Fairness Methods in Particle Physics Event Classification

Lydia Brenner, Tim van Erven,
Oliver Rieger, Wouter Verkerke, Karel de Vries

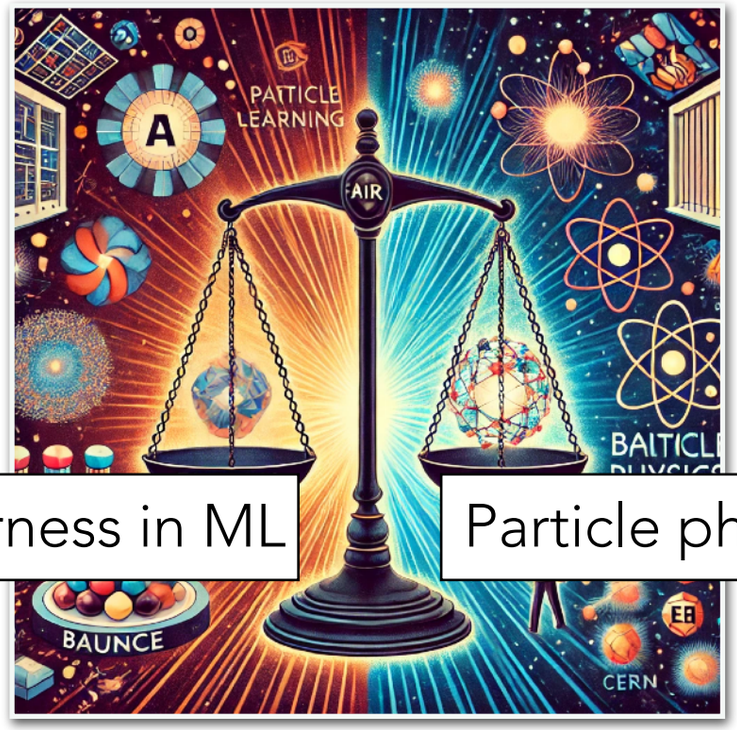
PHYSTAT workshop 2024 - Statistics meets ML
Imperial College London

What is this talk about?

In social sciences…
Fairness in ML *comprises the attempt to correct or eliminate algorithmic bias of gender, ethnicity, or sexual orientation from ML models*

In particle physics…
Searching for a resonant decay of a particle, ***mass-decorrelated classifiers*** *are employed, as the mass is used to perform final signal extraction fit*

Decorrelate the ***protected attribute*** *from the* ***ML model output***



Fairness in ML

Particle physics

Implicitly using fair classifiers to search for new particles

What is Fairness?

What is Fairness?

Fairness means everyone gets the same

Fairness means everyone gets what they need

As you can see the answer to this question depends on the problem...

Nikhef

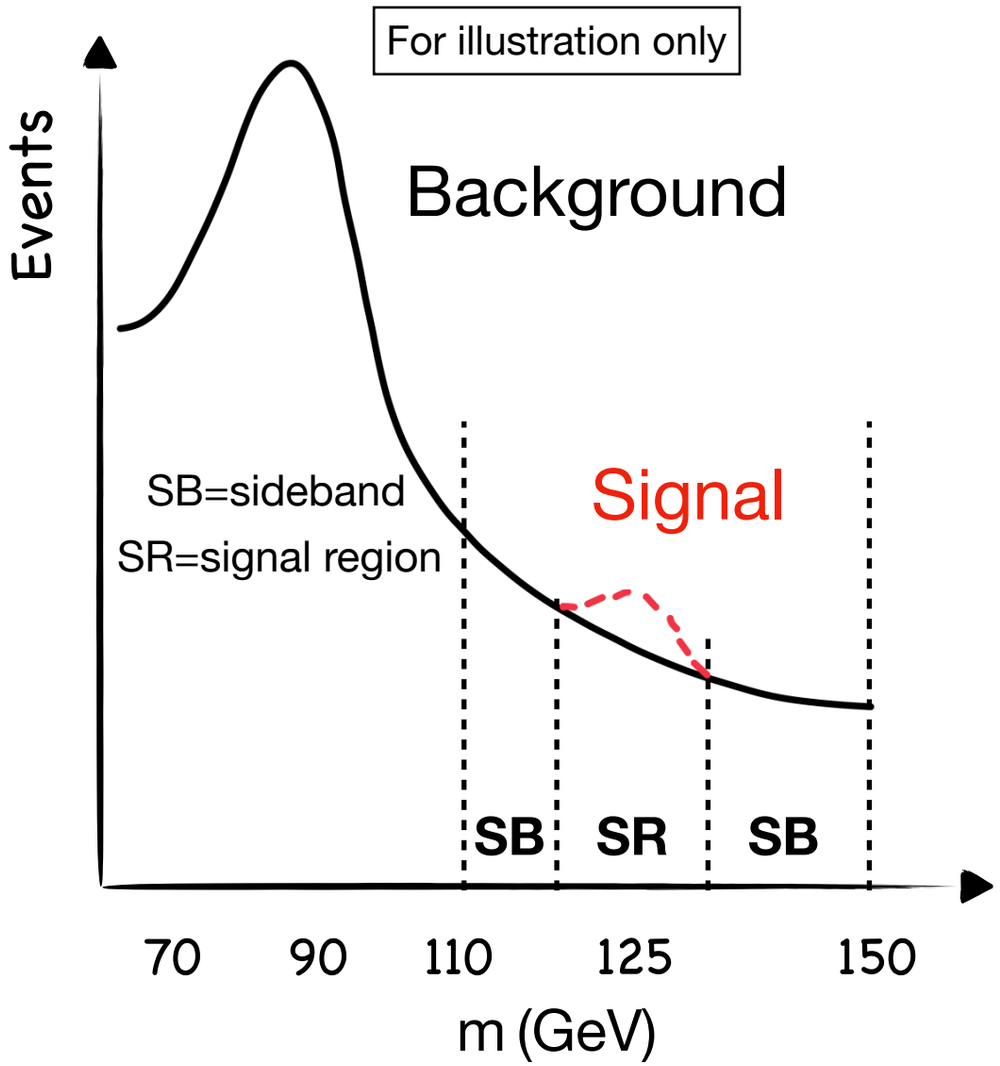
What to expect from this talk?

- Finding Fairness definition suitable for particle physics
- Proof-of-concept for applying, testing and comparing fairness methods
- Conclusions from a case study

Jari Egbers
Fairness in event classification of dimuon Higgs decays
Supervision:
W. Verkerke, T. Van Erven, O. Rieger

Karel de Vries
An Unexpected Application of Fairness to Higgs Boson Detection
Supervision:
W. Verkerke, T. Van Erven, O. Rieger

A bump hunt in particle physics

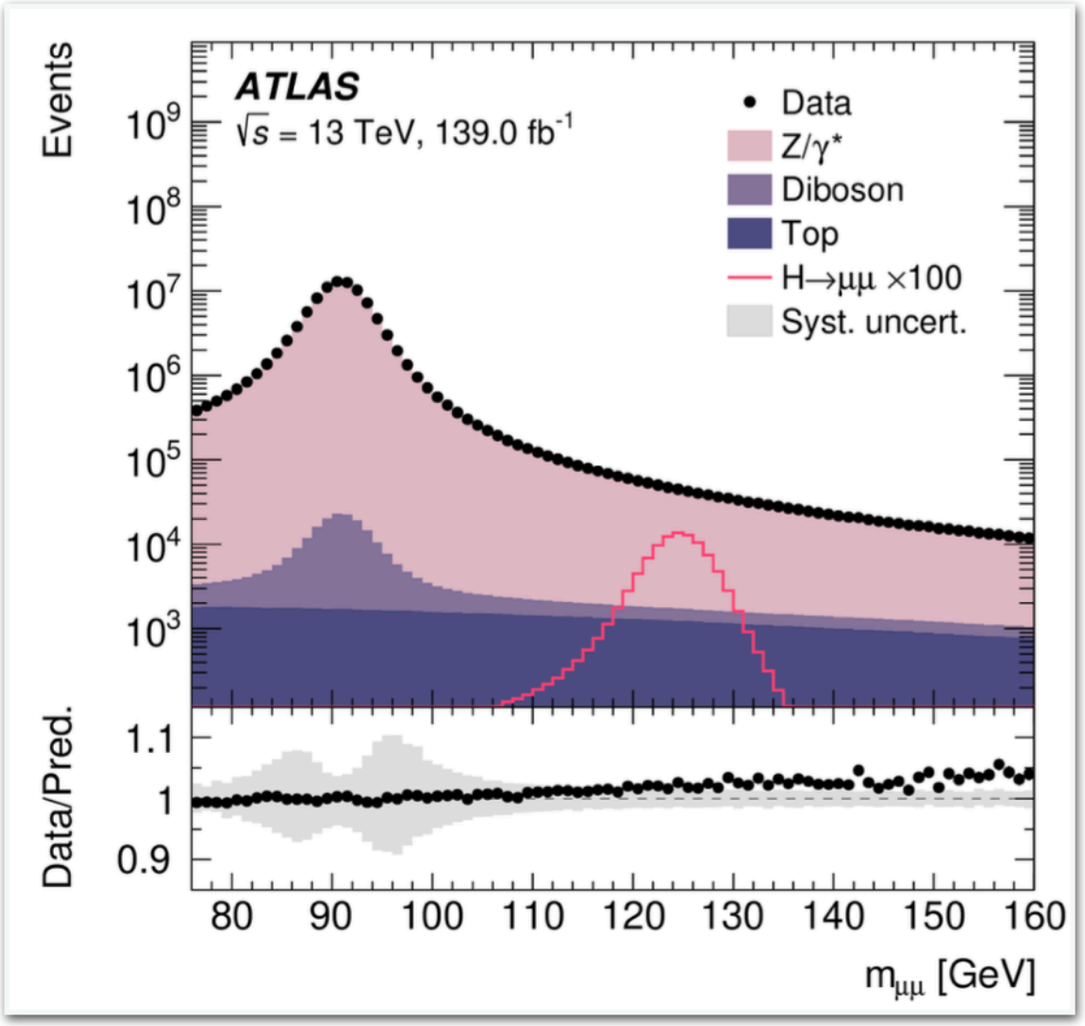


Characteristics of a bump hunt

- Search for narrow resonance at m_X on top of the falling invariant mass spectrum
- Signal extracted by sideband fit
  - Analytical function are used to model signal and background contribution
- Data-driven background estimation in signal region from signal-depleted sidebands

Nikhef

A bump hunt in particle physics



H→μμ as case study

- Precise muon momentum resolution
- Signal model well defined - position and size of peak predicted
- Backgrounds dominated by single process, which is relatively well understood

Challenges

- Low signal-to-background ratio
- Only small kinematic differences between signal and background

Nikhef

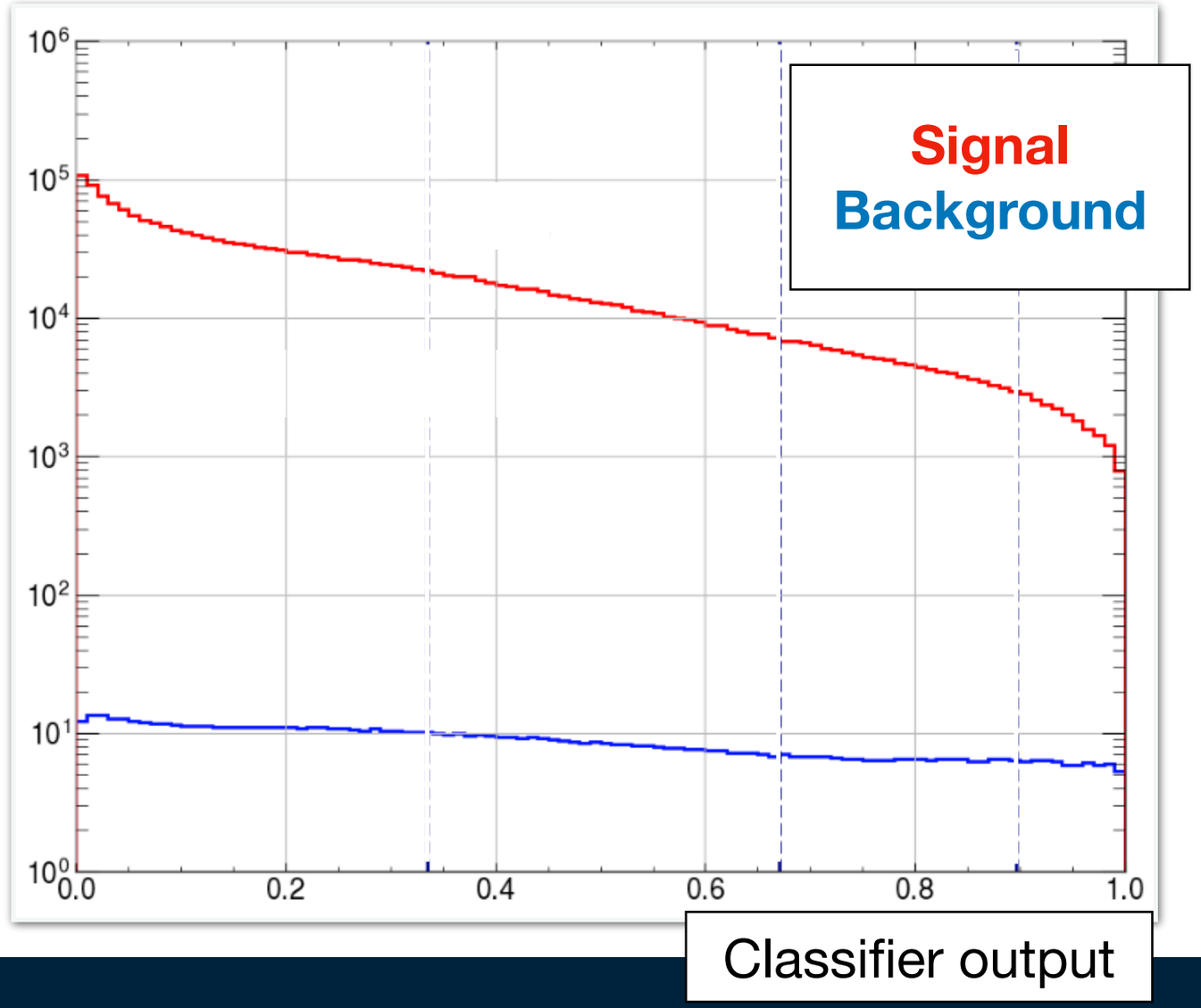
Optimising sensitivity of a bump hunt

Divide & Fit strategy

STEP 1: Optimization of S/B separation using **ML**

STEP 2: Subcategories based on classifier score

STEP 3: Simultaneous fit to $m_{\mu\mu}$ in subcategories



- Train ML model - Boosted Decision Tree (XGBoost)
- Input features - kinematic observables that distinguish signal and background

Nikhef

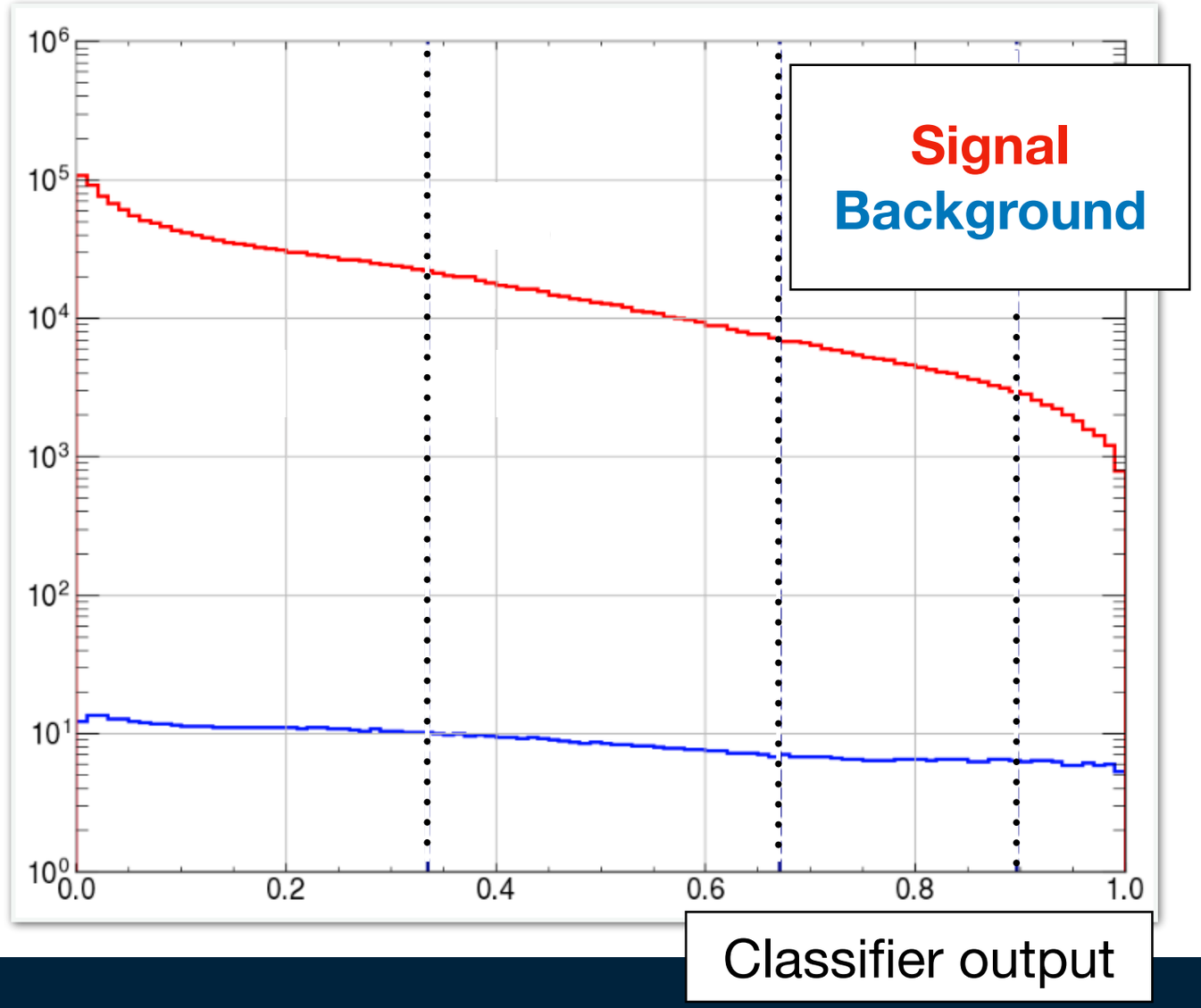
Optimising sensitivity of a bump hunt

Divide & Fit strategy

STEP 1: Optimization of S/B separation using **ML**

STEP 2: Subcategories based on classifier score

STEP 3: Simultaneous fit to $m_{\mu\mu}$ in subcategories



- Train ML model - Boosted Decision Tree (XGBoost)
- Input features - kinematic observables that distinguish signal and background
- Perform iterative algorithm for optimal boundaries that increase the signal-to-background ratio

Nikhef

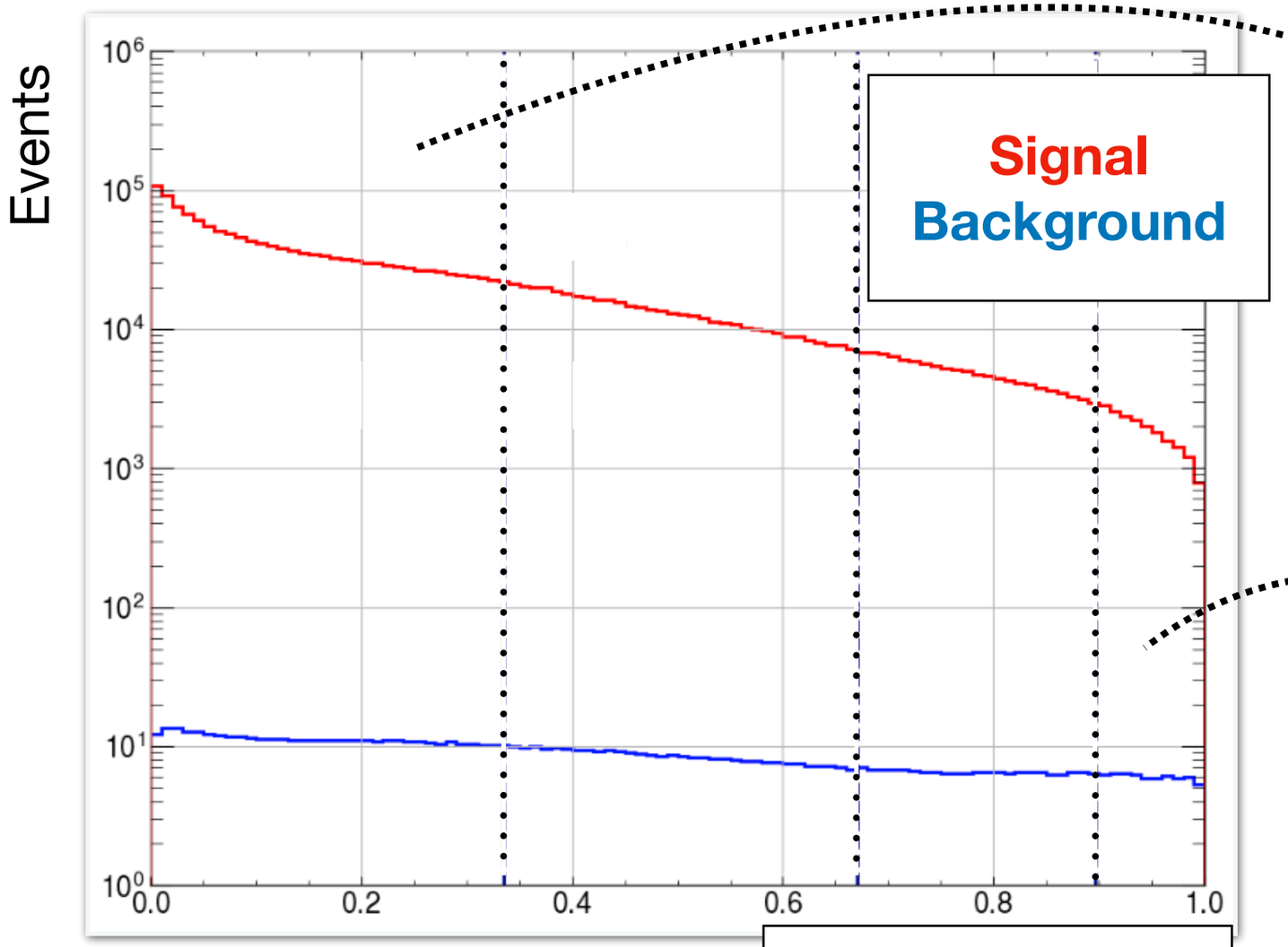
Optimising sensitivity of a bump hunt

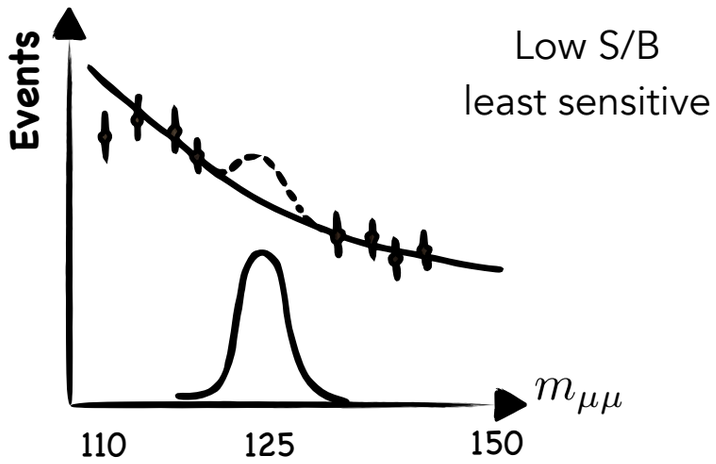
Divide & Fit strategy

STEP 1: Optimization of S/B separation using **ML**

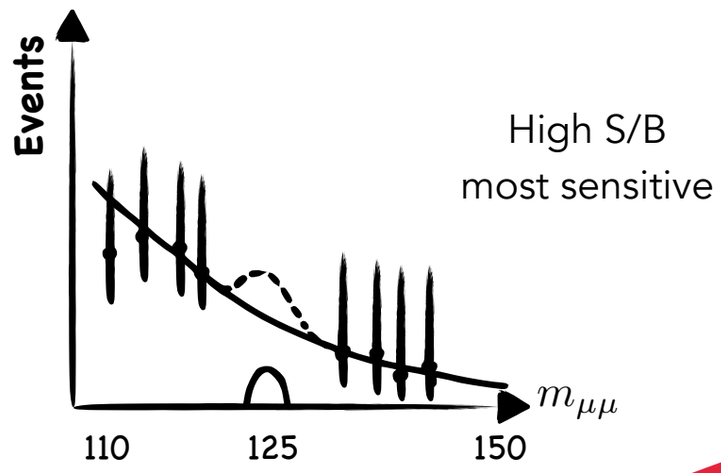
STEP 2: Subcategories based on classifier score

STEP 3: Simultaneous fit to $m_{\mu\mu}$ in subcategories





Low S/B
least sensitive



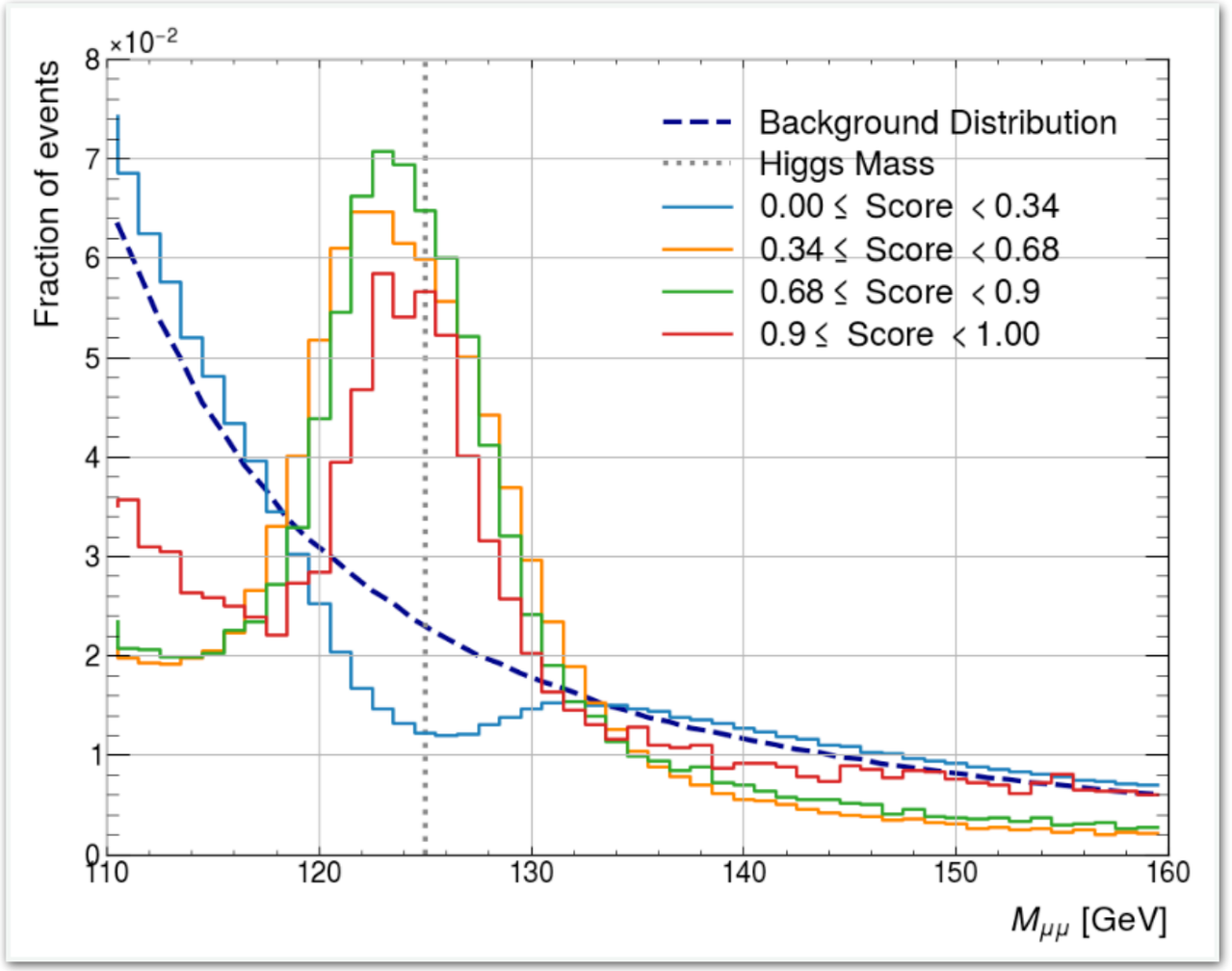
High S/B
most sensitive

Spurious signal due to mass sculpting

Classifier induced mass sculpting

- Shape of mass distribution shows peaks/dips for subcategories if classifier is not uncorrelated with mass
- *Example:*
  - Classifier trained using input features correlated with mass
  - Background-only sample shows peak

➡ Induces spurious signal!

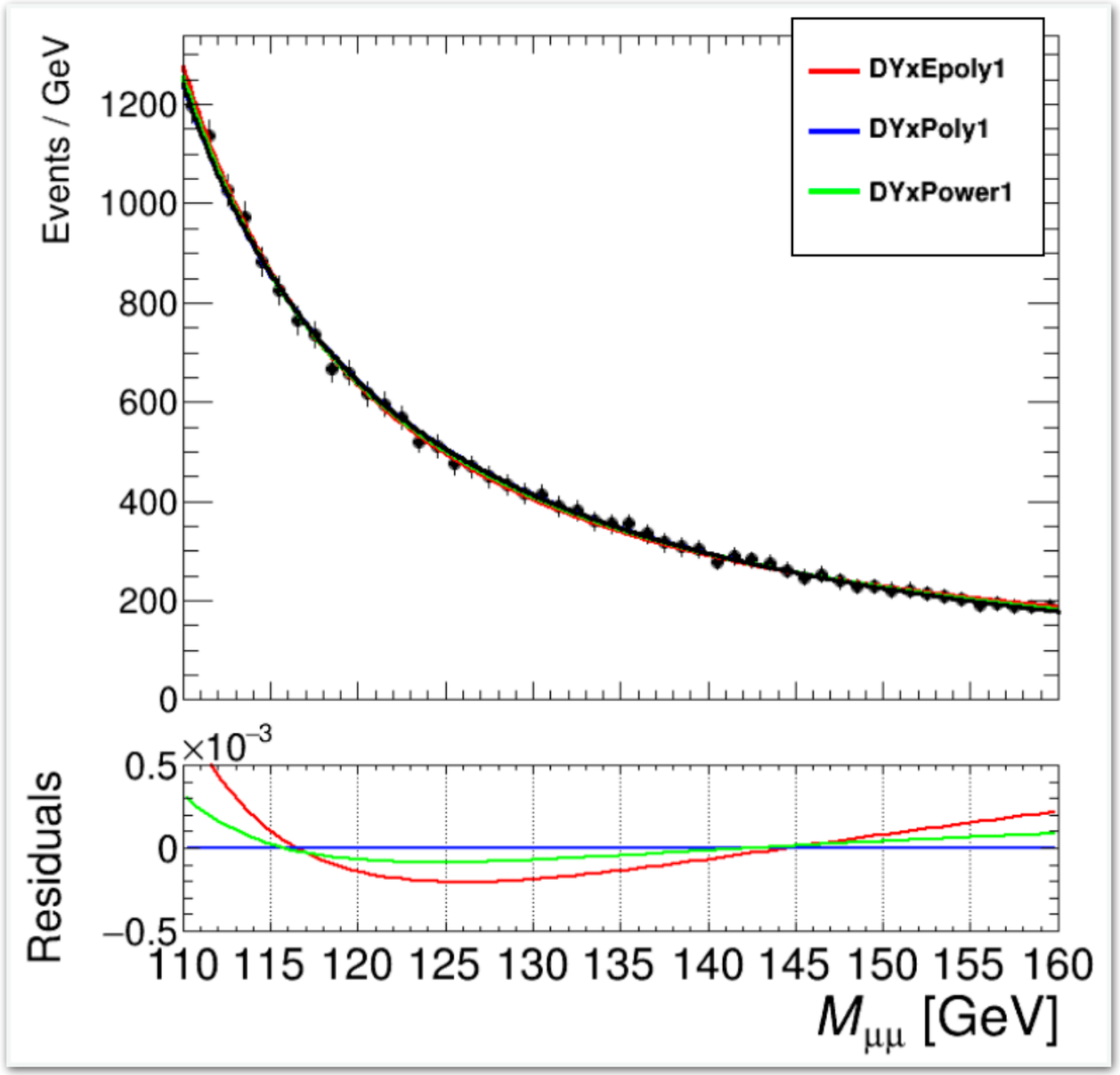


Spurious signal due to bkg function

Choice of background model can induce spurious signal

- Shape of background distribution unknown - can differ for ML-based subcategories
- *Example:*
  - Different models for background allow for larger/smaller signal contribution

➡ Induces spurious signal!



Nikhef

Fairness in Particle Physics

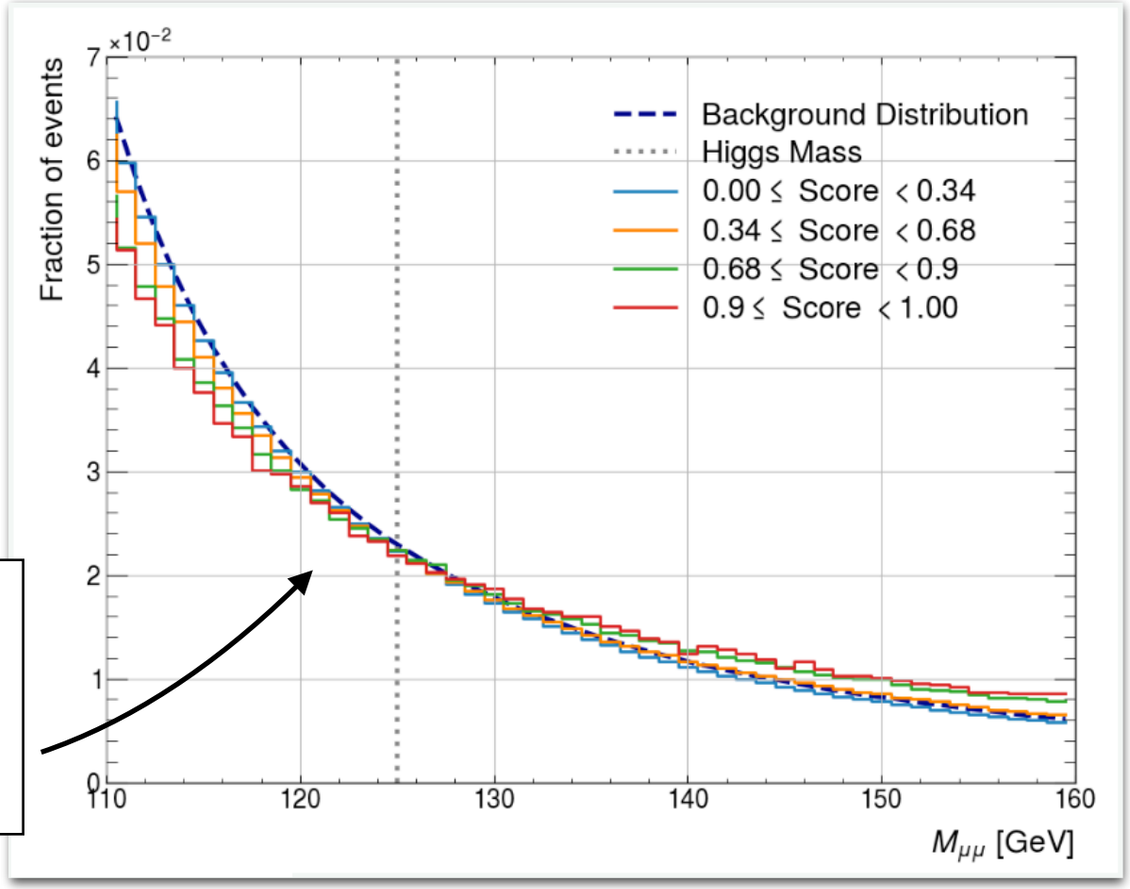
Equal Opportunity definition

A classifier $\hat{Y}$ satisfies equal opportunity with respect to a protected attribute A and the outcome $Y = 0$, if:

$$P(\hat{Y} = 0 | A = a, Y = 0) = P(\hat{Y} = 0 | A = b, Y = 0), \quad \forall a, b \in \mathcal{A}.$$

Generalise this for intervals of the classifier output

Interpretation:
Mass distribution in any interval of classifier output is (approximately) the same



Enforcing fairness on a classifier

The “classical” approach

- Input features largely uncorrelated with the protected attribute
- Training classifier on a rather flat phase-space of the protected attribute

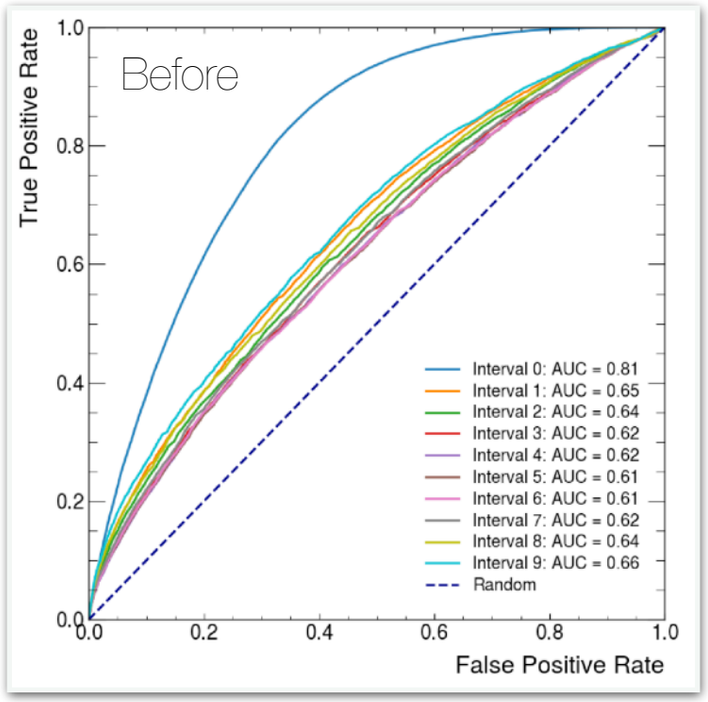
Examples:
narrow mass window, reweighting mass distribution, different signal samples

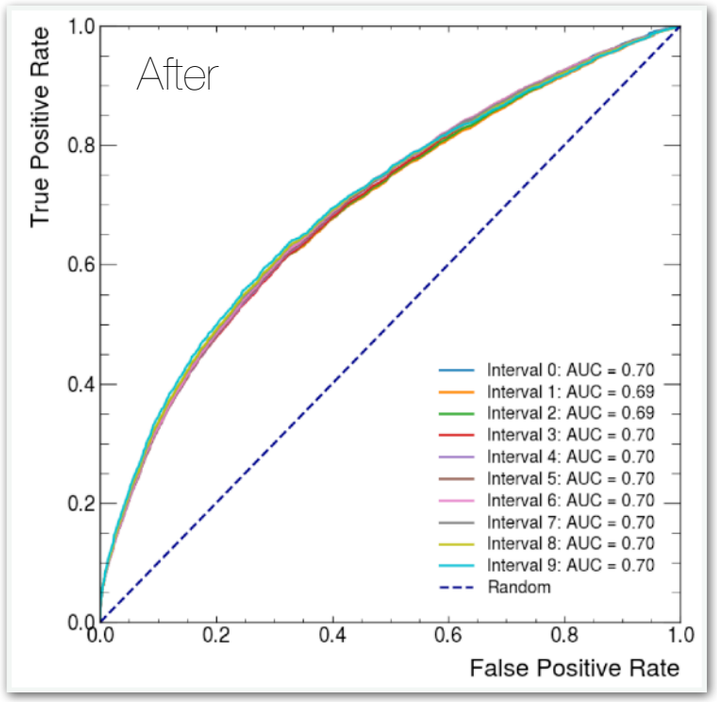
Nikhef

Enforcing fairness on a classifier

ROC Split (during training of ML model)

- Iterative algorithm trains classifier to satisfy EOP
  - Divide Mass into bins and determine AUC
  - Sample events from with $p_i = 2(1\text{-}AUC_i)$
  - Train model on new sample and repeat





Post integration methods

- Train classifier R with mass as input
- Integrate out mass

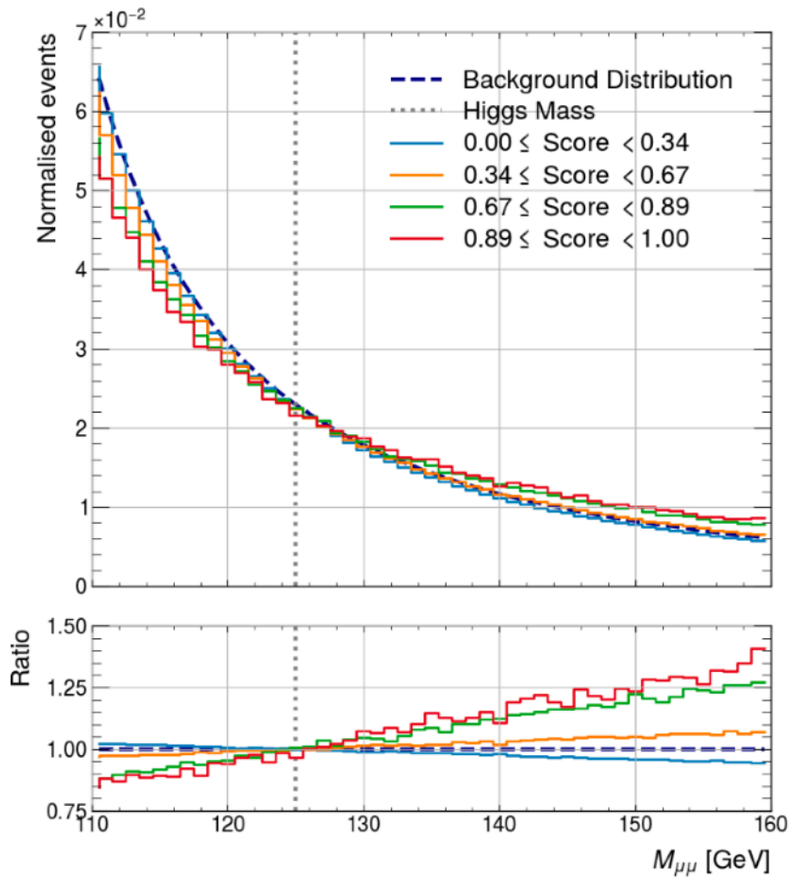
$$R_P(x) = \int R(M_{\mu\mu}, x)\ P(M_{\mu\mu})\ dM_{\mu\mu}$$

- *Effective, but strong impact on performance*
- *Helpful, if the correlation is small from the beginning*
- ***See Purvasha's talk for more sophisticated methods!***

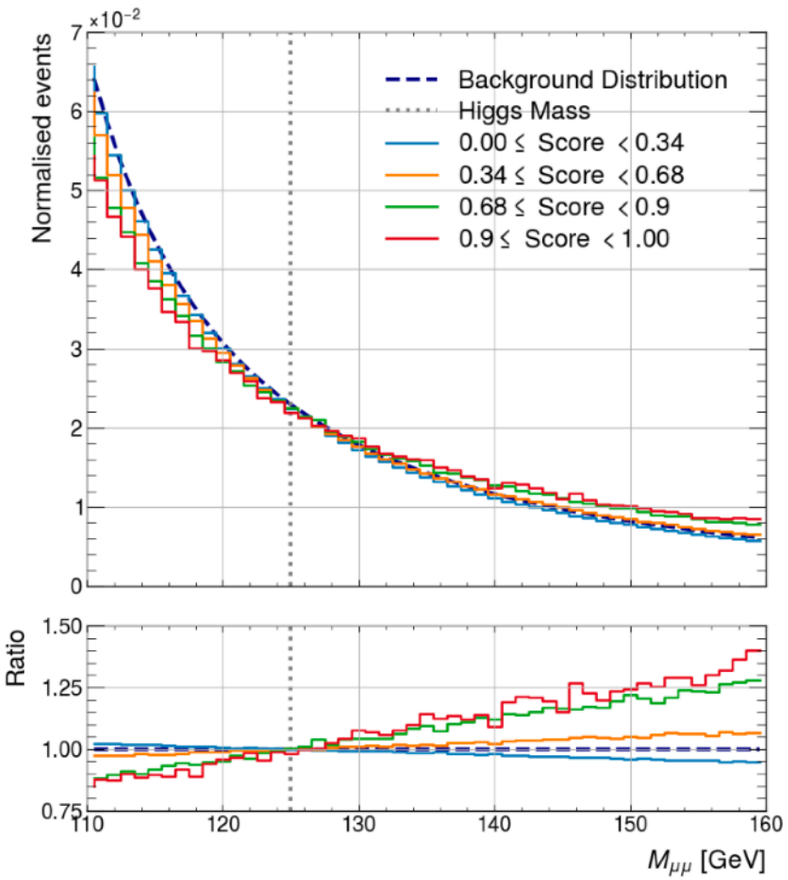
Nikhef

Evaluation of fairness

Results for a classical approach



Results for ROC Split



Nikhef

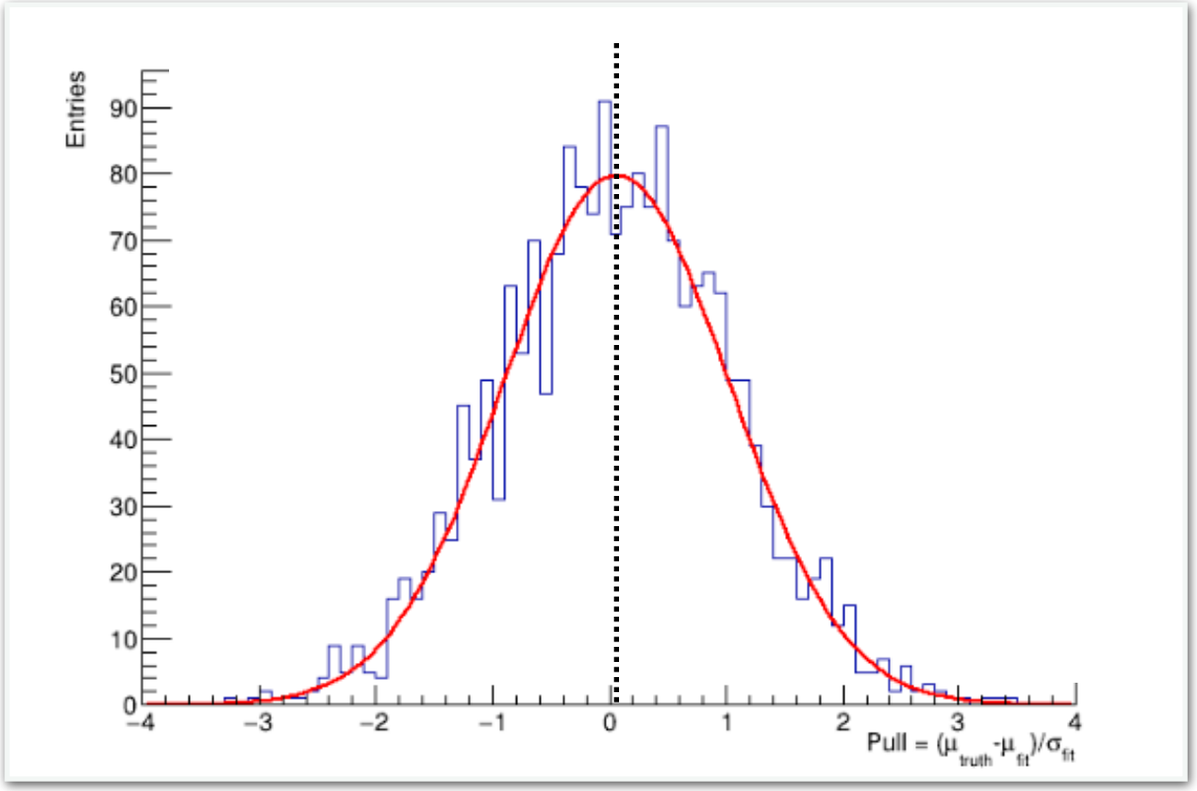
Evaluation of fairness

Toy-based spurious signal study

- To evaluate the impact of:
  1) Mass sculpting
  2) Unknown background distribution
- 2000 toys for (S+B, μ=1) or (B-only, μ=0)
- Pull distribution

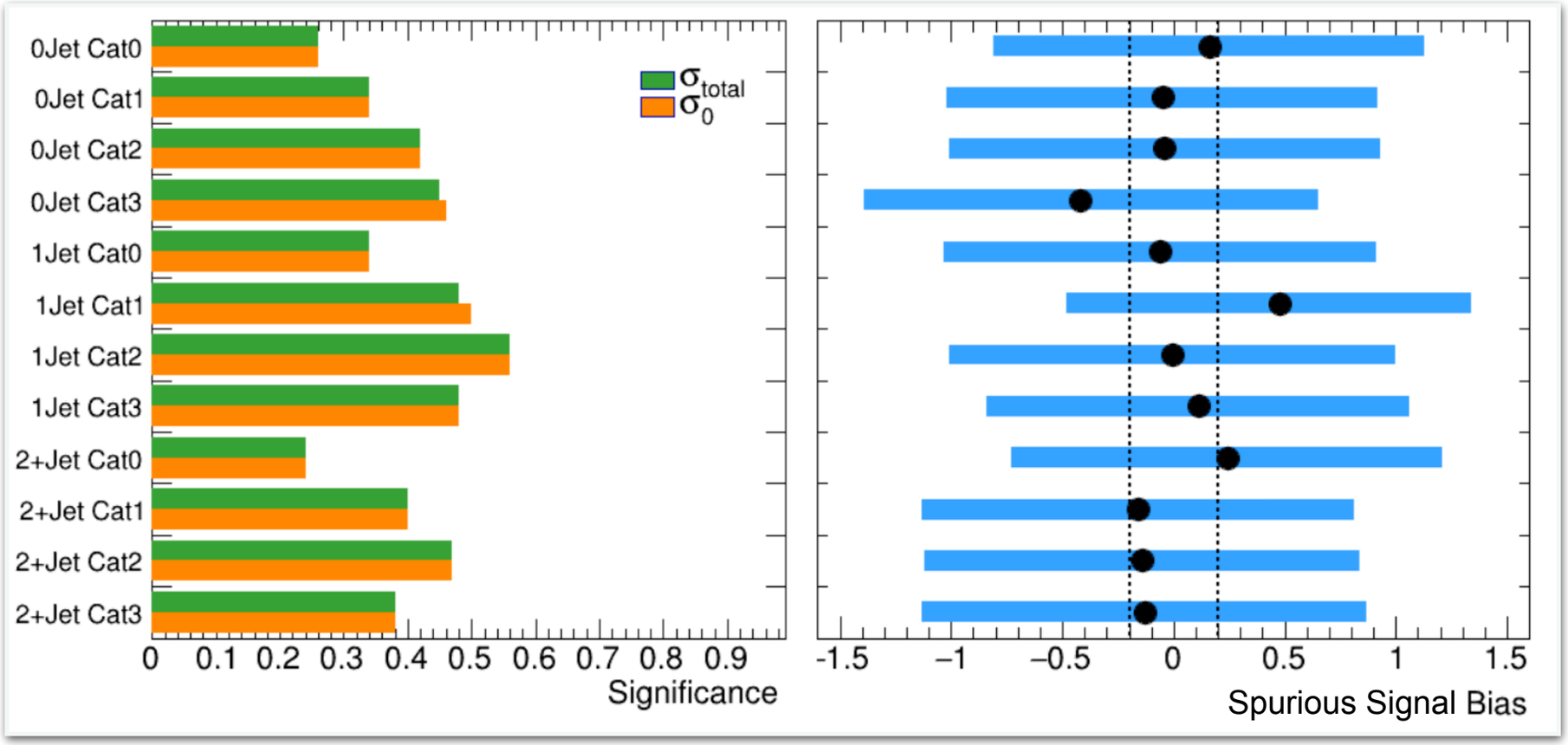
$$P = \frac{\mu_{toy} - \mu_{fit}}{\sigma_{fit}}$$

- Mean of pull indicates spurious signal
- *Note:* Impact of mass sculpting and choice of bkg function can not be disentangled



Nikhef

Case study results



- Calculate expected significance and spurious signal bias for each ML-based category
- The combined results of the analysis significance are comparable between the classical approach and analyses optimised using fairness methods
- In this case study, statistical uncertainties are larger than systematic uncertainty
- Impact potentially larger for cases when syst. unc. < stat. unc.

Nikhef

Take away messages..

- Fairness methods can be used to event classification problems in particle physics
- Effective methods to maintain fairness while preserving classifier performance
- Fairness can replace “classical” decorrelation methods

Outlook..

- Decorrelation/fairness techniques might become more important in the HL-LHC era when systematic uncertainties become dominant