Kernel PCA for Sne photometric classification

arXiv: astro-ph/1201.6676

Emille E. O. Ishida

Instituto de Astronomia, Geofisica e Ciencias Atmosfericas - Universidade de Sao Paulo (IAG/USP, Brazil)

In collaboration with:

Rafael S. de Souza
Korea Astronomy and Space Science Institute (KASI, Korea)

Summary

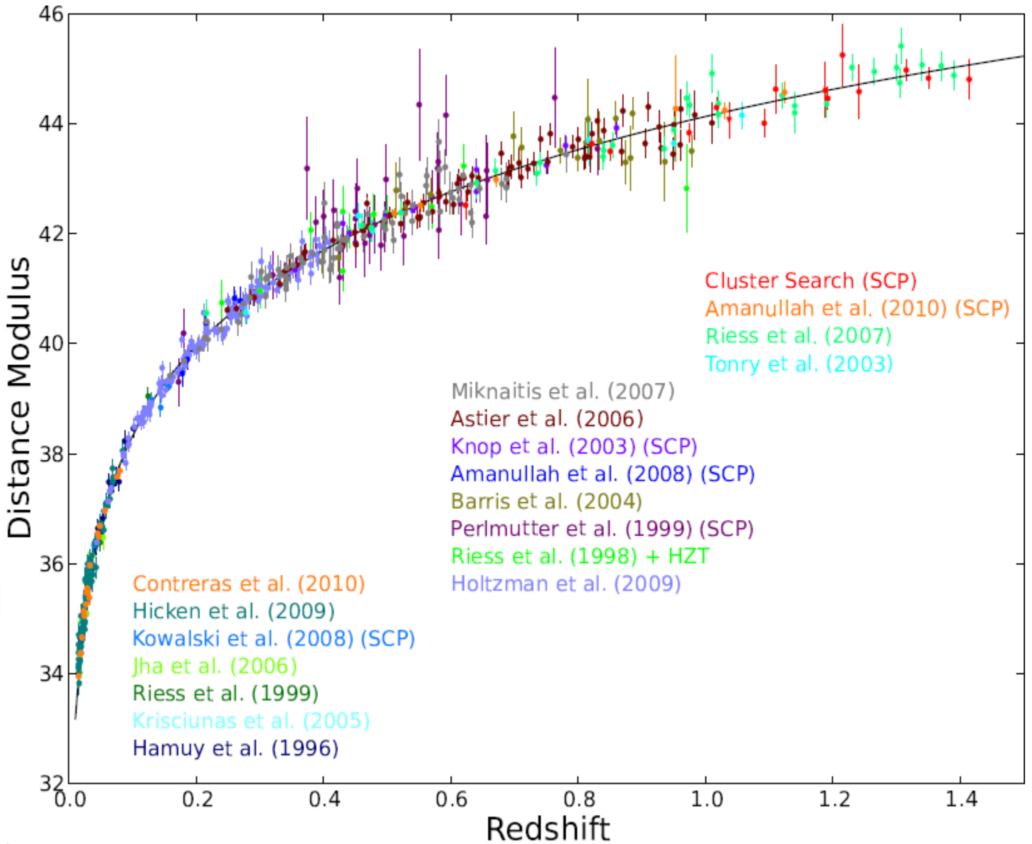
1. The accelerating Universe and Supernova Ia
2. The problem of supernova photometric classification
3. The Supernova Photometric Classification Challenge (SNPCC)
4. Applications of Principal Component Analysis (PCA)
5. Results from Kernel Principal Component Analysis (kPCA)
6. Perspectives
7. Final remarks

1. The accelerating Universe and type Ia SNe

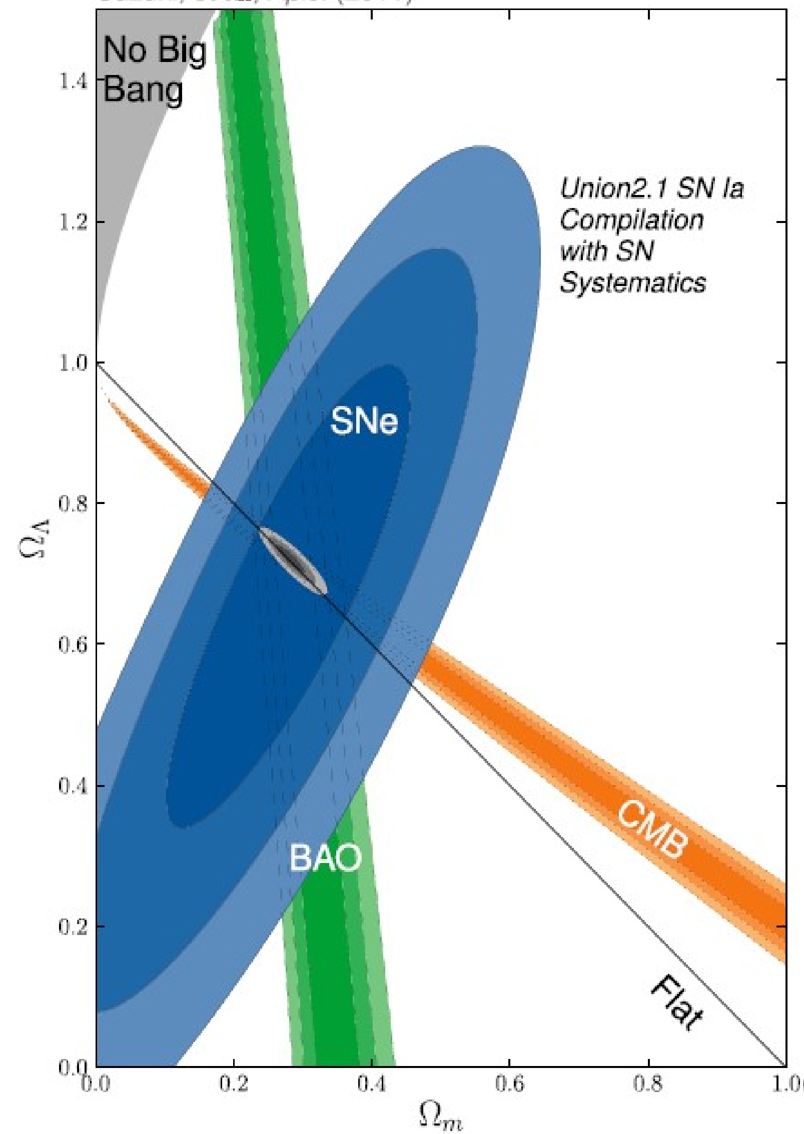
Current status of Sne samples

Union2.1 sample

580 type Ia SNe



Supernova Cosmology Project
Suzuki, et al., *Ap.J.* (2011)



FAPESP

1. The accelerating Universe and type Ia SNe

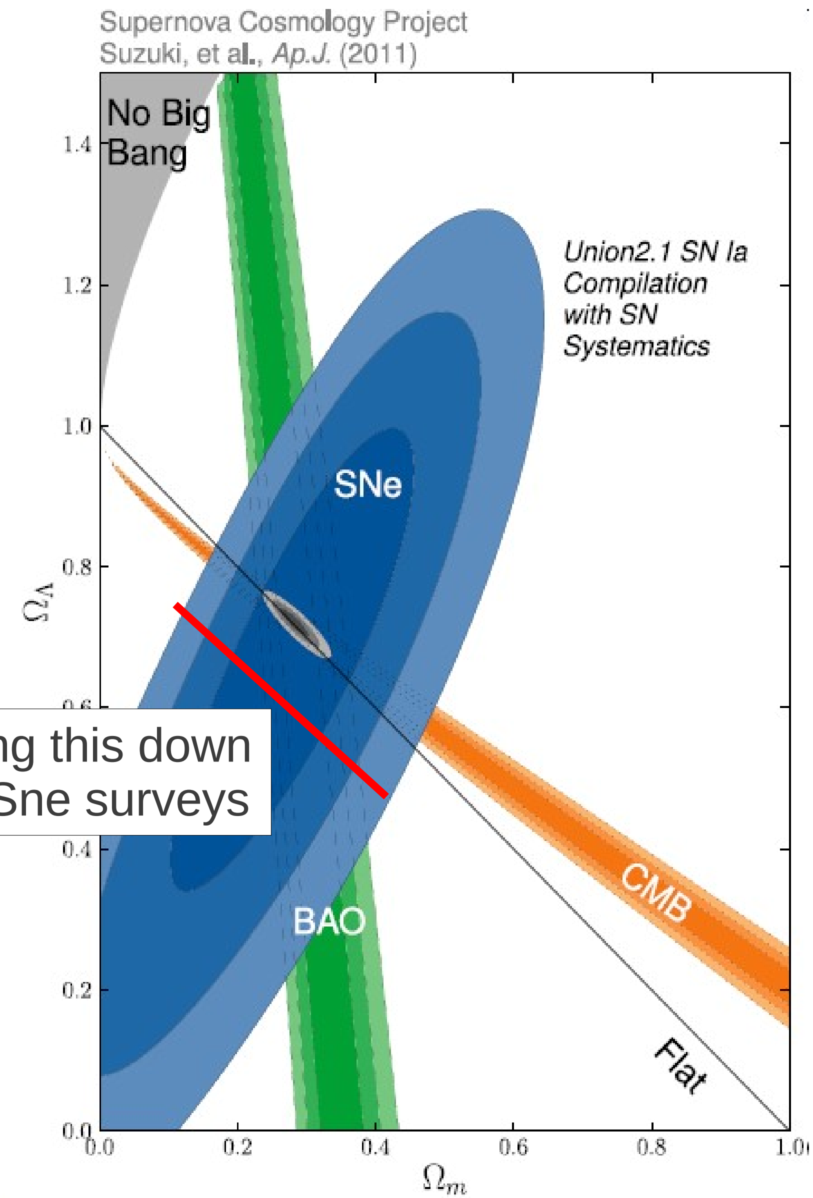
Current status of Sne samples

Sako *et al.*, 2011

SDSS, Sep-Nov of 2005-2007
~10.000 Sne candidates
(including other transients)

~500 spectroscopically confirmed

~ 3200 without spectroscopic confirmation



Narrowing this down is up to Sne surveys

1. The accelerating Universe and type Ia SNe

Message to take home:

The Sne photometric classification is not a problem of future large Sne surveys.

It is already here!

Our ability to extract cosmological information from current and future Sne surveys is highly dependent on it.

FAPESP

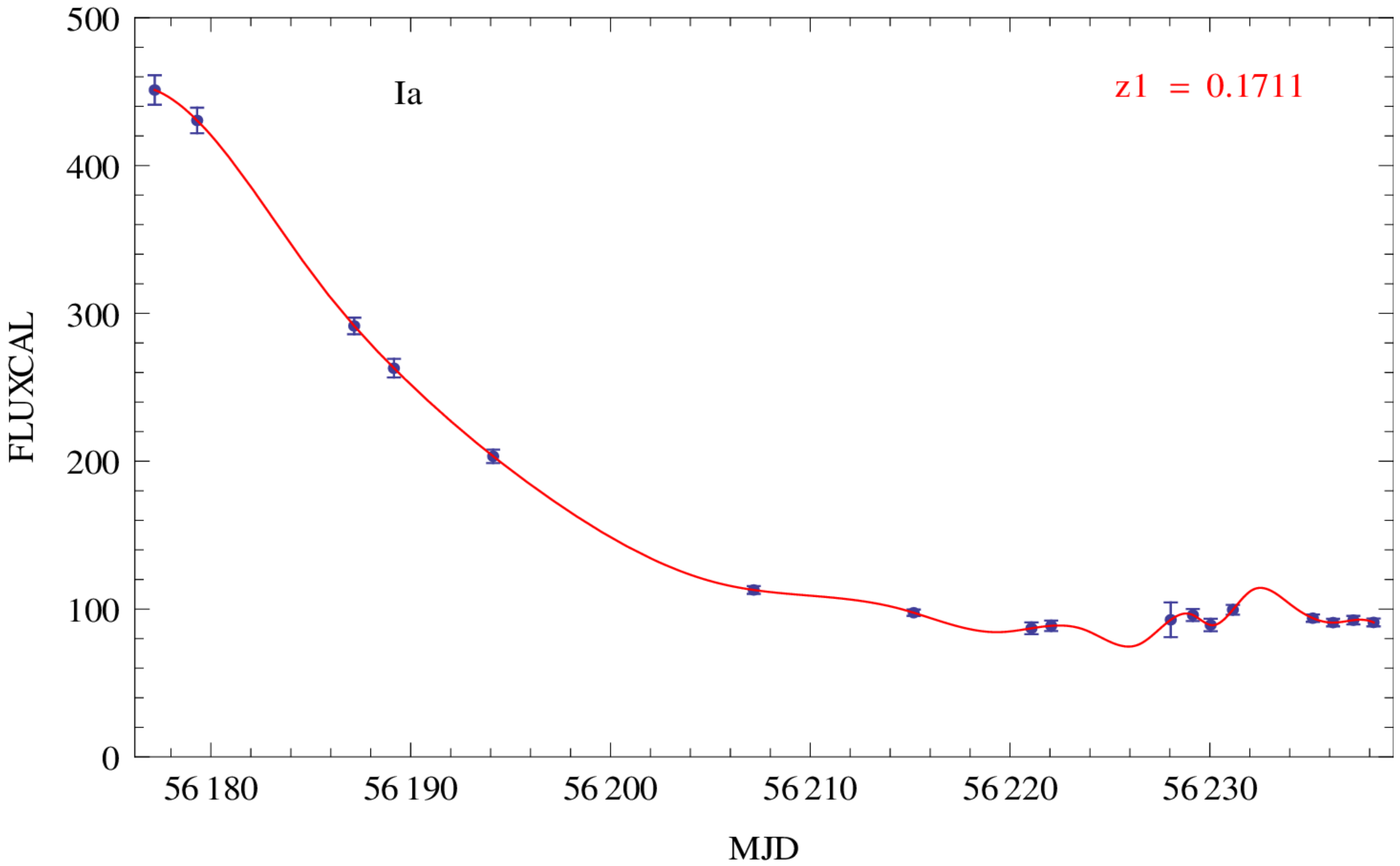
2. The Sne photometric classification problem

A glimpse on a simplified version of the problem:

Suppose the only transients that exist are supernovae of different classes

Ex: Ia, Ib, Ic, Ibc, IIn, IIP, IIL

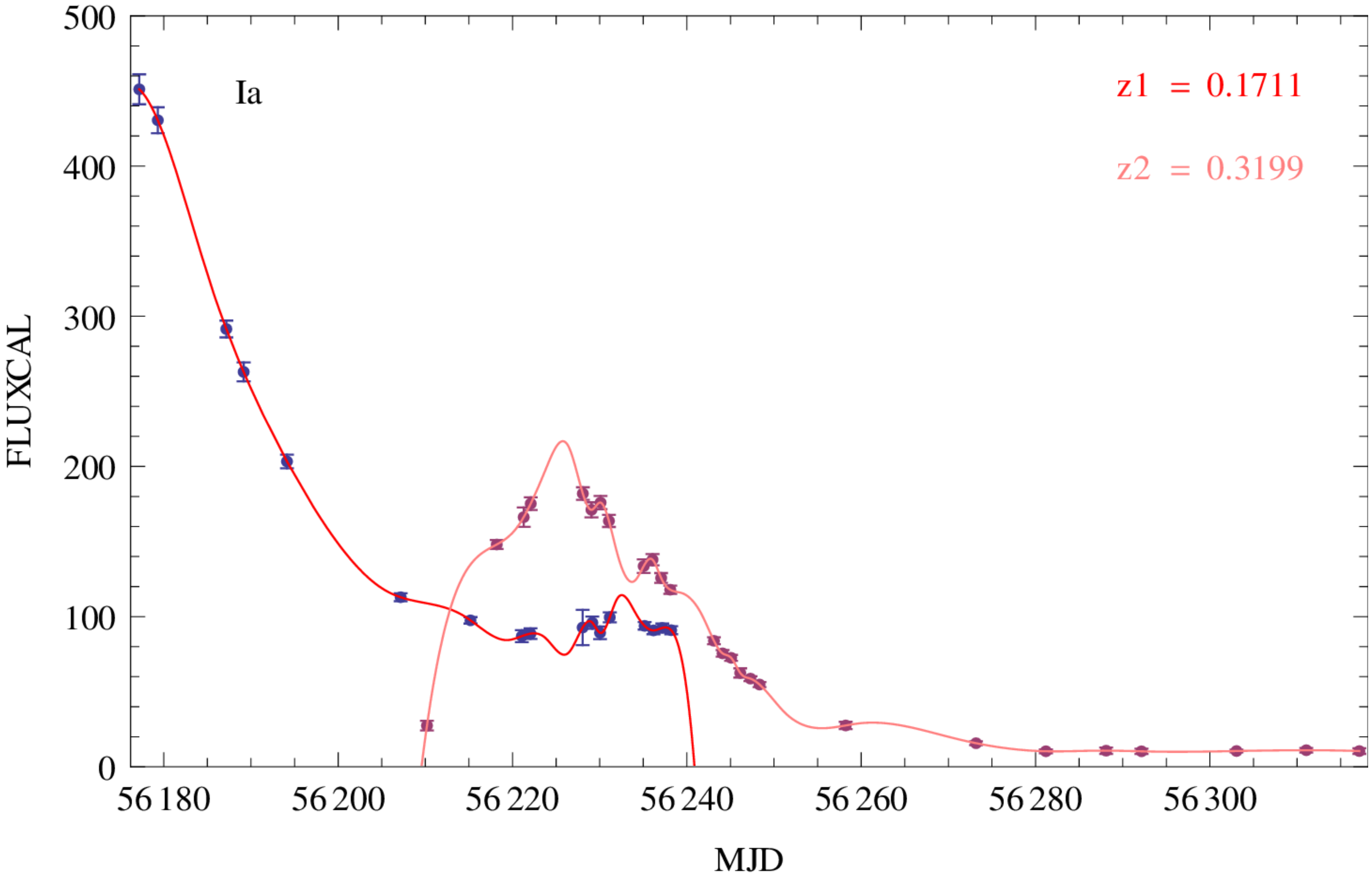
2. The Sne photometric classification problem



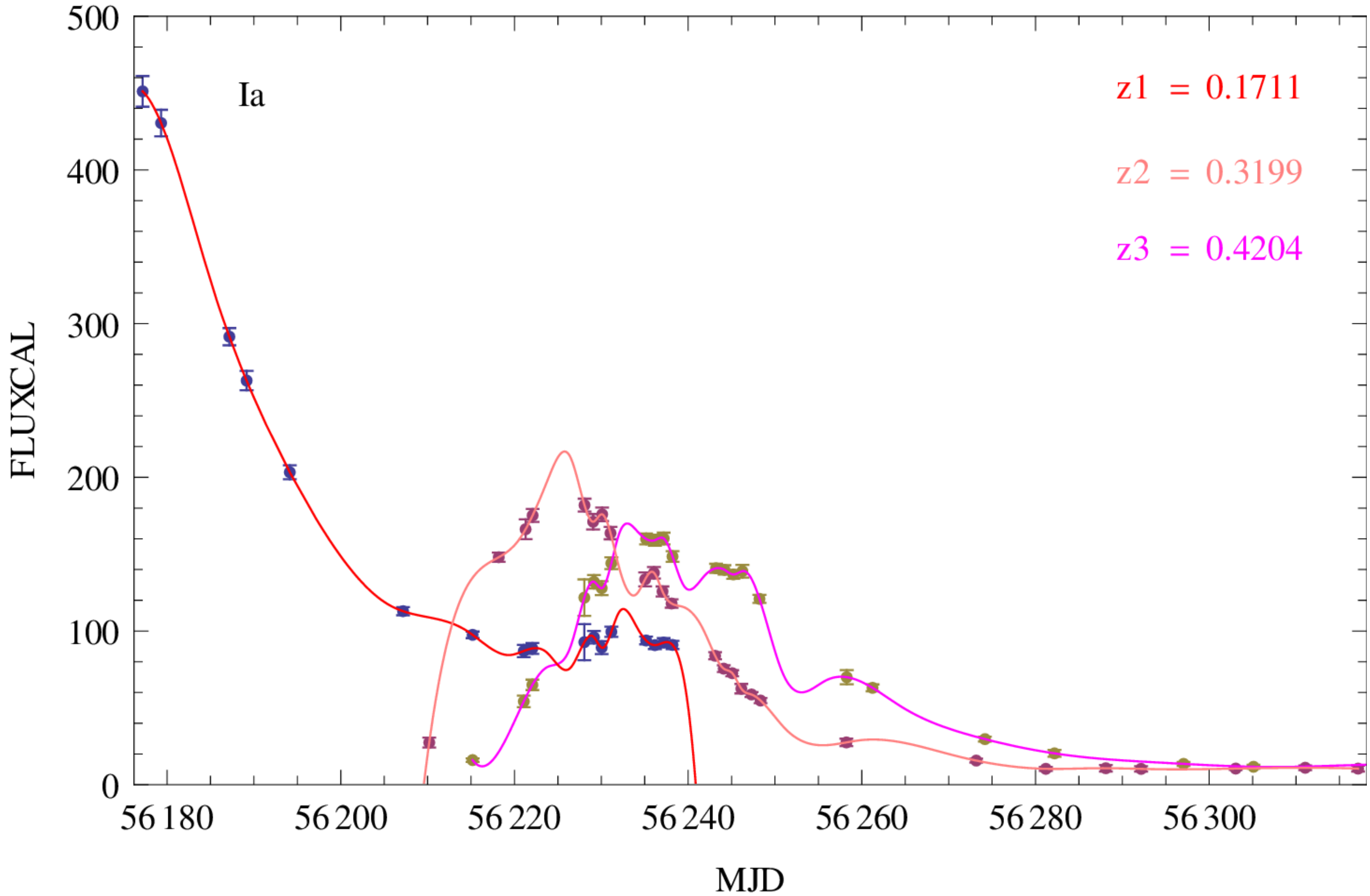
Constant time window

FAPESP

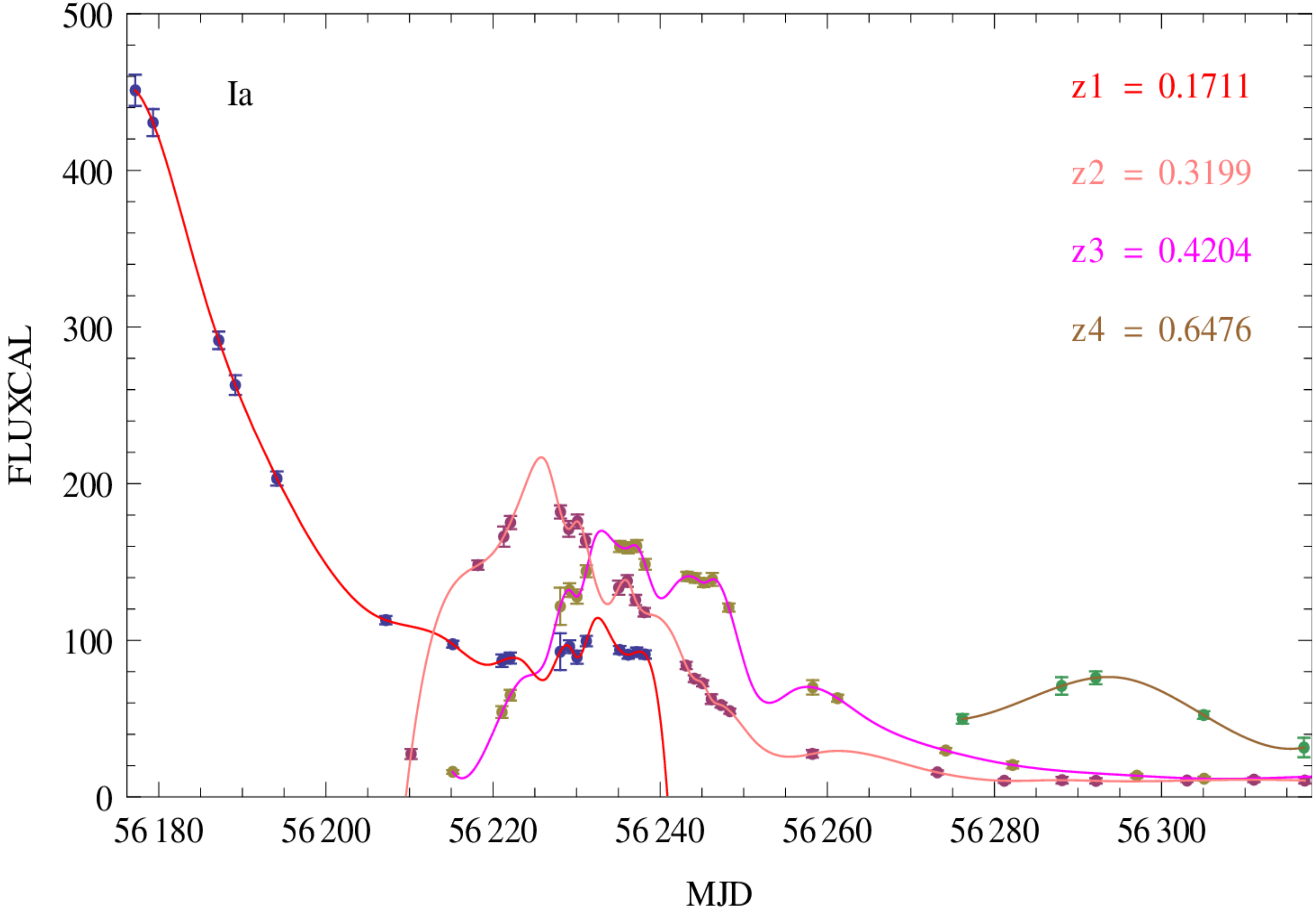
2. The Sne photometric classification problem



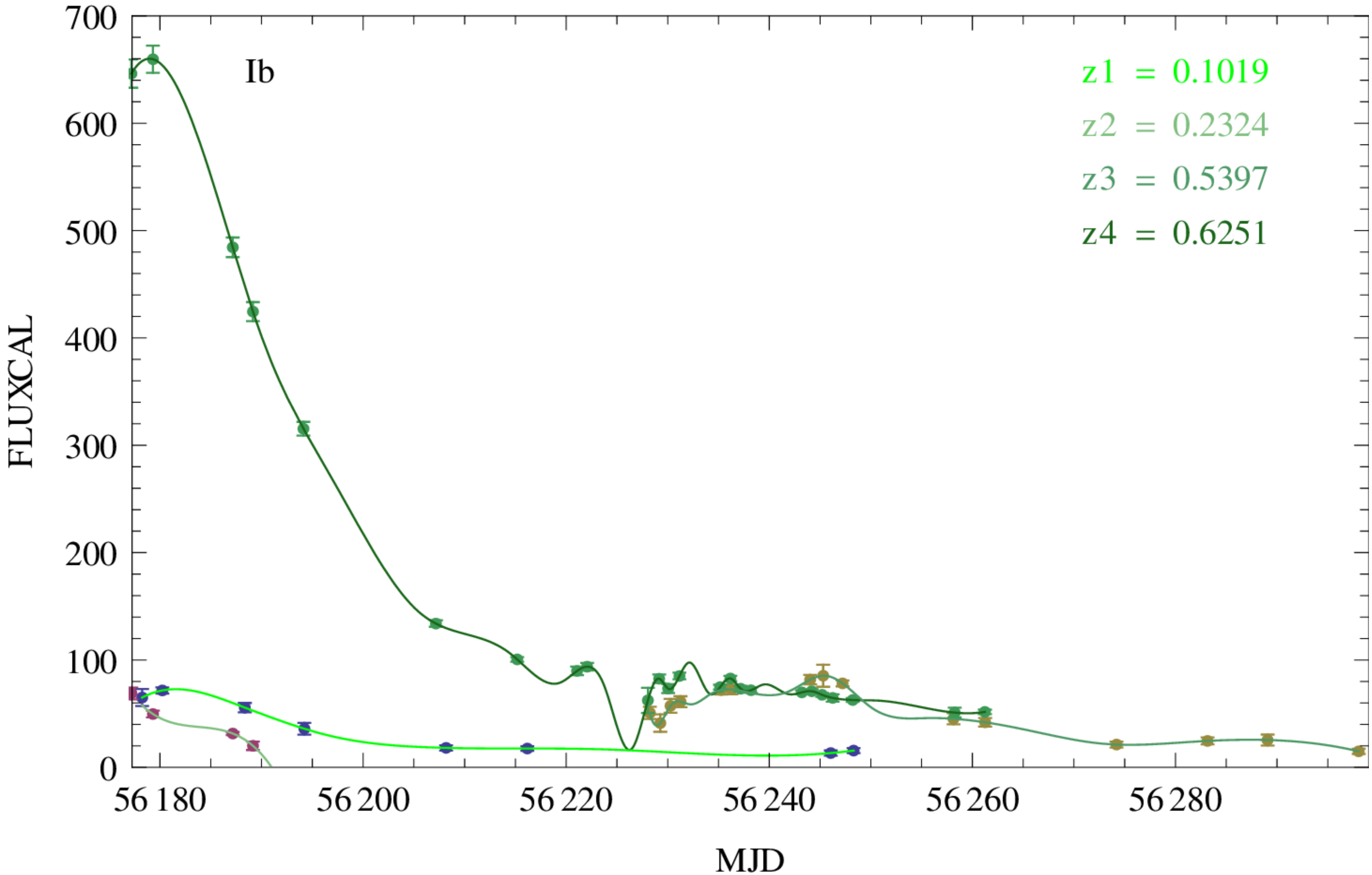
2. The Sne photometric classification problem



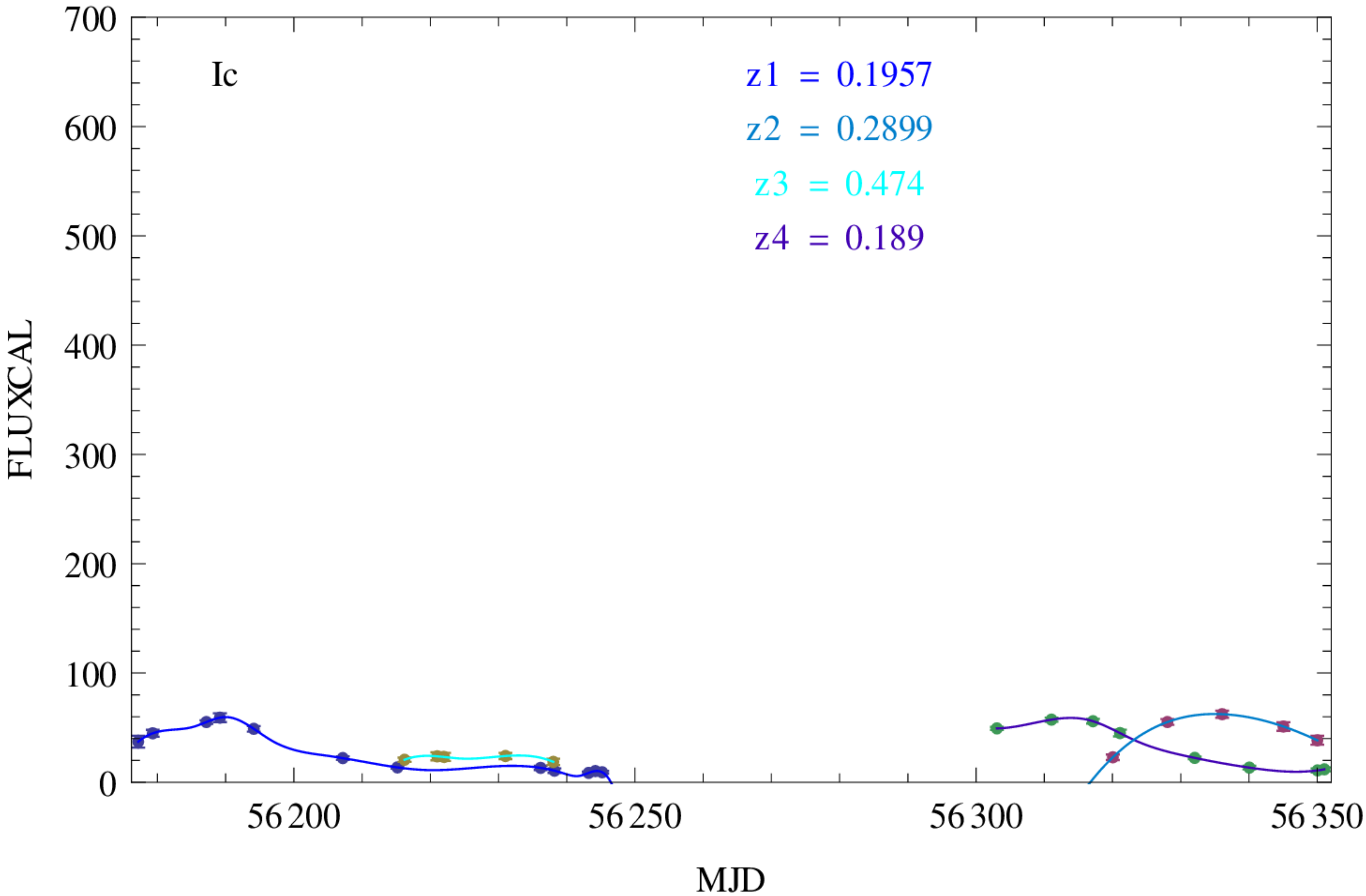
2. The Sne photometric classification problem



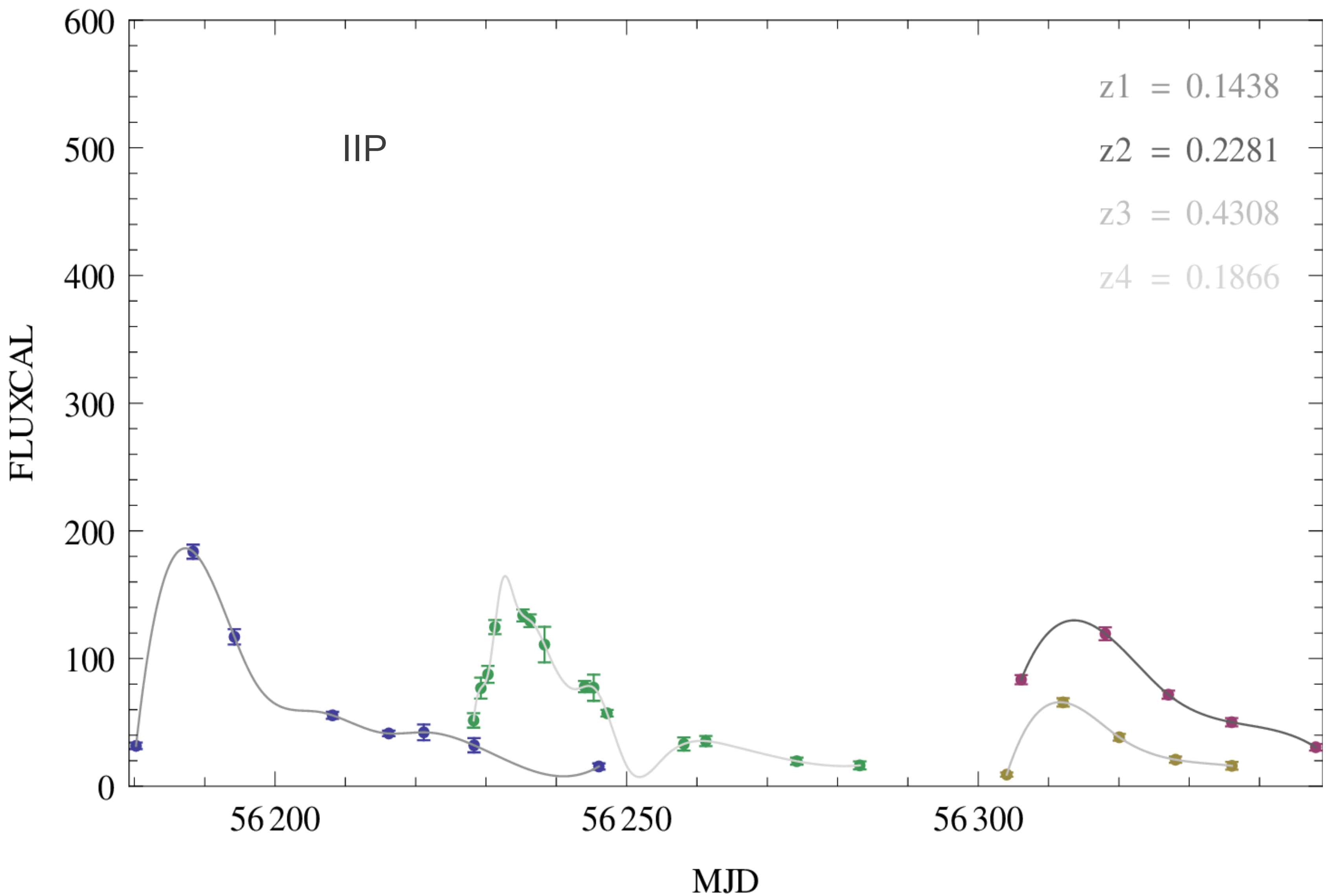
2. The Sne photometric classification problem



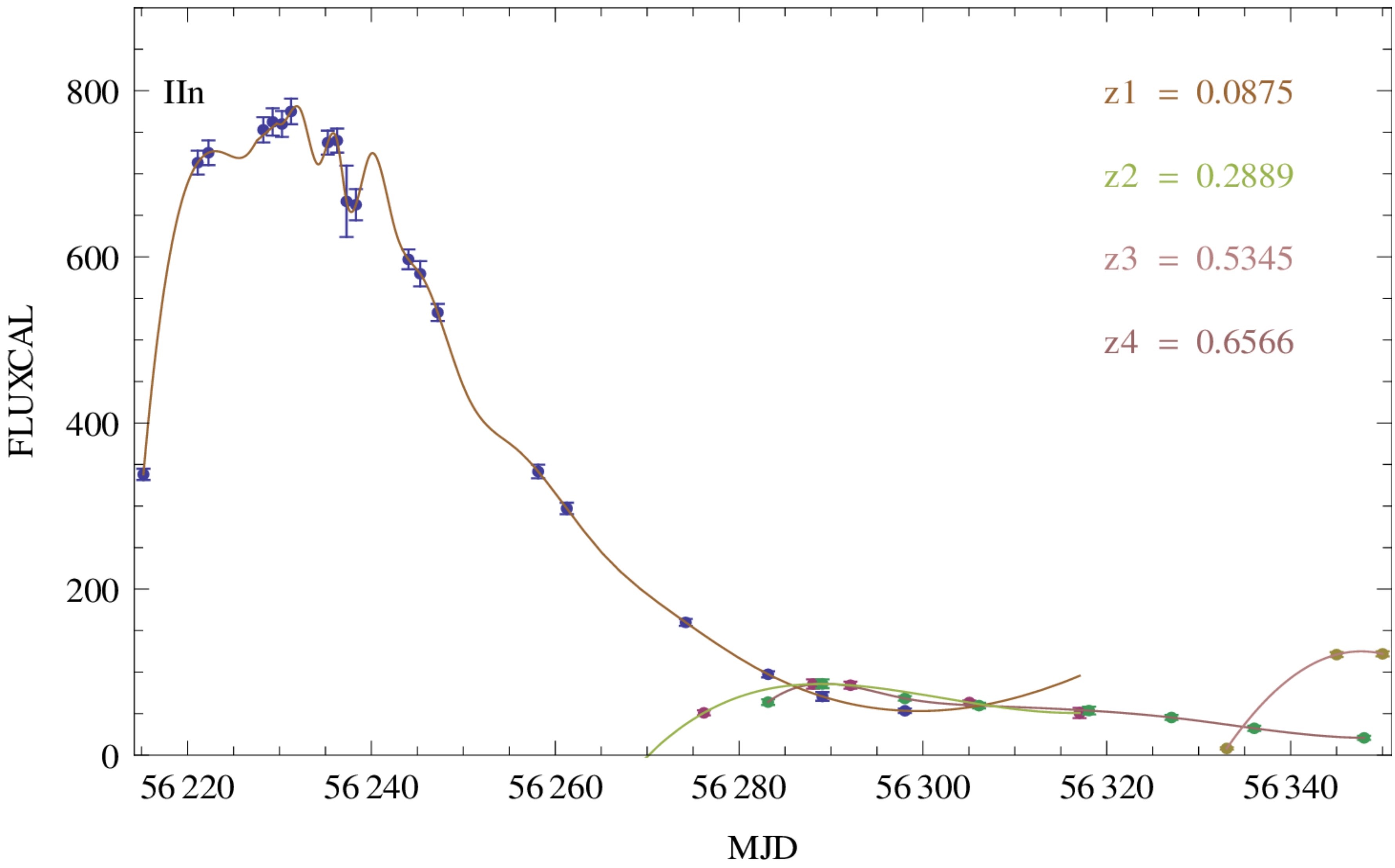
2. The Sne photometric classification problem



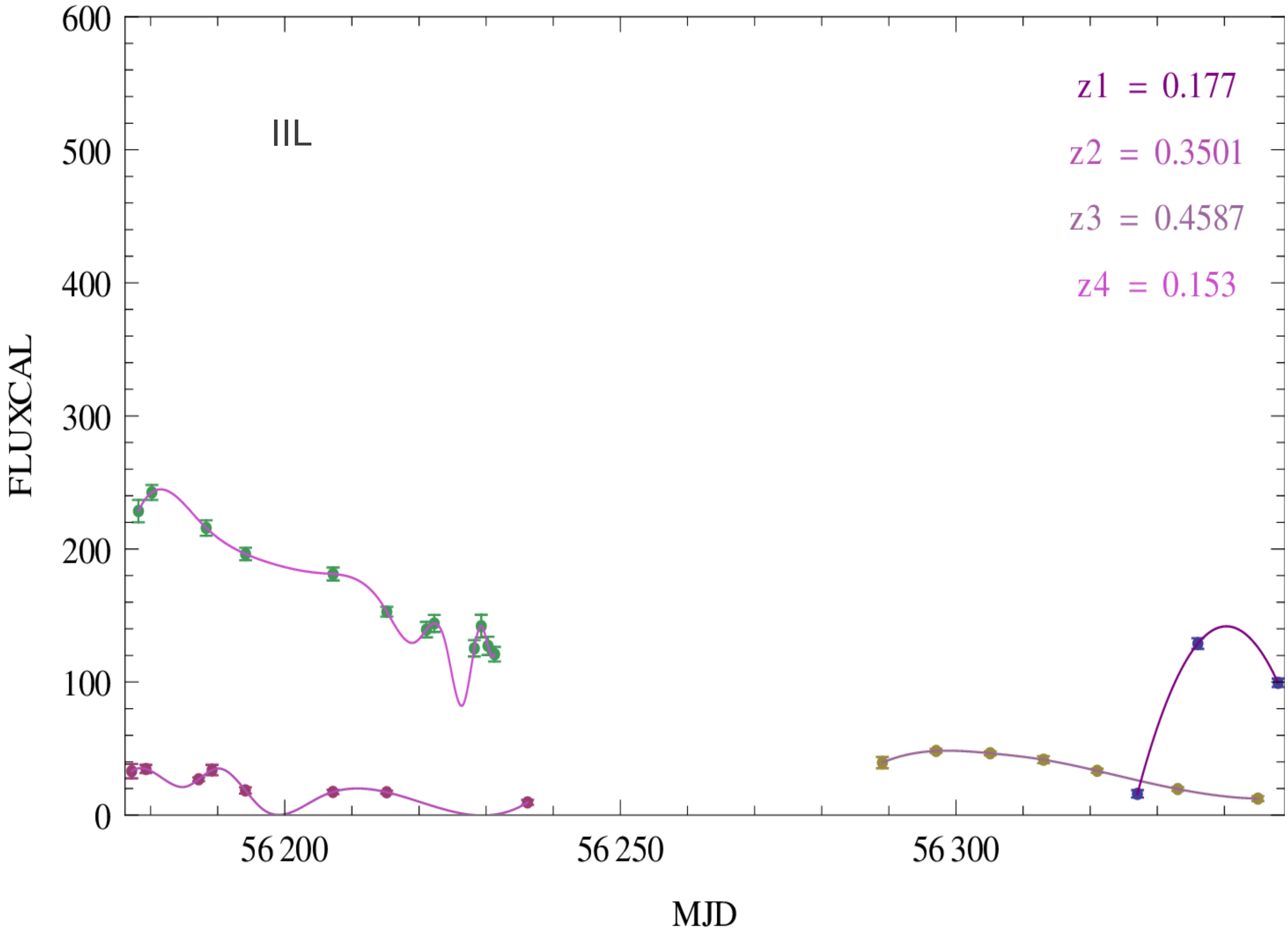
2. The Sne photometric classification problem



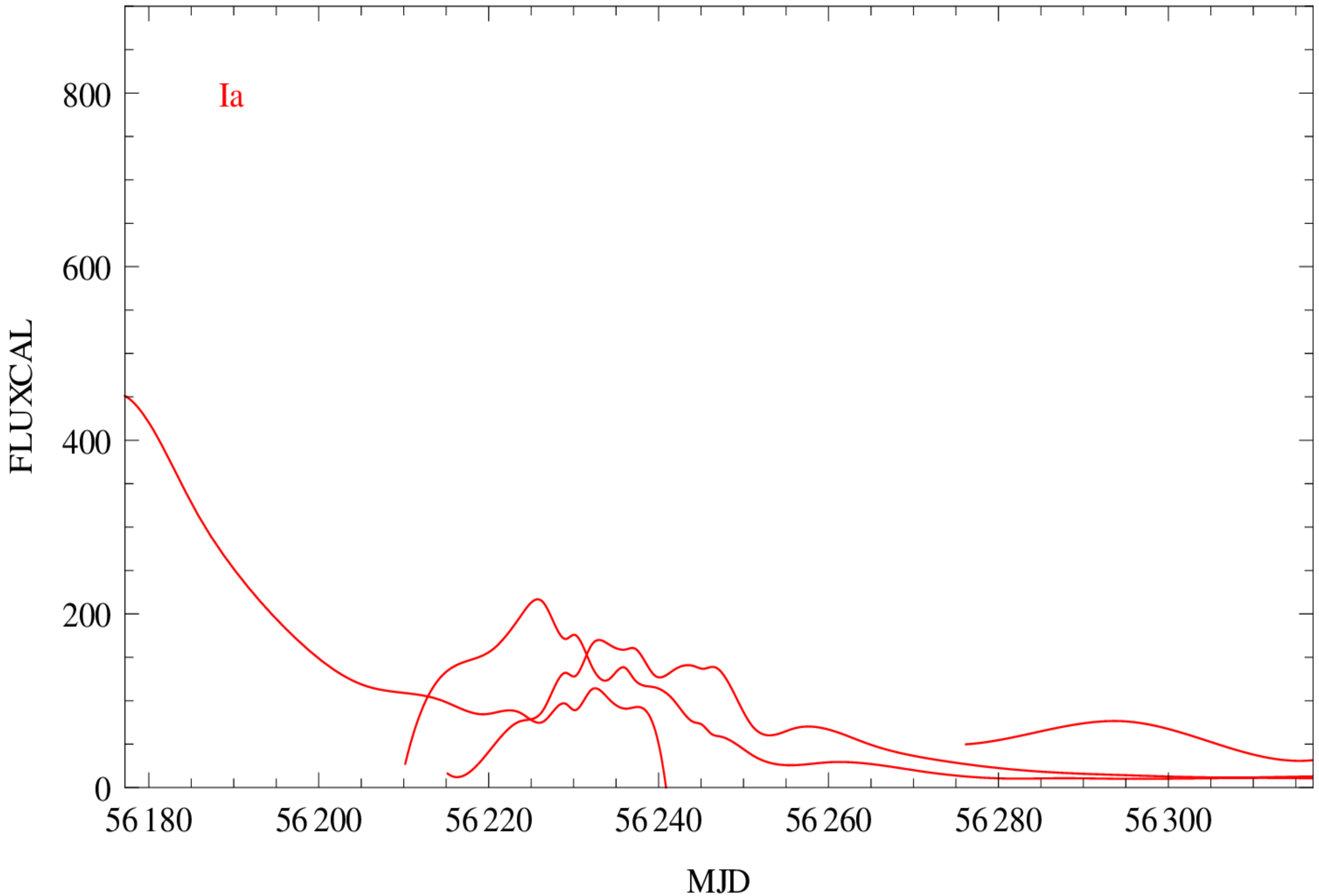
2. The Sne photometric classification problem



2. The Sne photometric classification problem

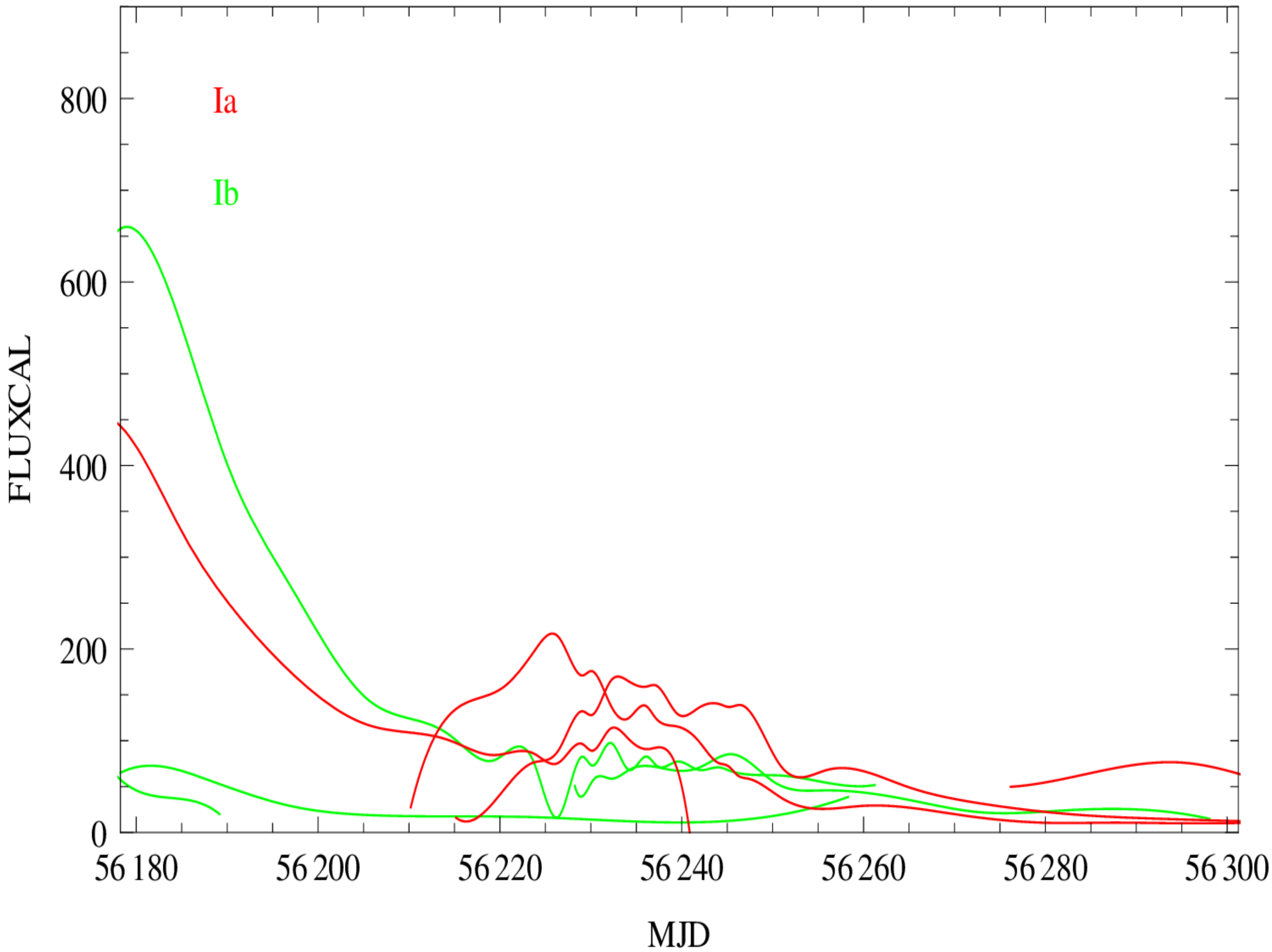


2. The Sne photometric classification problem

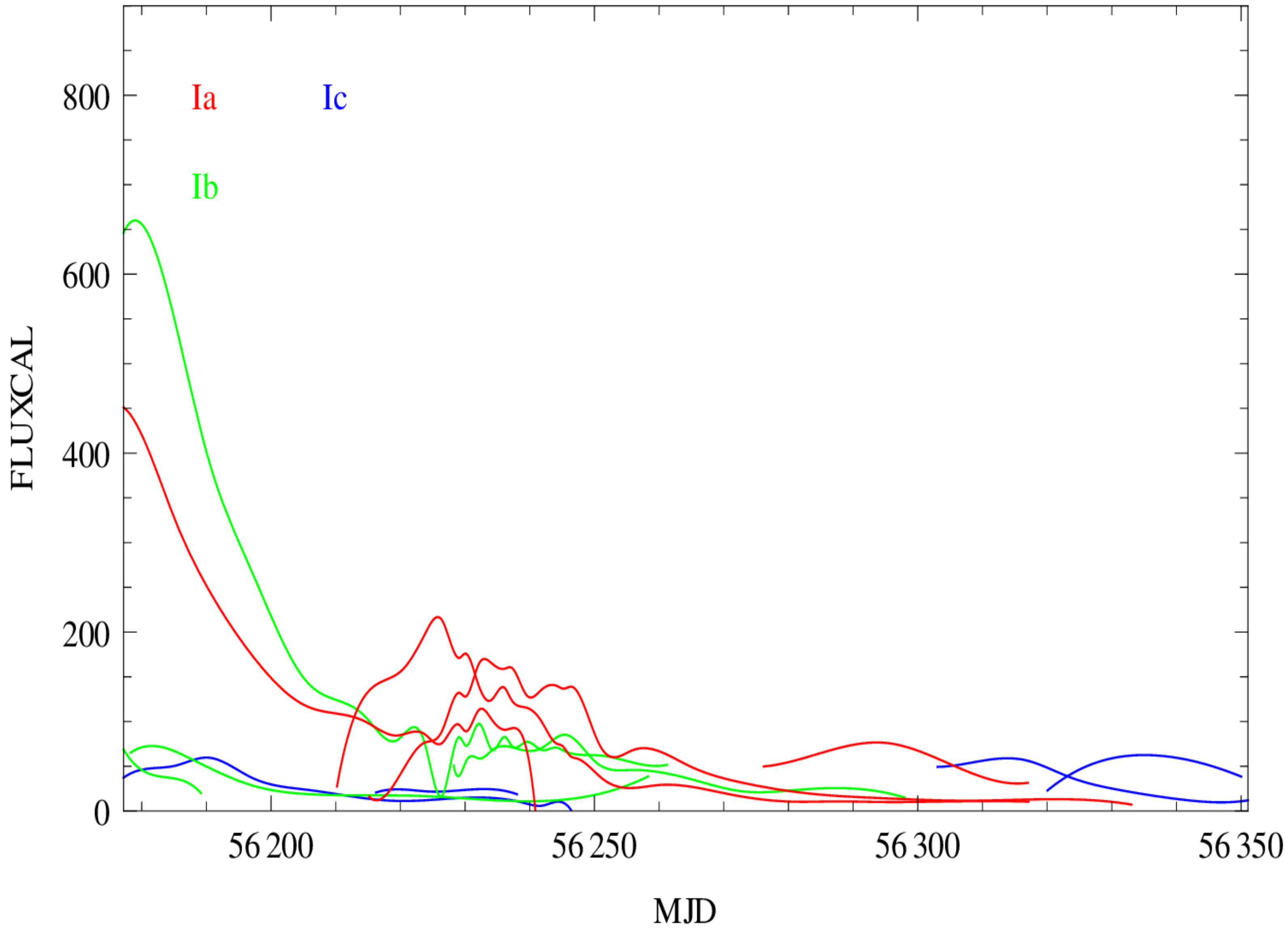


FAPESP

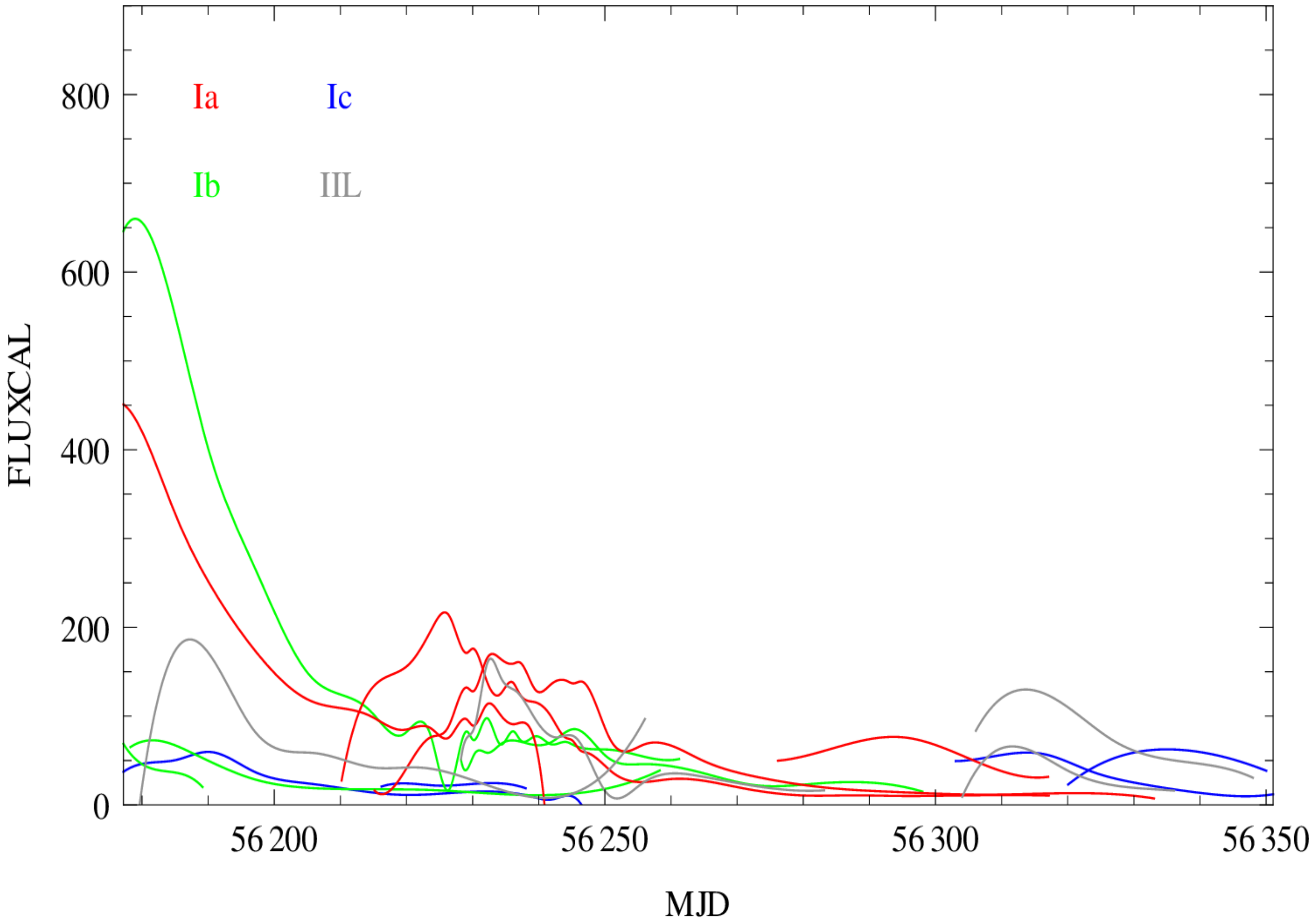
2. The Sne photometric classification problem



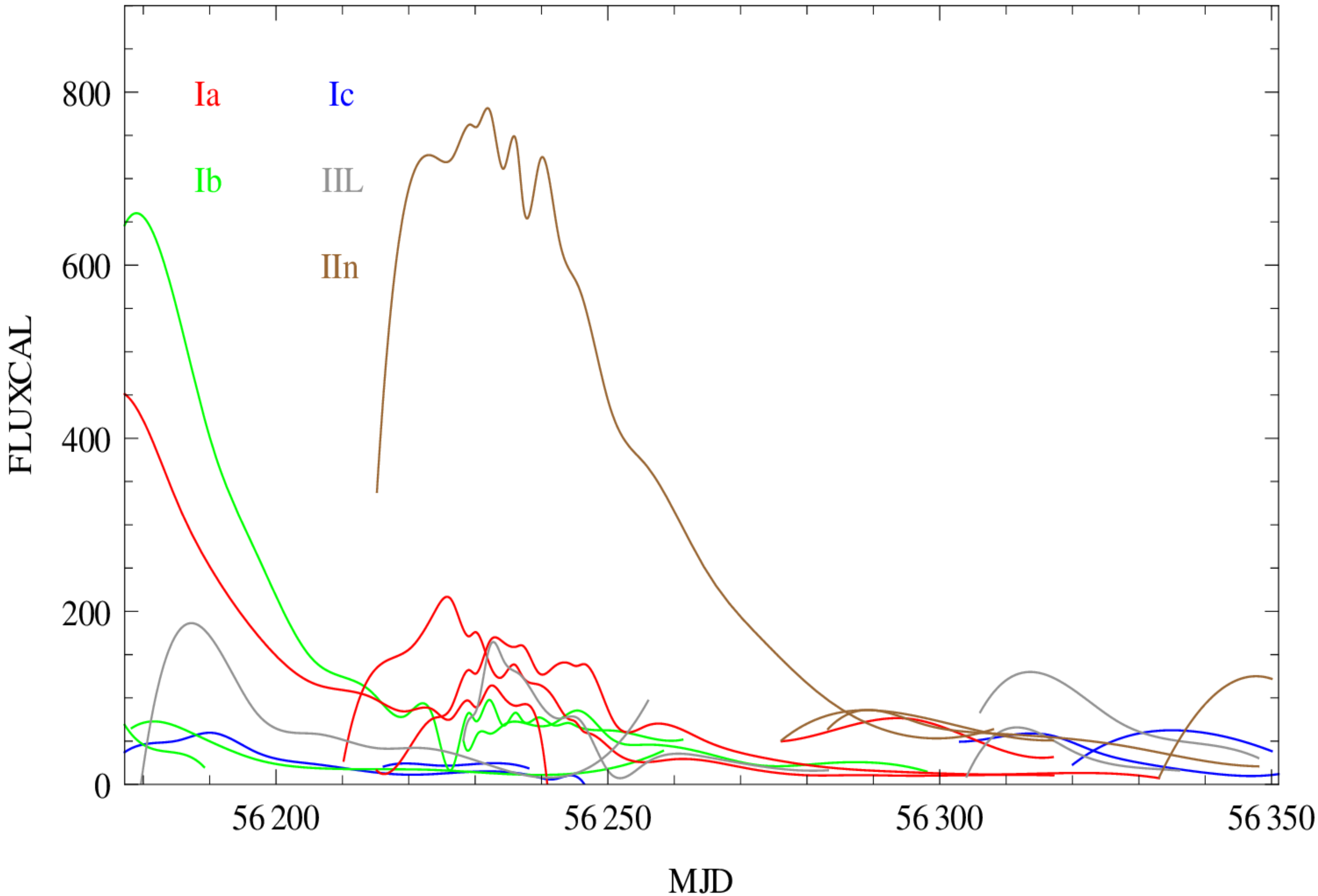
2. The Sne photometric classification problem



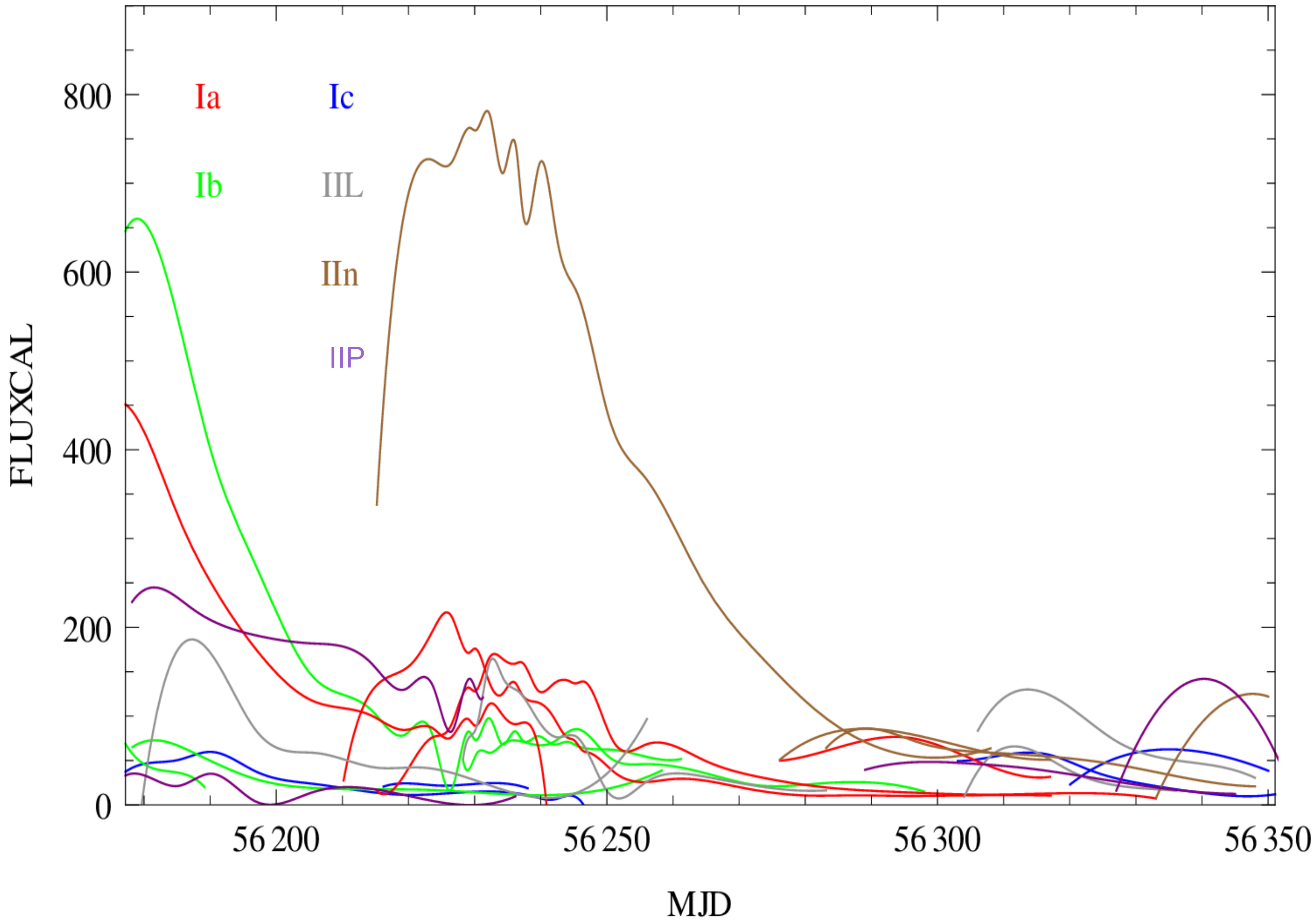
2. The Sne photometric classification problem



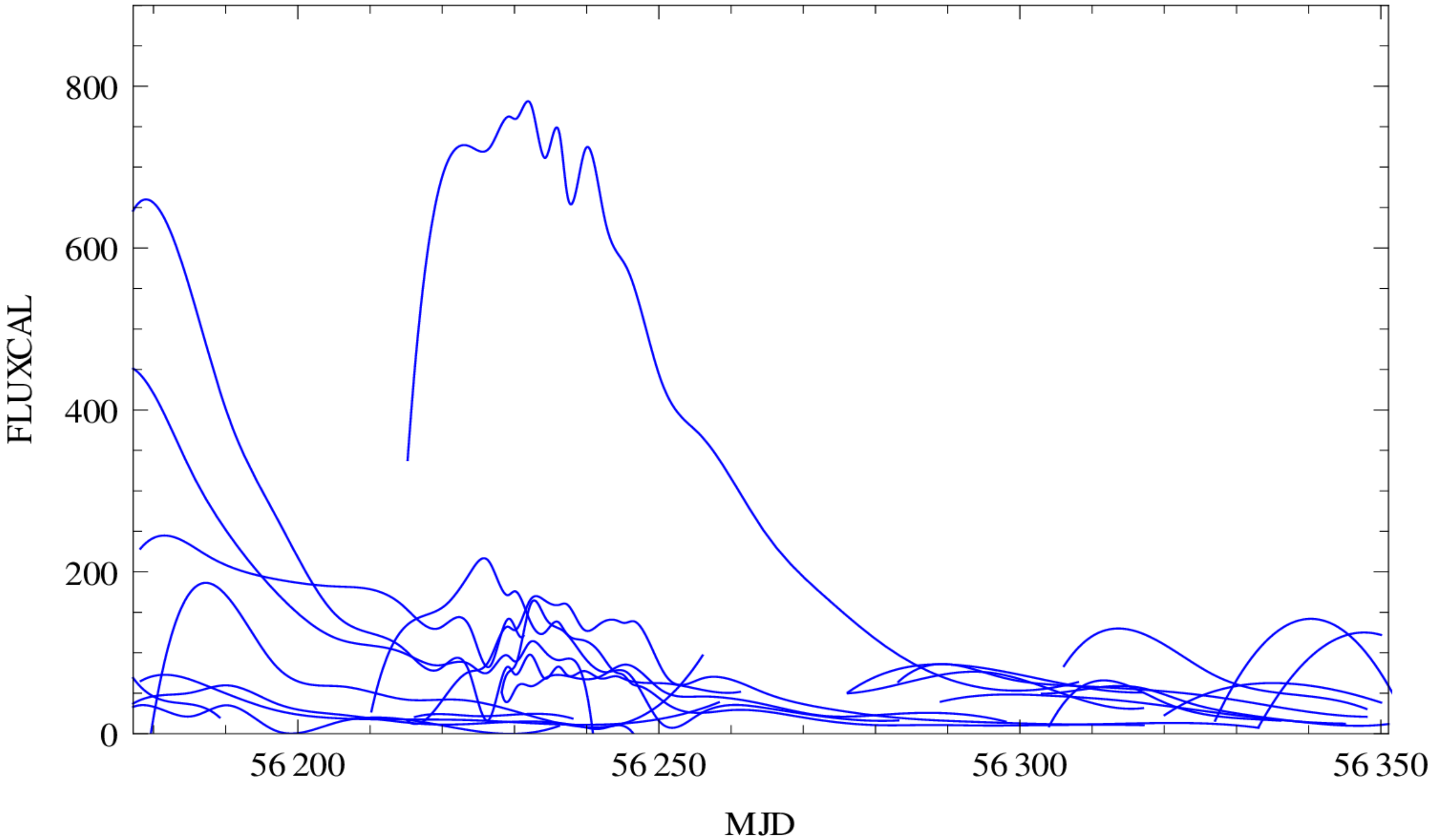
2. The Sne photometric classification problem



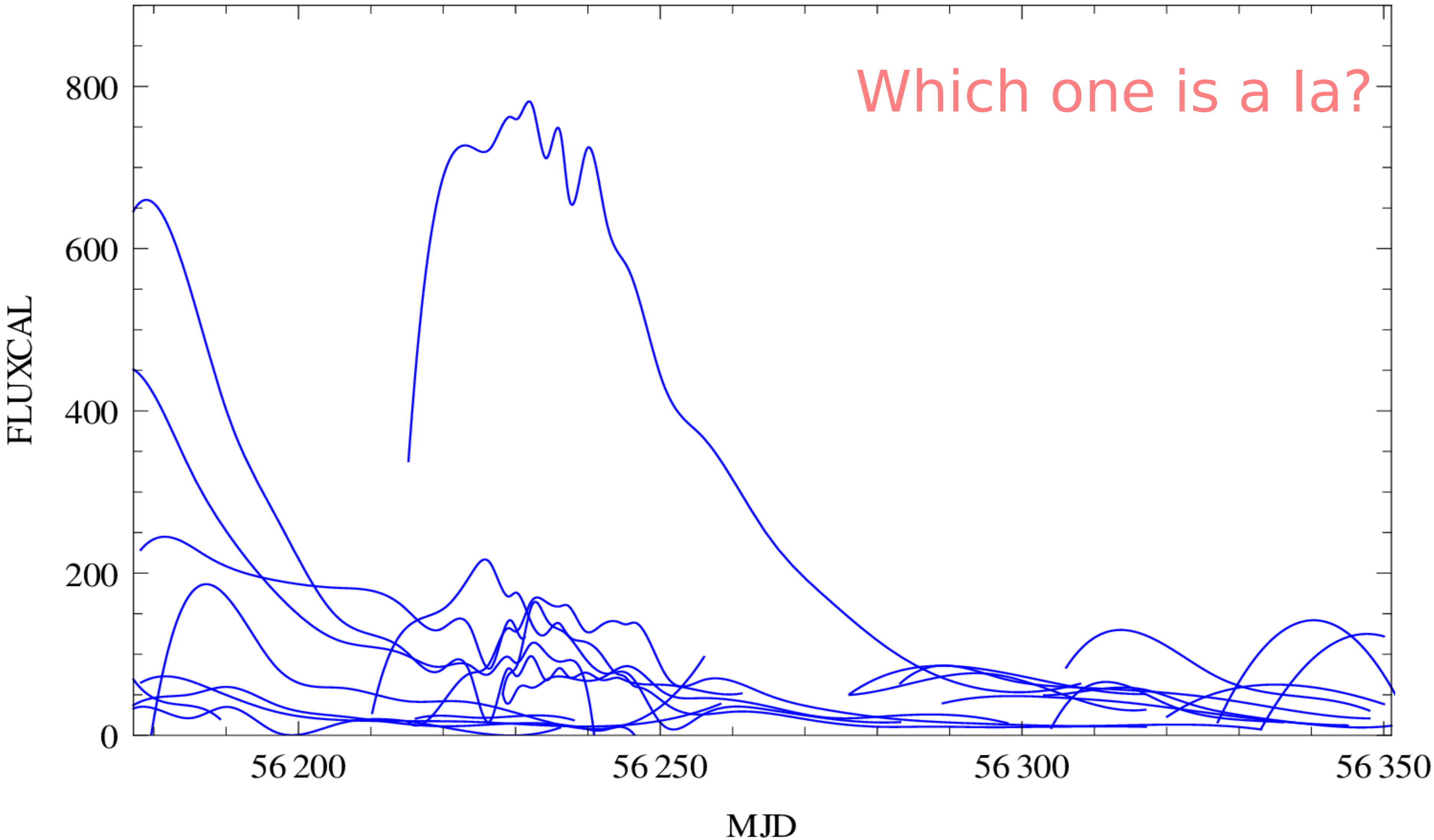
2. The Sne photometric classification problem



2. The Sne photometric classification problem



2. The Sne photometric classification problem



Our attempt:

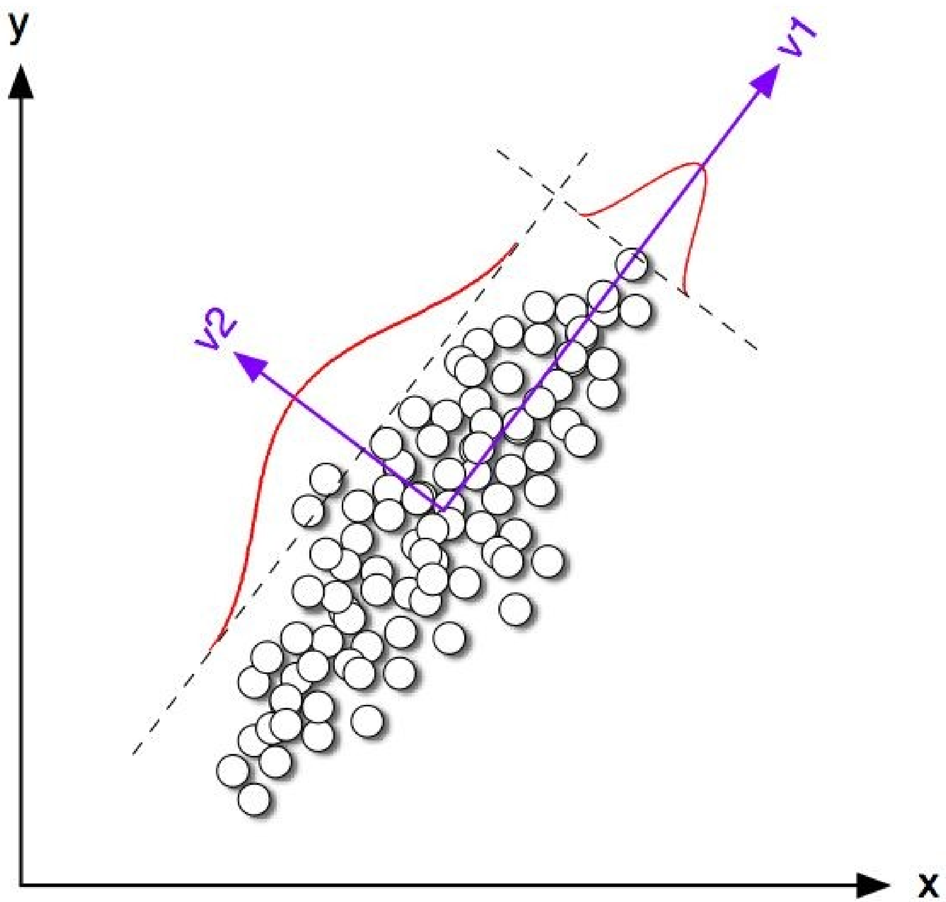
Kernel Principal Component Analysis (kPCA)

+

K=1 nearest neighbour

4. Principal Component Analysis (PCA)

The main goal of PCA is to reduce the dimensionality of the initial parameter space



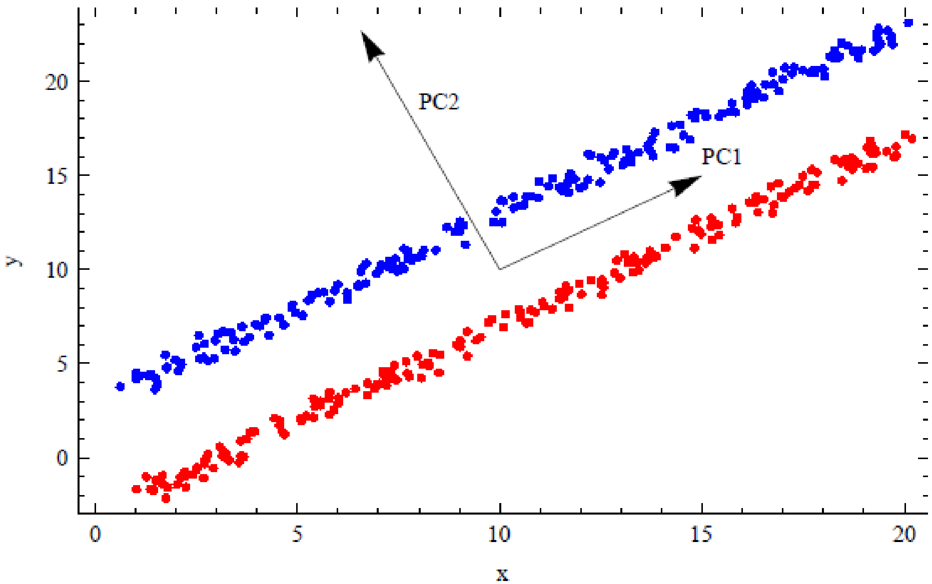
It looks for directions that maximize the variance of data points (information)

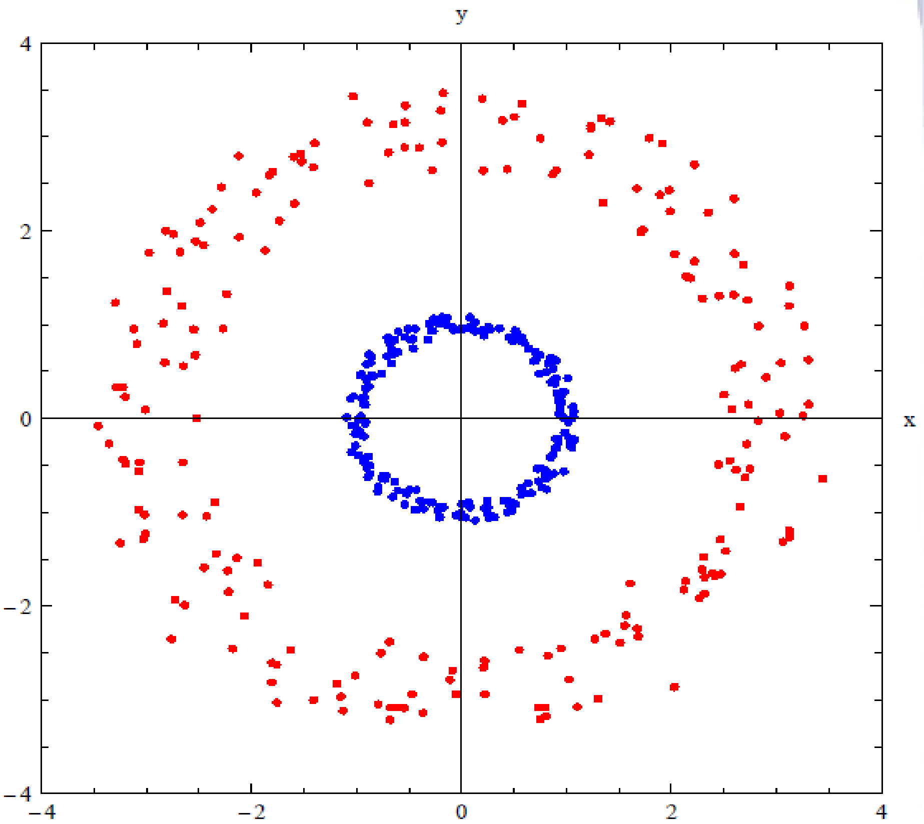
http://www.cs.cornell.edu/courses/cs322/2008sp/images/thumb_PCA.png

4. Principal Component Analysis (PCA)

PCA - limitations

Is not designed to capture non-linear structure

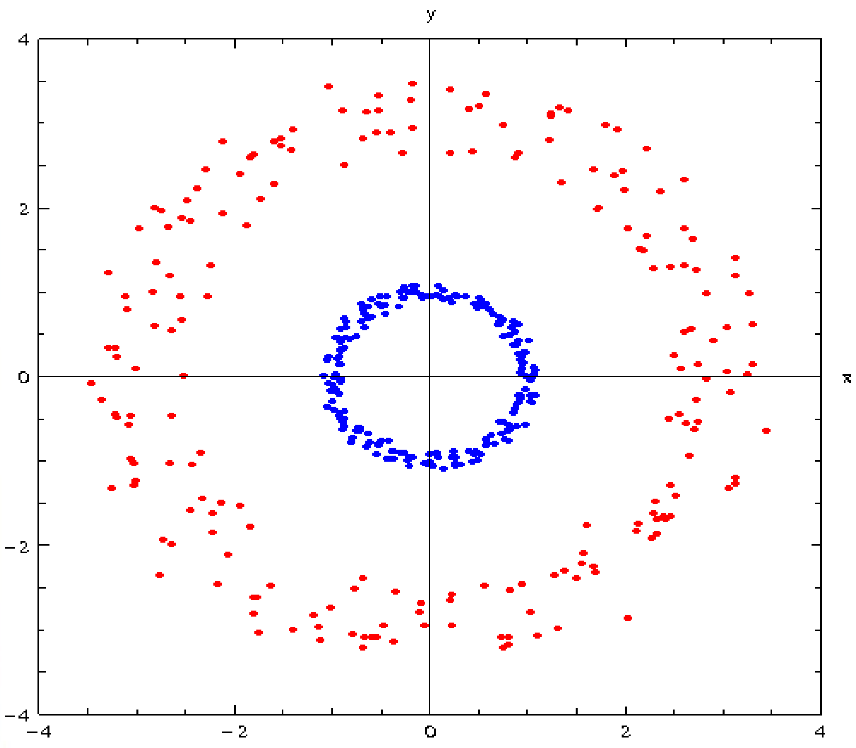




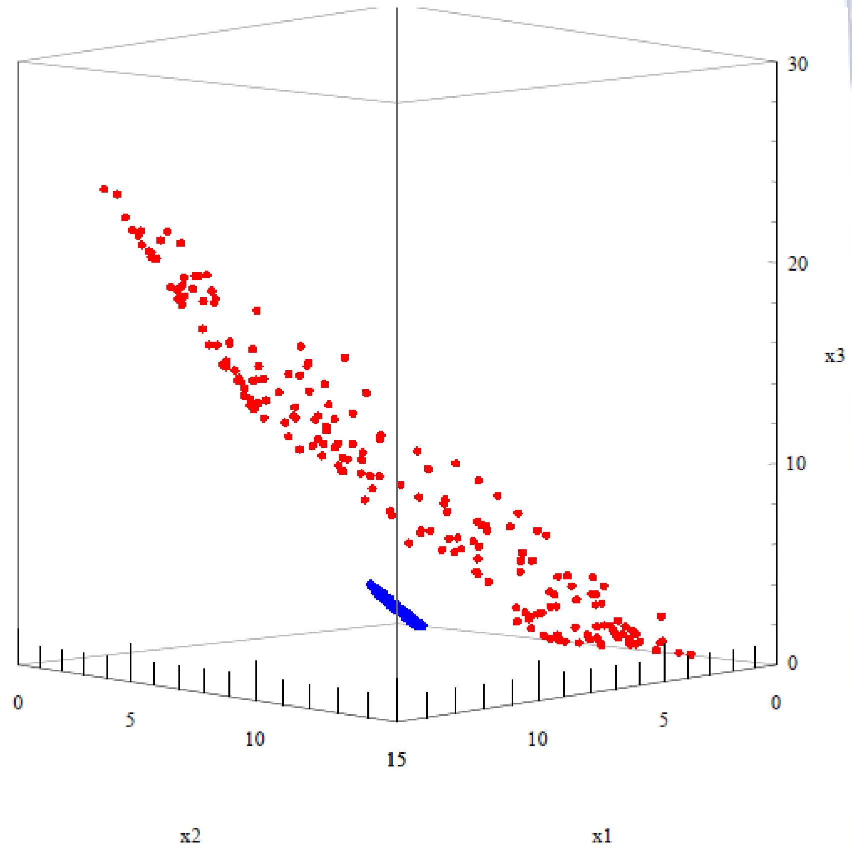
Does not care about labels

4. Kernel Principal Component Analysis (kPCA)

PCA - extensions



$$(x, y) \quad \rightarrow \quad z = \sqrt{x^2 + y^2}$$



Sometimes, going to higher dimensions might solve the problem

5. Kernel Principal Component Analysis (kPCA)

In the linear case, with

$\boldsymbol{x}_i \quad \rightarrow i-th$ data vector,

$$K_{ij} = \boldsymbol{x}_i^T \boldsymbol{x}_j$$

$\alpha_k \rightarrow k-th$ eigenvalue

$\boldsymbol{v_k} \rightarrow k-th$ eigenvector

$$\boldsymbol{v}_k^T \boldsymbol{n} = \sum_{i=1}^{N} \alpha_k K(\boldsymbol{x_i}, \boldsymbol{n})$$

where $\boldsymbol{n}$ is the data to be projected and

$$K(\boldsymbol{x}_i, \boldsymbol{n}) = \boldsymbol{x}_i^T \boldsymbol{n}$$

The kernel trick

Substitute every dot product: $\boldsymbol{x}_i^T \boldsymbol{x}_j$

by: $\Phi(\boldsymbol{x}_i) \cdot \Phi(\boldsymbol{x}_j)$

where: $\Phi : \mathbb{R} \rightarrow \mathbb{F}$

$\boldsymbol{x} \rightarrow \Phi(\boldsymbol{x})$

Gaussian kernel:

$$k(\mathrm{x}_i, \mathrm{x}_j) = \exp\left[\frac{\|\mathrm{x}_i - \mathrm{x}_j\|^2}{2\sigma^2}\right]$$

Distances between data points in higher dimensional space

5. Kernel Principal Component Analysis (kPCA)

Message to take home:

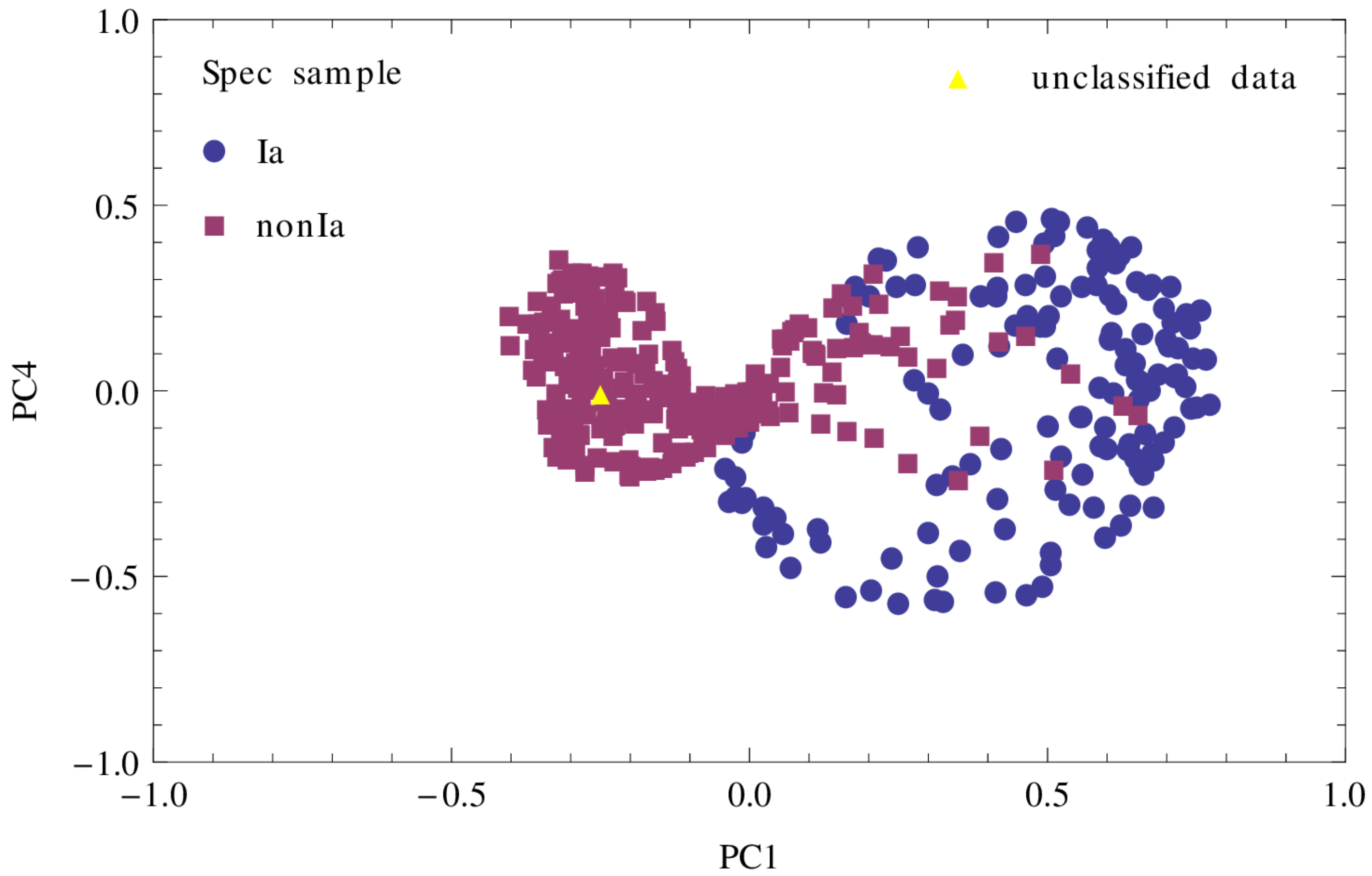
It is not necessary to know the mapping from the initial to the higher dimensional parameter space

5. kPCA

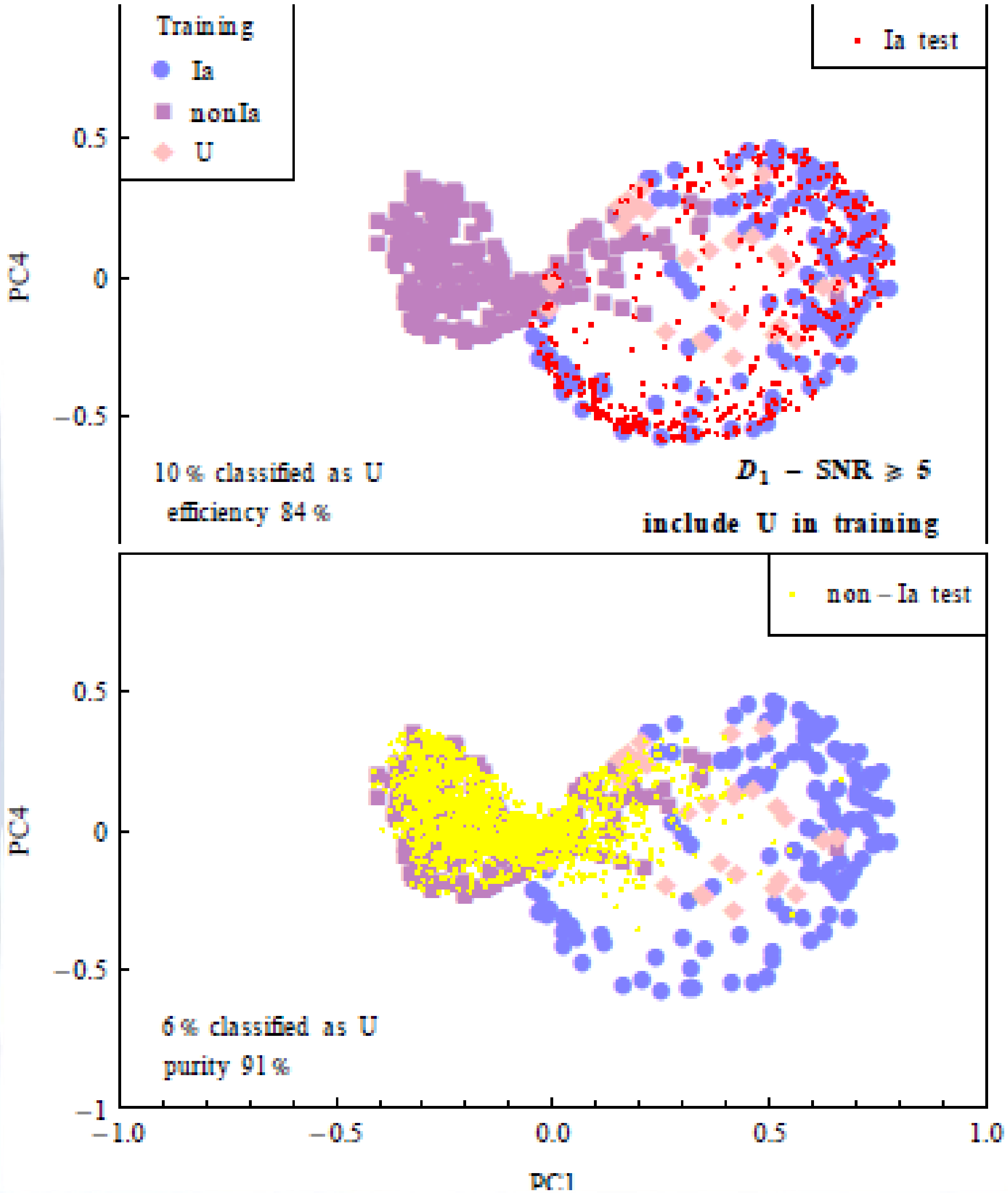
Diagonalize K

K=1
Nearest neighbor

Calculate projections



5. kPCA - results



Results

Post-SNPCC synthetic data

After cuts

FoM ~0.60

SC ~91%

Selection cuts:
{-3,+24} in r-band
At least 3 obs with SNR>5
in each band

FAPESP

5. kPCA - results

Important application:

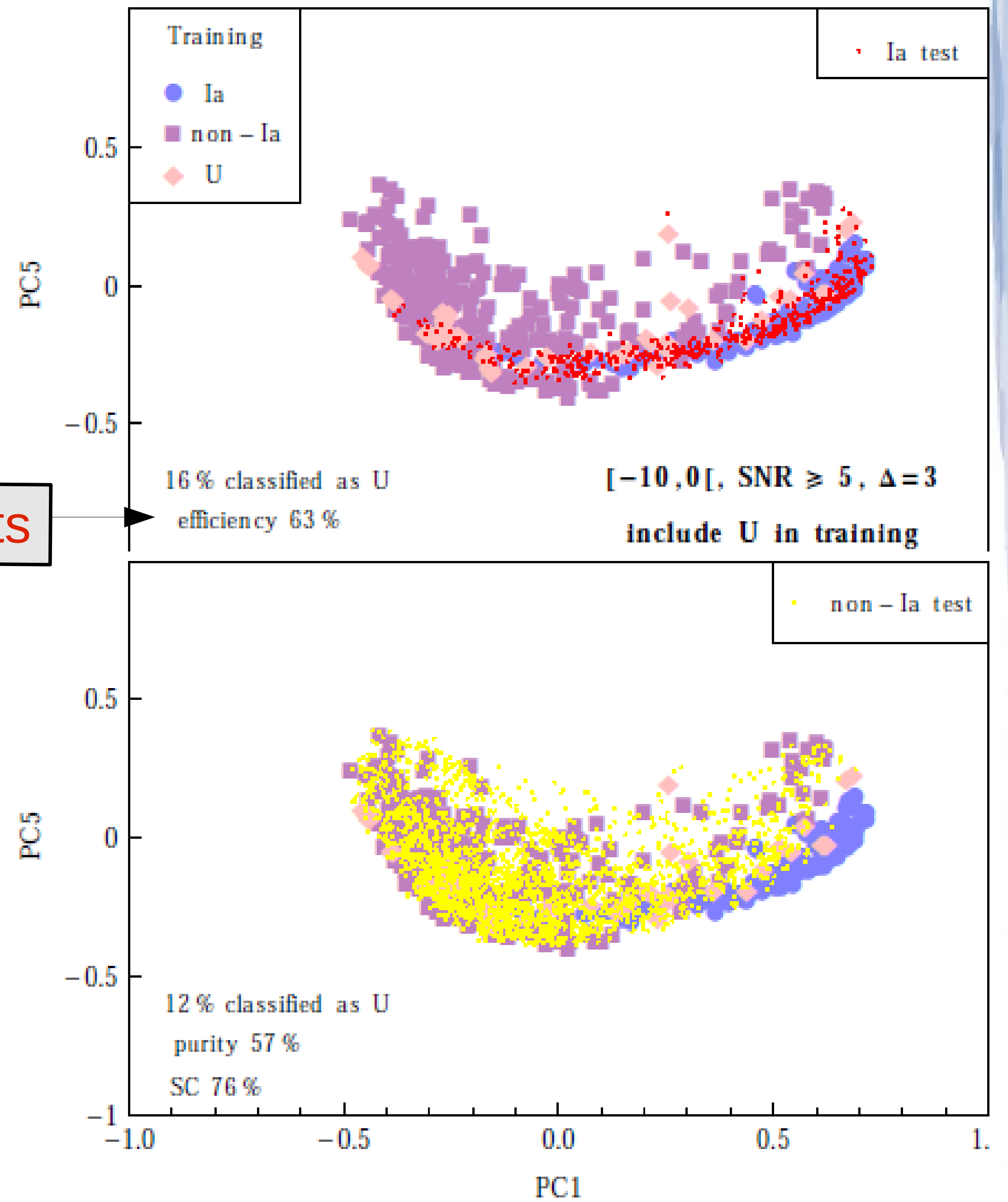
Use early epoch classification to guide spectroscopic follow-up resources

FAPESP

5. kPCA - results

Pre-maximum analysis

After cuts



5. kPCA - results

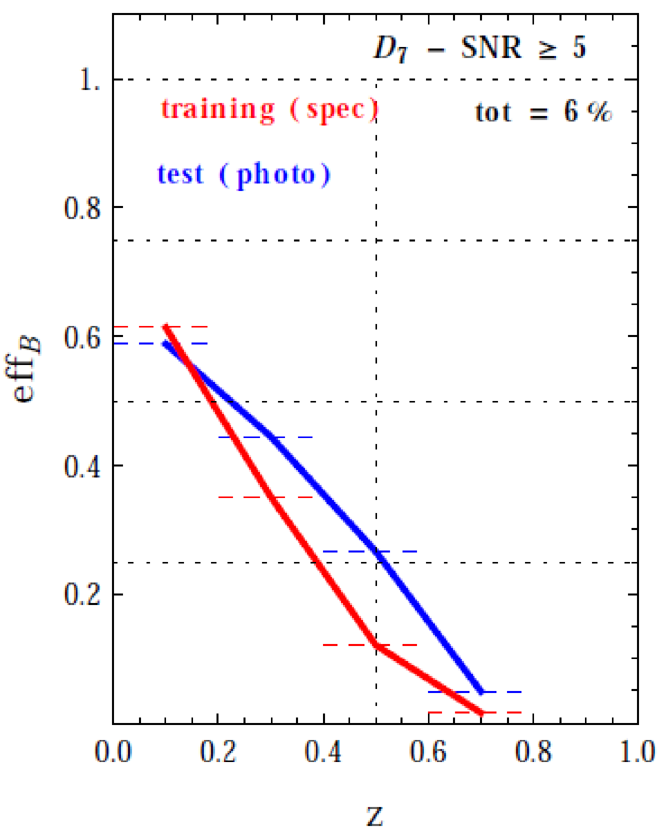
How these results compare with the others already in the literature?

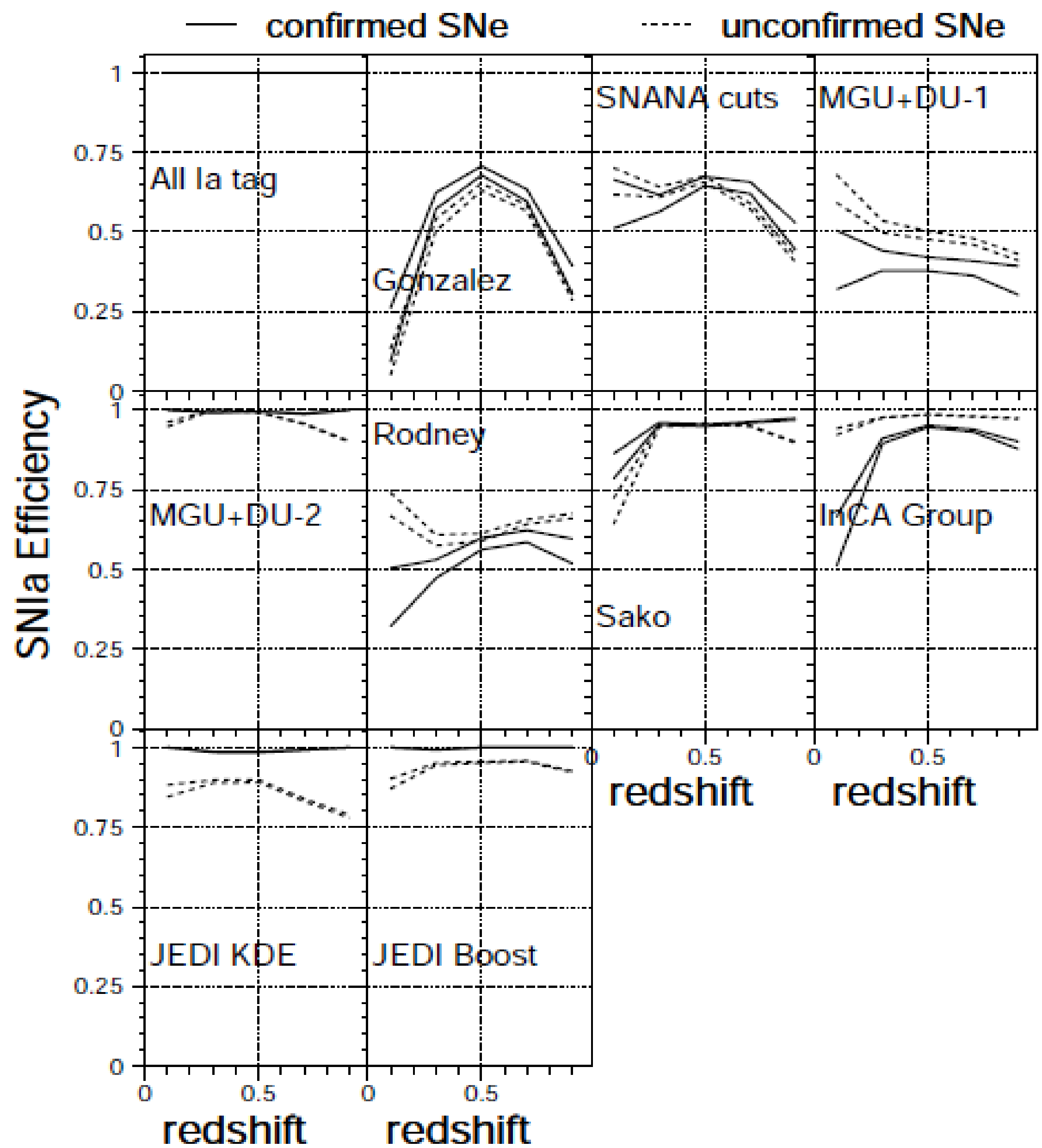
FAPESP

5. kPCA - results

Bad news: $\text{eff} = \frac{N_{\rm Ia}^{\rm SC}}{N_{Ia}^{\rm tot}}$

efficiency is not
good due to selection cuts





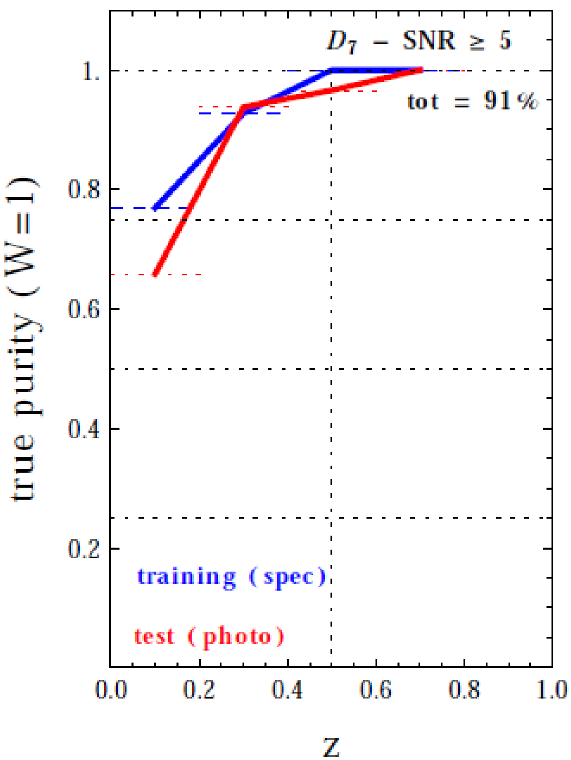
Results from the *Supernova Photometric Classification Challenge* (Kessler et al., 2010)

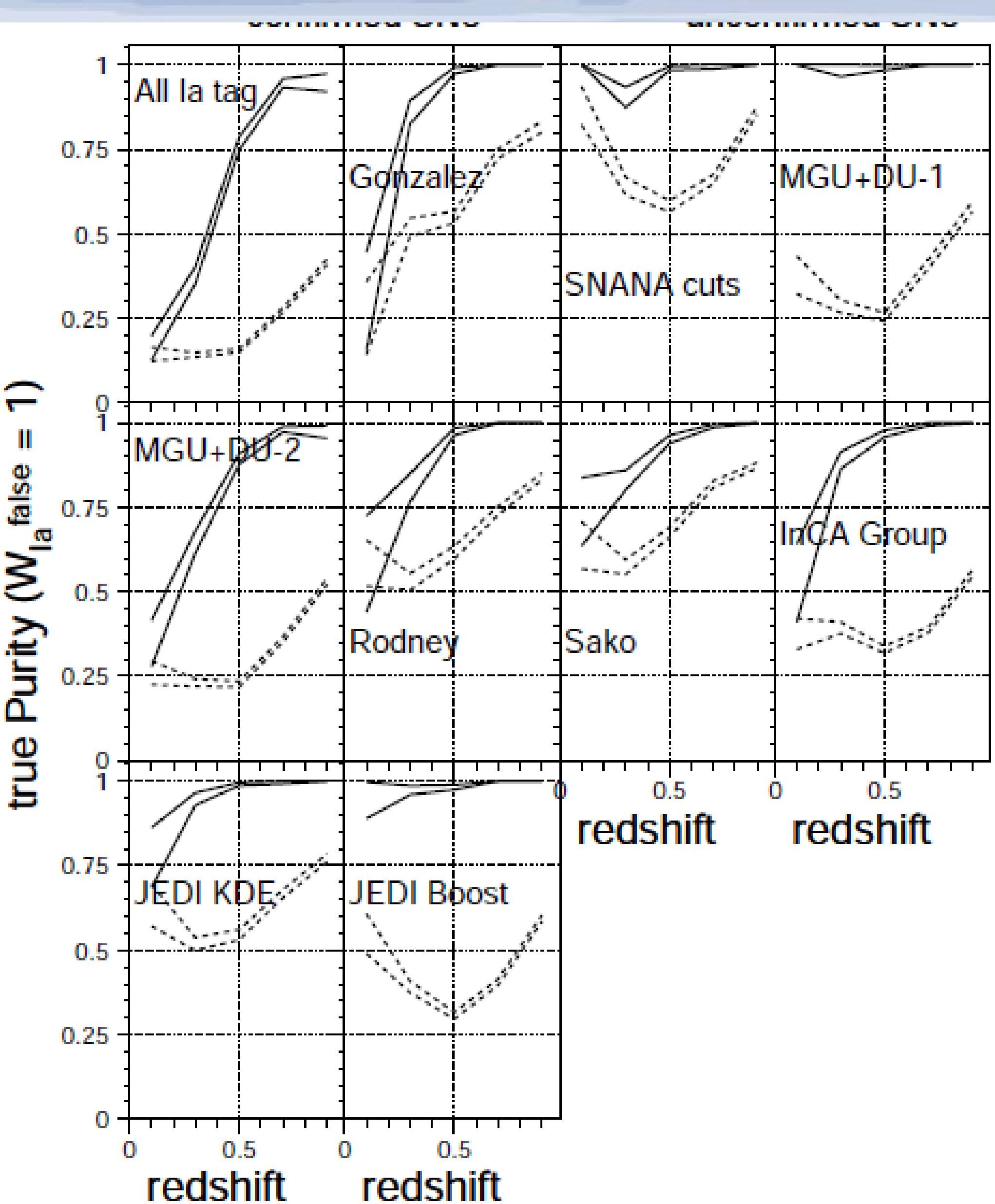
5. kPCA - results

Good news:

Purity is maximized!

$$\text{pur} = \frac{N_{\text{Ia}}^{\text{SC}}}{N_{\text{Ia}}^{\text{WC}} + N_{\text{Ia}}^{\text{SC}}}$$





Results from the *Supernova Photometric Classification Challenge* (Kessler et al., 2010)

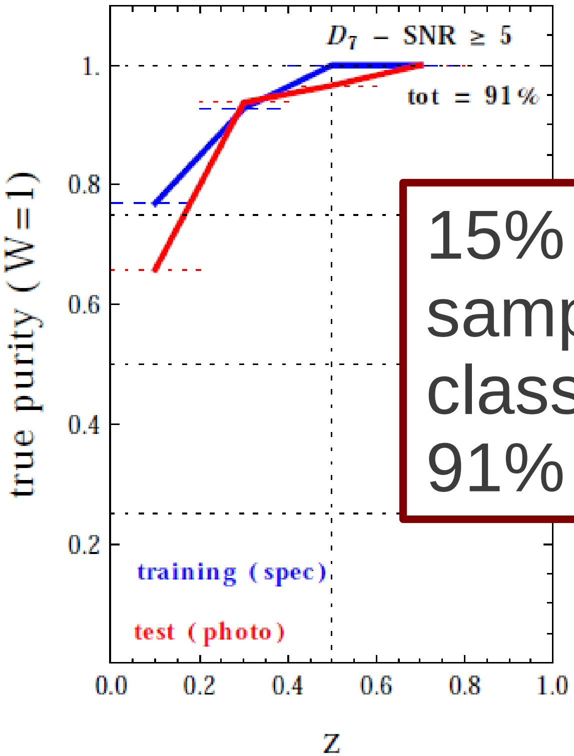
FAPESP

5. kPCA - results

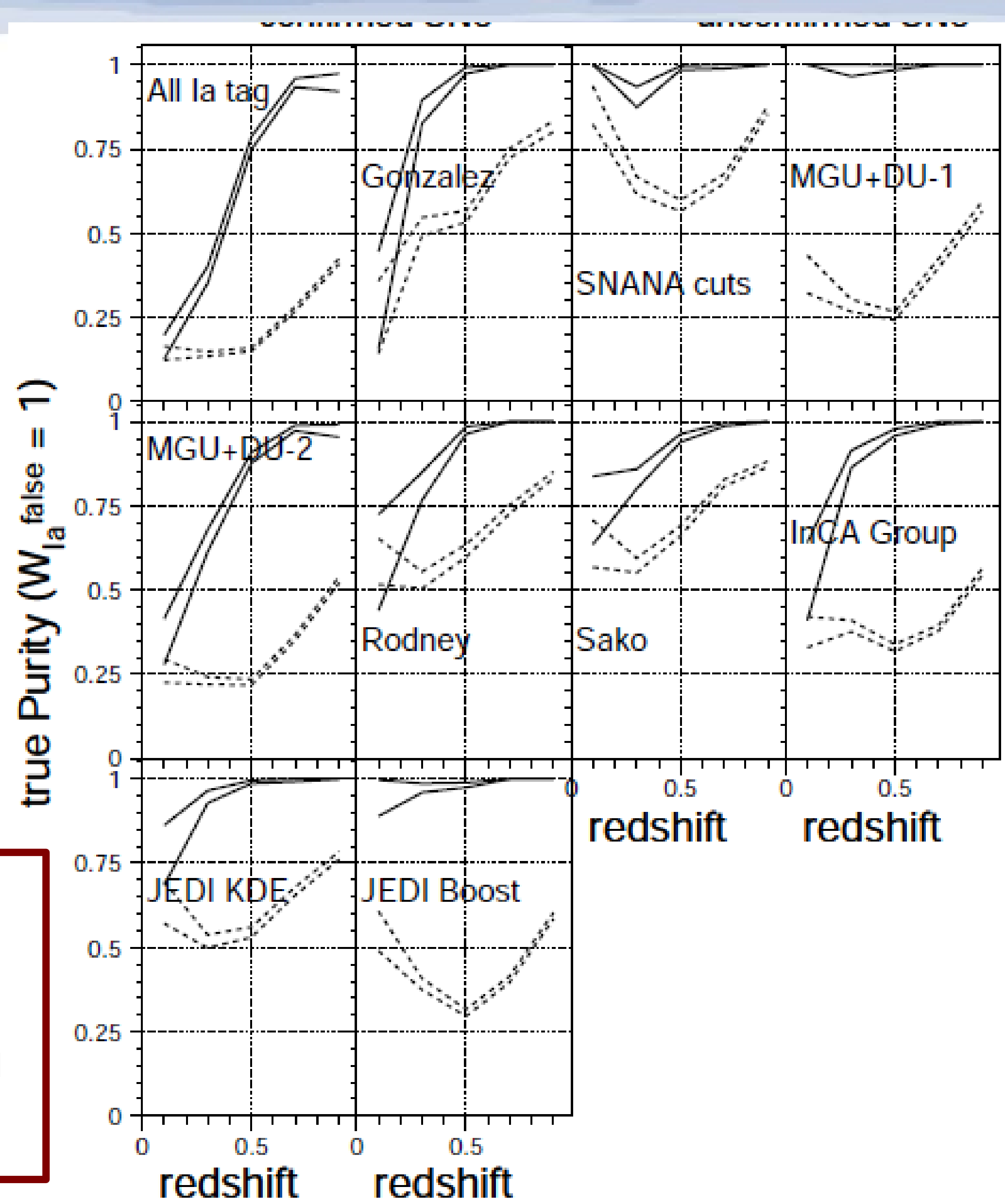
Good news:

Purity is maximized!

$$\mathrm{pur} = \frac{N_{\mathrm{Ia}}^{\mathrm{SC}}}{N_{\mathrm{Ia}}^{\mathrm{WC}} + N_{\mathrm{Ia}}^{\mathrm{SC}}}$$



15% original sample classified with 91% purity



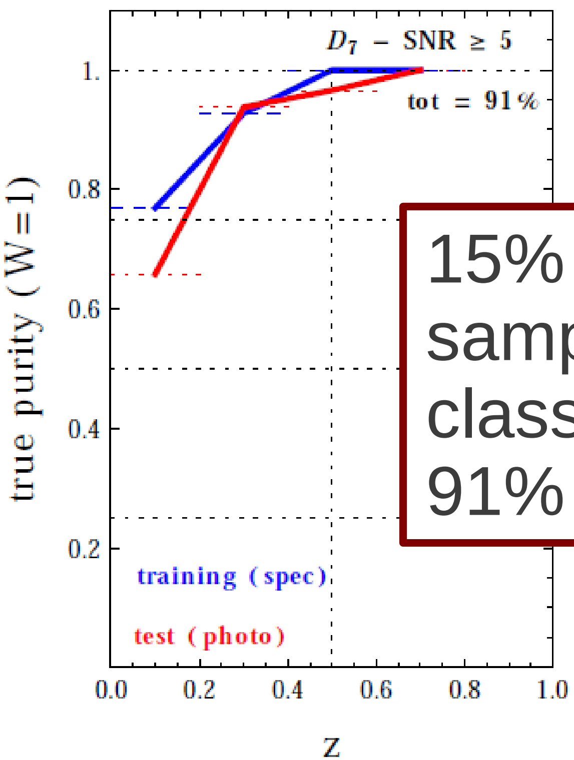
Results from the *Supernova Photometric Classification Challenge* (Kessler et al., 2010)

5. kPCA - results

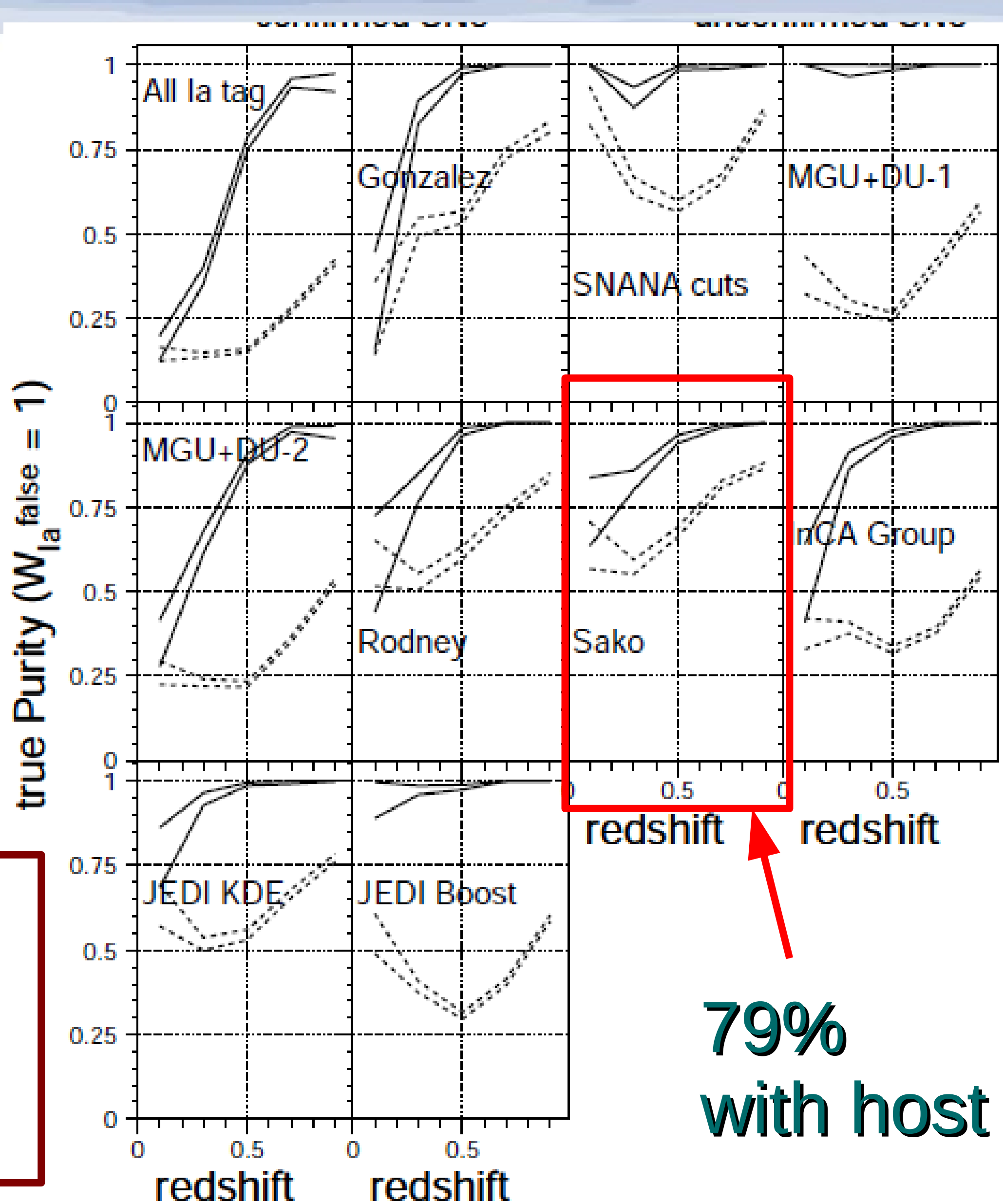
Good news:

Purity is maximized!

$$\text{pur} = \frac{N_{\text{Ia}}^{\text{SC}}}{N_{\text{Ia}}^{\text{WC}} + N_{\text{Ia}}^{\text{SC}}}$$



15% original sample classified with 91% purity



79%
with host z

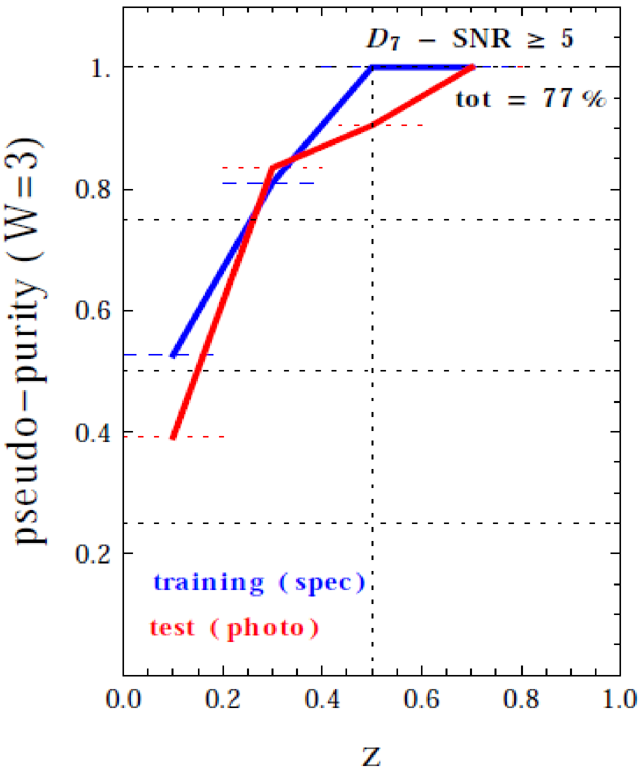
Results from the *Supernova Photometric Classification Challenge* (Kessler et al., 2010)

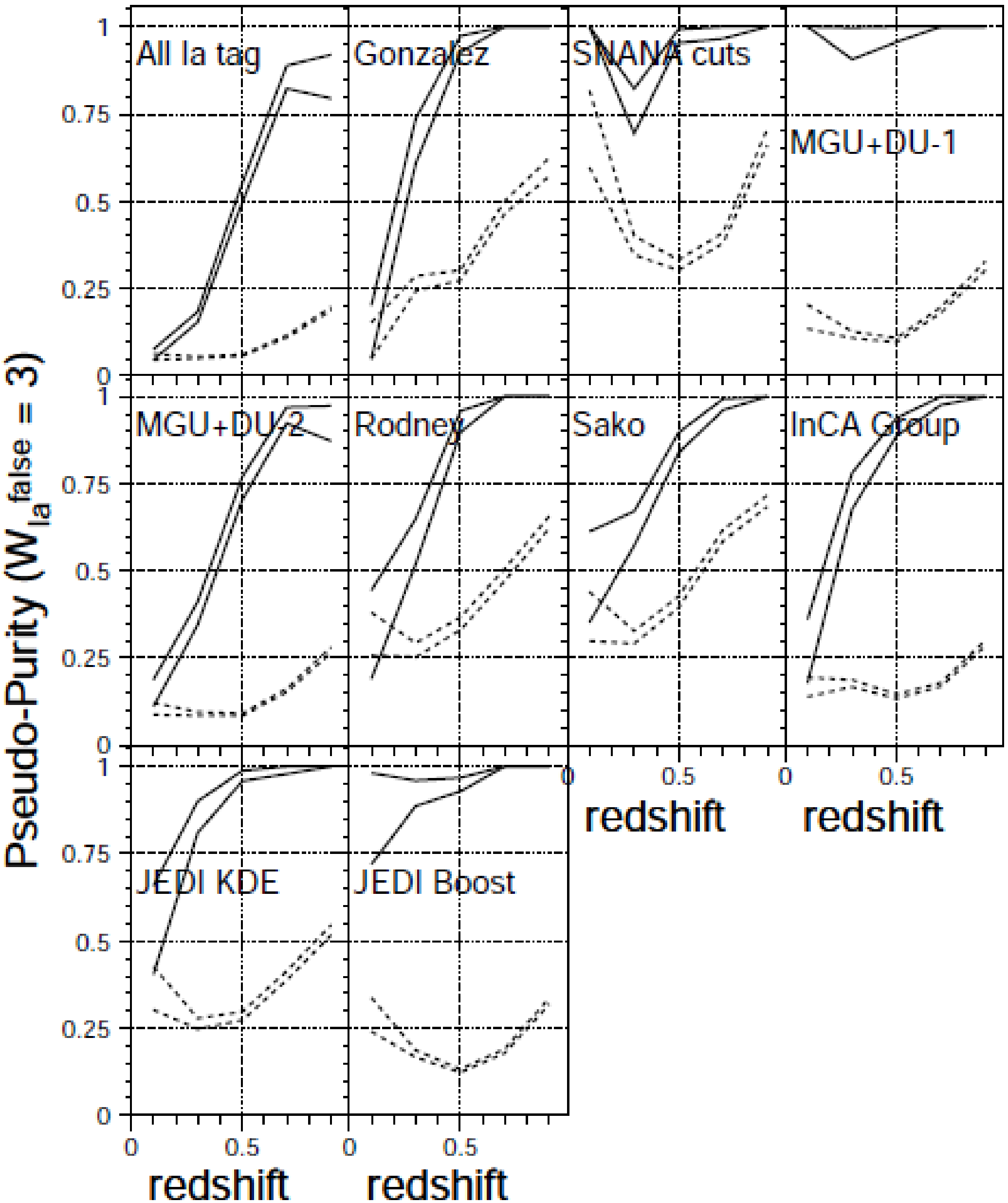
5. kPCA - results

Good news:

Purity is maximized!

$$\frac{N_{\rm Ia}^{\rm true}}{N_{\rm Ia}^{\rm true} + W_{\rm Ia}^{\rm false} N_{\rm Ia}^{\rm false}}$$



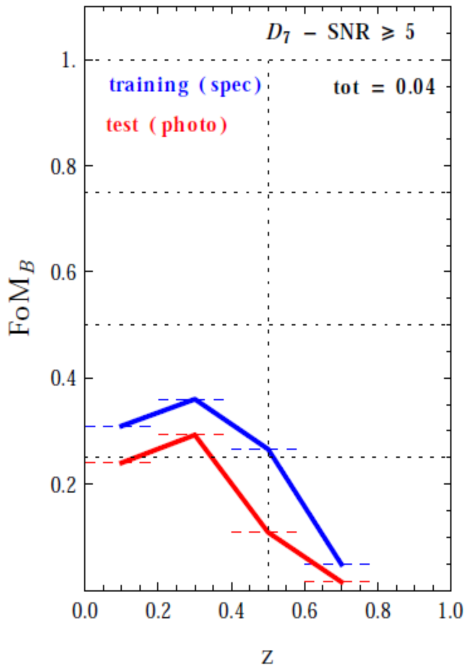


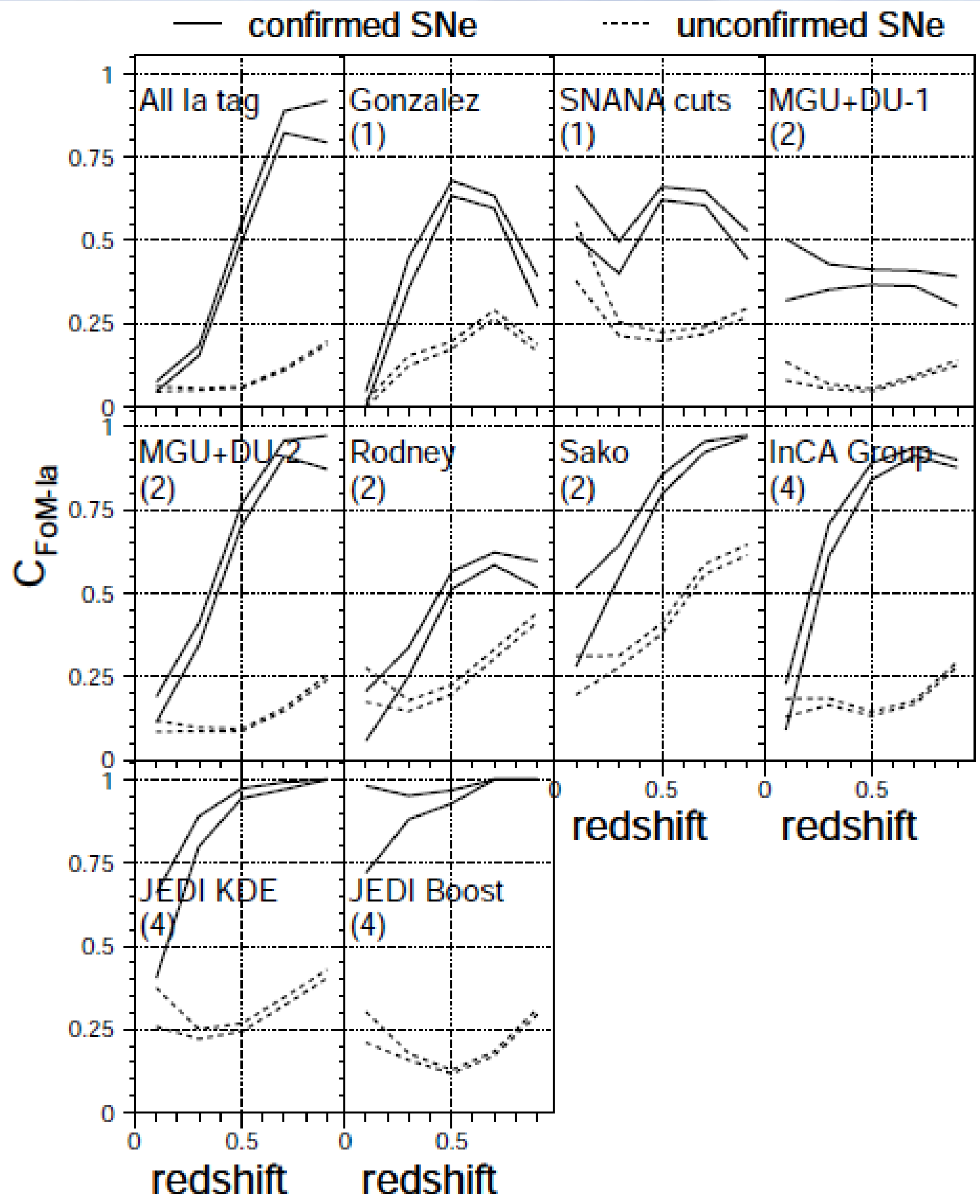
Results from the *Supernova Photometric Classification Challenge* (Kessler et al., 2010)

5. kPCA - results

Figure of Merit

$$\frac{N^{\rm true}_{\rm Ia}}{N^{\rm TOT}_{\rm Ia}} \times \frac{N^{\rm true}_{\rm Ia}}{N^{\rm true}_{\rm Ia} + W^{\rm false}_{\rm Ia} N^{\rm false}_{\rm Ia}}$$





How important it is to use the whole sample?

Need to test on cosmology fit

Results from the *Supernova Photometric Classification Challenge* (Kessler et al., 2010)

FAPESP

5. kPCA - results

Message to take home:

In the context of DES as in the SNPCC,

KPCA is able to classify 15% of the initial sample
With 91% purity

This is ideal for a first approach to cosmological use of Sne.

FAPESP

5. kPCA - results

Important remark!

Comparisons with SNPCC results must take into account:

1. Different analyzing conditions
2. No time restrictions
3. Previous experience with the outcomes of the SNPCC themselves
4. Our main goal was to maximize purity!

5. kPCA - results

Post-SNPCC results

Richards *et al.*, 2012

Diffusion map + random forest

90% purity, 8% efficiency
from a redshift limited spectroscopic sample (z < 0.4)

Ishida & de Souza, 2012

kPCA + k=1 nearest neighbor

98% purity, 7% efficiency
SNR > 5, and observations from -3 to +45
No need to change spectroscopic follow up strategy
Can be much better with interactive approach

Karpenka, Feroz and Hobson, 2012

Parametric fit + neural networks

82% purity, 82% efficiency
using 40% of the photometric sample as training (~8400 SNe)

6. Perspectives: looking for new detections

How to observe Population III Sne?

Populations III.1 and III.2 gamma-ray bursts: constraints on the event rate for future radio and X-ray surveys

R. S. de Souza[1,2], N. Yoshida[1], and K. Ioka[3]   A&A 533, A32 (2011)

Searching for the first stars with the Gaia mission

R. S. de Souza[1,2,3], A. Krone-Martins[4], E.E.O. Ishida[1,3], and B. Ciardi[3]

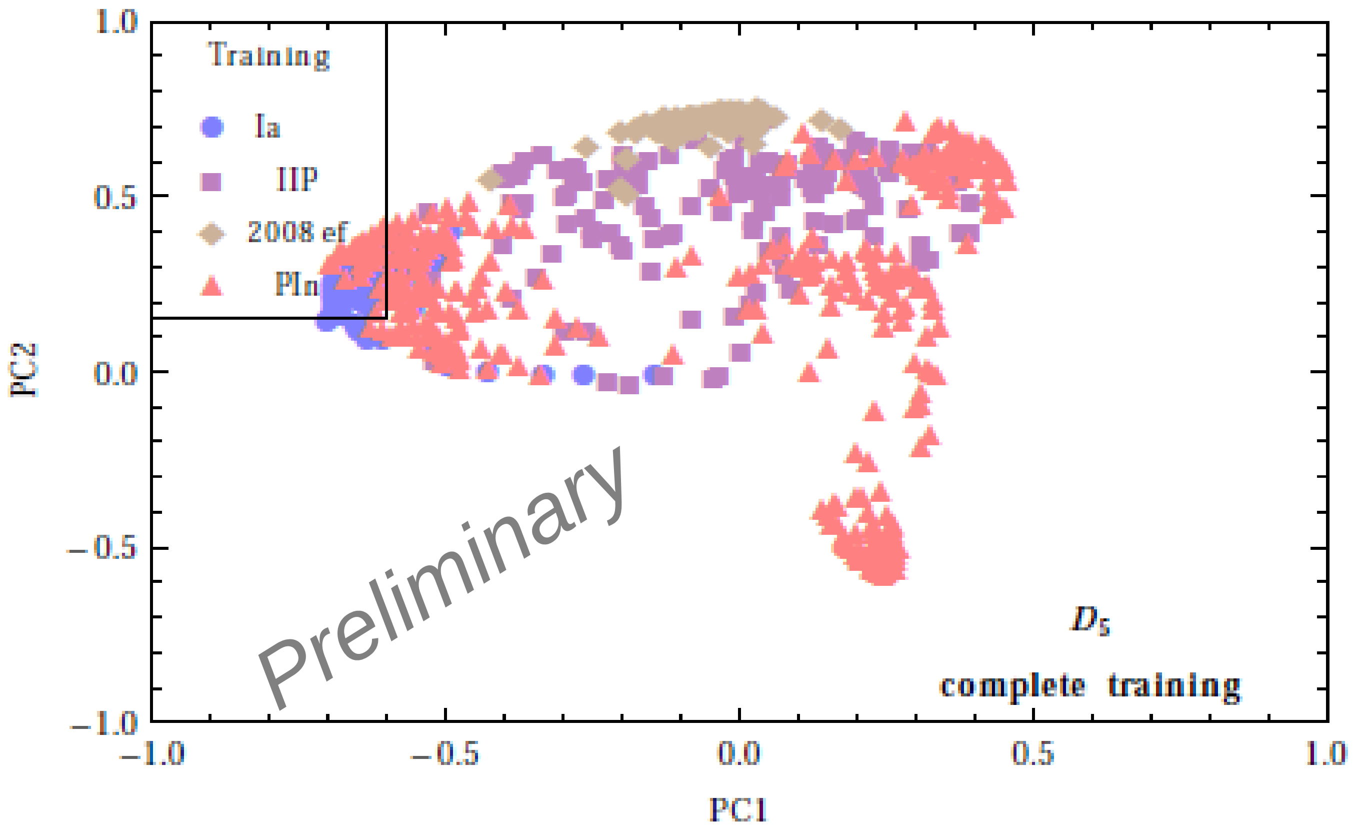
A&A, 545, id A102 (2012)

What about their Sne?

FAPESP

6. Perspectives: looking for new detections

Future application of kPCA + 1NN algorithm
Along with important SNANA modifications
And careful astrophysical modeling



De Souza, Ishida and Johson, in prep

FAPESP

7. Final Remarks

kPCA analysis provided interesting results using very simple classification algorithm and the most standard kernel function possible.

Ideal for a first approach to non-confirmed light curves into cosmological analysis

Promising technique to identify still non-observed objects based on their theoretical predictions

Thank you!