

A two-semester research project for the course
PHY_50P13_EP - Projet de recherche en laboratoire (2024-2025)

SEARCHING FOR SIGNATURES OF THE EBL ABSORPTION IN THE CTA INTERNAL SCIENCE DATA CHALLENGE

Maksim Kleimenov
scientific advisor Dr. Stephen Fegan

Contents

Introduction	2
1 Optical depth of the extragalactic medium	3
1.1 Distance element	4
1.2 Cross-section	5
1.3 Photon number density and specific intensity	6
1.4 Optical depth calculation	7
2 STeVCat and intrinsic spectra of the sources	9
2.1 The catalogue	9
2.2 Spectrum curvature	10
3 Bayesian approach to fitting experimental data	11
3.1 Single observation	11
3.2 MCMC for several independent observations	12
4 Single-parameter model of the EBL density	13
4.1 General spectrum model	13
4.2 Saldana-Lopez scaling	14
4.3 EBL density constraints	15
5 Multiparametric model of the EBL density	17
5.1 Choosing a kernel	17
5.2 Taking redshift into account	18
5.3 Setting the priors	19
5.4 EBL density estimations	20
Summary	22

Introduction

The extragalactic background light (EBL) is the diffuse radiation from all stellar and galactic processes throughout cosmic history that permeates the Universe. It occupies a broad range of wavelengths, from ultraviolet to gamma-rays; Its main components are cosmic optical background (COB), which originates from stellar radiation, and cosmic infrared background (CIB), which is understood to be the result of interstellar dust thermal emission [1]. The EBL plays a crucial role in astrophysical studies as it provides insights into galaxy formation, cosmic star formation history, and the large-scale structure of the Universe.

In the present work we use one of the primary observational effects of the EBL: its interaction with high-energy gamma rays ($E > 10$ GeV). When gamma rays from distant astrophysical sources travel through intergalactic space, they can undergo electron-positron pair production: $\gamma_{\text{HE}} + \gamma_{\text{EBL}} \rightarrow e^+ + e^-$ upon colliding with EBL photons [2]. This process leads to a characteristic energy-dependent attenuation of gamma-ray fluxes, which are used to probe the EBL intensity and spectral distribution.

High-energy gamma-ray sources, such as blazars, active galactic nuclei (AGN), and gamma-ray bursts (GRBs), serve as natural laboratories for studying EBL absorption. Their gamma-ray spectra, observed by ground-based Cherenkov telescopes (e.g., H.E.S.S., MAGIC, VERITAS [3, 4, 5]) and space-based detectors (e.g., Fermi-LAT [6]), exhibit signatures of EBL-induced attenuation. By analyzing spectral cutoffs and deviations from intrinsic power-law emissions, researchers can infer constraints on the EBL density and its evolution with redshift [7, 8].

The Cherenkov Telescope Array Observatory (CTAO), a next-generation experiment currently under construction, will consist of two Cherenkov telescope arrays in both hemispheres, allowing for full-sky coverage at extremely high energies [9]. As the construction works are expected to finish by the end of decade, the CTA collaboration members work on design and validation of the analysis tools. To establish conditions similar to those of future observations, the CTA collaboration performed a comprehensive simulation of the future experimental results, called CTA internal data challenge, with a variety of theoretical assumptions on various properties of the Universe, including a certain EBL model.

The present work was initially aimed to detect the EBL absorption embedded in the CTA internal data challenge. However, setting up the analysis tools required significant effort, and thus the research was concluded with detecting the EBL absorption from the STeVECat, a database of high-energy gamma-ray sources observations made by different telescopes [10].

In the first section we discuss the intergalactic medium optical depth from the theoretical perspective; the second section is dedicated to the discussion of the sources from the STeVECat and constraints on their physical spectra; section 3 reviews relevant notions and approaches in Bayesian analysis, which are used in section 4 for a single-parameter EBL analysis, based on the given EBL density profile; in section 5 we introduce the procedure to construct the EBL density from experimental data in a model-independent way, and report our results for the STeVECat experimental data.

1 Optical depth of the extragalactic medium

When high-energy photons travel from a certain source through space, they can scatter on low-energy photons by creating electron-positron pairs, which is illustrated in Figure 1.

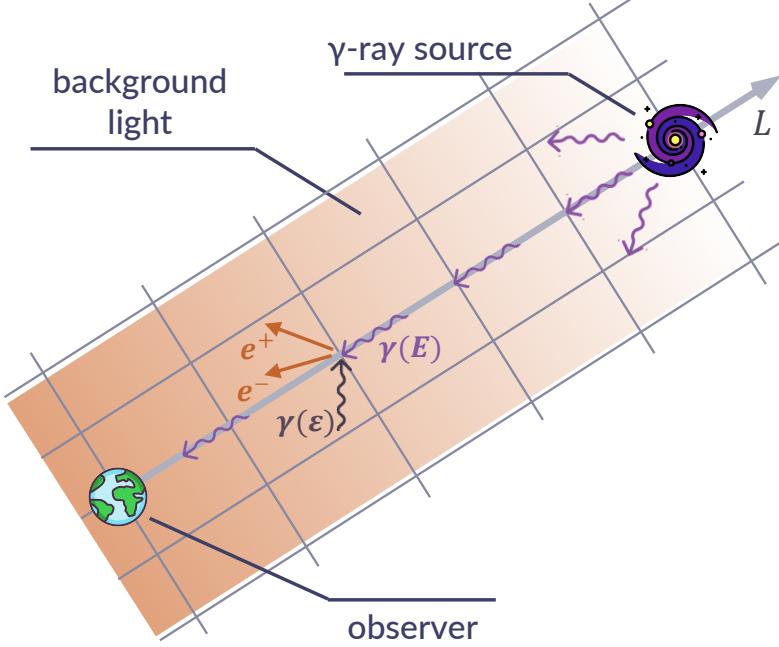


Figure 1: Gamma-rays on the EBL scattering in a «flat» space

If the scattering processes occurred in a flat stationary space, they could be described by the notion of *optical depth* [11]:

$$\tau_{\gamma\gamma}^{\text{flat}}(E) = \int_0^{L_0} dL \int_0^{\infty} d\epsilon n_{\epsilon}(L, \epsilon) \sigma_{\gamma\gamma}(E, \epsilon), \quad (1.1)$$

where $n_{\epsilon} = \partial n / \partial \epsilon$ is local background photon number density, $\sigma_{\gamma\gamma}$ is the scattering cross-section.

However, when looking at far-away gamma-ray sources, one needs to take the expansion of the universe into account. Expression (1.1) then changes to

$$\tau_{\gamma\gamma} = \int_0^{z_0} dz \frac{\partial L}{\partial z} \int_0^{\infty} d\epsilon n'_{\epsilon}(\epsilon, z) \sigma_{\gamma\gamma}(E', \epsilon), \quad (1.2)$$

where primes indicate values in the local reference frame at fixed redshift z , which implies $E' = (1+z)E$ by definition of redshift, and a more sophisticated transformation of n_{ϵ} , which will be discussed later.

In the following subsections a brief discussion of each component of the integral (1.2) is presented.

1.1 Distance element

To calculate distance element $dL = dz \frac{\partial L}{\partial z}$ we use standard FLRW metric [12]:

$$ds^2 = -c^2 dt^2 + a^2(t) [dL^2 + S_k^2(L)d\Omega^2]. \quad (1.3)$$

As only objects with $z < 1.0$ are taken into account, radiation contribution is negligible, and we write Friedman equation only with matter ($w_M = 0$) and Λ CMD ($w_\Lambda = -1$) terms as following:

$$\left(\frac{\dot{a}}{a}\right)^2 = H_0^2 \sum_{i=\{M,\Lambda\}} \Omega_{i,0} \left(\frac{a_0}{a}\right)^{3(1+w_i)} = H_0^2 \left[\Omega_M \left(\frac{a_0}{a}\right)^3 + \Omega_\Lambda \right]. \quad (1.4)$$

Co-moving coordinates of the source are considered constant, the light ray points directly to the Earth ($d\varphi = d\theta = 0$). For a photon $ds = 0$, and $cdt = a(t)dL$.

$$dL = \frac{c}{a(t)} dt = \frac{c}{a(t)} \frac{\partial t}{\partial z} dz = \frac{c}{a(t)} \frac{dz}{\dot{z}}.$$

By definition of redshift, $1 + z = a_0/a$, hence

$$\dot{z} = -\frac{a_0 \dot{a}}{a^2} = -(1+z) \frac{\dot{a}}{a} = -(1+z) H_0 \sqrt{\Omega_M (1+z)^3 + \Omega_\Lambda},$$

which leads directly to the final result

$$\frac{\partial L}{\partial z} = \frac{c}{H_0} \frac{1}{1+z} \frac{1}{\sqrt{\Omega_M (1+z)^3 + \Omega_\Lambda}} = \frac{c}{H_0} \frac{\partial l}{\partial z}. \quad (1.5)$$

Integrated equation (1.5) is visualized in Figure 2. Numerical values $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, and $H_0 = 7 \cdot 10^4 \text{ m/s Mpc}^{-1}$ were used.

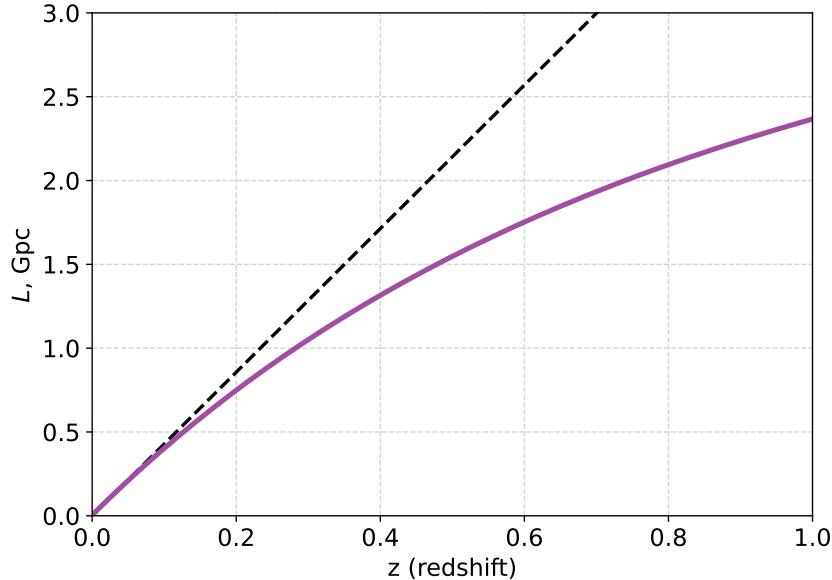


Figure 2: Dependence of distance on redshift

1.2 Cross-section

The total cross-section $\sigma_{\gamma\gamma}$ can be expressed from the integration of the differential cross-section $\partial\sigma/\partial\Omega$ with an additional term $(1 - \mu)/2$, where $\mu = \cos\theta$, angle between the two photons [13].

$$\sigma_{\gamma\gamma}(E, \epsilon) = \int_{-1}^1 d\mu \frac{1 - \mu}{2} \frac{\partial\sigma}{\partial\Omega} [\beta^2(E, \epsilon, \mu)] \quad (1.6)$$

The differential cross-section $\sigma_{\gamma\gamma}(\beta)$ is given in [14] by:

$$\frac{\partial\sigma}{\partial\Omega} = \frac{1}{2}\pi r_0^2(1 - \beta^2) \left[(3 - \beta^4) \ln \frac{1 + \beta}{1 - \beta} - 2\beta(2 - \beta^2) \right], \quad (1.7)$$

where r_0 is the classical electron radius, and β is the e^- (and e^+) velocity in the center-of-mass frame:

$$\beta^2 = 1 - \frac{2m_e^2c^4}{E\epsilon} \frac{1}{1 - \mu}.$$

This expression contains an explicit constraint on energies of interacting photons: β^2 has to be non-negative, which corresponds to a pair production threshold condition.

Differential cross-section as a function of the interaction angle for several energies of incident high-energy photon is shown in Figure 4.

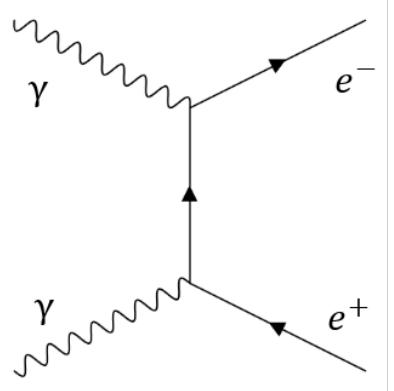


Figure 3: Feynman diagram for photon-photon scattering $\gamma\gamma \rightarrow e^+e^-$

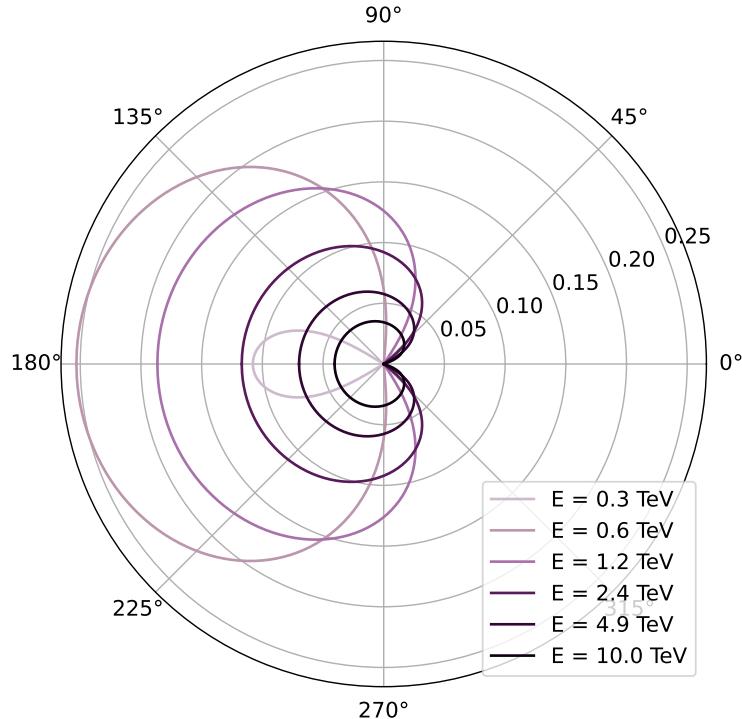


Figure 4: Differential cross-section of pair-production on a 1 eV background photon for various incoming photon energies E

In comoving frame, the background is spherically symmetric, and it turns out, that one can perform the integration (1.6) analytically with the use of the *particle-physics kernel* $P(\beta_{\max}^2(E, \epsilon))$, where $\beta_{\max}^2 = 1 - m_e^2 c^4 / E \epsilon$ [7].

Both methods were implemented in the computational scheme, which allows making faster calculations for homogeneous case and, in prospect, investigating non-homogeneous cases, e.g. modeling uneven matter distribution on the line-of-sight of the source. The results of these two approaches coincide with machine precision.

1.3 Photon number density and specific intensity

Experimental works usually provide measurements of *specific intensity multiplied by wavelength* $[\lambda I_\lambda]$ (further this value is called specific intensity), which is invariant under transformations of the reference value $\lambda \leftrightarrow \nu(\epsilon)$ [15, 16, 17].

Let us now deduce the connection between the observed specific intensity $[\lambda I_\lambda]$ and the *comoving* (or emitted) photon density $n'_\epsilon = \partial n / \partial \epsilon'$ at redshift z in assumption of *no absorption*.

By definition [11],

$$dE_\nu = I_\nu dA_\perp d\Omega d\nu dt.$$

Observed photon energy density u_ν is:

$$\frac{du_\nu}{d\Omega} = \frac{dE_\nu}{dV d\nu d\Omega},$$

with $dV = dA_\perp c dt$, where dt is the observer time.

Hence, if we consider all the photon flux,

$$u_\nu = \int_{4\pi} \frac{du_\nu}{d\Omega} d\Omega = \int_{4\pi} \frac{I_\nu dA_\perp d\Omega d\nu dt}{dA_\perp c dt d\nu d\Omega} d\Omega = \frac{1}{c} \int_{4\pi} I_\nu d\Omega = \frac{4\pi}{c} I_\nu.$$

By connecting photon energy density u_ν and photon number density n_ν , one writes

$$n_\nu = \frac{u_\nu}{h\nu} = \frac{4\pi}{hc} \frac{1}{\nu} I_\nu = \frac{4\pi}{hc} \frac{1}{\nu^2} [\nu I_\nu] = \frac{4\pi h}{c} \frac{1}{\epsilon^2} [\lambda I_\lambda]$$

We want to connect the measured EBL specific intensity $[\lambda I_\lambda] = [\nu I_\nu]$ with comoving energy photon density $n'_\epsilon = n'_\epsilon(z) = n'_\nu(z)/h$. The quantities with a prime correspond to the comoving reference frame, the observed quantities carry no prime.

$$u_\nu = u'_\nu \cdot (a/a')^{-4} = u'_\nu \cdot (1+z)^{-4}.$$

As photon frequency also changes due to expansion of the Universe, $\nu' = \nu(1+z)$. Thus,

$$n'_\nu = \frac{u'_\nu}{h\nu'} = \frac{u_\nu(1+z)^4}{h\nu(1+z)} = n_\nu(1+z)^3.$$

Thus, we find the desired relation between the photon number density at redshift z and the measured specific intensity:

$$n'_\epsilon = \frac{4\pi}{c} \epsilon^{-2} [\lambda I_\lambda](\epsilon, z) (1+z)^3. \quad (1.8)$$

1.4 Optical depth calculation

It is convenient to use logarithmic energy grid to perform numerical calculation of the EBL optical depth. For that, the z-independent part of (1.2) with regard to (1.8) can be rewritten as:

$$\tau_{\gamma\gamma}(E, z) = \frac{4\pi}{c} \ln 10 \frac{1}{(1+z)^3} \int d(\lg \epsilon) \frac{1}{\epsilon} [\lambda I_\lambda](\epsilon, z) \int d\mu \frac{1-\mu}{2} \sigma_{\gamma\gamma} [\beta(E, \epsilon, z, \mu)]. \quad (1.9)$$

To make an actual calculation, the EBL density profile from the Saldana-Lopez et al.[17] (SL) was used. We performed interpolation in the data domain, and made a log-scale extrapolation outside of it to obtain there values close to zero.

The cosmic microwave background, as dominant contribution to the overall cosmic photon density, was added to the total value as an independent component of the Planck-distribution-based shape (1.10).

$$n'_{\text{CMB}}(\lambda, z) = \frac{8\pi}{\lambda^4} \frac{1}{e^{hc/\lambda kT'} - 1}, \quad T' = (1+z)T \quad (1.10)$$

By inverting (1.8), one obtains the CMB specific intensity $[\lambda I_\lambda]_{\text{CMB}}$ in the following form:

$$[\lambda I_\lambda]_{\text{CMB}}(\lambda, z) = \frac{2hc(1+z)^3}{\lambda^4} \frac{1}{e^{hc/\lambda kT'} - 1}, \quad T' = (1+z)T$$

Both CMB and SL contributions to the total specific intensity are shown in Figure 5.

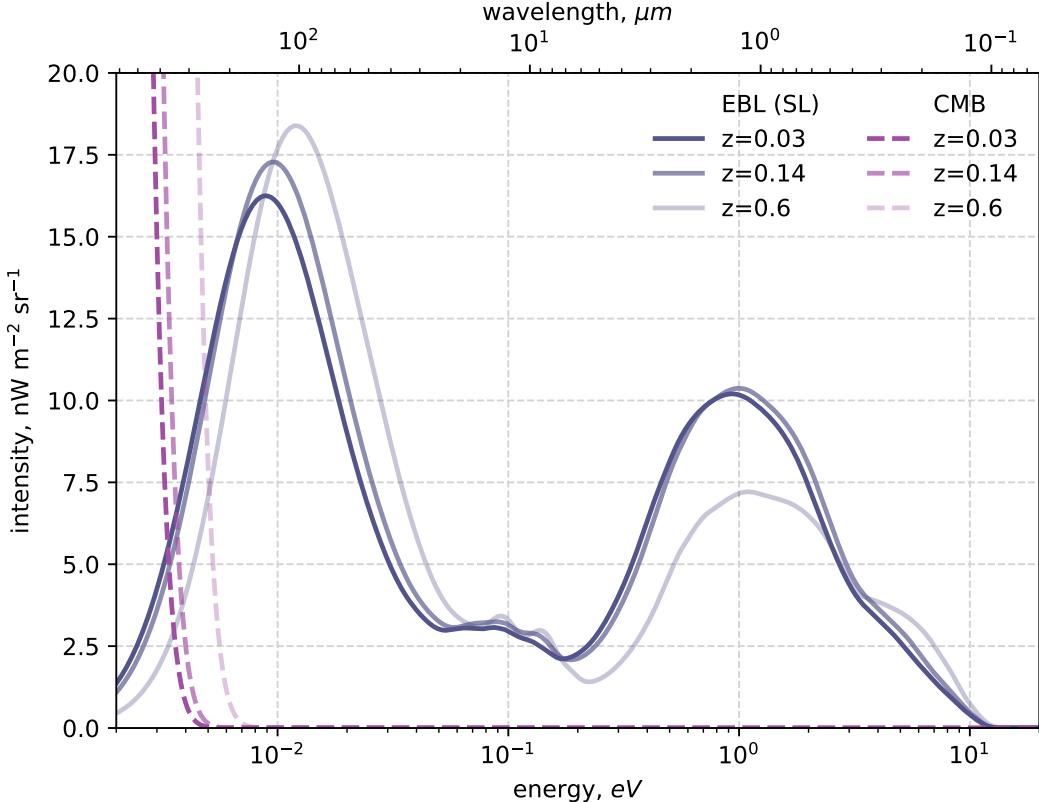


Figure 5: EBL specific intensity: the optical and infrared part (blue solid lines) from [17] and the CMB contribution (purple dashed lines) for different redshifts

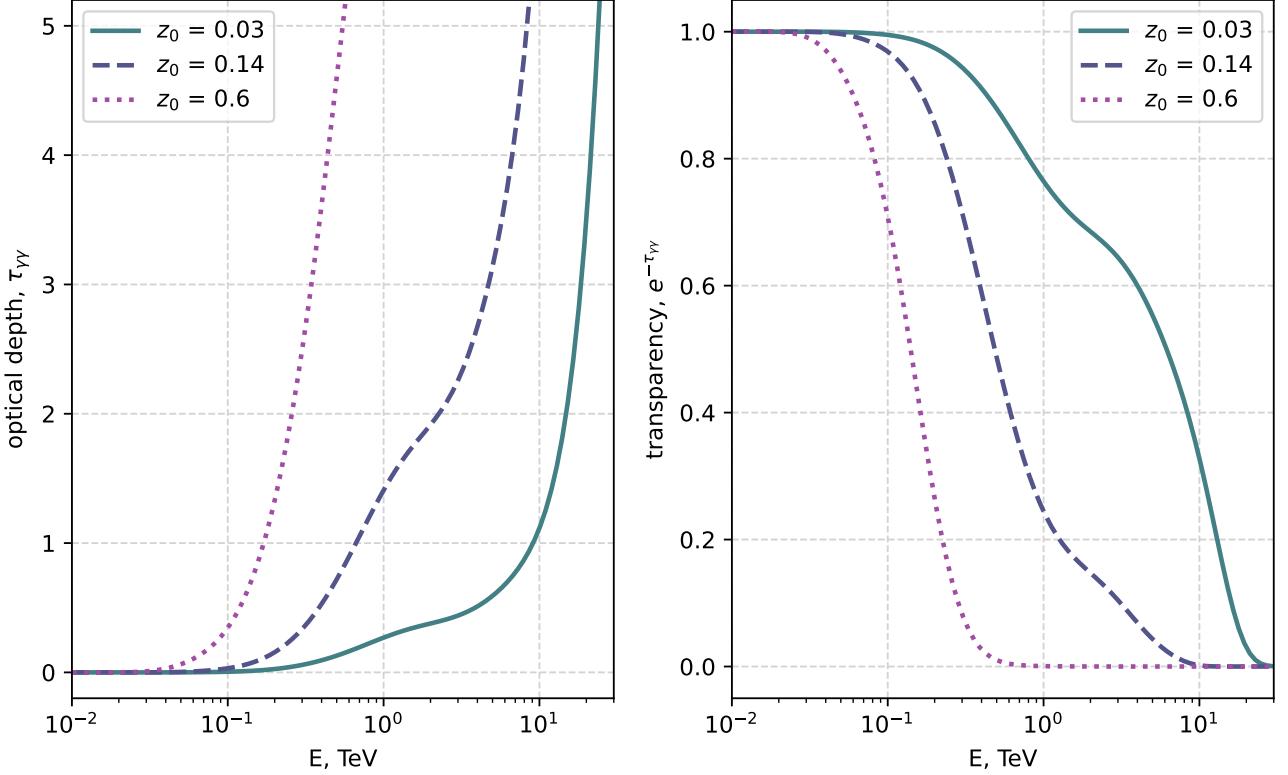


Figure 6: Optical depth of EBL $\tau_{\gamma\gamma}$ (left) and its transparency $e^{-\tau_{\gamma\gamma}}$ (right) dependence on energy for sources at different redshifts z_0 , calculated from (1.2) for the Saldana-Lopez photon density model.

The total specific intensity consists of two components: $[\lambda I_\lambda]_{\text{EBL}} = [\lambda I_\lambda]_{\text{CMB}} + [\lambda I_\lambda]_{\text{SL}}$, which gives rise to the optical depth $\tau_{\gamma\gamma}$ and corresponding transparency $e^{-\tau_{\gamma\gamma}}$ of extragalactic medium, shown in Figure 6.

One can see that at low redshifts, the extragalactic medium stays semi-transparent for photons of energies up to 10 TeV, whereas at redshift $z = 0.6$ the universe is opaque for all gamma-rays with energies higher than 1 TeV. This result allows to detect the EBL indirectly by measuring the spectra of far-away very-high energy gamma-ray sources and searching for an observational cut-off at the level of several TeV.

The obtained results were compared with the values from [17], and also with the standard `gamma-py` implementation of the Saldana-Lopez optical depth, the difference turned out to be of the order of interpolation errors and machine precision.

2 STeVCat and intrinsic spectra of the sources

2.1 The catalogue

STeVCat (Spectral TeV Extragalactic Catalog) is a catalog of extragalactic very high-energy gamma-ray sources, focusing on their spectral properties. It compiles spectral measurements from TeV observatories, providing a uniform dataset for studying the energy distributions and absorption effects of gamma rays due to the EBL.

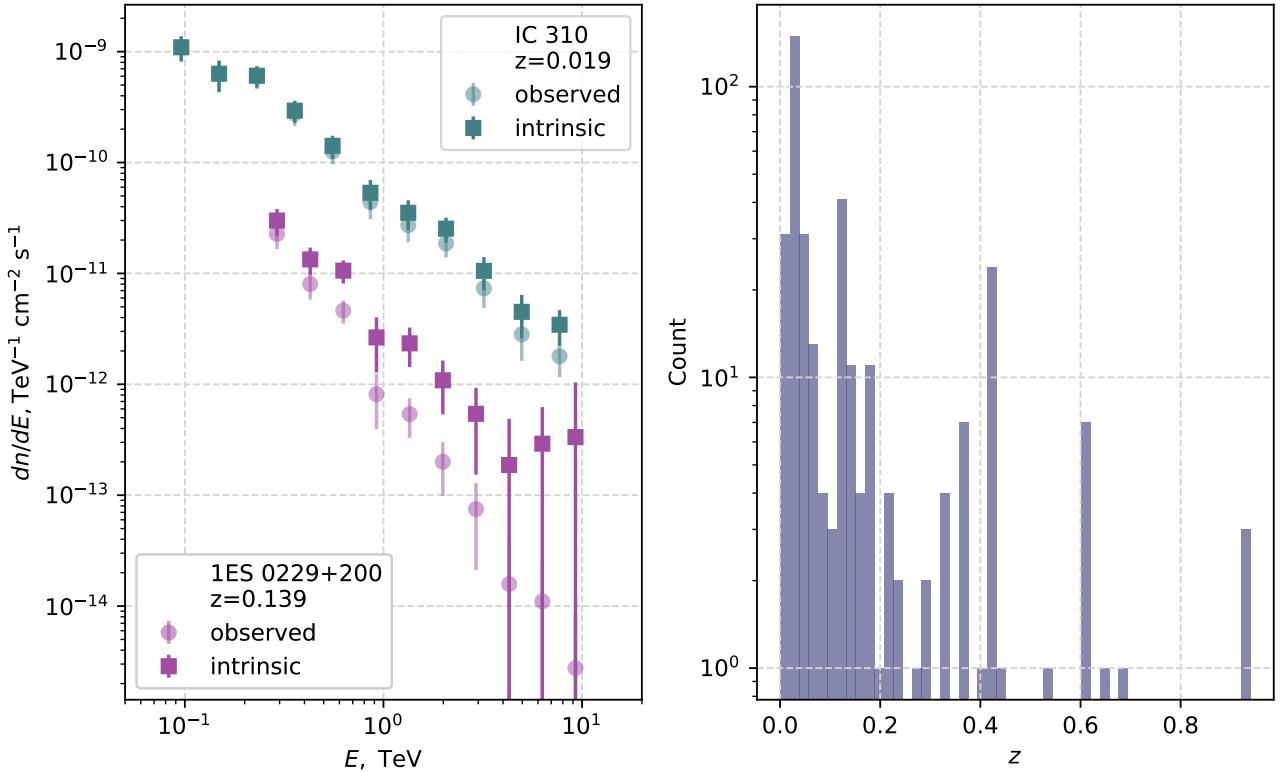


Figure 7: Left: A typical STeVCat source spectrum (left) for two objects: IC 310 ($z=0.019$) and for 1ES 0229+200 ($z=0.139$). Semi-transparent points with errors correspond to the observed flux, opaque ones stand for the reconstructed intrinsic spectrum of the sources with the use of Saldana-Lopes EBL opacity.

Right: Distribution of sources with different redshifts in the STeVCat.

In Figure 7 we show two typical STeVCat source spectra for two objects on different redshifts are shown. The semi-transparent points are taken directly from the STeVCat and represent the observed flux from the sources, and the colored points show the recalculated intrinsic spectrum with the use of the Saldana-Lopez EBL parametrization:

$$\Phi_{\text{int}}(E, z_0) = \Phi_{\text{obs}}(E) \cdot \exp\{+\tau_{\text{SL}}(E, z_0)\}.$$

From this figure we can clearly see, that the observed spectrum of a far-away source ($z = 0.139$) is more inclined at high-energies, in comparison to a closer source. These kind of signatures may indicate the existence of EBL absorption.

2.2 Spectrum curvature

To ensure that the Saldana-Lopez parametrization of the EBL density is physically valid, we investigated the curvature of the spectra. Based on our current understanding of acceleration and radiation processes in astrophysical objects, intrinsic spectra are generally not expected to have positive curvature [1, 7].

To test this, we employed a straightforward method: the intrinsic spectra were reconstructed from observations, as shown in Figure 7, and subsequently fitted with a parabola in the $(\lg E, \lg(dn/dE))$ space:

$$\lg(dn/dE) = -ax^2 + bx + c, \quad x = \lg E.$$

Table 1 presents the number of sources in STeVECat with positive curvature (negative a parameter) among 324 sources at different confidence levels for lower, mean, and high EBL density values. Notably, at the 5σ level, no sources exhibit unphysical spectra, and at the 4σ level, only under the assumption of high EBL absorption, there occur two observations with positively curved spectra.

	low	mean	high
0σ	95	103	112
1σ	41	44	51
2σ	19	22	23
3σ	6	6	10
4σ	0	0	2
5σ	0	0	0

Table 1: The number of sources in STeVECat with positive curvature (negative a parameter) among 324 sources at different confidence levels for lower, mean and high EBL density values from the Saldana-Lopez article [17].

The abovementioned results provide the first qualitative constraints on the EBL density.

To establish quantitative constraints, we required several statistical methods, which are the focus of the following section.

3 Bayesian approach to fitting experimental data

To detect the EBL, one must first develop a description of the EBL specific intensity, $[\lambda I_\lambda]_{\text{EBL}}$. Such a model is characterized by a set of parameters, $\boldsymbol{\alpha}$, which define the shape of the specific intensity curve.

Given a set of experimental observations, $\mathcal{D} = \{D_k\}_{k=1}^N$, it is necessary to constrain the possible values of the parameters $\boldsymbol{\alpha}$. To properly organize the comparison of complex spectral models with typically sparse experimental data, Bayesian analysis approaches are commonly used. The method applied in this work is described in the subsections below.

3.1 Single observation

In order to get a Bayesian credibility interval for parameter-vector $\boldsymbol{\alpha}$ in a group of parameters $\Theta = \{\boldsymbol{\alpha}, \boldsymbol{\theta}\}$ from a single observation D , one uses the famous Bayes's Theorem (B.Th.):

$$P(\Theta|D) \xrightarrow{\text{B.Th.}} \frac{P(D|\Theta)P(\Theta)}{P(D)}.$$

The distribution $P(\Theta)$ is considered to be known *a priori*, and is called *the prior distribution*.

As observation data is given, for all possible sets of parameters Θ the denominator will be the same, so one can omit its calculation and say

$$P(\Theta|D) \propto P(D|\Theta)P(\Theta). \quad (3.1)$$

To get the so-called *posterior* distribution (the one that takes data into consideration) of parameter $\boldsymbol{\alpha}$, one can integrate both parts by irrelevant parameters $\boldsymbol{\theta}$ and thus obtain

$$P(\boldsymbol{\alpha}|D) \propto \int d\boldsymbol{\theta} P(D|\boldsymbol{\alpha}, \boldsymbol{\theta})P(\boldsymbol{\alpha}, \boldsymbol{\theta}). \quad (3.2)$$

The first function under integral (3.2) is by definition the likelihood function, which can be constructed in the following way for gaussian errors

$$\mathcal{L}(D, \boldsymbol{\alpha}, \boldsymbol{\theta}) = P(D|\boldsymbol{\alpha}, \boldsymbol{\theta}) = \prod_{\text{obs}} \frac{1}{\sqrt{2\pi\sigma_{\text{obs}}^2}} \exp \left\{ -\frac{1}{2} \left(\frac{\Phi_{\text{mod}} - \Phi_{\text{obs}}}{\sigma_{\text{obs}}} \right)^2 \right\},$$

or can be modified to account for different lower and upper errorbars in the following form

$$\begin{aligned} \mathcal{L}(D, \boldsymbol{\alpha}, \boldsymbol{\theta})_{\text{asym}} = & \prod_{\text{obs}} \left[\frac{1}{\sqrt{2\pi\sigma_{\text{obs}}^+}} \exp \left\{ -\frac{1}{2} \left(\frac{\Phi_{\text{mod}} - \Phi_{\text{obs}}}{\sigma_{\text{obs}}^+} \right)^2 \right\} \mathbf{1}(\Phi_{\text{mod}} > \Phi_{\text{obs}}) + \right. \\ & \left. + \frac{1}{\sqrt{2\pi\sigma_{\text{obs}}^-}} \exp \left\{ -\frac{1}{2} \left(\frac{\Phi_{\text{obs}} - \Phi_{\text{mod}}}{\sigma_{\text{obs}}^-} \right)^2 \right\} \mathbf{1}(\Phi_{\text{mod}} < \Phi_{\text{obs}}) \right], \quad (3.3) \end{aligned}$$

where the standard indicator function is used $\mathbf{1}(x \in A) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{else.} \end{cases}$

The second function under integral (3.2) encodes our initial knowledge about the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$. This is referred to as a prior.

Given the prior and assuming a normal distribution of measurement errors in the data, one can construct the posterior distribution. From this, the expected values $\hat{\boldsymbol{\alpha}}$ can be determined, along with a corresponding credibility interval.

3.2 MCMC for several independent observations

Intuitively, a constraint from a single observation at a limited number of energies is insufficient to constrain a complex model. To draw reliable conclusions about the possible values of the parameters $\boldsymbol{\alpha}$ in the parameter set $\Theta = \{\boldsymbol{\alpha}, \boldsymbol{\theta}\}$, the model should be applied to multiple independent experimental measurements \mathcal{D} .

The posterior distribution is then marginalized over the observation-specific parameters $\boldsymbol{\theta}$, assuming that $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ are independent:

$$\begin{aligned} P(\boldsymbol{\alpha}|\mathcal{D}) &= \int d\boldsymbol{\theta} P(\boldsymbol{\alpha}, \boldsymbol{\theta}|\mathcal{D}) \stackrel{\text{B.Th.}}{=} \int d\boldsymbol{\theta} \frac{P(\mathcal{D}|\boldsymbol{\alpha}, \boldsymbol{\theta})P(\boldsymbol{\alpha})P(\boldsymbol{\theta})}{P(\mathcal{D})} \\ &\stackrel{\text{indep.}}{=} P(\boldsymbol{\alpha}) \int d\boldsymbol{\theta} \prod_{k=1}^N \frac{P(D_k|\boldsymbol{\alpha}, \boldsymbol{\theta}_k)P(\boldsymbol{\theta}_k)}{P(D_k)} \stackrel{\text{B.Th.}}{=} P(\boldsymbol{\alpha}) \int d\boldsymbol{\theta} \prod_{k=1}^N \frac{P(\boldsymbol{\alpha}, \boldsymbol{\theta}_k|D_k)}{P(\boldsymbol{\alpha})} = \\ &= P(\boldsymbol{\alpha}) \prod_{k=1}^N \int d\boldsymbol{\theta}_k \frac{P(\boldsymbol{\alpha}, \boldsymbol{\theta}_k|D_k)}{P(\boldsymbol{\alpha})} = P(\boldsymbol{\alpha}) \prod_{k=1}^N \frac{P(\boldsymbol{\alpha}|D_k)}{P(\boldsymbol{\alpha})} \quad (3.4) \end{aligned}$$

This shows that the posterior distribution can be computed at first for each source, and then the overall posterior results from multiplication of single-measurement posteriors. Analytical calculation of (3.4) is hardly possible, however one can obtain posteriors by sampling. In this work the Markov chain Monte-Carlo method is used due to its simplicity and universality.

Markov Chain Monte Carlo (MCMC) is a class of algorithms designed for sampling from complex probability distributions. These algorithms construct a Markov chain whose equilibrium distribution matches the target distribution.

In this work we used the MCMC method implemented in python library `emcee`, based on Metropolis-Hastings algorithm. Brief mathematical description of the method with some technical remarks on the library and its features are given in the article [18].

To sample the distribution, the method begins from a certain point in parametric space and step-by-step moves to the distribution with the maximal posterior probability $P(\boldsymbol{\alpha}, \boldsymbol{\theta}|\mathcal{D}) \propto \mathcal{L}(D, \boldsymbol{\alpha}, \boldsymbol{\theta})P(\boldsymbol{\alpha}, \boldsymbol{\theta})$. Surely, the equilibrium distribution should not depend on the starting point. To describe the step, at which the walker completely "forgets" its starting point, one uses the notion of *correlation time* τ_C . The method to estimate τ_C is present in the `emcee` library and was used in the present work to define the number of first steps to be cut to get the starting-point-independent distribution sample.

4 Single-parameter model of the EBL density

Experimental measurements for the Bayesian analysis were taken from STeVCat, a database of high-energy gamma-ray source observations collected by various telescopes.

The initial idea of the project was to use raw data from the CTA internal science data challenge. However, the acquisition of Bayesian techniques required more resources than planned in the beginning, leading to the decision to use already processed data from previous measurements.

4.1 General spectrum model

There exist several common models for parameterizing intrinsic spectra Φ_{int} of the sources: power law, log-parabolic, power law with exponential cutoff and the log parabola with exponential cutoff. One can define the generalized spectrum in the following way [19]:

$$\Phi_{\text{int}}(E, \eta, \gamma, \beta, \lambda) = \Phi_0 \cdot \exp\{\eta - \gamma \ln \varepsilon - \beta \ln^2 \varepsilon\} \cdot \exp\{-\lambda \varepsilon\}, \quad (4.1)$$

where $\varepsilon = E/E_0$, and $E_0^2 = E_{\min} E_{\max}$.

The observed spectrum differs from the intrinsic one by the Universe opacity $e^{-\tau_{\gamma\gamma}}$ factor, which depends on one or several parameters, which we denote as $\boldsymbol{\alpha}$:

$$\Phi_{\text{obs}}(E, \boldsymbol{\theta}, \boldsymbol{\alpha}) = \Phi_{\text{int}}(E, \eta, \gamma, \beta, \lambda) \cdot e^{-\tau_{\gamma\gamma}(E, \boldsymbol{\alpha})}, \quad (4.2)$$

where $\boldsymbol{\theta}$ is the source-specific set of parameters $\{\eta, \gamma, \beta, \lambda\}$.

The normalization constant Φ_0 in (4.1) is given by the following equality

$$\Phi_0^2 = \Phi_{\text{int, SL}}(E_{\min}) \cdot \Phi_{\text{int, SL}}(E_{\max}) = \Phi_{\text{obs}}(E_{\min}) e^{\tau_{\text{SL}}(E_{\min})} \cdot \Phi_{\text{obs}}(E_{\max}) e^{\tau_{\text{SL}}(E_{\max})}. \quad (4.3)$$

This allows us to take the prior for parametric normalization constant η symmetric around 0.

Prior distributions

For Bayesian approach, initial intrinsic spectrum to be a second-degree power law:

$$\Phi_{\text{int}}^{\text{std}}(E) = \Phi_0 \cdot \left(\frac{E}{E_0}\right)^{-2}$$

hence we assume $\langle \eta \rangle = 0$, $\langle \gamma \rangle = 2$, $\langle \beta \rangle = 0$, $\langle \lambda \rangle = 0$, where brackets denote initial values.

In accordance with [19], but using slightly different notation, we chose the following priors

$$P(\eta) = \mathcal{U}[0; 4], \quad P(\gamma) = \mathcal{U}[2; 3], \quad P(\beta) = \mathcal{U}[0; 2], \quad P(\lambda) = \mathcal{U}[0; 2], \quad (4.4)$$

where $\mathcal{U}[X, \delta X](x) = \frac{1}{2\delta X} \cdot \mathbf{1}(x \in [X - \delta X, X + \delta X])$, which means that X is the average (initial) value, and δX is the half-width of the uniform distribution.

4.2 Saldana-Lopez scaling

Sampling prerequisites

As a simple approach to constrain the EBL density, we scale the SL parmetrization:

$$\tau_{\text{mod}}(E, z; \alpha) = \alpha \cdot \tau_{\text{SL}}(E, z) \quad (4.5)$$

Value $\alpha = 0$ from (4.2) gives $\Phi_{\text{obs}} = \Phi_{\text{int}}$, which means no attenuation in extragalactic medium, and $\alpha = 1$ gives $\Phi_{\text{obs}} = e^{-\tau_{\text{SL}}} \Phi_{\text{int}}$.

For constraining the α -parameter apriori, we use a broad uniform distribution:

$$P(\alpha) = \mathcal{U}[1.5; 1.5]. \quad (4.6)$$

Independence of these simple priors (4.4) and (4.6) allows to modify the formula (3.1):

$$P(\alpha, |D_k) \propto \mathcal{L}(D_k | \alpha, \boldsymbol{\theta}_k) P(\alpha) P(\boldsymbol{\theta}_k) \propto \mathcal{L}(D_k, \alpha, \boldsymbol{\theta}_k) \mathbf{1}(\alpha \in I_\alpha) \mathbf{1}(\boldsymbol{\theta} \in I_{\boldsymbol{\theta}}).$$

This can also be rewritten in logarithmic space, as required by `emcee`:

$$\ln P(\alpha, \boldsymbol{\theta}_k | D_k) = \ln \mathcal{L}(D_k) = \mathcal{L}(D_k, \alpha, \boldsymbol{\theta}_k) + \ln \mathbf{1}(\alpha \in I_\alpha) + \sum_{j=1}^n \ln \mathbf{1}(\theta_j \in I_{\theta_j}) + \text{const}, \quad (4.7)$$

where under logarithm, indicator function degenerates into $\ln \mathbf{1}(x \in A) = \begin{cases} 0, & \text{if } x \in A, \\ -\infty, & \text{else.} \end{cases}$

Sampling results

By sampling a distribution for for a distant high energy peaked blazar (HBL) 1H 1013 + 498 ($z = 0.212$), by arranging 32 walkers of 10000 steps each, we obtain the marginalized posterior distribution $P(\alpha|D)$, shown parameter distributions, shown in Figure 8.

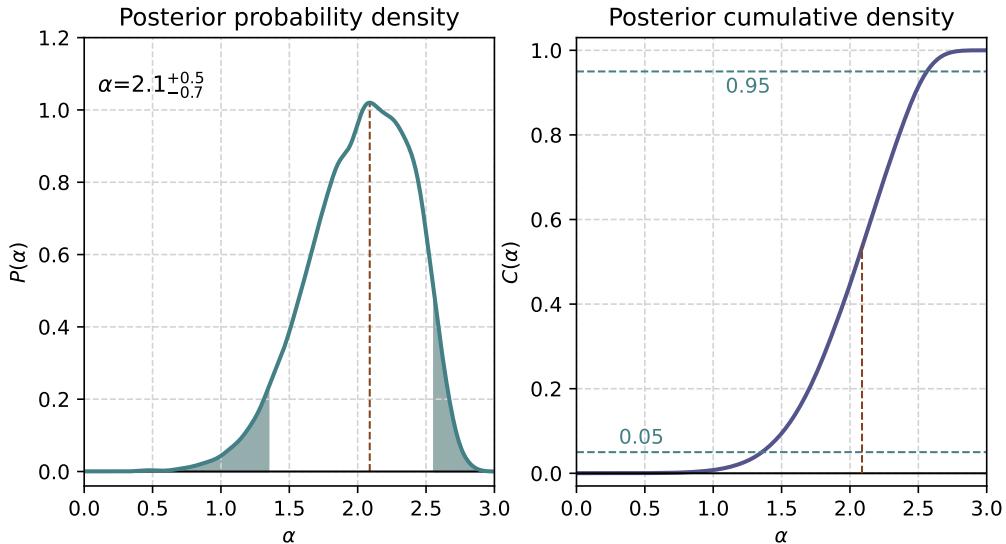


Figure 8: Scaling parameter α posterior density (left) and cumulative distribution function (right)

For the chosen source, we obtain $\alpha = 2.1^{+0.5}_{-0.7}$ at the 90% confidence level, which is consistent with the results of the same procedure applied to this source. To ensure the model's reliability, we also plotted the predicted intrinsic spectra along with the corresponding observed spectrum fit (see Figure 9) and compared them with [19].

With the sampling procedure for a single source established and validated, we proceed to defining constraints on the EBL density based on a set of observations from STeVECat.

4.3 EBL density constraints

In accordance with (3.4), to constrain the EBL normalization α in the one-parameter model (4.5), we calculate the posteriors $P(\alpha|D_k)$ for each source and construct the total posterior $P(\alpha|\mathcal{D})$ by multiplication. Figure 10 presents the distribution of single-source posterior mean and variance values.

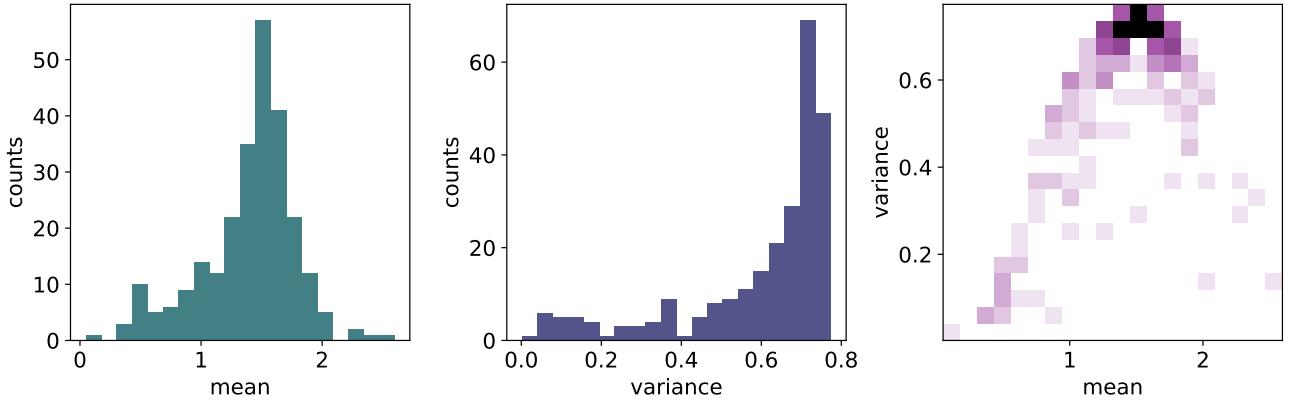


Figure 10: Means and variance distributions and correlation for a EBL scaling parametrization with STeVECat data.

The final step turns out to be less trivial, than it seems, as $P(\alpha|D_k)$ is not a continuous, but a sampled distribution. To proceed with the product (3.4), we use Gaussian Kernel Density Estimation (KDE) (4.8) of probability density for each source to get $P(\alpha|D_k)$ as a continuous function.

$$\hat{f}(x) = \frac{1}{nh} \sum_{l=1}^L K\left(\frac{x - x_l}{h}\right), \quad (4.8)$$

where $K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$ is the Gaussian kernel, and h controls smoothness.

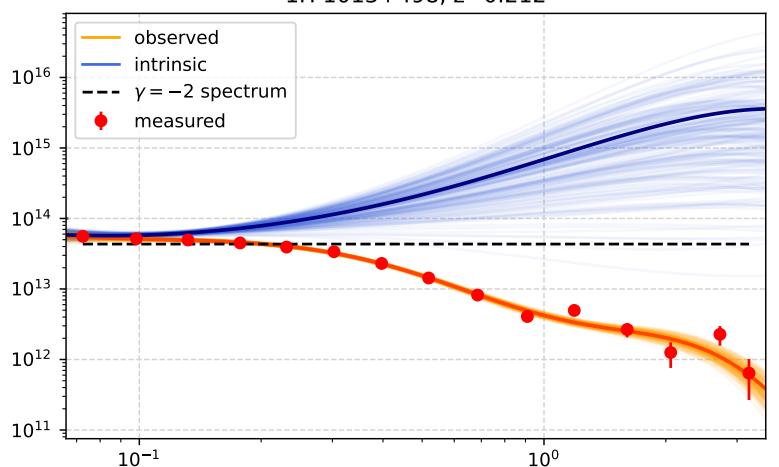


Figure 9: Measured (in red), sampled intrinsic (blue) and observed (orange) spectra of the 1H 1013 + 498

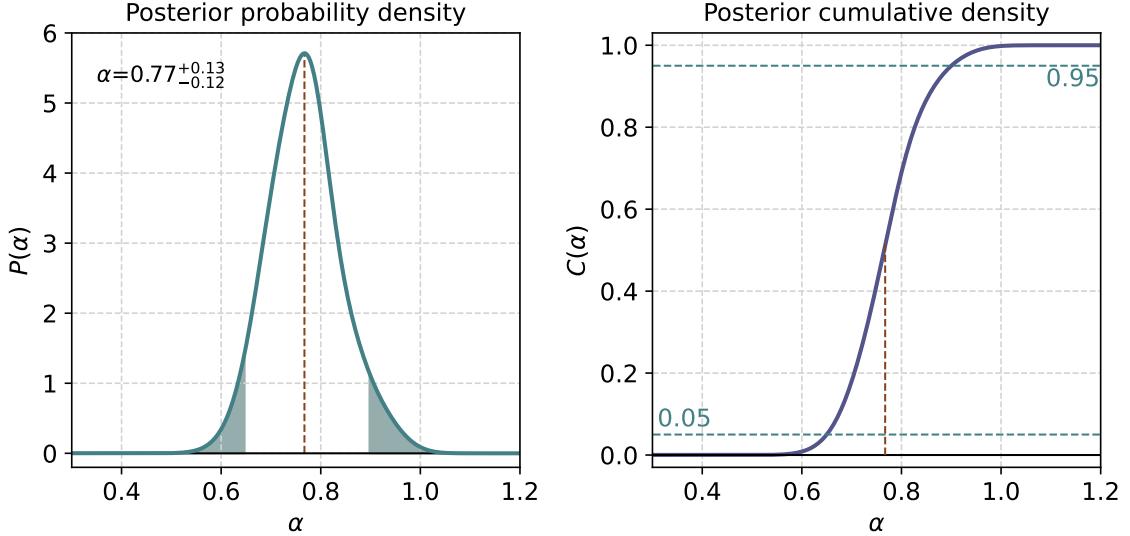


Figure 11: Scaling parameter α posterior probability density (left) and cumulative distribution function (right), computed from 258 independent samplings with 32 walkers each, 10000 steps per walker for the STeVCat observations having more than 5 experimental points.

The result of the above-described analysis is shown in Figure 11. By running the MCMC algorithm on 258 sources from STeVCat with more than five experimental points per observation, we reconstructed the following result at the 90% confidence level:

$$\alpha = 0.77^{+0.13}_{-0.12}.$$

The true value $\alpha = 1$ lies within the twice-scaled credibility interval, indicating no significant difference between the spectra. Additionally, we observe that 0 is excluded from the region within six lower errors, providing strong evidence that the EBL signature is clearly present in the STeVCat data.

5 Multiparametric model of the EBL density

The single-parameter EBL scaling approach relies on a specific model, which can be either theoretical [1] or phenomenological [19]. In this section, we discuss a method for constructing a model-independent estimation of the EBL density.

5.1 Choosing a kernel

For parametrization of the EBL specific intensity the following approach is used: kernel functions \mathcal{K}_i are chosen and total specific intensity is defined as a linear combinations

$$\lambda I_\lambda(l) = \sum_{k=1}^K a_i \mathcal{N}_k(l) \quad (5.1)$$

This approach allows to simplify a computationally expensive de-convolution procedure to solving a number of linear equations with unknown set of coefficients $\{a_i\}_{k=1}^K$.

Some authors choose gaussian functions as kernel functions for the specific intensity, which correlates strongly with the gaussian KDE approach 4.8, used in the previous section.

$$\mathcal{N}_k^0(l) = \mathcal{N}(l; l_k, \sigma_l) = \exp \left\{ -\frac{(l - l_k)^2}{2\sigma_l^2} \right\} \quad (5.2)$$

Their method is clear and produces a final distribution that closely matches experimental measurements. However, the resulting curve shows minor oscillations due to the parabola-based parametrization. To eliminate these oscillations while preserving the method's advantages, we will explore alternative basis functions.

B-splines

A B-spline of order $p + 1$ is a collection of piece-wise polynomial functions $B_{i,p}(t)$ of degree p in a variable t defined left-continuous. The values of t where the pieces of polynomial meet are known as knots, denoted t_0, t_1, \dots, t_m and sorted into non-decreasing order.

We use $B_{i,3}$, the most common set of basis functions, which are expressed in the following way:

$$\begin{pmatrix} B_{-2,3} \\ B_{-1,3} \\ B_{0,3} \\ B_{1,3} \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 8 & 12 & 6 & 1 \\ 4 & 0 & -6 & -3 \\ 4 & 0 & -6 & 3 \\ 8 & -12 & 6 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ u \\ u^2 \\ u^3 \end{pmatrix} \quad (5.3)$$

where $u = l - l_k$, and l_j is a coordinate of a grid knot, which gives

$$\mathcal{B}(l; l_k) = \sum_{i=-2}^1 B_{i,3}(l) \cdot \mathbf{1}(l \in [l_{j+i}, l_{j+1+i}]), \quad (5.4)$$

where $\mathbf{1}(l \in A)$ is a standard indicator function of the set A .

Hence, the EBL density is represented as

$$n_\epsilon(l) = \sum_{k=1}^K a_k \mathcal{B}(l; l_k) \quad (5.5)$$

From a parametric investigation, we concluded, that the optimal number of kernel functions are 8 and 17, as they provide the strongest drops in fitting errors (see Figure 12).

To show the differences between the (5.2) and (5.4) for the same set of coefficients $\{a_k\}_{k=1}^K$ for $K = 17$, we compared them in Figure 13. In comparison to a very smooth B-Spline approximation, the ExpParabolic curve has noticeable bumps that make the shape rather complex. This is not an issue of visualization in this particular work, similar portuberation features are seen, for example, in [7].

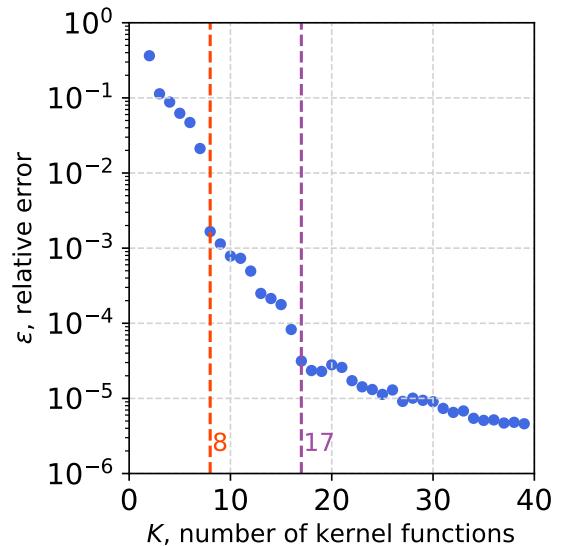


Figure 12: Search for the optimal number of B-Spline kernel functions for Saldana-Lopez density fitting

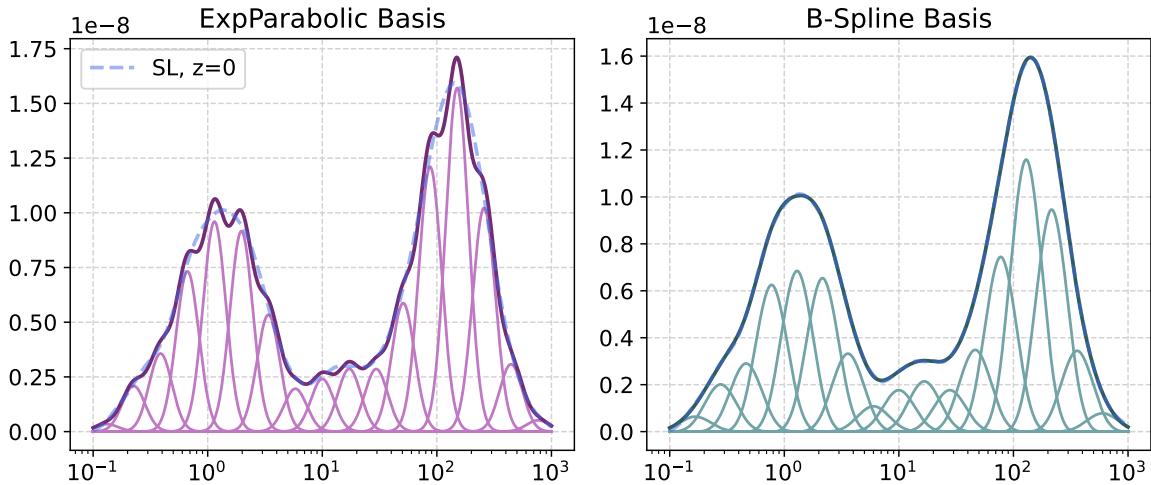


Figure 13: ExpParabolic (left, purple) and B-spline (right, green) basis behavior comparison for fitting the Saldana-Lopez (blue, dashed) density distribution, $K = 17$.

5.2 Taking redshift into account

The fitting pattern described in paragraph 5.1 is suitable for a 1D approximation. However, the EBL intensity $[\lambda I_\lambda]$ depends not only on wavelength, but also on redshift.

To simplify the calculations, one uses the variable separation-approach, considering the energy-dependence and energy

$$[\lambda I_\lambda](\lambda', z) = [\lambda I_\lambda](\lambda, 0) \cdot f_{\text{evol}}(z), \quad \lambda = \lambda'(1 + z) \quad (5.6)$$

The standard choice of the evolution function is $f_{\text{evol}}(z) = (1 + z)^{4-\chi}$ with $1 < \chi < 2$ [7].

However, an analysis of experimental observations [17] indicates that this approach fails to achieve the expected level of accuracy across a broad redshift range, $0 < z < 1$, which encompasses the sources of interest.

For this reason, we decided to complement the evolution function with a slow energy dependence (5.7), which can be physically justified by the origin of different EBL components [1].

$$f_{\text{evol}}(z, \lambda) = (1 + z)^{-\chi_0 + \chi_1 \ln(\lambda/\text{mkm})} \quad (5.7)$$

Comparison with the Saldana-Lopez EBL model shows that the addition of higher order terms of the form $\chi_n \ln^n(\lambda/\text{mkm})$ does not increase significantly the quality of the fit. The comparison of the two aforementioned approaches is given in Figure 14. The one-parameter description fails to follow the profile of the EBL density at redshifts significantly different from 0, whereas the two-parameter model (5.7) allows to fit the changing shape of the distribution both in optical and infrared regions.

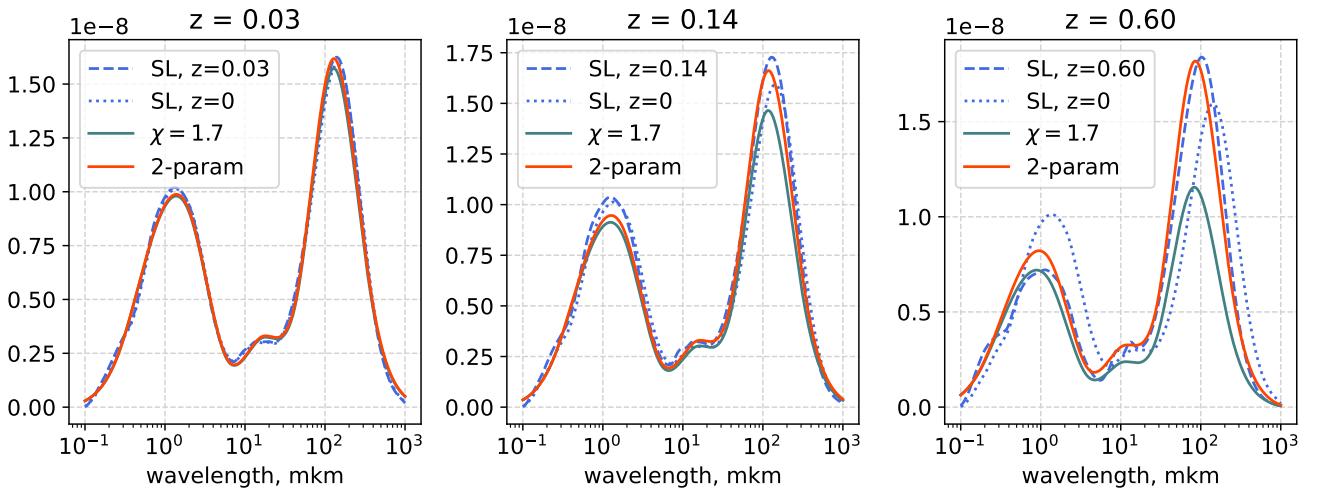


Figure 14: Saldana-Lopez EBL density at $z = 0$ (blue, dotted), its evolution (blue, dashed) and its parametrizations (5.6) in green, and (5.7) in red.

5.3 Setting the priors

Bayesian approach (3.4) requires to set the priors for the set of EBL density parameters. For that, we approximate the Saldana-Lopez profile with the B-Spline kernel, get the coefficients $\{\alpha_j^{\text{SL}}\}_{j=1}^8$. Then we calculate the mean value $\langle \alpha^{\text{SL}} \rangle$, maximal and minimal values $\alpha_{\max}^{\text{SL}}$, $\alpha_{\min}^{\text{SL}}$. The mean value of the prior stays fixed at $\langle \alpha^{\text{SL}} \rangle$, and its width is defined as following:

$$\delta\alpha_h = \max(\alpha_{\max}^{\text{SL}} - \langle \alpha^{\text{SL}} \rangle, \langle \alpha^{\text{SL}} \rangle - \alpha_{\min}^{\text{SL}}) \cdot (1 + h),$$

where $h \in [1, 2]$ is the roughness parameter.

The following formula is used to set the prior for all density profile parameters:

$$\alpha \sim \mathcal{U}[\langle \alpha^{\text{SL}} \rangle, \delta\alpha_h] \quad (5.8)$$

For numerical application we took $h = 0.5$.

5.4 EBL density estimations

Having a parametric model-independent description of the EBL (5.4), with a fixed evolution function (5.7), and priors given by (4.4) and (5.8), we were ready to start the EBL posterior distribution sampling with MCMC algorithm, implemented to `emcee`.

To obtain the posterior probability distribution, we act as follows [19]:

1. For each source k we produce 1D and 2D marginal samples of the $\{P(\alpha_i|D|k)\}$ and the $\{P(\alpha_i, \alpha_j|D_k)\}$. As in paragraph 4.3, to obtain the distributions we again use the gaussian-KDE, but now both in 1D and 2D.
2. From (3.4) we calculate the global posterior distributions $P(\alpha_i|\mathcal{D})$ and $P(\alpha_i, \alpha_j|\mathcal{D})$.
3. Using 1D distributions, we calculate $\langle\alpha_j\rangle$ and σ_i^2 ; with the 2D distributions, we compute the covariance matrices from $\langle\alpha_i\rangle$, $\langle\alpha_j\rangle$, σ_i , σ_j .
4. Approximate posterior with multivariate gaussian distribution.

For sampling we used two different initial conditions: in the first one, each parameter was set to start from $\langle\alpha^{\text{SL}}\rangle$ (results are shown in Figure 15 in red); in the second one initial sampling values were set to be equal to α_j^{SL} .

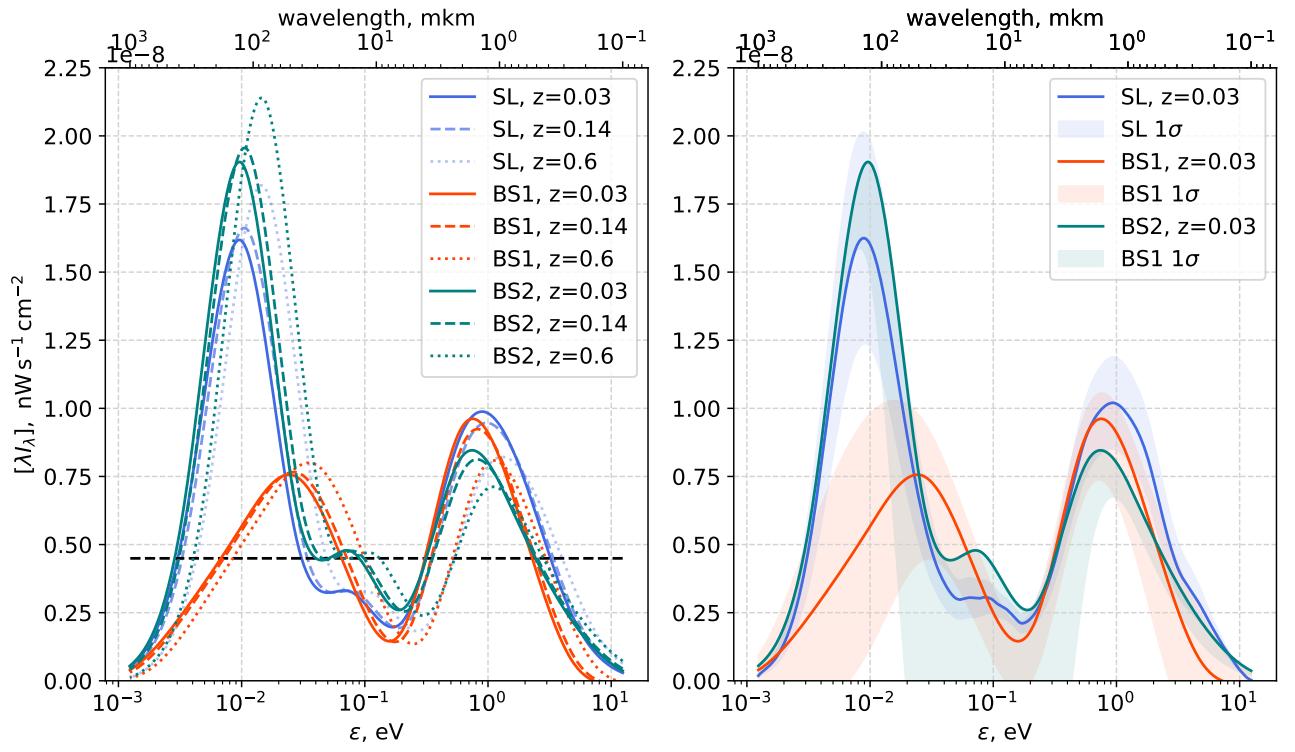


Figure 15: Left: the Saldana-Lopez EBL profile (blue), the B-Spline-based EBL profile for different redshifts for $\langle\alpha^{\text{SL}}\rangle$ initial conditions (red), and the profile for α_j^{SL} starting points (green). Right: the Saldana-Lopez EBL profile and obtained B-Spline fit with corresponding errors at 1σ confidence level for $z = 0.03$

Again, 32 walkers were used, but this time we increased the number of steps to 40,000 for each source, which led to a significant rise in computational time, from 2 minutes (in the one-parameter model) to 20 minutes per source.

Figure 15 presents the results after running 40,000 steps of the MCMC. The optical contribution to the EBL (COB) was successfully fitted, and the results show a strong correlation with the Saldana-Lopez model. However, in the low-energy infrared region (CIB), a significant discrepancy is evident. This can be explained by the fact that the CIB photons are significantly lower in energy, meaning only ultrahigh-energy gamma-ray photons can effectively interact with and be absorbed by this background. On the other hand, intrinsic source spectra are usually power-laws in energy, which makes ultrahigh-energy photon density decrease with energy, thus leading to a lower sensitivity of the observed spectra to CIB in comparison to COB photons.

Another remark about the sampling results can be made. Theoretically, Markov Chain Monte Carlo (MCMC) sampling is expected to be independent of the initial data as $n_{\text{steps}} \rightarrow \infty$. However, we observe that the two different starting points of the MCMC walkers lead to different final posterior distributions. This indicates that the sampling was stopped before reaching the algorithm's correlation time τ_C . For a model with p parameters, the correlation time asymptotically scales as $O(p^3)$ for linearly independent parameters and $O(p^4)$ for strongly correlated ones [19].

In the one-parameter sampling for each source, there were 5 parameters (1 for the EBL density and 4 for the source spectrum). In the multi-parameter model, there are 12 highly-correlated parameters, which results in a scaling factor between 15 and 35 in comparison with the first sampling. Running 258 chains with 300,000 steps each turned out to be unrealistic on a personal computer, highlighting the need for a computational cluster to achieve more satisfactory sampling results.

Summary

In the present report, we investigated the indirect method for measuring the EBL density through gamma-ray observations of distant high-energy gamma-ray sources.

The first qualitative constraint on the EBL density was derived using theoretical expectations of intrinsic source spectrum shapes. From the discussion in paragraph 2.2, we concluded that high values of the EBL density would violate the strong physical requirement for spectrum curvature to be non-positive, as intrinsic spectra are generally expected to have negative curvature in the absence of significant absorption.

To move toward quantitative constraints, we applied Bayesian methods and Markov Chain Monte Carlo sampling to the STeVECat high-energy gamma-ray observations, a database of gamma-ray source data gathered by multiple telescopes. This allowed for a more robust analysis by taking into account uncertainties and model parameters in a systematic manner.

We employed a single-parameter approach to model the Saldana-Lopez EBL density with a scaling factor. The resulting value, $\alpha = 0.77^{+0.13}_{-0.12}$, was found to be consistent with the Saldana-Lopez density and provides strong support for the presence of extragalactic background light absorption.

In section 5, we introduced a model-independent approach to further constrain the EBL density. Using a functional basis constructed from third-degree B-splines, we fitted the decomposition coefficients to the observed data. This approach produced results that were in good agreement with the Saldana-Lopez model in the optical region but revealed significant differences in the low-frequency infrared region.

This discrepancy can be attributed to the fact that the Cosmic Infrared Background photons are much less energetic, and only ultrahigh-energy gamma-ray photons are able to efficiently interact with and be absorbed by this background, further emphasizing the role of high-energy photons in probing the EBL.

The present work provides a clear direction for further development. The next step in this research will be to use the established methods to search for signatures of EBL absorption in the Cherenkov Telescope Array internal data challenge, where more extensive and high-quality datasets will allow for refined constraints on the EBL density.

References

- [1] Eli Dwek and Frank Krennrich. “The extragalactic background light and the gamma-ray opacity of the universe”. In: *Astroparticle Physics* 43 (Mar. 2013), pp. 112–133. DOI: 10.1016/j.astropartphys.2012.09.003.
- [2] A. I. Nikishov. “Absorption of High-Energy Photons in the Universe”. In: *Soviet Physics JETP* 14 (1962). Translated from Zh. Eksp. Teor. Fiz. 41, 570–574 (1961), pp. 394–398.
- [3] H.E.S.S. Collaboration. *H.E.S.S. - The High Energy Stereoscopic System*. Accessed: 2025-03-11. 2025. URL: <https://www.mpi-hd.mpg.de/HESS/>.
- [4] MAGIC Collaboration. *MAGIC - Major Atmospheric Gamma Imaging Cherenkov Telescopes*. Accessed: 2025-03-11. 2025. URL: <https://magic.mpp.mpg.de/>.
- [5] VERITAS Collaboration. *VERITAS - Very Energetic Radiation Imaging Telescope Array System*. Accessed: 2025-03-11. 2025. URL: <https://veritas.sao.arizona.edu/>.
- [6] Fermi-LAT Collaboration. *Fermi Large Area Telescope (Fermi-LAT)*. Accessed: 2025-03-11. 2025. URL: <https://fermi.gsfc.nasa.gov/science/instruments/lat.html>.
- [7] J. Biteau and D. A. Williams. “The Extragalactic Background Light, the Hubble Constant, and Anomalies: Conclusions from 20 years of TeV Gamma-ray Observations”. In: *The Astrophysical Journal* (2015). DOI: 10.1088/0004-637X/812/1/60.
- [8] Lucas Gréaux, Jonathan Biteau, and Mireia Nievas Rosillo. *The Cosmological Optical Convergence: Extragalactic Background Light from TeV Gamma Rays*. 2024. arXiv: 2410.07011 [astro-ph.HE].
- [9] Cherenkov Telescope Array Observatory. *Cherenkov Telescope Array Observatory (CTA)*. Accessed: 2025-03-11. 2025. URL: <https://www.ctao.org/>.
- [10] Lucas Gréaux et al. *STeVECat, the Spectral TeV Extragalactic Catalog*. Version 1. Zenodo, July 2023. DOI: 10.5281/zenodo.8152245.
- [11] G.B. Rybicki and A.P. Lightman. *Radiative Processes in Astrophysics*. WILEY-VCH Verlag GmbH & Co. KGaA, 1986.
- [12] Barbara Sue Ryden. *Introduction to cosmology*. Addison-Wesley, 2003.
- [13] Robert J. Gould and Gérard P. Schréder. “Pair Production in Photon-Photon Collisions”. In: *Phys. Rev.* 155 (5 Mar. 1967), pp. 1404–1407. DOI: 10.1103/PhysRev.155.1404.
- [14] J.M.Jauch and F.Rohrlich. *The Theory of Photons and Electrons*. Addison-Wesley Publishing Company, Inc., Reading Massachusetts, 1955.
- [15] S. K. Andrews et al. “Modelling the cosmic spectral energy distribution and extragalactic background light over all time”. In: *Monthly Notices of the Royal Astronomical Society* 474.1 (Nov. 2017), pp. 898–916. DOI: 10.1093/mnras/stx2843.
- [16] Alberto Franceschini and Giulia Rodighiero. “The extragalactic background light revisited and the cosmic photon-photon opacity”. In: *Astron. Astrophys.* 603 (2017), A34. DOI: 10.1051/0004-6361/201629684.
- [17] A. Saldana-Lopez et al. “An observational determination of the evolving extragalactic background light from the multiwavelength HST/CANDELS survey in the Fermi and CTA era”. In: *Monthly Notices of the Royal Astronomical Society* 507 (2021), pp. 5144–5160. DOI: 10.1093/mnras/stab2393.
- [18] Daniel Foreman-Mackey et al. “emcee: The MCMC Hammer”. In: *Publications of the Astronomical Society of the Pacific* 125.925 (Mar. 2013), pp. 306–312. DOI: 10.1086/670067.
- [19] Lucas Gréaux. “Gamma-ray cosmology – Conclusions from three decades of extragalactic gamma-ray astronomy and perspectives for CTAO”. PhD thesis. Université Paris-Saclay, Sept. 2024. URL: https://perso.crans.org/greaux/documents/Thesis_Greaux.pdf.