Are classification metrics good proxies for science output?

Colloque national Dark Energy - 5ème édition

Paris - 13 October 2021

Emille E. O. Ishida

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

and Alex I. Malz (Ruhr-University, Germany), for the RESSPECT team

1. Context:

What is the goal of classification tasks?

- Organize knowledge
- Understand patterns of behavior
- Predict possible outcomes
- Separate objects for further scientific analysis

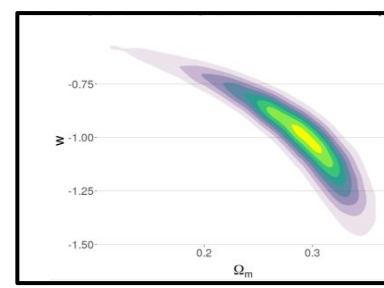
Discovery and Classification in Astronomy - by Steven Dick - Cambridge University Press (2013)

1. Context:

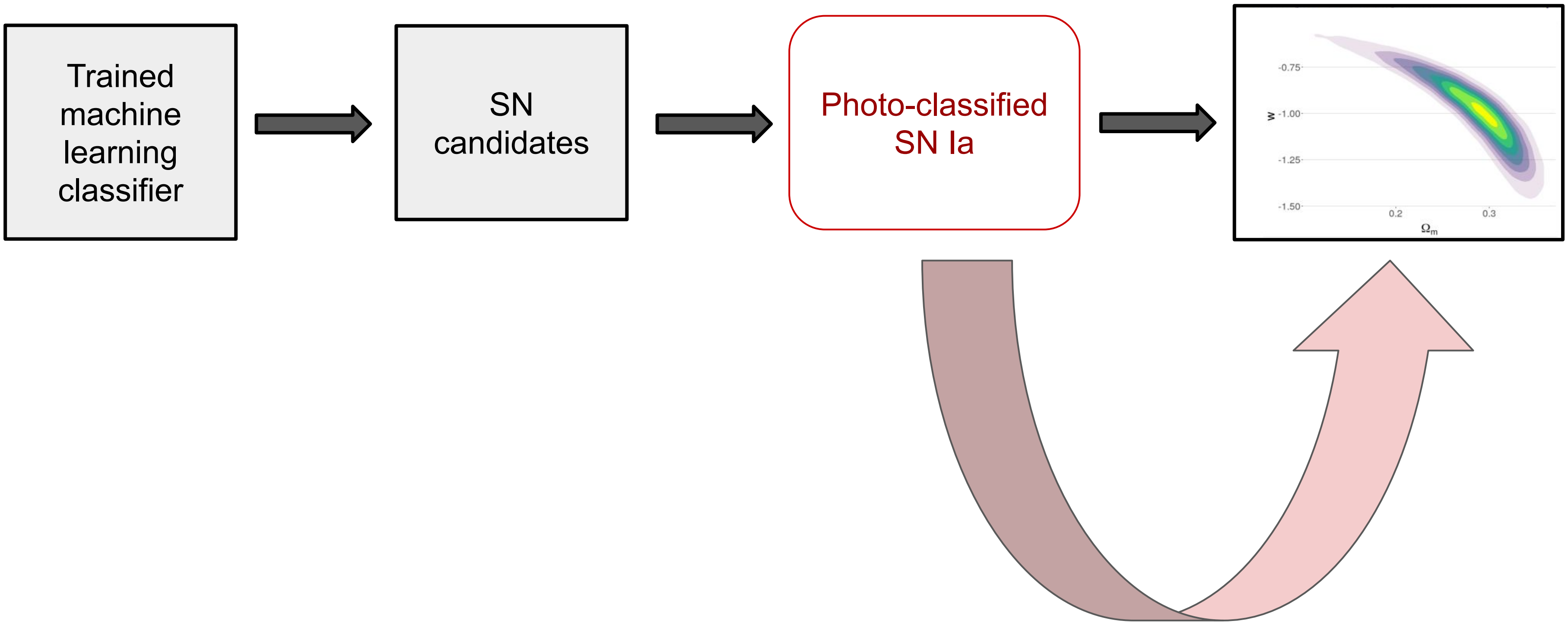
What is the goal of classification tasks?

- Organize knowledge
- Understand patterns of behavior
- Predict possible outcomes
- Separate objects for further scientific analysis

Trained machine learning classifier → SN candidates → Photo-classified SN Ia →



2. Evaluating classification results



Hypothesis:

Good classification results will result in better science results

2. Experiment set-up

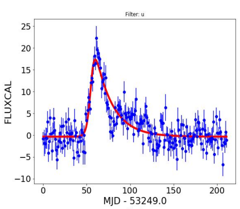
Goal: to understand the impact of different contaminants in a cosmology results

PS:1 There is no real classifier

4.a - wfit

w-best-fit

1 - LCs



2 - SALT2 fit

x1, c, mB

3 - SALT2mu

mu, muerr

4.b - Stan

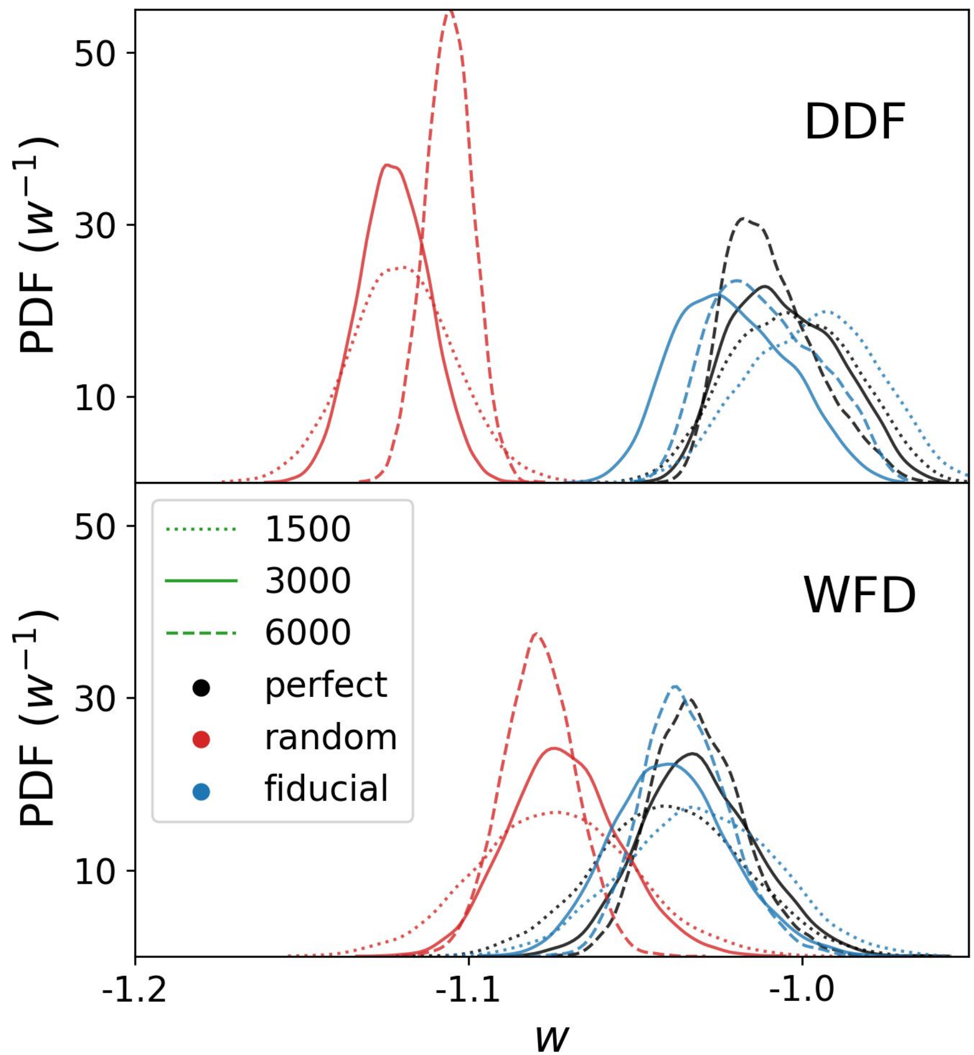
w-posterior

PS2: Very simplified cosmological analysis

- PLAsTiCC data
- Study each different contaminant at a time
- Separate analysis for DDF and WFD
- Compare traditional classification metrics with properties from posteriors

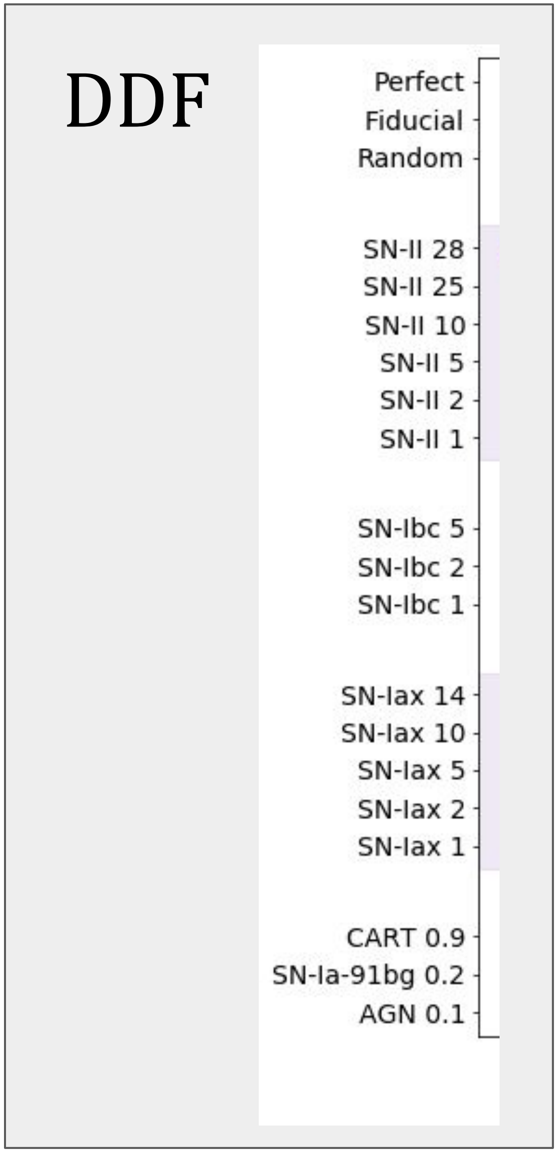
3. Total number of objects

- **Perfect:** all real SN Ia
- **Random**: sampling from the set of objects surviving SALT2 fit (14% cont - DDF, 10% cont WFD)
- **Fiducial:** objects classified as SNIa by Avocado (Boone, 2019) which survived SALT2 fit -- (Ibc, II, Iax ~ 5%, 3% cont - WFD)



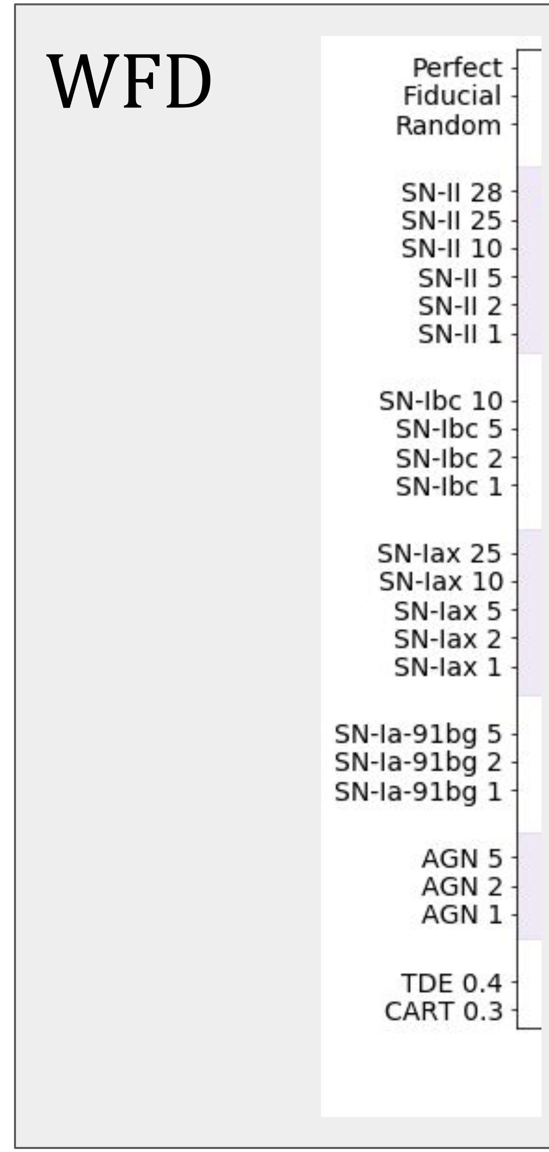
Malz and the RESSPECT team, 2021, under DESC and COIN internal review

4. Samples surviving SALT2 fit

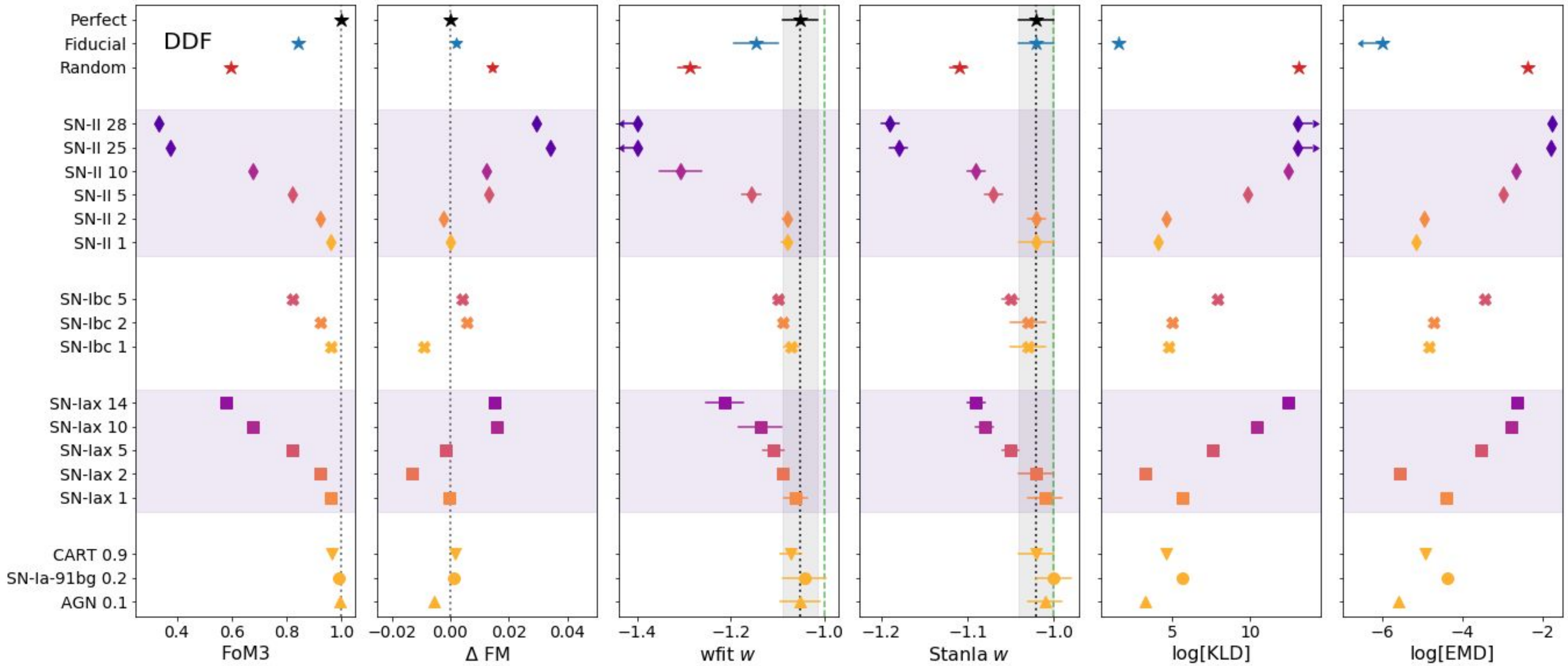


PS:1 There is no real classifier

(apart from Avocado)

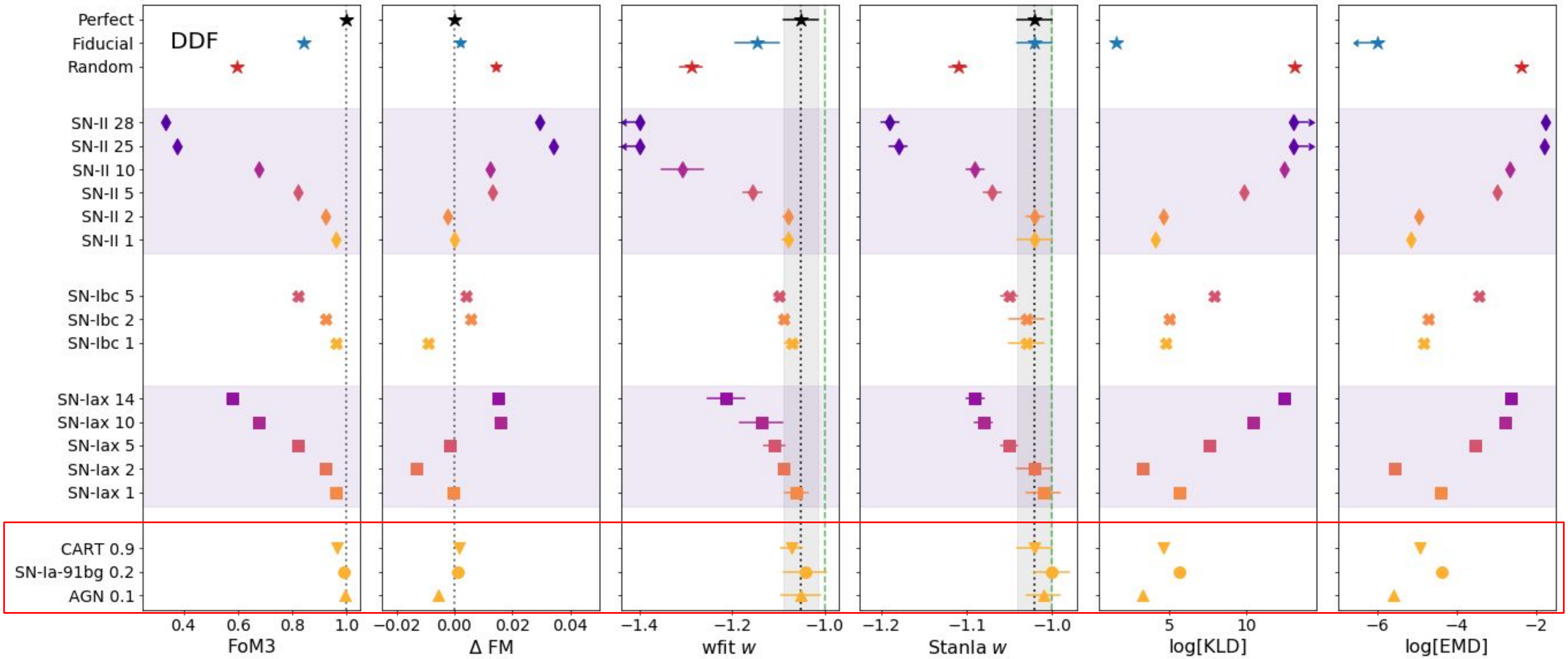


5. Results: DDF



Malz and the RESSPECT team, 2021, under DESC and COIN internal review

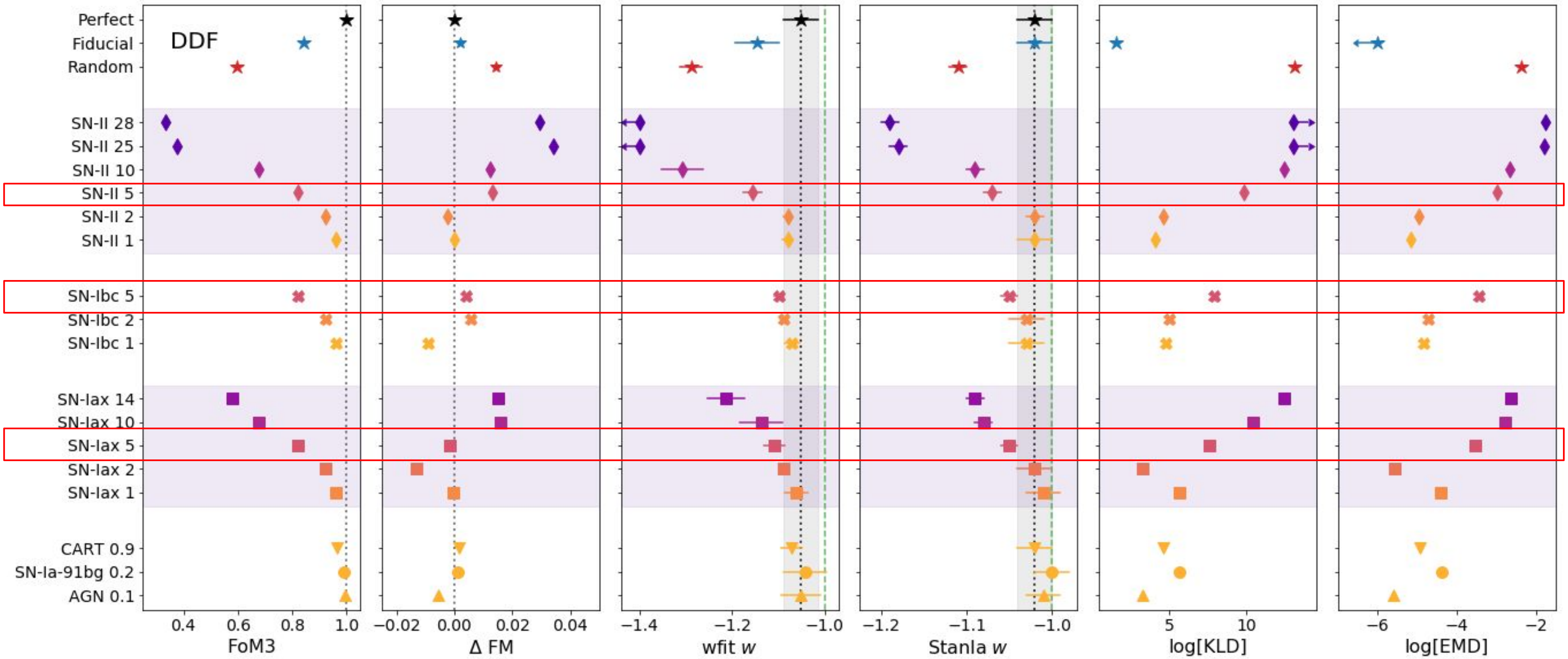
5. Results: DDF



1. Metrics based on posteriors are sensitive to small contaminations

Malz and the RESSPECT team, 2021, under DESC and COIN internal review

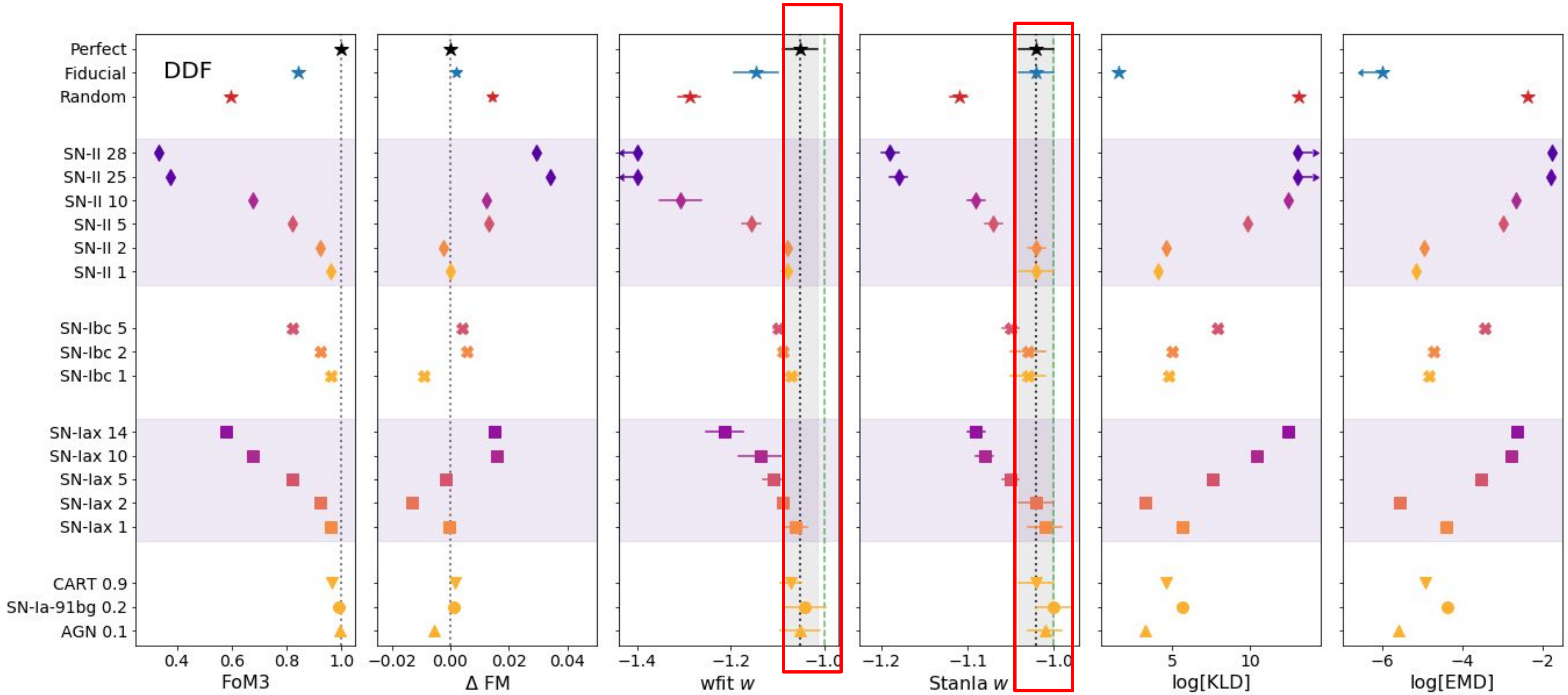
5. Results: DDF



1. Metrics based on posteriors are sensitive to small contaminations
2. 5% contamination is already too much

Malz and the RESSPECT team, 2021, under DESC and COIN internal review

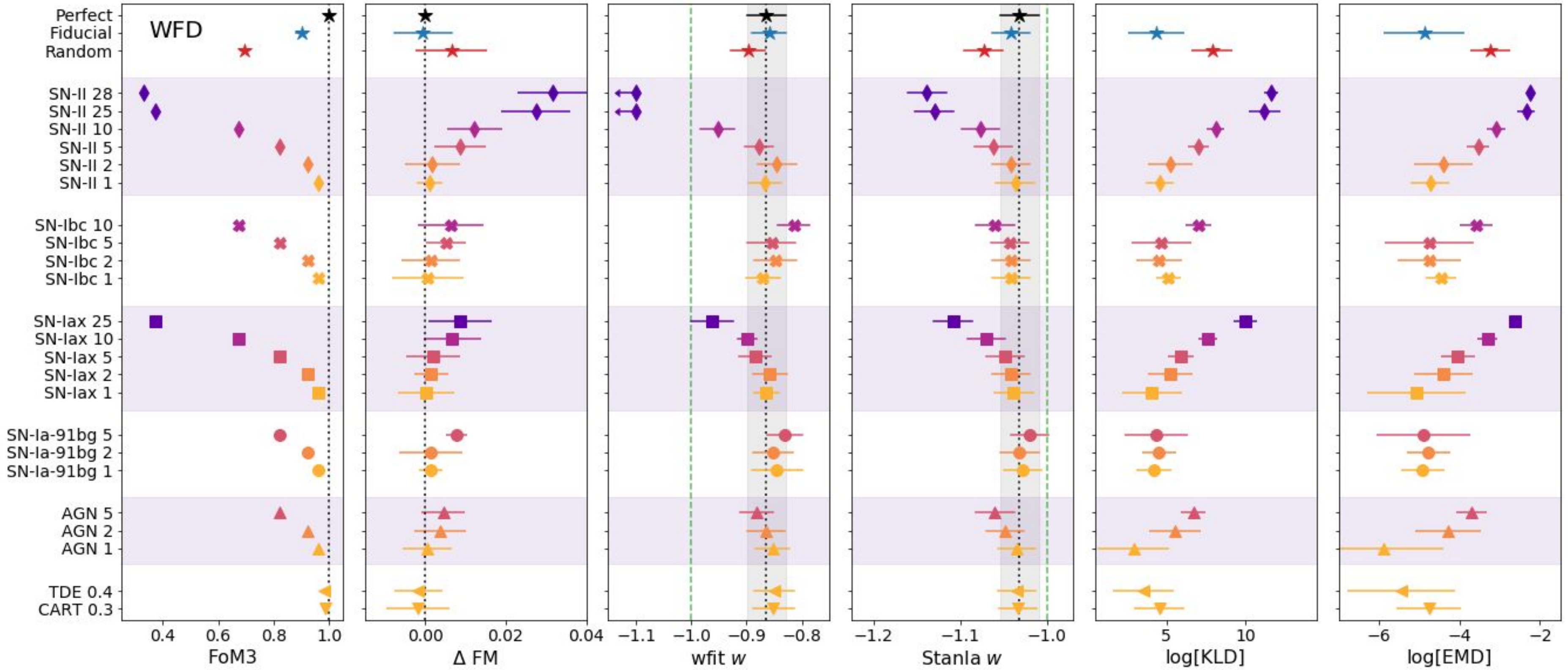
5. Results: DDF



1. Metrics based on posteriors are sensitive to small contaminations
2. 5% contamination is already too much
3. Perfect borders the simulated model

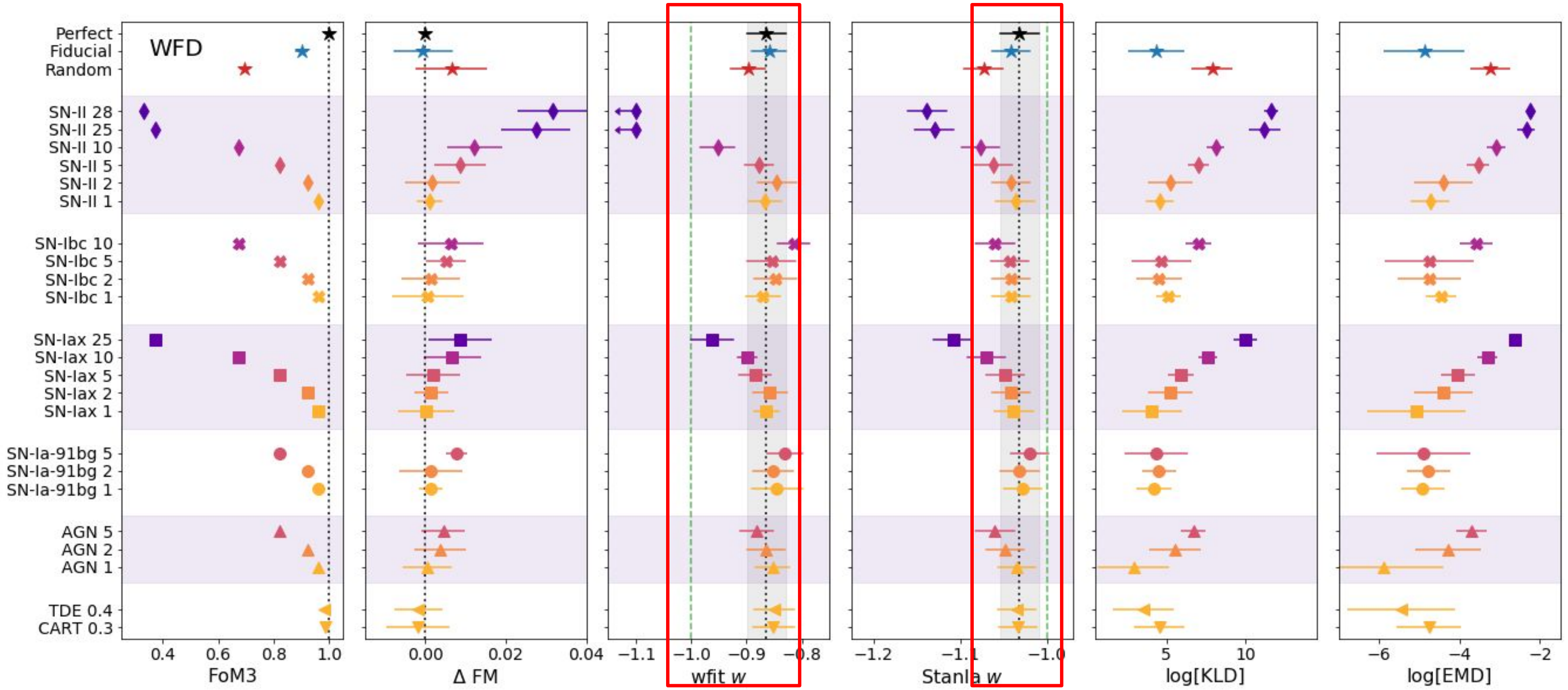
Malz and the RESSPECT team, 2021, under DESC and COIN internal review

6. Results: WFD



Malz and the RESSPECT team, 2021, under DESC and COIN internal review

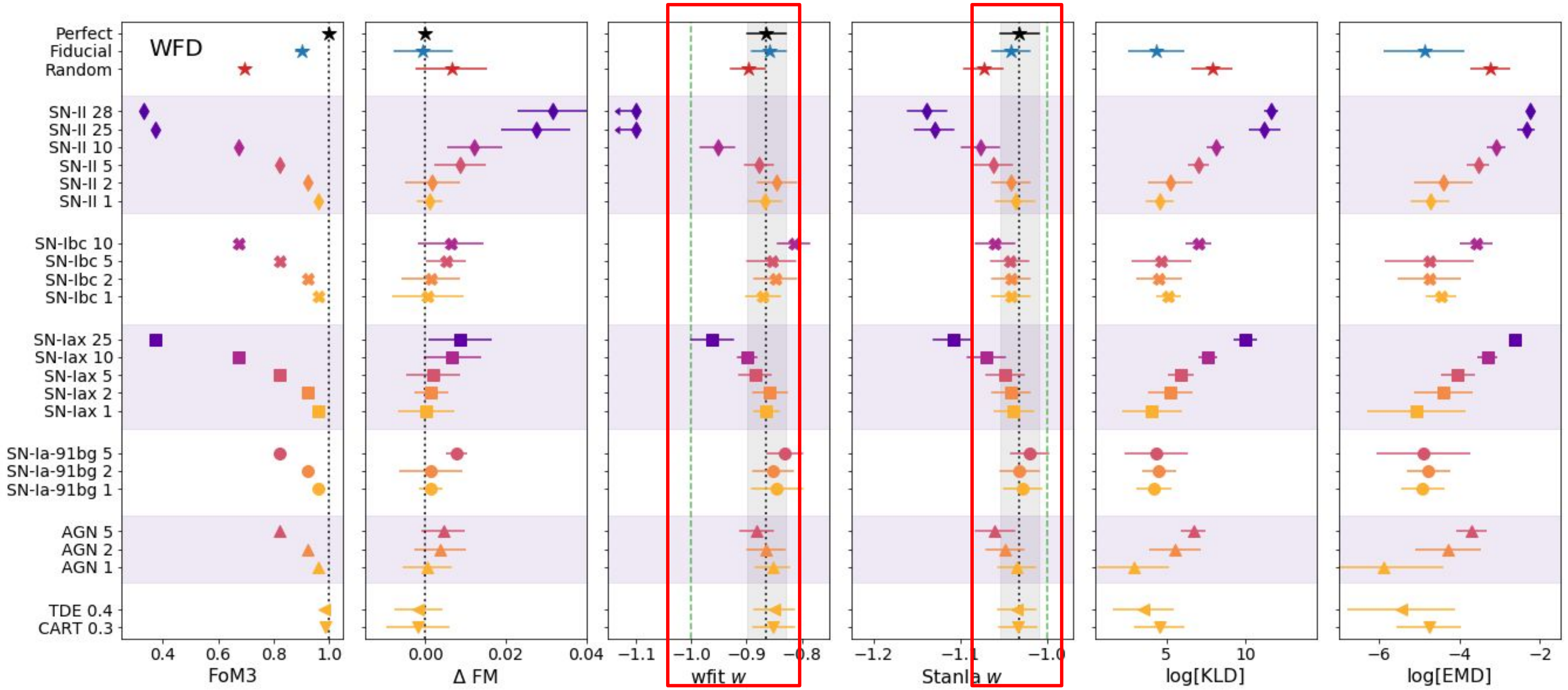
6. Results: WFD



1. Perfect results are further from simulated model + larger error bars

Malz and the RESSPECT team, 2021, under DESC and COIN internal review

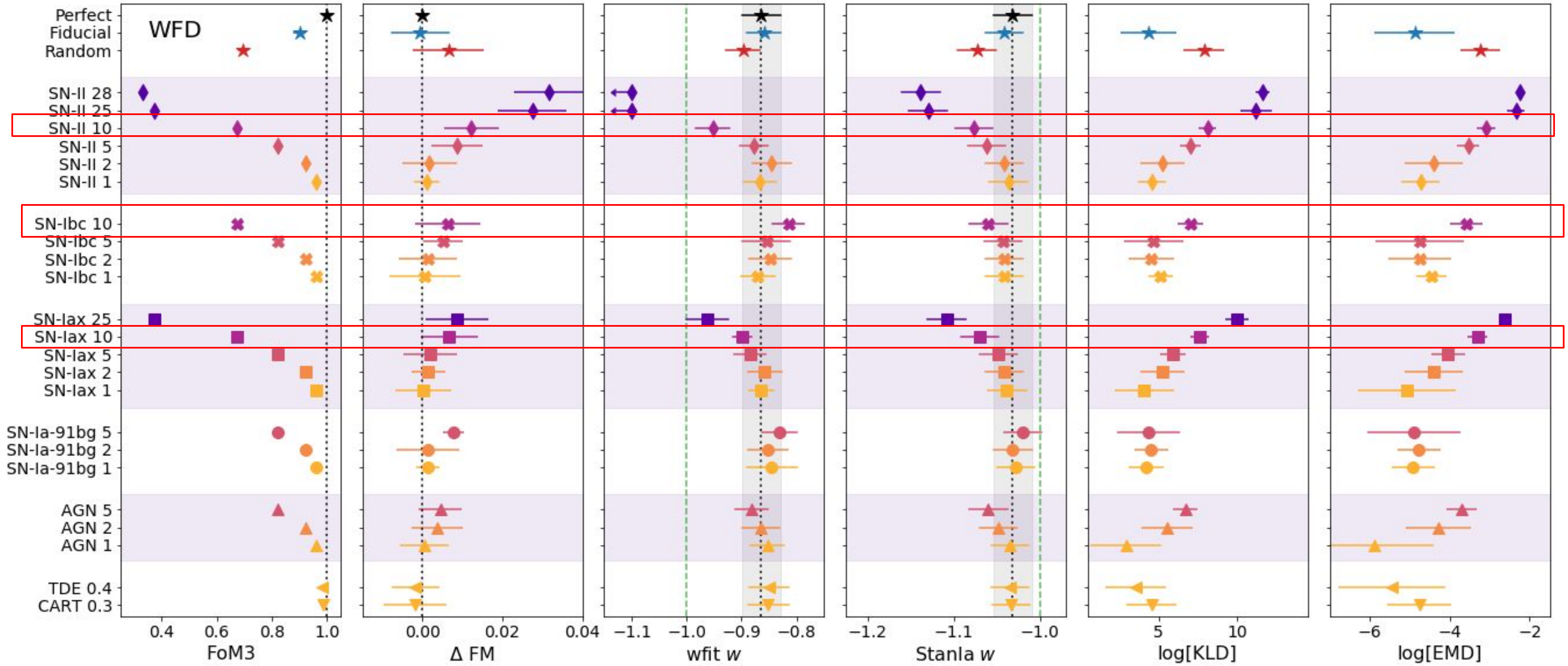
6. Results: WFD



1. Perfect results are further from simulated model + larger error bars

Malz and the RESSPECT team, 2021, under DESC and COIN internal review

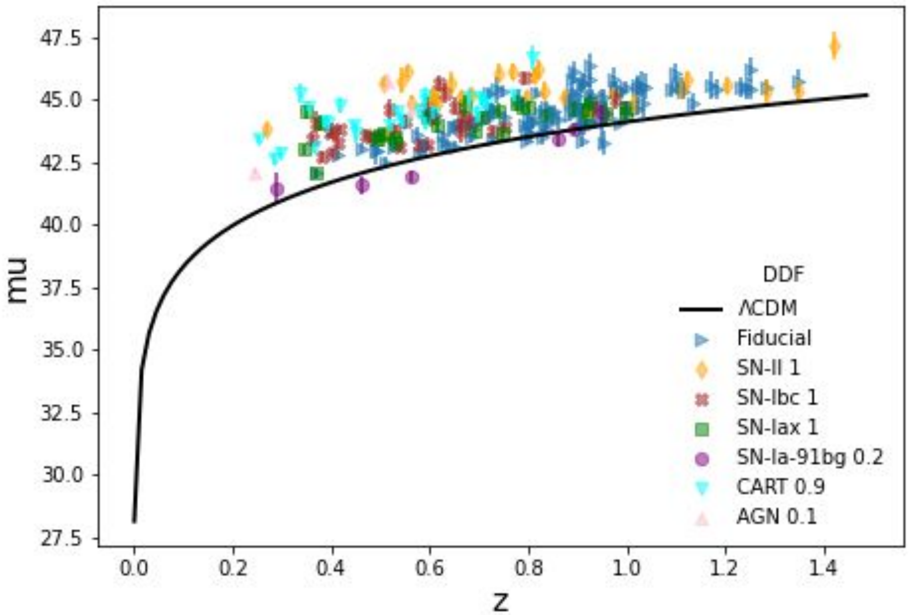
6. Results: WFD

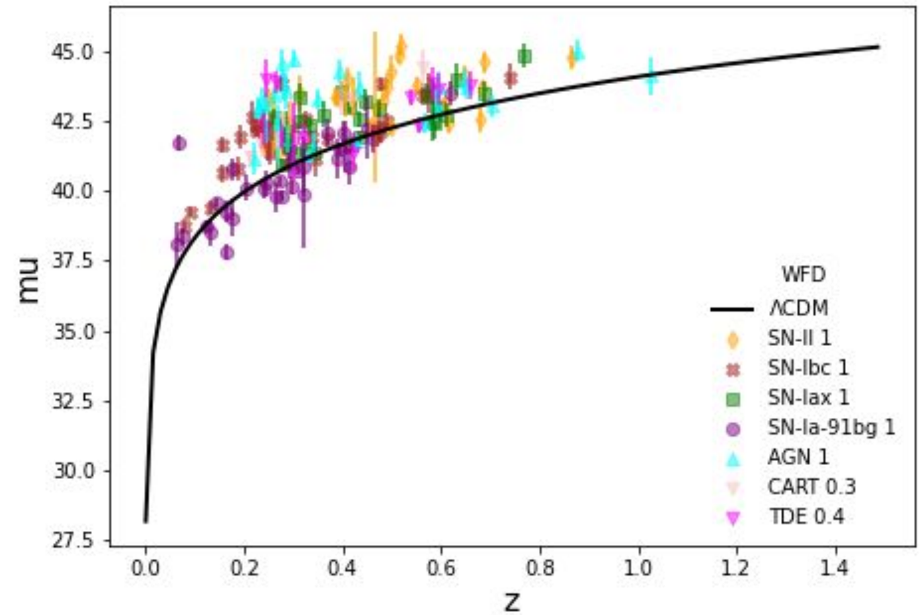


1. Perfect results are further from simulated model + larger error bars
2. 10% contamination borders the simulated model

Malz and the RESSPECT team, 2021, under DESC and COIN internal review

6. Results: Hubble diagram





Malz and the RESSPECT team, 2021, under DESC and COIN internal review

Summary

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

Classification metrics not necessarily follow impact on final scientific results

Summary

Classification metrics not necessarily follow impact on final scientific results

For SN Ia cosmology: the class of the contaminants matters

Summary

Classification metrics not necessarily follow impact on final scientific results

For SN Ia cosmology: the class of the contaminants matters

We should aim for classifiers specifically designed for the science question at hand

Summary

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

Classification metrics not necessarily follow impact on final scientific results

For SN Ia cosmology: the class of the contaminants matters

We should aim for classifiers specifically designed for the science question at hand

The full posteri analysis will be part of the cosmology metric of RESSPECT

Extra slides

Recommendation System for Spectroscopic Follow-up -- RESSPECT

Cosmology results from photometrically classified SN IA

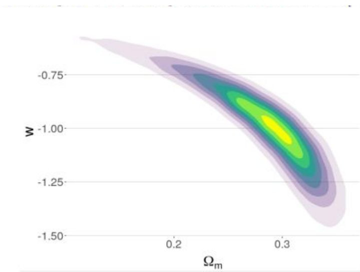


Photo-classified SN Ia

SN candidates

Trained machine learning classifier

2. To improve this!

learning algorithm

+

Training sample

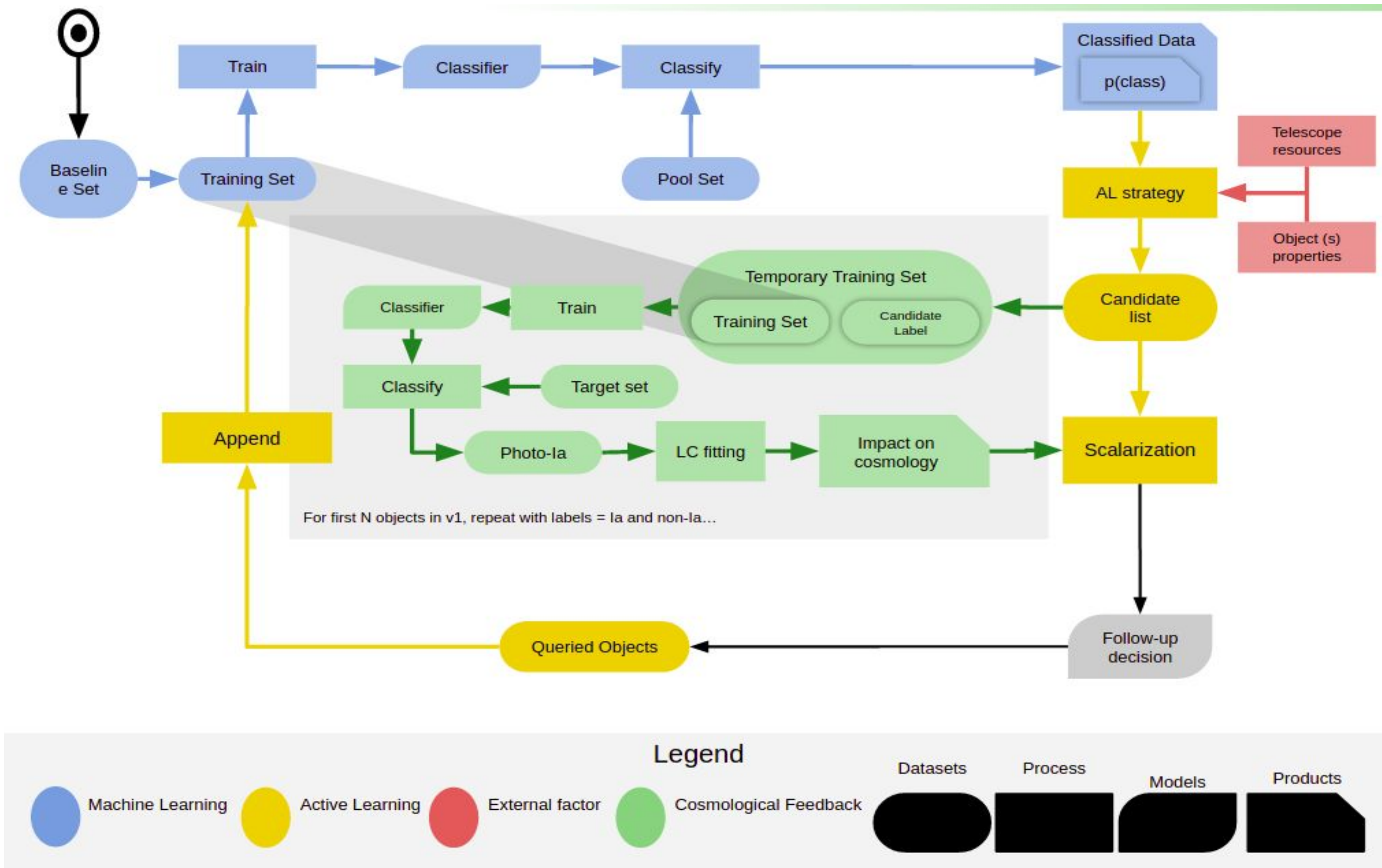
1. Optimize this!

LSST DESC & COIN

RESSPECT

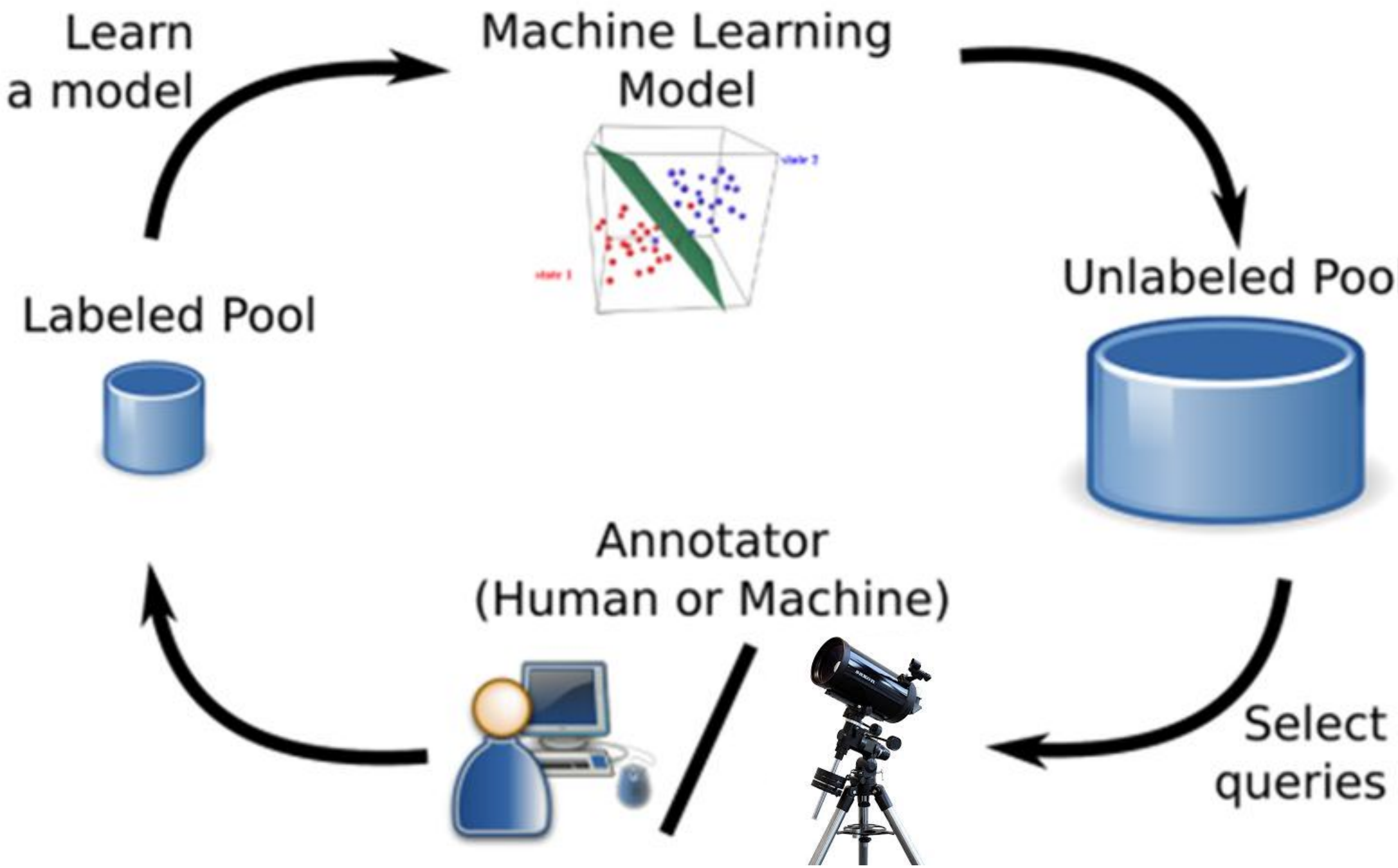
Recommendation System for Spectroscopic Follow-up

The RESSPECT pipeline

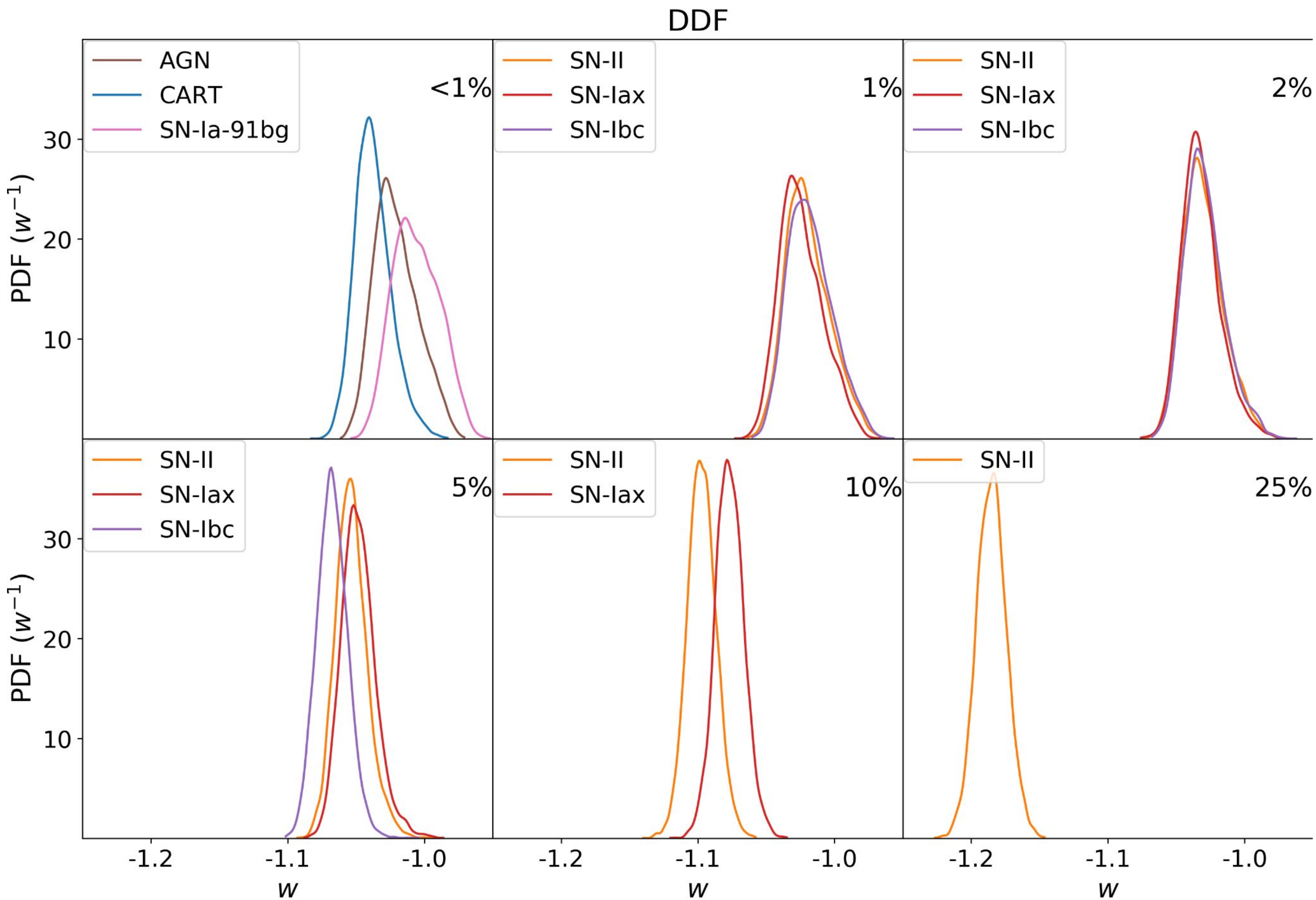


Active Learning

Optimal classification, minimum training

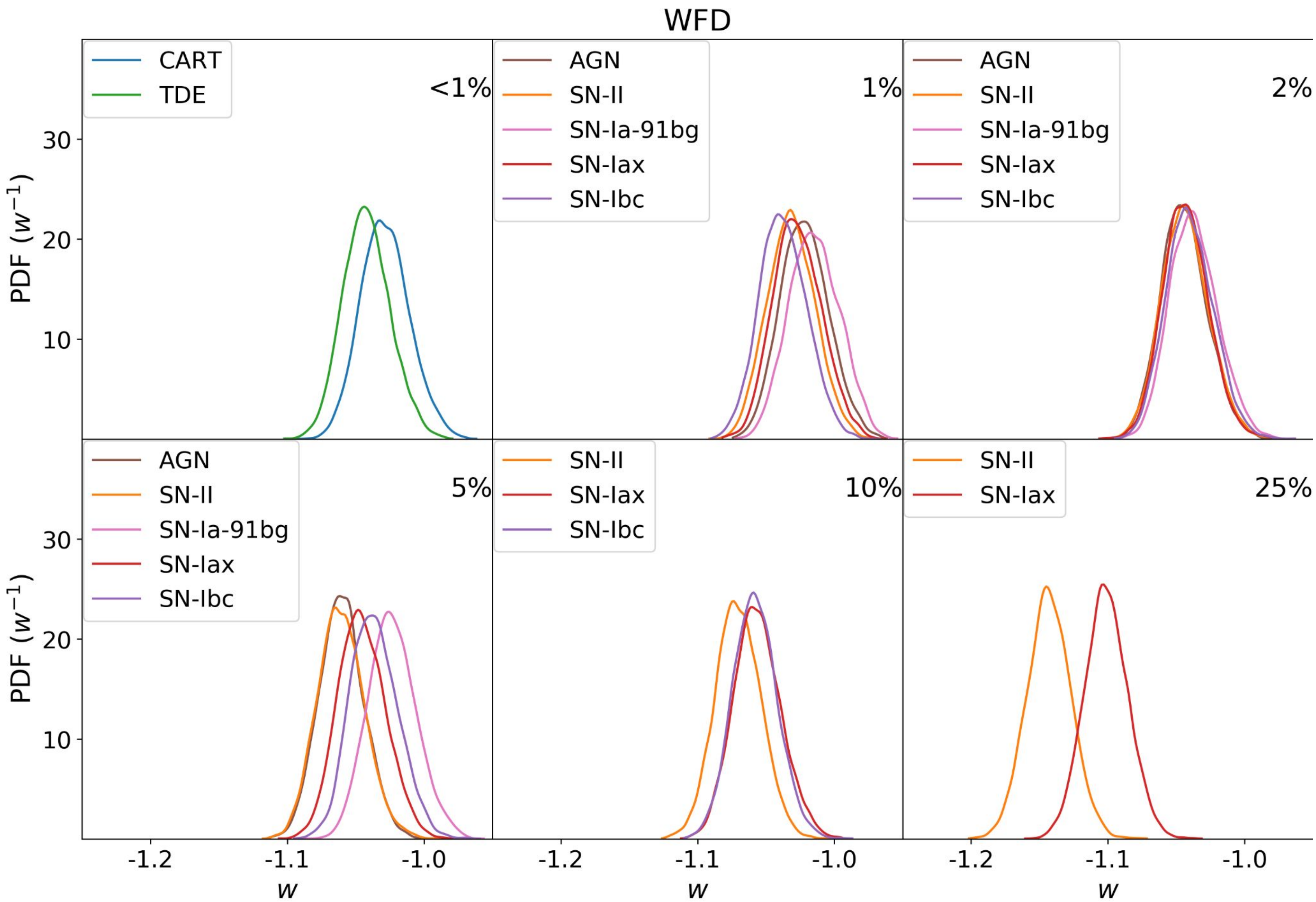


7. Posteriors: DDF



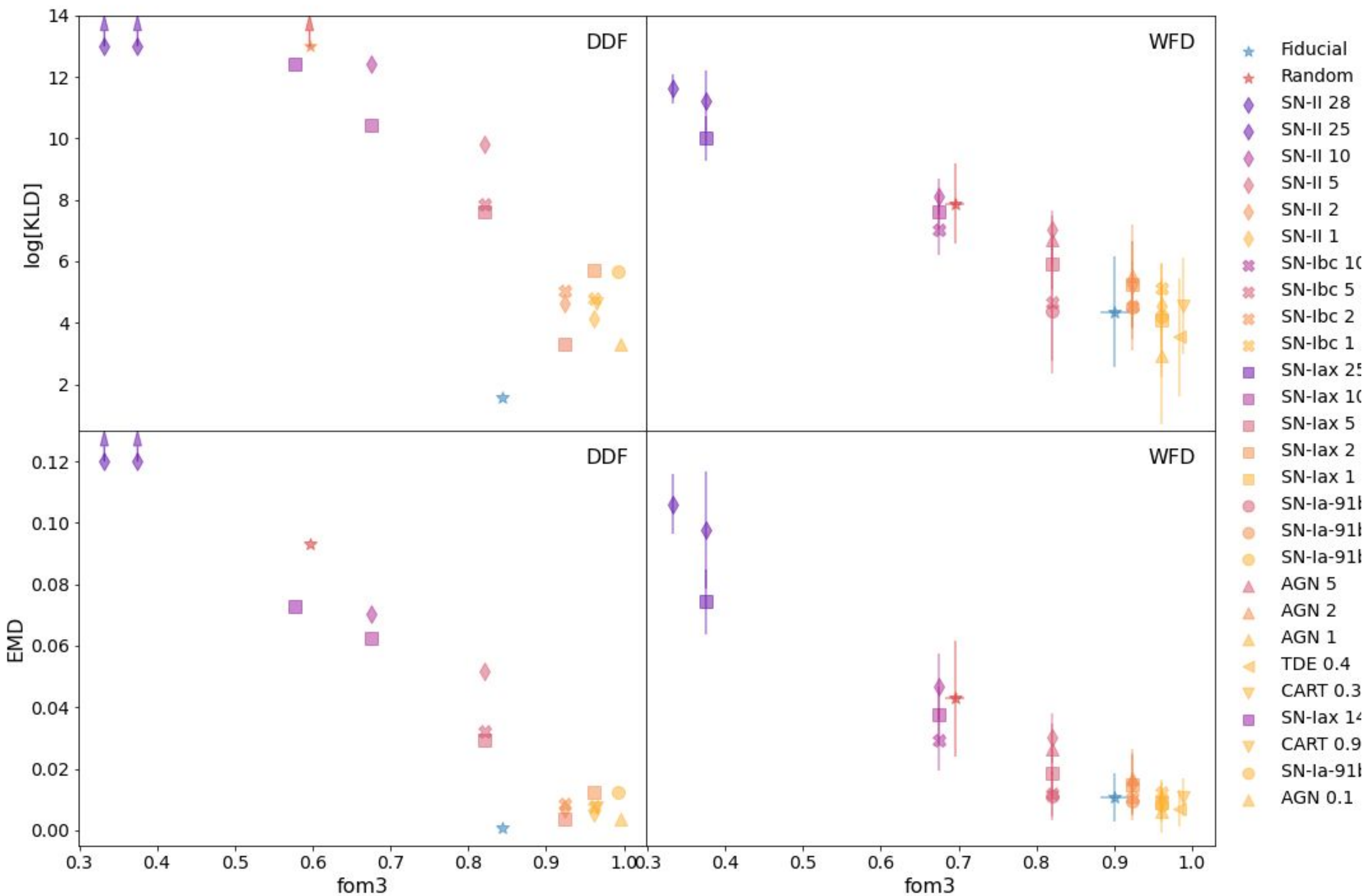
Malz and the RESSPECT team, 2021, under DESC and COIN internal review

7. Posteriors: WFD



Malz and the RESSPECT team, 2021, under DESC and COIN internal review

6. Metrics comparison



Stan model

Physical Constants:

$H_0 = 70$ km/s/Mpc

$c = 3 \times 10^5$ km/s

Model relationships:

$$E(z) = \int_0^z \frac{1}{\sqrt{\Omega_m(1+z)^3 + (1-\Omega_m)(1+z)^{3(w+1)}}}$$

$$\mu_{\text{th}}(z) = 25 + 5\log_{10}\left[\frac{c}{H_0}(1+z)E(z)\right]$$

Priors:

$\Omega_m \sim \mathcal{U}(0.299, 0.301)$

$w \sim \mathcal{N}(-1, 0.2)$

Likelihood:

$\mu \sim \mathcal{N}(\mu_{\text{th}}, \mu_{\text{err}}^2)$