RAPID, MACHINE-LEARNED RESOURCE ALLOCATION: APPLICATION TO HIGH-REDSHIFT GAMMA-RAY BURST FOLLOW-UP

A. N. MORGAN1, JAMES LONG2, JOSEPH W. RICHARDS1,2, TAMARA BRODERICK2, NATHANIEL R. BUTLER1,3, AND JOSHUA S. BLOOM1

1 Department of Astronomy, University of California, Berkeley, CA 94720-3411, USA; amorgan@astro.berkeley.edu
2 Department of Statistics, University of California, Berkeley, CA 94720-3860, USA
3 Department of Physics, Arizona State University, Tempe, AZ 85287, USA

Received 2011 October 13; accepted 2011 December 1; published 2012 February 3

ABSTRACT

As the number of observed gamma-ray bursts (GRBs) continues to grow, follow-up resources need to be used more efficiently in order to maximize science output from limited telescope time. As such, it is becoming increasingly important to rapidly identify bursts of interest as soon as possible after the event, before the afterglows fade beyond detectability. Studying the most distant (highest redshift) events, for instance, remains a primary goal for many in the field. Here, we present our Random Forest Automated Triage Estimator for GRB redshifts (RATE GRB-\(z\)) for rapid identification of high-redshift candidates using early-time metrics from the three telescopes onboard Swift. While the basic RATE methodology is generalizable to a number of resource allocation problems, here we demonstrate its utility for telescope-constrained follow-up efforts with the primary goal to identify and study high-\(z\) GRBs. For each new GRB, RATE GRB-\(z\) provides a recommendation—based on the available telescope time—of whether the event warrants additional follow-up resources. We train RATE GRB-\(z\) using a set consisting of 135 Swift bursts with known redshifts, only 18 of which are \(z > 4\). Cross-validated performance metrics on these training data suggest that \(\sim 56\%\) of high-\(z\) bursts can be captured from following up the top 20\% of the ranked candidates, and \(\sim 84\%\) of high-\(z\) bursts are identified after following up the top \(\sim 40\%\) of candidates. We further use the method to rank 200+ Swift bursts with unknown redshifts according to their likelihood of being high-\(z\).

Key words: gamma-ray burst: general – methods: data analysis – methods: statistical

Online-only material: machine-readable tables

1. INTRODUCTION

As the most luminous electromagnetic explosions, gamma-ray bursts (GRBs) offer a unique probe into the distant universe—but only if their rapidly fading afterglows are observed before dimming beyond detectability (e.g., Wijers et al. 1998; Miralda-Escude 1998; Lamb & Reichart 2000; Kawai 2008; McQuinn et al. 2008). Since the launch of the Swift satellite in 2004 November (Gehrels et al. 2004), more than 170 long-duration Swift GRBs have had measured redshifts, but only a handful fall into the highest redshift range that allows for the probing of the earliest ages of the universe, up to less than a billion years after the big bang (Figure 1). With a limited budget of large-aperture telescope time accessible for deep follow-up, it is becoming increasingly important to rapidly identify these GRBs of interest in order to capture the most interesting events without spending available resources on more mundane events.

Along with quasars (e.g., Mortlock et al. 2011) and NIR-dropout Lyman-break galaxies (e.g., Bouwens et al. 2010, 2011), GRBs have been established as among the most distant objects detectable in the universe, with a spectroscopically confirmed event at \(z = 8.2\) (GRB 090423; Tanvir et al. 2009; Salvaterra et al. 2009) and a photometric candidate at \(z = 9.4\) (GRB 090429B; Cucchiara et al. 2011b). Such observations can provide valuable constraints on star formation in the early universe, illuminate the locations and properties of some of the earliest galaxies and stars, and probe the epoch of reionization (e.g., Tanvir & Jakobsson 2007, and references therein). Further, the relatively simple spectra of GRB afterglows compared to other cosmic lighthouses make it easier to both identify their redshifts and extract useful spectral features such as neutral hydrogen absorption signatures for the study of cosmic reionization (e.g., Miralda-Escude 1998; Barkana & Loeb 2004; Totani et al. 2006; McQuinn et al. 2008). However, such benefits can only be realized if spectra are obtained with large-aperture telescopes before the afterglow fades beyond the level required to obtain a useful signal, typically within a day after the GRB.

As such, there has been a long-standing effort to extract a measure of a GRB’s redshift from its early-time, high-energy signal, with a primary goal of the rapid identification of high-\(z\) candidates. This might appear in principle to be a straightforward exercise; for instance, distant GRBs should on average appear fainter and of longer duration than nearby events due to distance and cosmological time dilation, respectively. In practice, however, the large intrinsic diversity of GRBs, as well as thresholding effects, confounds the straightforward use of early-time observations in divulging redshift and other important properties. While much effort has gone into tightening the correlations between high-energy properties in order to homogenize the sample for use as a luminosity (and hence distance/redshift) predictor (e.g., Amati et al. 2002; Ghirlanda et al. 2004; Firmani et al. 2006; Schaefer 2007), there has been significant debate as to whether some of these relations are actually due to thresholding effects specific to the detectors rather than intrinsic physical properties of the GRBs (e.g., Friedman & Bloom 2005; Butler et al. 2007, 2009). Regardless, whether or not these inferred relationships are actually physical or simply detector effects would not affect their utility as a detector-specific parameter prediction tool. By restricting ourselves to Swift events only, we avoid the uncertainty of whether certain correlations remain when using different detectors.
Figure 1. Redshift distribution of the 135 long-duration Swift GRBs in our sample (Table 2). For the purposes of this study, “high” redshift is defined as those bursts with redshifts larger than $z = 4$, which corresponds to approximately $1\sigma$ above the mean of the distribution. In our sample, 18 bursts fall into this category (black, and in inset). In determining age since the big bang, we assume a cosmology with $h = 0.71$, $\Omega_m = 0.3$, and $\Omega_L = 0.7$. Solid lines show the cumulative number of GRBs as a function of redshift for high-$z$ bursts (gray) and all bursts (black).

With this in mind, we set out to search for indications of high-redshift GRBs in the rich, mostly homogeneous data set provided by 6+ years of GRB observations by the three telescopes onboard Swift (Burst Alert Telescope, BAT, Barthelmy et al. 2005; X-Ray Telescope, XRT, Burrows et al. 2005; Ultraviolet/Optical Telescope, UVOT, Roming et al. 2005). Past studies exploring high-$z$ indicators have used hard cuts on certain features such as UVOT afterglow detection, burst duration, and inferred hydrogen column density (e.g., Grupe et al. 2007; vanden Berk et al. 2008; Ukwatta et al. 2009), regression on such features (Koen 2009, 2010), and combinations of potential GRB luminosity indicators (Xiao & Schaefer 2009, 2011). In this work, we take a different approach by utilizing supervised machine learning algorithms, specifically Random Forest (RF) classification, to make follow-up recommendations for each event automatically and in real time. Particular attention is paid to careful treatment of performance evaluation by using cross-validation (Section 4), a robust methodology to guard against overfitting and the circular practice of testing hypotheses using the same data that suggested (and constrained) them.

The primary driving force of this study is simple: given limited follow-up time available on telescopes, we want to maximize the time spent on high-$z$ GRBs. To this end, we provide a deliverable metric, explained in Section 3.2, to assist in the decision-making process on whether to follow up a new GRB. Real-time distribution of this metric is available for each new Swift trigger via the Web site and RSS feed.

The structure of this paper is as follows: in Section 2, we outline the collation of the data, and describe the particular GRB features utilized in redshift classification. In Section 3, the RF algorithm is detailed, along with some specific challenges posed by this particular data set. Performance metrics of the classifiers are presented in Section 4, and in Section 5 we discuss the results of testing the classifiers on additional GRBs, both with and without known redshifts. Finally, our conclusions are given in Section 6.

2. DATA COLLECTION

The Swift BAT constantly monitors 1.4 sr on the sky over the energy range 15–150 keV. GRB triggering can occur either by a detection of a large gamma-ray rate increase in the BAT detectors (“rate trigger”), or a fainter, long-duration event recovered after on-board source reconstruction reveals a new significant source (“image trigger”). A rough ($\sim 3$ arcmin) position is determined, and if there are no overriding observing constraints, the spacecraft slews to allow the XRT and UVOT to begin observations, typically between 1 and 2 minutes after the trigger. The XRT observes in the energy range of 0.2 and 10 keV and detects nearly all of the GRBs it can observe rapidly enough, providing positional accuracies of 2–5 arcsec within minutes. The UVOT is a 30 cm aperture telescope that can observe in the range of 170–650 nm. Due to the relatively blue response of this telescope, it cannot detect highly reddened sources due to either dusty environments or (more relevant to this analysis) high-redshift origins.

At each stage in the data collection process, information is sent to astronomers on the ground via the GRB Coordinates Network (GCN) providing rapid early-time metrics. The more detailed full data are sent to the ground in ~90 minute intervals starting between roughly 1 and 2 hr after the burst. For our data set, we have collected data after various levels of processing directly from GCN notices, online tables, and automated pipelines (Butler & Kocevski 2007; Butler et al. 2007) that process and refine the data into more useful metrics. Tens of

---

4 For the purposes of this study, “high redshift” corresponds to all $z > 4.0$: a compromise between only keeping the most interesting events and having enough data to train on. However, we have explored performance of different redshift cuts; see Section 4.3.

5 http://rate.grbz.info/

6 http://rate.grbz.info/rss.xml

7 http://gcn.gsfc.nasa.gov/

8 http://swift.gsfc.nasa.gov/docs/swift/archive/grb_table.html/
attributes and their estimated uncertainties (when available) are parsed from the various sources and collated into a common format.

In order to evaluate our full data set in an unbiased way, we restricted ourselves to using features which have been generated for all possible\(^9\) past events and are automatically generated for future events. This is the primary reason we do not include potentially useful features such as relative spectral lag (e.g., Ukwatta et al. 2010, 2012, and references therein) which has been utilized as a redshift indicator with smaller and pre-

Table 1

| Feature                      | Type               | Reference  |
|------------------------------|--------------------|------------|
| BAT rate trigger?            | BAT prompt         | GCN notices|
| \(\sigma_{\text{BAT}}\)      | BAT prompt         | GCN notices|
| \(R_{\text{peak,BAT}}\)     | BAT prompt         | GCN notices|
| \(t_{\text{BAT}}\)          | BAT prompt         | GCN notices|
| UVOT detection?             | NFI prompt         | GCN notices|
| \(N_{\text{H},\text{PC}}\)  | Processed          | Butler & Kocevski (2007) |
| \(\alpha\)                  | Processed          | Butler et al. (2007) |
| \(E_{\text{peak}}\)        | Processed          | Butler et al. (2007) |
| \(S\)                       | Processed          | Butler et al. (2007) |
| \(S/N_{\text{max}}\)       | Processed          | Butler et al. (2007) |
| \(T_{90}\)                  | Processed          | Butler et al. (2007) |
| \(P_{\text{z}>4}\)         | Processed          | Butler et al. (2010) |

\(^9\) Even with the restriction of observation by all three Swift telescopes, certain features derived from model fits are nonetheless in calculable for certain GRBs from the available data. See Section 3.1.1 for how our algorithm treats missing values.

\(^10\) For 14 events in our test set, the SWIFT\_BAT\_POSITION notice was not available on the online repository, primarily due to satellite downlink problems at the time of discovery. For these events, the relevant parameters were extracted directly from the Swift\_TDRSS database (http://heasarc.nasa.gov/W3Browse/all/swifttdrss.html).

\(^{11}\) \(T_{90}\) alone is not a strong enough discriminator to definitively assign a particular GRB to one class or another ("short" versus "long"; see Levesque et al. 2010 for discussion). In this study, we will accept the few errant bursts from the "short" class included in our sample as additional noise in our method.

\(^{12}\) The reason for these missing data is almost always due to observing constraints from the GRB being too close to the Sun, Moon, or Earth at the time of discovery. Not removing these bursts would introduce a bias in the sample due to the fact that events without a rapid XRT position are less likely to lead to an afterglow discovery and, hence, redshift determination. A total of 15 bursts with known \(z\) were removed because of this.
in Section 5.1. Exploratory data analysis shows preliminary indications of which of these features will be most useful for classification. Figure 2 shows several two-dimensional slices of the feature space, with the high-z bursts highlighted.

### 3. CLASSIFICATION METHODOLOGY

The resource allocation approach we have taken here naturally manifests itself as a classification problem: deciding whether or not to follow up a new event is simply a two-class problem of “observe” or “do not observe,” and the methodology presented here can be applied to any problem that can be broken up in this way. This was the primary motivation of using classification instead of a regression or “pseudo-z” approach for this study. The primary disadvantage of classification for the particular problem of high-redshift identification is that all instances above and below the class division (chosen here to be $z = 4$) are treated equally; e.g., a burst with $z = 4.01$ has the same

---

#### Table 2

| GRB     | $\hat{Q}_{\text{train}}$ | $z$     | References                  |
|---------|--------------------------|---------|-----------------------------|
| 050223  | 4.30e-01                 | 0.5915  | Berger & Shin (2006)        |
| 050315  | 3.57e-01                 | 1.949   | Kelson & Berger (2005)      |
| 050318  | 6.86e-01                 | 1.44    | Berger & Mulchaey (2005)    |
| 050319  | 5.90e-01                 | 3.2425  | Fynbo et al. 2005a; Jakobsson et al. 2006a; Fynbo et al. 2009a |
| 050416A | 7.68e-01                 | 0.6535  | Cenko et al. (2005)         |

(This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content.)

#### Table 3

| GRB     | $\alpha$ | $E_{\text{peak}}$ (keV) | $S$ (erg cm$^{-2}$) | $S/N_{\text{max}}$ | $N_{\text{H,pc}}$ (10$^{22}$ cm$^{-2}$) | $T_{90}$ (s) | $\sigma_{\text{BAT}}$ | $R_{\text{peak,BAT}}$ (counts s$^{-1}$) | Rate Trigger | $t_{\text{BAT}}$ (s) | UVOT Detect | $P_{z>4}$ |
|---------|----------|--------------------------|---------------------|---------------------|---------------------------------------|--------------|----------------------|-------------------------------------|--------------|----------------------|------------|----------|
| 050223  | $-1.74e+00$ | 6.70e+01                | 8.75e-07             | 1.34e+01            | $-2.37e-01$                           | 1.74e+01    | 9.00e+00            | 7.26e+02                           | Yes          | 8.19e+00            | No         | 1.74e-01 |
| 050315  | $?$       | 4.33e+01                | 4.32e-06             | 4.37e+01            | 9.60e-02                             | 9.46e+01    | 8.00e+00            | 2.60e+02                           | Yes          | 1.02e+00            | No         | 9.27e-02 |
| 050318  | $-1.22e+00$ | 5.01e+01                | 1.41e-06             | 4.90e+01            | 1.80e-02                             | 3.10e+01    | 9.00e+00            | 2.05e+02                           | Yes          | 5.12e-01            | Yes        | 6.29e-02 |
| 050319  | $-2.00e+00$ | 4.47e+01                | 1.87e-06             | 1.82e+01            | 1.50e-02                             | 1.54e+02    | 1.00e+01            | 2.63e+02                           | Yes          | 1.02e+00            | Yes        | 1.48e-01 |
| 050416A | $-7.24e-01$ | 1.50e+01                | 3.40e-07             | 1.75e+01            | 2.34e-01                             | 2.91e+00    | 1.10e+01            | 1.65e+02                           | Yes          | 5.12e-01            | Yes        | 4.35e-03 |

(This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content.)

---

Figure 2. Plot of a selection of early-time Swift features (Table 1) against each other. The gray points show the full distribution of Swift GRBs. Bursts with known redshifts are black, and the 18 known events with redshifts greater than 4 are overplotted in red. In the histogram text boxes, $N$ shows how many instances of that feature in total are shown (anything less than the full number of instances is due to the value of that feature being unknown for certain instances), and Max shows the maximum number of instances in any particular bin.
influence on our inference about “high” bursts as a burst with $z = 8$.\textsuperscript{13} However, classification has advantages over regression in that it is a conceptually much simpler problem, and most of the difficulties encountered due to the unbalanced, small data set of interest here would only be aggravated by an extension to regression. Further, our approach capitalizes on the fact that one of our predictors (lack of UVOT detection) is itself a binary feature with an understood physical connection to redshift.\textsuperscript{14}

### 3.1. Random Forest Classification

A supervised classification algorithm uses a set of training data of known class to estimate a function for assigning data points to classes based on their features. The statistics and machine learning communities have developed many classification algorithms, including Support Vector Machines (SVMs), Naïve Bayes, Neural Networks, and Gaussian Mixture Models. We use RF (Breiman 2001) for its ability to select important features, resist overfitting the data, model nonlinear relationships, handle categorical variables, and produce probabilistic output. These strengths, along with a record of attaining very high classification accuracy relative to other algorithms, have led to widespread use of RF in the astronomy community (e.g., Bailey et al. 2007; Carliles et al. 2010; Dubath et al. 2011; O’Keefe et al. 2009; Richards et al. 2011). In this work, we utilized custom R software built around the randomForest package to generate classifiers and evaluate performance.

RF is an ensemble classifier that averages together the outputs from many decision trees, a common example of which is Classification and Regression Trees (CART; Breiman et al. 1984). In RF, the decision trees are constructed by recursive binary splitting of the high-dimensional feature space, where each split is performed with respect to a particular feature. For example, the decision tree might split the data on feature $S/N_{\text{max}}$ using value 100, in which case all observations with $S/N_{\text{max}} > 100$ are placed in one group and the rest placed in the second group. As these are binary splits, for convenience we henceforth refer to observations going “left” or “right” of each split as an analogue for the decision made at that split. For each split, the feature and specific split point are chosen so as to best separate the observations into the classes by using some objective function. We use the Gini index, a standard objective function for classification (Breiman et al. 1984). At any given node in a tree and some proposed split $s$, let $N_{l,h} = \text{number of high-priority (in our case, high-$z$) events that go to the left of the split and } N_{l,l} = \text{number of low-priority events that go left. Define } N_{r,h} \text{ and } N_{r,l}\text{ similarly, replacing left with right. Let } N_{l} = N_{l,h} + N_{l,l}\text{, the total number of observations that go left. Similarly define, } N_{r} = N_{r,l} + N_{r,h}\text{, for the total number of observations that go right. The Gini criterion is defined as}

$$\frac{N_{l}}{N_{l} + N_{r}} \left( \frac{N_{l,h}}{N_{l}} \right) \left( \frac{N_{l,l}}{N_{l}} \right) + \frac{N_{r}}{N_{l} + N_{r}} \left( \frac{N_{r,h}}{N_{r}} \right) \left( \frac{N_{r,l}}{N_{r}} \right), \quad (1)$$

\textsuperscript{13} This of course would not be an issue when applying the RATE methodology to a problem with more well-defined class boundaries, such as prioritizing follow-up of a particular rare class of transient event.

\textsuperscript{14} Bursts with a UVOT detection must be $z < 5$ due to the Lyman cutoff. This is due to the fact that photons with wavelengths smaller (thus higher energy) than the Lyman limit of $\lambda = 912\ \text{Å}$ would be almost completely absorbed by neutral gas in the host galaxy and intergalactic star-forming regions. A redshift of $z = 5$ might therefore be considered a natural cutoff point for the high-$z$ class, but due to so few training events at this high redshift ($N_{z>5} = 8$), we opted for the more conservative cutoff point of $z = 4$ ($N_{z>4} = 18$).

and the split that minimizes this value over the random subset of features considered at each node\textsuperscript{15} is chosen. For instance, in the ideal case where the split on a particular feature completely separates all the instances of the two classes from each other, the Gini index reaches a minimum of 0. The splitting is done recursively, continuing down each subgroup until all of the observations in each final group (“terminal node”) are of a single class. The process is known as “growing a tree” because each split can be visualized as generating two branches from a single branch to produce a tree-like structure. Once a tree is constructed from the training data, each new observation starts at the root node (the top split in the tree) and, recursively, the splitting rules determine the terminal node to which the observation belongs. The observation is assigned to the class of the terminal node.

To create the RF classifier, a sufficiently large\textsuperscript{16} number of decision trees are constructed, resulting in a “forest.” Each decision tree is generated from an independent bootstrap sample (Efron 1982). Samples are drawn with replacement from the original data set, resulting in a new data set of the same size as the original, with on average 2/3 of the original observations present at least once. Additionally, only a random subset of the features is eligible for splitting at each node. Many decision trees are grown with each tree slightly different due to the bootstrap sampling and random selection of features at each split. RF classifies new observations by averaging the outputs of each tree in the ensemble.

Training observations can be classified by using all trees where that observation was not used in the bootstrap sampling stage. This produces estimates of error rates and class probabilities for each observation that are not overfitted to the training data. Error rates and probabilities computed using this method are known as “out-of-bag” estimates.

#### 3.1.1. Missing Feature Values

As mentioned in Section 2, certain features, namely $\alpha$ and $N_{\text{UVOT,pc}}$, were occasionally unable to be determined from model fits to the data and are thus missing for certain observations. We handle missing values by imputation, where missing values for features are assigned estimated values. For missing values of continuous features, we assigned the median of all observations for which that feature is non-missing. Missing categorical features are assigned the mode of all observations for which the feature is non-missing. This is one of the simplest imputation methods and has the advantage of being transparent and computationally cheap. We experimented with a more sophisticated imputation method, MissForest, that iteratively predicts the missing values of each feature given all the other features (Stekhoven & Bühlmann 2011), but as it produced similar error rates to median imputation, we opted for the latter, simpler approach in our final classifier.

#### 3.1.2. Class Imbalance

A further challenge in this data set is the imbalance between classes. We are training on 135 bursts, only 18 of which are in the...
high-z class—an asymmetry present in many resource allocation problems where the goal is to prioritize the rarer events. Without modification, standard machine learning classification algorithms applied to imbalanced data sets attain notoriously suboptimal performance (Chawla et al. 2004), and often result in simply classifying all unknown events as the more common class. As we care more about correctly classifying the rarer events, misclassifications of high-z events must be punished more strongly than vice versa. In RF, classes may be weighted in order to overcome the imbalance by altering the splits chosen by Gini and the probabilities assigned to classes in the terminal nodes of each tree (Chen et al. 2004).

We utilized the classwt option in the randomForest package, which accounts for class weights in the Gini index calculation (Equation (1)) when splitting at the nodes (A. Liaw 2011, private communication), similar to weighting techniques used in single CART trees (Breiman et al. 1984). If we are weighting high-priority observations (e.g., \( z > 4 \) GRBs) by \( w_h \) and low-priority observations by \( w_l \), we let

\[
\begin{align*}
N'_{l,h} &= w_h N_{l,h}, \\
N'_{l,l} &= w_l N_{l,l}, \\
N'_{r,h} &= w_h N_{r,h}, \\
N'_{r,l} &= w_l N_{r,l}.
\end{align*}
\]

Let \( N'_r = N'_{l,l} + N'_{l,h} \), the weighted total number of observations that go left. Similarly, define \( N'_r = N'_{r,l} + N'_{r,h} \) for the weighted total number of observations that go right. The Gini criterion (Equation (1)) is evaluated with the weighted values, and the split that minimizes this value is chosen. We tested a variety of weight choices by fixing \( w_l \) to be unity and varying \( w_h \) over a range of values. The results of this test are presented in Section 4.1, which demonstrates the effects of class weight choice on classifier performance.

3.2. RATE GRB-z: Random Forest Automated Triage Estimator for GRB Redshifts

With the background above in hand, we now describe our resource allocation algorithm and its utility for the prioritization of high-z GRB follow-up. In our application, the data are described in Section 2 and the classes are high- and low-redshift GRBs, with \( z = 4 \) as the boundary between the classes. Our primary goal is to provide a decision for each new GRB: should we devote further resources to this event or not? This decision may be different for each astronomer, as it is dependent on the amount of follow-up time available. Implicit in this goal is the desire to follow up on as many truly high-redshift bursts as possible, under a set of given telescope time constraints. Directly using the results of an off-the-shelf classifier for this task (i.e., strictly following up on events labeled as “high priority”) is suboptimal. If too few events are labeled as high priority, there would be an underutilization of available resources. If too many are being labeled as high priority, simply following up on the first ones available would preclude any prioritization of events within this high-priority class.

These issues can be avoided by instead tailoring the follow-up decision to the resources available (in this case, the available telescope time devoted to high-z GRB observations). The RATE method works as follows: let \( Q \) be the fraction of events one has resources to follow up on.\(^{17}\) First, we construct an RF classifier using the training data with known response (in this case redshift). We compute the probability of each training event being high priority using out-of-bag probabilities (see Section 3.1). For each new event, we obtain a probability of it being high priority using the RF classifier, and compute the fraction of training bursts that received a higher probability of being high priority than this new burst. A new burst is assigned rank \( n \), with \( n - 1 \) training events having a lower probability of being high priority. Then, for \( N \) total training bursts, we obtain a learned probability rank for the new event of \( \hat{Q} := n/(N + 1) \). This leads to a simple decision metric for each new event. If \( \hat{Q} \) is less than the desired fraction of events a particular observer wishes to follow up (\( \hat{Q} < Q \)), follow-up observations are recommended. For instance, if one can afford to follow up on \( ~30\% \) of all observable GRBs, the desired follow-up fraction is \( Q = 0.3 \), and follow-up would be recommended for all events assigned a \( \hat{Q} < 0.3 \). An illustration of this process in action is shown in Figure 3. The desired fraction of follow-up events \( Q \) can be dynamically changed without penalty; if the amount of available resources changes, one simply needs to raise or lower this cutoff value accordingly.

4. VALIDATION OF CLASSIFIER PERFORMANCE

Our training data consist of 135 bursts, 18 of which are high redshift (\( z > 4 \)). Our primary measure of performance is efficiency, defined here as the fraction of high bursts that we follow up on relative to the number of total high-z GRBs that occurred (\( N_{\text{high observed}}/N_{\text{total high}} \)). A secondary performance measure is purity, the number of followed-up events that were actually high-z (\( N_{\text{high observed}}/N_{\text{total observed}} \)). We measure performance using 10-fold cross-validation (Kohavi 1995), where 90\% of the data is used to construct a classifier and predict on the remaining 10\% of events. Each line in the following performance plots is the cross-validated performance averaged across 100 trials of 10-fold cross-validation in order to reduce variability due to randomness in training/test subset selection.

4.1. Comparison of Weight Choices

As described in Section 3.1.2, one of the primary challenges in learning on this data set is the simple fact that there are comparatively few high-z events on which to train. If simply getting the most classifications correct were the primary performance metric, as it is in many classification problems, classifying all new events as low redshift would be considered a strong classifier since so few events are in the high-z class. However, since our objective is to identify the best candidates of this rare class, we punish misclassifications of high-z GRBs more heavily to achieve higher efficiency and purity (outlined above) for a given fraction of followed-up events.

Thus, in selecting the best weight for our classifier, we compared the efficiency and purity of high-z classification for various choices of the weight \( w_h \) using the feature set shown in Table 1. While the relative probability ranking

\(^{17}\) As telescope resources are allocated by number of hours and not number of objects, we implicitly assume here that an equal amount of resource time will be allocated to each follow-up event. This is not in general the case, as objects that turn out to be particularly interesting may have additional resources spent on them. However, a user’s estimate of \( Q \) can always be adjusted without penalty as available resources change.
Figure 3. Example of the RATE GRB-\(z\) process.

The Astrophysical Journal, 746:170 (16pp), 2012 February 20
Morgan et al.

Figure 4. Bumps plot showing the cross-validated ranking prediction \(\hat{Q}\) for each GRB in the training set over a variety of weight choices. Each line corresponds to an individual GRB, colored by its observed redshift. Bursts with \(z > 4\) are plotted with a thicker line. The clustering of high-\(z\) events toward low \(\hat{Q}\) is clear, illustrating the predictive power of the classifier. The relative ranking of events remains largely stable over different penalization weights, but performance improvements at higher weights are apparent in Figure 5, which level off after a weight of 10.

of the GRBs stayed relatively stable over weight choices (Figure 4), a clear trend emerges when comparing classification performance (Figure 5). As expected, punishing misclassifications of the smaller, more desirable high-\(z\) class causes more of these rare events to be correctly identified. Beyond a weight of 10, however, a ceiling is reached where further weight increases show zero change in classification performance. This is therefore the weight chosen for all subsequent performance comparisons.

4.2. Effects of Feature Selection

As mentioned in Section 2, early testing indicated that the addition of too many features rapidly degraded the predictive power of the final classifier. This is due to a manifestation of the so-called curse of dimensionality known as the Hughes Phenomenon (Hughes 1968), where for a fixed number of training instances, the predictive power decreases as the dimensionality increases. This appears to contradict the conventional wisdom
that RF does not overfit, and thus it is better to use many features. However, we note that resistance to overfitting is different from signal being drowned in noise. With enough noisy features, correlations between class and a useless feature will happen purely by chance, preventing true relationships from being found.

To visualize this effect for our data, we took our nominal feature set and continually added features with no predictive power (random samples from the uniform distribution) to quantify the degradation in performance of the resultant classifiers. The random features were re-generated for each of the 100 trials, and the cross-validated results are shown in Figure 6. The fact that even a small number of useless features causes a noticeable decrease in performance highlights the importance of attribute selection. However, we note that too much fine tuning of attribute feature selection choices—such as testing all combinations of features and seeing which one gives the best performance—would overfit to the data and give an underestimate of the true error.

4.3. Final Classifier

Taking into account the above issues of multiple feature set choices, the deleterious effect of useless features, and the performance with various weight choices to help with imbalance, we have developed a classifier which we believe to be robust and powerful. The full feature set utilized is shown
in Table 1, and the weight chosen is described in Section 4.1. The final cross-validated estimates of $Q$ for the training data are shown alongside the corresponding redshifts in Table 2. By referencing a particular point on the $x$-axis of Figure 7 (left panel), one can determine what fraction of high bursts can be detected for a particular amount of telescope follow-up time. For example, if we are able to follow up on 20% of all GRBs detected by Swift, then the bursts recommended for follow up by our classifier will contain on average 56% ± 6% of all GRBs with redshift greater than 4 that occur. Following up on ~40% of all bursts will yield 84% ± 6% of all GRBs with redshift greater than 4, and following up on the top 50% of candidates will result in nearly all of the high-$z$ events being observed (96% ± 4%).

Purity is shown in the right panel of Figure 7, which describes how many of the followed-up bursts will actually be high redshift. Following up on 20% of all bursts would result in 37% ± 4% of the followed-up events being high redshift, and 28% ± 2% of followed-up bursts would be high redshift if 40% of GRBs were followed up on.

As the high/low class division of $z = 4$ was relatively arbitrary, for completeness we also re-trained the classifier and calculated performance results using cutoff values of $z = 3.5$ (Figure 8) and $z = 3$ (Figure 9). Note that while the sample size of “high” events more than doubles by lowering the cutoff value to $z = 3$, the resultant efficiency decreases significantly. We attribute this effect to a decrease in the predictive power of certain attributes at lower redshift. For instance, the $z > 3$ population has proportionally many more instances of UVOT detections in its “high-$z$” class than the $z > 4$ population, which reduces its effectiveness as a discriminating feature.

4.4. Feature Importance

There are several complications in identifying the relative importance of features in contributing to selecting high-$z$ candi-
from the mean value across all seeds.

However, the removal of a few of the individual features does cause a degradation in performance larger than what would be expected by random, indicating that these features are both useful and not completely redundant. Most of the features do not cause a significant change in performance once removed from the data set. However, the removal of a few of the individual features does cause a degradation in performance larger than what would be expected by random, implying that these features are both important and not completely redundant. Note that the relative change in both purity and efficiency is equal in both plots, as only the numerator of each metric is changing ($N_{\text{high observed}}$), but we show both values for consistency.

5. DISCUSSION

5.1. Calibration on GRBs with Unknown Redshifts

A natural application of our methodology is to use it to predict the follow-up metric $Q$ for the remaining majority of long-duration *Swift* GRBs with no known redshift, providing a
list of the top candidates predicted to be high-$z$. This application is precisely how RATE GRB-$z$ could be used in practice on new events, albeit one at a time rather than on many at once. We caution that due to the natural selection effect of GRBs with measured redshifts having a higher likelihood of being brighter events, the bursts with unknown redshifts are likely to comprise a somewhat different redshift distribution than our training data set. The primary consequence of this is the interpretation of the user-desired follow-up fraction $Q$ and the prioritization parameter $\hat{Q}$. In principle, the classifier was calibrated such that, over time, a fraction $Q$ of new events will have affirmative follow-up recommendations (that is, events such that $\hat{Q} \leq Q$). However, this will not necessarily be the case if the full redshift distribution of GRBs makes up a different population than our training data.

To test this, we calculated $\hat{Q}$ for each of the remaining 212 GRBs with unknown redshift that met our culling criteria outlined in Section 2. From this, we could calculate the fraction of GRBs followed up ($\hat{Q} \leq Q$) for each cutoff value of $Q$. The results of this test are shown in Figure 11. For the chosen weight of 10 (see Section 4.1), the $Q$-values are well calibrated with the final follow-up recommendations. The resultant $\hat{Q}$ priorities are listed in Table 4. These values can be interpreted as a ranking of which of these past events without secure redshift determinations are most likely to be at high redshift.

5.2. Validation Set: Application to Recent GRBs

Since the cutoff date in our training set (2010 June 21) until 2011 September 1, there have been 15 long-duration Swift GRBs with reliable redshifts from which we constructed an independent validation set to test our method. The feature values for these GRBs are presented in Table 5. While none of these events was over our high-redshift cutoff value of $z = 4$, it is still possible, though challenging, to use low-$z$ events (either by direct redshift measurement or by the identification of a coincident blue host galaxy) as a consistency test. We would expect that the purity at a given $Q$ would be lower than the fraction of recommended follow-up events ($Q \leq \hat{Q}$) without a secure low-$z$ determination. For instance, $Q = 0.2$ has a purity of $37\% \pm 4\%$, so no more than $\sim 63\%$ of events with $Q < 0.2$ should be definitely low redshift.

The validation GRBs were run through the RATE GRB-$z$ classifier, and their resultant $\hat{Q}$-values are shown in Table 6 along with their corresponding redshifts. The smallest $Q$-value of these events is $\sim 0.3$, meaning that none of these

---

**Table 4**

| GRB    | $\hat{Q}$ | $\alpha$ | $E_{\text{peak}}$ (keV) | $S$ (erg cm$^{-2}$) | $S/N_{\text{max}}$ | $N_{\text{H,pc}}$ (10$^{22}$ cm$^{-2}$) | $T_{90}$ (s) | $\sigma_{\text{BAT}}$ (counts s$^{-1}$) | $R_{\text{peak,BAT}}$ (counts s$^{-1}$) | Rate trigger | $t_{\text{BAT}}$ (s) | UVOT detect | $P_{>4}$ |
|--------|-----------|----------|--------------------------|--------------------|-------------------|--------------------------------|------------|--------------------------------|-------------------------------|-------------|-------------|-------------|---------|
| 050215A | 3.19e-01  | -1.29e+00| 4.14e+02                 | 1.34e-06           | 1.02e+01          | 5.E+1                        | 6.65e+01  | 9.00e+00                        | 6.94e+02                     | Yes          | 5.12e-01   | No          | 5.67e-02 |
| 050215B | 1.78e-01  | ?        | 3.01e+02                 | 2.86e-07           | 1.44e+01          | 5.70e-02                      | 8.50e+01  | 8.00e+00                        | 3.80e+02                     | Yes          | 2.05e+00   | No          | 1.06e-01 |
| 050219A | 4.22e-01  | 1.87e-02 | 1.00e+02                 | 4.91e-06           | 5.08e+01          | 9.10e-02                      | 2.58e+01  | 8.00e+00                        | 1.93e+02                     | Yes          | 1.02e+00   | No          | 1.12e-01 |
| 050219B | 7.33e-01  | -8.94e-01| 1.12e+02                 | 1.94e-05           | 7.19e+01          | 8.80e-02                      | 2.09e+01  | 1.70e+01                        | 4.09e+02                     | Yes          | 1.02e+00   | No          | 2.73e-02 |
| 050326  | 7.04e-01  | -1.04e+00| 3.41e+02                 | 1.70e-05           | 1.33e+02          | 3.80e-02                      | 3.02e+01  | 2.10e+01                        | 1.84e+04                     | Yes          | 5.12e-01   | No          | 5.67e-02 |

(This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content.)

---

18 One of the bursts with a measured redshift, GRB 110328A, had very unusual properties and was determined to be a potential Tidal Disruption Event (Bloom et al. 2011; Levan et al. 2011a), and was thus also excluded from the validation set.
Table 5
Validation Data

| GRB   | $\hat{Q}$  | $\alpha$ | $E_{\text{peak}}$ (keV) | $S$ (erg cm$^{-2}$) | $S/N_{\text{max}}$ | $N_{\text{hit,PC}}$ (10$^{22}$ cm$^{-2}$) | $T_{90}$ (s) | $\sigma_{\text{BAT}}$ | $R_{\text{peak,BAT}}$ (counts s$^{-1}$) | Rate trigger | $t_{\text{BAT}}$ (s) | UVOT detect | $P_{>4}$ |
|-------|------------|----------|--------------------------|---------------------|---------------------|---------------------------------|-------------|-------------------|---------------------------------|--------------|-----------------|-------------|--------|
| 100728B | 6.07e-01   | -1.64e+00 | 8.19e+01                 | 2.54e-06            | 2.06e+01            | 3.90e-02                        | 1.15e+01   | 9.07e+00        | 1.47e+02                        | Yes          | 1.02e+00      | Yes         | 1.01e-01|
| 100814A | 6.81e-01   | -1.11e+00 | 1.35e+02                 | 9.33e-06            | 9.80e+01            | ?                                | 1.77e+02   | 1.91e+01        | 8.34e+02                        | Yes          | 1.02e+00      | Yes         | 1.80e-01|
| 100816A | 9.33e-01   | -5.71e-01 | 1.42e+02                 | 2.71e-06            | 5.80e+01            | 1.13e-01                        | 2.50e+00   | 2.29e+01        | 1.42e+03                        | Yes          | 1.02e+00      | Yes         | 5.55e-02|
| 100901A | 4.00e-01   | -1.55e+00 | 1.28e+02                 | 3.41e-06            | 1.78e+01            | 4.00e-02                        | 4.59e+02   | 7.70e+00        | 4.50e+02                        | Yes          | 8.19e+00      | Yes         | 2.25e-01|
| 100906A | 1.00e+00   | -1.66e+00 | 1.57e+02                 | 1.37e-05            | 1.36e+02            | ?                                | 1.17e+02   | 1.05e+01        | 1.91e+02                        | Yes          | 5.12e-01      | Yes         | 7.39e-02|
| 101219B | 6.30e-01   | -1.89e+00 | 4.97e+01                 | 3.75e-06            | 1.00e+01            | -8.00e-03                       | 4.18e+01   | 7.63e+00        | 8.44e+02                        | No           | 6.40e+01      | Yes         | 1.07e-01|
| 110205A | 3.19e-01   | -1.39e+00 | 9.75e+01                 | 1.98e-05            | 1.50e+02            | 1.10e-02                        | 2.77e+02   | 1.00e+01        | 1.48e+03                        | No           | 6.40e+01      | Yes         | 1.45e-01|
| 110213A | 9.33e-01   | -1.82e+00 | 6.70e+01                 | 8.77e-06            | 3.10e+01            | 4.00e-02                        | 4.31e+01   | 1.21e+01        | 2.05e+02                        | Yes          | 1.02e+00      | Yes         | 5.32e-02|
| 110422A | 1.00e+00   | -6.23e-01 | 1.11e+02                 | 5.17e-05            | 2.10e+02            | 1.58e-01                        | 2.67e+01   | 7.19e+00        | 8.20e+01                        | Yes          | 1.28e-01      | Yes         | 2.49e-02|
| 110503A | 9.33e-01   | -8.18e-01 | 1.42e+02                 | 1.43e-05            | 6.27e+01            | 2.60e-02                        | 9.31e+00   | 2.04e+01        | 1.26e+03                        | Yes          | 1.02e+00      | Yes         | 1.89e-02|
| 110715A | 9.33e-01   | -1.06e+00 | 8.94e+01                 | 1.40e-05            | 2.02e+02            | 1.64e-01                        | 1.31e+01   | 1.19e+01        | 1.47e+02                        | Yes          | 1.28e-01      | Yes         | 9.70e-03|
| 110726A | 5.04e-01   | -2.97e-01 | 4.27e+01                 | 2.07e-07            | 1.51e+01            | -4.90e-02                       | 5.40e+00   | 8.60e+00        | 2.24e+02                        | Yes          | 1.02e+00      | Yes         | 1.14e-01|
| 110731A | 1.00e+00   | -1.19e+00 | 4.06e+02                 | 1.25e+05            | 1.30e+02            | 7.20e-02                        | 4.66e+01   | 2.46e+01        | 2.32e+03                        | Yes          | 1.02e+00      | Yes         | 5.09e-02|
| 110801A | 9.33e-01   | -1.84e+00 | 6.07e+01                 | 6.85e-06            | 3.56e+01            | 2.90e-02                        | 4.00e+02   | 7.83e+00        | 3.50e+02                        | Yes          | 4.10e+00      | Yes         | 1.98e-01|
| 110808A | 5.56e-01   | ?         | 2.59e+01                 | 4.27e-07            | 1.01e+01            | 2.17e-01                        | 3.94e+01   | 7.19e+00        | 4.26e+02                        | Yes          | 8.19e+00      | Yes         | 1.06e-01|
events would have been recommended for high-\(z\) follow-up for anyone wishing to observe fewer than 30\% of events. While these values are certainly consistent with our expected purity, it is not particularly constraining, as it would have been very unlikely for this almost random selection of GRBs to violate this constraint by chance alone, even if the classifier had no predictable power.

A more constraining test is the identification of high-\(z\) events with high \(\hat{Q}\) for comparison with the expected efficiency. Two events not included in our training set have had recent high-\(z\) identifications: GRB 090429B with strong photometric evidence for being \(z \approx 9.4\) (Cucchiara et al. 2011b), and the spectroscopic identification of GRB 111008A at \(z = 4.99\) (Leytan et al. 2011b; Wiersema et al. 2011). The former has a \(\hat{Q}\)-value of \(\sim 0.185\), consistent with the expected efficiency. However, GRB 111008A has a \(\hat{Q}\) of \(\sim 0.638\), a value above which we would have expected to find no more than 1\% of high-\(z\) events. This outlier seems likely due to the extreme brightness of the event (among the brightest \(\sim 10\%\) of \textit{Swift} bursts in the observer frame, and top \(\sim 3\%\) in the rest frame). Indeed, compared to all 18 high-\(z\) events in the training set, GRB 111008A has the most extreme values toward the “wrong” end of three of the highly important features identified in Section 4.4 (\(\alpha, P_{3-4},\) and \(R_{\text{peak BAT}}\)) and also has the fourth largest \(S/N_{\text{max}}\). In later iterations of \textit{RATE-GRB-\(z\)}, this event (and all new GRBs with secure redshifts) will be added to the training data to re-generate the classifier and further improve its robustness against such outliers.

### 5.3. Comparison to Previous Efforts

Extracting indications of redshift from promptly available information has been a continuing goal of GRB studies since their cosmological origins were discovered nearly 15 years ago. Several potential luminosity indicators were pursued with the optimistic goal of using GRBs as standard candles for cosmological studies. The efficacy of individual indicators toward this goal proved to be limited, and a physical origin of the relations has been contested, with authors attributing them instead to detector thresholding or other selection effects (Butler et al. 2007, 2009, 2010; Shahmoradi & Nemiroff 2011). While these studies have ruled out the majority of such relations as intrinsic to GRBs themselves, prompt properties can still be used as redshift indicators if the systematics are properly accounted for.

Several recent studies have attempted to use combinations of features to determine “pseudo-redshifts” for GRBs. In an extension of work by Schaefer (2007), Xiao & Schaefer (2009, 2011) used a combination of six purported luminosity relations. Further, Koen (2009, 2010) has explored linear regression as a tool for predicting GRB redshifts using the data set from Schaefer (2007). As data derived from multiple satellites were used, these studies are particularly vulnerable to the detector selection effects mentioned above.

Some works avoided the complications of regression and instead focused upon the simple selection of high-\(z\) candidates for follow-up purposes. Campana et al. (2007) utilized a sample of \textit{Swift}-only bursts (thus avoiding detector effect biases) and used hard cuts on three features (\(T_{90},\) lack of UVOT detection, and high galactic latitude) for high-\(z\) candidate selection. Salvaterra et al. (2007) extended upon this work with the additional feature of peak photon flux. Several issues prevent a direct comparison among the various methods of the effectiveness at separating high-\(z\) events. These include the usage of different features from each study, which is complicated by the lack of uniformity of features being created for each. Further, the techniques above strictly constrain the manner in which each feature influences the output, whereas our method is fully non-parametric and therefore more flexible. However, the largest concern is accurate reporting of predictive performance. In particular, we caution against the circular practice of measuring the performance of methods by applying them to the same events from which the luminosity relations were formed. In order to prevent overestimating the accuracy of a predictive model, one needs to test on data independent from the training set, such as with cross-validation.

Finally, the \textit{RATE} method differs from previous efforts in that it casts the problem as one of optimal resource allocation under limited follow-up time. Prior techniques are not explicitly calibrated to suit this purpose. Direct classification methods will either under- or overutilize available resources. Past regression or “pseudo-\(z\)” methods are not explicitly calibrated to a particular follow-up decision (i.e., at what “pseudo-\(z\)” does one decide to follow up?), though it would be possible in principle to correct for this using a transformation which ensures that the desired follow-up fraction corresponds to the actual fraction of bursts followed up (e.g., Figure 11). In contrast, the \textit{RATE} technique is by design applicable to any available resource reserves, and is generally extendable to any transient follow-up prioritization problem.

### Table 6

| GRB       | \(\hat{Q}\)  | \(z\)   | References                      |
|-----------|-------------|--------|--------------------------------|
| 100728B   | 5.63e-01    | 2.106  | Flores et al. (2010)            |
| 100814A   | 7.04e-01    | 1.44   | O’Meara et al. (2010)           |
| 100816A   | 9.33e-01    | 0.8035 | Tanvir et al. (2010a, 2010c)    |
| 100901A   | 4.22e-01    | 1.408  | Chornock et al. (2010a)         |
| 100906A   | 9.19e-01    | 1.727  | Tanvir et al. (2010b)           |
| 101219B   | 6.07e-01    | 0.5519 | de Ugarte Postigo et al. (2011c)|
| 110205A   | 3.19e-01    | 2.22   | Cenko et al. (2011)             |
| 110213A   | 8.89e-01    | 1.46   | Milne & Cenko (2011)            |
| 110422A   | 9.33e-01    | 1.770  | Malesani et al. (2011); de Ugarte Postigo et al. (2011a) |
| 110503A   | 9.33e-01    | 1.613  | de Ugarte Postigo et al. (2011b)|
| 110715A   | 8.89e-01    | 0.82   | Piranomonte et al. (2011)       |
| 110726A   | 5.56e-01    | 1.036  | Cucchiara et al. (2011a)        |
| 110731A   | 9.33e-01    | 2.83   | Tanvir et al. (2011)            |
| 110801A   | 8.67e-01    | 1.858  | Cabrera Lavers et al. (2011)    |
| 110808A   | 5.63e-01    | 1.348  | de Ugarte Postigo et al. (2011d)|
6. CONCLUSIONS

In this paper, we presented the RATE GRB-ζ method for allocating follow-up telescope resources to high-redshift GRB candidates using RF classification on early-time Swift metrics. The RATE method is generalizable to any prioritization problem that can be parameterized as “observe” or “do not observe,” and accommodates statistical challenges such as small data sets, imbalanced classes, and missing feature values. The issue of resource allocation is becoming increasingly important in the era of data-driven transient surveys such as Palomar Transient Factory, Pan-STARRS, and Large Synoptic Survey Telescope which provide extremely high discovery rates without a significant increase in follow-up resources. With enough training instances of any object of interest for a given transient survey, the RATE method can be applied to prioritize follow-up of future high-priority candidates.

In the RATE GRB-ζ application, our robust, cross-validated performance metrics indicate that by observing just 20% of bursts, one can capture 56% ± 6% of z > 4 events with a sample purity of 37% ± 4%. Further, following up on half of all events will yield nearly all (96% ± 4%) of the high-z events. The method provides a simple decision point for each new event: if the prioritization value Q is smaller than the percentage of events a user wishes to allocate resources to, then follow-up is recommended. These rapid predictions, combined with the more traditional photometric dropout technique from simultaneous multi-filter NIR observatories (such as PAIRITEL, GROND, and the upcoming RATIR), offer a robust tool in more efficiently informing GRB follow-up decisions. To facilitate the dissemination of high-redshift GRB predictions to the community, we have set up a Web site (http://rate.grbz.info) with Q-values for past bursts, and an RSS feed (http://rate.grbz.info/rss.xml) to provide real-time results from our classifier on new events.

This work was sponsored by an NSF-CDI grant (award no. 0941742) “Real-time Classification of Massive Time-series Data Streams” (PI: J. S. Bloom). This publication has made use of data obtained from the Swift interface of the High-Energy Astrophysics Archive (HEASARC), provided by NASA’s Goddard Space Flight Center. A. Morgan gratefully acknowledges support from an NSF Graduate Research Fellowship. T. Broderick was funded by a National Science Foundation Graduate Research Fellowship. We are extremely grateful to the Swift team for the rapid public dissemination, calibration, and analysis of the Swift data. We thank Daniel Perley for useful conversations and for the creation and maintenance of the very useful http://grbox.net Web site. We also thank Scott Barathelmy for his invaluable efforts in creating and maintaining the GCN system.

Facilities: Swift (BAT, XRT, UVOT)

REFERENCES

Amati, L., Frontera, F., Tavani, M., et al. 2002, A&A, 390, 81
Antonelli, L. A., Maund, J. R., Palazzi, E., et al. 2010, GRB Coordinates Network, 10620, 1
Bailey, S., Aragon, C., Romano, R., et al. 2007, ApJ, 665, 1246
Band, D. L., Norris, J. P., & Bonnell, J. T. 2004, ApJ, 613, 484
Barkana, R., & Loeb. A. 2004, ApJ, 601, 64
Barthelmy, S. D., Barbier, L. M., Cummings, J. R., et al. 2005, Space Sci. Rev., 120, 143
Berger, E. 2006, GRB Coordinates Network, 5962, 1
Berger, E., Cenko, S. B., Steidel, C., Reddy, N., & Fox, D. B. 2005, GRB Coordinates Network, 3368, 1
Berger, E., Foley, R., Simcoe, R., & Irwin, J. 2008a, GRB Coordinates Network, 8434, 1
Berger, E., Fox, D. B., & Cucchiara, A. 2007, GRB Coordinates Network, 6470, 1
Berger, E., Fox, D. B., Cucchiara, A., & Cenko, S. B. 2008b, GRB Coordinates Network, 8355, 1
Berger, E., & Gladders, M. 2006, GRB Coordinates Network, 5170, 1
Berger, E., Kulkarni, S. R., Rau, A., & Fox, D. B. 2006, GRB Coordinates Network, 4815, 1
Berger, E., & Mulchaey, J. 2005, GRB Coordinates Network, 3122, 1
Berger, E., & Rauch, M. 2008, GRB Coordinates Network, 8542, 1
Berger, E., & Shin, M.-S. 2006, GRB Coordinates Network, 5283, 1
Bloom, J. S., Giannios, D., Metzger, B. D., et al., 2011, Science, 333, 203
Bloom, J. S., Perley, D. A., & Chen, H. W. 2006, GRB Coordinates Network, 5826, 1
Bouwens, R. J., Illingworth, G. D., Labbe, I., et al. 2011, Nature, 469, 504
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2010, ApJ, 709, L133
Breiman, L. 2001, Mach. Learn., 45, 5
Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. 1984, Classification and Regression Trees (Monterey, CA: Wadsworth)
Burrows, D. N., Hill, J. E., Nousek, J. A., et al. 2005, Space Sci. Rev., 120, 165
Butler, N. R., Bloom, J. S., & Poznanski, D. 2010, ApJ, 711, 495
Butler, N. R., & Kocevski, D. 2007, ApJ, 663, 407
Butler, N. R., Kocevski, D., & Bloom, J. S. 2009, ApJ, 694, 76
Butler, N. R., Kocevski, D., Bloom, J. S., & Curtis, J. L. 2007, ApJ, 671, 656
Cabrera Lavers, A., de Ugarte Postigo, A., Castro-Tirado, A. J., et al. 2011, GRB Coordinates Network, 12234, 1
Campana, S., Tagliaferri, G., Malesani, D., et al. 2007, A&A, 464, L25
Cargile, S., Badavari, T., Heinis, S., Priebe, C., & Stalay, A. S. 2010, ApJ, 712, 511
Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. 2006, Measurement Error in Nonlinear Models: A Modern Perspective (2nd ed.; Boca Raton, FL: Chapman & Hall/CRC)
Cenko, S. B., Berger, E., Djorgovski, S. G., Mahabal, A. A., & Fox, D. B. 2006, GRB Coordinates Network, 5155, 1
Cenko, S. B., Bloom, J. S., Perley, D. A., et al. 2010a, GRB Coordinates Network, 10389, 1
Cenko, S. B., Gezari, S., Small, T., Fox, D. B., & Chornock, R. 2007, GRB Coordinates Network, 6322, 1
Cenko, S. B., Hora, J. L., & Bloom, J. S. 2011, GRB Coordinates Network, 11638, 1
Cenko, S. B., Kulkarni, S. R., Gal-Yam, A., & Berger, E. 2005, GRB Coordinates Network, 3542, 1
Cenko, S. B., Perley, D. A., Junkkarinen, V., et al. 2009, GRB Coordinates Network, 9518, 1
Cenko, S. B., Perley, D. A., Morgan, A. N., et al. 2010b, GRB Coordinates Network, 10752, 1
Chawla, N. V., Japkowicz, N., & Kotcz, A. 2004, SIGKDD Explor. Newsl., 6, 1
Chen, C., Liaw, A., & Breiman, L. 2004, Technical Report
Chen, H.-W., Helsby, J., Shectman, S., Thompson, I., & Crane, J. 2009, GRB Coordinates Network, 10038, 1
Chen, H.-W., Prochaska, J. X., Bloom, J. S., & Thompson, I. B. 2005, ApJ, 634, L25
Chornock, R., Berger, E., Fox, D., et al. 2010a, GRB Coordinates Network, 11164, 1
Chornock, R., Berger, E., Levesque, E. M., et al. 2010b, arXiv:1004.2262
Chornock, R., Cenko, S. B., Griffith, C. V., et al. 2009a, GRB Coordinates Network, 9151, 1
Chornock, R., Cucchiara, A., Fox, D., & Berger, E. 2010c, GRB Coordinates Network, 10466, 1
Chornock, R., Perley, D. A., Cenko, S. B., & Bloom, J. S. 2009b, GRB Coordinates Network, 9243, 1
Chornock, R., Perley, D. A., & Cobb, B. E. 2009c, GRB Coordinates Network, 10100, 1
Cucchiara, A., Bloom, J. S., & Cenko, S. B. 2011a, GRB Coordinates Network, 12202, 1
Cucchiara, A., Fox, D., Levan, A., & Tanvir, N. 2009, GRB Coordinates Network, 10202, 1
Cucchiara, A., Fox, D. B. 2008b, GRB Coordinates Network, 7654, 1
Cucchiara, A., & Fox, D. B. 2010, GRB Coordinates Network, 10624, 1
Cucchiara, A., Fox, D. B., & Berger, E. 2006a, GRB Coordinates Network, 4729, 1
Cucchiara, A., Fox, D. B., & Cenko, S. B. 2007a, GRB Coordinates Network, 7124, 1
Cucchiara, A., Fox, D. B., Cenko, S. B., & Berger, E. 2008a, GRB Coordinates Network, 8346, 1
Cucchiara, A., Fox, D. B., Cenko, S. B., & Berger, E. 2008b, GRB Coordinates Network, 8448, 1
