

Constraining binary neutron star population synthesis models using short gamma-ray burst data

A. L. De Santis^{1,2,*}, S. Ronchini^{1,2}, F. Santoliquido^{1,2}, and M. Branchesi^{1,2}

¹ Gran Sasso Science Institute (GSSI), I-67100 L'Aquila, Italy

² INFN, Laboratori Nazionali del Gran Sasso, I-67100 Assergi, Italy

Received 24 February 2026 / Accepted 14 May 2026

ABSTRACT

The landmark multi-messenger observations of the binary neutron star (BNS) merger GW170817 provided firm evidence that such mergers can produce short gamma-ray bursts (sGRBs). However, the limited number of BNS detections by current gravitational wave observatories raises the question of whether BNS mergers alone can account for the full observed sGRB population. We analysed a comprehensive set of 64 BNS population synthesis models with a Monte Carlo-based framework to reproduce the properties of sGRBs detected by *Fermi*-GBM over the past 16 years. We considered three jet geometry scenarios: a universal structured jet calibrated to GW170817, a universal top-hat jet, and a non-universal top-hat jet with distributions of core opening angles. Our results show that models characterised by low local BNS merger rates ($R_{\text{BNS}}(0) \lesssim 50 \text{ Gpc}^{-3} \text{ yr}^{-1}$) predict too few observable sGRBs to reproduce the *Fermi*-GBM population, effectively disfavoured as sole progenitors. Even when relaxing the assumptions on jet geometry, low-rate models remain viable only for wide jets ($\theta_c \geq 15^\circ$), in tension with the narrow jet cores ($\theta_c \approx 6^\circ$) inferred from sGRB afterglow observations. In contrast, models with local merger rates of order $R_{\text{BNS}}(0) \approx 100 \text{ Gpc}^{-3} \text{ yr}^{-1}$ successfully reproduce the observed sGRB population, assuming a plausible fraction of BNS mergers launch relativistic jets and realistic jet geometries. This analysis highlights the power of combining binary neutron star population synthesis models with sGRB observations and comparing them with current gravitational wave constraints to assess the role of BNS mergers as sGRB progenitors and to identify any emerging tensions between these populations.

Key words. astroparticle physics – gravitational waves – gamma-ray burst: general

1. Introduction

The era of multi-messenger astronomy, ushered in by the joint detection of gravitational waves (GWs) and electromagnetic radiation from the binary neutron star (BNS) merger GW170817 (Abbott et al. 2017b,c; Goldstein et al. 2017; Abbott et al. 2017a; Savchenko et al. 2017), has opened unprecedented opportunities to study the physics of compact objects. This event solidified the long-held hypothesis that BNS mergers are the progenitors of short gamma-ray bursts (sGRBs). The connection between BNS mergers and sGRBs, first proposed over three decades ago (Blinnikov et al. 1984; Eichler et al. 1989), is now a cornerstone of multi-messenger astronomy, with the model being developed and solidified by subsequent works (Narayan et al. 1992; Mochkovitch et al. 1993; Nakar 2007) before the first direct, unambiguous confirmation from GW170817 and GRB 170817A. However, the robustness and distinctiveness of this connection remain to be quantitatively established. Gravitational wave observations are opening new possibilities for understanding this relationship, as they now provide constraints on the BNS merger population (The LIGO Scientific Collaboration 2023, 2025). These constraints can be directly compared with the sGRB sample, which is based on more than three decades of observations (von Kienlin et al. 2020; Lien et al. 2016).

A key uncertainty in connecting the BNS and sGRB populations is the efficiency with which BNS mergers produce an sGRB. This efficiency is often encapsulated in the jet fraction,

f_j : the fraction of mergers that successfully launch a relativistic jet powerful enough to break out of the surrounding ejecta. Previous attempts to constrain f_j have yielded a wide range of values, from a few tens of percent (e.g. Salafia et al. 2022; Bhattacharjee et al. 2024) to 100% (Howell et al. 2019). The value of f_j is highly uncertain, as it depends on complex physics such as the merger outcome (e.g. long-lived neutron star versus prompt black hole), the interaction of the jet with merger debris (e.g. Gottlieb et al. 2018; Shibata & Hotokezaka 2019; Pavan et al. 2025), and the properties of the central engine (Ciolfi 2020; Bamber et al. 2024; Hayashi et al. 2025). Similarly, the merger environment itself can shape the angular structure and energetics of an sGRB jet (Pavan et al. 2023). The jet structure is the main uncertainty when determining f_j , as for a given rate of observed sGRBs, a small jet fraction can be compensated by a wide angular structure or, vice-versa, a large fraction with a narrow structure.

GW170817 represented a critical milestone by demonstrating that BNS mergers are capable of launching relativistic jets and that these jets possess an intrinsically structured angular profile. The decisive evidence for the presence of a relativistic jet came from very long baseline interferometry (VLBI), which measured both the apparent displacement of the radio source and its angular size (Mooley et al. 2018; Ghirlanda et al. 2019). Multi-wavelength observations of GW170817 over several months provided the first direct observational evidence against a simple top-hat jet model (characterised by a uniform cone of emission) in favour of a structured relativistic jet (Rossi et al. 2002) observed approximately 20 degrees off-axis

* Corresponding author: alessio.desantis@gssi.it

(Lazzati et al. 2018; D’Avanzo et al. 2018; Margutti et al. 2018; Troja et al. 2018; Lamb et al. 2019).

Based on the GW170817 observations, population studies often adopt the simplifying assumption of universality, using a structured jet profile calibrated to GRB 170817A (e.g. Ronchini et al. 2022). However, it remains unclear whether the jet structure inferred from GW170817 is representative of the broader sGRB population. Moreover, the assumption of universality itself may not be valid, as variations in merger properties and central-engine physics could naturally give rise to a diversity of relativistic outflows.

In this work, we investigate how GW and electromagnetic observations constrain BNS populations, obtained via population synthesis, as progenitors of sGRBs and how fundamental assumptions about jet physics affect these constraints. We structured our analysis as a comparative study. First, we establish a baseline using a physically motivated universal structured jet model calibrated to the best-fit structure of GW170817 (Ghirlanda et al. 2019). Second, we test how our conclusions change when adopting a universal top-hat jet model. Finally, we explore the effect of relaxing the assumption of universality by allowing jet properties to vary across the population. By comparing the results from these distinct scenarios, we can assess the robustness of our conclusions and identify which properties of the BNS population can be constrained irrespective of the jet physics.

While many studies on sGRB population adopt an empirical approach by convolving the star formation history with a delay-time distribution (e.g. Du et al. 2025; Pracchia & Salafia 2026) or by directly parametrizing the merger rate density (Salafia et al. 2023), our work takes a different path by directly testing the end-to-end predictions from population synthesis models. The aim of these models is to predict the properties and rates of BNS mergers based on stellar and binary evolution theory (e.g. Dominik et al. 2012; Belczynski et al. 2018). Here, we used the comprehensive set of models from Iorio et al. (2023), which provide distinct merger rate histories based on different assumptions about key evolutionary phases. While many models can produce local rates consistent with GW observations ($R_{\text{BNS}}(0) \approx 7.6\text{--}250 \text{ Gpc}^{-3} \text{ yr}^{-1}$; The LIGO Scientific Collaboration 2025), a critical tension arises when trying to simultaneously explain the observed sGRB rate. Assuming BNS mergers are the sole sGRB progenitors, any model predicting a low intrinsic BNS merger rate can only be reconciled with observations if the jet fraction is correspondingly high. This creates a testable scenario where models requiring a non-physical fraction (f_j larger than 1) can be disfavoured. Observed sGRBs carry information beyond their event rate. The full observed properties can be used to constrain the underlying source population and emission physics (e.g. Ghirlanda et al. 2016; Ronchini et al. 2022). In this work, we present a systematic framework that accounts for uncertainties in both relativistic jet and BNS population properties with the goal of connecting theoretical BNS population-synthesis predictions to the observed sGRB properties. We used a Markov chain Monte Carlo (MCMC) analysis. The code Multimessenger Analysis with GWs and GRBs in Python (MAGGPY)¹, an optimised implementation based on Ronchini et al. (2022, $\sim 100\times$ speed-up), allowed us to simulate synthetic sGRB catalogues for each of the 64 analysed BNS population models. The MCMC sampling was used to adjust the free parameters of our GRB emission model in order to find the best fit to the observed

Fermi-GBM catalogue and simultaneously yielded a posterior distribution for the required jet fraction f_j . This enabled us to assess which populations and BNS merger rates are most plausible based on their ability to reproduce the observed sGRB population while remaining consistent with GW-inferred local BNS rate constraints.

The paper is structured as follows. In Sect. 2 we outline our methodology. We first describe the general framework for modelling sGRB emission and introduce the specific jet structure models considered in this work: our fiducial structured jet, the comparative top-hat jet, and non-universal variations. We then present the theoretical BNS population models and the *Fermi*-GBM sGRB observations. Finally, we describe the statistical approach used to construct synthetic sGRB catalogues, starting from the BNS population able to reproduce the observed data. Section 3 presents a comparative analysis of the outcomes, directly evaluating how the constraints on BNS populations and jet properties shift depending on the assumed jet model. The results and conclusions are discussed in Sects. 4 and 5, respectively. Throughout the article, we use a Lambda cold dark matter cosmology with Hubble constant $H_0 = 67.66 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and present matter fraction $\Omega_m = 0.31$, from *Planck*-2018 (Planck Collaboration VI 2020).

2. Methodology

In this section we detail the framework of our analysis designed to connect BNS population synthesis models with observed sGRB data. We specify the properties of the prompt emission model, present the geometric configurations for the jets, and explain the methods used to simulate populations for comparison with observational data.

2.1. Short GRB prompt emission model

To connect a BNS merger population to an observable sGRB population, we model the prompt gamma-ray emission expected from each merger. This model is built upon the intrinsic properties of the central engine and the spectral and temporal evolution of the emission.

2.1.1. Intrinsic engine properties

We characterise each burst by three intrinsic, on-axis properties: total radiated energy, characteristic timescale, and rest-frame peak energy. While many studies model the luminosity (L_{iso}) function directly (e.g. Wanderman & Piran 2015; Salafia et al. 2023; Ghirlanda et al. 2016), we model the distributions of total radiated energy E_t and peak time t_p since these quantities are fundamentally linked ($L_{\text{iso}} \sim E_t/t_p$). The total energy converted into radiation, E_r , is drawn from a power law with index k and low-energy exponential cut-off E^* :

$$P(E_t|k, E^*) = \frac{k}{\Gamma(1 - k^{-1})E^*} \left(\frac{E_t}{E^*}\right)^{-k} \exp\left[-\left(\frac{E_t}{E^*}\right)^{-k}\right]. \quad (1)$$

This form captures the observed distribution for sGRBs while providing a turnover at low energies to ensure the distribution is physically realistic (e.g. Pescalli et al. 2016).

The peak time of the burst t_p is drawn from a log-normal distribution with median μ_t and standard deviation σ_t . Similarly, the on-axis rest-frame peak energy of the νF_ν spectrum, E_p , is also drawn from a separate log-normal distribution following Ronchini et al. (2022). For clarity, both variables $x = t_p, E_p$ are drawn from $\log_{10}(x) = \mathcal{N}(\log_{10}(\mu_x), \sigma_x)$.

¹ github.com/LudoDe/MAGGPY

2.1.2. Spectral and temporal properties

For an observer at redshift z and viewing angle θ_v , the observed peak energy at the peak time is $E_{p,\text{pk}}(\theta_v, z) = R_E(\theta_v)E_p/(1+z)$, and the observed peak time is $\tau_p(z) = t_p(1+z)$. Here, $R_E(\theta_v)$ represents the relativistic Doppler factor (see [Ascenzi et al. 2020](#); [Ronchini et al. 2022](#)). Although the observed emission duration in MeV energies depends on inclination (see Appendix A), the peak time itself is dominated by the central engine and is angle-independent. We defined the instantaneous photon energy spectrum in the observer frame as

$$N(E, t, \theta_v, z) = N_{\text{ph}}(t, \theta_v, z) \cdot f(E, E_p(t, \theta_v, z)) \left[\frac{\text{photons}}{\text{cm}^2 \text{ s keV}} \right], \quad (2)$$

where N_{ph} serves as the photon flux normalisation as defined in [Ronchini et al. \(2022\)](#). The term $f(E, E_p)$ describes the time-evolving spectral shape. For this, we adopt a smoothly broken power law (SBPL) model. While most sGRBs in the *Fermi*-GBM catalogue are statistically well-fit by a simpler cut-off power-law (CPL or Compton) model ([von Kienlin et al. 2020](#)), this is often a consequence of low signal-to-noise ratio at high energies. Detailed analysis of bright sGRB by [Ravasio et al. \(2019\)](#) revealed a clear power-law decay above the spectral peak accurately captured by the SBPL form. Additionally, recent magnetohydrodynamic BNS simulations ([Rudolph et al. 2024](#)) revealed a broken power law spectrum for emitted photons. Based on this evidence, we consider the SBPL model to be a more representative description of the entire sGRB population (see Appendix A for a comparison with other models). The normalised SBPL spectral shape is given by [Ravasio et al. \(2018\)](#) as

$$f(E, E_p) = C_n \left[\left(\epsilon \frac{E}{E_p} \right)^{-\alpha n} + \left(\epsilon \frac{E}{E_p} \right)^{-\beta_s n} \right]^{-\frac{1}{n}}. \quad (3)$$

The term ϵ is a function of the spectral indices: $\epsilon = \left(-\frac{2+\alpha}{2+\beta_s} \right)^{\frac{1}{n(\alpha-\beta_s)}}$. We fixed the model's high energy index following [Ravasio et al. \(2019\)](#). The low-energy index was instead assumed to be $\alpha = -2/3$. The high-energy index is $\beta_s = -2.6$, and the smoothness parameter is $n = 2$. The constant C_n normalises the spectral shape. We adopted the convention $f(E = E_p) = 1$, which sets $C_n = \sqrt[3]{2}$. The temporal evolution of the light curve and peak energy follows the model in [Ronchini et al. \(2022\)](#), [Ierardi et al. \(2026\)](#).

From this evolving spectrum, we derived the key observables for comparison with the *Fermi*-GBM catalogue. The observed peak photon flux was calculated at the peak time τ_p and integrated over the GBM band between 50–300 keV (the most sensitive energy band of NaI detectors):

$$F_p(\theta_v, z) = \int_{E_1/(1+z)}^{E_2/(1+z)} N(t = \tau_p, E, \theta_v, z) dE \left[\frac{\text{photons}}{\text{cm}^2 \text{ s}} \right], \quad (4)$$

where $E_1 = 50$ keV and $E_2 = 300$ keV. This quantity is additionally calculated on a 64 ms timescale to match the catalogue data. The bolometric energy light curve $F(t)$ is given by integrating the energy spectrum at each time step:

$$F(t, \theta_v, z) = \int_0^\infty E N(E, t, \theta_v, z) dE \left[\frac{\text{erg}}{\text{cm}^2 \text{ s}} \right]. \quad (5)$$

From this, we defined the cumulative fluence within the GBM band, $S(t) = \int_0^t \int_{E_1/(1+z)}^{E_2/(1+z)} E N(E, t', \theta_v, z) dE dt'$, from which the total fluence and the duration T_{90} were calculated.

2.2. Predicting the observed sGRB rate

We defined a fundamental parameter, f_j , connecting the BNS merger population to the observed sGRB sample, given by the ratio between the intrinsic sGRB rate (N_{sGRB}) and the BNS merger rate (N_{BNS}):

$$f_j = N_{\text{sGRB}}/N_{\text{BNS}}. \quad (6)$$

When $f_j \leq 1$, this quantity can be interpreted as the fraction of binary BNS mergers that are able to produce a successful relativistic jet, namely a jet that is able to break out from the post merger ejecta. In our framework we treat it as a free parameter inferred during the MCMC sampling, which ensures that all populations reproduce the same observed GRB rate. This approach allows f_j to explore values greater than unity as a diagnostic tool to test the viability of the BNS population in reproducing the sGRB data. If the posterior distribution for f_j lies predominantly above 1, it indicates that the BNS population model under consideration is insufficient to account for the observed sGRB rate.

We derive the predicted sGRB rate observed by *Fermi*-GBM by convolving the instrument's duty cycle, the jet production efficiency, the source population properties, and the detector sensitivity. In the most general case the detection efficiency Φ depends on the redshift, the intrinsic engine parameters θ_p (e.g. t_p or E_p), and the viewing angle θ_v . The predicted rate $R_{\text{sGRB}}^{\text{pred}}$ is

$$R_{\text{sGRB}}^{\text{pred}} = \epsilon_{\text{GBM}} \cdot f_j \cdot \int \Phi(\theta_p, \theta_v, z) \frac{R(z)}{1+z} \frac{dV}{dz} d\theta_p d\cos(\theta_v) dz, \quad (7)$$

where $\epsilon_{\text{GBM}} \approx 0.60$ is the combined effective instrument efficiency and field-of-view ([Burns et al. 2016](#)), and $d\cos(\theta_v)$ represents an isotropic distribution of viewing angles.

For the universal top-hat jet model (Sect. 2.3), the detection efficiency factorises because the emission is uniform within a cone of semi-aperture angle θ_c and negligible outside. We note that this sharp-edge approximation neglects the $1/\Gamma$ relativistic beaming spread of the emission pattern, which could allow detections slightly outside θ_c . However, for highly relativistic outflows ($\Gamma \gg 1$), this contribution is subdominant and the factorised approximation remains valid for population-level counting. In this specific case, an observer detects a burst only if $\theta_v \leq \theta_c$ and the on-axis flux is above the threshold ([Matsumoto et al. 2019](#)). This allowed us to evaluate the integral of the geometric beaming independently:

$$R_{\text{sGRB}}^{\text{pred}} = \epsilon_{\text{GBM}} \cdot f_j \cdot \underbrace{\int_0^{\theta_c} \sin \theta_v d\theta_v}_{\langle f_b \rangle} \cdot \int \Phi(\theta_p, z) \frac{R_{\text{BNS}}(z)}{1+z} \frac{dV}{dz} d\theta_p dz, \quad (8)$$

where $\langle f_b \rangle = (1 - \cos \theta_c)$ is the associated beaming factor. In our non-universal models (Sect. 2.3), where θ_c follows a distribution $p(\theta_c)$, the average beaming factor generalises to the probability that a random viewing angle is less than the core of a randomly sampled jet:

$$\langle f_b \rangle = \iint \Theta(\theta_c - \theta_v) \sin \theta_v p(\theta_c) d\theta_v d\theta_c, \quad (9)$$

where Θ is the Heaviside step function. By requiring $R_{\text{sGRB}}^{\text{pred}}$ matches the observed rate $R_{\text{sGRB}}^{\text{obs}}$, we could infer the necessary jet fraction f_j for each BNS population model for different geometrical assumptions.

2.3. Jet structure models

The intrinsic energy described in Sec 2.1 is not emitted isotropically but is channelled into a relativistic jet. The geometry of relativistic jets in GRBs is poorly constrained through observations and serves as a major source of uncertainty in population studies. We investigate three distinct assumptions for the jet structure.

2.3.1. Universal structured jet

Our primary model assumes a universal, axisymmetric structured jet, strongly motivated by the afterglow of GRB 170817A. Here, the energy and bulk Lorentz factor decrease with angular distance θ from the jet's core. We assumed angular profiles for the isotropic-equivalent kinetic energy and the bulk Lorentz factor given by

$$\frac{dE}{d\Omega} = \frac{E_c}{1 + (\theta/\theta_c)^{s_1}} \quad (10)$$

$$\Gamma(\theta) = 1 + \frac{\Gamma_c - 1}{1 + (\theta/\theta_c)^{s_2}}. \quad (11)$$

Following Ronchini et al. (2022), we adopted fixed structural parameters consistent with the best-fit values for GRB 170817A (Ghirlanda et al. 2019): a core opening angle $\theta_c = 3.4^\circ$, an on-axis core Lorentz factor $\Gamma_c = 500$, and power law indices $s_1 = s_2 = 4$. The apparent properties depend on the viewing angle θ_v , with scaling factors for peak energy and flux calculated following the formalism of Ascenzi et al. (2020). In this framework, the MCMC samples seven free parameters: two for the energy distribution (power law index and cut-off k, E^*), two for peak energy and spread ($\log_{10} \mu_E, \sigma_E$), two for peak time and spread ($\log_{10} \mu_t, \sigma_t$), and the jet fraction (f_j).

2.3.2. Universal top-hat jet

To assess the impact of the assumed jet profile, we also analysed a simplified universal top-hat jet model. As mentioned above this model introduces a degeneracy between the jet fraction and the opening angle given by the efficiency $f_j(1 - \cos(\theta_c)) = \text{const.}$ This allowed us to determine geometries and jet fractions that are compatible with observed jet opening angles. In this framework, we do not use a temporal evolution of the light curve but instead sample the intrinsic peak luminosity L from a distribution analogous to Eq. (1). The observed peak flux is then calculated as $F_p = L/(4\pi d_L(z)^2 \mathcal{K}(E_p, z))$, where $\mathcal{K}$ is the k-correction factor (Salafia et al. 2023; Poolakkil et al. 2021). For this model, the MCMC samples six free parameters: two for luminosity power law index and cut-off (k, L^*), two for peak energy and spread ($\log_{10} \mu_E, \sigma_E$), the jet fraction (f_j) and the core opening angle (θ_c).

2.3.3. Exploring non-universality

Finally, we relaxed the assumption of universality to test whether a diverse jet population significantly alters our constraints on BNS models. In this framework, we extended the top-hat jet model by allowing the core opening angle, θ_c , to vary across the population rather than being a single fixed value. We investigated two distinct scenarios for the population-wide distribution of θ_c , as illustrated in Fig. A.3. The first scenario is a *flat model*, where θ_c is drawn from a uniform distribution between a minimum of 1° and a variable upper bound, θ_c^{max} . This model represents a scenario with no preferred physical scale for the opening

angle, limited only by the maximum width of the jet geometry. The second scenario is a *log-normal model*, characterised by a median angle θ_c^{med} and a fixed shape parameter $\sigma_{\log_{10}} = 0.5$. This width is chosen to provide a realistic degree of diversity that covers approximately an order of magnitude in core angles, consistent with the heterogeneity observed in sGRB afterglow studies (Rouco Escorial et al. 2023).

To ensure physical consistency, both distributions produce minimum angles of $\theta_c^{\text{min}} = 1^\circ$ and specifically for the log-normal case, we cut angles larger than $\theta_c \geq 45^\circ$. Crucially, in these non-universal models, the isotropic equivalent peak luminosity L is sampled independently of θ_c . This implies that wider jets in our framework possess larger total jet energies ($E_{\text{tot}} \propto L\theta_c^2$; Salafia et al. 2022), rather than assuming a constant energy reservoir that would imply $L \propto \theta_c^{-2}$. We also explicitly tested this alternative scenario, confirming that it does not significantly alter our inferred posteriors (see Appendix B for details). In these non-universal models, the MCMC simultaneously samples the jet fraction f_j and the specific distribution parameters (either θ_c^{max} or θ_c^{med}). As in the universal top-hat model scenario, this approach allowed us to directly quantify the trade-off between the underlying BNS merger rate and the required jet geometry.

2.4. Fermi-GBM data

Historically, GRBs are classified into two populations: ‘short-hard’ and ‘long-soft’. This bimodality, first identified in Konus data (Mazets & Golenetskii 1981), was unambiguously established with the BATSE catalogue (Kouveliotou et al. 1993) through the separation of events in the duration-hardness plane.

To calibrate the BNS population with the sGRB observations, we selected sGRB data from the *Fermi*-GBM burst catalogue² covering a time span of 16 years (July 2008 – May 2025; see Appendix C for more details). Our approach to select the sample is similar to that of Ghirlanda et al. (2016), Salafia et al. (2023), aiming to construct a well-defined dataset by applying specific cuts on GRB duration and peak flux. As a proxy for duration, we adopt T_{90} , defined as the time interval containing 5% to 95% of the total measured fluence. Following established literature (e.g. Paciesas et al. 1999; Lien et al. 2016; von Kienlin et al. 2020), we applied the conventional threshold of $T_{90} < 2$ s to isolate the sGRB population. Although this threshold may include contamination from short-duration collapsars (single stars; Zhang et al. 2009; Bromberg et al. 2013), we verified that its impact on our population-level results is minimal. In fact, using a more restrictive $T_{90} < 1$ s cut, we find no significant changes in the distribution of GRB observables. Given the inherent difficulty in ensuring progenitor purity solely through duration (see e.g. Giudice et al. 2025), we retain the 2 s limit to maximise the sample size and minimise statistical uncertainty. To mitigate selection effects, we imposed a peak photon flux cut of $F_p^{\text{lim}} = 4 \text{ ph cm}^{-2} \text{ s}^{-1}$ (measured on a 64 ms timescale in the 50–300 keV band). We further restricted our sample to events with observed spectral peak energies (E_p) between 50 keV and 10 MeV, corresponding to the GBM detection band (see Fig. C.1 for both cuts). The peak photon flux F_p^{lim} cut is chosen to ensure the cumulative peak flux distribution follows the $N(>F) \propto F^{-3/2}$ power law expected for a uniform distribution in Euclidean space (Ghirlanda et al. 2016; Ronchini et al. 2022). Notably, while the *Fermi*-GBM catalogue provides E_p values derived from Comptonised model fits, our framework

² <https://heasarc.gsfc.nasa.gov/W3Browse/fermi/fermigbrst.html>

employs an SBPL. As noted in the literature (e.g. Gruber et al. 2014), Comptonised fits return systematically higher E_p values ($\sim 6\%$ on average in our sample) compared to models with high-energy tails, such as the Band function or an SBPL. However, given the large individual statistical uncertainties on E_p , this minor systematic offset is well within the intrinsic scatter of the population, and this discrepancy does not introduce significant bias (see Sect. 2.1 for a more thorough discussion). After the cut in peak flux, our sample contains 310 sGRBs. This corresponds to an observed rate of $R_{\text{sGRB}}^{\text{obs}} = 18.61 \text{ yr}^{-1}$. This rate is less than the one used in Ronchini et al. (2022), as a lower value for F_p^{lim} (0.5) was used in that work. After applying the quality cut on the peak energy, we are left with 221 events. For each of these events, we retrieved the catalogued values for the fluence in the 50–300 keV band, T_{90} , peak flux in the 50–300 keV band and peak energy. We adopt this band, since the T_{90} of each GRB is computed only in that band in the *Fermi*-GBM catalogue. We highlight that even if we impose a cut on the measured peak energy to ensure a good fit quality, that cut is relative only to the GRBs considered for the comparison between the predicted and observed distributions. However, in estimating the true GBM detection rate, we impose no quality cuts and consider all events that triggered the detector, regardless of data quality.

2.5. Theoretical populations of BNS mergers

In this work, we use a comprehensive set of 16 binary population synthesis models from Iorio et al. (2023). These models explore the large parameter space of stellar and binary evolution uncertainties. Here, we briefly recap how the merger rate density and its redshift evolution are obtained and highlight the assumptions (see Sects. 2.5.1 and 2.5.2) that have the most significant impact on the merger rate density.

The models are generated using the SEVN code, which evolves a total of $\sim 10^9$ binary systems across 15 metallicity bins spanning the range $Z \in [10^{-4}, 3 \times 10^{-2}]$. The primary masses, $m_{\text{ZAMS},1}$, are drawn from a Kroupa initial mass function (IMF; Kroupa 2001) over the range $[5, 150] M_{\odot}$, with a minimum secondary mass of $m_{\text{ZAMS},2} = 2.2 M_{\odot}$. We used the semi-analytic code COSMORATE (Santoliquido et al. 2020, 2021) to compute the cosmic merger rates from the merged binaries simulated with SEVN (Spera et al. 2019; Mapelli et al. 2020). The source frame merger rate density, $\mathcal{R}(z)$, was computed by convolving the delay-time distribution of the binaries with the cosmic star-formation history:

$$\mathcal{R}(z) = \int_z^{z_{\text{max}}} \int_{Z_{\text{min}}}^{Z_{\text{max}}} \psi(z') p(z', Z) \mathcal{F}(z', z, Z) dZ \frac{dt(z')}{dz'} dz', \quad (12)$$

where $t(z)$ is the look-back time at redshift z . The term $\mathcal{F}(z', z, Z)$ is defined as

$$\mathcal{F}(z', z, Z) = \frac{1}{M_{\text{pop}}(Z)} \frac{dN(z', z | Z)}{dt(z)}, \quad (13)$$

where $M_{\text{pop}}(Z)$ is the total initial mass of the simulated stellar population, normalised to the full IMF mass range. The term $\frac{dN(z', z | Z)}{dt(z)}$ represents the rate of binary compact object mergers forming from stars with initial metallicity Z at redshift z and merging at redshift z' extracted from the SEVN catalogues. We adopted the analytical fit for the star formation rate density, $\psi(z)$, from Madau & Fragos (2017):

$$\psi(z) = a \frac{(1+z)^b}{1 + [(1+z)/c]^d} [M_{\odot} \text{ yr}^{-1} \text{ Mpc}^{-3}], \quad (14)$$

with parameters $a = 0.01$, $b = 2.6$, $c = 3.2$, and $d = 6.2$. The metallicity distribution $p(Z | z)$ was assumed to follow a log-normal distribution:

$$p(Z | z) = \frac{1}{\sqrt{2\pi\sigma_Z^2}} \exp \left\{ -\frac{[\log(Z/Z_{\odot}) - \langle \log Z(z)/Z_{\odot} \rangle]^2}{2\sigma_Z^2} \right\}, \quad (15)$$

with $\sigma_Z = 0.1$. From the merger rate density, we defined two key quantities for our analysis, the local BNS merger rate density, $R_{\text{BNS}}(0) \equiv \mathcal{R}(z = 0)$, and the total integrated rate, Λ , defined as

$$\Lambda = \int_0^{z_{\text{max}}} \frac{\mathcal{R}(z)}{1+z} \frac{dV}{dz} dz, \quad (16)$$

where $\frac{dV}{dz}$ is the differential co-moving volume.

In our analysis, we assumed the Fiducial population (Model F, $\alpha_{\text{CE}} = 1$, $\sigma_z = 1$) from Iorio et al. (2023) as our baseline. This model yields a local merger rate density of $R_{\text{BNS}}(0) = 112 \text{ Gpc}^{-3} \text{ yr}^{-1}$, which lies near the centre of the 90% credible interval reported in the GWTC-4 catalogue (The LIGO Scientific Collaboration 2025). A summary of all analysed BNS merger populations is provided in Appendix D.

2.5.1. Common envelope

The common envelope (CE) phase is a critical, yet highly uncertain, stage in binary evolution (Webbink 1984; Livio & Soker 1988) and can change predicted BNS merger rates by orders of magnitude (Ivanova et al. 2013; Broekgaarden et al. 2022; Santoliquido et al. 2021; Mandel et al. 2020). In the CE phase, both stars orbit within a shared and extended envelope. Drag forces remove orbital energy and angular momentum, shrinking the orbit. If the orbital energy deposited into the envelope is sufficient to unbind it, the envelope is ejected, leaving behind a tighter post-CE binary. Otherwise, the stars coalesce. The CE ejection efficiency, α_{CE} , sets how effectively orbital energy unbinds the envelope. Low α_{CE} (e.g. 0.5 and 1) requires closer inspirals to eject the envelope, increasing premature mergers during CE while surviving binaries have tighter orbits and shorter GW-driven delay times. Iorio et al. (2023) finds that for $\alpha_{\text{CE}} \leq 1$, premature coalescences suppress BNS formation overall. High α_{CE} (e.g. 3 and 5) favours CE survival at wider separations. Across models, varying α_{CE} shifts the local BNS merger rate density from fewer than one to over $300 \text{ Gpc}^{-3} \text{ yr}^{-1}$ (Figs. 19, 20, 22 in Iorio et al. 2023), confirming α_{CE} as the dominant parameter in determining the theoretical BNS merger rate (Santoliquido et al. 2021). Within this work we consider the values $\alpha_{\text{CE}} = 0.5, 1, 3, 5$.

2.5.2. Natal kicks, mass transfer stability, and other assumptions

The velocity kicks imparted to neutron stars at birth are important for binary evolution. Models drawing kicks from a Maxwellian distribution (see Hobbs et al. 2005) with dispersion $\sigma = 150 \text{ km/s}$ (model K150) yield higher BNS survival and merger rates than models with $\sigma = 265 \text{ km/s}$ (K σ 265), where disruptions are more frequent.

The criteria that determine whether mass transfer from one star to another is stable or instead initiates a CE phase can also alter BNS formation channels and rates. The fiducial model assumes mass transfer is always stable for main sequence and Hertzsprung gap donors. The QCBSE model instead adopts the standard stability criteria from Hurley et al. (2002), widely used

in codes like MOBSE (Giacobbo & Mapelli 2018). The QCBB model goes further, assuming mass transfer is also always stable for pure-Helium star donors, which allows for stable mass transfer from pure-Helium stars in scenarios that might otherwise be considered unstable and thus enhances the BNS merger rate by providing an additional formation channel, following Mandel et al. (2020).

Several other physical parameters were tested by Iorio et al. (2023), but they were found to have a comparatively minor direct impact on the overall BNS merger rates. These include variations in Roche-lobe overflow accretion efficiency (model RBSE), the specific supernova engine model (SND), the pair-instability supernova model (F19), and the treatment of tides (NT, NTC). While these factors can influence other aspects of compact binary populations or specific BNS properties, their effect on the total BNS merger numbers is secondary to the parameters discussed above.

We analysed sixteen population synthesis models from Iorio et al. (2023). For each, we considered four values of α_{CE} resulting in a total of 64 BNS merger populations.

2.6. Analysis framework

From each BNS population, we generated a corresponding synthetic sGRB population calibrated to reproduce the *Fermi*-GBM observations. We employ a hierarchical Bayesian analysis to infer the properties of the sGRB population from the observed data. For each of the three jet structure models described in Sect. 2.3, we defined a posterior probability distribution for the model's free parameters, θ . We then sampled this posterior using a MCMC algorithm, implemented with the EMCEE package (Foreman-Mackey et al. 2013). The specific set of parameters θ and their priors are detailed below for each model.

2.6.1. Universal structured jet model

This model requires a time-evolving light curve and spectrum. The MCMC samples seven free parameters, $\theta = \{k, E^*, \mu_E, \sigma_E, \mu_t, \sigma_t, f_j\}$. Their descriptions and prior ranges are given in Table 1.

2.6.2. Universal top-hat jet model

In the universal top-hat jet model, the core half-opening angle θ_c is assumed to be the same for all the sGRBs. The abundance of observed sGRB is given by the efficiency, $\epsilon = f_j(1 - \cos \theta_c)$. This approach allowed us to quantify how specific geometric assumptions impact the required BNS abundance (f_j). This model does not include time evolution and samples the intrinsic peak luminosity directly. It has six free parameters, $\theta = \{k, L^*, \mu_E, \sigma_E, \theta_c, f_j\}$. While we kept the same prior of f_j as before, we added a flat prior $\theta_c \in [1^\circ, 25^\circ]$. This choice is motivated by observational constraints on sGRB beaming. Population studies of late-time afterglows find a median opening angle of $\langle \theta_c \rangle \approx 6^\circ$ with a tail to wider jets, including measurements and lower limits of $\theta_c \geq 10^\circ$ in $\approx 30\%$ of the sample in Rouco Escorial et al. (2023). Early X-ray afterglow analyses further indicate that most sGRBs are viewed within or very near their jet opening angle, consistent with compact cores and disfavoured very large core angles (O'Connor et al. 2024). Thus, $\theta_c \leq 25^\circ$ is conservative and comfortably includes most of the observed wide core tail without truncating plausible values, while avoiding prior support for unrealistically large cores. All parameters and their priors are detailed in Table 1.

2.6.3. Non-universal top-hat jet models

The non-universal models expand upon the structured jet and universal top-hat jet frameworks by allowing the core opening angle, θ_c , to vary on an event-by-event basis, according to an underlying population distribution. This approach introduces an additional parameter, P_θ , to the MCMC sampling, which describes the characteristic scale of the core geometry. The sampled parameter set for these models is $\theta = \{k, L^*, \mu_E, \sigma_E, \theta_c, f_j, P_\theta\}$, as detailed in Table 1. For the Flat model, P_θ represents the upper bound θ_c^{max} . In this scenario, the opening angle was sampled from $\mathcal{U}(1^\circ, \theta_c^{\text{max}})$, effectively testing a population with no preferred scale up to a maximum cut-off ($\theta_c^{\text{max}} < 25^\circ$). For the Log-Normal model, the parameter P_θ corresponds to the median angle θ_c^{med} . As noted in the previous section, we fixed the dispersion to $\sigma_{\log_{10}} = 0.5$ to ensure the population spans roughly one order of magnitude in jet core angle width. To maintain physical consistency and avoid negative or non-physically large angles, the distribution was truncated and renormalised to the range $[1^\circ, 45^\circ]$. In both cases, the simultaneous inference of the jet fraction f_j and the geometric hyperparameter allowed us to study the degeneracy between the BNS merger rate and the intrinsic beaming of the sGRB population. We highlight that although the MAGGPY framework is designed to allow users to define and test arbitrary jet structures, the present analysis focuses exclusively on the three models described above, as they provide a representative coverage of the geometric uncertainties most relevant to the sGRB population.

2.6.4. Likelihood

For each set θ drawn from the priors, we generated a synthetic catalogue of sGRBs. From a population of BNS mergers with redshifts and viewing angles, the sGRB/BNS ratio f_j (see Eq. (6)) determines the number of successful GRBs. We then drew the intrinsic properties for each event and applied the corresponding jet model to derive the observable quantities. For the structured jet model we derived four key observables for comparison with the *Fermi*-GBM data: fluence (S_i), peak flux ($F_{p,64,i}$), peak energy ($E_{p,i}$), and duration ($T_{90,i}$). For the top-hat models, which lacks a temporal profile, the comparison was performed using only the peak flux and peak energy. The final set of selected bursts, $N_{\text{det,sim}}$, are those that satisfy the selection criteria from our observational sample.

The comparison between simulated catalogue and real *Fermi*-GBM data is based on two components: the observable distribution shape and the event rate. To compare shapes, we used the two-sample Cramér-von Mises (CvM) test, which integrates the squared difference between the Empirical Cumulative Distribution Functions (ECDFs) of the simulated and observed data. This integral approach makes the test sensitive to discrepancies across the entire distribution, rather than a single local maximum. The choice of CvM over the Kolmogorov-Smirnov (KS) statistic used in previous works (Ronchini et al. 2022) is motivated by its superior stability and higher statistical power in Monte Carlo sampling (Stephens 1974). While both statistics are consistent and converge to the same limit, the CvM statistic exhibits lower variance across stochastic simulations (D'Agostino & Stephens 1986). This provides a p-value, $p_{\text{CvM},X}$, for each observable X . To constrain the rate, we used a Poisson likelihood, $P_{\text{Poisson}} = \text{Poisson}(N_{\text{det,sim}}|\mu_{\text{exp}})$, where μ_{exp} is the expected number of detections. Given the complexity of the detector effects, constructing a direct analytical likelihood

Table 1. Free parameters and prior distributions for the sGRB emission models.

Parameter	Description	Prior	Models
<i>Common parameters</i>			
k	Power-law index (Energy or Luminosity)	$\mathcal{U}(1.5, 12)$	All
$\log_{10}(\mu_E/\text{keV})$	Median spectral peak energy	$\mathcal{U}(2.5, 4.0)$	All
σ_E	Standard deviation of $\log_{10} E_p$	$\mathcal{U}(0.05, 1.0)$	All
f_j	Jet launching fraction	$\mathcal{U}(0, 10)$	All
<i>Universal structured jet parameters</i>			
$\log_{10}(E^*/\text{erg})$	Characteristic energy cut-off	$\mathcal{U}(46, 50)$	Structured
$\log_{10}(\mu_t/\text{s})$	Median peak time	$\mathcal{U}(-2.0, 0.3)$	Structured
σ_t	Standard deviation of $\log_{10} t_p$	$\mathcal{U}(0.05, 1.8)$	Structured
<i>Top-hat jet parameters</i>			
$\log_{10}(L^*/\text{erg s}^{-1})$	Characteristic luminosity cut-off	$\mathcal{U}(48, 54)$	All Top-Hat
θ_c (deg)	Universal core half-opening angle	$\mathcal{U}(1, 25)$	Univ. Top-Hat
θ_c^{max} (deg)	Maximum core opening angle (P_θ)	$\mathcal{U}(1, 25)$	Flat Non-Univ.
θ_c^{med} (deg)	Median core opening angle (P_θ)	$\mathcal{U}(1, 25)$	Log-Norm Non-Univ.

Notes. Common parameters are sampled in every model configuration; model-specific parameters are included only when the corresponding jet structure is selected. $\mathcal{U}(a, b)$ denotes a uniform prior within $[a, b]$.

is not feasible. Combining these metrics, we obtained the likelihood

$$\log \mathcal{L}(\theta) = \sum_X \log(p_{\text{CeM},X}) + \log(P_{\text{Poisson}}), \quad (17)$$

where the sum over X includes the relevant observables. The MCMC maximises this function to find the sets θ that best reproduce the observed sGRB population. Details regarding the convergence of our chains and posterior predictive checks are provided in Appendix E.

3. Results

Here, we present the outcomes of our analysis, performed independently for each of the 64 BNS population synthesis models. By comparing the predictions derived under different assumptions for the sGRB jet structure, we assess the physical viability of each BNS model and the robustness of our conclusions. The primary diagnostic is the posterior distribution of the jet fraction f_j , which is constrained by the requirement that the model reproduces the observed *Fermi*-GBM sGRB population. By allowing $f_j > 1$, we derived a quantitative metric for model viability, which is the degree to which a model fails to produce enough progenitors to match the *Fermi*-GBM rate.

3.1. Structured jet population

Applying our analysis to all 64 BNS population synthesis models and assuming a structured jet as described in Sect. 2.3, we obtained the posterior distributions of the jet fraction, f_j , shown in Fig. 1. These distributions form the basis for evaluating the physical viability of each population.

A wide range of f_j posterior distributions is observed, reflecting the significant variation in predicted merger rates across the models spanning orders of magnitude. Models predicting intrinsically high merger rates are compatible with relatively smaller jet fractions. Typically, higher merger rates correspond to larger α_{CE} . However, for several models, a trend is observed in which $\alpha_{\text{CE}} = 5$ yields lower merger rates than $\alpha_{\text{CE}} = 3$. This behaviour is primarily due to the higher efficiency of CE ejection, which leaves the binary system at a wider separation after the CE phase,

thereby requiring a longer time to merge. Conversely, for models predicting low intrinsic merger rates (e.g. those with low α_{CE} or high kicks like the model K265) the posterior distributions of f_j are widely spread above the physical threshold of unity. This suggests a scenario for these models where the predicted number of BNS mergers is insufficient to explain the sGRB observations. As detailed in Appendix F, we allowed $f_j > 1$ rather than enforcing a strict physical boundary because it provides a quantifiable metric of the missing mergers required by low-rate models, enabling a clearer distinction between marginal and significant model tensions that would otherwise be obscured by saturation at the prior limit. To better understand the relationship between jet fraction and BNS merger rate, we plot the median posterior value of f_j against the local BNS merger, $R_{\text{BNS}}(0)$, and the total event rate obtained by integrating over redshift, Λ , for each population synthesis model in Fig. 2 (top row). Here, the anti-correlation between rate and jet fraction is more clearly visible. This correlation exhibits scatter, which is more pronounced when plotting against the local merger rate than the total integrated rate. This scatter arises from models predicting different redshift evolutions, which alters the total number of detectable events in a way that is not captured by the local rate alone.

We also quantified the fraction of the posterior of f_j that lies in the non-physical region by plotting the quantile evaluated at the boundary $f_j = 1$, i.e. $P(f_j \leq 1)$, shown in the bottom row of Fig. 2. Throughout this work, we defined a BNS population as physical if the median of its jet fraction posterior satisfies $f_j \leq 1$ (equivalent to requiring $P(f_j \leq 1) \geq 0.5$). These can be considered as the preferred models, i.e. those for which BNS abundance is enough to reconcile with sGRB observations. Notably, we find that models predicting $R_{\text{BNS}}(0) \lesssim 100 \text{ Gpc}^{-3} \text{ yr}^{-1}$ exhibit a progressive shift of the jet fraction f_j posterior above unity. This indicates that under the assumption that BNS mergers are the sole progenitors of sGRBs, explaining the observations would require either a non-physical scenario in which more than one GRB is produced per BNS merger or the presence of additional progenitor channels.

This finding provides a powerful, independent constraint when compared with GW observations. The shaded vertical region in Fig. 2 shows the 90% credible interval for the BNS merger rate from GWTC-4 (The LIGO Scientific Collaboration

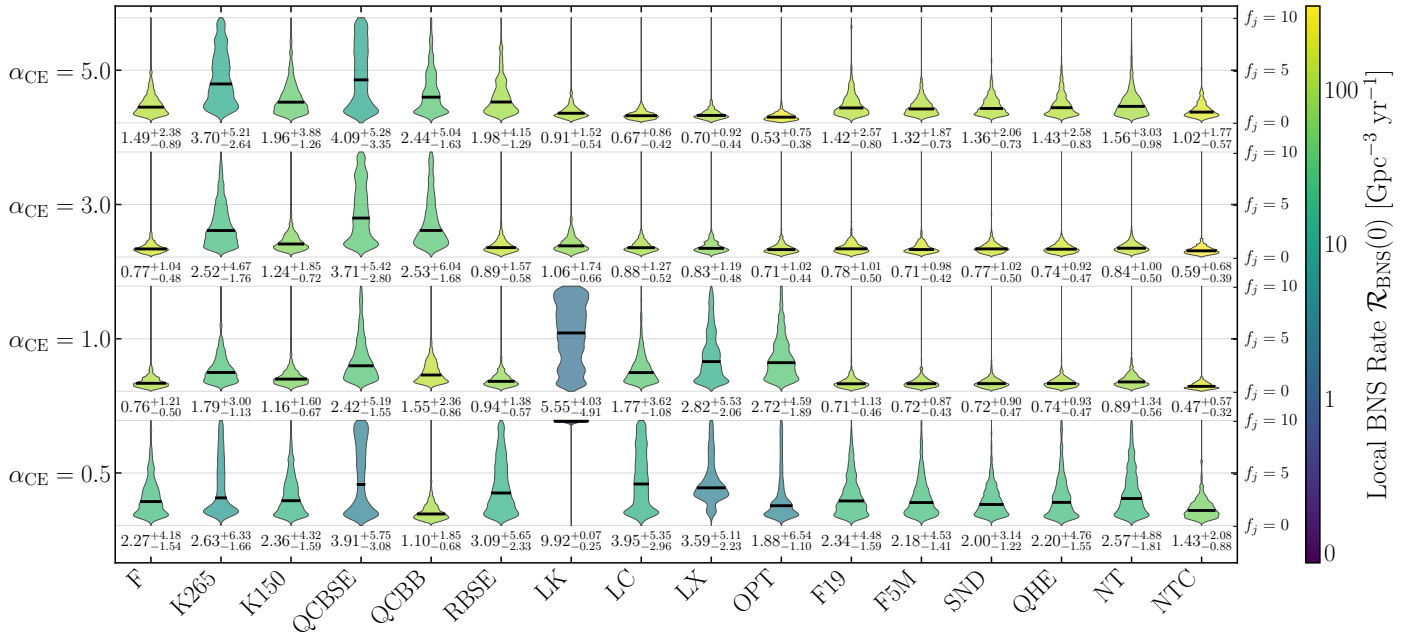


Fig. 1. One-dimensional marginalised posterior distributions of the jet fraction f_j for each considered BNS population model, assuming a universal jet structure calibrated to GRB170817A. Black lines mark medians; labels also provide the 90% C.I. values. Axes show models (x -axis) and α_{CE} (y -axis). The colours indicate the local merger rate.

2023). While many of the analysed models fall within this broad interval, our sGRB analysis effectively establishes a multi-messenger viability window for the BNS mergers as the progenitors of the entire sGRB population.

We can define a conservative lower bound at $R_{\text{BNS}}(0) \lesssim 50 \text{ Gpc}^{-3} \text{ yr}^{-1}$ for which the majority of the found posterior distribution is within non-physical bounds and for which the MCMC analysis indicates that at least a factor of 2–3 larger BNS merger rate would be required to reproduce the observed sGRB rate. Conversely, the GWTC-4 bounds $R_{\text{BNS}}(0) \leq 250 \text{ Gpc}^{-3} \text{ yr}^{-1}$ give an upper limit on the BNS local rate.

We note that the range for the BNS local rate from GWTC-4 was obtained by adding only a small portion of the O4 run (almost the first 7 months) to the first three LVK observing runs, and that, to date (Feb 2026), no significant BNS candidate has been released in low-latency in the subsequent one year and half of the O4 run. If the final analysis of O4 data and future GW observations push the allowed parameter space towards BNS merger rates lower than the floor implied by our sGRB constraints, it would become increasingly difficult to account for the full sGRB population with BNS mergers alone. Such a scenario would strengthen the case for a significant contribution from alternative progenitors, such as neutron star-black hole (NSHB) mergers, or that different geometric hypotheses including much wider jet cores are necessary. We therefore arrive to the conclusion that under the assumption of a universal GW170817-like jet structure for the sGRB population and that BNS mergers constitute the sole dominant progenitor channel, BNS populations with local merger rates $R_{\text{BNS}}(0) \lesssim 50 \text{ Gpc}^{-3} \text{ yr}^{-1}$ are difficult to reconcile with sGRB observations.

3.2. Universal top-hat jet

We repeat the analysis using a simplified universal top-hat model. This model constrains the overall efficiency $\epsilon = f_j(1 - \cos(\theta_c))$, which combines the jet fraction and the beaming angle. This allowed us to test various values of jet core aperture. In

Appendix E we show the posterior distributions for the fiducial BNS population (Fig. E.3). The corresponding median values and confidence intervals are provided in the right-hand columns of Table E.1. We find that the overall efficiency ϵ must be low to reproduce the observed data. The posterior distributions of the parameters k , μ_E , and σ_E closely match those found for the structured model (see Table E.1 and Fig. E.2). The L^* posterior is consistent with findings from other works on the GRB luminosity distribution, such as Wanderman & Piran (2015, $\sim 10^{51} \text{ erg/s}$).

The results for the 64 population models are summarised in Fig. 3. Although the sampler explores f_j and θ_c separately, we visualise the efficiency ϵ as the observed rate is directly proportional to this parameter. As expected, models with lower BNS rates require an overall higher efficiency ϵ to reproduce GRB observables. Owing to the degeneracy between the jet fraction and the jet aperture, this implies that such populations would require jet fractions exceeding unity or unrealistically large jet cores. To visualise this effect and to enable a direct comparison with our structured jet model, we condition the two-dimensional posterior of $P(f_j, \theta_c)$ on three fixed values of the opening angle ($\theta_c = 5^\circ, 10^\circ$, and 20°). Figure 4 demonstrates that for both narrow and wide jet assumptions, BNS models with very low intrinsic merger rates yield f_j posteriors that extend significantly above unity.

As in the structured jet case, for each assumed value of the opening angle θ_c , we computed the fraction of the posterior that is physical (i.e. $P(f_j \leq 1 | \theta_c)$) to provide a more direct assessment of which models can reasonably reproduce the observed sGRB rates. A model is considered physical if the median of its jet fraction distribution is ≤ 1 , which corresponds to $P(f_j \leq 1) \geq 0.5$. The quantiles associated with the 3 analysed geometries are shown in Fig. 5, where we observe that for lower opening angles ($\theta_c \approx 5^\circ$), models with local rates $R_{\text{BNS}}(0) \lesssim 50 \text{ Gpc}^{-3} \text{ yr}^{-1}$ have almost their entire posterior distribution spread above $f_j > 1$, in agreement with what found in the universal structured jet scenario. To reconcile these low-rate models with observations, significantly larger opening angles ($\theta_c \gtrsim 10^\circ$ – 20°) are required. This creates tension with

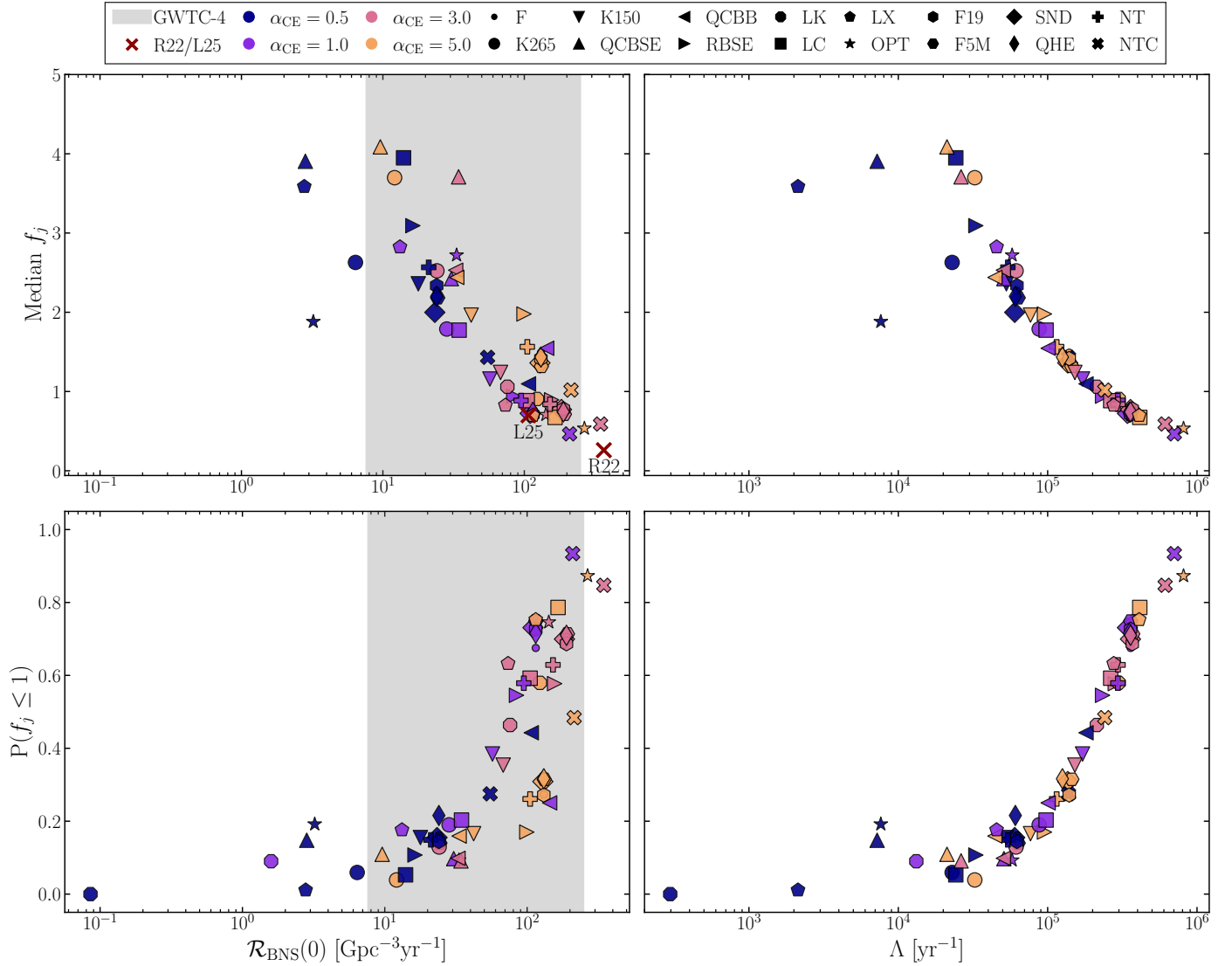


Fig. 2. Jet fraction statistics for the universal structured jet model assuming the GW170817/GRB170817A structure. *Top row:* Median jet fraction f_j versus the predicted local BNS merger rate density $R_{\text{BNS}}(0)$ (left) and total integrated rate Λ (right). *Bottom row:* Quantile of the posterior distribution of f_j at $f_j = 1$ as a function of local BNS rate density (left) and total integrated rate Λ (right). Symbols are coloured according to the α_{CE} parameter, and each symbol denotes a different population. The shaded vertical region indicates the 90% credible interval for the BNS merger rate from GWTC-4 (The LIGO Scientific Collaboration 2025). The cited fractions from the works of Ronchini et al. (2022) and Loffredo et al. (2025) are given as R22 and L25 for comparison.

observations of sGRB jet breaks, suggesting a low median opening angle $\langle \theta_c \rangle \approx 6^\circ$ (Rouco Escorial et al. 2023). While the top-hat model confirms that the geometric degeneracy allows low rates to be compensated for by wide jets, the tight geometry, inferred for GW170817 and applied to the entire BNS population, implies that BNS merger rates must be relatively high to remain physically viable as the sole progenitors of sGRBs.

3.3. Non-universality

We further relaxed the assumption of universality by allowing the jet core angle θ_c in our top-hat jet model to be drawn from a population-wide distribution (uniform and log-normal as defined in Sect. 2.3) for θ_c , treating the distribution's parameters as additional free parameters in the analysis. The posterior distributions from this analysis are shown in Fig. 6 for our fiducial population. We only show the posteriors for the geometry and jet-

fraction as the other parameters are almost identical to Fig. E.3. Both models exhibit a similar degeneracy between the parameters P_θ and f_j . One immediate result is that the log-normal model's extended tail allows a substantial fraction of the f_j posterior to lie below unity, even for small sampled θ_c^{med} angles. In contrast, the flat model places a large portion of f_j posterior above unity for almost all small θ_c^{max} angles. To systematically compare these non-universal results with the universal top-hat model and observational constraints, we introduce a unified geometric metric described below.

3.4. Narrowest geometry

To benchmark all our BNS models against observations, regardless of the assumed top-hat jet model, we quantify here the trade-off between the intrinsic merger rate and the jet geometry. The observed sGRB rate depends on the GRB population beaming

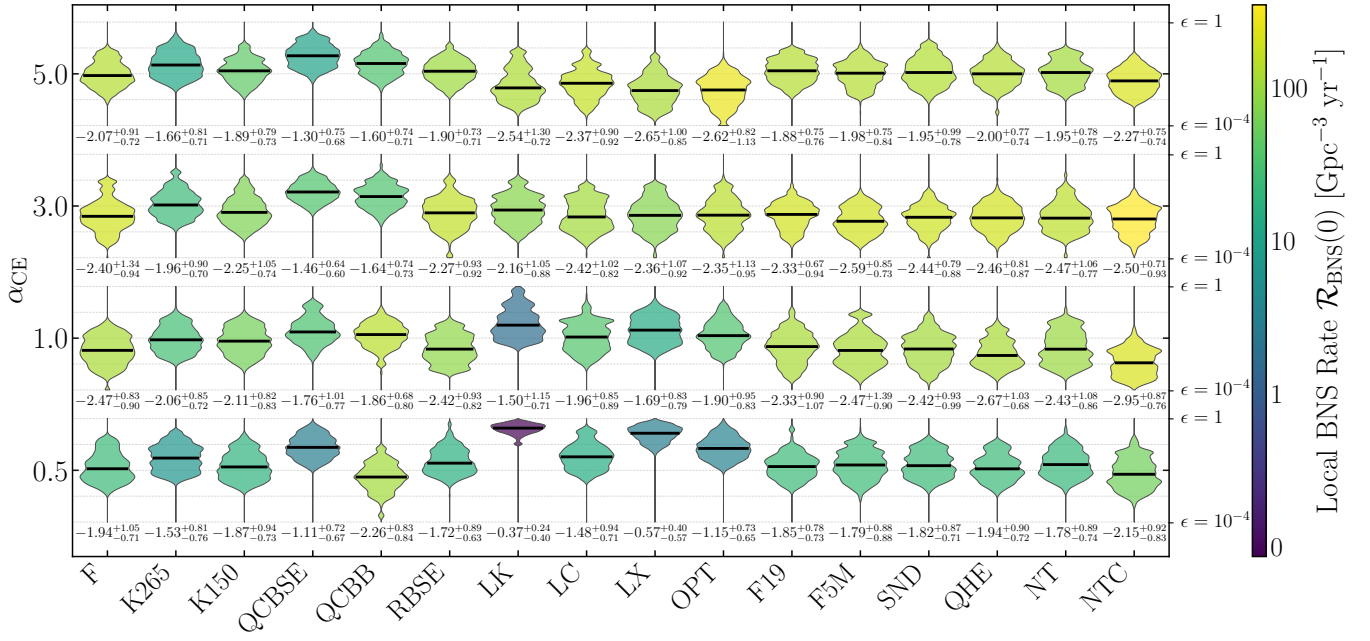


Fig. 3. Same as Fig. 1 but with the one-dimensional marginalised distributions of $\epsilon = f_j(1 - \cos(\theta_c))$ for all the BNS populations, considering the universal top-hat jet model. For clarity $\log(\epsilon)$ is shown.

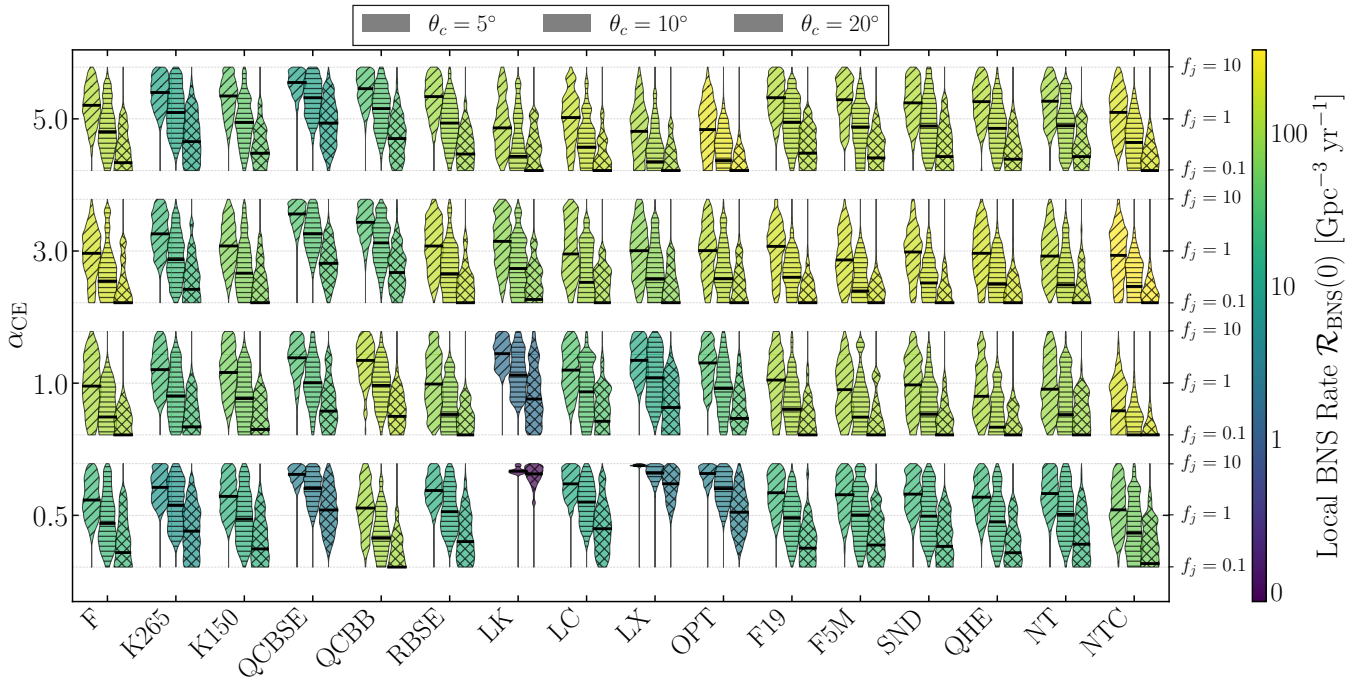


Fig. 4. Same as Fig. 1 but assuming a universal top-hat jet with three different opening angles, $\theta_c = 5^\circ$, 10° , and 20° . These posteriors were obtained from the conditioned distribution $P(f_j | \theta_c)$ (see Fig. 3). The model LK with $\alpha_{CE} = 0.5$ does not have any samples below $f_j \leq 10$ for $\theta_c = 5^\circ$.

factor $\langle f_b \rangle$ and jet fraction f_j such that for a fixed population model (see Sect. 2.2), the predicted detection rate scales as

$$R_{\text{obs}} \propto f_j \langle f_b(P_\theta) \rangle,$$

where P_θ represents the parameters of the jet opening angle distribution (e.g. θ_c for the universal top-hat, θ_c^{max} for the flat model, or θ_c^{med} for the log-normal model). Because the observed rate is fixed by *Fermi*-GBM data, a strict degeneracy exists as

$f_j \times \langle f_b \rangle = \epsilon$, where ϵ is a constant specific to each BNS population. If we impose that f_j takes only physical values (median $f_j \leq 1$), there exists a minimum beaming factor $\langle f_b \rangle_{\text{min}}$ required to reproduce the observed event rate. We translate this into a minimum characteristic opening angle, θ^* , defined as the narrowest population median angle $\tilde{\theta}_c$ that satisfies the physical condition

$$\theta^* = \min_{\tilde{\theta}_c} \{ \tilde{\theta}_c | P(f_j \leq 1) \geq 0.5 \}. \quad (18)$$

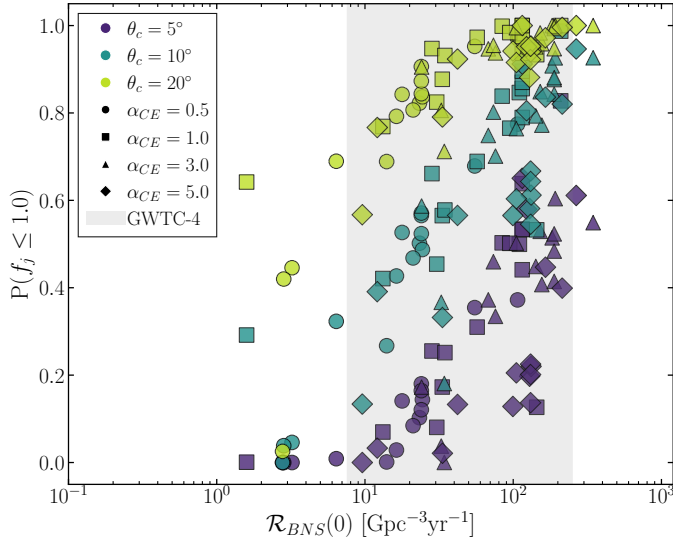


Fig. 5. Same as the bottom row of Fig. 2 but assuming a universal top-hat with three different aperture angles.

We calculated θ^* for three distinct distribution geometries:

1. *Universal top-hat:* $P(\theta_c) \sim \delta(\theta - \theta_c)$ distribution where $\langle f_b \rangle = 1 - \cos \theta_c$ and the median $\tilde{\theta}_c = \theta_c$.
2. *Flat distribution:* $P(\theta_c) \sim \mathcal{U}(\theta_c^{\min}, \theta_c^{\max})$, where

$$\langle f_b \rangle = 1 - \frac{\sin \theta_c^{\max} - \sin \theta_c^{\min}}{\theta_c^{\max} - \theta_c^{\min}}$$

and with a population median $\tilde{\theta}_c = (\theta_c^{\max} + \theta_c^{\min})/2$.

3. *Log-normal distribution:* Defined by a median θ_c^{med} and shape σ , where

$$\langle f_b \rangle = \int (1 - \cos \theta_c) p(\theta_c | \theta_c^{\text{med}}, \sigma) d\theta_c$$

and with $\tilde{\theta}_c = \theta_c^{\text{med}}$.

For each posterior sample (f_j, P_θ) we can calculate $\epsilon = f_j \times \langle f_b \rangle$ and derive the median $\langle \epsilon \rangle$. Numerically we solved $\langle f_b(P_\theta) \rangle - \langle \epsilon \rangle = 0$. From the minimum P_θ we define θ^* from the associated population median $\tilde{\theta}_c$. An example on how θ^* is recovered for the universal top-hat model in the case of our fiducial population is shown in Fig. G.1. The figure shows how the $f_j - \theta_c$ posterior follows the predicted line at $\langle \epsilon \rangle = \langle f_j \rangle (1 - \cos(\theta_c)) = \text{const.}$ We then calculated the θ^* value for each of the 64 populations across all three geometric assumptions as shown in Fig. 7 and compared these values with the jet opening angles inferred from sGRB afterglow observations (Rouco Escorial et al. 2023). Under the universal top-hat assumption (Fig. 7, top left panel), BNS population models predicting lower intrinsic merger rates require significantly wider θ^* values to compensate for the lower event count. Models with $R_{\text{BNS}}(0) \approx 100 \text{ Gpc}^{-3} \text{ yr}^{-1}$ yield θ^* values close to the observed median of $\langle \theta_c \rangle \approx 6^\circ$ reported in Rouco Escorial et al. (2023). As the authors of Rouco Escorial et al. (2023) mention, we are aware that estimates of $\langle \theta_c \rangle$ might be biased by selection effects (see the many lower limits set therein) we nonetheless use it as a reference point, while showing the trend of opening angles compared to local rates. In contrast, lower-rate models are forced to assume geometries significantly wider than typical observed values. This trend persists for the non-universal *Flat* model (Fig. 7,

top right panel). As merger rates decrease, the required population median θ_c moves beyond the 1σ upper bound of afterglow observations. For some populations, the local rate is so low that no numerical value of θ^* satisfies the condition of Eq. (18), since the required value would lie above the maximum value of the prior $\theta_c^{\max}$, effectively excluding these populations as viable progenitors under this assumption. To remain compatible with the narrow geometries inferred from X-ray observations ($\theta_c \leq 10^\circ$), the underlying BNS population must possess a local rate $R_{\text{BNS}}(0) \geq 50 \text{ Gpc}^{-3} \text{ yr}^{-1}$. The non-universal *Log-Normal* model (Fig. 7, bottom left panel) offers more flexibility due to its extended tail, allowing lower-rate BNS models to have a median $f_j \leq 1$ by populating the distribution with wide-angle jets. However, this comes at the cost of physical plausibility. As shown in the bottom right panel of Fig. 7, for models with $R_{\text{BNS}}(0) \lesssim 20 \text{ Gpc}^{-3} \text{ yr}^{-1}$, more than $\sim 10\%$ of the sGRB population would require core angles $\theta_c \geq 15^\circ$ to maintain a physical jet fraction. This result, under the log-normal assumption, highlights the presence of a subgroup of BNS populations that, in order to have a f_j distribution with a median below one, require a distribution of aperture angles containing a substantial fraction that stands in tension with current observational constraints on sGRB jet collimation.

A comprehensive summary of the physical viability and geometric consistency for all 64 BNS population models across the four analysed jet scenarios is provided in Table H.1. The specific criteria used to evaluate these models, including the treatment of non-physical jet fractions and observational tensions, are further detailed in Appendix H.

4. Discussion

Our analysis demonstrates that sGRB observations provide powerful independent constraints on BNS models. By calibrating the sGRB/BNS ratio (f_j) to the observed *Fermi*-GBM rate, we identify a clear tension between low-rate progenitor models and electromagnetic observations in terms of sGRB local rate, its evolution, and the inferred jet opening angle. In the following, we summarise and discuss the results obtained for the various jet models considered.

- Under the assumption of a universal structured jet consistent with GW170817/GRB170817A, BNS population models predicting $R_{\text{BNS}}(0) \lesssim 50 \text{ Gpc}^{-3} \text{ yr}^{-1}$ are effectively ruled out, because the number of BNS mergers is insufficient to account for the full population of observed sGRBs (the inferred median is $f_j > 1$). BNS population models with a local merger rate of $R_{\text{BNS}}(0) \approx 100 \text{ Gpc}^{-3} \text{ yr}^{-1}$, consistent with the current GWTC-4 constraints, are able to reproduce the observed sGRB population provided that the jet fractions are $f_j \approx 0.7 - 0.8$.
- The use of a universal top-hat model allowed us to quantify the degeneracy between merger rates and jet geometry. We find that, in order to maintain physically plausible jet fractions, low-rate models are forced to adopt opening angles $\theta_c \gtrsim 15^\circ - 20^\circ$, which are in significant tension with the median value of $\gtrsim 6^\circ$, inferred from the observation of jet break in afterglow light curves of sGRBs. In contrast, populations with $R_{\text{BNS}}(0) \approx 100 \text{ Gpc}^{-3} \text{ yr}^{-1}$ lie remarkably close to the observed median, allowing for physically plausible jet fractions without requiring wide-angle geometries.
- This conclusion remains robust when relaxing the assumption of universality by adopting a flat or log-normal distributions for the jet core opening angle. These non-universal jet models show that low-rate BNS population scenarios can

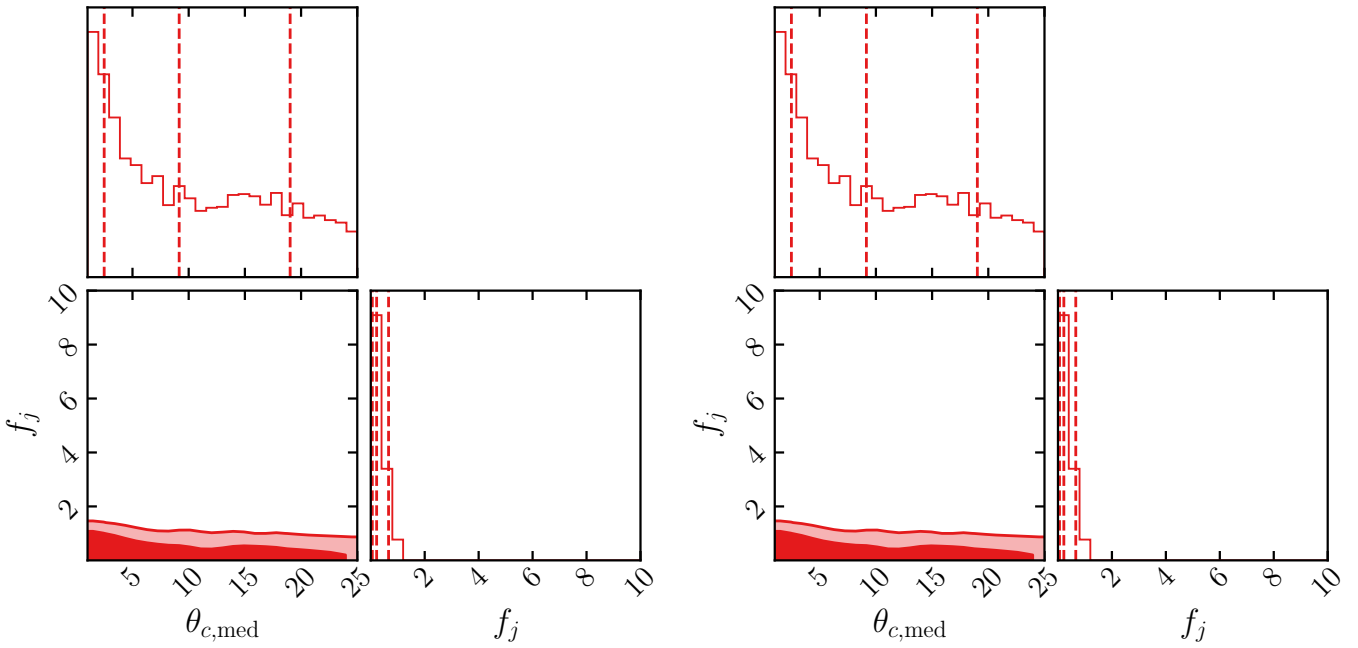


Fig. 6. Two-dimensional marginal posterior distribution $P(f_j, \theta_{c,\max})$ for the non-universal flat model (left) and $P(f_j, \theta_{c,\text{med}})$ for the non-universal log-normal model (right) using our fiducial BNS population. The remaining parameters remain consistent with the results of a universal top-hat (see Fig. E.3).

reproduce the observed sGRB counts by forcing the jet opening angle distribution towards systematically wider jets. For instance, a low-merger-rate BNS population assuming a log-normal jet model requires more than 10-20% of the systems to have core angles $\theta_c > 15^\circ$. Such a requirement is not supported by the scarcity of wide-jet observations in sGRB catalogues.

Our finding of physically viable BNS populations capable of reproducing sGRB observations which tend to favour jet-fraction close to unity, is in good agreement with other recent population studies that have taken different approaches to the problem, such as the work by [Salafia et al. \(2023\)](#), which utilised a more granular treatment of detector efficiency and inferred local sGRB rates above $100 \text{ Gpc}^{-3} \text{ yr}^{-1}$, implying a high jet production efficiency. Constraining the jet fraction, has physical implications beyond binary evolution. The ability of a BNS merger remnant to launch a relativistic jet is thought to depend critically on the nature of the central engine ([Ciolfi 2020](#)), which is in turn governed by the binary's total mass and the neutron star equation of state (EoS; e.g. [Giacomazzo & Perna 2013](#); [Ruiz et al. 2021](#)). Our finding of high jet production efficiency disfavours scenarios where prompt collapse to a black hole systematically suppresses jet production or alternatively implies that even prompt-collapse systems can launch jets efficiently.

We emphasise that the results of our analysis, shown in the different figures, provide a framework for assessing the validity of BNS merger population models in light of the progressively improving observational constraints expected from GW observations and complementary electromagnetic measurements of sGRBs. However, we can already conclude that if future GW observations continue to push the BNS rate towards the lower rates, the discrepancy between the required wide jets and the observed narrow ones ([Rouco Escorial et al. 2023](#)) would suggest that either BNS mergers are not the sole progenitors of sGRBs or our understanding of jet observational biases is incomplete.

4.1. Intrinsic local sGRB rate densities

Using our theoretical and statistical framework, we examine the intrinsic local sGRB rate densities (regardless of the viewing angles) inferred from our models and compare them with values reported in the literature. For each BNS population, we define the rate of successfully launched sGRB jets as $R_{\text{sGRB}} = R_{\text{BNS}}(0) \times f_j$, where $R_{\text{BNS}}(0)$ is the local BNS merger rate predicted by population synthesis and f_j is the jet fraction inferred via our analysis. This value represents the all-sky intrinsic local density of sGRB jets produced per year. The resulting distributions for the intrinsic local sGRB rate density, assuming our universal structured jet model, are presented in Fig. 8 (left panel). We contextualise these results by comparing them to recent derived constraints. [Salafia et al. \(2023, S23\)](#) estimated rates using a flux-limited sample similar to ours, finding $R_{\text{sGRB}} \approx 180_{-145}^{+660} \text{ Gpc}^{-3} \text{ yr}^{-1}$, while [Rouco Escorial et al. \(2023, RE23\)](#) report a rate $R_{\text{sGRB}} \approx 1786_{-1507}^{+6346} \text{ Gpc}^{-3} \text{ yr}^{-1}$ ($R_{\text{sGRB}} \approx 361_{-217}^{+4367} \text{ Gpc}^{-3} \text{ yr}^{-1}$ for the mock sGRB sample) based on the observed local rate and the jet opening angles determined from the afterglows. We observe that our inferred rates for physically viable populations (specifically those where their median is $f_j \leq 1$) generally sit at the lower end or below the median estimates of [S23](#) and [RE23](#). The higher intrinsic local sGRB rates favoured by other studies would increase the tension with the hypothesis that BNS mergers are the sole progenitors of sGRBs, as they would require non-physical jet fractions (median $f_j > 1$). The local sGRB rates are inferred from sGRBs observed out to high redshift, and this tension may also indicate that the redshift distributions assumed in those studies do not accurately reflect the redshift evolution of BNS mergers.

The sensitivity of our conclusions to the assumed jet geometry is illustrated in Fig. 8 (right panel), where we show the f_j posterior distributions for our fiducial BNS population across the four analysed jet models. While the universal structured and top-hat models yield comparable constraints, the non-universal

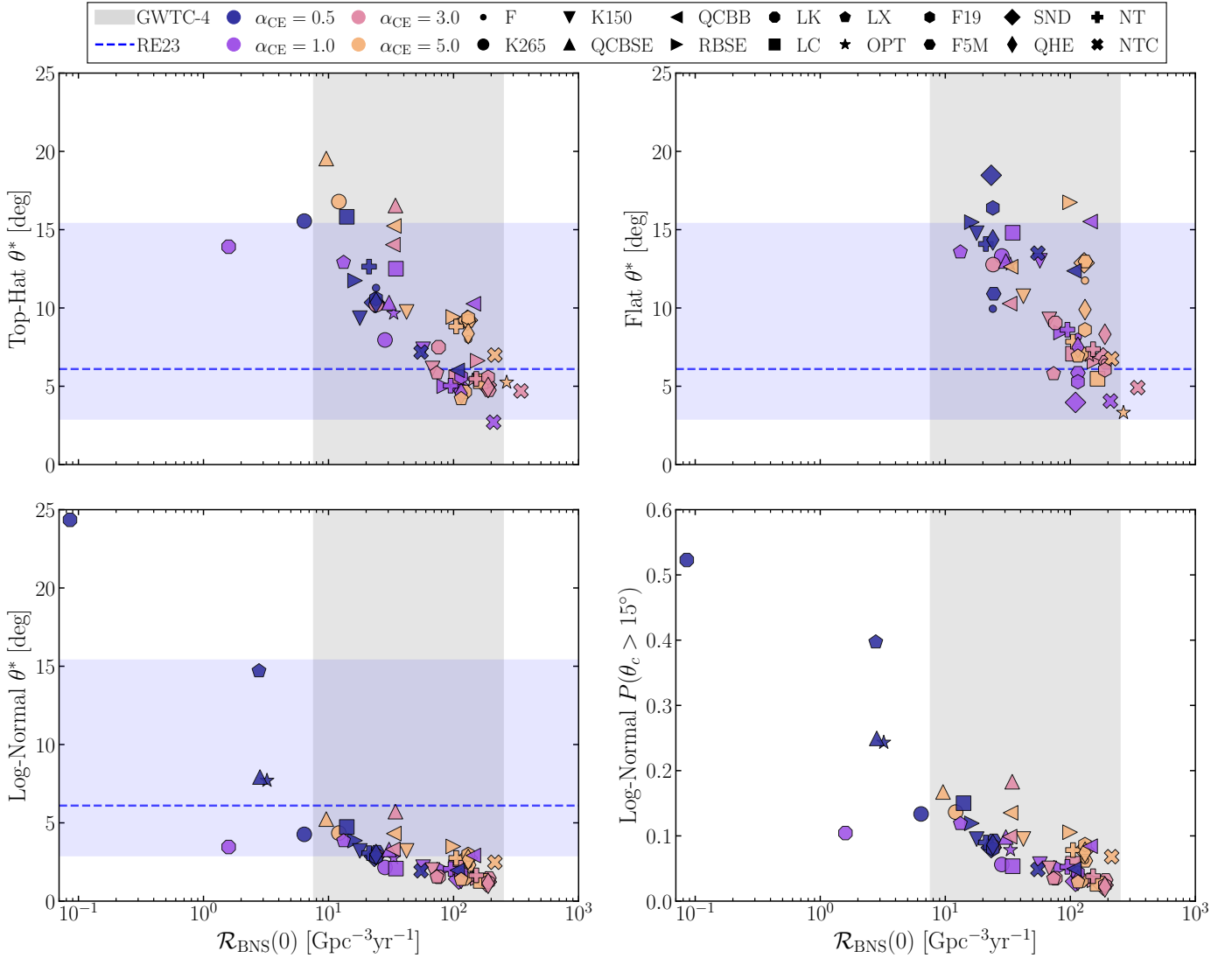


Fig. 7. Minimum characteristic jet opening angle, θ^* , required to keep the median jet fraction physical (median $f_j \leq 1$) as a function of the local BNS merger rate $R_{\text{BNS}}(0)$ for the universal top-hat (top left), non-universal flat (top right), and non-universal log-normal (bottom left). Each symbol denotes a different population. The horizontal dashed line and shaded region indicate the median and 90% credible interval of the aperture angle distribution derived from Rouco Escorial et al. (2023). Bottom right panel: Fraction of the population with wide jets ($\theta_c > 15^\circ$) for the log-normal model. The vertical grey band represents the GWTC-4 90% C.I. for the BNS merger rate. The plots do not display all 64 analysed populations, as some require θ^* larger than the prior upper limit of 25° . See Table H.1 for a comprehensive list of all the populations.

models diverge significantly based on the shape of their opening angle distribution. Notably, the log-normal model yields a posterior distribution located almost entirely below $f_j = 1$. This demonstrates that a geometric distribution with a long tail towards wide opening angles can effectively alleviate the tension for lower-rate models. The presence of a sub-population of wide jets increases the average beaming factor, thereby reducing the total number of progenitors required to match the observed *Fermi*-GBM rate. Conversely, the flat model results in a posterior centred well above unity, indicating a much stronger tension. This confirms that for most angles the MCMC is unable to infer physical jet fractions, even BNS models with moderate merger rates may fail to reproduce the observed event counts without non-physical efficiencies. Since $R_{\text{sGRB}} \propto f_j$, Fig. 8 (right panel) can be used to estimate what is the impact of the assumption of jet structure on the inferred sGRB local rate.

4.2. Delay-time distribution of BNS populations

Here, we analyse the distributions of delay times between the formation and the merger for BNSs, to compare them with the recent findings of Pracchia & Salafia (2026) in Fig. 9. The delay times are calculated starting from the catalogues of Iorio et al. (2023) and using the code COSMORATE (see Sect. 2.5). While early studies suggested long delay times (a few gigayears) for sGRBs, Pracchia & Salafia (2026) demonstrated that correcting for selection effects in flux-incomplete samples leads to significantly shorter delays, with average values $\langle \tau_d \rangle \approx 10 - 800$ Myr. Figure 9 show the average time delay corresponding to each of our BNS populations. All models but one (QCB with $\alpha_{\text{CE}} = 1$) exhibit short delay times ($\langle \tau_d \rangle < 1$ Gyr) in agreement with the 90% credible intervals reported by Pracchia & Salafia (2026). We also clarify how our time delay average values are derived directly from the BNS population synthesis catalogues and are

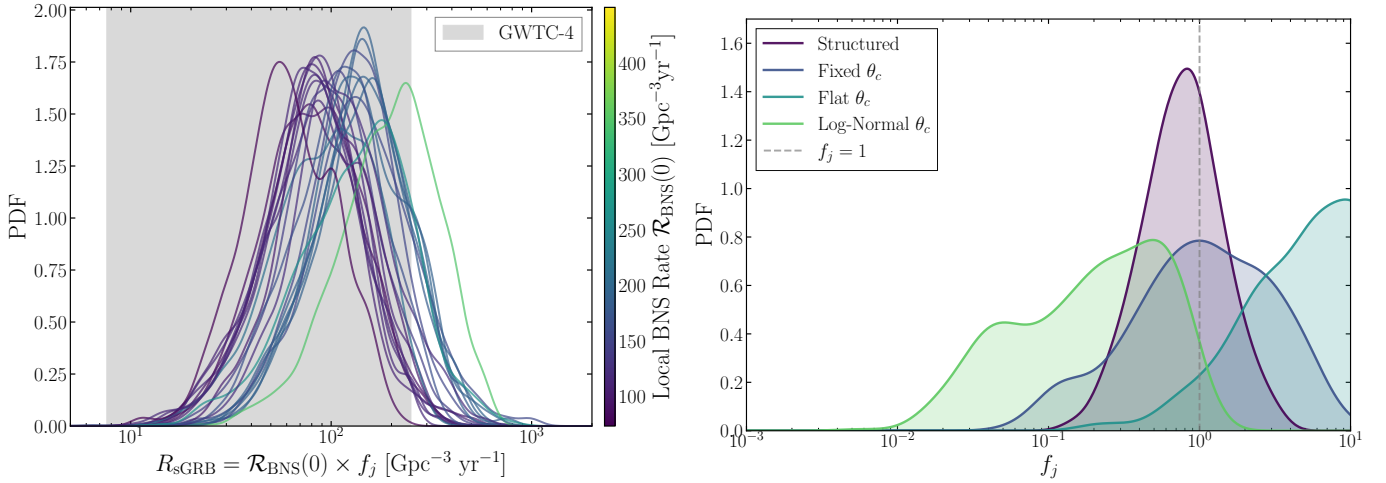


Fig. 8. *Left:* Inferred intrinsic local sGRB rate R_{sGRB} for all the BNS population models that allow physical jet fractions (with median $f_j \leq 1$), under the assumption of a universal structured jet. *Right:* Comparison of the inferred f_j posterior distributions for the fiducial BNS population across the four analysed jet structures: universal structured, universal top-hat, non-universal flat, and non-universal log-normal.

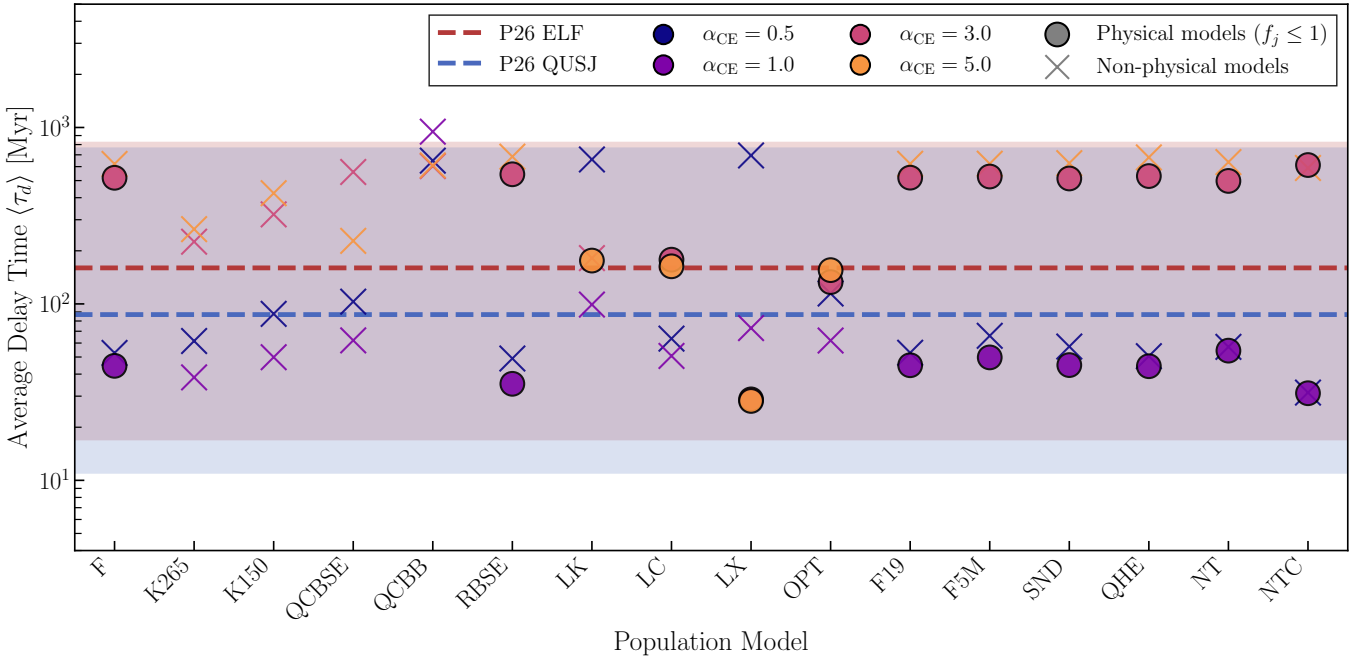


Fig. 9. Average delay time $\langle \tau_d \rangle$ for each of the 64 BNS population models compared to the median and 90% credible intervals inferred in Pracchia & Salafia (2026). The blue and red lines and ranges correspond to two models by Pracchia & Salafia (2026): a quasi-universal structured jet model and an empirical luminosity function model, respectively. To highlight the relationship between delay times and model viability, we mark with circles the populations that yield a physical jet fraction (with median $f_j \leq 1$) under the universal structured jet assumption and use crosses for those that are considered non-physical.

independent of our sGRB inference. In Appendix I we also show the redshift evolution of the average delay time.

4.3. Caveats

Progenitor purity. A central goal of our analysis is to test the hypothesis that BNS mergers are sufficient to explain the entire observed sGRB population. We use the inferred jet fraction as a diagnostic for this test. Any scenario that requires more SGRBs than BNS progenitors ($f_j > 1$) is interpreted as evidence against the BNS-unique progenitor hypothesis, implying that the BNS population under consideration is not numerous enough

to account for the observed sGRB rate. This could suggest a contribution from an alternative channel, such as NSBH mergers (Colombo et al. 2025). However, the association between NSBH mergers and sGRBs has not yet been firmly established observationally and remains more difficult to quantify. The outcome of NSBH mergers is highly model-dependent and sensitive to the mass ratio, the black hole spin, and the neutron star EoS (Foucart 2012; Krüger & Foucart 2020). Numerical simulations and analytical fits suggest that only a subset of the NSBH population, specifically those with high BH spins and/or low mass ratios, yields enough remnant mass to power a jet. These stringent requirements often result in a majority of systems

failing to launch a jet (Sarin et al. 2022; Clarke et al. 2025; Biscoveanu et al. 2023), implying smaller $f_j \ll 1$ for the NSBH population. Regarding the population abundance, while some synthesis models predict NSBH rates to be lower than BNS rates by an order of magnitude (Iorio et al. 2023), current observational constraints from GWTC-4 indicate that the two rates may be comparable. Specifically, GWTC-4 reports a BNS merger rate of $7.6\text{--}250 \text{ Gpc}^{-3} \text{ yr}^{-1}$ (90% C.I.) and an NSBH merger rate of $9.1\text{--}84 \text{ Gpc}^{-3} \text{ yr}^{-1}$. Despite this, because only a small fraction of the NSBH population is expected to produce significant ejected mass and a subsequent jet, we assume that their contribution would not significantly alter the inferred jet fractions presented in this work especially for comparatively higher rate models. We also acknowledge that the traditional classification distinguishing long and short GRBs has blurred by recent discoveries of long-duration bursts with kilonova counterparts, such as GRB 211211A and GRB 230307A (Rastinejad et al. 2022; Levan et al. 2023; Mei et al. 2022; Troja et al. 2022). This suggests that some compact object mergers may be classified as long ($T_{90} > 2 \text{ s}$) and thus excluded from our sample. While a detailed treatment of this contamination is beyond the scope of this work, it underscores the need for further observations and improved classification criteria to quantify it.

Emission model. For the structured jet we model the sGRB pulse with a simplified temporal profile (linear rise, power-law decay) and a fixed spectral shape (SBPL with fixed indices). While physically motivated, this does not capture the full diversity and complexity of observed GRB light curves, which often show multiple pulses. This simplification primarily affects the goodness-of-fit for the T_{90} and fluence distributions. While our likelihood is robust to some of this variation, more sophisticated, time-resolved emission models could provide a more detailed fit to the data. Nevertheless, because our analysis is calibrated to observational data, we expect this to have a negligible impact on our conclusions.

Systematic uncertainties. The analysis is subject to systematic uncertainties stemming from observational data and instrumental effects. The *Fermi*-GBM detection efficiency and sky exposure (ϵ_{GBM}) are approximations, and the measurement errors for catalogued observables like E_p are not fully propagated in our comparison. Given the large quantity of data spanning over a decade of measurements this is secondary to the major uncertainties in population synthesis and jet modelling. Our predictions of the local sGRB are compatible with Salafia et al. (2023), where the authors performed a similar analysis leveraging a more sophisticated treatment of the detector's response. Even with these additional considerations they infer local sGRB rates that are in tension with the bounds from GWTC-4, showing that lower BNS merger rates are disfavoured regardless of the level of refinement in instrument response modelling.

5. Conclusions

In this work, we have presented a comprehensive multi-messenger framework connecting 64 state-of-the-art BNS population synthesis models to the 16-year sGRB archive from *Fermi*-GBM. By treating the jet launching fraction, f_j , as a free parameter inferred directly from observations, we systematically evaluated the viability of BNS populations under three distinct jet structure assumptions: a universal structured jet calibrated to

GW170817, a universal top-hat jet, and non-universal populations with diverse opening angles.

Our analysis identified a significant tension between BNS population models predicting low local merger rates ($R_{BNS}(0) \lesssim 50 \text{ Gpc}^{-3} \text{ yr}^{-1}$) and the observed sGRB event count. Under the assumption of a GW170817-like structured jet, these models consistently require non-physical jet fractions (median $f_j > 1$) to match observations, implying that the underlying BNS population is insufficient to serve as the sole progenitor of sGRBs. This effectively establishes a lower bound for the merger rate derived purely from electromagnetic observations.

We further demonstrated that this tension cannot be resolved by simply invoking different jet geometries. Using both universal and non-universal top-hat models, we found that low-rate populations can only reproduce the observed sGRB rate if the population is dominated by wide jets ($\theta_c \gtrsim 15^\circ$). This requirement is very difficult to reconcile with the narrow core angles ($\langle \theta_c \rangle \approx 6^\circ$) inferred from sGRB afterglows and constraints from GRB 170817A (Ghirlanda et al. 2016; Rouco Escorial et al. 2023). Even when allowing for a log-normal distribution of opening angles, low-rate models are forced to populate the distribution's tail with a fraction of wide jets that is incompatible with current sGRB observational limits.

Consequently, our analysis favours BNS models with local merger rates of $R_{BNS}(0)$ around $100 \text{ Gpc}^{-3} \text{ yr}^{-1}$. In the context of binary evolution parameters, this preference points towards models assuming standard or high CE efficiencies ($\alpha_{CE} \geq 1$) and moderate natal kicks, which are necessary to maintain a sufficient population of merging binaries. Conversely, models characterised by low ejection efficiencies ($\alpha_{CE} \leq 0.5$) or high natal kicks (e.g. $\sigma = 265 \text{ km/s}$) are disfavoured, as they suppress the merger rate below the one required by electromagnetic observations (see Appendices H and D for specific models). The favoured populations naturally reproduce the spectral and temporal properties of *Fermi*-GBM sGRBs with physically plausible jet fractions ($f_j \approx 0.8$) and opening angle distributions consistent with sGRB afterglow observations. The inference of relatively high jet production efficiencies suggests that the conditions for launching a relativistic jet are satisfied in the majority of BNS mergers, disavouring scenarios where prompt collapse to a black hole systematically suppresses jet formation.

In conclusion, this study highlights the importance of combining GW and electromagnetic observations to constrain the physics of binary evolution. While current GW measurements primarily set an upper bound on the BNS merger rate, sGRB population statistics provide a complementary lower bound. As future GW observing runs further tighten the constraints on the local BNS rate density, the allowed parameter space will progressively shrink, enabling the relation between the jet fraction and the merger rate to become a precise probe of the physical conditions governing BNS mergers.

Acknowledgements. The authors thank Gor Oganessian, Biswajit Banerjee, and Matteo Pracchia for insightful discussions and constructive comments that helped improve this manuscript. M.B. and S.R. acknowledge support from the Astrophysics Center for Multi-messenger Studies in Europe (ACME), funded under the European Union's Horizon Europe Research and Innovation Program, Grant Agreement No. 101131928. F.S. has been funded by the European Union – NextGenerationEU under the Italian Ministry of University and Research (MUR) – CUP D13C25000700001. F.S. has been funded by the European Union – NextGenerationEU under the Italian Ministry of University and Research (MUR) ‘Decreto per l’assunzione di ricercatori internazionali post-dottorato PNRR’ – Missione 4 ‘Istruzione e Ricerca’ Componente 2 ‘Dalla Ricerca all’Impresa’ del PNRR – Investimento 1.2 ‘Finanziamento di progetti presentati da giovani ricercatori’ – CUP D13C25000700001.

References

- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017a, *ApJ*, **848**, L13
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017b, *Phys. Rev. Lett.*, **119**, 161101
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017c, *ApJ*, **848**, L13
- Ascenzi, S., Oganessian, G., Salafia, O. S., et al. 2020, *A&A*, **641**, A61
- Bamber, J., Ruiz, M., Shapiro, S., & Tsokaros, A. 2024, in *APS April Meeting Abstracts*, APS Meeting Abstracts, 2024, D05.001
- Belczynski, K., Askar, A., Arca-Sedda, M., et al. 2018, *A&A*, **615**, A91
- Berger, E. 2014, *ARA&A*, **52**, 43
- Bhattacharjee, S., Banerjee, S., Bhalerao, V., et al. 2024, *MNRAS*, **528**, 4255
- Biscoveanu, S., Burns, E., Landry, P., & Vitale, S. 2023, *RNAAS*, **7**, 136
- Blinnikov, S. I., Novikov, I. D., Perevodchikova, T. V., & Polnarev, A. G. 1984, *SvAL*, **10**, 177
- Broekgaarden, F. S., Berger, E., Stevenson, S., et al. 2022, *MNRAS*, **516**, 5737
- Bromberg, O., Nakar, E., Piran, T., & Sari, R. 2013, *ApJ*, **764**, 179
- Burns, E., Connaughton, V., Zhang, B.-B., et al. 2016, *ApJ*, **818**, 110
- Ciolfi, R. 2020, *MNRAS*, **495**, L66
- Clarke, T. A., Lasky, P. D., & Thrane, E. 2025, *ApJ*, **984**, 27
- Colombo, A., Salafia, O. S., Ghirlanda, G., et al. 2025, *A&A*, **704**, A260
- D’Agostino, R. B., & Stephens, M. A. 1986, *Goodness-of-Fit Techniques*, 1st edn. (New York: Routledge)
- D’Avanzo, P., Campana, S., Salafia, O. S., et al. 2018, *A&A*, **613**, L1
- Dominik, M., Belczynski, K., Fryer, C., et al. 2012, *AJ*, **759**, 52
- Du, Y. F., Yorgancıoglu, E. S., Yi, S. X., Cao, T. Y., & Zhang, S. N. 2025, *MNRAS*, **541**, 798
- Eichler, D., Livio, M., Piran, T., & Schramm, D. N. 1989, *Nature*, **340**, 126
- Farmer, R., Renzo, M., de Mink, S. E., Marchant, P., & Justham, S. 2019, *AJ*, **887**, 53
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, **125**, 306
- Foucart, F. 2012, *Phys. Rev. D*, **86**, 124007
- Ghirlanda, G., Salafia, O. S., Pescalli, A., et al. 2016, *A&A*, **594**, A84
- Ghirlanda, G., Salafia, O. S., Paragi, Z., et al. 2019, *Sci*, **363**, 968
- Giacobbo, N., & Mapelli, M. 2018, *MNRAS*, **480**, 2011
- Giacobbo, N., & Mapelli, M. 2020, *ApJ*, **891**, 141
- Giacomazzo, B., & Perna, R. 2013, *ApJ*, **771**, L26
- Giudice, I. F., Izzo, L., Martone, R., et al. 2025, *A&A*, **701**, A83
- Goldstein, A., Veres, P., Burns, E., et al. 2017, *ApJ*, **848**, L14
- Gottlieb, O., Nakar, E., Piran, T., & Hotokezaka, K. 2018, *MNRAS*, **479**, 588
- Gruber, D., Goldstein, A., Weller von Ahlefeld, V., et al. 2014, *ApJS*, **211**, 12
- Hayashi, K., Kiuchi, K., Kyutoku, K., Sekiguchi, Y., & Shibata, M. 2025, *Phys. Rev. Lett.*, **134**, 211407
- Hobbs, G., Lorimer, D. R., Lyne, A. G., & Kramer, M. 2005, *MNRAS*, **360**, 974
- Howell, E. J., Ackley, K., Rowlinson, A., & Coward, D. 2019, *MNRAS*, **485**, 1435
- Hurley, J. R., Tout, C. A., & Pols, O. R. 2002, *MNRAS*, **329**, 897
- Ierardi, A., Oganessian, G., Ascenzi, S., et al. 2026, *A&A*, **708**, A190
- Iorio, G., Mapelli, M., Costa, G., et al. 2023, *MNRAS*, **524**, 426
- Ivanova, N., Justham, S., Chen, X., et al. 2013, *A&A Rev.*, **21**, 59
- Kouveliotou, C., Meegan, C. A., Fishman, G. J., et al. 1993, *ApJ*, **413**, L101
- Kroupa, P. 2001, *MNRAS*, **322**, 231
- Krüger, C. J., & Foucart, F. 2020, *Phys. Rev. D*, **101**, 103002
- Lamb, G. P., Lyman, J. D., Levan, A. J., et al. 2019, *ApJ*, **870**, L15
- Lazzati, D., Perna, R., Morsony, B. J., et al. 2018, *Phys. Rev. Lett.*, **120**, 241103
- Levan, A. J., Gompertz, B. P., Malesani, D. B., et al. 2023, *GRB Coordinates Network*, **33569**, 1
- Lien, A., Sakamoto, T., Barthelmy, S. D., et al. 2016, *ApJ*, **829**, 7
- Livio, M., & Soker, N. 1988, *ApJ*, **329**, 764
- Loffredo, E., Hazra, N., Dupletsa, U., et al. 2025, *A&A*, **697**, A36
- Madau, P., & Fragos, T. 2017, *ApJ*, **840**, 39
- Mandel, I., Müller, B., Riley, J., et al. 2020, *MNRAS*, **500**, 1380
- Mapelli, M., Giacobbo, N., Santoliquido, F., & Bressan, A. 2020, *AJ*, **888**, 76
- Margutti, R., Alexander, K. D., Xie, X., et al. 2018, *ApJ*, **856**
- Matsumoto, T., Nakar, E., & Piran, T. 2019, *MNRAS*, **486**, 1563
- Mazets, E. P., & Golenetskii, S. V. 1981, *Ap&SS*, **75**, 47
- Mei, A., Banerjee, B., Oganessian, G., et al. 2022, *Nature*, **612**, 236
- Mochkovitch, R., Hernanz, M., Isern, J., & Martin, X. 1993, *Nature*, **361**, 236
- Mooley, K. P., Deller, A. T., Gottlieb, O., et al. 2018, *Nature*, **561**, 355
- Nakar, E. 2007, *Phys. Rep.*, **442**, 166
- Narayan, R., Paczynski, B., & Piran, T. 1992, *ApJ*, **395**, L83
- O’Connor, B., Beniamini, P., & Gill, R. 2024, *MNRAS*, **533**, 1629
- Paciesas, W. S., Meegan, C. A., Pendleton, G. N., et al. 1999, *ApJS*, **122**, 465
- Pavan, A., Ciolfi, R., Kalinani, J. V., & Mignone, A. 2023, *MNRAS*, **524**, 260
- Pavan, A., Ciolfi, R., Dreas, E., & Kalinani, J. V. 2025, *MNRAS*, **540**, 1345
- Pescalli, A., Ghirlanda, G., Salvaterra, R., et al. 2016, *A&A*, **587**, A40
- Planck Collaboration VI. 2020, *A&A*, **641**, A6, [Erratum: *A&A* 652, C4 (2021)]
- Poolakkil, S., Preece, R., Fletcher, C., et al. 2021, *ApJ*, **913**, 60
- Pracchia, M., & Salafia, O. S. 2026, arXiv e-prints [arXiv:2601.03861]
- Rastinejad, J. C., Gompertz, B. P., Levan, A. J., et al. 2022, *Nature*, **612**, 223
- Ravasio, M. E., Oganessian, G., Ghirlanda, G., et al. 2018, *A&A*, **613**, A16
- Ravasio, M. E., Ghirlanda, G., Nava, L., & Ghisellini, G. 2019, *A&A*, **625**, A60
- Ronchini, S., Branchesi, M., Oganessian, G., et al. 2022, *A&A*, **667**, A168
- Rossi, E., Lazzati, D., & Rees, M. J. 2002, *MNRAS*, **332**, 945
- Rouco Escorial, A., Fong, W., Berger, E., et al. 2023, *ApJ*, **959**, 13
- Rudolph, A., Tamborra, I., & Gottlieb, O. 2024, *ApJ*, **961**, L7
- Ruiz, M., Tsokaros, A., & Shapiro, S. L. 2021, *Phys. Rev. D*, **104**, 124049
- Salafia, O. S., Colombo, A., Gabrielli, F., & Mandel, I. 2022, *A&A*, **666**, A174
- Salafia, O. S., Ravasio, M. E., Ghirlanda, G., & Mandel, I. 2023, *A&A*, **680**, A45
- Santoliquido, F., Mapelli, M., Bouffanais, Y., et al. 2020, *ApJ*, **898**, 152
- Santoliquido, F., Mapelli, M., Giacobbo, N., Bouffanais, Y., & Artale, M. C. 2021, *MNRAS*, **502**, 4877
- Sarin, N., Lasky, P. D., Vivanco, F. H., et al. 2022, *Phys. Rev. D*, **105**, 083004
- Savchenko, V., Ferrigno, C., Kuulkers, E., et al. 2017, *ApJ*, **848**, L15
- Shibata, M., & Hotokezaka, K. 2019, *Annu. Rev. Nucl. Part. Sci.*, **69**, 41
- Spera, M., Mapelli, M., Giacobbo, N., et al. 2019, *MNRAS*, **485**, 889
- Stephens, M. A. 1974, *J. Am. Stat. Assoc.*, **69**, 730
- The LIGO Scientific Collaboration, the Virgo Collaboration, theKAGRA Collaboration, et al. 2023, *Phys. Rev. X*, **13**, 011048
- The LIGO Scientific Collaboration, the Virgo Collaboration, & the KAGRA Collaboration. 2025, arXiv e-prints [arXiv:2508.18083]
- Troja, E., Piro, L., Ryan, G., et al. 2018, *MNRAS*, **478**, L18
- Troja, E., Fryer, C. L., O’Connor, B., et al. 2022, *Nature*, **612**, 228
- von Kienlin, A., Meegan, C. A., Paciesas, W. S., et al. 2020, *AJ*, **893**, 49
- Wanderman, D., & Piran, T. 2015, *MNRAS*, **448**, 3026
- Webbink, R. F. 1984, *ApJ*, **277**, 355
- Xu, X.-J., & Li, X.-D. 2010, *ApJ*, **722**, 1985
- Zhang, B., Zhang, B.-B., Virgili, F. J., et al. 2009, *ApJ*, **703**, 1696

Appendix A: GRB modelling

In this appendix we provide supplementary details regarding the sGRB emission model and the statistical validation of our results. Figure A.1 compares different spectral models, highlighting the SBPL used in this work, which avoids the exponential cut-off of the Comptonised model while providing a smoother transition than the Band function. Figure A.2 depicts the qualitative time evolution of the peak energy and light curve for different viewing angles, showing how inclination primarily affects the post-peak decay slope. Finally, regarding the non-universal jet models discussed in Sect. 3.3, we visualise the probability density functions used for the jet core opening angle θ_c in Fig. A.3. The analysis samples the characteristic parameter (P_θ) for these distributions (either $\theta_c^{\max}$ or θ_c^{med}).

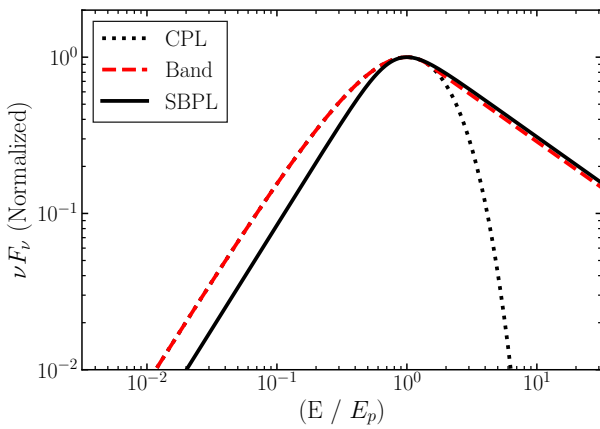


Fig. A.1. Comparison between different spectral models, normalised to 1 at their maxima. The low and high energy index used is $\alpha = -2/3$ and $\beta = -2.59$ for all models.

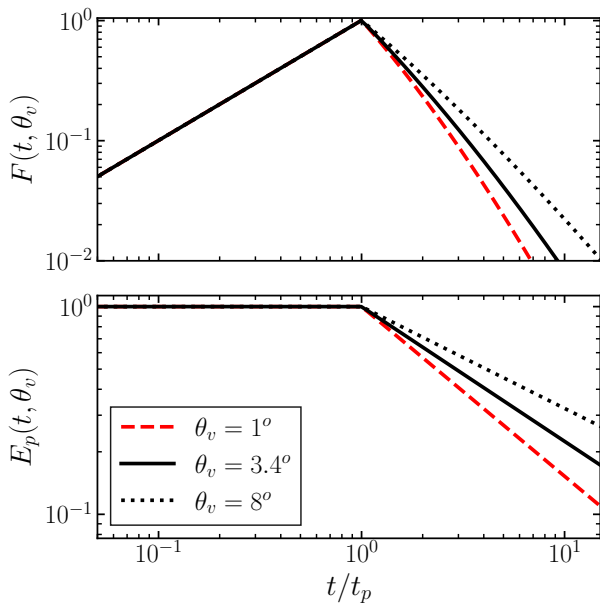


Fig. A.2. Qualitative behaviour of the rest frame peak energy and light curve used in our model with respect to peak time.

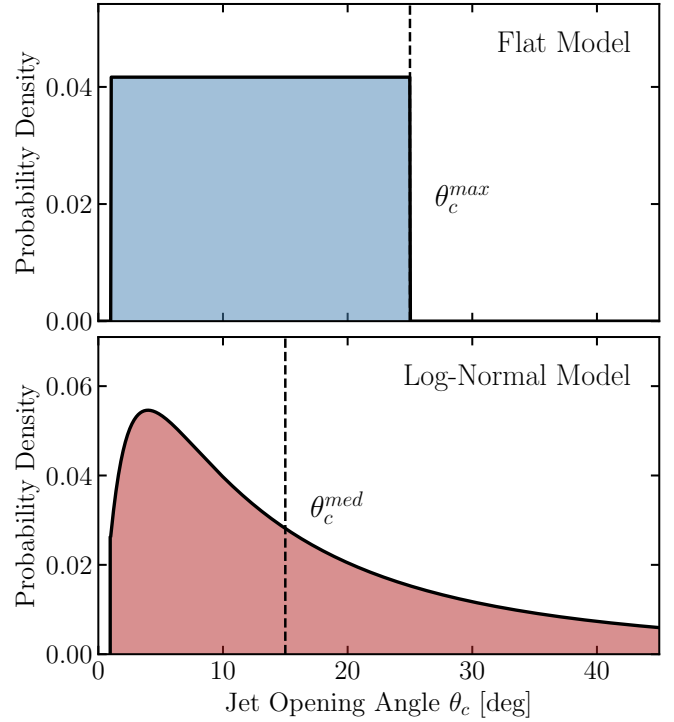


Fig. A.3. Example distributions for the two non-universal population models in our analysis.

Appendix B: Impact of the constant energy assumption on non-universal jets

In our main analysis of non-universal jet geometries (Sect. 3.3), we treat the isotropic equivalent luminosity L and the jet opening angle θ_c as independently sampled parameters. This implies an underlying population where wider jets possess intrinsically higher true total energies. To verify that our constraints are robust against this assumption, we tested an alternative ‘constant energy’ scenario where the total energy reservoir of the central engine is conserved. In this configuration, the apparent isotropic luminosity of each simulated burst is scaled for wider jets by a factor of $(1 - \cos \theta_c)$. Therefore the extraction of luminosity in the MCMC is conditional on the extraction of θ_c and this formally translates in assuming a probability distribution of the jet luminosity in the form $P(L|k, L^*(\theta_c))$, where $L^*(\theta_c) \propto 1 - \cos \theta_c$.

To determine if this coupling alters our constraints, we reran our MCMC framework for one of the models in higher tension with observations, the low-rate K265 model with high kicks ($\sigma = 265$ km/s) and $\alpha_{\text{CE}} = 0.5$, to study how this variation would affect the tension. We performed this test for both the Flat and Log-Normal non-universal jet distributions. We also performed a test on our Fiducial population and we obtained similar results.

Figure B.1 shows the joint posterior distributions for jet fraction, angle and energetics, comparing our standard baseline analysis against our analysis introducing the ‘constant energy’ assumption. We keep the nomenclature L^* for the luminosity parameter, noting that by adding the shift the physical meaning is not the same.

Apart from a shift in the posterior of L^* , we find that introducing the ‘constant energy’ assumption modifies the inferred median parameters by less than 15% compared to the baseline, with the remaining parameters (k, μ_E, σ_E) matching even more

closely. Most importantly, the constraints on the jet fraction f_j and the characteristic jet aperture remain effectively unchanged.

Appendix C: *Fermi*-GBM data

We provide the distributions of the *Fermi*-GBM data used in this work in Fig. C.1. As discussed in Sect. 2.4 we have very similar cuts to previous works (Salafia et al. 2023; Ronchini et al. 2022). The figure displays the complementary cumulative distribution functions for the four main observables which are peak energy, T_{90} , fluence, and peak flux. The vertical dashed lines indicate the specific cuts applied to define our sGRB sample ($F_p^{\text{lim}} \geq 4 \text{ ph cm}^{-2} \text{ s}^{-1}$, $T_{90} < 2 \text{ s}$, and $E_p \in [50 \text{ keV}, 10 \text{ MeV}]$). We additionally plot a dotted power law line with index $-3/2$ which is the behaviour expected from uniformly distributed GRBs (von Kienlin et al. 2020). This power law is used to define the peak flux cut to minimise selection effects.

Appendix D: Description of population synthesis models

In this work we analysed the binary population synthesis models generated by Iorio et al. (2023) using the SEVN code. These models explore the large parameter space of binary evolution uncertainties. While the original work provides a comprehensive technical description, we present here a simplified summary of the model variations most relevant to the formation of BNS mergers and their subsequent sGRB production. Table D.1 lists the model acronyms used throughout this paper and summarises the major physical assumptions and differences among the populations.

Appendix E: MCMC convergence and posterior predictive checks

To ensure the MCMC simulations reached a converged posterior distribution, we monitored the integrated autocorrelation time, τ , as defined in Foreman-Mackey et al. (2013). We required the total chain length to be at least 50 times greater than τ , a criterion typically satisfied within 30,000 to 40,000 iterations, indicating well-sampled posteriors. With convergence established, we performed a posterior predictive check to assess the model's goodness-of-fit. We generated synthetic sGRB catalogues by randomly sampling parameter sets from the converged chains and comparing the resulting Empirical Cumulative Distribution Functions (ECDFs) for the four key observables against the *Fermi*-GBM data. We show in Fig. E.1 the simulated median and 90% credible intervals having very good agreement with the observed distributions. The test is done on our Fiducial population. This confirms that the model successfully reproduces the fundamental features of the sGRB population. The resulting posterior probability distributions for the universal structured jet and the universal top-hat models are shown in Fig. E.2 and Fig. E.3, with relative medians summarised in Table E.1.

It is important to note that the resulting median characteristic energy posterior has a median at $E^* \sim 10^{48} - 10^{49} \text{ erg}$. This value may appear low compared to the canonical beaming-corrected kinetic energies of sGRBs, which are often inferred to be in excess of 10^{50} erg (e.g. Berger 2014). This apparent discrepancy arises directly from our model's parameter definitions. Our E^* represents the characteristic value of the radiated energy, distributed as a cut-off power-law, not the total beaming-corrected kinetic energy of the outflow. Taking into account this

definition, our model produces a population of synthetic bursts whose observable fluences and fluxes are consistent with the distributions seen by *Fermi*-GBM.

Appendix F: Impact of physical priors on the inferred jet fraction

In our primary analysis, we allowed the jet fraction parameter f_j to explore values well in excess of unity (setting the prior upper bound $f_{\text{max}} = 10$). While physically impossible, treating f_j as an unbounded effective parameter is essential for quantifying the tension between BNS population models and sGRB observations.

In order to see the effect of imposing a strict physical prior ($f_j \leq 1$), we repeated the analysis with this constraint and we found that for models with low intrinsic merger rates ($R_{\text{BNS}}(0) \lesssim 50 \text{ Gpc}^{-3} \text{ yr}^{-1}$), the posterior f_j saturates against the upper boundary of 1. In contrast, models with higher intrinsic rates find their optimal f_j well within the physical boundary. Figure F.1 illustrates this effect for the universal structured jet. With an unbounded prior, the median f_j continuously rises as the BNS rate decreases. This allowed us to distinguish between a model that requires $f_j \approx 1$ (corresponding to a physical scenario where all BNS mergers produce a jet) and one requiring $f_j \gg 1$ (which represents a fundamental breakdown of the sole BNS-progenitor hypothesis).

Appendix G: Example narrowest geometry

We give a visual example on how to recover the narrowest geometry for a universal top-hat model in Fig. G.1. Given that for all models the product between jet fraction f_j and beaming factor f_b is monotone as a function of the structure parameter (θ_c , θ_c^{max} and θ_c^{med}), we can find the minimum value of these parameters to obtain a physical jet fraction.

Appendix H: Summary of model viability criteria

In this appendix we detail the criteria used to construct the summary of results presented in Table H.1. The 64 BNS population models, characterised by different binary evolution prescriptions and CE efficiencies (α_{CE}), are evaluated based on two primary constraints. One is physical viability. A model is marked as physical ($\checkmark$) if the median of its inferred jet fraction posterior distribution lies within the physical regime (median $f_j \leq 1$), which is equivalent to $P(f_j \leq 1) \geq 0.5$. If the median $f_j > 1$, the model is marked as non-physical ($\times$). For models with non-universal structure we fixed the median θ_c at the median value of Rouco Escorial et al. (2023, $\sim 6^\circ$). The other is geometric consistency. For models where the jet structure parameters are free (universal top-hat, non-universal flat, and non-universal log-normal), we assessed if the geometry required to maintain a physical f_j is consistent with afterglow observations. We used the 90% credible interval upper limit from Rouco Escorial et al. (2023) of 15.4° . We classified the models using the following criteria:

- For the universal top-hat and flat models, the index a is applied if the minimum characteristic opening angle θ^* required to satisfy median $f_j \leq 1$ exceeds 15.4° .
- For the log-normal model, the index b is applied if more than 10% of the inferred jet population is required to have an opening angle $\theta_c > 15^\circ$ to satisfy the rate requirement.

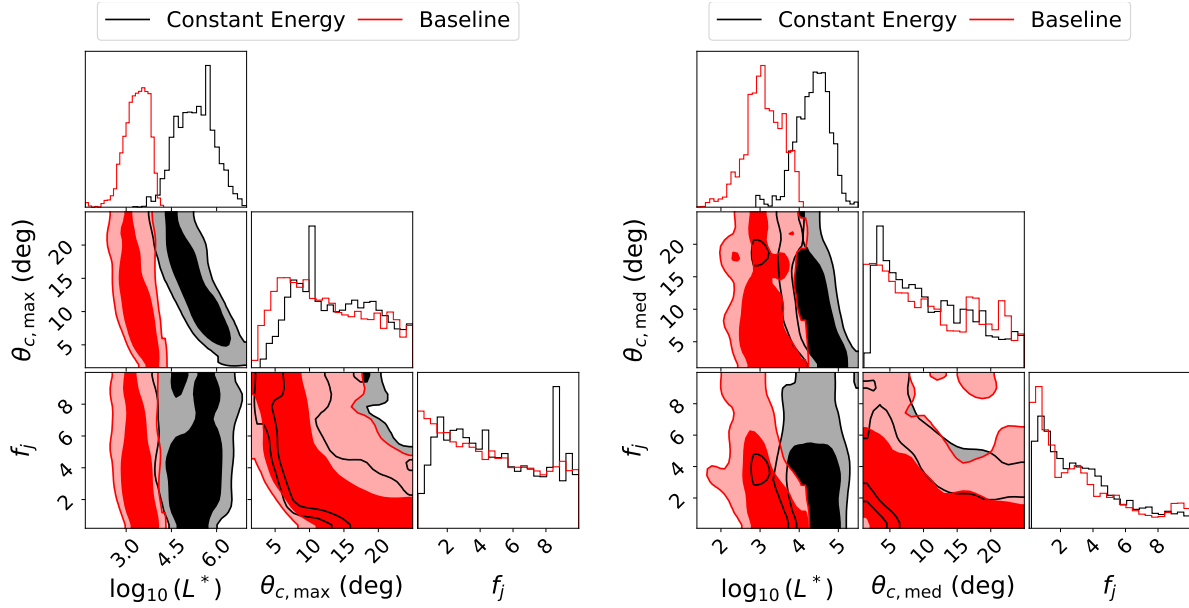


Fig. B.1. Posterior distributions for the K265 ($\alpha_{CE} = 0.5$) population, comparing our baseline model against the constant energy assumption where luminosity scales with $1 - \cos(\theta_c)$. The 1 and 2σ contours are shown. *Panel (a):* Flat model comparison. *Panel (b):* Log-normal model comparison.

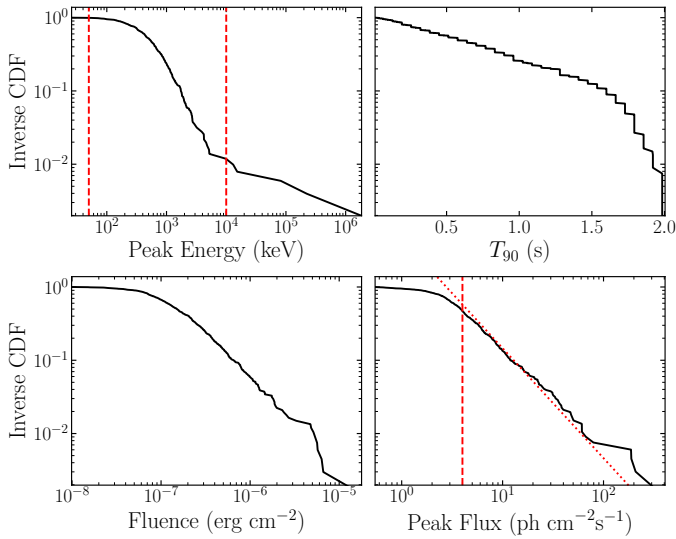


Fig. C.1. Inverse cumulative distributions of the observables retrieved from the *Fermi*-GBM catalogue. The additional quality cuts are shown with vertical red dashed lines. The dotted line shows a power law with index $-3/2$.

The combination of these marks allows for a rapid identification of viable BNS populations which are those that provide enough progenitors to match the sGRB rate while maintaining jet geometries consistent with observed afterglows.

Appendix I: Redshift distribution of the average time delay

In Fig. I.1 (for physical models) and Fig. I.2 (for the remaining models), we show the redshift evolution of the average delay time, $\langle\tau_d\rangle$, for each of our 64 models. Across all models, we observe that the delay times decrease by approximately two orders of magnitude as redshift increases. While the phys-

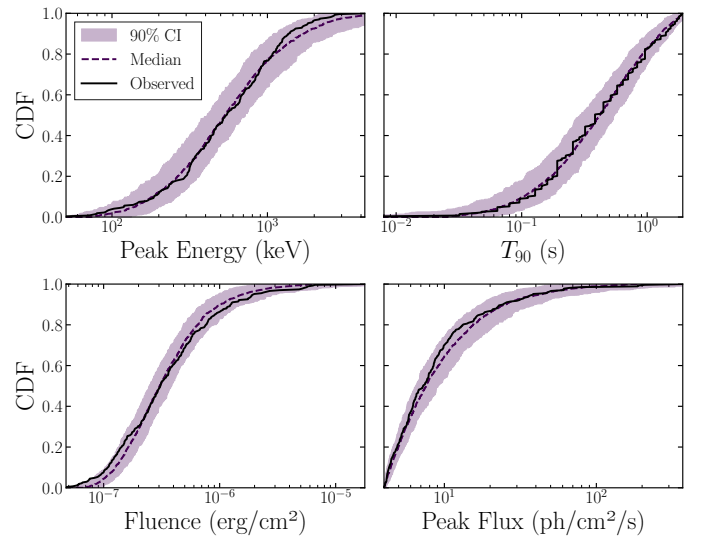


Fig. E.1. Cumulative distributions of the observables for real *Fermi*-GBM sGRBs (solid line) and simulated sGRBs associated with our fiducial BNS population.

ical models exhibit the best agreement with the population averages derived by [Pracchia & Salafia \(2026\)](#), it is crucial to note that the average delay time is sensitive to the distribution tail. Specifically, $\langle\tau_d\rangle$ can be skewed by the large delay times found at low redshifts ($z < 1$). Since this redshift range corresponds to the volume where sGRBs are most likely to be detected, the observed population is inherently biased towards these longer delay times, this underscores the importance of accounting for selection effects when constructing sGRB catalogues from fixed population synthesis models. A similar conclusion is reached by [Pracchia & Salafia \(2026\)](#), who note that selection effects, specifically when applying a peak flux cut that is too low, can bias the sample towards larger delay times.

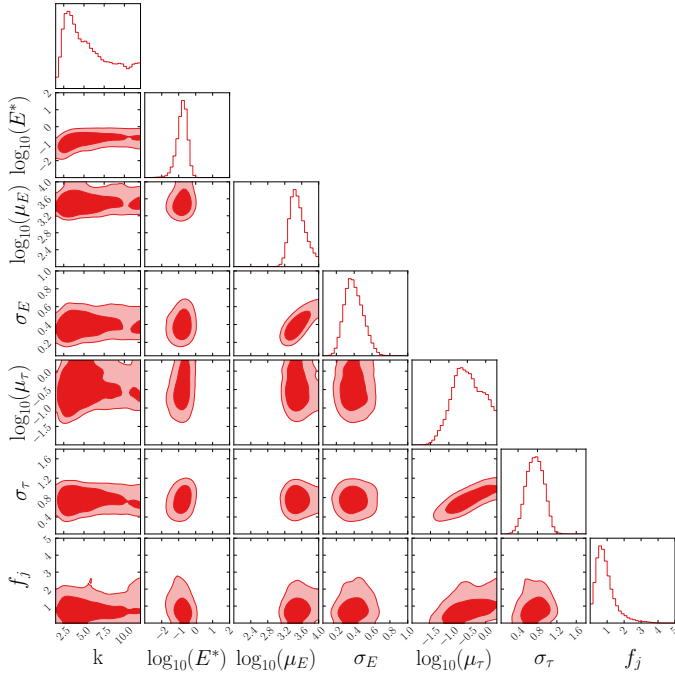
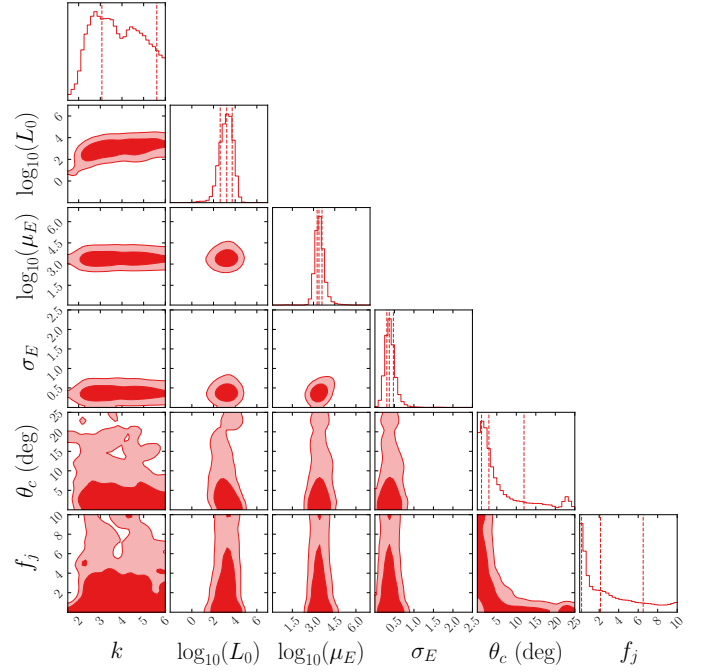
Table D.1. Primary physical assumption varied for each BNS population synthesis models from Iorio et al. (2023).

Model	Physics Var- ied	Description & Impact on Rates
F	Fiducial	The baseline model using standard assumptions for stellar winds and mass transfer, using a mass-dependent supernova kick model that allows for relatively high survival rates (Giacobbo & Mapelli 2020).
<i>Supernova Natal Kicks</i>		
K265	High Kicks	Neutron stars receive strong birth kicks (Maxwellian $\sigma = 265$ km/s). Significantly reduces the merger rate (by factor ~ 4 vs F) as binaries are disrupted.
K150	Low Kicks	Neutron stars receive weaker birth kicks ($\sigma = 150$ km/s), increasing the rate compared to K265, but still yields fewer mergers than F.
<i>Mass Transfer Stability</i>		
QCBSE	Standard Sta- bility	Uses restrictive criteria to decide if mass transfer is stable. Drastically suppresses the BNS rate (by ~ 7) as giant donors undergo unstable transfer and merge prematurely.
QCBB	Alternative Stability	Assumes mass transfer from pure-Helium stars is always stable. This avoids for some stars a final, often fatal CE phase where they would otherwise merge prematurely, allowing more systems to survive and eventually merge as BNSs.
RBSE	Accretion Effi- ciency	Assumes the companion star accretes matter less efficiently during mass transfer, giving a moderate reduction in merger rates compared to F.
<i>Common Envelope Physics</i>		
LK, LC, LX	High Binding Energy	LK and LC assume tightly bound envelopes. This suppresses BNS formation as binaries fail to eject the envelope and merge as stars. Similarly, LX uses the binding energy prescriptions from Xu & Li (2010) which also reduces merger rates.
OPT	Optimistic CE	Allows Hertzsprung gap donors to survive a CE phase. While it allows survival where standard models fail, this model is built upon the restrictive QCBSE physics.
<i>Other Assumptions</i>		
F19	Pair Instability	Uses Pair Instability prescriptions from Farmer et al. (2019). This mainly affects the black hole mass spectrum and has negligible impact on BNS merger rates.
F5M	Sampling Check	A high-resolution run of F. Used to verify that merger rates are not artefacts of statistical sampling noise.
SND	Supernova Engine	Uses a delayed supernova mechanism. This yield comparable total rates to F but alters the mass distribution of the compact objects by allowing some progenitors to not directly collapse into a black-hole.
QHE	Homogeneous Evolution	Allows rapidly rotating stars to mix chemically. Total merger rates are similar to F, though formation channels differ as these stars do not form Red Giants and remain compact.
NT	No Tides	Disables tidal forces. This results in a minor reduction in merger rates compared to F.
NTC	Circularization	Allows binaries to merge sooner during Roche-lobe overflow. Increases the merger rate (~ 2) by preventing collisions in eccentric orbits.

Notes. All models listed above are combined with four variations of the CE efficiency parameter, $\alpha_{\text{CE}} \in \{0.5, 1.0, 3.0, 5.0\}$, resulting in the 64 total populations analysed in this work.

Table E.1. Fiducial BNS population's median posterior values and 90% credible intervals for the sGRB emission model parameters.

Structured jet		Universal top-hat jet	
Parameter	Value	Parameter	Value
k	$5.23^{+4.16}_{-2.33}$	k	$6.74^{+3.55}_{-3.23}$
$\log_{10}(E^*/10^{49} \text{ erg})$	$-0.75^{+0.18}_{-0.31}$	$\log_{10}(L^*/10^{49} \text{ erg s}^{-1})$	$3.23^{+0.47}_{-0.57}$
$\log_{10}(\mu_E/\text{keV})$	$3.48^{+0.22}_{-0.13}$	$\log_{10}(\mu_E/\text{keV})$	$3.42^{+0.37}_{-0.13}$
σ_E	$0.39^{+0.11}_{-0.08}$	σ_E	$0.39^{+0.13}_{-0.08}$
$\log_{10}(\mu_t/\text{s})$	$-0.50^{+0.48}_{-0.38}$	$\theta_c \text{ (deg)}$	$2.74^{+4.38}_{-1.66}$
σ_t	$0.76^{+0.17}_{-0.18}$		
f_j	$0.78^{+0.61}_{-0.36}$	f_j	$2.87^{+4.43}_{-2.36}$

**Fig. E.2.** Posterior of the universal structured model for the fiducial population. The 1 and 2σ contours are shown.**Fig. E.3.** Posterior of the universal top-hat model for the fiducial population. The 1 and 2σ contours are shown.

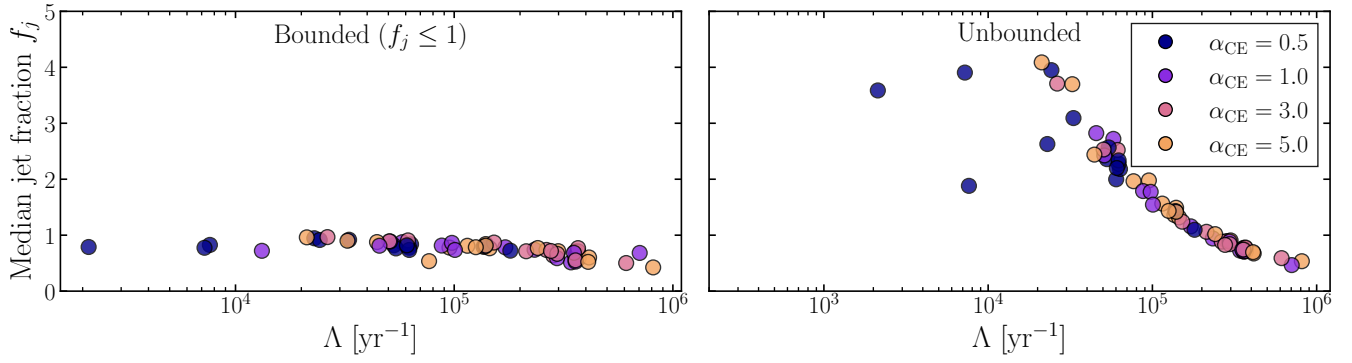


Fig. F.1. Comparison of the inferred median jet fraction f_j as a function of total BNS rate Λ . The left panel shows bounded models, the right unbounded models. We zoom in on the region $f_j \in [0, 5]$ for clarity. Both cases assume a universal structured jet.

Table H.1. Physical viability and geometric consistency of BNS population models.

α_{CE}	Universal top-hat				Non-univ. flat				Non-univ. log-normal				Universal structured			
	0.5	1.0	3.0	5.0	0.5	1.0	3.0	5.0	0.5	1.0	3.0	5.0	0.5	1.0	3.0	5.0
F	$\times^a$	$\checkmark$	$\checkmark$	$\times$	$\times$	$\checkmark$	$\checkmark$	$\times$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\checkmark$	$\checkmark$	$\times$
K265	$\times^a$	$\times$	$\times$	$\times^a$	$\times$	$\times$	$\times$	$\times$	$\checkmark^b$	$\checkmark$	$\checkmark$	$\checkmark^b$	$\times$	$\times$	$\times$	$\times$
K150	$\times$	$\times$	$\checkmark$	$\times$	$\times$	$\times$	$\times$	$\times$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\times$	$\times$	$\times$
QCBSE	$\times^{a,c}$	$\times$	$\times^a$	$\times^a$	$\times$	$\times$	$\times$	$\times$	$\times^b$	$\checkmark$	$\checkmark^b$	$\checkmark^b$	$\times$	$\times$	$\times$	$\times$
QCBF	$\checkmark$	$\times$	$\times$	$\times$	$\checkmark$	$\checkmark^a$	$\times$	$\times$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark^b$	$\times$	$\times$	$\times$	$\times$
RBSE	$\times$	$\times^a$	$\checkmark$	$\times$	$\times^a$	$\checkmark$	$\checkmark$	$\times^a$	$\checkmark^b$	$\checkmark$	$\checkmark$	$\checkmark^b$	$\times$	$\checkmark$	$\checkmark$	$\times$
LK	$\times^{a,c}$	$\times^c$	$\times$	$\checkmark$	$\times$	$\times$	$\checkmark$	$\checkmark$	$\times^b$	$\checkmark^b$	$\checkmark$	$\checkmark$	$\times$	$\times$	$\times$	$\checkmark$
LC	$\times^a$	$\times$	$\checkmark$	$\checkmark$	$\times$	$\times$	$\checkmark$	$\checkmark$	$\checkmark^b$	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\times$	$\checkmark$	$\checkmark$
LX	$\times^{a,c}$	$\times$	$\checkmark$	$\checkmark$	$\times$	$\times$	$\checkmark$	$\checkmark$	$\times^b$	$\checkmark^b$	$\checkmark$	$\checkmark$	$\times$	$\times$	$\checkmark$	$\checkmark$
OPT	$\times^{a,c}$	$\times$	$\checkmark$	$\times^c$	$\times$	$\times$	$\checkmark$	$\checkmark$	$\times^b$	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\times$	$\checkmark$	$\checkmark$
F19	$\times^a$	$\checkmark$	$\checkmark$	$\times$	$\times^a$	$\checkmark$	$\checkmark$	$\times$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\checkmark$	$\checkmark$	$\times$
F5M	$\times^a$	$\checkmark$	$\checkmark$	$\times^a$	$\times$	$\checkmark$	$\checkmark$	$\times$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\checkmark$	$\checkmark$	$\times$
SND	$\times$	$\times^c$	$\checkmark$	$\times$	$\times^a$	$\checkmark$	$\checkmark$	$\times$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\checkmark$	$\checkmark$	$\times$
QHE	$\times$	$\checkmark$	$\checkmark$	$\times$	$\times$	$\checkmark$	$\checkmark$	$\times$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\checkmark$	$\checkmark$	$\times$
NT	$\times$	$\times^c$	$\times^c$	$\times$	$\times$	$\checkmark$	$\checkmark$	$\times$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\checkmark$	$\checkmark$	$\times$
NTC	$\times$	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\times$	$\checkmark$	$\checkmark$	$\times$

Notes. Checkmarks ($\checkmark$) indicate physical median jet fractions ($f_j \leq 1$). Crosses ($\times$) indicate non-physical scenarios with median $f_j > 1$. The aperture angle is fixed at 6.1° for top-hat models and the distribution median for non-universal models, following Rouco Escorial et al. (2023).
^aRequired $\theta^* > 15.4^\circ$. ^bMore than 10% of the inferred jet population exceeds $\theta_c = 15^\circ$. ^cRequired θ^* exceeded prior bounds.

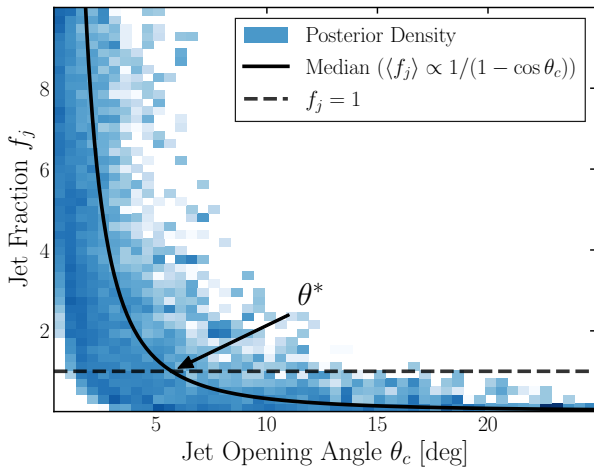


Fig. G.1. Narrowest geometry for the fiducial population using a universal top-hat jet. A solid line overlays the joint posterior $P(\theta_c, f_j)$ to show the median f_j as a function of θ_c . The intersection $f_j = 1$ defines the minimum characteristic opening angle, θ^* .

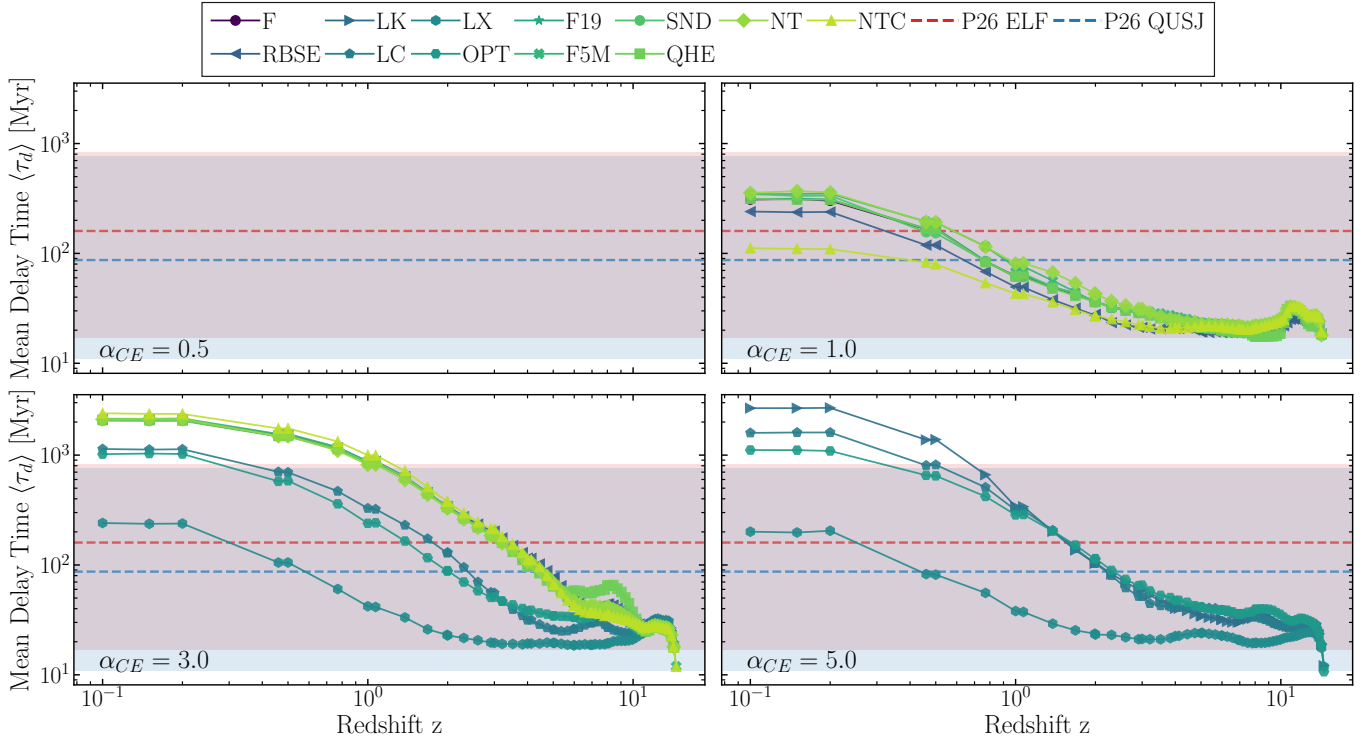


Fig. I.1. Average delay time $\langle\tau_d\rangle$ evolution with redshift for physical population (median $f_j \leq 1$) assuming the universal structured jet compared to the median and 90% credible intervals inferred in Pracchia & Salafia (2026). Plots with erratic spikes have a very small number of mergers. See Table D.1 for a description of each model variation (top right of each plot). See Fig. I.2 for non physical populations. For consistency with Fig. I.2 we keep the panel for $\alpha_{CE} = 0.5$ even though no physical models exist for that value in the universal structured jet case.

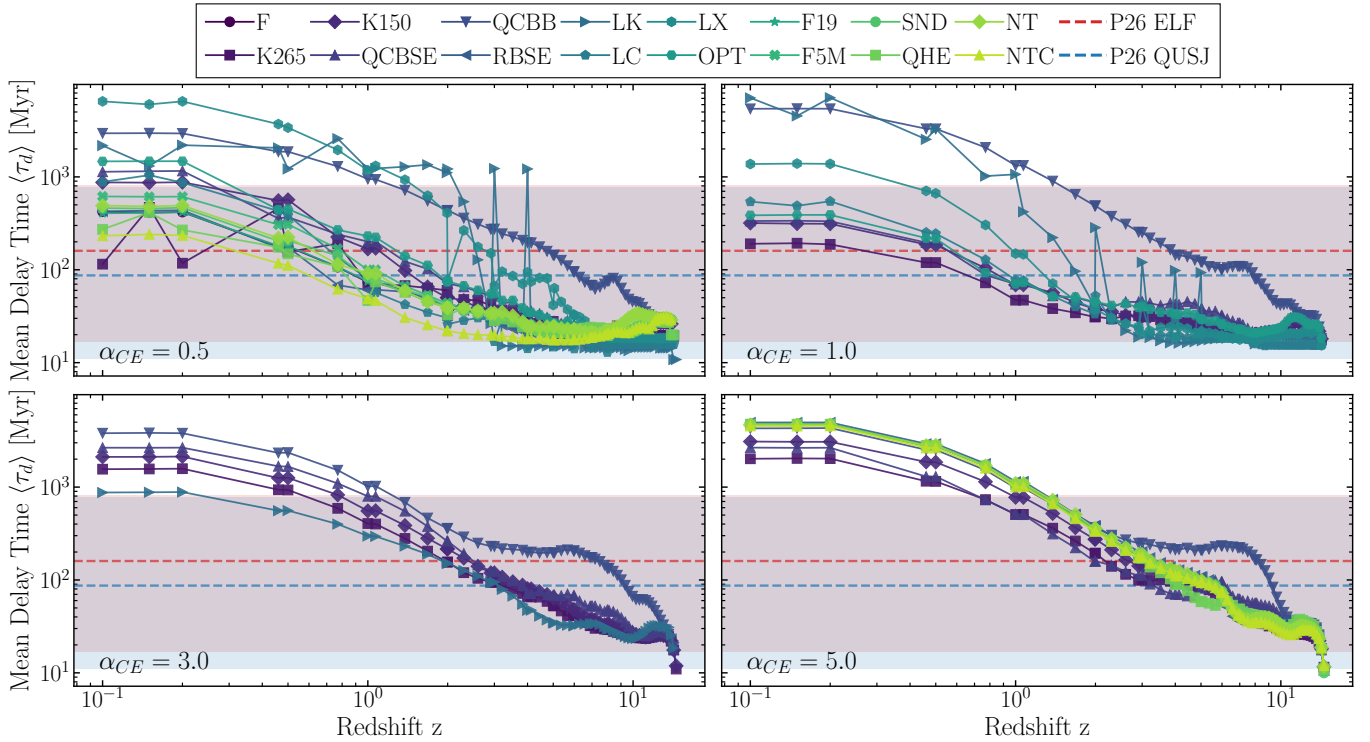


Fig. I.2. Same as Fig. I.1 but for non physical populations (median $f_j > 1$).