Journal of High Energy Astrophysics 18 (2018) 21–34

Contents lists available at ScienceDirect

Journal of High Energy Astrophysics

www.elsevier.com/locate/jheap

Possible correlations between gamma-ray burst and its host galaxy offset

Fei-Fei Wang [a], Yuan-Chuan Zou [a,*], Yu Liu [a], Bin Liao [a], Reetanjali Moharana [b]

[a] School of Physics, Huazhong University of Science and Technology, Wuhan 430074, China
[b] Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel

ARTICLE INFO

Article history:
Received 23 January 2018
Accepted 18 March 2018
Available online xxxx

Keywords:
Gamma-ray bursts
Statistics
Host galaxy offsets
T_{90}
T_{R45}
$E_{\gamma,\mathrm{iso}}$

ABSTRACT

We collected the information of 304 gamma-ray bursts (GRBs) from the literature, and analyzed the correlations among the host galaxy offsets (the distance from the site of the GRB to the center of its host galaxy), $T_{90,\mathrm{i}}$ (the duration T_{90} in rest-frame), $T_{\mathrm{R45,i}}$ (the duration T_{R45} in rest-frame), $E_{\gamma,\mathrm{iso}}$ (the isotropic equivalent energy), $L_{\gamma,\mathrm{iso}}$ ($= E_{\gamma,\mathrm{iso}}/T_{90,\mathrm{i}}$, the isotropic equivalent luminosity) and L_{pk} (peak luminosity). We found that $T_{90,\mathrm{i}}$, $T_{\mathrm{R45,i}}$, $E_{\gamma,\mathrm{iso}}$, L_{pk} have negative correlation with offset, which is consistent with origin of short GRBs (SGRBs) and long GRBs (LGRBs). On separate analysis, we found similar results for $\log E_{\gamma,\mathrm{iso}} - \log(\mathrm{offset})$ and $\log L_{\mathrm{pk}} - \log(\mathrm{offset})$ relations in case of SGRBs only, while no obvious relation for LGRBs. There is no correlations between offset and $L_{\gamma,\mathrm{iso}}$. We also put the special GRB 170817A and GRB 060218A on the plots. The two GRBs both have low luminosity and small offset. In the log(offset) $-$ $\log T_{90,\mathrm{i}}$ plot, we found GRB 170817A locates in between the two regions of SGRBs and LGRBs and it is the outlier in the offset $- E_{\gamma,\mathrm{iso}}$, offset $- L_{\gamma,\mathrm{iso}}$ and offset $- L_{\mathrm{pk}}$ plots. Together with GRB 060218A being an outlier in all plots, it indicates the speciality of GRBs 170817A and 060218A, and might imply more subgroups of the GRB samples.

1. Introduction

Gamma-ray bursts (GRBs) are widely accepted to have two categories, short GRBs (SGRBs) having duration shorter than 2 s and long GRBs (LGRBs) with duration longer than 2 s (Kouveliotou et al., 1993). SGRBs are thought to be from the merger of compact object binaries involving at least one neutron star (Eichler et al., 1989; Paczynski, 1991; Narayan et al., 1992), and have a broad range of spatial host galaxy distribution (Zhang et al., 2017b). The origin of LGRBs are most-likely to be the collapse of rapidly-rotating massive stars (MacFadyen and Woosley, 1999), hence expected to be inside the star forming region. Consequently, the offsets of the location in the host galaxy of LGRBs are mostly smaller than those of SGRBs.

In the past few decades, there have been many studies on host galaxy offsets of GRBs. For example, Bloom et al. (2002) studied host galaxy offsets for LGRBs. The result was consistent with the expected distribution of massive stars, confirming the core-collapse model as the origin of LGRBs. Fong et al. (2010) presented the first comprehensive analysis of *Hubble Space Telescope* (*HST*) observations of ten SGRBs host galaxies. Their result showed an median at 5 kpc for SGRBs host galaxy offsets, which is about 5 times larger than LGRBs. There was no evidence of differences between SGRBs with and without extended emission. The host galaxy offsets are in good agreement with neutron star binary mergers (see also Church et al., 2011). However, Malesani et al. (2007) noticed that SHBs (short hard GRBs) with extended emission are more easier to detect their optical counterparts. This has been explained as an environmental property by Troja et al. (2008), as SHBs with extended emission seem to occur closer to their host galaxies, in denser interstellar environments. This also implies that SGRBs progenitors have an intrinsically different behavior, due to their association with different origins such as black hole (BH)-neutron star (NS) and NS–NS merger. Troja et al. (2008) showed that SGRBs with extended hard X-ray emissions that have small projected physical offsets may be due to NS–BH mergers, while those without extended hard X-ray emission components that have bigger projected physical offsets may be due to NS–NS mergers. The correlation between X-ray absorption column densities and host galaxy offsets gives another evidence that SGRBs possibly have two distinct populations (Kopač et al., 2012). Furthermore, some negative correlations are found between the broadband afterglow emissions and SGRBs host galaxy offsets (Zhang et al., 2017b). This is because the afterglow emission depends on the circum-burst medium and it decreases with the distance to the host galaxy cen-

* Corresponding author.
E-mail address: zouyc@hust.edu.cn (Y.-C. Zou).

https://doi.org/10.1016/j.jheap.2018.03.001

ter, providing more evidences that SGRBs with larger host galaxy offsets prefer lower circum-burst densities (Fong et al., 2015).

To investigate the properties of the host galaxies and the connection to the GRBs, we collect all the possible sample from the literature about the offsets, durations of the GRBs (T_{90} (time duration from 5% photon counts to 95% photon counts) and T_{R45} (defined in Reichart et al., 2001)), the isotropic equivalent γ-ray energy $E_{\gamma,\rm iso}$, and the 1 s time binned peak luminosity $L_{\rm pk}$. In this work we analyze these data and present our results for the relations found for SGRBs, LGRBs and combination of them. The paper is organized as the follow: the data is collected and described in §2, the statistics is performed in §3, and conclusion and discussion is given in §4.

2. The GRB sample

We selected 304 GRBs from different instruments, and collected their trigger time, instrument, redshift z, offset, T_{90}, T_{R45}, $E_{\gamma,\rm iso}$ and $L_{\rm pk}$ values from different published papers. All the information is provided in Table 3 and 4. $E_{\gamma,\rm iso}$ and $L_{\rm pk}$ are in rest-frame 1–10^4 keV energy band, and $L_{\rm pk}$ is in 1 s time bin (except GRB 170817A in 50 ms time bin). We also calculated isotropic equivalent luminosity $L_{\gamma,\rm iso}$ in the rest-frame 1–10^4 keV energy band, which is $L_{\gamma,\rm iso} = (1+z)E_{\gamma,\rm iso}/T_{90}$. For $L_{\rm pk}$, sometimes the energy band is not in rest-frame 1–10^4 keV energy band, like Deng et al. (2016). We changed the energy band using the spectral information. There are mainly three kinds of spectral models: Band model, cutoff power law (CPL) model and simple power law (SPL) model (more details in Li et al., 2016). In Table 4, we gave the GRB spectral information which need to change the energy band, as well as the $L_{\rm pk}$. For Band model, α, β and $E_{\rm pk}$ are low energy spectral index, high energy spectral index and peak energy, respectively. For CPL model, α is the spectral index for the power law band and $E_{\rm pk}$ is the cutoff energy. There is no β for CPL model, and we use "..." to remark β. Besides, we excluded some values with lower limit smaller than 0. For example, the offset of GRB 120119A is 0.104 ± 0.147 (Li et al., 2016).

The data are not complete, as not every GRB has all the observational values listed above, available. Some of the data have only the central values available without error bars. To keep the information of the central values, we need to impute the errors from other data. We used the R package *mice* to impute the error bars for the data that have just the central values, by multiple imputation with chained equations (MICE) (Rubin, 1987, 1996).

2.1. Error imputation

We use the R package *mice* to impute incomplete multivariate data by using the method MICE. MICE is a powerful tool for imputation and it has been widely used. Only the central values with missing error bars are imputed. The ones with missing central values are omitted in the statistical analysis. According to Rubin (1987, 1996), MICE includes three steps: generating multiple imputation, analyzing imputed data, and pooling analysis results.

The imputation model should also have three principles: accounting for the process that created the missing data, preserving the relations in the data and preserving the uncertainty about these relations. At first, we changed $T_{90,\rm i}$, $T_{R45,\rm i}$, $E_{\gamma,\rm iso}$ and host galaxy offset into their logarithmic values. Then we did 5 times imputation as suggested in Rubin (1996). It means every error bar which need to be imputed will have 5 imputed values, hence we have 5 complete set of data. We need to choose the imputation model first, because our data is missing at random (MAR) (Rubin, 1976), additionally, our data is numeric type. So we choose the predictive mean matching model (PMM) (Little, 1988), a general purpose semi-parametric imputation method.[1] We set a threshold at 0.25, which means the minimum proportion of usable cases for imputation is at least 0.25. An important step in multiple imputation is that, we want to assess whether imputations are plausible, then we have done diagnostic checks. We used following three indicators to assess the goodness of our imputation results.

1. Relative increase in variance due to missing data $r_{\rm m}$ (RIV): It is the ratio between imputation variance and the imputation variance of the 5 data sets, then multiplying the imputation time m. It stands for the increase fraction in variance due to missing data, the influence of the missing data is bigger when $r_{\rm m}$ is bigger. While smaller $r_{\rm m}$ indicates influence of the change of m is smaller, this is to say that missing data has smaller influence to the whole data parameters, hence the imputation results are more stable and the imputations are better.
$$r_m = \frac{(1+\frac{1}{m})\sigma_{\rm B}{}^2}{\sigma_{\rm W}{}^2}. \tag{1}$$
$\sigma_{\rm W}{}^2$ is within-imputation variance, it represents the mean of the variance for the m data sets.
$$\sigma_{\rm W}{}^2 = \frac{1}{m}\sum_{i=1}^{m}\sigma_{\rm i}{}^2.$$
$\sigma_{\rm B}{}^2$ is between-imputation variance, it represents the variance of the mean of m data sets.
$$\sigma_{\rm B}{}^2 = \frac{1}{m-1}\sum_{i=1}^{m}(\widehat{\theta_{\rm i}}-\widehat{\theta})^2$$
$\widehat{\theta_{\rm i}}$ is the mean of every complete data set, $\widehat{\theta} = \frac{1}{m}\sum_{i=1}^{m}\widehat{\theta_{\rm i}}$.
2. Fraction of missing information $\gamma_{\rm m}$ (FMI): This represents the influence of the missing data for the whole parameters(e.g. mean). Smaller FMI values indicate that the imputation results are more stable.
$$\gamma_{\rm m} = \frac{r_{\rm m}+\frac{2}{v_{\rm m}+3}}{r_{\rm m}+1} \tag{2}$$
$v_{\rm m} = (m-1)(1+\frac{1}{r_{\rm m}}{}^2)$ is the degree of freedom.
3. Relative efficiency (RE): is a comprehensive analysis of RIV and FMI. It represents the imputation fraction for missing information by MICE. The higher value of RE means the better result.
$$\mathrm{RE} = (1+\frac{\gamma_{\rm m}}{m})^{-1} \tag{3}$$

For analyzing imputed data and pooling analysis results, we use the mean of every imputed error bar, because we also need to calculate some values and plot scatter plots with error bars. As there are 5 candidate values for each parameter, we use the mean of them as the imputed error.

The imputation results are shown in Table 1. From the results, we can see that RIV and FMI are very close to 0, which means our

[1] We compared the correlations between the central values and the related errors, and found the PMM is reliable in the error imputation. For example, the positive error of $T_{90,\rm i}$ is $T_{90,\rm i,1}$. Before imputation, the linear regression between $T_{90,\rm i}$ and $T_{90,\rm i,1}$ is $T_{90,\rm i} = (1.26\pm0.05)+(-9.28\pm1.58)\times T_{90,\rm i,1}$, and the Pearson coefficient is -0.33 ± 0.05 with p-value 1.2×10^{-8}. After the imputation, the linear regression between $T_{90,\rm i}$ and $T_{90,\rm i,1}$ is $T_{90,\rm i} = (1.27\pm0.05)+(-9.7\pm1.58)\times T_{90,\rm i,1}$, and the Pearson coefficient is -0.33 ± 0.05 with p-value 1.6×10^{-9}. The results do not change too much, which means PMM model can preserve the relations in the data and preserve the uncertainty about these relations.