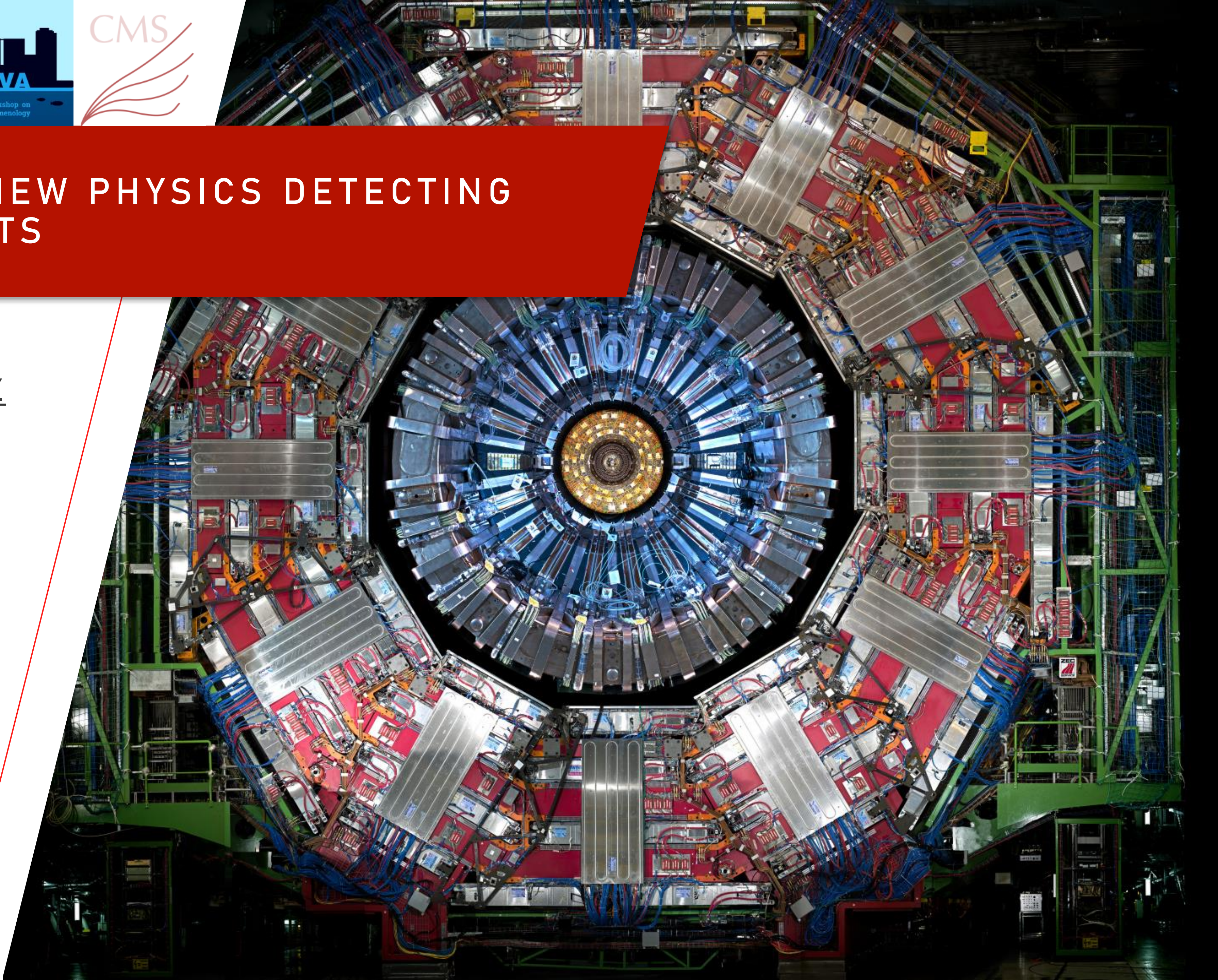
SEARCHING FOR NEW PHYSICS DETECTING ANOMALIES IN JETS

Roberto Seidita - ETHZ

On behalf of the CMS Collaboration



CASTING A WIDE NET

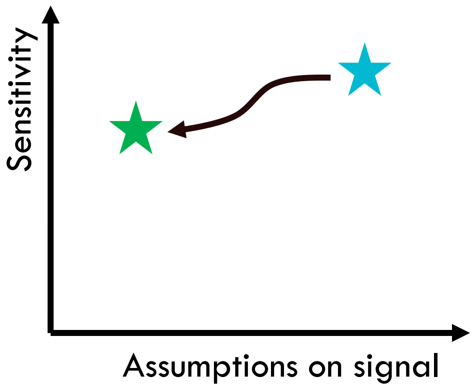
The space of plausible physics beyond the SM is vast

Peak performance via supervised ML-based searches:

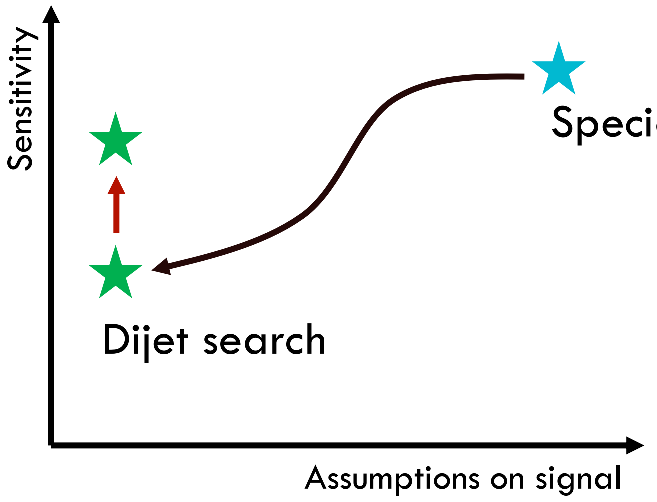
- Optimal sensitivity for specific target signals, but
- Sub-optimal to no sensitivity to signals not yet thought of

Unsupervised methods offer a natural complement:

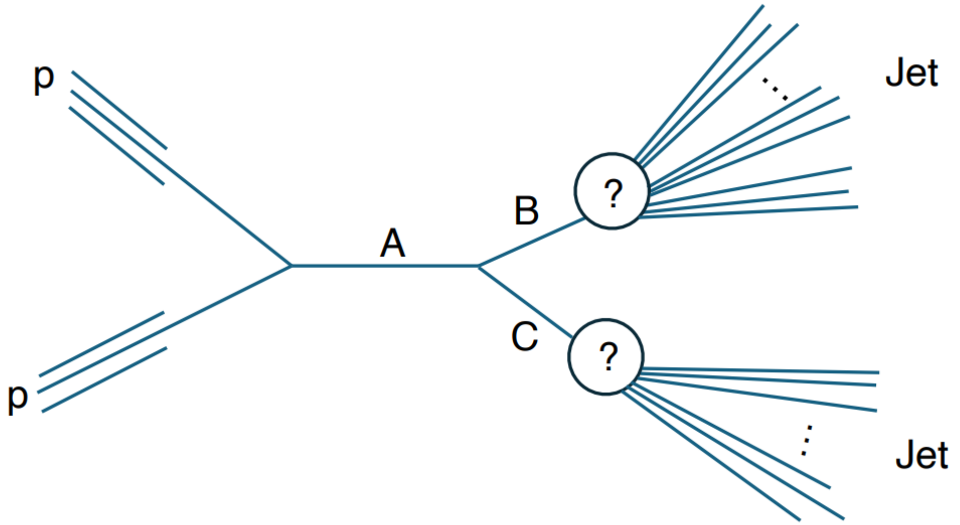
- Less sensitive then targeted searches, but
- Can reach a wider chunk of model space



ENHANCING THE DIJET SEARCH



Not a new concept: e.g., (pair) dijet searches are always exploring the energy frontier in the most model agnostic way possible

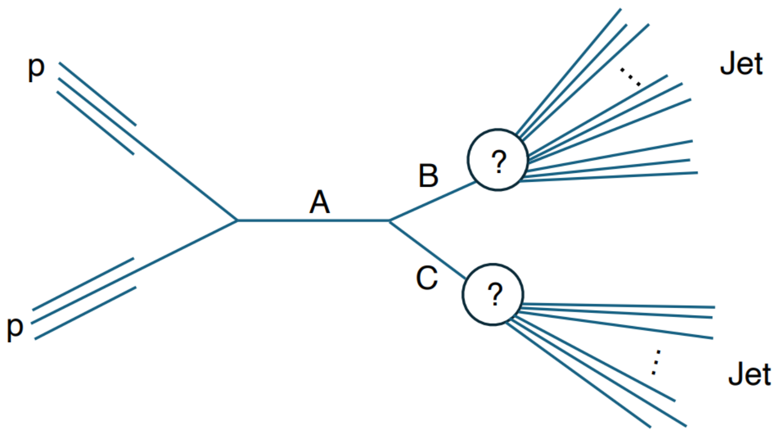


[CMS-PAS-EXO-22-026]()

Dijet searches only exploit event kinematics (on purpose) → can we re-introduce jet substructure into the picture without spoiling the generality of the search?

DEFINITIONS AND SELECTIONS

Jet $\equiv$ anti-k_T jets with $R = 0.8$ (AK8), clustered from particle-flow (PF) candidates with pileup-per-particle-identification (PUPPI) weights



Signal region:

- At least 2 jets with $p_T > 300$ GeV and $\eta < 2.5$
- Invariant mass of leading pair $m_{jj} > 1455$ GeV
- Pseudorapidity gap of leading pair $|\Delta\eta_{jj}| < 1.3$

Leading pair: B and C candidates

Ensures efficient triggering

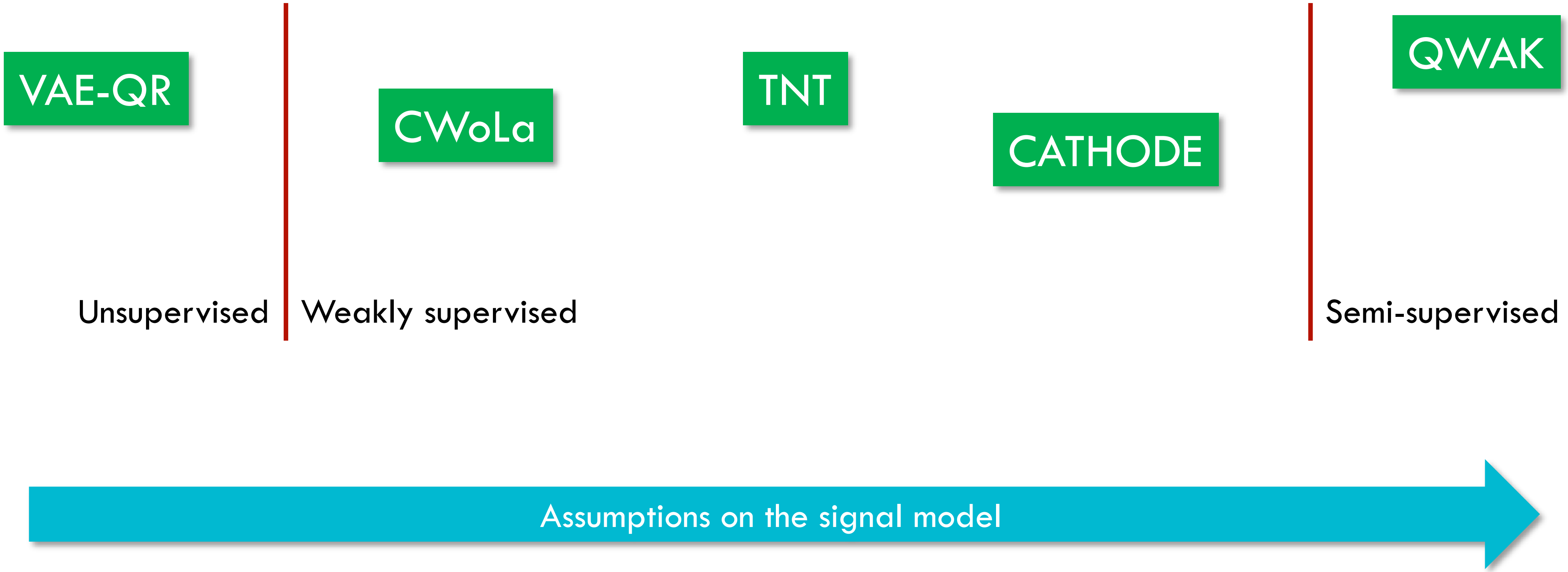
Suppresses t-channel QCD

Control region:

- $2.0 < |\Delta\eta_{jj}| < 2.5$

Techniques

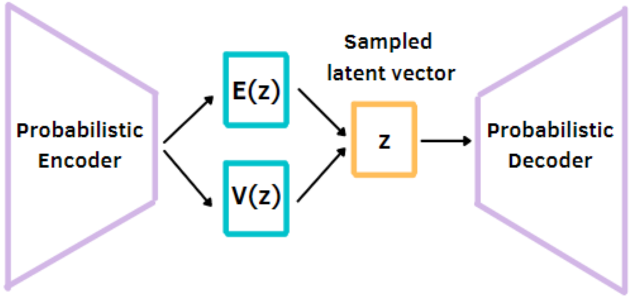
THE TECHNIQUES



VAE-QR

arXiv:1312.6114

$$\begin{pmatrix} p_x^1 & p_y^1 & p_z^1 \\ \vdots & \ddots & \vdots \\ p_x^{100} & p_y^{100} & p_z^{100} \end{pmatrix} \equiv X \longrightarrow$$



$$\longrightarrow \hat{X}$$

- Up to 100 constituents per jet, padded
- Ordered via CA reclustering
- Trained on data in high $|\Delta\eta|$ control region to minimize the reconstruction error
- Quantile regression model learns threshold on anomaly score as a function of m_{jj}

$\min\left(\mathcal{L}_{j_1}(X,\hat{X}), \mathcal{L}_{j_2}(X,\hat{X})\right)$

m_{jj}

Quantile regression model $\longrightarrow \mathcal{T}(m_{jj})\Big|_{\epsilon=1-Q}$

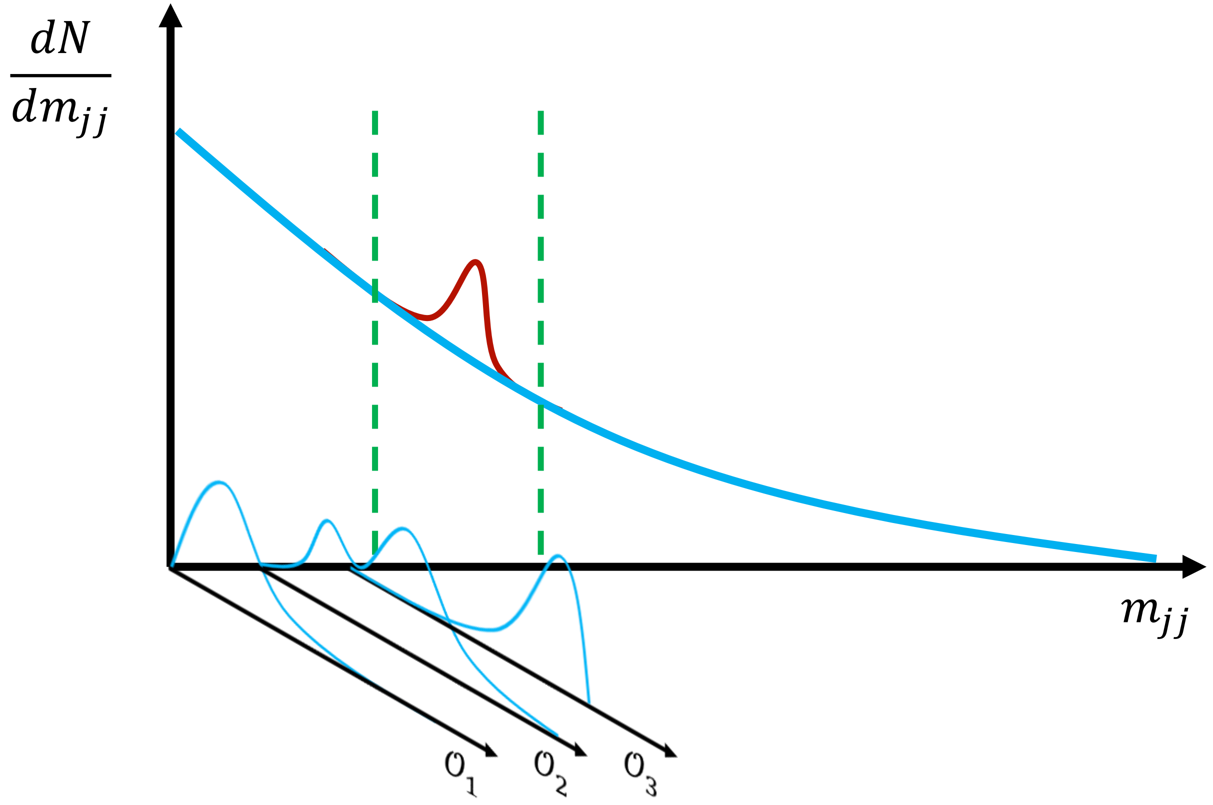
Events in which both jets pass $\mathcal{T}(m_{jj})$ are selected for the fit of the m_{jj} spectrum

See Florian's talk for a deeper discussion on such a stategy

CLASSIFICATION WITHOUT LABELS (CWOLA)

10.1007/JHEP10(2017)174

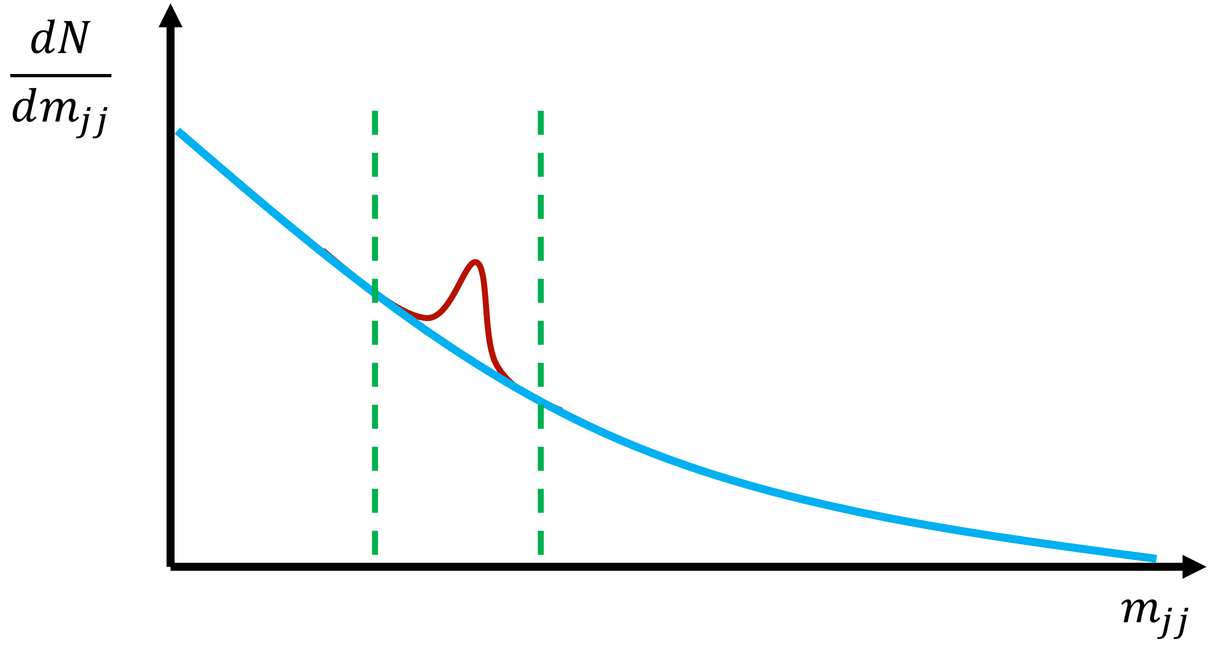
New assumption: signal is a narrow peak in m_{jj}

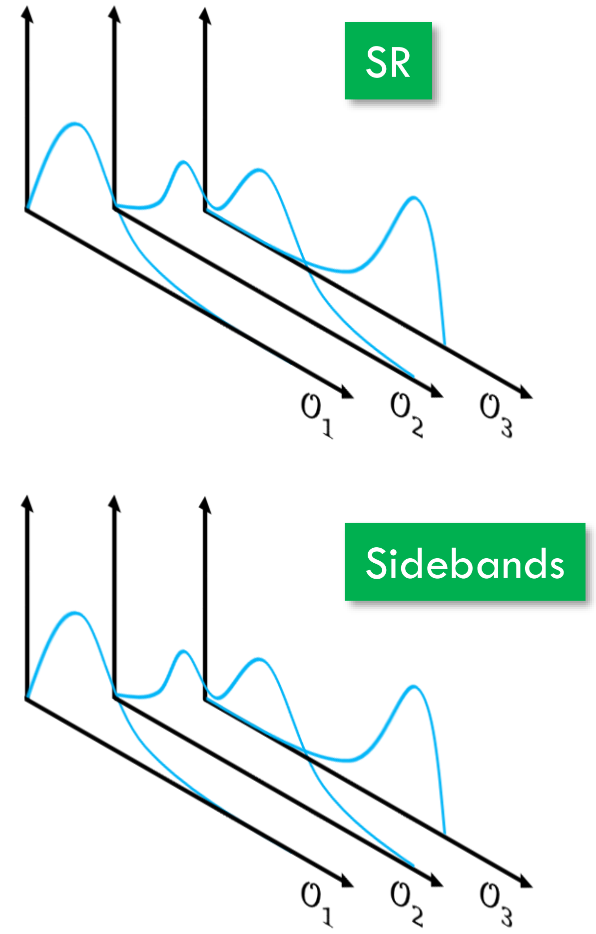


CLASSIFICATION WITHOUT LABELS (CWOLA)

10.1007/JHEP10(2017)174

New assumption: signal is a narrow peak in m_{jj}





If there is a signal peaked in m_{jj} with different JSS, a classifier between the SR and the sidebands will pick it up

CLASSIFICATION WITHOUT LABELS (CWOLA)

Two classifiers are trained for the leading and sub-leading jet in the event → allows for flattening in p_T during training by reweighting

Each classifier takes as input 7 JSS observables:

- Soft-drop mass m_{SD}
- N-subjettiness ratios τ_{21}, τ_{32}, and τ_{43}
- Number of PF candidates in the jet n_{PF}
- Lepton subject fraction with 3 subjets LSF_3
- Max deepCSV b-tagging score of the two leading subjets

The classifier score is then used to build a (potentially) signal-enriched region and a fit on m_{jj} is done

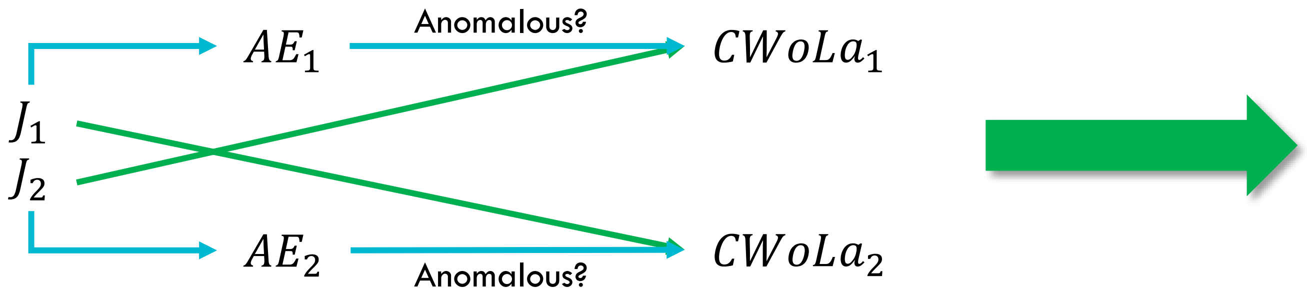
TAG 'N TRAIN (TNT)

10.1007/JHEP01(2021)153

The performance of a CWoLa classifier strongly depends on the purity of the SR (assuming the signal exists)

We can enhance the purity via jet-based unsupervised taggers e.g., autoencoders trained in m_{jj} sidebands

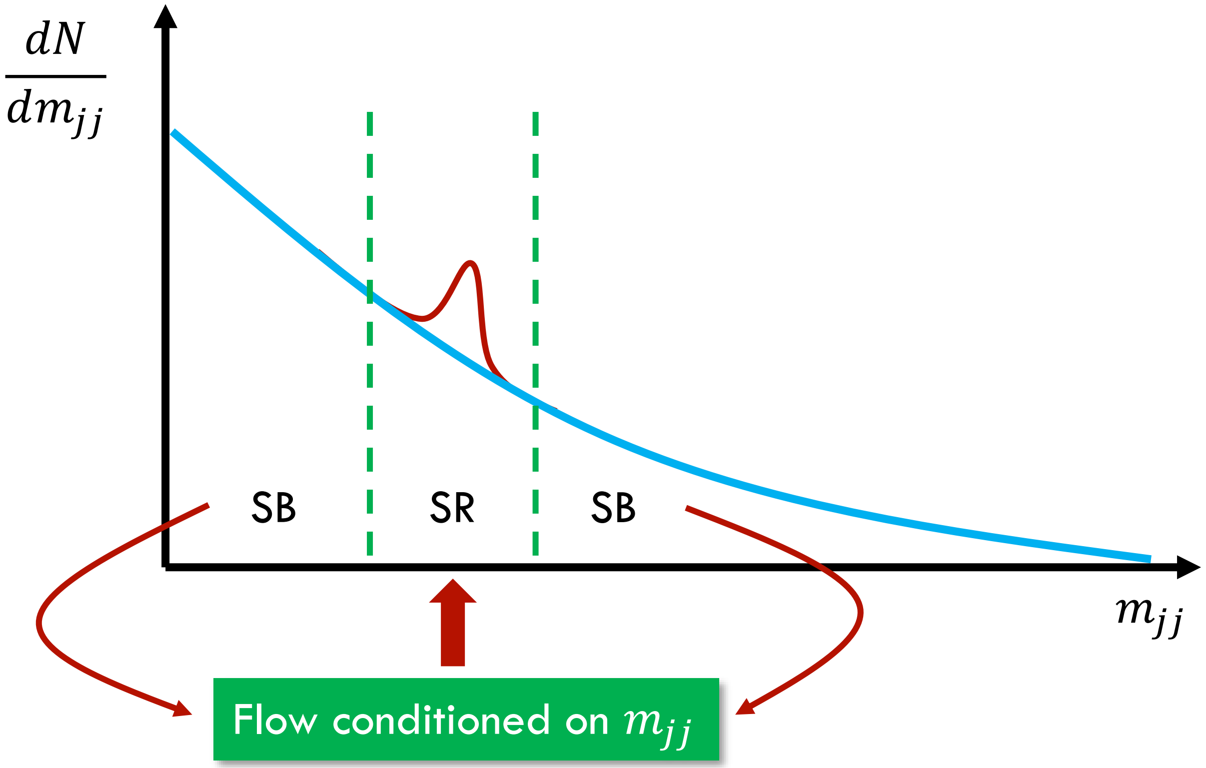
Key idea: for background, whether a jet passes the anomaly threshold is entirely random; for signal, if one jet passes the threshold, it should be more likely for the other one to pass it too



Enhances the performance of CWoLa without additional assumptions

CATHODE

10.1103/PhysRevD.106.055006



- Train a normalizing flow on the sidebands, conditioned on m_{jj}
- Use the flow to generate a sample in the SR (a fancy interpolation if you will)
- Inputs to flow: m_{J_1}, $\Delta m_{J_1,J_2}$, $\tau_{41}^{J_{1(2)}}$
- Separate model with addition of deepCSV scores (CATHODE-b)
- Train the CWoLa classifier not on SR vs SB, but on generated vs real events in the SR

QUASI-ANOMALOUS KNOWLEDGE (QUAK)

10.1007/JHEP06(2021)030

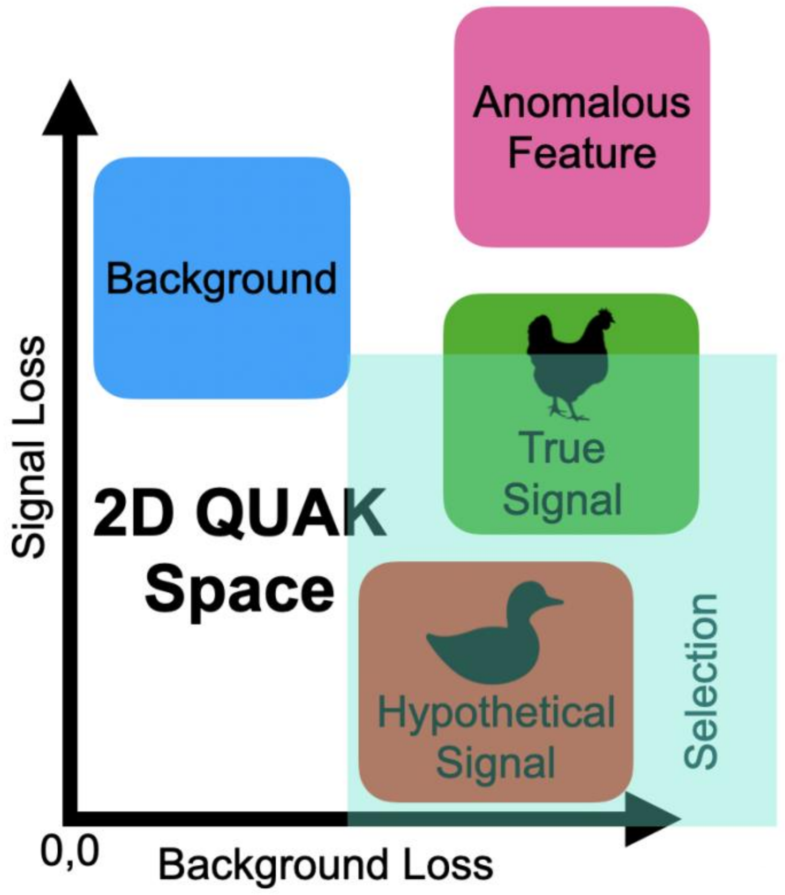
Same core concept of unsupervised learning: train a model on the background, get some form of anomaly score

One key addition: unsupervised model(s) trained on a (set of) signal hypotheses (MC)

Each score is a dimension of the QUAK space

In principle:

- True signal is among the set of hypotheses → approach sensitivity of supervised searches
- True signal is completely different from all signal hypotheses → same sensitivity as fully unsupervised search
- Something in between → something in between



QUASI-ANOMALOUS KNOWLEDGE (QUAK)

10.1007/JHEP06(2021)030

Flow

Experimental data $\rightleftarrows$ $\mathcal{N}(0, I_N)$

Used in the normalizing direction, a flow becomes a density estimator

Can be done with autoencoders; here normalizing flows are used instead

One flow is trained on each of a set of 6 signal hypotheses with varying B and C particle masses

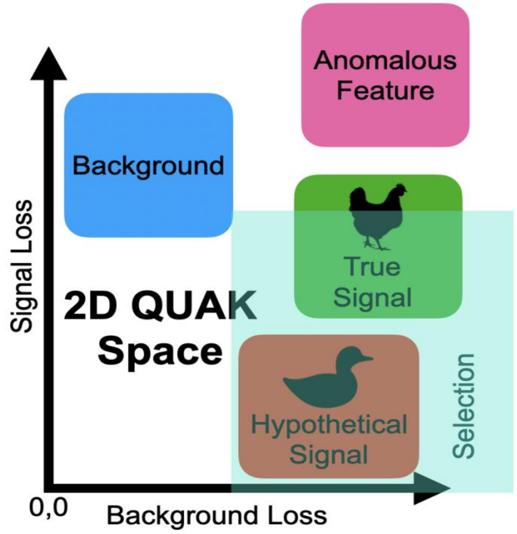
One extra flow trained on the background (MC)

Same inputs as CWoLa, with the following exceptions:

- LSF $\rightarrow \tau_S = \sqrt{\tau_{21}}/\tau_1$
- $m_{SD} \rightarrow \rho = M/p_T$

Signal-like score are combined into a signed L5 norm:

$$L5 = \mathrm{sgn}(\Sigma_5)\,|\Sigma_5|^{\frac{1}{5}}, \qquad \Sigma_5 = \sum_{i=1}^{k} \mathrm{sgn}(\ell_i)|\ell_i|^5$$



Signal extraction

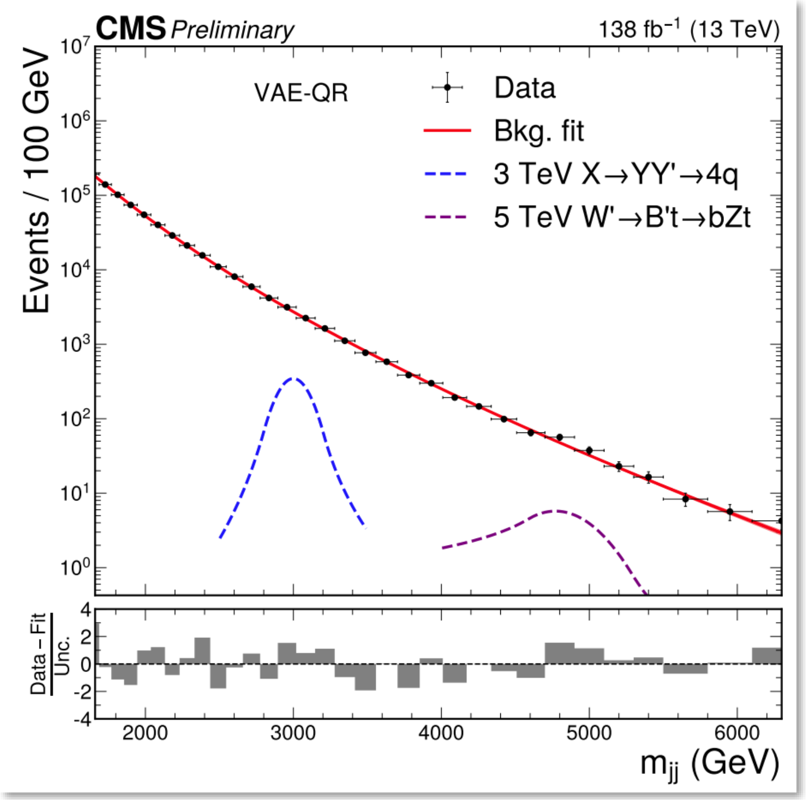
FIT PROCEDURE

Once potentially interesting events are selected by each of the AD techniques, we want to set limits!

Parametric fit to the m_{jj} shape:

- Signal: double-sided crystal ball fit to each signal hypothesis
- Background: falling spectrum modelled as $\frac{dN}{dm_{jj}} = \frac{P_0(1-x)^{P_1}}{(x)^{P_2+P_3\log(x)+P_4\log^2(x)}}$, where $x = m_{jj}/\sqrt{s}$

Each AD method gives a different spectrum → F-test for choice of fit function parameters repeated each time



Performance

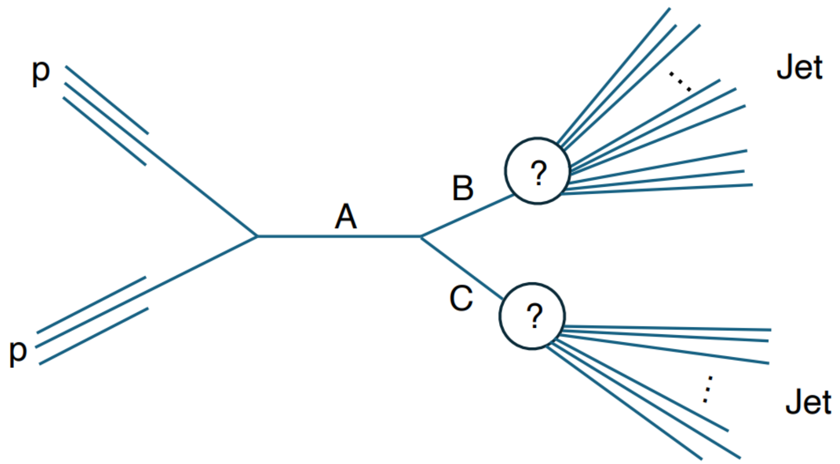
BENCHMARK SIGNAL MODELS

Six signal models are considered as benchmarks for sensitivity:

- 2+1 prong: $Q^* \to qW'(\to qq)$
- 2+2 prong: $X \to \mathrm{Y}(\to qq)\mathrm{Y}'(\to qq)$
- 3+3 prong: $W' \to tB'(\to bZ)$
- 2+4 prong: $W_{KK} \to RW(\to WW)W$
- 5+5 prong: $Z' \to T'(\to tZ)T'(\to tZ)$
- 6+6 prong: $Y \to HH \to 4t$

For each model, a grid with different masses is considered:

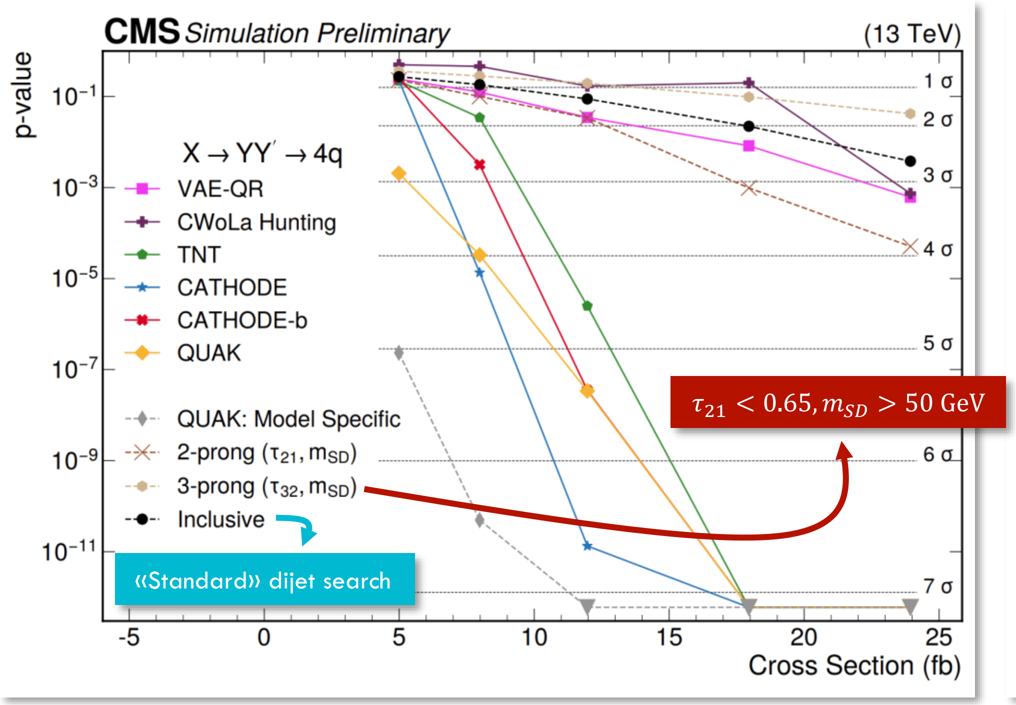
- $M_A = 3, 5$ TeV
- $M_{B,C} = 25, 80, 170, 400$ GeV

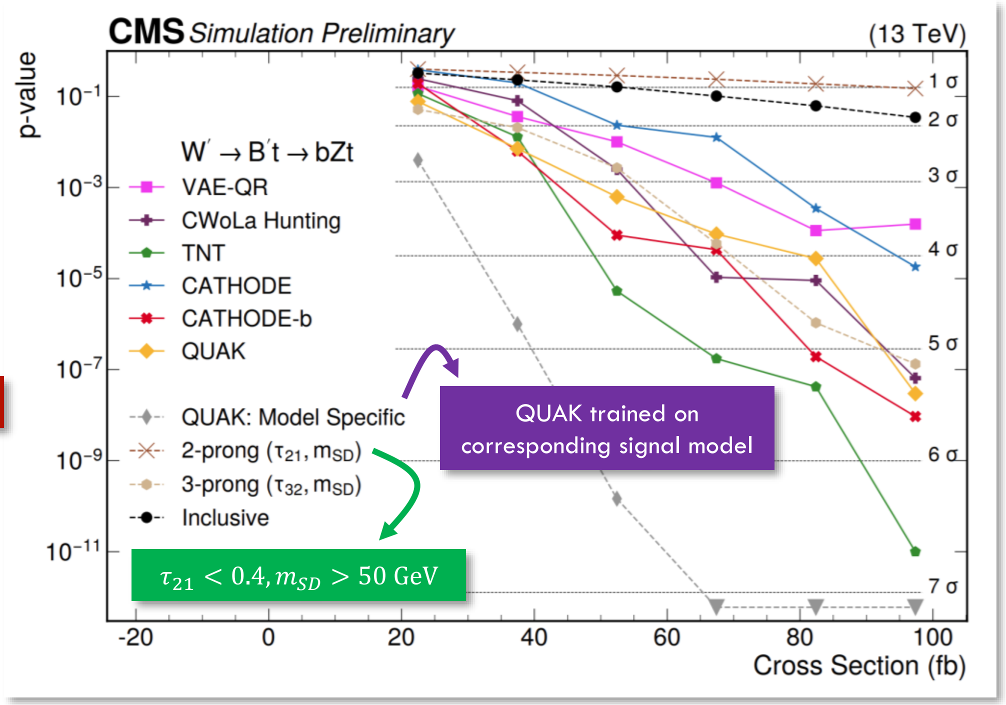


EVALUATING AD STRATEGIES

A mock dataset is produced with a realistic mixture of background processes (mainly QCD) and injections of signals with varying cross sections

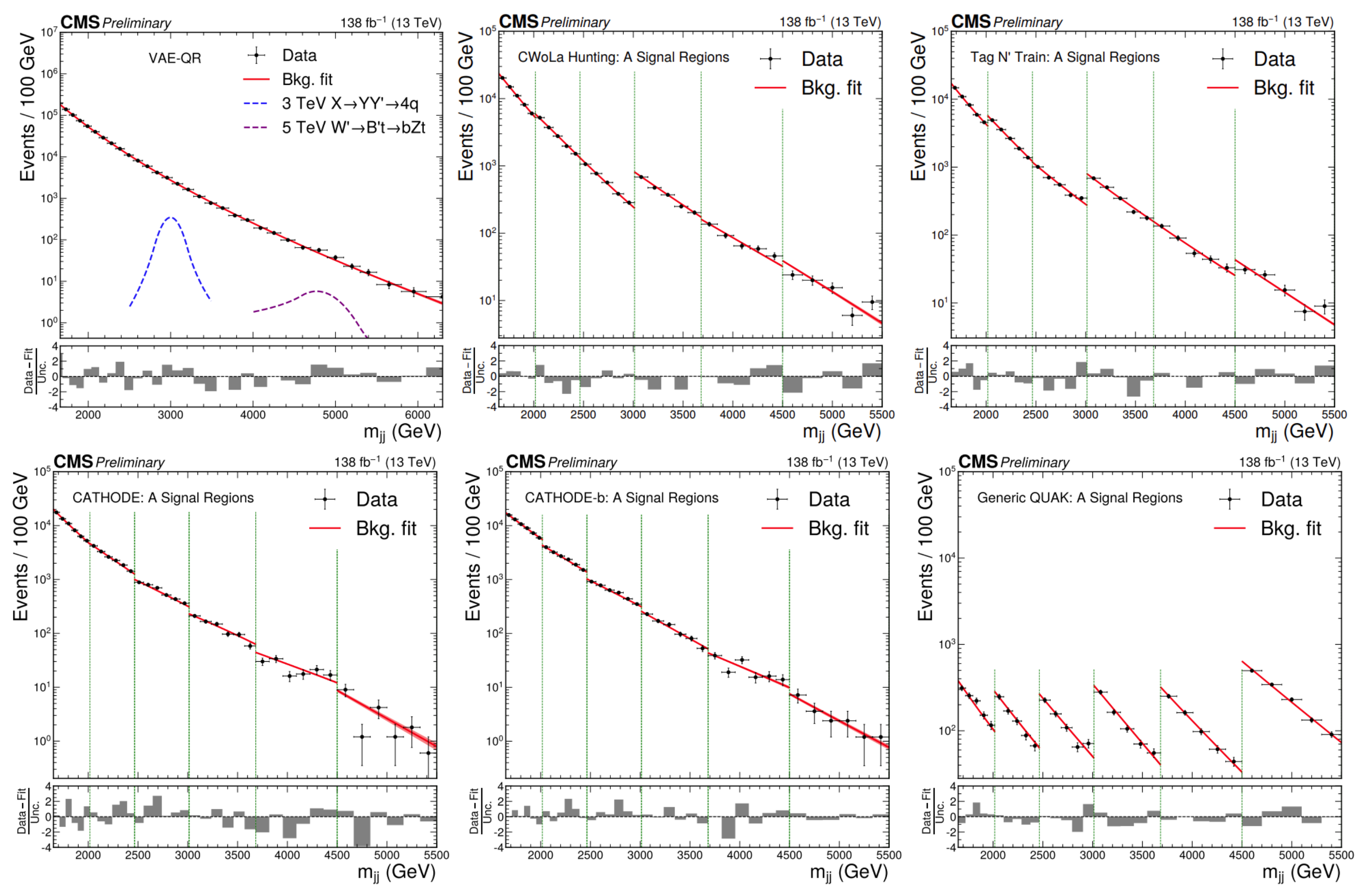
A background only sample is also used to check that none of the methods produces artificial excesses





Did we find new physics?

DID WE FIND NEW PHYSICS?



No ☹

A few excesses of max ~ 3σ **local** significance

Nothing significant

LIMITS AND IMPROVEMENT

$m_A = 3$ TeV

Signal Model (3 TeV)	Daughter Masses (GeV)	Method	Exp. (Obs.) Limit (fb)	Improvement wrt Inclusive
$Q* \to qW'$	25	*CWoLa Hunting*	61.1 (30.1)	0.3
$Q* \to qW'$	80	*CATHODE*	50.0 (95.2)	0.4
$Q* \to qW'$	170	*VAE-QR*	52.5 (37.5)	0.4
$Q* \to qW'$	400	*CWoLa Hunting*	45.8 (24.3)	0.5
$X \to YY' \to 4q$	25/25	*CATHODE*	8.0 (9.9)	0.9
$X \to YY' \to 4q$	25/80	*CATHODE*	7.6 (13.2)	0.9
$X \to YY' \to 4q$	25/170	*CATHODE*	10.3 (18.4)	0.7
$X \to YY' \to 4q$	25/400	*VAE-QR*	13.6 (12.5)	0.6
$X \to YY' \to 4q$	80/80	*CATHODE*	4.2 (8.0)	1.6
$X \to YY' \to 4q$	80/170	*CATHODE*	5.7 (11.4)	1.2
$X \to YY' \to 4q$	80/400	*CATHODE*	6.0 (7.3)	1.2
$X \to YY' \to 4q$	170/170	*CATHODE*	3.7 (6.8)	1.9
$X \to YY' \to 4q$	170/400	*VAE-QR*	4.4 (4.0)	1.7
$X \to YY' \to 4q$	400/400	*VAE-QR*	2.1 (1.9)	4.2
$W' \to B't \to bZt$	25	*TNT*	25.2 (17.4)	1.5
$W' \to B't \to bZt$	80	*TNT*	22.3 (14.6)	1.5
$W' \to B't \to bZt$	170	*TNT*	12.2 (7.3)	2.1
$W' \to B't \to bZt$	400	*VAE-QR*	15.2 (11.4)	1.8
$W_{KK} \to RW \to 3W$	170	*TNT*	25.1 (20.1)	1.4
$W_{KK} \to RW \to 3W$	400	*CWoLa Hunting*	23.8 (25.0)	1.5
$Z' \to T'T' \to tZtZ$	400	*QUAK*	28.3 (13.9)	2.7
$Y \to HH \to 4t$	400	*QUAK*	7.7 (3.7)	3.5

$m_A = 5$ TeV

Signal Model (5 TeV)	Daughter Masses (GeV)	Method	Exp. (Obs.) Limit (fb)	Improvement wrt Inclusive
$Q* \to qW'$	25	*QUAK*	3.5 (3.1)	0.7
$Q* \to qW'$	80	*QUAK*	3.2 (2.8)	0.8
$Q* \to qW'$	170	*QUAK*	3.3 (3.6)	0.8
$Q* \to qW'$	400	*QUAK*	3.9 (9.9)	0.7
$X \to YY' \to 4q$	25/25	*QUAK*	1.7 (1.6)	0.5
$X \to YY' \to 4q$	25/80	*QUAK*	1.3 (1.3)	0.7
$X \to YY' \to 4q$	25/170	*QUAK*	1.1 (1.1)	0.8
$X \to YY' \to 4q$	25/400	*VAE-QR*	1.0 (3.4)	0.9
$X \to YY' \to 4q$	80/80	*TNT*	1.1 (1.2)	0.8
$X \to YY' \to 4q$	80/170	*QUAK*	0.9 (1.0)	0.9
$X \to YY' \to 4q$	80/400	*VAE-QR*	0.9 (3.0)	0.9
$X \to YY' \to 4q$	170/170	*CATHODE*	0.7 (0.7)	1.2
$X \to YY' \to 4q$	170/400	*VAE-QR*	0.7 (2.3)	1.2
$X \to YY' \to 4q$	400/400	*VAE-QR*	0.4 (1.1)	2.3
$W' \to B't \to bZt$	25	*TNT*	4.4 (6.2)	1.3
$W' \to B't \to bZt$	80	*TNT*	3.9 (5.7)	1.4
$W' \to B't \to bZt$	170	*TNT*	2.8 (3.5)	1.6
$W' \to B't \to bZt$	400	*TNT*	2.7 (3.8)	1.6
$W_{KK} \to RW \to 3W$	170	*TNT*	6.1 (7.2)	0.8
$W_{KK} \to RW \to 3W$	400	*VAE-QR*	5.4 (18.6)	0.9
$Y \to HH \to 4t$	400	*TNT*	1.5 (2.3)	2.5

These techniques can give large improvements over the baseline strategies while remaining model agnostic

NO ONE SIZE FITS ALL

No technique sticks out as better than all others across the board

Best performing "traditional" method

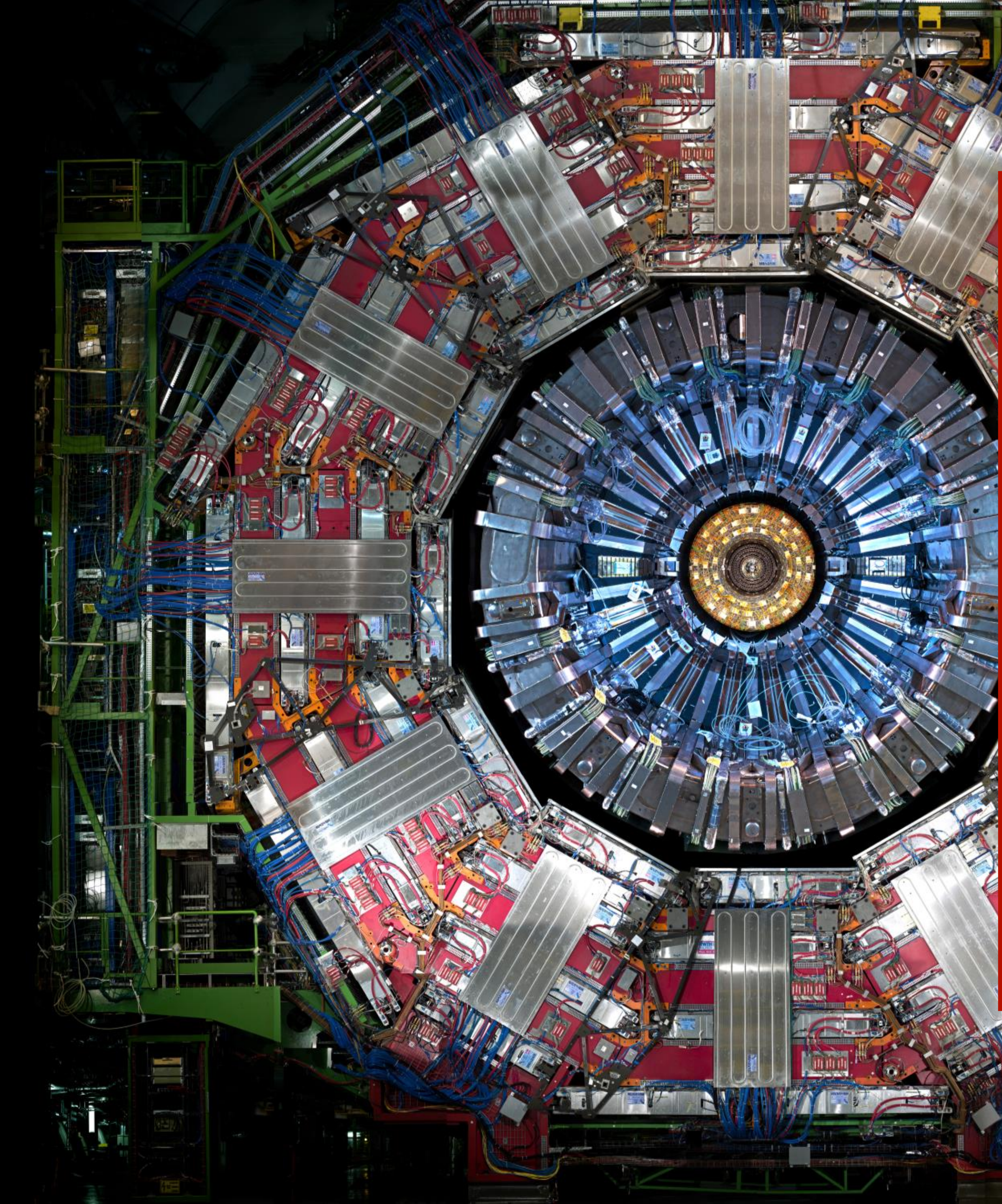
Greatest improvement over best performing "traditional" method

Prongs	Process $A \to BC$	Comparison Method	Improvements in Expected Sensitivities 95% CL Limit $m_A = 3$ TeV	95% CL Limit $m_A = 5$ TeV	5σ discovery $m_A = 3$ TeV
2+2	$\mathrm{X} \to \mathrm{YY}' \to 4\mathrm{q}$	2-prong $(\tau_{21}, m_{\mathrm{SD}})$	1.4 (*CATHODE*)	1.1 (*CATHODE*)	2.3 (*QUAK*)
3+3	$\mathrm{W}' \to \mathrm{B}'\mathrm{t} \to \mathrm{bZt}$	3-prong $(\tau_{32}, m_{\mathrm{SD}})$	1.2 (*VAE-QR*)	1.3 (*TNT*)	2.1 (*TNT*)
4+2	$\mathrm{W}_{\mathrm{KK}} \to \mathrm{RW} \to 3\mathrm{W}$	Inclusive	1.5 (*CWoLa Hunting*)	0.9 (*VAE-QR*)	2.6 (*CWoLa Hunting*)
6+6	$\mathrm{Y} \to \mathrm{HH} \to 4\mathrm{t}$	3-prong $(\tau_{32}, m_{\mathrm{SD}})$	1.5 (*QUAK*)	1.4 (*TNT*)	3.3 (*TNT*)

Improvement in expected upper limit

Improvement in 5σ discovery threshold

I.e., what cross section would have given 5σ expected



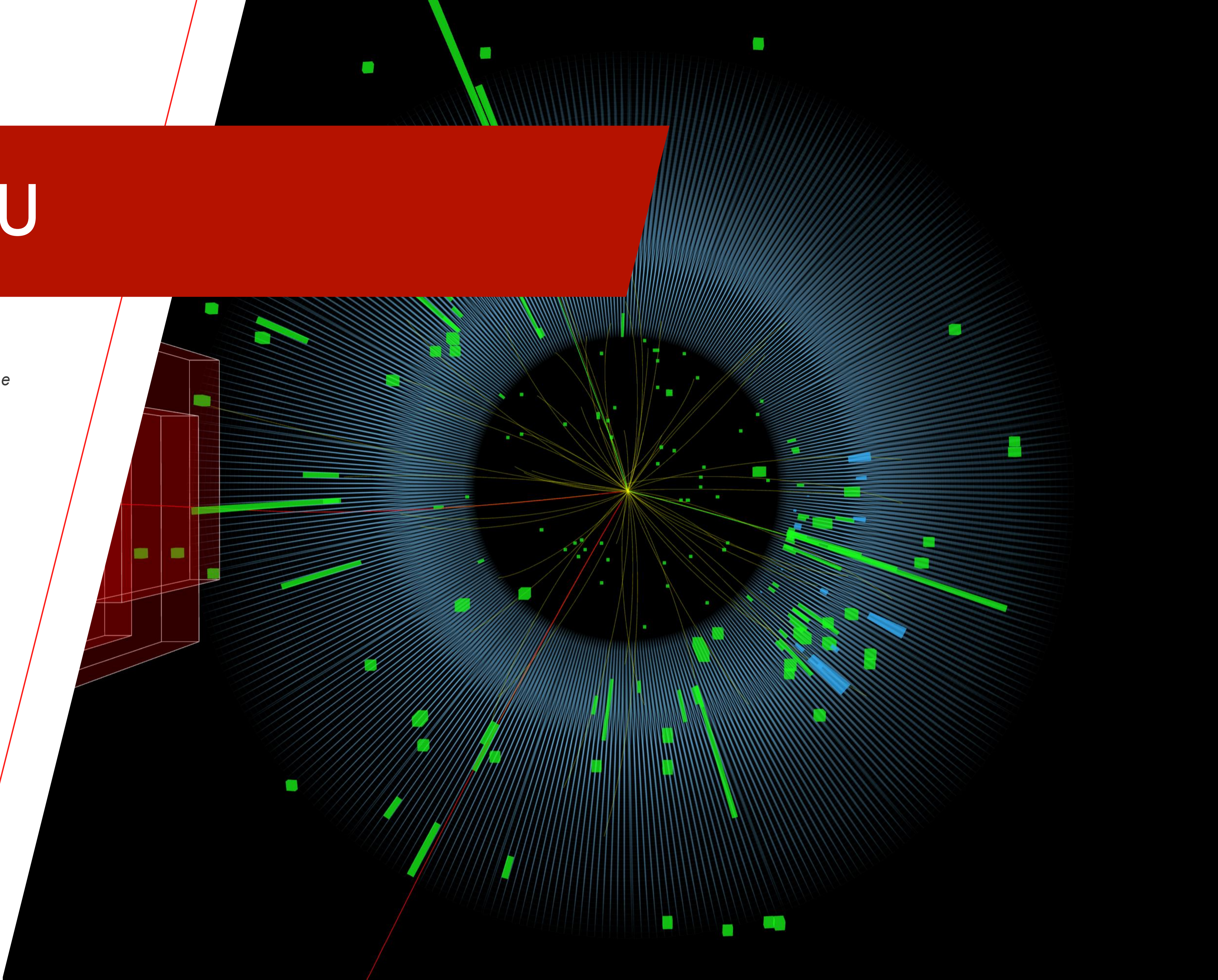
SUMMARY

- CMS has performed a model-agnostic search for new physics manifesting as a dijet resonance
- The use of machine-learning-based anomaly detection techniques allows the analysis to exploit jet substructure to enhance the sensitivity while remaining agnostic
- No significant excess observed, but demonstrated enhanced sensitivity

THANK YOU

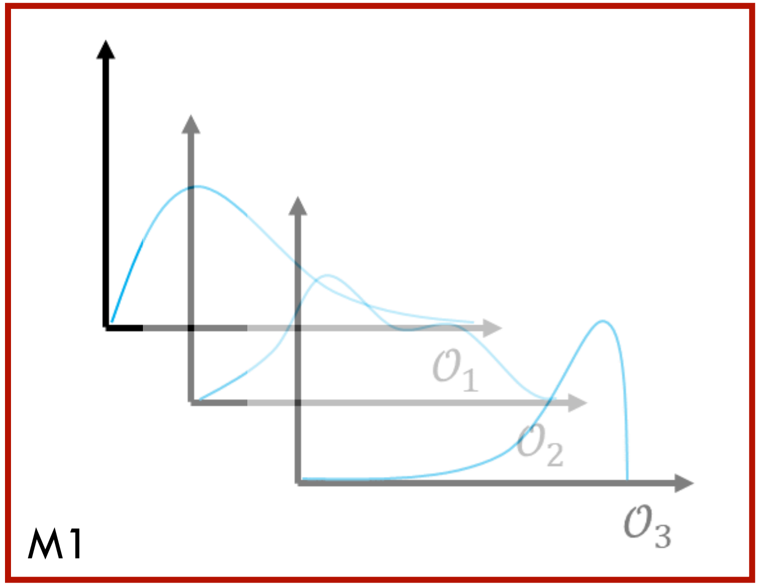
« Ce qui est admirable, ce n’est pas que le champ des étoiles soit si vaste, c’est que l’homme l’ait mesuré. »

Jacques Anatole François Thibault

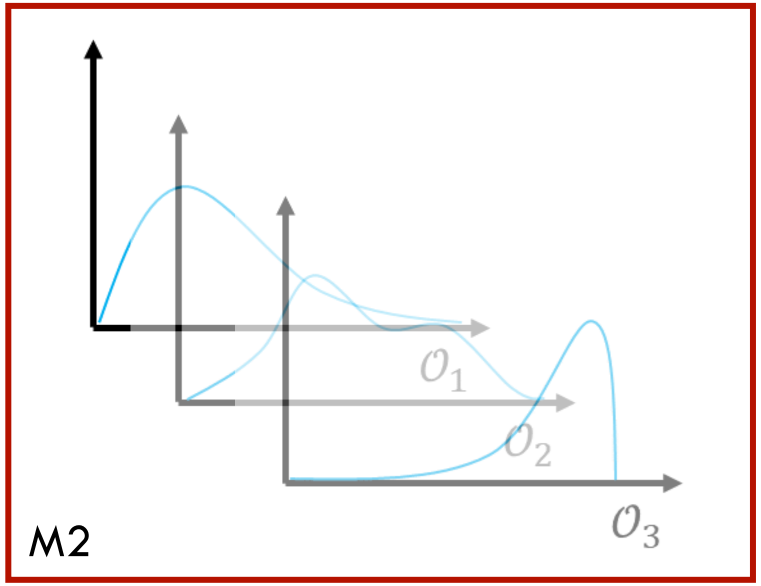


BACKUP

THE WEAKLY SUPERVISED PARADIGM



You can tell these apart



$$L_{M1/M2} = \frac{p_{M1}}{p_{M2}} = \frac{f_1 p_S + (1 - f_1) p_B}{f_2 p_S + (1 - f_2) p_B} = \frac{f_1 L_{S/B} + (1 - f_1)}{f_2 L_{S/B} + (1 - f_2)}$$

Learning to distinguish **uneven** mixtures of S and B means learning to distinguish S and B