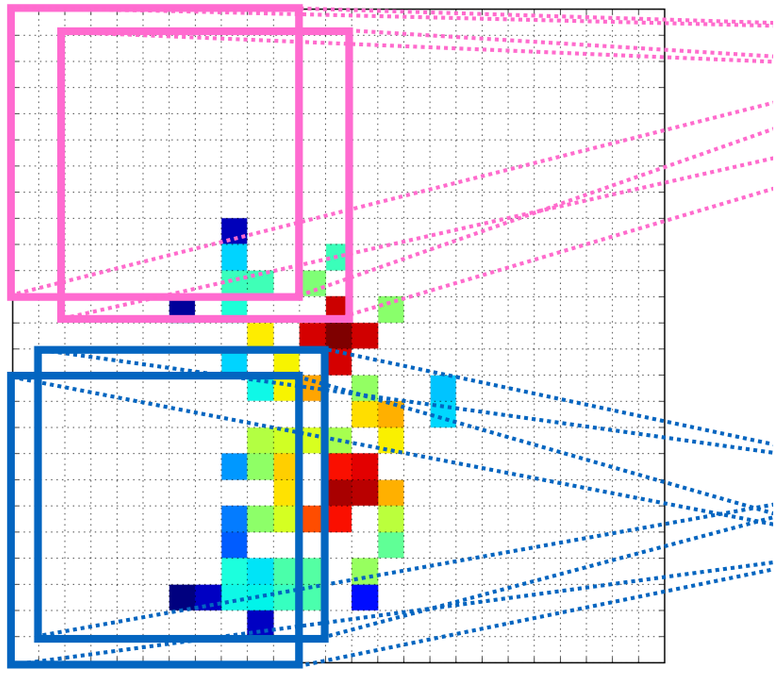
The LHC Olympics 2020:
A Community Challenge for Anomaly Detection in High Energy Physics

Benjamin Nachman

Lawrence Berkeley National Laboratory

bpnachman.com
bpnachman@lbl.gov
@bpnachman
bnachman



LHC Reinterpretation Workshop

Feb. 18, 2021

https://lhco2020.github.io/homepage/ 2101.08320

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]

Lawre

bpnac

bpnachr

tion

21

(Brief) Motivation

Theoretical and **experimental** questions motivate a deep exploration **of the fundamental structure of nature**

Dark matter

Hierarchy problem

Strong CP

Dark energy

Flavor puzzles

Baryogengesis

We have performed thousands of hypothesis tests & have no significant evidence for physics beyond the Standard Model

Three possibilities

(Brief) Motivation

Theoretical and **experimental** questions motivate a deep exploration **of the fundamental structure of nature**

Dark matter

Hierarchy problem

Strong CP

Flavor puzzles

Dark energy

Baryogengesis

We have performed thousands of hypothesis tests & have no significant evidence for physics beyond the Standard Model

Three possibilities

(1) There is nothing new at LHC energies

(Brief) Motivation

Theoretical and **experimental** questions motivate a deep exploration **of the fundamental structure of nature**

Dark matter

Strong CP

Hierarchy problem

Dark energy

Flavor puzzles

Baryogengesis

We have performed thousands of hypothesis tests & have no significant evidence for physics beyond the Standard Model

Three possibilities

(1) There is nothing new at LHC energies

(2) Patience! (new physics is rare)

(Brief) Motivation

Theoretical and **experimental** questions motivate a deep exploration **of the fundamental structure of nature**

Dark matter

Hierarchy problem

Strong CP

Flavor puzzles

Dark energy

Baryogengesis

We have performed thousands of hypothesis tests & have no significant evidence for physics beyond the Standard Model

Three possibilities

(1) There is nothing new at LHC energies

(2) Patience! (new physics is rare)

(3) We are not looking in the right place

(Brief) Motivation

Theoretical and **experimental** questions motivate a deep exploration **of the fundamental structure of nature**

Dark matter

Hierarchy problem

Strong CP

Dark energy

Flavor puzzles

Baryogengesis

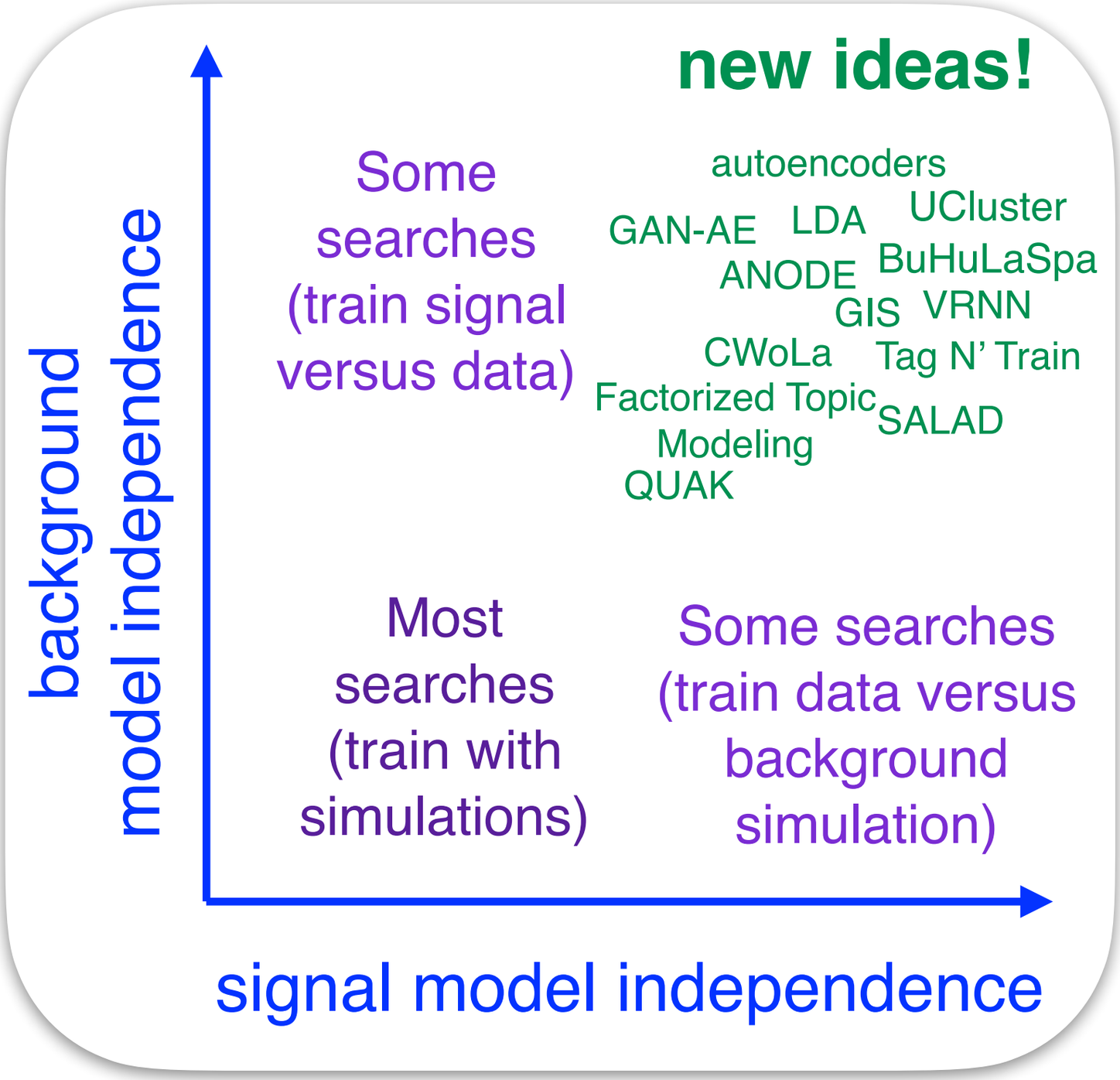
We have performed thousands of hypothesis tests & have no significant evidence for physics beyond the Standard Model

Three possibilities

This is what motivated this challenge!

(3) We are not looking in the right place

New Methods



There are many new ideas that make use of modern machine learning

The goal is to learn **directly from data**, injecting as little bias as possible

*N.B. this is just for signal sensitivity - there is **also model dependence** for determining the background*

Adapted from BN and D. Shih, 2001.04990

The Challenge

We provide a list of particle for each event (700 particles with the 3-vector of each particle)

1 dataset for R&D with labeled signal and background

3 black boxes with unlabeled data

The particle-level + detector-level simulation for background in the black boxes was modified for each dataset (think Pythia/Herwig, etc.)

Actually, all of the parameters are now public on Zenodo

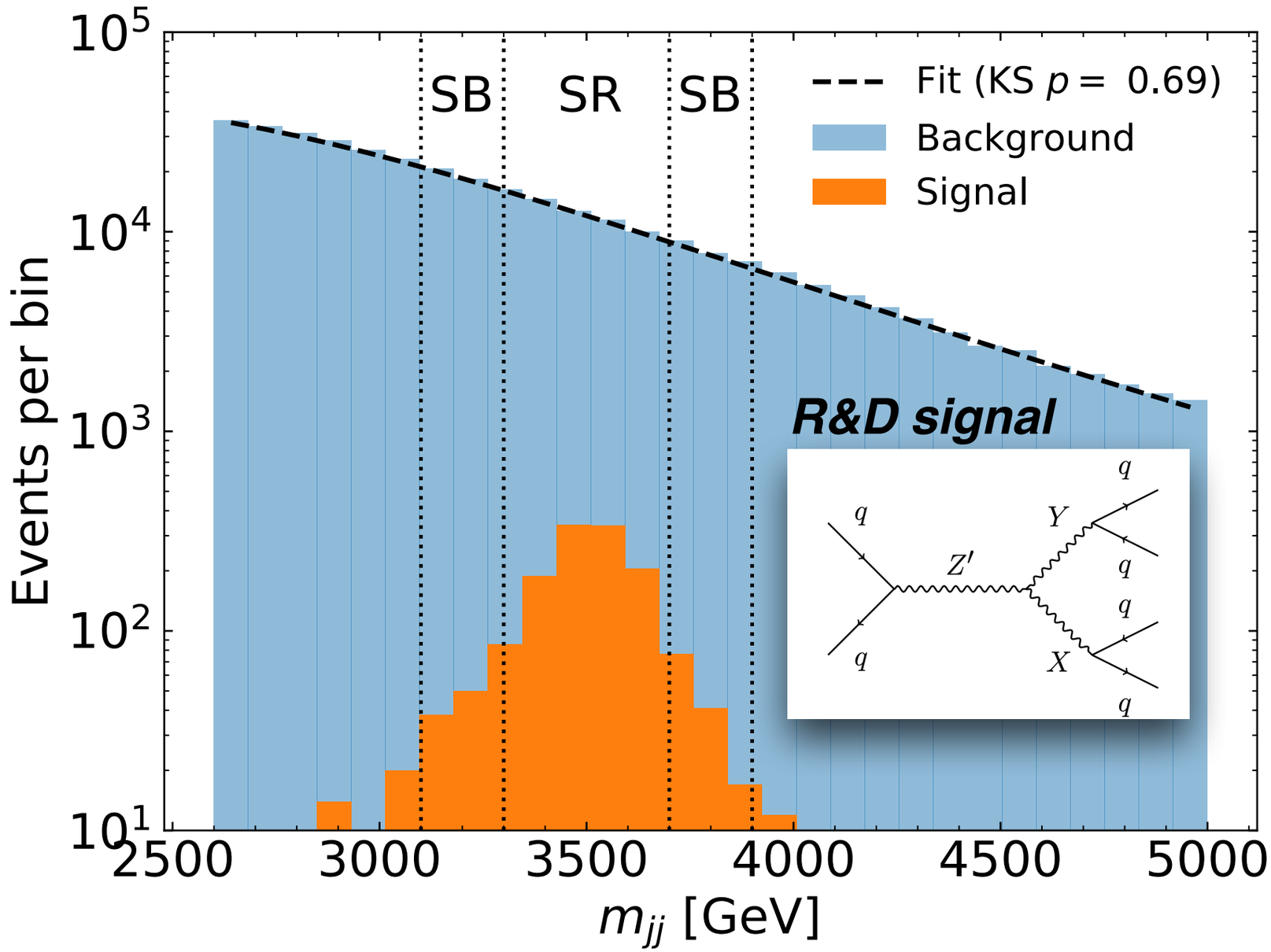
The Challenge

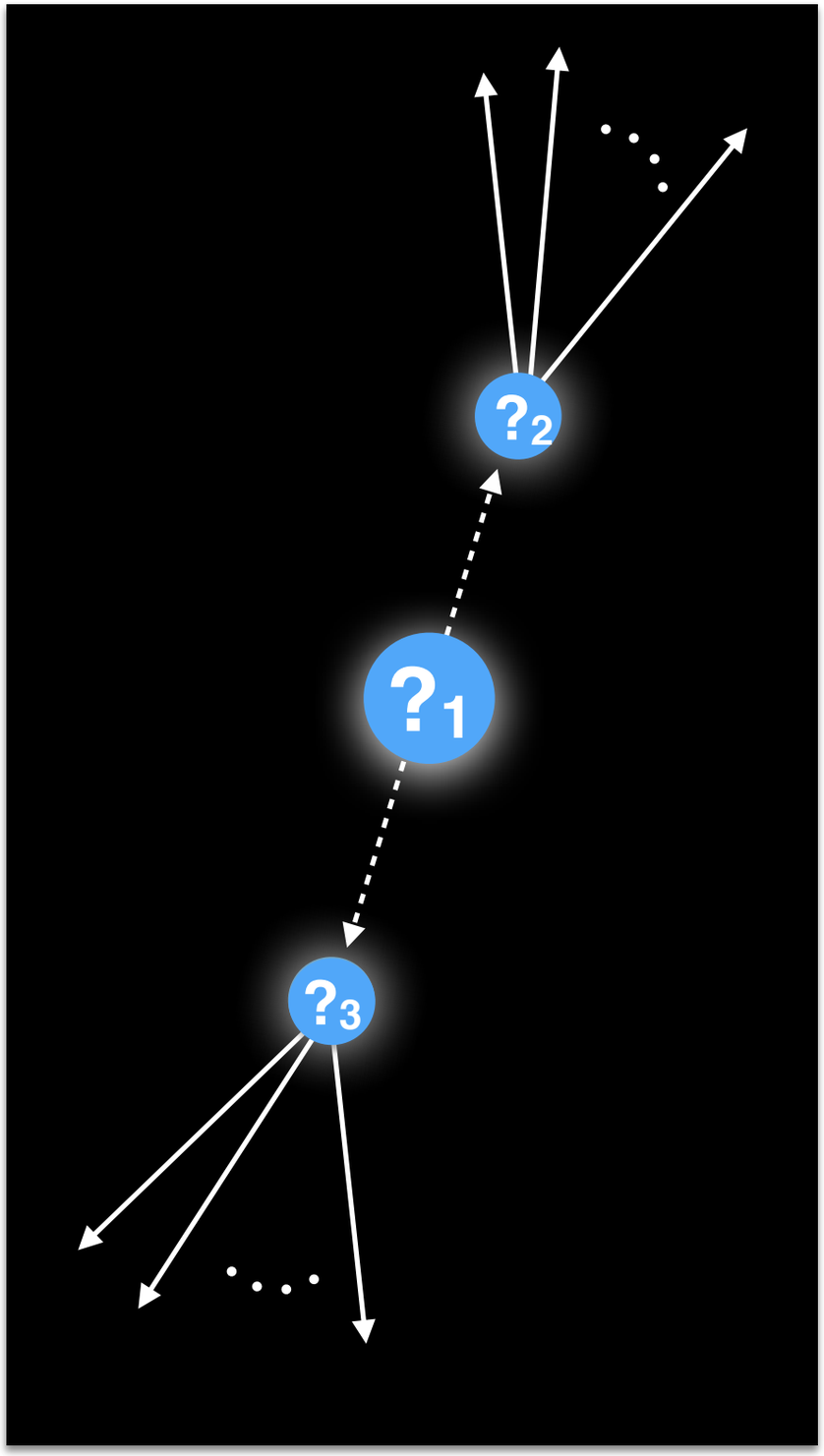
Section	Short Name	Method Type	Results Type
3.1	VRNN	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.2	ANODE	Unsupervised	(iii)
3.3	BuHuLaSpa	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.4	GAN-AE	Unsupervised	(i) (BB2-3) and (ii) (BB1)
3.5	GIS	Unsupervised	(i) (BB1)
3.6	LDA	Unsupervised	(i) (BB1-3)
3.7	PGA	Unsupervised	(ii) (BB1-2)
3.8	Reg. Likelihoods	Unsupervised	(iii)
3.9	UCluster	Unsupervised	(i) (BB2-3)
4.1	CWoLa	Weakly Supervised	(ii) (BB1-2)
4.2	CWoLa AE Compare	Weakly/Unsupervised	(iii)
4.3	Tag N' Train	Weakly Supervised	(i) (BB1-3)
4.4	SALAD	Weakly Supervised	(iii)
4.5	SA-CWoLa	Weakly Supervised	(iii)
5.1	Deep Ensemble	Semisupervised	(i) (BB1)
5.2	Factorized Topics	Semisupervised	(iii)
5.3	QUAK	Semisupervised	(i) (BB2,3) and (ii) (BB1)
5.4	LSTM	Semisupervised	(i) (BB1-3)

BB = black box; (i) = blinded, (ii) = unblinded

(I'll explain what supervision is momentarily)

The dataset





Dijet final state (allow for data-driven background + complex final state).

Timeline

April 4, 2019: R&D data released

Nov. 19, 2019: Black boxes released

Jan. 16, 2020: Winter Olympics + BB1 unblinded

July 16-17, 2020: Summer Olympics + BB2/3 unblinded

July 20, 2020: Invite contributions to community paper

Oct. 1, 2020: Contributions due

Jan. 20, 2021: Paper posted to arXiv

Solutions

I don't have time to cover all of them - please see the paper for details! I'll just highlight some general ideas.

Section	Short Name	Method Type	Results Type
3.1	VRNN	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.2	ANODE	Unsupervised	(iii)
3.3	BuHuLaSpa	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.4	GAN-AE	Unsupervised	(i) (BB2-3) and (ii) (BB1)
3.5	GIS	Unsupervised	(i) (BB1)
3.6	LDA	Unsupervised	(i) (BB1-3)
3.7	PGA	Unsupervised	(ii) (BB1-2)
3.8	Reg. Likelihoods	Unsupervised	(iii)
3.9	UCluster	Unsupervised	(i) (BB2-3)
4.1	CWoLa	Weakly Supervised	(ii) (BB1-2)
4.2	CWoLa AE Compare	Weakly/Unsupervised	(iii)
4.3	Tag N' Train	Weakly Supervised	(i) (BB1-3)
4.4	SALAD	Weakly Supervised	(iii)
4.5	SA-CWoLa	Weakly Supervised	(iii)
5.1	Deep Ensemble	Semisupervised	(i) (BB1)
5.2	Factorized Topics	Semisupervised	(iii)
5.3	QUAK	Semisupervised	(i) (BB2,3) and (ii) (BB1)
5.4	LSTM	Semisupervised	(i) (BB1-3)

BB = black box; (i) = blinded, (ii) = unblinded

Solutions

Supervision refers to the type of label information provided to the ML during training.

Unsupervised = no labels
Weakly-supervised = noisy labels
Semi-supervised = partial labels
Supervised = full label information

Section	Short Name	Method Type	Results Type
3.1	VRNN	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.2	ANODE	Unsupervised	(iii)
3.3	BuHuLaSpa	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.4	GAN-AE	Unsupervised	(i) (BB2-3) and (ii) (BB1)
3.5	GIS	Unsupervised	(i) (BB1)
3.6	LDA	Unsupervised	(i) (BB1-3)
3.7	PGA	Unsupervised	(ii) (BB1-2)
3.8	Reg. Likelihoods	Unsupervised	(iii)
3.9	UCluster	Unsupervised	(i) (BB2-3)
4.1	CWoLa	Weakly Supervised	(ii) (BB1-2)
4.2	CWoLa AE Compare	Weakly/Unsupervised	(iii)
4.3	Tag N' Train	Weakly Supervised	(i) (BB1-3)
4.4	SALAD	Weakly Supervised	(iii)
4.5	SA-CWoLa	Weakly Supervised	(iii)
5.1	Deep Ensemble	Semisupervised	(i) (BB1)
5.2	Factorized Topics	Semisupervised	(iii)
5.3	QUAK	Semisupervised	(i) (BB2,3) and (ii) (BB1)
5.4	LSTM	Semisupervised	(i) (BB1-3)

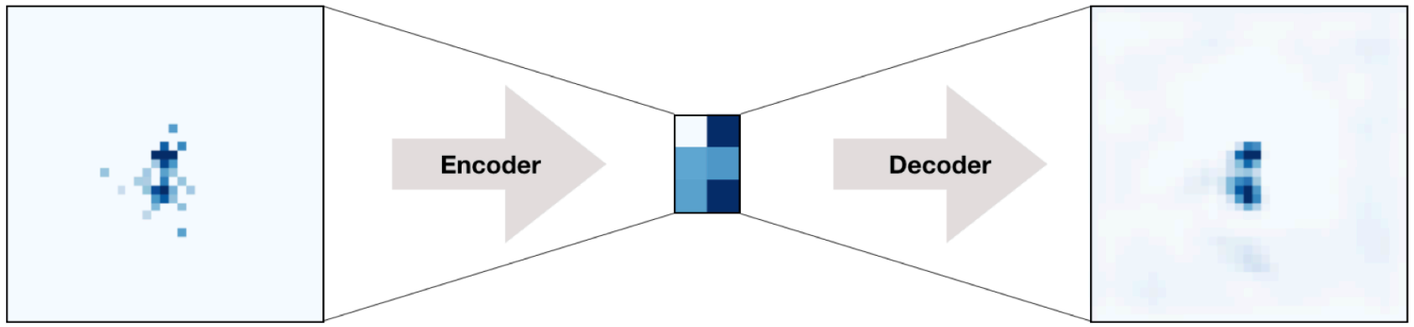
These categories are not exact and the boundaries are not rigid!

*N.B. Not everyone agrees on the boundary between semi-supervised and weakly supervised.

Solutions: Unsupervised

Unsupervised = no labels

Typically, the goal of these methods is to look for events with low *p(background)*



Section	Short Name	Method Type	Results Type
3.1	VRNN	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.2	ANODE	Unsupervised	(iii)
3.3	BuHuLaSpa	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.4	GAN-AE	Unsupervised	(i) (BB2-3) and (ii) (BB1)
3.5	GIS	Unsupervised	(i) (BB1)
3.6	LDA	Unsupervised	(i) (BB1-3)
3.7	PGA	Unsupervised	(ii) (BB1-2)
3.8	Reg. Likelihoods	Unsupervised	(iii)
3.9	UCluster	Unsupervised	(i) (BB2-3)
4.1	CWoLa	Weakly Supervised	(ii) (BB1-2)
4.2	CWoLa AE Compare	Weakly/Unsupervised	(iii)
4.3	Tag N' Train	Weakly Supervised	(i) (BB1-3)
4.4	SALAD	Weakly Supervised	(iii)
4.5	SA-CWoLa	Weakly Supervised	(iii)
5.1	Deep Ensemble	Semisupervised	(i) (BB1)
5.2	Factorized Topics	Semisupervised	(iii)
5.3	QUAK	Semisupervised	(i) (BB2,3) and (ii) (BB1)
5.4	LSTM	Semisupervised	(i) (BB1-3)

One strategy (autoencoders) is to try to compress events and then uncompress them. When x = uncompres(compress(x)), then x probably has low p(x).

M. Farina, Y. Nakai, D. Shih, 1808.08992; T. Heimel, G. Kasieczka, T. Plehn, J. Thompson, 1808.08979; + many more

Solutions: Weakly-supervised

Weakly-supervised = noisy labels

Typically, the goal of these methods is to look for events with high *p(possibly signal-enriched)/p(possibly signal-depleted)*

Section	Short Name	Method Type	Results Type
3.1	VRNN	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.2	ANODE	Unsupervised	(iii)
3.3	BuHuLaSpa	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.4	GAN-AE	Unsupervised	(i) (BB2-3) and (ii) (BB1)
3.5	GIS	Unsupervised	(i) (BB1)
3.6	LDA	Unsupervised	(i) (BB1-3)
3.7	PGA	Unsupervised	(ii) (BB1-2)
3.8	Reg. Likelihoods	Unsupervised	(iii)
3.9	UCluster	Unsupervised	(i) (BB2-3)
4.1	CWoLa	Weakly Supervised	(ii) (BB1-2)
4.2	CWoLa AE Compare	Weakly/Unsupervised	(iii)
4.3	Tag N' Train	Weakly Supervised	(i) (BB1-3)
4.4	SALAD	Weakly Supervised	(iii)
4.5	SA-CWoLa	Weakly Supervised	(iii)
5.1	Deep Ensemble	Semisupervised	(i) (BB1)
5.2	Factorized Topics	Semisupervised	(iii)
5.3	QUAK	Semisupervised	(i) (BB2,3) and (ii) (BB1)
5.4	LSTM	Semisupervised	(i) (BB1-3)

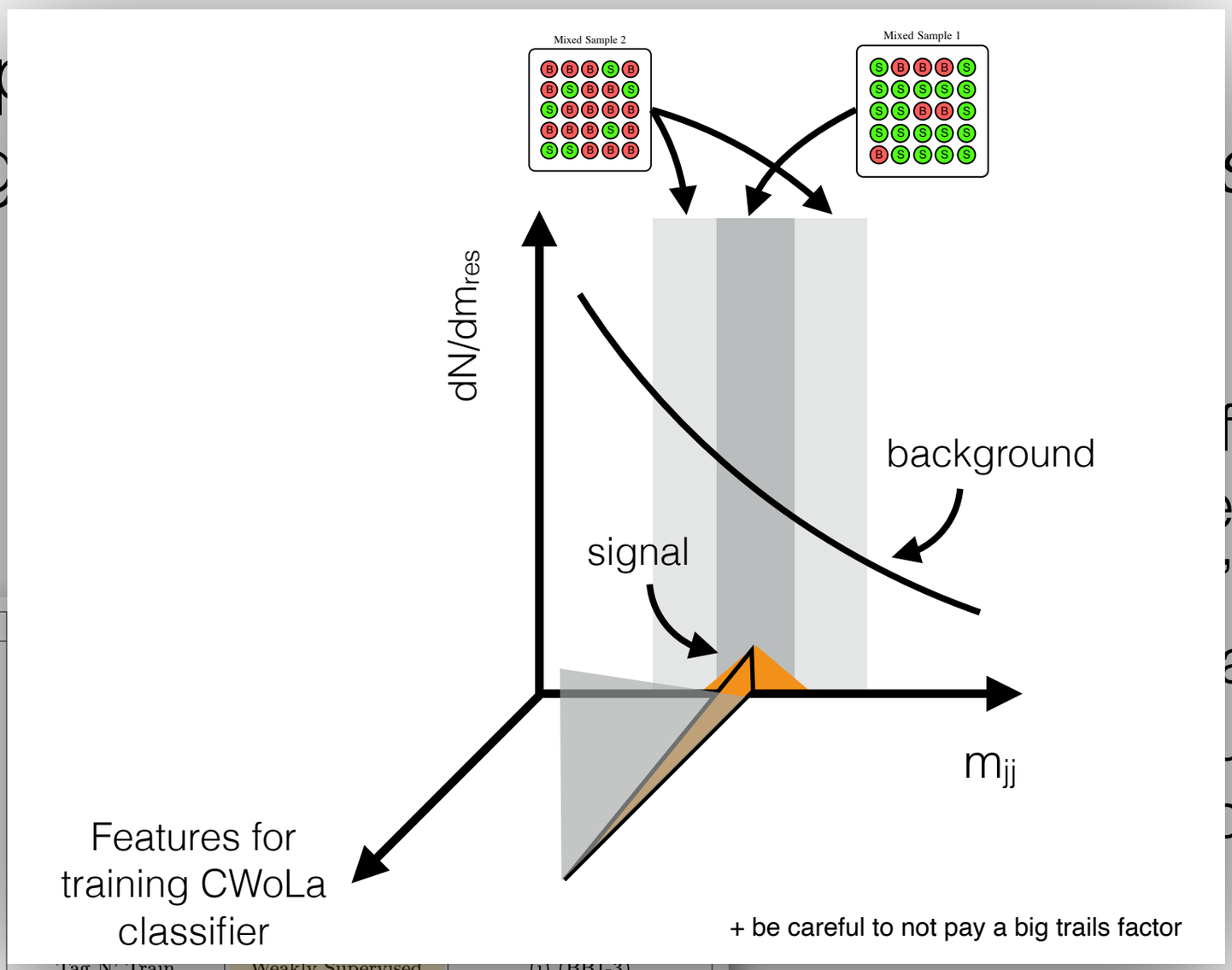
e.g. Classification Without Labels (CWoLa), events in a signal region are labeled "signal" and events in a sideband are labeled "background". These labels are "noisy" but a classifier trained with them can detect the presence of a signal.

E. Metodiev, BN, J. Thaler, 1708.02949; J. Collins, K. Howe, BN, 1805.02664

Solutions: Weakly-supervised

Weakly-supervised = noisy labels

Typ to look for events with
hig ssibly signal-depleted)



fication Without Labels
ents in a signal region are
" and events in a sideband
ackground". These labels
ut a classifier trained with
ct the presence of a signal.

Section			
3.1			
3.2			
3.3			
3.4			
3.5			
3.6			
3.7			
3.8			
3.9			
4.1			
4.2			
4.3	Tag N' Train	Weakly Supervised	(i) (BB1-3)
4.4	SALAD	Weakly Supervised	(iii)
4.5	SA-CWoLa	Weakly Supervised	(iii)
5.1	Deep Ensemble	Semisupervised	(i) (BB1)
5.2	Factorized Topics	Semisupervised	(iii)
5.3	QUAK	Semisupervised	(i) (BB2,3) and (ii) (BB1)
5.4	LSTM	Semisupervised	(i) (BB1-3)

E. Metodiev, BN, J. Thaler, 1708.02949; J. Collins, K. Howe, BN, 1805.02664

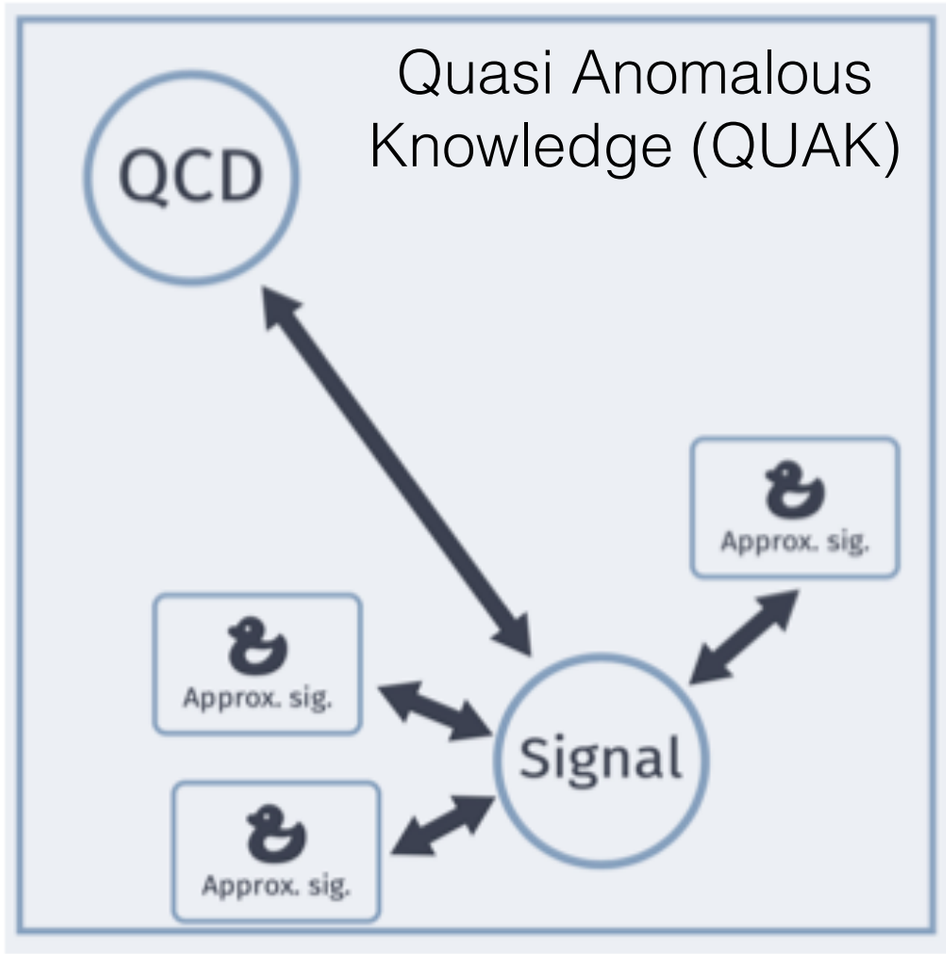
Solutions: Semi-supervised

Semi-supervised = partial labels

Typically, these methods use some signal simulations to build signal sensitivity

(We did not give bonus points for the best acronyms !)

Section	Short Name	Method Type	Results Type
3.1	VRNN	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.2	ANODE	Unsupervised	(iii)
3.3	BuHuLaSpa	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.4	GAN-AE	Unsupervised	(i) (BB2-3) and (ii) (BB1)
3.5	GIS	Unsupervised	(i) (BB1)
3.6	LDA	Unsupervised	(i) (BB1-3)
3.7	PGA	Unsupervised	(ii) (BB1-2)
3.8	Reg. Likelihoods	Unsupervised	(iii)
3.9	UCluster	Unsupervised	(i) (BB2-3)
4.1	CWoLa	Weakly Supervised	(ii) (BB1-2)
4.2	CWoLa AE Compare	Weakly/Unsupervised	(iii)
4.3	Tag N' Train	Weakly Supervised	(i) (BB1-3)
4.4	SALAD	Weakly Supervised	(iii)
4.5	SA-CWoLa	Weakly Supervised	(iii)
5.1	Deep Ensemble	Semisupervised	(i) (BB1)
5.2	Factorized Topics	Semisupervised	(iii)
5.3	QUAK	Semisupervised	(i) (BB2,3) and (ii) (BB1)
5.4	LSTM	Semisupervised	(i) (BB1-3)



S. Park, D. Rankin, S.-M. Udrescu, M. Yunus, P. Harris, 2011.03550

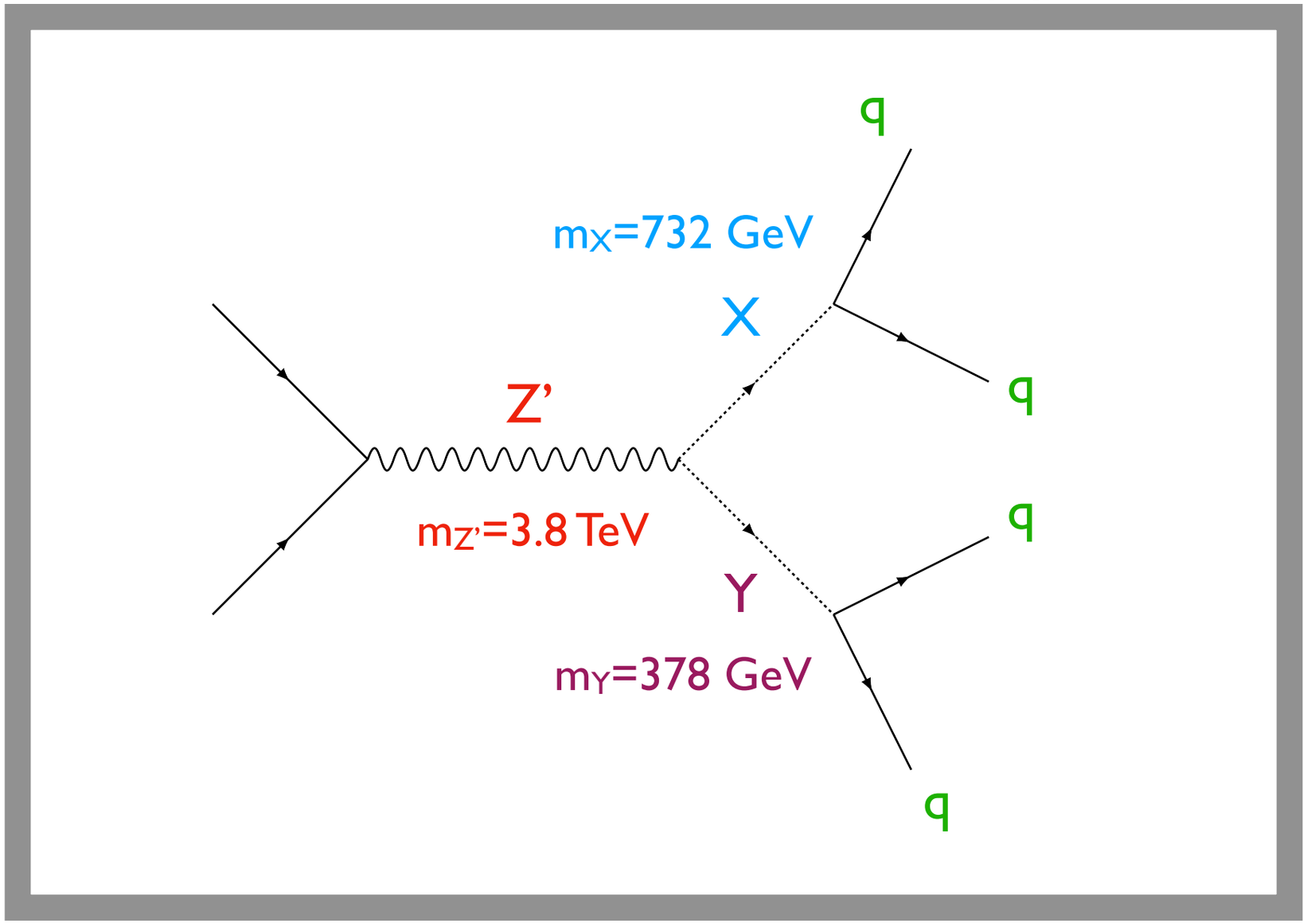
Results on the blinded* challenge

Black Box 1

*There are a lot more results in the summary paper (!)

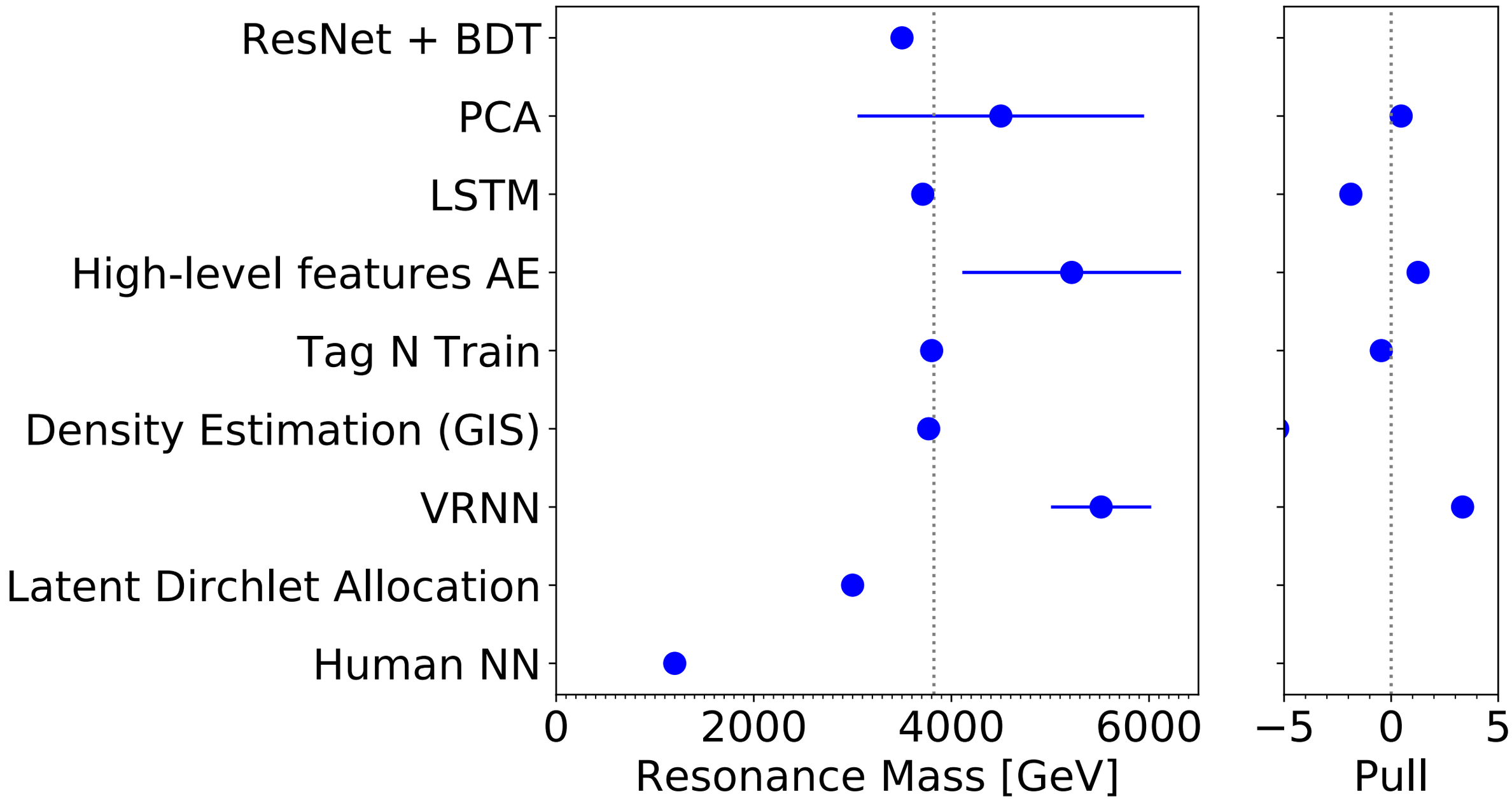
Results on the blinded challenge

Black Box 1



834 events. Same topology as R&D dataset
(not known to participants)

Results on the blinded challenge



…and the winners are…

We refrained from ranking the results … all methods are an important contribution to this growing research area.

However, two submissions clearly stood out:

Conditional density estimation for anomaly detection (GIS)

George Stein, Uroš Seljak, Biwei Dai, He Jia

BERKELEY CENTER *for* COSMOLOGICAL PHYSICS

Tag N' Train (CWoLa + Autoencoder)

Oz Amram and Cristina Mantilla Suarez

JOHNS HOPKINS UNIVERSITY

Results on the blinded challenge

Black Box 2

Results on the blinded challenge

Black Box 2

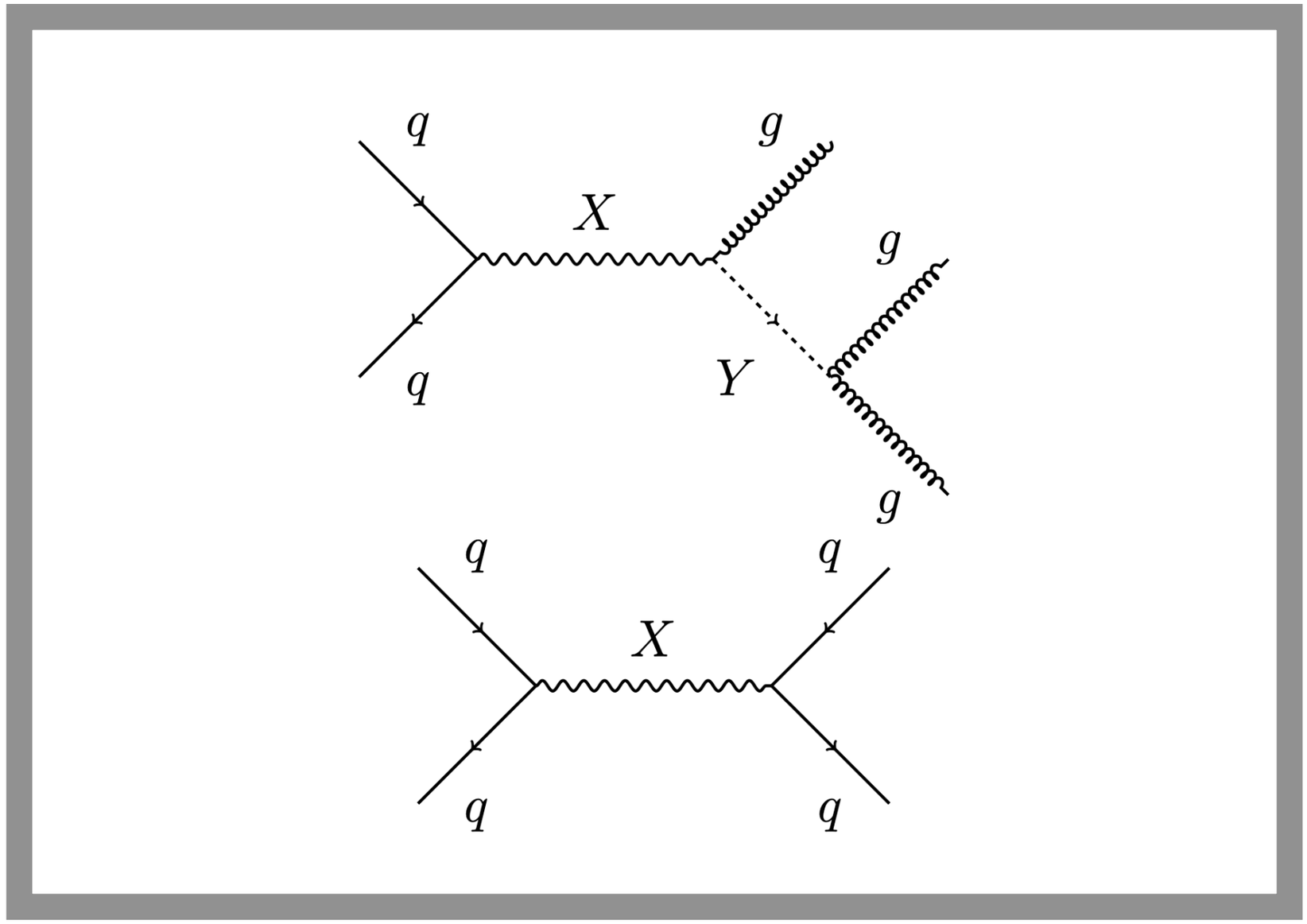
Empty! Multiple teams still reported (fake) signals…

Results on the blinded challenge

Black Box 3

Results on the blinded challenge

Black Box 3



This one was difficult - there are **two decay modes**. Without finding both, the signal is not significant. **No one** convincingly found this signal (**more work** is needed!)

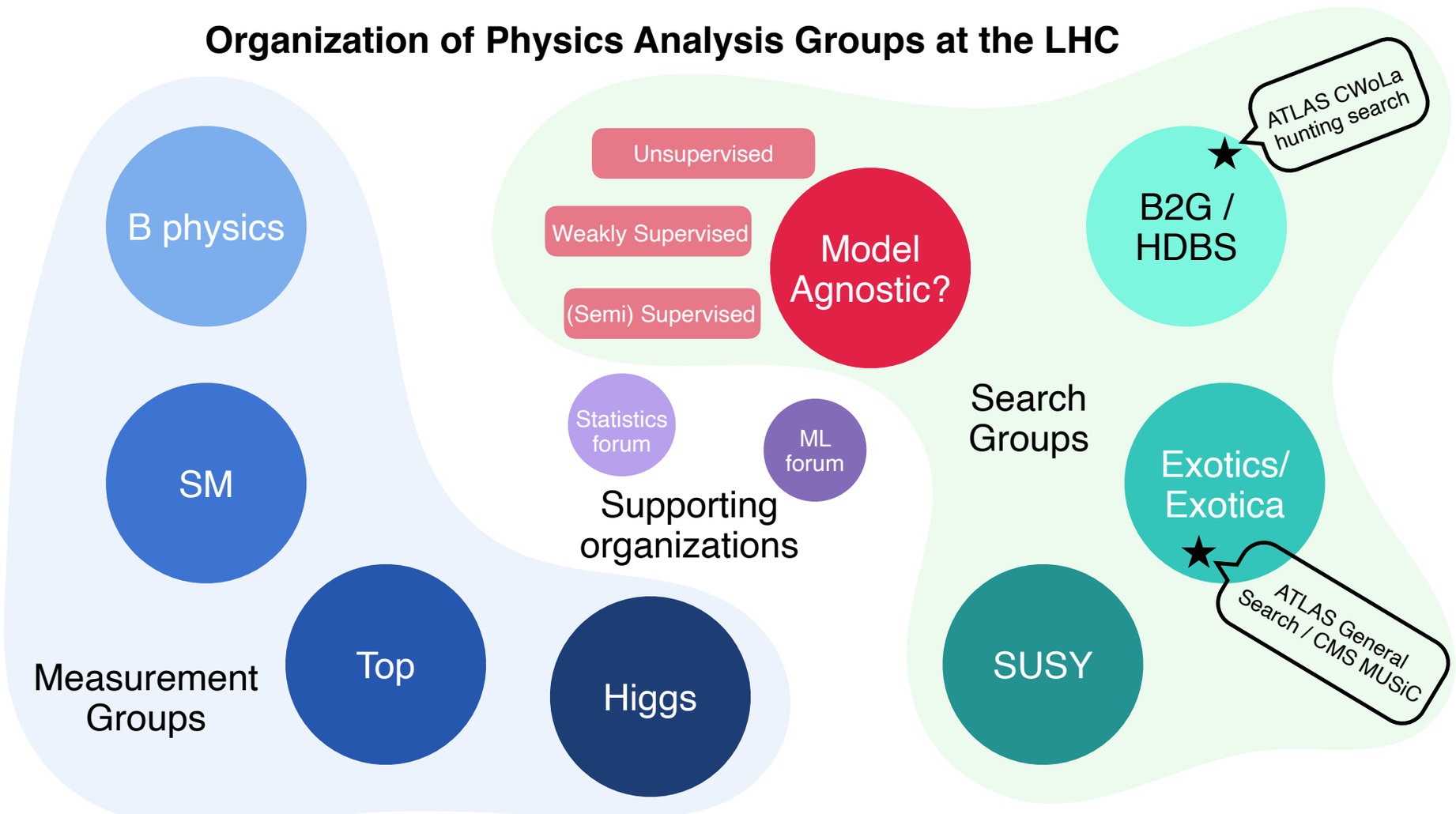
Results

That was a quick overview - please see the paper for many more details. Each section has the same format and is relatively brief (often further details are provided in standalone papers)

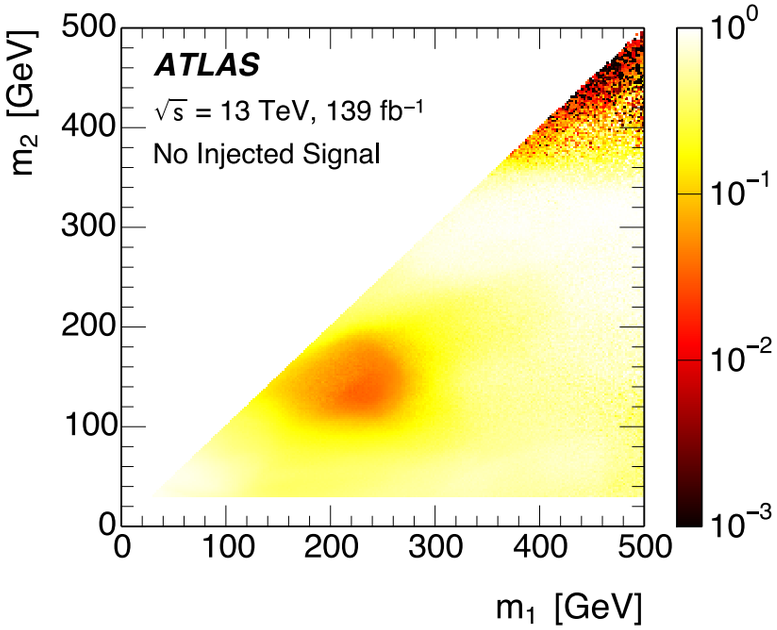
Section	Short Name	Method Type	Results Type
3.1	VRNN	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.2	ANODE	Unsupervised	(iii)
3.3	BuHuLaSpa	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.4	GAN-AE	Unsupervised	(i) (BB2-3) and (ii) (BB1)
3.5	GIS	Unsupervised	(i) (BB1)
3.6	LDA	Unsupervised	(i) (BB1-3)
3.7	PGA	Unsupervised	(ii) (BB1-2)
3.8	Reg. Likelihoods	Unsupervised	(iii)
3.9	UCluster	Unsupervised	(i) (BB2-3)
4.1	CWoLa	Weakly Supervised	(ii) (BB1-2)
4.2	CWoLa AE Compare	Weakly/Unsupervised	(iii)
4.3	Tag N' Train	Weakly Supervised	(i) (BB1-3)
4.4	SALAD	Weakly Supervised	(iii)
4.5	SA-CWoLa	Weakly Supervised	(iii)
5.1	Deep Ensemble	Semisupervised	(i) (BB1)
5.2	Factorized Topics	Semisupervised	(iii)
5.3	QUAK	Semisupervised	(i) (BB2,3) and (ii) (BB1)
5.4	LSTM	Semisupervised	(i) (BB1-3)

Outlook

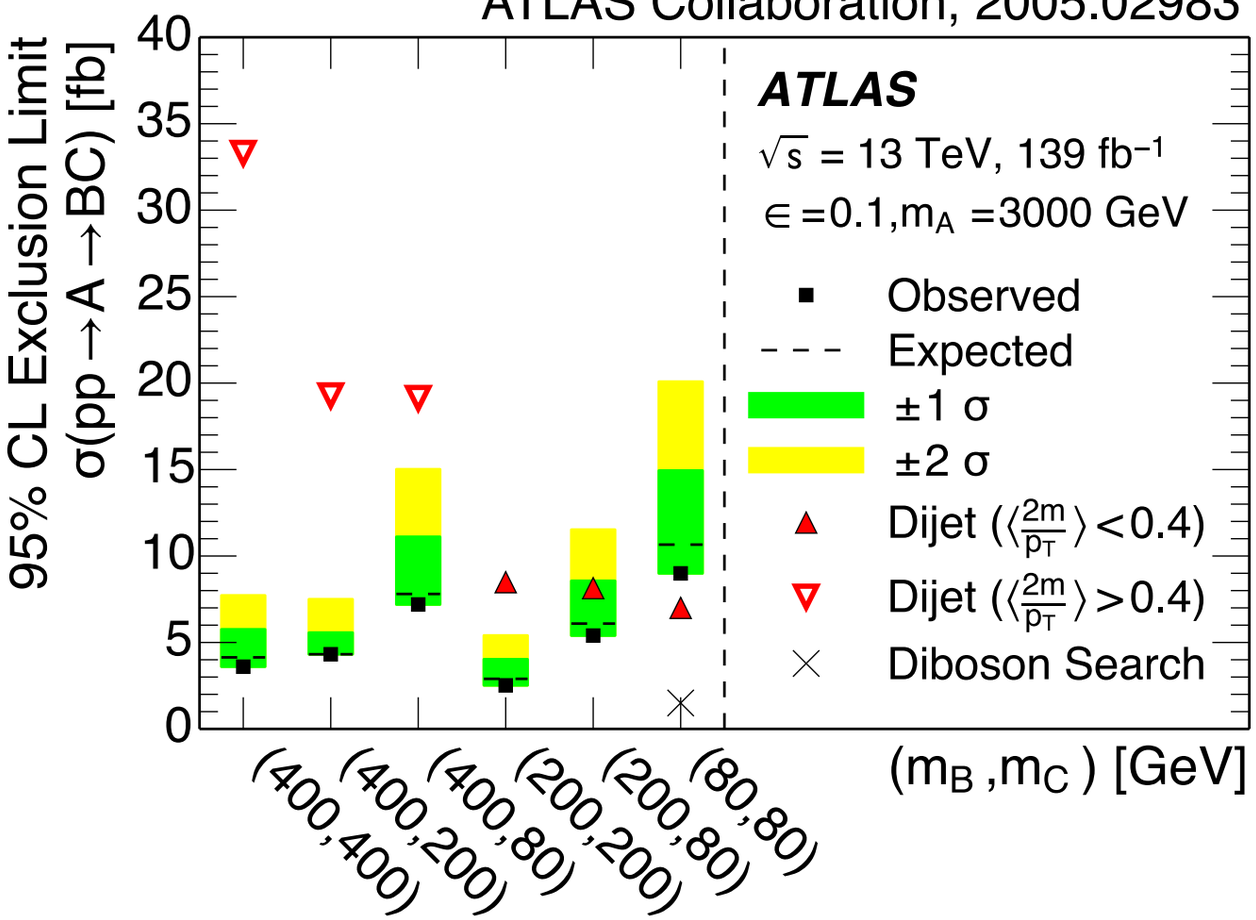
This challenge identified new methods and exposed gaps in coverage. It is clear that we will need **multiple approaches** to achieve broad coverage.



Outlook: the path to LHC data



ATLAS Collaboration, 2005.02983



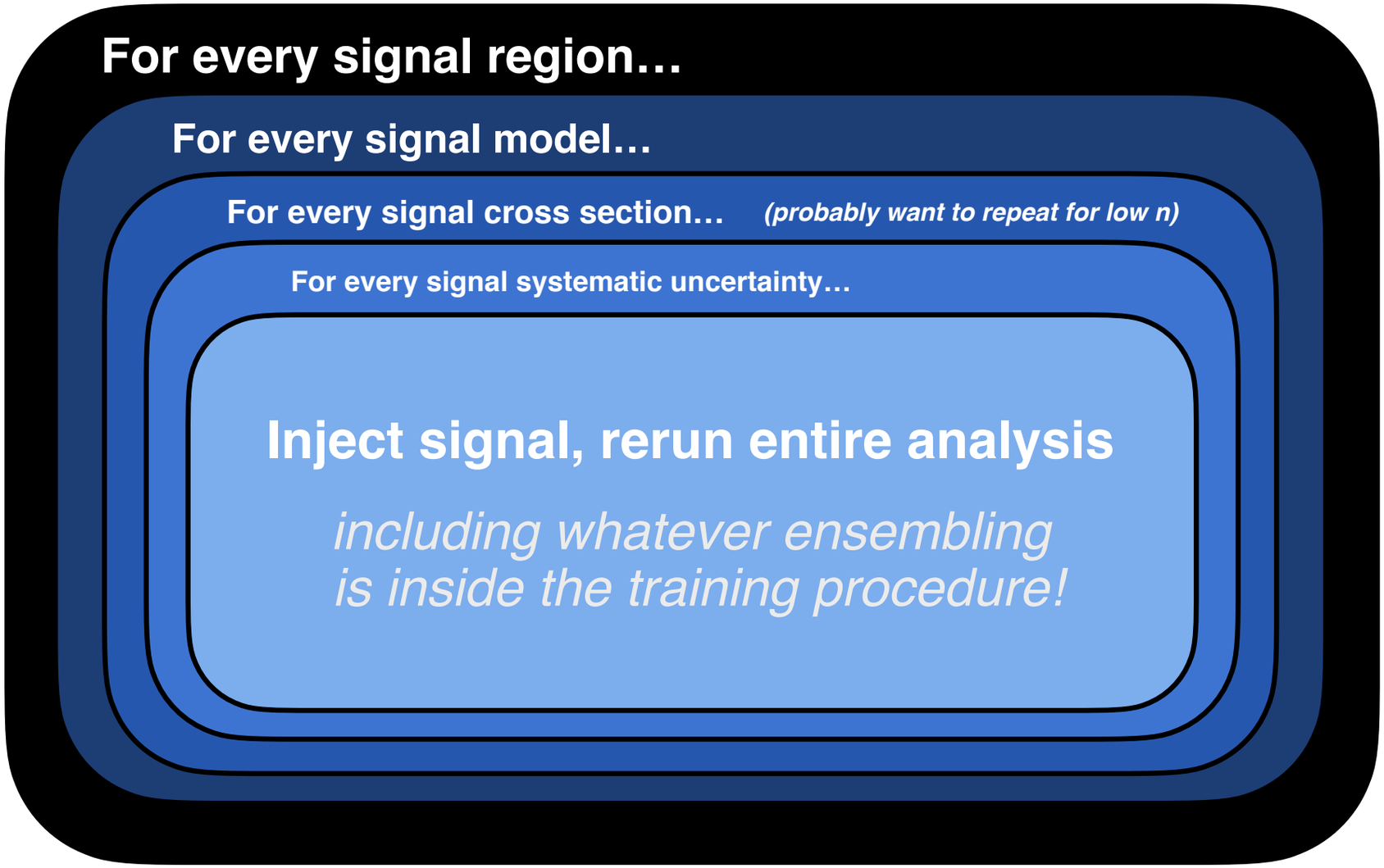
While we clearly need more R&D, it is also time to start deploying these methods on LHC data !

There are always new challenges with real data

I am proud to say that ATLAS recently completed a first analysis and I am hopeful that this is the start of many more to come!

A last thought: reinterpretation

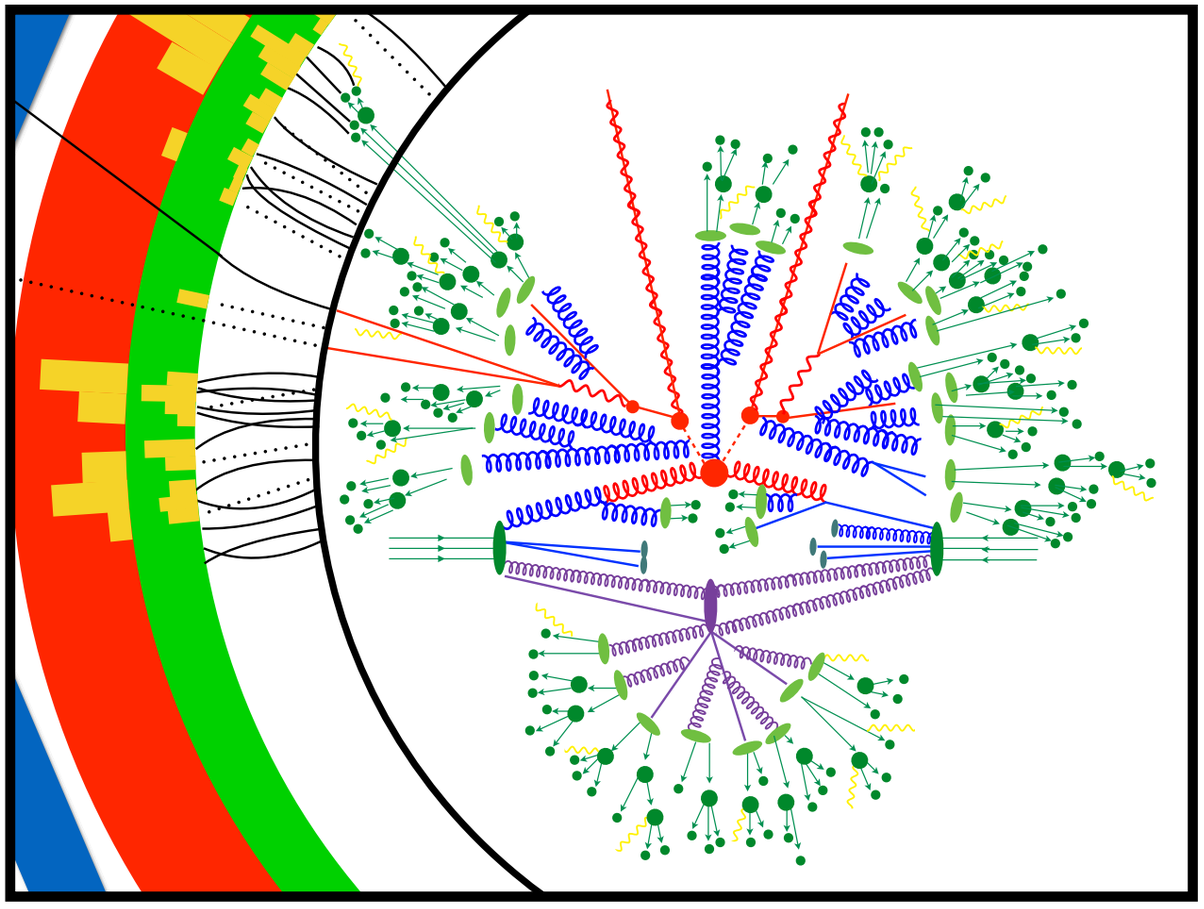
After all, this is a reinterpretation workshop …
these analyses are notoriously **difficult to reinterpret** as the "event selection" often depends on the data (unique to these analyses!).



This is an interesting challenge we will have to address in the future.

Summary

It is critical that we complement the current search program with **model agnostic methods**



Machine learning provides many powerful solutions, which were prototyped on the **LHC Olympics dataset**

…and these tools will increase the LHC discovery potential!

Questions?