hep-ph/2307.11157

The Interplay of Machine Learning-based Resonant Anomaly Detection Methods

Radha Mastandrea

In collaboration with T. Golling, G. Kasieczka, C. Krause, B. Nachman, J. A. Raine, D. Sengupta, D. Shih, and M. Sommerhalder

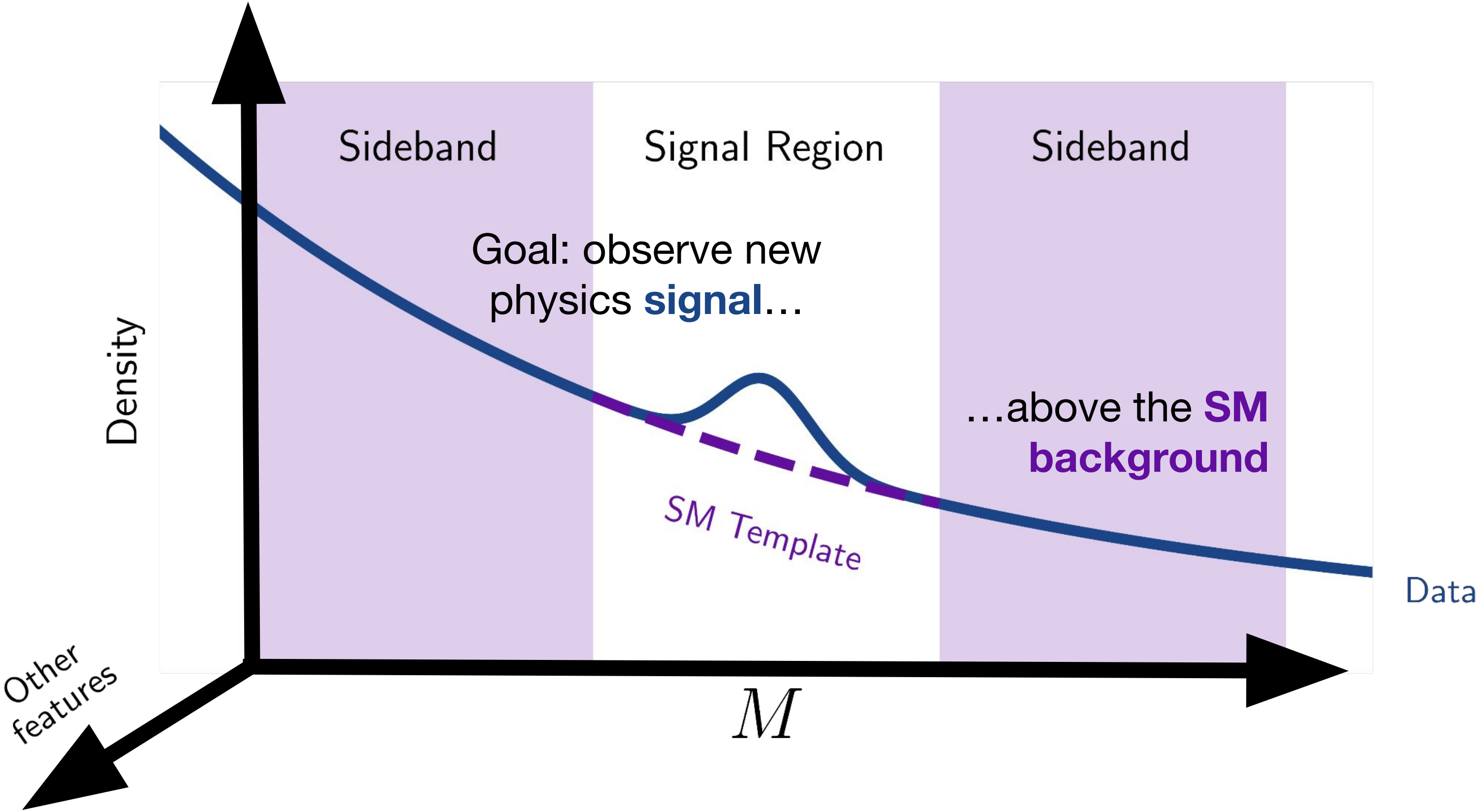
ML4Jets 2023

08/11/2023

BERKELEY LAB
Lawrence Berkeley National Laboratory

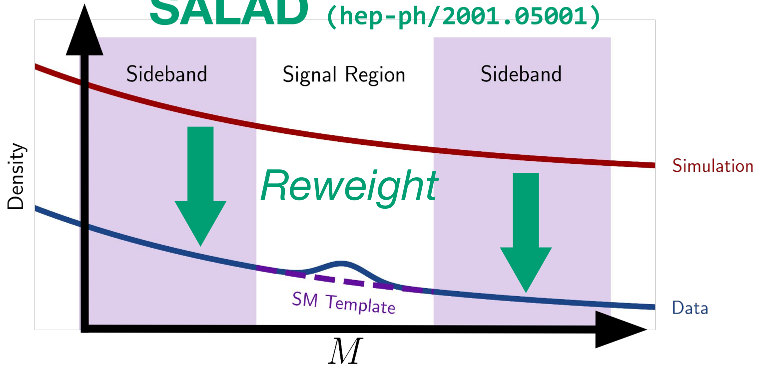
Berkeley
UNIVERSITY OF CALIFORNIA

Resonant anomaly detection as a search strategy

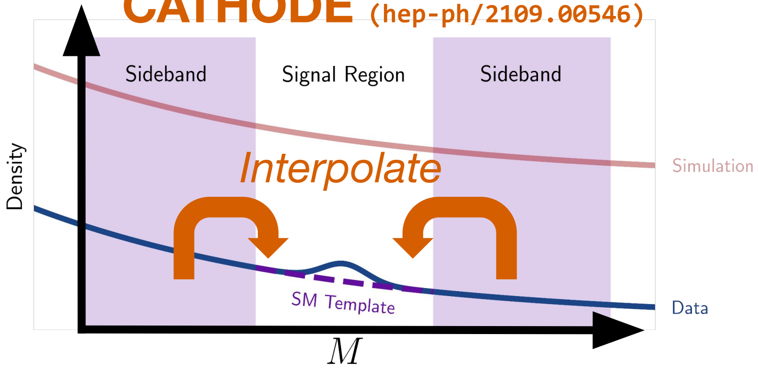


Many* ML techniques can construct the SM Template

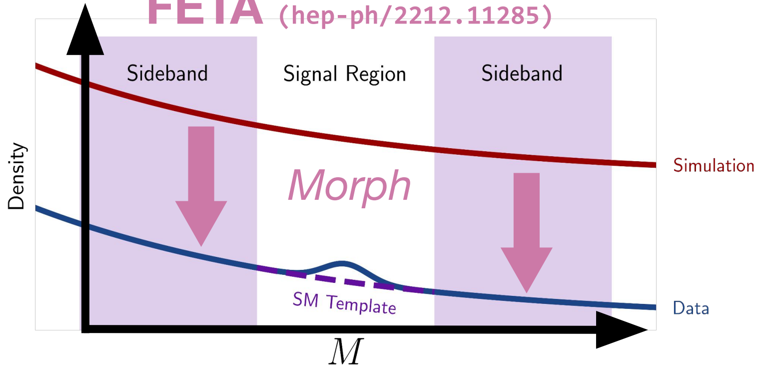
SALAD (hep-ph/2001.05001)



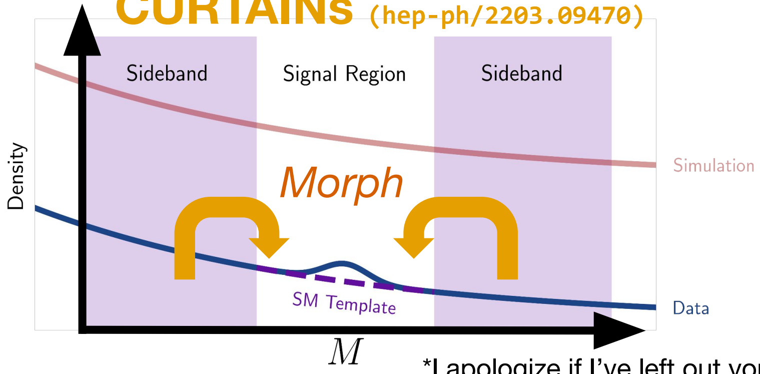
CATHODE (hep-ph/2109.00546)



FETA (hep-ph/2212.11285)

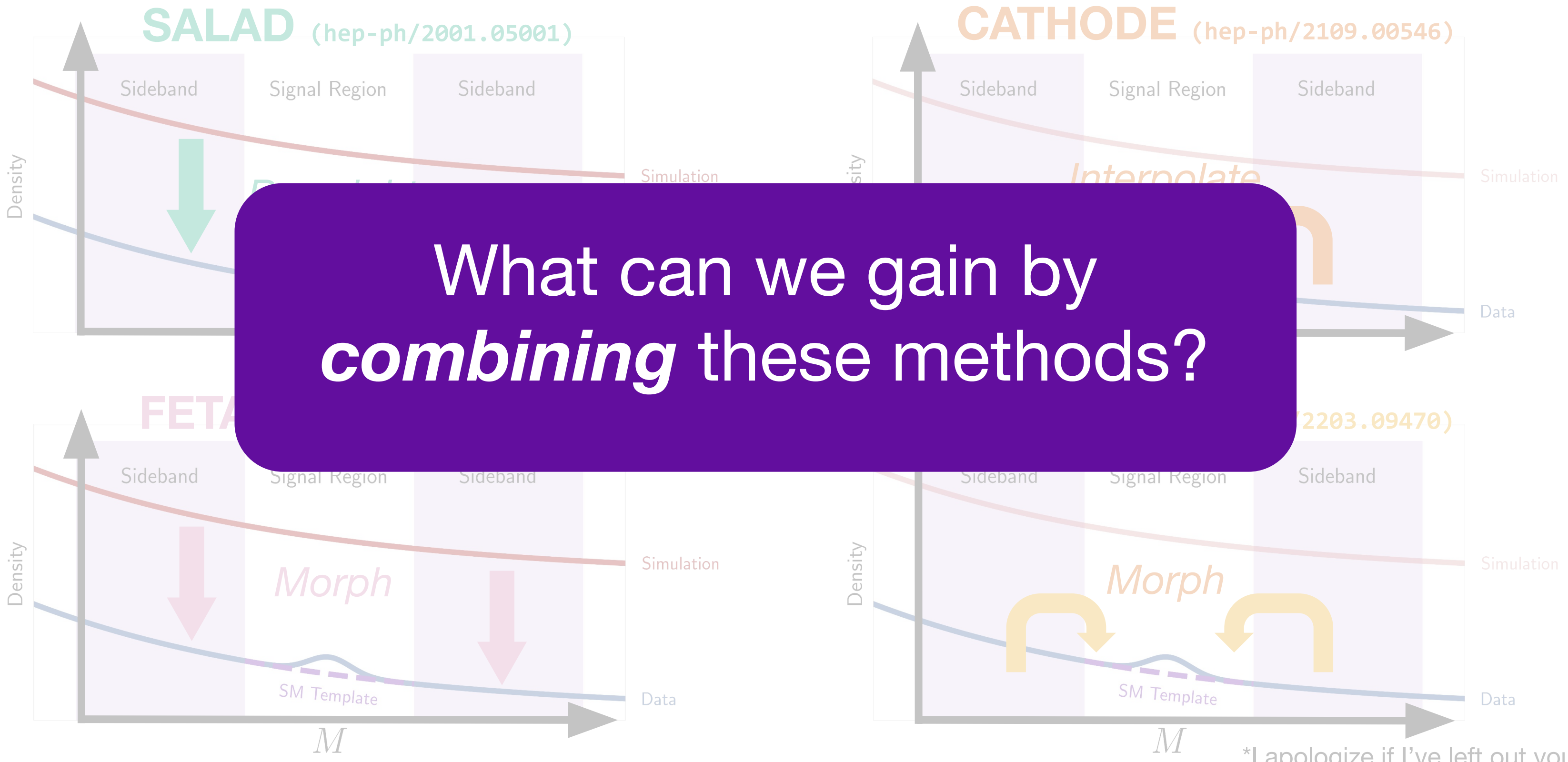


CURTAINs (hep-ph/2203.09470)



*I apologize if I've left out your favorite!

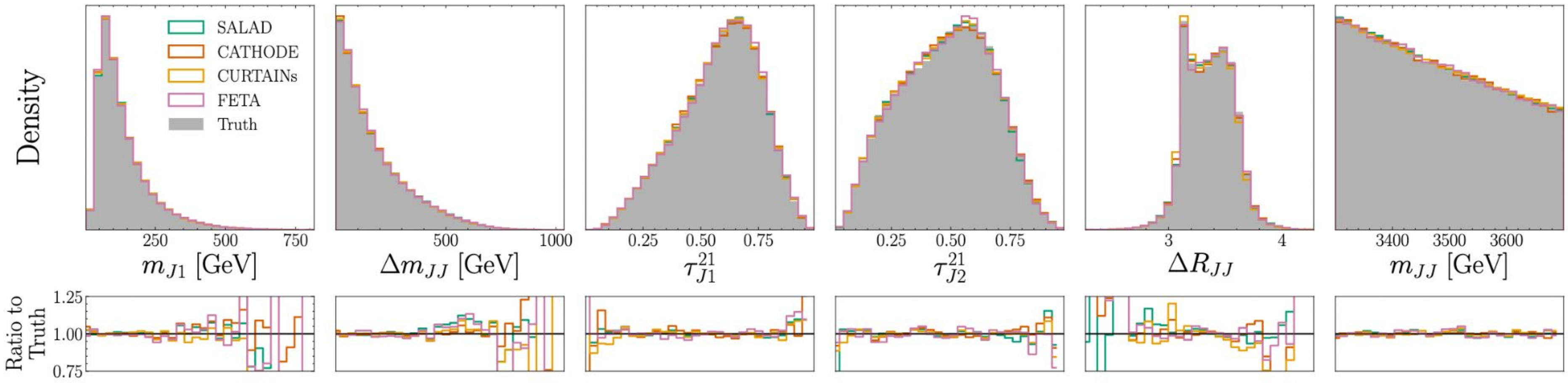
Many* ML techniques can construct the SM Template



What can we gain by ***combining*** these methods?

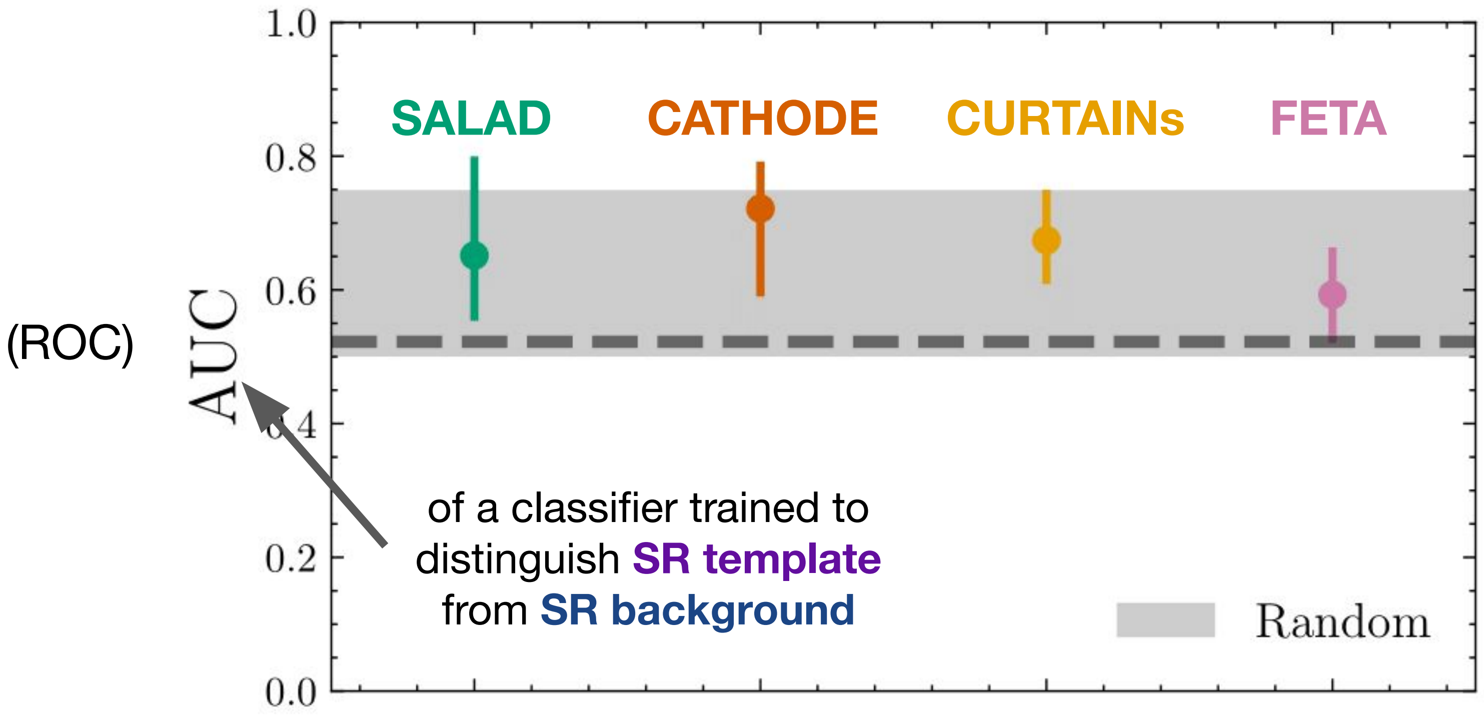
*I apologize if I've left out your favorite!

Study 0: do the methods reproduce the expected marginals?

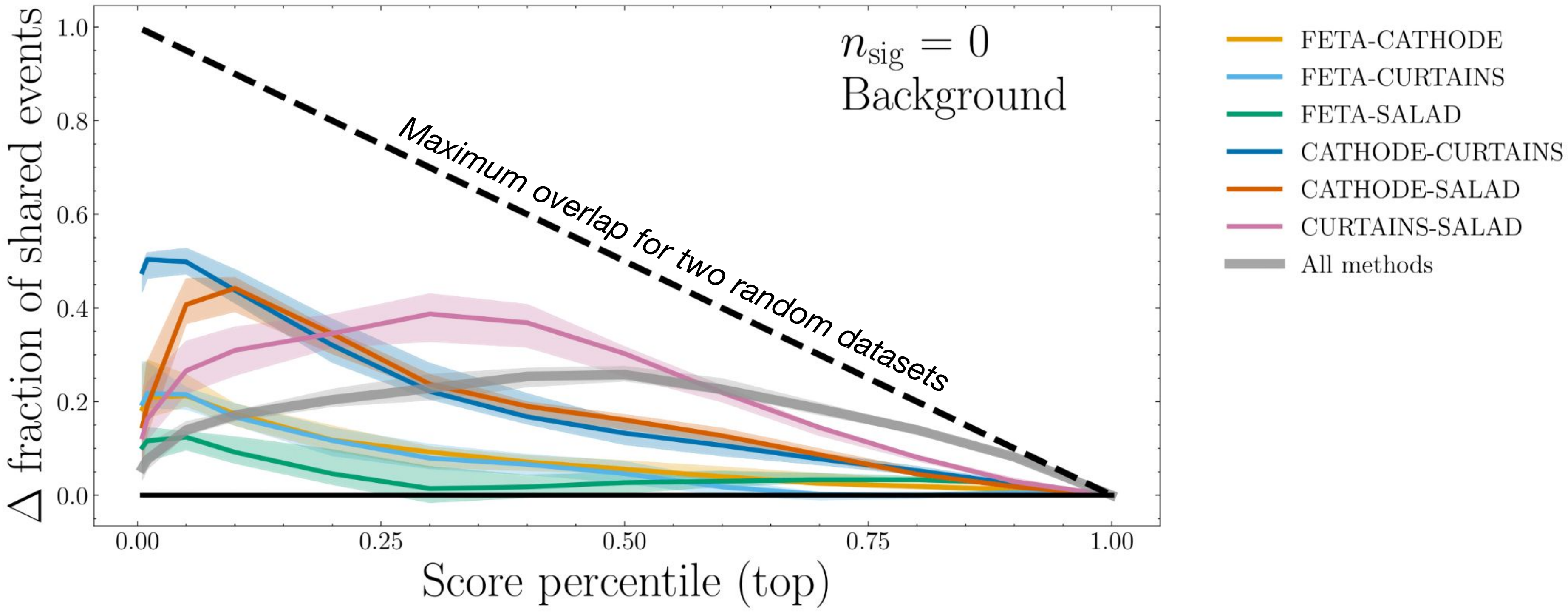


Dataset from the LHC Olympics

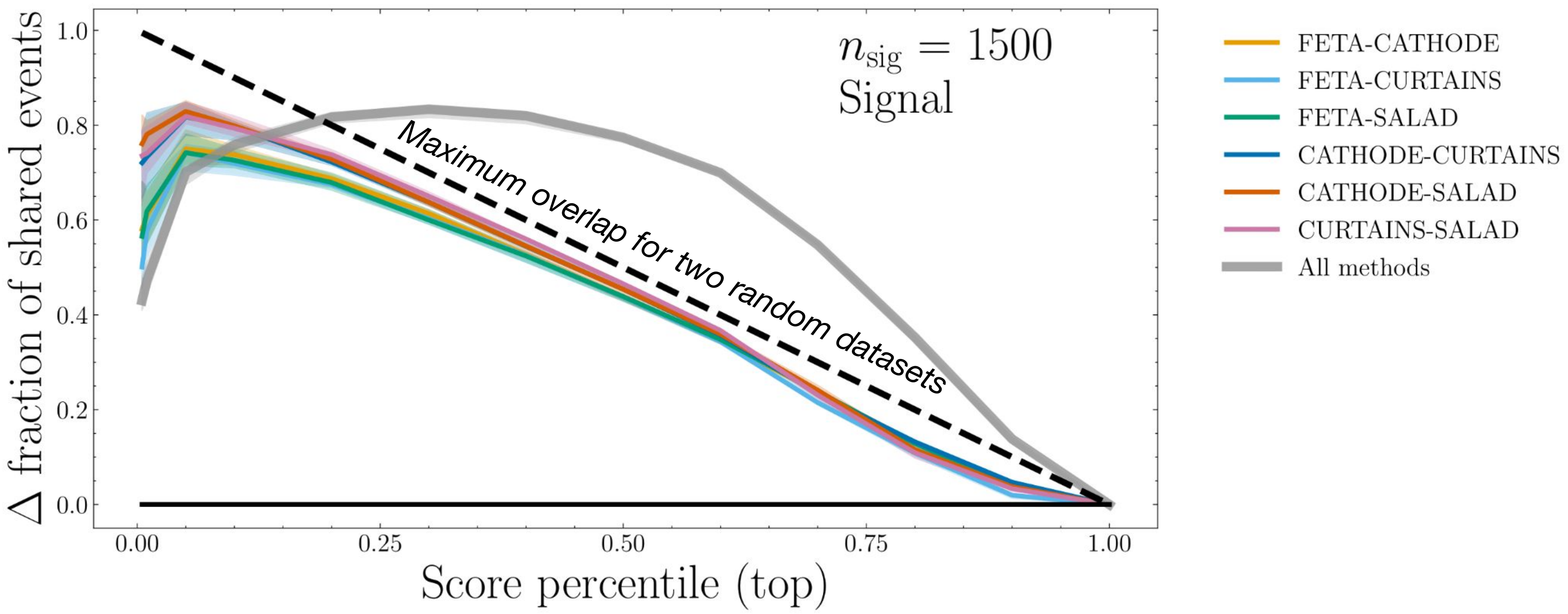
Study 1: are the samples good proxies for SM background?



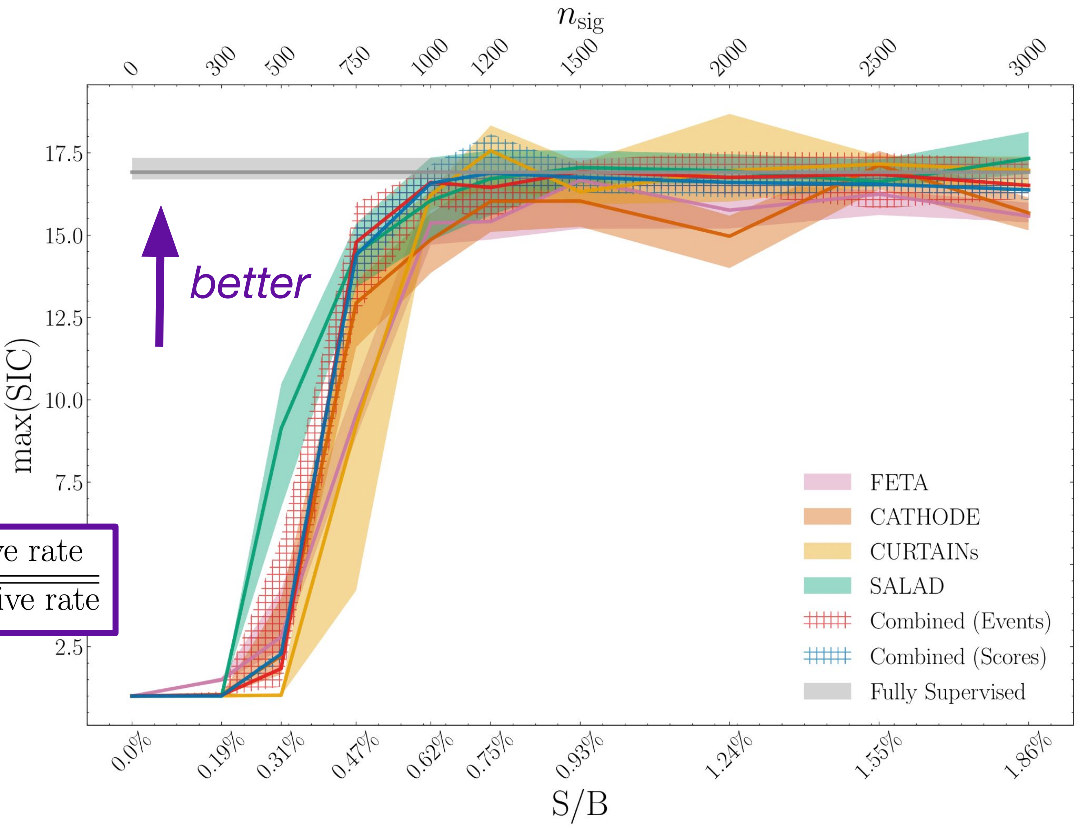
Study 2: do the samples agree on “anomalous” background?



Study 3: do the samples agree on signal?

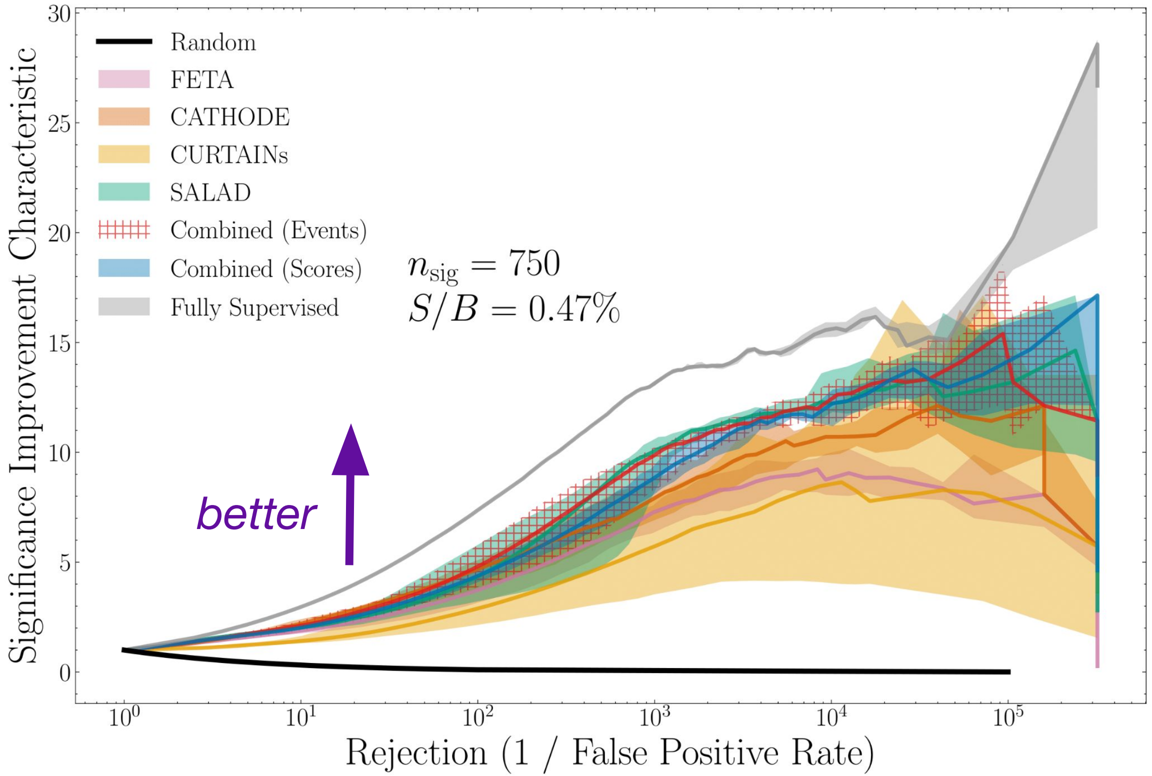


Combination appears to stabilize and improve performance…



$$\text{SIC} = \frac{\text{true positive rate}}{\sqrt{\text{false positive rate}}}$$

…across a range of signal efficiencies



Closing thoughts

- The LHC Olympics dataset has been almost exclusively used for resonant AD. **We should be testing on a variety of signal models!**
- SALAD appears* to beat the combined methods, but reweighting needs regions of overlapping support.
  - What signals **could break individual methods?**
  - Would the combination of samples still perform well on these models?
  - Note that sample combination can be **weighted** (though not explored here).

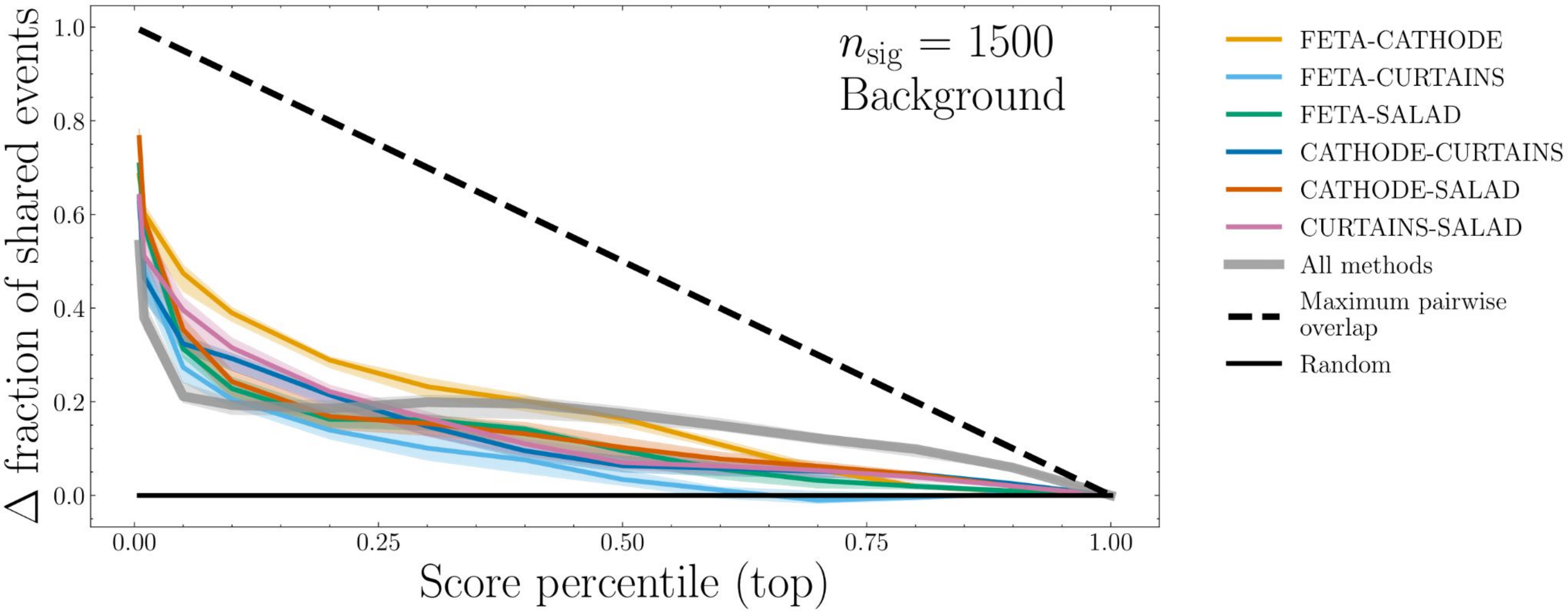
*This SALAD victory is likely a fluctuation – see backups!

Backup slides

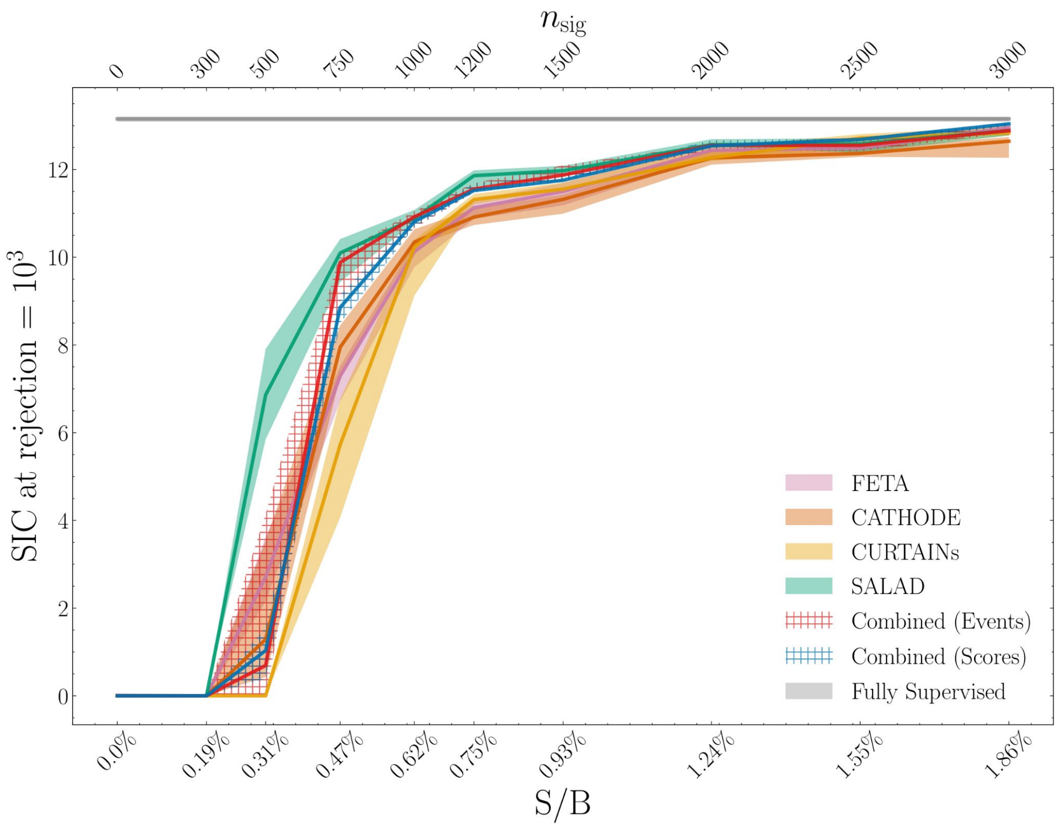
Number of generated samples

Method	Training data	Validation data	# samples	Oversampling
SALAD	793k SIM, 696k DAT	198K SIM, 174K DAT	1,045k	N/A
CATHODE	696k DAT	174K DAT	400k	3
CURTAINs	373k DAT	93k DAT	1,887k	4
FETA	793k SIM, 696k DAT	198K SIM, 174K DAT	732k	6

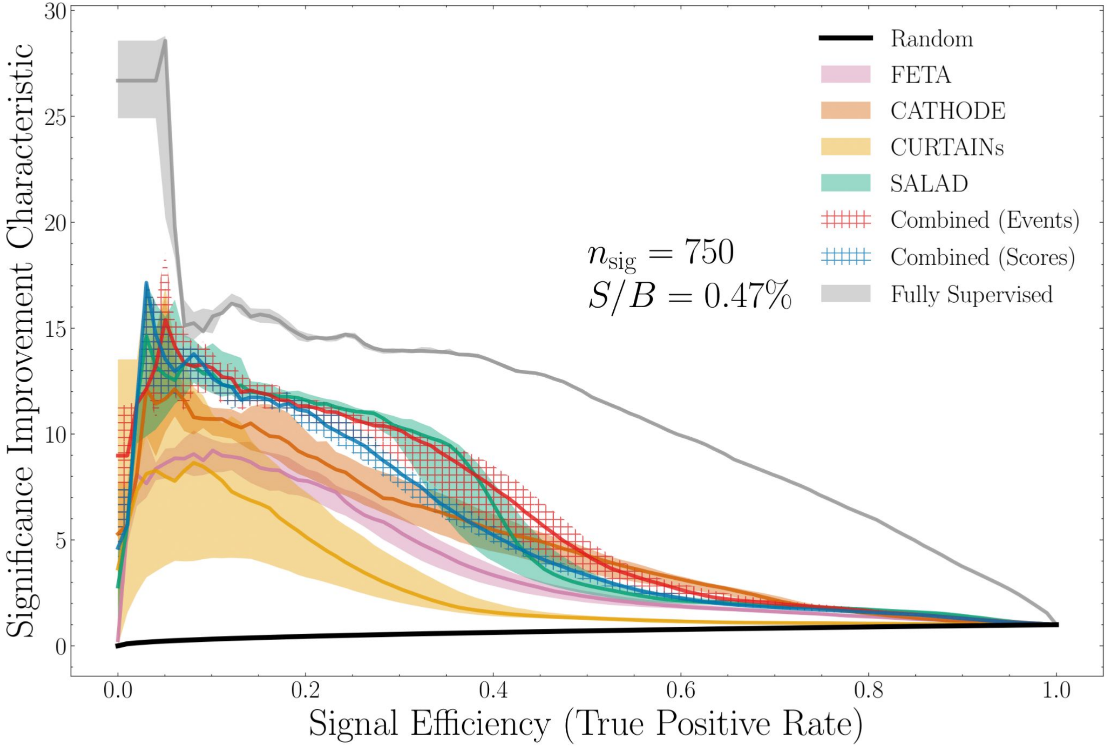
Background overlaps do not agree when there is signal



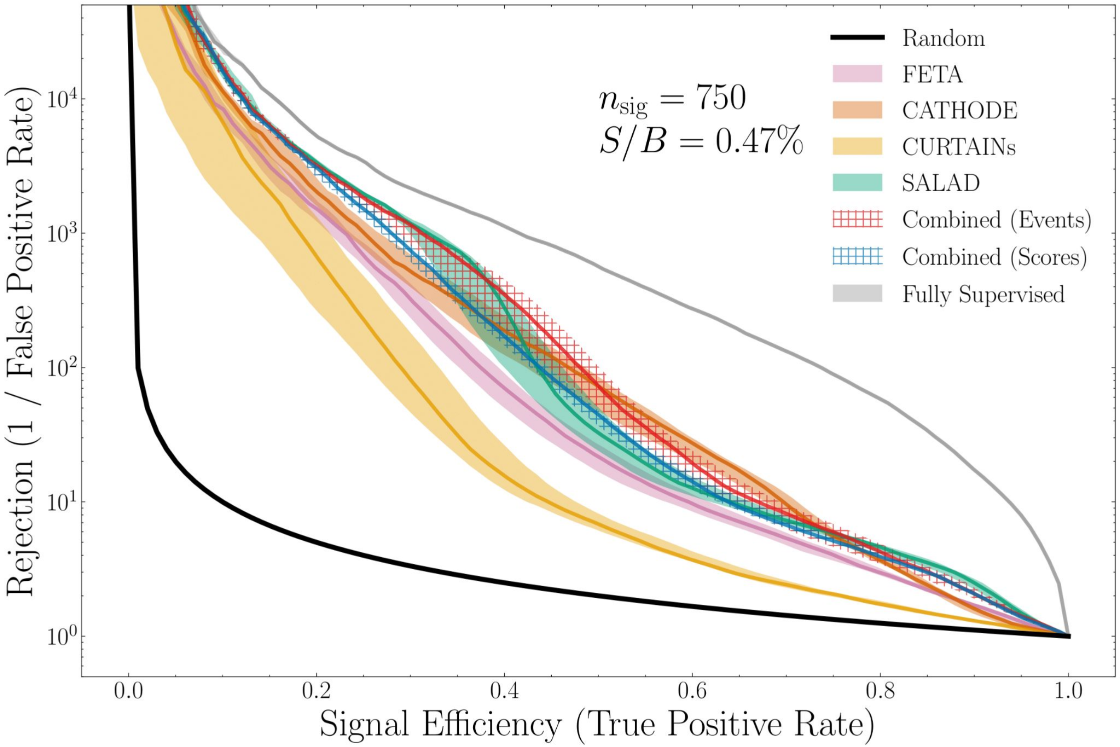
SIC at rejection = 1000



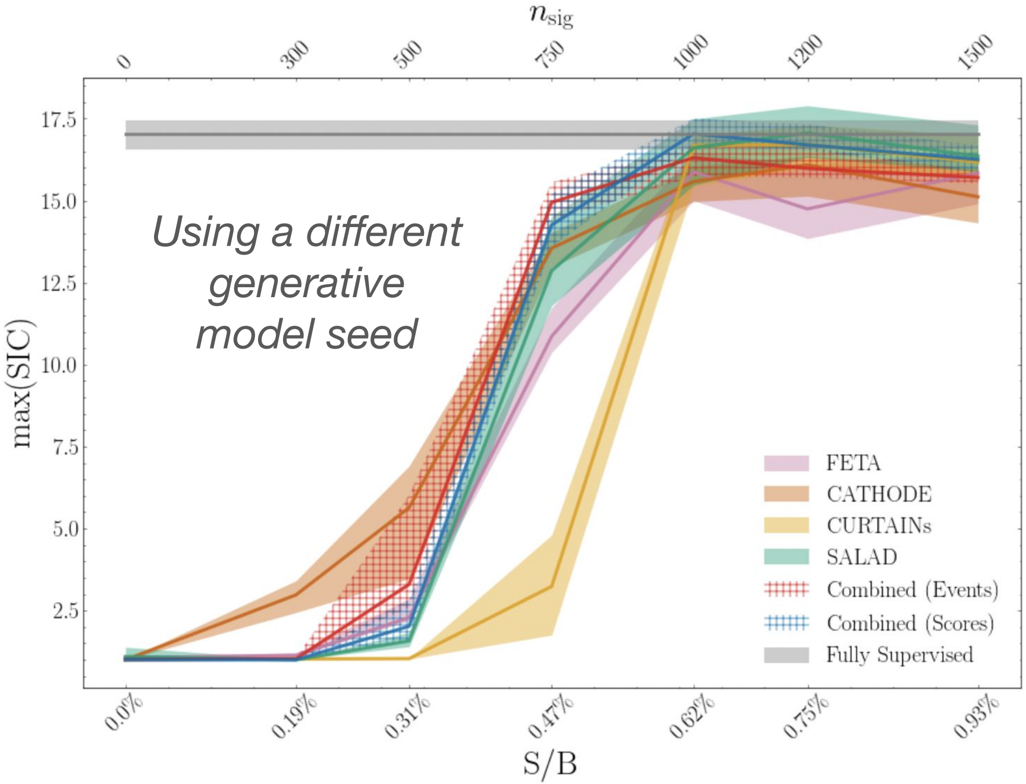
Classifier SIC for n_{sig} = 750

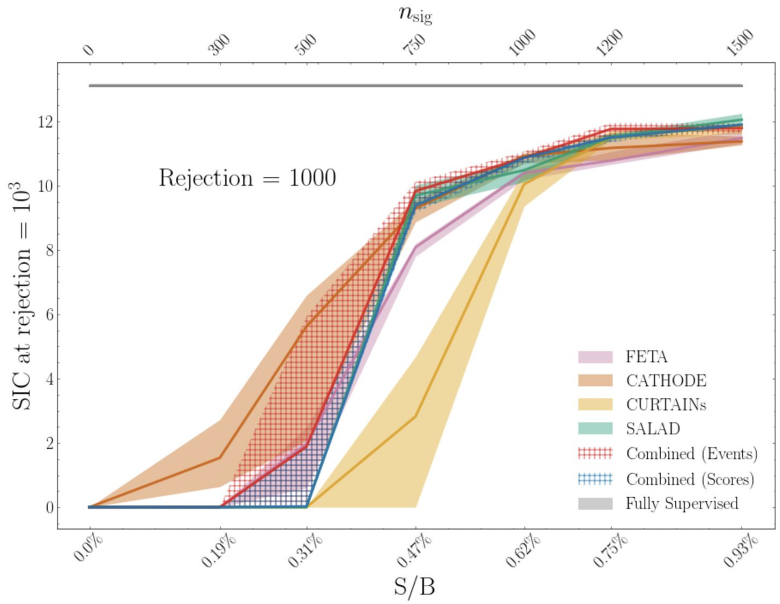


Classifier rejection for n_{sig} = 750

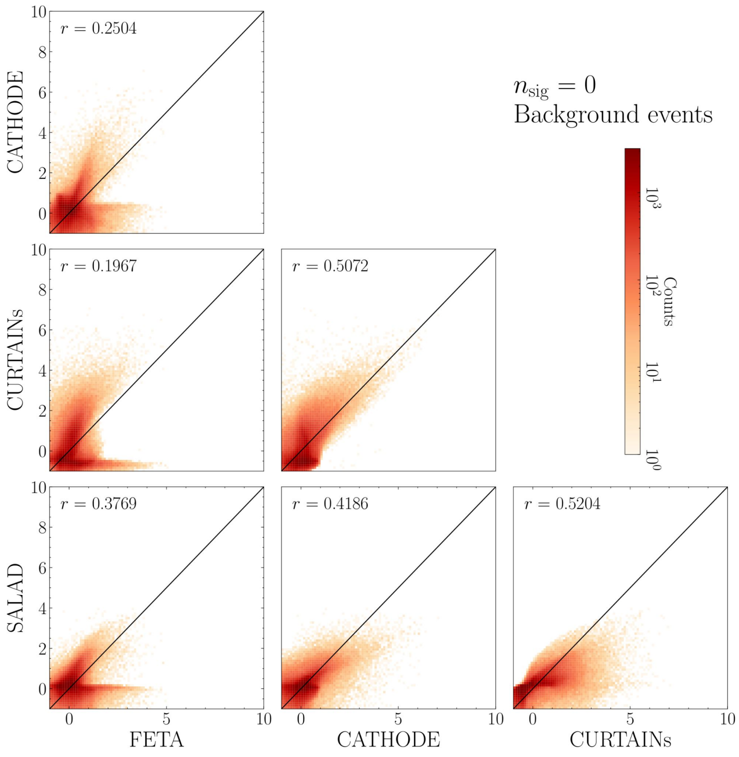


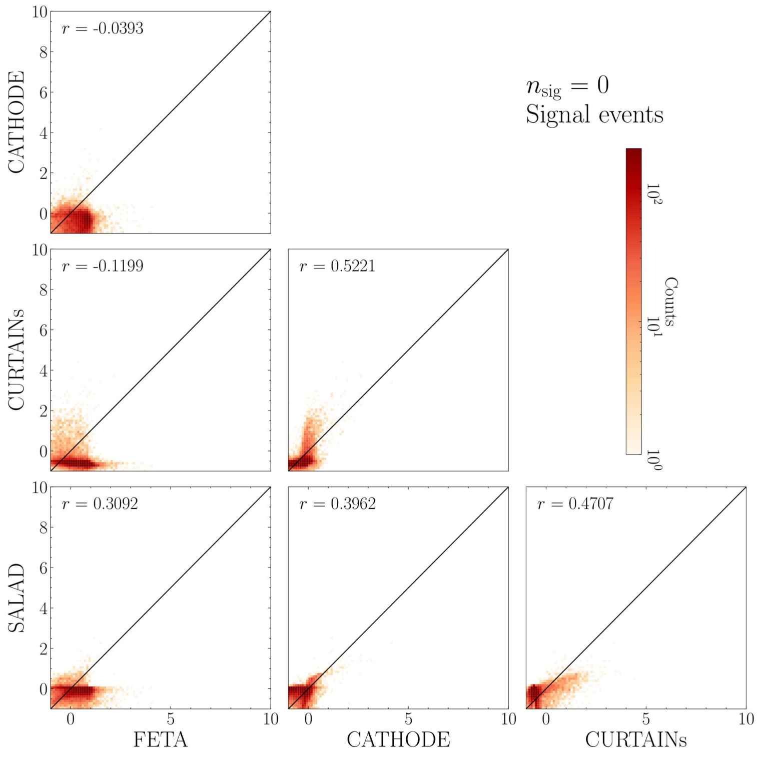
Around n_{sig} = 500, the AD task breaks down





Correlations of scores: background only





Correlations of scores: n_{sig} = 1500