

1

## **Part I**

2

# **Theory introduction**



<sup>3</sup> Chapter **1**

<sup>4</sup> **Standart model**



<sup>5</sup> Chapter **2**

<sup>6</sup> **Theory of pp collisions**



<sup>7</sup> Chapter **3**

<sup>8</sup> **PDF fits and DLGAP formalism**



<sup>9</sup> Chapter **4**

<sup>10</sup> **Physics of WZ bosons**

<sup>11</sup> **4.1 Theory**

<sup>12</sup> **4.2 Cross-section measurement**



13

## **Part II**

14

# **Experimental setup**



15

# Chapter **5**

16

## **LHC**



<sup>17</sup>

Chapter

**6**

<sup>18</sup>

## **ATLAS experiment**



19

Chapter

**7**

20

## **Software**



## 22 Monte Carlo

23 The Monte Carlo(MC) method was invented by scientists working on the atomic bomb in the 1940s.  
 24 Its core idea is to use random samples of parameters or inputs to explore the behavior of a complex  
 25 system or process. Nowadays, MC are essential part of both theoretical and experimental particle  
 26 physics research. This chapter gives an overview of ATLAS experiment simulation scheme, simulation  
 27 methods and software used. Also, a techniques for fast simulation will be discussed.

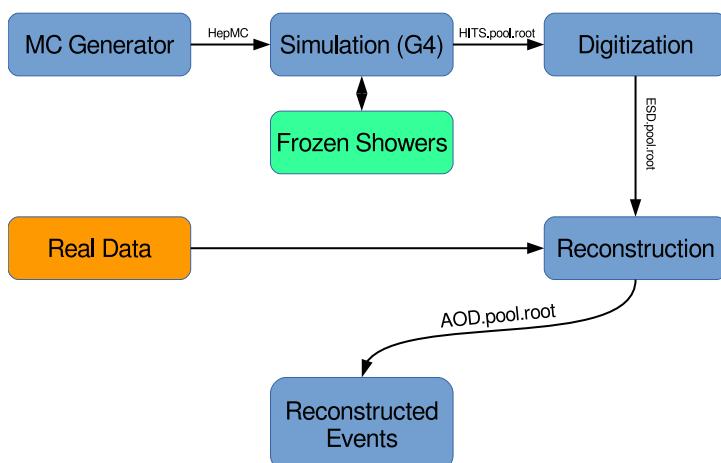
28 **8.1 Monte Carlo simulation at ATLAS experiment**

Fig. 8.1: Diagram of the ATLAS MC production chain

29 MC is allowing to make different analysis, such as compare data with predictions, study detector  
 30 or selection algorithms performance. All of this applications are requiring MC precision. Simulation  
 31 software expects to use precise physics models for sampling and have large enough statistics, to  
 32 exclude statistical uncertainties (usually 5 times more, than expected in a data). ATLAS simulation  
 33 software is integrated into Athena and usually used during large production of events. Simulation  
 34 chain is generally divided into 4 main steps (Figure 8.1):

35 **Event generation** Simulation of hard interaction and a resulting high-energy particles parameters.  
 36 This step is independent of ATLAS detector geometry.

37 **Simulation** Simulation of energy depositions ("hits") done by final state particles in ATLAS detector.

38 **Digitalization** Simulation of detector response using "hits" information: first, inputs to the read out  
 39 drivers (ROD's), called "digits" are constructed, then, ROD functionality is emulated. Detector  
 40 noise effects are added at this stage.

41 **Reconstruction** Production of the Analysis Object Data (AOD) files, which are containing sufficient  
 42 information for physics analysis. This stage is identical for both data and MC

43 This scheme allows to use computing resources more efficiently, than with a single-step simulation,  
 44 and simplifies software validation, since it is possible to reuse files from previous stages. In the  
 45 following sections event generation and simulation will be described in more details

46 **8.1.1 Event generators**

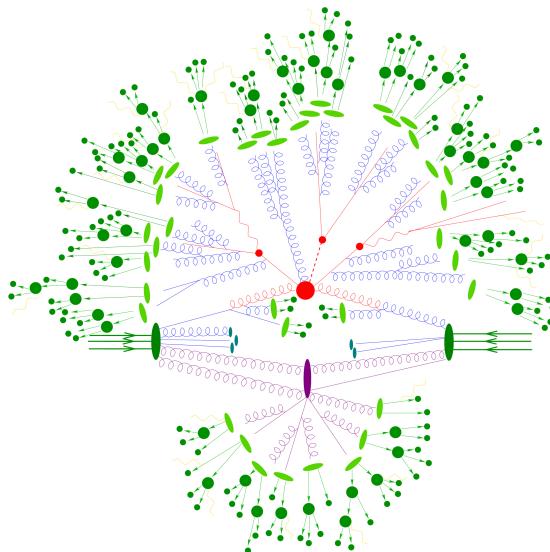


Fig. 8.2: Schematic view of a  $t\bar{t}H$  event produced in a pp-collision: the hard scattering is shown as a red blob with the solid and dashed lines as the resulting three particles. Independently happening multi-particle interactions are indicated by the violet blob. Parton showers are shown with curly lines. Hadronization yields hadrons as shown in light green, while the final state particle are dark green.

47 The outcome of the hard interaction could be simple scattering of the hadron elementary con-  
 48 stituents, their annihilation into new resonances or a combination of two. In any case, the final state  
 49 has a large particles multiplicity. The main goal of event generator is to provide a complete picture of  
 50 this final states: description of the particle types and momentia on event-by-event basis. The fac-  
 51 torisation theorem [?] allows to make event generation in independent stages, which are dominated  
 52 by different dynamics:

53 **Modelling of hard subprocess** Hard subprocess is happening at the smallest scales of times and  
 54 distance, where all of the colliding partons are considered free. Process of interest is simulated  
 55 by selecting production channels and calculating corresponding matrix elements (ME) in the  
 56 desired level of accuracy in perturbation theory . Most of the generators have leading order  
 57 or next to leading order ME in  $\alpha_s$ .

58 **Parton showering** Quarks and gluons from hard process can radiate secondary quarks and gluons,  
59 resulting on the dozens of additional partons associated with the event. This process calculated  
60 as step-by-step evolution of momentum transfer scales from highest (hard subprocess), to  
61 the lowest (around 1 GeV). There is a possibility of double counting between showers and hard  
62 subprocess. This can be avoided using matching approach, for which higher order corrections  
63 to ME are integrated with parton showers, or merging strategy, where jet resolution scale is  
64 used as a threshold between matrix elements and parton showers.

65 **Hadronisation** Final stable color-neutral particles, what can be detected in experiment, are formed  
66 during hadronisation. This occurs at larger nonperturbative scales and usually implemented  
67 using different phenomenological models.

68 **Modelling underlying event** Parallel to the main process other collisions of partons can occur,  
69 called underlying event. These additional interaction can produce partons which contribute to  
70 the final state. This is one of the least understood aspect of hadronic collisions.

71 Schematic plan of simulation of ttbar event is shown in Figure ???. The hard scattering itself is shown  
72 as a red blob with the solid and dashed lines indicating the resulting particles, which themselves decay  
73 further. Underlying event is indicated by the violet blob. Parton showers are shown with curly lines.  
74 Hadronization yields hadrons as shown in light green, while the final state particles are dark green.  
75 The current analysis uses samples generated with the following generators:

76 Powheg [?] Powheg is generator with NLO ME [?], that can be interfaced with other generator(such  
77 as Pythia or Herwig) for higher precision of showering.

78 Pythia [?] Pythia is a general purpose generator for hadronic, hadron-lepton and leptonic collisions.  
79 It can model initial and final state showers, hadronisation and decays, underlying event (via  
80 multi parton interactions). Pythia contains library with around 240 processes with LO ME. It  
81 uses Lund String model [?] for hadronisation.

82 Herwig [?] Herwig is a LO general purpose event generator for simulation lepton-lepton, hadron-  
83 lepton and hadron-hadron collisions. The main difference between Pythia and Herwig is that  
84 it uses angular ordering in the parton showers and also models the hadronisation step based  
85 on the cluster fragmentation

86 Sherpa [?] Sherpa is a generator with tree level of matrix elements, featuring its own implementation  
87 of parton shower and hadronisation models.

88 Photos [?] Precision tool for QED radiative corrections in W and Z decays.

89 Tauola [?] Generator, used to describe leptonic and semi-leptonic  $\tau$ -decays.

### 90 8.1.2 Simulation in Geant4

91 After event generation, simulation software obtains hardware response for final state particles. The  
92 main method used by ATLAS, referred to as *Full Simulation*, makes use of the Geant4 [?]. It is C++  
93 based toolkit for the simulation of the passage of particles through matter. It is used in a wide range  
94 of experiments in high energy and nuclear physics.

95 Geant4 can simulate complex detector structures with sensitive detector material and correspond-  
96 ing infrastructure. It can also calculate basic properties of materials, like radiation and interaction

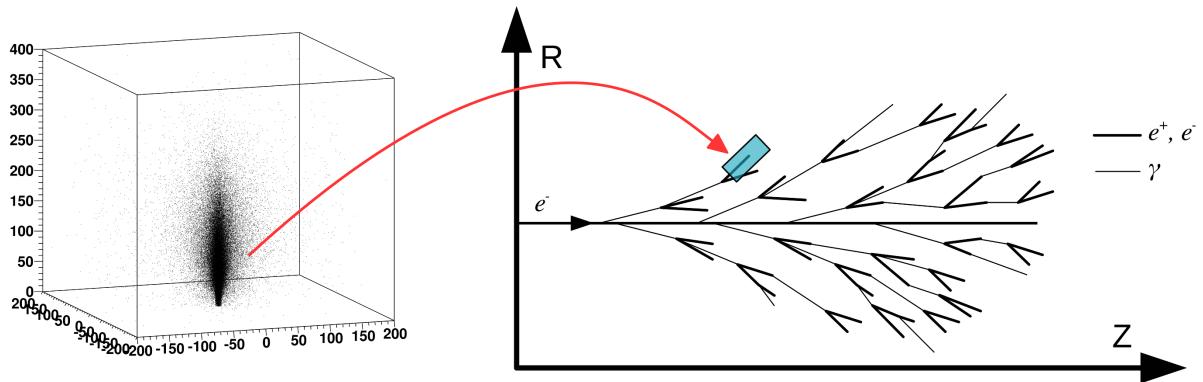


Fig. 8.3: Diagram showing the shower substitution of the low-energy particle, during the high-energy particle simulation.

length. For detector Geant4 stores "hits" information - snapshots of physical interactions. In Geant4 events and particles are simulated separately and each particle is moved step by step. Size of this step is chosen to preserve both CPU performance and required precision. Physics is treated as a set of discrete processes. They could be handled either at rest, along step or after it. Geant4 package has different models and approximations for hadronic and electromagnetic processes. Some of them are not approximate, but computationally fast. It allows to choose set of the models, called physics list, depending on particular requirements. There are several reference physics lists, that are validated for each new release of Geant4 software. ATLAS experiment is using one of this lists.

It is necessary to have mass MC production for each data taking, what is taking most of the resources. Uncertainties of some of Run-I analyses are dominated by available MC statistics.

It is possible to improve in CPU usage by tuning physics list or working on a magnetic field parametrisation. Also there are long term developments for multi-threading and vectorisation of the code. Yet, Run-2 has a higher pileup and luminosity, so even more MC events are needed. This means that fast and accurate simulation approach is essential. During simulation largest time is spent on calorimeters. This is the motivation for development of fast calorimetry techniques.

There are two main methods used in ATLAS:

- Parametrisation of calorimeter cells response. Spacial energy response is simulated using longitudinal and lateral energy profiles.
- Frozen Showers. This technique will be described more detailed in the following section

### 8.1.3 Frozen Showers

Main principle of this method is described by its name. It is using pre-simulated "frozen" showers generated in full simulation and stored in a library. Particles below minimum energy thresholds are killed and replaced with these showers. All of the other particles are simulated using full simulation. This process is schematically shown in a Figure 8.3.

The library itself organized as follows: the header contains basic simulation parameters, like Geant4, geometry and ATLAS software release version and physics list used. Showers are stored in a bins of positional variables (see sec. 8.1.3), while energy remain unbinned. Each shower stores lateral and transverse size and information about energy, time and positions of the hits.

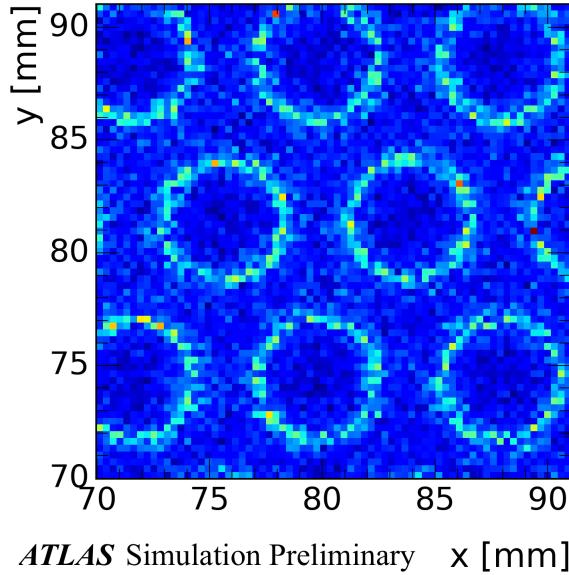


Fig. 8.4: Average energy response in a calorimeter vs x,y for electrons with energy less than 1 GeV

### 125 Production use of Frozen Showers

126 During simulation, if an energy of a particle falls below cut-off energy, the particle algorithm examines  
 127 resulting shower containment. It checks that particle is far from the edges of calorimeter, so what  
 128 shower will be by 90% inside calorimeter. This depends also on a energy of particle, because  
 129 shower sizes are growing with energy. When particle is removed and substituted by shower taken  
 130 from corresponding eta and distance bin with the closest energy found. Energies of the hits in shower  
 131 found are scaled to fully correspond to particle energy. Additionally, shower direction is changed to  
 132 the direction of the particle.

Frozen Showers have been used in ATLAS Monte-Carlo production since run-1. This method is applicable for all LAr calorimeters in ATLAS, but currently it is enabled for simulation of forward calorimeters (FCAL), since it is showing the smallest differences, compared to the other fast simulation methods (e.g parametrisation). This is because of large of non-uniformly distributed sensitive material, which is giving different responce, than a dead material (Figure 8.5). Resolution of a calorimeter can be written as:

$$\frac{\sigma}{E} \approx \frac{1}{\sqrt{E}} \oplus \frac{1}{E} \oplus const, \quad (8.1)$$

where symbol  $\oplus$  indicates a quadratic sum. The first term is 'stochastic term', which includes intrinsic shower fluctuations, second takes into account readout noise effects and pile-up fluctuations. Constant term derives from non-uniformities in a detector, what are causing large fluctuation of the energy loss. Resolution of high-energy electrons is mostly dominated by this term. Example of these fluctuations is shown on a Figure 8.4. Circles are corresponding to a LAr gaps inside FCAL. It can be seen, that particles inside sensitive material are having more energetic showers, than particles in a dead material. It is possible to capture this structure by introducing distance to a closest rod center:

$$D = \min(\sqrt{dx^2 + dy^2}, \sqrt{(step_x - dx)^2 + dy^2}, \sqrt{(step_x/2 - dx)^2 - (step_y - dy)^2}), \quad (8.2)$$

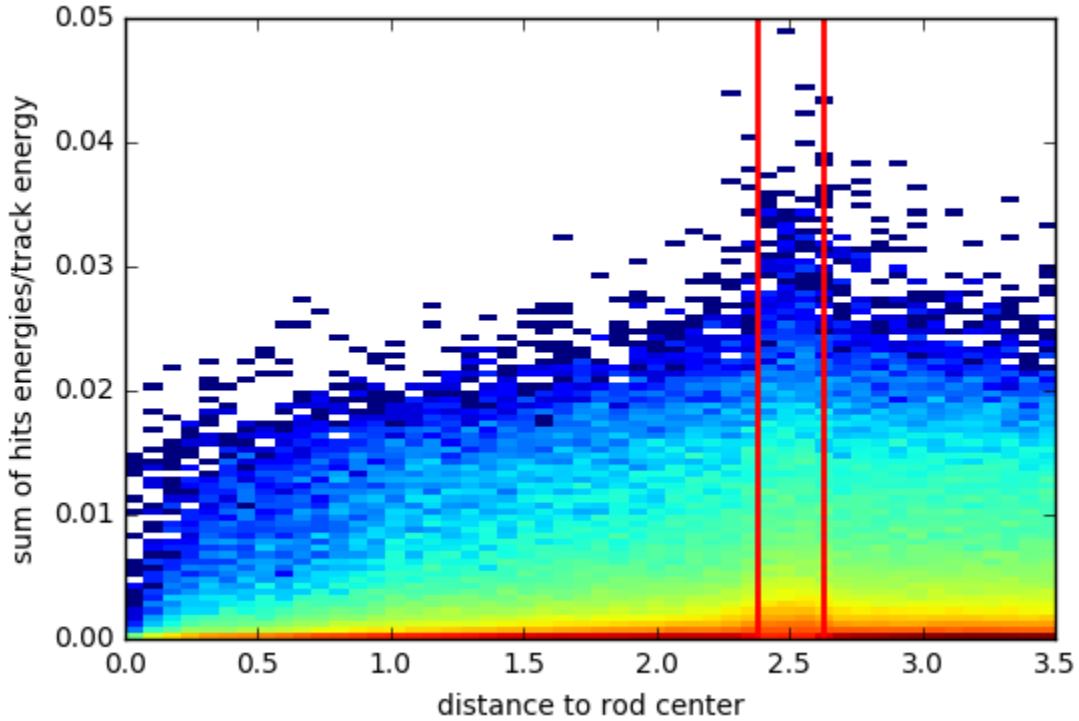


Fig. 8.5: Energy response for electrons in a calorimeter for all electrons in a library

where  $dx$  and  $dy$  are the distances to a rod center in a  $x$  and  $y$  plane respectively. They are calculated as:

$$dy = y - n[] \quad (8.3)$$

Dependency of summed energy of shower on the distance defined above is shown on a Figure 8.5. Gap is marked by the red lines. Size of this differences between sensitive material and dead material depends on a initial particle momentia (Figure 8.6 a and b). For electrons with energy greater than 500 GeV they are almost negligible. Additionally, at higher energies gain in a CPU time is moderate, while library becomes bigger. This is reason for an upper limit to be set at 1000 MeV

Performance of frozen showers is also depending on a lower limit of a method. Distribution of shower energies, used for production of high-energetic electron (1000 GeV in that case), is shown on a Figure . More than 50% of them are having energy less than 20 MeV. Studies have showed, that Frozen showers are slower, than a standard Geant4 simulation for showers with energy 3 MeV. This is happening due to library non-binned structure for energy. This makes search of closest energy shower in a library slower, than simulation of shower with zero or one hit in a sensitive detector.

#### 144 **Generation of Frozen Showers library**

145 In a Frozen Shower method there are separate libraries for each particle and subdetector used.  
146 Showers should cover fully energy and pseudorapidity region and be able to describe data, that is

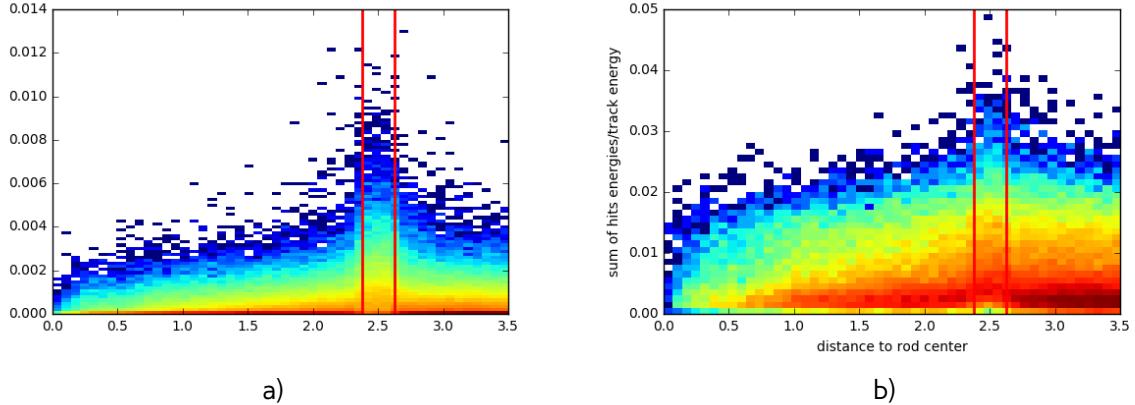


Fig. 8.6: Energy response for electrons in a calorimeter for a) electrons with energy less than 100 MeV b) electrons with energy bigger, than 100 MeV

The general frozen showers parameters	
Detectors used	FCAL1, FCAL2, FCAL3
Type of the particle	photons, electrons, neutrons
Energy range	$E_\gamma < 10 \text{ MeV}$ , $E_e < 1000 \text{ MeV}$ , $T_n < 100 \text{ MeV}$
Containment requirement	$\Delta E_{\text{shower}} > 98\%$
The library post-processing parameters	
Generation clustering cutoff	$(\Delta R_{\text{cluster}})^2 < 25 \text{ mm}$
Generation truncation cutoff	$R_{\text{hit}}^2 < 50000 \text{ mm}$ , $\Delta E_{\text{shower}} < 1\%$

Table 8.1: Main parameters used for the frozen shower libraries in FCAL

needed during simulation. This is why 2 stages simulation approach have been used. The first stage takes initial particle parameters from a physical processes (ttbar or a single electron).

The first stage is to take initial particle parameters, that later will be used in a library from a physical process. This is done using simulation of some process (e.g. ttbar or single electron). Every time, when particle becomes eligible for Frozen Showers, it parameters are saved in a HepMC format. Particles inside calorimeter tend to cluster tightly around initial track, so random truncation of initial particles is used to obtain better detector coverage. On the second stage, this primary particles are propagated through the calorimeter using standart Atlas simulation infrastructure. Resulted shower parameters are saved in a library. This procedure allows to take into account sampling fluctuations and charge-collection effects on a hit information automatically. Additionally, in order to save disc space as well as a memory consumption, hit information is compressed. This compression is done in a two steps, hit merging and truncation:

- if the distance between any two hits is smaller, than a given parameter  $R_{\min}$ , then hits are merged into one deposit at the energy weighted center of them. This process is done iteratively.
- hits whose energies are below the fraction  $f$  of the total energy sum of all hits, are truncated. The energy of remaining hits is rescaled back to preserve the total deposited energy.

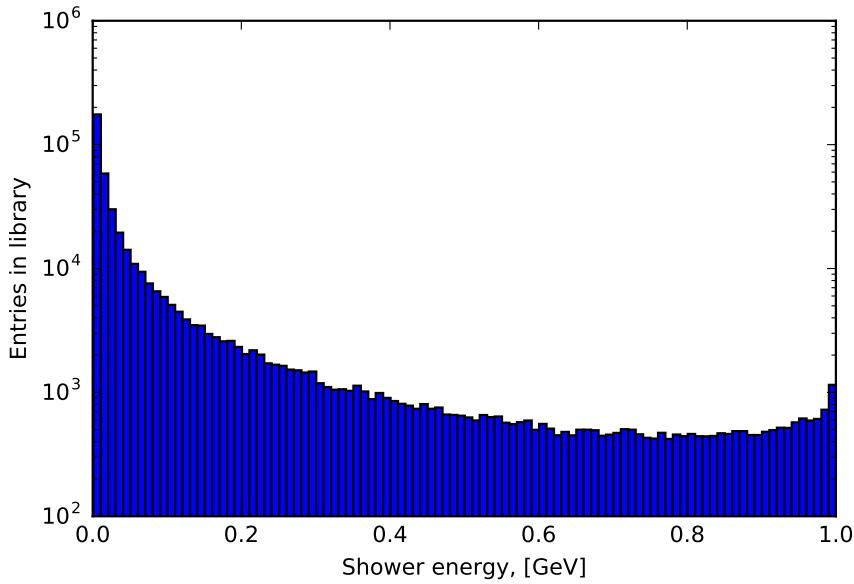


Fig. 8.7: Distribution of shower energy used in production of 1000 GeV electrons.

164 Unfortunately, for a Frozen Showers, generated for Run-1 monte-carlo, additional tuning of electron  
 165 libraries was needed. This was done using reconstructed energy of electrons. Frozen Showers tend  
 166 to underestimate fluctuations of energy loss, that is leading to a smaller electron resolution for a  
 167 high energies. Correction is done by enlarging bin, corresponding to a gap position. Also, correction  
 168 of the mean shift is done by scaling energy response of all showers. After this frozen showers are  
 169 showing good agreement with full simulation. This procedure needs to be done every time, when  
 170 something is changing in software. Because tuning is done manually, lots of manpower is needed  
 171 for each Monte Carlo mass production campaign.

## 172 Distance binning problem

173 As it was mentioned before, process of library generation can be complicated and take a lot of the  
 174 time because of the needed tuning. In this subchapter possible ways to improve frozen showers  
 175 performance have been studied.

176 As it was mentioned before, that there are two type of material used in a FCAL. Showers within  
 177 them are giving different response, what is affecting overall reconstructed electron energy resolution.  
 178 At the first generations distance bin have been corresponding to LAr gap or dead material positions.  
 179 During tuning bin with LAr was enlarged to gain a better agreement with full simulation. So, one of  
 180 the basic ideas to improve frozen showers performance is to change a size of LAr gap in a library  
 181 generation.

182 It was decided to treat showers, that have been born near LAr gap and crossed it on a radiation  
 183 length, in a same way with showers in sensitive material gap, and call them sensitive material showers.  
 184 Oppositely, showers, that haven't crossed LAr gap, are called dead showers. This model leads to a  
 185 bigger gap width by a definition. One of the possible ways to find this bin position automatically is  
 186 to use machine learning tools.

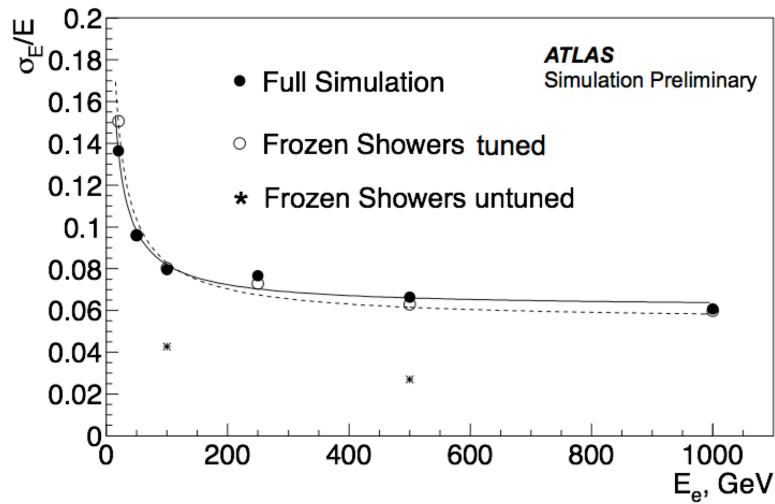


Fig. 8.8: Electron resolution for full simulation, tuned and untuned frozen showers

187 Machine Learning is a set of algorithms, what allows computers to learn and give a predictions  
 188 without being specifically programmed. This is a modern field of computer science, that is wildly  
 189 used in a different fields like computer vision, natural language processing, data science etc. There  
 190 are two main types of machine learning algorithms: supervised, where example of desired output  
 191 is given by the "teacher" and the goal is to learn a general rule, that maps inputs to outputs and  
 192 unsupervised learning, then there are no labels given to algorithm, and algorithms is discovering  
 193 hidden patterns in data. Initial data parameters of interest, that are used in algorithm to learn are  
 194 called features. It is important to have right proper set of features and good training sample.

195 From a geometrical point of view, one of the main parameter is a direction of the shower. Eta  
 196 momentum distribution is showed on a Figure 8.9 . Most of the showers are collinear to an electron  
 197 direction. Because of this it was decided to use as a training sample simulation results for electrons  
 198 with energies less than 1 GeV and momentum uniformly distributed between eta 3.0 and 4.0. This  
 199 allowing to study equally low and high energy showers equally.

200 From our definition of 2 classes of showers, it is simple to construct a pre-labelled training sample.  
 201 This is done by reducing initial sample and taking showers near rod center and inside liquid argon  
 202 gap. Output of this classifier, that was trained on with sample with shower features, such as energy  
 203 response and number of hits, than can be used to expand our labels to a full distance range. Then  
 204 it can be used as an input to a second classifier, which will separate two types of showers using  
 205 particle parameters, such as energy and distance to a rod center. For a first step decision trees have  
 206 showed good classification efficiency (around 97%). For a second classifier support vector machines  
 207 have been used. This method is trying to reconstruct a hyperplane, that is dividing two classes.  
 208 Outputs of both of this classifiers are shown on a figure . New gap position is determined using  
 209 borders of hyperplane. This procedure is giving expected from the initial model results. Gap is wider,  
 210 than and original one. It is also getting bigger with bigger energy, because of the radiation length  
 211 growth. Validation results for two different eta bins are shown on figure a) and b). In a bin this new  
 212 binning is performing better, than original one without any additional tuning. Unfortunately this is not  
 213 true for all of the bins, as we can see on a figure b). This eta bin have showed worst performance  
 214 for a new binning, but it is performing still better, than original binning without tuning.

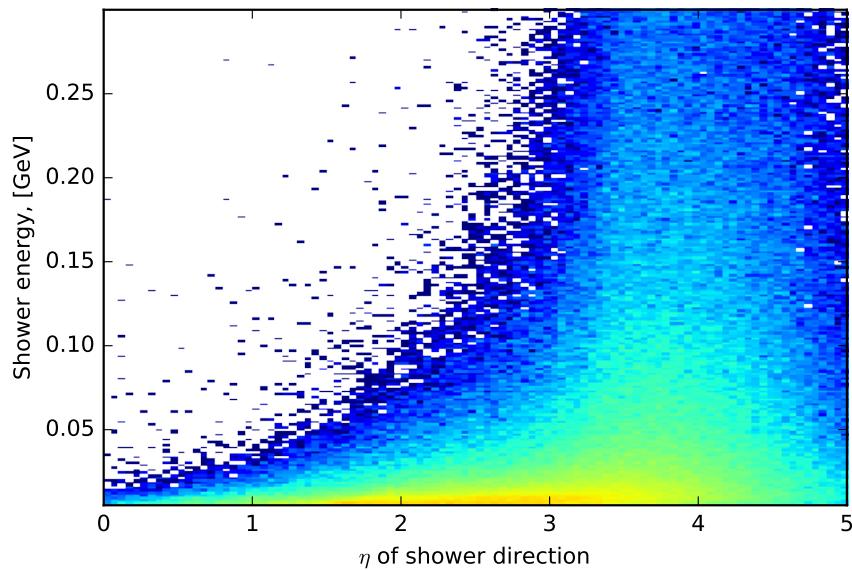


Fig. 8.9: Distribution of showers used in production of 1000 GeV electrons on shower energy vs  $\eta_{momentum}$  plane.

215     This binning was used in a production of new libraries for Monte Carlo in a Run-2. It is planned to  
216     use more precise training sample for a future iterations of this procedure for improving performance  
217     of outlying eta bins.

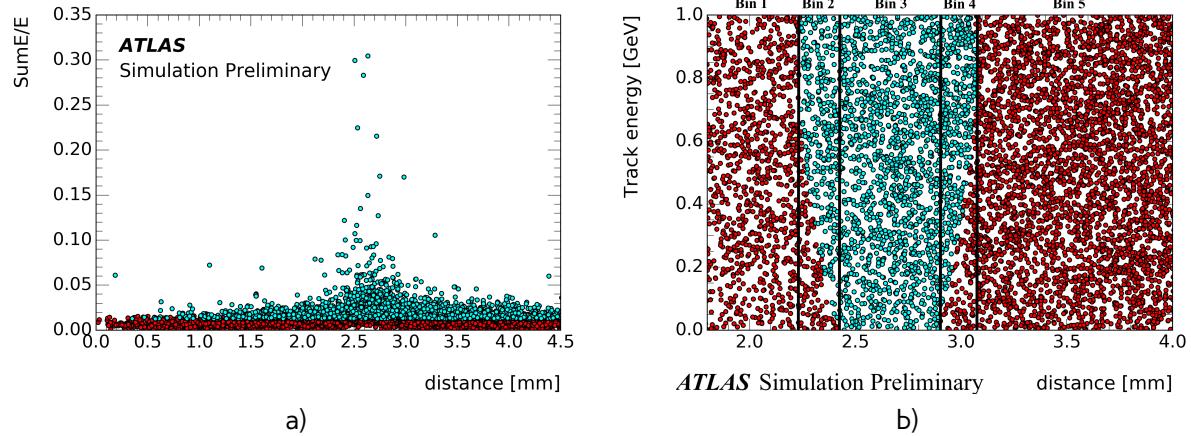


Fig. 8.10: Results of machine learning for a) first classifier b) second classifier. Cyan dots are corresponding to sensitive material showers, red - dead material showers

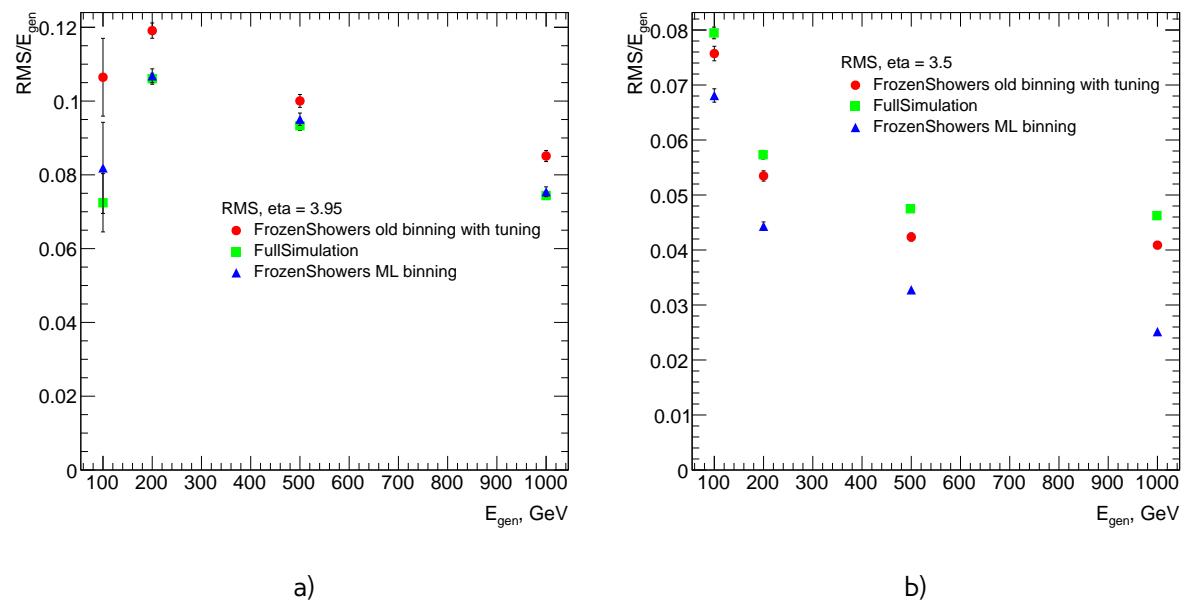


Fig. 8.11: Resolution of reconstructed electrons for full simulation, new libraries with ML binning and old tuned libraries with original binning for a) eta = 3.95 b) eta = 3.5



218

# Chapter 9

219

## DataSample



220

## **Part III**

221

# **Measurement of cross-section**



222 Chapter **10**

223 **Selection**

224 Selection criteria is the set of requirements, that is applied both on data and MC. Analysis is de-  
225 pending on a selection, that can separate process of interest (signal) from other processes. For  
226  $pp \rightarrow W \rightarrow ev/\mu\nu$  and  $pp \rightarrow Z/\gamma^* \rightarrow ee/\mu\mu$  selection criteria can be divided into 3 groups: data  
227 quality, lepton and boson cuts. In this chapter all of them will be discussed and a cut flow presented  
228 In this chapter selection criteria for  $pp \rightarrow W \rightarrow ev/\mu\nu$  and  $pp \rightarrow Z/\gamma^* \rightarrow ee/\mu\mu$  are presented.

229 **10.1 Data quality cuts**

Table 10.1: Analysis selection

Event selection	
Single lepton trigger	
Good Run List	
Reject events with LAr errors	
Number of tracks at primary vertex $\geq 3$	
Electron Selection	
$P_T > 20\text{GeV}$	$P_T > 20\text{GeV}$
$ \eta  < 2.47$	$ \eta  < 2.5$
excluding $1.37 <  \eta  < 1.52$	
OQ cut	staco reconstruction chain
Medium electron identification	Medium muon identification
$\text{PtCone}20 < 0.1$	$\text{PtCone}20 < 0.1$
W boson selection	
$\text{EtMiss} > 25 \text{ GeV}$	
$M_T > 45 \text{ GeV}$	$66 < M_{ee} < 116 \text{ GeV}$
Z boson selection	

230 For a measurement we must use the data with a proper quality. Unfortunately not all of the events  
231 satisfy this criteria. One of the possible source of the problems could be that LHC was not in a  
232 stable beam mode, or parts of the detector have been switched off, or event had too many noisy  
233 cells. The information about luminosity blocks, that need to be excluded is stored in a "Good Run  
234 List". Events, where LAr calorimeter was malfunctioning are excluded by LAr quality criteria. Events

235 are furthermore required to have at least one primary vertex from a hard scattering with at least 2  
236 associated tracks reconstructed.

237 **10.2 Lepton quality cuts**

238 Online selection of events is based on a single lepton trigger, depending on a flavor of analysis.  
239 For electron analysis it is required to have EF\_e15\_loose1 trigger, which records electrons with  $E_T >$   
240  $7\text{GeV}$ . This trigger is also using additional "loose" isolation requirements to exclude jets, that are  
241 misidentified as electrons. In muon channel lowest single lepton trigger available used (EF\_mu10). It  
242 records events with muons  $E_T > 10\text{GeV}$ . Moreover, matching between trigger and lepton is required.  
243 All of the analysis are using similar selection criteria, applied on a leptons. All of the leptons must  
244 satisfy requirement  $P_T > 20\text{GeV}$  Electron candidates are required to be within pseudorapidity range  
245  $|\eta| < 2.47$ . Candidates within the transition region between the barrel and endcap electromagnetic  
246 calorimeters,  $1.37 < |\eta| < 1.52$ , are removed. Additionally, for better multijet background rejection  
247 medium identification and PtCone20  $< 0.1$  criterias are applied.

248 Muons are satisfying following criteria: they should be reconstructed by a staco algorithm in a  
249 muon spectrometer and ... within range  $|\eta| < 2.5$  . Set of medium requirements is applied. They  
250 must also satisfy PtCone20  $< 0.1$  isolation criteria

251 **10.3 Boson selection**

252 Events, contained W boson are required to have exactly one selected lepton. Events, where there are  
253 additional "good" leptons are rejected. Missing transverse energy is required to be  $E_{T\text{Miss}} > 25\text{GeV}$ .  
254 W boson, formed out of etMiss and lepton should have transverse mass  $M_T > 45\text{GeV}$ . After the  
255 full selection total number of events in electron channel is ..(.. and .. for  $e^+$  and  $e^-$  respectively.

256 The reconstructed lepton pair in case of Z boson analysis is required to invariant mass between  
257 66 and 116 GeV. Both upper and bottom limits allows to exclude regions with high background  
258 contamination and low statistics.

259 Full set of cuts is summarized in a table 10.1.

260

# Chapter 11

## 261 Monte Carlo corrections

262 Monte Carlo plays important role in cross-section measurement. It is constantly undergoing correction  
 263 to data, in order to obtain a required precision. Part of this corrections have been described in  
 264 a chapter 8. Unfortunately, not everything can be taken into account during simulation itself. This  
 265 leads to a differences between data and monte carlo, that needs to be accounted for. There are two  
 266 possible methods to correct monte carlo without regenerating it. First on is to apply event weight, so  
 267 what each mc event can contribute to non 1 entries in a histogram. This is called reweighting. Second  
 268 one is to smear MC. It is using random number to alter reconstructed 4-vectors. This chapter de-  
 269 scribes all additional corrections, what have been applied on MC in this analysis. All of this correction  
 270 are introducing additional systematic error, that will be discussed in the chapter ??

### 271 11.1 Lepton efficiency corrections

272 Lepton detection efficiency at ATLAS detector can be divided into three components:

- 273 • The reconstruction efficiency  $\epsilon_{rec}$  is a probability to reconstruct lepton as a lepton of this  
 flavor.
- 275 • The identification efficiency  $\epsilon_{id|rec}$  is the probability that a reconstructed lepton survives iden-  
 tification requirements.
- 277 • The trigger efficiency  $\epsilon_{trig|rec,id}$  is the probability, that lepton satisfy trigger requirements.

The full efficiency for a single lepton can be written as:

$$\epsilon_{total} = \epsilon_{rec} \times \epsilon_{id|rec} \times \epsilon_{trig|rec,id} \quad (11.1)$$

278 All of this efficiencies are measured using Tag and Probe method in  $Z \rightarrow ll$  decays. This is allowing  
 279 to insure, that all of the reconstructed lepton candidates are coming from an actual leptons. One of  
 280 the leptons from Z boson, called "probe", is initially selected with all of the cuts, minus one under  
 281 study. Second one, called "probe" satisfies more tighter selection with additional cut, such as, for  
 282 example, trigger matching.

Reconstruction efficiency is assosiated with algorithm used to perform reconstruction. This is causing difference between electrons and muons efficiencies. In electron case it is a probability to reconstruct an elec tron with an electromagnetic calorimeter as an electron. Muon reconstruction efficiency is given by:

$$\epsilon_{reco,muon} = \epsilon_{reco,muon|ID} \cdot \epsilon_{ID} \approx \epsilon_{reco,muon|ID} \cdot \epsilon_{ID|MS}, \quad (11.2)$$

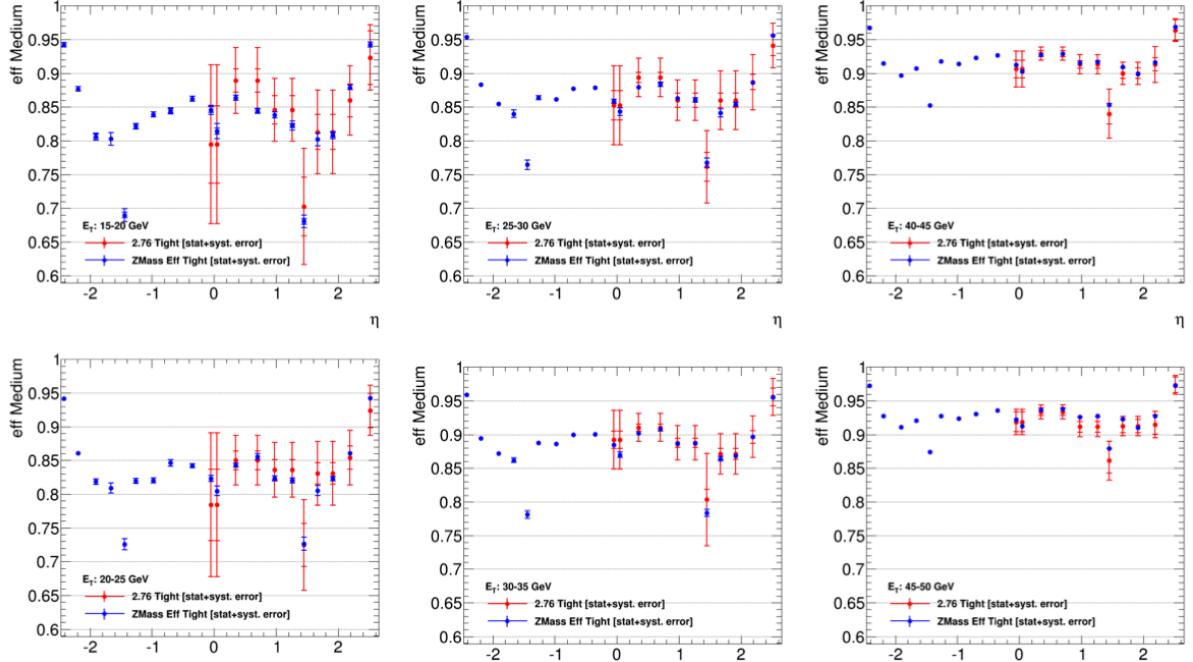


Fig. 11.1: Comparison of electron efficiencies as calculated for 8TeV (blue points) and 2.76TeV (red points) for MC simulation. Efficiencies are shown as a function of pseudorapidity ( $\eta$ ) for different electron  $E_T$  bins. Both statistical and systematic uncertainties are shown.

where  $\epsilon_{reco,muon|ID}$  is a conditional probability that muon reconstructed in ID is also reconstructed using MS as a combined muon, and  $\epsilon_{ID}$  is a probability that muon is reconstructed as an ID track. This quantity cannot be measured directly and therefore is replaced by  $\epsilon_{ID|MS}$ , that can be measured by tag-and-probe method. uncertainty in this analysis.

Simulation samples are corrected to match data efficiencies by a scale-factor :

$$SF_{reco,id,trig} = \frac{\epsilon_{reco,id,trig}^{data}}{\epsilon_{reco,id,trig}^{MC}} \quad (11.3)$$

Each of the scale factors calculated in a  $p_t$  and  $\eta$  bins and has an associated statistical and systematical uncertainty component. Statistical component is connected to a size of  $Z \rightarrow ll$ , which is in our case is around 500 event per each lepton flavor. This means that precise calculation of scaling factors based on this data is difficult.

It is possible to use scale factors for 8 TeV 2012 data. The main difference between this data samples are center of mass energy and a pile-up conditions (10 in 2012 and less than 1 in 2013). This effects have been studied on a  $Z \rightarrow ee$  sample. Fig. 11.1 shows that all of the differences in a scale factors are negligible and fully covered by the statistical error. This justifies the usage of 8 TeV scaling factors with increased

## 11.2 Muon Trigger SF

Unfortunately, single muon trigger haven't been presented in a 2012 data, so muon trigger scale factor needed to be derived from a 2.76 TeV data. Size of the Z sample is not enough to make scale

299 factors both in  $P_T^\mu$  and  $\eta$  bins.

300 Since the  $P_T^\mu$  cut is significantly higher, than trigger threshold,  $P_T^\mu$  dependency on the efficiency  
 301 can be considered flat. On another hand,  $\eta$  dependence of the scale factor can give a significant  
 302 difference. Binning in on  $\eta$  is chosen from a detector point of view. And there it goes. Muon trajectory  
 303 is bend in the different directions in a magnetic field, that can lead to a small differences in a trigger  
 304 efficiencies. Possible charge dependency of the scale factors have been also studied.

305 Trigger efficiencies for data and MC in  $\eta$  bins are shown on a Fig. 12.10. Total scale factors are  
 306 presented in a Tab. 11.1. Scale factors for  $\mu^+$  and  $\mu^-$  are more, than  $3\sigma$  away from each other, that  
 307 is a clear sign of a charge dependency.

308 Effect of applying scale factors on

Table 11.1: Muon trigger scale factors

	SF	SF stat.error
$\mu$	0.988	0.011
$\mu^+$	1.012	0.015
$\mu^-$	0.964	0.015

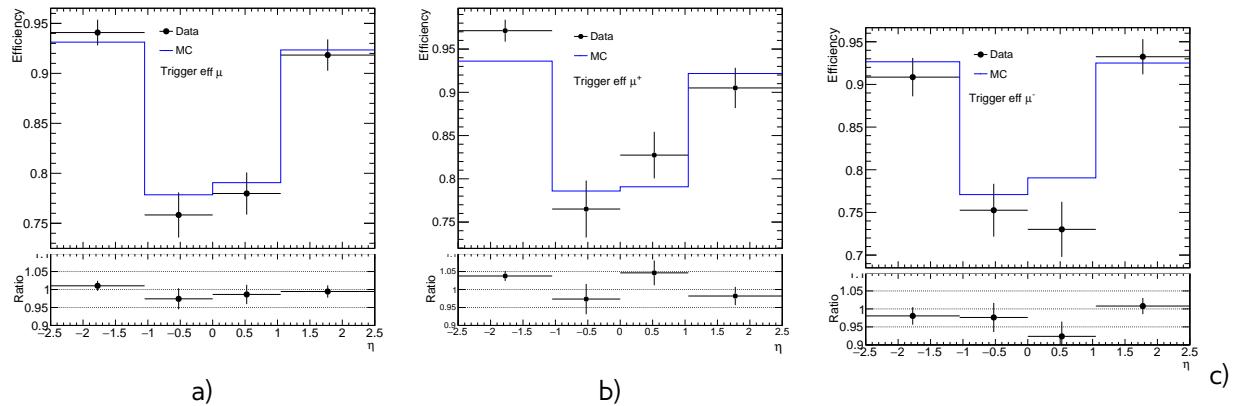


Fig. 11.2: Trigger scale factors for a)  $\mu$  b)  $\mu^+$  c)  $\mu^-$

### 309 11.3 Electron energy scale and resolution

310 Electrons clusters tend to shift in a reconstructed energy compared to a truth energy of initial  
 311 electron. Correction of this shift is done on a both data and MC as a 3 step process:

- 312 • Electronic calibration, that transfers a raw signal from a readout to a cluster energy deposit.
- 313 • MC based calibration. It corrects effects of energy loss in the material in front of calorimeter  
 314 and leakage into the hadronic calorimeter. This calibration is applied on both data and MC.
- 315 • Correction of calorimeter cell response in data. This is allowing to get right response in non-  
 316 optimal HV-regions and exclude biases in a calorimeter electronics reconstruction.

Energy shift is parameterised, as:

$$E^{data} = E^{MC}(1 + \alpha_i), \quad (11.4)$$

where  $E^{data}$  and  $E^{MC}$  are the energies in data and simulation, respectively and  $\alpha_i$  is a mean shift in a given bin  $i$  in  $\eta$ . Effect of this miscalibration on a reconstructed mass of Z boson is:

$$m_{i,j}^{data} = m_{i,j}^{MC}(1 + \alpha_{i,j}), \quad \alpha_{i,j} \sim \frac{\alpha_i + \alpha_j}{2} \quad (11.5)$$

317 neglecting second order terms.  $m_{i,j}^{data}$  and  $m_{i,j}^{MC}$  are reconstructed mass of Z boson in a  $i$  and  $j$  bins  
318 of  $\eta$  for data and MC respectively.

There is also a need to correct difference in a electron resolution. It can be described by a formula 8.1. It is assumed, that sampling and noise terms are modeled well by MC and the main difference is coming from a constant term. So, the electron resolution correction then can be written as:

$$\frac{\sigma_E}{E_i}^{Data} = \frac{\sigma_E}{E_i}^{MC} \oplus c_i \quad (11.6)$$

319 where  $c_i$  is  $\eta$  dependent relative resolution correction. Similarly to a energy scale correction it is  
320 possible to derive resolution correction factor by a comparing  $m_{i,j}^{data}$  and  $m_{i,j}^{MC}$  distribution.

321 Correction values of  $\alpha_i$  and  $c_i$  are obtained via  $\chi^2$  fit on a invariant mass electrons for data and  
322 MC. Resulting energy scale is applied on a data, while resolution is corrected for MC. The resulting  
323 scale is validated on a  $J/\psi \rightarrow ee$  and  $Z \rightarrow ee\gamma$

## 324 11.4 Muon momentum correction

Muon momentum resolution is depending on a  $\eta$ ,  $\phi$  and  $p_T$  of the muon [?]. There is an empirical formula to describe it inside the detector (ID or MS):

$$\frac{\sigma_{Det}(p_T)}{p_T} = \frac{r_0^{Det}(\eta, \phi)}{p_T} \oplus r_1^{Det}(\eta, \phi) \oplus r_2^{Det}(\eta, \phi) \cdot p_T \quad (11.7)$$

325 The first term origins from fluctuations of energy loss in transversed material. Second  $r_1^{Det}$  is com-  
326 ing from magnetic field inhomogeneities and local displacements. Third term  $r_2^{Det}$  describes intrinsic  
327 resolution effects.

Similarly to electrons, overall energy scale shift between data and MC parameterised as:

$$p_T^{data} = p_T^{MC} + s_0^{Det}(\eta, \phi) + s_1^{Det}(\eta, \phi) \cdot p_T^{MC}, \quad (11.8)$$

328 where  $s_0^{Det}(\eta, \phi)$  is coming from the imperfect knowledge of energy losses for muons passing through  
329 detector.

This leads to a total correction formula:

$$p_T^{Cor, Det} = \frac{p_T^{MC, Det} + \sum_{n=0}^1 s_n^{Det}(\eta, \phi)(p_T^{MC, Det})^n}{1 + \sum_{m=0}^2 \Delta r_m^{Det}(\eta, \phi)(p_T^{MC, Det})^{m-1} g_m}, \quad (11.9)$$

330 where  $g_m$  are normally distributed random variables with mean 0 and width 1. Because small amount  
 331 of material between interaction point and the ID,  $\Delta r_0^{ID}(\eta, \phi)$  and  $s_0^{ID}(\eta, \phi)$  are set to 0. Missalignment  
 332 effect for an MS is corrected on a simulation level by adding a random smearing to an alignment  
 333 constants. This is allowing to set  $\Delta r_2^{MS}(\eta, \phi)$  to 0 during a fit.

The correction factors are extracted using  $Z \rightarrow \mu\mu$  candidates events with requirement on a two combined muons. For correction invariant mass distribution  $m_{\mu\mu}^I D$  and  $m_{\mu\mu}^{MS}$  are considered individually within a specific  $\eta - \phi$  region of fit. Combined muon parameters are used to obtain angles  $\eta, \phi$ . The correction extraction is performed first for an ID and then for MS with addition of the fit variable:

$$\rho = \frac{p_T^{MS} - p_T^{ID}}{p_T^{ID}}, \quad (11.10)$$

334 which represents  $p_T$  imbalance in ID and MS.

In a second step corrections are propagated to the combined momentum, using a weight average:

$$p_T^{Cor,CB} = f \cdot p_T^{Cor,ID} + (1 - f) \cdot p_T^{Cor,MS}, \quad (11.11)$$

335 where weight  $f$  is derived from mc.



336

# Chapter 12

## Missing Transverse Energy reconstruction and correction

339 Atlas detector has almost  $4\pi$  coverage. This is allowing to calculate imbalance of energies inside  
 340 calorimeter, especially transversal part of it called  $E_T^{miss}$ . Neutrino from a  $W \rightarrow l\nu$  decay is leaving  
 341 detector, without interacting with it, that is causing large energy imbalance in a detector.

Standard reconstruction of  $E_T^{miss}$  at ATLAS experiment uses transverse energy deposits in the calorimeter, energy losses in cryostat and reconstructed muons for a calculation:

$$E_{x(y)}^{miss} = E_{x(y)}^{miss,calo} + E_{x(y)}^{miss,cryo} + E_{x(y)}^{miss,muon}. \quad (12.1)$$

Calorimeter term is using information from reconstructed physics objects for calibration of cell response. The total transverse energy in calorimeter is defined as:

$$E_{x(y)}^{miss} = E_{x(y)}^{miss,e} + E_{x(y)}^{miss,\gamma} + E_{x(y)}^{miss,\tau} + E_{x(y)}^{miss,jets} + E_{x(y)}^{miss,SoftTerm} + E_{x(y)}^{miss,\mu}. \quad (12.2)$$

342 where each term is calculated as the negative sum of the calibrated reconstructed objects, projected  
 343 onto the x and y directions. Each jet with energy  $P_T > 20$  GeV is corrected for a pile-up and a jet  
 344 energy scale is applied. Soft term is calculated from topoclusters and tracks, that are not associated  
 345 with high-pt objects. To avoid double counting muon energy loss in calorimeter is subtracted  
 346 from  $E_T^{miss}$ . The  $E_T^{miss}$  muon term is calculated from the momenta of muons measured in a range  
 347 of pseudorapidity. Since pileup gives a significant effect on a  $E_T^{miss}$  performance several methods of  
 348 pileup suppression are used. This procedure was optimised for 8 TeV runs and using a calibration  
 349 constants from it. This can cause problems with 2.76 TeV low pileup run. As a examine showed this  
 350 is not optimal procedure in this case. Control plots for W production in electron and muon channels  
 351 are shown on a Fig. 12.3. Where a big discrepancies in a muon and electron channel, that cannot be  
 352 accounted to multijet background.

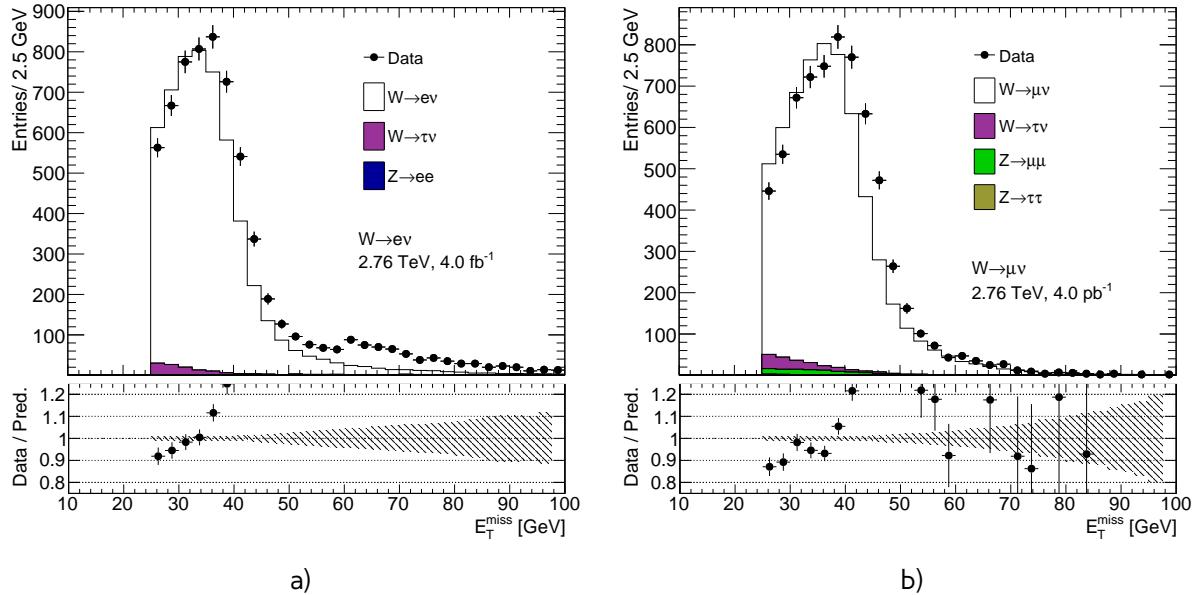


Fig. 12.1: Data and MC comparison for  $E_T^{\text{miss}}$  calculated by standard ATLAS algorithm for a)  $W \rightarrow e\nu$  b)  $W \rightarrow \mu\nu$  events

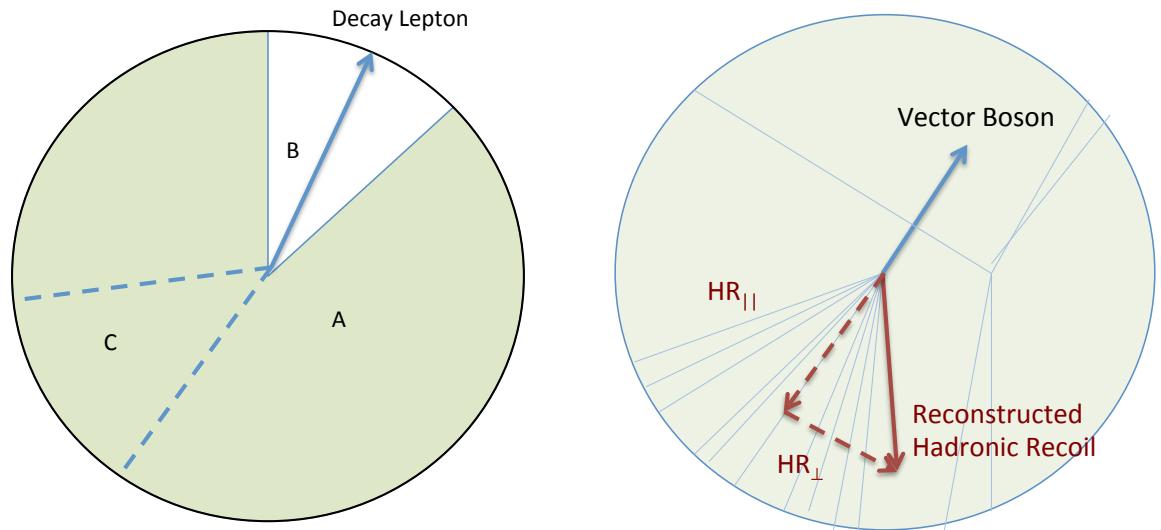


Fig. 12.2: a) Definition of different zones in the calculation of the cluster-based hadronic recoil. , b) Parallel and perpendicular projection of the hadronic recoil with the respect to the transverse momentum of the vector boson

---

## 353 12.1 Hadron Recoil Calculation

Second way of calculating  $E_T^{miss}$  was developed specifically for a W and Z decays by W mass measurements group. This procedure is using this fact, that a transverse momentum of a W-boson has to be balanced with initial (quark-gluon) state radiation, because initial sum of transverse momentum is zero:

$$\vec{P}_T^W = \vec{P}_T^{lep} + \vec{P}_T^\nu = \sum \vec{P}_T^{ISRquarks,gluon}, \quad (12.3)$$

where  $\sum \vec{P}_T^{ISRquarks,gluon}$  is a transverse momentum of partons from initial state radiation, also called hadronic recoil (HR). Therefore,  $E_T^{miss}$  can be determined as:

$$E_T^{miss} = P_T^\nu = -HR + p_T^l \quad (12.4)$$

This procedure assumes, that recoil is arises from one single leading jet, and the rest is coming from a soft hadronic activity. This hadron recoil is computed as a vector sum of calorimeter clusters:

$$HR = \sum_{i=0}^{N_{topo}} \vec{p}_T^{topo} \quad (12.5)$$

while a scalar sum of all transverse energies is corresponding to the hadronic activity of the event:

$$\sum E_T = \sum_{i=0}^{N_{topo}} E_T^{topo} \quad (12.6)$$

- 354 To avoid double counting of lepton energy losses in calorimeter, the clusters inside cone with radius
- 355  $dR = 0.2$  are excluded from this calculation. To compensate soft activity inside this cone, clusters
- 356 are then compensated by replacement cone (Fig. 12.2). This cone is defined as cone at the same
- 357 pseudorapidity, but different  $\phi$ . It should be far from any other lepton and hadron recoil direction.
- 358 Each cone is then rotated to a direction of the original lepton direction. This definition is not taking
- 359 into account jet reconstruction aspects. This is allowing to get a better data MC agreement (Fig. ??).

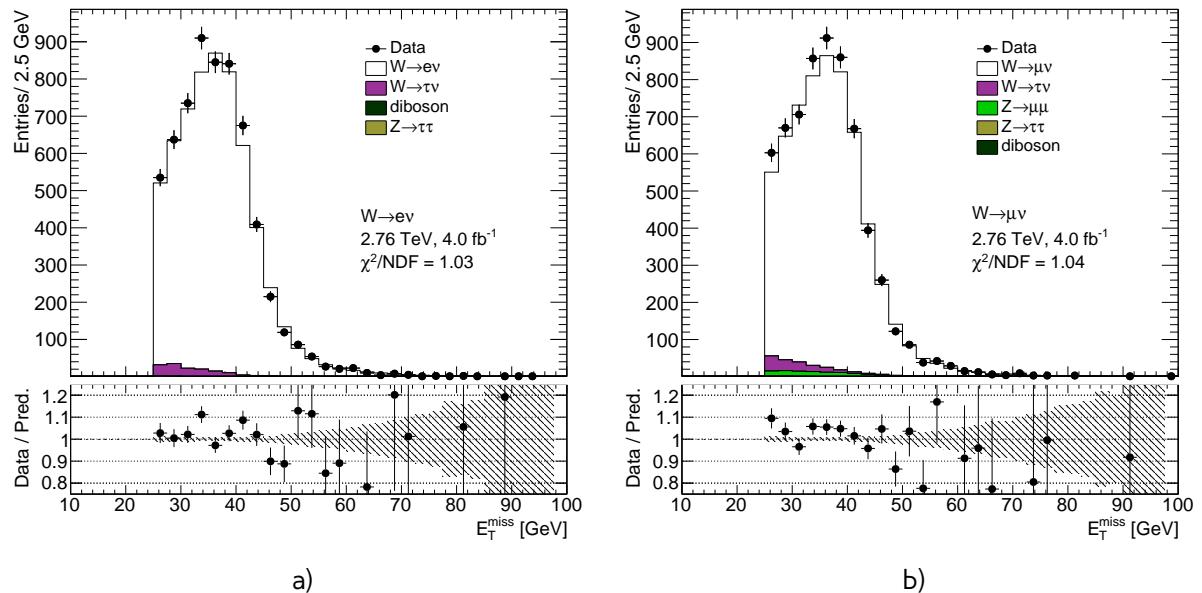


Fig. 12.3: Data and MC comparison for  $E_T^{\text{miss}}$  calculated from hadron recoil for a) $W \rightarrow e\nu$  b) $W \rightarrow \mu\nu$  events

## 360 12.2 Hadron Recoil calibration

361  $E_T^{miss}$  affects significantly on a W boson measurement, so its important to have good understanding  
 362 of sources of a possible differences in a hadron recoil reconstruction in a data and monte carlo.

The hadron recoil algorithm performance can be studied in MC through the projection of  $\vec{H}R$  on  
 the direction of the transverse momentum of the vector boson, as shown on Fig. 12.2. This projection  
 can be divided into perpendicular  $u_{\perp}$  and a parallel  $u_{\parallel}$  component as follows:

$$u_{\parallel} = \vec{v}_{xy} \cdot \vec{H}Ru_{\perp} = v_x \cdot HR_y - v_y \cdot HR_x, \quad (12.7)$$

363 where  $\vec{v}_{xy}$  is a transverse component of vector boson direction and  $v_x$  and  $v_y$  are the projections on  
 364 x and y plane respectively. In ideal case this  $u_{\parallel} = p_T^{bos}$  and  $u_{\perp} = 0$ . However calorimeter resolution is  
 365 causing relatively wide distributions for this projections (Fig. ??). Parallel component  $u_{\parallel}$  is sensitive  
 366 to a possible bias in the hadron recoil, while perpendicular  $u_{\perp}$  can be used for a determination  
 367 of resolution discrepancies. The mean and width of this distributions can depend on a different  
 368 variables, such as mean number of interactions in event, hadronic activity, boson  $P_T^{bos}$  etc. Typical  
 369 resolution of measured hadron recoil is <something>

370 It is convenient to use Z boson decays for a hadron recoil calibration, since its transverse momen-  
 371 tum can be determined not only by a hadron recoil, but also from its decay products. Zpt resolution  
 372 coming from lepton reconstruction is 3-4 times better, than from a hadron recoil. This is allowing  
 373 to treat leptonically reconstructed  $P_T^Z$  as a truth  $P_T$  of the boson and compare directly  $u_{\perp}$  and  $u_{\parallel}$   
 374 in data and MC. Small size of the Z sample in 2.76TeV data will lead to a high statistics error for  
 375 this distributions. Also, calibration constants can be also derived from W boson decays through the  
 376 indirect measurements. These corrections can be biased by a possible truth boson  $P_T$  mismodelling.

377 First step in a hadron recoil calibration procedure aims to correct differences in a pile-up modelling  
 378 in the event. Additional interaction can have a significant effect on a  $E_T^{miss}$  and  $\sum E_T$  distributions.  
 379 It is usually accounted scaling average number of interaction per bunch crossing to match a data.  
 380 However, ATLAS simulation is suited for an high pile-up runs, so this quantity is modelled discretely  
 381 in case of 2.76 TeV analysis (Fig. 12.4), what makes the corrections to match data impossible.

382 The combined Z and W boson determination procedure have been used. This section describes  
 383 a procedure of calibrating bias and resolution mismodelling in a hadron recoil, that was adapted for  
 384 2.76 TeV data.

### 385 12.2.1 Hadron recoil resolution correction

386 There are two possible ways of correcting hadron recoil resolution in a 2.76 TeV data.

387 Event activity plays an important role in a  $E_T^{miss}$  reconstruction. Since  $\sum E_T$  and hadron recoil  
 388 resolution values are correlated, the possible mismodelling of event activity can lead to a differences  
 389 between data and monte carlo (Fig. ??). It could be corrected by reweighting  $\sum E_T$  distribution to match  
 390 a data (blue arrow on a Fig.). Remaining differences can be corrected on a second step. On another  
 391 hand it is also possible to neglect second order effects on  $E_T^{miss}$  from  $\sum E_T$  distribution and directly  
 392 correct difference between data and MC (red arrow on a Fig.).

#### 393 Sumet distribution correction

Distribution of  $\sum E_T$  events are shown on a Fig. ???. There is a clear sign of shift in this distribution in a  
 both channels. Unfortunately, size of the Z sample is not sufficient for correcting this discrepancies.  
 The determination of sumet reweighting constants uses W boson decays. This procedure should

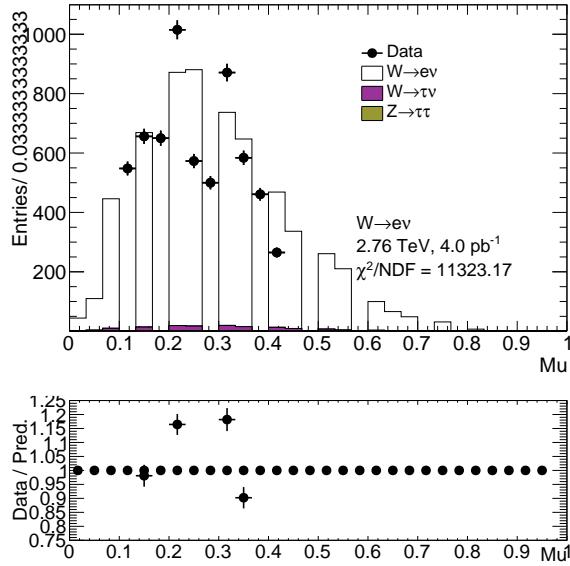


Fig. 12.4: Pileup

leave the truth boson Pt spectrum untouched. In order to do so, the correction factor are derived inside pt bins as follows:

$$SF^{channel} = \frac{\sum E_T^{data, selection}}{\sum E_T^{MC, no cuts}}, \quad (12.8)$$

where  $\sum E_T^{data, selection}$  and  $\sum E_T^{MC, no cuts}$  is a  $\sum E_T$  distribution inside  $p_T^{W, rec}$  bin without any cuts. Oppositely, in MC  $\sum E_T^{MC, no cuts}$  is taken without any cuts. Scale factors are determined separately for each signal process for a W boson decays. In order to increase statistics in data the combination of  $W \rightarrow ee$  and  $W \rightarrow \mu\nu$  processes is used. In order to have a smother correction and not try to account data fluctuation ratio inside each bin can be parameterised by a polynomial degree 2 inside each  $p_T^{W, rec}$  bin (Fig. ??). Total SF obtained by this procedure are shown on a Fig. ???. The distribution of  $\sum E_T$  after correction is shown on a Fig. 12.8. Reconstructed boson pt spectrum is leaving almost untouched, while this procedure still intoduses some shift in a truth boson pt spectrum(Fig. 12.7 ). Effect on the resoultion of  $u_\perp$  is shown on a Fig. ??.

Statistical error of this fluctuations can be estimated from polynomial parameters obtained from fit using bootstrap method. Inside each bin parameters are varied within its fit uncertanty as:

$$fitparametersnew = fitparameters + gaus^{2D}(cov.matrix), \quad (12.9)$$

where *fitparameters* is a vector of best fit parameters and *gaus*<sup>3D</sup> is a 3D gaus, that takes covariance matrix from fit results. This method is allowing to take into account correlations between parameters. This procedure is repeated 25 times for each bin, that gives us set of 25 scale factors, that are later used for error determination.

Sytematical error can be studied by applying lower order of approximation on a SF or not applying it at all. The overall effect on a  $C_W$  for a different methods is shown in a Tab. ???. Results are dominated by a systematics error. Hoever, there is a difference in a sign of the effect for a different flavours of the analysis. This cannot be explained from a physical point of view, so it was decided not to use this corrections in a final analysis.

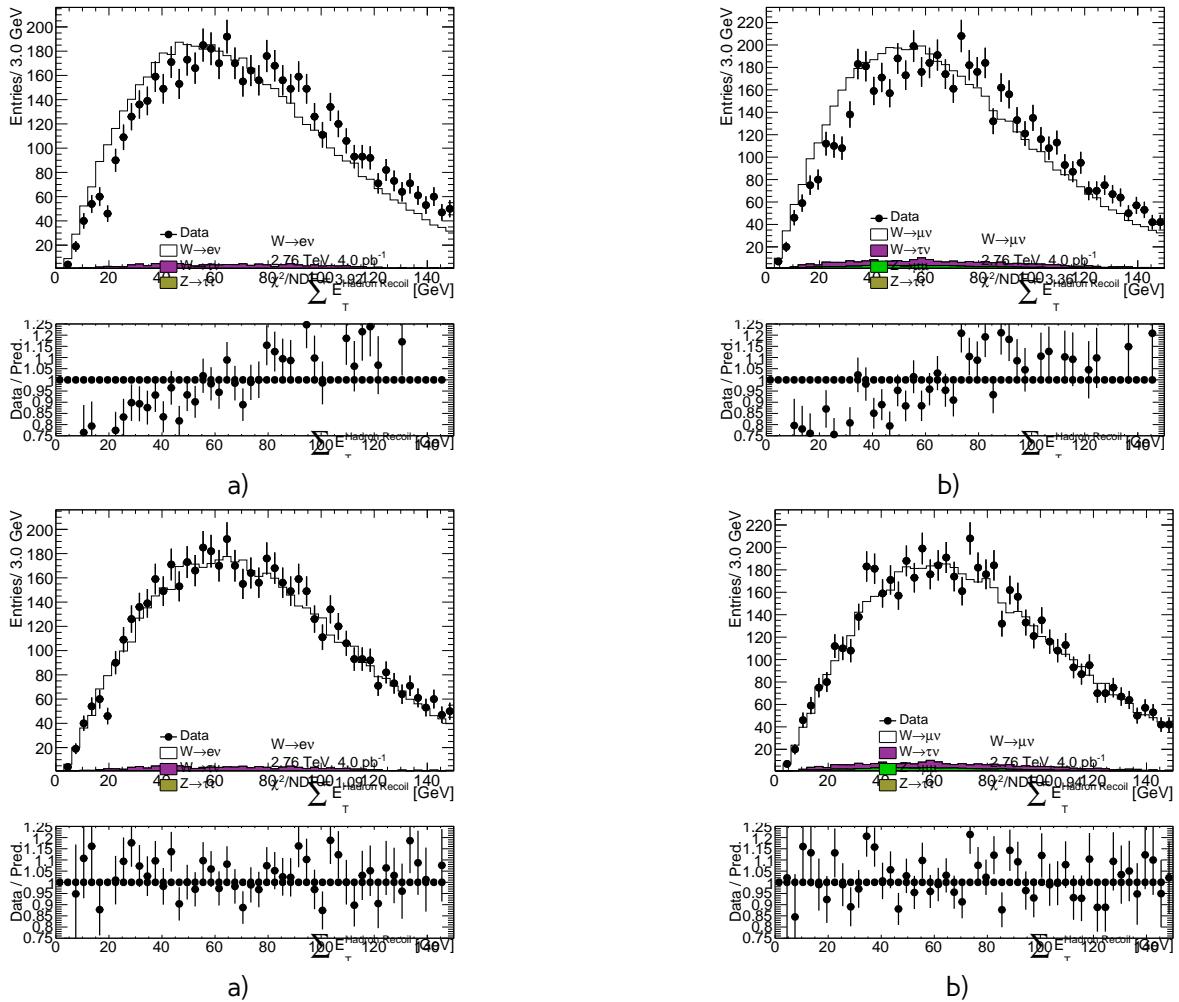


Fig. 12.5

#### 412 Resolution correction using Z events

413 Another way to check resolution effects is to use  $u_{\perp}$  and  $u_{\parallel} - \frac{Z}{T}$  distributions in a Z events. This  
 414 correction assumes, that any resolution mismodelling reflects discrepancies in the  $\sum E_T$  distribution,  
 415 while mismodelling of resolution at a given  $\sum E_T$  is a subleading. There is a clear difference in a rms  
 416 of this distributions between data and MC, that cannot be accounted as a statistical error in data.  
 417 Difference in resolutions is consistent for  $u_{\perp}$  and  $u_{\parallel} - \frac{Z}{T}$  distributions, but depends on a flavour of  
 418 the analysis. The resolution mismodelling is corrected by adding up a gaus to each component of a  
 419 hadron recoil:

$$u'_{\parallel} = u_{\parallel} + \text{Gaus}(0, d\sigma), \quad u'_{\perp} = u_{\perp} + \text{Gaus}(0, d\sigma), \quad (12.10)$$

where  $d\sigma$  is a difference in a resolutions calculated as:

$$d\sigma = \sqrt{\sigma_{data}^2 - \sigma_{MC}^2} \quad (12.11)$$

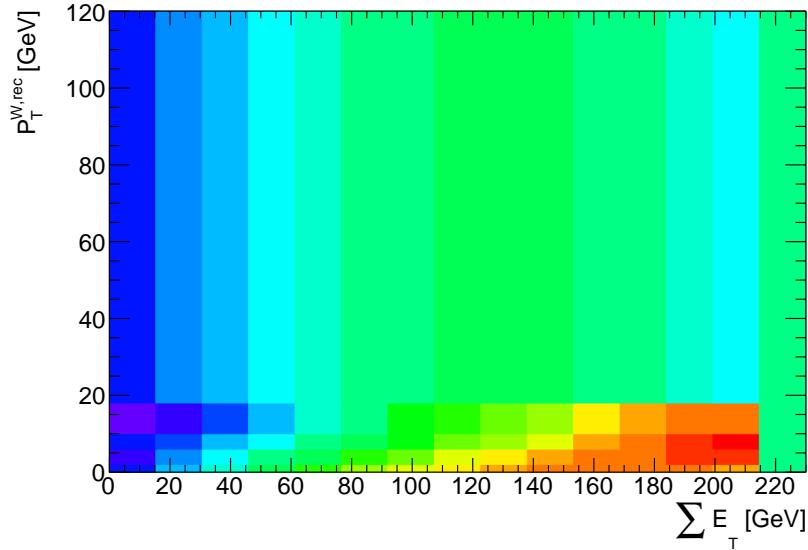


Fig. 12.6: Correction factors for  $W \rightarrow e\nu$

420 Systematic error of this  $d\sigma$  is taken as an statistical error for  $\sigma_{data}$ . Overall effect on a  $C_W$  depending  
 421 on a  $d\sigma$  is shown on a Fig. 12.9. Due to a random nature of this correction, effect is not stable  
 422 for a small  $d\sigma$ . Stability of this correction can be tested by repeating this procedure several times  
 423 with different random seed number. Overall effect on a  $C_W$  and corresponding systematical and  
 424 statistical error are summarized in Tab. ???. The effect of stability of the correction is a subleading.

### 425 12.2.2 Hadron recoil bias correction

As it was mentioned before, it is possible to use both Z and W boson sample for hadron recoil bias determination. Correction factor  $SF_{HR,bias}$  is applied as:

$$u_{\parallel}^{MC,cor} = u_{\parallel}^{MC} \cdot SF_{HR,bias}, \quad (12.12)$$

426 and can be obtained by scanning the impact of the scaling factor on the Data to MC agreement of  
 427 the distributions that are dominated by the recoil scale uncertainties. Since W boson has no second  
 428 source of  $P_T^W$  measurments, determination of the hadron recoil bias should use the distributions,  
 429 that are not sensitive to a truth  $P_T^W$  spectrum. One of the optimal choises is a  $M_T^W$  distribution.  
 430 Transverse mass distribution for a different scale choises is shown on a Fig. ???. Multijet background  
 431 is not included, because it shape and number of events is depending on a hadron recoil scale and  
 432 thus can introduce additional systematics.

433 The first way to determine correction factor is using a difference in the mean of transverse mass  
 434 in data and MC. Statistical error of this determination is an error of the mean in the data. The  
 435 precision of this method is low, is it is mainly used as a cross-check.

Second way is calculating  $\chi^2$  for each correction factor. The ideal correction factor is determined by fitting  $\chi^2$  distribution by the function:

$$\chi^2 = \frac{(x - sf_{best})^2}{\sigma_{sf}^2} + \chi_0^2, \quad (12.13)$$

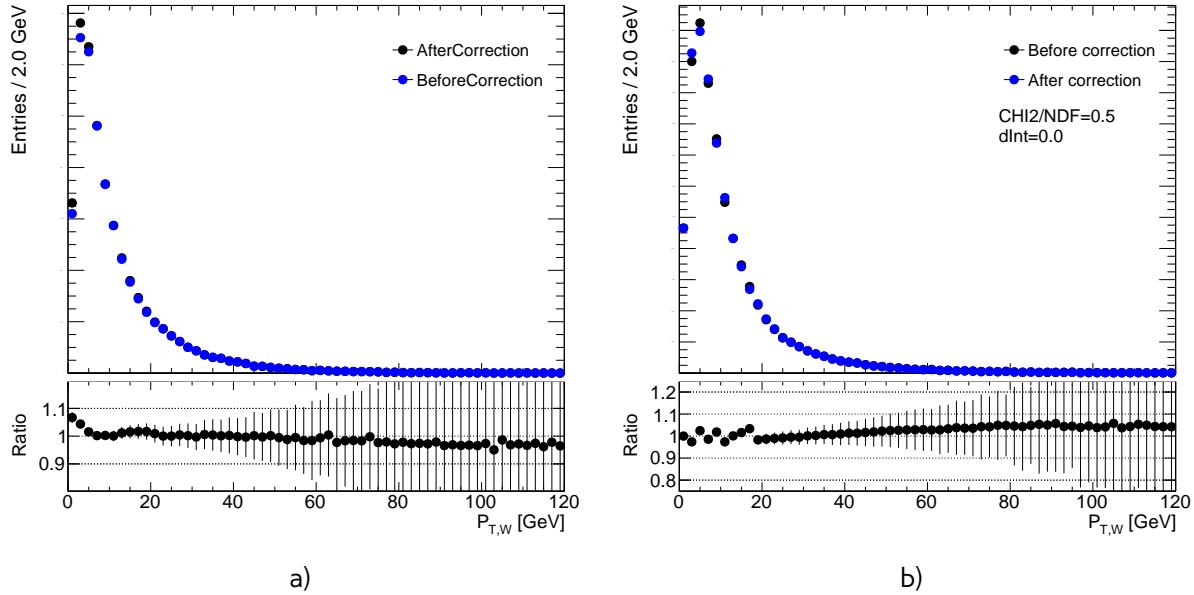


Fig. 12.7

Method	SF	error
Mean $M_T^W$	1.10	0.2
$M_T^W \chi^2$	1.01	0.07
$u_{\parallel} \chi^2$	1.00	0.014
Total	1.010	0.013

436 where  $s f_{best}$  is the best scale factor and  $\sigma_{sf}$  is a statistical error of this parameter. Distribution of  $\chi^2$   
 437 and a fit in combined W channel is shown on a Fig. ??.

438 Because of the possible mismodelling of the tail  $M_T^W$  distribution it is not included in a  $\chi^2$  cal-  
 439 culation, leaving a free choise of the parameter of the cutoff. It is also possible to exclude regions  
 440 with high multijet background contamination by applying a tighter cut on a  $M_T^W$ . This fit range is  
 441 introducing one source of systematic error. Effect of the range on value determination is shown on a  
 442 Fig. ???. Similarly to a W channel, scale correction in a Z sample can be determined from distribution  
 443  $\frac{u_{\parallel}}{p_T^W}$ , shown on a Fig. ???. Since there is no choise of the range and dependency on  $P_T^{bos}$  modelling,  
 444 there is just one source of uncertainty.

445 Results on a hadron scale factros and it's errors are shown in a Table ???. The results are consistent  
 446 within 1 sigma.

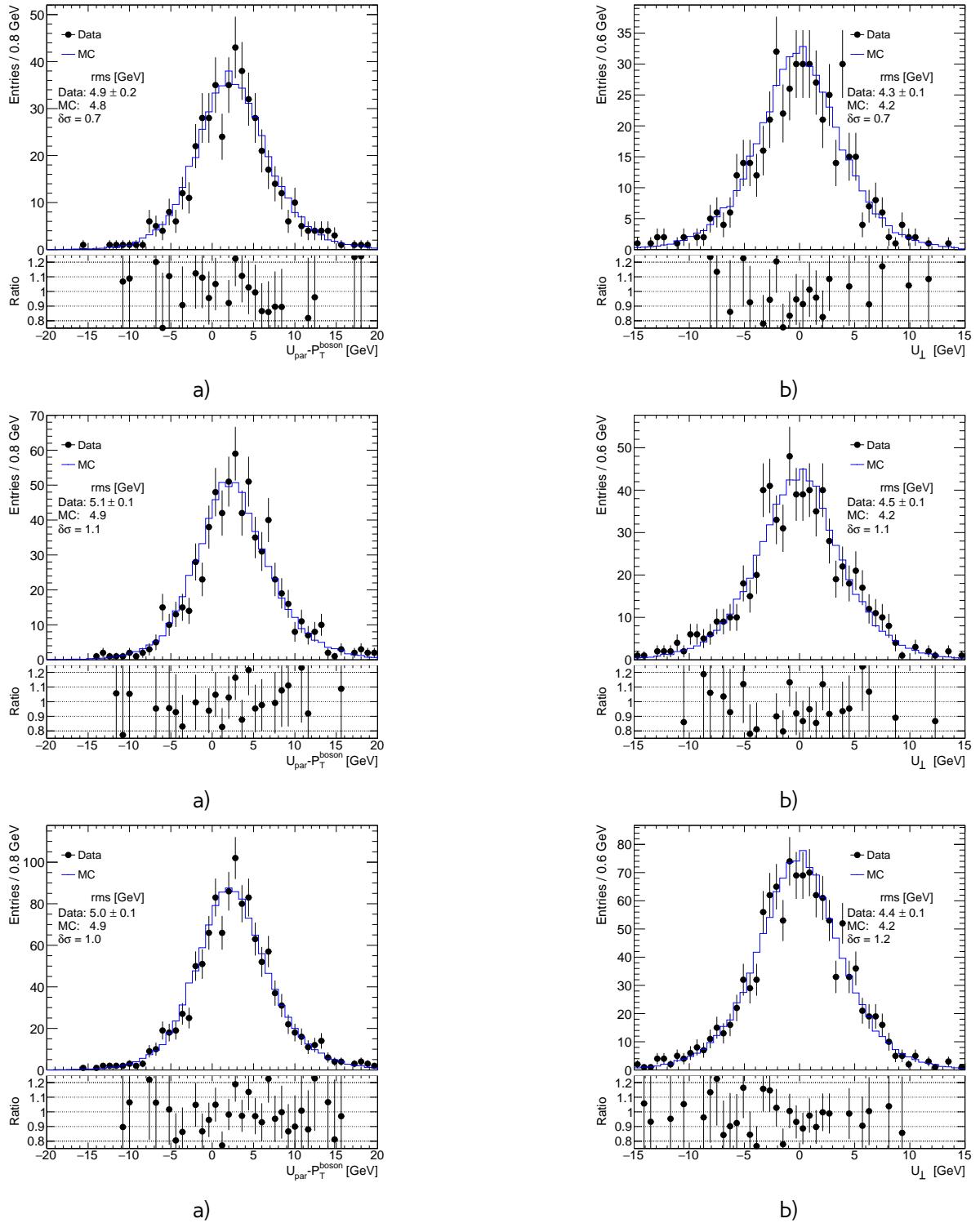


Fig. 12.8

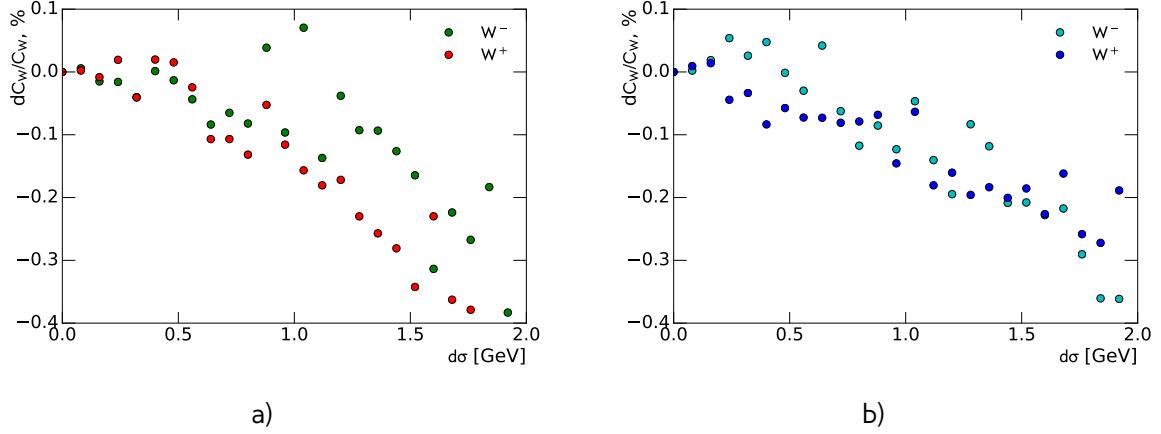


Fig. 12.9: Effect on a  $C_W$  for a different  $d\sigma$  for a)  $W \rightarrow e\nu$  b)  $W \rightarrow \mu\nu$  channel

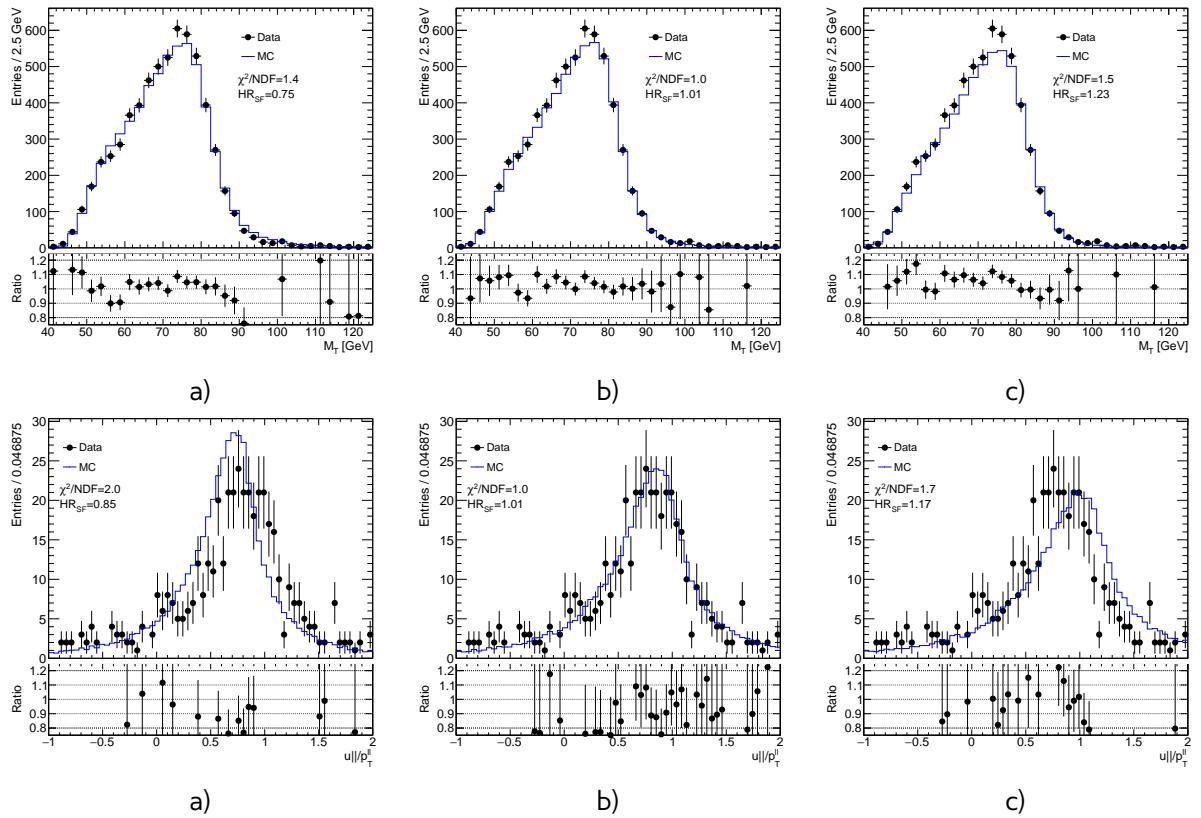


Fig. 12.10: Trigger scale factors for a)  $\mu$  b)  $\mu^+$  c)  $\mu^-$

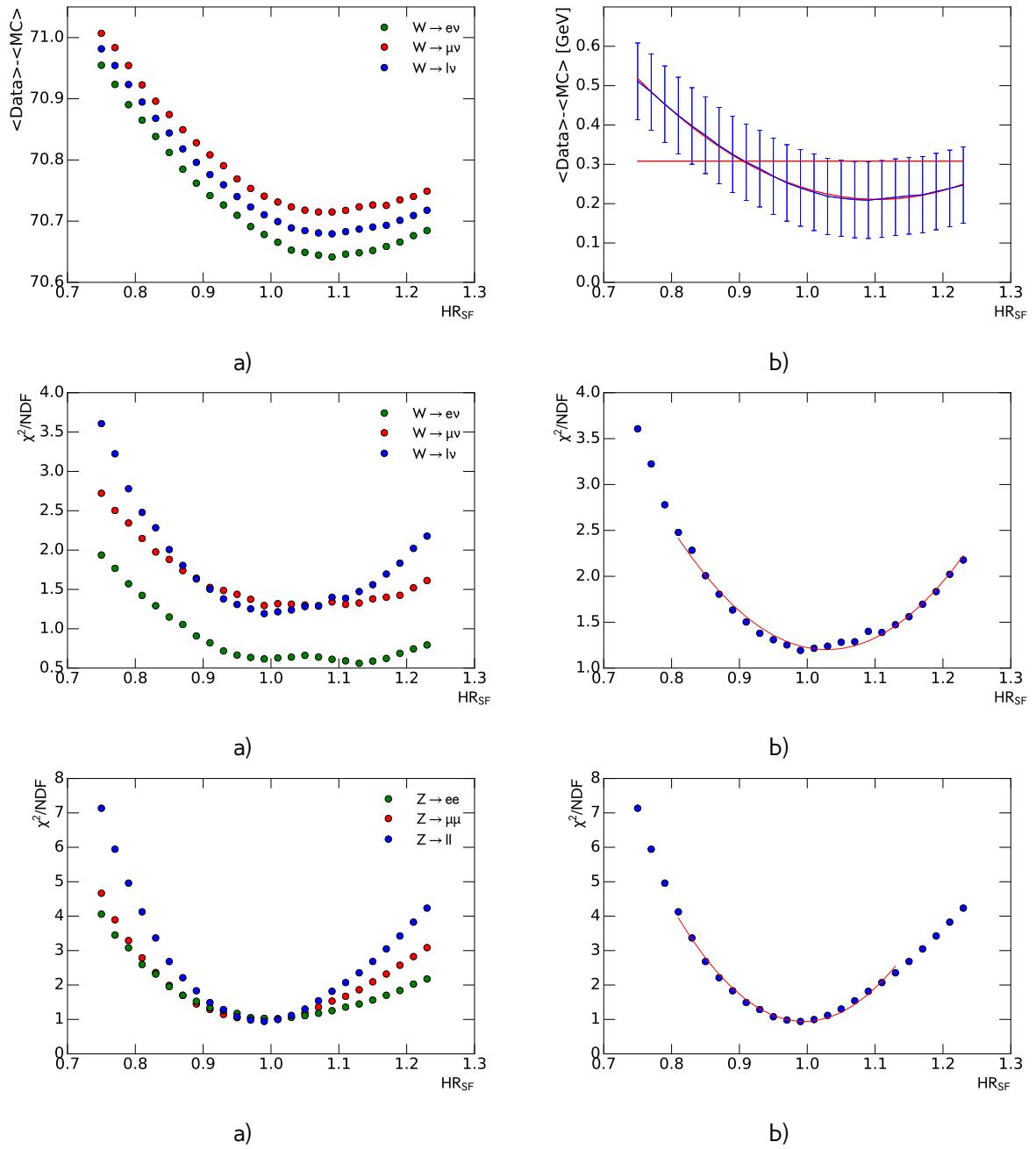


Fig. 12.11: Effect on a  $C_W$  for a different  $d\sigma$  for a)  $W \rightarrow e\nu$  b) $W \rightarrow \mu\nu$  channel

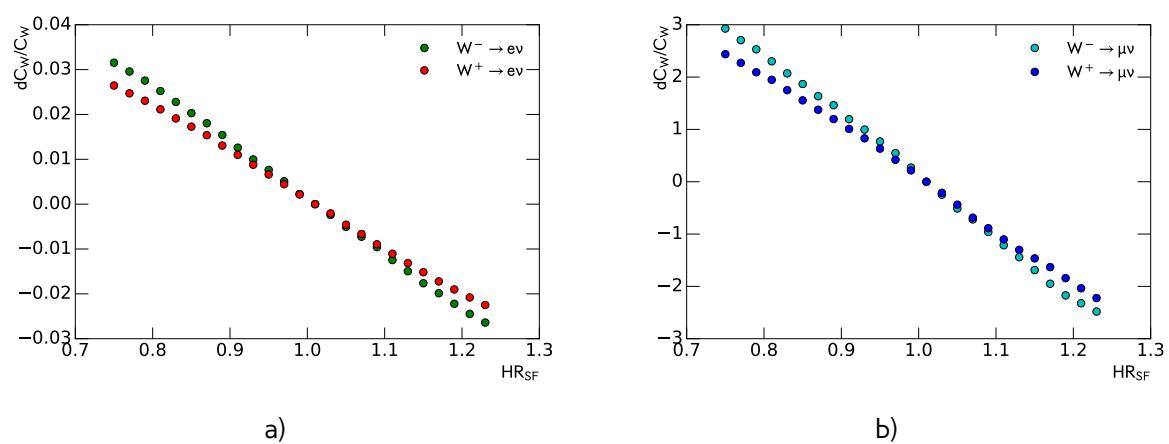


Fig. 12.12: Effect on a  $C_W$  for a different  $d\sigma$  for a)  $W \rightarrow e\nu$  b) $W \rightarrow \mu\nu$  channel



447

# Chapter 13

## 448 Background estimation

449 After the event selection described in chapter 10 the background contribution is around % for W  
 450 and % for Z analysis (which is with this statistics is negligible). Main backgrounds for W analysis  
 451 are coming from:

- 452 • Processes with  $\tau$  lepton, misidentified as a electron or muon + missing energy from neutrino
- 453 • Z decays with one missing lepton.
- 454 • QCD processes. In electron channel this is mostly coming from jets faking electrons, while in a  
 455 muon channel it consists mostly of a real muons produced in decays of heavy-flavor mesons.

456 Most of the backgrounds are estimated using MC. They are normalized using highest cross-section  
 457 order available. The total list of simulated backgrounds and its cross-section is shown in a Table 13.1.  
 458 QCD background is estimated using data driven method.

Table 13.1: Background processes with their associated cross sections and uncertainties. The quoted cross sections are used to normalise estimates of expected number of events

Process	$\sigma \cdot BR$ [pb]	Order
$W^+ \rightarrow l\nu$	2116(1.9)	NNLO
$W^- \rightarrow l\nu$	1267(1.0)	NNLO
$Z \rightarrow ll$	303(0.2)	NNLO
$t\bar{t}$	7.41	LO
$WW$	0.6	LO
$ZZ$	0.7	LO
$WZ$	0.2	LO
$DY \rightarrow ee$	2971	LO
$DY \rightarrow \mu\mu$	2971	LO

### 459 13.1 QCD background estimation

460 There is a small probability, that jet can fake W-boson decay with isolated lepton and  $E_T^{miss}$  through  
 461 the energy mismeasurement in the event. Event selection is suppressing this type of the background,  
 462 but not fully eliminating it. Due to a large jet production cross-section and complex composition,

463 generation of MC events becomes impractical. This is why data driven technique for QCD back-  
 464 ground estimation have been used. In our case contribution of the QCD background in a Z sample  
 465 is negligible, so it is estimated just for a  $W \rightarrow e\nu$  and  $W \rightarrow \mu\nu$  processes.

466 Data driven method allows to have model independent predictions with small statistical uncertainty.  
 467 This method is using *QCD* enriched region, where signal events are suppressed. This is usually done  
 468 by reversing identification or isolation criteria. It is assumed, that shape of the qcd background is  
 469 staying the same in the signal region. Normalization can be derived in a control region through the  
 470 template fit.

471 This section describes method of QCD background determination, that have been used in 2.76  
 472 TeV data.

## 473 13.2 Template selection

A study have been performed to determine appropriate template selection. Identification criteria are inverted in order to suppress the signal events. Because of the origins of the QCD backgrounds, missing transverse energy  $E_T^{miss}$  should be smaller in a QCD, than in a signal region. Releasing  $E_T^{miss}$  cut is allowing to gain a bigger statistics for a QCD template. The template sample can have a contributions from other backgrounds (mostly coming from  $W \rightarrow l\nu$ ). In order to avoid double counting, they are subtracted from a template. The total number of events in the template can be defined as:

$$N_{template} = N_{data}^{bkg\ enriched} - \sum_j^{MC} N_{MC_j}^{bkg\ enriched}, \quad (13.1)$$

474 where  $N_{data}^{bkg\ enriched}$  and  $N_{MC_j}^{bkg\ enriched}$  are number of the events in a background enriched sample in data  
 475 and MC respectively.

476 For electron flavour, template is build by requiring the electron candidate to fail Medium isolation  
 477 criteria, but to pass loose selection. The resulting shape of the QCD background is shown on a Fig.  
 478 ???. Events are selected to pass looser trigger <>, which requires on electron candidate that passing  
 479  $p_T^{lep} > 10$  GeV and loose ID criteria. The total number of events in a templates is <somethig>. The  
 480 stability of the template can be studied by reversing different identification criteria. As it can be  
 481 seen on a Fig. ???, ID criteria is almost not affecting shape of the QCD background.

482 It is impossible to use similar procedure in muon channel, since the resulting statistics of template  
 483 is small (Fig. ??). Another way of defining QCD template is using the properties of the process, that  
 484 is resulting in a fake muons. Fake muon are mostly coming from a heavy flavour decays. If a charged  
 485 hadron comes through the HCAL, it can leave a track in MS and be identified as a muon. Most of  
 486 the time there are multiple tracks in the ID. So the template muon is selected as a muon, that has  
 487 track in the ID, but has no track in MS. Effect on the shape of this selection can be studied using  
 488 smaller sample of  $b\bar{b}$  and  $c\bar{c}$  MC samples. Additionally, checks on a differences in a shape between  
 489 signal and template region have been performed (Fig. ??). This is totally justifies this choice of the  
 490 template selection. Shape of the background should not depend on the charge of the analysis, so it  
 491 was decided to use template, combined in a both channel. The resulting template is shown on a Fig.  
 492 ???. Total number of events in a template <something>.

493 

### 13.3 Methodology of the template sample normalization

As it was mentioned before, multijet events tend to have smaller  $E_T^{miss}$ , than a signal. It was decided to use  $E_T^{miss}$  distribution with the released  $E_T^{miss}$  cut for a finding a template shape normalization. The normalisation is found through the  $\chi^2$  fit of the template and backgrounds to the data. The following composite model have been used for estimation:

$$M(E_T^{miss}) = \sum_{i=1}^{N-1} f_i F_i(E_T^{miss}) + (1 - \sum_{i=1}^{N-1} f_i) \cdot F_{qcd}(E_T^{miss}), \quad (13.2)$$

494 where  $F_i(E_T^{miss})$  and  $F_{qcd}(E_T^{miss})$  are the probability density functions of MC samples and QCD back-  
495 ground template respectively. Fit parameters  $f_i$  are the fractions of MC within fit region. In order  
496 to eliminate systematics, coming from cross-section uncertainty, with signal fractions are left freely  
497 and background MC fractions are varied within 5% uncertainty.

Normalisation scale for QCD events is calculated from obtained fit parameters as:

$$scale = \frac{(1 - \sum f_i) \cdot N_{Data}^{fit}}{N_{template}}, \quad (13.3)$$

498 where  $\sum f_i$  is a sum of all fractions in the fit,  $N_{Data}^{fit}$  is a number of data events in a fit histogram  
499 and  $N_{template}$  is a number of event in a template. The fit is performed separately for  $W^+$  and  $W^-$ .  
500 Additionally, fit in uncharged  $W$  channel is used as a cross-check of the fit. The results of the fitting  
501 procedure are shown on a Fig. . Total number of events and fit uncertainty are shown in a Tab. ??.  
502 The overall fraction of QCD events is lower, than in 7 TeV data <reference to a 7 TeV paper>, what  
503 is agreeing with expectations.

 504 

### 13.4 Systematic Uncertainty from the Multi-jet Background 505 Estimation

The uncertainty of multi-jet background can esimation can be divided into 3 main components:

$$\delta_{QCD} = \sqrt{\delta_{fit\,unc} + \delta_{fit\,bias} + \delta_{template}}, \quad (13.4)$$

506 where  $\delta_{fit\,unc}$  is the uncertainty for a scale from a  $\chi^2$  fit. The second  $\delta_{fit\,bias}$  is coming from an effect  
507 from arbitrary choise of binning and fit range. This error is estimated by repeating fit for a different  
508 bin and range choises. Fird uncertainty is due to a potential bias in the template as a result of the  
509 template choise and a template statistics itself.



510 Chapter **14**

511 **Uncertainties**

512 **14.1 Toy MC method**

513 **14.2 Experimental systematic uncertainties**

514 table about methods

515 **14.2.1 Hadron Recoil correction uncertainty**

516 **14.3 Theoretical uncertainty**

517     • PDF

518     • Parton shower and matrix element



519

# Chapter 15

520

## Control distributions



521 Chapter **16**

522 **Results of the Cross Section Measurement**

523 **16.1 Cross-Section measurement definition**

524 **16.2 Fiducial phase-space**

525 **16.3 Comparation with Theoretical Predictions**



526

## **Part IV**

527

## **PDF fits**

