

SUPERNOVA PHOTOMETRIC CLASSIFICATION CHALLENGE

RICHARD KESSLER,^{1,2} ALEX CONLEY,³ SAURABH JHA,⁴ STEPHEN KUHLMANN⁵*Challenge Released on Jan 29, 2010. Last update: September 24, 2013*

ABSTRACT

We have publicly released a blinded mix of simulated SNe, with types (Ia, Ib, Ic, II) selected in proportion to their expected rate. The simulation is realized in the *griz* filters of the Dark Energy Survey (DES) with realistic observing conditions (sky noise, point spread function and atmospheric transparency) based on years of recorded conditions at the DES site. Simulations of non-Ia type SNe are based on spectroscopically confirmed light curves that include *unpublished* non-Ia samples donated from the Carnegie Supernova Project (CSP), the Supernova Legacy Survey (SNLS), and the Sloan Digital Sky Survey-II (SDSS-II). We challenge scientists to run their classification algorithms and report a type for each SN. A spectroscopically confirmed subset is provided for training. The goals of this challenge are to (1) learn the relative strengths and weaknesses of the different classification algorithms, (2) use the results to improve classification algorithms, and (3) understand what spectroscopically confirmed sub-sets are needed to properly train these algorithms. The challenge is available at www.hep.anl.gov/SNchallenge, and the due date for classifications is May 1, 2010.

Subject headings: supernova light curve fitting and classification

1. MOTIVATION

To explore the expansion history of the universe, increasingly large samples of high quality SNe Ia light curves are being used to measure luminosity distances as a function of redshift. With increasing sample sizes, there are not nearly enough resources to spectroscopically confirm each SN. Currently, the world's largest samples are from the Supernova Legacy Survey (SNLS: Astier et al. (2006)) and the Sloan Digital Sky Survey-II (SDSS-II: Frieman et al. (2008)), each with more than 1000 SNe Ia, yet less than half of their SNe are spectroscopically confirmed. The numbers of SNe are expected to increase dramatically in the coming decade: thousands for the Dark Energy Survey (DES: Bernstein et al. (2009)) and a few hundred thousand for the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS)⁶ and the Large Synoptic Survey Telescope (LSST: Ivezić et al. (2008); LSST Science Book (2009)). Since only a small fraction of these SNe will be spectroscopically confirmed, photometric identification is crucial to fully exploit these large samples.

In the discovery phase of accelerated cosmological expansion, results were based on tens of high-redshift SNe Ia, and some samples included a significant fraction of events that were not classified from a spectrum (Riess et al. 1998; Perlmutter et al. 1999; Tonry et al. 2003; Riess et al. 2004). While human judgment played a significant role in classifying these SNe without a spectrum, more formal methods of

photometric classification have been developed over the past decade: Poznanski et al. (2002); Dahlen & Goobar (2002); Sullivan et al. (2006); Johnson & Crots (2006); Poznanski et al. (2007); Kuznetsova & Connolly (2007); Rodney & Tonry (2009). Some of these methods have been used to select candidates for spectroscopic follow-up observations, but these methods have not been used to select a significant photometric SN Ia sample for a Hubble diagram analysis. In short, cosmological parameter estimates from much larger and recent surveys are based solely on spectroscopically confirmed SNe Ia (SNLS: Astier et al. (2006), ESSENCE: Wood-Vasey et al. (2007), CSP: Freedman et al. (2009), SDSS-II: Kessler et al. (2009)).

The main reason for the current reliance on spectroscopic identification is that vastly increased spectroscopic resources have been used in these more recent surveys. In spite of these increased resources, more than half of the discovered SNe do not have a spectrum and therefore photometric methods will eventually be needed to classify the majority of the SNe. There are two difficulties limiting the application of photometric classification. First is the lack of adequate non-Ia data for training algorithms. Many classification algorithms were developed using non-Ia templates⁷ constructed from averaging and interpolating a limited amount of spectroscopically confirmed non-Ia data, and therefore the impact of the non-Ia diversity has not been well studied. The second difficulty is that there is no standard testing procedure, and therefore it is not clear which classification methods work best.

To aid in the transition to using photometric SN classification, we have released a public “SN Photometric Classification Challenge” to the community, hereafter called SNPhotCC. The SNPhotCC consists of a blinded mix of simulated SNe, with types (Ia, Ib, Ic, II) selected in proportion to their expected rate. The challenge is for scientists to run their classification algorithms and re-

¹ Department of Astronomy and Astrophysics, The University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637

² Kavli Institute for Cosmological Physics, The University of Chicago, 5640 South Ellis Avenue Chicago, IL 60637

³ Center for Astrophysics and Space Astronomy, University of Colorado, Boulder, CO, 80309-0389, USA

⁴ Department of Physics and Astronomy, Rutgers University, 136 Frelinghuysen Road, Piscataway, NJ 08854

⁵ Argonne National Laboratory, 9700 S. Cass Avenue, Lemont, IL 60437

⁶ <http://pan-starrs.ifa.hawaii.edu/public>

⁷ http://supernova.lbl.gov/nugent/nugent_templates.html

port a type for each SN. A spectroscopically confirmed sub-set is provided so that algorithms can be tuned with a realistic training set. The goals of this challenge are to (1) learn the relative strengths and weaknesses of the different classification algorithms, (2) use the SNPhotCC results to improve the algorithms, and (3) understand what spectroscopically confirmed sub-sets are needed to properly train these algorithms.

To address the paucity of non-Ia data, the CSP, SNLS and SDSS-II have contributed *unpublished* spectroscopically confirmed non-Ia light curves. These data are high-quality multi-band light curves, not just junk that nobody cares about, and therefore we are grateful to the donating collaborations. This non-Ia sample is likely to undersample the potential variety in the upcoming surveys like DES and LSST, but we anticipate that this challenge will be a useful step away from the overly-simplistic studies that have relied on a handful of non-Ia templates.

The outline of this release-note is as follows. A description of the simulation is given in §2, and instructions for participants are in §3. Comments on the evaluations and posting of results are given in §4.

2. THE SIMULATION

The simulation is realized in the *griz* filters of the Dark Energy Survey (DES). The sky-noise, point-spread function and atmospheric transparency are evaluated in each filter based on years of observational data from the ESSENCE project at the Cerro Tololo Inter-American Observatory (CTIO). For the five SN fields (3 sq deg each), the cadence is based on allocating 10% of the DES photometric observing time and most of the non-photometric time. The cadence used in this publicly available simulation was generated by the Supernova Working Group within the DES collaboration.⁸ Since the DES plans to collect data during 5 months of the year, incomplete light curves from temporal edge effects are included; i.e., the simulated explosion times extend well before the start of each survey season, and extend well beyond the end of the season.

Simulated SNe Ia are based on models empirically derived from data. In addition to the model parameters, we have applied tweaks to simulate the anomalous Hubble scatter. While these tweaks are invented ad-hoc, they have not been ruled out with current observations. Simulated non-Ia SNe are based on observed multi-color light curves (from CSP, SNLS, and SDSS) that have been smoothed in each passband, and then K-corrected to the appropriate redshift and filters.

A spectroscopically confirmed subset is based on observations on a 4 meter class telescope with a limiting *r*-band magnitude of 21.5, and an 8 meter class telescope with a limiting *i*-band magnitude of 23.5. The subset is randomly selected, and the number of spectroscopically confirmed SNe (~ 1000) corresponds to the combined resources of the SNLS & SDSS-II surveys. While this number of spectroscopic identifications may be optimistic, this allows for further study on how the training quality depends on the size of the spectroscopic sample.

⁸ Although two of us (RK & SK) are members of the DES, we have not included other DES colleagues in any discussions about this challenge, and we have made our best efforts to prevent our DES collaborators from obtaining additional information beyond that contained in this note.

For the challenge that includes the host-galaxy photometric redshift, the photo-*z* estimates are based on simulated galaxies (for DES) analyzed with the methods in Oyaizu et al. (2008a,b). The average host-galaxy photo-*z* resolution is 0.03.

Two simple selection criteria have been applied. First, each object must have at least one observation with a signal to noise ratio (S/N) above 5 (in any filter). Second, there must be at least 5 observations after explosion, and there is no S/N requirement on these observations. These requirements are relatively loose because part of the challenge is to determine the optimal selection criteria. The total number of simulated SNe that satisfy these loose selection requirements is 2×10^4 , and corresponds to the 5 seasons planned for the DES.

3. TAKING THE CHALLENGE

Two independent challenges have been generated: one with a host-galaxy photo-*z*, and another without any redshift information. In addition to these challenges based on the entire light curve, there is also an early-epoch challenge based on the first six observations (in any filter) with $S/N > 4$. On the night of the sixth observation, all observations made this night are included. Among the four challenges available, you may take any of them or all of them.

The simulated light curves can be downloaded from the SNPhotCC website.⁹ The filter response functions are given in the files `DES_[griz].dat`. The file with the “.LIST” suffix provides a list of all data files to analyze. The data files are self-documented and visual inspection should be adequate for preparing a parsing algorithm. The calibrated fluxes are defined as

$$\text{FLUXCAL} = 10^{(-0.4 \cdot m + 11)} + \text{noise} \quad (1)$$

where *m* is the modeled AB-magnitude of the SN, and the noise contributions¹⁰ include Poisson fluctuations, sky noise, and CCD noise. The observed magnitudes are not provided because they are not defined when noise fluctuations result in a negative flux; for fitting, we recommend translating model-magnitudes into fluxes as defined in Eq. 1.

For tuning your algorithms, the spectroscopically confirmed sub-sample is identified by the `SNTYPE` keyword (see Table 1), and the corresponding redshift is given by the `REDSHIFT_SPEC` keyword. For the majority of SNe that do not have spectroscopic identification, the type and spectroscopic redshift are set to -9 . For the host-galaxy photo-*z* sample, the photo-*z* is given by the `HOST_GALAXY_PHOTO-Z` keyword. For the early-epoch challenge, process only the observations that appear before the “DETECTION:” keyword.

A valid challenge submission must contain three items: (1) an answer list containing the type for each SN, (2) a brief description of your method, and (3) an estimate of the CPU resources. For a group effort, a team name is recommended. These submission items are discussed below in more detail.

For each challenge that you participate in, your answer list must contain four columns:

⁹ www.hep.anl.gov/SNchallenge

¹⁰ The noise has been scaled from photoelectrons into FLUXCAL units.

TABLE 1
INTEGER CODES FOR SN TYPES.

SN-type	integer code
Ia	1
II (IIn, IIP, IIL)	2 (21, 22, 23)
Ibc (Ib, Ic)	3 (32, 33)
other	66
rejected	-1

SNID TYPE PHOTOZ PHOTOZ_ERROR

where

- SNID is the SN integer id
- TYPE is the integer SN-type code returned by your classifier (see Table 1). You can report either a general type (1,2,3 for Ia,II,Ibc), or a specific sub-type.
- PHOTOZ is photo- z value returned by your classifier.
- PHOTOZ_ERROR is the uncertainty

If your code does not return a useful photo- z value, just set -9 in the last two columns. A valid answer list must contain entries in all four columns and for each SN; invalid answer files will be returned. In addition to the answer file, please provide a brief description of your technique. A reference to either a refereed journal article or arXiv posting is adequate, but please describe any modifications from the referenced article. Finally, include the processing time, the number of light curves analyzed (i.e., that are not rejected by selection cuts) and a description of your computing processor hardware.

In addition to thinking about your classification algorithm, you should also think about appropriate selection cuts to reject SNe that are difficult to classify. Set the SN type to -1 for rejected SNe. As described in §4, our evaluation generally penalizes incorrect classifications more than it penalizes the loss from selection cuts.

To maximize the utility of this challenge, please respect the following guidelines. While you can use the spectroscopically confirmed subset to train your algorithms, please use your program to report classifications from this subset; i.e., do not just report the spectroscopic SN type. A useful diagnostic in the evaluation will be to compare the classification performance from the training subset to that from the rest of the sample. In a similar spirit, do *not* use the spectroscopic redshift (REDSHIFT_SPEC) to report classifications. Finally, for the early-epoch challenge, use only the spectroscopically confirmed sub-sample for tuning your algorithms; i.e., do *not* use the full set of (unconfirmed) light curves.

Don't hesitate to report problems or suggestions, including methods for evaluation. Missing information and updates will be appended to §5 and re-posted to the arXiv. You should periodically check this arXiv posting for updates.

Finally, the due date is **May 1, 2010**.

4. POSTING & EVALUATING THE CHALLENGE RESULTS

Classification results from the participants will be posted publicly along with our initial evaluations and the answer key. Anyone can therefore evaluate the algorithms using their choice of figure-of-merit (FoM). We will also provide additional information about the simulation strategy, along with details for each simulated SN. For non-Ia type SNe based on K-correcting unpublished light curves, the level of detail that we release will be determined solely by the donating collaborations. Shortly before posting the answer key, we will ask the donating collaborations for instructions on what details can be released.

We finish with a discussion of ideas on how to evaluate the results. Ideally, we would like to assign a single number (FoM) for each algorithm. To make more refined comparisons, the FoM can be tabulated as a function of redshift or any other variable of interest.

We begin the discussion by considering the FoM for a Ia rate measurement based on photometric identification. After selection requirements have been applied, let $N_{\text{Ia}}^{\text{true}}$ be the number of correctly typed SNe Ia, and $N_{\text{Ia}}^{\text{false}}$ be the number of non-Ia that are incorrectly typed as an SN Ia. A simple classification FoM is the square of the signal-to-noise ratio (S/N) divided by the total number of SNe Ia ($\mathcal{N}_{\text{Ia}}^{\text{TOT}}$) before selection cuts,

$$\mathcal{C}_{\text{FoM-Ia}} \equiv \frac{1}{\mathcal{N}_{\text{Ia}}^{\text{TOT}}} \times \frac{(N_{\text{Ia}}^{\text{true}})^2}{N_{\text{Ia}}^{\text{true}} + W_{\text{Ia}}^{\text{false}} N_{\text{Ia}}^{\text{false}}} \\ = \epsilon_{\text{Ia}} \times [N_{\text{Ia}}^{\text{true}} / (N_{\text{Ia}}^{\text{true}} + W_{\text{Ia}}^{\text{false}} N_{\text{Ia}}^{\text{false}})] , \quad (2)$$

where $W_{\text{Ia}}^{\text{false}}$ is the false-tag weight (penalty factor) described below, ϵ_{Ia} is the SN Ia efficiency that includes both selection and typing requirements, and $N_{\text{Ia}}^{\text{true}} = \epsilon_{\text{Ia}} \mathcal{N}_{\text{Ia}}^{\text{TOT}}$. Since $\mathcal{N}_{\text{Ia}}^{\text{TOT}}$ is a constant that is independent of the analysis, we have divided out this term so that $0 \leq \mathcal{C}_{\text{FoM-Ia}} \leq 1$, with $\mathcal{C}_{\text{FoM-Ia}} = 1$ corresponding to the theoretically optimal analysis.

The FoM in Eq. 2 is the product of two terms. The first term is the efficiency for selecting and classifying type Ia SNe, and the second term is the Ia purity (when $W_{\text{Ia}}^{\text{false}} = 1$), the fraction of classified Ia that really are SNe Ia. In the ideal case where the average of $N_{\text{Ia}}^{\text{false}}$ is perfectly determined, $W_{\text{Ia}}^{\text{false}} = 1$ and the naive statistical uncertainty is the only contribution to the FoM. In practice, uncertainties in determining the false-tag rate lead to $W_{\text{Ia}}^{\text{false}} > 1$. For example, suppose that $N_{\text{Ia}}^{\text{false}}$ is scaled from a spectroscopically confirmed subset containing a fraction (ϵ_{spec}) of the total number of SNe; in this case, $W_{\text{Ia}}^{\text{false}} = 1 + 1/\epsilon_{\text{spec}}$ is much larger than 1 if the spectroscopic subset is small. It may be possible to reduce $W_{\text{Ia}}^{\text{false}}$ using other methods to determine $N_{\text{Ia}}^{\text{false}}$, such as fitting the tails in the distance-modulus residuals. For SN-cosmology applications, a proper determination of $W_{\text{Ia}}^{\text{false}}$ is beyond the scope of this classification challenge, but suggestions are welcome on setting an appropriate value for the evaluations.

Next we illustrate the FoM with a numerical example in which the false-tag rate is determined from a spectroscopic sub-sample with $\epsilon_{\text{spec}} = 0.2$, and $W_{\text{Ia}}^{\text{false}} = 6$. Consider a sample with 50% type Ia and 50% non-Ia, and $\epsilon_{\text{spec}} = 0.2$. Assume that the classification algo-

rithm correctly identifies half of the SNe, while for the other half the classification works so poorly that it is equivalent to making random guesses with a 50% probability of guessing correctly. If the ambiguous half is rejected, then $\epsilon_{\text{Ia}} = 0.5$, the purity term is 100% (since $N_{\text{Ia}}^{\text{false}} = 0$), and $\mathcal{C}_{\text{FoM-Ia}} = 0.5$. Now consider an analysis strategy without selection requirements. The efficiency term increases to $\epsilon_{\text{Ia}} = 75\%$ since 25% of the SNe Ia are rejected by incorrect classifications. However, since the false-classification rate increases to $N_{\text{Ia}}^{\text{false}}/N_{\text{Ia}}^{\text{true}} = 1/3$ the purity term drops to $1/(1 + 6 \cdot 1/3) = 1/3$ and the net FoM drops to $\mathcal{C}_{\text{FoM-Ia}} = 1/4$. An algorithm that simply makes a random guess on all SNe results in $\mathcal{C}_{\text{FoM-Ia}} = 1/14$. The point of this exercise is to illustrate the importance of selection criteria, and that forcing a classification on every SN candidate is not necessarily the optimal strategy.

5. POST-RELEASE UPDATES

- **February 7, 2010:** for the spectroscopically confirmed subset, sub-types are given as indicated in Table 1. Participants can either report a general classification (i.e., 1,2,3 \rightarrow Ia,II,Ibc) or report a specific sub-type (e.g., IIn, Ic, etc.). Download the updated challenge data files only if you need the sub-types.

- **March 14, 2010:** Fixed bug in which about 1% of the SNe have pathological late-time magnitudes. Download data files after date-stamp above.
- **March 24, 2010:** Fixed bug in which a few dozen non-Ia SNe have pathological magnitudes at all epochs.
- **April 13, 2010:** Fixed two bugs related to type II SNe. First, the wrong redshift was mistakenly used for one of the observed IIP, resulting in a 2 mag overestimate of its brightness. Second, for another type II SN the absolute mag was mistakenly set 0.3 mag too bright. While the generated fraction of these buggy SNe was small, their contribution to the challenge sample after requiring $S/N > 5$ was relatively large; therefore the updated sample has ~ 1400 fewer SNe.
- **April 27, 2010:** No bug-fixes, but we have decided to fix $W_{\text{Ia}}^{\text{false}} = 3$ for the $\mathcal{C}_{\text{FoM-Ia}}$ calculation, and allow participants to optimize accordingly. Also, to help check for buggy submissions, please include your evaluation of the Ia-purity and Ia-efficiency for the spectroscopically confirmed subset.

REFERENCES

- Astier, P. et al. 2006, *A&A*, 447, 31
 Bernstein, J. P., Kessler, R., Kuhlmann, S., & Spinka, H. 2009, *ArXiv:0906.2955*
 Dahlen, T. & Goobar, A. 2002, *PASP*, 114, 284
 Freedman, W. et al. 2009, *ApJ*, 704, 1036
 Frieman, J. A. et al. 2008, *AJ*, 135, 338
 Ivezić, Ž. et al. 2008, *arXiv:0805.2366*
 Johnson, B. D. & Crotts, A. P. S. 2006, *AJ*, 132, 756
 Kessler, R. et al. 2009, *ApJS*, 185, 32
 Kuznetsova, N. V. & Connolly, B. M. 2007, *ApJ*, 659, 530
 LSST Science Book, A. 2009, in *arXiv:0912.0201*
 Oyaizu, H. et al. 2008a, *ApJ*, 674, 768
 —. 2008b, *ApJ*, 689, 709
 Perlmutter, S. et al. 1999, *ApJ*, 483, 565
 Poznanski, D., Maoz, D., & Gal-am, A. 2007, *AJ*, 134, 1285
 Poznanski, D. et al. 2002, *PASP*, 114, 833
 Riess, A. et al. 1998, *AJ*, 116, 1009
 Riess, A. G. et al. 2004, *ApJ*, 607, 665
 Rodney, S. A. & Tonry, J. L. 2009, *AJ*, 707, 1064
 Sullivan, M. et al. 2006, *AJ*, 131, 969
 Tonry, J. L. et al. 2003, *ApJ*, 594, 1
 Wood-Vasey, W. M. et al. 2007, *ApJ*, 666, 694