Photometric classification of SNe Ia in the SuperNova Legacy Survey with supervised learning

Anais Möller
CAASTRO Postdoc at the Australian National University (ANU)

Photometric Classification of SNIa Workshop, University of Chicago
May 28th, 2016

real data application

SNLS deferred photometric pipeline

a complete photometric SNIa pipeline

different classification methods

we classify SNe Ia using only photometric info

- sequential cuts with host-galaxy photometric z & SALT2

  G. Bazin et al. A&A 534, A43 (2011)

- supervised learning with SN z & general fitter

  Möller et al. in preparation (2016)

- supervised learning with host-galaxy photometric z & SALT2

SNLS

the SuperNova Legacy Survey

- based on the **C**anada **F**rance **H**awaii **T**elescope
- **MegaCam** : 36 CCD mosaic
- 4 broadband **filters g,r,i,z**
- 4 **fields** of 1 square degree
- **rolling search** mode
- spectroscopic follow up (Keck, Gemini and VLT)
- observations: 2003-2008
  - **SNLS3** analysed and published
  - **SNLS5** currently being processed (complete SNLS data set)

SNLS pipelines

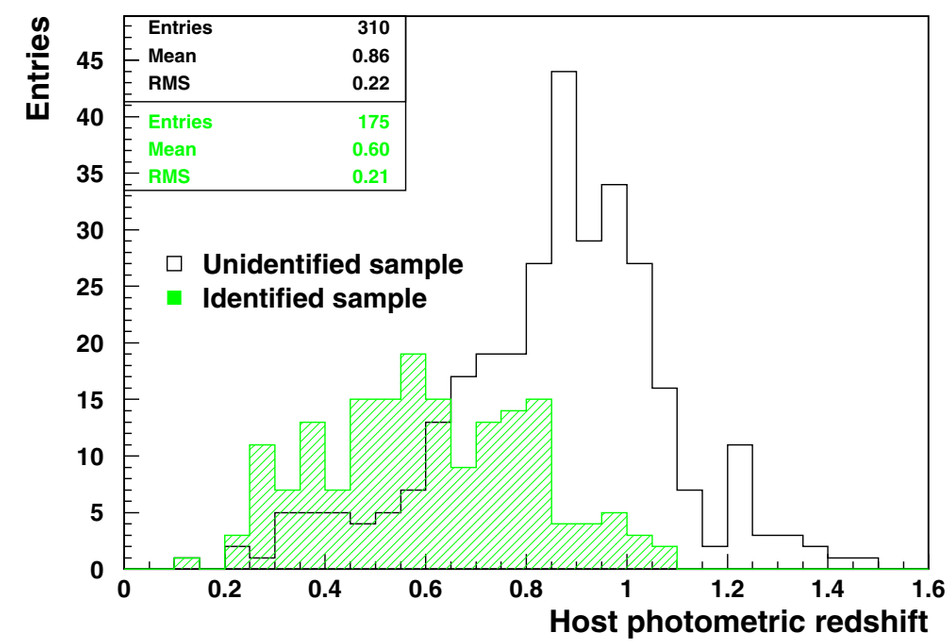
standard SNLS approach *e.g. Astier et al. 2005, Guy et al. 2010, used in JLA Betoule et al. 2015*

- real time detection, uses spectroscopy for typing and redshift
- 0.15 < z <1.1

deferred photometric pipeline

developed in the SNLS Saclay group (France)

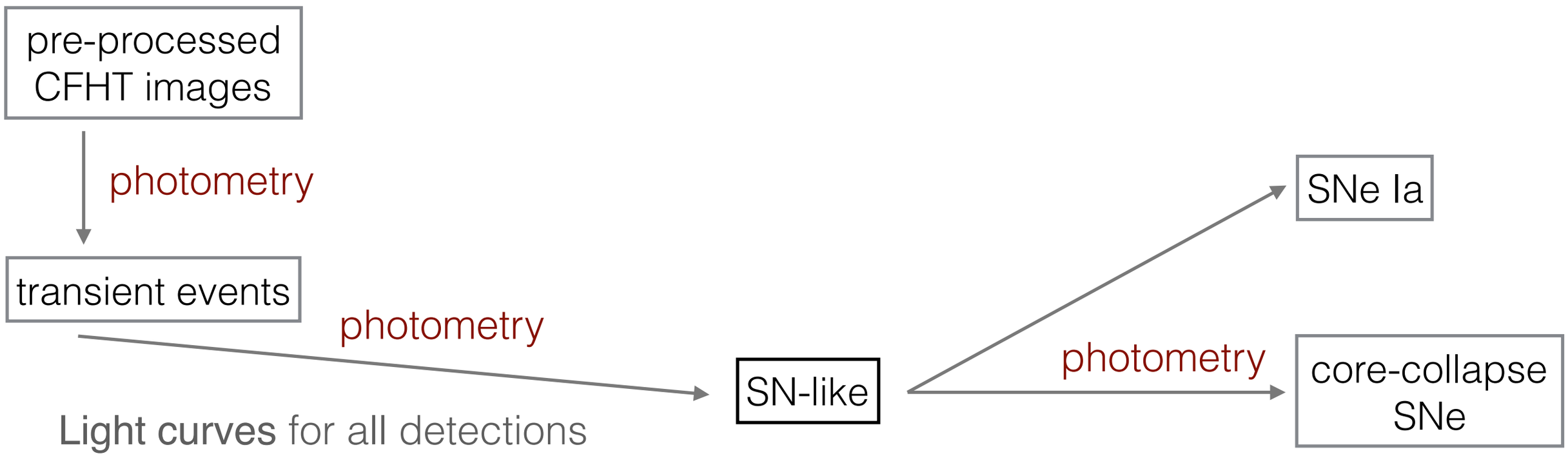
- no spectroscopy required
- differed detection of all kinds of transient events
- larger number of detections
- larger redshift coverage
- sensitive to other SN types



feasibility of detecting and classifying SNe Ia shown with SNLS3 data

*G. Bazin et al. A&A 534, A43 (2011) **Photometric selection** of Type Ia supernovae in the Supernova Legacy Survey.*
*G.Bazin et al. A&A 499, Issue 3,(2009) The **core-collapse rate** from the Supernova Legacy Survey.*

SNLS photometric pipeline



Light curves for all detections are constructed.

Preselection cuts (SN-like):

-One significant flux variation

SN-like variation

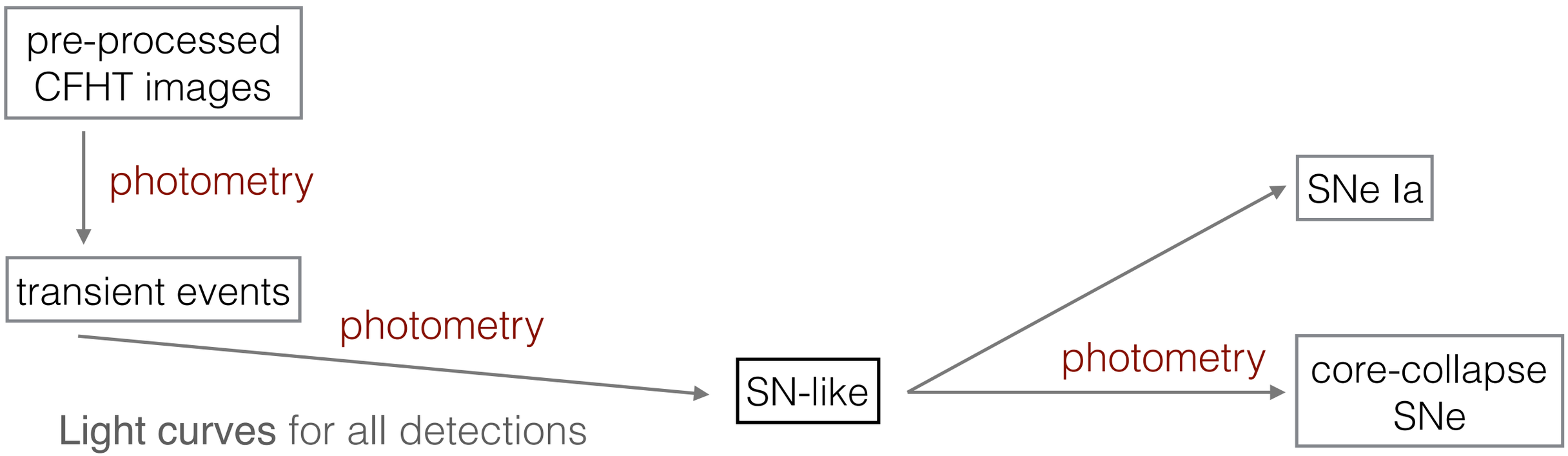
Star rejection

-Quality cuts on light curves

SNLS3 original pipeline

300,000 -> 1,500

SNLS photometric pipeline



Light curves for all detections are constructed.

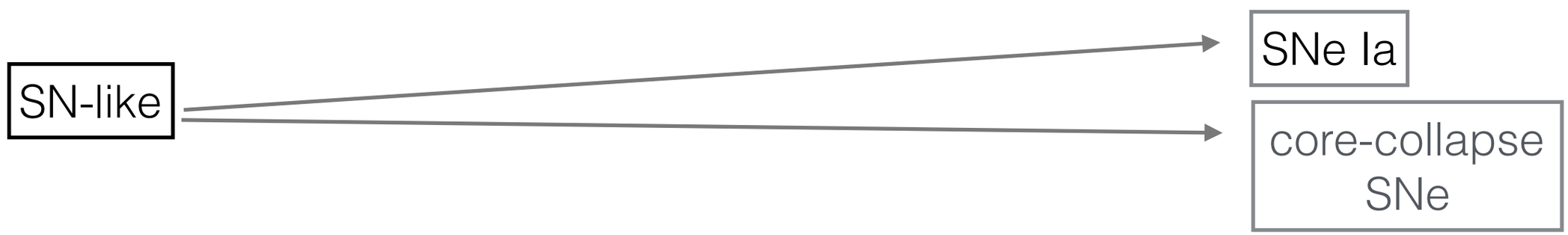
Preselection cuts (SN-like):
-One significant flux variation
SN-like variation
Star rejection
-Quality cuts on light curves

SNLS3 original pipeline
300,000 -> 1,500

this pipeline was designed to be sensitive to other SN-types.

<u>SN-like cuts are loose.</u>

SN classification



Goal:

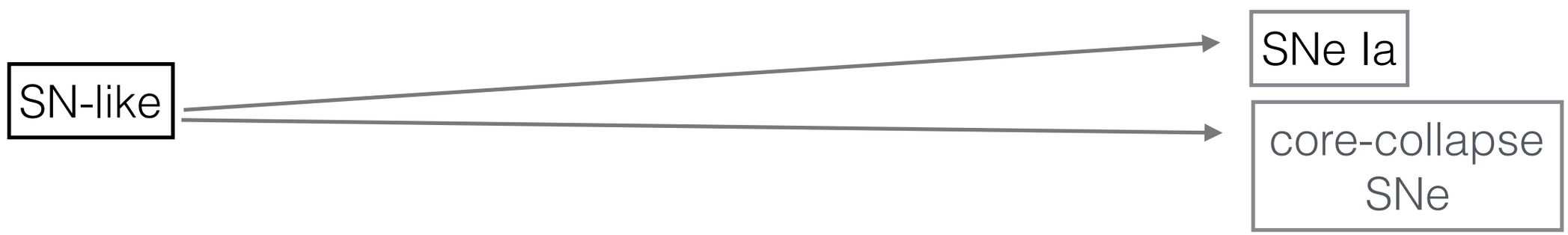
classify a large SN Ia sample with high-purity.

How:

1. extract information:
   A. light-curve features: SALT2, general LC fitter
   B. redshift: host-galaxy photometric, SN-redshift
2. classification strategy: sequential cuts, machine learning

this looks like our usual classification challenge

SN classification



Goal:

classify a large SN Ia sample with high-purity.

How:

1. extract information:
   A. light-curve features: SALT2, general LC fitter
   B. redshift: host-galaxy photometric, SN-redshift
2. classification strategy: sequential cuts, machine learning

are we forgetting something?

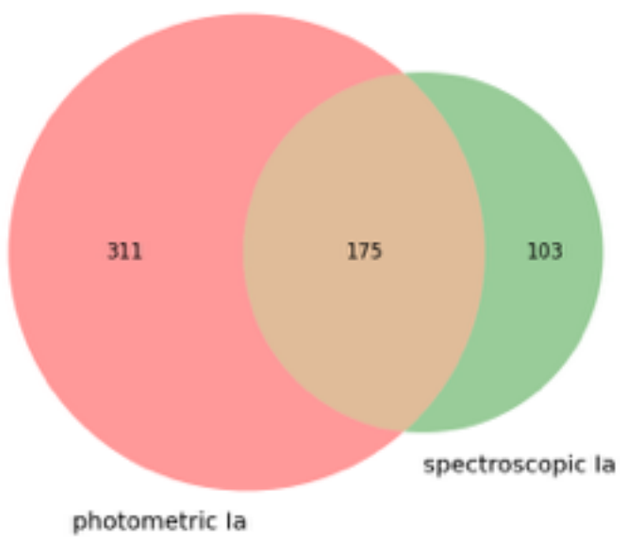
real data, a complete pipeline.

Classification starts from SN-like so we must be aware of backgrounds and selection cuts

SN classification

evaluated using simulations and data

SNLS3
contains large spectroscopically and photometrically identified SNe samples



the core-collapse photometric sample was determined by *Bazin et al. 2009* in a specific analysis.

classification
signal= SNIa with correct z
background={SNIa with incorrect z, core-collapse SNe}

method 1

How:

G. Bazin et al. A&A 534, A43 (2011)

1. extract information:
   A. light-curve features SALT2: C, x1, magnitudes and chi^2 of the fit.
   B. redshift host-galaxy photometric: event coordinates matched to a host galaxy using an external catalogue.

      Ilbert et al. 2006: 520,000 galaxies with an AB magnitude brighter than 25 in i_M .

      ~4% resolution , catastrophic assignment ($\Delta z/(1 + z) > 0.15$): **3.7%**

2. **selection cuts:**
   - sampling
   - poor fit :fitted bands chi^2 of the fit (-7 spe Ia), unfitted bands (- 10 spe Ia)

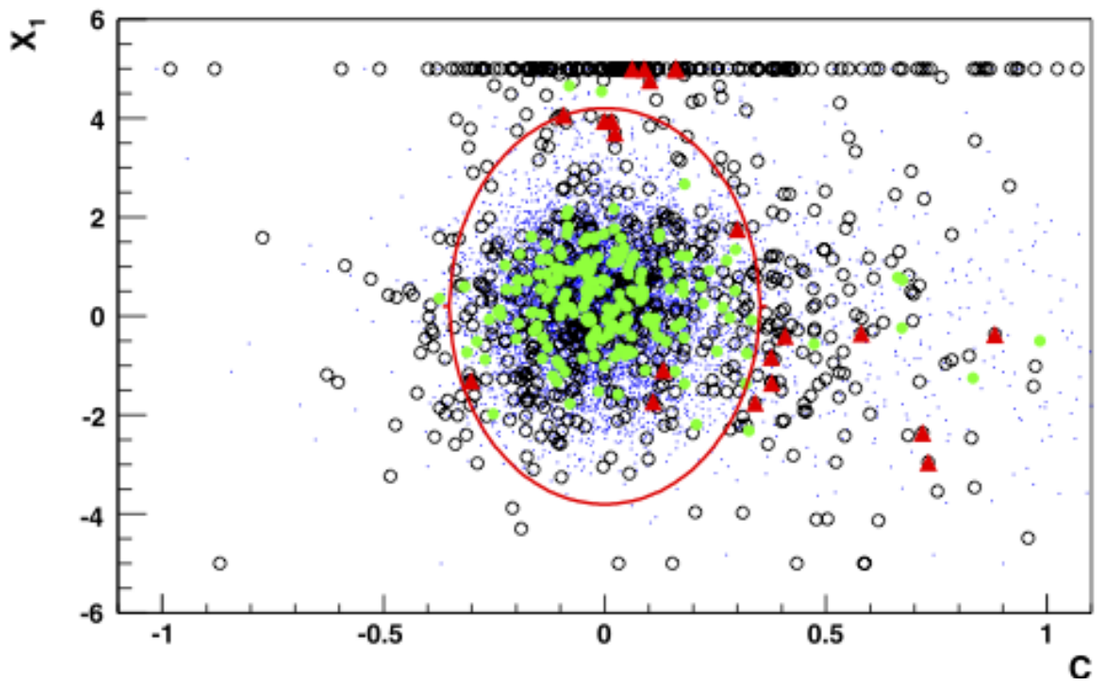
3. classification strategy: sequential cuts

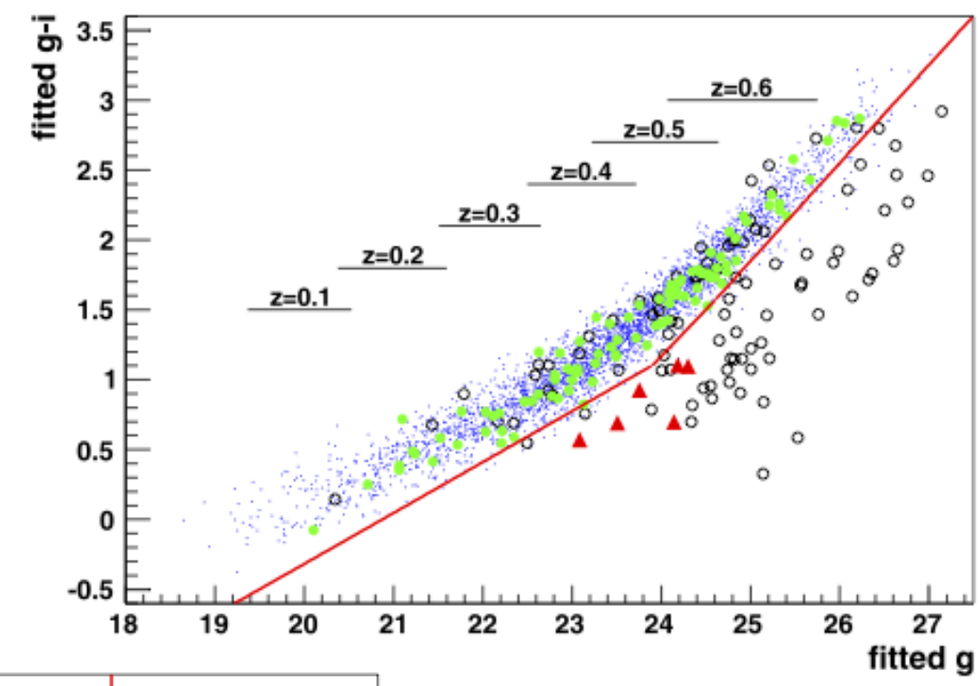
method 1

SNLS3

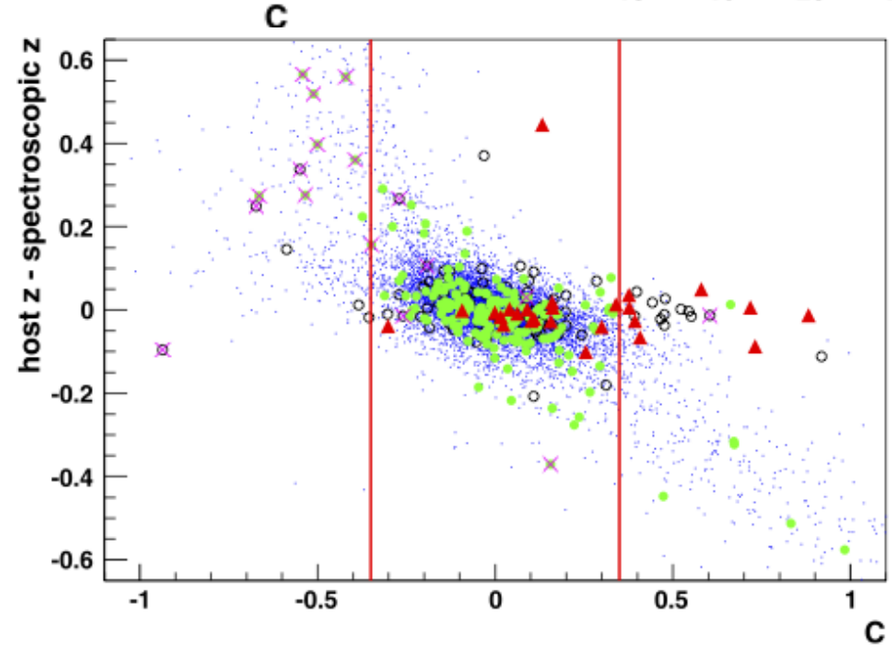
G. Bazin et al. A&A 534, A43 (2011)

zgal + SALT2 + sequential cuts





- SNLS3 data
- spectroscopic SNe Ia
- spectroscopic CC
- simulated SNe Ia



method 1

SNLS3

Simulation

purity	contamination		efficiency
SNe Ia	bad redshift SNe Ia	core-collapse	SNIa
$94.4 \pm 0.5\%$	$0.65 \pm 0.08\%$	$4.9 \pm 0.5\%$	$29.9 \pm 0.3\%$

SNLS3 data

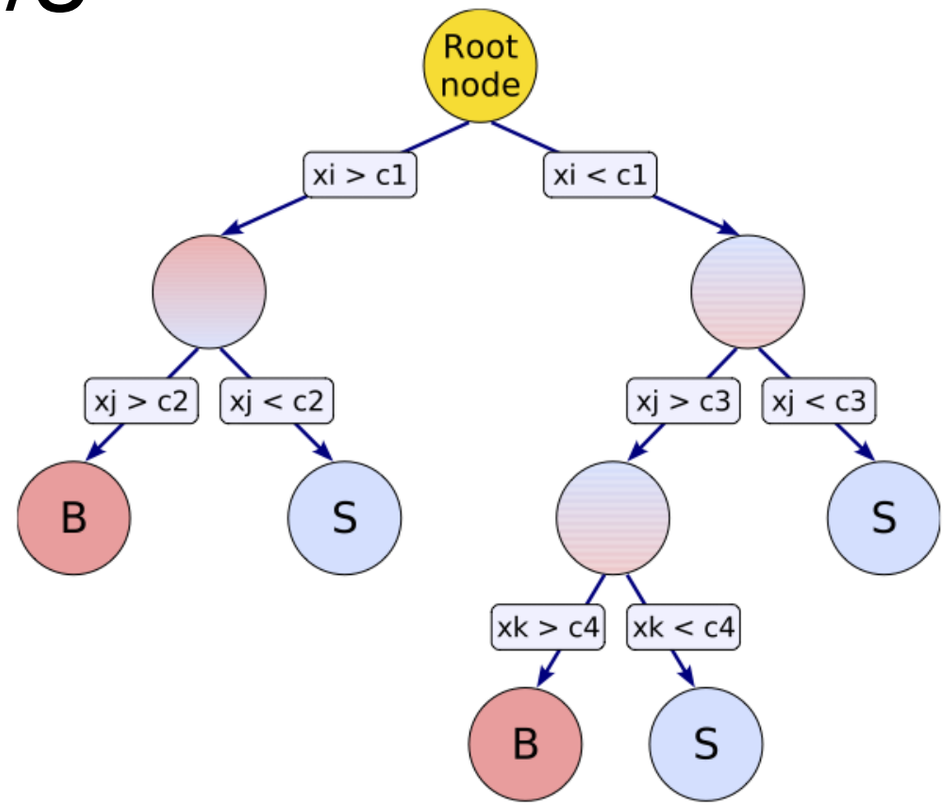
# events	# spectroscopic SNe Ia	# spectroscopic CC	# photometric CC
486	175	0	0

the core-collapse photometric sample was determined by *Bazin et al. 2009* in a specific analysis.

BUT host-galaxy redshift assignment (**zgal**) **efficiency** is **83%**

method 2

introductions 1/3

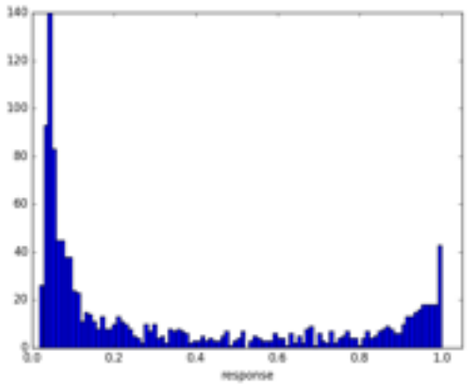


Boosted Decision Trees (BDT)

- supervised classification methods
- **decision trees** make successive rectangular cuts in the parameter space
- improving model:
  - boosting: making the model more complex AdaBoost, XGBoost
  - averaging methods: Random Forest, Bagging

redshift + features → BDT response

N-dimensional problem 1-dimensional problem



method 2

introductions 2/3

SN redshift (zpho)

Palanque-Delabrouille et al. 2010

- estimated **directly** from **SN light curves**.

light curves are fitted using **SALT2** searching **minimisation of the reduced χ^2**

1. for successive values of the redshift separated by $\delta z = 0.1$ with color and stretch set with gaussian priors.

2. setting the color and stretch parameters free and performing a new redshift scan around first step's redshift with the same step δz as above.

- set up on SNLS3 data

- *~2% resolution* , *catastrophic assignment* ($\Delta z/(1 + z) > 0.15$): 1.4%

method 2

introductions 3/3

general type fitter (gen fit)

SN photometric redshift algorithm uses SALT2. Advisable to use a light curve fitter which is:

- independent from SALT2
- able to fit other types of SNe. e.g. the empirical model used for SN-like selection.

$$f^k = A^k \frac{e^{-(t-t_0^k)/\tau_{fall}^k}}{1+e^{-(t-t_0^k)/\tau_{rise}^k}} + c^k$$

k= filter

$$t_{max}^k = t_0^k + \tau_{rise}^k ln(\tau_{fall}^k/\tau_{rise}^k - 1)$$

method 2

Möller et al. (2016) in preparation

How:

1. extract information:
   A. light-curve features general fitter
   B. redshift SN-redshift

2. **selection cuts** (adapted to our analysis)**:**
   - sampling
   - poor fit : chi^2 of the SN-z fit, checked twice

3. classification strategy: supervised learning

method 2

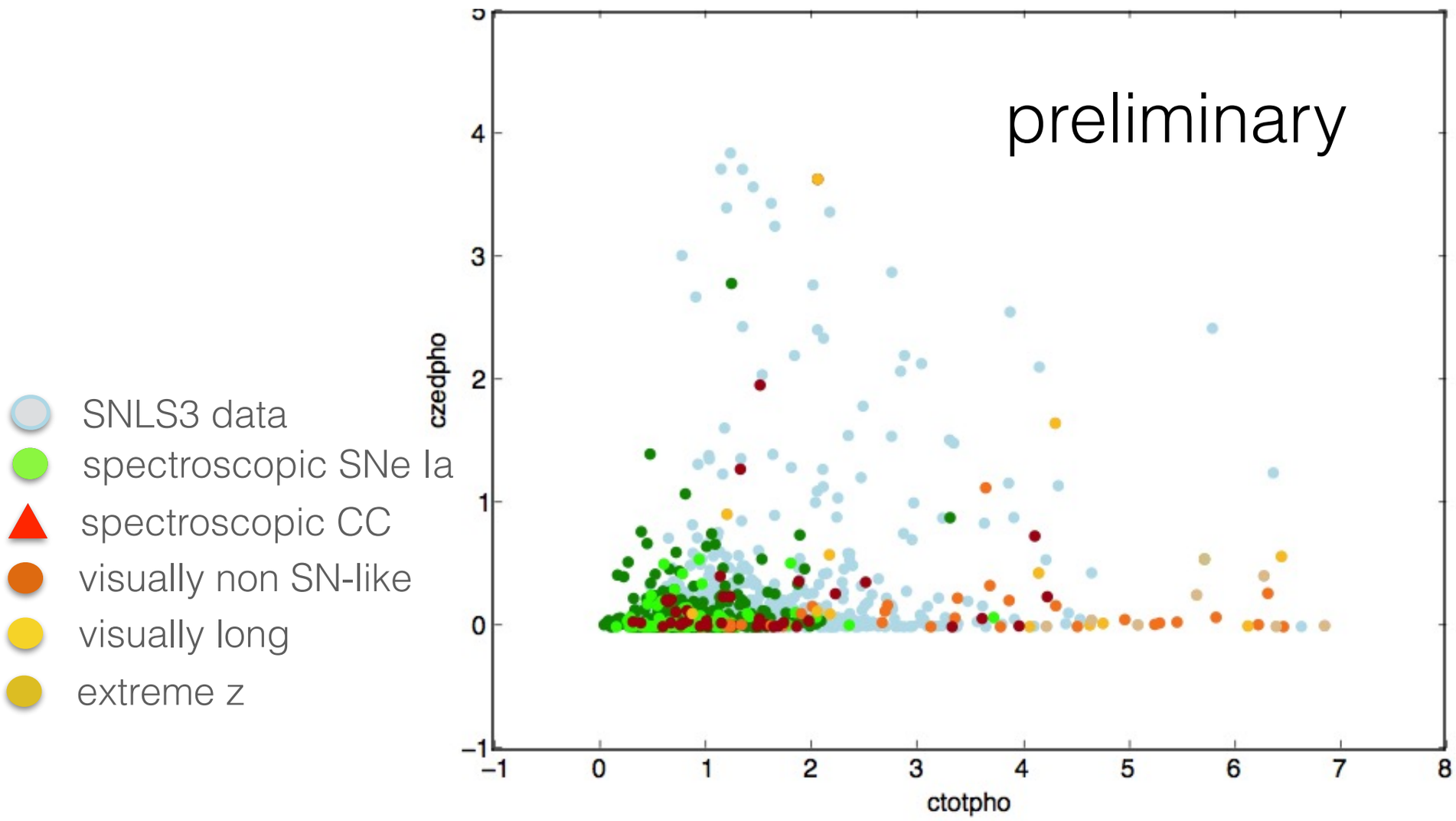


Figure 2: SNLS3 data: χ^2 of redshift fit against χ^2 ot total fit. Classified events are shown in: type Ia in (bright=spectroscopic, dark=photometric) green, core-collapse SNe in red. Visual inspection of light-curves were used to determine if some events were not SN-like (orange) and long events similar to type II SNe (yellow). In beige we show events with extreme photometric redshifts (when compared to spectroscopic redshifts).

method 2

How:

1. extract information:
   A. light-curve features general fitter
   B. redshift SN-redshift

2. selection cuts (adapted to our analysis):
   - sampling
   - poor fit : chi^2 of the SN-z fit, checked twice

3. classification strategy: supervised learning
   - setting parameters
   - selecting features
   - crossvalidation

   using scikit learn packages

-Random Forest

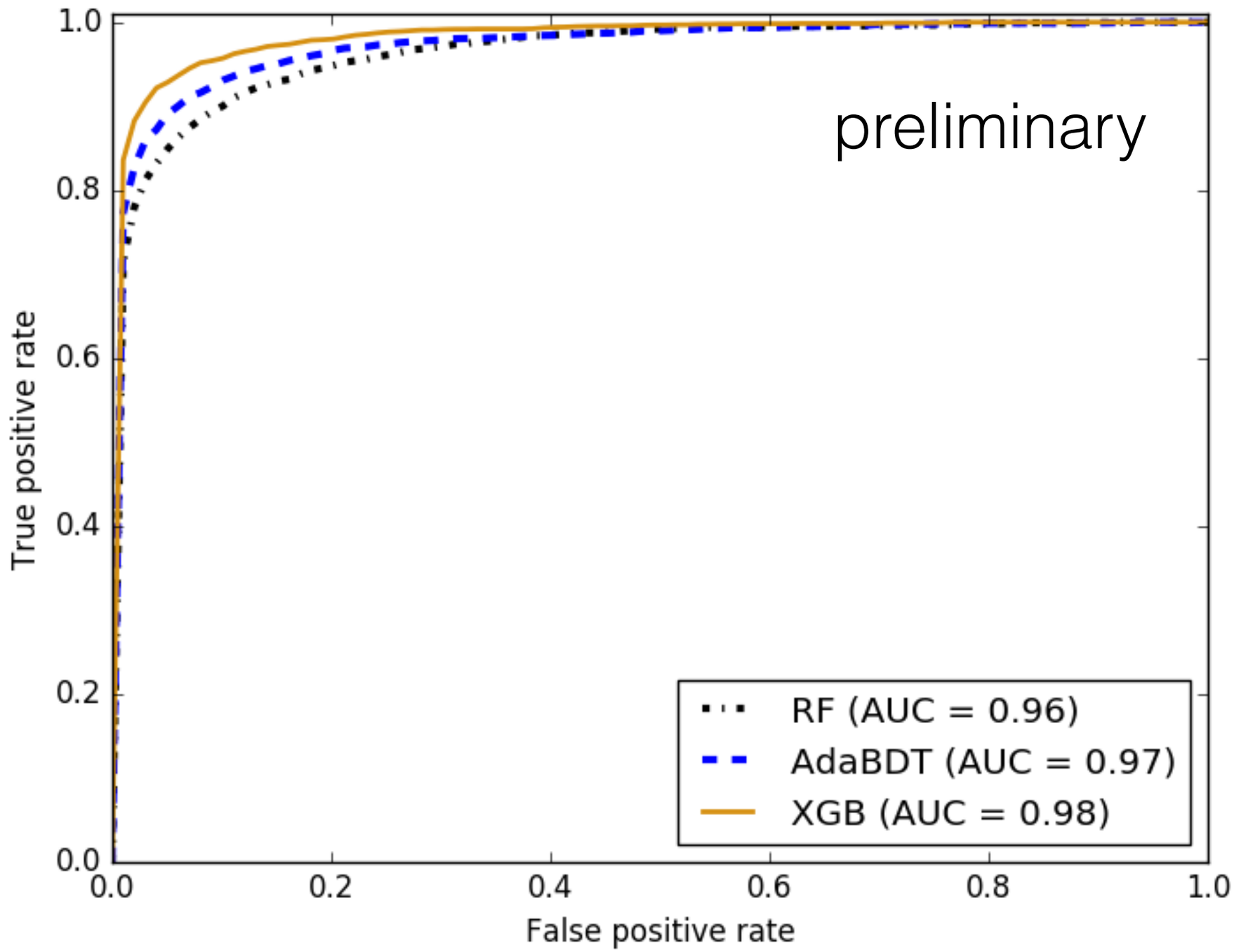
-AdaBoost

-XGBoost (extreme gradient boosting)

. http://xgboost.readthedocs.org/

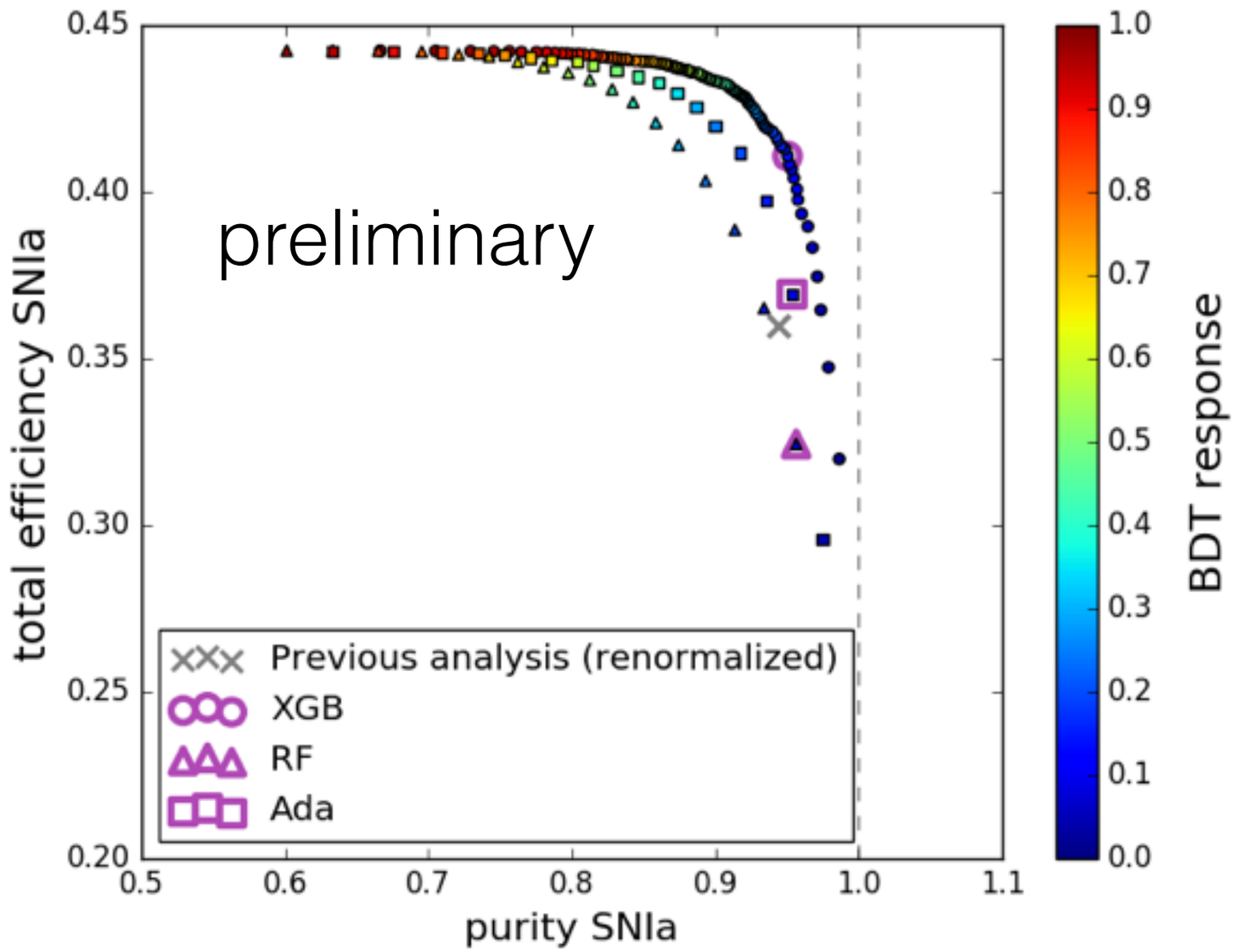
method 2

Simulations



method 2

Simulations



we want to compare classified samples -> **95% purity**

method 2

Simulations 95% purity

		AdaBoost	Random Forest	XGBoost
total efficiency	Ia	36.9 ± 0.6	32.4 ± 0.7	41.1 ± 0.7
purity	Ia	95.6 ± 0.5	95.6 ± 0.4	95.3 ± 0.4
contamination	Ia bad z	0.53 ± 0.09	0.29 ± 0.07	0.60 ± 0.09
contamination	CC	3.9 ± 0.4	4.1 ± 0.4	4.0 ± 0.4

preliminary

data

	AdaBoost	Random Forest	XGBoost
photometric sample	478	549	677
spectroscopic Ia	166	198	223
photometric Ia	318	364	444
spectroscopic CC	1	0	2
photometric CC	1	1	6

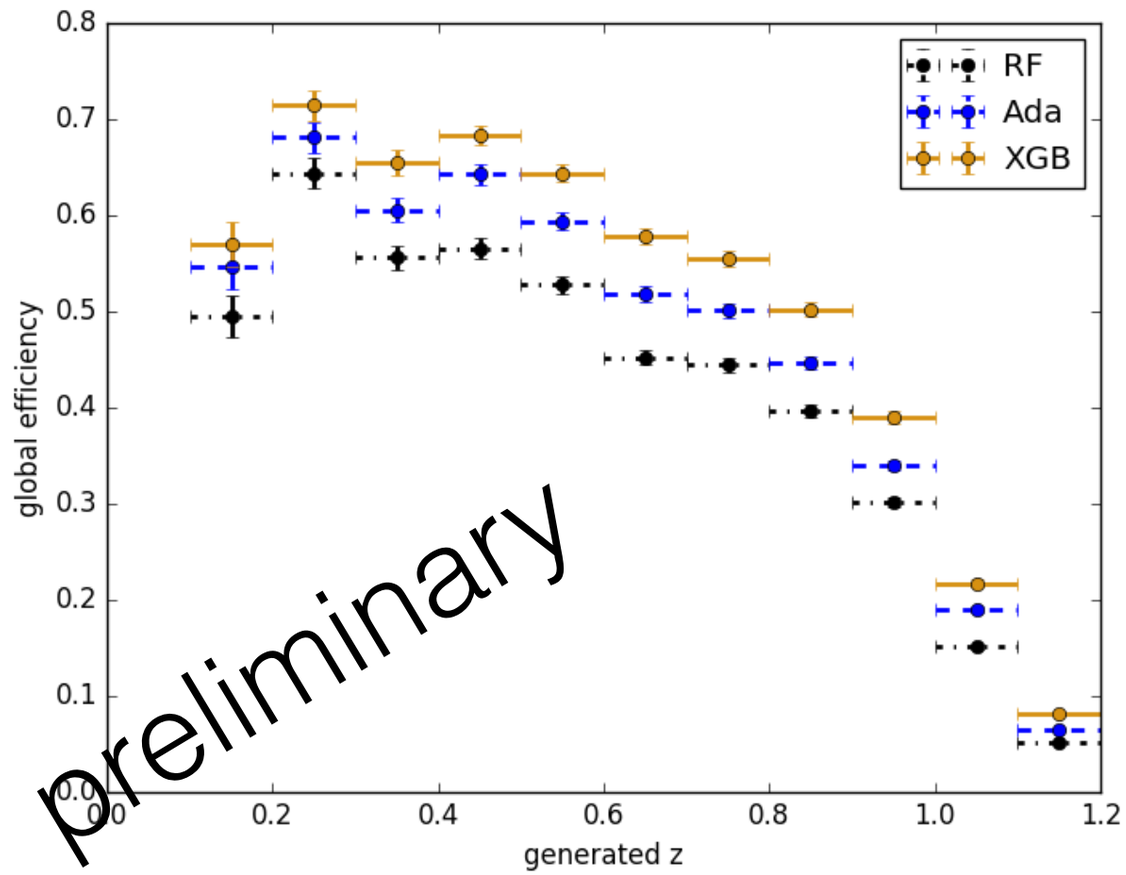
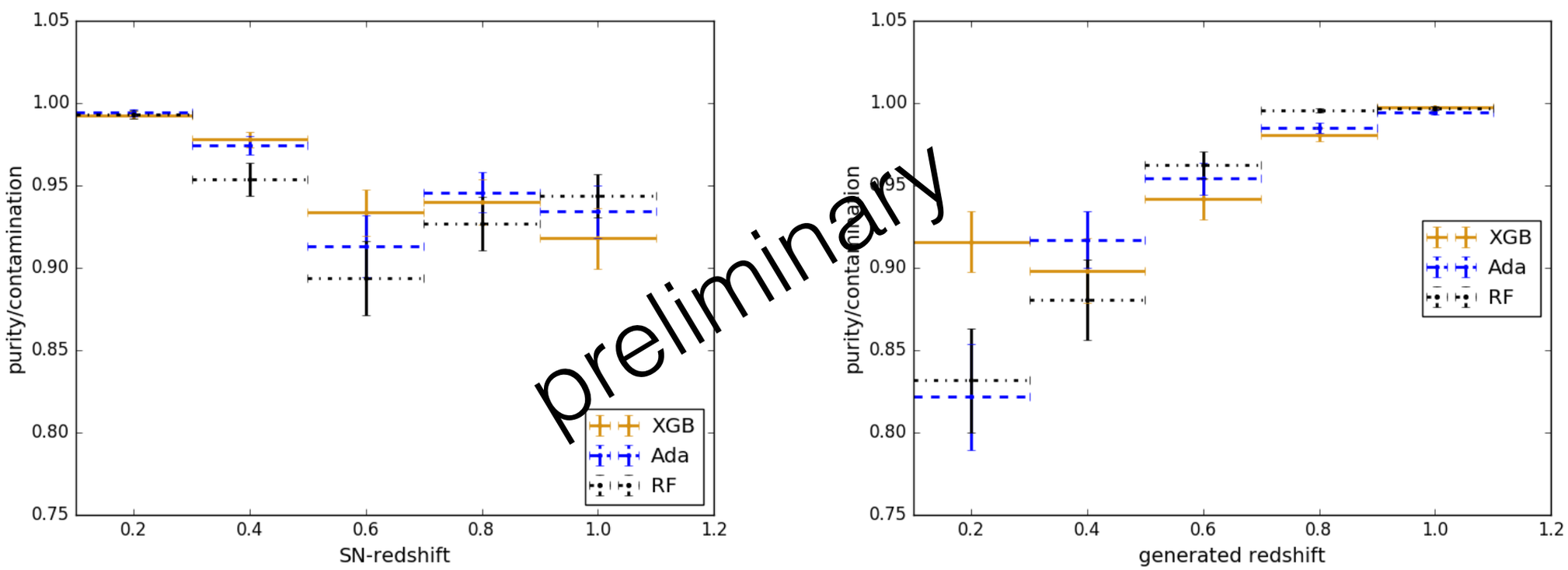


Figure 7: Selection efficiency from synthetic SN Ia light curves as a function of the generated redshift, for different classification methods with a given purity of 95%.

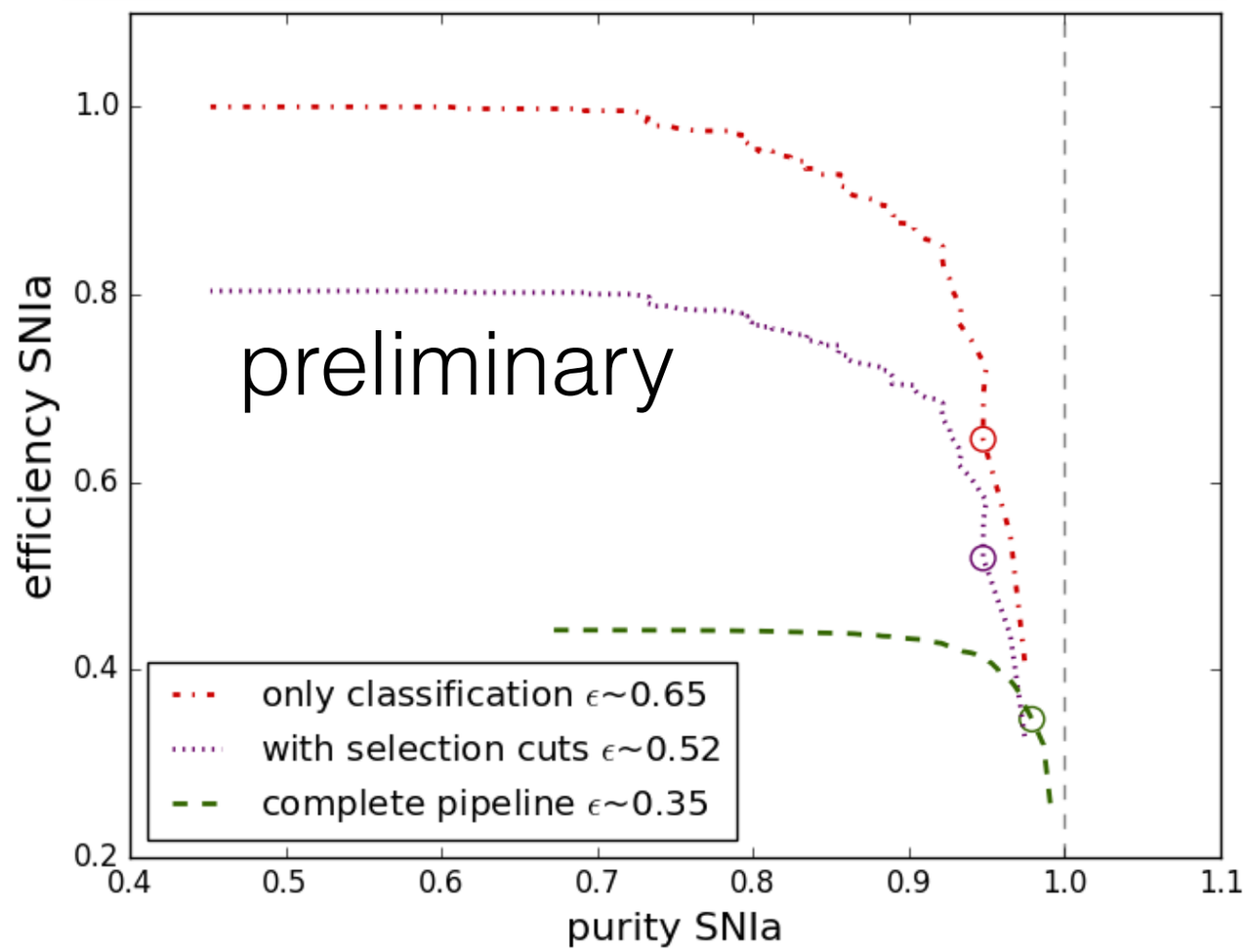


method 2

SN-z, general fit, XGB 98% purity

cut	events	spectroscopic Ia	spectroscopic CC	photometric Ia	photometric CC	simulated $Ia\%$
SN-like	1483	246	33	486	109	0.50
selected	1193	238	21	481	77	0.47
classified	529	205	0	374	1	0.35

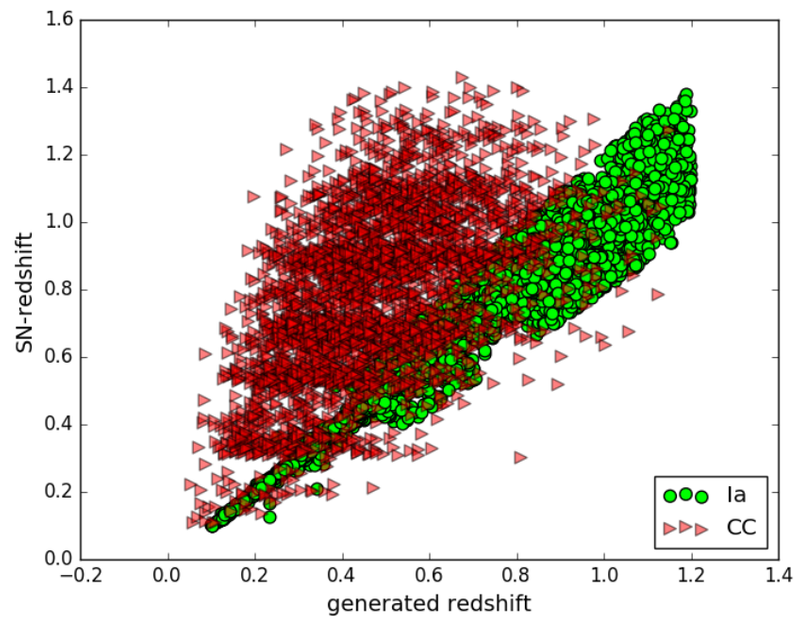
data



Simulations

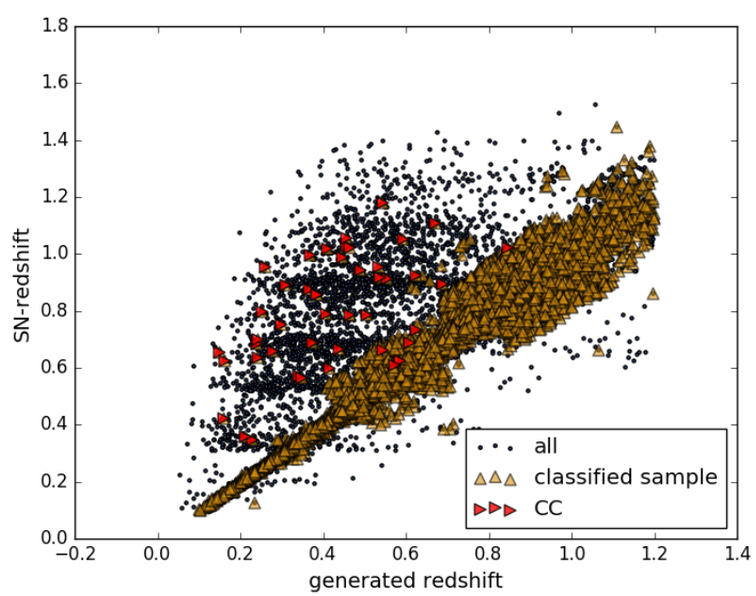
method 2

preliminary

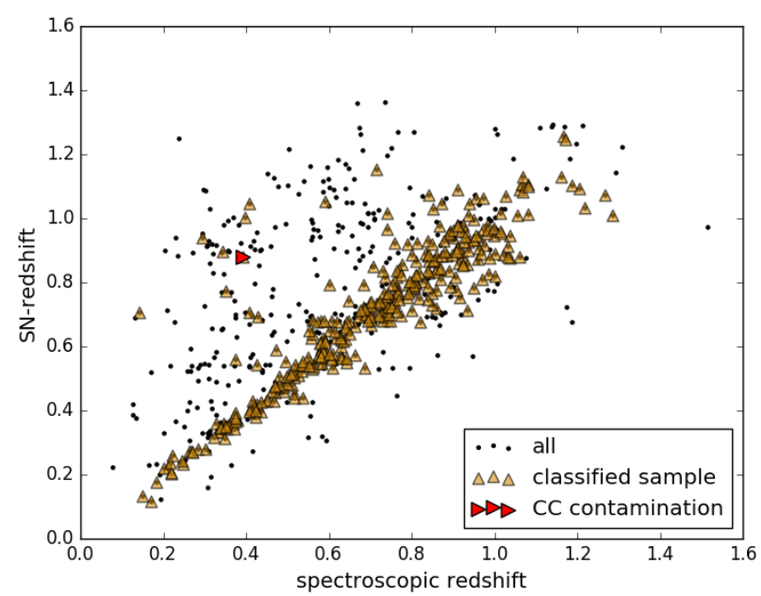


Simulations

Simulations



(a) synthetic SNe in our photometric sample



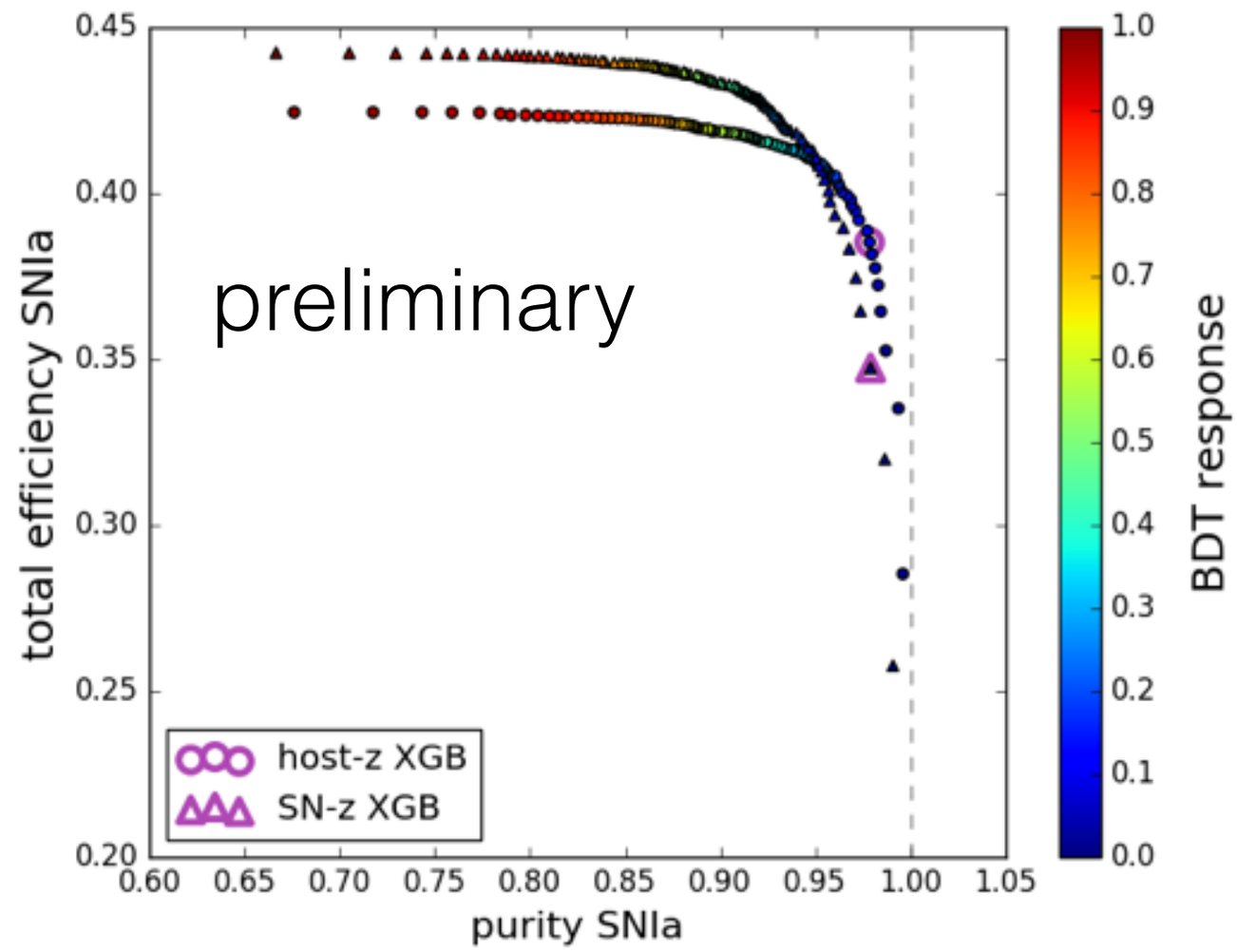
(b) SNLS3 SNe in our photometric sample

data

method “3”

host-galaxy photometric z, SALT2 fit, XGB

efficiency 38.2 +- 0.7
purity 98.0 +- 0.3



take away and discussion

SNLS deferred photometric pipeline:a unique opportunity for testing classification with real data.

selection cuts

- necessary to reduce other non-modelled backgrounds.
- reduces efficiency but is the price to pay if we want to trust our sample.

SN-z classification

- can achieve easily a purity above 95%
- we are able to disentangle core-collapse events that have colors and photometry coherent with type Ia (therefore they have a correct SN-z). May be attributed to general fitter features.

our BDT methods can also be used in SALT2+host-galaxy z classification

to be applied on SNLS5. crosschecking w. larger spectroscopic sample!

other: cosmology analysis errors, SNIa diversity?

lead by V. Ruhlmann-Kleider