

# Contents

|  |           |
|--|-----------|
| <b>I Thesis overview</b>   | <b>1</b>  |
| <b>II Theory introduction</b>  | <b>3</b>  |
| <b>1 Theoretical introduction</b>  | <b>5</b>  |
| 1.1 Standard model . . . . .   | 5         |
| 1.2 Theory of pp collisions . . . . .  | 5         |
| 1.3 Proton structure . . . . .   | 5         |
| <b>2 Physics of W and Z bosons in pp collisions</b>                            | <b>7</b>  |
| <b>3 Methodology of Cross-Section Measurement</b>                              | <b>9</b>  |
| 3.1 Fiducial volume definition . . . . .                                       | 9         |
| 3.2 Results combination . . . . .  | 10        |
| 3.3 PDF fits . . . . .   | 10        |
| <b>III Experimental setup</b>  | <b>11</b> |
| <b>4 The Large Hadron Collider</b>   | <b>13</b> |
| <b>5 The ATLAS experiment</b>  | <b>15</b> |
| 5.1 Inner detector . . . . .   | 15        |
| 5.2 Calorimeters . . . . .   | 15        |
| 5.2.1 Forward calorimeters . . . . .   | 15        |
| 5.3 Muon system . . . . .  | 15        |
| 5.4 The magnet system . . . . .  | 15        |
| 5.5 Trigger system . . . . .   | 15        |
| 5.6 Atlas software . . . . .   | 15        |
| <b>6 Event reconstruction</b>  | <b>17</b> |
| 6.1 Tracks and vertexes . . . . .  | 17        |
| 6.2 Electron reconstruction and identification . . . . .                       | 17        |
| 6.3 Muon reconstruction and identification . . . . .                           | 17        |
| 6.4 Missing transverse energy reconstruction . . . . .                         | 17        |
| 6.4.1 Standard Missing Transverse Energy reconstruction . . . . .              | 17        |
| 6.4.2 Reconstruction of Missing Transverse Energy from hadron recoil . . . . . | 19        |
| <b>7 Monte Carlo simulation</b>  | <b>21</b> |
| 7.1 ATLAS chain of MC production . . . . .                                     | 21        |
| 7.2 Event generators . . . . .   | 22        |

|  |           |
|--|-----------|
| 7.3 Simulation in Geant4 . . . . .   | 23        |
| <b>8 Frozen Showers</b>  | <b>25</b> |
| 8.1 Problem description . . . . .  | 26        |
| 8.2 Generation and use in simulation . . . . .                                   | 30        |
| 8.2.1 Libraries tuning . . . . .   | 30        |
| 8.3 Machine learning based bin finding procedure . . . . .                       | 32        |
| 8.3.1 Machine Learning . . . . .   | 32        |
| 8.3.2 Model description . . . . .  | 34        |
| 8.3.3 Interpretation of results . . . . .  | 34        |
| 8.4 Validation of the new libraries . . . . .                                    | 37        |
| 8.5 Plans for a future developments . . . . .                                    | 37        |
| <b>9 Data and Monte-Carlo samples</b>  | <b>39</b> |
| 9.1 Data sample . . . . .  | 39        |
| 9.2 Monte-Carlo samples . . . . .  | 39        |
| <b>IV The Measurement</b>  | <b>41</b> |
| <b>10 Selection</b>  | <b>43</b> |
| 10.1 Data quality cuts . . . . .   | 43        |
| 10.2 Lepton quality cuts . . . . .   | 44        |
| 10.3 Boson selection . . . . .   | 44        |
| 10.4 Cut flow . . . . .  | 44        |
| <b>11 Monte Carlo corrections</b>  | <b>45</b> |
| 11.1 Lepton efficiency corrections . . . . .                                     | 45        |
| 11.1.1 Muon Trigger SF . . . . .   | 47        |
| 11.2 Electron energy scale and resolution . . . . .                              | 49        |
| 11.3 Muon momentum correction . . . . .  | 50        |
| <b>12 Hadron recoil calibration</b>  | <b>53</b> |
| 12.1 Hadron recoil resolution correction . . . . .                               | 54        |
| 12.1.1 Sumet distribution correction . . . . .                                   | 55        |
| 12.1.2 Resolution correction using Z events . . . . .                            | 58        |
| 12.2 Hadron recoil bias correction . . . . .                                     | 60        |
| <b>13 Background estimation</b>  | <b>65</b> |
| 13.1 QCD background estimation . . . . .   | 66        |
| 13.1.1 Template selection . . . . .  | 67        |
| 13.1.2 Methodology of the template sample normalization . . . . .                | 67        |
| 13.1.3 Systematic Uncertainty from the Multi-jet Background Estimation . . . . . | 69        |
| 13.2 Background-subtracted W and Z candidate events . . . . .                    | 71        |
| <b>14 Control distributions</b>  | <b>73</b> |

|   |           |
|---|-----------|
| <b>15 Uncertainties in the cross-section measurement</b>      | <b>83</b> |
| 15.1 Methods of uncertainties propagation . . . . .           | 83        |
| 15.2 Experimental systematic uncertainties . . . . .          | 83        |
| 15.2.1 Electron energy scale and resolution . . . . .         | 84        |
| 15.2.2 Muon energy scale and resolution . . . . .             | 84        |
| 15.2.3 Muon and electron efficiency toy Monte-Carlo . . . . . | 84        |
| 15.3 Theoretical uncertainty . . . . .                        | 86        |
| 15.4 Correlation between uncertainties . . . . .              | 87        |
| 15.4.1 Toy MC correlations . . . . .                          | 87        |
| 15.4.2 Correlations between PDF's eigenvectors . . . . .      | 87        |
| <b>V Results</b>  | <b>89</b> |
| <b>16 Results of the Cross Section Measurement</b>            | <b>91</b> |
| 16.1 Combined results . . . . .                               | 91        |
| 16.2 Comparation with Theoretical Predictions . . . . .       | 91        |
| <b>17 PDF fits results</b>                                    | <b>93</b> |
| <b>18 Summary</b>   | <b>95</b> |



1

## **Part I**

2

# **Thesis overview**



3

## **Part II**

4

# **Theory introduction**

3



<sup>5</sup> Chapter **1**

<sup>6</sup> **Theoretical introduction**

<sup>7</sup> **1.1 Standard model**

<sup>8</sup> **1.2 Theory of pp collisions**

<sup>9</sup> **1.3 Proton structure**



10 Chapter **2**

11 **Physics of W and Z bosons in pp collisions**



12 Chapter **3**

13 **Methodology of Cross-Section Measurement**

The production of W and Z bosons is measured in a fiducial volume using the equation:

$$\sigma_{fid} = \frac{N - B}{C_{W/Z} L_{int}}, \quad (3.1)$$

14 where

- 15 • N is the number of candidates measured in a data  
16 • B is the number of background events  
17 •  $L_{int}$  is the integrated luminosity corresponding to a run selections and trigger requirements  
18 • Efficiency correction factor  $C_W$ .

The efficiency scale factor  $C_W$  is calculated from MC and defined as :

$$C_{W/Z} = \frac{N_{MC,rec}}{N_{MC,gen,cut}}, \quad (3.2)$$

19 where  $N_{MC,rec}$  are sums of weights of events after simulation, reconstruction and selection,  $N_{MC,gen,cut}$   
20 - are taken on the generator level after fiducial cuts.

The total W and Z cross-sections are calculated as:

$$\sigma_{tot} = \sigma_{W/Z} \cdot BR(W/Z \rightarrow l\nu/l\bar{l}) = \frac{\sigma_{fid}}{A_{W/Z}} = \frac{N - B}{A_{W/Z} C_{W/Z} L_{int}}, \quad (3.3)$$

where the acceptance  $A_{W/Z}$  is used to extrapolate the cross-section measurement in the fiducial volume to the full kinematic region. It is derived from MC and affected only by the theoretical uncertainties. It is obtained as follows:

$$A_{W/Z} = \frac{N_{MC,gen,cut}}{N_{MC,gen,all}}, \quad (3.4)$$

21 where  $N_{MC,gen,all}$  are the sum of weights of all generated MC events

22 **3.1 Fiducial volume definition**

In order to combine results fiducial region is defined commonly for a different flavors of the analysis:

$$W \rightarrow l\nu : p_T^l > 20 GeV, |\eta^l| < 2.5, p_T^\nu < 20 GeV, M_T^W > 40 GeV \quad (3.5)$$

$$Z \rightarrow ll : p_T^l > 20\text{GeV}, |\eta^l| < 2.5, 66\text{GeV} < m_{ll} < 116\text{GeV} \quad (3.6)$$

<sup>23</sup> Additionally, for a future combination, 13 TeV analysis fiducial phase space have been used:

$$W \rightarrow l\nu : p_T^l > 20\text{GeV}, |\eta^l| < 2.5, p_T^\nu < 20\text{GeV}, M_T^W > 40\text{GeV} \quad (3.7)$$

$$Z \rightarrow ll : p_T^l > 20\text{GeV}, |\eta^l| < 2.5, 66\text{GeV} < m_{ll} < 116\text{GeV} \quad (3.8)$$

<sup>24</sup> **3.2 Results combination**

<sup>25</sup> **3.3 PDF fits**

26

## **Part III**

27

# **Experimental setup**



<sup>28</sup>

Chapter

**4**

<sup>29</sup>

## The Large Hadron Collider



30

# Chapter **5**

## 31 **The ATLAS experiment**

32 **5.1 Inner detector**

33 **5.2 Calorimeters**

34 **5.2.1 Forward calorimeters**

35 **5.3 Muon system**

36 **5.4 The magnet system**

37 **5.5 Trigger system**

38 **5.6 Atlas software**



39 Chapter

# 6

## 40 Event reconstruction

### 41 6.1 Tracks and vertexes

### 42 6.2 Electron reconstruction and identification

### 43 6.3 Muon reconstruction and identification

### 44 6.4 Missing transverse energy reconstruction

45 Atlas detector has almost  $4\pi$  coverage. This allows to calculate imbalance of energies inside calorimeter,  
46 especially transversal part of it called  $E_T^{miss}$ . In W-analyses  $E_T^{miss}$  is used as a proxy for neutrino  
47 from a  $W \rightarrow l\nu$  decay. It leaves detector without interacting with it and that causes large energy  
48 imbalance in a detector. In this section two methods of  $E_T^{miss}$  reconstruction and the reasons for  
49 using non-standard one will be discussed.

#### 50 6.4.1 Standard Missing Transverse Energy reconstruction

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:

$$E_{x(y)}^{miss} = E_{x(y)}^{miss,calo} + E_{x(y)}^{miss,cryo} + E_{x(y)}^{miss,muon}. \quad (6.1)$$

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

$$E_{x(y)}^{miss} = E_{x(y)}^{miss,e} + E_{x(y)}^{miss,\gamma} + E_{x(y)}^{miss,\tau} + E_{x(y)}^{miss,jets} + E_{x(y)}^{miss,SoftTerm} + E_{x(y)}^{miss,\mu}. \quad (6.2)$$

51 where each term is calculated as a negative sum of the calibrated reconstructed objects, projected  
52 onto the x and y directions. Each jet with energy  $P_T > 20$  GeV is corrected for a pile-up and a jet  
53 energy scale is applied. Soft term is calculated from topoclusters and tracks, that are not associated  
54 with high-pt objects. To avoid double counting, muon energy loss in the calorimeter is subtracted  
55 from  $E_T^{miss}$ . The  $E_T^{miss}$  muon term is calculated from the momenta of muons measured in a range  
56 of pseudorapidity. Since pileup has a significant effect on a  $E_T^{miss}$  performance several methods of  
57 pileup suppression are used.

58 The runs at 2.76 TeV are characterized by a low pileup (mean number of interaction per bunch  
59 crossing  $< 1.0$ ), so the usage of a procedure optimized for high pileup 8 TeV runs may not be optimal.  
60 It was examined and figured out, that there are big discrepancies between  $E_T^{miss}$  distributions for data

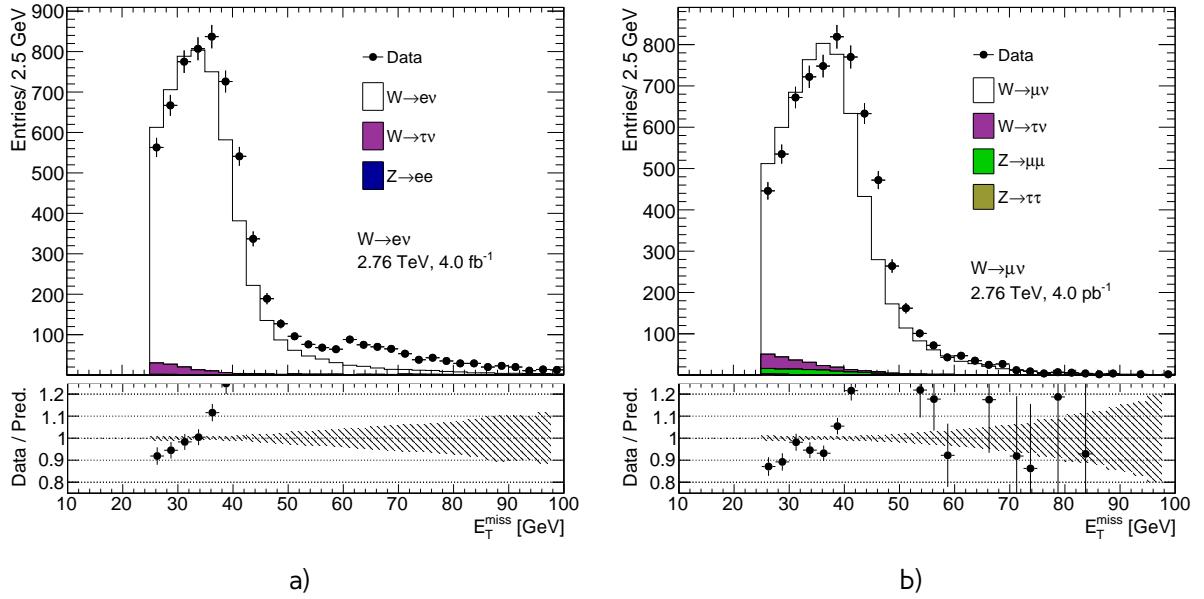


Fig. 6.1: Missing transverse energy distribution for a) the  $W \rightarrow e\nu$  selection and b) the  $W \rightarrow \mu\nu$  selection from Chap. 10.  $E_T^{\text{miss}}$  calculated using the standard ATLAS algorithm. The expected contributions from all backgrounds are estimated with Monte Carlo simulations, except for QCD background that is not included. All Monte-Carlo corrections from Chap. 11 are applied. There are visible discrepancies between data and MC, that cannot be explained by the contribution of QCD background, which is expected mainly in the low  $E_T^{\text{miss}}$  region (Sec. 13.1).

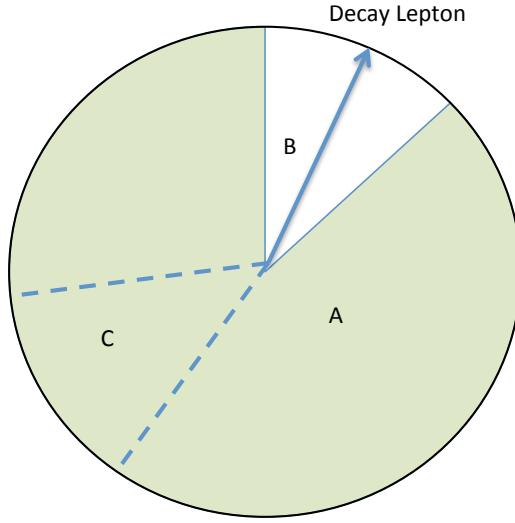


Fig. 6.2: Definition of different zones in the calculation of the cluster-based hadronic recoil. Zone B is excluded from hadron recoil calculation because it contains decay lepton. To describe properly overall acitivity it is replaced by the zone C, rotated in the direction of B. Zone A corresponds to the rest of the calorimeter.

and MC simulation, as shown on a Fig. 6.1, where the missing transverse energy for data is compared to signal and background MC predictions.

The differences are visible in both electron and muon channels and cannot be explained by the (missing on the control plots) contributions from the QCD background, which is expected mainly in the low  $E_T^{\text{miss}}$  region (see Sec. 13.1).

#### 6.4.2 Reconstruction of Missing Transverse Energy from hadron recoil

Different way of  $E_T^{\text{miss}}$  calculation was developed for W and Z decays by W mass measurements group [?]. This procedure based on the requirement of balance in transverse momentum of a W-boson and the initial (quark-gluon) state radiation.:

$$\vec{P}_T^W = \vec{P}_T^l + \vec{P}_T^\nu = \sum \vec{P}_T^{\text{ISRquarks,gluon}}, \quad (6.3)$$

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

$$E_T^{\text{miss}} = P_T^\nu = -HR + p_T^l \quad (6.4)$$

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:

$$HR = \sum_{i=0}^{N_{\text{topo}}} \vec{p}_T^{\text{topo}} \quad (6.5)$$

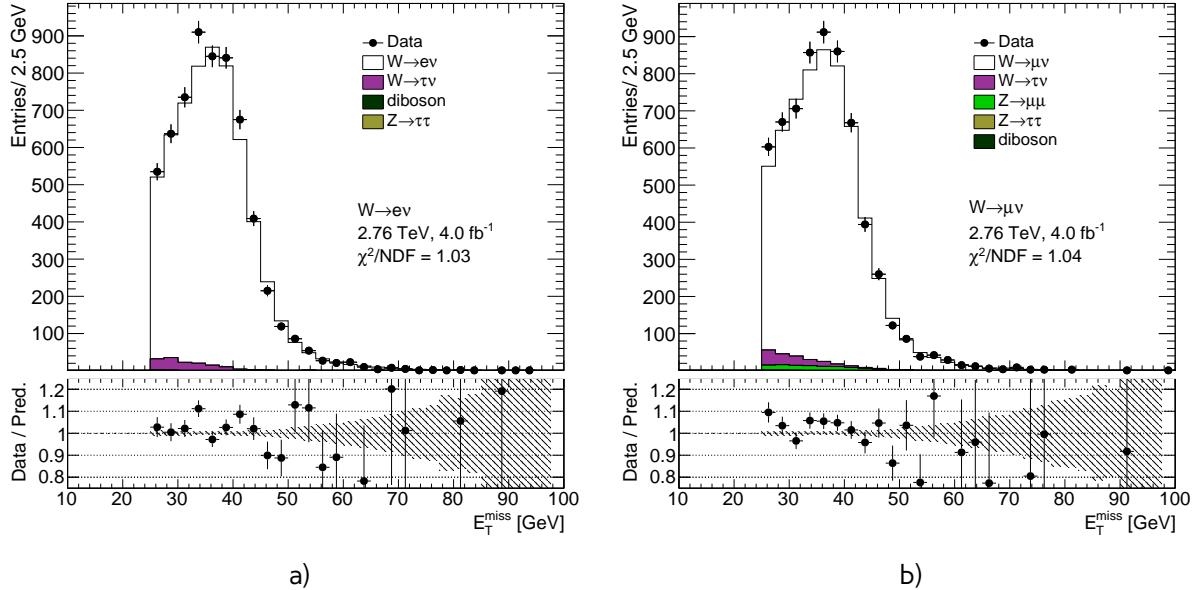


Fig. 6.3: Data and MC comparison for  $E_T^{\text{miss}}$  calculated from hadron recoil for a)  $W \rightarrow e\nu$  b)  $W \rightarrow \mu\nu$  events

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

$$\sum E_T = \sum_{i=0}^{N_{\text{topo}}} E_T^{\text{topo}} \quad (6.6)$$

- 67 To avoid double counting of lepton energy losses in calorimeter, the clusters inside cone with radius
- 68  $dR = 0.2$  are excluded from this calculation. To compensate soft activity inside this cone, clusters
- 69 are then compensated by replacement cone (Fig. 6.2). This cone is defined as cone at the same
- 70 pseudorapidity, but different  $\phi$ . It should be far from any other lepton and hadron recoil direction.
- 71 Each cone is then rotated to a direction of the original lepton direction. This definition is not taking
- 72 into account jet reconstruction aspects. This is allowing to get a better data MC agreement (Fig. 6.3).

## 74 Monte Carlo simulation

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

## 81 7.1 ATLAS chain of MC production

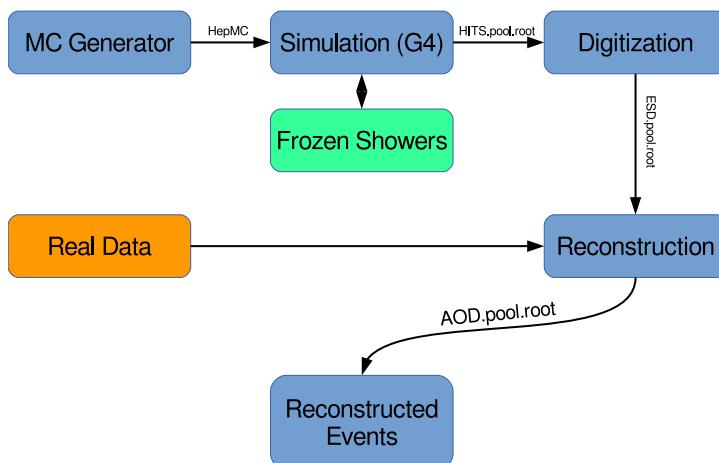


Fig. 7.1: Diagram of the ATLAS MC production chain

82 Monte Carlo method allows to perform different analysis, generate predictions for comparisons  
 83 with data, study detector or selection algorithms performance. All of these applications require  
 84 accurate MC predictions. Simulation software expects to use precise physics models for sampling  
 85 and have large enough statistics, to exclude statistical uncertainties (usually 5 times more, than  
 86 expected in a data). ATLAS simulation software is integrated into Athena framework and usually used  
 87 during large production of events. Simulation chain is generally divided into 4 main steps (Figure 7.1):

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

90 **Simulation** Simulation of energy depositions ("hits") which are produced by a final state particles.

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

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

96    Additionally, the pileup effects are added to MC by overlaying simulation of the hard interactions  
 97       with simulation of soft inelastic scatterings. This scheme allows to use computing resources more  
 98       efficiently, than with a single-step simulation, and simplifies software validation, since it is possible  
 99       to reuse files from previous stages. In the following sections event generation and simulation will be  
 100      described in more details.

101    

## 7.2 Event generators

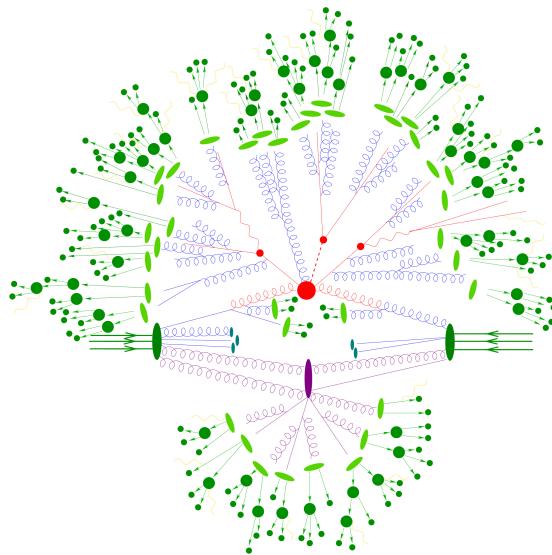


Fig. 7.2: Schematic view of a  $t\bar{t}H$  event produced in a pp-collision: the hard scattering is shown as a red blob with the solid and dashed lines as the resulting three particles. Independently happening multi-particle interactions are indicated by the violet blob. Parton showers are shown with curly lines. Hadronization yields hadrons as shown in light green, while the final state particle are dark green.

102    The outcome of the hard interaction could be simple scattering of the hadron elementary con-  
 103       stituents, their annihilation into new resonances or a combination of two. This can lead to a final  
 104       state with a large particles multiplicity. The main goal of event generator is to provide a complete  
 105       picture of this final states: description of the particle types and momentia on event-by-event ba-  
 106       sis. The factorisation theorem [?] allows to make event generation in independent stages, which are  
 107       dominated by different dynamics. Schematic plan of simulation of ttbar event is shown in Figure ??:

108    **Modelling of hard subprocess** Hard subprocess happens at the smallest times and distances,  
 109       where the colliding partons are considered free. Process of interest is simulated by selecting

110 production channels and calculating corresponding matrix elements (ME) in the desired level of  
 111 accuracy in perturbation theory . Most of the generators have leading order or next to leading  
 112 order ME in  $\alpha_s$ .

113 **Parton showering** Quarks and gluons from hard process can radiate secondary quarks and gluons,  
 114 resulting in the dozens of additional partons associated with the event. This process is  
 115 calculated as step-by-step evolution of momentum transfer scales from highest (hard subprocess), to the lowest (around 1 GeV). There is a possibility of double counting between showers  
 116 and hard subprocess. This can be avoided using matching approaches, for which higher order  
 117 corrections to ME are integrated with parton showers, or merging strategy, where jet resolution  
 118 scale is used as a threshold between matrix elements and parton showers.  
 119

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

123 **Modelling underlying event** Parallel to the main process other collisions of partons can occur,  
 124 called underlying event. These additional interactions can produce partons which contribute to  
 125 the final state. This is one of the least understood aspects of hadronic collisions.

126 The current analysis uses samples generated with the following generators:  
 127 Powheg [?] Powheg is a generator with NLO ME [?], that can be interfaced with other generators (such  
 128 as Pythia or Herwig) for higher precision of showering.  
 129 Pythia [?] Pythia is a general purpose generator for hadronic, hadron-lepton and leptonic collisions.  
 130 It can model initial and final state showers, hadronisation and decays, underlying event (via  
 131 multi-parton interactions). Pythia contains a library with around 240 processes with LO ME. It  
 132 uses Lund String model [?] for hadronisation.

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

137 Sherpa [?] Sherpa is an event generator, that uses tree-level leading order matrix element for a hard  
 138 scattering and featuring its own implementation of parton shower and hadronisation models.

139 Photos [?] Program used for generation of QED radiative corrections in W and Z decays.

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

## 141 7.3 Simulation in Geant4

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

146 Geant4 can simulate complex detector structures with sensitive detector material and corresponding  
 147 infrastructure. It can also calculate basic properties of materials, like radiation and interaction

length. For detector Geant4 stores "hits" information - snapshots of physical interactions. In Geant4 events and particles are simulated separately and each particle is moved in steps. Size of each step is chosen to preserve both CPU performance and required precision.

Physics interactions are 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 approximate and 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 uses one of this lists.

Most of the computing resources are taken by a mass MC production, required for each data taking periods. 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 replacing a complex magnetic field maps by a 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 spend on calorimeters. This is the motivation for development of fast calorimetry techniques.

There are two main methods used at 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 Chap. 8

170 

## Frozen Showers

171 As it was mentioned in a previous chapter, fast simulation techniques are the essential part of  
 172 Monte-Carlo production at ATLAS experiment. Typical time for a simulation of  $t\bar{t}$  event is around 1  
 173 minute, and most of the time is spent on a simulation of particle interaction in calorimeters. This  
 174 motivates a development of fast calorimetry techniques.

175 Frozen showers is currently the main fast calorimeter simulation approach used at ATLAS experiment.  
 176 In this chapter we will discuss main principles, difficulties and current developments in this  
 177 method.

178 Frozen shower method uses pre-simulated "frozen" showers instead of the full simulation. This  
 179 is allowing to reduce time spent on a simulation of a large amount of low energy sub showers.  
 180 This method is allowing to have a 25% speedup in simulation. It is required to have an in advance  
 181 generation of a libraries for each detector and particle used in this method. Each shower from a  
 182 library stores its lateral and transverse size and list of all energy depositions inside sensitive  
 183 material (hits) with information about their energy, position and time. During a simulation, if energy  
 184 of secondary electron falls below cutoff energy it is replaced by a shower from a library, as shown  
 185 on a Fig. 8.1. Main parameters used in ATLAS simulation are summarized in a Tab. 8.1.

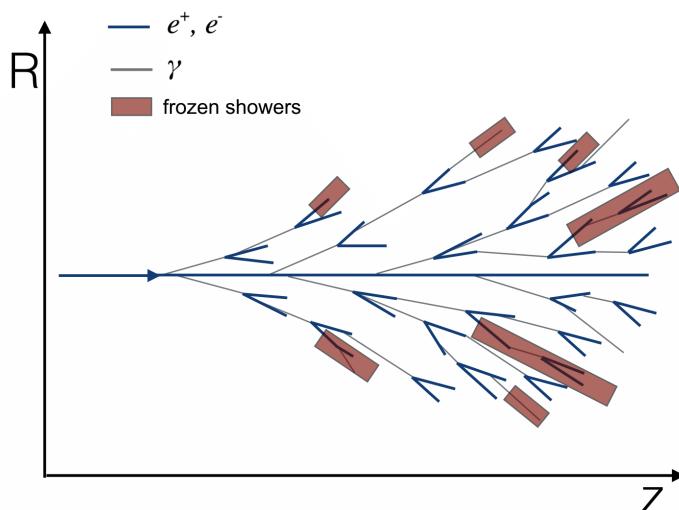


Fig. 8.1: Diagram showing the shower substitution of the low-energy particle, during the high-energy particle simulation. Some of the showers from a particles, substituted by frozen showers method marked by a red squares

Table 8.1: Main parameters used for the frozen shower libraries

| The general frozen showers parameters |  |
|---------------------------------------|--|
| Detectors used                        | FCAL1, FCAL2   |
| Type of the particle                  | photons, electrons, neutrons   |
| Energy range                          | $E_\gamma < 10 \text{ MeV}$ , $E_e < 1000 \text{ MeV}$ , $T_n < 100 \text{ MeV}$ |

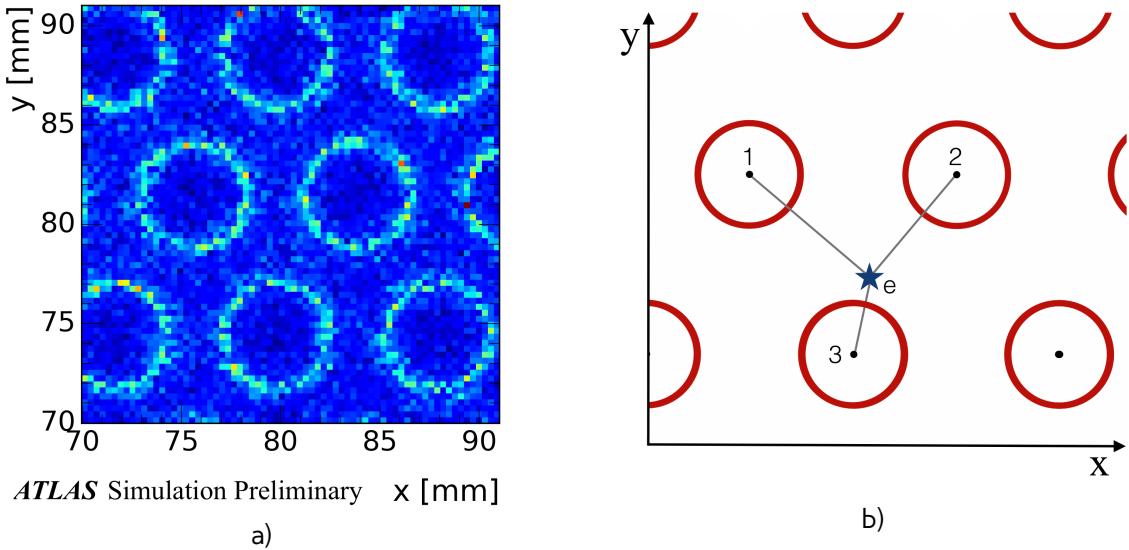


Fig. 8.2: a) Shower energy response histogram in the  $x$  vs  $y$  plane for electrons, generated with uniformly distributed  $x$  and  $y$  and energy less than 1 GeV. Light circles are corresponding to a showers, started inside a LAr gaps with on average higher energy response, while the dark parts are corresponding to dead material respectively with smaller sum of the "hits" energy. b) Distance to a closest rod center scheme, where  $d_{rod} = \min(d(1,e), d(2,e), d(3,e))$ . Rod centers and liquid argon gaps are shown by black and red circles respectively.

## 8.1 Problem description

Fast simulation of forward calorimeters (FCAL) is a complicated task due to its complex structure. As it was mentioned in a Sec. 5.2.1 FCAL consists of hexagonal absorber cells with anode tube and cathod rod in the cell center and liquid argon in the gap between rod and tube. In order to simulate resolution of high energetic electrons, good fast simulation technique should take this feature of large amount of non-uniformly distributed sensitive material.

Resolution of electron inside 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, causing large fluctuation of the energy loss. Resolution of high-energy electrons is mostly dominated by the constant term.

Fluctuations due to a detector design are visible in a simulation of small energy electrons, gener-

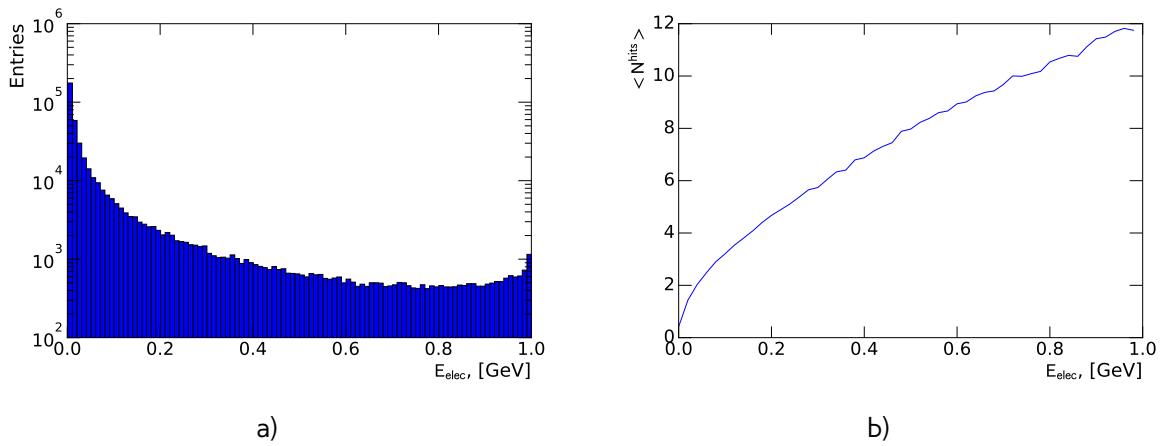


Fig. 8.3: Distribution a) electron energies and b) mean number of hits in a shower vs energy of electron for electrons with energy less than 1 GeV coming from initial electron with energy 1 TeV.

ated inside a different points in forward calorimeter. Shower energy  $E^{\text{shower}}$  distribution in the x vs y plane is showed on Fig. 8.2, where shower energy is defined as:

$$E^{\text{shower}} = \sum E_i^{\text{hits}}, \quad (8.2)$$

where  $E_i^{\text{hits}}$  is an energy of i-th shower deposit inside sensitive material. Periodic structure resembles the calorimeter design, where light circles are corresponding to gaps with liquid argon. It could be reduced to a 1-d problem by introducing  $d_{\text{rod}}$  distance to a closest rod center.

Typical electron substituted by a frozen showers coming from a simulation of high energy electrons have a relatively small energy (Fig. 8.3 a)). Mean number of depositions in a sensitive material in a "frozen" shower is around 5 and this value rises with electron energy (Fig. 8.3 b). Fig. 8.4 presents a distribution showers for electrons with energy below 1 GeV coming from initial electrons with energy 1 TeV in the distance to a closest rod center vs shower energy plane. Liquid argon gap is marked by a red lines. There is a clear difference in a showers energies between electrons born in a sensitive and dead material. Difference in a shower properties are also visible for number of hits (Fig. 8.5 a) and standard deviation energy of hits in shower (Fig. 8.5 b) distributions. Size of this differences depends on a electron energies and higher for a smaller energies (Fig. 8.6 a) and less significant for a higher energies (Fig. 8.6 b). This fact combined with energy distribution states an importance of proper simulation non-uniformities for showers coming from a small energy electrons.

On another hand, use of the frozen showers in a small energy region can be suboptimal because of the small number of energy depositions in a shower. For electrons with energy less than 30 MeV 90% of the showers has zero depositions and just 0.5% of showers are having more than 1 hits (Fig. 8.7). Below this energy single energy spot model have showed better performance in simulation.

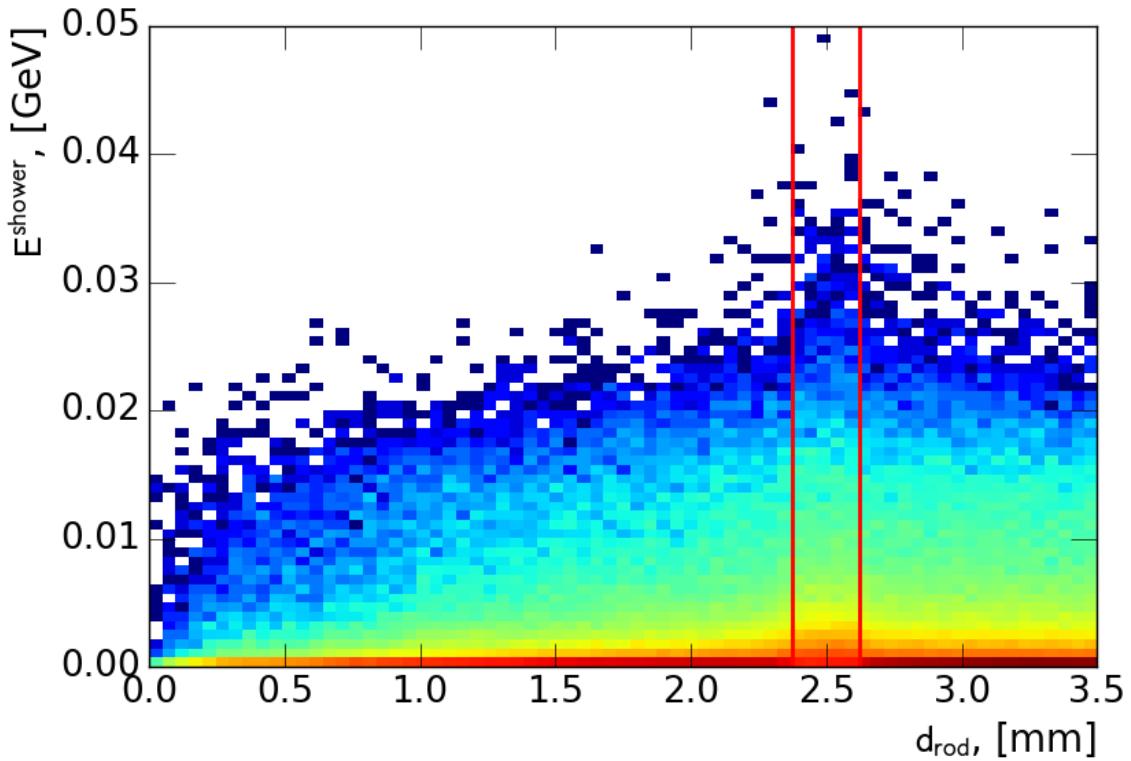


Fig. 8.4: Distribution of electron showers for electrons with energy less than 1 GeV coming from initial electron with energy 1 TeV in distance to a closest rod center vs shower energy plane. Position of a liquid argon gap is noted by a red lines. There is visible difference in shower properties between showers inside and outside of the liquid argon gaps

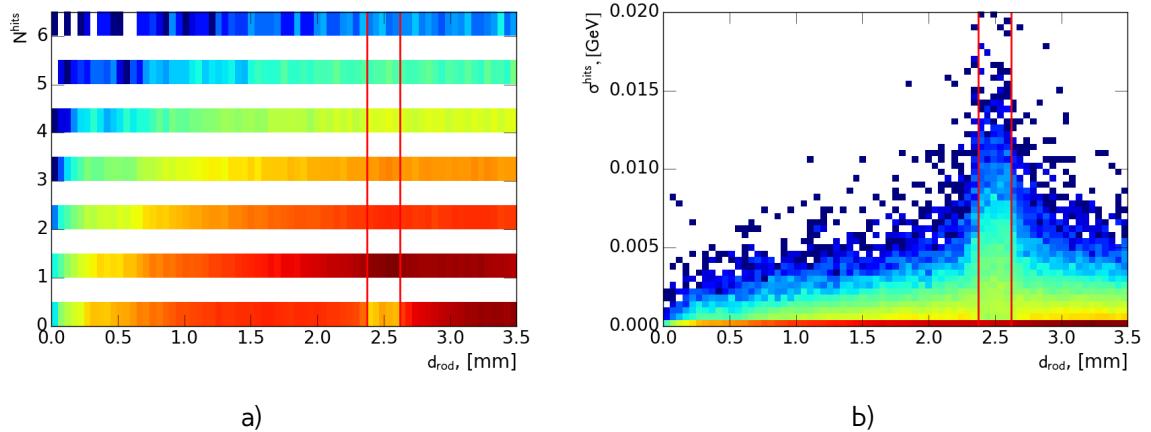


Fig. 8.5: Distribution of electron showers for electrons with energy less than 1 GeV coming from initial electron with energy 1 TeV in distance to a closest rod center vs a) number of hits in a shower plane and b) standard deviation of hits in a shower energy. Position of a liquid argon gap is noted by a red lines. There is visible difference in shower properties between showers inside and outside of the liquid argon gaps

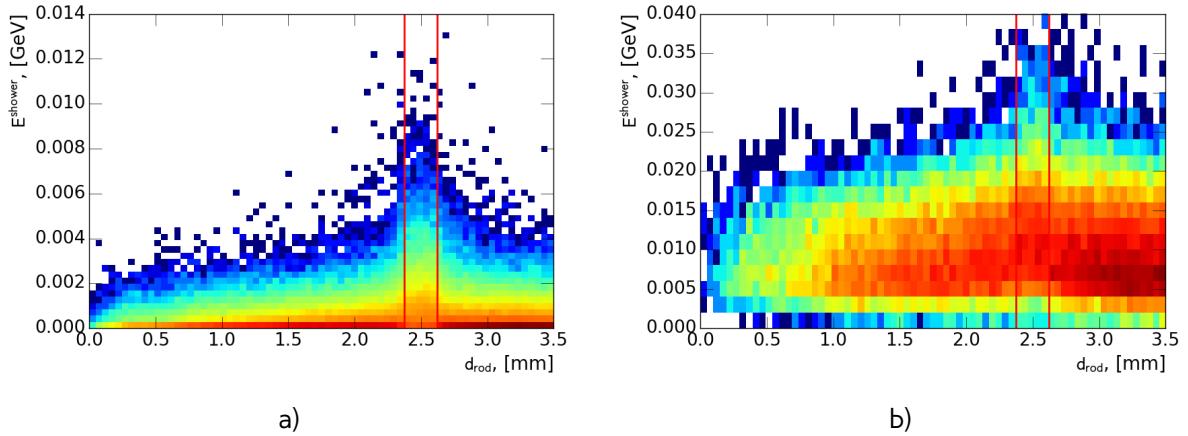


Fig. 8.6: Distribution of electron showers for electrons a) less than 100 MeV and b) higher than 300 GeV coming from initial electron with energy 1 TeV in distance to a closest rod center vs shower energy plane. Position of a liquid argon gap is noted by a red lines. Size of the difference in a shower properties depends on the energy of the electrons and higher for smaller energies

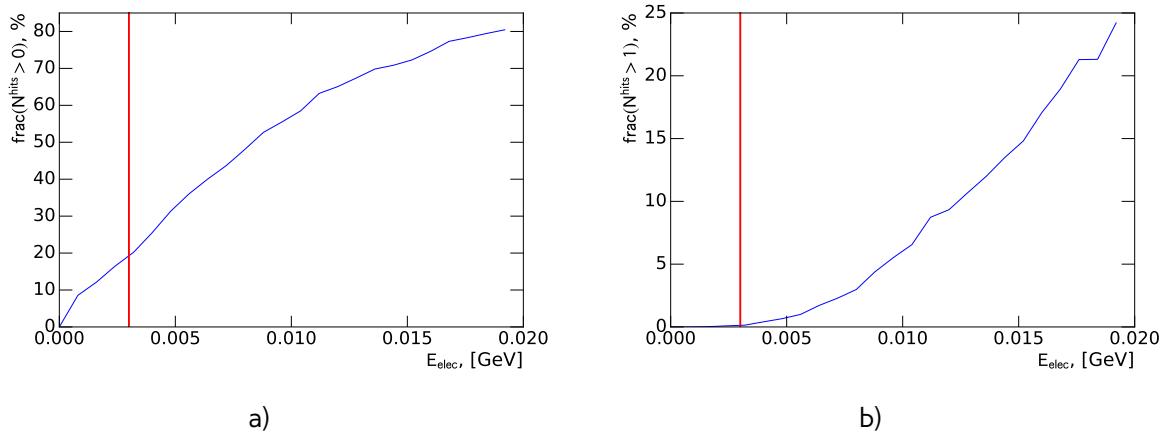


Fig. 8.7: Distribution of fraction of showers with a)at least 1 b) at least 2 depositions inside sensitive material depending on a initial electron energy. Red line denotes 30 MeV lower limit for a frozen showers method.

## 214 8.2 Generation and use in simulation

215 As it was mentioned in introduction frozen showers method consist of the 2 stages: generation of  
216 libraries and use in simulation. Generation needs to be repeated for each significant change in a  
217 physics processes description in Geant4 or in a description of detector. Showers are stored in a  
218 library in pseudorapidity and distance to a closest rod center bins, while energy remains unbinned.  
219 Distance binning was introduced to describe fluctuations from Sec. 8.1. Position of the liquid argon  
220 gap bin is corresponding to a real gap position.

221 In order to have proper energy distribution for a generation of library particles coming from  
222 a simulation of physical process (usually  $t\bar{t}$  or high energy electrons) are used. For each particle  
223 eligible for frozen showers use parameters are saved in a HepMC (reference) format for a later  
224 use. On a second stage, these particles are propagated through the calorimeter using standard  
225 ATLAS simulation infrastructure. Each hit is saved as a shower inside library in a corresponding  
226 pseudorapidity and distance bin.

227 Additionally, in order to save disc space as well as a memory consumption, hit information is  
228 compressed. This compression is done in a two steps:

229 **Hit merging** if the distance between any two hits is smaller, than a given parameter  $R_{min}$ , then  
230 hits are merged into one deposit at the energy weighted center of them. This process is done  
231 iteratively.

232 **Truncation** hits whose energies are below the fraction  $f$  of the total energy sum of all hits, are  
233 truncated. The energy of remaining hits is rescaled back to preserve the total deposited energy.

During simulation, if an energy of a particle falls below cut-off energy, the particle algorithm examines resulting shower containment. It checks that particle is far from the edges of calorimeter, so what shower will be by 90% inside calorimeter. This depends also on a energy of particle, because shower sizes are growing with energy. The algorithm searches for a shower with the closes energy in a corresponding pseudorapidity and distance bin. Shower is a rotated in a direction of particle. In order to correct differences in energy, each hit in a shower is scaled as:

$$E_{hit}^{new} = E_{hit} \cdot \frac{E_{part}}{E_{part,lib}}, \quad (8.3)$$

234 where  $E_{hit}$  is the energy of the hit,  $E_{part}$  is the energy of the particle and  $E_{part,lib}$  is the energy of  
235 the particle from a library. Then particle is removed and substituted by a resulting shower. Later,  
236 reconstruction algorithm uses hits from a frozen shower as a usual energy depositions in a sensitive  
237 material.

### 238 8.2.1 Libraries tuning

239 Frozen showers method gives a good agreement with a full simulation for a shower shape vari-  
240 ables(Sec. 8.4), however resolution of the reconstructed electrons is around 2 times smaller(Fig.  
241 8.8), than in a full simulation. It can be interpreted as a lack of the showers from liquid argon gap in  
242 a simulation. Most of this effect is coming from an electron libraries. This means that this libraries  
243 are requiring additional reconstruction-based tuning after generation.

244 Usual tuning consists of 2-step manual procedure:

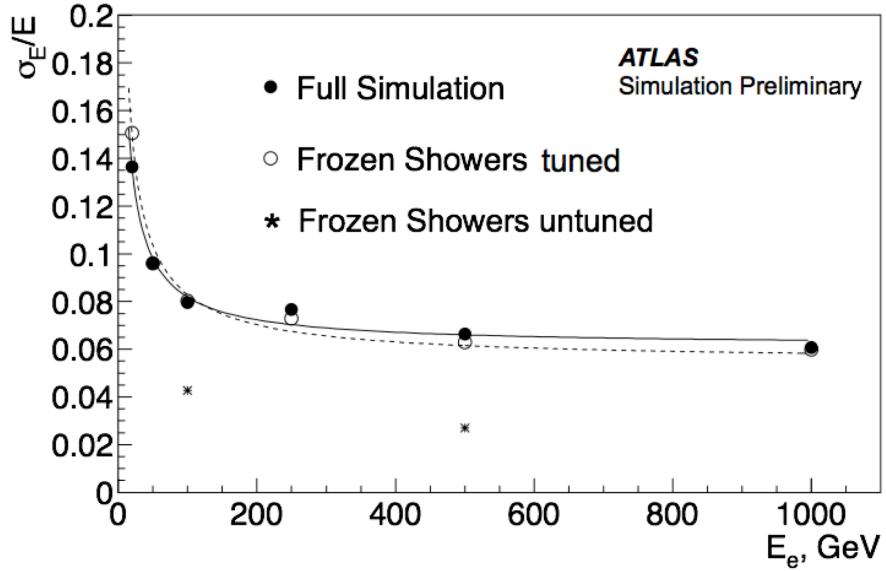


Fig. 8.8: Electron resolution for full simulation, tuned and untuned frozen showers. Electrons simulated with frozen showers libraries before tuning (star points) have twice smaller resolution, than an electrons from full simulation (circles). Tuning (black dots) is allowing to gain better agreement with full simulation.

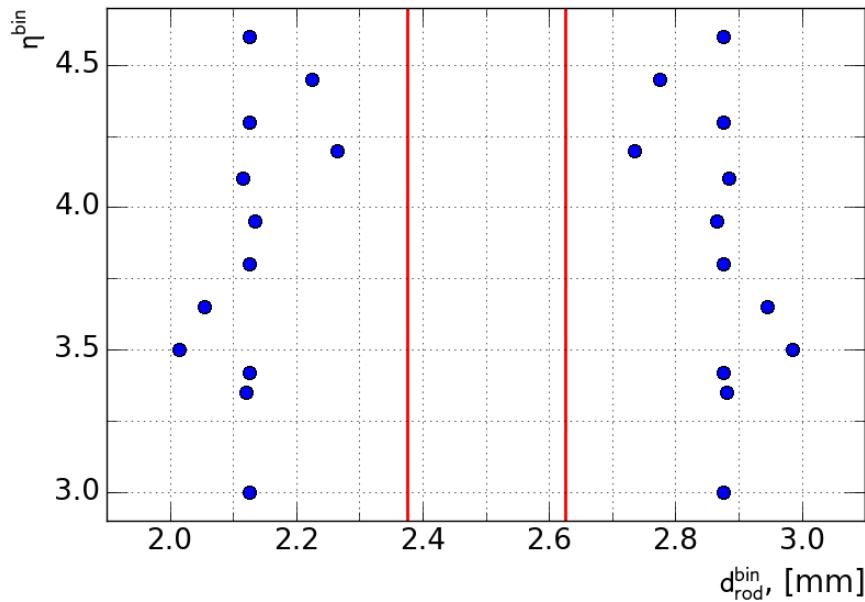


Fig. 8.9: Position of gap bins for different  $\eta$  bins in old libraries after tuning. Dots are corresponding to a limits of each bin. Red lines are denoting original position of bins, that are corresponding to a position of a liquid argon gap in the calorimeter.

245 **Changing bin width** On this stage width of the liquid argon bin is enlarged. This is causing larger  
246 size of the fluctuations, that leads to a higher resolution and a mean energy of reconstructed  
247 electrons

248 **Shower energy scaling** In order to correct introduced shift in a mean energy shower energy is  
249 reduced by rescaling all of the hits in a shower.

250 It is repeated iteratively in each pseudorapidity bin separately till the desired agreement is obtained.  
251 The resulting bin positions are shown on a Fig. 8.9. This method allows to have a relatively good  
252 agreement with full simulation (black dots on Fig. 8.8). However, it is necessary to repeat this  
253 procedure for each new libraries generation and requires significant tuning effort, that makes it not  
254 optimal.

## 255 **8.3 Machine learning based bin finding procedure**

256 Since frozen showers are planned to be used in Run-2 Monte-Carlo production, there is a need  
257 for a more automatic procedure of library generation with proper electrons resolution. One of the  
258 possible ways is to choose different position of liquid argon bin during libraries generation using  
259 machine learning tools. In this section automatic bin finding procedure will be discussed.

### 260 **8.3.1 Machine Learning**

261 Machine learning is a set of algorithms, what allows computers to learn and give a predictions without  
262 being specifically programmed. This is a modern field of computer science, that is wildly used in a  
263 different fields like computer vision, natural language processing, data science etc. There are two  
264 main types of machine learning algorithms: supervised, where example of desired output is given by  
265 the "teacher" and the goal is to learn a general rule, that maps inputs to outputs and unsupervised  
266 learning, then there are no labels given to algorithm, and algorithms is discovering hidden patterns  
267 in data. Initial data parameters of interest, that are used in algorithm to learn are called features.

268 Machine learning algorithms can be used for solving a classification problem, where each event  
269 should be identified to one of the specified classes. Since the first introduction of the machine  
270 learning classifiyng algorithms called perceptron by a Rosenblatt [\[ref\]](#) many different algrothims  
271 have been invented. In this analysis decision trees and support vector machines have been used.

### 272 **Binary decision trees**

273 Binary decision trees are one of the most commonly used machine learning algorithms for a clas-  
274 sification problems in a particle physics. Schematic representation of this algorithm is shown on a  
275 Fig. 8.10 a. Each node is corresponding to the one of the internal input variables and connected to  
276 two branches, that are split in the respect to the a variable. The first node is called a root node.  
277 Depth of the tree is a number of edges from the node to the tree's root node. Leaf node represents  
278 classification or decision. For each node the best feature is selected and it's cut value are obtained  
279 by calculating each possible variation in the feature set and then ranking them. One of the main  
280 advantages of the decision trees is a simplicity of visualization and interpretation.

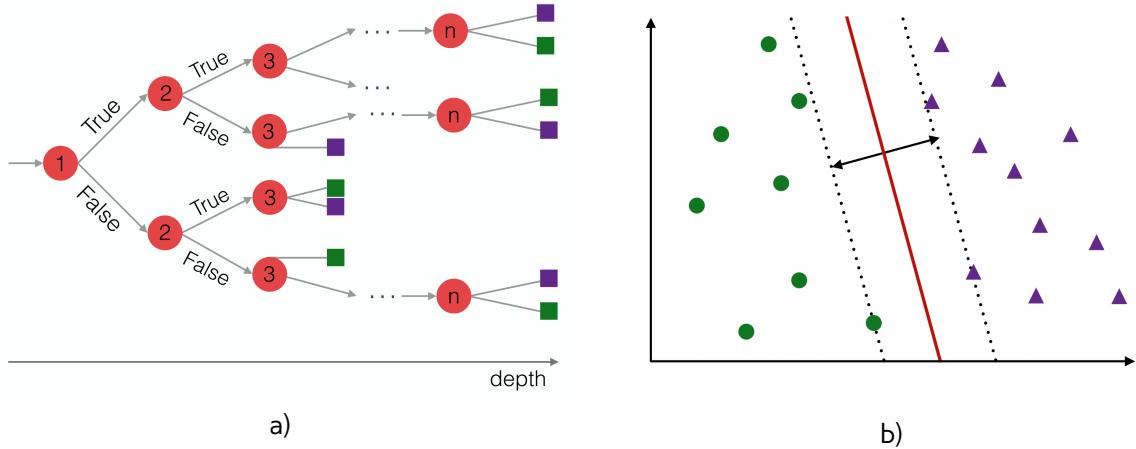


Fig. 8.10: Schematic representation of machine learning algorithms, used in the analysis for a classification. Green figures are representing first class of the events, whereas violet ones belong to a second class. a) Representation of a binary decision tree structure: red circles are corresponding to a node, that are split in the respect to the one of the features. Squares represent leafs, where all of the events are classified to a certain class. Depth of the tree is calculated as a number of edges from the node to the trees first node. b) Representation of the SVM algorithm. Dividing hyperplane is shown by a solid line. The dashed lines represent the maximum margin boundaries

### 281 Support vector machines

Support vector machines (SVM) is a supervised machine learning algorithm which can be used for classification problems. In this algorithm each event is represented in a  $p$ -dimensional parameter space. Classification is performed by finding a the hyper-plane that differentiate the two classes with the largest separation (Fig. 8.10 b). The hyperplane can be written as the set of points  $\vec{x}$  in a parameter space satisfying:

$$\vec{w} \cdot \vec{x} - b = 0, \quad (8.4)$$

282 where  $\vec{w}$  is the normal vector to the hyperplane and the parameter  $\frac{b}{\|\vec{w}\|}$  determines the offset of  
283 the hyperplane from the origin along the normal vector  $\vec{w}$ . The margin hyperplanes are described  
284 by equations:

$$\vec{w} \cdot \vec{x} - b = 1, \quad (8.5)$$

$$\vec{w} \cdot \vec{x} - b = -1, \quad (8.6)$$

285 where  $\frac{2}{\|\vec{w}\|}$  is the distance between these 2 hyperplanes, so planes with the maximum margin between  
286 should have the minimum  $\|\vec{w}\|$ .

287 Because we want to prevent each point to fall into the margin, we following constrain should be  
288 satisfied:

$$\vec{w} \cdot \vec{x} - b \geq 1 \text{ where } y_i = 1, \quad (8.7)$$

$$\vec{w} \cdot \vec{x} - b \leq -1 \text{ where } y_i = -1, \quad (8.8)$$

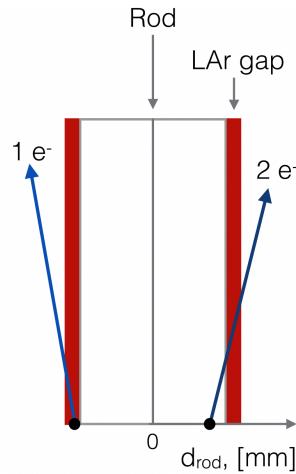


Fig. 8.11: Schematic representation of the model. Electron 1 is created in a liquid argon gap. Electron 2 is created near liquid argon gap and crosses it. This causing a smearing of sensitive material showers distribution. Electrons created in a sensitive material tend to create more energetic showers, than electrons from a dead material. However, electrons, shown on this scheme, may give similar shower and therefore be not distinguishable.

These equations can be rewritten as:

$$y_i(\vec{w} \cdot \vec{x} - b) \geq 1 \quad (8.9)$$

It is also possible to construct a non-linear classifier by replacing dot-product with a different *kernel* function. In this thesis, a radial basis function (RBF) kernel is used:

$$K_{rbf}(\vec{x}_i, \vec{x}_j) = e^{-\gamma|\vec{x}_i - \vec{x}_j|^2} \quad \gamma > 0, \quad (8.10)$$

where parameter  $\gamma$  adjusts the width of the kernel.

### 8.3.2 Model description

As it was mentioned in a previous sections, modules in FCAL are consisting of different types of material and showers started inside dead material are usually having smaller energies, than a sensitive material once. However, electron, created in a dead material, can cross a liquid argon gap and give a hit there (Fig. 8.13). It was decided to treat this electrons together with electrons created in sensitive material, and call a caused showers as a sensitive material showers. Oppositely, showers that did not crossed a liquid argon gap, are called a dead material showers. This model leads to a bigger gap width by a definition.

### 8.3.3 Interpretation of results

. As it was mentioned before, that there are two type of material used in a FCAL. Showers within them are giving different response, what is affecting overall reconstructed electron energy resolution.

At the first generations distance bin have been corresponding to LAr gap or dead material positions. During tuning bin with LAr was enlarged to gain a better agreement with full simulation. So,

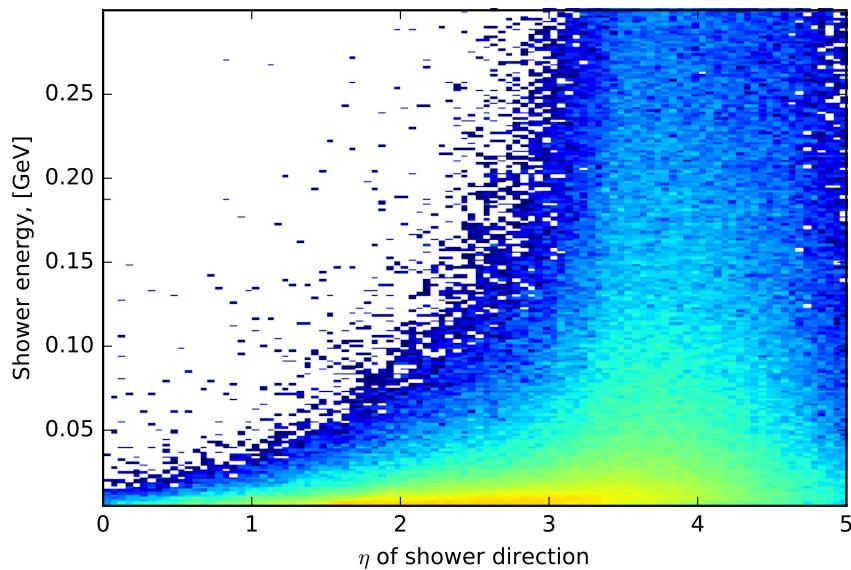


Fig. 8.12: Distribution of showers used in production of 1000 GeV electrons on shower energy vs  $\eta_{momentum}$  plane.

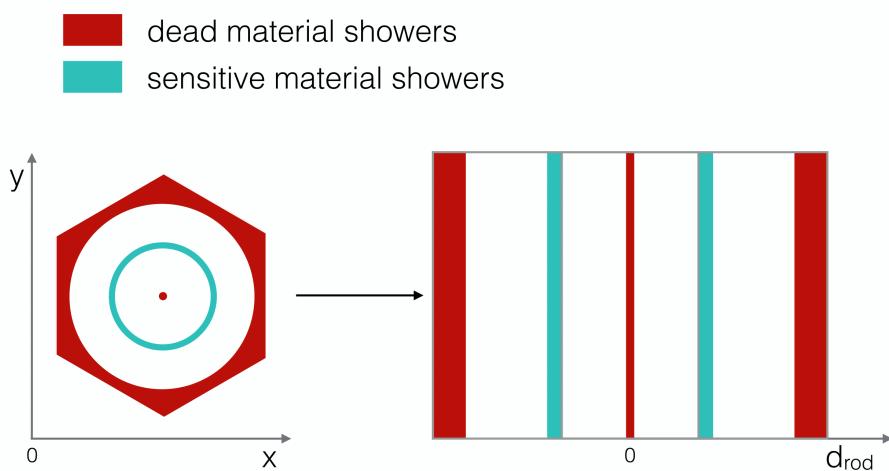


Fig. 8.13: Schematic representation of preselected data for a first classifier in  $x$  vs  $y$  plane and distance plane. Electrons, created near rod center and on the borders of the module have low probability to cross the sensitive material, while oppositely all of the electrons created inside liquid argon gap are sensitive material showers.

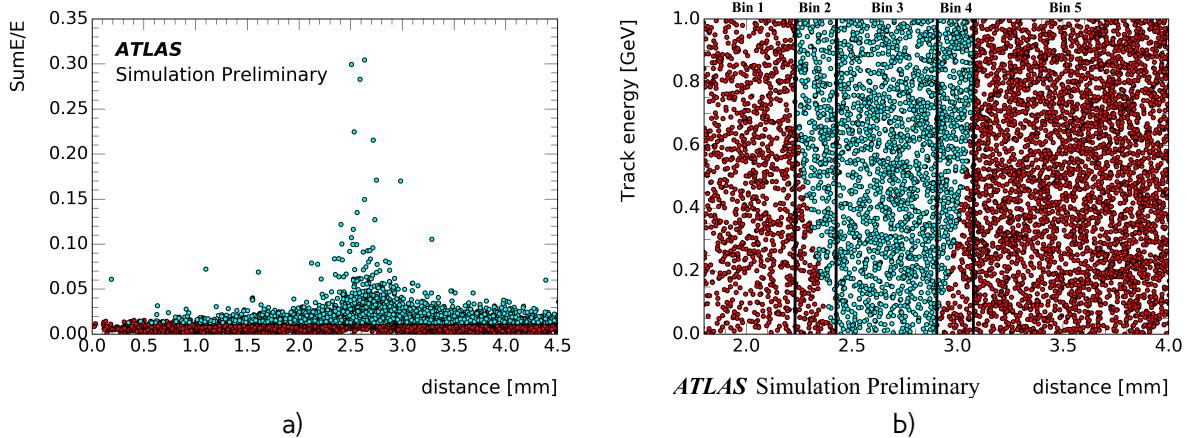


Fig. 8.14: Results of machine learning for a) first classifier b) second classifier. Cyan dots are corresponding to sensitive material showers, red - dead material showers

303 one of the basic ideas to improve frozen showers performance is to change a size of LAr gap in a  
 304 library generation.

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

#### 310 Machine Learning

311 From a geometrical point of view, one of the main parameter is a direction of the shower. Eta  
 312 momentum distribution is showed on a Figure 8.12 . Most of the showers are collinear to an electron  
 313 direction. Because of this it was decided to use as a training sample simulation results for electrons  
 314 with energies less than 1 GeV and momentum uniformly distributed between eta 3.0 and 4.0. This  
 315 allowing to study equally low and high energy showers equally.

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

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

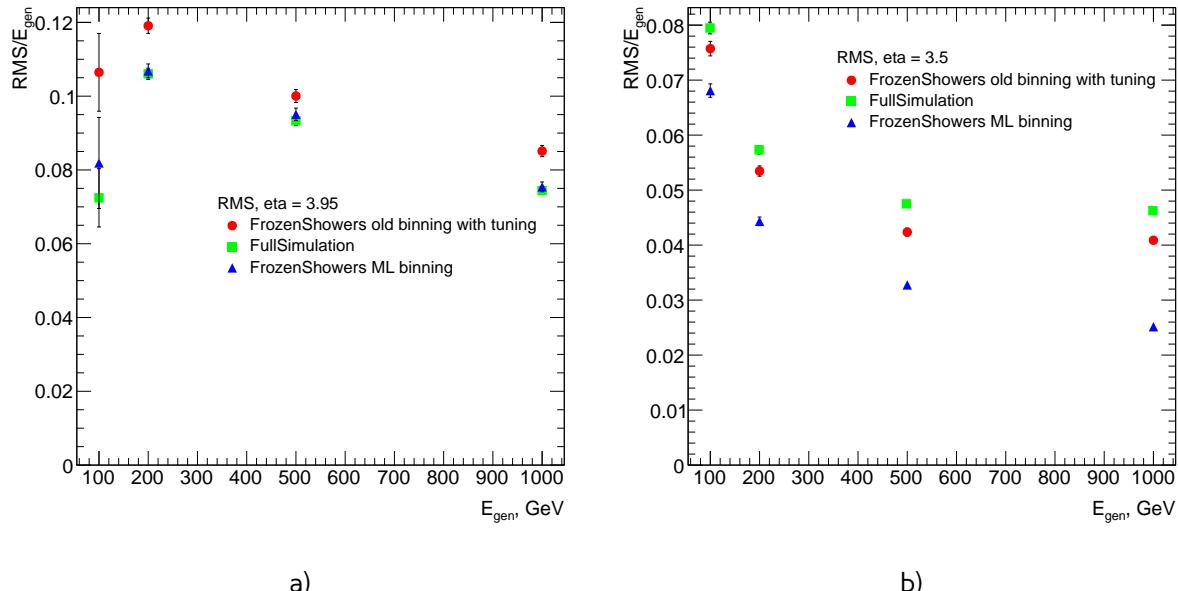


Fig. 8.15: Resolution of reconstructed electrons for full simulation, new libraries with ML binning and old tuned libraries with original binning for a) eta = 3.95 b) eta = 3.5

333 of outlying eta bins.

## 334 8.4 Validation of the new libraries

## 335 8.5 Plans for a future developments



336

# Chapter 9

## 337 **Data and Monte-Carlo samples**

### 338 **9.1 Data sample**

### 339 **9.2 Monte-Carlo samples**



340

## **Part IV**

341

# **The Measurement**



342 Chapter **10**

343 **Selection**

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

348 **10.1 Data quality cuts**

Table 10.1: Selection criteria

| Event selection                             |                              |
|---|------------------------------|
| Single lepton trigger                       |                              |
| Good Run List                               |                              |
| Reject events with LAr errors               |                              |
| Number of tracks at primary vertex $\geq 3$ |                              |
| Electron Selection                          |                              |
| EF_e15_loose1                               | EF_mu10                      |
| $P_T^l > 20\text{GeV}$                      | $P_T^l > 20\text{GeV}$       |
| $ \eta^l  < 2.47$                           | $ \eta^l  < 2.5$             |
| excluding $1.37 <  \eta^l  < 1.52$          |                              |
| OQ cut                                      | staco reconstruction chain   |
| Medium electron identification              | Medium muon identification   |
| $P_T^{cone,20} < 0.1$                       | $P_T^{cone,20} < 0.1$        |
| W boson selection                           |                              |
| $E_T^{miss} > 25 \text{ GeV}$               |                              |
| $M_T^W > 40 \text{ GeV}$                    | $66 < M^Z < 116 \text{ GeV}$ |
| Z boson selection                           |                              |

349 Data taking conditions are important in the analysis because of the possible biases. In order  
350 to preserve high data quality some events must be rejected. The reason may be unstable beam  
351 conditions, disabled parts of the detector or events with too many noisy cells. Number of runs, than  
352 can be used for the analysis are stored in a so-called Good Run List (GRL), which in the addition to  
353 the run numbers contains information about luminosity blocks.

354 Events, where LAr calorimeter was malfunctioning are excluded by LAr quality criteria. Furthermore  
355 required to have at least one primary vertex from a hard scattering with at least 2 tracks, that are  
356 reconstructed from this vertex.

357 Online selection of events is based on a single lepton trigger, depending on a lepton flavor. For  
358 electron analysis EF\_e15\_loose1 trigger is used, which records electrons with  $E_T > 15$  GeV. This trigger  
359 uses additional "loose" isolation requirements to exclude jets, that are misidentified as electrons. In  
360 the muon channel lowest single lepton trigger is EF\_mu10. It records events with muons  $E_T > 10$  GeV.

## 361 **10.2 Lepton quality cuts**

362 Both analyses use similar selection criteria, applied on a leptons. The leptons must satisfy require-  
363 ment  $P_T^l > 20$  GeV and match to the event trigger.

364 Electron candidates are required to be within pseudorapidity range  $|\eta^l| < 2.47$ . The electron candi-  
365 dates found within the transition region between the barrel and the endcap electromagnetic calorime-  
366 ters,  $1.37 < |\eta^l| < 1.52$ , are removed. Additionally, for a better multijet background rejection medium  
367 identification and  $P_T^{cone,20} < 0.1$  criterias are applied. The object quality (OQ) cut applied in order to  
368 remove events from runs where there dead front end boards in the calorimeter.

369 Muons have to satisfy the following criteria: they should be reconstructed by a staco algorithm in  
370 a muon spectrometer and fall within range  $|\eta^l| < 2.5$ . Set of medium requirements is applied. They  
371 must also satisfy  $P_T^{cone,20} < 0.1$  isolation criteria.

## 372 **10.3 Boson selection**

373 The events, containing W boson candidates are required to have exactly one lepton, fulfilling the  
374 lepton selection. Missing transverse energy, used as a proxy for a neutrino from W decay is required  
375 to be  $E_T^{miss} > 25$  GeV. The transverse mass, calculated from the lepton and missing transverse energy  
376 as in Eq. ?? has to be bigger than 40 GeV ( $M_T^W > 40$  GeV).

377 The reconstructed lepton pair in case of Z boson analysis is required to have the invariant mass  
378 between 66 and 116 GeV. Both upper and bottom limits allow to exclude regions with high background  
379 contamination.

380 The full set of cuts is summarized in a Tab. 10.1.

## 381 **10.4 Cut flow**

382 The effect of each selection can be studied using cut-flows, which show the number of events  
383 passing each set of cuts in a sequential order. Cut flows for W and Z analysis are shown in a Tab.  
384 ?? and Tab. ?? respectively.

385

# Chapter 11

## Monte Carlo corrections

387 Monte Carlo plays important role in the cross-section measurement. It is constantly being improved,  
 388 in order to obtain a better precision in data description. Part of these corrections have been described  
 389 in Chap. 7. Unfortunately, not everything can be taken into account during simulation itself. This  
 390 leads to differences between data and Monte Carlo, that need to be accounted for. There are two  
 391 possible methods to correct Monte Carlo without regenerating it. First one is to apply event weights,  
 392 so that each MC event can contribute by a non one entry to a histogram. This procedure called  
 393 event reweighting. Second one is MC smearing. It uses random numbers to alter the reconstructed  
 394 4-vectors. This chapter describes all additional corrections, that have been applied on MC samples in  
 395 this analysis. All of these correction are introducing additional systematic error, that will be discussed  
 396 in the Chap. 15.

### 11.1 Lepton efficiency corrections

398 The efficiency of lepton detection at ATLAS detector can be divided into three components:

- 399 • The reconstruction efficiency  $\epsilon_{rec}$  is a probability to reconstruct lepton as a lepton of this  
 400 flavor.
- 401 • The identification efficiency  $\epsilon_{id|rec}$  is the probability that a reconstructed lepton survives iden-  
 402 tification requirements.
- 403 • 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:

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

404 All these efficiencies are measured using tag-and-probe method in  $Z \rightarrow ll$  decays. One of the  
 405 leptons from Z boson, called "probe", is initially selected with all of the cuts, except the one under  
 406 study. Second one, called "probe", satisfies more tighter selection with some additional cut, e.g.  
 407 trigger matching.

408 The reconstruction efficiency is associated with the algorithm used in the event reconstruction  
 409 process. This is causing differences between electrons and muons efficiencies. In the electron case  
 410 the reconstruction efficiency is depicted as a probability to reconstruct an electron which has de-  
 411 posited its energy in electromagnetic calorimeter cluster as an electron candidate.

Muon reconstruction efficiency is given by:

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

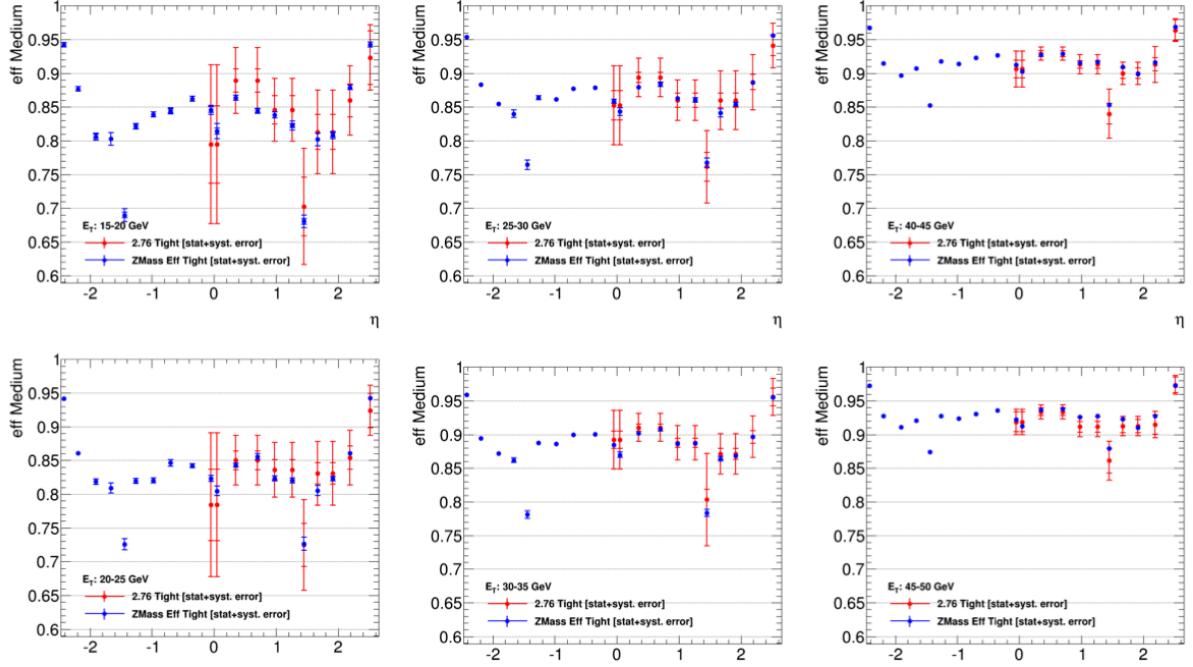


Fig. 11.1: Comparison of electron efficiencies as calculated for 8TeV (blue points) and 2.76TeV (red points) for MC simulation. Efficiencies are shown as a function of pseudorapidity ( $\eta$ ) for different electron  $E_T$  bins. Both statistical and systematic uncertainties are shown.

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

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

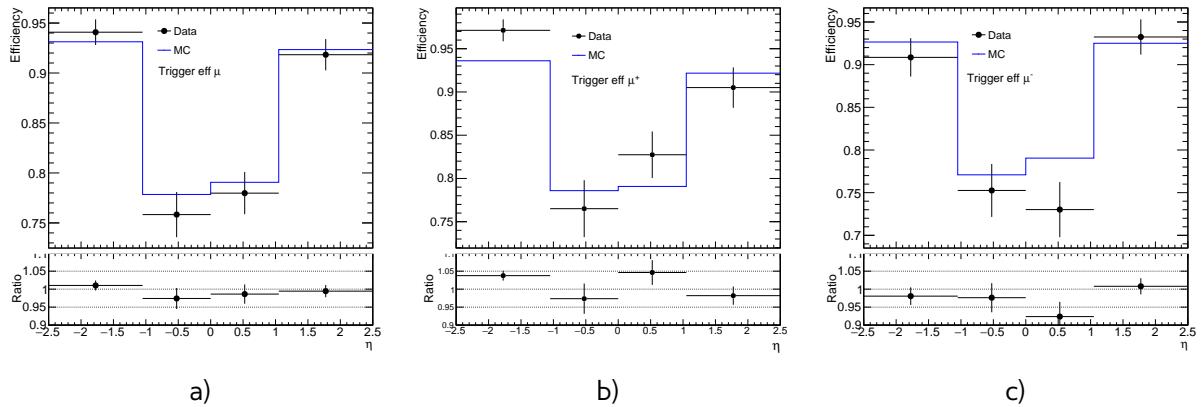
$$SF_{reco,id,trig} = \frac{\epsilon_{reco,id,trig}^{data}}{\epsilon_{reco,id,trig}^{MC}} \quad (11.3)$$

417 The scale factors are calculated in a  $P_T^l$  and  $\eta^l$  bins and have associated statistical and systematic  
 418 uncertainty components. The statistical component is connected to a size of  $Z \rightarrow ll$ , which is in  
 419 our case is around 500 event per each lepton flavor. This makes statistical error the dominant one  
 420 and means that precise calculation of scaling factors based on this data is difficult.

421 It is possible however to use scale factors for 8 TeV 2012 data. The main difference between these  
 422 data samples are center of mass energy and the pile-up conditions (10 in 2012 and less than 1 in 2013).  
 423 This effects have been studied on a  $Z \rightarrow ee$  sample. Fig. 11.1 shows the scale factors for different  
 424  $P_T^l$  ranges as a function of  $\eta^l$ . The differences in the scale factors are negligible and fully covered  
 425 by the statistical errors. This justifies the usage of 8 TeV scaling factors with increased uncertainty  
 426 (that is considered to be fully statistical) in the analysis at 2.76 TeV.

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 efficiencies distribution for a)  $\mu$  b)  $\mu^+$  c)  $\mu^-$  as a function of pseudorapidity

### 11.1.1 Muon Trigger SF

Unfortunately, single muon trigger haven't been present in the 2012 data, so muon trigger scale factor had to be derived from the 2.76 TeV data. The size of the Z sample is not big enough to make tje scale factors in  $P_T^l$  and  $\eta^l$  bins.

Since the  $P_T^\mu$  cut is significantly higher, than the trigger threshold, efficiency  $P_T^\mu$  dependency can be considered flat. On the another hand,  $\eta$  dependence are expected. Binning in  $\eta$  is motivated by a detector construction. Muon trajectory is bend in a magnetic field. That can lead to small differences in a trigger efficiencies for different muon charges. Possible charge dependency of the scale factors have been also studied.

Trigger efficiencies for data and MC in  $\eta$  bins are shown in a Fig. 12.11. Total scale factors are presented in Tab. ???. Scale factors for  $\mu^+$  and  $\mu^-$  are more than  $3\sigma$  away from each other, that is a clear indicator of a charge dependency.

Effect of applying different scale factors on muon for W analysis is shown on Fig. 11.3 - 11.5. Best agreement with data is achieved by applying single bin scale factor. This motivates a choice of single bin charge dependent scale factor for this analysis.

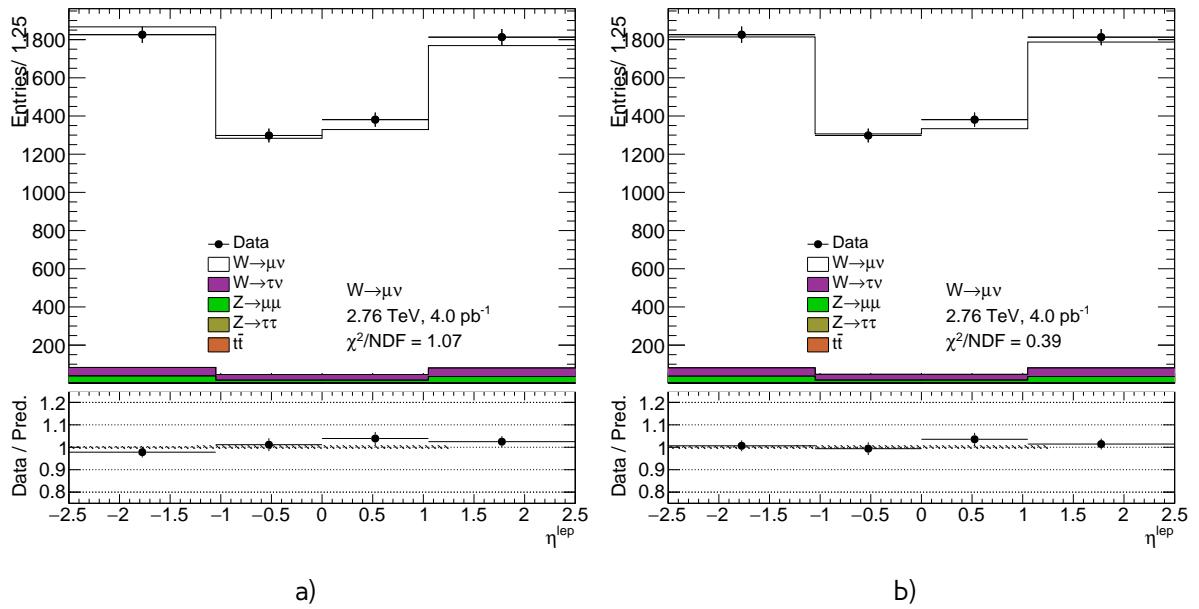


Fig. 11.3: Muon pseudorapidity distribution from the  $W \rightarrow \mu\nu$  selection with a) binned b) not binned charge dependent trigger scale factor applied

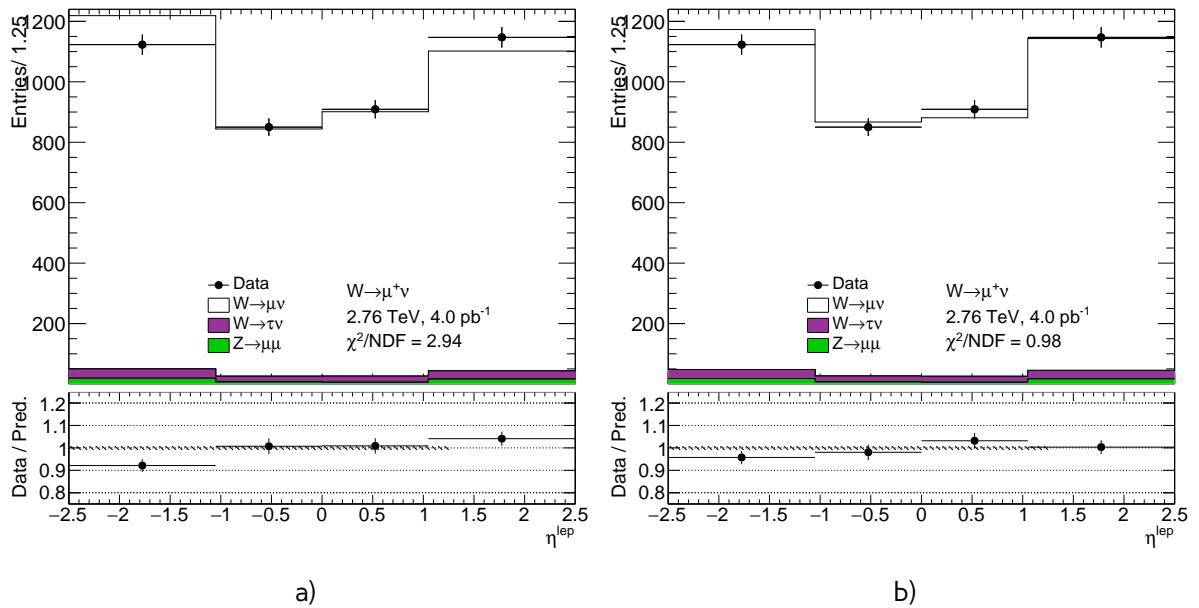


Fig. 11.4: Muon pseudorapidity distribution from the  $W \rightarrow \mu^+\nu$  selection with a) binned b) not binned charge dependent trigger scale factor applied

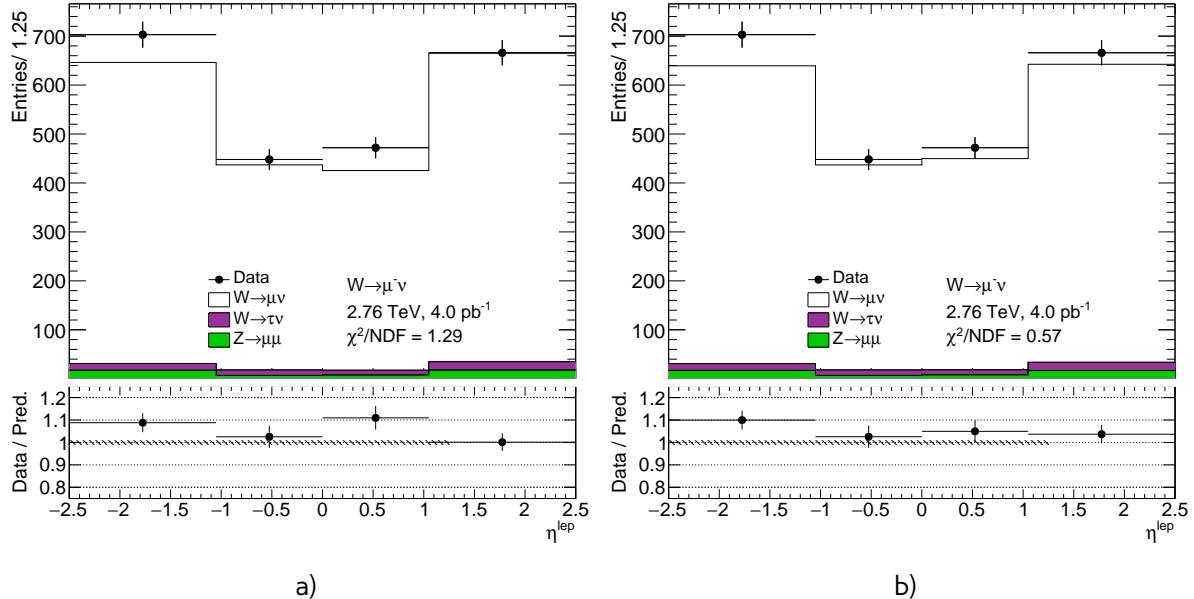


Fig. 11.5: Muon pseudorapidity distribution from the  $W \rightarrow \mu^- \nu$  selection with a) binned b) not binned charge dependent trigger scale factor applied

## 11.2 Electron energy scale and resolution

The reconstructed electron clusters energy tend to be shifted in comparison to the true energy of the initial electron. The correction of this shift is done in both data and MC as a 3 step procedure:

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

The 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  $\eta$ . The effect of this miscalibration on a reconstructed mass of Z boson neglecting second order terms is:

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

where  $m_{i,j}^{data}$  and  $m_{i,j}^{MC}$  are reconstructed mass of Z boson in a  $i$  and  $j$  bins in  $\eta$  for data and MC respectively.

It is also needed to correct the difference in the electron resolution. It can be described by Eq. 8.1. It is assumed, that sampling and noise terms are well modeled by MC and the main difference is

coming from a constant term. So, the electron resolution correction can be written as:

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

where  $c_i$  is an  $\eta$  dependent relative resolution correction. Similarly to an energy scale correction it is possible to derive resolution correction factor by a comparison of  $m_{i,j}^{data}$  and  $m_{i,j}^{MC}$  distributions.  $\alpha_i$  and  $c_i$  correction values are obtained via the  $\chi^2$  fit on a invariant mass electrons for data and MC. The resulting energy scale is applied on a data, while the resolution is corrected in MC. The resulting scale is validated using a  $J/\psi \rightarrow ee$  and  $Z \rightarrow ee\gamma$  samples.

### 11.3 Muon momentum correction

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

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

The first term origins from the fluctuations of the energy loss in the transversed material. The second term is coming from the magnetic field inhomogenities and the local displacements. Third term describes the intrinsic resolution effects.

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

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

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

This leads to a total correction formula:

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

where  $g_m$  are normally distributed random variables with mean 0 and width 1. Due to small amount 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 effect of the MS is corrected on a simulation level by adding a random smearing to the alignment constants. This allows to set  $\Delta r_2^{MS}(\eta, \phi)$  to 0 during a fit.

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

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

which represents the  $p_T$  imbalance between ID and MS.

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

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

469 where the weight  $f$  is derived from the MC.



470

# Chapter 12

## Hadron recoil calibration

472  $E_T^{miss}$  affects significantly the W boson measurement, so it is important to have a solid understanding  
 473 of possible sources differences in the hadron recoil reconstruction in 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 in Fig. ???. This projection can be divided into perpendicular  $u_\perp$  and parallel  $u_\parallel$  component as follows:

$$u_\parallel = \vec{v}_{xy} \cdot \vec{H}R \quad (12.1)$$

$$u_\perp = v_x \cdot HR_y - v_y \cdot HR_x, \quad (12.2)$$

474 where  $\vec{v}_{xy}$  is a unitary vector along the transverse component of a vector boson momentum and  $v_x$   
 475 and  $v_y$  are its projections on x and y axis respectively. In the case of the true kinematics  $u_\parallel = p_T^{bos}$   
 476 and  $u_\perp = 0$ . However the calorimeter resolution is causing relatively wide distributions for these  
 477 projections. The parallel component  $u_\parallel$  is sensitive to a possible bias in the hadron recoil, while  
 478 the perpendicular  $u_\perp$  can be used for determination of the resolution discrepancies. The mean  
 479 and the width of these distributions can depend on different variables, such as a mean number of  
 480 interactions in event, hadronic activity, boson  $P_T^{bos}$  etc.

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

488 First step in a hadron recoil calibration procedure aims to correct differences in a pile-up modelling  
 489 in the event. Additional interaction can have a significant effect on a  $E_T^{miss}$  and  $\sum E_T$  distributions.  
 490 It is usually accounted scaling average number of interaction per bunch crossing to match a data.  
 491 However, ATLAS simulation is suited for high pile-up runs, so this quantity is modelled discretely  
 492 in case of 2.76 TeV analysis (Fig. 12.1), what makes the corrections to match data impossible.

493 The combined Z and W boson determination procedure have been used. This section describes  
 494 a procedure of calibrating bias and resolution mismodelling in a hadron recoil, that was adapted for  
 495 2.76 TeV data.

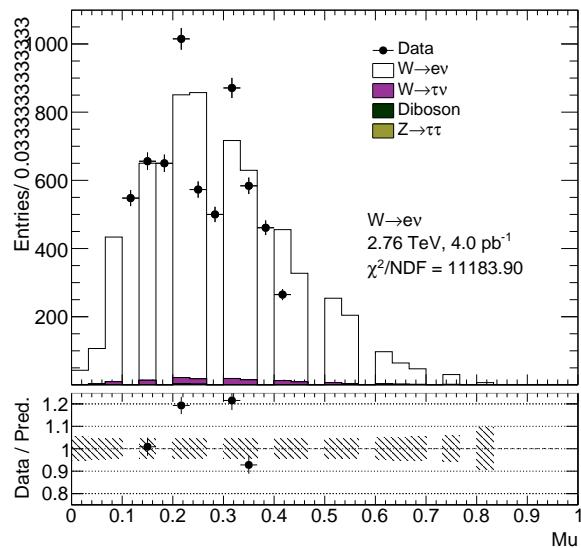
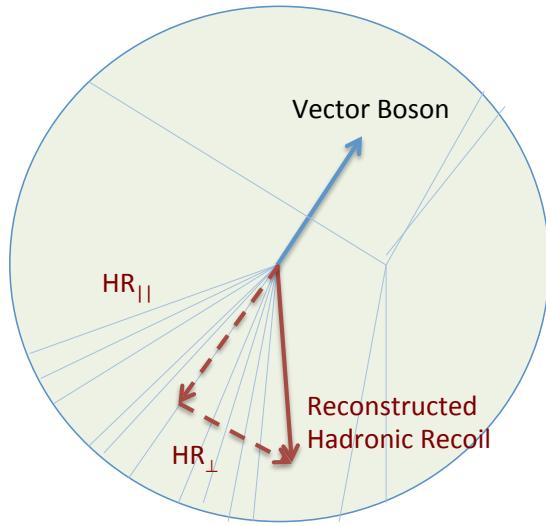


Fig. 12.1: Pileup

## 12.1 Hadron recoil resolution correction

Event activity plays an important role in a  $E_T^{miss}$  reconstruction. Since  $\sum E_T$  and hadron recoil resolution values are correlated, the possible mismodelling of event activity can lead to a difference between data and monte carlo  $E_T^{miss}$  resolution (Fig. 12.2). There are two possible ways of resolution correction in a 2.76 TeV data. It could be corrected by reweighting  $\sum E_T$  distribution to match a data. Remaining differences can be corrected on a second step. On another hand it is also possible to neglect second order effects on  $E_T^{miss}$  from  $\sum E_T$  distribution and directly correct difference between data and MC.

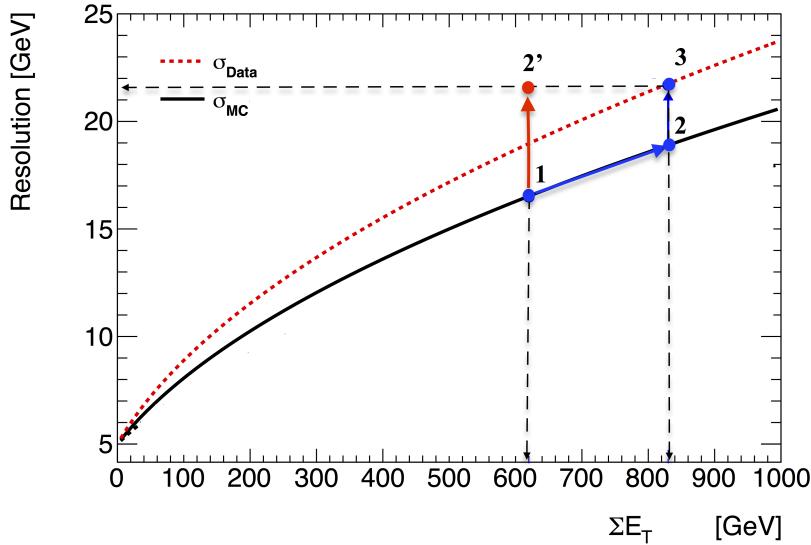


Fig. 12.2: Schematic view of the correction procedure: this figure illustrates the resolution of  $u_{\perp}$  as a function of  $\sum E_T$ . The dotted curve represents data resolution ( $\sigma_{data}$ ), solid black is a nominal MC ( $\sigma_{MC}$ ). Blue line from point 1 to point 2 corresponds to a  $\sum E_T$  correction. Red line from point 1 to point 2' corresponds to a direct correction of resolution mismodelling

### 504 12.1.1 Sumet distribution correction

Distribution of  $\sum E_T$  events are shown on a Fig. 12.3. 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. Since  $\sum E_T$  and  $P_T^W$  are correlated correction factors are derived inside  $p_T^{W,rec}$  as follows:

$$SF^{channel} = \frac{\sum E_T^{data, selection}}{\sum E_T^{MC, no cuts}}, \quad (12.3)$$

505 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. Because  
 506 of the small data statistics combination of  $W \rightarrow e\nu$  and  $W \rightarrow \mu\nu$  processes is used. Oppositely, in  
 507 MC  $\sum E_T^{MC, no cuts}$  is taken without any cuts. Scale factors are determined separately for each signal  
 508 process for a W boson decays. Example of correction factors for a different  $p_T^{W,rec}$  bins are shown  
 509 on a Fig.12.5 Ratio inside each bin can be parameterised by a polynomial degree 2 inside each  $p_T^{W,rec}$   
 510 bin. Total SF obtained by this procedure are shown on a Fig. 12.6. The distribution of  $\sum E_T$  after  
 511 correction is shown on a Fig. 12.8. Reconstructed boson pt spectrum is leaving almost untouched,  
 512 while this procedure still intoduses some shift in a truth boson pt spectrum(Fig. 12.7 ). Effect on the  
 513 resoultion of  $u_{\perp}$  is shown on a Fig. ??.

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

$$fit\ parameters_{new} = fit\ parameters + gaus^{2D}(cov.matrix), \quad (12.4)$$

514 where *fit parameters* is a vector of best fit parameters and *gaus*<sup>3D</sup> is a 3D gaus, that takes co-

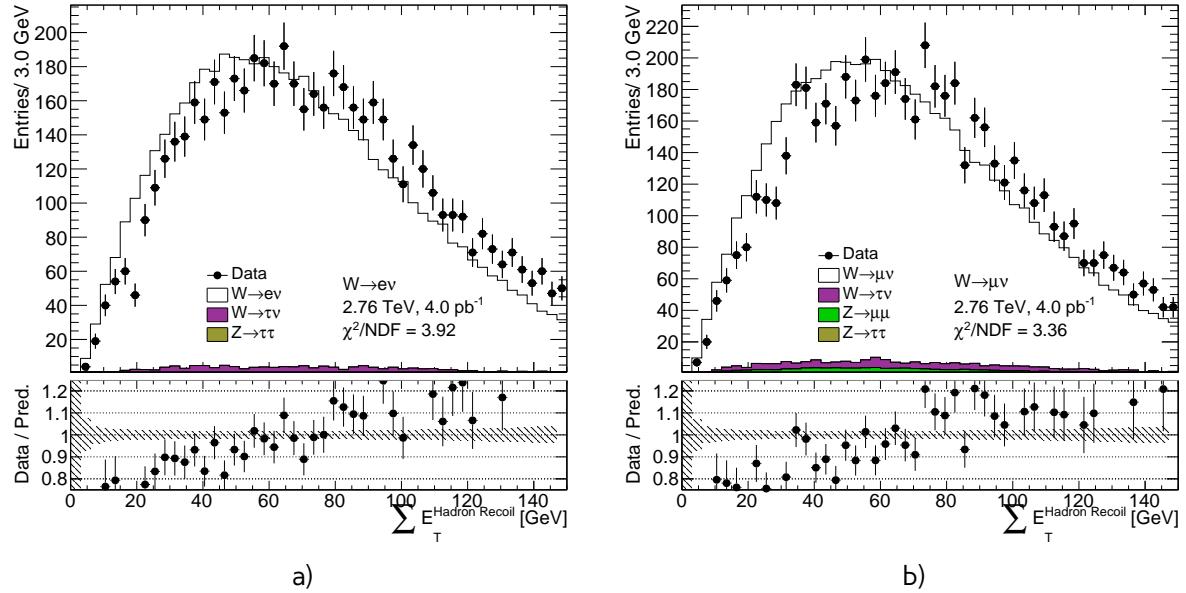


Fig. 12.3: Distribution of  $\sum E_T$  from a)  $W \rightarrow e\nu$  and b)  $W \rightarrow \mu\nu$  events before correction

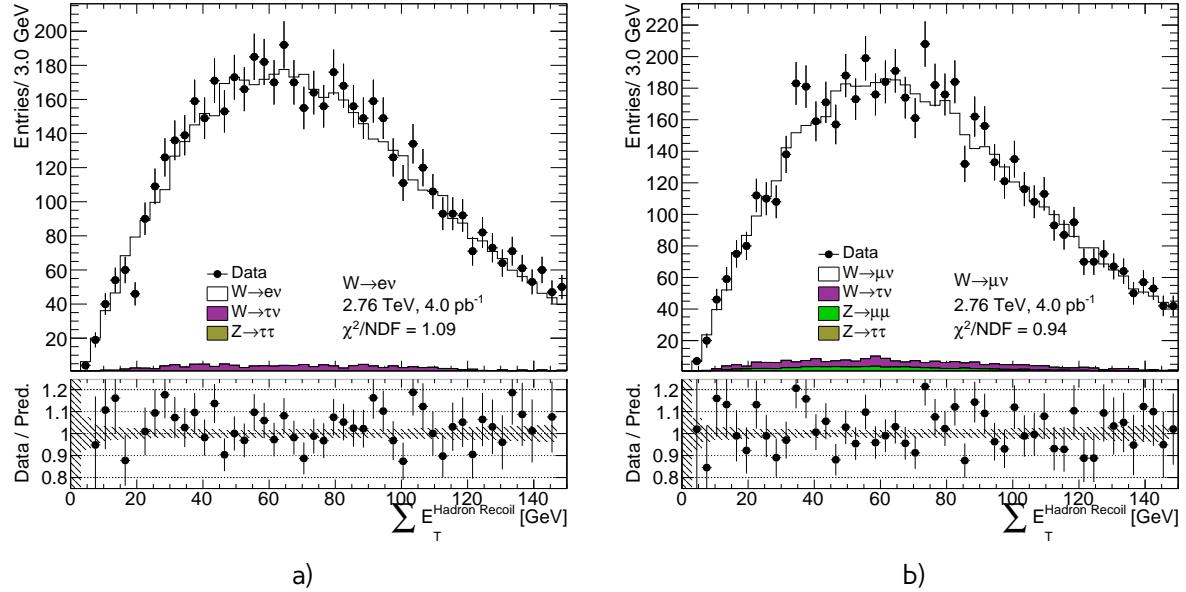


Fig. 12.4: Distribution of  $\sum E_T$  from a)  $W \rightarrow e\nu$  and b)  $W \rightarrow \mu\nu$  events after correction

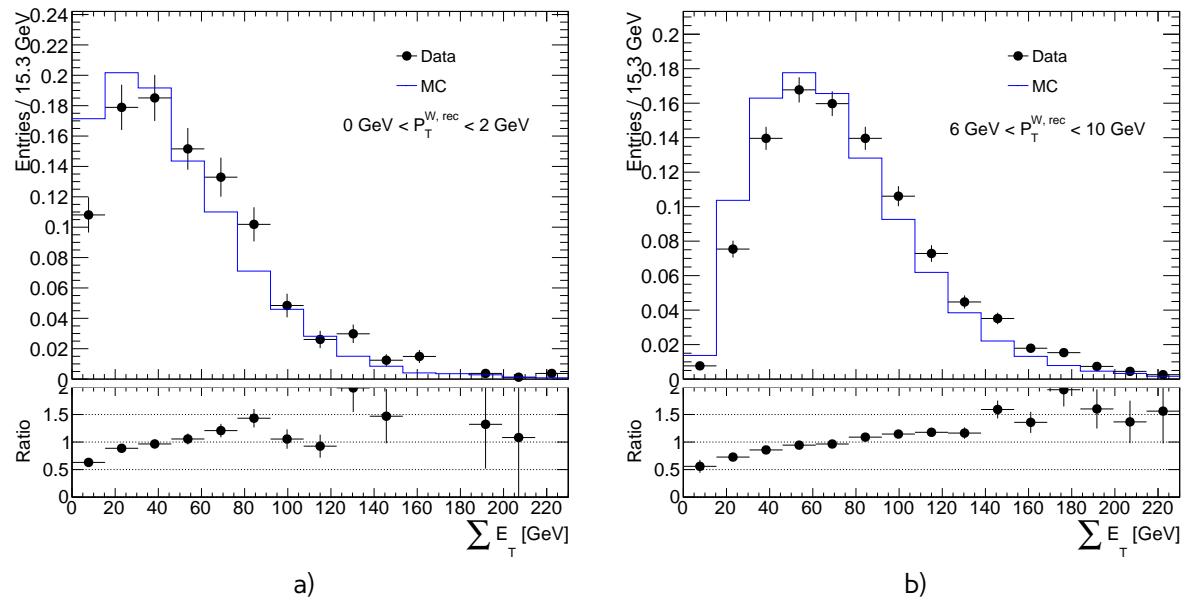


Fig. 12.5: Distribution of  $\sum E_T$  for different  $p_T^{W,rec}$  bins

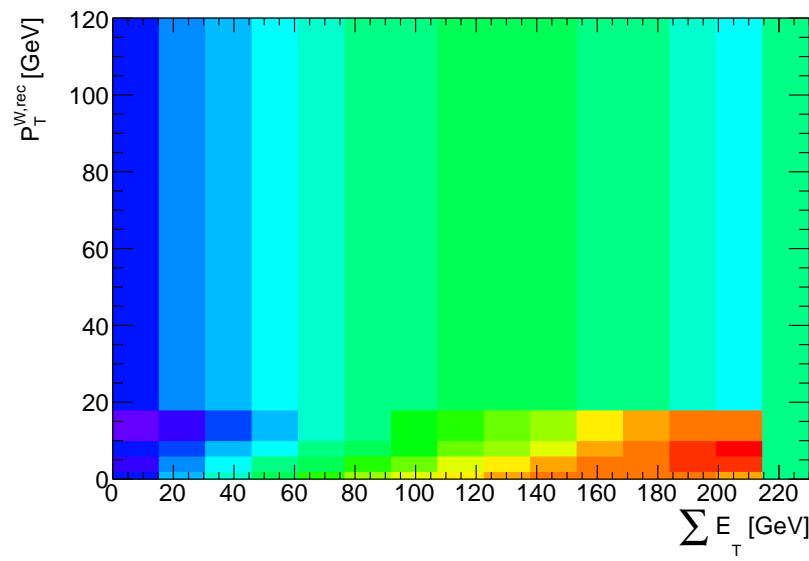


Fig. 12.6: Correction factors for  $W \rightarrow e\nu$

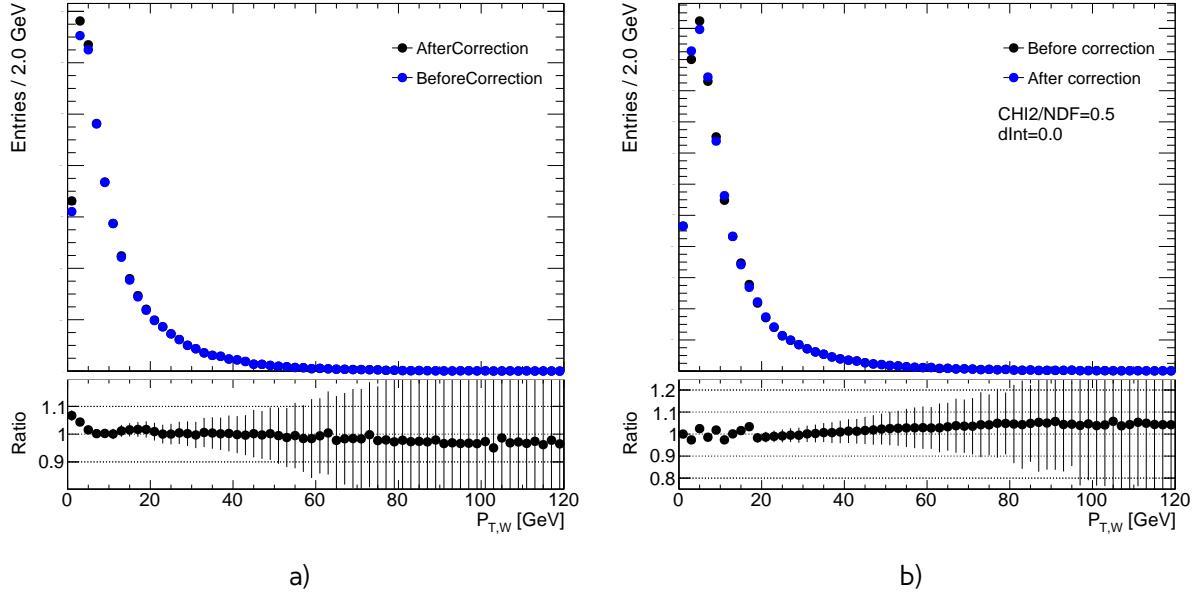


Fig. 12.7

 Table 12.1: Effect of  $\sum E_T$  correction on a  $C_W$  for a different channels and methods

| Channel                    | $\delta C_W$<br>polynomial order 2 | $\delta C_W$<br>polynomial order 1 | $\delta C_W$<br>Toy MC |
|----------------------------|------------------------------------|------------------------------------|------------------------|
| $W^+ \rightarrow e^+\nu$   | 0.39%                              | 0.31%                              | 0.03%                  |
| $W^- \rightarrow e^-\nu$   | 0.33%                              | 0.22%                              | 0.03%                  |
| $W^+ \rightarrow \mu^+\nu$ | -0.20%                             | -0.28%                             | 0.03%                  |
| $W^- \rightarrow \mu^-\nu$ | -0.21%                             | -0.27%                             | 0.03%                  |

515 variance matrix from fit results. This method is allowing to take into account correlations between  
 516 parameters. This procedure is repeated 25 times for each bin, that gives us set of 25 scale factors,  
 517 that are later used for error determination.

518 Systematical error can be studied by applying lower order of approximation on a SF or not applying it  
 519 at all. The overall effect on a  $C_W$  for a different methods is shown in a Tab. 12.1. Results are dominated  
 520 by a systematics error. However, there is a difference in a sign of the effect for a different flavors  
 521 of the analysis. This cannot be explained from a physical point of view, so it was decided not to use  
 522 this corrections in a final analysis.

### 523 12.1.2 Resolution correction using Z events

524 Another way to check resolution effects is to use  $u_\perp$  and  $u_\parallel - p_T^Z$  distributions in a Z events. This  
 525 correction assumes, that any resolution mismodelling reflects discrepancies in the  $\sum E_T$  distribution,  
 526 while difference in resolution at a given  $\sum E_T$  is a subleading. There is a clear difference in a rms  
 527 of this distributions between data and MC, that cannot be accounted as a statistical error in data.  
 528 Difference in resolutions is consistent for  $u_\perp$  and  $u_\parallel - p_T^Z$  distributions, but depends on a flavor of

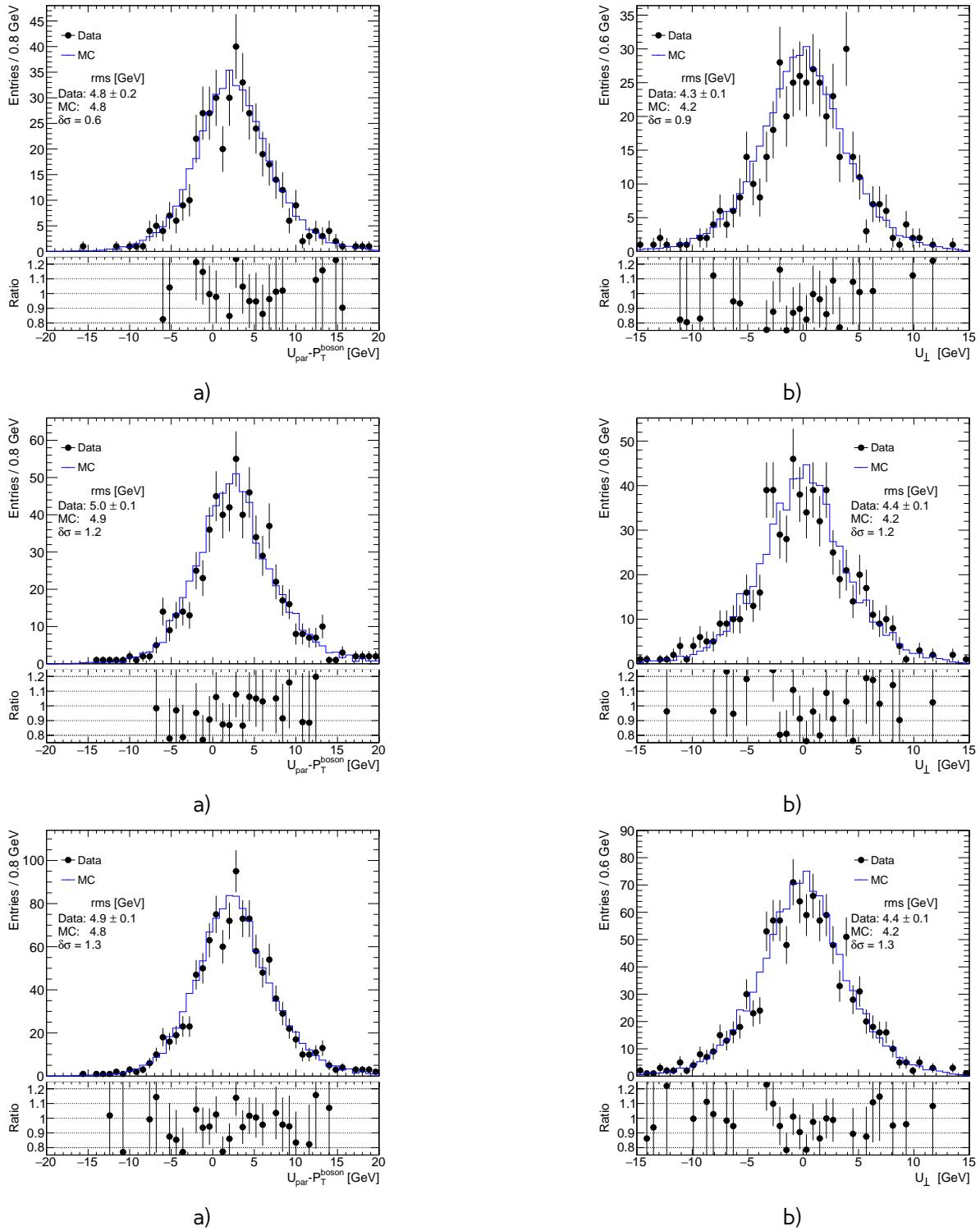


Fig. 12.8

529 the analysis. The resolution is corrected by smearing with a Gaussian distribution each component  
 530 of a hadron recoil:

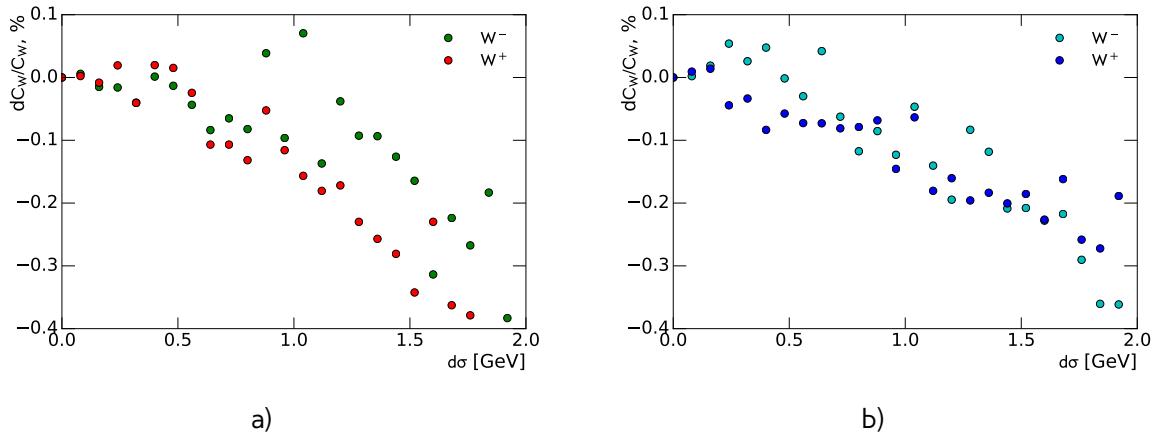


Fig. 12.9: Effect on a  $C_W$  for a different  $d\sigma$  for a)  $W \rightarrow e\nu$  b)  $W \rightarrow \mu\nu$  channel

Table 12.2

| 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 |

$$u'_\parallel = u_\parallel + Gaus(0, d\sigma) \quad (12.5)$$

$$u'_\perp = u_\perp + Gaus(0, d\sigma), \quad (12.6)$$

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

$$d\sigma = \sqrt{\sigma_{data}^2 - \sigma_{MC}^2} \quad (12.7)$$

531 Systematic error of this  $d\sigma$  is taken as an statistical error for  $\sigma_{data}$ . Overall effect on a  $C_W$  depending  
 532 on a  $d\sigma$  is shown on a Fig. 12.9. Due to a random nature of this correction, effect is not stable for  
 533 a small  $d\sigma$ . Stability of this correction can be tested by repeating this procedure several times with  
 534 different random seed number. Due to a not stable nature of this correction the overall systematics  
 535 coming from resolution mismodelling is assumed to be 0.2% for each W channel.

## 12.2 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:

$$u_{\parallel}^{MC,cor} = u_{\parallel}^{MC} \cdot SF_{HR,bias}, \quad (12.8)$$

and can be obtained by scanning the impact of the scaling factor on the Data to MC agreement of the distributions that are dominated by the recoil scale uncertainties. Since W boson has no second

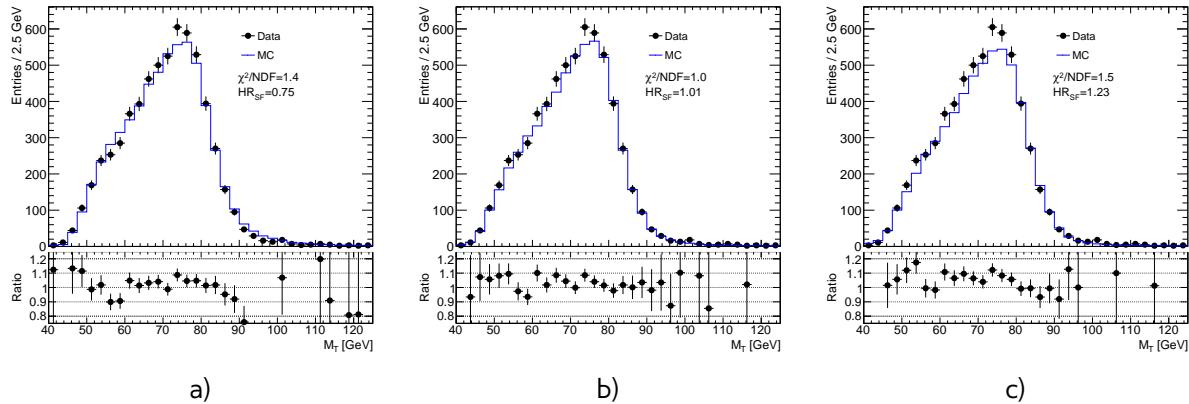


Fig. 12.10

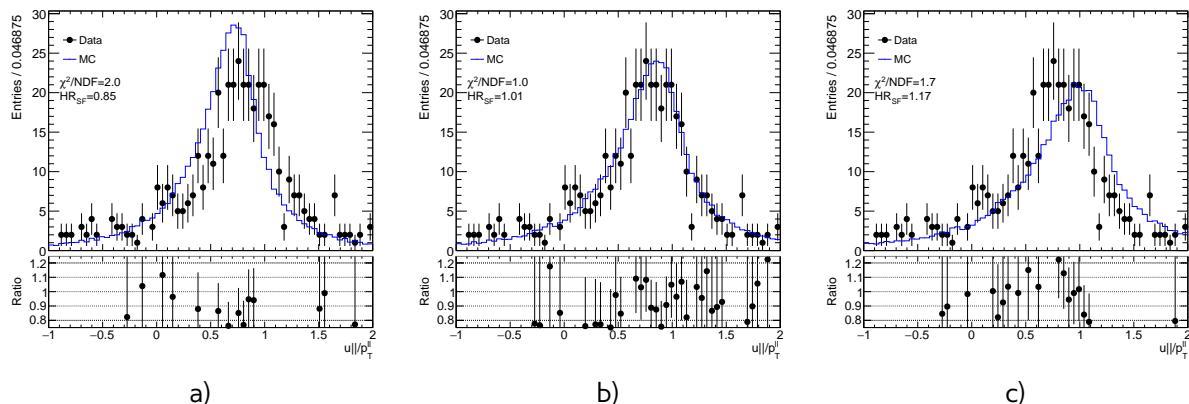


Fig. 12.11

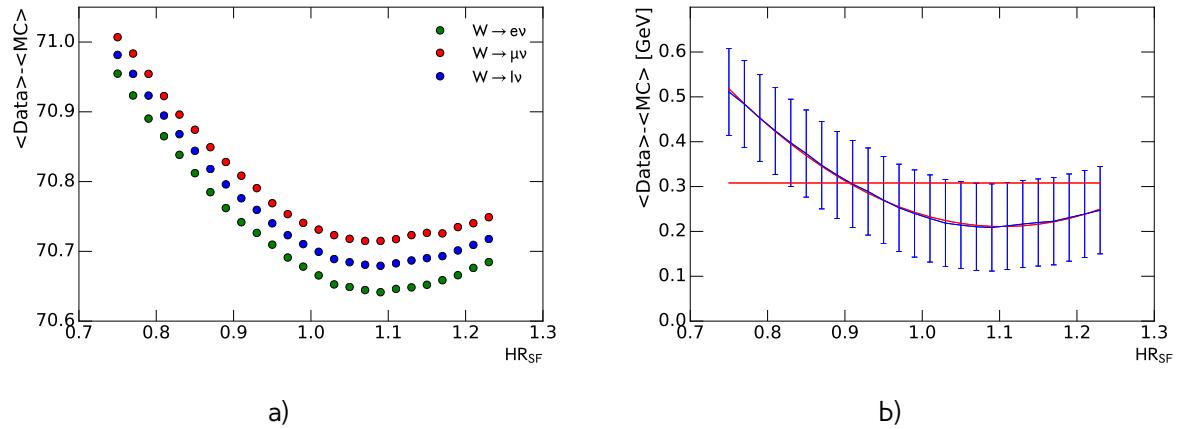


Fig. 12.12

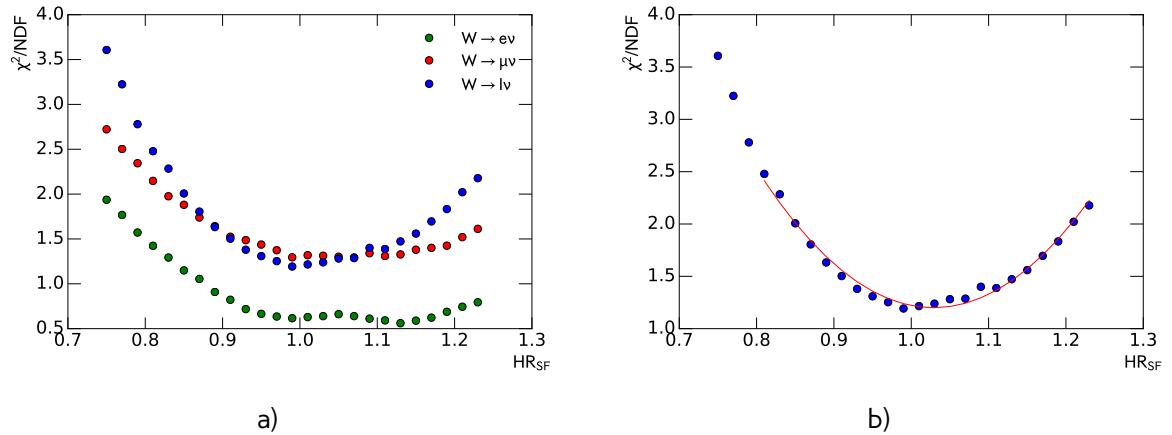


Fig. 12.13

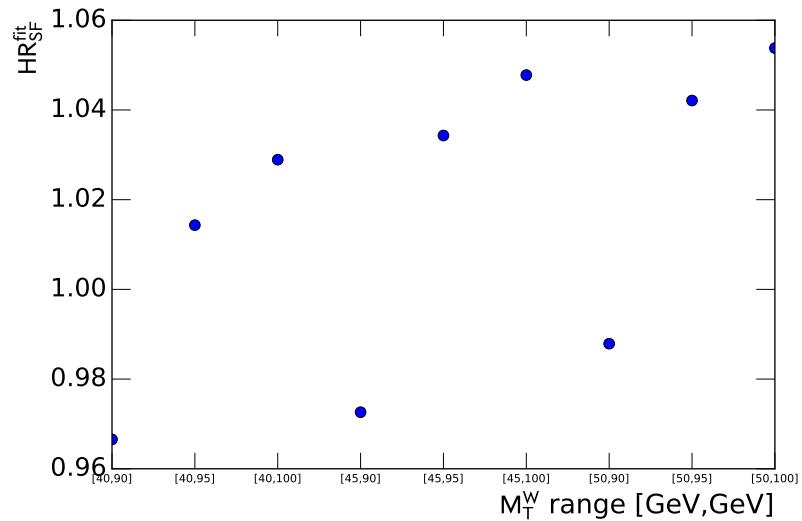


Fig. 12.14

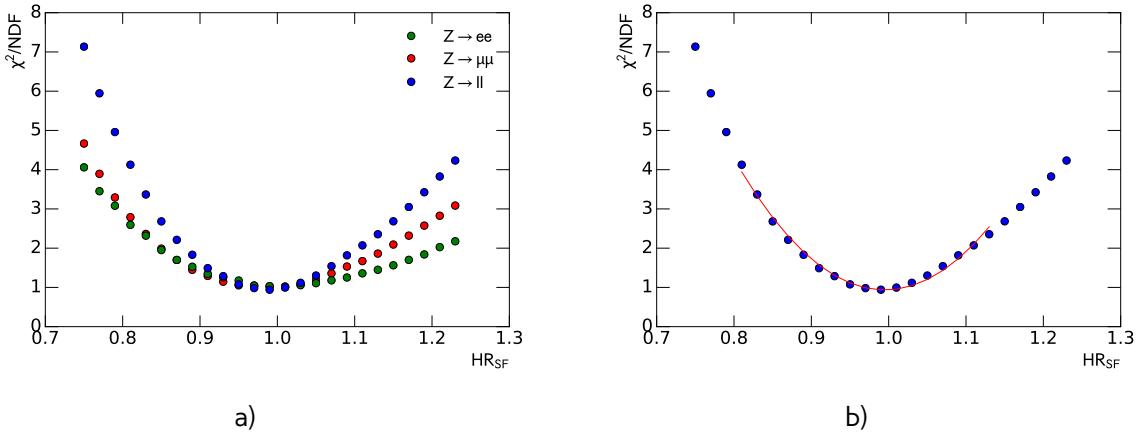


Fig. 12.15: Effect on a  $C_W$  for a different  $d\sigma$  for a)  $W \rightarrow e\nu$  b)  $W \rightarrow \mu\nu$  channel

539 source of  $P_T^W$  measurements, determination of the hadron recoil bias should use the distributions, that  
 540 are not sensitive to a truth  $P_T^W$  spectrum. One of the optimal choices is a  $M_T^W$  distribution. Transverse  
 541 mass distribution for a different scale choices is shown on a Fig. 12.10. Multijet background is not  
 542 included, because its shape and number of events is depending on a hadron recoil scale and thus  
 543 can introduce additional systematics.

544 The first way to determine correction factor is using a difference in the mean of transverse mass  
 545 in data and MC. Statistical error of this determination is an error of the mean in the data. The  
 546 precision of this method is low, as 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:

$$\chi^2 = \frac{(x - sf_{best})^2}{\sigma_{sf}^2} + \chi^2_0, \quad (12.9)$$

547 where  $sf_{best}$  is the best scale factor and  $\sigma_{sf}$  is a statistical error of this parameter. Distribution of  $\chi^2$   
 548 and a fit in combined  $W$  channel is shown on a Fig. 12.13.

549 Because of the possible mismodelling of the tail  $M_T^W$  distribution it is not included in a  $\chi^2$  cal-  
 550 culation, leaving a free choice of the parameter of the cutoff. It is also possible to exclude regions  
 551 with high multijet background contamination by applying a tighter cut on a  $M_T^W$ . This fit range is  
 552 introducing one source of systematic error. Effect of the range on value determination is shown on  
 553 a Fig. 12.14.

554 Similarly to a  $W$  channel, scale correction in a  $Z$  sample can be determined from distribution  $\frac{u_{ll}}{p_T^{ll}}$ ,  
 555 shown on a Fig. 12.15. Since there is no choice of the range and dependency on  $P_T^{bos}$  modeling, there  
 556 is just one source of uncertainty.

557 Results on a hadron scale factors and its errors are shown in a Table 12.2. The results are consistent  
 558 within 1 sigma.

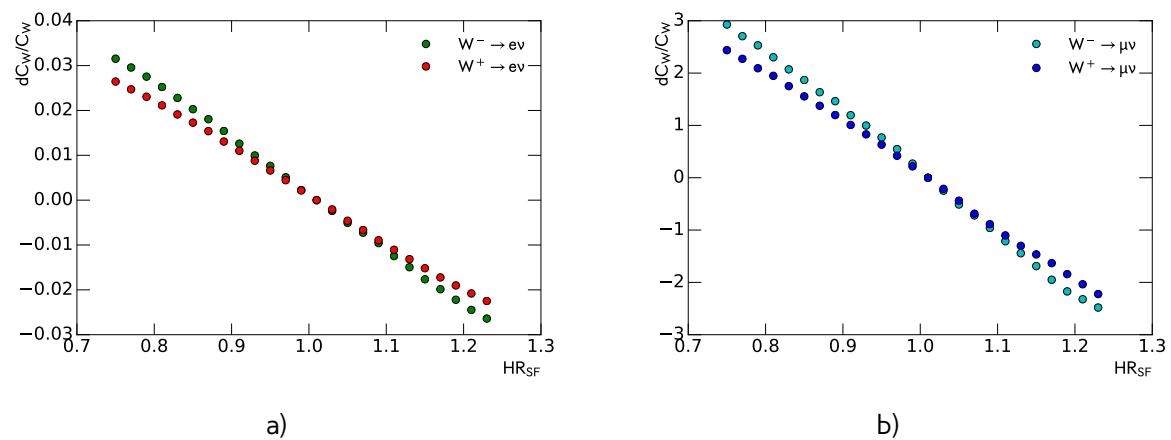


Fig. 12.16: Effect on a  $C_W$  for a different  $d\sigma$  for a)  $W \rightarrow e\nu$  b) $W \rightarrow \mu\nu$  channel

559

# Chapter 13

## Background estimation

561 After the event selection described in chapter 10 the background contribution is around 4% for  
 562 W-analysis and 0.2% for Z analysis (which is with this statistics is negligible). Main backgrounds for  
 563 W analysis are coming from:

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

568 Most of the backgrounds are estimated using MC. They are normalized using highest cross-section  
 569 order available. The total list of simulated backgrounds and its cross-section is shown in a Table 13.1.  
 570 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  |
| $Z \rightarrow \tau\tau$ | 303                    | LO    |
| $t\bar{t}$               | 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    |

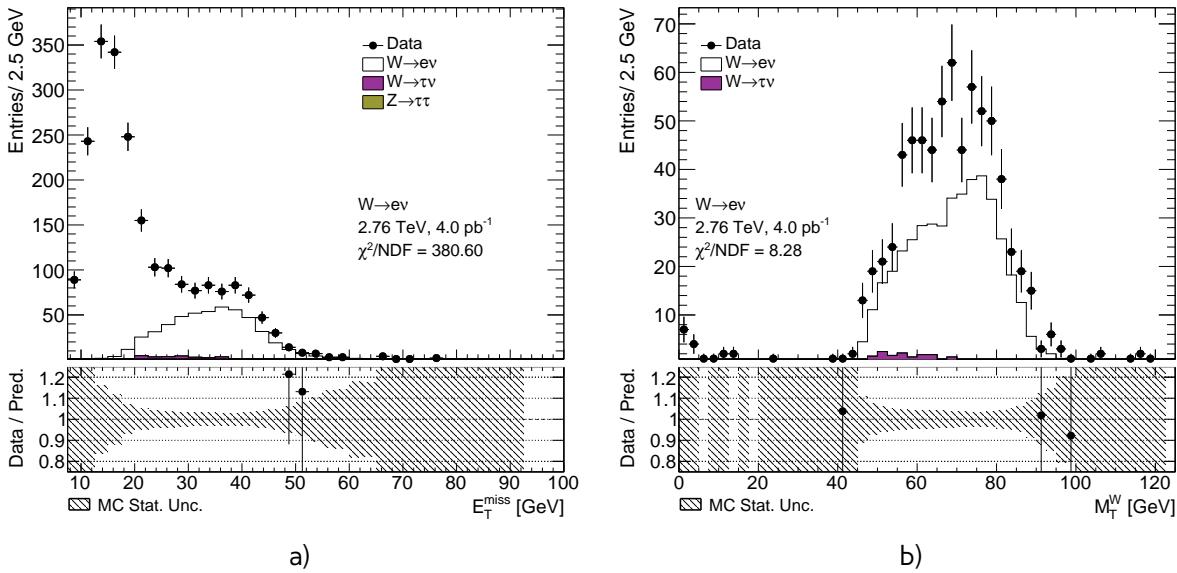


Fig. 13.1: Distribution for a)  $E_T^{\text{miss}}$  b)  $M_T^W$  template selection for  $W \rightarrow e\nu$  events

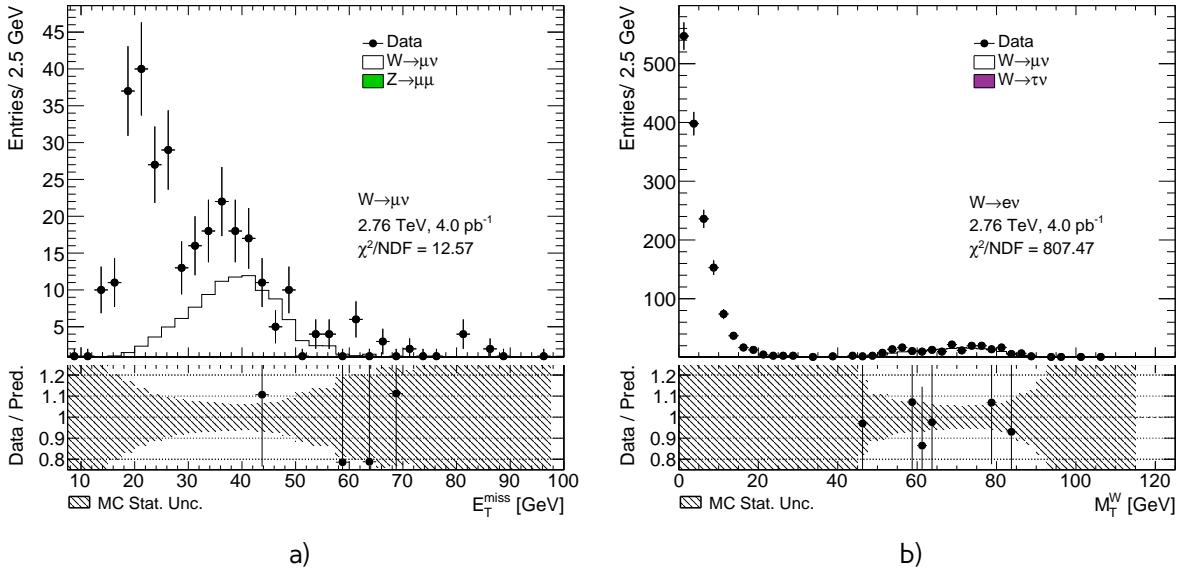


Fig. 13.2: Distribution for a)  $E_T^{\text{miss}}$  b)  $M_T^W$  template selection for  $W \rightarrow \mu\nu$  events

### 13.1 QCD background estimation

There is a small probability, that jet can fake W-boson decay with isolated lepton and  $E_T^{\text{miss}}$  through the energy mismeasurement in the event. Event selection is suppressing this type of the background, but not fully eliminating it. Due to a large jet production cross-section and complex composition, generation of MC events becomes impractical. This is why data driven technique for QCD background estimation have been used. In our case contribution of the QCD background in a Z sample is negligible, so it is estimated just for a  $W \rightarrow e\nu$  and  $W \rightarrow \mu\nu$  processes.

Data driven method allows to have model independent predictions with small statistical uncertainty. This method is using QCD enriched region, where signal events are suppressed. This is usually done

580 by reversing identification or isolation criteria. It is assumed, that shape of the qcd background is  
 581 staying the same in the signal region. Normalization can be derived in a control region through the  
 582 template fit.

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

### 585 13.1.1 Template selection

586 A study have been performed to determine appropriate template selection. Because of the origins  
 587 of the QCD backgrounds, missing transverse energy  $E_T^{miss}$  should be smaller in a QCD, than in a signal  
 588 region. Releasing  $E_T^{miss}$  cut is allowing to gain a bigger statistics for a QCD template. It is also possible  
 589 to release  $M_T^W$  cut. Most of the multijet background event should contribute than in small  $M_T^W$   
 590 region. The template sample can have a contributions from other backgrounds (mostly coming from  
 591  $W \rightarrow l\nu$ ). Best template selection is allowing to have big data statistics and small EWK background  
 592 contribution at the same time. In order to supress signal additionally reversed ID or Isolation criteria  
 593 is applied.

594 For electron flavor, template is requiring for electron candidate to fail Medium isolation criteria,  
 595 but to pass loose selection. Control distribution for a different template selection in electron channel  
 596 are shown on a Fig. 13.1. Released  $E_T^{miss}$  cut is allowing to have a better template statistics.

597 In a muon channel template selection build by inverting isolation criteria ( $PtCone20 > 0.1$ ). In case  
 598 of  $W \rightarrow \mu\nu$  the qcd background template can be achieved by releasing  $M_T^W$  cut (Fig. 13.2).

In order to avoid double counting, EWK backgrounds are substracted from a template. The total  
 number of events in the template can be defined as:

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

599 where  $N_{data}^{bkg\ enriched}$  and  $N_{MC_j}^{bkg\ enriched}$  are number of the events in a background enriched sample in data  
 600 and MC respectivelly. The resulting template statistic is 1348 and 1509 events for  $W \rightarrow e\nu$  and  
 601  $W \rightarrow \mu\nu$  respectively.

### 602 13.1.2 Methodology of the template sample 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:

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

603 where  $x$  is a fit variable ( $E_T^{miss}$  and  $M_T^W$  for  $W \rightarrow e\nu$  and  $W \rightarrow \mu\nu$  respectivelly),  $F_i(x)$  and  $F_{qcd}(x)$   
 604 are the probability density functions of MC samples and QCD background template respectivelly. Fit  
 605 parameters  $f_i$  are the fractions of MC within fit region. In order to eliminate systematics, coming  
 606 from cross-section uncertainty, with signal fractions are left freely and and background MC fractions  
 607 are varied within 5% uncertanty.

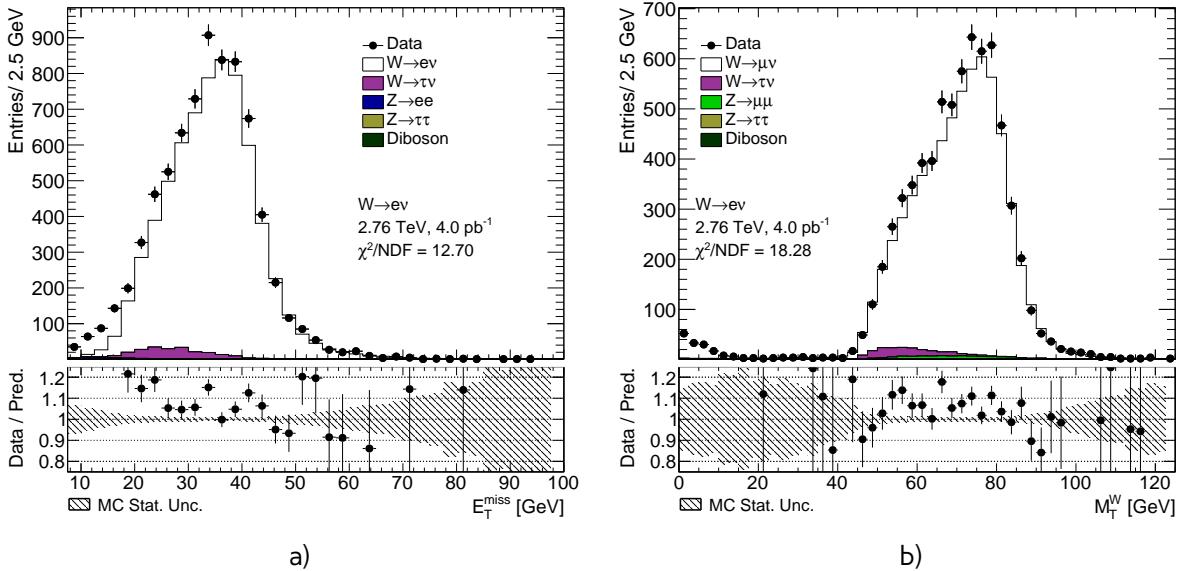


Fig. 13.3: Distributions used for multijet background estimation for a)  $W \rightarrow e\nu$  b)  $W \rightarrow \mu\nu$

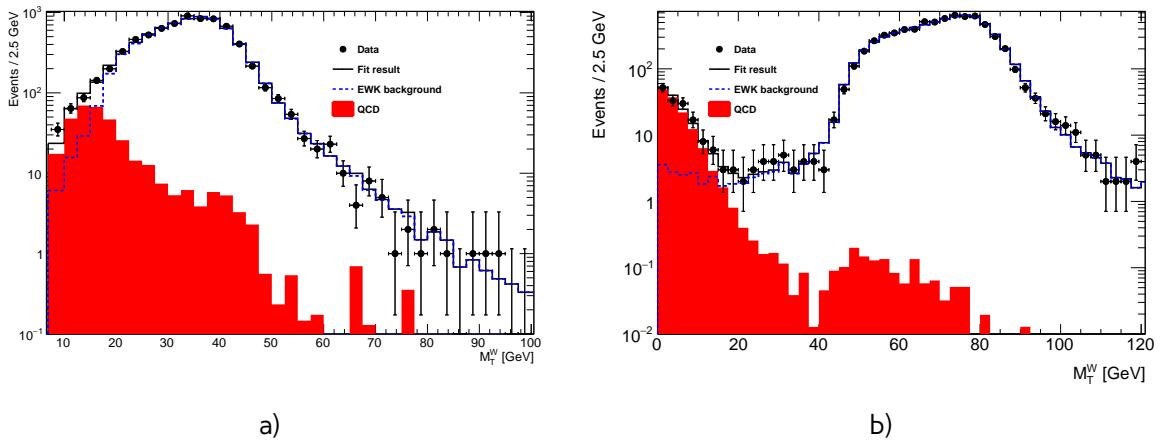


Fig. 13.4: The multijet background estimation for a)  $W \rightarrow e\nu$  using reversed ID cut and released  $E_T^{\text{miss}}$  cut b)  $W \rightarrow \mu\nu$  using released  $M_T^W$  cut and  $b\bar{b} + c\bar{c}$  template

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

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

where  $\sum f_i$  is a sum of all fractions in the fit,  $N_{\text{Data}}^{\text{fit}}$  is a number of data events in a fit histogram and  $N_{\text{template}}$  is a number of event in a template. The fit is performed separately for  $W^+$  and  $W^-$ . Additionally, fit in uncharged  $W$  channel is used as a cross-check of the fit. The results of the fitting procedure are shown on a Fig. 13.4 .

Table 13.2: Results of QCD background estimation for  $W \rightarrow e\nu$  and corresponding error

| Charge | $N_{QCD}$ | $\delta N_{fit\,unc}$ | $\delta N_{MC}$ | $\delta N_{fit\,bias}$ |
|--------|-----------|-----------------------|-----------------|------------------------|
| $W^+$  | 38.3      | 7.0                   | 7.0             | 5.0                    |
| $W^-$  | 21.5      | 0.7                   | -9.4            | 4.0                    |
| $W$    | 66.1      | 21.2                  | 4.2             | 10.                    |
| Total  | 31.0      | 6.1                   | 8.6             | 4.7                    |

### 612 13.1.3 Systematic Uncertainty from the Multi-jet Background Estimation

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

$$\delta_{QCD} = \sqrt{\delta_{fit\,unc}^2 + \delta_{MC}^2 + \delta_{fit\,bias}^2 + \delta_{template}^2}, \quad (13.4)$$

613 where  $\delta_{fit\,unc}$  is the uncertainty for a scale from a  $\chi^2$  fit.

The second component  $\delta_{MC}$  is coming from a possible mismodelling of MC in a fitted region. It can be estimated by comparison of separate fit results for  $W$ ,  $W^+$  and  $W^-$ . Number of multijet background events should not depend on a charge of the analysis, so it is expected to have:

$$N_{QCD}^W = 0.5 \cdot N_{QCD}^{W^+} = N_{QCD}^{W^-} \quad (13.5)$$

614 Standard deviation of this 3 numbers is taken as systematic uncertainty. Since in  $W \rightarrow \mu\nu$  channel  
615 QCD template is fitted in a region without any additional EWK background this component is 0.

616 Third uncertainty is due to a potential bias in the template as a result of the template choise  
617 and a template statistics itself. For estimation of this uncertainty different template selections have  
618 been used. For  $W \rightarrow e\nu$  channel different reversed isolation criteria have been tried (Fig. 13.5 a)).  
619 The overall discrepancies can be considered negligible. For  $W \rightarrow \mu\nu$  channel template variations  
620 are estimated using fits with  $b\bar{b} + c\bar{c}$  MC samples. Fig. 13.5 b) compares data template with template  
621 obtained using signal selection with released  $M_T^W$  cut and template selection. Results for a different  
622 template fits are presented in a Tab 13.3

623 The third component  $\delta_{fit\,bias}$  is coming from an effect from arbitrary choise of bin size . This  
624 error is estimated by repeating fit for a different binnings. This component is assumed negligible in  
625  $W \rightarrow \mu\nu$  case, because of the small number of events.

626 Results of QCD background uncertainty estimation for  $W \rightarrow e\nu$  and  $W \rightarrow \mu\nu$  are shown in  
627 a Tab. 13.2 and 13.3 respectively. The overall number of QCD background events is estimated as  
628  $31.00 \pm 13.0$  for  $W^+ \rightarrow e^+\nu$  and  $W^- \rightarrow e^-\nu$  and  $1.50 \pm 0.9$  for  $W^+ \rightarrow \mu^+\nu$  and  $W^- \rightarrow \mu^-\nu$ . The  
629 overall fraction of QCD events is lower, than in 7 TeV data <reference to a 7 TeV paper>, what is  
630 agreeing with expectations.

Table 13.3: Results of QCD background estimation for  $W \rightarrow \mu\nu$  using different templates and its fit error

| Charge            | $N_{QCD}$<br>data template | $N_{QCD}$<br>$b\bar{b} + c\bar{c}$ template selection | $N_{QCD}$<br>$b\bar{b} + c\bar{c}$ signal selection |
|-------------------|----------------------------|---|---|
| $W^+$             | 2.48                       | 0.73  | 1.34  |
| $W^-$             | 2.48                       | 0.73  | 1.35  |
| $W$               | 4.97                       | 1.47  | 2.70  |
| Total per channel | 2.48                       | 0.73  | 1.35  |
| Fit error         | 0.60                       | 0.73  | 0.19  |

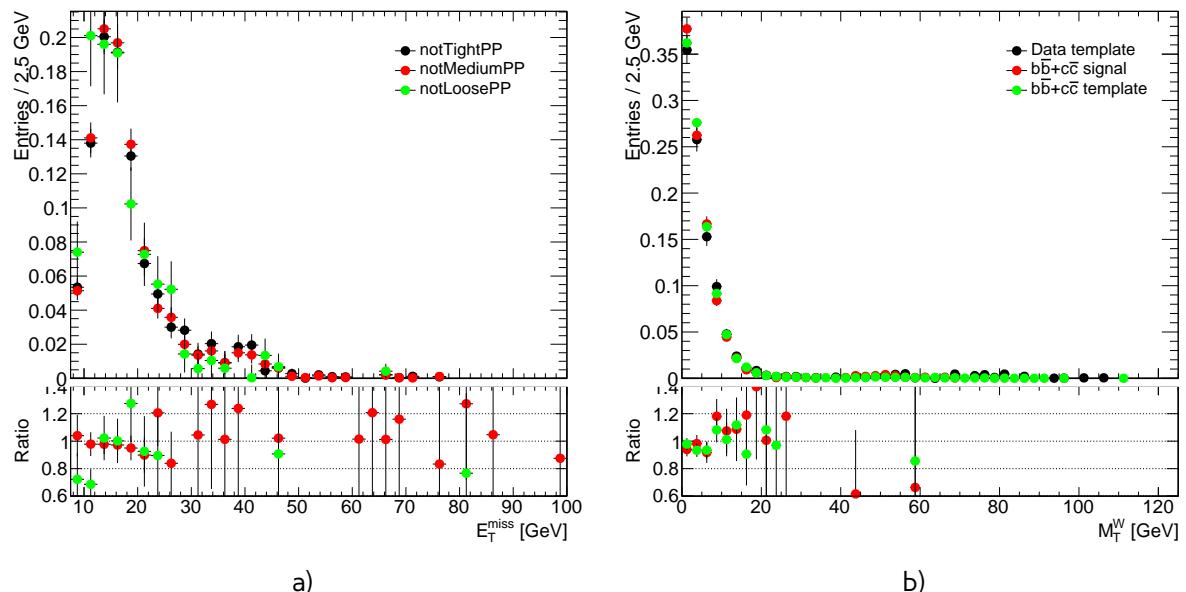


Fig. 13.5: Data and MC comparison for  $E_T^{\text{miss}}$  calculated by standard ATLAS algorithm for a)  $W \rightarrow e\nu$   
b)  $W \rightarrow \mu\nu$  events

Table 13.4: Numbers of observed candidate events for the  $W \rightarrow l\nu$  channel, electroweak (EW) plus top, and data- derived QCD background events, and background-subtracted signal events

| I       | Observed candidates | Background (EWK + top) | Background (Multijet) | Background-subtracted data $N_W^{sig}$ |
|---------|---------------------|------------------------|-----------------------|--|
| $e^+$   | 3914                | $108.1 \pm 4.6$        | $31.00 \pm 13.0$      | $3774.9 \pm 62.6 \pm 4.6 \pm 13.0$     |
| $e^-$   | 2209                | $74.2 \pm 2.4$         | $31.00 \pm 13.0$      | $2103.8 \pm 47.0 \pm 2.4 \pm 13.0$     |
| $\mu^+$ | 4365                | $150.5 \pm 4.8$        | $1.50 \pm 0.9$        | $4213.0 \pm 66.1 \pm 4.8 \pm 0.9$      |
| $\mu^-$ | 2460                | $105.6 \pm 2.6$        | $1.50 \pm 0.9$        | $2352.9 \pm 49.6 \pm 2.6 \pm 0.9$      |

Table 13.5: Numbers of observed candidate events for the  $Z \rightarrow ll$  channel, electroweak (EW) plus top and background-subtracted signal events

| I     | Observed candidates | Background (EWK + top) | Background-subtracted data $N_Z^{sig}$ |
|-------|---------------------|------------------------|--|
| $e$   | 430                 | $1.2 \pm 0.0$          | $428.8 \pm 20.7 \pm 0.0$               |
| $\mu$ | 646                 | $1.5 \pm 0.0$          | $644.5 \pm 25.4 \pm 0.0$               |

## 631 13.2 Background-subtracted $W$ and $Z$ candidate events

632 Tables 13.4 and 13.5 summarize the number of background events for  $W$  and  $Z$  selections respectiv-  
 633 ely. Uncertainties on a number of EWK+top events are coming from a statistics, cross-section uncer-  
 634 tainty (if given) and 3% of luminosity determination uncertainty. For multijet background uncertainty  
 635 is coming from a method and described in a subsection 13.1.3. For the background-subtracted events  
 636 the statistical uncertainty is quoted first, followed by the total systematic uncertainty, derived from  
 637 the EW+top and multijet background ones, considering the sources as uncorrelated.



# Chapter 14

# Control distributions

640 Kinematic distribution after all cuts (Section ??) and corrections applied on MC (Section ??), are  
 641 presented in this chapter. Distributions for  $W \rightarrow l\nu$  are split in charge and shown on a Figs. 14.1-  
 642 14.12. Distributions for  $Z \rightarrow l^+l^-$  analysis are shown on a Fig. 14.13-14.17. This plots are also showing  
 643 the systematic and statistical uncertainty as a shaded band. The uncertainties are including all of  
 644 the sources, described in a ??, except for uncertainty coming from shape variation due to a PDF  
 645 reweighing and QCD background . All of uncorrelated sources are summed in a quadrature.

The expected background contributions are estimated using MC simulations, apart from QCD background.

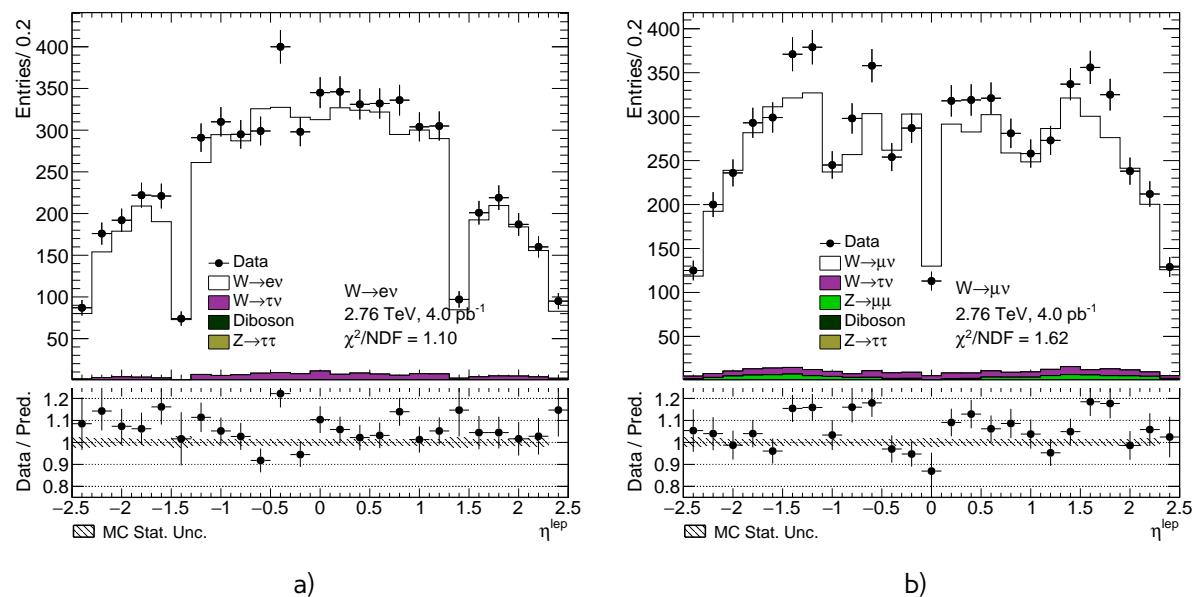


Fig. 14.1: Lepton pseudorapidity distribution from the a)  $W \rightarrow e\nu$  selection and b) the  $W \rightarrow \mu\nu$  selection.

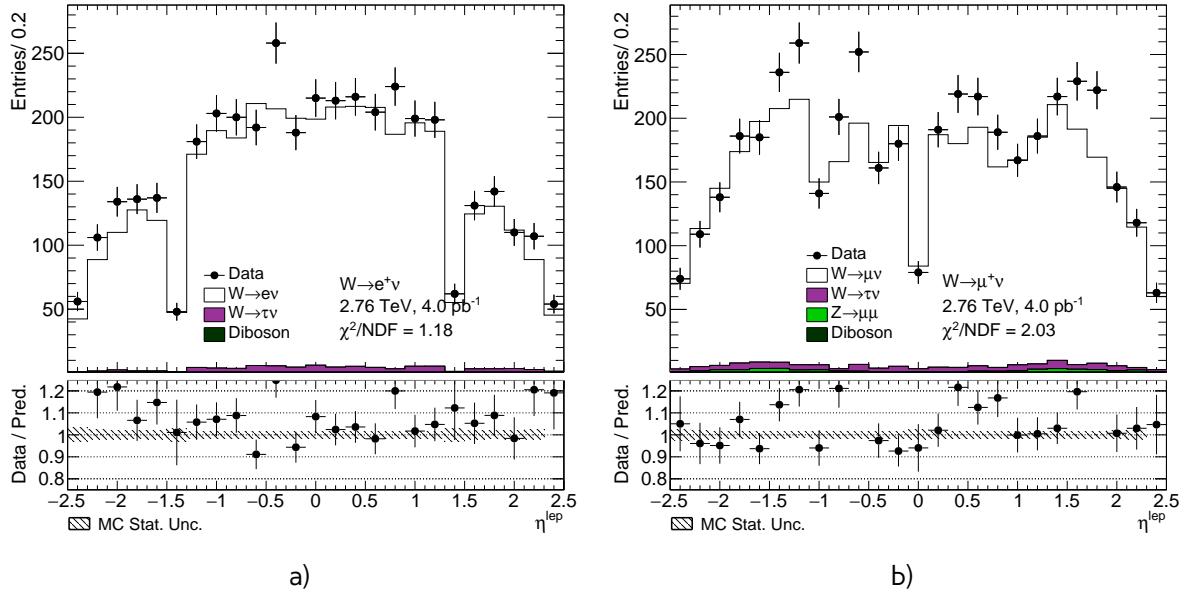


Fig. 14.2: Lepton pseudorapidity distribution from the a)  $W^+ \rightarrow e^+\nu$  selection and b) the  $W^+ \rightarrow \mu^+\nu$  selection.

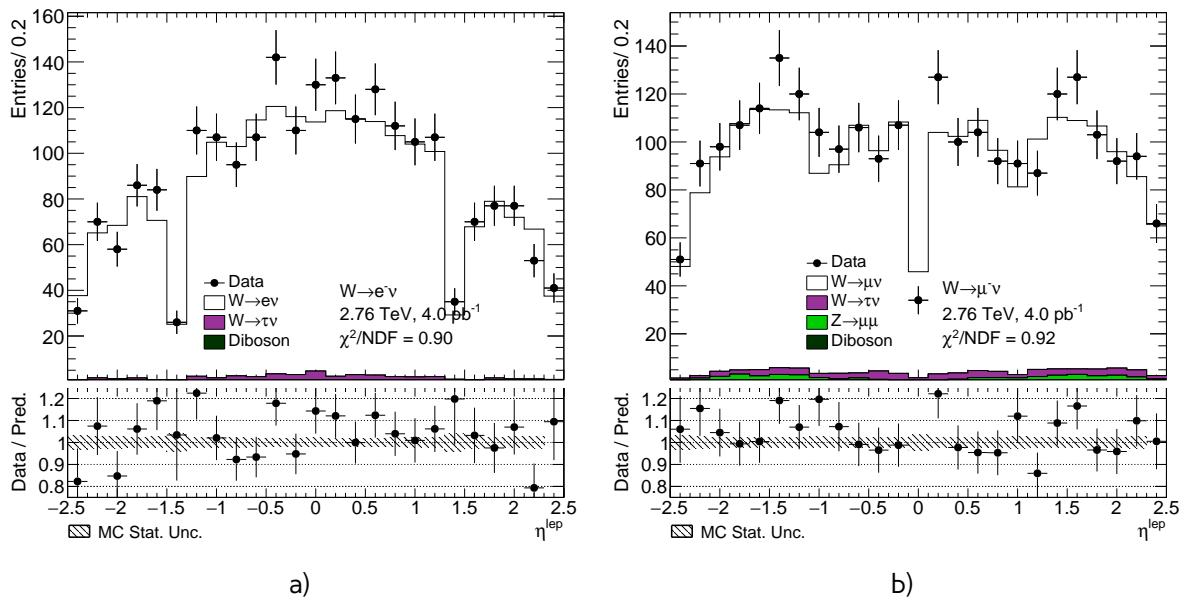


Fig. 14.3: Lepton pseudorapidity distribution from the a)  $W^- \rightarrow e^-\nu$  selection and b) the  $W^- \rightarrow \mu^-\nu$  selection.

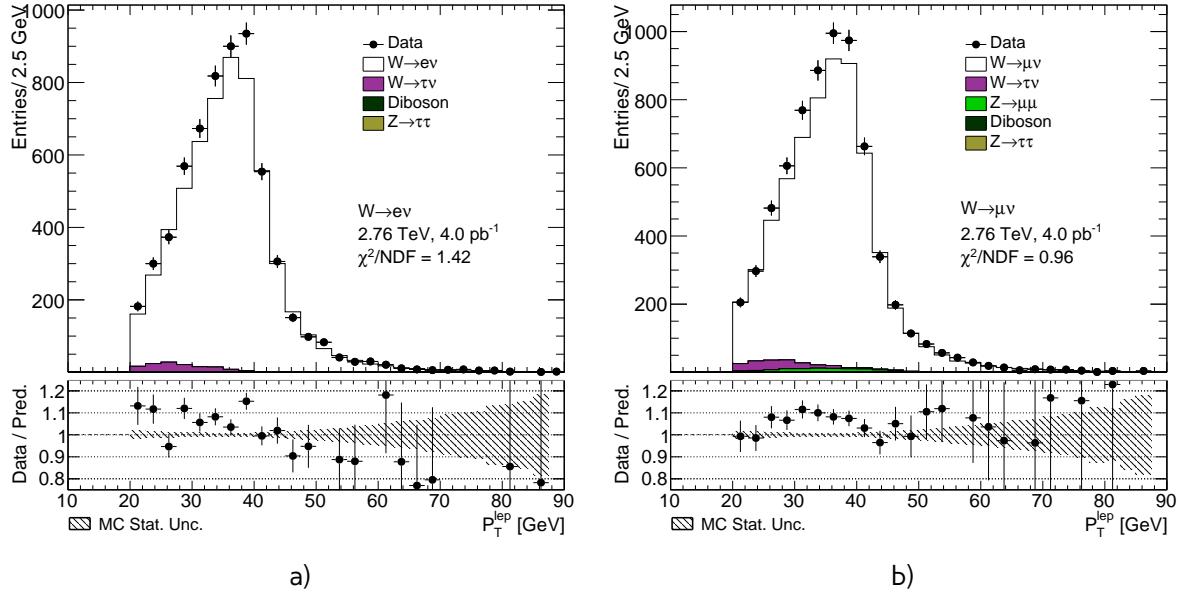


Fig. 14.4: Lepton transverse momentum distribution from the a)  $W \rightarrow e\nu$  selection and b) the  $W \rightarrow \mu\nu$  selection.

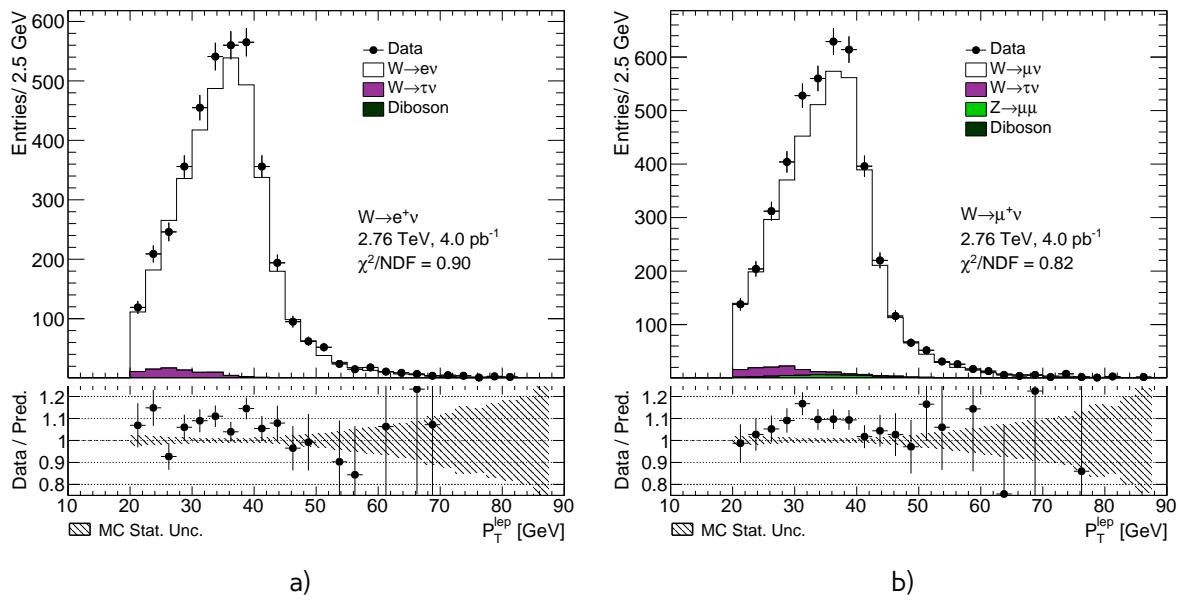


Fig. 14.5: Lepton transverse momentum distribution from the a)  $W^+ \rightarrow e^+\nu$  selection and b) the  $W^+ \rightarrow \mu^+\nu$  selection.

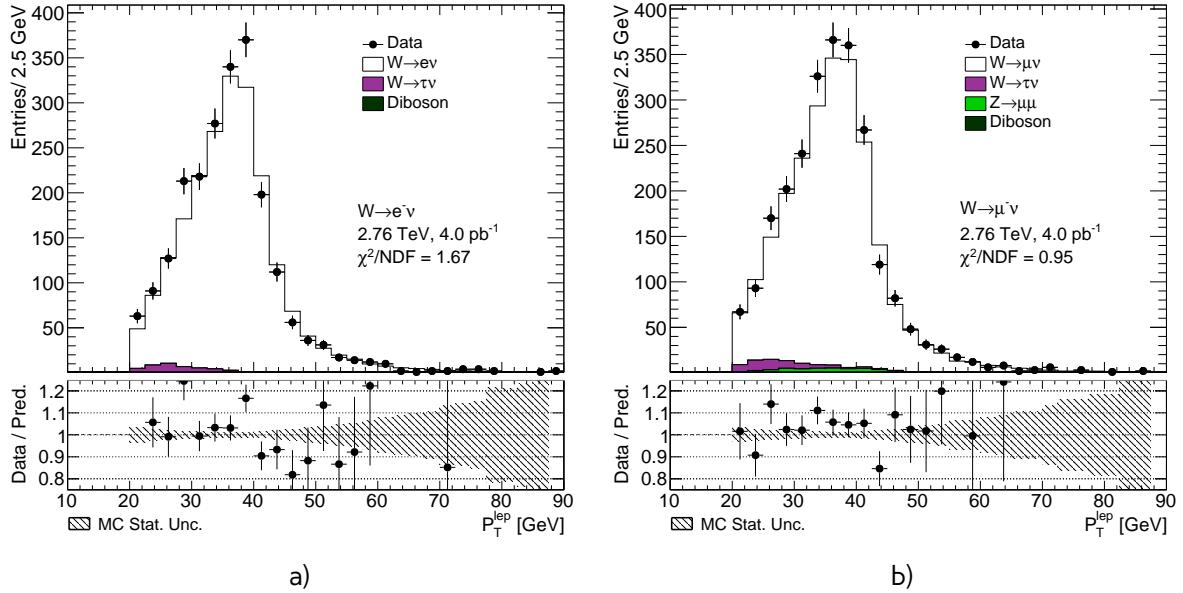


Fig. 14.6: Lepton transverse momentum distribution from the a)  $W^- \rightarrow e^- \nu$  selection and b) the  $W^- \rightarrow \mu^- \nu$  selection.

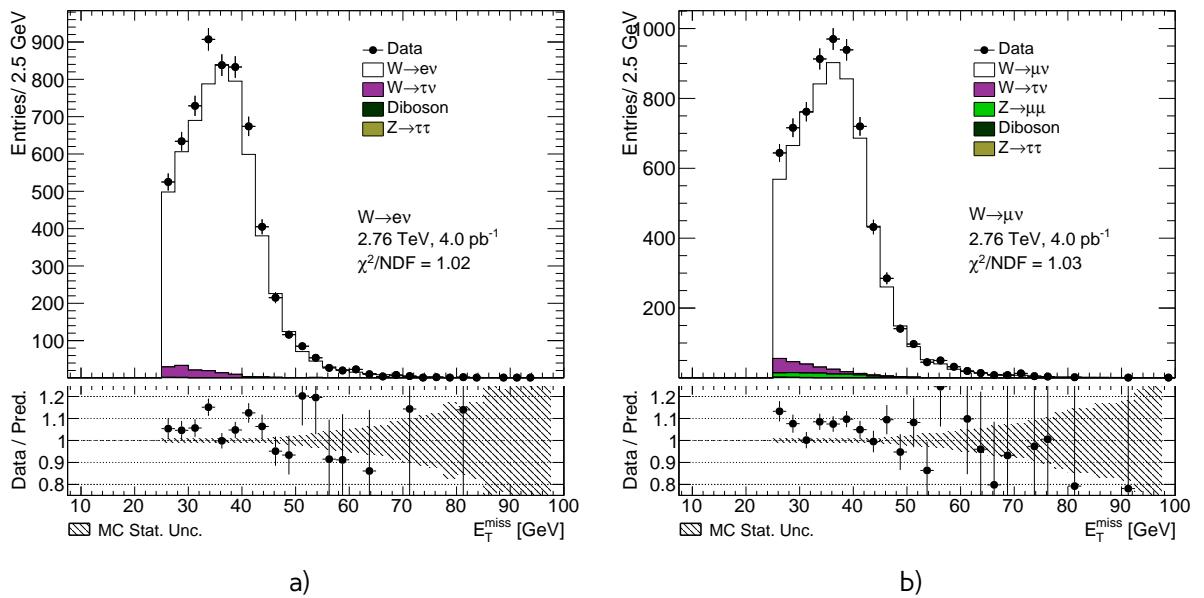


Fig. 14.7: Missing transverse energy distribution from the a)  $W \rightarrow e\nu$  selection and b) the  $W \rightarrow \mu\nu$  selection.

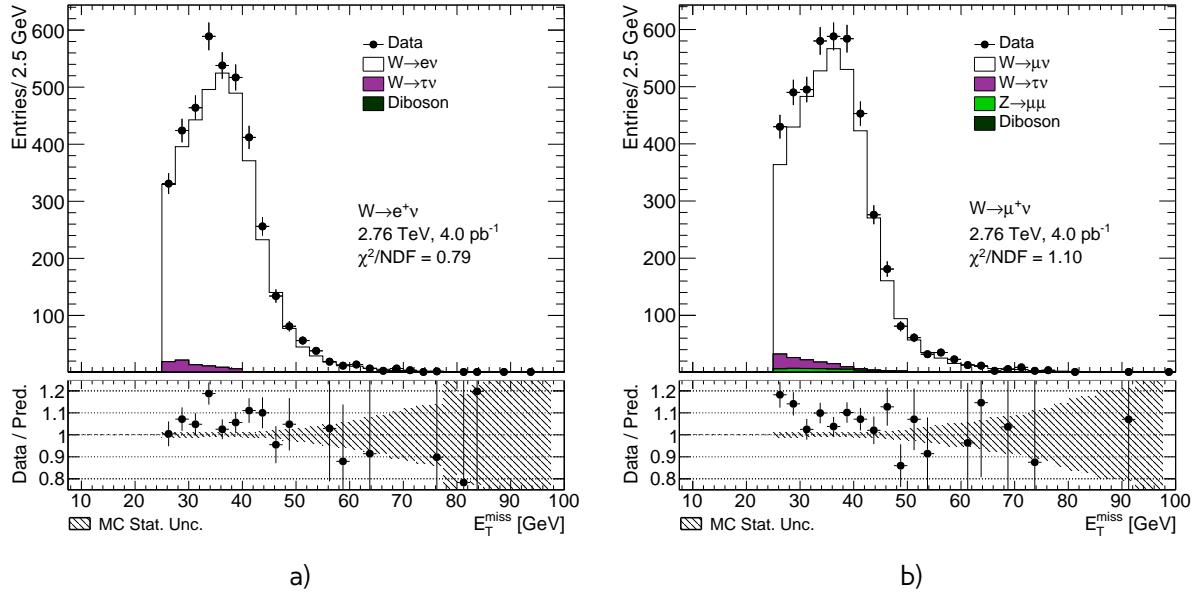


Fig. 14.8: Missing transverse energy distribution from the a)  $W^+ \rightarrow e^+\nu$  selection and b) the  $W^+ \rightarrow \mu^+\nu$  selection.

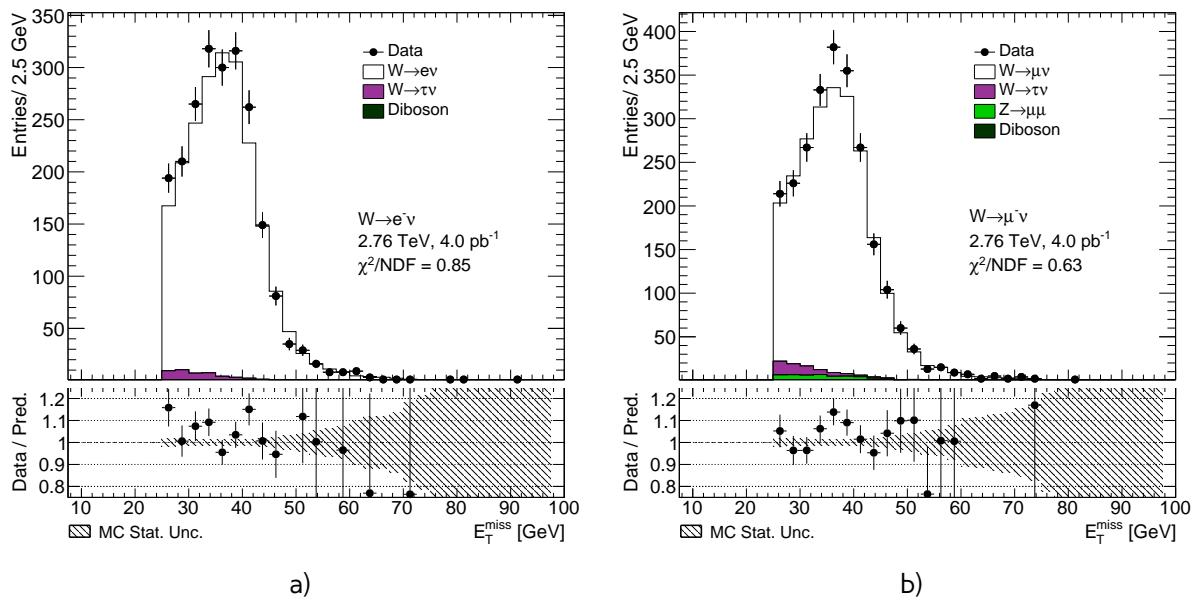


Fig. 14.9: Missing transverse energy distribution from the a)  $W^- \rightarrow e^-\nu$  selection and b) the  $W^- \rightarrow \mu^-\nu$  selection.

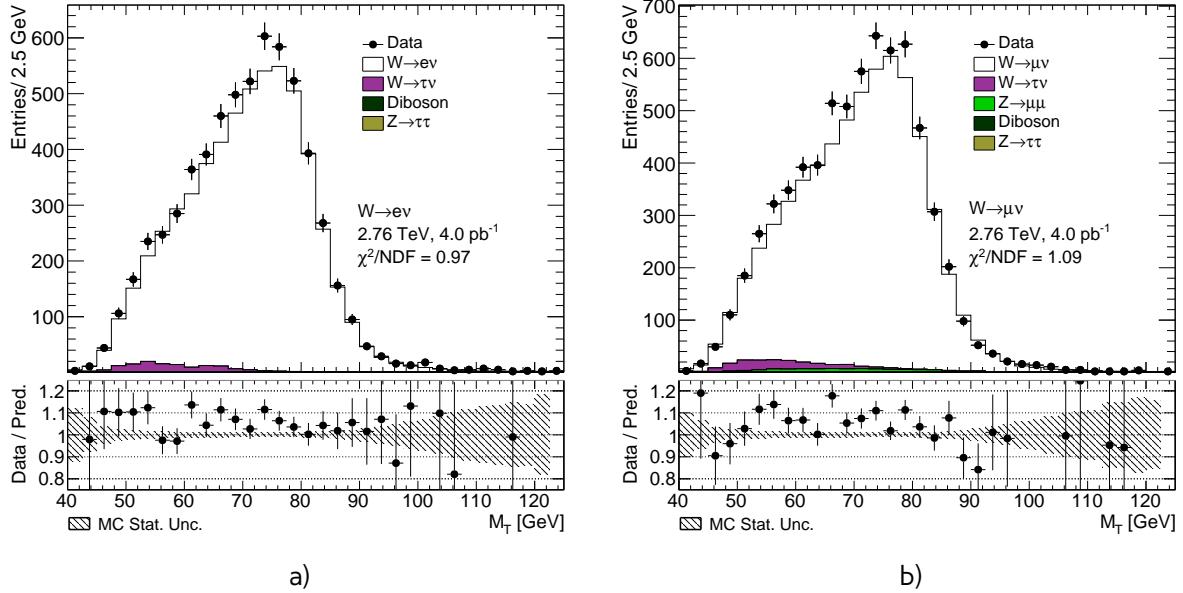


Fig. 14.10: Transverse mass distribution distribution from the a)  $W \rightarrow e\nu$  selection and b) the  $W \rightarrow \mu\nu$  selection.

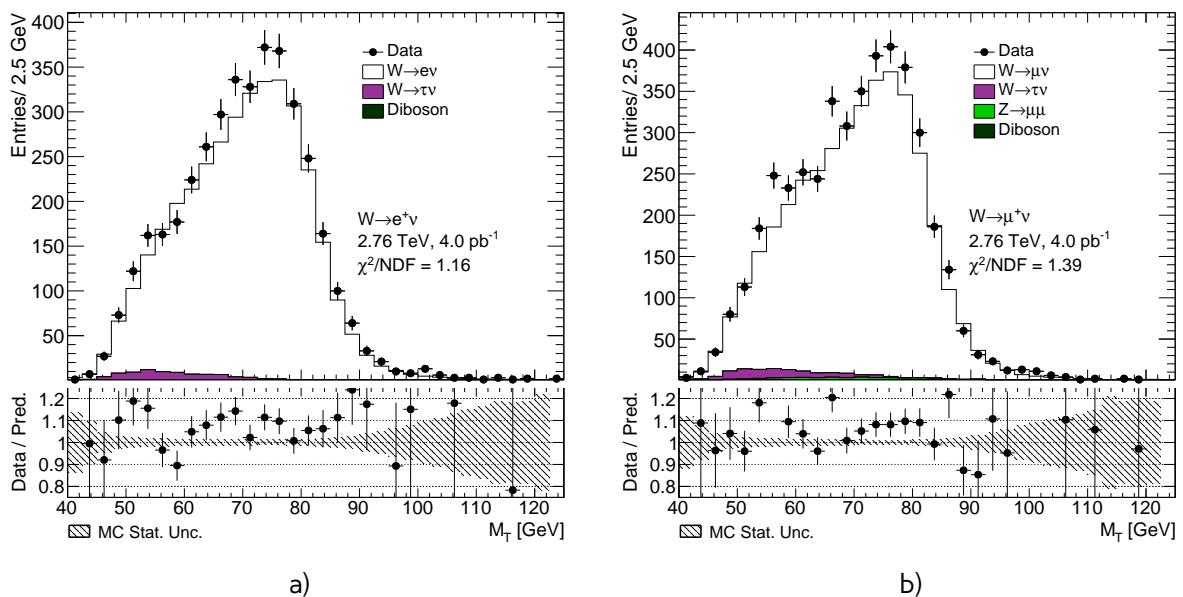


Fig. 14.11: Transverse mass distribution distribution from the a)  $W^+ \rightarrow e^+\nu$  selection and b) the  $W^+ \rightarrow \mu^+\nu$  selection.

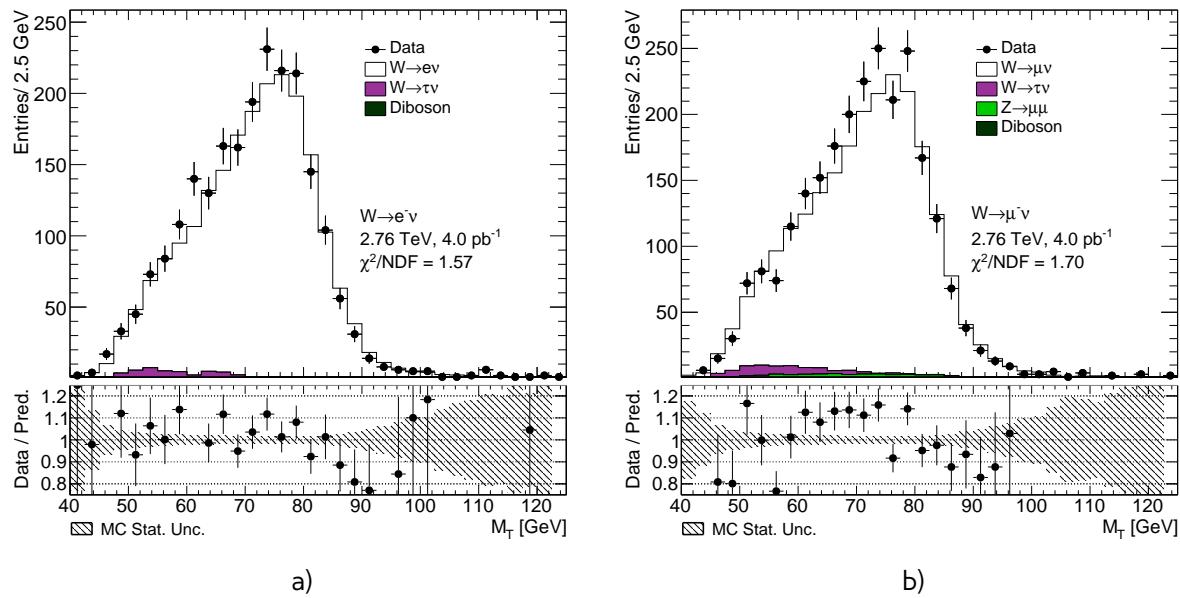


Fig. 14.12: Transverse mass distribution distribution from the a)  $W^- \rightarrow e^-\nu$  selection and b) the  $W^- \rightarrow \mu^-\nu$  selection.

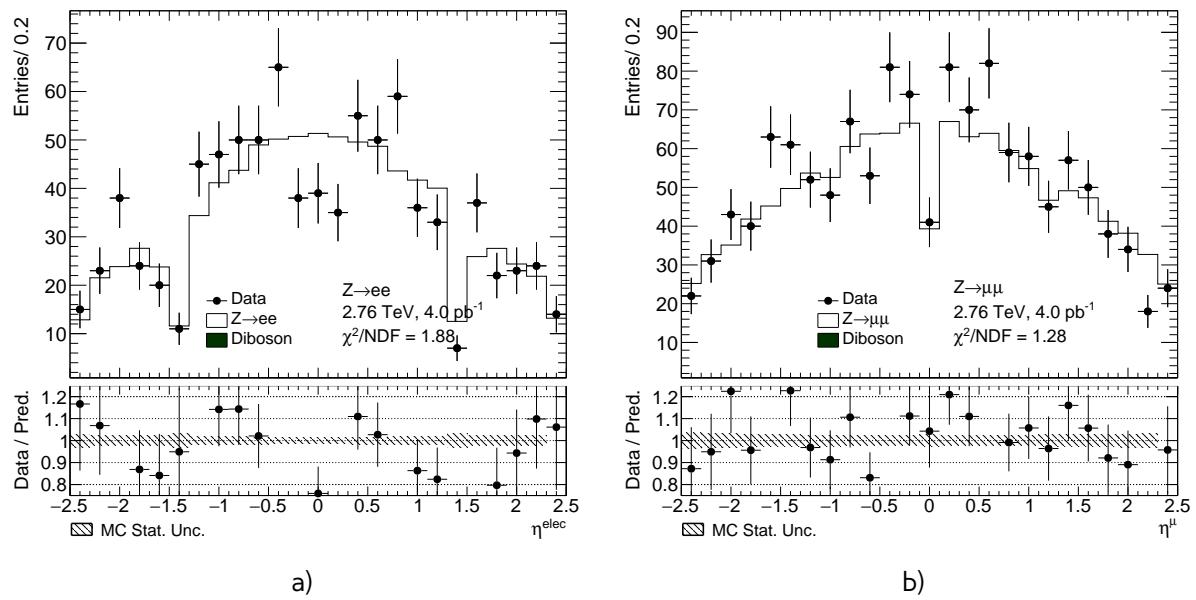


Fig. 14.13: Lepton pseudorapidity distributions from the a)  $Z \rightarrow e^+e^-$  b)  $Z \rightarrow \mu^+\mu^-$

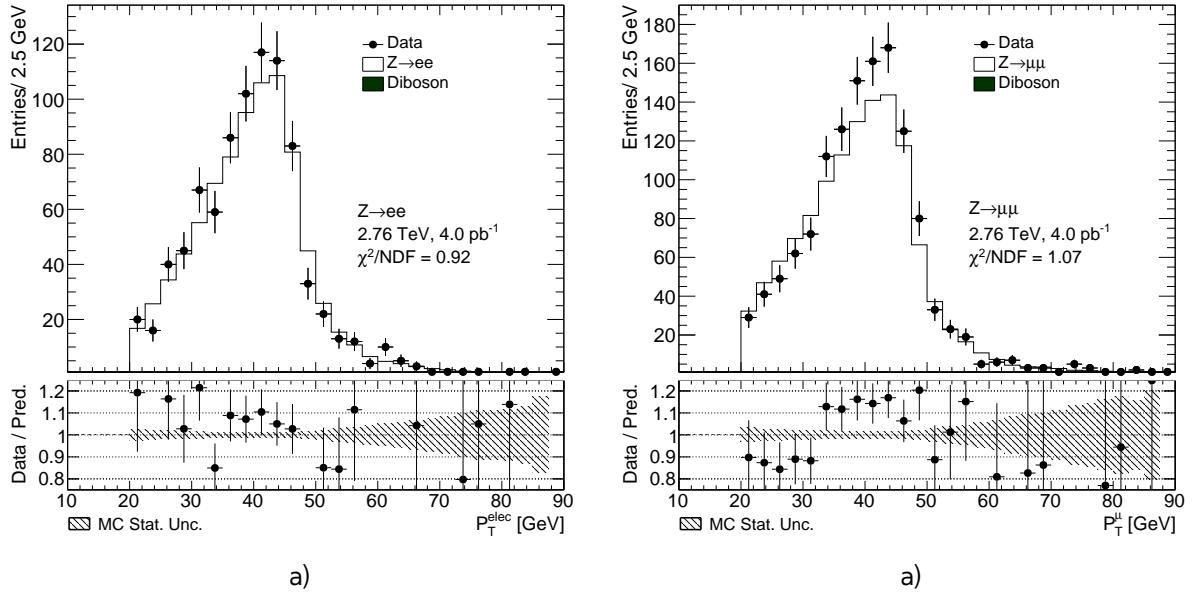


Fig. 14.14: Lepton transverse momentum distributions from the a)  $Z \rightarrow e^+e^-$  b)  $Z \rightarrow \mu^+\mu^-$

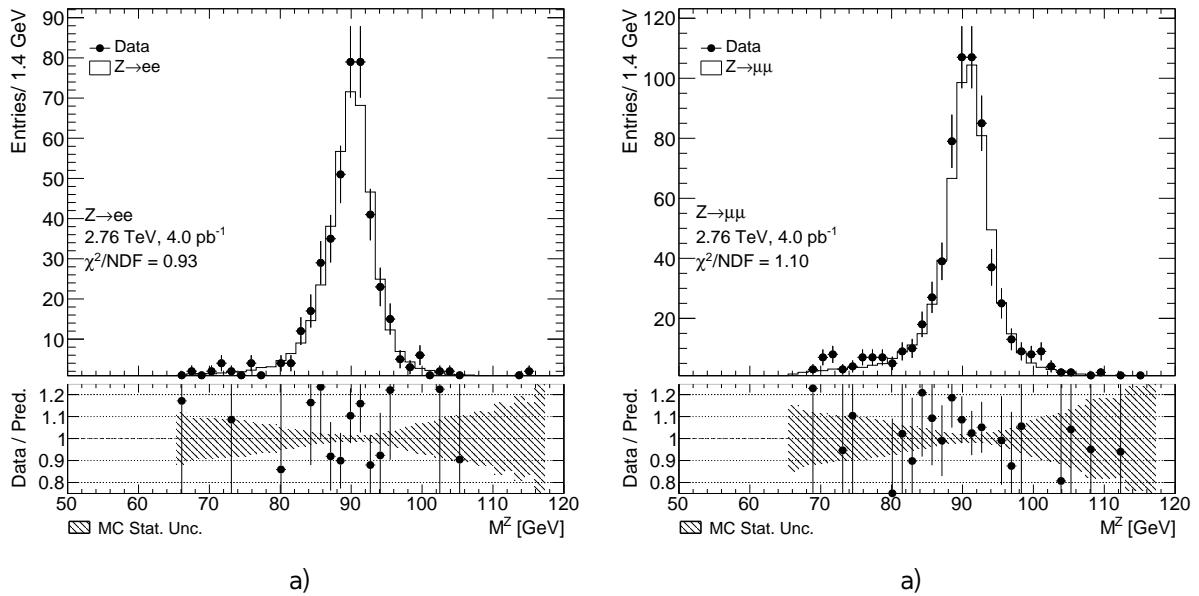


Fig. 14.15: Dilepton mass distribution distributions from the a)  $Z \rightarrow e^+e^-$  b)  $Z \rightarrow \mu^+\mu^-$

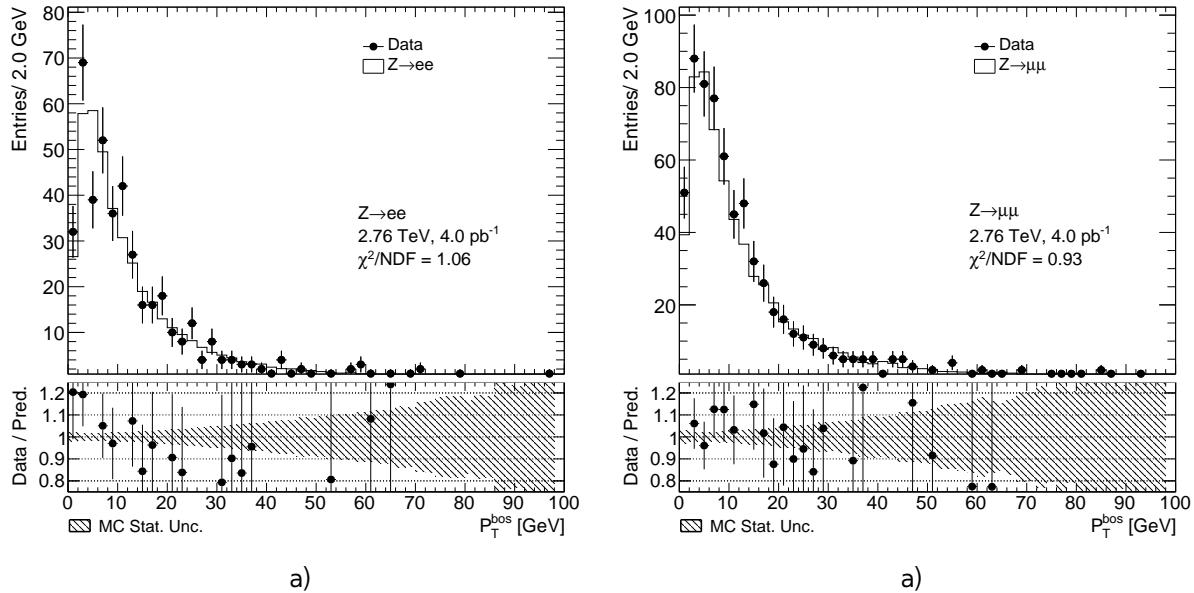


Fig. 14.16: Z boson transverse momentum distributions from the a)  $Z \rightarrow e^+e^-$  b)  $Z \rightarrow \mu^+\mu^-$

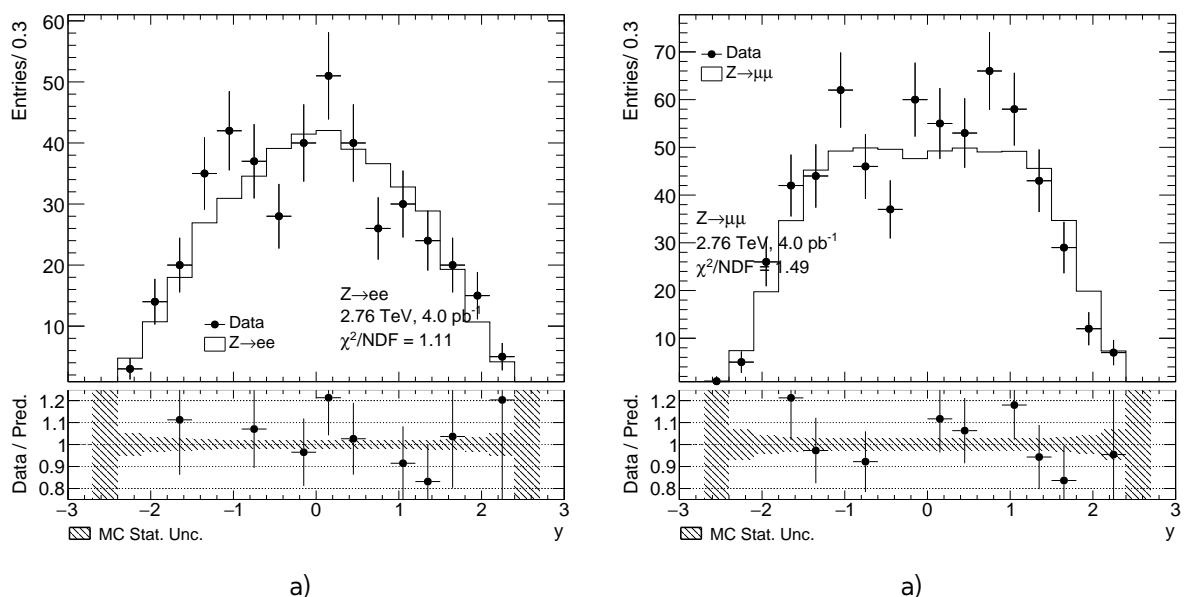


Fig. 14.17: Z boson rapidity distribution from the a)  $Z \rightarrow e^+e^-$  b)  $Z \rightarrow \mu^+\mu^-$



648

# Chapter 15

## 649 Uncertainties in the cross-section measurement

650 Cross-section measurement relies on theoretical models and corrections, used in Monte-Carlo. Thus,  
 651 their intrinsic uncertainties should be propagated to a final result. This chapter discusses main  
 652 methods of uncertainties measurements and sources on  $C_{W,Z}$  and  $A_{W,Z}$  correction factors.

### 653 15.1 Methods of uncertainties propagation

The offset method changes a correction by a  $\pm 1\sigma$  of its systematic uncertainty. The contribution of each correction's uncertainty on the observable (e.g.  $C_{W,Z}$ ,  $A_{W,Z}$  or a cross-section) is taken as a symmetric approximation:

$$U_i^{offset} = \frac{\sigma_i^{up} - \sigma_i^{down}}{2}, \quad (15.1)$$

654 where  $\sigma_i^{up(down)}$  - the change in a observable due to the shift of the correction on  $\sigma$  up or down.

Another method used for a uncertainties propagation is a toy MC method, that uses a pseudo experiments with modified input corrections. For a scale factors binned  $p_T$  and  $\eta$  uncertainties inside each bin can be divided to a correlated and uncorrelated systematic components and statistical error. For each pseudo-experiment, a table of new scale factors is filled, where inside each bin a scale factor is randomly varied as:

$$SF_i^{Toy_n} = SF_i + Gauss(0, \Delta SF_i^{uncorr+stat}) + \sum \Delta SF_i^{corr} \cdot Gauss(0, 1), \quad (15.2)$$

655 where  $SF_i^{Toy_n}$  is a new scale factor in i-th bin,  $\Delta SF_i^{uncorr+stat}$  - is the quadratic sum of uncorrelated  
 656 and statistical errors and  $\Delta SF_i^{corr}$  is a correlated error.

The overall effect on a observable is calculated as a standard deviation of the values in a pseudo-experiments:

$$U_i = \sqrt{\frac{\sum_{Toy_n=1}^N \sigma_i^2}{N} + \frac{\sum_{Toy_n=1}^N \sigma_i^2}{N}} \quad (15.3)$$

657 The number N of pseudo experiments should be sufficiently large to avoid possible bias in the  
 658 uncertainty estimation.

### 659 15.2 Experimental systematic uncertainties

660 Sources of experimental uncertainties, methods of estimation and their effect on a  $C_{W,Z}$  are sum-  
 661 marized in a Tab. 15.2. Systematical errors coming from a hadron recoil calculation are discussed in  
 662 a Sec. 12.

### **15.2.1 Electron energy scale and resolution**

663 Electron energy scale correction, described in Sec. 11.2 has associated uncertainties coming from  
664 <reference>:

- 666 • Statistical component of the scale uncertainty  
667 • Uncertainty from the possible bias of the calibration method  
668 • Scale uncertainty from the choice of generator  
669 • Uncertainty from the presampler energy scale  
670 • Imperfect knowledge of the material in front of EM calorimeter.

671 The uncertainty contribution from each component is estimated using offset method. The total  
672 energy scale uncertainty is the quadratic sum of the components <reference>.

### **15.2.2 Muon energy scale and resolution**

- 674 • MS modelling  
675 • ID modelling  
676 • overall scale

677 The uncertainty contribution from each component is estimated using offset method. The total  
678 energy scale uncertainty is the quadratic sum of the components

### **15.2.3 Muon and electron efficiency toy Monte-Carlo**

680 In case of 2.76 TeV analysis scale factor errors are considered to be enlarged for a statistical and  
681 uncorrelated components, so correlated error is assumed to be negligible. The toy MC experiments  
682 are performed for electron reconstruction, identification and trigger scale factors and muon recon-  
683 struction + identification scale factors. In the current analysis 30 pseudo-experiments are used with  
684 a combined toy MC method. Plots for  $C_W$  effect on a cross-section. Correlation

Table 15.1

| Source of uncertainty        | Method | $\delta C_W/C_W(\%)$<br>$W^+ \rightarrow e\nu$ | $\delta C_W/C_W(\%)$<br>$W^- \rightarrow e\nu$ | $\delta C_W/C_W(\%)$<br>$W^+ \rightarrow \mu\nu$ | $\delta C_W/C_W(\%)$<br>$W^+ \rightarrow \mu\nu$ | $\delta C_Z/C_Z(\%)$<br>$Z \rightarrow ee$ | $\delta C_Z/C_Z(\%)$<br>$Z \rightarrow \mu\mu$ |
|------------------------------|--------|--|--|--|--|--|--|
| Electron reconstruction      | Toy MC | 0.11   | 0.09   | -  | -  | 0.12                                       | -  |
| Electron identification      | Toy MC | 0.32   | 0.30   | -  | -  | 0.54                                       | -  |
| Electron trigger efficiency  | Toy MC | 0.14   | 0.13   | -  | -  | 0.00                                       | -  |
| Muon reco+id                 | Toy MC | -  | -  | 0.03   | 0.02   | -  | 0.03   |
| Muon trigger                 | Offset | -  | -  | 1.01   | 1.01   | -  | 0.02   |
| Electron energy scale        | Offset |  |  | -  | -  |  | -  |
| - Statistical error          | Offset |  |  | -  | -  |  | -  |
| - Bias in method             | Offset |  |  | -  | -  |  | -  |
| - Scale uncertainty          | Offset |  |  | -  | -  |  | -  |
| - Presampler energy scale    | Offset |  |  | -  | -  |  | -  |
| - Material knowledge         | Offset |  |  | -  | -  |  | -  |
| Electron energy resolution   | Offset | 0.05   | 0.03   | -  | -  | 0.03                                       | -  |
| Muon energy scale            | Offset | -  | -  | 0.05   | 0.05   | -  | 0.03   |
| Muon energy resolution total | Offset | -  | -  | Wplusmumu<br>Wminmumu                            |  | -  | Zmumu  |
| - Muon ID energy scale       | Offset | -  | -  | 0.02   | 0.01   | -  | 0.01   |
| - Muon MS energy scale       | Offset | -  | -  | 0.01   | 0.00   | -  | 0.01   |
| Hadron recoil scale          | Offset |  |  |  |  | -  | -  |
| Hadron recoil resolution     | Offset |  |  |  |  | -  | -  |
| EWK + $t\bar{t}$ background  |        |  |  |  |  |  |  |
| QCD                          |        |  |  |  | -  | -  |  |
| PDF error                    |        |  |  |  |  |  |  |
| Total                        |        |  |  |  |  |  |  |
| Statistics                   |        |  |  |  |  |  |  |
| Luminosity                   |        |  |  |  |  |  |  |

Table 15.2

| Source of uncertainty       | Method | $\delta C_W/C_W(\%)$<br>$W^+ \rightarrow e\nu$ | $\delta C_W/C_W(\%)$<br>$W^- \rightarrow e\nu$ | $\delta C_W/C_W(\%)$<br>$W^+ \rightarrow \mu\nu$ | $\delta C_W/C_W(\%)$<br>$W^+ \rightarrow \mu\nu$ |
|-----------------------------|--------|--|--|--|--|
| Electron reconstruction     | Toy MC | 0.11   | 0.09   | -  | -  |
| Electron identification     | Toy MC | 0.32   | 0.30   | -  | -  |
| Electron trigger efficiency | Toy MC | 0.14   | 0.13   | -  | -  |
| Muon reco+id                | Toy MC | -  | -  | 0.03   | 0.02   |
| Muon trigger                | Offset | -  | -  | 0.03   | 0.02   |
| Electron energy scale       | Offset | Wplusenu                                       | Wminenu  | -  | -  |
| Electron energy resolution  | Offset |  |  | -  | -  |
| Muon energy scale           | Offset | -  | -  | Wplusmunu  | Wminmunu   |
| Muon energy resolution      | Offset | -  | -  |  |  |
| Hadron recoil scale         | Offset | Wplusenu                                       | Wminenu  | Wplusmunu  | Wminmunu   |
| Hadron recoil resolution    | Offset |  |  |  |  |
| Total                       |        |  |  |  |  |

| Source of uncertainty       | Method | $\delta C_Z/C_Z(\%)$<br>$Z \rightarrow ee$ | $\delta C_Z/C_Z(\%)$<br>$Z \rightarrow \mu\mu$ |
|-----------------------------|--------|--|--|
| Electron reconstruction     | Toy MC | 0.12                                       | Zmumu  |
| Electron identification     | Toy MC | Zee  | Zmumu  |
| Electron trigger efficiency | Toy MC | 0.00                                       | Zmumu  |
| Muon reco+id                | Toy MC | Zee  | Zmumu  |
| Muon trigger                | Offset | Zee  | Zmumu  |
| Electron energy scale       | Offset | Zee  | Zmumu  |
| Electron energy resolution  | Offset | 0.03                                       | Zmumu  |
| Muon energy scale           | Offset | Zee  | Zmumu  |
| Muon energy resolution      | Offset | Zee  | Zmumu  |
| Total                       |        |  |  |

### 15.3 Theoretical uncertainty

- Theoretical uncertainties on the predictions are mostly dominated by a imperfect knowledge of the proton PDF's. They are affecting both  $A_{W,Z}$  and  $C_{W,Z}$ . Error coming from an arbitary choice of PDF set is estimated by PDF reweighting <reference> of original MC generated using <something> to a one of the 4 pdf sets: CT10 <reference>, ATLAS-epWZ12 <reference>, abkm09 <reference> and NNPDF23 <reference>. The error is calculated as a standard deviance for all of the sets.
- Systematic uncertainty within one pdf set is evaluated using CT10 NLO set. This set contains 52 asosiated error sets, corresponding to a 90% C.L. limits along 26 eigenvectors. The resulting

| PDF Set      | $\delta A_W / A_W (\%)$  | $\delta C_W / C_W (\%)$ | $\delta A_W / A_W (\%)$ | $\delta C_W / C_W (\%)$  |
|--------------|--------------------------|-------------------------|-------------------------|--------------------------|
|              | $W^+ \rightarrow e\nu$   |                         |                         | $W^- \rightarrow e\nu$   |
| CT10         |                          |                         |                         |                          |
| ATLAS-epWZ12 |                          |                         |                         |                          |
| abkm09       |                          |                         |                         |                          |
| NNPDF23      |                          |                         |                         |                          |
|              | $W^+ \rightarrow \mu\nu$ |                         |                         | $W^- \rightarrow \mu\nu$ |
| CT10         |                          |                         |                         |                          |
| ATLAS-epWZ12 |                          |                         |                         |                          |
| abkm09       |                          |                         |                         |                          |
| NNPDF23      |                          |                         |                         |                          |

52 variation are separately added in a quadrature as:

$$\delta_X = \frac{1}{2} \cdot \sqrt{\sum_{i=1}^N (X^+ - X^-)^2} \quad (15.4)$$

- The uncertainties arising from the choice of generator and parton showering model are considered small. They can be calculated as a difference in the acceptances  $A_{W,Z}$  for MC samples, generated using same PDF set, but different models for showering and matrix element, namely Powheg + Pythia and Sherpa

## 15.4 Correlation between uncertainties

### 15.4.1 Toy MC correlations

A correlation coefficient between two observables  $o_1$  and  $o_2$  can be estimated as:

$$\rho_{12} = \frac{1}{\sigma(o_1)\sigma(o_2)} \cdot \frac{1}{N-1} \sum_{i=1}^N (o_1^i - \bar{o}_1)(o_2^i - \bar{o}_2) \quad (15.5)$$

Using cholesky transformation this uncertainty can be propagated to a 3 eigenvectors:

### 15.4.2 Correlations between PDF's eigenvectors

Correlations for  $A_{W,Z}$  and  $C_{W,Z}$  for a CT10nlo set have been estimated, since they could affect error on a total W boson cross section measurement and PDF fits. Given two processes X and Y (

$$\delta_{XY}^2 = \delta_X^2 + \delta_Y^2 + 2\delta_X\delta_Y\rho_{XY} \quad (15.6)$$

$$\rho_{XY} = \frac{1}{4\delta_X\delta_Y} \sum (X^+ - X^-) \cdot (Y^+ - Y^-) \quad (15.7)$$



700

## **Part V**

701

# **Results**



702 Chapter **16**

703 **Results of the Cross Section Measurement**

704 **16.1 Combined results**

705 **16.2 Comparation with Theoretical Predictions**



706

# Chapter 17

707

## PDF fits results



708

# Chapter 18

709

## Summary