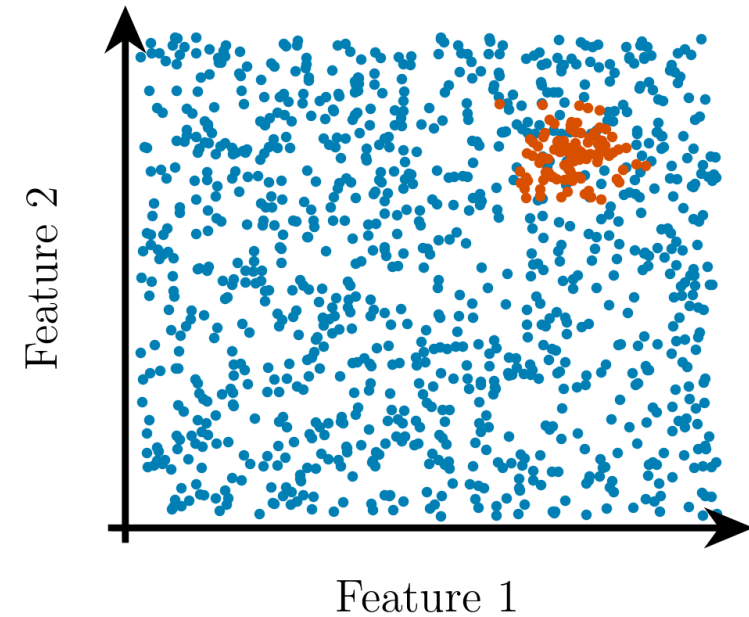
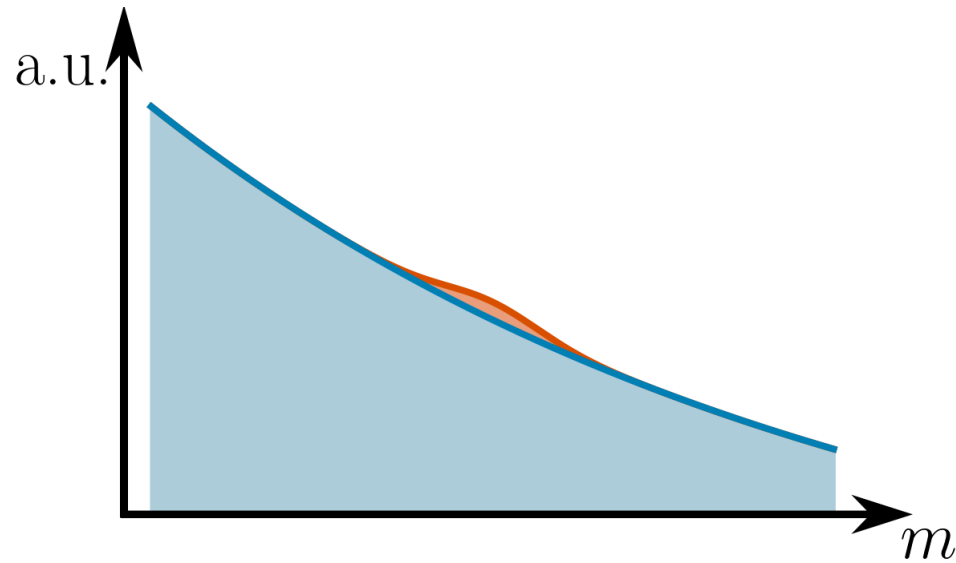
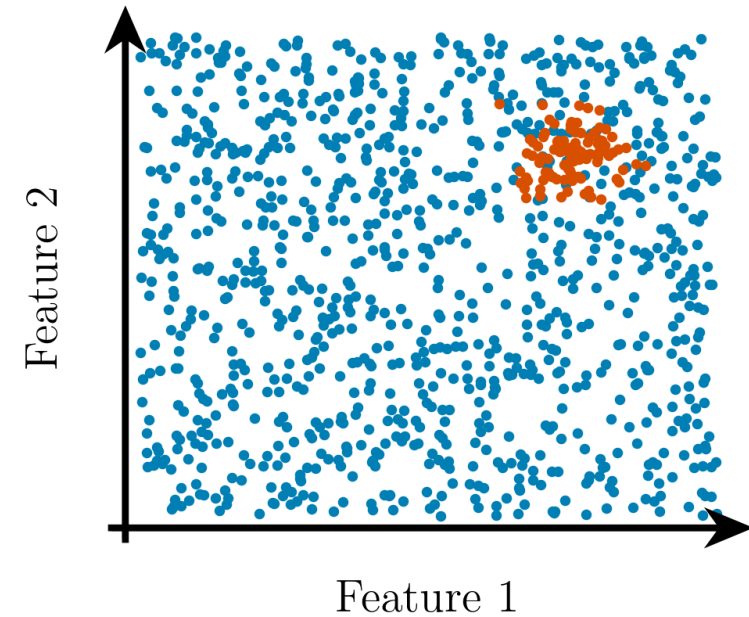
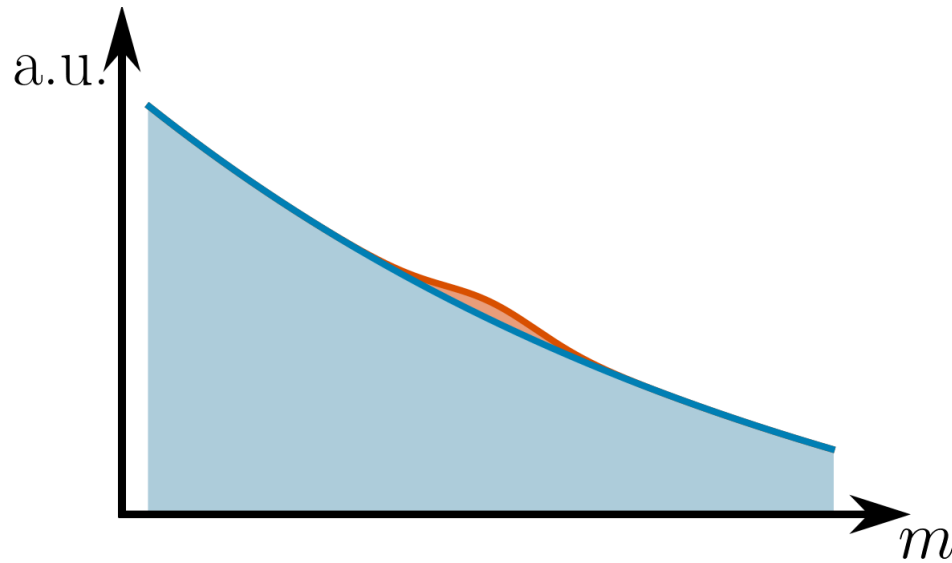


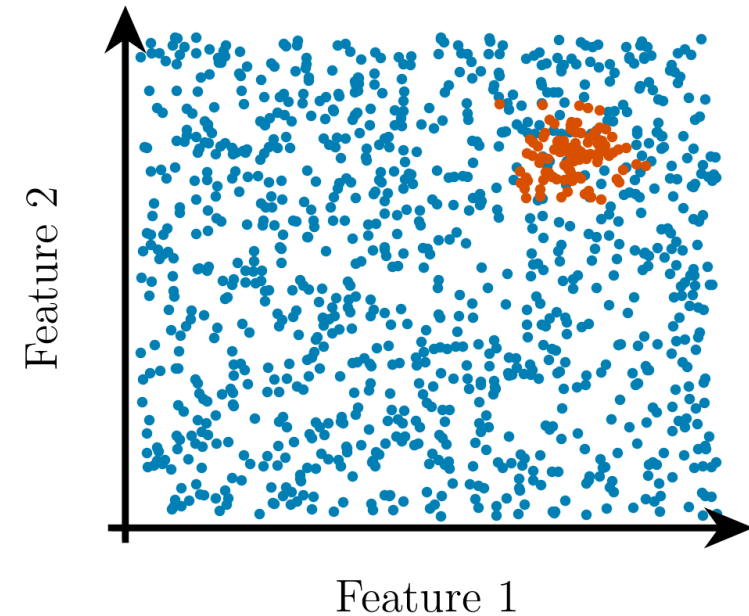
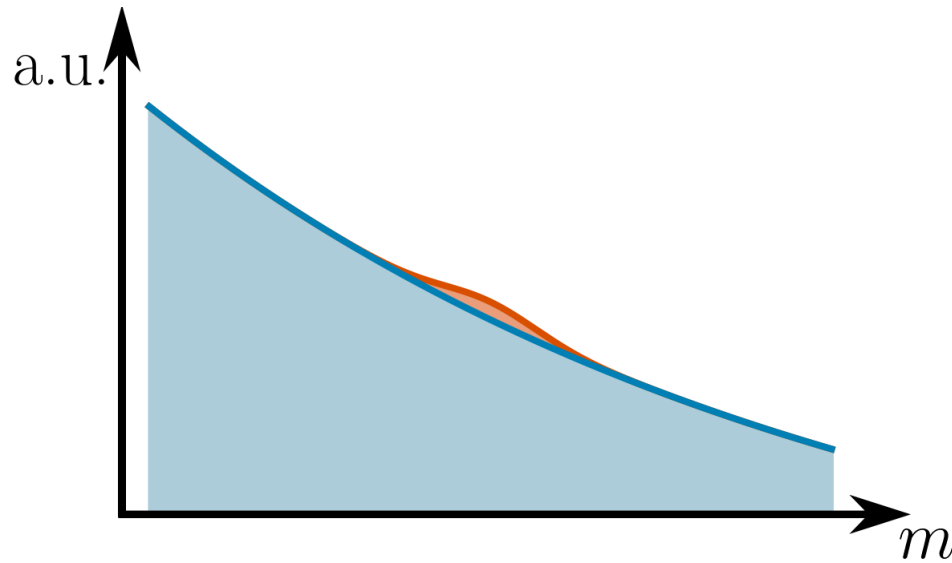
Anomaly Detection: From Methods to Applications

Marie Hein

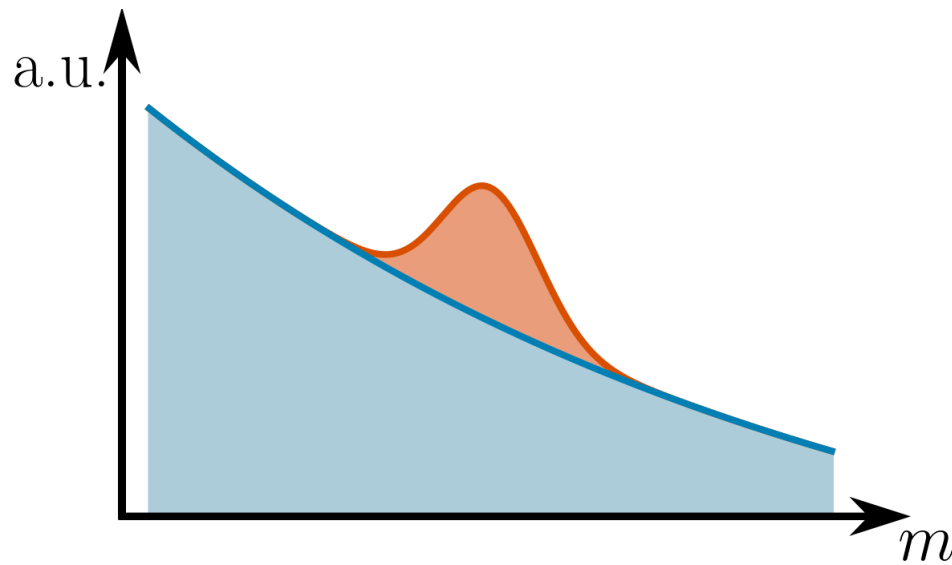
ML4Jets 2026, Vienna



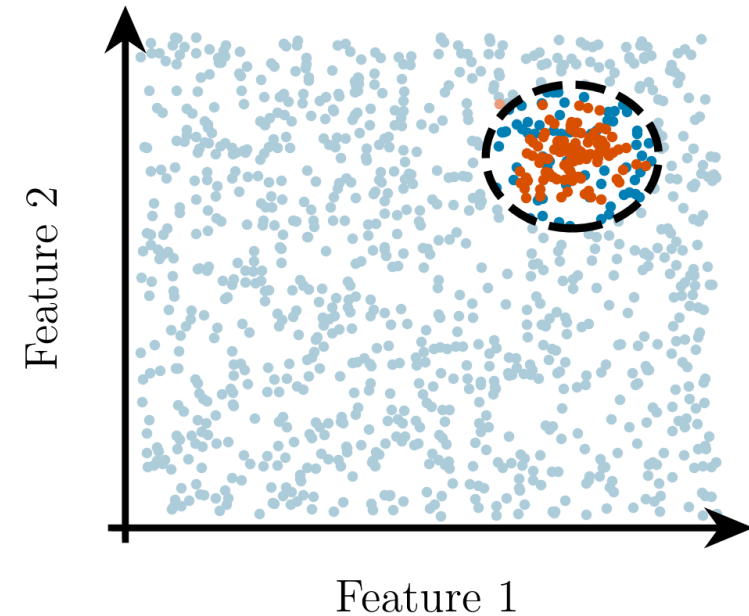




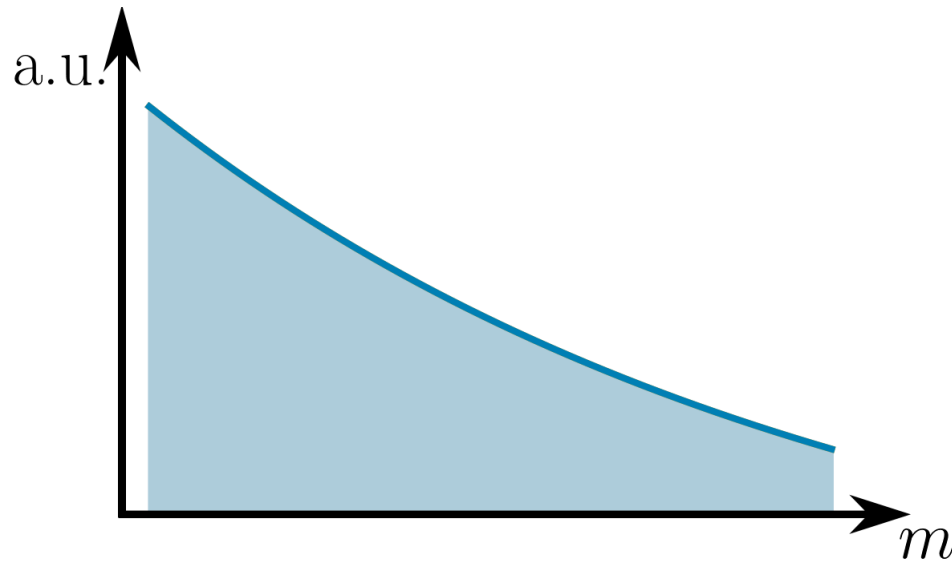
Inclusive



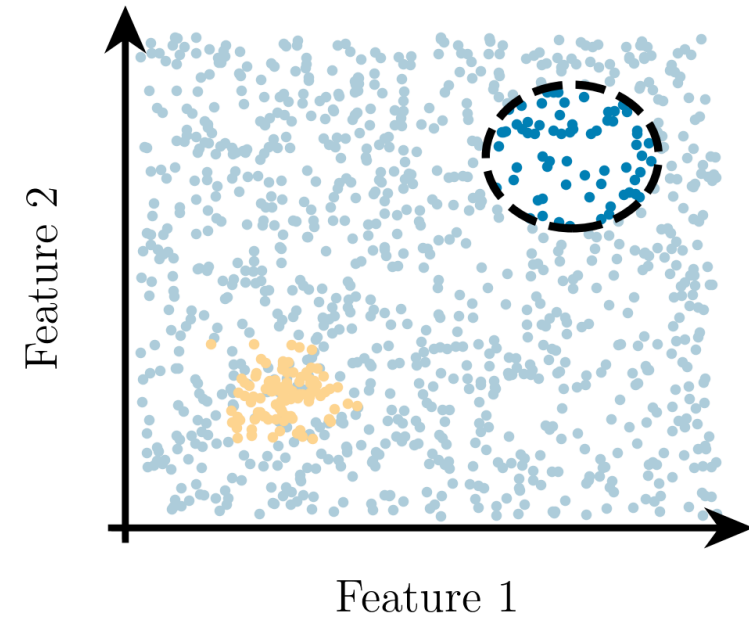
Inclusive



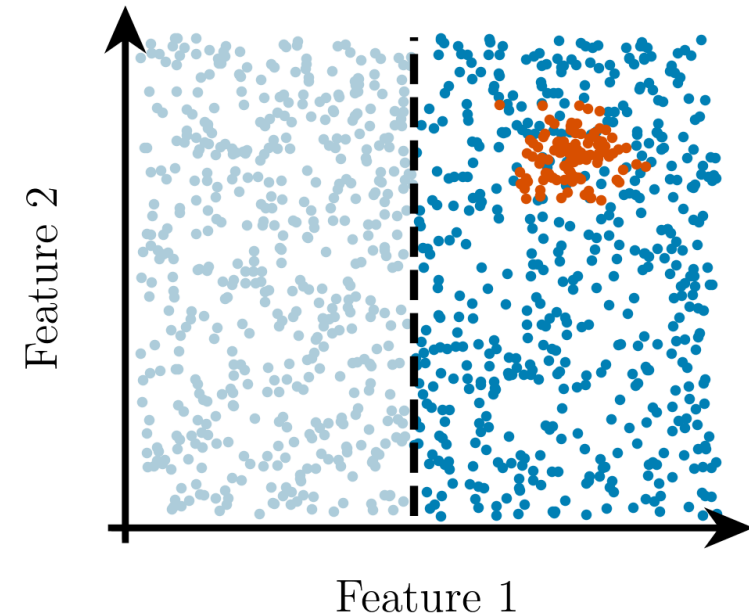
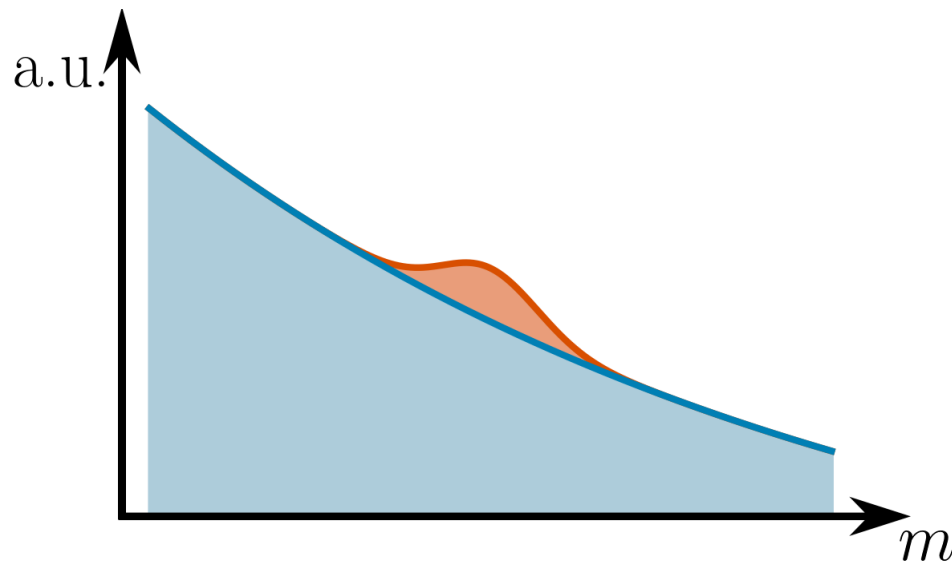
Model-specific



Inclusive



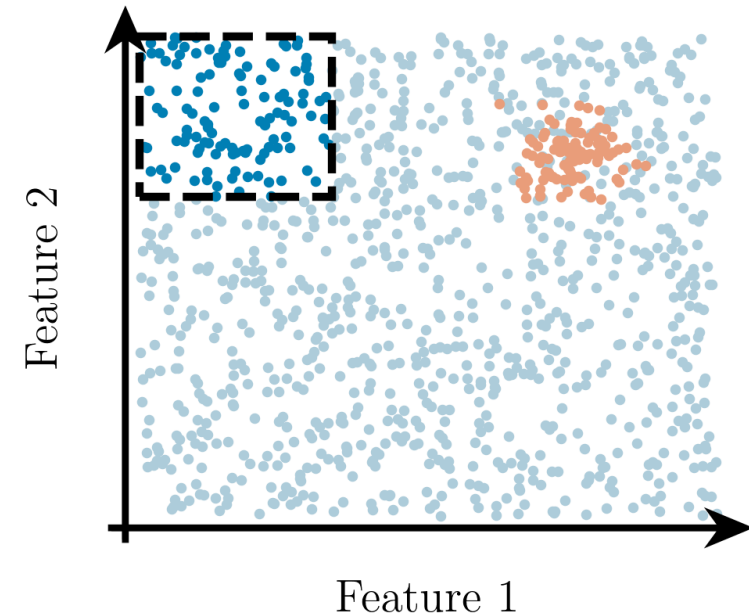
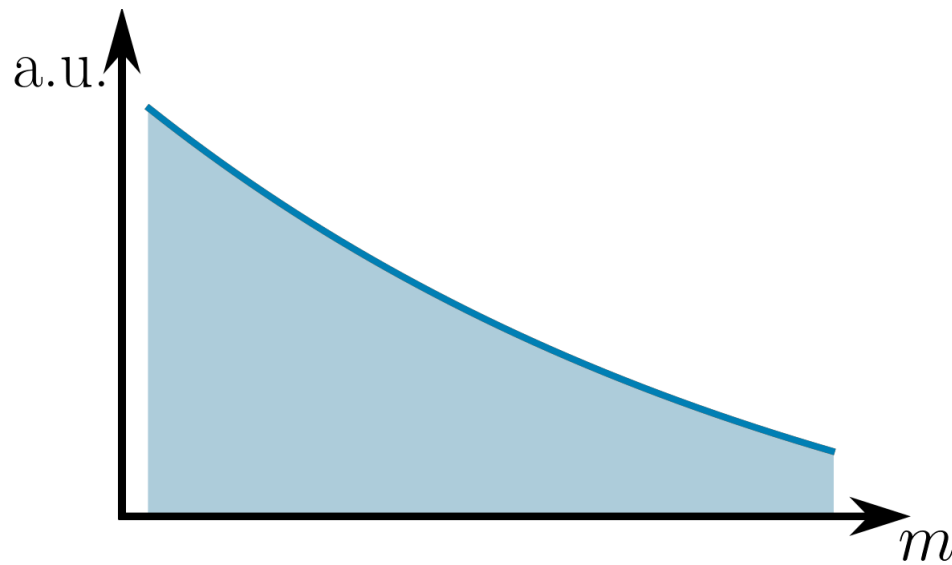
Model-specific



Inclusive

Single cut

Model-specific

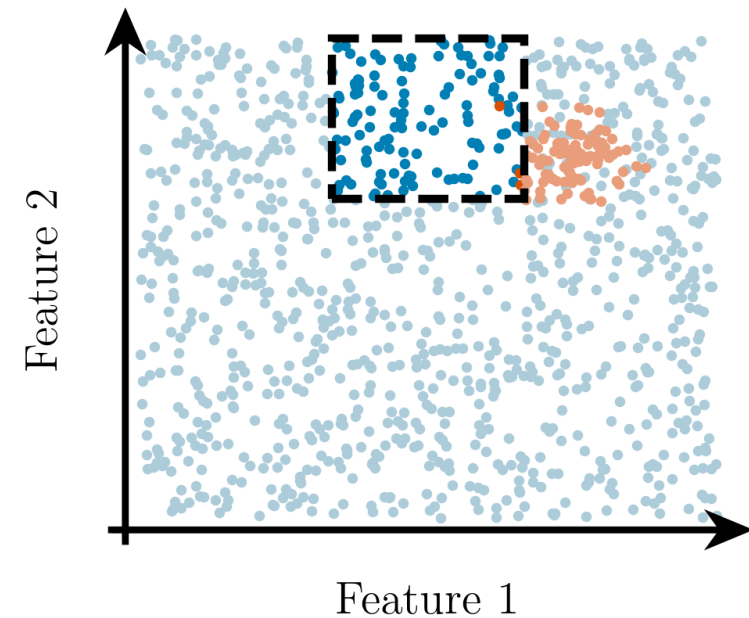
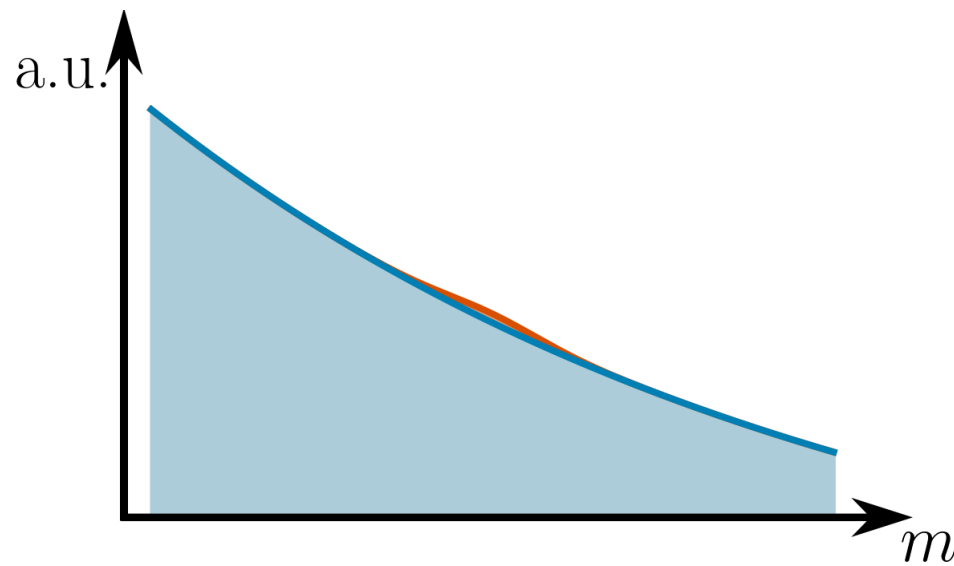


Inclusive

Single cut

Model-specific

Anomaly detection

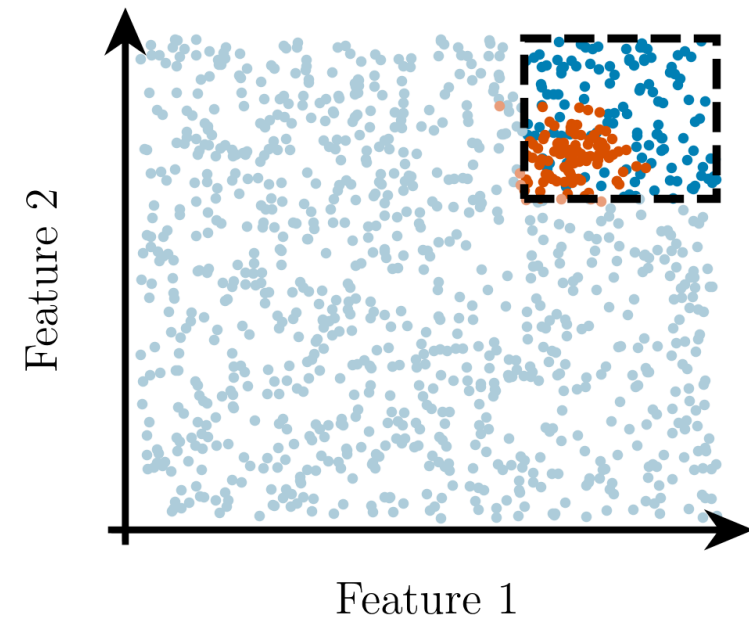
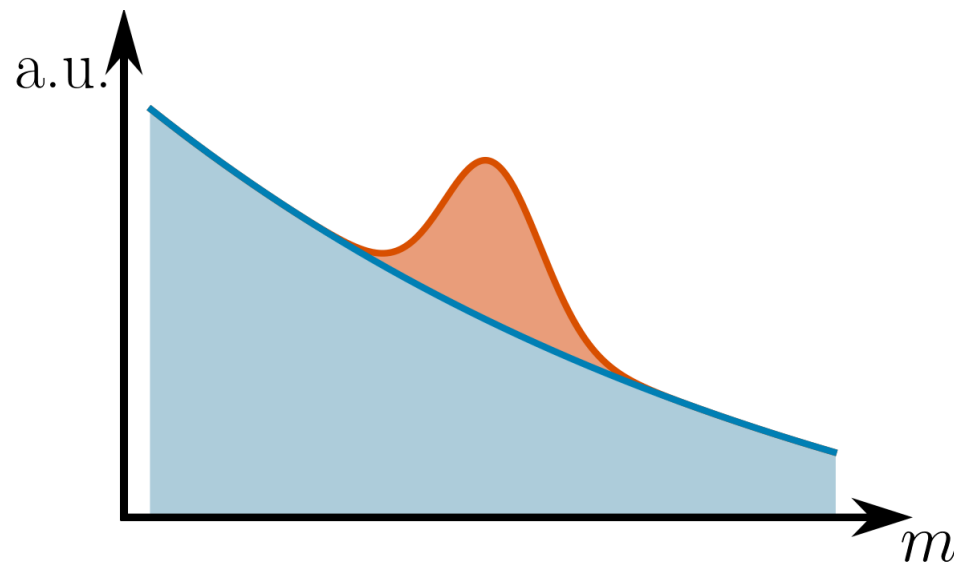


Inclusive

Single cut

Model-specific

Anomaly detection



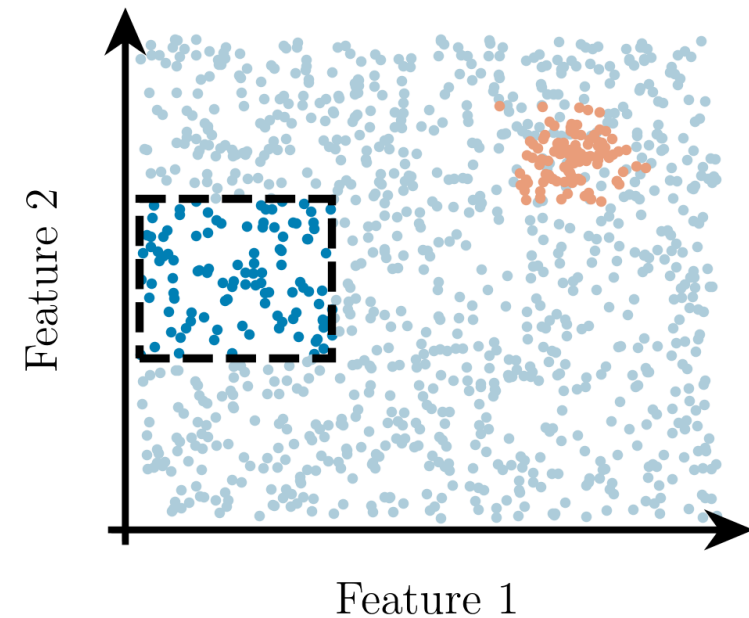
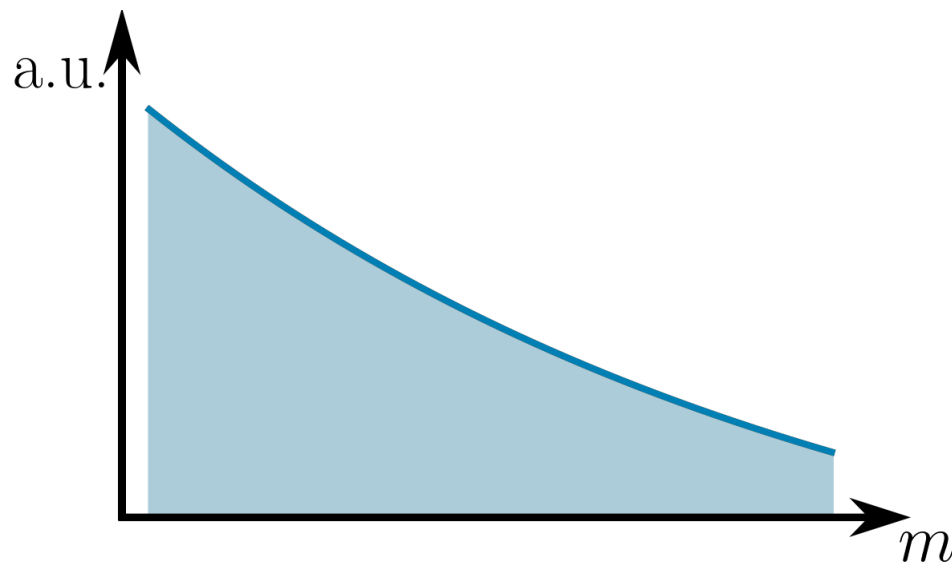
Inclusive

Single cut

Model-specific

Anomaly detection

Anomaly Detection

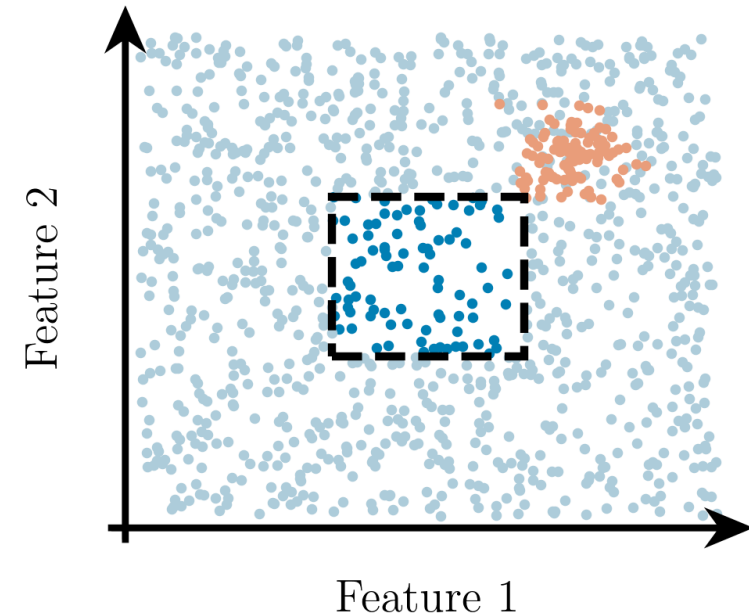
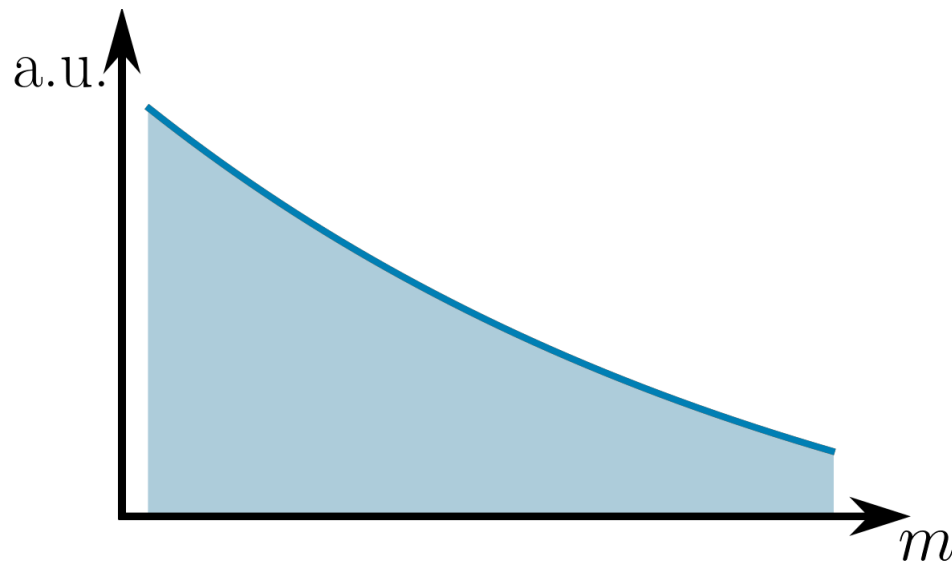


Inclusive

Single cut

Model-specific

Anomaly detection

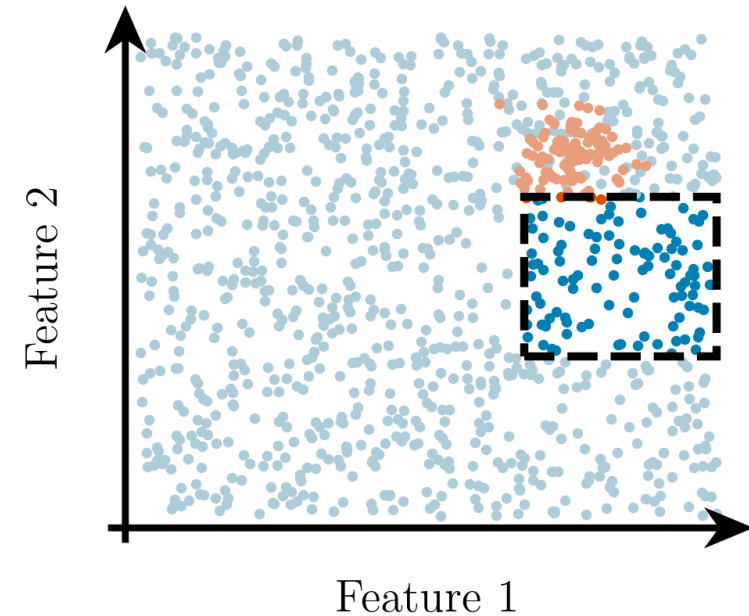
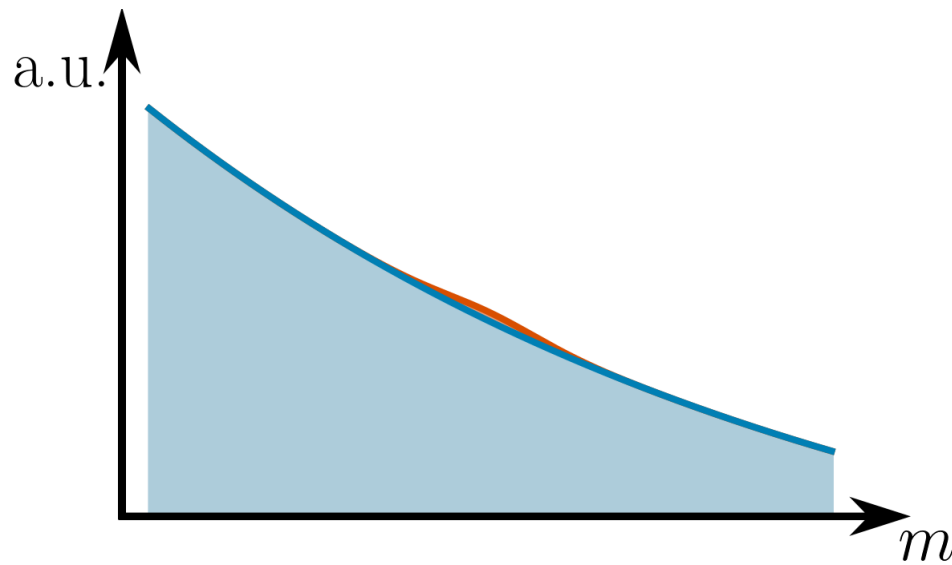


Inclusive

Single cut

Model-specific

Anomaly detection



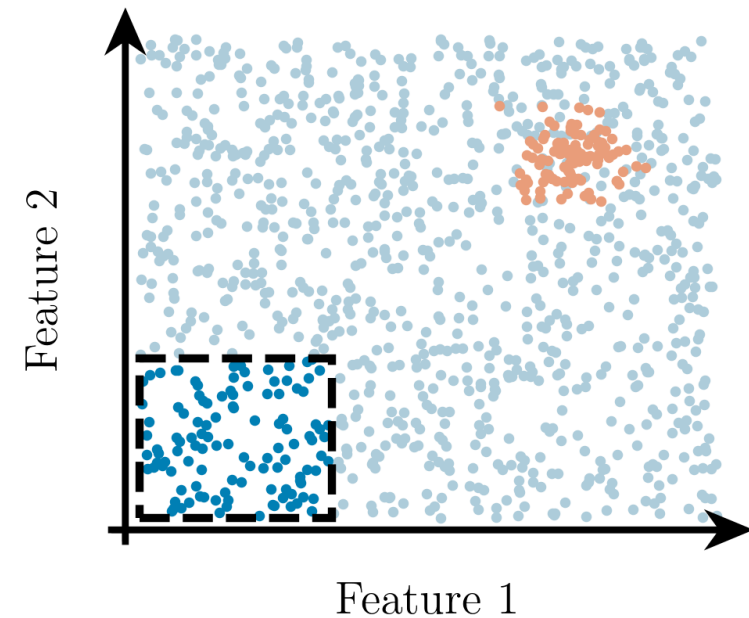
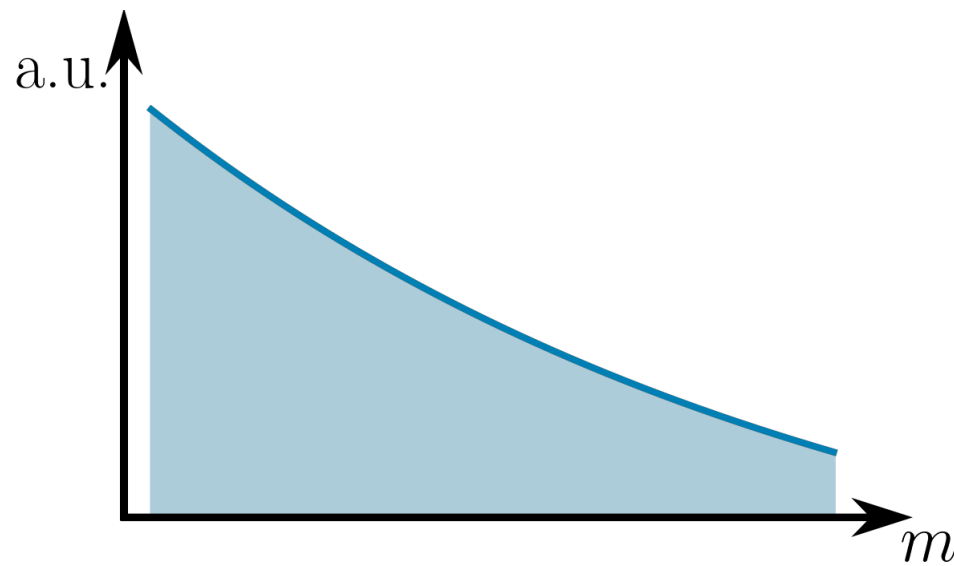
Inclusive

Single cut

Model-specific

Anomaly detection

Anomaly Detection

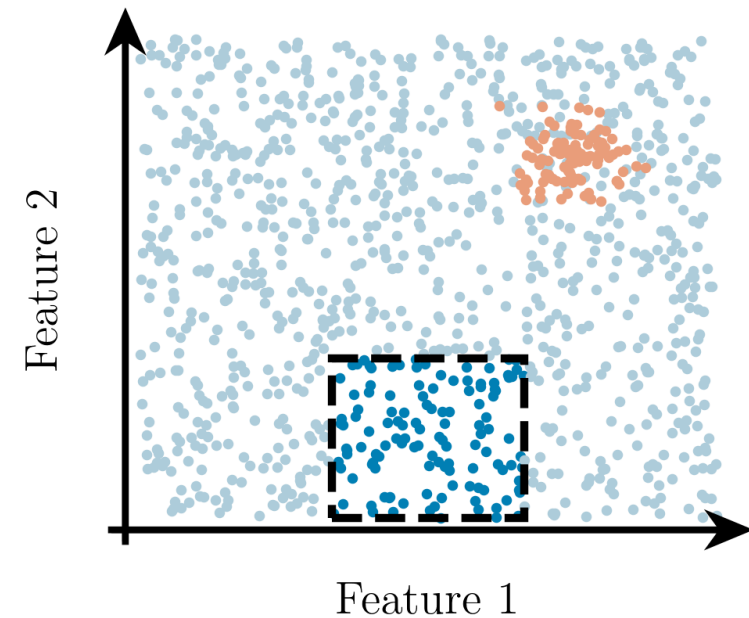
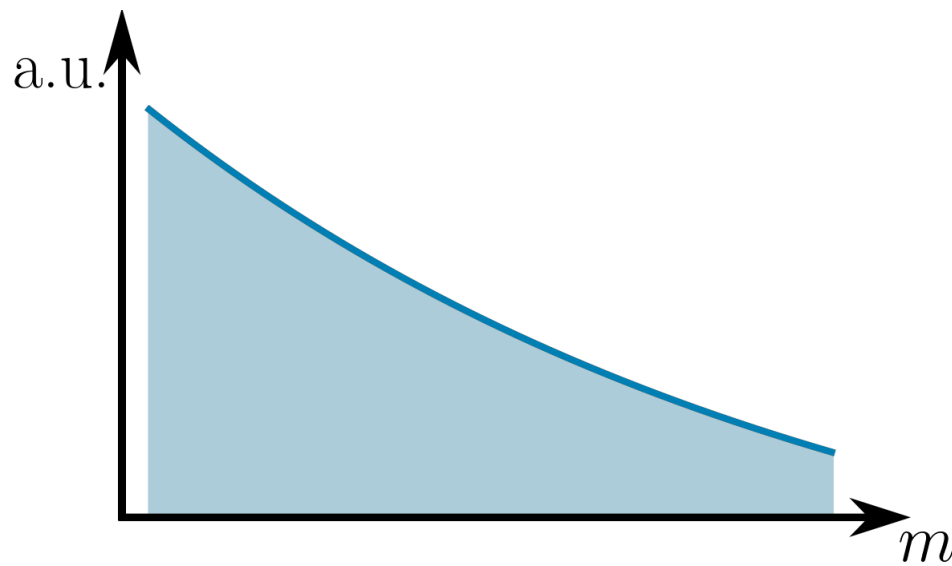


Inclusive

Single cut

Model-specific

Anomaly detection



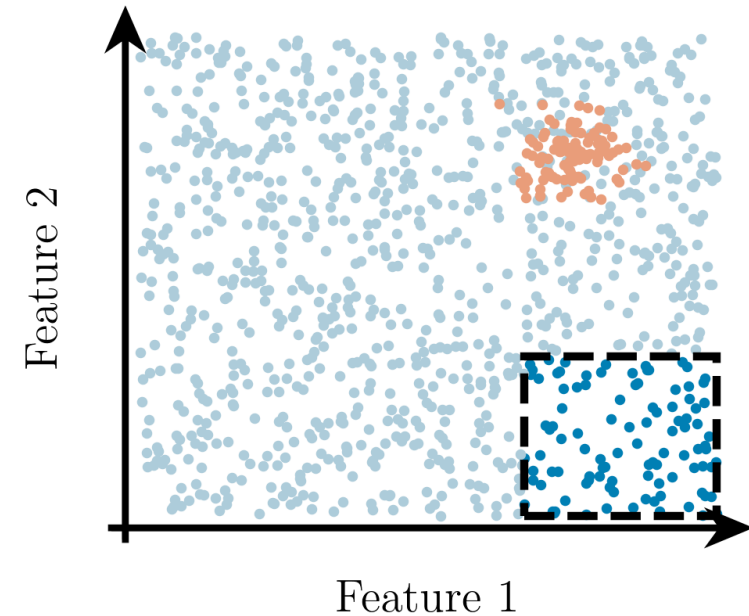
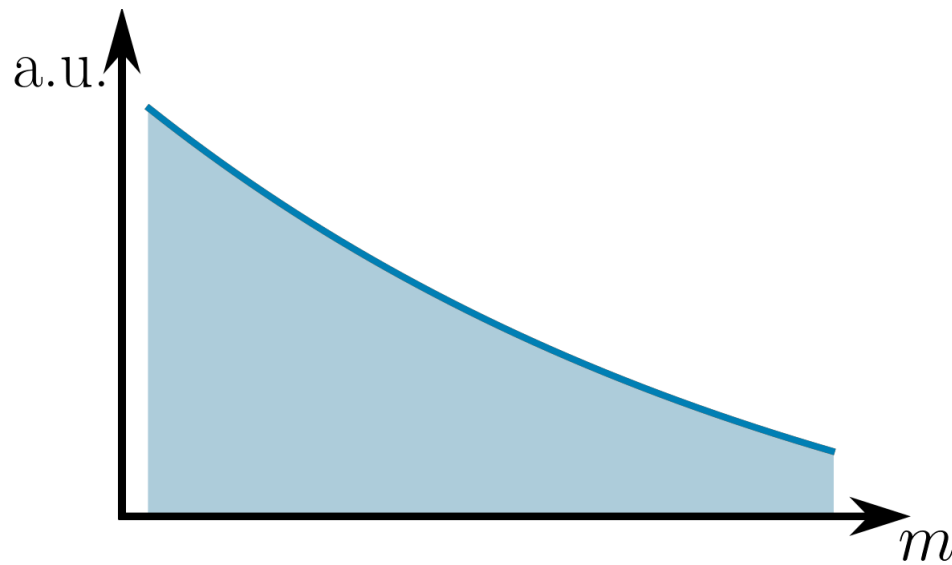
Inclusive

Single cut

Model-specific

Anomaly detection

Anomaly Detection



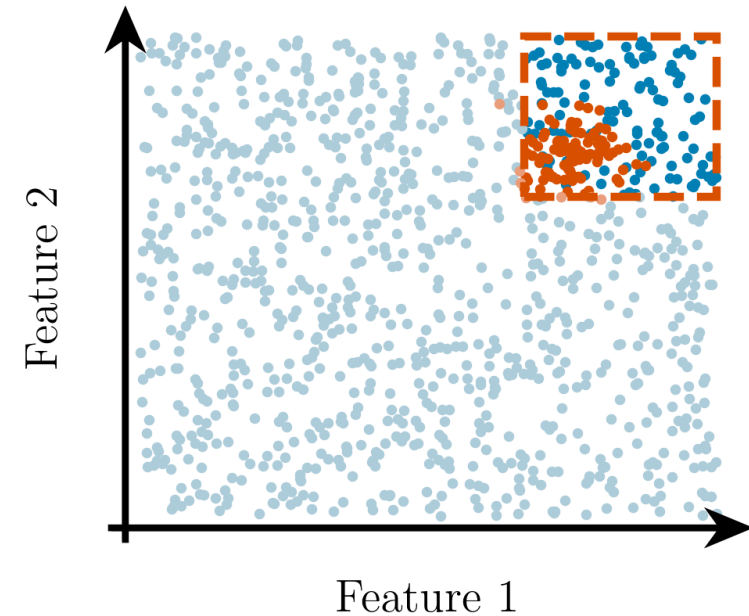
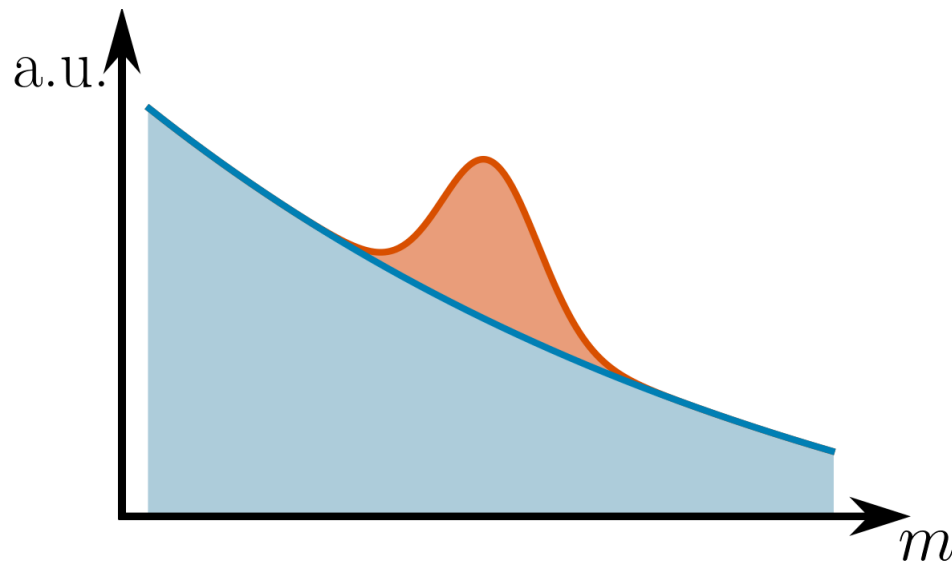
Inclusive

Single cut

Model-specific

Anomaly detection

Anomaly Detection



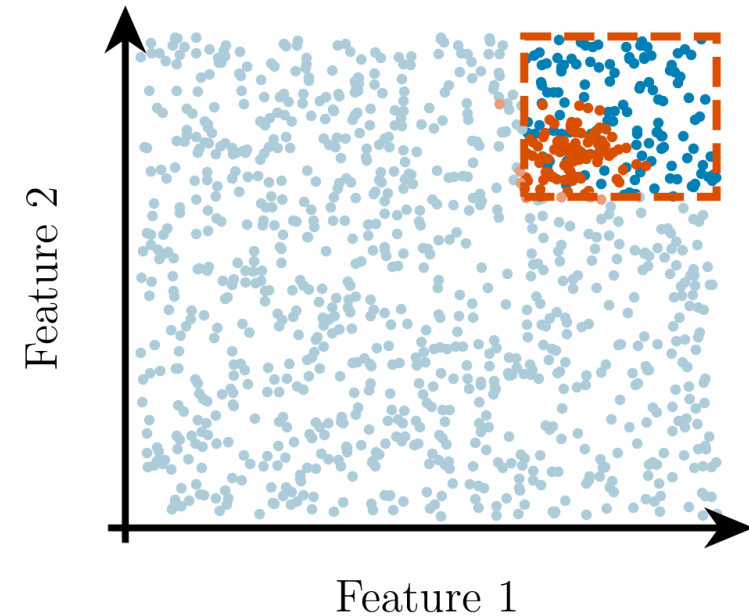
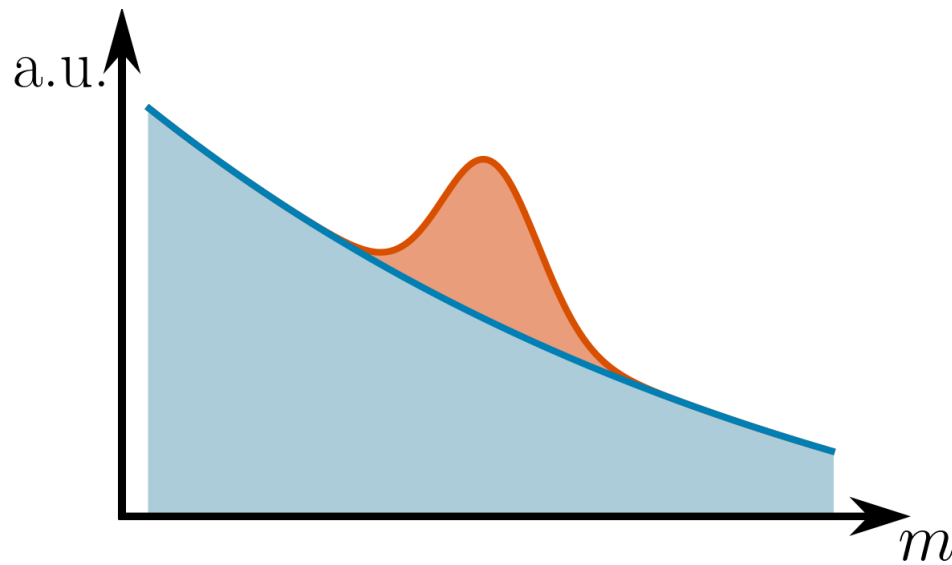
Inclusive

Single cut

Model-specific

Anomaly detection

Anomaly Detection



Inclusive

Single cut

Model-specific

Anomaly detection

The Spectrum of Anomaly Detection

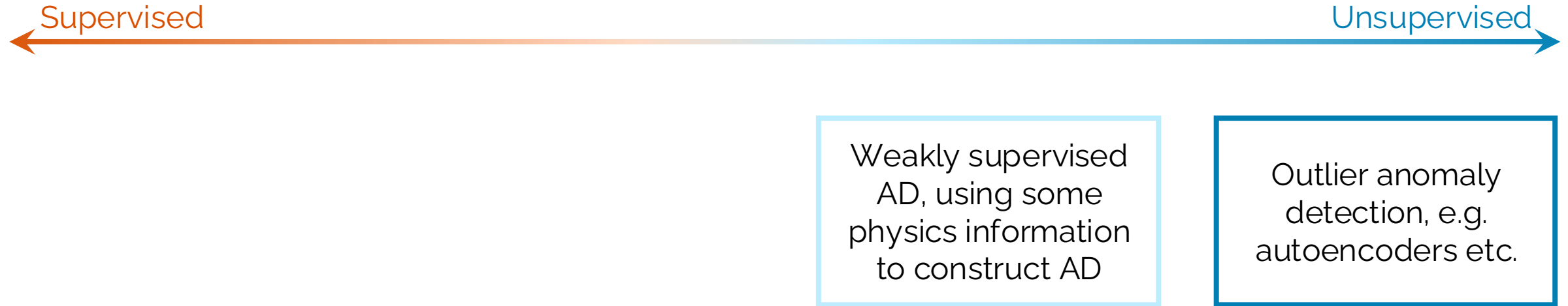


The Spectrum of Anomaly Detection

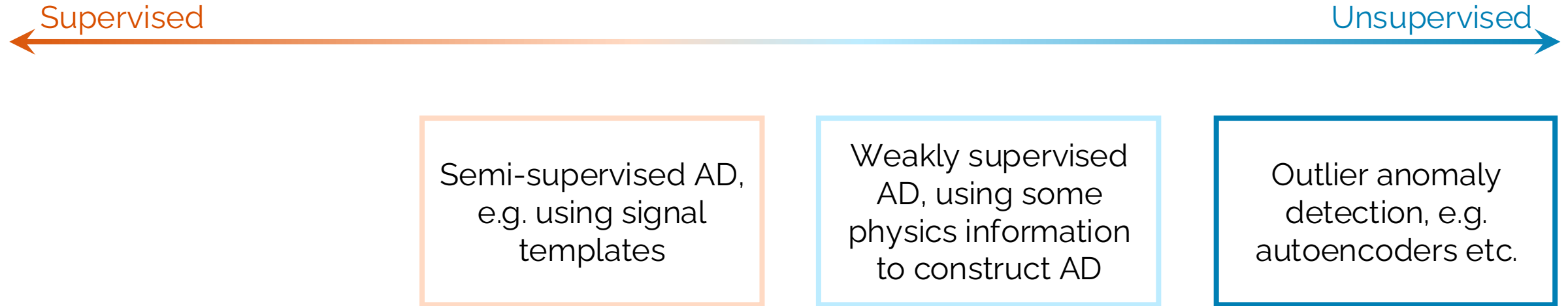


Outlier anomaly
detection, e.g.
autoencoders etc.

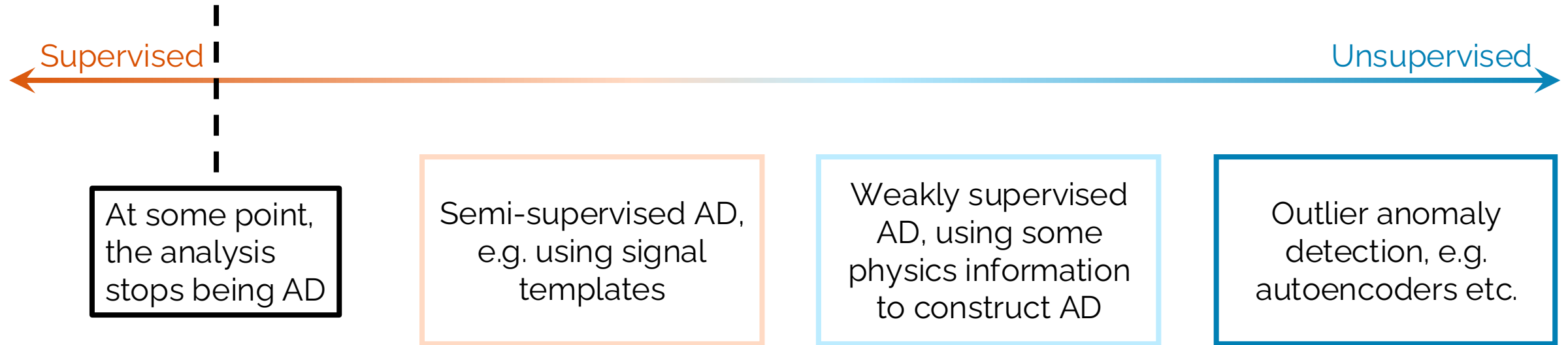
The Spectrum of Anomaly Detection



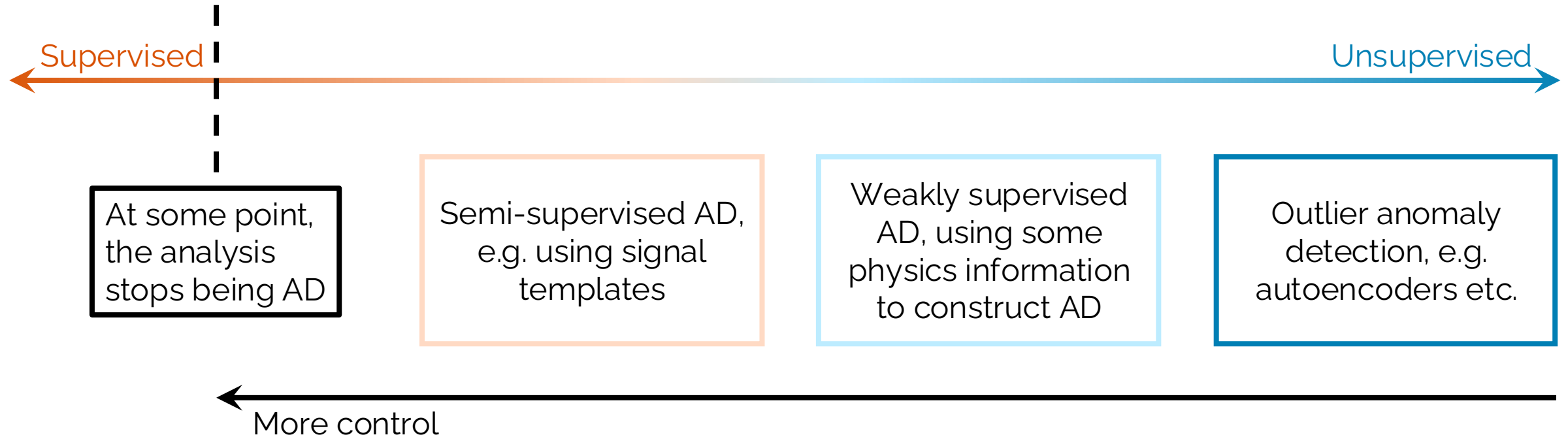
The Spectrum of Anomaly Detection



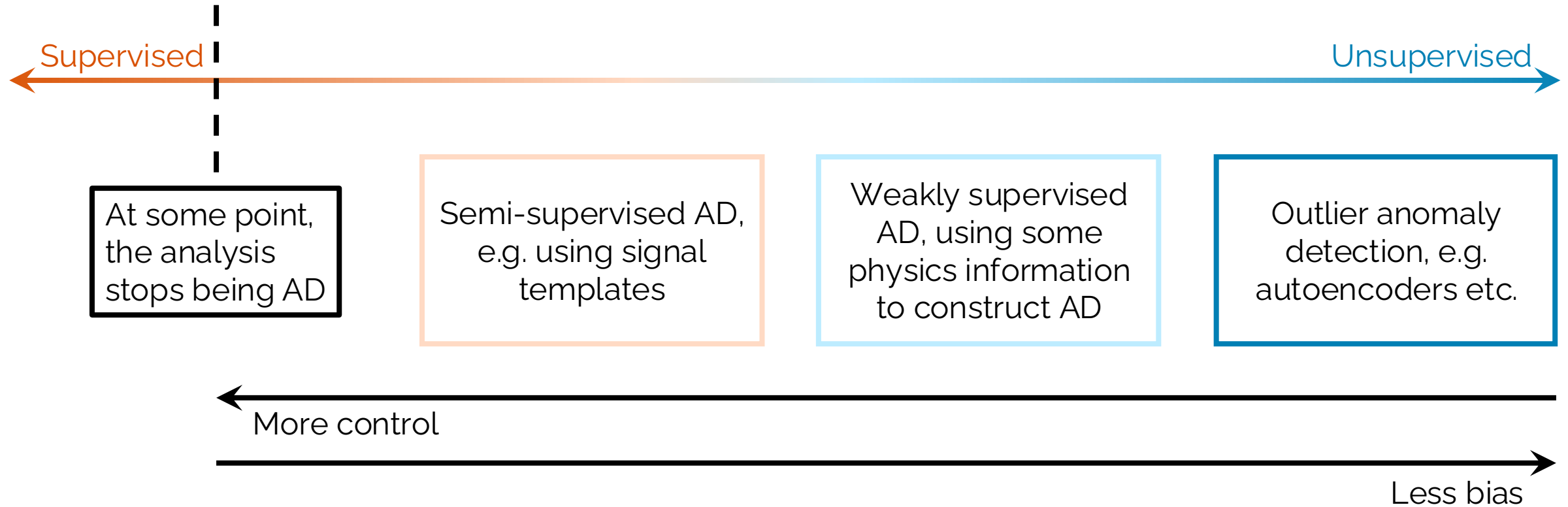
The Spectrum of Anomaly Detection



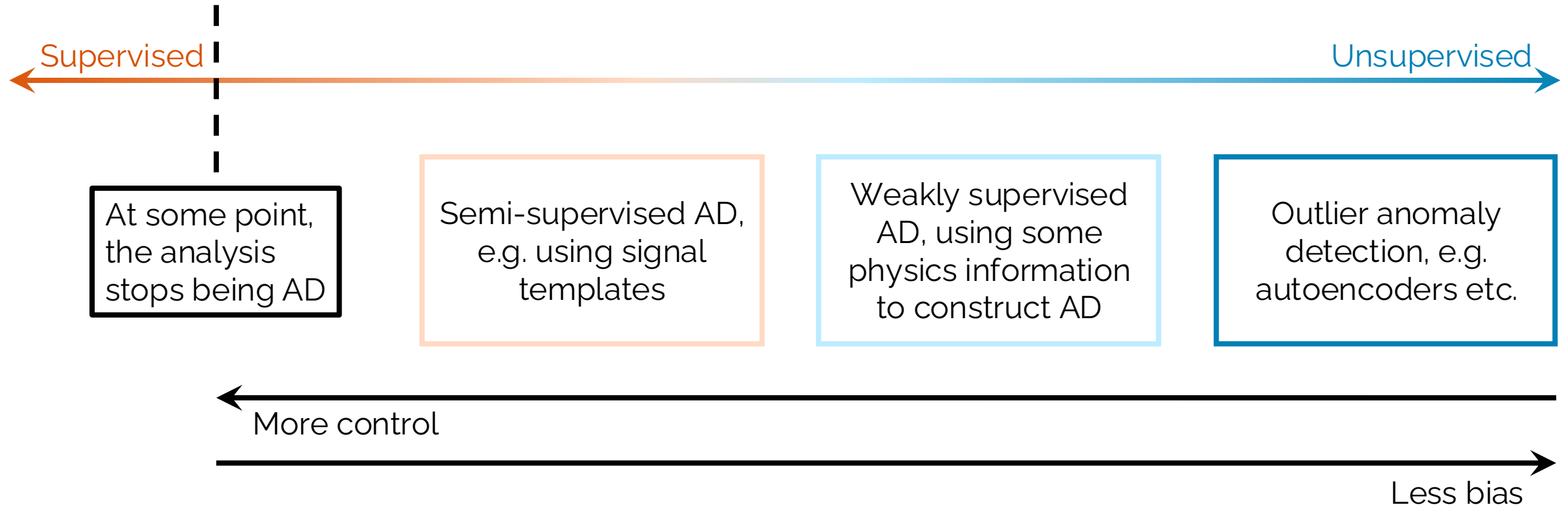
The Spectrum of Anomaly Detection



The Spectrum of Anomaly Detection

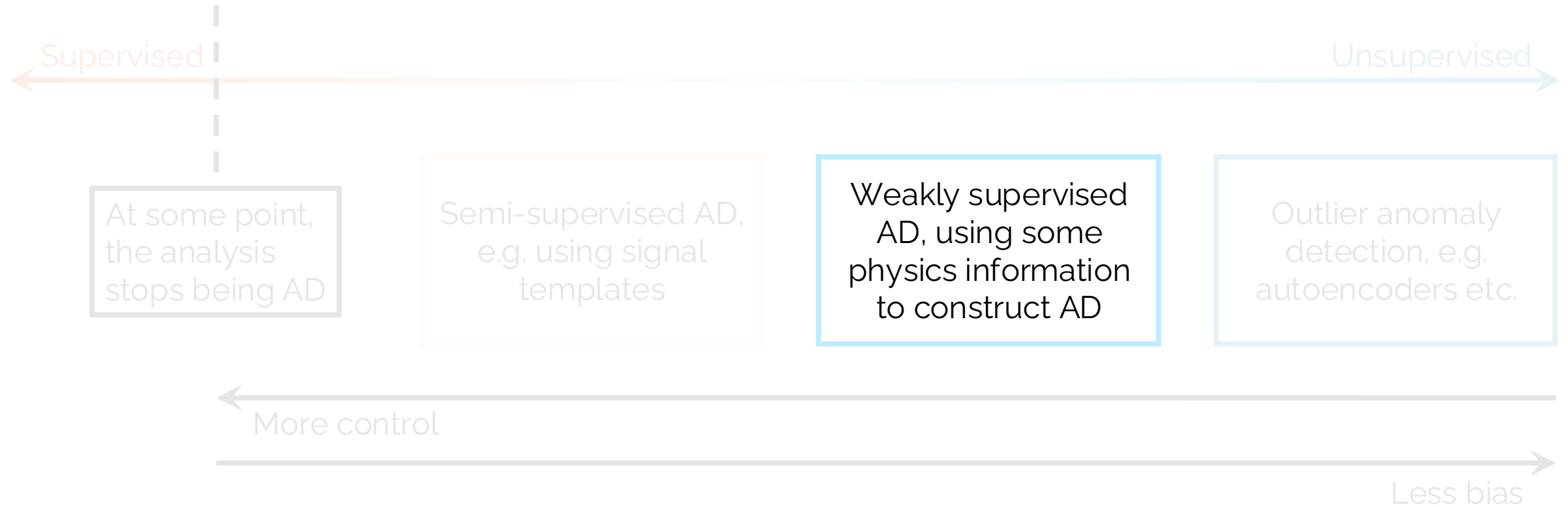


The Spectrum of Anomaly Detection



Many different AD methods have been constructed in the last decade
→ Current development moving away from method development to application and detailed studies

The Spectrum of Anomaly Detection



Many different AD methods have been constructed in the last decade
→ Current development moving away from method development to application and detailed studies

“Classification without labels: Learning from mixed samples in high energy physics” [\[1709.02949\]](#), E. Metodiev, B. Nachman, J. Thaler

- Optimal classifier

$$R_{\text{optimal}}(x) = \frac{p_S(x)}{p_B(x)}$$

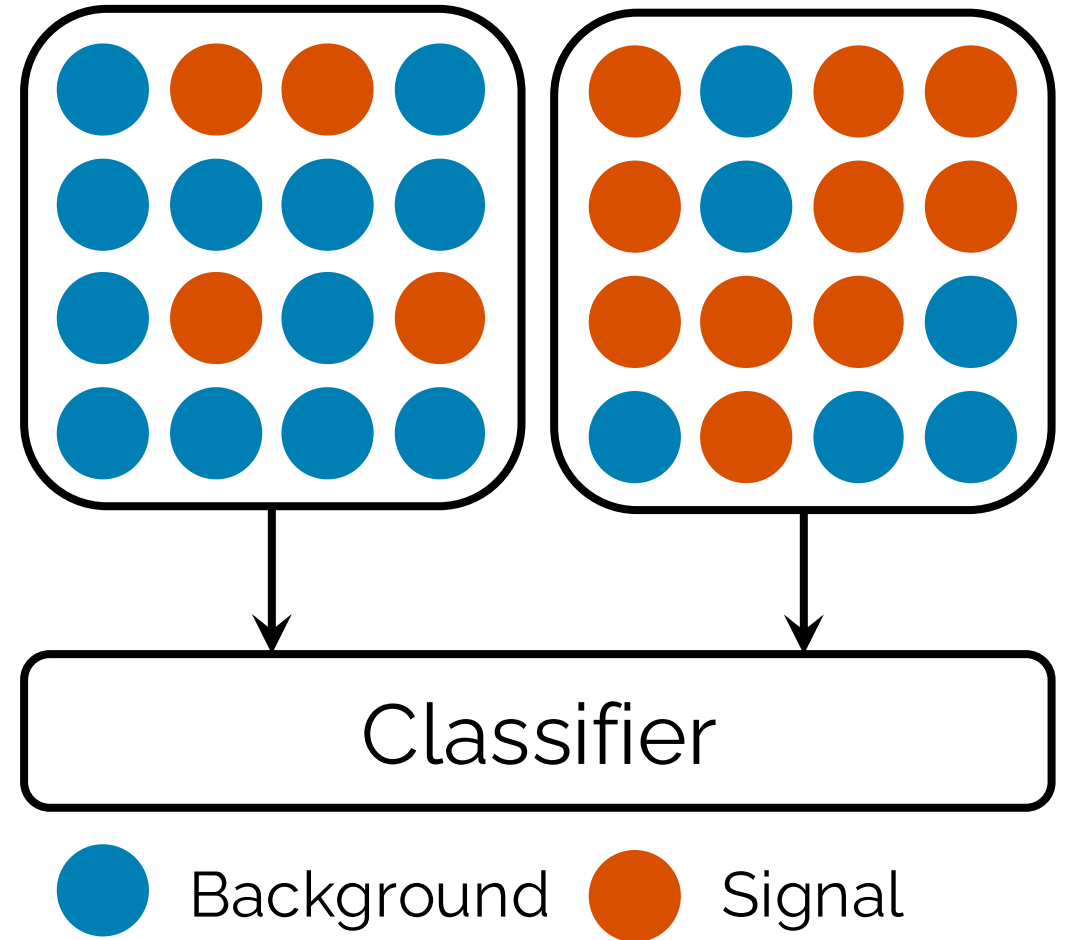
- For mixed datasets with signal fractions f_i

$$R_{\text{mixed}}(x) = \frac{f_1 R_{\text{optimal}}(x) + (1 - f_1)}{f_2 R_{\text{optimal}}(x) + (1 - f_2)}$$

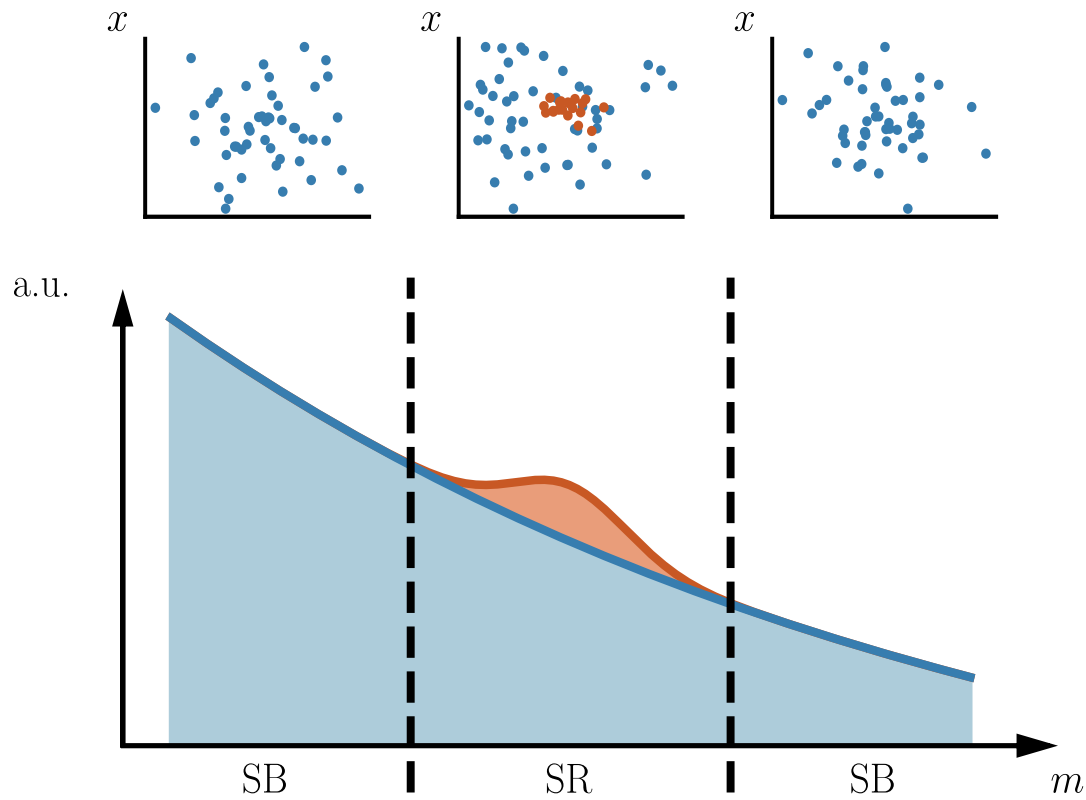
→ Monotonically increasing function of

$R_{\text{optimal}}(x)$ as long as $f_1 > f_2$

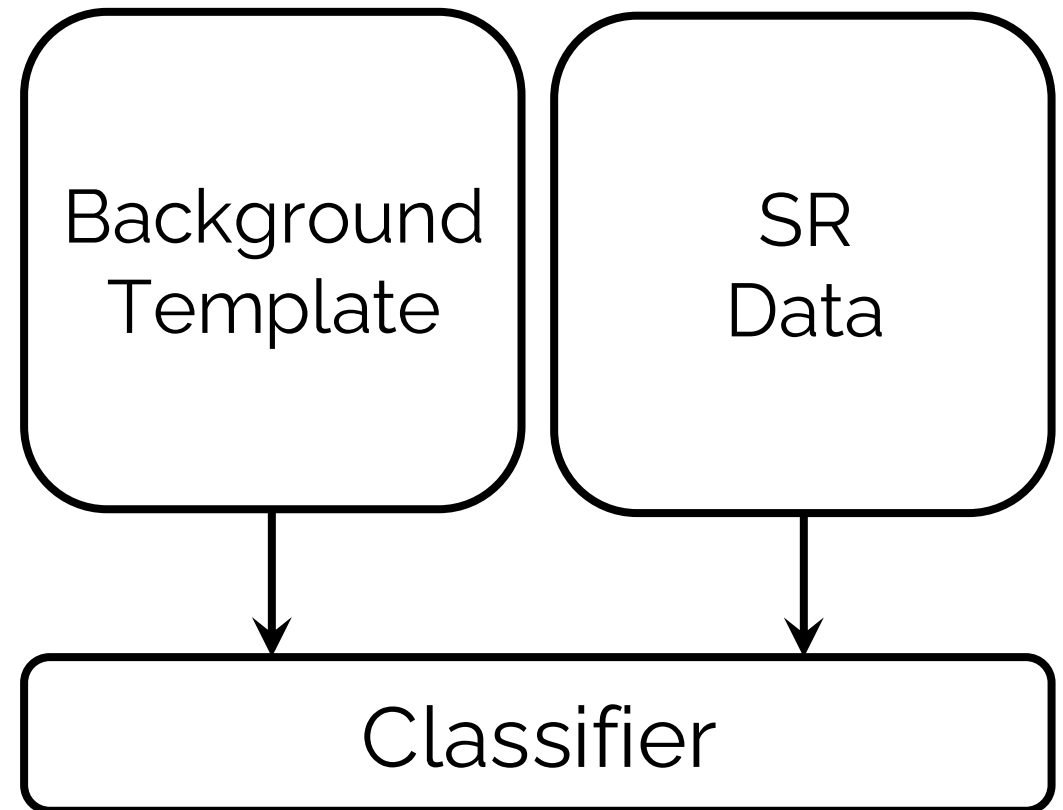
→ Same decision boundaries



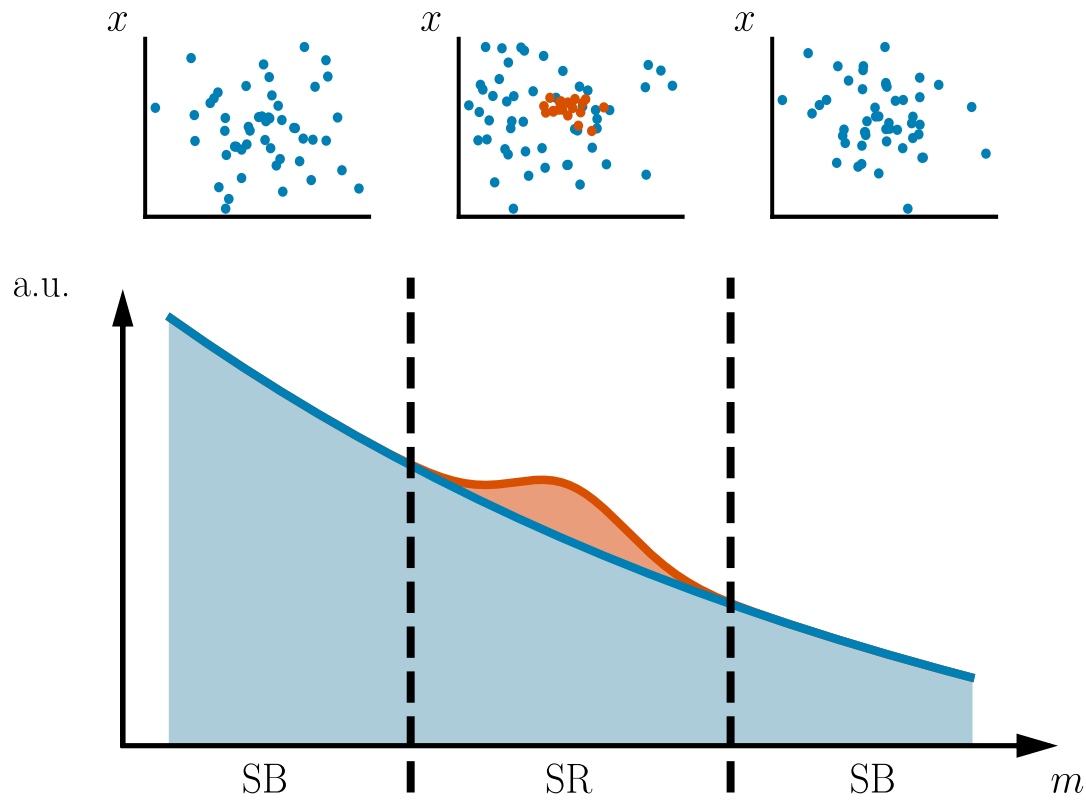
Application to Resonance Searches



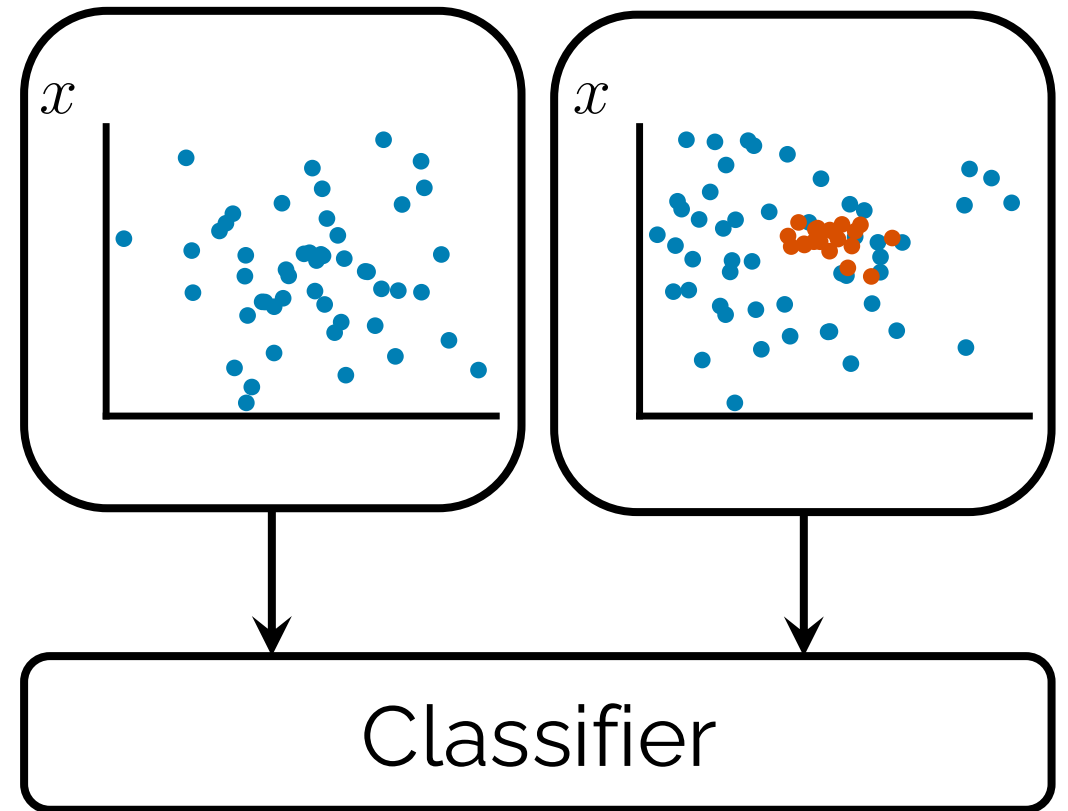
Recreated from [\[2109.00546\]](#)



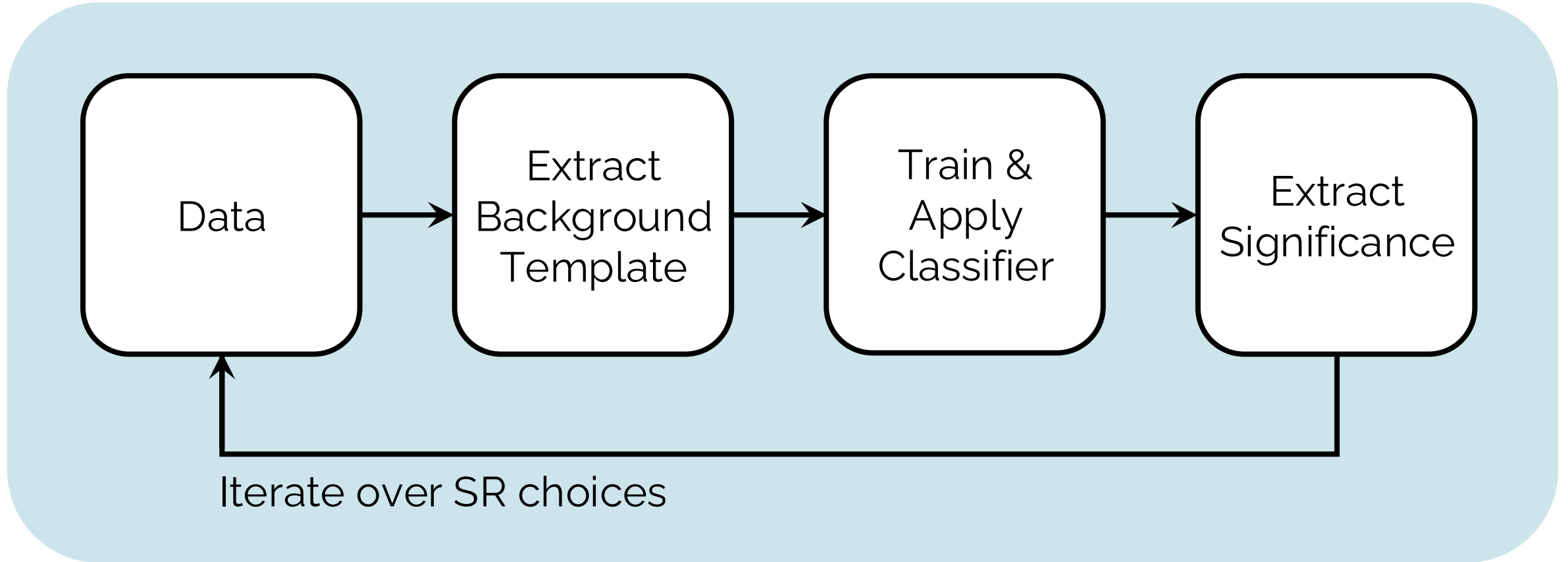
Application to Resonance Searches



Recreated from [\[2109.00546\]](#)



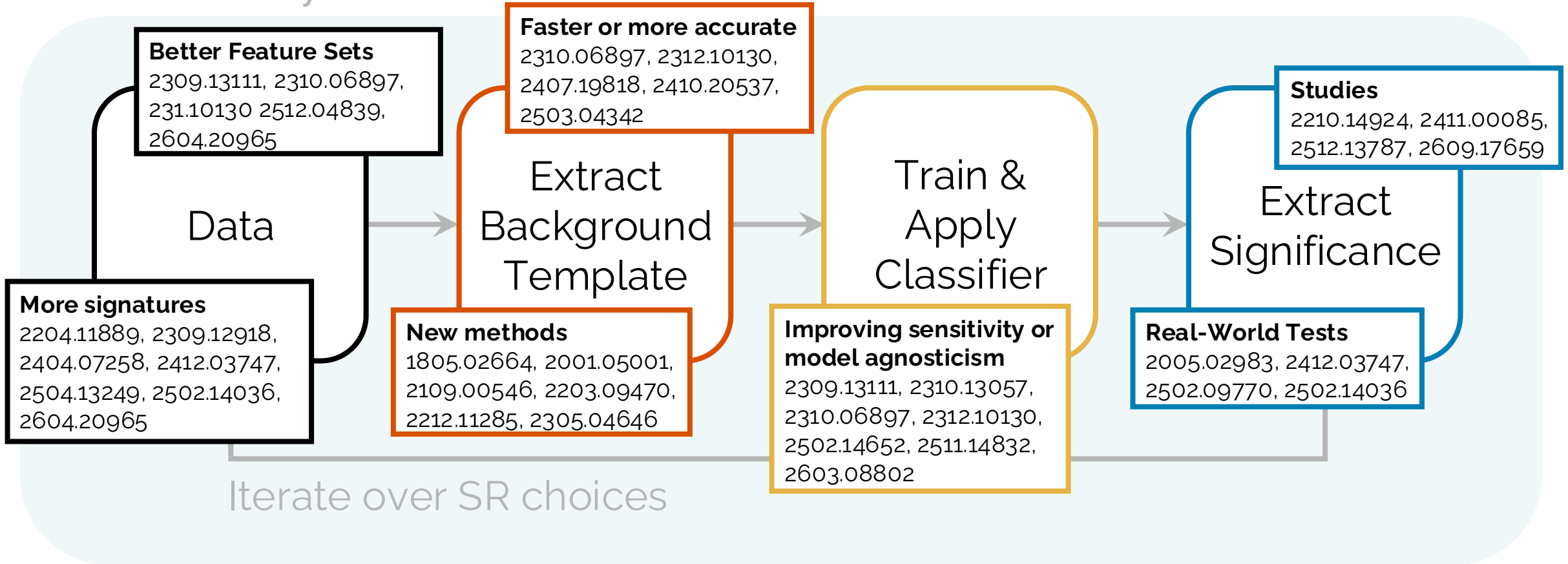
Full Analysis Chain



Recent Progress

Warning: This slide is not exhaustive!

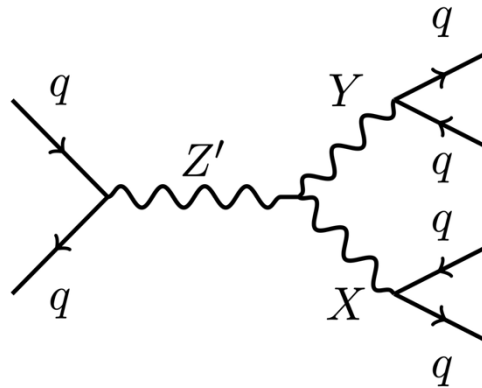
Full Analysis Chain



Data: Where We Started

Benchmark Datasets such as **LHCO Olympics R&D** data set [\[2101.08320\]](#)

→ Compare method performance between different papers



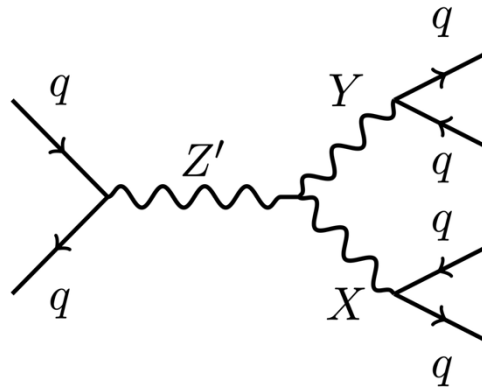
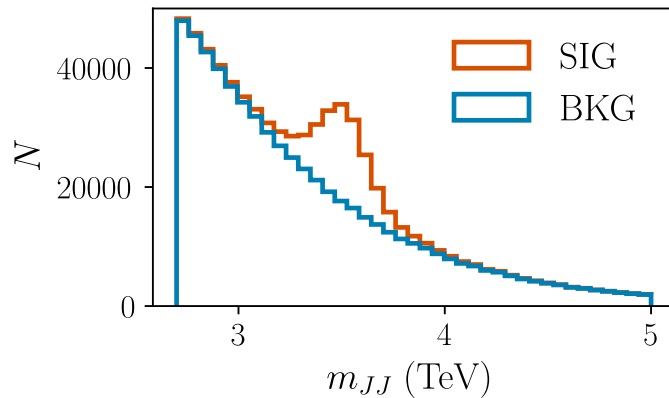
Data: Where We Started

Benchmark Datasets such as **LHCO Olympics R&D** data set [\[2101.08320\]](#)

→ Compare method performance between different papers

Signature

Resonance (dijet)



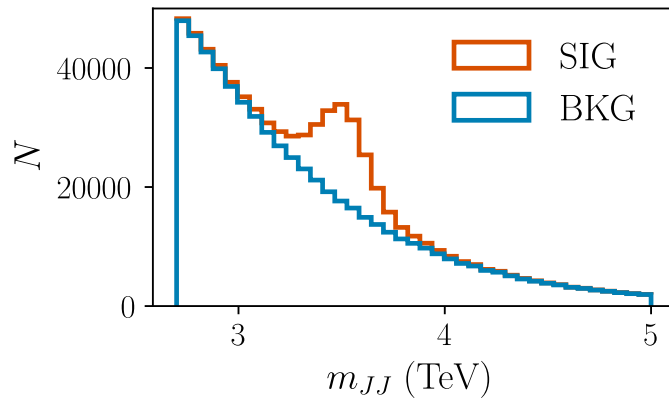
Data: Where We Started

Benchmark Datasets such as **LHCO Olympics R&D** data set [\[2101.08320\]](#)

→ Compare method performance between different papers

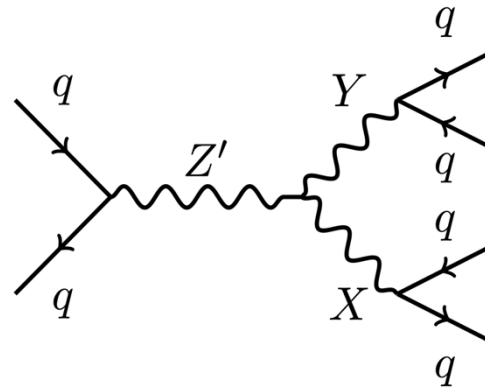
Signature

Resonance (dijet)



Signal

2-prong jets



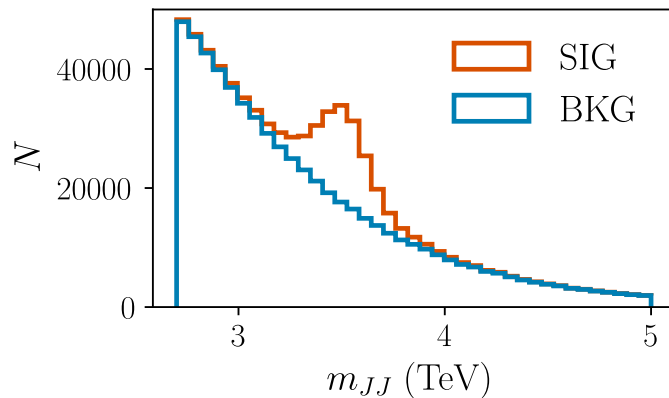
Data: Where We Started

Benchmark Datasets such as **LHCO Olympics R&D** data set [\[2101.08320\]](#)

→ Compare method performance between different papers

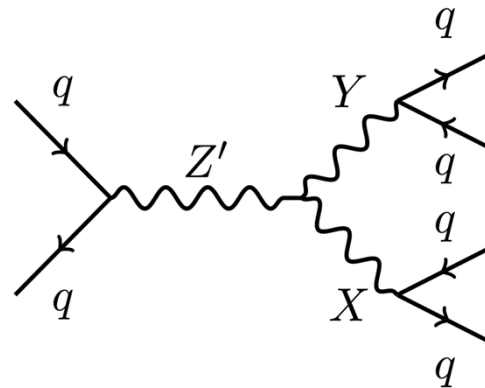
Signature

Resonance (dijet)



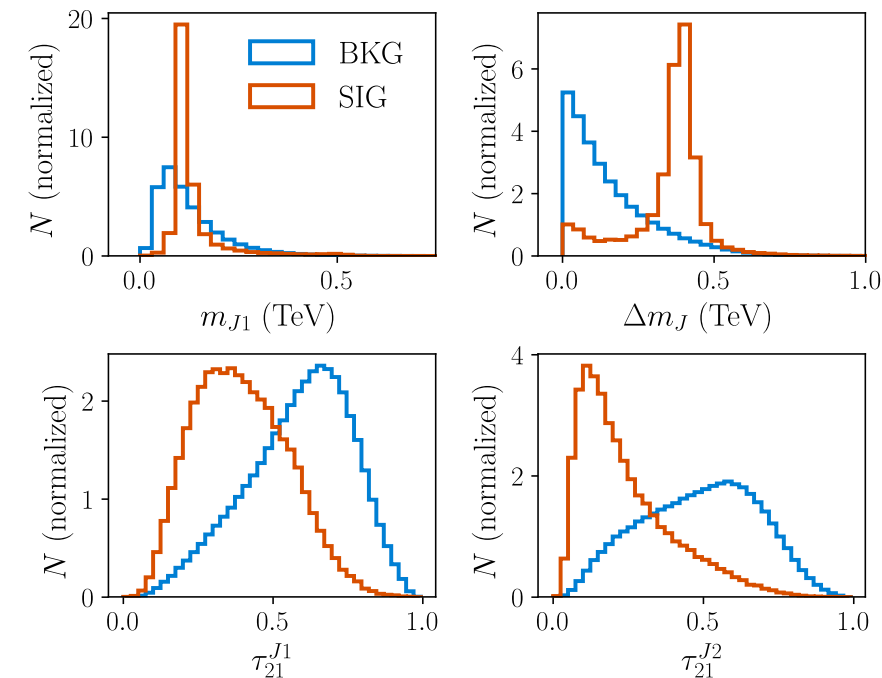
Signal

2-prong jets



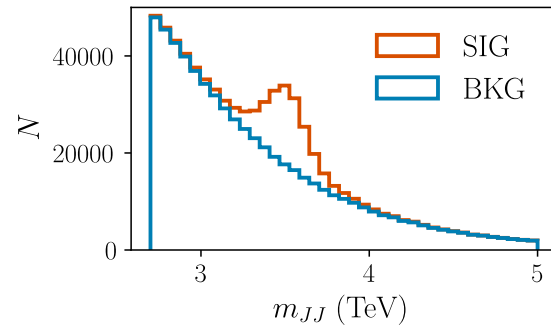
Feature Space

Few high-level observables



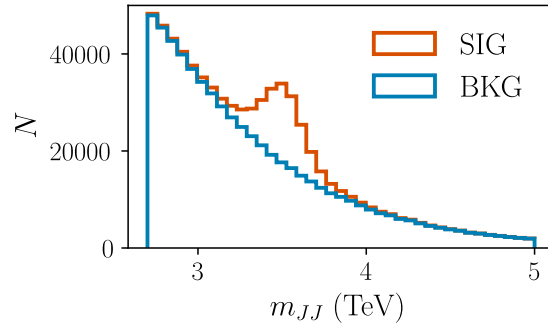
Data: Broadening the Signature Range

Starting point:
Dijet Resonance



Data: Broadening the Signature Range

Starting point:
Dijet Resonance



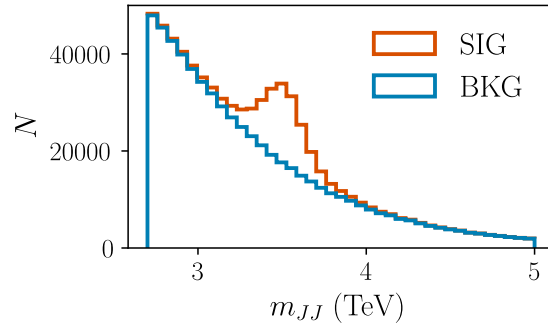
Mixture of dijet and trijet decays in resonant
signal

LHCO Black Box 3, resonant G_{KK} with two decay
modes:

- $G_{KK} \rightarrow Rq$, where $R \rightarrow qq$
- $G_{KK} \rightarrow qq$
→ Need to define a resonant mass carefully

Data: Broadening the Signature Range

Starting point:
Dijet Resonance

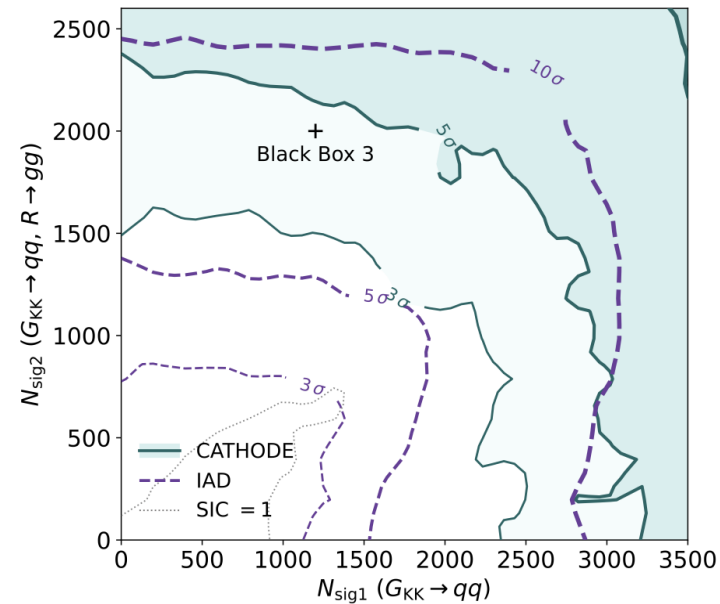


Mixture of dijet and trijet decays in resonant signal

LHCO Black Box 3, resonant G_{KK} with two decay modes:

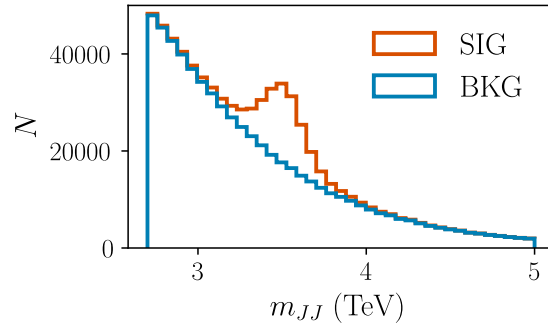
- $G_{KK} \rightarrow Rq$, where $R \rightarrow qq$
- $G_{KK} \rightarrow qq$
- Need to define a resonant mass carefully

[2609.00132]



Data: Broadening the Signature Range

Starting point:
Dijet Resonance

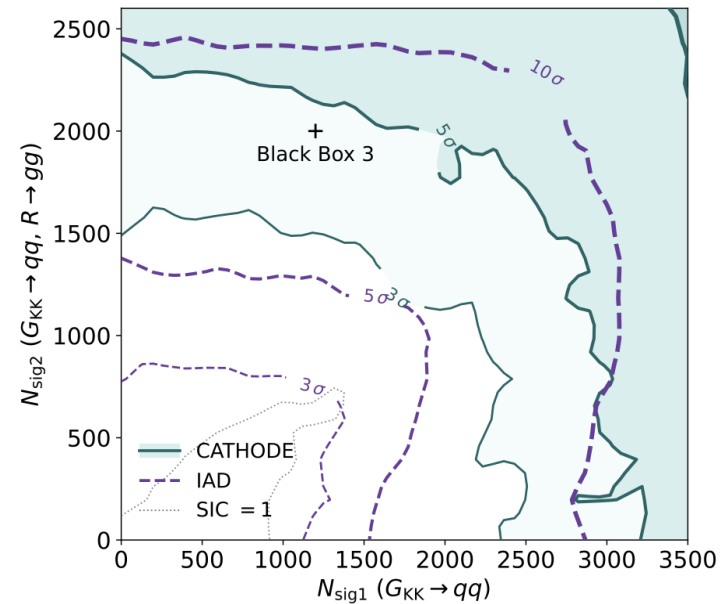


Mixture of dijet and trijet decays in resonant
signal

LHCO Black Box 3, resonant G_{KK} with two decay
modes:

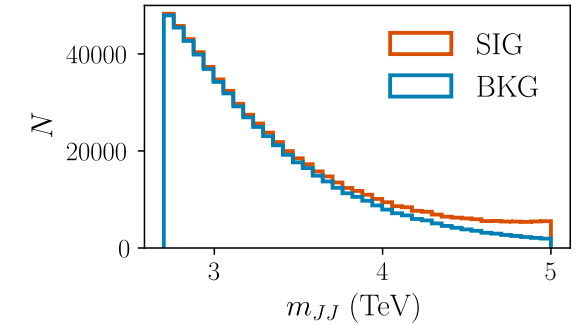
- $G_{KK} \rightarrow Rq$, where $R \rightarrow qq$
 - $G_{KK} \rightarrow qq$
- Need to define a resonant mass carefully

[2609.00132]



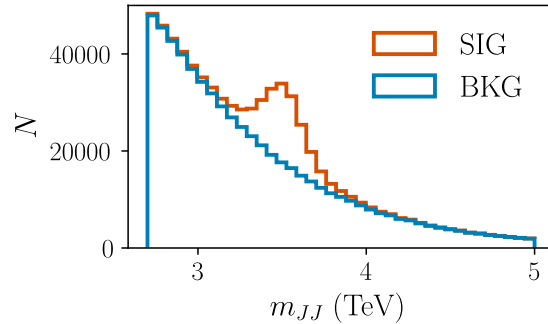
Tail-based anomalies

[2309.12918, 2404.07258]



Data: Broadening the Signature Range

Starting point:
Dijet Resonance

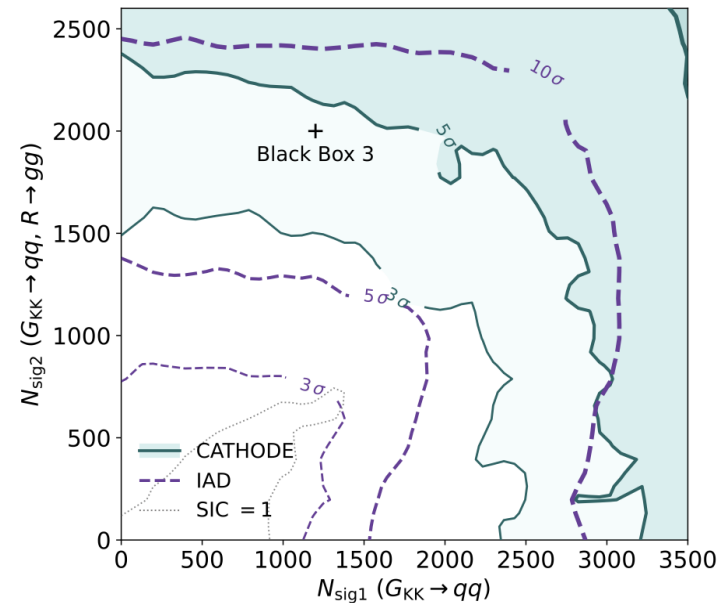


Mixture of dijet and trijet decays in resonant signal

LHCO Black Box 3, resonant G_{KK} with two decay modes:

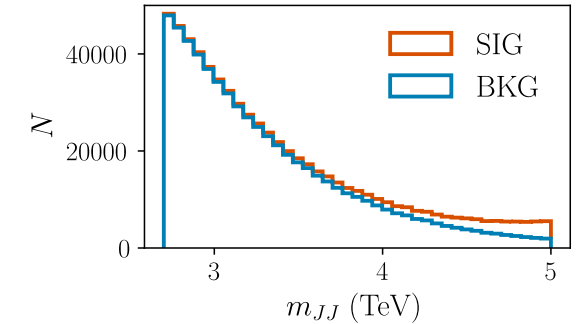
- $G_{KK} \rightarrow Rq$, where $R \rightarrow qq$
 - $G_{KK} \rightarrow qq$
- Need to define a resonant mass carefully

[2609.00132]



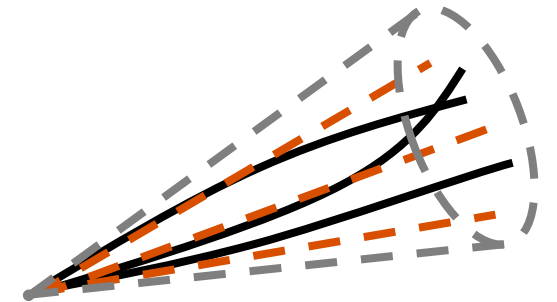
Tail-based anomalies

[2309.12918, 2404.07258]



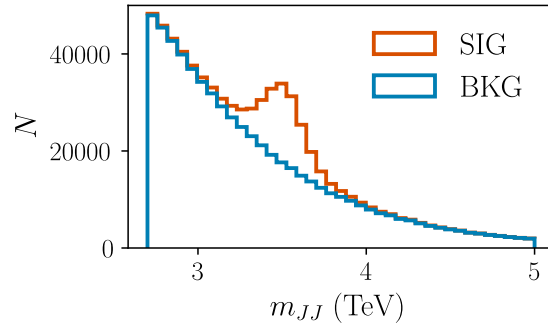
Dark sector jets

[2202.00686, 2204.11889]



Data: Broadening the Signature Range

Starting point:
Dijet Resonance

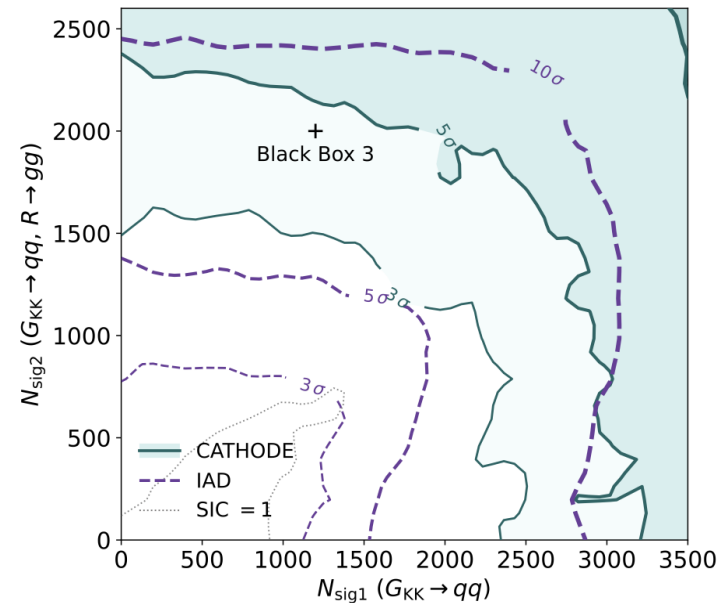


Mixture of dijet and trijet decays in resonant
signal

LHCO Black Box 3, resonant G_{KK} with two decay
modes:

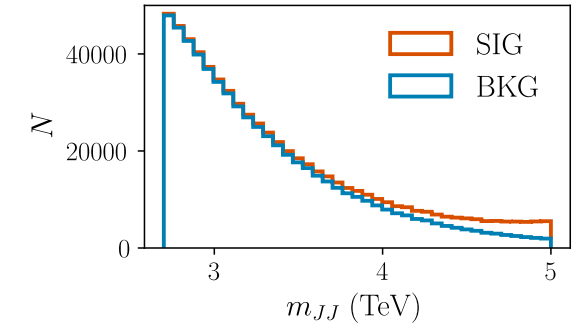
- $G_{KK} \rightarrow Rq$, where $R \rightarrow qq$
 - $G_{KK} \rightarrow qq$
- Need to define a resonant mass carefully

[2609.00132]



Tail-based anomalies

[2309.12918, 2404.07258]

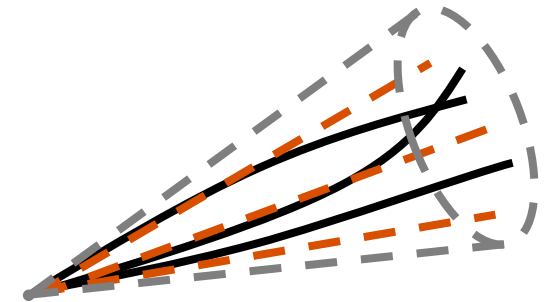


Related talks

→ Dennis' talk for Higgs+X
models

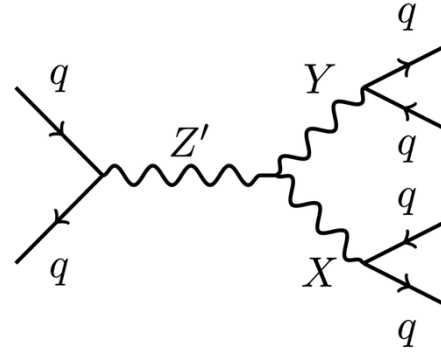
Dark sector jets

[2202.00686, 2204.11889]



Data: Studying Sensitivity to More Signals

Starting point:
2-prong jets

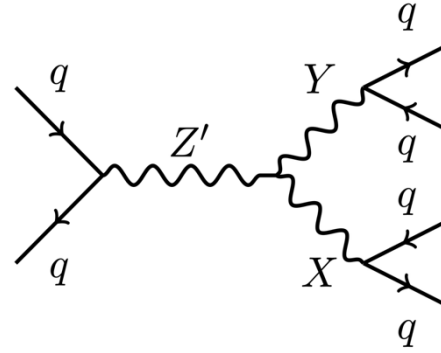


Data: Studying Sensitivity to More Signals

Testing Signal Sensitivity

- To estimate model agnostic sensitivity, need to consider more signals

Starting point:
2-prong jets

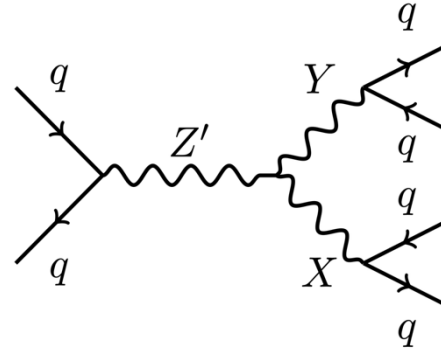


Data: Studying Sensitivity to More Signals

Testing Signal Sensitivity

- To estimate model agnostic sensitivity, need to consider more signals
- Even small changes to the signal can lead to large reduction in sensitivity
→ LHCO signal with $X/Y \rightarrow qqq$

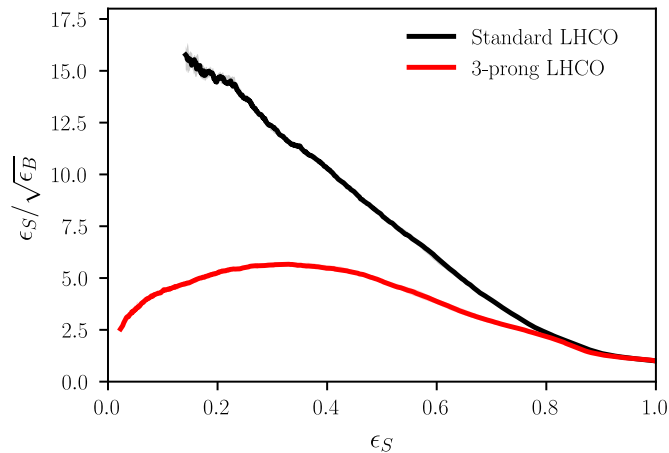
Starting point:
2-prong jets



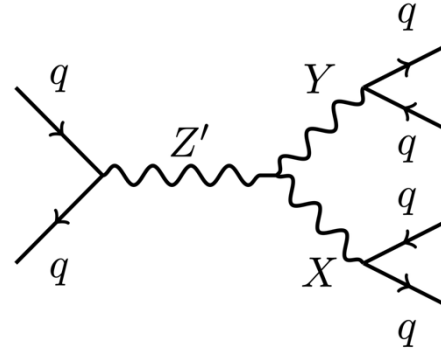
Data: Studying Sensitivity to More Signals

Testing Signal Sensitivity

- To estimate model agnostic sensitivity, need to consider more signals
- Even small changes to the signal can lead to large reduction in sensitivity
→ LHCO signal with $X/Y \rightarrow qq\bar{q}$



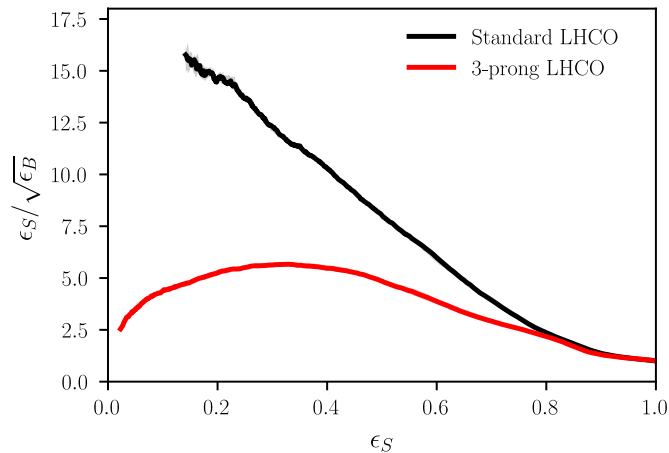
Starting point:
2-prong jets



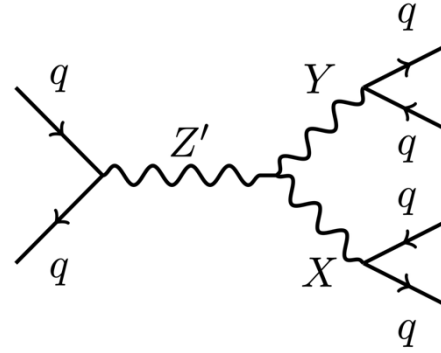
Data: Studying Sensitivity to More Signals

Testing Signal Sensitivity

- To estimate model agnostic sensitivity, need to consider more signals
- Even small changes to the signal can lead to large reduction in sensitivity
→ LHCO signal with $X/Y \rightarrow qq\bar{q}$



Starting point:
2-prong jets

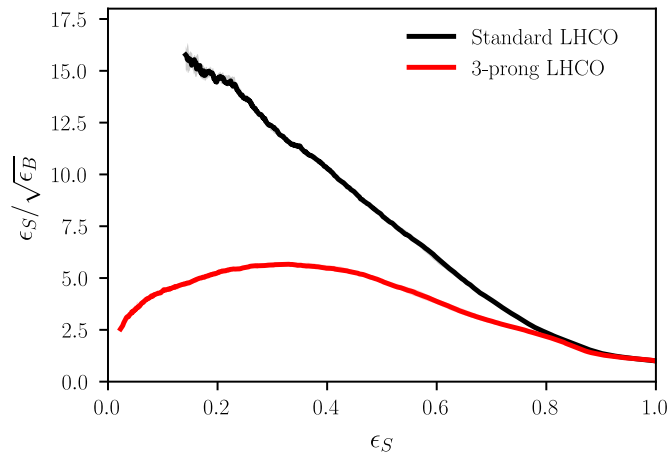


- To be sensitive to many models, need to use larger feature sets

Data: Studying Sensitivity to More Signals

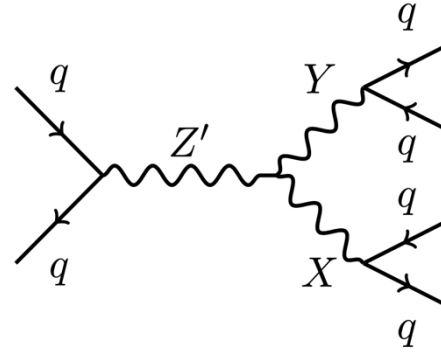
Testing Signal Sensitivity

- To estimate model agnostic sensitivity, need to consider more signals
- Even small changes to the signal can lead to large reduction in sensitivity
→ LHCO signal with $X/Y \rightarrow qq$

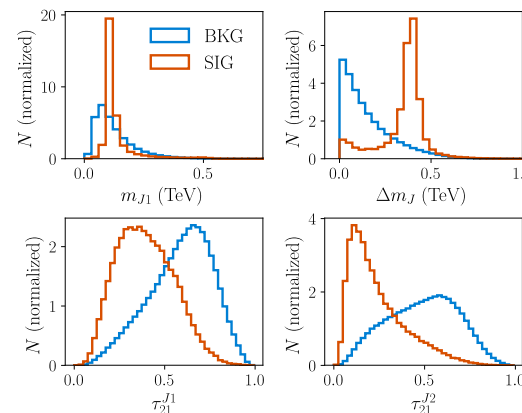


- To be sensitive to many models, need to use larger feature sets

Starting point:
2-prong jets



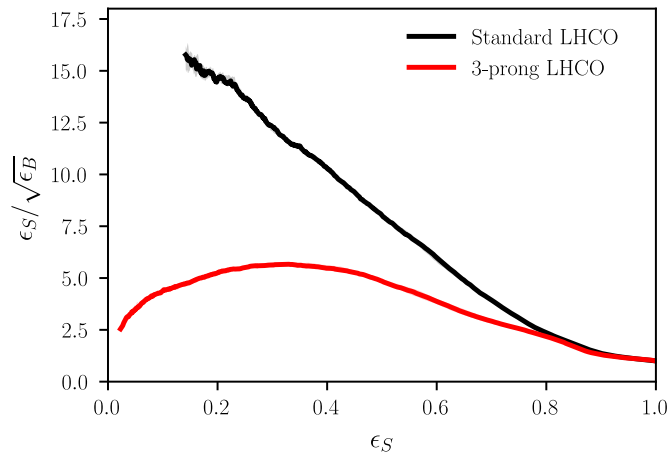
Starting point:
Few high-level features



Data: Studying Sensitivity to More Signals

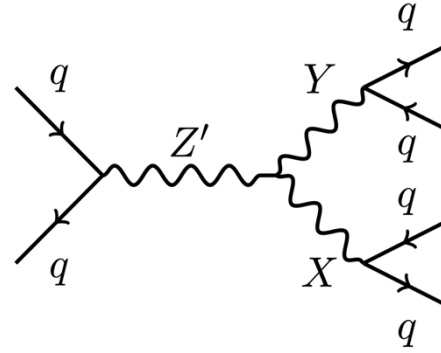
Testing Signal Sensitivity

- To estimate model agnostic sensitivity, need to consider more signals
- Even small changes to the signal can lead to large reduction in sensitivity
→ LHCO signal with $X/Y \rightarrow qq$



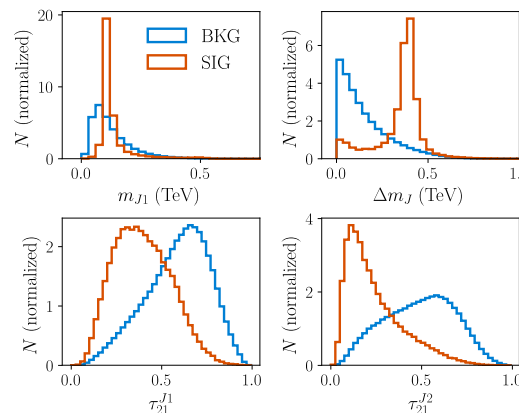
- To be sensitive to many models, need to use larger feature sets

Starting point:
2-prong jets



Low-level features

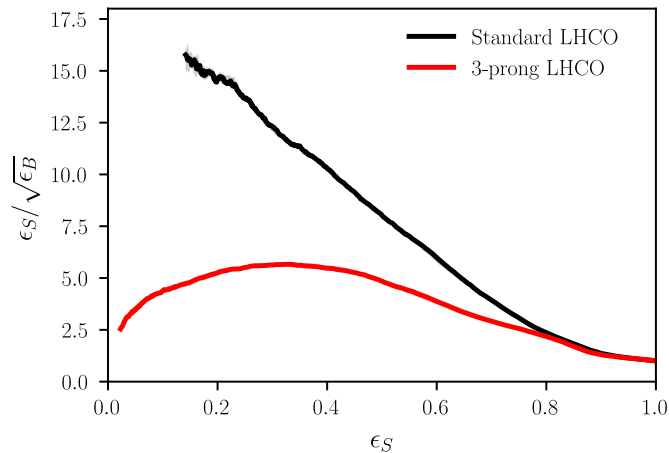
Starting point:
Few high-level features



Data: Studying Sensitivity to More Signals

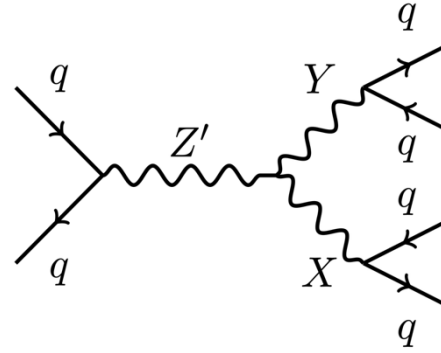
Testing Signal Sensitivity

- To estimate model agnostic sensitivity, need to consider more signals
- Even small changes to the signal can lead to large reduction in sensitivity
→ LHCO signal with $X/Y \rightarrow qq$



- To be sensitive to many models, need to use larger feature sets

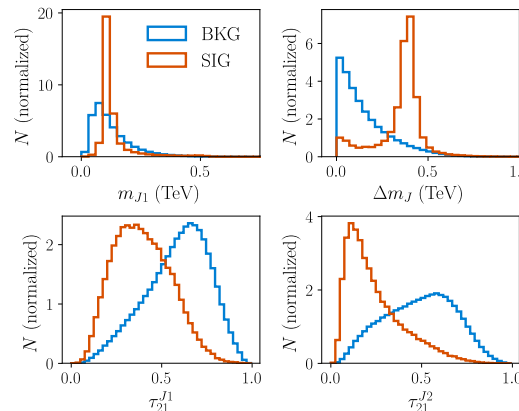
Starting point:
2-prong jets



Low-level features

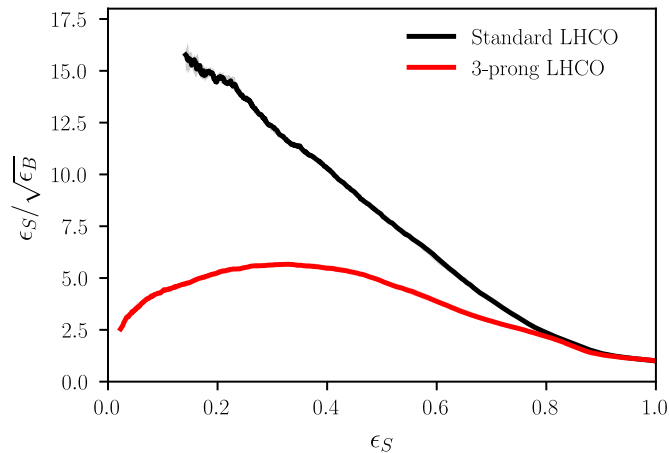
- “The dream” → minimal model bias, particle level information should contain everything

Starting point:
Few high-level features



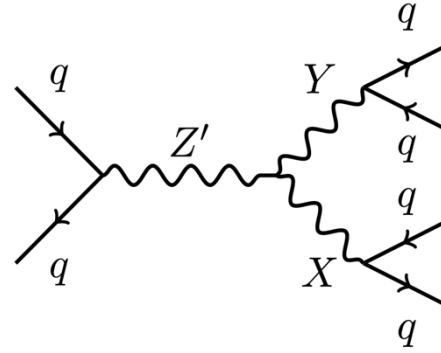
Testing Signal Sensitivity

- To estimate model agnostic sensitivity, need to consider more signals
- Even small changes to the signal can lead to large reduction in sensitivity
→ LHCO signal with $X/Y \rightarrow qq$

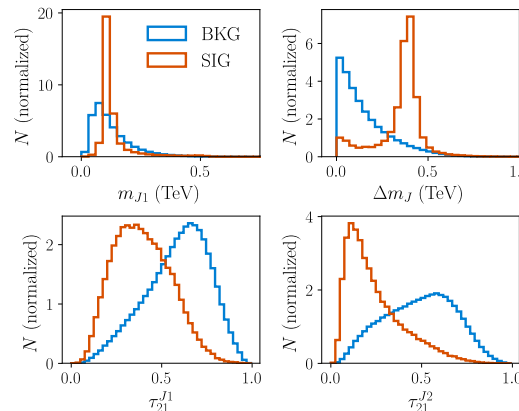


- To be sensitive to many models, need to use larger feature sets

Starting point:
2-prong jets



Starting point:
Few high-level features

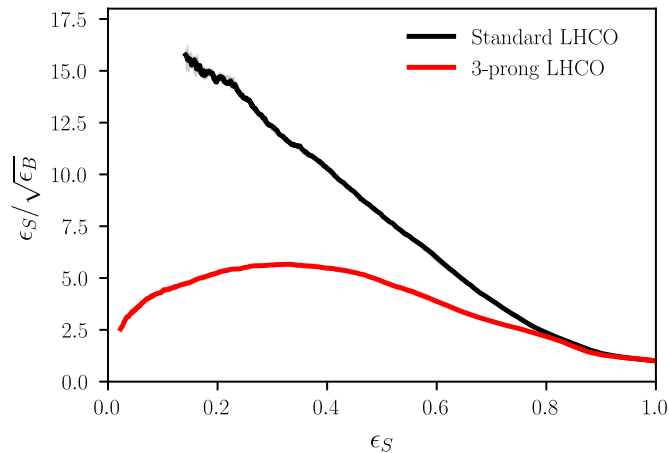


Low-level features

- “The dream” → minimal model bias, particle level information should contain everything
- Problem: need model to derive informative features but very few signal events
→ Possible solution: foundation models
→ *Humberto's talk yesterday (look at the slides and chat to him)*

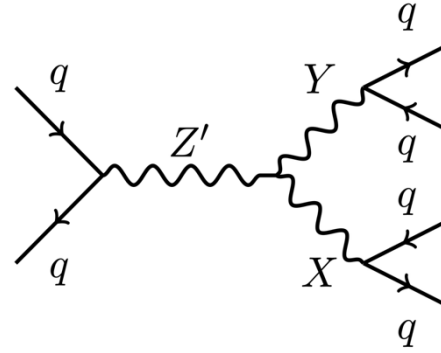
Testing Signal Sensitivity

- To estimate model agnostic sensitivity, need to consider more signals
- Even small changes to the signal can lead to large reduction in sensitivity
→ LHCO signal with $X/Y \rightarrow qq$

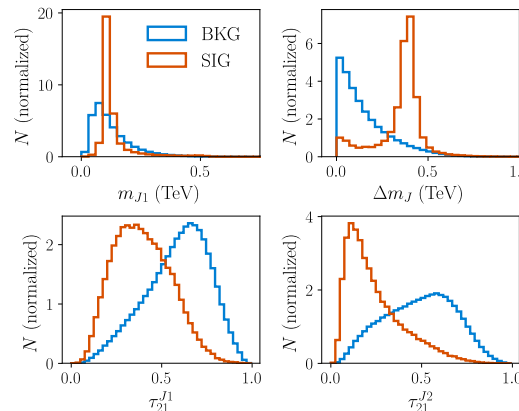


- To be sensitive to many models, need to use larger feature sets

Starting point:
2-prong jets



Starting point:
Few high-level features



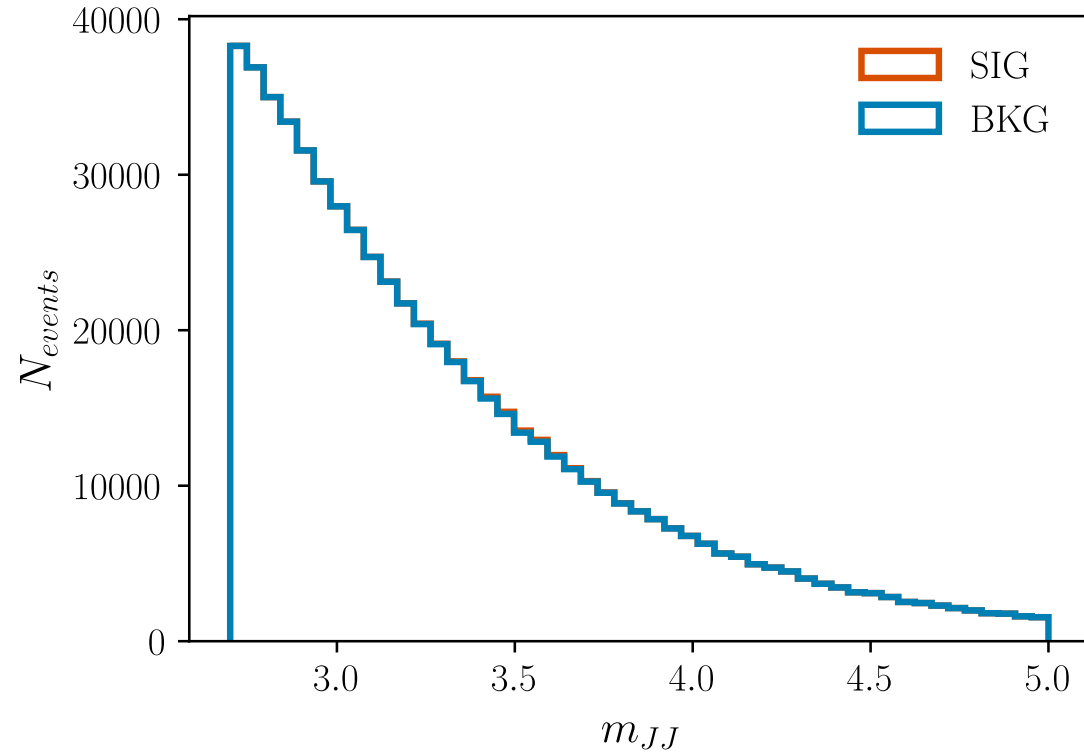
Low-level features

- “The dream” → minimal model bias, particle level information should contain everything
- Problem: need model to derive informative features but very few signal events
→ Possible solution: foundation models
→ *Humberto's talk yesterday (look at the slides and chat to him)*

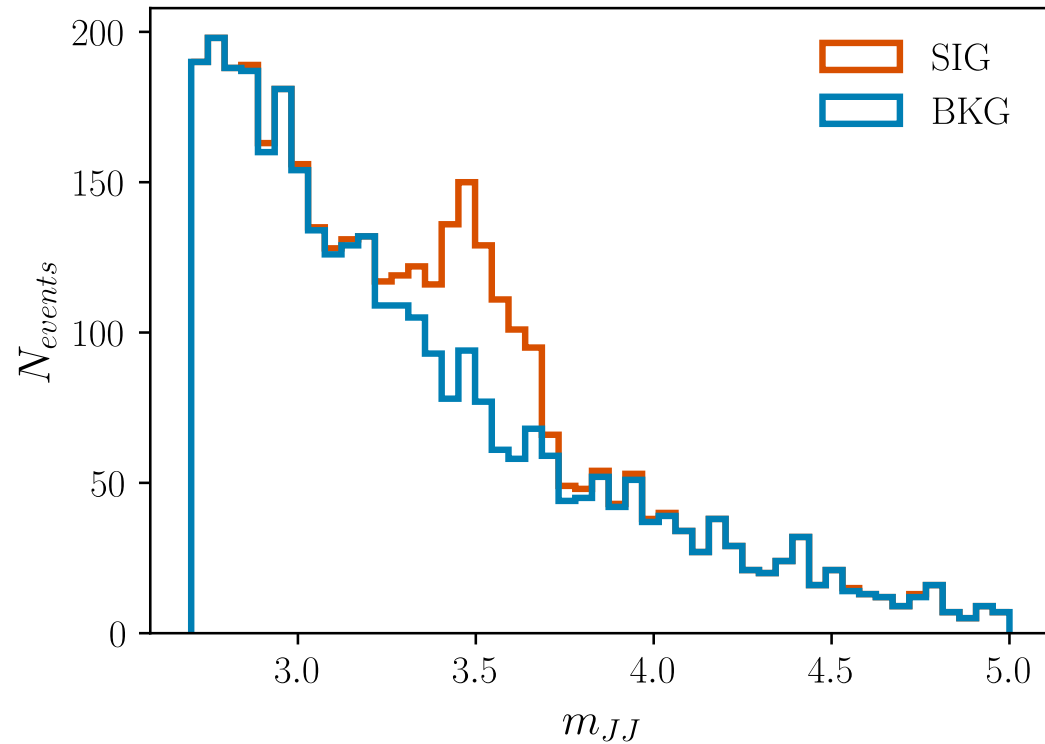
More high-level features

- Features designed to be useful
→ Can typically find low signal injections and find many signals with large feature sets
→ *Lukas' talk*

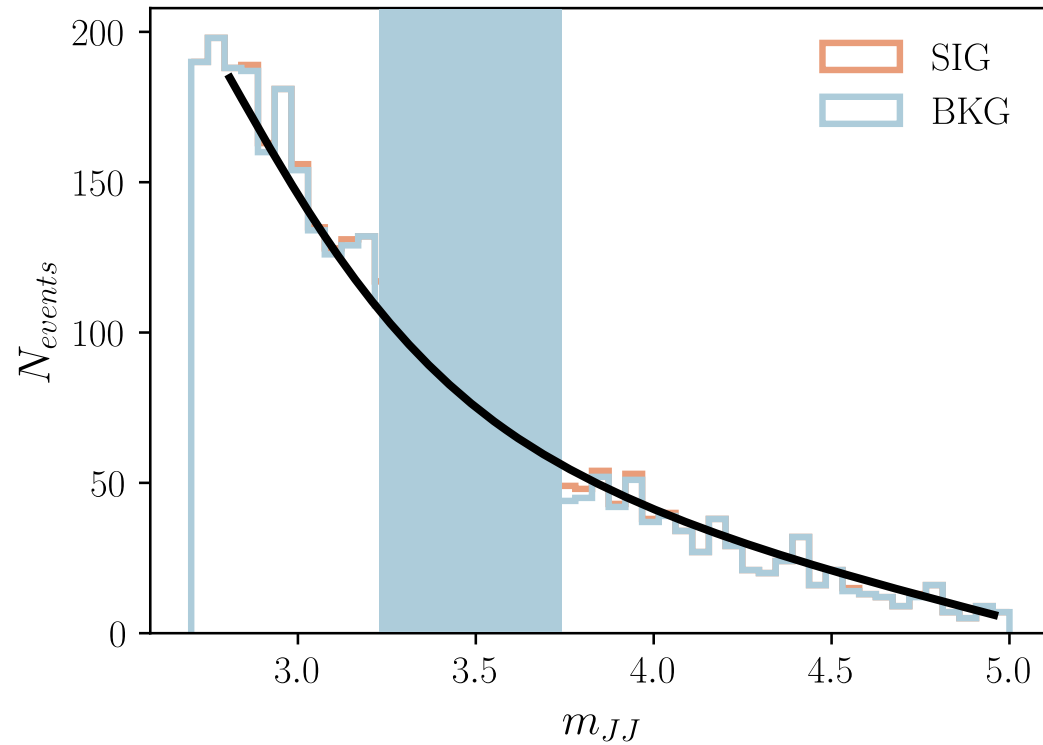
Background Sculpting: Problem



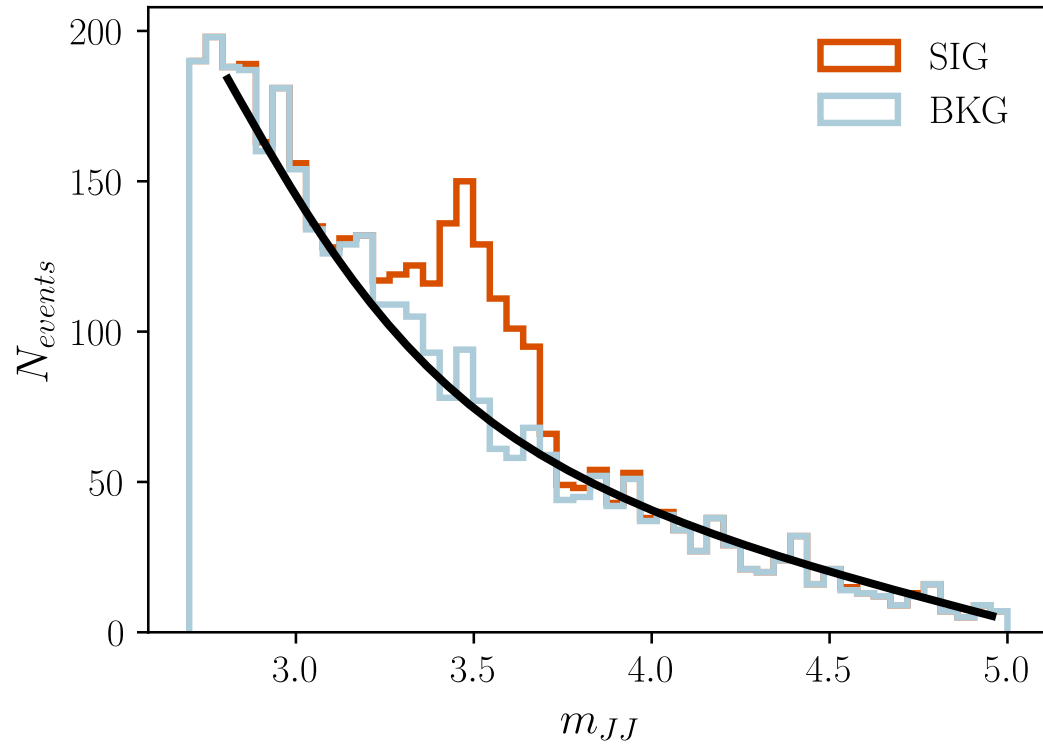
Background Sculpting: Problem



Background Sculpting: Problem

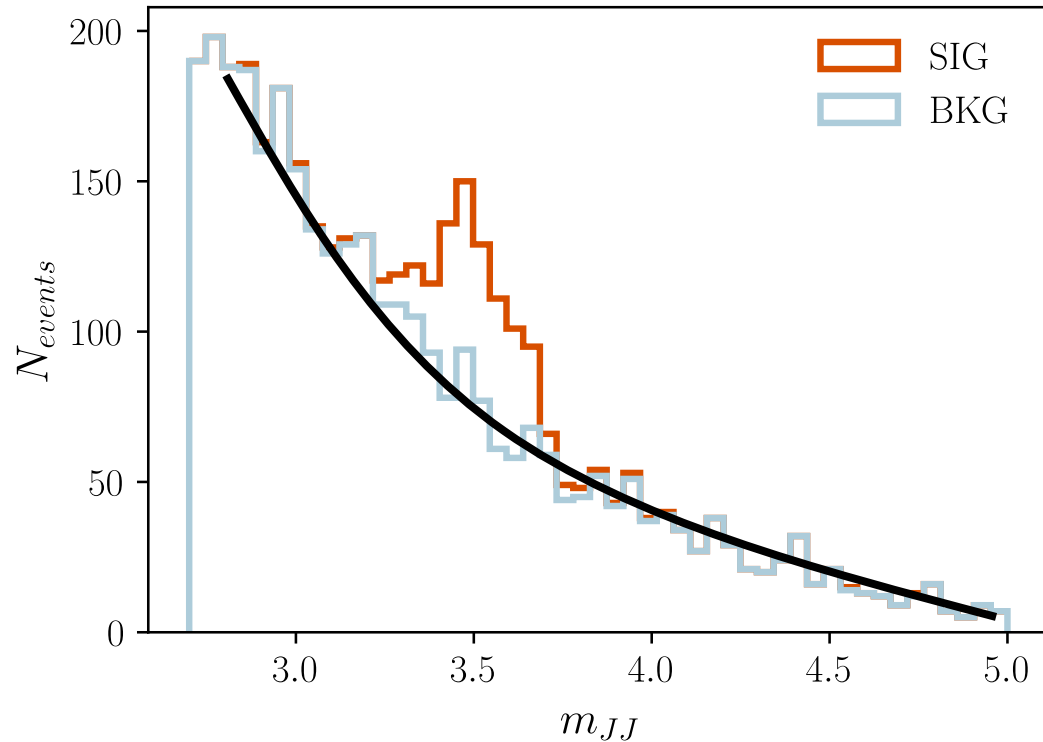


Background Sculpting: Problem

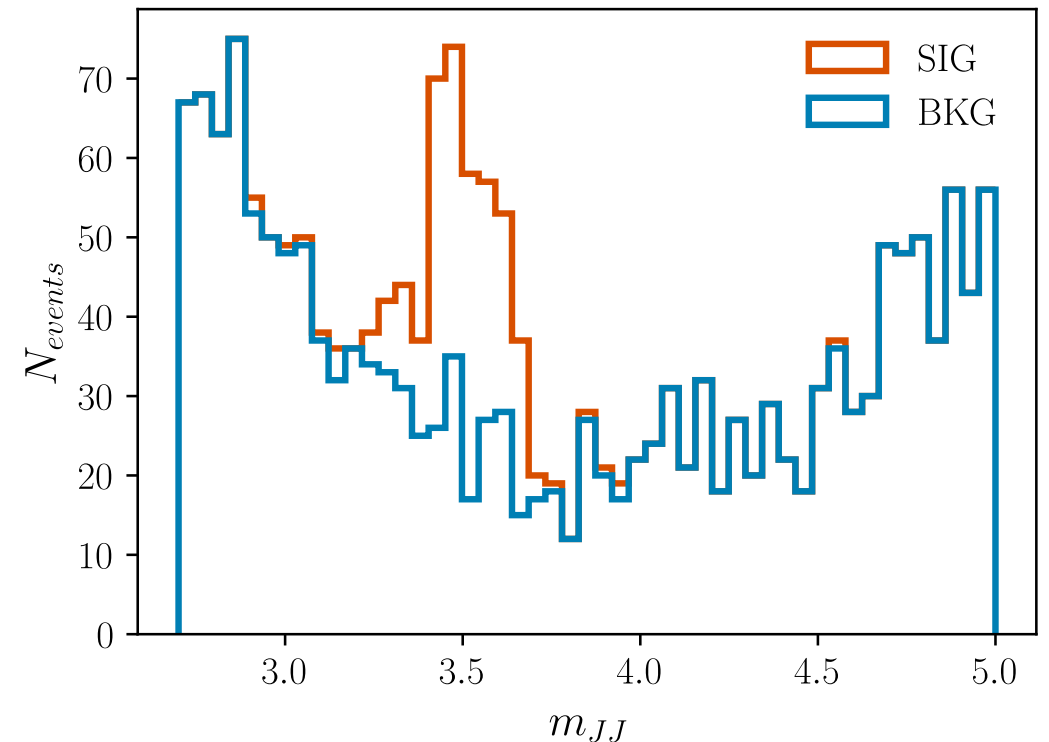


Background Sculpting: Problem

Features uncorrelated with m_{JJ}

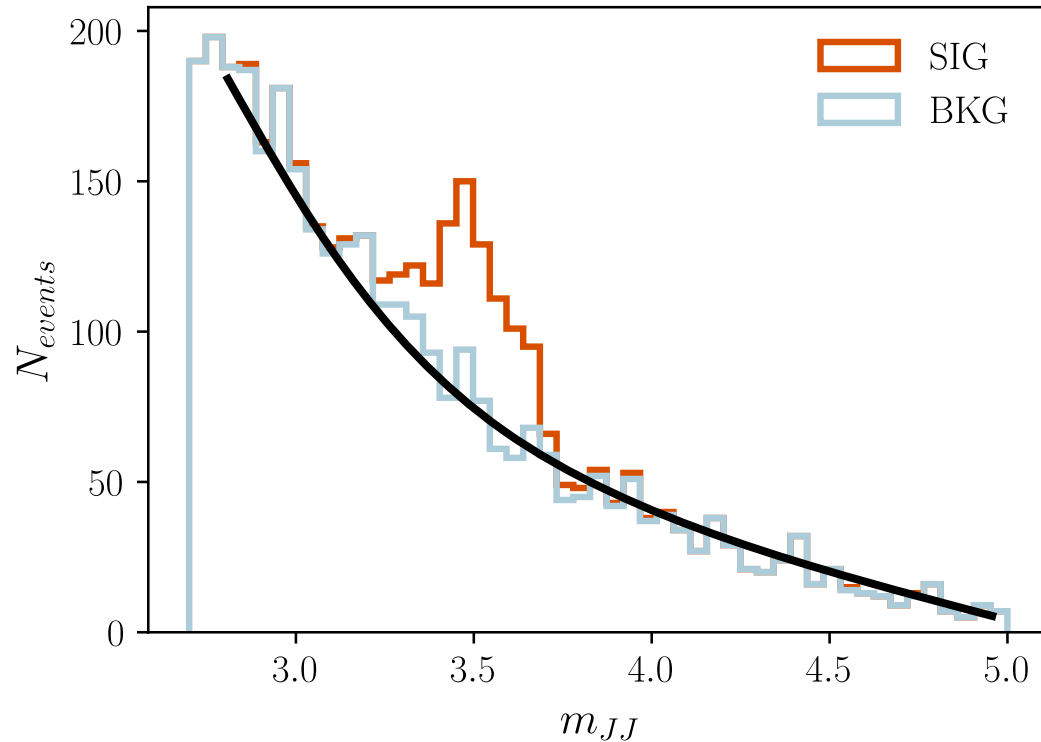


Features correlated with m_{JJ}

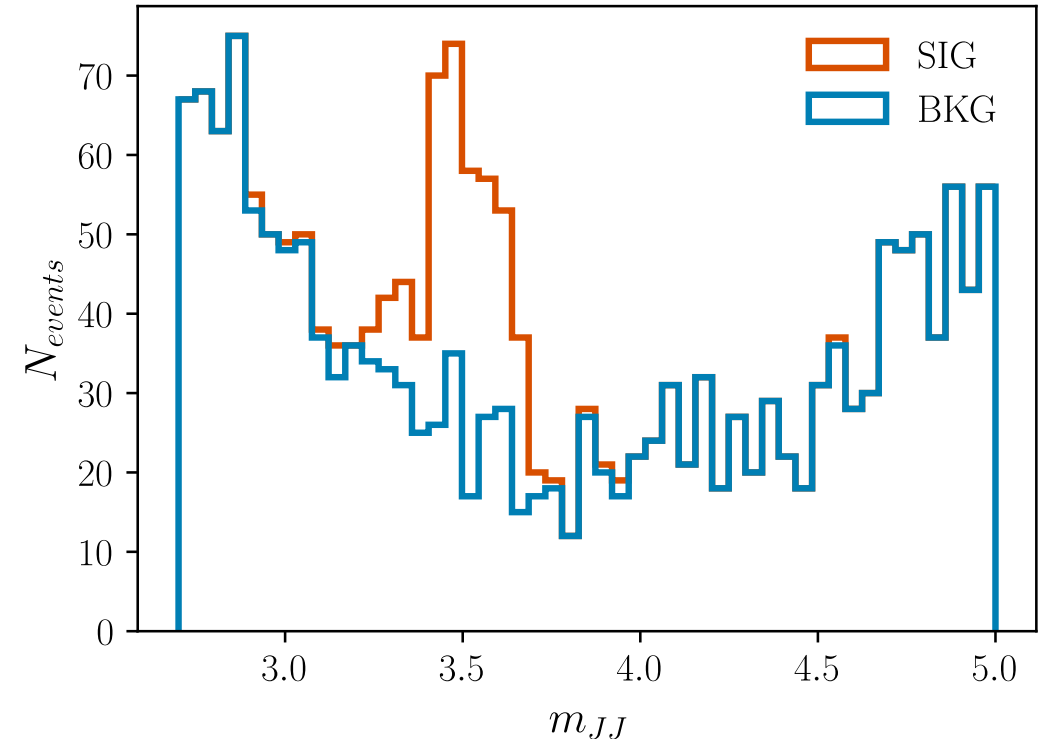


Background Sculpting: Problem

Features uncorrelated with m_{JJ}



Features correlated with m_{JJ}



In some cases, fit becomes impossible. In others,
sensitivity suffers as fit function becomes more flexible.

Background Sculpting: Solutions

Background Sculpting: Solutions

Decorrelate training input

Background Sculpting: Solutions

Decorrelate training input

- Choose feature set explicitly to reduce correlations

[2502.09770]

Decorrelate training input

- Choose feature set explicitly to reduce correlations

[2502.09770]

- Actively decorrelate input

→ LaCATHODE [2210.14924]
allows this for DE-based
background templates

Background Sculpting: Solutions

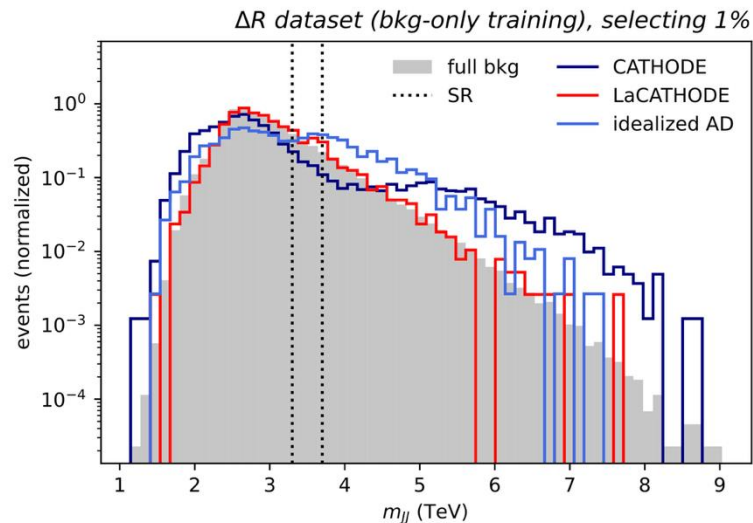
Decorrelate training input

- Choose feature set explicitly to reduce correlations

[\[2502.09770\]](#)

- Actively decorrelate input

→ LaCATHODE [\[2210.14924\]](#)
allows this for DE-based
background templates



Background Sculpting: Solutions

Decorrelate training input

- Choose feature set explicitly to reduce correlations

[\[2502.09770\]](#)

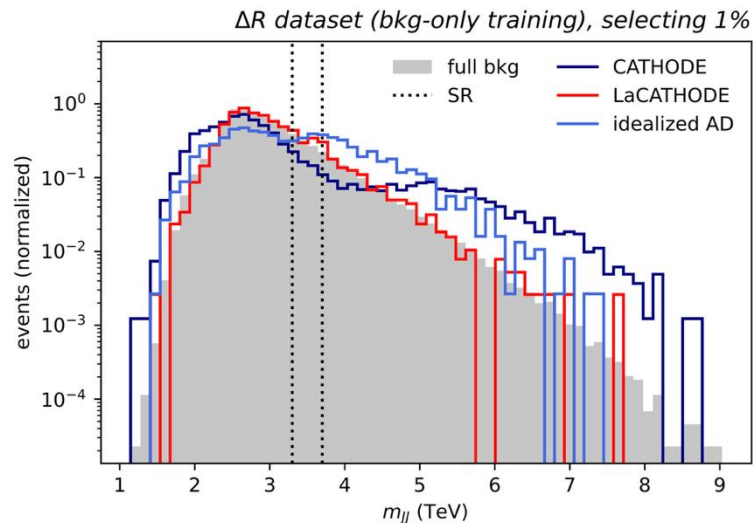
- Actively decorrelate input

→ LaCATHODE [\[2210.14924\]](#)
allows this for DE-based
background templates

Decorrelate cut

- Use quantile regression to make mass-dependent cut

[\[2412.03747\]](#)



Background Sculpting: Solutions

Decorrelate training input

- Choose feature set explicitly to reduce correlations

[\[2502.09770\]](#)

- Actively decorrelate input

→ LaCATHODE [\[2210.14924\]](#)
allows this for DE-based background templates

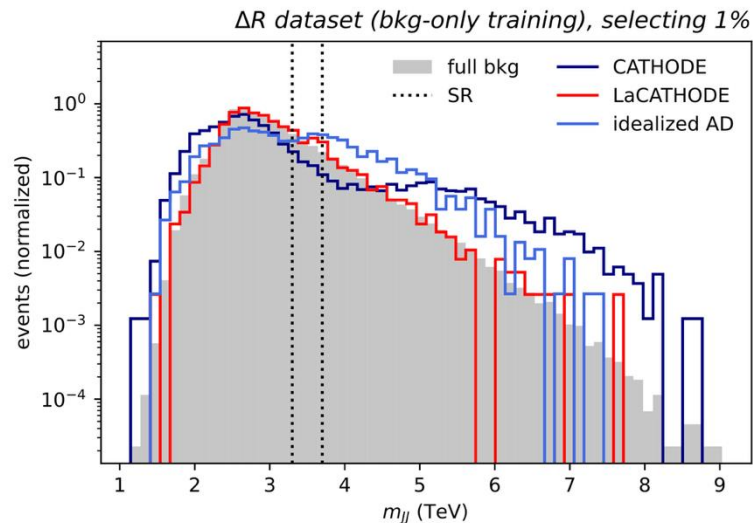
Decorrelate cut

- Use quantile regression to make mass-dependent cut

[\[2412.03747\]](#)

Decorrelate the score itself

→ Vincent's talk



Background Sculpting: Solutions

Decorrelate training input

- Choose feature set explicitly to reduce correlations

[\[2502.09770\]](#)

- Actively decorrelate input

→ LaCATHODE [\[2210.14924\]](#)
allows this for DE-based background templates

Decorrelate cut

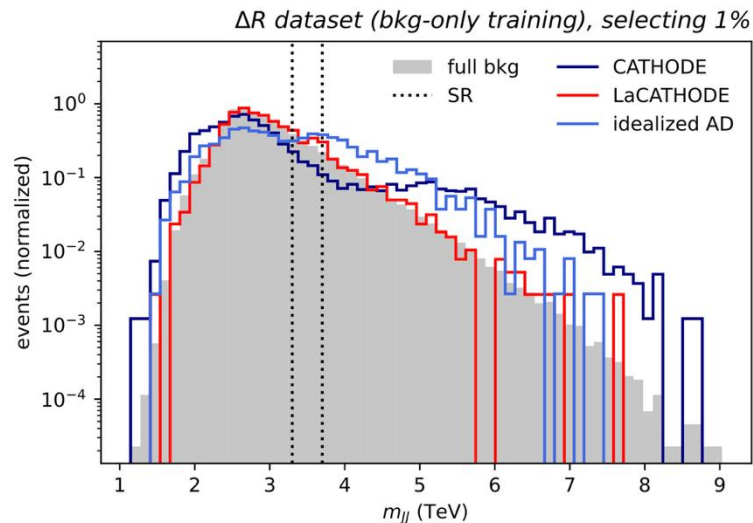
- Use quantile regression to make mass-dependent cut

[\[2412.03747\]](#)

Decorrelate the score itself

→ Vincent's talk

Avoid the fit



Background Sculpting: Solutions

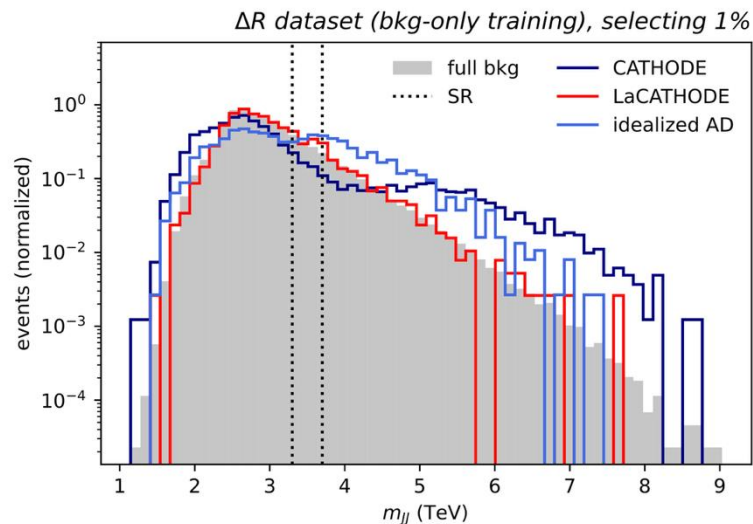
Decorrelate training input

- Choose feature set explicitly to reduce correlations

[\[2502.09770\]](#)

- Actively decorrelate input

→ LaCATHODE [\[2210.14924\]](#)
allows this for DE-based background templates



Decorrelate cut

- Use quantile regression to make mass-dependent cut

[\[2412.03747\]](#)

Decorrelate the score itself

→ Vincent's talk

Avoid the fit

- Compare background template and signal region data using cut-and-count analysis [\[2411.00085\]](#)
- Problem: need to estimate systematic shift for realistic BT

Background Sculpting: Solutions

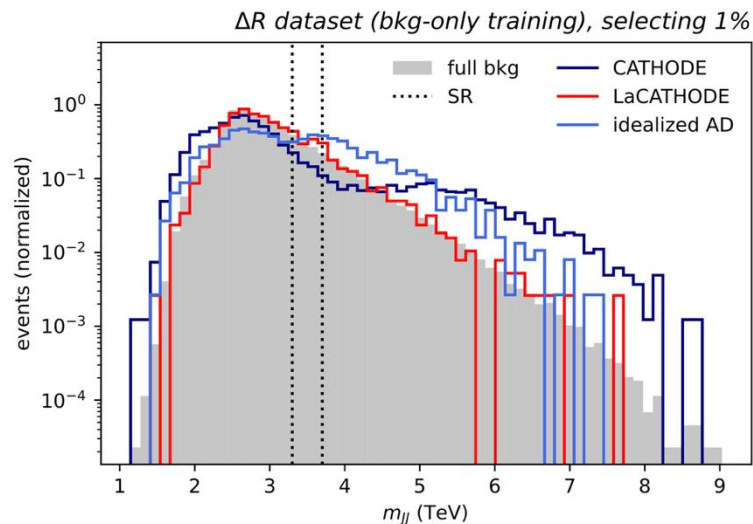
Decorrelate training input

- Choose feature set explicitly to reduce correlations

[\[2502.09770\]](#)

- Actively decorrelate input

→ LaCATHODE [\[2210.14924\]](#)
allows this for DE-based background templates



Decorrelate cut

- Use quantile regression to make mass-dependent cut

[\[2412.03747\]](#)

Decorrelate the score itself

→ Vincent's talk

Avoid the fit

- Compare background template and signal region data using cut-and-count analysis [\[2411.00085\]](#)
- Problem: need to estimate systematic shift for realistic BT

→ Mark's talk

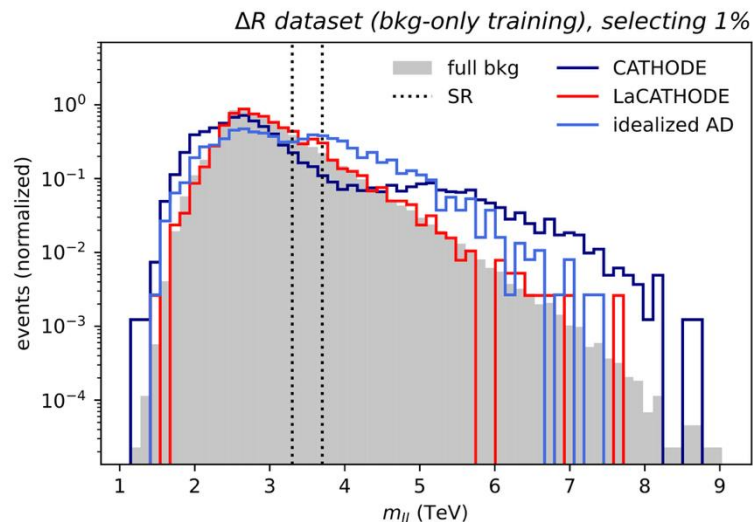
Decorrelate training input

- Choose feature set explicitly to reduce correlations

[\[2502.09770\]](#)

- Actively decorrelate input

→ LaCATHODE [\[2210.14924\]](#)
allows this for DE-based background templates



Decorrelate cut

- Use quantile regression to make mass-dependent cut

[\[2412.03747\]](#)

Decorrelate the score itself

→ Vincent's talk

Avoid the fit

- Compare background template and signal region data using cut-and-count analysis [\[2411.00085\]](#)
- Problem: need to estimate systematic shift for realistic BT

→ Mark's talk

Related topic: Distribution shift in trigger AD

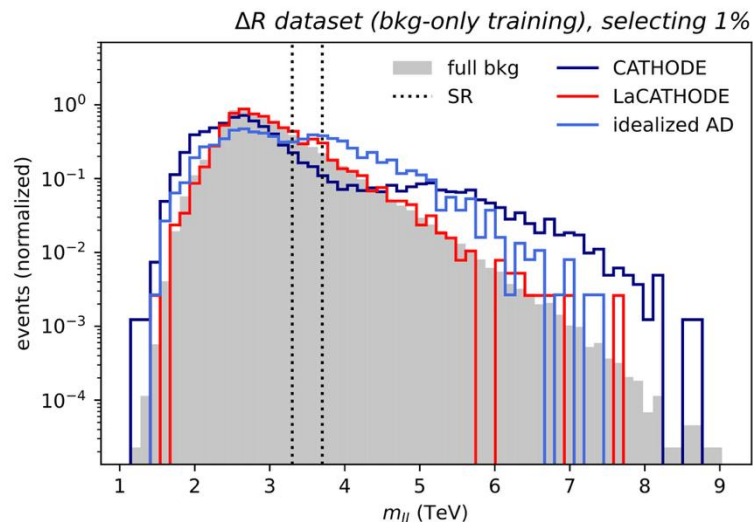
Decorrelate training input

- Choose feature set explicitly to reduce correlations

[\[2502.09770\]](#)

- Actively decorrelate input

→ LaCATHODE [\[2210.14924\]](#)
allows this for DE-based background templates



Decorrelate cut

- Use quantile regression to make mass-dependent cut

[\[2412.03747\]](#)

Decorrelate the score itself

→ Vincent's talk

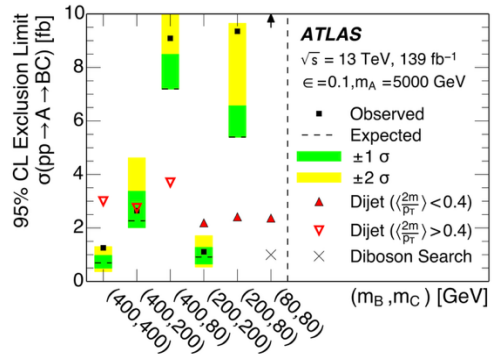
Avoid the fit

- Compare background template and signal region data using cut-and-count analysis [\[2411.00085\]](#)
 - Problem: need to estimate systematic shift for realistic BT
- Mark's talk

Related topic: Distribution shift in trigger AD

→ Amishi's talk

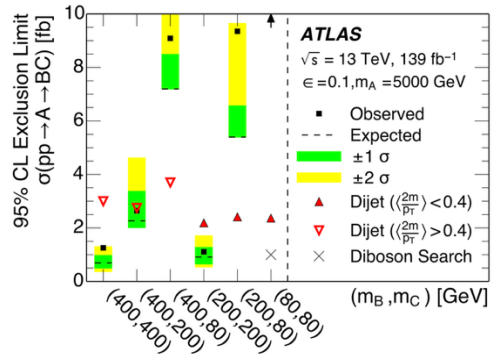
ATLAS '20 [2005.02983]



1st generation search:

- Simple background template
- Just 2 features
- Simple signal topology

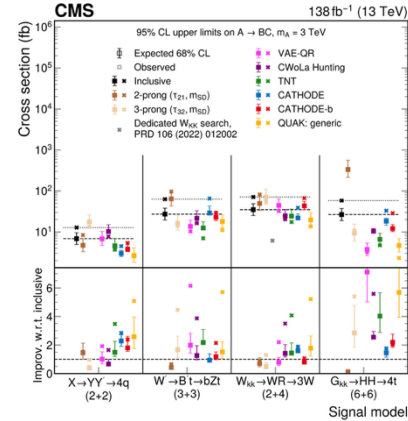
ATLAS '20 [2005.02983]



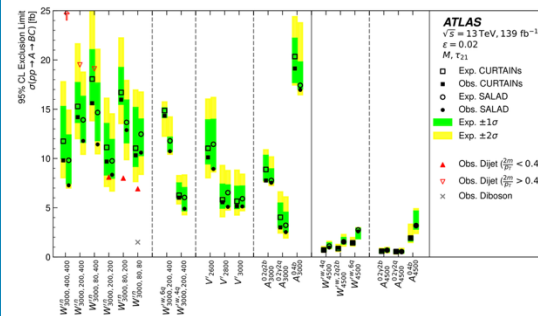
1st generation search:

- Simple background template
- Just 2 features
- Simple signal topology

CMS '24 [2412.03747]



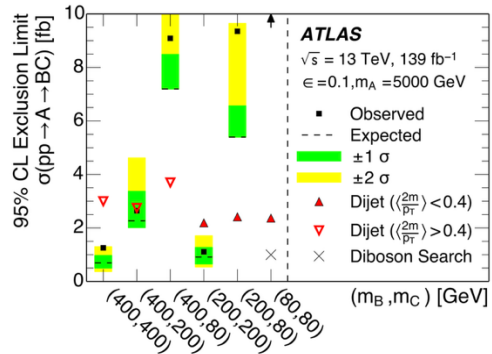
ATLAS '25 [2502.09770]



2nd generation searches improve on:

- Signal variety
- Feature set size
- Background template quality
- Fighting background sculpting

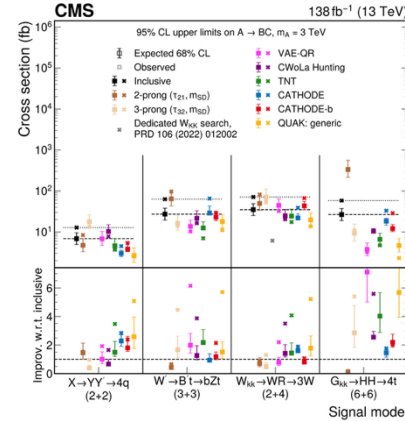
ATLAS '20 [2005.02983]



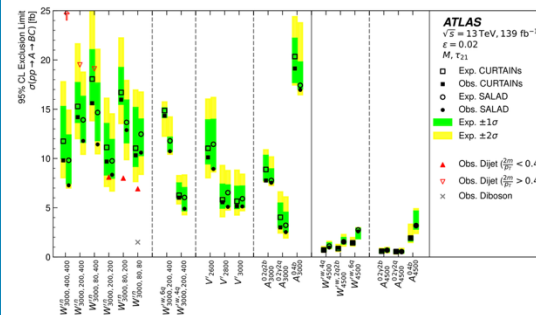
1st generation search:

- Simple background template
- Just 2 features
- Simple signal topology

CMS '24 [2412.03747]



ATLAS '25 [2502.09770]



2nd generation searches improve on:

- Signal variety
- Feature set size
- Background template quality
- Fighting background sculpting

Newer developments and learnings can be taken into account for 3rd generation searches

→ Tore's talk on new CMS developments

Anomaly Detection Task Force

- Task Force within the BSM WG
- First Meeting in May 2026

Marco Letizia
Theory



Tobias Golling
ATLAS



Roberto Seidita
CMS



David Shih
Theory



Oz Amram
CMS



Marie Hein
Theory/CMS



Anomaly Detection Task Force

- Task Force within the BSM WG
- First Meeting in May 2026
- **Goals:**
 - Be a contact point between theory and the experiments
 - Collect and organize community priorities
 - Propose standards for analyses
 - Identify limitations of AD

Marco Letizia
Theory



Tobias Golling
ATLAS



Roberto Seidita
CMS



David Shih
Theory



Oz Amram
CMS



Marie Hein
Theory/CMS



Anomaly Detection Task Force

- Task Force within the BSM WG
- First Meeting in May 2026
- **Goals:**
 - Be a contact point between theory and the experiments
 - Collect and organize community priorities
 - Propose standards for analyses
 - Identify limitations of AD
- **Waypoint:**
 - Working on AD white paper summarizing state of the art and open questions

Marco Letizia
Theory



Tobias Golling
ATLAS



Roberto Seidita
CMS



David Shih
Theory



Oz Amram
CMS



Marie Hein
Theory/CMS



Kick-Off Meeting and Community Survey

July 14th 2026

- Wanted to get an overview of community opinions on potential of AD and what is needed

Kick-Off Meeting and Community Survey

July 14th 2026

- Wanted to get an overview of community opinions on potential of AD and what is needed
- Before meeting conducted community survey
 - 55 responses
 - Asked about background of respondents as well as potential, limitations and needs of AD analyses



Survey

Kick-Off Meeting and Community Survey

July 14th 2026

- Wanted to get an overview of community opinions on potential of AD and what is needed
- Before meeting conducted community survey
 - 55 responses
 - Asked about background of respondents as well as potential, limitations and needs of AD analyses
- During the meeting
 - Presented the survey and ATLAS AD guidelines
 - Had an active discussion about the needs of the community



Survey

Indico

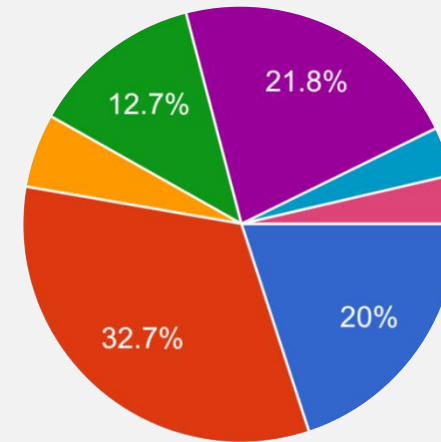


Kick-Off Meeting and Community Survey

July 14th 2026

- Wanted to get an overview of community opinions on potential of AD and what is needed
- Before meeting conducted community survey
 - 55 responses
 - Asked about background of respondents as well as potential, limitations and needs of AD analyses
- During the meeting
 - Presented the survey and ATLAS AD guidelines
 - Had an active discussion about the needs of the community

Which of these would add the most value if made reusable and standardized across the community?



- Shared, validated background-estimation tools
- Common benchmarks & datasets for comparison
- Shared feature/representation backbones
- Shared statistical & look-elsewhere tools
- Shared reinterpretation & analysis-preservation tools
- Turnkey, easy-to-adopt search pipeline
- Other



Survey

Indico



Signal Benchmarks and Common Pool

September 10th 2026

First topical meeting of the AD task force

16:00	→ 16:10	Introduction	🕒 10m
Sprecher: David Shih, Dr. Marco Letizia (INFN e Università Genova (IT)), Marie Hein (RWTH Aachen University), Oz Amram (Fermi National Accelerator Lab. (US)), Roberto Seidita (ETH Zurich (CH)), Tobias Golling (Université de Genève (CH))			
16:10	→ 16:30	ATLAS Efforts	🕒 20m
Sprecher: Theresa Reisch (Université de Genève (CH))			
ATL-COM-PHYS-2...			
16:30	→ 16:45	Dark Shower Models	🕒 15m
Sprecher: Kevin Pedro (Fermi National Accelerator Lab. (US))			
dark_showers.mp4 Dark Showers Tas...			
16:45	→ 17:05	Theory Perspective	🕒 20m
Sprecher: Guilherme Guedes (CERN)			
AD_meeting.pdf theory_perspectiv... theory_perspectiv...			
17:05	→ 17:35	Lightning Talks	🕒 30m
Sprecher: Andre Lessa (IFGW - UNICAMP), Andre Lessa (Universidade de Sao Paulo), Dennis Noll (Stanford University), Julia Lynne Gonski (SLAC National Accelerator Laboratory (US)), Frau Swagata Biswas (National Institute of Technology Durgapur), Tania Robens (Rudjer Boskovic Institute (HR))			
0910_jgonski_lhc... 260910_noll-higg... ADTaskForceMee... AD-Taskforce_Sw... andre_lessa.mp4			
dennis_noll.mp4 julia_gonski.mp4 model_spec_robe... swagata_biswas.... tania_robens.mp4			



Continuing to solicit input, e.g. with a

- Meeting with AD sceptics
 - Identify limitations of AD through listening to their concerns
 - Convince them that AD is useful
- Meeting on reinterpretation of AD analyses
- Pheno-experiment feedback session
 - Why are some things adopted and others not?
 - Talk about communication pathways
 - How to get from proof-of-concept to analyses

Future Plans & How to Contribute

Continuing to solicit input, e.g. with a

- Meeting with AD sceptics
 - Identify limitations of AD through listening to their concerns
 - Convince them that AD is useful
- Meeting on reinterpretation of AD analyses
- Pheno-experiment feedback session
 - Why are some things adopted and others not?
 - Talk about communication pathways
 - How to get from proof-of-concept to analyses

We want the task force to be a **useful resource for the community**, so we want your contribution!

17/09/2026



Mattermost invite to
LHC-BSM-WG
→ Join, then join
“LHC Anomaly
Detection Task
Force”

E-Group invite

