

Maria Giovanna Dainotti

Redshift estimator to support GRB rate vs SFR density

Main Collaborators: Aleksander Lenart (PhD, Jagiellonian), Nagataki (Riken), Iwasaki (NAOJ), Agnieszka Pollo (NCBJ), Vahe' Petrosian (Stanford), Fraija (UNAM), Dieter Hartmann, Nikita Khatiya, Druv Bahl (Clemson University)

1/10/2026, Cape Town, HEASA

GRB and SFH connection

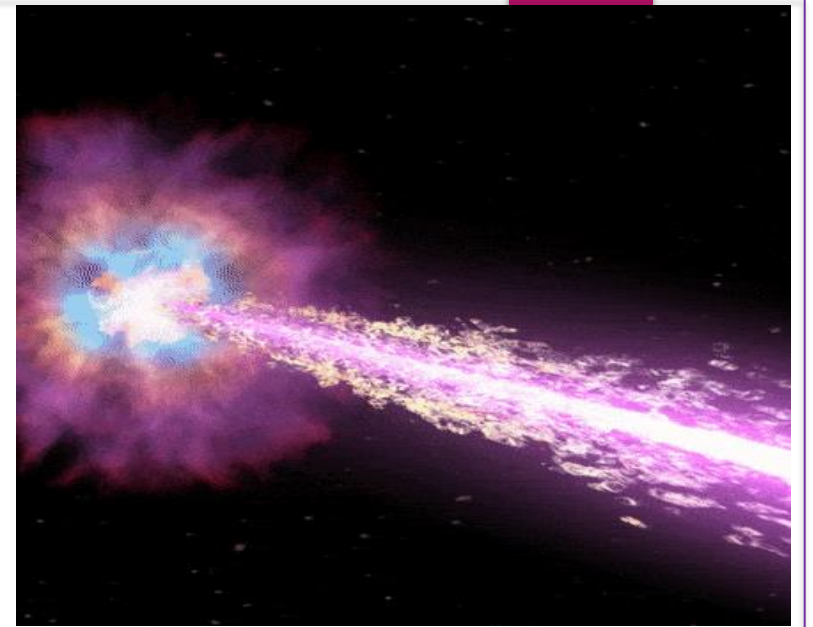
Traditionally

- ▶ Long duration GRBs, due to massive-star deaths
- ▶ Short duration GRBs, due to compact-object mergers

Both in principle they should trace the Star Formation rate density with different delays

Delay=time between star formation and GRB
(40 Myr for Long and up to Gyr for short)

We will tackle these delays



Is the GRB-SFH link so simple?

GRB rate deviate from the SFR Density at low- z and high- z

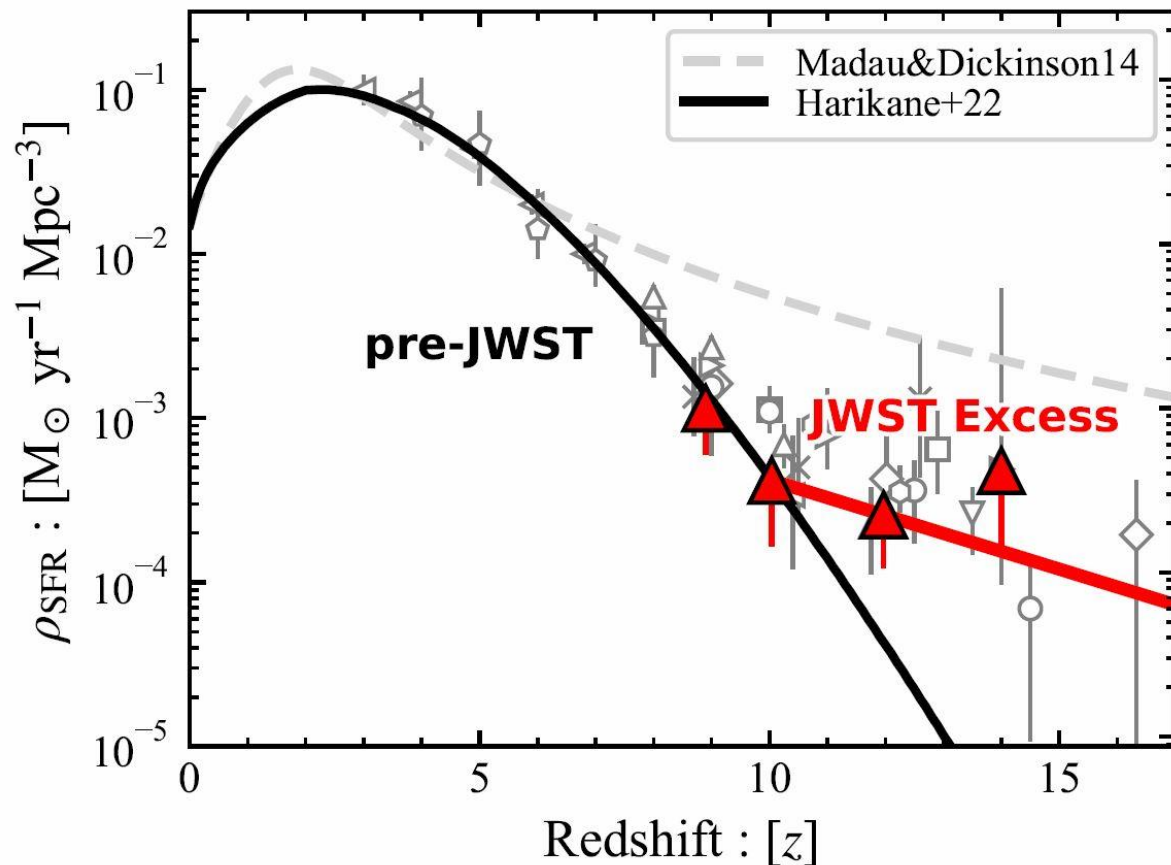
Key challenges

- ▶ Few GRBs have observed redshifts
- ▶ Non uniform (different surveys, etc) GRB host galaxy analysis
 - ▶ Progenitors are mixed in terms of the long/short duration

JWST galaxy excess and the GRB-SFR connection

4

JWST finds more UV-bright galaxies at $z \gtrsim 10$ than many models predict. The inferred SFR depends on the stellar population model and AGN light: some of a galaxy's brightness could come from gas accreting onto its central supermassive black hole.



UV-inferred SFR assumes a stellar population model

Harikane et al. (arXiv:2208.01612) Matsumoto et al. (2024, arXiv:2409.11468), Fig. 2

Reduced UV suppression

Weaker UV feedback (radiation affecting the gas that form stars) may allow more SF and more GRBs, if GRB efficiency does not evolve.

Top-heavy Initial Mass Function

More UV light per stellar mass and more massive star forms altering both the inferred SFR and GRB efficiency.

AGN contribution

AGN light can yield a galaxy excess without increasing the GRB rate.

Progenitor physics and selection biases affect GRB rates.

1) Is GRB rate a tracer of the SFR Density?

Hence, we need to answer:

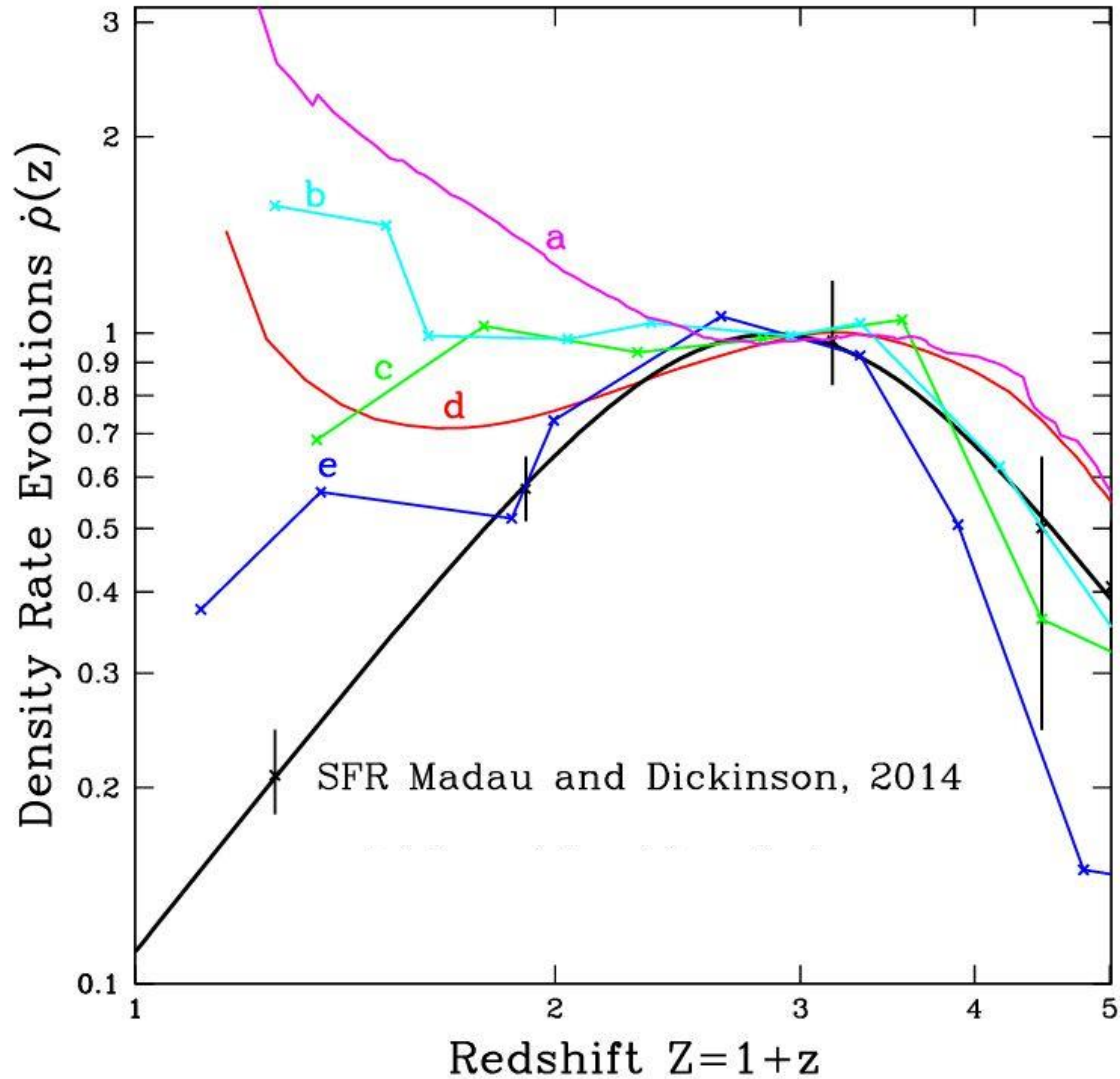
2) Host galaxy census: How do GRB host galaxies differ in their properties from the global star-forming population?

3) Progenitors: How can we understand GRB classes from the physics rather than from their duration?

Observable → Physical Classification

GRB deviates from the cosmic SFR

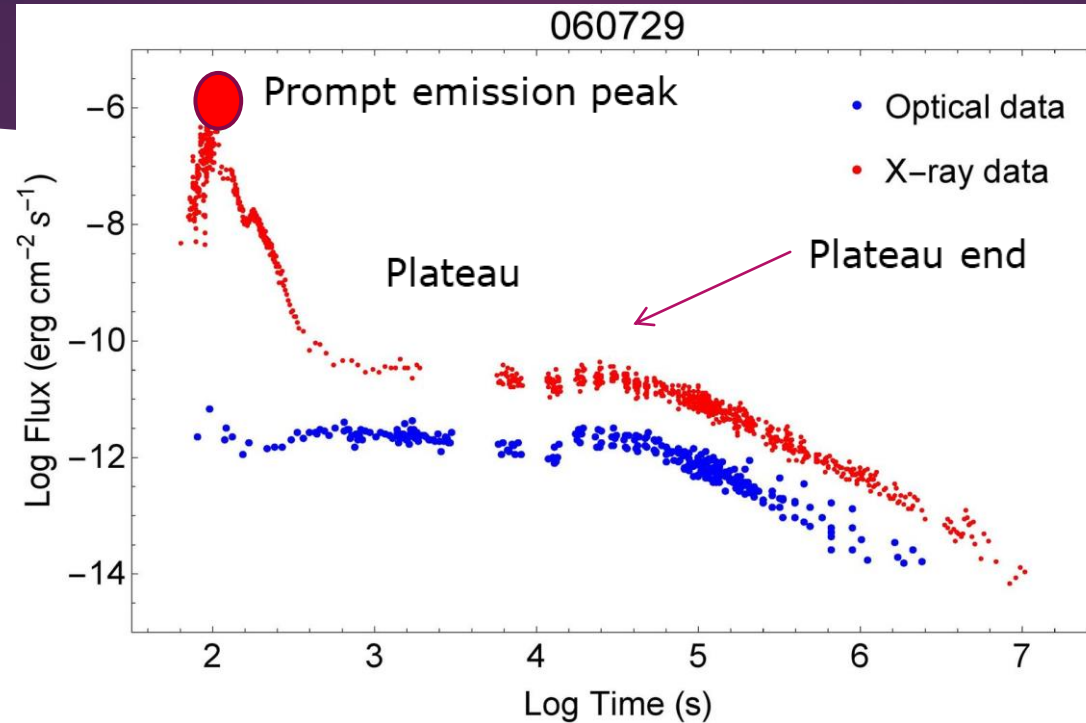
6



- (a) magenta, Lloyd-Ronning et al. (2019);
(b) cyan, Yu et al. (2015);
(c) green, Tsvetkova et al. (2017);
(d) red, Petrosian et al. (2015);
(e) blue, Pescalli et al. (2016).

We can also use a sample with plateau features

Standard is good!



Dainotti, Livermore, Kann, Li, Oates, Yi, Zhang, Gendre, Cenko, Fraija 2020, ApJL, 905, 26.

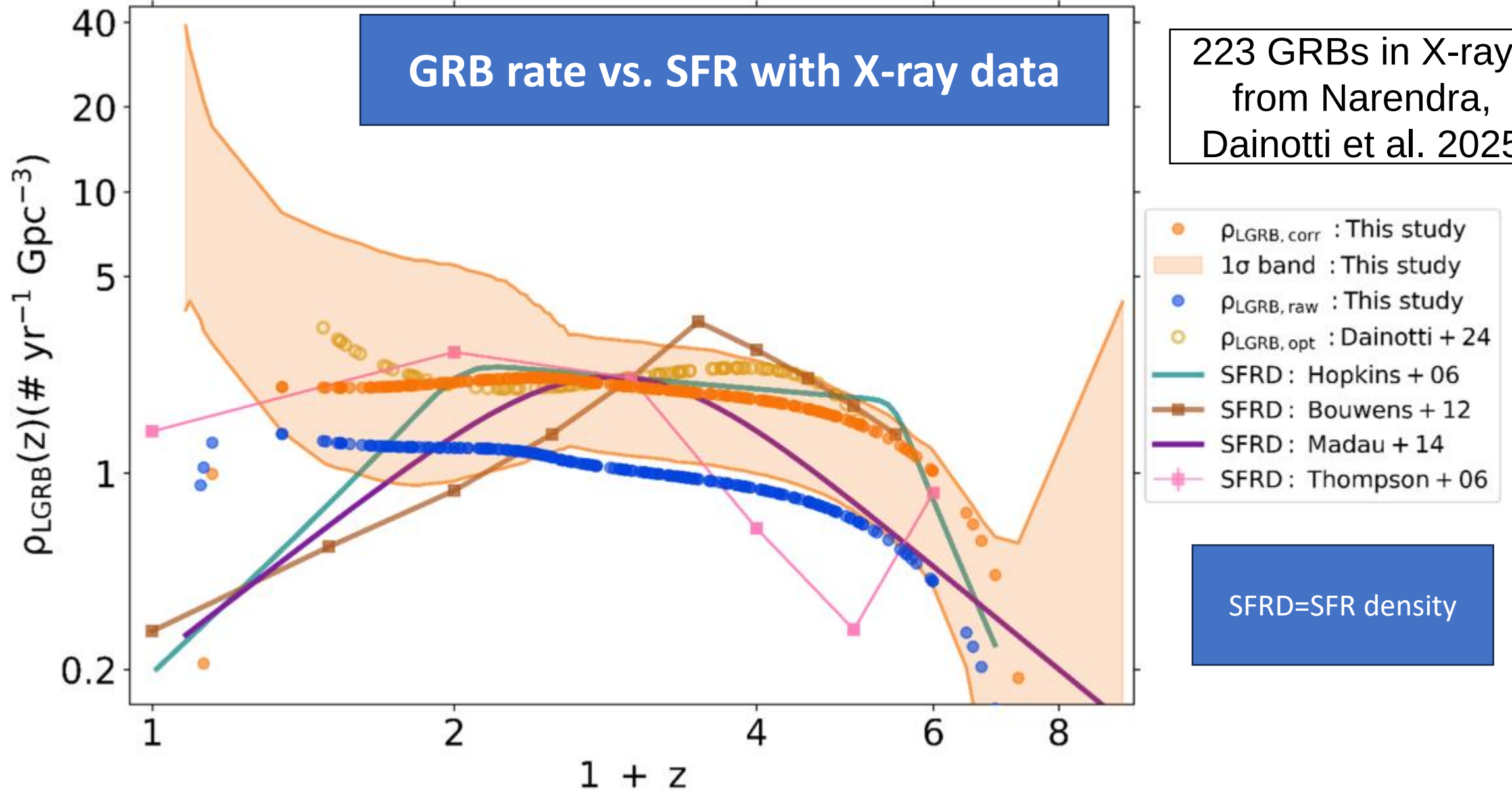
Burst Alert Telescope+ X-Ray Telescope+UVOT from Swift.

The blue line is the phenomenological Willingale model (R. Willingale et al. 2007)

- The less varied properties of the plateau compared to the prompt allows an independent analysis to check the deviation.

GRB rate vs. SFR with X-ray data

223 GRBs in X-rays
from Narendra,
Dainotti et al. 2025

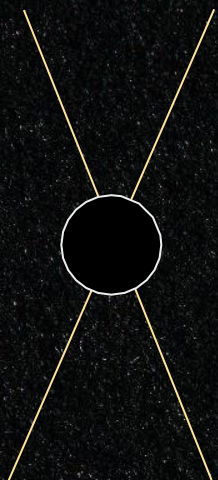


1- σ band is obtained with bootstrap method for 1000 samples where parameter distributions reproduce the actual observations

Khatiya, Dainotti, et al.
2024, ApJ, 990, 1, id.69, 11

Explain mismatch-I

Beaming evolution



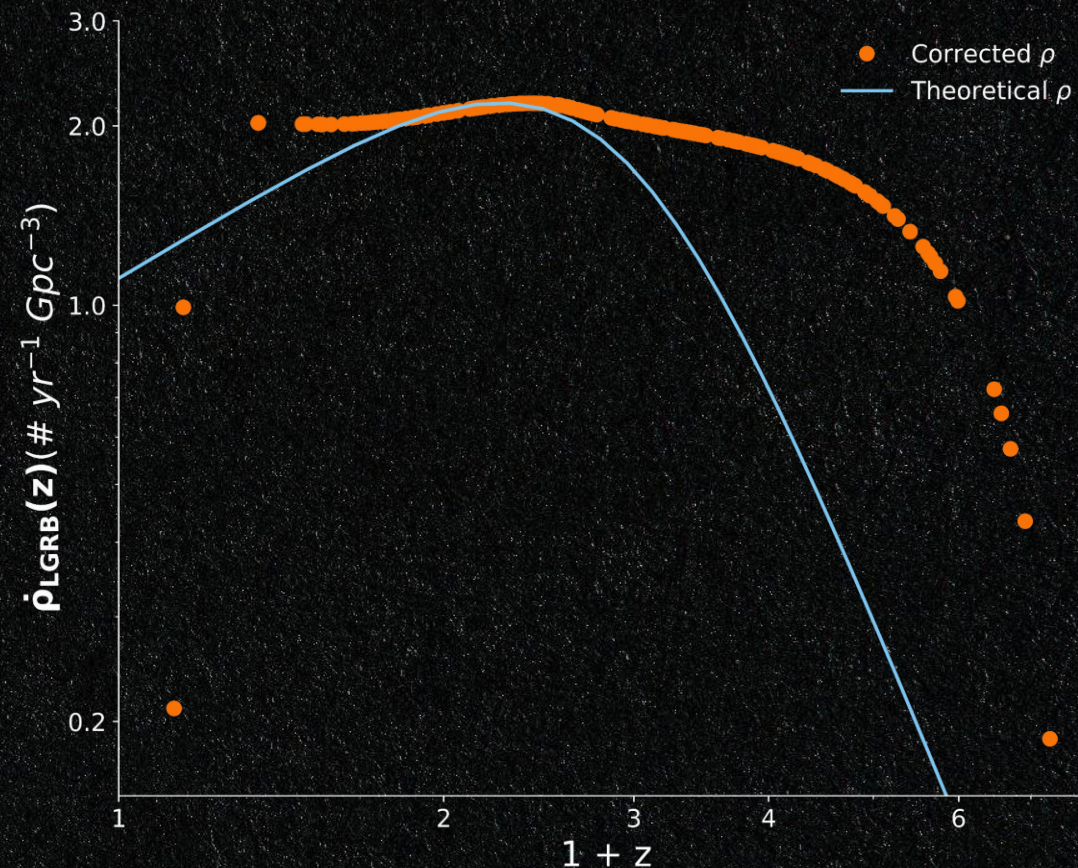
Low- $z \rightarrow$ Wider jets



High- $z \rightarrow$ Narrower jets

$$f_{\text{beam}}(z) \propto \theta_j^2 \propto (1+z)^{-1.6} \quad [\text{Ref: Lloyd-Ronning+19}]$$

Instead of fixed f_{beam} , $\rho_{\text{theory}}(z) \propto f_{\text{beam}}(z)$



Beaming evolution with z is not sufficient to explain the ρ_{corr}

Explain mismatch-II

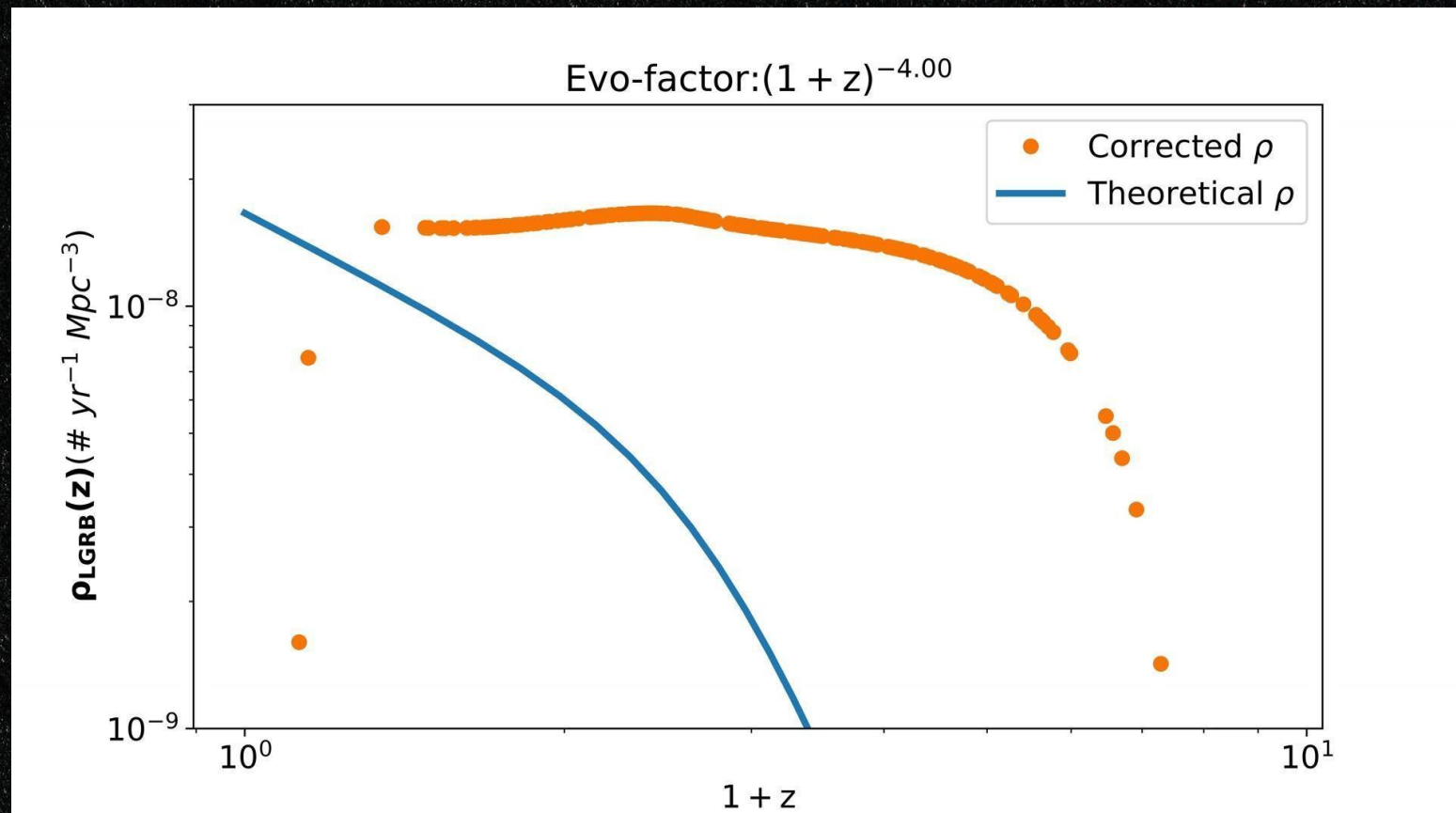
$$\rho(z) = \frac{d\sigma}{dz}(1+z) \left(\frac{dV}{dz} \right)^{-1}$$

$$\rho_{\text{LGRB}}(z) = \frac{F_1 * F_2 * \rho(z)}{\text{Years} * \text{Uptime} * \text{FOV}}$$

F1=GRBs removed by the analysis;
F2=fraction of GRBs with plateaus

Power law evolution absorbing all
unresolved progenitor properties:

$$\rho_{\text{theory}}(z) \propto (1+z)^k; k \rightarrow [-4, 4]$$



$(1+z)^k$ factor is not sufficient to explain the ρ_{corr}

To answer all these questions we need redshifts

WHY ARE WE DOING THIS?

- Currently, the SWIFT GRB catalog only has 26% of GRBs (~400) with measured redshift.
- ~ 9% of GRBs have measured redshifts of the total sample.
- Redshift measurements are crucial for the cosmological application of GRBs: connection between GRB rate and SFR
- This highlights the difficulty of the problem and why we needed to mature the methodology

WHAT ARE WE TRYING TO DO?

- We are training a supervised machine learning model on optical photometric GRB properties to reliably predict the redshift.
- Using this model, we will estimate the unknown redshift of GRBs.

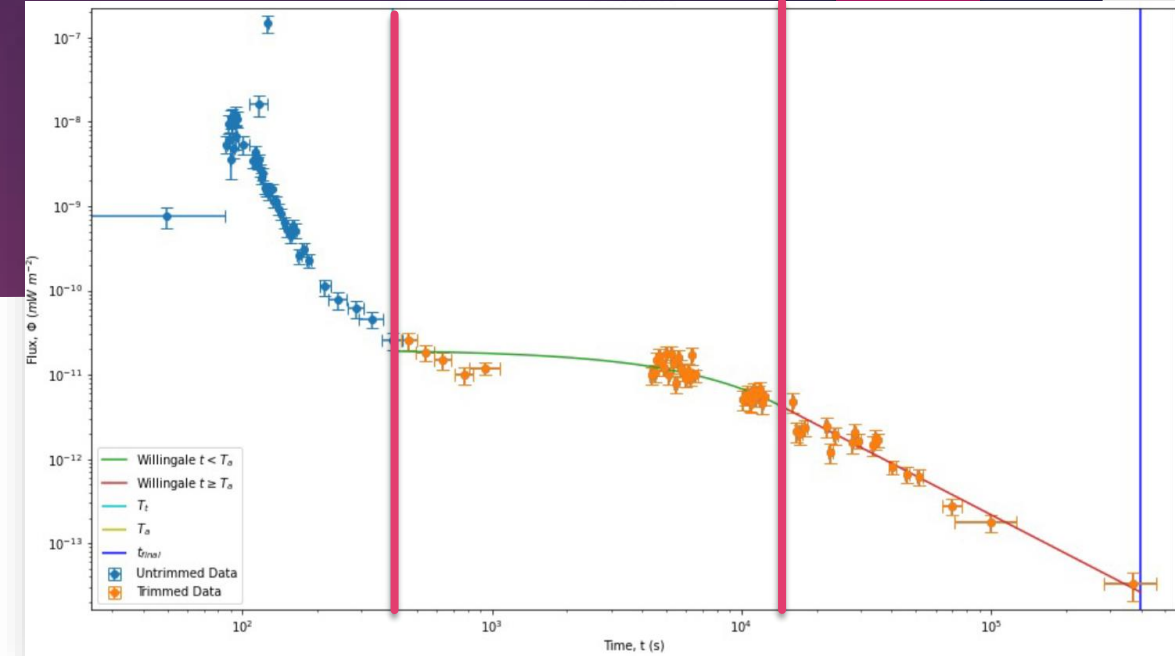
HOW ARE WE GOING DO IT?

- ▶ We are applying the SuperLearner machine learning model to GRBs.
- ▶ We use MICE to impute missing data
- ▶ Applying Optimal Transport Bias Correction
- ▶ For the first time, using optical plateau properties for GRB redshift estimation.

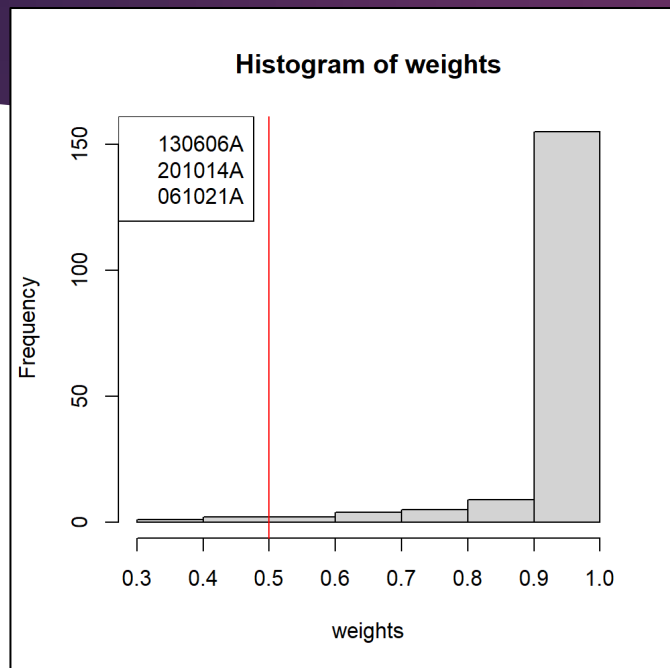
Data Sample

Dainotti, Narendra, 2024, **ApJL**,
967L,30

- ▶ We are using 179 GRBs which show the optical plateau in the lightcurve.
- ▶ These are taken from “*The Optical Two- and Three-dimensional Fundamental Plane Correlations for Nearly 180 Gamma-Ray Burst Afterglows with Swift/UVOT, RATIR, and the Subaru Telescope*” (Dainotti et al. 2022d)
 - The dataset contains GRBs from Swift/UVOT, RATIR and the Subaru.
 - We use 9 features in this analysis:
 - 4 features from the prompt emission:
 - Fluence,
 - T90,
 - Peak flux,
 - Photon Index
 - 5 features from the plateau emission:
 - Time (T_a), Flux (F_a),
 - Temporal index (α), Spectral index (β) at the end of plateau &
 - Hydrogen column density (NH)

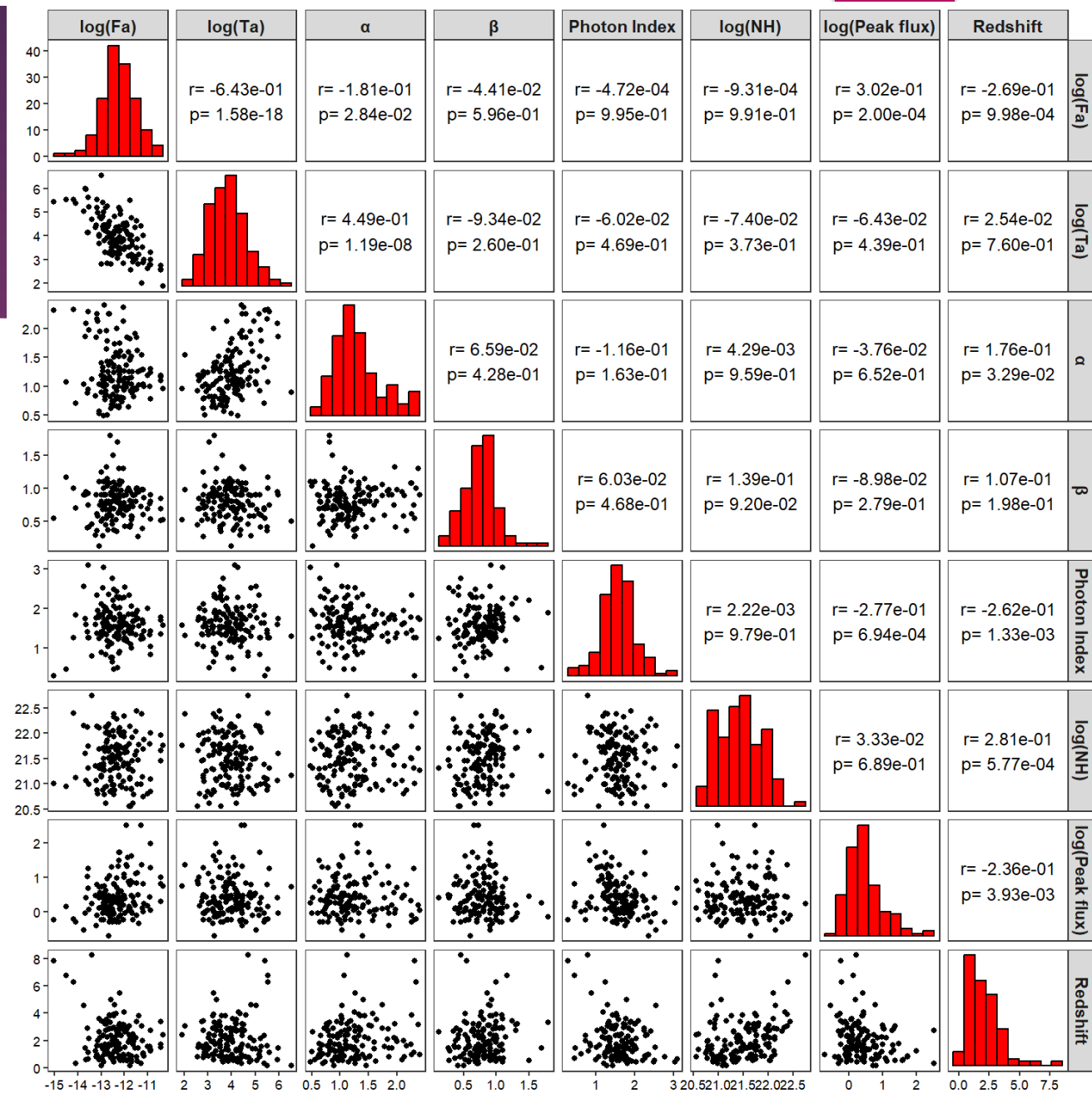


Data Sample: Outlier



Finally, we remove 3 GRBs as outliers using the Robust Linear Model with M-estimator, by selecting a cutoff weight of 0.5.

Training set= **176 GRBs**

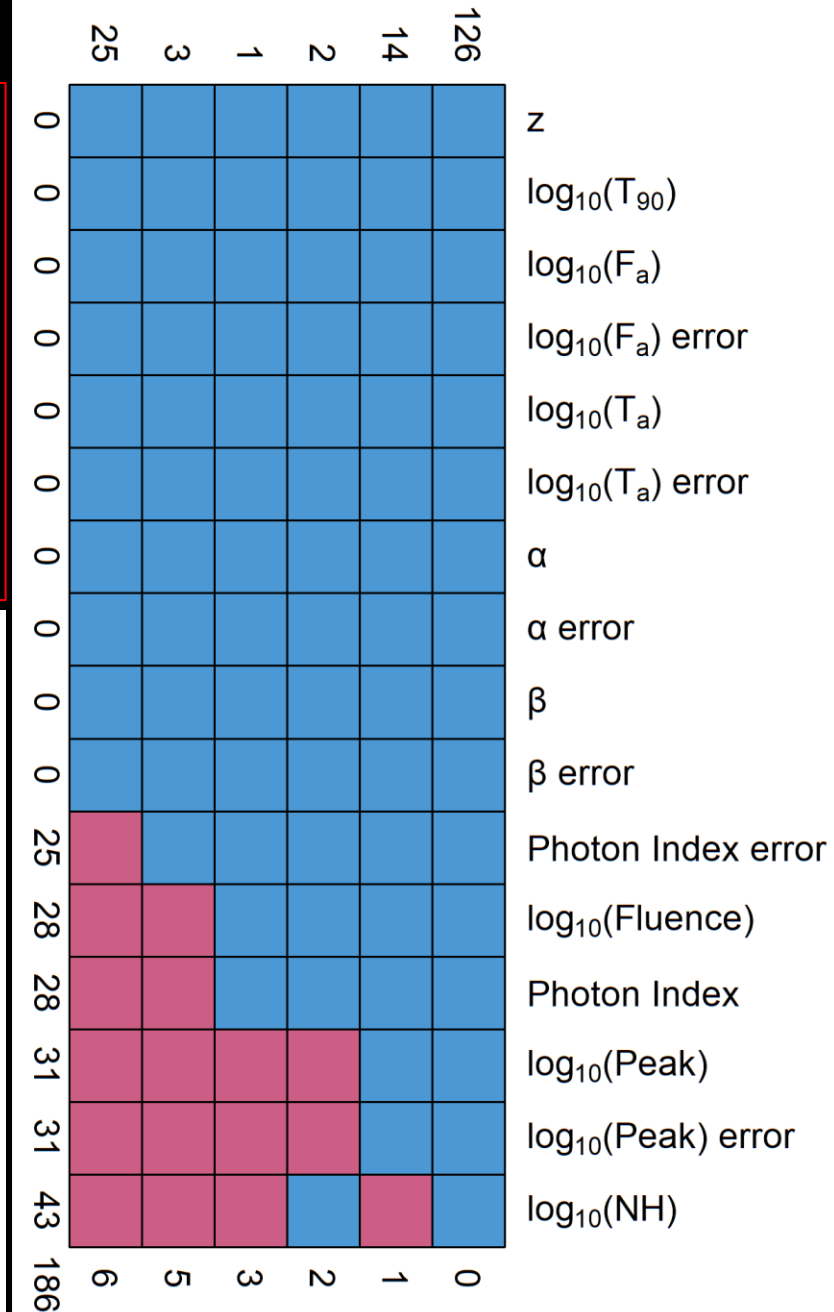
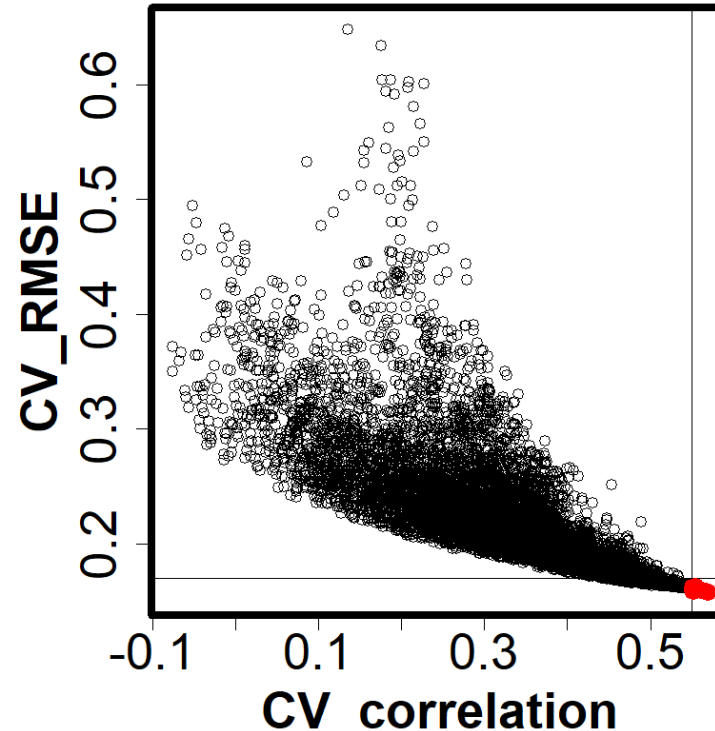


MULTIVARIATE IMPUTATION BY CHAINED EQUATION (MICE)

- 26% of 171 (45 GRBs) have missing data.
- This is a method used for imputing missing data.
- MICE uses the data present to infer the missing data.
- We use Predictive Mean Matching method for imputation.
- Previous publications have demonstrated its effectiveness

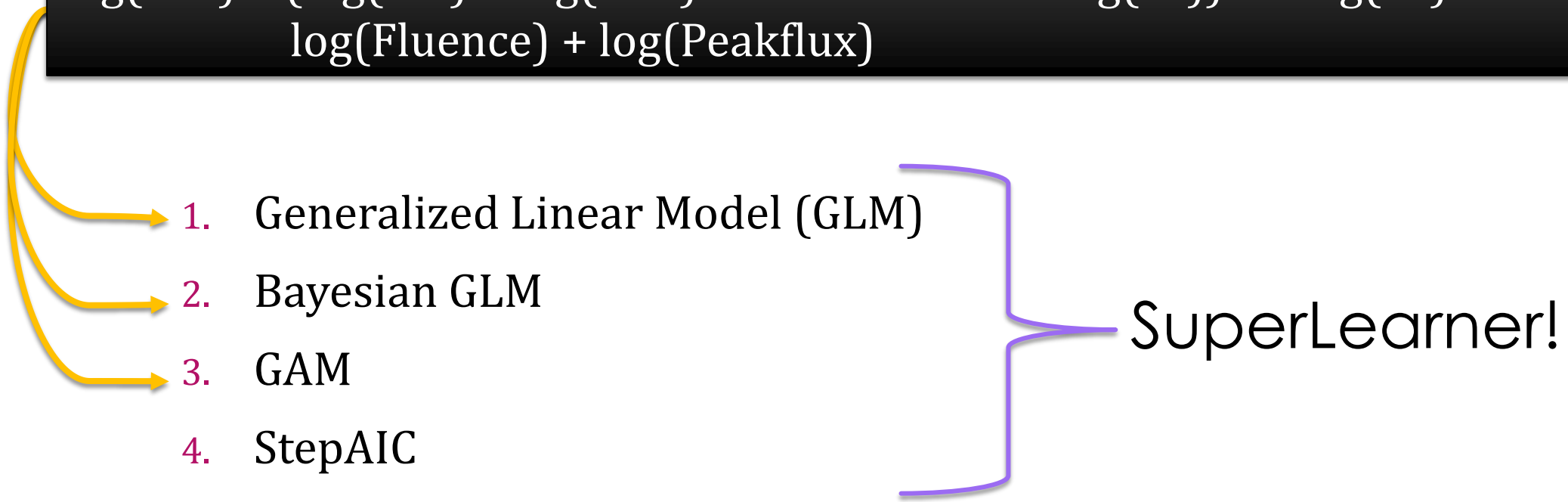
FORMULA ESTIMATION

- We combine 7 predictors to generate 16000+ formula
- Select a subset of formulas based on cross-validated results
- Further tested on a separate test set
- Formula with best correlation, RMSE and MAD are picked.



Methodology

$$\log(z+1) = (\log(\text{NH}) + \log(\text{T90}) + \text{PhotonIndex} + \log(\text{Fa}))^2 + \log(\text{Ta}) + \alpha + \beta + \log(\text{Fluence}) + \log(\text{Peakflux})$$

- 
1. Generalized Linear Model (GLM)
 2. Bayesian GLM
 3. GAM
 4. StepAIC

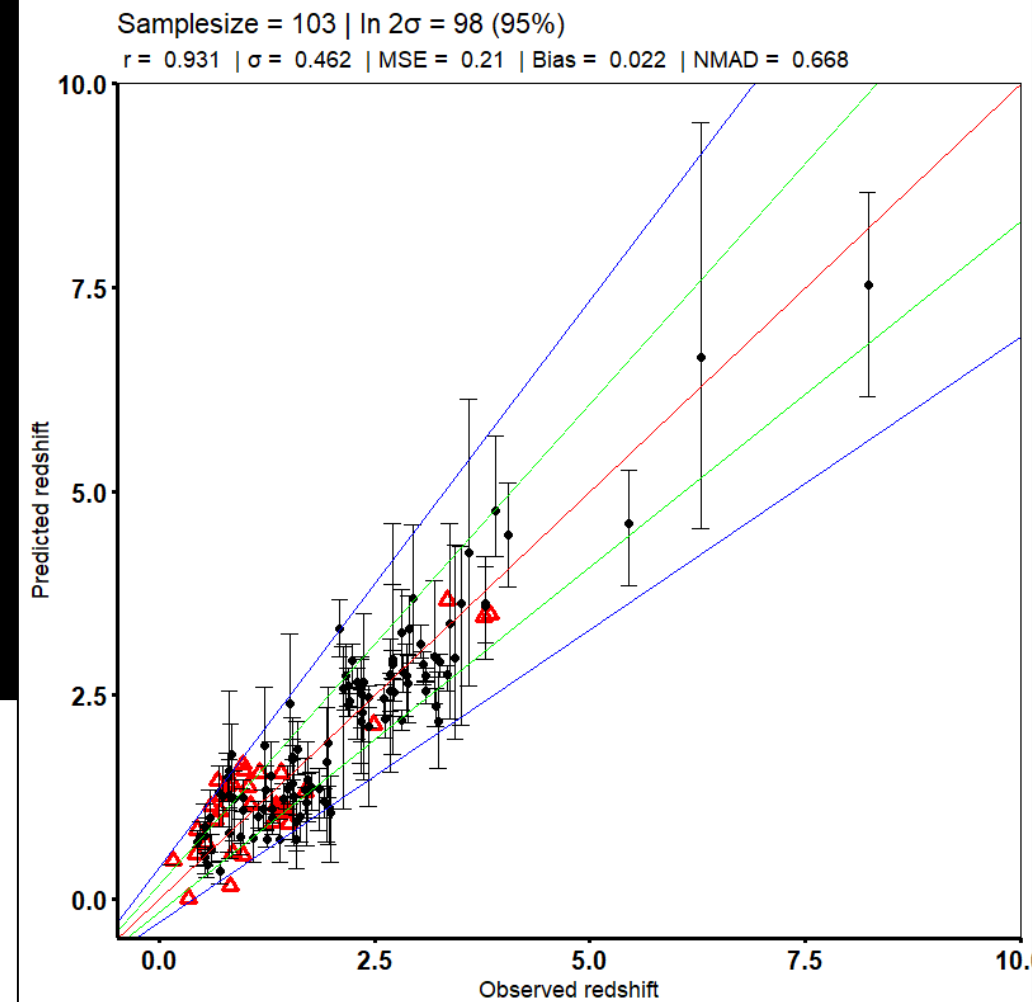
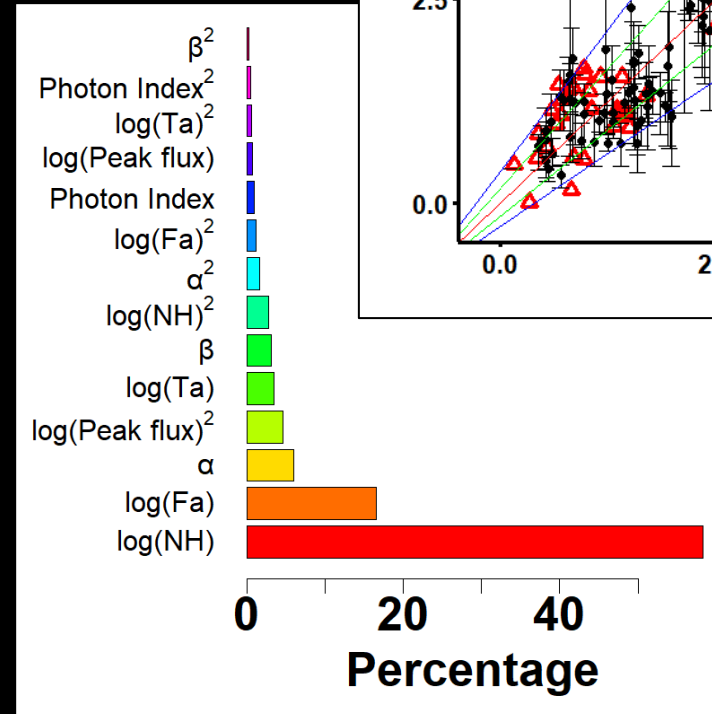
SuperLearner!

- We did test many other ML models, but these 4 obtained the highest weights by SuperLearner consistently.
- For GLM and bayes GLM to use the formula we had to develop custom functions in SuperLearner.
- For the results we perform 10-fold cross validation 100 times and take their average.

RESULTS

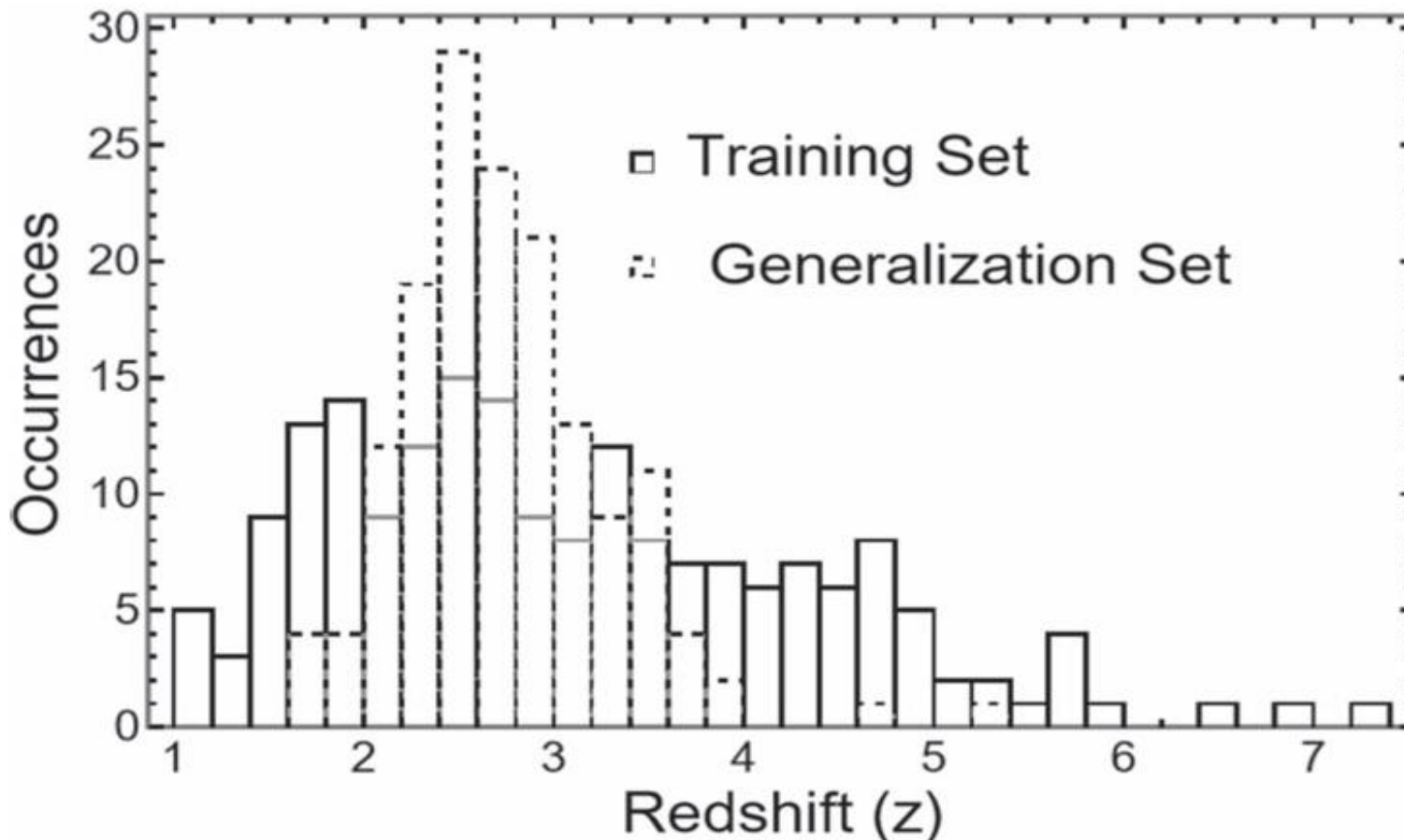
CV RESULTS

- This is after bias correction+removing outliers
- **Pearson correlation 0.93**
- **Mean square error 0.21**
- 61% higher correlation than Ukwatta et. al 2016.
- Less number of predictors (7 vs 11).
- $\log(\text{NH})$ and $\log(\text{Fa})$ most important redshift predictors.
- **Two out of top 3 predictors are plateau.**
- 103 GRBs used for LF & DRE



The optical sample and the Dainotti correlation parameters are recovered

The generalization and the training set



Based on the X-ray redshift sample

Dainotti, ... et al. 2024,
ApJS, 271, 1, id.22.

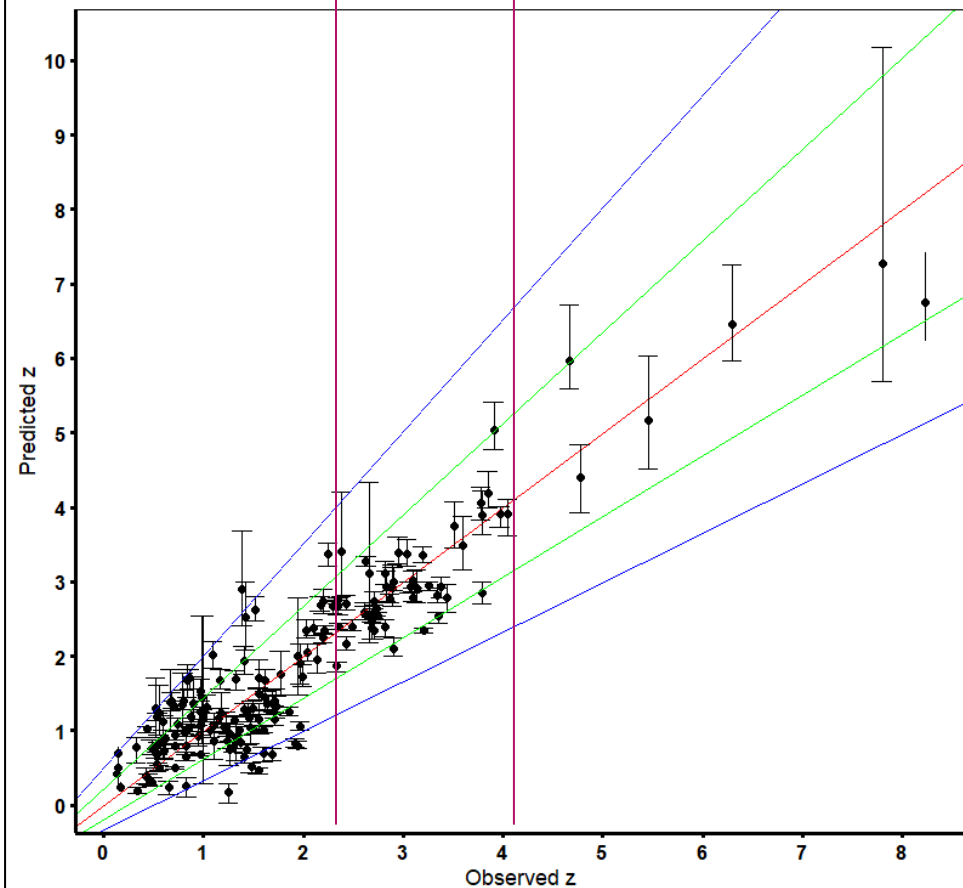
Similar differences in
Aldowma & Razzaque
2024MNRAS.529.2676A

**Distributions of $\log T_a$
and $\log N_H$ are
different between the
training set and the
generalization set**

After Optimal Transport Bias Correction (the X-ray sample) → Dainotti, et al 2024ApJS..271...22D

Remove
outlier

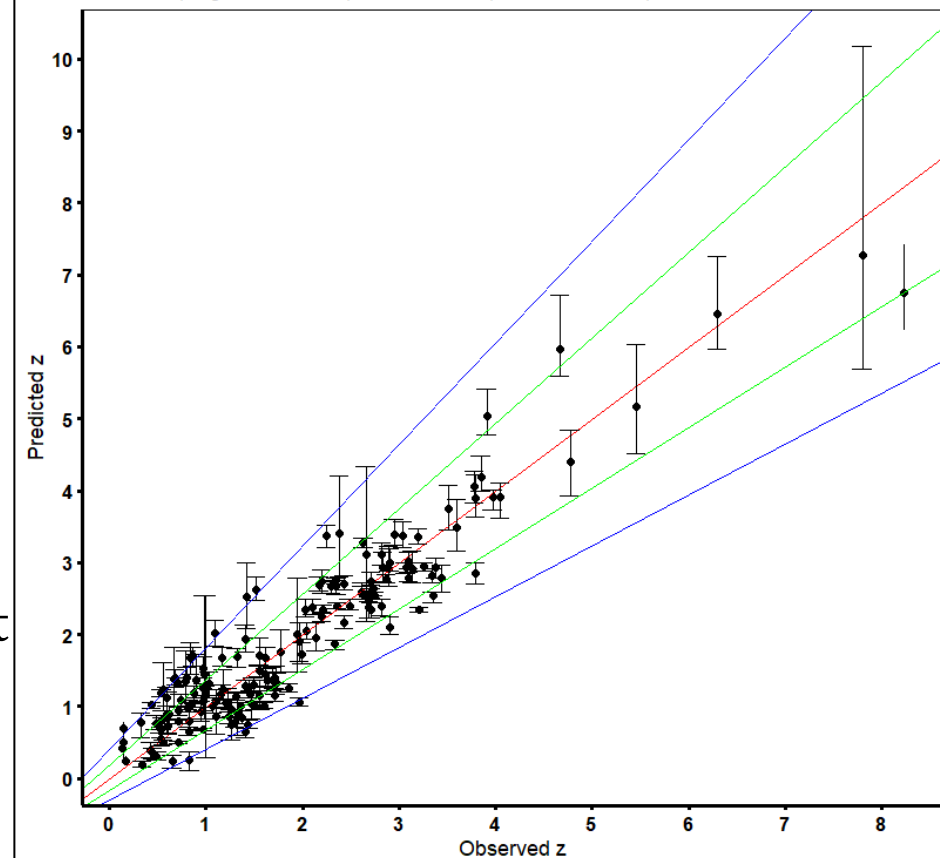
Sample size = 171 | In 2sigma = 162 (95%) | In sigma = 120 (70%)
r = 0.922 | Sigma = 0.508 | RMS = 0.51 | Bias = 0.003 | NMAD = 0.667



162 GRBs:
Pearson correlation,
r=0.94, RMSE = 0.46
Bias = 0.03

Similar results are
obtained independently
Aldowma & Razzaque
2024MNRAS.529.2676A
Comparing with Ukwatta et
al. 2016,
we have **56% increase of
the Pearson correlation**

Sample size = 162 | In 2sigma = 149 (92%) | In sigma = 112 (69%)
r = 0.94 | Sigma = 0.454 | RMS = 0.46 | Bias = -0.028 | NMAD = 0.64



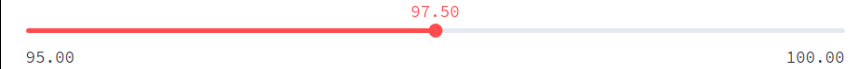
THE USABLE WEB-APP

- a webapp
- Expanded methodology with more data
- Applying to GRBs without redshift
- Narendra, Dainotti, .. al. A&A, 2025A&A...698A..92N

GAM/GLM Formula Generator

Input: GRBs with redshift Output: GAM / GLM formulas and relevant plots.

Choose correlation quantile cut-off



Choose RMSE quantile cut-off



Choose a file (.csv or .txt)

Redshift Estimator

Please enter the generalization data.

Choose your input method:

- ☒ Upload a file
- ☐ Enter parameters manually
- ☐ Apply MICE
- ☐ Do you already have modelled data?

Choose a file (.csv or .txt)

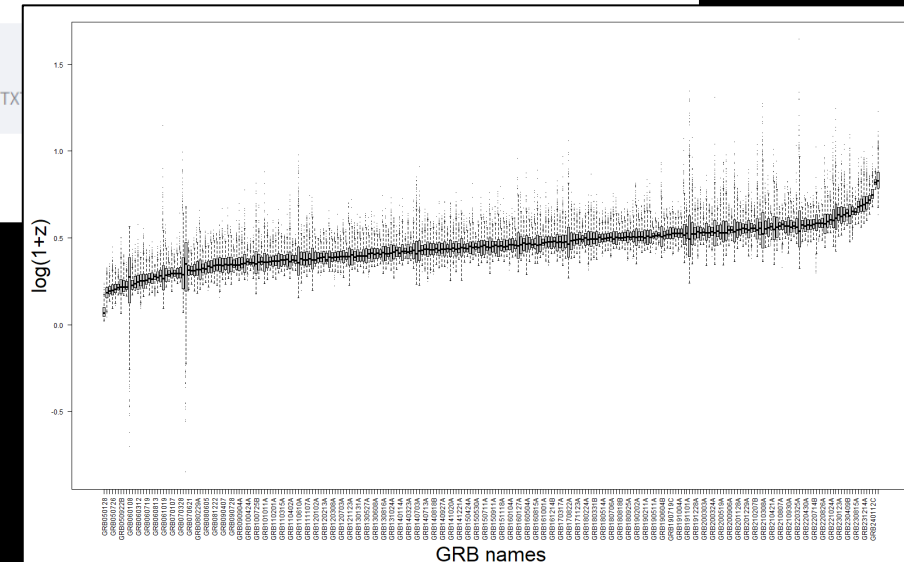


Drag and drop file here

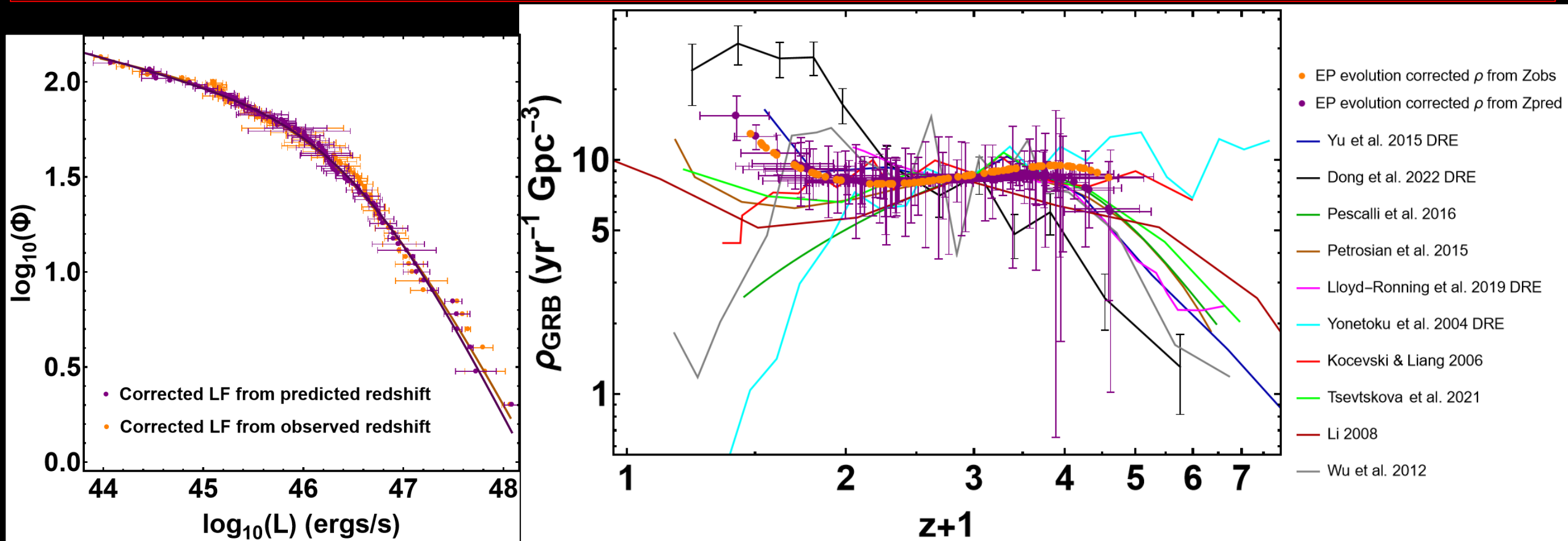
Limit 200MB per file • CSV, TXT

Submit

Browse files

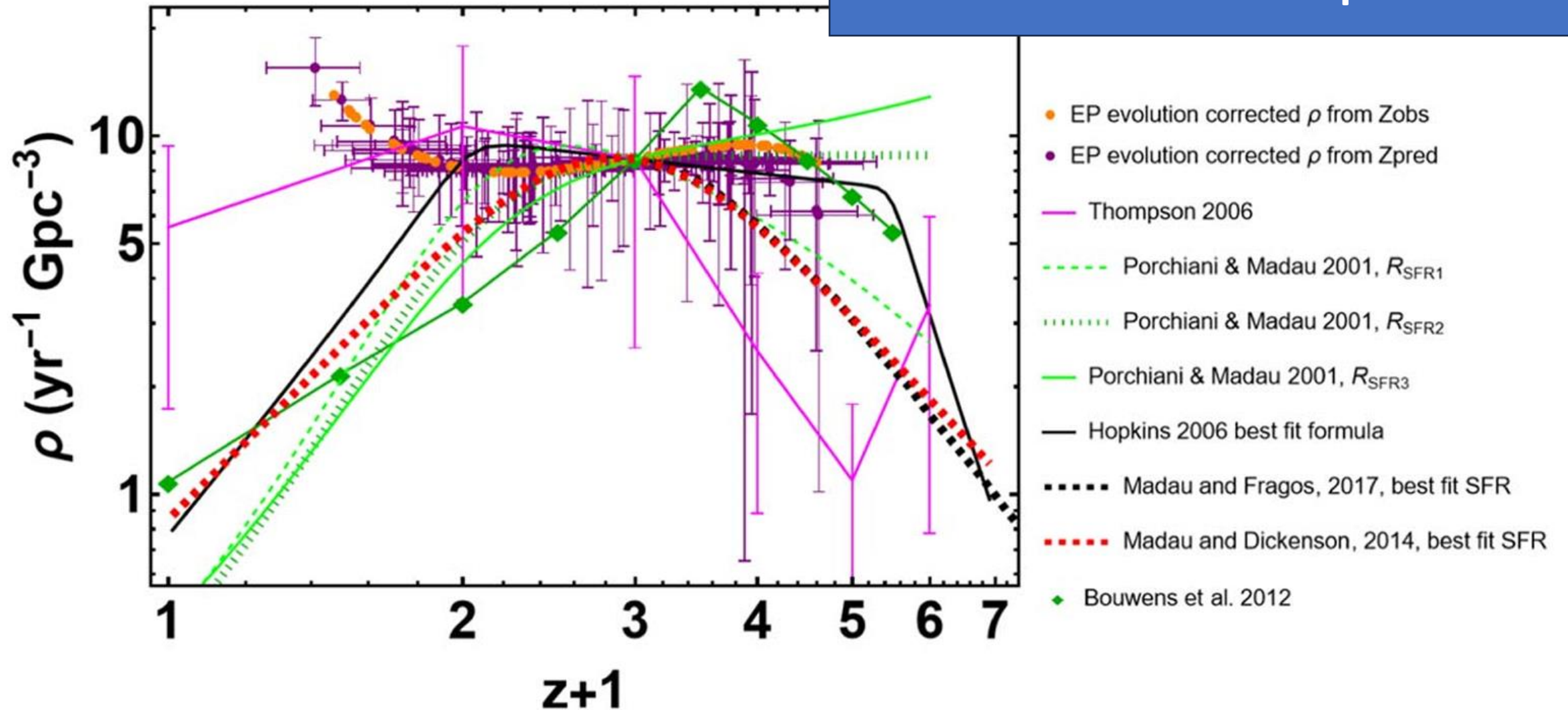


1. Using plateau features for GRB redshift prediction is effective
2. The LF from z_{obs} and z_{pred} are compatible in 1σ !
3. LGRB DRE discrepancy with SFR at low- z



THESE PREDICTIONS CAN BE USED FOR DERIVING the rates: 11 Gpc $^{-3}$ yr $^{-1}$ at $z < 1$. One order of magnitude higher than the predicted values

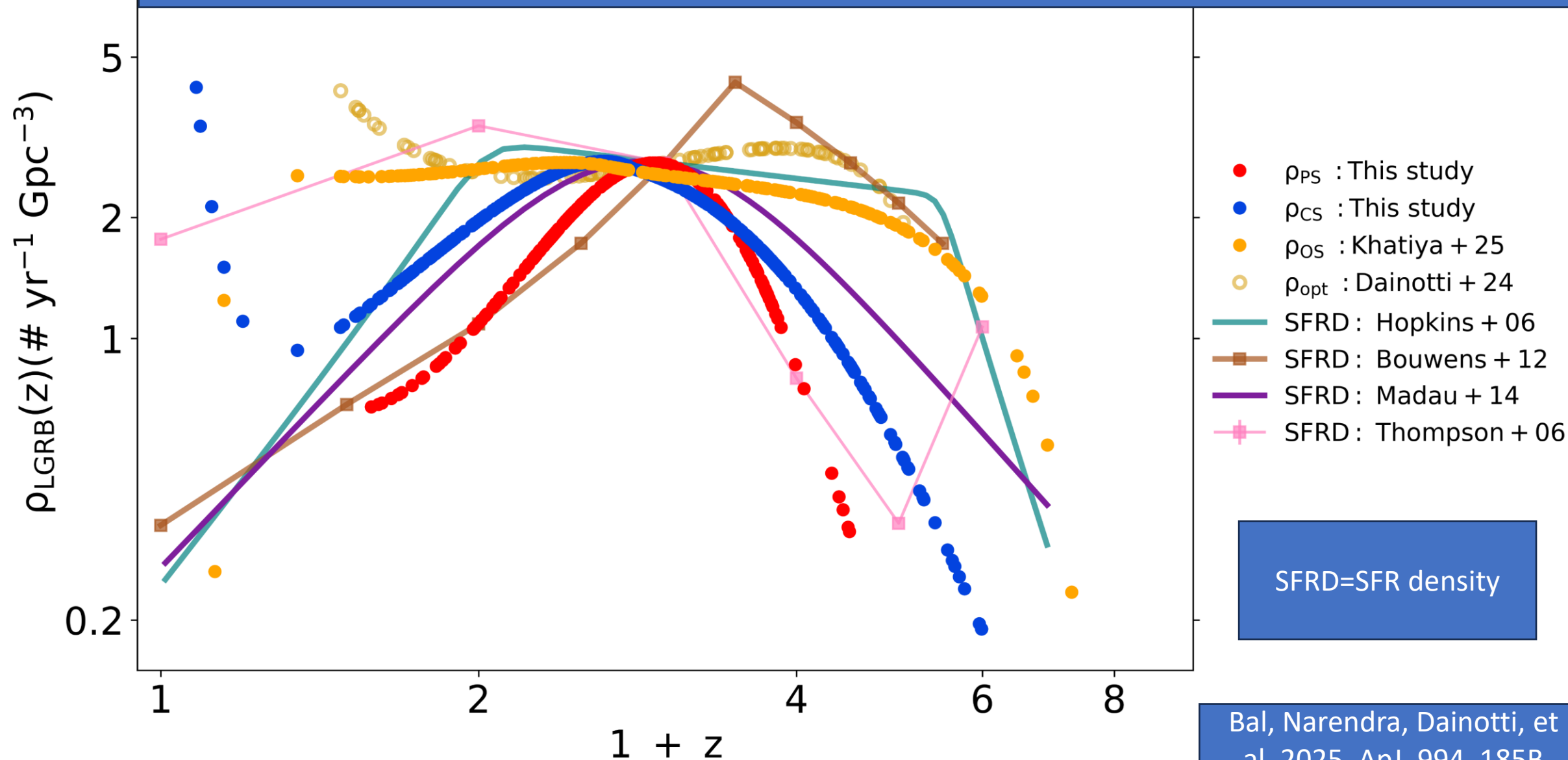
GRB rate vs. SFR with optical data



Dainotti et al. 2024, ApJL, 967, L30

Optical data=103 GRBs

Combined data (451 GRBs)=
Observed redshifts (223 GRBs) + inferred redshifts (228 GRBs)



Superlearner applied for Classification

24

GRB Redshift Classifier to Follow up High-redshift GRBs Using Supervised Machine Learning

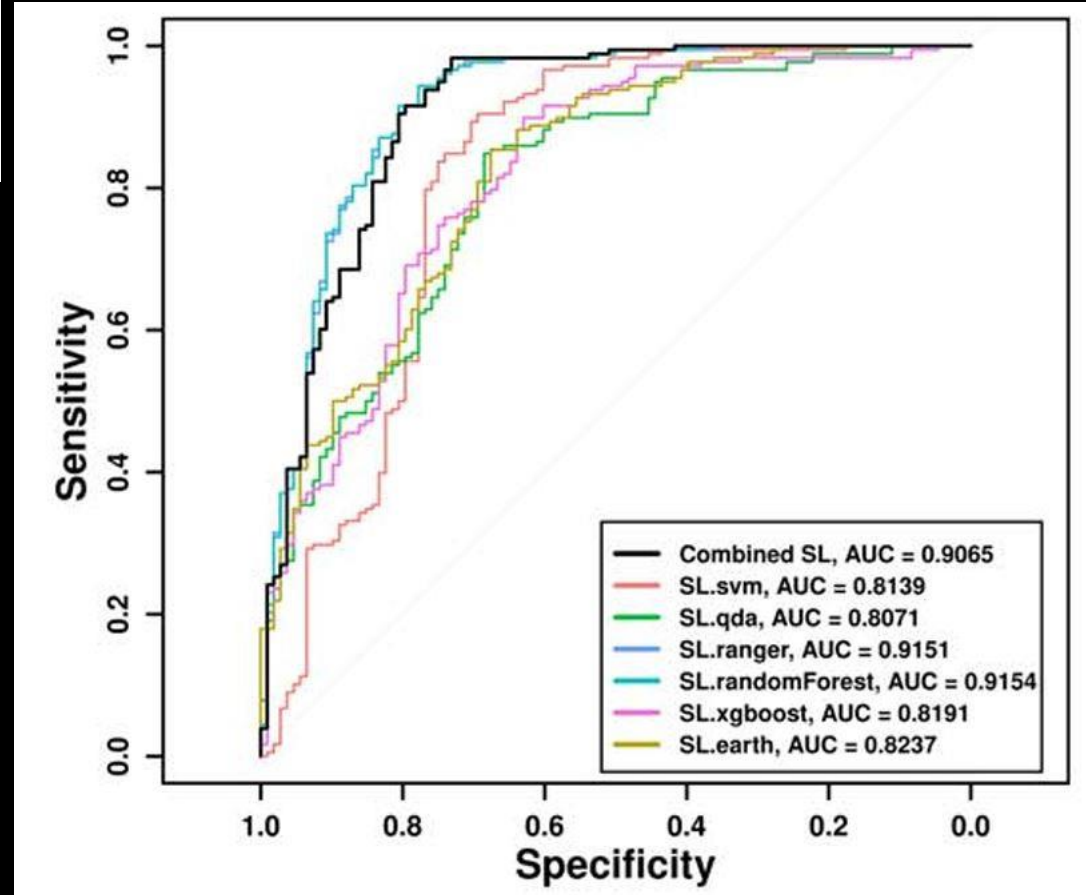
Show affiliations

Dainotti, Maria Giovanna  ; Bhardwaj, Shubham  ; Cook, Christopher ; Ange, Joshua  ; Lamichhane, Nishan  ; Bogdan, Malgorzata  ; McGee, Monnie  ; Nadolsky, Pavel  ; Sarkar, Milind  ; Pollo, Agnieszka  ; Nagataki, Shigehiro 

The Astrophysical Journal Supplement Series,
Volume 277, Issue 1, id.31, 34.

Sensitivity = $TP / (TP + FN)$,
TP=true positive, FN=false
negative

A sensitivity of 87% and 89% with a balanced sampling for both $z_t = 3.0$ and $z_t = 3.5$, respectively → 9% and 11% increase in the sensitivity over random forest used alone.



Conclusions:

This discrepancy persists across samples and wavelengths

- Optical (103 GRBs): Dainotti et al. 2024, ApJL, 967, L30.
- X-ray (223 GRBs): Khatiya, Dainotti et al. 2024, ApJ, 990, 69.
- Observed + inferred z (451 GRBs) – Bal, Narendra, Dainotti et al. 2026, ApJ, 994, 185B,

Champati, S., Petrosian, V., & Dainotti, M. G. (2026), *ApJ*, 1005, 77

More data are needed: future perspective transfer learning between optical via the catalog and X-ray data

The largest GRB Optical Repository: GRB LC package

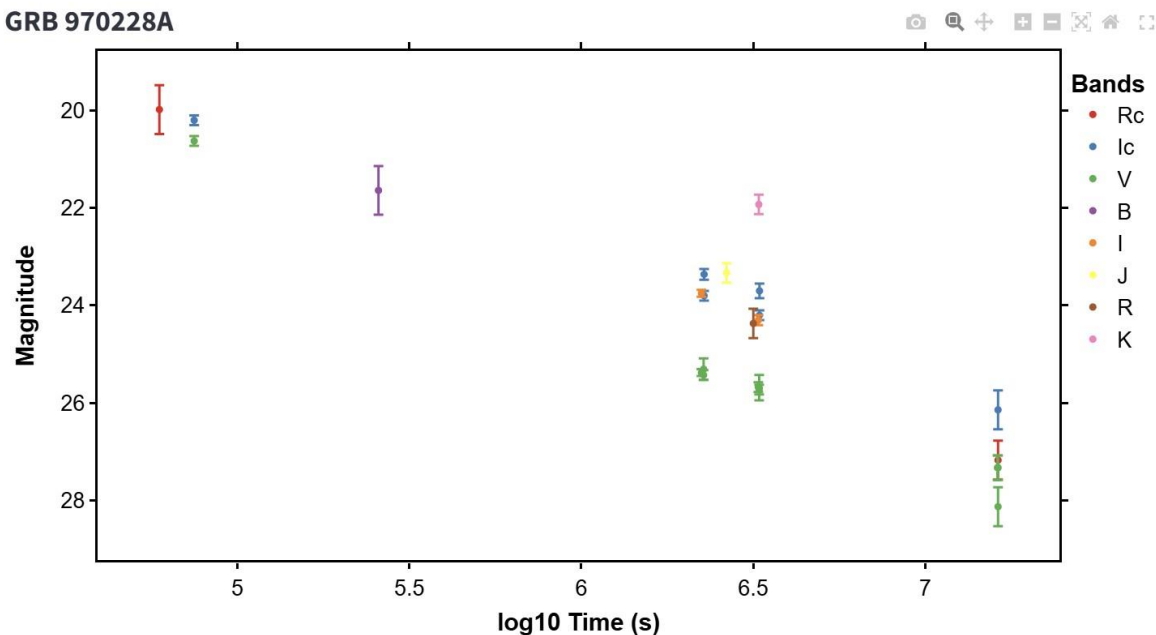
Dainotti et al., 2024, MNRAS, 2024, 533, 4023.

**535 GRBs with
redshift**
Easily searchable
455 Telescopes
570 Instruments

Gamma Ray Bursts Optical Afterglow Repository

We present the individual optical/IR photometry of each one of the 535 GRBs gathered in [Dainotti et al. 2024](#) as well the spectral information of each observation.

GRB 970228A



☐ Show raw data before homogenisation of photometric system and extinction correction

Calculate β_{opt}

Download data

- **Large to small apertures:**
14 GRBs from Subaru,
1 from Kiso at $z=3.6$, and
other Japanese Telescopes.
- **Re-calibration**
- **GRBs observed by Kiso**
(Dainotti, PI) since 2022.
- **Joint KISO and SVOM.**
- **Product: accessible via**
web-app

OPEN ACCESS

Gamma-Ray Bursts as Distance Indicators by a Statistical Learning Approach

Maria Giovanna Dainotti^{16,1,2,3} , Aditya Narendra^{16,4,5} , Agnieszka Pollo^{5,6} , Vahé Petrosian^{7,8,9} , Malgorzata Bogdan^{10,11} , Kazunari Iwasaki^{1,2,12} , Jason Xavier Prochaska¹³ , Enrico Rinaldi¹⁴ , and David Zhou¹⁵ 

OPEN ACCESS

Inferring the Redshift of More than 150 GRBs with a Machine-learning Ensemble Model

Maria Giovanna Dainotti^{1,2,3,4,5} , Elias Taira⁶, Eric Wang⁷, Elias Lehman⁸ , Aditya Narendra^{9,10}, Agnieszka Pollo^{10,11} , Grzegorz M. Madejski¹², Vahe Petrosian¹² , Malgorzata Bogdan^{13,14}, Apratim Dey¹⁵  [Show full author list](#)

OPEN ACCESS

Predicting the Redshift of Gamma-Ray Loud AGNs Using Supervised Machine Learning. II

Aditya Narendra¹ , Spencer James Gibson², Maria Giovanna Dainotti^{3,4} , Malgorzata Bogdan^{5,6}, Agnieszka Pollo^{1,7}, Ioannis Liodakis⁸, Artem Poliszczuk⁷, and Enrico Rinaldi^{9,10} 

THE ASTROPHYSICAL JOURNAL, 990:69 (11pp), 2025 September 1

© 2025. The Author(s). Published by the American Astronomical Society.

OPEN ACCESS

Probing Evolution of Long Gamma-Ray Burst Properties through Their Cosmic Formation History

Nikita S. Khatiya¹ , Maria Giovanna Dainotti^{2,3,4,5,6} , Aditya Narendra^{7,8} , Dhruv S. Bal¹ , Aleksander Ł. Lenart⁸ , and Dieter H. Hartmann¹ 

<https://doi.org/10.3847/1538-4357/adf219>



Thank you!

