GEFÖRDERT VOM

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

Bundesministerium für Bildung und Forschung

RUTGERS
UNIVERSITY | NEW BRUNSWICK

Resonant anomaly detection without background sculpting

Manuel Sommerhalder[1],
Anna Hallin[2], Gregor Kasieczka[1], Tobias Quadfasel[1], and David Shih[2]

[1]University of Hamburg
[2]Rutgers University

manuel.sommerhalder@uni-hamburg.de

ML4Jets

November 3, 2022

Searches for new physics

Compelling motivation for existence of physics beyond the standard model (BSM)

Usually searching for BSM physics at the LHC via dedicated searches:

- pick specific BSM **signal model**
- optimize selections to enhance this signal
- check compatibility with SM background–only vs signal+background hypothesis

No BSM processes so far measured at the LHC. Why?

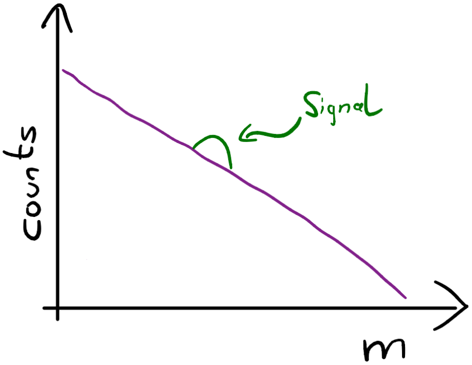
- no new physics at LHC energy scale?
- not yet searched for the right model?
- cannot cover all models with a dedicated search

$\rightarrow$ Need data-driven **model-independent** searches $\rightarrow$ anomaly detection

Resonant anomaly detection

Base assumptions:

- feature m with smooth background
- look for small signal localized in m



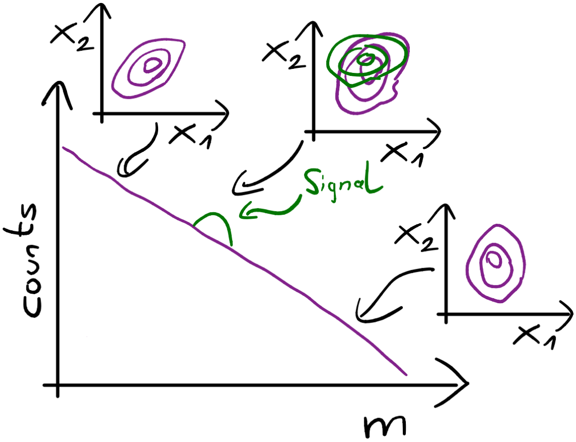
Resonant anomaly detection

Base assumptions:

- feature m with smooth background
- look for small signal localized in m

Auxiliary features:

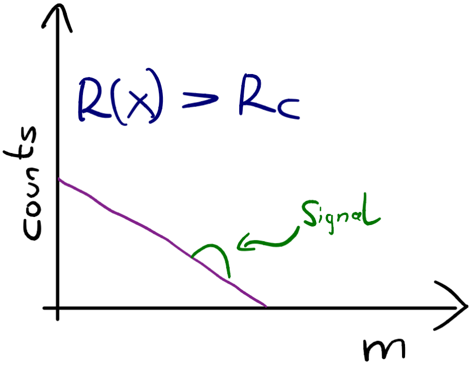
- additional dimesions x
- in general correlated with m
- signal shape different in x



Resonant anomaly detection

Enhanced bump hunt:

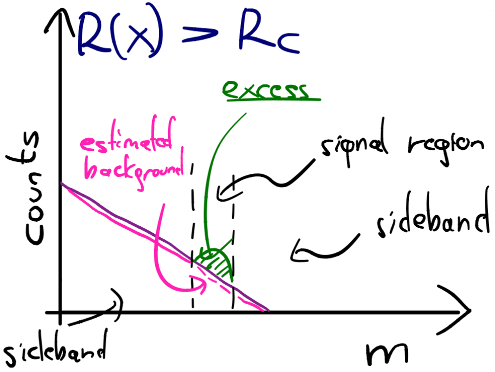
- find ML model $R(x)$ s.t. $R(x) > R_c$ enhances signal fraction



Resonant anomaly detection

Enhanced bump hunt:

- find ML model $R(x)$ s.t. $R(x) > R_c$ enhances signal fraction
- fit background from sidebands (SB)
- compare to data in signal region (SR)



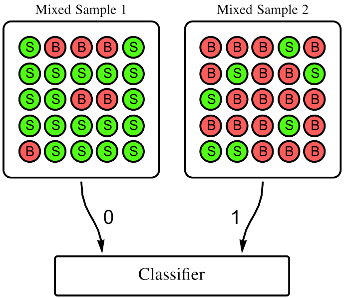
Weak supervision

Classification without labels (CWoLa)[a]

- a classifier trained on samples with mixed S/B ratio will still learn to distinguish signal and background

Combining with bump hunt (CWoLa Hunting)[b]

- train classifier on x from SR data vs SB data
- assuming $p_{\text{bkg}}(x|\text{SR}) \approx p_{\text{data}}(x|\text{SB})$
- then we learn $R(x) = \frac{p_{\text{data}}(x|\text{SB})}{p_{\text{bkg}}(x|\text{SR})} = \frac{p_{\text{data}}(x|\text{SR})}{p_{\text{bkg}}(x|\text{SR})}$
- breaks down in case of significant correlations between x and m



from ^{a}

---

[a] E. Metodiev, B. Nachman, J. Thaler, JHEP 10 (2017) 174
[b] J. Collins, K. Howe, B. Nachman, Phys. Rev. D 99, 014038 (2019)

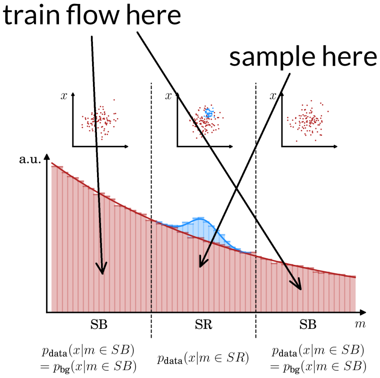
Dealing with correlated features

Classifying Anomalies THrough Outer Density Estimation (CATHODE)[a]:

- train normalizing flow on SB to learn $p_{\text{bkg}}(x|m \in \text{SB})$
- interpolate into SR and sample events bkg-like events from this $p_{\text{bkg}}(x|\text{m} \in \text{SR})$
- train classifier between SR data and bkg-like samples
- learns $R(x) = \frac{p_{\text{data}}(x|\text{SR})}{p_{\text{bkg}}(x|\text{SR})}$ also under correlations

Conditional normalizing flow:

- learn invertible map $f(x;m)$ from x to gaussian latent space z for every m

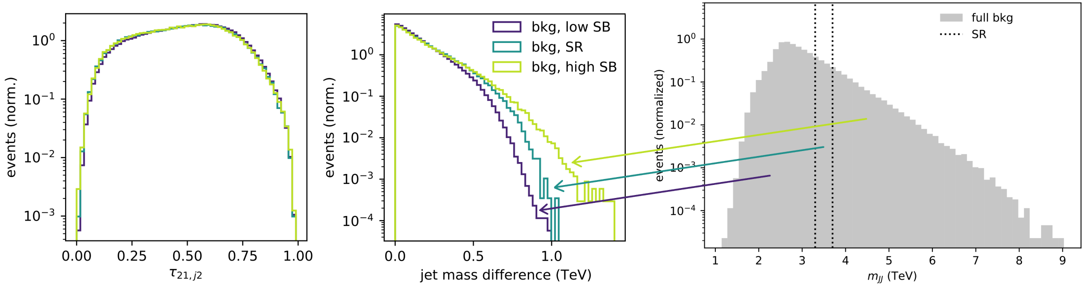


adjusted from [a]

[a] M.S., Hallin, et al., Phys. Rev. D 106, 055006 (2022)

Testing on LHCO R&D dataset

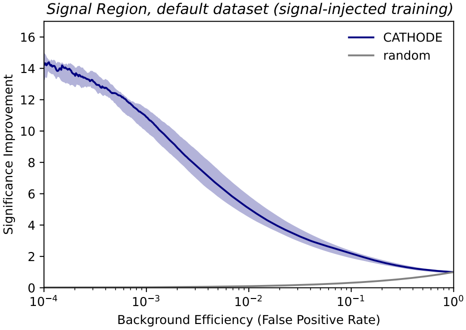
- LHCO R&D dataset[1]: simulated dijet resonance search
- signal: $Z'_{3.5\text{TeV}} \to X_{500\text{GeV}} Y_{100\text{GeV}}$
- resonant feature m: m_{jj}
- auxiliary features x: m_{j1}, Δm_j, $\tau_{21,j1}$, $\tau_{21,j2}$
- minor differences between SR and SB x

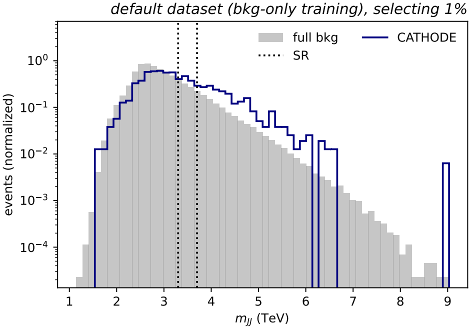


[1] arXiv:2101.08320

Testing on LHCO R&D dataset

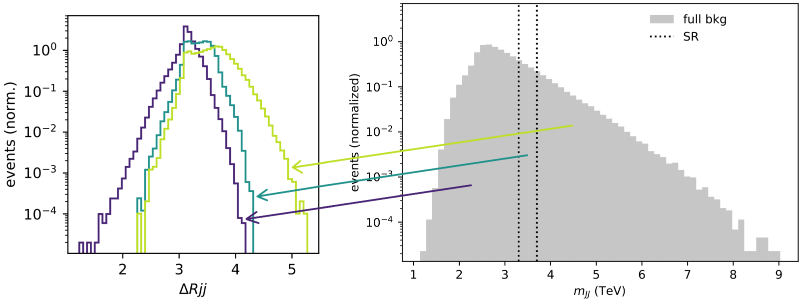
- high signal sensitivity
- significance improvement = significance after cut over sign. before cut
- slight change in m_{jj} bkg distribution after $R(x) > R_c(x)$
- looks still well fittable for enhanced bump hunt





Testing on LHCO R&D dataset – adding another feature

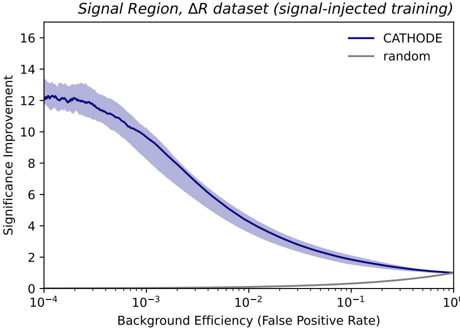
- adding angular dijet separation ΔR_{jj} as input feature[2]
- more pronounced differences between SR and SB



[2] first suggested by Raine et al., arXiv:2203.09470

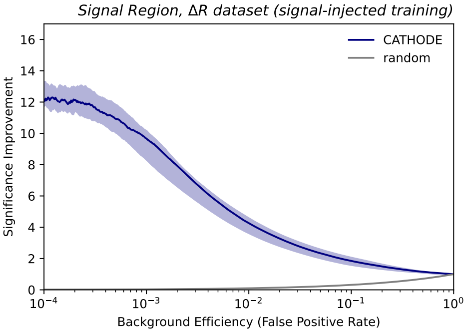
Testing on LHCO R&D dataset – adding another feature

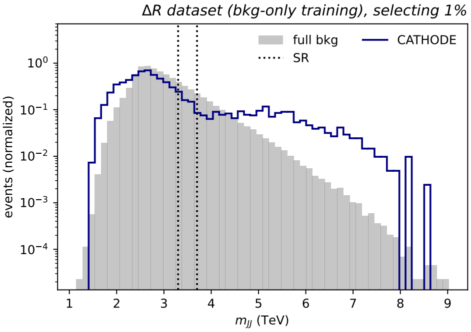
- still high signal sensitivity



Testing on LHCO R&D dataset – adding another feature

- still high signal sensitivity
- heavy change in m_{jj} bkg distribution after $R(x) > R_c(x) \rightarrow$ sculpting
- difficult to fit bkg from SB here :(

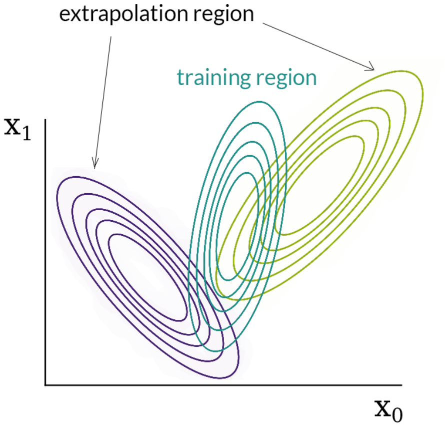




Background sculpting

When x depends on m:

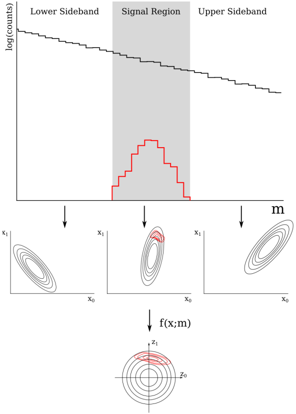
- background m shape changes if we cut on learned $R(x) = R(x(m))$
- evaluating SR-trained $R(x)$ on SB can cause uncontrolled extrapolation
- smooth bkg fit in bump hunt increasingly difficult when m is sculpting



Moving to latent space

Latent CATHODE (LaCATHODE)

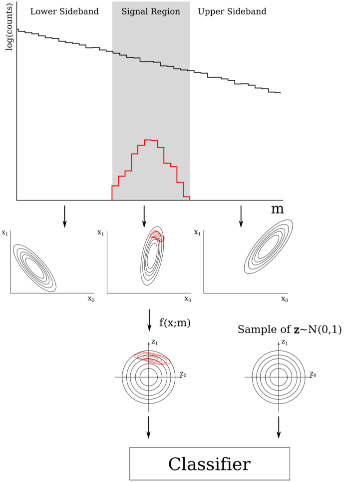
- use flow to map SR data to latent space $z = f(x; m)$



Moving to latent space

Latent CATHODE (LaCATHODE)

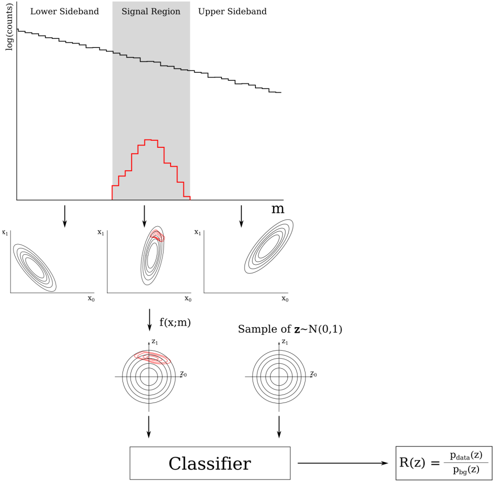
- use flow to map SR data to latent space $z = f(x; m)$
- perform classification task in latent space
- sample bkg from unit gaussian



Moving to latent space

Latent CATHODE (LaCATHODE)

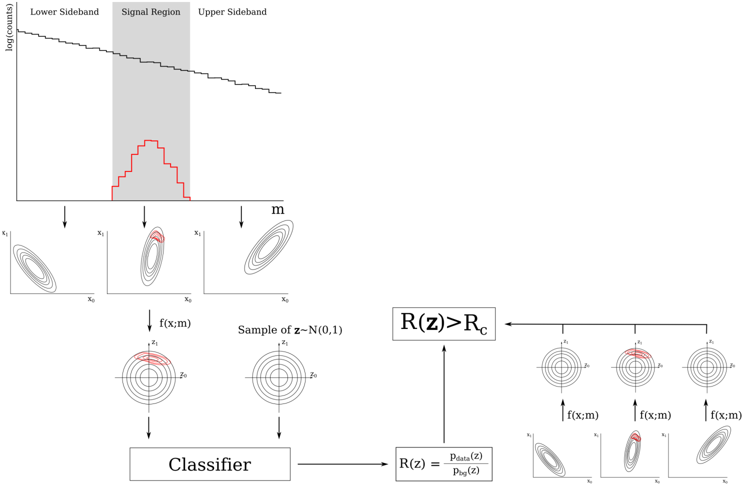
- use flow to map SR data to latent space $z = f(x; m)$
- perform classification task in latent space
- sample bkg from unit gaussian
- learning $R(z) = \frac{p_{\text{data}}(z|\text{SR})}{p_{\text{bkg}}(z|\text{SR})} = \frac{p_{\text{data}}(x|\text{SR})}{p_{\text{bkg}}(x|\text{SR})}$



Moving to latent space

Latent CATHODE (LaCATHODE)

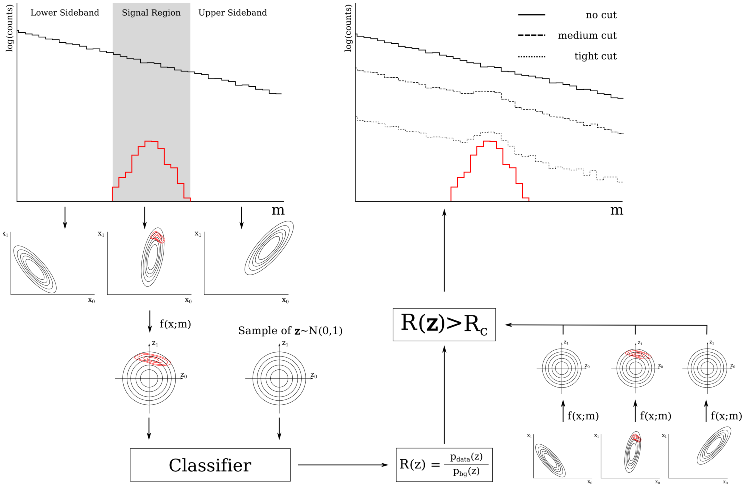
- use flow to map SR data to latent space $z = f(x; m)$
- perform classification task in latent space
- sample bkg from unit gaussian
- learning $R(z) = \frac{p_{\text{data}}(z|\text{SR})}{p_{\text{bkg}}(z|\text{SR})} = \frac{p_{\text{data}}(x|\text{SR})}{p_{\text{bkg}}(x|\text{SR})}$
- mapping SB to same latent space for $R(z) > R_c$



Moving to latent space

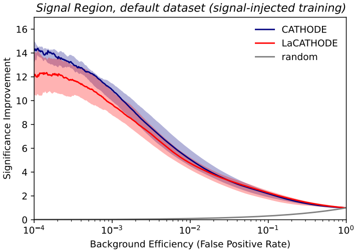
Latent CATHODE (LaCATHODE)

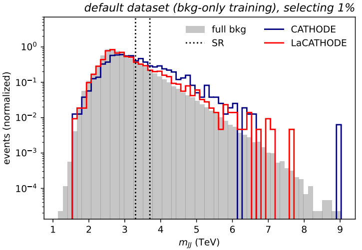
- use flow to map SR data to latent space $z = f(x; m)$
- perform classification task in latent space
- sample bkg from unit gaussian
- learning $R(z) = \frac{p_{\text{data}}(z|\text{SR})}{p_{\text{bkg}}(z|\text{SR})} = \frac{p_{\text{data}}(x|\text{SR})}{p_{\text{bkg}}(x|\text{SR})}$
- mapping SB to same latent space for $R(z) > R_c$



Testing on LHCO R&D dataset – LaCATHODE

without ΔR_{jj}

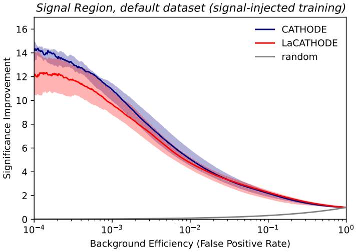


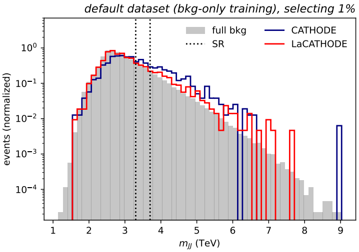


Testing on LHCO R&D dataset – LaCATHODE

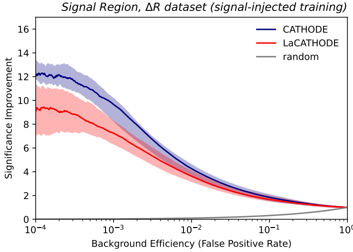
- LaCATHODE eliminates sculpting
- retains much of CATHODE signal sensitivity
- some reduction with ΔRjj, but perfect bkg stability will pay off in actual bump hunt

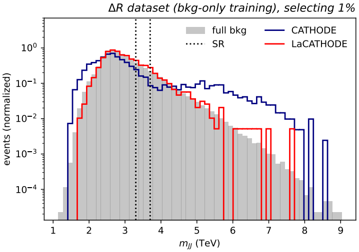
without ΔR_{jj}





with ΔR_{jj}





Summary & conclusion

- bump hunt–enhancing anomaly detection methods can sculpt the background when features are correlated
- the background estimation becomes increasingly difficult with more correlation
- La(tent)CATHODE solves this issue by performing weakly supervised classification in a decorrelated latent space
- LaCATHODE is no more complex than classic CATHODE
- we observe significantly less sculpting with LaCATHODE than with other protocols but similar signal sensitivity

→ *read more at arXiv:2210.14924*

BACKUP

LHCO 2020 R&D dataset

- Benchmarking on the LHC Olympics 2020 challenge R&D dataset
  - Black box challenge whose now uncovered signals are still used for benchmarking
  - Generic LHC-like DELPHES simulation
  - *arXiv:2101.08320*
- Background: 1M simulated QCD multijet events
- Signal: 100k Z'$\rightarrow$XY events where X$\rightarrow$qq and Y$\rightarrow$qq
- $m_{Z'} = 3.5$ TeV, $m_X = 500$ GeV, $m_Y = 100$ GeV
- Select 2 most massive jets in each event and use their dijet mass m_{jj} to search for resonances
- Use 4 additional "high-level" variables:
  - Lower jet mass m_{j1} & mass difference $\Delta m_{1,2}$
  - Jet subjettiness ratios $\tau_{21,1}$ and $\tau_{21,2}$

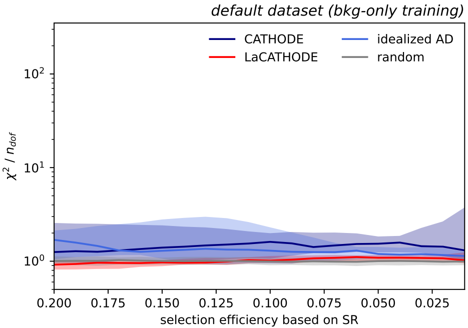
Quantifying background sculpting

- quantifying difference between m_{jj} spectra before and after cut using

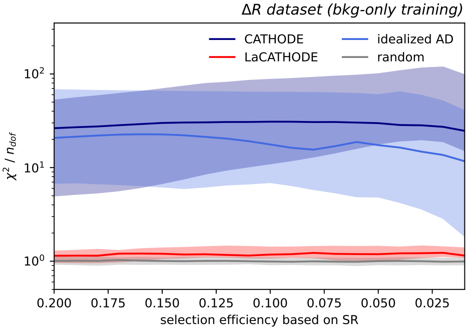
$$\chi^2/n_{\text{dof}}, \quad \text{with} \quad \chi^2 = \sum_{i=1}^{n_{\text{bins}}} \frac{\left(n_i^{cut} - n_i^{uncut}\right)^2}{\sigma_i^2}, \quad n_{\text{dof}} = n_{\text{bins}} - 1$$

- choosing binning such that every bin has > 10 expected events

without ΔR_{jj}

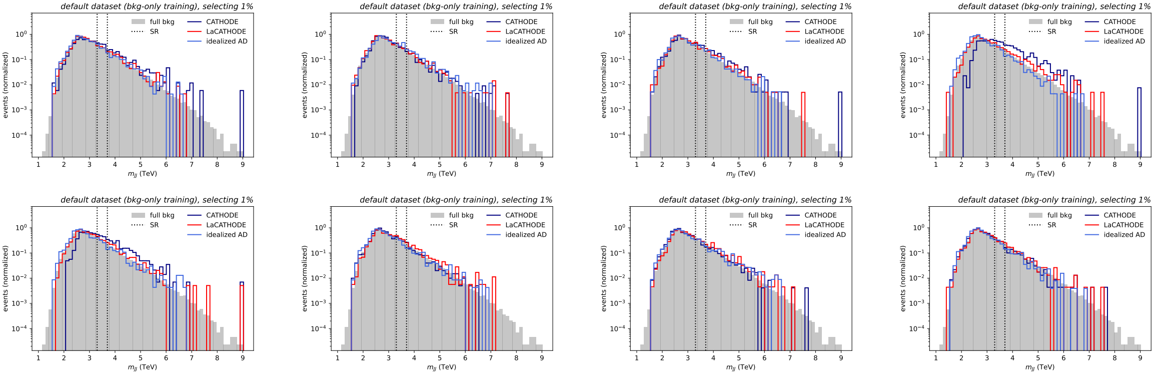


with ΔR_{jj}



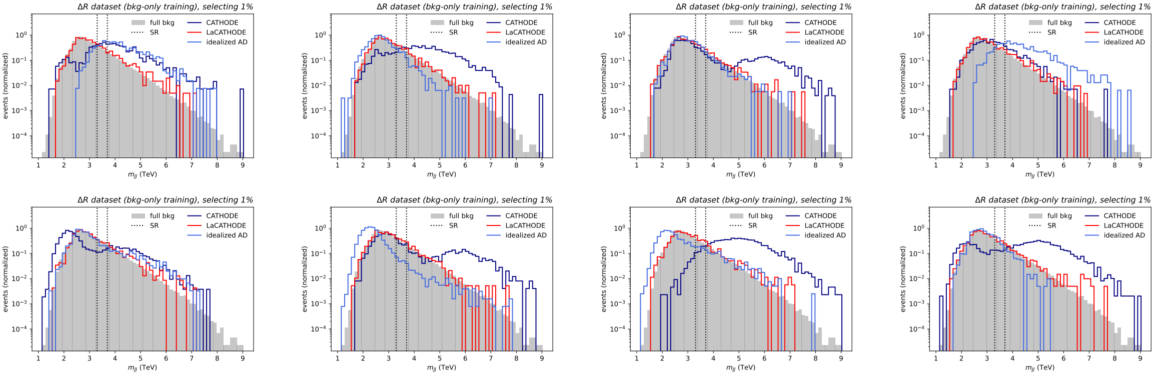
Varying runs on LHCO R&D dataset – default features only

Retraining the (La)CATHODE pipeline multiple times:



Varying runs on LHCO R&D dataset – adding dR

Retraining the (La)CATHODE pipeline multiple times using ΔR_{jj} input:



$\rightarrow$ extrapolation behavior is unstable between retrainings, except for LaCATHODE