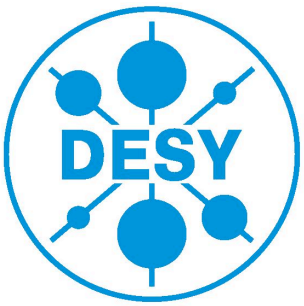


Background studies for the measurement of the Drell-Yan cross section using 2012 ATLAS data

Sebastian Keckert, University of Siegen, Germany

September 5, 2013

Abstract

This report presents background studies for the process $Z \rightarrow e^+e^-$. The analysis is performed using $\sqrt{s} = 8\,\text{TeV}$ ATLAS data collected in 2012 corresponding to an integrated luminosity of $\int L\,\mathrm{d}t = 20.3\,\text{fb}^{-1}$. In the E_T^{miss} observable a significant discrepancy between data and simulations is observed and a data calibration method is studied. Assuming a dependence proportional to E_T^{miss} the analysis is performed by dividing the di-lepton invariant mass into five bins and by using the full mass range. In conclusion the impact of the calibration on an additional $t\bar{t}$ background scale is studied.

Contents

1 Introduction

The Large Hadron Collider (LHC) at CERN [1] (European Organization for Nuclear Research), Geneva, has recently become popular for the discovery of a Higgs boson. However, that has not been its only purpose. The data collected with six detectors is used to study the properties of the fundamental particles and forces and to answer a variety of questions, e. g. of physics beyond the Standard Model, the asymmetry of matter and anti-matter in the universe and many others.

The LHC is a proton-proton storage ring operated in 2012 with a centre of mass energy $\sqrt{s} = 8\,\text{TeV}$. The proton beams collide at four interaction points where four big experiments ALICE, ATLAS, CMS and LHCb are located. ALICE (A Large Ion Collider Experiment) is specialized in analyzing the quark-gluon plasma produced in lead-ion collisions. ATLAS (A Torodial LHC ApparatuS) and CMS (Compact Muon Solenoid) are general-purpose detectors which cover a broad field of particle physics research. Despite similarities in the physics program the design of the two detectors differs significantly by which means different priorities are set. The data used in this analysis was collected by the ATLAS detector which will be explained in more detail in section 1.2. LHCb (LHC beauty) is a forward detector which specializes in the study of b-quarks. Near the ATLAS collision point LHCf (LHC forward) is built to study particles produced very close to the beam direction. With it models used to describe the interactions of high-energy cosmic rays with the atmosphere can be tested. Another small-sized experiment is TOTEM (TOTal Electric and Diffractive cross-section Measurement). Its detectors are located near CMS and measure the total cross section of pp interactions.

The studies presented in this report have been done in the ZeeD (Z to ee DESY) group at DESY. The physical basics of this topic are explained in section 1.1 while section 1.3 briefly describes the tools and software used in the analysis.

Section 2 introduces the data set and the data selection criteria. Also the studied sources of background processes in this analysis are explained. The report then focuses on the E_T^{miss} observable which leads to the motivation of the calibration performed in this study. Afterwards the calibration method is explained and the results are presented. The report ends with a summary and an outlook in section 4.

1.1 The Drell-Yan process

The Drell-Yan process denotes the production of a vector boson (γ, W or Z) by the annihilation of a quark and an anti-quark and the subsequent decay into a lepton pair. It is named after Sidney David Drell und Tung-Mow Yan who first suggested this process in 1970 [2].

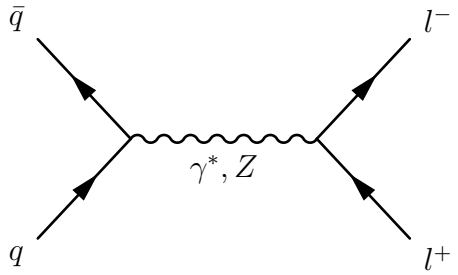


Figure 1: The Drell-Yan process $q\bar{q} \rightarrow \gamma^\star/Z \rightarrow l^+l^-$.

The measurement of the Drell-Yan process plays an important role in Standard Model (SM) physics. It can be used as testing ground for Quantum Chronodynamics (QCD) since theoretical calculation are available up to next-to-next-to-leading order (NNLO). Crucial ingredients for these calculation are the momentum distribution functions of partons in the proton (PDFs). A variety of different PDFs is available from phenomenological analyses which also can be tested.

The study of cross section and branching ratios of the Z boson and comparisons to $W^\pm$ bosons are useful to improve the understanding of the electroweak force. The SM can be tested here, too, e. g. in the field of lepton universality. Beyond that the precise understanding of the Z boson is important for studying the Higgs boson as it can appear as a decay product. Furthermore the Z boson is an important source of background for searches for New Physics e. g. in Supersymmetry searches.

In this analysis only the decay channel $Z \rightarrow e^+e^-$ is considered.

1.2 The ATLAS Experiment

The ATLAS (A Torodial LHC ApparatuS) detector is one of the two big general-purpose particle detectors at the LHC. It has a barrel-like shape in order to measure particles almost regardless of their direction. The detector is built out of several different subdetectors which are arranged onion-like around the interaction point (s. fig. 2).

The innermost part (Inner Detector, ID) consists of the tracking system. Several layers of silicon pixel and strip detectors form the Semi-Conductor Tracker (SCT) and provide high precision spatial resolution. The SCT is surrounded by the Transition Radiation Tracker which also gives tracking information and contributes to electron identification. The ID is enclosed by a superconduction solenoid which generates a 2 T axial magnetic field to measure the momentum of charged particles.
The next layer contains two kinds of calorimeters. The purpose of the liquid argon (LAr) electromagnetic (EM) calorimeter and the hadronic calorimeter (HCAL) is to measure the energy of particles by absorbing them inside the LAr. These calorimeters work via periodic sampling. This means that they consist of alternating active and passive layers where particles are only measured (sampled) within the active layers. To cover a rapidity range as big as possible the calorimeters are split into different parts. In the

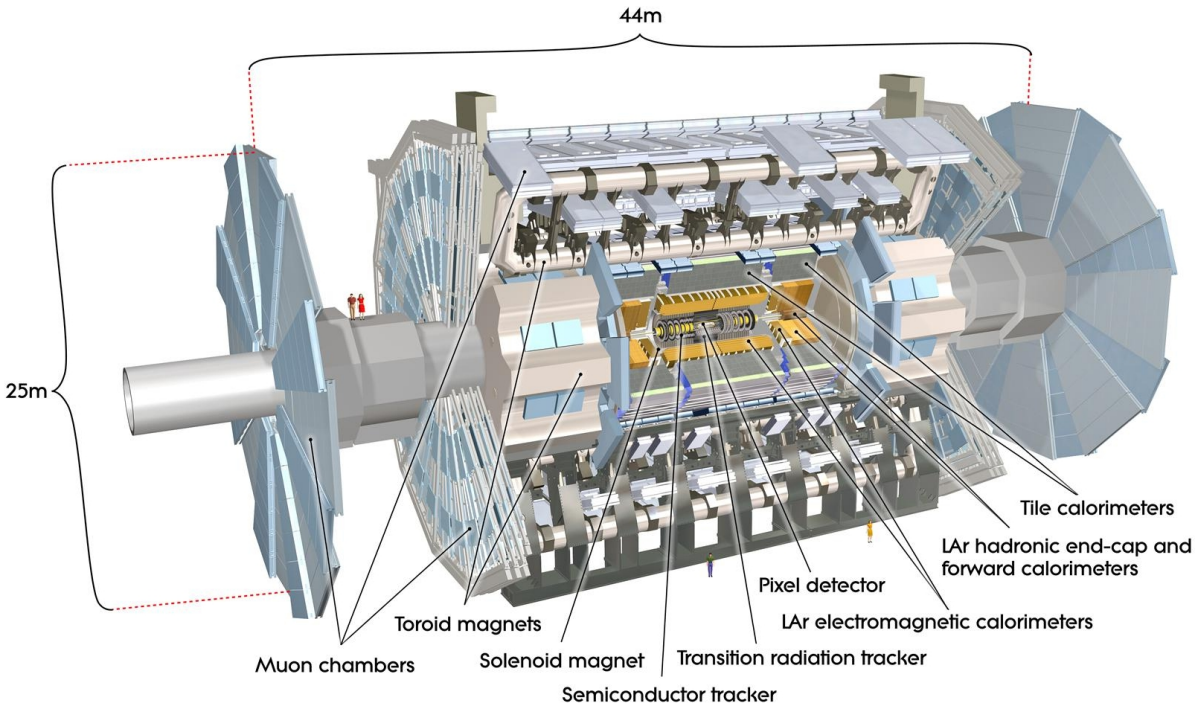


Figure 2: A schematic view of the ATLAS detector [3].

central $|\eta|$ range the calorimeters are build barrel-like around the ID. For the forward part ($|\eta| \gtrsim 1.5$) end-cap components are installed around the beam pipe.
The HCAL is surrounded by the muon spectrometer (MS). The MS is based on three large superconducting toroids with coils arranged in an eight-fold symmetry. The MS precisely measures the momentum of muons escaping the detector by the help of Monitored Drift Tubes. At large pseudo-rapidities Cathode Strip Chambers are used.

In pp collisions large numbers of particles are produced and detected resulting in the production of a large amount of data. Since it is impossible to readout and store all events a three-level trigger system is installed consisting of Level-1 (L1), Level-2 (L2) and the Event Filter (EF). This tigger system reduces the event rate to approximately 200 Hz before data transfer to mass storage.
Electrons are triggered in the pseudo-rapidity range $|\eta_e| < 2.5$ where the electromagnetic calorimeter is finely segmented. The analysis on the Z cross section in the electron channel is done in two fiducial regions. In the central-central analysis (ZCC) both electrons' pseudo-rapidities are required to be in that range. For the central-forward analysis (ZCF) one central electron and one electron with $2.5 < |\eta_e| < 4.9$ is required. The analysis presented in this report is based only on ZCC data [4].

1.3 Analysis software

The analysis presented in this report is performed using a tool called ZeeD [5]. It is ATHENA-based so all ATHENA services are available, but ZeeD is specialized to analyze the Z boson decay to an electron-positron pair. ATHENA is an ATLAS-specific control framework used for data analysis. In this analysis preprocessed data is used which is available in the "ZeeD TTree" data format. The advantage of this is a factor of 10 CPU improvement and a factor of 200 disc space reduction [6]. The cuts applied to data in order to select the signal events are listed in table 1.

2 Data selection and background contributions

2.1 Data selection

The analysis is performed using $\sqrt{s} = 8\,\text{TeV}$ ATLAS data collected in 2012 corresponding to an integrated luminosity of $\int L\,\mathrm{d}t = 20.3\,\text{fb}^{-1}$. In order to select the signal events the following cuts are applied:

ZeeD selection cut	description
ExistZ	2 electrons with invariant mass $> 30\,\text{GeV}$ required
ZmassZ	Invariant mass of electron pair within $66 - 150\,\text{GeV}$
IsEMMediumPPBothElecZ	Electron identification cut
ChargeBothElecOppositeZ	Electrons have opposite charges
PtMinBothElecZ	The transverse momentum (p_T) of electrons $> 20\,\text{GeV}$
EtaMaxBothElecZ	The pseudo-rapidity $\lvert\eta\rvert$ of the two electrons < 2.47 (central region of the detector)
MaxTwoGoodElec	Not more than two identified electrons in the event
NtracksAtPrimVtx	Cut on number of tracks at the primary vertex
EtaCrackBothElecZ	Both electrons are outside a crack in the calorimeter defined by $1.37 < \lvert\eta\rvert < 1.52$
AuthorBothElecZ	Electons are reconstructed by a cluster algorithm
OQMaps	Both electrons must not hit malfunctioning regions of the calorimeter
ZeeDCutLArEventVeto	LAr hole veto

Table 1: Signal selection cuts.

2.2 Background contributions

After the data selection by application of the above listed cuts the event sample contains contributions from events which do not contain the process $Z \to e^+e^-$. These events are called background and have to be subtracted in order to get correct and precise results. As in the procedure of signal selection events with an electron and a positron fulfilling special kinematic constraints are required, two sources of background are possible: On

the one hand there can be processes with electrons fulfilling the same cuts, which are called 'real' electrons. On the other hand it is possible that due to detector inefficiencies and uncertainties objects can be mis-reconstructed as electrons which are then called 'fake'. In this analysis the total background is divided into two categories – QCD background and electroweak (EW) background – which are explained in the following:

QCD background events originate from hadronic jets. The reconstruction of jets can lead to fake electrons or there are real electrons e. g. from semi-leptonic decays which fulfill all signal selection cuts. In general the cross sections for these processes are so big that Monte Carlo simulations with sufficient statistics cannot be computed. Therefore a data-driven method is used to subtract the QCD background.

The Electroweak background summarizes the following sources of background:

- $t\bar{t} \rightarrow b\bar{b}\,W^+W^-$
- $Z \rightarrow \tau^+\tau^-$
- $W^\pm \rightarrow e^\pm \overset{(-)}{\nu_e}$
- $W^\pm \rightarrow \tau^\pm \overset{(-)}{\nu_\tau}$
- ZZ, WZ, ZZ

The cross sections of these processes are small enough to compute Monte Carlo (MC) simulations. The studies described in the following only focus on the electroweak background.

2.3 Electroweak background

To compute MC samples for all electroweak background processes the following MC generators and parton density functions (PDFs) have been used:

process	MC generator	PDF
$Z \rightarrow e^+e^-$ (Signal MC)	Powheg+Pythia8	CT10
$t\bar{t}$	MC@NLO+Herwig	CT10
$Z \rightarrow \tau^+\tau^-$	Powheg+Pythia8	CT10
$W^\pm \rightarrow e^\pm \overset{(-)}{\nu_e}$	Sherpa	CT10
$W^\pm \rightarrow \tau^\pm \overset{(-)}{\nu_\tau}$	Sherpa	CT10
ZZ	Herwig	CTEQ6L1
WZ	Herwig	CTEQ6L1
WW	Herwig	CTEQ6L1

Table 2: MC generators and PDFs to simulate the signal and the $t\bar{t}$ and electroweak background.

In order to be compatible with the data the simulated distributions have to be scaled to data luminosity (L_{data}). This is done according to the formula

$$scale = \frac{L_{\text{data}} \cdot \sigma_{\text{MC}}}{N_{\text{MC}}} \tag{1}$$

where σ_{MC} denotes the cross section of the simulated process and N_{MC} contains the total number of simulated events. The di-electron mass distribution for the data events and EW background sources is shown in figure 3.

Due to the presence of a real Z boson the shapes of the background contributions ZZ and WZ are similar to the signal shape and therefore become dominant around the Z mass peak.
The production of a $t\bar{t}$ pair contributes as background if both W bosons from the t and $\bar{t}$ decays themselves decay into an electron and a positron. Due to the broad momentum distributions in these multi-body decays all kinematic cuts can be fulfilled. This leads to a background contribution slowly decreasing as a function of the di-electron mass, which becomes especially dominant in the high mass region.
For masses smaller than the Z mass peak the process $Z \rightarrow \tau^+\tau^-$ is important. To fulfill the cuts both τ have to decay into electrons. The production of four neutrinos in these decays leads to missing mass which shifts the di-lepton invariant mass to lower values.
The contributions from W or WW events are relatively small. W events would contribute to the background as 'fake' electrons, which are suppressed by the electron identification requirements.

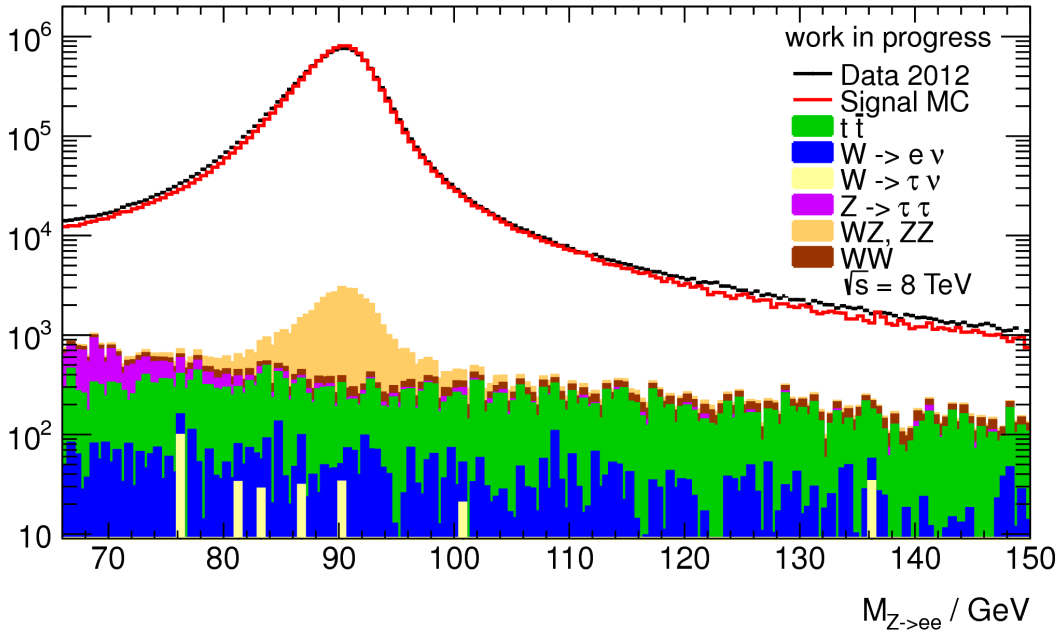


Figure 3: Di-electron mass distribution, simulations scaled to data luminosity.

2.4 Missing transverse energy (E_T^{miss})

For further optimization of the background subtraction the missing transverse energy (E_T^{miss}) is studied. Figure 4 shows the distributions of signal and background.

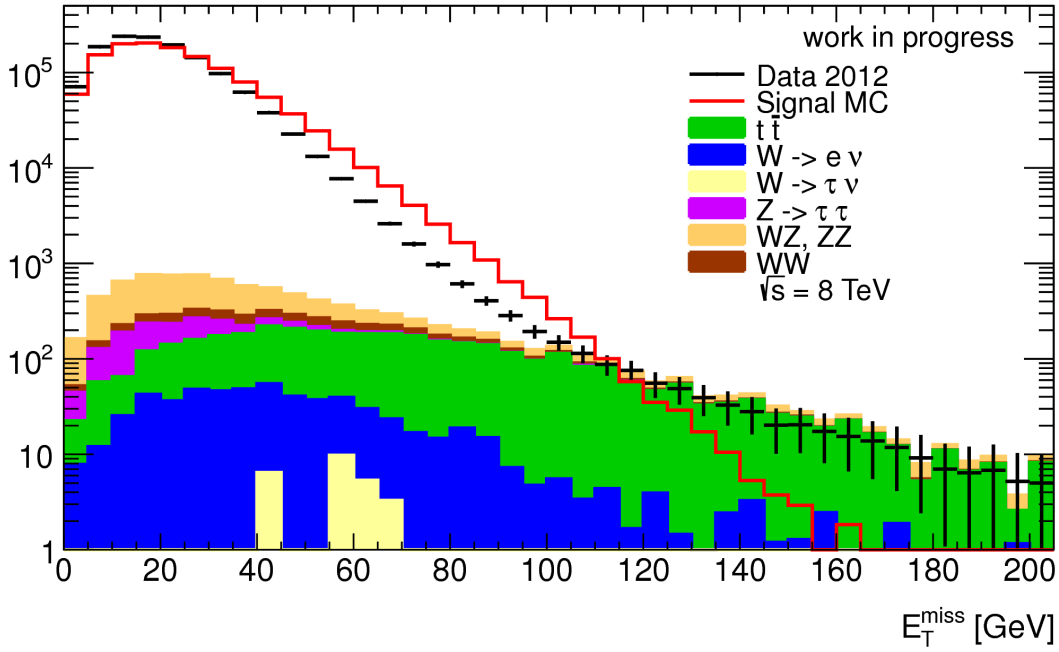


Figure 4: Distribution of missing transverse energy.

In leading order the Drell-Yan process $Z \rightarrow e^+e^-$ does not lead to any missing energy since both electrons are fully absorbed in the calorimeter of the detector. At $\sqrt{s} = 8\,\text{TeV}$ a significant fraction of these events show a Z boson with high transverse momentum (p_T^Z) due to QCD radiation. If the energy of these jets is not reconstructed correctly, e. g. due to detector effects or high energetic neutrinos from semi-leptonic decays, missing energy is observed. As expected, the signal dominates at small E_T^{miss} and drops quickly with increasing E_T^{miss}. Also the background contributions containing real Z bosons drop in a similar way.
For the $t\bar{t}$ background the picture is quite different: Here two energetic neutrinos are produced which lead to a greater amount of E_T^{miss}. The shape of this background is also different, instead of a quick drop a broad distribution is observed. Around 50 GeV the distribution shows a peak, which is the expected average missing energy carried by the two neutrinos if the $t\bar{t}$ pair is produced at rest.

By comparing data to the sum of signal MC and all background distributions (s. fig. 5) a significant discrepancy can be seen. At low E_T^{miss} in the range of 0 to 20 GeV the amount of detected events is above the expectation from simulations while from 40 GeV on too few events are observed. To correct this discrepancy a data calibration has been studied which is presented in the following section.

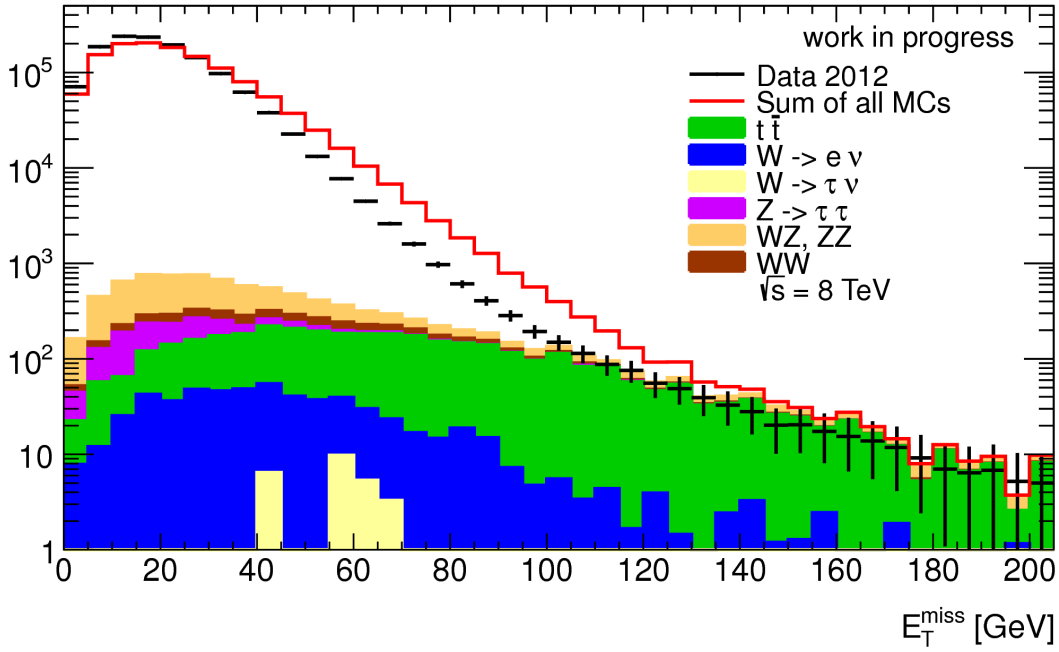


Figure 5: E_T^{miss}, the red line shows the sum of signal MC and all background simulations.

3 Calibration

3.1 Motivation

As described in section 2.4 a significant discrepancy between data and simulations is present in the observable E_T^{miss}. It is assumed that the signal MC is trustworthy while the reconstruction of E_T^{miss} can have some deficiencies. Therefore a calibration is needed. The dependecies of this calibration are a priori unknown and have to be studied. In this analysis a dependence proportional to E_T^{miss} is assumed.

3.2 Method and expectation

A calibration as proposed above can be implemented by multiplying all reconstructed values of E_T^{miss} with a constant calibration factor. A calibration factor of 1.0 will return an unchanged distribution. To find the factor fitting best a number of different calibrations is applied to the data and the resulting set of distributions is studied.
To determine the set of calibration factors needed for this study an estimate is made in the range of small E_T^{miss} where the background contributions are comparatively small: For three different intervals $0-40$, $0-60$ and $0-80\,\mathrm{GeV}$ (s. fig. 6) the mean values for signal MC and data distribution are calculated while all background simulations are subtracted from data. The ratio $\frac{\mathrm{mean}^{\mathrm{MC}}}{\mathrm{mean}^{\mathrm{data}}}$ indicates the 'shift' between data and MC and is used as an expectation for the calibration constant. Table 3 shows the values obtained in the different intervals.

interval	ratio
$0 - 40\,\text{GeV}$	1.079
$0 - 60\,\text{GeV}$	1.124
$0 - 80\,\text{GeV}$	1.141

Table 3: Ratio of mean values obtained in different intervals.

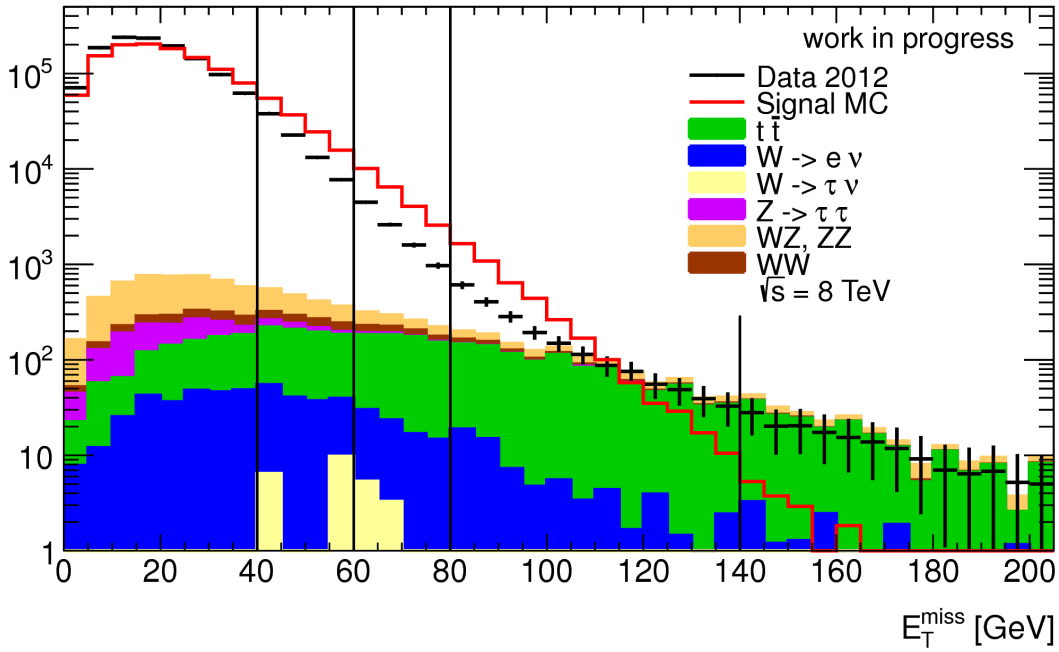


Figure 6: E_T^{miss}, vertical black lines at 40, 60 and 80 GeV mark the right borders of the intervals used for the calculation of mean values. The vertical black line at 140 GeV marks the left border of the range used for scaling the $t\bar{t}$ background.

The following set of calibration factors is used:

1.0 1.025 1.05 1.075 1.1 1.125 1.15 1.175 1.2 1.25 1.3 1.35 1.4 1.45 1.5

3.3 Scaling of the $t\bar{t}$ background

As is visible in figures 5 and 6 the $t\bar{t}$ background becomes dominant at high E_T^{miss}. To optimize the matching of this background contribution to the data distribution and therefore optimize the background subtraction an additional scaling factor is calculated: In the range of 140 to 200 GeV (see fig. 6) all other background simulations and the signal MC are subtracted from data. Then the integral is computed for both data and $t\bar{t}$ background and the scale factor results in

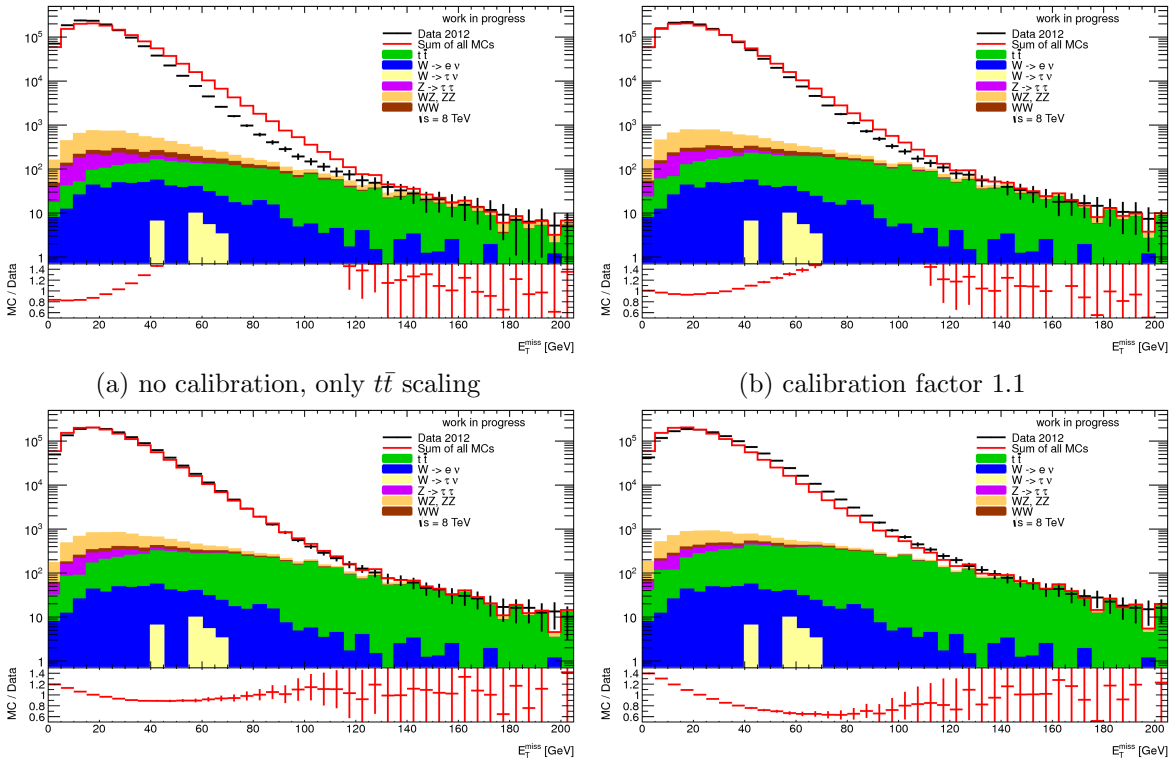
$$scale_{t\bar{t}} = \frac{N_{\text{data}} - N_{\text{MC and EW bg., excl. } t\bar{t}}}{N_{t\bar{t}}} \tag{2}$$

3.4 Results

To study the calibration described above the analysis of E_T^{miss} is performed using the full data set and bins of the invariant mass of the electron-positron pair (mass bins). These bins are defined as follows:

- $66 - 76\,\text{GeV}$
- $76 - 91\,\text{GeV}$
- $91 - 106\,\text{GeV}$
- $106 - 116\,\text{GeV}$
- $116 - 150\,\text{GeV}$

In Figure 7 data distributions for different calibration factors for the full mass range are shown. The additional scaling of the $t\bar{t}$ background as described in section 3.3 has also been applied.



(a) no calibration, only $t\bar{t}$ scaling

(b) calibration factor 1.1

(c) calibration factor 1.2

(d) calibration factor 1.3

Figure 7: E_T^{miss} distributions for different calibrations including additional $t\bar{t}$ scaling.

A calibration factor of around 1.1 or 1.2 improves the matching between data and MC in both regions described in section 2.4. At low E_T^{miss} the number of recorded data events is reduced while in the middle region the falling slope is shifted to higher values of E_T^{miss}. The additional scaling of the $t\bar{t}$ distribution leads to a good agreement of data and simulation in the tail region. A calibration of 1.3 or bigger seems to be too big and has a negative effect.
For a quantiativ analysis a χ^2-test is performed to compare the data with the sum of all MC distributions. Figure 8 shows the reduced χ^2/ndf for different calibration factors for the full mass range and all mass bins. To obtain the optimal values for the calibration, parabolas are fitted to the data points and the minima are calculated. The results are listed in table 4, the uncertainties are derived by propagation of the statistical errors of the fitted parabolas.

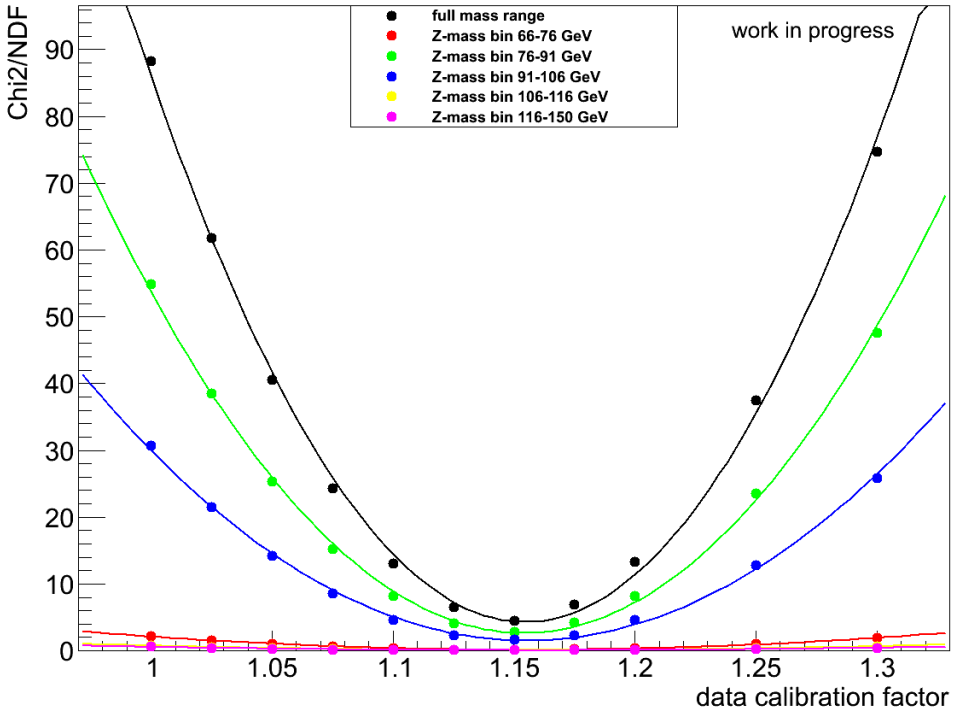


Figure 8: χ^2 distributions for different calibrations in mass bins and the full mass range.

mass bin	minimum
$66 - 76\,\text{GeV}$	1.154 ± 0.039
$76 - 91\,\text{GeV}$	1.154 ± 0.024
$91 - 106\,\text{GeV}$	1.155 ± 0.027
$106 - 116\,\text{GeV}$	1.155 ± 0.045
$116 - 150\,\text{GeV}$	1.161 ± 0.079
full mass range	1.154 ± 0.029

Table 4: Minima of parabola fits to χ^2 distributions.

All results are in very good agreement with each other nevertheless in the sidebands the statistical errors are quite big and lead to unrealistic small values of χ^2.
The $t\bar{t}$ scaling factors for each calibration are shown in figure 9. Without applying the calibration to data all scaling factors are significantly smaller than 1 which supports the application of a calibration to the data. To determine the $t\bar{t}$ scaling factors corresponding to the optimal calibrations as listed above (s. fig. 4) a linear fit is performed to the data points. The results are listed in table 5. The errors are obtained by varying the calibration factor by 1σ.

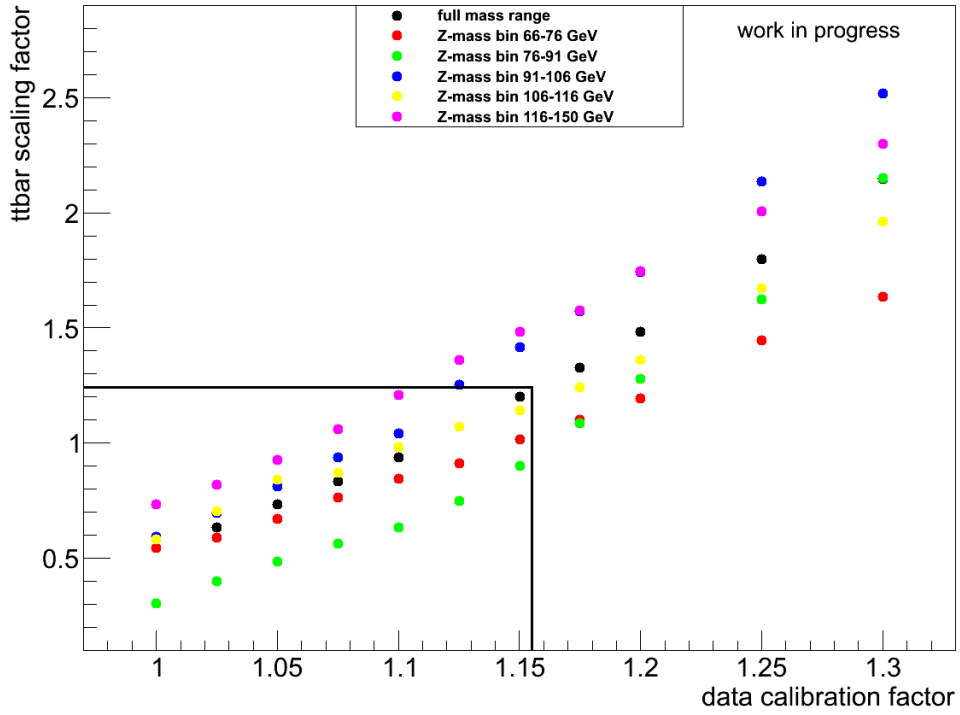


Figure 9: $t\bar{t}$ scaling factors for different calibrations in mass bins and the full mass range. The scaling factor corresponding to a calibration of 1.154 for the full mass range is marked by a black line.

mass bin	$t\bar{t}$ scaling factor
$66 - 76\,\text{GeV}$	1.03 ± 0.14
$76 - 91\,\text{GeV}$	0.96 ± 0.16
$91 - 106\,\text{GeV}$	1.44 ± 0.19
$106 - 116\,\text{GeV}$	1.18 ± 0.17
$116 - 150\,\text{GeV}$	1.53 ± 0.41
full mass range	1.23 ± 0.16

Table 5: $t\bar{t}$ scaling factors for the optimal calibration factors.

4 Conclusion and outlook

In this analysis the electroweak background in the $Z \rightarrow e^+e^-$ channel has been studied with focus on the E_T^{miss} observable. By comparing data to the sum of signal MC and all background distributions a significant discrepancy could be seen which led to the implementation of an energy-dependent calibration. This calibration was assumed to be proportional to E_T^{miss}.

The optimal calibration factor has been determined by performing χ^2 tests in mass bins and in the full mass range. The results are well compatible with each other and led to an optimal calibration factor of around 1.15.

Additionally the $t\bar{t}$ background has been rescaled to fit the data in the tail region of the E_T^{miss} distribution, defined as $140 - 200\,\text{GeV}$. Without applying the calibration to data all scaling factors are significantly smaller than 1 which supports the application of a calibration to the data. After application of the calibration the $t\bar{t}$ scaling factors for the different mass bins scatter around the value obtained for the full mass range of 1.23 ± 0.16. The scaling factors of all mass bins are consistent within 2σ.

In future steps these scaling factors can be applied to the $t\bar{t}$ background and the impact of this on other observables in the $Z \rightarrow e^+e^-$ analysis can be studied. Also different parametrizations of the calibration and their impact on the χ^2 tests can be studied. For future studies the QCD background has be to taken into account since it is not negligible.

I would like to acknowledge Alexandre Glazov and Elena Yatsenko for answering my questions and the valuable guidance. I learned a lot and really enjoyed my stay at DESY. Many thanks!

References

[1] European Organization for Nuclear Research, CERN, http://home.web.cern.ch

[2] Massive Lepton-Pair Production in Hadron-Hadron Collisions at High Energies, *S.D. Drell, T.-M. Yan*

[3] The ATLAS Experiment, http://www.atlas.ch

[4] Measurement of the inclusive $W^{\pm}$ and Z/γ^{*} cross sections in the e and μ decay channels in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ with the ATLAS detector, *The ATLAS Collaboration*

[5] ZeeD Analysis Tool, https://wiki-zeuthen.desy.de/ATLAS/Projects/ZeeD

[6] Measurement of the Z Boson production with the ATLAS Experiment at the LHC *M. Karnevskiy*