

# 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 Standart 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>III Experimental setup</b>   | <b>9</b>  |
| <b>3 The Large Hadron Collider</b>  | <b>11</b> |
| <b>4 The ATLAS experiment</b>   | <b>13</b> |
| 4.1 Inner detector . . . . .  | 13        |
| 4.2 Calorimeters . . . . .  | 13        |
| 4.2.1 Forward calorimeters . . . . .  | 13        |
| 4.3 Muon system . . . . .   | 13        |
| 4.4 The magnet system . . . . .   | 13        |
| 4.5 Trigger system . . . . .  | 13        |
| 4.6 Atlas software . . . . .  | 13        |
| <b>5 Event reconstruction</b>   | <b>15</b> |
| 5.1 Tracks and vertexes . . . . .   | 15        |
| 5.2 Electron reconstruction and identification . . . . .                    | 15        |
| 5.3 Muon reconstruction and identification . . . . .                        | 15        |
| <b>6 Monte Carlo simulation</b>   | <b>17</b> |
| 6.1 ATLAS chain of MC production . . . . .                                  | 17        |
| 6.2 Event generators . . . . .  | 18        |
| 6.3 Simulation in Geant4 . . . . .  | 19        |
| <b>7 Frozen Showers</b>   | <b>21</b> |
| 7.1 Problem description . . . . .   | 22        |
| 7.2 Frozen showers generation and use in a Monte-Carlo production . . . . . | 25        |
| 7.3 Library generation . . . . .  | 27        |
| 7.4 Use of Frozen Showers in MC production . . . . .                        | 27        |
| 7.5 Tuning of Frozen Showers libraries . . . . .                            | 27        |

|  |           |
|--|-----------|
| 7.6 Validation of the new libraries . . . . .                                    | 29        |
| 7.7 Plans for a future . . . . .   | 29        |
| <b>8 Data and Monte-Carlo samples</b>  | <b>31</b> |
| 8.1 Data sample . . . . .  | 31        |
| 8.2 Monte-Carlo samples . . . . .  | 31        |
| <b>IV The Measurement</b>  | <b>33</b> |
| <b>9 Selection</b>   | <b>35</b> |
| 9.1 Data quality cuts . . . . .  | 35        |
| 9.2 Lepton quality cuts . . . . .  | 36        |
| 9.3 Boson selection . . . . .  | 36        |
| 9.4 Cut flow . . . . .   | 36        |
| <b>10 Monte Carlo corrections</b>  | <b>37</b> |
| 10.1 Lepton efficiency corrections . . . . .                                     | 37        |
| 10.1.1 Muon Trigger SF . . . . .   | 38        |
| 10.2 Electron energy scale and resolution . . . . .                              | 39        |
| 10.3 Muon momentum correction . . . . .  | 40        |
| <b>11 Missing Transverse Energy reconstruction and correction</b>                | <b>43</b> |
| 11.1 Standard Missing Transverse Energy reconstruction . . . . .                 | 43        |
| 11.2 Reconstruction of Missing Transverse Energy from hadron recoil . . . . .    | 45        |
| 11.3 Hadron Recoil calibration . . . . .   | 46        |
| 11.3.1 Hadron recoil resolution correction . . . . .                             | 47        |
| 11.3.2 Hadron recoil bias correction . . . . .                                   | 52        |
| <b>12 Background estimation</b>  | <b>59</b> |
| 12.1 QCD background estimation . . . . .   | 60        |
| 12.1.1 Template selection . . . . .  | 61        |
| 12.1.2 Methodology of the template sample normalization . . . . .                | 61        |
| 12.1.3 Systematic Uncertainty from the Multi-jet Background Estimation . . . . . | 63        |
| 12.2 Background-subtracted W and Z candidate events . . . . .                    | 65        |
| <b>13 Control distributions</b>  | <b>67</b> |
| <b>14 Methodology of Cross-Section Measurement</b>                               | <b>77</b> |
| 14.1 Fiducial volume definition . . . . .  | 77        |
| <b>15 Uncertainties in the cross-section measurement</b>                         | <b>79</b> |
| 15.1 Methods of uncertainties propagation . . . . .                              | 79        |
| 15.2 Experimental systematic uncertainties . . . . .                             | 79        |
| 15.2.1 Electron energy scale and resolution . . . . .                            | 80        |
| 15.2.2 Muon energy scale and resolution . . . . .                                | 80        |
| 15.2.3 Muon and electron efficiency toy Monte-Carlo . . . . .                    | 80        |
| 15.3 Theoretical uncertainty . . . . .   | 82        |

|  |           |
|--|-----------|
| 15.4 Correlation between uncertainties . . . . .         | 83        |
| 15.4.1 Toy MC correlations . . . . .                     | 83        |
| 15.4.2 Correlations between PDF's eigenvectors . . . . . | 83        |
| <b>16 Results of the Cross Section Measurement</b>       | <b>85</b> |
| 16.1 Cross-section combination . . . . .                 | 85        |
| 16.2 Comparation with Theoretical Predictions . . . . .  | 85        |
| <b>17 PDF fits results</b>                               | <b>87</b> |
| <b>18 Summary</b>  | <b>89</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

## **Part III**

13

# **Experimental setup**



14

# Chapter 3

## 15 The Large Hadron Collider



16 Chapter **4**

17 **The ATLAS experiment**

18 **4.1 Inner detector**

19 **4.2 Calorimeters**

20 **4.2.1 Forward calorimeters**

21 **4.3 Muon system**

22 **4.4 The magnet system**

23 **4.5 Trigger system**

24 **4.6 Atlas software**



25

# Chapter **5**

## 26 **Event reconstruction**

27 **5.1 Tracks and vertexes**

28 **5.2 Electron reconstruction and identification**

29 **5.3 Muon reconstruction and identification**



30 Chapter 6

31 **Monte Carlo simulation**

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

38 **6.1 ATLAS chain of MC production**

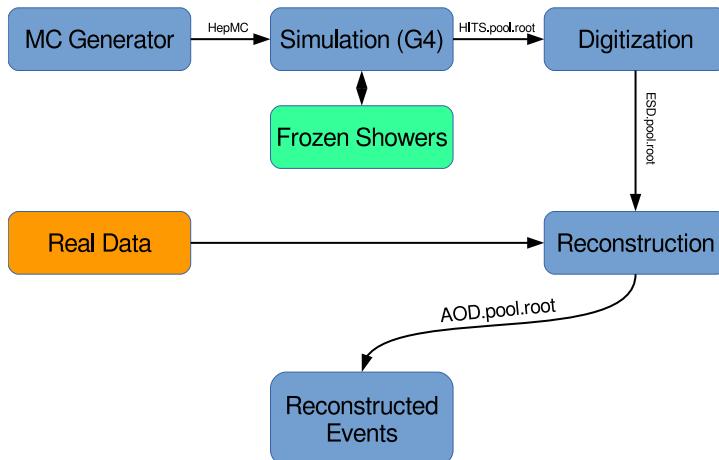


Fig. 6.1: Diagram of the ATLAS MC production chain

39 Monte Carlo method allows to perform different analysis, generate predictions for comparisons  
40 with data, study detector or selection algorithms performance. All of these applications require  
41 accurate MC predictions. Simulation software expects to use precise physics models for sampling  
42 and have large enough statistics, to exclude statistical uncertainties (usually 5 times more, than  
43 expected in a data). ATLAS simulation software is integrated into Athena framework and usually used  
44 during large production of events. Simulation chain is generally divided into 4 main steps (Figure 6.1):

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

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

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

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

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

58 

## 6.2 Event generators

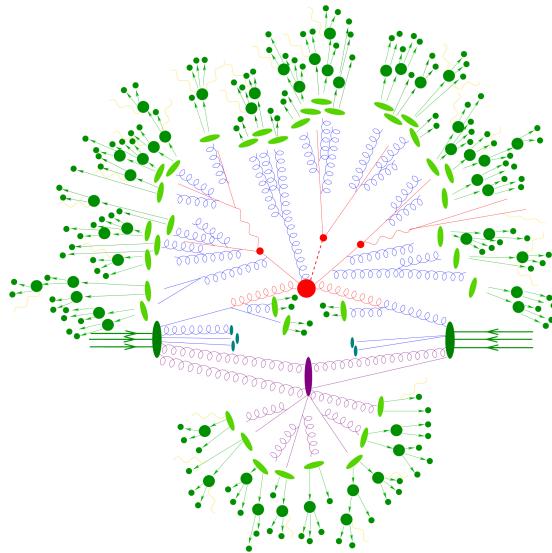


Fig. 6.2: Schematic view of a  $t\bar{t}H$  event produced in a  $pp$ -collision: the hard scattering is shown as a red blob with the solid and dashed lines as the resulting three particles. Independently happening multi-particle interactions are indicated by the violet blob. Parton showers are shown with curly lines. Hadronization yields hadrons as shown in light green, while the final state particles are dark green.

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

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

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

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

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

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

83 The current analysis uses samples generated with the following generators:

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

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

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

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

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

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

## 98 **6.3 Simulation in Geant4**

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

103 Geant4 can simulate complex detector structures with sensitive detector material and corresponding  
104 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. 7

127 

## Frozen Showers

128 As it was mentioned in a previous chapter, fast simulation techniques are the essential part of  
 129 Monte-Carlo production at ATLAS experiment. Typical time for a simulation of 1  $t\bar{t}$  event is around 1  
 130 minute, and most of the time is spend on a simulation of particle interaction in calorimeters. This  
 131 is a main motivation of development of fast calorimetry techniques. Frozen showers is currently the  
 132 main fast calorimeter simulation approach used at ATLAS experiment. In this chapter we will discuss  
 133 main principles and current developments in optimization of this method.

134 Frozen shower method uses pre-simulated "frozen" showers instead of the full simulation. This is  
 135 allowing to reduce time spend on a simulation of a large amount of low energy sub showers. Typical  
 136 number of frozen showers used in a simulation of 1  $t\bar{t}$  event is. This method is allowing to have  
 137 a 25% speedup.

138 This method requiring in advance generation of a libraries for each detector and particle used in  
 139 this method. Later, during fast simulation, if a particle energy falls below a cutoff it is replaced by a  
 140 shower from a library. Main parameters used in ATLAS simulation are summarized in a Tab. 7.1.

Table 7.1: Main parameters used for the frozen shower libraries in FCAL

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

## 7.1 Problem description

Fast simulation of forward calorimeters (FCAL) is a complicated task due to its complex structure. As it was mentioned in a Sec. 4.2.1 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:

$$\frac{\sigma}{E} \approx \frac{1}{\sqrt{E}} \oplus \frac{1}{E} \oplus \text{const}, \quad (7.1)$$

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, generated inside a different points in forward calorimeter. Shower energy  $E^{shower}$  distribution in the x vs y plane is showed on Fig. 7.1, where shower energy is defined as:

$$E^{shower} = \sum E_i^{hits}, \quad (7.2)$$

where  $E_i^{hits}$  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 distance to a closest rod center. Fig. 7.2 presents a

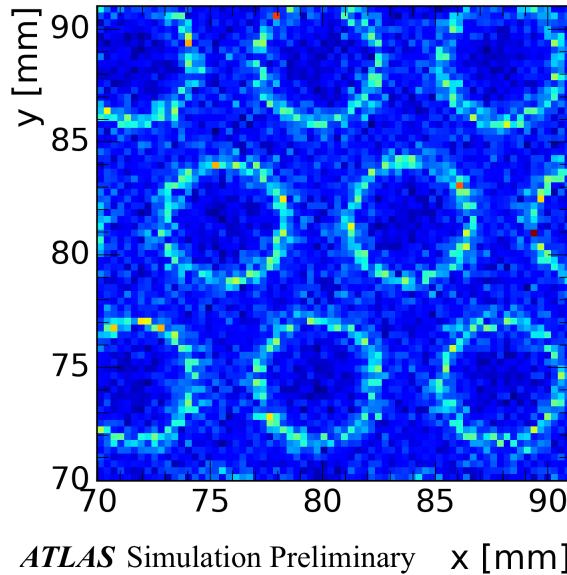


Fig. 7.1: Shower energy response histogram in the x vs y plane for electrons, generated with uniformly distributed x and y and energy less than 1 GeV. Light circles are corresponding to a showers, started inside a LAr gaps with on average higher energy response, while the dark parts are corresponding to dead material respectively with smaller sum of the "hits" energy.

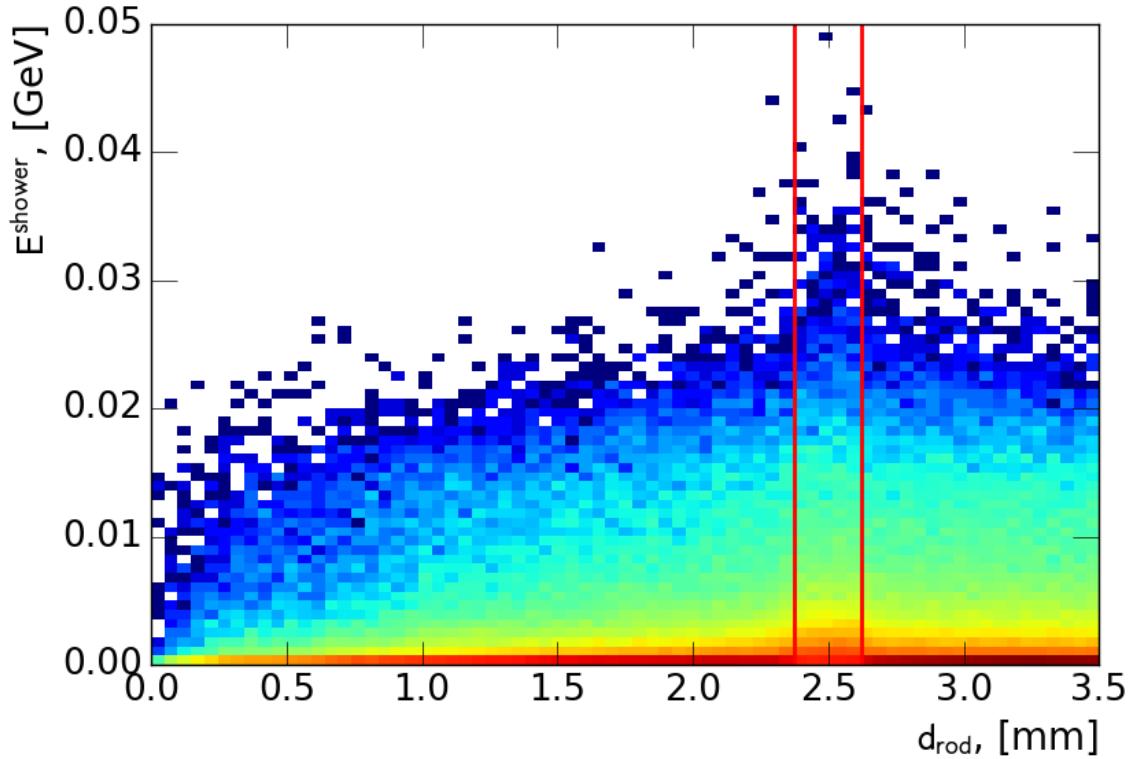


Fig. 7.2: Distribution of electron showers for electrons with energy less than 1 GeV coming from initial electron with energy 1 TeV in distance to a closest rod center vs shower energy plane. Position of a liquid argon gap is noted by a red lines. There is visible difference in shower properties between showers inside and outside of the liquid argon gaps

154 distribution showers for electrons with energy below 1 GeV coming from initial electrons with energy  
 155 1 TeV in the distance to a closest rod center vs shower energy plane. Liquid argon gap is marked by  
 156 a red lines. There is a clear difference in a showers energies between electrons born in a sensitive  
 157 and dead material. Difference in a shower properties are also visible for number of hits (Fig. 7.3 a)  
 158 and standard deviation energy of hits in shower (Fig. 7.3 b) distributions. Size of this differences  
 159 depends on a electron energies and higher for a smaller energies (Fig. 7.4 a) and less significant  
 160 for a higher energies (Fig. 7.4 b). In order to simulate resolution of electrons appropriately , frozen  
 161 showers method should simulate this distributions close to the nominal.

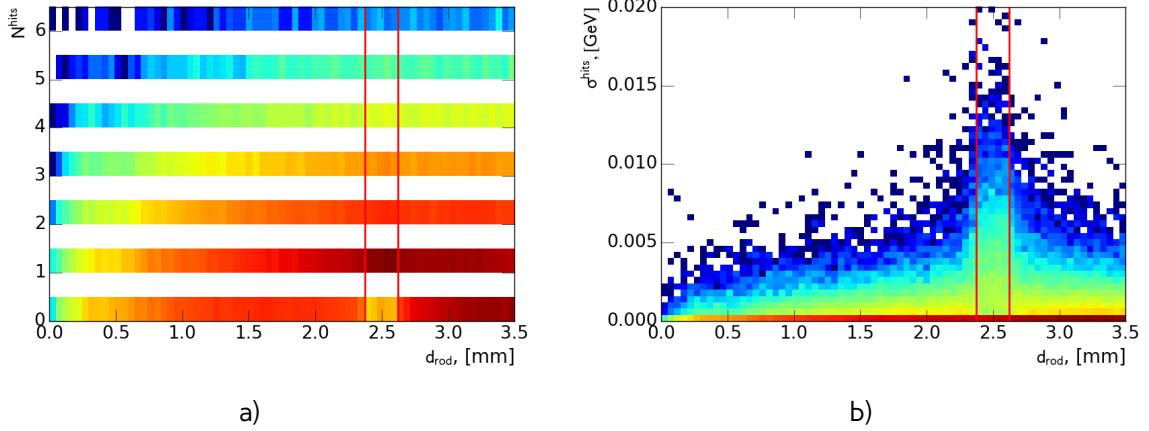


Fig. 7.3: Distribution of electron showers for electrons with energy less than 1 GeV coming from initial electron with energy 1 TeV in distance to a closest rod center vs a) number of hits in a shower plane and b) standard deviation of hits in a shower energy. Position of a liquid argon gap is noted by a red lines. There is visible difference in shower properties between showers inside and outside of the liquid argon gaps

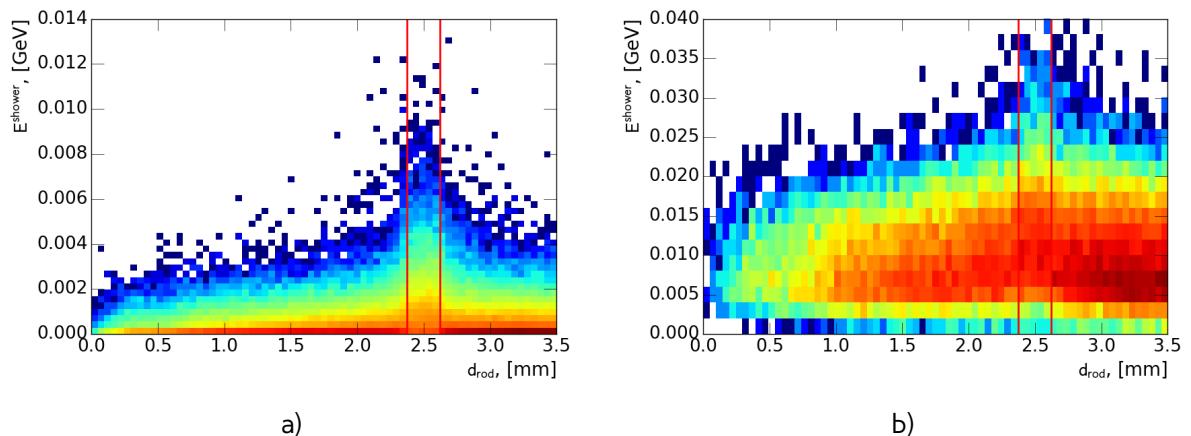


Fig. 7.4: Distribution of electron showers for electrons a) less than 100 MeV and b) higher than 300 GeV coming from initial electron with energy 1 TeV in distance to a closest rod center vs shower energy plane. Position of a liquid argon gap is noted by a red lines. Size of the difference in a shower properties depends on the energy of the electrons and higher for smaller energies

## 7.2 Frozen showers generation and use in a Monte-Carlo production

Frozen showers is the main method used in ATLAS experiment for a fast simulation of physics inside calorimeter. It replaces a full simulation of a shower with a "frozen" shower from a library. This method consists of 2 stages: library generation and a production use. Because most of the time is spend on a simulation of electromagnetic showers inside foward calorimeters due to their complex structure (Sec. ??) and a large particle multiplicity, frozen showers are used in simulation of FCAL1 and FCAL2. Main parameters of a frozen shower libraries are summarized in Tab. 7.1.

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

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 dependes also on a energy of particle, because shower sizes are growing with energy. When particle is removed and substituted by shower taken from corresponding eta and distance bin with the closest energy found. Energies of the hits in shower found are scaled to fully correspond to particle energy. Additionally, shower direction is changed to the direction of the particle.

Frozen Showers have been used in ATLAS Monte-Carlo production since run-1. This method is applicable for all LAr calorimeters in ATLAS, but currently it is enabled for simulation of forward calorimeters (FCAL).

It uses "frozen" showers generated using full simulation. This showers are stored in a Particles below minimum energy thresholds are killed and replaced with these showers. All of the other particles are simulated using full simulation. This process is schematically shown in a Figure 7.5.

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

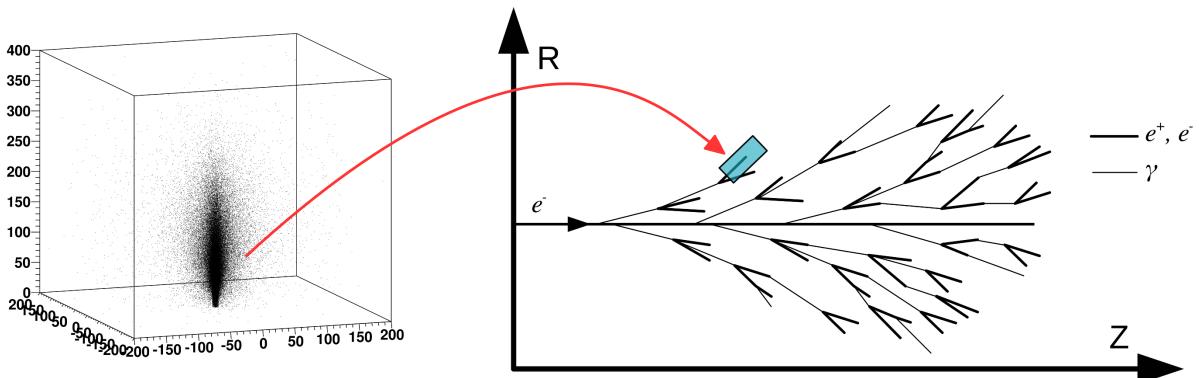


Fig. 7.5: Diagram showing the shower substitution of the low-energy particle, during the high-energy particle simulation.

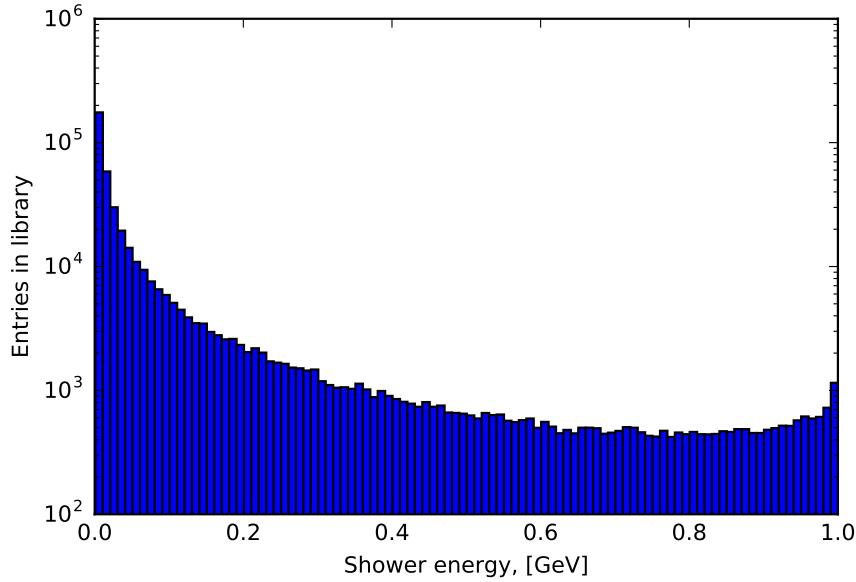


Fig. 7.6: Distribution of shower energy used in production of 1000 GeV electrons.

shower in a library slower, than simulation of shower with zero or one hit in a sensitive detector.  
 In a Frozen Shower method there are separate libraries for each particle and subdetector used.  
 Showers should cover fully energy and pseudorapidity region and be able to describe data, that is  
 needed during simulation. This is why 2 stages simulation approach have been used. The first stage  
 takes initial particle parameters from a physical processes (ttbar or a single electron).

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

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

Unfortunately, for a Frozen Showers, generated for Run-1 monte-carlo, additional tuning of electron  
 libraries was needed. This was done using reconstructed energy of electrons. Frozen Showers tend  
 to underestimate fluctuations of energy loss, that is leading to a smaller electron resolution for a  
 high energies. Correction is done by enlarging bin, corresponding to a gap position. Also, correction

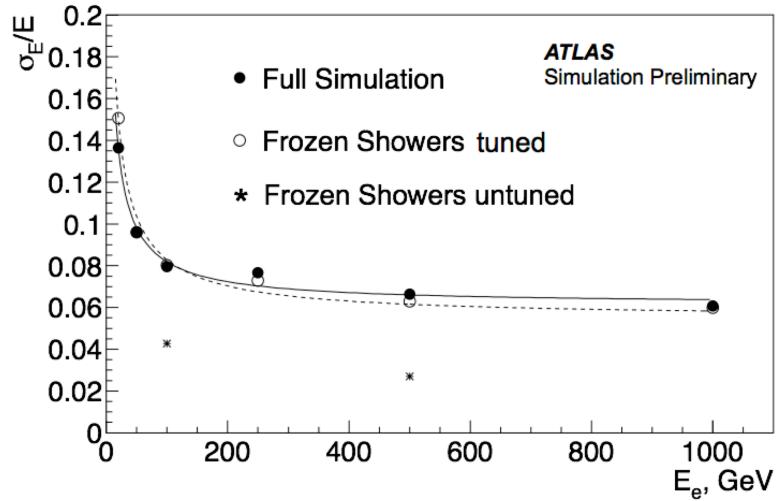


Fig. 7.7: Electron resolution for full simulation, tuned and untuned frozen showers

216 of the mean shift is done by scaling energy response of all showers. After this frozen showers are  
 217 showing good agreement with full simulation. This procedure needs to be done every time, when  
 218 something is changing in software. Because tuning is done manually, lots of manpower is needed  
 219 for each Monte Carlo mass production campaign.

## 220 7.3 Library generation

## 221 7.4 Use of Frozen Showers in MC production

## 222 7.5 Tuning of Frozen Showers libraries

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

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

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

237 Machine Learning is a set of algorithms, what allows computers to learn and give a predictions  
 238 without being specifically programmed. This is a modern field of computer science, that is wildly

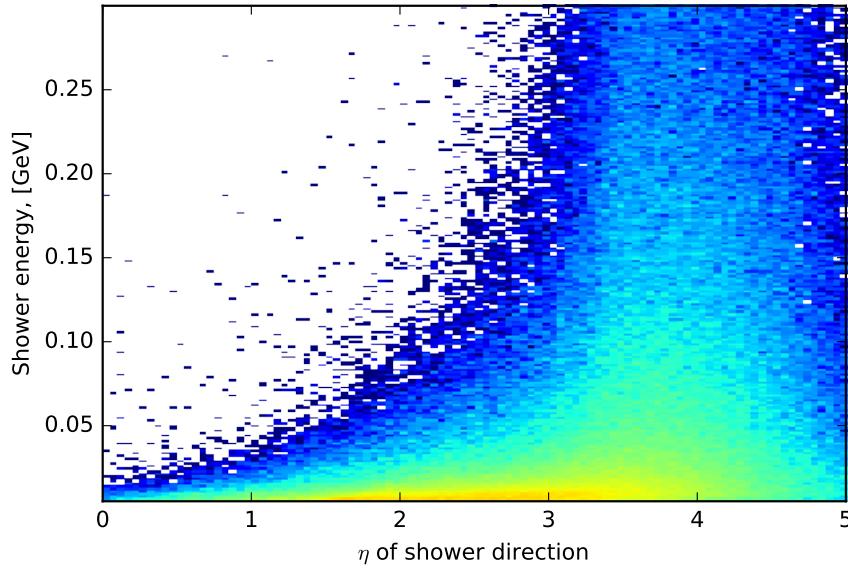


Fig. 7.8: Distribution of showers used in production of 1000 GeV electrons on shower energy vs  $\eta_{momentum}$  plane.

used in a different fields like computer vision, natural language processing, data science etc. There are two main types of machine learning algorithms: supervised, where example of desired output is given by the "teacher" and the goal is to learn a general rule, that maps inputs to outputs and unsupervised learning, then there are no labels given to algorithm, and algorithms is discovering hidden patterns in data. Initial data parameters of interest, that are used in algorithm to learn are called features. It is important to have right proper set of features and good training sample.

From a geometrical point of view, one of the main parameter is a direction of the shower. Eta momentum distribution is showed on a Figure 7.8 . Most of the showers are collinear to an electron direction. Because of this it was decided to use as a training sample simulation results for electrons with energies less than 1 GeV and momentum uniformly distributed between eta 3.0 and 4.0. This allowing to study equally low and high energy showers equally.

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

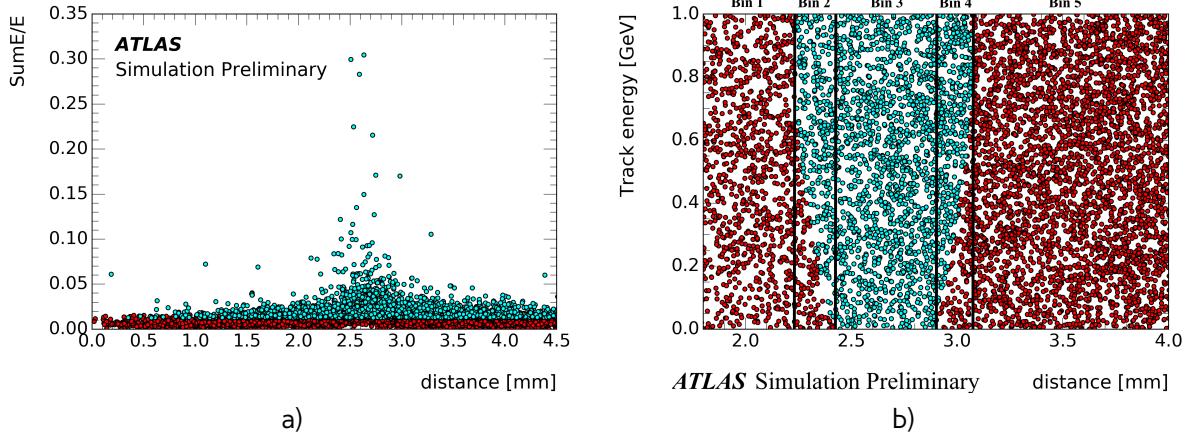


Fig. 7.9: Results of machine learning for a) first classifier b) second classifier. Cyan dots are corresponding to sensitive material showers, red - dead material showers

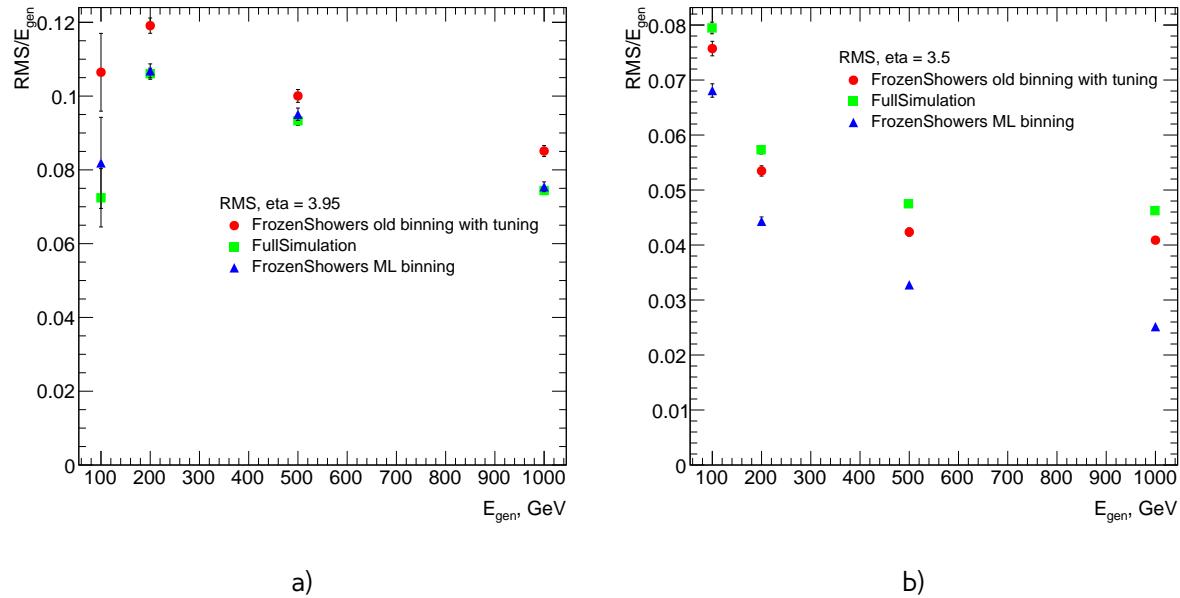


Fig. 7.10: Resolution of reconstructed electrons for full simulation, new libraries with ML binning and old tuned libraries with original binning for a)  $\eta = 3.95$  b)  $\eta = 3.5$

for a new binning, but it is performing still better, than original binning without tuning.

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

## 268 7.6 Validation of the new libraries

## 269 7.7 Plans for a future



270

Chapter

# 8

## 271 Data and Monte-Carlo samples

### 272 8.1 Data sample

### 273 8.2 Monte-Carlo samples



274

## **Part IV**

275

# **The Measurement**



276

# Chapter 9

## 277 Selection

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

### 283 9.1 Data quality cuts

Table 9.1: Analysis selection

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

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

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

291 **9.2 Lepton quality cuts**

292 Online selection of events is based on a single lepton trigger, depending on a flavor of analysis.  
293 For electron analysis it is required to have EF\_e15\_loose1 trigger, which records electrons with  $E_T > 7GeV$ . This trigger is also using additional "loose" isolation requirements to exclude jets, that are  
294 misidentified as electrons. In muon channel lowest single lepton trigger available used (EF\_mu10). It  
295 records events with muons  $E_T > 10GeV$ . Moreover, matching between trigger and lepton is required.

296 All of the analysis are using similar selection criteria, applied on a leptons. All of the leptons must  
297 satisfy requirement  $P_T > 20GeV$  Electron candidates are required to be within pseudorapidity range  
298  $|\eta| < 2.47$ . Candidates within the transition region between the barrel and endcap electromagnetic  
299 calorimeters,  $1.37 < |\eta| < 1.52$ , are removed. Additionally, for better multijet background rejection  
300 medium identification and PtCone20 < 0.1 criterias are applied.

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

305 **9.3 Boson selection**

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

310 The reconstructed lepton pair in case of Z boson analysis is required to invariant mass between  
311 66 and 116 GeV. Both upper and bottom limits allows to exclude regions with high background  
312 contamination and low statistics.

313 Full set of cuts is summarized in a table 9.1.

314 **9.4 Cut flow**

315 Chapter 10

316 Monte Carlo corrections

317 Monte Carlo plays important role in cross-section measurement. It is constantly undergoing correction  
318 to data, in order to obtain a required precision. Part of this corrections have been described in  
319 a chapter 6. Unfortunately, not everything can be taken into account during simulation itself. This  
320 leads to a differences between data and monte carlo, that needs to be accounted for. There are two  
321 possible methods to correct monte carlo without regenerating it. First on is to apply event weight, so  
322 what each mc event can contribute to non 1 entries in a histogram. This is called reweighting. Second  
323 one is to smear MC. It is using random number to alter reconstructed 4-vectors. This chapter de-  
324 scribes all additional corrections, what have been applied on MC in this analysis. All of this correction  
325 are introducing additional systematic error, that will be discussed in the chapter ??

326 **10.1 Lepton efficiency corrections**

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

- 328 • The reconstruction efficiency  $\epsilon_{rec}$  is a probability to reconstruct lepton as a lepton of this  
329 flavor.
- 330 • The identification efficiency  $\epsilon_{id|rec}$  is the probability that a reconstructed lepton survives iden-  
331 tification requirements.
- 332 • 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:

$$\epsilon_{total} = \epsilon_{rec} \times \epsilon_{id|rec} \times \epsilon_{trig|rec,id} \quad (10.1)$$

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

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

$$\epsilon_{reco,muon} = \epsilon_{reco,muon|ID} \cdot \epsilon_{ID} \approx \epsilon_{reco,muon|ID} \cdot \epsilon_{ID|MS}, \quad (10.2)$$

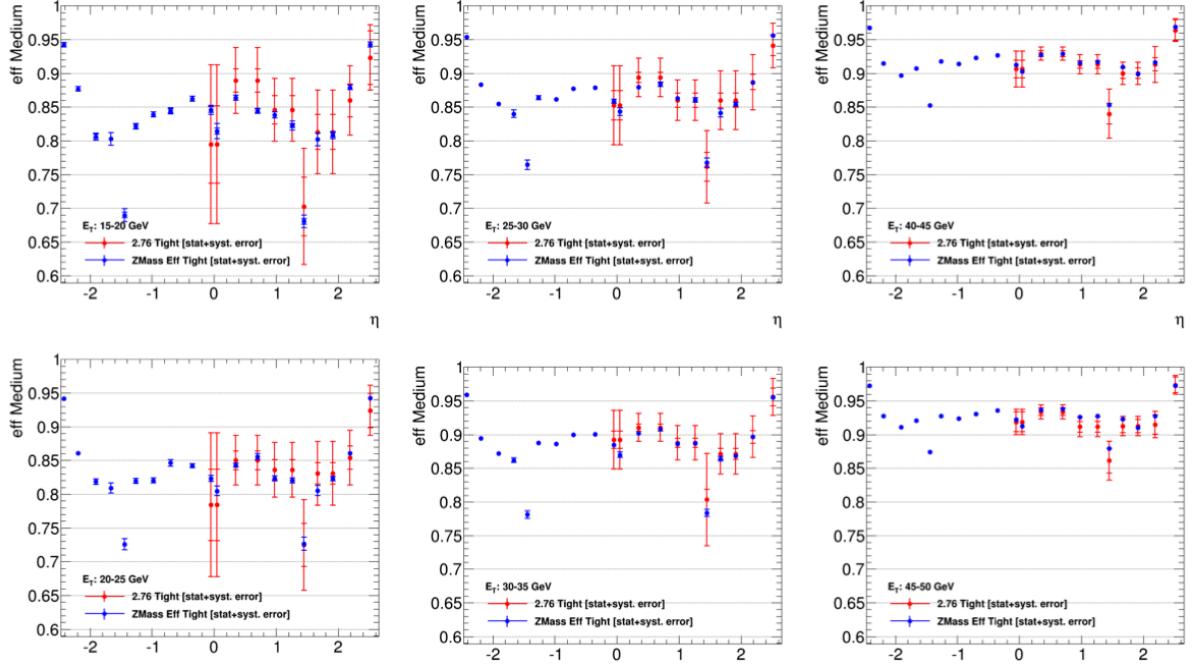


Fig. 10.1: Comparison of electron efficiencies as calculated for 8TeV (blue points) and 2.76TeV (red points) for MC simulation. Efficiencies are shown as a function of pseudorapidity ( $\eta$ ) for different electron  $E_T$  bins. Both statistical and systematic uncertainties are shown.

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

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

$$SF_{reco,id,trig} = \frac{\epsilon_{reco,id,trig}^{data}}{\epsilon_{reco,id,trig}^{MC}} \quad (10.3)$$

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

346 It is possible to use scale factors for 8 TeV 2012 data. The main difference between this data  
 347 samples are center of mass energy and a pile-up conditions (10 in 2012 and less than 1 in 2013).  
 348 This effects have been studied on a  $Z \rightarrow ee$  sample. Fig. 10.1 shows that all of the differences in a  
 349 scale factors are negligible and fully covered by the statistical error. This justifies the usage of 8 TeV  
 350 scaling factors with increased

### 351 10.1.1 Muon Trigger SF

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

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

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

360 Trigger efficiencies for data and MC in  $\eta$  bins are shown on a Fig. 11.14. Total scale factors are  
 361 presented in a Tab. 10.1. Scale factors for  $\mu^+$  and  $\mu^-$  are more, than  $3\sigma$  away from each other, that  
 362 is a clear sign of a charge dependency.

363 Effect of applying scale factors on

Table 10.1: Muon trigger scale factors

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

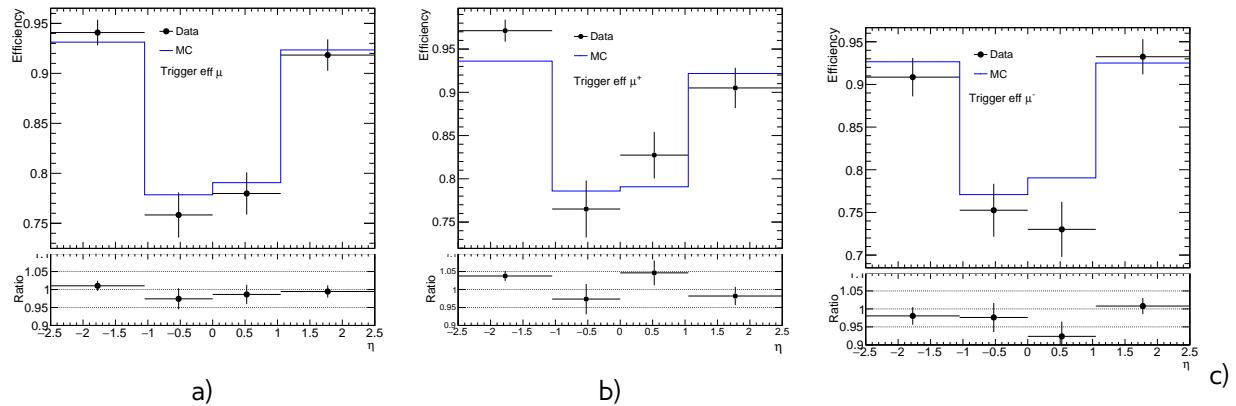


Fig. 10.2: Trigger scale factors for a)  $\mu$  b)  $\mu^+$  c)  $\mu^-$

## 364 10.2 Electron energy scale and resolution

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

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

Energy shift is parameterised, as:

$$E^{data} = E^{MC}(1 + \alpha_i), \quad (10.4)$$

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

$$m_{i,j}^{data} = m_{i,j}^{MC}(1 + \alpha_{i,j}), \quad \alpha_{i,j} \sim \frac{\alpha_i + \alpha_j}{2} \quad (10.5)$$

<sup>372</sup> neglecting second order terms.  $m_{i,j}^{data}$  and  $m_{i,j}^{MC}$  are reconstructed mass of Z boson in a  $i$  and  $j$  bins  
<sup>373</sup> of  $\eta$  for data and MC respectively.

There is also a need to correct difference in a electron resolution. It can be described by a formula 7.1. It is assumed, that sampling and noise terms are modeled well by MC and the main difference is coming from a constant term. So, the electron resolution correction then can be written as:

$$\frac{\sigma_E}{E_i}^{Data} = \frac{\sigma_E}{E_i}^{MC} \oplus c_i \quad (10.6)$$

<sup>374</sup> where  $c_i$  is  $\eta$  dependent relative resolution correction. Similarly to a energy scale correction it is  
<sup>375</sup> possible to derive resolution correction factor by a comparing  $m_{i,j}^{data}$  and  $m_{i,j}^{MC}$  distribution.

<sup>376</sup> Correction values of  $\alpha_i$  and  $c_i$  are obtained via  $\chi^2$  fit on a invariant mass electrons for data and  
<sup>377</sup> MC. Resulting energy scale is applied on a data, while resolution is corrected for MC. The resulting  
<sup>378</sup> scale is validated on a  $J/\psi \rightarrow ee$  and  $Z \rightarrow ee\gamma$

### <sup>379</sup> 10.3 Muon momentum correction

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

$$\frac{\sigma_{Det}(p_T)}{p_T} = \frac{r_0^{Det}(\eta, \phi)}{p_T} \oplus r_1^{Det}(\eta, \phi) \oplus r_2^{Det}(\eta, \phi) \cdot p_T \quad (10.7)$$

<sup>380</sup> The first term origins from fluctuations of energy loss in transversed material. Second  $r_1^{Det}$  is com-  
<sup>381</sup> ing from magnetic field inhomogeneities and local displacements. Third term  $r_2^{Det}$  describes intrinsic  
<sup>382</sup> resolution effects.

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

$$p_T^{data} = p_T^{MC} + s_0^{Det}(\eta, \phi) + s_1^{Det}(\eta, \phi) \cdot p_T^{MC}, \quad (10.8)$$

<sup>383</sup> where  $s_0^{Det}(\eta, \phi)$  is coming from the imperfect knowledge of energy losses for muons passing through  
<sup>384</sup> detector.

This leads to a total correction formula:

$$p_T^{Cor,Det} = \frac{p_T^{MC,Det} + \sum_{n=0}^1 s_n^{Det}(\eta, \phi)(p_T^{MC,Det})^n}{1 + \sum_{m=0}^2 \Delta r_m^{Det}(\eta, \phi)(p_T^{MC,Det})^{m-1} g_m}, \quad (10.9)$$

385 where  $g_m$  are normally distributed random variables with mean 0 and width 1. Because small amount  
 386 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  
 387 effect for an MS is corrected on a simulation level by adding a random smearing to an alignment  
 388 constants. This is allowing to set  $\Delta r_2^{MS}(\eta, \phi)$  to 0 during a fit.

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

$$\rho = \frac{p_T^{MS} - p_T^{ID}}{p_T^{ID}}, \quad (10.10)$$

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

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

$$p_T^{Cor,CB} = f \cdot p_T^{Cor,ID} + (1 - f) \cdot p_T^{Cor,MS}, \quad (10.11)$$

390 where weight  $f$  is derived from mc.



391

# Chapter 11

## Missing Transverse Energy reconstruction and correction

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

### 11.1 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:

$$E_{x(y)}^{miss} = E_{x(y)}^{miss,calo} + E_{x(y)}^{miss,cryo} + E_{x(y)}^{miss,muon}. \quad (11.1)$$

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

$$E_{x(y)}^{miss} = E_{x(y)}^{miss,e} + E_{x(y)}^{miss,\gamma} + E_{x(y)}^{miss,\tau} + E_{x(y)}^{miss,jets} + E_{x(y)}^{miss,SoftTerm} + E_{x(y)}^{miss,\mu}. \quad (11.2)$$

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

406 This procedure was optimised for 8 TeV runs and using a calibration constants from it. This can  
 407 cause problems with 2.76 TeV low pileup run. As a examine showed this is not optimal procedure  
 408 in this case. Control plots for W production in electron and muon channels are shown on a Fig. 11.1.  
 409 Where a big discrepancies in a muon and electron channel, that cannot be accounted to multijet  
 410 background.

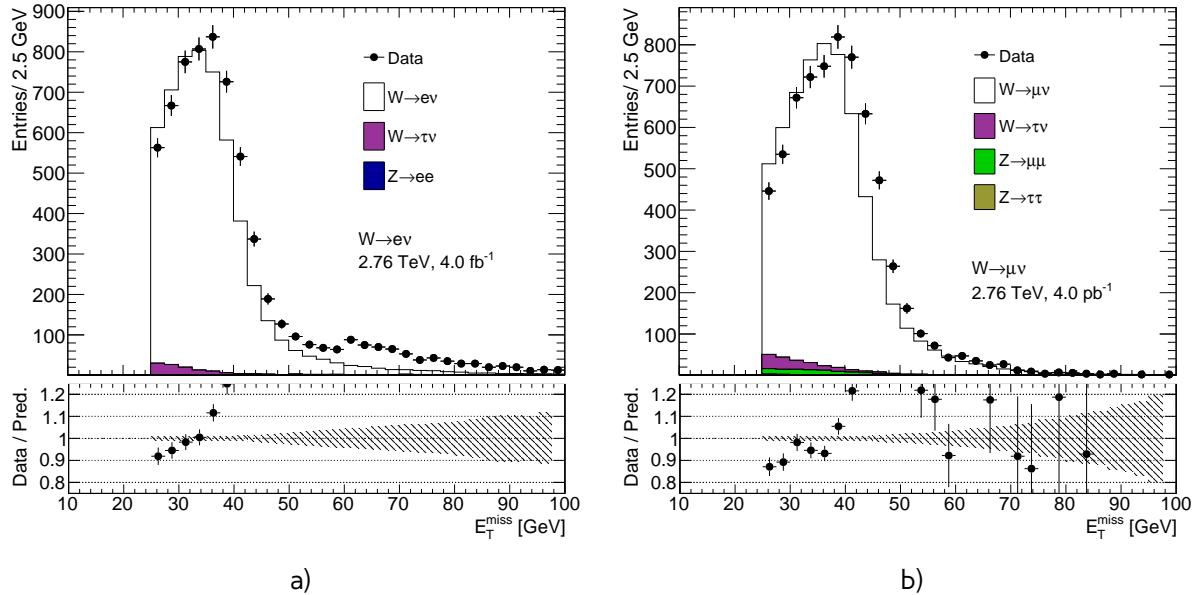


Fig. 11.1: Data and MC comparison for  $E_T^{\text{miss}}$  calculated by standard ATLAS algorithm for a)  $W \rightarrow e\nu$  b)  $W \rightarrow \mu\nu$  events

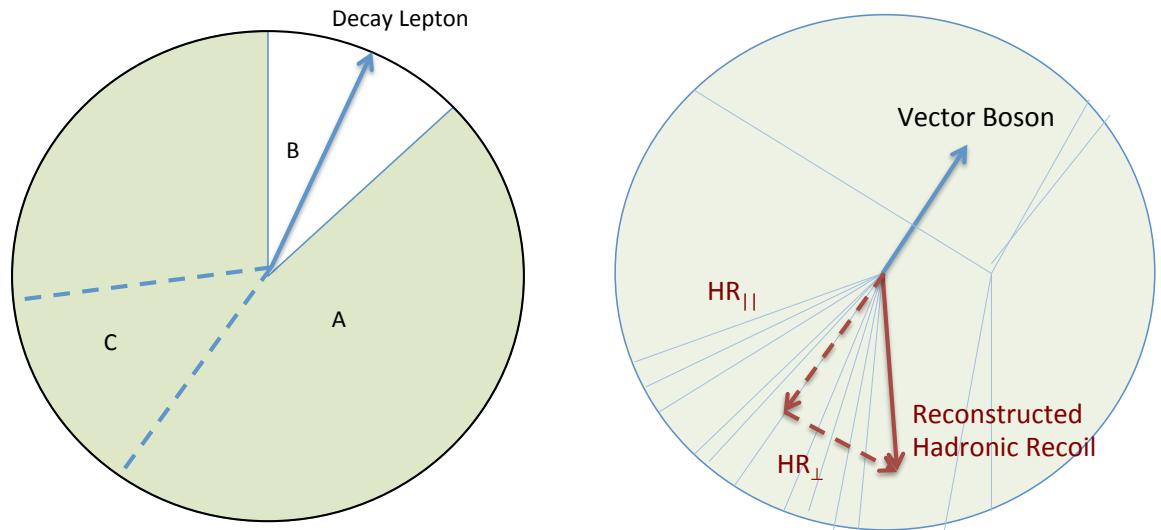


Fig. 11.2: a) Definition of different zones in the calculation of the cluster-based hadronic recoil. , b) Parallel and perpendicular projection of the hadronic recoil with the respect to the transverse momentum of the vector boson

411 **11.2 Reconstruction of Missing Transverse Energy from hadron  
412 recoil**

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

$$\vec{P}_T^W = \vec{P}_T^{lep} + \vec{P}_T^\nu = \sum \vec{P}_T^{ISRquarks,gluon}, \quad (11.3)$$

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

$$E_T^{miss} = P_T^\nu = -HR + p_T^l \quad (11.4)$$

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

$$HR = \sum_{i=0}^{N_{topo}} \vec{p}_T^{topo} \quad (11.5)$$

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

$$\sum E_T = \sum_{i=0}^{N_{topo}} E_T^{topo} \quad (11.6)$$

413 To avoid double counting of lepton energy losses in calorimeter, the clusters inside cone with radius  
414  $dR = 0.2$  are excluded from this calculation. To compensate soft activity inside this cone, clusters  
415 are then compensated by replacement cone (Fig. 11.2). This cone is defined as cone at the same  
416 pseudorapidity, but different  $\phi$ . It should be far from any other lepton and hadron recoil direction.  
417 Each cone is then rotated to a direction of the original lepton direction. This definition is not taking  
418 into account jet reconstruction aspects. This is allowing to get a better data MC agreement (Fig. 11.3).

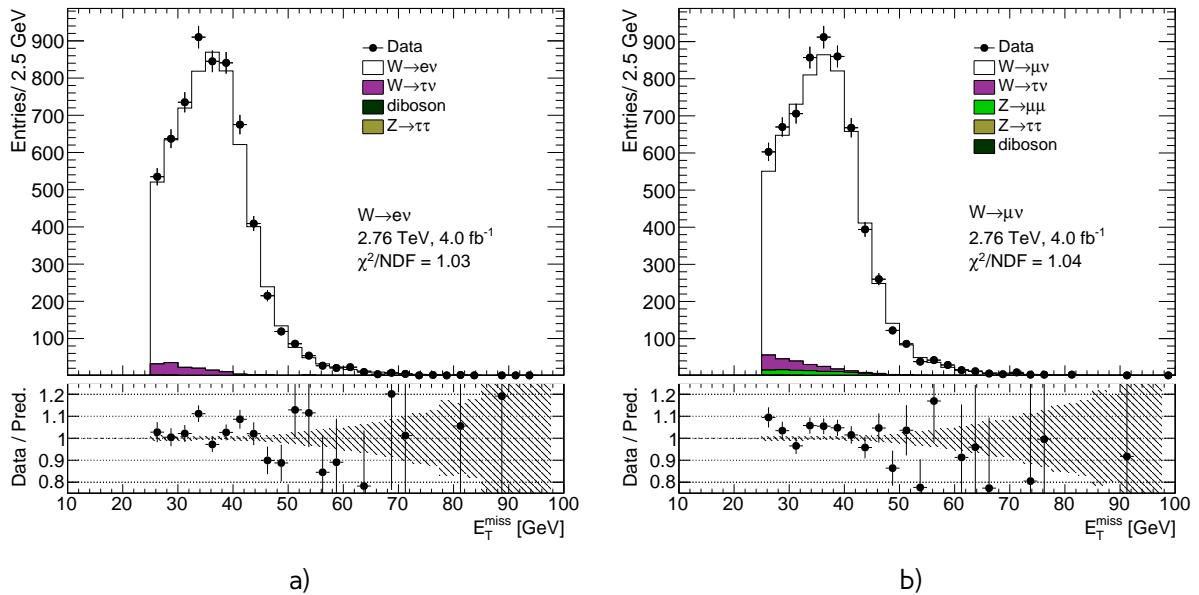


Fig. 11.3: Data and MC comparison for  $E_T^{\text{miss}}$  calculated from hadron recoil for a)  $W \rightarrow e\nu$  b)  $W \rightarrow \mu\nu$  events

### 11.3 Hadron Recoil calibration

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

The hadron recoil algorithm performance can be studied in MC through the projection of  $\vec{H}\vec{R}$  on the direction of the transverse momentum of the vector boson, as shown on Fig. 11.2. This projection can be divided into perpendicular  $u_{\perp}$  and a parallel  $u_{\parallel}$  component as follows:

$$u_{\parallel} = \vec{v}_{xy} \cdot \vec{H}\vec{R} \quad (11.7)$$

$$u_{\perp} = v_x \cdot H R_y - v_y \cdot H R_x, \quad (11.8)$$

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

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

First step in a hadron recoil calibration procedure ames to correct differences in a pile-up modelling

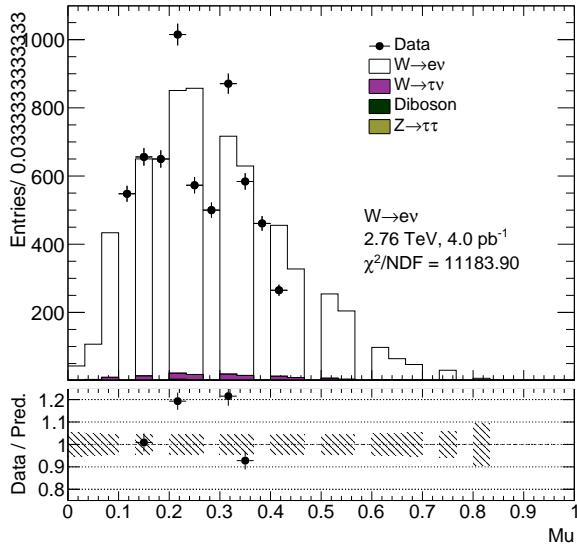


Fig. 11.4: Pileup

in the event. Additional interaction can have a significant effect on a  $E_T^{miss}$  and  $\sum E_T$  distributions. It is usually accounted scaling average number of interaction per bunch crossing to match a data. However, ATLAS simulation is suited for an high pile-up runs, so this quantity is modelled discretely in case of 2.76 TeV analysis (Fig. 11.4), what makes the corrections to match data impossible.

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

#### 11.3.1 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 differences between data and monte carlo  $E_T^{miss}$  resolution (Fig. 11.5). 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.

#### Sumet distribution correction

Distribution of  $\sum E_T$  events are shown on a Fig. 11.6. There is a clear sign of shift in this distribution in 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:

$$SF^{channel} = \frac{\sum E_T^{data,selection}}{\sum E_T^{MC,no\ cuts}}, \quad (11.9)$$

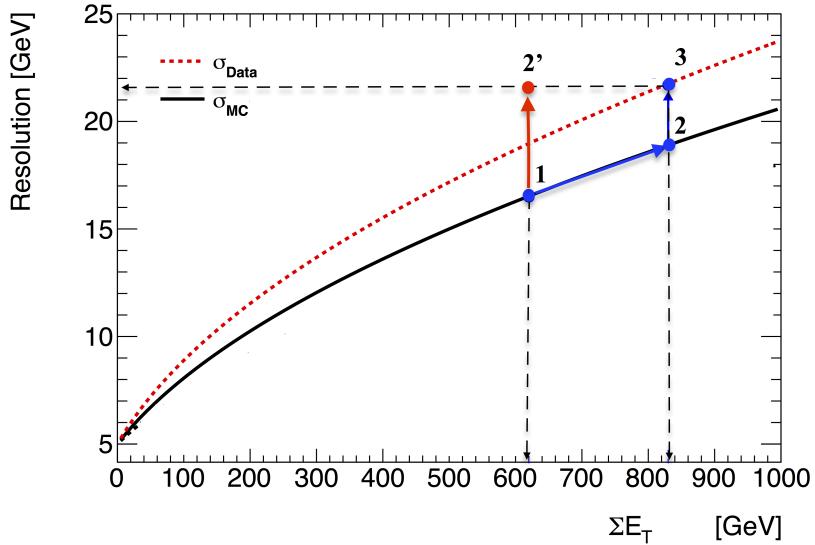


Fig. 11.5: Schematic view of the correction procedure: this figure illustrates the resolution of  $u_{\perp}$  as a function of  $\sum E_T$ . The dotted curve represents data resolution ( $\sigma_{data}$ ), solid black is a nominal MC ( $\sigma_{MC}$ ). Blue line from point 1 to point 2 corresponds to a  $\sum E_T$  correction. Red line from point 1 to point 2' corresponds to a direct correction of resolution mismodelling

453 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  
 454 of the small data statistics combination of  $W \rightarrow e\nu$  and  $W \rightarrow \mu\nu$  processes is used. Oppositely, in  
 455 MC  $\sum E_T^{MC, no cuts}$  is taken without any cuts. Scale factors are determined separately for each signal  
 456 process for a W boson decays. Example of correction factors for a different  $p_T^{W, rec}$  bins are shown on  
 457 a Fig.11.8 Ratio inside each bin can be parameterised by a polynomial degree 2 inside each  $p_T^{W, rec}$  bin.  
 458 Total SF obtained by this procedure are shown on a Fig. 11.9. The distribution of  $\sum E_T$  after correction  
 459 is shown on a Fig. 11.11. Reconstructed boson pt spectrum is leaving almost untouched, while this  
 460 procedure still introduces some shift in a truth boson pt spectrum(Fig. 11.10 ). Effect on the resolution  
 461 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:

$$fit\ parameters_{new} = fit\ parameters + gaus^{2D}(cov.matrix), \quad (11.10)$$

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

466 Systematical error can be studied by applying lower order of approximation on a SF or not applying it  
 467 at all. The overall effect on a  $C_W$  for a different methods is shown in a Tab. 11.1. Results are dominated  
 468 by a systematics error. However, there is a difference in a sign of the effect for a different flavors  
 469 of the analysis. This cannot be explained from a physical point of view, so it was decided not to use  
 470 this corrections in a final analysis.

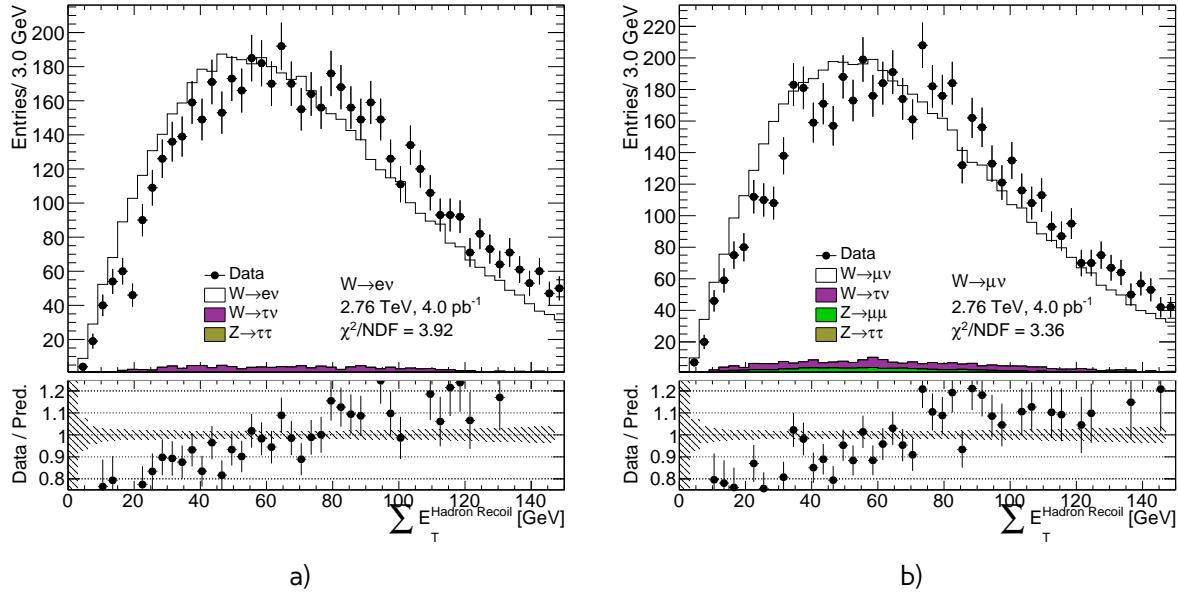


Fig. 11.6: Distribution of  $\sum E_T$  from a)  $W \rightarrow e\nu$  and b)  $W \rightarrow \mu\nu$  events before correction

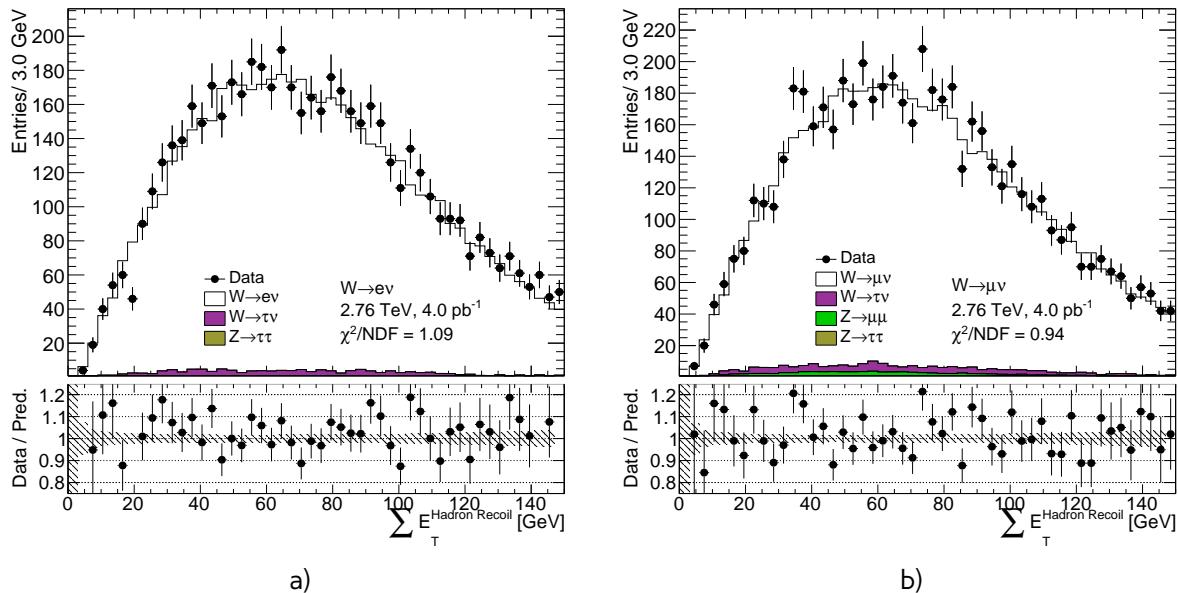


Fig. 11.7: Distribution of  $\sum E_T$  from a)  $W \rightarrow e\nu$  and b)  $W \rightarrow \mu\nu$  events after correction

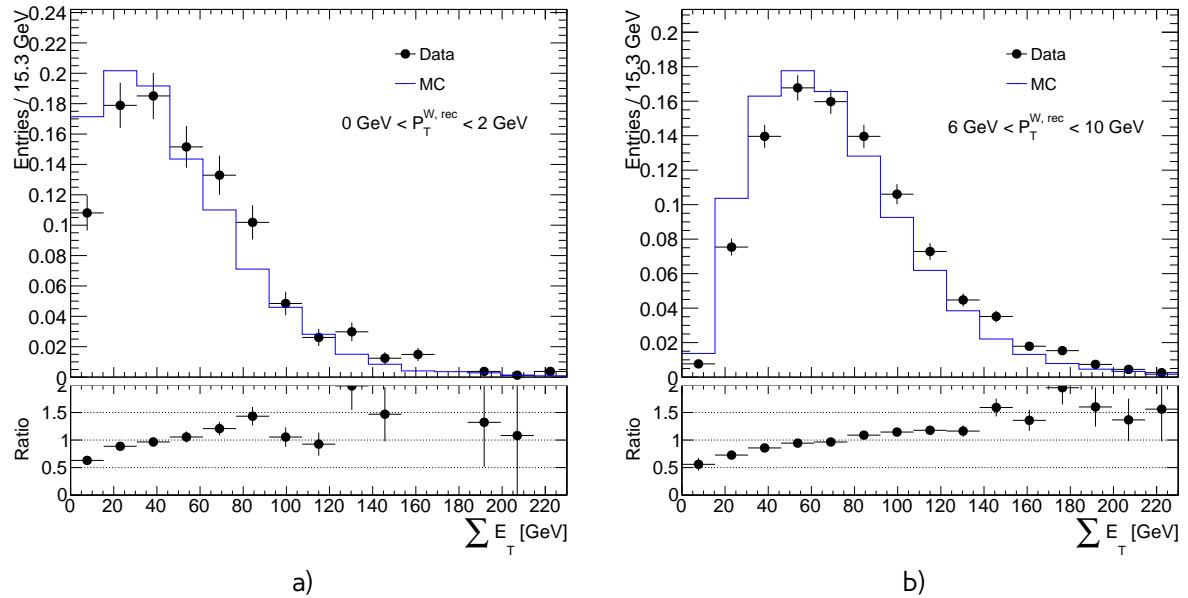


Fig. 11.8: Distribution of  $\sum E_T$  for different  $p_T^{W,\text{rec}}$  bins

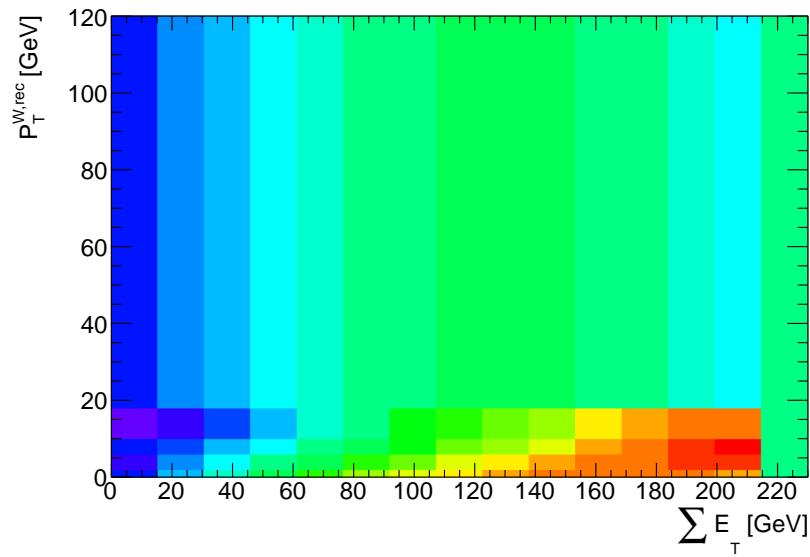


Fig. 11.9: Correction factors for  $W \rightarrow e\nu$

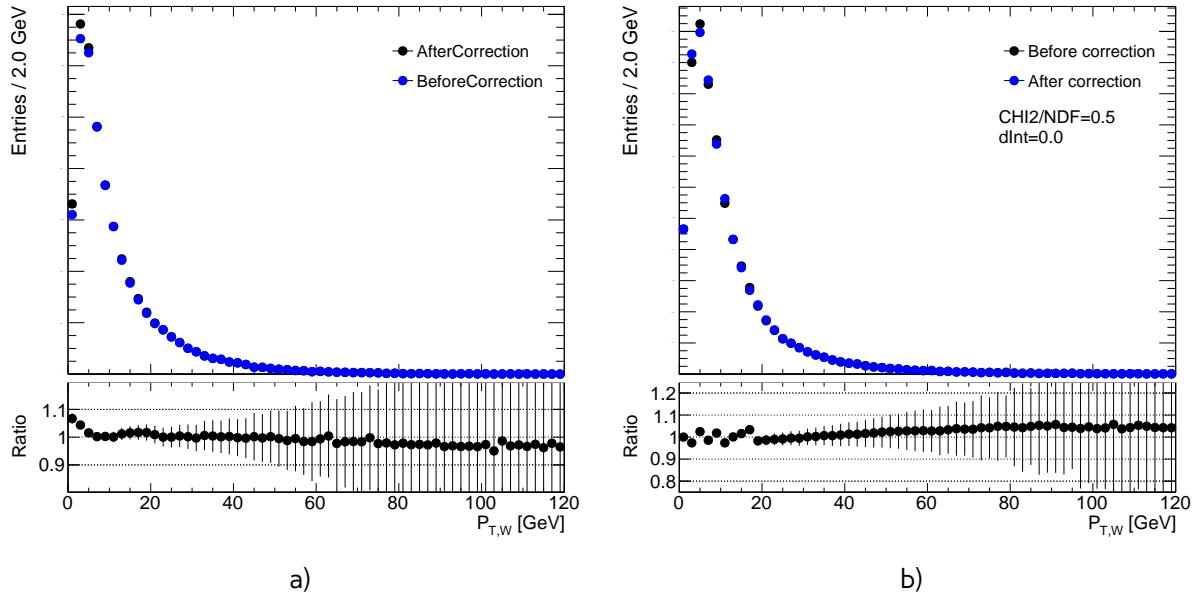


Fig. 11.10

Table 11.1: 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%                  |

Table 11.2

| 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 |

### 471 Resolution correction using Z events

472 Another way to check resolution effects is to use  $u_{\perp}$  and  $u_{\parallel} - p_T^Z$  distributions in a Z events. This  
 473 correction assumes, that any resolution mismodelling reflects discrepancies in the  $\sum E_T$  distribution,  
 474 while difference in resolution at a given  $\sum E_T$  is a subleading. There is a clear difference in a rms  
 475 of this distributions between data and MC, that cannot be accounted as a statistical error in data.  
 476 Difference in resolutions is consistent for  $u_{\perp}$  and  $u_{\parallel} - p_T^Z$  distributions, but depends on a flavor of  
 477 the analysis. The resolution is corrected by smearing with a Gaussian distribution each component  
 478 of a hadron recoil:

$$u'_{\parallel} = u_{\parallel} + \text{Gaus}(0, d\sigma) \quad (11.11)$$

$$u'_{\perp} = u_{\perp} + \text{Gaus}(0, d\sigma), \quad (11.12)$$

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

$$d\sigma = \sqrt{\sigma_{data}^2 - \sigma_{MC}^2} \quad (11.13)$$

479 Systematic error of this  $d\sigma$  is taken as an statistical error for  $\sigma_{data}$ . Overall effect on a  $C_W$  depending  
 480 on a  $d\sigma$  is shown on a Fig. 11.12. Due to a random nature of this correction, effect is not stable for  
 481 a small  $d\sigma$ . Stability of this correction can be tested by repeating this procedure several times with  
 482 different random seed number. Due to a not stable nature of this correction the overall systematics  
 483 coming from resolution mismodelling is assumed to be 0.2% for each W channel.

### 484 11.3.2 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:

$$u_{\parallel}^{MC,cor} = u_{\parallel}^{MC} \cdot SF_{HR,bias}, \quad (11.14)$$

485 and can be obtained by scanning the impact of the scaling factor on the Data to MC agreement of  
 486 the distributions that are dominated by the recoil scale uncertainties. Since W boson has no second  
 487 source of  $P_T^W$  measurments, determination of the hadron recoil bias should use the distributions, that  
 488 are not sensitive to a truth  $P_T^W$  spectrum. One of the optimal choises is a  $M_T^W$  distribution. Transverse  
 489 mass distribution for a different scale choises is shown on a Fig. 11.13. Multijet background is not  
 490 included, because it shape and number of events is depending on a hadron recoil scale and thus  
 491 can introduce additional systematics.

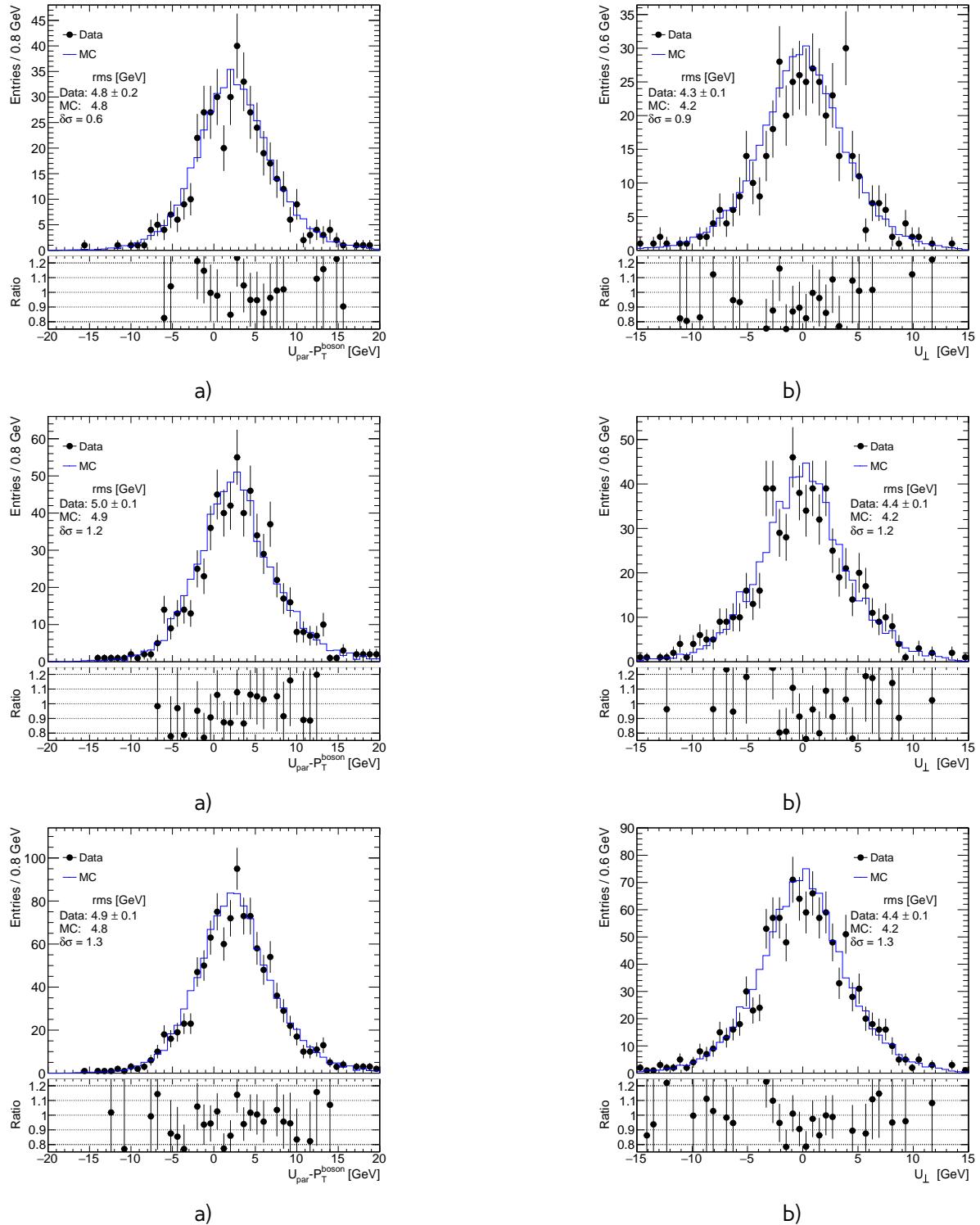


Fig. 11.11

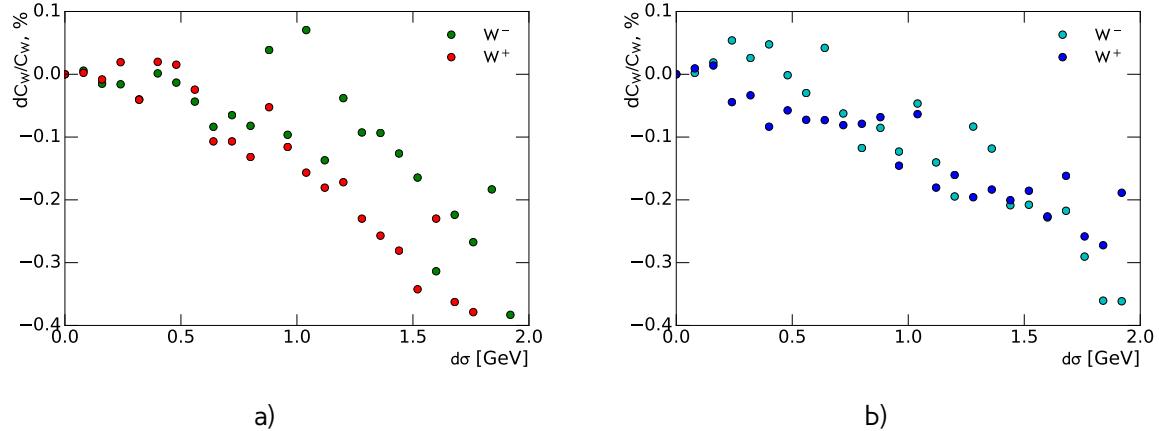


Fig. 11.12: Effect on a  $C_W$  for a different  $d\sigma$  for a)  $W \rightarrow e\nu$  b)  $W \rightarrow \mu\nu$  channel

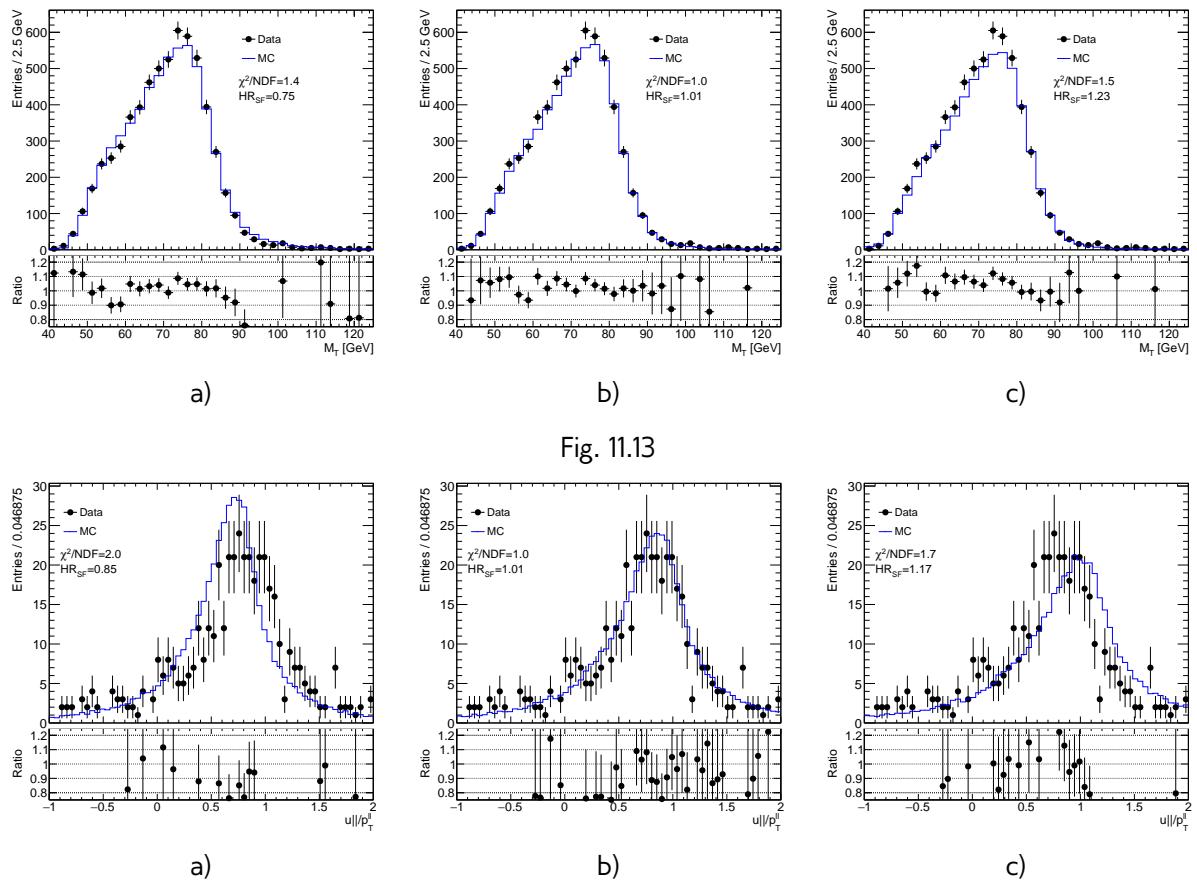


Fig. 11.14

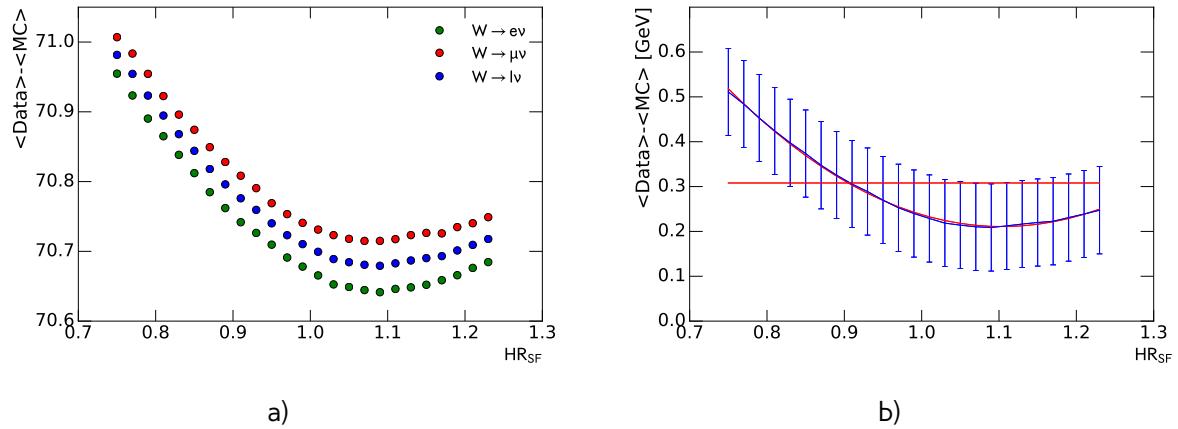


Fig. 11.15

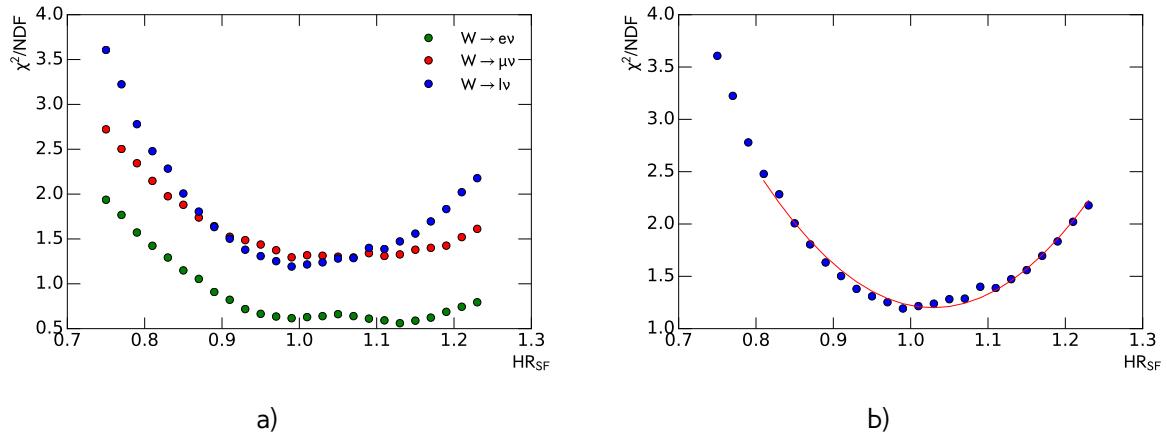


Fig. 11.16

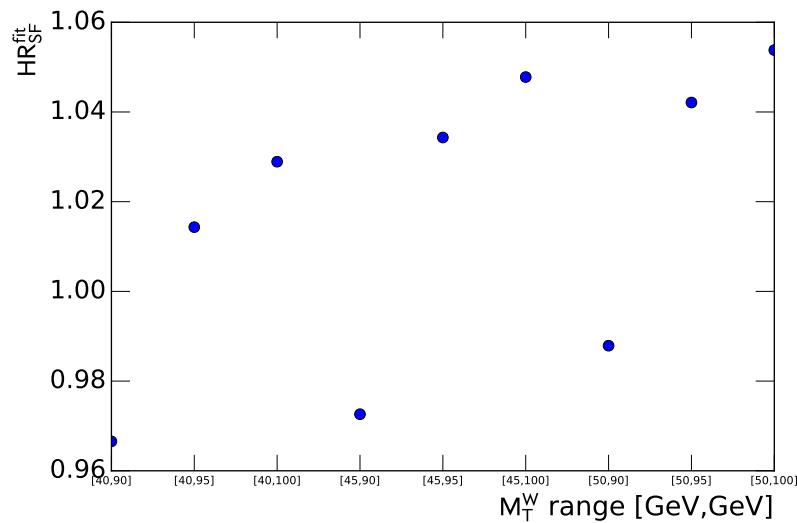


Fig. 11.17

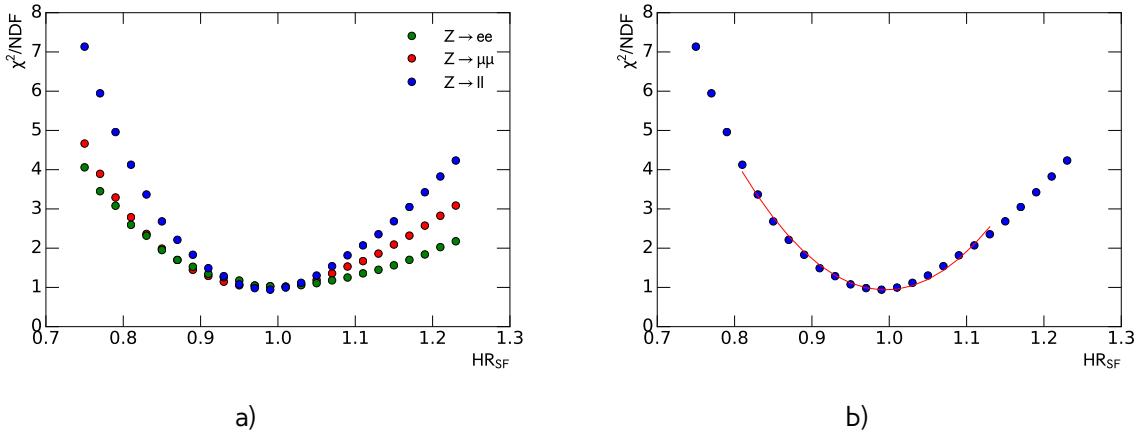


Fig. 11.18: Effect on a  $C_W$  for a different  $d\sigma$  for a)  $W \rightarrow e\nu$  b) $W \rightarrow \mu\nu$  channel

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

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

$$\chi^2 = \frac{(x - sf_{best})^2}{\sigma_{sf}^2} + \chi_0^2, \quad (11.15)$$

495 where  $sf_{best}$  is the best scale factor and  $\sigma_{sf}$  is a statistical error of this parameter. Distribution of  $\chi^2$   
 496 and a fit in combined W channel is shown on a Fig. 11.16.

497 Because of the possible mismodelling of the tail  $M_T^W$  distribution it is not included in a  $\chi^2$  cal-  
 498 culation, leaving a free choice of the parameter of the cutoff. It is also possible to exclude regions  
 499 with high multijet background contamination by applying a tighter cut on a  $M_T^W$ . This fit range is  
 500 introducing one source of systematic error. Effect of the range on value determination is shown on  
 501 a Fig. 11.17.

502 Similarly to a W channel, scale correction in a Z sample can be determined from distribution  $\frac{u_{||}}{p_T^{ll}}$ ,  
 503 shown on a Fig. 11.18. Since there is no choice of the range and dependency on  $P_T^{bos}$  modeling, there  
 504 is just one source of uncertainty.

505 Results on a hadron scale factors and its errors are shown in a Table 11.2. The results are consistent  
 506 within 1 sigma.

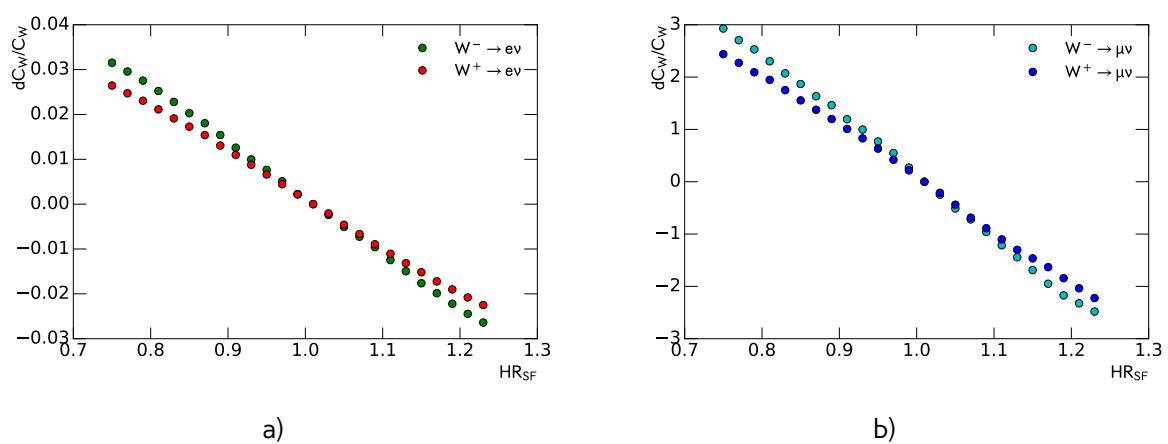


Fig. 11.19: Effect on a  $C_W$  for a different  $d\sigma$  for a)  $W \rightarrow e\nu$  b) $W \rightarrow \mu\nu$  channel



507

# Chapter 12

## 508 Background estimation

509 After the event selection described in chapter 9 the background contribution is around 4% for  
 510 W-analysis and 0.2% for Z analysis (which is with this statistics is negligible). Main backgrounds for  
 511 W analysis are coming from:

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

516 Most of the backgrounds are estimated using MC. They are normalized using highest cross-section  
 517 order available. The total list of simulated backgrounds and its cross-section is shown in a Table 12.1.  
 518 QCD background is estimated using data driven method.

Table 12.1: 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    |

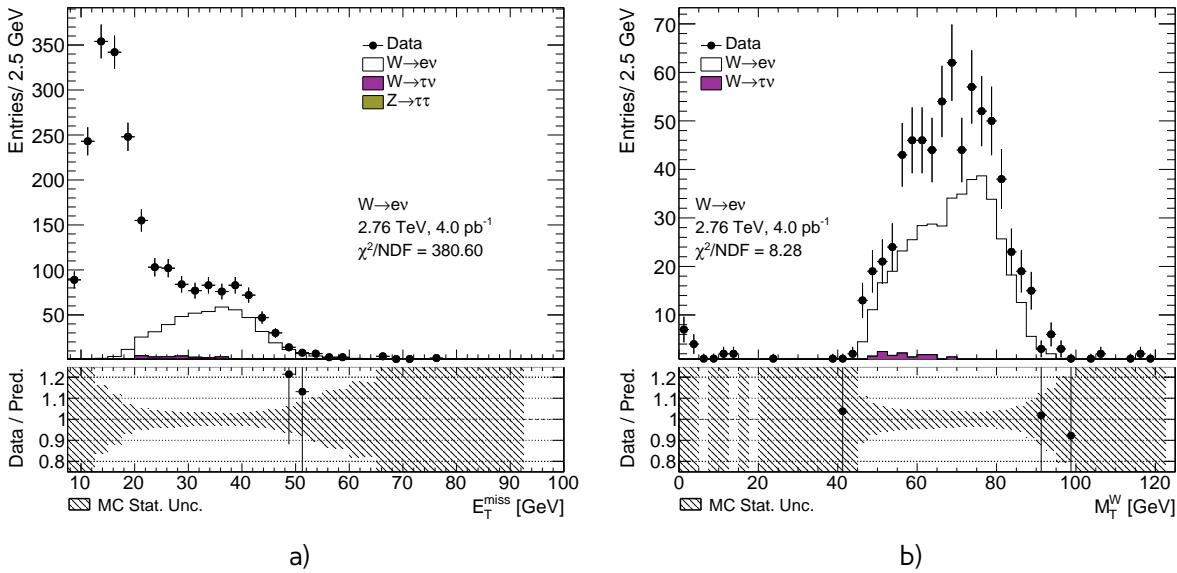


Fig. 12.1: Distribution for a)  $E_T^{\text{miss}}$  b)  $M_T^W$  template selection for  $W \rightarrow e\nu$  events

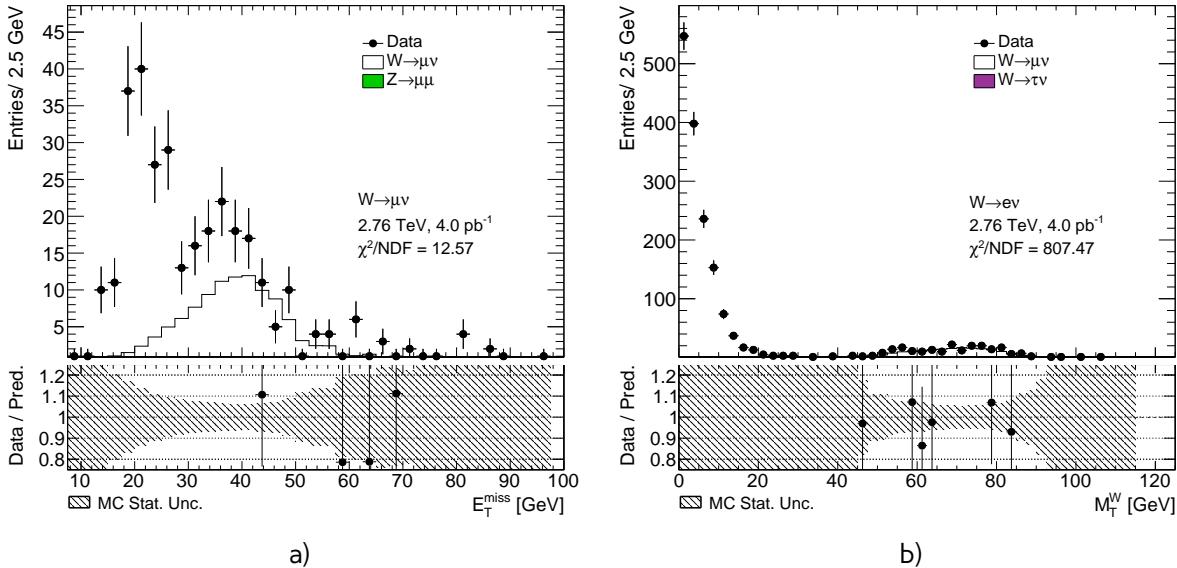


Fig. 12.2: Distribution for a)  $E_T^{\text{miss}}$  b)  $M_T^W$  template selection for  $W \rightarrow \mu\nu$  events

## 12.1 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

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

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

### 533 12.1.1 Template selection

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

542 For electron flavor, template is requiring for electron candidate to fail Medium isolation criteria,  
 543 but to pass loose selection. Control distribution for a different template selection in electron channel  
 544 are shown on a Fig. 12.1. Released  $E_T^{miss}$  cut is allowing to have a better template statistics.

545 In a muon channel template selection build by inverting isolation criteria ( $PtCone20 > 0.1$ ). In case  
 546 of  $W \rightarrow \mu\nu$  the qcd background template can be achieved by releasing  $M_T^W$  cut (Fig. 12.2).

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:

$$N_{template} = N_{data}^{bkg\ enriched} - \sum_j^{MC} N_{MC_j}^{bkg\ enriched}, \quad (12.1)$$

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

### 550 12.1.2 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:

$$M(x) = \sum_{i=1}^{N-1} f_i F_i(x) + (1 - \sum_{i=1}^{N-1} f_i) \cdot F_{qcd}(x), \quad (12.2)$$

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

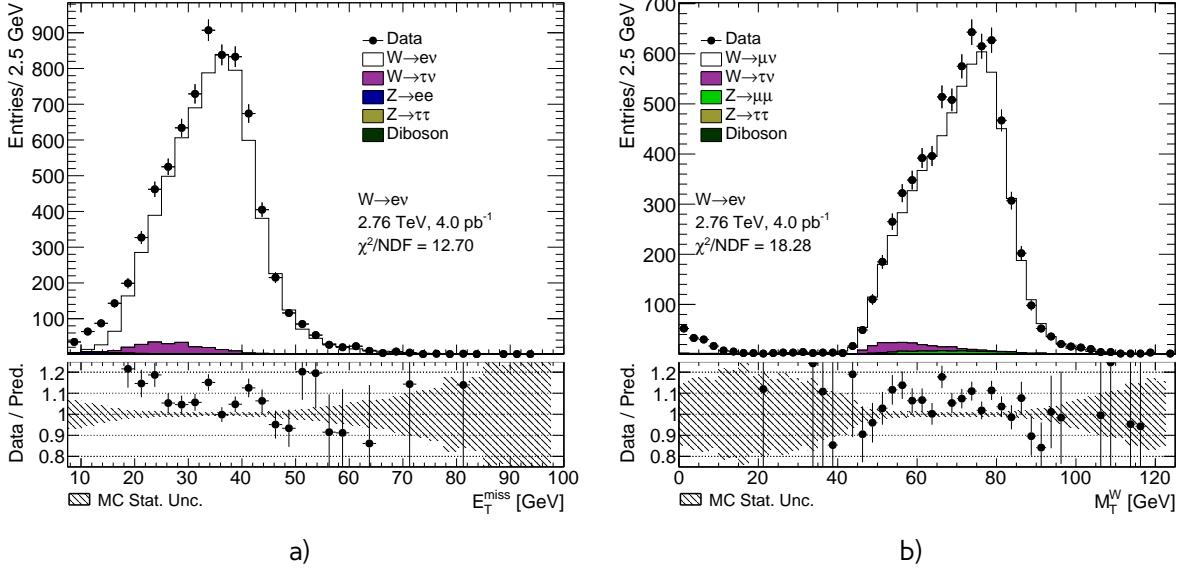


Fig. 12.3: Distributions used for multijet background estimation for a)  $W \rightarrow e\nu$  b)  $W \rightarrow \mu\nu$

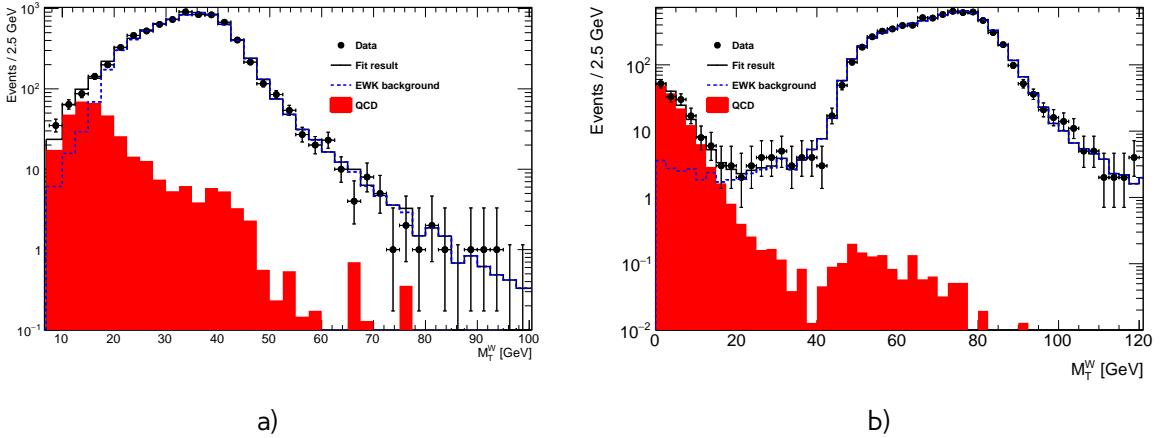


Fig. 12.4: The multijet background estimation for a)  $W \rightarrow e\nu$  using reversed ID cut and released  $E_T^{miss}$  cut b)  $W \rightarrow \mu\nu$  using released  $M_T^W$  cut and  $bb\bar{b} + cc\bar{c}$  template

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

$$scale = \frac{(1 - \sum f_i) \cdot N_{Data}^{fit}}{N_{template}}, \quad (12.3)$$

where  $\sum f_i$  is a sum of all fractions in the fit,  $N_{Data}^{fit}$  is a number of data events in a fit histogram and  $N_{template}$  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. 12.4 .

Table 12.2: 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                    |

### 560 12.1.3 Systematic Uncertainty from the Multi-jet Background Estimation

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

$$\delta_{QCD} = \sqrt{\delta_{fit\,unc}^2 + \delta_{MC}^2 + \delta_{fit\,bias}^2 + \delta_{template}^2}, \quad (12.4)$$

561 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:

$$N_{QCD}^W = 0.5 \cdot N_{QCD}^{W^+} = N_{QCD}^{W^-} \quad (12.5)$$

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

564 Third uncertainty is due to a potential bias in the template as a result of the template choise  
565 and a template statistics itself. For estimation of this uncertainty different template selections have  
566 been used. For  $W \rightarrow e\nu$  channel different reversed isolation criteria have been tried (Fig. 12.5 a)).  
567 The overall discrepancies can be considered negligible. For  $W \rightarrow \mu\nu$  channel template variations  
568 are estimated using fits with  $b\bar{b} + c\bar{c}$  MC samples. Fig. 12.5 b) compares data template with template  
569 obtained using signal selection with released  $M_T^W$  cut and template selection. Results for a different  
570 template fits are presented in a Tab 12.3

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

574 Results of QCD background uncertainty estimation for  $W \rightarrow e\nu$  and  $W \rightarrow \mu\nu$  are shown in  
575 a Tab. 12.2 and 12.3 respectively. The overall number of QCD background events is estimated as  
576  $31.00 \pm 13.0$  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  
577 overall fraction of QCD events is lower, than in 7 TeV data <reference to a 7 TeV paper>, what is  
578 agreeing with expectations.

Table 12.3: 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  |

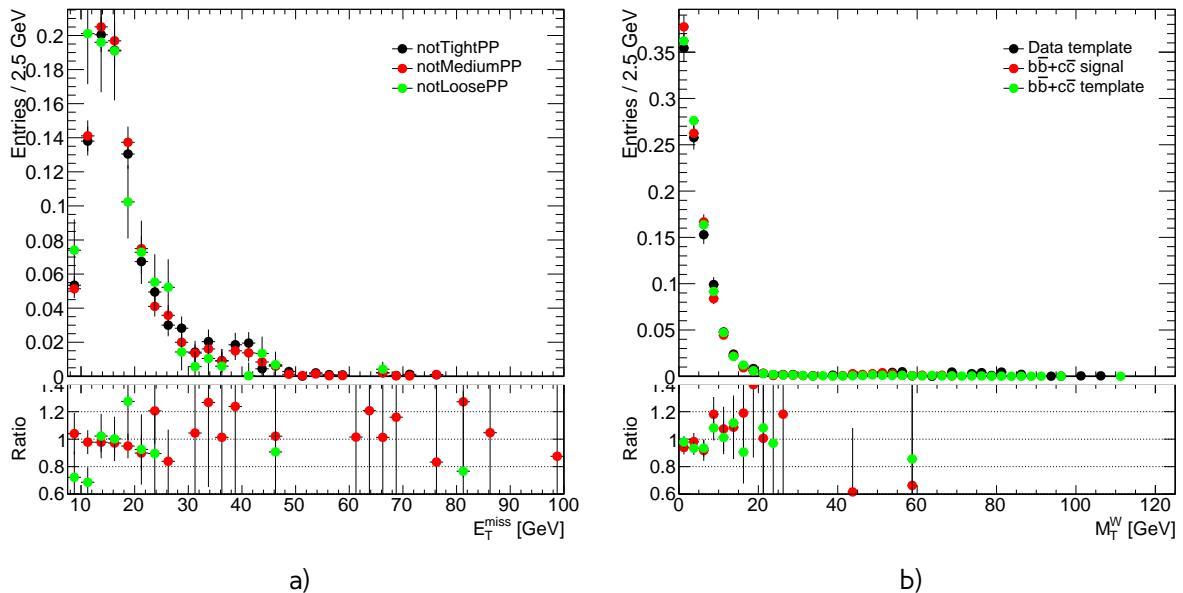


Fig. 12.5: Data and MC comparison for  $E_T^{\text{miss}}$  calculated by standard ATLAS algorithm for a)  $W \rightarrow e\nu$   
b)  $W \rightarrow \mu\nu$  events

Table 12.4: 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 12.5: 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                 | $1.5 \pm 0.0$          | $644.5 \pm 25.4 \pm 0.0$               |

## 579 12.2 Background-subtracted $W$ and $Z$ candidate events

580 Tables 12.4 and 12.5 summarize the number of background events for  $W$  and  $Z$  selections respectiv-  
 581 ely. Uncertainties on a number of EWK+top events are coming from a statistics, cross-section uncer-  
 582 tainty (if given) and 3% of luminosity determination uncertainty. For multijet background uncertainty  
 583 is coming from a method and described in a subsection 12.1.3. For the background-subtracted events  
 584 the statistical uncertainty is quoted first, followed by the total systematic uncertainty, derived from  
 585 the EW+top and multijet bacgrkound ones, considering the sources as uncorrelated.



# Chapter 13

## 587 Control distributions

588 Kinematic distribution after all cuts (Section ??) and corrections applied on MC (Section ??), are  
 589 presented in this chapter. Distributions for  $W \rightarrow l\nu$  are split in charge and shown on a Figs. 13.1-  
 590 13.12. Distributions for  $Z \rightarrow l^+l^-$  analysis are shown on a Fig. 13.13-13.17. This plots are also showing  
 591 the systematic and statistical uncertainty as a shaded band. The uncertainties are including all of  
 592 the sources, described in a ??, except for uncertainty coming from shape variation due to a PDF  
 593 reweighing and QCD background . All of uncorrelated sources are summed in a quadrature.

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

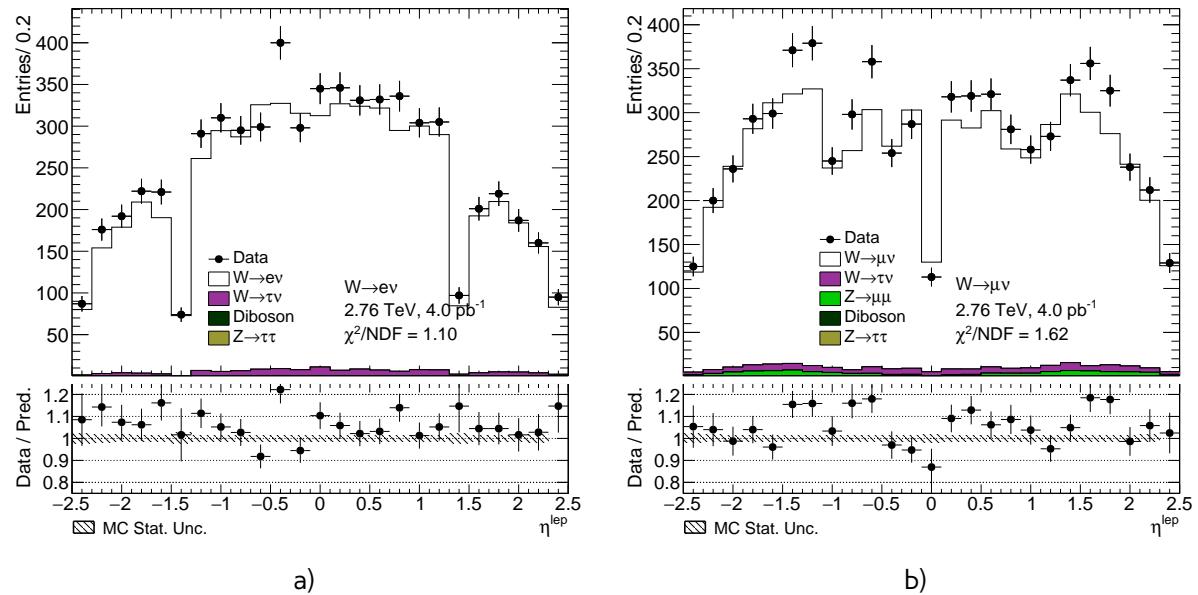


Fig. 13.1: Lepton pseudorapidity distribution from the a)  $W \rightarrow e\nu$  selection and b) the  $W \rightarrow \mu\nu$  selection.

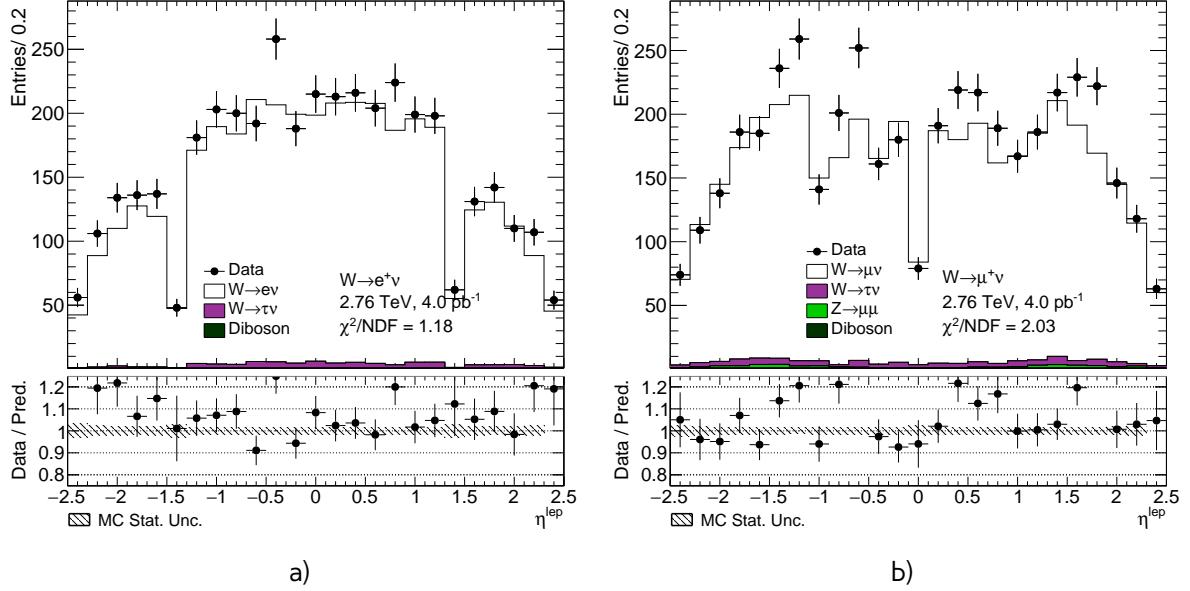


Fig. 13.2: Lepton pseudorapidity distribution from the a)  $W^+ \rightarrow e^+\nu$  selection and b) the  $W^+ \rightarrow \mu^+\nu$  selection.

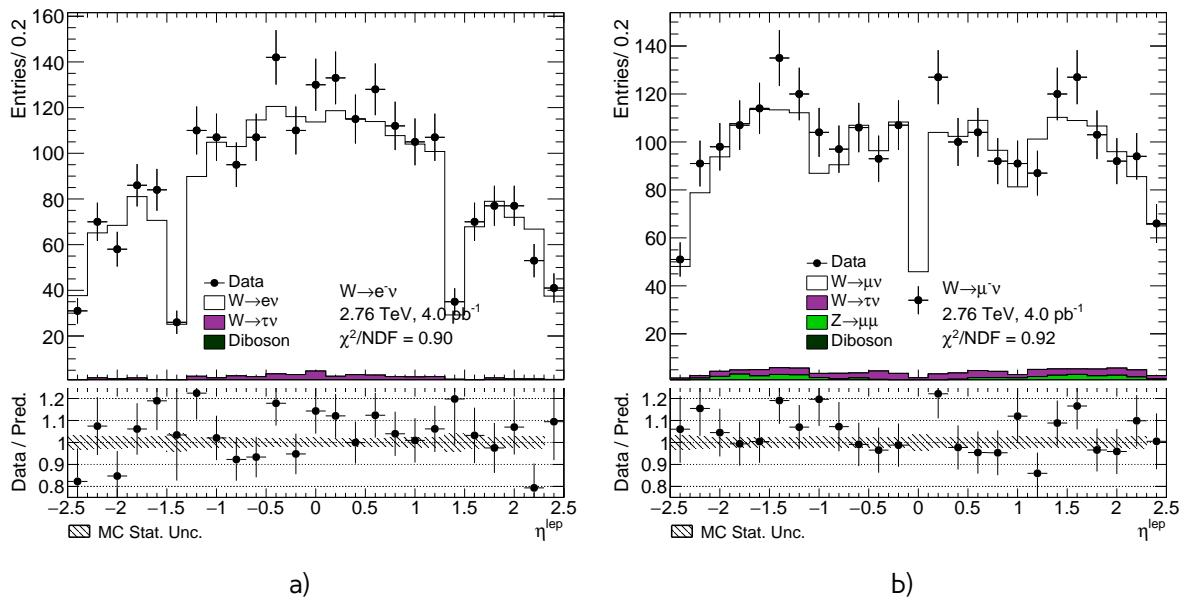


Fig. 13.3: Lepton pseudorapidity distribution from the a)  $W^- \rightarrow e^-\nu$  selection and b) the  $W^- \rightarrow \mu^-\nu$  selection.

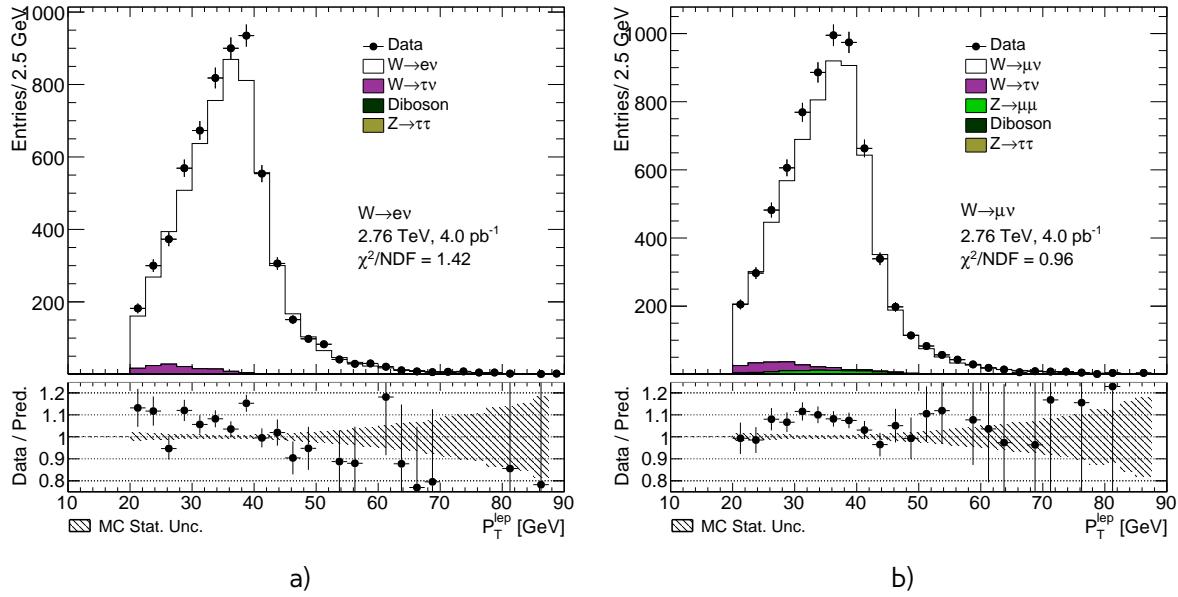


Fig. 13.4: Lepton transverse momentum distribution from the a)  $W \rightarrow e\nu$  selection and b) the  $W \rightarrow \mu\nu$  selection.

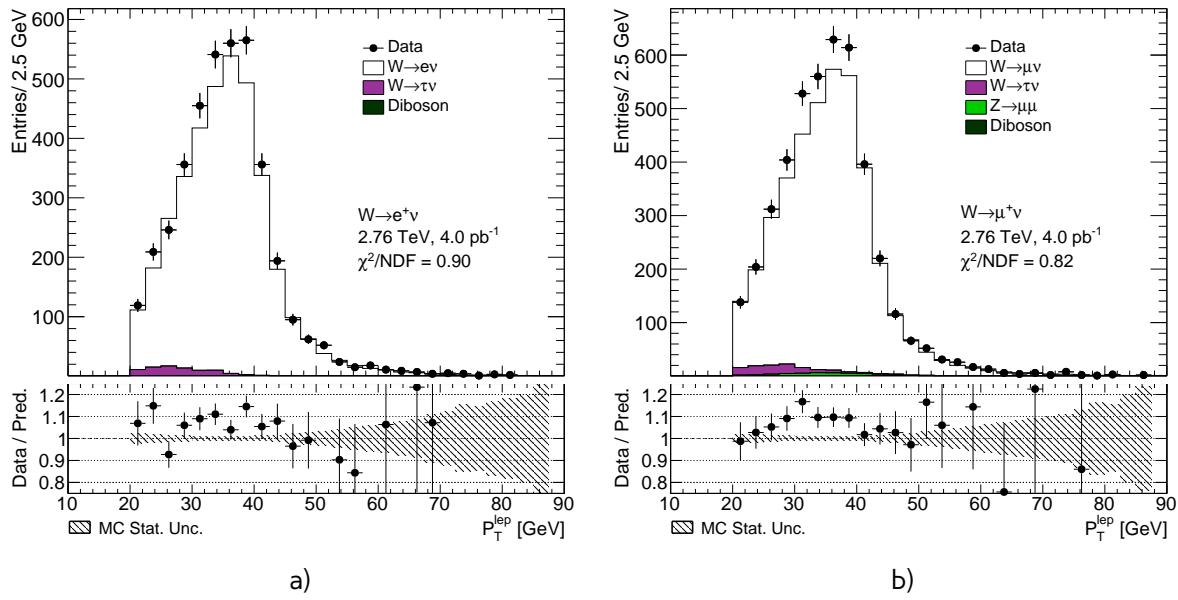


Fig. 13.5: Lepton transverse momentum distribution from the a)  $W^+ \rightarrow e^+\nu$  selection and b) the  $W^+ \rightarrow \mu^+\nu$  selection.

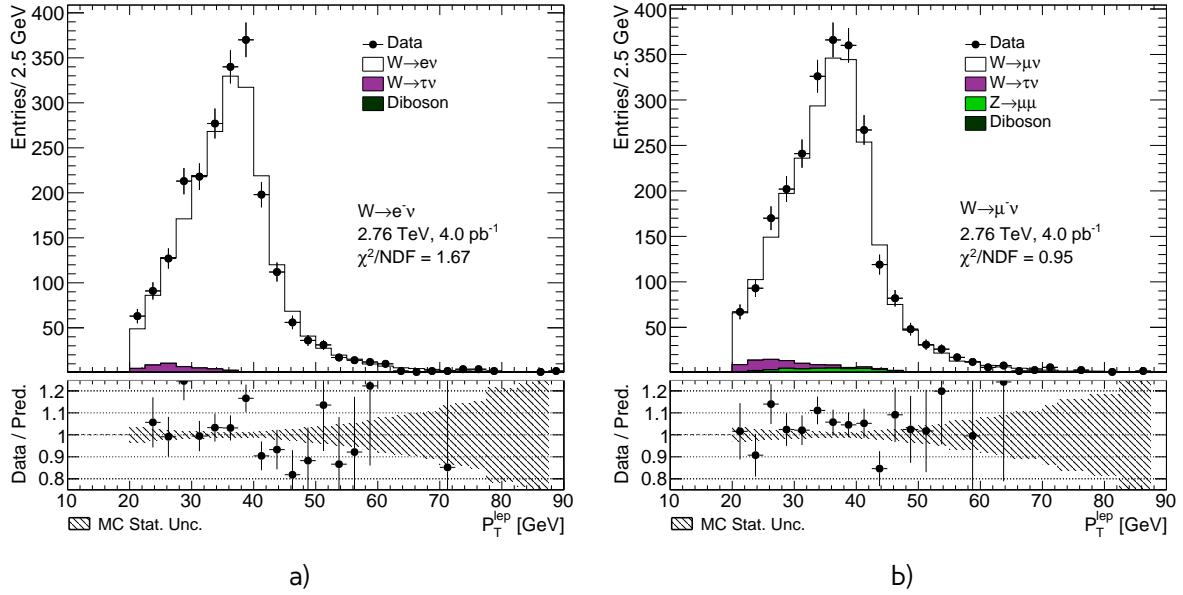


Fig. 13.6: Lepton transverse momentum distribution from the a)  $W^- \rightarrow e^- \nu$  selection and b) the  $W^- \rightarrow \mu^- \nu$  selection.

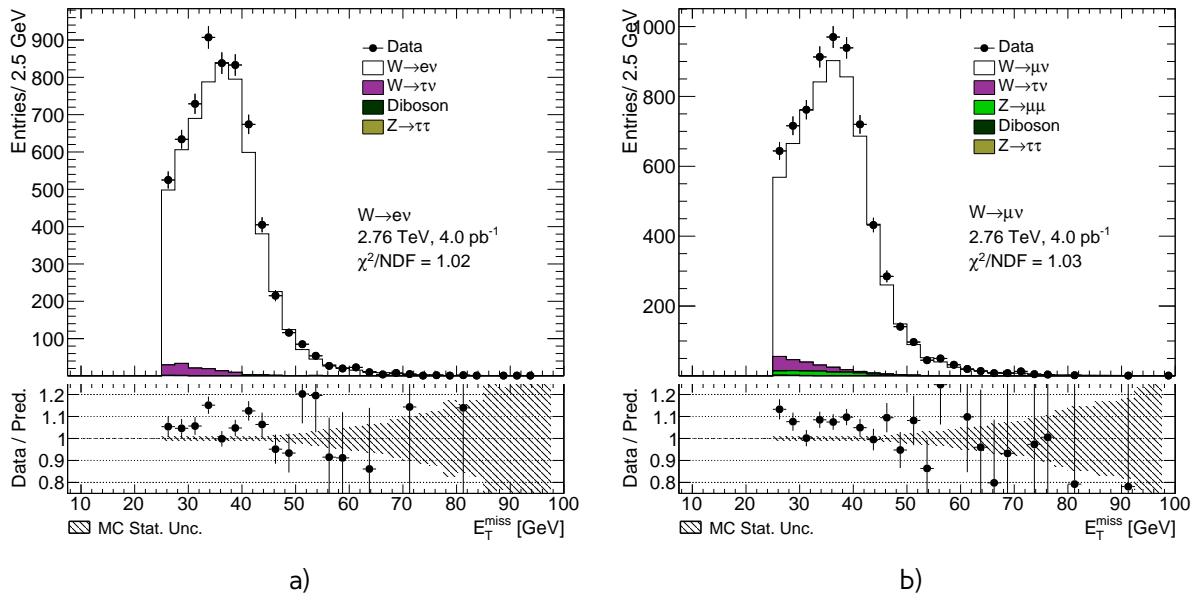


Fig. 13.7: Missing transverse energy distribution from the a)  $W \rightarrow e \nu$  selection and b) the  $W \rightarrow \mu \nu$  selection.

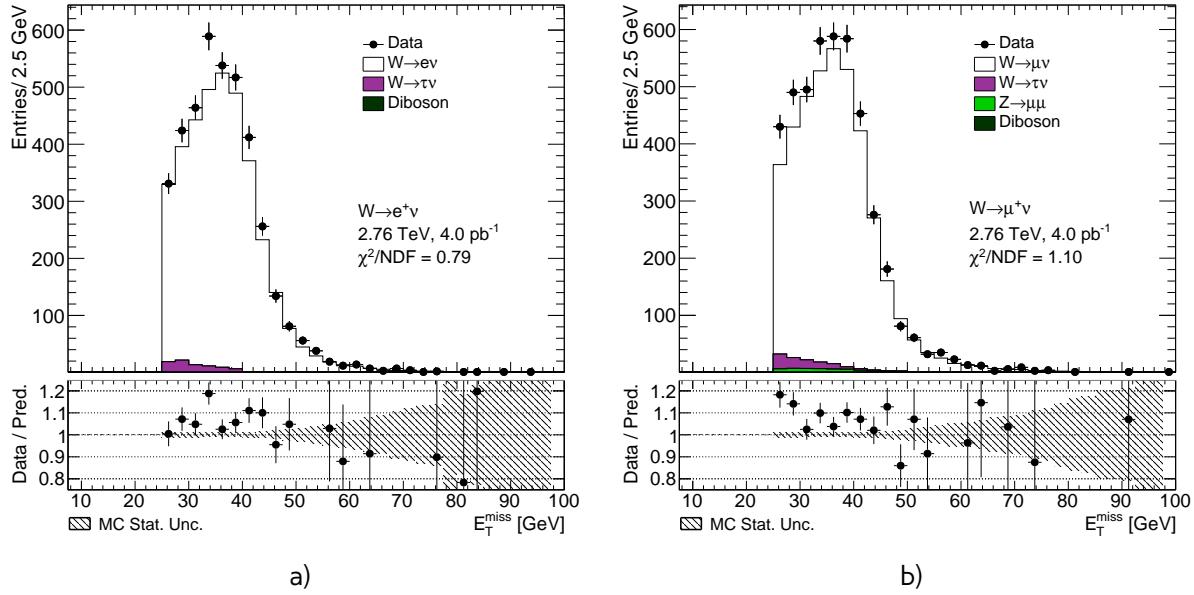


Fig. 13.8: Missing transverse energy distribution from the a)  $W^+ \rightarrow e^+\nu$  selection and b) the  $W^+ \rightarrow \mu^+\nu$  selection.

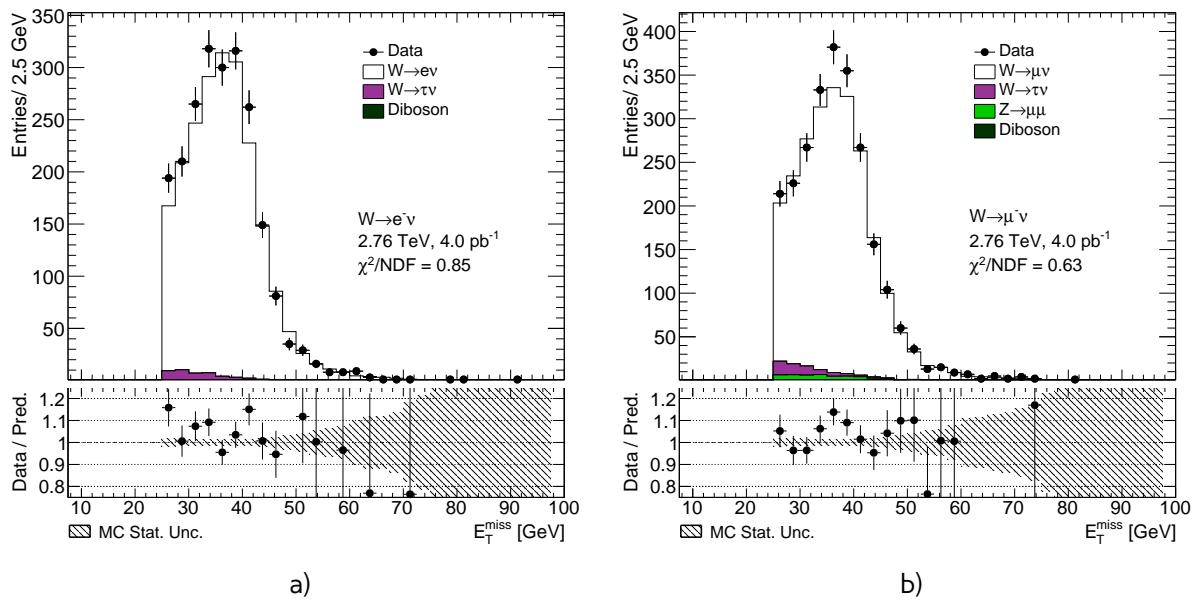


Fig. 13.9: Missing transverse energy distribution from the a)  $W^- \rightarrow e^-\nu$  selection and b) the  $W^- \rightarrow \mu^-\nu$  selection.

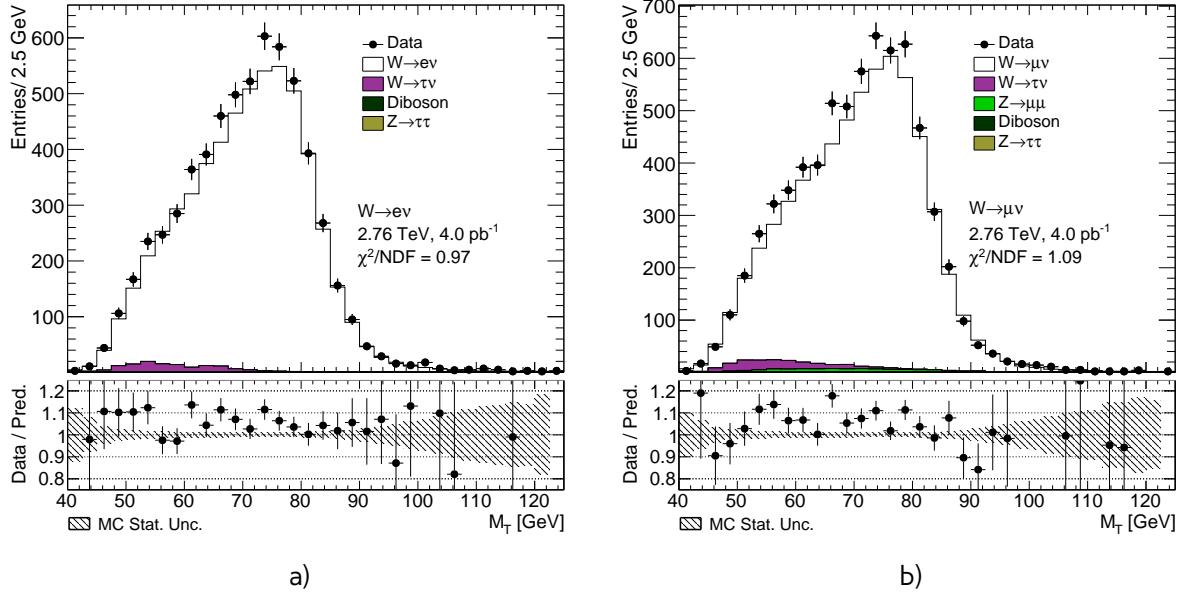


Fig. 13.10: Transverse mass distribution distribution from the a)  $W \rightarrow e\nu$  selection and b) the  $W \rightarrow \mu\nu$  selection.

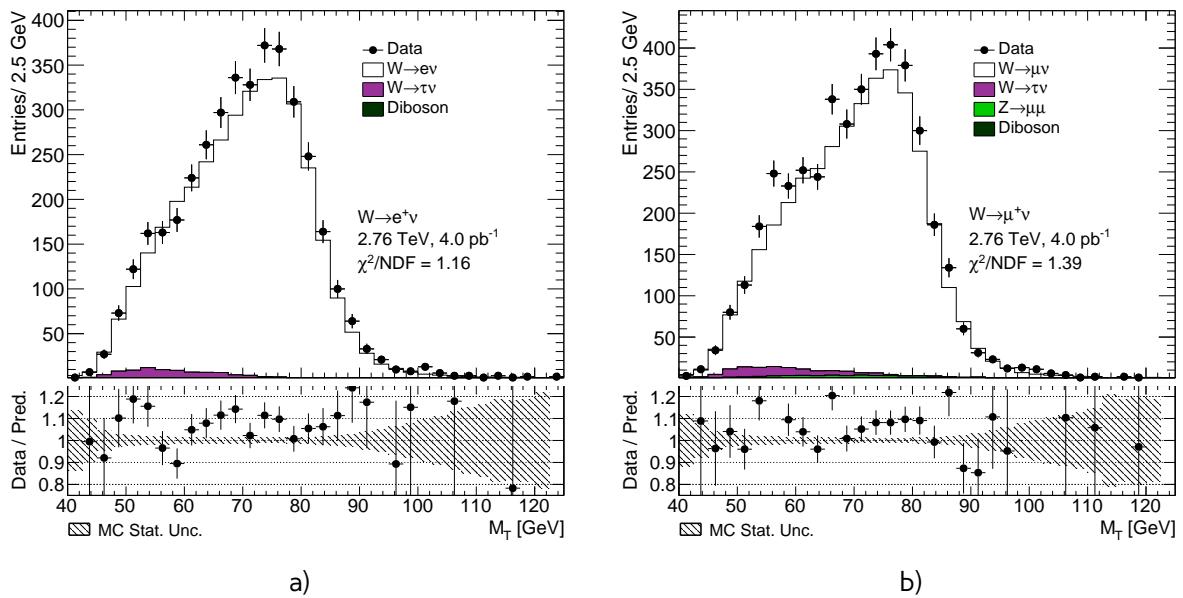


Fig. 13.11: Transverse mass distribution distribution from the a)  $W^+ \rightarrow e^+\nu$  selection and b) the  $W^+ \rightarrow \mu^+\nu$  selection.

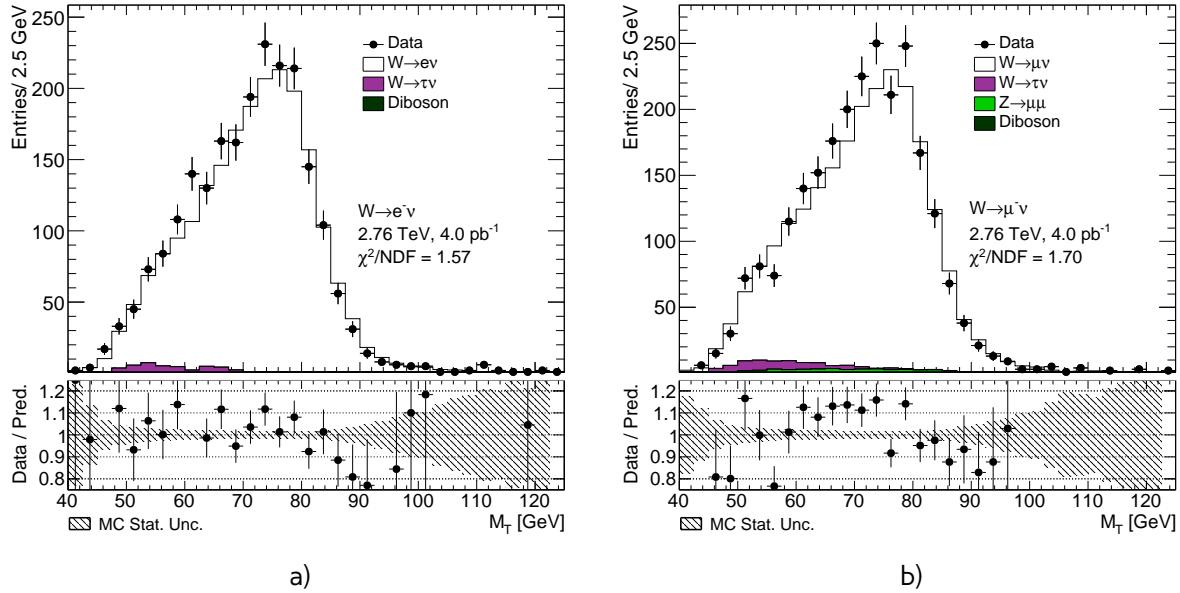


Fig. 13.12: Transverse mass distribution distribution from the a)  $W^- \rightarrow e^-\nu$  selection and b) the  $W^- \rightarrow \mu^-\nu$  selection.

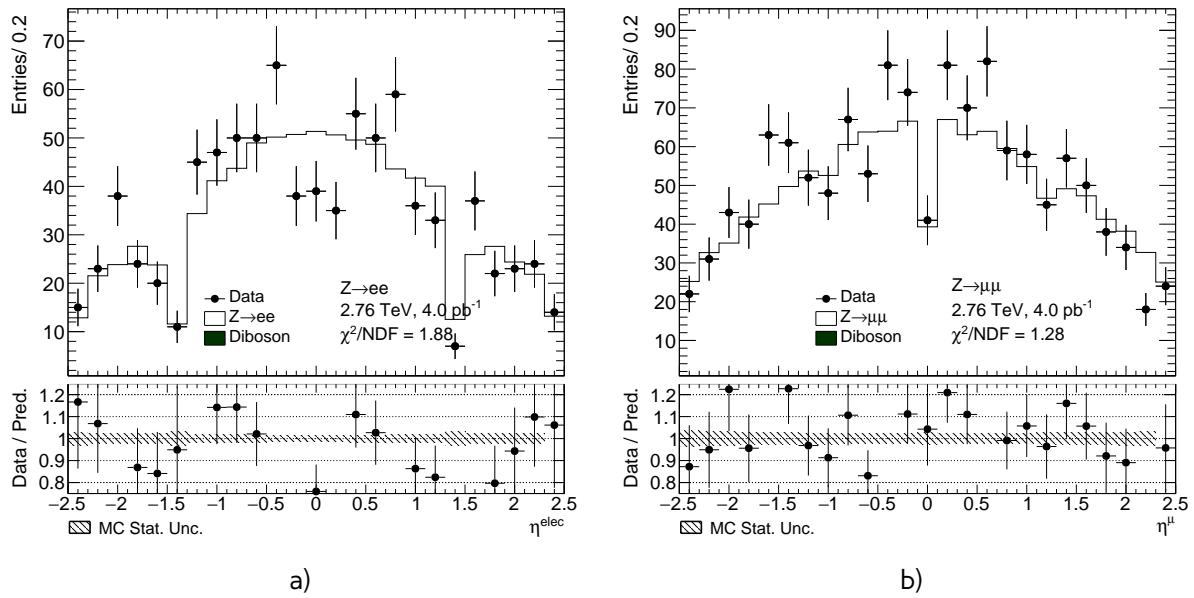


Fig. 13.13: Lepton pseudorapidity distributions from the a)  $Z \rightarrow e^+e^-$  b)  $Z \rightarrow \mu^+\mu^-$

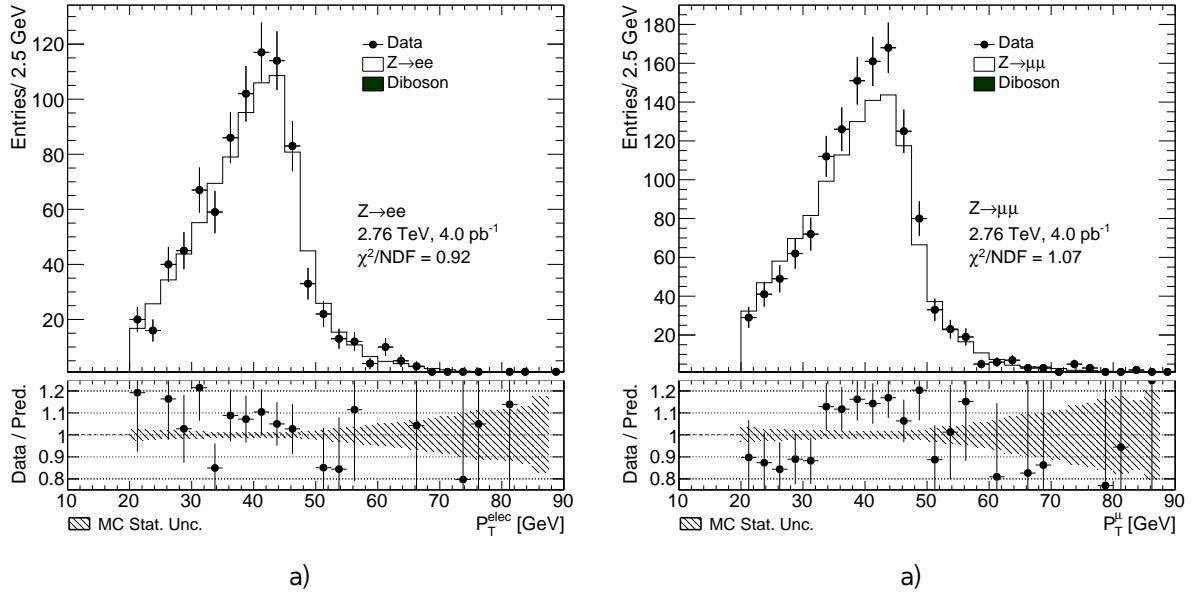


Fig. 13.14: Lepton transverse momentum distributions from the a)  $Z \rightarrow e^+e^-$  b)  $Z \rightarrow \mu^+\mu^-$

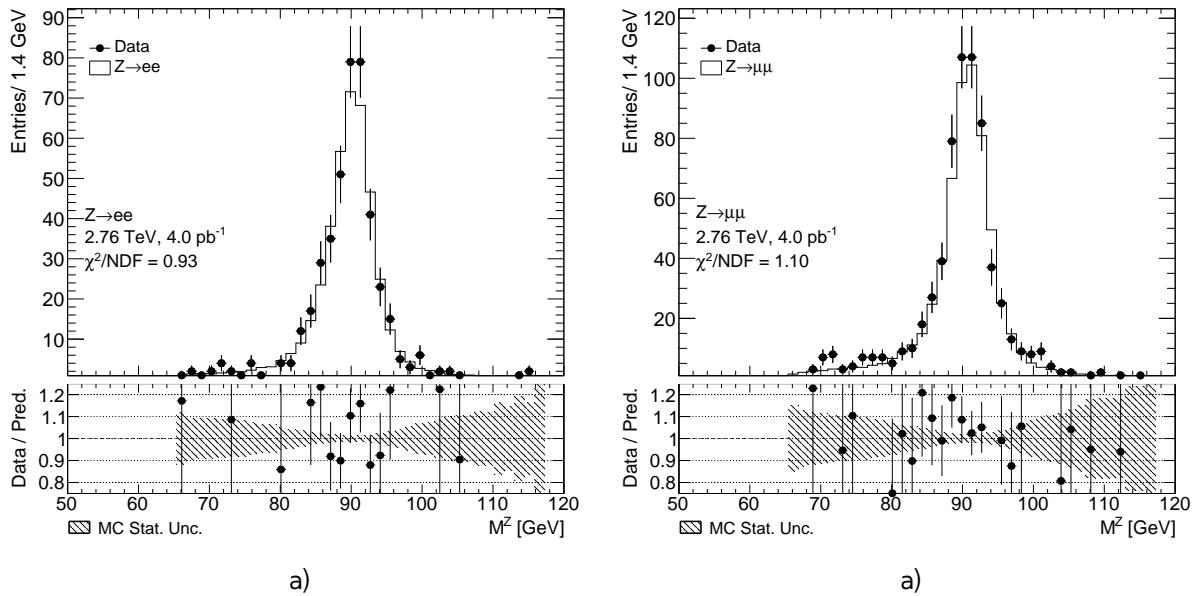


Fig. 13.15: Dilepton mass distribution distributions from the a)  $Z \rightarrow e^+e^-$  b)  $Z \rightarrow \mu^+\mu^-$

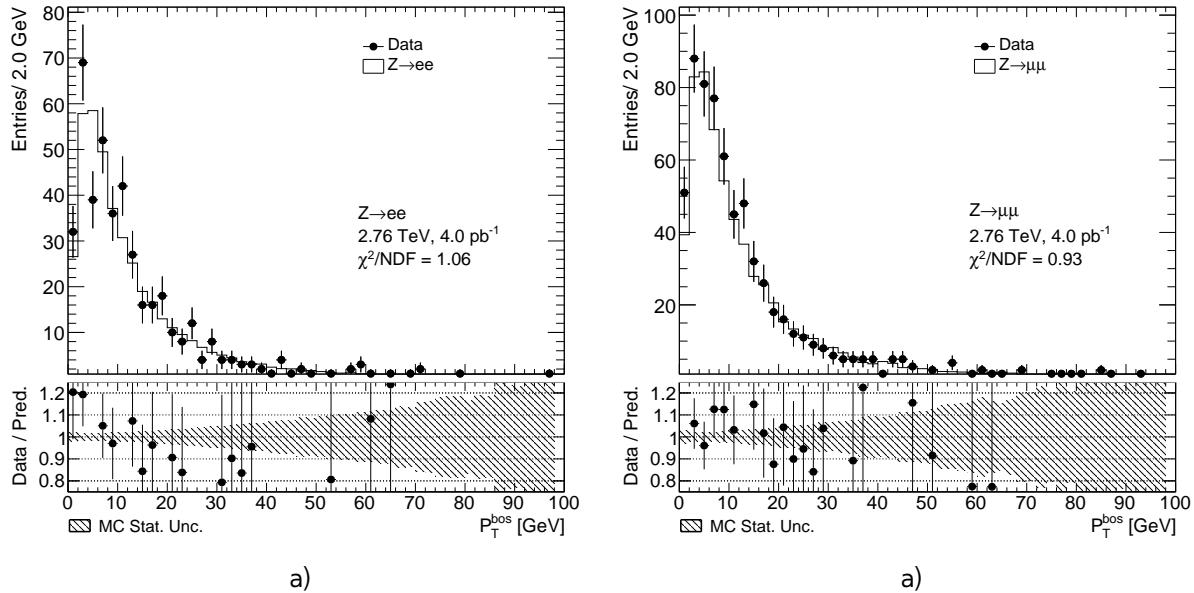


Fig. 13.16: Z boson transverse momentum distributions from the a)  $Z \rightarrow e^+e^-$  b)  $Z \rightarrow \mu^+\mu^-$

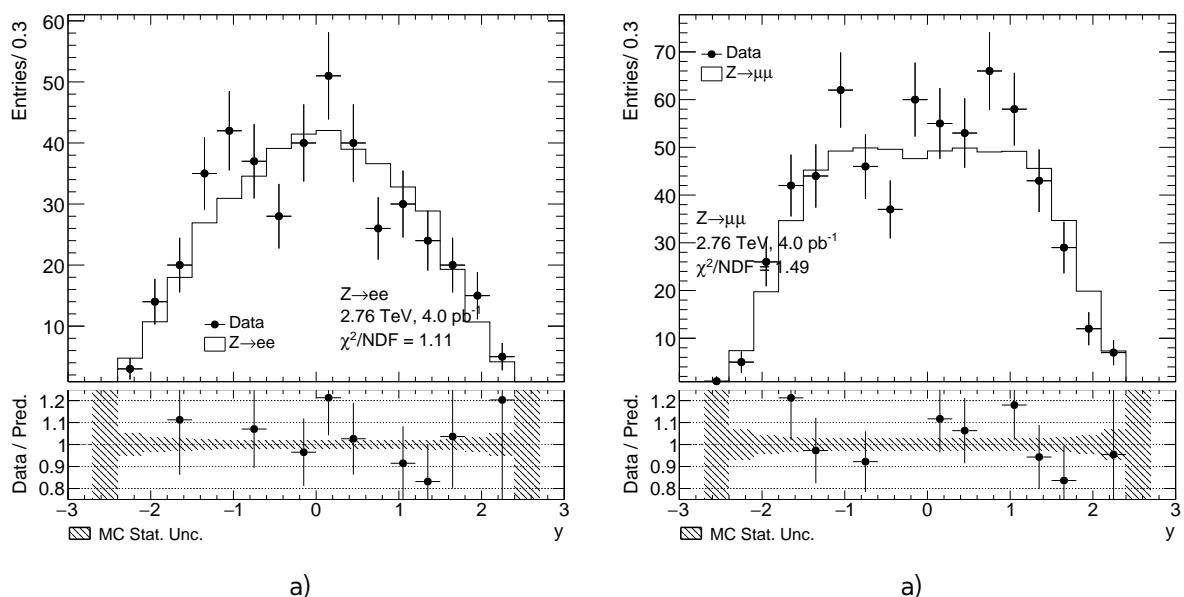


Fig. 13.17: Z boson rapidity distribution from the a)  $Z \rightarrow e^+e^-$  b)  $Z \rightarrow \mu^+\mu^-$



596 Chapter **14**

597 **Methodology of Cross-Section Measurement**

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

$$\sigma_{fid} = \frac{N - B}{C_{W/Z} L_{int}}, \quad (14.1)$$

598 where

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

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

$$C_{W/Z} = \frac{N_{MC,rec}}{N_{MC,gen,cut}}, \quad (14.2)$$

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

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

$$\sigma_{tot} = \sigma_{W/Z} \cdot BR(W/Z \rightarrow l\nu/l\bar{l}) = \frac{\sigma_{fid}}{A_{W/Z}} = \frac{N - B}{A_{W/Z} C_{W/Z} L_{int}}, \quad (14.3)$$

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:

$$A_{W/Z} = \frac{N_{MC,gen,cut}}{N_{MC,gen,all}}, \quad (14.4)$$

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

606 **14.1 Fiducial volume definition**

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

$$W \rightarrow l\nu : p_T^l > 20 GeV, |\eta^l| < 2.5, p_T^\nu < 20 GeV, M_T^W > 40 GeV \quad (14.5)$$

$$Z \rightarrow ll : p_T^l > 20\text{GeV}, |\eta^l| < 2.5, 66\text{GeV} < m_{ll} < 116\text{GeV} \quad (14.6)$$

607 Additionally, for a future combination, 13 TeV analysis fiducial phase space have been used:

$$W \rightarrow l\nu : p_T^l > 20\text{GeV}, |\eta^l| < 2.5, p_T^\nu < 20\text{GeV}, M_T^W > 40\text{GeV} \quad (14.7)$$

$$Z \rightarrow ll : p_T^l > 20\text{GeV}, |\eta^l| < 2.5, 66\text{GeV} < m_{ll} < 116\text{GeV} \quad (14.8)$$

608

# Chapter 15

## 609 Uncertainties in the cross-section measurement

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

### 613 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:

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

614 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:

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

615 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  
 616 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:

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

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

### 619 15.2 Experimental systematic uncertainties

620 Sources of experimental uncertainties, methods of estimation and their effect on a  $C_{W,Z}$  are summarized in a Tab. 15.2. Systematical errors coming from a hadron recoil calculation are discussed in  
 621 a Sec. 11.3.

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

623 Electron energy scale correction, described in Sec. 10.2 has associated uncertainties coming from  
624 <reference>:

- 626 • Statistical component of the scale uncertainty  
627 • Uncertainty from the possible bias of the calibration method  
628 • Scale uncertainty from the choice of generator  
629 • Uncertainty from the presampler energy scale  
630 • Imperfect knowledge of the material in front of EM calorimeter.

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

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

- 634 • MS modelling  
635 • ID modelling  
636 • overall scale

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

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

640 In case of 2.76 TeV analysis scale factor errors are considered to be enlarged for a statistical and  
641 uncorrelated components, so correlated error is assumed to be negligible. The toy MC experiments  
642 are performed for electron reconstruction, identification and trigger scale factors and muon recon-  
643 struction + identification scale factors. In the current analysis 30 pseudo-experiments are used with  
644 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 | -  | -  | Wplusmunu  | Wminmunu   | -  | Zmumu  |
| - Muon ID energy scale       | Offset | -  | -  |  |  | -  | 0.01   |
| - Muon MS energy scale       | Offset | -  | -  | 0.02   | 0.01   | -  | 0.01   |
| Hadron recoil scale          | Offset |  |  | 0.01   | 0.00   | -  | -  |
| Hadron recoil resolution     | Offset |  |  |  |  | -  | -  |
| Background                   |        |  |  |  |  |  |  |
| PDF error                    |        |  |  |  |  |  |  |
| Statistics                   |        |  |  |  |  |  |  |
| Total                        |        |  |  |  |  |  |  |

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 (

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

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



660

# Chapter 16

## 661 Results of the Cross Section Measurement

662 16.1 Cross-section combination

663 16.2 Comparation with Theoretical Predictions



664

# Chapter 17

665 **PDF fits results**



666

# Chapter 18

667

## Summary