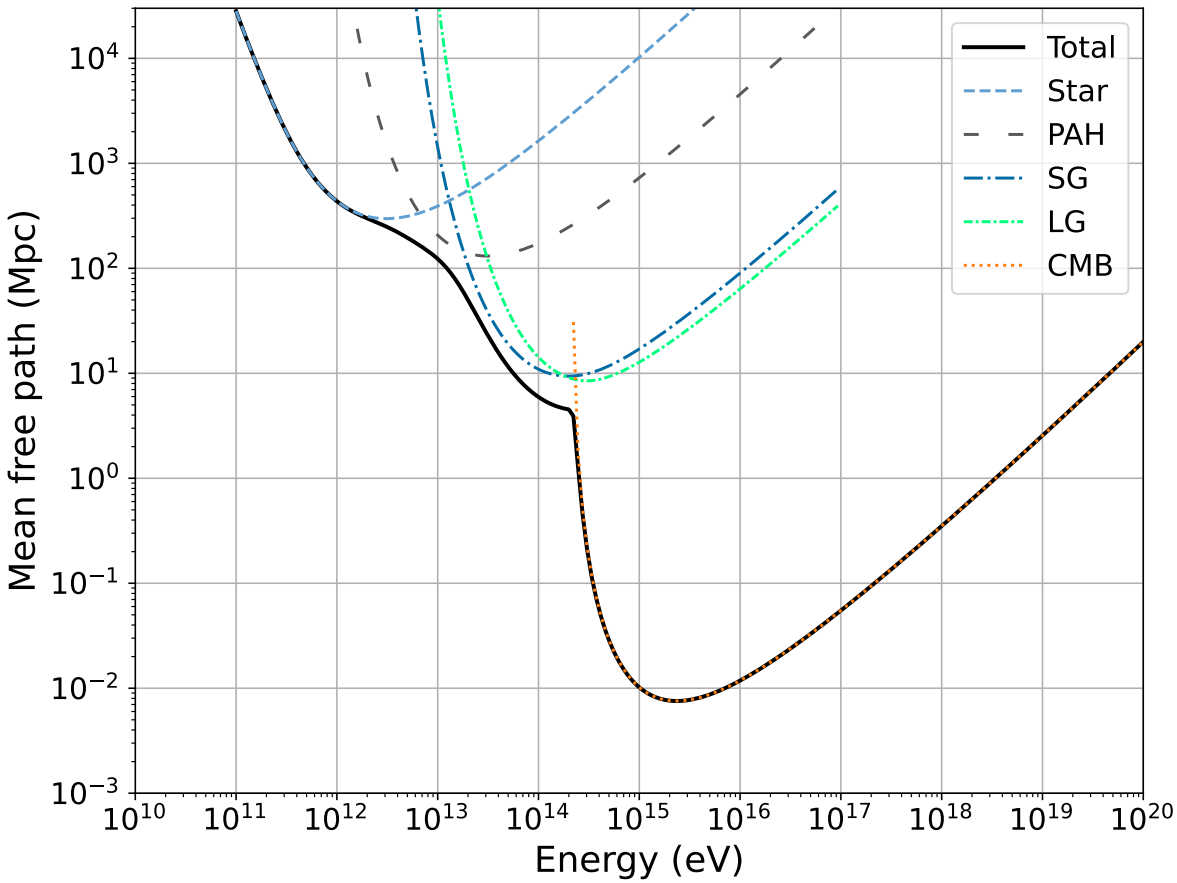

Figure 6.4: Mean free path for photons as a function of their energy, considering the Breit-Wheeler interaction with the EBL or CMB, at $z = 0$. EBL computed according to the median dust fractions inferred from the HMC result with Mkn 501 modelled as LP; the individual contributions from the stellar part and dust components (PAH, SG, LG) are also shown. The CMB follows a Planck spectrum with temperature $T_0 = 2.726\,\mathrm{K}$.

small grains, but the corresponding free mean path rapidly drops below 100 Mpc.

6.4.1 Discussing intrinsic spectral parameters

In terms of the intrinsic spectral parameters of the sources, Figure 6.5 reveals that the median spectral indices (including a and Γ of LP and PLC models) and flux normalisation values are very consistent in all three MCMC samplings performed. Therefore, although Markarian 501's inclusion and model choice has a big impact on EBL constraints, they do not interfere on the marginal distribution of spectral parameters for the remaining sources. With this in mind, Tables 6.2 to 6.5 present the resulting median values and uncertainties from the LP case for all 65 spectra.

To discuss the overall adequacy of the posterior distribution in describing all the spectra, we have selected a random sample of 100 models from the posterior and have computed the residuals in relation to the data. The plot on the left of Figure 6.6 presents

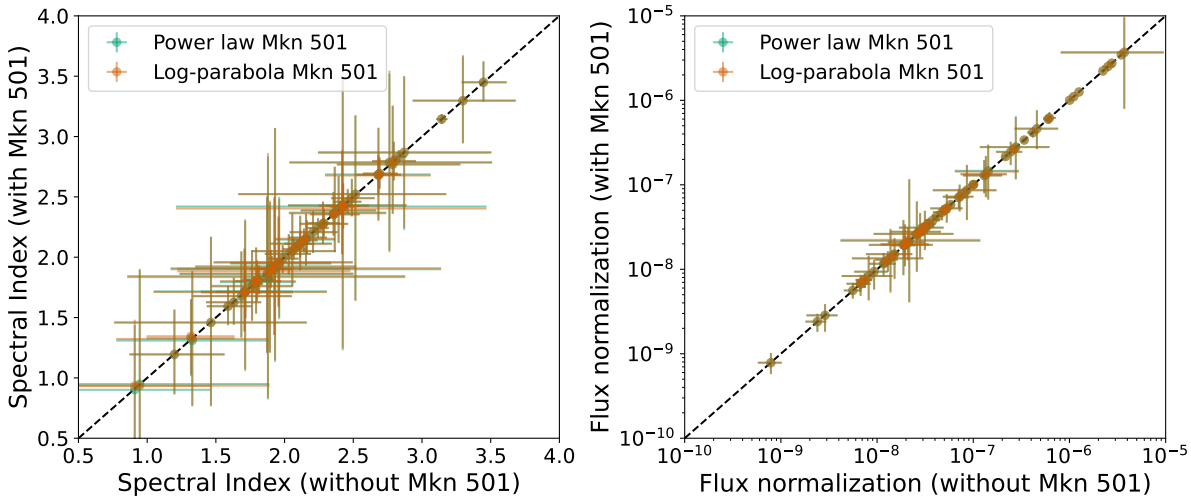

Figure 6.5: Comparison between the inferred spectral indices (left) and flux normalization parameters (right) for the three HMC simulations performed (without Mkn 501 flare data, or including it as a power-law or log-parabola). The diagonal dashed line is the identity.

the resulting histogram with an unbinned likelihood fit. We see that the standard deviation is close but statistically below 1, while the mean is above zero, so in average the models seem to underestimate the data. We should mention that the low uncertainties come partially from the high statistics from the combined residuals of all sources and the 100 models. Also, naturally, we do not expect a perfect agreement to the data, as the F10 model may not exactly describe the true EBL and analogously the flux parametrization regarding the intrinsic spectra. Not to mention the fact that the uncertainties on the measured fluxes are likely to have non-Gaussian contributions. Nevertheless, we find that the simultaneous inference of EBL and spectral parameters was able to adequately describe the observed fluxes, with no significant outliers. On the right of Figure 6.6, we see the resulting distribution of spectral indices. This reveals that most of the sources have a spectral index near $\Gamma = 2$, with the maximum value being 3.45 and the lowest one being 0.90. Six of them have a median index smaller than 1.5: 1ES 0229+200 (HESS/2005-2006), 1ES 0440+009 (VERITAS/2008-2011), 3C 279 (MAGIC/2008), H1426+428 (HEGRA/2002), PKS 0447-439 (HESS/2009) and RBS 0413 (VERITAS/2009). However, only H1426+428, with $\Gamma = 0.90^{+0.53}_{-0.50}$ has more than 1σ difference to $\Gamma = 1.5$. This result essentially agrees within the statistical uncertainty with Biteau & Williams (2015), as they have found $\Gamma = 1.37 \pm 0.30$ for the same source.

Finally, Figures E.1 to E.6 of section E.3 in Appendix E show the fit for all remaining spectral data. Besides the median fit with residuals, we show the 100 random models used to build the histogram of Figure 6.6 and the de-absorbed data with uncertainty propagation from the dust fractions. In what follows, we are going to discuss some particular sources, comparing to other works on the literature. We have selected only a few cases for which it was possible to draw more direct comparisons, so this should not be

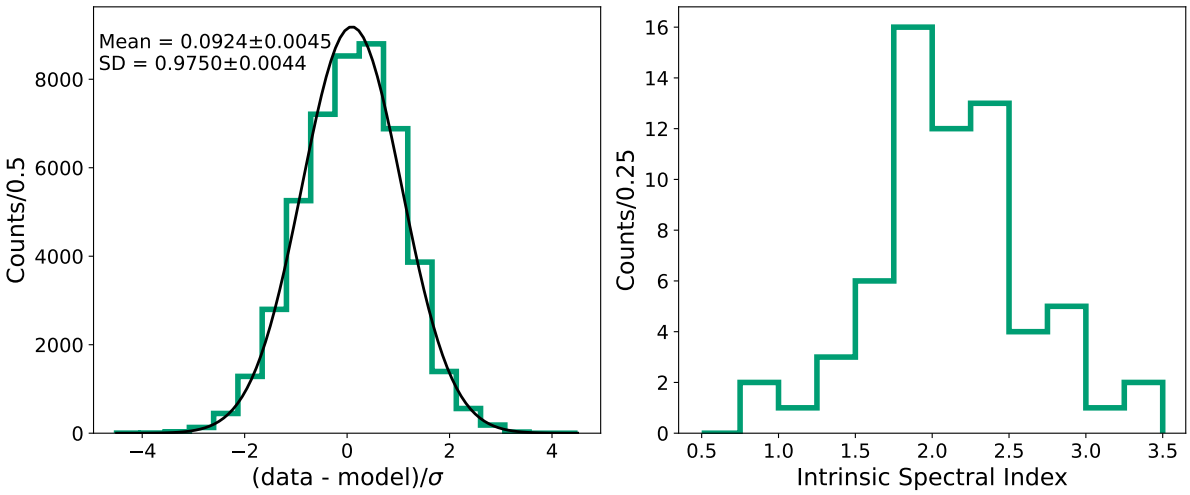

Figure 6.6: *Left*: Pull distribution from the global fit of 100 models sampled from the posterior probability. A Gaussian fit is presented, alongside the mean and standard deviation values. *Right*: the distribution of median spectral indices of all sources; the a and Γ parameters from the LP and PLC models were also included.

taken as a complete inspection of the results. The posterior distribution of the intrinsic parameters off all sources can be found in Figures E.8 (Mkn 501) and E.9–E.31 of section E.4 in Appendix E.

6.5 Markarian 501

In Figure 6.7, we show the fit to the Mkn 501 flare data considering both the PL and LP models. Besides the curve constructed from the median values of the parameters, we also present the de-absorbed data with the intrinsic spectrum by removing the corresponding EBL attenuation. The new flux points take into account the uncertainty propagation coming from the covariance matrix of the dust fractions. As seen in the residuals and χ^2 values produced by the median fit, both PL and LP models can adequately represent the observed data. Besides, the intrinsic emission is qualitatively the same, which is compatible with the fact that the LP model resulted in a curvature parameter b statistically consistent with zero (see table 6.2 for the exact value). This reinforces a robust inference of the intrinsic parameters of Mkn 501, as even with the different parametrizations and EBL parameters, both results still agree. Comparing to the results by de Matos Pimentel & Moura-Santos (2019), our PL case resulted in a spectral index $\Gamma = 2.31^{+0.13}_{-0.13}$, while theirs in $\Gamma = 2.05 \pm 0.39$, which is also consistent to the fact that they inferred a higher PAH fraction. In the LP scenario, our value of $a = 2.06^{+0.57}_{-0.58}$ and $b = 0.18^{+0.43}_{-0.39}$ are compatible too with their result of $a = 1.96 \pm 0.48$ and $b = 0.16 \pm 0.60$.

In other works, we find a similar agreement with our results. Adopting their own EBL model, Franceschini et al. (2008) found $\Gamma = 1.94$, although using the flux points

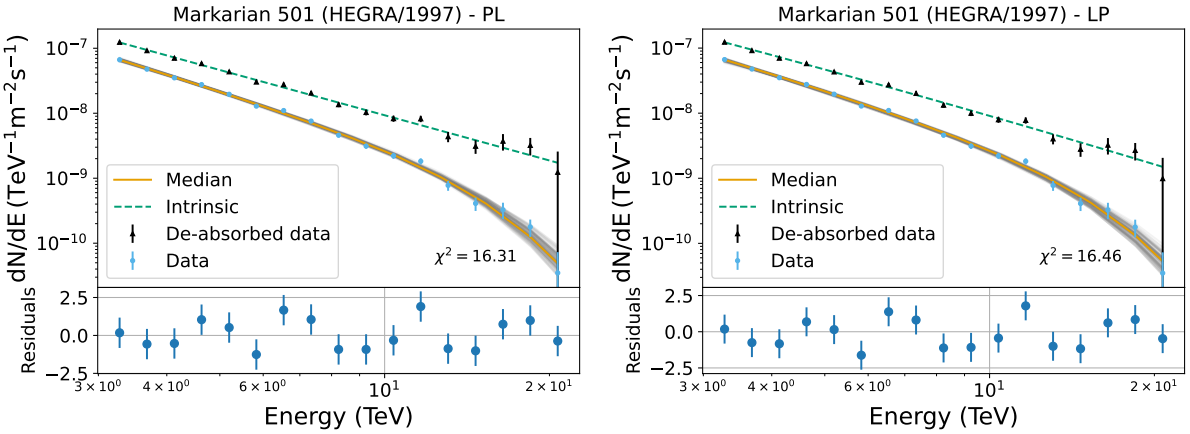

Figure 6.7: Spectrum of Mkn 501 (HEGRA/1997) fitted as a power law (left) and log-parabola (right) model. The residuals correspond to the curve constructed with the median values of the parameters (orange line). The de-absorbed points take into account the uncertainty propagation of EBL dust fractions and their covariance matrix. In grey, there are plotted 100 random models sampled from the posterior distribution.

prior to the reanalysis by Aharonian et al. (2001). As pointed by D11, the Mkn 501 flare spectrum prior to the reanalysis can suggest an upturn in its intrinsic emission after EBL correction, with a clear deviation from the PL line. Such behaviour disappears with the reanalysis data, as we can also confirm in Figure 6.7, since the last energy bin follows the expected fall in flux. Our result is also consistent to the spectral index value found by D11, of $\Gamma = 2.37^{+0.24}_{-0.20} \pm 0.04$, considering also systematic errors.

Finally, we point out that the Markarian 501 data in quiescent state, as observed by VERITAS (2009) (Acciari et al. 2011a), has a possibly softer spectral index, of $\Gamma = 2.49^{+0.16}_{-0.15}$. This result is compatible with the value of $\Gamma = 2.39 \pm 0.65$ proposed by Zhong et al. (2018), when using the average EBL model from Dwek & Krennrich (2005).

6.6 Other sources

- A total of 7 spectra from Mkn 421 were considered to have some kind of intrinsic curvature and in all cases we found that the curvature parameter is non-negligible. For the PLC spectra, in particular, all have an energy cut-off well defined below 4 TeV. Specifically for the MAGIC (2004–2005) data, our results are in good agreement with Albert et al. (2007b). Using a scaled version of the EBL from Primack et al. (2005), they found a photon index $\Gamma = 2.20 \pm 0.08$ and a cut-off energy $E_{\rm cut} = (1.44 \pm 0.27)$TeV, both consistent with our results of $\Gamma = 2.242^{+0.080}_{-0.079}$ and $E_{\rm cut} = 1.64^{+0.43}_{-0.29}$ TeV.

- The H.E.S.S. (2006) spectrum of PKS 2155–304 was analysed by Abramowski et al. (2013) to determine upper limits on the strength of the coupling between axion-like particles (ALPs) and photons. By considering Franceschini et al. (2008) EBL

model and the LP parametrization of the intrinsic spectrum, they found a photon index $a = 3.18 \pm 0.03\,(\text{stat}) \pm 0.20\,(\text{syst})$ and a curvature parameter $b = 0.32 \pm 0.02\,(\text{stat}) \pm 0.05\,(\text{syst})$. Although our photon index closely matches their result, with $a = 3.145^{+0.040}_{-0.040}$, we found a relatively higher curvature, of $b = 0.749^{+0.068}_{-0.068}$. However, considering the systematic errors, both results are probably not in significant tension.

- The time-averaged spectrum of PKS 2005–489, measured by H.E.S.S. between 2004 and 2007, was described by H. E. S. S. Collaboration et al. (2010a) using a PL corrected by the EBL model from Franceschini et al. (2008). They have found an intrinsic spectral index of $\Gamma = 2.69 \pm 0.16$, which is essentially the same result as ours, of $\Gamma = 2.68^{+0.17}_{-0.15}$.

- H 1246+428 data from HEGRA (1999-2000) is presented and discussed by Aharonian, F. et al. (2002). They affirm that the X-ray spectrum of this source, up to 100 keV, favours a spectral index of 1.9, which they assume to represent the TeV intrinsic emission; also taking into consideration two different EBL models to describe the observed data. Curiously, we have obtained in our analysis — without any a priori assumptions — the same spectral index, although with a relatively large uncertainty, of $\Gamma = 1.9^{+1.2}_{-0.8}$.

- The high redshift ($z = 0.536$) source 3C 279 was analysed by MAGIC Collaboration et al. (2008) to probe the CGRH. Using two different EBL models, they find two "extreme" values for the intrinsic photon index, of $\Gamma = 2.9 \pm 0.9\,(\text{stat}) \pm 0.5\,(\text{syst})$ (low EBL, Primack et al. 2005) and $\Gamma = 0.5 \pm 1.2\,(\text{stat}) \pm 0.5\,(\text{syst})$ (high EBL, Stecker et al. 2006). Our result, of $\Gamma = 0.94^{+0.96}_{-0.66}$, is in between both values, although closer to the high EBL case. Overall, the uncertainties are large. Notably, D11 presents a much higher photon index for 3C 279, of $\Gamma = 3.78^{+0.10}_{-0.08} \pm 0.88$, which can be partially due to differences in the EBL evolution, as this is a factor that may impact the interpretation of more distant sources. In any case, a "flattening" of the spectrum in the higher energy bins is identified, as we can see in the bottom-right panel of Figure E.7 in Appendix E. According to D11, an external photon field feeding the inverse Compton mechanism may be required to physically explain this behaviour.

- The source 1ES 1011+496 was observed by MAGIC (2007) and the analysis from Albert et al. (2007c) points to a PL spectrum with photon index $\Gamma = 3.3 \pm 0.7$, taking into account the EBL attenuation from Kneiske et al. (2002b). Our result suggests a lower value of $\Gamma = 2.87^{+0.63}_{-0.63}$, but still compatible under the uncertainties.

- The time-average spectrum of 1ES 2344+514 in quiescent state observed from 2007-2008 and the flare data in December of 2007, both measured by VERITAS, were

analysed by Acciari et al. (2011c). Considering the EBL opacity from Franceschini et al. (2008), they estimated intrinsic spectral indices of $\Gamma \approx 2.5$ and $\Gamma \approx 2.1$ for the low and high flux states, respectively. Our results are very similar, as we have obtained $\Gamma = 2.46^{+0.10}_{-0.10}$ and $\Gamma = 2.11^{+0.24}_{-0.22}$.

- Discussing EBL constraints from the H.E.S.S. (2005–2006) observations of 1ES 0229+200, Aharonian et al. (2007b) identifies that the EBL models from Stecker et al. (2006) result in harder spectral indices of $\Gamma = 0.6 \pm 0.3$ or below, which would require more extreme assumptions to explain this emission. The model from Primack et al. (2005), on the other hand, results in $\Gamma = 1.92 \pm 0.22$. However, they also reject this scenario as this EBL model falls below lower limits in the mid-IR given by Spitzer. Our result of $\Gamma = 1.35^{+0.31}_{-0.31}$ lies in between both cases and it is statistically compatible with $\Gamma = 1.5$.

- The spectrum of 1ES 0347–121 measured by H.E.S.S. (2006) was analysed by Aharonian et al. (2007a) using the EBL model from Primack et al. (2005) and the upper limit curve from Aharonian et al. (2006b), obtaining the spectral index values of $\Gamma = 2.10 \pm 0.21$ and $\Gamma = 1.69 \pm 0.22$, respectively. Our result more closely matches the second one, as we have found $\Gamma = 1.71^{+0.36}_{-0.32}$.

- The analysis by Aleksić et al. (2014b) of IC 310, observed by MAGIC (2009–2010), provides the intrinsic spectral indices of $\Gamma = 1.81 \pm 0.13\,(\text{stat}) \pm 0.20\,(\text{syst})$ and $\Gamma = 1.85 \pm 0.11\,(\text{stat}) \pm 0.20\,(\text{syst})$ for the low and high states, respectively, considering the EBL attenuation from D11. Our result is compatible to both, with the corresponding median values of $\Gamma = 1.80^{+0.17}_{-0.16}$ and $\Gamma = 1.81^{+0.11}_{-0.11}$.

6.7 Discussion

By combining multiple spectra of different AGNs, it was possible to obtain more robust constraints on the mid-IR region of the EBL. The intrinsic spectral parameters we have inferred are generally consistent to other results found in the literature, as well as the resulting EBL is compatible to other models in the ∼10–100 µm wavelength range. Improved constraints on the far-IR, on the other hand, are limited by current measurements, as only the detection of nearby sources at energies of dozens of TeV — such as Markarian 501 flare spectrum — can contribute to it. This suggests the important role of CTA in expanding the data, given its improved sensitivity at VHEs. The HMC method can be a useful tool for incorporating this new data, as it performs very well in a space with a high number of parameters, although, in its simplest implementation, it may be challenging to manually tune its efficiency.

Name	Survey/Year - Label	$N_0(*)\times10^{-6}$	Γ	$E_{\rm cut}$ (**)
Markarian 421	MAGIC/2004-2005	$0.417^{+0.062}_{-0.053}$	$2.242^{+0.080}_{-0.079}$	$1.64^{+0.43}_{-0.29}$
	VERITAS/2008 - High A	$2.53^{+0.39}_{-0.32}$	$1.81^{+0.12}_{-0.12}$	$2.06^{+0.59}_{-0.40}$
	VERITAS/2008 - Low	$1.006^{+0.050}_{-0.047}$	$2.097^{+0.038}_{-0.038}$	$3.27^{+0.45}_{-0.36}$
	VERITAS/2008 - Very Low	$0.62^{+0.10}_{-0.09}$	$2.11^{+0.11}_{-0.11}$	$1.76^{+0.49}_{-0.33}$
Name	Survey/Year - Label	$N_0(*)\times10^{-6}$	a	b
Markarian 421	VERITAS/2008 - High B	$2.23^{+0.12}_{-0.13}$	$2.03^{+0.16}_{-0.17}$	$0.62^{+0.27}_{-0.26}$
	VERITAS/2008 - Middle	$1.107^{+0.019}_{-0.019}$	$2.359^{+0.022}_{-0.022}$	$0.315^{+0.049}_{-0.051}$
	VERITAS/2008 - Very High	$3.46^{+0.14}_{-0.14}$	$1.99^{+0.11}_{-0.12}$	$0.83^{+0.21}_{-0.20}$
Markarian 501	HEGRA/1997	$1.58^{+0.91}_{-0.59}$	$2.06^{+0.57}_{-0.58}$	$0.18^{+0.43}_{-0.39}$
PKS 2155–304	H.E.S.S./2006	$1.256^{+0.021}_{-0.021}$	$3.145^{+0.040}_{-0.040}$	$0.749^{+0.068}_{-0.068}$

Table 6.2: References of the IACT spectra utilized for PLC and LP sources, alongside the median values and uncertainties (68% interval) of their respective parameters; (*) in $\mathrm{TeV}^{-1}\mathrm{m}^{-2}\mathrm{s}^{-1}$ and (**) in TeV.

Name	Survey/Year - Label	$N_0(*)\times10^{-8}$	Γ
1ES 0229+200	HESS/2005–2006	$3.06^{+0.58}_{-0.59}$	$1.35^{+0.31}_{-0.31}$
	VERITAS/2010–2012	$2.57^{+0.33}_{-0.34}$	$1.59^{+0.17}_{-0.15}$
1ES 0347–121	HESS/2006	$3.81^{+0.71}_{-0.72}$	$1.71^{+0.25}_{-0.22}$
1ES 0414+009	HESS/2005–2009	$2.1^{+1.0}_{-0.8}$	$1.71^{+0.36}_{-0.32}$
	VERITAS/2008–2011	$0.83^{+0.61}_{-0.40}$	$1.32^{+0.56}_{-0.55}$
1ES 0806+524	MAGIC/2011 - Low	$2.8^{+1.9}_{-1.4}$	$1.71^{+0.60}_{-0.66}$
	MAGIC/2011 - High	$7.2^{+2.8}_{-2.4}$	$2.05^{+0.31}_{-0.29}$
	VERITAS/2006-2008	$1.9^{+1.9}_{-1.1}$	$1.8^{+1.0}_{-1.0}$
1ES 1011+496	MAGIC/2007	$2.6^{+3.8}_{-1.7}$	$2.87^{+0.63}_{-0.63}$
1ES 1101-232	HESS/2004-2005	$3.44^{+0.62}_{-0.63}$	$1.63^{+0.20}_{-0.19}$
1ES 1215+303	MAGIC/2011	$1.92^{+0.56}_{-0.52}$	$2.43^{+0.19}_{-0.19}$
	VERITAS/2011	$1.17^{+0.40}_{-0.35}$	$2.68^{+0.40}_{-0.38}$
1ES 1218+304	VERITAS/2008-2009	$10.0^{+1.2}_{-1.1}$	$1.91^{+0.10}_{-0.10}$
	VERITAS/2007	$7.1^{+1.9}_{-1.8}$	$1.68^{+0.37}_{-0.35}$
	MAGIC/2005	$8.7^{+8.5}_{-4.9}$	$1.88^{+0.59}_{-0.68}$
1ES 1312-423	HESS/2004-2010	$0.67^{+0.19}_{-0.19}$	$1.95^{+0.39}_{-0.36}$
1ES 1727+502	VERITAS/2013	$4.74^{+0.93}_{-0.95}$	$1.76^{+0.30}_{-0.29}$
1ES 1959+650	VERITAS/2007-2011	$10.11^{+0.87}_{-0.89}$	$2.14^{+0.097}_{-0.099}$
	MAGIC/2006	$4.28^{+0.57}_{-0.57}$	$2.21^{+0.18}_{-0.17}$
1ES 2344+514	VERITAS/2007-2008	$5.26^{+0.36}_{-0.36}$	$2.46^{+0.10}_{-0.10}$
	VERITAS/2007 - High	$26.5^{+3.1}_{-3.1}$	$2.11^{+0.24}_{-0.22}$
	MAGIC/2005-2006	$2.66^{+0.42}_{-0.43}$	$2.69^{+0.13}_{-0.13}$
1RXS J101015.9	HESS/2006-2010	$0.94^{+0.36}_{-0.36}$	$1.96^{+0.53}_{-0.45}$
3C 279	MAGIC/2008	370^{+610}_{-290}	$0.94^{+0.96}_{-0.66}$

Table 6.3: References of the IACT spectra utilized for PL sources, alongside the median values and uncertainties (68% interval) of their respective parameters; (*) in $\mathrm{TeV}^{-1}\mathrm{m}^{-2}\mathrm{s}^{-1}$.

Name	Survey/Year - Label	$N_0(*)\times10^{-8}$	Γ
3C 66A	VERITAS/2008 - Low	14^{+16}_{-8}	$1.93^{+0.57}_{-0.57}$
	VERITAS/2008 - High	28^{+36}_{-16}	$1.87^{+0.63}_{-0.65}$
4C +21.35	MAGIC/2010	46^{+31}_{-20}	$2.39^{+0.28}_{-0.27}$
AP Librae	HESS/2010-2011	$0.75^{+0.15}_{-0.15}$	$2.35^{+0.19}_{-0.17}$
BL Lacertae	VERITAS/2011	$12.9^{+6.8}_{-5.0}$	$3.30^{+0.38}_{-0.36}$
Centaurus A	HESS/2004-2008	$0.241^{+0.056}_{-0.058}$	$2.77^{+0.48}_{-0.41}$
H 1426+428	HEGRA/1999-2000	$13.4^{+8.6}_{-6.2}$	$1.9^{+1.2}_{-0.8}$
	HEGRA/2002	$2.0^{+1.5}_{-0.9}$	$0.93^{+0.55}_{-0.51}$
H 2356-309	HESS/2004-2007	$2.02^{+0.28}_{-0.28}$	$1.90^{+0.16}_{-0.15}$
IC 310	MAGIC/2012	$21.6^{+1.7}_{-1.6}$	$1.776^{+0.067}_{-0.064}$
	MAGIC/2009-2010 - High	$4.97^{+0.53}_{-0.52}$	$1.81^{+0.11}_{-0.11}$
	MAGIC/2009-2010 - Low	$0.67^{+0.12}_{-0.13}$	$1.80^{+0.17}_{-0.16}$
M87	HESS/2005	$1.22^{+0.16}_{-0.16}$	$2.16^{+0.15}_{-0.14}$
	HESS/2004	$0.28^{+0.10}_{-0.10}$	$2.43^{+0.46}_{-0.41}$
	MAGIC/2005-2007	$0.56^{+0.12}_{-0.11}$	$2.15^{+0.21}_{-0.22}$
	MAGIC/2008	$2.88^{+0.42}_{-0.43}$	$2.28^{+0.12}_{-0.11}$
	VERITAS/2007	$0.74^{0.13}_{-0.14}$	$2.27^{+0.19}_{-0.17}$
Markarian 180	MAGIC/2006	$1.51^{+0.83}_{-0.74}$	$2.52^{+0.65}_{-0.88}$
Markarian 421	MAGIC/2006	$33.9^{+2.3}_{-2.4}$	$2.068^{+0.088}_{-0.084}$
	VERITAS/2008 - High C	276^{+17}_{-17}	$2.422^{+0.090}_{-0.085}$

Table 6.4: Same as table 6.3; (*) in $\mathrm{TeV}^{-1}\mathrm{m}^{-2}\mathrm{s}^{-1}$.

Name	Survey/Year - Label	$N_0(*)\times10^{-8}$	Γ
Markarian 501	VERITAS/2009	$8.1^{+1.1}_{-1.2}$	$2.49^{+0.16}_{-0.15}$
NGC 1275	MAGIC/2009-2014	$0.078^{+0.024}_{-0.021}$	$3.45^{+0.17}_{-0.16}$
PKS 0447-439	HESS/2009	$24.5^{+7.9}_{-7.4}$	$1.20^{+0.37}_{-0.33}$
PKS 1441+25	MAGIC/2015	$5.3^{+2.5}_{-1.7}$	$2.80^{+0.16}_{-0.16}$
PKS 1510-089	HESS/2009	$2.2^{+9.3}_{-1.8}$	$2.4^{+1.1}_{-1.2}$
PKS 2005-489	HESS/2004-2007	$1.48^{+0.13}_{-0.14}$	$2.68^{+0.17}_{-0.15}$
PKS 2155-304	HESS/2005-2007	$5.84^{+0.37}_{-0.37}$	$2.831^{+0.056}_{-0.054}$
	MAGIC/2006	$59.9^{+6.3}_{-6.6}$	$2.42^{+0.12}_{-0.11}$
RBS 0413	VERITAS/2009	$3.1^{+1.9}_{-1.4}$	$1.46^{+0.71}_{-0.69}$
RGB J0152+017	HESS/2007	$1.31^{+0.34}_{-0.37}$	$2.36^{+0.38}_{-0.32}$
RGB J0710+591	VERITAS/2008-2009	$3.31^{+0.65}_{-0.67}$	$1.80^{+0.28}_{-0.27}$
RX J0648.7+1516	VERITAS/2010	$1.3^{+1.7}_{-0.8}$	$2.79^{+0.76}_{-0.74}$

Table 6.5: Same as table 6.3; (*) in $\mathrm{TeV}^{-1}\mathrm{m}^{-2}\mathrm{s}^{-1}$.

Chapter 7

A Bayesian approach to signal estimation in gamma-ray astronomy

This chapter discusses a Bayesian approach to signal estimation in On/Off measurements in gamma-ray astronomy. We present some techniques for gamma-hadron separation in gamma-ray data analysis, in section 7.1, and proceed to explain the Frequentist and Bayesian methodologies for signal estimation, in section 7.2, comparing their similarities and differences. Finally, we present a Bayesian method from D'Amico et al. (2021)), originally applied to the MAGIC telescope data, that allows signal estimation without performing selection cuts, so we test this approach using Monte Carlo in the context of H.E.S.S. measurements, in section 7.5. Finally, we apply this methodology to the Crab Nebula observations with H.E.S.S. in section 7.6. The bulk of the work presented in this chapter is the outcome of five months period spent at the Erlangen Centre for Astroparticle Physics.

7.1 Gamma-hadron separation for IACTs

Differently from instruments orbiting in space, ground-based telescopes for gamma rays are subject to strong background contamination due to cosmic rays. While a space telescope such as Fermi-LAT can shield itself from cosmic rays, with its anti-coincidence detector, allowing the rejectiion of the majority of charged particles, the indirect measurement technique from IACTs — collecting Cherenkov light produced by atmospheric showers — encounters an overwhelming dominance of cosmic ray events, of the order of 10^4 more numerous than gamma rays (Catanese & Weekes 1999). Consequently, for the success of IACTs, it is essential to have an analysis procedure that can distinguish and separate gamma rays from background observations.

It all starts with event reconstruction. In a proceedings of 1985, Hillas (1985) presented what it became known as the Hillas parameters: a geometric parametrization of

the shower image in a IACT camera. The main idea behind it is that the electromagnetic air shower, initially produced by a gamma ray, generates a cone of Cherenkov light that results in an approximately elliptical image at the telescope's mirror. The geometric properties of this ellipse, in particular in combination with a simultaneous measurement from other telescope(s), allows the reconstruction of key characteristics of the primary gamma ray. For instance, the energy can be obtained by the size of ellipse — compared to look up tables from Monte Carlo simulations — and the arrival direction can be reconstructed by the stereoscopic imaging of multiple telescopes, identifying the intersection of the major semi-axis of all ellipses. This reconstruction procedure is tuned to gamma rays, as the hadronic background produces much more irregular shapes, due to multiple channels of interaction and the variety of particles that can be created during the development of the shower. Such discrepancies, however, are useful to quantitatively distinguish gamma and hadron events. This can be done directly with Hillas parameters, or using machine learning methods, as it is currently done in H.E.S.S.

H.E.S.S., the High Energy Stereoscopic System (Aharonian et al. 2006c), consists of five IACTs located in Namibia, with the first of the telescopes operating since 2002. Alongside MAGIC and VERITAS, it is part of the current generation of IACTs, resulting in a vast contribution to gamma-ray astronomy. The gamma-hadron separation procedure in H.E.S.S. involves the use of Boosted Decision Trees (BDT) (Breiman et al. 1984) to classify events with a likeliness of originating from a gamma ray or a cosmic ray. A decision tree is a chain of binary decision splits, where at each node a criterion is used to select an event as signal- or background-like, based on input parameters. To determine these criteria and the number of decision steps, the tree is subject to a training process, in which known signal and background samples are used. From them, at each node, the variable and cut parameter which result in the best separation are determined, until there is no improvement or it is not possible to separate further. The process returns "leaves" at which the decision of classifying as signal or background is made. Finally, the training is extended to a "forest" of trees, in a process called "boosting", dealing with the statistical fluctuations from the training sample. As explained in more detail by Ohm et al. (2009), the training is based on Hillas parameters, such as the length, width, size and other information such as the depth of the shower maximum and the spread of the energy reconstruction. The samples are generated from Monte Carlo simulations and the training is performed in a set of energy and redshift bands. The weighted votes of all trees is then summarised in the BDT output as a number between -1 and 1.

From the distribution of the BDT output for gamma and background events, selection cuts can be made in any observed data to remove a substantial amount of background and improve the precision of the signal estimation — a topic we are going to further explore in section 7.5. However, as these distributions change with energy and zenith angle, a single cut value would result in different gamma-efficiencies — the proportion of signal

events surviving the cut. Therefore, the event reconstruction process in H.E.S.S. computes the efficiency for the BDT outputs, using it as the separation variable for spectral analyses, which ensures a fixed efficiency cut. This number, always in the range between 0 and 1, is what we are going to call ζ-BDT and it was first presented and implemented by Ohm et al. (2009). Currently, the ζ-BDT classification is an important piece of H.E.S.S. data analyses, but other methods such as deep learning are also being used to perform gamma-hadron separation directly from the shower images (e.g. Shilon et al. 2019).

7.2 Signal Estimation in On/Off Analysis

The separation procedure is never perfect, so a statistical method is required to estimate the excess signal of a given observation and also attribute uncertainties to these results. In the discussion that follows, we work with what is known as On/Off analysis. This method is mostly used for point-like sources, such that the data is only binned in energy (instead of also including the spatial degrees of freedom as is the case of an extended source) and it is possible to clearly define an "On" region for the observation, in which, presumably, only the target source contributes to the signal emission. Naturally, the On region also contains background events, from an unknown background emission rate. Then, the background can be estimated from observing a control region (Off) in which presumably the same background levels of the On region are found and no other signal sources.

Since the On and Off regions are independent and the signal and background counts follow Poisson distributions, the likelihood of observing $N_{\rm on}$ and $N_{\rm Off}$ in these respective regions is

$$L(N_{\rm on}, N_{\rm off}|s, b; \alpha) = \frac{(s+\alpha b)^{N_{\rm on}}}{N_{\rm on}!} e^{-(s+\alpha b)} \frac{b^{N_{\rm off}}}{N_{\rm off}!} e^{-b}, \tag{7.1}$$

where s and b are the corresponding expected signal and background counts, while α is a normalisation constant between both regions. In general, this parameter can be written as a ratio between the detector acceptances A — the probability of registering a gamma-ray-like event reconstructed at a given position and energy (Berge et al. 2007) —, as

$$\alpha = \frac{\int_{\rm on} A_{\rm on}(\psi_x, \psi_y, \phi_z, E, t)\, \mathrm{d}\psi_x\, \mathrm{d}\psi_y\, \mathrm{d}\phi_z\, \mathrm{d}E\, \mathrm{d}t}{\int_{\rm off} A_{\rm off}(\psi_x, \psi_y, \phi_z, E, t)\, \mathrm{d}\psi_x\, \mathrm{d}\psi_y\, \mathrm{d}\phi_z\, \mathrm{d}E\, \mathrm{d}t} \tag{7.2}$$

where the integration in the On and Off regions are over the positions in the field of view (ψ_x, ψ_y), the zenith angle ϕ_z, event energy E and exposure time t. Usually these regions are defined to have the same acceptances, so α becomes a ratio between the different exposures (area $\times$ time).

The likelihood in eq. 7.1 is the basis of any inference analysis, but one can take

different routes. Here, we discuss two distinct approaches: the frequentist (or "standard") way and the Bayesian method. We compare these methodologies and discuss the implementation of a Bayesian procedure to estimate the signal excess that is going to avoid selection cuts in the data, called BASiL (D'Amico et al. 2021). The following presentation is based on D'Amico et al. (2021) and D'Amico (2022).

7.2.1 Standard Maximum Likelihood Estimation

In the frequentist approach, our goal is to obtain an estimate of the true signal value and to quantify the reliability of the procedure if the experiment were to be repeated infinitely many times. Therefore, we need to define an estimator, compute the statistic value from observed data and compare it to its frequency distribution given by the infinite repetition. The maximum likelihood method applied to eq. 7.1 can provide an estimation of the expected signal and background counts, as

$$\hat{s} = N_{\rm on} - \alpha N_{\rm off} \tag{7.3}$$

$$\hat{b} = N_{\rm off}. \tag{7.4}$$

In our case, the background is simply a nuisance parameter, so when comparing alternative hypotheses, we want to attribute an uncertainty to the signal. To do this, we can construct a likelihood ratio with the profile likelihood. The likelihood ratio is the statistic with maximum power (D'Amico 2022), being the one that minimises the type II error — the probability of not rejecting a hypothesis that is false —, according to the Neyman–Pearson lemma (Neyman & Pearson 1933). Explicitly, it is given by

$$\frac{L(N_{\rm on}, N_{\rm off}|s, \tilde{b}; \alpha)}{L(N_{\rm on}, N_{\rm off}|\hat{s}, \hat{b}; \alpha)}, \tag{7.5}$$

where $L(N_{\rm on}, N_{\rm off}|s, \tilde{b}; \alpha)$ is the profile likelihood, in which the background is set to the value that maximises the likelihood conditional to s. In particular, the first-order maximisation condition for b implies

$$\frac{\partial L}{\partial b} = \frac{(s+\alpha b)^{N_{\rm on}}}{N_{\rm on}!} e^{-(s+\alpha b)} \frac{b^{N_{\rm off}}}{N_{\rm off}!} e^{-b} \left[\frac{\alpha N_{\rm on}}{s+\alpha b} + \frac{N_{\rm off}}{b} - (1+\alpha) \right] = 0, \tag{7.6}$$

so we can solve the equation above to find

$$\tilde{b} = \frac{N + \sqrt{N^2 + 4(1+1/\alpha) s N_{\rm off}}}{2(1+\alpha)}, \tag{7.7}$$

where $N \equiv N_{\rm On} + N_{\rm off} - (1+1/\alpha)s$.

From the likelihood ratio, we can further construct the statistic

$$\lambda(s) = -2\ln\left(\frac{L(N_{\rm on}, N_{\rm off}|s, \tilde{b}; \alpha)}{L(N_{\rm on}, N_{\rm off}|\hat{s}, \hat{b}; \alpha)}\right), \tag{7.8}$$

which, in the large data sample limit, has the asymptotic property of following a chi-squared distribution with one degree of freedom, according to the Wilk's theorem (Wilks 1938), i.e.

$$\lambda \sim \chi^2(1). \tag{7.9}$$

Using this property and the χ^2 distribution, we can construct confidence intervals of $(1-\delta)\times 100\%$, meaning that in $(1-\delta)\times 100\%$ of the times this procedure is repeated, the interval would cover the true signal value. For instance, choosing $\lambda(s) = 1$, we would get the $1\sigma \approx 68.27\%$ interval, while for $\lambda(s) = 4$, it corresponds to $2\sigma \approx 95.45\%$. For a large number of counts, the uncertainty in the expected signal can be readily derived from the expression

$$\sigma = \sqrt{N_{\rm on} + \alpha^2 N_{\rm off}}. \tag{7.10}$$

Also, we can compute the significance of the source detection by setting $s = 0$ and defining

$$\mathrm{TS} \equiv \lambda(0) \tag{7.11}$$

as the test-statistic. Due to its asymptotic behaviour,

$$S \equiv \sqrt{\mathrm{TS}} \tag{7.12}$$

is described by a Gaussian distribution and can be readily interpreted as "the number of sigma", or significance, of the measurement. Such approach was first implemented in the context of gamma-ray astronomy by Li & Ma (1983) and it became the basis of subsequent analyses in the field.

7.2.2 Bayesian Inference of Signal Excess

In the Bayesian perspective, the probability is seen as the degree of knowledge regarding some parameters, given the information we may have. Therefore, the goal is to construct the probability distribution of the signal (called the posterior distribution), from which we can compute statistics and *credible* intervals to summarise its aspects — corresponding to the probability of finding the parameter in a given range. In other words, differently from the frequentist approach, the "true" signal is not simply an unknown fixed

value, but a random variable to which we can attribute a probability distribution[1].

From basic probability rules, the Bayes' theorem informs us that the posterior distribution of the signal s is given by

$$p(s|N_{\rm on}, N_{\rm off}; \alpha) = \frac{p(N_{\rm on}, N_{\rm off}|s; \alpha)p(s)}{p(N_{\rm on}, N_{\rm off}; \alpha)}. \tag{7.13}$$

To explicit show the background dependency, we can insert b by "integrating in", as

$$p(N_{\rm on}, N_{\rm off}|s; \alpha) = \int {\rm d}b\, p(N_{\rm on}, N_{\rm off}|s, b; \alpha)p(b), \tag{7.14}$$

where $p(N_{\rm on}, N_{\rm off}|s, b; \alpha)$ is simply the likelihood of eq. 7.1, so we are going to replace it by the L notation, while $p(b)$ is some prior probability. Then, the posterior becomes

$$p(s|N_{\rm on}, N_{\rm off}; \alpha) = \frac{\int {\rm d}b\, L(N_{\rm on}, N_{\rm off}|s, b; \alpha)p(b)p(s)}{p(N_{\rm on}, N_{\rm off}; \alpha)}. \tag{7.15}$$

Following D'Amico et al. (2021), we can further simplify this expression by assuming improper priors $p(b), p(s) =$ constant for $s, b > 0$ and $p(b) = p(s) = 0$ otherwise. Since the evidence term $p(N_{\rm on}, N_{\rm off}; \alpha)$ is also constant over the parameters, we can write

$$p(s|N_{\rm On}, N_{\rm Off}; \alpha) \propto \int_0^\infty {\rm d}b\, L(N_{\rm On}, N_{\rm Off}|s, b; \alpha), \tag{7.16}$$

meaning the posterior distribution is obtained by marginalising the likelihood over the background. This integration can be performed by using the binomial identity

$$(s + \alpha b)^{N_{\rm on}} = \sum_{N_{\rm s}=0}^{N_{\rm on}} \frac{N_{\rm on}!}{(N_{\rm on} - N_{\rm s})!N_{\rm s}!} s^{N_{\rm s}} (\alpha b)^{N_{\rm on} - N_{\rm s}}, \tag{7.17}$$

so the likelihood becomes

$$p(N_{\rm on}, N_{\rm off}|s, b; \alpha) \propto \sum_{N_{\rm s}=0}^{N_{\rm on}} \frac{(N_{\rm on} + N_{\rm off} - N_{\rm s})!}{(1 + 1/\alpha)^{-N_{\rm s}}(N_{\rm on} - N_{\rm s})!} \frac{s^{N_{\rm s}}}{N_{\rm s}!} e^{-s} \frac{(b(1+\alpha))^{N_{\rm on}+N_{\rm off}-N_{\rm s}}}{(N_{\rm on} + N_{\rm off} - N_{\rm s})!} e^{-b(1+\alpha)}, \tag{7.18}$$

after grouping the terms and neglecting constant ones. Then, the integration over b is of the form

$$\int_0^\infty {\rm d}x\, x^n e^{-\alpha x} = \frac{n!}{\alpha^{n+1}}, \tag{7.19}$$

[1]This is why for frequentist *confidence* intervals, it is the probability of coverage that is relevant, while Bayesian confidence intervals are interpreted directly as the probability of the parameter being found in a given interval.

which cancels some factors and results in

$$p(s|N_{\rm on}, N_{\rm off}; \alpha) \propto \sum_{N_{\rm s}=0}^{N_{\rm on}} \frac{(N_{\rm on} + N_{\rm off} - N_{\rm s})!}{(N_{\rm on} - N_{\rm s})!(1+1/\alpha)^{-N_{\rm s}}} \frac{s^{N_{\rm s}}}{N_{\rm s}!} e^{-s}. \tag{7.20}$$

From this expression, it is possible to estimate the signal by computing the mode of the distribution and corresponding credible intervals. For a 68% interval (1σ), for instance, we need to find the limits $[s_{\rm inf}, s_{\rm sup}]$ such that

$$\int_{s_{\rm inf}}^{s_{\rm sup}} {\rm d}s \, p(s|N_{\rm on}, N_{\rm off}; \alpha) = 0.68, \tag{7.21}$$

when the distribution is normalised, with the condition that

$$p(s_{\rm inf}|N_{\rm on}, N_{\rm off}; \alpha) = p(s_{\rm sup}|N_{\rm on}, N_{\rm off}; \alpha). \tag{7.22}$$

This choice is somewhat arbitrary, but we do it for consistency with D'Amico et al. (2021). In the Bayesian context, the credible intervals have the direct interpretation of informing the region in which the parameter falls with a given probability.

7.3 Frequentist and Bayesian Comparison

We can compare the Frequentist and Bayesian approaches in two distinct situations, in which their similarities and differences will become apparent. First, for a case of high excess counts coming from On and Off regions with large statistic, we expect an agreement between both methods, as the asymptotic condition for the likelihood ratio is satisfied and the Bayesian priors are not relevant ($s \gg 0$). This is clear from Figure 7.1, where we present a comparison between the posterior distribution of s (blue curve) and the $\lambda(s)$ profile (orange curve) for a situation with $N_{\rm on} = 500$, $N_{\rm off} = 800$ and $\alpha = 0.5$. In this scenario, not only the maximum likelihood estimate of $\hat{s} = 100$ (central vertical dotted line) coincides with the mode of the posterior, but also the 68.27% and 95.45% Bayesian credible intervals (shaded areas) match the respective confidence intervals (width of the orange curve when setting $\lambda(s) = 1$ and $\lambda(s) = 4$). We can also see that the asymptotic expression for the uncertainty σ of the signal (eq. 7.10), represented as the other vertical dotted lines for $\pm 1\sigma$ and $\pm 2\sigma$, essentially coincide with the corresponding probability intervals — as expected in this high counts case. Finally, the black points describe the discrete probability distribution of observing $N_{\rm s}$ signal counts in the On region, which can be derived from eq. 7.20 (see D'Amico et al. 2021).

With a low number of counts, some statistical properties of the Frequentist method may not be valid and the approach requires some ad hoc adjustments — as mentioned by D'Amico et al. (2021) — to provide physically meaningful results. Figure 7.2 shows

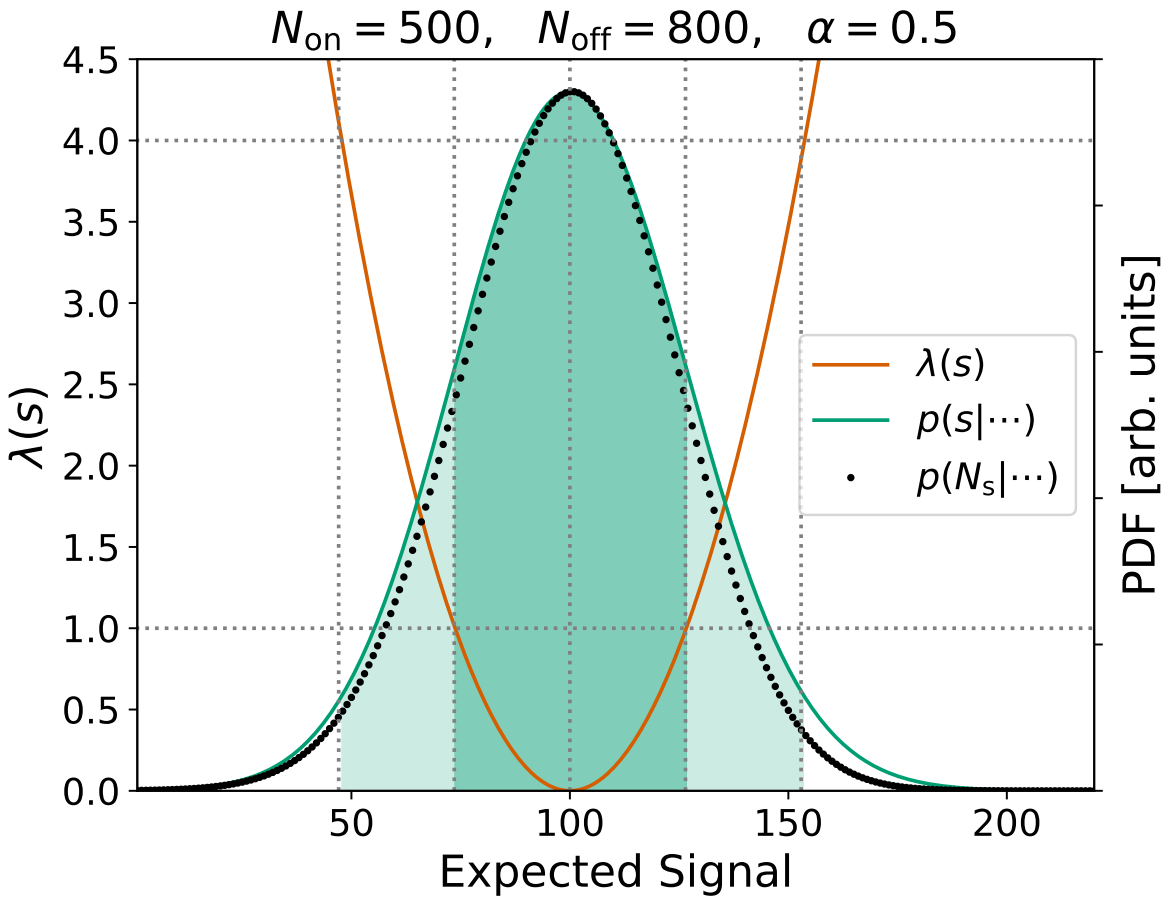

Figure 7.1: Posterior probability of the expected signal counts, s, in arbitrary units (bluish-green curve), compared to the discrete probability of observing N_s counts in the On region (black dots) and the frequentist statistic $\lambda(s)$, orange curve, as defined in eq. 7.8. The shaded area marks the 68.27% (darker) and 95.45% (lighter) credible intervals from the posterior distribution. The vertical dotted lines define the maximum likelihood estimate $\hat{s} = 100$ (central line) and the $\pm 1\sigma$ and $\pm 2\sigma$ values according to eq. 7.10.

a situation with low statistics consistent with no signal in the On region. It reveals discrepant results of the two approaches, as the maximum likelihood estimate provides a negative signal rate. Furthermore, while the confidence intervals cover negative values, the credible interval gives a more straightforward answer to upper limits in the expected signal, as the prior cuts the posterior distribution for values below zero, automatically putting the mode in $s = 0$. It is also interesting to notice that the asymptotic expression for σ, in particular for the $\pm 2\sigma$ range (vertical lines), does not coincide anymore with the 95.45% confidence interval constructed from $\lambda(s) = 4$.

7.4 The BASiL Method

Due to high background levels, the improvement of precision in the Standard method requires selection cuts in the data. In the case of H.E.S.S., we saw that ζ-BDT

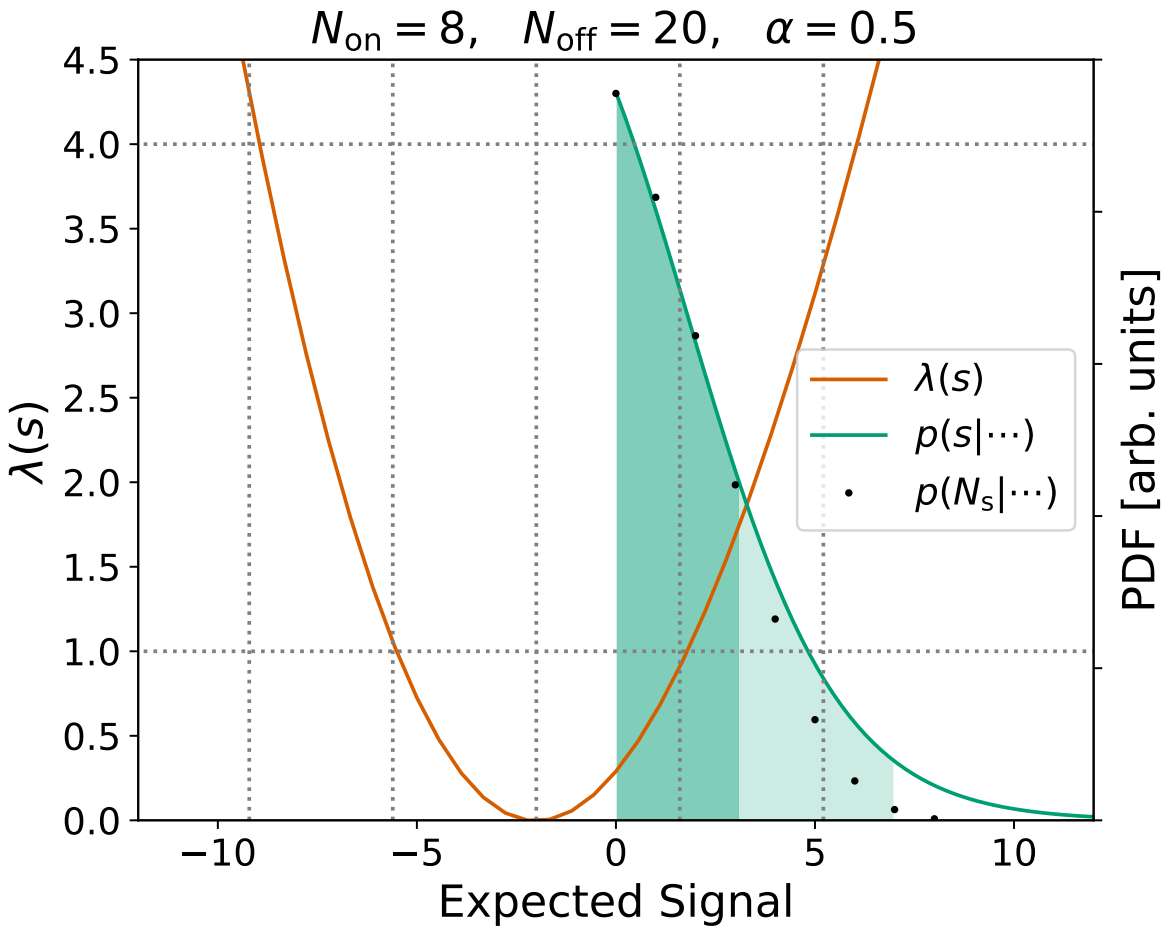

Figure 7.2: Same as Figure 7.1, but for low counts in the On/Off regions ($\hat{s} = -2$).

is an important variable for gamma-hadron separation. For strong and steep spectrum sources, for instance, the elimination of events with ζ-BDT > 0.84 maximises the Li & Ma (1983) significance of detection[2] (Ohm et al. 2009). Other cuts can also be optimised for different circumstances. These selection cuts imply that part of the signal events can also be lost. This can particularly impact faint sources or VHE events, in which already few signal counts are expected. As seen in Section 7.3, the Standard method can require ad hoc adjustments to produce physical results in such low counts limit. In the Bayesian approach, however, it is possible to include information regarding gamma-hadron separation variables and estimate the signal with better precision, *without* performing selection cuts in the data. This method was introduced by D'Amico et al. (2021), known as Bayesian Analysis including Single-event Likelihoods, or BASiL.

BASiL works by introducing in the posterior distribution, through the likelihood, a matrix or list of event variables, $\mathbf{X}$. This matrix corresponds to the relevant variables for

[2]Alongside a cut of events with squared angular distance $\theta^2 > 0.0125\,\mathrm{deg}^2$, the separation between the source and the reconstructed event direction.

gamma-hadron separation for all On events, as $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{N_{\rm on}}\}$, where, for instance,

$$\mathbf{x}_i = (E, \theta^2, \zeta\text{-BDT})_i \tag{7.23}$$

for an event i with energy E and squared angular distance θ^2 in relation to the source position. Analogously to the Bayesian derivation of section 7.2.2, we need to compute the posterior distribution

$$p(s|N_{\rm On}, N_{\rm Off}, \mathbf{X}; \alpha) \propto \int_0^\infty {\rm d}b\, p(N_{\rm On}, N_{\rm Off}, \mathbf{X}|s, b; \alpha), \tag{7.24}$$

with the inclusion of the single-event variables. From the product probability rule,

$$p(N_{\rm on}, N_{\rm off}, \mathbf{X}|s, b; \alpha) = p(\mathbf{X}|N_{\rm on}, s, b; \alpha) L(N_{\rm on}, N_{\rm off}|s, b; \alpha), \tag{7.25}$$

where the second term on the right-hand side is simply eq. 7.1 and

$$p(\mathbf{X}|N_{\rm on}, s, b; \alpha) = \prod_{i=1}^{N_{\rm on}} p(\mathbf{x}_i|N_{\rm on}, s, b; \alpha), \tag{7.26}$$

as the events are independent. Furthermore, since the event can only originate from a gamma ray (γ) or the cosmic-ray background ($\bar{\gamma}$), the probability $p(\mathbf{x}_i|N_{\rm on}, s, b; \alpha)$ can be decomposed in the disjoint cases

$$p(\mathbf{X}|N_{\rm on}, s, b; \alpha) = \prod_{i=1}^{N_{\rm on}} \left[p(\mathbf{x}_i|\gamma) p(\gamma|s, b; \alpha) + p(\mathbf{x}_i|\bar{\gamma}) p(\bar{\gamma}|s, b; \alpha) \right], \tag{7.27}$$

with $p(\mathbf{x}_i|\gamma)$ being the probability of observing $\mathbf{x}_i$ values given the event is a gamma ray, while $p(\mathbf{x}_i|\bar{\gamma})$ respectively for the background. D'Amico et al. (2021) uses the prior probabilities

$$p(\gamma|s, b; \alpha) = \frac{s}{s + \alpha b} = 1 - p(\bar{\gamma}|s, b; \alpha), \tag{7.28}$$

such that the probability for $\mathbf{X}$ becomes

$$p(\mathbf{X}|N_{\rm on}, s, b; \alpha) = \frac{1}{(s + \alpha b)^{N_{on}}} \prod_{i=1}^{N_{on}} \left[p(\mathbf{x}_i|\gamma) s + p(\mathbf{x}_i|\bar{\gamma}) \alpha b \right], \tag{7.29}$$

formally resulting in

$$p(\mathbf{X}|N_{\rm on}, s, \alpha b) = \sum_{N_{\rm s}=0}^{N_{\rm on}} C(\mathbf{X}, N_{\rm s}) \frac{s^{N_{\rm s}} (\alpha b)^{N_{\rm on} - N_{\rm s}}}{(s + \alpha b)^{N_{\rm on}}}, \tag{7.30}$$

where the combinatorial factor $C(\mathbf{X}, N_\mathrm{s})$ is the probability of observing $\mathbf{X}$ given N_s signal events in the On region. It encapsulates the sum of all possible product combinations of $p(\mathbf{x}_i|\gamma)$ and $p(\mathbf{x}_i|\bar{\gamma})$, considering N_s signal events. For instance, for only 3 events, we would have

$$\begin{aligned}
C(\mathbf{X},0) &= p(\mathbf{x}_1|\bar{\gamma})p(\mathbf{x}_2|\bar{\gamma})p(\mathbf{x}_3|\bar{\gamma}) \\
C(\mathbf{X},1) &= p(\mathbf{x}_1|\gamma)p(\mathbf{x}_2|\bar{\gamma})p(\mathbf{x}_3|\bar{\gamma}) + p(\mathbf{x}_1|\bar{\gamma})p(\mathbf{x}_2|\gamma)p(\mathbf{x}_3|\bar{\gamma}) + p(\mathbf{x}_1|\bar{\gamma})p(\mathbf{x}_2|\bar{\gamma})p(\mathbf{x}_3|\gamma) \\
C(\mathbf{X},2) &= p(\mathbf{x}_1|\gamma)p(\mathbf{x}_2|\gamma)p(\mathbf{x}_3|\bar{\gamma}) + p(\mathbf{x}_1|\gamma)p(\mathbf{x}_2|\bar{\gamma})p(\mathbf{x}_3|\gamma) + p(\mathbf{x}_1|\bar{\gamma})p(\mathbf{x}_2|\gamma)p(\mathbf{x}_3|\gamma) \\
C(\mathbf{X},3) &= p(\mathbf{x}_i|\gamma)p(\mathbf{x}_i|\gamma)p(\mathbf{x}_i|\gamma).
\end{aligned} \tag{7.31}$$

Inserting eq. 7.30 in eq. 7.24 to perform the background integration, a similar procedure to Section 7.2.2 can be done, resulting in

$$p(s|N_\mathrm{on}, N_\mathrm{off}, \mathbf{X}; \alpha) \propto \sum_{N_\mathrm{s}=0}^{N_\mathrm{on}} \frac{(N_\mathrm{on}+N_\mathrm{off}-N_\mathrm{s})!}{(N_\mathrm{on}-N_\mathrm{s})!(1+1/\alpha)^{-N_\mathrm{s}}} \frac{C(\mathbf{X}, N_\mathrm{s})}{\binom{N_\mathrm{on}}{N_\mathrm{s}}} \frac{s^{N_\mathrm{s}}}{N_\mathrm{s}!} e^{-s}, \tag{7.32}$$

which is the BASiL posterior distribution. In the next section we investigate how this method can be implemented in practice, testing it with Monte Carlo generated data.

7.5 Monte Carlo results

Firstly, we create a random set of On and Off events, given some expected signal and background rates. We start by fixing the expected signal counts to $s = 100$ and the background to $b = 3000$, with $\alpha = 1/3$. Then, to obtain the measured counts in the On and Off regions, we can sample from Poisson distributions with averages s, αb and b. This produces the number of signal counts, N_s, and background counts, N_bkg, in the On region ($N_\mathrm{on} = N_\mathrm{s} + N_\mathrm{bkg}$) and the number of counts in the Off region, N_off, respectively. Secondly, we attribute to these events two variables: the squared angular distance, θ^2, and ζ-BDT. Depending if it is originally a signal or background event, their values will follow different distributions, which depend on the instrument characteristics and observation conditions. Therefore, we assume a point source observed by H.E.S.S. with a zenith angle $\sim$45° and an offset between the gamma ray and the telescope axis of 0.5° to generate the signal (gamma) events distribution using the H.E.S.S. Analysis Package (HAP)[3]. For the background, we simply use real data from Off region measurements under similar observation conditions. Figure 7.3 presents the expected θ^2 and ζ-BDT distributions from signal and background events, in an arbitrary energy range between 0.5–1 TeV. From these distributions, we can finally sample the event variables for the N_s signal and $N_\mathrm{bkg} + N_\mathrm{off}$ background counts.

The process of generating the On/Off observations was repeated 1000 times and,

[3]Internal tool developed by the H.E.S.S. collaboration.

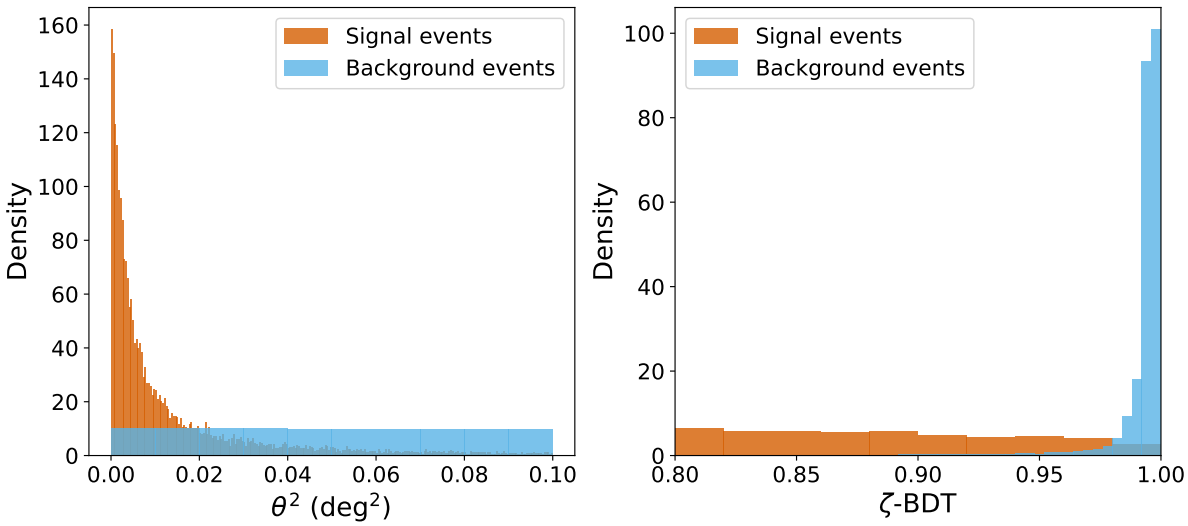

Figure 7.3: Distribution of squared-angular distance (θ^2) and ζ-BDT values for signal (gamma rays) and background events. The distribution of ζ-BDT extends to $0 \leq \zeta\text{-BDT} \leq 1$, but we only show a limited range for visualisation purposes. Signal events were simulated considering a point source observed at zenith angle $\sim$45° and an offset between the incoming gamma ray and the telescope axis of 0.5°, while background events come from real observations of Off regions.

for each realisation, the Standard and BASiL approaches were applied to estimate the expected signal. This provides a distribution of $\hat{s}$ for both methods. To compare them, we compute the relative error

$$\eta \equiv \frac{\hat{s} - s \cdot \epsilon}{s \cdot \epsilon}, \tag{7.33}$$

where $s = 100$ is the true value and ϵ is the efficiency from selection cuts, i.e., the fraction of signal events that survives a given cut. For the Standard method, we apply cuts on θ^2 and ζ-BDT, so it is expected that a fraction $1 - \epsilon$ of the signal events are lost, so that one needs to take the efficiency into account to avoid biasing the estimation of the true signal rate. For BASiL, $\epsilon = 1$ always, as no cuts are applied. Following D'Amico et al. (2021), we call the standard deviation and the average of η as the "precision" and "bias", respectively, i.e. we define

$$\text{Precision} \equiv \sigma(\eta) \tag{7.34}$$

$$\text{Bias} \equiv \langle \eta \rangle. \tag{7.35}$$

We first analyse θ^2 and ζ-BDT separately. That is, we only apply selection cuts on one of these variables (for the Standard method), or incorporate information from only one single event variable (for BASiL). Figure 7.4, then, presents the precision and bias for both methods as a function of ϵ of the Standard approach, considering ζ-BDT on the left and θ^2 on the right. When applying the Standard method, there is an optimum cut that minimises the dispersion of the inferred signal, which happens around a value of

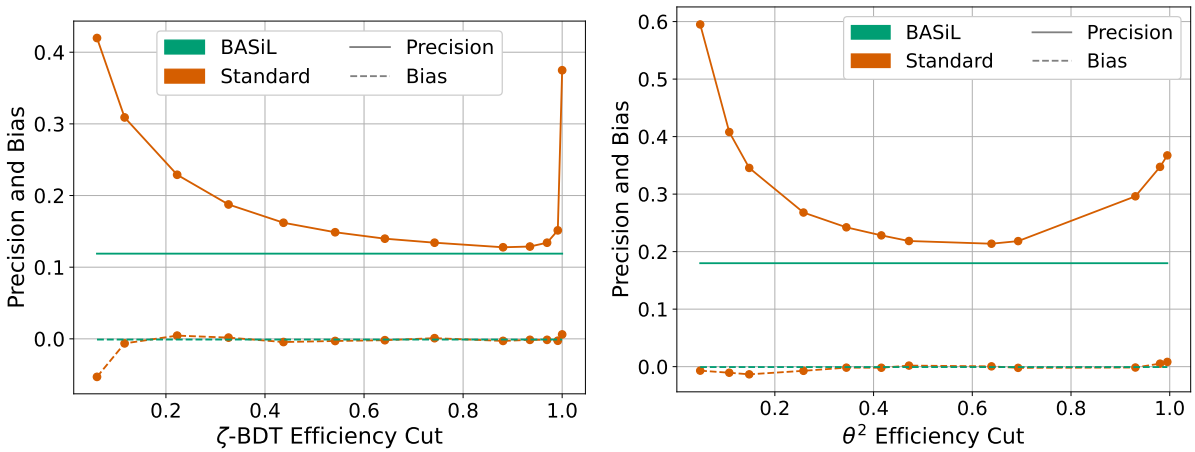

Figure 7.4: Precision and bias comparison of the Standard and BASiL methods, by varying the efficiency cut on ζ-BDT (left) and on θ^2 (right). The true signal was fixed at $s = 100$, with background $b = 300$ and $\alpha = 1/3$.

ζ-BDT $= 0.84$ ($\epsilon \approx 0.88$) and $\theta^2 = 0.1\,\mathrm{deg}^2$ ($\epsilon \approx 0.64$). Under this condition, the bias is negligible, which is a property that is maintained for various efficiency cuts, unless the cut is very extreme ($\epsilon \approx 0$) or no selection cuts are applied ($\epsilon \approx 1$). As expected, the precision degrades drastically in both extremes. On the other hand, the BASiL method is able to achieve a comparable bias while having improved uncertainty (lower precision value) and retaining all the signal events data. Noticeably, this improvement is better for the θ^2 variable. Since ζ-BDT has probably a higher gamma-hadron separation power, its choice for selection variable already results in a good precision for the Standard method (as efficiency cuts around 80% essentially remove all background), so less additional information is gained by incorporating the probability distributions of this parameter. Nevertheless, there is a clear improvement with respect to the Standard method, which can be further extended by combining the information of both selection variables in the analysis.

We can compare the two approaches in different situations of background dominance. While in the previous example, the "signal-to-noise" ratio, as defined by SNR $= s/(\alpha b)$, was at 10%, we now explore the case of various SNR values, fixing $b = 300$ and varying s from 1 to 100. In the case of the Standard approach, we have kept the cut values on ζ-BDT $= 0.84$ and $\theta^2 = 0.0125\,\mathrm{deg}^2$, respectively, which correspond to the standard choice analysis in H.E.S.S. Figures 7.5 and 7.6 show the results considering the ζ-BDT and θ^2 variables. On the right of each figure, we also present the relative improvement in the precision, defined as the the difference of precision between the Standard and BASiL methods divided by the Standard's value. In both cases we can see that the dispersion on the estimated signal in the BASiL approach is always smaller than in the Standard method, but the improvement becomes less prominent as we increase the SNR. This reinforces the idea that the Bayesian method would be particularly useful for dealing with cases with a low number of photon counts, such as for faint sources or VHE measurements.

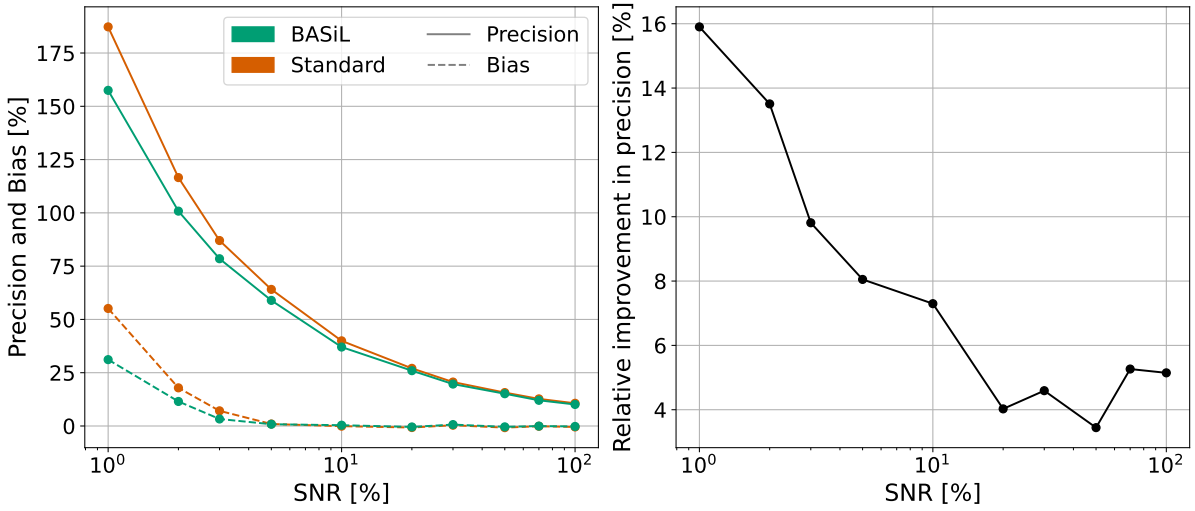

Figure 7.5: Precision and bias comparison of the Standard and BASiL methods, by varying the signal-to-noise ratio (fixing $\alpha b = 100$), and considering only the ζ-BDT variable (for selection cuts in case of the Standard approach, or its probability distribution for BASiL). On the right, the difference between the Standard and BASiL precision divided by the Standard one.

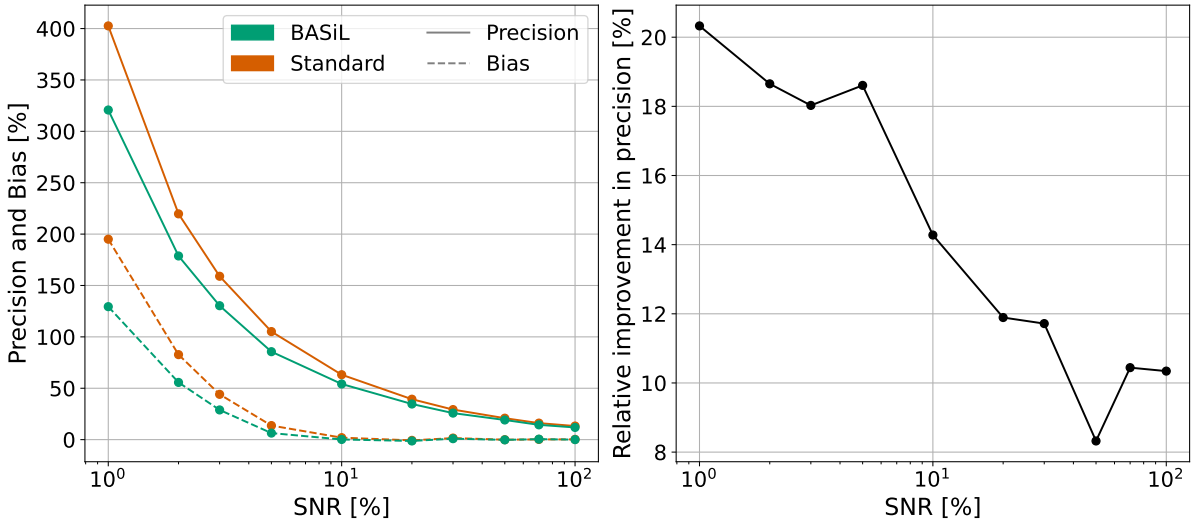

Figure 7.6: Precision and bias comparison of the Standard and BASiL methods, by varying the signal-to-noise ratio (fixing $\alpha b = 100$), and considering only the θ^2 variable (for selection cuts in case of the Standard approach, or its probability distribution for BASiL). On the right, the difference between the Standard and BASiL precision divided by the Standard one.

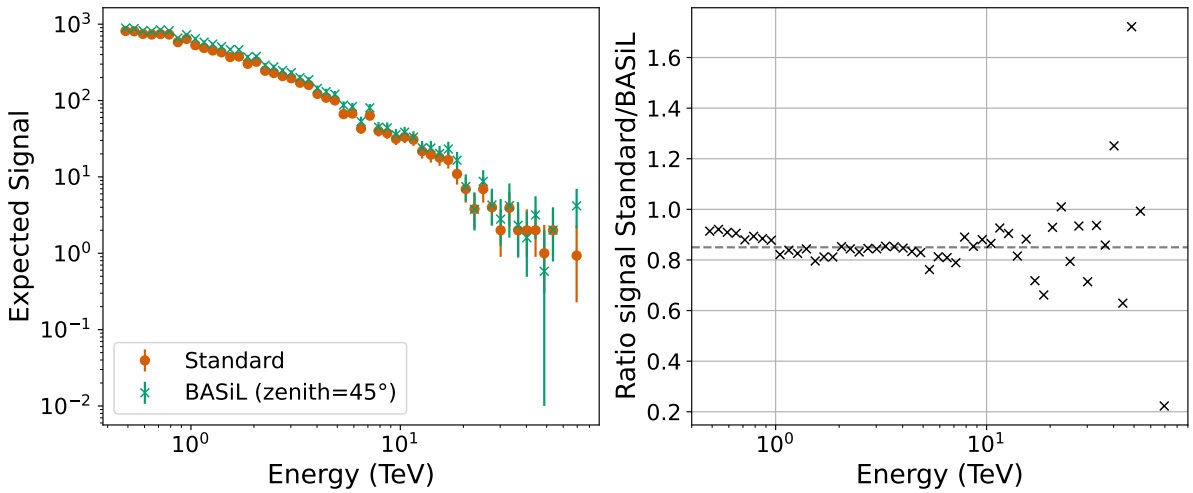

Figure 7.7: *Left*: Expected signal counts for the H.E.S.S. Crab data when analysing with the BASiL and Standard methods, adopting ζ-BDT as the gamma-hadron separation variable. A selection of ζ-BDT < 0.84 events was performed in the frequentist case and the points are not corrected due to the efficiency in the cut. *Right*: the ratio between the estimated signal in the Standard and BASiL approaches. The dashed line corresponds to the mean efficiency of $\epsilon = 0.85$.

The increased bias for low SNR comes from the fact that the distribution of the estimated signal is not symmetric, as negative signal values are not allowed (for the Standard case, if $\hat{s} < 0$, we have considered $\hat{s} = 0$ for computing the bias and precision).

7.6 Application to the Crab Nebula

The Crab Nebula is a supernova remnant in our Galaxy extensively studied in gamma-ray astronomy (Aharonian et al. 2006c; Aleksić et al. 2015a), being an ideal test case for any statistical analysis on the subject. A set of 57 H.E.S.S. observations, totalling around 27 h of live time, was selected. The source was observed with an offset of 0.5°, while a cut in $\theta^2 = 0.0125\,\mathrm{deg}^2$ relative to Crab's position defines the On region adopted. For the analysis presented here, we focus on ζ-BDT as the gamma-hadron separation variable and implement the Standard and BASiL analysis for comparison.

We estimate the expected signal s on 51 energy bins in observed energy. For the Standard analysis we select the events for which ζ-BDT < 0.84, while for BASiL we consider the ζ-BDT gamma and background distributions restricted in the respective energy bins (and the a zenith angle of 45°) to construct the posterior distribution. The uncertainties are computed according to the likelihood profile $\lambda(s) = 1$ and the respective credible interval for the Bayesian method. Figure 7.7, then, presents the expected signal from both methodologies, but without applying the correction due to the efficiency cut in the Standard one. On the right of the same figure, we show the ratio between the two estimates, which fluctuates around the mean efficiency of ∼0.85, represented by the

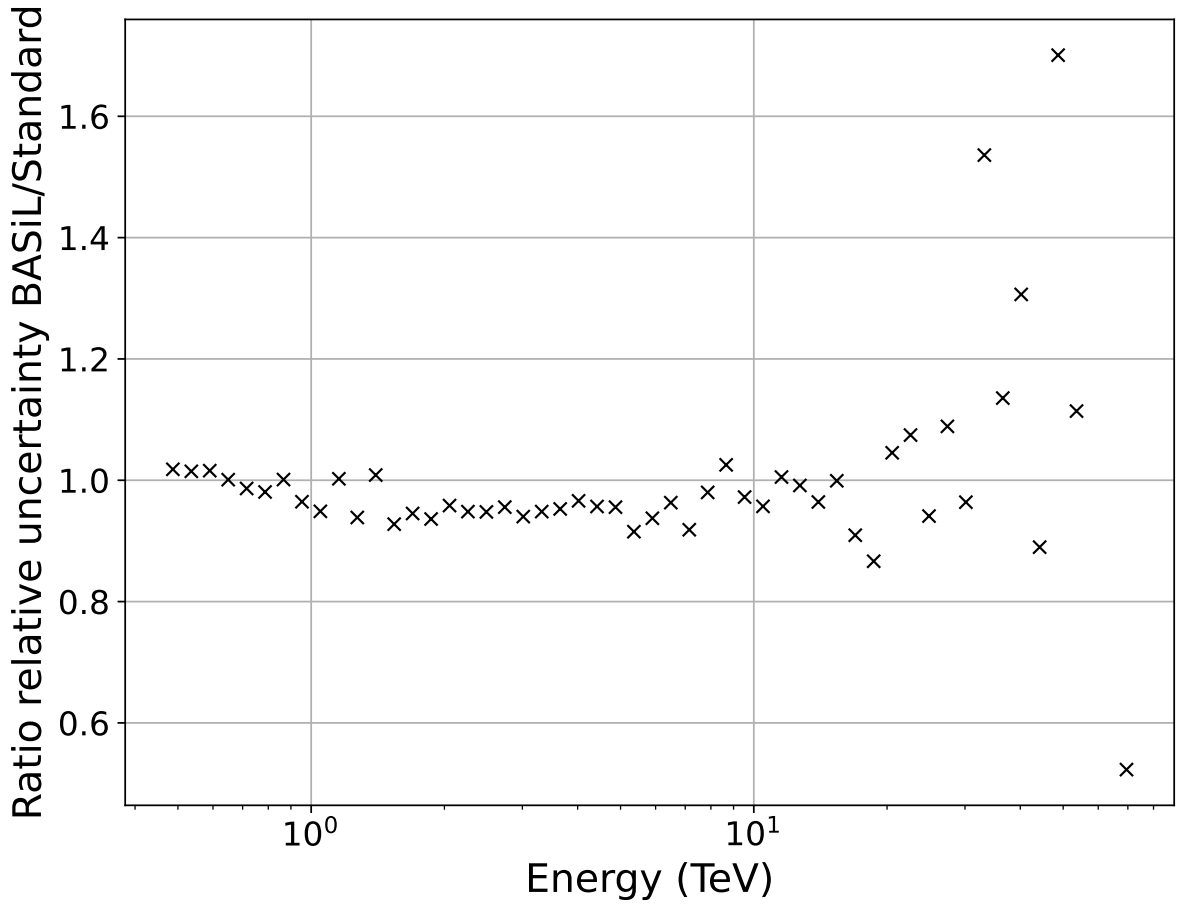

Figure 7.8: Ratio between relative uncertainties (uncertainty over estimated signal) of BASiL and Standard methods for the Crab analysis.

horizontal dashed line. Finally, Figure 7.8 reveals the general improvement in uncertainty of the BASiL method, by comparing the ratio of the relative uncertainties (width between upper and lower limits divided by the estimated signal) in both approaches. We find that BASiL results in a ∼5–6% smaller uncertainty in the range between 1–10 TeV, while we recover similar uncertainties for energies lower than 1 TeV. Besides, for the low energy bins, we have high signal excess counts, so it is expected that we identify an equivalence between the methods. For the highest energies, above 10 TeV, however, there is a lot of fluctuation introduced due to the low signal counts. In fact, we expect an increase in bias and dispersion for low SNR, according to the Monte Carlo analysis, so this behaviour should make sense.

7.7 Discussion

The BASiL method is a new way of estimating the signal in On/Off observations of gamma-ray sources. Its primary advantage is to avoid selection cuts on the data, improving the resolution of the estimation in an unbiased way while keeping more signal

events. The main challenge comes from correctly describing the probability distribution of gamma-hadron separation variables and dealing with the increased computational effort of calculating the posterior probability of the signal, comparatively to the Standard methodology based on maximum likelihood. The results here presented do not incorporate the instrument response, so further extensions could apply unfolding or forward folding methods (D'Amico 2022) to be able to extract the true spectral flux from the event observations. Then, this analysis can be also combined with EBL studies, as the improved signal estimation (smaller uncertainties and no selection cuts) may reflect in better constraints on the opacity of the extragalactic medium.

Chapter 8

Conclusion

The background radiation content of the Universe emerges from a complex variety of physical processes and a large range of scales, being an important tracer of how things formed and evolved throughout the history of the Cosmos. One of its components, the Extragalactic Background Light, is the direct and indirect product of star formation, comprising essentially by stellar emission that escapes the interior of galaxies and radiation reprocessed by dust. Its spectral shape today is naturally a result of the integrated emission history since the beginning of structure formation, influenced by the cosmological setting of how the Universe expands. However, direct measurements of the EBL have observational challenges, especially due to the dominant foregrounds, such as the zodiacal light, diffuse galactic emission and stars in our Galaxy. This leaves particular gaps in the measured EBL intensity, mostly in the mid-IR range, that require alternative methods to probe it.

An indirect way of constraining the EBL comes from exploring its connection to very energetic gamma rays. Extragalactic VHE photons are produced by AGNs through leptonic or hadronic processes and most of the current detections come from blazars of the BL Lac type. During the propagation of gamma rays, they may interact with EBL photons through electron-positron pair production (the Breit-Wheeler process), resulting in a suppression of the emitted flux. In this sense, by making assumptions on the intrinsic emission, one can infer limits on the opacity of the extragalactic medium (and consequently on the intensity of the EBL). At the same time, to know the EBL spectral energy distribution is necessary to interpret observations of VHE gamma rays from extragalactic sources, as one needs to disentangle intrinsic characteristics from propagation features.

Given this interdependency, we have investigated simultaneous constraints on the EBL and the intrinsic gamma-ray emission of multiple AGNs. This was possible by defining the posterior probability distribution of EBL and spectral parameters of sources, conditional to the adopted models and data, and sampling it through MCMC methods. Once we have obtained a representative sample of the posterior distribution, it is

easy to marginalise over a subset of parameters. For instance, we can integrate over all sources' spectral parameters to obtain a more robust inference on the EBL. Conversely, the marginalisation over EBL parameters allows the incorporation of its uncertainties to the unknown intrinsic emission. Furthermore, MCMC methods are useful for sampling distributions with an arbitrary number of dimensions, so we have explored the incorporation in the likelihood of various spectra from multiple sources, seeking to break degeneracies in the description of the extragalactic opacity by having a larger data sample.

With a sample of highly significant synthetic sources, we were able to show that it is possible to achieve reasonable constraints on the IR range of the EBL — particularly in the mid-IR, where the VHE data is more sensitive —, while also simultaneously recovering the spectral index of power law sources. The most significant sources dominate these constraints, but the addition of more data in the likelihood can reduce the uncertainties, although this improvement becomes less impacting after a certain number of sources, if the inclusion is performed in decreasing order of detection significance. Moreover, in the context of the F10 model, we were capable of constraining the PAH temperature with a precision of dozens of Kelvin. By increasing the observation time of the sources, significant reductions in the uncertainties were observed, reinforcing that future measurements with CTA, given its improved sensitivity, will greatly contribute to better EBL constraints. The increased observation time will also lead to better comparisons between different EBL models. Fixing the D11 model as the true EBL, we have noticed an underestimation of the spectral indices (particularly for harder sources), possibly due to the fact that D11 predicts a much larger far-IR intensity that cannot be accounted by changing the normalised dust fractions in F10 model. Furthermore, we have not identified significant improvements in this comparison by turning the PAH temperature a free parameter. Nevertheless, it was possible to achieve a good description of the mid-IR range of D11 model if the data is not sensitive to longer EBL wavelengths, revealing that the F10 model with free dust fractions has some flexibility to adapt to different EBL opacities. During this analysis, we have presented a way of analytically marginalising the likelihood over the flux normalisation parameters, which can be applicable to a range of different parametrizations of the intrinsic spectrum and it is useful for improving the efficiency of MCMC tools (from the smaller number of dimensions to sample), especially if the focus is on EBL constraints.

The same methodology used to probe the EBL can be further extended to constrain cosmological parameters, such as the Hubble constant. In general, the optical depth depends on H_0 through the redshift path integration and the cosmological dependency on the EBL density. Therefore, we have recomputed the emissivities coming from the stellar and dust components to build an expanded optical depth grid as a function also of H_0, based on the F10 model. Essentially, changes in the optical depth for different expansion rates come from the stellar emission, as different H_0 values affect the relation between

time intervals and redshift, modifying the age of stellar populations at a given redshift and consequently the overall emissivity. Performing the analysis with the synthetic sample, it was possible to obtain very consistent results from the low observation time data. However, even in an ideal scenario without systematic errors, a shift of a few $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ in the inferred H_0 value can appear when increasing the observation time — although we could not make a thorough study of the contribution coming from the numerical precision of the optical depth grid onto the final value of the Hubble constant. Furthermore, if the wrong EBL model is used, a more significant bias in the Hubble constant can arise, revealing that systematic errors in the analysis are the biggest challenge for achieving competitive constraints on H_0. On the other hand, this method is independent of other cosmological analyses and can be further refined if different EBL models are incorporated (to more robustly estimate the systematic uncertainties).

As pointed out by de Matos Pimentel & Moura-Santos (2019), current extragalactic VHE data is primarily attenuated by the stellar component of the EBL. Therefore, it has a limited scope for probing the dust-dominated IR band. However, we have found that it is possible to achieve robust and improved constraints on the mid-IR range (controlled by the PAH fraction) when combining multiple sources. We have selected 65 spectra from 36 different AGN and have simultaneously inferred EBL and intrinsic parameters of all of them. The resulting uncertainty on the PAH fraction is three to five times smaller than what was obtained by de Matos Pimentel & Moura-Santos (2019) when analysing the constraints from Markarian 501 (HEGRA/1997) data. Furthermore, our result is statistically consistent for different assumptions on the intrinsic spectrum of Mkn 501 (PL or LP) and even when we remove its data from the likelihood, revealing that the combination of the 64 other spectra can successfully probe the mid-IR EBL intensity. TOn the other hand, the flare data of Mkn 501 (HEGRA/1997) showed to be essential for constraining the far-IR part of the EBL, as their removal implied near flat marginal distributions for the SG and LG fractions. Consequently, the presence or absence of intrinsic curvature to describe the flux of Mkn 501 significantly impact these constraints, as an intrinsic suppression of the flux accommodates more EBL configurations, increasing the uncertainties. Nevertheless, in our analysis, the PL and LP models — with their respective EBL constraints — produced very similar intrinsic fluxes for Mkn 501, as the curvature parameter is consistent to zero.

Our results revealed an EBL intensity approximately in between the F10 and D11 models and also compatible with Finke et al. (2022) model, up to $\sim$100 µm. We have also presented the CGRH with the corresponding uncertainties in the dust fractions, as well as the spectral index distribution of the sources and a table containing the intrinsic parameter information for all spectra, which can be used by anyone interested in analysing a given source. In general, we obtain a good global fit to all spectra, as no particular outliers are present, and the intrinsic parameters are generally consistent with other results

from the literature. Finally, the analysis with a larger data sample was made possible by developing an HMC code, as it drastically improved the efficiency given the larger dimensionality of the parameter space. Although the tuning of the algorithm can be particular challenging, it can be simply extended when other spectra become available. The analysis can be further refined by including some knowledge on the systematic errors of the data, or testing different parametrizations of the intrinsic fluxes.

This dissertation also included the preliminary steps for applying a Bayesian method for signal estimation in gamma-ray observations. The BASiL methodology presented by D'Amico et al. (2021) allows the classical On/Off analysis to be performed without selection cuts, which has the advantage of keeping more signal events in the data, while improving the precision of the estimate. It also can complement the usual maximum likelihood (frequentist) approach by being better suited for low SNR data. This can be particularly promising for VHE measurements, as we are typically dealing with a low rate of signal events. Here, we have tested the methodology with Monte Carlo data in the context of H.E.S.S. and did an application to Crab Nebula real data. In the range between 1–10 TeV, we have obtained a consistent improvement in the uncertainty of the estimated signal, comparatively to the Standard analysis. Topics that could be explored are the implementation of BASiL according to the gamma-hadron separation analysis of CTA, as well as EBL studies with data analysed with this new method.

Appendix A

Breit-Wheeler QED cross section at tree level

Here we are going to present an overall sketch on how the Breit-Wheeler cross section can be computed from QED using the Feynman diagram formalism. In leading order, the only two diagrams that contribute to the invariant amplitude can be seen in Figure A.1. We are denoting the 4-momentum of the photons and fermions by k and p, respectively, with corresponding indices $1, 2, +, -$ to distinguish between the photons (1 or 2), positron ($+$) and electron ($-$). Applying the Feynman rules (Peskin 2018), we must insert a factor $ig\gamma^\mu$ for each vertex (in which $g^2 = 4\pi\alpha$, α is the fine-structure constant and γ^μ are the gamma, or Dirac, matrices); the factor $i/(\not{q} - mc)$ for the fermion propagator; spinors $\bar{u}^{(s_-)}$ and $v^{(s_+)}$ for the outgoing electron and positron, respectively with spins s_- and s_+; and a polarization 4-vector ϵ_μ for each incoming photon. Then, for the diagram on the *right* and making use of the Einsten summation notation, the amplitude is

$$\begin{aligned} -i\mathcal{M}_R &= \bar{u}^{(s_-)}(ig\gamma^\mu\epsilon_{1\mu})\frac{i}{\not{p}_- - \not{k}_1 - mc}(ig\gamma^\nu\epsilon_{2\nu})v^{(s_+)} \\ &= -ig^2\bar{u}_a^{(s_-)}(\gamma^\mu_{ab}\epsilon_{1\mu})\frac{[\not{p}_- - \not{k}_1 + mc]_{bc}}{(p_- - k_1)^2 - (mc)^2}(\gamma^\nu_{cd}\epsilon_{2\nu})v_d^{(s_+)}, \end{aligned} \tag{A.1}$$

where we have omitted for notation clarity the momentum dependence of each spinor and on the second line we explicitly show the spinor indices (Roman alphabet). This can be further simplified using the special transverse gauge (tilde notation), in which $\tilde{\epsilon}_1 \cdot p_- = \tilde{\epsilon}_2 \cdot p_- = 0$. In this case,

$$\bar{u}^{(s_-)}\tilde{\not{\epsilon}}_1(\not{p}_1 + mc) = \bar{u}^{(s_-)}(-\not{p}_- + mc)\tilde{\not{\epsilon}}_1 = 0, \tag{A.2}$$

using Dirac's equation and the fact that

$$\tilde{\not{\epsilon}}_1\not{p}_- = \tilde{\epsilon}_{1\mu}p_{-,\nu}\gamma^\mu\gamma^\nu = \tilde{\epsilon}_{1\mu}p_{-,\nu}(2g^{\mu\nu} - \gamma^\nu\gamma^\mu) = 2\tilde{\epsilon}_1 \cdot p_- - \not{p}_-\tilde{\not{\epsilon}}_1 = -\not{p}_-\tilde{\not{\epsilon}}_1, \tag{A.3}$$

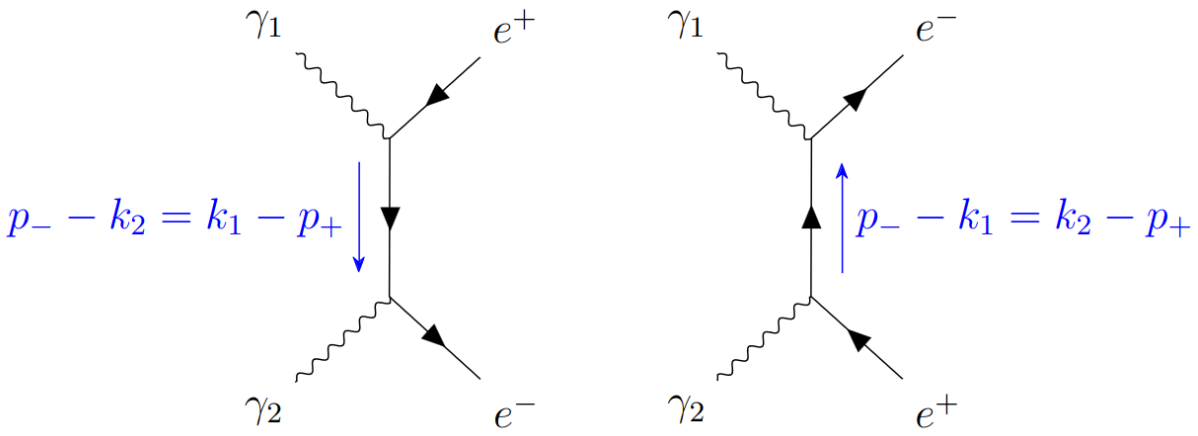

Figure A.1: Feynamn diagrams for the Breit-Wheeler process at leading order.

as the algebra of the gamma matrices is defined by their anticommutation relation $\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}\mathbb{1}$. Therefore,

$$-i\mathcal{M}_R = -ig^2 \bar{u}_a^{(s_-)} (\gamma^\mu_{ab} \tilde{\epsilon}_{1\mu}) \frac{[-\not{k}_1]_{bc}}{(p_- - k_1)^2 - (mc)^2} (\gamma^\nu_{cd} \tilde{\epsilon}_{2\nu}) v_d^{(s_+)}. \tag{A.4}$$

Then, for the diagram on the *left*,

$$\begin{aligned} -i\mathcal{M}_L &= \bar{u}^{(s_-)}(p_-)(ig\gamma^\rho \epsilon_{2\rho}) \frac{i}{\not{p}_- - \not{k}_2 - mc} (ig\gamma^\sigma \epsilon_{1\sigma}) v^{(s_+)}(p_+) \\ &= -ig^2 \bar{u}_i^{(s_-)} (\gamma^\rho_{ij} \epsilon_{2\rho}) \frac{[-\not{k}_2]_{jk}}{(p_- - k_2)^2 - (mc)^2} (\gamma^\sigma_{kl} \epsilon_{1\sigma}) v_l^{(s_+)}, \end{aligned} \tag{A.5}$$

once again using the special gauge, but droping the tilde notation for now. For a more rigorous treatment, you should also check that there is no relative negative sign between both diagrams/amplitudes, which can be seen by constructing the 4-point correlation function in the generating functional formalism and applying Wick's theorem (contractions of fields). Since both diagrams have the same structure (only reversed fermion arrows), makes sense to not expect any relative sign[1].

The next step is to compute the amplitude squared

$$\overline{|\mathcal{M}|}^2 = (\mathcal{M}_R + \mathcal{M}_L)(\mathcal{M}_R + \mathcal{M}_L)^\dagger = |\mathcal{M}_R|^2 + |\mathcal{M}_L|^2 + \mathcal{M}_R \mathcal{M}_L^\dagger + \mathcal{M}_L \mathcal{M}_R^\dagger. \tag{A.6}$$

[1] An example in which a relative negative sign is needed is the Bhabha scattering (electron-positron scattering) at lowest order.

The computation of the first term results in

$$\begin{aligned}|\mathcal{M}_R|^2 = g_e^4 (u_h^{(s_-)} \bar{u}_a^{(s_-)})(\gamma^\mu_{ab}\epsilon_{1\mu}) \frac{[-\not{k}_1]_{bc}}{(p_- - k_1)^2 - (mc)^2} (\gamma^\nu_{cd}\epsilon_{2\nu}) \\ \times (v_d^{(s_+)} \bar{v}_e^{(s_+)})(\gamma^\beta_{ef}\epsilon^*_{2\beta}) \frac{[-\not{k}_1]_{fg}}{(p_- - k_1)^2 - (mc)^2} (\gamma^\alpha_{gh}\epsilon^*_{1\alpha}),\end{aligned} \tag{A.7}$$

but we can sum over the spins to apply the properties

$$\sum_s u^s(p)\bar{u}^s(p) = \not{p} + mc \tag{A.8}$$

$$\sum_s v^s(p)\bar{v}^s(p) = \not{p} - mc, \tag{A.9}$$

and use that

$$(p_- - k_1)^2 = p_-^2 + k_1^2 - 2p_- \cdot k_1 = m^2c^2 + 0 - 2p_- \cdot k_1, \tag{A.10}$$

so

$$\sum_s |\mathcal{M}_R|^2 = g_e^4 \,\mathrm{Tr}\left[(\not{p}_- + mc)\frac{\not{\epsilon}_1 \not{k}_1 \not{\epsilon}_2}{2p_- \cdot k_1}(\not{p}_+ - mc)\frac{\not{\epsilon}^*_2 \not{k}_1 \not{\epsilon}^*_1}{2p_- \cdot k_1}\right]. \tag{A.11}$$

A similar reasoning can be applied to the other terms of eq. A.6. Grouping everything and averaging over the spins, we find

$$\frac{1}{4}\sum_{s_-,s_+} \overline{|\mathcal{M}|}^2 = \frac{g^4}{4}\,\mathrm{Tr}\left[(\not{p}_+ - mc)\left(\frac{\tilde{\not{\epsilon}}^*_2 \not{k}_1 \tilde{\not{\epsilon}}^*_1}{2p_- \cdot k_1} + \frac{\tilde{\not{\epsilon}}^*_1 \not{k}_2 \tilde{\not{\epsilon}}^*_2}{2p_- \cdot k_2}\right)(\not{p}_- + mc)\left(\frac{\tilde{\not{\epsilon}}_1 \not{k}_1 \tilde{\not{\epsilon}}_2}{2p_- \cdot k_1} + \frac{\tilde{\not{\epsilon}}_2 \not{k}_2 \tilde{\not{\epsilon}}_1}{2p_- \cdot k_2}\right)\right]. \tag{A.12}$$

To compute the trace, some properties are required. First of all, the trace of an odd number of gamma matrices is zero. Other relevant relations are that on-shell photons ($\not{k}\not{k} = k^2 = 0$) are transverse, so

$$\tilde{\not{\epsilon}}_1 \not{k}_1 = \tilde{\epsilon}_{1\mu} k_{1\nu} \gamma^\mu \gamma^\nu = \tilde{\epsilon}_{1\mu} k_{1\mu}(2g^{\mu\nu} - \gamma^\nu\gamma^\mu) = 2\underbrace{\tilde{\epsilon}_1 \cdot k_1}_{=0} - \not{k}_1 \tilde{\not{\epsilon}}_1 = -\not{k}_1 \tilde{\not{\epsilon}}_1 \tag{A.13}$$

and the normalisation of the polarisation vector imposes $\tilde{\not{\epsilon}}^*_1 \tilde{\not{\epsilon}}_1 = -1$. Finally,

$$\mathrm{Tr}\left[\gamma^\mu\gamma^\nu\gamma^\lambda\gamma^\sigma\right] = 4(g^{\mu\nu}g^{\lambda\sigma} - g^{\mu\lambda}g^{\nu\sigma} + g^{\mu\sigma}g^{\nu\lambda}), \tag{A.14}$$

is also a useful property. If you are brave enough, after a lengthy computation, you should

arrive at the deceivingly simple result of

$$2\left[\frac{k_1\cdot k_2}{p_-\cdot k_2}+\frac{k_2\cdot k_1}{p_-\cdot k_1}-4(\tilde{\epsilon}_1\cdot\tilde{\epsilon}_2)^2\right]. \tag{A.15}$$

It is useful to notice that

$$\begin{aligned}(k_1+k_2)^2 &= k_1^2+k_2^2+2k_1\cdot k_2 = 2k_1\cdot k_2\\ &= (p_-+p_+)^2 = p_-^2+p_+^2+2p_-\cdot p_+ = 2m^2c^2+2p_-\cdot p_+,\end{aligned} \tag{A.16}$$

so

$$k_1\cdot k_2 = m^2c^2+p_-\cdot p_+ = m^2c^2+p_-\cdot(k_1+k_2-p_-) = p_-\cdot k_1+p_-\cdot k_2. \tag{A.17}$$

Therefore, we find

$$2\left[\frac{p_-\cdot k_1}{p_-\cdot k_2}+\frac{p_-\cdot k_2}{p_-\cdot k_1}-4(\tilde{\epsilon}_1\cdot\tilde{\epsilon}_2)^2+2\right] \tag{A.18}$$

for the trace term in eq. A.12.

To compute the product $(\tilde{\epsilon}_1\cdot\tilde{\epsilon}_2)^2$, it is easier to go to a general gauge. The gauge symmetry implies that the photon field transforms as $A_\mu\to A'_\mu = A_\mu+\partial_\mu\lambda$. By choosing $\lambda = iae^{-ik\cdot x}$, we see that

$$\partial_\mu\lambda = ak_\mu e^{-ik\cdot x}. \tag{A.19}$$

Taking the plane wave solutions, $A_\mu\sim\epsilon_\mu e^{-ik\cdot x}$, a gauge transformation is equivalent to changing the polarisation vectors $\epsilon_\mu\to\epsilon'_\mu = \epsilon_\mu+ak_\mu$. Then, we may write the relation between the special gauge ($\epsilon' = \tilde{\epsilon}$) and an arbitrary gauge (ϵ) as $\tilde{\epsilon}_\mu = \epsilon_\mu - ak_\mu$. Imposing that $\tilde{\epsilon}_1\cdot p_- = 0 = \tilde{\epsilon}_2\cdot p_-$, we find

$$\tilde{\epsilon}_1 = \epsilon_1-\frac{\epsilon_1\cdot p_-}{k_1\cdot p_-}k_1 \tag{A.20}$$

$$\tilde{\epsilon}_2 = \epsilon_2-\frac{\epsilon_2\cdot p_-}{k_2\cdot p_-}k_2. \tag{A.21}$$

In this general gauge, we can compute $(\tilde{\epsilon}_1\cdot\tilde{\epsilon}_2)^2$ and average over the polarisation values (λ), which corresponds to performing the replacement

$$\sum_\lambda\epsilon^\lambda_\mu(k)\epsilon^\lambda_\nu(k)\to -g_{\mu\nu}. \tag{A.22}$$

This leads to the following substitutions

$$(\epsilon_1 \cdot \epsilon_2)^2 = \epsilon_1^\mu \epsilon_{\mu 2} \epsilon_1^\nu \epsilon_{\nu 2} \to -g^{\mu\nu}(-g_{\mu\nu}) = 4 \quad \text{(A.23)}$$

$$(\epsilon \cdot p)^2 = \epsilon^\mu p_\mu \epsilon^\nu p_\nu \to -g^{\mu\nu} p_\mu p_\nu = -p^2 \quad \text{(A.24)}$$

$$(\epsilon_1 \cdot \epsilon_2)(\epsilon_2 \cdot p)(\epsilon_1 \cdot k) = \epsilon_1^\mu \epsilon_{2\mu} \epsilon_{2\nu} p^\nu \epsilon_1^\alpha k_\alpha \to -g^{\mu\alpha}(-g_{\mu\nu}) p^\nu k_\alpha = g^{\mu\alpha} p_\mu k_\alpha = p \cdot k \quad \text{(A.25)}$$

$$(\epsilon \cdot k)(\epsilon \cdot p) = \epsilon^\mu k_\mu \epsilon^\nu p_\nu \to -g^{\mu\nu} k_\mu p_\nu = -k \cdot p. \quad \text{(A.26)}$$

Since the other terms in eq. A.18 do not depend on the polarisation, they receive an extra factor of 4 after the summation. By combining all results, the total invariant amplitude becomes

$$\begin{aligned} |\mathcal{M}|^2 &\equiv \frac{1}{4} \sum_{s,\lambda} \overline{|\mathcal{M}|}^2 \\ &= 2g^4 \left[\frac{p_- \cdot k_1}{p_- \cdot k_2} + \frac{p_- \cdot k_2}{p_- \cdot k_1} - m^4 c^4 \frac{(k_1 \cdot k_2)^2}{(k_1 \cdot p_-)^2 (k_2 \cdot p_-)^2} + 2m^2 c^2 \frac{(k_1 \cdot k_2)}{(k_1 \cdot p_-)(k_2 \cdot p_-)} \right]. \end{aligned} \quad \text{(A.27)}$$

Having the invariant amplitude, we can compute the cross section. The differential cross section for a two particle scattering is

$$d\sigma = |\mathcal{M}|^2 \frac{c^2 \hbar^2 S}{4 E_1 E_2} \frac{c}{|\mathbf{v}_1 - \mathbf{v}_2|} \frac{c \mathrm{d}^3 \mathbf{p}_-}{(2\pi)^3 2 E_-} \frac{c \mathrm{d}^3 \mathbf{p}_+}{(2\pi)^3 2 E_+} (2\pi)^4 \delta^4(p_+ + p_- - k_1 - k_2), \quad \text{(A.28)}$$

where S is a statistical factor for correcting if there are identical particles in the final state, so here $S = 1$. In the centre of momentum frame, the particle energies are the same, so $E_- = E_+ \equiv E$ and $E_1 = E_2 \equiv E_\gamma$, besides $\mathbf{v}_2 = -\mathbf{v}_1 = -\mathbf{c}$. The integral over $\mathbf{p}_+$ can be easily done due to the delta function. Inserting also the integration symbol over $\mathbf{p}_-$, we get

$$\sigma = \int \frac{c^4 \mathrm{d}^3 \mathbf{p}_-}{(2\pi)^3 2E} |\mathcal{M}|^2 \frac{\hbar^2}{4 E_\gamma^2} \frac{1}{2} \frac{(2\pi)}{2E} \delta(E/c + E/c - E_\gamma/c - E_\gamma/c), \quad \text{(A.29)}$$

where now $\mathbf{p}_+ + \mathbf{p}_- = \mathbf{k}_1 + \mathbf{k}_2$. In the centre of momentum frame, $\mathbf{p}_+ = -\mathbf{p}_-$, so the Mandelstam variable $s = (p_- + p_+)^2 c^2$ is $s = 4E^2$. Then,

$$\sigma = \int \frac{c^4 \mathrm{d}^3 \mathbf{p}_-}{(2\pi)^2} \frac{1}{2} |\mathcal{M}|^2 \frac{\hbar^2}{4 E_\gamma^2 s} \delta(2E/c - 2E_\gamma/c). \quad \text{(A.30)}$$

For the electron, $E = \sqrt{|\mathbf{p}_-|^2 c^2 + m^2 c^4}$, so we need to compute

$$\begin{aligned}\sigma &= \frac{c^4\hbar^2}{4\pi^2}\frac{1}{2}\int \mathrm{d}|\mathbf{p}|_- \,\mathrm{d}\phi\,\mathrm{d}\cos\theta\,|\mathbf{p}_-|^2\frac{1}{4E_\gamma^2 s}|\mathcal{M}|^2\delta\left(\frac{2}{c}\sqrt{|\mathbf{p}_-|^2c^2+m^2c^4}-\frac{2}{c}E_\gamma\right)\\ &= \frac{c^4\hbar^2}{4\pi}\int \mathrm{d}|\mathbf{p}|_- \,\mathrm{d}\cos\theta\,|\mathbf{p}_-|^2\frac{1}{4E_\gamma^2 s}|\mathcal{M}|^2\delta\left(\frac{2}{c}\sqrt{|\mathbf{p}_-|^2c^2+m^2c^4}-\frac{2}{c}E_\gamma\right),\end{aligned} \tag{A.31}$$

choosing spherical coordinates and already integrating over the azimuth angle ϕ, as the invariant amplitude does not depend on it. To perform the $|\mathbf{p}_-|$ integration, we use the following property of the delta function

$$\int_{-\infty}^{\infty} f(x)\delta(g(x))\,\mathrm{d}x = \sum_a \frac{1}{|g'(a)|}\int_{-\infty}^{\infty} f(x)\delta(x-a)\,\mathrm{d}x\,, \tag{A.32}$$

where a are the zeros of $g(x)$, so we can define

$$g(x) = \frac{2}{c}\sqrt{x^2c^2+m^2c^4}-\frac{2}{c}E_\gamma \Rightarrow g'(x) = \frac{2xc}{\sqrt{x^2c^2+m^2c^4}}. \tag{A.33}$$

For $g(x)=0$, then $E = E_\gamma$ and $|\mathbf{p}_-|^2c^2 = E_\gamma^2 - m^2c^4$, so

$$g'(a) = \frac{2\sqrt{E_\gamma^2-m^2c^4}}{E_\gamma} = 2\sqrt{1-\frac{m^2c^4}{E_\gamma^2}}. \tag{A.34}$$

Therefore, integrating $|\mathbf{p}_-|$ in eq. A.31 with the above property and using $s = 4E^2$ results in

$$\sigma = \frac{c^2\hbar^2}{4\pi}\frac{1}{8}\int_{-1}^{1}\mathrm{d}\cos\theta\sqrt{1-\frac{4m^2c^4}{s}}\frac{1}{s}|\mathcal{M}|^2. \tag{A.35}$$

It is useful to relate s to the speed $\beta = v/c$ of the electron/positron in the centre-of-momentum frame. Since $\beta = pc/E$, then

$$\beta = \sqrt{1-\frac{4m^2c^4}{s}} \Rightarrow s = \frac{4m^2c^4}{1-\beta^2}, \tag{A.36}$$

so

$$\sigma = \frac{c^2\hbar^2}{4\pi}\frac{1}{32m^2c^4}\int_{-1}^{1}\mathrm{d}\cos\theta\,\beta(1-\beta^2)|\mathcal{M}|^2. \tag{A.37}$$

To perform this integration, we need to identify the dependence of the invariant

amplitude (eq. A.27) to the spherical coordinate angle θ. For instance,

$$p_- \cdot k_1 = \frac{E^2}{c^2} - \mathbf{p}_- \cdot \mathbf{k} = \frac{E^2}{c^2} - |\mathbf{p}_-||\mathbf{k}| \cos\theta = \frac{E^2}{c^2} - \frac{E}{c}|\mathbf{p}_-| \cos\theta = \frac{s}{4c^2}(1 - \beta \cos\theta) \tag{A.38}$$

$$p_- \cdot k_2 = \frac{E^2}{c^2} - |\mathbf{p}_-||\mathbf{k}| \cos(\pi - \theta) = \frac{E^2}{c^2} + \frac{E}{c}|\mathbf{p}_-| \cos\theta = \frac{s}{4c^2}(1 + \beta \cos\theta). \tag{A.39}$$

The other terms can be similarly computed, resulting in

$$\sigma = \frac{c^2\hbar^2 g^4}{m^2c^4 64\pi} \int_{-1}^{1} \mathrm{d}\cos\theta \, \beta(1-\beta^2) \left[\frac{2 + 2\beta^2 \cos^2\theta}{1 - \beta^2 \cos^2\theta} + \frac{4(1-\beta^2)}{(1 - \beta^2 \cos^2\theta)} - \frac{4(1-\beta^2)^2}{(1 - \beta^2 \cos^2\theta)^2} \right], \tag{A.40}$$

The constants on the front can be rearranged as

$$\frac{c^2\hbar^2 g^4}{m^2c^4 64\pi} = \frac{\pi}{2} \left(\frac{\alpha\hbar c}{mc^2} \right)^2 \frac{1}{2}. \tag{A.41}$$

Finally, the integration on $\cos\theta$ can be performed using standard integration techniques. This cancels a factor $\beta/2$ and results in

$$\sigma = \frac{\pi}{2} \left(\frac{\alpha\hbar c}{mc^2} \right)^2 (1-\beta^2) \left[(3 - \beta^4) \ln\left(\frac{1+\beta}{1-\beta} \right) - 2\beta(2 - \beta^2) \right], \tag{A.42}$$

which is the same expression as eq. 3.23 of Chapter 3.

Appendix B

Marginal Likelihood

If the modelled flux of gamma ray sources, $\Phi_{\rm mod}$, is parametrized such that the flux normalisation parameter can be factored out, as

$$\Phi_{\rm mod}^{(j)} = N_0^{(j)} \tilde{\phi}_{\rm mod}, \tag{B.1}$$

where $\tilde{\phi}_{\rm mod}$ is the remaining parametrization of the flux, then the assumption of Gaussian errors and independence of the spectra allows an analytical marginalisation of the posterior probability distribution over N_0. For a total of N spectra in the likelihood, this means performing the integral

$$p(\mathbf{\Omega}_r|D,I) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\prod_{j=1}^{N} \mathrm{d}N_0^{(j)} \right) \frac{p(D|\mathbf{\Omega},I)p(\mathbf{\Omega}|I)}{p(D|I)}, \tag{B.2}$$

where $\mathbf{\Omega}_r$ is the set of parameters after integrating out the $N_0^{(j)}$ normalisation variables, with $j = 1, \ldots, N$. We also use an improper prior such that $p(\Omega|I) = 0$ if any $N_0^{(j)} < 0$. As the prior and evidence are constants during the integration, our task turns into computing the marginal likelihood

$$p(D|\mathbf{\Omega}_r,I) = \int_{0}^{\infty} \cdots \int_{0}^{\infty} \left(\prod_{j=1}^{N} \mathrm{d}N_0^{(j)} \right) \frac{1}{Z} \exp\left(-\frac{1}{2}\chi^2 \right), \tag{B.3}$$

where Z is the probability normalisation and

$$\chi^2 = \sum_{j=1}^{N} \sum_{i=1}^{n_j} \left[\frac{\Phi_{\rm obs}^{(j)}\left(E_i^{(j)}\right) - \Phi_{\rm mod}^{(j)}\left(E_i^{(j)};\mathbf{\Omega}\right)}{\sigma\left(E_i^{(j)}\right)} \right]^2 . \tag{B.4}$$

Due to the independence of the normalisation variables, we arrive at a product of

integrals

$$p(D|\mathbf{\Omega}_r, I) = \frac{1}{Z}\prod_{j=1}^{N}\left\{\int_0^\infty \mathrm{d}N_0^{(j)} \exp\left[-\frac{1}{2}\sum_{i=1}^{n_j}\left(\frac{\Phi_{\mathrm{obs}}^{(j)}\left(E_i^{(j)}\right) - N_0^{(j)}\tilde{\phi}_{\mathrm{mod}}^{(j)}\left(E_i^{(j)};\mathbf{\Omega}_r\right)}{\sigma\left(E_i^{(j)}\right)}\right)^2\right]\right\}. \tag{B.5}$$

For each integral, by "completing the square", a Gaussian integral can be performed. Simplifying the notation by droping the j dependence, we realise that

$$\sum_i\left(\frac{\Phi(E_i) - N_0\tilde{\phi}(E_i)}{\sigma(E_i)}\right)^2 = \sum_i \frac{N_0^2\tilde{\phi}^2(E_i) - 2N_0\Phi(E_i)\tilde{\phi}(E_i) + \Phi^2(E_i)}{\sigma^2(E_i)} \tag{B.6}$$

resulting in the expression

$$AN_0^2 - 2N_0B + C, \tag{B.7}$$

where we define

$$A \equiv \left(\sum_i \frac{\tilde{\phi}^2(E_i)}{\sigma^2(E_i)}\right) \tag{B.8}$$

$$B \equiv \left(\sum_i \frac{\Phi(E_i)\tilde{\phi}(E_i)}{\sigma^2(E_i)}\right) \tag{B.9}$$

$$C \equiv \left(\sum_i \frac{\Phi^2(E_i)}{\sigma^2(E_i)}\right), \tag{B.10}$$

which assume different values for each source j (so we are actually defining A_j, B_j and C_j). By completing the square,

$$AN_0^2 - 2N_0B + C = A\left(N_0^2 - 2N_0\frac{B}{A}\right) + C = A\left(N_0 - \frac{B}{A}\right)^2 + C - \frac{B^2}{A}, \tag{B.11}$$

so the marginal likelihood becomes

$$p(D|\mathbf{\Omega}_r, I) = \frac{1}{Z}\prod_{j=1}^{N}\left\{\int_0^\infty \mathrm{d}N_0^{(j)} \exp\left[-\frac{A_j}{2}\left(N_0^{(j)} - \frac{B_j}{A_j}\right)^2\right]\exp\left(-\frac{1}{2}\left(C_j - \frac{B_j^2}{A_j}\right)\right)\right\}$$

and is simplified to

$$p(D|\mathbf{\Omega}_r, I) = \frac{1}{Z}\exp\left[-\frac{1}{2}\sum_{j=1}^{N}\left(C_j - \frac{B_j^2}{A_j}\right)\right]\prod_{j=1}^{N}\left\{\int_0^\infty \mathrm{d}N_0^{(j)} \exp\left[-\frac{A_j}{2}\left(N_0^{(j)} - \frac{B_j}{A_j}\right)^2\right]\right\}.$$

Changing variables

$$N_0^{(j)} \to u = \sqrt{\frac{A_j}{2}} \left(N_0^{(j)} - \frac{B_j}{A_j} \right) \tag{B.12}$$

and performing the Gaussian integral, we arrive at

$$p(D|\mathbf{\Omega}_r, I) = \frac{1}{Z} \exp\left[-\frac{1}{2} \sum_{j=1}^{N} \left(C_j - \frac{B_j^2}{A_j} \right) \right] \prod_{j=1}^{N} \sqrt{\frac{\pi}{2A_j}} \left[1 + \mathrm{erf}\left(\frac{B_j}{\sqrt{2A_j}} \right) \right], \tag{B.13}$$

which can be computationally implemented for MCMC purposes.

Appendix C

Complementary Material for Chapter 4

True EBL: F10+P19 (dust fractions)						
	Obs. Time < 5 h			Obs. Time ≥ 5 h		
N.S.	a.f	ESS	ϵ	a.f.	ESS	ϵ
2	0.654	3×10^5	8.38×10^{-3}	0.644	1.8×10^5	1.08×10^{-2}
3	0.559	1.8×10^5	1.09×10^{-2}	0.555	1.2×10^5	1.34×10^{-2}
4	0.485	2.25×10^5	9.84×10^{-3}	0.486	1.125×10^5	1.39×10^{-2}
5	0.428	2.7×10^5	9.01×10^{-3}	0.428	1.35×10^5	1.27×10^{-2}
6	0.371	2.1×10^5	1.02×10^{-2}	0.377	1.099×10^5	1.42×10^{-2}
7	0.337	2×10^5	1.05×10^{-2}	0.335	1.314×10^5	1.30×10^{-2}
8	0.285	2×10^5	1.05×10^{-2}	0.290	1×10^5	1.49×10^{-2}
9	0.263	2.4×10^5	9.59×10^{-3}	0.254	1.2×10^5	1.36×10^{-2}
10	0.224	1.995×10^5	1.05×10^{-2}	0.225	1.08×10^5	1.43×10^{-2}
11	0.190	1.71×10^5	1.14×10^{-2}	0.201	1.098×10^5	1.42×10^{-2}
12	0.173	2×10^5	1.05×10^{-2}	0.174	1.2×10^5	1.35×10^{-2}
True EBL: F10+P19 (dust fractions + PAH temperature)						
2	0.416	3.6×10^5	7.73×10^{-3}	0.370	2.4×10^5	9.47×10^{-3}
3	0.448	2.4×10^5	9.52×10^{-3}	0.511	2.4×10^5	9.52×10^{-3}
4	0.305	1.5×10^5	1.21×10^{-2}	0.427	1.44×10^5	1.23×10^{-2}
5	0.386	1.8×10^5	1.11×10^{-2}	0.377	1.4×10^5	1.25×10^{-2}
6	0.330	1.8×10^5	1.11×10^{-2}	0.319	1.05×10^5	1.45×10^{-2}
7	0.287	1.44×10^5	1.24×10^{-2}	0.286	1.224×10^5	1.34×10^{-2}
8	0.249	1.8×10^5	1.34×10^{-2}	0.247	1.144×10^5	1.39×10^{-2}
9	0.217	1.68×10^5	1.15×10^{-2}	0.213	8.54×10^4	1.61×10^{-2}
10	0.192	1.55×10^5	1.19×10^{-2}	0.186	1.152×10^5	1.38×10^{-2}
11	0.168	1.7×10^5	1.14×10^{-2}	0.168	1.22×10^5	1.34×10^{-2}
12	0.148	1.6×10^5	1.17×10^{-2}	0.145	1.1×10^5	1.41×10^{-2}

Table C.1: Statistical MCMC information about the analysis performed with simulated spectra with EBL absorption from F10+P19. For each MCMC run we present the number of sources (N.S.), the acceptance fraction (a.f.), the effective sample size (ESS) and the precision ϵ, adopting a confidence interval of 95%.

True EBL: D11 (dust fractions)						
	Obs. Time $< 5\,\mathrm{h}$			Obs. Time $\geq 5\,\mathrm{h}$		
N.S.	a.f	ESS	ϵ	a.f.	ESS	ϵ
2	0.662	3.6×10^5	7.65×10^{-3}	0.663	3.6×10^5	7.65×10^{-3}
3	0.568	2.248×10^5	9.78×10^{-3}	0.573	1.8×10^5	1.09×10^{-2}
4	0.477	1.5×10^5	1.20×10^{-2}	0.487	1.8×10^5	1.10×10^{-2}
5	0.432	1.758×10^5	1.12×10^{-2}	0.429	1.5×10^5	1.21×10^{-2}
6	0.385	1.68×10^5	1.14×10^{-2}	0.396	1.76×10^5	1.12×10^{-2}
7	0.340	2×10^5	1.05×10^{-2}	0.336	1.4×10^5	1.26×10^{-2}
8	0.288	1.42×10^5	1.25×10^{-2}	0.283	1.16×10^5	1.38×10^{-2}
9	0.254	1.368×10^5	1.27×10^{-2}	0.252	1.14×10^5	1.39×10^{-2}
10	0.227	1.71×10^5	1.14×10^{-2}	0.221	1.188×10^5	1.36×10^{-2}
11	0.192	1.2×10^5	1.36×10^{-2}	0.197	1.2×10^5	1.36×10^{-2}
12	0.169	1.152×10^5	1.38×10^{-2}	0.174	1.188×10^5	1.36×10^{-2}
True EBL: D11 (dust fractions + PAH temperature)						
2	0.548	3.6×10^5	7.73×10^{-3}	0.583	2.25×10^5	9.78×10^{-3}
3	0.378	3.68×10^5	7.69×10^{-3}	0.504	1.73×10^5	1.12×10^{-2}
4	0.331	2.8×10^5	8.85×10^{-3}	0.427	1.5×10^5	1.21×10^{-3}
5	0.300	2.8×10^5	8.87×10^{-3}	0.357	1.125×10^5	1.40×10^{-3}
6	0.252	2.4×10^5	9.59×10^{-3}	0.333	1.76×10^5	1.12×10^{-2}
7	0.290	2.4×10^5	9.59×10^{-3}	0.281	1.168×10^5	1.37×10^{-2}
8	0.249	1.995×10^5	1.05×10^{-2}	0.260	1.596×10^5	1.18×10^{-2}
9	0.213	1.7×10^5	1.14×10^{-2}	0.231	1.71×10^5	1.14×10^{-2}
10	0.191	1.76×10^5	1.12×10^{-2}	0.196	1.2×10^5	1.36×10^{-2}
11	0.161	1.32×10^5	1.29×10^{-2}	0.173	1.296×10^5	1.30×10^{-2}
12	0.145	1.525×10^5	1.41×10^{-2}	0.150	1.22×10^5	1.34×10^{-2}

Table C.2: Same as table C.1, but for spectra attenuated by D11's EBL model.

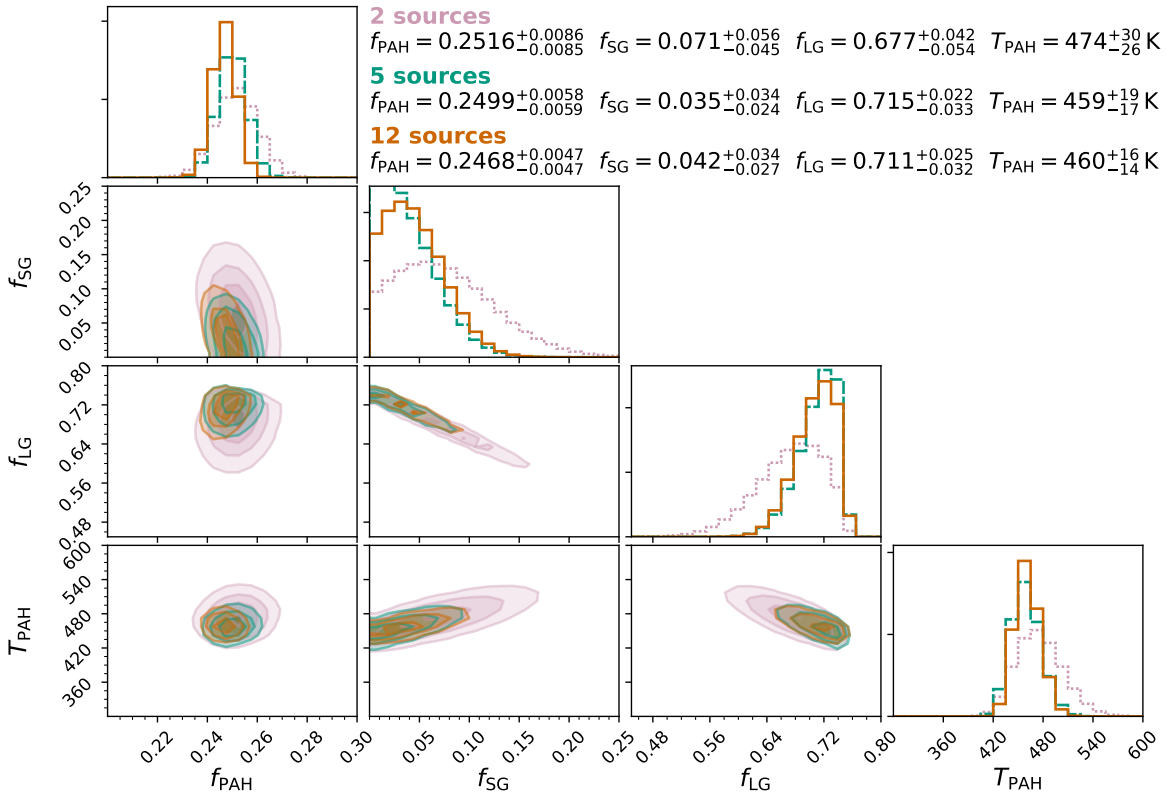

Figure C.1: One and two dimensional marginal distributions of EBL parameters, including the temperature of the PAH component, T_{PAH}, using the set of long observation time attenuated by F10+P19 model.

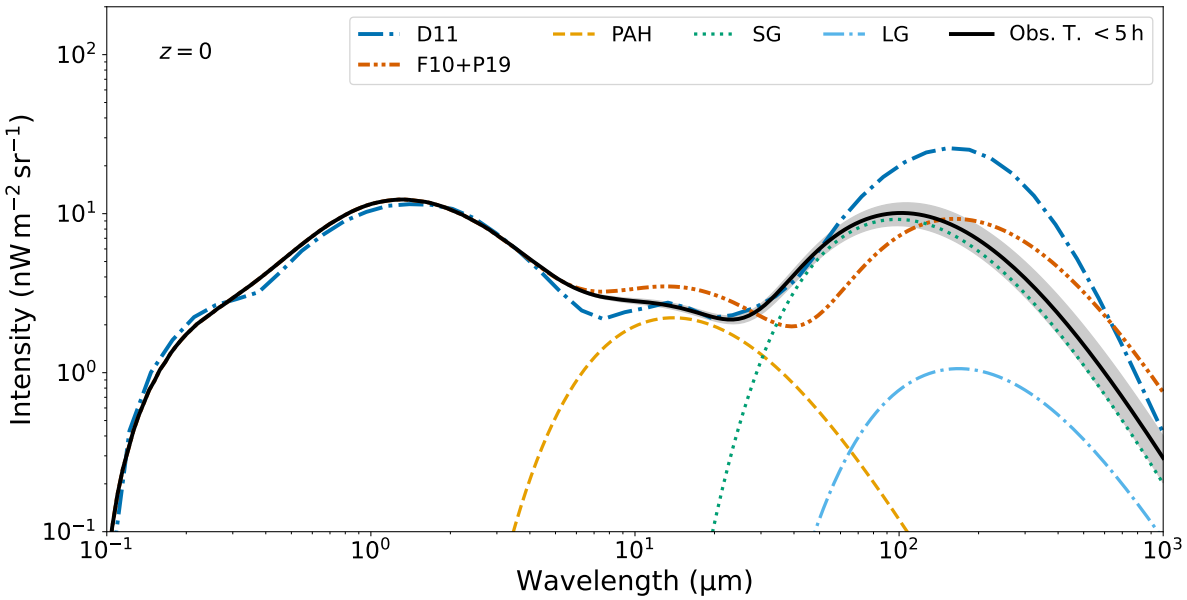

Figure C.2: EBL's local spectral intensity as inferred with the data set of short observational time and sources attenuated by D11, considering the PAH temperature as a free parameter. The individual PAH, SG and LG contributions are shown, alongside the F10+P19 and D11 nominal curves.

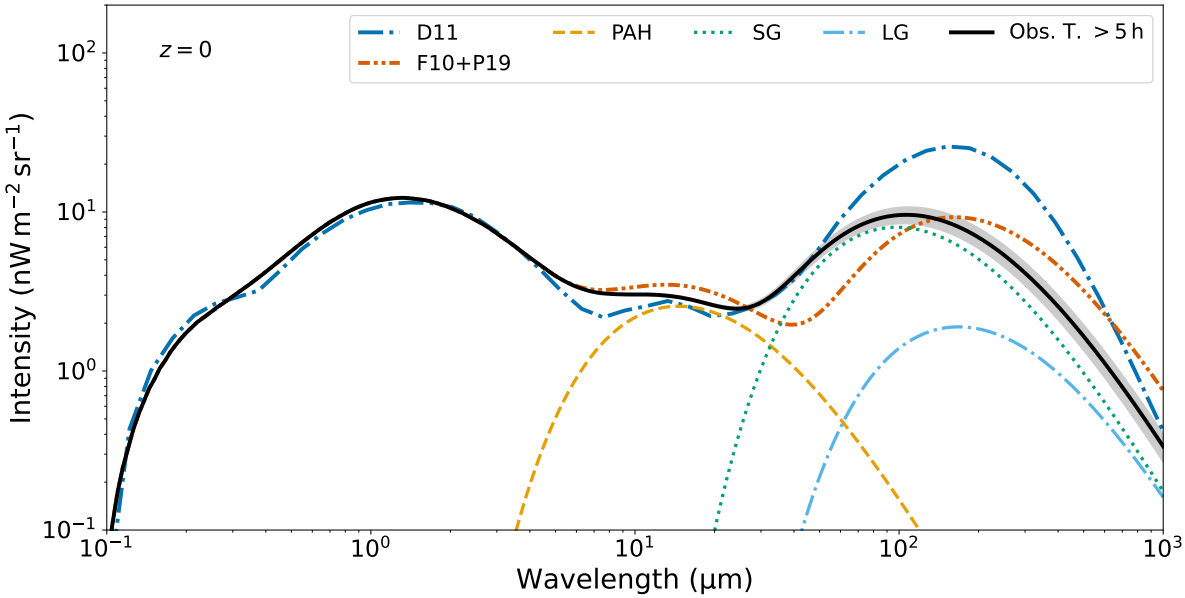

Figure C.3: EBL's local spectral intensity as inferred with the data set of long observational time and sources attenuated by D11, considering the PAH temperature as a free parameter. The individual PAH, SG and LG contributions are shown, alongside the F10+P19 and D11 nominal curves.

Appendix D

Complementary Material for Chapter 5

True EBL: F10+P19 (dust fractions + Hubble constant)						
	Obs. Time $< 5\,\mathrm{h}$			Obs. Time $\geq 5\,\mathrm{h}$		
N.S.	a.f	ESS	ϵ	a.f.	ESS	ϵ
2	0.540	7.2×10^5	5.47×10^{-3}	0.566	7.2×10^5	5.47×10^{-3}
3	0.478	7.2×10^5	5.50×10^{-3}	0.480	3.6×10^5	7.78×10^{-3}
4	0.413	6×10^5	6.04×10^{-3}	0.421	2.24×10^5	9.89×10^{-3}
5	0.352	3.6×10^5	7.82×10^{-3}	0.362	1.68×10^5	1.14×10^{-2}
6	0.309	3×10^5	8.57×10^{-3}	0.324	2.1×10^5	1.02×10^{-2}
7	0.267	2.52×10^5	9.36×10^{-3}	0.266	1.2×10^5	1.36×10^{-2}
8	0.222	1.68×10^5	1.15×10^{-2}	0.235	9.24×10^4	1.55×10^{-2}
9	0.211	2.52×10^5	9.36×10^{-3}	0.201	8.04×10^4	1.66×10^{-2}
10	0.180	1.8×10^5	1.09×10^{-2}	0.179	8.4×10^4	1.62×10^{-2}
11	0.159	2.16×10^5	1.01×10^{-2}	0.161	9.18×10^4	1.55×10^{-2}
12	0.138	1.68×10^5	1.14×10^{-2}	0.139	8.4×10^4	1.62×10^{-2}
True EBL: D11 (dust fractions + Hubble constant)						
2	0.542	6.72×10^5	5.66×10^{-3}	0.485	3.36×10^5	8×10^{-3}
3	0.473	5.6×10^5	6.23×10^{-3}	0.436	2.24×10^5	9.86×10^{-3}
4	0.407	3.36×10^5	8.10×10^{-3}	0.382	2.24×10^4	9.89×10^{-3}
5	0.348	2.24×10^5	9.91×10^{-3}	0.328	1.68×10^5	1.14×10^{-2}
6	0.303	1.68×10^5	1.14×10^{-2}	0.287	1.32×10^5	1.29×10^{-2}
7	0.269	1.68×10^5	1.14×10^{-2}	0.271	1.44×10^5	1.24×10^{-2}
8	0.247	1.68×10^5	1.14×10^{-2}	0.235	1.3×10^5	1.3×10^{-2}
9	0.212	1.365×10^5	1.27×10^{-2}	0.222	1.68×10^5	1.15×10^{-2}
10	0.183	1.302×10^5	1.3×10^{-2}	0.187	1.12×10^5	1.14×10^{-2}
11	0.157	1.08×10^5	1.43×10^{-2}	0.160	9.15×10^4	1.55×10^{-2}
12	0.137	1.088×10^5	1.42×10^{-2}	0.134	8×10^4	1.66×10^{-2}

Table D.1: Statistical MCMC information about the cosmology analysis performed with simulated spectra with EBL absorption from F10+P19. For each MCMC run we present the number of sources (N.S.), the acceptance fraction (a.f.), the effective sample size (ESS) and the precision ϵ, adopting a confidence interval of 95%.

Appendix E

Complementary Material for Chapter 6

E.1 Partial derivatives for HMC

For sampling the posterior distribution $p(\mathbf{\Omega}|D, I)$, the HMC requires solving Hamilton's equations for each parameter Ω_i. The rate of change of the conjugate momenta is given by

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = -\frac{\partial H}{\partial \Omega_i} = \frac{\partial}{\partial \Omega_i} \ln(p(\mathbf{\Omega}|D, I)), \tag{E.1}$$

where

$$p(\mathbf{\Omega}|D, I) \propto \frac{1}{Z} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{N} \sum_{i=1}^{n_j} \left[\frac{\Phi_{\mathrm{obs}}^{(j)}\left(E_i^{(j)}\right) - \Phi_{\mathrm{mod}}^{(j)}\left(E_i^{(j)}; \mathbf{\Omega}\right)}{\sigma\left(E_i^{(j)}\right)} \right]^2 \right\}, \tag{E.2}$$

as we adopt constant priors. Therefore, we need to compute the partial derivatives with respect to the EBL parameters (dust fractions) and the intrinsic spectral parameters. Simplifying the notation and substituting back at Hamilton's equation,

$$\frac{\mathrm{d}p_k}{\mathrm{d}t} = -\frac{1}{2} \frac{\partial}{\partial \Omega_k} \sum_{j=1}^{N} \sum_{i=1}^{n_j} \left[\frac{\Phi_{\mathrm{obs},i}^{(j)} - \Phi_{\mathrm{intr},i}^{(j)} e^{-\tau_i^{(j)}}}{\sigma} \right]^2, \tag{E.3}$$

expressing the modelled flux as the intrinsic part times the exponential attenuation from the optical depth. The index i refers to the energy bins and the index j to a given source.

First, for the dust fractions, $\Omega_k = \{f_{\rm PAH}, f_{\rm SG}\} \equiv f_k$, we obtain

$$\begin{aligned}\frac{\mathrm{d}p_k}{\mathrm{d}t} &= -\frac{1}{2}\frac{\partial}{\partial f_k}\sum_{j=1}^{N}\sum_{i=1}^{n_j}\left[\frac{\Phi^{(j)}_{\mathrm{obs},i} - \Phi^{(j)}_{\mathrm{intr},i}e^{-\tau_i^{(j)}}}{\sigma}\right]^2 \\ &= \sum_{j=1}^{N}\sum_{i=1}^{n_j}\left[\frac{\Phi^{(j)}_{\mathrm{obs},i} - \Phi^{(j)}_{\mathrm{intr},i}e^{-\tau_i^{(j)}}}{\sigma}\right]\frac{\Phi^{(j)}_{\mathrm{intr},i}}{\sigma}e^{-\tau_i^{(j)}}\left(-\frac{\partial\tau_i^{(j)}}{\partial f_k}\right) \\ &= -\sum_{j=1}^{N}\sum_{i=1}^{n_j}\left[\frac{\Phi^{(j)}_{\mathrm{obs},i} - \Phi^{(j)}_{\mathrm{mod},i}}{\sigma^2}\right]\Phi^{(j)}_{\mathrm{mod},i}\left(\tau^{(j)}_{i,k} - \tau^{(j)}_{i,\mathrm{LG}}\right),\end{aligned} \tag{E.4}$$

using that

$$\tau_i^{(j)} = \tau^{(j)}_{i,\mathrm{star}} + f_{\rm PAH}\tau^{(j)}_{i,\mathrm{PAH}} + f_{\rm SG}\tau^{(j)}_{i,\mathrm{SG}} + (1 - f_{\rm PAH} - f_{\rm SG})\tau^{(j)}_{i,\mathrm{LG}}. \tag{E.5}$$

For the intrinsic spectral parameters, we consider the PL, LP and PLC models. The derivative with respect to the flux normalisation parameters can be written in general as

$$\begin{aligned}\frac{\mathrm{d}p_k}{\mathrm{d}t} &= -\frac{1}{2}\frac{\partial}{\partial N_0^{(k)}}\sum_{j=1}^{N}\sum_{i=1}^{n_j}\left[\frac{\Phi^{(j)}_{\mathrm{obs},i} - N_0^{(j)}\tilde{\Phi}^{(j)}_{\mathrm{mod},i}}{\sigma}\right]^2 \\ &= \sum_{i=1}^{n_j}\left[\frac{\Phi^{(k)}_{\mathrm{obs},i} - N_0^{(k)}\tilde{\Phi}^{(k)}_{\mathrm{mod},i}}{\sigma}\right]\frac{\tilde{\Phi}^{(k)}_{\mathrm{mod},i}}{\sigma} \\ &= \sum_{i=1}^{n_j}\left[\frac{\Phi^{(k)}_{\mathrm{obs},i} - \Phi^{(k)}_{\mathrm{mod},i}}{\sigma^2}\right]\tilde{\Phi}^{(k)}_{\mathrm{mod},i},\end{aligned} \tag{E.6}$$

where the tilde notation refers to the rest of the modelled flux, after factoring out N_0. For the remaining spectral parameters, ω_k, we write in general

$$\begin{aligned}\frac{\mathrm{d}p_k}{\mathrm{d}t} &= -\frac{1}{2}\frac{\partial}{\partial\omega_k}\sum_{j=1}^{N}\sum_{i=1}^{n_j}\left[\frac{\Phi^{(j)}_{\mathrm{obs},i} - \Phi^{(j)}_{\mathrm{mod},i}}{\sigma}\right]^2 \\ &= \sum_{i=1}^{n_j}\left[\frac{\Phi^{(k)}_{\mathrm{obs},i} - \Phi^{(k)}_{\mathrm{mod},i}}{\sigma^2}\right]\frac{\partial\Phi^{(k)}_{i,\mathrm{mod}}}{\partial\omega_k}.\end{aligned} \tag{E.7}$$

For the PL model, $\omega_k = \Gamma_k$. Since

$$\left(\frac{E}{E_0}\right)^{-\Gamma} = e^{-\Gamma\ln(E/E_0)}, \tag{E.8}$$

then

$$\frac{\partial \Phi_{i,\mathrm{mod}}^{(k)}}{\partial \Gamma_k} = -\ln(E_i/E_0)\Phi_{i,\mathrm{mod}}^{(k)}. \tag{E.9}$$

For the LP one, the parameter a_k, we get the same relation

$$\frac{\partial \Phi_{i,\mathrm{mod}}^{(k)}}{\partial a_k} = -\ln(E_i/E_0)\Phi_{i,\mathrm{mod}}^{(k)}, \tag{E.10}$$

while for the curvature parameter b_k, we notice

$$\left(\frac{E}{E_0}\right)^{-b\log(E/E_0)} = e^{-b\log(E/E_0)\ln(E/E_0)} \tag{E.11}$$

where log is the logarithm in base 10. Then,

$$\frac{\partial \Phi_{i,\mathrm{mod}}^{(k)}}{\partial b_k} = -\frac{[\ln(E_i/E_0)]^2}{\ln(10)}\Phi_{i,\mathrm{mod}}^{(k)}. \tag{E.12}$$

Finally, for the PLC model, we have the same derivative of Γ_k as the PL, while for the cut-off energy E_{cut} we obtain

$$\frac{\partial \Phi_{i,\mathrm{mod}}^{(k)}}{\partial E_{\mathrm{cut},k}} = \frac{E_i}{E_{\mathrm{cut},k}^2}\Phi_{i,\mathrm{mod}}^{(k)}. \tag{E.13}$$

E.2 Data references

Name	Survey/Year - Label	Reference
Markarian 421	MAGIC/2004-2005	Albert et al. (2007b)
	VERITAS/2008 - High A	Acciari et al. (2011b)
	VERITAS/2008 - Low	Acciari et al. (2011b)
	VERITAS/2008 - Very Low	Acciari et al. (2011b)
Name	Survey/Year - Label	Reference
Markarian 421	VERITAS/2008 - High B	Acciari et al. (2011b)
	VERITAS/2008 - Middle	Acciari et al. (2011b)
	VERITAS/2008 - Very High	Acciari et al. (2011b)
Markarian 501	HEGRA/1997	Aharonian et al. (2001)
PKS 2155-304	HESS/2006	Abramowski et al. (2013)

Table E.1: References of the IACT spectra utilised for PLC and LP sources.

Name	Survey/Year - Label	Reference
1ES 0229+200	HESS/2005-2006	Aharonian et al. (2007b)
	VERITAS/2010-2012	Aliu et al. (2014)
1ES 0347-121	HESS/2006	Aharonian et al. (2007a)
1ES 0414+009	HESS/2005-2009	H. E. S. S. Collaboration et al. (2012a)
	VERITAS/2008-2011	Aliu et al. (2012b)
1ES 0806+524	MAGIC/2011 - Low	Aleksić et al. (2015b)
	MAGIC/2011 - High	Aleksić et al. (2015b)
	VERITAS/2006-2008	Acciari et al. (2009a)
1ES 1011+496	MAGIC/2007	Albert et al. (2007c)
1ES 1101-232	HESS/2004-2005	Aharonian et al. (2006b)
1ES 1215+303	MAGIC/2011	Aleksić et al. (2015b)
	VERITAS/2011	Aliu et al. (2013b)
1ES 1218+304	VERITAS/2008-2009	Acciari et al. (2010a)
	VERITAS/2007	Acciari et al. (2009b)
	MAGIC/2005	Albert et al. (2006a)
1ES 1312-423	HESS/2004-2010	HESS Collaboration et al. (2013)
1ES 1727+502	VERITAS/2013	Archambault et al. (2015)
1ES 1959+650	VERITAS/2007-2011	Aliu et al. (2013a)
	MAGIC/2006	Tagliaferri et al. (2008)
1ES 2344+514	VERITAS/2007-2008	Acciari et al. (2011c)
	VERITAS/2007 - High	Acciari et al. (2011c)
	MAGIC/2005-2006	Albert et al. (2007a)
1RXS J101015.9	HESS/2006-2010	H. E. S. S. Collaboration et al. (2012b)
3C 279	MAGIC/2008	MAGIC Collaboration et al. (2008)

Table E.2: References of the IACT spectra utilised for PL sources.

Name	Survey/Year - Label	Reference
3C 66A	VERITAS/2008 - Low	Abdo et al. (2011)
	VERITAS/2008 - High	Abdo et al. (2011)
4C +21.35	MAGIC/2010	Aleksić et al. (2011)
AP Librae	HESS/2010-2011	H. E. S. S. Collaboration et al. (2015)
BL Lacertae	VERITAS/2011	Arlen et al. (2013)
Centaurus A	HESS/2004-2008	Aharonian et al. (2009)
H 1426+428	HEGRA/1999-2000	Aharonian et al. (2003)
	HEGRA/2002	Aharonian et al. (2003)
H 2356-309	HESS/2004-2007	H. E. S. S. Collaboration et al. (2010b)
IC 310	MAGIC/2012	Aleksić et al. (2014a)
	MAGIC/2009-2010 - High	Aleksić et al. (2014b)
	MAGIC/2009-2010 - Low	Aleksić et al. (2014b)
M87	HESS/2005	Aharonian et al. (2006a)
	HESS/2004	Aharonian et al. (2006a)
	MAGIC/2005-2007	Aleksić et al. (2012b)
	MAGIC/2008	Albert et al. (2008)
	VERITAS/2007	Acciari et al. (2008)
Markarian 180	MAGIC/2006	Albert et al. (2006b)
Markarian 421	MAGIC/2006	Acciari et al. (2009c)
	VERITAS/2008 - High C	Acciari et al. (2011b)

Table E.3: Continuation of table E.2.

Name	Survey/Year - Label	Reference
Markarian 501	VERITAS/2009	Acciari et al. (2011a)
NGC 1275	MAGIC/2009-2014	Ahnen et al. (2016)
PKS 0447-439	HESS/2009	H. E. S. S. Collaboration et al. (2013a)
PKS 1441+25	MAGIC/2015	Ahnen et al. (2015)
PKS 1510-089	HESS/2009	H. E. S. S. Collaboration et al. (2013b)
PKS 2005-489	HESS/2004-2007	H. E. S. S. Collaboration et al. (2010a)
PKS 2155-304	HESS/2005-2007 MAGIC/2006	H. E. S. S. Collaboration et al. (2010c) Aleksić et al. (2012a)
RBS 0413	VERITAS/2009	Aliu et al. (2012a)
RGB J0152+017	HESS/2007	Aharonian et al. (2008)
RGB J0710+591	VERITAS/2008-2009	Acciari et al. (2010b)
RX J0648.7+1516	VERITAS/2010	Aliu et al. (2011)

Table E.4: Continuation of table E.3.

E.3 Spectral fit

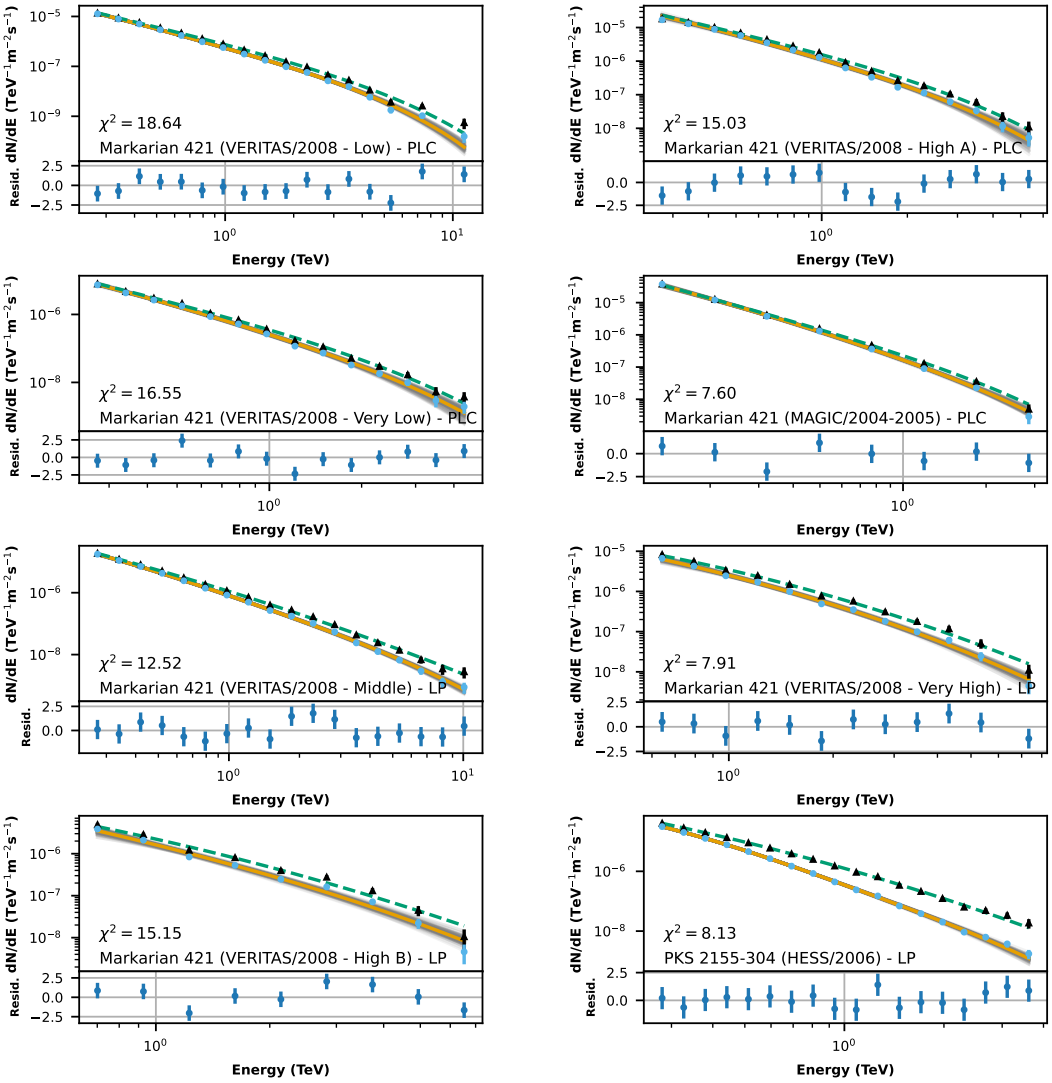

Figure E.1: Flux data (blue points) and fit from the median values of EBL and source parameters (orange solid curve), with the respective residuals (in units of uncertainty). The black points are the de-absorbed spectrum and the dashed green line is the intrinsic emission. Around the median fit, in grey, is a sample of 100 random models taken from the posterior distribution.

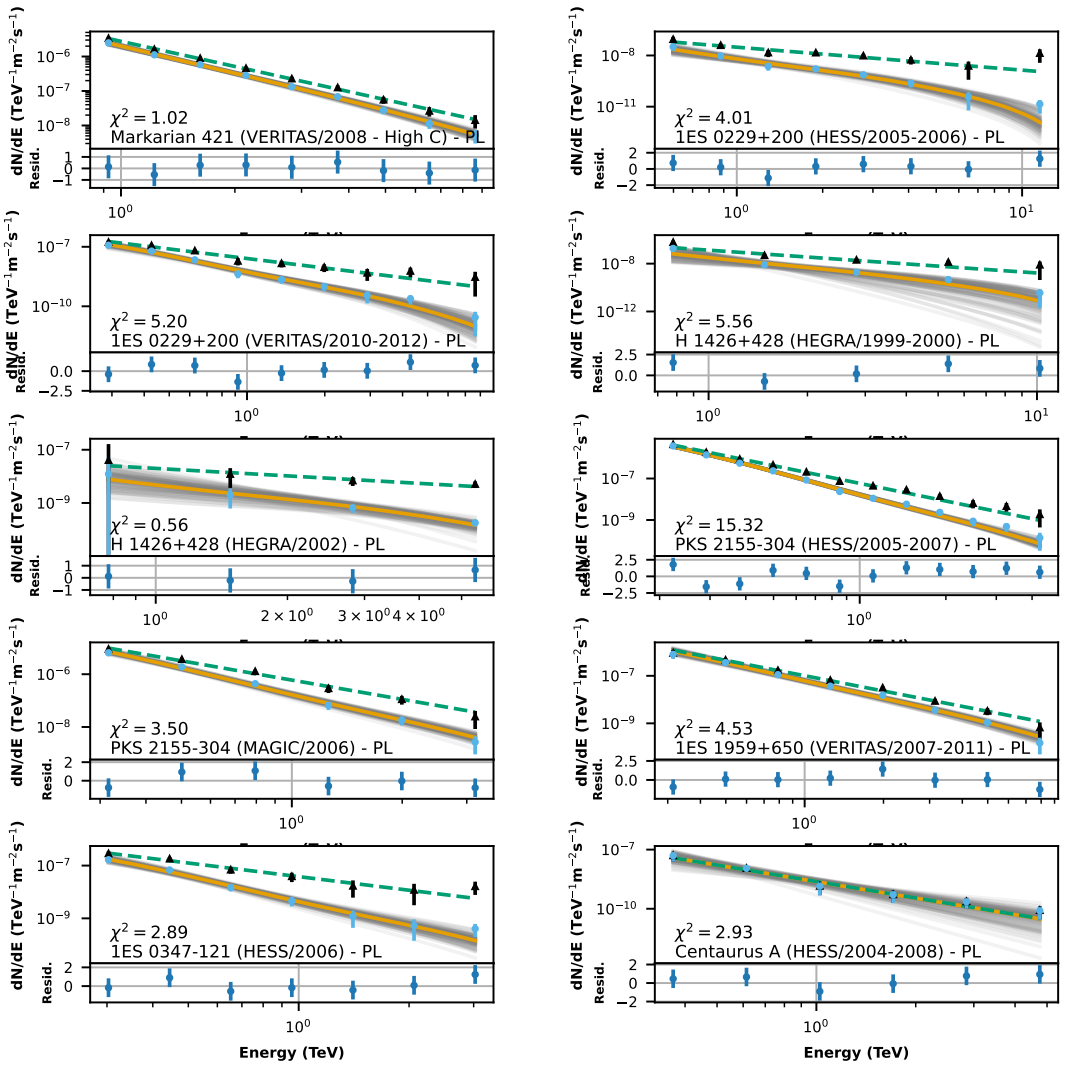

Figure E.2: Same as Figure E.1, but for PL sources.

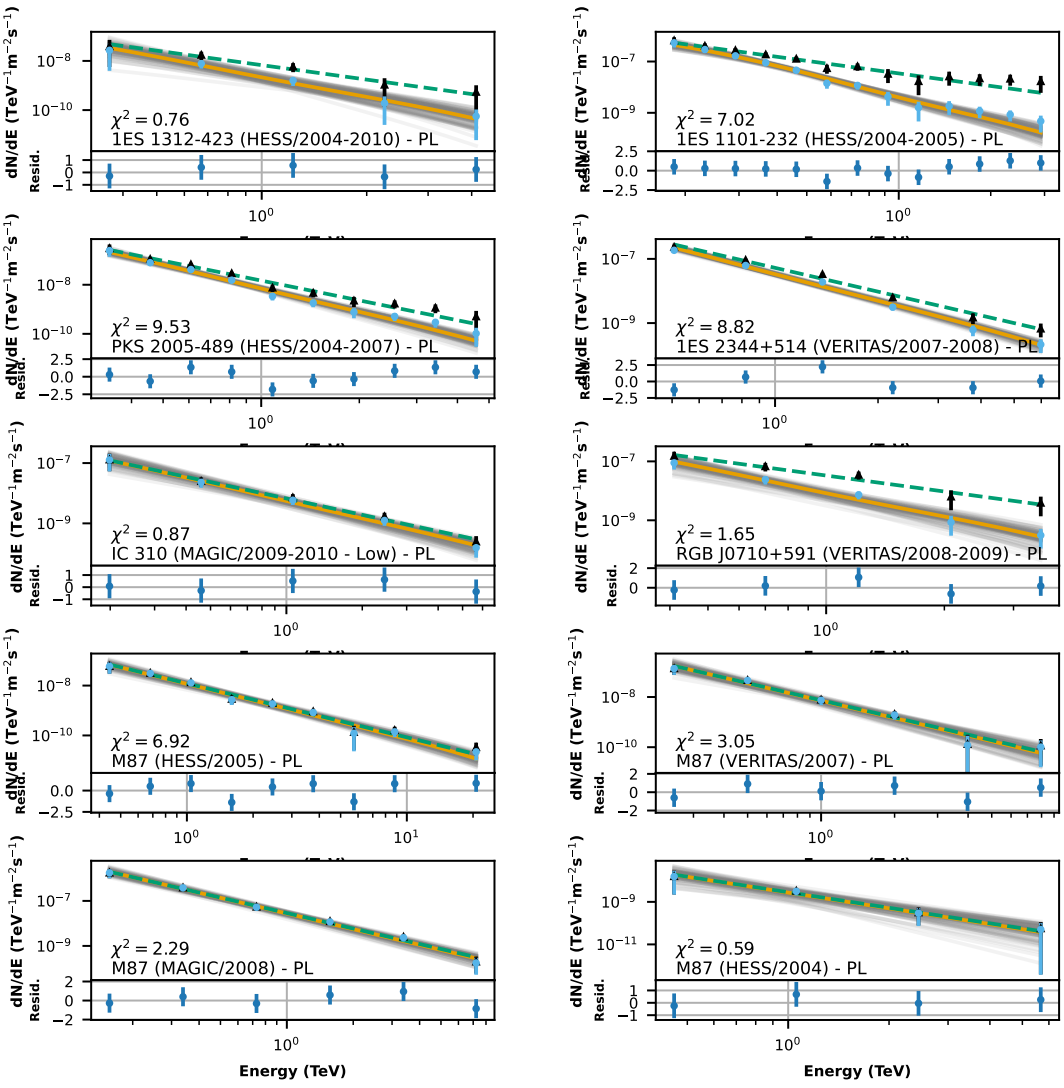

Figure E.3: Continuation of Figure E.2.

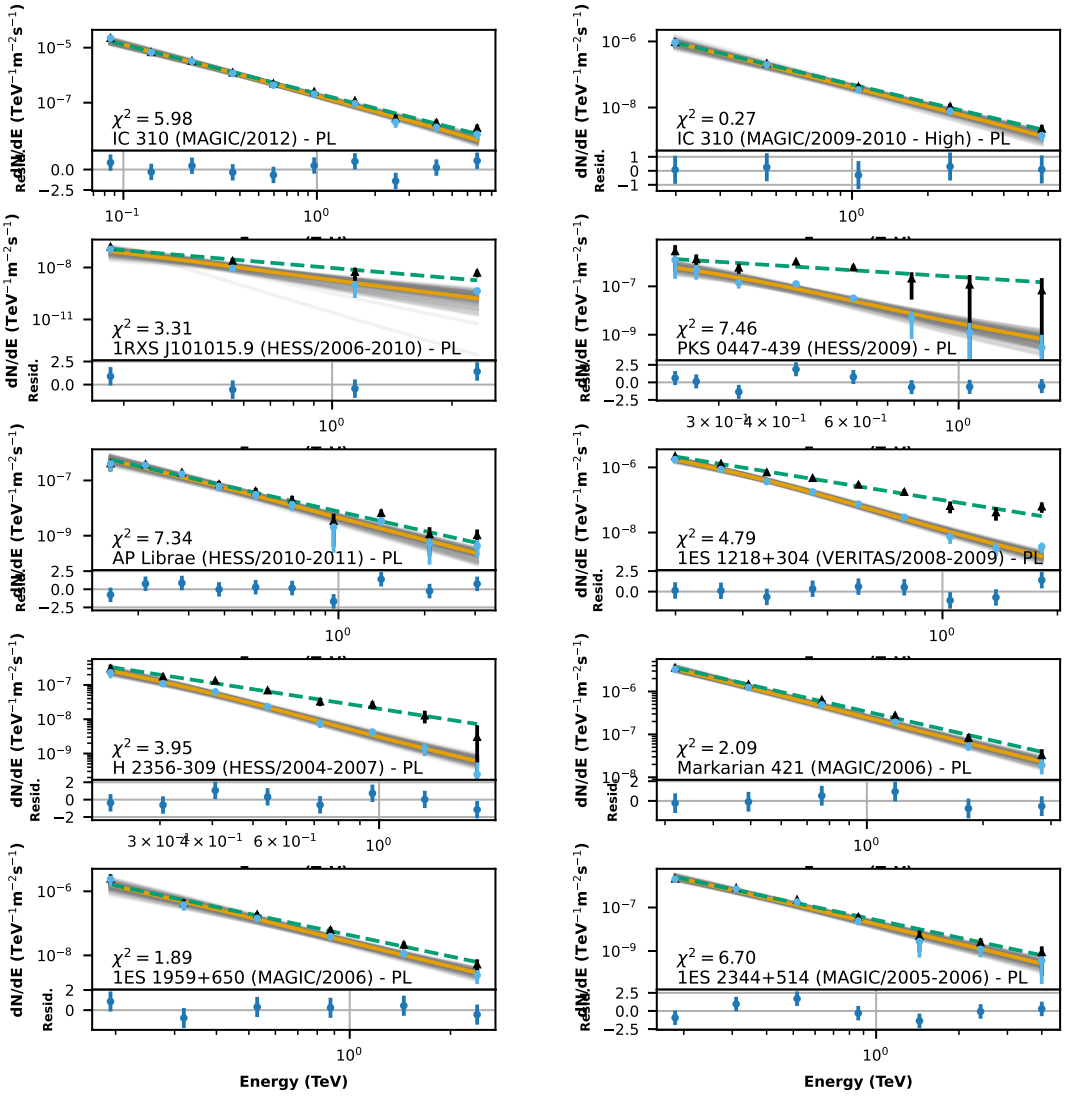

Figure E.4: Continuation of Figure E.3.

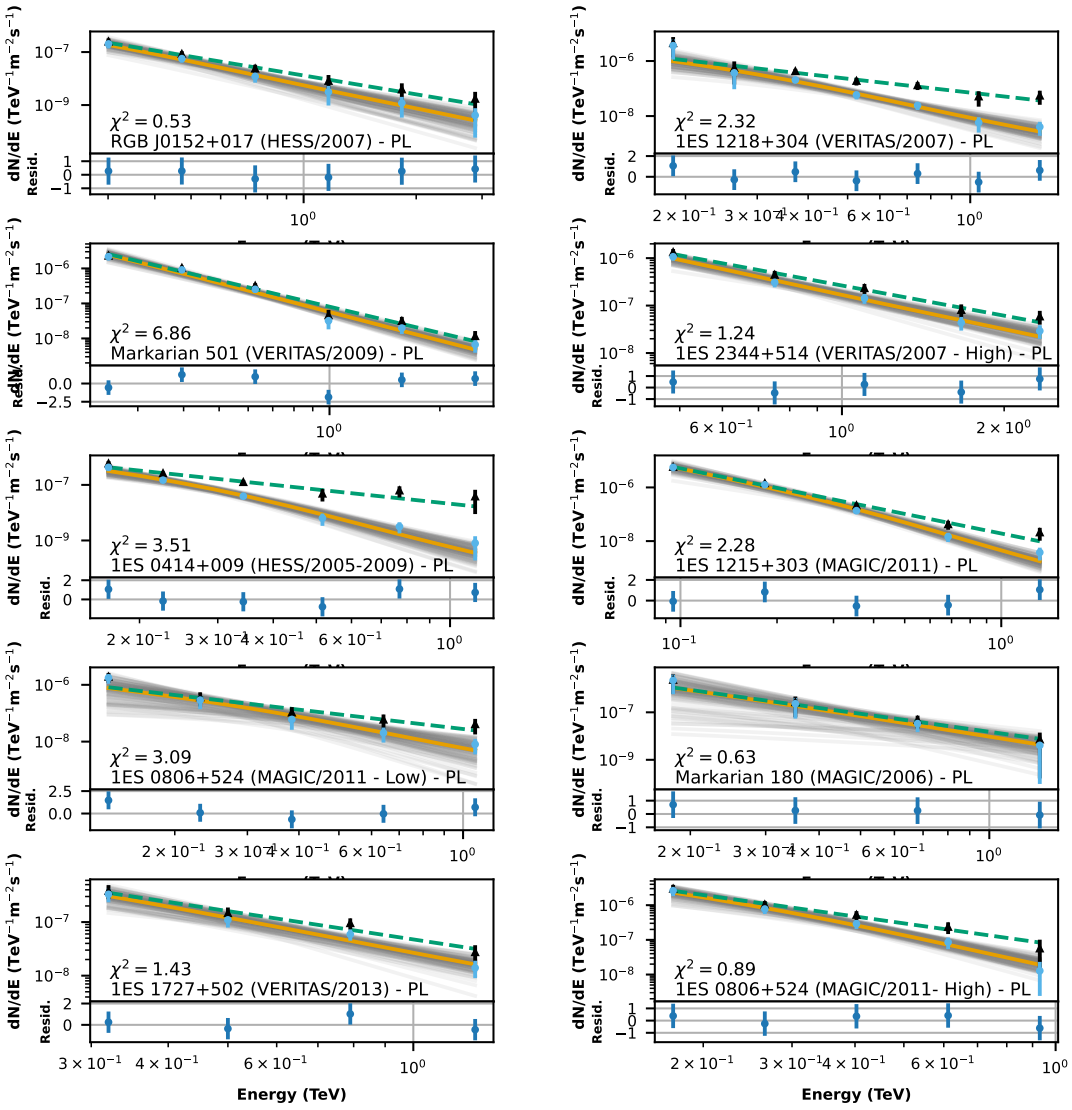

Figure E.5: Continuation of Figure E.4.

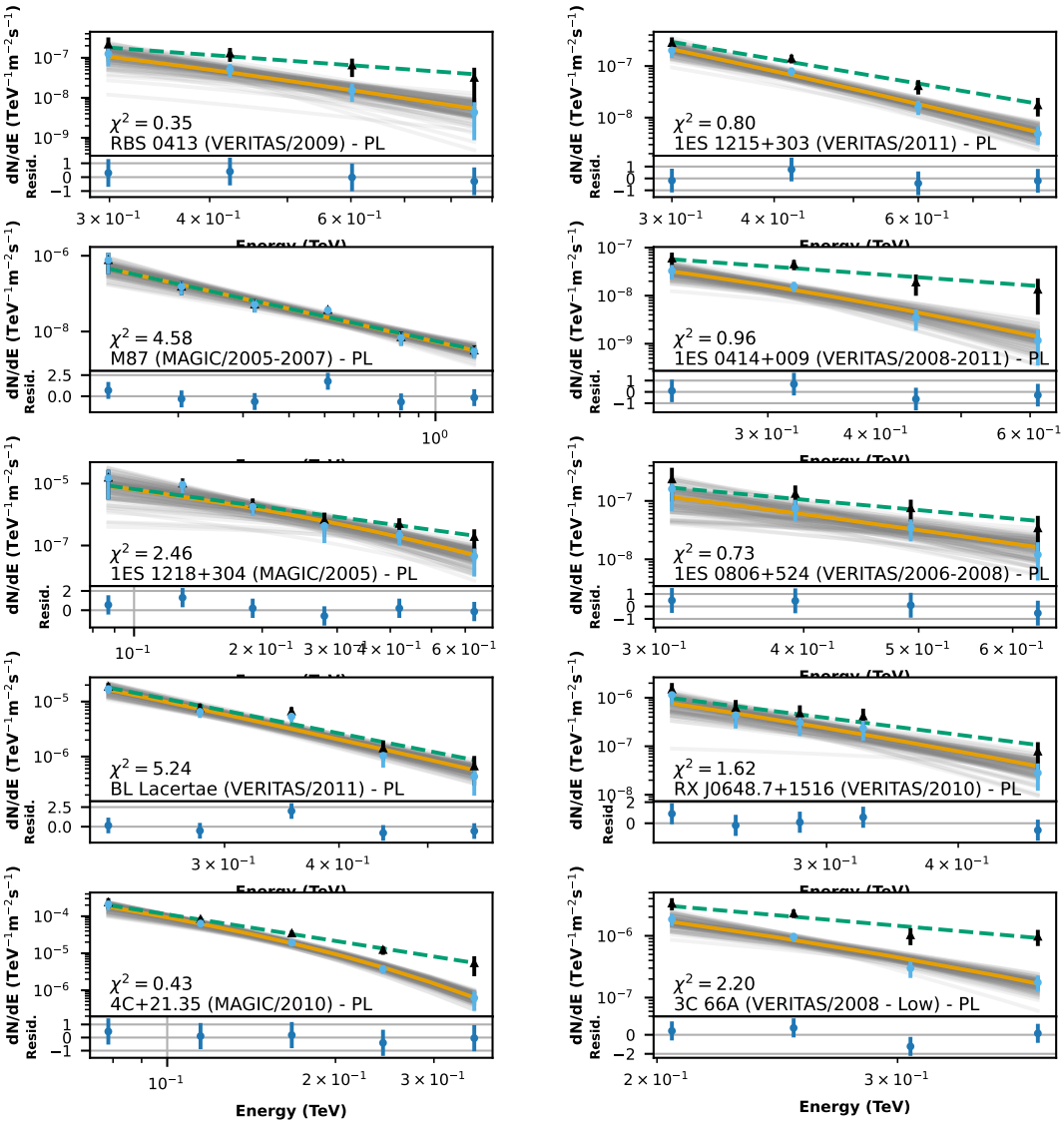

Figure E.6: Continuation of Figure E.5.

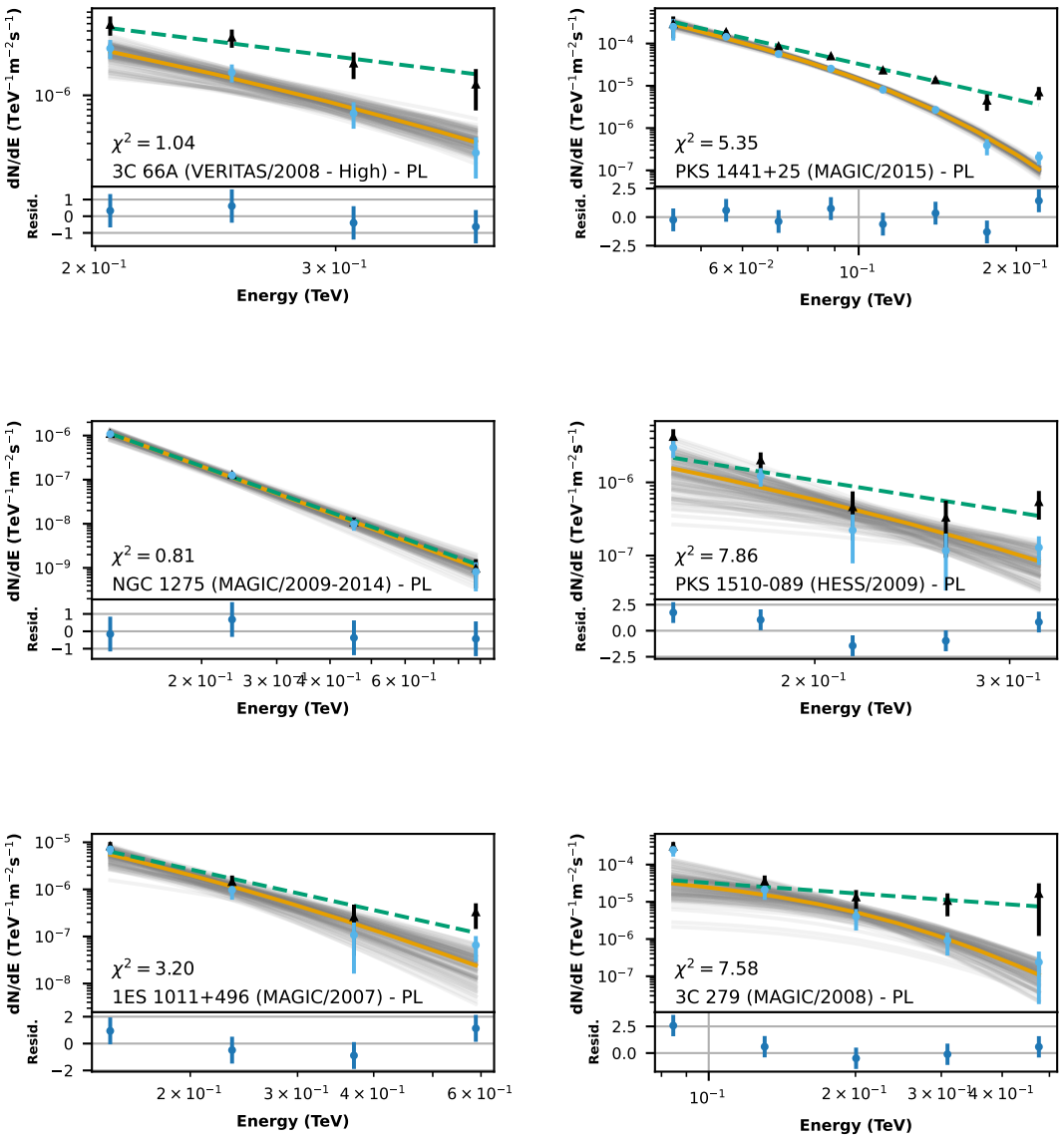

Figure E.7: Continuation of Figure E.6.

E.4 Marginal distributions of source parameters

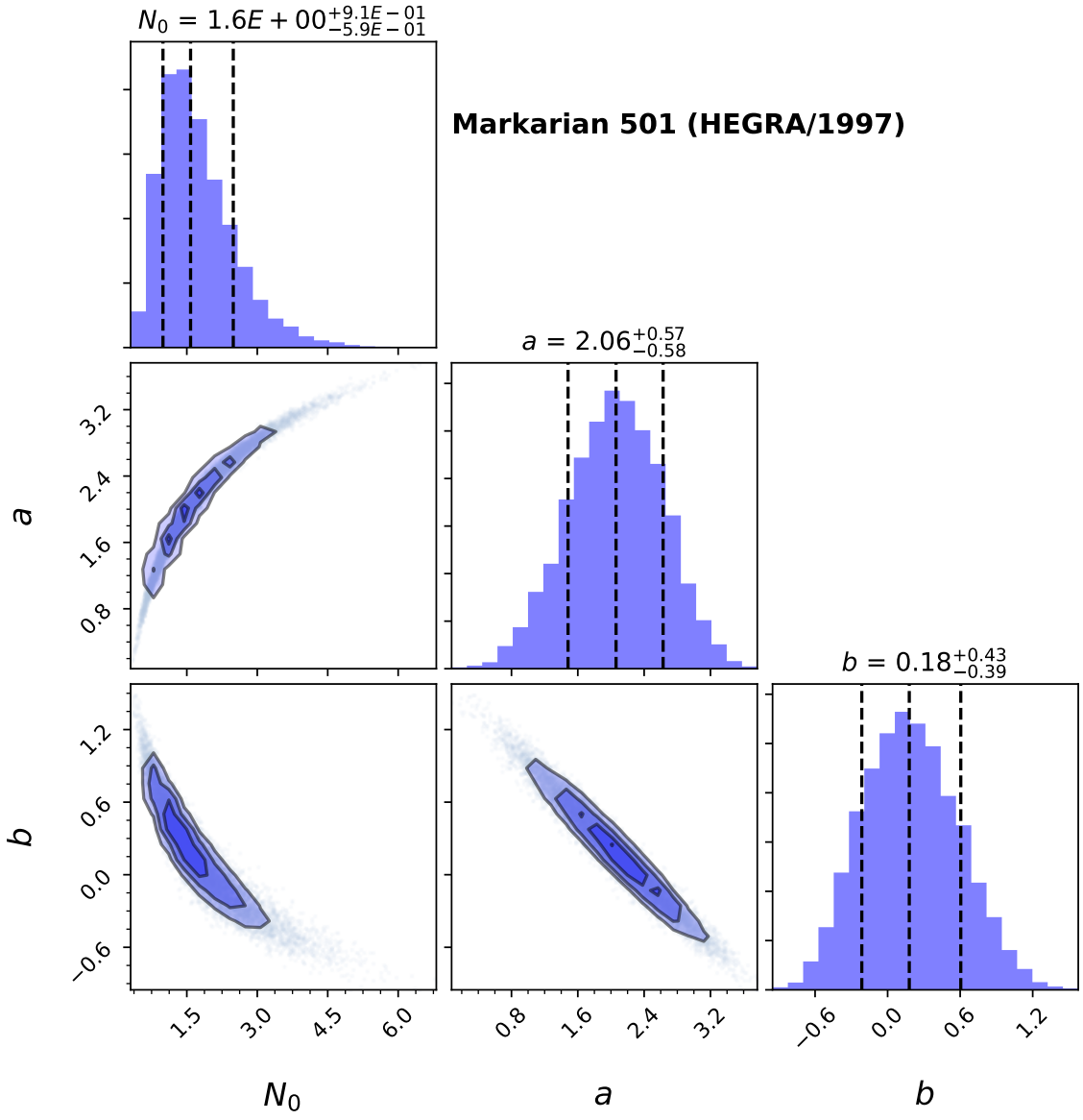

Figure E.8: Marginal distributions of intrinsic source parameters of Markarian 501 (HEGRA/1997) modelled as LP. The vertical dashed lines are the median and a 68% confidence interval, with respective values on top of each histogram.

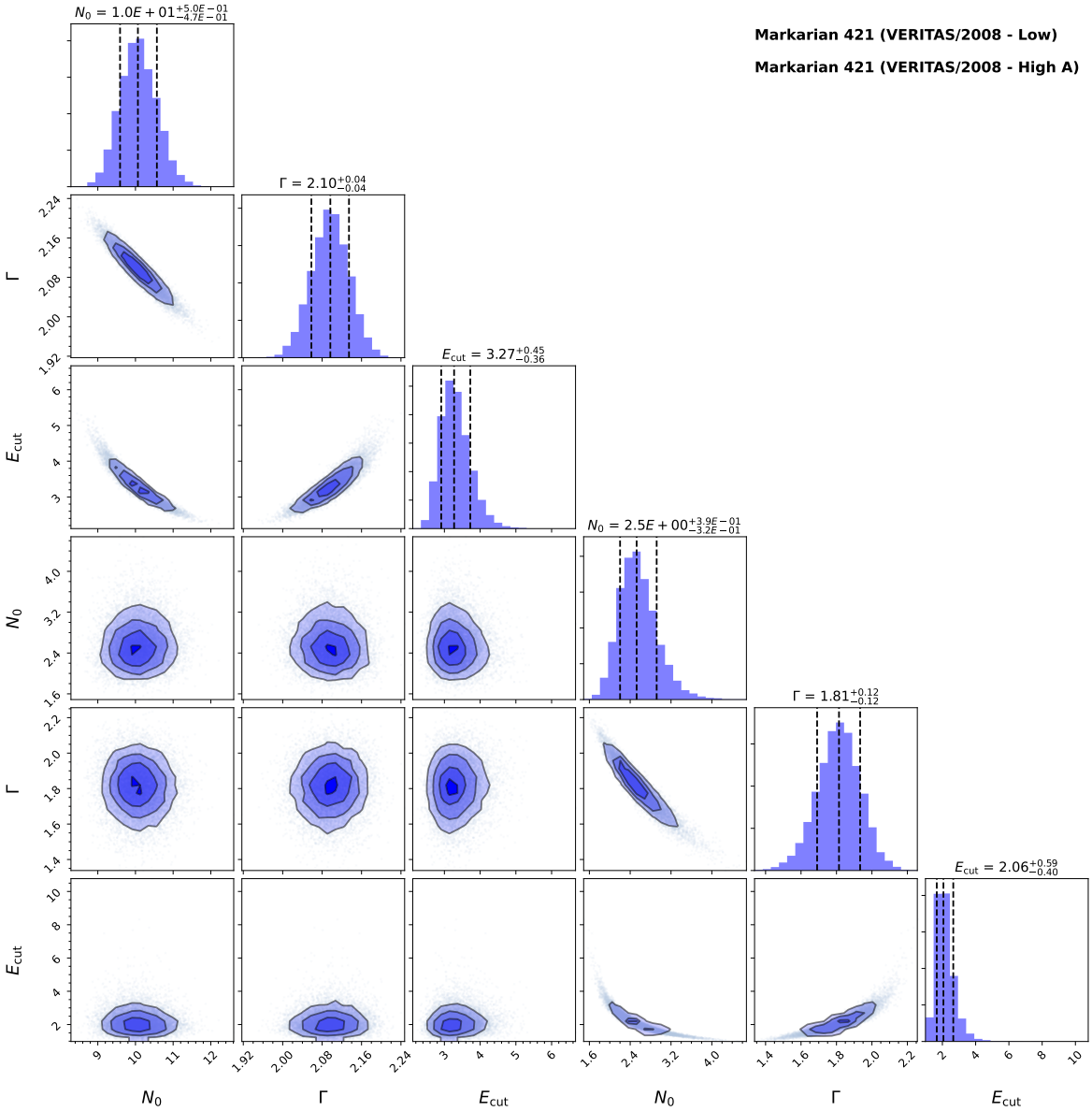

Figure E.9: Marginal distributions of intrinsic source parameters. The text indicates the order of the data, so the upper parameters correspond to the first source and the bottom ones to the last source. These are PLC sources.

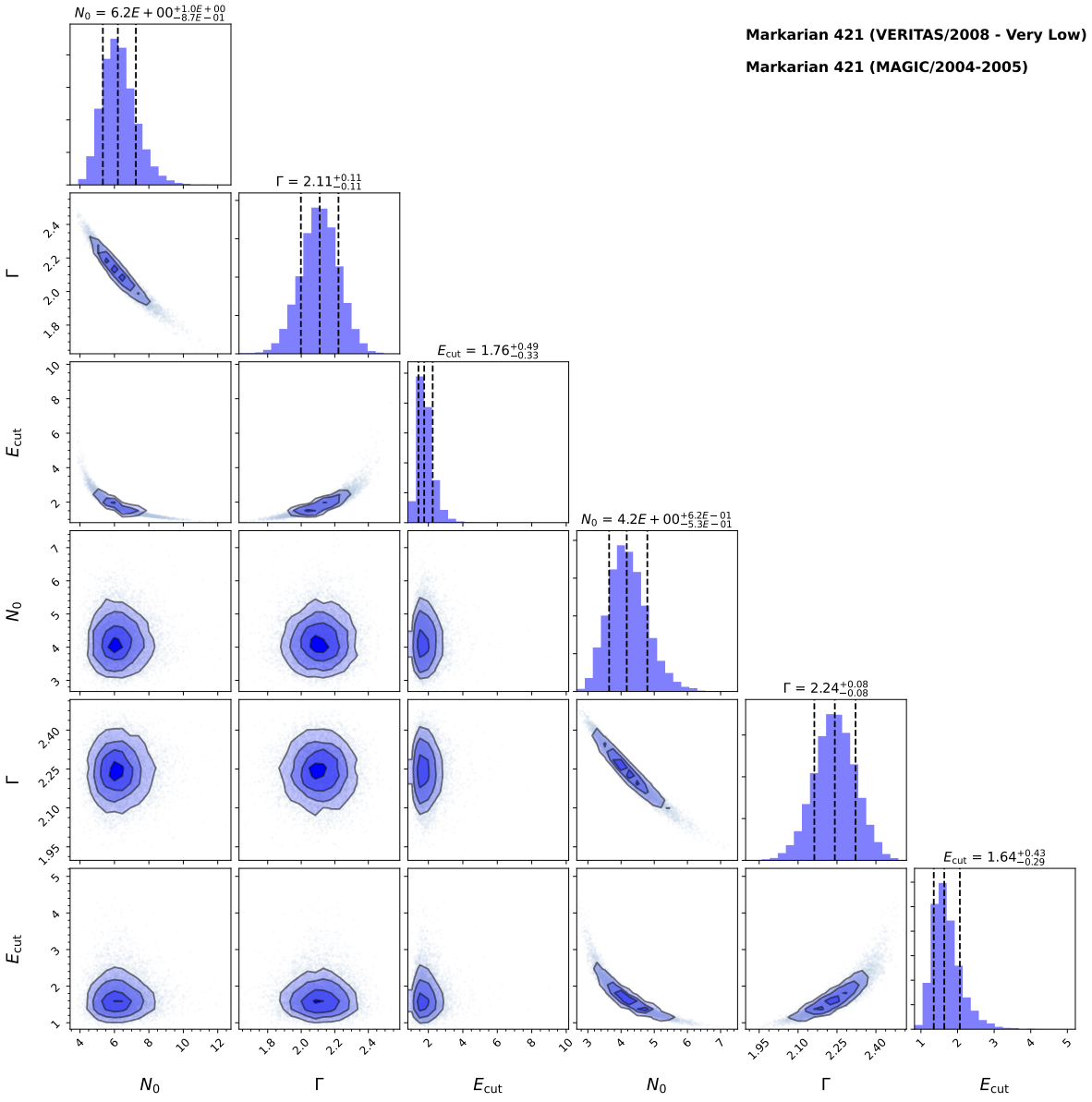

Figure E.10: Same as Figure E.9.

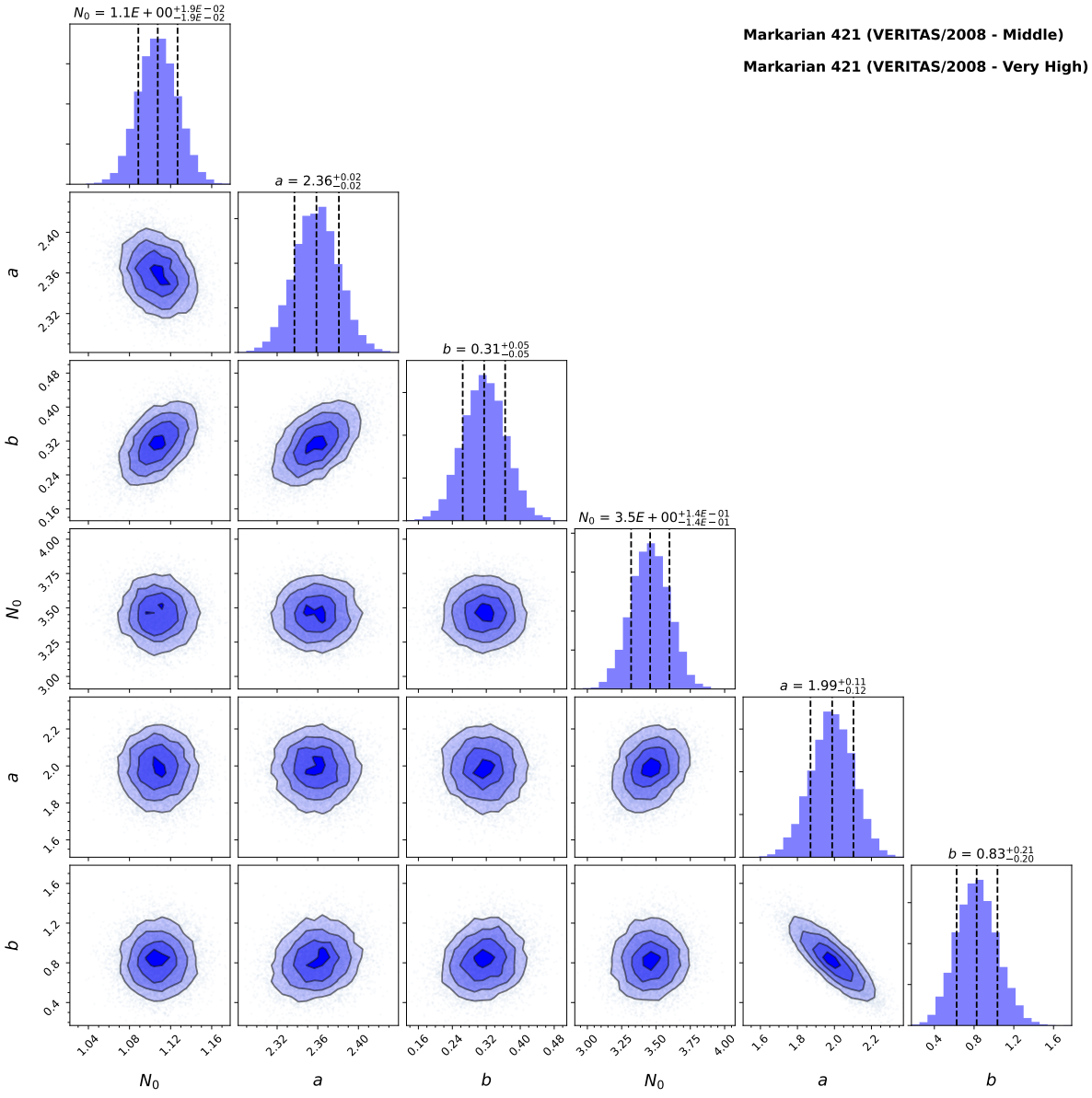

Figure E.11: Same as Figure E.9, but LP sources.

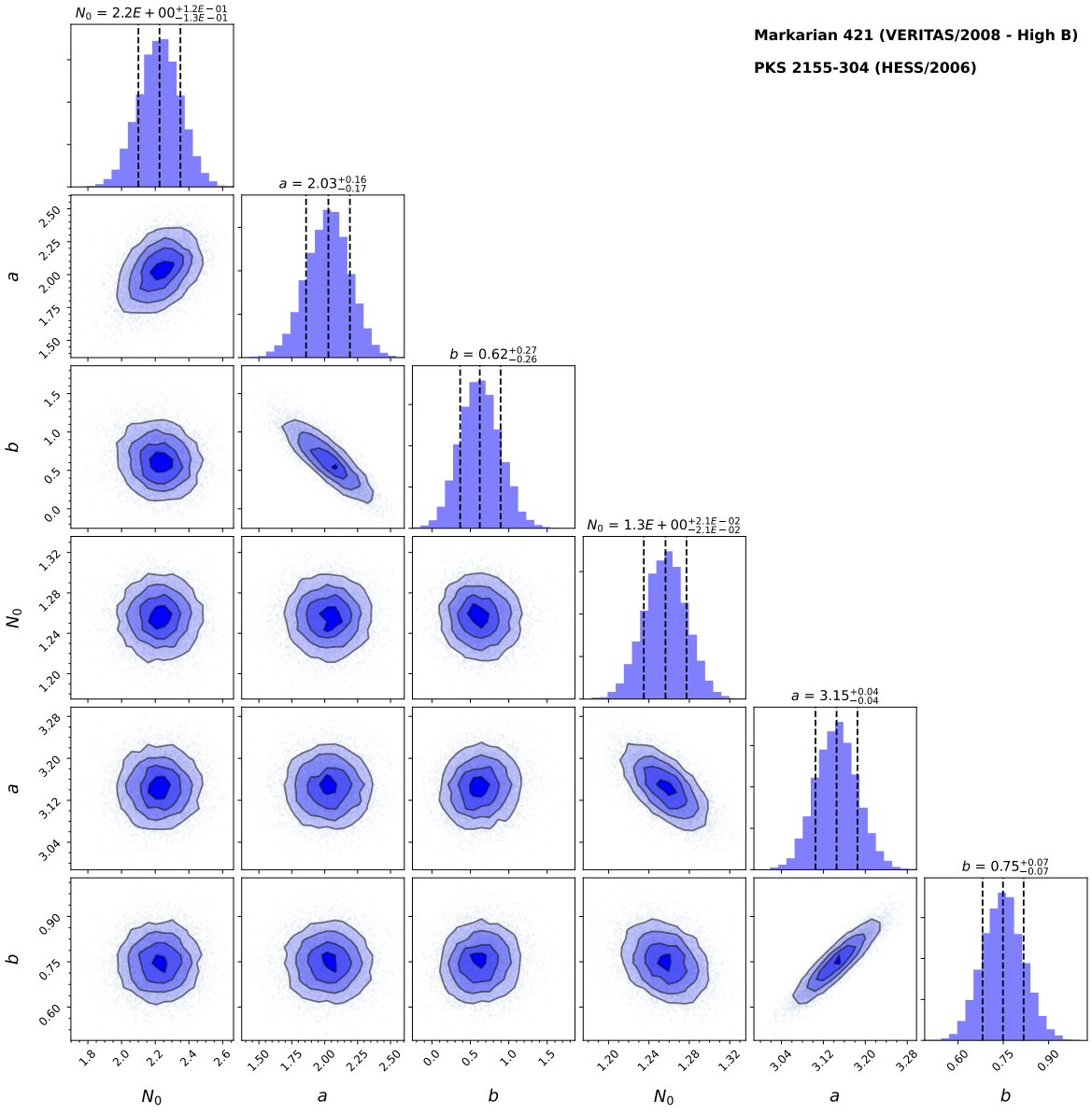

Figure E.12: Same as Figure E.9, but LP sources.

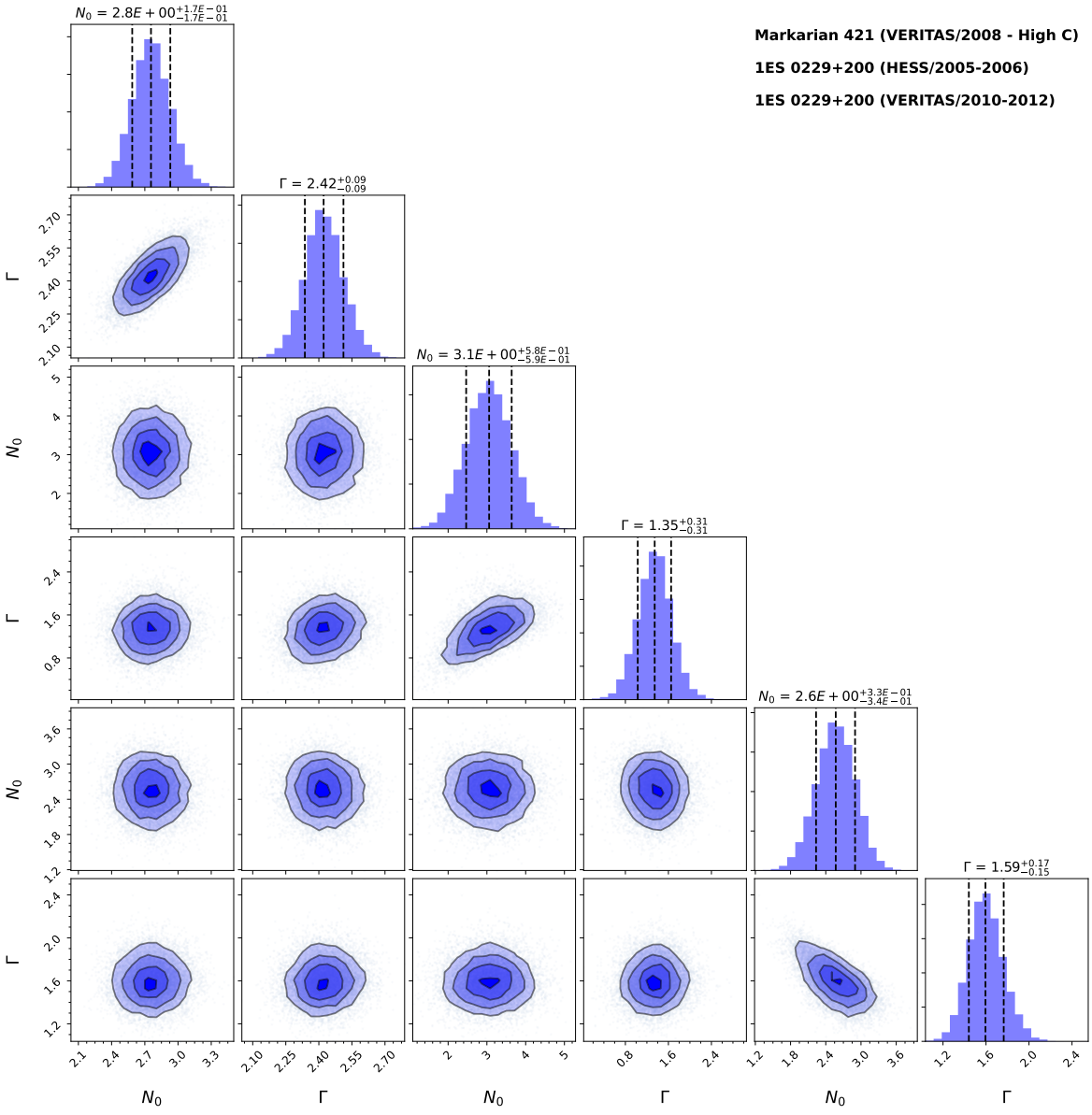

Figure E.13: Same as Figure E.9. These are PL sources.

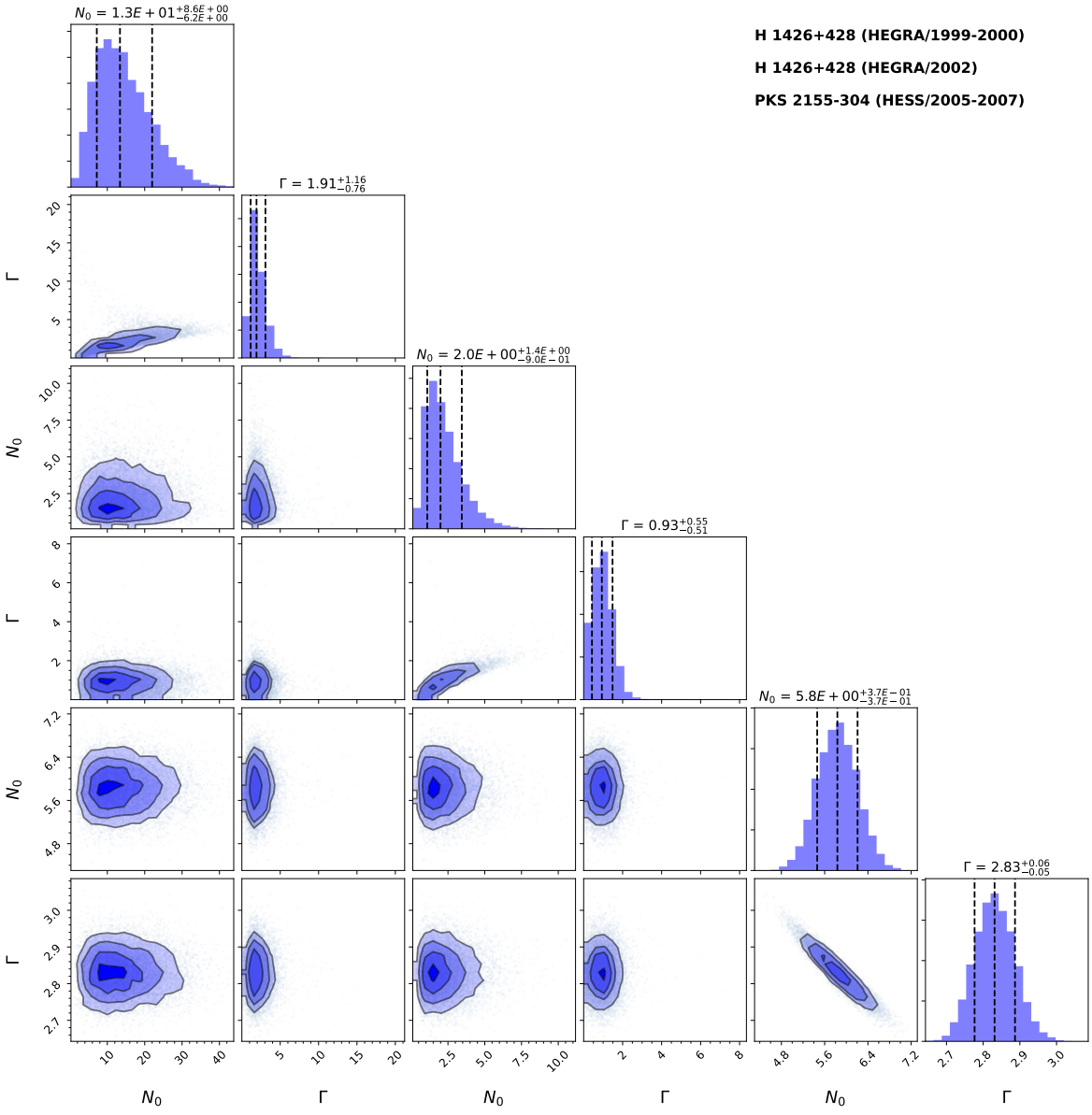

Figure E.14: Same as Figure E.9. These are PL sources.

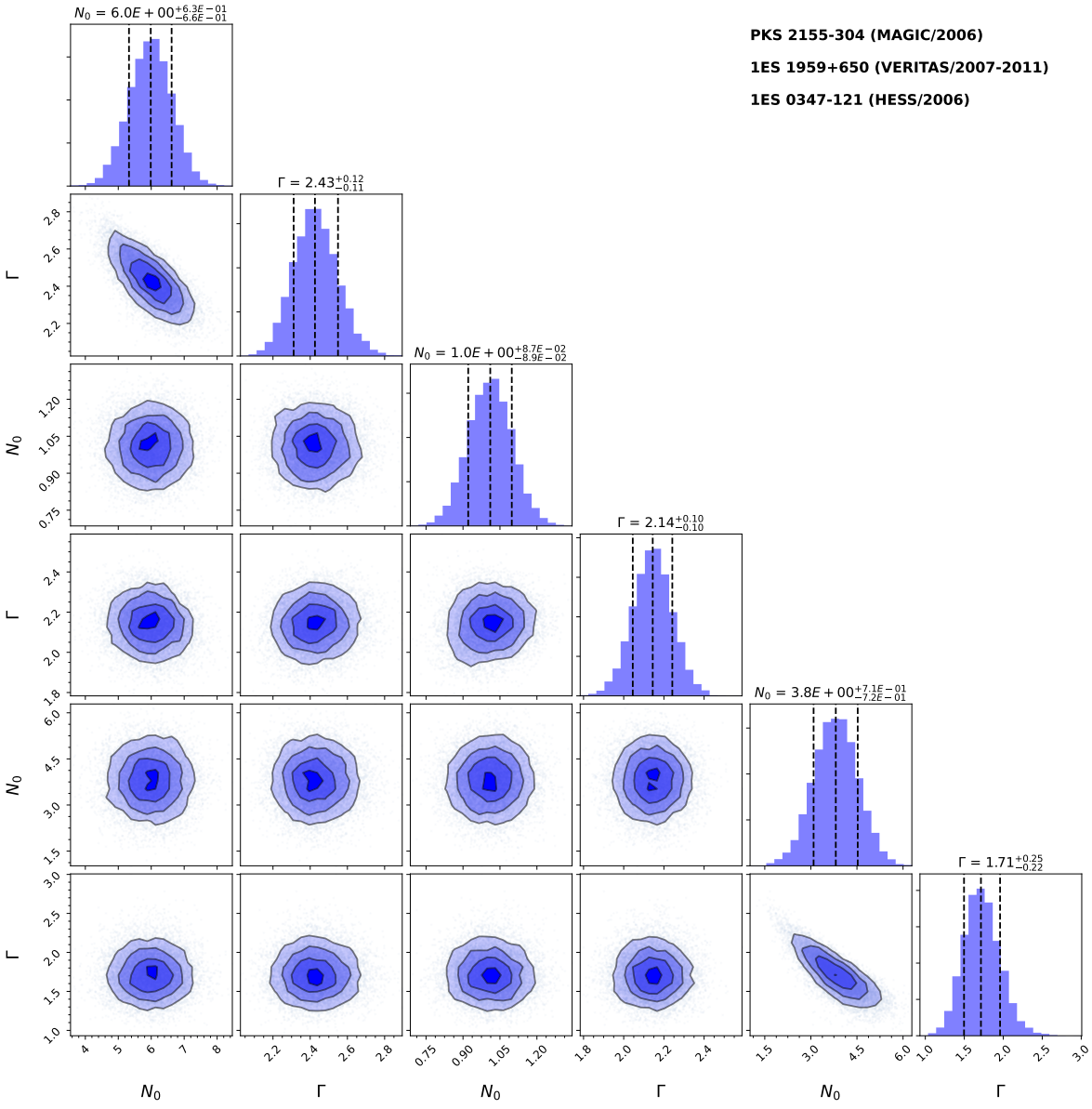

Figure E.15: Same as Figure E.9. These are PL sources.

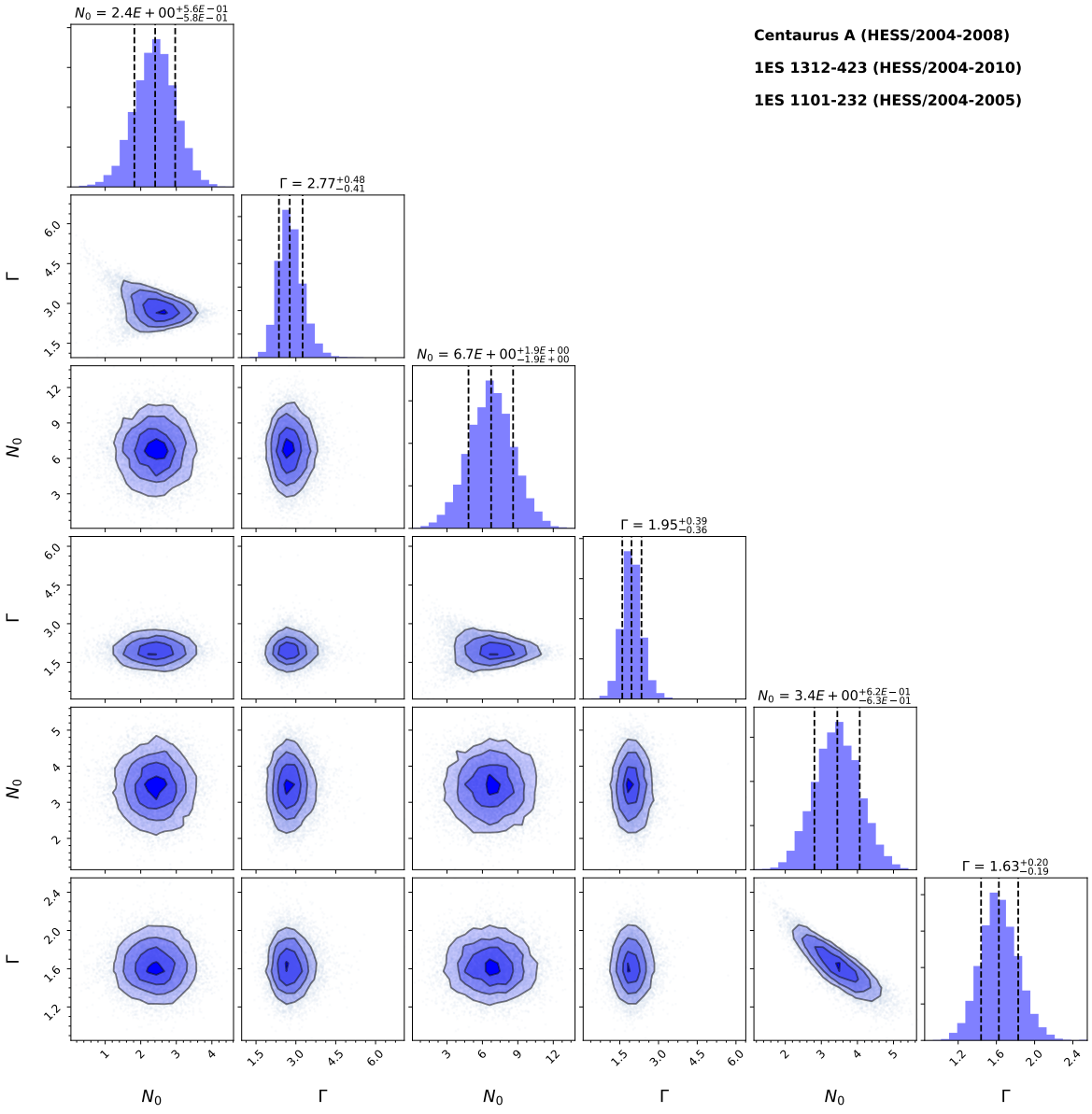

Figure E.16: Same as Figure E.9. These are PL sources.

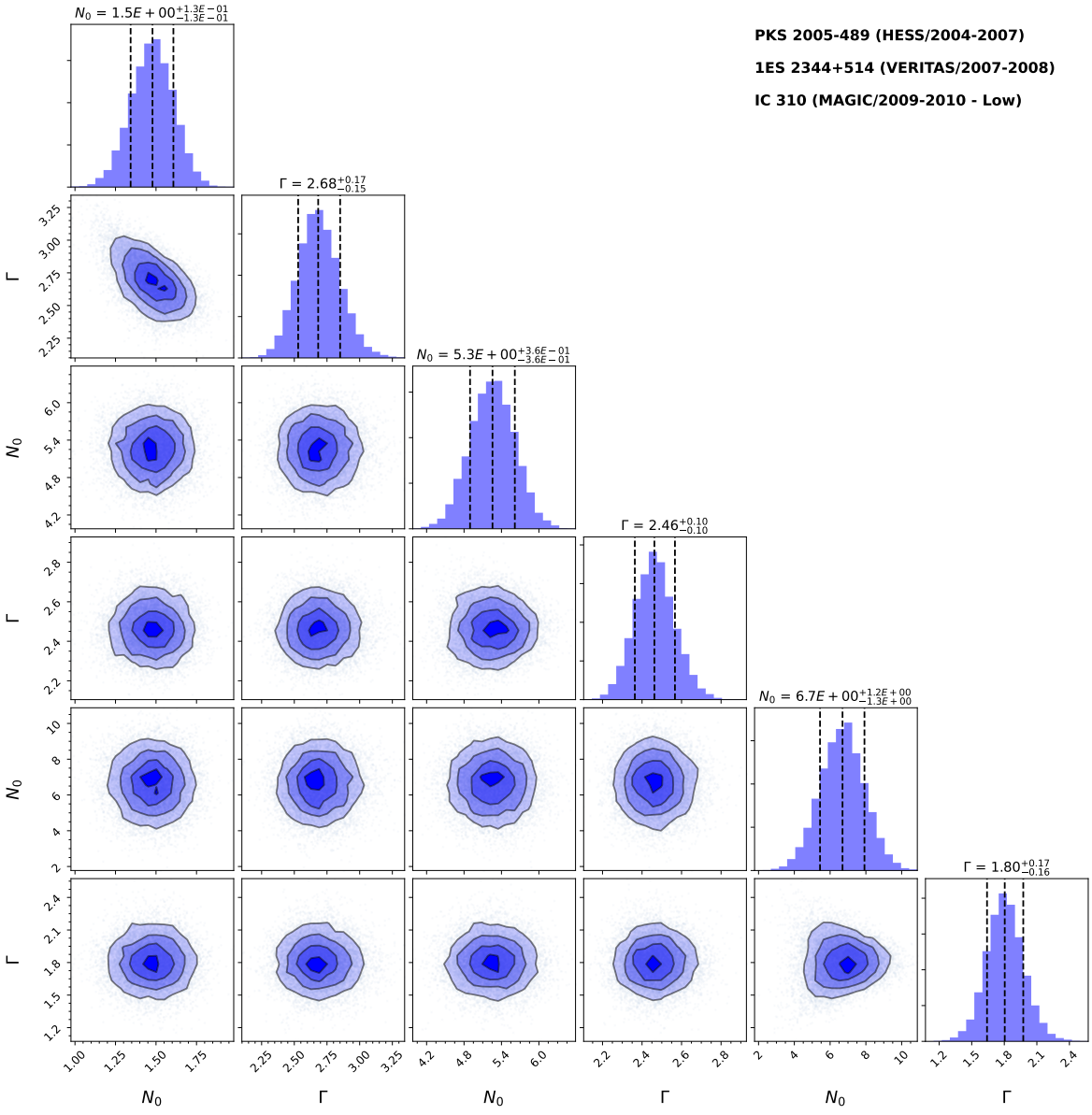

Figure E.17: Same as Figure E.9. These are PL sources.

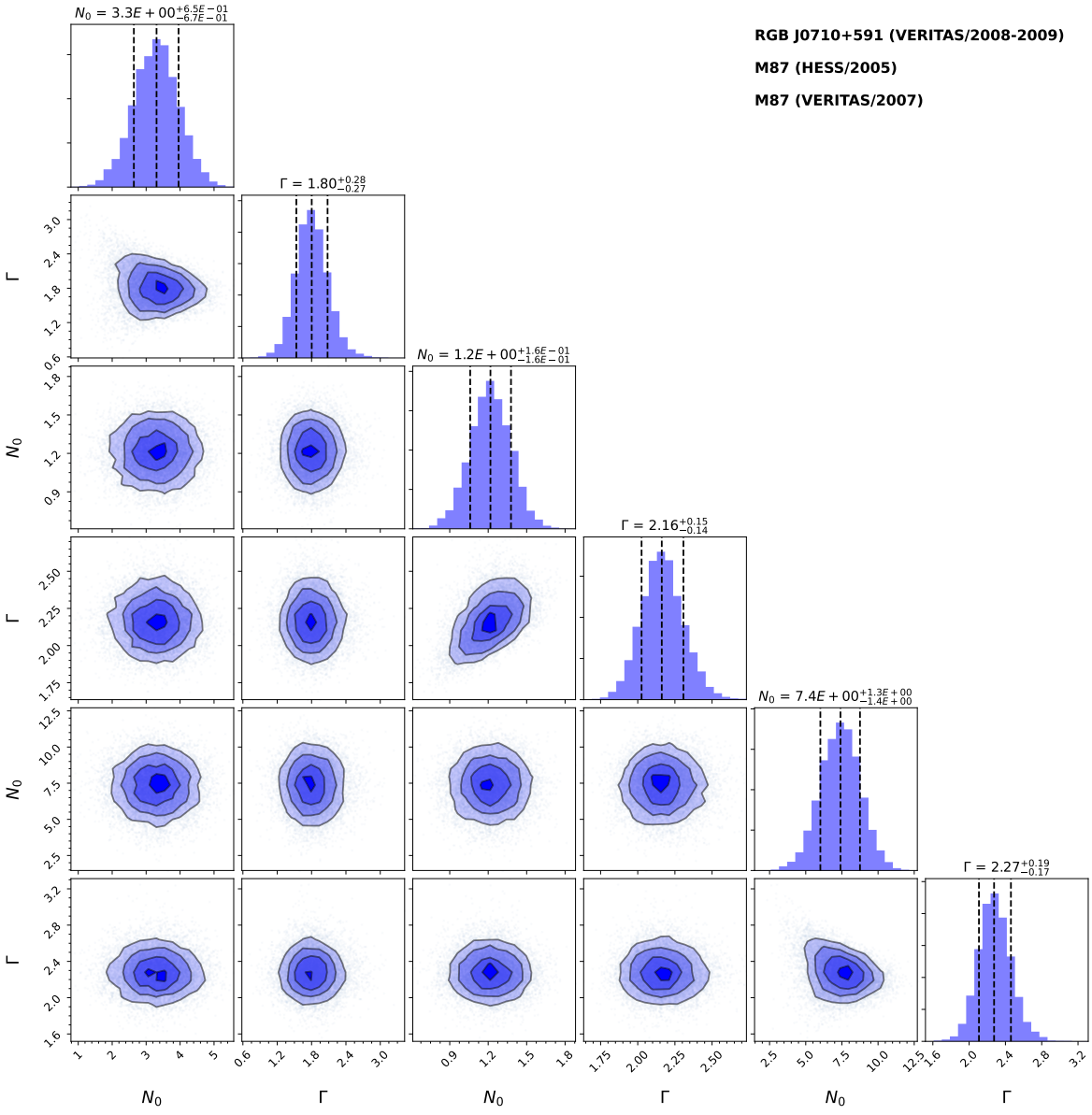

Figure E.18: Same as Figure E.9. These are PL sources.

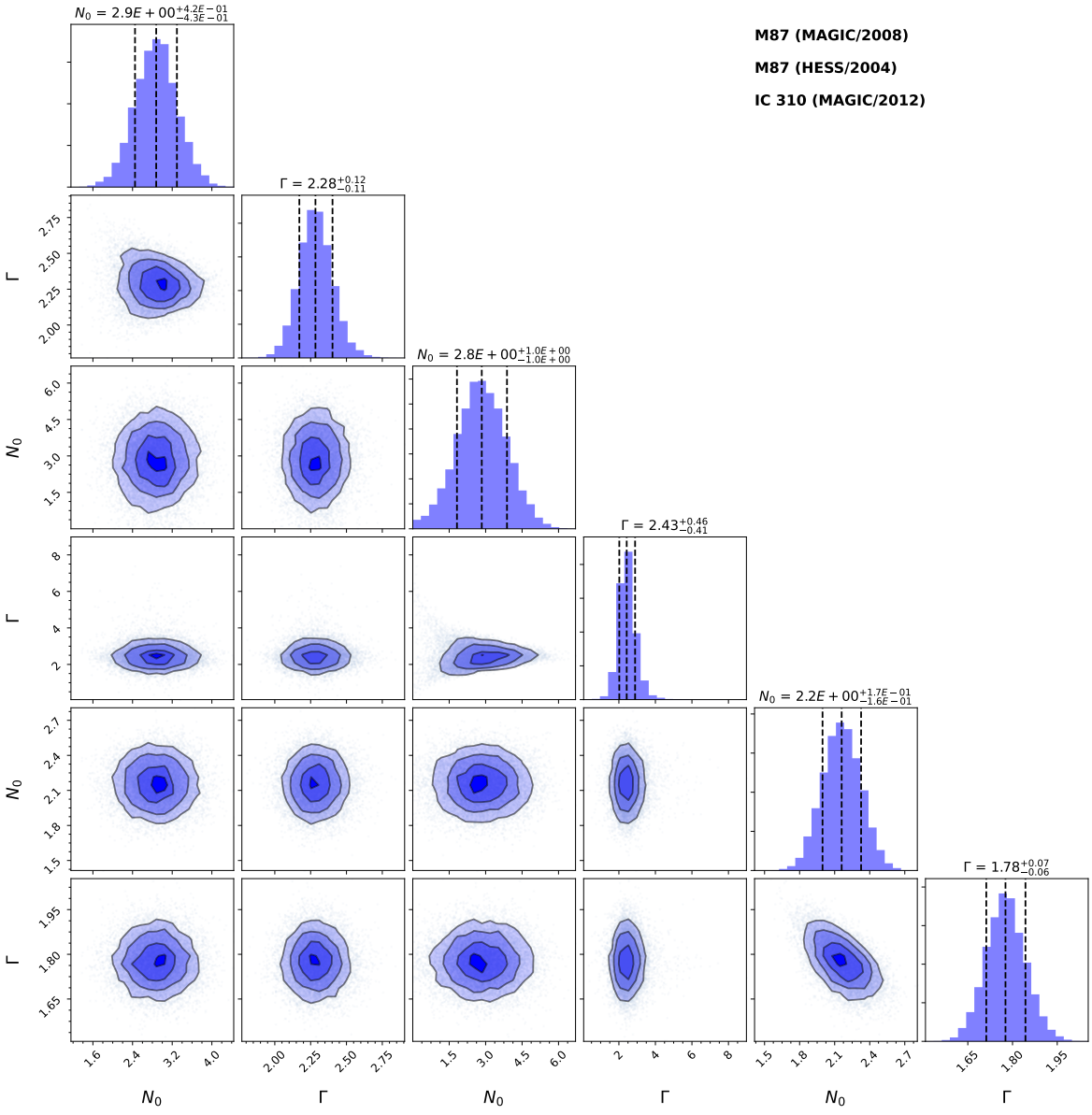

Figure E.19: Same as Figure E.9. These are PL sources.

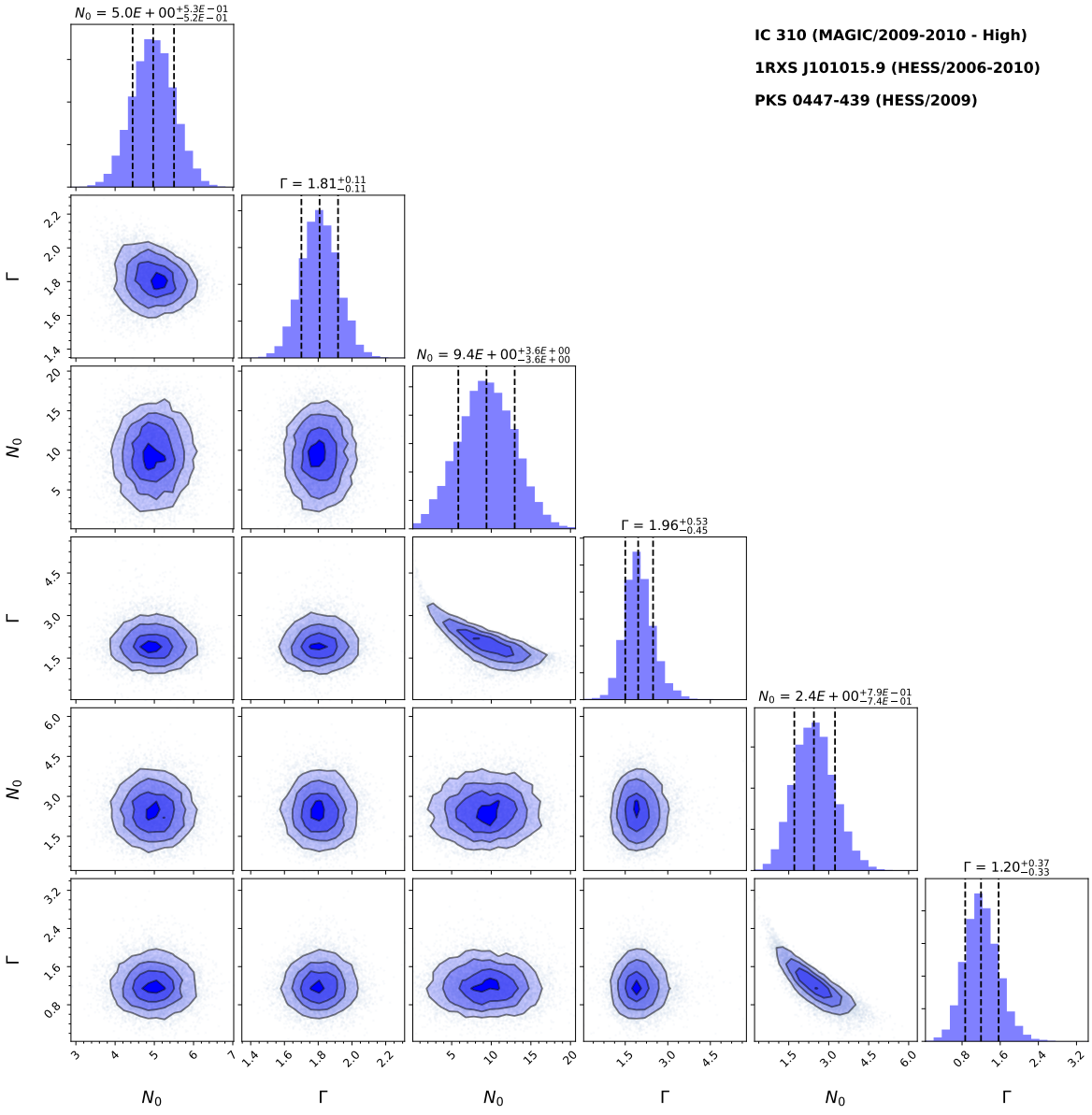

Figure E.20: Same as Figure E.9. These are PL sources.

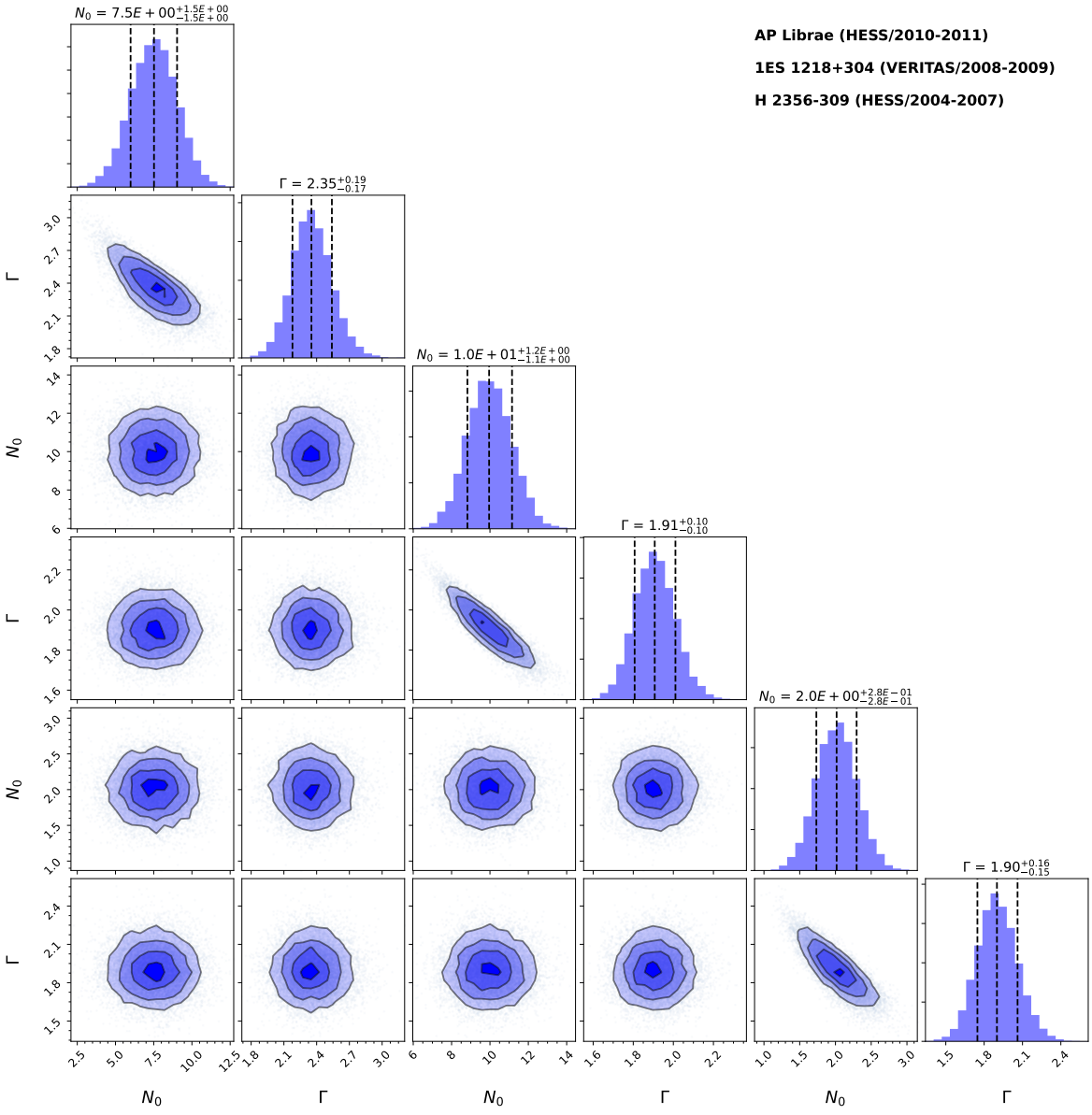

Figure E.21: Same as Figure E.9. These are PL sources.

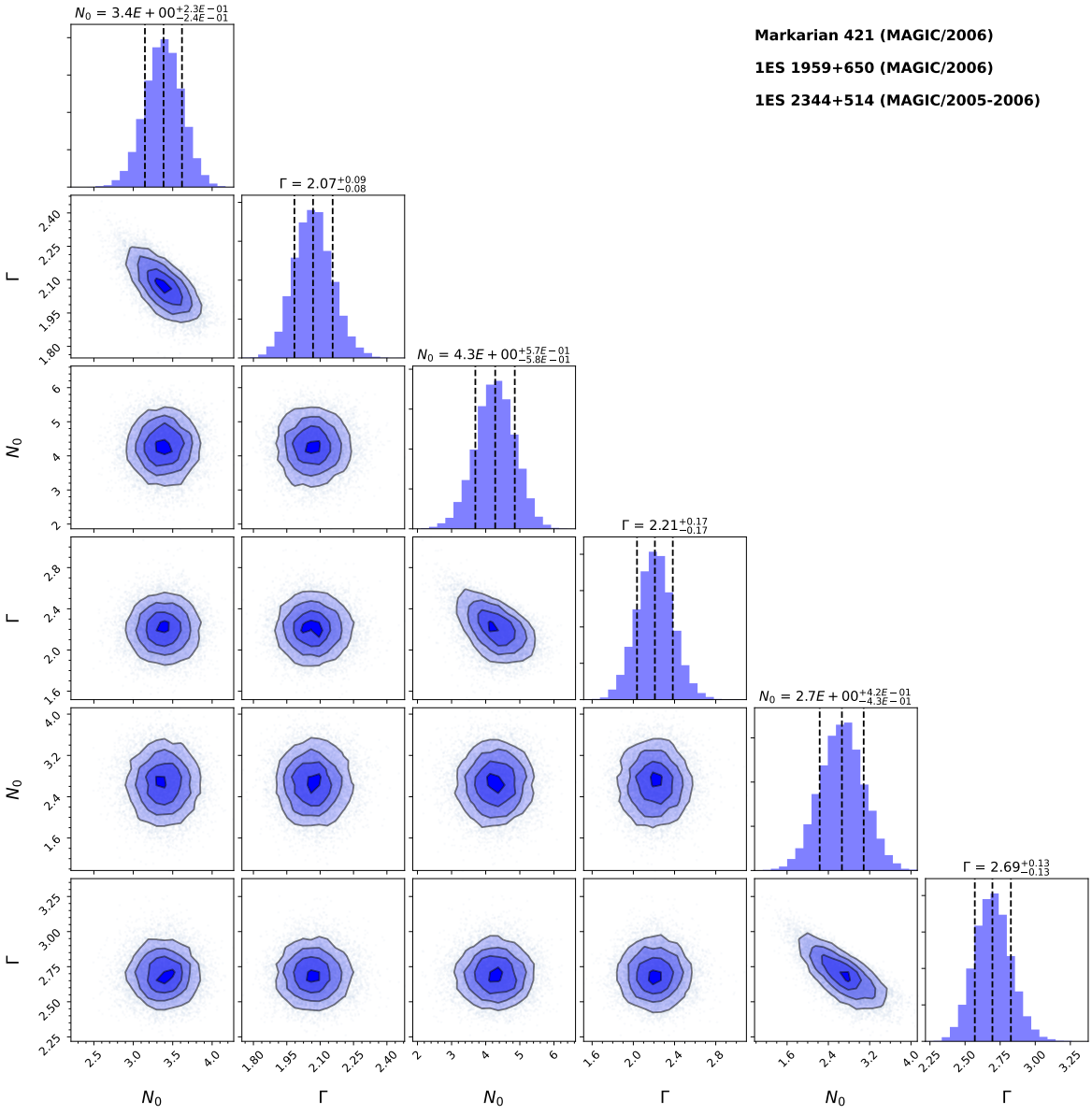

Figure E.22: Same as Figure E.9. These are PL sources.

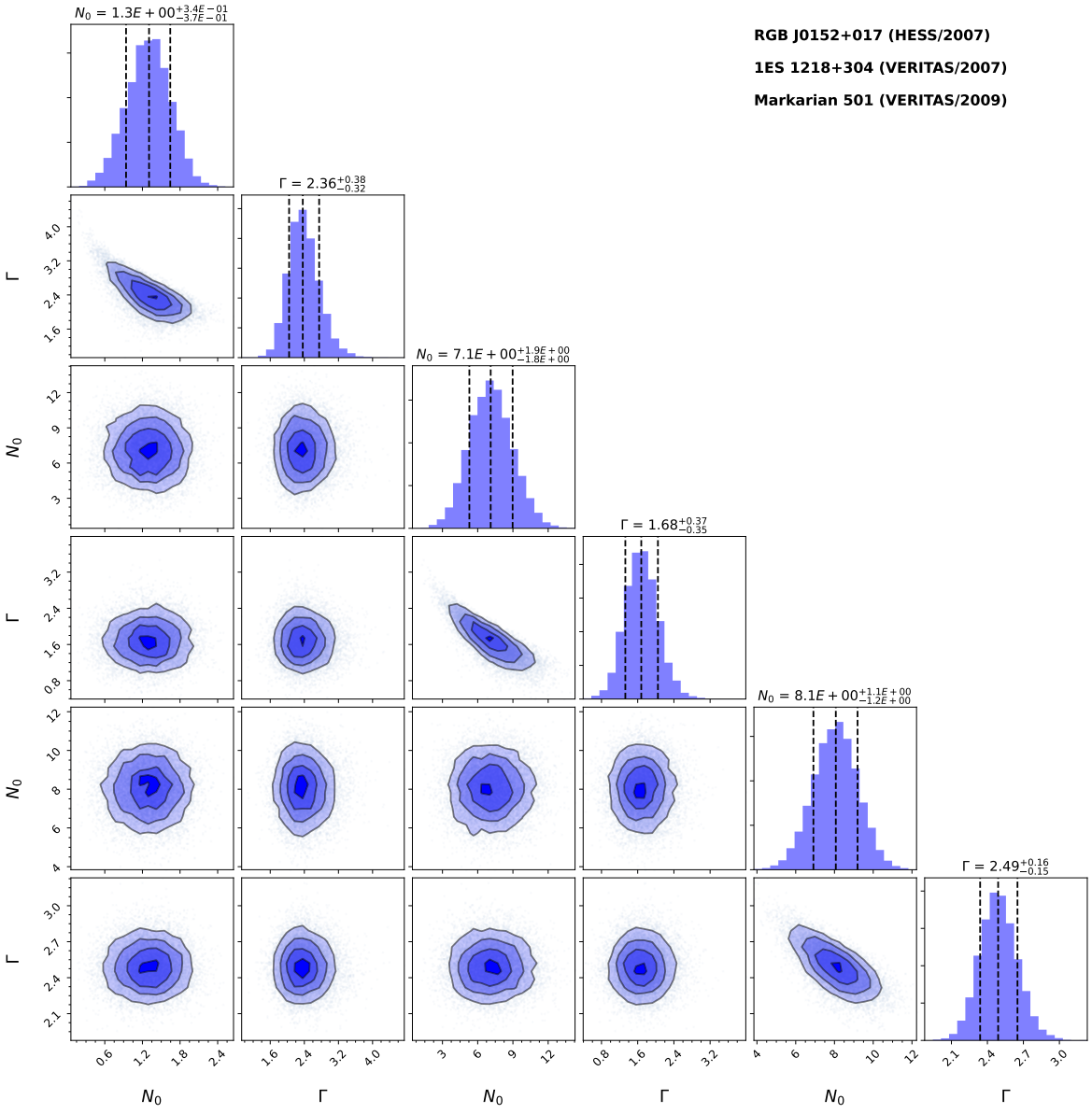

Figure E.23: Same as Figure E.9. These are PL sources.

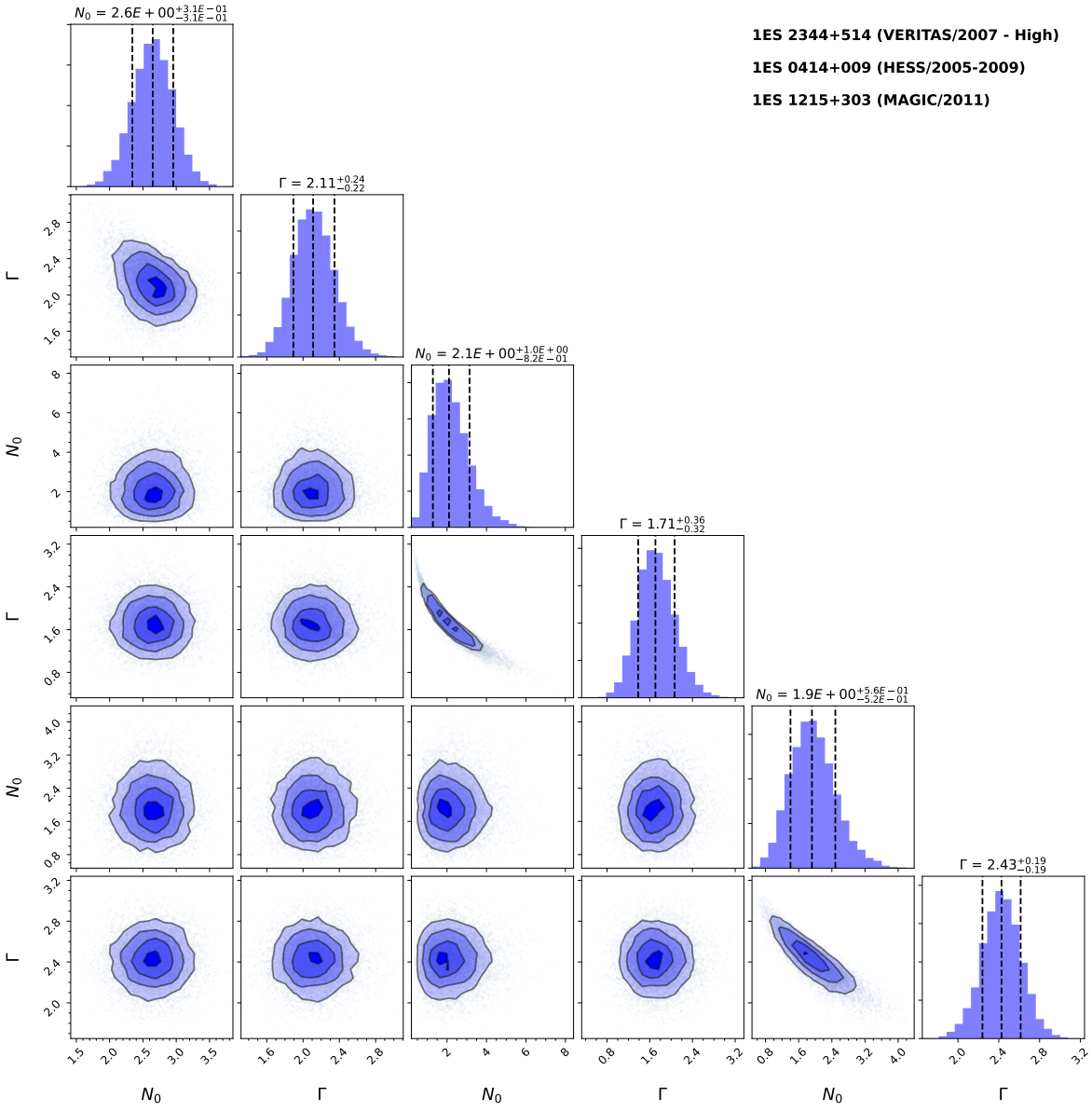

Figure E.24: Same as Figure E.9. These are PL sources.

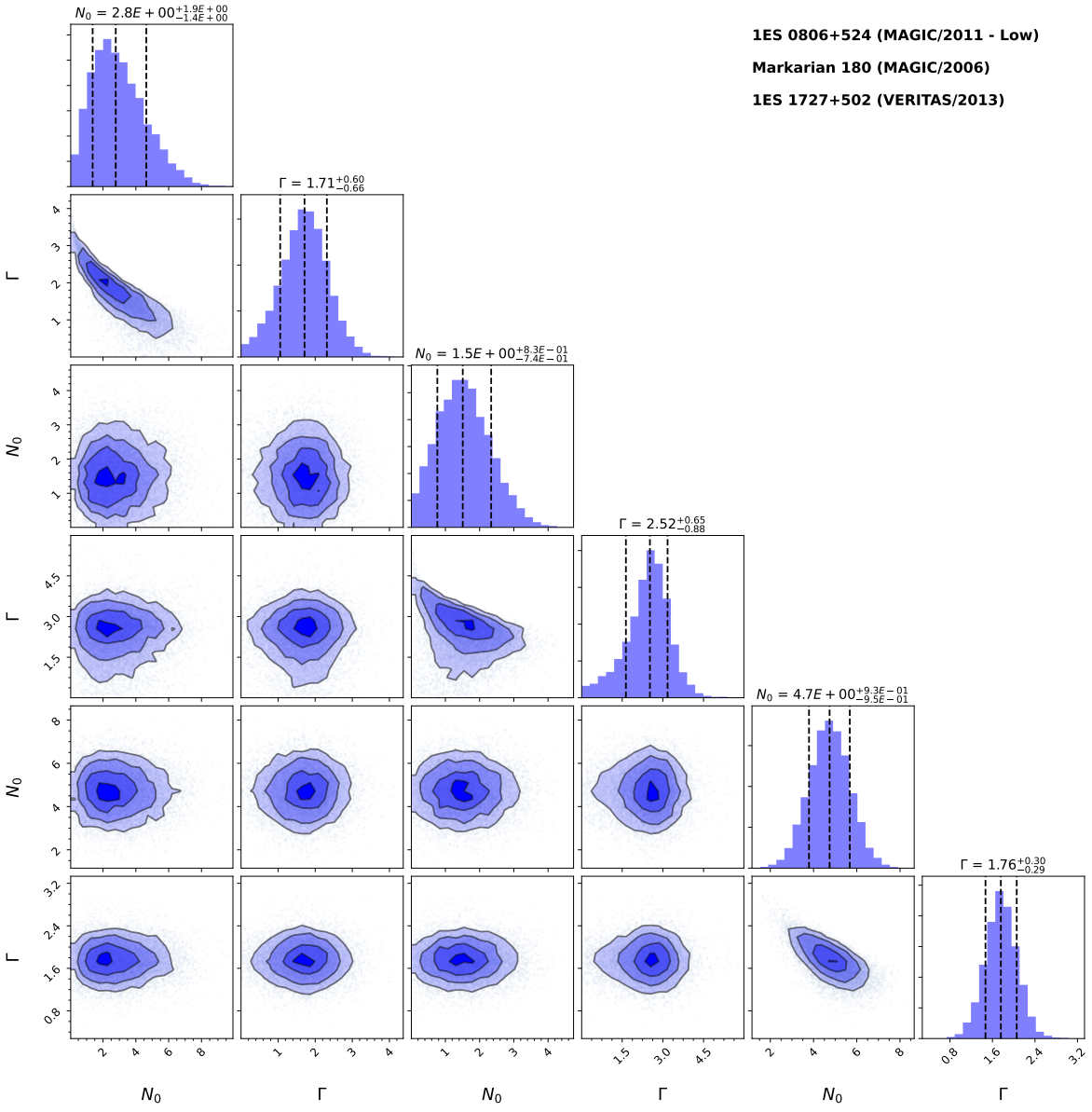

Figure E.25: Same as Figure E.9. These are PL sources.

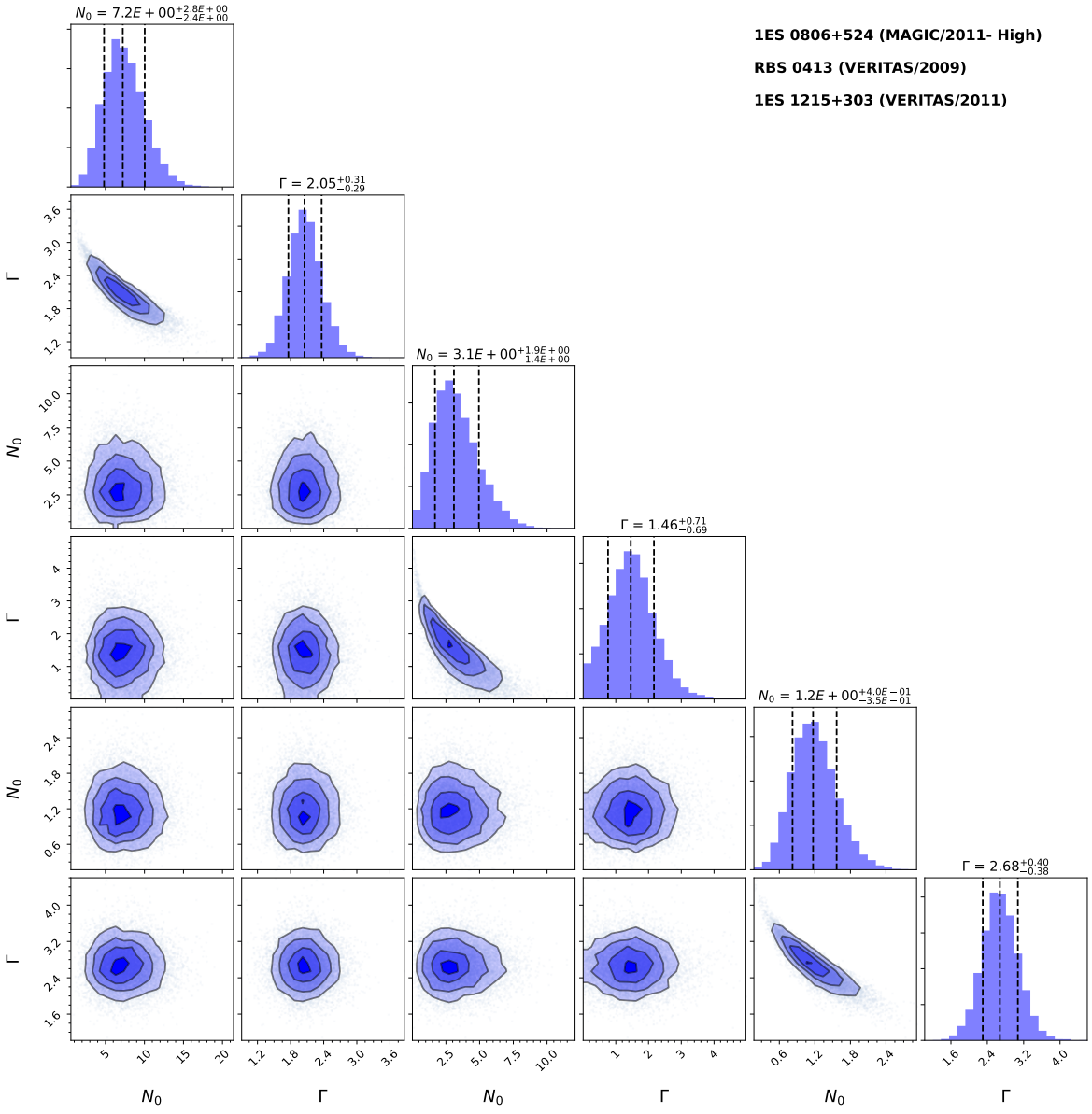

Figure E.26: Same as Figure E.9. These are PL sources.

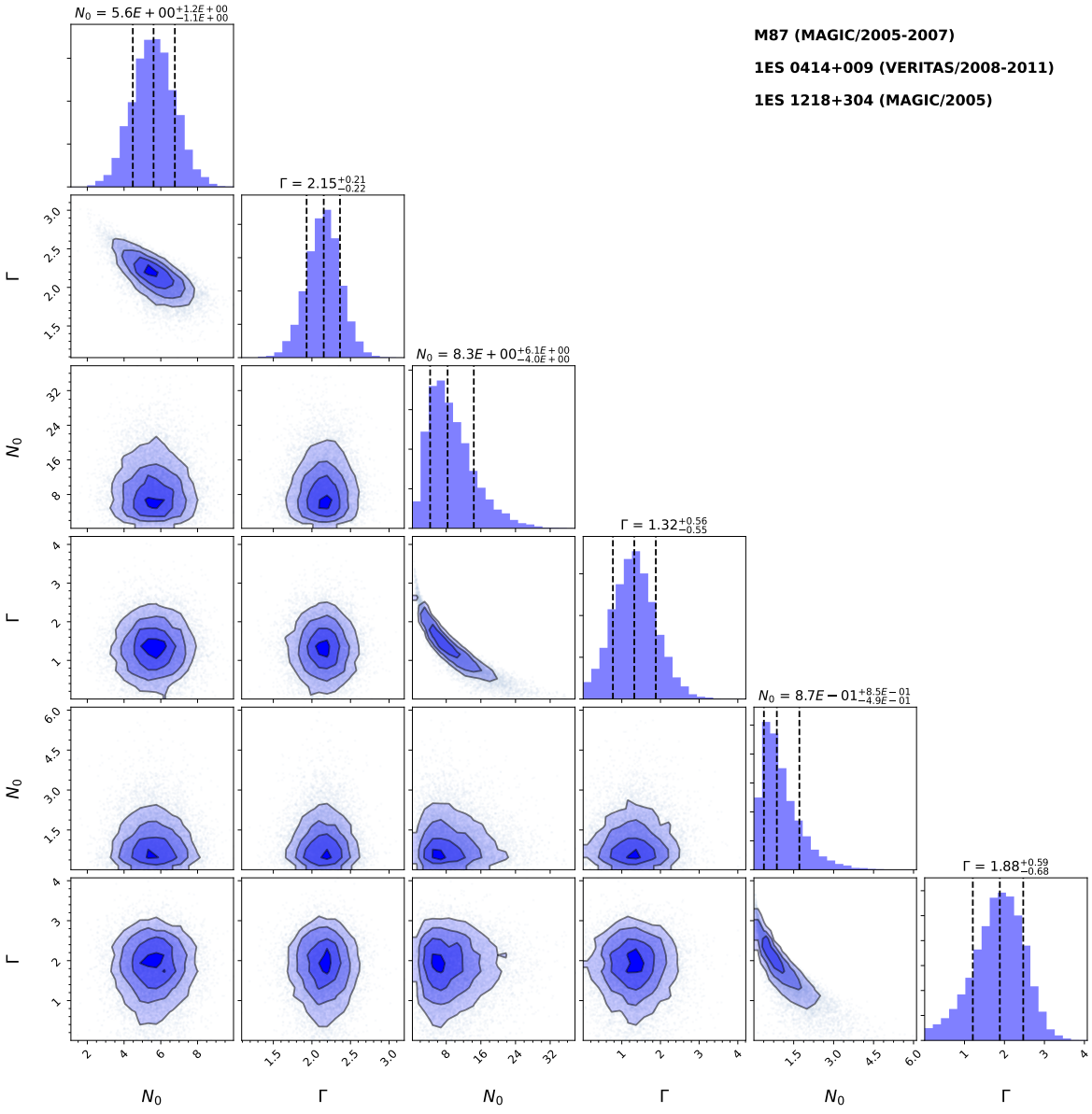

Figure E.27: Same as Figure E.9. These are PL sources.

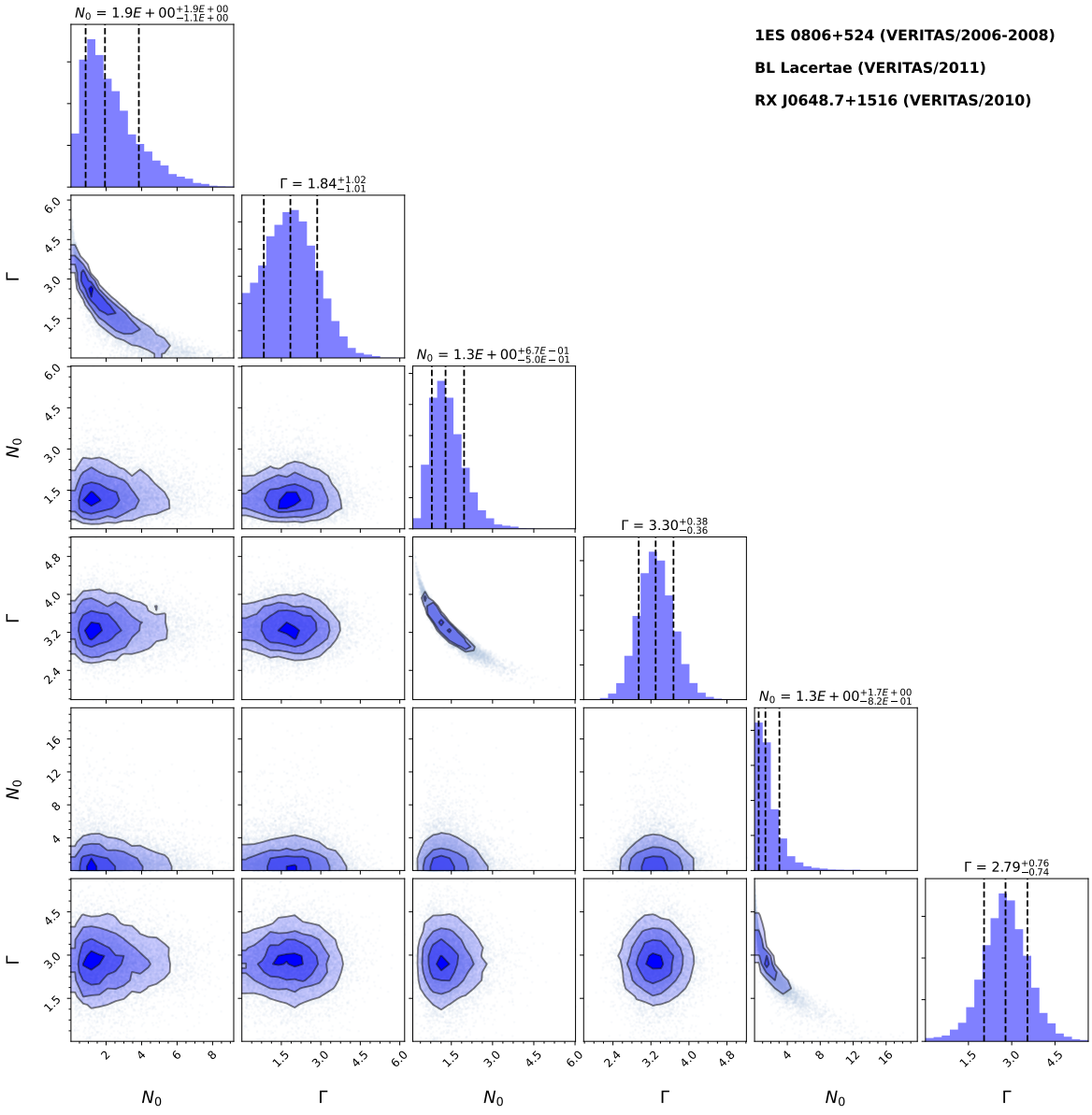

Figure E.28: Same as Figure E.9. These are PL sources.

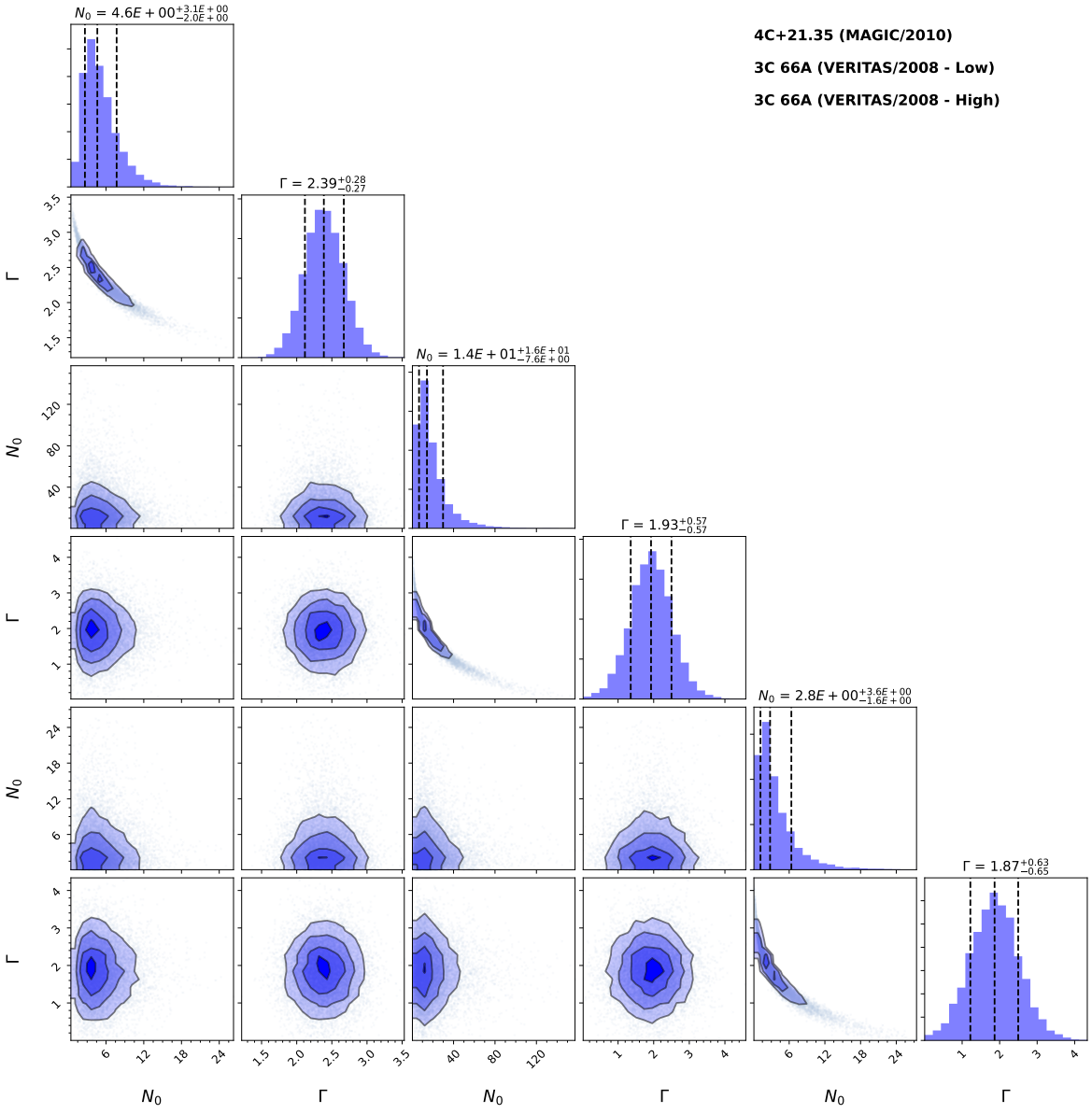

Figure E.29: Same as Figure E.9. These are PL sources.

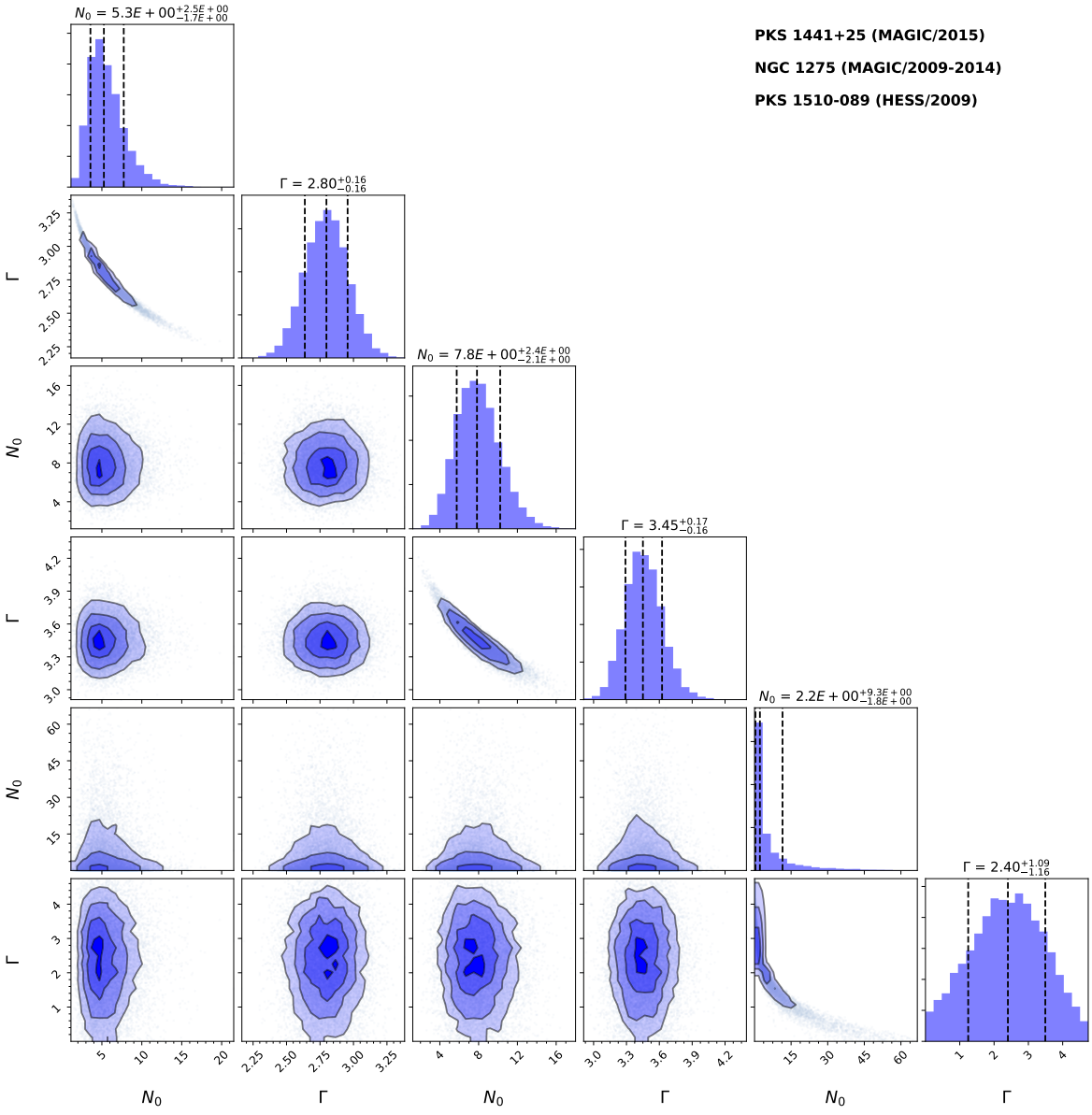

Figure E.30: Same as Figure E.9. These are PL sources.

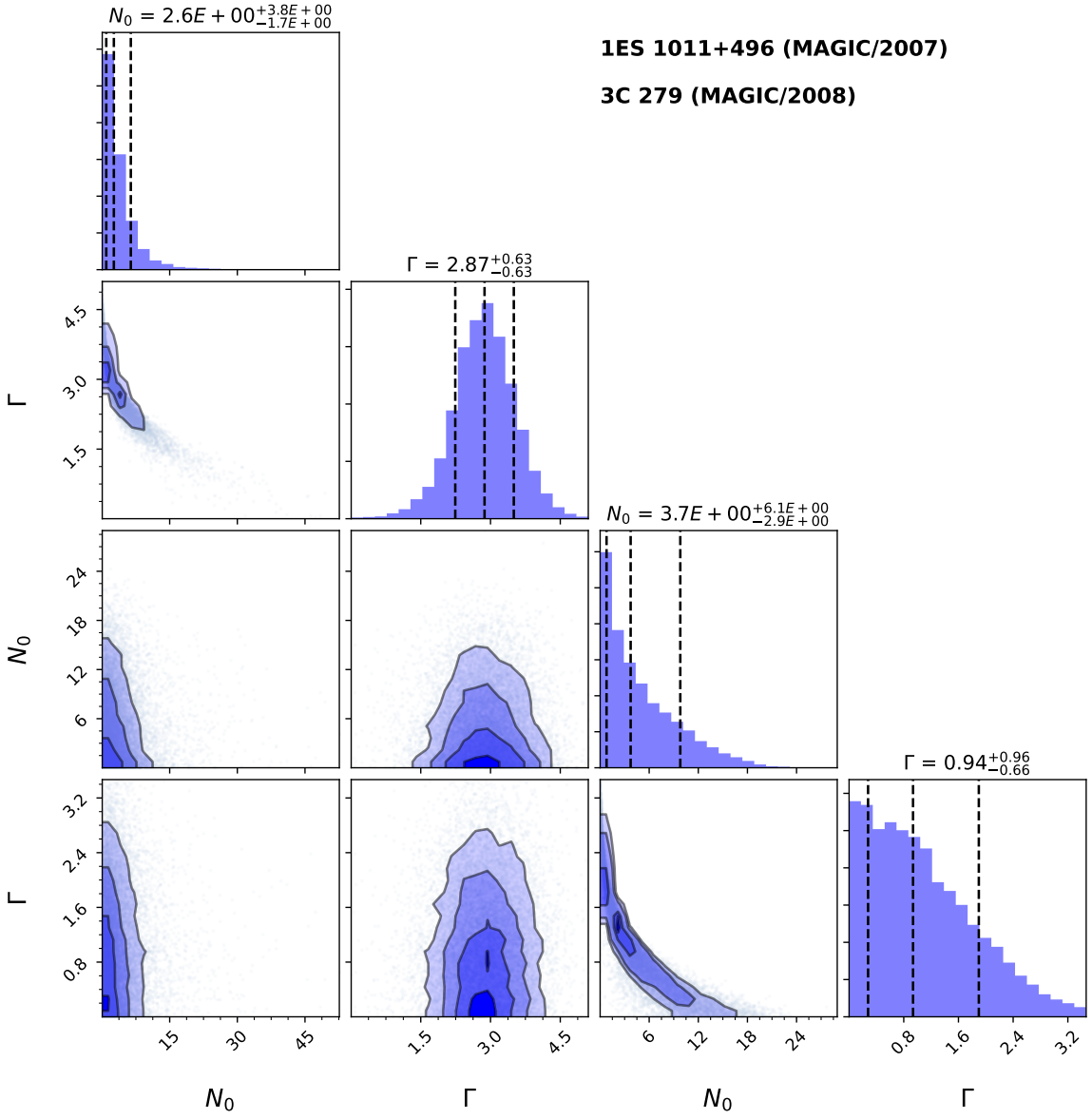

Figure E.31: Same as Figure E.9. These are PL sources.

Bibliography

Abdalla H., et al., 2021, J. Cosmology Astropart. Phys., 2021, 048

Abdo A. A., et al., 2011, ApJ, 726, 43

Abeysekara A. U., et al., 2019, ApJ, 885, 150

Abramowski A., et al., 2013, Phys. Rev. D, 88, 102003

Acciari V. A., et al., 2008, ApJ, 679, 397

Acciari V., et al., 2009a, ApJ, 690, L126

Acciari V. A., et al., 2009b, ApJ, 695, 1370

Acciari V. A., et al., 2009c, ApJ, 703, 169

Acciari V. A., et al., 2010a, ApJ, 709, L163

Acciari V. A., et al., 2010b, ApJ, 715, L49

Acciari V. A., et al., 2011a, ApJ, 729, 2

Acciari V. A., et al., 2011b, ApJ, 738, 25

Acciari V. A., et al., 2011c, ApJ, 738, 169

Acharyya A., et al., 2019, Astroparticle Physics, 111, 35

Ackermann M., et al., 2015, ApJ, 799, 86

Aharonian, F. et al., 2002, A&A, 384, L23

Aharonian F. A., et al., 2001, A&A, 366, 62

Aharonian F., et al., 2003, A&A, 403, 523

Aharonian F., et al., 2006a, Science, 314, 1424

Aharonian F., et al., 2006b, Nature, 440, 1018

Aharonian F., et al., 2006c, A&A, 457, 899

Aharonian F., et al., 2007a, Astron. Astrophys., 473, L25

Aharonian F., et al., 2007b, A&A, 475, L9

Aharonian F., et al., 2008, A&A, 481, L103

Aharonian F., et al., 2009, ApJ, 695, L40

Ahnen M. L., et al., 2015, ApJ, 815, L23

Ahnen M. L., et al., 2016, A&A, 589, A33

Ajello M., et al., 2014, ApJ, 780, 73

Albert J., et al., 2006a, ApJ, 642, L119

Albert J., et al., 2006b, ApJ, 648, L105

Albert J., et al., 2007a, ApJ, 662, 892

Albert J., et al., 2007b, ApJ, 663, 125

Albert J., et al., 2007c, ApJ, 667, L21

Albert J., et al., 2008, ApJ, 685, L23

Aleksić J., et al., 2011, ApJ, 730, L8

Aleksić J., et al., 2012a, A&A, 544, A75

Aleksić J., et al., 2012b, A&A, 544, A96

Aleksić J., et al., 2014a, Science, 346, 1080

Aleksić J., et al., 2014b, A&A, 563, A91

Aleksić J., et al., 2015a, Journal of High Energy Astrophysics, 5, 30

Aleksić J., et al., 2015b, MNRAS, 451, 739

Aliu E., et al., 2011, ApJ, 742, 127

Aliu E., et al., 2012a, ApJ, 750, 94

Aliu E., et al., 2012b, ApJ, 755, 118

Aliu E., et al., 2013a, ApJ, 775, 3

Aliu E., et al., 2013b, ApJ, 779, 92

Aliu E., et al., 2014, Astrophys. J., 782, 13

Archambault S., et al., 2015, ApJ, 808, 110

Arlen T., et al., 2013, ApJ, 762, 92

Ascasibar Y., Yepes G., Gottlöber S., Müller V., 2002, A&A, 387, 396

Atwood W. B., et al., 2009, Astrophys. J., 697, 1071

Baldry I. K., Glazebrook K., 2003, ApJ, 593, 258

Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57

Berge D., Funk S., Hinton J., 2007, A&A, 466, 1219

Bernstein R. A., 2007, The Astrophysical Journal, 666, 663

Bernstein R. A., Freedman W. L., Madore B. F., 2002, ApJ, 571, 85

Bernstein R. A., Freedman W. L., Madore B. F., 2005, The Astrophysical Journal, 632, 713

Betancourt M., 2017, arXiv e-prints, p. arXiv:1701.02434

Béthermin M., et al., 2012, A&A, 542, A58

Biteau J., Meyer M., 2022, arXiv e-prints, p. arXiv:2202.00523

Biteau J., Williams D. A., 2015, ApJ, 812, 60

Blanch O., Martinez M., 2005, Astroparticle Physics, 23, 598

Blumenthal G. R., Gould R. J., 1970, Rev. Mod. Phys., 42, 237

Boettcher M., 2010, arXiv e-prints, p. arXiv:1006.5048

Boggess N. W., et al., 1992, ApJ, 397, 420

Bondi H., Gold T., 1948, MNRAS, 108, 252

Breiman L., Friedman J. H., Olshen R. A., Stone C. J., 1984, CA: Wadsworth International Group

Breit G., Wheeler J. A., 1934, Phys. Rev., 46, 1087

Catanese M., Weekes T. C., 1999, PASP, 111, 1193

Cerruti M., 2020, Galaxies, 8, 72

Cherenkov Telescope Array Consortium et al., 2019, Science with the Cherenkov Telescope Array, doi:10.1142/10986.

Cooray A., 2016, Royal Society Open Science, 3, 150555

D'Amico G., 2022, Universe, 8, 90

D'Amico G., Terzić T., Strišković J., Doro M., Strzys M., van Scherpenberg J., 2021, Phys. Rev. D, 103, 123001

Das S., Wandelt B. D., Souradeep T., 2015, J. Cosmology Astropart. Phys., 2015, 050

De Angelis A., Galanti G., Roncadelli M., 2013, MNRAS, 432, 3245

Dermer C. D., 2015, Mem. Soc. Astron. Italiana, 86, 13

Dermer C. D., Giebels B., 2016, Comptes Rendus Physique, 17, 594

Desai A., Helgason K., Ajello M., Paliya V., Domínguez A., Finke J., Hartmann D., 2019, ApJ, 874, L7

Desert F. X., Boulanger F., Puget J. L., 1990, A&A, 237, 215

Domínguez A., Prada F., 2013, ApJ, 771, L34

Domínguez A., Finke J. D., Prada F., Primack J. R., Kitaura F. S., Siana B., Paneque D., 2013, Astrophys. J., 770, 77

Domínguez A., et al., 2019, ApJ, 885, 137

Domínguez A., et al., 2011, MNRAS, 410, 2556

Driver S. P., Popescu C. C., Tuffs R. J., Graham A. W., Liske J., Baldry I., 2008, ApJ, 678, L101

Driver S. P., et al., 2016, ApJ, 827, 108

Dwek E., Krennrich F., 2005, Astrophys. J., 618, 657

Dwek E., Krennrich F., 2013, Astroparticle Physics, 43, 112

Dwek E., et al., 1998, ApJ, 508, 106

Edelstein J., Bowyer S., Lampton M., 2000, ApJ, 539, 187

Eggleton P. P., Fitchett M. J., Tout C. A., 1989, ApJ, 347, 998

Fermi-LAT Collaboration et al., 2018, Science, 362, 1031

Finke J. D., Razzaque S., Dermer C. D., 2010, ApJ, 712, 238

Finke J. D., Ajello M., Domínguez A., Desai A., Hartmann D. H., Paliya V. S., Saldana-Lopez A., 2022, ApJ, 941, 33

Fixsen D. J., 2009, ApJ, 707, 916

Fixsen D. J., et al., 2011, ApJ, 734, 5

Flegal J. M., Gong L., 2013, arXiv e-prints, p. arXiv:1303.0238

Flegal J. M., Hughes J., Vats D., Dai N., Gupta K., Maji U., 2021, mcmcse: Monte Carlo Standard Errors for MCMC. Riverside, CA, and Kanpur, India

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, Publications of the Astronomical Society of the Pacific, 125, 306

Forest E., Ruth R. D., 1990, Physica D Nonlinear Phenomena, 43, 105

Franceschini A., Rodighiero G., Vaccari M., 2008, A&A, 487, 837

Gendron-Marsolais M., et al., 2020, MNRAS, 499, 5791

Gilmore R. C., Somerville R. S., Primack Joel R. an d Domínguez A., 2012, Monthly Notices of the Royal Astronomical Society, 422, 3189

Gong L., Flegal J. M., 2016, Journal of Computational and Graphical Statistics, 25, 684

Goodman J., Weare J., 2010, Communications in Applied Mathematics and Computational Science, 5, 65

Gould R. J., Schréder G. P., 1967, Phys. Rev., 155, 1404

Guy J., Renault C., Aharonian F. A., Rivoal M., Tavernet J. P., 2000, Astron. Astrophys., 359, 419

H. E. S. S. Collaboration et al., 2010a, A&A, 511, A52

H. E. S. S. Collaboration et al., 2010b, A&A, 516, A56

H. E. S. S. Collaboration et al., 2010c, A&A, 520, A83

H. E. S. S. Collaboration et al., 2012a, A&A, 538, A103

H. E. S. S. Collaboration et al., 2012b, A&A, 542, A94

H. E. S. S. Collaboration et al., 2013a, A&A, 552, A118

H. E. S. S. Collaboration et al., 2013b, A&A, 554, A107

H. E. S. S. Collaboration et al., 2015, A&A, 573, A31

HESS Collaboration et al., 2013, MNRAS, 434, 1889

Harrison E., 1964, Nature, 204, 271

Hastings W. K., 1970, Biometrika, 57, 97

Hauser M. G., Dwek E., 2001, Ann. Rev. Astron. Astrophys., 39, 249

Hauser M. G., et al., 1998, ApJ, 508, 25

Hillas A. M., 1985, in 19th Intern. Cosmic Ray Conf-Vol. 3. No. OG-9.5-3

Hogg D. W., Foreman-Mackey D., 2018, ApJS, 236, 11

Hopkins A. M., Beacom J. F., 2006, Astrophys. J., 651, 142

Huijser D., Goodman J., Brewer B. J., 2015, arXiv e-prints, p. arXiv:1509.02230

Ichiki K., Inoue S., Takahashi K., 2008, ApJ, 682, 127

Kashlinsky A., 2005, Phys. Rep., 409, 361

Kashlinsky A., Mather J. C., Odenwald S., Hauser M. G., 1996, ApJ, 470, 681

Kneiske T. M., Mannheim K., Hartmann D. H., 2002a, A&A, 386, 1

Kneiske T. M., Mannheim K., Hartmann D. H., 2002b, A&A, 386, 1

Knödlseder J., et al., 2016, A&A, 593, A1

Knutsen H., 1997, European Journal of Physics, 18, 295

Lang R. G., Martínez-Huerta H., de Souza V., 2019, Phys. Rev. D, 99, 043015

Lauer T. R., et al., 2021, ApJ, 906, 77

Li T. P., Ma Y. Q., 1983, ApJ, 272, 317

MAGIC Collaboration et al., 2008, Science, 320, 1752

Madau P., Pozzetti L., 2000, Mon. Not. Roy. Astron. Soc., 312, L9

Martínez-Huerta H., Lang R. G., de Souza V., 2020, Symmetry, 12, 1232

Matsumoto T., et al., 2005, ApJ, 626, 31

Matsuura S., et al., 2017, Astrophys. J., 839, 7

Mattila K., 2006, Mon. Not. Roy. Astron. Soc., 372, 1253

Mazin D., Raue M., 2007, A&A, 471, 439

Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E., 1953, J. Chem. Phys., 21, 1087

Meyer M., 2022, ebltable, doi:10.5281/zenodo.7312062, `https://doi.org/10.5281/zenodo.7312062`

Meyer M., Raue M., Mazin D., Horns D., 2012, Astron. Astrophys., 542, A59

Murthy J., Hall D., Earl M., Henry R. C., Holberg J. B., 1999, ApJ, 522, 904

Murthy J., Henry R. C., Sujatha N. V., 2010, ApJ, 724, 1389

Neal R., 2011, in , Handbook of Markov Chain Monte Carlo. pp 113–162, doi:10.1201/b10905

Netzer H., 2015, ARA&A, 53, 365

Neyman J., Pearson E. S., 1933, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 231, 289

Nitu I. C., Bevins H. T. J., Bray J. D., Scaife A. M. M., 2021, Astroparticle Physics, 126, 102532

Observatory C. T. A., Consortium C. T. A., 2016, CTAO Instrument Response Functions - version prod3b-v2, doi:10.5281/zenodo.5163273, `https://doi.org/10.5281/zenodo.5163273`

Ohm S., van Eldik C., Egberts K., 2009, Astroparticle Physics, 31, 383

Orr M. R., Krennrich F., Dwek E., 2011, ApJ, 733, 77

Peskin M. E., 2018, An introduction to quantum field theory. CRC press

Pilbratt G. L., et al., 2010, A&A, 518, L1

Planck Collaboration et al., 2020a, A&A, 641, A1

Planck Collaboration et al., 2020b, A&A, 641, A6

Primack J. R., Bullock J. S., Somerville R. S., 2005, in Aharonian F. A., Völk H. J., Horns D., eds, American Institute of Physics Conference Series Vol. 745, High Energy Gamma-Ray Astronomy. pp 23–33 (arXiv:astro-ph/0502177), doi:10.1063/1.1878394

Protheroe R. J., Biermann P. L., 1996, Astroparticle Physics, 6, 45

Razzaque S., Dermer C. D., Finke J. D., 2009, ApJ, 697, 483

Riess A. G., et al., 2022, ApJ, 934, L7

Ryden B., 2017, Introduction to cosmology. Cambridge University Press

Salamon M. H., Stecker F. W., de Jager O. C., 1994, ApJ, 423, L1

Saldana-Lopez A., Domínguez A., Pérez-González P. G., Finke J., Ajello M., Primack J. R., Paliya V. S., Desai A., 2020, arXiv e-prints, p. arXiv:2012.03035

Schneider P., 2014, Extragalactic astronomy and cosmology: an introduction. Springer

Shilon I., et al., 2019, Astroparticle Physics, 105, 44

Sokal A., 1997, in , Functional integration. Springer, pp 131–192

Stecker F. W., Malkan M. A., Scully S. T., 2006, ApJ, 648, 774

Tagliaferri G., et al., 2008, ApJ, 679, 1029

Treister E., et al., 2006, ApJ, 640, 603

Vats D., Flegal J. M., Jones G. L., 2015, arXiv e-prints, p. arXiv:1512.07713

Vavryčuk V., 2016, Ap&SS, 361, 198

Wakely S. P., Horan D., 2008, in International Cosmic Ray Conference. pp 1341–1344

Werner M. W., et al., 2004, ApJS, 154, 1

Wesson P., 1989, Science Progress (1933-), pp 133–146

Wesson P. S., 1991, ApJ, 367, 399

Wesson P. S., Valle K., Stabell R., 1987, ApJ, 317, 601

Wilks S. S., 1938, Ann. Math. Stat., 9, 60

Zemcov M., Immel P., Nguyen C., Cooray A., Lisse C. M., Poppe A. R., 2017, Nature Communications, 8, 15003

Zhong W., Liu W. G., Zheng Y. G., 2018, Ap&SS, 363, 179

de Matos Pimentel D. R., Moura-Santos E., 2019, J. Cosmology Astropart. Phys., 2019, 043