W-like measurement of the Z boson mass using dimuon events collected in pp collisions at $\sqrt{s} = 7$ TeV

The CMS Collaboration

Abstract

The mass of the Z boson is measured using a sample of $Z \rightarrow \mu\mu$ events, where one of the two muons is removed from the event to form the $W^{\text{like}}$ candidate. This procedure provides a proof of principle and a quantitative validation of analysis techniques developed for a high-precision measurement of the W boson mass in $W \rightarrow \mu\nu$ events. The study is made on the basis of a dimuon data sample collected by CMS at $\sqrt{s} = 7$ TeV, corresponding to an integrated luminosity of 4.7 fb$^{-1}$. A set of $2 \times 10^5$ $Z \rightarrow \mu\mu$ candidates is used to extract the result. The Z mass is extracted through the $W^{\text{like}}$ lepton $p_T$, transverse mass and transverse missing energy distributions, and is compatible with the world-average value, $M_Z^{\text{PDG}} = 91187.6 \pm 2.1$ MeV. The lowest uncertainty is obtained when using the $W^{\text{like+}}$ transverse mass, for which $M_Z^{W^{\text{like+}}} = 91206 \pm 36$ (stat.) $\pm 30$ (syst.) MeV.
1 Introduction and physics motivation

The standard model (SM) quantum corrections to the mass of the W boson, $M_W$, are dominated by contributions dependent on the masses of the top quark, $M_{\text{top}}$, and of the Higgs boson, $M_H$, as well as the fine-structure constant $\alpha$ [1]. Therefore, combining precise measurements of $M_W$, $M_{\text{top}}$, and $M_H$ provides a critical test of the nature and consistency of the SM. The presently available W mass measurements lead to a world average of $M_W = 80.385 \pm 0.015$ GeV [2]. The world-average top mass is $173.34 \pm 0.76$ GeV [2], not yet including the most precise $M_{\text{top}}$ direct measurement [3], which has an uncertainty of 0.66 GeV. After the discovery of the Higgs boson, a global electroweak fit [4] predicts $M_W = 80.358 \pm 0.008$ GeV, a result with an uncertainty smaller than the combination of all direct measurements. Given the presently available accuracies on $M_{\text{top}}$ and $M_H$, the mass of the W boson should be measured with a precision of 6 MeV or better, to significantly probe the consistency of the SM [4]. Reaching such a high accuracy represents a major challenge to the CMS experiment, implying a truly outstanding understanding of the detector’s performance and limitations, to be reached via multiple analysis stages.

The analysis presented in this document constitutes a first milestone towards a high-precision measurement of the W mass with the CMS experiment, to be made using $W \rightarrow \mu\nu$ events. This analysis stage consists of the measurement of the Z boson mass using a sample of so-called $W^{\text{like}}$ events, i.e. $Z \rightarrow \mu\mu$ events where one of the two muons has been removed so as to mimic the topology of $W \rightarrow \mu\nu$ decays. This first analysis is restricted to pp collisions collected by CMS in 2011, at $\sqrt{s} = 7$ TeV. Beyond the exact value of the measured parameter, denoted as $M_{Z^{\text{like}}}$, the most important outcome of the effort reported in this document is a vastly improved understanding of the CMS detection capabilities, as represented by the systematic uncertainties related to the measurements of the muon momentum and of the missing transverse momentum of the event. This exercise represents a proof of principle, showing that the analysis procedure is reliable and thereby validating the tools and techniques that will be applied in the W boson mass measurement. The chief remaining differences in the two cases are the event selection criteria, as well as the treatment of the background and of most of the theory systematic uncertainties.

This document is structured as follows. We start by giving an overview of the analysis procedure in Section 2, and a description of the CMS detector in Section 3. In Section 4, we describe the data samples used, the respective triggers, the event selection criteria, the muon efficiencies, and the backgrounds. The CMS detector performance for the muon momentum scale and resolution is addressed in Section 5, and the performance for the charged-track-based missing transverse energy ($\not{E}_T$) in Section 6. Section 7 describes the analysis techniques and Section 8 presents the resulting systematic uncertainties. Finally, Section 9 provides the obtained results.

2 Analysis overview

The present analysis is based on the 7 TeV pp data sample collected by CMS in 2011, using a single-muon trigger, already providing statistical uncertainties similar to those of the latest $M_W$ Tevatron publications [5–8]. We use state-of-the-art Monte Carlo (MC) event generators, together with the most recent PDF sets, already including some LHC measurements. The detector effects are simulated in detail (“full simulation”) and the resulting events are processed with reconstruction algorithms including the most up-to-date calibrations, so that we can exploit the exceptionally good tracking capabilities of the CMS detector. Improved calibration techniques allow us to improve muon momentum scale and resolution uncertainties considerably in this analysis compared to previous CMS results.
The $W_{\text{like}}$ kinematics of $Z$ events are defined by removing one of the two muons from the reconstructed $Z$, thereby mimicking the (undetected) neutrino emitted in $W \to \mu\nu$ decays. The presence of the neutrino prevents the direct reconstruction of the boson invariant mass. The transverse momentum of the (charged) muon, $p_T^\mu$, can be very precisely measured but, on its own, does not encode all the available information on the boson mass. The transverse momentum of the neutrino is estimated through the $E_T$ of the event, related to $p_T^\mu$ and to the boson transverse recoil, $u_T$. The $u_T$ model for $W$ events can be calibrated by measuring the balance between the recoil and boson $p_T^Z$ in $Z$ events. A wise choice of $E_T$ definition and the large $Z$ boson statistics provide a recoil calibration with an accuracy similar to that of the Tevatron analyses.

We combine the muon and neutrino information into a $W$ “transverse mass”,

$$m_T = \sqrt{2p_T^\mu E_T (1 - \cos \Delta \phi)},$$

where $\Delta \phi$ is the angle, in the transverse plane, between the muon momentum direction and the $E_T$ direction. We use a binned likelihood maximization to independently fit the measured muon $p_T$, $E_T$, and boson $m_T$ distributions to MC templates, with $M_{Z_{\text{like}}}$ as the free parameter. During the development of the analysis, all mass fits have been performed on simulated samples.

### 3 The CMS detector

The central feature of the Compact Muon Solenoid (CMS) apparatus [9] is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the superconducting solenoid volume are a silicon pixel and strip tracker, a lead tungsten crystal electromagnetic calorimeter (ECAL), and a brass/scintillator hadron calorimeter (HCAL), each composed of a barrel and two end-cap sections. Muons are measured in gas-ionization detectors embedded in the steel flux-return yoke outside the solenoid. Extensive forward calorimetry complements the coverage provided by the barrel and end-cap detectors. Muons are measured in the pseudorapidity range $|\eta| < 2.4$, with detection planes made using three technologies: drift tubes, cathode strip chambers, and resistive plate chambers. Matching muons to tracks measured in the silicon tracker results in a relative $p_T$ resolution for muons with $20 < p_T < 100\text{ GeV}$ of 1.3–2.0% in the angular region used for this analysis. This analysis benefits from a new precise calibration of the muon momentum scale, presented in Section 5.

While most CMS analyses use the particle-flow event reconstruction [10] to measure the missing transverse momentum, in this analysis we only use charged tracks produced from the same primary vertex as the muon that triggered the event. This is discussed in Section 6.

### 4 Event samples, selection criteria, and backgrounds

The analysis is performed using a data sample collected by CMS in 2011, corresponding to an integrated luminosity of 4.7 fb$^{-1}$. The $Z \to \mu\mu$ candidates are reconstructed from an event sample collected with a trigger requiring a single isolated muon with $p_T > 24\text{ GeV}$. The background MC samples consist of $Z \to \tau\tau$, inclusive dibosons ($WW, WZ, ZZ$), $t\bar{t}$ and single top processes, as well as $W \to l\nu$. They were generated with MadGraph 5 [11] and PYTHIA 6 [12], interfaced with the CTEQ6L [13] PDF set. The signal MC sample was produced with POWHEG [14–17] and PYTHIA 8 [18] with the default tune 4C, interfaced with the NNPDF 2.3 [19] PDF set at NLO.

The event selection requires that at least one well-measured vertex is found. To mimic the
selection intended for W events, we only impose high quality requirements on the muon used for the mass measurement.

Before rejecting events on the basis of kinematic quantities, we calibrate the measured (data) and simulated (MC) muon $p_T$ using the momentum scale corrections described in Section 5, and the simulated missing transverse energy using the recoil corrections described in Section 6.

The selected events must pass the trigger requirement. The two muons must have opposite sign and the dimuon invariant mass must be above 50 GeV. The selection of $Z$ events requires both muons to be of high quality, isolated, and have $d_{xy} < 0.2$ cm, where $d_{xy}$ is the distance of closest approach between the muon and the beam line, in the transverse plane. An event enters in the positive (negative) $W$-like sample if the $\mu^+ (\mu^-)$ is matched to the trigger and fulfills the acceptance conditions $|\eta| < 0.9$, $p_T > 30$ GeV, while the $\mu^- (\mu^+)$ is only required to have $p_T > 10$ GeV and $|\eta| < 2.1$.

In view of measuring $M_Z^{W_{\text{like}}}$, the analysis only uses the transverse recoil of the boson and the transverse component of the muon momentum. A narrow kinematic region, defined to mimic the phase space expected to be selected in the $W$ mass analysis, selects the final sample of $W$-like events: $30 < p_T^{\mu} < 55$ GeV, $30 < E_T^{W_{\text{like}}} < 55$ GeV, $60 < m_T(p_T^{\mu},E_T^{W_{\text{like}}}) < 100$ GeV, $|\not p_T^{W_{\text{like}}}| < 15$ GeV, $p_T^Z < 30$ GeV.

The muon trigger, identification, and reconstruction efficiencies have been determined with the “Tag and Probe” method [20]. Scale factors between the measured and simulated efficiencies were determined and applied as corrections to simulated data in the analysis.

The background contamination in the $W_{\text{like}}$ distributions is very small, below the per-mil level, thanks to the very clean $Z$ peak and to the low recoil cut. Therefore, the background contribution has been evaluated from MC simulation. The background estimation will become a more important part of the analysis when analyzing $W$ event samples.

Since $Z$ events with an even event number are used for the recoil calibration, as will be discussed in Section 6, the fits are performed using only $W_{\text{like}}$ events with odd event number. We select 181 985 events in the positive $W_{\text{like}}$ sample and 180 554 in the negative $W_{\text{like}}$ sample. Around 47% of the events are common among the two samples.

The $J/\psi$ and $Y(1S)$ dimuons used in the calibration of the muon momentum and resolution were collected with a trigger requiring an opposite-sign muon pair with rapidity within $|y(\mu\mu)| < 1.25$ and dimuon vertex fit $\chi^2$ probability larger than 0.5%. Furthermore, events with the two muons bending towards each other in the magnetic field were rejected. The $J/\psi$ dimuons have invariant mass within $2.8 < M < 3.35$ GeV and dimuon $p_T > 9.9$ GeV for 93% of the sample (and $p_T > 12.9$ GeV for the remainder) [21]. The $Y(1S)$ dimuons have $8.5 < M < 11.5$ GeV and $p_T > 5$ or 7 GeV, depending on the instantaneous luminosity [22]. The selection of $J/\psi$ and $Y(1S)$ events requires both muons to be of high quality, have pseudorapidity within $|\eta| < 2.4$, $p_T > 4$ GeV, and $d_{xy} < 0.2$ cm. The final calibration sample consists of 3.5 M (1 M) $J/\psi$ ($Y(1S)$) events.

5 Muon momentum scale and resolution

The muon momentum scale is the dominant experimental challenge in measuring the $W$ and $W_{\text{like}}$ masses. In the phase space of the analysis, a transverse momentum scale uncertainty of 5 MeV implies roughly a mass uncertainty of 10 MeV. This implies understanding the momentum scale of 40 GeV muons with an accuracy much better than one per mil. To achieve this
goal, the standard CMS calibration is improved upon with the procedure explained below.

The improved muon calibration is derived using the $J/\psi$ and $Y(1S)$ dimuon decays, providing a statistically independent sample from the $W_{\text{like}}$ measurement. Since the momentum range of the quarkonia samples is very different from the $W$ and $W_{\text{like}}$ ones, the challenge is to find a physically motivated model that describes the detector well in the whole range where the muon momentum measurement precision is dominated by the inner tracker measurement (below 200 GeV).

Three effects are accounted for in the muon momentum calibration to correct the curvature of the muon ($k = 1/p_T$): small variations of the magnetic field, residual misalignment effects, and imperfect modeling of the material resulting in different energy loss. The magnetic field is a multiplicative factor to the curvature ($A$) while the misalignment is an additive factor ($M$) with opposite sign for opposite muon charge. The energy loss correction is an additive term ($\epsilon$) to the muon momentum ($p$) resulting in a term that includes angular dependence. The corrected curvature, $k^c$, is calculated as

$$k^c = (A - 1)k + qM + \frac{k}{1 + k\epsilon\sin\theta},$$

where $\theta$ and $q$ are the polar angle and the charge of the muon, and the three terms correspond to the three effects described above.

The magnetic field residual angular dependence is modeled with a parabolic function in muon pseudorapidity ($A_1 + A_2\eta^2$), while for the energy loss due to the differences in the material budget description, the correction ($\epsilon$) is derived (and then applied) in 12 pseudorapidity bins. For the misalignment, the residual weak modes are described by the first terms of a Fourier series, in $\phi$, in six pseudorapidity bins. The total number of parameters in the fitting model is 44.

The calibration is implemented using a Kalman filter [23]. For each event, the dimuon invariant mass is calculated and its event-by-event uncertainty is estimated by propagating the uncertainties of the two tracks using their full covariance matrices. The event-by-event uncertainties are first corrected to correspond to the real resolution by accounting for additional multiple scattering and hit position uncertainties. The mass of each event is compared with a target mass and, using the Kalman formulation, the parameters of interest are estimated ($A$, $M$, $\epsilon$). The target mass is given by the arithmetic average of the simulated events after final state radiation and before detector effects. A full covariance matrix is also derived and used to propagate the statistical uncertainty of the calibration to the $W_{\text{like}}$ mass measurement.

Corrections are derived for both data and simulation. They have small values: $A$ differs from unity by less than 0.0005, $M$ is less than $10^{-4}$ GeV$^{-1}$, and $\epsilon$ is of the order of 4 MeV, values to be compared to the typical momentum of muons from $Z$ or $W$ decays, around 40 GeV.

The muon momentum resolution is also corrected. While not important for the $W_{\text{like}}$ measurement (dominated by the resolution of $E_T$) the muon momentum resolution needs to agree between data and simulation within 10% since large resolution differences can result in different shapes, affecting both the closure tests of the method and the muon momentum spectra in the analysis. For the resolution calibration a fit is performed using the $J/\psi$ sample, correcting the resolution for multiple scattering and hit position effects in different bins of muon pseudorapidity. This correction results in a 10% relative agreement of the resolution in data and simulation.

To estimate the closure of the calibration technique, an independent fit is implemented using
the J/ψ, Y(1S), and Z resonances, to measure the difference between the dimuon mass scales obtained in data \( \frac{m_{\text{DATA}}}{m_{\text{true}}^\text{corr}} \) and in simulation \( \frac{m_{\text{MC}}^\text{corr}}{m_{\text{true}}^\text{corr}} \), where \( m_{\text{true}} \) is the generator-level mass after final state radiation. The results are shown in Fig. 1. An agreement at the 0.2 per-mil level is achieved for the J/ψ and Υ(1S) samples, used in the calibration, and also for the Z sample, which was not used.

Figure 1: Closure of the calibration of the relative scale (data with respect to MC) for J/ψ, Y(1S), and Z dimuons, as a function of absolute pseudorapidity (left) and \( p_T \) (right) of the positive muon, after applying the calibration corrections measured with the J/ψ and Y(1S) samples.

6 Track-based recoil and missing transverse energy

The \( W \rightarrow \mu\nu \) events are characterized by the momentum of the charged muon, directly measurable in the detector, and by the undetected energy of the neutrino, which can be inferred from the \( \not{E}_T \), a global-event variable defined in the plane transverse to the beam line. The \( \not{E}_T \) is estimated from the measured muon momentum and the measured hadronic recoil,\[ \not{u}_T = -\not{E}_T - \sum \vec{p}_T^\mu. \] (3)

In principle, the hadronic recoil directly reflects the hadronic activity balancing the boson \( p_T \). In practice, however, this quantity is also influenced by other effects, such as the underlying event, multiple-parton interactions, and pileup collisions. To reach an accurate control of the \( \not{E}_T \) variable we need to find a precise and reliably-calibrated measurement of the hadronic recoil. To measure the W boson mass with an uncertainty of 10–20 MeV we need to understand the recoil with a precision of half a percent.

The recoil calibration is performed using \( Z \rightarrow \mu\mu \) events, which have a very low background contamination while having production and decay kinematics very similar to the W events. Half of the Z sample (even event number) is used for the recoil calibration and the rest for the W-like mass measurement.

The \( Z \rightarrow \mu\mu \) events do not have intrinsic \( E_T \), so that any measured \( E_T \) results from mis-measurements. In these events, Eq. 3 becomes\[ \not{u}_T = -\not{E}_T - \sum \vec{p}_T^\mu. \] (4)
where $\sum \vec{p}_T^\mu$ is the vector sum of the $p_T$ of the two muons.

To effectively study the properties of the hadronic recoil, we partially disentangle the hadronic activity recoiling against the boson $p_T$ from the other effects by projecting the recoil vector along the directions parallel ($u_\parallel$) and perpendicular ($u_\perp$) to the boson $p_T$ direction: $u_\parallel$ should be proportional to the boson $p_T$, the proportionality coefficient depending on the $E_T$ definition; $u_\perp$ is expected to be distributed around zero.

The optimal $E_T$ choice was obtained by using all the reconstructed charged tracks compatible with the primary vertex (PV), requiring longitudinal impact parameter with respect to the reconstructed primary vertex $dz(track,PV)<0.1$ cm. While this definition, called tkMET, has the drawback of only retaining 40% of the hadronic recoil probed with the more widely used pfMET variable [24], it has the advantages of exhibiting a better data-MC agreement and of being essentially insensitive to pileup. More importantly, tkMET provides — in the presence of pileup — the best discriminating power for the transverse mass Jacobian peak.

The left panel of Fig. 2 compares the resolution of $u_\perp$ for the tkMET and pfMET observables in Z events, as a function of the number of reconstructed vertices in the event, showing that tkMET is essentially insensitive to pileup. The resolution has been corrected for the corresponding recoil response [24], defined as $\langle u_\parallel \rangle / p_T$. The right panel of the same figure shows instead the performance of the tkMET for the transverse mass distribution in $W \to \mu^+ \nu$ events, when compared to different $E_T$ definitions, at generation and reconstruction level. The generated pfMET includes all stable particles within $|\eta| < 5.0$, while the generated tkMET selects the stable charged particles with $|\eta| < 2.4$.

In view of minimizing the systematic uncertainties caused by parton distributions functions (PDFs) and polarization differences when applying the calibration to W events, the recoil calibration is performed in bins of boson rapidity. The $u_\parallel$ and $u_\perp$ distributions (generically denoted by $u_i$) are modeled empirically by a sum of three Gaussians, whose parameters are polynomial functions of $p_T^Z$. The models obtained from fitting the different (data and simulated) event samples are used to derive corrections that can be used to transform the original recoil values of a
source event sample into corrected values matching the distribution of a target event sample. This is achieved using probability integral transforms of the models for the source and target distributions.

The calibration corrections are applied to the simulation using the direction of the generated boson transverse momentum to define the axis in the transverse plane needed for the definition of $u_\parallel$ and $u_\perp$.

The left panel of Fig. 3 shows one example of the fit to calibration data events using the sum of three Gaussians to the $u_\parallel$ distribution. The right panel shows the recoil distributions in data and simulation.

Figure 3: Left: Example of the fit to calibration data events using the sum of three Gaussians to the $u_\parallel$ distribution. Right: Comparison of the recoil distributions in data and simulation, and the pull distribution including both statistical and systematic experimental uncertainties (described in Section 8).

7 Analysis techniques and theory inputs

Several theory variations and MC correction factors have been implemented for the measurement of the Z boson mass using the Wlike procedure. Besides event weights related to PDF alternatives, binned distributions are also used to reweight the shape of the boson transverse momentum and polarization.

7.1 W and Z boson production

The Z boson leptonic decays can be triggered and selected at the LHC with high purity. They are used for calibration purposes and their differential cross sections provide precise information about PDFs and the production processes. The transverse momentum distribution of the Z boson can be accurately measured and used to tune non-perturbative parameters in the Monte Carlo generators. This measurement indirectly constrains the transverse momentum distribution of the W boson.

In the last years, new matrix element generators have been developed, such as POWHEG, based
on a consistent matching of next-to-leading order (NLO) QCD corrections with Parton Shower Monte Carlo (PSMC) programs, like PYTHIA 8. They have reached very high reliability and robustness and are widely employed in analyses of LHC data. In particular, in the last two years MC event generators for Drell–Yan (DY) production have been realized in the context of POWHEG, with NLO QCD and NLO electroweak (EW) corrections interfaced to QCD and QED PSMC for both single W [25] and single Z [17] production. These tools can be successfully used to predict exclusive quantities, like the W boson or lepton transverse momentum, together with the details of the full event and its recoil properties.

7.2 Event reweighting

Reweighting procedures save computing time that would otherwise be needed to produce simulated events. Without such tools, the total number of simulated events would easily exceed tens of billions. The simulated Monte Carlo events are reweighted (a posteriori) with weights \( w_i \), so that they can be used as if they had been generated with different parameters (e.g. a different \( W \) or \( Z \) mass), and/or different PDFs. The systematic uncertainties due to PDFs are estimated with a reweighting tool provided by the LHAPDF package, as described in Ref. [26]. For each event generated with a given PDF member \( S_0 \), a PDF weight for the PDF member \( S_i \) can be computed, a posteriori, from the momentum exchanged in the collision (\( Q \)) and from the flavor and fraction of proton energy of each incoming parton \( (x_j, f) \),

\[
  w_i(x_{1,f}, x_{2,f}, Q) = \frac{f(x_{1,f}, Q; S_i) f(x_{2,f}, Q; S_i)}{f(x_{1,f}, Q; S_0) f(x_{2,f}, Q; S_0)}.
\]  

A very fast procedure to emulate different boson mass hypotheses for the same simulated events is to use mass shape weights, computed as

\[
  w_i = \frac{BW(m_i, m_V^{\text{new}})}{BW(m_i, m_V^{\text{old}})},
\]

where \( m_V \) is the pole mass of the vector boson propagator and the BW functions are standard fixed-width relativistic Breit-Wigner mass distributions [2],

\[
  BW(m, m_V) \propto \frac{m^2}{(m^2 - m_V^2)^2 + m^4 \Gamma_V^2 / m_V^2},
\]

where \( \Gamma_V \) is the \( V \) boson width. The procedure has been compared with a full matrix element reweighting and no significant differences have been found.

7.3 Corrections for boson \( p_T \) and polarization of the signal sample

The PYTHIA 8 QCD tune 4C, interfaced to the POWHEG matrix element to produce the signal sample, does not describe well the boson transverse momentum. This distribution is therefore reweighted to data, in bins of 0.5 GeV, to the measured distribution in Z data events. In addition, the default settings of the POWHEG event generator show discrepancies when compared to the measured angular coefficients of DY events, as reported in Ref. [27]. The \( \cos \theta^* \) defined in the Collins–Soper frame (CS) [28] is reweighted to data, as a function of the absolute value of the Z rapidity. Since the boson \( p_T \) and angular coefficient reweightings are performed in the final fit phase space at reconstruction level, no systematic uncertainty is assigned.
Figure 4 compares data and simulation after reweighting for the Z mass and rapidity in the positive \(W^{\text{like}}\) sample, and for the Z \(p_T\) and \(\cos \theta^*\) in the CS frame in the negative \(W^{\text{like}}\) case, together with pull distributions. The uncertainty used to compute the pulls includes the statistical uncertainty of the data sample and systematic experimental uncertainties, described in Section 8.

Figure 4: Comparison between data and simulation after reweighting, for the Z mass and rapidity in the positive \(W^{\text{like}}\) sample (top), and the Z \(p_T\) and \(\cos \theta^*\) in the CS frame for the negative \(W^{\text{like}}\) case (bottom), together with pull distributions. The uncertainty used to compute the pulls includes the statistical uncertainty of the data sample and systematic experimental uncertainties, described in Section 8.

### 7.4 Mass fits and statistical treatment

The \(W^{\text{like}}\) fits are performed in the ranges 32–45 GeV (lepton \(p_T\) and \(E_T\)) and 65–100 GeV (\(m_T\)), scaled by the ratio \(m_{Z}^{\text{PDG}} / m_{W}^{\text{PDG}} = 1.134\) to retain a phase space similar to that intended for the W mass fits.
All the fits involving mass measurements are performed with a binned-template likelihood-ratio fitting procedure. The distribution to be fitted is generated as a discrete function of the fit parameter, using simulated events. These simulated distributions are referred to as “templates”. For each value of the fit parameter, the simulated distribution is compared to the data distribution and the logarithm of a binned likelihood \( \ln L \) is calculated. The best-fit value of the parameter maximizes the likelihood (or minimizes \( -\frac{1}{2} \ln L / L_0 \), where \( L_0 \) is the reference likelihood, defined as the minimum likelihood), and the \( \pm 1\sigma \) confidence intervals are defined by the increase of \( -\ln L \) by one unit. The approximation for \( \ln n! \) only affects the shape of the likelihood around the minimum and not the position of the minimum. The procedure was validated by fitting simulated data (“pseudoexperiments”). We reduce the effect of finite template statistics by fitting \( -\ln L \) to a parabola, and extracting the best-fit value and the confidence intervals using this parabola.

We independently fit the measured \( m_T, p_T, \) and \( \not{E}_T \) distributions to a sum of background and simulated signal templates, fixing the normalization of the sum to the number of data events. The fit minimizes the negative log likelihood as a function of the template parameter \( M^{\text{W like}}_W \). The likelihood is calculated in \( M^{\text{W like}}_Z \) steps of 10 MeV for MC studies, while 4 MeV steps are used in data. To check that the likelihood fit is unbiased, a closure test is performed on simulated events and no bias is observed. Another test has been carried out to check that the statistical precision of the fit scales with the square root of the number of collected events, as expected.

The statistical correlation between the different fitting variables used in the analysis has been estimated with a non-parametric technique known as “jackknife delete-d resampling” [29, 30]. The MC events have been used to build 2000 pseudo-samples, with which the values of the \( \text{W like} \) mass are computed. The corresponding correlation matrix is reported in Table 1. The off-diagonal terms have an absolute uncertainty of about 5%.

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Lepton transverse momentum ((p_T))</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Transverse mass ((m_T))</td>
<td>0.67</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>3. Missing transverse energy ((\not{E}_T))</td>
<td>0.34</td>
<td>0.70</td>
<td>1.00</td>
</tr>
</tbody>
</table>

As expected, the transverse mass is highly correlated with both lepton transverse momentum and missing transverse energy, while \( p_T \) and \( \not{E}_T \) are practically uncorrelated.

As described in Section 4, the \( \text{W like} \) positive and negative event samples have about half of their events in common, naturally leading to the effect of the statistical uncertainties for the \( \text{W like} \) negative and positive measurements to be partly correlated. This effect will not be present in the \( W \) case by construction as the \( W^+ \) and \( W^- \) samples will be independent.

# 8 Systematic uncertainties

The systematic uncertainty associated with the modeling of the muon efficiencies is evaluated assuming uncorrelated bin-by-bin statistical uncertainties and 1% systematic uncertainties of the “Tag and Probe” fits.

Two sources of systematic uncertainties are considered for the calibration of the lepton energy scale and resolution. The first is the deviation from perfect closure in Fig. 1. The second is the statistical uncertainty of the calibration sample, where the 44 parameters of the covariance
matrix are diagonalized and varied individually, both for data and MC corrections. The two uncertainties are then summed in quadrature.

Two sources of systematic uncertainties for the mass fits are associated with the recoil corrections. The first is the propagation of the statistical uncertainty of the recoil fits due to the limited statistical accuracy of the calibration sample. The fit covariance matrices for the 21 parameters on $u_\parallel$ and the 15 on $u_\perp$ are diagonalized, and the effect of each eigenvector variation is evaluated. The most significant contribution to this uncertainty is the statistical uncertainty of the data sample. The second uncertainty reflects the deviation from the perfect closure of the calibration fits. This is estimated with an alternative model based on an adaptive kernel estimation probability density function. A third uncertainty reflecting the background modeling in the recoil fit model is evaluated by using recoil corrections obtained without considering the background component and is found to be negligible.

The associated PDF uncertainties are evaluated with the NNPDF 2.3 at NLO set, through a MC-like approach: we test all 100 NNPDF members and compute the standard deviation.

The systematic uncertainty associated with the QED modeling is evaluated by comparing the templates obtained by reweighting the invariant mass distributions with different configurations at generator level, in the full phase space, after final state radiation. The central choice is $\text{POWHEG}$ NLO EW+QCD interfaced to $\text{PYTHIA}$ 8 for both QCD and QED showers, while the alternative configuration is obtained by switching off the NLO EW contribution.

The correction factors used to reweight the simulation to data, described in Section 7.3, have been independently estimated on the events with odd and even event number. The differences in the fit results have been assigned as a systematic uncertainty.

The expected uncertainties are collected in Table 2, symmetrizing the largest value between the $\pm 1\sigma$ variations.

Table 2: Uncertainties on $M_{W_{\text{like}}}^+$, in MeV, obtained with the three fitting variables. Both the $M_{W_{\text{like}}}^+$ and $M_{W_{\text{like}}}^-$ cases are reported.

<table>
<thead>
<tr>
<th>Sources of uncertainty</th>
<th>$p_T$</th>
<th>$m_T$</th>
<th>$E_T$</th>
<th>$p_T$</th>
<th>$m_T$</th>
<th>$E_T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lepton efficiencies</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Lepton calibration</td>
<td>14</td>
<td>13</td>
<td>14</td>
<td>12</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>Recoil calibration</td>
<td>0</td>
<td>9</td>
<td>13</td>
<td>0</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>Total experimental syst. uncertainties</td>
<td>14</td>
<td>17</td>
<td>19</td>
<td>12</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>Alternative data reweightings</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>14</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>PDF uncertainties</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>QED radiation</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>23</td>
<td>23</td>
<td>24</td>
</tr>
<tr>
<td>Simulated sample size</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Total other syst. uncertainties</td>
<td>24</td>
<td>25</td>
<td>27</td>
<td>28</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td>Total systematic uncertainties</td>
<td>28</td>
<td>30</td>
<td>32</td>
<td>30</td>
<td>32</td>
<td>34</td>
</tr>
<tr>
<td>Statistics of the data sample</td>
<td>40</td>
<td>36</td>
<td>46</td>
<td>39</td>
<td>35</td>
<td>45</td>
</tr>
<tr>
<td>Total stat.+syst.</td>
<td>49</td>
<td>47</td>
<td>56</td>
<td>50</td>
<td>48</td>
<td>57</td>
</tr>
</tbody>
</table>

9 Results

The comparison between data and simulation is reported in Fig. 5 for the lepton $p_T$, transverse mass, and $E_T$, in the positive and negative $W_{\text{like}}$ cases, for the corresponding best fit mass
values, together with pull distributions. The uncertainty used to compute the pulls includes the statistical uncertainty of the data sample and systematic experimental uncertainties, described in Section 8.

The results of the fits to the data are shown in Fig. 6, where the experimental uncertainties (on the recoil and the lepton) are quoted separately from the others. The three observables ($p_T$, $m_T$, and $E_T$) are correlated to each other (as shown in Table 1) and 47% of the events are included in both the positive and negative event samples, so that each of the six measurements can only be considered individually. The observable expected to provide the most precise measurement of the W mass [7, 8, 31], $m_T$, provides a fitted mass that differs from the PDG value by 18 MeV or $-20$ MeV, if we use the positive or negative event samples, respectively. These differences are well within the uncertainties affecting the present results.

The systematic uncertainties on the lepton momentum and recoil calibrations reflect the present status of the calibrations and may improve in the future by refining the calibration models. The uncertainty related to PDF reflects the knowledge of the parton densities relevant for $Z$ production, that benefits from precise measurements of the $Z$ transverse momentum and rapidity. This uncertainty will be different, and presumably larger, when evaluated for the measurement of the W mass. A first attempt to decompose the various sources of PDF uncertainties has been carried out by the ATLAS collaboration [32], identifying the W polarization and charm-initiated processes as the most relevant.

As discussed in Section 7.1, this analysis benefits from reweighting the $Z$ simulated sample with the measured distributions of the boson transverse momentum, rapidity, and decay angle. In this way several uncertainties can be safely neglected. On the other hand, when performing the measurement of the W boson mass, the modeling of such aspects will require an extrapolation from the measurements with the $Z$ boson to the expectations for the W boson. While, in general, single uncertainties can be easily extrapolated, for instance the different initial states which are accounted for in the Monte Carlo and PDFs, the full extrapolation is a challenging theoretical issue. The major difficulty arises from correlations among PDF, boson $p_T$ and polarization, and the underlying event, which imply correlated systematic uncertainties in the evaluation of the PDF uncertainties, the systematics of the matrix element, the resummed parts of the calculations, and the parton shower model. The statistical uncertainty of the simulated sample size plays a subdominant role in this context, while it will become more critical for the W mass analysis, given the larger cross section.

In this analysis the largest systematic uncertainty arises from the QED modeling, which is evaluated very conservatively switching on and off the NLO EW contributions in POWHEG. Preliminary studies comparing the QED radiation emitted by the PYTHIA 8 and PHOTOS [33] showers, suggest that PHOTOS would be preferable to model the photon emissions occurring at low angle with respect to the leptons [34]. These studies will also assess the impact of residual effects not yet included in the calculations, motivating, in the longer term, a less conservative choice to estimate this systematic uncertainty.

10 Summary

We have shown that the muon $p_T$ and the recoil can be calibrated with a precision suitable to pursue an accurate measurement of the W mass at the LHC, even in the presence of large pileup. As a proof of principle, the analysis technique has been used to measure the mass of the $Z$ boson after removing one of its decay muons, a fundamental step towards the W mass measurement.
Figure 5: Comparison between data and simulation for the lepton $p_T$ (top), transverse mass (middle), and $E_T$ (bottom) in the positive (left) and negative (right) $W_{\text{like}}$ cases, together with pull distributions. The uncertainty used to compute the pulls includes the statistical uncertainty of the data sample and systematic experimental uncertainties, described in Section 8.
Figure 6: Difference between the fitted mass and $M_{Z}^{PDG}$ obtained with each of the three observables, together with the corresponding uncertainties, separating the statistical and experimental systematic terms. Each of the six measurements can only be considered individually because of the expected correlations among all of them, as detailed in the text.

References


