TABLE OF CONTENTS

INTRODUCTION

T his chapter
BlaBla

1.1 Physics

1.2 CMS Detector

CHARMONIA

This chapter
BlaBla

2.1 Theory

2.2 Analysis

2.3 Results

3

W BOSON PRODUCTION IN PROTON-LEAD COLLISIONS

This chapter describes the measurement of the production of W bosons in proton-lead (pPb) collisions at a nucleon-nucleon center-of-mass energy $\sqrt{s_{\mathrm{NN}}} = 8.16$ TeV with the CMS detector. As mentioned in Section 1.1, the primary processes that contribute at LHC energies to the production of W bosons are $\mathrm{u}\overline{\mathrm{d}} \to \mathrm{W}^+$ and $\mathrm{d}\overline{\mathrm{u}} \to \mathrm{W}^-$. Once produced, W bosons can decay weakly to either leptons or quarks. The semi-muonic decay ($\mathrm{W} \to \mu\nu_\mu$) is of particular interest, since muons lose negligible energy in the medium and the surrounding background is small.

The individual W^+ and W^- production rates are expected to be modified relative to proton-proton (pp) collisions, due to the different content of up and down quarks in the proton compared to the lead nuclei. The distribution of partons inside the nuclei can also be modified by parton shadowing (or depletion), which can alter the W boson yields at the LHC. Precise measurements of W production in pPb collisions can therefore provide strong constrains to the nuclear parton distribution functions (nPDF).

The analysis is currently been reviewed. Previous results have been published by the CMS collaboration on pPb data at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV collected in 2013 [1]. A similar study has been performed by the ALICE [2] and the ATLAS [3] collaborations at pPb collision energies of 5.02 TeV.

The datasets and simulated samples are described in Section 3.2. In Section **??**, the event selection is presented. The relevant backgrounds are introduced in Section **??**.

Finally, the systematic uncertainties are discussed in Section **??** and the results are detailed in Section 3.8.

3.1   Theory

3.2 Dataset and simulated samples

3.2.1 Dataset

The primary dataset used for the W analysis is PASingleMuon. The dataset contains events from pPb collisions at 8.16 TeV recorded by the CMS detector requiring at least one identified muon. The data were reconstructed with the CMS software version 8.0.30 and was thoroughly validated by the CMS collaboration. Only runs passing the data quality requirements were processed. The total integrated luminosity of the recorded data corresponds to 173.4 nb^{-1}, currently known within 5% [4].

In the first part of the pPb run (labelled as $Pb-p$), the proton was heading towards negative pseudorapidity $-\eta$, according to the CMS detector convention [5], with an energy of 6.5 TeV and colliding a lead nuclei at 2.56 TeV [6]. In the second part of the pPb run (labelled as $p-Pb$), the proton was going towards $+\eta$, still boosted compared to the lead nuclei. The integrated luminosity of the $Pb-p$ and $p-Pb$ runs were 62.6 nb^{-1} and 110.8 nb^{-1}, respectively.

3.2.2 Next-to-leading order simulations

Fully reconstructed Monte Carlo (MC) simulated samples are used to describe the signal and the different sources of Electro-Weak (EWK) backgrounds. The MC samples were generated at Next-to-Leading order (NLO) using the POsitive Weight Hardest Emission Generator (POWHEG) [7–9] version 2. To account for Quantum Chromo-Dynamic (QCD) and EWK corrections, the POWHEG-BOX packages `W_ew-BMMNP` [10] and `Z_ew-BMMNPV` [11] were used to generate the $pp \to W \to l\nu_l$ and $pp \to Z/\gamma* \to l^+l^-$ processes, respectevely. The $pp \to t\bar{t}$ was generated using the POWHEG-BOX package `hvq` [12], which is a heavy flavour quark generator at NLO QCD.

Since the data is based on pPb collisions, the simulation of both proton-proton (pp) and proton-neutron (pn) collisions is crucial. In order to simulate pPb collisions and also include nuclear modifications to the PDFs, the POWHEG framework was modified. This was done by first generating pp collisions using the CT14 parton distribution function (PDF) [13]. Then each PDF was corrected by applying the EPPS16 nuclear modification factors derived for Pb ions [14]. And finally, the u-quark and d-quark PDFs were scaled considering the number of neutrons and protons in the Pb nucleus. The nuclear modification and normalization of the quark PDFs is shown in Eq. (3.1).

$$f_{Pb}^{\mathrm{u}} = \frac{Z}{A}\left[R_s^{\mathrm{u}}f_{\mathrm{p}}^{\overline{\mathrm{u}}} + R_v^{\mathrm{u}}\left(f_{\mathrm{p}}^{\mathrm{u}} - f_{\mathrm{p}}^{\overline{\mathrm{u}}}\right)\right] + \frac{A-Z}{A}\left[R_s^{\mathrm{d}}f_{\mathrm{p}}^{\overline{\mathrm{d}}} + R_v^{\mathrm{d}}\left(f_{\mathrm{p}}^{\mathrm{d}} - f_{\mathrm{p}}^{\overline{\mathrm{d}}}\right)\right]$$

$$f_{Pb}^{\mathrm{d}} = \frac{Z}{A}\left[R_s^{\mathrm{d}}f_{\mathrm{p}}^{\overline{\mathrm{d}}} + R_v^{\mathrm{d}}\left(f_{\mathrm{p}}^{\mathrm{d}} - f_{\mathrm{p}}^{\overline{\mathrm{d}}}\right)\right] + \frac{A-Z}{A}\left[R_s^{\mathrm{u}}f_{\mathrm{p}}^{\overline{\mathrm{u}}} + R_v^{\mathrm{u}}\left(f_{\mathrm{p}}^{\mathrm{u}} - f_{\mathrm{p}}^{\overline{\mathrm{u}}}\right)\right]$$

(3.1)

where f represent the quark PDF, R_s (R_v) is the nuclear modification factor for sea (valence) quarks, A is the Pb mass number and Z is the Pb atomic number.

The parton showering is performed by hadronizing the POWHEG events in PYTHIA 8.212 [15] with the CUETP8M1 underlying event (UE) tune [16, 17]. The decay of τ particles is handled in PYTHIA using the TAUOLA C++ Interface 1.1.5 [18], including final state radiative (FSR) QED corrections using PHOTOS 2.15 [19]. The full CMS detector response is simulated in all MC samples, based on GEANT4 [20], considering a realistic aligment and calibration of the beam spot and the different sub-detectors of CMS, tuned on data. The MC events are reconstructed with the standard CMS pp reconstruction software used during 2016 data taking.

To consider a more realistic distribution of the underlying environment present in the pPb collisions, the MC signal events were embedded in a minimum bias sample generated with EPOS [21], taking into account the pPb boost direction. The EPOS MC samples were tuned to reproduce the global event properties of the pPb data such as the charged-hadron transverse momentum spectrum and the particle multiplicity [22]. The list of simulated samples and the corresponding cross sections used are summarized in Table 3.1.

3.2.3   Center of Mass Frame

Due to the energy difference between the pPb colliding beams, the nucleon-nucleon center-of-mass (CM) frame was not at rest with respect to the laboratory (LAB) frame. Massless particles emitted at a pseudorapidity η_{CM} in the CM frame experienced a longitudinal boost as described in equation 3.2.

$$|\Delta\eta_{CM}| = \frac{1}{2} \times \left|\ln\left(\frac{Z_{Pb} \times A_p}{Z_p \times A_{Pb}}\right)\right| = \frac{1}{2} \times \ln\left(\frac{208}{82}\right) = 0.465$$

(3.2)

The shift between the CM and LAB frames depends on the orientation of the beams, and are applied in the following way:

- Pb $-$ p (1st run): $\eta_{LAB} = \eta_{CM} - 0.465$

- p $-$ Pb (2nd run): $\eta_{LAB} = \eta_{CM} + 0.465$

Process	Criteria	Generator	PDF	Cross section (nb)	Events
$pPb \to W^+ \to \mu^+ \nu_\mu$		POWHEG +PYTHIA	CT14+EPPS16	1213.4 (NLO)	982714
$Pbp \to W^+ \to \mu^+ \nu_\mu$		POWHEG +PYTHIA	CT14+EPPS16	1214.1 (NLO)	981874
$pPb \to W^- \to \mu^- \bar{\nu}_\mu$		POWHEG +PYTHIA	CT14+EPPS16	1082.2 (NLO)	995726
$Pbp \to W^- \to \mu^- \bar{\nu}_\mu$		POWHEG +PYTHIA	CT14+EPPS16	1083.4 (NLO)	998908
$pPb \to W^+ \to \bar{\tau} \nu_\tau$		POWHEG +TAUOLA	CT14+EPPS16	1146.3 (NLO)	481125
$Pbp \to W^+ \to \bar{\tau} \nu_\tau$		POWHEG +TAUOLA	CT14+EPPS16	1147.4 (NLO)	500000
$pPb \to W^- \to \tau \bar{\nu}_\tau$		POWHEG +TAUOLA	CT14+EPPS16	1026.3 (NLO)	495450
$Pbp \to W^- \to \tau \bar{\nu}_\tau$		POWHEG +TAUOLA	CT14+EPPS16	1019.4 (NLO)	498092
$pPb \to Z/\gamma* \to \mu^+ \mu^-$	$10 < M_{Z/\gamma*} < 30$	POWHEG +PYTHIA	CT14+EPPS16	1182.2 (NLO)	997120
$Pbp \to Z/\gamma* \to \mu^+ \mu^-$	$10 < M_{Z/\gamma*} < 30$	POWHEG +PYTHIA	CT14+EPPS16	1168.0 (NLO)	1000000
$pPb \to Z/\gamma* \to \mu^+ \mu^-$	$30 < M_{Z/\gamma*} < \infty$	POWHEG +PYTHIA	CT14+EPPS16	266.3 (NLO)	1000000
$Pbp \to Z/\gamma* \to \mu^+ \mu^-$	$30 < M_{Z/\gamma*} < \infty$	POWHEG +PYTHIA	CT14+EPPS16	266.3 (NLO)	1000000
$Pbp \to Z/\gamma* \to \tau \bar{\tau}$	$10 < M_{Z/\gamma*} < 30$	POWHEG +TAUOLA	CT14+EPPS16	1143.7 (NLO)	464494
$Pbp \to Z/\gamma* \to \tau \bar{\tau}$	$30 < M_{Z/\gamma*} < \infty$	POWHEG +TAUOLA	CT14+EPPS16	259.4 (NLO)	498444
$pPb \to t\bar{t}$		POWHEG +PYTHIA	CT14+EPPS16	45 (CMS)	99578
$Pbp \to t\bar{t}$		POWHEG +PYTHIA	CT14+EPPS16	45 (CMS)	100000

Table 3.1: Simulated NLO samples used for the W boson measurement in pPb at 8.16 TeV. The listed cross sections are the POWHEG generator cross sections scaled by 208 (atomic number of Pb ion), except for $t\bar{t}$ cross sections which are taken from the latest measurement in pPb at 8.16 TeV by CMS [23].

3.2.4 Combining pPb runs

Since the CMS detector is symmetric with respect to the beam orientation, the $p-Pb$ and $Pb-p$ samples are merged in order to maximize the statistics of the MC simulations and the data. This is done by first flipping the sign of the pseudorapidity of particles from the $Pb-p$ sample measured in the laboratory frame, and then combine them with the events from the $p-Pb$ sample. In the case of the MC samples, the events are reweighed before merging them, so that the simulated luminosity match the integrated luminosity recorded in each proton-lead run. The combined sample is labelled as "pA" and corresponds to pPb collisions with the proton always going toward $+\eta$.

3.3   Event selection

The W boson yields are measured in the semi-muonic decay channel. The signal events, determined by the process $W \rightarrow \mu \nu_\mu$, are characterised by a high transverse momentum (p_T) muon and the presence of missing tranverse energy ($\not{E}_T$), originated from the undetected neutrino. Moreover, events with similar characteristics can also be produced by other background processes, such as multi-jet events, Drell-Yan dilepton events or EWK decays. This section explain the different selections implemented to supress the background and enhance the signal.

3.3.1   pPb global filter

In order to ensure that the samples are not contaminated by non-collision events, such as cosmics or electronic noise, a pPb Global Event Filter (GEF) is applied. The different selections included in the pPb GEF are described below:

- Primary Vertex filter: Requires a primary vertex reconstructed from at least two tracks, within a longitudinal (transverse) distance of 25 cm (2 cm) of the nominal interaction point.

- HF Coincidence filter: Requires at least one tower on each side of the interaction point in the Hadron-Forward (HF) calorimeter with an energy deposit per tower of at least 3 GeV.

- Beam-Scraping filter: Requires at least 25% of tracks in the event to be high quality tracks.

The impact of the GEF was checked both in data and simulation. Only 0.08% of events in data and 0.06% of events in W boson simulation, passing all the analysis cuts summarized in Section 3.3.6, were removed by the filter.

3.3.2   Missing Transverse Energy

Since neutrinos are not detected in the CMS detector, their presence is characterized by a particle momentum imbalance in the transverse plane, also called Missing Transverse Energy (MET).

The MET is defined as the negative vectorial sum of the transverse momemtum of all reconstructed particles. Its coordinates along the the x and y axes in the CMS coordinate system [5], can be computed as:

$$\vec{E}_{\mathrm{T}} \;\; = \;\; - \sum_{particles} \vec{p}_{\mathrm{T}}$$

$$E_{\mathrm{T}} \;\; = \;\; \left| \vec{E}_{\mathrm{T}} \right|$$

The particle-flow algorithm (PF) [24] is used to identify the particles in the event and estimate the E_{T}. The PF algorithm is designed to reconstruct all stable particles such as electrons, muons, photons, charged hadrons and neutral hadrons, by taking into account all CMS sub-detectors. The outcome is an optimal determination of each particle's type, momentum and energy. This set of particles is then used to measure the $\vec{E}_{\mathrm{T}}$. The performance of the MET reconstruction in pp data has been documented in [25, 26].

Ideally, in an event where no neutrinos are produced, the E_{T} should be equal to zero. But since the momentum of particles is not measured with perfect precision, the sum of the reconstructed particles $\vec{p}_{\mathrm{T}}$ does not cancel completely due to the resolution of the detector. In order to correct for the differences in the MET resolution between simulation and data, the resolution of the hadronic recoil component is smeared in the simulation to match the data as explained in Section 3.4.2.

3.3.3 Trigger

The muon triggers are divided in two systems: the Level-1 (L1) trigger and the High Level Trigger (HLT).

The L1 triggers are hardware-based and were updated during 2016 (LHC Run 2) using μTCA technology. The L1 muon trigger system is divided in 3 sub-systems: Endcap, Overlap and Barrel Muon Track Finders. The muon candidates found by each track finder are sent to the L1 Global Muon Trigger (μGMT), which process them and selects the best muon candidates. Afterwards, the L1 Global Trigger (μGT) takes the decision to either accept or reject the event based on the information provided by the μGMT and the calorimeters. In order to only consider real collisions, all pPb L1 trigger algorithms were required to be in coincidence with a bunch crossing identified by the Beam Pick-up Timing for eXperiments (BPTX) detector. The technical information of the upgraded L1 muon trigger system is described in detail in [27, 28].

The HLT are software-based and are only applied once an event is accepted by the L1 trigger system. The HLT muon triggers implemented for pPb data are divided in 3 levels: the `HLT_PAL1` trigger takes as input all events fired by the L1 muon trigger and does not

apply any further cuts, the `HLT_PAL2` trigger reconstructs the muon candidates found in the muon sub-detectors using a more sophisticated algorithm compared to L1, and the `HLT_PAL3` trigger uses the muon tracks reconstructed by combining the inner tracker hits and muon tracks found by `HLT_PAL2`. The HLT muon reconstruction algorithms were identical to the ones used during the 2016 pp runs. A complete description of the CMS HLT trigger system can be found in [29].

For the W boson analysis, events that passed the muon trigger `HLT_PAL3Mu12_v*` are used. This trigger requires a fully reconstructed online muon with $p_\mathrm{T} > 12$ GeV/c. The HLT was seeded by the L1 trigger path `L1_SingleMu7`, which pass events with at least one L1 muon with $p_\mathrm{T} > 7$ GeV/c. The muon trigger was unprescaled both at L1 and HLT during the entire data taking period.

If in a given event, the main analysis trigger `HLT_PAL3Mu12_v1` fired and a reconstructed muon is matched to the HLT L3 muon that fired the trigger, the reconstructed muon is considered trigger matched. The matching criteria between the reconstructed and the HLT muon is done by requiring a $\Delta R(\mu_{reco}, \mu_{HLT}) < 0.1$.

3.3.4 Muon selection

Muon candidates are identified using a tight selection optimised for muons with high p_T. The tight selection requires muon candidates to be reconstructed globally from hits in the muon stations and the tracker, been identified with the Particle-Flow algorithm [24] and pass the following criteria:

- The muon track fit has at least a $\chi^2/\mathrm{NDF} < 10$, ensuring a reasonable fit quality.

- The muon track has at least one hit in the muon detectors, making sure that the information from the tracker and the muon system is consistent.

- The muon track segments are matched to at least two muon stations, making the selection consistent with the muon trigger logic.

- The tranverse impact parameter (longitudinal distance) of the muon track is consistent with the primary vertex within 2 mm (5 mm), to reduce cosmic background and muons from decays in flight.

- The muon track has at least one hit in the pixel detector to further supress muons from decays in flight.

- The muon track includes hits in at least six tracker layers to guarantee a good p_T measurement.

Apart from the muon tight selection, muon candidates are also required to be isolated in order to reduce the contribution from multi-jet background. Muons are considered isolated if the sum of the p_T of all PF particles (excluding the muon), within a cone of $\Delta R\left(\mu, PF\right) < 0.3$, is less than 15% of the muon p_T. The muon isolation variable is defined in Eq. (3.3).

$$\text{isolation} = \left(\sum_{\substack{\text{charged hadrons}}}^{\mathrm{DR}<0.3} p_\mathrm{T} + \sum_{\substack{\text{neutral hadrons}}}^{\mathrm{DR}<0.3} p_\mathrm{T} + \sum_{\substack{\text{photons}}}^{\mathrm{DR}<0.3} p_\mathrm{T} \right) / p_\mathrm{T}^{\mu} \tag{3.3}$$

Since more than one muon candidate can be reconstructed in an event, the muon candidate with the highest p_T (i.e. the leading muon), is used. The event is only kept if the leading muon has a $p_\mathrm{T} > 25$ GeV/c and $|\eta| < 2.4$. Any difference in the performance of the muon cuts observed between simulation and data, is corrected in simulation through the use of the Tag-And-Probe scale factors described in Section 3.5.1.

3.3.5 Drell-Yan Veto

A Drell-Yan veto is applied to suppress the contribution from $Z/\gamma* \rightarrow \mu^+\mu^-$ background events. This veto removes events that contain at least two opposite sign muons with $p_\mathrm{T} > 15$ GeV/c passing the tight muon identification and the relative muon isolation cut $iso < 0.15$, defined in Section 3.3.4 .

The probability that Drell-Yan events survive the veto is checked using a Drell-Yan simulated sample. The denominator of the Drell-Yan Veto effiency is filled with the muons passing all the W boson analysis cuts (see Section 3.3.6), while the numerator is filled with the same muons as long as the event pass the Drell-Yan veto. The MC survival probability is shown in Fig. 3.1.

3.3.6 Summary of Event Selection

In summary, the W candidate selection consists of the detection of a high p_T muon, passing all identification criteria explained in Section 3.3.4. The leading muon is required to have a $p_\mathrm{T} > 25$ GeV/c, and be trigger matched (see Section 3.3.3). The QCD background is reduced by requiring the leading muon to be isolated. Since high p_T muons can also come from Drell-Yan or resonance decays, a dimuon veto is applied, to remove this potential source of background.
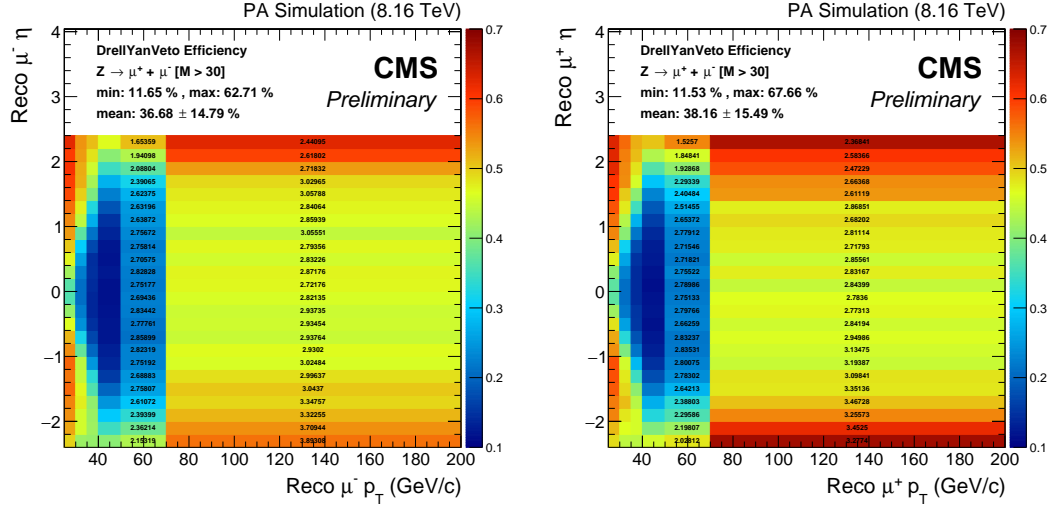
Figure 3.1: Survival probablity of single muons from Drell-Yan ($M > 30$ GeV/c^2) MC sample as a function of the reconstructed muon η and p_T, separated in negative (left) and positive (right) charged muons. The pPb and Pbp MC samples are combined as described in Section 3.2.4. The relative statistical efficiency uncertainties scaled by 100 are shown for the two highest p_T bins. Reconstructed muons are required to be within $p_T > 25$ GeV/c and $|\eta| < 2.4$, be trigger matched and pass the isolation and tight selection criteria.

The other signature of a W event is a high p_T neutrino, estimated through the $\not{E}_T$. No explicit cut is applied on $\not{E}_T$. The missing transverse energy is directly used to build templates and extract the yields by fitting the signal and background components. The conditions used to define the signal and background regions of interest are illustrated in Fig. 3.2.
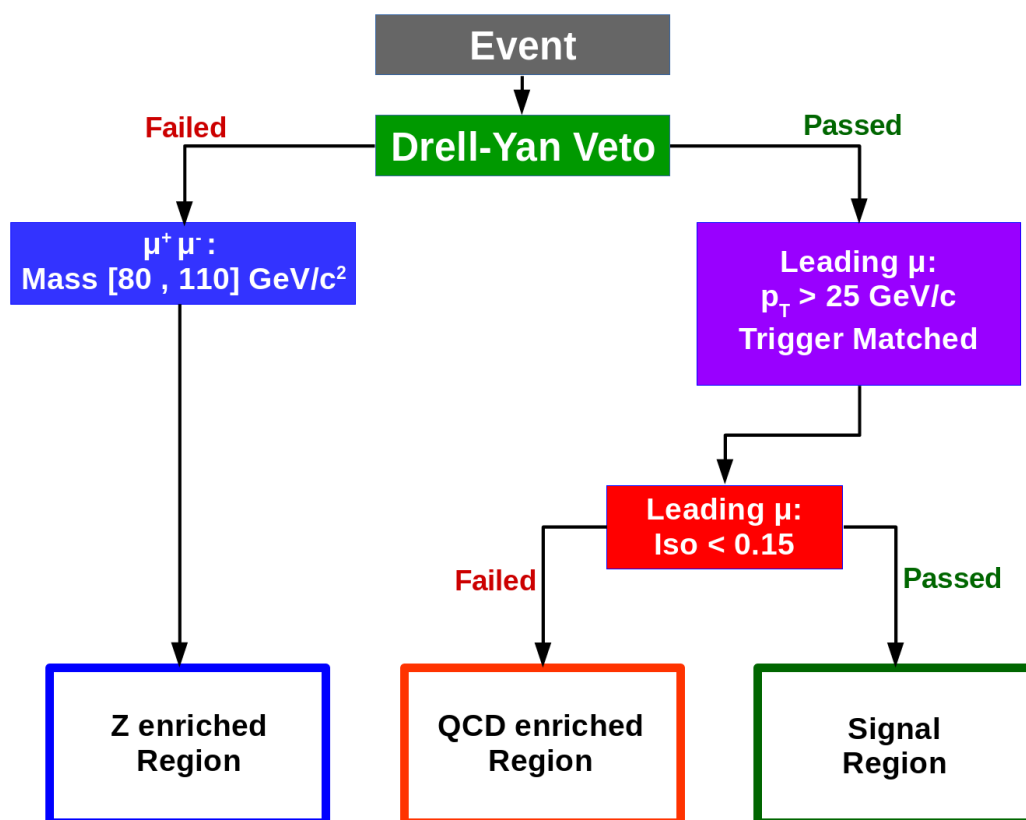
Figure 3.2: Flowchart illustrating the way the events are classified

3.4 Corrections

3.4.1 Event Activity Reweighing

The pPb underlying event (UE) activity has been checked in a $Z/\gamma* \to \mu^+\mu^-$ enriched sample of data and MC, by comparing the distribution of the number of tracks and the energy deposited in the Hadron-Forward (HF) calorimeters. The official MC samples embedded in EPOS minimum bias events are not currently able to reproduce the UE present in pPb data, as can be observed in Fig. 3.3.

Since the muon isolation and the MET are computed by summing over particles, they are correlated with the UE. Therefore, any disagreement in the modelling of the event activity can have an impact on the muon efficiency and the signal extraction. The disagreement between the simulation and the data can be caused by the presence of hard probes such as W bosons, which bias the event activity towards higher multiplicity compared to minimum bias events.

In order to correct the mismodelling of the UE, the simulated distribution of the energy deposited in both sides of the HF calorimeters is reweighed to match the data in a Z enriched region. As can be seen in Fig. 3.3, the HF reweighing significantly improves the MC-data agreement of the MET distribution for Z boson selected events.

3.4.2 Corrections for MET

As mentioned in Section 3.3.2, the MET is defined as the sum of the p_T of all the particles reconstructed with the PF algorithm, including those produced by the underlying events. Therefore, the first step to improve the simulation of the MET consist of correcting the modelling of the event activity from the EPOS minimum bias events by reweighing the HF energy distribution as explained in Section 3.4.1. The low MET region in MC agrees better with the data distribution after applying the HF energy reweighing.

The width of the MET distribution in $Z \to \mu^+\mu^-$ events represents the MET resolution due to the p_T measurement of the PF particles. Differences in the emulation of the detector conditions is one of the main sources of disagreement between simulation and data in the high MET region of a Z boson enriched sample. The simulated MET resolution describes better the data after correcting the hadronic recoil present in MC events.
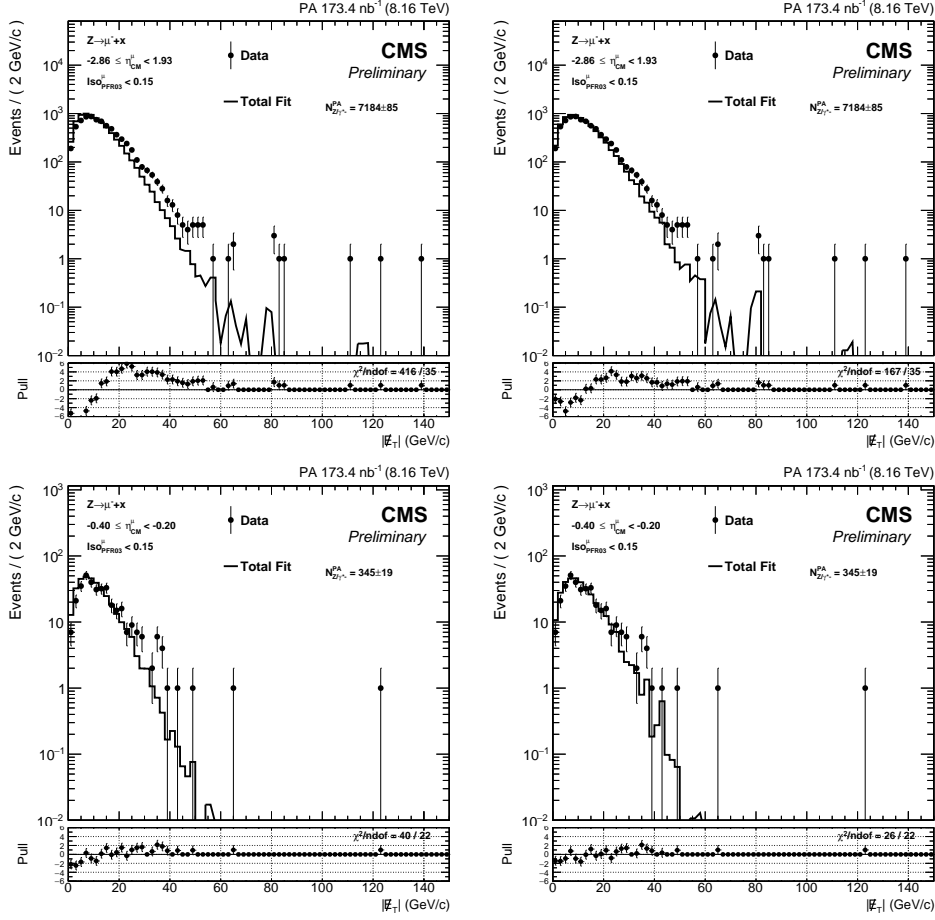
Figure 3.3: A comparison of the MET distribution in data and MC for Z boson selected events. The top plots correspond to the full pseudorapidity range in the analysis while the bottom ones correspond to an specific pseudorapidity bin. The left plots use the PF MET RAW without HF reweighting in MC, while the right ones the MC events are reweighed.

3.4.3 Definition of recoil

The hadronic recoil of $Z \to \mu^+ \mu^-$ events is defined as the vector sum of the p_T of all PF candidates excluding the muons from the boson decay. Considering the definition of MET, the recoil vector can be expressed as:

$$\vec{u}_T = -\vec{E\!\!\!/}_T - \vec{q}_T, \tag{3.4}$$

where $\vec{u}_T$ is the hadronic recoil in the transverse plane to the beam direction and $\vec{q}_T$ is the transverse momentum of the Z boson.
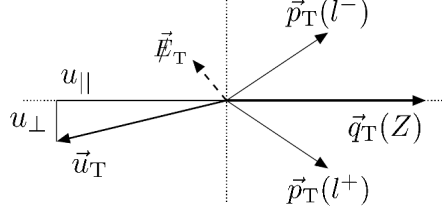
17

Figure 3.4: Definition of the recoil vector and its components for $Z \to \mu^+\mu^-$ events.

In the case of $W \to \mu\nu_\mu$ events, the recoil is defined as:

$$\vec{u}_T = -\vec{\slashed{E}}_T - \vec{p}_T^\mu,$$

(3.5)

where $\vec{p}_T^\mu$ is the p_T of the muon.

The recoil vector is projected along the transverse plane in the boson direction. The parallel and perpendicular components of the recoil vector with respect to the boson $\vec{q}_T$ are labelled as $u_\parallel$ and $u_\perp$, respectively. Fig. 3.4 describes the recoil and its components.

The boson transverse momentum vector $\vec{q}_T$ is determined in MC using the reconstructed decay products whenever possible. In the $W \to \mu\nu_\mu$ MC samples, the $\vec{q}_T$ is derived from the sum of the p_T of the reconstructed muon and the generated neutrino, while for the $W \to \tau\nu_\tau$, $Z/\gamma* \to \tau\bar{\tau}$ and $t\bar{t}$, the $\vec{q}_T$ is taken from the generated boson p_T. In the case of $Z/\gamma* \to \mu^+\mu^-$, if the sub-leading muon falls outside of the coverage of the CMS detector, the Z boson q_T is computed from the sum of the reconstructed leading muon p_T and the generated sub-leading muon p_T, otherwise the q_T is equal to the sum of both reconstructed muons p_T.

The recoil corrections are extracted using a Z control sample. In order to compute them, the parallel and perpendicular components of the recoil vector are determined event by event in data and MC, and their distributions are sorted in bins of q_T. The recoil distributions are fitted with an unbinned maximum likelihood fit. A combination of two Gaussian functions is chosen to better describe the shape of the recoil distributions. Examples of the distributions of the parallel and perpendicular recoil components, denoted as u_1 and u_2 respectively, are shown in Fig. 3.5 for data and simulation. Furthermore, the fits performed using the double Gaussian functions, and their corresponding pull distributions are also shown.
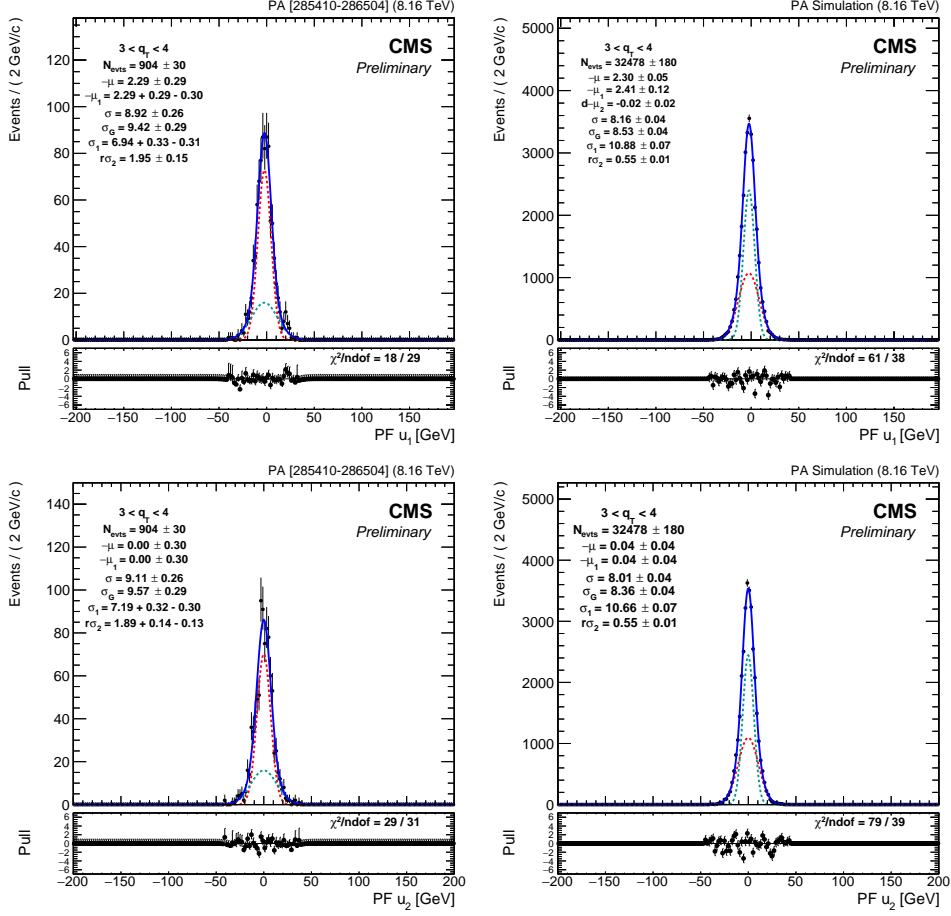
Figure 3.5: Distributions of the parallel (top) and perpendicular (bottom) components of the recoil in data (left) and MC (right). The fit function is based on a weighted sum of two Gaussian distributions. The plots correspond to the q_T bin [3, 4] GeV.

3.4.3.1 Recoil Scale

The mean pamareters (μ_1 and μ_2) of the parallel component of the recoil (u_1) are extracted in different bins of q_T by fitting the recoil distribution as shown in Fig. 3.5. Afterwards, the profile of the average recoil as a function of q_T is fitted using the following function:

$$-\mu_{1,2}(q_T) = (c_0 + c_1 q_T)(1 + erf(\alpha q_T^{\beta}))\qquad(3.6)$$

The fits of the average values of the parallel component of the recoil versus q_T are shown in Fig. 3.6. In addition, the average mean value given by $\mu = f \cdot \mu_1 + (1-f) \cdot \mu_2$, is also shown. The difference between data and the simulation as a function of q_T is used on an event by event basis to correct the scale of the recoil in MC.
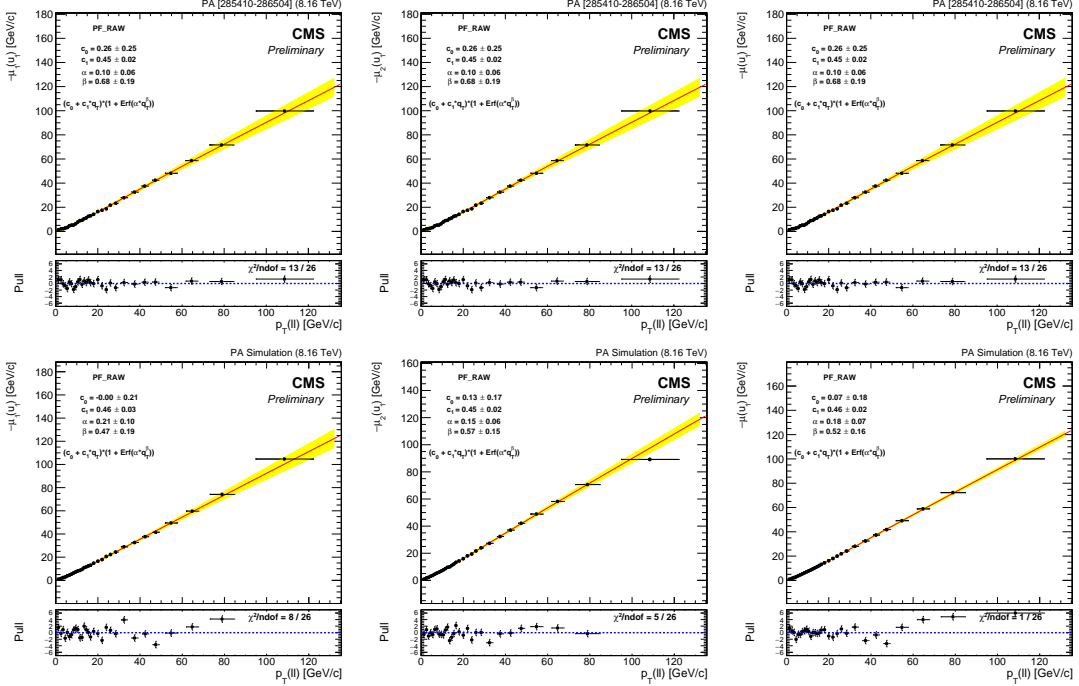
Figure 3.6: Fits for the μ_1 (left), μ_2 (middle) and weighted average μ (right) values of the parallel recoil component versus q_T. The plots on the top correspond to data while the plots in the bottom correspond to $Z \to \mu^+\mu^-$ MC.

The average values of the perpendicular component of the recoil (u_2) from data and MC Z samples are fitted with a constant function (c_0). The outcome is shown in Fig. 3.7. One observe that the average perpendicular component is consistent with zero, and so, it is set to zero in the correction procedure.

3.4.3.2 Recoil Resolution

The gaussian widths (σ_1 and σ_2) of the parallel and perpendicular distributions of the recoil are also extracted from the recoil fits for each q_T bin. The recoil resolution is parametrised as a function of q_T using the following formula:

$$\sigma_{1,2}(q_T) = \sqrt{s_0^2 + s_1^2 \cdot q_T^\alpha}, \tag{3.7}$$

The fits to the resolution profile of the parallel and perpendicular components of the recoil are presented in Fig. 3.8 and Fig. 3.9, respectively. Moreover, the distribution of the weighted average of the two gaussian widths, $\sigma = f \cdot \sigma_1 + (1-f) \cdot \sigma_2$, is also shown.
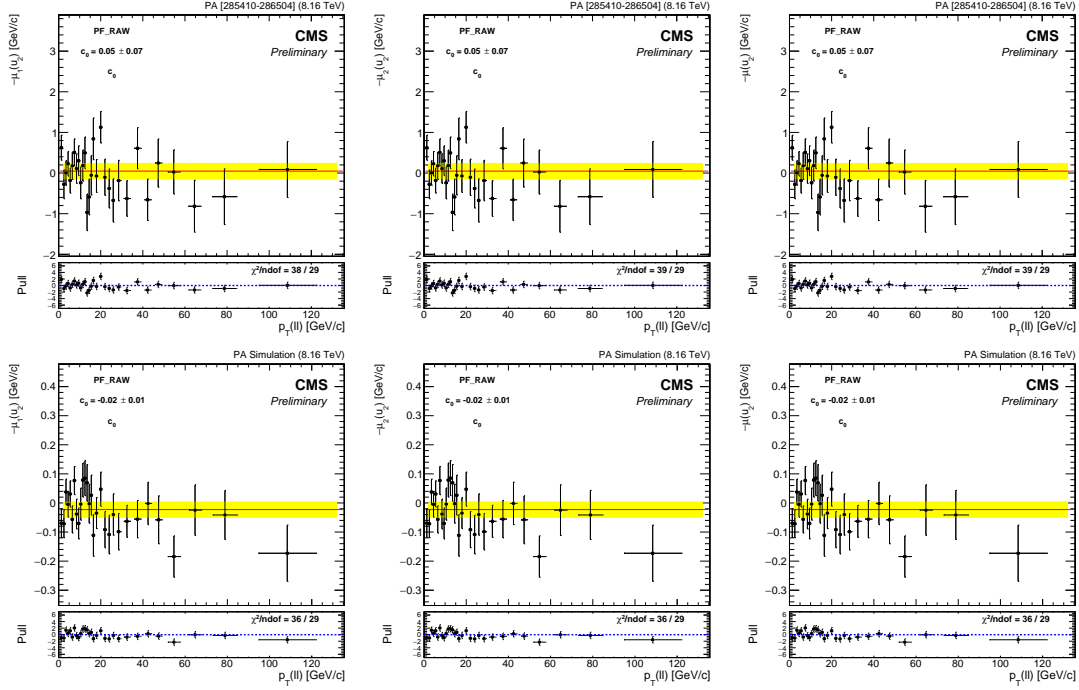
Figure 3.7: Fits for the μ_1 (left), μ_2 (middle) and weighted average μ (right) values of the perpendicular recoil component versus q_T. The plots on the top correspond to data while the plots in the bottom correspond to $Z \to \mu^+\mu^-$ MC.

3.4.3.3   Recoil Correction

The simulated recoil distribution is corrected to match the scale and resolution observed in data. This can be done in two ways:

- Smearing method : Each component of the recoil distribution is smeared event-by-event using a gaussian probability distribution function defined as:

$$u_\parallel^{corr} = Gauss(u_\parallel - \mu_\parallel^{MC}(q_T) + \mu_\parallel^{data}(q_T), \sqrt{\sigma_\parallel^{data}(q_T)^2 - \sigma_\parallel^{MC}(q_T)^2}) \quad , \qquad (3.8)$$

$$u_\perp^{corr} = Gauss(u_\perp, \sqrt{\sigma_\perp^{data}(q_T)^2 - \sigma_\perp^{MC}(q_T)^2}) \quad , \qquad (3.9)$$

where μ and σ are the weighted average values obtained in the previous subsections.

- Scaling method : The parallel and perpendicular components of the simulated recoil are scaled in two possible ways:
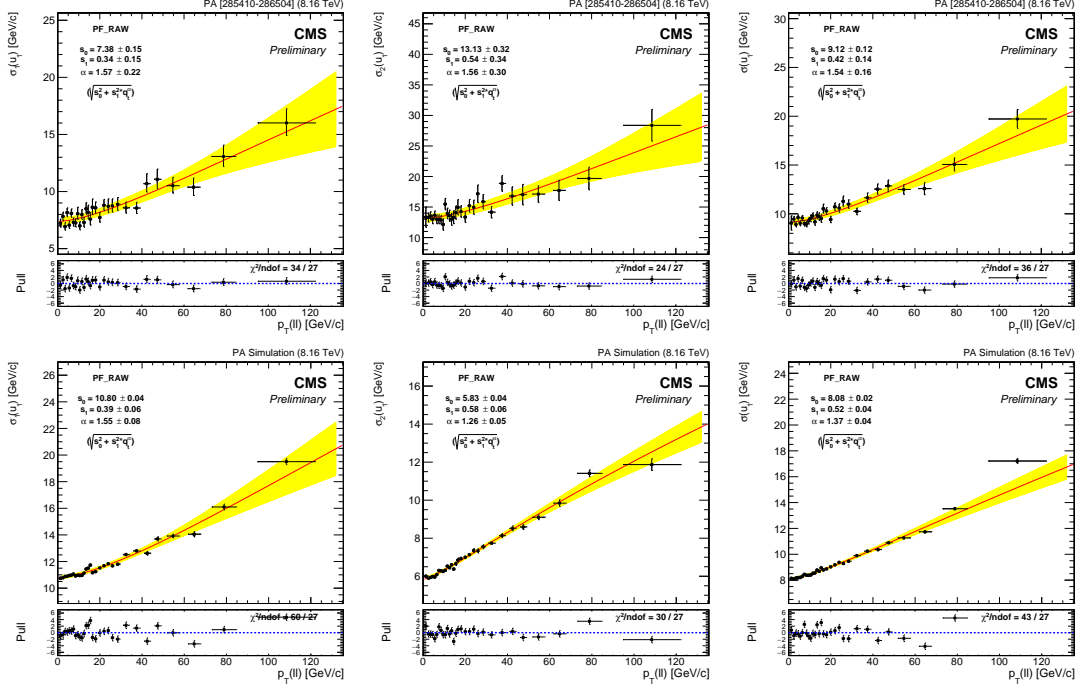
Figure 3.8: Fits for the σ_1 (left), σ_2 (middle) and weighted average σ (right) values of the parallel recoil component versus q_T. The plots on the top correspond to data while the plots in the bottom correspond to $Z \to \mu^+ \mu^-$ MC.

– General case: Considering the full information of the double gaussian PDF used to fit the recoil distributions in data and MC, by using their Cumulative Distribution Functions (CDF).

$$u_{\parallel,\perp}^{corr} = CDF_{data}^{-1}[CDF_{MC}[u_{\parallel,\perp}^{MC}]] \tag{3.10}$$

– Gaussian case: Approximating the recoil distributions by assuming a single gaussian PDF using the weighted average values μ and σ derived before.

$$u_{\parallel}^{corr} = (u_{\parallel} - \mu_{\parallel}^{MC}(q_T))\frac{\sigma_{\parallel}^{data}(q_T)}{\sigma_{\parallel}^{MC}(q_T)} + \mu_{\parallel}^{data}(q_T), \tag{3.11}$$

$$u_{\perp}^{corr} = u_{\perp}\frac{\mu_{\perp}^{data}(q_T)}{\sigma_{\perp}^{MC}(q_T)} \tag{3.12}$$

Once the recoil is corrected, the MET distribution is recalculated as:

$$\text{MET}_{corr} = \left|\vec{p}_T^{\mu} + \vec{u}_{corr}\right| \tag{3.13}$$
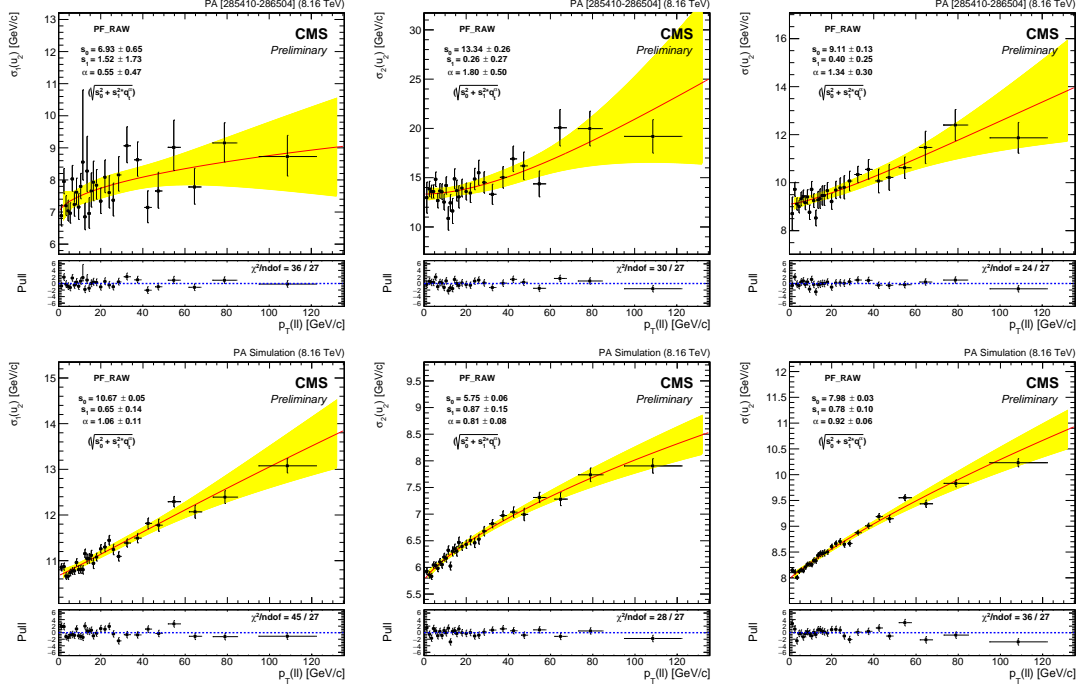
Figure 3.9: Fits for the σ_1 (left), σ_2 (middle) and weighted average σ (right) values of the recoil perpendicular component versus q_T. The plots on the top correspond to data while the plots in the bottom correspond to $Z \to \mu^+ \mu^-$ MC.

3.4.3.4 Recoil Correction: Control Region

The recoil corrections are checked by applying them on the same MC sample used to derive the recoil corrections (Z boson enriched). The data and simulation (including HF reweighing) corrected MET distributions are shown in Fig. 3.10 in the inclusive (left) and in a selected (right) pseudorapidity region. As can be observed, the agreement between simulation and data is greatly improved after applying the recoil correction using the gaussian scaling method.

3.4.3.5 Recoil Correction: Signal Region

The W signal in data is extracted following the procedure in Section 3.6. The recoil corrections are applied to the W boson MC using the 3 methods detailed in Section 3.4.3.3, in order to determine which one works better. Fig. 3.4.3.5 shows the data fits in the W region with the scaling method in the general case (left), gaussian case (middle) and the smearing method (right), for the full pseudorapidity region used in the analysis. The gaussian scaling provides the best fits, so it is used as nominal correction method.
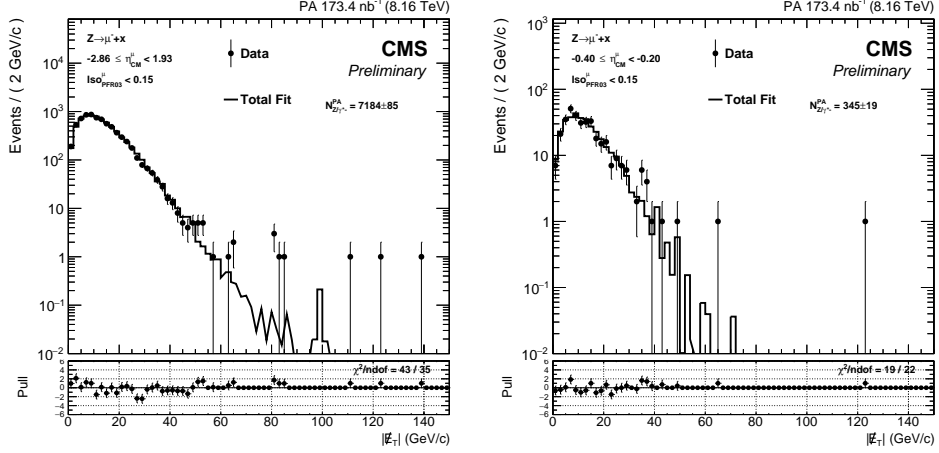
Figure 3.10: A comparison of the MET distribution in data and MC for Z boson selected events. The left plot corresponds to the full pseudorapidity range in the analysis while the right one corresponds to an pecific pseudorapidity bin. The recoil correction using the gaussian scaling method has been applied to MC.
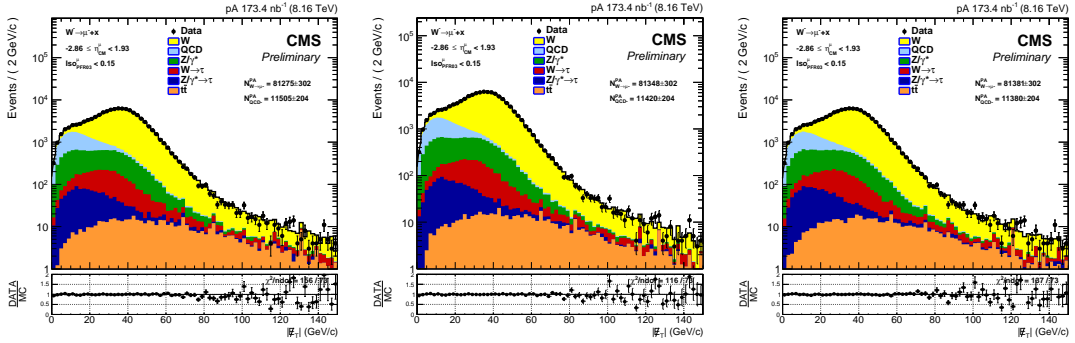

Figure 3.11: A comparison of the MET distribution in data and MC for W boson selected events in the full pseudorapidity range in the analysis. Different methods to apply the recoil corrections in MC are used in each plot. The left plot corresponds to the scaling method in the general case, the middle one corresponds to the scaling method in the gaussian case, and the right one to the smearing method.

3.5 Signal efficiency

The signal truth efficiency is estimated using the $W \rightarrow \mu\nu_\mu$ NLO MC samples since they contain the full history of the events, including the generation and reconstruction of the particles. The energy deposited in the HF calorimeters is reweighed in MC to match the distribution observed in data as explained in Section 3.4.1.

A reconstructed muon is considered an offline muon if it passes all the W boson

analysis cuts detailed in Section 3.3.6. Among the selection criteria, an offline muon is required to pass the isolation and tight identification cuts defined in Section 3.3.4, be trigger matched, and have a $p_T > 25$ GeV/c and $|\eta| < 2.4$.

The muon truth efficiency is defined as the fraction of generated muons matched to an offline muon around a cone of $\Delta R < 0.05$. All the generated muons are required to be inside the analysis kinematic region ($p_T > 25$ GeV/c and $|\eta| < 2.4$) and come from a W boson decay. The muon truth efficiency is described in Eq. (3.14).

$$\epsilon^{\mu^{\pm}}_{offline}(p_T^{gen}, \eta^{gen}) = \frac{N^{\mu^{\pm}}_{gen, p_T > 25}(p_T^{gen}, \eta^{gen})[\text{Matched to } \mu^{\pm}_{offline}]}{N^{\mu^{\pm}}_{gen, p_T > 25}(p_T^{gen}, \eta^{gen})} \tag{3.14}$$

The muon efficiencies of the p−Pb and Pb−p MC samples are calculated individually and later combined following the strategy described in Section 3.2.4. After flipping the sign of the muon η_{LAB} in the Pb−p sample, the combined efficiency is determined as presented in Eq. (3.15), by taking into account the corresponding recorded integrated luminosities of each run (see Section 3.2.1). The combined efficiency is labelled as pA muon efficiency.

$$\epsilon^{\mu^{\pm}}_{PA}(p_T^{gen}, \eta^{gen}) = \frac{\sigma^{MC}_{pPb} \times \mathscr{L}_{pPb} \times \epsilon^{\mu^{\pm}}_{pPb}(p_T^{gen}, \eta^{gen}) + \sigma^{MC}_{Pbp} \times \mathscr{L}_{Pbp} \times \epsilon^{\mu^{\pm}}_{Pbp}(p_T^{gen}, -\eta^{gen})}{\sigma^{MC}_{pPb} \times \mathscr{L}_{pPb} + \sigma^{MC}_{Pbp} \times \mathscr{L}_{Pbp}} \tag{3.15}$$

The statistical uncertainty of the muon efficiencies is determined using the ROOT class TEfficiency [30]. The uncertainty is estimated by applying Bayesian statistics using a Jeffrey's prior and considering a 68% credible interval. The results of the $W \to \mu\nu_\mu$ truth efficiency extracted from the combined pA MC samples are shown in Fig. 3.12.

3.5.1 Corrected muon efficiency

The simulation of the CMS detector is very precise but still far from fully describing all the detector conditions observed in real data. In order to compensate for the imperfections in the simulations, the muon MC efficiency is corrected by applying a set of muon scale factors derived using the Tag-and-Probe method [31].

The Tag-and-Probe (TnP) method is a data-driven technique widely used to compute efficiencies of physical objects, such as muons, produced from known mass resonances (e.g. J/ψ, Z boson). One advantage of the TnP method is that it can be applied to both MC and real data, allowing to asses the differences between the data and MC efficiencies. For high p_T muons ($p_T > 15$ GeV/c), the $Z \to \mu^+\mu^-$ decays are used to build a clean sample.

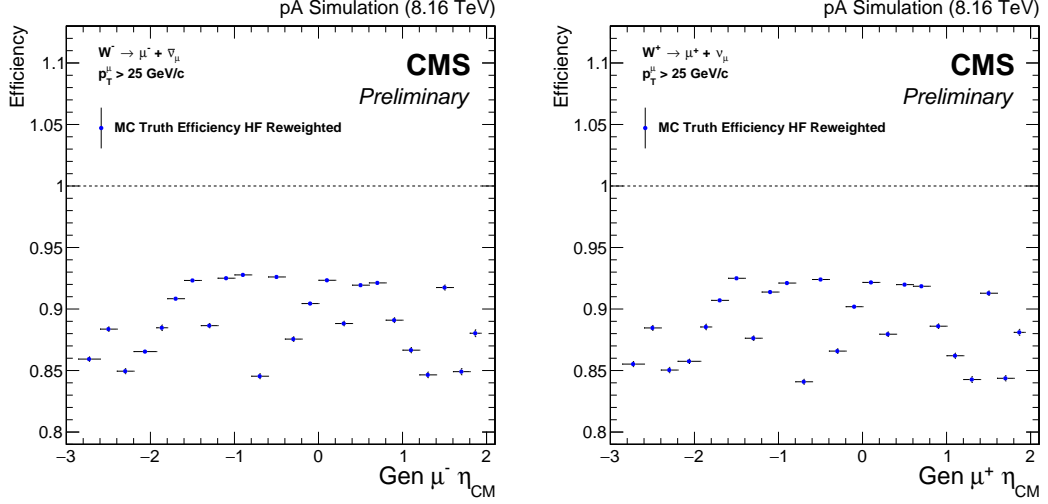Figure 3.12: Truth efficiency derived from $W \to \mu\nu_\mu$ NLO MC sample as a function of the generated muon η_{CM}, separated in negative (left) and positive (right) charged muons. The event activity of the MC samples have been reweighed. Plots corresponds to Eq. (3.14), and the p−Pb and Pb−p MC efficiencies are combined according to Eq. (3.15). Only MC statistical errors are included.

In each event, one muon is classified as the "tag" if it has a $p_T > 15$ GeV/c and passes all the analysis cuts (trigger matching, isolation and identification), ensuring that it is a real muon, while the remaining muon is labelled as the "probe". Subsequently, the invariant mass distribution of the tag-probe dimuons is fitted in three cases: all pairs, failed pairs and passing pairs, depending on wether the probe muon passes or fails the selection criteria. The efficiency is calculated as a function of the probe p_T and η_{LAB} in the laboratory frame by dividing the extracted yields of the passing pairs over all the pairs for various kinematic ranges of the probe. Finally, the TnP efficiencies in data and MC are compared, and the ratio of the two efficiencies is used to correct the simulations. More information about the pPb TnP scale factors can be found in [32].

In this case, the muon efficiencies derived in Section 3.5 are corrected by applying event by event the TnP scale factors as a function of muon η_{LAB} and p_T. The statistical and systematic components of the TnP correction uncertainties are derived by applying the following set of variations:

1. TnP statistical uncertainty:

   a) For muon ID and isolation: 100 scale factor variations derived from toy MC.

b) For trigger: Up and down $1\sigma_{stat}$ variations.

2. TnP systematic uncertainty:

   a) For muon ID, isolation and trigger: Up and down by $1\sigma_{syst}$ variations.

   b) For muon ID and isolation: Using the scale factor from the Data/MC bins instead of the one derived from fitting the ratio.

   c) For isolation: An uncertainty of **0.34%** to account for the impact on the isolation efficiency of the different level of activity and pile-up between data and MC.

   d) For stand-alone muon reconstruction: An uncertainty of **0.6%** to account for mismodelings in the STA efficiency.

The uncertainties on the muon efficiency due to the TnP corrections are calculated in three different ways, as explained below.

1. Taking the RMS between the 100 efficiencies corrected using the toy MC scale factors. Used in 1.a.

2. Taking the maximum difference between the efficiency corrected using the nominal and the varied TnP scale factors. Used in 1.b, 2.a and 2.b.

3. Applying the TnP scale factor uncertainty as a relative uncertainty on the corrected efficiency. Used in 2.c and 2.d.

The total TnP uncertainty is obtained by summing in quadrature all the statistical and systematic TnP uncertainties. The TnP corrected efficiencies including their total uncertainties are shown in Fig. 3.13.

3.6 Signal extraction

The signal and background yields are extracted by fitting the E_T distribution in data. The main background sources come from QCD and electroweak processes. The QCD background includes events where a high p_T muon is produced from a semi-leptonic decay of quarks (multi-jet background). Most of these muons come from b quark decays and light meson (pion and kaon) decays in flight. The dominant electroweak processes are the $Z/\gamma* \to \mu^+\mu^-$ (Drell-Yan) and $W \to \tau\nu_\tau$ as mentioned in Section 3.6.2. The shape
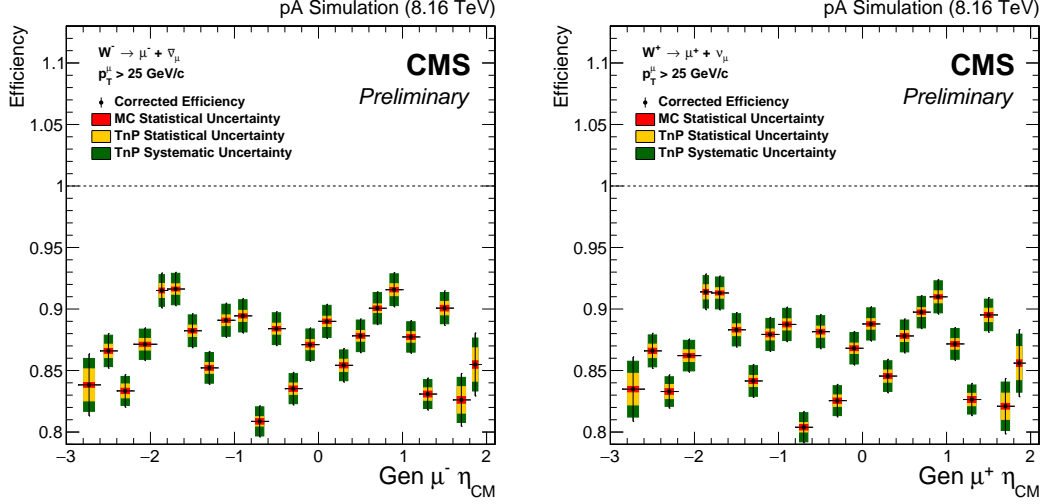
Figure 3.13: Muon corrected efficiency derived from $W \to \mu\nu_\mu$ POWHEG MC sample as a function of the generated muon η_{CM}, separated in negative (left) and positive (right) charged muons. The muon efficiency has been corrected by applying the Tag and Probe scale factors event by event. The red, yellow and green boxes represents the uncertainty on the efficiency due to the MC statistics, TnP statistics and TnP systematics, respectively. The $p-Pb$ and $Pb-p$ MC efficiencies are combined according to Eq. (3.15).

of the signal and background is estimated by using MET templates to describe the signal and the electroweak backgrounds, and a MET functional form derived from data to describe the multi-jet background as explained in Section 3.6.1.

The simulated samples of $Z/\gamma* \to \mu^+\mu^-$, $Z/\gamma* \to \tau\overline{\tau}$, $W \to \mu\nu_\mu$ and $W \to \tau\nu_\tau$ are normalized to the total integrated luminosity recorded in data using the NLO cross sections derived from the POWHEG MC samples scaled by the number of nucleons in the Pb ion. The normalization is done by reweighing all MC events by a global factor defined as $w_{MC} = (\sigma \times \mathcal{L}_{data})/N_{gen}$. In the case of the $t\bar{t}$ MC, the samples are normalized using $\sigma_{t\bar{t}} = 45$ nb^{-1}, which corresponds to the total cross section measured in pPb collisions at 8.16 TeV recently published by CMS [23].

Several corrections are applied to the MC samples before producing the templates, in order to improve the description of the data. First, the event activity in the simulations is reweighed using the energy deposited in the HF calorimeters as explained in Section 3.4.1. Afterwards, since the muon η_{CM} distribution is used to bin the data, the MC muon efficiency is corrected by applying the nominal TnP scale factors event by event, as mentioned in Section 3.5.1. And finally, the resolution of the hadronic recoil component of the MET in MC is corrected as shown in Section 3.4.3, improving the agreement of

the MET distribution between data and MC. Once the signal and the EWK MC samples are fully corrected, the MC templates are extracted in each muon η_{CM} bin, by creating a histogram of the MET distribution in bins of 2 GeV/c.

3.6.1 QCD background

The QCD background is described using a data-driven method. The functional form of the multi-jet $\not{E}_T$ distribution is derived by fitting the MET in a region dominated by non-isolated muons selected after inverting the nominal muon isolation cut mentioned in Section 3.3.4. All analysis cuts are applied except for the muon isolation cut. The general strategy is to first determine the dependence of the QCD functional form with respect to the muon isolation and then extrapolate the QCD shape down to low muon isolation values (signal region).

The nominal QCD $\not{E}_T$ functional form is based on a modified version of the Rayleigh distribution as shown in Eq. (3.16).

$$f_{QCD}(x) = x e^{-x^2 / \left(2 \left(\sigma_0 + \sigma_1 \tilde{x} + \sigma_2 \left(2\tilde{x}^2 - 1 \right) \right)^2 \right)} \quad , \text{ where } \quad \tilde{x} = (x/50) - 1 \qquad (3.16)$$

The free parameters are σ_0 , σ_1 and σ_2. The fit parameters have been arranged following a Chebyshev polynomial of second order, to reduce the correlation between parameters and improve the stability of the fits.

In order to derive the muon isolation dependence of the QCD parameters, the $\not{E}_T$ distribution function is fitted over the following five muon isolation bins : [0.4 , 0.5 , 0.6 , 0.7 , 0.8 , 0.9]. The normalization of the QCD background is left free in the fits to the data.

The bins with lower muon isolation values (iso < 0.4) are discarded due to the large contamination from EWK processes. The dependence of the three parameters σ_0 , σ_1 and σ_2 with the muon isolation variable is fitted by a linear function. The results of the linear fits are then used to extrapolate the shape of the QCD background to the signal region (average muon isolation of 0.03), as shown in Fig. 3.14. The values of the parameters corresponding to this extrapolation are given in Table 3.2.

The dependence on η_{CM} of the QCD shape has been studied. The extrapolated results do not vary significantly with respect to η_{CM} and have been found to be consistent with the η-inclusive results, within the fit uncertainties. Therefore, the η-inclusive extrapolated results for each muon charge are used to fix the QCD background shape when fitting the signal.

29

Parameter	$QCD \to \mu^-$	$QCD \to \mu^+$
σ_0	14.59±0.18	14.72±0.17
σ_1	6.30±0.24	6.76±0.23
σ_2	0.49±0.15	0.47±0.13

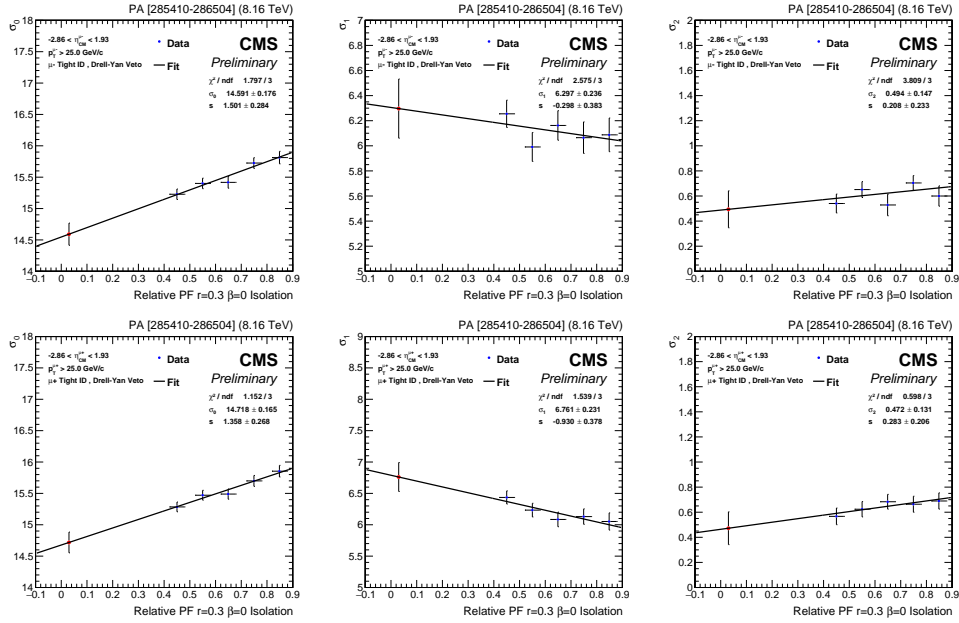Table 3.2: QCD shape parameters in the inclusive η_{CM} bin, extrapolated to the signal region (iso = 0.03).

Figure 3.14: Extrapolation of the parameters of the QCD background shape to the signal region: σ_0 (left), σ_1 (middle) and σ_2 (right), separated in negative (top) and positive (bottom) muon decays. Only the results inclusive in muon η_{CM} are shown.

3.6.2   Electroweak background

Among all the electroweak background sources, the dominant one in the W signal region is the Drell-Yan process. The Drell-Yan background is mainly made of $Z \to \mu^+ \mu^-$ events, where one of the muons is produced outside of the detector coverage ($|\eta| < 2.4$) or did not pass the selection criteria. The other type of EWK sources considered in this analysis are $Z/\gamma* \to \tau\bar{\tau}$, $W \to \tau\nu_\tau$ and $t\bar{t}$. The EWK background represents around 13.3% of the events in the signal region, divided as: $Z/\gamma* \to \mu^+ \mu^-$ (9.4%), $W \to \tau\nu_\tau$ (2.4%), $Z/\gamma* \to \tau\bar{\tau}$ (1.0%) and $t\bar{t}$ (0.5%). Other EWK processes such as double boson decays have been checked to contribute less than 0.03%, so they are not considered.

The contribution of each EWK background is described using $\mathbb{E}_T$ templates derived from the MC samples. The MET templates are produced following the procedure explained at the beginning of this section.

3.6.3 Fit model

The raw $W \to \mu \nu_\mu$ yields are extracted by performing an unbinned log-likelihood fit of the observed $\mathbb{E}_T$ distribution in each muon η_{CM} bin. The fits are done using a combination of binned templates and a continous functional form. The data analysis framework RooFit v3.60 [33] is used to make the fits.

The total fit model includes five contributions: the signal $W \to \mu \nu_\mu$ template ($\mathscr{T}_W$), the EWK backgrounds $Z/\gamma* \to \mu^+\mu^-$, $W \to \tau\nu_\tau$, $Z/\gamma* \to \tau\bar{\tau}$ and $t\bar{t}$ templates ($\mathscr{T}_{EWK}$), and the QCD data-driven functional form ($\mathscr{F}_{QCD}$). The model used to fit the data is shown in Eq. (3.17).

$$N_W \times \left(\mathscr{T}_W + r_{Z/\gamma* \to \mu^+\mu^-} \cdot \mathscr{T}_{Z/\gamma* \to \mu^+\mu^-} + r_{W \to \tau\nu_\tau} \cdot \mathscr{T}_{W \to \tau\nu_\tau} + r_{Z/\gamma* \to \tau\bar{\tau}} \cdot \mathscr{T}_{Z/\gamma* \to \tau\bar{\tau}} + r_{t\bar{t}} \cdot \mathscr{T}_{t\bar{t}} \right) + N_{QCD} \times \mathscr{F}_{QCD}$$

$$(3.17)$$

The signal and the EWK background shapes are defined based on $\mathbb{E}_T$ templates, evaluated from MC simulations. Since the fits become too unstable if the EWK background yields are left free, the following criteria is used to constrain their yields. First, the $Z/\gamma* \to \mu^+\mu^-$, $W \to \tau\nu_\tau$, $Z/\gamma* \to \tau\bar{\tau}$ and the $t\bar{t}$ raw yields are expressed in terms of $N_{EWK} = r_{EWK} \times N_W$. Afterwards, the ratio of raw yields (r_{EWK}) is fixed to its corresponding value derived from MC after having normalized all the MC samples to the recorded integrated luminosity and applied all analysis corrections and selection cuts. This criteria assume that the nuclear modification of the electroweak sources scales in the same way as the W signal, which is expected at least for the $W \to \tau\nu_\tau$ and the $Z/\gamma*$ processes.

The QCD contribution, as explained in section 3.6.1, is taken into account by means of a functional form depending on three parameters. For the fit of the $\mathbb{E}_T$ distribution the σ_0, σ_1 and σ_2 are fixed to the extrapolated values mentioned in Table 3.2.

The $\mathbb{E}_T$ distribution is fitted separately for $W^+ \to \mu^+\nu_\mu$ and $W^- \to \mu^-\bar{\nu}_\mu$ candidates. Only the signal raw yield (N_W) and the QCD normalization (N_{QCD}) are left free when fitting the W signal region. The fits are done in both the inclusive selected events and in bins of muon pseudorapidity in the center of mass frame η_{CM}.

3.6.4 Raw yields

The results of the fits to the data in each of the different muon η_{CM} bins are summarized in Table 3.3 for W$^-$ and in Table 3.4 for W$^+$.

η_{CM} Range	Total	Fitted	Signal	$Z/\gamma* \to \mu^+\mu^-$	$W \to \tau\nu_\tau$	$Z/\gamma* \to \tau\bar{\tau}$	$t\bar{t}$	QCD	p-value (χ^2)
-2.86 , -2.60	5210	5210	4056 ± 66	563 ± 9	137 ± 2	46 ± 1	3 ± 0	405 ± 40	0.42
-2.60 , -2.40	4308	4308	3407 ± 60	465 ± 8	104 ± 2	37 ± 1	4 ± 0	292 ± 36	0.31
-2.40 , -2.20	4273	4273	3290 ± 60	450 ± 8	103 ± 2	37 ± 1	5 ± 0	388 ± 38	0.26
-2.20 , -1.93	6423	6423	4936 ± 74	659 ± 10	158 ± 2	63 ± 1	12 ± 0	595 ± 48	0.53
-1.93 , -1.80	3140	3140	2430 ± 52	309 ± 7	80 ± 2	29 ± 1	8 ± 0	285 ± 34	0.51
-1.80 , -1.60	4822	4822	3682 ± 64	438 ± 8	119 ± 2	46 ± 1	19 ± 0	518 ± 43	0.46
-1.60 , -1.40	4727	4727	3644 ± 64	395 ± 7	119 ± 2	39 ± 1	18 ± 0	512 ± 43	0.85
-1.40 , -1.20	4521	4521	3600 ± 64	347 ± 6	111 ± 2	46 ± 1	21 ± 0	397 ± 40	0.03
-1.20 , -1.00	4626	4626	3672 ± 65	312 ± 6	120 ± 2	51 ± 1	23 ± 0	448 ± 42	0.57
-1.00 , -0.80	4722	4722	3769 ± 66	284 ± 5	121 ± 2	46 ± 1	31 ± 1	470 ± 43	0.87
-0.80 , -0.60	4198	4198	3434 ± 63	247 ± 5	105 ± 2	46 ± 1	29 ± 1	337 ± 39	0.82
-0.60 , -0.40	4648	4648	3743 ± 66	253 ± 4	122 ± 2	55 ± 1	37 ± 1	438 ± 43	0.09
-0.40 , -0.20	4344	4344	3484 ± 64	233 ± 4	112 ± 2	51 ± 1	34 ± 1	430 ± 41	0.16
-0.20 , +0.00	4474	4474	3522 ± 65	267 ± 5	114 ± 2	44 ± 1	37 ± 1	490 ± 43	0.01
+0.00 , +0.20	4643	4643	3661 ± 65	318 ± 6	116 ± 2	48 ± 1	39 ± 1	461 ± 43	0.00
+0.20 , +0.40	4638	4638	3538 ± 64	343 ± 6	112 ± 2	52 ± 1	43 ± 1	549 ± 44	0.00
+0.40 , +0.60	4718	4718	3538 ± 63	396 ± 7	117 ± 2	47 ± 1	38 ± 1	583 ± 44	0.30
+0.60 , +0.80	4552	4552	3382 ± 62	456 ± 8	106 ± 2	49 ± 1	34 ± 1	525 ± 43	0.64
+0.80 , +1.00	4637	4637	3331 ± 61	496 ± 9	106 ± 2	43 ± 1	39 ± 1	622 ± 44	0.00
+1.00 , +1.20	4612	4612	3281 ± 60	545 ± 10	108 ± 2	45 ± 1	27 ± 0	605 ± 44	0.54
+1.20 , +1.40	4053	4053	2761 ± 54	532 ± 11	78 ± 2	39 ± 1	22 ± 0	621 ± 42	0.11
+1.40 , +1.60	4251	4251	2924 ± 56	625 ± 12	100 ± 2	39 ± 1	21 ± 0	541 ± 43	0.16
+1.60 , +1.80	3844	3844	2511 ± 51	620 ± 13	79 ± 2	35 ± 1	15 ± 0	584 ± 41	0.06
+1.80 , +1.93	2640	2640	1724 ± 42	443 ± 11	55 ± 1	22 ± 1	9 ± 0	387 ± 34	0.44

Table 3.3: Raw yields of W$^- \to \mu^-\bar{\nu}_\mu$ and background processes, extracted from the nominal fits for each muon η_{CM} bin in the combined pPb and Pbp collision system. All analysis cuts are applied including the muon $p_T > 25$ GeV/c cut. All uncertainties shown are statistical only.

η_{CM} Range	Total	Fitted	Signal	$Z/\gamma* \to \mu^+\mu^-$	$W \to \tau\nu_\tau$	$Z/\gamma* \to \tau\bar{\tau}$	$t\bar{t}$	QCD	p-value (χ^2)
-2.86 , -2.60	4465	4465	3361 ± 59	589 ± 10	68 ± 1	46 ± 1	3 ± 0	397 ± 38	0.05
-2.60 , -2.40	4234	4234	3253 ± 58	533 ± 10	66 ± 1	37 ± 1	4 ± 0	341 ± 36	0.04
-2.40 , -2.20	4377	4377	3350 ± 60	503 ± 9	61 ± 1	37 ± 1	11 ± 0	415 ± 38	0.96
-2.20 , -1.93	6847	6846	5266 ± 76	723 ± 10	103 ± 1	55 ± 1	14 ± 0	686 ± 49	0.41
-1.93 , -1.80	3592	3591	2769 ± 56	340 ± 7	56 ± 1	30 ± 1	8 ± 0	388 ± 36	0.55
-1.80 , -1.60	5421	5417	4311 ± 70	496 ± 8	95 ± 2	51 ± 1	15 ± 0	449 ± 44	0.48
-1.60 , -1.40	5343	5343	4382 ± 70	451 ± 7	97 ± 2	46 ± 1	17 ± 0	350 ± 42	0.90
-1.40 , -1.20	5129	5125	4198 ± 70	386 ± 6	100 ± 2	42 ± 1	22 ± 0	377 ± 43	0.12
-1.20 , -1.00	5382	5380	4475 ± 72	344 ± 6	102 ± 2	56 ± 1	26 ± 0	378 ± 43	0.23
-1.00 , -0.80	5467	5464	4499 ± 73	314 ± 5	101 ± 2	51 ± 1	30 ± 0	469 ± 45	0.05
-0.80 , -0.60	4738	4737	3971 ± 69	251 ± 4	91 ± 2	43 ± 1	28 ± 0	352 ± 41	0.18
-0.60 , -0.40	5349	5349	4446 ± 73	263 ± 4	101 ± 2	49 ± 1	36 ± 1	454 ± 45	0.16
-0.40 , -0.20	5027	5023	4154 ± 70	246 ± 4	91 ± 2	47 ± 1	35 ± 1	450 ± 44	0.12
-0.20 , +0.00	5161	5160	4279 ± 71	276 ± 5	101 ± 2	47 ± 1	38 ± 1	419 ± 43	0.05
+0.00 , +0.20	5473	5471	4359 ± 72	315 ± 5	102 ± 2	55 ± 1	40 ± 1	600 ± 47	0.46
+0.20 , +0.40	5175	5175	4192 ± 70	343 ± 6	100 ± 2	49 ± 1	35 ± 1	457 ± 44	0.08
+0.40 , +0.60	5482	5481	4345 ± 71	409 ± 7	94 ± 2	44 ± 1	34 ± 1	555 ± 47	0.21
+0.60 , +0.80	5722	5719	4477 ± 72	482 ± 8	100 ± 2	51 ± 1	39 ± 1	569 ± 47	0.07
+0.80 , +1.00	6061	6061	4665 ± 73	570 ± 9	101 ± 2	48 ± 1	36 ± 1	641 ± 48	0.01
+1.00 , +1.20	5814	5814	4415 ± 70	609 ± 10	105 ± 2	41 ± 1	34 ± 1	610 ± 47	0.36
+1.20 , +1.40	5365	5362	4057 ± 67	574 ± 9	88 ± 1	35 ± 1	23 ± 0	585 ± 45	0.01
+1.40 , +1.60	5768	5768	4309 ± 69	680 ± 11	95 ± 2	41 ± 1	21 ± 0	622 ± 47	0.00
+1.60 , +1.80	5320	5320	3980 ± 65	667 ± 11	83 ± 1	34 ± 1	16 ± 0	541 ± 44	0.00
+1.80 , +1.93	3600	3599	2664 ± 54	456 ± 9	66 ± 1	20 ± 0	9 ± 0	384 ± 37	0.00

Table 3.4: Raw yields of $W^+ \to \mu^+\nu_\mu$ and background processes, extracted from the nominal fits for each muon η_{CM} bin in the combined pPb and Pbp collision system. All analysis cuts are applied including the muon $p_T > 25$ GeV/c cut. All uncertainties shown are statistical only.

3.6.5 Corrected yields

The raw yields extracted from the fits are corrected taking into account the efficiency of the detector. The corrected yields are derived from the raw yields after dividing by the MC efficiency corrected with the Tag-And-Probe method as described in Section 3.5.1. The formula used to derived the corrected yields is shown in Eq. (3.18).

$$N^{\pm}_{corr} = \frac{N^{\pm}_{raw}}{\epsilon^{MC}_{corr}} \tag{3.18}$$

The statistical uncertainty of the corrected yields are computed based on error propagation according to the following equation:

$$\delta N^{\pm}_{corr} = \frac{\delta N^{\pm}_{raw}}{\epsilon^{MC}_{corr}} \tag{3.19}$$

The results of the corrected yields with respect to the muon η_{CM} are summarized in Table 3.5 for W$^-$ and in Table 3.6 for W$^+$.

η_{CM} Range	Raw Yield	Efficiency (%)	Corrected Yield
-2.86 , -2.60	4056 ± 66	83.83 ± 0.21	4839 ± 78
-2.60 , -2.40	3407 ± 60	86.60 ± 0.22	3935 ± 70
-2.40 , -2.20	3290 ± 60	83.34 ± 0.23	3947 ± 72
-2.20 , -1.93	4936 ± 74	87.14 ± 0.17	5664 ± 85
-1.93 , -1.80	2430 ± 52	91.51 ± 0.20	2655 ± 57
-1.80 , -1.60	3682 ± 64	91.63 ± 0.15	4018 ± 70
-1.60 , -1.40	3644 ± 64	88.23 ± 0.23	4130 ± 73
-1.40 , -1.20	3600 ± 64	85.22 ± 0.22	4224 ± 75
-1.20 , -1.00	3672 ± 65	89.09 ± 0.20	4122 ± 73
-1.00 , -0.80	3769 ± 66	89.45 ± 0.21	4214 ± 74
-0.80 , -0.60	3434 ± 63	80.87 ± 0.24	4246 ± 78
-0.60 , -0.40	3743 ± 66	88.40 ± 0.22	4234 ± 75
-0.40 , -0.20	3484 ± 64	83.53 ± 0.24	4171 ± 77
-0.20 , +0.00	3522 ± 65	87.10 ± 0.23	4043 ± 74
+0.00 , +0.20	3661 ± 65	89.00 ± 0.22	4114 ± 73
+0.20 , +0.40	3538 ± 64	85.43 ± 0.24	4141 ± 75
+0.40 , +0.60	3538 ± 63	87.82 ± 0.23	4028 ± 72
+0.60 , +0.80	3382 ± 62	90.07 ± 0.21	3755 ± 68
+0.80 , +1.00	3331 ± 61	91.57 ± 0.17	3638 ± 66
+1.00 , +1.20	3281 ± 60	87.73 ± 0.21	3740 ± 68
+1.20 , +1.40	2761 ± 54	83.09 ± 0.24	3323 ± 66
+1.40 , +1.60	2924 ± 56	90.07 ± 0.23	3246 ± 62
+1.60 , +1.80	2511 ± 51	82.61 ± 0.28	3040 ± 62
+1.80 , +1.93	1724 ± 42	85.49 ± 0.33	2016 ± 49

Table 3.5: Corrected yields of $W^- \to \mu^- \overline{\nu}_\mu$, given for each muon η_{CM} bin in the combined pPb and Pbp collision system. All analysis cuts are applied including the muon $p_T >$ 25 GeV/c cut. The muon efficiency has been corrected by applying the Tag and Probe scale factors event by event. All uncertainties shown are statistical only.

η_{CM} Range	Raw Yield	Efficiency (%)	Corrected Yield
-2.86 , -2.60	3361 ± 59	83.49 ± 0.24	4026 ± 71
-2.60 , -2.40	3253 ± 58	86.60 ± 0.24	3757 ± 67
-2.40 , -2.20	3350 ± 60	83.29 ± 0.24	4022 ± 72
-2.20 , -1.93	5266 ± 76	86.21 ± 0.18	6108 ± 88
-1.93 , -1.80	2769 ± 56	91.40 ± 0.18	3030 ± 61
-1.80 , -1.60	4311 ± 70	91.31 ± 0.16	4721 ± 76
-1.60 , -1.40	4382 ± 70	88.31 ± 0.21	4962 ± 79
-1.40 , -1.20	4198 ± 70	84.15 ± 0.24	4988 ± 83
-1.20 , -1.00	4475 ± 72	87.94 ± 0.23	5089 ± 82
-1.00 , -0.80	4499 ± 73	88.75 ± 0.23	5069 ± 82
-0.80 , -0.60	3971 ± 69	80.39 ± 0.24	4940 ± 85
-0.60 , -0.40	4446 ± 73	88.16 ± 0.22	5043 ± 83
-0.40 , -0.20	4154 ± 70	82.55 ± 0.24	5032 ± 85
-0.20 , +0.00	4279 ± 71	86.81 ± 0.22	4929 ± 82
+0.00 , +0.20	4359 ± 72	88.79 ± 0.21	4910 ± 81
+0.20 , +0.40	4192 ± 70	84.55 ± 0.24	4958 ± 82
+0.40 , +0.60	4345 ± 71	87.81 ± 0.22	4948 ± 81
+0.60 , +0.80	4477 ± 72	89.75 ± 0.21	4989 ± 80
+0.80 , +1.00	4665 ± 73	90.99 ± 0.16	5127 ± 80
+1.00 , +1.20	4415 ± 70	87.16 ± 0.23	5065 ± 80
+1.20 , +1.40	4057 ± 67	82.64 ± 0.24	4909 ± 81
+1.40 , +1.60	4309 ± 69	89.52 ± 0.20	4813 ± 77
+1.60 , +1.80	3980 ± 65	82.10 ± 0.25	4848 ± 79
+1.80 , +1.93	2664 ± 54	85.60 ± 0.29	3112 ± 63

Table 3.6: Corrected yields of $W^+ \rightarrow \mu^+ \nu_\mu$, given for each muon η_{CM} bin in the combined pPb and Pbp collision system. All analysis cuts are applied including the muon $p_T > 25$ GeV/c cut. The muon efficiency has been corrected by applying the Tag and Probe scale factors event by event. All uncertainties shown are statistical only.

3.7 Systematic uncertainties

This section presents the different sources and the procedure employed to determine the systematic uncertainties in the measurement of the W boson production in pPb.

3.7.1 Luminosity

The integrated luminosity of the 2016 pPb data sample processed is $173.4 \pm 5\%$ nb^{-1} as presented in Section 3.2.1. Since the integrated luminosity cancels in asymmetries, it only affects the measurement of the $W \to \mu \nu_\mu$ differential cross sections. In this case, the systematic uncertainty is taken as a global uncertainty of 5% and the bin-to-bin correlation is +100%.

3.7.2 Muon efficiency

3.7.2.1 MC statistics

A systematic uncertainty is assigned to account for the limited statistics in the MC samples used to compute the muon efficiency. The MC statistical uncertainty of the muon efficiencies is computed using Bayesian statistics as explained in Section 3.5 and then propagated to all the observables. The uncertainty is considered to be fully uncorrelated.

3.7.2.2 Theory model

The NLO model used to generate the MC samples can impact the measurement of the muon efficiencies. The main sources of theory uncertainties include the choice of the nuclear parton distribution function (EPPS16+CT14), renormalization scale, factorization scale and α_s. The theory uncertainties are assumed to be correlated in both charge and η_{CM}.

Since the PDFs are not calculable from first principles but are determined experimentally, the inclusion of any PDF introduces an additional systematic error. Therefore, it is important to determine the impact of a change of PDF on the muon efficiencies.

The procedure used to derive the theory uncertainties of the PDF variations consist of reweighing the simulation using the weights produced by POWHEG after applying each PDF set. The PDF sets are accessed through the LHAPDF6 [34] framework and consist of 56 CT14 PDFs, 40 EPPS16 nuclear corrections and 2 CT14 α_s variations. Once the MC samples are reweighed with each PDF set, the efficiencies are recomputed and used to recalculate all the observables.

The nPDF uncertainty is determined by combining the EPPS16+CT14 variations of the observables using the Hessian approach as recommended by the EPPS16 authors (see Eq. 29 of Ref. [35]). The EPPS16+CT14 sets corresponds to 90% confidence interval (CL) variations, so the uncertainties are scaled to 68%.

Furthermore, the α_s uncertainty is determined by using the EPPS16+CT14 nPDFs associated to the $\alpha_s\left(m_Z^2\right)$ values: 0.1170 and 0.1190. The variations on the observables are combined following the prescription recommended in the PDF4LHC15 [36], which basically corresponds to taking the average per bin of the two variations. The uncertainties are converted to 68% CL after rescaling them by 0.0015/0.0010 as described in PDF4LHC15 [36].

And finally, the uncertainty due to the renormalization (μ_R) and factorization (μ_F) scales, is computed by varying the two scales in POWHEG using the following sets:

$$(\mu_R, \mu_F) = [\,(0.5, 0.5)\,,\,(1.0, 0.5)\,,\,(0.5, 1.0)\,,\,(1.0, 2.0)\,,\,(2.0, 1.0)\,,\,(2.0, 2.0)\,]$$

The MC samples are reweighed event by event using the POWHEG weights produced with each set of scales, then the efficiencies are recomputed and the observables are recalculated for each varied efficiency. The variations on the observables are combined by taking the envelope (i.e. the maximum variation in each η_{CM} bin). The uncertainties are considered 68% CL.

3.7.2.3 Tag-and-Probe correction

The main source of systematic uncertainty in the measurement of the muon efficiency arises from the application of the TnP scale factors used to correct the truth efficiency. As mentioned in Section **??**, the statitstical and systematic uncertainties of the TnP correction is composed by the different components that enters the muon efficiency: trigger, identification, isolation, reconstruction, event activity and pileup.

Since this analysis measures the forward-backward ratios and the charge asymmetry of W bosons, it is crucial to consider the correlation between the different TnP uncertainties as a function of muon η_{CM} and charge. The statistical TnP variations are considered correlated only within the η_{LAB} bins (see [32]) in which they were derived (η_{LAB} bin in the case of trigger, $|\eta_{LAB}|$ bin in the case of identification and isolation), while no correlation is assumed with respect to the muon charge. The systematic TnP variations are expected to be fully correlated as a function of muon charge and uncorrelated between the different η_{CM}. The correlations of each source of TnP uncertainty are listed in Table 3.7.

Source of uncertainty	correlation in muon η	correlation in muon charge
Trigger, stat.	yes per η range	no
ID, stat.	yes per $|\eta|$ range	no
Iso, stat.	yes per $|\eta|$ range	no
Trigger, syst.	no	yes
ID, syst.	no	yes
Iso, syst.	no	yes
ID, syst. binned	no	yes
Iso, syst. binned	no	yes
HF+PU	no	yes
STA	no	yes

Table 3.7: Summary of the correlations per source of tag and probe uncertainty, as a function of muon pseudorapidity (in the laboratory frame) and charge. The η (for trigger) or $|\eta|$ (for identification and isolation) ranges are the one used to define the scale factors. When a correlation per η or $|\eta|$ range is mentioned, the different ranges are uncorrelated between them.

To compute the uncertainties, the muon charge asymmetry and the forward-backward ratios are recalculated for each efficiency derived by varying the TnP scale factors. The TnP uncertainties are then determined by taking the difference between the value obtained with the varied TnP SF and its nominal value, combining the uncertainties as explained in Section **??**. If the TnP source is correlated in muon charge or pseudorapidity, the corresponding W yields are varied at the same time. Moreover, for the $W^{\pm}$ cross sections, the statistical and systematic TnP uncertainties are calculated by propagating the uncertainties on the corrected muon efficiency due to the TnP variations.

3.7.3 Choice of binning

Since the W signal distribution is parametrised using a binned MC template with a bin width of 2 GeV/c, the uncertainty on the signal template is determined by decreasing the bin width to 1 GeV/c. The systematic uncertainty is assigned for each η_{CM} bin by taking the maximum difference between the nominal and the varied observables. The correlations bin-to-bin and between the positively and negatively charged muons is taken to be 100 %.

3.7.4   QCD background

The systematic uncertainty in the QCD background originates from the uncertainty in the modeling of the multi-jet MET distribution in the signal region. The nominal procedure consist on fixing the parameters of the modified Rayleigh distribution from the fits extrapolated from data as explained in Sec. 3.6.1. In order to estimate the uncertainty of the mismodeling of the QCD shape, both the parameters and the functional form are varied as explained in the following subsections.

3.7.4.1   QCD model parameters

The first source of systematic uncertainty reflects the possible mis-modelling of the QCD shape due to the η_{CM} dependence of the QCD parameters. In order to check this, the parameters of the nominal QCD model are set free but constrained to the weighted RMS and mean value of the extrapolated results extracted in each η_{CM} bin. The values used to constrain the σ_0, σ_1 and σ_2 parameters are shown in Table 3.8. The average over η_{CM} bins of the difference between the final observables derived from the constrained fits and the nominal results is taken as the systematic error. This source of uncertainty is taken to be fully uncorrelated.

Parameter	$QCD \to \mu^-$	$QCD \to \mu^+$
σ_0	14.49±0.95	14.78±0.52
σ_1	6.20±0.89	6.77±0.91
σ_2	0.50±0.73	0.48±0.57

Table 3.8: QCD shape parameters, extracted from the weighted mean and RMS in muon η_{CM} of the extrapolated values to the signal region.

Another systematic variation consists of changing the muon isolation point used to extrapolate the QCD parameters. In the nominal case, the isolation point of 0.03 was determined from the average muon isolation value in data inside the signal region. As an alternative case, the isolation distribution is checked in a QCD MC sample passing all the analysis cuts and the average isolation value is determined to be approximately 0.08. Thus, the QCD parameters are recomputed by extrapolating them to an isolation point of 0.08, and the fits are redone by fixing the QCD parameters to the extrapolated values in the η inclusive bin as in the nominal case. The new QCD parameters are listed in Table 3.9. Since each bin varies independently, the uncertainty is taken to be fully uncorrelated.

Parameter	$QCD \to \mu^-$	$QCD \to \mu^+$
σ_0	14.67	14.79
σ_1	6.28	6.71
σ_2	0.50	0.49

Table 3.9: QCD shape parameters, extrapolated to the average isolation in a QCD MC sample passing all analysis cuts (iso = 0.08).

3.7.4.2 QCD functional form

To assign a systematic uncertainty due to the assumed functional form for modelling the QCD MET distribution, the shape of the QCD is described using a different model. The alternative MET distribution function used, taken from [1], is shown in Eq. (3.20).

$$f(x) = (x + x_0)^{\alpha} \exp(\beta \sqrt{x + x_0}) \tag{3.20}$$

The extrapolation procedure explained in Section ?? is redone using the alternative model, and the results are summarized in the Appendix ??. All the fits are remade using the new QCD functional form fixed to the QCD parameters extrapolated in the inclusive η_{CM} bin. The difference between the final observables derived using the alternative QCD PDF and the nominal results is taken as the systematic error due to mis-modeling of the QCD shape. The bin-to-bin correlation for this uncertainty is taken to be fully uncorrelated.

3.7.5 EWK background

The different sources of EWK background are described using templates derived from the MC POWHEG samples. Since the simulations are scaled to the luminosity recorded by CMS by using the cross-section of each EWK process, a systematic uncertainty is assigned to each source by varying their cross sections as explained below.

3.7.5.1 $Z/\gamma* \to \mu^+\mu^-$ background

When performing the fits, the ratio of the $Z/\gamma* \to \mu^+\mu^-$ background over the signal yields is fixed to the corresponding value extracted from the simulation. The MC sample is normalized to match the integrated luminosity in data using the $Z/\gamma* \to \mu^+\mu^-$ and $W \to \mu\nu_\mu$ NLO cross sections derived from POWHEG.

The uncertainty on the W/Z ratio is estimated using MCFM [37] at NLO with CT14+EPPS16 [13, 14]. More precisely, the ratio is between the $W \to \mu\nu_\mu$ cross section and the $Z/\gamma* \to \mu^+\mu^-$ ($M \in [15, 600]\,\mathrm{GeV}/c^2$) cross section, both with muon kinematic cuts ($p_T^\mu > 25$ GeV/c for W and $p_T^\mu > 15$ GeV/c for $Z/\gamma*$). This PDF uncertainty is propagated in the standard way for Hessian PDF sets (Eq. 20 of Ref. [36]). The results for the full CT14+EPPS16 uncertainty are the following:

$$W^+/Z = 3.455^{+0.038}_{-0.044} \tag{3.21}$$

$$W^-/Z = 2.784^{+0.021}_{-0.020} \tag{3.22}$$

corresponding to a relative uncertainty of 0.8% for Z/W^- and 1.3% for Z/W^+.

Since the cross sections in the muon channel depend on the branching ratio associated to each process, their uncertainty has to be also taken into account. The values of the branching ratios are extracted from the Particle Data Group (PDG) database [38], and correspond to $\mathrm{BR}(Z \to \mu^+\mu^-) = (3.366 \pm 0.007)\%$ and $\mathrm{BR}(W \to \mu\nu_\mu) = (10.63 \pm 0.15)\%$, which gives a relative uncertainty on the ratio of Z/W branching ratios of 1.4%. Summing in quadrature the MCFM uncertainties with the ones derived from the branching ratios, one gets a total relative uncertainty for Z/W^+ of 1.6% and for Z/W^- of 1.9%. To be conservative the systematic variation is fixed to 2% overall.

Therefore, the systematic uncertainty due to the $Z/\gamma* \to \mu^+\mu^-$ background is determined by varying the $Z/\gamma* \to \mu^+\mu^-$ cross section by 2% up and down when performing the fits. The systematic uncertainty in each η_{CM} bin is derived by taking the maximum difference between the nominal and the up/down variations. The bin-to-bin correlations in charge and pseudorapidity are taken to be 100%.

3.7.5.2 $Z/\gamma* \to \tau\bar{\tau}$ background

The ratio of the $Z/\gamma* \to \tau\bar{\tau}$ background over the signal raw yields is also fixed to the values derived from the corrected MC samples. Its uncertainty is considered to be the same as the 2% uncertainty determined for the $Z/\gamma* \to \mu^+\mu^-$ background as detailed in Section **??**.

To account for the difference in the Z branching ratios to leptons, the current values are taken from the PDG [38]. They correspond to $\mathrm{BR}(Z \to \tau\bar{\tau}) = (3.370 \pm 0.008)\%$ and $\mathrm{BR}(Z \to \mu^+\mu^-) = (3.366 \pm 0.007)\%$, which represents a relative uncertainty on the ratio of $Z \to \tau\bar{\tau}/Z \to \mu^+\mu^-$ branching ratios of 0.3%. Since the relative uncertainty of the Z

branching ratios is very small compared to the $Z/\gamma* \to \mu^+\mu^-$ background uncertainty, it is not considered.

Hence, the $Z/\gamma* \to \tau\bar{\tau}$ cross section is varied by 2% up and down when performing the fits. The systematic uncertainty in each η_{CM} bin is derived by taking the maximum difference between the nominal and the up/down variations. The bin-to-bin correlations are taken to be 100% both in charge and pseudorapidity.

3.7.5.3 $W \to \tau\nu_\tau$ background

The $W \to \tau\nu_\tau$ and the $W \to \mu\nu_\mu$ MC samples are first normalized to the data integrated luminosity using the POWHEG NLO cross sections. Afterwards, the ratio of the $W \to \tau\nu_\tau$ background and the signal raw yields is fixed to values from the simulations.

The current results of the W leptonic branching ratios, as reported in the PDG [38], corresponds to BR($W \to \mu\nu_\mu$) = (10.63 ± 0.15)% and BR($W \to \tau\nu_\tau$) = (11.38 ± 0.21)%, which gives a relative uncertainty on the ratio of $W \to \tau\nu_\tau/W \to \mu\nu_\mu$ cross sections of 2.3%.

To determine the systematic error, the ratio of $W \to \tau\nu_\tau$ to signal yields is varied up and down by ± 2.3%, and the MET fits are computed again. The maximum difference between the varied up or down observable and the nominal observable value in each muon η_{CM} bin is taken as the systematic error. The bin-to-bin correlations in muon pseudorapidity and charge are taken to be 100%.

3.7.5.4 $t\bar{t}$ background

The $t\bar{t}$ simulation is normalized to the integrated luminosity measured in data using $\sigma_{t\bar{t}} = 45 \pm 8$ nb^{-1}, which corresponds to the total $t\bar{t}$ cross-section measured by CMS in pPb collisions at 8.16 TeV [23].

The systematic error related to the $t\bar{t}$ background normalization is computed by varying up and down the total cross-section by its relative uncertainty (± 18%) and repeating all the MET fits. The maximum difference between the varied up or down observable and the nominal observable value in each muon η_{CM} bin is taken as the systematic error. The bin-to-bin correlations in muon pseudorapidity and charge are taken to be 100%.

3.7.6 Event activity reweighting

The modelling of the underlying event (UE) activity present in the simulated pPb collisions is improved by reweighting the distribution of the energy deposited in the HF

calorimeters, as explained in Section **??**.

The UE activity is also correlated with other global variables such as the track multiplicity. Therefore, the systematic error on the UE reweighting is determined by reweighting the distribution of the number of tracks in MC to match what is observed in data, and computing again the fits and the efficiency. The variations on the observables with respect to the nominal results is assigned as the systematic uncertainty in each muon η_{CM} bin. This source of uncertainty is considered uncorrelated.

3.7.7 Recoil Correction

Systematic errors are assigned for uncertainties in the recoil corrections used to correct the MET distributions in MC to match the data. We can classify the uncertainties in the recoil corrections in *statistical* and *systematic* components. The statistical component arises from the uncertainties (statistical) in the recoil scale and resolution obtained from the fits of the recoil distributions in the different q_T bins. The systematic component arises from the following sources:

- the recoil correction method used to correct the MET distributions in MC

- the functions used to parametrise the q_T dependence of the recoil scale and resolution

- the assumptions made on shape used to fit the recoil distributions in each q_T bin

- the uncertainty on the energy scale of the PF candidates used to compute the MET

- a possible rapidity dependence of the recoil corrections

3.7.7.1 Statistical component

As we can observe in Figs. **??**, **??**, **??** and **??**, the statistical uncertainties on the recoil scale in the different q_T bins is very small, so we do not need to take them into account. In order to estimate the uncertainty associated to the recoil resolution, the resolution parameter in each q_T bin is randomly smeared using a Gaussian function centred in the parameter value and with sigma equal to the parameter uncertainty. The q_T dependence is parametrised again using the nominal functions mentioned in Sections **??** and **??**. The procedure is repeated a hundred times, and the corrections are applied to the MC MET distributions, to perform the measurements again.

The variations on the MC and data recoil resolutions are computed separately, and the RMS of the measured asymmetry and cross section distributions in each muon η_{CM}, derived from the hundred variations is computed in each case. The quadratic sum of the RMS due to the MC and data recoil resolution variations is taken as the systematic uncertainty.

3.7.7.2 Systematic component

In order to compute the uncertainty related to the functions used to parametrise the q_T dependence of the recoil scale and resolution, we use different fitting functions in both data and MC. The functions used are: a second order polynomial for the recoil scale and resolution for the parallel component of the recoil, and a first order polynomial for the recoil scale and a second order polynomial for the recoil resolution in the case of the perpendicular component of the recoil. The corrections are applied to the MC MET distributions, to perform the measurements again, and the difference with respect to the nominal results is assigned as systematic uncertainty.

The uncertainty on the assumptions made on the shape of the recoil distributions in each q_T bin can be estimated using a different function. Instead of using a weighted sum of two Gaussian functions we use the weighted sum of a Breit-Wigner and a Gaussian functions, both in data and MC. The resulting q_T dependence of the recoil scale and resolution are fitted with the nominal functions in Sections **??** and **??**, and the corrections are applied to the MC MET distributions, to perform the measurements again. The difference with respect to the nominal results is assigned as systematic uncertainty.

Futhermore, the uncertainty on the energy of the raw PF jets, that enters the MET distribution, is estimated using the tools provided by the JetMET POG (code). The procedure consists of varying by one sigma the energy of the PF jets using the JES uncertainty. The JES uncertainties are extracted from the latest pp AK4PF JEC for 2016 MC and data as mentioned in the JetMET twiki. Once the variations are performed, the PF MET is recomputed and the recoil distributions are fitted again using the nominal functions. The new sets of recoil corrections are then used to correct again the MC, which is then used to extract the W yields. The maximum difference between the nominal and the up/down variations of the observables is taken as the systematic uncertainty in each bin.

And finally, the uncertainty on the method used to apply the recoil correction is determined by using the smearing method instead of the scaling guassian used as nominal.

3.7.8 Summary of systematic uncertainties

The maximum systematic uncertainty for each category among all η_{CM} bins is summarized in the Table 3.10. The systematic errors are shown for each observable, including the W cross sections, W charge asymmetry and the forward-backward ratios.

Systematic Variation	$W^- B \times d\sigma/d\eta_{CM}(nb)$	$W^+ B \times d\sigma/d\eta_{CM}(nb)$	$W^- R_{FB}$	$W^+ R_{FB}$	$W R_{FB}$	$(N^+ - N^-)/(N^+ + N^-)$
Binning	0.001	0.002	0.002	0.002	0.002	0.001
EWK Background	0.004	0.003	0.002	0.001	0.001	0.000
Efficiency	0.030	0.032	0.026	0.037	0.030	0.011
Event Activity	0.012	0.008	0.010	0.010	0.006	0.006
QCD Background	0.012	0.007	0.016	0.008	0.009	0.006
Recoil Correction	0.004	0.003	0.004	0.003	0.002	0.002
Total Systematic Unc.	0.033	0.033	0.030	0.038	0.031	0.013
Statistical Unc.	0.024	0.020	0.026	0.029	0.019	0.015

Table 3.10: Maximum error of the measured observables determined for each category in the combined pPb and Pbp collision system. The uncertainties of the cross-sections are relative while for the asymmetries are absolute.

3.7.9 Covariance matrix

The covariance matrix of the systematic uncertainties is computed by combining all the bins of each observable to account for the bin-to-bin correlations. In the case of the $W^\pm$ cross sections and the $W^\pm$ forward-backward ratios, the matrix is made using 48 bins (24 pseudorapidity bins times 2 charge bins), while for the charge asymmetry and the charge-inclusive R_{FB} only 24 bins are considered.

For a given (i,j) entry of the matrix, the covariance is calculated as the error in bin i times the error in bin j. If the uncertainty is uncorrelated, the off-diagonal elements are set to zero. The total covariance matrix is computed by summing the matrices of each uncertainty.

The total correlation matrix of each observable is derived from the total covariance matrix by using the following formula $corr(i,j) = cov(i,j)/(\sqrt{cov(i,i) * cov(j,j)})$. The corresponding correlation matrices are displayed in Fig. 3.16. The black lines are used to distinguish the different bins of muon charge which are ordered from top to bottom as: Minus-Minus , Minus-Plus, Plus-Minus and Plus-Plus.
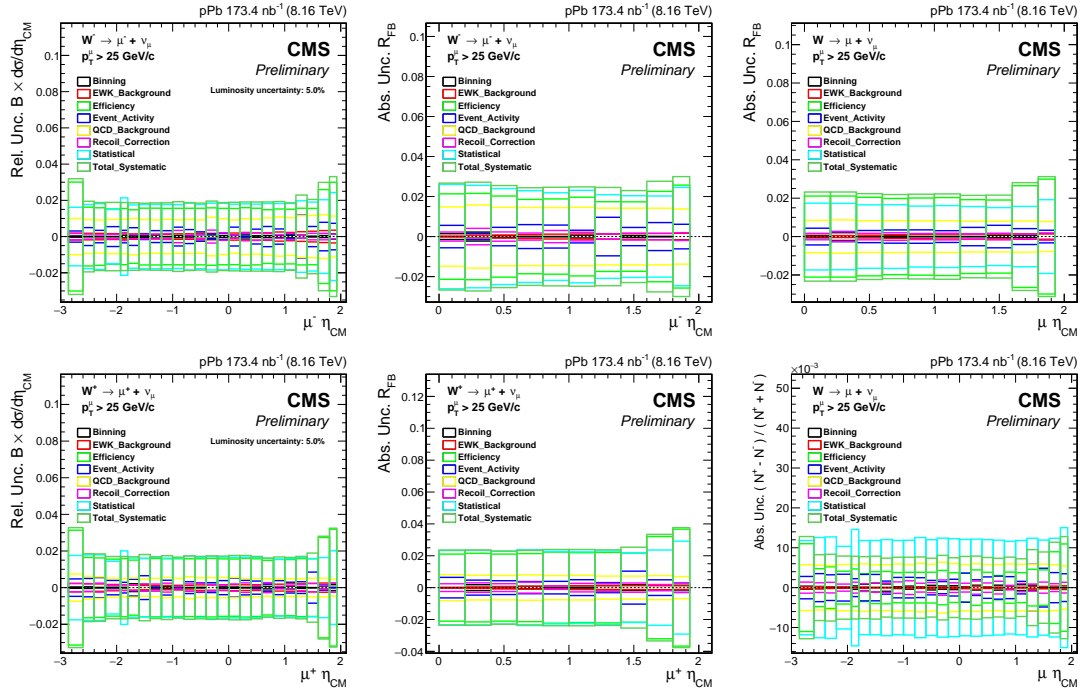
Figure 3.15: Uncertainty corresponding to each category as function of the muon η_{CM}. The plots are divided as: $W^- \to \mu^- \overline{\nu}_\mu$ (top-left) and $W^+ \to \mu^+ \nu_\mu$ (bottom-left) cross sections, W^- (top-middle) and W^+ (bottom-middle) R_{FB}, and finally the W R_{FB} (top-right) and W charge asymmetry (bottom-right). The uncertainties of the cross sections are relative while for the asymmetries are absolute. The global luminosity uncertainty of $\pm5.0\%$ is not included.
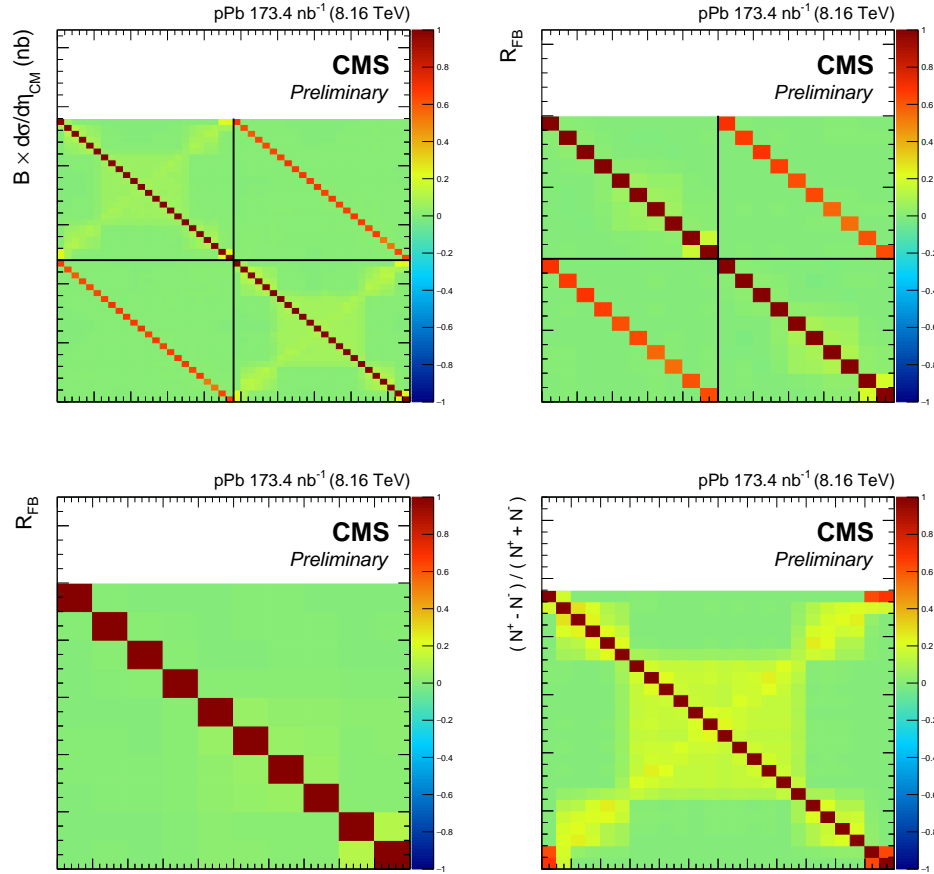
Figure 3.16: Correlation matrix of the systematic uncertainties. The plots are divided as: $W^{\pm}$ cross section (top-left) , $W^{\pm}$ R_{FB} (top-right) , charge-inclusive R_{FB} (bottom-left) , and charge asymmetry (bottom-right). The lines in the top plots are used to separate the different muon charge bins.

3.8 Results

CONCLUSION

BIBLIOGRAPHY

[1] **CMS** Collaboration, V. Khachatryan *et al.*, "Study of W boson production in pPb collisions at $\sqrt{s_{NN}}$ = 5.02 TeV", *Phys. Lett.* **B750** (2015) 565–586, arXiv:1503.05825. [Cited on pages 5 and 41.]

[2] **ALICE** Collaboration, K. Senosi, "Measurement of W-boson production in p-Pb collisions at the LHC with ALICE", *PoS* **Bormio2015** (2015) 042, arXiv:1511.06398. [Cited on page 5.]

[3] **ATLAS** Collaboration, M. Duman*ƒçiƒ*á, "*W* and *Z* boson production in 5.02 TeV *pp* and *p*+Pb collisions with the ATLAS detector", *Nucl. Part. Phys. Proc.* **289-290** (2017) 193–196, arXiv:1704.00298. [Cited on page 5.]

[4] **CMS Collaboration** Collaboration, "Luminosity calibration for the pPb/Pbp 2016 runs", Tech. Rep. CMS-LUM-AN-17-002, CERN, Geneva, 2017. [Cited on page 7.]

[5] **CMS** Collaboration, S. Chatrchyan *et al.*, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004. [Cited on pages 7 and 10.]

[6] E. Todesco and J. Wenninger, "Large Hadron Collider momentum calibration and accuracy", *Phys. Rev. Accel. Beams* **20** (2017), no. 8, 081003. [Cited on page 7.]

[7] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, arXiv:0709.2092. [Cited on page 7.]

[8] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2004) 040, arXiv:hep-ph/0409146. [Not cited.]

[9] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, arXiv:1002.2581. [Cited on page 7.]

[10] L. Barze, G. Montagna, P. Nason, O. Nicrosini, and F. Piccinini, "Implementation of electroweak corrections in the POWHEG BOX: single W production", *JHEP* **04** (2012) 037, `arXiv:1202.0465`. [Cited on page 7.]

[11] L. Barze, G. Montagna, P. Nason, O. Nicrosini, F. Piccinini, and A. Vicini, "Neutral current Drell-Yan with combined QCD and electroweak corrections in the POWHEG BOX", *Eur. Phys. J.* **C73** (2013), no. 6, 2474, `arXiv:1302.4606`. [Cited on page 7.]

[12] S. Frixione, P. Nason, and G. Ridolfi, "A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction", *JHEP* **09** (2007) 126, `arXiv:0707.3088`. [Cited on page 7.]

[13] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C. P. Yuan, "New parton distribution functions from a global analysis of quantum chromodynamics", *Phys. Rev.* **D93** (2016), no. 3, 033006, `arXiv:1506.07443`. [Cited on pages 7 and 42.]

[14] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado, "EPPS16: Nuclear parton distributions with LHC data", *Eur. Phys. J.* **C77** (2017), no. 3, 163, `arXiv:1612.05741`. [Cited on pages 7 and 42.]

[15] T. Sjostrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An Introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159–177, `arXiv:1410.3012`. [Cited on page 8.]

[16] **CMS** Collaboration, V. Khachatryan *et al.*, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J.* **C76** (2016), no. 3, 155, `arXiv:1512.00815`. [Cited on page 8.]

[17] **CMS** Collaboration, S. Chatrchyan *et al.*, "Study of the underlying event at forward rapidity in pp collisions at $\sqrt{s}$ = 0.9, 2.76, and 7 TeV", *JHEP* **04** (2013) 072, `arXiv:1302.2394`. [Cited on page 8.]

[18] N. Davidson, G. Nanava, T. Przedzinski, E. Richter-Was, and Z. Was, "Universal Interface of TAUOLA Technical and Physics Documentation", *Comput. Phys. Commun.* **183** (2012) 821–843, `arXiv:1002.0543`. [Cited on page 8.]

[19] P. Golonka and Z. Was, "PHOTOS Monte Carlo: A Precision tool for QED corrections in *Z* and *W* decays", *Eur. Phys. J.* **C45** (2006) 97–107, `arXiv:hep-ph/0506026`. [Cited on page 8.]

[20] **GEANT4** Collaboration, S. Agostinelli *et al.*, "GEANT4: A Simulation toolkit", *Nucl. Instrum. Meth.* **A506** (2003) 250–303. [Cited on page 8.]

[21] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner, "EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider", *Phys. Rev.* **C92** (2015), no. 3, 034906, `arXiv:1306.0121`. [Cited on page 8.]

[22] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Pseudorapidity distributions of charged hadrons in proton-lead collisions at $\sqrt{s_{\mathrm{NN}}}$ = 5.02 and 8.16 TeV", `arXiv:1710.09355`. [Cited on page 8.]

[23] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Observation of top quark production in proton-nucleus collisions", `arXiv:1709.07411`. [Cited on pages 9, 28, 43, and 57.]

[24] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Particle-flow reconstruction and global event description with the cms detector", *JINST* **12** (2017) P10003, `arXiv:1706.04965`. [Cited on pages 11 and 12.]

[25] **CMS Collaboration** Collaboration, "Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector", Tech. Rep. CMS-PAS-JME-16-004, CERN, Geneva, 2016. [Cited on page 11.]

[26] **CMS** Collaboration, S. S. Ghosh, "Performance of MET reconstruction and pileup mitigation techniques in CMS", *Nucl. Part. Phys. Proc.* **273-275** (2016) 2512–2514, `arXiv:1502.05207`. [Cited on page 11.]

[27] M. Jeitler, J. Lingemann, D. Rabady, H. Sakulin, and A. Stahl, "Upgrade of the CMS Global Muon Trigger", *IEEE Trans. Nucl. Sci.* **62** (2015), no. 3, 1104–1109. [Cited on page 11.]

[28] J. P. Lingemann, A. Stahl, and H. Sakulin, "Upgrade of the global muon trigger at the CMS experiment", PhD thesis, RWTH Aachen University, Sep 2016. Presented 14 Sep 2016. [Cited on page 11.]

[29] **CMS** Collaboration, V. Khachatryan *et al.*, "The CMS trigger system", *JINST* **12** (2017), no. 01, P01020, `arXiv:1609.02366`. [Cited on page 12.]

[30] R. Brun and F. Rademakers, "Root – an object oriented data analysis framework", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **389** (1997), no. 1, 81 – 86, New Computing Techniques in Physics Research V. [Cited on page 25.]

[31] **CMS Collaboration** Collaboration, "Generic tag and probe tool for measuring efficiency at cms with early data", Tech. Rep. CMS-MUO-AN-09-111, CERN, Geneva, 2009. [Cited on page 25.]

[32] **CMS Collaboration** Collaboration, "Muon performance studies in 2016 pPb data", Tech. Rep. CMS-HIN-AN-17-137, CERN, Geneva, 2017. [Cited on pages 26 and 38.]

[33] W. Verkerke and D. P. Kirkby, "The RooFit toolkit for data modeling", *eConf* **C0303241** (2003) MOLT007, `arXiv:physics/0306116`, [,186(2003)]. [Cited on page 31.]

[34] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, "LHAPDF6: parton density access in the LHC precision era", *Eur. Phys. J.* **C75** (2015) 132, `arXiv:1412.7420`. [Cited on page 37.]

[35] K. J. Eskola, H. Paukkunen, and C. A. Salgado, "EPS09: A new generation of NLO and LO nuclear parton distribution functions", *JHEP* **04** (2009) 065, `arXiv:0902.4154`. [Cited on page 38.]

[36] J. Butterworth *et al.*, "PDF4LHC recommendations for LHC Run II", *J. Phys.* **G43** (2016) 023001, `arXiv:1510.03865`. [Cited on pages 38 and 42.]

[37] R. Boughezal, J. M. Campbell, R. K. Ellis, C. Focke, W. Giele, X. Liu, F. Petriello, and C. Williams, "Color singlet production at NNLO in MCFM", *Eur. Phys. J.* **C77** (2017), no. 1, 7, `arXiv:1605.08011`. [Cited on page 42.]

[38] **Particle Data Group** Collaboration, C. Patrignani *et al.*, "Review of Particle Physics", *Chin. Phys.* **C40** (2016), no. 10, 100001. [Cited on pages 42 and 43.]

TABLE   **Page**

LIST OF FIGURES

FIGURE   **Page**