

# Early SN module with Active Learning

*Marco Leoni*

Fink collaboration meeting

20 May 2022

Annecy

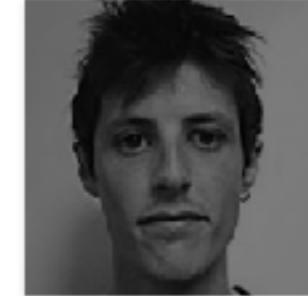


# Introduction

**Who :**



*A. Moller*  
ANU, Swinburne



*J. Peloton*  
IJCLAB, UPSaclay

*E. Ishida*

LPC, Clermont



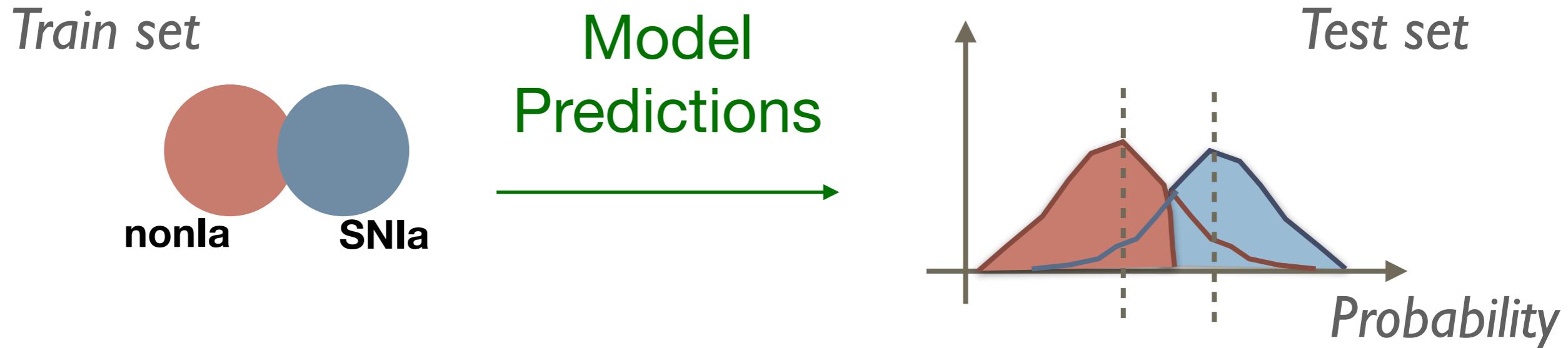
**What :** Early discovery of supernovae (no need to say why SN are relevant here !)

**Why Machine Learning :** huge amount of data from LSST hence  
the need for making ‘automatic’ reliable predictions

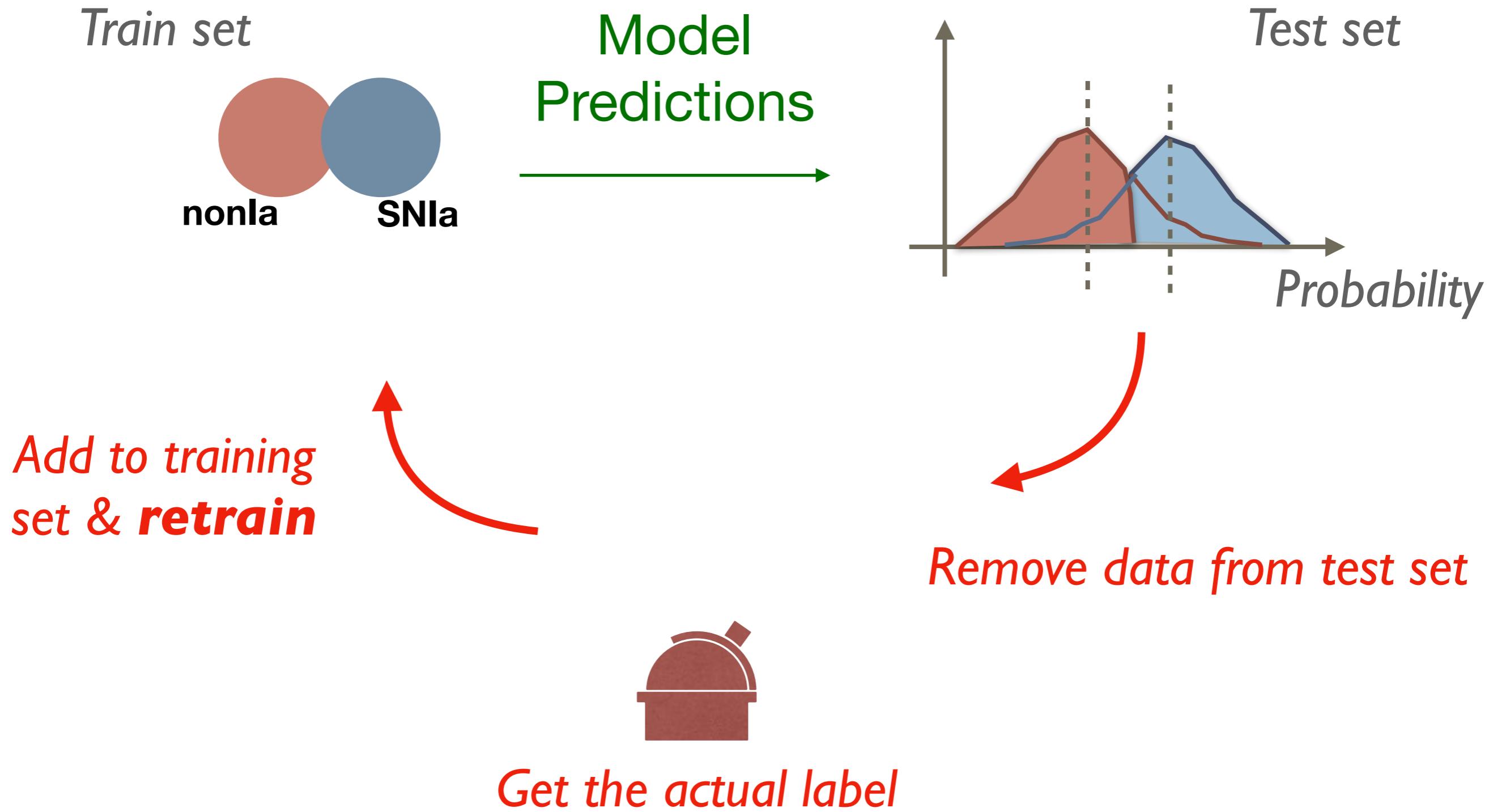
**When :** focus in the last couple of years using ZTF data

**Where :** <https://arxiv.org/pdf/2111.11438.pdf>

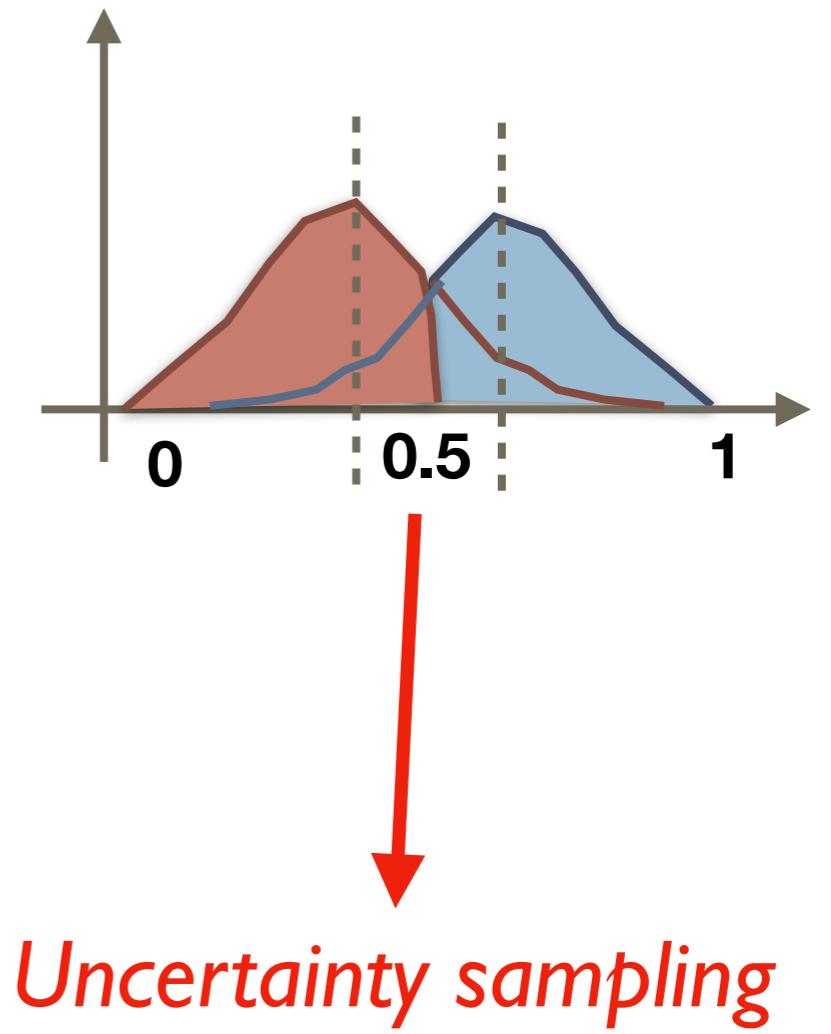
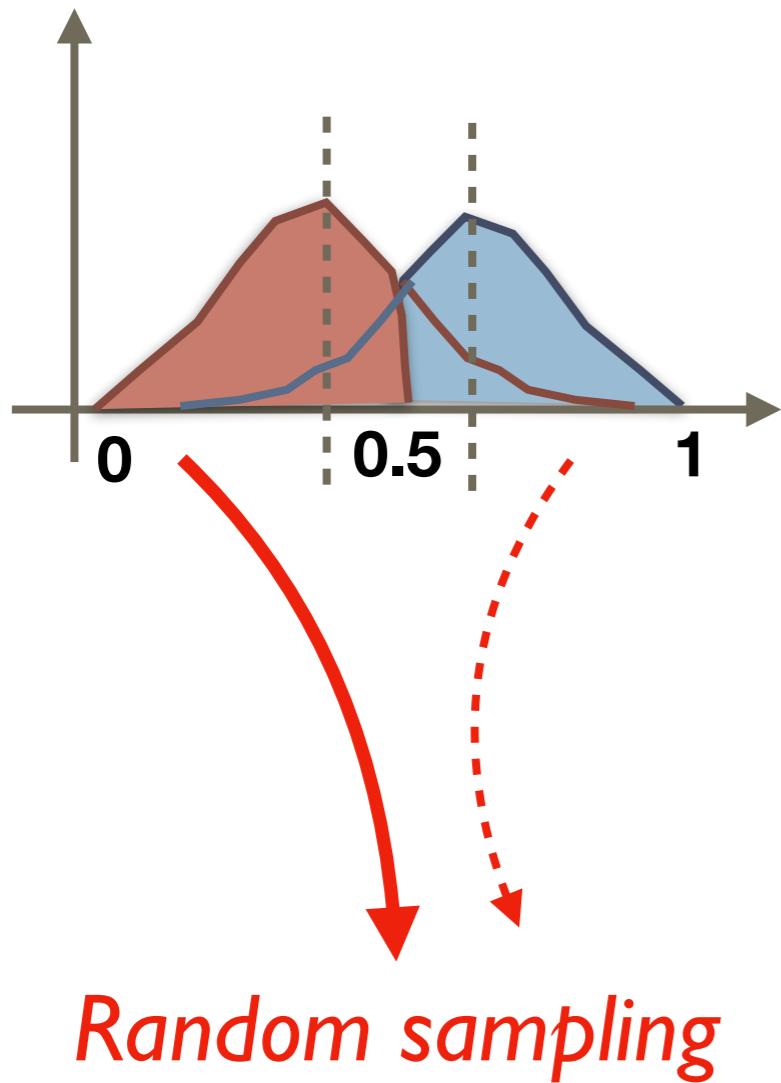
# What is active ML ?



# What is active ML ?

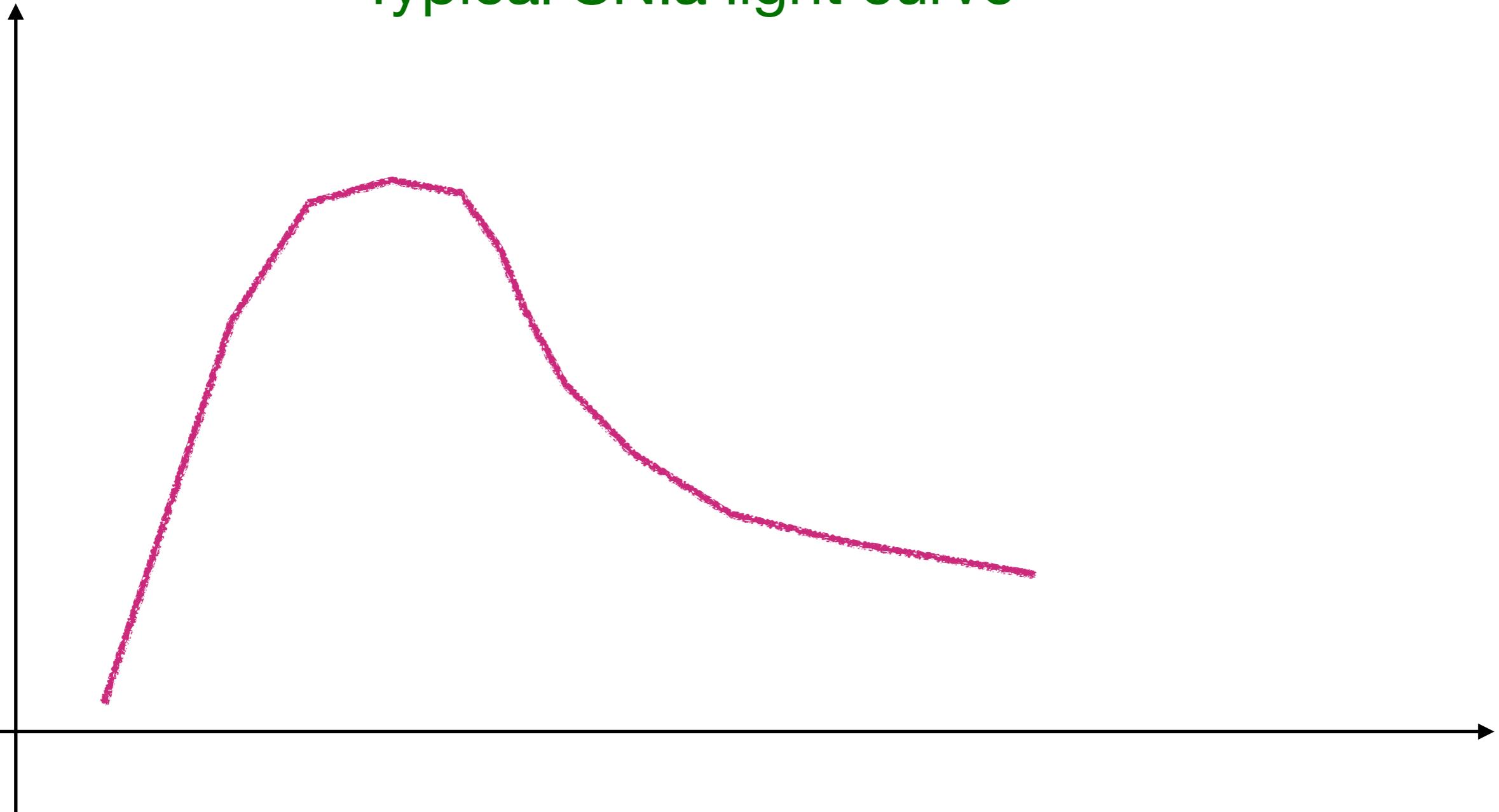


# Uncertainty vs Random Sampling



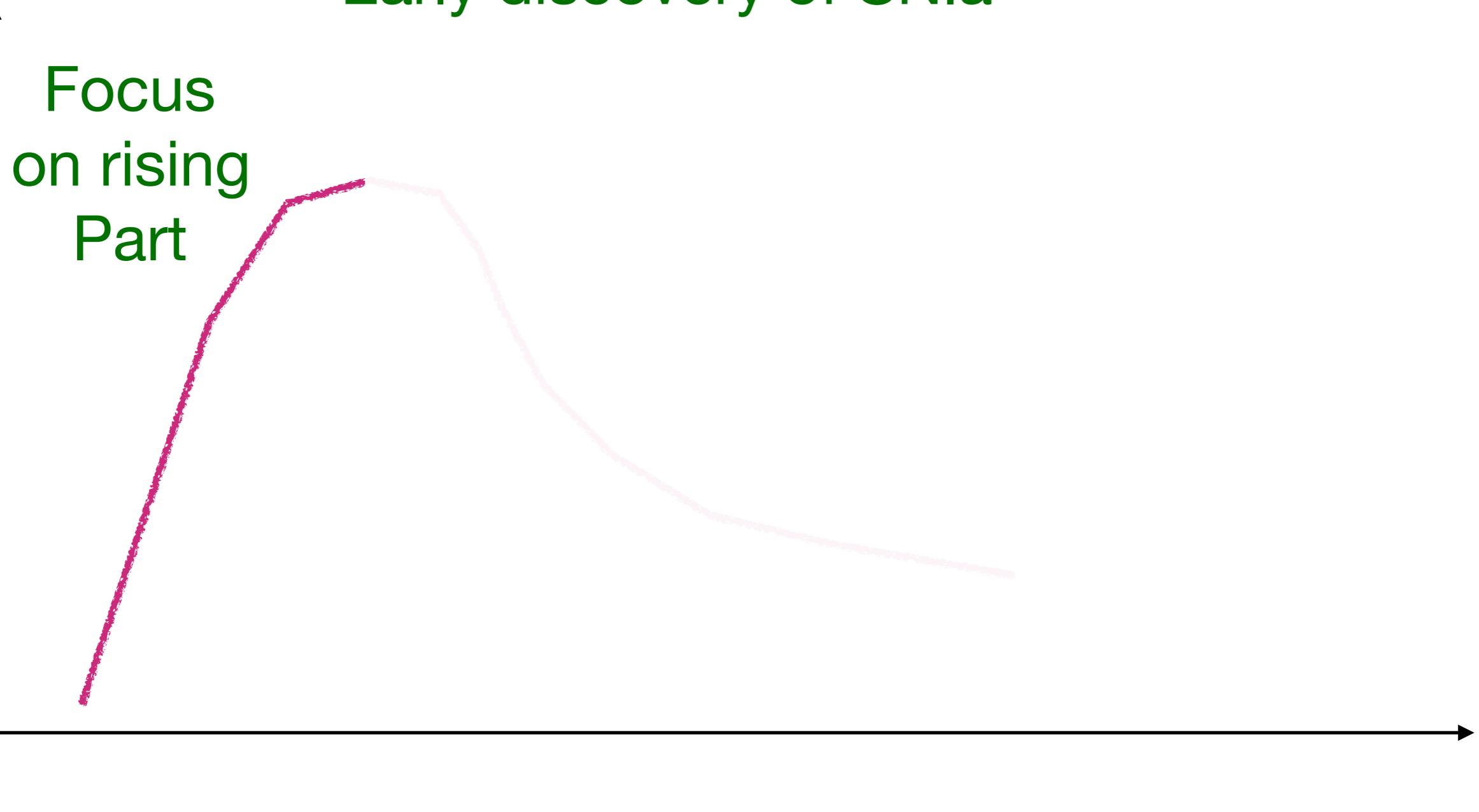
# Learning from light curves

Typical SNIa light curve



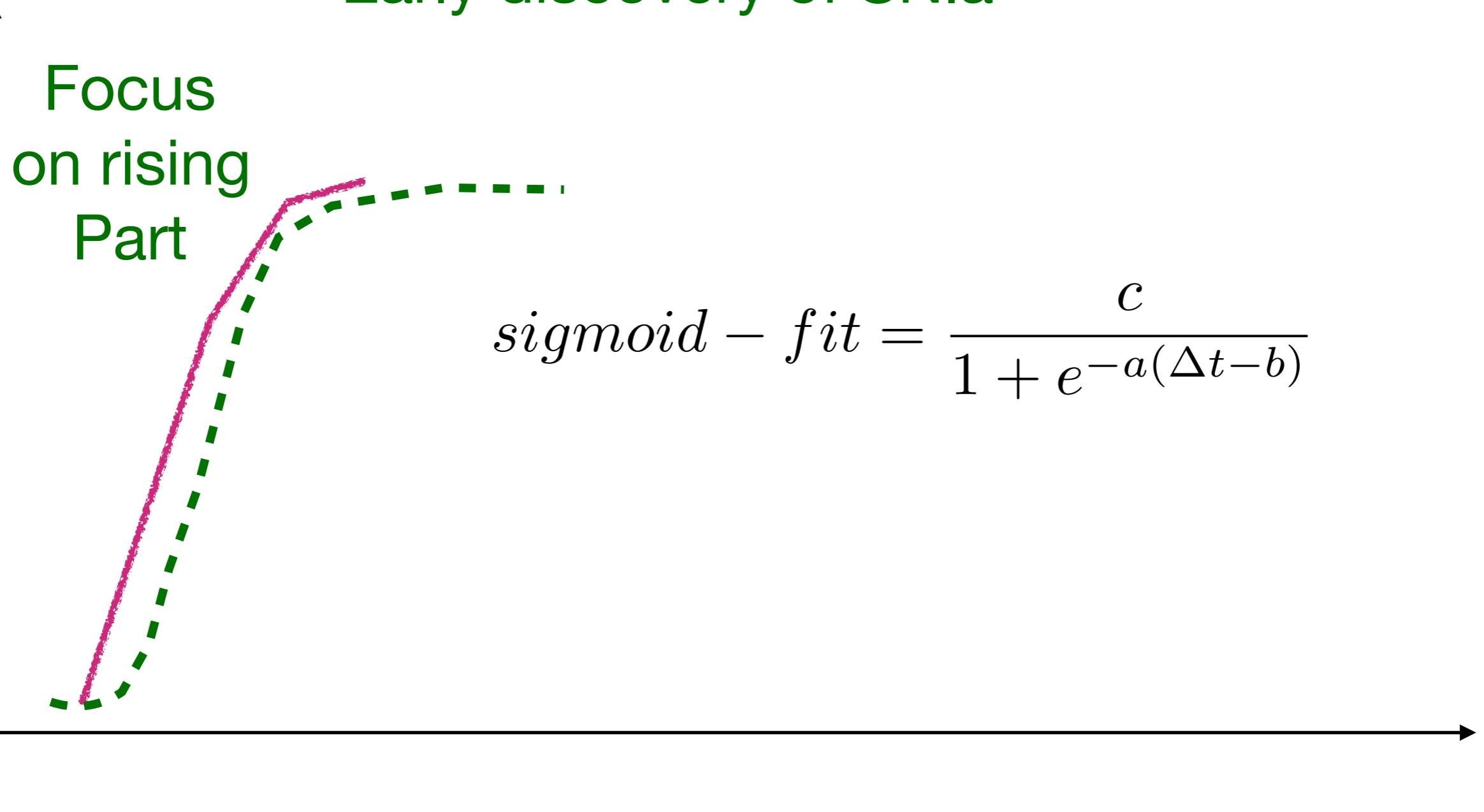
# Learning from light curves

Early discovery of SNIa

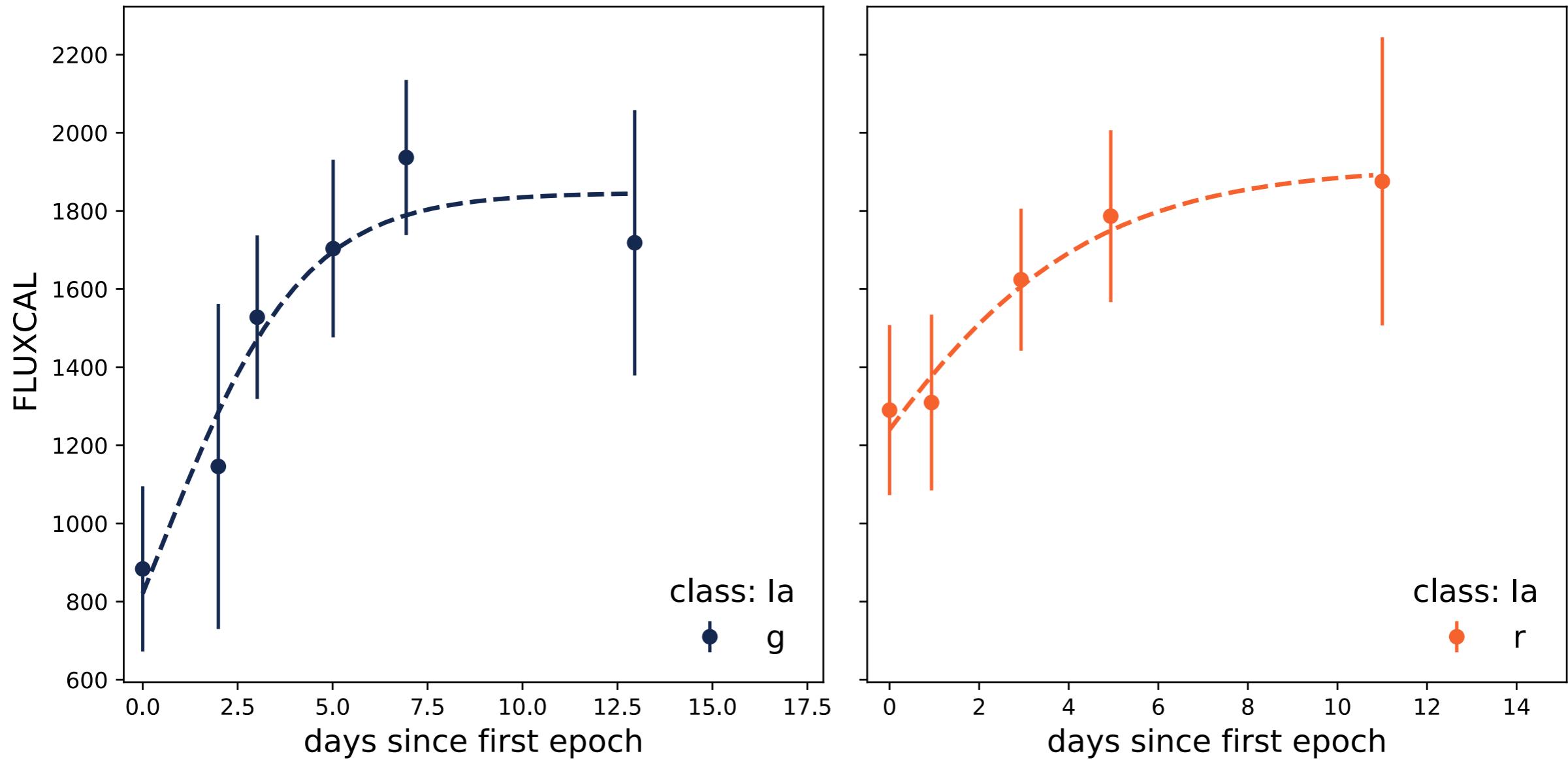


# Learning from light curves

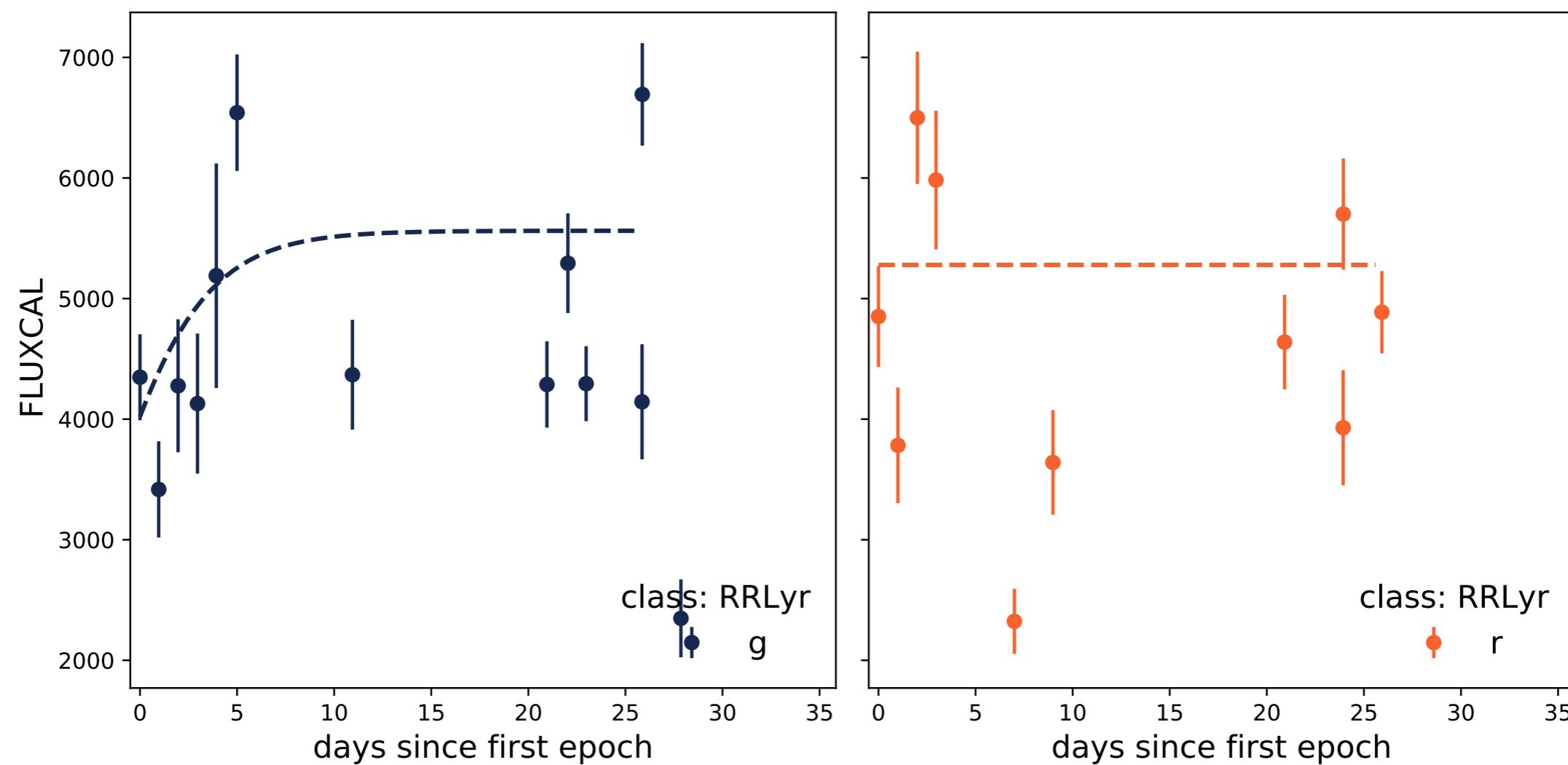
Early discovery of SNIa



# Actual data for SNIa



# ...and for nonla



Luckily nonla objects are very diverse hence the fit is pretty bad in describing such light curves.

A feature  $\chi^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i}$  describes the goodness of the fit

# Features

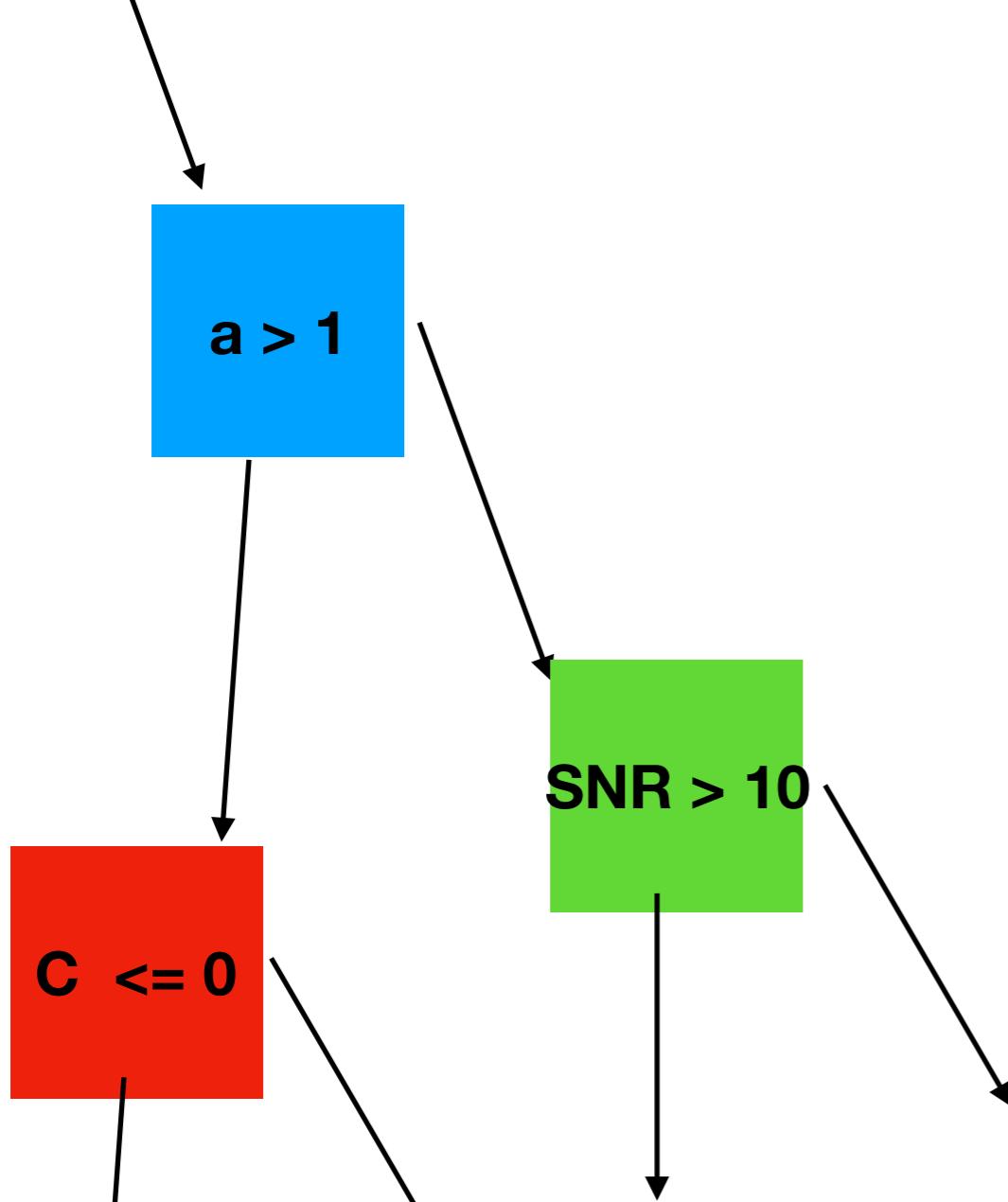
- I)  $\text{sigmoid-fit} = \frac{c}{1 + e^{-a(\Delta t - b)}}$  3 features +
- II)  $\chi^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i}$  1 feature +
- III)  $SNR = \frac{1}{N} \sum_{i=1}^N \frac{o_i}{\Delta o_i}$  1 feature+
- IV)  $N$  points in the raising part 1 feature =

---

6 features which we use to train  
RF model

# Training using random forests

an ensemble of  
decision trees



*Parameters for our training*

$N\text{-trees} = 1000$

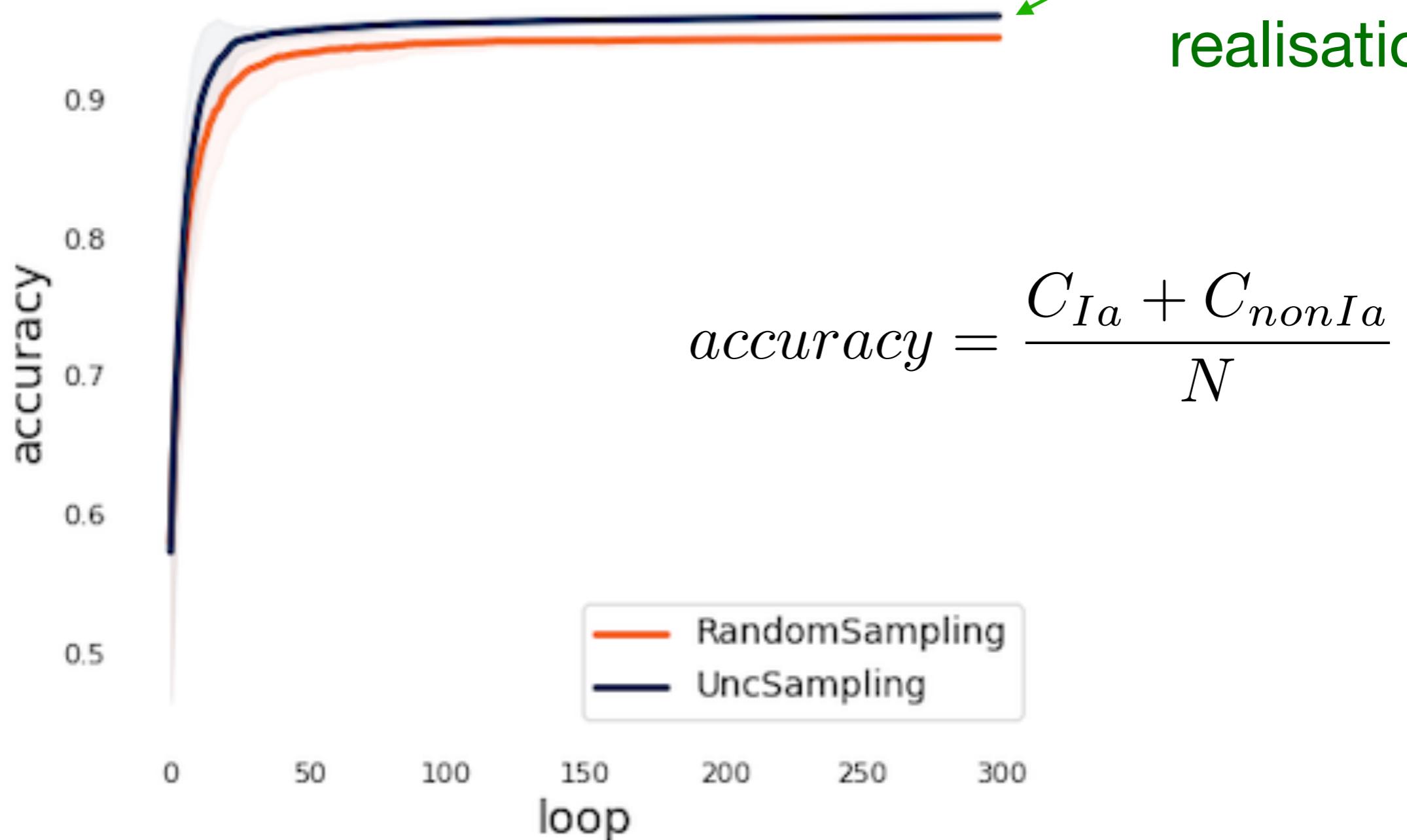
*Train Sample  $\sim 10$  alerts  
(half of which SNIa)*

*Test Sample  $\sim 23000$  alerts  
(vast majority of nonla)*

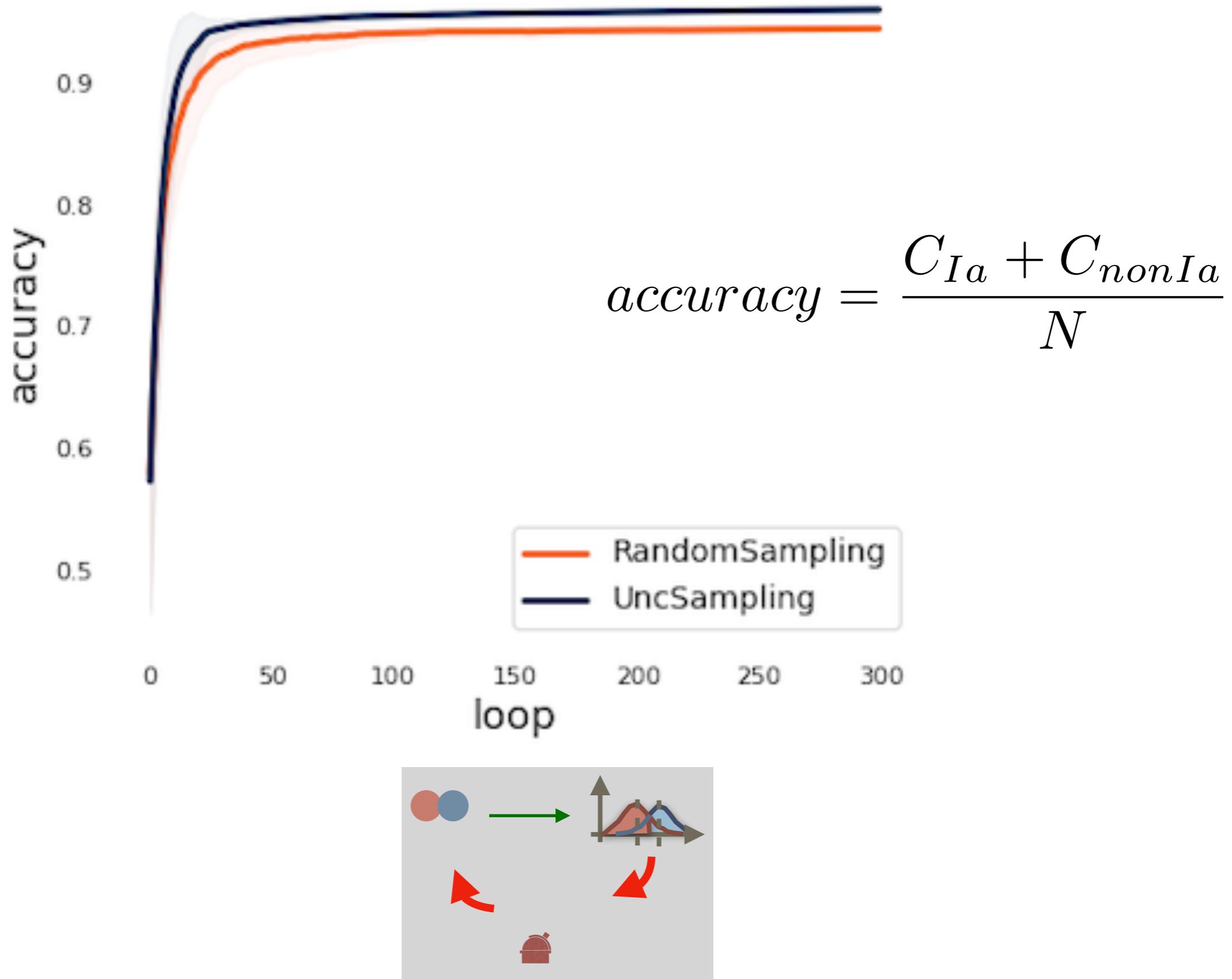
*300 steps of active learning*

# Results : accuracy

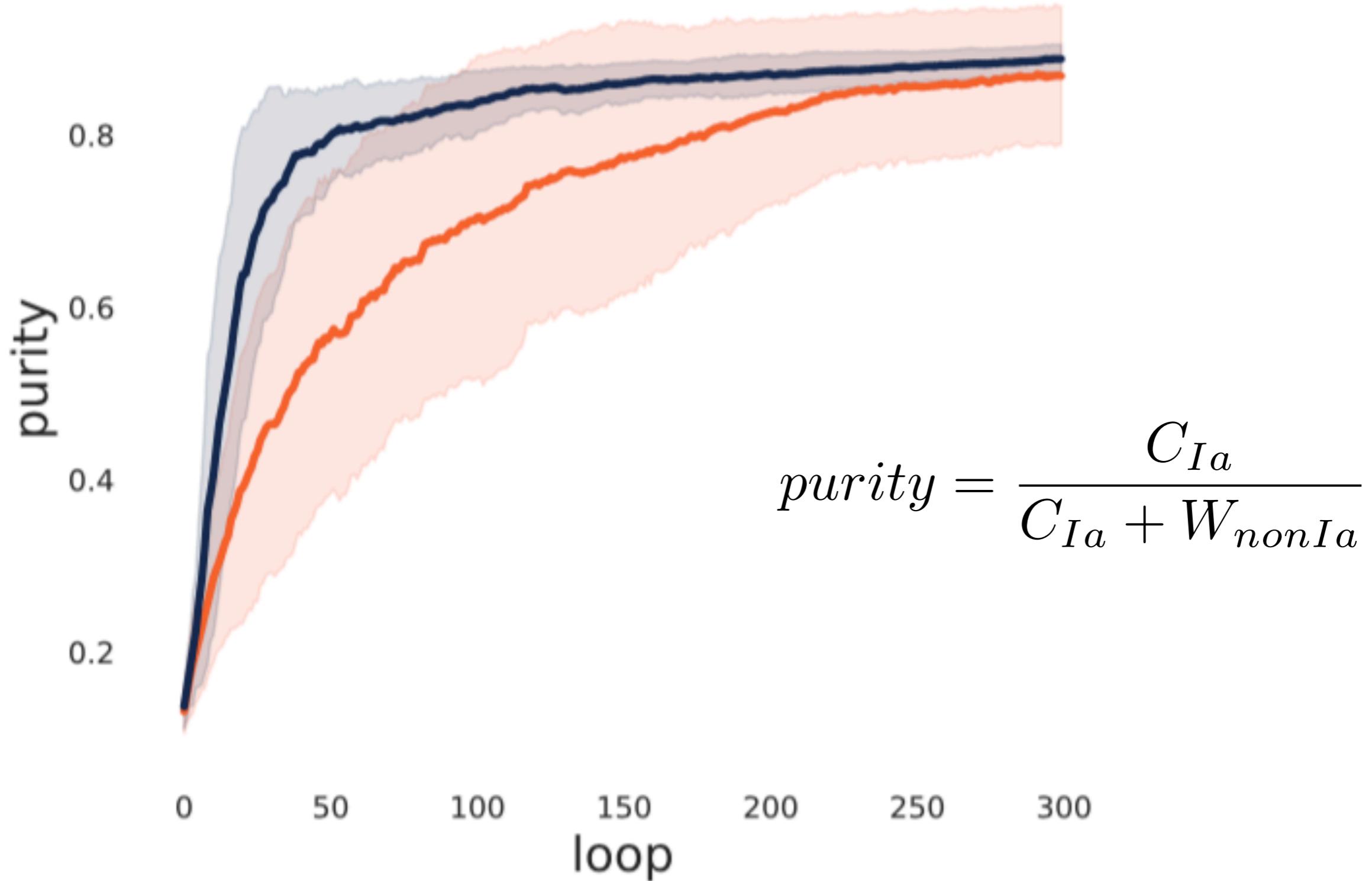
average over  
100 different  
realisations



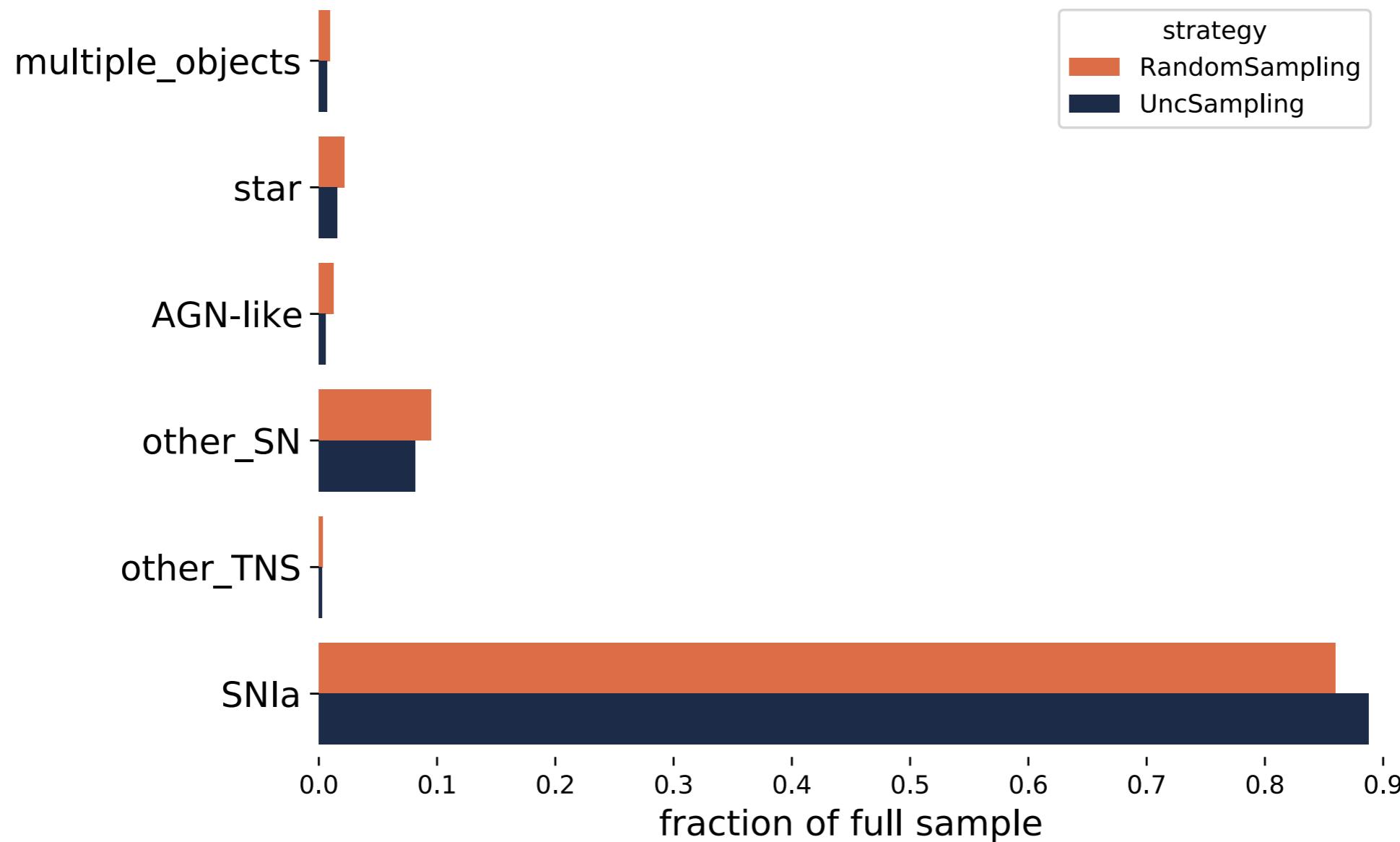
# Results : accuracy



# Results : purity



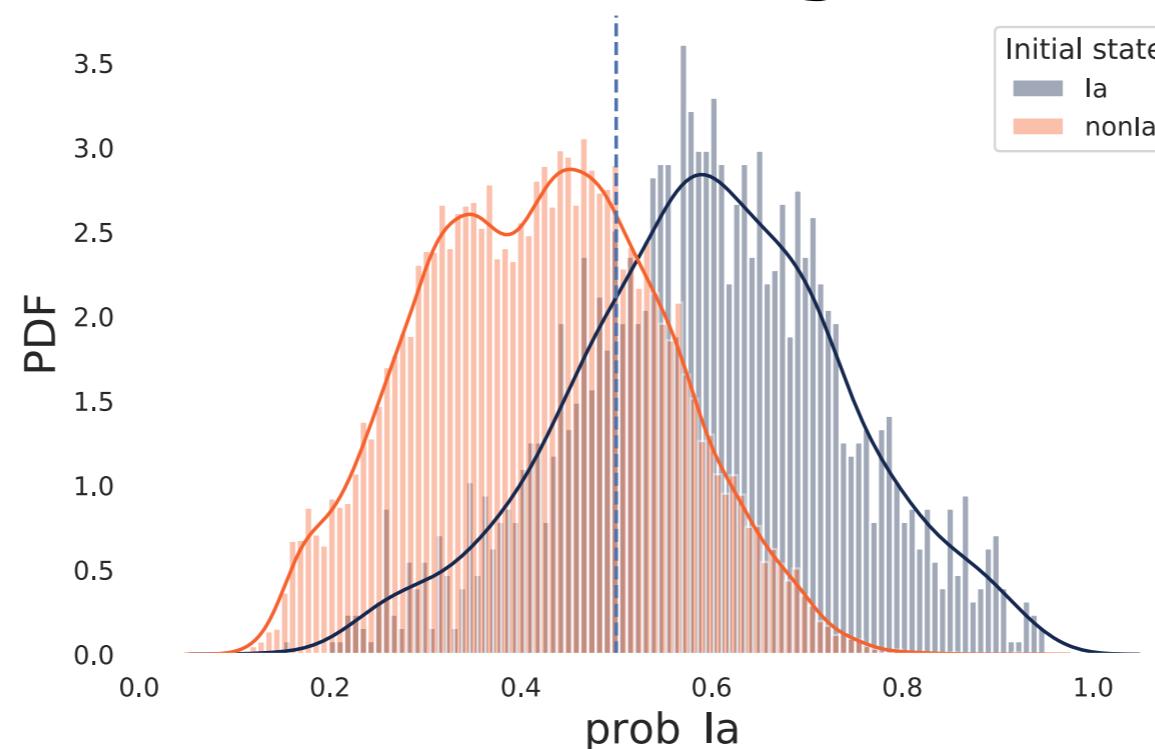
# Results : actual classes



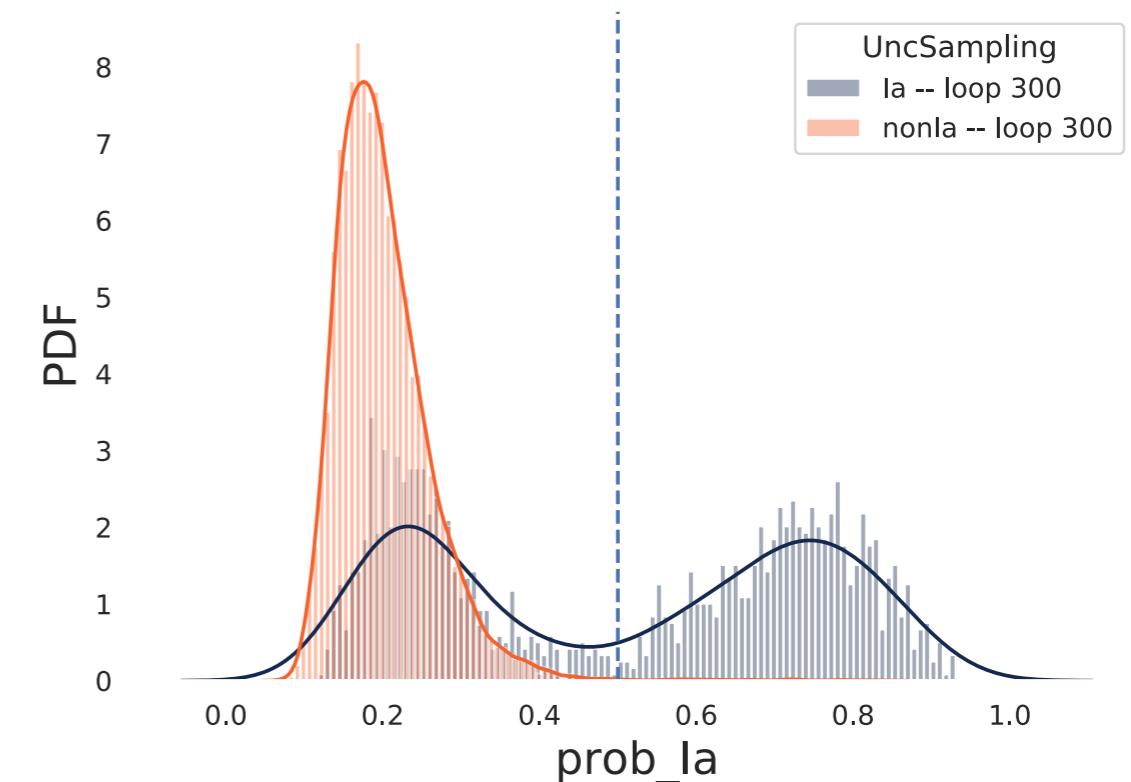
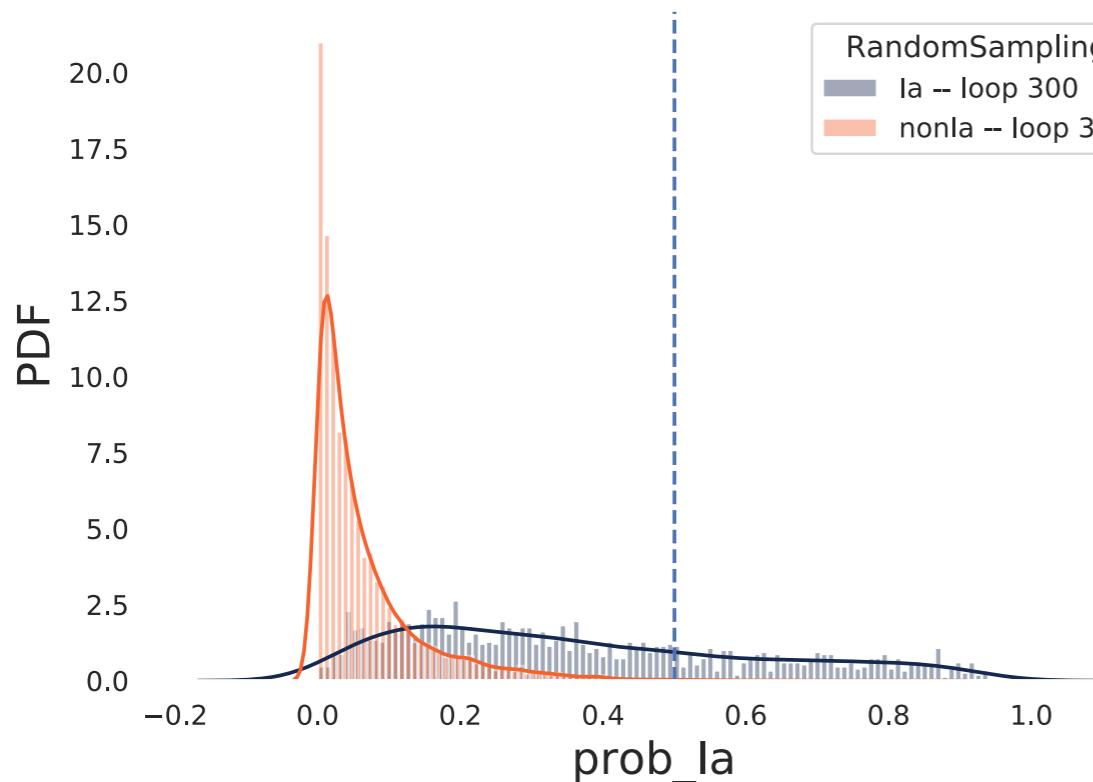
# Results : probability distribution

**Loop = 0**

•  
•  
•



**Loop = 99**



# Results : in prod

predictions are matched with those from Bayesian Neural Network models (Anais).

If both models agree predictions are sent out.

Over the period (01/November/2020 to 31/October/2021) :

806 candidates were proposed

66% of which : followed-up & spectroscopically confirmed  
86% of which : correctly classified as SNIa

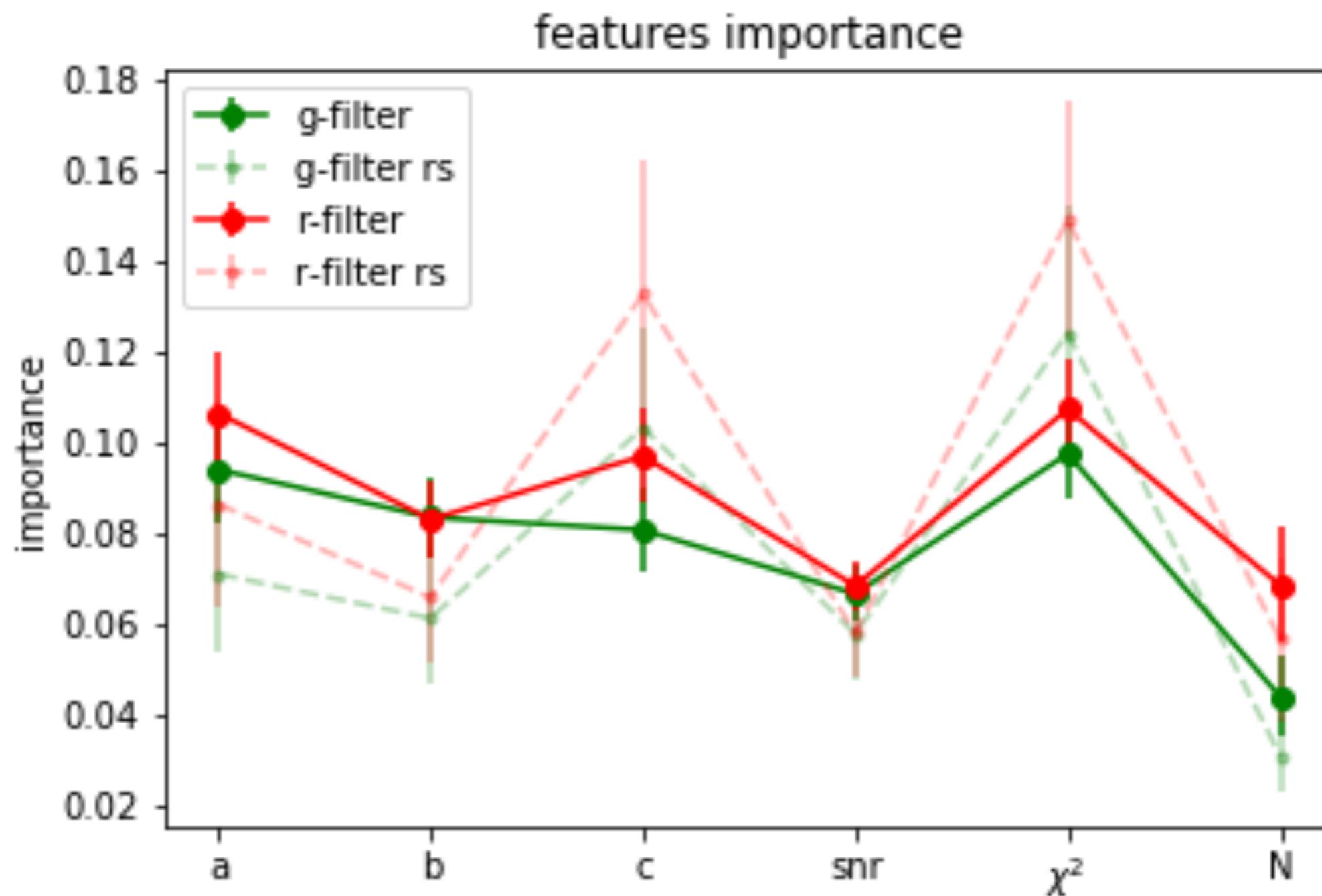
**Thanks for your attention**

# Backup

# Early discovery: difficulties

- i) from ZTF simulations, we have to reconstruct a dataset of rising light curves.
- ii) *Bazin's* fit does not work well with a few data points (more later)
- iii) sometimes there are seasonal gaps in the time series (telescope off for a season etc. )
- iv) some of the data points for the flux are negative (with values below -10)

# Features importance



# Estimate of how many days before max ?

1. typical length of a light curve last : 30 data points ?
2. typical length of rising part ? (1/3 of the light curve i.e. 10 points)
3. Typical prediction sent out ? (probably 5 points on average)
4. Typical N points before max ? ( $10 - 5 = 5?$  )
5. Confirm with predicted data from the broker ?