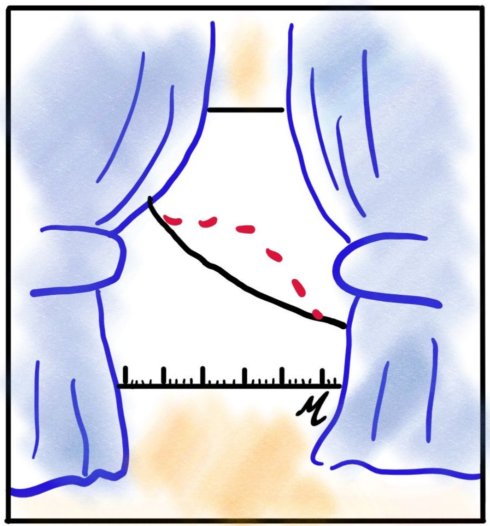
CURTAINs



Johnny Raine, Samuel Klein, Debajyoti Sengupta, Tobias Golling

University of Geneva

5th Inter-experiment Machine Learning Workshop

What is CURTAINS?

With CURTAINS, we are…

Constructing **U**nobserved **R**egions by **T**ransforming **A**djacent **IN**terval**s**

- **Data driven** method for generating background template with a set of features
- Account for correlations between features and resonant variable

Bump hunts

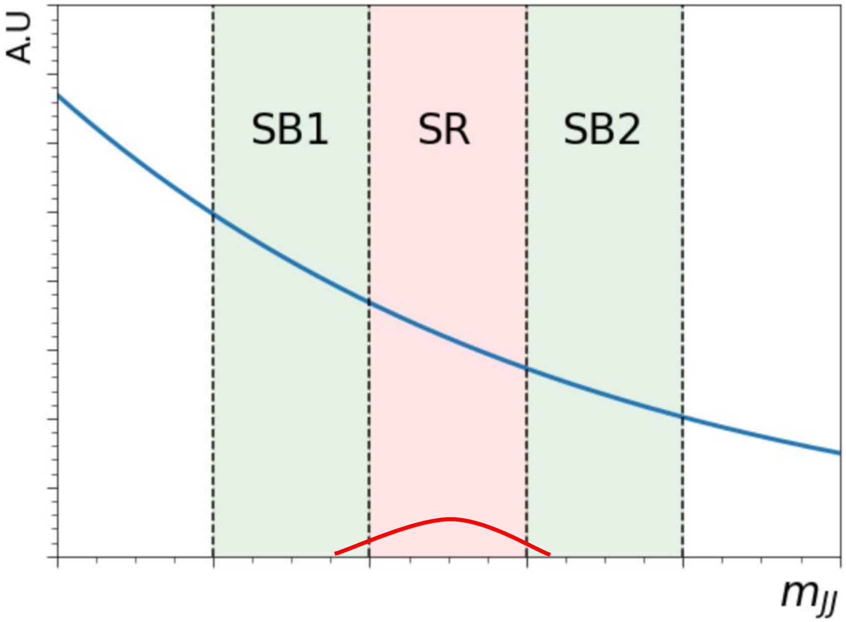
Assumption: Signal localised in some feature

(Usually, invariant mass)

⇒ Shows up as a **bump** in the spectrum

Method:

1. Split spectrum into '**Side Bands**'
2. Fit the distribution in Side Bands
3. **Interpolate** into the **Signal region**
4. Look for an excess!



So what's wrong?

New Physics is rare!

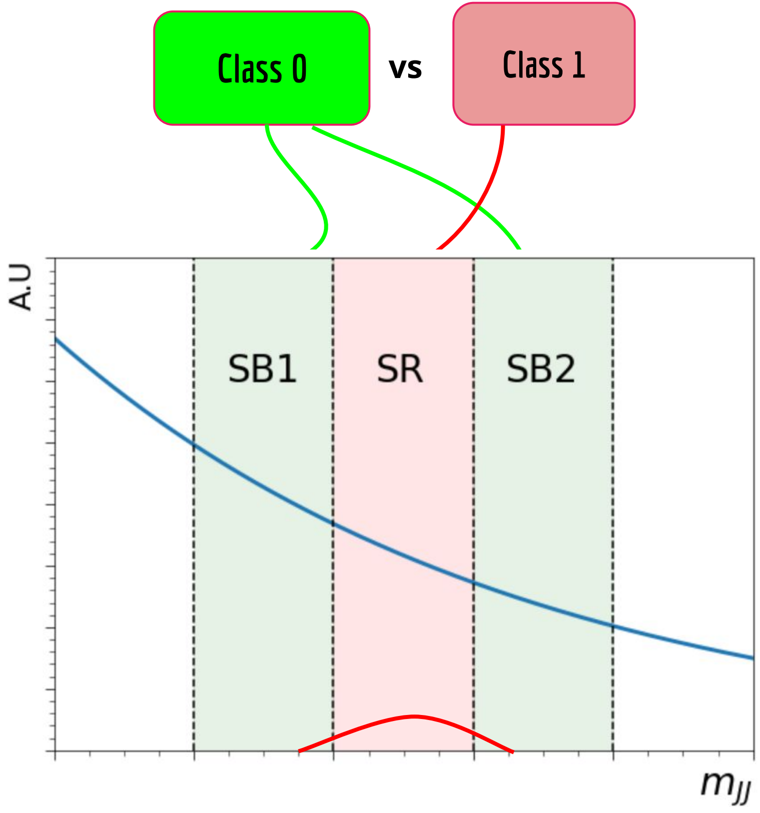
⇒ Increase sensitivity: use additional observables

CWoLa: Optimal Classifier for two different admixtures of classes ~ Optimal Classifier for fully supervised case.

Problem: Observables *strongly correlated to the mass.*

~ Differences even in the absences of signal!

What if we want to use these correlated variables?



Can we learn a conditional transformation of the features?

What would a sample from side-band look like if it had the mass of a sample in signal region?

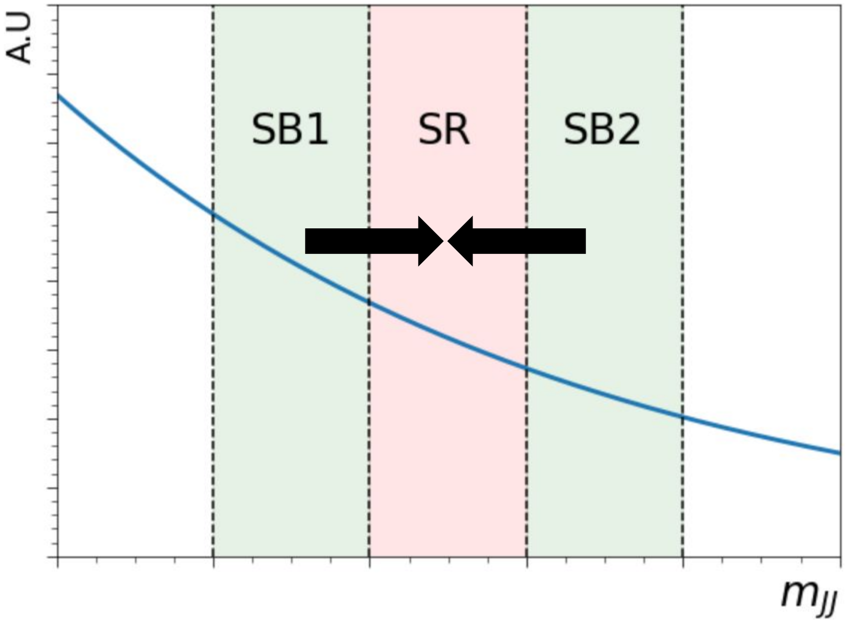
Idea + Motivation

Transform data from **the sidebands** into the **signal region**

⇒ **Small correction**

Transformed side bands provide a background template.

Train a classifier to separate background template from signal.

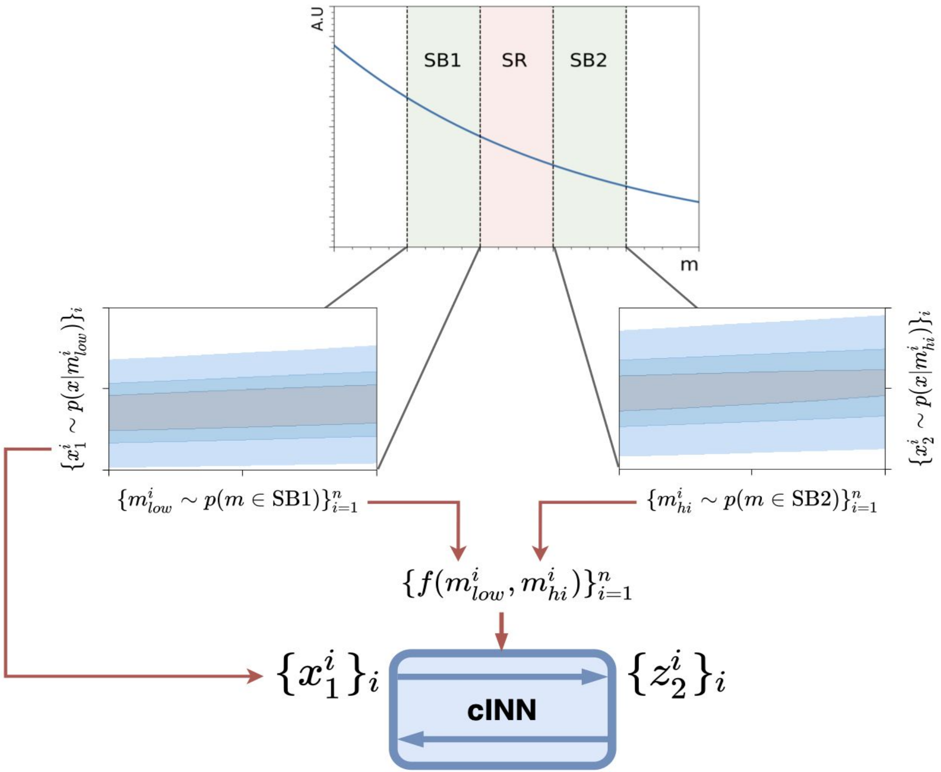


Learning to Transform

We can train the model to map from one sideband to the other

- Use a conditional invertible neural network (cINN).
- Many schemes available~, we use **rational quadratic splines**.

Train to transform batches of $\{x_1\}$ to $\{x_2\}$ → backprop → update weights.

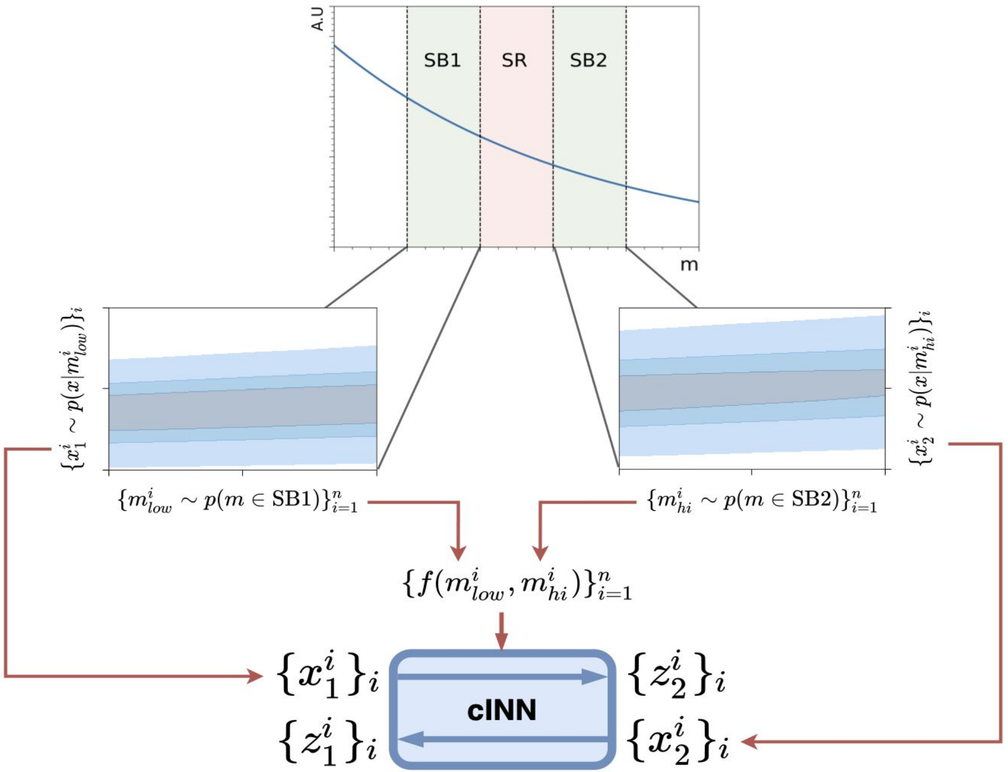


Learning to Transform

And in both directions!

Train to transform batches of $\{x_2\}$ to $\{x_1\}$ → backprop → update weights.

Alternate between these passes over an epoch.

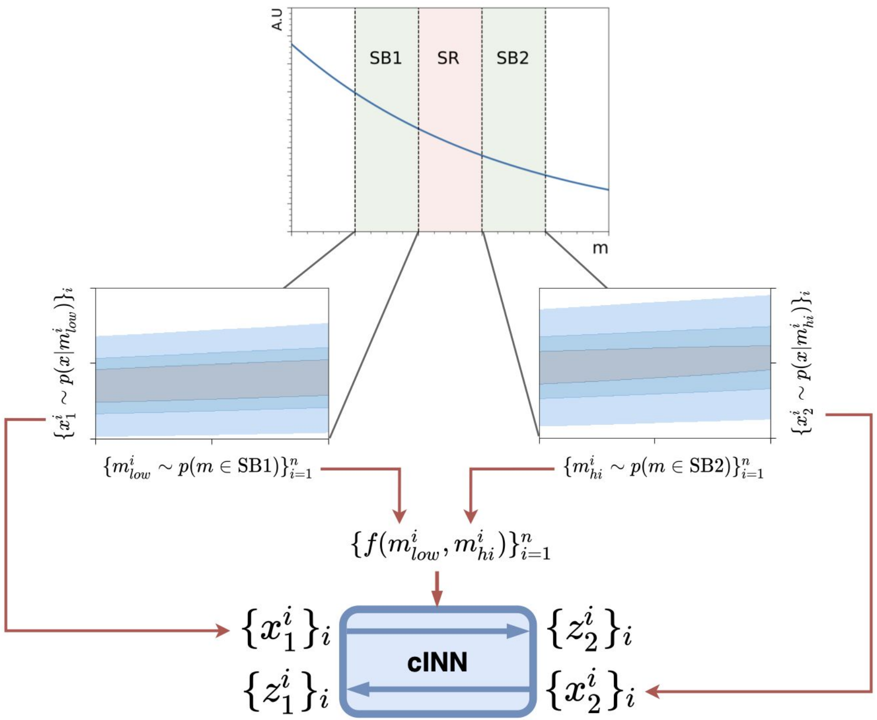


Application

cINN conditioned on the mass difference of input and target ~ Δm

To **create bg template in SR~**

1. Sample target mass from fit pdf; compute Δm
2. Perform forward (backward) pass through cINN to get SR template from SB1 (SB2).



But why the mass difference?

Each sideband can be split further.

Mapping from SB1’ to SB1” is the same as mapping from SB1” to SR

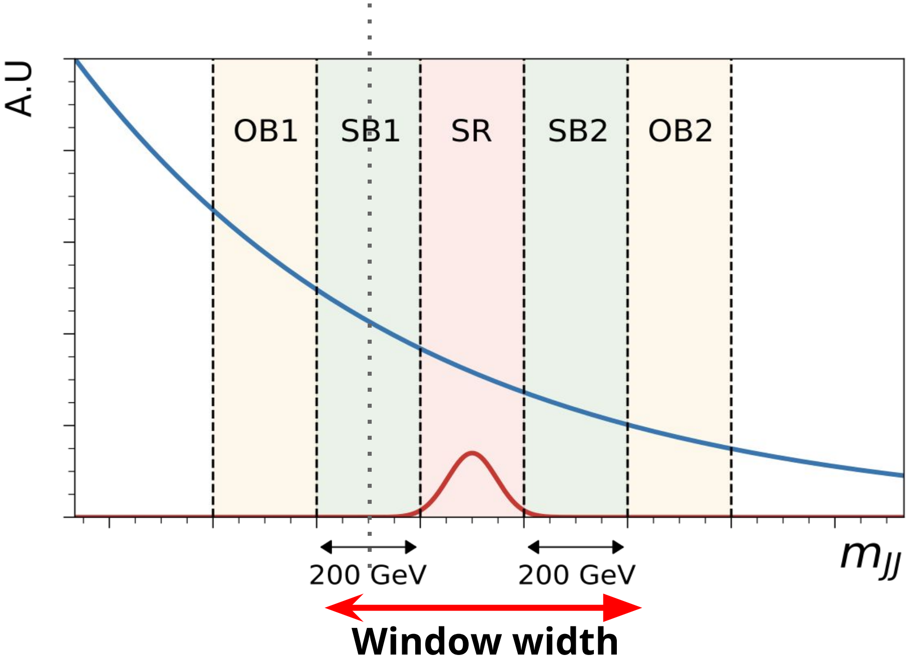
Conditioning variable: Δm

- (min ~0, max ~ window width)

∴ Conditioning for SB → SR:

- **Never out of distribution**!

Alternate between inter and intra-band training throughout training.



Distance Measures:

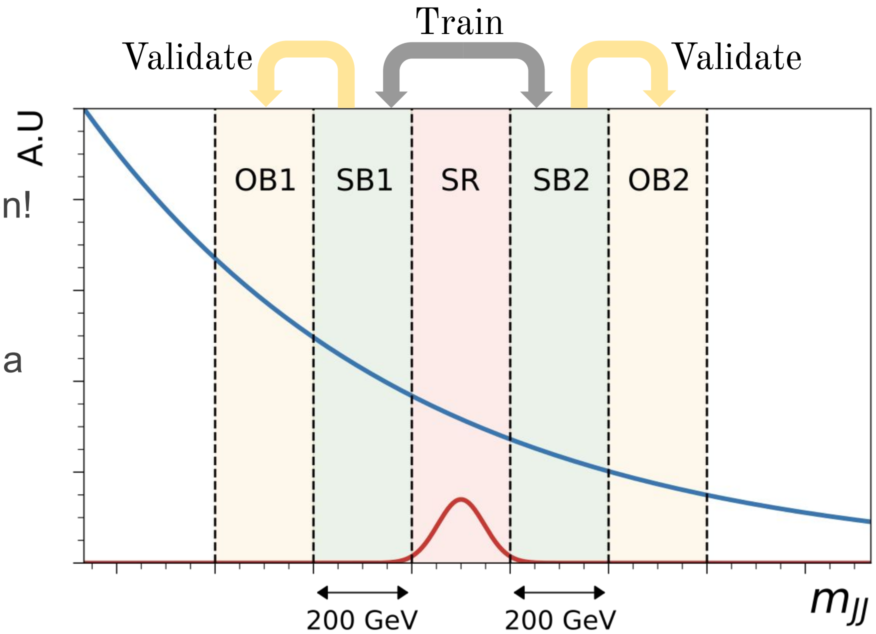
How to estimate transformations of **distributions** over features?

- We don’t have pairs, we want to shift one distribution to another.
- **Optimal transport loss:**
    - Sinkhorn loss between transformed events and target events.
- Trained over multiple epochs with shuffling of batches.

Curtains validation

Are our transformers any good?

- Training on narrow bands ~ Have OBs for validation!
- After training, Validate transformers on SB → OB.
- Invertible transformations, and choice of conditioning implies ~ SB→OB performance gives a reasonable bound on SB → SR performance.



Other methods

- CWoLa:
    - requires the additional features to be uncorrelated with the resonant variable.
- ANODE:
    - Learn conditional (mass) density of data and background in SR and do classification.
- CATHODE:
    - generate sample events from the trained, interpolated background density estimator → do classification.
    - has been shown to outperform the other approaches for building templates.

Dataset: LHCO R&D

QCD dijet and signal samples. ~ dijet mass as the resonant variable.

Features used for training:

- Leading jet mass
- Difference in jet masses
- Subjetiness ratios
- ΔR_{jj}*

*Strongly correlated with the dijet mass.

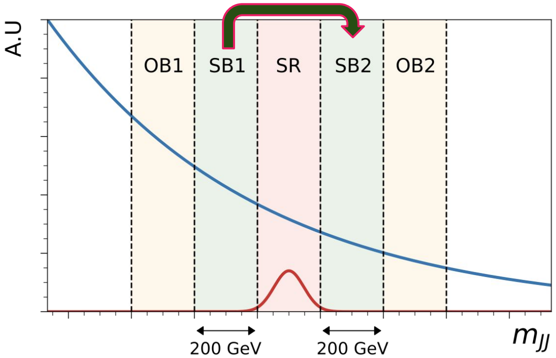
Transformation Performance

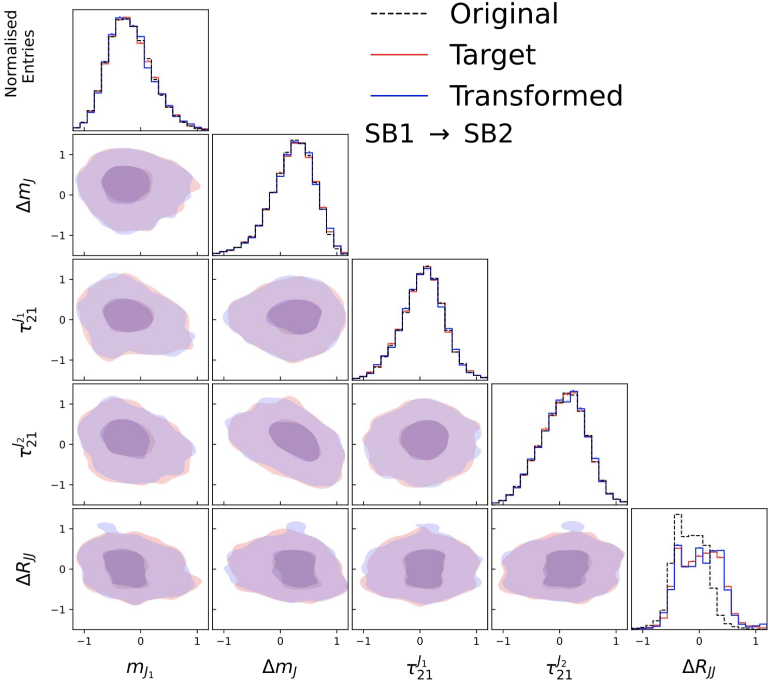
SB1: [3.2-3.4] TeV, SB2: [3.6, 3.8] TeV (no doping)

- Transformed features (with correlations) from SB1 → SB2.
  - Features in diagonal
  - Correlations in the off-diagonal.

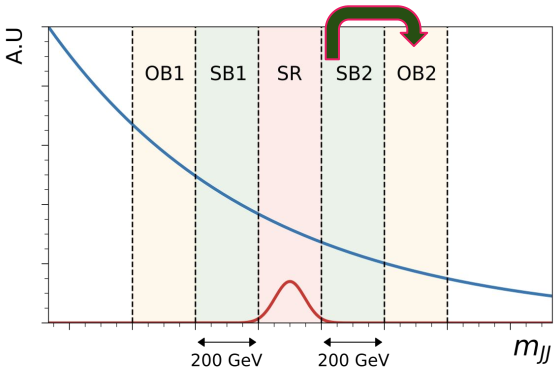
Near perfect overlap of transformed and target!

Just as good in the inverse direction!





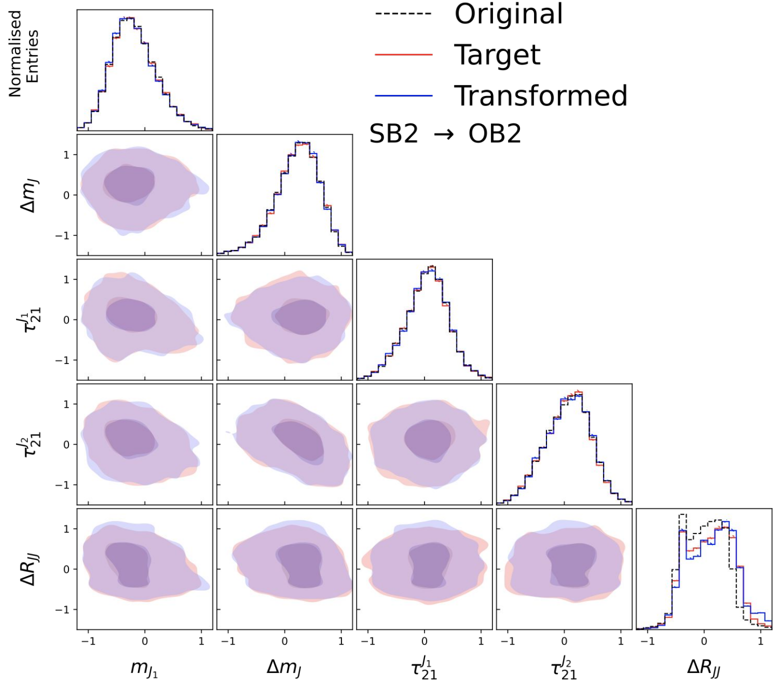
Validation regions



Narrow bands mean - we have the option to validate our transformations!

- SB2 [3.6, 3.8] TeV → OB2 [3.8, 4.0] TeV shown here.

Great overlap of Transformed and target distributions!



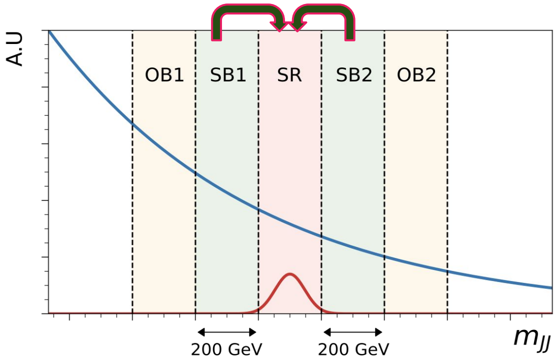
Signal region

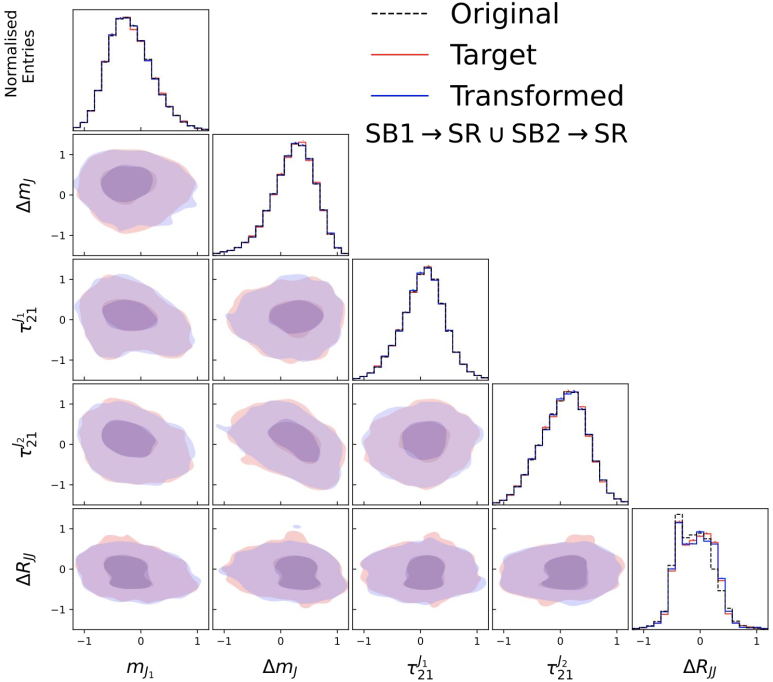
Generated bg template in SR ~ used in classifier training (and eventually bump hunt)

- Needs to accurate.

Can’t compare in real analysis

- Transformed SB1→ SR and SB2 → SR
- Near perfect closure on all features and correlations!
- Generated bg template is good!





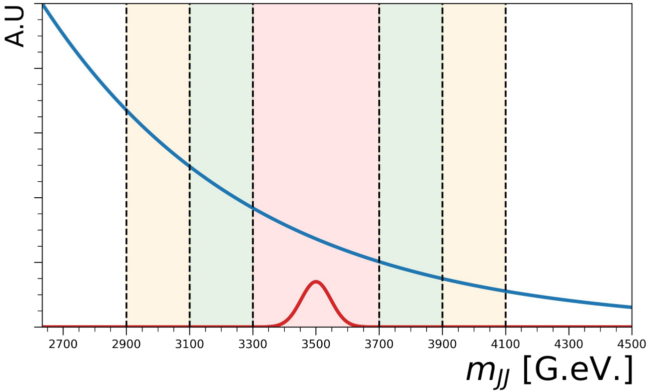
Expected Significance Improvement

Widen SR bin (200 → 400 GeV) to fully contain the peak:

- This is the same SR window as CATHODE.

Train CURTAINs in this window definition → Generate bg template in SR → Train a classifier on SR Data vs Template.

Measure Significance Improvement: $\varepsilon_s/\sqrt{\varepsilon_B}$



Classifier Setup

Architecture:

- Three hidden layers with 32 nodes
- Trained for 20 epochs
- Learning rate scheduler
- No early stopping
- Ten classifiers trained per run

Features:

- All features used during transformer training!

Comparisons

ROCs and Significance Improvement Characteristic (SIC) curves:

<u>Supervised</u>: QCD vs signal with labels.

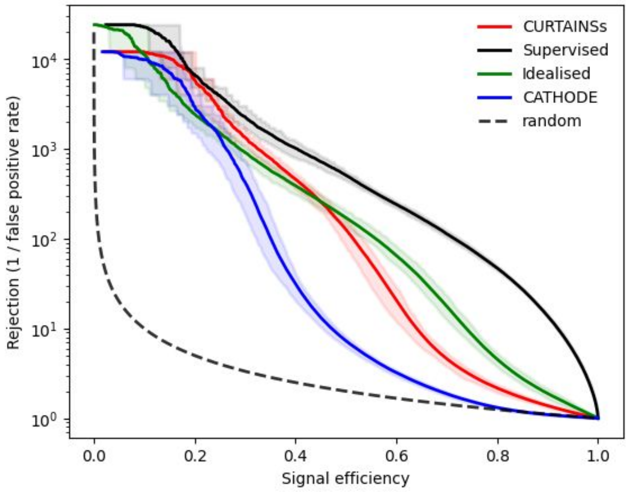
- *Given the data, what's the best classification performance*?
- *Fully supervised - 120k QCD vs 3k Signal (weighted loss).*

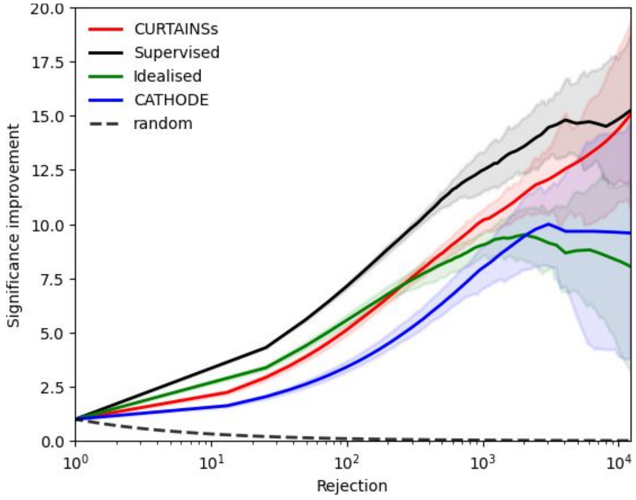
<u>Idealised</u>: QCD vs QCD+Signal.

- *Given the data, what's the best classification possible in the CWoLa setting?*
- *50% of SR data as class 0, the rest doped with signal is class 1.*

<u>CATHODE*/CURTAINS</u>: Generated Bg template vs SR data.

CURTAINs ~ Almost as good as Idealised performance!

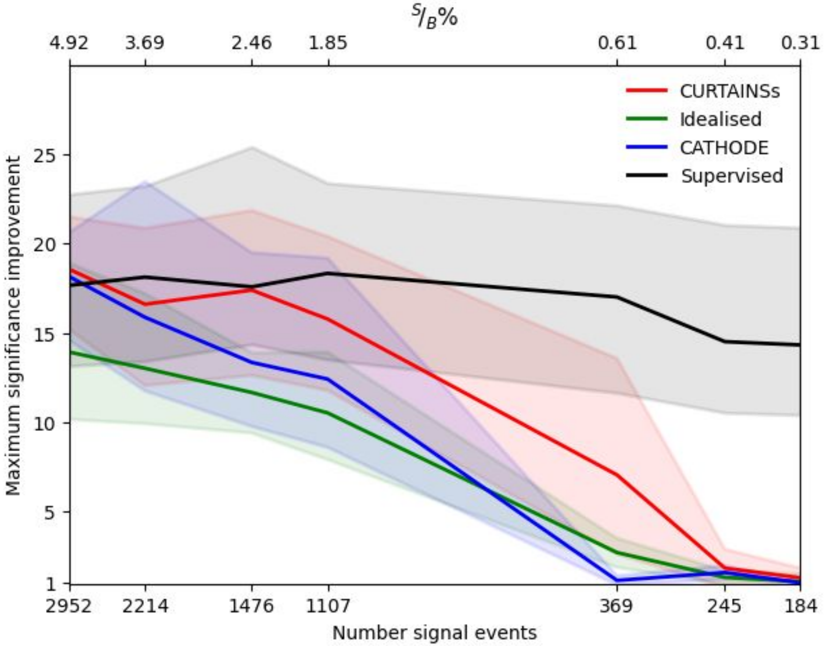




Maximum Significance Improvement

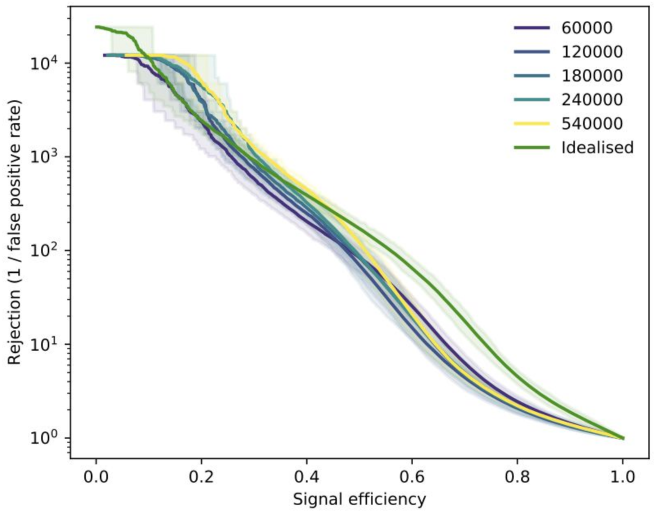
…as a function of signal strength:

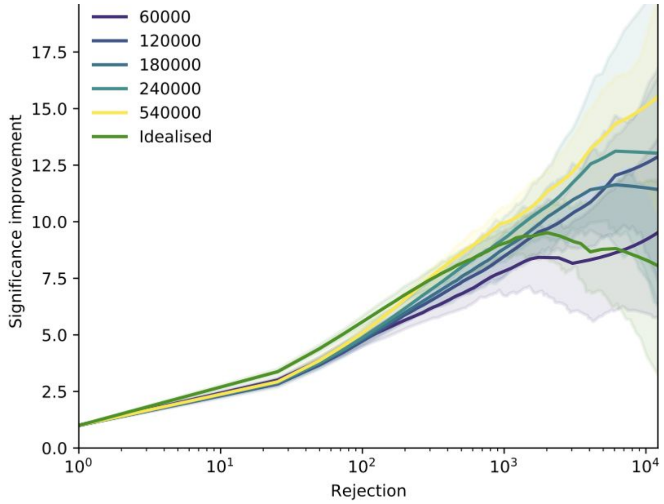
- CURTAINS performance excellent across all signal strengths!
- Generally better than ‘Idealised’ since it can *oversample*:
    - i.e. can generate loads of stats for bg template!



Curtains and oversampling

- Transform data in SB to multiple targets in SR
  ⇒ Oversampling.
- Check performance of CURTAINs as a function of oversampling.

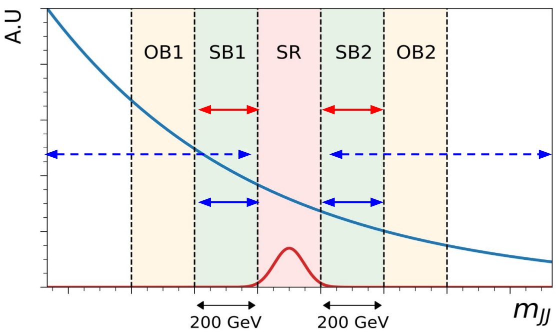




CURTAINS already matches Idealised at ~2X oversampling!

- Note: SR here contains 120k events.

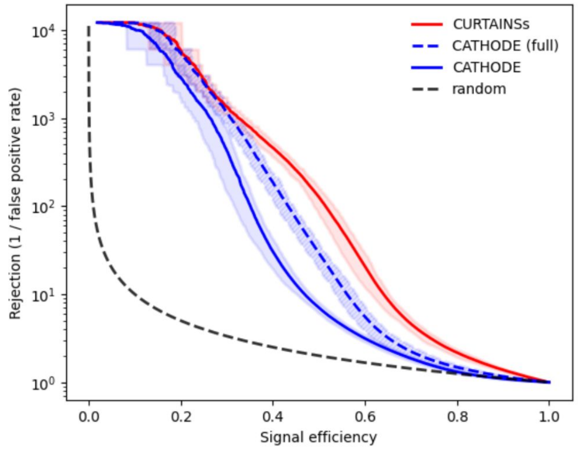
CATHODE full width sidebands

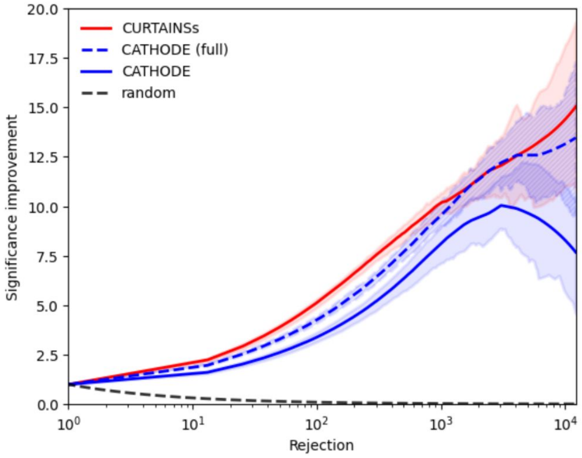


CATHODE designed to use everything except SR as side bands → CATHODE(full)

- CATHODE (full) recovers performance with more data.
- In this setting, we lose the possibility to do validation!

CURTAINs designed to work locally → Fewer data required → Retains regions of validation.

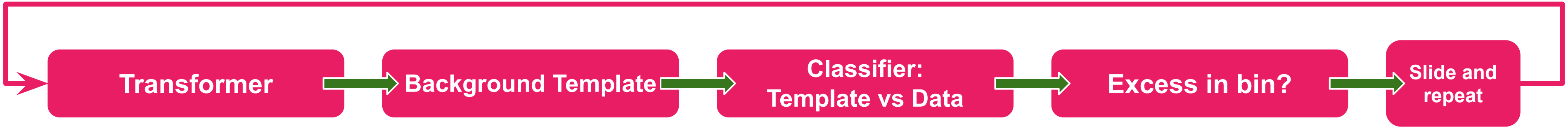




CURTAINS for a bump hunt

No one knows where the signal is!

- Need to slide windows



- Learn a transformer to go from SB to SB.

- Validate on SB → OB

- Generate template by transforming SB → SR

- Train a classifier on generated bg template and true SR data.

- Apply cuts to reject x% of generated background template.

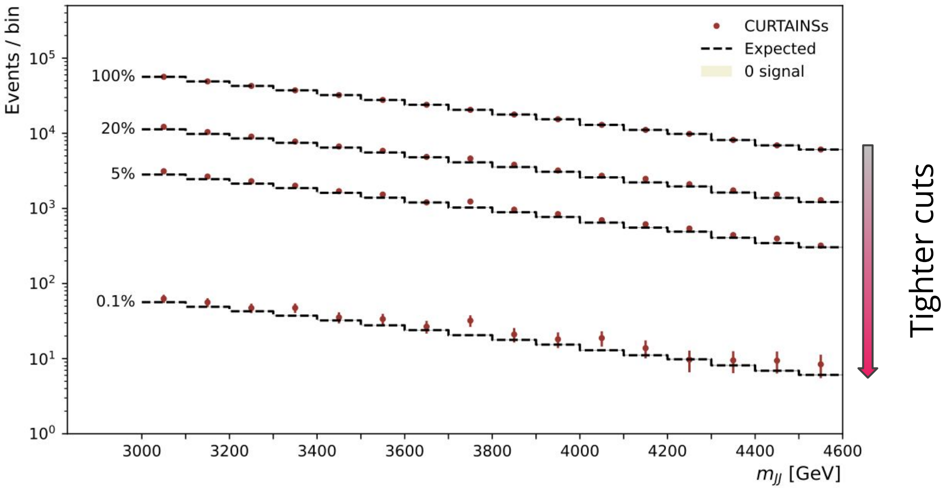
- If there is signal in data - cut would reveal excesses over expected data.

- Generated background template needs to be good!

Bump hunt

CURTAINs applied to a dataset with **no signal**.

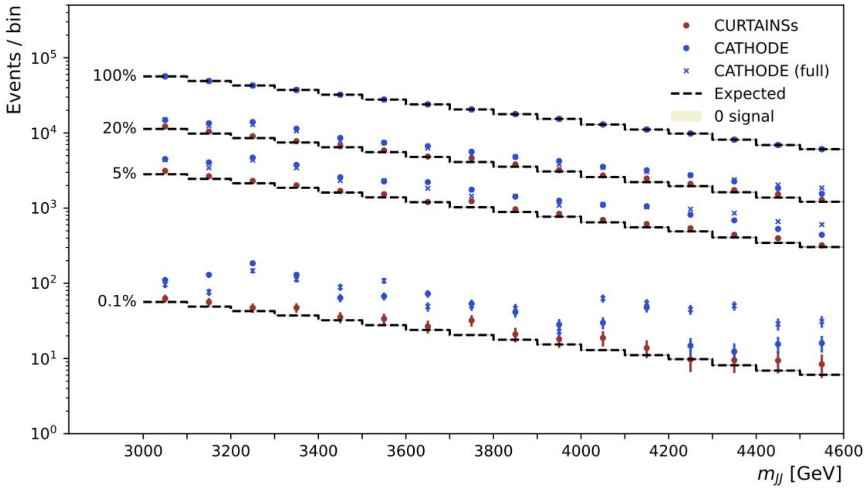
- New CURTAINs model trained every 200 GeV window
- Post classifier training, determine cuts to reject x% bg template
- Cut on true SR Data and see if there's any excess



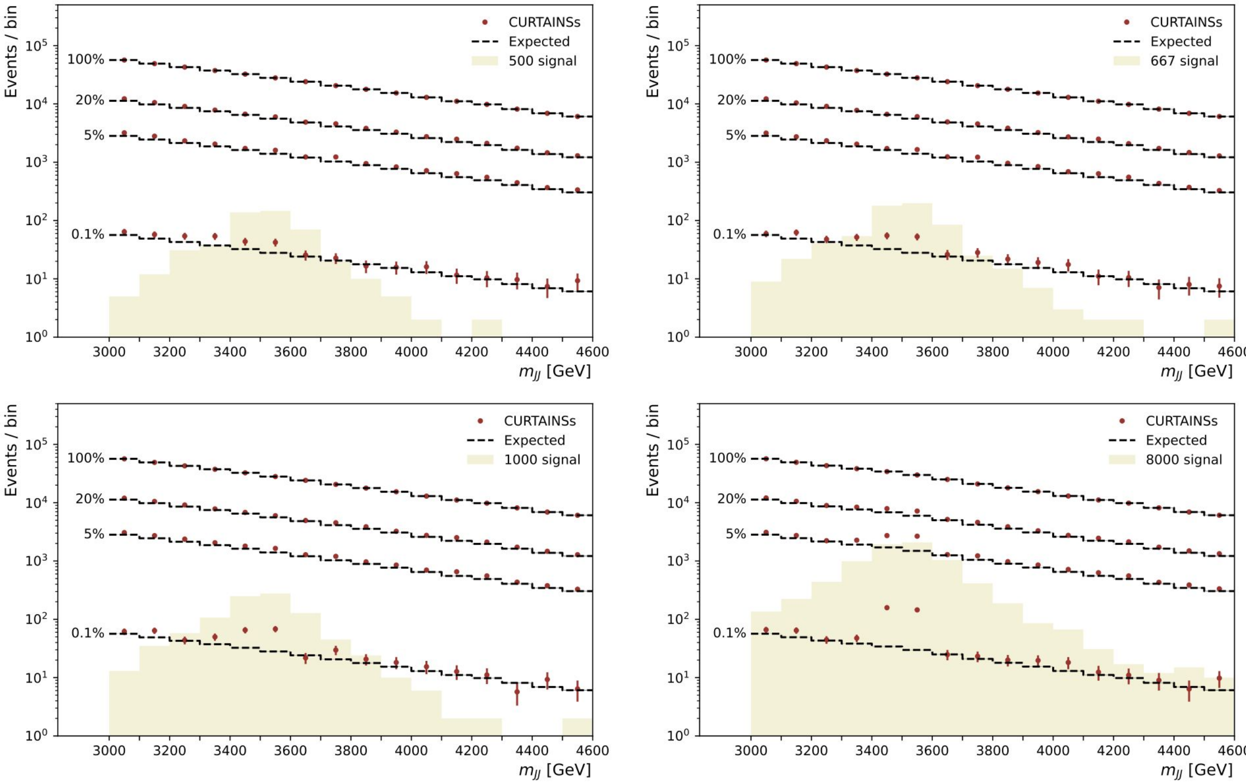
When no signal is present, no bumps appear!

Bump hunt vs CATHODE

CATHODE (and CATHODE (full)) present spurious bumps in the absence of any signal.



Bump hunt - Injecting signal



- 60,000 bg events in the bin [3400, 3600] GeV.
- Inject varying number of signals
  - More signal ⇒ Higher bump observed.
- But even as low as 667 injected signals (~300 in [3.4 - 3.6] TeV) **CURTAINs finds a bump**!

Summary

CURTAINs provides a robust method for constructing background templates that

1. produces no spurious excesses when used in a bump hunt
2. finds bumps even at small signal injection
3. Can be trained locally and on data only.

CURTAINs does not need to

4. sample from any prior distribution
5. be applied to ranges outside of the training distribution.

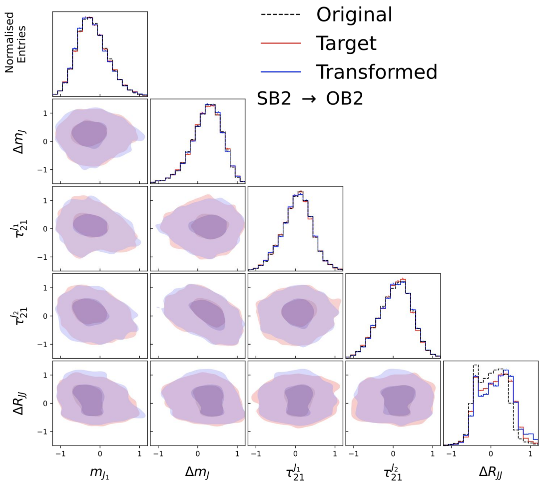
Preprint available on the arxiv.

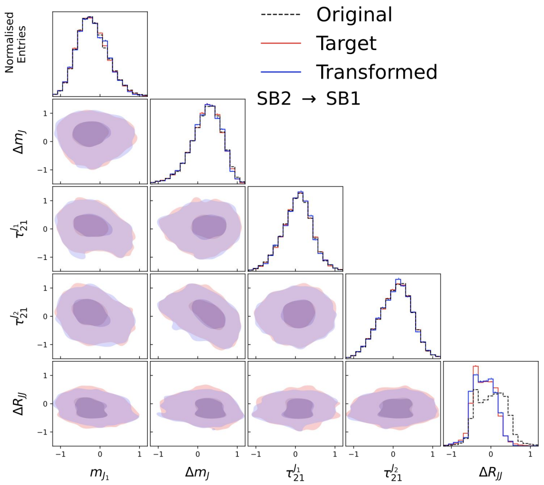
Backup

Transformations

CURTAINs transformations for SB2 → OB2 and SB2 → SB1.
For both regions we achieve excellent overlap.

Note: SB2 → OB2 is (also) used as validation for the transformer.
- Expected to be a harder transformation as this is out of distribution compared to training.



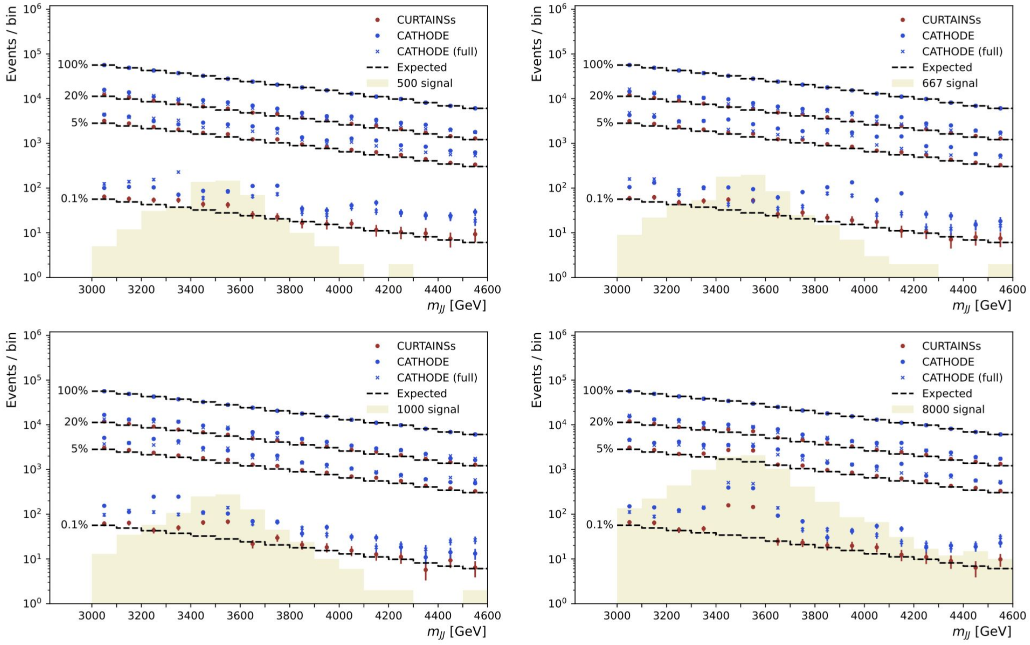


Cathode comparisons

Bump hunt with various level of signal injections.

CURTAINs correctly produces a bump at the signal peak and gets better with higher signal strengths.

CATHODE (and CATHODE full) predicts excesses across all bins, and erroneous excesses at higher mass values.



Comparisons

The signal region contains 120,000 background events in total.

Model	Bkg Events	Signal Events
Supervised	120,000	3,000
Idealised	60,000	60,000 + 3,000
Curtains	540,000	60,000 + 3,000
Cathode	540,000	60,000 + 3,000

CURTAINs can do oversampling ~ Transform data from SB multiple times to different target mass values.

CURTAINS v CATHODE v CATHODE (Full)

CATHODE with extended sidebands recovers performance, and approaches the performance of CURTAINs at high rejection values.