Classification of Gamma-Ray Burst durations using robust model-comparison techniques

Soham Kulkarni[1] • Shantanu Desai[2]

Abstract Gamma-Ray Bursts (GRBs) have been conventionally bifurcated into two distinct categories dubbed "short" and "long", depending on whether their durations are less than or greater than two seconds respectively. However, many authors have pointed to the existence of a third class of GRBs with mean durations intermediate between the short and long GRBs. Here, we apply multiple model comparison techniques to verify these claims. For each category, we obtain the best-fit parameters by maximizing a likelihood function based on a weighted superposition of lognormal distributions. We then do model-comparison between each of these hypotheses by comparing the chi-square probabilities, Akaike Information criterion (AIC), and Bayesian Information criterion (BIC). We uniformly apply these techniques to GRBs from Swift (both observer and intrinsic frame), BATSE, BeppoSAX, and Fermi-GBM. We find that the Swift GRB distributions (in the observer frame) show evidence for three categories at about 2.4σ from difference in chi-squares and show decisive evidence in favor of the two components using both AIC and BIC. For all the other datasets, evidence for three components is either very marginal or not favored.

1 Introduction

Gamma-ray bursts (GRBs) are short-duration energetic cosmic explosions with prompt emission between KeV-

Soham Kulkarni

Shantanu Desai

[1,2]Department of Physics, IIT Hyderabad, Kandi, Telangana-502285, India

GeV energies, and were first detected by the Vela military satellites in late 1960s and continue to be detected at the rate of about one per day by many different satellites (Zhang et al. 2016a). Data from the Burst and Transient Source Explorer (BATSE) onboard the Compton Gamma Ray Observatory (CGRO) was analyzed by Kouveliotou et al. (1993), and led to establishing the conventional classification of GRBs into short (T90 < 2 s) and long (T90 > 2 s) classes, where T90 is the time which encompasses 90% of the bursts fluence, and is used as a proxy for the duration of a GRB. Most classification studies of GRBs have been done using T90, although other measures have also been proposed (Zhang 2006; Li et al. 2016). The progenitors of long GRBs consist of supernovae related to the collapse of massive stars (Woosley & Bloom 2006) and those of short GRBs are thought to be binary compact object mergers (Nakar 2007). There are however exceptions to this general picture (Zhang 2006; Bromberg et al. 2013).

It has been observed that T90 exhibits lognormal distributions, which were thereafter fit to short and long GRBs (McBreen et al. 1994; Koshut et al. 1996; Kouveliotou et al. 1996; Horváth 2002). The existence of an intermediate-duration GRB class, with T90 in the range 2-10s in the BATSE dataset was first put forward (Horváth 1998; Mukherjee et al. 1998). This was confirmed from further analysis of the complete BATSE dataset (Horváth 2002; Chattopadhyay et al. 2007; Zitouni et al. 2015). Evidence for a third lognormal component was also found in Swift/BAT data (Horváth et al. 2008; Zhang & Choi 2008; Huja et al. 2009; Horváth et al. 2010). This was recently confirmed for the October 2015 Swift GRB catalog by Horváth & Tóth (2016), who pointed that three lognormal distributions provide a better fit to the data than two with 99.9999% confidence level. Tarnopolski (2016a) finds that for the similar Swift GRB dataset consisting of 947

GRBs, three groups are favored in the observer frame, whereas two in the rest frame. However, these results have been disputed by Yang et al. (2016), who find that after applying the Gaussian mixture model on T90 and hardness ratio on Swift GRBs with redshifts, two components are favored compared to three or more in both the observer and intrinsic frame. Another recent analysis of BATSE, Swift, and Fermi/GBM by applying Gaussian Mixture model on T90 finds that BATSE, Fermi and Swift data (for GRBs with measured redshifts in observer and rest frame) are better fitted by two components in the observer frame (Zhang et al. 2016b). The same analysis finds that the Swift data is consistent with three components in the observer frame. Hence, there is no uniform consensus among the authors inspite of analyzing the same GRB datasets.

To resolve these conflicting results, we use multiple model-comparison techniques on the distribution of T90 to find out the optimum number of GRB classes. These include both frequentist hypothesis testing methods as well as Bayesian procedures such as Akaike Information Criterion and Bayesian Information Criterion. These model comparison techniques have been applied to a variety of problems in astrophysics and particle physics (See Shafer (2015); Desai & Liu (2016); Desai (2016) and references therein). We apply these methods to data from multiple detectors including BATSE, Fermi-GBM, BeppoSAX, and Swift.

The outline of this paper is as follows. In Sect. 2 we discuss the methodology used to obtain the best-fit parameters for the mean GRB duration and its variance, after positing two and three classes of GRBs. In Sect 3, we discuss various techniques used for model comparison. We then present results for various GRB datasets in Sect 4, including a very brief comparison with previous results. We conclude in Sect. 5.

2 Parameter Estimation

2.1 Datasets

Herein, we consider the GRB datasets available from BATSE[1] 4B catalog (Paciesas et al. 1999), Swift[2] (Lien et al. 2016), Fermi-GBM[3] (Narayana Bhat et al. 2016) and BeppoSAX[4] (Frontera et al. 2009). The number of GRBs analyzed for the model comparison are 2036 from BATSE, 927 from Swift, 1901 from Fermi, and 1003

from BeppoSAX. These detectors account for almost all the GRBs discovered in the past three decades. We did not consider other catalogs such as those from RHESSI, INTEGRAL etc, as they contained less than 500 GRBs.

2.2 Fitting method

We have applied the same Maximum Likelihood(ML) method as proposed in Horváth & Tóth (2016) (see also Horváth (2002); Horváth et al. (2008); Horváth (2009)) for fitting the data and obtain the best-fit parameters. As is done in a ML method, we select a probability density function and define a log-likelihood function, which is to be maximized over the parameter space by varying the free parameters. We model the probability density function to be a superposition of k lognormal gaussian distributions, where k is the total number of GRB classes. Also, we associate a weight w_j for each k, which indicates the number of GRBs of that particular type found in our dataset. For a probability density function $f(x, \theta)$, where θ is the set of parameters required for defining the probability function, the log-likelihood will be defined as:

$$\mathcal{L} = \sum_{i=1}^{N} \ln \sum_{j=1}^{k} w_j f_j(x_i, \theta) \tag{1}$$

where x_i are the sample datapoints, w_j is the number of GRB categories, and N is the total number of GRBs analyzed. As stated above we take the k lognormal distributions of the form:

$$f(x, \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \log T_{90})^2}{2\sigma^2}\right) \tag{2}$$

under the condition that

$$\sum_{i=1}^{k} w_i = N \tag{3}$$

The maximization is done by implementing an optimization algorithm (`SLSQP` and `COBYLA`) included in the `SciPy` package of Python. We did not bin the data during the optimization process. We should also point out that optimization of the likelihood in Eq. 1 is mathematically similar to Gaussian Mixture Model (GMM), if the weights in Eq. 3 are normalized to unity and the covariance matrix is diagnol (Ivezić et al. 2014). The parameters of the GMM can be found by the Expectation-Maximization (EM) Algorithm (Ivezić et al. 2014). The GMM and the corresponding parameter estimation using the EM algorithm have been also applied to GRB datasets using both T90 (Zhang et al. 2016b) as well as using T90 vs hardness ratio (Yang et al. 2016). We

---

also tried to estimate the best-fit parameters by applying the EM algorithm after normalizing the weights to unity, instead of the total number of GRBs. However, the best likelihood model is still obtained by using the optimization algorithm in `SciPy` and in the rest of the paper, we report the best-fit values from this.

3 Model Comparison

The comparison of models on the basis of best-fit likelihood (or minimum χ^2) is not a good way to do hypothesis testing or select the optimum model after finding the best-fit parameters for each model. As we increase the number of free parameters, it is obvious that the likelihood will increase, but it leads to over-fitting. Therefore, the additional free parameters need to be penalized so as to avoid getting a bad result. To address these issues, a number of both frequentist and Bayesian model-comparison techniques have been used over the past decade to determine the best model which fits the observational data (Liddle 2004). Here, we use multiple analysis methods, such as the frequentist hypothesis testing (based on χ^2 probabilities) and information criterion based tests such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for model comparison. AIC and BIC have also been previously used for GRB classification by a number of authors (Mukherjee et al. 1998; Tarnopolski 2016a,b; Yang et al. 2016; Zhang et al. 2016b). Frequentist model comparison after binning the data has been used by Zitouni et al. (2015); Tarnopolski (2015). More information about AIC and BIC and its application to astrophysical problems can be found in Liddle (2004, 2007); Shi et al. (2012); Shafer (2015).

3.0.1 Chi-Square Test

In order to construct a frequentist model comparison test, we calculate the reduced χ^2 to compare different models. The reduced χ^2 is equal to χ^2/ν, where ν is the total degrees of freedom and χ^2 is defined as follows:

$$\chi^2 = -\sum_{i=1}^{N} \ln \sum_{j=1}^{k} \frac{w_j}{\sqrt{2\pi}\sigma_j} \exp\left(\frac{-(x_j - \log T_{90})^2}{2\sigma_j^2}\right); \quad (4)$$

under the condition that

$$\sum_{i=1}^{k} w_i = 1 \qquad (5)$$

This is essentially the same expression (modulo the minus sign) that we are using for the calculation of the

likelihood in Eq. 1, except that the weights in Eq. 5 are now normalized to unity instead of the total number of GRBs. We have also verified using numerical simulations that for a distribution of two Gaussians, $\chi^2/\nu \sim 1$ for the best-fit input parameters, where χ^2 is defined in Eq. 4. After obtaining the best-fit model parameters for each hypothesis, we compare the χ^2 probability, after taking into account the total degrees of freedom. The χ^2 probability is equal to $\Gamma(\nu/2, \chi^2/2)/\Gamma(\nu/2)$ (Press et al. 1992), where Γ is the incomplete Gamma function and ν is the total degrees of freedom. The preferred model is the one with higher value of χ^2 probability. If two models are nested, then according to Wilk's theorem (Wilks 1938), the difference in χ^2 between the two models satisfies a χ^2 distribution with degrees of freedom equal to the difference in the number of free parameters for the two hypotheses (Lyons 2016). Since a model with two Gaussian components is a special case of a model with three components, we can apply Wilk's theorem to assess the statistical significance of the better model.

3.0.2 AIC

The Akaike Information Criterion (AIC) is used for model comparison, when we need to penalize for any additional free parameters to avoid overfitting. A preferred model in this test is the one with the smaller value of AIC between the two hypothesis. The AIC is given by:

$$AIC = 2p + 2\ln L \qquad (6)$$

where p is the number of free parameters in the model and L is the likelihood. The AIC defined in Eq. 6 is good when the ratio N/p is very large i.e. > 40 (Burnham & Anderson 2004). For a smaller value of the ratio, a first order correction is included and the expression is:

$$AIC = 2p + 2\ln L + \frac{2p(p+1)}{N-p-1} \qquad (7)$$

As all our datasets have a ratio of N/p greater than 40, we don't need to worry about this correction. The absolute value of AIC is usually not of interest. The goodness of fit between two hypothesis (A) and (B) is described by the difference of the AIC values and is given by,

$$\Delta AIC = AIC_A - AIC_B, \qquad (8)$$

where AIC_A - AIC_B correspond to the AIC values for the hypothesis A and B. Burnham & Anderson (2004) have provided qualitative strength of evidence rules to assess the significance of a model based on the ΔAIC values between the two models. If ΔAIC> 5, then it is considered strong evidence against the model with higher AIC and ΔAIC> 10 is considered as decisive evidence against the model with higher AIC (Liddle 2007). Values of ΔAIC< 5 correspond to weak evidence.

3.0.3 BIC

The Bayesian Inference Criterion (BIC) is also used for penalizing the use of extra parameters. As in the case of AIC, the model with the smaller value of BIC is the preferred model. The penalty in the BIC test is harsher than that in the case of AIC and is given by:

$$BIC = p \ln N + 2 \ln L \tag{9}$$

The logarithmic term and the number of free parameters act as a very harsh measure needed for the BIC test. The goodness of fit used for hypothesis testing between two models A and B is given by:

$$\Delta BIC = BIC_A - BIC_B \tag{10}$$

Similar to AIC, the model with lower value of BIC is favored. To assess the significance of a model, strength of evidence rules have also been proposed based on ΔBIC (Kass & Raftery 1995), which are approximately the same as those for AIC.

4 Results

We apply all the techniques discussed in the previous sections to GRB datasets from various detectors. For data from each of the GRB detectors, we find the mean value of T90 and its standard deviation by positing that the data has two as well as three components, followed by maximizing the likelihood in Eq. 1 for both the hypotheses. For these best-fit parameters, we then implement all the three different model-comparison techniques outlined in Sect. 3. We now present our results for BATSE, BeppoSAX, Fermi-GBM and Swift.

4.1 BATSE

The current BATSE GRB (Paciesas et al. 1999) catalogue contains 2036 GRBs detected between 1991 and 2000. The fits for the data for $k = 2$ and $k = 3$ are shown graphically in Figs. 1 and 2 respectively. From
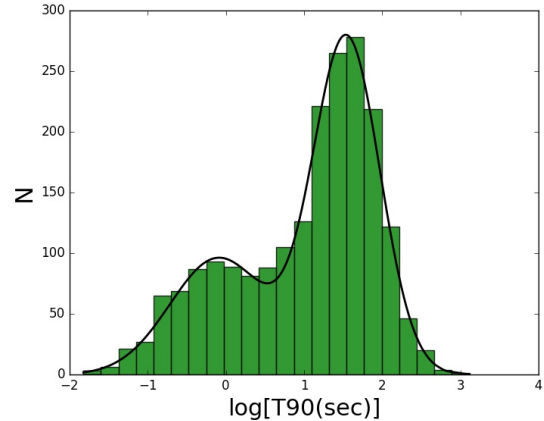
Fig. 1 A fit for the 2-component model for BATSE GRBs. Details of the fits can be found in Table 1.
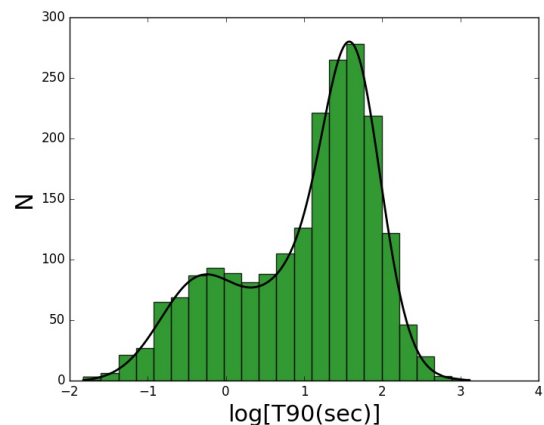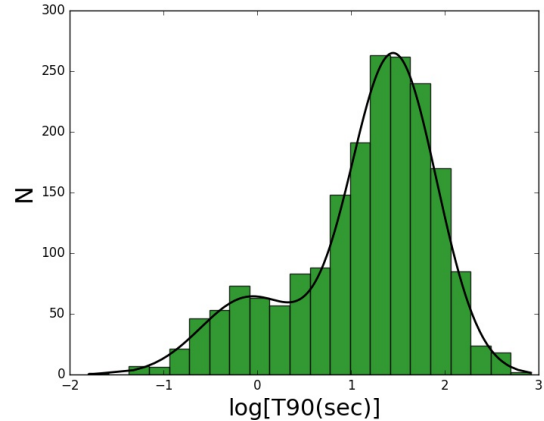

Fig. 2 A fit for the 3-component model for BATSE GRBs. Details of the fits can be found in Table 1.

Table 1 Model Comparison Parameters for BATSE GRBs. The first column contains the total number of GRB classes and the next three indicate the best-fit values for the logarithm of the mean T90 (μ_{T90}), its standard deviation (σ_{T90}), total number of GRBs (w_i) in each category after positing both two and three types of GRBs. These are obtained by maximizing Eq. 1. $\mathcal{L}$, $p(\chi^2, \nu)$, χ^2/dof, AIC , BIC represent the likelihood, χ^2 probability for ν DOF, reduced χ^2, Akaike and Bayesian Information criterion respectively. The last three columns indicate the p-value, ΔAIC, and ΔBIC between the two models, which are used for model comparison. (We have not used the likelihoods for model comparison). In the table the preferred value for every test is highlighted in bold. We find that AIC and χ^2 probability favor three components, whereas BIC favors two. However with all these model comparison techniques, the statistical significance is marginal.

k	μ_{T90}	σ_{T90}	w_i	$\mathcal{L}$	χ^2/ν	$P(\chi^2, \nu)$	AIC	BIC	p-value	$\Delta(AIC)$	$\Delta(BIC)$
2	-0.093	0.62	681	13076	1.1987	1.04e-10	4879.5	**4902**			
	1.542	0.43	1355								
3	-0.361	0.5	459	13080.9	1.1974	**1.3e-10**	**4873**	4907	0.108(1.2σ)	6.5	5
	1.09	0.65	692								
	1.63	0.36	885								

the figures, we can see that both the fits are indistinguishable by eye. When we fit the BATSE dataset for three components, we find that 459 GRBs belong to the short, 692 to the intermediate, and 885 to the long category. While fitting for two components, we find that 681 GRBs belong to the short category and 1355 to the long category. The detailed results of model comparison are tabulated in Table no. 1. This table contains the likelihood, reduced χ^2, AIC, and BIC for both the hypotheses. Here, we find that ΔAIC = 6.5, with the smaller value for $k = 3$. This corresponds to strong evidence for three components. When we compare the χ^2 probability, we find that $k = 3$ has a higher value, which implies that it is a better fit. To assess the statistical significance of $k = 3$ model compared to $k = 2$ model, we apply Wilk's theorem and find that the p-value is equal to 0.108. This implies that there is 10.8% probability that the third component is a statistical fluctuation. This p-value corresponds to 1.2σ Gaussian significance (Press et al. 1992). However, the $k = 2$ model has a lower value of BIC and ΔBIC between the two models is equal to 5, corresponding to weak evidence.

Therefore in summary, two of the three model comparison techniques (AIC and frequentist test) prefer $k = 3$ and one of them (based on BIC) prefer $k = 2$. However, in all the three cases, the significance of one model with respect to the other is marginal and none of these tests pass the 5σ criterion (usually used in high-energy physics) to decisively pick one model over the other. We note that from similar likelihood analysis of the BATSE data and comparison of likelihoods, Horváth (2002) found evidence for three GRB classes and the probability that the third group is a fluctuation is 0.5%. On the other hand, a recent GMM-based analysis of BATSE T90 dataset showed evidence for two components with ΔBIC=13 in favor of two component model (Zhang et al. 2016b).

Fig. 3 A fit for the 2-component model for FERMI GBM GRBs. Summary of the fits can be found in Table 2.
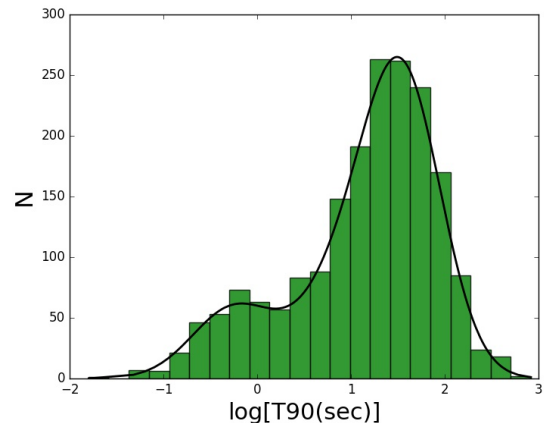
Fig. 4 A fit for the 3-component model for FERMI GBM GRBs. Summary of the fits can be found in Table 2.

Table 2 Model Comparison Parameters Fermi-GBM GRBs. The explanation of all columns can be found in Table 1. In the table the preferred value for every test is highlighted in bold. We find that AIC and χ^2 probability favor three components, whereas BIC favors two. However with all these model comparison techniques, the statistical significance is marginal.

| k	μ_{T90}	σ_{T90}	w_i	$\mathcal{L}$	χ^2/ν	$P(\chi^2, \nu)$	AIC	BIC	p-value	$\Delta(AIC)$	$\Delta(BIC)$
2	-0.0851	0.52	406	12313.5	1.075	4.74e-03	4086.7	**4109**			
	1.45	0.46	1495								
3	-0.251	0.443	309	12315.9	**1.0748**	**4.808e-03**	**4085**	4119	0.32(0.5σ)	1.7	10
	1.133	0.531	691								
	1.589	0.402	901								

4.2 Fermi-GBM

The Fermi-GBM catalogue (as of Sept. 2016) currently has 1901 GRBs (Narayana Bhat et al. 2016). When we fit the data for three components we find that 309 are short, 691 are intermediate, and 901 belong to long type. On positing two components, we find that 406 GRBs belong to the short category and 1495 belong to the long category. The model fits to the data shown in Figs. 3 and 4 are not much different compared to those for BATSE GRBs. Similar to BATSE, both AIC and the frequentist comparison test favor the $k = 3$ case over the $k = 2$ case, whereas BIC prefers two components. The model comparison values from Table 2 are $\Delta AIC = 1.7$ which favor the $k = 3$ model very weakly as compared to the $k = 2$ model. Finally, $\Delta BIC = 10$, which corresponds to strong evidence. The frequentist test using χ^2 probability prefers three components with p-value $= 0.3$, which only corresponds to 0.5σ significance.

Therefore, to summarize, AIC and frequentist model comparison for Fermi-GBM GRBs prefer three components, whereas BIC prefers two. However, the statistical significance of all the three tests is marginal, thus implying that both hypothesis cannot be easily distinguished. Our results also agree with the analysis of Tarnopolski (2015), who also compared χ^2 probabilities by carrying out a binned analysis of the duration distribution. He concluded that although the χ^2 probability for a 3-component fit is more than a 2-component one, the p-values range from 14-77% for different values of the binning. Therefore from this analysis, there is no evidence that the third peak is statistically significant. The GMM based analysis by Zhang et al. (2016b) shows a preference for two components with $\Delta BIC=13$.

4.3 BeppoSAX:

The BeppoSAX catalogue (Frontera et al. 2009) has a total of 1003 GRBs detected between 1996 and 2001.
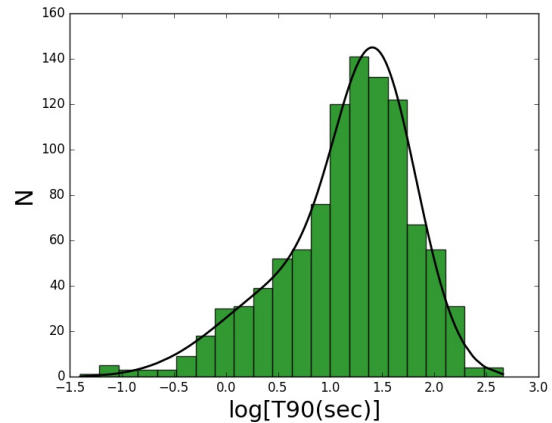
Fig. 5 A fit for the 2-component model for BeppoSAX GRBs. Summary of the fits can be found in Table 3.
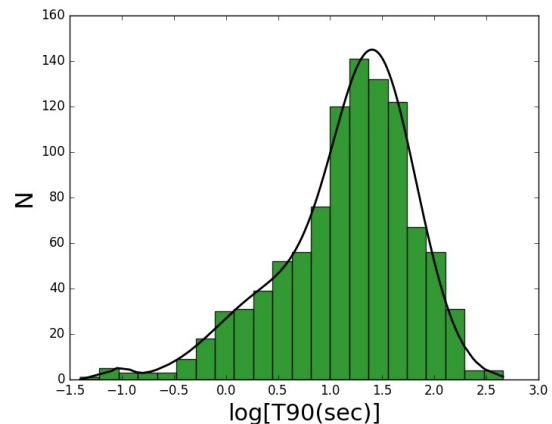
Fig. 6 A fit for the 3-component model for BeppoSAX GRBs. Summary of the fits can be found in Table 3.

Table 3 Model Comparison Parameters for BeppoSAX GRBs. Explanation of all columns can be found in Table 1. The preferred value for every test is highlighted in bold. We find that BIC and χ^2 probability favor two components, whereas AIC favors three. However with all these model comparison techniques, the statistical significance is marginal.

k	μ_{T90}	σ_{T90}	w_i	$\mathcal{L}$	χ^2/ν	$P(\chi^2,\nu)$	AIC	BIC	p-value	$\Delta(AIC)$	$\Delta(BIC)$
2	0.626	0.668	356	6011.6	0.9207	**0.00185**	1847	**1867**			
	1.45	0.392	647								
3	-1.013	0.1	6						0.256(0.6σ)	1.5	8
	0.4307	0.530	259	6014.7	0.9195	0.00177	**1845.5**	1875			
	1.43	0.404	738								

The results of our fits for two and three components are shown in Figs. 5 and Figs. 6 respectively. A tabular summary of our model comparison tests can be found in Table 3. After doing a three-component fit, we find that the GRBs from BeppoSAX are mainly divided into intermediate and long GRB segments (259 and 738 respectively) leaving only 6 GRBs in the short segments. For a two-component fit, we find that 356 GRBs belong to short category and 647 to long. From the data presented in Table no. 3, we do not have a strong consensus to decide the preferred model. Both the χ^2 probability and BIC prefer $k = 2$, whereas AIC shows a preference for $k = 3$. We find that $\Delta\text{AIC} = 1.5$ corresponding to weak evidence. The χ^2 probability gives a p-value of 0.256, corresponding to only 0.6σ significance. The $\Delta\text{BIC} = 8$ in favor of the $k = 2$ model corresponding to strong evidence.

Therefore, none of the three model comparison tests provide a decisive evidence for the three-component model over the two-component one or vice-versa. A likelihood analysis of the BeppoSAX data by Horváth (2009) showed evidence for three components, but the probability that this is a fluctuation was only 3.7%.

4.4 Swift GRBs:

We analyzed 927 Swift (BAT) GRBs detected between Nov. 2004 and Sept. 2016 (Lien et al. 2016). The results from our likelihood fits for $k = 2$ and $k = 3$ are shown in Figs. 7 and 8 respectively. We see from Figs. 7 and 8 that the $k = 3$ model gets preferred over the $k = 2$ model but only slightly. After fitting for two components, we find that 200 and 727 belong to the short and long category respectively. On doing the same for three components, we find that 74, 266, and 587 belong to short, intermediate, and long class respectively. From Table 4, we find that the Swift catalogue prefers the $k = 3$ case over $k = 2$ case with all the three tests. We find that both $\Delta BIC = 18$ and $\Delta AIC = 28$ are greater than 10, which corresponds to decisive evidence for a 3-component model compared
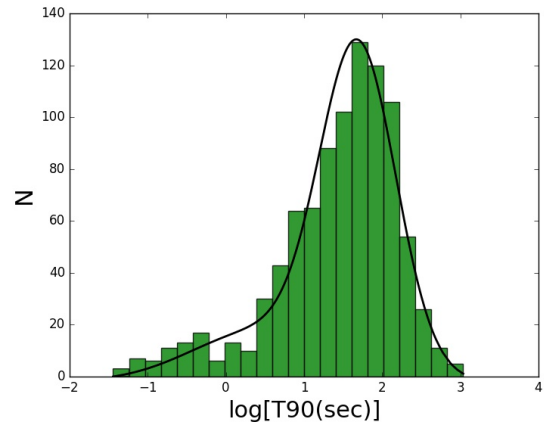
Fig. 7 A fit for the 2-component model for Swift GRBs. Summary of the fits can be found in Table 4.
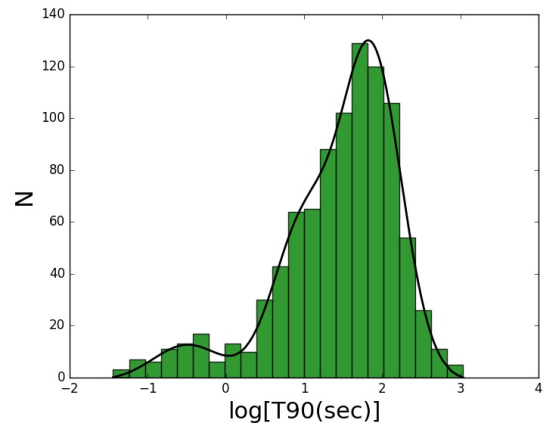
Fig. 8 A fit for the 3-component model for Swift GRBs. Summary of the fits can be found in Table 4.

Table 4 Model comparison parameters for Swift GRBs. Explanation of all the columns can be found in Table 1. The preferred value for every test is highlighted in bold. All the three tests favor three components compared to two. Both BIC and AIC point to decisive evidence (in terms of significance) for the three components and the significance of three component model compared to two is 2.36σ.

k	μ_{T90}	σ_{T90}	w_i	$\mathcal{L}$	χ^2/ν	$p(\chi^2,\nu)$	AIC	BIC	$p-value$	$\Delta(AIC)$	$\Delta(BIC)$
2	0.422	0.869	200	5354.3	1.075	0.002	1997	2016			
	1.388	0.428	727								
3	-0.492	0.441	74	5371.9	1.061	**0.00388**	**1969**	**1998**	$0.009(2.36\sigma)$	28	18
	0.98	0.385	266								
	1.857	0.388	587								

a 2-component one. From our frequentist model comparison test, the χ^2 probability is higher for the three-component model with a p-value of 0.009 corresponding to 2.36σ significance. More data is necessary to see if the significance enhances with increased data sample.

Therefore, both the information criterion based model comparison tests point to decisive (or very strong evidence) for a trifurcated GRB data sample (based on its duration). The significance of the third component from the frequentist model comparison test is 2.36σ. As mentioned in the introduction, a large number of groups have analyzed the Swift data over the past decade. Using a maximum likelihood analysis Horváth & Tóth (2016) finds that three distributions fit the data better than two with 99.9999% (4.75σ) significance. On comparing maximum likelihood, AIC, and BIC, Tarnopolski (2016b) also finds three distributions fit the data better than two. The results from GMM based analysis are also consistent with three distributions with ΔBIC of about 6 (Zhang et al. 2016b). Therefore, our results agree with recent T90 based classifications of the Swift dataset by other authors.

4.5 Swift GRBs in rest frame

We now carry out a similar classification of the intrinsic durations of the GRBs by taking into account the measured redshifts. The Swift GRB catalog consists of 323 detections with redshifts. All the other detectors have less than 100 GRBs with measured redshifts. So we restrict this analysis to the intrinsic $T90$ distribution of only the Swift GRB sample. The intrinsic durations for a GRB is given by:

$$T_{90_{int}} = \frac{T_{90_{obs}}}{1+z} \tag{11}$$

where as indicated, $T_{90_{obs}}$ are the measured T90 values, $T_{90_{int}}$ the intrinsic T90 values, and z being the redshift for the GRB. The model fits for the intrinsic T90 for the Swift GRBs are shown in Fig. 9 and Fig. 10 respectively for the $k = 2$ and $k = 3$ case and a tabular summary of the model comparison results in Tab. 5.

On fitting the data to two components, we find that 163 and 160 GRBs belong to the short and long category respectively. On doing the same for three components, 1, 185, and 137 fit in short, intermediate, and long categories respectively.

The intrinsic T90 distribution for the Swift GRBs weakly prefers the $k = 2$ case over $k = 3$ case with all the three model-comparison tests. The value of ΔAIC is equal to 1.7, which amounts to weak evidence. Similarly, ΔBIC = 9.8, which corresponds to strong evidence. The frequentist model comparison test also shows a preference for $k = 3$ model with a p-value of 0.518 corresponding to less than 0.02σ. In summary, we can say that the preferred model at first look is the $k = 2$ model for the intrinsic GRB case but its significance is low with all the model comparison tests used. When a similar analysis of the intrinsic T90 distribution of 347 Swift GRBs was done by Tarnopolski (2016b), he found that AIC points to three components (albeit with very weak evidence against two), and BIC yielded a very strong support for two components. The GMM-based analysis also showed evidence for two components with ΔBIC of about 6.

5 Conclusions

The main goal of this paper was to investigate the existence of an intermediate class of Gamma Ray Bursts, in addition to the pre-existing short and long class of GRBs, as previously argued by several authors. We did a comprehensive analysis of the T90 distributions of GRBs from all the major instruments used to detect them in the past three decades, by fitting the data to two as well as three lognormal distributions. We then conducted three statistical tests to ascertain the best model among these two hypotheses. These tests include AIC, BIC, and a frequentist model comparison test based on χ^2 probability. The statistical significance from the information criterion based tests was obtained

Table 5 Model comparison parameters for the intrinsic durations detected by Swift GRBs after incorporating the measured redshifts. Explanation of all the columns is same as in Table 1. The preferred values are highlighted in bold. All tests prefer two components. However the significance is marginal in all the cases.

k	μ_{T90}	σ_{T90}	w_i	$\mathcal{L}$	χ^2/ν	$P(\chi^2,\nu)$	AIC	BIC	p-value	$\Delta(AIC)$	$\Delta(BIC)$
2	0.761	0.826	163	1521	1.0823	**8.75e-3**	**698.5**	**713**			
	1.439	0.4163	160								
3	-1.783	0.1	1						$0.518(< 0.02\sigma)$	1.7	9.8
	0.8239	0.7865	185	1522.1	1.085	8.49e-3	700.2	722.8			
	1.485	0.3798	137								

Table 6 Summary of model comparison tests for all the different GRB datasets analyzed. We find that only for the Swift GRBs (in the observer frame) three components are preferred with very decisive evidence using information-criterion-based tests and 2.36σ significance from frequentist model comparison tests.

Dataset	p-value (from χ^2 probability)		ΔAIC		ΔBIC	
	Model preferred	Magnitude	Model preferred	Magnitude	Model preferred	Magnitude
BATSE	3	0.109	3	6.5	2	5
Fermi	3	0.321	3	1.7	2	10
BeppoSAX	2	0.256	3	1.5	2	8
Swift	3	0.009 (2.36σ)	3	28	3	18
Intrinsic Swift	2	0.518	2	1.7	2	9.8
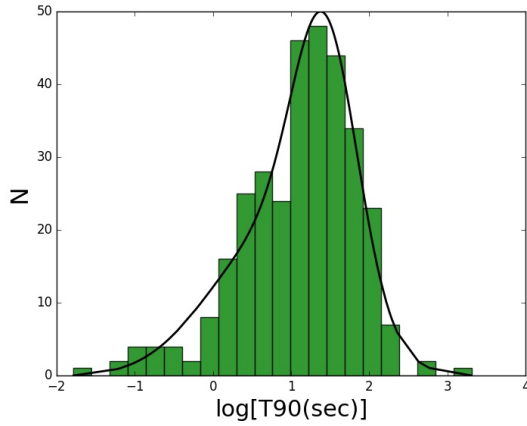
Fig. 9 A fit for the 2-component model for intrinsic red shifted Swift GRBs. Summary of the fits can be found in Table 5.
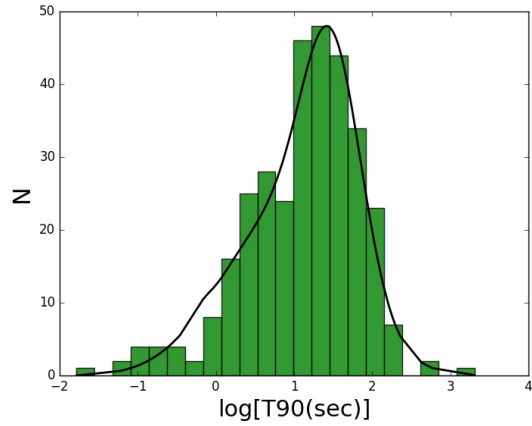
Fig. 10 A fit for the 3-component model for intrinsic red shifted Swift GRBs. Summary of the fits can be found in Table 5.

using empirical strength of evidence rules. From the frequentist test, significance was obtained by using Wilk's theorem.

Our results for each of the detectors are as follows. A tabular summary of all these results can be found in Table 6.

1. For the BATSE dataset, we find that the frequentist model-comparison test and AIC prefer three components, whereas BIC prefers two. However, the significance from all the three tests is marginal and hence the evidence for the third component is weak.

2. The results of model comparison tests for Fermi-GBM are same as that for BATSE. Both AIC and the frequentist model comparison test prefer three components, whereas BIC prefers two. However, the statistical significance from all these tests is quite weak and no decisive evidence can be made.

3. For BeppoSAX GRBs, AIC prefers three components, whereas BIC and frequentist model comparison tests prefer two. However the statistical significance of each of these tests is marginal and no decisive evidence can be made either way.

4. For Swift GRBs, all three tests favor three components. The statistical significance of the third component from the frequentist model comparison test is about 2.4σ. Both ΔAIC and ΔBIC value points to decisive evidence for a third component.

5. Since a large number of Swift GRBs have measured redshifts, we redid the classification on the intrinsic T90 distribution for the Swift GRBs. All the three tests favor two GRB components. The statistical significance though is very marginal.

Therefore in conclusion, we find that none of the detectors show consistent results in accord with previous findings (Gehrels et al. 2009; Zhang et al. 2016b). From our model comparison tests we find that only for Swift GRBs, the observation duration is consistent with three components at about 2.4σ level using frequentist model comparison tests and this is corroborated by BIC and AIC which also point to very strong evidence for three components. However, when we carry out the same test with the intrinsic T90 distribution for Swift GRBs, we do not find evidence for the third component. For all the other detectors, the evidence for the third component is either very marginal or disfavored.

References

Bromberg, O., Nakar, E., Piran, T., & Sari, R. 2013, Astrophys. J., 764, 179

Burnham, K. P., & Anderson, D. R. 2004, Sociological methods & research, 33, 261

Chattopadhyay, T., Misra, R., Chattopadhyay, A. K., & Naskar, M. 2007, Astrophys. J., 667, 1017

Desai, S. 2016, EPL (Europhysics Letters), 115, 20006

Desai, S., & Liu, D. W. 2016, Astroparticle Physics, 82, 86

Frontera, F., Guidorzi, C., Montanari, E., et al. 2009, Astrophys. J. Suppl. Ser., 180, 192

Gehrels, N., Ramirez-Ruiz, E., & Fox, D. B. 2009, Annu. Rev. Astron. Astrophys., 47, 567

Horváth, I. 1998, Astrophys. J., 508, 757

—. 2002, Astron. Astrophys., 392, 791

—. 2009, Astrophys. Space Sci., 323, 83

Horváth, I., Bagoly, Z., Balázs, L. G., et al. 2010, Astrophys. J., 713, 552

Horváth, I., Balázs, L. G., Bagoly, Z., & Veres, P. 2008, Astron. Astrophys., 489, L1

Horváth, I., & Tóth, B. G. 2016, Astrophys. Space Sci., 361, 155

Huja, D., Mészáros, A., & Řípa, J. 2009, Astron. Astrophys., 504, 67

Ivezić, Ž., Connolly, A., Vanderplas, J., & Gray, A. 2014, Statistics, Data Mining and Machine Learning in Astronomy (Princeton University Press)

Kass, R. E., & Raftery, A. E. 1995, Journal of the American Statistical Association, 90, 773

Koshut, T. M., Paciesas, W. S., Kouveliotou, C., et al. 1996, Astrophys. J., 463, 570

Kouveliotou, C., Koshut, T., Briggs, M. S., et al. 1996, in American Institute of Physics Conference Series, Vol. 384, American Institute of Physics Conference Series, ed. C. Kouveliotou, M. F. Briggs, & G. J. Fishman, 42–46

Kouveliotou, C., Meegan, C. A., Fishman, G. J., et al. 1993, Astrophys. J. Lett., 413, L101

Li, Y., Zhang, B., & Lü, H.-J. 2016, ArXiv e-prints, arXiv:1608.03383

Liddle, A. R. 2004, Mon. Not. R. Astron. Soc., 351, L49

—. 2007, Mon. Not. R. Astron. Soc., 377, L74

Lien, A., Sakamoto, T., Barthelmy, S. D., et al. 2016, Astrophys. J., 829, 7

Lyons, L. 2016, ArXiv e-prints, arXiv:1607.03549

Mazets, E. P., Golenetskii, S. V., Ilinskii, V. N., et al. 1981, Astrophys. Space Sci., 80, 3

McBreen, B., Hurley, K. J., Long, R., & Metcalfe, L. 1994, Mon. Not. R. Astron. Soc., 271, 662

Mukherjee, S., Feigelson, E. D., Jogesh Babu, G., et al. 1998, Astrophys. J., 508, 314

Nakar, E. 2007, Phys. Rep., 442, 166

Narayana Bhat, P., Meegan, C. A., von Kienlin, A., et al. 2016, Astrophys. J. Suppl. Ser., 223, 28

Paciesas, W. S., Meegan, C. A., Pendleton, G. N., et al. 1999, Astrophys. J. Suppl. Ser., 122, 465

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical recipes in FORTRAN. The art of scientific computing

Shafer, D. L. 2015, Phys. Rev. D, 91, 103516

Shi, K., Huang, Y. F., & Lu, T. 2012, Mon. Not. R. Astron. Soc., 426, 2452

Tarnopolski, M. 2015, Astron. Astrophys., 581, A29

—. 2016a, Astrophys. Space Sci., 361, 125

—. 2016b, New Astron., 46, 54

Wilks, S. S. 1938, Annals Math. Statist., 9, 60

Woosley, S. E., & Bloom, J. S. 2006, Annu. Rev. Astron. Astrophys., 44, 507

Yang, E. B., Zhang, Z. B., & Jiang, X. X. 2016, Astrophys. Space Sci., 361, 257

Zhang, B. 2006, Nature, 444, 1010

Zhang, B., Lü, H.-J., & Liang, E.-W. 2016a, ArXiv e-prints, arXiv:1611.01948

Zhang, Z.-B., & Choi, C.-S. 2008, Astron. Astrophys., 484, 293

Zhang, Z.-B., Yang, E.-B., Choi, C.-S., & Chang, H.-Y. 2016b, Mon. Not. R. Astron. Soc., 462, 3243

Zitouni, H., Guessoum, N., Azzam, W. J., & Mochkovitch, R. 2015, Astrophys. Space Sci., 357, 7