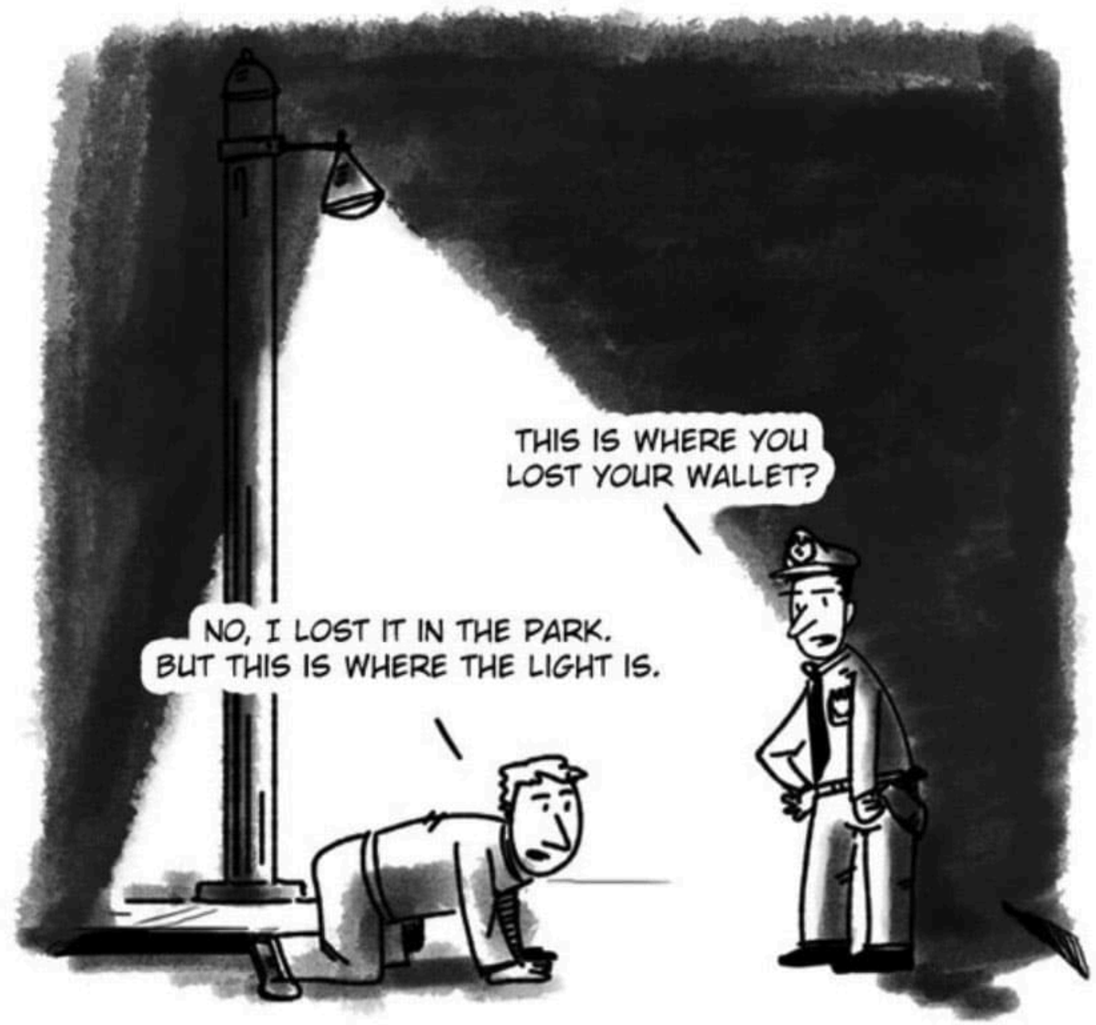
Expecting the Unexpected at the LHC and Beyond

David Shih

December 8, 2021

Israel Joint Particle Physics Seminar

Where (the f***) is the new physics??



Despite thousands of searches for new physics at the LHC, nothing but limits and null results so far.

What if new physics is hiding in the data but we haven't looked in the right places yet?

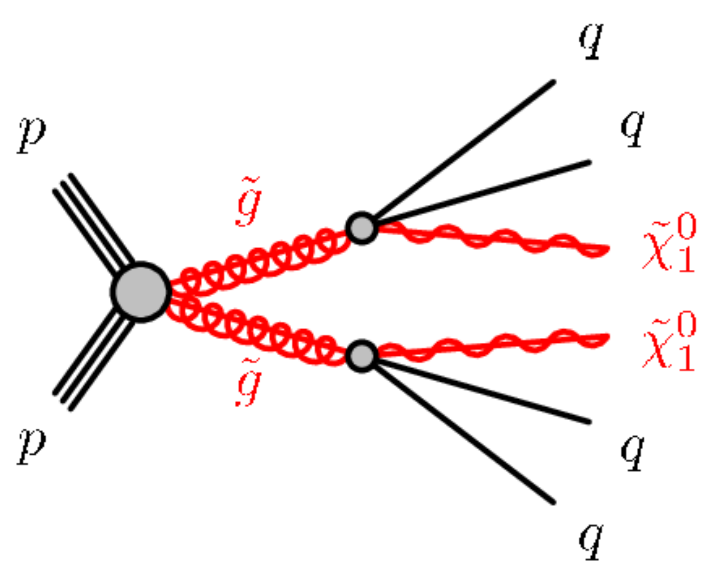
The most common approach

Model specific searches

Most NP searches at the LHC are heavily optimized with specific signals in mind (SUSY, extra dimensions, …)

ATLAS jets+MET 2010.14293

	BDT-GGd1	BDT-GGd2	BDT-GGd3	BDT-GGd4
N_{j}	≥ 4			
$\Delta\phi(j_{1,2,(3)}, \boldsymbol{p}_{\mathrm{T}}^{\mathrm{miss}})_{\mathrm{min}}$	> 0.4			
$\Delta\phi(j_{i>3}, \boldsymbol{p}_{\mathrm{T}}^{\mathrm{miss}})_{\mathrm{min}}$	> 0.4			
$E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}}(N_{\mathrm{j}})$	> 0.2			
m_{eff} [GeV]	> 1400		> 800	
BDT score	> 0.97	> 0.94	> 0.94	> 0.87
$\Delta m(\tilde{g}, \tilde{\chi}_1^0)$ [GeV]	1600–1900	1000–1400	600–1000	200–600



Kinematic cuts (or BDTs) optimized using simulations of signal AND background.

Of course, we should continue to perform these model-specific searches, because NP could always be right around the corner…

But we probably can't cover every possible model this way…

The landscape of NP signatures remains largely unexplored!

TABLE I. Existing two-body exclusive final state resonance searches at $\sqrt{s}$ = 8 TeV. The ∅ symbol indicates no existing search at the LHC.

	e	μ	τ	γ	j	b	t	W	Z	h
e	±∓[4],±±[5]	±±[5, 6] ±∓[6, 7]	[7]	∅	∅	∅	∅	∅	∅	∅
μ		±∓[4],±±[5]	[7]	∅	∅	∅	∅	∅	∅	∅
τ			[8]	∅	∅	∅	[9]	∅	∅	∅
γ				[10]	[11–13]	∅	∅	[14]	[14]	∅
j					[15]	[16]	[17]	[18]	[18]	∅
b						[16]	[19]	∅	∅	∅
t							[20]	[21]	∅	∅
W								[22–25]	[23, 24, 26, 27]	[28–30]
Z									[23, 25, 31]	[28, 30, 32, 33]
h										[34–37]

From Craig, Draper, Kong, Ng & Whiteson 1610.09392

										BSM $\to$ SM$_1$ $\times$ SM$_1$				BSM $\to$ SM$_1$ $\times$ SM$_2$			BSM $\to$ complex			
	e	μ	τ	q/g	b	t	γ	Z/W	H	q/g	γ/π^0's	b	$\cdots$	tZ/H	bH	$\cdots$	$\tau qq'$	eqq'	$\mu qq'$	$\cdots$
e	[37,38]	[39,40]	[39]	∅	∅	∅	[41]	[42]	∅	∅	∅	∅		∅	∅	∅	∅	[43,44]	∅	
μ		[37,38]	[39]	∅	∅	∅	[41]	[42]	∅	∅	∅	∅		∅	∅	∅	∅	∅	[43,44]	
τ			[45,46]	∅	[47]	∅	∅	∅	∅	∅	∅	∅		∅	∅	∅	[48,49]	∅	∅	
q/g				[29,30,50,51]	[52]	∅	[53,54]	[55]	∅	∅	∅	∅		∅	∅	∅	∅	∅	∅	
b					[29,52,56]	[57]	[54]	[58]	[59]	∅	∅	∅		[60]	∅	∅	∅	∅	∅	
t						[61]	∅	[62]	[63]	∅	∅	∅		[64]	[60]	∅	∅	∅	∅	
γ							[65,66]	[67–69]	[68,70]	∅	∅	∅		∅	∅	∅	∅	∅	∅	
Z/W								[71]	[71]	∅	∅	∅		∅	∅	∅	∅	∅	∅	
H									[72,73]	[74]	∅	∅		∅	∅	∅	∅	∅	∅	
BSM $\to$ SM$_1$ $\times$ SM$_1$: q/g										∅	∅	∅		∅	∅	∅	∅	∅	∅	
γ/π^0's											[75]	∅		∅	∅	∅	∅	∅	∅	
b												[76,77]		∅	∅	∅	∅	∅	∅	
$\vdots$																				
$\vdots$																				

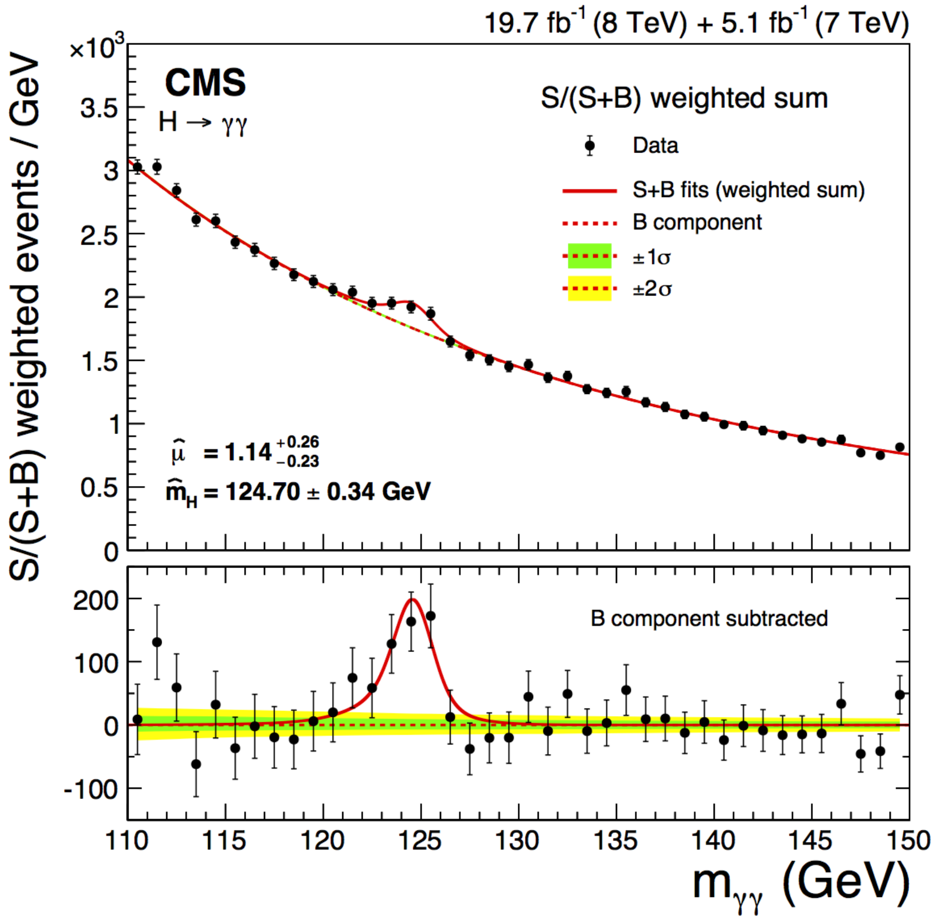
From Kim, Kong, Nachman & Whiteson 1907.06659

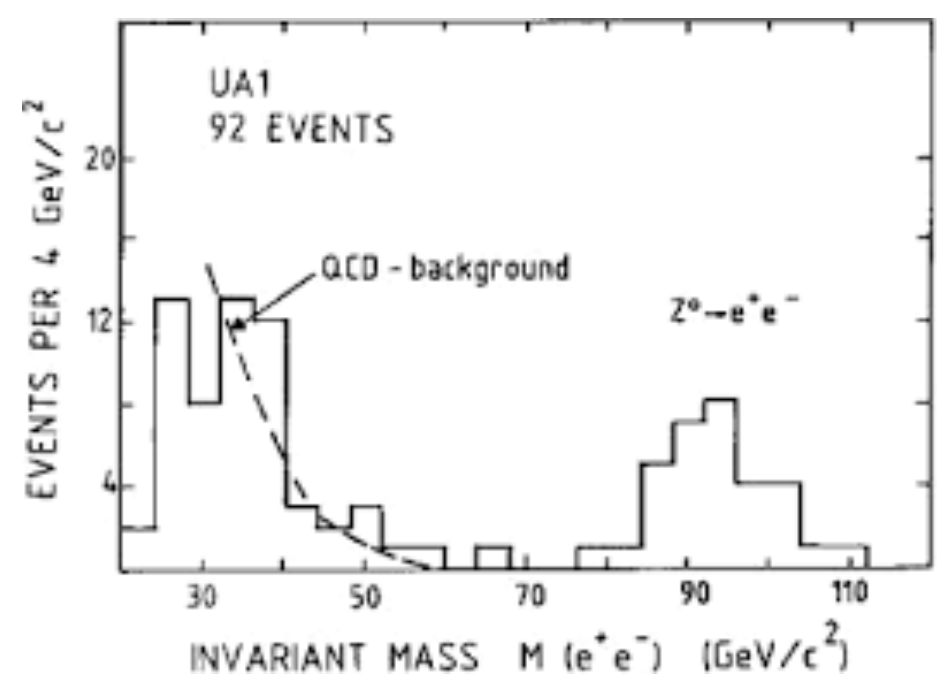
Existing model-independent approaches

"the bump hunt"

Idea: assume signal is localized in some feature (usually invariant mass) while background is smooth.

Interpolate from **sidebands** into **signal region**, search for an excess.





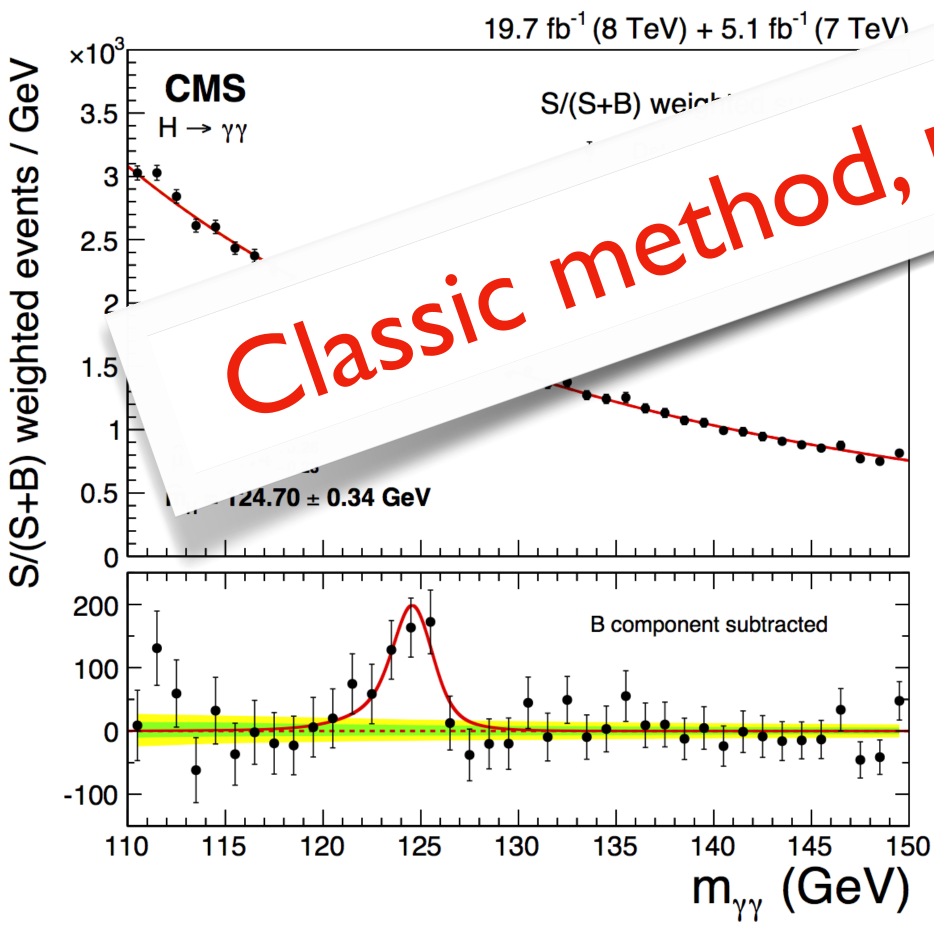
Existing model-independent approaches

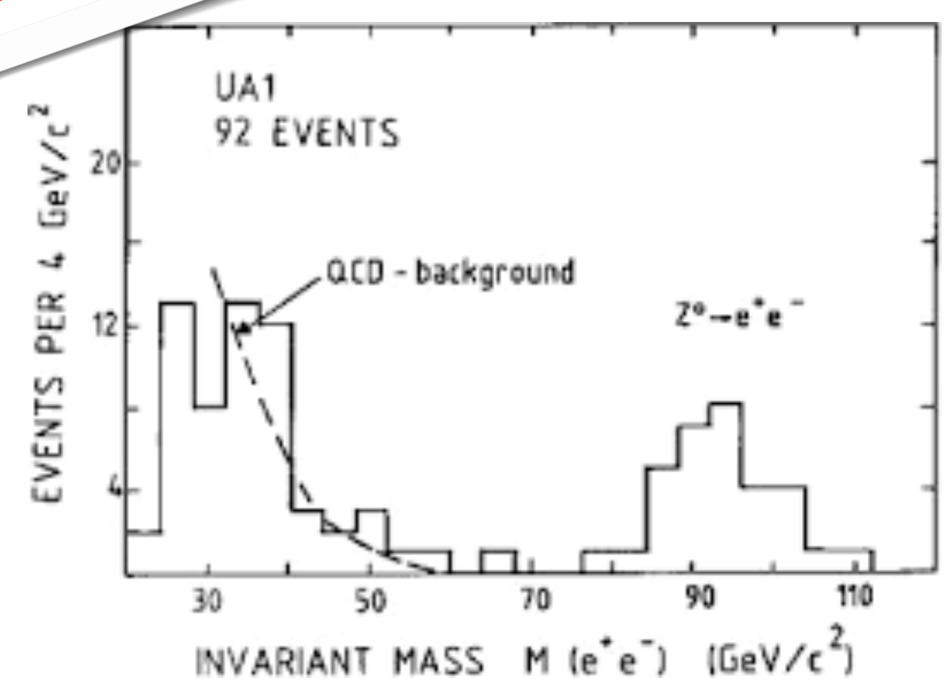
"the bump hunt"

Idea: assume signal is localized in some feature (usually invariant mass) while background is smooth.

Interpolate from **sidebands** into **signal re** … excess.

Classic method, used in many discoveries.

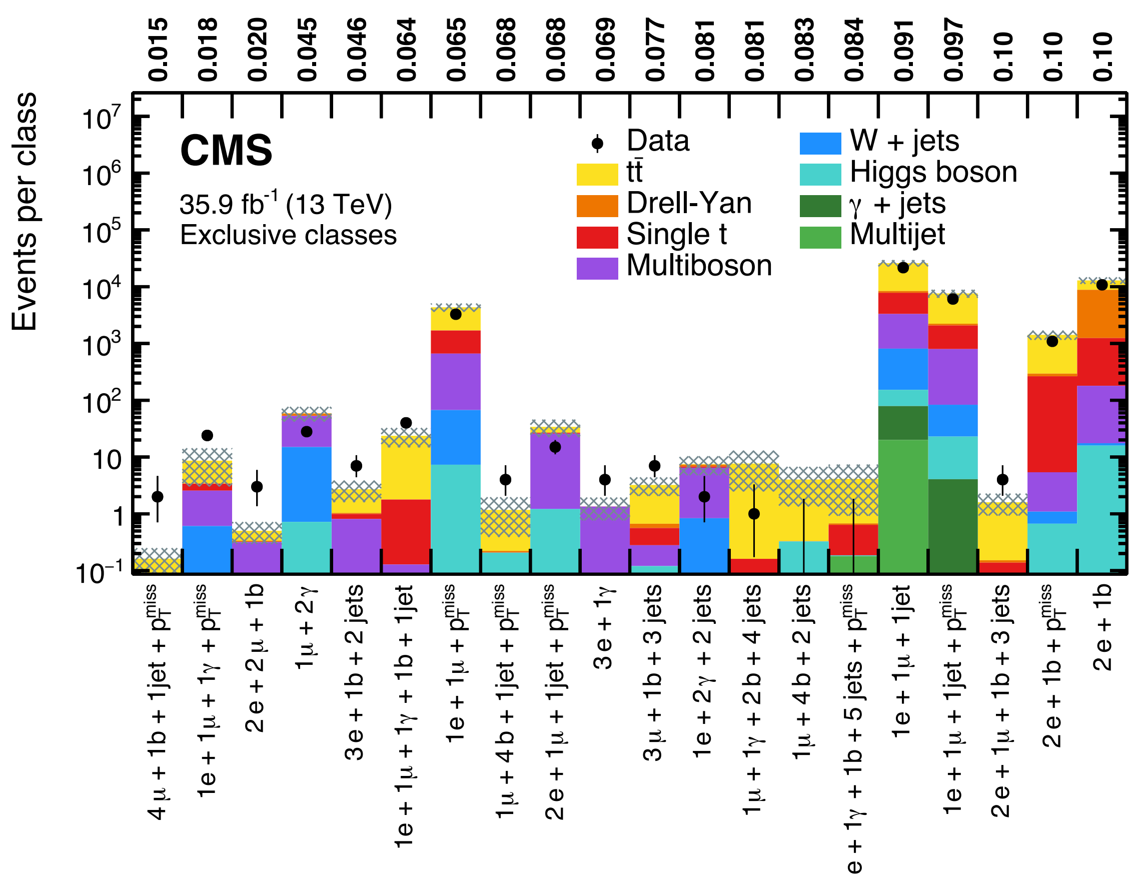




Existing model-independent searches

"the general search"

Idea: divide the phase space up into thousands of bins, compare **data to SM simulation** in each one



CMS

"MUSIC"

CMS-PAS-EXO-14-016

ATLAS

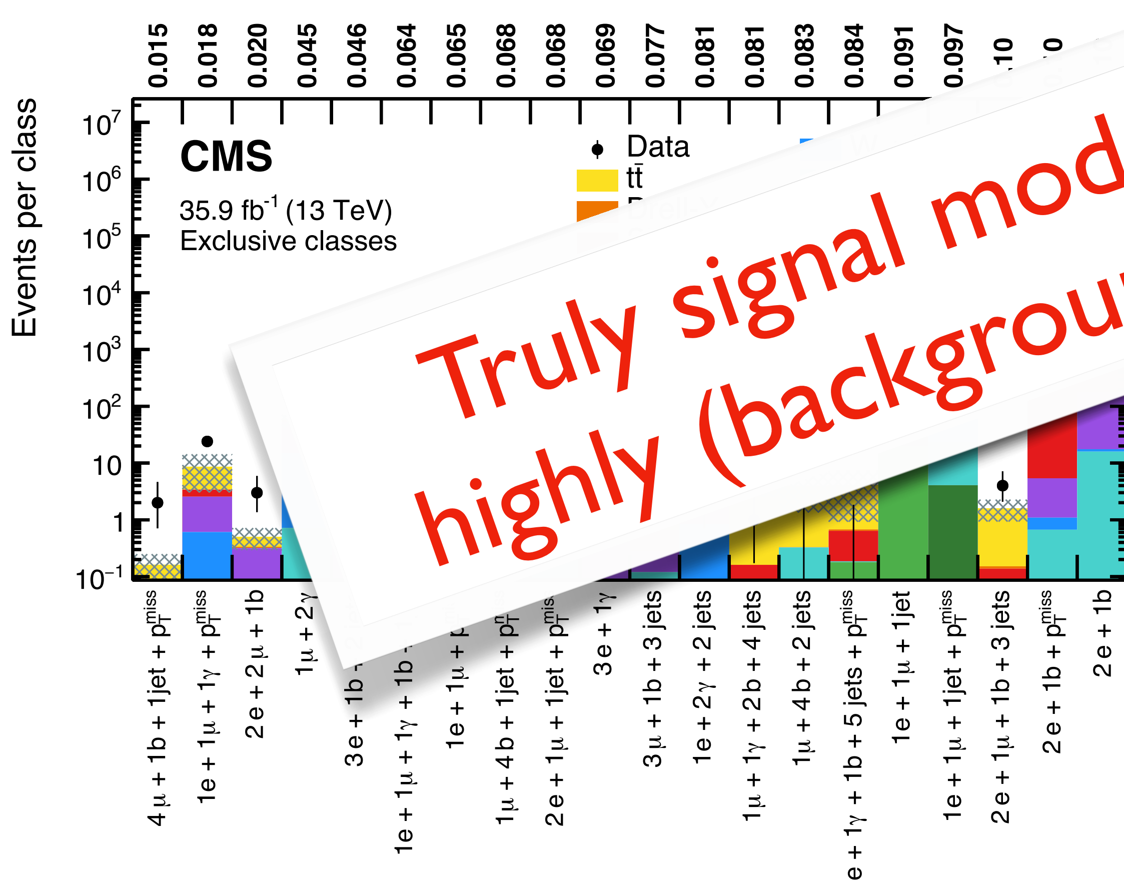
"Model independent general search"

1807.07447
EPJC 79:120 (2019)

Existing model-independent searches

"the general search"

Idea: divide the phase space up into thousands of bins, com
data to SM simulation in each one



Truly signal model independent, but still highly (background) simulation dependent

"MUSIC"

CMS-PAS-EXO-14-016

ATLAS

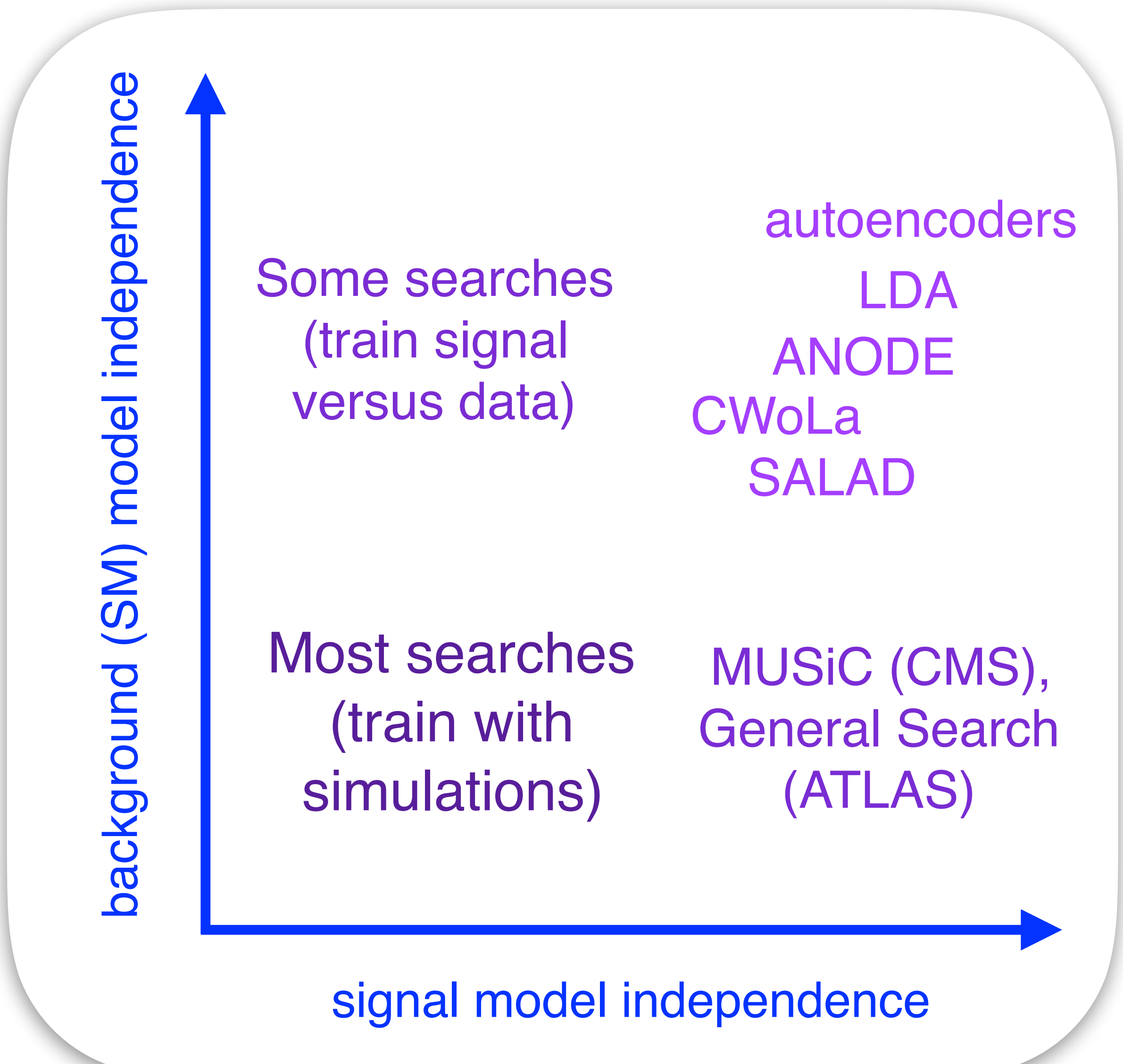
"Model independent general search"

1807.07447
EPJC 79:120 (2019)

New paradigms for model-agnostic searches

Can modern advances in machine learning open up new avenues for model-independent searches?

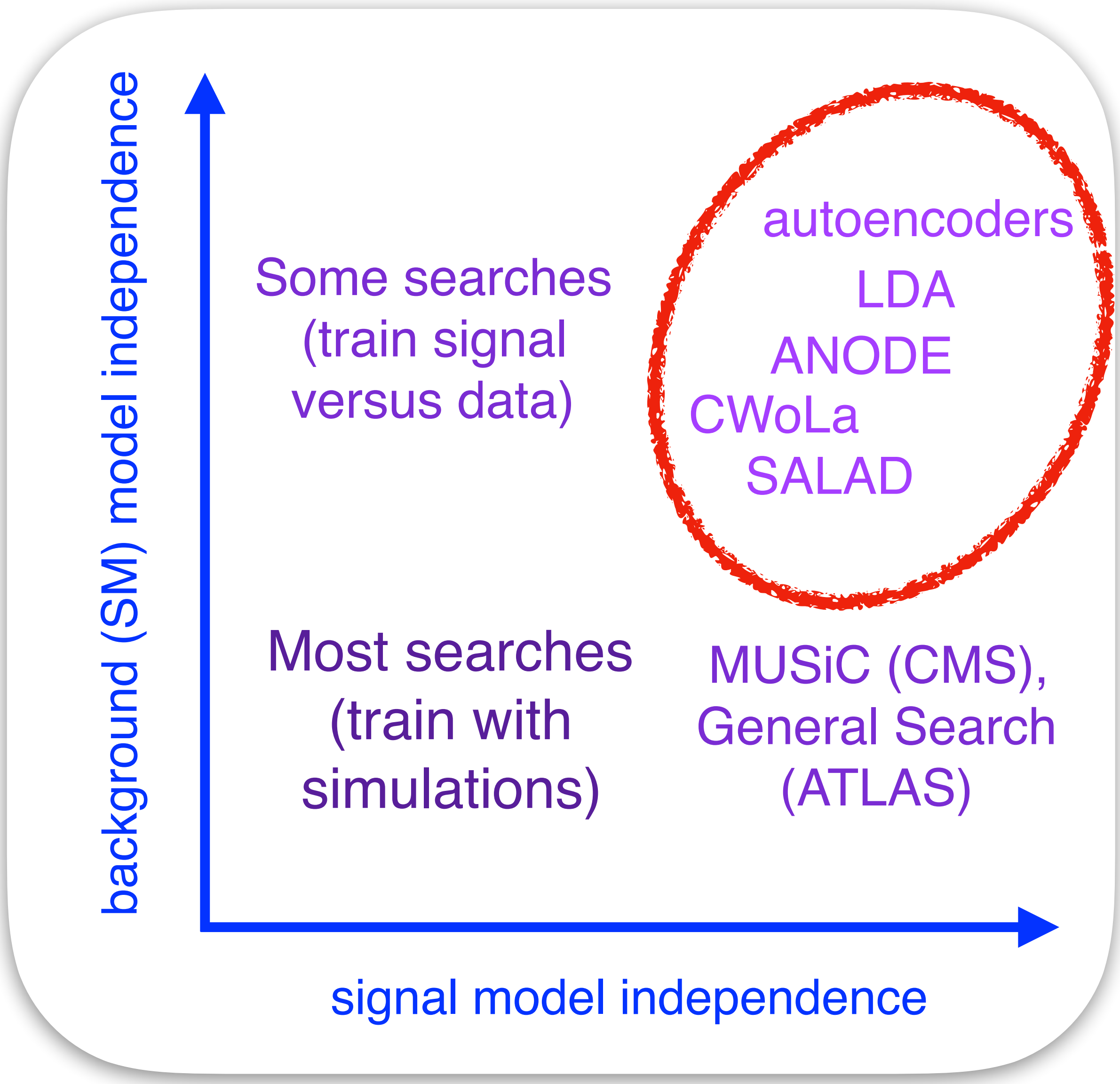
from Nachman & DS 2001.04990



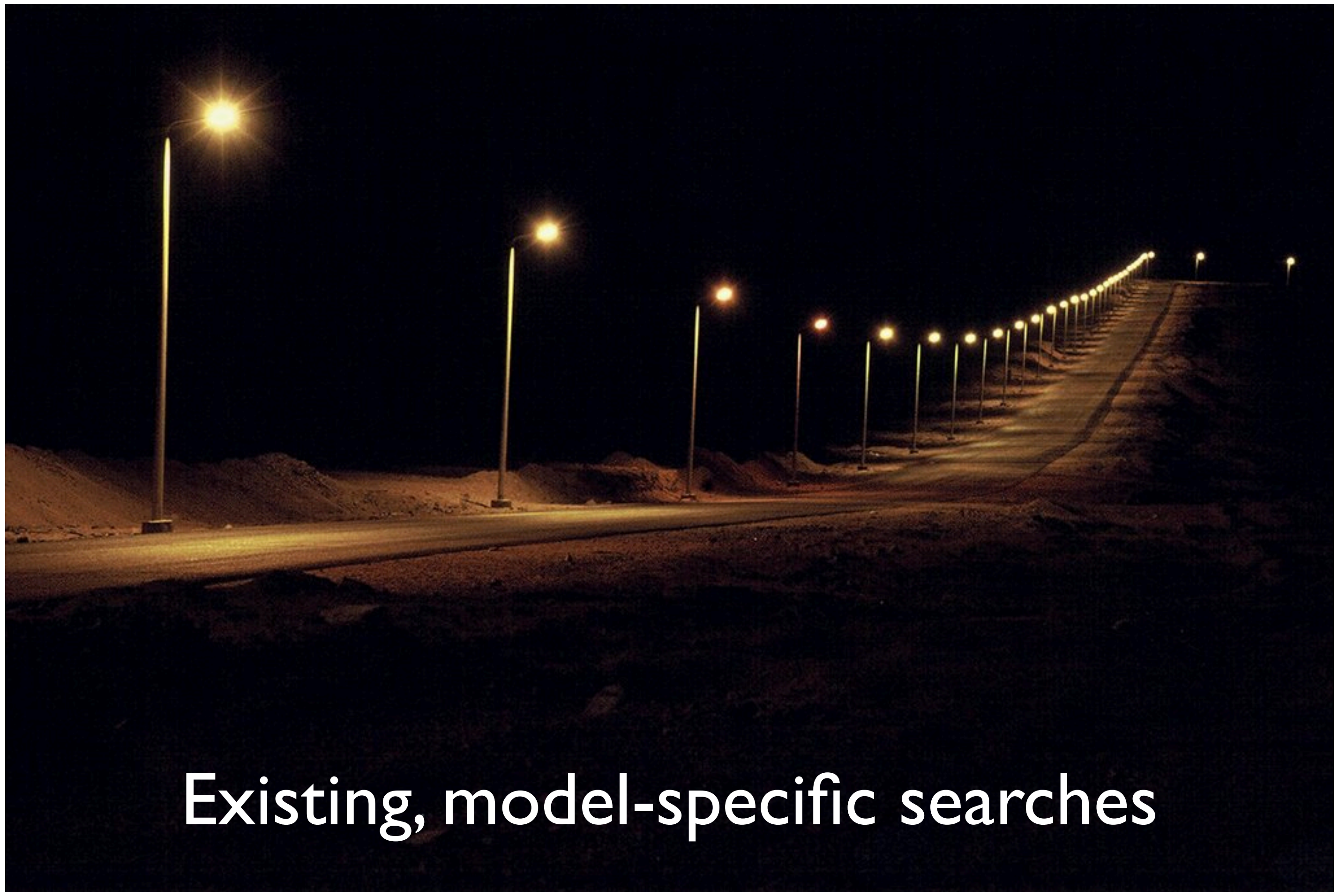
New paradigms for model-agnostic searches

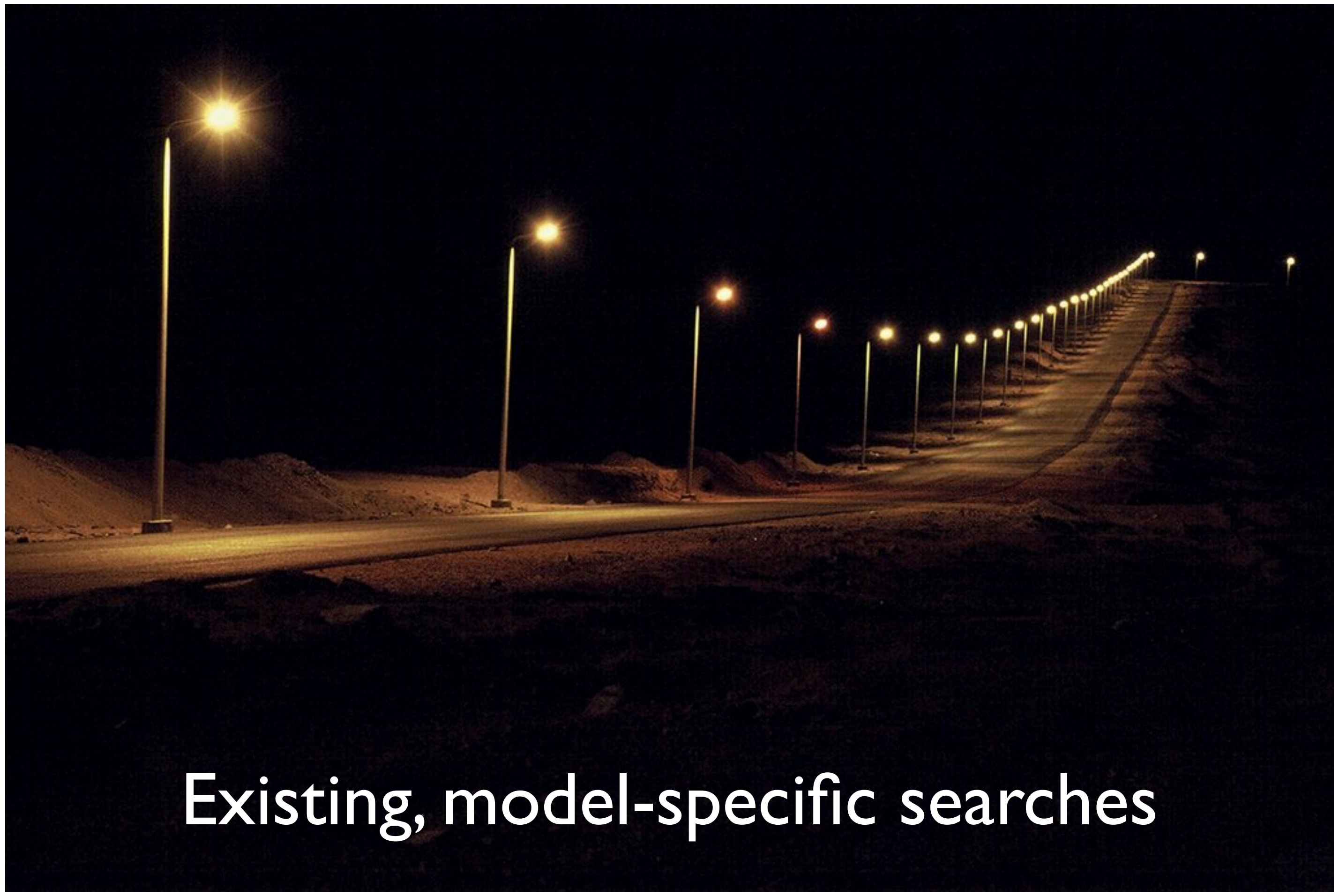
Can modern advances in machine learning open up new avenues for model-independent searches?

from Nachman & DS 2001.04990



Many new ideas recently!

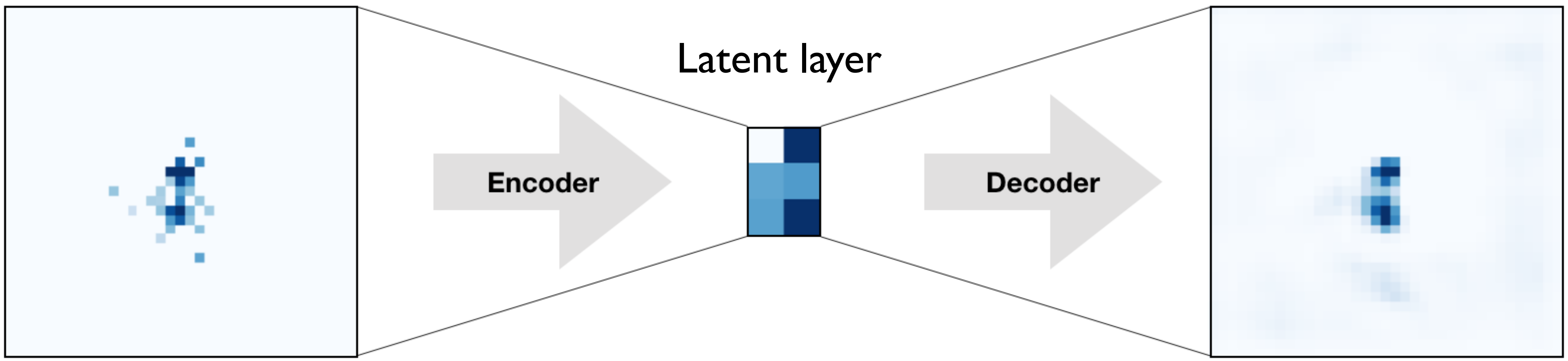

Existing, model-specific searches


Existing, model-specific searches


Model-agnostic searches with modern ML

Searching for NP with deep autoencoders

Farina, Nakai & DS 1808.08992; Heimel, Kasieczka, Plehn & Thompson 1808.08979



An autoencoder maps an input into a "latent representation" and then attempts to reconstruct the original input from it.

The encoding is lossy, so the reconstruction is not perfect.

Many real world applications of autoencoders, including anomaly detection, fraud detection, denoising, compression, generation, density estimation

See also:
Hajer et al "Novelty Detection Meets Collider Physics" 1807.10261
Cerri et al "Variational Autoencoders for New Physics Mining at the Large Hadron Collider" 1811.10276
and many more….!

Searching for NP with deep autoencoders

Farina, Nakai & DS 1808.08992; Heimel, Kasieczka, Plehn & Thompson 1808.08979

Loss function for autoencoder:

"reconstruction error"

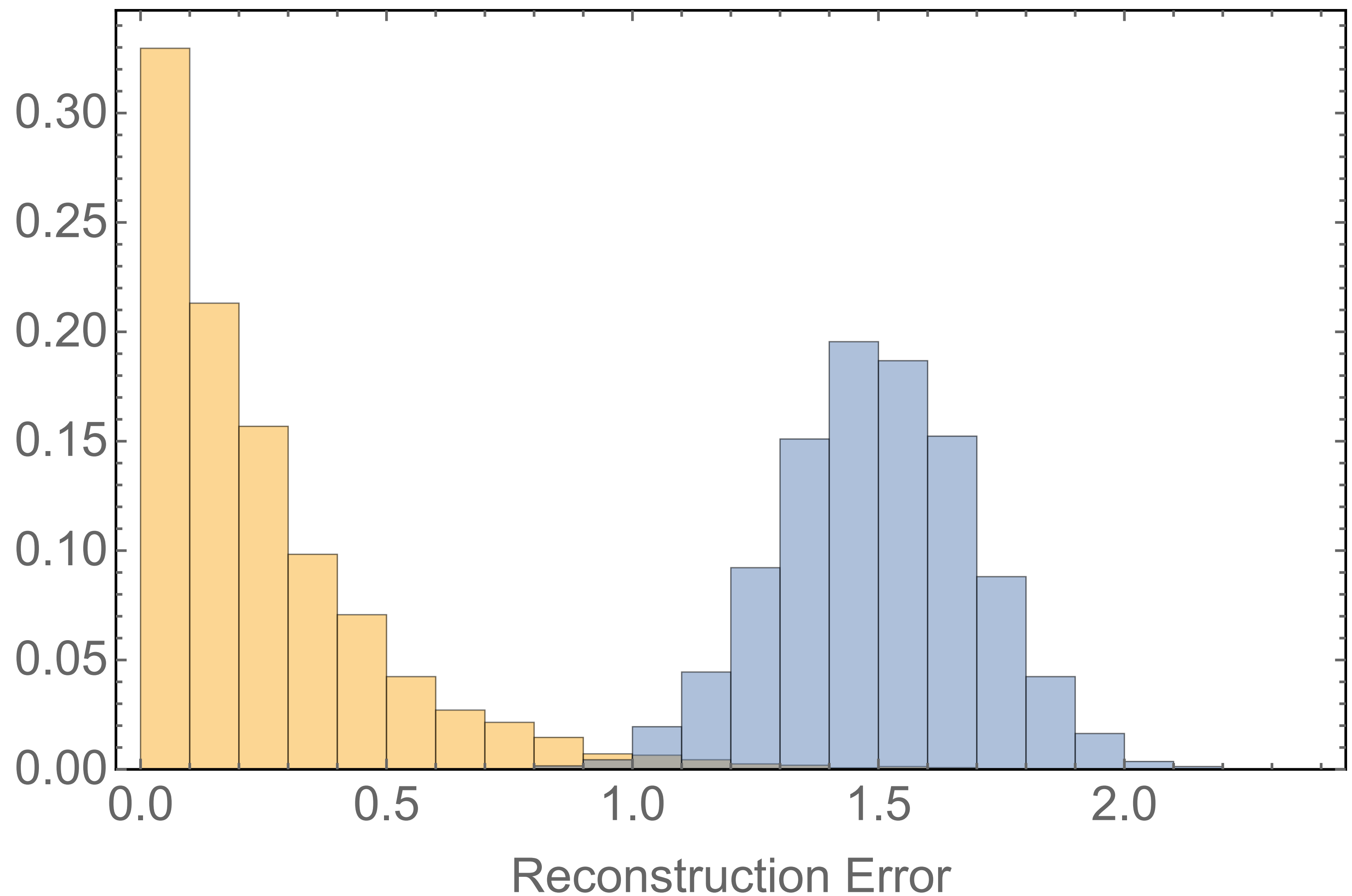
$$L = \frac{1}{N}\sum_{i=1}^{N}(x_i^{in} - x_i^{out})^2$$

Searching for NP with deep autoencoders

Farina, Nakai & DS 1808.08992; Heimel, Kasieczka, Plehn & Thompson 1808.08979

Loss function for autoencoder: $$L = \frac{1}{N}\sum_{i=1}^{N}(x_i^{in} - x_i^{out})^2$$

"reconstruction error"

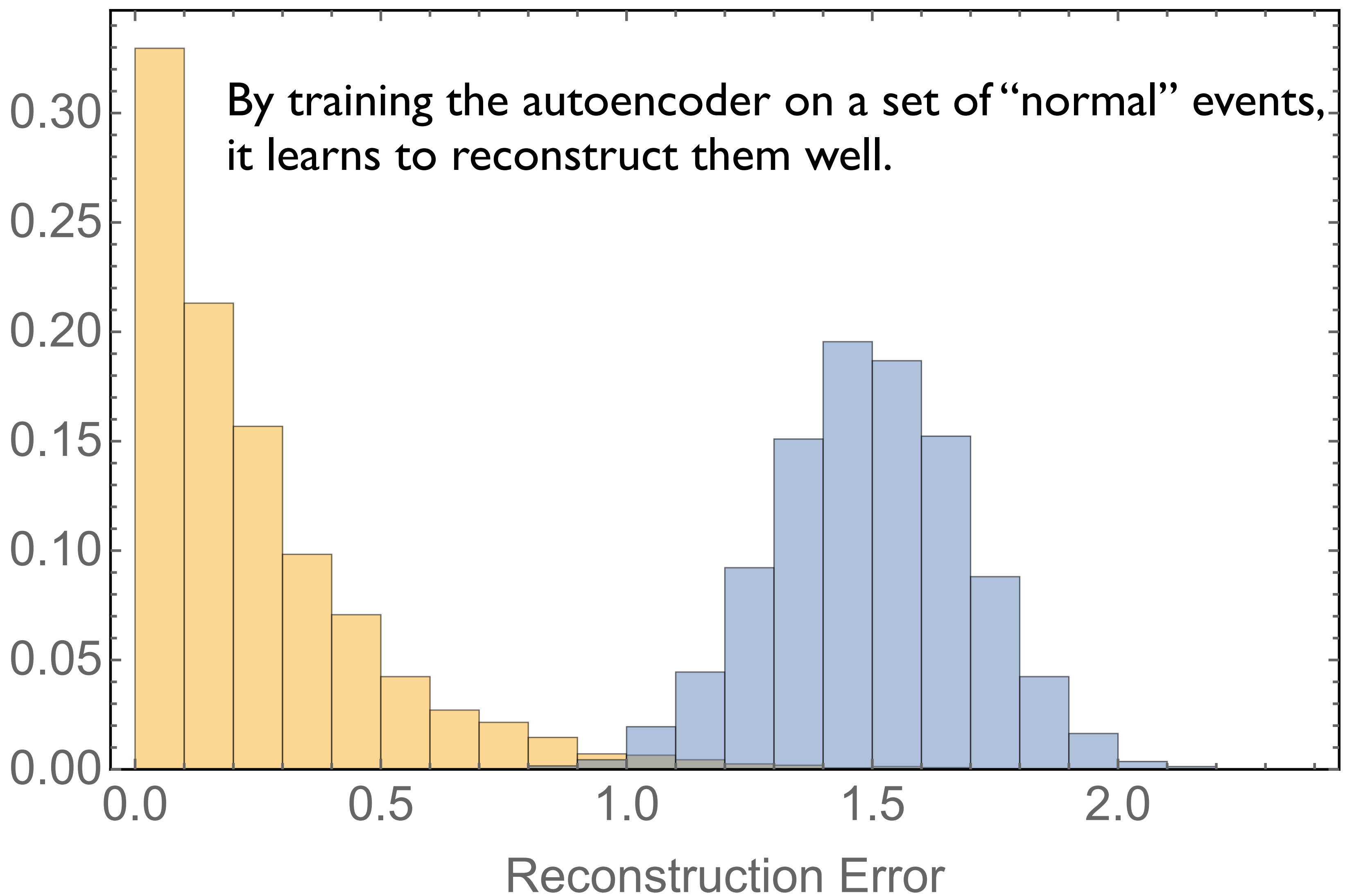


Searching for NP with deep autoencoders

Farina, Nakai & DS 1808.08992; Heimel, Kasieczka, Plehn & Thompson 1808.08979

Loss function for autoencoder: $$L = \frac{1}{N}\sum_{i=1}^{N}(x_i^{in} - x_i^{out})^2$$

"reconstruction error"



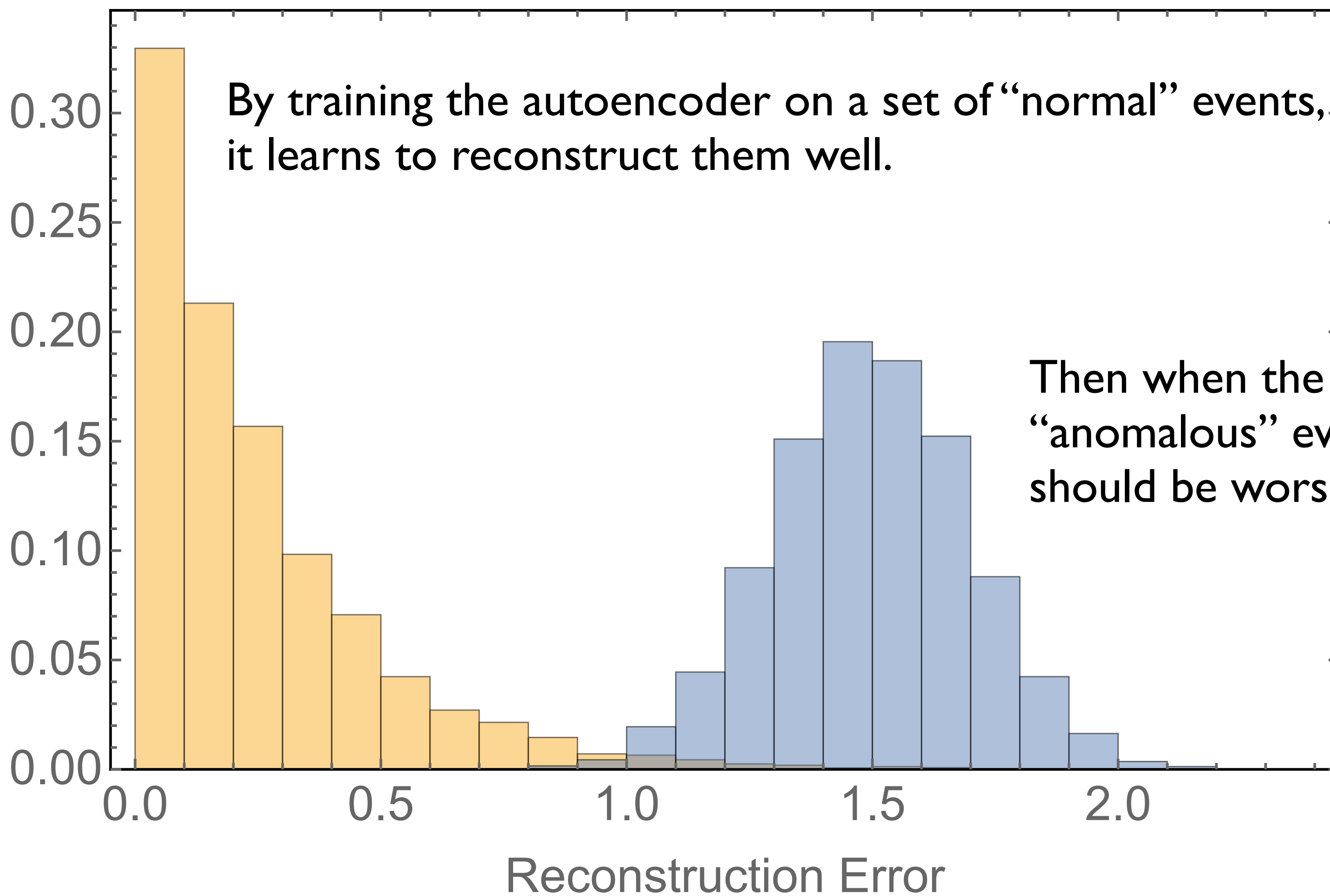
Searching for NP with deep autoencoders

Farina, Nakai & DS 1808.08992; Heimel, Kasieczka, Plehn & Thompson 1808.08979

Loss function for autoencoder: $$L = \frac{1}{N}\sum_{i=1}^{N}(x_i^{in} - x_i^{out})^2$$

"reconstruction error"

By training the autoencoder on a set of "normal" events, it learns to reconstruct them well.



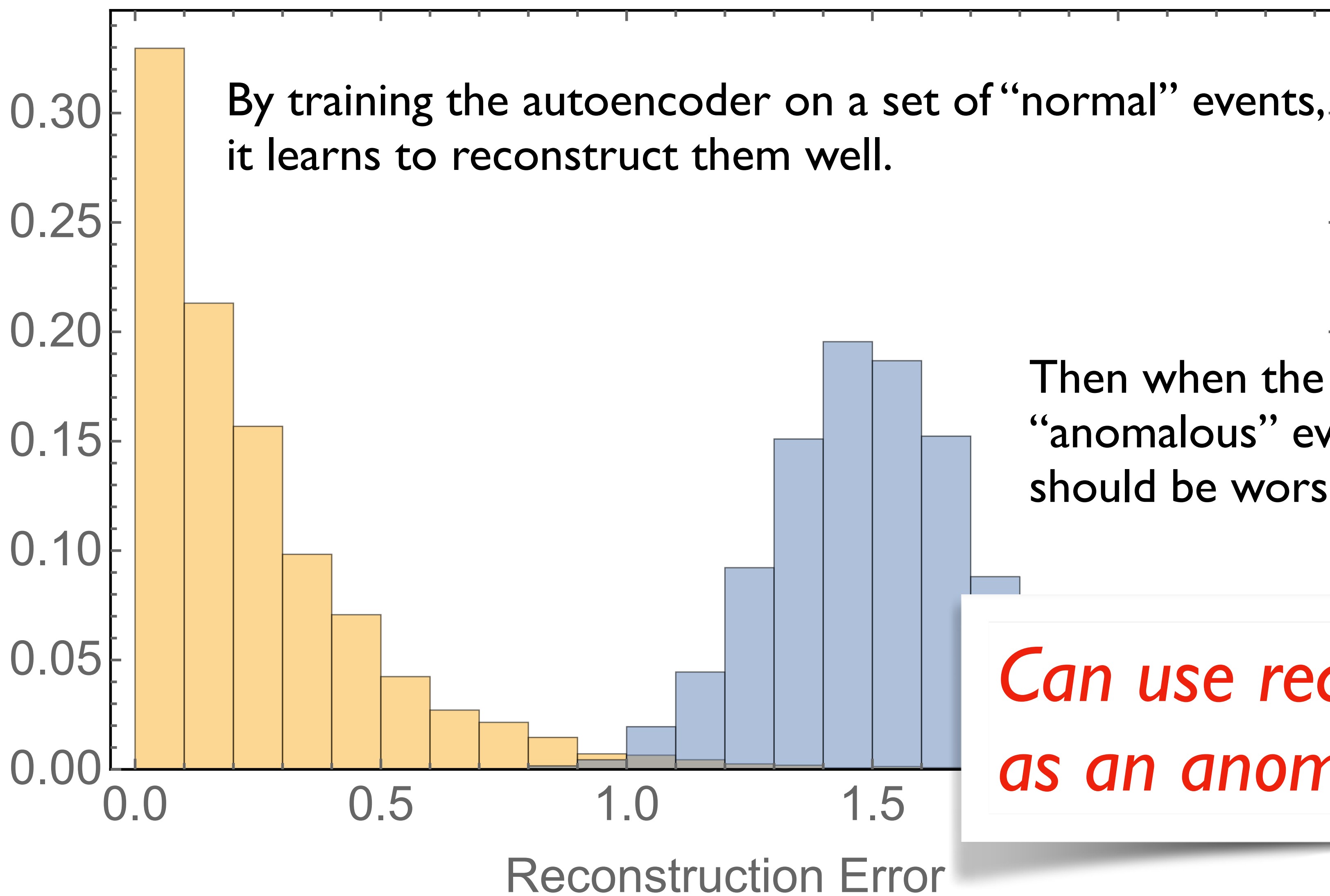
Then when the autoencoder encounters "anomalous" events, its performance should be worse.

Searching for NP with deep autoencoders

Farina, Nakai & DS 1808.08992; Heimel, Kasieczka, Plehn & Thompson 1808.08979

Loss function for autoencoder: $$L = \frac{1}{N}\sum_{i=1}^{N}(x_i^{in} - x_i^{out})^2$$

"reconstruction error"



By training the autoencoder on a set of "normal" events, it learns to reconstruct them well.

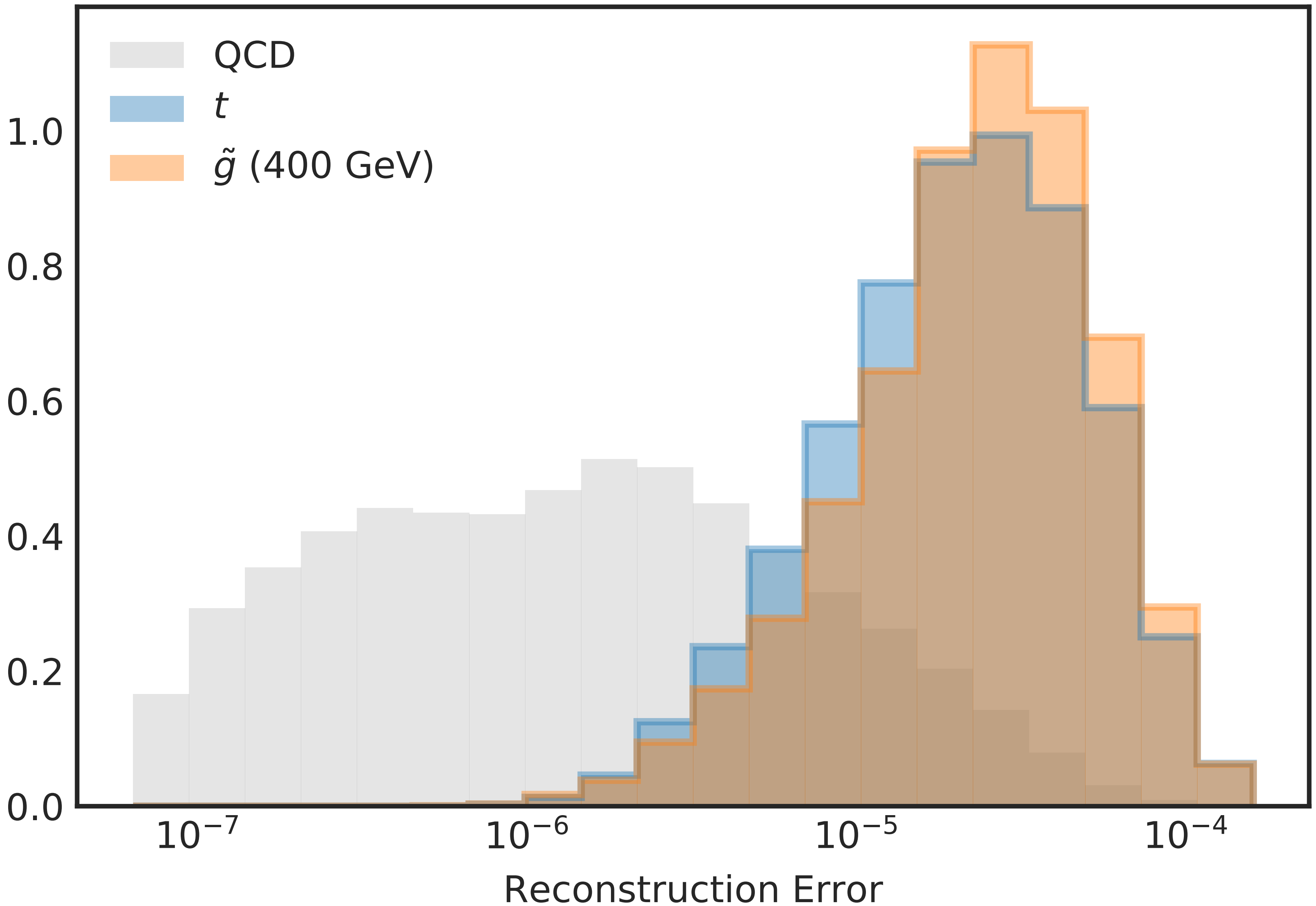
Then when the autoencoder encounters "anomalous" events, its performance should be worse.

Can use reconstruction error as an anomaly score!

Searching for NP with deep autoencoders

Farina, Nakai & DS 1808.08992; Heimel, Kasieczka, Plehn & Thompson 1808.08979

Train the AE on QCD backgrounds only.



It works as an anomaly detector!

Autoencoder pros and cons

Pros:

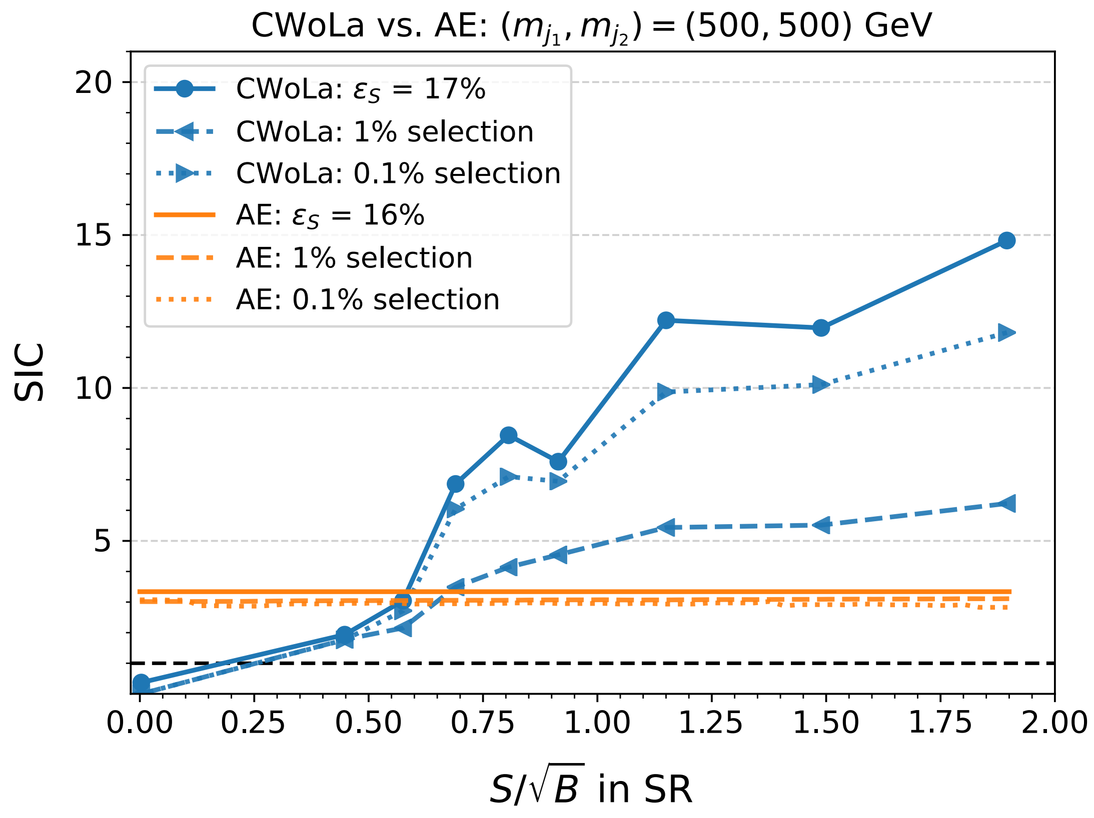
- Fully unsupervised — truly signal model independent, can find rare signals

Autoencoder pros and cons

Pros:

- Fully unsupervised — truly signal model independent, can find rare signals

Nice complementarity with weakly-supervised approaches
(Collins, Martin-Ramiro, Nachman & DS 2104.02092)

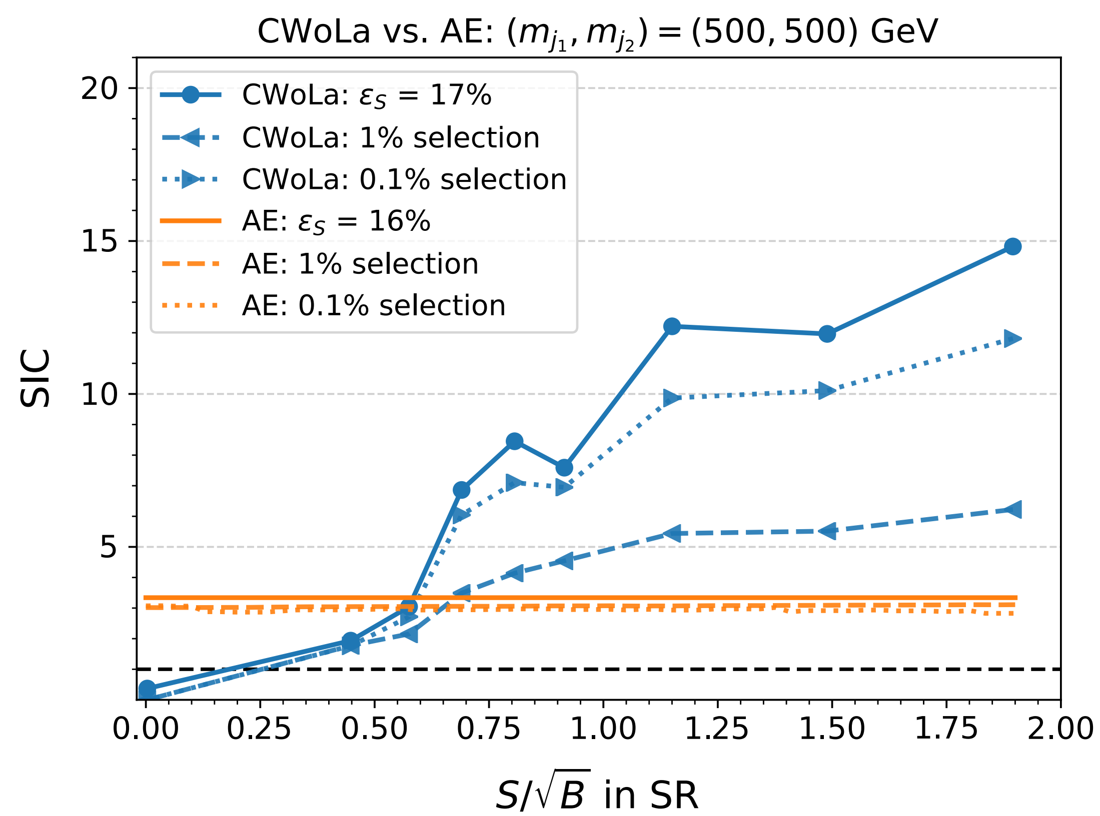


Autoencoder pros and cons

Pros:

- Fully unsupervised — truly signal model independent, can find rare signals

Nice complementarity with weakly-supervised approaches
(Collins, Martin-Ramiro, Nachman & DS 2104.02092)



AE as "early detection system"

Autoencoder pros and cons

Cons:

- Uncontrolled, not very sensitive, optimality not guaranteed — the AE will find what it finds…

- The reconstruction loss can fail to detect anomalies if they are "simpler" than the background
  (T. Weber Bachelor Thesis [G. Kasieczka]; Dillon, Plehn, Sauer & Sorrenson 2104.08291; Finke, Kraemer, Morandini, Mueck & Oleksiyuk 2104.09051)

- Not obvious how to perform background estimation with AE anomaly detection

  - Can combine with bump hunt at cost of model-independence
    (Farina, Nakai & DS 1808.08992; Heimel, Kasieczka, Plehn & Thompson 1808.08979)

Double Decorrelated Autoencoder

Mikuni, Nachman & DS (2111.06417)

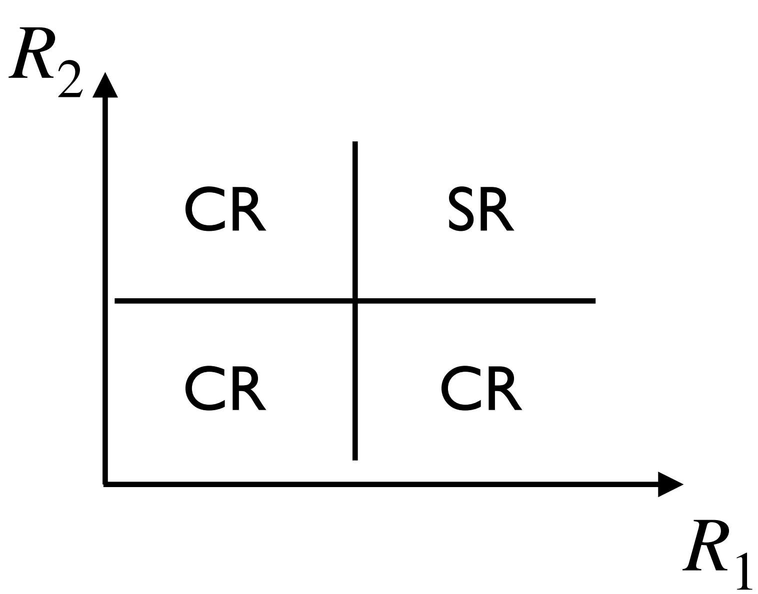
Idea for fully unsupervised anomaly detection with background estimation:

- Train ***two*** autoencoders and force them to be statistically independent of one another:

$$L[f_1, f_2, g_1, g_2] = \sum_i R_1(x_i)^2 + \sum_i R_2(x_i)^2 + \lambda\,\mathrm{DisCo}^2[R_1(X), R_2(X)]$$

"DisCo Decorrelation"
Kasieczka & DS
2001.05310

- Can use ABCD method for fully data-driven background estimation



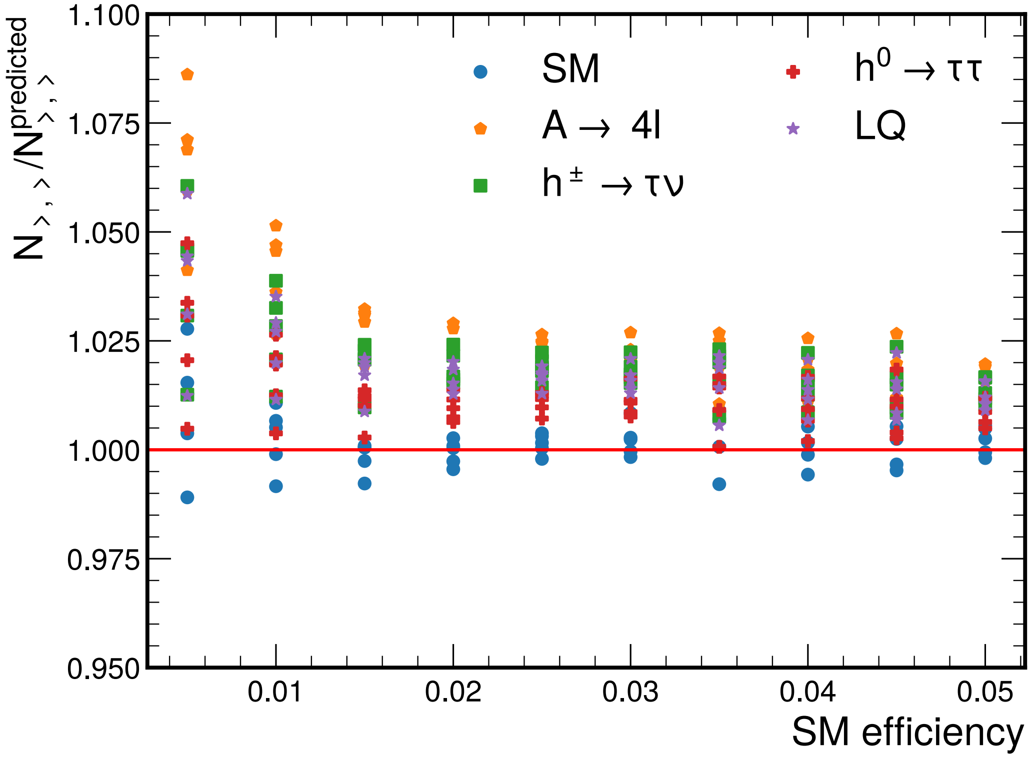
$$N^{\text{predicted}}_{>,>}(\vec{c}) = \frac{N_{>,<}(\vec{c})\,N_{<,>}(\vec{c})}{N_{<,<}(\vec{c})}$$

Double Decorrelated Autoencoder

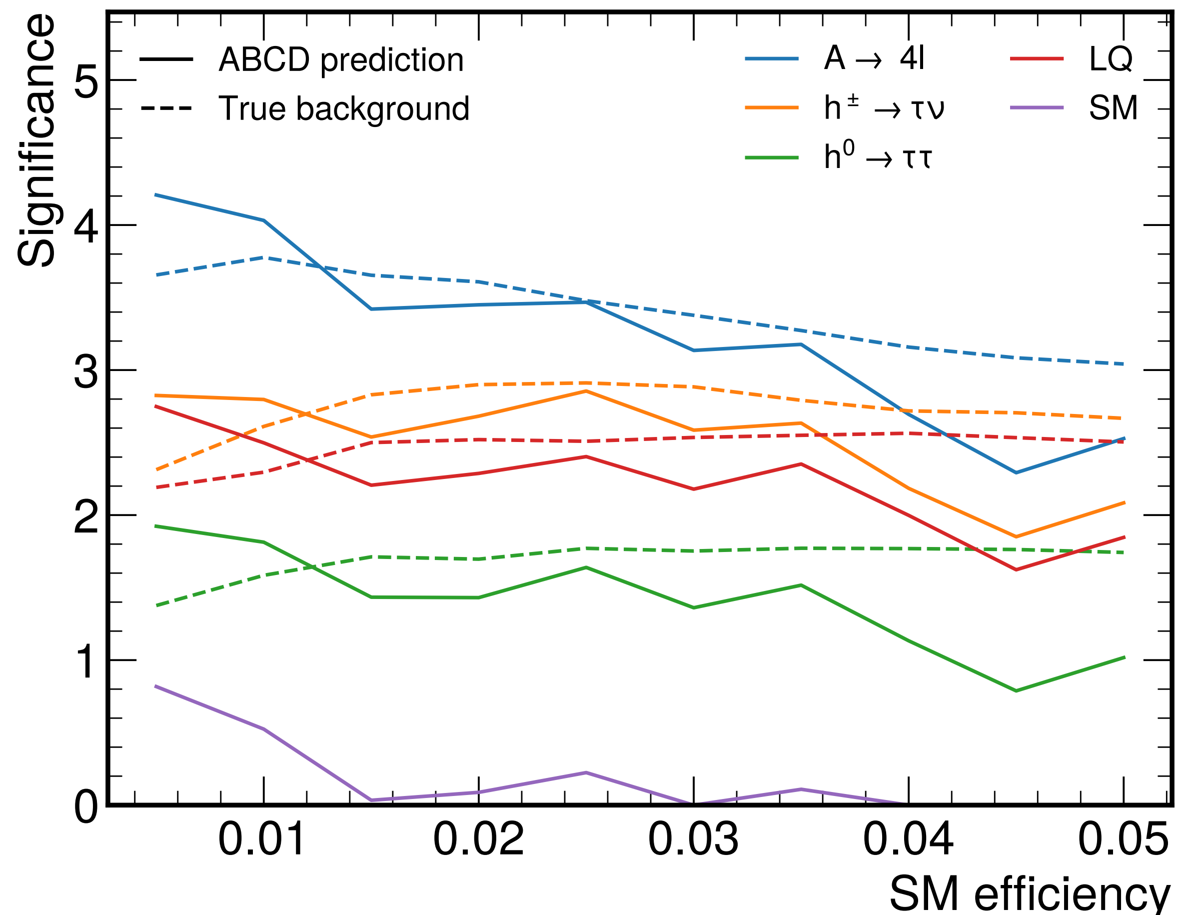
Mikuni, Nachman & DS (2111.06417)

The method works!

(Initial NP significance: 0.8σ)



ADC2021 1-lepton dataset
Govorkova et al 2107.02157



First complete strategy for unsupervised, non-resonant anomaly detection

Can also be used online as an anomaly trigger

LHC Olympics 2020

https://doi.org/10.5281/zenodo.3547721

In 2019, Gregor Kasieczka, Ben Nachman and I initiated the LHC Olympics 2020 Challenge.

It consisted of three "black boxes" of simulated data (bg dominated!):

- 1 million events each
- 4-vectors of every reconstructed particle (all hadronic) in the event
- Particle ID, charge, etc not included
- Single R=1 jet trigger pT>1.2 TeV

The goal of the challenge was for participants to analyze each box and

1. Decide whether or not it contains new physics
2. Characterize the new physics, if it's there

Many new approaches inspired by **LHCO2020**

The LHC Olympics 2020

A Community Challenge for Anomaly Detection in High Energy Physics

Gregor Kasieczka (ed),[1] Benjamin Nachman (ed),[2,3] David Shih (ed),[4] Oz Amram,[5] Anders Andreassen,[6] Kees Benkendorfer,[2,7] Blaz Bortolato,[8] Gustaaf Brooijmans,[9] Florencia Canelli,[10] Jack H. Collins,[11] Biwei Dai,[12] Felipe F. De Freitas,[13] Barry M. Dillon,[8,14] Ioan-Mihail Dinu,[5] Zhongtian Dong,[15] Julien Donini,[16] Javier Duarte,[17] D. A. Faroughy[10] Julia Gonski,[9] Philip Harris,[18] Alan Kahn,[9] Jernej F. Kamenik,[8,19] Charanjit K. Khosa,[20,30] Patrick Komiske,[21] Luc Le Pottier,[2,22] Pablo Martín-Ramiro,[2,23] Andrej Matevc,[8,19] Eric Metodiev,[21] Vinicius Mikuni,[10] Inês Ochoa,[24] Sang Eon Park,[18] Maurizio Pierini,[25] Dylan Rankin,[18] Veronica Sanz,[20,26] Nilai Sarda,[27] Uroš Seljak,[2,3,12] Aleks Smolkovic,[8] George Stein,[2,12] Cristina Mantilla Suarez,[5] Manuel Szewc,[28] Jesse Thaler,[21] Steven Tsan,[17] Silviu-Marian Udrescu,[18] Louis Vaslin,[16] Jean-Roch Vlimant,[29] Daniel Williams,[9] Mikaeel Yunus[18]

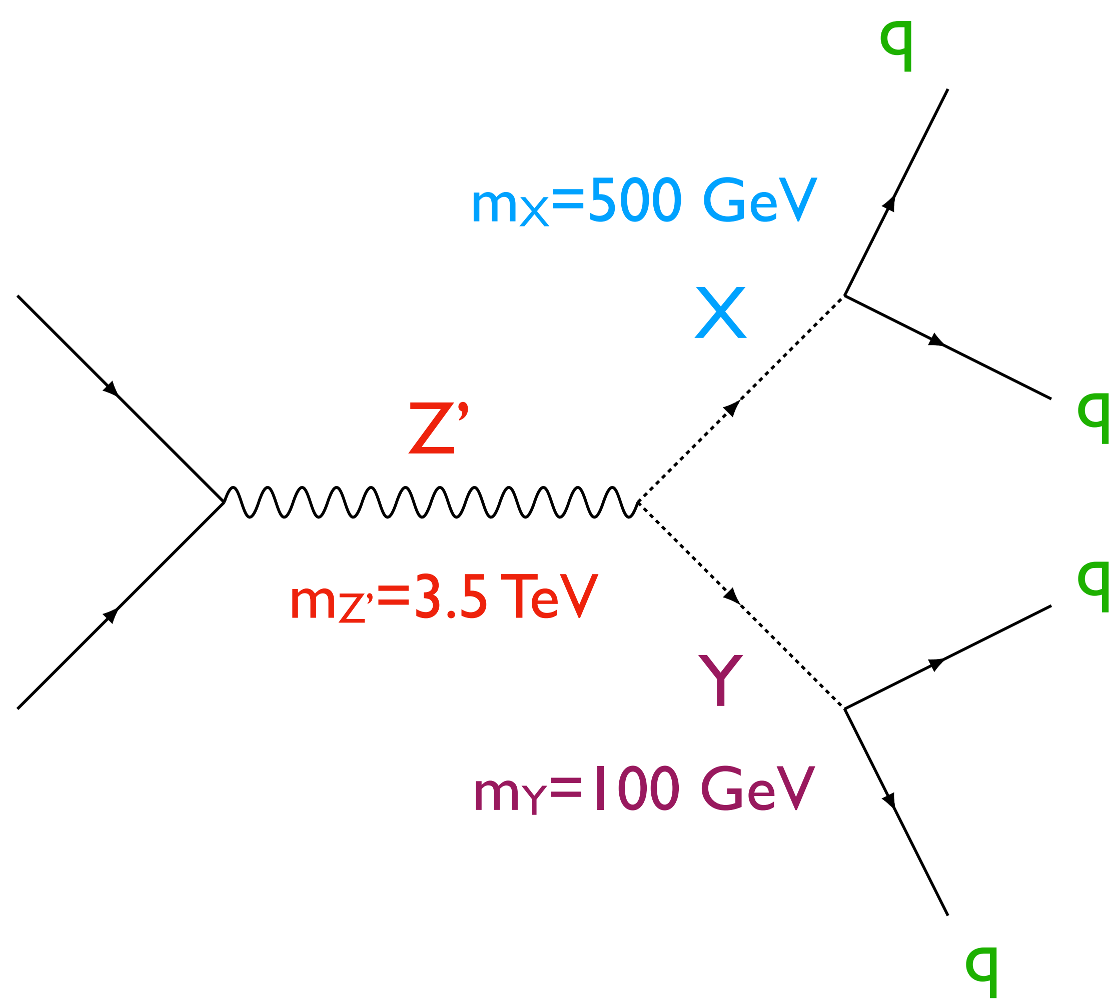
arxiv: 2101.08320

Many new approaches inspired by **LHCO2020**

LHC Olympics 2020: R&D Dataset

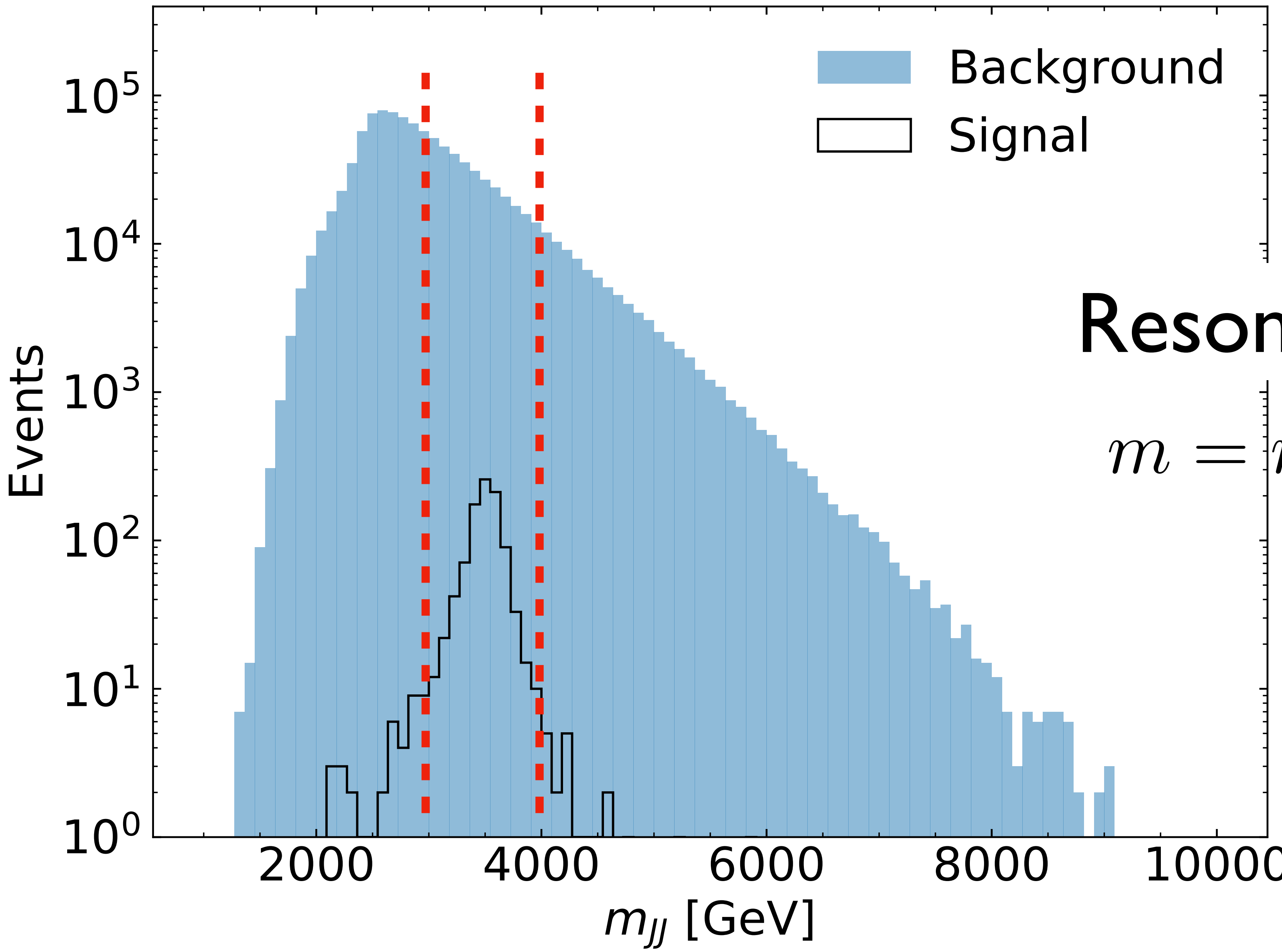
https://doi.org/10.5281/zenodo.2629072

Prior to the challenge, we also released a labeled R&D dataset consisting of 1M QCD dijet events and 100k signal events



No explicit search at the LHC for this scenario!

LHC Olympics 2020: R&D Dataset



Resonant feature

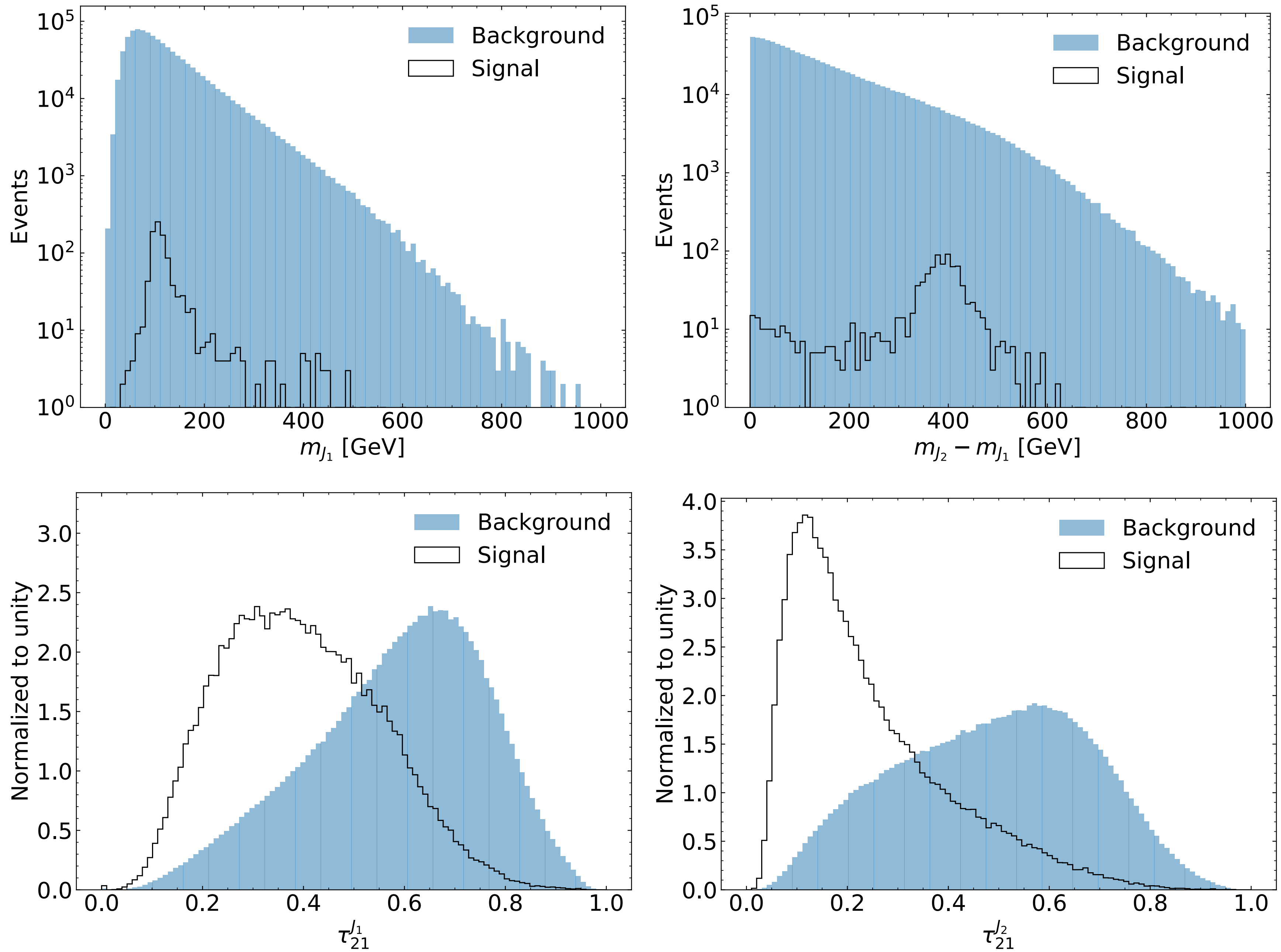
$$m = m_{Z'} = m_{JJ}$$

Benchmark signal strength:

S=500, B=500,000, B_{SR}=61,000

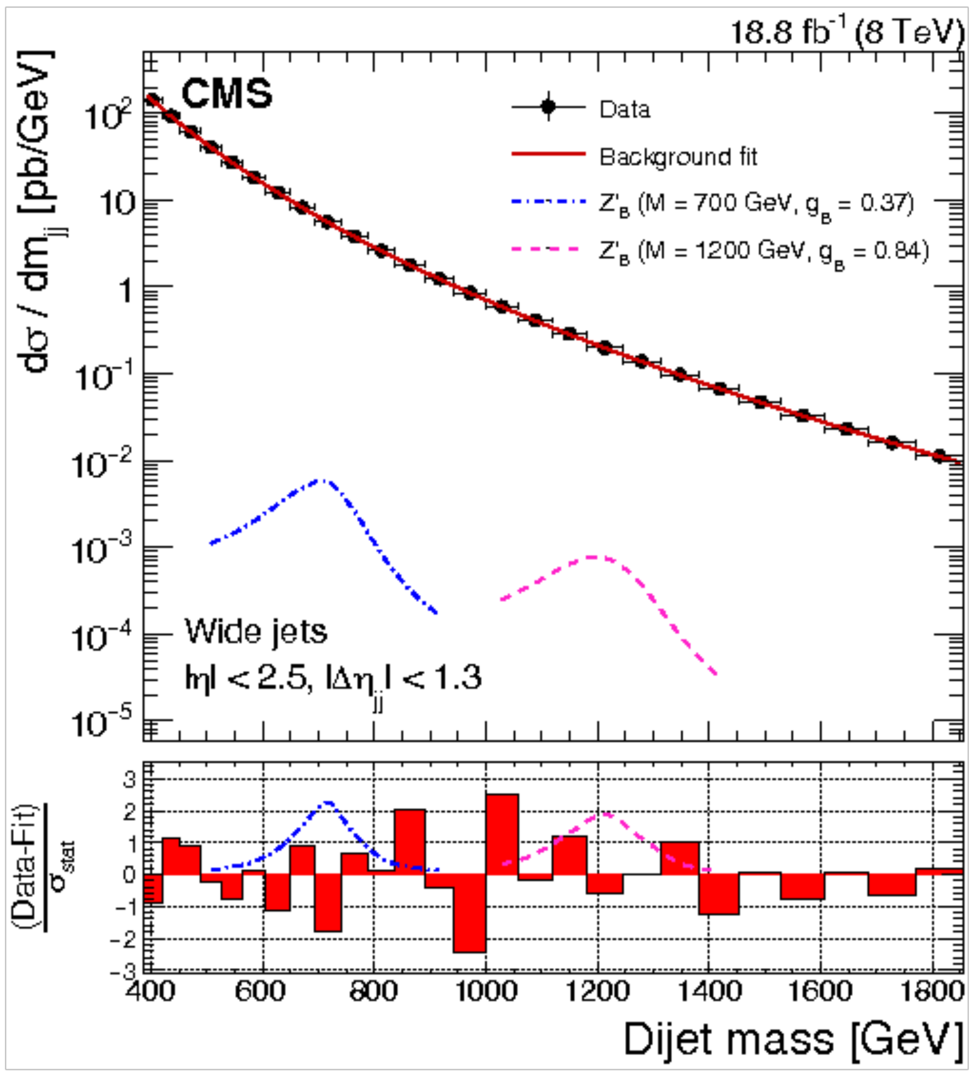
$S/B_{SR} \sim 6\times10^{-3}$, $S/\sqrt{B_{SR}} \sim 1.5$

LHC Olympics 2020: R&D Dataset

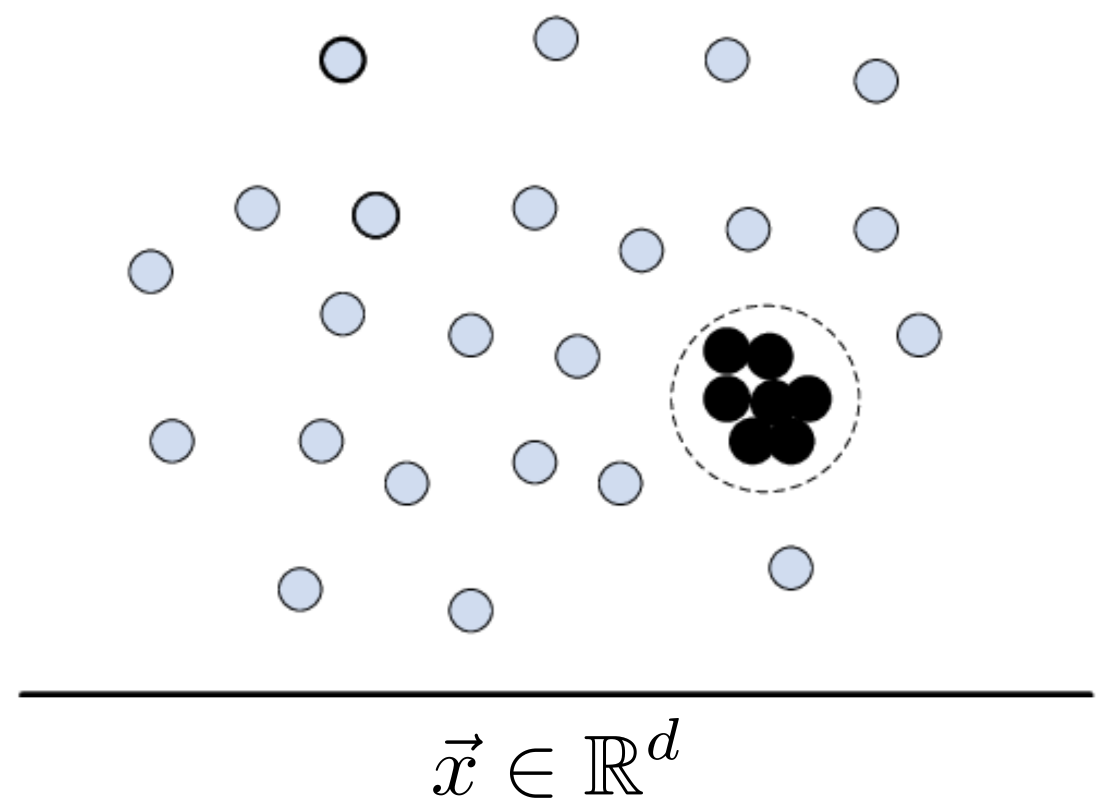


Additional features: $x = (m_{J_1}, m_{J_2}, \tau_{21}^{J_1}, \tau_{21}^{J_2})$

Enhancing the bump hunt



primary resonant feature (mJJ)



additional features

Q: If the signal is localized in additional features, can we find it in a model-independent way?

A general optimal strategy

Claim: the optimal model-agnostic discriminant would be (Neyman & Pearson)

$$R(x) = \frac{P_{data}(x)}{P_{bg}(x)}$$

"Idealized Anomaly Detector"

A general optimal strategy

Claim: the optimal model-agnostic discriminant would be (Neyman & Pearson)

$$R(x) = \frac{P_{data}(x)}{P_{bg}(x)}$$

"Idealized Anomaly Detector"

Proof:

A general optimal strategy

Claim: the optimal model-agnostic discriminant would be (Neyman & Pearson)

$$R(x) = \frac{P_{data}(x)}{P_{bg}(x)}$$

"Idealized Anomaly Detector"

Proof:

$$P_{data}(x) = \epsilon_s P_{sig}(x) + (1 - \epsilon_s) P_{bg}(x)$$

A general optimal strategy

Claim: the optimal model-agnostic discriminant would be (Neyman & Pearson)

$$R(x) = \frac{P_{data}(x)}{P_{bg}(x)}$$

"Idealized Anomaly Detector"

Proof:

$$P_{data}(x) = \epsilon_s P_{sig}(x) + (1 - \epsilon_s) P_{bg}(x)$$

$$R(x) = (1 - \epsilon_s) + \epsilon_s \frac{P_{sig}(x)}{P_{bg}(x)}$$

A general optimal strategy

Claim: the optimal model-agnostic discriminant would be (Neyman & Pearson)

$$R(x) = \frac{P_{data}(x)}{P_{bg}(x)}$$

"Idealized Anomaly Detector"

Proof:

$$P_{data}(x) = \epsilon_s P_{sig}(x) + (1 - \epsilon_s) P_{bg}(x)$$

$$R(x) = (1 - \epsilon_s) + \epsilon_s \frac{P_{sig}(x)}{P_{bg}(x)}$$

So R(x) is monotonic with signal-to-background likelihood ratio
regardless of unknown, arbitrary signal strength and probability density

Idea: data vs MC classifier

D'Agnolo, Wulzer et al (1806.02350, 1912.12155)

Train a neural network to classify data vs MC simulation of the SM.

If the NN classifier is optimal, its output should be (monotonic with) the data vs MC likelihood ratio (Neyman-Pearson).

Idea: data vs MC classifier

D'Agnolo, Wulzer et al (1806.02350, 1912.12155)

Train a neural network to classify data vs MC simulation of the SM.

If the NN classifier is optimal, its output should be (monotonic with) the data vs MC likelihood ratio (Neyman-Pearson).

$$R_{classifier}(x) \approx \frac{P_{data}(x)}{P_{MC}(x)}$$

"The likelihood-ratio trick"

Idea: data vs MC classifier

D'Agnolo, Wulzer et al (1806.02350, 1912.12155)

Train a neural network to classify data vs MC simulation of the SM.

If the NN classifier is optimal, its output should be (monotonic with) the data vs MC likelihood ratio (Neyman-Pearson).

$$R_{classifier}(x) \approx \frac{P_{data}(x)}{P_{MC}(x)}$$

"The likelihood-ratio trick"

The modern ML version of the "general search"

Idea: data vs MC classifier

D'Agnolo, Wulzer et al (1806.02350, 1912.12155)

Train a neural network to classify data vs MC simulation of the SM.

If the NN classifier is optimal, its output should be (monotonic with) the data vs MC likelihood ratio (Neyman-Pearson).

$$R_{classifier}(x) \approx \frac{P_{data}(x)}{P_{MC}(x)}$$

"The likelihood-ratio trick"

The modern ML version of the "general search"

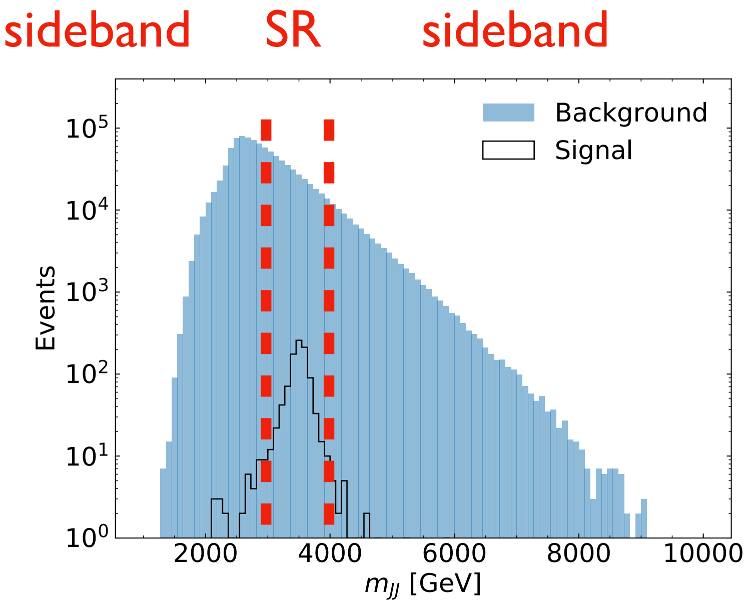
But generally we cannot rely on MC simulations for accurate SM background predictions…

How to get $P_{data}(x)$ and $P_{bg}(x)$ in a data-driven way?

Idea: data vs sideband classifier

Collins, Howe & Nachman 1805.02664, 1902.02634

"CWoLa Hunting"

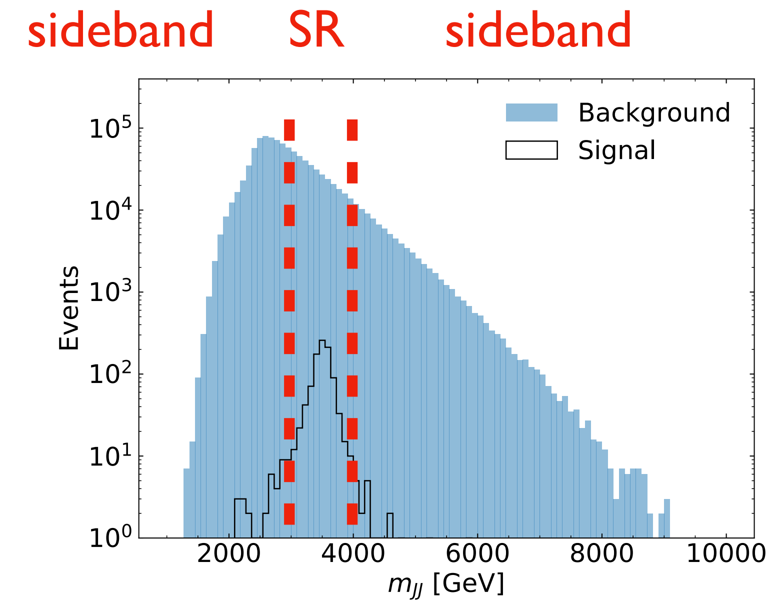


Key idea: use signal region and sidebands to learn $P_{data}(x)$ and $P_{bg}(x)$!

Idea: data vs sideband classifier

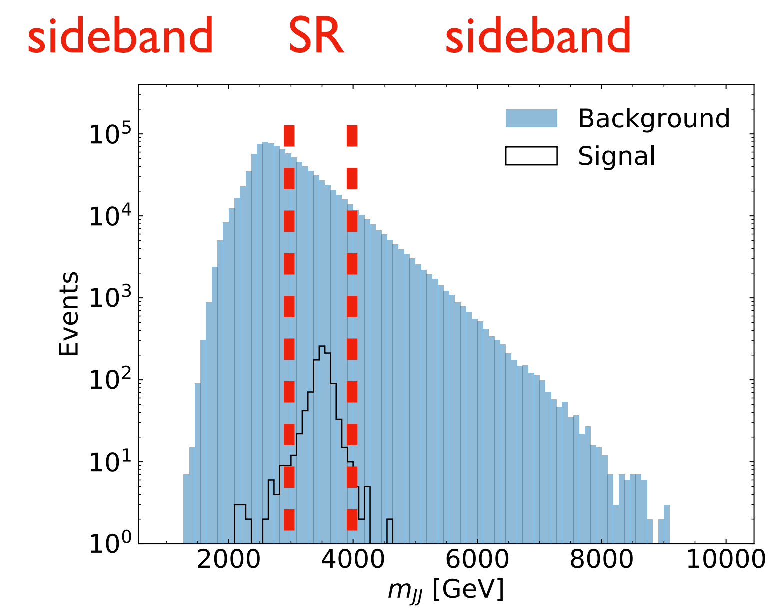
Collins, Howe & Nachman 1805.02664,1902.02634

“CWoLa Hunting”



Key idea: use signal region and sidebands to learn $P_{data}(x)$ and $P_{bg}(x)$!

Train a NN classifier on SR vs SB data, learn



Idea: data vs sideband classifier

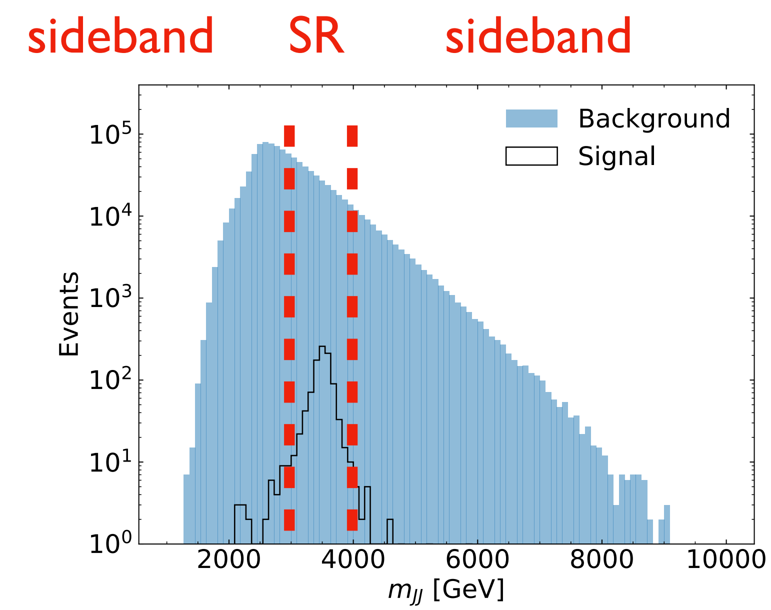
Collins, Howe & Nachman 1805.02664,1902.02634

"CWoLa Hunting"

Key idea: use signal region and sidebands to learn $P_{data}(x)$ and $P_{bg}(x)$!

Train a NN classifier on SR vs SB data, learn

$$R_{classifier}(x) \approx \frac{P_{data}(x\,|\,m \in SR)}{P_{data}(x\,|\,m \in SB)} = \frac{P_{data}(x\,|\,m \in SR)}{P_{bg}(x\,|\,m \in SB)}$$



Idea: data vs sideband classifier

Collins, Howe & Nachman 1805.02664,1902.02634

"CWoLa Hunting"

Key idea: use signal region and sidebands to learn $P_{data}(x)$ and $P_{bg}(x)$!

Train a NN classifier on SR vs SB data, learn

$$R_{classifier}(x) \approx \frac{P_{data}(x\,|\,m \in SR)}{P_{data}(x\,|\,m \in SB)} = \frac{P_{data}(x\,|\,m \in SR)}{P_{bg}(x\,|\,m \in SB)}$$

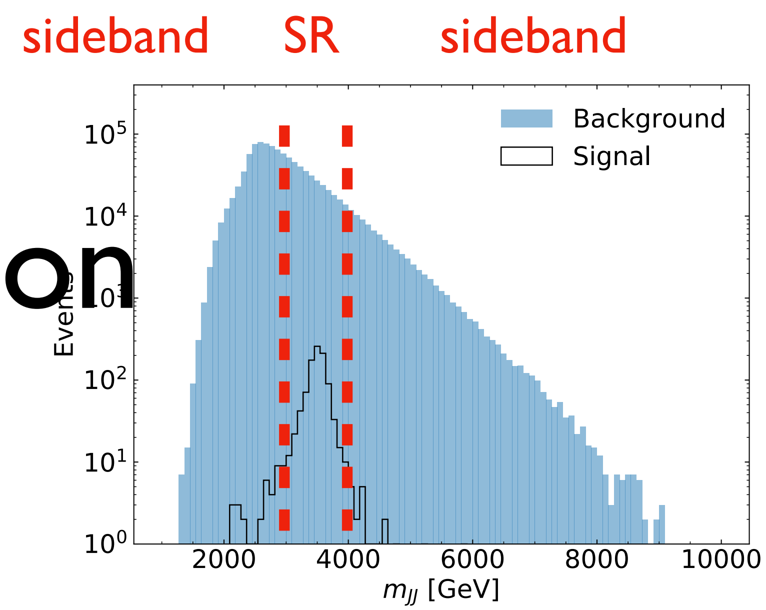
If bg x in SB statistically identical to bg x in SR (x and mJJ are uncorrelated in the bg), then $P_{bg}(x\,|\,m \in SB) = P_{bg}(x\,|\,m \in SR)$

and the classifier gives the desired likelihood ratio.

Idea: conditional density estimation

Nachman & DS 2001.04990

"ANODE"



Idea: use *neural density estimation* to learn numerator and denominator separately.

Train separate ***masked autoregressive flows (MAFs)*** on SR and SB events to learn $P_{data}(x\,|\,m \in SR)$ and $P_{data}(x\,|\,m \in SB) = P_{bg}(x\,|\,m \in SB)$ directly from the data.

The sideband MAF automatically interpolates into the SR, giving an estimate of $P_{bg}(x\,|\,m \in SR)$.

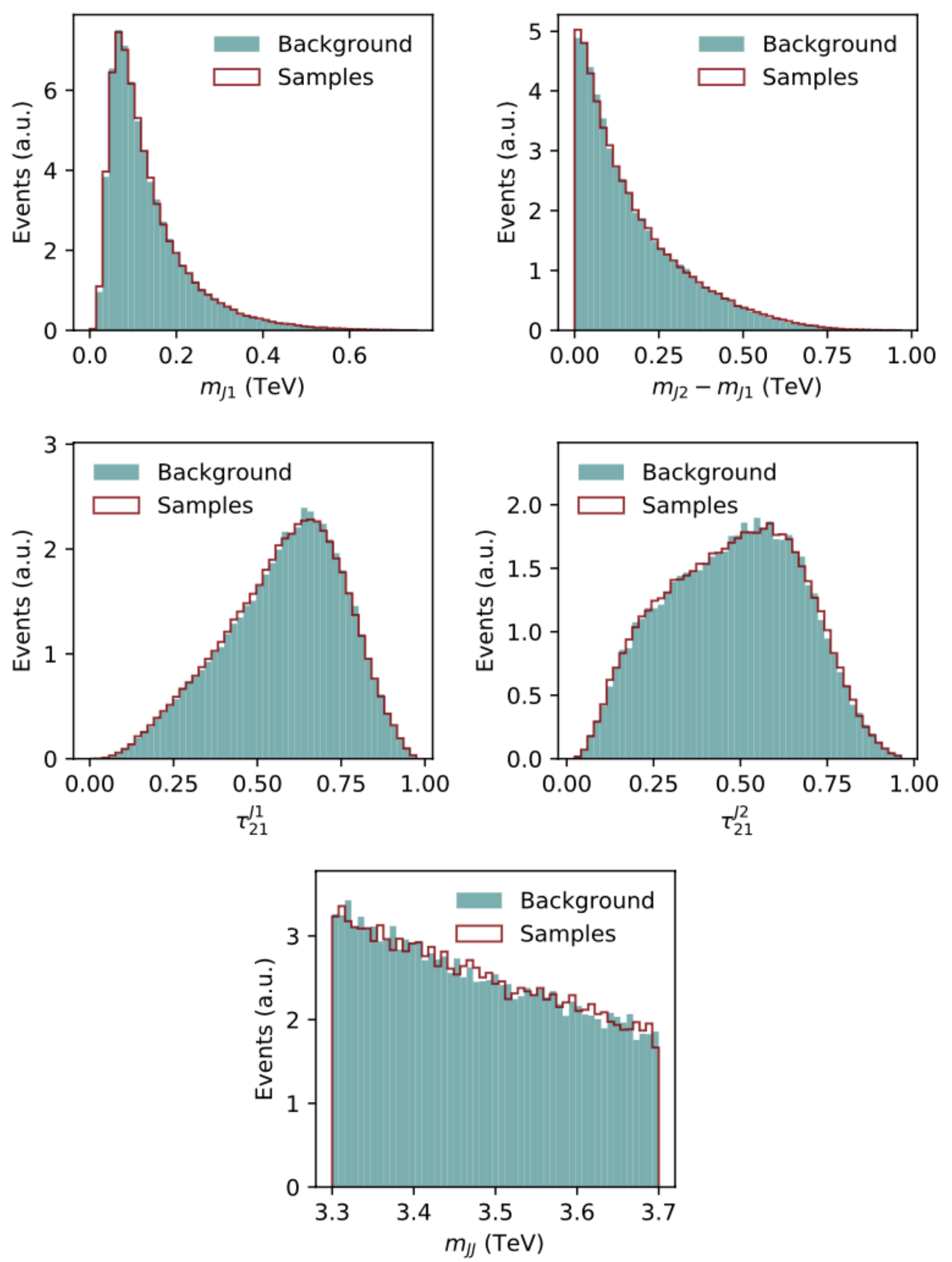
Pros: robust against correlations!

Cons: density estimation much harder than classification

Idea: density estimation+classification

Hallin, Isaacson, Kasieczka, Krause, Nachman, Quadfasel, Schlaffer, DS & Sommerhalder (2109.00546)

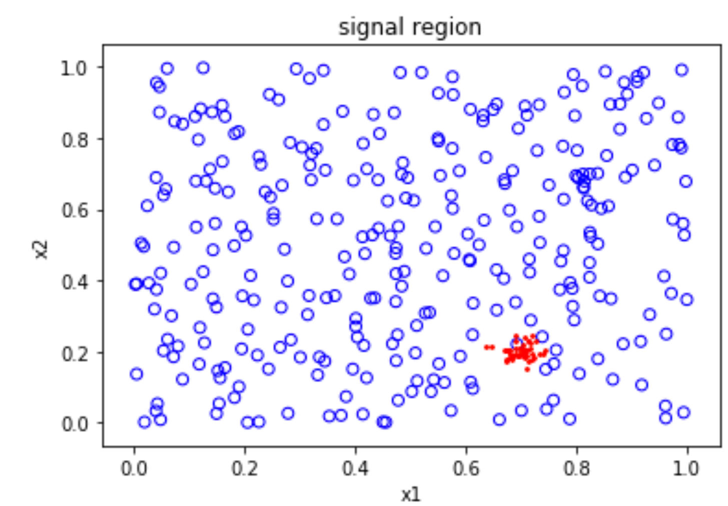
"CATHODE"



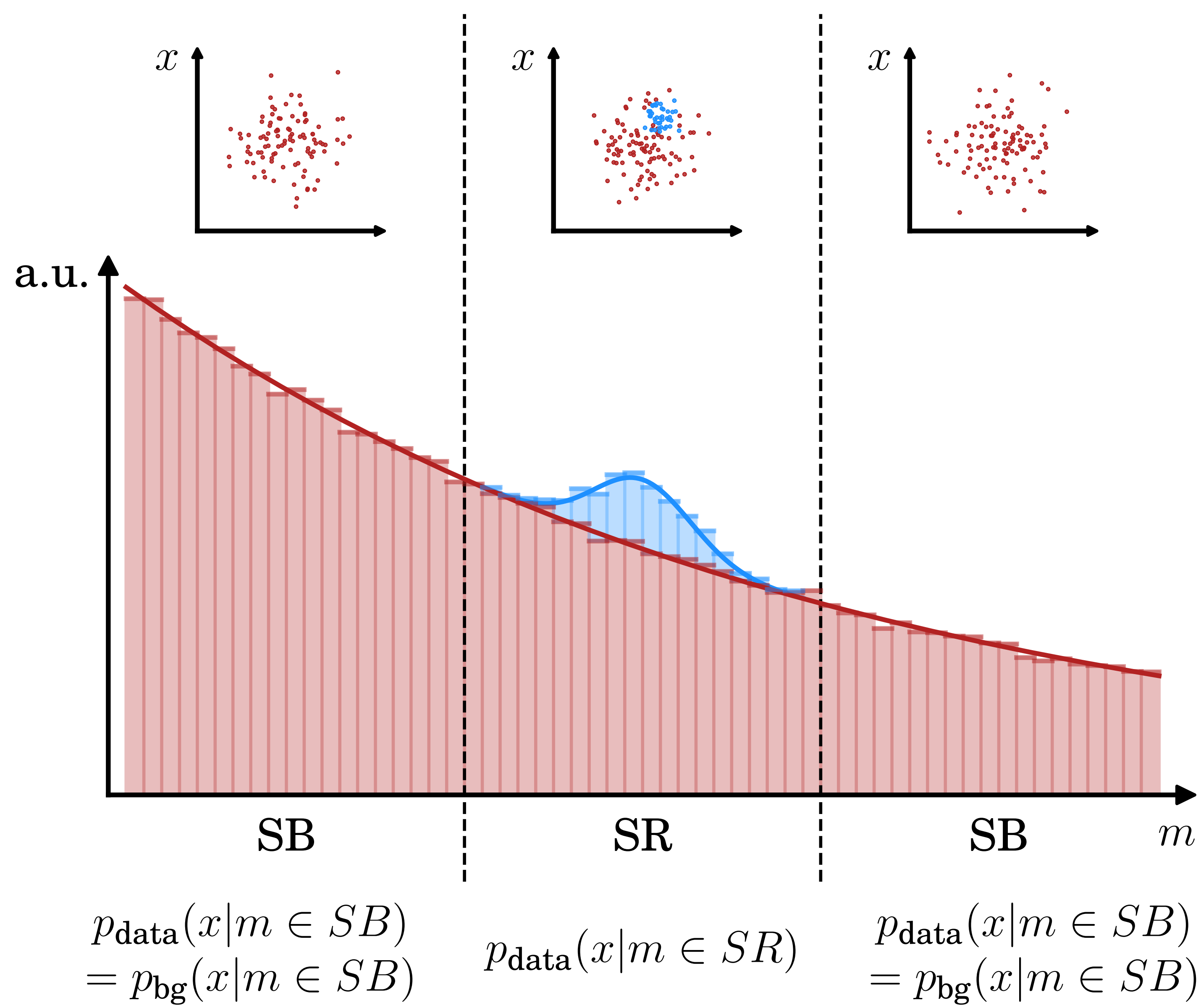
We realized: can also sample from trained density estimator.

So why not sample background events from the interpolated outer MAF and train a fully data-driven data-vs-background classifier?!

Pros: Robust against correlations *and* don't have to learn density estimator for $P_{data}(x \mid m \in SR)$



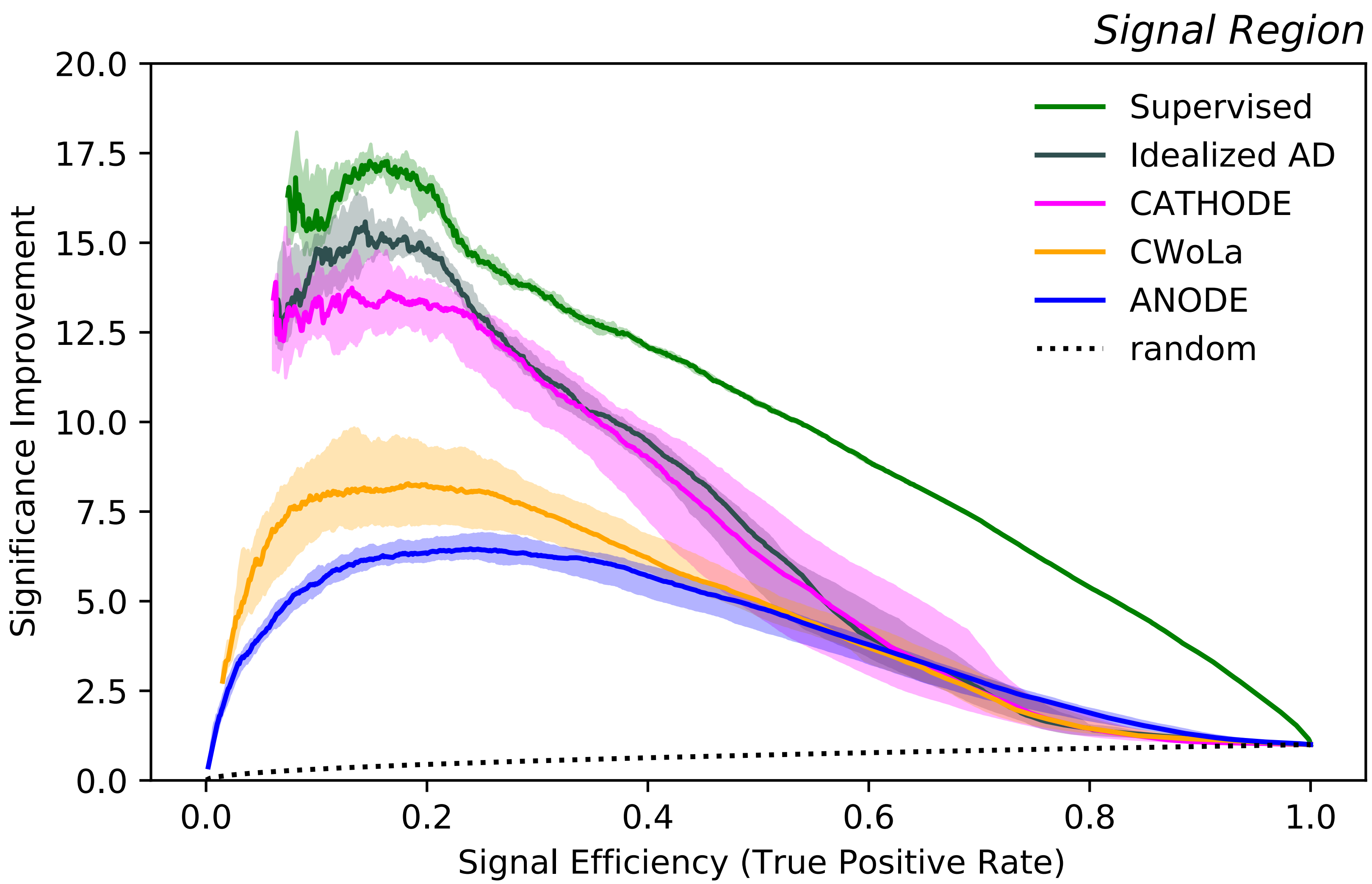
Summary of methods



- CWoLa Hunting: **classifier** between SB and SR data
- ANODE: two **conditional density estimators** on SB and SR data; interpolate SB density estimator into SR
- CATHODE: **conditional density estimator** on SB data; sample interpolated SB density estimator in SR; **classifier** between sampled events and data in SR

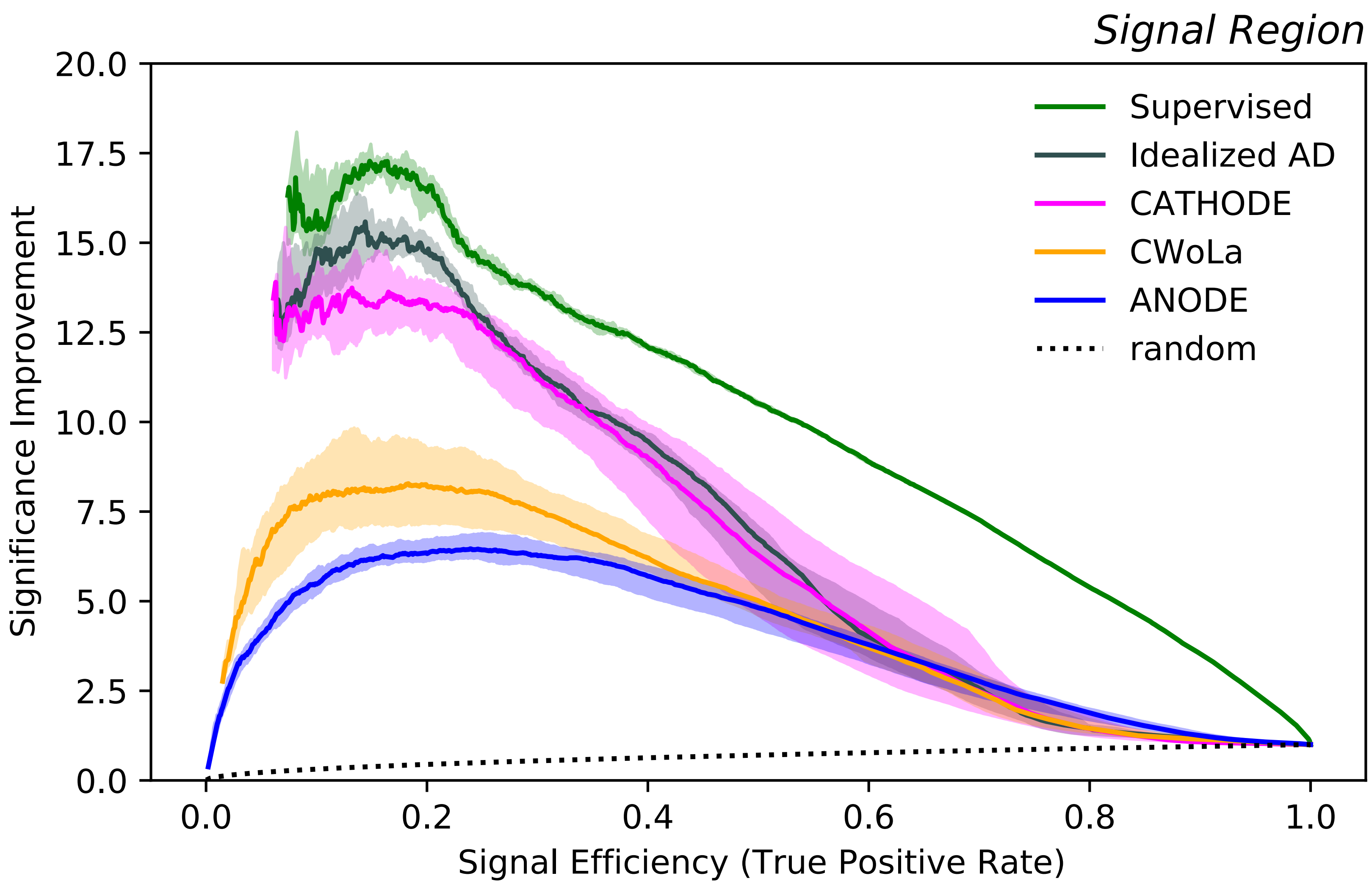
Comparison of methods

Significance improvement characteristic (SIC): $\epsilon_S/\sqrt{\epsilon_B}$



Comparison of methods

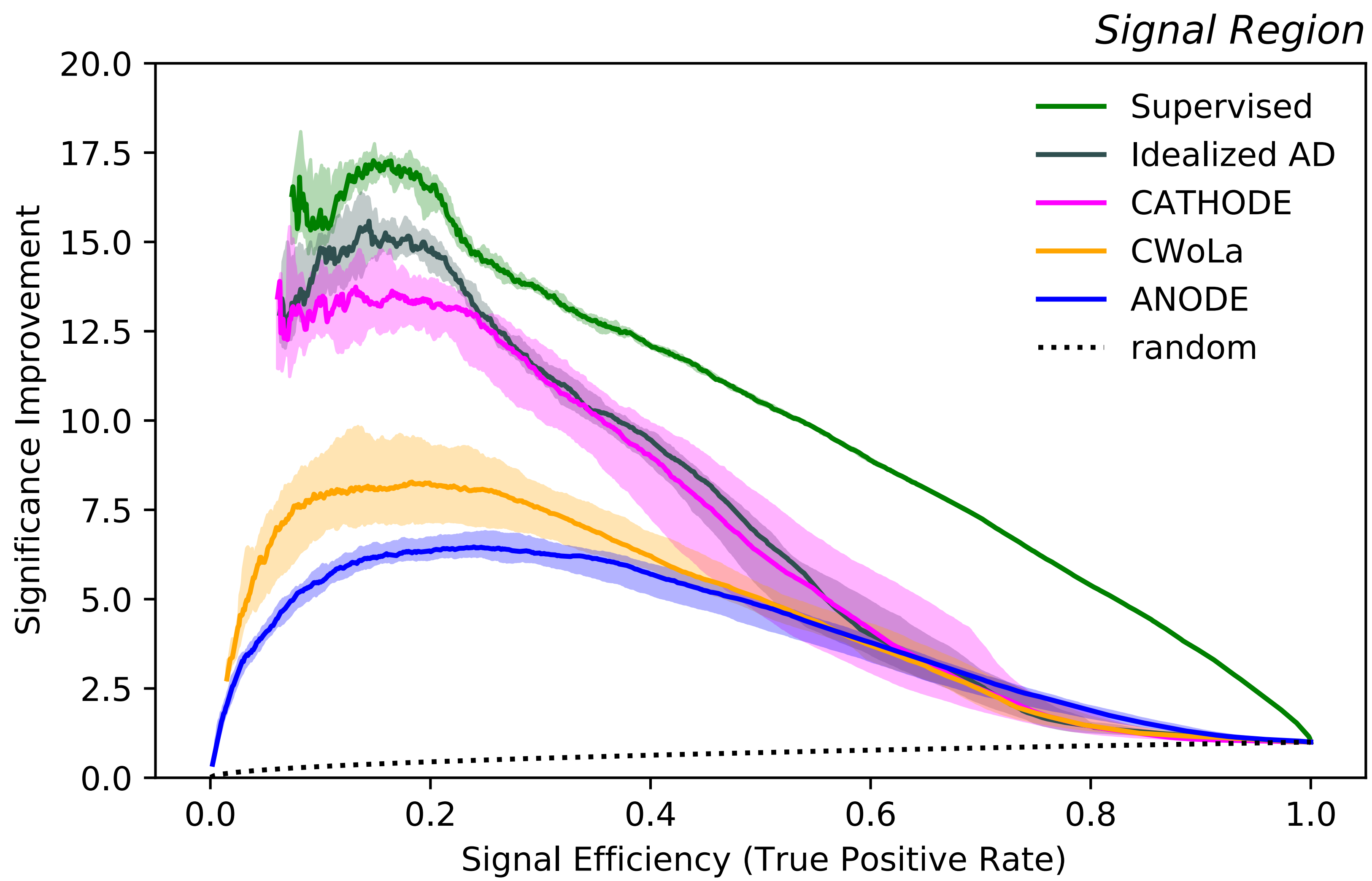
Significance improvement characteristic (SIC): $\epsilon_S/\sqrt{\epsilon_B}$



CATHODE outperforms CWoLa and ANODE and nearly saturates the idealized anomaly detector!

Comparison of methods

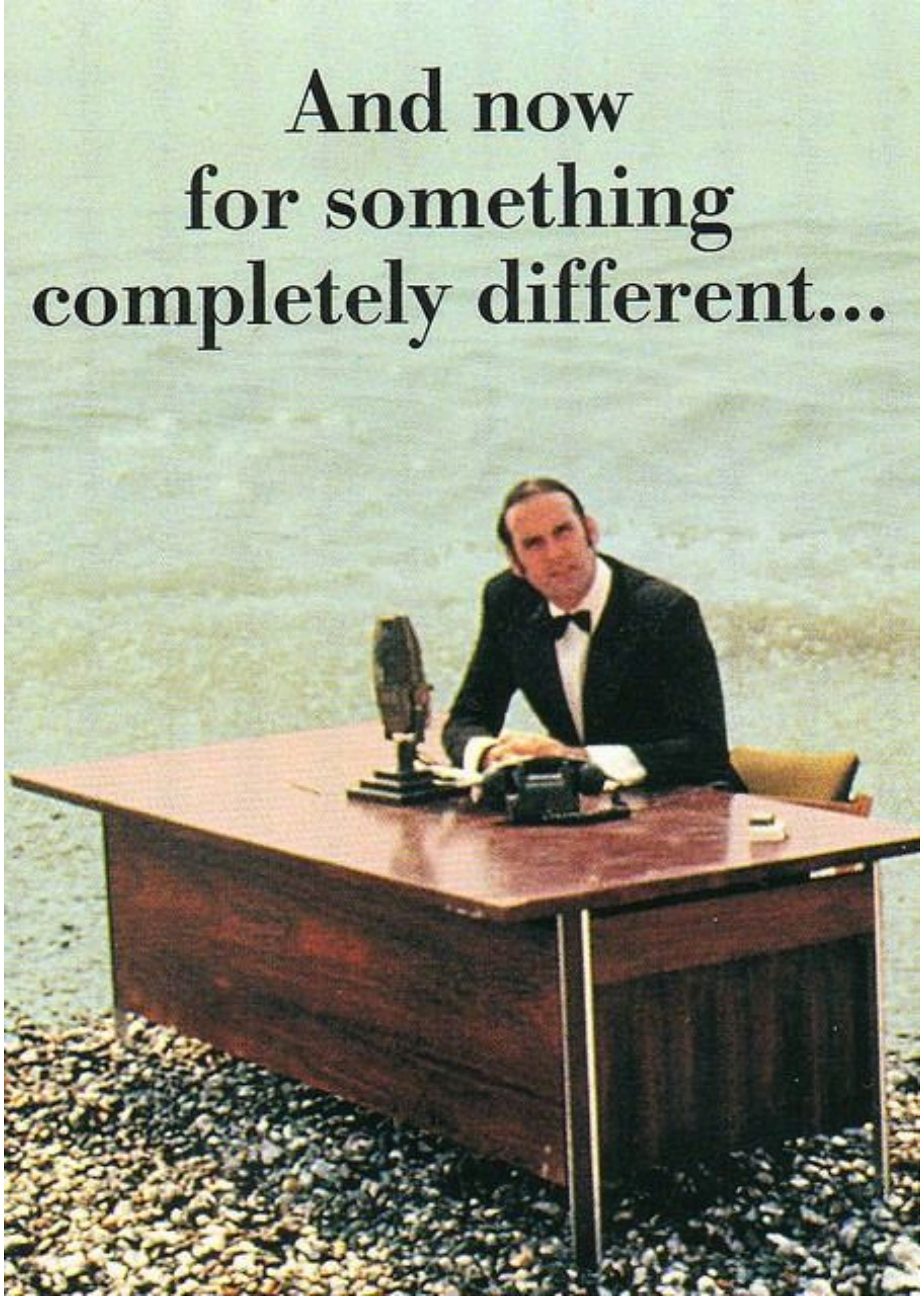
Significance improvement characteristic (SIC): $\epsilon_S / \sqrt{\epsilon_B}$



CATHODE outperforms CWoLa and ANODE and nearly saturates the idealized anomaly detector!

Initial significance was ~1.5σ

==> <u>**a ~30σ anomaly could be hiding in the data right now**</u>

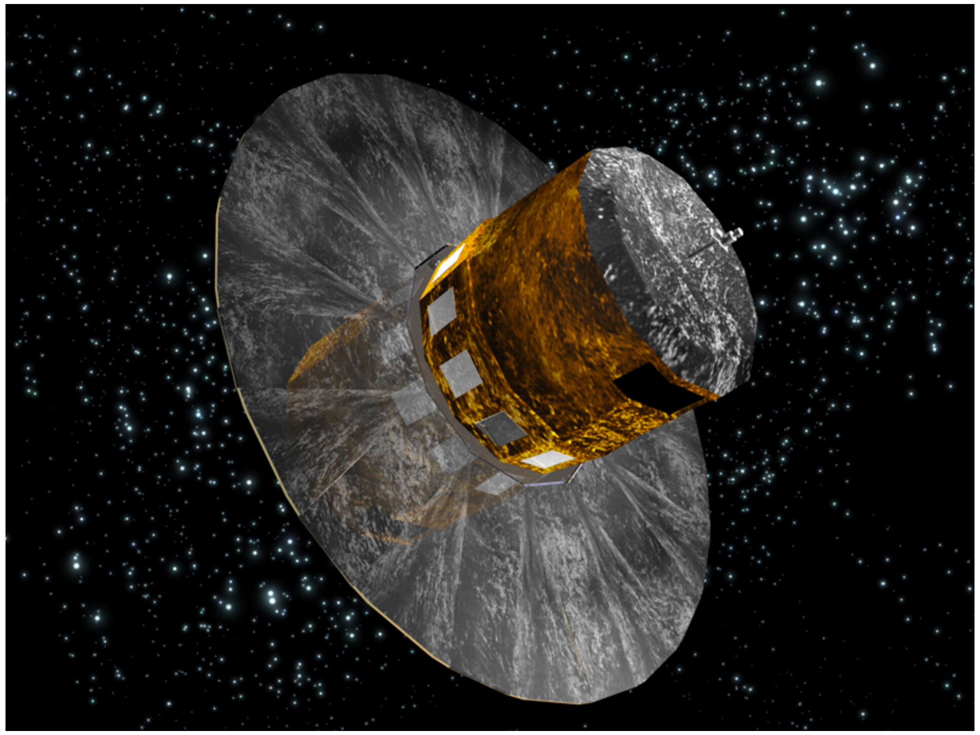

And now
for something
completely different...

Via Machinae: ANODE IN SPACE

DS, Buckley, Necib & Tamanas 2104.12789

ANODE is a completely general method for identifying multivariate overdensities in data using sidebands. It could have many diverse applications.

We are currently using it to find **stellar streams** in Gaia data.

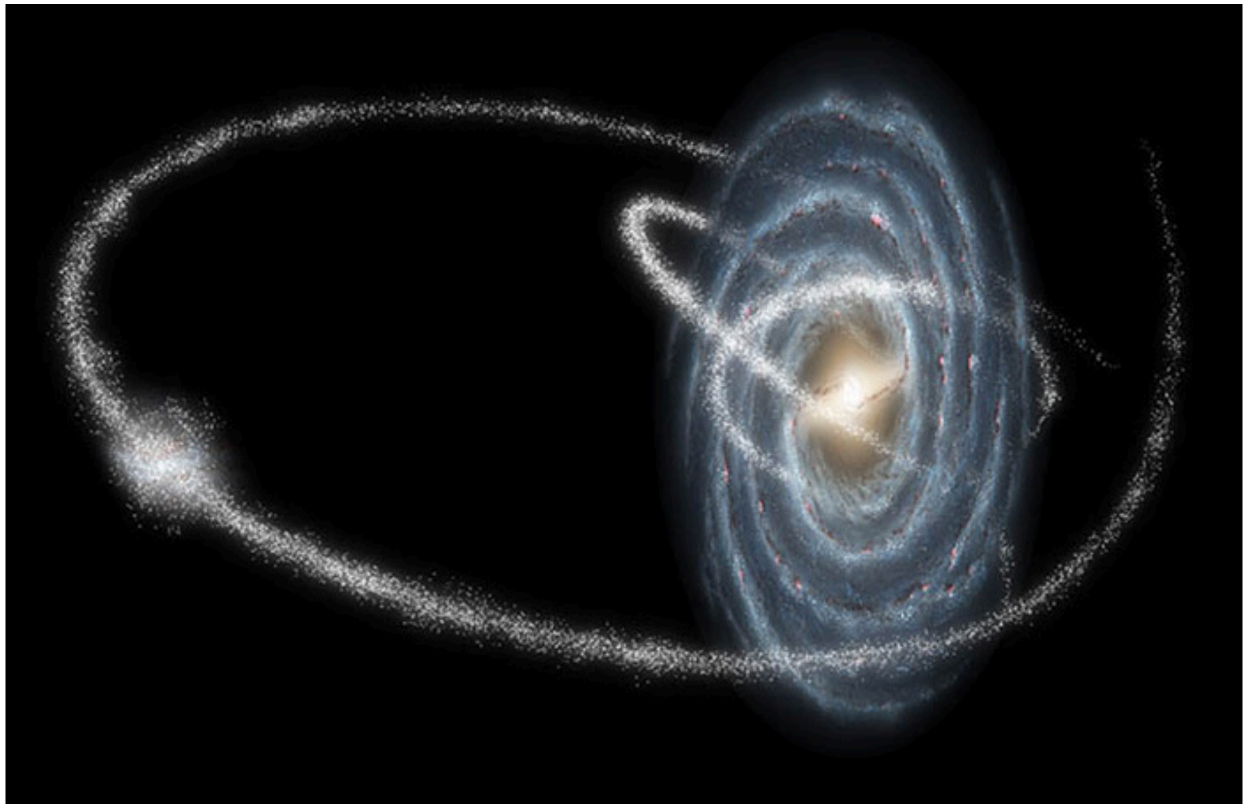


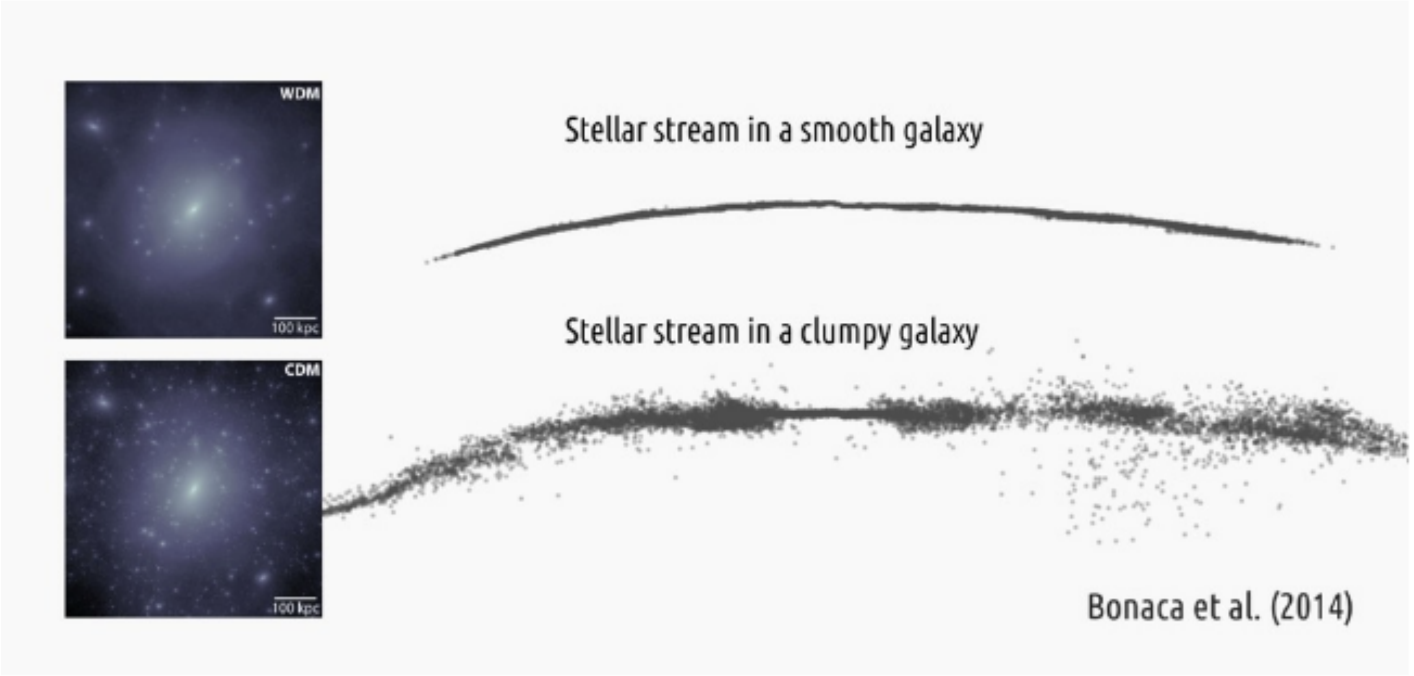
Gaia satellite:

- Launched in 2013; extended to 2025
- Mission: map out the full 6d phase space of the stars in our galaxy
- Angular positions, velocities, color and magnitude of over 1 billion stars in our galaxy
- radial positions and velocities for a smaller subset of nearby stars (not used in this work)

Stellar streams

Cold stellar streams are tidally-stripped remnants of globular clusters and dwarf galaxies, falling into and orbiting our galaxy.



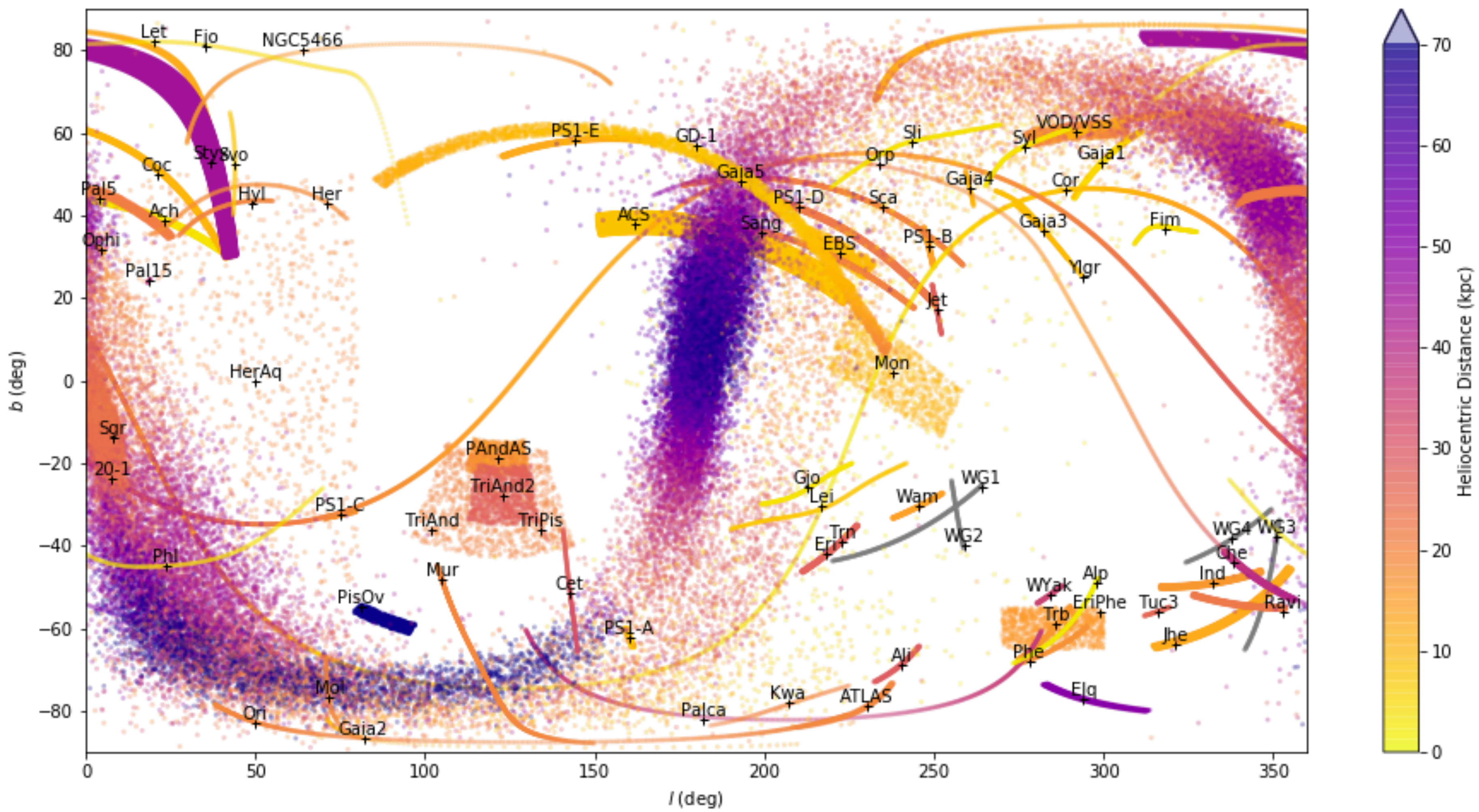


They are very interesting objects of study for astrophysicists and particle physicists.

In particular, they could be unique probes into dark matter substructure.

Stream finding

https://github.com/cmateu/galstreams

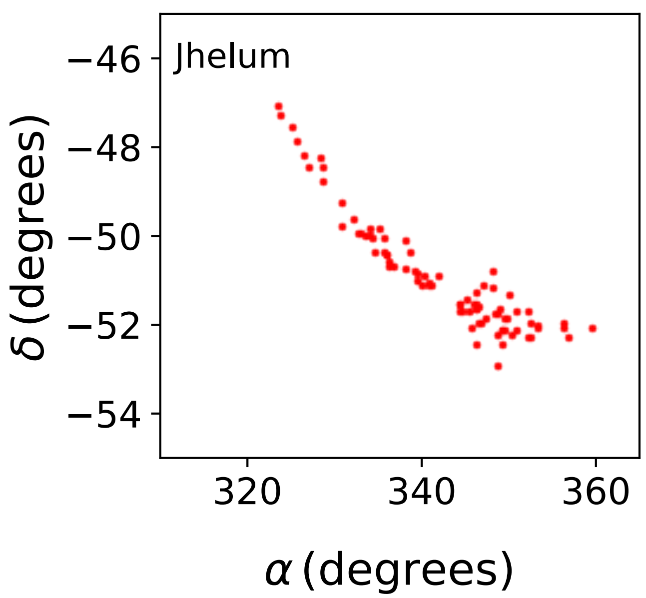


Existing methods (eg STREAMFINDER, Malhan & Ibata 2018) have found many new streams in the Gaia data, but they make a number of model-dependent assumptions (form of the galactic potential, orbits, isochrones, …).

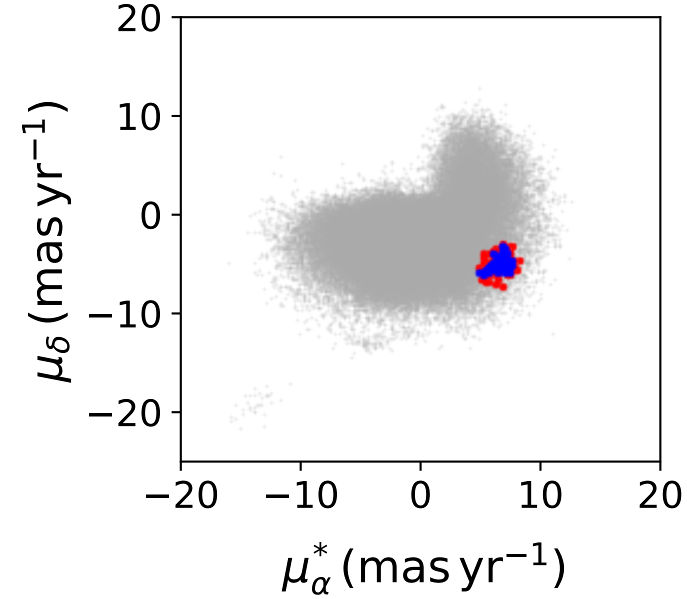
We were interested in whether unsupervised ML could be used to find streams more model-independently.

Stream finding

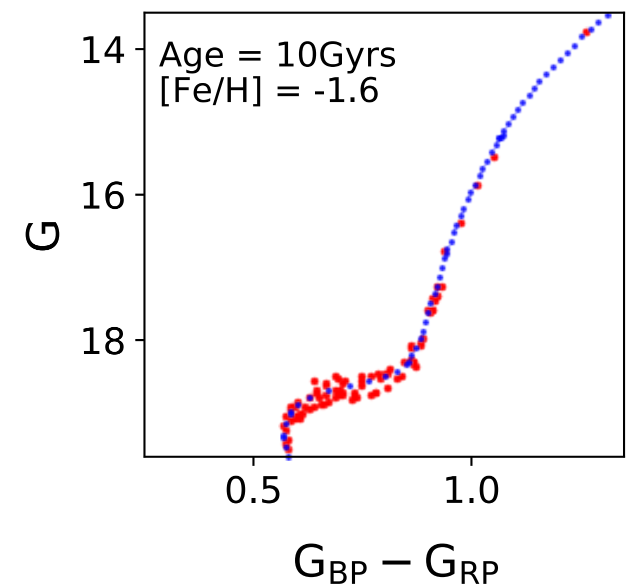
from Malhan et al 2018



position



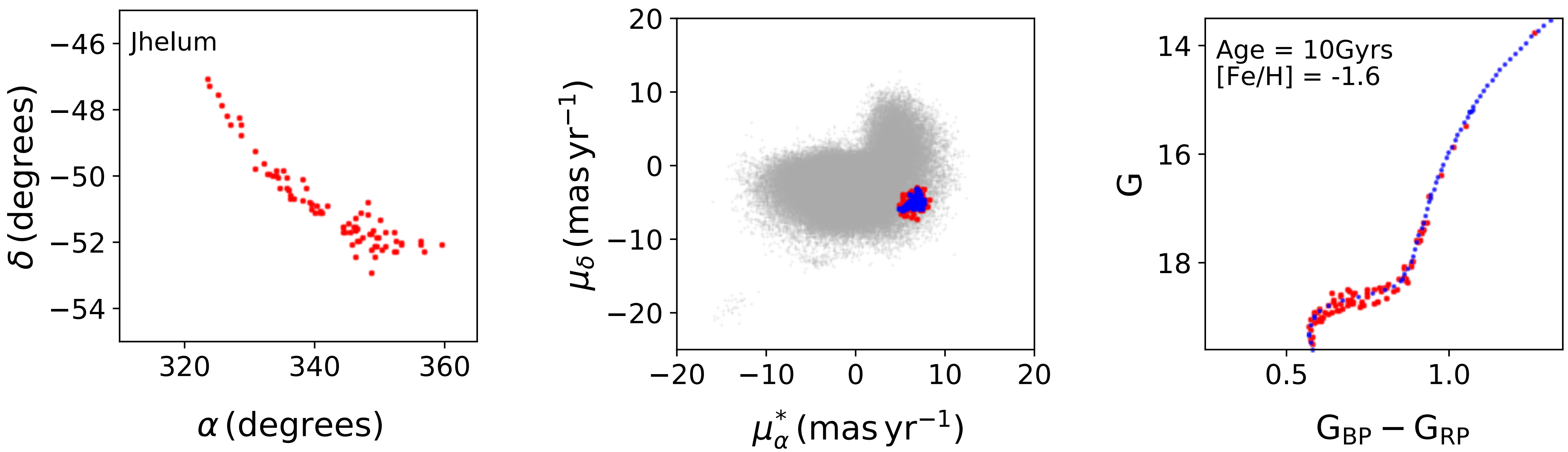
velocity



photometry

Stream finding

from Malhan et al 2018

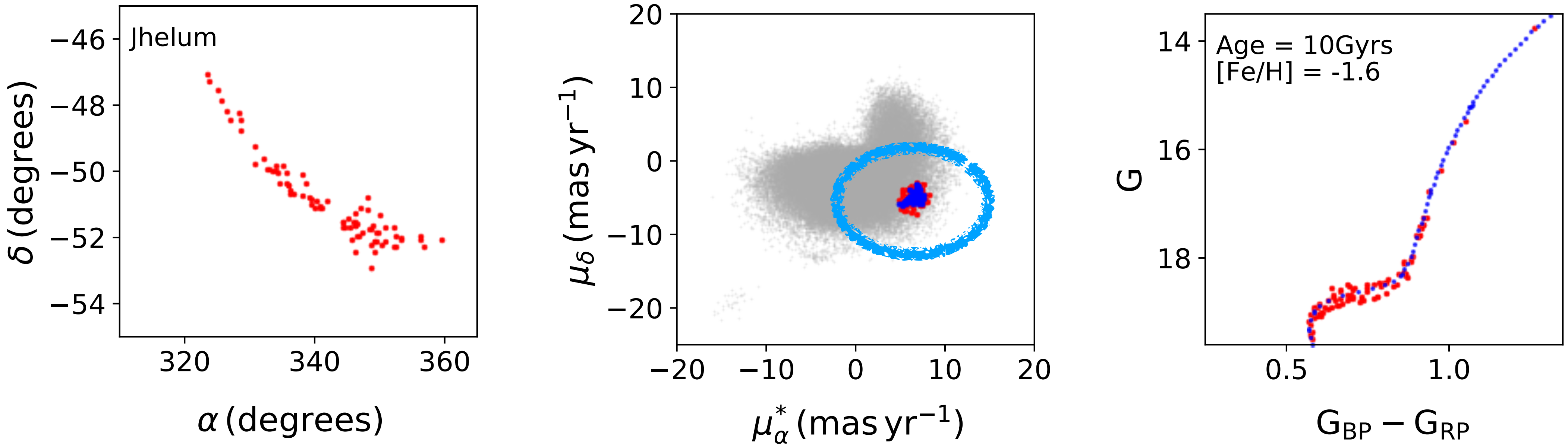


position velocity photometry

Idea: streams are local overdensities in position, velocity and photometric space.

Stream finding

from Malhan et al 2018



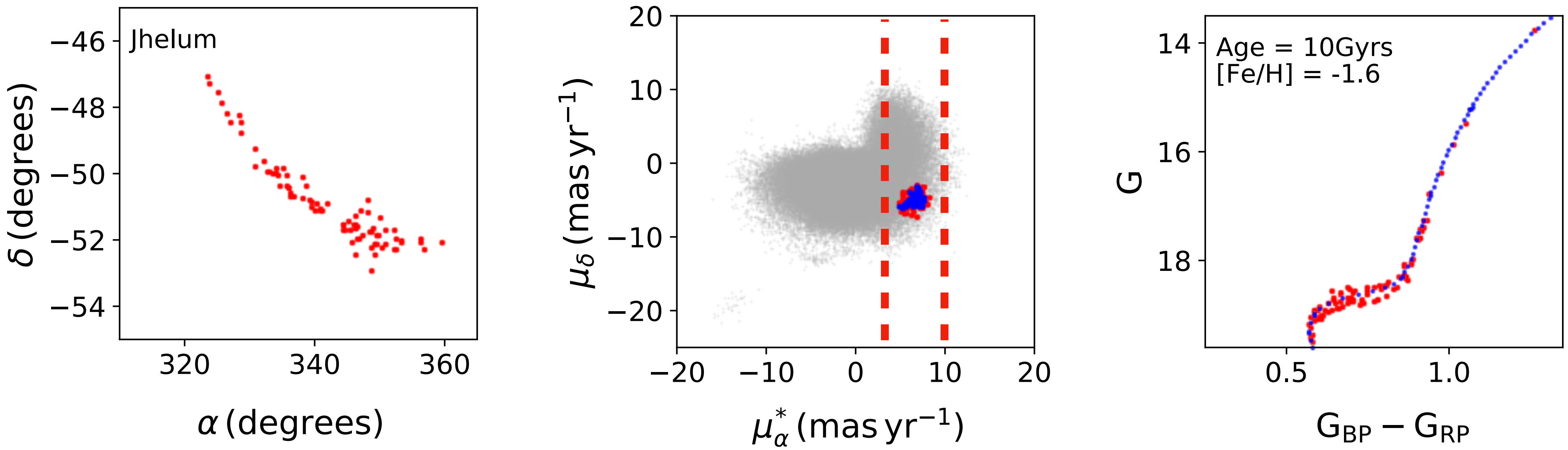
position velocity photometry

Idea: streams are local overdensities in position, velocity and photometric space.

Since they are **cold**, the stars in the stream are clustered in velocity.

Stream finding

from Malhan et al 2018



position velocity photometry

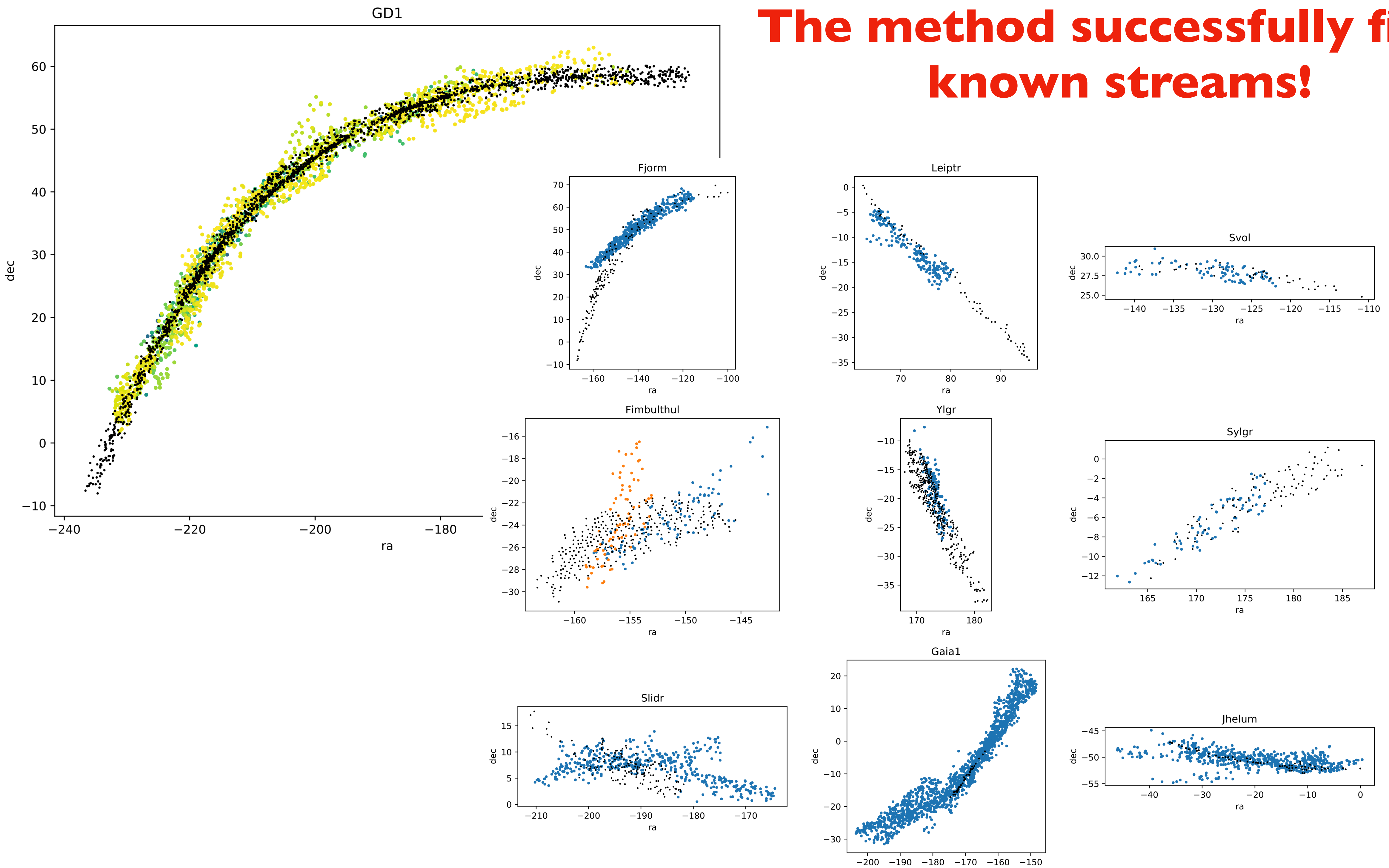
Idea: streams are local overdensities in position, velocity and photometric space.

Since they are **cold**, the stars in the stream are clustered in velocity.

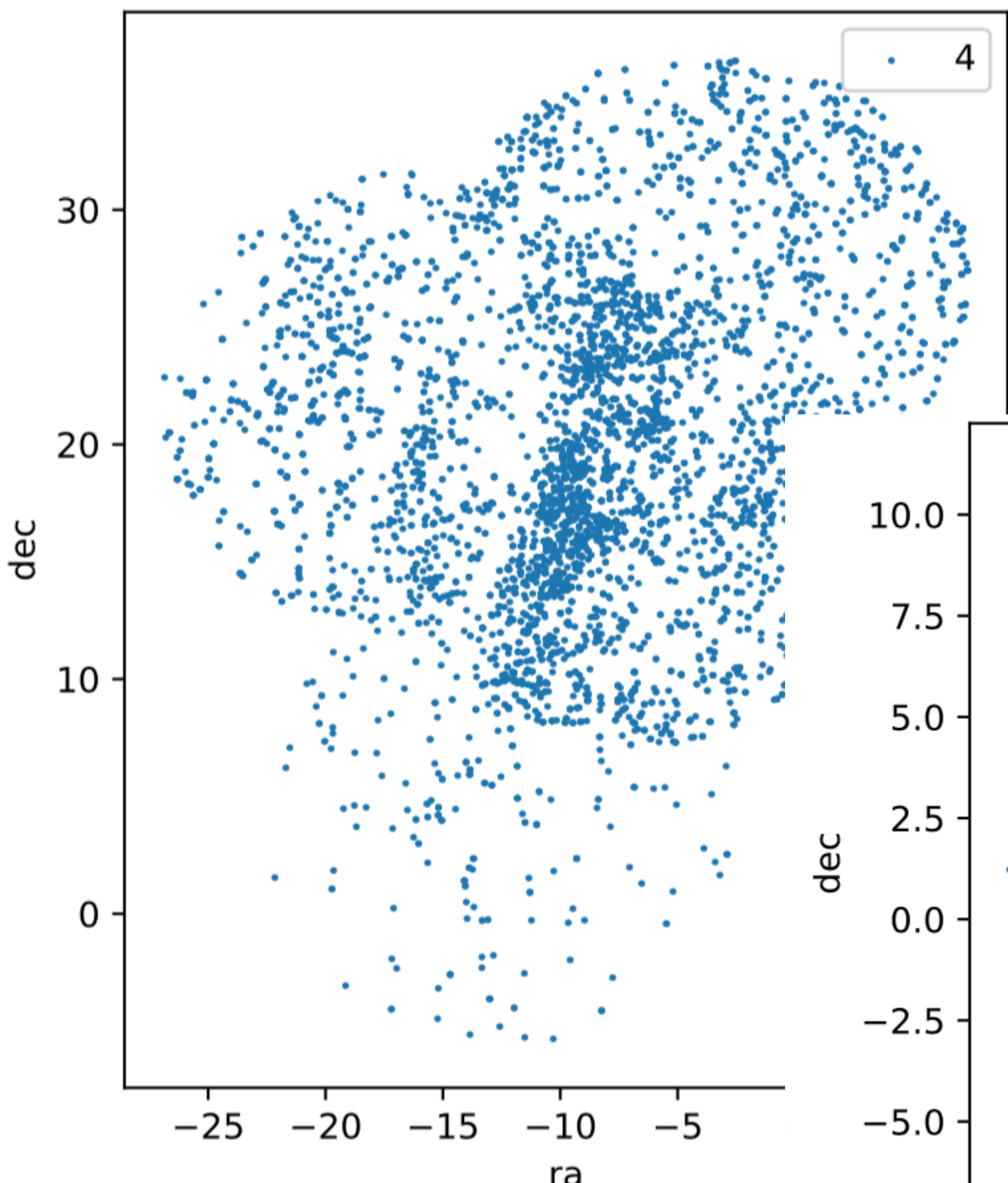
Can sideband in one of the velocities and use ANODE to look for local overdensities!

Via Machinae: preliminary results

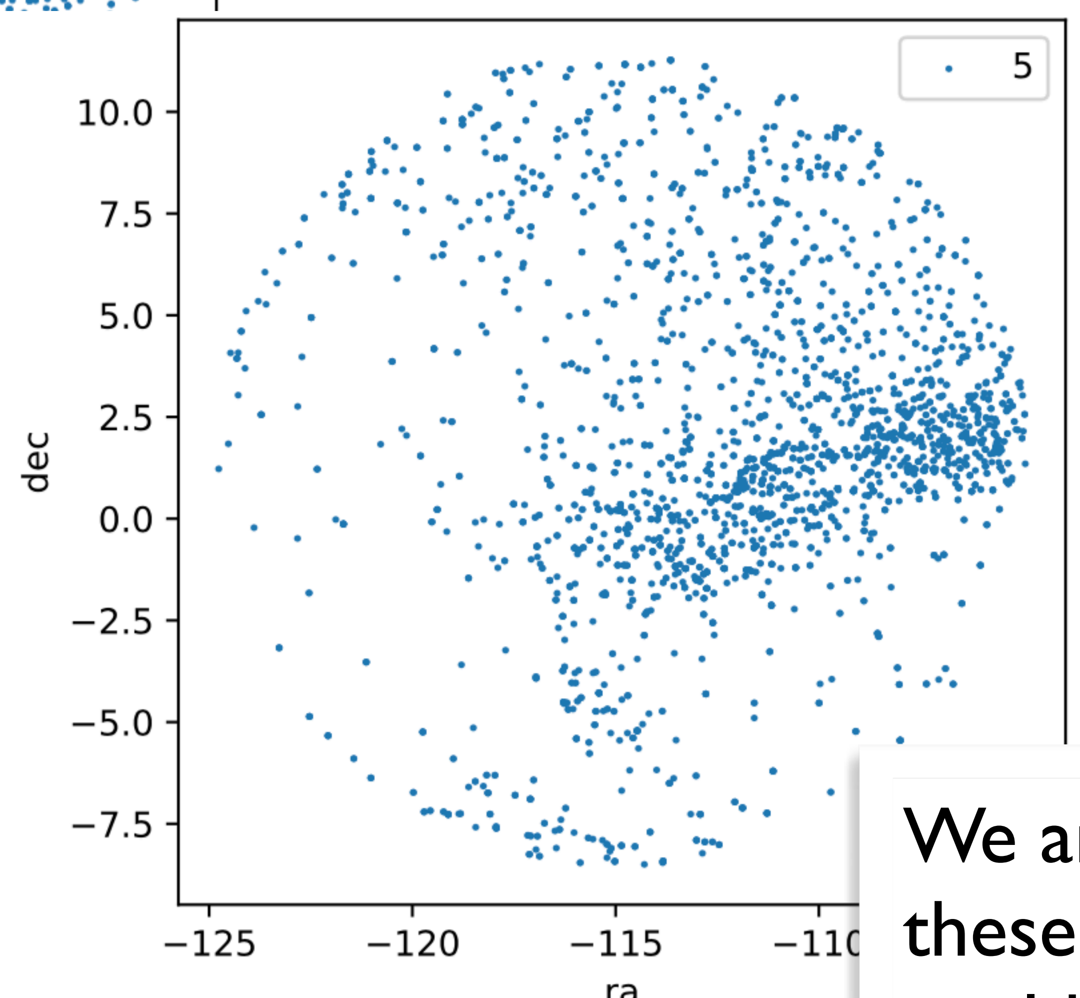
The method successfully finds known streams!

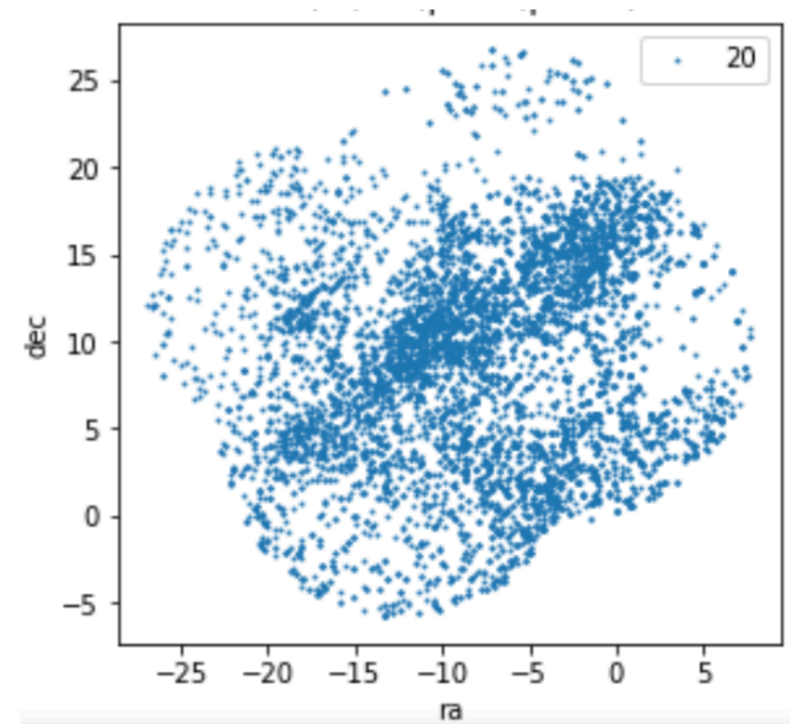


Via Machinae: preliminary results



The method also finds many other stream candidates



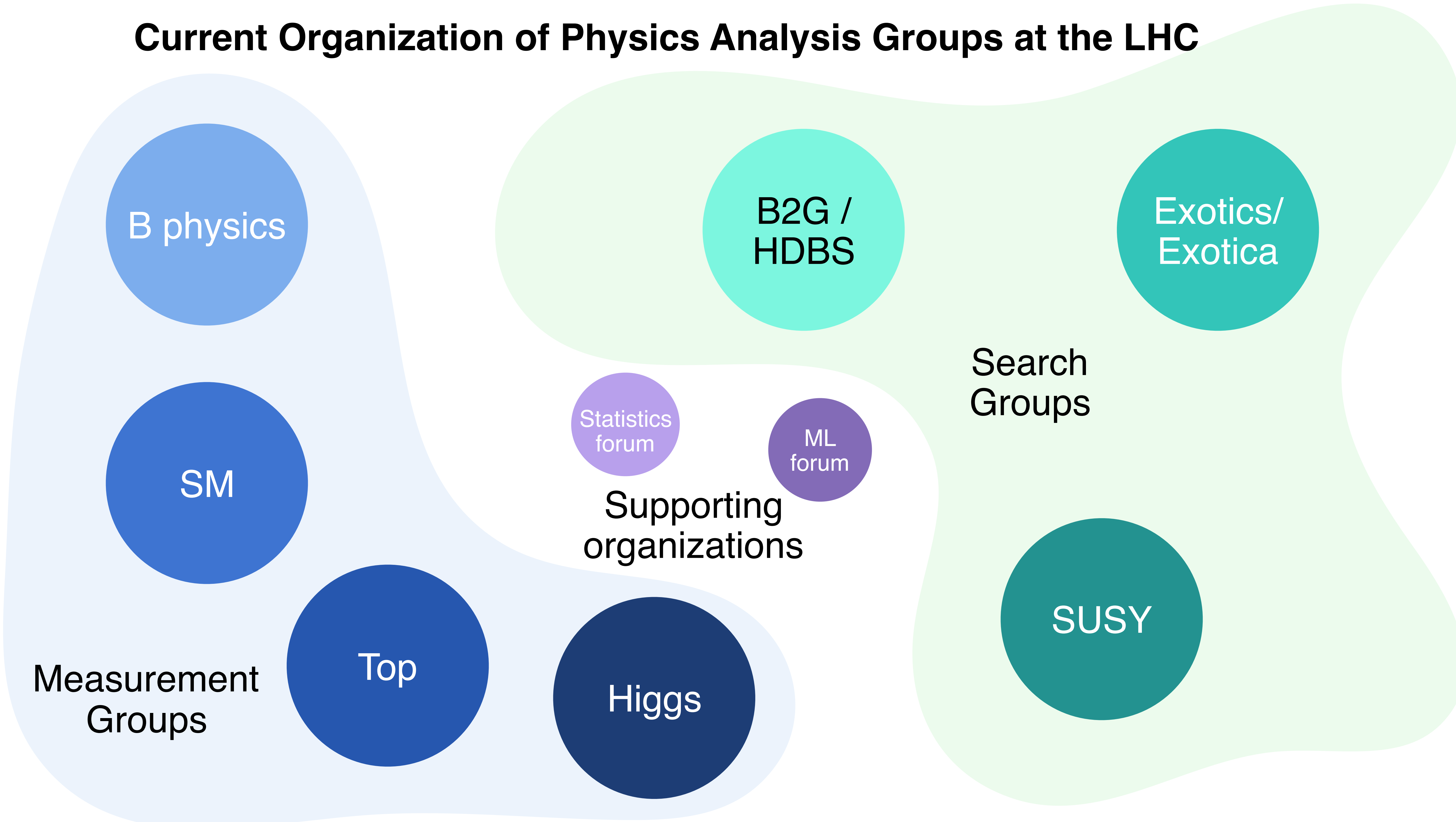


We are currently investigating these further to see if they could be real. Stay tuned!

Summary and Outlook

- Advances in machine learning are opening up new and exciting avenues for model independent new physics searches at the LHC.
- Ideas and methods inspired by LHC problems are being applied successfully to other fields such as astronomy and astrophysics.
- The LHC Olympics 2020 provided a very useful testing ground for the development and common benchmarking of new approaches.
- Much work remains to be done in order to port these ideas over to ATLAS and CMS and implement them as actual analyses on real data.
- We need more ideas for model-independent searches at the LHC. This is just the beginning!

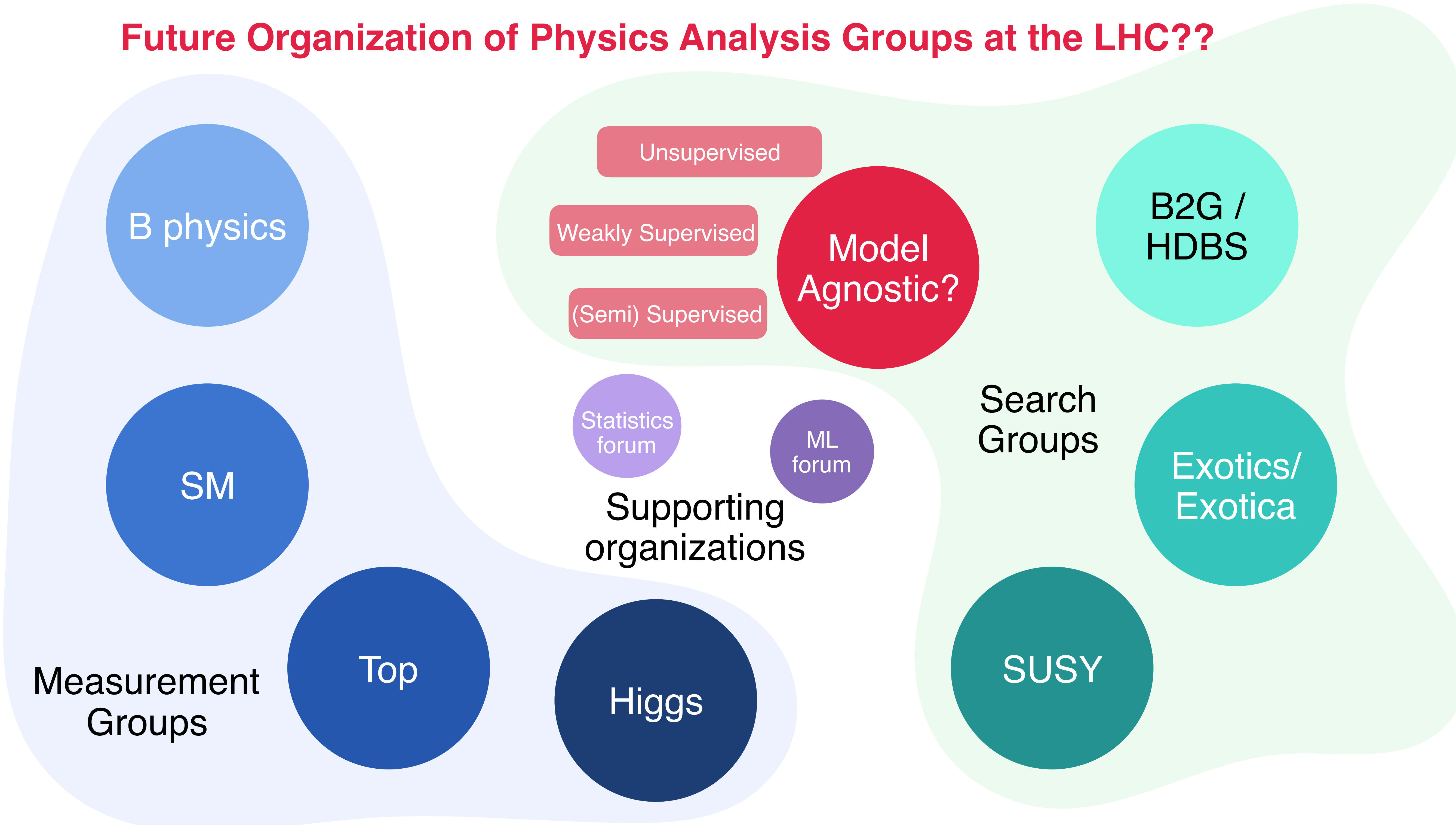
Current Organization of Physics Analysis Groups at the LHC



Q: Why is there no model independent search group???

A vision for the future...

Future Organization of Physics Analysis Groups at the LHC??



from G. Kasieczka, B. Nachman, DS (eds), et al 2101.08320

Thanks for your attention!

Additional slides

Masked Autoregressive Flows

(Papamakarios, Pavlakou, Murray 1705.07057)

Idea: learn an *invertible* transformation between data (x) to base space (z) with simple (eg normal) distribution.

$$P(x) = P(z)|\frac{\partial f}{\partial z}|^{-1}$$

"normalizing flow"

Need fast evaluation of Jacobian

Idea: assume *autoregressive* transformation of the form

$$\begin{aligned}
x_1 &= z_1 \\
x_2 &= z_2 * \sigma_2(z_1) + \mu_2(z_1) \\
x_3 &= z_3 * \sigma_3(z_1, z_2) + \mu_3(z_1, z_2) \\
&\dots \\
x_d &= z_d * \sigma_d(z_1, \dots, z_{d-1}) + \mu_d(z_1, \dots, z_{d-1})
\end{aligned}$$

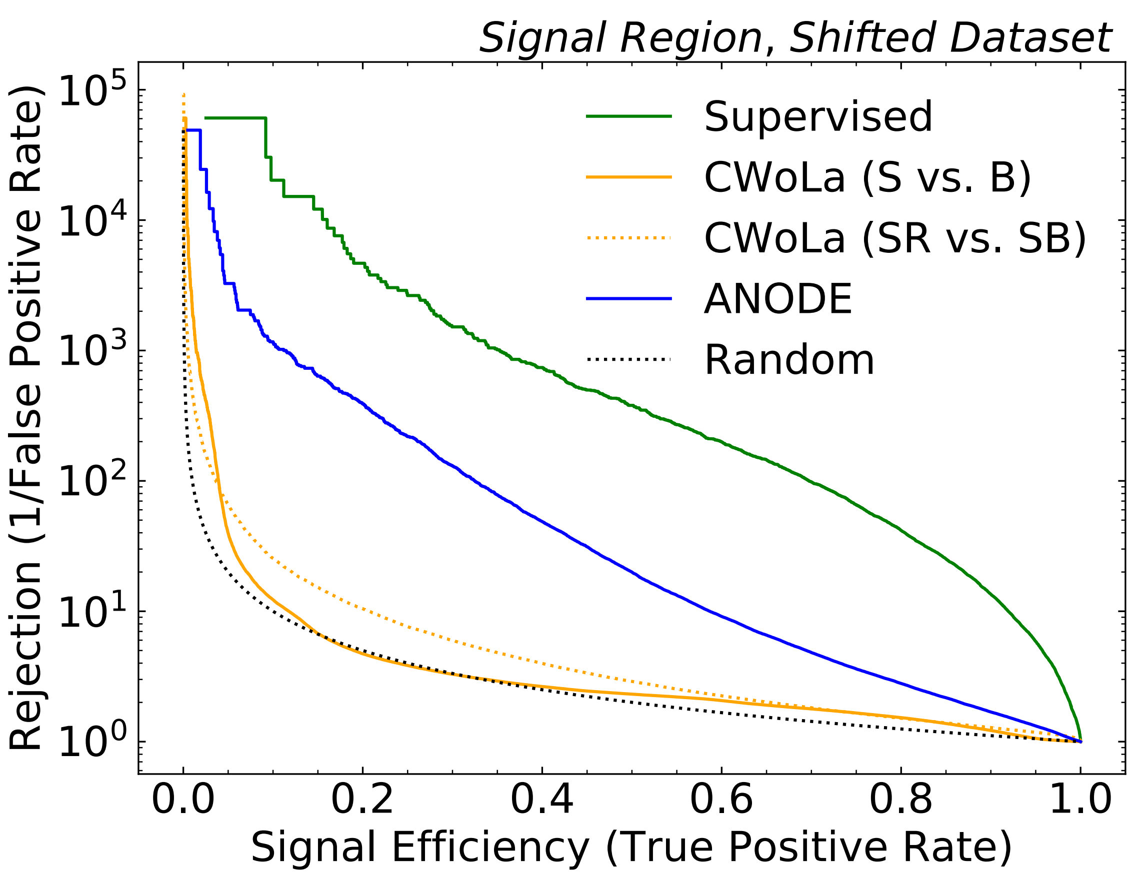
Jacobian is upper triangular!
Can be evaluated quickly!

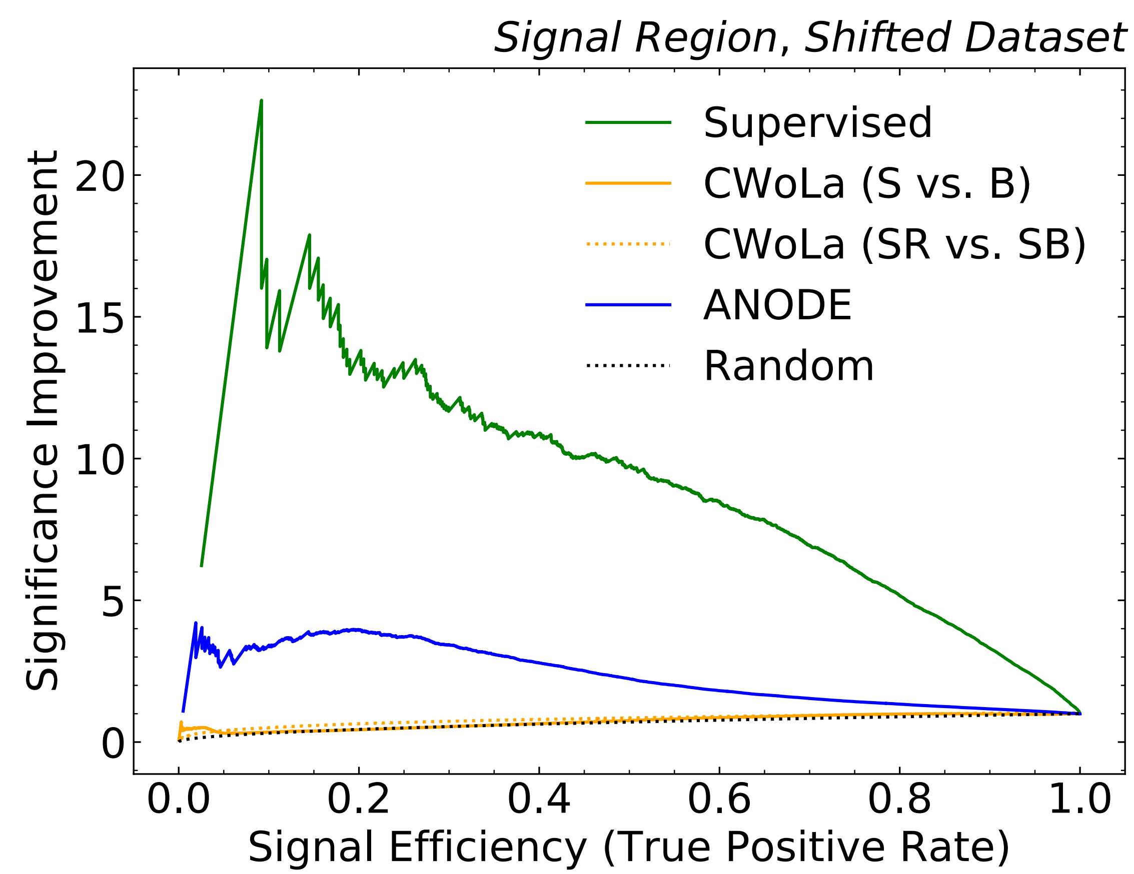
Idea: transformations can be *stacked* to greatly increase model expressivity

ANODE: Results on LHCO R&D Dataset

Ben Nachman & DS 2001.04990

Can also consider performance on a feature set which is not independent of m. We introduced artificial correlations just as proof of concept: $m_{J_{1,2}} \to m_{J_{1,2}} + c\, m_{JJ}$



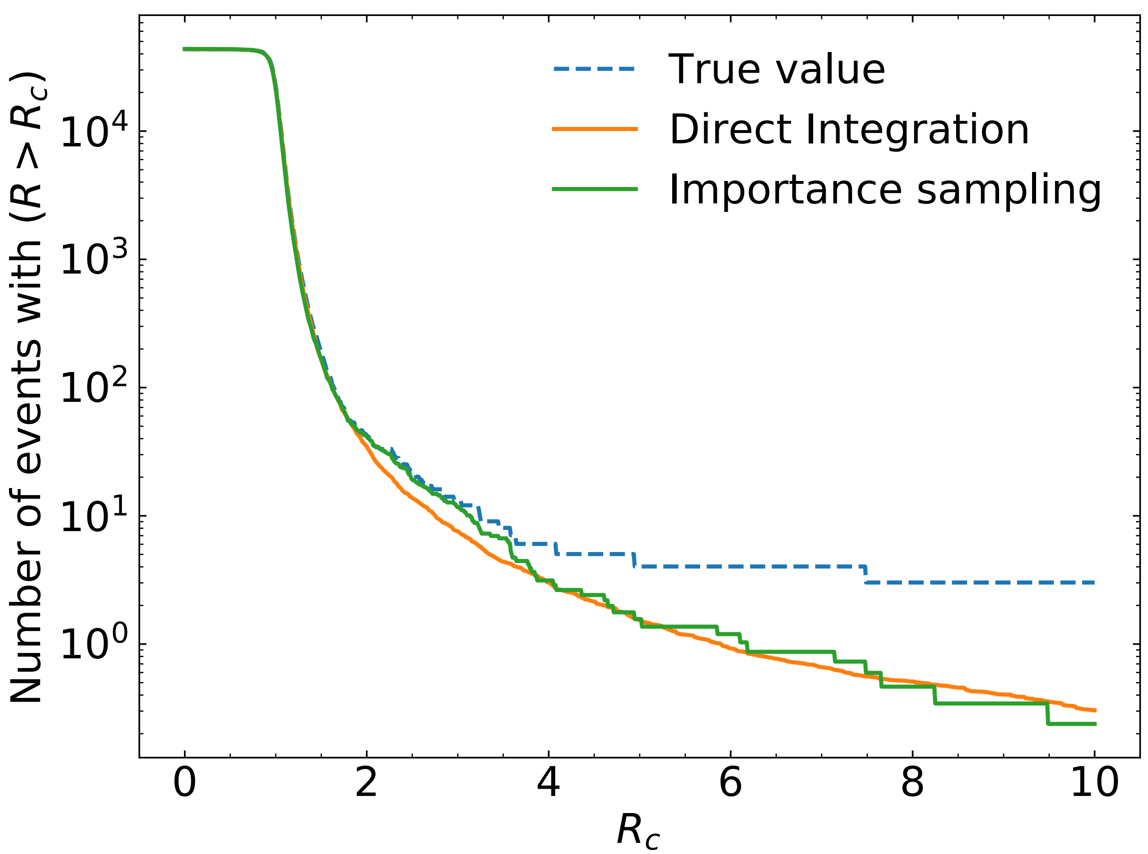


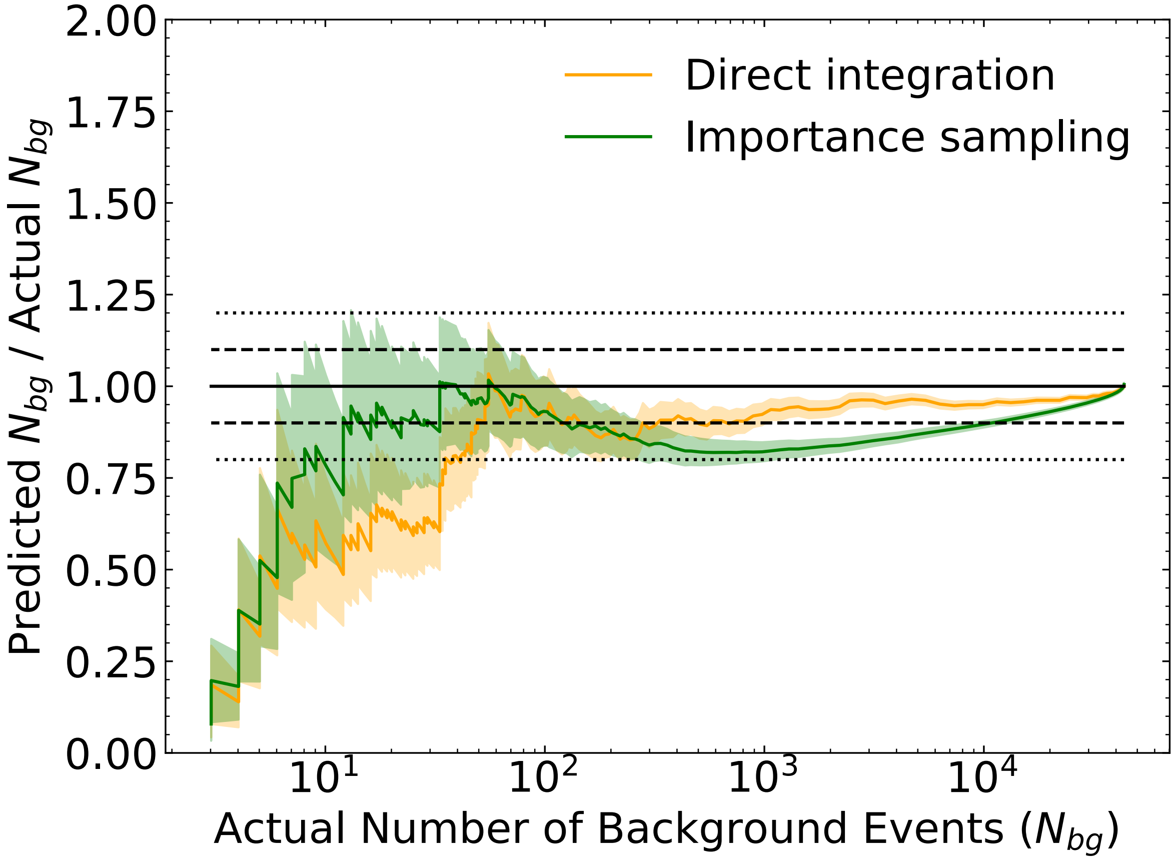
ANODE is robust while CWoLa completely fails!

ANODE: Results on LHCO R&D Dataset

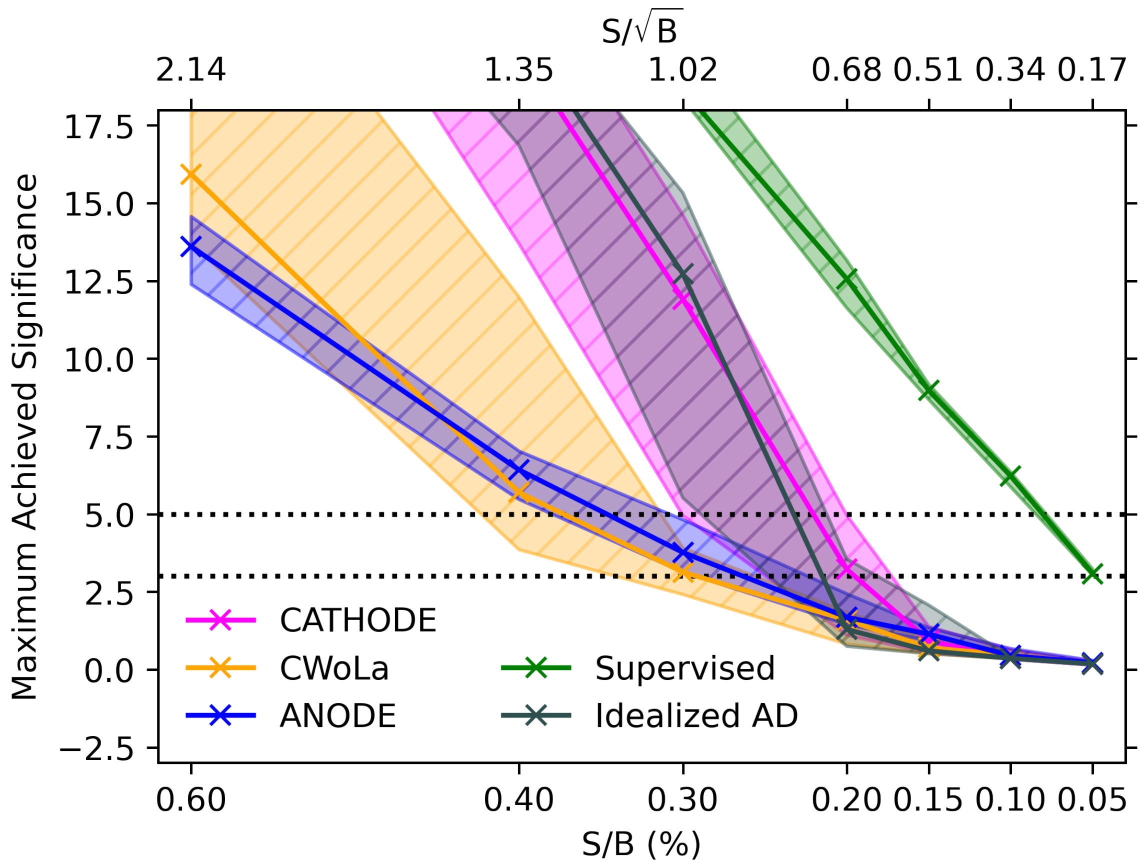
Nachman & DS 2001.04990

Novel aspect of ANODE: can estimate backgrounds directly with P(x|bg; m∈SR)

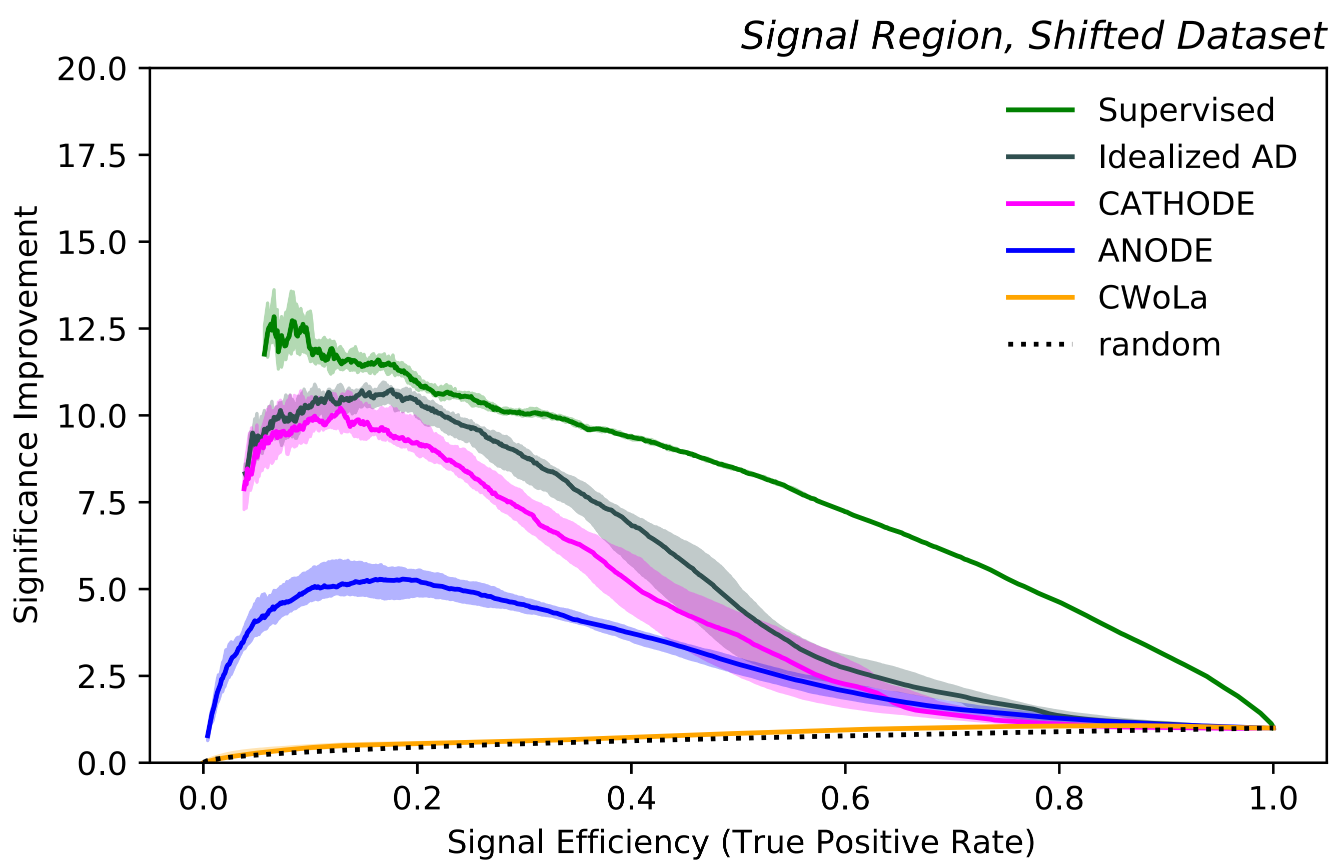




CATHODE: more results



CATHODE saturates idealized AD across wide range of signal strengths



CATHODE remains near-optimal even in the presence of correlations

LHC Olympics 2020: Results

- 9 groups submitted results on box 1
- 5 groups submitted results on boxes 2 and 3
- (A number of additional groups could not finish the challenge in time but got results on the R&D dataset, or on the black boxes after unblinding)
- Two workshops:
  - “Winter Olympics” — special session of the ML4Jets conference, January 2020, NYU [box 1 opened]
  - “Summer Olympics” — virtual anomaly detection mini-workshop, July 2020, “Hamburg” [boxes 2 & 3 opened]

Box 1

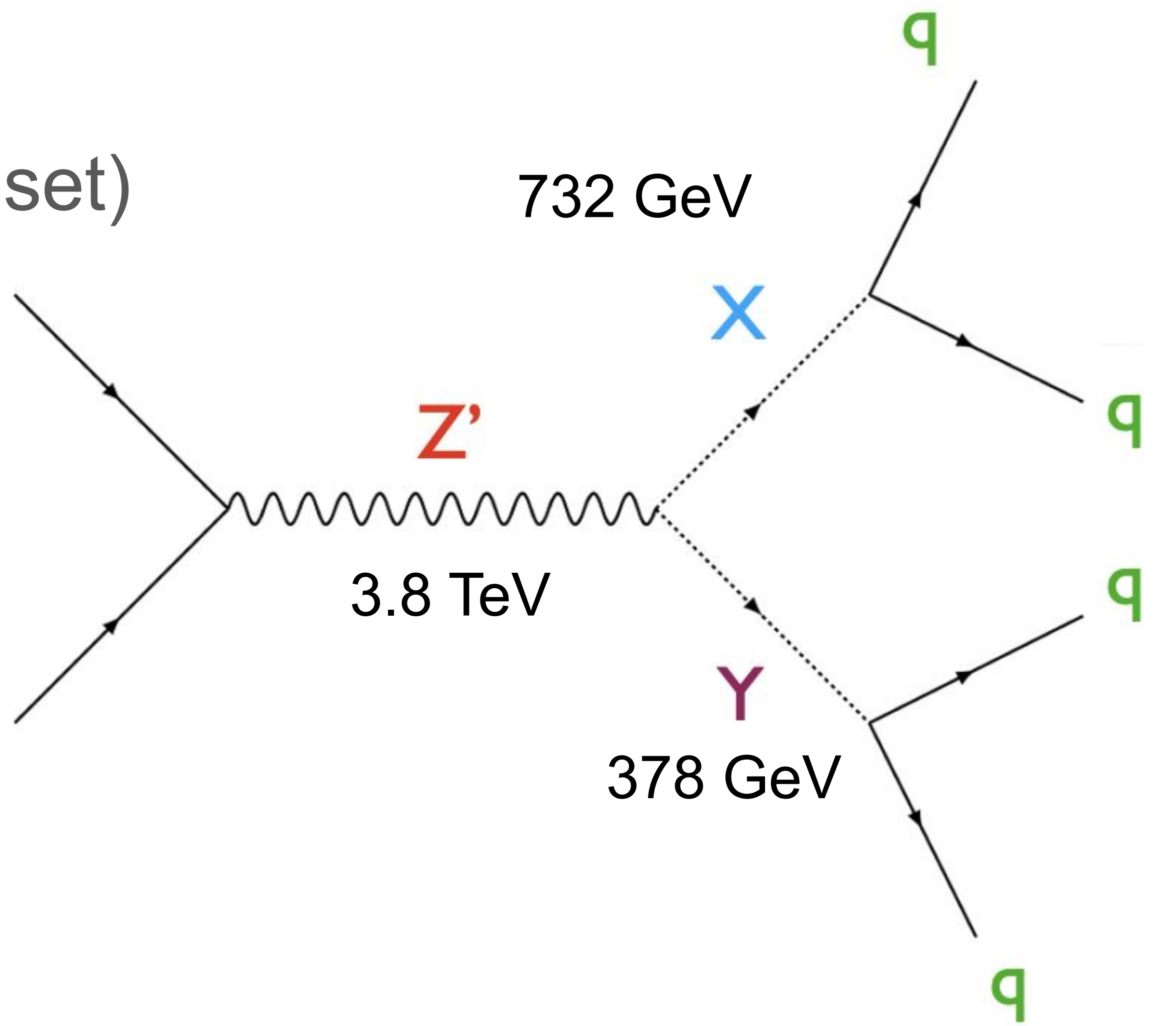
Signal: 834 events

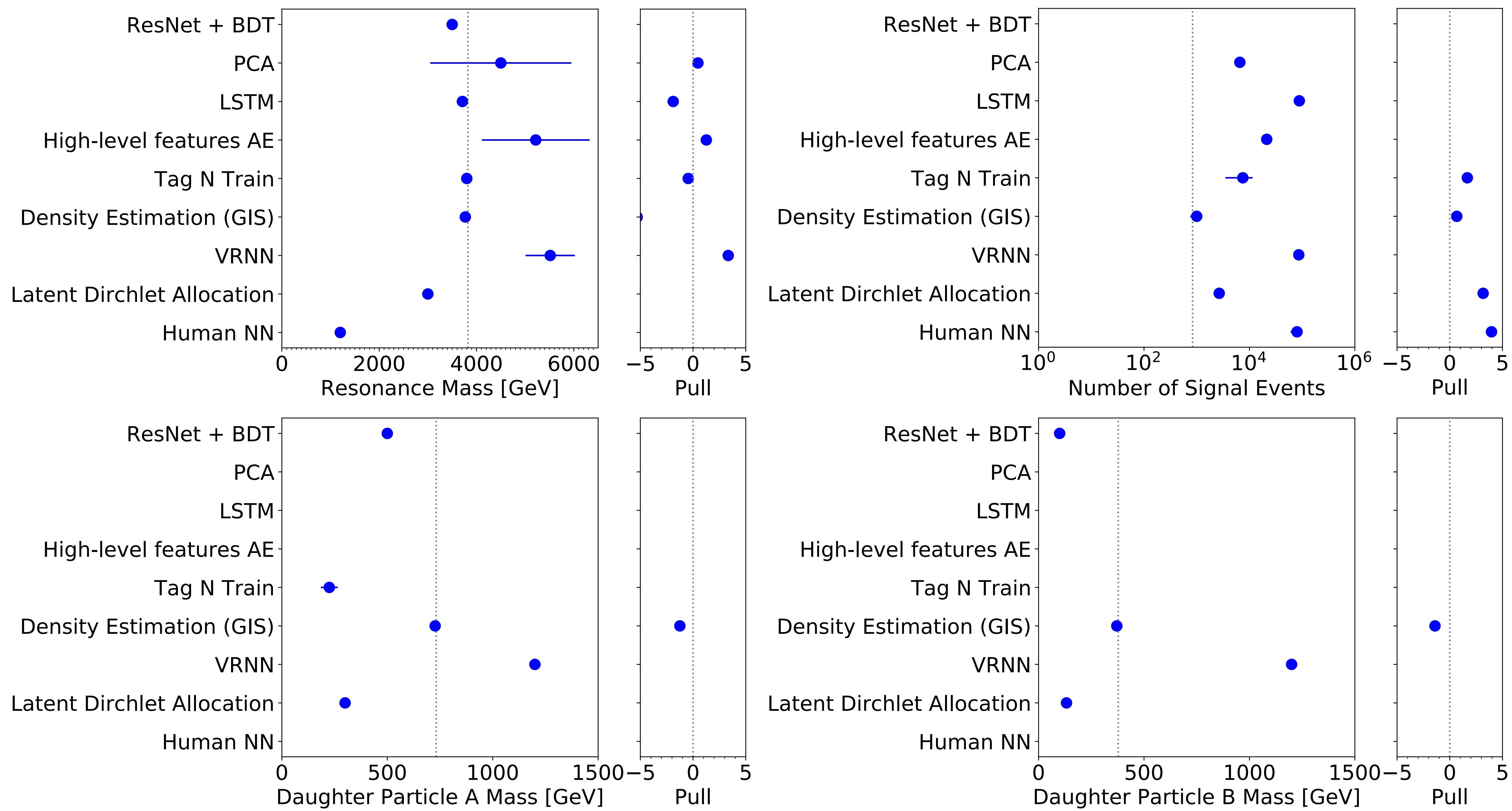
Z’->XY; X,Y->qq
(same topology as R&D dataset)

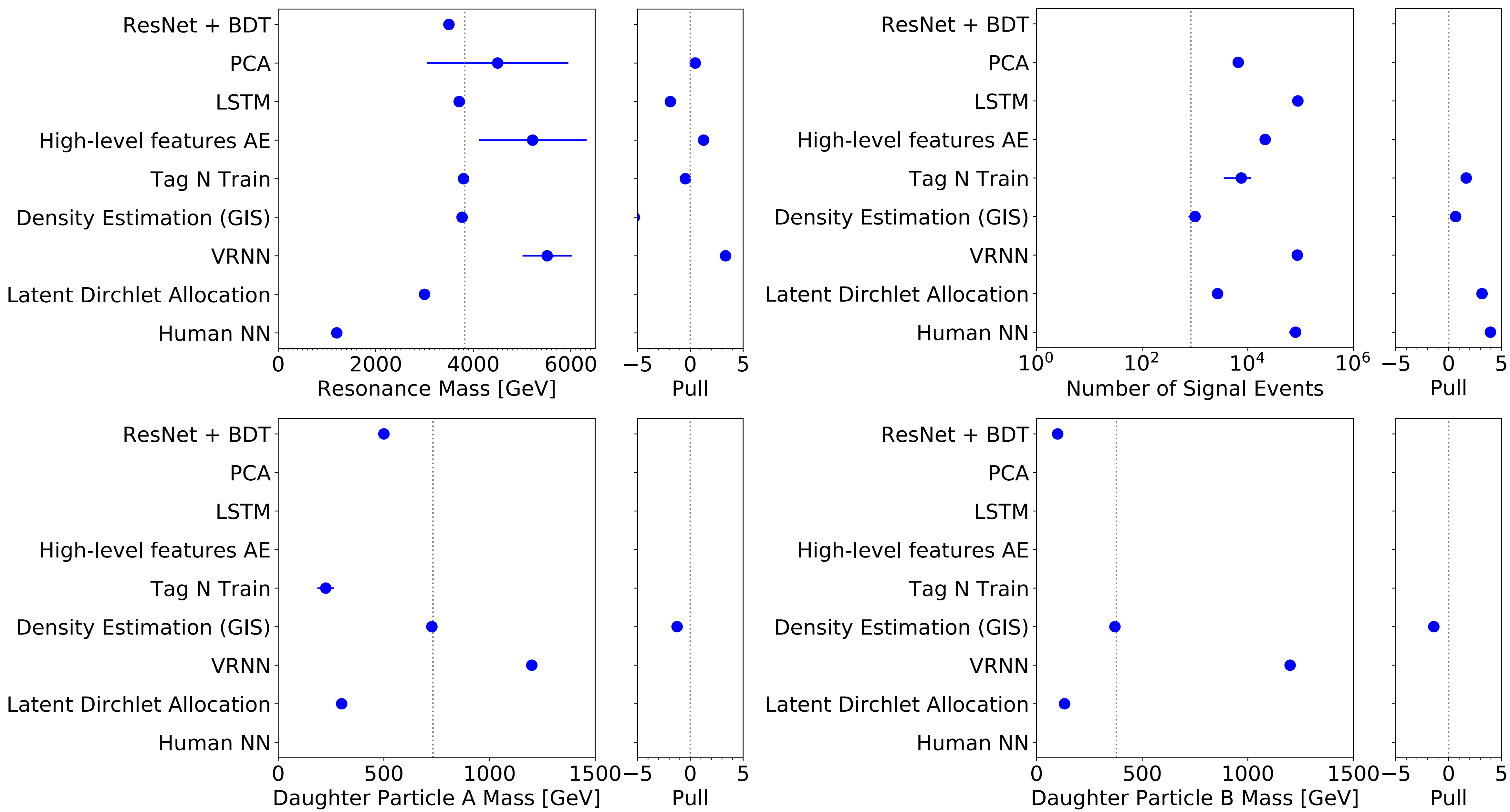
mZ’ = 3823 GeV

mX = 732 GeV

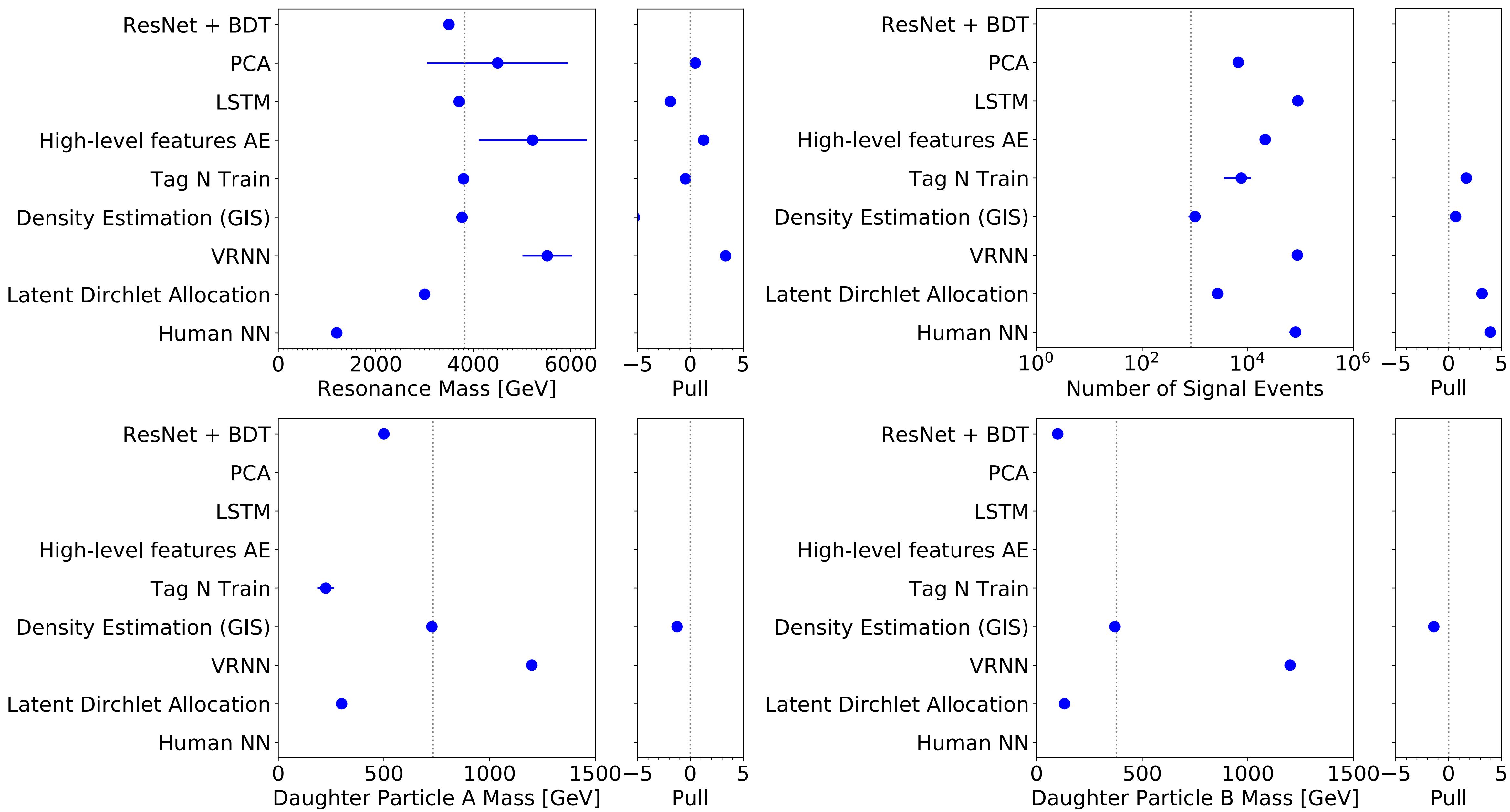
mY = 378 GeV




ResNet + BDT
PCA
LSTM
High-level features AE
Tag N Train
Density Estimation (GIS)
VRNN
Latent Dirchlet Allocation
Human NN
Resonance Mass [GeV]
Pull
Number of Signal Events
Daughter Particle A Mass [GeV]
Daughter Particle B Mass [GeV]



Two approaches clearly stood out:



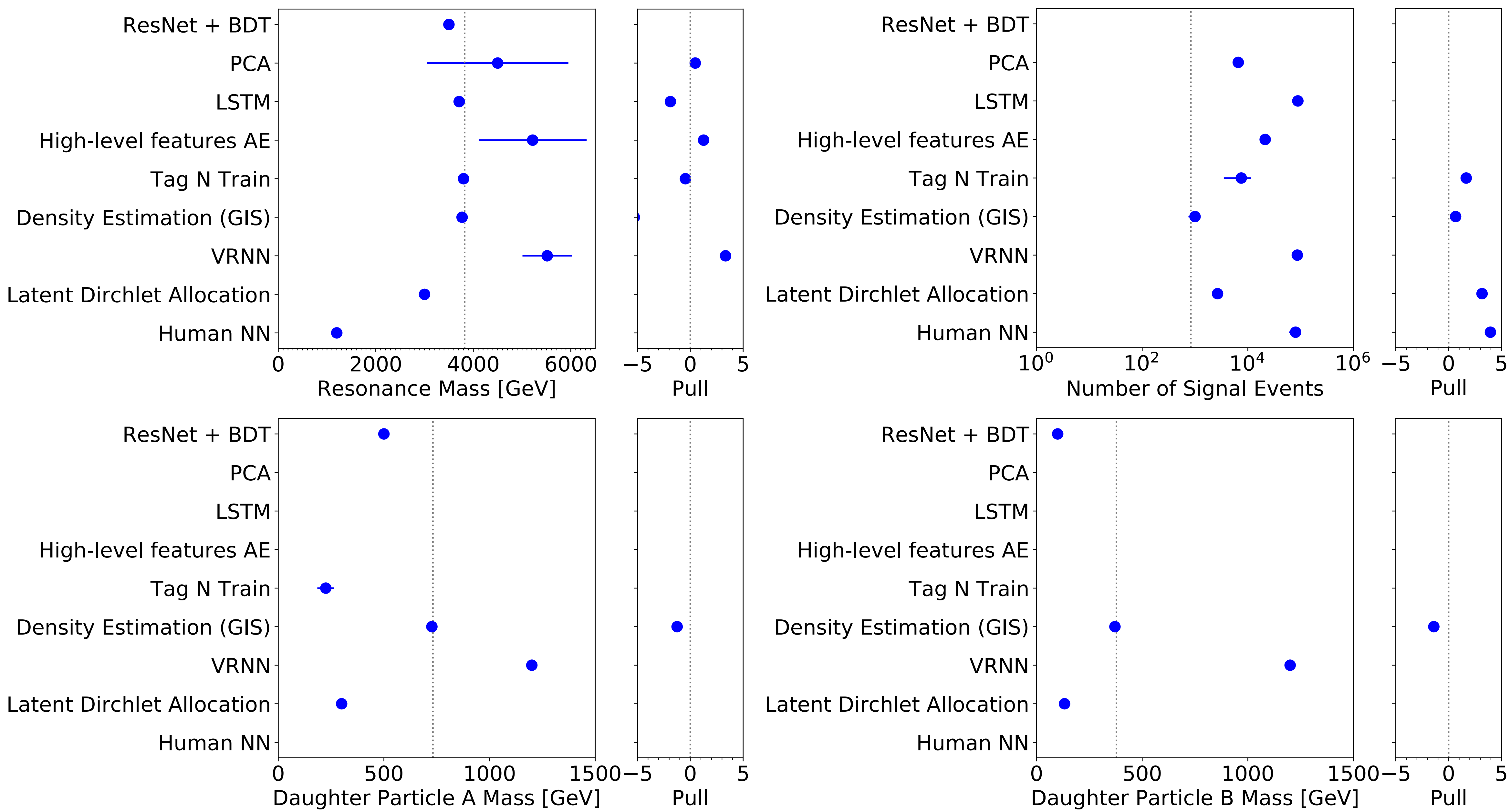
Conditional density estimation for anomaly detection

George Stein, Uros Seljak, Biwei Dai, He Jia

BERKELEY CENTER *for* COSMOLOGICAL PHYSICS

Two approaches clearly stood out:

Used the ANODE method with a novel density estimator!



Tag N' Train

Oz Amram & Cristina Mantilla Suarez (Johns Hopkins)

JOHNS HOPKINS
UNIVERSITY

Two approaches clearly stood out:

Used a combination of autoencoders and CWoLa hunting

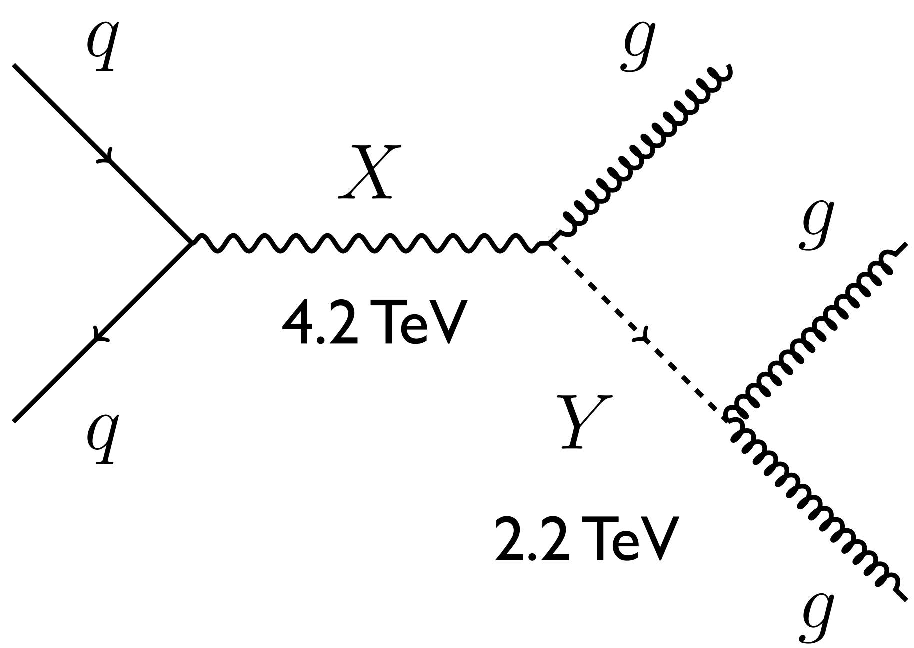
Box 2

No signal! QCD background only.

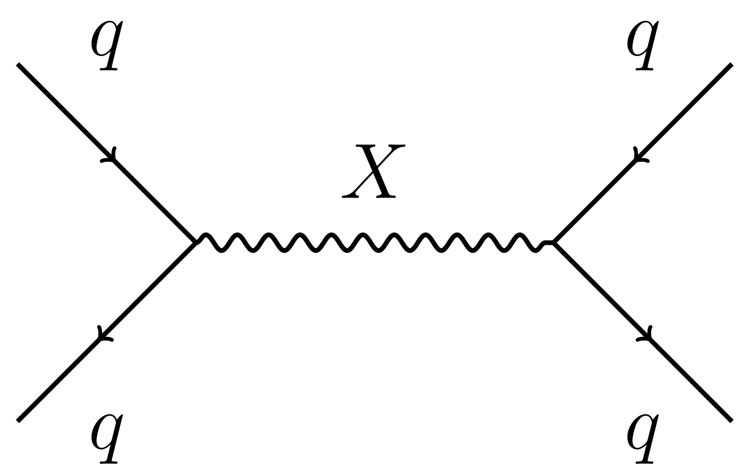
4 of the 5 submissions found false positives…

Clearly a matter of concern / area of future improvement for anomaly detection approaches!

Box 3



2000 events



1200 events

No jet substructure.

Two decay modes of X resonance. Need to combine to reach discovery significance.

No approach succeeded in finding the signal.