Optimize training samples for future supernova surveys using Active Learning

Cosmostat Day on Machine Learning in Astrophysics
26 January 2018

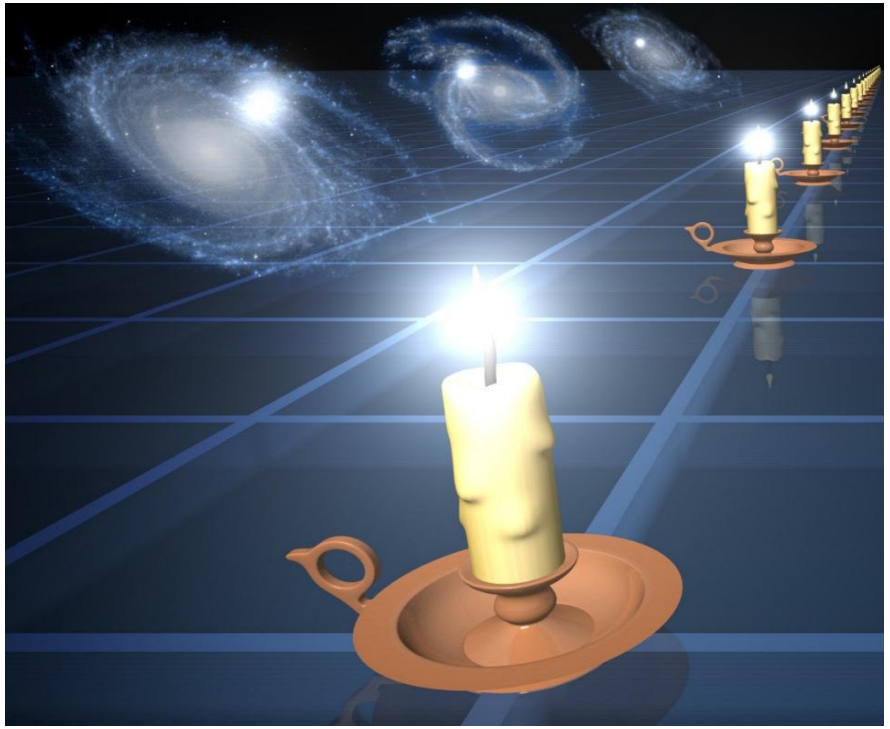
Emille E. O. Ishida

Laboratoire de Physique de Clermont - Université Clermont-Auvergne
Clermont Ferrand, France

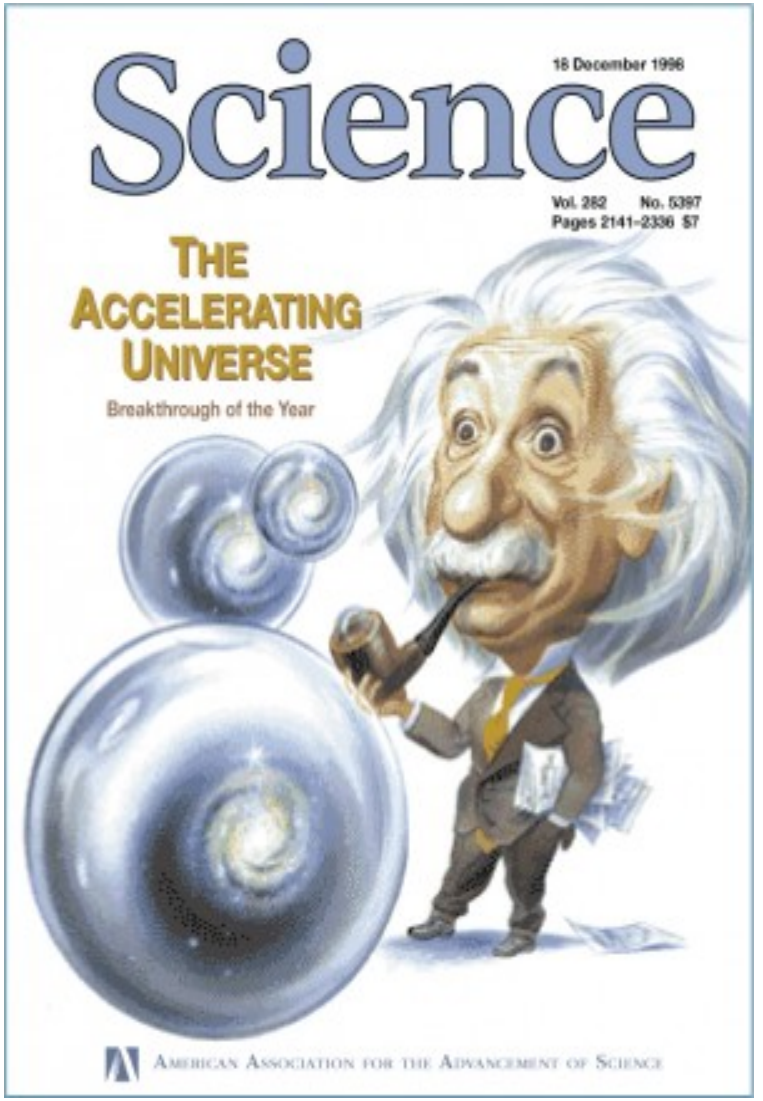
Supernova Cosmology

Measuring the accelerated expansion of the Universe

2011

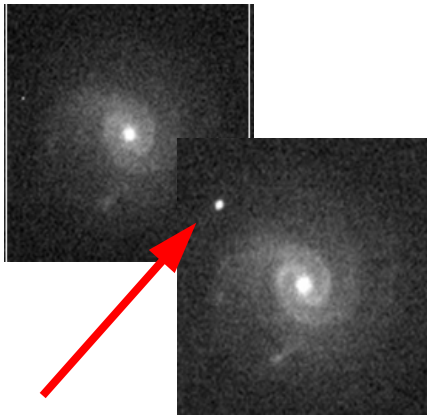


https://www.extremetech.com/wp-content/uploads/2016/01/magnetars-3.jpg

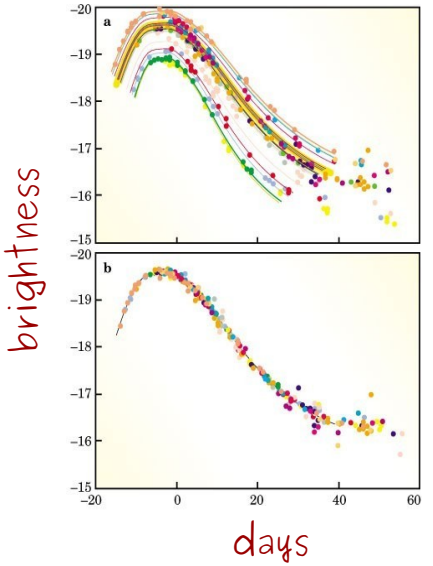


Supernova Cosmology

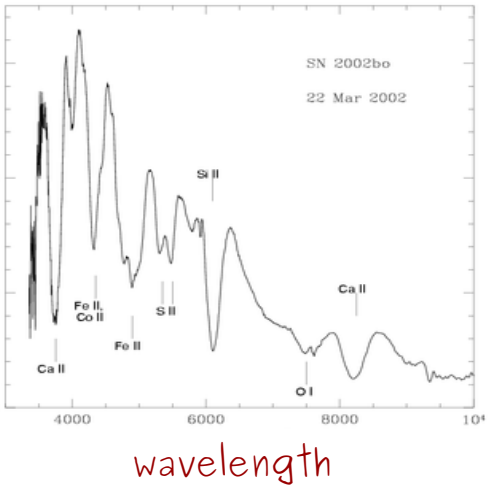
1. detection



2. photometry

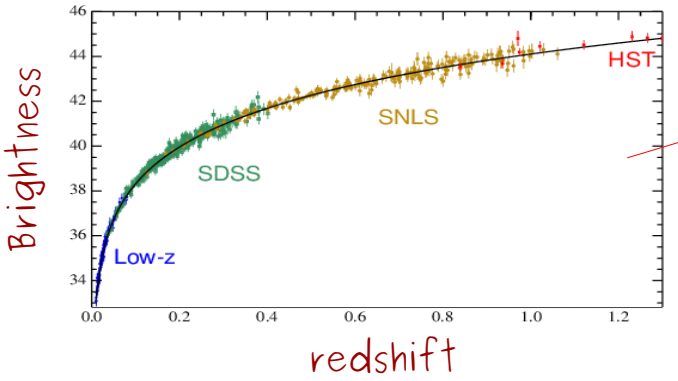


3. spectroscopy



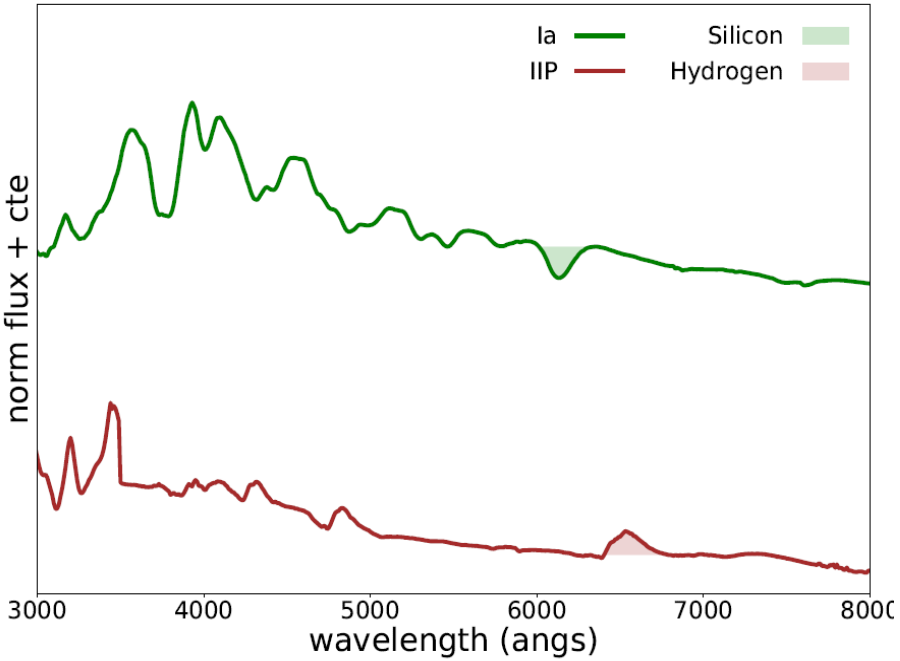
Distance (redshift) + classification

4. standardization + cosmological fit



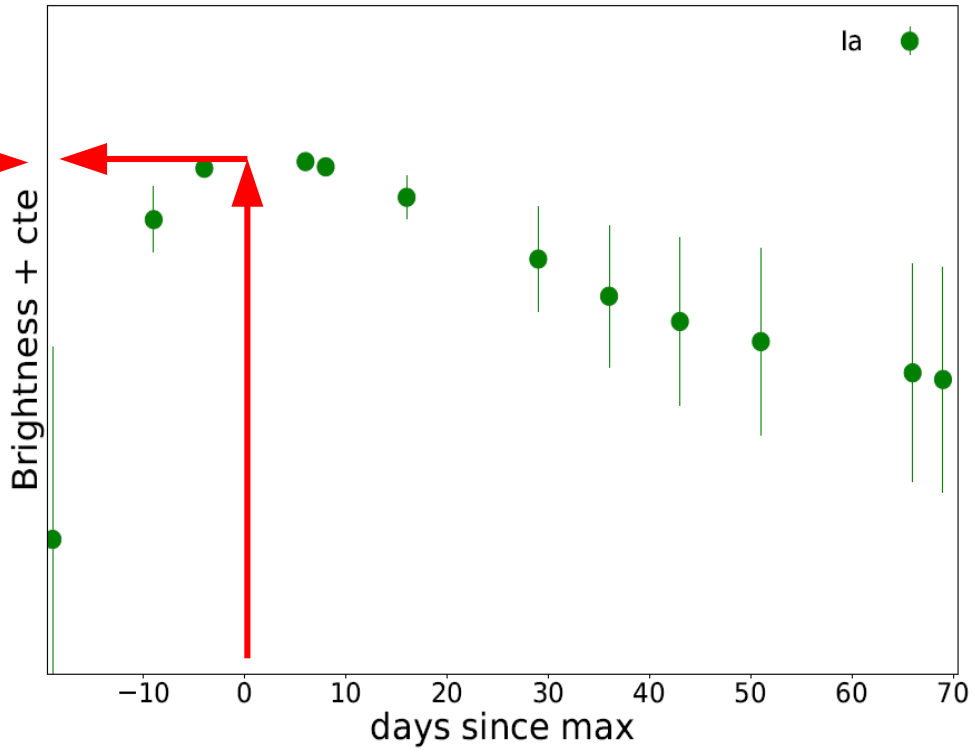
year	Number of supernova
1998	42
2014	740
2025	> 10 000

Type Ia Supernovae – how to identify them?



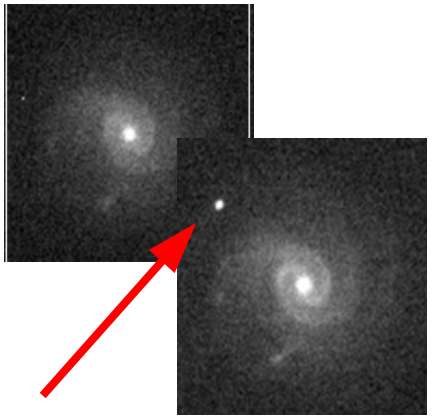
Spectrum at maximum light

Light-curve

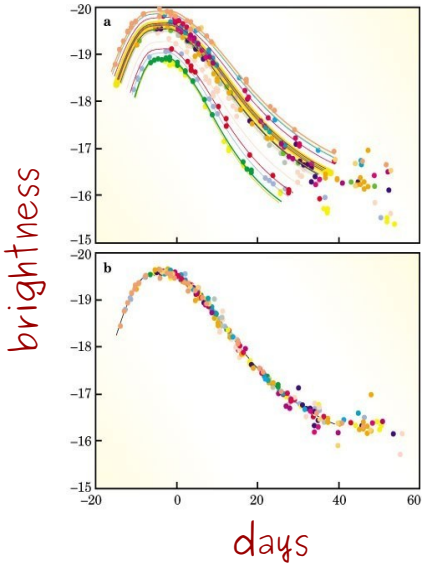


Supernova Cosmology

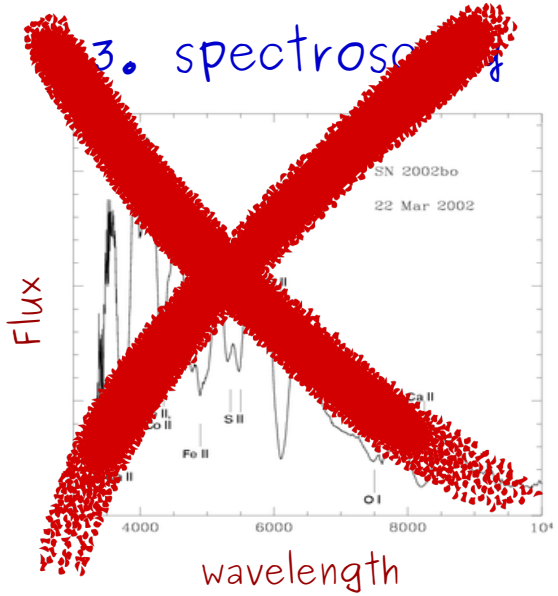
1. detection



2. photometry

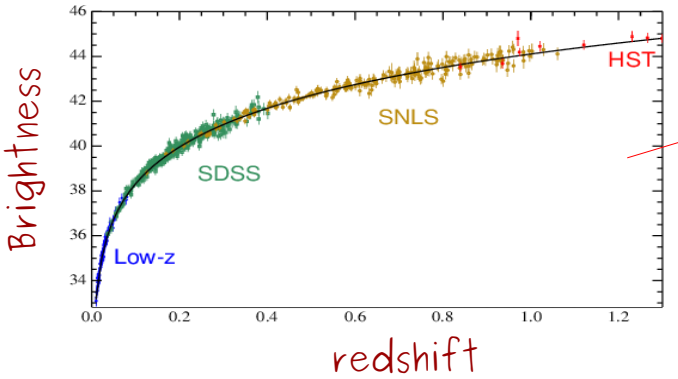


3. spectroscopy



Distance (redshift) + classification

4. standardization + cosmological fit



year	Number of supernova
1998	42
2014	740
2025	> 10 000

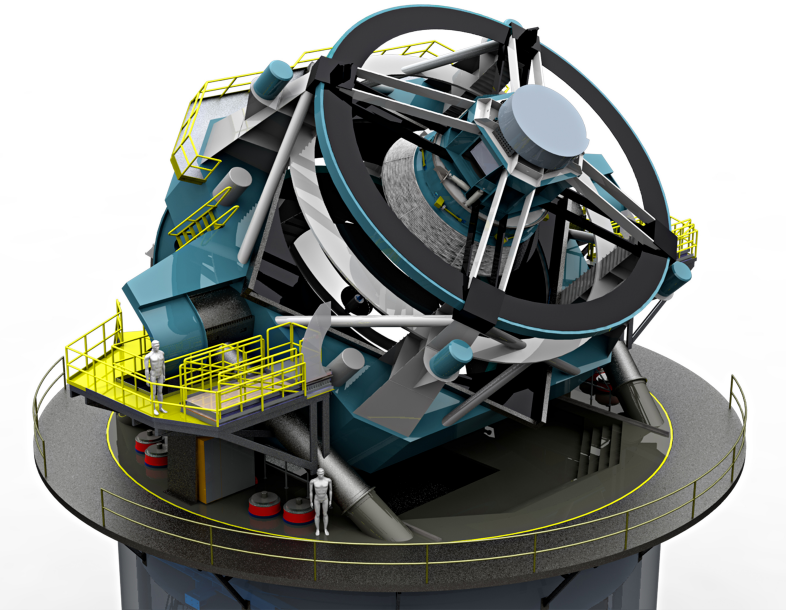
Big Data (in astronomy) → Large Scale Sky Surveys

year	Number of supernova
1998	42
2014	740
2025	> 10 000

2 million alerts/day
15 TB/day

40 nights of LSST

entire Google database

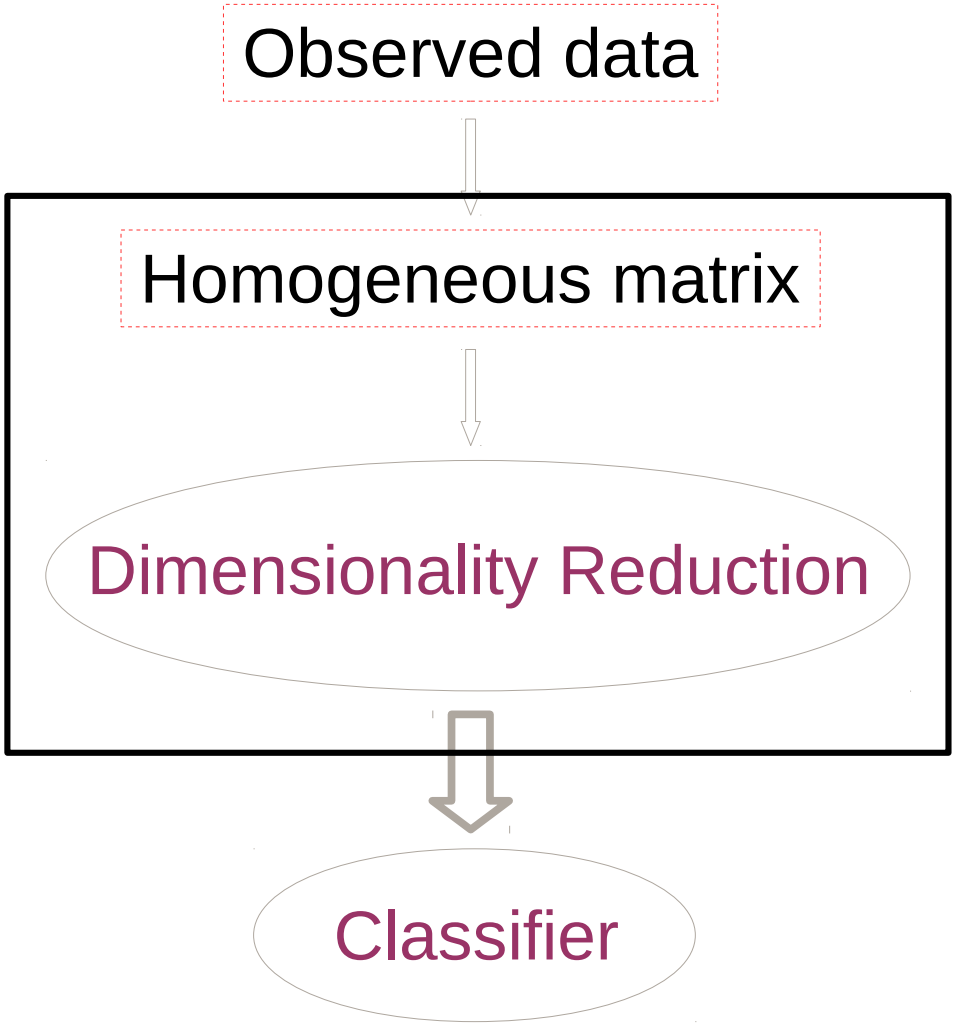


https://www.lsst.org/

Large Synoptic Survey Telescope

ML for Supernova classification

Traditional strategy



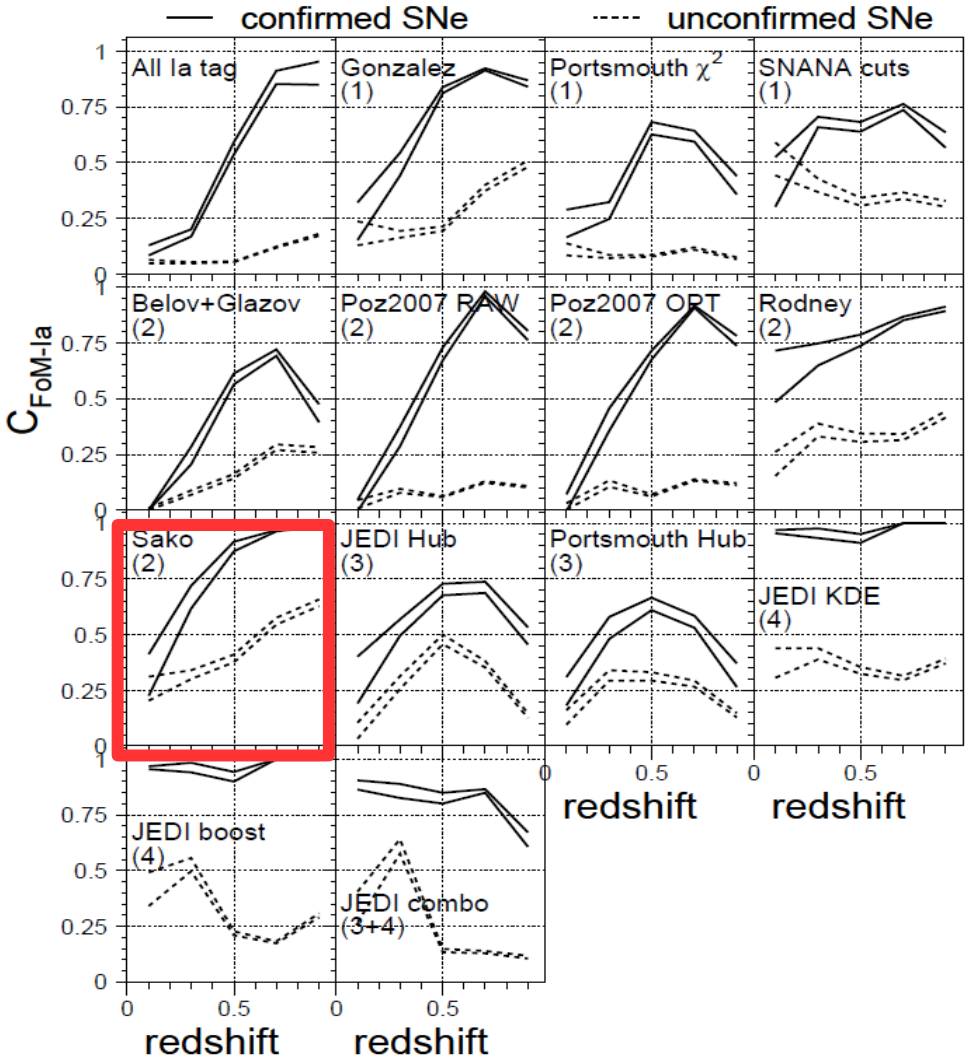
Pre-processing

Parametric fit

ML for Supernova classification

The Supernova Photometric Classification Challenge

Kessler et al., 2010, PASP, Volume 122, Issue 898, pp. 1415



Outcomes:

1 – None of the methods obviously outperformed the others

2 – SNID had better overall metric

3- An updated data set was released to the community

Spectroscopy x Photometry

High quality

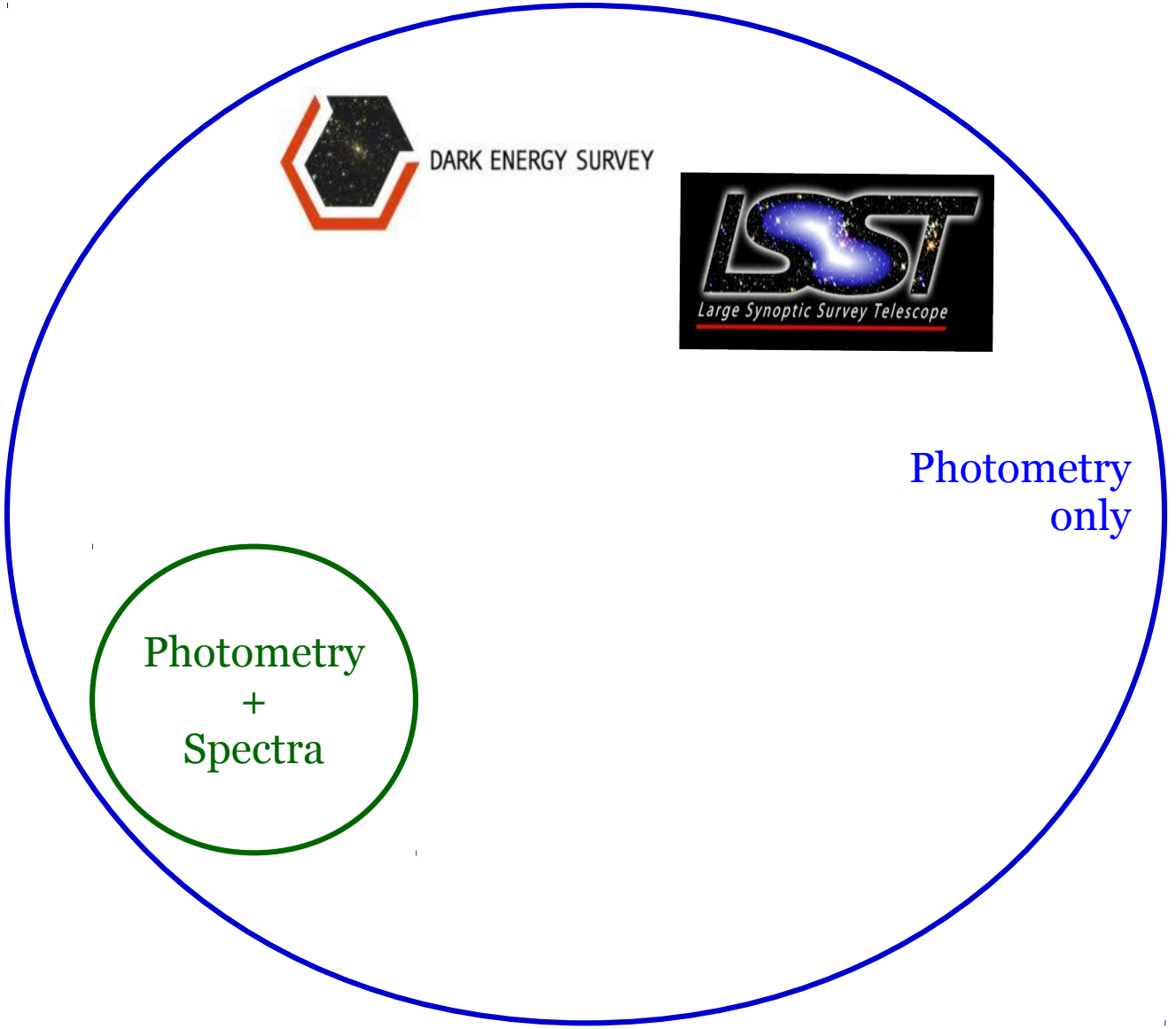
Expensive

Scarce

Low quality

Cheap

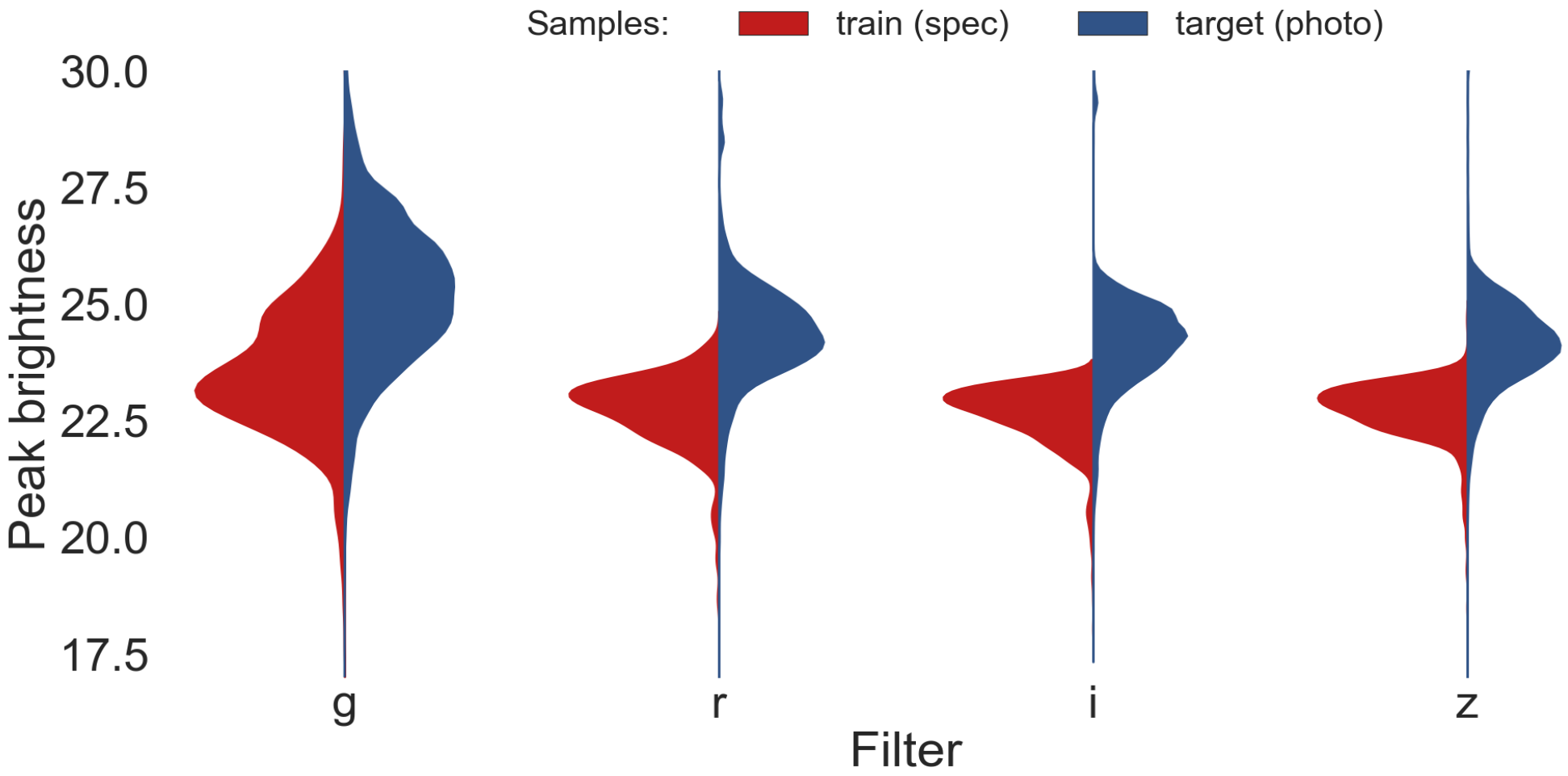
Abundant



PS: Machine Learning methods do not extrapolate!

ML for Supernova classification

Representativeness

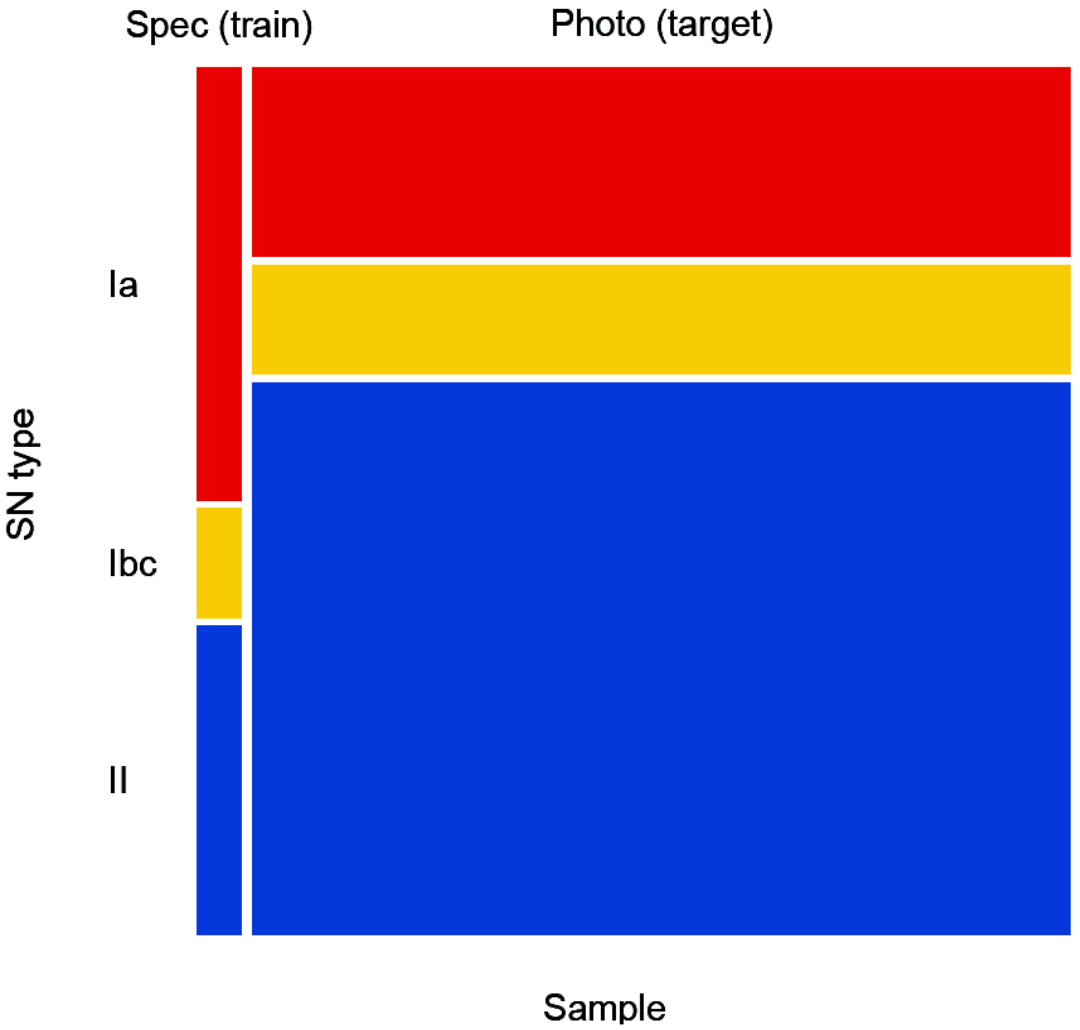


From COIN Residence Program #4, ***Ishida*** *et al., 2018 – in prep*

The Data: post-SNPCC simulations – *Kessler et al., 2010*

ML for Supernova classification

Representativeness

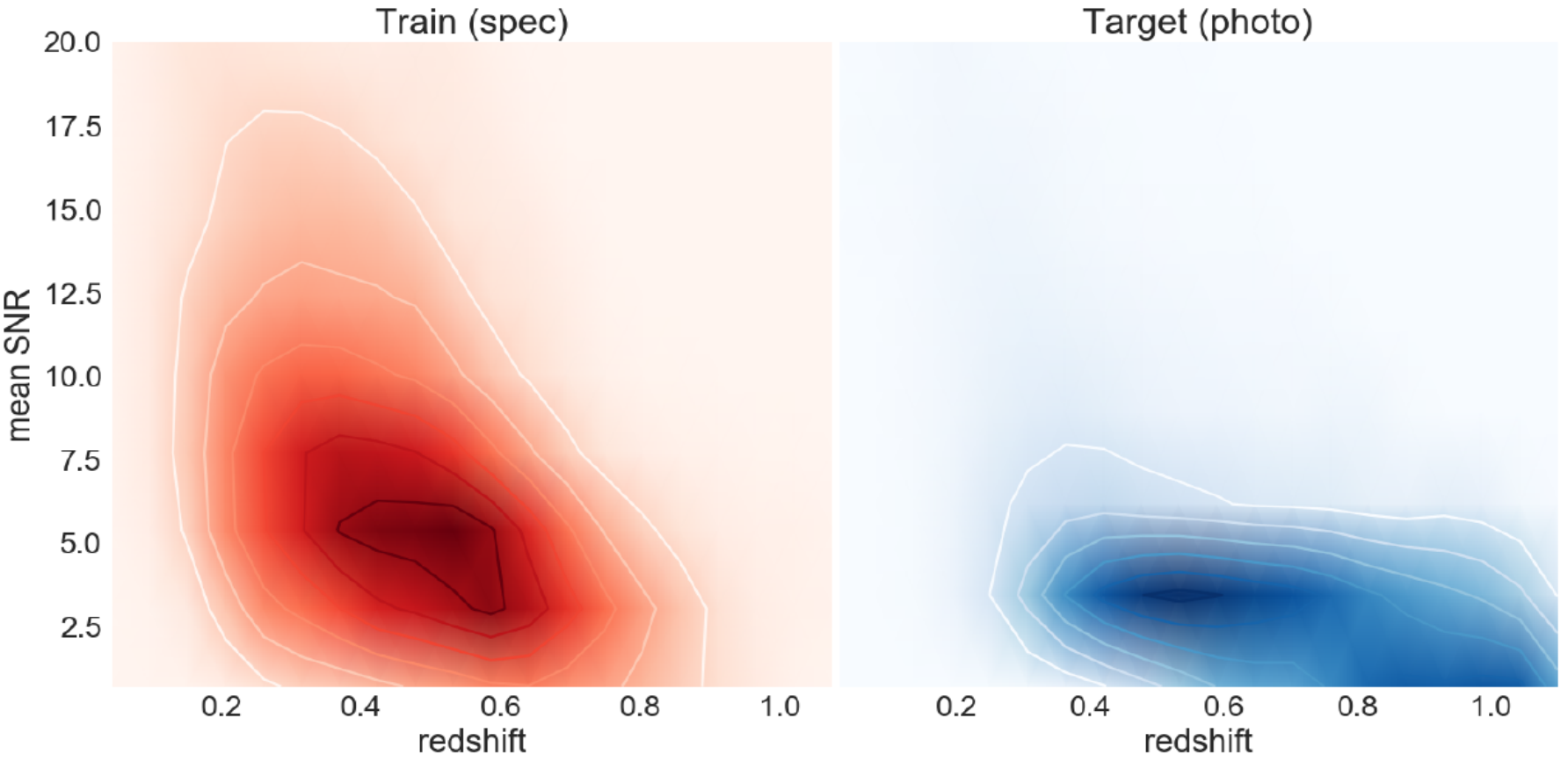


From COIN Residence Program #4, ***Ishida*** *et al., 2018 – in prep*

The Data: post-SNPCC simulations – *Kessler et al., 2010*

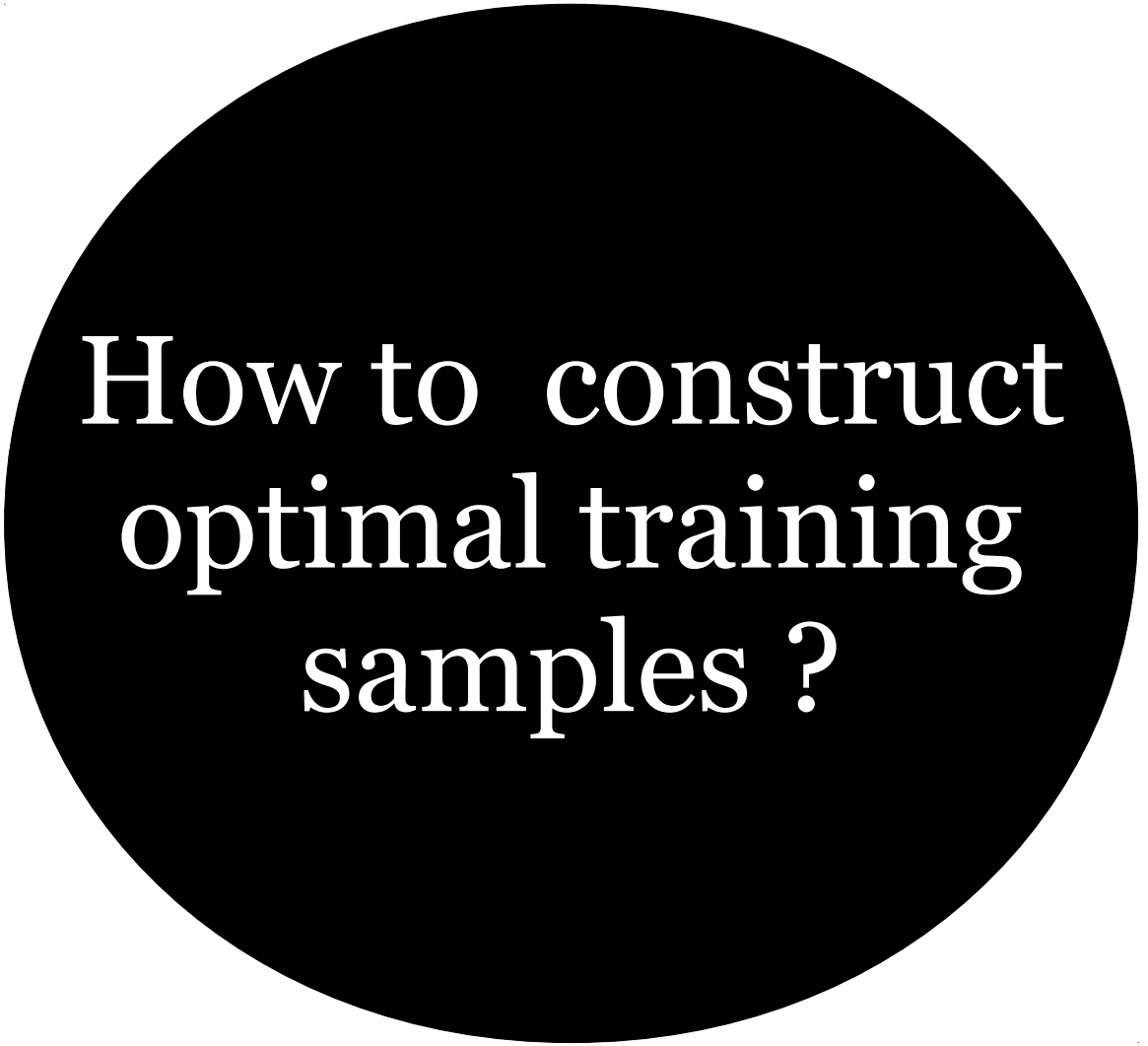
ML for Supernova classification

Representativeness



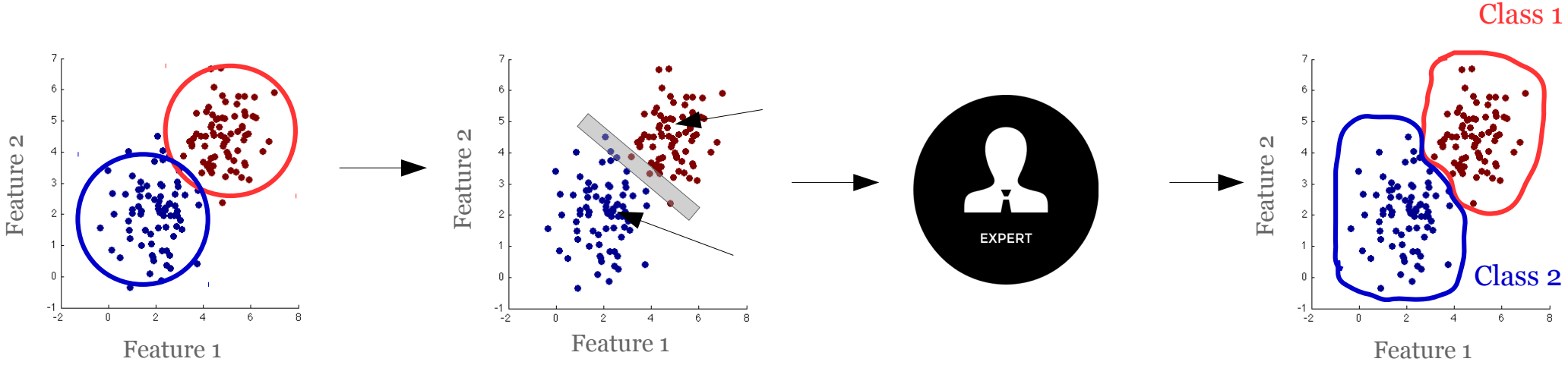
From COIN Residence Program #4, ***Ishida*** *et al., 2018 – in prep*

The Data: post-SNPCC simulations – *Kessler et al., 2010*


How to construct optimal training samples ?

Active Learning *or Optimal Experimental Design*

"Can machines learn with **fewer labeled** training instances if they are allowed to ask questions?"



Step 1

Standard Classification

Step 2

Identify problematic as well as representative elements

Step 3

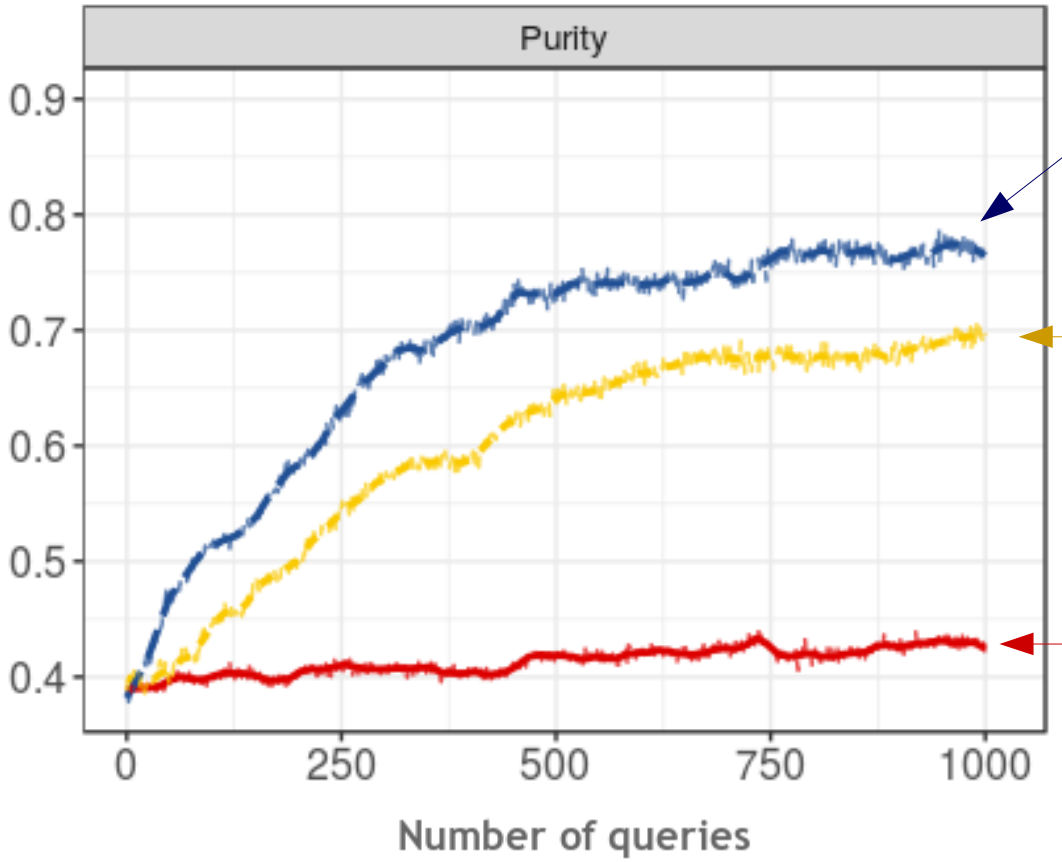
Make a query

Step 4

Classify

Active Learning in Astronomy

the case for Supernova Classification



Active Learning

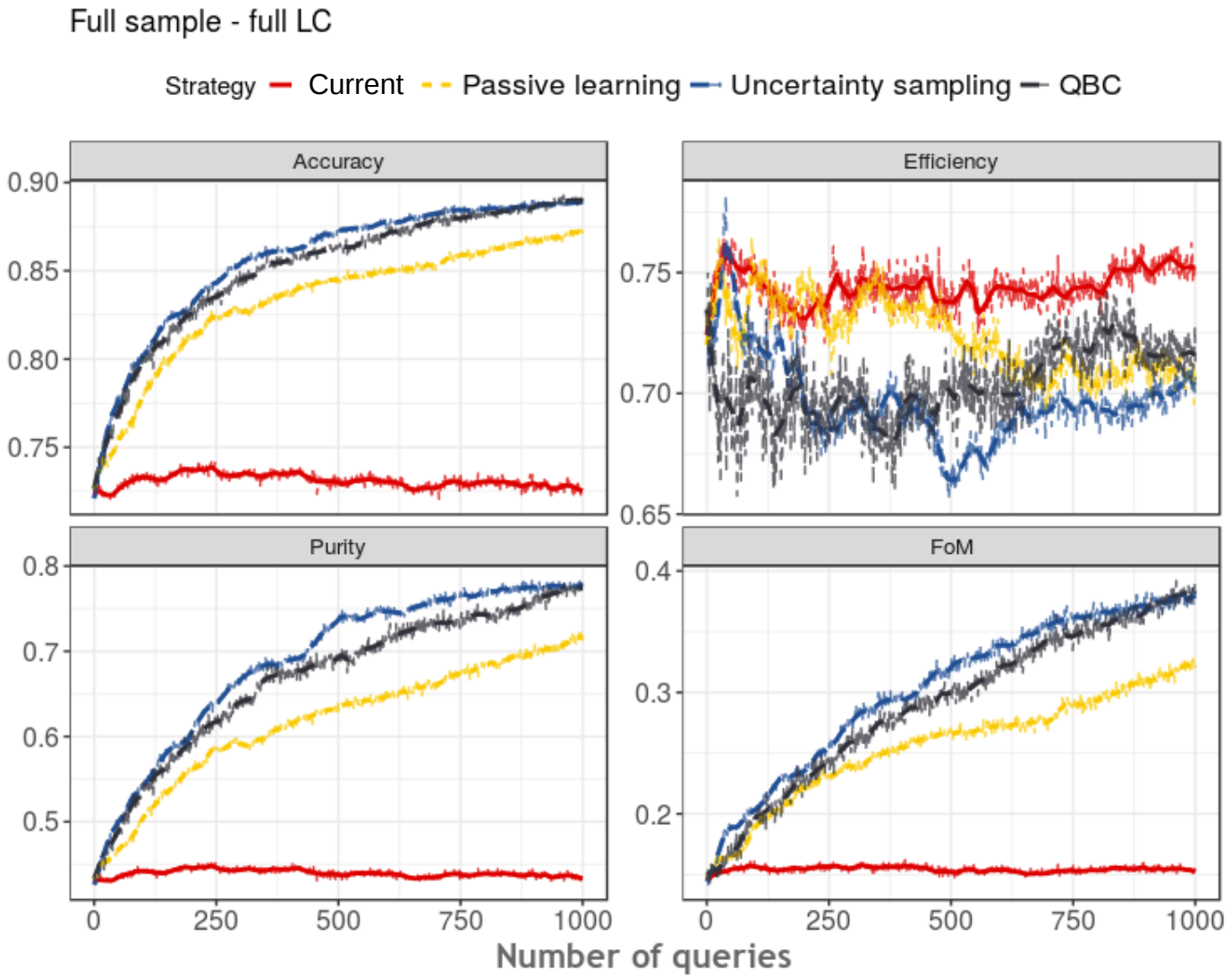
Passive Learning

Current strategy

From COIN Residence Program #4, ***Ishida*** *et al., 2018 – in prep*

Active Learning in Astronomy

the case for Supernova Classification



$$\text{eff} = \frac{N_{\text{Ia}}^{\text{SC}}}{N_{Ia}^{\text{tot}}}$$

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

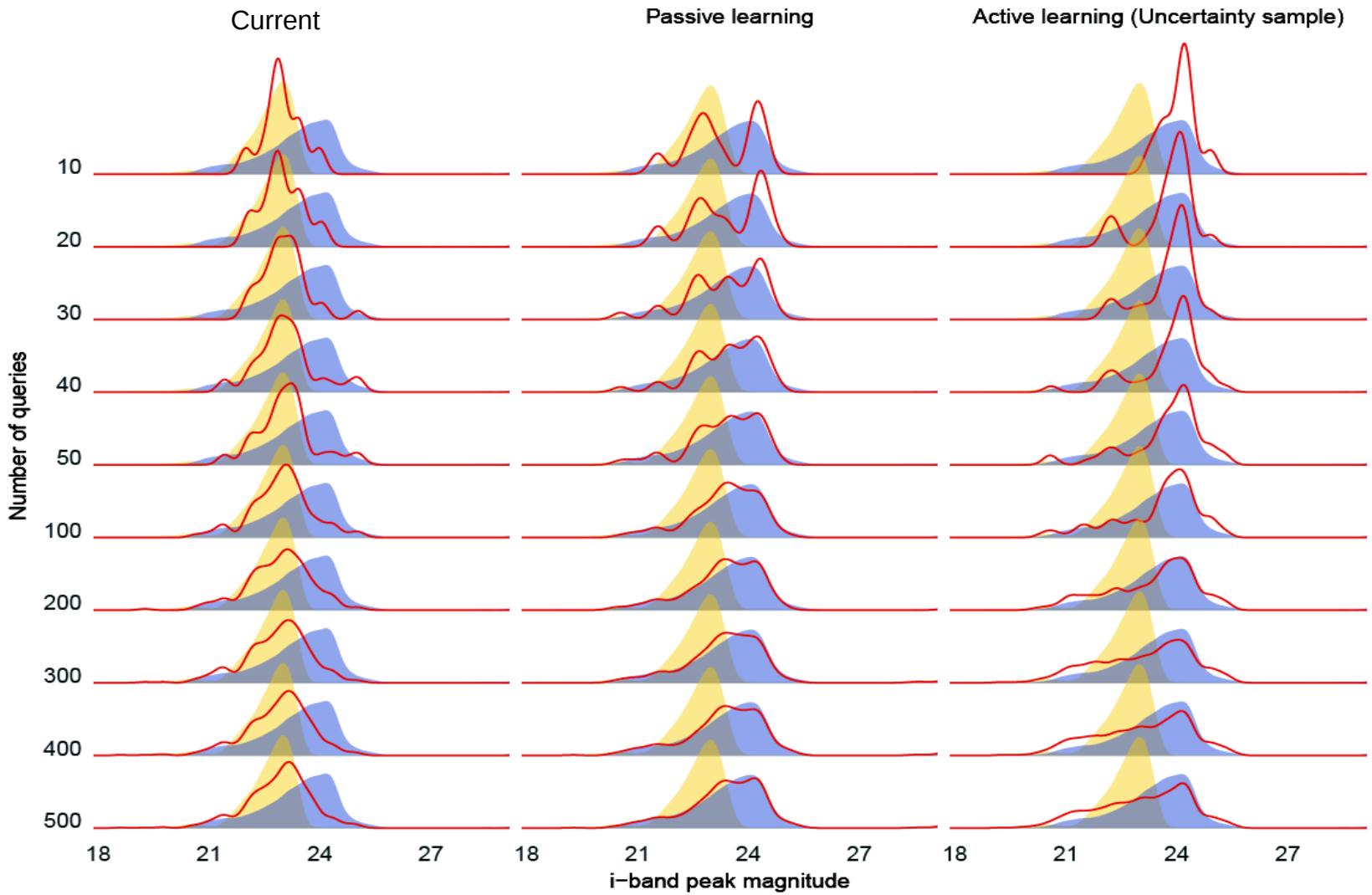
$$\text{ppur} = \frac{N_{\text{Ia}}^{\text{SC}}}{N_{\text{Ia}}^{\text{SC}} + W N_{\text{nonIa}}^{\text{WC}}},$$

$$\text{FoM} = \text{eff} \times \text{ppur},$$

From COIN Residence Program #4, ***Ishida*** *et al., 2018 – in prep*

Active Learning in Astronomy

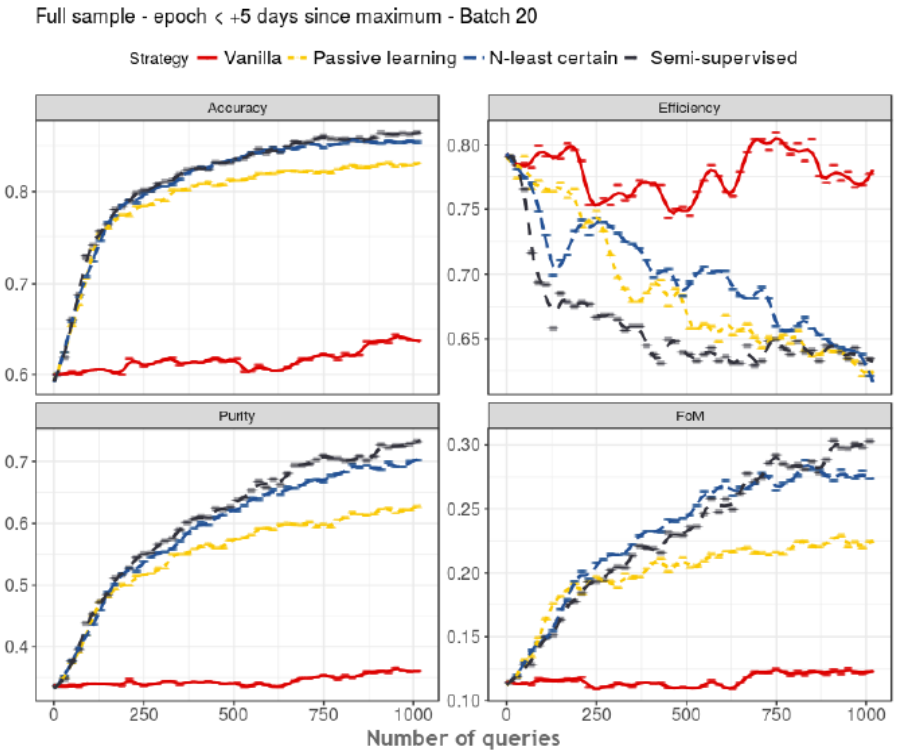
the case for Supernova Classification

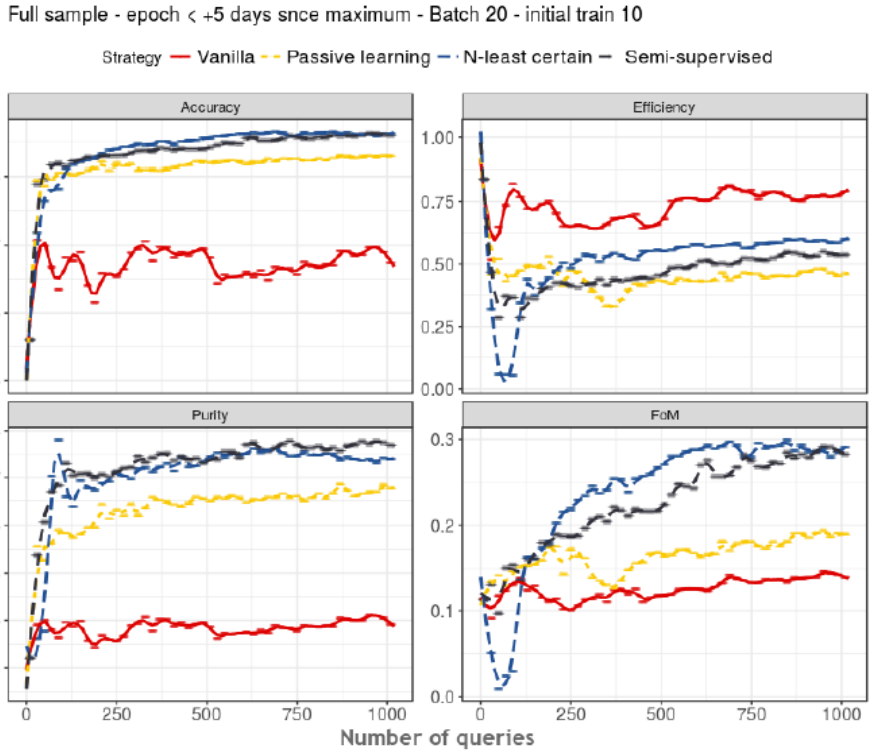


From COIN Residence Program #4, ***Ishida*** *et al., 2018 – in prep*

Active Learning in Astronomy

the case for Supernova Classification



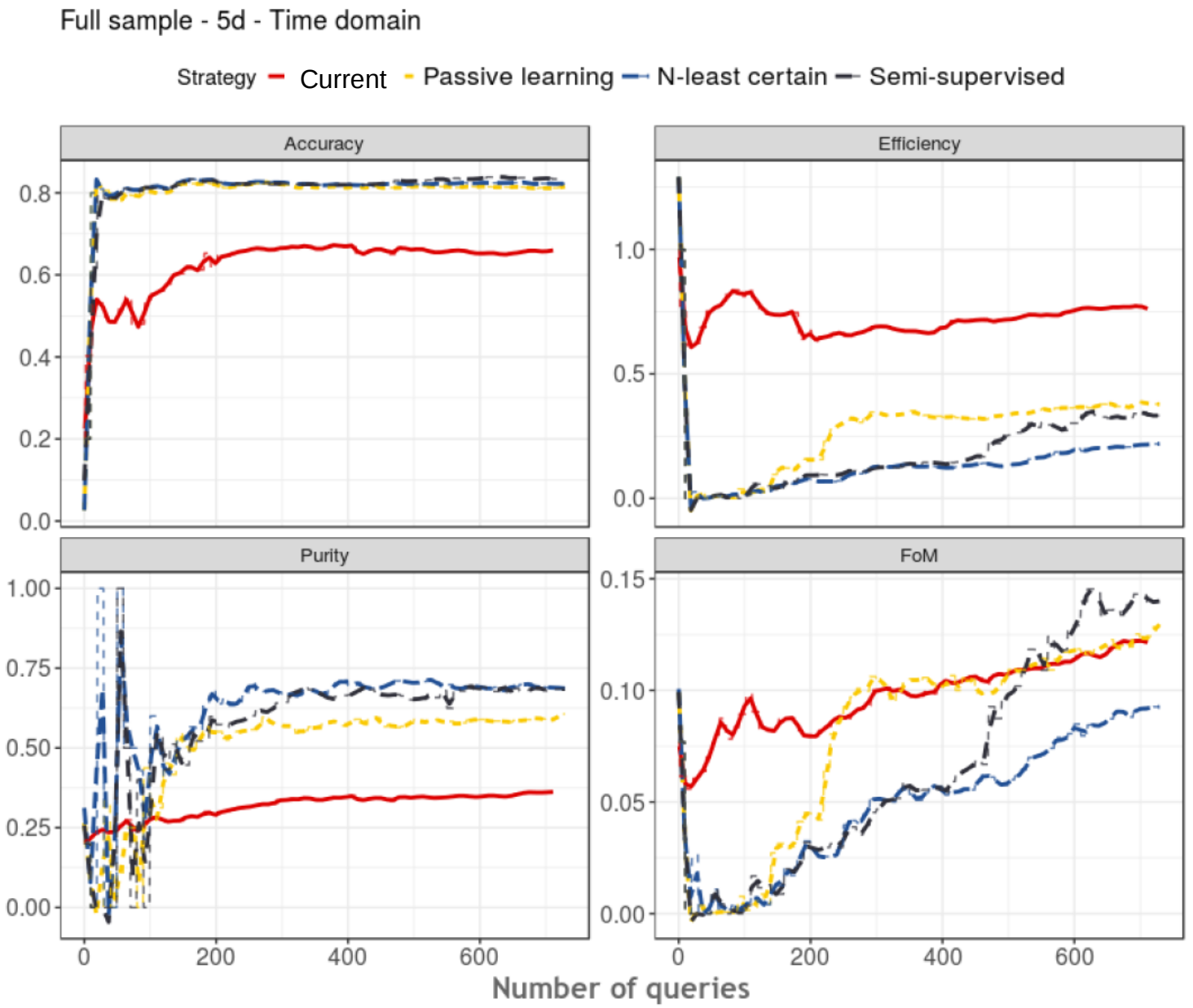


From COIN Residence Program #4, ***Ishida*** *et al., 2018 – in prep*

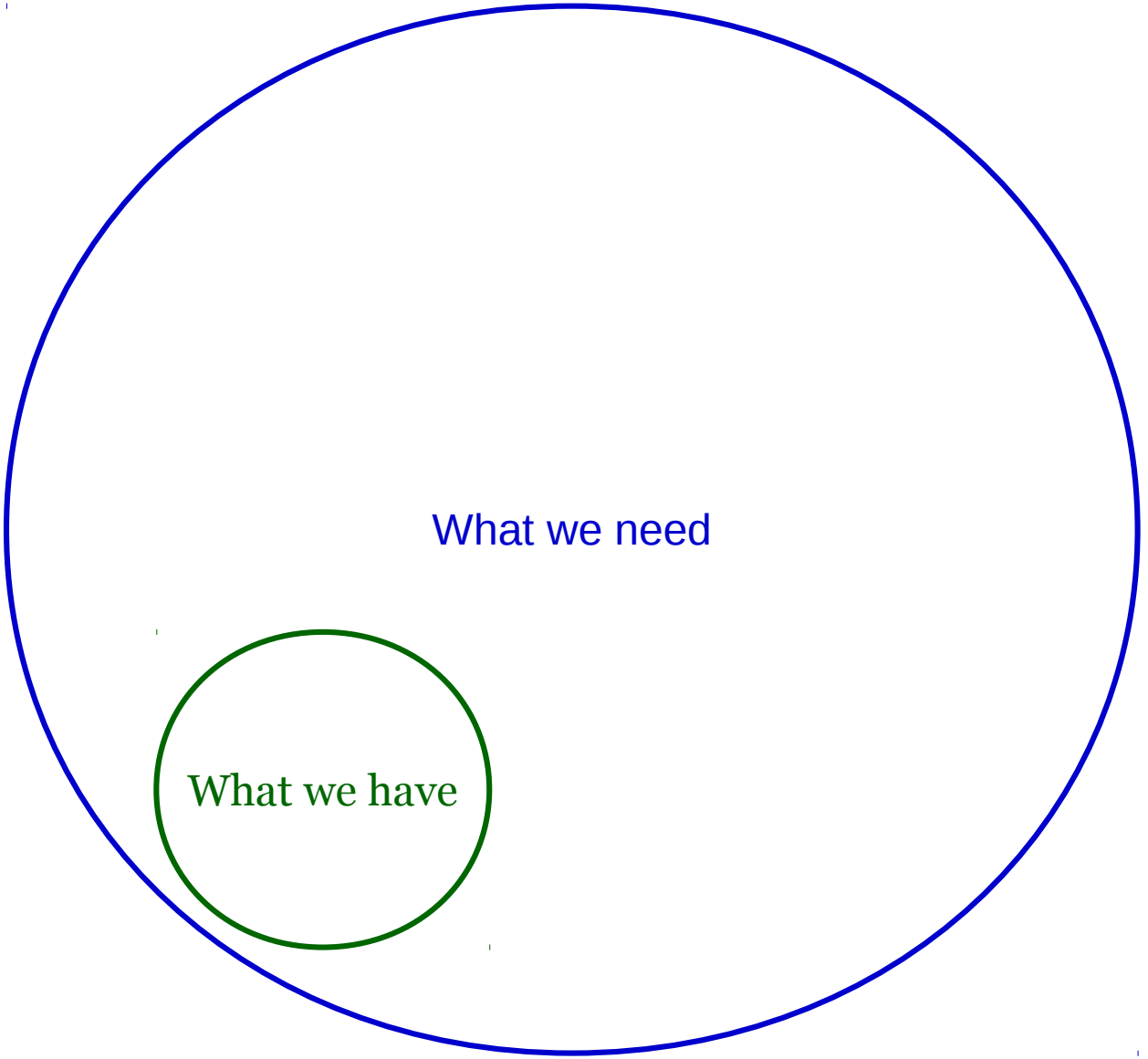
Active Learning in Astronomy

the case for Supernova Classification

Time evolution
No training



Summary



Active Learning

designed for astronomical data

"How do we optimize machine learning results with **a minimum number of labeled** training instances?"

https://github.com/COINtoolbox

COIN Residence Program #4

20 - 27 August 2017

Clermont Ferrand, France

France

Brazil

UK

Hungary

France / Brazil

Germany / USA

Portugal / Brazil

Portugal / Colombia

USA

France / Venezuela

Brazil

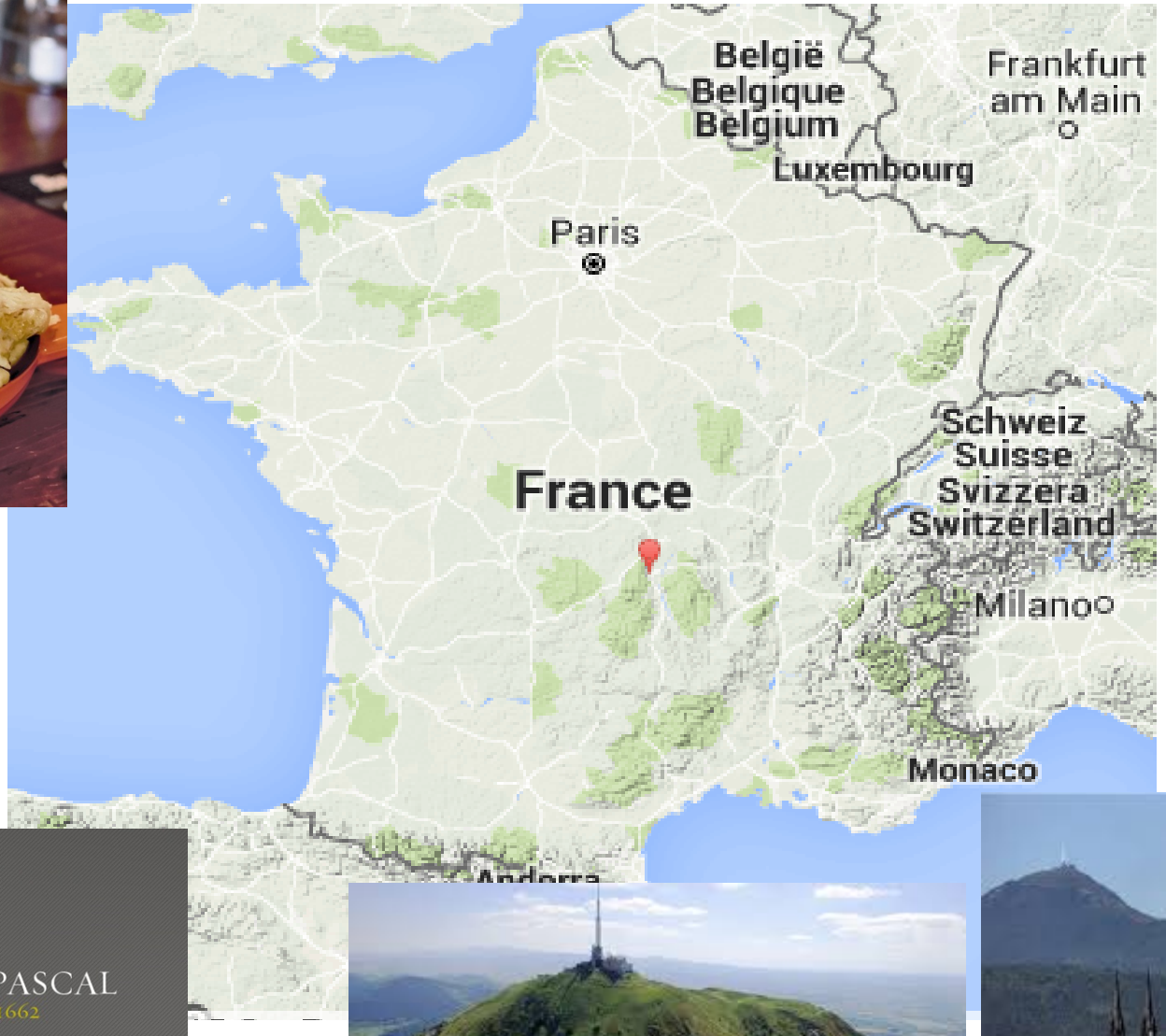
USA / Brazil

Sponsors:

Clermont Ferrand, France

We are hiring!


Truffade



COIN
Cosmostatistics Initiative

LSST
Large Synoptic Survey Telescope

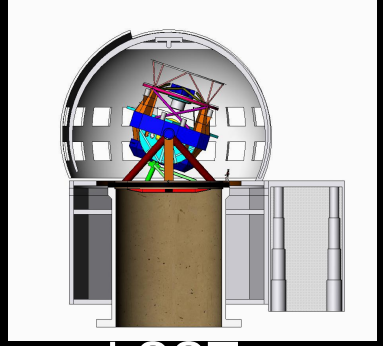

Puy-de-Dome


Merci

Extra slides

Hubble

Keck



LSST

VLT

real size

Field of View

size x 15

0.056 deg

0.2 deg

3.5 deg

0.5 deg

0.45 deg

Primary Mirror

2,4 m

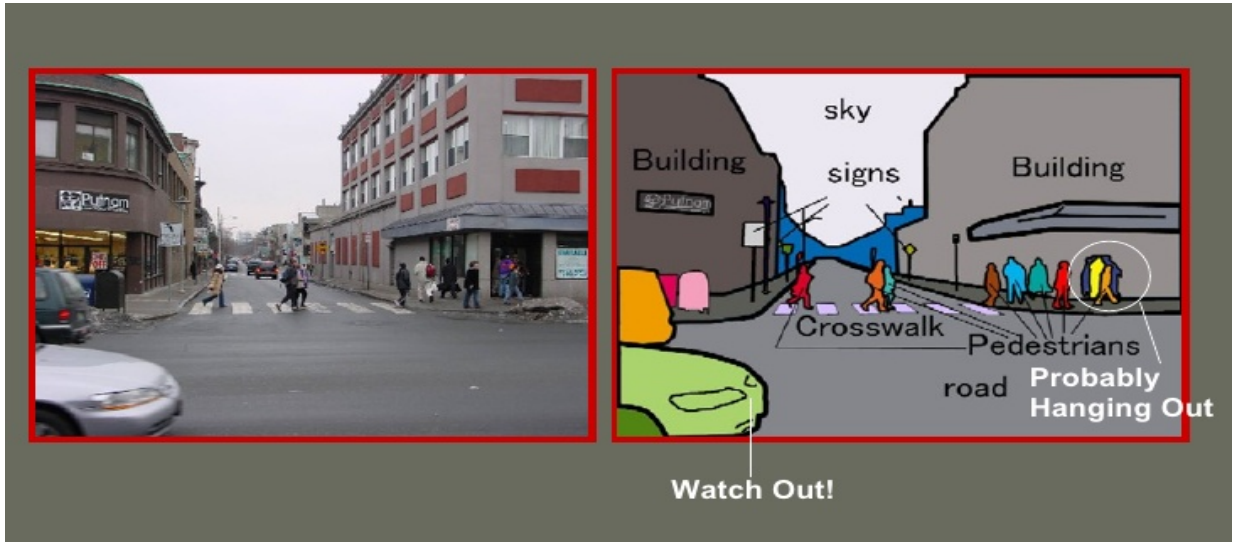
10 m

8.4 m
(6.5 eff.)

8.2 m

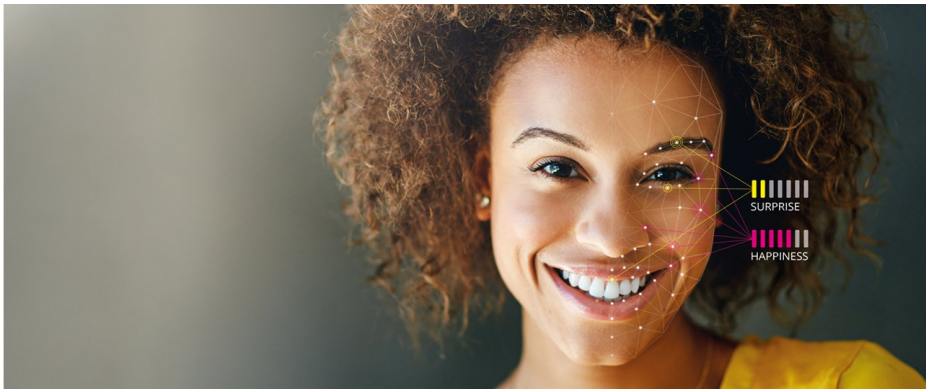
Slide from Pierre Antilogus

Big Data → Machine Learning



Credit: Stan Bileschi - CBCL

http://www.nviso-insights.com/en



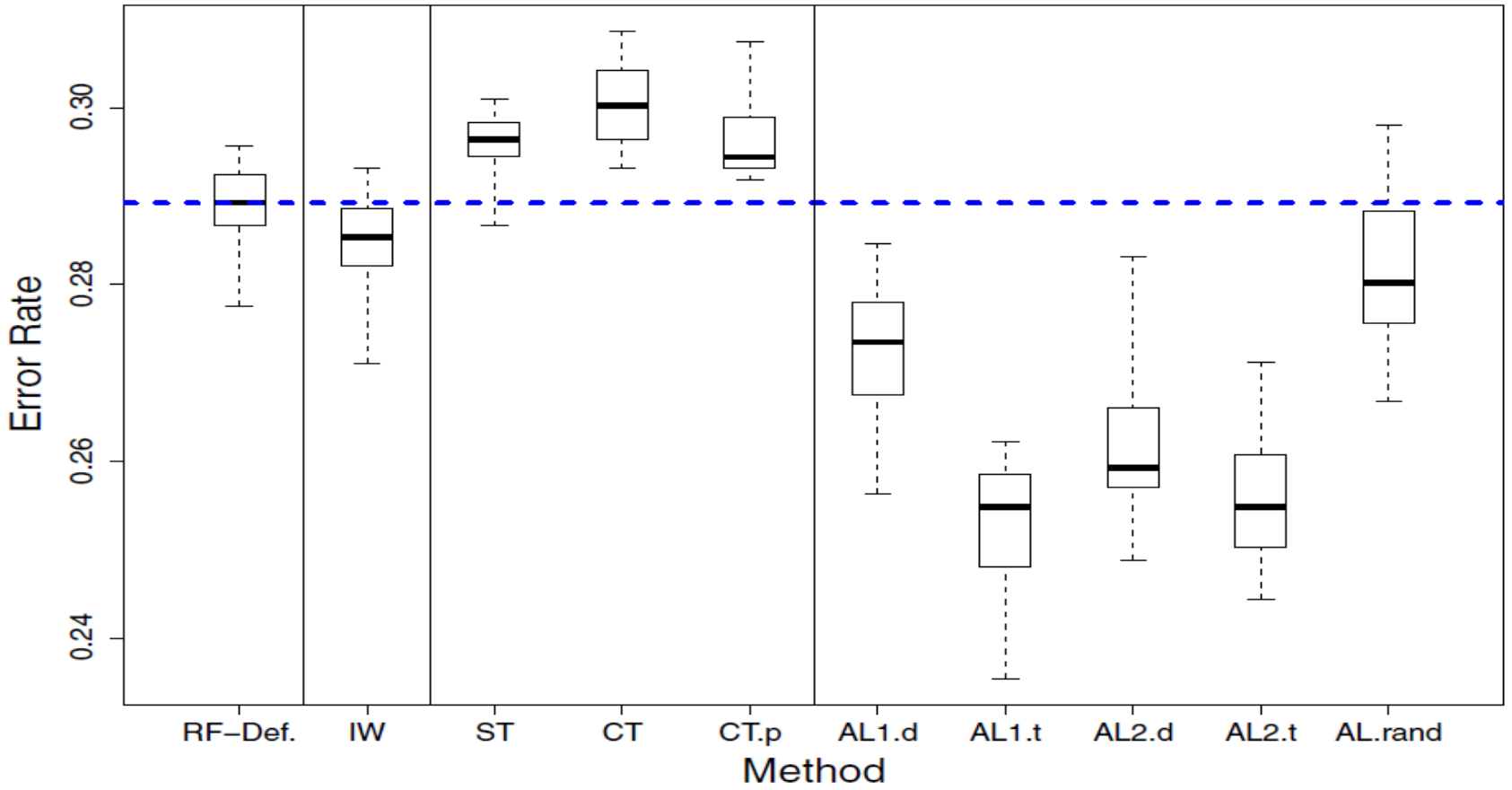
Background: Active Learning in Astronomy

ACTIVE LEARNING TO OVERCOME SAMPLE SELECTION BIAS: APPLICATION TO PHOTOMETRIC VARIABLE STAR CLASSIFICATION

JOSEPH W. RICHARDS[1,2], DAN L. STARR[1], HENRIK BRINK[3], ADAM A. MILLER[1], JOSHUA S. BLOOM[1], NATHANIEL R. BUTLER[1], J. BERIAN JAMES[1,3], JAMES P. LONG[2], AND JOHN RICE[2]

supervised classification

THE ASTROPHYSICAL JOURNAL, 744:192 (19pp), 2012 January 10



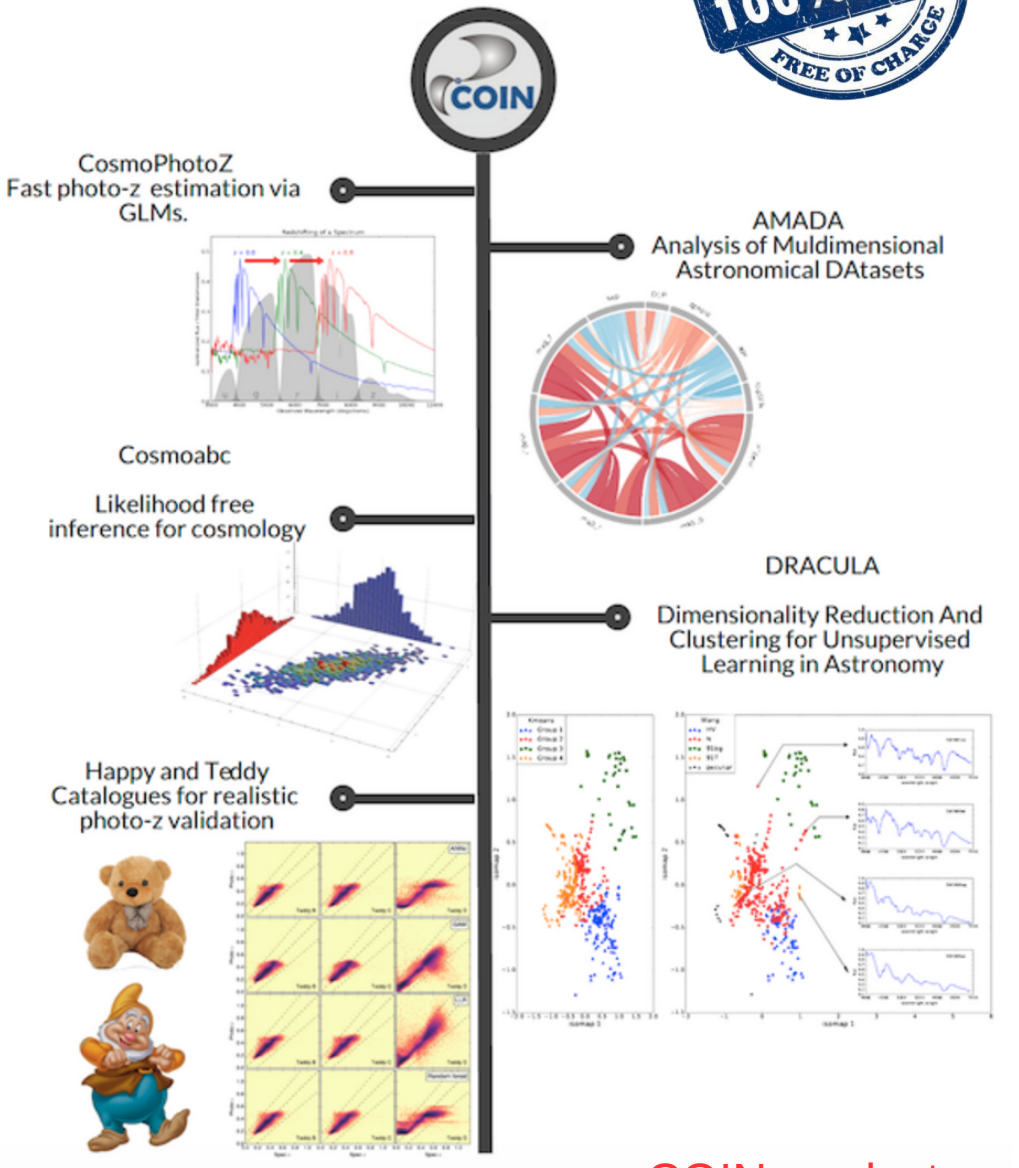
COIN products

Rafael S. de Souza (head) - statistics

Alberto Krone-Martins astrometry

Emille E. O. Ishida SN cosmology

60 researchers from **15** countries



Scientific outcomes

In 4 years

	Paper	Citation
1	GLM I	de Souza *et al.*, 2015
2	GLM II	Elliott *et al.*, 2015
3	GLM III	de Souza *et al.*, 2015
4	AMADA	de Souza & Ciardi, 2015
5	CosmoABC	Ishida *et al.*, 2015
6	DRACULA	Sasdelli *et al.*, 2016
7	AGNlogit	de Souza *et al.*, 2016
8	PhotoZ	Beck *et al.*, 2017
9	AGNgmm	de Souza *et al.*, 2017

1	CosmoPhotoZ	de Souza *et al.*, 2014,
2	AMADA	de Souza & Ciardi, 2015
3	CosmoABC	Ishida *et al.*, 2015
4	DRACULA	Aguena *et al.*, 2015

+ 1 galaxy catalog
+ 1 GMM tutorial
+ 2 photoz catalogs

COIN products are open source!

Infographic by Rafael S. de Souza