Interpretation of the diffuse astrophysical neutrino flux in terms of the blazar sequence

Andrea Palladino,[1] Xavier Rodrigues,[1] Shan Gao,[1] and Walter Winter[1]

[1] *DESY, Platanenallee 6, 15738 Zeuthen, Germany*

(Dated: August 23, 2018)

ABSTRACT

We study if the diffuse astrophysical neutrino flux can come from blazar jets – a subclass of Active Galactic Nuclei (AGNs) – while it, at the same time, respects the blazar stacking limit based on source catalogs and is consistent with the observation from TXS 0506+056. We compute the neutrino flux from resolved and unresolved sources using an averaged, empirical relationship between electromagnetic spectrum and luminosity, known as the *blazar sequence*, for two populations of blazars (BL Lacs and FSRQs). Using a source model with realistic neutrino flux computations, we demonstrate that blazars can indeed power the diffuse neutrino flux at the highest energies and obey the stacking limit at the same time, and we derive the conditions for the baryonic loading (proton versus γ-ray luminosity) evolving over the blazar sequence. Under the hypothesis that low-luminosity blazars power the diffuse astrophysical neutrino flux, we find that the dominant contribution of the diffuse flux up to PeV energies must come from unresolved BL Lacs with baryonic loadings larger than about 10^5 – while only a very small contribution may come from resolved high-luminosity BL Lacs or FSRQs, which can be directly tested by the stacking limit. We find that the blazar TXS 0506+056 is on the verge of these populations in our baseline scenario, at a relatively high luminosity and redshift; as a consequence we predict about 0.4 γ-ray-neutrino associations per year from the whole population, dominated by BL Lacs with $L_\gamma \simeq 10^{45}$ erg/s and $z \sim 0.1$.

1. INTRODUCTION

The discovery of a diffuse flux of high-energy neutrinos by IceCube (Aartsen et al. 2013) has triggered many questions, one of the most relevant being "what is the source of the high-energy neutrinos detected by IceCube?". Several candidate classes have been proposed to interpret the IceCube data, such as blazars (Protheroe 1997; Essey et al. 2010; Murase et al. 2014; Padovani et al. 2016a; Esmaili et al. 2016; Righi et al. 2018a,b; Murase et al. 2018), Gamma-Ray Bursts (GRBs) (Paczynski & Xu 1994; Waxman & Bahcall 1997; Vietri 1998; Meszaros & Waxman 2001; Hummer et al. 2012; Murase & Ioka 2013), starburst galaxies (Loeb & Waxman 2006; Stecker 2007; Tamborra

andrea.palladino@desy.de

xavier.rodrigues@desy.de

shan.gao@desy.de

walter.winter@desy.de

et al. 2014), and dark matter decay (Esmaili & Serpico 2013; Chianese et al. 2016). Some of them are already strongly constrained by measurements, based on the lack of correlations between high-energy neutrinos and known sources. For example, it is known that the contribution of GRBs to the IceCube neutrino signal is of the order of a few percent (Aartsen et al. 2017a), whereas the contribution of starburst galaxies may be larger but is also insufficient to explain the observed neutrinos, due to constraints from the extragalactic γ-ray background (Murase et al. 2013; Bechtol et al. 2017). A certain contribution to the neutrino flux may also come from the Galactic plane (Palladino & Vissani 2016; Palladino et al. 2016) but the latest measurements of IceCube and ANTARES suggest that this contribution cannot be greater than $\sim 10\%$ (Albert et al. 2018).

In this work, we study the contribution of blazars to the astrophysical neutrino flux. Blazars are a sub-class of Active Galactic Nuclei (AGNs), which are supermassive black holes that accrete matter from the host galaxy, launching relativistic jets and emitting non-thermal ra-

diation. In the case of blazars, the relativistic jet points in the direction of Earth. Assuming that cosmic rays (CRs) may be accelerated in the jet, we calculate the neutrino emission from CR interactions with the blazar photon fields. To obtain the photon spectrum we rely on the so-called *blazar sequence* (Ghisellini et al. 2017), an empirical relation between the luminosity of a blazar and the photon spectrum emitted.

Considering the high-energy starting events (HESE) in IceCube, the contribution of resolved blazars to the astrophysical neutrino flux cannot be larger than $\sim 20\% - 25\%$ (Aartsen et al. 2017b) based on the missing association with sources in γ-ray catalogs, or even less based on theoretical considerations (Palladino & Vissani 2017). On the other hand, there have been indications of associations of individual neutrino events with AGNs (Resconi et al. 2017; Padovani et al. 2016b; Kadler et al. 2016), which means that it is plausible that a few of IceCube's events stem from resolved blazars. More recently, direct evidence has been found of a correlation between IceCube neutrinos and the object TXS 0506+056 (Aartsen et al. 2018a,b), which represents a breakthrough discovery in multi-messenger astrophysics. This observation may support the AGN origin of the diffuse neutrino flux, possibly powered by unresolved objects.

In this work we show that low-luminosity blazars can provide the dominant contribution to the high-energy neutrinos with energy between a few hundreds of TeV and a few PeV, while the contribution of very bright sources to the neutrino flux must be highly suppressed in order to respect the blazar stacking limit (Aartsen et al. 2017b). This result allows us to draw conclusions about the possible baryonic loading of blazars evolving over the blazar sequence. We also discuss explicitly the role of TXS 0506+056, and what we can learn from the population study for future such observations.

In Section 2.1 we evaluate the neutrino spectra from blazars numerically, taking into account the differences between the two blazar sub-classes, namely BL Lacs and Flat-Spectrum Radio Quasars (FSRQs). In Section 2.2 we discuss the cosmic evolution of BL Lacs (Ajello et al. 2014) and FSRQs (Ajello et al. 2012), which is necessary to compute the expected diffuse neutrino flux at Earth. We present our results in Section 3, where we compute the diffuse flux of neutrinos under three different hypotheses: *i)* constant baryonic loading, *ii)* proportionality between the luminosity of blazars in neutrinos (L_ν) and in γ-rays (L_γ), and *iii)* a baryonic loading that scales with the source luminosity. A discussion is presented in Section 4, focusing on *i)* source characteristics predicted by the model as a function of γ-ray

luminosity, *ii)* implications of the multiplet constraint, *i.e.*, the non-observation of two neutrino events from the same source, *iii)* implications of our model to potential neutrino sources such as TXS 0506+056 and *iv)* PKS 0502+049 (He et al. 2018). We conclude the work with Section 5.

2. METHODS

We now describe the methods used to compute the neutrino flux from blazars. In Section 2.1 we present the radiation model for the calculation of neutrino spectra across the blazar sequence, in Section 2.2 we address the cosmic evolution of blazars, in Section 2.3 we explain the procedure to calculate the diffuse neutrino flux observed at Earth, and in Section 2.4 we introduce the special blazar TXS 0506+056.

2.1. *Neutrinos from the blazar sequence*

The calculation of neutrino production in blazars follows closely that in Rodrigues et al. (2018), which is based on the model introduced in Murase et al. (2014). Some details of the present calculation, including key differences to the previous implementation of the model, are discussed in Appendix A. The basic assumption is that CR protons are accelerated in the relativistic jet of the blazar during a period of flaring activity. We assume the same jet speed for all the sources, corresponding to a Lorentz factor $\Gamma = 10$. We assume the source is in a persistent flaring state, corresponding to a duty cycle of 100% (see Section 2.4 for a discussion on the implications of different duty cycle assumptions). Cosmic rays are injected isotropically during a period of time corresponding to a typical flare, $t' = 10^6$ s,[1] in a spherical region of size $r'_{\rm blob} = ct'_{\rm flare}$. This is the region represented in dark red in Fig. 1.

Cosmic rays are assumed to undergo second-order Fermi acceleration, with an acceleration rate given by

$$t'^{-1}_{\rm acc}(E', B') = \eta \frac{ZeB'c}{E'}, \qquad (1)$$

where Z is the atomic number of the nucleus being accelerated, E' is its energy, and B' is the magnetic field strength in the jet (assumed to scale with the blazar luminosity, see Appendix A). Moreover, η is the acceleration efficiency, which is assumed to be $\eta = 10^{-3}$ in this work. That low value of acceleration efficiency is motivated by matching the primary energy to the cutoff of the IceCube neutrinos, and implies that there is

---

[1] Primed quantities will refer to the rest frame of the jet plasma, while unprimed quantities may refer to the source or the observer's frame.
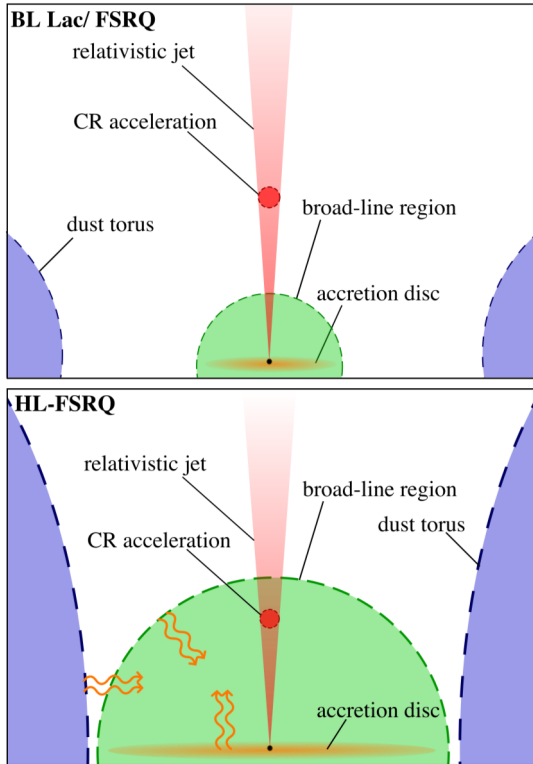
Figure 1. Schematics of the blazar model (not to scale). Top: BL Lac or low-luminosity FSRQ. In these sources the relativistic jet blob, where CRs are injected, lies outside the broad line region (BLR), and therefore CRs hardly interact with the photon fields produced by the accretion disk or the gas surrounding the BLR – which are de-boosted in the blob frame. Bottom: high-luminosity FSRQ (HL-FSRQ), where the acceleration region lies in the BLR and the external photon spectra are relevant for the interactions in the blob frame. If we assume that the luminosity of the source increases with the size of the accretion disk, then in FSRQs brighter than $L_\gamma = 2.4 \times 10^{48}$ erg/s the jet blob lies inside the BLR in this model, and the efficiency of photo-hadronic interactions is increased by the external photon fields of the FSRQ. Figure adopted from Rodrigues et al. (2018).

no connection to ultra-high-energy cosmic rays (as *e.g.* in Rodrigues et al. (2018), where a higher value of η was used). We assume CRs are accelerated to a power-law energy distribution in the jet:

$$\frac{dN}{dE'} = E'^{-2} \exp\left(-\frac{E'}{E'_{\max}}\right) , \qquad (2)$$

where $E'_{\max}$ is the maximum energy achieved by the cosmic rays in the source. This is the energy at which cooling or interaction processes become more efficient than acceleration, or the acceleration becomes limited by the size of the blob.

The interaction of CRs with a target photon field is simulated using the NEUCOSMA code (Baerwald et al.

2012). The main process responsible for neutrino emission is photo-meson production, where in most cases the neutrinos produced carry around 5% of the energy of the primary. The amount of CRs injected in each source is a free parameter of the model; we quantify this quantity by means of the **baryonic loading** of the source, defined as

$$\xi \equiv \frac{L_{\mathrm{CR}}}{L_\gamma} , \qquad (3)$$

where L_{CR} is the total luminosity of injected CRs and L_γ is the γ-ray luminosity of the source, defined here in the range $0.1 - 100$ GeV, roughly corresponding to the Fermi-LAT observation range.

The spectral energy distribution (SED) of each source (used as the target photon field for hadronic interactions) is determined according to the most recent implementation of the blazar sequence (Ghisellini et al. 2017), an empirical relation based on the Fermi 3LAC blazar catalog (Ackermann et al. 2015) that attributes an average SED to each blazar luminosity L_γ, distinguishing between BL Lacs and FSRQs. The SEDs are shown in the left panels of Fig. 2: the upper panels refer to BL Lacs, and the bottom panels to FSRQs. All SEDs with $L_\gamma > 10^{48}$ erg/s and $L_\gamma < 10^{44}$ erg/s have been extrapolated by renormalizing the brightest and dimmest SEDs (respectively) provided in Ghisellini et al. (2017). As we can see in the top left panel of Fig. 2, high-frequency peaked BL Lacs (HBLs), *i.e.*, BL Lacs with synchrotron and inverse Compton peaks at higher energies, are generally dimmer sources, while low-frequency peaked BL Lacs (LBLs) are brighter. For FSRQs, on the other hand, there is no strong relationship between the frequency of non-thermal emission bands and the jet luminosity (bottom left panel of Fig. 2). Besides synchrotron and inverse Compton emission, FSRQs typically exhibit bright broad lines from atomic emission of the gas surrounding the accretion disk, as well as thermal bumps from an accretion disk and a dusty torus. If we assume that the size of the broad line region (BLR, green region in Fig. 1) increases with luminosity, then the radiation zone in the jet is contained within reach of the external photon fields for high-luminosity (HL) FSRQs (bottom panel of Fig. 1, *cf.*, Murase et al. (2014)). These external fields, seen in the SEDs of the brightest FSRQs in Fig. 2, enhance CR interactions both inside the jet and during the propagation of escaping CRs through the BLR and near the dusty torus. This is taken into account by means of two additional radiation regions in the modeling of high-luminosity FSRQs (see Appendix A).

In the right panels of Fig. 2 we show the emitted neutrino spectra obtained with our blazar model. The integral of the neutrino spectrum of each blazar yields
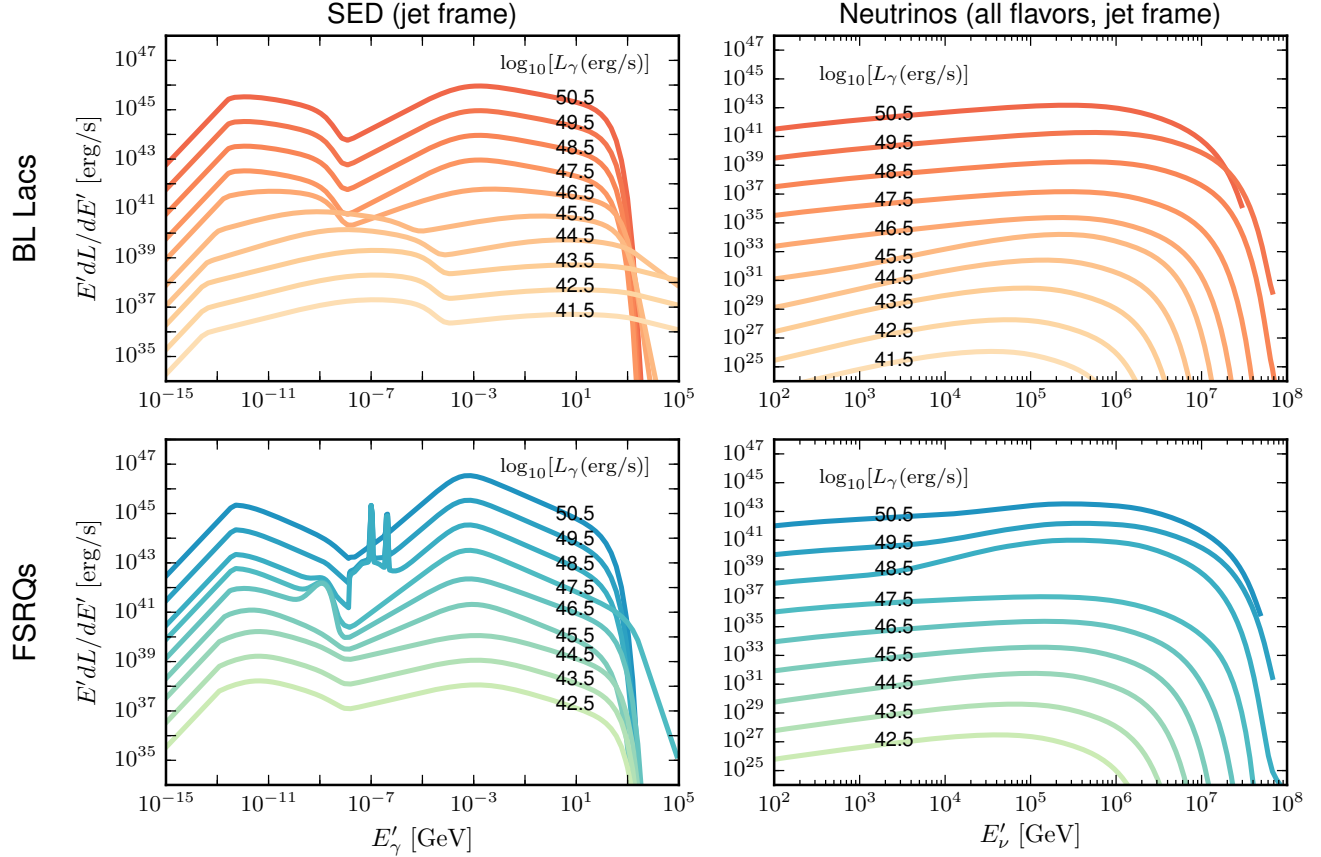
Figure 2. Left: spectral energy distributions (SEDs) for BL Lacs (top) and FSRQs (bottom) used in this work, given in the rest frame of the jet. The non-thermal SEDs are those from the blazar sequence (Ghisellini et al. 2017), and the external components energy spectra are based on Murase et al. (2014). All non-thermal SEDs with $L_\gamma > 10^{48}$ erg/s and $L_\gamma < 10^{44}$ erg/s have been extrapolated by renormalizing the brightest and dimmest SEDs (respectively) provided in Ghisellini et al. (2017). Right: neutrino luminosity spectra calculated for each blazar luminosity, for a baryonic loading of $\xi = 1$, also in the rest frame of the jet (see Appendix A for details). The label of each curve indicates the logarithm of the γ-ray luminosity of the jet L_γ in erg/s, in the source comoving frame.

the total neutrino luminosity of the source, L_ν. This quantity is directly proportional to the baryonic loading ξ of the source, since the emitted neutrinos originate from CR interactions (Fig. 2 refers to the special case $\xi = 1$). In general, the emitted neutrino luminosity L_ν of a blazar may be written as

$$L_\nu = \mathcal{K}\, L_\gamma, \qquad (4)$$

where $\mathcal{K} \equiv \epsilon_\nu \times \xi$ is the product of the neutrino production efficiency and the baryonic loading, defined in Eq. (3). The neutrino production efficiency $\epsilon_\nu \equiv L_\nu/L_{\rm CR}$ quantifies the amount of energy from CRs converted into neutrinos in the source (which is independent of the baryonic loading). The quantity $\mathcal{K} = L_\nu/L_\gamma$ will be used throughout this work to express the ratio between the neutrino and γ-ray luminosity of a given source.

Using the results of Fig. 2, we can calculate the neutrino production efficiency of each source considered in this work. This is shown in the top panel of Fig. 3 as a function of the γ-ray luminosity of the source. The neutrino production efficiency increases monotonically with the γ-ray luminosity of the blazar, since higher luminosities imply higher photon densities in the radiation zone in the jet. Note the abrupt increase in the efficiency of FSRQs with $L_\gamma \gtrsim 3 \times 10^{48}$ erg/s, which is due to interactions with external fields, producing additional neutrinos.

In the bottom panel of Fig. 3 we show the maximum energy achieved by protons accelerated in the jet, which is highest for blazars of $L_\gamma \sim 3 \times 10^{49}$ erg/s. Above this luminosity, photo-hadronic interactions of high-energy CRs dominate over acceleration, and the maximum CR energy starts decreasing with luminosity.
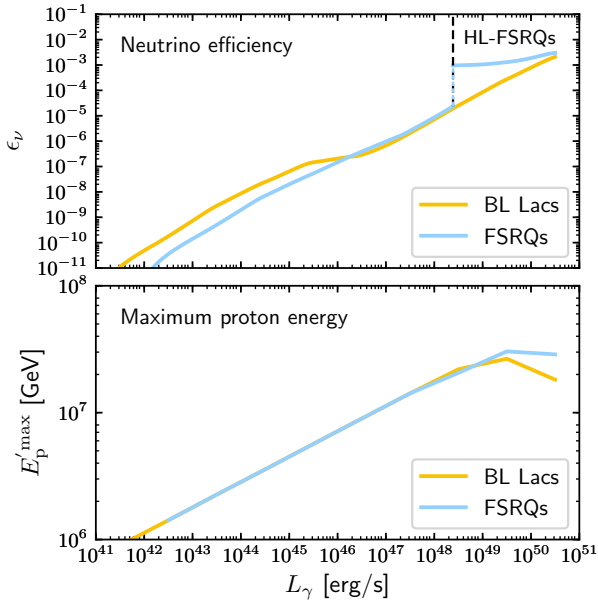
Figure 3. Top: neutrino production efficiency of BL Lacs (yellow) and FSRQs (blue) of the blazar sequence as a function of the γ-ray luminosity of the source. Bottom: maximum energy of accelerated protons in the jet of BL Lacs (yellow) and FSRQs (blue) as a function of luminosity. We assume that the radiation zone in the jet is exposed to external photon fields from molecular and thermal emission in the case of FSRQs with luminosity above 3×10^{48} erg/s (HL-FSRQs); this is responsible for the jump in neutrino efficiency (see Appendix A).

2.2. *Source evolution*

The calculation of the cumulative neutrino flux from blazars, reported in this work, is based on the distribution of BL Lacs and FSRQs in Ajello et al. (2014, 2012). In these two works the distribution of blazars is parametrized in order to reproduce the sources observed by Fermi-LAT and the observed diffuse γ-ray background (Abdo et al. 2010). The parametrization is also useful to evaluate the contribution of unresolved sources, *i.e.*, the sources below the Fermi-LAT sensitivity threshold (but expected from a theoretical point of view) that contribute to the diffuse γ-ray flux.

In Appendix B we give the details of the parameterization by Ajello et al., which we use in Section 2.3 to compute the cumulative neutrino flux from blazars. The distribution describes the number of sources per redshift and luminosity and is characterized by 12 parameters (which are different for BL Lacs and FSRQs), reported in Tab. 2.[2].
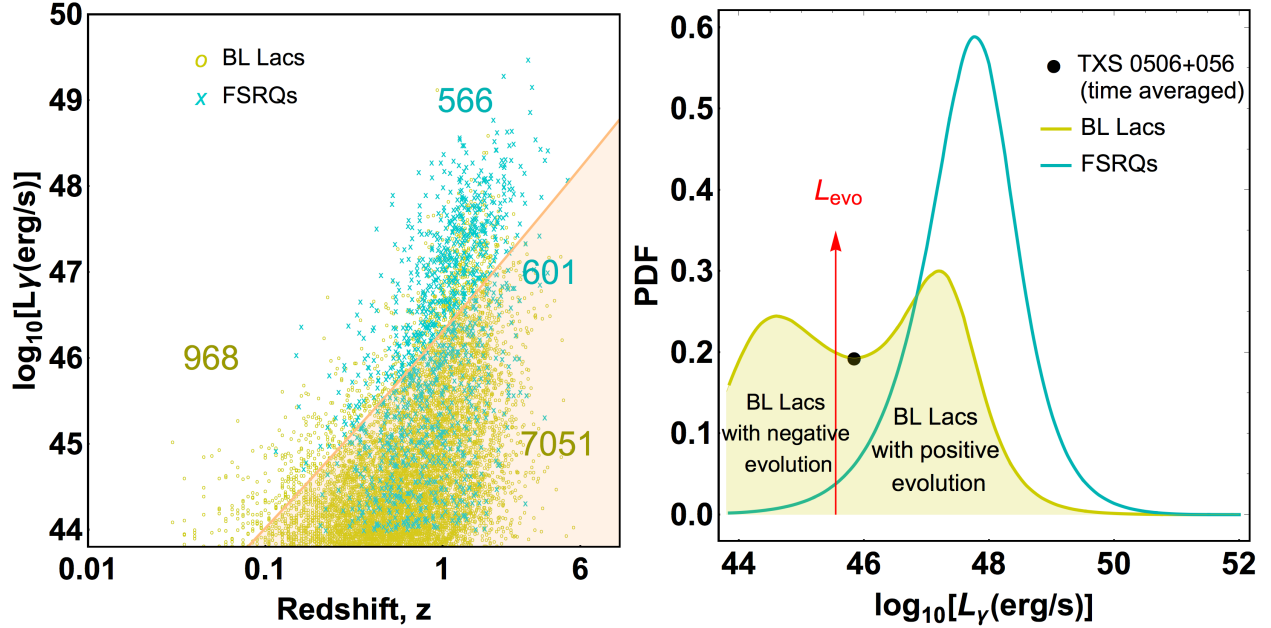
The distributions of BL Lacs and FSRQs are reported in Fig. 4. In these figures we represent a set of 9186 sources, obtained from a Monte Carlo extraction in agreement with the distribution presented in Appendix B. The total number of sources is already contained in the parametrization by Ajello et al. We represent the source evolution as a function of redshift z and luminosity $\ell \equiv \log_{10}[L_\gamma \ (\mathrm{erg/s})]$. The threshold that separates resolved and unresolved sources has been chosen to reproduce the ~ 1500 identified sources contained in the 3LAC catalog (Ackermann et al. 2015) and it corresponds to a flux $\phi_\gamma = 4 \times 10^{-12}$ erg cm^{-2} s^{-1}. It is interesting to notice that about 50% of FSRQs are resolved, whereas more than 85% of the expected BL Lacs are not resolved but they are expected to contribute to the total flux of γ-rays and neutrinos.

In the right panel of Fig. 4 we show the distribution of BL Lacs and FSRQs as a function of their luminosity, obtained integrating on the redshift z. We notice that most FSRQs have a luminosity between $L_\gamma = 10^{46}$ and 10^{50} erg/s, whereas BL Lacs are characterized by two populations with different characteristics: a low-luminosity population (roughly between $L_\gamma = 10^{44} - 10^{46}$ erg/s) and a high-luminosity population above $L_\gamma \simeq 10^{46}$ erg/s. These two populations of BL Lacs are characterized by different cosmological evolutions: the low-luminosity BL Lacs have a negative evolution, *i.e.*, many of these objects are nearby, while high-luminosity BL Lacs have a positive evolution, *i.e.*, are more abundant at high redshifts. Following the parametrization by Ajello et al. we find that the evolution changes from negative to positive when the luminosity is equal to $L_\gamma^{\mathrm{evo}} \simeq 3.5 \times 10^{45}$ erg/s. A summary of these characteristics is reported in Tab. 1, together with the relative contributions from different blazar classes to the resolved and unresolved γ-ray flux.

In this work we focus on astrophysical neutrinos, whereas the distribution by Ajello et al. was originally created to study the γ-ray emission. Indeed using this distribution, it is possible to evaluate the γ-ray flux expected from BL Lacs and FSRQs. As stated before, a non-negligible fraction of this γ-ray flux can be produced by unresolved sources. Therefore it is quite natural to expect a similar behavior for neutrinos, and this aspect will be discussed in depth in the next sections. Moreover, if the contribution of unresolved sources to the neutrino flux were large enough, the lack of correlations

---

[2] In Ajello et al. (2014, 2012) Eq. (21) is reported with the wrong signs of p_1 and p_2. This produces an incorrect distribution

which is particularly relevant for low-luminosity BL Lacs. In this work we use this function considering the correct signs. Please note that the erratum has been reported in a later work in 2015. See the second footnote of Ajello et al. (2015).

Figure 4. Left panel: the cosmic evolution of BL Lacs and FSRQs as a function of redshift z and logarithm of the γ-ray luminosity L_γ, following the distributions given in Ajello et al. (2014, 2012). The orange curve roughly separates resolved sources (above) from unresolved sources (below). The yellow and cyan numbers denote the number of resolved and unresolved BL Lacs and FSRQs respectively. Right: distribution of BL Lacs and FSRQs as a function of luminosity, obtained by integrating over the redshift. We notice that FSRQs are accumulated at high luminosity, whereas BL Lacs are characterized by two different populations. The luminosity $L_\gamma^{\mathrm{evo}} = 3.5 \times 10^{45}\mathrm{erg/s}$ is indicated.

between neutrinos and known γ-ray sources would become less problematic.

2.3. The expected flux of astrophysical neutrinos

The expected flux of astrophysical neutrinos produced by blazars can be determined using the neutrino flux from a single source (identified by its luminosity and its redshift) and the distribution of blazars in the universe.

The neutrino flux at Earth Φ_s produced by a single source with γ-ray luminosity L_γ at redshift z is given by the following expression:

$$\frac{d\Phi_s}{dE}(\ell, z, E, \xi(\ell)) = \frac{1}{4\pi D_c(z)^2}\left[\frac{1}{E}\frac{dL_\nu}{dE}(\ell, E(1+z))\right]\times\xi(\ell),$$

$$(5)$$

where $\xi(\ell)$ is the baryonic loading of the specific source, the neutrino luminosity spectra EdL_ν/dE are represented in Fig. 2 for a baryonic loading $\xi = 1$ and $D_c(z)$ is the comoving distance, defined in Appendix B. The relation between the total neutrino luminosity of the blazar and its γ-ray luminosity is given by Eq. (4).

The cumulative flux of neutrinos at Earth is given by convolving the single-source neutrino flux with the distribution of BL Lacs and FSRQs over L_γ and redshift z, as follows:

$$\frac{d\Phi_{\mathrm{tot}}}{dE}(E, \xi(\ell)) = \left[\int_{z_1}^{z_2}\int_{\ell_1}^{\ell_2}\frac{dN}{dzd\ell}\times\frac{d\Phi_s}{dE}(\ell, z, E, \xi(\ell))\ dzd\ell\right]_{\mathrm{BL\ Lacs}}$$
$$+ \left[\int_{z_1}^{z_2}\int_{\ell_1}^{\ell_2}\frac{dN}{dzd\ell}\times\frac{d\Phi_s}{dE}(\ell, z, E, \xi(\ell))\ dzd\ell\right]_{\mathrm{FSRQs}},$$

$$(6)$$

where the integration limits ℓ_1, ℓ_2, z_1 and z_2 are given in Appendix B. On the other hand, to calculate the contribution of resolved and unresolved blazars separately, the γ-ray luminosity must be integrated either in the range $[\ell_1, \ell_{\mathrm{vis}}]$ (unresolved) or $[\ell_{\mathrm{vis}}, \ell_2]$ (resolved), where ℓ_{vis} is the luminosity threshold for the observation by the Fermi-LAT, as discussed in section 2.2, corresponding to a flux of $\phi_\gamma^{\mathrm{vis}} = 4 \times 10^{-12}$ erg cm^{-2} s^{-1}. Therefore ℓ_{vis} is a function of the redshift, namely $\ell_{\mathrm{vis}}(z) = 4\pi \times \phi_\gamma^{\mathrm{vis}} \times D_c(z)^2(1+z)^2$.

It is important to remark that in the previous procedure it is implicitly contained the hypothesis that a source emits *always* an average luminosity L_γ.

2.4. Neutrinos from the blazar TXS 0506+056

On September 2017 a high-energy neutrino, event IceCube-170922A, has been detected by IceCube (Aartsen et al. 2018b). After this detection an alert was produced (see Aartsen et al. (2017c) for information on the

Table 1. Summary of some characteristics of BL Lacs and FSRQs: the evolution, the number of resolved and unresolved BL Lacs and FSRQs, and the resolved versus unresolved γ-ray flux contribution, listed separately for each population.

	Evolution	# resolved sources	# unresolved sources	Resolved flux	Unresolved flux
Low luminosity BL Lacs	Negative	359	6070	64%	36%
High luminosity BL Lacs	Positive	609	981	90%	10%
FSRQs	Positive	566	601	97%	3%
All blazars	—	1534	7652	88%	12%

IceCube realtime alert system) and different γ-ray telescopes have detected a flaring state of the source TXS 0506+056 (Aartsen et al. 2018a), which lies in a direction consistent with that of the IceCube event. That source is a BL Lac at redshift $z^{\mathrm{TXS}} = 0.3365 \pm 0.0010$, with a luminosity of $L_\gamma \simeq 0.6 \times 10^{46}$ erg/s in the quiescent state, that increased up to $L_\gamma \simeq 2 \times 10^{46}$ erg/s during the flare. Assuming a duty cycle (flaring versus non-flaring activity, see discussion in Murase et al. (2018)) of about 10%, this corresponds to an average luminosity of about $L_\gamma \simeq 0.7 \times 10^{46}$ erg/s. We will use the time-averaged luminosity to study the characteristics of TXS 0506+056 within the blazar population, and both the time-averaged and the flaring luminosity to refer to the neutrino production (see Section 4.3). The average luminosity is shown in the right panel of Fig. 4.

This discovery represents a breakthrough in the field of multi-messenger astronomy, since it is the first evidence of correlation between high-energy neutrinos and γ-rays. Several theoretical papers have been written after this discovery attempting to explain the neutrino emission with blazar radiation models (Ahnen et al. 2018; Gao et al. 2018; Keivani et al. 2018; Padovani et al. 2018; Cerruti et al. 2018; Liao et al. 2018). It is therefore important to analyze the role of this source in the context of our general description, which focuses on the entire blazar population. Note that, looking back at historical data, IceCube has also found an excess from the same direction between September 2014 and March 2015, with respect to atmospheric backgrounds. This constitutes 3.5σ evidence of neutrino emission from the direction of TXS 0506+056, independent and prior to the 2017 flaring episode (Aartsen et al. 2018b). On the other hand, the FSRQ PKS 0502+049 displayed flaring activity around the time of the neutrino flare observed by IceCube between 2014–2015 flare, and its position is compatible with the angular resolution of the events (Padovani et al. 2018; He et al. 2018), which makes this source another interesting test case for our model. In Sections 4.3 and 4.4 we discuss the implications of the current work to both these sources.
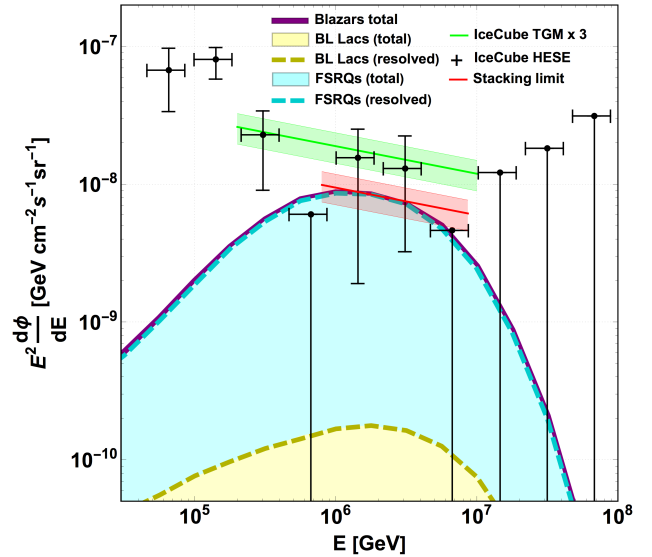

Figure 5. Scenario 1: constant baryonic loading for all sources. The black points represent the IceCube HESE (all flavor, Aartsen et al. (2015)), the green line represents the flux of throughgoing muons (x3, Aartsen et al. (2016)) and the red line represents the current IceCube blazar stacking limit (Aartsen et al. 2017b). The shaded yellow and cyan regions denote the contribution of BL Lacs and FSRQs, respectively, to the total flux of neutrinos, whereas the dotted yellow and cyan curves denote the contribution from resolved sources only. The purple solid curve represents the total neutrino flux expected from blazars. In this case most of the flux is powered by resolved sources, particularly FSRQs.

3. RESULTS

We test three different hypotheses for the baryonic loading and therefore for the relation between the the γ-ray luminosity and the neutrino luminosity defined in Eq. (4):

Scenario 1: Constant baryonic loading: We assume a constant baryonic loading $\xi(\ell) = \tilde{\xi}$ for all sources.

Scenario 2: Constant ratio L_ν/L_γ: We assume that the ratio between neutrino luminosity and γ-ray luminosity is constant, *i.e.*, $L_\nu/L_\gamma \equiv \mathcal{K} = const$. This assumption is model-independent and
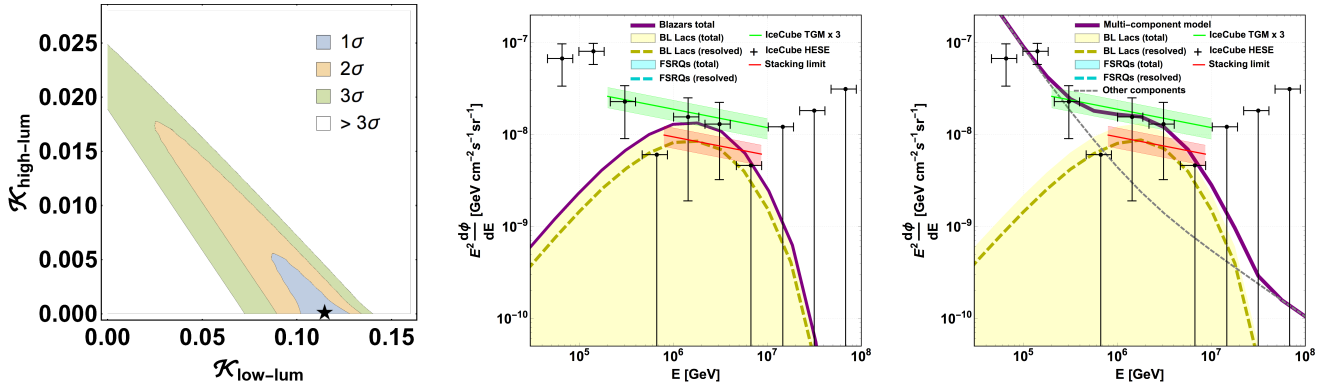
Figure 6. Scenario 2: constant ratio $\mathcal{K} = L_\nu/L_\gamma = \epsilon_\nu \times \xi$. Left: scan of the values of $\mathcal{K} = \epsilon_\nu \times \xi$ for low- and high-luminosity sources, taking into account both IceCube observations and the stacking limit. The number of standard deviations σ is calculated for 2 degrees of freedom. The best fit is marked with a black star. Middle: theoretical expectation for the neutrino flux, as described in Fig. 5, compared to the IceCube results. The flux of neutrinos is obtained considering the best fit values of $\epsilon_\nu \times \xi$, i.e., $\epsilon_\nu \times \xi = 10.5\%$ for low-luminosity sources and $\epsilon_\nu \times \xi = 0\%$ for high-luminosity sources. In this scenario, the neutrino flux is powered by both unresolved and resolved sources. Right: the same as the middle panel, including the effect of other possible contributions (multi-component model), such as the residual atmospheric background and Galactic neutrinos, as discussed in Palladino & Winter (2018).

has been used in previous works, such as Kadler et al. (2016); Righi et al. (2017); Halzen & Kheirandish (2016) to evaluate the flux of neutrinos from BL Lacs. In the context of our model, it implies that the product $\xi \times \epsilon_\nu = \mathcal{K}$ (cf., Eq. (4)) is constant, which means that $\xi \propto (\epsilon_\nu)^{-1}$. We allow for different values for low-luminosity BL Lacs and high-luminosity BL Lacs/FSRQs (for the purposes of this analysis we consider high-luminosity sources as one group), reflecting the potentially different baryonic loadings.

Scenario 3: Baryonic loading evolving with L_γ:
We assume that the baryonic loading changes continuously as a function of ℓ; this function is assumed to be universal for BL Lacs and FSRQs. This is a generalization of the second scenario.

The theoretical predictions will be compared with the through-going muon flux (Aartsen et al. 2016), with the high-energy starting events (Aartsen et al. 2015) above 100 TeV and with the blazar stacking limit (Aartsen et al. 2017b). As discussed in depth in Palladino & Vissani (2016); Palladino et al. (2016); Palladino & Winter (2018), the IceCube data below 100 TeV can be affected by the presence of Galactic neutrinos and residual atmospheric background (both conventional and prompt neutrinos). For this reason, we choose the throughgoing muons as reference for the extragalactic neutrino flux.

We now discuss the results obtained in the three scenarios described above.

3.1. *Scenario 1: Constant baryonic loading*

The diffuse flux obtained choosing a constant baryonic loading is represented in Fig. 5, where a baryonic loading $\tilde{\xi} = 150$ has been chosen in order to match and not overshoot the present IceCube stacking limit for blazars (Aartsen et al. 2017b). In Fig. 5, the cumulative flux is represented using a purple solid curve, whereas the contributions from BL Lacs and FSRQs are shown using the shaded yellow and cyan regions, respectively. From the same plot we can also see the flux produced by resolved sources (dashed curves), which in this case is not distinguishable from the total flux produced by BL Lacs and FSRQs. This implies that:

> *Assuming a constant baryonic loading, the flux of neutrinos is fully powered by resolved sources. Therefore, it is not possible to simultaneously interpret the IceCube observations and obey the stacking limit.*

A higher baryonic loading would produce a higher neutrino flux, improving the agreement with the observations but violating the stacking limit. Therefore, assuming that astrophysical neutrinos are produced by blazars, a constant baryonic loading is not a viable assumption to interpret the astrophysical neutrinos. Even if one allowed for two different constant values for BL Lacs and FSRQs, the result would not change because the neutrino flux is determined by the resolved sources; therefore, it is not possible to saturate the IceCube measurement obeying the stacking limit at the same time.

3.2. *Scenario 2: Constant ratio L_ν/L_γ*

As a second scenario, we consider a constant ratio $\mathcal{K} \equiv L_\nu/L_\gamma$. We assume two different values for low-

luminosity BL Lacs ($\mathcal{K}_{\text{low-lum}}$) and high-luminosity BL Lacs/FSRQs ($\mathcal{K}_{\text{high-lum}}$). This assumption comes from the following consideration: in order to power the neutrino flux detected by IceCube without violating the stacking limit, the contribution of unresolved sources (mainly low-luminosity sources) has to be enhanced and, at the same time, the contribution of resolved sources (mainly high-luminosity sources) has to be suppressed. As a splitting point, we choose L_γ^{evo}, which possibly separates (for BL Lacs) two different populations; see right panel of Fig. 4 and Section 2.2. Therefore we have:

$$L_\nu = \begin{cases} \mathcal{K}_{\text{low-lum}}\, L_\gamma & \text{for} \quad L_\gamma \leq L_\gamma^{\text{evo}} \\ \mathcal{K}_{\text{high-lum}}\, L_\gamma & \text{for} \quad L_\gamma > L_\gamma^{\text{evo}} \end{cases}$$

where $L_\gamma^{\text{evo}} = 3.2 \times 10^{45}$ erg/s.

Since in scenario 2 the value of $\mathcal{K} = \epsilon_\nu \times \xi$ is constant for all sources, for our source model this implies that the baryonic loading scales as $\xi(\ell) \propto (\epsilon_\nu(\ell))^{-1}$, $i.e.$, the baryonic loading $\xi(\ell)$ decreases with luminosity in a way that it exactly compensates for the increasing neutrino production efficiency $\epsilon_\nu(\ell)$. Note that the behavior of the baryonic loading depends on the particular model of neutrino production in the source, while the assumption $L_\nu \propto L_\gamma$ is not model dependent.

We compute the total and resolved fluxes from BL Lacs and FSRQs and we compare them with the measured through-going muon flux (Aartsen et al. 2016) and with the present stacking limit (Aartsen et al. 2017b), for each value of $\mathcal{K}_{\text{low-lum}}$ and $\mathcal{K}_{\text{high-lum}}$. The scan of these two parameters is represented in the left panel of Fig. 6, where we show the 1σ, 2σ, 3σ, and $\geq 3\sigma$ regions. The best fit is given by the following set of parameters:

$$\mathcal{K}_{\text{low-lum}} = 10.5\%; \quad \mathcal{K}_{\text{high-lum}} = 0\%. \tag{7}$$

The interpretation of that result is that the neutrino flux must be powered by low-luminosity sources in order to avoid overshooting the stacking limit. Within 1σ, $\mathcal{K}_{\text{high-lum}} \leq 0.5\%$ is allowed, providing an upper limit to the contribution of high-luminosity sources.

Within scenario 2, we can interpret the IceCube data at the highest energies without violating the stacking limit, as illustrated in the middle panel of Fig. 6. From this plot we notice that the contribution of FSRQs is absent while BL Lacs (particularly those with low luminosity) provide the dominant contribution to the diffuse neutrino flux (yellow shaded region). Combining this result with the idea of the multiple components presented in Palladino & Winter (2018), which may power the neutrino flux at lower energies, we obtain the total neutrino flux (shown in the right panel of Fig. 6), as combination of $i)$ the neutrino flux from blazars plus

$ii)$ the neutrino flux coming from residual background (atmospheric neutrinos, $\phi(E) \sim E^{-3.7}$ below 100 TeV) and Galactic neutrinos ($\phi(E) \sim E^{-2.6}$ with energy cutoff at 150 TeV); we denote these additional neutrinos as "other components". The additional component has not been re-fitted in this paper, but it is exactly the same reported in Palladino & Winter (2018). It is important to remark that in principle the atmospheric background should have been already subtracted in the IceCube data points. On the other hand, in Palladino & Winter (2018) we discuss the possibility that a certain residual background can still affect the IceCube measurement, even after the subtraction of the background (see particularly Fig. 2 of Palladino & Winter (2018).). The residual atmospheric background may become relevant especially below 100 TeV in the high-energy starting events analysis.

We notice from the right panel of Fig. 6 that the combination of scenario 2 and these additional components allows for an interpretation of the shape of the spectrum measured by IceCube.

3.3. Scenario 3: Baryonic loading evolving with luminosity as a power law

Here we generalize the previous scenario and assume that the baryonic loading scales with the luminosity ℓ as a continuous function, defined as follows:

- for low luminosity sources, we roughly replicate $\mathcal{K}_{\text{low-lum}} \equiv L_\nu/L_\gamma = 10.5\%$ suggested by the best fit result of scenario 2. Therefore the baryonic loading is given by $\xi(\ell) \simeq \mathcal{K}_{\text{low-lum}}/\epsilon_\nu(\ell)$ for our production model;

- for high luminosity sources we use the information $\mathcal{K}_{\text{high-lum}} \equiv L_\nu/L_\gamma < 0.5\%$ from scenario 2, and we derive an upper limit for the baryonic loading, indicated above;

- for the region in between, we use the information from TXS 0506+056 in the model presented in Gao et al. (2018), where the baryonic loading of the source TXS 0506+056 is calculated to be $\xi \simeq 3 \times 10^4$ during the flare.

The information provided above can be checked looking at the middle panel of Fig. 8, where the function $\mathcal{K}$ is represented as a function of luminosity. Let us notice that at low luminosity the value of $\mathcal{K}$ for FSRQs is not equal to $\simeq 10\%$, as stated before. On the other hand we have to take into account that at low luminosity there are no FSRQs at all (see right panel of Fig. 4); therefore this fact cannot produce any effect on the calculation.
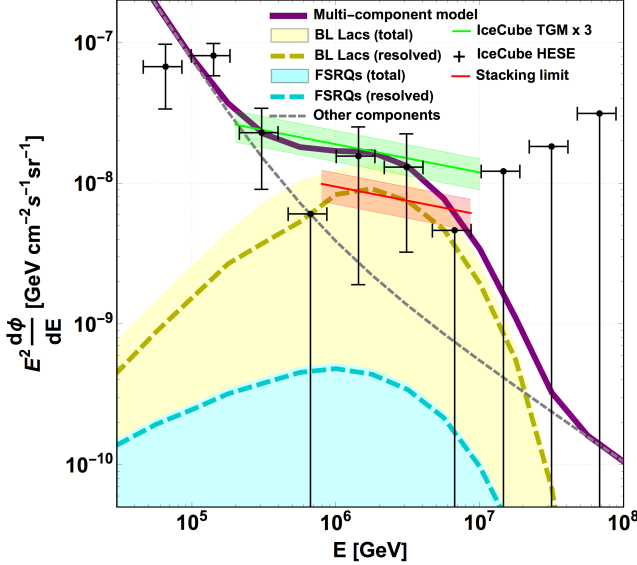
Figure 7. Scenario 3: baryonic loading changing continuously with luminosity, evolving from a baryonic loading larger than 10^5 for low-luminosity sources to a baryonic loading smaller than 10 for high-luminosity sources, see main text. As in scenario 2, the neutrino flux is powered by both unresolved and resolved sources, with a dominant contribution from low-luminosity BL Lacs.

The inclusion of the information from TXS 0506+056 is relevant for our purpose, since TXS 0506+056 has been identified as the first possible source of an IceCube neutrino event (Aartsen et al. 2018b,a). Although it may not be representative for the whole population, it is so far the only direct evidence of a correlation between neutrinos and γ-rays, and therefore it is the best available piece of information at this point. As a consequence, our model is consistent with TXS 0506+056 by construction.

As in scenario 2, we can describe the through-going muon flux without violating the stacking bound. The results and the main conclusion are similar to the ones obtained in the second scenario, as it can be observed in Fig. 7. However, this scenario allows for a small contribution by FSRQs, which was not present in the last scenario. The actual function obtained for the baryonic loading is represented by the blue curve in the left panel of Fig. 8. On the other hand, in this case we have chosen the upper limit (within 1σ) of the baryonic loading for high-luminosity sources; therefore the contribution of FSRQs can also be lower or even negligible, as can be seen by the blue shaded region in the left panel of Fig. 8, representing the uncertainty in the baryonic loading. Note that the TXS 0506+056 lies on the verge between low-luminosity sources, which dominate the diffuse neutrino flux, and the flux-limited high-luminosity

sources; it is therefore a rather special case in the context of our model.

In conclusion, the result from the second and third scenarios is comparable: the contribution of high-luminosity blazars must be suppressed in order to not violate the stacking limit, whereas the contribution of low-luminosity blazars has to be dominant. This conclusively means that:

> *Low-luminosity BL Lacs can efficiently produce neutrinos and must have a baryonic loading higher than 10^5. On the contrary, FSRQs should have a lower baryonic loading (about 10) and therefore they are not efficient in neutrino production.*

From here on this scenario will represent our baseline model and its implications will be discussed in the next section.

4. DISCUSSION

In this section we discuss some aspects to test the plausibility of our baseline model (scenario 3), namely: *i)* the evolution with γ-ray luminosity of the baryonic loading and neutrino luminosity of blazars, including a comparison of the baryonic loading with the Eddington luminosity; *ii)* the discussion concerning the multiplet constraint, *i.e.*, the absence of two neutrinos coming from the same source; *iii)* the comparison between our general picture and the expectations from the blazar TXS 0506+056, which rises at present great interest due to the recent evidence for a correlation between one of its flares and one IceCube neutrino detected in September 2017; *iv)* a discussion concerning the object PKS 0502+049 as neutrino emitter.

4.1. *Source parameters with γ-ray luminosity*

In order to test the plausibility of our result, we check if the baryonic loading that we obtain is compatible with the Eddington luminosity. The Eddington luminosity is the maximum luminosity that a body (such as a star) can achieve when there is balance between the outward radiation force and the inward gravitational force. For accreting black holes of mass M, such as AGNs, it is given by:

$$L_{\text{edd}} = \frac{4\pi G m_p c}{\sigma_{\text{T}}} M \simeq 1.26 \times 10^{38} \left(\frac{M}{M_\odot} \right) \text{ erg/s.} \quad (8)$$

where G is the gravitational constant, m_p is the proton mass and σ_{T} is the Thomson cross section. We choose a Lorentz factor $\Gamma = 40$ (see Gao et al. (2017)) and $M \in [10^8, 10^{10}]M_\odot$, which are typical black hole masses for blazars (Ghisellini & Tavecchio 2008; Ghisellini et al.
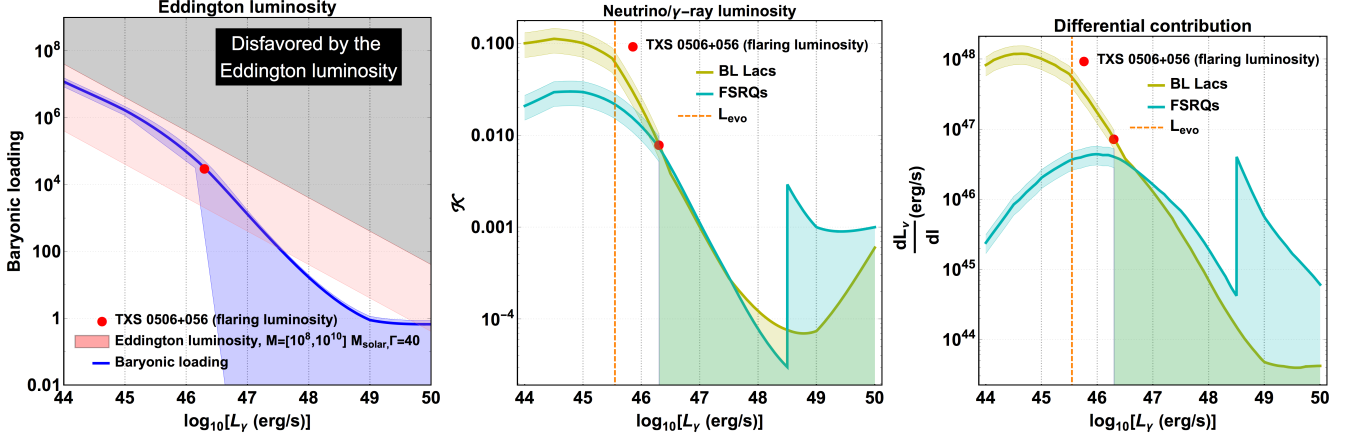
Figure 8. Left: baryonic loading as a function of L_γ (best-fit as blue curve, uncertainties represented by blue band). The region disfavored by the Eddington luminosity is represented in gray. The red region represents the Eddington luminosity assuming a black hole mass of 10^8-$10^{10} M_\odot$ and a Lorenz factor $\Gamma = 40$. Middle: neutrino to gamma-ray luminosity as a function of L_γ. Right: same quantity as in the middle panel multiplied by $dN/d\ell$, the differential number of sources of luminosity ℓ. All panels given for scenario 3.

2010) (see also Fig. 7 of Yu et al. (2015)). Note that the Eddington luminosity is no hard limit on the expected luminosity of blazars and may be exceeded, especially during flares.

For a jet dominated by baryons, we can use the Eddington luminosity to estimate the maximal plausible baryonic loading by comparing to the physical jet luminosity in γ-rays (corrected by the beaming factor) as

$$\xi(\ell) \lesssim \frac{L_{\rm edd}}{L_\gamma/(2\Gamma^2)} . \qquad (9)$$

The comparison between the baryonic loading $\xi(\ell)$ obtained in Section 3.3 and the Eddington luminosity limit is shown in the left panel of Fig. 8, where both the best-fit (blue curve) and the uncertainties (blue band) are represented. From this plot, we notice that the baryonic loading obtained is compatible with the Eddington luminosity for the Lorentz factor and the black hole masses that we have discussed. Note that the quantity $L_{\rm edd}$ may also vary with L_γ, due to the fact that the black hole mass and the ejected luminosity are slightly correlated, as found in Yu et al. (2015). On the other hand, this behavior is contained in the uncertainties that we have used, *i.e.*, within the red region in the left panel of Fig. 8; therefore this comparison is sufficient for the purposes of this study.

In the middle panel of Fig. 8 we show the ratio L_ν/L_γ for our baseline scenario 3. The ratio L_ν/L_γ evolves from approximately 10% to 0.01% for BL Lacs (yellow curve). FSRQs exhibit a similar behavior up to $L_\gamma \simeq 3 \times 10^{48}$ erg/s, where L_ν/L_γ strongly increases due to the increasing neutrino production efficiency because of external radiation fields becoming relevant. It is noteworthy that L_ν/L_γ strongly decreases in the range

relevant for TXS 0506+056, compared to lower luminosities. This observation will be relevant later on in this analysis.

The right panel of Fig. 8 shows the differential contribution to the neutrino flux, where the function $dL_\nu/d\ell$ is obtained as follows:

$$\frac{dL_\nu}{d\ell} = \int_0^{z_{\rm max}} dz \, 10^\ell \epsilon_\nu(\ell) \times \xi(\ell) \times \frac{dN}{d\ell dz} \frac{1}{(1+z)^2} . \quad (10)$$

Here $z_{\rm max} = 6$ for both BL Lacs and FSRQs. Let us recall from Eq. (4) that $10^\ell \epsilon_\nu(\ell) \times \xi(\ell) = L_\nu(\ell)$, where $10^\ell = L_\gamma$. The factor $1/(1+z)^2$ refers to the fact that frequency and energy are redshifted. It is clear from this plot that the most important contribution to the high-energy neutrino flux comes from low-luminosity blazars, especially BL Lacs.

The results obtained in the second and third scenarios also have effects on the SED modeling, since they suggest that low-luminosity objects are rich in hadrons, which means the second hump of the SED may be produced by π^0 decays. For intermediate-luminosity objects the situation is unclear and both hadronic and hybrid (lepto-hadronic) models may be sufficient to explain the γ-ray and neutrino emission at the same time. High-luminosity objects (such as FSRQs) have a low baryonic loading and, in principle, they could be in agreement with hybrid or even purely leptonic models, since the contribution of these sources to the neutrino flux can be negligible, without affecting the quality of the result, in both scenarios 2 and 3.

4.2. *Multiplet constraint*

In this section we discuss the multiplet constraint, refering to the constraint from the non-detection of two

or more events from the same source. Namely, we discuss whether the fact that IceCube has not yet observed any multiplet can test our model. This aspect has been emphasized in Murase & Waxman (2016); Murase et al. (2018), where it was remarked that the absence of neutrino multiplets (or point sources) constrains the contribution of BL Lacs at the level of 10% in the 0.1-1 PeV energy range. On the other hand, the result of this paper is not directly comparable with the one obtained in Murase et al. (2018), since in that work the number density of the sources scales as $L^{-3/2}$, where L is the average luminosity of the source. In our paper, instead, we have seen in the right panel of Fig.4 that this behavior is not always valid; moreover, following our scenario 3, the ratio L_ν/L_γ is not constant but varies from 1% to less than 0.01%, going from low-luminosity to high-luminosity sources. Therefore, the non-linear behavior between γ-rays and neutrinos suggests to use a different approach to evaluate the constraint coming from the absence of multiplets.

Let us consider the 9186 sources coming from the distribution by Ajello et al. Each source will contribute with a certain weight to the total neutrino flux expected from blazars and the weight w_i depends on the luminosity L_ν and on the redshift z, as follows:

$$w_i = \frac{L_\nu^i}{D_c(z^i)^2(1+z^i)^2},$$

where the relation between L_ν and L_γ has been discussed in Sections 3.2 and 3.3 and $D_c(z)$ is defined in Appendix B. Let us remark that more than 90% of the neutrino flux from blazars is powered by BL Lacs with $L_\gamma \leq 3.5 \times 10^{45}$erg/s (about 6300 sources), following our baseline model (scenario 3).

Using the previous information we can perform several simulations to evaluate the expected number of multiplets, given a certain number of observed signal events. Up to now, 36 throughgoing muons[3] have been observed by IceCube (Aartsen et al. 2017d) and 2/3 of them are expected to be signal events, with an uncertainty of about 30%. It translates in a tension between our model and the absence of multiplets of $1\sigma - 1.7\sigma$, as it can be seen looking at the orange band of Fig. 9. Therefore, the absence of multiplets cannot be considered an issue for our model at this stage.

On the other hand, the next IceCube generation (Aartsen et al. 2014), in which the exposure is expected

---

[3] In order to evaluate the expected number of multiplets we need events with good angular resolution, therefore we use the throughgoing muons dataset in this analysis since the average angular resolution is $\sim 1°$.
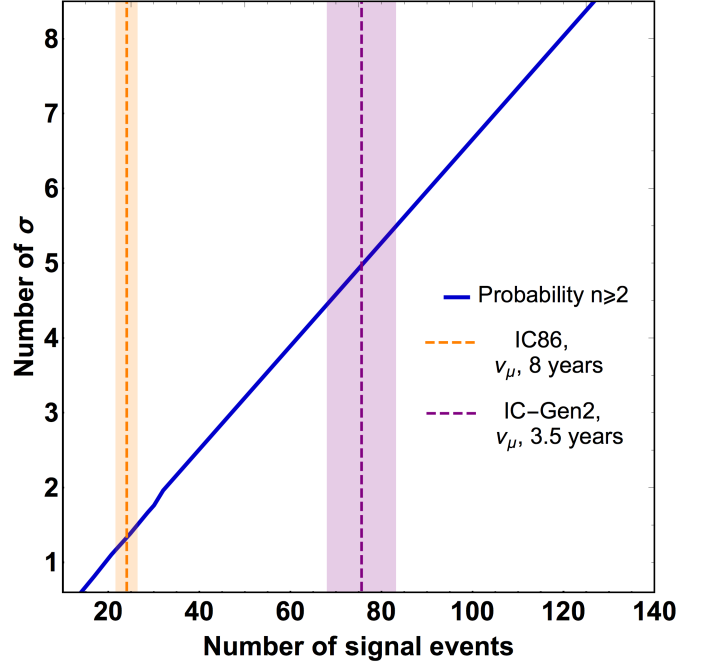

Figure 9. Multiplet limit (number of standard deviations σ of exclusion of the model due to the non-observation of multiplets) as a function of the observed number of signal events in the through-going muon dataset. In this analysis we use our baseline model (scenario 3) and we consider all 9186 blazars contained in the distribution by Ajello et al. We show as an orange band the present number of signal detected using throughgoing muons (Aartsen et al. 2016). The purple band represents the number of years required for IceCube-Gen2 to reach the 5 σ level (Aartsen et al. 2014). The band is due to the uncertainty on the contribution of the atmospheric background (mainly prompt neutrinos) to the throughgoing muons, that is about 10% (see Eq.12 of (Palladino & Vissani 2017)).

to be 6-7 times larger than the present one, can test and eventually rule out our model in about 3.5 years (purple band).

4.3. Comparison with TXS 0506+056

We here discuss the implications of the neutrino associated with TXS 0506+056 in the context of our model.

Expected number of events from TXS 0506+056:—The luminosity of neutrinos is expected to be about 1% of the γ-ray luminosity, see middle panel of Fig. 8. This is in agreement with the expectations reported in Gao et al. (2018), in which the BL Lac has been modeled using a lepto-hadronic hybrid model. Moreover, we can evaluate the neutrino flux expected from this source, following the same procedure used in Section 2.3 to evaluate the cumulative neutrino flux expected from the entire blazar population. Using the effective area for point sources reported in Aartsen et al. (2018b), we obtain a (time-
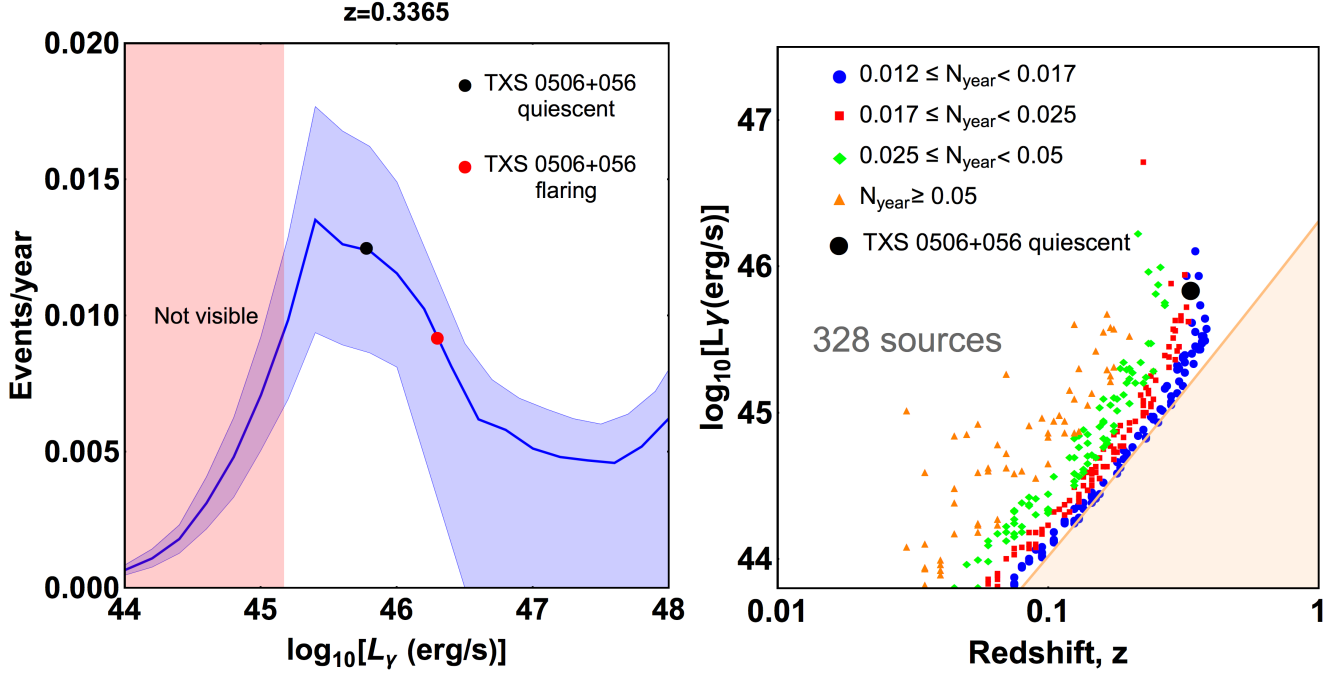
Figure 10. Left: expected number of events produced by a single source at the same redshift as TXS 0506+056 as a function of the γ-ray luminosity. The shaded pink region denotes the unresolved sources at redshift $z = 0.3365$. For a TXS-like source, the (time-averaged luminosity) number is equal to 0.012 events per year. Right: distribution of resolved sources that can produce at least the same number of events per year as TXS 0506+056. The different symbols indicate ranges of the expected number of events per year from the source.

averaged) expected number of 0.012 events per year from TXS. In the left panel of Fig. 10 we show the expected number of events/year for objects with the same redshift as TXS but different luminosities. From this plot it is clear that objects with $10^{45} \lesssim L_\gamma(\mathrm{erg/s}) \lesssim 3 \times 10^{46}$ are the best candidates to produce high-energy neutrinos, due to the fact that for lower luminosities the efficiency $\epsilon(L_\gamma)$ is lower (see Fig. 3) whereas for higher luminosities the baryonic loading $\xi(L_\gamma)$ is too low (see Fig. 7) and the neutrino flux is suppressed.

Expected number of correlations per year:—In the right panel of Fig. 10, we show the distribution of sources that produce a number of events equal or grater compared to the TXS source. Using the BL Lac distribution described in Section 2.2 we find that there are 328 visible objects capable of producing at least as many neutrinos per year as TXS. Using the effective area of Aartsen et al. (2018b) and integrating over redshift and luminosity, we find that the average number of events produced by these sources is 0.047 events/year for each source, which is roughly four times higher than that expected from TXS. Therefore, future correlations will be rather expected at somewhat lower luminosities and redshifts.

In order to estimate how many correlations per year we expect in IceCube, we have to take into account that:

- the exposure of the 3LAC catalog is 4 years (Ackermann et al. 2015) and we have to divide the total number of sources by this number;

- assuming that these 328 sources are isotropically distributed, IceCube can only detect the ones that are visible from the Northern hemisphere, via throughgoing muons. In particular, the alert system is active for neutrinos above 500 TeV (Aartsen et al. 2017c) and at this energy only neutrinos coming from declination $0° \leq \delta \leq 30°$ can cross the Earth. Therefore, IceCube is roughly sensitive to 1/4 of the sky or, equivalently, to 1/4 of the sources.

Following the previous considerations, the expected number of correlations per year N_c in IceCube is roughly

$$N_c \sim 0.047 \times \frac{328}{4} \times \frac{1}{4} = 0.96 \; \frac{\text{correlations}}{\text{year}} \quad (11)$$

The alert system is active since about two years (Aartsen et al. 2017c), and therefore the expected number of correlations is roughly equal to two, following our baseline model. Up to now, only one correlation has been observed, which is perfectly consistent with our expectation within the Poissonian uncertainty.

This estimate is based on time-averaged fluxes, and does not take into account blazar flares. If the neutrinos are only produced during flares, the duty cycle has to be taken into account, see *e.g.* discussion in Murase et al. (2018). If the blazar spends, for instance, 10% of its time in flaring activity, only 10% of the blazars are flaring at any given time, and can be used for correlation studies, suppressing the second number in Eq. (11). Assuming that during the flaring state the γ-ray luminosity increases of a factor 10, we can evaluate the expected number of neutrino events; it *does not always increase linearly*. For low luminosity objects, $L_\nu \propto L_\gamma$, but for TXS-like blazar the increase of L_γ is compensated by a decrease of the baryonic loading; therefore L_ν does not increase for this kind of objects (see left panel of Fig. 10). In fact, taking into account the distribution of luminosities and expected event rates, we obtain 0.4 expected correlations per year in our model for a duty cycle of 10% and a flaring luminosity 10 times higher than the average one. We furthermore expect from these considerations that BL Lacs with $L_\gamma \simeq 10^{45}$ erg/s and $z \sim 0.1$ will dominate future such associations.

4.4. *Comparison with PKS 0502+049*

For the sake of completeness, we tested the object PKS 0502+049 in the context of our model. In He et al. (2018) the possibility that this object has emitted high-energy neutrinos, during the flaring state observed between 2014 and 2015, is discussed. See also Padovani et al. (2018) for a direct comparison between this object and TXS 0506+056 as neutrino emitters. This source is an FSRQ at redshift $z = 0.954$ and He et al. (2018) state that to reproduce the γ-ray flare and the neutrino flare the object must exceed the Eddington luminosity by two orders of magnitude during the flaring state. Under this assumption the flaring luminosity would be $L_\gamma = 9.5 \times 10^{48}$ erg/sec. Using the previous luminosity and redshift and under scenario 3, we find that for such an object the neutrino luminosity would be $L_\nu = 7 \times 10^{-5} L_\gamma$ and it would produce a neutrino signal corresponding to the $\sim 40\%$ of the number of events per year expected from TXS 0506+056. Therefore, the association between the neutrinos observed by IceCube in 2014-2015 and the PKS 0502+049 is less plausible than the association between the same neutrinos and the TXS 0506+056 but it cannot be ruled out based on our model.

5. SUMMARY AND CONCLUSION

In this work we have studied the possibility that the diffuse astrophysical neutrino flux at the highest energies is powered by blazars, which are Active Galactic Nuclei

viewed in the direction of the jet. A major obstacle has been the AGN stacking limit, which constrains the blazar contribution from the non-observation of correlations between high-energy neutrinos and known (observed) blazars, while unresolved sources may at the same time power most of the diffuse neutrino flux. Using a phenomenological relationship between spectral energy distribution (SED) of blazars and γ-ray luminosity, known as *blazar sequence*, and a realistic neutrino production model based on these SEDs, we have derived the implications for blazars assuming that diffuse flux and stacking limit have to be described at the same time.

We have demonstrated that the choice of a constant baryonic loading over the blazar sequence does not allow for a description of the neutrino data because, fixing the baryonic loading, high-luminosity objects are very efficient neutrino producers; as a consequence, resolved sources will dominate the neutrino flux, and the stacking limit will constrain the blazar contribution to the diffuse flux. In order to avoid that, we have allowed the baryonic loading to change as a function of luminosity. We have analyzed two different possibilities: in the first one the ratio between luminosity in neutrinos and luminosity in γ-rays is constant (implying that the baryonic loading decreases with luminosity in a not continuous way), in the second one the baryonic loading scales continuously with the γ-ray luminosity and also the expectation from TXS 0506+056 are taken into account. We have found that the baryonic loading of low-luminosity objects has to be higher than 10^5, whereas the baryonic loading of high-luminosity sources (both BL Lacs and FSRQs) has to be lower than ~ 100. It is also possible that high-luminosity BL Lacs and FSRQs do not contribute to the neutrino flux at all, which means that these sources may not be cosmic ray accelerators. Low-luminosity objects can then power all the neutrino flux above few hundreds of TeV, while the contribution of high-luminosity objects is limited by the stacking limit.

In order to test the plausibility of our results, we have demonstrated that the baryonic loading obtained roughly satisfies the Eddington limit and the constraint coming from the non-observation of multiplets.

The recent observation of neutrinos from TXS 0506+056 can be interpreted in our baseline scenario. We find that the (time-averaged) expected number of neutrino events from an object at this redshift and luminosity is about 0.01 per year. Taking into account the redshift distribution, we find on average 0.05 events per blazar per year from about 80 Fermi-LAT detected objects per year, which implies that somewhat larger event rates are expected from objects with higher neutrino luminosity and lower redshifts. This yields about

one expected neutrino–blazar association per year if the sky-coverage of IceCube is taken into account. If the neutrino emission is only produced during flares, we predict about 0.4 expected associations per for a duty cycle (flaring time versus total time) of 10% and a correspondingly higher luminosity during the flare (compared to the averaged one). These associations are likely to come from BL Lacs with lower luminosities $L_\gamma \sim 10^{45}$ erg/s and redshifts $z \sim 0.1$ in our model.

In conclusion, we have demonstrated that the observed astrophysical flux of throughgoing muons and the stacking limit for blazars are almost perfectly described if one accepts large enough baryonic loadings for unresolved (low-luminosity) objects, while high-luminosity BL Lacs and FSRQs are disfavored as the main contrib-

utors to the diffuse flux of high-energy neutrinos. As a consequence, we expect signatures of hadronic processes in the SED of low-luminosity objects, such as in X-ray or TeV γ-ray data (Gao et al. 2017). From the population study, we expect further associations similar to the one to TXS 0506+056 at a rate of about 0.4 to 1 per year. However, associations with blazars with lower redshift and luminosity than TXS will be more likely.

REFERENCES

Aartsen, M., et al. 2013, Science, 342, 1242856, doi: 10.1126/science.1242856

—. 2014. https://arxiv.org/abs/1412.5106

Aartsen, M. G., et al. 2015, Astrophys. J., 809, 98, doi: 10.1088/0004-637X/809/1/98

—. 2016, Astrophys. J., 833, 3, doi: 10.3847/0004-637X/833/1/3

—. 2017a, Astrophys. J., 843, 112, doi: 10.3847/1538-4357/aa7569

—. 2017b, Astrophys. J., 835, 45, doi: 10.3847/1538-4357/835/1/45

—. 2017c, Astropart. Phys., 92, 30, doi: 10.1016/j.astropartphys.2017.05.002

—. 2017d. https://arxiv.org/abs/1710.01191

—. 2018a, Science, 361, eaat1378, doi: 10.1126/science.aat1378

—. 2018b, Science, 361, 147, doi: 10.1126/science.aat2890

Abdo, A. A., et al. 2010, Astrophys. J. Suppl., 188, 405, doi: 10.1088/0067-0049/188/2/405

Ackermann, M., et al. 2015, Astrophys. J., 810, 14, doi: 10.1088/0004-637X/810/1/14

Ahnen, M. L., et al. 2018, Astrophys. J. Lett., doi: 10.3847/2041-8213/aad083

Ajello, M., et al. 2012, Astrophys. J., 751, 108, doi: 10.1088/0004-637X/751/2/108

—. 2014, Astrophys. J., 780, 73, doi: 10.1088/0004-637X/780/1/73

—. 2015, Astrophys. J., 800, L27, doi: 10.1088/2041-8205/800/2/L27

Albert, A., et al. 2018. https://arxiv.org/abs/1808.03531

Baerwald, P., Hümmer, S., & Winter, W. 2012, Astropart. Phys., 35, 508. https://arxiv.org/abs/1107.5583

Bechtol, K., Ahlers, M., Di Mauro, M., Ajello, M., & Vandenbroucke, J. 2017, Astrophys. J., 836, 47, doi: 10.3847/1538-4357/836/1/47

Cerruti, M., Zech, A., Boisson, C., et al. 2018. https://arxiv.org/abs/1807.04335

Chianese, M., Miele, G., Morisi, S., & Vitagliano, E. 2016, Phys. Lett., B757, 251, doi: 10.1016/j.physletb.2016.03.084

Dermer, C. D., Cerruti, M., Lott, B., Boisson, C., & Zech, A. 2014, Astrophys. J., 782, 82, doi: 10.1088/0004-637X/782/2/82

Elvis, M., Wilkes, B. J., McDowell, J. C., et al. 1994, Astrophys. J. Sup., 95, 1, doi: 10.1086/192093

Esmaili, A., Palladino, A., & Vissani, F. 2016, EPJ Web Conf., 116, 11002, doi: 10.1051/epjconf/201611611002

Esmaili, A., & Serpico, P. D. 2013, JCAP, 1311, 054, doi: 10.1088/1475-7516/2013/11/054

Essey, W., Kalashev, O. E., Kusenko, A., & Beacom, J. F. 2010, Phys. Rev. Lett., 104, 141102, doi: 10.1103/PhysRevLett.104.141102

Fossati, G., Maraschi, L., Celotti, A., Comastri, A., & Ghisellini, G. 1998, Mon. Not. Roy. Astron. Soc., 299, 433, doi: 10.1046/j.1365-8711.1998.01828.x

Gao, S., Fedynitch, A., Winter, W., & Pohl, M. 2018. https://arxiv.org/abs/1807.04275

Gao, S., Pohl, M., & Winter, W. 2017, Astrophys. J., 843, 109, doi: 10.3847/1538-4357/aa7754

Ghisellini, G., Righi, C., Costamante, L., & Tavecchio, F. 2017, Mon. Not. Roy. Astron. Soc., 469, 255, doi: 10.1093/mnras/stx806

Ghisellini, G., & Tavecchio, F. 2008, mnras, 387, 1669, doi: 10.1111/j.1365-2966.2008.13360.x

Ghisellini, G., Tavecchio, F., Foschini, L., et al. 2010, mnras, 402, 497, doi: 10.1111/j.1365-2966.2009.15898.x

Halzen, F., & Kheirandish, A. 2016, Astrophys. J., 831, 12, doi: 10.3847/0004-637X/831/1/12

He, H.-N., Inoue, Y., Inoue, S., & Liang, Y.-F. 2018. https://arxiv.org/abs/1808.04330

Hummer, S., Baerwald, P., & Winter, W. 2012, Phys. Rev. Lett., 108, 231101, doi: 10.1103/PhysRevLett.108.231101

Inoue, Y., & Tanaka, Y. T. 2016, Astrophys. J., 828, 13, doi: 10.3847/0004-637X/828/1/13

Kadler, M., et al. 2016, Nature Phys., 12, 807, doi: 10.1038/nphys3715,10.1038/NPHYS3715

Keivani, A., et al. 2018. https://arxiv.org/abs/1807.04537

Liao, N.-H., Xin, Y.-L., Liang, Y.-F., et al. 2018. https://arxiv.org/abs/1807.05210

Loeb, A., & Waxman, E. 2006, JCAP, 0605, 003, doi: 10.1088/1475-7516/2006/05/003

Meszaros, P., & Waxman, E. 2001, Phys. Rev. Lett., 87, 171102, doi: 10.1103/PhysRevLett.87.171102

Murase, K., Ahlers, M., & Lacki, B. C. 2013, Phys.Rev., D88, 121301, doi: 10.1103/PhysRevD.88.121301

Murase, K., Inoue, Y., & Dermer, C. D. 2014, Phys. Rev., D90, 023007, doi: 10.1103/PhysRevD.90.023007

Murase, K., & Ioka, K. 2013, Phys. Rev. Lett., 111, 121102, doi: 10.1103/PhysRevLett.111.121102

Murase, K., Oikonomou, F., & Petropoulou, M. 2018. https://arxiv.org/abs/1807.04748

Murase, K., & Waxman, E. 2016, Phys. Rev., D94, 103006, doi: 10.1103/PhysRevD.94.103006

Paczynski, B., & Xu, G. H. 1994, Astrophys. J., 427, 708, doi: 10.1086/174178

Padovani, P., Giommi, P., Resconi, E., et al. 2018

Padovani, P., Resconi, E., Giommi, P., Arsioli, B., & Chang, Y. L. 2016a, Mon. Not. Roy. Astron. Soc., 457, 3582, doi: 10.1093/mnras/stw228

—. 2016b, Mon. Not. Roy. Astron. Soc., 457, 3582, doi: 10.1093/mnras/stw228

Palladino, A., Spurio, M., & Vissani, F. 2016, JCAP, 1612, 045, doi: 10.1088/1475-7516/2016/12/045

Palladino, A., & Vissani, F. 2016, Astrophys. J., 826, 185, doi: 10.3847/0004-637X/826/2/185

—. 2017, Astron. Astrophys., 604, A18, doi: 10.1051/0004-6361/201730739

Palladino, A., & Winter, W. 2018, doi: 10.3204/PUBDB-2018-01376

Protheroe, R. J. 1997, ASP Conf. Ser., 121, 585. https://arxiv.org/abs/astro-ph/9607165

Resconi, E., Coenders, S., Padovani, P., Giommi, P., & Caccianiga, L. 2017, Mon. Not. Roy. Astron. Soc., 468, 597, doi: 10.1093/mnras/stx498

Righi, C., Tavecchio, F., & Guetta, D. 2017, Astron. Astrophys., 598, A36, doi: 10.1051/0004-6361/201629412

Righi, C., Tavecchio, F., & Inoue, S. 2018a. https://arxiv.org/abs/1807.10506

Righi, C., Tavecchio, F., & Pacciani, L. 2018b. https://arxiv.org/abs/1807.04299

Rodrigues, X., Fedynitch, A., Gao, S., Boncioli, D., & Winter, W. 2018, Astrophys. J., 854, 54, doi: 10.3847/1538-4357/aaa7ee

Stecker, F. W. 2007, Astropart. Phys., 26, 398, doi: 10.1016/j.astropartphys.2006.08.002

Tamborra, I., Ando, S., & Murase, K. 2014, JCAP, 1409, 043, doi: 10.1088/1475-7516/2014/09/043

Tavecchio, F., Ghisellini, G., Ghirlanda, G., Foschini, L., & Maraschi, L. 2010, Mon. Not. Roy. Astron. Soc., 401, 1570, doi: 10.1111/j.1365-2966.2009.15784.x

Vietri, M. 1998, Phys. Rev. Lett., 80, 3690, doi: 10.1103/PhysRevLett.80.3690

Waxman, E., & Bahcall, J. N. 1997, Phys. Rev. Lett., 78, 2292, doi: 10.1103/PhysRevLett.78.2292

Yu, X., Zhang, X., Zhang, H., et al. 2015, Astrophys. Space Sci., 357, 14, doi: 10.1007/s10509-015-2309-4

A. SOURCE MODEL

We present here some details of the source model used in the present work. While most features are similar to Rodrigues et al. (2018), some parameter are different, which are emphasized in the following discussion. It is also worth noting that in Rodrigues et al. (2018) the model was applied to a previous implementation of the blazar sequence (Fossati et al. 1998), based mainly on radio and X-ray observations. The new implementation of the blazar sequence used in this work (Ghisellini et al. 2017), on the other hand, was constructed based on the more recent Fermi 3LAC blazar catalog (Ackermann et al. 2015), as discussed in Section 2.1. Moreover, in Rodrigues et al. (2018) only one sequence is considered, where all low-luminosity sources are BL Lacs and all high-luminosity sources are FSRQs, while in the present work we consider two sequences, one for BL Lacs and another for FSRQs, spanning the same luminosity range.

The magnetic field strength in the jet is in this work considered to be a soft power-law of the blazar luminosity, $B' \sim L_\gamma^{1/5}$. This scaling was employed in Appendix A of Rodrigues et al. (2018) because it can yield values of B' for BL Lacs in the range 0.1-1 G, and for FSRQs in the range 1 G to a few G, which are in agreement with previous results (Tavecchio et al. 2010; Dermer et al. 2014). The scaling is assumed to be the same for BL Lacs and FSRQs, which yields $B' = 6.3$ G for the brightest FSRQ discussed in this work (see Fig. 2), and $B' = 0.1$ G for the dimmest BL Lac. In contrast, in the main text of Rodrigues et al. (2018) a stronger scaling was assumed, $B \sim L_\gamma^{1/2}$, which yields much weaker magnetic field strengths in the jet of low-luminosity sources.

In Fig. 11 we explore the photo-hadronic interactions in the jet of two blazars from the blazar sequence; one a high-luminosity FSRQ, and the other a BL Lac of intermediate luminosity (bottom). We compare some of their features with an FSRQ and a BL Lac of similar luminosities, but with the assumptions considered in Rodrigues et al. (2018). In this discussion, as well as throughout this work, we consider the case of CR escape by Bohm-like diffusion, where the rate of escape is proportional to the Larmor radius of the CRs of a given energy. As shown in Rodrigues et al. (2018), the assumption for the particular CR escape mechanism only affects marginally the neutrino spectra ejected by the blazar.

In the left panels of Fig. 11 we show the photon density spectrum in the jet. Note that in the case of the FSRQ (top panel), the external radiation is relativistically boosted into the jet blob, as discussed above ($E_\gamma' = \Gamma E_\gamma$, cf. Fig. 2). For high-luminosity FSRQs (see end of this section), additional components of the SED are considered in addition to non-thermal radiation (assumed to be produced by electrons in the jet). These external components are an infra-red bump from a dusty torus of temperature 3000 K, an X-ray bump from the corona of the accretion disk (Elvis et al. 1994), and two broad lines emitted by molecular gas on the edge of the BLR (Murase et al. 2014) (cf. Fig. 1). The broad line emission, as well as a component of the radiation from the dusty torus and the accretion disk, are assumed to isotropize in the BLR (Rodrigues et al. 2018), and that component is relativistically boosted into the jet frame. This assumption has also been considered in previous works dealing with hadronic blazar jet models (Murase et al. 2014).

In the middle panels of Fig. 11 we show the rates of processes undergone by protons in the jet of the FSRQ (top) and the BL Lac (bottom). The acceleration rate displayed as a solid red line is the one considered in this work, i.e., an acceleration efficiency of $\eta = 10^{-3}$ is used (cf. Eq. (1)). In contrast, we show in dashed red the case of ultra-efficient proton acceleration, as discussed in the main text of Rodrigues et al. (2018). In that work, low-luminosity jets were considered to have lower magnetic field strengths, which counteracts the effect of a higher acceleration efficiency in the calculation of the acceleration rate (cf. Eq. (1)). The maximum energy achieved by protons in the source (which is proportional to the maximum energy of the emitted neutrinos) is the energy at which acceleration becomes dominated by cooling or hadronic interactions (including adiabatic cooling, synchrotron losses, electron-positron pair production and photo-meson production). The adiabatic cooling timescale is assumed to scale with the size of the region, whose inverse timescale is represented by the gray line; note that even if the plasma blob cools slower than adiabatically, the size of the blob plays the same role as an adiabatic cooling term in that it limits cosmic-ray acceleration to the energy where the particle Larmor radius reaches the size of the region, when it cannot be efficiently accelerated further. Note also that the maximum proton energy marked in the figure corresponds to the acceleration efficiency considered in this work, while it would be 2-3 orders of magnitude higher for $\eta = 1$.

As we can see by the solid curves in the right panels of Fig. 11, the FSRQ (top) produces neutrinos with a maximum of ~ 10 PeV in the source frame (~ 1 PeV in the jet rest frame), and the BL Lac (bottom) produces neutrinos with a maximum of 1 PeV in the source frame. If an acceleration efficiency of $\eta = 1$ were considered (dashed curves), the FSRQ neutrino spectrum would instead peak at 1 EeV and the BL Lac at 50 PeV. This shows that the low

acceleration efficiency considered in this work is essential to obtain a neutrino spectrum whose cutoff is compatible with the maximum energy of the observed astrophysical neutrinos. In addition, note that diffusive shock acceleration models (Inoue & Tanaka 2016) indicate that high acceleration efficiencies may be difficult to achieve in relativistic shocks, as it is the case of blazar jets. On the other hand, with such low acceleration efficiencies, these sources cannot power the diffuse ultra-high-energy CR spectrum (*cf.* bottom panel of Fig. 3).

Besides the acceleration efficiency, the blob size is another parameter of the model that affects the peak energy and normalization of the ejected neutrino spectrum. While the value used in this work reflects a source with a one-day variability timescale, $r'_{\rm blob} = \Gamma c \times 1\,{\rm day} = 3 \times 10^{16}$ cm, other values could be considered. In low-luminosity blazars, the maximum neutrino energy then scales $E_\nu^{\rm max} \sim r'_{\rm blob}$, because cosmic-ray acceleration is limited only by the blob size (see bottom center panel of Fig. 11); in high-luminosity objects, this scaling is weaker because the maximum energy is limited by photo-hadronic interactions (see upper center panel of Fig. 11). The total neutrino luminosity produced by the source scales mainly with the optical thickness to photo-meson production. The optical thickness grows linearly with $r'_{\rm blob}$ and with $n_\gamma \sim V'^{-1} \sim r'^{-3}_{\rm blob}$; thus, the normalization of the emitted neutrino spectrum scales as $r'^{-2}_{\rm blob}$.

Finally, the model geometry (briefly discussed in Section 2.1) follows the principles discussed in detail in Rodrigues et al. (2018). The jet blob is at a distance to the black hole $r_{\rm diss} = r'_{\rm blob}/\sin(\theta_{\rm jet}) = \Gamma r'_{\rm blob}$ (see Fig. 1), which corresponds to the dissipation radius of the jet, assumed the same for all blazars. At the same time, the sizes of the BLR and dusty torus ($r_{\rm BLR}$ and $r_{\rm DT}$), are assumed to scale with $L_{\rm disk}^{1/2}$ (Murase et al. 2014; Rodrigues et al. 2018), where $L_{\rm disk}$ the accretion disk luminosity. In turn, $L_{\rm disk}$ and L_γ also scale together as $L_\gamma \propto L_{\rm disk}^{0.683}$ (Murase et al. 2014; Rodrigues et al. 2018). Therefore, the jet blob lies inside the BLR only in FSRQs with luminosity $L_\gamma > 2.4 \times 10^{48}$ erg/s. In these cases, we model the BLR and the dusty torus as two additional radiation zones. The CRs that escape the jet then interact with the photons fields present in either zone before leaving the source. For FSRQs with $r_{\rm BLR} < r_{\rm diss} < r_{\rm DT}$, *i.e.*, with luminosity $3 \times 10^{46} < L_\gamma\,({\rm erg/s}) < 2.4 \times 10^{48}$, radiation from the dusty torus is accounted for in the calculation of CR interactions in the jet, but a one-zone model is employed, such as for BL Lacs and dimmer FSRQs (this is the case for those SEDs in Fig. 2 in which the thermal IR bump, but no optical bumps or emission lines, are present). This approximation is justified by the conclusion in Rodrigues et al. (2018) that for a purely diffusive cosmic-ray escape, the neutrinos produced during CR propagation through the BLR typically contribute negligibly to the total neutrino fluence. Rather, the flux is dominated either by the neutrinos produced directly in the jet, or during CR propagation through the thermal field from the dust torus.

B. THE DISTRIBUTION OF BL LACS AND FSRQS

The parametrization in Ajello et al. (2014, 2012) describes the differential source distribution

$$\frac{d^3 N}{dV_c\, dL_\gamma\, d\Gamma},$$ (12)

which represents the number of sources N per comoving volume V_c, L_γ is the emitted luminosity between 0.1 and 100 GeV, and Γ is the slope of the γ-ray flux, assuming power law spectra in that energy interval.

For our purpose, it is convenient to write the previous expression as a function of the redshift z and $\ell \equiv \log_{10}[\frac{L_\gamma}{\rm erg/s}]$. The comoving volume is given by:

$$V_c(z) = \frac{4}{3}\pi\, D_c(z)^3.$$ (13)

The comoving distance $D_c(z)$ is defined as follows:

$$D_c(z) = D_H \times d(z)$$ (14)
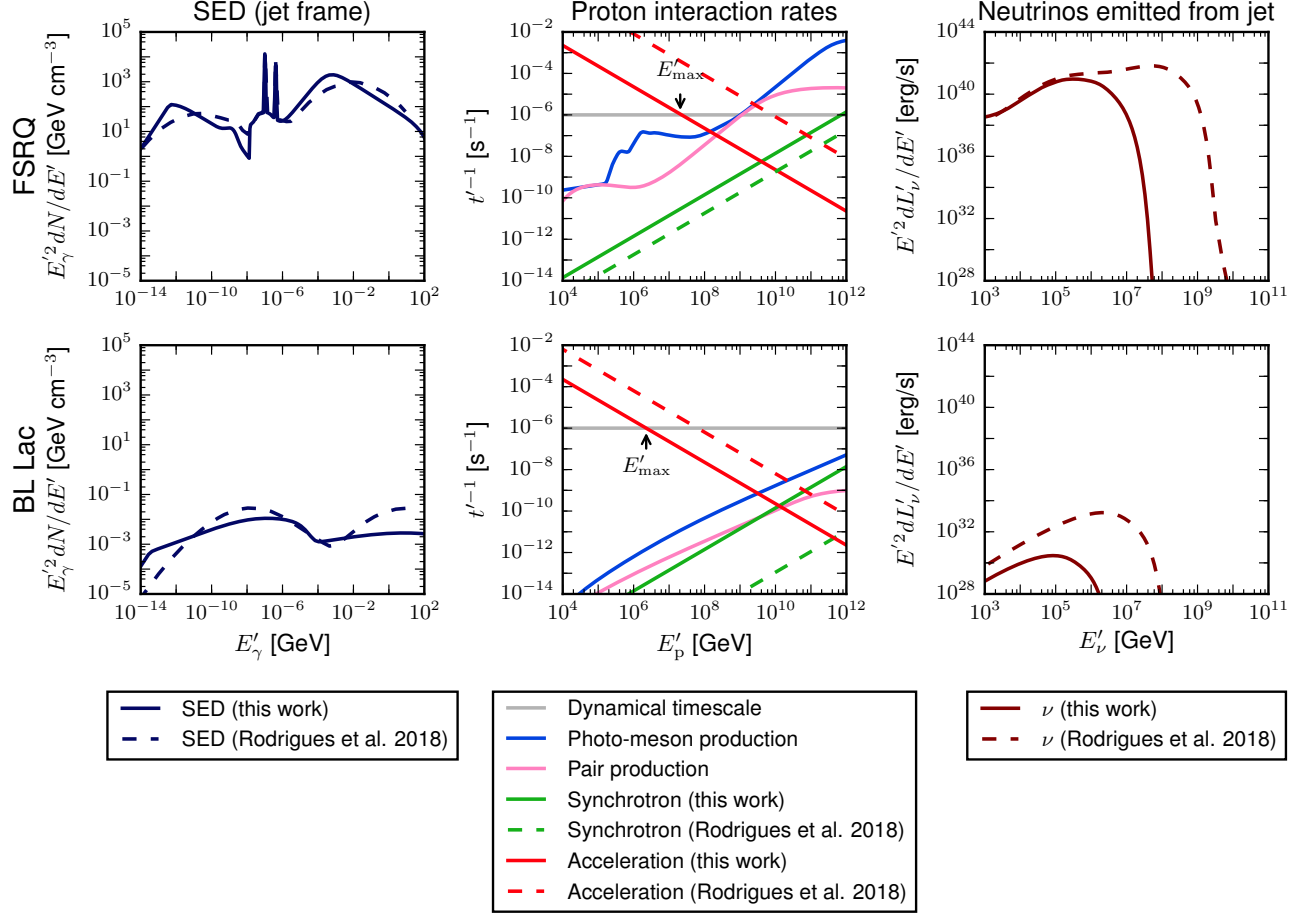
where $H_0 = 4.22$ Gpc is the Hubble distance. The function $d(z)$ is given by

$$d(z) = \int_0^z \frac{dz'}{h(z')},$$ (15)

where $h(z) = \sqrt{\Omega_\Lambda + \Omega_m(1+z)^3}$, with $\Omega_\Lambda = 0.73$, $\Omega_m = 0.27$. Based on the previous equations, the relation between the comoving volume and the redshift is

$$J_z(z) \equiv \frac{dV_c}{dz} = 4\pi\, D_H^3\, \frac{d^2(z)}{h(z)}.$$ (16)

Figure 11. Left: the solid curves represent the photon density spectra in the jet blob in an FSRQ of luminosity $10^{48.5}$ erg/s (top) and a BL Lac with $10^{44.5}$ erg/s (bottom). Note how the external fields of the FSRQ appear boosted in the jet frame due to the relativistic motion of the jet (*cf.* Fig. 2). For comparison, we plot in dashed the SEDs of an FSRQ and BL Lac of similar luminosities considered in Rodrigues et al. (2018) (an FSRQ of $10^{48.8}$ erg/s and a BL Lac of $10^{44.6}$ erg/s). Middle: cooling and interaction rates of protons in the jet for the same sources. The solid red line is the acceleration efficiency considered in this work, $\eta = 10^{-3}$ (Eq. (1)). For comparison, the red dashed line refers to a ultra-efficient proton acceleration ($\eta = 1$), discussed in Rodrigues et al. (2018). On the other hand, in that work lower magnetic fields were considered (0.9 G compared to 2 G in this work for the FSRQ and 7 mG compared to 250 mG in this work for the BL Lac). This counteracts the higher acceleration efficiency in the calculation of the acceleration rates, and also reflects in lower synchrotron cooling rates (compare solid and dashed yellow lines). Right: all-flavor neutrino luminosity spectrum produced in the jet, considering the SED and acceleration efficiency used in this work (solid) and in Rodrigues et al. (2018) (dashed). Note that in the case of the BL Lac, the suppressed neutrino production below 10^5 GeV compared to Rodrigues et al. (2018) is simply due to the higher magnetic field, which increases proton synchrotron cooling.

The Jacobian obtained from the transformation $L_\gamma \to \ell$ is given by:

$$J_\ell(\ell) \equiv \frac{dL}{d\ell} = \ln(10) \ L_\gamma \,. \tag{17}$$

Combining the two previous expressions with Eq. (12) we obtain:

$$\frac{dN}{dz d\ell d\Gamma} = J_z(z) \times J_\ell(\ell) \times \frac{dN}{dV_c dL_\gamma d\Gamma} \,. \tag{18}$$

The parametrization by Ajello et al. is characterized by several parameters, which we report in Tab. 2. Only the first

Table 2. Table of the parameters that define the distribution of BL Lacs and FSRQs. Only the first two parameters are dimensional, as indicated.

	A [Gpc^{-3}]	L_*[erg/s]	γ_1	γ_2	μ_*	β	σ	p_1^*	p_2^*	τ	z_c^*	α
BL Lacs	3.39	$10^{47.4472}$	0.27	1.86	2.10	0.0646	0.26	2.24	-7.37	4.92	1.34	0.0453
FSRQs	3.06	$10^{47.9243}$	0.21	1.58	2.44	0	0.18	7.35	-6.51	0	1.47	0.21

two parameters, namely A and L_*, are dimensional, respectively in Gpc^{-3} and erg/s. In this discussion, the quantity L_γ is always in units of erg/s.

The distribution in the slope Γ is assumed to be Gaussian,

$$g(\Gamma, L_\gamma) = \exp\left[-\frac{(\Gamma - \mu(L_\gamma))^2}{2\sigma^2}\right], \tag{19}$$

where $\mu(L_\gamma) = \mu_* + \beta[\log_{10}(L_\gamma) - 46]$.

The distribution in luminosity is given by a double power law:

$$f(L_\gamma) = \frac{1}{\left(\frac{L_\gamma}{L_*}\right)^{\gamma_1} + \left(\frac{L_\gamma}{L_*}\right)^{\gamma_2}}. \tag{20}$$

The distribution in z is described by an evolution factor, also a double power law:

$$e(\omega, L_\gamma) = \frac{1}{\omega^{-p_1(L_\gamma)} + \omega^{-p_2}} \quad \text{with } \omega(z, L_\gamma) = \frac{1 + z}{1 + z_c(L_\gamma)}, \tag{21}$$

where $z_c(L_\gamma) = z_c^* \times (L_\gamma/10^{48})^\alpha$ and $p_1(L_\gamma) = p_1^* + \tau[\log_{10}(L_\gamma - 46)]$. **Note: in Ajello et al. (2012, 2014) the previous equation is reported with a wrong positive sign for both p_1 and p_2. This mistake has been corrected later, in the footnote of Ajello et al. (2015). The wrong sign produces important difference especially in the distribution of low luminosity BL Lacs, that accumulates at high redshift using the wrong distribution.**

Summarizing, we have:

$$\frac{d^3N}{dV_c dL_\gamma d\Gamma} = \frac{A}{\ln(10)L_\gamma} \, g(\Gamma, \ell) f(\ell) e(\omega, \ell), \tag{22}$$

which, using Eq. (18), becomes:

$$\frac{d^3N}{dz\, d\ell\, d\Gamma} = \mathcal{N} \times g(\Gamma, \ell) \, f(\ell) \, e(\omega, \ell) \times \frac{d(z)^2}{h(z)}. \tag{23}$$

The coefficient $\mathcal{N} = 4\pi D_H^3 \times A$ is a dimensionless number.

Our purpose is to compute the diffuse high-energy neutrino flux from blazars. Since the neutrino spectra at the source only depends on the luminosity, we can integrate over the slope Γ (we cannot integrate over the redshift z, since the redshift enters in the computation of the neutrino flux at Earth). Henceforth we will use the distribution of BL Lacs and FSRQs defined as follows:

$$\frac{d^2N}{dz\, d\ell} = \int_{\Gamma_1}^{\Gamma_2} \frac{d^3N}{dz\, d\ell\, d\Gamma} d\Gamma. \tag{24}$$

The intervals of integration for BL Lacs are $z \in [0.03, 6]$ and $\ell \in [43.85, 52]$, and for FSRQs $z \in [0, 6]$ and $\ell \in [44, 52]$. The distributions $\frac{d^2N}{dz d\ell}$ of BL Lacs and FSRQs are represented in the left panel of Fig. 4.