Identification-efficiency Scale-Factors

for forward electrons in the ATLAS experiment @ 13 TeV

Lukas Bayer
Hamburg, 25.04.2024

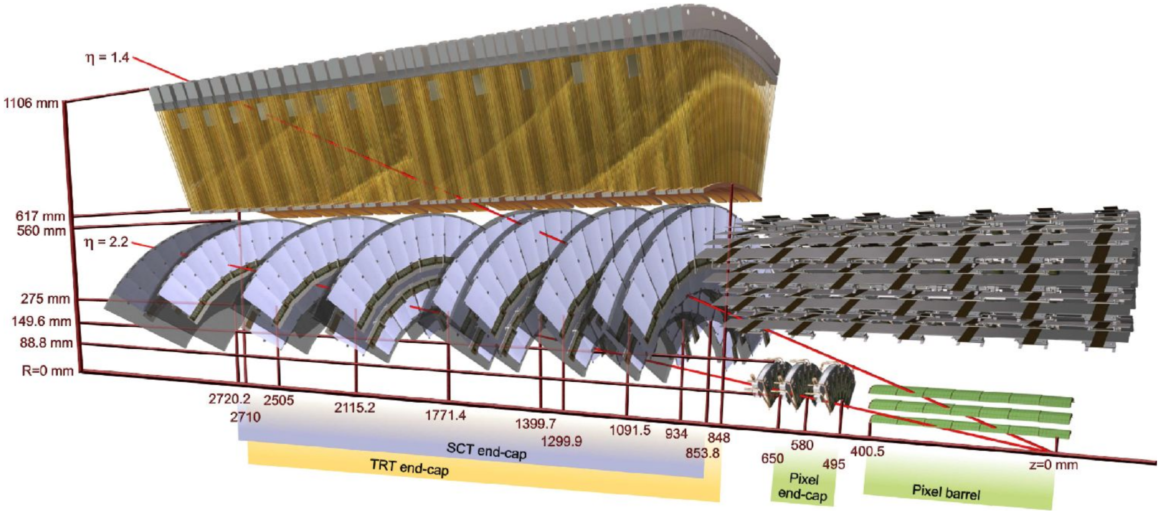
HELMHOLTZ

Geometry

of the ATLAS experiment

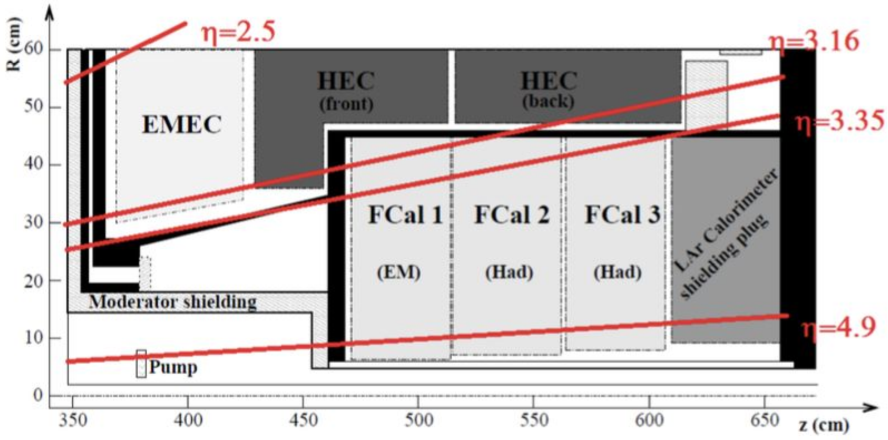
Central Region (|η| < 2.5)

- Silicon pixel and microstrip detectors
- Transition radiation tracker
- Electromagnetic calorimeter



Forward Region (|η| > 2.5)

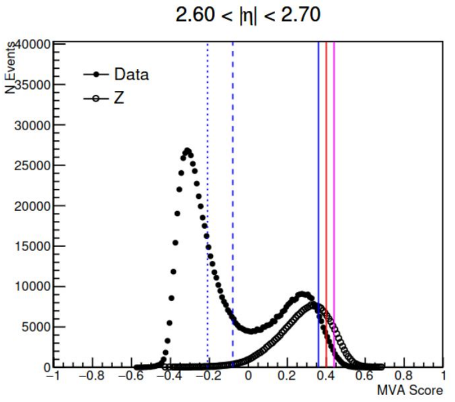
- Electromagnetic end-cap calorimeter (EMEC)
  - 2.5 < |η| < 3.2
- Forward calorimeter (FCAL)
  - 3.2 < |η| < 4.9

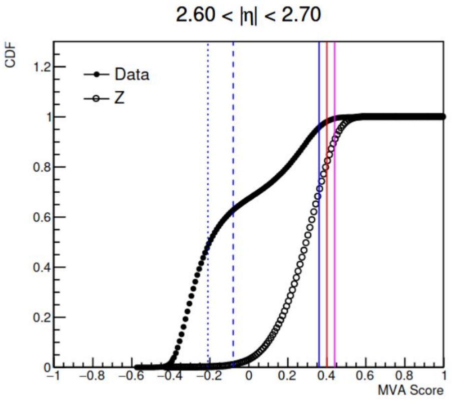


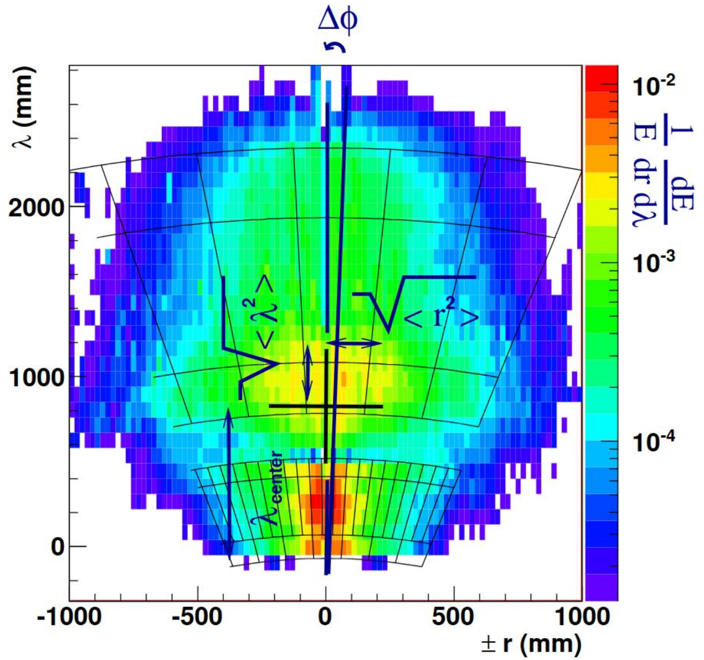
Electron Identification

in the forward region

- No tracking information, only calorimeter
- Rely on shower shape to discriminate electrons from jets etc.
- Calculate electron likelihood via multi-variate analysis
- Choose threshold as identification criterion
  - (very very loose, very loose, loose, medium, tight)





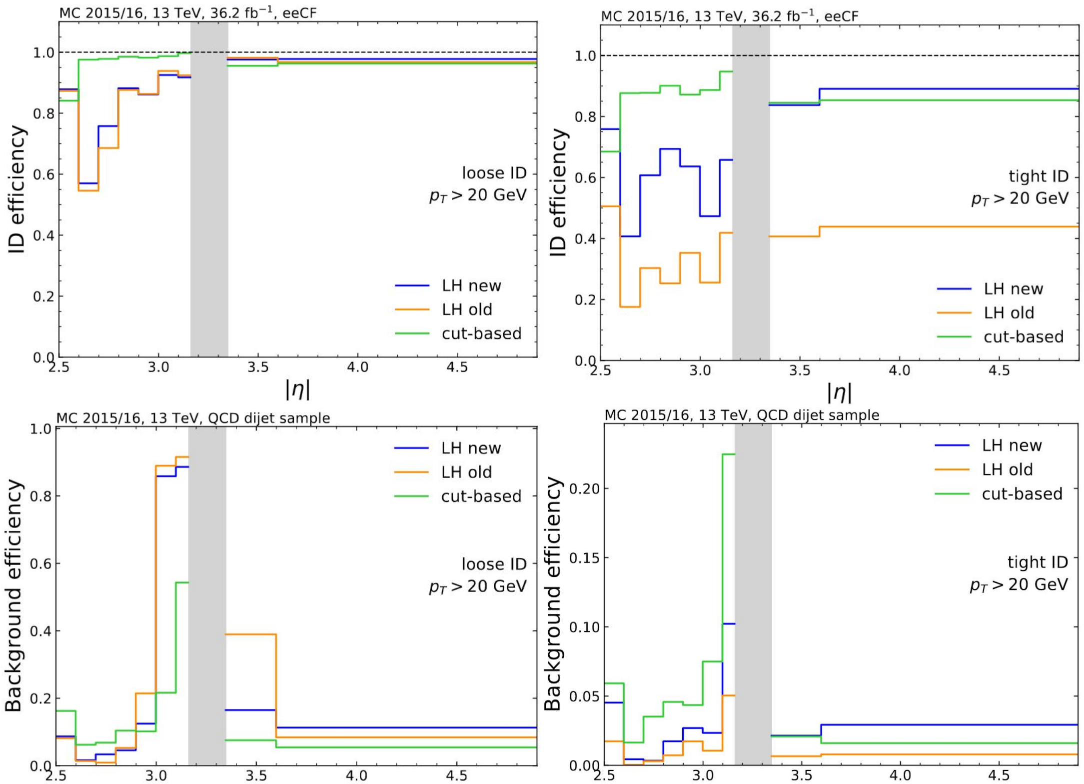


LH Working Point Versions

ID and Background Efficiencies

loose ID

tight ID



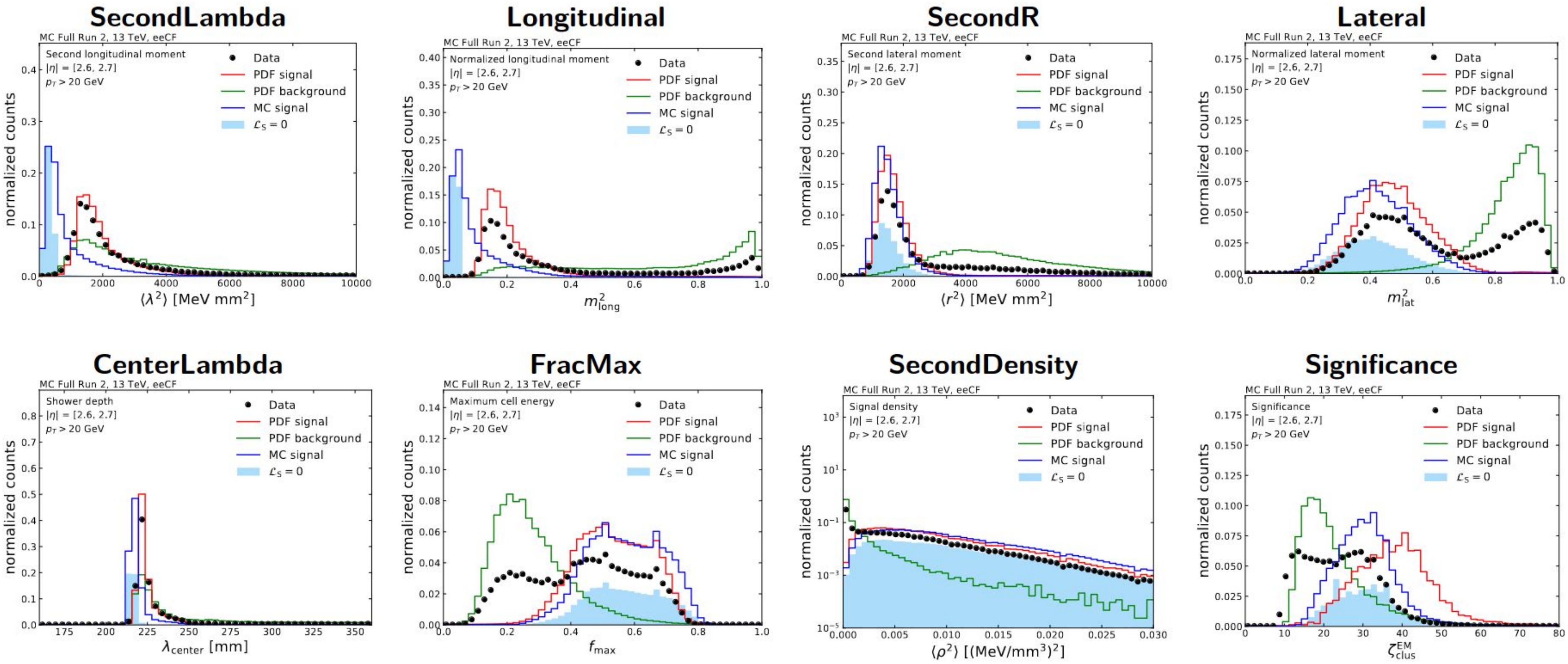
- Versions:
  - LH new: mc16_20190729
    - (used here)
  - LH old: mc16_20180822
    - (current standard)
  - cut-based: mc15_20150812
- tuned with 2015/16 physics_Main STDM4 data and mc16 Monte-Carlo
- ID efficiency estimated using MCTruthPartClassifier
- Background efficiency estimated using QCD dijet sample to select fake forward electrons

https://indico.cern.ch/event/984796/contributions/4180608/attachments/2171408/3666061/210114_fwd_electron_identification_mh.pdf

https://indico.cern.ch/event/999892/contributions/4199454/subcontributions/326621/attachments/2180678/3683535/210129_update_mh.pdf

Shower Shapes

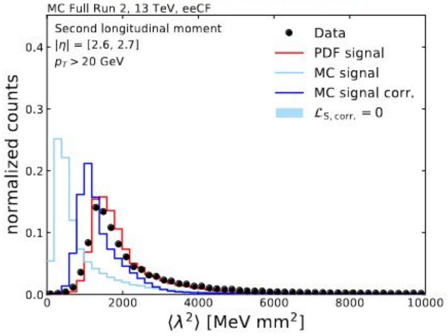
before applying fudge factors (in EMEC)



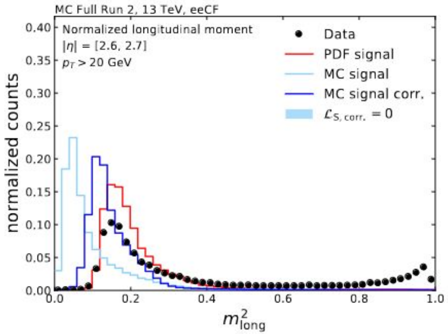
Shower Shapes

after applying fudge factors (in EMEC)

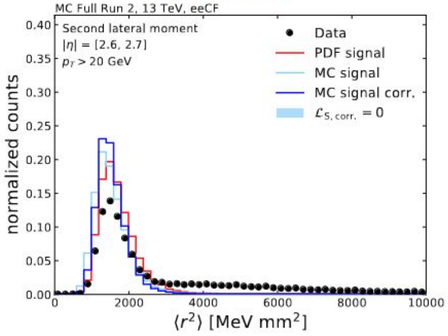
SecondLambda



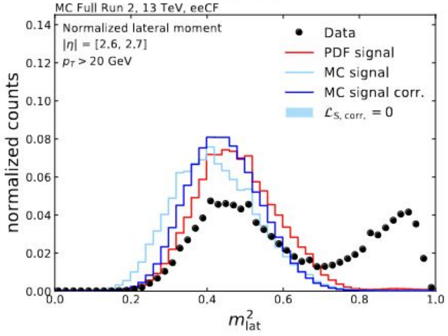
Longitudinal



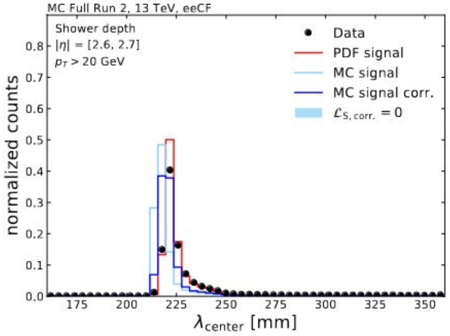
SecondR



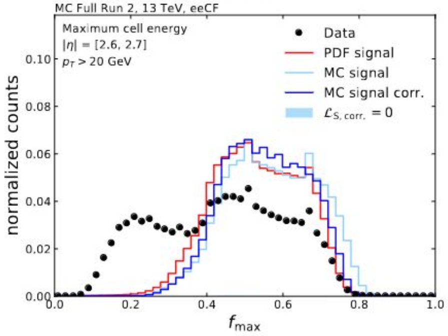
Lateral



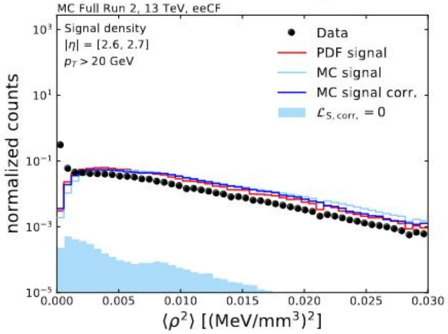
CenterLambda



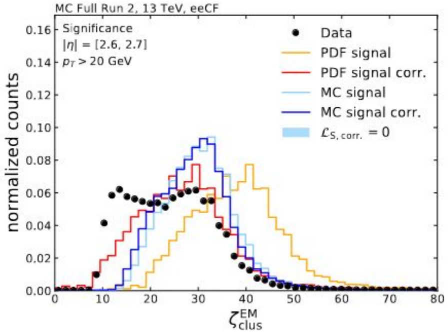
FracMax



SecondDensity



Significance



Efficiency Correction
Motivation

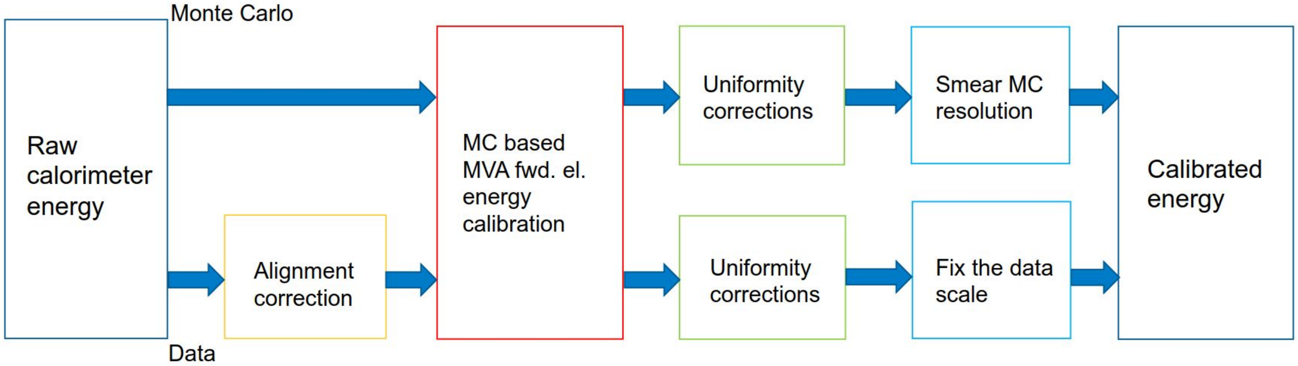
- Residual shower shape differences after applying fudge factors
  → lead to ID efficiency differences between data and MC
  → need to correct Monte-Carlo efficiency by applying scale-factors to events
- shower shapes and real/fake-electron discrimination power depend on electron energy and traversed material / detector parts
  → determine identification efficiency in bins of p_t and $|\eta|$
- electron identification efficiency a priori unknown in data
  → need to extract using fit or counting method after background subtraction
  → need well calibrated sample of real forward electrons!

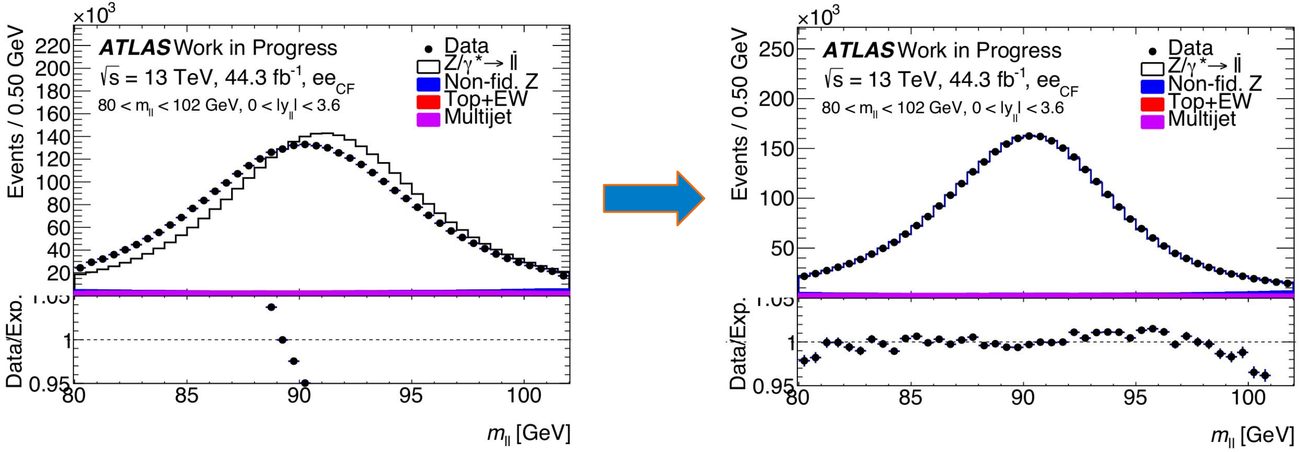
Efficiency:

$$\epsilon_S^{ID} = \frac{N_S^{\mathrm{passID}}}{N_S^{\mathrm{all}}}$$

Forward Electron Energy

Calibration Process





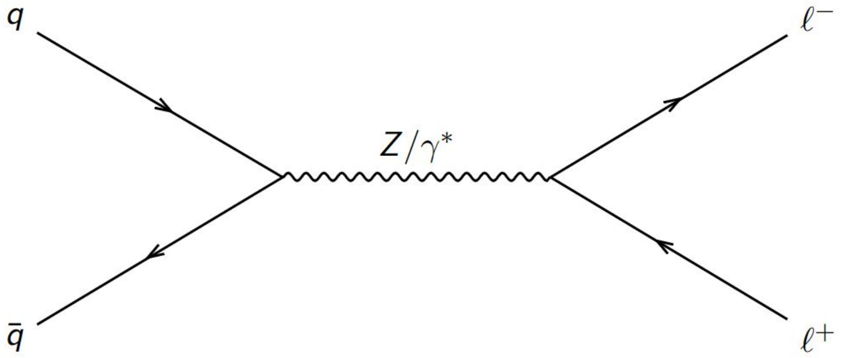
- correcting the raw energy of forward electrons using multi-variate-analysis
- harmonizing the energy response of calorimeter cells
- matching the mass spectrum of the Z boson in data and Monte-Carlo (in-situ calibration)

- performed by Craig Wells
- for more details, see: https://cds.cern.ch/record/2789786?ln=en

Tag and Probe Method

to get unbiased sample of forward electrons

- select electron pair from known resonance:



- no charge information for forward electron!
  - significant amount of background in selected sample
  - need to model and fit signal / background

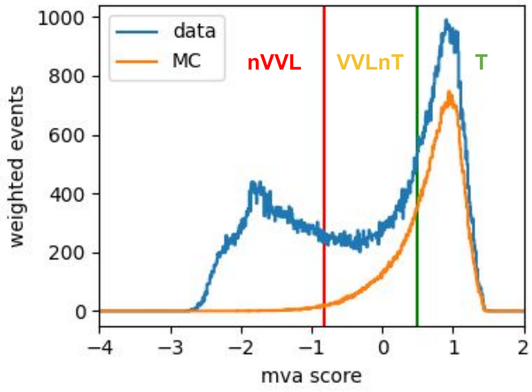
	Tag Electron (central)	Probe Electron (forward)
$\lvert\eta\rvert$	< 2.47 (excluding crack)	> 2.5 < 4.9
p_t	> 27 GeV	> 20 GeV
ID	**tight + isolation**	**no requirement**
trigger	HLT_e26_lhtight_nod0_ivarloose HLT_e60_lhmedium_nod0 HLT_e140_lhloose_nod0	
isEE_reco	passed	
DQ	passed	
MC truth-matching	passed	

Efficiency Fit

Basic model

- Signal
  - Monte-Carlo approach
  - Shape modified by ID selection
  - Size modified by ID efficiency

3 ID regions:



example using 2017 data
$3.4 < |\eta^{fwd}| < 3.5$
30 GeV < p_t^{fwd} < 35 GeV

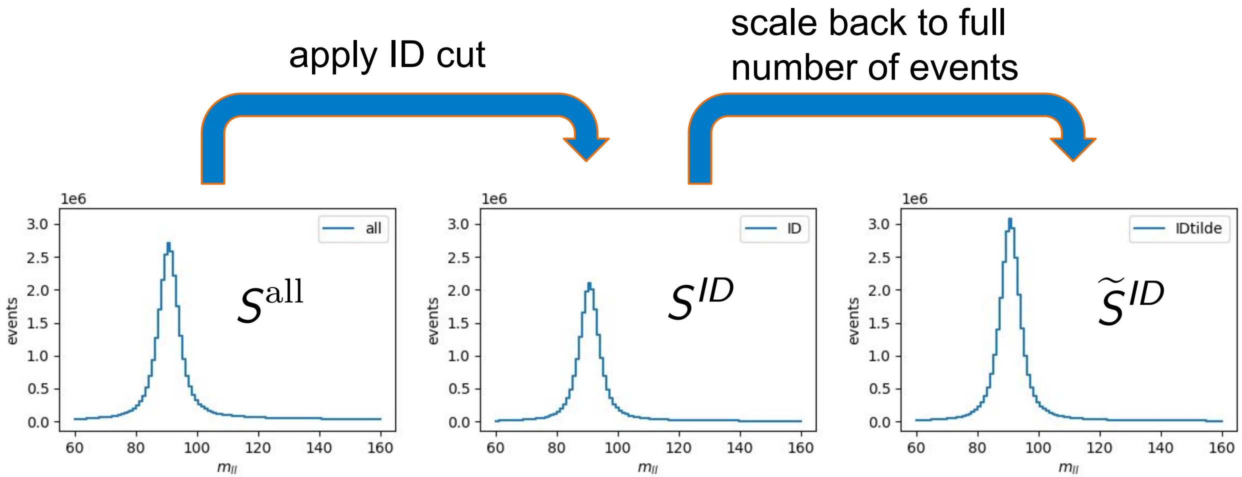
$$D^{ID} = S^{\mathrm{all}} \cdot \epsilon_S^{ID} + B$$

fit parameters in blue

Efficiency Fit

Basic model: Signal

- Signal
  - Monte-Carlo approach
  - Shape modified by ID selection
  - Size modified by ID efficiency
  - Normalized to match data



$$D^{ID}(m_{\ell\ell}) = \mu\widetilde{S}^{ID}(m_{\ell\ell})\epsilon_S^{ID} + B(m_{\ell\ell})$$

normalization correction

$$\widetilde{S}^{ID}(m_{\ell\ell}) = \frac{S^{ID}(m_{\ell\ell})}{\epsilon_S^{ID}}$$

Efficiency Fit

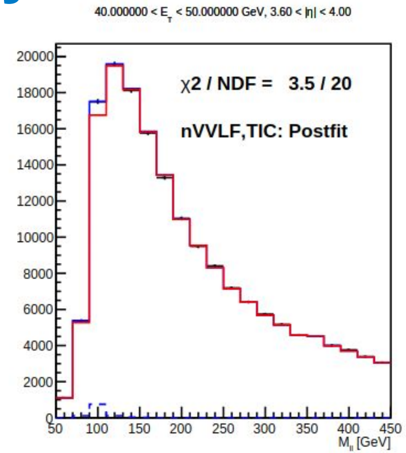
Basic model: Background

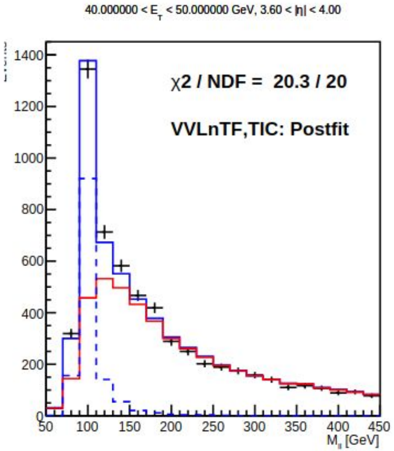
- Signal
  - Monte-Carlo approach
  - Shape modified by ID selection
  - Size modified by ID efficiency
  - Normalized to match data
- Background
  - Shape and size a priori unknown
  - Data-driven approach
  - Expect shape to be independent of chosen ID
  - Expect normalization to depend heavily on chosen ID

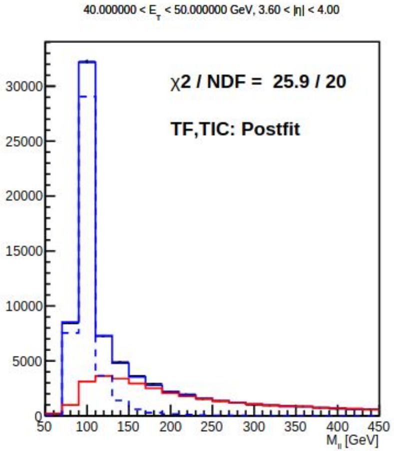
$$D^{ID}(m_{\ell\ell}) \quad = \quad \mu\widetilde{S}^{ID}(m_{\ell\ell})\epsilon_S^{ID} + B^{!VVL}(m_{\ell\ell})\frac{\epsilon_B^{ID}}{1-\epsilon_B^{VVL}}$$

Efficiency Fit

In 3 Regions







$$D^{!VVL}(m_{\ell\ell}) = \mu\widetilde{S}^{!VVL}(m_{\ell\ell})(1-\epsilon_S^{VVL}) + B^{!VVL}(m_{\ell\ell})$$

- dominated by background
- use as initial parameters for background shape

$$D^{VVL!T}(m_{\ell\ell}) = \mu\widetilde{S}^{VVL!T}(m_{\ell\ell})(\epsilon_S^{VVL}-\epsilon_S^{T}) + B^{!VVL}(m_{\ell\ell})\frac{\epsilon_B^{VVL}-\epsilon_B^{T}}{1-\epsilon_B^{VVL}}$$

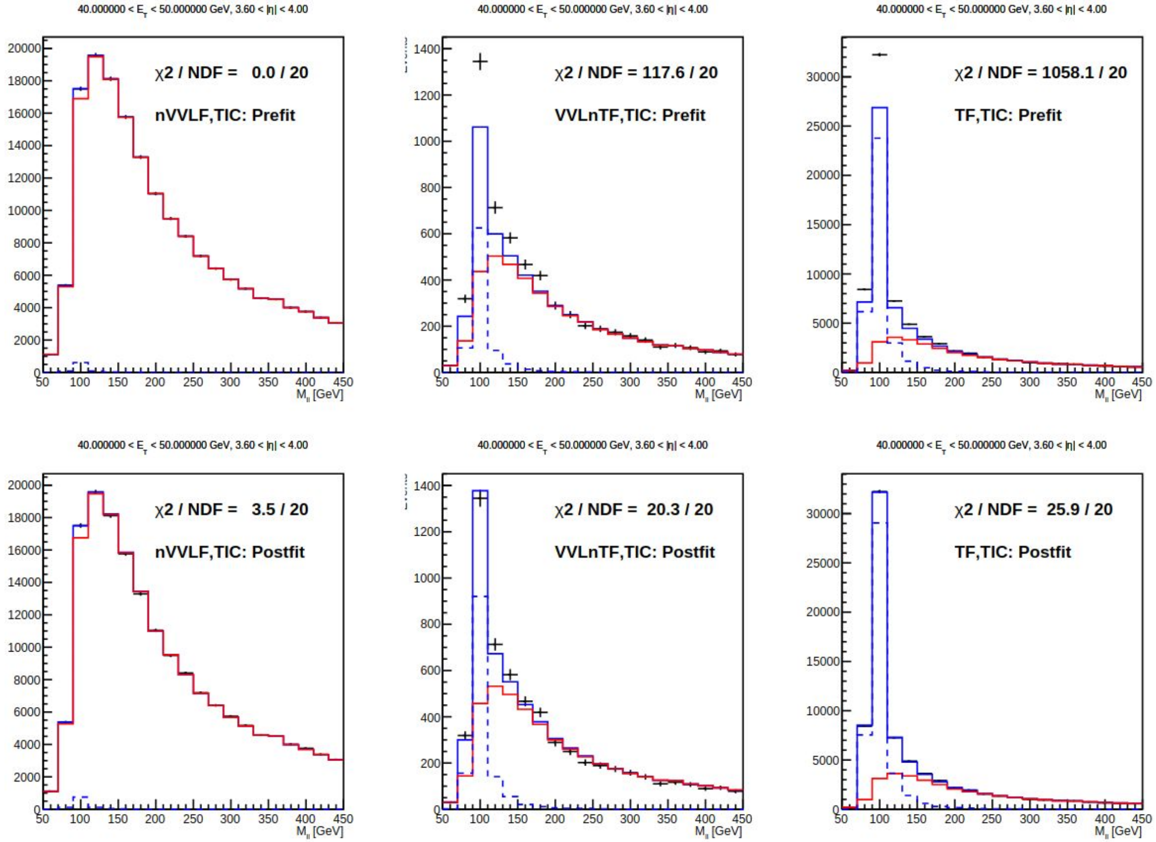
- significant contributions to both background and signal
- increasing statistics and # data points to help the fit
  - (the model has too many d.o.f. to be fitted in 2 regions)

$$D^{T}(m_{\ell\ell}) = \mu\widetilde{S}^{T}(m_{\ell\ell})\epsilon_S^{T} + B^{!VVL}(m_{\ell\ell})\frac{\epsilon_B^{T}}{1-\epsilon_B^{VVL}}$$

- dominated by signal
- use to fit the POI

Results: fit

example: 2017 data, $3.6 < |\eta^{fwd}| < 4.0$, 40 GeV < p_t^{fwd} < 50 GeV

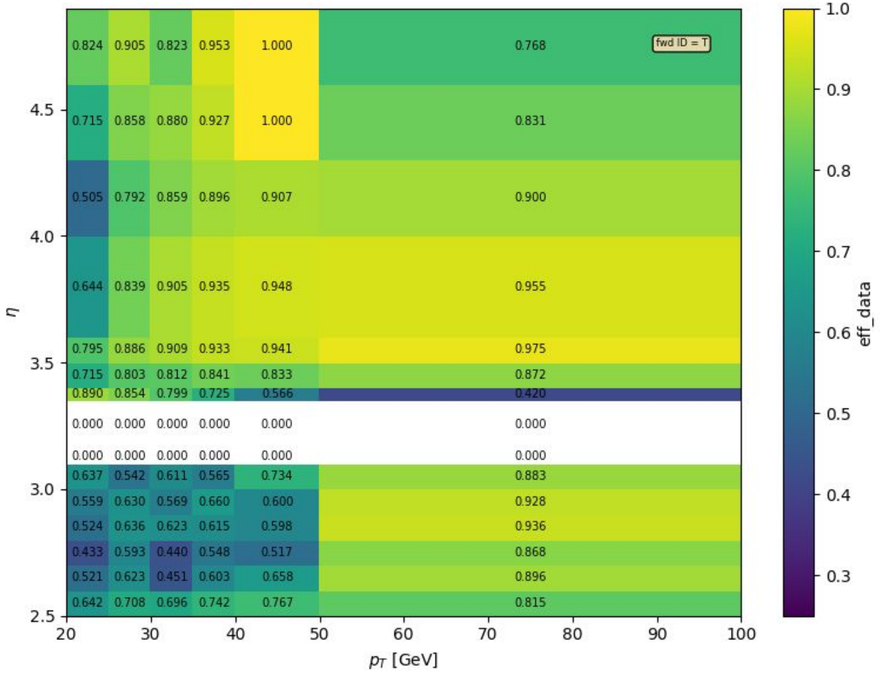


— Data
— Background estimate
- - Signal estimate
— Total estimate

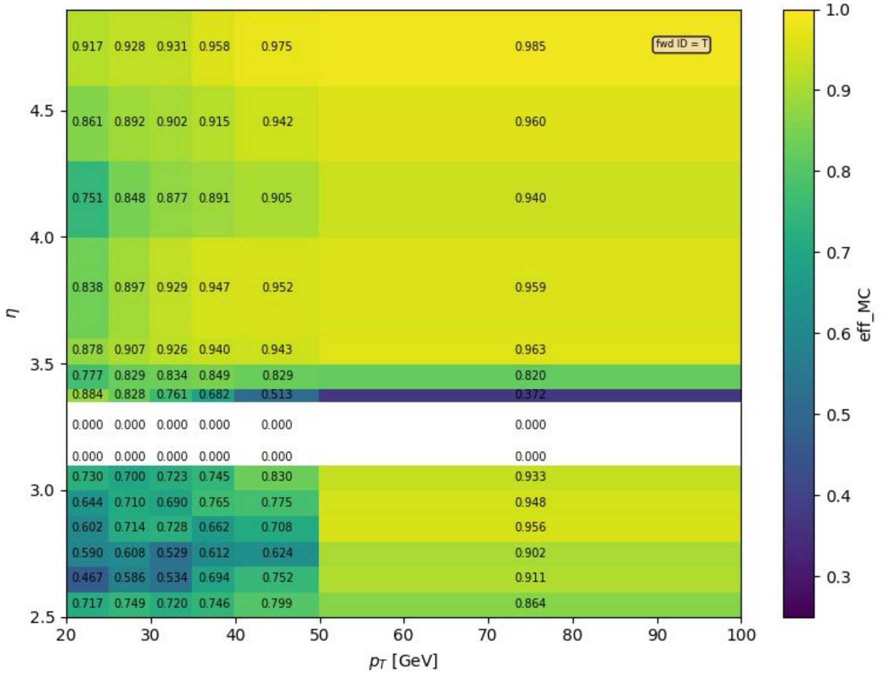
Results: fwd tight Efficiency

2017

data

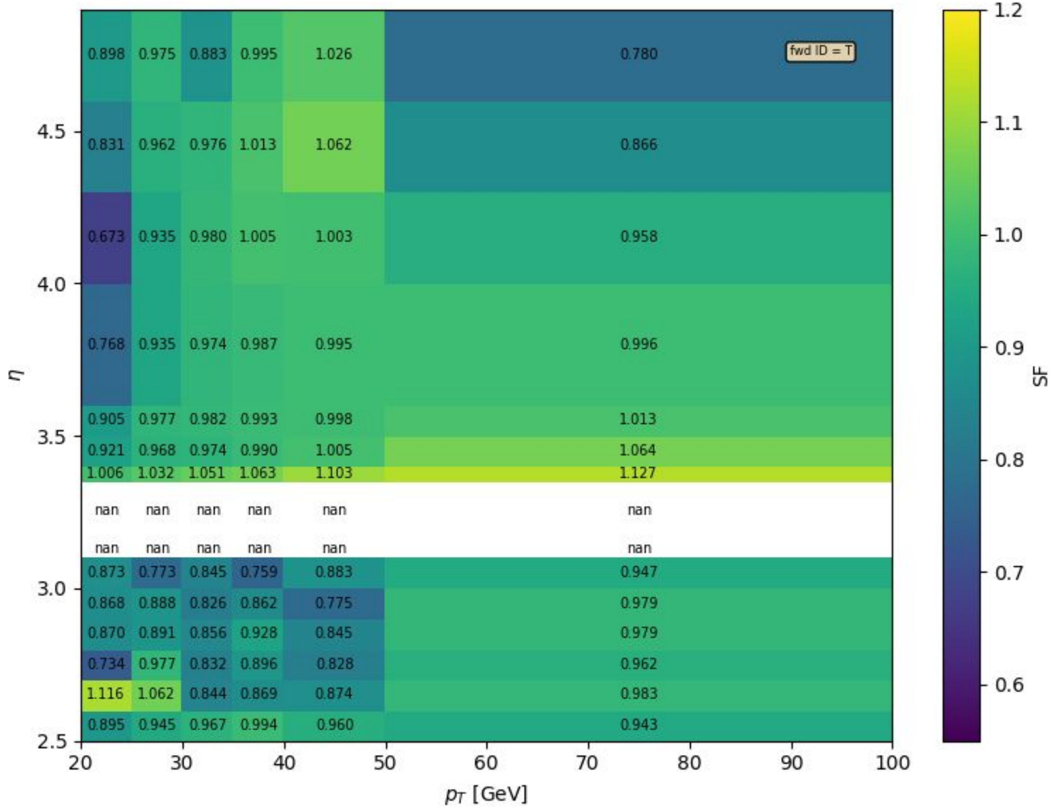


Monte-Carlo



Results: fwd tight Scale-Factor (SF)

2017

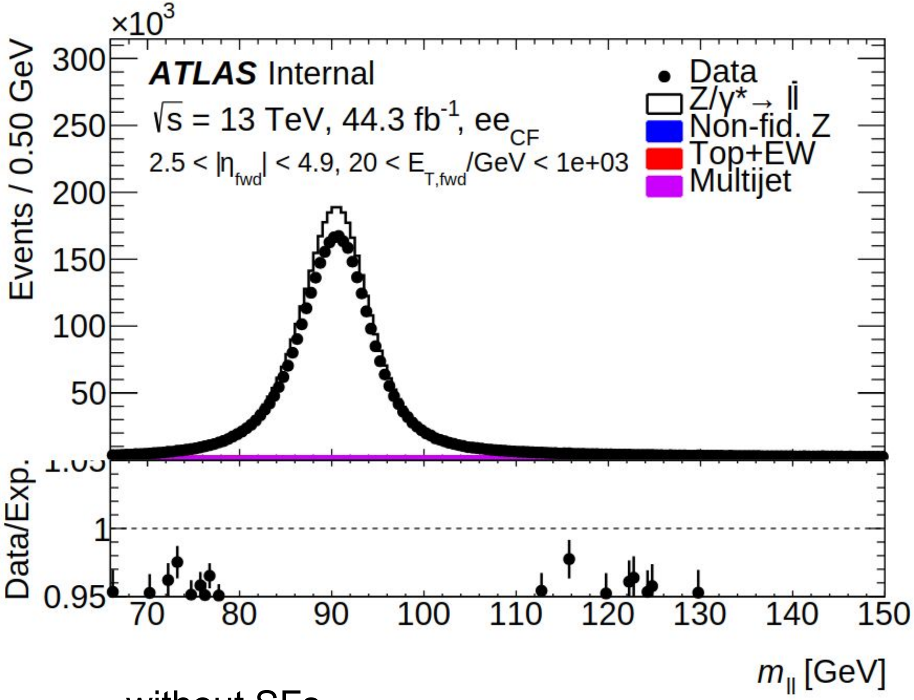


Scale-Factor:

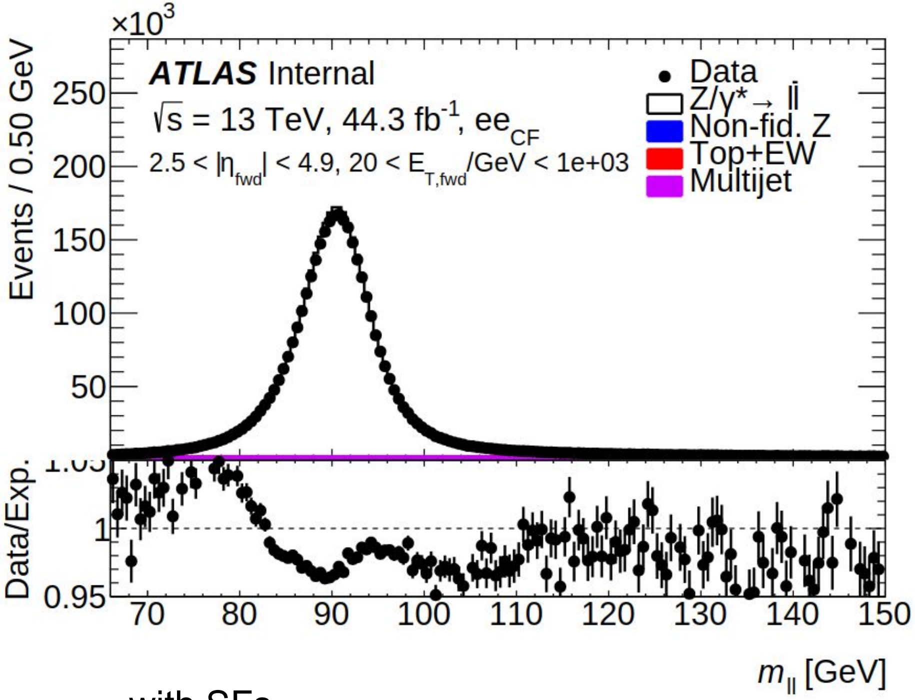
$$\mathrm{SF} = \frac{\epsilon_S^{ID}(\mathrm{data})}{\epsilon_S^{ID}(\mathrm{MC})}$$

Results: control plots

example: 2017 data, mll



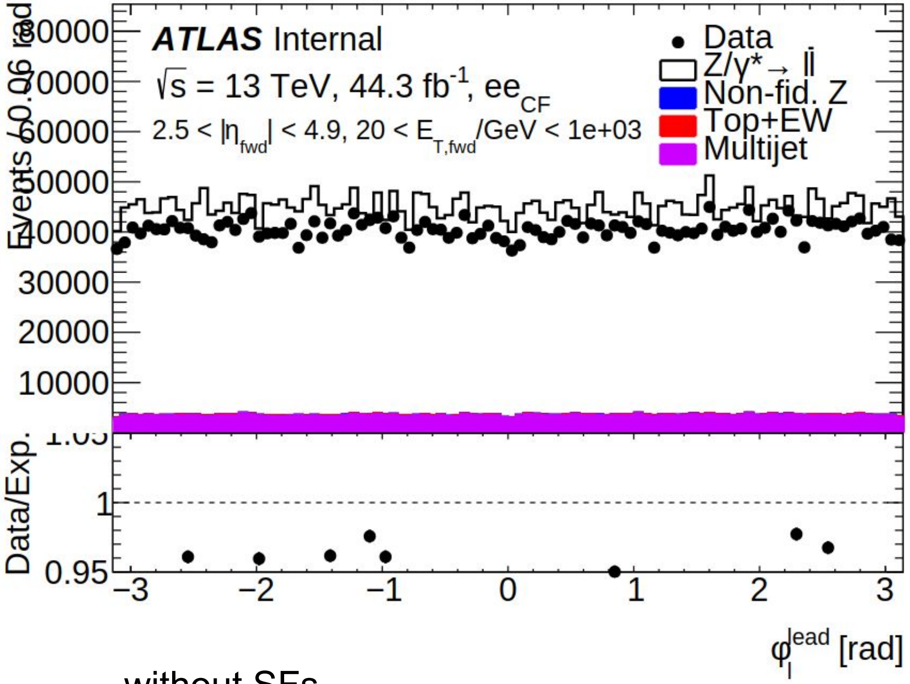
without SFs



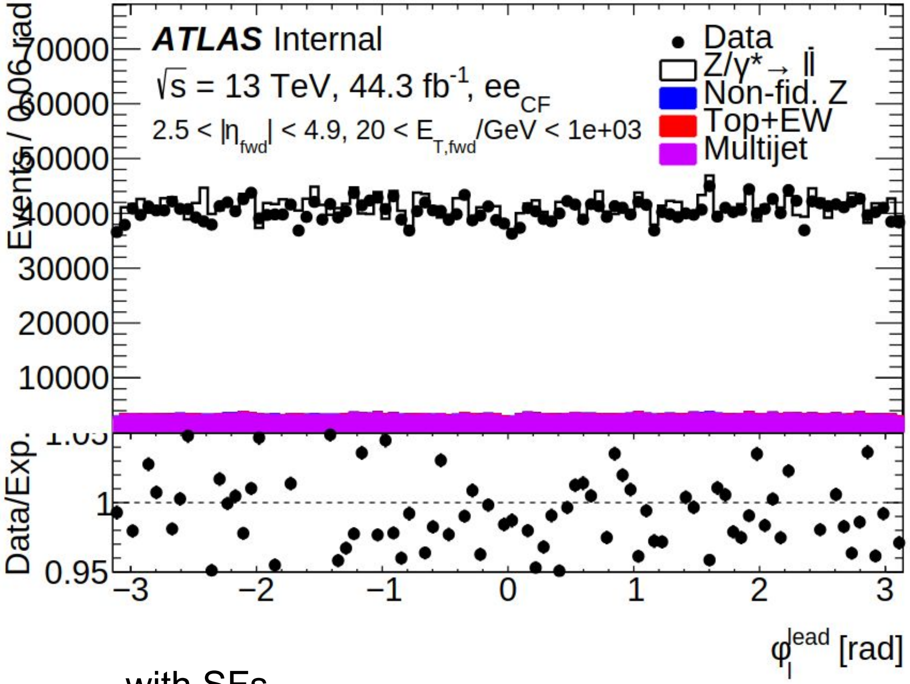
with SFs

Results: control plots

example: 2017 data, phi lead



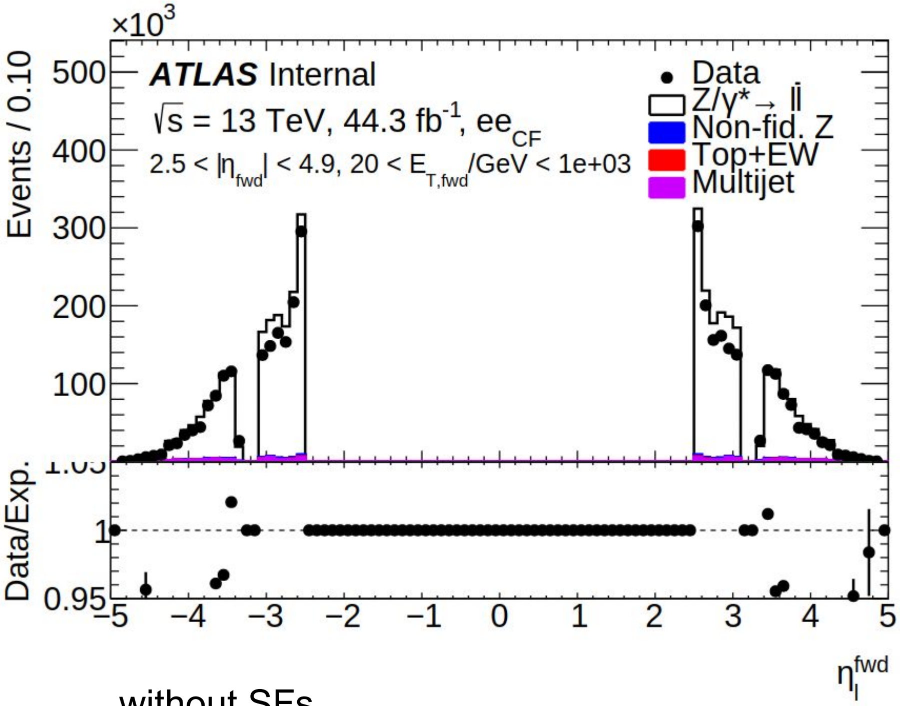
without SFs



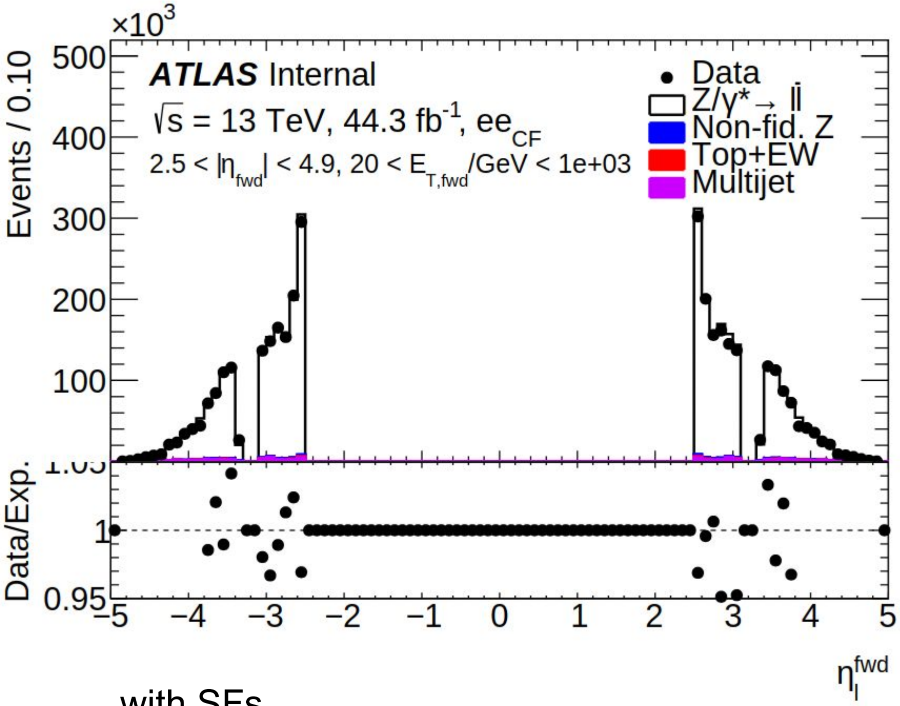
with SFs

Results: control plots

example: 2017 data, eta fwd



without SFs



with SFs

Systematic Uncertainties

BkgExtrap

	Tag Electron	Probe Electron Signal Region	Probe Electron Control Region
ID	tight + isolation (TIC)	tight (T)	failing very very loose (nVVL) → failing very loose (nVL)

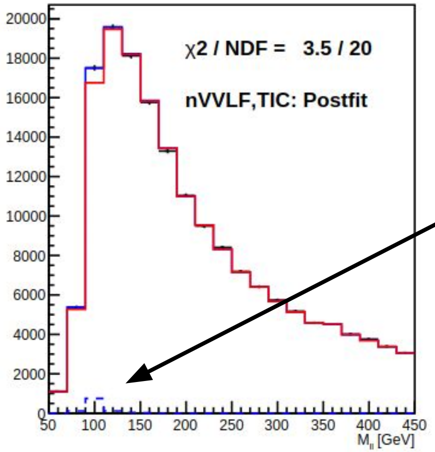
CenID

	Tag Electron	Probe Electron Signal Region	Probe Electron Control Region
ID	tight + isolation (TIC) → tight (T)	tight (T)	failing very very loose (nVVL)

Signal Contamination

VVL inefficiency estimated via Monte-Carlo → vary by ± 50%

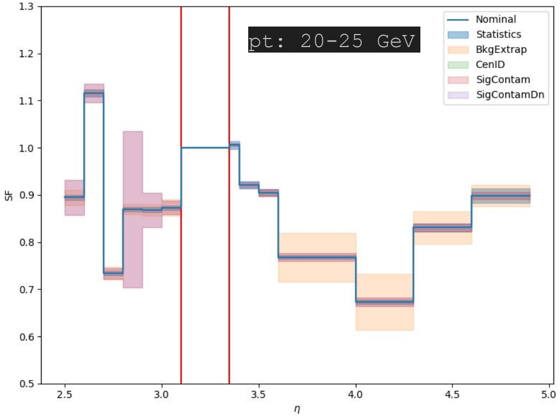
$$D^{!VVL}(m_{\ell\ell}) \quad = \quad \mu\widetilde{S}^{!VVL}(m_{\ell\ell})(1-\epsilon_S^{VVL}) + B^{!VVL}(m_{\ell\ell})$$

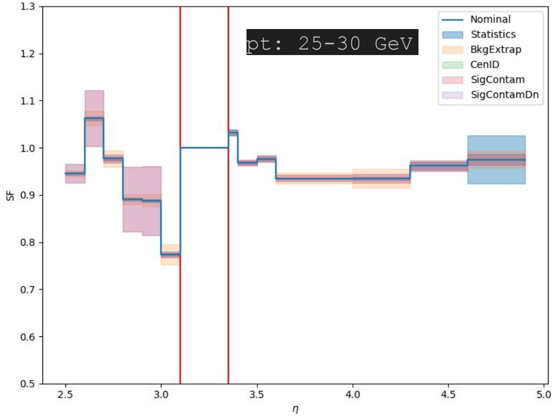


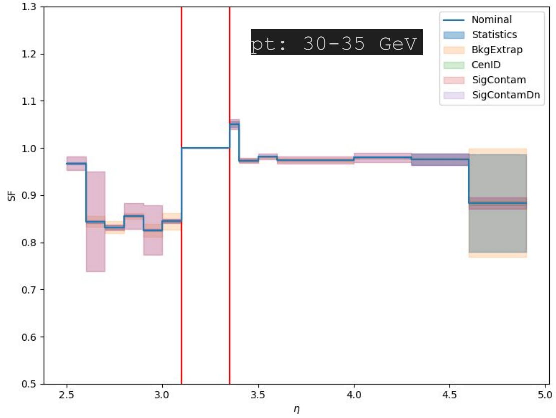
signal contamination in control region

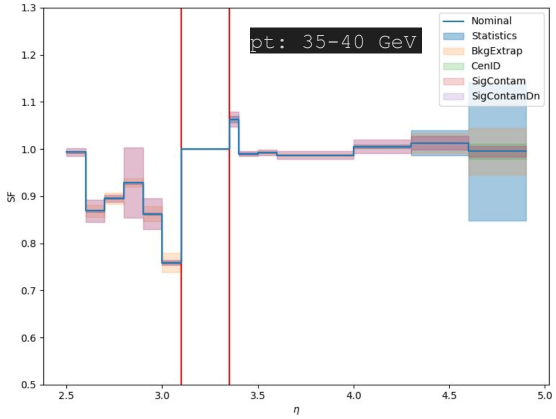
Systematic Uncertainties

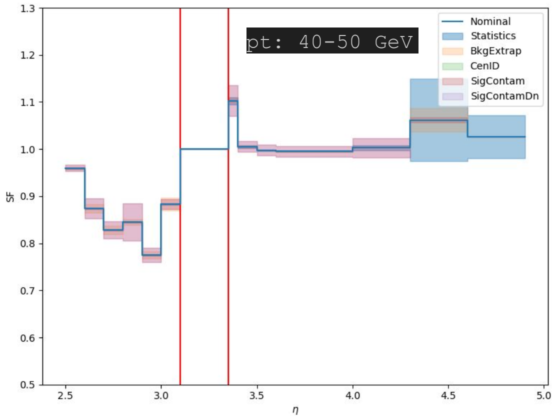
2017

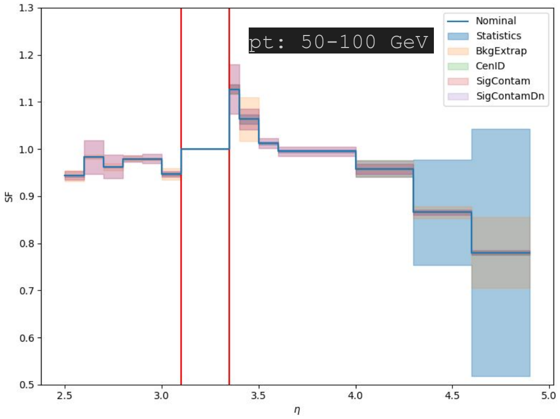












Comparison

to forward electron ID SF calculation by Luxin Zhang (University of Science and Technology of China)

- different shower-shape correction (only affects EMEC)
- different electron energy calibration
- different η binning
  - using η instead of |η| → not assuming forward-backward symmetry of detector / efficiency
  - but much coarser binning in η (6 vs 14 bins)
  - # events per bin in same order of magnitude
- different methodology to determine SFs:
  - Luxin: subtract background and count events
    - in narrow region around Z peak

$$\epsilon_S^{ID} = \frac{N_S^{\mathrm{passID}}}{N_S^{\mathrm{all}}}$$

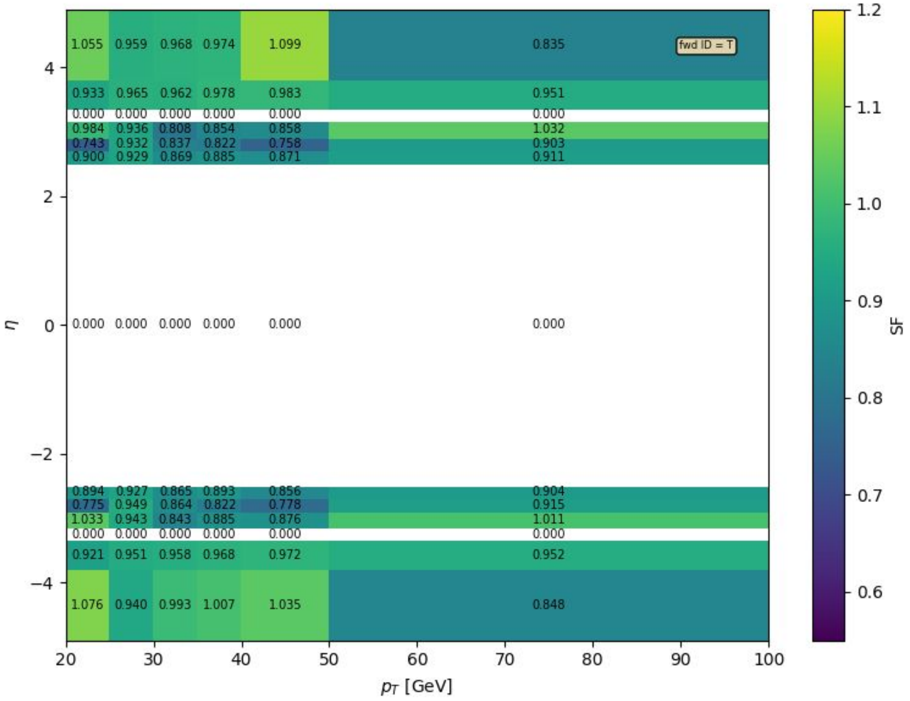
  - Lukas: perform simultaneous fit in 3 ID regions
    - in wide mll window
      → less sensitive to calibration errors

$$D^{ID}(m_{\ell\ell}) = \mu \widetilde{S}^{ID}(m_{\ell\ell})\epsilon_S^{ID} + B^{!VVL}(m_{\ell\ell})\frac{\epsilon_B^{ID}}{1-\epsilon_B^{VVL}}$$

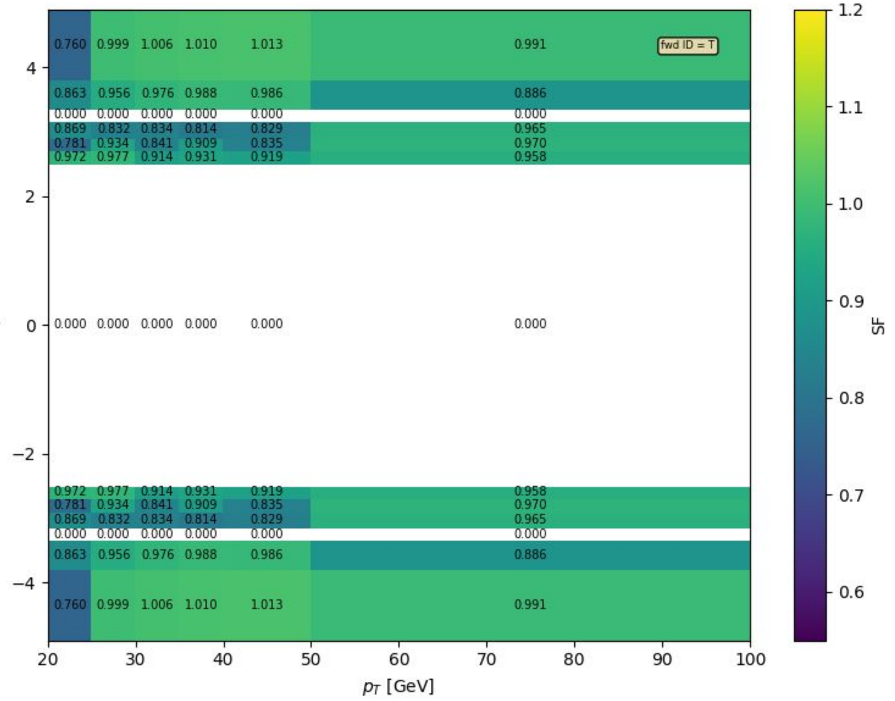
Comparison

of forward electron ID SFs

Luxin's SFs



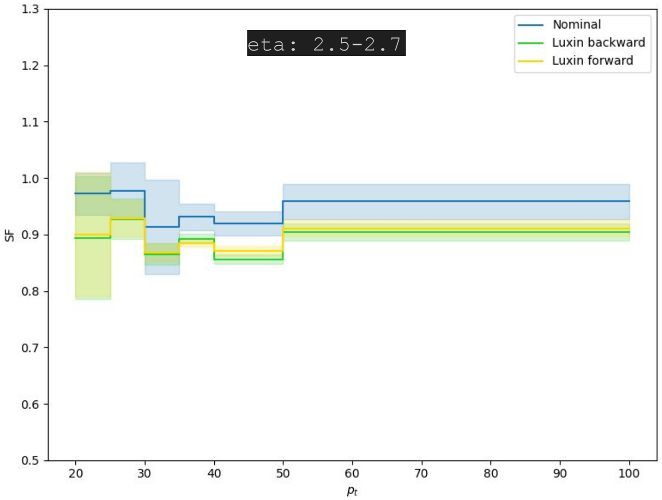
Lukas' SFs (recomputed with coarse binning)

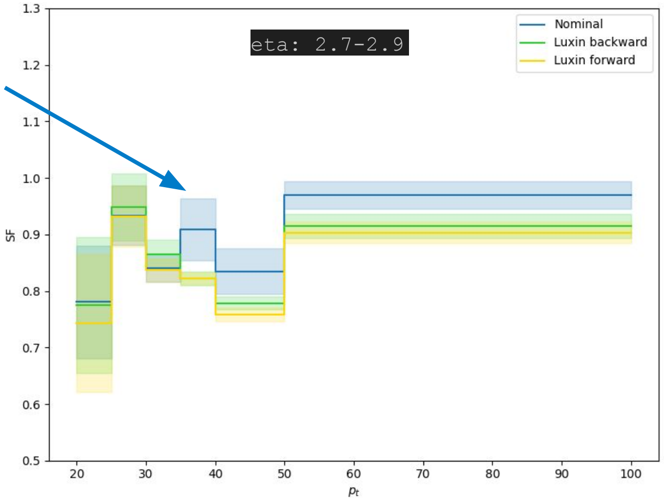


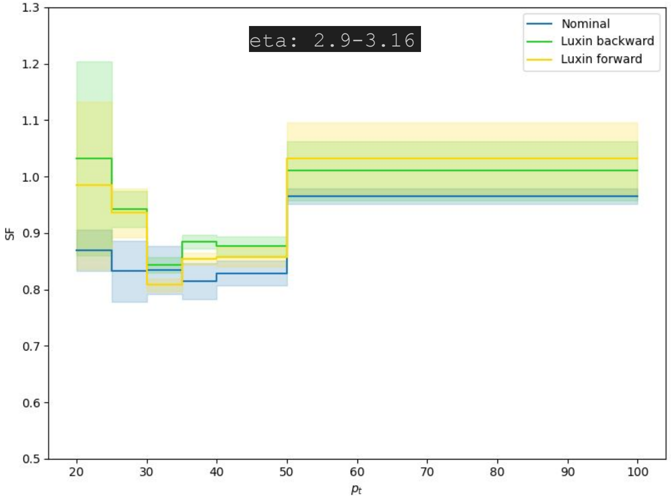
Comparison

in the EMEC

uncertainty dominated by signal contamination



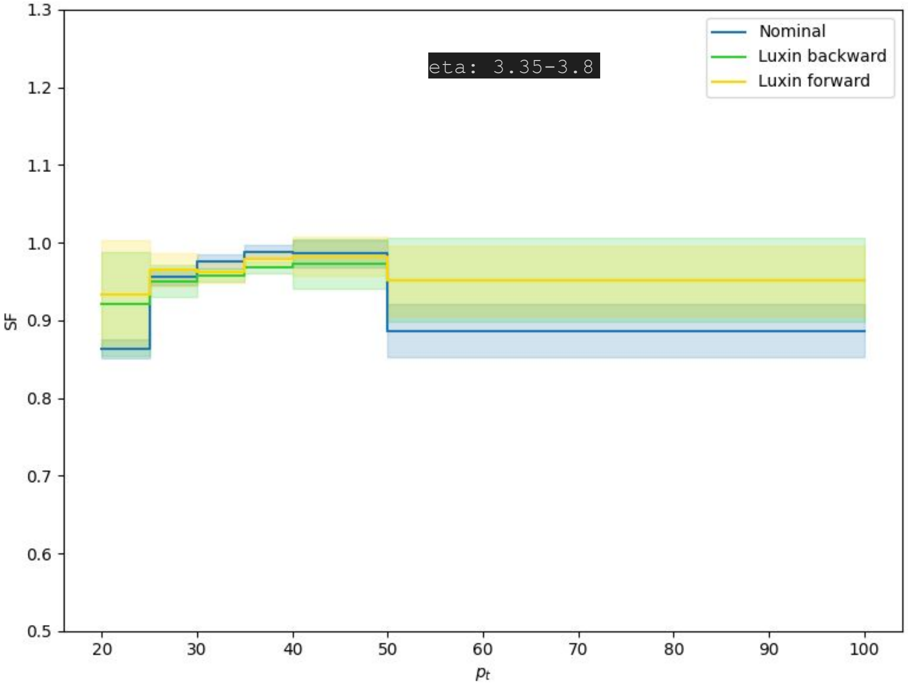


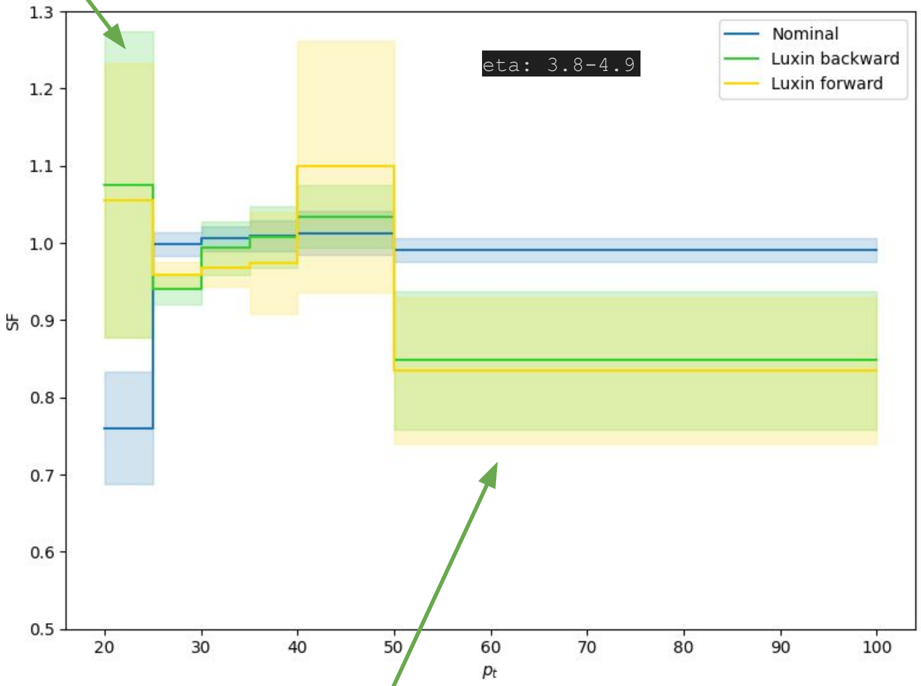


Comparison

in the FCAL

uncertainty dominated by
MJ estimate





uncertainty dominated by MJ
estimate and integral range

Summary

and conclusion

- successfully computed full-Run2 tight-ID efficiency scale-factors for forward electrons using
  - tag-and-probe method to select sample
  - simultaneous fit in 3 ID regions to determine efficiencies

- results for 2017 presented here (2015/16 and 2018 in Backup)
  - SFs are able to resolve significant data-to-Monte-Carlo differences in control plots

- comparison to Luxin’s SFs
  - observed differences due to calibration, shower-shape corrections and methodology of SF computation
  - our SFs have finer binning
  - assume forward-backward symmetry
    - justified since forward and backward SFs agree with each other within uncertainties
  - more suited for our analysis of Z four-fold cross-section measurement and extraction of angular coefficients
    - less dependent on background model
    - less sensitive to energy miscalibration
    - [analysis glance: https://atlas-glance.cern.ch/atlas/analysis/analyses/details.php?ref_code=ANA-STDM-2018-46]

Thank you
for your attention

Contact

Deutsches Elektronen-
Synchrotron DESY

www.desy.de

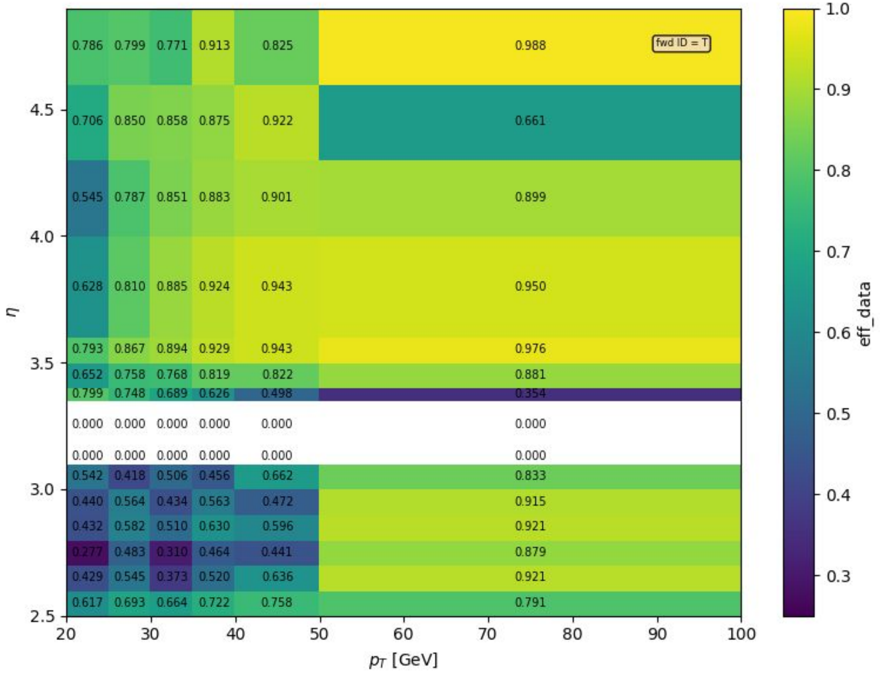
Lukas Bayer
DESY ATLAS Standard-Model Group

lukas.bayer@desy.de

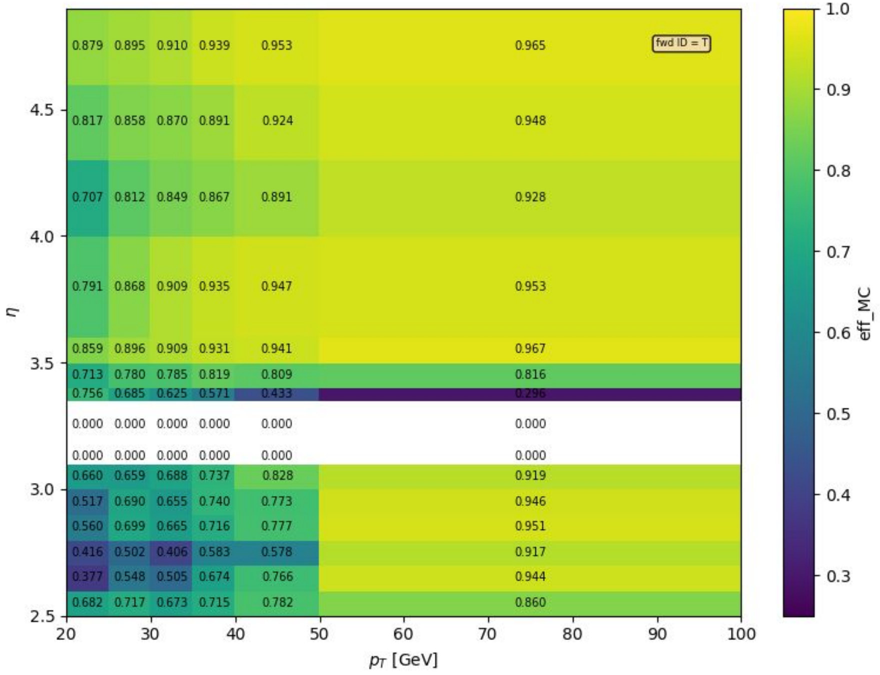
Backup: fwd tight Efficiency

2015/16

data



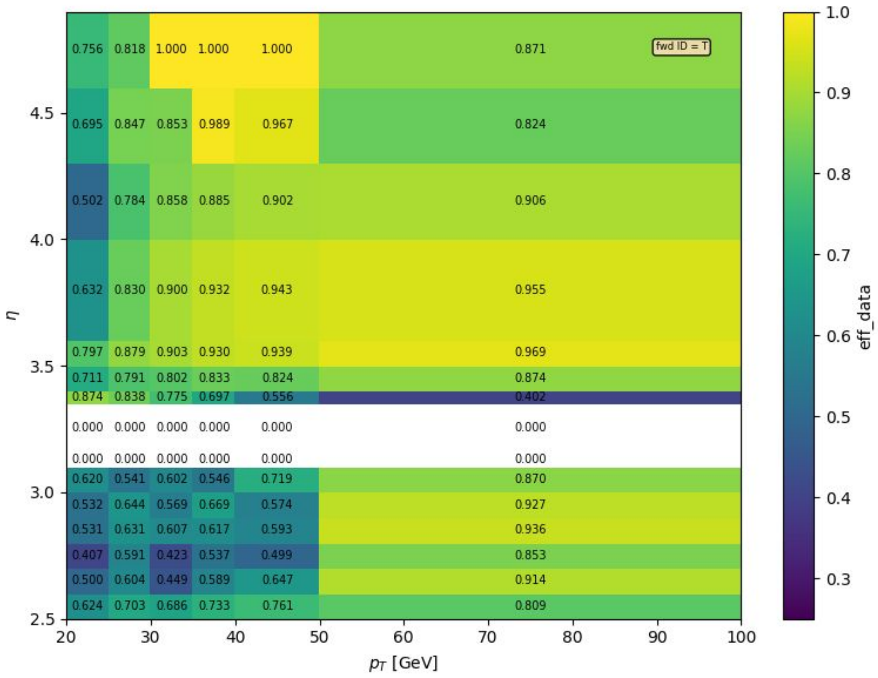
Monte-Carlo



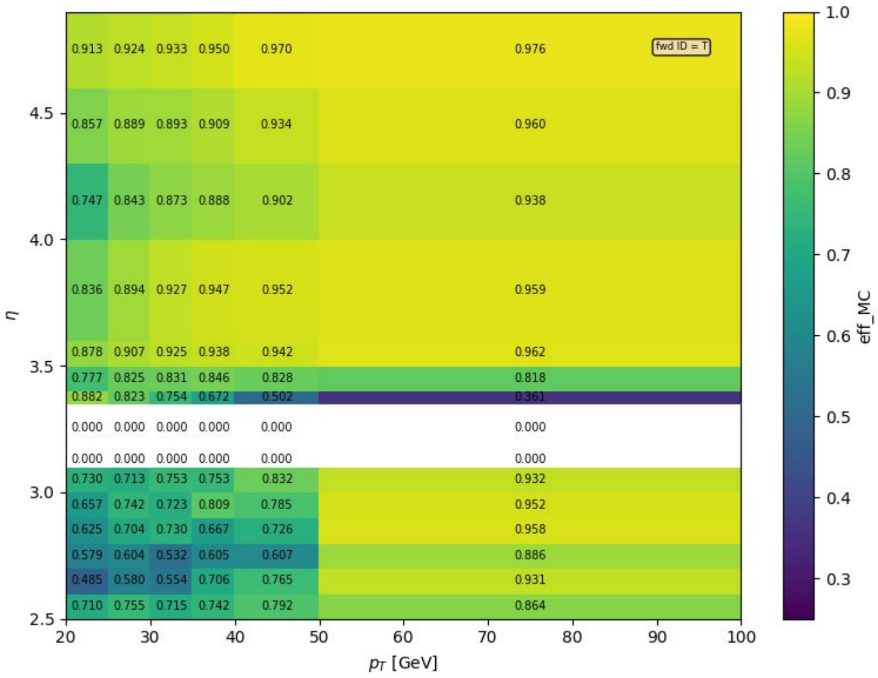
Backup: fwd tight Efficiency

2018

data

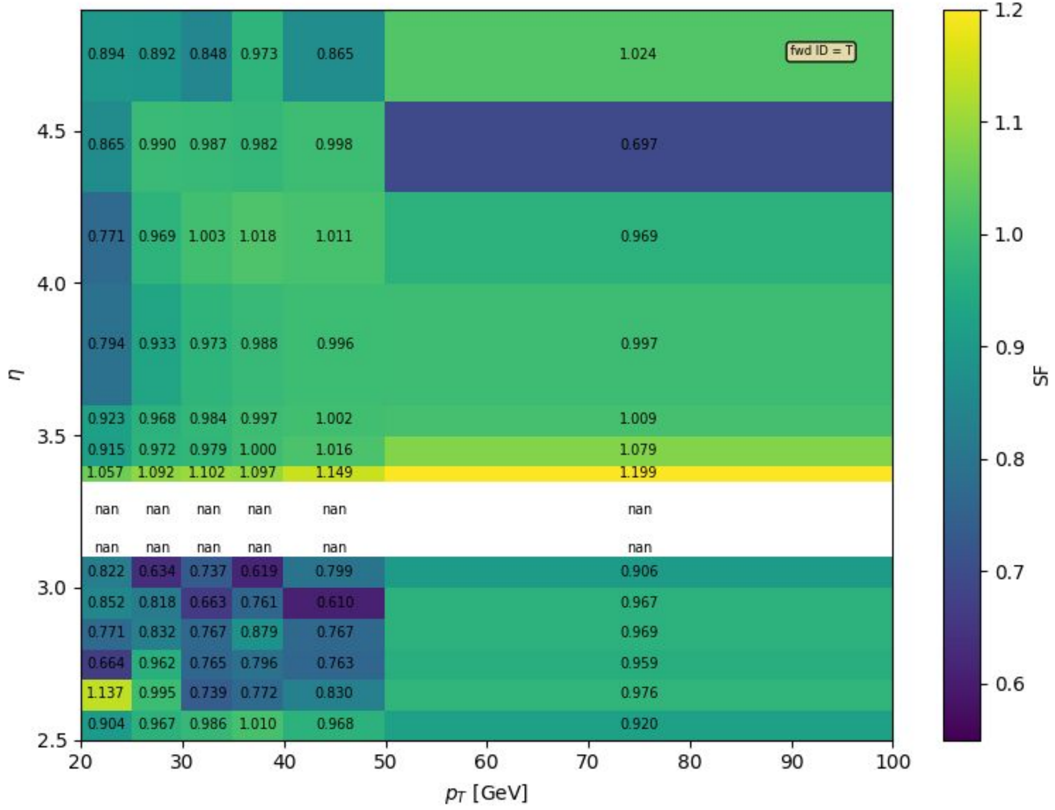


Monte-Carlo



Backup: fwd tight Scale-Factor (SF)

2015/16

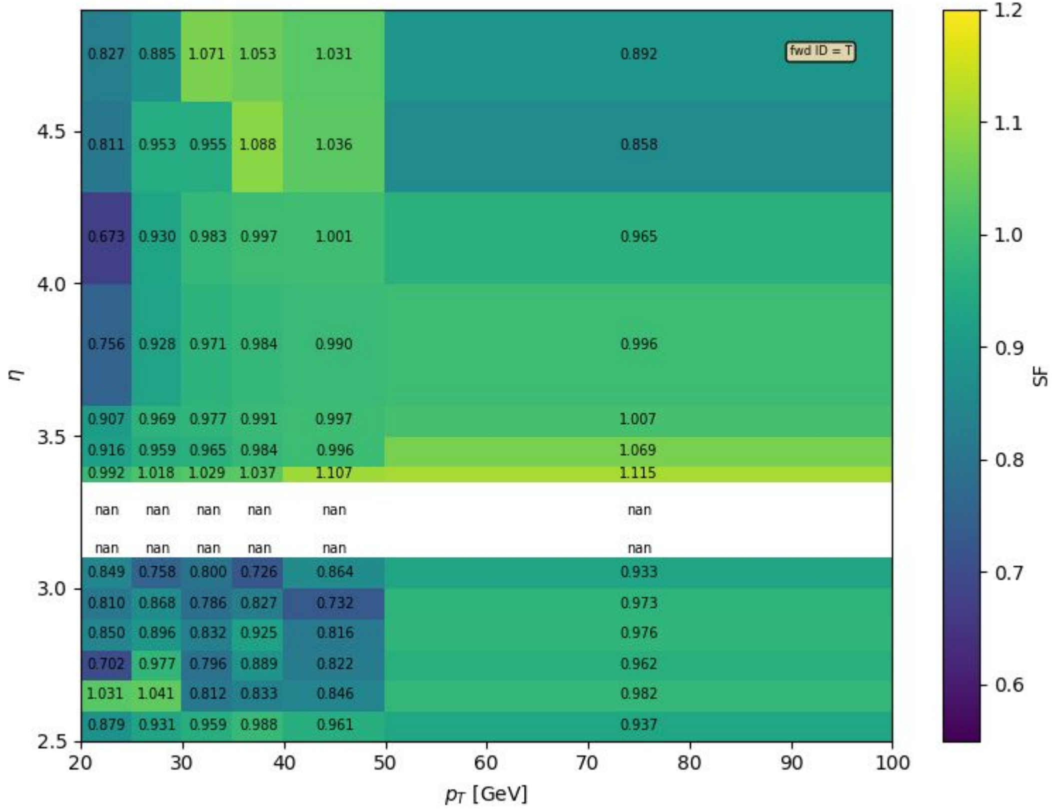


Scale-Factor:

$$\mathrm{SF} = \frac{\epsilon_S^{ID}(\mathrm{data})}{\epsilon_S^{ID}(\mathrm{MC})}$$

Backup: fwd tight Scale-Factor (SF)

2018

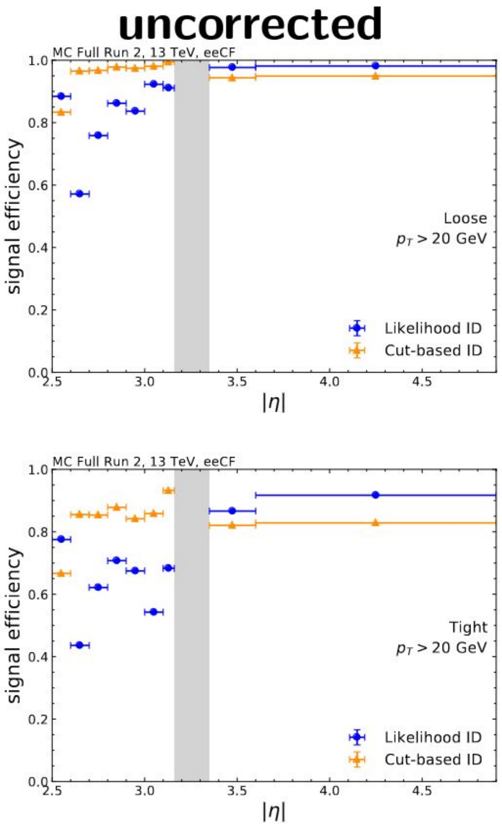


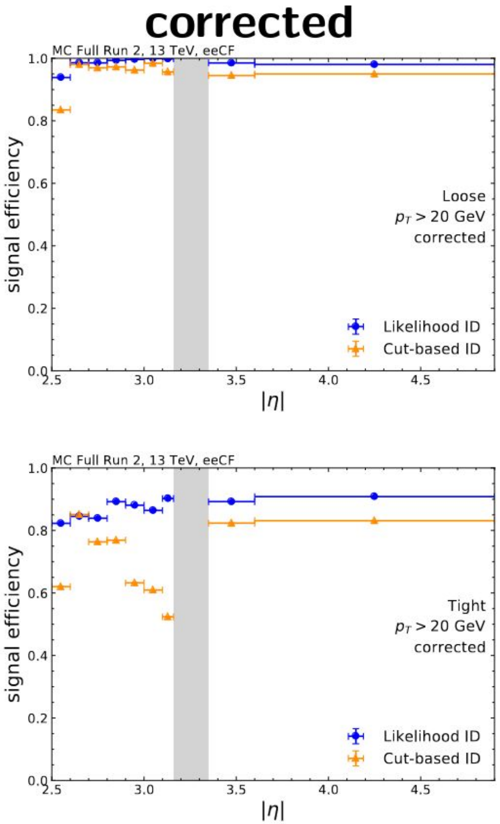
Scale-Factor:

$$\mathrm{SF} = \frac{\epsilon_S^{ID}(\mathrm{data})}{\epsilon_S^{ID}(\mathrm{MC})}$$

Backup: Shower Shapes

effect of shower shape fudging on signal efficiency



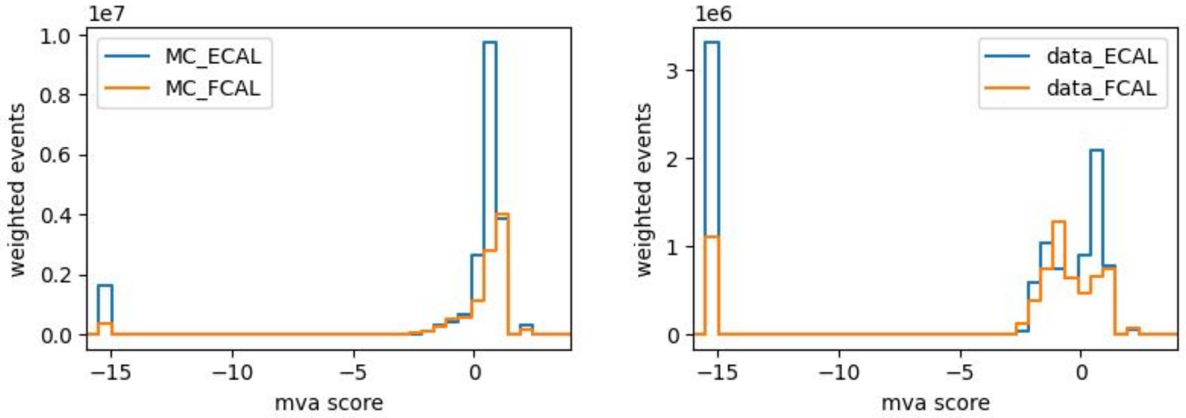


Backup: Systematic Uncertainties

why is the signal contamination uncertainty large in the ECAL?

underflow bin (-15) in mva likelihood

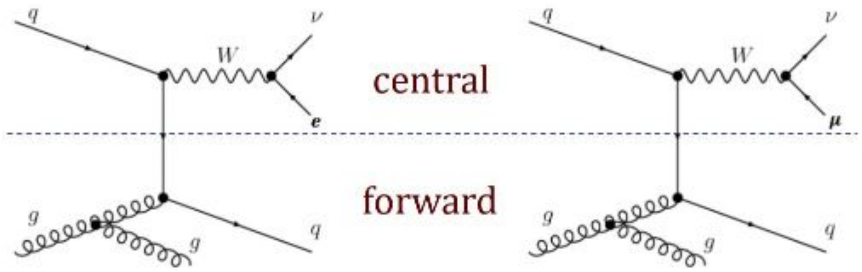
- originates from failed events in which no mva score could be assigned
- larger in data than Monte-Carlo
- larger in EMEC than FCAL
- events fail VVL
- not respected in ID working point definitions



Backup: Background Estimation

in Luxin Zhang’s SF calculation

μe reweight method for W+jets, ttbar, WW…

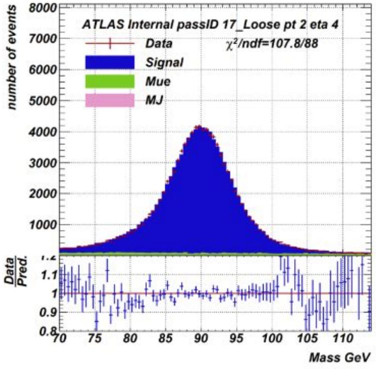


A = B*C/D

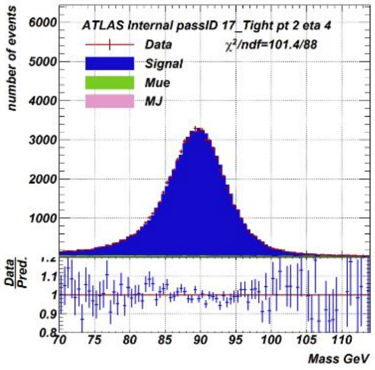
A W+jet , ttbar, WW… ee events Signal region	**B** W+jet , ttbar, WW… μe events Signal region
C W+jet ee events Control reigon	**D** W+jet μe events Control region

template fit for multijet (MJ) estimation

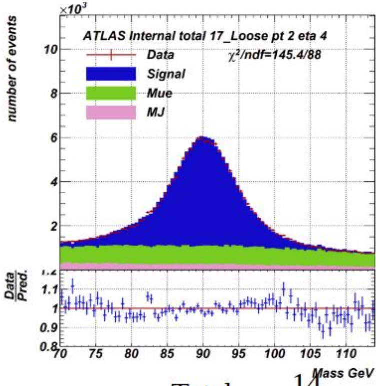
- multijet template:
  - reverse cen iso (but keep ID)
  - reverse fwd ID loose
- fit mll range 70-114 GeV
- for each p_t and η bin
- for each ID region



Loose



Tight



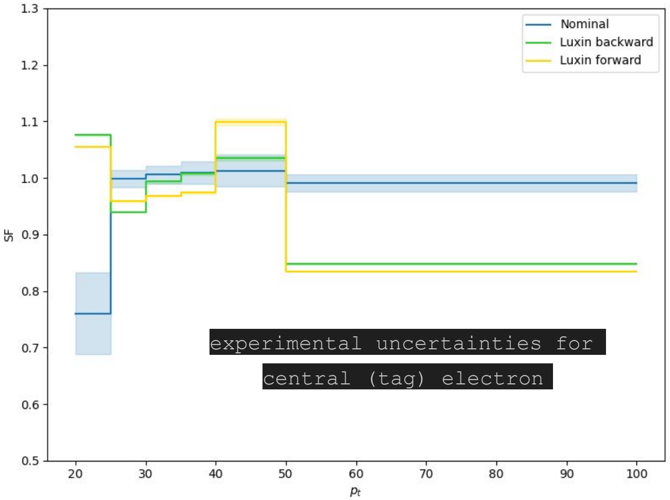
Total

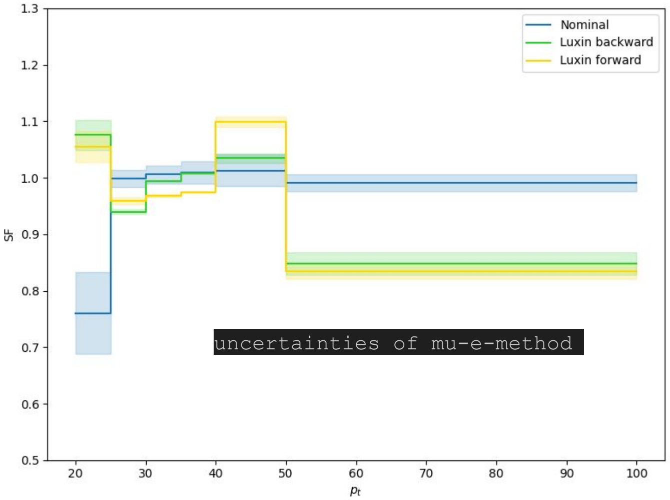
14

Backup: Systematics

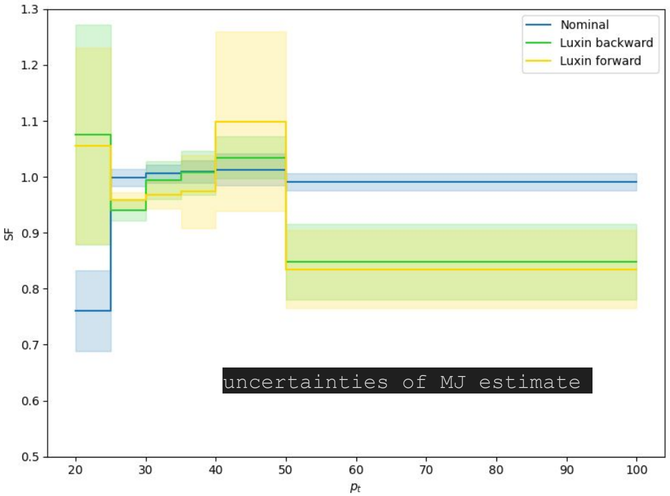
categories of uncertainties

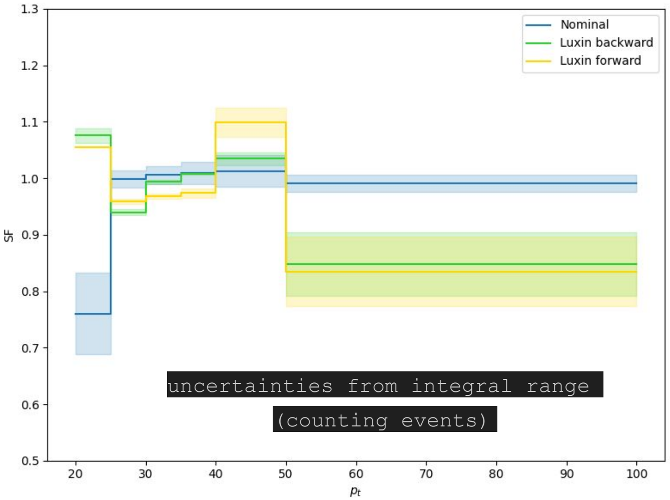
eta: 3.8-4.9





[these plots show only those uncertainties of Luxin's results, that correspond to a certain source, and the full uncertainty of my results for comparison]

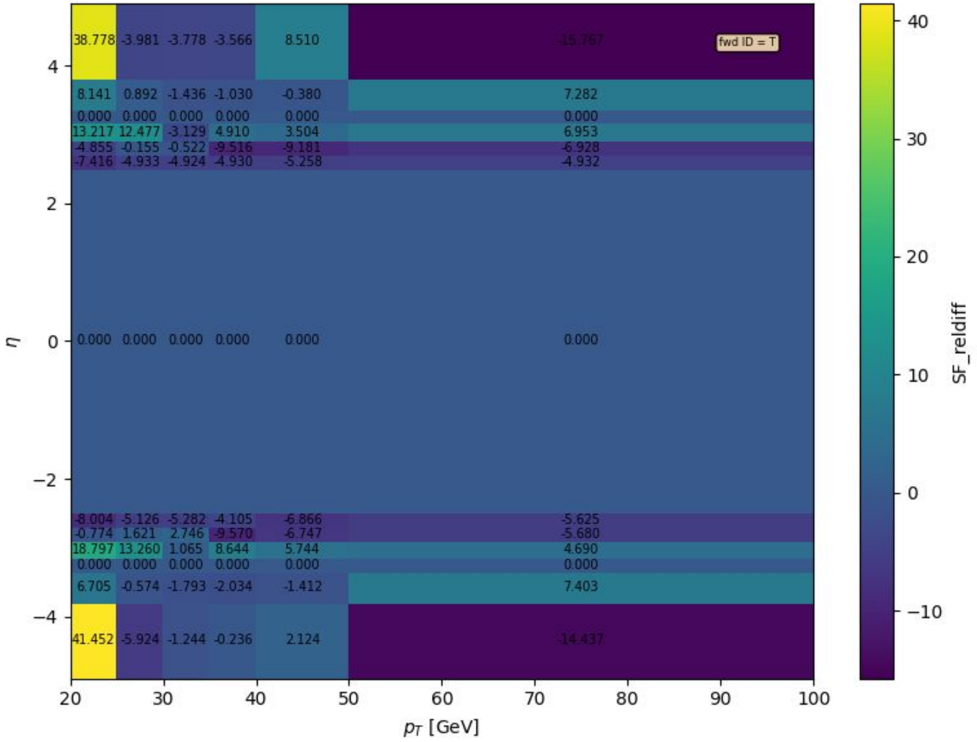




Backup: Relative Differences

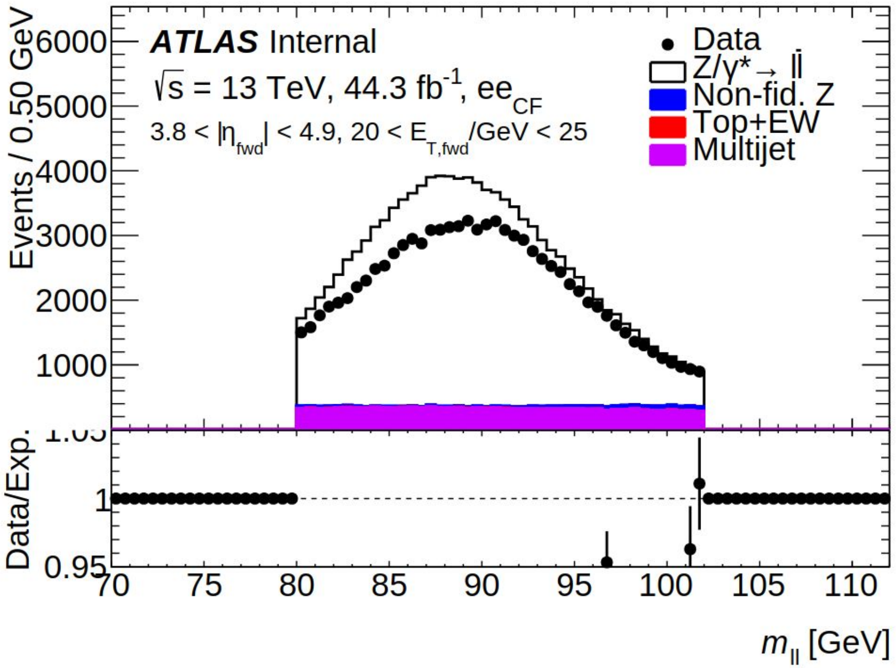
of forward electron ID SFs

relative differences

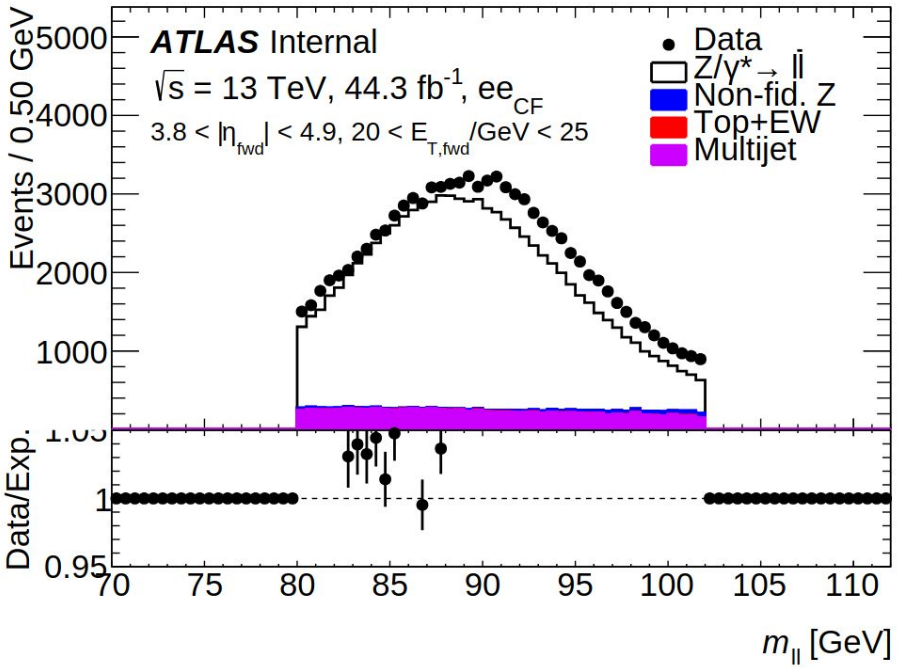


Backup: Control Plots

without SFs

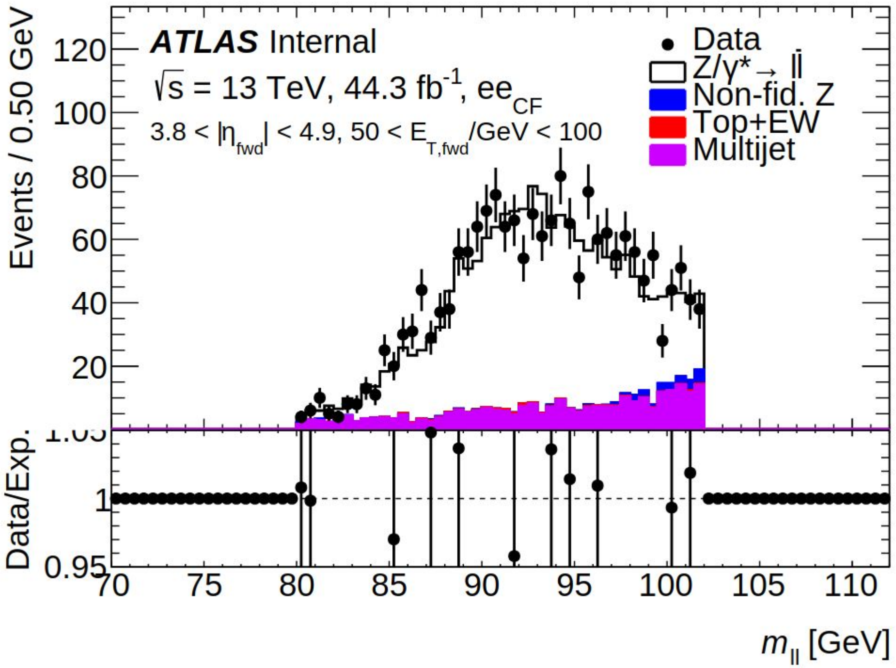


with SFs



Backup: Control Plots

without SFs



with SFs