LHC Olympics 2020

Gregor Kasieczka
gregor.kasieczka@uni-hamburg.de
Twitter: @GregorKasieczka

PHYSTAT-Anomalies, May 24/25 2022

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Emmy Noether-Programm
Deutsche Forschungsgemeinschaft DFG

PIER
Partnership of Universität Hamburg and DESY

Bundesministerium für Bildung und Forschung

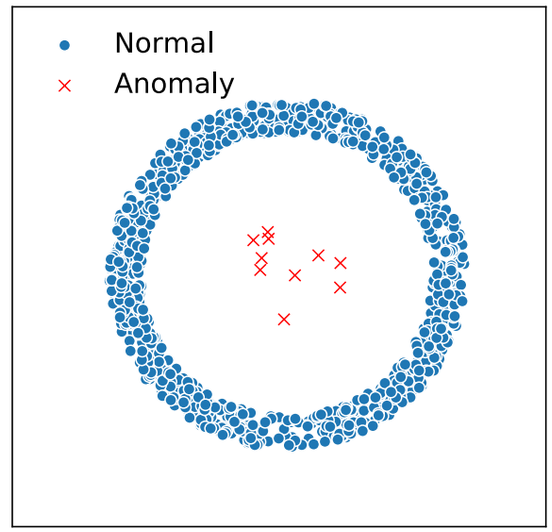
Overview

- At this point in the workshop, do not need to motivate searching for anomalies
- Introduce LHC Olympics: Ideas behind, setup, methods, and results
- Some other comments on open issues for anomaly detection

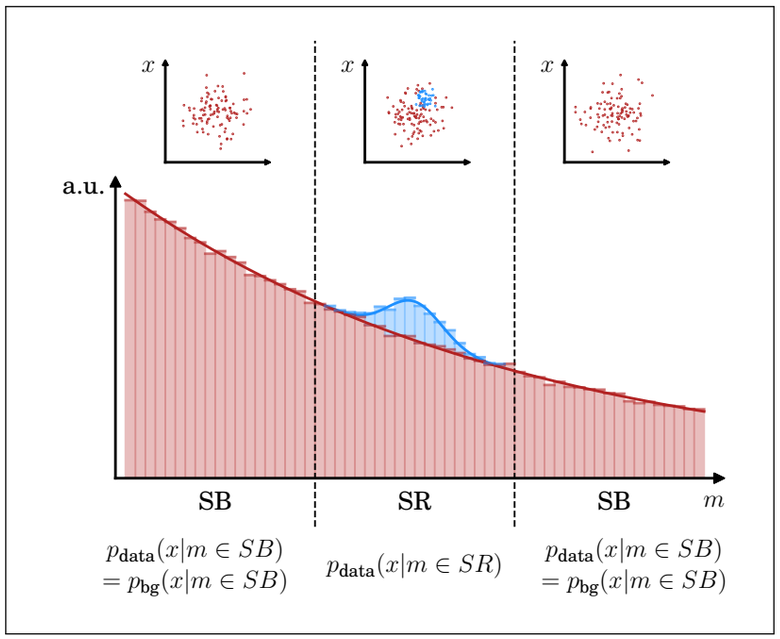
(With many thanks to Ben & David as LHCO co-organizers and all participants!!)

Types of anomalies

- **Outliers/Point anomalies**: Datapoints far away from regular distribution
- Examples:
  - Detector malfunctions
  - Background-free search



- **Group anomlies**: Individual examples not interesting, but signal is an overdensity with respect to background
- Examples:
  - Resonance searches
  - Transient signals in time series
- ***Focus of LHC Olympics***

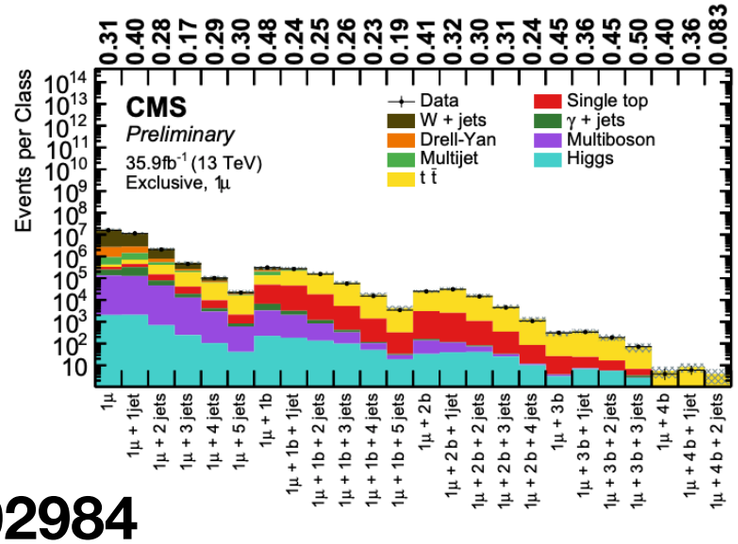


Approaches

Use simulation to estimate backgrounds?

Yes

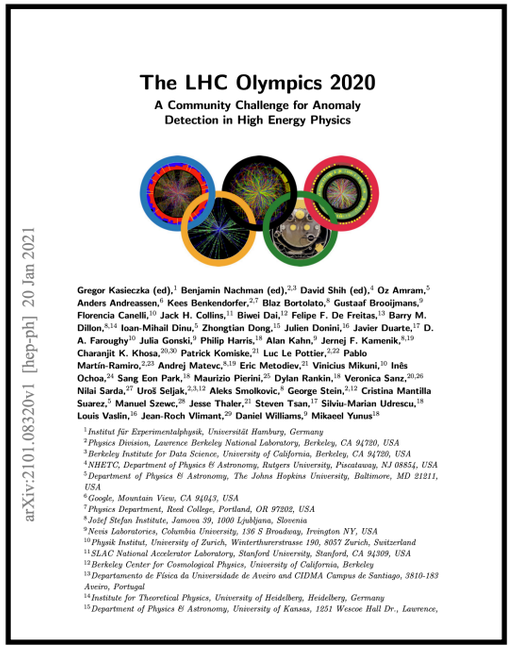
- Systematically compare simulation and recorded data, look for differences
- Con: Relies on imperfect simulation, Maximally background model dependent
- Pro: Sensitive to all types of anomalies
- e.g. MUSIC



arXiv: 2010.02984

No

- Estimate background from data
- Con: Need to make assumptions about signal model
- Pro: No reliance on simulation
- Focus of LHC Olympics


The LHC Olympics 2020

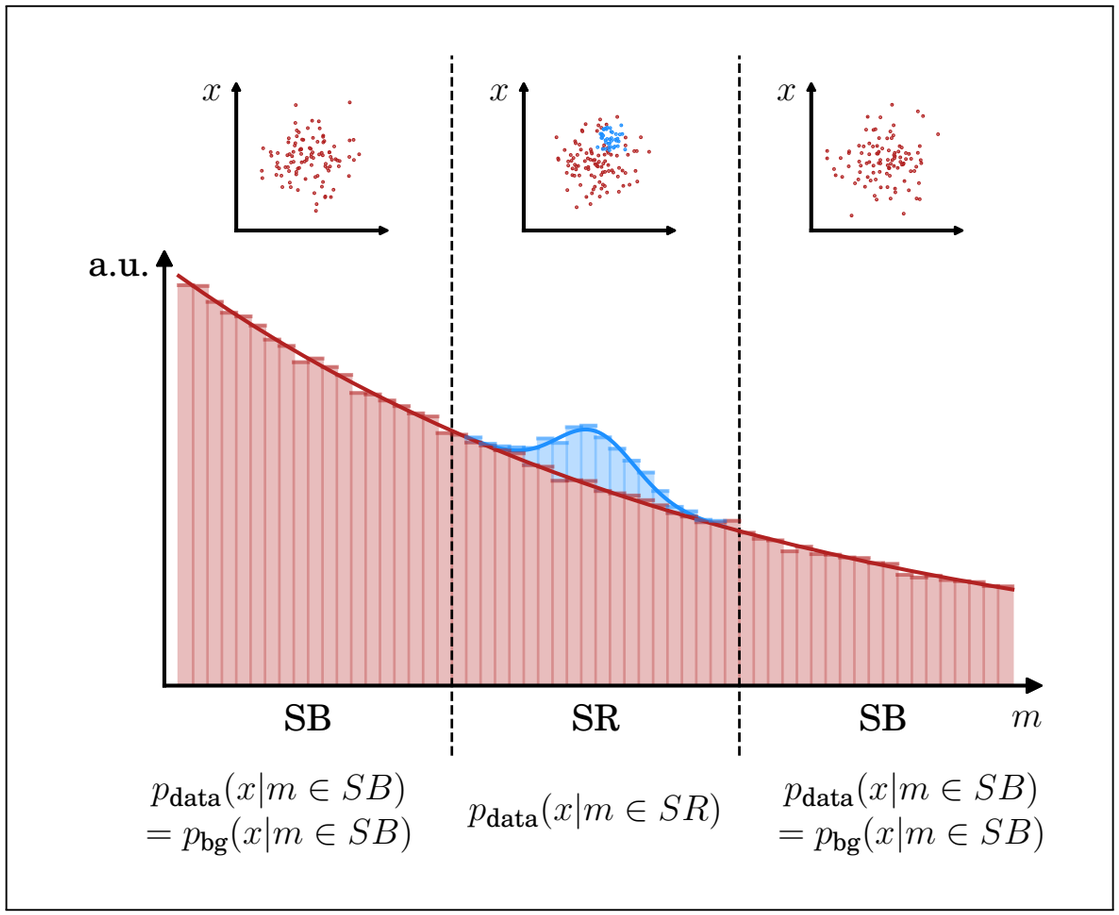
A Community Challenge for Anomaly Detection in High Energy Physics

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,[9] 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]

[1]Institut für Experimentalphysik, Universität Hamburg, Germany
[2]Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[3]Berkeley Institute for Data Science, University of California, Berkeley, CA 94720, USA
[4]NHETC, Department of Physics & Astronomy, Rutgers University, Piscataway, NJ 08854, USA
[5]Department of Physics & Astronomy, The Johns Hopkins University, Baltimore, MD 21211, USA
[6]Google, Mountain View, CA 94043, USA
[7]Physics Department, Reed College, Portland, OR 97202, USA
[8]Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia
[9]Nevis Laboratories, Columbia University, 136 S Broadway, Irvington NY, USA
[10]Physik Institut, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
[11]SLAC National Accelerator Laboratory, Stanford University, Stanford, CA 94309, USA
[12]Berkeley Center for Cosmological Physics, University of California, Berkeley
[13]Departamento de Física da Universidade de Aveiro and CIDMA Campus de Santiago, 3810-183 Aveiro, Portugal
[14]Institute for Theoretical Physics, University of Heidelberg, Heidelberg, Germany
[15]Department of Physics & Astronomy, University of Kansas, 1251 Wescoe Hall Dr., Lawrence,

Assumptions

- **Rarity:** Pr(anomaly) ≪ Pr(normal)
- **Overlap:** max x $p(x|\text{anomaly})/p$ $(x|\text{normal}) < \infty$
- **Resonance:** Pr($|m - m0| > \delta|$anomaly) ≈ 0 for some feature m (often a mass) and fixed $m0$, δ
- **Smoothness:** p $(x|m$, normal) varies slowly with m so that one can use data with $|m - m0| > \delta$ to estimate $p(x|m,\text{normal})$ for $|m - m0| < \delta$



Overall Strategy

- Define an anomaly score **a**
  - Should be high for anomalous (signal-like) and low for background-like data

- Use a selection **a** to create an anomaly-enriched dataset

- Estimate background from data

- Compare predicted background to number of observed with high **a** (+potentially other selection criteria)

Idea behind LHC Olympics: Provide dataset that allows exercising all these steps

History

- Ancient LHC Olympics
  - Four workshops 2005—2007 / prior starting to LHC
  - Assumed new physics would be plenty, focus on characterisation of new physics models

First LHCO lost to sands of time:
http://ph-dep-th.web.cern.ch/ph-dep-th/content2/workshops/lhcOlympics/lhcolympicsI.html

CERN Accelerating science — Sign in — Directory

Department of Theoretical Physics — ABOUT — PEOPLE — EVENTS — VISIT — TH GUIDE — EDI — JOBS

Page not found

Page not found, or something went wrong with the web server.

4th LHC Workshop

Princeton, 22 March, 2007

Information

The LHC Olympics is a collective effort by theorists to train themselves in establishing a correspondence between theoretical models and experimental signatures using collider simulation. This is done by developing and distributing user-friendly versions of simulation and data analysis tools, and via a series of black box exercises, in which participants are challenged to disentangle simulated LHC data sets. The first three LHC Olympics workshops were held at CERN in July 2005 and February 2006, and the KITP in August 2006 .

The fourth LHC Olympics meeting will take place during the workshop "Physics at LHC: From Experiment to Theory," held from March 21-24, 2007, in Princeton. The workshop will bring together high energy physicists with expertise ranging from formal theory to collider phenomenology, simulation and experiment. The meeting is held jointly with the second workskop in the "Monte Carlo Tools for Beyond the Standard Model Physics" (MC4BSM) series.

Data Challenge

Using event generators in combination with the PGS-4 detector simulation, a number of black box data sets have been generated. The boxes are presented as a data challenge to participants, who are invited to try to figure out the underlying theoretical model. This round of the LHCO has two new black box data sets.

A) The SLAC LHC Olympics Black Box contains signal only -- no background -- for 20 fb^-1 of LHC data. The black box' creators are John Conley, Michael Peskin, and Tommer Wizansky.

B) The Cornell/Harvard LHC Olympics Black Box represents about 10 fb^-1 of new physics data. This black box has been created by Patrick Meade, Peter Onyisi, Maxim Perelstein, and Matt Reece.

An explanation of the LHCO format and suggestions for how to get started on analyzing them can be found on the black box web pages the LHCO wiki . The login for the wiki is "olympian" and password is "blackbox." All LHCO participants are free to use the wiki to add comments, ask questions, make suggestions, etc.

The programs used to generate the boxes are in continuing development. To keep apprised of possible updates to the boxes, it is advisable to regularly check the BB webpages and the LHCO wiki for announcements.

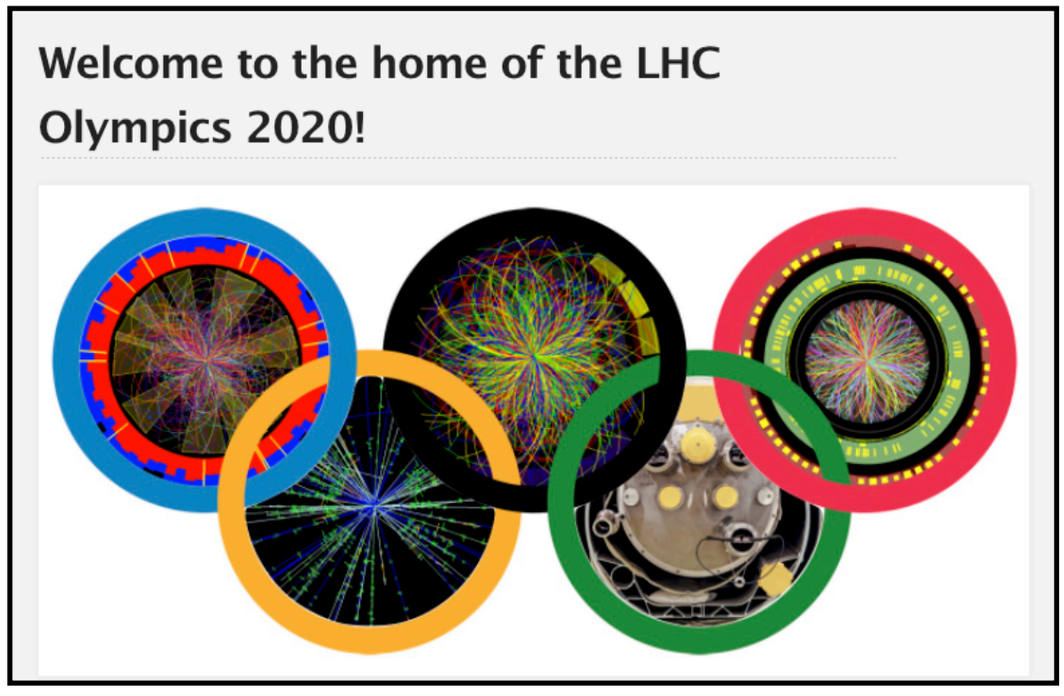
Motivation

- Encourage development and comparison of model-agnostic search strategies
    - Focus on group anomalies, data-driven searches

- Provide a complete package, balance details vs accessiblity

- Datasets:
    - One R&D dataset for algorithm development
    - Three black box datasets (BB1-BB3)
        - Unblinded over time

- Timeline:
    - Spring 2019: Release R&D dataset (link)
    - Autumn 2019: Release BB datasets (link)
    - January 2020: Winter Olympics as part of ML4Jets, unblinding of BB1 (link)
    - July 2020: (Virtual) Summer Olympics, unblinding of BB2 and BB3 (link)
    - LHC Olympics paper (https://arxiv.org/abs/2101.08320) public



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

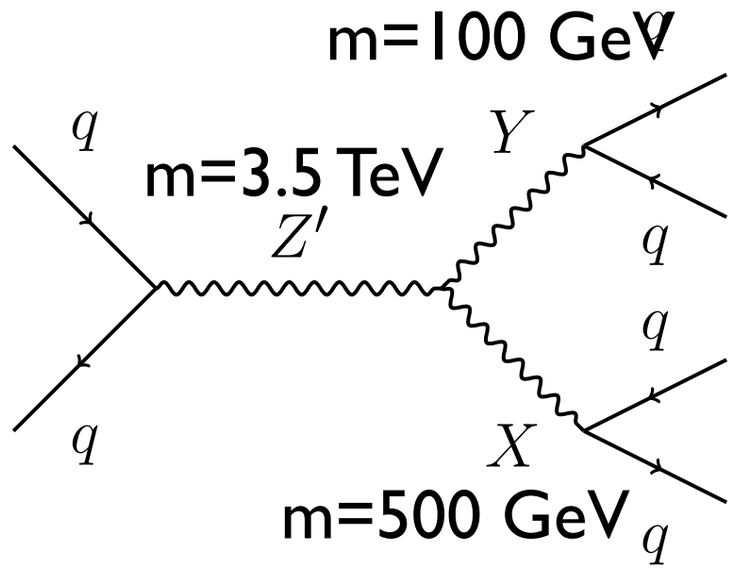
Datasets

- Data format:
  - 3-vectors of reconstructed particles in the event
    - pt, eta, phi, (m assumed to be zero)
    - Leading 700 particles (zero-passed otherwise)
    - -> **2100 dimensional input space**
  - Also provided a much lower dimensional representation
    - Clustering into two jets and using mass/substructure
    - O(10) dimensional input space
- No other quantities (e.g. flavor tagging) included
- Single R=1 jet trigger pT>1.2 TeV
- Generation with Pythia/Herwig; detector simulation with Delphes

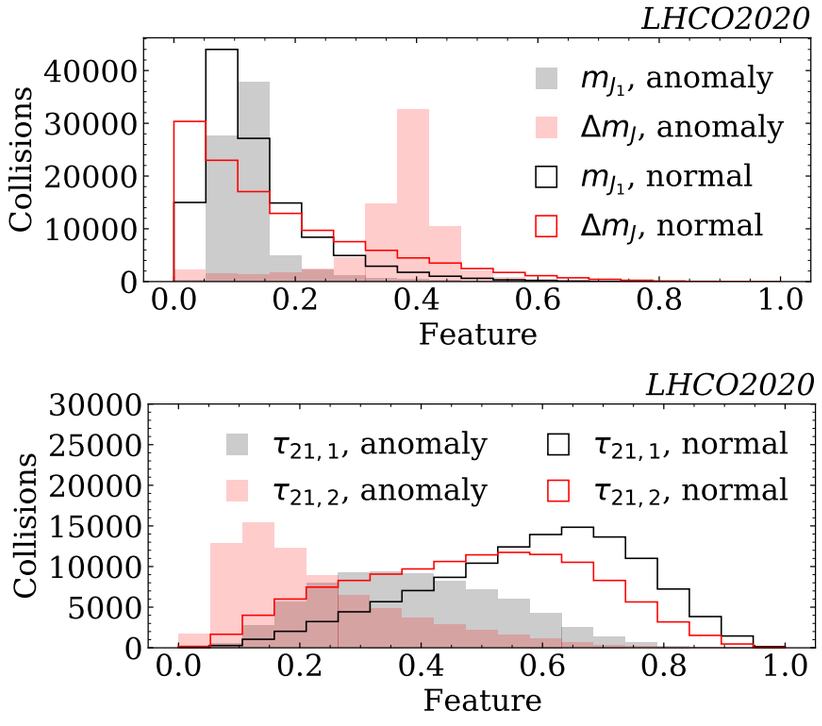
Setting	R&D	BB1	BB3
Tune:pp	14	3	10
PDF:pSet	13	12	5
TimeShower:alphaSvalue	0.1365	0.118	0.16
SpaceShower:alphaSvalue	0.1365	0.118	0.16
TimeShower:renormMultFac	1	0.5	2
SpaceShower:renormMultFac	1	0.5	2
TimeShower:factorMultFac	1	1.5	0.5
SpaceShower:factorMultFac	1	1.5	0.5
TimeShower:pTmaxMatch	1	2	1
SpaceShower:pTmaxMatch	0	2	1

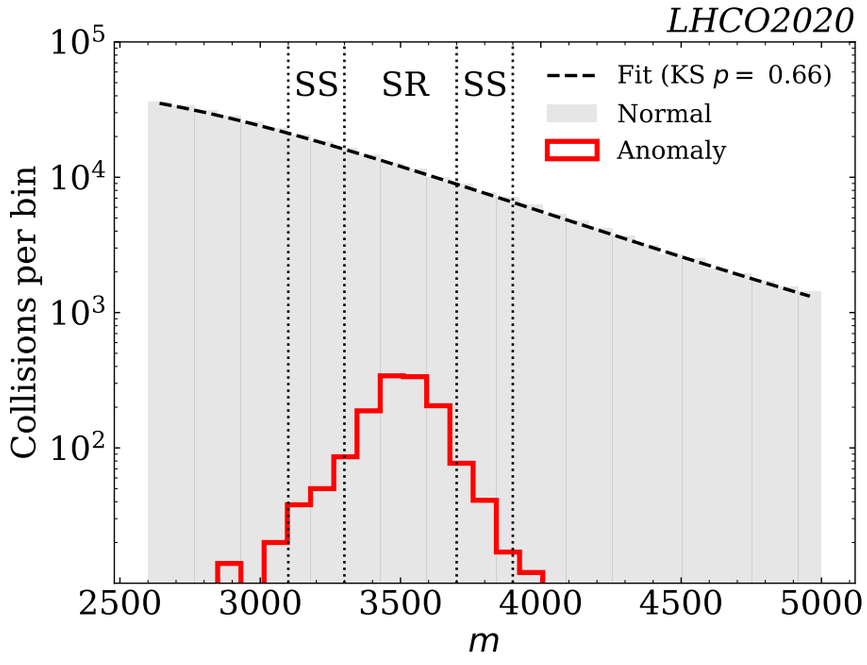
R&D dataset

- For building and testing methods
- 1M background examples (Standard Model), 100k signal examples (signal, see Feynman diagram on the right)
- Labels provided



- Relatively simple signal
  - Known to differ in previously mentioned features from background distribution
- Unrealistically high S/B





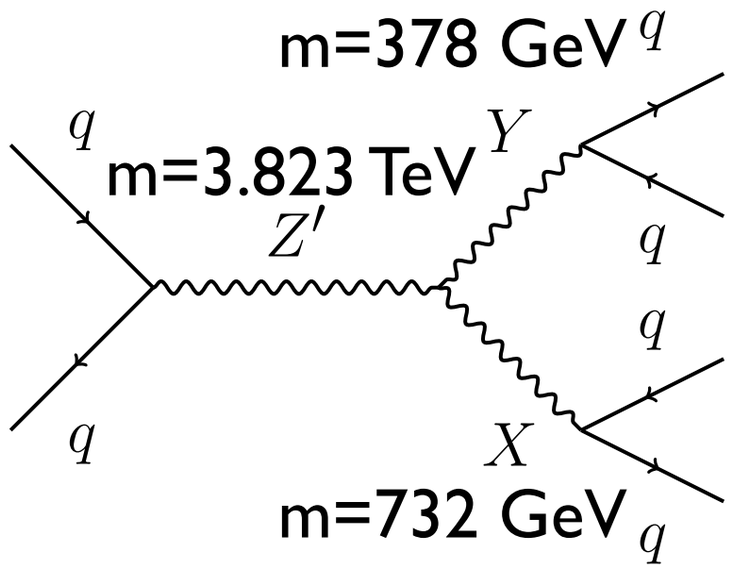
2107.02821

Challenge datasets: BB1

- All contain total of 1M examples; might contain signal;
  no labels provided during ‘content’ phase (labels available no)
- All used different simulation parameters for background (to avoid unrealistic exploits)

- BB1: 834 signal examples
  Same event topology as R&D dataset, different masses

  might be easy?



Challenge datasets: BB2

- All contain total of 1M examples; might contain signal;
  no labels provided during ‘content’ phase (labels available no)
- All used different simulation parameters for background (to avoid unrealistic exploits)
- Additional pure-background sample provided (again with a different tune)

- BB2: 0 signal examples; Herwig++ instead of Pythia for background

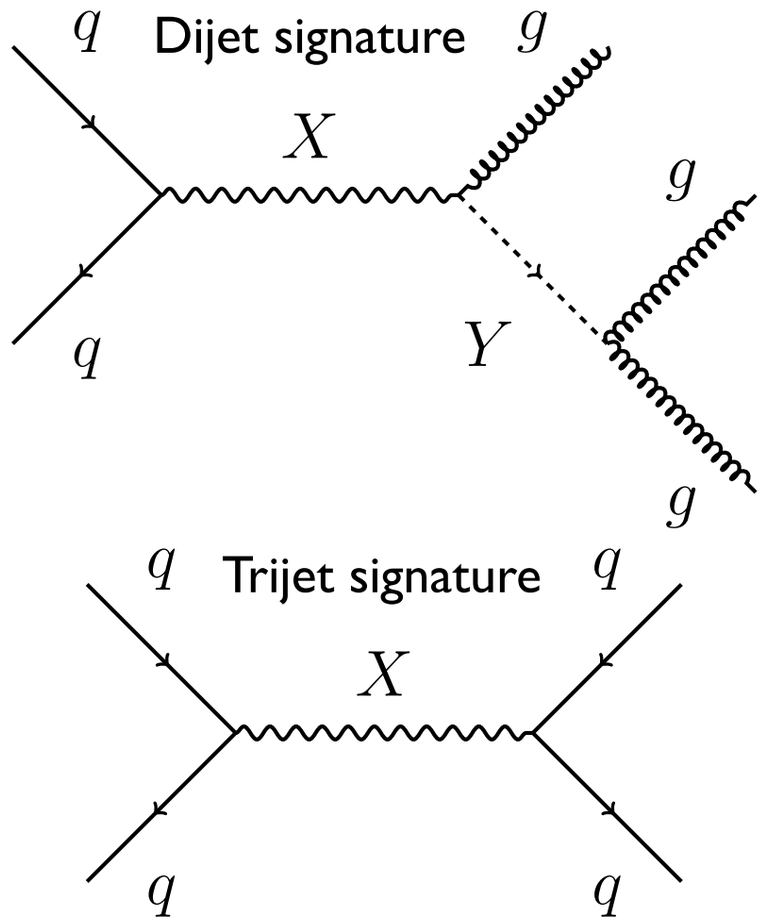
test for false positives

Challenge datasets: BB3

- All contain total of 1M examples; might contain signal;
  no labels provided during ‘content’ phase (labels available no)
- All used different simulation parameters for background (to avoid unrealistic exploits)
- Additional pure-background sample provided (again with a different tune)

- BB3:
  mX = 4.2 TeV and two decay modes:
  1200 signal events in di-jet signature
  2000 signal events in tri-jet signature
  (finding individual excess should not yield significance)

should be challenging



Evaluation criteria

- **What you should report:**
  - A p-value associated with the dataset having no new particles (null hypothesis).
  - As complete a description of the new physics as possible. For example: the masses and decay modes of all new particles (and uncertainties on those parameters).
  - How many signal events (+uncertainty) are in the dataset (before any selection criteria).
  - Partial submissions in only a subset of the categories are welcome!

(Goal not to necessarily pick ‘one winner’ but to get a useful understanding of anomaly detection capabilities)

Overview of Methods

(No labels)

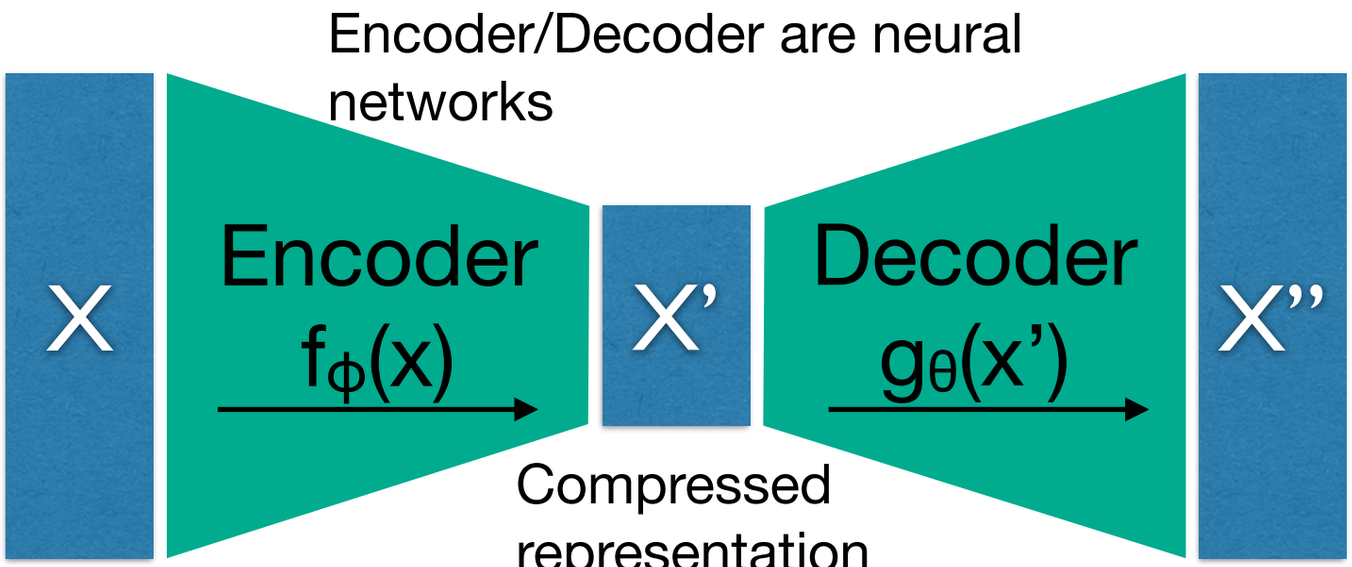
(Noisy labels)

(Partial / full labels)

Some examples and trends in the following.
For exhaustive discussion, refer to
2101.08320

Unsupervised - Autoencoders

- Several autoencoder-type learning approaches
- Underlying assumption is that an autoencoder trained on background dominated sample will have bad reconstruction performance for previously unseen signal



$$\mathcal{L}(x) = ||x - g_\theta(f_\phi(x))||_2$$

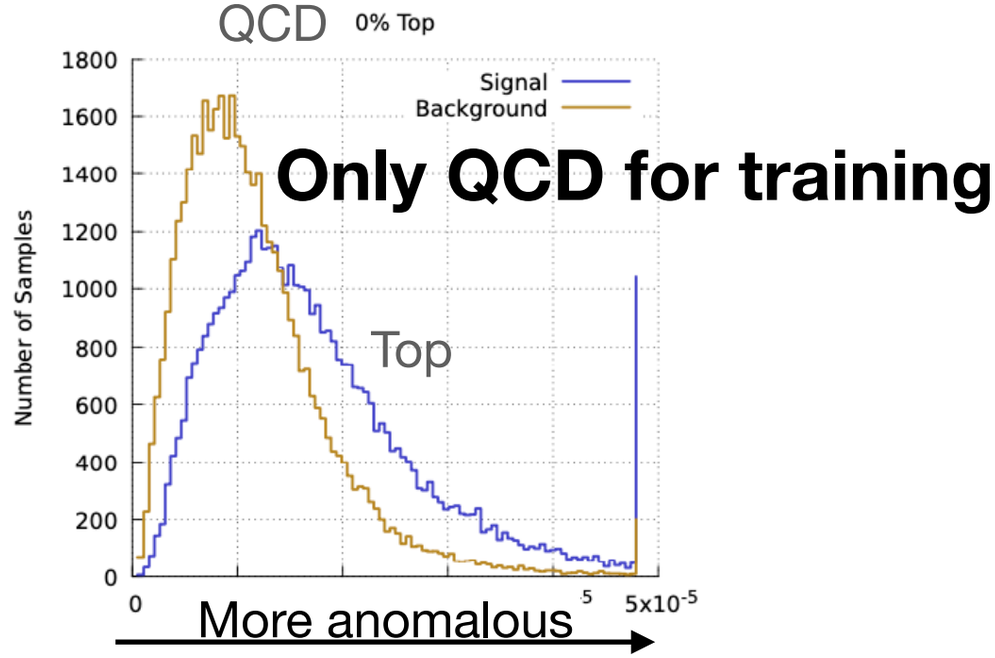
$$a(x) = \mathcal{L}(x)$$

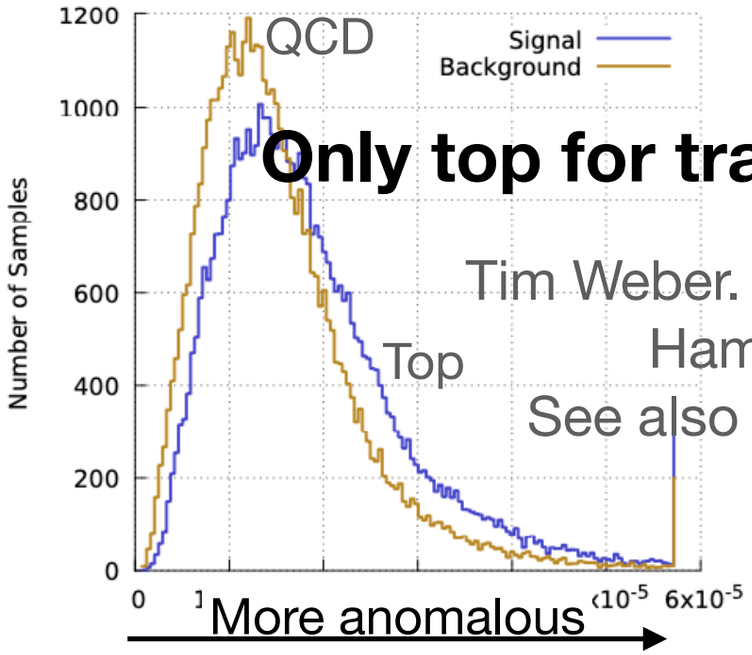
- Differences in data representation
  Data space vs latent space anomaly detection
  Different latent space prior distributions

Limitations

Complexity

- If anomalies are much simpler (therefore easier to reconstruct):
  a(x) will still be lower, despite never encountered in training
- Observed with naive AE in QCD vs top
  - Train on tops only; top still considered anomaly wrt/ QCD



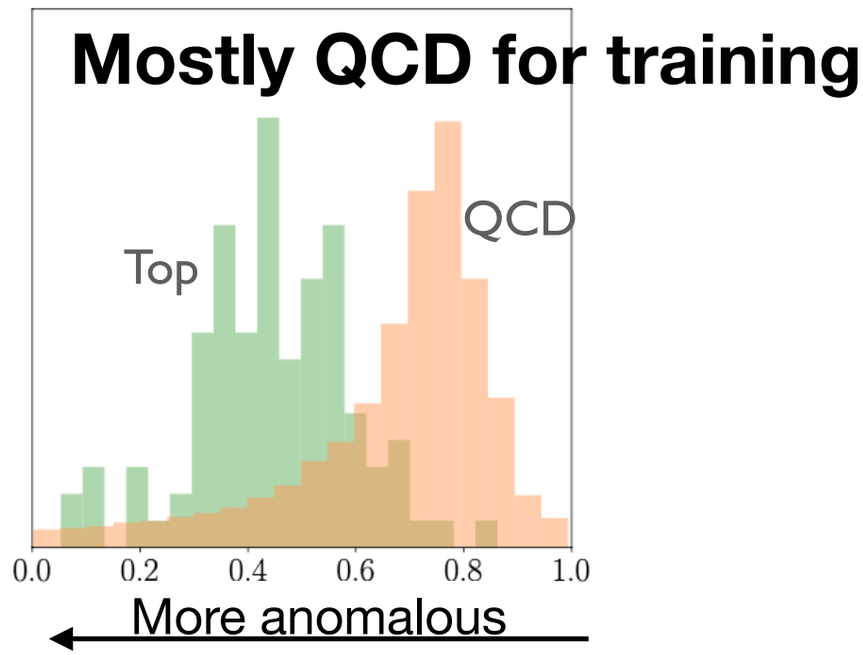


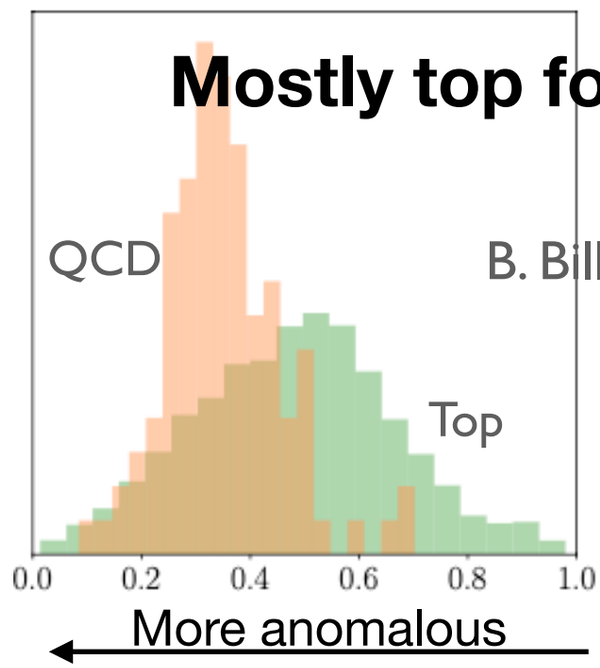
Tim Weber. MSc thesis. Hamburg, 2019;
See also 2104.09051

Limitations

Complexity

- If anomalies are much simpler (therefore easier to reconstruct):
  a(x) will still be lower, despite never encountered in training
- Observed with naive AE in QCD vs top
  - Train on tops only; top still considered anomaly wrt/ QCD
  - Can be overcome (e.g. by structuring the latent space)





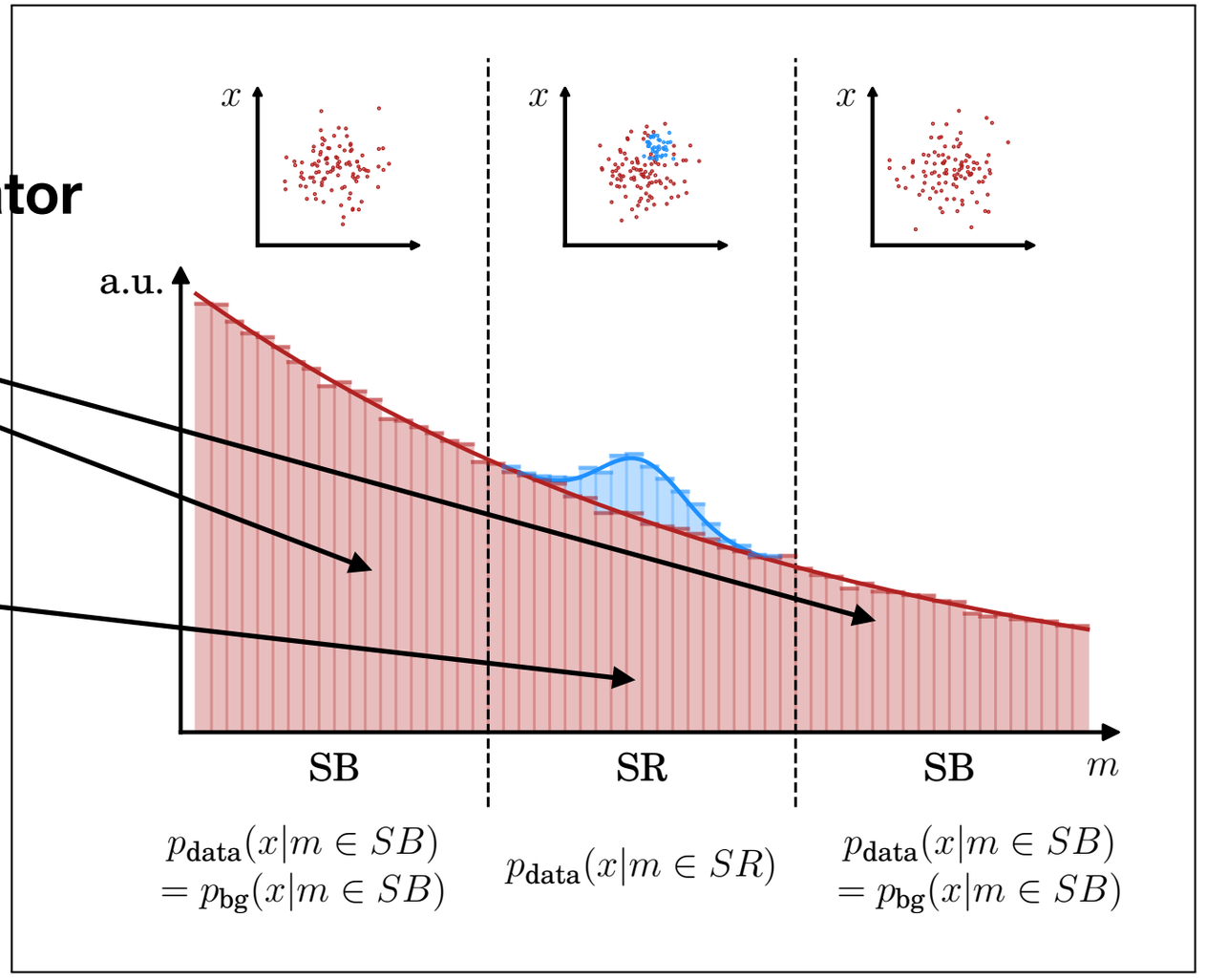
B. Billon, T Plehn, et al
2104.08291

Unsupervised - Density Estimation

Train density estimator here

and sample here

(Corresponds to training a background model on data!)

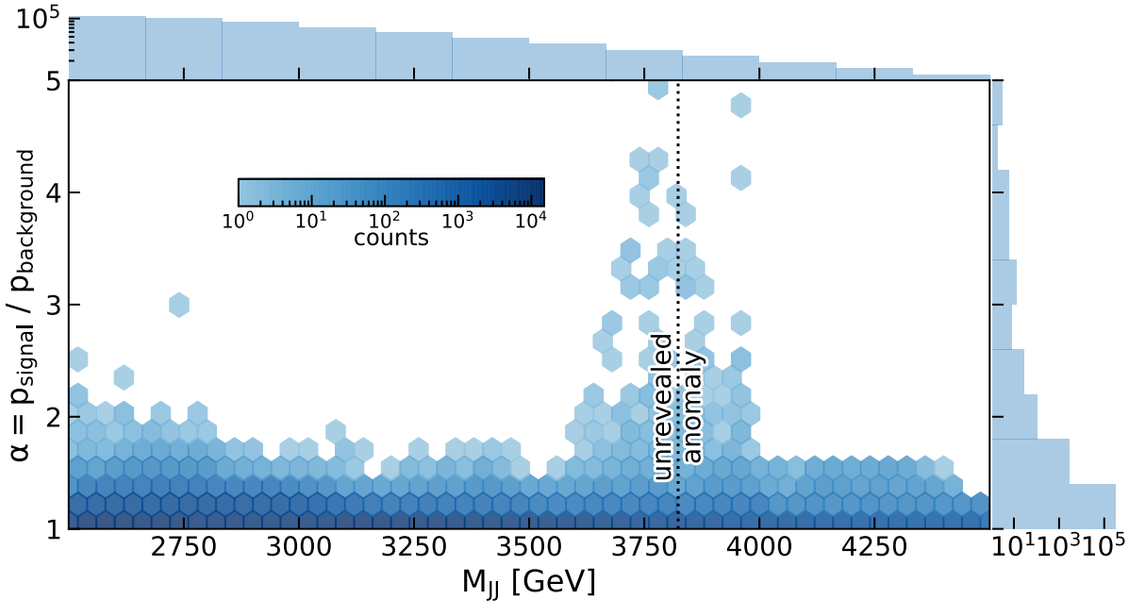


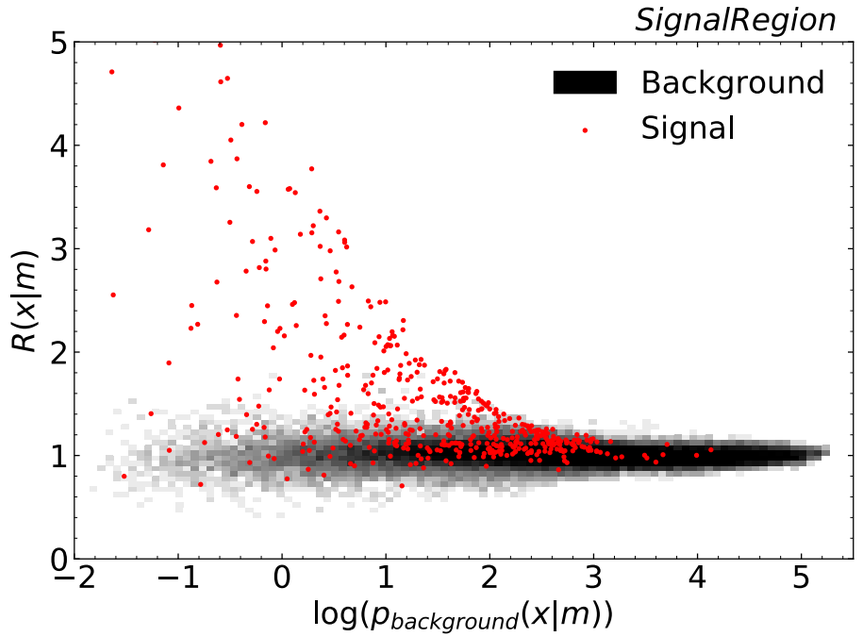
- Conditional transport and sampling
  - Train density estimator (e.g. conditional normalising flow) in sideband
  - Interpolate to signal region
  - Sample data there
  - This produces *‘extrapolated-background’*

Unsupervised - Density Estimation

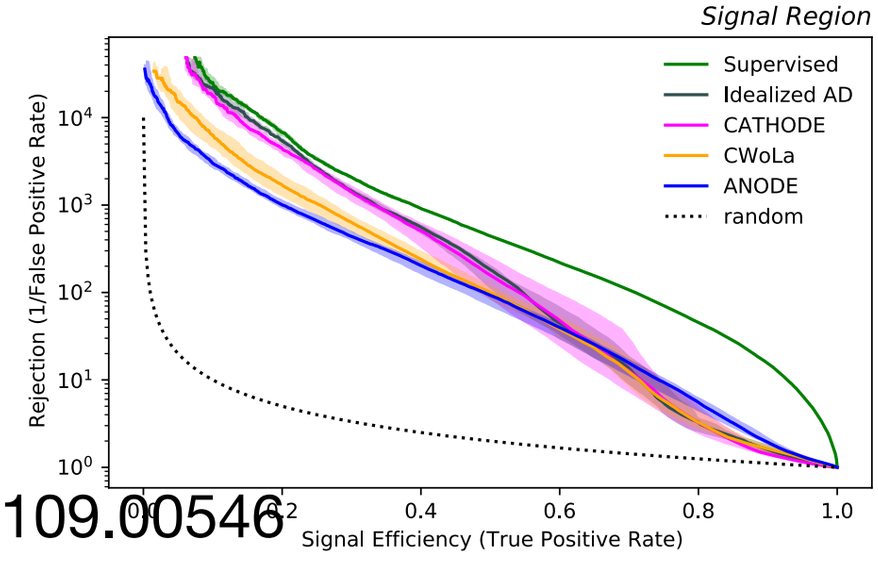
- Compare extrapolated-background to actual data
- Either by also training a density estimator in signal region and building the likelihood ratio (GIS-approach, ANODE)

$$R(x|m) = \frac{p_{\text{data}}(x|m)}{p_{\text{background}}(x|m)}.$$



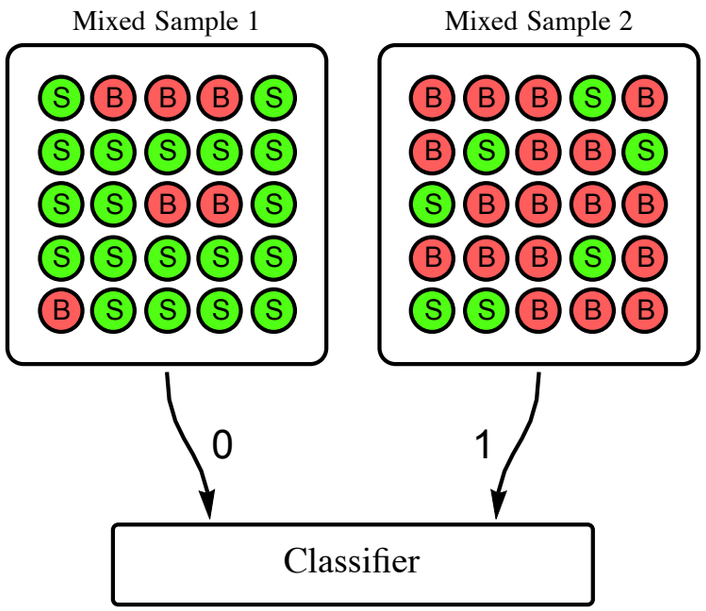


- Or by training a classifier between extrapolated-background and actual data (e.g. CATHODE / 2109.00546 or CURTAINS / 2203.09470) (post-LHCO)



2109.00546

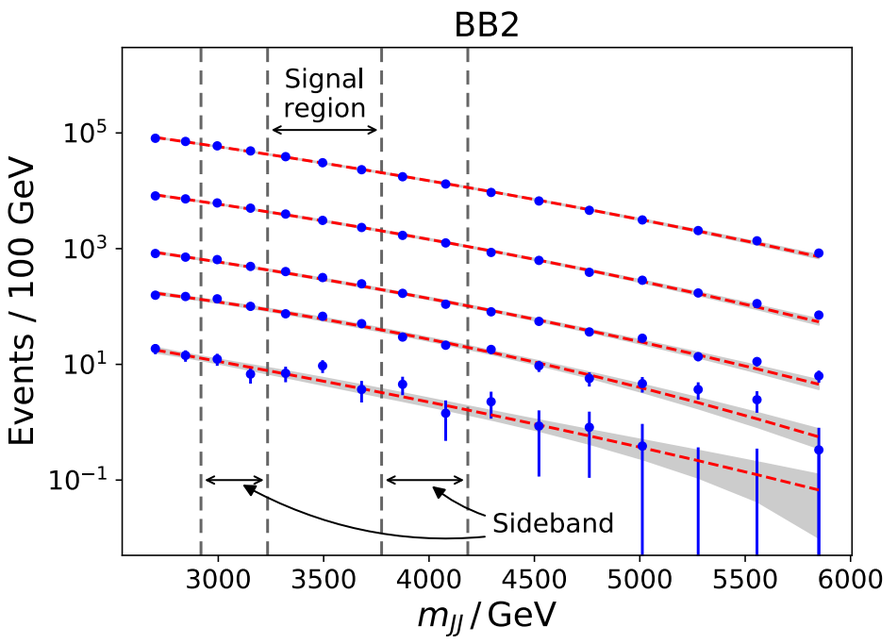
Weak Supervision: CWola Hunting

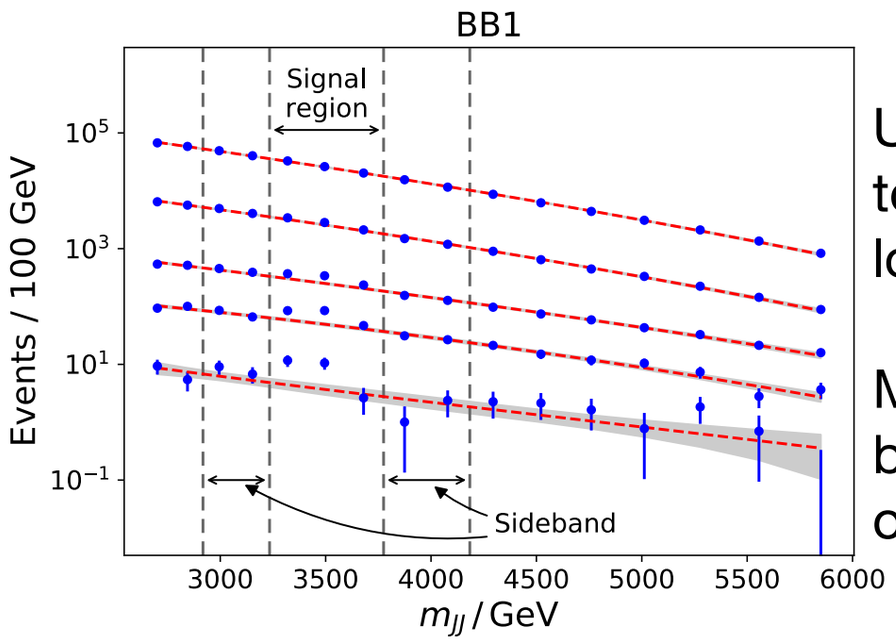


$$L_{M_1/M_2} = \frac{p_{M_1}}{p_{M_2}} = \frac{f_1\, p_S + (1-f_1)\, p_B}{f_2\, p_S + (1-f_2)\, p_B} = \frac{f_1\, L_{S/B} + (1-f_1)}{f_2\, L_{S/B} + (1-f_2)}$$

Important observation:

A classifier (i.e. a neural network) trained to distinguish two mixed samples learns to distinguish the components





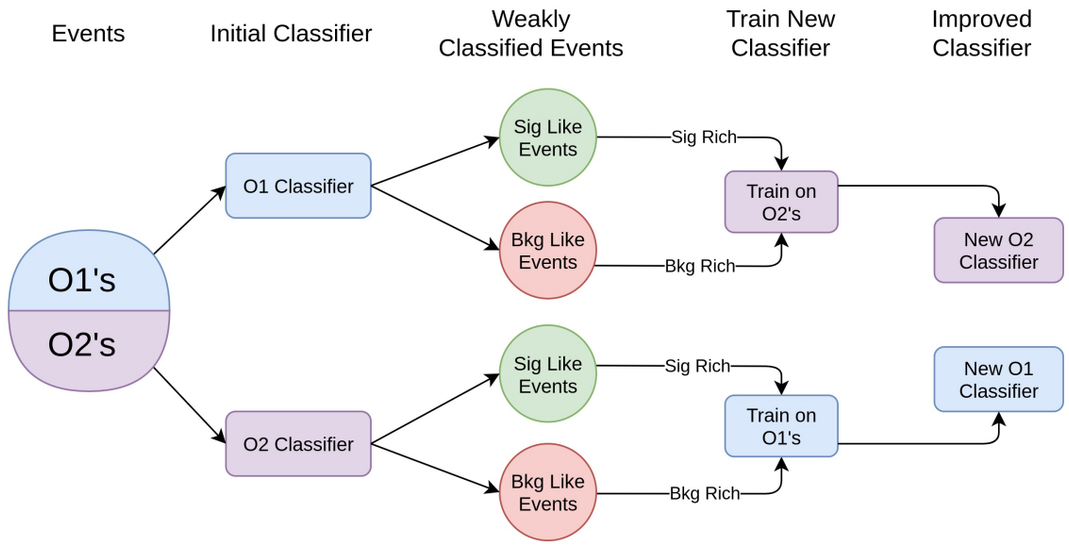
Use increasingly tight selections to identify localised signal.

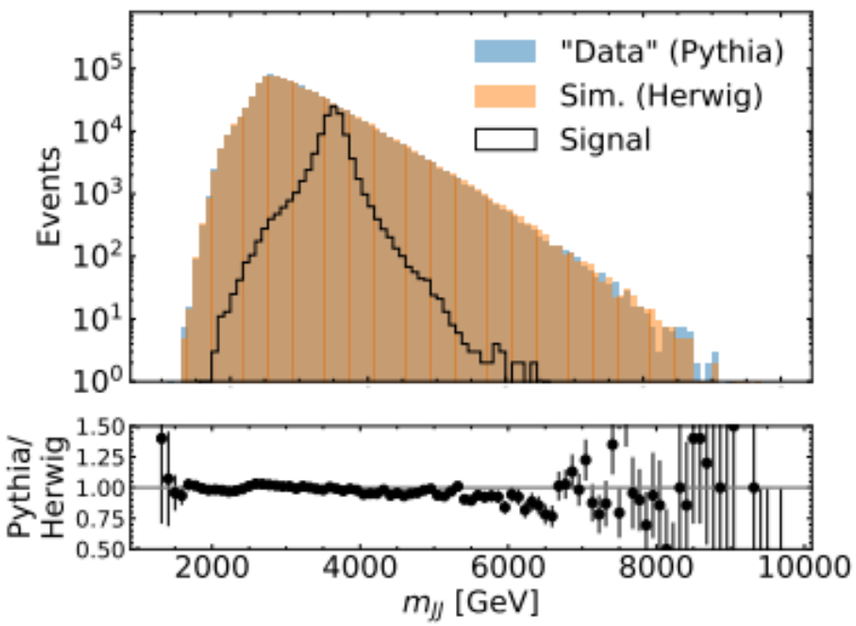
Major downside: Correlations between mass and other features

1708.02949, 1902.02634,
see also talk by Ines on ATLAS study

Weak Supervision: TNT and SALAD

TNT (2002.12376):
Interesting signal might contain two anomalous jets per event. Use per-jet classifiers to build enriched datasets for training.





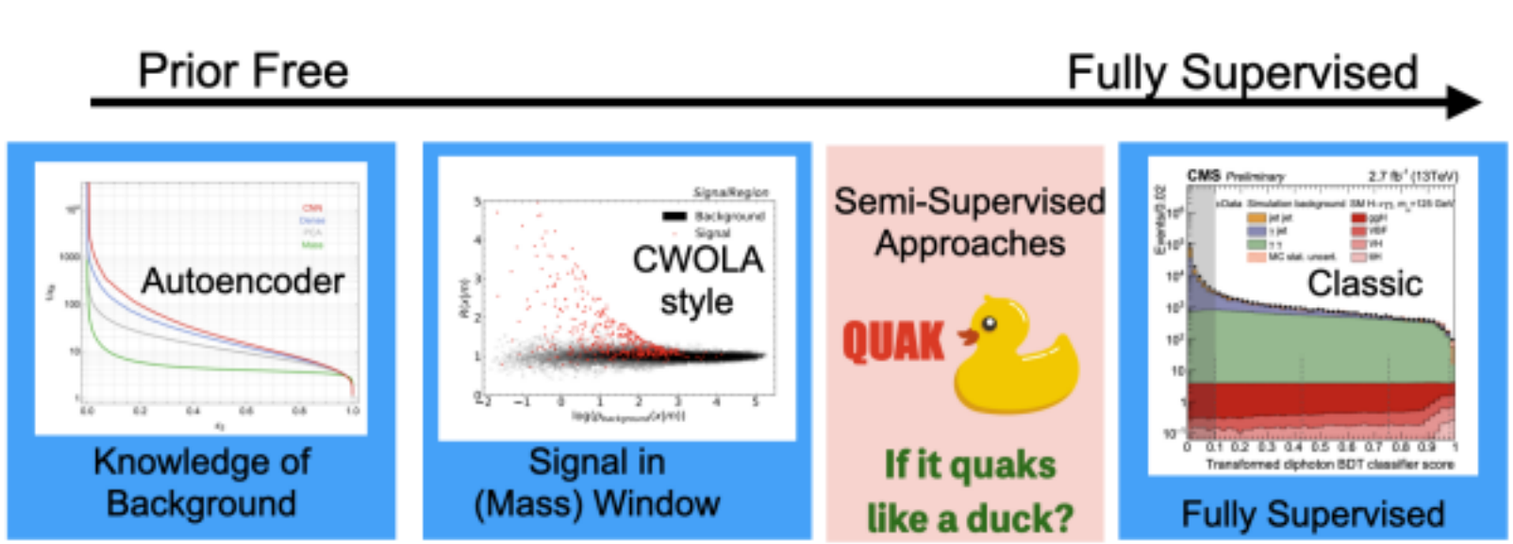
SALAD (2001.05001):
Use classifier-based reweighting (DCTR approach / 1907.08209) to learn mapping background simulation in sideband to data. Apply in signal-region and treat non-closure as anomaly

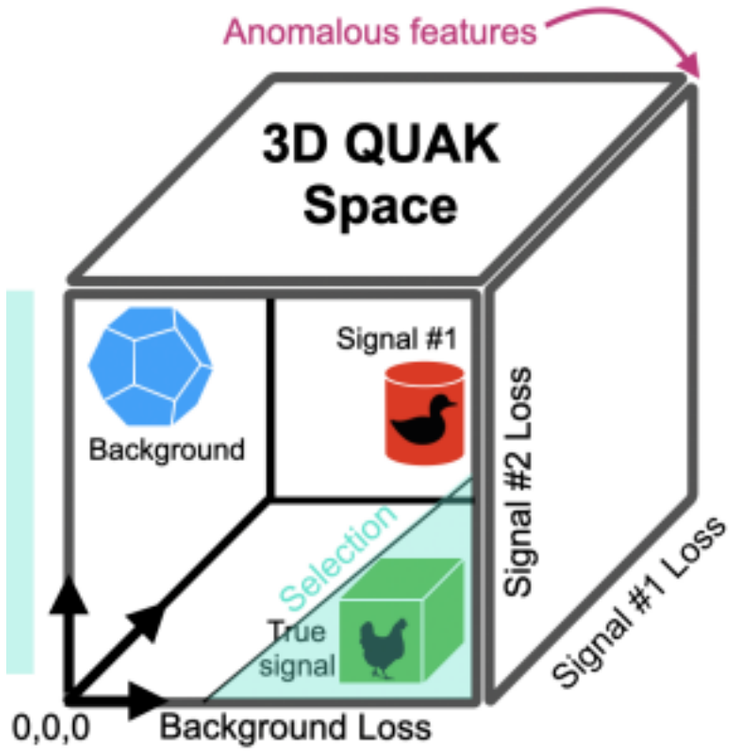
$$w(x|m) \equiv \frac{f(x)}{1-f(x)} = \frac{p(x|\text{data})}{p(x|\text{simulation})} \times \frac{p(\text{data})}{p(\text{simulation})}$$

Semi-supervised

Signal-classifier based training
Train a classifier on a potential signal (or cocktail of potential signals!) and use like in a fully supervised search.

QUAK (2011.03550):
Combine potential signals (supervised) and unlabelled data. Essentially use different signal priors to build a latent space in which to search for anomalies.



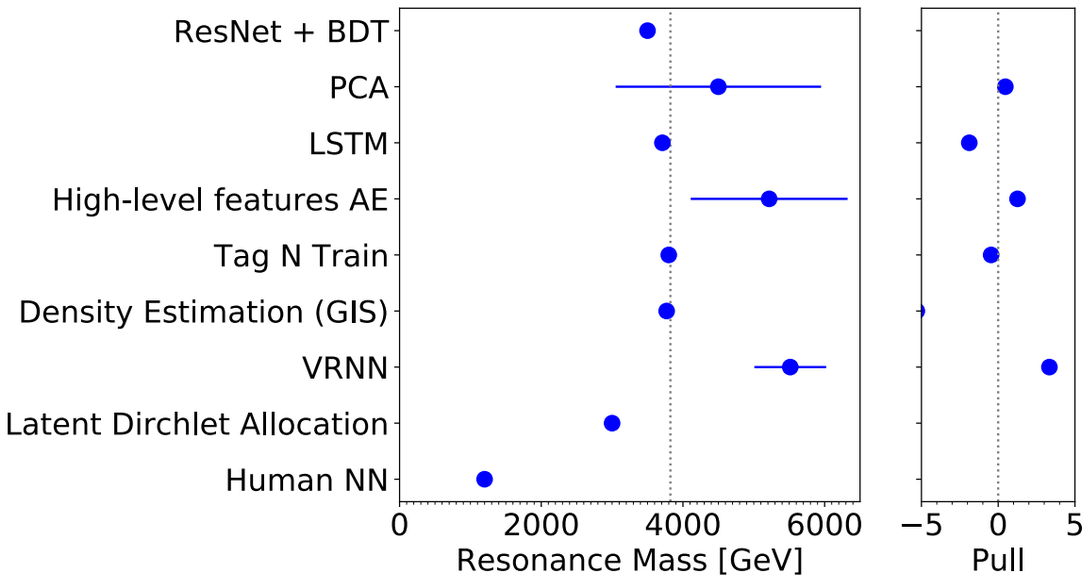


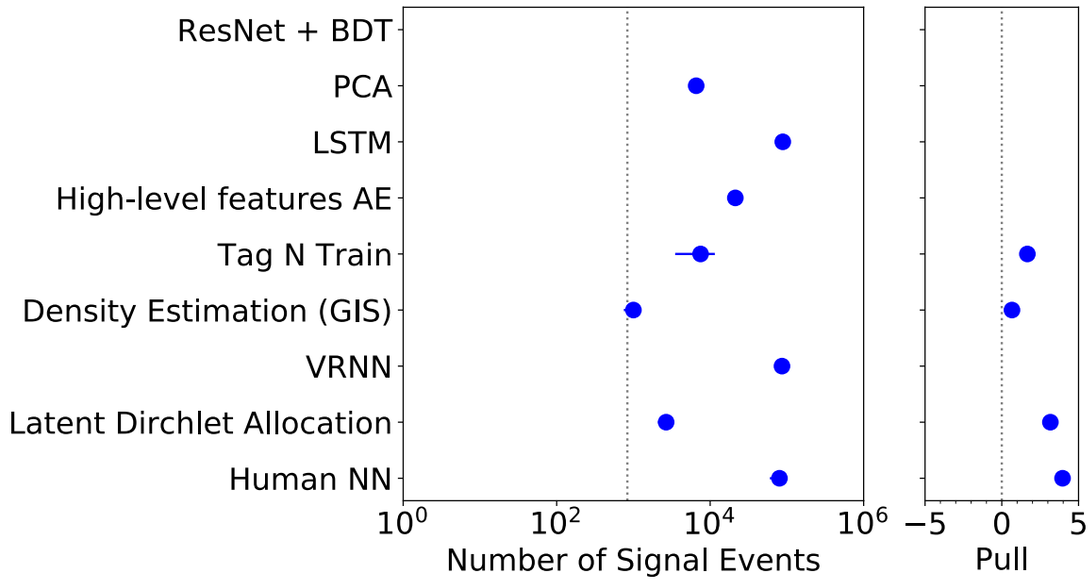
Reporting of results

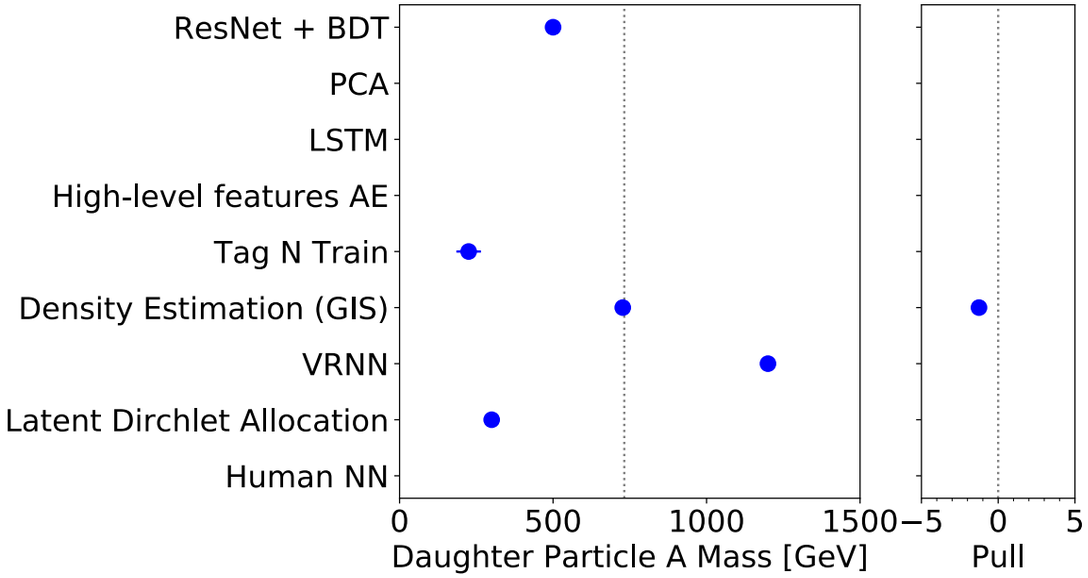
Section	Short Name	Method Type	Results Type
3.1	VRNN	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.2	ANODE	Unsupervised	(iii)
3.3	BuHuLaSpa	Unsupervised	(i) (BB2,3) and (ii) (BB1)
3.4	GAN-AE	Unsupervised	(i) (BB2-3) and (ii) (BB1)
3.5	GIS	Unsupervised	(i) (BB1)
3.6	LDA	Unsupervised	(i) (BB1-3)
3.7	PGA	Unsupervised	(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)

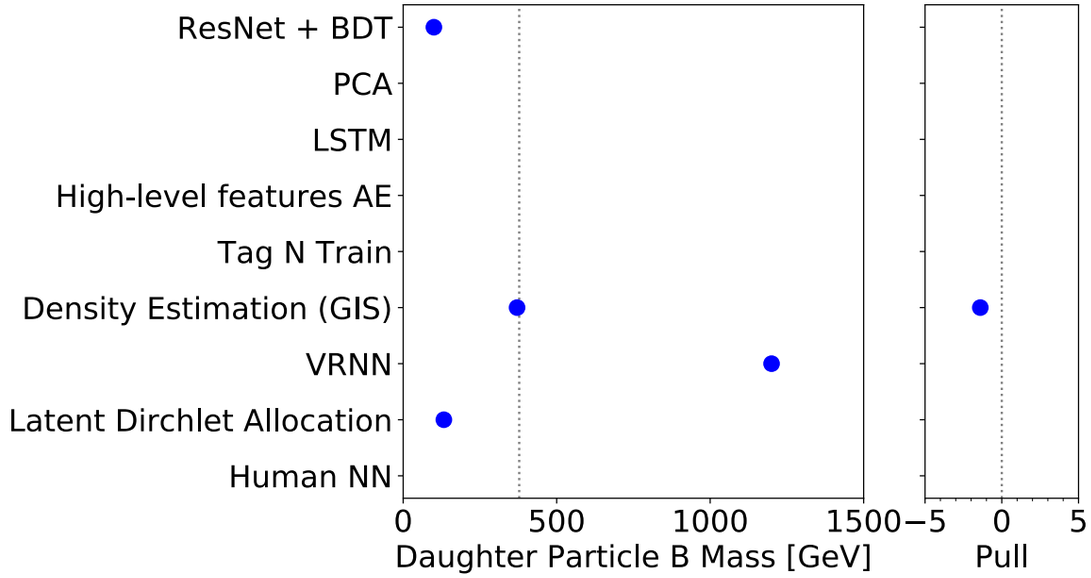
i) during challenge phase
ii) after challenge phase
iii) R&D dataset used

Results - BB1









(Shown are results during challenge)

Several approaches identified resonance; density estimation also found correct properties

Results - BB2

Reminder: no signal injected

Some methods reported resonances in the tail of the mass distribution (around 4.5 TeV)

Difficult to predict for edges of phase space

Latent Dirichlet Allocation: Our method extracts signal descriptions which appear convincing, however the classifier does not identify a bump in the invariant mass spectra. Without this we were unable to determine that a signal was present. The di-jet description extracted consisted of one jet of mass 350-400 GeV and another of mass 150-200 GeV. If the production of these states was non-resonant, we would be unable to find the signal with our method. Or if more than just di-jets were relevant to reconstruct the invariant mass, we would also not be able to find it. **Otherwise, we determine that no signal was present in the data.**

Results - BB3

Reminder: di-jet and tri-jet topologies

Different observations claimed, none identified the correct signal.

Lessons learned

- Anomaly detection is difficult
  - Even for “anomalies” close to already considered signals
  - Even more so for “exotic” signals
  - Value in blind studies

- Robust uncertainty quantification needed - especially for tails of distributions

- Many methods used “sidebanding” in invariant mass + learning some anomaly detector.
  - Less reliance on peaks annd
  - less reliance on one ‘lucky’ (physically inspired) variable desirable

- Did not discuss data representation: image vs point cloud vs … vs high-level features
  - Will bias anomaly detection performance. Need to understand better.

Other open issues

- Strategies to assess the quality of anomaly detection techniques without (or at least with less) dependence on specific signal models? Right now, the strategy seems to be to compare the ability of ADs to find some benchmark signals.

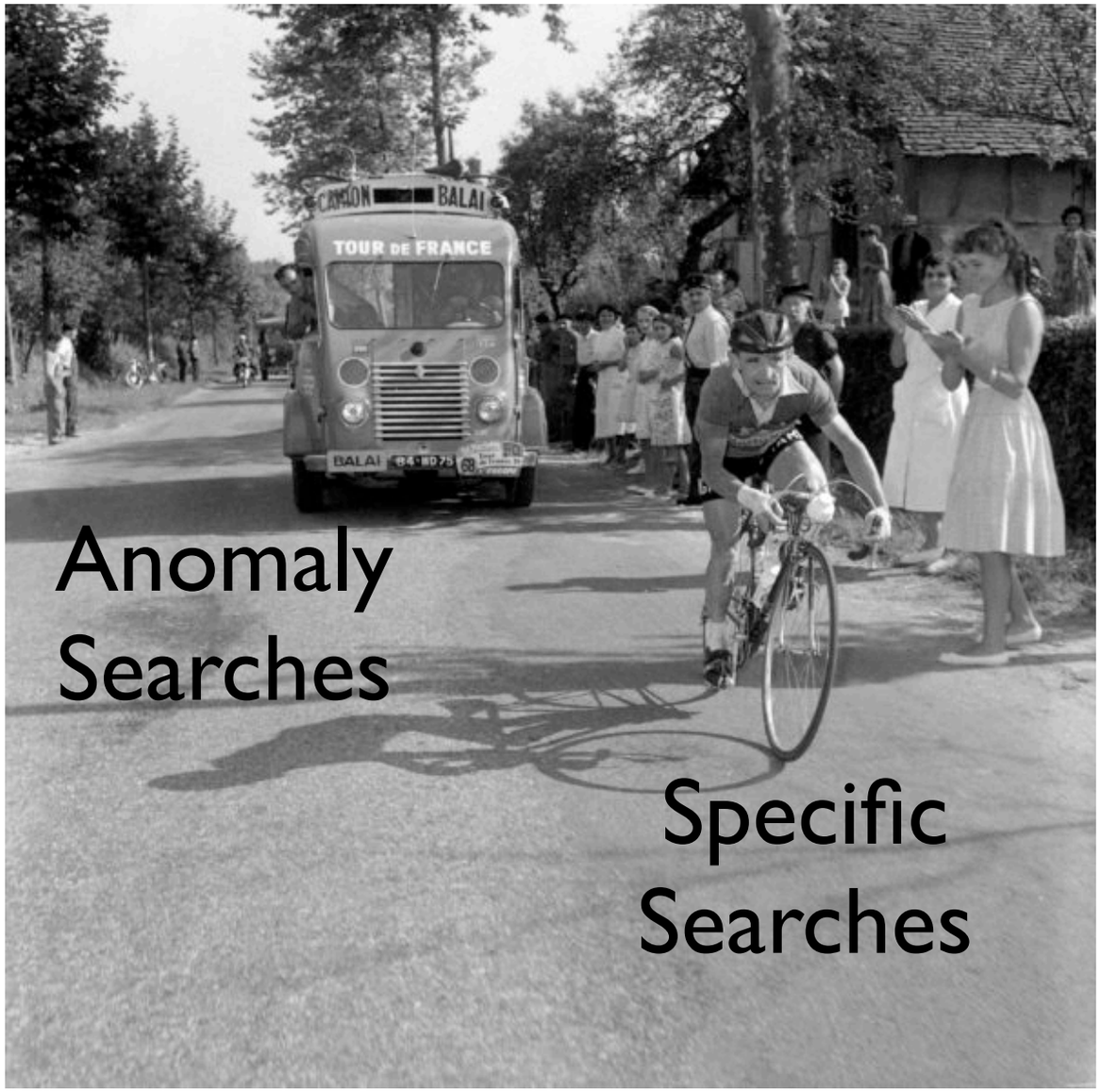
- Can there be robust methods to set exclusion limits with ‘data only’ anomaly dectors (i.e. methods where all final trainings are carried out on data - as opposed to training on simulation) without injecting signal events into the data.

- How to to publish the on-data-trained anomaly detectors in such a way that allows ex-post analysis by people outside the experiment whether the training result is compatible / rules out a given new physics signals.

- Methods to go from an observed anomaly (ie. a signal like excess in some region of data) to an interepretation in terms of physics models can still be improved as well.

Conclusions

- Exciting space of anomaly detection in LHC physics
- First successes and breadth of ideas but better understanding and more applications needed
- Potential to search for more signatures with less people, even while some conceptual issues are being resolved
- Also applies to other areas (e.g. data quality / detector operations)



Thank you!