LSST DESC & COIN

RESSPECT

Recommendation System for Spectroscopic Follow-up

Project update

Emille E. O. Ishida on behalf of the RESSPECT team

LSST France, 4 February 2020

Context:

- There will be a lot of photometric candidates without spectroscopic confirmation

Context:

- There will be a lot of photometric candidates without spectroscopic confirmation
- We would like to use them for something

Context:

- There will be a lot of photometric candidates without spectroscopic confirmation
- We would like to use them for something
- ML-based SN photometric classification has attracted a lot of attention

Context:

- There will be a lot of photometric candidates without spectroscopic confirmation
- We would like to use them for something
- ML-based SN photometric classification has attracted a lot of attention
- ML-based strategies rely on good training sets

Context:

- There will be a lot of photometric candidates without spectroscopic confirmation
- We would like to use them for something
- ML-based SN photometric classification has attracted a lot of attention
- ML-based strategies rely on good training sets
- COIN suggested an Active Learning Strategy

CRP #4, Clermont Ferrand 2017

Context:

- There will be a lot of photometric candidates without spectroscopic confirmation
- We would like to use them for something
- ML-based SN photometric classification has attracted a lot of attention
- ML-based strategies rely on good training sets
- COIN (CRP #4 team) suggested an Active Learning Strategy
- DESC is interested in cosmology

COIN
Cosmostatistics Initiative

LSST DESC
Dark Energy Science Collaboration

Context:

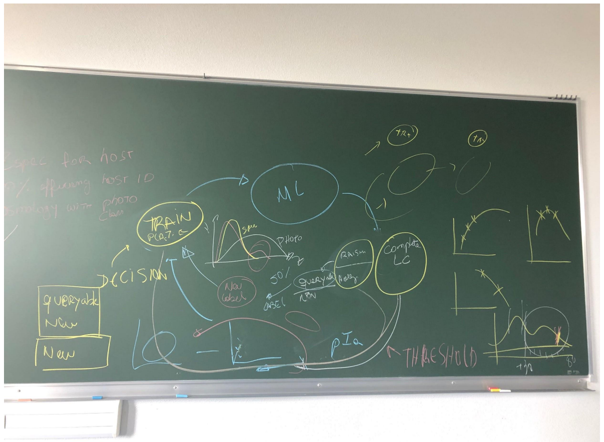
- There will be a lot of photometric candidates without spectroscopic confirmation
- We would like to use them for something
- ML-based SN photometric classification has attracted a lot of attention
- ML-based strategies rely on good training sets
- COIN (CRP #4 team) suggested an Active Learning Strategy
- DESC is interested in cosmology
- April/2019 an inter-collaboration agreement was signed

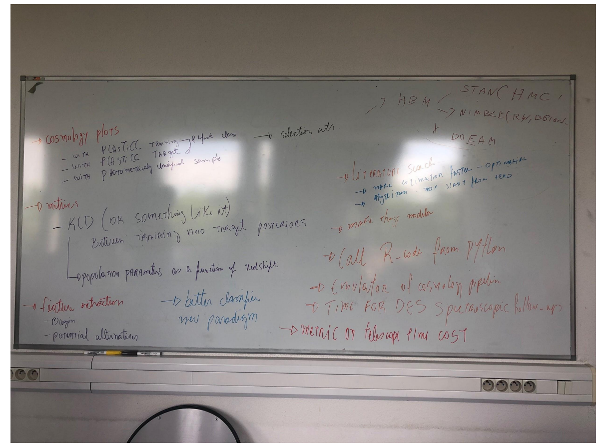
COIN
Cosmostatistics Initiative

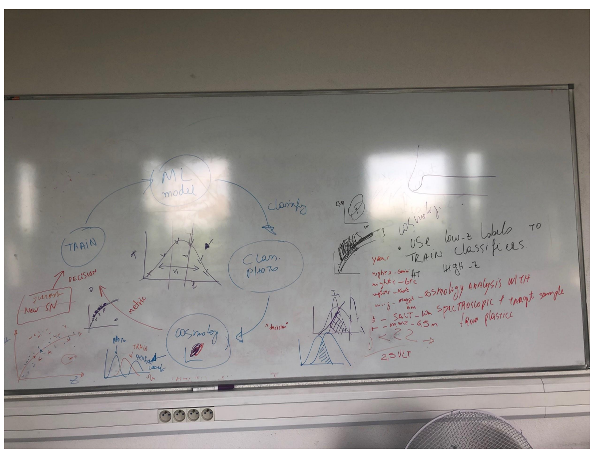
LSST DESC
Dark Energy Science Collaboration

LSST DESC & COIN
RESSPECT
Recommendation System for Spectroscopic Follow-up

1st RESSPECT meeting:
July 2019 - Clermont Ferrand, France







https://cosmostatistics-initiative.org/focus/resspect1/

Goal:

Build a recommendation system that optimize photometric supernova cosmology results

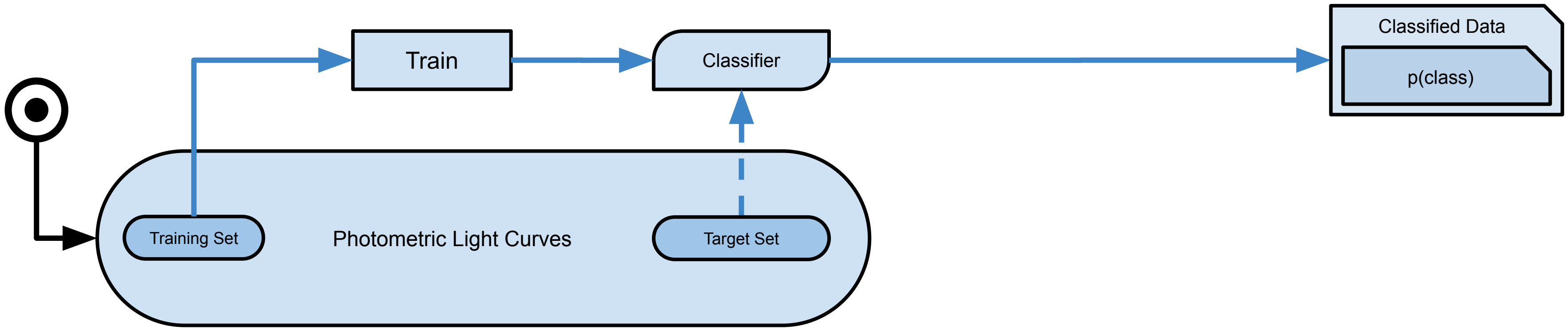
Current work structure:

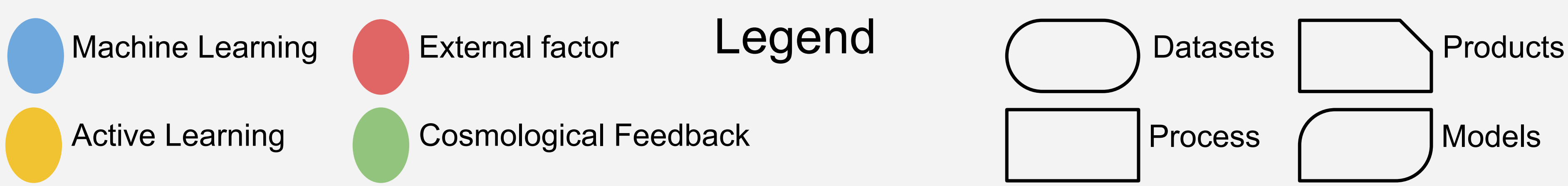
Number of researchers subscribed to the project: **29**
(14 DESC, 10 COIN, 5 both)

Minutes document in confluence
From July-December 2019: 1 face-to-face meeting and
3 telecons

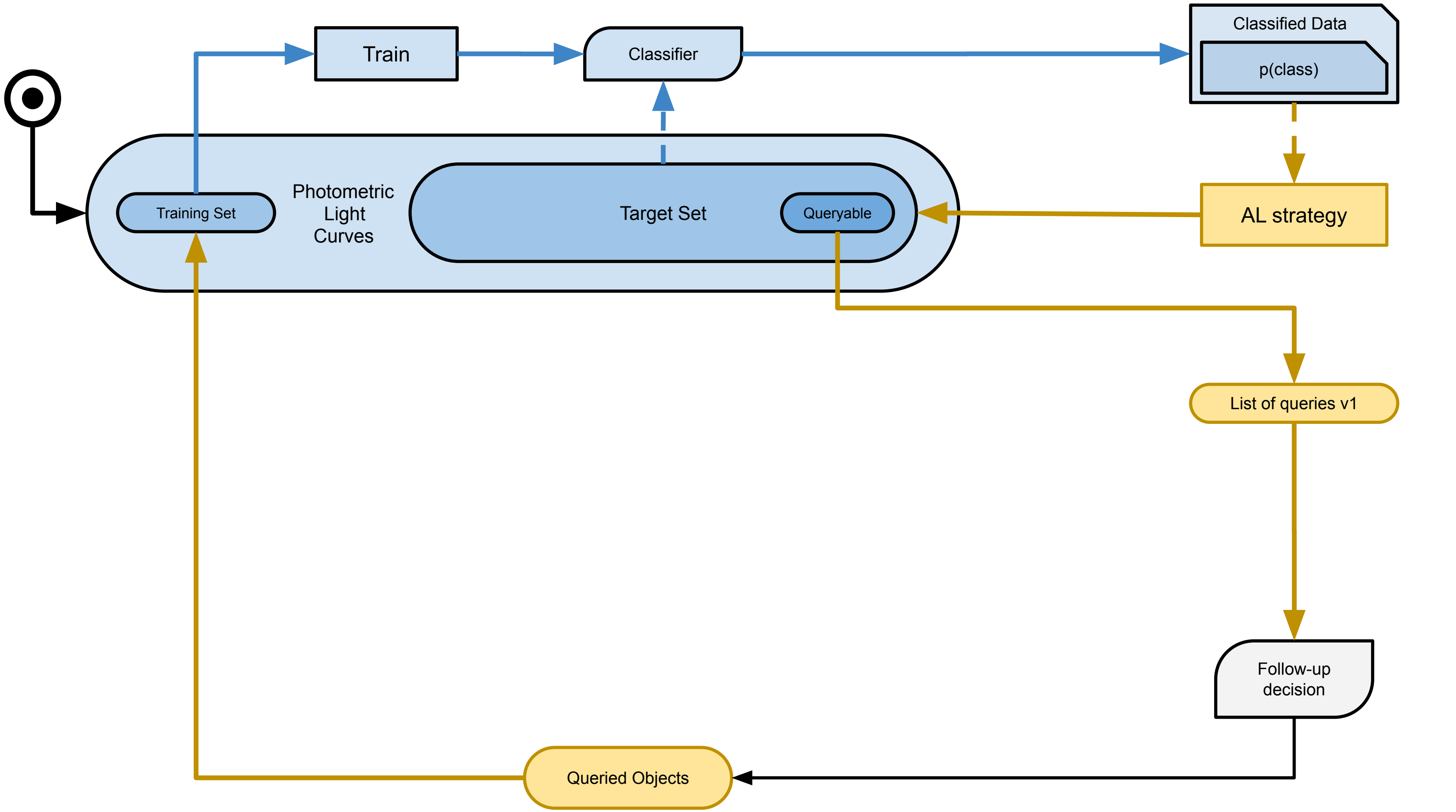
- Project design : Bruno Quint, Alex Malz
- Simulations: David Jones, Mi Dai, Anais Moller, Maria Vincenzi
- Classifiers: Sreevarsha Sreejith, Noble Kennamer, Anais Moller, Bruno Quint
- Active Learning Strategy: Noble Kennamer
- Spectroscopic requirements: Santiago Gonzalez-Gaitan, Anais Moller, Lluis Galbany, Alberto Krone-Martins, Kara Ponder
- Cosmology based metric: Kara Ponder
- Full cosmology pipeline: Rafael S. de Souza, David Jones, Mi Dai

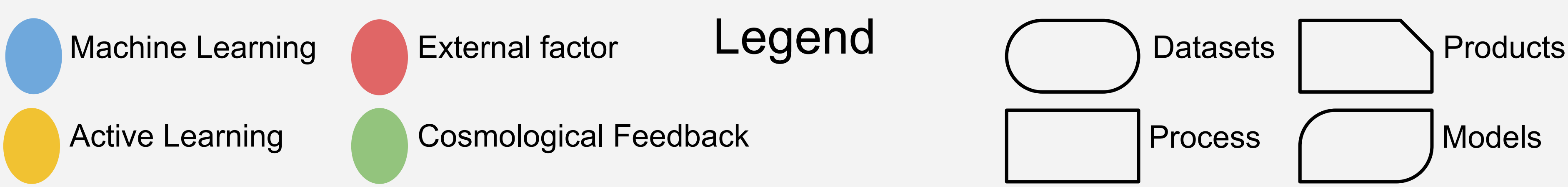
Machine Learning



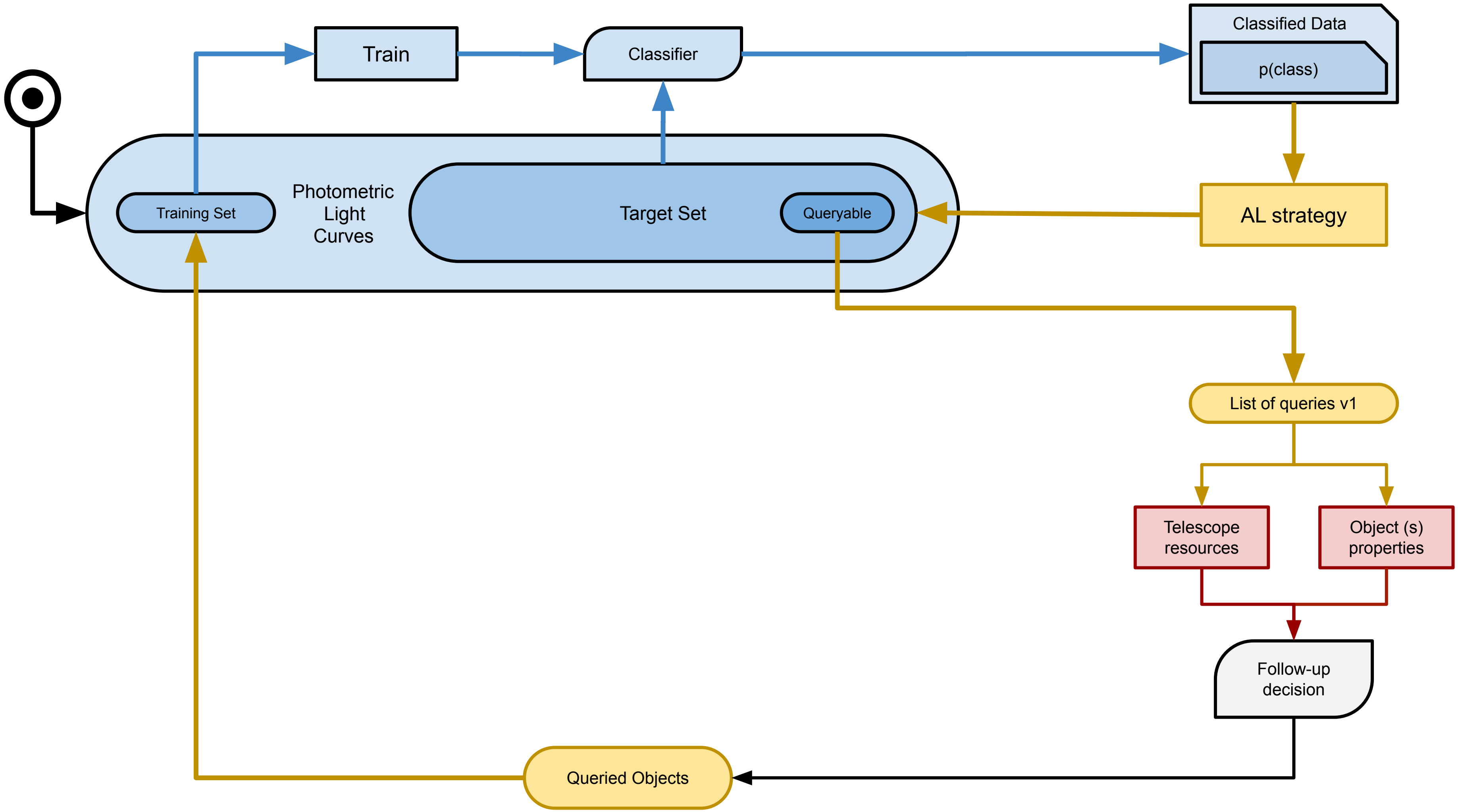


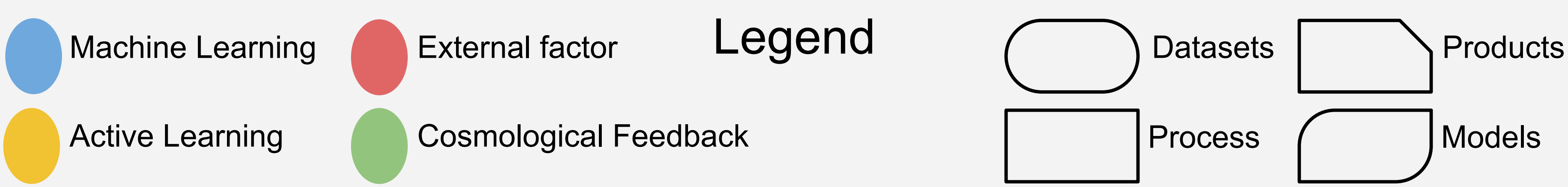
Active Learning



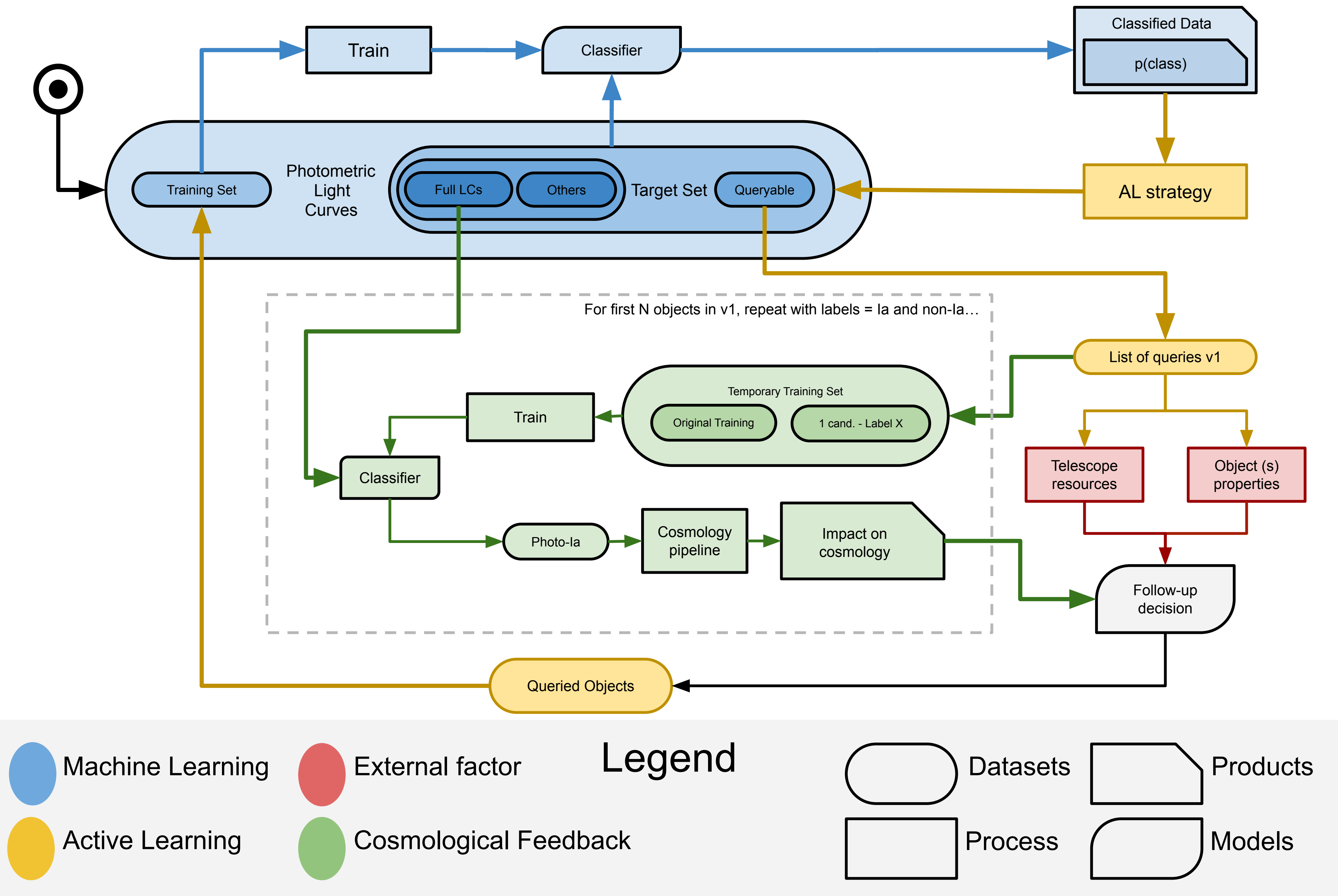


External Factors





Cosmological Feedback



Simulations:

by David Jobes (UCSC)
Mi Dai (Rutgers U.)
and Maria Vincenzi (U. Portsmouth)

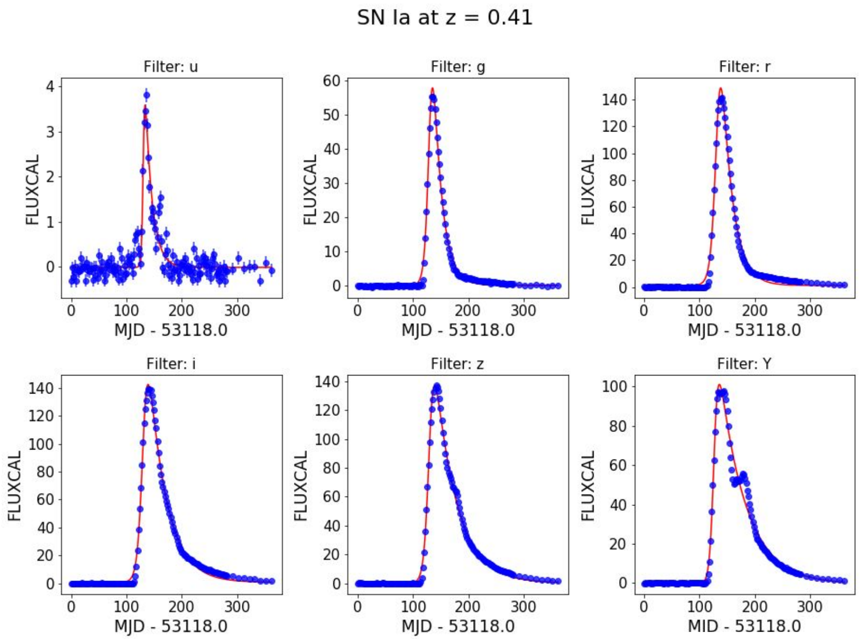
- **Stage 0**: Perfect observation conditions and 1-day cadence.
- **Stage 1**: Realistic observation conditions and 1-day cadence.
- **Stage 2**: Realistic observation conditions and uniform cadence (3, 5, 10 days).
- **Stage 3**: Realistic observation conditions and a couple of proposed LSST cadences.

Table 1: Summary of the status of simulations.

<table>
<tr><th>Simulation</th><th>Models</th><th>Spec/
Train selection</th><th>Status</th></tr>
<tr><td>Stage 0</td><td rowspan="3">PLAsTiCC[1]: Iax, Ibc (MOSFIT), Ia, 91bg, IIn and II (NMF), ILOT, CART
Vincenzi et al. (2019): Ibc and II</td><td>i-band mag< 24</td><td>Done</td></tr>
<tr><td>Stage 1</td><td rowspan="2">PLAsTiCC SelecFunc</td><td rowspan="2">in validation</td></tr>
<tr><td>Stage 2</td></tr>
<tr><td>Stage 3</td><td>TBD</td><td>TBD</td><td>Not started</td></tr>
</table>

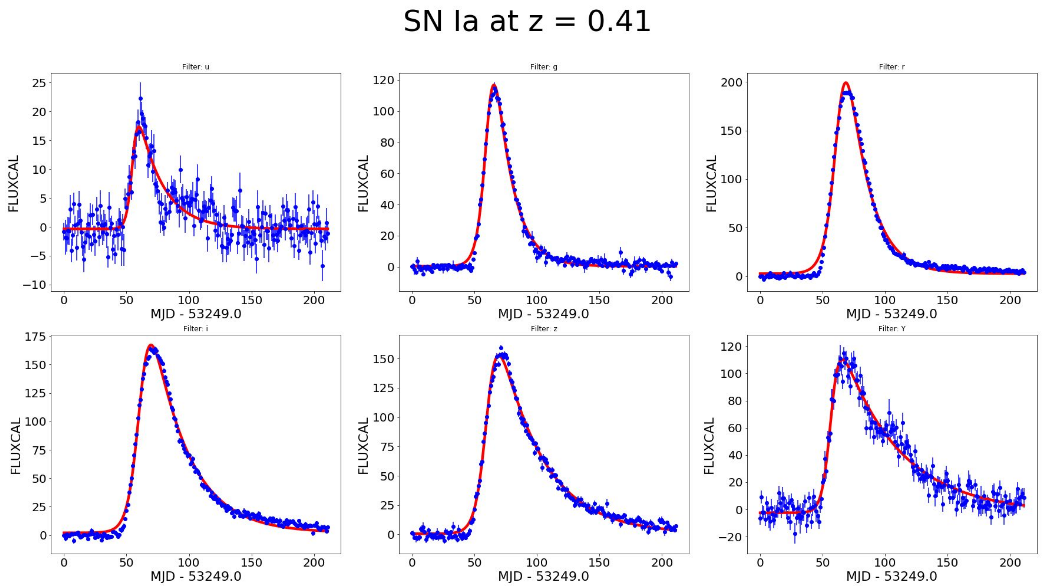
Simulations:

by David Jobes (UCSC)
and Mi Dai (Rutgers U.)

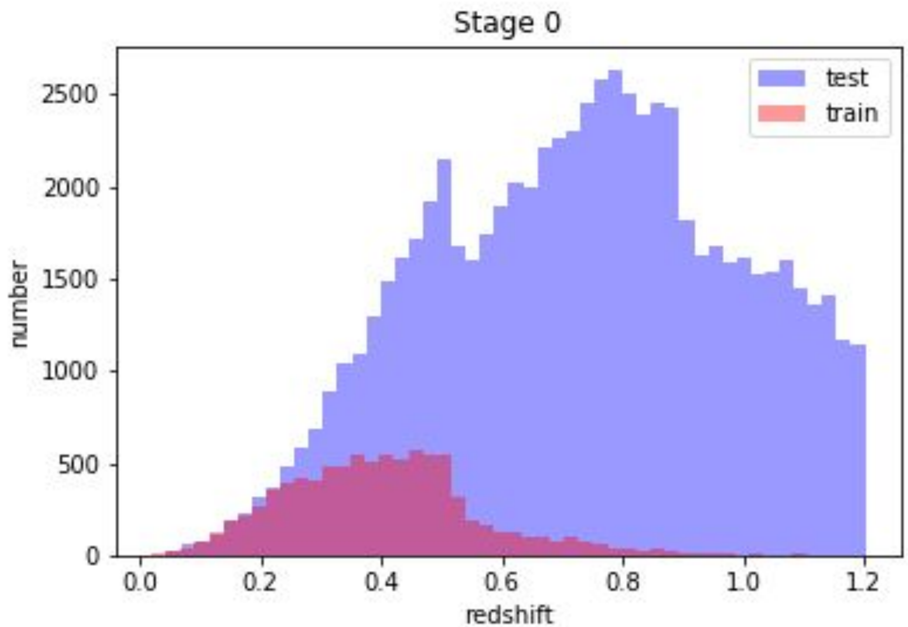


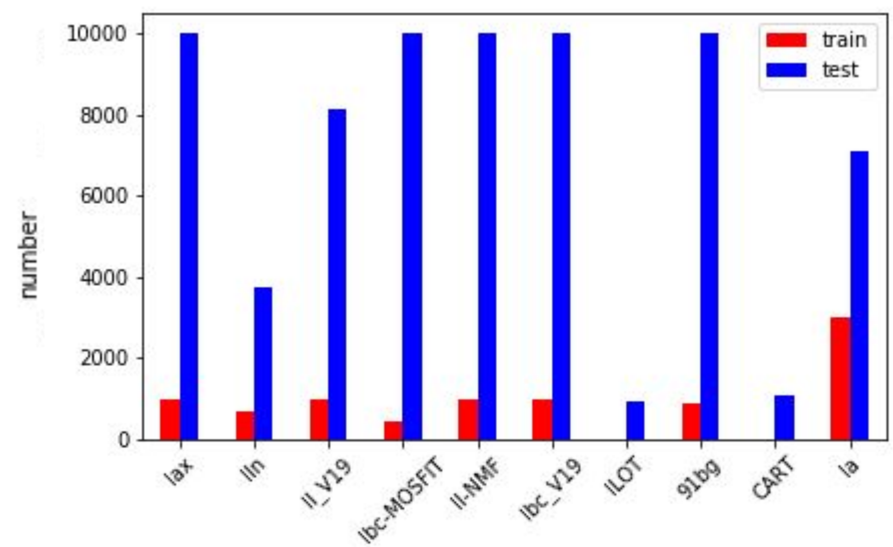
Stage 0

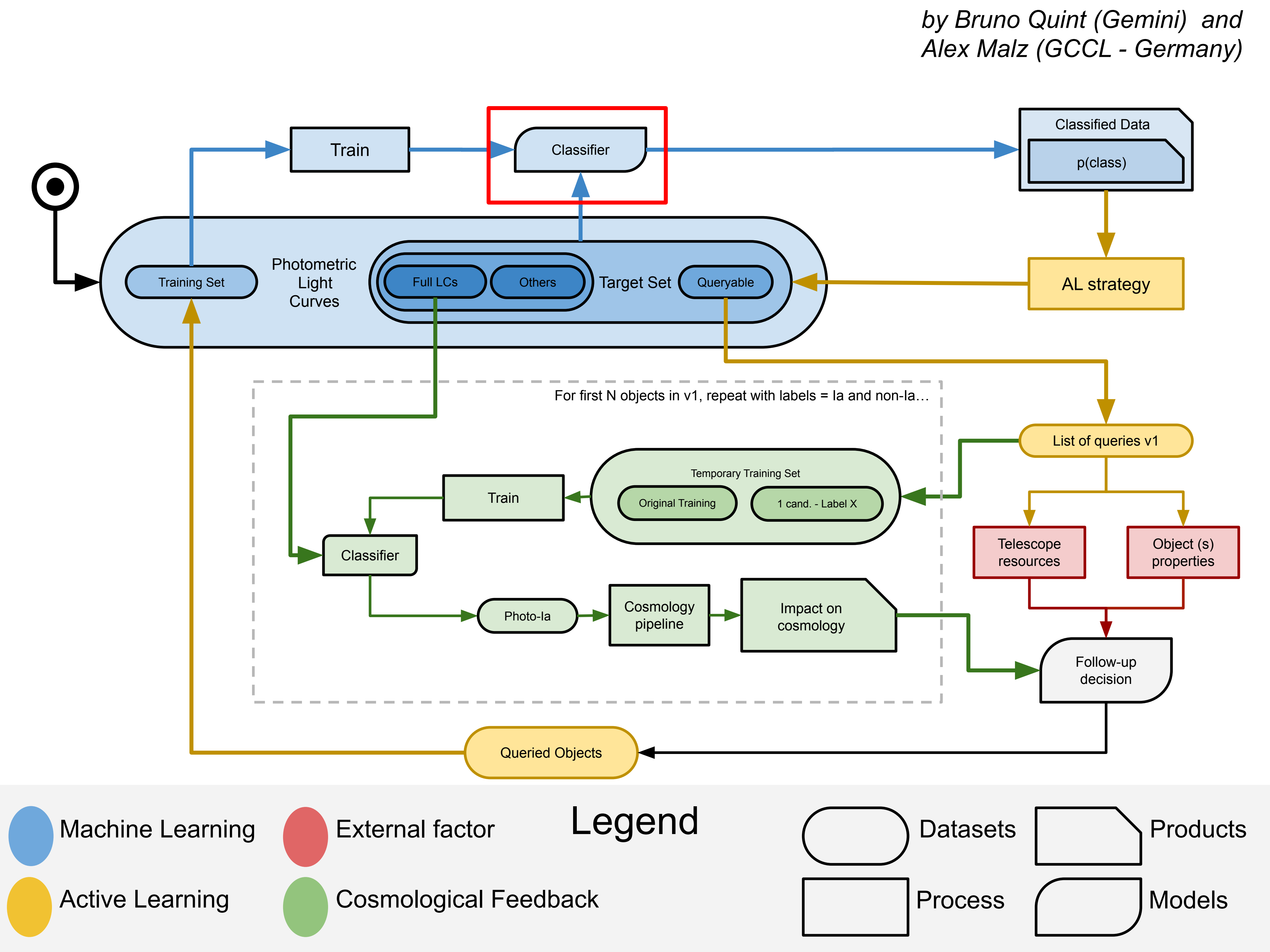
Stage 1



Simulations - Stage 0:







by Bruno Quint (Gemini) and
Alex Malz (GCCL - Germany)
Train
Classifier
Classified Data
p(class)
AL strategy
Training Set
Photometric Light Curves
Full LCs
Others
Target Set
Queryable
For first N objects in v1, repeat with labels = Ia and non-Ia…
List of queries v1
Temporary Training Set
Original Training
1 cand. - Label X
Train
Classifier
Telescope resources
Object (s) properties
Photo-Ia
Cosmology pipeline
Impact on cosmology
Follow-up decision
Queried Objects
Legend
Machine Learning
External factor
Datasets
Products
Active Learning
Cosmological Feedback
Process
Models

Classifiers on Stage 0 simulations:

by Sreevarsha Sreejith
LPC - Clermont

Tree base algorithms:
- Random Forest
- Gradient boosting
- Extremely randomized trees
- Deep Forest (Zhou and Feng, 2017))

Nearest Neighbors
Support Vector Machines
Multi-layer Precepton
Naive Bayes.

We confirmed previous results that t**ree-based methods** produced better results

1 classifier better than decision tree-based algorithms

by Noble Kennamer

Bazin fits
MC Drop-out
Mutual information
On SNPCC data

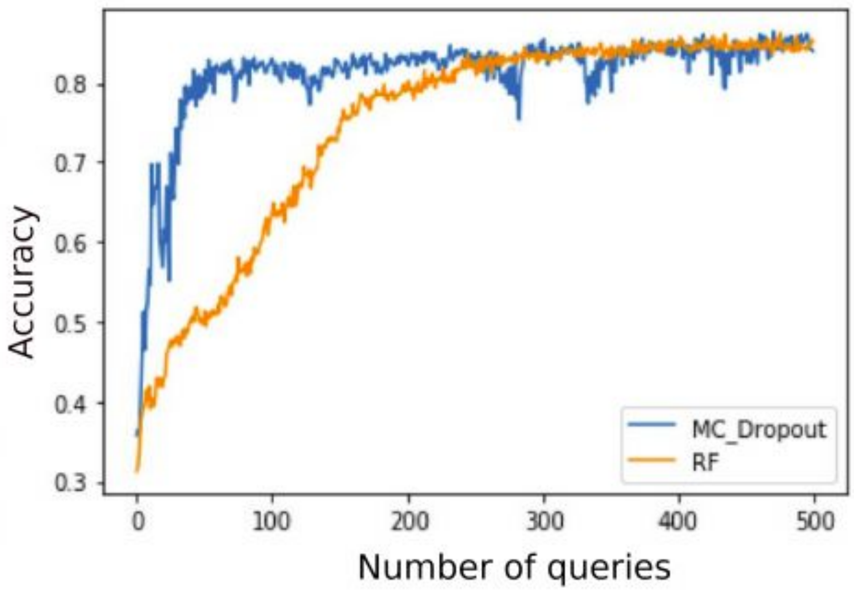
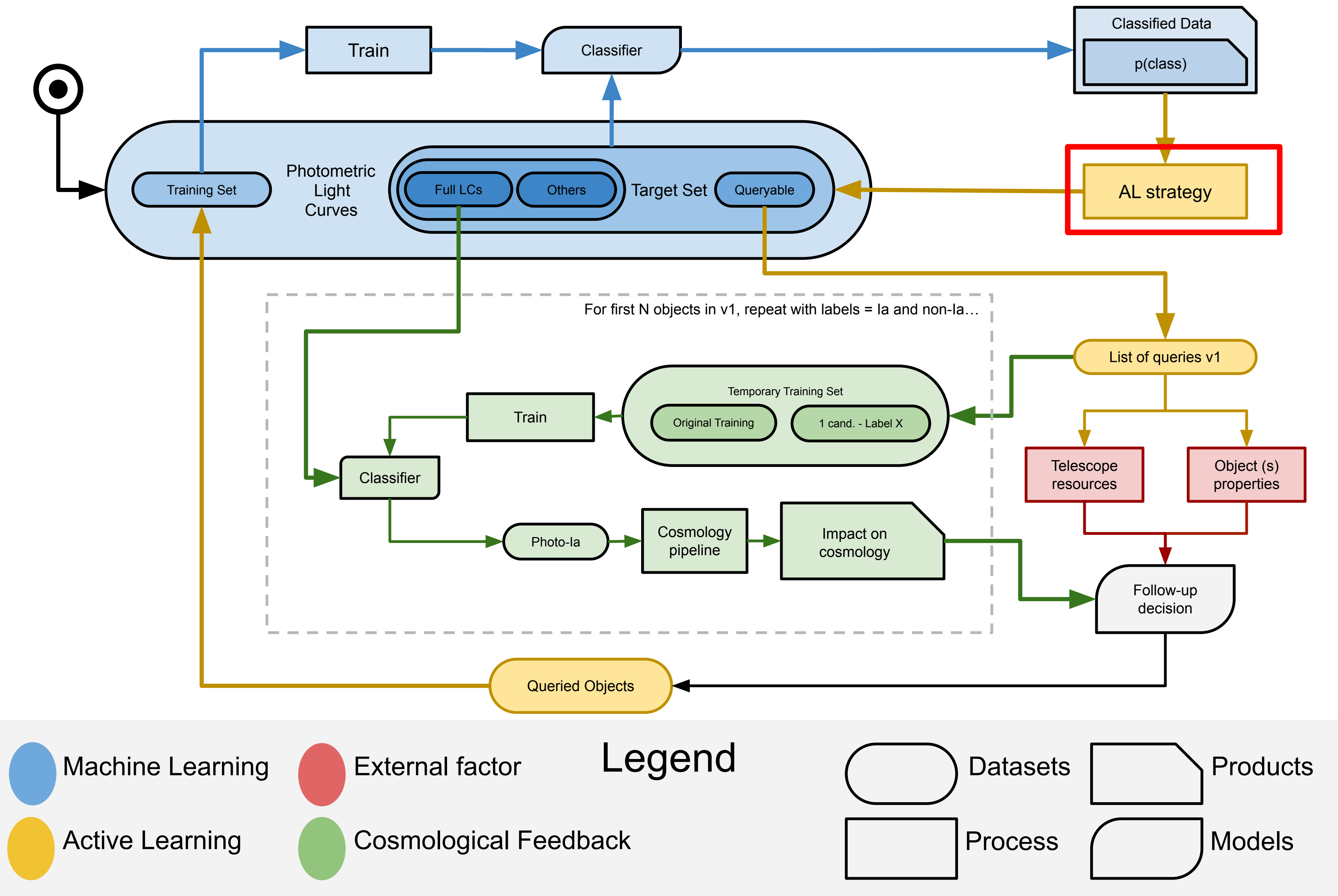


Figure 8: Results from using a MC-Dropout strategy in the context of active learning. This study was independently developed by the UCI group. Figure by Noble Kennamer.

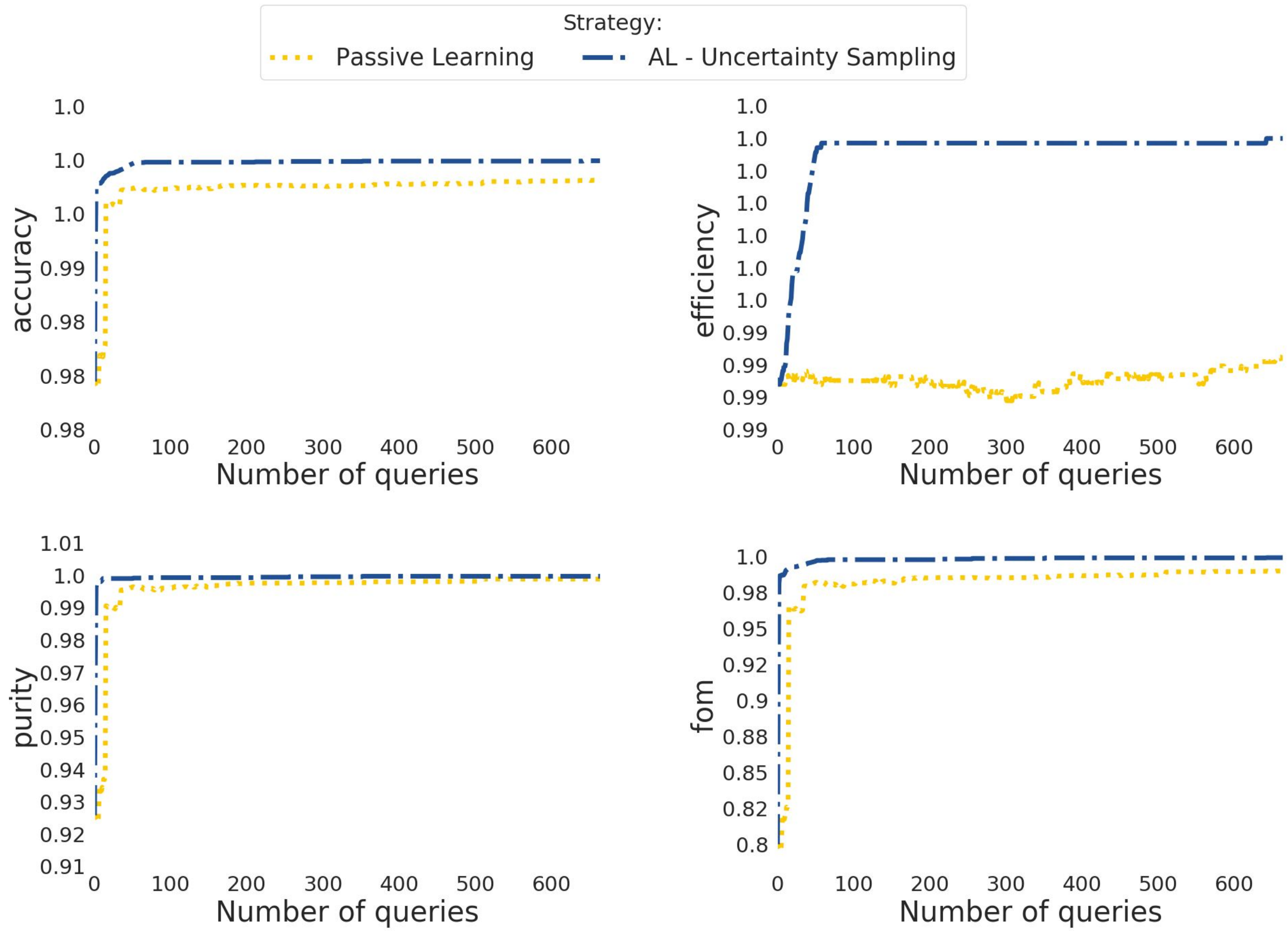
by Bruno Quint (Gemini) and Alex Malz (GCCL - Germany)



Classifiers on Stage 0 simulations:

Bazin fits
Random Forest
Uncertainty sampling

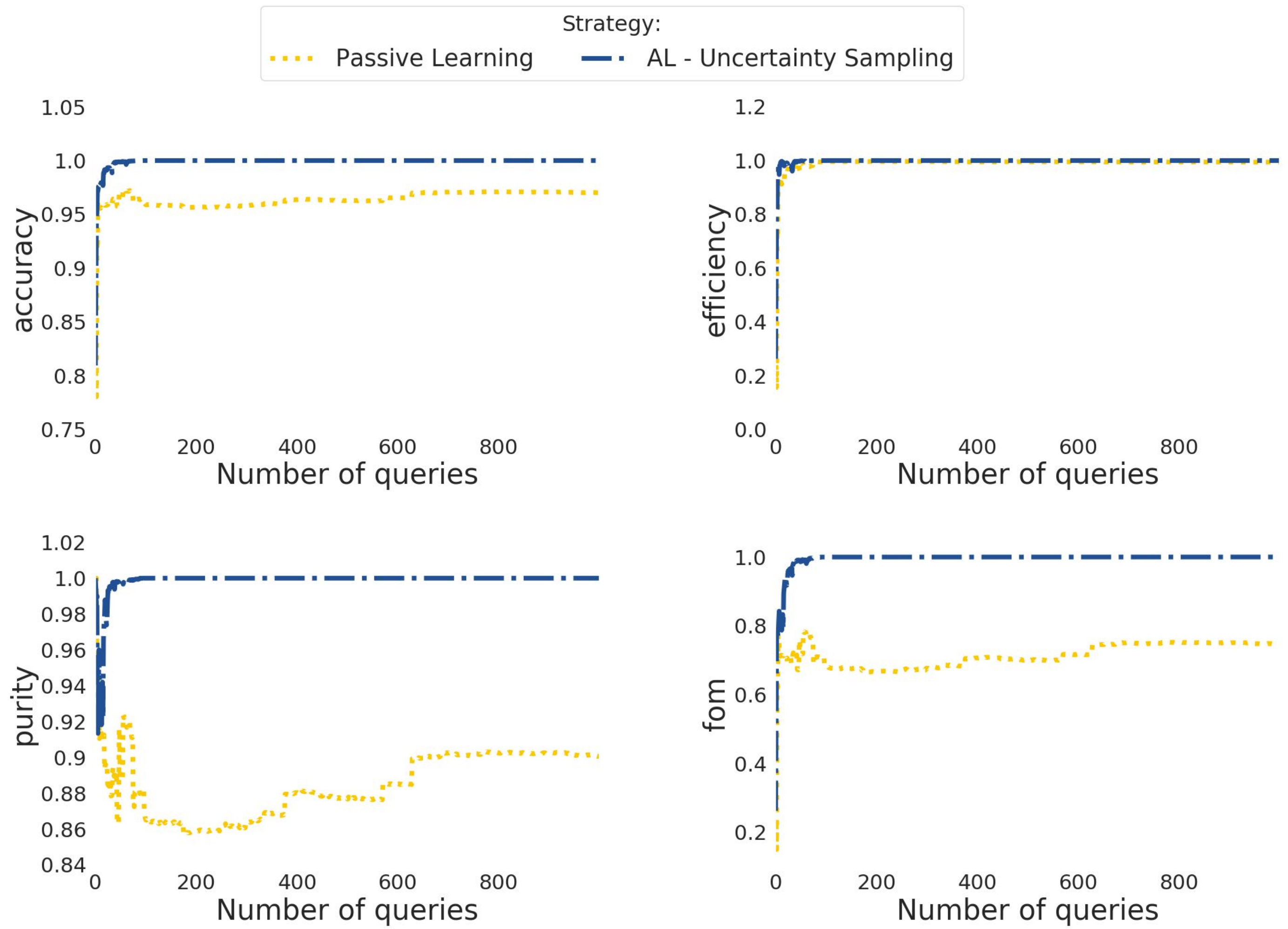
Training sample: 9 000 objects
Test sample: 74 000 objects

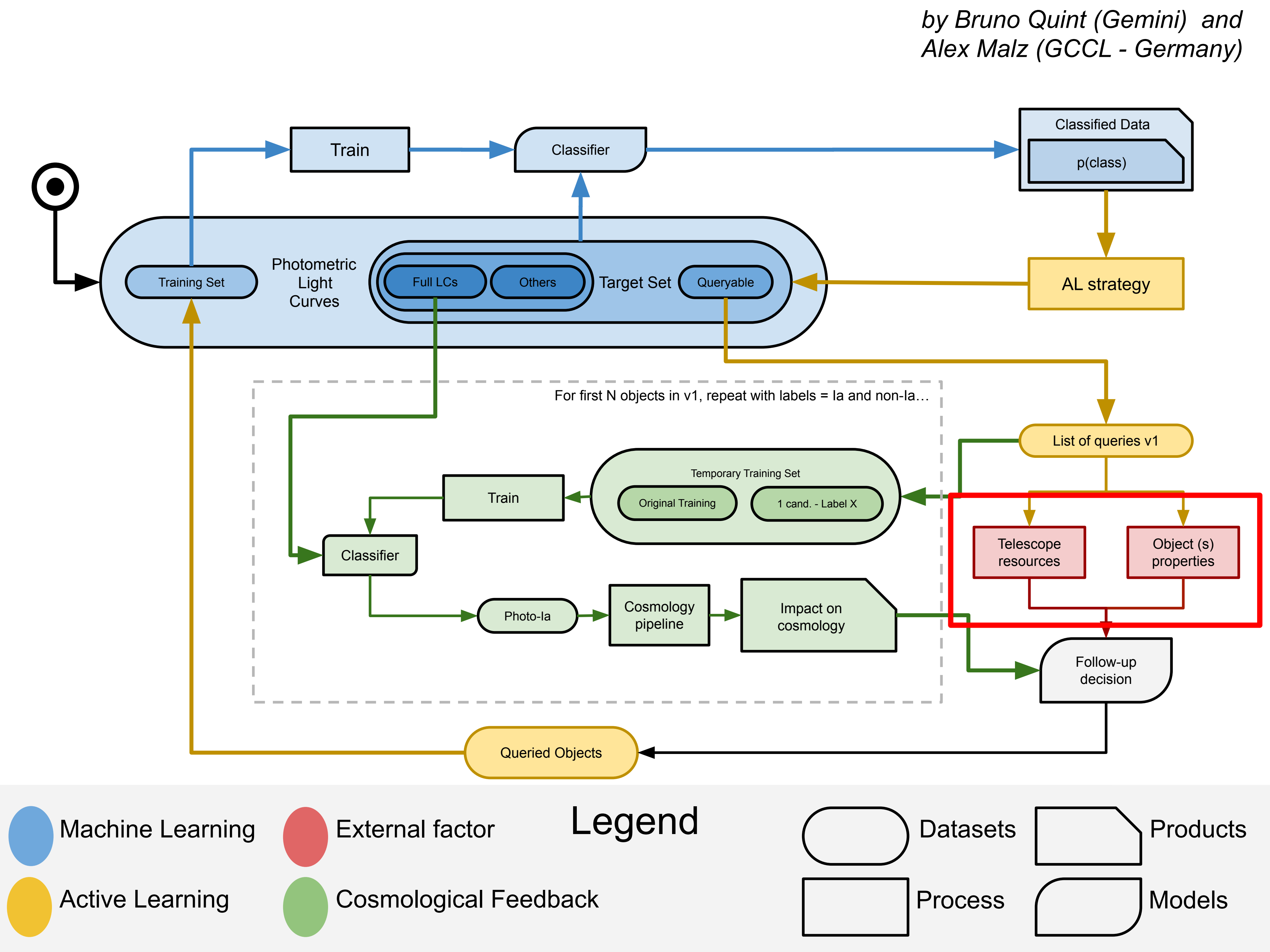


Classifiers on Stage 0 simulations:

Bazin fits
Random Forest
Uncertainty sampling

Training sample: 10 objects
Test sample: 81 000 objects




by Bruno Quint (Gemini) and
Alex Malz (GCCL - Germany)
Train
Classifier
Classified Data
p(class)
Training Set
Photometric Light Curves
Full LCs
Others
Target Set
Queryable
AL strategy
For first N objects in v1, repeat with labels = Ia and non-Ia…
List of queries v1
Temporary Training Set
Original Training
1 cand. - Label X
Train
Classifier
Telescope resources
Object (s) properties
Photo-Ia
Cosmology pipeline
Impact on cosmology
Follow-up decision
Queried Objects
Legend
Machine Learning
External factor
Active Learning
Cosmological Feedback
Datasets
Products
Process
Models

Spectroscopic requirements

- **Position in the sky** (object [RA, DEC] in comparison with field of view of chosen telescopes, moon);
- **Noisy oracle**s (probability that spectral classification can be wrong – maybe based on SNR);
- **Missed opportunity** (situations where spectra was requested but observation was not good enough to result in a reliable classification);
- **Time since maximum brightness** (this will require a classification probability and consequent estimation of the time of maximum brightness);
- ….

Spectroscopic requirements

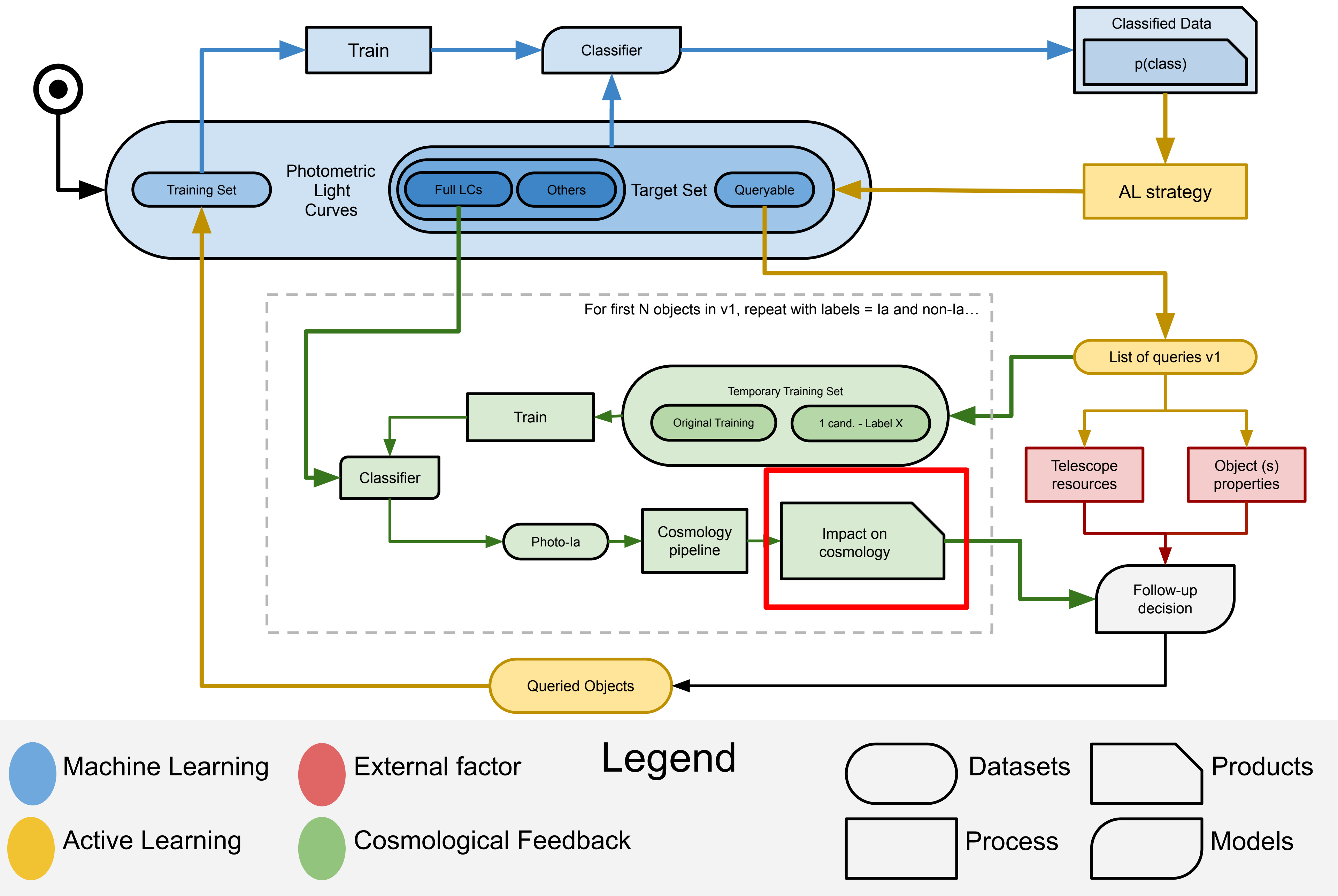
by Santiago Gonzalez Gaitan (U. Lisbon)

Table 2: Limiting epochs when spectroscopic observation is expected to be able to produce a classification. This is just an extract. Complete table contains information for other redshifts and can be found at `https://bit.ly/36SbWLD`. Table by Santiago Gonzalez Gaitan.

	z = 0.01				...
	Follow-up - 8m*		Follow-up - 4m^{x}		...
Type	Pre-max (days)	Post-max (days)	Pre-max (days)	Post-max (days)	...
SNIa	-18	80	-18	80	...
SNIa-91bg	-11	70	-11	nsi	...
SNIax	-14	80	-14	80	...
SNIa-supCh	-20	120	-20	120	...
SN-IIP	-7	150	-7	150	...
SN-IIL	-7	120	-7	120	...
SN-IIn	-30	100	-30	100	...
SN-IIb	-12	80	-12	80	...
SNIbc	-20	80	-20	80	...
SLSN-I	-35	220	-35	220	...
SLSN-II	-60	450	-60	450	...
...	...	...	...	...	...

* Assumes limiting i-band mag of 24, $t_{\mathrm{exp}} = 7200s$ for S/N $\sim$ 10 per 10 Å.
^{x} Assumes limiting i-band mag of 21.5, $t_{\mathrm{exp}} = 2000s$ for S/N $\sim$ 10 per 20 Å.

by Bruno Quint (Gemini) and
Alex Malz (GCCL - Germany)



by Kara Ponder (UC Berkeley)

Cosmology based metric

We need a fast way to compare the the impact on cosmological results based on different data sets.

Our first try is a fisher matrix-based comparison.

Figure 9: Example of the output from the metrics pipeline when applied to 2 idealized data sets. **Left**: Contours derived from data 2 data sets with the same number of points (500 each) and different error levels. Data set 1 (red) considered 10% of the errors from Stage 1 simulations and data set 2 (blue) considered full erros in the same simulation. **Right**: Diagnostic output using the approximation presented by Hees et al. (2019).

by Rafael S. de Souza (UNC and SHAO)

Cosmology pipeline

The fast metric test needs to be confirmed by a full cosmological fit.

A cosmology fit code was developed to serve as a baseline test for the metric

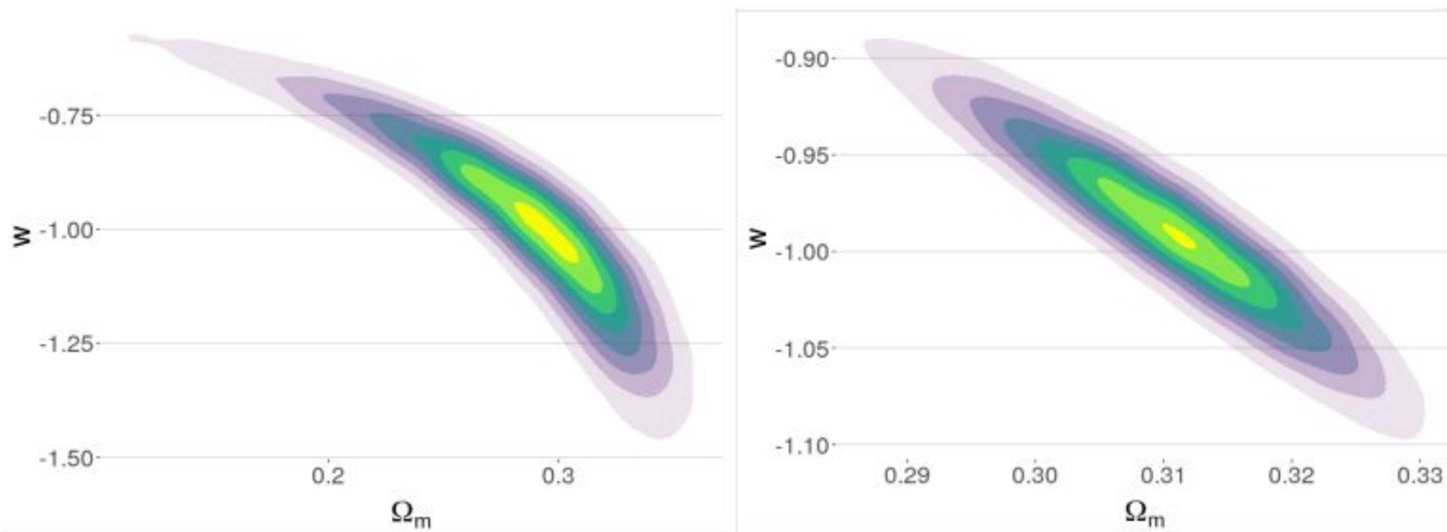


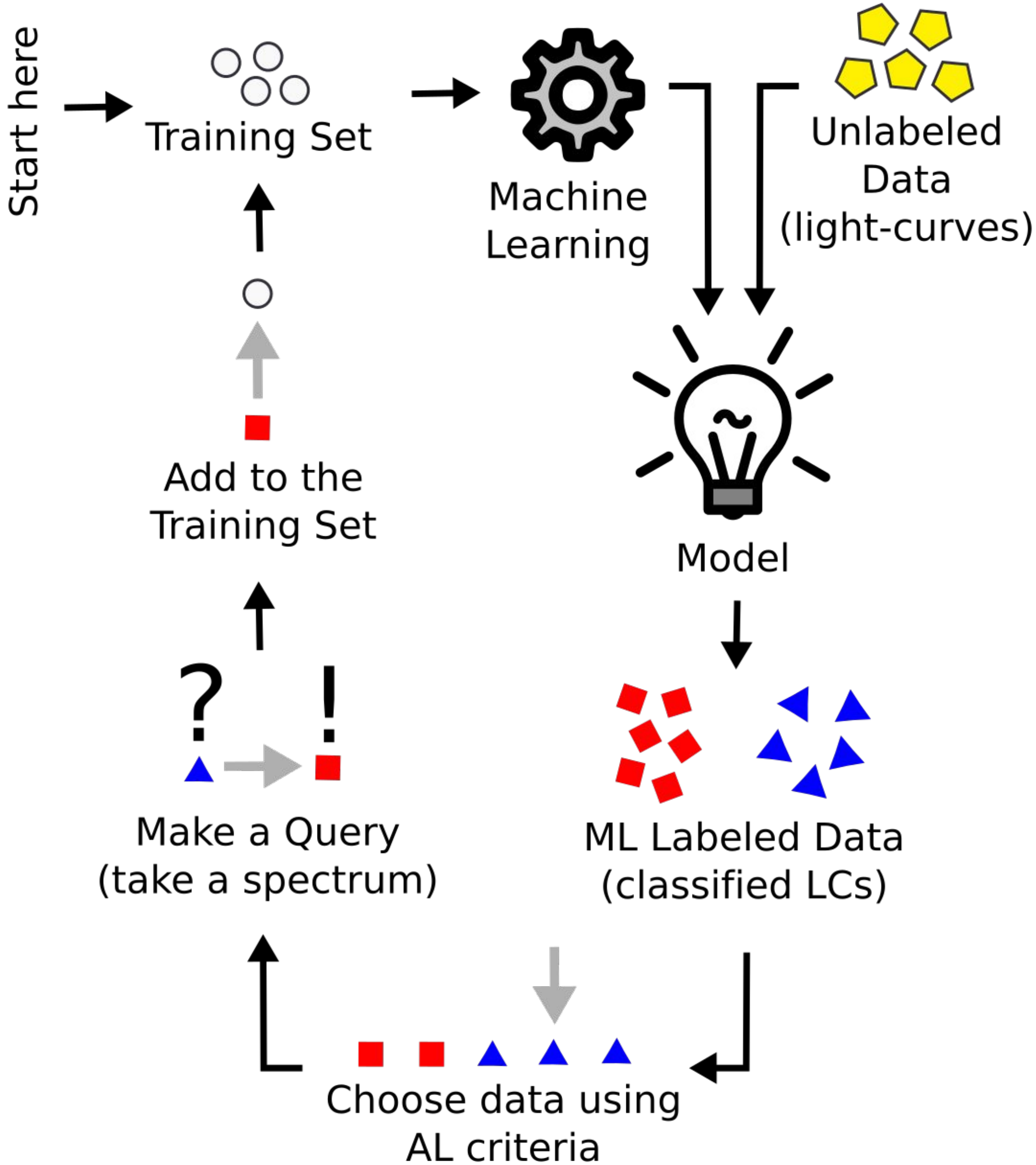
Figure 11: Joint posterior for w and Ω_m for 500 (left panel), and 10,000 (right panel) simulated SNe. We use fitted SALT2 parameter values and considered measurement errors only in m_B.

Second live meeting planned for March, 2020

Deep thanks to all the RESSPECT team!!!!

https://cosmostatistics-initiative.org/focus/resspect2/

Extra slides

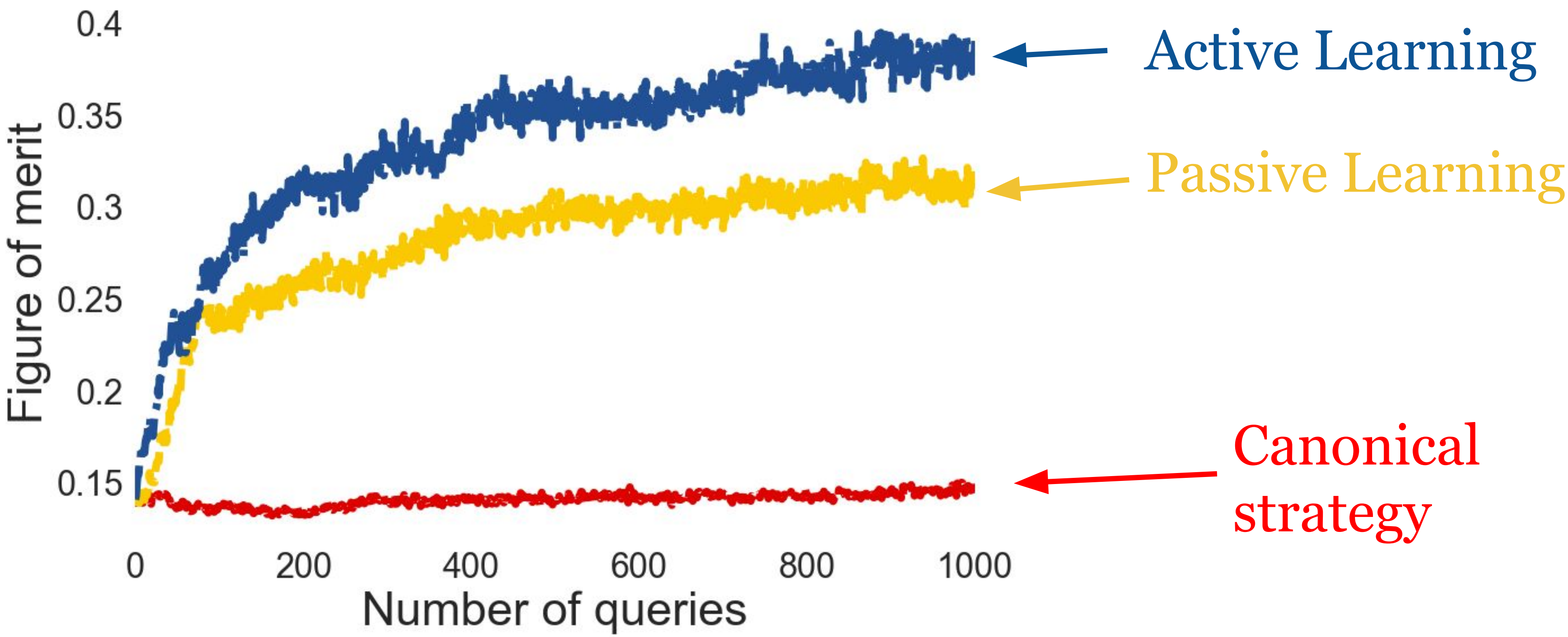


COIN
Cosmostatistics Initiative

Ishida et al., 2019, MNRAS
from COIN Residence Program #4

AL for SN classification

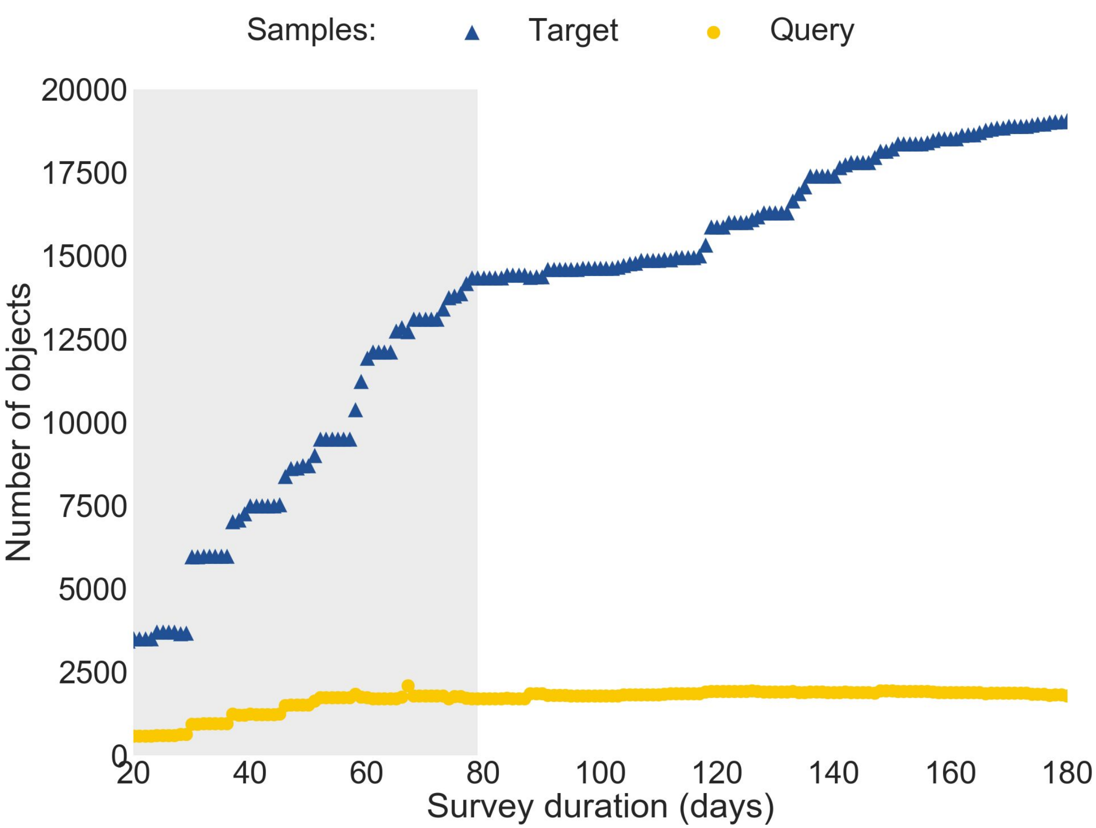
Static results



Ishida** et al., 2018 - arXiv:astro-ph/1804.03765 - from **CRP #4

SN are transients

Window of opportunity

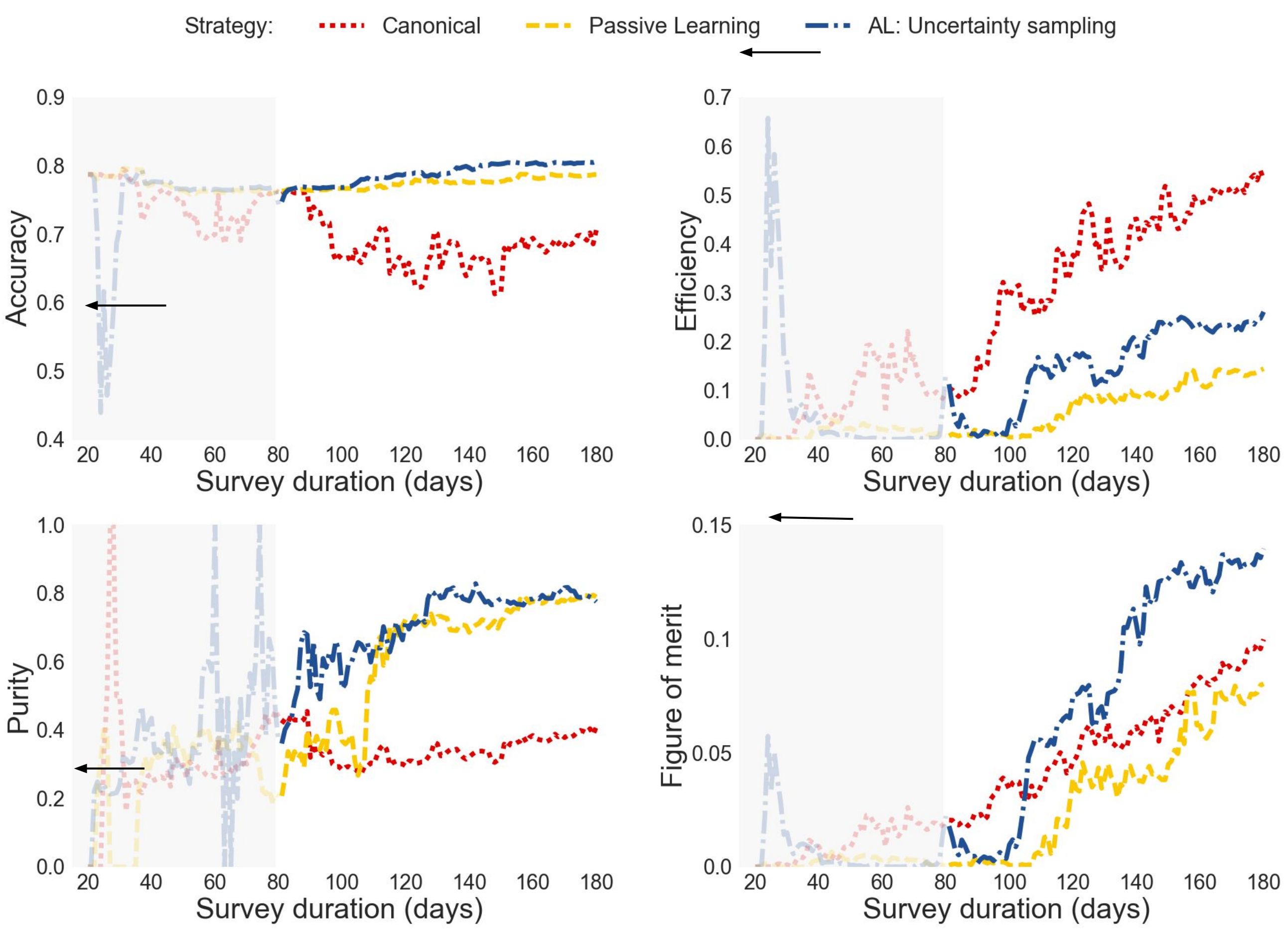


1. Feature extraction done daily **with available observed epochs until then.**

2. Query sample is also re-defined daily: objects with **r-mag < 24**

From COIN Residence Program #4, ***Ishida*** *et al., 2019, MNRAS, 483 (1), 2–18*

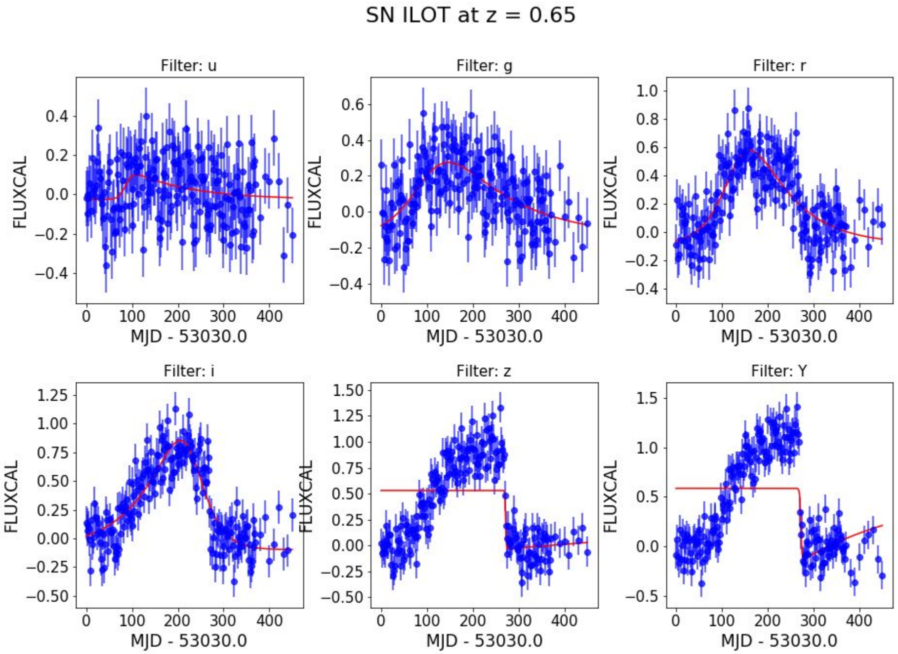
No initial training



The arrow shows <u>traditional</u> Full light-curve results with full SNPCC spec

From COIN Residence Program #4, ***Ishida*** *et al., 2019, MNRAS, 483 (1), 2–18*

Simulations:



Stage 0

Stage 1