The short gamma-ray burst population in a quasi-universal jet scenario

Om Sharan Salafia1,2, Maria Edvige Ravasio3,1, Giancarlo Ghirlanda1,2, and Ilya Mandel4,5

1 INAF – Osservatorio Astronomico di Brera, Via Emilio Bianchi 46, 23807 Merate (LC), Italy
2 INAF – sezione di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano (MI), Italy
3 Department of Astrophysics/IMAPP, Radboud University, 6525 AJ Nijmegen, The Netherlands
4 School of Physics and Astronomy, Monash University, Clayton, Victoria 3800, Australia
5 ARC Center of Excellence for Gravitational Wave Discovery – OzGrav, Hawthorn, Australia

Received 2 July 2023 / Accepted 7 September 2023

ABSTRACT

We present a model of the short gamma-ray burst (SGRB) population under a ‘quasi-universal jet’ scenario in which jets can differ somewhat in their on-axis peak prompt emission luminosity, $L_\gamma$, but share a universal angular luminosity profile, $\ell(\theta_0) = L(\theta_0)/L_\gamma$, as a function of the viewing angle, $\theta_0$. The model was fitted, through a Bayesian hierarchical approach inspired by gravitational wave (GW) population analyses, to three observed SGRB samples simultaneously: the Fermi/GBM sample of SGRBs with spectral information available in the catalogue (367 events); a flux-complete sample of 16 Swift/BAT SGRBs that are also detected by the GBM and have a measured redshift; and a sample of SGRBs with a binary neutron star (BNS) merger counterpart, which only includes GRB 170817A at present. Particular care was put into modelling selection effects. The resulting model, which reproduces the observations, favours a narrow jet ‘core’ with half-opening angle $\theta_0 = 2.1^{+2.8}_{-1.1}$ deg (uncertainties hereon refer to 90% credible intervals from our fiducial ‘full sample’ analysis) whose peak luminosity, as seen on-axis, is distributed as a power law, $p(L_\gamma) \propto L_\gamma^{-4}$, above a minimum isotropic-equivalent luminosity, $L^*_{\gamma} = 5^{+15}_{-0.1} \times 10^{51}$ erg s$^{-1}$. For viewing angles larger than $\theta_0$, the luminosity profile scales as a single power law, $\ell \propto \theta^{-\alpha_{\ell}}$, with $\alpha_{\ell} = 4.7^{+1.2}_{-1.2}$, with no evidence of a break, despite the model allowing for it. While the model implies an intrinsic ‘Yonetoku’ correlation between $L$ and the peak photon energy, $E_\gamma$, of the spectral energy distribution, its slope is somewhat shallower, $E_\gamma \propto L^{0.6 \pm 0.2}$ than the apparent one, and the normalisation is offset towards larger $E_\gamma$, due to selection effects. The implied local rate density of SGRBs (regardless of the viewing angle) is between about one hundred up to several thousand events per cubic gigaparsec per year, in line with the BNS merger rate density inferred from GW observations. Based on the model, we predict 0.2 to 1.3 joint GW+SGRB detections per year by the advanced GW detector network and Fermi/GBM during the O4 observing run.

Key words. relativistic processes – gamma-ray burst: general – methods: statistical

1. Introduction

Two main clusters of gamma-ray bursts (GRBs) have long been identified in the two-dimensional plane of duration versus hardness ratio$^1$ of the large sample collected by the Burst Alert and Transient Source Experiment (BATSE) on board the Compton Gamma-Ray Observatory (Kouveliotou et al. 1993). The bimodality in the BATSE GRBs is also apparent when considering the durations only (more precisely, the time, $T_{90}$, over which 5% to 95% of the background-subtracted counts are collected), with the histogram of the logarithms of the durations featuring two peaks separated by a valley at a duration $T_{90} = 2$ s. This has since become the customary separation between the ‘long’ (LGRB) and the ‘short’ (SGRB) events in the GRB class, although the actual position of the valley varies somewhat across different detectors (e.g. Bromberg et al. 2013). The evidence accumulated during the following decades painted a picture of two different progenitor systems: starting with GRB 980425 (Galama et al. 1998), a number of LGRBs have been firmly associated with type Ib/c core-collapse supernovae (e.g. Bloom et al. 2002; Malesani et al. 2004; Mirabal et al. 2006; Kann et al. 2011; Cano et al. 2014; D’Elia et al. 2018; Hu et al. 2021), securing the scenario of a massive star progenitor (Woosley 1993); the progenitors of SGRBs remained elusive for a longer time, though all hints consistently pointed (e.g. Nakar 2007; Fong & Berger 2013; Berger 2014; D’Avanzo 2015) to a compact binary merger progenitor (Eichler et al. 1989; Mochkovitch et al. 1993). Such a progenitor has been confirmed by the astounding association (Abbott et al. 2017a,b) of the SGRB 170817A (Goldstein et al. 2017) – detected by the Gamma-ray Burst Monitor (GBM; Meegan et al. 2009) on board the Fermi spacecraft and by the International Gamma-Ray Astrophysics Laboratory (INTEGRAL; Savchenko et al. 2017) – with the first-ever binary neutron star (BNS) merger detected by humankind, GW170817 (Abbott et al. 2017a), whose gravitational wave (GW) signal was captured on 17 August 2017 by the Advanced Laser Interferometer Gravitational wave Observatory (aLIGO; Aasi et al. 2015) and localised thanks to Advanced Virgo (Acernese et al. 2014).

$^1$ The source code and posterior samples behind this work are publicly available at https://github.com/omsharanalsalafia/sgrbpop.

$^2$ In X-ray and $\gamma$-ray astronomy, detectors typically identify ‘events’ (interactions between photons and the active part of the detector) in different energy channels. The hardness ratio is generally defined as the ratio of the event counts in a higher-energy (‘harder’) channel (or group of channels) to that in a lower-energy (‘softer’) channel.
The properties of SGRBs – especially their shorter duration and harder spectrum with respect to LGRBs (Ghirlanda et al. 2009; Calderone et al. 2015) – make them very hard to detect with current facilities: of the more than 280 GRBs revealed by Fermi/GBM every year, only about 40 are SGRBs. The softer sensitivity band of the Burst Alert Telescope (BAT) on board the Neil Gehrels Swift Observatory (Gehrels et al. 2004), together with its smaller field of view, allows it to identify and localise only a handful of SGRBs per year. Moreover, even when localised by Swift, the fraction of SGRBs that end up with a secure redshift determination is relatively low, due to a combination of a fainter X-ray ‘afterglow’ (Costa et al. 1997) – whose detection is a requirement for a precise localisation – and a larger typical offset from the host galaxy (Fong et al. 2013, 2022; Fong & Berger 2013; D’Avanzo et al. 2014; Berger 2014) with respect to LGRBs, which renders the host galaxy identification ambiguous in cases where multiple galaxies stand at similar offsets from the position of the afterglow. As a result, the intrinsic properties of the SGRB population are much more uncertain than for LGRBs. Indeed, a variety of attempts at constraining the properties of the SGRB population throughout the years, in some cases with very different methodologies and reference samples, has yielded varied results (e.g. Schmidt 2001; Guetta & Piran 2005, 2006; Virgili et al. 2011; Yonetoku et al. 2014; D’Avanzo et al. 2014; Wanderman & Piran 2015; Shahmoradi & Nemiroff 2015; Ghirlanda et al. 2016; Zhang & Wang 2018; Paul 2018; Tan & Yu 2020). Some of these results are in clear tension with each other: in this work, we consider Wanderman & Piran (2015; hereafter W15) and Ghirlanda et al. (2016; hereafter G16), which reach very different conclusions, as our benchmarks.

A prominent challenge in unveiling the intrinsic properties of SGRBs is the uncertainty about which processes shape the luminosity function, that is, the probability distribution from which the luminosity of each event in the population is sampled. Lipunov et al. (2001) and Rossi et al. (2002) were the first to realise that the highly relativistic nature of GRB jets would make their angular structure an important factor in determining the luminosity function, in addition to the intrinsic spread in luminosities. Indeed, if the energy density and the typical Lorentz factor of a GRB jet are functions of the angular separation from the jet axis, then the apparent energetics are viewing-angle-dependent by virtue of relativistic aberration effects (e.g. Salafia et al. 2015). Since jets are isotropically oriented in space, this naturally produces a large spread in the apparent luminosities, with a well-defined dependence on the angular structure (Pescalli et al. 2015). In the presence of a narrow distribution of intrinsic jet luminosities, the luminosity function is then mainly shaped by the angular structure (e.g. Salafia et al. 2020; Tan & Yu 2020).

An angular profile in the jet properties arises naturally in essentially any physically viable jet formation scenario (see Salafia & Ghirlanda 2022 for a recent review). Moreover, a non-trivial jet energy density angular profile is required (e.g. Mooley et al. 2018; Lamb et al. 2018, 2019; Ghirlanda et al. 2019; Takahashi & Ioka 2020, 2021; Beniamini et al. 2022) to explain the observed properties of the non-thermal afterglow of the SGRB associated with GW170817. Hence, the question of whether the observed distribution of SGRB properties can be traced back, at least in part, to the differing viewing angles is of particular relevance, as it would provide a route to a unification of these sources and a way to disentangle the intrinsic diversity in their properties (and hence those of their progenitor) from the apparent diversity due to extrinsic factors, particularly the viewing angle.

In the future, joint GW-GRB observations will provide direct information on the structure of SGRB jets thanks to the measurements of the inclination of the merging binary’s angular momentum (which is most likely a proxy of the jet viewing angle) that can be inferred from GW analysis (Williams et al. 2018; Hayes et al. 2020; Farah et al. 2020; Biscoveanu et al. 2020). Still, such information must be combined self-consistently with that encoded in the SGRB population observed in gamma-rays only. In this work, we describe our investigation of the population properties of SGRBs within a ‘quasi-universal’ jet scenario. We assumed that all SGRB jets share the same angular profile of luminosity as a function of the viewing angle, while we allowed for a spread in the on-axis luminosities (as in e.g. Tan & Yu 2020; Hayes et al. 2023). This produces a particular parametrisation of the SGRB population properties, which we derive and describe in detail in Sects. 2.2 and 2.3. In order to constrain the parameters of this model, we considered the sample of SGRBs detected by Fermi and carefully modelled the underlying selection effects (Sect. 2.4). This allowed us to fit the model to the data through a hierarchical Bayesian approach (Sect. 2.5). In Sect. 3 we describe in detail the results of the fit, and in Sect. 4 we discuss several implications.

Throughout this work, we assume a flat Friedmann-Lemaître-Robertson-Walker cosmology with Planck Collaboration XIII (2016) parameters, that is, $H_0 = 67.74 \text{ km Mpc}^{-1} \text{s}^{-1}$ and $\Omega_{\text{m},0} = 0.3075$.

## 2. Methodology

### 2.1. Apparent versus intrinsic structure

The dominant form of energy in GRB jets and the processes that dissipate such energy, leading to the observed ‘prompt’ gamma-ray emission are still a matter of debate (see Kumar & Zhang 2015, for a recent review). Still, regardless of the particular dissipation and emission process, the observed emission is affected by relativistic aberration effects in a way that depends on the Lorentz factor profile and the viewing angle (e.g. Woods & Loeb 1999; Salafia et al. 2015). For example, an observer looking, from a viewing angle $\theta_0$ (angle between the line of sight and the jet axis), at an axisymmetric jet with a bulk Lorentz factor profile $\Gamma(\theta)$ (where $\theta$ is the angle from the jet axis) that radiates an energy per unit solid angle $dE/d\Omega$ at $\theta$ would measure an isotropic-equivalent gamma-ray energy (Salafia et al. 2015),

$$E_{\text{iso}}(\theta_0) = \int_{0}^{\pi/2} \int_{0}^{2\pi} \frac{\delta_3(\theta, \phi, \theta_0)}{\Gamma(\theta)} \frac{dE}{d\Omega}(\theta) \ d\phi \sin \theta d\theta,$$

where $\phi$ is the azimuthal angle of a spherical coordinate system whose $z$-axis coincides with the jet axis, $\delta_3(\theta, \phi, \theta_0) = \Gamma^{-1}(\theta) [1 - \beta(\theta, \phi) \cdot e_{\text{LoS}}(\theta_0)]^{-1}$ is the Doppler factor, $\beta(\theta, \phi)$ is the local jet bulk velocity vector – with a magnitude $\beta(\theta) = \sqrt{1 - 1/\Gamma^2(\theta)}$ – and $e_{\text{LoS}}$ is a unit vector pointing to the observer. Hence, when considering the emitted energy in gamma-rays, the ‘apparent structure’ $E_{\text{iso}}(\theta_0)$ depends on the intrinsic structure ($T(\theta), dE/d\Omega$) through a functional that is not invertible in general. The situation is even more nuanced when considering the luminosity, as the effective angular profile $L_{\text{iso}}(\theta_0)$ depends also on the degree of overlap between different pulses (Salafia et al. 2016), and hence on the intrinsic variability.

For these reasons, in the absence of strong theoretical constraints on the intrinsic jet structure and on the dissipation and
emission processes, the most straightforward approach – which we adopt in this work – is that of parameterising directly the apparent luminosity structure \( L(\theta_c) \) (we drop the ‘iso’ suffix from here on for simplicity), which also reduces the number of parameters. An assessment of the intrinsic jet structures that are compatible with a given apparent luminosity profile can be then carried out a posteriori.

### 2.2. Statistical model of an SGRB population with a quasi-universal jet

Within our framework, we describe each SGRB by four physical quantities, namely its viewing angle \( \theta_v \), its peak isotropic-equivalent luminosity \( L \), the photon energy \( E_p \) at the peak of the spectral energy distribution (SED; i.e. the \( vF_v \) spectrum) and its redshift \( z \). These collectively represent what we refer to as the source parameter vector, \( \lambda_{\text{src}} = (\theta_v, L, E_p, z) \). For most events, the viewing angle is unknown and we therefore consider a reduced source parameter vector \( \lambda_{\text{src}} = (L, E_p, z) \). We assumed the diversity in these parameters to be the combined result of intrinsic heterogeneity in the physical properties of jets within the population and extrinsic diversity induced by the differing viewing angles under which these jets are observed.

In order to represent the intrinsic heterogeneity in SGRB jets, we opted for parameterising the joint probability density \( P(L_c, E_{pc}) \) of their on-axis (‘core’) peak luminosity \( L_c \) and peak SED photon energy \( E_{pc} \) as follows. We assumed \( L_c \) to be distributed as a power law with index \( -A \), with a lower exponential cutoff below \( L_c^* \), namely

\[
P(L_c \mid A, L_c^*) = \frac{A}{(1 - 1/A)L_c^*} \exp\left[-\frac{L_c^*}{L_c}\right] \left(\frac{L_c}{L_c^*}\right)^{-A},
\]

where \( \Gamma \) indicates a Gamma function, and the integrated probability is normalised to unity. This probability density is defined in such a way that it peaks at \( L_c^* \) regardless of the value of \( A \). This choice of parametrisation follows from the fact that, in the quasi-universal jet scenario, the high end of the luminosity distribution reflects the luminosity distribution of jets seen close to on-axis (see Sect. 2.3), and the high-luminosity end has been found to be well described by a power law in most previous studies of the SGRB population (e.g. Schmidt 2001; Guetta & Piran 2006; Virgili et al. 2011; Yonetoku et al. 2014; D’Avanzo et al. 2016; Ghirlanda et al. 2016; Zhang & Wang 2018; Paul 2018; Tan & Yu 2020). The reason for introducing a low-end cut-off follows from the quasi-universality assumption (in which case low luminosities are due to off-axis viewing angles rather than to intrinsically weak jets) and can be physically linked to a minimum jet luminosity required for the jet to break out from the progenitor vestige (see Salafia & Ghirlanda 2022 for a recent review).

The probability distribution on \( E_{pc} \), conditional on \( L_c \), was assumed log-normal and centred at a \( L_c \)-dependent value \( \bar{E}_{pc} = E_{pc}^*(L_c/L_c^*)^\gamma \), where \( \gamma \) sets the slope of the relation. Hence,

\[
P(E_{pc} \mid L_c, E_{pc}^*, \gamma, \sigma_c) = \frac{\exp\left[-\frac{1}{2} \ln(E_{pc}) - \ln(E_{pc}^*)^2}{E_{pc} \sqrt{2\pi\sigma_c^2}},
\]

where \( \sigma_c \) sets the dispersion of \( E_{pc} \) around \( \bar{E}_{pc} \). This assumption allows for a ‘Yonetoku’ correlation\(^2\) with slope \( \gamma \) between the logarithms of the on-axis peak SED photon energy and the luminosity, which may be induced, for example, by the underlying emission process. The log-normal form of the scatter around the relation was chosen for its simplicity. The case with no correlation (hence with log-normally distributed values of \( E_{pc} \), un-correlated with \( L_c \)) is represented by \( \gamma = 0 \) and it is therefore naturally included. The joint probability density of the core quantities is

\[
P(L_c, E_{pc} \mid \lambda_{\text{pop}}) = P(L_c \mid \lambda_{\text{pop}})P(E_{pc} \mid L_c, \lambda_{\text{pop}}),
\]

where the population parameter vector \( \lambda_{\text{pop}} \) contains \( L_c^*, A, E_{pc}^*, \gamma, \sigma_c \), and all other parameters that fully specify the SGRB population model.

The next, key assumption of the model is that all jets share a universal ‘structure’ that specifies the dependence of \( L \) and \( E_p \) on the viewing angle \( \theta_v \). In practice, we assumed the viewing-angle-dependent luminosity and SED peak photon energy to be expressed as

\[
L(\theta_v, \lambda_{\text{pop}}) = L_c(\theta_v, \lambda_{\text{pop}}) \quad \text{and} \quad \quad E_p(\theta_v, \lambda_{\text{pop}}) = E_p(\theta_v, \lambda_{\text{pop}}),
\]

where \( \ell \) and \( \eta \) are functions of the viewing angle and of some parameters included in the \( \lambda_{\text{pop}} \) vector. These functions, which we assumed to be redshift-independent, define the universal apparent structure of the jet.

For a population of isotropically oriented jets, whose viewing angle probability distribution is \( P(\theta_v) = \sin \theta_v \), we have

\[
P(\theta_v, L, E_p \mid \lambda_{\text{pop}}) = P(L, E_p \mid \lambda_{\text{pop}})P(\theta_v) = \delta(L - L_c(\theta_v, \lambda_{\text{pop}}))\delta(E_p - E_{pc}(\theta_v, \lambda_{\text{pop}}))P(L_c, E_{pc} \mid \lambda_{\text{pop}}) \sin \theta_v.
\]

The induced joint luminosity and peak photon energy distribution, given the isotropic viewing angles, is then given by

\[
P(L, E_p \mid \theta_v) = \int_0^{\pi/2} \int_0^{\pi/2} P(L, E_p \mid \theta_v, \lambda_{\text{pop}}) \sin \theta_v \, d\theta_v.
\]

In general, this must be solved numerically, but in the next section we analyse two cases where the intrinsic dispersion is negligible (i.e. \( \sigma_c \rightarrow 0 \) and \( \ell \rightarrow \infty \)) and \( \ell \) and \( \eta \) take simple forms, so that an analytical integration is possible: this will help in demonstrating the main features of the \( P(L, E_p) \) distribution induced by such a quasi-universal structure scenario.

As a consequence of the assumption of redshift independence of the jet structure parameters, the probability distribution of the population source parameters is \( P_{\text{pop}}(L, E_p, z \mid \lambda_{\text{pop}}) = P(L, E_p \mid \lambda_{\text{pop}})P(z \mid \lambda_{\text{pop}}) \), where the redshift probability distribution can be expressed as

\[
P(z \mid \lambda_{\text{pop}}) \propto \frac{\dot{\rho}(z, \lambda_{\text{pop}}) \, dV}{1 + z}.
\]

Here \( dV/dz \) is the differential comoving volume and \( \dot{\rho}(z, \lambda_{\text{pop}}) \) is the SGRB rate density at redshift \( z \). We parametrise the latter as

---

\(^2\) The name follows from the apparent correlation between \( \log(L) \) and \( \log(E_p) \) in LGRBs originally found by Yonetoku et al. (2004).
a smoothly broken power law, namely

\[ \rho(z, \theta_0) = \frac{1 + z}{R_0} (1 + \frac{1 + z}{1 + R_0}), \]  

(9)

where \( a, b, \) and \( z_0 \) are free parameters and \( R_0 \) is the local rate density of SGRBs with any viewing angle. This parametrisation reflects the functional form of the fitting function to the cosmic star formation rate (CSFR) from Madau & Dickinson (2014). Allowing for variations in its parameters \( a, b, \) and \( z_0 \) covers a variety of possible evolutionary histories that rise, peak and decay with increasing \( (1 + z) \), in a way that is agnostic of the progenitor nature, but can still accommodate, for example, typical rate evolutions obtained by convolving the cosmic star formation history with a merger delay time distribution under the assumption of a compact binary merger progenitor (as done in e.g. D’Avanzo et al. 2014; Wanderman & Piran 2015).

2.3. Apparent jet structure models and the implied luminosity-peak energy distribution

2.3.1. Luminosity function

A simple and widely adopted parametric form for the jet structure is a Gaussian one, namely

\[ \ell = \exp \left( -\frac{1}{2} \left( \frac{\theta}{\theta_c} \right)^2 \right), \]

\[ \eta = \exp \left( -\frac{1}{2} \left( \frac{\theta}{\theta_{c,p}} \right)^2 \right). \]

(10)

In the absence of a dispersion in the core quantities, which formally corresponds to the limit \( \sigma_\ell \to 0 \) and \( A \to \infty \), and in the case where \( \ell \) and \( \eta \) are monotonic, Eq. (7) reduces to a change of variables from \( \theta \) to either \( L \) or \( E_p \) applied to the viewing angle probability \( P(\theta_\ell, \alpha) = \sin \theta_\ell \), that is (Pescalli et al. 2015; Salafia & Ghirlanda 2022),

\[ P(L, E_p) = \left( \frac{L_c}{\theta_\ell} \right)^{\alpha} \sin \theta_\ell \delta \left( E_p - \eta(\theta_\ell, E_p) \right) \]

\[ \times \left( \frac{L}{L_c} \right)^{\alpha} \left( \frac{E_p}{E_p} \right)^{-\alpha} \delta \left( E_p - E_{p,c} \left( \frac{L}{L_c} \right)^{\alpha} \right). \]

(11)

where \( \ell^{-1} \) is the inverse function of \( \ell \) and \( \eta^{-1} \) is the inverse function of \( \eta \). In what follows, we show results using the first of the above equalities, which highlights the dependence on \( L \), but the results using the second equality are entirely analogous and can be obtained by exchanging \( L \leftrightarrow E_p \), \( \theta_\ell \leftrightarrow \theta_{c,p} \), and \( L_c \leftrightarrow E_{p,c} \). In the Gaussian apparent structure case this yields, for \( L < L_c \) and \( E_p < E_{p,c} \),

\[ P(L, E_p) = \frac{\theta_\ell}{\theta_{c,p}} \frac{\sin \theta_\ell \sqrt{2 \ln(L_c/L)}}{\sqrt{2 \ln(L_c/L)}} \delta \left( E_p - E_{p,c} \left( \frac{L}{L_c} \right)^{\alpha} \right) \]

\[ \approx \frac{\theta_\ell^2}{\ell} \delta \left( E_p - E_{p,c} \left( \frac{L}{L_c} \right)^{\alpha} \right). \]

(12)

where the last approximate equality is valid for \( \theta_\ell \ll \pi/2 \), which corresponds to \( L \gg L_c \exp(-1/2\theta_\ell^2) \) (or \( E_p \gg E_{p,c} \exp(-1/2\theta_\ell^2) \)). For typical values \( \theta_\ell \ll 1 \) (or \( \theta_{c,p} \ll 1 \)), the exponential factor is tiny and hence the approximation applies to essentially all relevant luminosities and \( E_p \)’s. The luminosity function induced by a Gaussian universal apparent structure is therefore uniform in \( \log(L) \), and the same applies to the \( E_p \) distribution.

The effect of a non-zero dispersion in the core quantities is equivalent to that of convolving the zero-dispersion \( P(L, E_p) \) probability density distribution with the probability density distribution of the core luminosity and peak SED photon energy \( P(L_c, E_{p,c}) \). In that sense, the probability density distribution of the core quantities acts essentially as a smoothing kernel: in the Gaussian apparent structure case, it introduces a smooth transition to a power law fall-off above \( L_c^* \) and \( E_{p,c}^* \). In more physical terms, and for typical parameters relevant to the quasi-universal jet scenario, the high-end of the luminosity function is shaped by the distribution of core luminosities, while below \( L_c^* \) the luminosity function is set by the jet (apparent) structure. The left-hand panel in Fig. 1 shows the luminosity function \( \phi(L) = dP/d\ln(L) = L \int P(L, E_p) dE_p \) for an example Gaussian apparent structure case, demonstrating the effect of introducing a core luminosity dispersion with three different values of the slope \( \alpha \).

We can get some further insight by adopting a power law apparent structure model with a “uniform core” within \( \theta_\ell \leq \theta_c \), namely

\[ \ell = \min \left[ 1, \left( \frac{\theta_c}{\theta_\ell} \right)^{\alpha} \right], \]

\[ \eta = \min \left[ 1, \left( \frac{\theta_c}{\theta_{c,p}} \right)^{\alpha} \right], \]

(13)

again with no dispersion. The change of variables approach can be applied to \( \theta_\ell > \theta_c \), where the structure is monotonic; for \( \theta_c \leq \theta_\ell \) it is sufficient to note that all observers see \( L = L_c \) and \( E_p = E_{p,c} \), and the probability of having a viewing angle in this range is \( 1 - \cos \theta_c \). Hence, the \( L - E_p \) distribution is

\[ P(L, E_p) = \frac{\theta_\ell}{\alpha L} \frac{\sin \theta_\ell \sqrt{2 \ln(L_c/L)}}{\sqrt{2 \ln(L_c/L)}} \delta \left( E_p - E_{p,c} \left( \frac{L}{L_c} \right)^{\alpha} \right) \]

\[ \approx \frac{\theta_\ell^2}{\ell} \delta \left( E_p - E_{p,c} \left( \frac{L}{L_c} \right)^{\alpha} \right). \]

(14)

for \( L_{\min}^* \leq L < L_c \) and \( E_{p,\min}^* \leq E_p < E_{p,c} \), where \( L_{\min}^* = L_c (2\theta_\ell/\pi)^{\alpha} \) and \( E_{p,\min}^* = E_{p,c}(2\theta_\ell/\pi)^{\alpha/2} \), and the last approximate equality is valid for \( L \gg L_{\min}^* \) and \( E_p \gg E_{p,\min}^* \). Again, the alternate form, which highlights the dependence on \( E_p \), can be obtained by the substitutions \( L \leftrightarrow E_p \), \( \alpha L \leftrightarrow \alpha E_p \), and \( L_c \leftrightarrow E_{p,c} \). For \( L \geq L_{\min}^* \geq E_p \geq E_{p,c} \), we have

\[ P(L, E_p) = (1 - \cos \theta_c) (L - L_c) \delta(E_p - E_{p,c}). \]

(15)

While in this limiting zero-dispersion case the jet core clearly extends over a zero-measure region of the \( (L, E_p) \) plane, a non-zero dispersion spreads this over a more physically sound, finite region of the plane. Hence, a power law apparent structure induces a power law luminosity function with a slope \( L^{-2/\alpha} \) in the logarithm of \( L \) (Pescalli et al. 2015). Similar conclusions apply to the induced \( E_p \) distribution. The slope parameter \( \alpha L \), together with the core half-opening angle \( \theta_c \), control the extent of the luminosity function as a consequence of the limited physical viewing angle range \( 0 \leq \theta_c \leq \pi/2 \). The right-hand panel in
Fig. 1 shows the luminosity function in an example power law case with $\alpha_L = 3$.

The above derivation also shows that in a quasi-universal structured jet scenario $\phi(L)$ can never attain a positive slope, except at the low-luminosity end, or in a narrow luminosity range close to the core for small dispersions (unless $\alpha_L < 0$, which however does not seem a likely physical possibility). This is a feature of quasi-universal jet models.

2.3.2. $L - E_p$ correlation induced by the jet structure

Since both $\ell$ and $\eta$ are functions of the viewing angle, the quasi-universal structured jet model inherently implies a Yonetoku correlation (Yonetoku et al. 2004) between $L$ and $E_p$ within the observed population (e.g. Salafia et al. 2015), in addition to any intrinsic correlation that may hold between the core quantities $L_c$ and $E_{p,c}$. The shape of the induced correlation can be obtained by eliminating $\theta_c$ in the average apparent structure functions, to obtain $\eta(\ell)$, and is already apparent in the Dirac-delta functions in Eqs. (12) and (14). In the Gaussian average apparent structure case, it yields $(E_p/E_{p,c}) = (L/L_c)^{(\theta_c/\theta_p)}$, so that the slope of the correlation is set by the ratio of the scale parameters $\theta_c$ and $\theta_p$, over which $L$ and $E_p$ decay with the viewing angle. Similarly, in the power law case the relation is $(E_p/E_{p,c}) = (L/L_c)^{\theta_p/\theta_c}$, where the ‘smoothness’ parameter is set to $s = 4$, which makes the transitions between the power law branches relatively sharp, to compensate the fact that the intrinsic dispersion of core quantities tends to smooth out the induced breaks in the luminosity function. The implied $L - E_p$ correlation has two branches, with slopes $\alpha_L/\alpha_p$ and $\beta_p/\beta_c$, the break being around $L \sim L_p^* = L_c^{\alpha_p/\alpha_L} \theta_p/\theta_c$ and $E_p \sim E_{p,w}^* = E_{p,c}^{\theta_p/\theta_c} \theta_p/\theta_c$. Figure 2 demonstrates the features of such a model in detail, including the induced $L$ and $E_p$ probability distributions, based on an example choice of parameters.

With this choice of average apparent jet structure functions, the model features a total of 14 free parameters. We list these parameters in Table 1, along with brief definitions and with information on the priors adopted on each of them in the analysis, discussed later in the text.

2.3.3. Average apparent structure model adopted in this work

In this study, in order to endow the average apparent structure model with a high degree of flexibility (in the absence of strong theoretical constraints on the expected shape), we adopted a double smoothly broken power law model with a nearly constant core within $\theta_c < \theta_c$ and a break at a wider angle $\theta_c$, that is,

$$
\ell(\theta_c) = \left[ 1 + \left( \frac{\theta_c}{\theta_p} \right)^{\frac{1}{\alpha_L}} \right]^{-\alpha_L/s} \left[ 1 + \left( \frac{\theta_c}{\theta_p} \right)^{\frac{1}{\beta_p}} \right]^{-\beta_p/s} \frac{\theta_c}{\theta_p} \theta_p^{\alpha_p/\alpha_L} \theta_c^{\beta_c/\beta_p},
$$

$$
\eta(\ell) = \left[ 1 + \left( \frac{\theta_p}{\theta_c} \right)^{\frac{1}{\alpha_L}} \right]^{-\alpha_p/s} \left[ 1 + \left( \frac{\theta_p}{\theta_c} \right)^{\frac{1}{\beta_p}} \right]^{-\beta_p/s} \frac{\theta_p}{\theta_c} \theta_c^{\alpha_c/\alpha_L} \theta_p^{\beta_c/\beta_p},
$$

where the ‘smoothness’ parameter is set to $s = 4$, which makes the transitions between the power law branches relatively sharp, to compensate the fact that the intrinsic dispersion of core quantities tends to smooth out the induced breaks in the luminosity function. The implied $L - E_p$ correlation has two branches, with slopes $\alpha_L/\alpha_p$ and $\beta_p/\beta_c$, the break being around $L \sim L_p^* = L_c^{\alpha_p/\alpha_L} \theta_p/\theta_c$ and $E_p \sim E_{p,w}^* = E_{p,c}^{\theta_p/\theta_c} \theta_p/\theta_c$. Figure 2 demonstrates the features of such a model in detail, including the induced $L$ and $E_p$ probability distributions, based on an example choice of parameters.

With this choice of average apparent jet structure functions, the model features a total of 14 free parameters. We list these parameters in Table 1, along with brief definitions and with information on the priors adopted on each of them in the analysis, discussed later in the text.

2.4. Sample definition and selection effect modelling

In order to compare a population model with an observed sample, the selection effects that shape the latter must be taken into account. Here we describe our sample choice and the procedure that we employed to model the underlying selection effects. We adopted a description of the SGRB photon spectrum as a cut-off power law (Ghirlanda et al. 2004), $dN/dE(E, E_{p,\text{obs}}, \alpha) \propto E^\alpha \exp[-(2 + \alpha)E/E_{p,\text{obs}}]$, where $N$ represents the rate of photons hitting the detector, $E$ is the photon energy, and $E_{p,\text{obs}} = E_p/(1 + z)$. We set the low-energy photon index to $\alpha = -0.4$, which is the median of the values reported in the *Fermi*GBM online catalogue for SGRBs. The peak photon flux in the $E_0 - E_1$
keV band is then defined as

$$p(E_{\text{p}}, E_{\text{p}, \text{z}}) = \frac{L}{4\pi d_L^2(E_{\text{p}, \text{z}})} \left( \frac{E_{\text{p}}}{E_{\text{p}, \text{z}}} \right)^{\alpha} \frac{dN}{dE(E_{\text{p}}, \alpha) dE}.$$  

(17)

where $d_L$ is the luminosity distance and

$$E_{\text{p}, \text{z}}(E_{\text{p}, \text{z}}, \theta_0) = \frac{1}{\Omega_{\text{z}}(1+z)^2} \frac{dN}{dE(E_{\text{p}, \text{z}}, \alpha)} dE.$$  

(18)

We stress here again that we extended the customary $1-10^5$ keV pseudo-bolometric rest-frame band to the wider $0.1-10^7$ keV to include possible cases with very high $E_{\text{p}}$. In practice, we pre-computed $E_{\text{p}, \text{z}}(50-300)$ for Fermi/GBM and $E_{\text{p}, \text{z}}(15-150)$ for Swift/BAT over a uniformly spaced two-dimensional grid in $(\log E_{\text{p}}, \log z)$ and then used two-dimensional linear interpolation to recover it and obtain the photon flux from Eq. (17) (or equivalently to obtain $L$ from $p$ and $E_{\text{p}, \text{obs}}$, by inverting the equation).

3 While Fermi/GBM is sensitive over a larger energy band, and the results in the catalogue usually refer to the $10-10^5$ keV band, the $50-300$ keV band is where most of the online GRB trigger algorithms look for excess (von Kienlin et al. 2020), so that the flux in that band is the most relevant for what concerns the modelling of the GBM detection – see Appendix C.

2.4.1. Observer-frame sample: Fermi/GBM SGRBs with spectral information

We considered three reference SGRB samples: (i) SGRBs detected by Fermi/GBM, with available spectral information in the public catalogue; (ii) SGRBs detected by both Swift/BAT and Fermi/GBM, with a number of additional cuts to reach a high completeness in redshift; and (iii) SGRBs detected by Fermi/GBM with a GW counterpart, which currently includes only GRB 170817A/GW170817. In what follows, we describe in detail the selection cuts of each sample and our approach to the modelling of selection effects.
Table 1. Population model parameters and adopted priors.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_c$</td>
<td>Isotropic, 0.01 rad $\leq \theta_c &lt; \theta_w \leq \pi/2$ rad</td>
<td>Core half-opening angle</td>
</tr>
<tr>
<td>$\theta_w$</td>
<td>Uniform-in-log, $3 \times 10^{51}$ erg s$^{-1} \leq L^*_c \leq 10^{55}$ erg s$^{-1}$</td>
<td>‘break’ angle</td>
</tr>
<tr>
<td>$L^*_c$</td>
<td>Uniform, $0 \leq \alpha_L &lt; 6$</td>
<td>Low-end cutoff luminosity for on-axis observers ($\theta_c = 0$)</td>
</tr>
<tr>
<td>$\beta_L$</td>
<td>Uniform, $-3 \leq \beta_L &lt; 6$</td>
<td>Slope of $\ell$ at intermediate viewing angles $\theta_c &lt; \theta_L &lt; \theta_w$</td>
</tr>
<tr>
<td>$E^*_{p,c}$</td>
<td>Uniform-in-log, $10^{5}$ keV $\leq E^*_{p,c} \leq 10^{5}$ keV</td>
<td>Low-end cutoff $E_p$ for on-axis observers ($\theta_c = 0$)</td>
</tr>
<tr>
<td>$\alpha_{E_p}$</td>
<td>Uniform, $0 \leq \alpha_{E_p} &lt; 6$</td>
<td>Slope of $\eta$ at intermediate viewing angles $\theta_c &lt; \theta_L &lt; \theta_w$</td>
</tr>
<tr>
<td>$\beta_{E_p}$</td>
<td>Uniform, $-3 \leq \beta_{E_p} &lt; 10$</td>
<td>Slope of $\eta$ at large viewing angles $\theta_c &gt; \theta_w$</td>
</tr>
<tr>
<td>$A$</td>
<td>Uniform, $1.5 \leq A &lt; 5$</td>
<td>Slope of $L_e$ distribution</td>
</tr>
<tr>
<td>$\sigma_c$</td>
<td>Uniform-in-log, $0.3 \leq \sigma_c &lt; 3$</td>
<td>$E_p$ dispersion around $E_p = E^<em>_{p,c}(L_c/L^</em>_c)^\sigma$</td>
</tr>
<tr>
<td>$y$</td>
<td>Uniform, $-3 &lt; y &lt; 3$</td>
<td>Rate density evolution slope at low redshift, $\rho \propto (1 + z)^y$</td>
</tr>
<tr>
<td>$a$</td>
<td>Uniform, $-1 &lt; a &lt; 5$</td>
<td>Rate density decay slope at high redshift, $\rho \propto (1 + z)^{-a}$</td>
</tr>
<tr>
<td>$b$</td>
<td>Uniform, $1 &lt; b &lt; 10$</td>
<td>Redshift of rate density peak</td>
</tr>
<tr>
<td>$\zeta_p$</td>
<td>Uniform, $0.1 &lt; \zeta_p &lt; 3$</td>
<td></td>
</tr>
</tbody>
</table>

Notes. (1)Isotropic = uniform in the subtended solid angle, i.e. $\pi(\theta) = \sin \theta$. This selection criterion significantly reduces the sample size: out of a total of 367 SGRBs in the raw sample, only 212 have $p_{50-300} > 3.5 \text{cm}^{-2} \text{s}^{-1}$. Most importantly, the discarded events possibly probe the luminosity function down to lower luminosities, which is where most of the useful information on the jet structure resides in a quasi-universal jet scenario. Last, but not least, the only SGRB with reliable viewing angle information, that is, GRB 170817A, is not included in this flux-limited sample. In order to access the flux-incomplete part of the Fermi/GBM SGRB sample, we carefully constructed a detection probability $P_{\text{det}}(L, E_p, z)$ by simulating the response of the GBM NaI detectors to SGRBs with a broad range of characteristics, as described in detail in Appendix C. In this study, we compare the results obtained by using either of the strategies in modelling the selection effects: hereafter, we refer to the reduced sample of Fermi/GBM SGRBs with $p_{50-300} > 3.5 \text{cm}^{-2} \text{s}^{-1}$ as the ‘flux-limited sample’, and to the analysis adopting the associated simplified selection effect model as the ‘flux-limited sample analysis’. Conversely, the analysis performed adopting the simulated Fermi/GBM detection efficiency is referred to as the ‘full sample analysis’.

As a further countermeasure against possible biases, we applied an additional quality cut to both the above samples: we removed events with best-fit $E_{\text{p,obs}} > 10 \text{MeV}$ or $E_{\text{p,obs}} < 50 \text{keV}$, which fall outside the spectral range where the effective area of the GBM detectors is optimal: this removes two events whose uncertainty on $E_{\text{p,obs}}$ is very large, reducing the flux-limited sample to 210 events. In the full sample analysis, we also removed another ten events with a low best-fit peak 64 ms photon flux $p_{50-300} < 1 \text{cm}^{-2} \text{s}^{-1}$, all of which have very large errors on both $p_{50-300}$ and $E_{\text{p,obs}}$. This reduces the ‘full’ sample to 355 events. To reflect these further quality cuts, we updated our model detection probabilities by multiplying them by $\Theta(p_{50-300} < 1 \text{cm}^{-2} \text{s}^{-1})\Theta(E_{\text{p,obs}} > 10 \text{MeV})\Theta(p_{50-300})$.  

2.4.2. Rest-frame sample: Flux-complete sample of Swift/BAT SGRBs also observed by Fermi/GBM

For some of the events in our Fermi/GBM observer-frame sample, redshift information is available and can be used to constrain the population parameters better than can be done using the observer-frame information ($p_{50-300}$, $E_{\text{p,obs}}$) only. In order
Table 2. Rest-frame SGRB sample.

<table>
<thead>
<tr>
<th>GRB</th>
<th>GBM name</th>
<th>$T_{90}$/s</th>
<th>$z$</th>
<th>Redshift source (*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>080905A</td>
<td>GB0809054999</td>
<td>0.960 ± 0.345</td>
<td>0.1218 ± 0.0003</td>
<td>Fong et al. (2022; Gold)</td>
</tr>
<tr>
<td>090510</td>
<td>GB090510016</td>
<td>0.960 ± 0.138</td>
<td>0.903 ± 0.002</td>
<td>Rau et al. (2009), Levan (2009)</td>
</tr>
<tr>
<td>100117A</td>
<td>GB1001178799</td>
<td>0.256 ± 0.834</td>
<td>0.914 ± 0.0004</td>
<td>Fong et al. (2022; Gold)</td>
</tr>
<tr>
<td>100206A</td>
<td>GB1002066563</td>
<td>0.176 ± 0.072</td>
<td>0.40 ± 0.002</td>
<td>Fong et al. (2022; Gold)</td>
</tr>
<tr>
<td>100625A</td>
<td>GB1006257773</td>
<td>0.240 ± 0.276</td>
<td>0.452,002</td>
<td>Fong et al. (2022; Silver)</td>
</tr>
<tr>
<td>111117A</td>
<td>GB1111177510</td>
<td>0.432 ± 0.082</td>
<td>2.211 ± 0.001</td>
<td>Fong et al. (2022; Silver)</td>
</tr>
<tr>
<td>130515A</td>
<td>GB1305150566</td>
<td>0.256 ± 0.091</td>
<td>0.80 ± 0.01</td>
<td>Fong et al. (2022; Silver)</td>
</tr>
<tr>
<td>130912A</td>
<td>GB1309123588</td>
<td>0.512 ± 0.143</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>131004A</td>
<td>GB1310049046</td>
<td>1.152 ± 0.590</td>
<td>0.717 ± 0.002</td>
<td>Fong et al. (2022; Silver)</td>
</tr>
<tr>
<td>160821B</td>
<td>GB1608219373</td>
<td>1.088 ± 0.977</td>
<td>0.1619 ± 0.0002</td>
<td>Fong et al. (2022; Silver)</td>
</tr>
<tr>
<td>170127B</td>
<td>GB1701276344</td>
<td>1.728 ± 1.346</td>
<td>2.28 ± 0.14 (***)</td>
<td>Fong et al. (2022; Silver)</td>
</tr>
<tr>
<td>180204A</td>
<td>GB1802041092</td>
<td>1.152 ± 0.991</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>180727A</td>
<td>GB1807275994</td>
<td>0.896 ± 0.286</td>
<td>1.95 ± 0.50 (***)</td>
<td>Fong et al. (2022; Gold)</td>
</tr>
<tr>
<td>191031D</td>
<td>GB1910318914</td>
<td>0.256 ± 0.023</td>
<td>1.93 ± 1.40 (***)</td>
<td>Fong et al. (2022; Silver)</td>
</tr>
<tr>
<td>200219A</td>
<td>GB2002193175</td>
<td>1.152 ± 1.032</td>
<td>0.48 ± 0.02</td>
<td>Fong et al. (2022; Gold)</td>
</tr>
<tr>
<td>200411A</td>
<td>GB2004111887</td>
<td>1.440 ± 0.506</td>
<td>0.82 ± 0.18 (***)</td>
<td>Fong et al. (2022; Bronze)</td>
</tr>
<tr>
<td>210221D</td>
<td>GB2102219634</td>
<td>0.144 ± 0.066</td>
<td>1.046 ± 0.002</td>
<td>de Ugarte Postigo et al. (2020)</td>
</tr>
<tr>
<td>210323A</td>
<td>GB2103239186</td>
<td>0.960 ± 0.781</td>
<td>0.733 ± 0.001</td>
<td>Fong et al. (2022; Gold)</td>
</tr>
</tbody>
</table>

Notes. The sample consists of SGRBs belonging to the extended S-BAT4 sample (original sample: D'Avanzo et al. 2014; extended sample: Ferro et al. 2023, D’Avanzo et al., in prep.) and also detected by Fermi/GBM with $T_{90} < 2$s. (*)Fong et al. (2022) host galaxy class in parentheses, where available, based on the probability $P_e$ of a chance association with the SGRB. Gold: $P_e \leq 0.02$; Silver: $0.02 < P_e \leq 0.1$; Bronze: $0.1 < P_e \leq 0.2$. (***)Photometric redshift posterior samples retrieved from the BRIGHT online catalogue at https://bright.ciera.northwestern.edu/.

to avoid biases, on the other hand, any additional selection effects at play in the sub-sample with known redshift must be accounted for in the inference, that is, in the $P_{\text{det}}$ model. The redshift determination is a complex process that involves multiple facilities and depends not only on the prompt emission properties, but also on those of the afterglow. Therefore, modelling the associated selection effects is prohibitive. On the other hand, Salvaterra et al. (2012) and D’Avanzo et al. (2014) showed that it is possible to construct a sample with a selection that is easier to model, but that leads to a high redshift completeness. The selection involves two cuts that do not bias the redshift distribution, namely (i) a cut on the foreground interstellar dust extinction $A_V$ and (ii) a cut on the Swift X-Ray Telescope (XRT) slew time; plus a cut on the BAT 64 ms peak flux in the 15–150 keV band, $p_{15-150} > 3.5$ cm$^{-2}$ s$^{-1}$, to ensure flux completeness. The original SGRB sample constructed in this way, known as the S-BAT4 (D’Avanzo et al. 2014), included 16 events, 11 of which had a measured redshift. Thanks to a considerable effort spent by the community, and in particular by Fong et al. (2022) and Nugent et al. (2022), in identifying SGRB host galaxies and measuring their redshifts, it has been recently possible to construct an extended ‘S-BAT4ext’ sample (Ferro et al. 2023; D’Avanzo et al., in prep.) with a more than doubled size and an increased redshift completeness. For this work, we adopt the S-BAT4ext sub-sample of 18 events that have been jointly detected by Fermi/GBM with $T_{90} < 2$s (for consistency with our observer-frame sample). Of these, 16 have either a spectroscopic (12 events) or photometric (4 events) redshift measurement, as listed in Table 2. For these SGRBs we performed an independent analysis of their peak spectra, based on publicly available Fermi/GBM data, as described in Appendix B, obtaining posterior samples of their bolometric flux $F$ and observed peak photon energy $E_{p,\text{obs}}$, adopting a prior $\pi(F, E_{p,\text{obs}}) \propto F^{-1}$. For the 12 events with a spectroscopic redshift, these were converted into samples of $P(L, E_p, z | d_j)$ by simply fixing the redshift at the best fit value. For the four events with photometric redshift, we obtained samples of $P(L, E_p, z | d_j)$ by computing $L_{\text{det}} = 4\pi c^2 F_{\text{m}}$ and $E_{p,\text{det}} = (1 + z)E_{p,\text{obs}}$, where $[z]_{\text{det}}$ are $N_s$ photometric redshift posterior samples from the Broadband Repository for Investigating Gamma-ray burst Host galaxies Traits (BRIGHT) catalogue (Nugent et al. 2022), $\{d_j\}_{j=1}^{N_s}$ are the corresponding luminosity distances under our assumed cosmology and $[(E_m, E_{p,\text{obs}})]_{m=1}^{N_m}$ are $M_s$ posterior samples from the spectral analysis. In both cases, the effective prior on the source parameters is $\pi(L, E_p, z) \propto (1 + z)^{-1}L^{-1}$. The two events with unknown redshift were not included in the analysis.

The careful selection adopted to construct this sample allows us to model the underlying selection effects by multiplying the GBM detection efficiency, $P_{\text{det,GBM}}$, by the simple BAT detection efficiency:

$$P_{\text{det,BAT}} = \Theta(p_{15-150}(L, E_p, z) - p_{\text{lim,BAT}}),$$

where $p_{\text{lim,BAT}} = 3.5$ cm$^{-2}$ s$^{-1}$ is numerically identical to $p_{\text{lim,GBM}}$ by pure chance.

2.4.3. Viewing angle sample: GRB 170817A/GW170817

The third and last sample we considered is that of Fermi/GBM SGRBs that have a GW counterpart produced by the inspiral of a BNS merger, from which a measurement of $\theta$ can be obtained under the assumption that the jet is launched along the direction of the total angular momentum. At present, the sample clearly consists of the single event GRB 170817A with its counterpart GW170817. We use $d_{\text{HOST}}$ to indicate the available data regarding the prompt emission and the GW signal of the event, and with $d_{\text{HOST}}$ the available data of the host galaxy NGC4993 (Coulter et al. 2017; Hjorth et al. 2017; Cantoriello et al. 2018). We took the host galaxy spectroscopic redshift $z_\text{host} = 0.009783$ (Coulter et al. 2017; Hjorth et al. 2017) as the redshift of the
formed by Cantiello et al. 2018). We then computed the posterior of statistical and systematic errors from the measurements per-

\[ P(\theta_r, r | d_{G17}) = \int_0^\infty P(\theta_r, r | d_{G17}) P(r | d_{\text{HOST}}) \, dr, \]

where \( P(\theta_r, r | d_{G17}) \) is the joint posterior on \( \theta_r \) and \( r \) from the GW analysis. In practice, we obtained samples of the posterior probability density in Eq. (21) by re-sampling the publicly available posterior samples from the low-spin prior analysis of GW170817 performed by Abbott et al. (2019) with a weight equal to the right-hand side of Eq. (21) evaluated at the luminosity distance of each sample. The viewing angle posterior probability density obtained from a kernel density estimate on the resulting samples is shown in Fig. 5.

We modelled the selection effects acting on this sample as the product of \( P_{\text{det,GBM}}(L, E_p, z) \) times a GW detection efficiency \( P_{\text{det,GW}}(\theta_r, z) \). Since the time-volume surveyed by aLIGO and Advanced Virgo so far is dominated by that of the third observing run, O3 (LIGO Scientific Collaboration 2021), we constructed the detection efficiency assuming, for the sake of simplicity, the GW network sensitivity of O3, neglecting periods with a lower sensitivity. To do so, we retrieved the dataset made publicly available by the LIGO Scientific Collaboration et al. (2023) that contains information on the response of online GW search pipelines to a large number of simulated signals injected into O3 data. We re-weighted the BNS merger injections in that dataset to reflect a population with a primary mass distribution \( p(m_1) \propto m_1^{-2} \) between \( m_{1,\text{min}} = 1.2 \, M_\odot \) and \( m_{1,\text{max}} = 2.1 \, M_\odot \).
Advanced Virgo detectors in our case, \( \pi(\lambda_{\text{pop}}) \) is the prior on the population parameters, and \( P(d) = \prod_{i=1}^{N} P(d_i) \) is a normalisation factor. We indicate with \( N_i = N(d_i, \lambda_{\text{pop}}) \) the numerator of the fraction after the product symbol in the above equation, that is,

\[
N_i = \int P(d_i \mid \lambda_{\text{src}}) P_{\text{pop}}(\lambda_{\text{src}} \mid \lambda_{\text{pop}}) \, d\lambda_{\text{src}},
\]  

(26)

and with \( D = D(\lambda_{\text{pop}}) \) the denominator, namely

\[
D = \int P_{\text{pop}}(\lambda_{\text{src}} \mid \lambda_{\text{pop}}) P_{\text{det}}(\lambda_{\text{src}}) \, d\lambda_{\text{src}}.
\]  

(27)

This equation contains the detection efficiency \( P_{\text{det}}(\lambda_{\text{src}}) \), which must be chosen consistently with the selection effects acting on each of the considered samples.

2.5.1. Numerator \( N_i \) for observer-frame sample events

For events in our observer-frame sample, which have unknown redshifts, we approximated the posterior probability on the source parameters \( P(L, \lambda_p, z \mid d) = P(d \mid \lambda_{\text{src}}) \pi(L, \lambda_p, z) \mid P(d) \) by neglecting the uncertainty on the measured peak photon flux \( p_i = p_{[50-300]} \) and peak photon energy \( E_{p, \text{obs}, i} \) (which is justified by the large sample size and by the quality cuts discussed in the previous section), and by assuming the posterior to simply reflect the prior \( \pi(z) \) along the \( z \) axis. In other words, we assumed \( P(L, \lambda_p, z \mid d) \propto \pi(z) \) when keeping \( L \) and \( E_p \) fixed: this is equivalent to stating that no redshift information is available. This leads to

\[
P(d_i \mid \lambda_{\text{src}}) \sim \left( L - 4\pi d_i^2 E_p \right) \delta \left( E_p - (1 + z) E_{p, \text{obs}, i} \right) P(d_i).
\]  

(28)

The appropriate prior \( \pi(L, E_p) \) was obtained by applying a coordinate transform to a uniform prior in both \( p_i \) and \( E_{p, \text{obs}, i} \) for these events, that is,

\[
\pi(L, E_p) = \left| \left| \begin{array}{c} \frac{\partial \lambda_p}{\partial L} \\ \frac{\partial \lambda_p}{\partial E_p} \end{array} \right| \right| \pi(p_i, E_{p, \text{obs}, i}) \propto \frac{p_i}{(1 + z) \hat{L}_i(z)},
\]  

(29)

where \( \left| X \right| \) represents the determinant of matrix \( X \), and \( \hat{L}_i = 4\pi d_i^2 (1 + z) E_{p, \text{obs}, i} \). The integral in the numerator of Eq. (25) is then

\[
N_i \sim P(d_i) \int_0^{\infty} \frac{1 + z}{p_i} \hat{L}_i(z) \frac{E_{p, \text{obs}, i} \mid \lambda_{\text{pop}}}{1 + z} \, dz,
\]  

(30)

and we note that the \( P(d_i) \) term eventually cancels out in Eq. (25) (as in M19).

2.5.2. Numerator \( N_i \) for rest-frame sample events

For events in the rest-frame sample, in order to carry out the integral over \( z, L \) and \( E_p \) in \( N_i \), we employed the same Monte Carlo approximation as M19, that is,

\[
N_i \sim \frac{P(d_i)}{N_i} \sum_{j=1}^{N_i} \frac{P_{\text{pop}}(L_{i,j}, E_{p,i,j} | z_{i,j} \mid \lambda_{\text{pop}})}{\pi(E_{p,i,j} | \lambda_{\text{pop}})}.
\]  

(31)

where \( \{L_{i,j}, E_{p,i,j}, z_{i,j}\}_{j=1}^{N_i} \) represent a total of \( N_i \) samples of the luminosity, peak photon energy, and redshift posterior of the

**Fig. 5.** GRB 170817A viewing angle posterior probability distribution, assuming the jet to be aligned with the GW170817 binary total angular momentum. The blue line shows the posterior probability constructed using the posterior samples from the low-spin-prior GW analysis (Abbott et al. 2019), while the red line shows the result of conditioning on the host galaxy distance (Cantiello et al. 2018), as explained in the main text.

(similar to the preferred distribution from the GWTC-3 population analysis, Abbott et al. 2023) and a flat secondary mass distribution \( p(m_2 \mid m_1) \propto \Theta(m_1 - m_2) \Theta(m_2 - m_{1,\text{min}}) \). Assuming the inclination \( i \) and the viewing angle \( \theta_i \), to be related by

\[
\theta_i(i) = \left\{ \begin{array}{ll} 0 & 0 \leq \theta < \pi/2 \\ \pi - \theta & \pi/2 < \theta < \pi \end{array} \right.,
\]  

(23)

we binned the injected signals into a number of two-dimensional bins in \((\theta_i, z)\) space centred at \((\theta_i, z_i)\). Calling \( w_k \) and \( \rho_i \) the weight and network signal-to-noise ratio (S/N) associated with the \( k \)-th injected signal, we estimated \( P_{\text{det}, \text{GW}}(\theta_i, z) \) at the centre of each bin as

\[
P_{\text{det}, \text{GW}}(\theta_i, z) \sim \frac{\sum_{k \in I_{i,j}} w_k \Theta(p_k \geq 12)}{\sum_{k \in I_{i,j}} w_k},
\]  

(24)

where \( I_{i,j} \) represents the set of indices \( k \) of injections whose viewing angle and redshift fall into the bin centred at \((\theta_i, z_i)\). The GW detection efficiency on the rest of the \((\theta_i, z)\) space was obtained by two-dimensional linear interpolation of the resulting values. The relatively high cut \( \rho \geq 12 \) includes GW170817 (Abbott et al. 2017a) and ensures that the detection can be represented by a simple cut in \( \text{S/N} \), in analogy with the flux completeness cuts discussed previously.

### 2.5. Inference on the population properties

Within a Bayesian hierarchical approach, the posterior probability on the population parameters \( \lambda_{\text{pop}} \) can be written as (Eqs. (7) and (8) in Mandel et al. 2019, hereafter M19)

\[
P(\lambda_{\text{pop}} \mid d) = \frac{\pi(\lambda_{\text{pop}})}{P(d)} \prod_{i=1}^{N} \left[ \frac{P(d_i \mid \lambda_{\text{src}}) P_{\text{pop}}(\lambda_{\text{src}} \mid \lambda_{\text{pop}}) \, d\lambda_{\text{src}}}{\int P_{\text{pop}}(\lambda_{\text{src}} \mid \lambda_{\text{pop}}) P_{\text{det}}(\lambda_{\text{src}}) \, d\lambda_{\text{src}}} \right],
\]  

(25)

where the index \( i \) runs over the \( N \) events in the sample, \( d_i \) (the \( i \)-th element of \( d \)) representing the corresponding data in the detectors (i.e. Fermi/GBM, plus either Swift/BAT or the aLIGO and
GRB. In cases with a photometric redshift, the \((L, E_p)\) posterior was obtained by combining the results from our analysis of the spectrum at the peak of the GRB with the redshift posterior samples from Nugent et al. (2022), obtained from the BRIGHT online catalogue\(^4\). In cases with a spectroscopic redshift, we simply kept \(z_{i,j}\) fixed at the best-fit redshift. The prior \(\pi(L, E_p)\) in Eq. (31) is proportional to \((1 + z)^{-1}L^{-1}\), as discussed in Sect. 2.4.2.

### 2.5.3. Numerator \(N_i\) for viewing angle sample events

For the viewing angle sample, that is, for GRB 170817A, it was necessary to proceed differently for the full sample and flux-limited sample analyses: the peak photon flux of the GRB, \(p_{50-300}\) = \(2.02\) cm\(^{-2}\) s\(^{-1}\), is below the completeness cut \(p_{\text{lim,GBM}} = 3.5\) cm\(^{-2}\) s\(^{-1}\) adopted in the flux-limited sample analysis; hence, in this case, the simple treatment of GBM selection effects is not adequate. In the full sample analysis, this is not a problem since the GBM detection efficiency model from Appendix C accounts for the smooth decrease in the detection efficiency at low peak photon fluxes. In the full-sample case, the correct form of \(N_i\) can be obtained by reformulating our population model by including \(\theta_i\) among the source parameters, \(N_{\text{LC}} = \{(L, E_p, z, \theta_i)\}\). The population probability then becomes 

\[
N_i = \prod_{j=1}^{N_i} P(d_i | L, E_p, z, \theta_i) \times P(L, E_p, \theta_i | A_{\text{pop}}) \sin \theta_i P(z | A_{\text{pop}}) \, dL \, dE_p \, dz \, d\theta_i. \tag{32}
\]

It is straightforward to verify that whenever \(d_i\) does not contain information on the viewing angle this leads to exactly the same definition of \(N_i\) as before.

For GRB 170817A, we have (see Sect. 2.4.3)

\[
P(d_{g17}, d_{\text{HOST}} | L, E_p, \theta_i) = \frac{P(\theta_i | d_{g17}, d_{\text{HOST}})}{\pi(\theta_i)} \times \frac{P(L, E_p, \theta_i | z)}{P(L, E_p, z)} P(z | A_{\text{pop}}) \, dL \, dE_p \, dz \, d\theta_i.
\]

where the prior \(\pi(\theta_i) = \sin \theta_i\) corresponds to that used in the GW analysis. This leads to

\[
N_{g17} = \prod_{i=1}^{N_i} P(\theta_i | d_{g17}, d_{\text{HOST}}) \frac{P(L, E_p, \theta_i | z)}{\pi(\theta_i)} \times \frac{P(L, E_p, \theta_i | z)}{P(L, E_p, z)} P(z | A_{\text{pop}}) \, dL \, dE_p \, dz \, d\theta_i.
\]

Approximating again the integral with a Monte Carlo sum, we obtained

\[
N_{g17} = \frac{P(z | A_{\text{pop}})}{N_{g17} N_{g17}} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} P(L, E_p, \theta_i, \theta_j | A_{\text{pop}}) \frac{P(L, E_p, \theta_i, \theta_j | z)}{\pi(\theta_i)} \times \frac{P(L, E_p, \theta_i, \theta_j | z)}{P(L, E_p, \theta_i, \theta_j | z)} P(z | A_{\text{pop}}) \, dL \, dE_p \, dz \, d\theta_i.
\]  \tag{33}

where \(\{(L, E_p, \theta_i)\}_{i=1}^{N_i}\) are a total of \(N_i\) samples from our analysis of the peak spectrum of GRB 170817A and \(\{(\theta_i, \theta_j)\}_{i=1}^{N_j}\) are \(N_j\) posterior samples of the probability in Eq. (22). Clearly, the term \(P(d_{g17}, d_{\text{HOST}})\) eventually cancels out in Eq. (25) as before.

In the flux-limited sample analysis, the information on the viewing angle, luminosity and peak photon energy of

\[\text{GRB 170817A/GW170817 can still be used to condition the apparent structure parameters before the flux-limited sample analysis is performed, because the structure must be consistent with what has been observed in that event. The starting point is}
\]

\[
P(A_{\text{pop}} | d_{g17}, d_{\text{HOST}}) = \int \int \int P(A_{\text{pop}} | L, E_p, \theta_i) P(L, E_p, \theta_i | d_{g17}, d_{\text{HOST}}) \, dL \, dE_p \, d\theta_i. \tag{34}
\]

\[\text{Application of Bayes’ theorem gives}
\]

\[
P(A_{\text{pop}} | L, E_p, \theta_i) = \frac{P(L, E_p, \theta_i | A_{\text{pop}}, \pi(A_{\text{pop}}))}{\pi(\theta_i) \pi(L, E_p))}
\]

but we have \(P(L, E_p, \theta_i | A_{\text{pop}}) = P(L, E_p | \theta_i, A_{\text{pop}}) \sin \theta_i \pi(\theta_i) = \sin \theta_i,\) which therefore leads to \(P(A_{\text{pop}} | L, E_p, \theta_i) = P(L, E_p | \theta_i, A_{\text{pop}}) \pi(A_{\text{pop}})/\pi(L, E_p).\) Substitution of this into Eq. (34) leads to

\[
P(A_{\text{pop}} | d_{g17}, d_{\text{HOST}}) = \pi(A_{\text{pop}}) \int \int P(L, E_p, \theta_i | d_{g17}, d_{\text{HOST}}) \times \frac{P(L, E_p | \theta_i, A_{\text{pop}})}{\pi(L, E_p)} \, dL \, dE_p \, d\theta_i. \tag{36}
\]

This is similar to \(N_{g17},\) but the redshift information is not used. This can again be approximated by Monte Carlo integration over samples drawn from the posterior \(P(L, E_p, \theta_i | d_{g17}, z_{\text{fit}}),\) namely

\[
P(A_{\text{pop}} | d_{g17}, d_{\text{HOST}}) \sim \frac{\pi(A_{\text{pop}})}{N_{g17} N_{g17}} \sum_{k=1}^{N_k} \sum_{j=1}^{N_j} P(L_k, E_{p,k} | \theta_{i,j}, A_{\text{pop}}) \frac{P(L_k, E_{p,k}, \theta_{i,j}, z_{\text{fit}})}{\pi(L_k, E_{p,k}, \theta_{i,j}, z_{\text{fit}})}.
\]  \tag{37}

This result can then be used as a prior in the analysis of the observer-frame and rest-frame samples. The final posterior on \(A_{\text{pop}}\) then becomes

\[
P_{\text{flux-limited}}(A_{\text{pop}} | d, d_{g17}, d_{\text{HOST}}) \sim \frac{\pi(A_{\text{pop}})}{P(d)} \left[ \prod_{i=1}^{N_i} \frac{N_i}{N_{g17} N_{g17}} \sum_{j=1}^{N_j} \sum_{k=1}^{N_k} P(L_k, E_{p,k} | \theta_{i,j}, A_{\text{pop}}) \frac{P(L_k, E_{p,k}, \theta_{i,j}, z_{\text{fit}})}{\pi(L_k, E_{p,k}, \theta_{i,j}, z_{\text{fit}})} \right].
\]

Hence, the posterior takes a similar form as in the full-sample analysis case, the difference being in a missing \(P(z | A_{\text{pop}})/\mathcal{D}\) factor.

### 2.5.4. Denominator \(\mathcal{D}\) for events in the three samples

The denominator \(\mathcal{D}\) in the above expressions represents the fraction of events in the population that pass the sample selection criteria (M19), that is, the ‘accessible’ events according to the selection effects model. For the observer-frame sample events, the denominator takes the form

\[
\mathcal{D}(A_{\text{pop}}) = \int \int \int P_{\text{pop}}(L, E_p, z | A_{\text{pop}}) P_{\text{det,GBM}}(L, E_p, z) \, dL \, dE_p \, dz.
\]  \tag{38}

---

\(^4\) https://bright.ciera.northwestern.edu/
For the rest-frame sample, it is averaged over all walkers, is around 650. A corner plot showing the density of the samples in the 14-dimensional parameter space is shown in Fig. A.1, while summary statistics for each parameter are reported in Table 3. The results presented in the next section are constructed using 1000 random posterior samples from these chains, after discarding the first half as burn-in.

3. Results

Generally speaking, the full sample and flux-limited sample analyses yielded quite similar results, as demonstrated by the detailed posterior probability density distributions (Fig. A.1) and the summary in Table 3. The main notable differences in the full sample analysis versus the flux-limited sample analysis are a preference for a slightly larger on-axis SED peak photon energy $E_{\text{pc}z}$: a steeper slope, $a$, of the rate density evolution before peak; and a slightly better constrained redshift $z_\text{p}$ of the peak of the rate density evolution. The posterior distributions from the two analyses show large overlaps for all parameters. In what follows, we present a thorough description of the results, further highlighting the differences in the results from the two analyses when relevant.

3.1. Apparent structure

First of all, we focus on the apparent jet structure. Figure 6 shows the constraint we obtained on the ‘average apparent jet structure’ $L = L_\text{c}\ell$ (central panel) and $E_p = E_{p,\text{c}}\eta$ (bottom panel) from the two analyses, with the insets showing the posterior distribution on the core dispersion parameters $A$ and $\sigma_c$. In this and the following figures, solid lines refer to the full-sample analysis and dashed lines to the flux-limited sample analysis. The main panel at the top shows the posterior probability density on the logarithm of $\theta_c$ (in grey) and of $\theta_{\text{th}}$ (in green). The figure shows that the preferred apparent structure features a narrow core of $\sim 2 \sim 3$ deg, outside of which the luminosity falls off approximately as $\theta_c^{-4.7}$ and the SED peak photon energy as $\theta_{\text{th}}^{-2}$. The posterior on the transition angle $\theta_{\text{th}}\eta$ rails against the upper boundary of its physical range, and the slope $\beta_1$ and $\beta_2$, at larger viewing angles are not well constrained by the available data (see Table 3). This indicates that the data are consistent with an apparent structure described by a single power law outside the core: this may disfavour somewhat the presence of a distinct dissipation mechanism that dominates the gamma-ray emission at viewing angles in the range $20 \leq \theta_{\text{th}}/\text{deg} \leq 30$ (e.g. cocoon shock breakout as proposed by Gottlieb et al. 2018) in the majority of SGRBs, unless the apparent structure of the combined emission components follows the same power law. A caveat here is that there are most likely not more than a handful of events in our samples that are observed at such large viewing angles (see Sects. 4.2 and 4.4), and therefore this result must be taken with a grain of salt.

The structure is by construction consistent with the GRB 170817A luminosity and $E_p$ at the relevant viewing angle, as shown by the light green contours in the figure. These contours represent the 50% and 90% credible regions of ($\theta_c, L$) and ($\theta_{\text{th}}, E_p$) for GRB 170817A, constructed by combining the viewing angle information from the GW analysis conditioned on the host galaxy distance (Sect. 2.4.3) and our GRB 170817A spectral analysis at peak (Appendix B).
Table 3. Constraints on population model parameters from our analysis.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Full-sample</th>
<th>Flux-limited sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_{\ell}$/deg</td>
<td>$2.1^{+0.4}_{-0.4}$</td>
<td>$3.0^{+0.4}_{-0.4}$</td>
</tr>
<tr>
<td>$\theta_{\phi}$/deg</td>
<td>$1.4^{+0.4}_{-0.4}$</td>
<td>$2.1^{+0.4}_{-0.4}$</td>
</tr>
<tr>
<td>$L_\phi^*/10^{52}$ erg s$^{-1}$</td>
<td>$63.4^{+51.5}_{-51.5}$</td>
<td>$64.5^{+23.2}_{-23.2}$</td>
</tr>
<tr>
<td>$L_{\phi}/10^{49}$ erg s$^{-1}$</td>
<td>$0.5^{+0.7}_{-0.1}$</td>
<td>$0.5^{+0.7}_{-0.1}$</td>
</tr>
<tr>
<td>$\alpha_L$</td>
<td>$4.7^{+1.2}_{-1.4}$</td>
<td>$5.4^{+1.2}_{-1.4}$</td>
</tr>
<tr>
<td>$\beta_L$</td>
<td>$1.6^{+0.9}_{-0.9}$</td>
<td>$1.9^{+0.9}_{-0.9}$</td>
</tr>
<tr>
<td>$E_{pc}$/MeV</td>
<td>$17.7^{+15.3}_{-10.5}$</td>
<td>$4.5^{+15.3}_{-2.3}$</td>
</tr>
<tr>
<td>$\alpha_{E_p}$</td>
<td>$1.9^{+0.1}_{-0.9}$</td>
<td>$1.5^{+0.1}_{-0.9}$</td>
</tr>
<tr>
<td>$\beta_{E_p}$</td>
<td>$2.5^{+0.4}_{-0.4}$</td>
<td>$1.5^{+0.4}_{-0.4}$</td>
</tr>
<tr>
<td>$A$</td>
<td>$3.2^{+0.7}_{-0.4}$</td>
<td>$2.9^{+0.7}_{-0.4}$</td>
</tr>
<tr>
<td>$\sigma_{E_p}$/dex</td>
<td>$0.4^{+0.1}_{-0.1}$</td>
<td>$0.4^{+0.1}_{-0.1}$</td>
</tr>
<tr>
<td>$y$</td>
<td>$0.3^{+0.0}_{-0.0}$</td>
<td>$0.0^{+0.0}_{-0.0}$</td>
</tr>
<tr>
<td>$a$</td>
<td>$4.6^{+0.4}_{-0.4}$</td>
<td>$3.8^{+1.0}_{-1.0}$</td>
</tr>
<tr>
<td>$b$</td>
<td>$5.3^{+2.2}_{-3.8}$</td>
<td>$5.5^{+2.2}_{-3.8}$</td>
</tr>
<tr>
<td>$z_p$</td>
<td>$2.2^{+0.8}_{-0.6}$</td>
<td>$2.3^{+0.8}_{-0.6}$</td>
</tr>
<tr>
<td>$\log(L_\phi/\text{erg s}^{-1})$</td>
<td>$45.5^{+1.7}_{-2.0}$</td>
<td>$45.7^{+1.7}_{-2.0}$</td>
</tr>
<tr>
<td>$2/\alpha_L$</td>
<td>$0.4^{+0.2}_{-0.1}$</td>
<td>$0.4^{+0.2}_{-0.1}$</td>
</tr>
<tr>
<td>$2/\beta_L$</td>
<td>$0.4^{+0.1}_{-0.1}$</td>
<td>$0.4^{+0.1}_{-0.1}$</td>
</tr>
<tr>
<td>$\alpha_{E_p}/\alpha_L$</td>
<td>$0.4^{+0.1}_{-0.1}$</td>
<td>$0.3^{+0.1}_{-0.1}$</td>
</tr>
<tr>
<td>$\beta_{E_p}/\beta_L$</td>
<td>$0.1^{+0.2}_{-0.1}$</td>
<td>$0.1^{+0.2}_{-0.1}$</td>
</tr>
<tr>
<td>$R_0$/Gpc$^{-3}$ yr$^{-1}$</td>
<td>$740^{+130}_{-630}$</td>
<td>$180^{+30}_{-145}$</td>
</tr>
<tr>
<td>$\log(R_0/L &gt; 10^{50}$ erg s$^{-1})/\text{Gpc}^{-3}$ yr$^{-1}$</td>
<td>$3.6^{+0.9}_{-2.5}$</td>
<td>$1.3^{+0.9}_{-2.7}$</td>
</tr>
</tbody>
</table>

Notes. For each parameter and for each analysis setup (full sample; flux-limited sample), the table reports the median of the marginalised posterior probability density and the symmetric 90% credible interval. A horizontal line separates actual model parameters from derived ones. The latter are: the typical luminosity at the ‘break’ angle, $L_\phi = \bar{L}(\theta_\phi)$; the slope $2/\alpha_L$ of the luminosity function $\phi(L) = dP/d\ln L$ for jets seen between $\theta_\phi$ and $\theta_\psi$; the slope $2/\beta_L$ for jets seen at larger viewing angles; the slope $\alpha_{E_p}/\alpha_L$ of the Yonetoku correlation for off-axis jets with $\theta_\psi < \theta_\phi < \theta_\psi$; the slope of the Yonetoku correlation for $\theta_\phi > \theta_\psi$. The uncertainty band on the typical luminosity $L_\phi^*$ at $\theta_\psi < \theta_\phi$ is particularly narrow because of the chosen prior on $L_\phi^*$ (see the discussion in Sect. 4.5). The larger uncertainty at $\theta_\psi \sim \theta_\phi$ arises because of the combined uncertainties on $L_\phi^*$, $\theta_\phi$, and $\alpha_L$. The jet structure functions can also be projected onto the ($L_\phi$, $E_p$) plane: Fig. 7 shows the median ($\bar{L}_\phi$, $\bar{E}_p$) relation and the corresponding 90% credible region for the two analyses. In both cases, the bulk of the SGRBs with known redshift (shown by red crosses) are close to the upper-right end of the relation, which indicates that they are observed close to on-axis in the model (more precisely, close to the edge of the core; see also Sect. 4.2). Interestingly, the relation lies above most of the SGRBs in the known-redshift sub-sample, indicating that selection effects play a major role in how the plane is populated, according to the model. We expand on this in Sect. 4.3.

The inset in Fig. 7 shows the posterior probability density on the $y$ parameter that sets the correlation between the on-axis luminosity and peak SED photon energy. Despite our model allowing for such a correlation, the posterior is fully compatible with $y = 0$, that is, the absence of an intrinsic correlation between $L_\phi$ and $E_{pc}$.

3.2. Luminosity function

Our analysis does not directly constrain the local rate density, $R_0$, of SGRBs, because in our inference framework it is effectively only a normalisation factor as long as its prior is uniform in the logarithm (M19; Fishbach et al. 2018). In order to derive the local rate density implied by our results, we required the observed rate of events with $p > p_{lim}$ to be equal to $R_{\text{obs}}(p_{\text{lim}})$, the number of SGRBs with $p > p_{\text{lim}}$ in our sample, where $p_{\text{lim}} = 0.59$ is a factor that corrects for the accessible field of view and the duty cycle of GBM (Burns et al. 2016), and $\eta_{\text{GBM}} = 13\text{ yr}$ is the Fermi mission duration at the time of the last SGRB in the observer-frame sample. Given $\Lambda_{\text{pop}}$, the local rate density is then

$$R_0 = \frac{R_{\text{obs}}(p_{\text{lim}})}{P(L, E_p) | \Lambda_{\text{pop}}| \int P(L, E_p) | \Lambda_{\text{pop}}| dL dE_p dz}, \quad (42)$$

where $P(L, E_p)$ is the flux-limited form from Eq. (19) and $\rho(z)/R_0$ is independent of $R_0$ (see Eq. (9)). We applied the above expression to our population posterior samples $\Lambda_{\text{pop}}^{(i)}$ to obtain an equal number of $(R_0, \phi(L, \Lambda_{\text{pop}}^{(i)}))$ samples. In turn, this allowed us to construct samples of the posterior distribution of the luminosity function, $(R_0, \phi(L, \Lambda_{\text{pop}}^{(i)}))$, where $\phi(L, \Lambda_{\text{pop}}) = \int P(L, E_p) | \Lambda_{\text{pop}}| dE_p dz$. Figure 8 shows the luminosity function of SGRBs obtained in this way from the full-sample analysis (solid red) and the flux-limited sample analysis (dashed orange). The corresponding results from W15 and G16 (their fiducial model ‘a’) are

---

We are effectively neglecting here the Poisson error on the observed rate, which has a small impact on $R_0$ with respect to the uncertainty on $\Lambda_{\text{pop}}$. 

---

Fig. 6. Average apparent jet structure. The two larger panels show the constraints on the SGRB average apparent structure obtained from our analysis: solid lines show the medians of the projected posterior distributions of the jet structure functions at fixed $\theta$, from the full sample analysis (red: $L = L^*_p(\theta)$; blue: $E_p = E^*_p(\theta)$), while dashed lines refer to the flux-limited sample (orange: $L_c$; cyan: $E_p$). The shaded region around each line encompasses the corresponding 90% credible range at each fixed $\theta$. Dotted lines show reference power law trends. The light green contours show the 50% and 90% credible regions of $(L, \theta)$ (middle panel) and $(E_p, \theta)$ (bottom panel) for GRB 170817A (Sect. 2.4.3). The insets show the posterior distributions on the parameters $A$ and $\sigma_\nu$ that determine the dispersion around the average structure within the population. The top smaller panel shows the posterior distributions on the logarithms of the transition angles, $\ln(\theta_p)$ (grey) and $\ln(\theta_w)$ (green), with the results for the full sample shown as filled areas, and those for the flux-limited sample shown as dashed lines.

Fig. 7. Average apparent jet structure in the $L - E_p$ plane. The solid turquoise line connects the points $(\langle L(\theta) \rangle, \langle E_p(\theta) \rangle)$ at varying viewing angles, where $\langle \cdot \rangle$ represents the median, from the full sample analysis. The turquoise shaded area encompasses the corresponding 90% credible range. Grey lines show 100 posterior samples of the lines $(\langle L(\theta) \rangle, \langle E_p(\theta) \rangle)$ from the same analysis. The dashed green line and yellow area show the corresponding results for the flux-limited sample analysis. The crosses mark the positions of the Fermi/GBM SGRBs with known redshifts (all in red except for GRB 170817A, which is shown in blue). The inset shows the posterior probability density on the $y$ parameter that sets the slope of the correlation between the on-axis luminosity and the on-axis SED peak photon energy, with the results of the two analyses colour coded as in the main panel.

Fig. 8. Luminosity function. The median of the posterior probability density of $R_p(L) = dR_p/d\ln(L)$ (i.e., the local rate density per unit logarithm of the peak luminosity) from the full sample analysis is shown with a solid red line, while the dashed orange line refers to the flux-limited sample analysis. The shaded regions encompass the symmetric 90% credible interval at each fixed $L$. We show for comparison the corresponding results from W15 (blue) and G16 (grey – their fiducial model ‘a’). The credible region of W15 is computed assuming uncorrelated errors on their parameters.
The total local rate density, \( R_0 \), of SGRBs (including all viewing angles) is not well constrained by our analyses, because of the difficulty in determining the actual extent of the luminosity function with the available data. The full-sample analysis yields \( R_0 = 740^{+670}_{-330} \) Gpc\(^{-3}\) yr\(^{-1}\) (median and symmetric 90\% credible interval), while the flux-limited sample analysis gives \( R_0 = 180^{+145}_{-60} \) Gpc\(^{-3}\) yr\(^{-1}\) (see Fig. 9, left-hand panel). Both are compatible with the local BNS merger rate derived from GW observations (Abbott et al. 2023), \( R_{0,\text{BNS}} = 10 - 1700 \) Gpc\(^{-3}\) yr\(^{-1}\), and also with other recent estimates of the total, collimation-corrected SGRB rate based on different methods (e.g. Rouco Escorial et al. 2022, see Mandel & Broekgaarden 2022 for a review and further references). The fact that the derived SGRB rate leans towards the high-end of the BNS merger rate uncertainty interval can be interpreted as an indication that the fraction of BNS mergers that yield a jet must be high, in agreement with the results of Salafia et al. (2022), Sarin et al. (2022), Beniamini et al. (2019), and Ghirlanda et al. (2019).

In order to compare our local rate density with those presented in the literature, we also computed the rate density of events above a minimum luminosity \( L_{\text{min}} = 10^{50} \) erg s\(^{-1}\). The right-hand panel in Fig. 9 shows the result from our two analyses, which yield \( R_0(L > 10^{50} \) erg s\(^{-1}\)) = \( 3.6^{+6.9}_{-2.5} \) Gpc\(^{-3}\) yr\(^{-1}\) (full sample) and \( 1.3^{+3.4}_{-0.7} \) Gpc\(^{-3}\) yr\(^{-1}\) (flux-limited sample), compared with those obtained by integrating the W15 and G16 luminosity functions over the same luminosities. All results are in agreement with each other, placing the local rate density of luminous SGRBs around one event per Gpc\(^3\) yr, with roughly one order of magnitude of uncertainty.

Our model also constrains the evolution of the rate density with redshift, \( \rho(z) \), using the parametrisation given in Eq. (9). Figure 10 shows the symmetric 90\% credible interval of the posterior probability distribution of \( \rho(z) \), at each fixed \( z \), considering only events with \( L \geq 10^{50} \) erg s\(^{-1}\) (solid red: full-sample analysis; dashed orange: flux-limited sample analysis). Lowering the minimum luminosity would increase the uncertainty on the normalisation, but leave the shape identical, as a consequence of our assumption of no evolution of the jet structure parameters with redshift. For comparison, we show the corresponding results from W15 and G16, where the shaded regions only account for the uncertainty on the local rate density, and thus represent an underestimate of the actual uncertainty in these models. The shape of the constraint is in qualitative agreement with the result of G16, who find an evolution that is compatible with the expectations from compact binary merger progenitors but is in strong disagreement with the sharp cut-off in the rate density at \( z > 0.9 \) found by W15. We believe that this disagreement stems from a bias induced by the limited treatment of selection effects in that work, with a similar impact.
as that described in Bryant et al. (2021) for non-parametric methods.

The low-redshift scaling of the SGRB rate, \((1 + z)^a\) with \(a = 4.6^{+0.4}_{-0.8}\), is steeper than that of the CSFR at low redshift, CSFR \(\propto (1 + z)^{2.7}\) (Madau & Dickinson 2014), while the constraint on the peak of the SGRB rate \(z_p \sim 2.2^{+0.8}_{-0.6}\) indicates a preference for a rate that peaks at larger redshift than the CSFR (whose peak is at \(z \sim 1.9\)). This is difficult to explain with either very short delay times between star formation and BNS mergers (which would suggest that the SGRB rate should trace the CSFR) or long delay times, which would shift the peak of the SGRB to lower redshift than the CSFR peak (though the uncertainty on the SGRB redshift peak could allow for the long delay time interpretation). This apparent discrepancy could point to a redshift-dependent evolution of the yield of merging BNSs per unit star formation or a redshift-dependent evolution of the fraction of BNS mergers yielding SGRBs. Such effects could plausibly be caused by metallicity-dependent variations in stellar and binary evolution, including in NS masses. We note, however, that three out of four SGRBs with a photometric redshift in our rest-frame sample, namely 170127B, 180727A and 191031D, have median redshift larger than 1.9. If we remove the four SGRBs with a photometric redshift from our sample, we obtain generally similar results (with larger error bars) except for the redshift evolution, whose preferred low-redshift slope and peak become more consistent with the Madau & Dickinson (2014) CSFR (but clearly with larger error bars: \(a \sim 3 \pm 1.5, z_p \sim 1.9 \pm 1\)), which could reconcile the result with the expectations for BNS mergers with short delay times. Thus, the redshift distribution is sensitive to the reliability of these photometric redshifts.

4. Discussion

4.1. Jet intrinsic structures compatible with the derived apparent structure

4.1.1. Jet total luminosity and energy

Keeping in mind that \(L\) represents the luminosity at the peak of the light curve, the actual time-averaged gamma-ray luminosity can be written as \((L) = \varepsilon L_{\text{iso}}\), where we take the reference value \(\varepsilon = 0.3 \xi_{-0.5}\), which is the median of the peak flux to average flux ratios in the GBM sample. The prompt emission conversion efficiency is \(\varepsilon_f = 0.1 \varepsilon_{-1}\), where the reference value is based on the results of Beniamini et al. (2016). Therefore, the two factors compensate each other to some extent, and the jet core total (i.e. prior to dissipation that leads to gamma-ray emission) average isotropic-equivalent energy output rate is \((L_{\text{tot}}) = \left(\varepsilon \xi_{-0.5}\right) L_{\text{iso}} = 9 \times 10^{51} \varepsilon_{-0.5} L_{51.5} \nu_{-1} Edward\, \text{erg}\, s^{-1}\). If the jet duration is \(T = 1 T_{0.5}\), the core isotropic-equivalent jet energy is then, by definition, \(E_{\text{jet},51.5} \sim (L_{\text{jet}}) T = 9 \times 10^{51} \varepsilon_{-0.5} L_{51.5} T_{0.5}\) erg. This is compatible with typical estimates of the core-isotropic-equivalent energy of the GRB 170817A jet, which fall in the range \(10^{51.3} - 10^{52.0}\) erg (see e.g. Fig. S6 in Ghirlanda et al. 2019), and in line with cosmological SGRBs in general, which typically fall in a similar range (e.g. Rouco Escorial et al. 2022; Fong et al. 2015). The total jet energy is of the order of \(E_{\text{tot}} \sim \theta_{2}^2 E_{\text{jet},51.5} \sim 2 \times 10^{59} (\theta_{2}/3\, \text{deg})^2 \xi_{-0.5} L_{51.5} T_{0.5}\) erg. Also, this value is in line with those inferred from afterglow modelling of cosmological SGRBs (Rouco Escorial et al. 2022).

4.1.2. Angular structure

As discussed in Sect. 2.1, the relationship between the intrinsic jet structure and the apparent luminosity angular profile is not straightforward and dependent on the underlying dissipation and emission mechanism. Nevertheless, we can get some

Fig. 10. Rate density evolution. The figure shows the redshift evolution of the rate density from both our full sample (solid red lines) and flux-limited sample (dashed orange lines) analysis, limited to events above a minimum luminosity \(L_{\text{min}} = 10^{50}\) erg s\(^{-1}\). The corresponding evolutions from W15 and G16 are shown in blue and grey, respectively. The thick lines show the median of the posterior predictive distribution at each fixed redshift, while the shaded areas encompass the symmetric 90% credible interval.

3.4. Comparison with the three reference samples

Figure 11 compares the distributions of 64 ms peak photon flux \(p_{(50-300)}\) and observed peak photon energy \(E_{\text{obs}}\) distributions predicted by our model (using the median of the \(z_{\text{pop}}\) posterior distribution) with the observer-frame Fermi/GBM sample. For the flux-limited sample analysis, we limit the comparison to the sub-sample of events with \(p_{(50-300)} > P_{\text{lim},\text{GBM}}\). The figure demonstrates an excellent agreement both in the joint \(p_{(50-300)} - E_{\text{obs},\text{lim}}\) distribution and in the individual distributions. The fact that the shape of the low-end of the inverse cumulative \(\log(N_{\text{obs}}) - \log(p_{(50-300)})\) distribution is well reproduced (panel b in the figure) indirectly demonstrates the ability of our Fermi/GBM detection efficiency model to accurately reproduce the selection effects of the full sample.

In Fig. 12 we also compare the distribution of \(L, E_{p}\) and \(z\) of our rest-frame sample (Sect. 2.4.2), in red, with those predicted by the population model with the parameter constraints from the full sample analysis (blue) and flux-limited sample analysis (grey). In panels a.1, a.2 and b we show the 90% credible bands that stem from the \(z_{\text{pop}}\) uncertainty. In this case, we apply the detection efficiency model described in Appendix C to both the full sample and flux-limited sample analysis results in order to compute the predicted distributions. In panel a.0 we additionally show the only event in our viewing angle sample, GRB 170817A (orange cross), along with the 50%, 90%, 99% and 99.9% containment contours (shades of orange) of the distribution of SGRBs detected by Fermi/GBM and with a BNS merger counterpart detected by aLIGO and Advanced Virgo with the O3 sensitivity, as predicted by the population model with the best-fit parameters from the full sample analysis.
Fig. 11. Observer-frame constraints and best-fit model predictions. Panel a.0 shows the predicted distribution of Fermi/GBM SGRBs on the \((p_{[50-300]}, E_{\text{iso}})\) plane for our best-fit model, with progressively lighter contours containing 50%, 90%, 99% and 99% of the events. Red dots show the observed data reported in the Fermi/GBM catalogue that pass our additional quality cuts, while grey points show the events that are discarded. We additionally show the error bars with thin grey lines for those events with a relative error larger than 50% on either quantity, or both. Panels a.1 and a.2 show the predicted (solid blue line) and observed (solid red cumulative histogram, with the pink region showing the 90% confidence region that stems from statistical uncertainties on spectral fitting parameters) cumulative distributions of \(p_{[50-300]}\) (panel a.1) and \(E_{\text{iso}}\) (panel a.2) for events that pass the quality cuts. Panel b shows the inverse cumulative distribution of \(p_{[50-300]}\), highlighting the behaviour at the high-flux end, which follows the expected \(p_{[50-300]}\) trend. Panel b uses the same conventions as panels a.1 and a.2, except the shaded pink region shows the one-sigma equivalent Poisson error.

In the internal shocks scenario, for instance, a large efficiency \(\xi\) of the peak luminosity to the average luminosity does not depend on the viewing angle (i.e. the light curve shape is preserved when changing the viewing angle), and (ii) the observed duration of the emission is dominated by the central engine activity time, and hence it is also independent of the viewing angle. Under these assumptions, which likely hold only in a limited range of viewing angles close to the jet core, the peak luminosity scales with the viewing angle in the same way as the isotropic-equivalent energy (Eq. (1)).

Under these assumptions, which likely hold only in a limited range of viewing angles close to the jet core, the peak luminosity scales with the viewing angle in the same way as the isotropic-equivalent energy (Eq. (1)). We further assumed a power law profile for the jet total isotropic-equivalent energy, \(E_{\text{tot,iso}}\), and a bulk Lorentz factor, \(\Gamma\), that is, we set \(E_{\text{tot,iso}}(\theta) \propto (\theta/\theta_c)^{-\alpha_E}\) and \(\Gamma(\theta) \propto (\theta/\theta_c)^{-\alpha_\Gamma}\) for \(\theta > \theta_c\). It is likely that the gamma-ray efficiency \(\epsilon_\gamma\) is also a function of the angle from the jet axis\(^8\), so that we also set \(\epsilon_\gamma(\theta) \propto (\theta/\theta_c)^{-\alpha_\epsilon}\), with \(s_\epsilon > 0\). Therefore, the isotropic-equivalent energy radiated in gamma-rays at each angle goes as \(E_{\text{rad,iso}}(\theta) \propto (\theta/\theta_c)^{-\alpha_\epsilon-\epsilon}\).

In principle, using Eq. (1) one can derive the profile of the observed isotropic-equivalent energy \(E_{\gamma,\text{iso}}(\theta_c)\) (and hence peak luminosity \(L_c\), given our assumptions) as a function of the viewing angle given the slopes \(s_E\), \(s_\epsilon\), \(s_\Gamma\), the core bulk Lorentz factor \(\Gamma_c\) and the core half-opening angle \(\theta_c\). On the other hand, we can simplify the problem by noting that at relatively small viewing angles the emission is dominated by material along the line of sight, provided that \(\Gamma_c\) is relatively large, say \(\Gamma_c \gtrsim 100\). In this regime, \(E_{\gamma,\text{iso}}(\theta_c) \sim E_{\text{rad,iso}}(\theta = \theta_c)\) (Rossi et al. 2002); thus, \(\alpha_L \sim s_E + s_\epsilon\) and hence \(s_{\epsilon} \leq \alpha_L\). In the same regime, \(E_{\gamma,\text{iso}}(\theta_c) \sim \Gamma(\theta = \theta_c)E_{\gamma}(\theta = \theta_c)\). The \(E_{\gamma}(\theta)\) here is the comoving peak SED photon energy and is likely positively correlated with \(\Gamma\), and therefore \(s_E \leq \alpha_{\epsilon}\). Using the upper end of the 90% credible intervals for \(\alpha_L\) and \(\alpha_{\epsilon}\) from our full sample analysis, these arguments therefore lead to the upper limits \(s_E \leq 6\) and \(s_{\epsilon} \leq 3\). We stress again that, given the assumptions, these results only hold for viewing angles close to the jet core.

\(^8\) In the internal shocks scenario, for instance, a large efficiency requires a large average bulk Lorentz factor and high Lorentz factor contrast between colliding shells. Since the bulk Lorentz factor is expected to decrease away from the core, the efficiency should decrease as well.
While these limits clearly conflict with a ‘top-hat’ jet structure (which would correspond to $s_E, s_T \to \infty$), they are still in agreement with the rather steep kinetic energy profiles and shallower Lorentz factor profiles found in studies of the GRB 170817A afterglow (e.g. Hotokezaka et al. 2019; Ghirlanda et al. 2019; Mooley et al. 2022). The approximate $\theta^{-3}$ scaling of the jet isotropic-equivalent kinetic energy found in recent numerical simulations of SGRB jets (e.g. Gottlieb et al. 2020, 2021, 2022) is also compatible with these limits, even though it seems to conflict with the former findings based on the GRB 170817A afterglow.

### 4.2. Viewing angles of Fermi/GBM SGRBs with known redshifts

Through our population model it is possible to derive a viewing angle probability for any SGRB using only the information on its luminosity and spectral peak energy. Here we focus on SGRBs with a measured redshift in our rest-frame sample. The posterior probability on the viewing angle $\theta_{\nu,i}$ of the $i$-th SGRB in the sample (represented by data $d_i$ in the data vector $d$) is

$$P(\theta_{\nu,i} \mid d) = \sum_i \mathcal{P}(L_i, E_{p,i} \mid d) \mathcal{P}(\theta_{\nu,i} \mid \lambda_{p,i}) \mathcal{P}(\lambda_{p,i} \mid d) \frac{dL_i}{dE_{p,i}}$$

$$= \int P(L_i, E_{p,i} \mid d_i) \mathcal{P}(\theta_{\nu,i} \mid \lambda_{p,i}) \mathcal{P}(\lambda_{p,i} \mid d) \, d\lambda_{p,i} \, dL_i \, dE_{p,i}$$

$$= \int P(L_i, E_{p,i} \mid d_i) \mathcal{P}(\theta_{\nu,i} \mid \lambda_{p,i}) \mathcal{P}(\lambda_{p,i} \mid d) \mathcal{P}(\lambda_{p,i} \mid \lambda_{p,\text{pop}})$$

$$\times \mathcal{P}(\lambda_{p,\text{pop}} \mid d) \, d\lambda_{p,\text{pop}} \, dL_i \, dE_{p,i}$$

$$= \frac{\sin \theta_{\nu,i}}{N_i N_p} \sum_j \sum_k \frac{\mathcal{P}(L_j, E_{p,j} \mid \theta_{\nu,i}, \lambda_{p,j}) \mathcal{P}(\lambda_{p,j} \mid \lambda_{p,\text{pop}})}{\mathcal{P}(L_j, E_{p,j} \mid \lambda_{p,\text{pop}})}.$$

(43)
where the last equality follows from Monte Carlo approximation of the integrals, with \( \{(L_{i,j}, \theta_{i,j})\}_{i,j=1}^{N_i} \) being samples from the \( P(L_i, E_p | d_i) \) posterior obtained from the spectral analysis of the SGRB, and \( \{\lambda_{\text{pop},k}^i\}_{i=1}^{N_i} \) being samples from the population posterior.

Figure 13 shows the resulting population-informed viewing angle posterior probability densities for the SGRBs in our rest-frame sample. The constraints from the full sample and flux-limited sample analyses are in general agreement, with most jets likely viewed a few degrees from the jet axis. Focussing on the full sample analysis results, four SGRBs have population-informed viewing angles that are larger than 2 deg at 95% credibility: GRB 080905A, GRB 131004A, GRB 160821B, and, unsurprisingly, GRB 170817A. The median and 90% credible interval of our population-informed viewing angle estimate for GRB 160821B is \( \theta_v = 5.2 \pm 2 \) deg, which is compatible with the estimate \( \theta_v = 10^{+4}_{-1} \) deg by Troja et al. (2019) based on afterglow modelling. The population-informed estimate for GRB 170817A is \( \theta_v = 23.5 \pm 3 \) deg, in excellent agreement with afterglow-based estimates that include the information on the centroid proper motion from Very Long Baseline Interferometry imaging (e.g. Mooley et al. 2018, 2022; Hotokezaka et al. 2019; Ghirlanda et al. 2019; Fernández et al. 2022; Govoreanu-Segal & Nakar 2023), which find viewing angles in the range \( 15 \leq \theta_v/\text{deg} \leq 25 \), and also with the estimate \( \theta_v = 18 \pm 8 \) deg by Mandel (2018, 68% credible interval) based on a similar method as that employed in Sect. 2.4.3, but which includes a marginalisation over the cosmological parameters.

### 4.3. The impact of selection effects on the \( L–E_p \) plane

Our result, shown in Fig. 7, that the bulk of the SGRB population is located upwards of the apparent SGRB Yonetoku correlation came with a surprise to us. In order to demonstrate the role played by selection effects in this context, we show with grey shades in Fig. 14 the detection efficiency that represents the selection effects acting on our rest-frame sample (Sect. 2.4.2), averaged over redshift, that is,

\[
\langle P_{\text{det}} \rangle (L, E_p) = \int P_{\text{det,GBM}}(L, E_p, \lambda) P_{\text{det,BAT}}(L, E_p, \lambda) P(\lambda | \lambda_{\text{pop}}) \, d\lambda,
\]

where \( P_{\text{det,GBM}}(L, E_p, \lambda) \) is that from the full sample analysis (but the discussion remains unchanged when adopting that from the flux-limited sample analysis) and in this discussion we keep \( \lambda_{\text{pop}} \) fixed at the median of the full sample analysis posterior. The purple contours in the figure contain 50%, 90%, 99%, 99.9%, 99.99% and 99.999% of SGRBs in the Universe according to our model. For reference we also show the \( \langle L(\theta_v), E_p(\theta_v) \rangle \) relation with a thick, purple dashed line. This ‘intrinsic’ distribution of SGRBs in the \( L–E_p \) plane is distorted by selection effects into the distribution represented by cyan contours, which contain 50%, 90%, 99%, 99.9%, 99.99% and 99.999% of SGRBs that pass the rest-frame sample selection criteria according to the model. The distribution represented by cyan contours can be understood as the product of the intrinsic (purple) distribution times the redshift averaged detection efficiency (grey filled contours). As in previous figures, red crosses mark the positions of SGRBs with measured \( \lambda \) in our rest-frame sample, while the blue cross marks GRB 170817A.

Focussing on SGRBs with \( L > 10^{50} \text{erg s}^{-1} \), these reside in a part of the plane where the \( \langle P_{\text{det}} \rangle \) contours are roughly parallel and equally spaced. In other words, the gradient

\[
\nabla_{\ln L, \ln E_p} \ln \langle P_{\text{det}} \rangle (L, E_p) = \left( \frac{\partial \ln \langle P_{\text{det}} \rangle}{\partial \ln L}, \frac{\partial \ln \langle P_{\text{det}} \rangle}{\partial \ln E_p} \right)
\]

is roughly constant across the region occupied by the observed SGRBs in the plane, with a GBM detection efficiency that decreases by almost four orders of magnitude along the direction of this gradient. The variation in the density of the SGRBs in the rest-frame sample along the same direction, on the other hand, is clearly much less than four orders of magnitude. This suggests that the intrinsic density of SGRBs in this plane must
increase steeply in the direction opposite to the $\langle P_{\text{det}} \rangle$ gradient, in order for the variation in the intrinsic density of SGRBs to compensate the dramatic decrease in the detection efficiency. Hence, selection effects likely play a major role in shaping the observed $L-E_p$ correlation in SGRBs. This conclusion and also, intriguingly, the slope and dispersion of the intrinsic $L-E_p$ correlation we obtained, are in agreement with those found by Palmero & Daigne (2021) for LGRBs (see Sect. 5.1 of that paper). This might be taken as an indication of a common universal luminosity and $E_p$ angular profile in the two populations.

On a different note, it is worth stressing that the existence of a tail of SGRBs with low luminosities $L \lesssim 10^{50}$ erg s$^{-1}$ but very high SED peak photon energies $E_p \gtrsim 3$ MeV, predicted by our population model and visible in the figure, must be taken with a grain of salt. Our parametrisation is constructed in such a way that the dispersion of $E_p$ around the viewing-angle-dependent average $\bar{E}_p(\theta_v)$ is symmetric and identical at all viewing angles, so that the low-$L$, high-$E_p$ tail is merely a result of the choice of parametrisation, being unobservable with the current instrumentation and therefore not observationally constrained.

4.4. Joint SGRB + GW detection predictions

We can produce predictions for the rate of coincident detections of SGRBs and GWs by Fermi/GBM and the ground-based GW detector network from the population model, leveraging the fact that it accounts for the jet luminosity as a function of the viewing angle. We focus here on a network consisting of the two aLIGO detectors (Hanford and Livingston) plus the Virgo detector (an ‘HLV’ network), and assume the projected sensitivities in the upcoming O4 observing run$^9$. We assumed all SGRBs to be produced by BNS mergers with non-spinning components and the jets to be aligned with the orbital angular momentum. In practice, we constructed an HLV O4 GW detection efficiency $P_{\text{det, GW}}(z, \theta_v)$ as a function of redshift and viewing angle as follows: we binned the simulated BNS mergers from Colombo et al. (2022; selecting only those that produce a jet according to their criteria, which are $\sim 50\%$ in their population) in the $(\theta_v, z)$ plane and we computed the fraction $f_{\text{GW, i}}$ with a network S/N net $\geq 12$ in each bin (we assumed 100 duty cycle for all detectors for simplicity). We then estimated $P_{\text{det, GW}}(z, \theta_v)$ by linearly interpolating the $f_{\text{GW, i}}$’s on a grid with nodes corresponding to the centres of the bins. Samples of the cumulative joint detection rate were then computed based on 100 random population posterior samples as

$$N_{\text{GBM+GW}}(<z, \lambda_{\text{pop}}) = \int_0^z \int_0^{\pi/2} P(L, E_p | \theta_v, \lambda_{\text{pop}}) P_{\text{det, GBM}}(L, E_p, \theta_v) \sin \theta_v d\theta_v dL dE_p dz'$$

and similarly

$$N_{\text{GBM+GW}}(> \theta_v, \lambda_{\text{pop}}) = \int_{\theta_v}^{\pi/2} \int_0^{\pi/2} P(L, E_p | \theta_v, \lambda_{\text{pop}}) P_{\text{det, GBM}}(L, E_p, \theta_v) \sin \theta_v d\theta_v dL dE_p dz'$$

Figure 15 shows the median and symmetric 90% credible range over the population posterior samples of $N_{\text{GBM+GW}}(< z)$ (left-hand panels) and $N_{\text{GBM+GW}}(> \theta_v)$ (right-hand panels) using posterior samples from the full sample analysis (top panels) and flux-limited sample analysis (bottom panels). The plots also show the same result for the Fermi/GBM detection only (i.e. setting $P_{\text{det, GW}} = 1$) for comparison. All rates are normalised to the Fermi/GBM observed SGRB detection rate of 40 yr$^{-1}$ and hence include the limited field of view and duty cycle of the instrument. The total observed SGRB + GW rates are $N_{\text{GBM+GW}} = 0.53^{+0.75}_{-0.34}$ yr$^{-1}$ for the full sample analysis and

---

$^9$ https://dcc.ligo.org/T2000043-v3/public. We conservatively did not consider KAGRA (Somiya 2012) due to its much lower expected sensitivity.
To investigate the effect of relaxing that assumption, we estimated by Colombo et al. (2022), that the spectral parameters (photon index, observed peak photon flux from the spectral analysis) in the Fermi/GBM alone (blue: median; cyan: 90% credible range) or in coincidence with a network consisting of the two LIGO and the Virgo detectors with the projected O4 sensitivity (red: median; orange: 90% credible range). The right-hand panels show the corresponding rates $N_{\text{obs}}(> \theta_i)$ for events seen at a viewing angle larger than $\theta_i$.

### 4.5. Effect of decreasing the lower bound of the prior on $L_\ast^p$

As stated in Sect. 2.6, we chose the lower bound of the prior on the typical on-axis luminosity $L_\ast^p$ to be consistent with our assumption that the high end of the luminosity function is shaped by on-axis events, while intermediate and low luminosities are due to off-axis events and depend on the apparent structure. In order to investigate the effect of relaxing that assumption, we ran the full sample analysis one additional time with a much looser bound $L_\ast^p \geq 10^{48} \text{erg s}^{-1}$. As expected (given the fact that the marginalised posterior on $L_\ast^p$ relies against the lower bound in the results described above), this results in a posterior that shows a mild preference for $L_\ast^p \sim 10^{50} \text{erg s}^{-1}$, but with significant posterior support all the way down to $10^{48} \text{erg s}^{-1}$ and up to $10^{52} \text{erg s}^{-1}$, showing that the typical on-axis luminosity is poorly constrained by the available data within the proposed model. On the other hand, the lower $L_\ast^p$, the larger the local rate $R_0$ needed to reproduce the GBM observed SGRB rate: if we additionally require $R_0$ $\leq 1700 \text{Gpc}^{-3} \text{yr}^{-1}$, to reflect the upper limit on the BNS merger rate from the GW population analysis (Abbott et al. 2023), then the posterior becomes again consistent with that obtained with the original prior. We thus conclude that our prior, in addition to being a consequence of the assumed quasi-universal jet scenario, has a similar impact as the requirement that the SGRB local rate does not exceed the BNS merger rate. This suggests that future improved constraints on this rate through GW observations will positively impact our ability to constrain the SGRB population properties, including their typical apparent jet structure.

### 4.6. Difficulties in defining a ‘clean’ sample of GRBs from compact binary mergers

When selecting our sample, we focused on events with $T_{90}$ shorter than 2 seconds for simplicity, and for the practical reason that the spectral parameters (photon index, observed $E_p, \text{obs}$, peak photon flux from the spectral analysis) in the Fermi/GBM online catalogue are given with 64 ms binning only for events with $T_{90} < 2 \text{s}$. On the other hand, this selection may result in a sample that includes events from both the main progenitor classes, that is, compact binary mergers and collapsars (e.g.
Zhang et al. 2009; Bromberg et al. 2013), with the contamination from the latter class being difficult to quantify. Even if we did not test explicitly for the dependence of our results on this choice, the fact that the high-end of the luminosity function we obtain agrees relatively well with that obtained by W15 (who adopted a much more restrictive criterion in an attempt to select a sample of pure ‘non-collapsar’ GRBs) lends support to the conclusion that any potential contamination from collapsars in our sample does not impact the results significantly, given the present uncertainties.

Conversely, some GRBs with $T_{90} > 2$ s are now widely accepted as being the result of a compact binary merger rather than a collapsar (e.g. GRB211211A; Rastinejad et al. 2022; Mei et al. 2022; Gompertz et al. 2023). Hence, the definition of a clean sample of GRBs from compact binary mergers is not straightforward. We believe that the best approach to this kind of problem in the future will be that of modelling the entire GRB population as a mixture of the two classes, jointly fitting the two sub-populations in a hierarchical model, similarly to what is currently done for binary black hole merger GW population analyses (e.g. Bouffanais et al. 2019; Wong et al. 2021; Zevin et al. 2021).

As a final note, we caution that the duration of an SGRB must eventually increase with the viewing angle: at large enough viewing angles, the observed duration of a single pulse can exceed the duration of the central engine activity (e.g. Salafia et al. 2016). Therefore, our sample selection could contain a bias against events with a large viewing angle. In order to address this problem, the duration of the emission, and its dependence on the viewing angle, must be included in the model as well.

4.7. Peak luminosity versus average luminosity

Population studies of GRBs, whose luminosity varies erratically without a clear-cut minimum variability timescale during the prompt emission, must confront the issue of defining the relevant luminosity whose distribution is to be modelled. Given the stochastic nature of the light curves, the time-averaged luminosity is arguably the most relevant quantity from a physical point of view; on the other hand, the detectability (and hence the selection effects) depends more closely on the peak luminosity, so that from a practical point of view this is the most important quantity to be modelled, which has therefore become the standard in the field. To our knowledge, the relation between the two distributions, that is, the peak luminosity function $\phi(L) = dP/d \ln L$ and the average luminosity function $\Phi(\langle L \rangle) = dP/d \ln \langle L \rangle$, has never been investigated. Quite clearly, since $L = \xi^{-1} \langle L \rangle$ and the factor $\xi^{-1} \geq 1$ can differ between two SGRBs with the same $\langle L \rangle$ due to the stochasticity in the light curve, then the $L$ distribution is necessarily broader than the $\langle L \rangle$ one. In that sense, the ‘core’ luminosity dispersion in our model, parametrised as in Eq. (2), accounts at least partially for the dispersion due to these effects. At large viewing angles, on the other hand, the broadening of individual pulses is likely going to smooth out the light curve (Salafia et al. 2016), hence reducing this kind of scatter. Hence, another improvement over our approach could be that of including a viewing-angle-dependent scatter in $L$ (and similarly in $E_p$), which may improve the recovery of the actual average luminosity and peak photon energy profiles $\xi$ and $\eta$.

5. Summary and conclusions

The observations of GW170817 and the GRB170817A afterglow provided clear support for the presence of a jet viewed off-axis and endowed with a non-trivial angular structure. The inferred properties of the jet’s core were found to be consistent with those typically derived from the afterglows of SGRBs. This prompted the question of whether jets underlying SGRBs could be very similar to each other on average, with a large part of the diversity due to the geometric choice of a viewing angle rather than intrinsic variations in jet structure and energetics. In this work we have shown that a good description of the observed SGRB population can be obtained within such a scenario. The implied typical jet properties are consistent with those inferred from the GRB 170817A afterglow and from the larger population of SGRBs with a known distance, adding stability to the foundations of a unification programme for SGRBs under the quasi-universal jet scenario, and more generally to our physical understanding of these phenomena.

The inferred jet feature structures a $\sim 2$ deg uniform core within which the observer sees a large typical SED peak photon energy ($E_p \sim 5$ MeV) and luminosity ($L \sim 5 \times 10^{51}$ erg s$^{-1}$). Outside the core, the luminosity falls off with the viewing angle as a steep power law (slope $\alpha_L \sim -4.7$), while $E_p$ decreases as a relatively shallow power law ($\alpha_{E_p} \sim 2$). No evidence for a break in these power laws has been found with the present data and analysis approach. While we find no clear support for a correlation between the on-axis luminosity, $L_o$, and the on-axis peak SED photon energy, $E_{p,\text{on}}$, the combined viewing angle dependence of $L$ and $E_p$ induces a correlation for events viewed outside the core, $L \propto E_{p,\text{on}}^{0.4}$. In the observed sample, we find that this correlation is distorted by selection effects.

The inferred local rate density of SGRBs (at all viewing angles) is compatible with that of BNS mergers as inferred from GW population studies, suggesting they are the dominant progenitors. The model shows a preference for a strong rate density evolution with redshift: the rate density steeply increases as $\dot{\rho}(z) \propto (1 + z)^{16}$ at low redshifts, plateaus towards a maximum near $z \sim 2.2$, and declines at higher redshifts, where it is poorly constrained. These results, on the other hand, may be driven by the rather large redshifts of three out of four SGRBs with a photometric redshift in our rest-frame sample: if photometric redshifts are excluded from the analysis, the redshift evolution becomes consistent with that of the CSFR, and hence with progenitor binaries that merge rapidly after formation. Based on the model and on the projected sensitivity of the aLIGO and Advanced Virgo network, we predict around 0.2 to 1.3 joint SGRB and GW detections per year during the O4 observing run.

Through the population model, it is possible to derive a population-informed viewing angle estimate for every SGRB whose intrinsic luminosity and peak photon energy are reasonably constrained. The estimates obtained for SGRBs with a known redshift in our sample indicate that most of them are viewed close to the edge of the core (either just within the core or slightly outside it), with a few exceptions with a somewhat larger viewing angle. The largest viewing angle is clearly that of GRB 170817A, for which we estimate $\theta_v = 23^{+3}_{-3}$ deg, in excellent agreement with the estimates based on the afterglow and the superluminal motion seen in Very Long Baseline Interferometry observations.

A unification of SGRBs under a quasi-universal jet scenario would call for a relatively narrow progenitor parameter space, which can eventually help in pinpointing the long-debated jet-launching mechanism and the nature of the central engine. As demonstrated by the amount of information contained in the single GW170817 event, future multi-messenger observations of BNS mergers and their jets will be of the utmost importance to the success of this programme.
Appendix A: Additional details on the results

A.1. Corner plot

Figure A.1 shows the full, 14-parameter corner plot of both the full sample analysis (magenta) and the flux-limited sample analysis (light blue).

Fig. A.1. Corner plot of the posterior probability densities from the two analyses. The full sample analysis is shown in magenta, while the flux-limited sample analysis is shown in blue. The histograms on the diagonal represent the marginalised posterior probability densities constructed from the posterior samples, with the solid vertical lines marking the medians and the vertical dashed lines delimiting the symmetric 90% credible interval (i.e. the 5th and the 95th percentiles – note that these are not shown if they differ by less than one bin size from the nearest edge of the allowed range). The contours in the remaining panels show the one, two, three and four sigma credible areas from the two-dimensional marginalised joint posterior probability densities, with the dots showing the intersections of the medians of the corresponding one-dimensional marginalised posterior probability densities.
Appendix B: Spectral analysis of Fermi/GBM SGRBs with known redshifts

For each Fermi/GBM SGRB with a known redshift in our sample, we analysed the spectrum at the peak flux of the light curve binned with a 64 ms timescale. As starting time of the interval selected for the spectral analysis, we used the one reported in the GBM Catalogue (referred to as ‘Flux_64_Time’). The spectral data files and corresponding latest response matrix files (rsp2) were obtained from the online High Energy Astrophysics Science Archive Research Center (HEASARC) archive. We used the public software GTBURST to extract the spectral data. As part of the standard procedure, we selected the spectral data of the two most illuminated NaI detectors with a viewing angle smaller than 60° and the most illuminated BGO detector. In particular, we considered the energy channels in the range 10–900 keV for the NaI detectors, and 0.3–40 MeV for the BGO detector. We used intercalibration factors among the detectors, scaled to the most illuminated NaI and free to vary within 30%. To model the background, we manually selected time intervals before and after the burst and modelled them with a polynomial function whose best-fitting order is automatically found by GTBURST.

The spectral analysis has been performed with the public software XSPEC (v. 12.12.1). In the fitting procedure, we used the PG-Statistic, valid for Poisson data with a Gaussian background. Each peak flux spectrum was analysed with a cutoff power law model, typically used for the analysis of SGRBs spectra, and it consists of three parameters: the low-energy photon index \( \alpha \), the characteristic energy \( E_{\text{cut}} \) (from which the peak energy of the spectrum can be derived as \( E_p = E_{\text{cut}}(2 - \alpha) \)) and the normalisation. To obtain the luminosity (and its uncertainties) directly from the fit, the cutoff power law model is multiplied by the CLUMIN function available in XSPEC\(^{11}\), which computes the luminosity of a specific model component, provided the characteristic energy \( E_{\text{cut}} \) of the source-frame energy band over which luminosity is calculated is 1 keV-10 MeV. Since the CLUMIN model is used, the normalisation of the cutoff power law model has been kept fixed to 1 for all the spectra analysed. The parameters left free to vary in each fit are the following: the luminosity in the 1 keV-10 MeV energy range, the low-energy photon index \( \alpha \) and the characteristic energy \( E_{\text{cut}} \), alongside the two inter-calibration factors for the GBM detectors. Best-fit values and confidence ranges on these parameters have been derived within XSPEC, through the built-in Markov chain Monte Carlo algorithm (using the CHAIN command). The 50% and 90% contours for \( L_{\text{peak},\text{iso}} \) and \( E_p \) derived from the spectral analysis of each SGRB analysed in this work are reported in Fig. 4.

Appendix C: Fermi/GBM short gamma-ray burst detection efficiency

In order to construct a reliable estimate of the SGRB detection efficiency of Fermi/GBM, a detailed simulation of its response to an event of that class is needed. The onboard trigger algorithms of GBM, described for example in von Kienlin et al. 2020 (vK20 hereafter), monitor the counts (binned over a given timescale \( \Delta t \)) recorded in a subset of the 8 energy channels of the NaI detectors\(^{12}\). Each algorithm looks for an excess in the background-subtracted, binned counts \( C_a(t) - B(t) \) (where \( B(t) \) is an estimate of the counts due to the background), over a certain multiple \( n_{\text{th}} \) of the standard deviation \( \sigma_{\text{bkg}}(t) \) of the counts recorded in a time interval immediately preceding \( t \). Multiple algorithms operating on the same channels and with the same binning timescale are run in parallel, differing from each other only by small time offsets, which improves the triggering efficiency by minimising cases with sub-optimally placed bins. A burst trigger is initiated whenever an excess is recorded at a consistent time in at least two of the NaI detectors by any of the running algorithms.

Since a sufficient condition for a trigger is that the peak counts exceed the threshold, it is enough to simulate the peak counts generated by a putative source. On the other hand, since different algorithms operate with different binning timescales \( \Delta t \) and since SGRBs are highly variable, it is necessary that the peak counts generated by the same source, but measured with a different \( \Delta t \), reflect the actual peak count ratios induced by the temporal behaviour of the events of that class. The simulated counts must then be compared with realistic background counts, and this must be done in each of the NaI detectors and consistently for each of the triggering algorithms. In what follows, we describe our approach to these problems.

C.1. NaI background

In order to simulate a realistic background in each of the channels of the NaI detectors, we extracted the observed count rates from publicly available GBM daily data\(^{13}\), at a randomly sampled time, and multiplied them by \( \Delta t \) to get the counts. In particular, we considered ctime data relative to two selected days without triggers, 210530 and 220928, and extracted the count rates of the 12 NaI detectors in the eight energy channels at random times, excluding periods when the detectors were turned off. The resulting probability density distributions of count rates in the 50–300 keV band (channels 3-4, combined from all 12 detectors) is shown in Fig. C.1 (red histogram), along with that for the 25–50 keV band (channel 2), for comparison.

C.2. Detector response to an SGRB

We assumed the photon spectrum (number of SGRB photons per unit time, per unit area, per unit photon energy as measured at the Fermi/GBM position) of every SGRB to be time-independent and well described by the same cut-off power law as used in the population modelling, namely

\[
\frac{d^3N}{dr dA dE} \propto E^\alpha \exp\left(-\frac{(2 + \alpha)}{E_{p,\text{obs}}} \right).
\]  

which corresponds to the COMP model commonly used in Fermi/GBM data analyses (vK20). We use \( f(E) \) to indicate the above spectrum normalised to unity in the 50 – 300 keV band, that is,

\[
f(E, \alpha, E_{p,\text{obs}}) = \frac{\int_{50\text{keV}}^{300\text{keV}} \frac{d^3N}{dr dA dE} \, dE}{\int_{50\text{keV}}^{300\text{keV}} (d^3N/dr dA dE) \, dE}.
\]

\(^{10}\) https://heasarc.gsfc.nasa.gov/W3Browse/fermi/fermigrbst.html

\(^{11}\) For SGRBs with a photometric redshift, we used the CFLUX multiplicative model instead.

\(^{12}\) Algorithms that analyse the BGO detector data stream exist as well, but most of them require a trigger in the NaI detectors to be activated, or they have very restrictive triggering conditions, so they are effectively redundant and we neglect them here (vK20).

\(^{13}\) https://heasarc.gsfc.nasa.gov/W3Browse/fermi/fermigdays.html
Table C.1. Trigger algorithms considered in our framework1.

<table>
<thead>
<tr>
<th>Alg. Channels Band [keV]</th>
<th>∆t [s]</th>
<th>nσ</th>
<th>GBM algms.*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3-4</td>
<td>50-300</td>
<td>0.064</td>
</tr>
<tr>
<td>2</td>
<td>3-4</td>
<td>50-300</td>
<td>0.128</td>
</tr>
<tr>
<td>3</td>
<td>3-4</td>
<td>50-300</td>
<td>0.256</td>
</tr>
<tr>
<td>4</td>
<td>3-4</td>
<td>50-300</td>
<td>0.512</td>
</tr>
<tr>
<td>5</td>
<td>3-4</td>
<td>50-300</td>
<td>1.024</td>
</tr>
<tr>
<td>6</td>
<td>3-4</td>
<td>50-300</td>
<td>2.048</td>
</tr>
</tbody>
</table>

Notes. 1 Each row lists the identification number, monitored channels, binning timescale, ratio of threshold to background standard deviation and corresponding Fermi/GBM algorithm numbers for one of the triggering algorithms implemented in our framework. *Identification numbers of the Fermi/GBM algorithms (vK20) that are represented by each single algorithm in our framework.

Table C.2. Best-fit parameter values for the GBM NaI incidence-angle-dependent effective area ansatz function.

<table>
<thead>
<tr>
<th>Param.</th>
<th>32 keV</th>
<th>279 keV</th>
<th>663 keV</th>
</tr>
</thead>
<tbody>
<tr>
<td>c₁</td>
<td>1.19</td>
<td>0.503</td>
<td>0.280</td>
</tr>
<tr>
<td>c₂</td>
<td>-</td>
<td>0.427</td>
<td>0.366</td>
</tr>
<tr>
<td>c₃</td>
<td>0.00938</td>
<td>0.397</td>
<td>0.518</td>
</tr>
</tbody>
</table>

*The data were obtained from https://heasarc.gsfc.nasa.gov/W3Browse/fermi/fermiEvent.html, where 527 tridata files for bursts with $T_{90} < 2$ s were available. Some of the events had apparently incorrect background estimates, which were much larger than the actual counts recorded, resulting in negative background-subtracted counts at peak. Others had corrupted tridata files. After removing these problematic events, we were left with 449 valid SGRBs.
obtained in this way. We then computed trigdata at the three reference energies (coloured circles). The dashed lines show our assumed behaviour of $c$ over the channels monitored by the algorithm, $\bar{R}$ each simulated SGRB, we randomly picked one of the $t$ samples from a Poisson distribution with mean $P$ as explained in Appendix C.1, and represents a randomly picked matrix of NaI background counts computed the background-subtracted counts as $C$ (which is a good approximation given the background count distributions shown in Fig. C.1) and considered a trigger in the $k$-th detector if $(C_{t,\Delta}\text{ or } B_{t,\Delta}) \geq n_t \sigma_{\text{bkg,}\Delta,t}$. The simulated SGRB was marked as detected whenever at least two NaI detectors had at least one algorithm that triggered.

**C.3. Detection algorithms**

In our framework, each detection algorithm is characterised by a range of monitored channels, a binning timescale and a factor $n_t$ that sets the threshold for triggering. When compared to the actual triggering algorithms running on board Fermi/GBM (vK20), we do not consider a time offset, as our definition of $R_{\text{pflx}}(\Delta t)$ essentially corresponds to always taking the offset that maximises the peak count rates for each timescale. Hence, each of our algorithms effectively covers all the GBM algorithms with the same channels and binning timescales, but different offsets. Table C.1 reports the 6 algorithms that we implemented in our framework. These are limited to the 50-300 keV band because, by inspection of the trigdata files, we verified that essentially all the SGRBs detected by GBM triggered one of the algorithms operating in that band. We implemented all algorithms that operate on timescales $\Delta t \geq 64$ ms, because trigdata files do not provide count rates with a finer resolution. We do not consider this as a limiting factor, since the variability of the SGRB light curves over such short timescales is hardly important to our purposes.

**C.4. NaI detector effective area**

We decomposed the effective area of a NaI detector into its value for zero incidence angle $A_{\text{eff}}(0, E)$, times a dimensionless function that captures its variation with the incidence angle, $A_{\text{eff}}(\Theta, E) \equiv A_{\text{eff}}(0, E)/A(\Theta, E)$. We computed $A_{\text{eff}}(0, E)$ by linearly interpolating the measurements reported Fig. 11 of M09 on a log-log plane, as shown in panel (a) of Fig. C.4. To model $A(\Theta, E)$ we considered the measurements reported in Fig. 12 of

### Fig. C.4. NaI detector effective area model. Panel (a) shows the NaI detector effective area measurements (red crosses), for zero incidence angle, reported in M09. The pink line shows our adopted interpolation. Panel (b) shows the measured effective area (M09) for different photon incidence angles at three reference photon energies (black circles: 32 keV; blue squares: 279 keV; red triangles: 662 keV). The solid lines show our best-fitting model. Panel (c) shows the best-fit values of parameters $c_1$, $c_2$ and $c_3$ of our model (Eq. C.5) for the angular dependence of the effective area at the three reference energies (coloured circles). The dashed lines show our assumed behaviour of $c_i$ as a function of the photon energy.
Fig. C.6. *Fermi*/GBM detection efficiency for SGRBs. Panels (a) show the detection efficiency $\eta_{\text{det,3D}}$ for a fixed value of $E_{\text{p,obs}}$ (reported on top of each panel) and different photon indices $\alpha$ (values shown in the legend), as a function of the 64 ms peak photon flux $p_{\text{50–3000}}$. Panels (b) are obtained by averaging $\eta_{\text{det,3D}}$ over the observed distribution of $E_{\text{p,obs}}$ (panel b.1), that of $\alpha$ (panel b.2), or both (panel b.3). In panels b.1 and b.2, the solid lines are contours of constant averaged $\eta_{\text{det,3D}}$, with the values reported along each line.

M09 (shown in panel (b) of Fig. C.4), relative to the three reference photon energies $E_{\text{ref,1}} = 32$ keV, $E_{\text{ref,2}} = 279$ keV and $E_{\text{ref,3}} = 662$ keV, and normalised these data to the peak (which corresponds to $\Theta = 0$ in each case). We assumed an ansatz analytical form,

$$\hat{A}(\Theta, c_1, c_2, c_3) = \begin{cases} |\cos(\Theta)|^{c_1} & \cos(\Theta) \geq 0 \\ c_3 \cos(\Theta)^{c_2} & \cos(\Theta) < 0 \end{cases}$$

(C.5)

and fit it to the data at each reference photon energy by minimising the sum squares of the residuals, obtaining the values reported in Table C.2 (the best-fit models are shown by solid lines in panel (b) of Fig. C.4 – we note that $c_2$ remains unconstrained at $E_{\text{ref,1}}$, because $c_3 \sim 0$). In order to extend the model to other photon energies, we then assumed a linear dependence of the parameters $c_1$, $c_2$, and $c_3$ on the natural logarithm of the photon energy normalised to 32 keV, $\epsilon = \ln(E/32 \text{ keV})$, namely $c_1 = q_1 + m_1 \epsilon$, and obtained a good description of the data with $q_1 = 1.19$, $m_1 = -0.308$, $q_2 = 0.4$ and $m_2 = 0$, $q_3 = 0.00938$ and $m_3 = 0.173$, as shown in panel (c) of Fig. C.4 (in the case of $c_3$, since it is naturally positive definite, we took $c_3 = \max(q_3 + m_3 \epsilon, 0)$). We therefore used $\tilde{A}(\Theta, E) = \hat{A}(\Theta, q_1 + m_1 \epsilon, q_2 + m_2 \epsilon, \max(q_3 + m_3 \epsilon, 0))$.

C.5. Resulting detection efficiency

Within the framework described in the preceding sections, we computed the detection efficiency $\eta_{\text{det,3D}}(p_{\text{50–3000}}, E_{\text{p,obs}}, \alpha)$ at a number of points on a $(p_{\text{50–3000}}, E_{\text{p,obs}}, \alpha)$ three-dimensional grid. For each point of the grid, we simulated a large number of SGRBs with isotropic $(\theta, \phi)$ positions in the *Fermi* sky, with randomly sampled NaI backgrounds, by repeatedly following the procedure outlined above. Finally, we estimated $\eta_{\text{det,3D}}(p_{\text{50–3000}}, E_{\text{p,obs}}, \alpha)$ as the fraction of simulated SGRBs that yielded a detection over the total.
Panels (a) in Fig. C.6 show the resulting detection probability as a function of $E_{p, \text{obs}}$ and $\alpha$. Panels (b) additionally show the detection efficiency averaged over the observed distribution of $E_{p, \text{obs}}$ (panel b.1), $\alpha$ (panel b.2) or both (panel b.3). The averaging is done by constructing kernel density estimates $w(E_{p, \text{obs}})$ and $w(\alpha)$ of the distributions of these quantities from the best-fit values reported in the GBM spectral catalogue, limiting to the events for which a valid value for the corresponding COMP model parameter was available. The distributions are shown in Fig. C.5. The averaged $\eta_{\text{det}, \text{3D}}$ is then obtained as

$$
\eta_{\text{det}, \text{3D}, \alpha}(\alpha, E_{p, \text{50-300}}) = \int_{0}^{\infty} \eta_{\text{det}, \text{2D}}(\alpha, E_{p, \text{obs}}, E_{p, \text{50-300}}) w(E_{p, \text{obs}}) \, dE_{p, \text{obs}},
$$

$$
\eta_{\text{det}, \text{2D}, \text{3D}, \text{obs}}(E_{p, \text{obs}}, P_{50-300}) = \int_{0}^{\infty} \eta_{\text{det}, \text{2D}}(\alpha, E_{p, \text{obs}}, P_{50-300}) w(\alpha) \, d\alpha,
$$

$$
\eta_{\text{det}, \text{1D}, \text{obs}}(P_{50-300}) = \int_{0}^{\infty} \eta_{\text{det}, \text{2D}, \text{E}_{p, \text{obs}}}(E_{p, \text{obs}}, P_{50-300}) w(E_{p, \text{obs}}) \, dE_{p, \text{obs}}.
$$

(C.6)

The figure demonstrates that the dependence of the detection efficiency on the low-energy photon index is essentially negligible, except for extreme cases where $E_{p, \text{obs}}$ lies well below the $50-300$ keV band. For these cases, on the other hand, the values $P_{50-300}$ reported on the x-axis would correspond to unrealistically large bolometric fluxes. The dependence on $E_{p, \text{obs}}$ is relevant only for very low values of this quantity. For the purposes of our study, we define the detection probability, expressed as a function of the source intrinsic parameters $\lambda_{\text{src}} = (L, E_p, z)$, as

$$
P_{\text{det,GBM}}(L, E_p, z) = \eta_{\text{det}, \text{2D}, \text{E}_{p, \text{obs}}}(E_p/(1+z), p(L, E_p, z)),
$$

where the peak photon flux $p$ is computed as defined in Eq. 17, setting $\alpha = -0.4$. In practice, this is obtained by linear interpolation of $\eta_{\text{det}, \text{2D}, \text{E}_{p, \text{obs}}}$ over its grid. The result, for a number of fixed peak luminosities, is shown in Fig. C.7.