Non-resonant Anomaly Detection

with background extrapolation

Kehang Bai, Radha Mastandrea, Ben Nachman

ML4Jets Workshop 11/09/2023

BERKELEY LAB

UNIVERSITY OF OREGON

Current landscape of Anomaly Detection

Requirements for complete Anomaly Detection

| Successful background estimation | ➕ | Sensitive to signal |

Many proposals and all data results

| Background interpolation from sideband | ➕ | resonant signal |

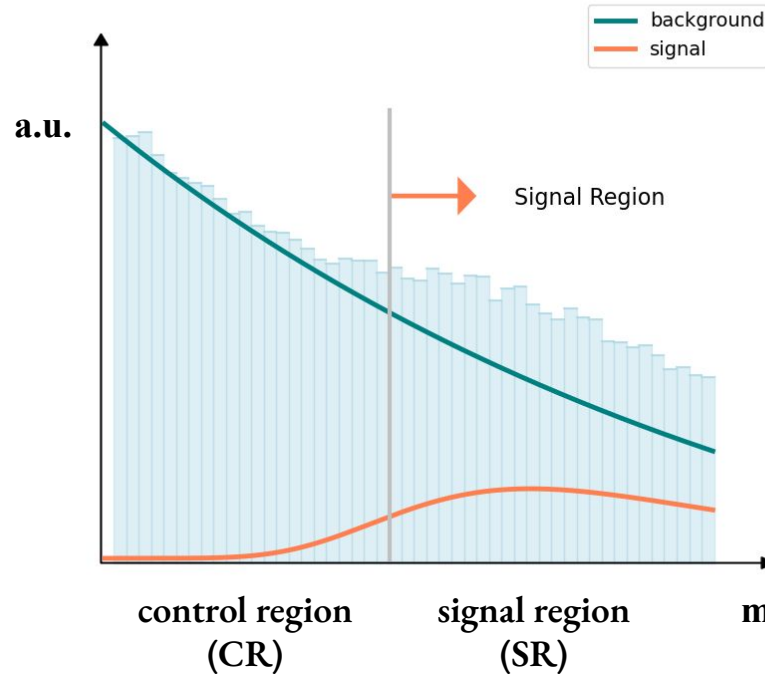Examples: SALAD, CATHODE, FETA, CURTAINs, ...

See Radha's talk for the interplay of resonant AD methods.

This proposal

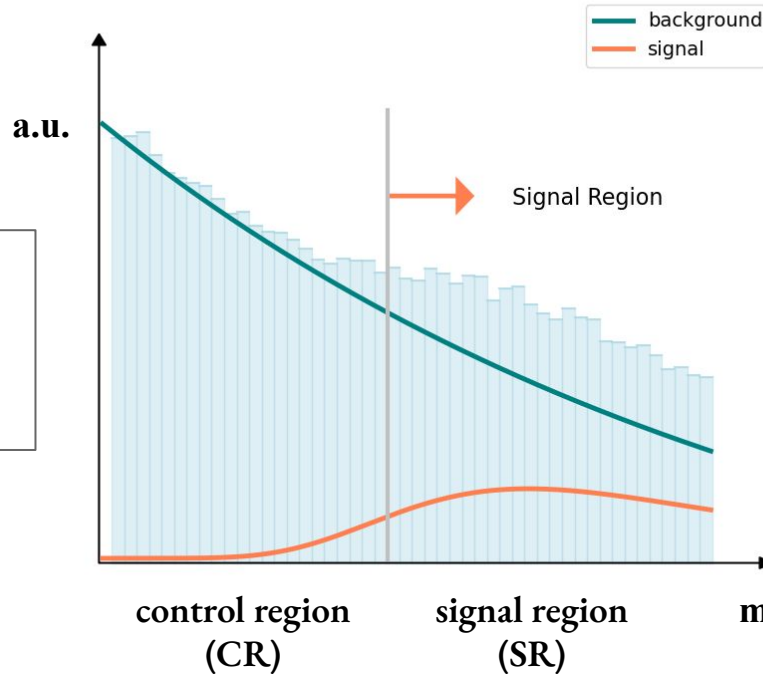| Background **extrapolation** from one side | **+** | **non-resonant** signal |

background
signal

a.u.

Signal Region

control region (CR)   signal region (SR)   m

This proposal

Background extrapolation from one side   **+**   **non-resonant signal**

See Gerrit's talk on tail-based AD.

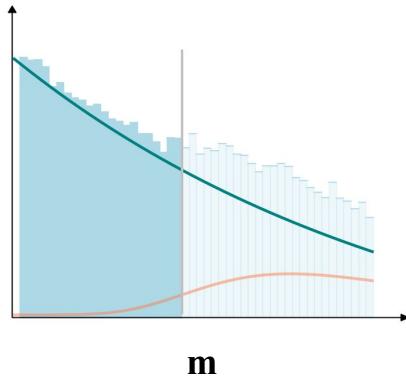Though using interpolation, it also used some non-resonance features.

The idea

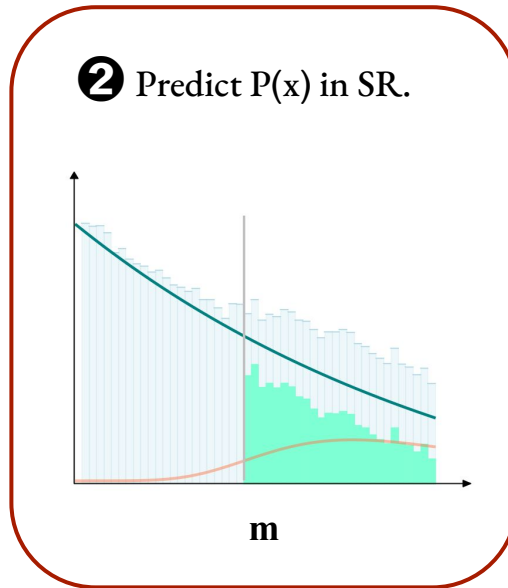➔ **Why don't we just extrapolate with existing methods?**

In principle we could use only MC-to-data reweighting (SALAD).

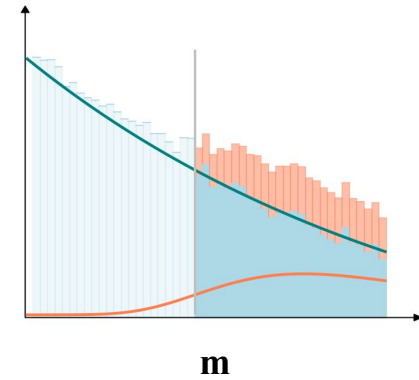However, density estimation with generative models are **less robust without a sideband.**

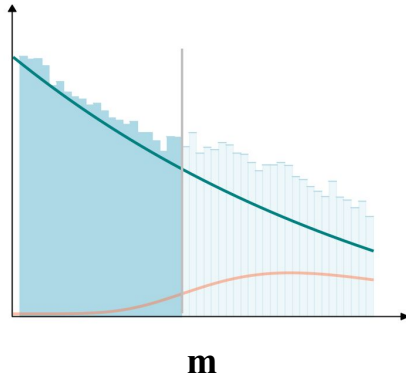❶ Learn P(x|m) in CR. ❷ Predict P(x) in SR. ❸ Classification in SR

The idea
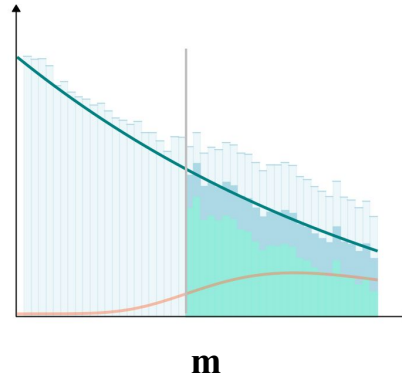
→ **Hybridize generated models with SALAD reweighting!**

Reweight conditional variables from MC to data in SR.

$$P_{\text{data}}(x, m) = P_{\text{data}}(x|m)P_{\text{data}}(m)$$
$$= P_{\text{data}}(x|m)P_{\text{MC}}(m)\frac{P_{\text{data}}(m)}{P_{\text{MC}}(m)}$$
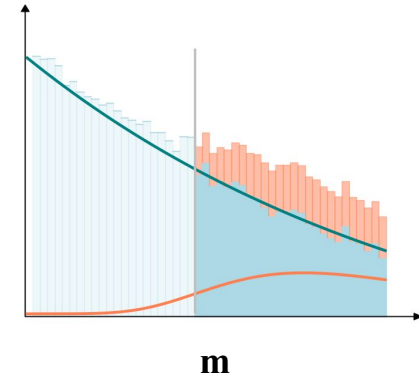$$= P_{\text{data}}(x|m)P_{\text{MC}}(m)w(m)$$

❶ Learn P(x|m) in CR.
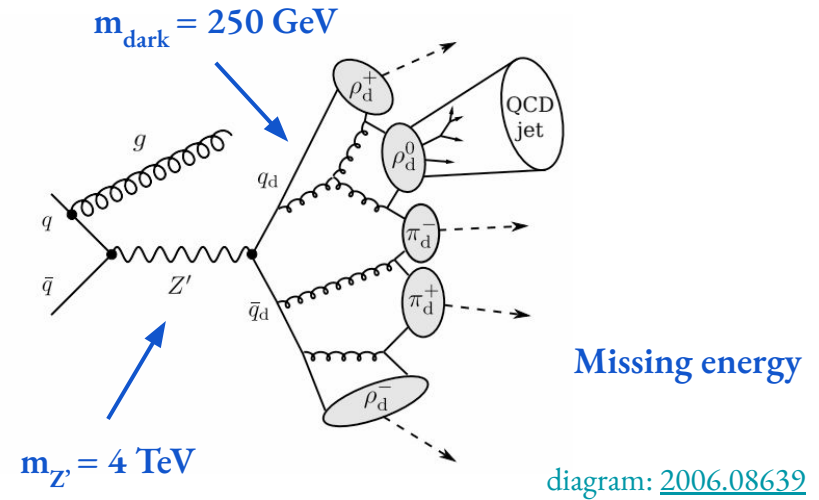
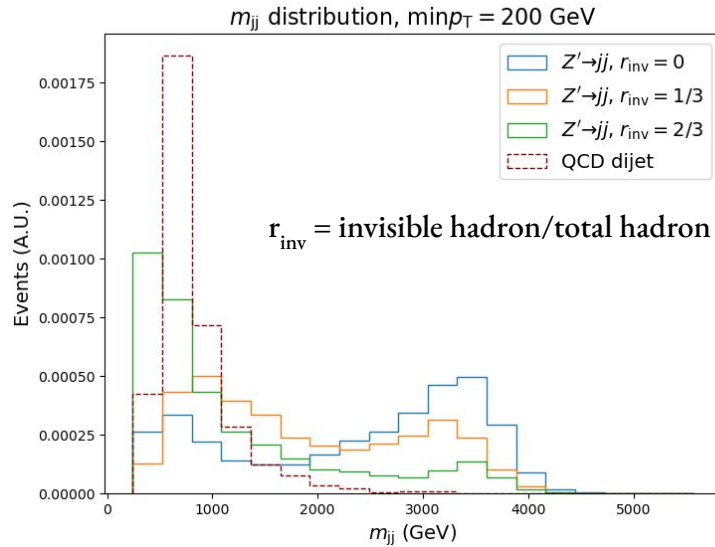❷ Predict P(x) in SR with reweighting.

❸ Classification in SR

m

m

m

Non−resonant signals

Case 1: large missing energy in the final states of a resonant production. Example here: semi-visible jets from Z'.

m_{jj} distribution, minp_T = 200 GeV

- $Z' \to jj$, $r_{inv} = 0$
- $Z' \to jj$, $r_{inv} = 1/3$
- $Z' \to jj$, $r_{inv} = 2/3$
- QCD dijet

r_{inv} = invisible hadron/total hadron

m_dark = 250 GeV

Missing energy

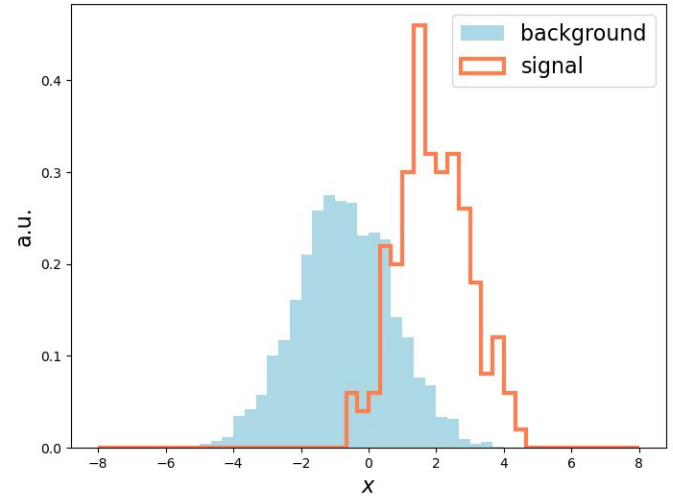**m_Z' = 4 TeV**

diagram: 2006.08639

Case 2: off-shell effects from heavy particles. Example: modifications to SMEFT coefficients, but not explored here.

First, let's take a look at a toy example

Context variables m1 and m2, used to define SR & CR.

A feature variable x.

→ Signal Region (SR)

→ Control Region (CR)
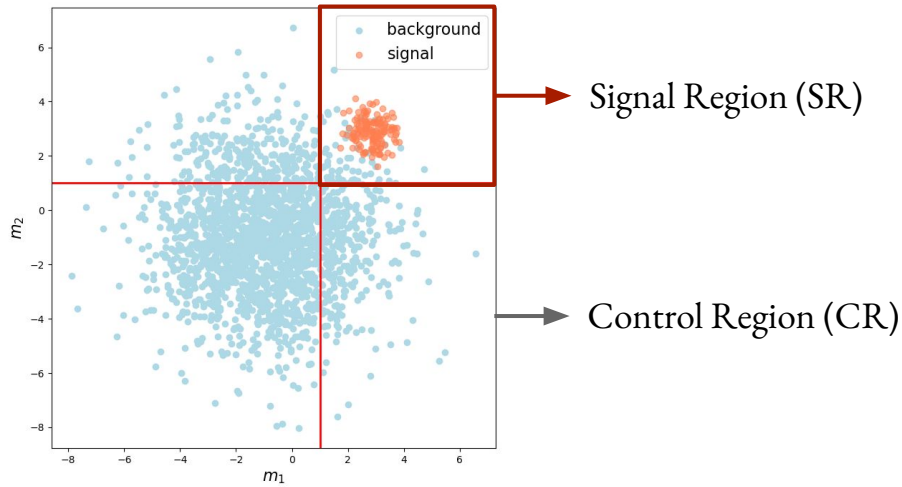
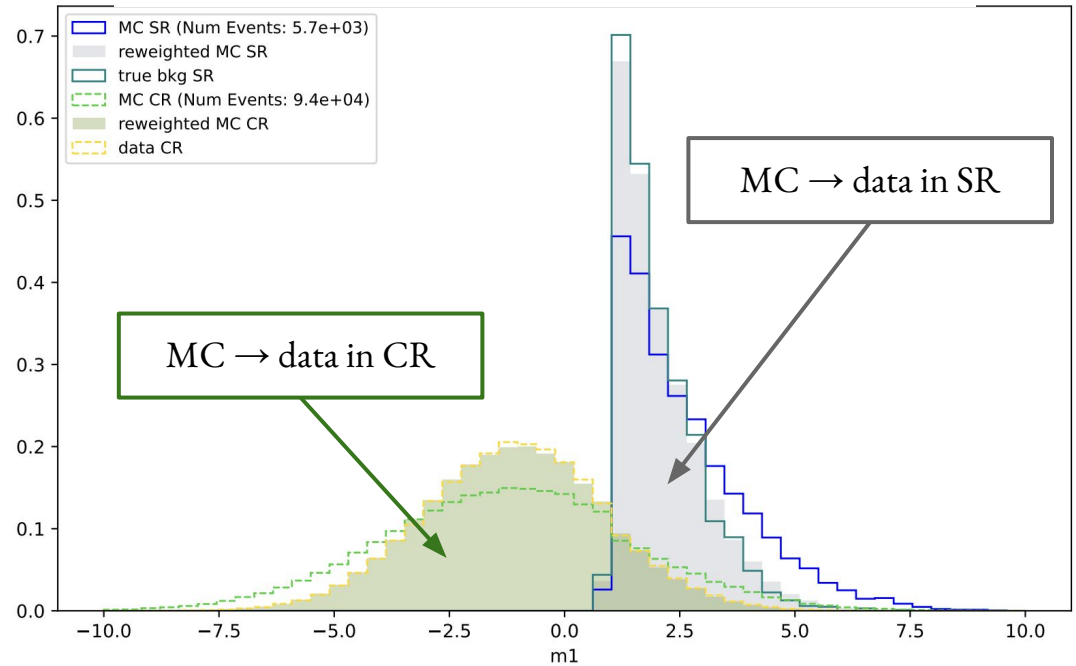**Test SALAD-only, CATHODE+SALAD, and FETA+SALAD.**

Anomaly detection methods

- Reweighting (SALAD)

Assumption:

$$w(m) = \frac{P_{data}(m \in CR)}{P_{MC}(m \in CR)}$$

$$= \frac{P_{data}(m \in SR)}{P_{MC}(m \in SR)}$$

Reweighted MC vs data for the toy background

Legend:
- MC SR (Num Events: 5.7e+03)
- reweighted MC SR
- true bkg SR
- MC CR (Num Events: 9.4e+04)
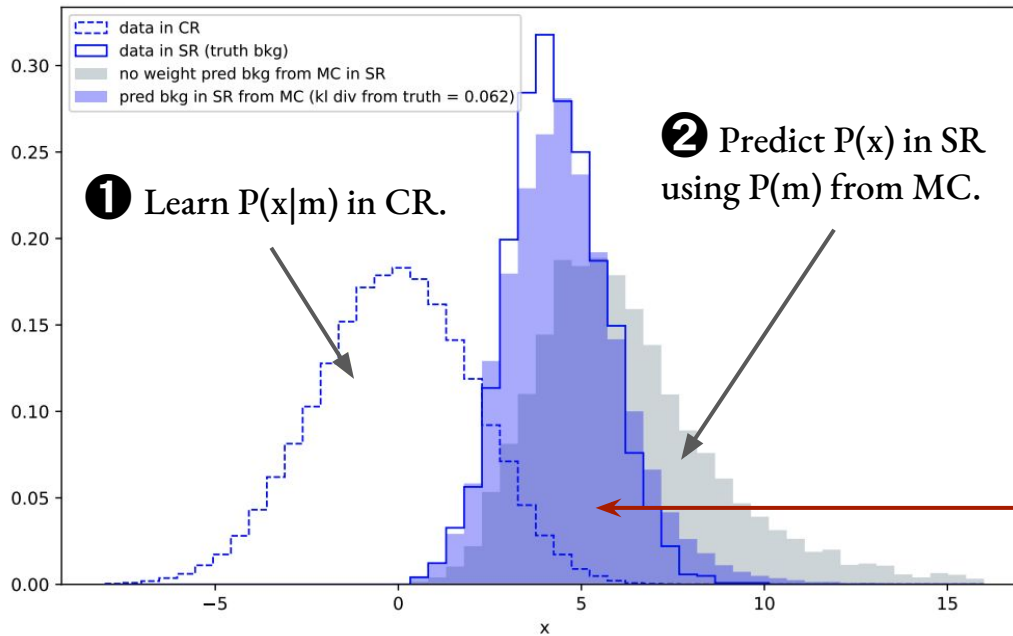- reweighted MC CR
- data CR

MC → data in SR

MC → data in CR

* This plot is used for illustration purpose only, not the final result.

Anomaly detection methods

- Density estimation (CATHODE & FETA + SALAD)

True vs predicted toy background

❶ Learn P(x|m) in CR.

❷ Predict P(x) in SR using P(m) from MC.

❸ Apply the weight function w(m).

Legend:
- data in CR
- data in SR (truth bkg)
- no weight pred bkg from MC in SR
- pred bkg in SR from MC (kl div from truth = 0.062)

* This plot is used for illustration purpose only, not the final result.

Results from the toy example

AD performance over increasing signal injections
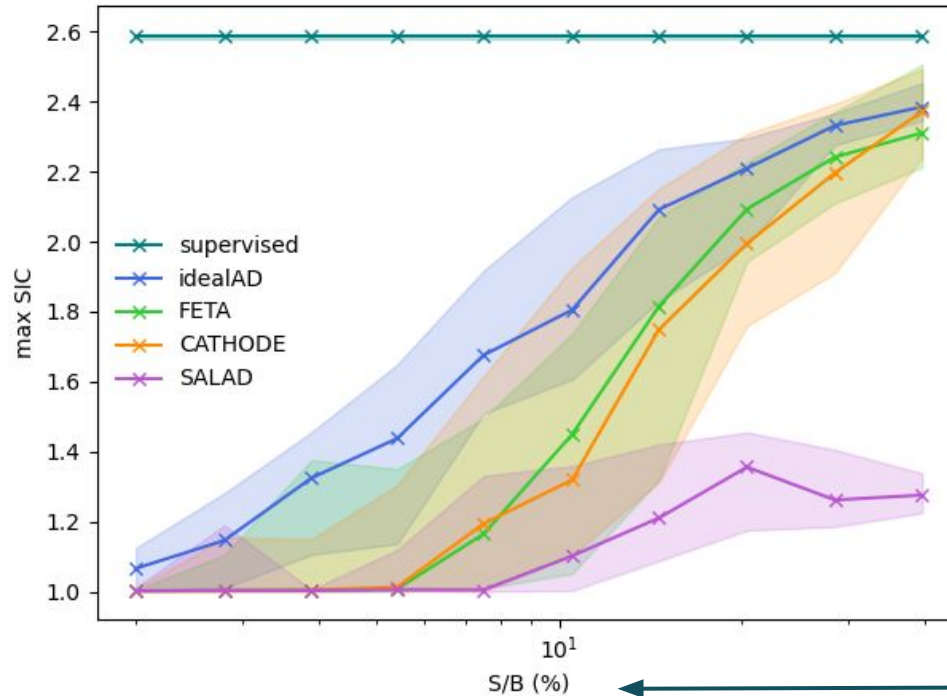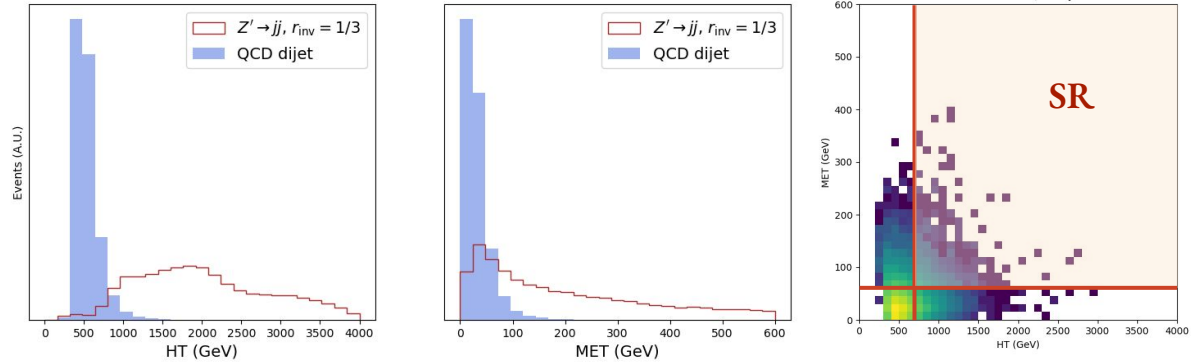

Maximum Significance Improvement
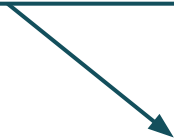
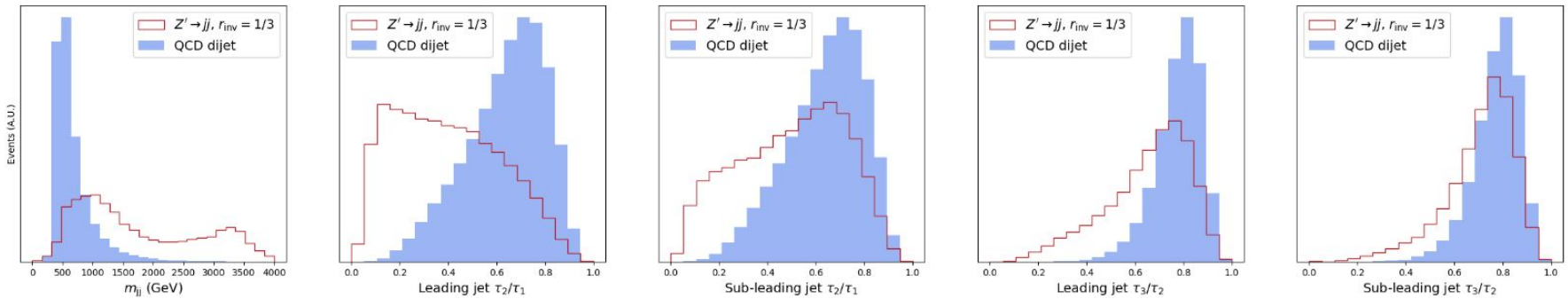Not yet thoroughly optimized, but early indication suggest the hybrid methods are better.

signal-to-background ratio
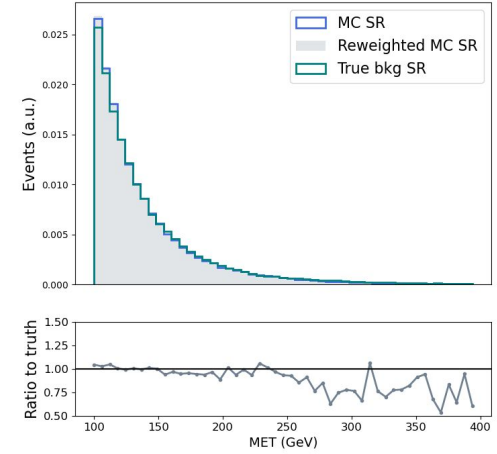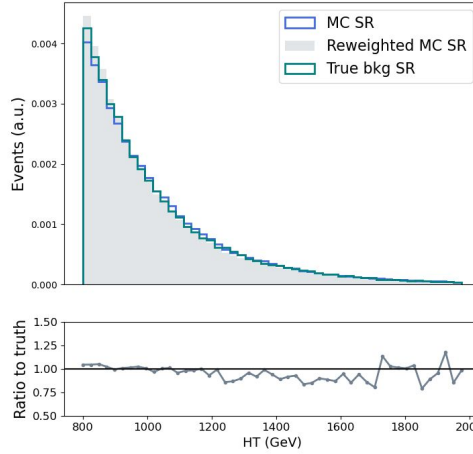
Physics case

Context variables: HT &MET
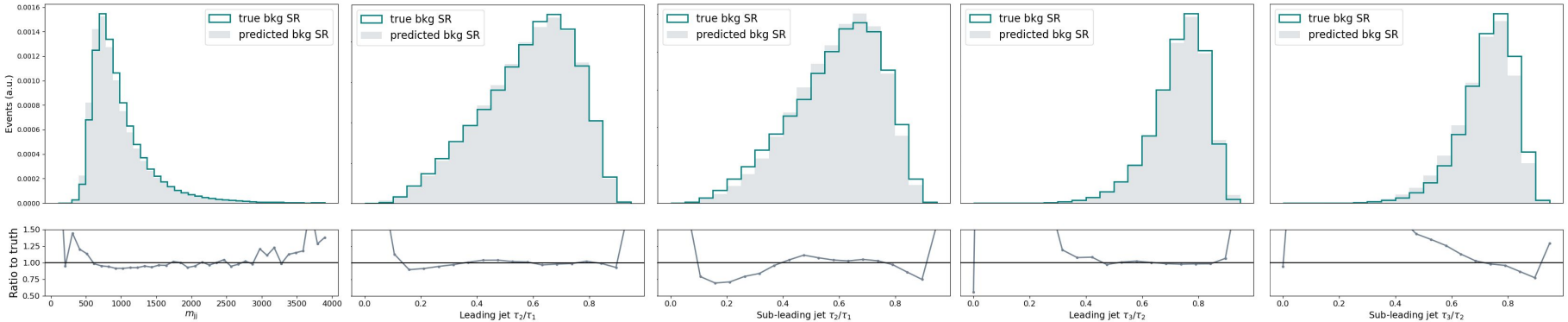

Feature variables: m$_{jj}$ & N-subjettiness
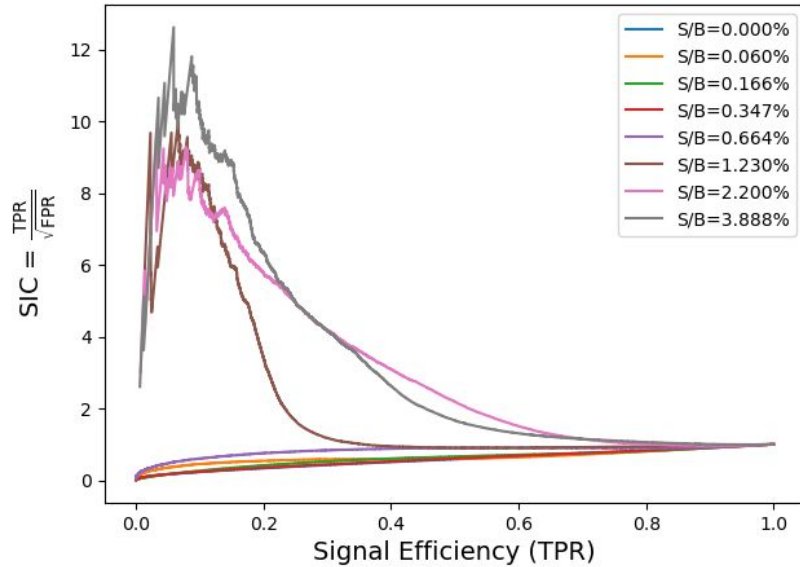
CATHODE + SALAD hybrid

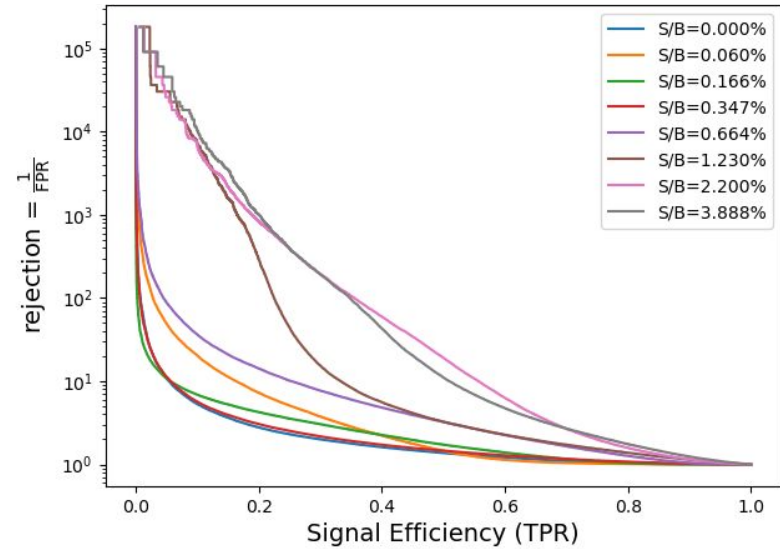**SALAD reweighting for contexts**

CATHODE extrapolation for features

CATHODE + SALAD hybrid

Very preliminarily performance plots, taken from one training

Significance Improvement

Rejection

Outlook

- **This is the first case of non-resonant AD using extrapolation!**

- **There are many interesting problems associated with extrapolation that are worth exploring in the future.**

- **Stay tuned for the full results!**

Thank you!

Kehang Bai

Back up

SR definition in this example

SR defined to be: HT > 600 GeV AND MET > 75 GeV

HT distribution, minp_T = 200 GeV — MET distribution, minp_T = 200 GeV — MET vs HT in QCD dijet