

How unbiased statistical methods lead to biased scientific discoveries: A case study of the Efron-Petrosian statistic applied to the luminosity-redshift evolution of Gamma-Ray Bursts

Christopher Bryant^{1*} Joshua Alexander Osborne^{1†} Amir Shahmoradi^{1,2‡}

¹*Department of Physics, College of Science, The University of Texas, Arlington, TX 76010, USA*

²*Data Science Program, College of Science, The University of Texas, Arlington, TX 76010, USA*

ABSTRACT

Statistical methods are frequently built upon assumptions that limit their applicability to certain problems and conditions. Failure to recognize these limitations can lead to conclusions that may be inaccurate or biased. An example of such methods is the non-parametric Efron-Petrosian test statistic used in the studies of truncated data. We argue and show how the inappropriate use of this statistical method can lead to biased conclusions when the assumptions under which the method is valid do not hold. We do so by reinvestigating the evidence recently provided by multiple independent reports on the evolution of the luminosity/energetics distribution of cosmological Long-duration Gamma-Ray Bursts (LGRBs) with redshift. We show that the effects of detection threshold has been likely significantly underestimated in the majority of previous studies. This underestimation of detection threshold leads to severely-incomplete LGRB samples that exhibit strong apparent luminosity-redshift or energetics-redshift correlations. We further confirm our findings by performing extensive Monte Carlo simulations of the cosmic rates and the luminosity/energy distributions of LGRBs and their detection process.

Key words: Gamma-Rays: Bursts – Gamma-Rays: observations – Methods: statistical – Methods: Monte Carlo

1 INTRODUCTION

Gamma-Ray Bursts (GRBs) are the most violent and energetic stellar explosions in the known universe. They radiate huge amounts of gamma-ray energy, comparable to the lifetime energy output of the sun, over a short period of time and are often followed by an afterglow at longer wavelengths (e.g., Mészáros 2006; Zhang 2007; Gehrels, Ramirez-Ruiz & Fox 2009). With their energy concentrated in a collimated beam, they can be seen at much higher redshifts than supernovae (SNe). Amongst other things,

GRBs are excellent tools to probe the Star Formation Rate (SFR) of the early as well as the recent universe.

GRBs are generally divided into two categories: Long-soft GRBs (LGRBs) with $T_{90} \gtrsim 2$ [s], and Short-hard GRBs (SGRBs) with $T_{90} \lesssim 2$ [s].¹ These values are based on population statistics of the *Compton Gamma Ray Observatory*'s Burst and Transient Source Experiment (BATSE) detector, which was decommissioned in 2000 (Kouveliotou et al. 1993; Shahmoradi & Nemiroff 2011; Shahmoradi 2013b,a; Shahmoradi & Nemiroff 2015). LGRBs are believed to be

* E-mail: christopher.bryant@mavs.uta.edu (CMB)

† E-mail: joshua.osborne@uta.edu (JAO)

‡ E-mail: a.shahmoradi@uta.edu (AS) (corresponding author)

¹ T_{90} is the duration over which a burst emits from 5–95% of its total measured gamma-ray photon flux.

the result of the collapse of massive stars into a black hole (Woosley 1993), while SGRBs are theorized to be the result of the merger of two neutron stars or of a neutron star and a black hole (Eichler et al. 1989).

Current research attempts to infer an accurate description and distribution profile of various GRB characteristics, in particular, the class of LGRBs due to their abundant redshift measurements compared to the SGRB class. A recent focus in the community has been on the potential cosmological evolution of LGRB luminosity/energetics $L_{\text{iso}}/E_{\text{iso}}$ with redshift, as well as estimating the cosmic rates of LGRBs. A popular technique to constrain these is based on the non-parametric method of Efron-Petrosian (Efron & Petrosian 1992; Petrosian 2002), which is widely used to study observational data sets subject to truncation and censorship (e.g., Kocevski & Liang 2006; Singal et al. 2011; Dainotti et al. 2013).

Yu et al. (2015) (hereafter Y15), Petrosian, Kitanidis & Kocevski (2015) (hereafter P15), and Lloyd-Ronning, Aykulalp & Johnson (2019) (hereafter L19) use the method of Efron-Petrosian to deduce the local (redshift-decorrelated) luminosity function $\psi(L_0)$ and cosmic GRB formation rate $\rho(z)$ to infer that local GRBs ($z < 1$) are in excess of the SFR. Pescalli et al. (2016) (hereafter P16) follows the same approach as the previous three, however, does not find an excess of GRBs relative to the SFR at low redshifts. Tsvetkova et al. (2017) (hereafter T17) simply deduces the luminosity function and GRB formation rate.

In this work, we hypothesize and provide evidence that the effects of detection threshold in the aforementioned studies might have been significantly underestimated. This underestimation of the detector threshold results in an artificial correlation between the luminosity/energetics of LGRBs and redshifts. This could, in turn, lead to the conclusion that the GRB formation rate is different from the SFR at any redshift.

This paper is organized as follows. Section 2 details the methodology used by the aforementioned papers to deduce luminosity/energy evolution $L(z)/E(z)$ given a sample of GRBs, which leads to $\psi(L_0)$ and $\rho(z)$. Section 3 is a re-analysis of their work (with the exception of P15, for whom we could not locate their data). Section 4 describes our own Monte Carlo simulation of a synthetic population of LGRBs. Finally, section 5 is a discussion of the results of our reanalysis and the implications of our Monte Carlo simulations.

2 METHODS

It is often the case in the analysis of astronomical data that one is faced with reconstructing the joint bivariate or multivariate distributions from truncated data. Truncation can be due to a multitude of factors, most importantly, the

Malmquist-types of biases in the population studies of GRBs (Shahmoradi & Nemiroff 2009, 2011; Shahmoradi 2013b; Shahmoradi & Nemiroff 2015, e.g.). The GRB luminosity and redshift distribution (L, z) is one such set of bivariate data. For simplicity, it is often assumed that the functional form of the total Luminosity Function is separable in the following form,

$$\Psi(z, L) = \rho(z)\psi(L) \quad (1)$$

where $\rho(z)$ is the GRB Formation Rate and $\psi(L)$ is the Luminosity Function. Efron & Petrosian (1992) developed a nonparametric technique for estimating $\rho(z)$ and $\psi(L)$ based on the c^- method of Lynden-Bell (1971). Luminosity is assumed to have a simple power law redshift dependence:

$$L(z) = g(z)L_0 = (1 + z)^\alpha L_0 \quad (2)$$

such that the resulting distribution of L_0 , and hence $\psi(L_0)$ (the local Luminosity Function), becomes independent of redshift.

Consider the data set seen in Figure 1a. Instead of dealing with the entire data set, we deal with subsets that can be constructed independent of the truncation limit that affects the full data set. Each of these subsets includes only the objects for a given range of luminosity and redshift. For the i th data point in the (L, z) data set, we can define J_i as,

$$J_i = \{j | L_j \geq L_i, z_j \leq z_i^{max}\}, \quad (3)$$

where L_i is the i th GRB luminosity and z_i^{max} is the maximum redshift at which a GRB with luminosity L_i can be observed due to the detector threshold limit. This produces the dashed black bounding box seen in Figure 1a. N_i represents the number of GRBs in this region: $N_i \equiv \text{sizeof}\{J_i\}$. This is the same as in the c^- method in Lynden-Bell (1971). R_i is the number of events that have redshift z_j less than z_i ,

$$R_i \equiv \text{sizeof}\{j \in J_i : z_j < z_i\}. \quad (4)$$

We expect the rank R_i of z_i to be uniformly distributed between 1 and N_i if L and z are independent of each other (Efron & Petrosian 1992). The Efron-Petrosian test statistic τ is then,

$$\tau \equiv \frac{\sum_i (R_i - E_i)}{\sqrt{\sum_i V_i}}, \quad (5)$$

where $E_i = (1 + N_i)/2$, $V_i = (N_i^2 - 1)/12$ are the expected mean and the variance of R_i , respectively. This is a specialized version of Kendall's τ statistic. The τ statistic represents the significance of the correlation in the bivariate distribution of the two quantities of interest by taking into account the effects of data truncation created by a detection threshold hard cutoff. It is normally distributed

about a mean of 0 with a standard deviation of 1 (Efron & Petrosian 1992). Hence, a τ of 2 implies a 2σ correlation. Once the existence of a correlation has been confirmed, one simply parameterizes it in some way to remove the correlation. As has been aforementioned, a functional form of L_{iso} as $L \rightarrow L_0 = L/g(z)$ has been chosen in majority of previous studies, where $g(z)$ has a simple power law dependence on z , $g(z) = (1+z)^\alpha$. Then, it is just a matter of varying α until $\tau = 0$.

In this work we use the following cosmological parameters: $H_0 = 71$ [$\text{km s}^{-1} \text{Mpc}^{-1}$], $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ (Shahmoradi & Nemiroff 2011; Shahmoradi 2013b; Shahmoradi & Nemiroff 2015, e.g.).

3 REANALYSES OF PAST WORK

In the following subsections, we attempt to regenerate and reanalyze the findings of several recent papers that present evidence in favor of a strong evolution of the GRB luminosity/energetics with redshift.

3.1 Yu et al. (2015) (Y15)

In their paper, Y15 find an excess of LGRBs at low redshift ($z < 1$), deviating from the SFR. They use the method of Efron & Petrosian (1992) to infer the luminosity function and the cosmic rates of LGRBs based on the observational data collected by *Swift*'s Burst Alert Telescope (BAT) gamma-ray detector. They find that the luminosity function of LGRBs evolves with redshift as Eq. (2) with $\alpha = 2.43^{+0.41}_{-0.38}$. This conclusion is based on the assumption of a flux lower limit of

$$F_{\min} = 2.0 \times 10^{-8} [\text{erg cm}^{-2} \text{s}^{-1}], \quad (6)$$

representing the detector threshold limit of *Swift*'s BAT.

To better understand the role of *Swift*'s BAT detector threshold on the conclusions drawn by Y15, here we attempt to reproduce their analysis of *Swift* data. Figures 1a - 1c depict the distributions of the observational LGRB sample used in Table 1 of Y15. Specifically, the bivariate distribution of the 1-second total isotropic peak flux (L_{iso}) and the redshifts ($1+z$) of LGRBs as shown in plot 1a exhibits an apparently strong correlation. However, much of this correlation is potentially due to the BAT detection threshold effects on the observational sample of LGRBs. To quantify and eliminate the effects of detector threshold, Y15 use the proposed non-parametric methodology of Efron & Petrosian (1992) by assuming a parametrization as seen in Eq. (2) for the luminosity-redshift dependence in the LGRB data, such that the resulting distribution of L_0 becomes independent of redshift. We note an apparent inconsistency between Eq. (5) that we have extracted from Efron &

Petrosian (1992) and Eq. (9) of Y15, where summations are taken in different places.

The value of exponent that we infer from the data set of Y15 is consistent with, although not the same as, their inferred α exponent. Y15 found a value of $\alpha = 2.43^{+0.41}_{-0.38}$ in their analysis. In our reanalysis of their work, we find a value of $\alpha = 2.15^{+0.34}_{-0.37}$. However we do believe that their assumption for the value of the flux lower limit of Eq. (6) is likely a severe underestimation of the detection threshold of *Swift*'s BAT. In reality the detection threshold limit of the BAT detector is far more complex than a simple hard cutoff, but rather, a fuzzy range arising from a multitude of factors.

The *Swift* satellite is very well known for its immensely complex triggering algorithm. To our knowledge, it is comprised of at least three separate detection mechanisms that complement each other (e.g., Fenimore et al. 2003):

(i) The first type of trigger is for short time scales (4 ms to 64 ms). These are traditional triggers (single background), for which about 25,000 combinations of time-energy-focal plane subregions are checked per second.

(ii) The second type of trigger is similar to HETE: fits multiple background regions to remove trends for time scales between 64 ms and 64 s. About 500 combinations for these triggering mechanisms are checked per second. For these rate triggers, false triggers and variable non-GRB sources are also rejected by requiring a new source to be present in an image.

(iii) The third type of trigger works on longer time scales (minutes) and is based on routine images that are made of the field of view.

The entire complexity of the detection mechanism of *Swift*'s BAT, as mentioned above, is summarized in a single value Eq. (6) in the work of Y15. The consequences of choosing this value is most apparent in Figures 1b and 1c, where the data set of detected LGRBs virtually ends before reaching the detection threshold line. This implies that the observational LGRB data set is not constrained by the assumed detection threshold of Y15, which is counterintuitive. We do expect the threshold to soft-truncate the data set, and this truncation likely occurs closer to the central peak of the histogram of Figure 1c.

This severe underestimation of the detection threshold of BAT by Y15 is very well seen in Figure 1d where we plot the redshift-corrected isotropic peak luminosity vs. redshift. The solid black line in this plot represents the redshift-corrected detection threshold. In agreement with our hypothesis in the previous paragraph, we observe that the resulting redshift-corrected detection threshold of Y15 resembles almost a flat line at high redshifts. This is another indication that the inferred relationship between $(1+z)$ and L_{iso} is likely heavily influenced by the improperly-modeled detection threshold of BAT.

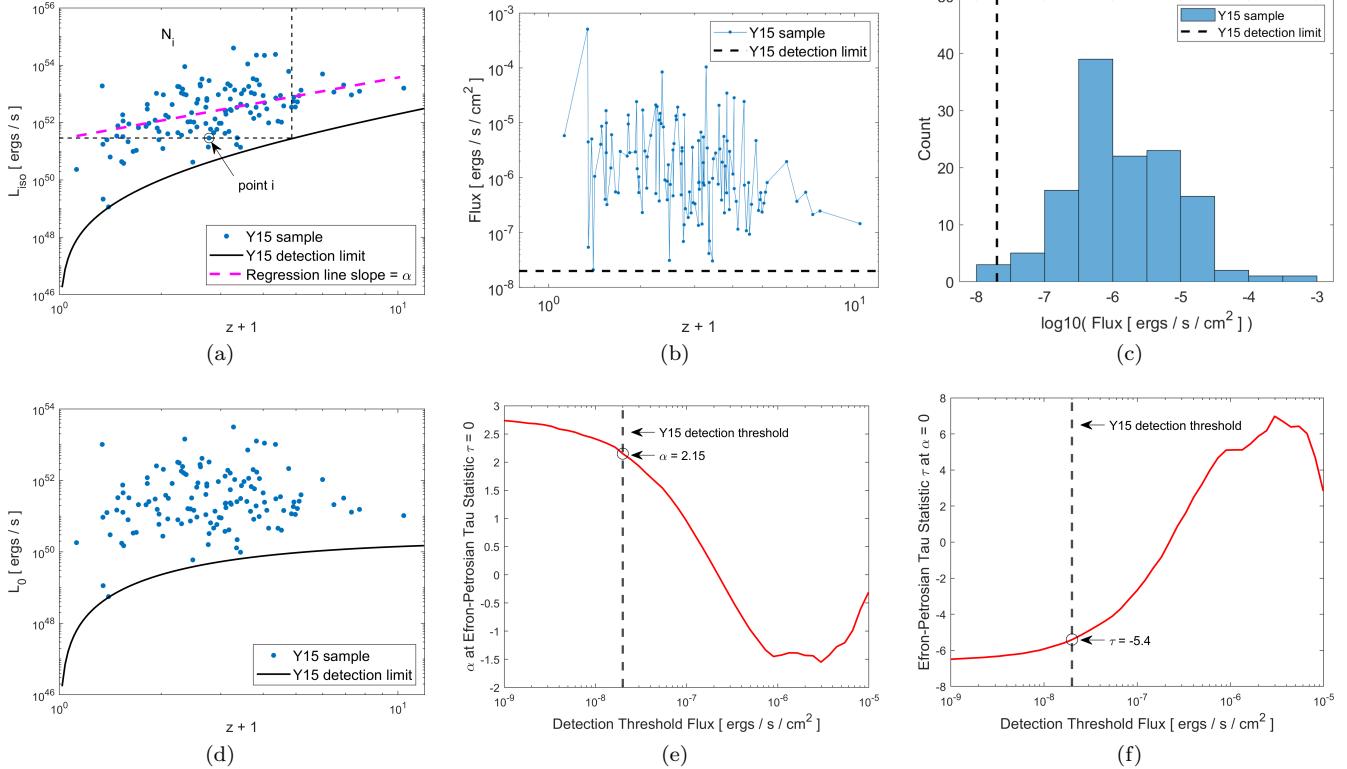


Figure 1. Plotting of the 127 GRBs from Y15. Plot (a) shows isotropic luminosity vs. redshift. The black line represents the observational limit of *Swift* assumed by Y15 to be 2.0×10^{-8} [$\text{erg cm}^{-2} \text{s}^{-1}$]. The purple line represents the linear regression through the data set whose slope is $\alpha = 2.15$. Plot (b) is the observer frame visualization of the Y15 data set, where the dashed line is the observational limit of Y15. Plot (c) is a histogram of flux, where the dashed line is the observational limit. Plot (d) shows redshift vs. $L_0 = L(z)/(1+z)^{2.15}$, the redshift-independent luminosity. Plot (e) shows a range of possible threshold limits vs. α values at $\tau = 0$. The intersection of this line with Y15's threshold limit is our value for their α . Plot (f) shows a range of possible threshold limits vs τ values at $\alpha = 0$. Assuming the detection threshold of Y15, a redshift-independent luminosity distribution is rejected at 5.4σ . However, choosing a more conservative detection threshold at 2×10^{-7} [$\text{erg cm}^{-2} \text{s}^{-1}$] yields no evidence for luminosity-redshift evolution.

A range of threshold values can be seen in Figures 1e and 1f, as well as their effects on α at $\tau = 0$ and, on τ at $\alpha = 0$, respectively.

To gain a better insight into the effects of detection threshold on observational data, we have reproduced parts of the results of Shahmoradi & Nemiroff (2015) in Figure 2. This figure illustrates well the subtle fuzzy effects of the BATSE detector threshold on the observed distribution of the energetics of BATSE LGRBs and SGRBs, including the observed peak flux distribution for which a sharp detection threshold cutoff is frequently assumed. The detection threshold causes a deviation in the observed data from the inferred underlying population, to begin just to the right of the histogram peak (the solid lines). This deviation becomes more significant as one moves leftward. Y15 chose their detector threshold to begin far to the left of the histogram peak, when in reality, it should have been chosen close to

the peak.

Finally, we turn our attention to Y15's Monte Carlo simulations, which seemingly confirms their results. Therein, they begin with their inferred value of α to simulate a distribution of LGRBs following the relationship of Eq. (2) with $\alpha = 2.43$. They find that the synthetic data and the observed data have similar distributions. This is, however, no surprise considering their simulation was based on the derived results of their observational analysis and the assumed potentially-underestimated detection threshold. In other words, their Monte Carlo simulation proves the self-consistency of the Efron-Petrosian statistic and their analysis, but falls short of verifying the accuracy of the detection threshold assumption made in their analysis. Therefore, this circular reasoning concludes that the two observed and synthetic distributions are similar without validating the accuracies of the assumptions made in their

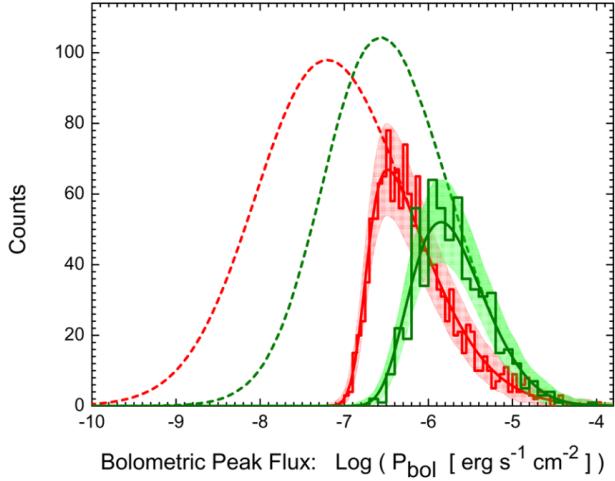


Figure 2. The red and green colors represent data and model for LGRBs and SGRBs, respectively, for BATSE catalogue GRBs. The solid curves illustrate the projection of the multivariate GRB world models of Shahmoradi (2013b); Shahmoradi & Nemiroff (2015) on the BATSE-catalog peak flux P_{bol} distribution, subject to the BATSE detection threshold. The colour-shaded areas represent the 90 percent prediction intervals for the distribution of BATSE data. The dashed lines represent the predicted underlying populations of LGRBs and SGRBs, respectively based on the multivariate GRB world models of Shahmoradi (2013b); Shahmoradi & Nemiroff (2015).

work.

Their primary conclusion of Y15, of an excess of GRBs at low redshift ($z < 1$) compared to the SFR, also contradicts previous studies based on the properties of GRB host galaxies. In point of fact, Vergani et al. (2015), Perley et al. (2015, 2016a,b), Kruhler et al. (2015), and Schulze et al. (2015) performed spectroscopic and multi-wavelength studies on the properties (stellar masses, luminosities, SFR, and metallicity) of GRB host galaxies of various complete GRB samples and compared them to those of the star-forming galaxies selected by galaxy surveys. Their collective results indicate that at low redshift ($z < 1$) only a small fraction of the star formation produces GRBs Pescalli et al. (2016).

3.2 Pescalli et al. (2016) (P16)

P16 proceeds in a similar fashion to Y15, beginning with the observational data set of LGRBs found in their Table B.1. We extract from this data set 81 LGRBs that have both redshift and isotropic peak luminosity (L_{iso}) values for our reanalysis. They use the Efron-Petrosian τ statistic to find an α value of $\alpha = 2.5$, consistent with the results of Y15. This provides the functional form of the luminosity evolution with redshift $L(z) = L_0(1 + z)^{2.5}$. From here they proceed with Lynden-Bell's c^- method to derive the cumulative luminosity function $\Phi(L_0)$ and the LGRB

formation rate $\rho(z)$.

These results, however, are predicated based on a value of detector threshold which P16 gives as,

$$P_{\text{lim}} = 2.6 \text{ [photons cm}^{-2} \text{ s}^{-1} \text{]}. \quad (7)$$

This corresponds to an instrument that is ~ 6 times less sensitive than *Swift*'s BAT (Salvaterra et al. 2012). In their work, P16 adopt a slightly different approach to modeling the flux limit of their sample. The quantities L_{lim} and z_{max} that are used in the Efron-Petrosian statistic are computed by adopting individual spectral and temporal properties of LGRBs and applying the corresponding K-corrections. This approach results in a small scatter in the energy flux-limit of their $L_{\text{iso}} - (z + 1)$ plane. However, they find that this non-uniqueness of the detection threshold has a very small impact in the computation of their τ statistic.

Given the lack of sufficient details about the approach proposed by P16 and the fact they find almost no difference between the traditional approach to computing the Efron-Petrosian statistic and their proposed method, here we follow the traditional formal technique for computing the τ statistic. However, since P16 do not provide an effective energy flux limit for their sample in units of $[\text{erg cm}^{-2} \text{ s}^{-1}]$, we searched for an effective energy-flux detection threshold that would yield a τ statistic comparable to what P16 find.

We find that an effective detection threshold of $F_{\text{min}} = 1.0 \times 10^{-7} \text{ [erg cm}^{-2} \text{ s}^{-1} \text{]}$ yields a regression slope of $L_{\text{iso}} - (z + 1)$ correlation of $\alpha = 2.53$ which is very close to the reported value by P16. Alternatively, we also compute the corresponding detection threshold energy flux by converting the photon flux threshold of Eq. (7) to an effective energy flux limit by assuming a Band model (Band et al. 1993) of the form,

$$\Phi(E) \propto \begin{cases} E^{\alpha_{\text{ph}}} e^{(-\frac{(2+\alpha_{\text{ph}})E}{E_p})} & \text{if } E \leq (E_p)(\frac{\alpha_{\text{ph}} - \beta_{\text{ph}}}{2+\alpha_{\text{ph}}}), \\ E^{\beta_{\text{ph}}} & \text{if otherwise.} \end{cases} \quad (8)$$

with low- and high-energy spectral indices of $\alpha_{\text{ph}} = -1$ and $\beta_{\text{ph}} = -2.25$, respectively, taken from Salvaterra et al. (2012) and an effective spectral peak energy fixed to the average observed spectral peak energy of the LGRB sample of P16 ($E_p = 574 \text{ keV}$). Using this approach, we obtain an effective energy flux limit ($F_{\text{min}} = 2.34 \times 10^{-7} \text{ [erg cm}^{-2} \text{ s}^{-1} \text{]}$) for the LGRB sample that results in a completely different value, $\alpha = 1.35$, for the regression slope of the $L_{\text{iso}} - (z + 1)$ relation than what is reported by P16.

Therefore, we conclude that the effective detection threshold ($F_{\text{min}} = 1.0 \times 10^{-7} \text{ [erg cm}^{-2} \text{ s}^{-1} \text{]}$) that we previously inferred directly from the Efron-Petrosian

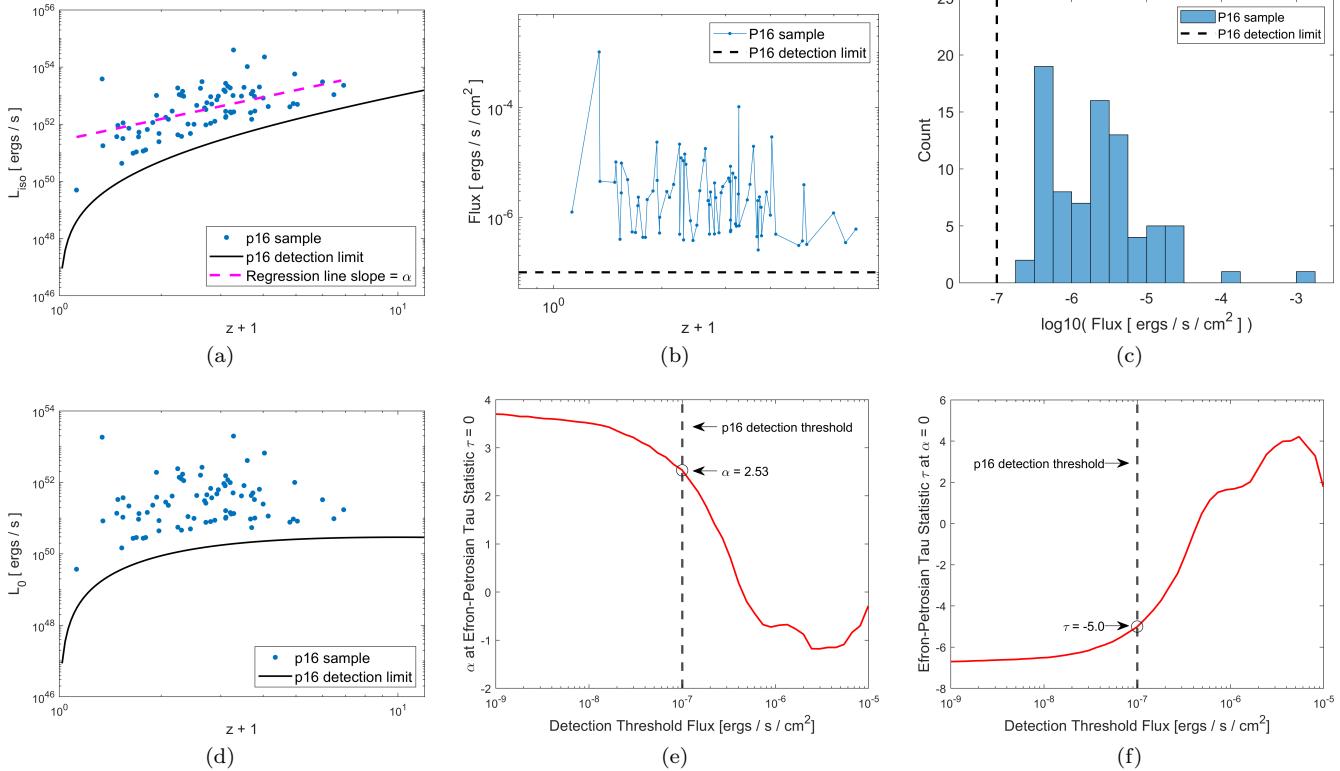


Figure 3. Plotting of the 81 GRBs from P16. Plot (a) shows redshift vs. isotropic luminosity. The black line represents the observational limit of *Swift*, which has been deduced from P16 to be 1.0×10^{-7} [$\text{erg cm}^{-2} \text{s}^{-1}$]. The purple line represents the linear regression through the data set whose slope is $\alpha = 2.53$. Plot (b) is the observer frame visualization of the P16 data set, where the dashed line is the observational limit. Plot (c) is a histogram of flux, where the dashed line is the observational limit. Plot (d) shows redshift vs. $L_0 = L(z)/(1+z)^{2.53}$, the redshift-independent luminosity. Plot (e) shows a range of possible threshold limits vs. α values at $\tau = 0$. The intersection of this line with P16's supposed threshold limit is our value for their α . Plot (f) shows a range of possible threshold limits vs τ values at $\alpha = 0$. Assuming the detection threshold inferred from the analysis of P16, a redshift-independent luminosity distribution is rejected at 5σ . However, choosing a more conservative detection threshold at 4×10^{-7} [$\text{erg cm}^{-2} \text{s}^{-1}$] yields no evidence for luminosity-redshift evolution.

statistic should likely resemble more the flux limit that is used but not clearly discussed or shown in the work of P16. However, the trade off in choosing this value is that the limit appears to be underestimated, as can be seen in Figures 3a - 3c. It is not possible to choose a value of E_p that both yields an α value in agreement with P16 and does not appear to underestimate the detector threshold flux limit.

Assuming this chosen value for F_{\min} is indeed an appropriate approximation for that used by P16, we are again faced with an underestimation of the true effective value of the detection threshold of Swift, similar to Y15. This can be clearly seen in Figure 3d where we plot the redshift-corrected isotropic peak luminosity vs. redshift, and the solid black line represents the redshift-corrected detection threshold. We observe that this threshold resembles almost a flat line at high redshifts, indicating that the inferred rela-

tionship between $(1+z)$ and L_{iso} is likely heavily influenced by the improperly-modeled detection threshold of BAT. A range of threshold values can be seen in Figures 3e and 3f, as well as their effect on α at $\tau = 0$ and τ at $\alpha = 0$, respectively.

In addition to potential underestimation of the detection threshold, the observational data in the work of P16 also appears to not have been homogeneously collected. Looking at Figure 3c, the histogram of data appears to be multimodal, implying the presence of some, yet-unknown, selection effects in the process of constructing this data set.

Finally, we turn our attention to P16's Monte Carlo simulation, which seems to confirm their results. Unlike Y15, P16 avoid a circular logic inference in their simulations by assuming different F_{\min} (5×10^{-8} [$\text{erg cm}^{-2} \text{s}^{-1}$]) and α (2.2) values from their methodology and results. They are able to successfully recover the GRB formation rate and

luminosity function that they adopted for their simulated sample.

They further test the consequences of sample incompleteness in two approaches. In the first approach they randomly remove a fraction of the bursts close to F_{\min} . In the second approach, they lower F_{\min} by a factor of 5, creating an underestimation of its value. Both approaches artificially create sample incompleteness. The result of both realizations of sample incompleteness is to flatten out the GRB formation rate at low redshift, creating the illusion of an excess of low-redshift GRBs relative to the SFR. This result contradicts the simulations and findings of Y15 and corroborates our conclusions in §3.1.

3.3 Tsvetkova et al. (2017) (T17)

In T17, the authors explore a data set of GRBs detected in the triggered mode of the *Konus-Wind* experiment. Beginning with 150 mixed-type GRBs, they prune the data set down to 137 by removing the Type I (short) GRBs as well as GRB 081203A. It is not explained why GRB 081203A is excluded.

Similar to Y15 and P16, they employ the Efron-Petrosian τ statistic to find the luminosity evolution, assuming a functional form as seen in Eq. (2). They find individual truncation limits to the $(1+z) - L_{\text{iso}}$ plane calculated for each burst separately, yielding an α value of $\alpha = 1.7$. They note that similar results were obtained by using the "monolithic" truncation limit of

$$F_{\min} = 2 \times 10^{-6} [\text{erg cm}^{-2} \text{s}^{-1}] . \quad (9)$$

Since both methods yield the same α value, we will use the single-valued limit, as we have in our previous analyses. When we do so, we obtain a value of $\alpha = 0.29$, which is significantly different from the reported value of 1.7. This should not be the case, so we turn to another method for obtaining α by visually matching the threshold cut from their Figure 8. Doing so requires a threshold of $1.1 \times 10^{-6} [\text{erg cm}^{-2} \text{s}^{-1}]$ and yields $\alpha = 1.36$. Still not the reported α value, we search for the threshold limit of $8.6 \times 10^{-7} [\text{erg cm}^{-2} \text{s}^{-1}]$ which correctly yields $\alpha = 1.7$. This can be seen in Figures 4a - 4c.

As can be seen in Figures 4b and 4c, the analysis of T17 also appears to suffer from an underestimation of the detector threshold limit of *Swift*. Again, we expect the limit to be closer to the central peak of the histogram, soft-truncating the data set. Otherwise, such a sharp drop in the count of LGRB events before reaching the detection threshold would have truly fundamental and revolutionary implications about the cosmic rates of LGRBs.

We note that T17's underestimation of the detector

threshold limit does not appear to be as severe as Y15 or P16. Once the luminosity evolution has been removed, the detector threshold cut in the $(1+z) - L_0$ plane does not become flat at high redshift, as can be seen in Figure 4d.

Also of note is the disparity in the significance of rejecting no luminosity-redshift evolution between our results and those of T17. Figure 4f shows our inferred significance ($\tau \sim 5.4\sigma$), which is hard to reconcile with the inferred significance $\tau = 1.2\sigma$ in T17.

3.4 Lloyd-Ronning, Aykutalp & Johnson (2019) (L19)

In L19, the authors use a data set taken from Wang et al. (2020) that consists of all publicly available observations of 6289 GRBs, from 1991 to 2016. They isolate those events with a measured redshift and observed duration of $T_{90} > 2s$ (LGRBs), which can yield an estimate of the total isotropic gamma-ray emission, E_{iso} . This leads to the selection of 376 LGRB events by L19, based upon which they proceed to construct the τ statistic, in similar fashion to the previous studies, to find the redshift evolution parameter, α . L19 choose a functional form of $E_{\text{iso}} = E_0(1+z)^\alpha$, and find that $\alpha = 2.3 \pm 0.5$. L19 report a value of,

$$F_{\min} = 2 \times 10^{-6} [\text{erg cm}^{-2}] . \quad (10)$$

used in their study. However, when we use this limit in our reanalysis, the threshold line cuts through the majority of the data set, and yields a value of $\alpha = -0.58$. Our inferred value for α is completely at odds with the value reported by L19. In order to obtain their value of $\alpha = 2.3$, we have to use a threshold limit of $7.0 \times 10^{-10} [\text{erg cm}^{-2}]$. This inferred detection threshold cut is almost two orders of magnitude below the data set.

To resolve the disagreement between our inferred value of detection threshold used by L19 and the reported value in their study, we instead settle on visually matching the threshold cut in Figure 1 of L19 to obtain a limit of $1.6 \times 10^{-7} [\text{erg cm}^{-2}]$. This yields Figure 5a which looks remarkably similar to the corresponding Figure 1 of L19. Assuming this detection threshold, we obtain a value of $\alpha = 1.34$ using the Efron-Petrosian τ statistic.

If our inferred the detection threshold is indeed the value used by L19 in their study, then Figures 5b and 5c, lead us to conclude that the detection threshold has been likely severely underestimated in the study of L19. The effective threshold represented by the dashed line in the histogram of Figure 5c is well to the left of the peak of the distribution of LGRB fluences.

Our conclusion in the above is further confirmed by Figure 5d, where we plot the redshift-corrected isotropic

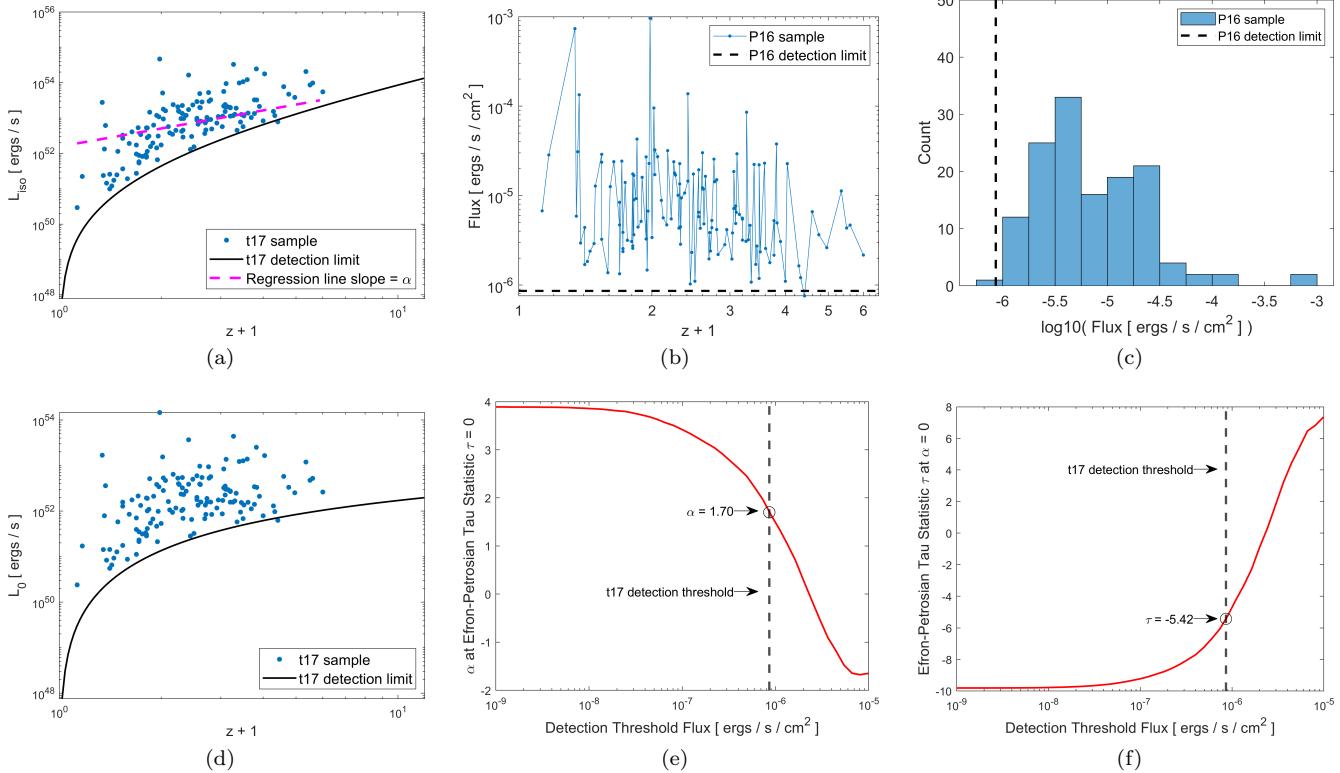


Figure 4. Plotting of the 137 GRBs from T17. Plot (a) shows redshift vs. isotropic luminosity. The black line represents the observational limit of *Swift*, which has been deduced from T17 to be 8.6×10^{-7} [$\text{erg cm}^{-2} \text{s}^{-1}$]. The purple line represents the linear regression through the data set whose slope is $\alpha = 1.70$. Plot (b) is the observer frame visualization of the T17 data set, where the dashed line is the observational limit. Plot (c) is a histogram of flux, where the dashed line is the observational limit. Plot (d) shows redshift vs. $L_0 = L(z)/(1+z)^{1.70}$, the redshift-independent luminosity. Plot (e) shows a range of possible threshold limits vs. α values at $\tau = 0$. The intersection of this line with T17's supposed threshold limit is our value for their α . Plot (f) shows a range of possible threshold limits vs τ values at $\alpha = 0$. Assuming the detection threshold inferred from the analysis of T17, a redshift-independent luminosity distribution is rejected at 5.4σ . However, choosing a more conservative detection threshold at 2×10^{-6} [$\text{erg cm}^{-2} \text{s}^{-1}$] yields no evidence for luminosity-redshift evolution.

effective energy vs. redshift. The solid black line in this plot represents the redshift-corrected detection threshold and almost resembles a flat line at high redshifts. This is another indication that the inferred relationship between $(1+z)$ and E_{iso} is likely heavily influenced by the improperly-modeled detection threshold. In this study, however, the effective threshold cut represent the combined effects of the detection thresholds of multiple satellites due to the heterogenous collection of data from multiple independent GRB catalogs. All of these raise the possibility that L19's single-valued threshold is likely a severe oversimplification of the complex merger of individual satellite thresholds, while none of the individual satellite thresholds might be well represented by a single-valued hard cutoff.

Figure 5e shows a range of possible threshold limits vs. α values at $\tau = 0$. One can see the red line approaches L19's value of $\alpha = 2.3$ on the far left edge of the graph. Figure

5f indicates that E_{iso} and $(1+z)$ are correlated at 6.15σ significance assuming our inferred threshold has been used by L19. Although this significance is not the same as the value reported by L19, 5.6σ , it is consistent. Note, however, that this inferred detection threshold still severely underestimates the actual combined effects of multiple detection thresholds on the heterogenous data set of L19.

4 MONTE CARLO SIMULATIONS OF LUMINOSITY-REDSHIFT EVOLUTION

To further investigate the effects of detection threshold underestimation on the inferred luminosity-redshift evolution of LGRBs, we also create a Monte Carlo universe of LGRBs. The premise of our simulation is that the LGRB world is devoid of any luminosity-redshift or energetics-redshift correlations. Therefore, the application of the Efron-Petrosian statistic on any collection of events measured from such

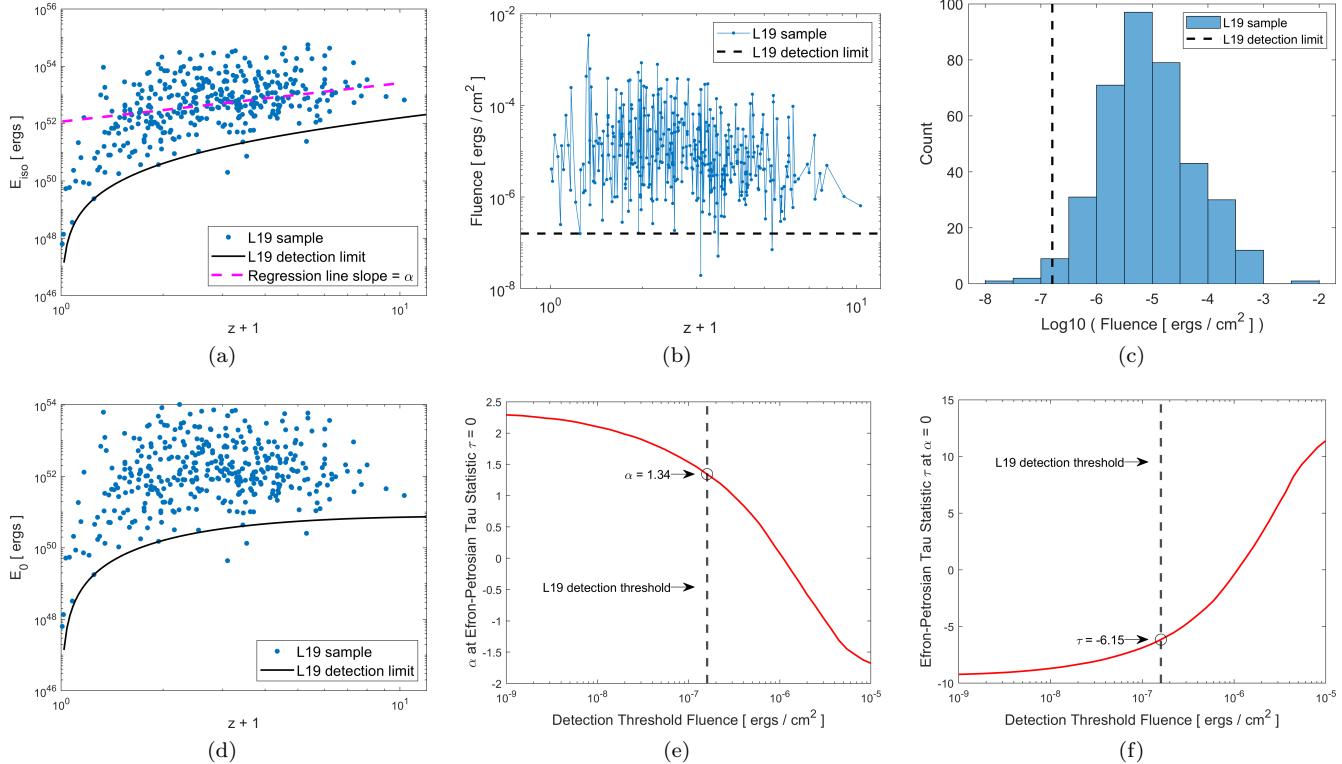


Figure 5. Plotting of the 376 GRBs from L19. Plot (a) shows redshift vs. isotropic emitted energy. The black line represents the observational limit of *Swift*, which has been deduced from L19 to be 1.6×10^{-7} [erg cm $^{-2}$]. The purple line represents the linear regression through the data set whose slope is $\alpha = 1.34$. Plot (b) is the observer frame visualization of the L19 data set, where the dashed line is the observational limit. Plot (c) is a histogram of fluence, where the dashed line is the observational limit. Plot (d) shows redshift vs. $E_0 = E(z)/(1+z)^{1.34}$, the redshift-independent effective energy. Plot (e) shows a range of possible threshold limits vs. α values at $\tau = 0$. The intersection of this line with L19's supposed threshold limit is our value for their α . Plot (f) shows a range of possible threshold limits vs τ values at $\alpha = 0$. Assuming the detection threshold inferred from the analysis of L19, a redshift-independent luminosity distribution is rejected at 6.15σ . However, choosing a more conservative detection threshold at $\sim 10^{-6}$ [erg cm $^{-2}$] yields no evidence for luminosity-redshift evolution.

Monte Carlo universe of LGRBs, subjected to a given simulated detection threshold, should also accurately predict no energetics/luminosity-redshift correlations for the intrinsic underlying population of LGRBs in the Monte Carlo universe.

$$\zeta(z) \propto \begin{cases} (1+z)^{\gamma_0} & z < z_0 \\ (1+z)^{\gamma_1} & z_0 < z < z_1 \\ (1+z)^{\gamma_2} & z > z_1 , \end{cases} \quad (11)$$

whose parameters,

$$(z_0, z_1, \gamma_0, \gamma_1, \gamma_2) = (0.97, 4.00, 3.14, 1.36, -2.92) , \quad (12)$$

4.1 The LGRB World Model

We follow the approach extensively discussed in Shahmoradi (2013b,a); Shahmoradi & Nemiroff (2015, 2019); Shahmoradi & Nemiroff (2019); Osborne, Shahmoradi & Nemiroff (2020), to construct our Monte Carlo Universe of LGRBs. Toward this, we assume that the intrinsic comoving rate density of LGRBs follows a piecewise power-law function of the form,

are adopted from Butler, Bloom & Poznanski (2010). For simplicity, but also as argued by Shahmoradi (2013b,a); Shahmoradi & Nemiroff (2015, 2019); Shahmoradi & Nemiroff (2019); Osborne, Shahmoradi & Nemiroff (2020), we consider a 4-dimensional Multivariate Lognormal Probability Density Function (PDF) for the joint distribution of the four main LGRB prompt gamma-ray emission characteristics: The total isotropic peak luminosity (L_{iso}), the total gamma-ray emission (E_{iso}), the intrinsic spectral

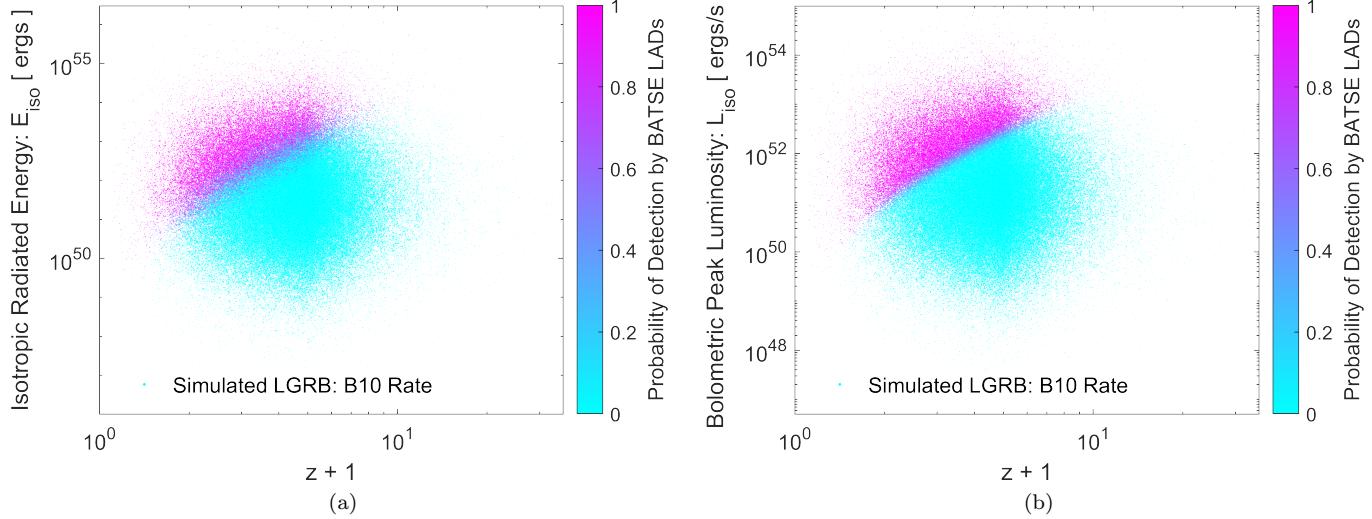


Figure 6. An illustration of the Monte Carlo universe of LGRBs constructed in §4.2. Each point in this plot represents one synthetic LGRB. The magenta color represents a high probability of detection while the cyan represents a low probability of detection.

peak energy (E_{pz}), and the intrinsic duration (T_{90z}).

We use the BATSE catalog of 1366 LGRBs (Paciesas et al. 1999; Shahmoradi & Nemiroff 2010; Goldstein et al. 2013) to construct a Bayesian hierarchical model (Shahmoradi 2017,?) of the cosmic distribution of LGRBs in the universe, subject to an accurate modeling of the the detection threshold of BATSE Large Area Detectors (LADs) and data uncertainties. Then, we use a variant of the adaptive Markov Chain Monte Carlo techniques to sample the resulting posterior distribution of the parameters of the hierarchical model (Shahmoradi 2019; Shahmoradi & Bagheri 2020; Shahmoradi & Bagheri 2020a; Shahmoradi, Bagheri & Osborne 2020; Shahmoradi & Bagheri 2020b,b). Details of model construction and sampling are extensively discussed in the aforementioned papers (e.g., Shahmoradi 2013b; Osborne, Shahmoradi & Nemiroff 2020).

4.2 The Monte Carlo Universe of LGRBs

Once the best-fit parameters of the LGRB world model are inferred, we create a Monte Carlo Universe of LGRBs by randomly and repeatedly generating LGRB events whose characteristics are distributed according to the LGRB World model constructed in §4.1. For each LGRB event synthesis, we use a set of model parameters randomly drawn from the posterior distribution of the LGRB world model parameters explored in §4.1. Then, each LGRB passes through the simulated LGRB detection process of the BATSE LADs.

An illustration of the resulting Monte Carlo Universe of LGRBs is provided in Figure 6. The two plots represent the joint distributions of $E_{\text{iso}}/L_{\text{iso}}$ and redshift. Clearly, the

BATSE LADs create a rather sharp cut on the synthesized $z - L_{\text{iso}}$ sample of LGRBs compared to the distribution of $z - E_{\text{iso}}$, which exhibits much fuzzier detection threshold effects. This is expected and reassuring, since the BATSE LADs primarily triggered on the peak photon flux at different timescales.

We note that the specific shape of the energetics or redshift distribution of LGRBs or the specific detection mechanism of LGRBs in our Monte Carlo simulations has no relevance or effects on our assessment of the utility and accuracy of the Efron-Petrosian statistic. All that is important here, is the lack of any a priori correlations between the energetics and the redshifts of LGRBs in our Monte Carlo simulations.

Using our Monte Carlo universe of LGRBs, we generate a random sample of 380 BATSE-detectable LGRBs. This sample size is comparable to the size of the observational data sets collected and analyzed in previous studies. We flag an LGRB as detectable by generating a uniform-random number between 0 and 1 and comparing it to the probability of detection of the LGRB. If the probability of detection is higher than the randomly generated number, we include the event in the sample of detected LGRBs for our analysis.

4.3 Analysis of Synthetic Monte Carlo Data

We start with the synthetic $E_{\text{iso}} - z$ sample shown in figure 7a where the black line represents the BATSE detector threshold at 50% detection probability and the color on each point represents the probability of that burst being detected

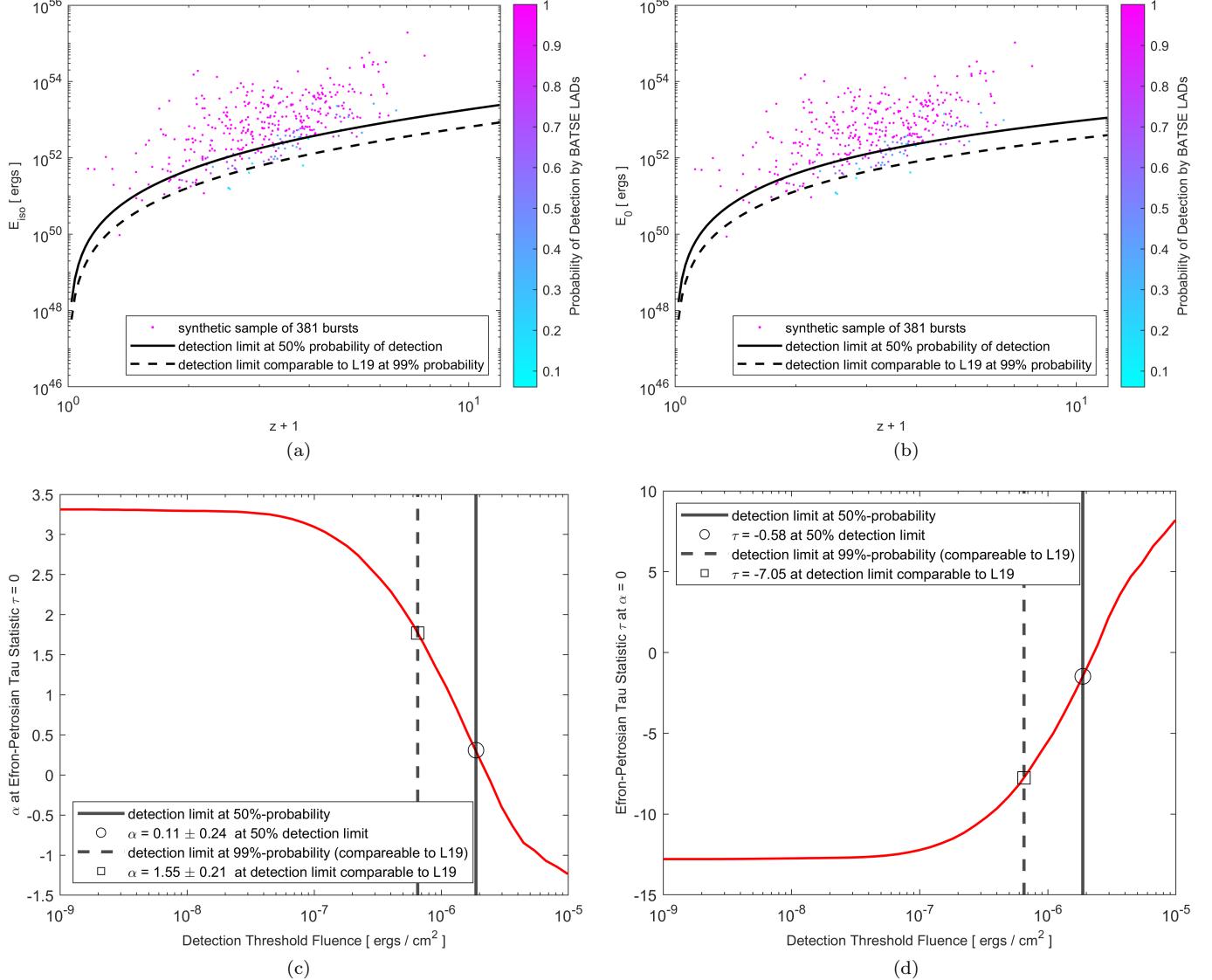


Figure 7. An illustration of the effects of detection threshold on the outcome of the Efron-Petrosian test statistic. Plot (a) Shows the synthetic data set used for our study of ($E_{\text{iso}}, z + 1$) correlation. The solid black line represents the detector threshold at 50% while the dashed black line represents a detector threshold comparable to that of L19 at 99% probability of detection. The color bar represents the probability of detection by the BATSE LADs where cyan and magenta represent 0% and 100% chances of detection, respectively. Plot (b) shows the redshift-evolution corrected data set based off of the value of alpha calculated using the detector threshold at 50% probability of detection. Plot (c) shows the alpha value calculated corresponding to $\tau = 0$ with varying detection threshold limits. The solid black line represents the detector threshold at 50%, the black circle is the average α value over 50 generated samples for the specific threshold used, while the dashed black line represents the detector threshold at 99% probability of detection, comparable to those of previous studies, and the black square is the average α over 50 generated samples at $\tau = 0$. Plot (d) displays the τ statistic at $\alpha = 0$. The black line represents the detector threshold at 50% detection probability and the dashed black line represents the detection threshold at 99% detection probability, comparable to those of previous studies. The circle and square in this figure are the average τ values over 50 generated samples.

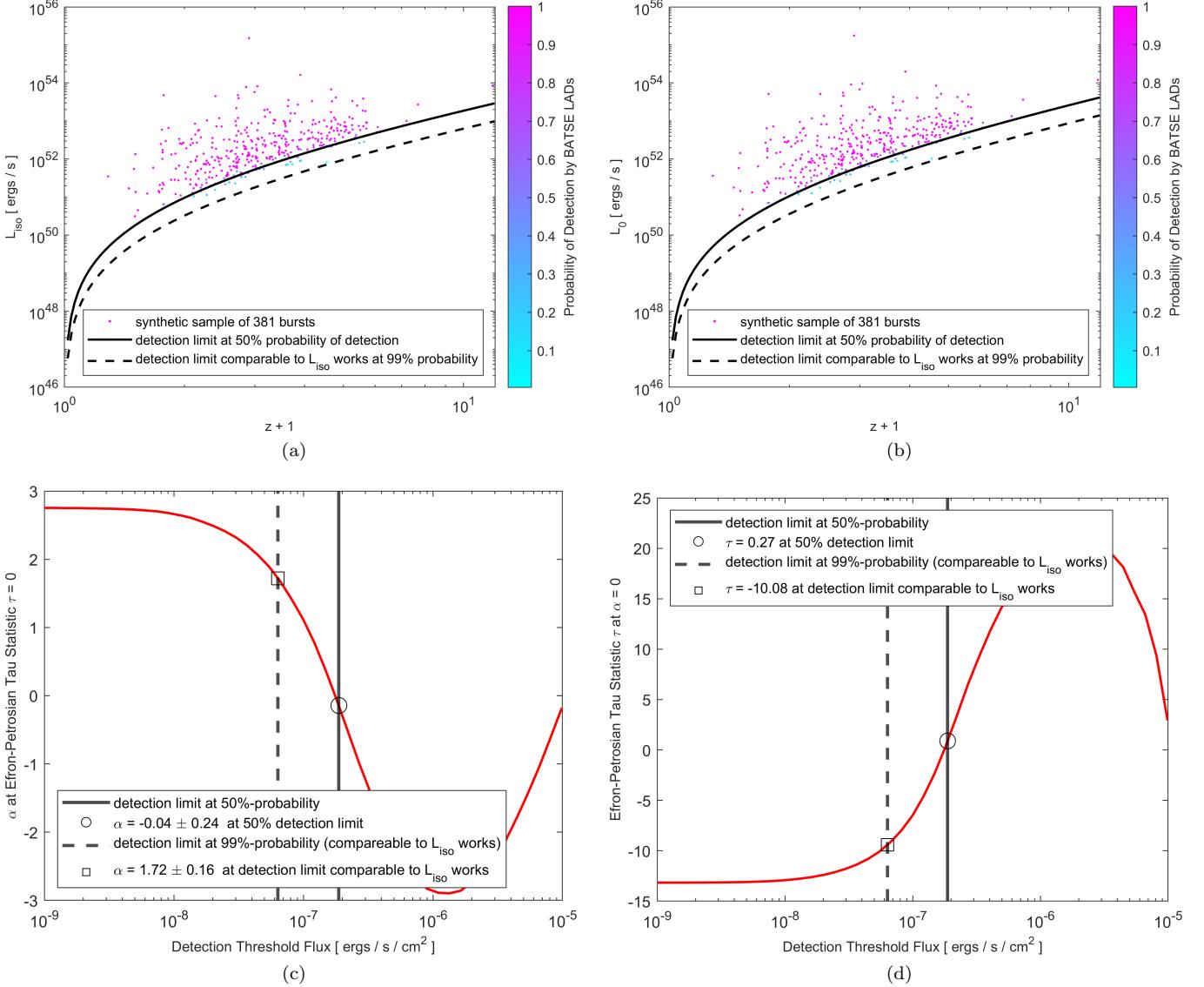


Figure 8. An illustration of the effects of detection threshold on the outcome of the Efron-Petrosian test statistic. Plot (a) Shows the synthetic data set used for our study of ($L_{\text{iso}}, z + 1$) correlation. The solid black line represents the detector threshold at 50% while the dashed line represents a detector threshold comparable to that of L19 at 99% probability of detection. The color bar represents the probability of detection by the BATSE LADs where cyan and magenta represent 0% and 100% chances of detection, respectively. Plot (b) shows the redshift-evolution corrected data set based off of the value of alpha calculated using the detector threshold at 50% probability of detection. Plot (c) shows the alpha value calculated corresponding to $\tau = 0$ with varying detection threshold limits. The solid black line represents the detector threshold at 50%, the black circle is the average α value over 50 generated samples for the specific threshold used, while the dashed black line represents the detector threshold at 99% probability of detection, comparable to those of previous studies, and the black square is the average α over 50 generated samples at $\tau = 0$. Plot (d) displays the τ statistic at $\alpha = 0$. The black line represents the detector threshold at 50% detection probability and the dashed black line represents the detection threshold at 99% detection probability, comparable to those of previous studies. The circle and square in this figure are the average τ values over 50 generated samples.

by BATSE. The corresponding lower limit on the fluence at 50% chance of detection is

$$F_{\min} = 1.88 \times 10^{-6} [\text{ ergs cm}^{-2}] . \quad (13)$$

We then apply the Efron-Petrosian statistic to our synthetic data set and find that $\alpha = 0.11 \pm 0.24$ with the detection threshold set at 50% probability of detection. This is reassuring as it implies that the Efron-Petrosian test statistic remains relatively unbiased even when the detection threshold is not a sharp cutoff. But this is true only for as long as the detection threshold is not significantly underestimated.

To further illustrate this, we consider a lower detection threshold, comparable to the value used in L19. We note that a direct application of the assumed detection threshold of L19 to our analysis is not possible since the data set used in L19 has been collected from multiple heterogeneous sources, as opposed to our synthetic homogeneously-detected LGRB sample. However, a detection threshold equivalent to that of L19 can be obtained in our analysis by noting that the detection threshold cutoff assumed in the study of L19 is above only 3 individual LGRB events. This comprises less than 1% of the entire data set of 376 LGRBs in L19.

We therefore, choose our detection threshold limit such that only 1% of our synthetic sample falls below the assumed detection threshold hard cutoff, similar to that of L19. This yields a value of $\alpha = 1.55 \pm 0.21$ which depicted illustrated in Figure 7c. This inferred non-zero correlation at 7σ significance clearly contradicts the fundamental assumption of our Monte Carlo simulation, and confirms our hypothesis that an underestimation the detection threshold can readily bias the Efron-Petrosian test statistic.

Next, we repeat the above analysis for the joint distribution of $z - L_{\text{iso}}$ with a detection threshold hard cutoff set at 50% probability of detection: $1.88 \times 10^{-7} [\text{ ergs cm}^{-2} \text{ s}^{-1}]$. We find $\alpha = -0.04 \pm 0.24$ at this probability of detection. However, when we use a detection threshold comparable to those of Y15, P16, and T17, we find $\alpha = 1.72 \pm 0.16$ at $> 10\sigma$ significance, again contradicting the a priori assumption of our Monte Carlos universe of LGRBs.

5 DISCUSSION

In this work we re-analyzed several previous studies on the evolution of the luminosity/energetics of LGRBs with redshift. To be consistent with the previous studies we used the method of Efron-Petrosian and the τ statistic to determine the exponent of the power-law relationship, α in $g(z) = (1 + z)^\alpha$, between the luminosity/energetics and redshifts of LGRBs.

Contrary to the previous studies, we conclude that the

effects of the detection threshold has been likely severely underestimated. We further confirm our conclusion via Monte Carlo simulation, where we assume no correlation between the energetics and the redshift of LGRBs. We then measure these simulated LGRBs via the BATSE detector (for its simplicity). Our finding is that **an underestimation of the effective detection threshold by even less than a factor of two can create artificial correlations between the redshifts and the luminosity/energetics of LGRBs**. The Monte Carlo simulation of P16 also shows that an underestimation of the detector flux limit can lead to apparent artificial correlations between luminosity and redshift. They further show that this effect can also give rise to an apparent overabundance of LGRBs at low redshift.

The regression slope (α , on a log-log plot) of the reported correlations between the redshift and the luminosity/energetics of previous studies also resembles their chosen detection threshold (Figures 1a, 3a, and 5a). This further corroborates our hypothesis that the observed correlations are an artifact of the individually chosen detection thresholds of the various gamma-ray detectors.

A more accurate study of the luminosity/energetics-redshift evolution requires a more careful modeling of the detection threshold of gamma-ray detectors, where the detection threshold is not a single cutoff on the distribution of LGRBs but rather a dispersed set of detection probabilities in the entire bivariate distribution. However, such a modeling approach is impossible with the original method of Efron-Petrosian and requires parametric modeling of the luminosity/energetics and redshift distribution as well as the detection threshold.

Le, Ratke & Mehta (2020) uses purely parametric methods to determine the GRB formation rate $\rho(z)$. They find that there is no deviation from the SFR at any redshift for the complete unbiased LGRB *Swift*-Perley and *Swift*-Ryan-b samples. They do, however, find an excess at low redshift ($z < 1$) for the *Swift*-Ryan 2012 sample, and conclude that the reason for this excess is either incomplete sample size, that $\rho(z)$ doesn't trace SFR at low redshift, or that it is simply unclear.

The premise of the previous studies has been to provide a nonparametric investigation of the luminosity/energetics vs redshift. However, upon performing the nonparametric correlation, the majority of these investigations rely on parametric fitting of the luminosity/energetics and redshift distributions, which convolutes the premise. Indeed, Lan et al. (2019) presents a fully parametric study of the redshift/energetics evolution and reports a potential correlation between the two, but nevertheless, their study is founded on the assumption of a simple hard cutoff of the detection threshold of LGRBs. A fully parametric study of the correlation which incorporates a more accurate and detailed de-

scription of the detection threshold of gamma-ray detectors remains to be done.

REFERENCES

- Band D. et al., 1993, *The Astrophysical Journal*, 413, 281
 Butler N. R., Bloom J. S., Poznanski D., 2010, *The Astrophysical Journal*, 711, 495
 Dainotti M. G., Singal J., Ostrowski M., et al., 2013, *The Astrophysical Journal*, 774, 157
 Efron B., Petrosian V., 1992, *The Astrophysical Journal*, 399, 345
 Eichler D., Livio M., Piran T., Schramm D. N., 1989, *Nature*, 340, 126
 Fenimore E., Palmer D., Galassi M., Tavener T., Barthelmy S., Gehrels N., Parsons A., Tueller J., 2003in , American Institute of Physics, pp. 491–493
 Gehrels N., Ramirez-Ruiz E., Fox D. B., 2009, *Annual Review of Astronomy and Astrophysics*, 47, 567
 Goldstein A., Preece R. D., Mallozzi R. S., Briggs M. S., Fishman G. J., Kouveliotou C., Paciesas W. S., Burgess J. M., 2013, *The Astrophysical Journal Supplement Series*, 208, 21
 Kocevski D., Liang E., 2006, *The Astrophysical Journal*, 642, 371
 Kouveliotou C., Meegan C. A., Fishman G. J., Bhat N. P., Briggs M. S., Koshut T. M., Paciesas W. S., Pendleton G. N., 1993, *The Astrophysical Journal*, 413, L101
 Krühler T. et al., 2015, *Astronomy & Astrophysics*, 581, A125
 Lan G.-X., Zeng H.-D., Wei J.-J., Wu X.-F., 2019, *Monthly Notices of the Royal Astronomical Society*, 488, 4607
 Le T., Ratke C., Mehta V., 2020, *Monthly Notices of the Royal Astronomical Society*, 493, 1479
 Lloyd-Ronning N. M., Aykutalp A., Johnson J. L., 2019, *Monthly Notices of the Royal Astronomical Society*, 488, 5823
 Lynden-Bell D., 1971, *Monthly Notices of the Royal Astronomical Society*, 155, 95
 Mészáros P., 2006, *Reports on Progress in Physics*, 69, 2259
 Osborne J. A., Shahmoradi A., Nemiroff R. J., 2020, arXiv e-prints, arXiv:2006.01157
 Paciesas W. S. et al., 1999, *The Astrophysical Journal Supplement Series*, 122, 465
 Perley D. et al., 2015, *The Astrophysical Journal*, 801, 102
 Perley D. et al., 2016a, *The Astrophysical Journal*, 817, 8
 Perley D. A. et al., 2016b, *The Astrophysical Journal*, 817, 7
 Pescalli A. et al., 2016, *Astronomy & Astrophysics*, 587, A40
 Petrosian V., 2002, in *International Astronomical Union Colloquium*, Vol. 184, Cambridge University Press, pp. 389–397
 Petrosian V., Kitanidis E., Kocevski D., 2015, *The Astrophysical Journal*, 806, 44
 Salvaterra R. et al., 2012, *The Astrophysical Journal*, 749, 68
 Schulze S. et al., 2015, *The Astrophysical Journal*, 808, 73
 Shahmoradi A., 2013a, arXiv preprint arXiv:1308.1097
 Shahmoradi A., 2013b, *The Astrophysical Journal*, 766, 111
 Shahmoradi A., 2017, arXiv preprint arXiv:1711.10599
 Shahmoradi A., 2017, arXiv e-prints, arXiv:1711.10599
 Shahmoradi A., 2019, *Bulletin of the American Physical Society*
 Shahmoradi A., Bagheri F., 2020, arXiv preprint arXiv:2008.09589
 Shahmoradi A., Bagheri F., 2020a, arXiv e-prints, arXiv:2008.09589
 Shahmoradi A., Bagheri F., 2020b, arXiv e-prints, arXiv:2009.14229
 Shahmoradi A., Bagheri F., Osborne J. A. e., 2020, arXiv e-prints, arXiv:2010.00724
 Shahmoradi A., Nemiroff R., 2009in , AIP, pp. 425–427
 Shahmoradi A., Nemiroff R., 2011, *Monthly Notices of the Royal Astronomical Society*, 411, 1843
 Shahmoradi A., Nemiroff R. J., 2010, *Monthly Notices of the Royal Astronomical Society*, 407, 2075
 Shahmoradi A., Nemiroff R. J., 2015, *Monthly Notices of the Royal Astronomical Society*, 451, 126
 Shahmoradi A., Nemiroff R. J., 2019, arXiv e-prints, arXiv:1903.06989
 Shahmoradi A., Nemiroff R. J., 2019, arXiv preprint arXiv:1903.06989
 Singal J., Petrosian V., Lawrence A., et al., 2011, *The Astrophysical Journal*, 743, 104
 Tsvetkova A. et al., 2017, arXiv preprint arXiv:1710.08746
 Vergani S. et al., 2015, *Astronomy & Astrophysics*, 581, A102
 Wang F., Zou Y.-C., Liu F., Liao B., Liu Y., Chai Y., Xia L., 2020, *The Astrophysical Journal*, 893, 77
 Woosley S. E., 1993, *The Astrophysical Journal*, 405, 273
 Yu H., Wang F., Dai Z., Cheng K., 2015, *The Astrophysical Journal Supplement Series*, 218, 13
 Zhang B., 2007, *Chinese Journal of Astronomy and Astrophysics*, 7, 1