UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

KIRCHHOFF-INSTITUT FÜR PHYSIK

ATLAS EXPERIMENT

“Data-Driven” BG Estimate

Tobias Heintz, 06/11/2025

Overview

BIB data: CR Triggers
with failed BIB removal

Main data: "normal"
CR Triggers

	Low-E_{T} selection:	Main data	BIB
Preselection:	trigger, ~~2~~ 1 clean jet~~s~~	40 743 867	2 200 854
	$\sum \Delta R_{\mathrm{min}} > 0.5$	28 248 024	1 399 351
Event cleaning:	~~low-E_{T} BDT > 0.05~~	1 288 596	44 035
	Trigger matching	1 138 961	36 266
	$-3 < t < 15\mathrm{ns}$	1 123 239	35 245
	$\log_{10}(E_{\mathrm{H}}/E_{\mathrm{EM}}) > -1.5$ for $\mathrm{jet}^{\mathrm{sig}_1}$, $\cancel{\mathrm{jet}^{\mathrm{sig}_2}}$, $\mathrm{jet}^{\mathrm{bib}_1}$, $\cancel{\mathrm{jet}^{\mathrm{bib}_2}}$	1 038 019	33 100
	$\lvert\eta\rvert \notin [1.45, 1.55]$ for $\mathrm{jet}^{\mathrm{sig}_1}$, $\cancel{\mathrm{jet}^{\mathrm{sig}_2}}$	976 805	31 292
Low-E_{T} selection:	$\cancel{H_{\mathrm{T}}^{\mathrm{miss}}/H_{\mathrm{T}} < 0.6}$	965 748	30 712
	$\cancel{p_{\mathrm{T}}(\mathrm{jet}^{\mathrm{sig}_{1l}}) > 80\,\mathrm{GeV}}$	315 530	10 048
	$\cancel{p_{\mathrm{T}}(\mathrm{jet}^{\mathrm{sig}_{2l}}) > 80\,\mathrm{GeV}}$	73 484	2810
	$\cancel{\sum_{\mathrm{jet}^{\mathrm{sig}_{1l}},\mathrm{jet}^{\mathrm{sig}_{2l}}}} \log_{10}(E_{\mathrm{H}}/E_{\mathrm{EM}}) \cancel{> 2}$ >1, for jet(sig1)	3375	93
	Low-E_{T} NN ~~product > 0.7~~ > 0.95	307	10
Region A:		23	0
Region B:		3	0
Region C:		220	7
Region D:		61	3

Is the low-ET selection what we actually want?

Skip ABCD stuff?
[Our signal yield will be likely $\lesssim 100$]

Scale	B2F	BIB
CR x 5 / 108.33	1880530	101580
eff_{CR}	1303786	64587
--	------------	------------
eff_{CR}	1152387	53192
SQRT[eff_{CR}]	1144406	52438
SQRT[eff_{CR}]	1100136	50817
SQRT[eff_{CR}]	1067205	49410
--	------------	------------
--	------------	------------
--	------------	------------
SQRT[eff_{CR}]	228712	8989
0.0072 (SM mJ), 0.0116 (BIB), 0.7531 (Signal)	**1647**	**104**

See details in the spreadsheet

1. Pre-Selection

	Low-E_{T} selection:	Main data	BIB
Preselection:	trigger, ~~2~~1clean jet~~s~~	40 743 867	2 200 854
	$\sum \Delta R_{\mathrm{min}} > 0.5$	28 248 024	1 399 351
Event cleaning:	~~low-E_{T} BDT > 0.05~~	1 288 596	44 035
	Trigger matching	1 138 961	36 266
	$-3 < t < 15$ns	1 123 239	35 245
	$\log_{10}(E_{\mathrm{H}}/E_{\mathrm{EM}}) > -1.5$ for $\mathrm{jet}^{\mathrm{sig}_1}$, $\cancel{\mathrm{jet}^{\mathrm{sig}_2}}$, $\mathrm{jet}^{\mathrm{bib}_1}$, $\cancel{\mathrm{jet}^{\mathrm{bib}_2}}$	1 038 019	33 100
	$\lvert\eta\rvert \notin [1.45, 1.55]$ for $\mathrm{jet}^{\mathrm{sig}_1}$, $\cancel{\mathrm{jet}^{\mathrm{sig}_2}}$	976 805	31 292
Low-E_{T} selection:	$\cancel{H_{\mathrm{T}}^{\mathrm{miss}}/H_{\mathrm{T}} < 0.6}$	965 748	30 712
	$\cancel{p_{\mathrm{T}}(\mathrm{jet}^{\mathrm{sig}_{1l}}) > 80\,\mathrm{GeV}}$	315 530	10 048
	$\cancel{p_{\mathrm{T}}(\mathrm{jet}^{\mathrm{sig}_{2l}}) > 80\,\mathrm{GeV}}$	73 484	2810
	$\cancel{\sum_{\mathrm{jet}^{\mathrm{sig}_{1l}},\mathrm{jet}^{\mathrm{sig}_{2l}}}} \log_{10}(E_{\mathrm{H}}/E_{\mathrm{EM}}) \cancel{> 2}$ >1, for jet(sig1)	3375	93
	Low-E_{T} NN ~~product > 0.7~~ > 0.95	307	10
~~**Region A:**~~		23	0
~~**Region B:**~~		3	0
~~**Region C:**~~		220	7
~~**Region D:**~~		61	3

- CR Trigger (low ET or high ET)
  - B2F = CR / 108 due to jet-MET correlation
  - What about their high ET trigger?
- 1 clean jet with pT > 40 GeV
  - 1 clean jet = 2 clean jets x 5, since
    $$\frac{\sigma(p_T^{j_1} > 40)}{\sigma(p_T^{j_1} > 40 \wedge p_T^{j_2} > 40)} \sim 5$$
  - Can we confirm this factor in data?
- ΔR_{min} summed over all clean jets > 0.5
  - Although having only one jet in our selection, I think we can leave this cut as is

2. Event Cleaning

	Low-E_T selection:	Main data	BIB
Preselection:	trigger, ~~2~~1 clean ~~jets~~	40 743 867	2 200 854
	$\sum \Delta R_{\min} > 0.5$	28 248 024	1 399 351
Event cleaning:	~~low-E_T BDT > 0.05~~	1 288 596	44 035
	Trigger matching	1 138 961	36 266
	$-3 < t < 15$ns	1 123 239	35 245
	$\log_{10}(E_H/E_{EM}) > -1.5$ for $\text{jet}^{\text{sig}_1}$, $\cancel{\text{jet}^{\text{sig}_2}}$, $\text{jet}^{\text{bib}_1}$, $\cancel{\text{jet}^{\text{bib}_2}}$	1 038 019	33 100
	$\lvert\eta\rvert \notin [1.45, 1.55]$ for $\text{jet}^{\text{sig}_1}$, $\cancel{\text{jet}^{\text{sig}_2}}$	976 805	31 292
Low-E_T selection:	~~$H_T^{\text{miss}}/H_T < 0.6$~~	965 748	30 712
	~~$p_T(\text{jet}^{\text{sig}_{1l}}) > 80$ GeV~~	315 530	10 048
	~~$p_T(\text{jet}^{\text{sig}_{2l}}) > 80$ GeV~~	73 484	2810
	$\cancel{\sum_{\text{jet}^{\text{sig}_{1l}},\text{jet}^{\text{sig}_{2l}}}} \log_{10}(E_H/E_{EM}) \cancel{> 2}$ >1, for jet(sig1)	3375	93
	Low-E_T NN ~~product > 0.7~~ > 0.95	307	10
~~**Region A:**~~		23	0
~~**Region B:**~~		3	0
~~**Region C:**~~		220	7
~~**Region D:**~~		61	3

- No BDT cut
- Trigger Matching
  - doesn't depend on 1/2 jets
- Jet timing
- Individual EMF cut on the signal and BIB jet candidates
- Reject the transition region
  - CR applies these cuts on two jets, we only on one jet
  - Scale eff.'s by SQRT

3. Low ET Selection

	Low-E_T selection:	Main data	BIB
Preselection:	trigger, ~~2~~1 clean jet~~s~~	40 743 867	2 200 854
	$\sum \Delta R_{\mathrm{min}} > 0.5$	28 248 024	1 399 351
Event cleaning:	~~low-E_T BDT > 0.05~~	1 288 596	44 035
	Trigger matching	1 138 961	36 266
	$-3 < t < 15$ns	1 123 239	35 245
	$\log_{10}(E_H/E_{EM}) > -1.5$ for $\mathrm{jet}^{\mathrm{sig}_1}$, ~~$\mathrm{jet}^{\mathrm{sig}_2}$~~, $\mathrm{jet}^{\mathrm{bib}_1}$, ~~$\mathrm{jet}^{\mathrm{bib}_2}$~~	1 038 019	33 100
	$\lvert\eta\rvert \notin [1.45, 1.55]$ for $\mathrm{jet}^{\mathrm{sig}_1}$, ~~$\mathrm{jet}^{\mathrm{sig}_2}$~~	976 805	31 292
Low-E_T selection:	~~$H_T^{\mathrm{miss}}/H_T < 0.6$~~	965 748	30 712
	~~$p_T(\mathrm{jet}^{\mathrm{sig}_1 l}) > 80\,\mathrm{GeV}$~~	315 530	10 048
	~~$p_T(\mathrm{jet}^{\mathrm{sig}_2 l}) > 80\,\mathrm{GeV}$~~	73 484	2810
	$\cancel{\sum_{\mathrm{jet}^{\mathrm{sig}_1 l}, \mathrm{jet}^{\mathrm{sig}_2 l}}} \log_{10}(E_H/E_{EM}) \cancel{> 2}$ >1, for jet(sig1)	3375	93
	Low-E_T NN ~~product > 0.7~~ > 0.95	307	10
~~**Region A:**~~		23	0
~~**Region B:**~~		3	0
~~**Region C:**~~		220	7
~~**Region D:**~~		61	3

- No cuts on HTmiss, pT of jet(s)
- Summed EMF cut
  - Quite tricky to disentangle, but if we assume that the main contribution comes from two equally displaced jets, then we potentially could use

$$\mathscr{P}\left(\sum_{j=1}^{2} \log_{10}\left(\frac{E_{\mathrm{had}}}{E_{\mathrm{EM}}}\right)_j > 2\right) \approx \mathscr{P}\left(\log_{10}\left(\frac{E_{\mathrm{had}}}{E_{\mathrm{EM}}}\right)_j > 1\right)^2$$

  - In this case, our BG would scale with SQRT of CR
- NN cut
  - In CR the NN score is used as input for the BDT, and a sharp cut is applied on the NN score product of the two jets
  - For B2F, we cut at NN score > 0.95, which reduces Background by
    - SM multijets:
      - S = 0.7531
      - B = 0.0072
      - S/B = 105
      - S/sqrt[B] = 8.9
    - BIB:
      - S = 0.7531
      - B = 0.0116
      - S/B = 65
      - S/sqrt[B] = 7.0

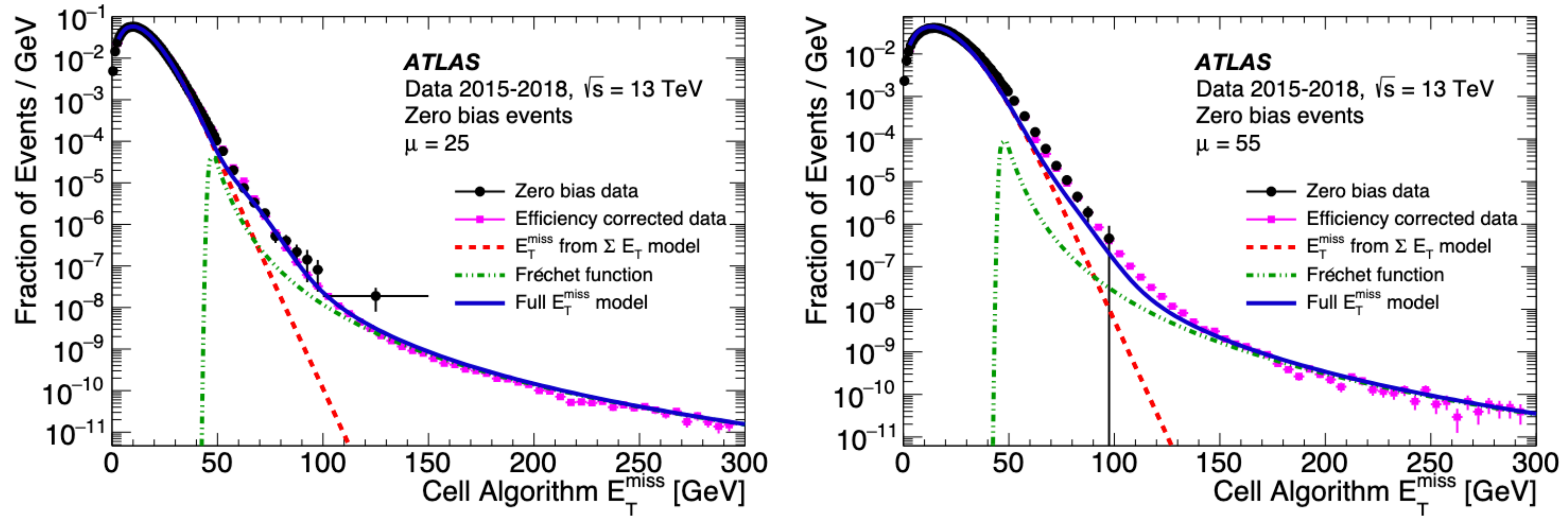
Backup

L1 Background Suppression

$$\frac{B_{\mathrm{B2F,R2}}}{B_{\mathrm{CR,R2}}} = \frac{1}{\pi} \times \frac{\int_{40}^{100} \mathrm{MET}}{\int_{0}^{100} \mathrm{MET}} \approx \frac{1}{\pi} \times 0.029 = 0.0094 \quad ✓$$

MET must be correlated in φ

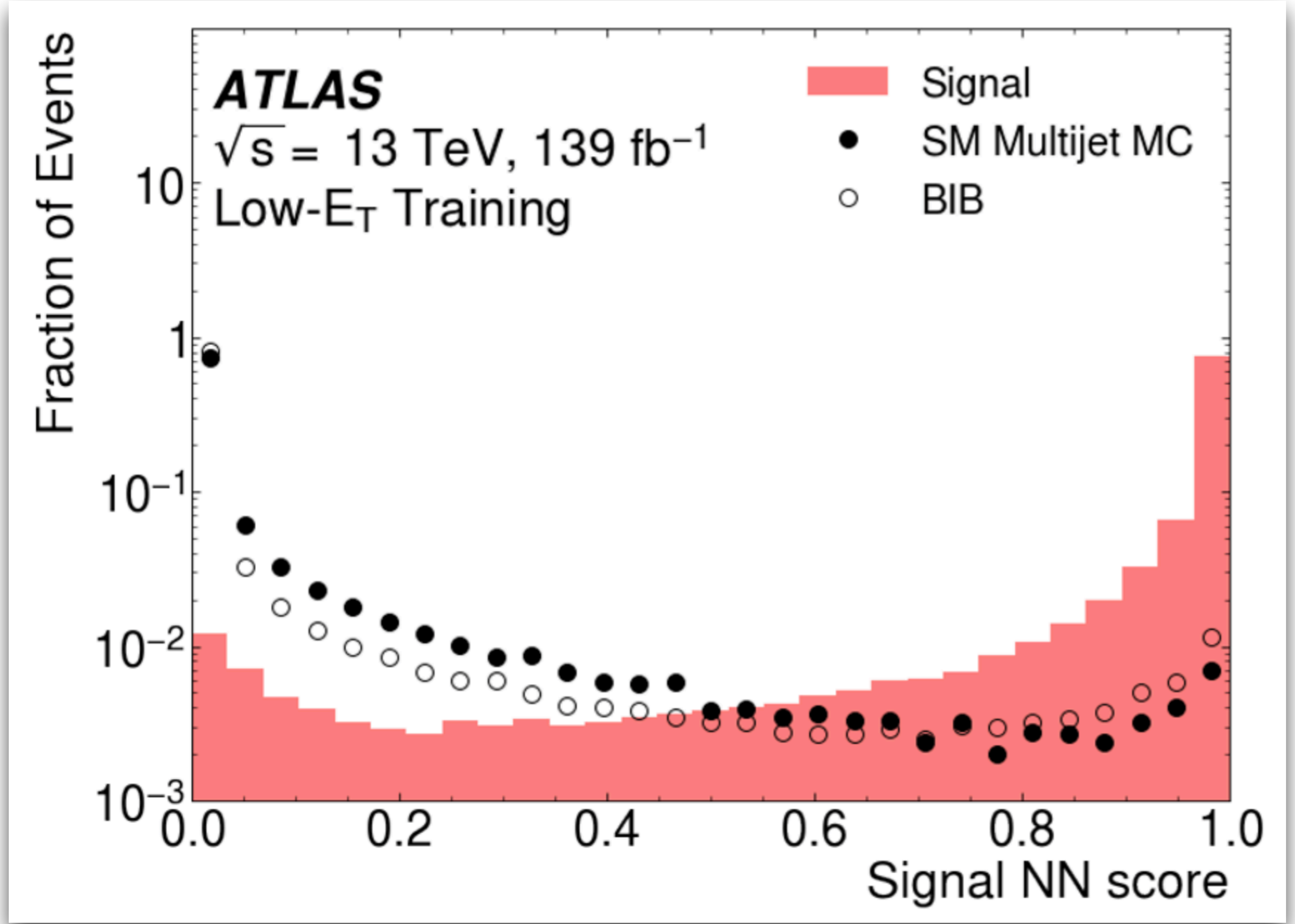
MET must exceed 40 GeV
→ We can look up Zerobias MET distributions in https://arxiv.org/pdf/2005.09554 and https://www.hepdata.net/record/ins1796953 (N.B. only for Run-2, i.e. PU ~ 55, and at HLT level)



(a) (b)

Figure 1: A comparison of the measured `cell` E_T^{miss} distribution with that predicted by the two-component model for two pile-up scenarios compared with data. The circular points show the data collected using zero bias triggers, but have insufficient luminosity to probe the higher E_T^{miss} portion of the distribution. The square points extend the measured distribution using L1 $E_T^{\mathrm{miss}} > 30$ GeV and L1 $E_T^{\mathrm{miss}} > 50$ GeV data. The uncertainties for the data points are statistical only, and much larger for the zero bias data due to the limited luminosity. The dashed (red) curve is the prediction from the calorimeter-resolution part of the model. The dash-dotted (green) curve is the high E_T^{miss} tail's probability distribution for the mean number of pp interactions μ in each figure. The solid (blue) curve is the full model prediction computed by combining the E_T^{miss} from these two individual sources shown in red and green, each calculated for $\mu = \langle\mu\rangle$. The black points show the unbiased E_T^{miss} distribution measured in data. (a) corresponds to a prediction for $\langle\mu\rangle = 25$ while (b) corresponds to $\langle\mu\rangle = 55$.

NN Signal Score



- Use low ET Signal NN score to optimise
$$\frac{S}{B} = \frac{\int_{x_{cut}}^{1} S\,\mathrm{d}x}{\int_{x_{cut}}^{1} MJ\,\mathrm{d}x}$$
- It turns out that just taking the last bin gives the best S/B:
  - S = 0.7531
  - B = 0.0072
  - S/B = 105
  - S/sqrt[B] = 8.9
- For reference: CR makes a (sharp) cut for the Signal Score product (of both jets) at 0.7, and additionally uses the NN signal score as input for the event level BDT
- For BIB: Best S/B with cut at 0.95:
  - S = 0.7531
  - B = 0.0116
  - S/B = 65
  - S/sqrt[B] = 7.0

How To Disentangle 2 jet Probabilities

- $\mathscr{P}_0 := \mathscr{P}\left(f_{j_1}(x) > \text{cut} \wedge f_{j_2}(x) > \text{cut}\right) = \ldots$
- $\ldots = \mathscr{P}\left(f_{j_1}(x) > \text{cut}\right) \times \mathscr{P}\left(f_{j_2}(x) > \text{cut} \,|\, f_{j_1}(x) > \text{cut}\right)$
- $\ldots = \mathscr{P}\left(f_{j_1}(x) > \text{cut}\right) \times \mathscr{P}\left(f_{j_2}(x) > \text{cut}\right)$
- $\ldots = \left[\mathscr{P}\left(f_j(x) > \text{cut}\right)\right]^2$
- $\Rightarrow \mathscr{P}\left(f_j(x) > \text{cut}\right) = \sqrt{\mathscr{P}_0}$

- That is 1 jet probability scales with sqrt of 2 jet probabilities
- More complicated for summed quantise, such as summed dRmin and EMF cuts