

Aristotle University of Thessaloniki
Faculty of Science
Department of Physics



Master's Thesis

Study of signal to background discrimination and
determination of the "signal strength" for the ZZ
boson production in proton-proton collisions at 13
TeV with the ATLAS detector

Christos Toukmenidis

Under supervision of

Prof. Konstantinos Kordas

Submitted in partial fulfillment of the requirements
for the Master degree in Subatomic Physics and Technological
Applications

AUTh October 2023

Acknowledgments

I would like to express my sincere gratitude to my supervisor Konstantinos Kordas, Professor at A UTh, for his invaluable guidance and support throughout the completion of this work.

As my time as a student at A UTh comes to an end upon the completion of this work, I want to express my deep appreciation to all my professors for sharing their valuable knowledge with me and also for fueling my passion for the subject even further.

In addition, I owe special thanks to my colleague and PhD candidate at A UTh, Angelos Tsiamis for his crucial help and advice on methodology and analysis techniques.

Moreover, I owe huge thanks to my family and friends for their belief in me and especially to my dear friend, Sofia. Her support and belief in me throughout the entire demanding process of studying and completing the current work has been truly priceless.

Lastly, I want to express my deep gratitude to Charis, my colleague and lifelong best friend, for being a constant source of support and joy throughout my academic journey and life, sharing in both our past experiences and those that are yet to come ☺.

Περίληψη

Η παρούσα διπλωματική εργασία αφορά στη μελέτη μεθόδων ανάλυσης πειραματικών δεδομένων με σκοπό τον διαχωρισμό μεταξύ σήματος και υποβάθρου κατά τη παραγωγή δύο Z μποζονίων. Η μελέτη πραγματοποιείται μέσω της ταυτοποίησης του καναλιού $ZZ \rightarrow ll\nu\nu$ όπου τα δύο Z μποζόνια διασπώνται σε 2 λεπτόνια και 2 νετρίνο. Η μελέτη αφορά δεδομένα συγκρούσεων πρωτονίων σε ενέργεια κέντρου μάζας $\sqrt{s} = 13$ TeV που συλλέχθηκαν από το πείραμα ATLAS στο CERN καθ' όλη τη διάρκεια της δεύτερης περιόδου (Run II) και που αντιστοιχούν σε ολοκληρωμένη φωτεινότητα 140.1 fb^{-1} .

Η εργασία ξεκινά με μια σύντομη περιγραφή του υπό μελέτη καναλιού σήματος και του ανιχνευτή ATLAS (Ενότητα 1). Στη συνέχειά γίνεται μια σύντομη περιγραφή της διαδικασίας επιλογής αντικειμένων και γεγονότων από τον ανιχνευτή καθώς και των δειγμάτων δεδομένων (data) και προσομοιώσεων Monte Carlo (Ενότητα 2).

Όσον αφορά τη μελέτη του σήματος $ZZ \rightarrow ll\nu\nu$ (Ενότητα 3), προκειμένου να πραγματοποιηθεί μια ακριβής εκτίμηση, ορίζεται ένας φασικός χώρος κατάλληλος για την εν λόγω μελέτη ο οποίος ονομάζεται *Signal Region*. Στη συνέχεια ο φασικός χώρος της Signal Region βελτιστοποιείται σε ότι αφορά τον διαχωρισμό μεταξύ σήματος και υποβάθρου προκειμένου να επιτευχθεί η καλύτερη δυνατή αναλογία σήματος έναντι του υποβάθρου. Κατά τη διαδικασία βελτιστοποίησης αξιοποιήθηκαν δείγματα προσομοιώσεων Monte Carlo. Η διαδικασία βελτιστοποίησης πραγματοποιήθηκε με τη χρήση διαφορετικών υπολογιστικών μεθόδων (Brute Force & Heuristic Particle Swarm Optimization), προκειμένου να διασταυρωθούν τα αποτελέσματα.

Σύμφωνα με τη Signal Region, κάθε μια από τις σημαντικές πηγές υποβάθρου (Non-resonant lepton pairs, 3-lepton και $Z+jets$), μελετήθηκε σε αντίστοιχες περιοχές στο φασικό χώρο (Ενότητα 4). Κάθε περιοχή έχει παρόμοια χαρακτηριστικά με της Signal Region, και ονομάζεται *Control Region*. Σε κάθε μία από τις Control Regions, αξιοποιούνται τόσο δεδομένα μετρήσεων (data) όσο και δείγματα προσομοιώσεων Monte Carlo και έχει οριστεί κατάλληλα ώστε να κυριαρχείται από ένα από τα υπό μελέτη υπόβαθρα. Σε κάθε μία από αυτές υπολογίζεται ο αντίστοιχος πολλαπλασιαστικός παράγοντας (*scaling factor*) ο οποίος λαμβάνει υπόψη του και διορθώνει την ασυμφωνία μεταξύ δεδομένων και Monte Carlo πρόβλεψης εκμεταλλευόμενος την αφθονία του εκάστοτε υποβάθρου στην περιοχή, προκειμένου να παρέχει μια ακριβή εκτίμηση αυτού.

Αφού υπολογιστούν όλοι οι πολλαπλασιαστικοί παράγοντες για κάθε σημαντική συνεισφορά υποβάθρου μέσω των Control Regions, αξιοποιούνται προκειμένου να παρέχουν μια ακριβή εκτίμηση του υποβάθρου στη Signal Region, διορθώνοντας την πρόβλεψη Monte Carlo για το υπόβαθρο. Κατ' αντίστοιχία με τις περιοχές υποβάθρου, ένας πολλαπλασιαστικός παράγοντας ms που ονομάζεται "παράγοντας σήματος" (Signal Strength), υπολογίζεται μέσω της Signal Region προκειμένου να ληφθεί υπόψη η ασυμφωνία μεταξύ δεδομένων και προσομοιώσης για το επιθυμητό σήμα.

Προκειμένου να αξιολογηθεί η αποτελεσματικότητα των διορθωτικών παραγόντων (scaling factors), ορίστηκε μια ανεξάρτητη περιοχή τόσο από τις περιοχές υποβάθρου όσο και από τη περιοχή σήματος, που ονομάζεται *Validation Region*. Στη περιοχή αυτή αξιοποιούνται όλοι οι πολλαπλασιαστικοί παράγοντες προκειμένου να διορθωθούν τα γεγονότα προσομοιώσεων και να αξιολογηθούν τα αποτελέσματα σε ότι αφορά τη συμφωνία τους με τα δεδομένα.

Ο παράγοντας σήματος ms είναι και αυτός που παρουσιάζει το μεγαλύτερο ενδιαφέρον καθώς μέσω αυτού μπορεί τελικά να πραγματοποιηθεί μια μέτρηση της ενερ-

γού διατομής του υπό μελέτη σήματος $ZZ \rightarrow \ell\ell\nu\nu$. Πριν τη μέτρηση της ενεργού διατομής, παρουσιάζεται η μελέτη των πηγών αβεβαιότητας του παράγοντα σήματος μ_S οι οποίες λήφθηκαν υπόψη (Sec.5).

Η τιμή για τον παράγοντα σήματος που λήφθηκε υπόψη για τη μέτρηση της ενεργού διατομής είναι $\mu_S = 1.006 \pm 0.041$ και είχε ως αποτέλεσμα τη μέτρηση $\sigma_{ZZ}^{fid} = 16.35 \pm 0.68 \text{ fb}$ για την ενεργό διατομή (Ενότητα 6).

Είναι σημαντικό να αναφερθεί ότι καθ' όλη την εκπόνηση της παρούσας εργασίας, το δείγμα δεδομένων στη περιοχή σήματος (Signal Region) αποκρύπτεται σε αντίθεση με τις περιοχές του φασικού χώρου που δημιουργήθηκαν με σκοπό τη μελέτη του υποβάθρου (Control Regions). Επομένως, η αντίστοιχη πρόβλεψη Monte Carlo για το σήμα στη Signal Region αξιοποιείται ως "ψευδό-δεδομένα" και έχει ως αποτέλεσμα τη τιμή $\mu_S = 1$ για τον παράγοντα σήματος.

Η εργασία τελειώνει με κάποια συμπεράσματα καθώς και αντίστοιχα σχόλια πέραν όσον αναφέρθηκαν κατά τη διάρκεια της μελέτης στις προηγούμενες ενότητες (Ενότητα 7).

Λέξεις-Κλειδιά: ATLAS, Παραγωγή μποζονίων Z, Ανάλυση Δεδομένων, Μελέτη Υποβάθρου, Μέτρηση Ενεργού Διατομής

Abstract

The present thesis concerns the study of methods for analyzing experimental data in order to distinguish signal from background in the production of two Z bosons. The study utilizes both Monte Carlo and data samples. The study is conducted through the identification of the channel $ZZ \rightarrow \ell\ell\nu\nu$ where the two Z bosons decay into 2 leptons and 2 neutrinos. The study utilizes proton-proton collision data at a center-of-mass energy of $\sqrt{s} = 13$ TeV collected by the ATLAS experiment at CERN throughout the second period (Run II) and corresponding to an integrated luminosity of 140.1 fb^{-1} .

The work begins with a brief description of the signal channel under investigation and the ATLAS detector (Sec.1). Subsequently, a brief description of the object and event selection from the detector, as well as the data and Monte Carlo simulation samples, is provided (Sec.2).

Regarding the study of the desired $ZZ \rightarrow \ell\ell\nu\nu$ signal channel (Sec.3), in order to make a precise estimation, a suitable phase space called *Signal Region*, is defined. The phase space of the Signal Region is optimized in terms of signal-to-background discrimination. Different computational methods (Brute Force & Heuristic Particle Swarm Optimization) were employed during the optimization process to cross-validate the results.

Based on the Signal Region phase space, each of the significant background sources (non-resonant lepton pairs, 3-leptons, and Z+jets) was studied in a corresponding suitable region in the phase space (Sec.4). Each region has similar characteristics to the Signal Region and is referred to as a *Control Region*. Each Control Region is defined to be dominated by one of the backgrounds under study. Both data and Monte Carlo simulation samples are used to perform the study. In each of these regions, a corresponding scaling factor is calculated to take under consideration and correct the discrepancy between data and Monte Carlo predictions.

Once the scaling factors have been estimated for every major background contribution, one from each Control Region, they are utilized to provide a precise estimation of the background in the Signal Region by correcting the Monte Carlo prediction accordingly. Similarly, a signal strength factor μ_S is calculated to account for the discrepancy between data and Monte Carlo prediction for the desired signal process.

To assess the effectiveness of the scaling factors, an independent region called *Validation Region* was defined, which is independent from both the background and signal regions. In this region, every scaling factor is applied to correct the simulation events, and the results are evaluated in terms of their agreement with the data.

The signal strength factor μ_S is of particular interest since it allows for a cross section measurement of the desired signal process $ZZ \rightarrow \ell\ell\nu\nu$. Before the cross section measurement, a study of the μ_S uncertainty sources taken under consideration is presented (Sec.5).

The value for the signal strength factor that was taken under consideration for the cross section measurement, is $\mu_S = 1.006 \pm 0.041$, resulting in a measurement $\sigma_{ZZ}^{fid} = 16.35 \pm 0.68 \text{ fb}$ for the cross section (Sec.6).

It is important to note that throughout this work, the data sample in the Signal Region is blinded, in contrast to the phase space regions created to study the background (Control Regions). Therefore, the corresponding Monte Carlo

prediction for the signal in the Signal Region is used as "pseudo-data," resulting in a value of $\mu_S = 1$ for the signal strength factor.

The work concludes with conclusions and remarks apart from the ones already made during the study in each particular section (Sec.7).

Keywords: ATLAS, Z Boson Production, Data Analysis, Background Study, Cross Section Measurement

Contents

1	Introduction	4
1.1	The ATLAS Detector	5
1.1.1	Overview of the ATLAS Detector	5
1.1.2	ATLAS Coordinate System and Observables	6
2	Object and Event Selection	8
2.1	Object Selection	8
2.1.1	Muons	8
2.1.2	Electrons	8
2.1.3	Jets	9
2.1.4	Missing Transverse Momentum	9
2.2	Data and Simulation Samples	10
2.2.1	Data Samples	10
2.2.2	Simulation Samples	10
2.3	Event Selection	11
2.4	Fiducial Phase Space	13
3	Study of the $ZZ \rightarrow \ell\ell\nu\nu$ Signal Process	15
3.1	Kinematic Variables Selection	15
3.2	Brute Force Optimization	19
3.3	Particle Swarm Optimization	21
3.4	Optimal Signal Region Definiton	26
4	Study of the Background	28
4.1	Control Regions	29
4.1.1	3 lepton Control Region (3ℓ CR)	29
4.1.2	Non-Resonant Control Region ($e\mu A$ & $e\mu B$ CR)	30
4.1.3	Z+jets Control Regions (Zjets0, Zjets1 & Zjets2 CR)	31
4.2	Direct estimation of Scaling Factors	33
4.3	Maximum Likelihood Analysis	39
4.4	Remarks on the Signal Strength Estimation	42
4.5	Validation Region	43
5	Study of Signal Strength Uncertainties	47
5.1	Uncertainty due to Data Statistics	47
5.2	Uncertainty due to MC Statistics	49
5.3	Uncertainty due to Control Regions	50
6	Fiducial Cross Section Measurement	53
7	Conclusion and Remarks	56
A	Parameter estimation using the Maximum Likelihood Method	58
A.1	The Maximum Likelihood Function	58
A.2	Uncertainty estimation assuming Gaussian Errors	59
B	Claiming discovery in a counting experiment	60
C	Signal Region Optimization with Boosted Decision Trees	63

D Particle Swarm Optimization Algorithm	68
D.1 Initialization	69
D.2 Search Space	69
D.3 Search Space Confinement	71
E Study of the Validation Region	74
F Event Yields Tables	77
F.1 Control Region Event Yields	77
F.2 Signal and Validation Region Event Yields	80
F.3 Control Region Variants Event Yields	81
G Regions as a function of kinematic variables	84
G.1 Additional Plots on Control Regions	84
G.2 Additional Plots on Signal Region	98
G.3 Additional Plots on Validation Region	99

List of Tables

2.1	Preliminary SR Definition	12
2.2	Fiducial Phase Space definition	13
3.1	Optimal Cuts on Preliminary SR	16
3.2	BRO SCAN I	19
3.3	BRO SCAN II	20
3.4	PSO Search Space I	24
3.5	PSO Scan Results	26
3.6	Optimal SR Definition	27
4.1	3ℓ CR Definition	30
4.2	Non-Resonant CR Definition	31
4.3	Zjets CR Definition	33
4.4	CR Dominant Background Abundances	34
4.5	Direct estimation of μ_S	38
4.6	MLE simultaneous fit results	42
4.7	Validation Region Definition	44
5.1	CR Variants	51
5.2	Summary of uncertainties on μ_S	52
6.1	Fiducial Volume signal yields	54
6.2	Quantities associated with cross section estimation	54
C.1	BDT Hyperparameters	63
C.2	Kinematic Variable Ranking with TMVA	66
E.1	Study of Validation Region Cuts	76
F.1	Event Yields in the 3ℓ CR	77
F.2	Event Yields in the NR CRs	78
F.3	Event Yields in Zjets CRs	79
F.4	Event Yields in the SR and VR	80
F.5	3ℓ CR Variants Yields	81
F.6	NR CR Variants Yields	82
F.7	Zjets CR Variants Yields	83

List of Figures

1.1 ZZ Production Feynman Diagrams	4
1.2 Layout of the ATLAS Detector	5
1.3 ATLAS Experiment Coordinate System	6
2.1 Representation of the ZZ signal process	11
2.2 Preliminary SR Pie Chart	12
3.1 Visualisation of a cut based on signal significance	16
3.2 Preliminary Signal Region vs Kinematic Variables 1	17
3.3 Preliminary Signal Region vs Kinematic Variables 2	18
3.4 PSO Algorithm Search Space	21
3.5 Progress of the PSO Algorithm	23
3.6 PSO Maximum Significance 2D Plots	25
3.7 Optimal SR Pie Chart	27
4.1 3-lepton CR Pie Chart	30
4.2 Non-resonant CR Pie Charts	31
4.3 Signal and Zjets Regions Plot	32
4.4 Z+jets CR Pie Charts	32
4.5 Visual Representation of μ_S estimation procedure	35
4.6 Direct Estimation CR vs $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$	36
4.7 Signal Region vs Optimization Kinematic Variables	37
4.8 Direct estimation scaling factor plot	39
4.9 Example of likelihood scan on μ_S	41
4.10 MLE simultaneous fit scaling factor plot	41
4.11 Validation Region Visualization	43
4.12 Validation Region Pie Chart	44
4.13 VR versus kinematic Variables 1	45
5.1 Uncertainty on μ_S due to Data Statistics plots	48
5.2 Uncertainty due to MC Statistics plots	50
6.1 Cross Section Measurement	55
A.1 Uncertainty obtained from $\ln\mathcal{L}$ plot	59
C.1 Preliminary Signal Region vs Kinematic Variables with TMVA . . .	64
C.2 Correlation Matrices	65
C.3 BTG Classifier Response	66
D.1 PSO Fundamental Criteria	68
D.2 Example of a PSO reaching a significance plateau	69
D.3 PSO Search Space Confinement	72
E.1 Validation Regions Study	74
E.2 Signal and background distribution in the VR	75
G.1 Non-Resonant and 3 ℓ CR vs E_T^{miss}	84
G.2 Z+jets CR vs E_T^{miss}	85
G.3 Non-Resonant and 3 ℓ CR vs E_T^{miss}/H_T	86
G.4 Z+jets CR vs E_T^{miss}/H_T	87
G.5 Non-Resonant and 3 ℓ CR vs $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$	88
G.6 Z+jets CR vs $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$	89
G.7 Non-Resonant and 3 ℓ CR vs $\Delta R_{\ell\ell}$	90
G.8 Z+jets CR vs $\Delta R_{\ell\ell}$	91
G.9 Non-Resonant and 3 ℓ CR vs p_T^Z	92
G.10 Z+jets CR vs p_T^Z	93
G.11 Non-Resonant and 3 ℓ CR vs n_{jets}	94

G.12 Z+jets CR vs n_{jets}	95
G.13 Non-Resonant and 3ℓ CR vs $\sigma(E_T^{miss})$	96
G.14 Z+jets CR vs $\sigma(E_T^{miss})$	97
G.15 SR versus kinematic variables	98
G.16 VR versus kinematic Variables 2	99

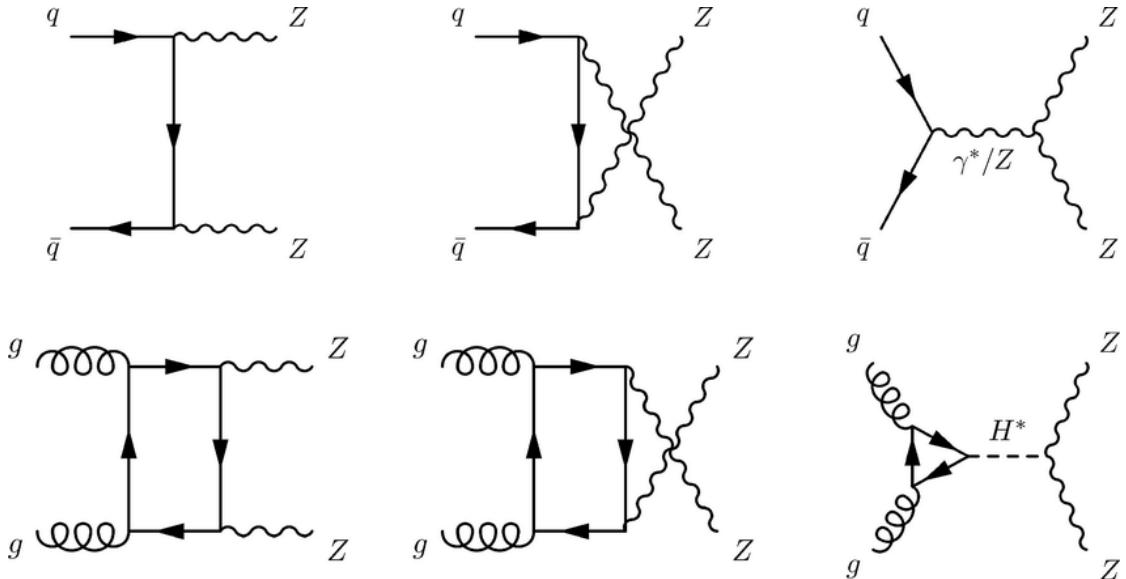
1 Introduction

The production of electroweak boson pairs provides a field for precision studies of the electroweak sector of the Standard Model (SM). The boson pair production can be exploited to further study the non-Abelian nature of the electroweak (EW) through their couplings, as well as their connection to the spontaneous symmetry breaking of the EW gauge symmetry.

Moreover, many phenomena predicted by theories beyond the SM may reveal themselves through the diboson production. Potential deviations from the predicted total and differential production cross sections may indicate a new resonance or coupling not predicted by the SM.

The production of two on-shell Z bosons has the smallest cross section among all diboson production processes. Hence, it is a very rare in the Standard Model (SM). However, the study of the ZZ channel is of high importance since it provides a leading field to investigate the *anomalous triple-gauge boson couplings* (aTGCs) and the off-shell study of the Higgs Boson. Moreover, the ZZ production appears as background to other diboson processes such as WZ.

Figure 1.1: [1] Lowest-order Feynman diagrams for the ZZ production process. The t-channel and u-channel quark-antiquark as well as the gluon-gluon processes, contribute to the ZZ production cross section. A ZZ boson pair can be produced via a Higgs boson as well (bottom right corner). The s-channel quark-antiquark diagram corresponds to an aTGC process not expected by the SM (top right corner).



Through the ZZ diboson production at the Large Hadron Collider (LHC), the electroweak sector of the SM can be studied and tested at the TeV energy scale. In the SM, Z boson pairs can be produced via non-resonant processes or in the decay of Higgs bosons. Any deviation from the SM expected cross section values could be indicative of new resonances decaying to Z bosons or other non-SM contributions.

In the Standard Model, the ZZ production happens via quark-antiquark t-channel and u-channel annihilation processes. The t-channel annihilation is the dominant process while the s-channel process does not exist in the Standard Model since it contains a neutral ZZZ or ZZ γ vertex. Moreover, the ZZ production

happens via loop-induced gluon-gluon fusion which contributes approximately 10% of the events for $\sqrt{s} = 13$ TeV.

The ZZ events are identified in the following two channels:

$$ZZ \rightarrow \ell^+ \ell^- \ell'^+ \ell'^- \quad \text{and} \quad ZZ \rightarrow \ell^+ \ell^- \nu \nu$$

where ℓ and ℓ' refer to leptons produced from different Z bosons with same or different flavor.

In the present work, the $2\ell 2\nu$ decay mode is the subject of study. The characteristic signature of the decay mode of interest $2\ell 2\nu$ is two high- p_T charged leptons and a large imbalance of transverse momentum between the initial and final state. This imbalance translates to missing transverse energy E_T^{miss} due to the neutrino pair passing through the detector without interacting with it at all.

The branching fraction of the $2\ell 2\nu$ decay mode is approximately six times larger than the 4ℓ decay mode. However, the channel of interest is characterised by large background contamination from numerous sources. Major background sources are Z+jets or WZ.

1.1 The ATLAS Detector

The ATLAS is a multipurpose particle detector with cylindrical geometry and forward-backward symmetry with respect to the interaction point. It extends 44 m in length and 25 m in height. In total it weighs 7000 tonnes and contains 100 million readout channels.

1.1.1 Overview of the ATLAS Detector

The detector consists of the Inner Detector (ID) surrounded by a superconducting solenoid, generating a 2 T axial magnetic field, electromagnetic and hadronic sampling calorimeters and a muon spectrometer (MS) with a toroidal magnetic field.

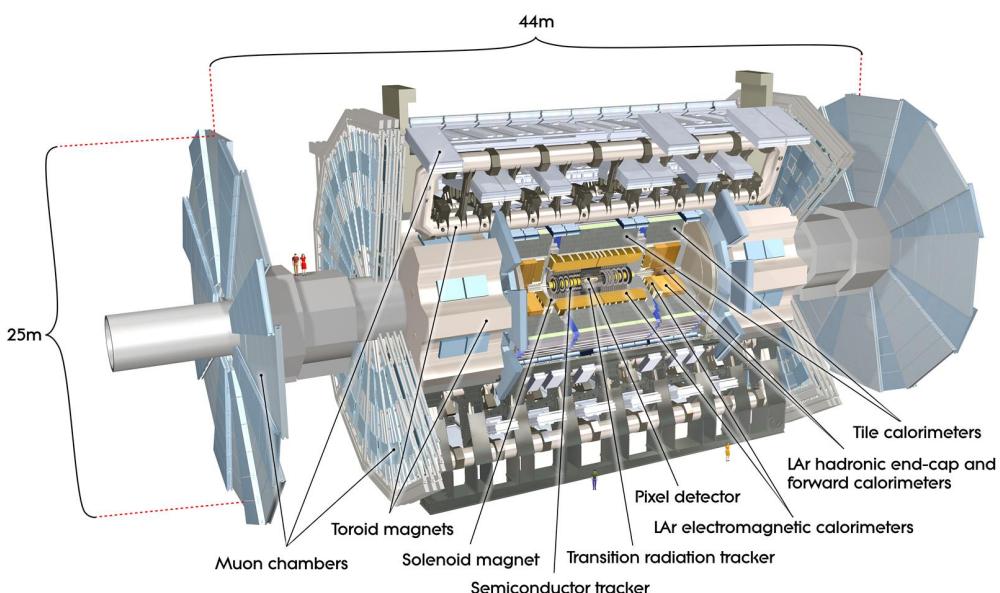


Figure 1.2: [2] Overall layout of the ATLAS detector and its components. It has a height of 25 m, a length of 44 m and weighs 7000 tonnes.

The Inner Detector (ID) is the innermost instrument of the detector enclosed within a cylindrical envelope of length ± 3512 mm and radius 1150 mm. The envelope is immersed in a 2 T axial magnetic field. The combination of the ID and the magnetic field provides precision tracking for charged particles in the pseudorapidity range $|\eta| < 2.5$. The ID consists of a silicon pixel detector, a silicon microstrip (SCT) detector and a straw tube transition radiation tracker (TRT). Utilising the combination of these parts allows for pattern recognition, momentum and vertex measurements, and electron identification. The electron identification is possible over the pseudorapidity range $|\eta| < 2.0$.

The calorimeter system consists of a number of electromagnetic and hadronic calorimeters with full azimuthal symmetry and coverage around the beam axis. The calorimeters closest to the beam-line consist of the barrel cryostat and two end-cap cryostats. The barrel contains the electromagnetic barrel calorimeter whereas, the two end-cap cryostats contain an electromagnetic end-cap calorimeter (EMEC), a hadronic end-cap calorimeter (HEC) and the forward calorimeter (FCal) each. The FCal is responsible for the pseudorapidity coverage of both electromagnetic and hadronic calorimeters up to $|\eta| < 4.9$. All these calorimeters use liquid argon as active material and lead as the absorber. The outer hadronic calorimeter consists of scintillator tiles as sampling medium and steel as absorber.

The outer part of the ATLAS detector consists of the muon spectrometer which is designed to detect and measure the momentum of charged particles exiting the barrel and end-cap calorimeters. It covers a pseudorapidity range of $|\eta| < 2.7$.

1.1.2 ATLAS Coordinate System and Observables

The ATLAS detector uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the center of the detector and the z-axis along the beampipe direction. The x-axis points from the IP to the center of the LHC ring. The y-axis points upwards.

For practical purposes, cylindrical coordinates (r, ϕ) are used in the plane $x - y$, transverse to the beam axis. The distance from the z-axis origin is represented by r while ϕ represents the azimuthal angle measured around the z-axis. In the transverse $x - y$ plane, the particle momentum and energy are measured by their transverse components p_T and E_T

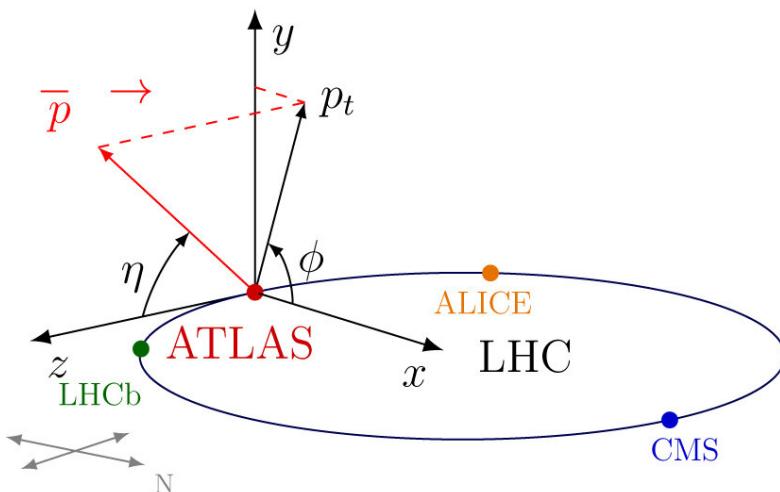


Figure 1.3: [3] Visualization of the coordinate system used by the ATLAS experiment.

Moreover, a polar angle θ is defined to express the angle between a particle and the beam axis. Using the polar angle θ an additional and rather important kinematic variable called *pseudorapidity* η is defined. The important property that characterises it is being *lorentz invariant*. Pseudorapidity is defined as follows:

$$\eta = -\ln(\tan(\theta/2)) \quad (1.1)$$

2 Object and Event Selection

2.1 Object Selection

Firstly, in order to remove non-collision background, events are required to contain at least one vertex formed from at least three associated tracks with $p_T > 0.4$ GeV. The vertex with the largest sum of p_T^2 computed from the associated tracks, is defined as the primary vertex.

2.1.1 Muons

Muons are reconstructed by matching the track in the muon spectrometer with the track in the inner detector (ID) using the hits of the ID track, the energy loss in the calorimeter and the hits of the track in the muon system. Baseline muons are required to have a transverse momentum $p_T > 7$ GeV and *loose* identification criteria [4]. After overlap removal, muons with $p_T > 20$ GeV are selected.

In order to suppress the contribution from cosmic and non-prompt muons, cuts on the muon impact parameters need to be applied. Specifically, a muon candidate is required to have a longitudinal impact parameter with respect to the primary vertex $|z_0 \cdot \sin\theta| < 0.5$ mm and a transverse impact parameter with respect to the primary vertex in significance units $|d_0/\sigma(d_0)| < 3$.

Muon candidates are required to be isolated from muons associated with jets. The isolation criteria are based on activity in the calorimeters and in the tracker respectively, within a cone of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.2$ around the muon. The isolation criteria used correspond to an isolation efficiency $\geq 99\%$ for muon with $p_T > 20$ GeV.

Finally, in order to account for effects in the detector resolution that are not well reproduced in MC samples, the transverse momentum of the muons is smeared and weights are applied to account for the difference in efficiency using software tools that provide the corresponding efficiency scaling factors as well.

2.1.2 Electrons

Electrons are reconstructed by matching the energy deposited in the EM calorimeters with the track signatures in the inner detector (ID).

The "medium" identification criteria for the central region $|\eta| < 2.7$ of the detector are used which are about 90% efficient for electrons with $p_T \approx 40$ GeV. The criteria are chosen according to calorimeter shower shape, ID track quality, calorimeter and tracking matching as well as the ratio of the energy measured in the calorimeter over the momentum measured in the tracker E/p [5]. The baseline electron candidates are required to have a transverse momentum of $p_T > 7$ GeV. After overlap removal, electrons with $p_T > 20$ GeV are selected.

In order to suppress the contribution from non-prompt electrons, a cut on impact parameters need to be applied. Specifically, candidate electrons are required to have a longitudinal impact parameter with respect to the primary vertex, $|z_0 \cdot \sin\theta| < 0.5$ mm and a transverse impact parameter with respect to the primary vertex in significance units $|d_0/\sigma(d_0)| < 5$.

Moreover, electrons are required to be isolated with respect to other tracks and calorimeter clusters. As for muons, a p_T dependent isolation within a ΔR cone requirement is applied which corresponds to an isolation efficiency $\geq 99\%$.

Finally, in order to account for differences between data and MC, energy calibration and smearing as well as the corresponding efficiency scaling factors are retrieved using software tools.

2.1.3 Jets

Jets are clustered from topological clusters in the calorimeters using the anti- k_t algorithm with radius parameter $R = 0.4$. After the jets are built, they are calibrated as described in [6].

The jets used in the analysis are required to have $p_T > 30$ GeV and to be reconstructed within the region $|\eta| < 4.5$. In order to reject jets originating from pileup

In order to further reduce the effect of pile-up, a cut on the jet-vertex-fraction (JVF) variable is applied for each jet. The JVF variable is defined as the scalar sum of the transverse momentum of the tracks associated to the jet that originates from the hard-scattering vertex, divided by the scalar sum of the transverse momentum of all tracks.

Jets are selected for analysis only if they pass the *loose* criteria for jet cleaning defined in [7]. These criteria correspond to an efficiency of selecting jets $\geq 99.5\%$ for $p_T > 20$ GeV.

Lastly, a veto on b-tagged jets with $p_T > 30$ GeV and $|\eta| < 2.5$ is applied in the analyses to reject $t\bar{t}$ contributions.

Additional criteria regarding the overlapping lepton tracks and jets have been considered in order to avoid reconstruction of multiple electron, muons and/or jets from the same detector signature. Electron candidates sharing an ID track with a selected muon are rejected, except if the muon is only calorimeter-tagged, in which case the muon is rejected instead. Electron candidates sharing their track or calorimeter cluster with a selected higher- p_T electron are rejected. Jets within $\Delta R = 0.4$ of a selected lepton are rejected as well.

2.1.4 Missing Transverse Momentum

The missing transverse momentum vector \vec{E}_T^{miss} is computed as the negative of the vector sum of the transverse momenta of all the measured and calibrated leptons and jets (*hard term*) as well as the tracks originating from the primary vertex but not associated with any of the leptons or jets (*soft term*). The soft-term is computed in a way minimising the pile-up impact in the E_T^{miss} reconstruction [8].

$$E_{x(y)}^{miss} = E_{x(y)}^{miss, e} + E_{x(y)}^{miss, \gamma} + E_{x(y)}^{miss, \tau} + E_{x(y)}^{miss, jets} + E_{x(y)}^{miss, \mu} + E_{x(y)}^{miss, soft} \quad (2.1)$$

The magnitude of the E_T^{miss} is then calculated by using the E_x^{miss} and E_y^{miss} components as follows:

$$E^{miss} = \sqrt{(E_x^{miss})^2 + (E_y^{miss})^2} \quad (2.2)$$

During the E_T^{miss} reconstruction process, the energy deposits are associated with every object with the following order: electrons (e), photons (γ), hadronically decaying tau-leptons (τ), jets and finally muons (μ). The total sum of these components comprises the hard term.

Any transverse momentum deposited but not associated with any of the aforementioned hard objects, is reconstructed as the soft term. During Run I the soft

term was reconstructed from calorimeter-based methods, known as the *Calorimeter Soft Term* (CST). However, during Run II, due to large pile-up contamination leading to bad E_T^{miss} resolution, the soft term is built exclusively from track-based methods, known as the *Track Soft Term* (TST). This way, improved PU rejection and better resolution is achieved.

2.2 Data and Simulation Samples

2.2.1 Data Samples

The present work analysis utilizes the full Run II data consisting of three different period data samples of proton-proton collisions at $\sqrt{s} = 13$ TeV collected by the ATLAS experiment during the period 2015 - 2018. The full Run II data sample corresponds to an integrated luminosity of 140.1 fb^{-1} . The data events are only selected if they pass the final *Good Run List* (GRL) released by the *Data Quality Group*.

A good E_T^{miss} measurement is critical for the extraction of the $ZZ \rightarrow l^+l^-\nu\nu$ signal given that the E_T^{miss} distinguishes the signal process of interest from the Drell-Yan background.

2.2.2 Simulation Samples

Signal Samples

The quark-quark and gluon-gluon QCD production of signal $\ell\ell\nu\nu$ processes are modelled using Sherpa 2.2.2 with the NNPDF3.0NNLO PDF set. Both the decays from Z and W bosons in the $\ell\ell\nu\nu$ final state, are included in the simulation.

The quark-quark EW production of signal processes $\ell\ell\nu\nu + \text{jets}$, is modelled using MadGraph with the NNPDF2.3LO PDF set. The Vector Boson Fusion (VBF) process is included in the simulation as well.

Background Samples

The diboson productions of $WW \rightarrow \ell\ell qq$ and $WZ \rightarrow \ell\ell qq$ are modelled using Powheg- Box with the CT10 PDF set. The diboson productions from $WZ \rightarrow \ell\ell\ell\nu$ are modelled with Sherpa 2.2.2 with the NNPDF3.0NNLOPDF. The WW are modelled with Powheg.

The VVV ($V = W,Z$) triboson productions, are modelled using Sherpa 2.2.2 with the NNPDF3.0NNLO PDF. The generation of $t\bar{t}$ events is modelled using Powheg-Box v2 with the NNPDF3.0 PDF. Single-top-quark events produced in Wt final states were generated with Powheg-Box v2 with the CT10 PDF set. Single top-quark production via s or t channels, was generated using Powheg-Box v1. This generator uses the four-flavour scheme for the NLO QCD matrix element (ME) calculations together with the fixed four-flavour PDF set CT10f4. The $t\bar{t}V$ processes were generated at LO.

The Z+jets process (with $Z/\gamma \rightarrow ee/\mu\mu/\tau\tau$) was modelled using Sherpa 2.2.1 with the NNPDF3.0NNLO PDF. The ME was calculated at NLO for up to two partons and at LO for up to four partons using Comix and OpenLoops.

The parton shower was modelled with Pythia 8.186 with the NNPDF2.3 PDF set and the A14 set of tuned parameters for all the samples except those from Sherpa, where parton shower was done within Sherpa.

The detector response for simulated events was simulated within a framework based on GEANT4. Furthermore, simulated events were processed with the same

reconstruction software used for data. In order to account for the different particle reconstruction efficiencies measured in data and simulation, correction factors are derived in dedicated measurements and applied to simulated events.

2.3 Event Selection

The first stage of event selection is to require data quality conditions by rejected LAr, Tile and SCT corrupted events as well as incomplete events. This rejection in both data and simulated events by rejecting events with at least one misidentified jet of non-collision background or detectors origin, with respect to the high-efficiency loose working point of selection jets. Moreover, events are required to have a primary vertex with at least two associated tracks as well as one lepton trigger.

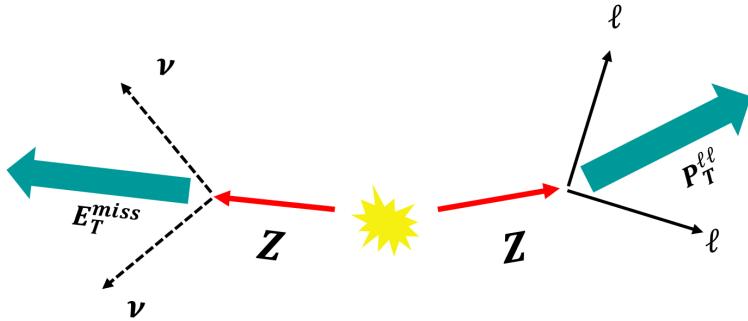


Figure 2.1: Visual representation of the desired ZZ signal process. The signature contains two high- p_T , same flavour, oppositely charged leptons and a large amount of missing transverse energy E_T^{miss} due to the neutrino system passing through the detector undetected.

In order to accurately identify as many signal events as possible, a suitable phase space must be defined following a *cut-based analysis*. That is, imposing cuts on certain kinematic variables that show potential in terms of separating the desired signal from the background events, in order to define a phase space maximally enhanced in signal events. This phase space is called *Signal Region* (SR).

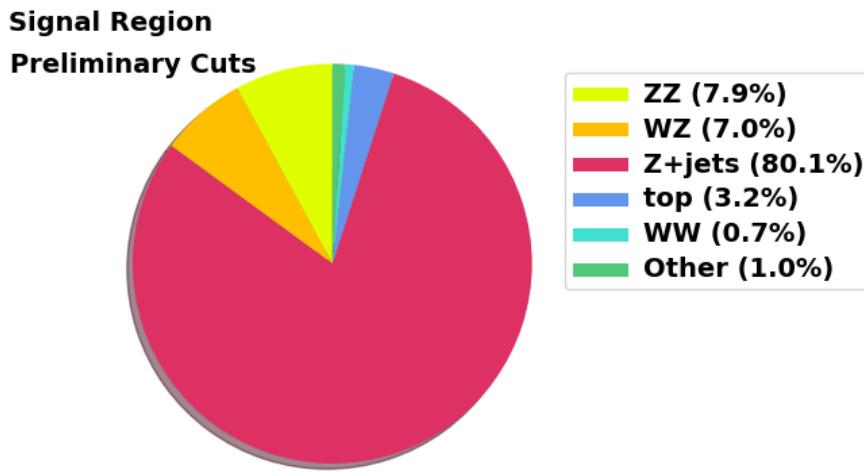
The process of interest $ZZ \rightarrow \ell^+\ell^-\nu\nu$ contains exactly two high- p_T , same flavour and oppositely charged leptons and a large amount of missing transverse energy E_T^{miss} due to the neutrino pair passing through the detector without interacting with it at all. Hence, events where additional leptons have been identified, are vetoed to reduce background contributions such as $ZZ \rightarrow \ell\ell\ell\ell$ and $WZ \rightarrow \ell\nu\ell\ell$. Additionally, a veto on b-jets is imposed in order to suppress the large top quark background contribution.

In order to suppress background contribution from non-resonant lepton pairs that do not include a Z boson such as $t\bar{t}$, $WW \rightarrow l\nu l\nu$, $Z \rightarrow \tau\tau$, a relatively strict mass window needs to be applied for the lepton pair invariant mass. Specifically, the lepton pair invariant mass m_{ll} needs to be compatible with the Z-boson mass that it originates from. For the channel of interest $ZZ \rightarrow \ell^+\ell^-\nu\nu$, the lepton pair is required to satisfy the condition $|m_{ll} - m_Z| < 15$ GeV resulting in a $76 < m_{ll} < 106$ GeV mass window. In practise, a more narrow mass window $80 < m_{ll} < 100$ GeV has been imposed.

Table 2.1: Preliminary selection criteria for the inclusive $\ell\ell\nu\nu$ process. The criteria correspond to the visual representation shown in Figure 2.2.

Inclusive $ZZ \rightarrow \ell\ell\nu\nu$ Criteria	
Variable	Criterion
Preselection	
Trigger	At least one lepton trigger
Vertex	At least one vertex with two or more tracks
Pair selection	
Third lepton	veto
b-jets	veto
Dilepton invariant mass	$80 < m_{ll} < 100$ GeV
Leading lepton p_T	$p_T^{\ell 1} > 30$ GeV
Subleading lepton p_T	$p_T^{\ell 2} > 20$ GeV
E_T^{miss}	$E_T^{miss} > 70$ GeV
$\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$	$\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss}) > 2.2$ radians
$\Delta R_{\ell\ell}$	$\Delta R_{\ell\ell} < 1.8$

Figure 2.2: Pie chart of the MC event yields corresponding to the preliminary Signal Region phase space defined by the criteria shown in Table 2.1.



Concerning the lepton pair, the Z boson is really massive meaning that high- p_T leptons are expected to originate from it. Consequently, a transverse momentum cut of $p_T^{\ell 1} > 30$ GeV and $p_T^{\ell 2} > 20$ GeV is applied for the leading and subleading lepton of the process, respectively.

Similarly, the neutrino pair originating from the other Z boson is expected to be highly energetic as well. As a result, a cut on missing transverse energy $E_T^{miss} > 70$ GeV, is imposed to further reject background contamination from Z + jets and non-resonant lepton pair background events.

Additional selection criteria have been imposed on angular variables. Since the ZZ system production is expected at a backward-forward topology, high values of the azimuthial angle difference $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$, between the lepton system and \vec{E}_T^{miss} , are expected. A cut has been set to $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss}) > 2.2$ radians to further reduce Z+jets contributions.

Finally, due to really high energy at the center of mass, the lepton pairs are ex-

pected to be highly boosted. Consequently, a small angular separation $\Delta R_{\ell\ell}$ in the $\eta - \phi$ plane is expected between them, expressed as $\Delta R_{\ell\ell} = \sqrt{(\Delta\phi_{\ell\ell})^2 + (\Delta\eta_{\ell\ell})^2} < 1.8$.

All the aforementioned criteria are presented in Table 2.1. These criteria form a preliminary SR phase space to study the signal process $ZZ \rightarrow \ell\ell\nu\nu$. However, the MC prediction resulting from this preliminary phase space, still contains mainly background events, with signal occupying only a $\sim 7.9\%$ of the phase space as shown in Figure 2.2. Consequently, further study of the Signal Region to optimize it in terms of signal-to-background discrimination is needed. That is, utilize different analysis methodologies to find the optimal cut values to impose on top of the preliminary SR cuts shown in Table 2.1 in order to further separate the signal from the background contamination. The Signal Region Optimization procedure is presented in section 3.

2.4 Fiducial Phase Space

The observations in the Signal Region are obviously subject to the various detector effects, reconstruction algorithms and selection criteria that can distort and bias the observed data. On the other hand, the cross section result needs to be more fundamental in the sense that it refers to a process produced by nature, regardless of detector or selection effects. Such effects may be, detector resolution, particles passing by regions of the detector that are not instrumented or efficiency in selection criteria. Hence, a measurement needs to refer to the same process produced by nature and predicted by theory in order to be compared to the theoretically predicted value.

Table 2.2: Fiducial phase space definition criteria for the $ZZ \rightarrow \ell\ell\nu\nu$ process. The E_T^{miss} is relaxed by ~ 1 RMS lower than the cut value at reconstruction level.

Fiducial Phase space criteria for $ZZ \rightarrow \ell\ell\nu\nu$
Leptons dressed with photons within $\Delta R(l, \gamma) < 0.1$
$ \eta^l < 2.5$
Truth $p_T^l > 30$ (20) GeV for leading (sub-leading) lepton
Veto on $\Delta R < 0.4$ between electrons and muons to remove electron overlap
$76 < M_{ll} < 106$ GeV
Truth $E_T^{miss} > 95$ GeV
Truth $\Delta R_{ll} < 1.8$
Truth $\Delta\varphi(E_T^{miss}, p_T^l) > 2.7$
Truth $E_T^{miss}/H_T > 0.65$

Consequently, by translating the observations made in the phase space defined by the SR to the *fiducial phase space* in *truth-level kinematics*, the observed quantities are corrected for the detector and selection effects. Hence, similar but different variables are utilized to define the fiducial phase space, in the sense that they are totally independent from the whole experimental procedure (Table 2.2). This way, a quantity corresponding to the process produced by nature is measured, and its value can be compared to the theoretically predicted one.

The fiducial phase space is defined as a region relatively close to the detector's event reconstruction cuts but larger. For the inclusive $ZZ \rightarrow \ell\ell\nu\nu$ process, the

fiducial phase space is defined with a more relaxed E_T^{miss} cut at ~ 1 RMS lower from the reconstruction level in order to account for the detector resolution. Moreover, dressed leptons are used for the fiducial phase space definition.

The entirety of factors contributing to the correction of the measurements in the SR used during analysis in relation to the fiducial phase space, can be represented by a correction factor. This factor can be denoted as C_{ZZ} and defined as the number of reconstructed signal events in the SR, N_{Signal}^{SR} over the corresponding signal events, $N_{Signal}^{fid.}$ in truth level kinematics:

$$C_{ZZ} = \frac{N_{Signal}^{SR}}{N_{Signal}^{fid.}} \quad (2.3)$$

The number of signal events in the fiducial volume is estimated using MC simulation samples of all the processes involved in the $pp \rightarrow ZZ \rightarrow \ell\ell\nu\nu$ production process. An accurate estimation of the C_{ZZ} is crucial in order to measure the *fiducial cross section* accurately (Sec.6).

3 Study of the $ZZ \rightarrow \ell\ell\nu\nu$ Signal Process

In order to provide an accurate measurement of the cross section of the desired $ZZ \rightarrow \ell\ell\nu\nu$ signal process, one needs to perform an optimization procedure of the Signal Region to achieve the highest abundance possible of the signal against the various background types. This enhancement of the signal in the SR is accomplished with a cut-based analysis. A study of the optimal cut values is performed and the additional cuts are imposed on top of the preliminary SR phase space defined in Table 2.1.

To optimize the SR, Monte Carlo simulation samples are utilized. Consequently, one uses the predicted value $s + b$ as the measurement, called *Asimov* data set. The quantities s and b represent the model prediction for the signal and background events respectively.

If the mean background b is sufficiently large and known with negligible uncertainty, one can use a test statistic called *median significance* based on the Asimov value $s + b$, to quantify the discriminating power of a variable cut value [9]:

$$Z = \sqrt{2 \left((s + b) \cdot \ln \left(1 + \frac{s}{b} \right) - b \right)} \quad (3.1)$$

This test statistic which is based on the model prediction value $s + b$ is the ultimate criterion used to discriminate signal from background in the present work. Higher significance values correspond to better signal-to-background discrimination. Throughout the rest of this work the median significance test statistic is referred as *Signal Significance*. A more elaborate description on claiming a discovery in counting experiments and the derivation of this formula is given in Appendix B.

For the inclusive ZZ process, both QCD $qq \rightarrow ZZ \rightarrow \ell\ell\nu\nu$ and $gg \rightarrow ZZ \rightarrow \ell\ell\nu\nu$ visualized in Figure 1.1 are considered in the signal model prediction. In addition, the rare EW $qq \rightarrow ZZjj \rightarrow \ell\ell\nu\nu jj$ process is considered.

3.1 Kinematic Variables Selection

As already mentioned, an additional set of cuts on top of the preliminary SR cuts shown in Table 2.1, is required to be imposed in the SR in order to further reduce the background contamination and maximize the signal significance.

In order to choose kinematic variables that are suitable for optimization, the SR signal and background events are plotted versus various variables (Fig.3.2 & 3.3). Following this methodology, a selection of the kinematic variables that can potentially have an impact in the signal-to-background discrimination procedure, can be achieved.

These plots provide a qualitative but really indicative representation of the discriminating power of each variable. For example, the variable E_T^{miss}/H_T (Fig.3.2b) exhibits good discriminating power.

In the bottom plots the signal significance given by equation (3.1) is presented. For a given bin, the signal significance is calculated by the integral of signal and background with the bin lower value as the lower integration limit, until the last bin (Fig.3.1).

The signal significance shown in plots (3.2) & (3.3), strongly depends on the choice of binning. For all the variables that are represented in GeV units, the same

binning has been chosen. The maximum significance corresponding to an optimal cut for each kinematic variable, is shown in Table 3.1. The results are in good agreement with the kinematic variable ranking obtained with the BDT analysis presented in Table C.2.

Figure 3.1: Visualisation of a cut on E_T^{miss}/H_T based on signal significance. The optimal cut value is the one for which the integrals of signal and background maximize the signal significance given by equation (3.1). In this example the optimal cut is $E_T^{miss}/H_T > 0.8$ visualized by the shaded area.

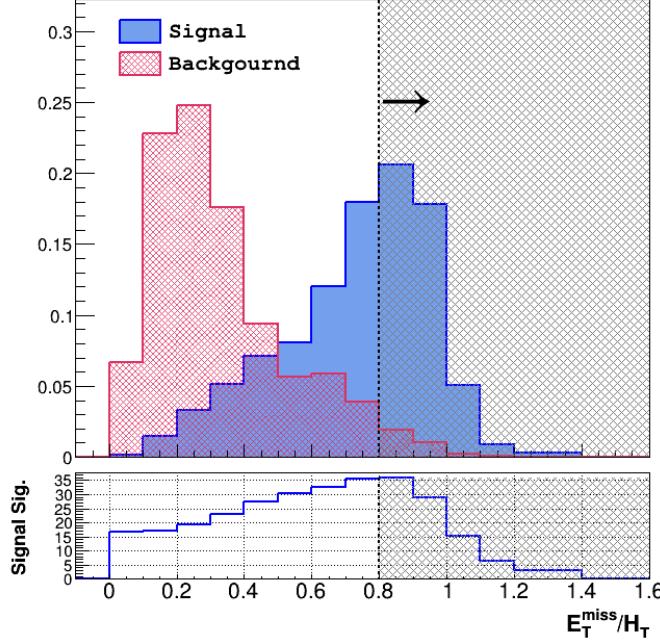


Table 3.1: Optimal cut values for the kinematic variables shown in Figures 3.2 and 3.3 in descending signal significance order. The results are in good agreement with the ranking results obtained with the BDT analysis presented in Table C.2. Only MC samples have been used.

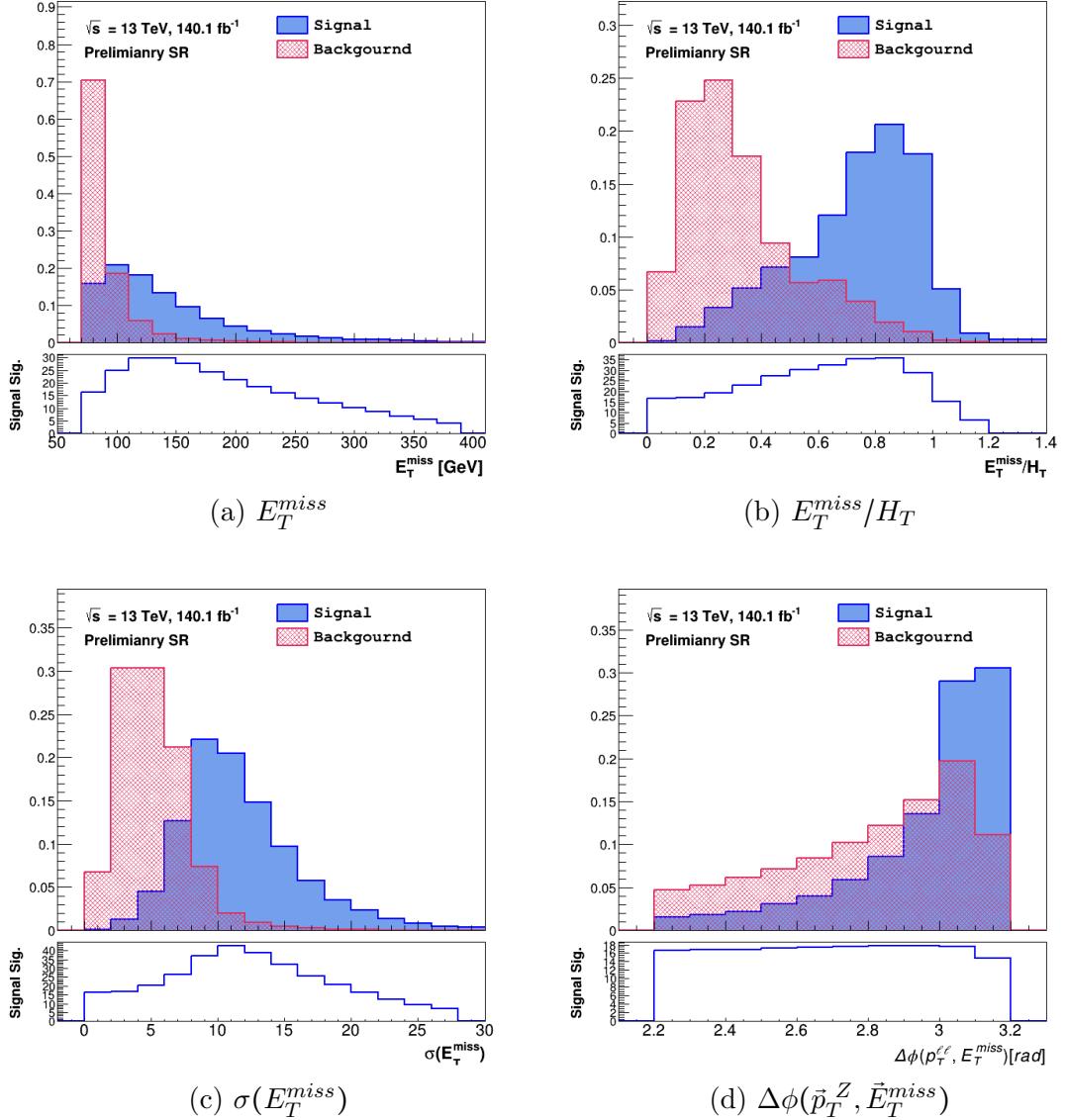
Variable	Optimal Cut	Significance
$\sigma(E_T^{miss})$	10	42.79
E_T^{miss}/H_T	0.8	35.95
E_T^{miss} [GeV]	130	30.34
$\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$	2.9	17.7
p_T^Z [GeV]	70	16.64
$p_T^{\ell 1}$ [GeV]	40	16.61
n_{jets}	0	16.60
$p_T^{\ell 2}$ [GeV]	20	16.59

The bin where the integral of the signal significance has a higher value, indicates the area near to which, a cut on the variable should be optimal. Moreover, variables with generally higher values of signal significance, are better discriminators of signal and background.

According to the remarks made, the variables of Figure 3.2, have better discriminating power as opposed to the variables of Figure 3.3. Additional confirma-

tion is obtained from the BDT analysis presented in Appendix C according to the kinematic variables ranking shown in Table C.2.

Figure 3.2: Plots for the kinematic variables E_T^{miss} , E_T^{miss}/H_T , $\sigma(E_T^{miss})$ and $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ in the Signal Region. The plots have been created using the preliminary SR defined by the cuts shown in Table 2.1. The bottom plots correspond to the integrated signal significance given by (3.1).

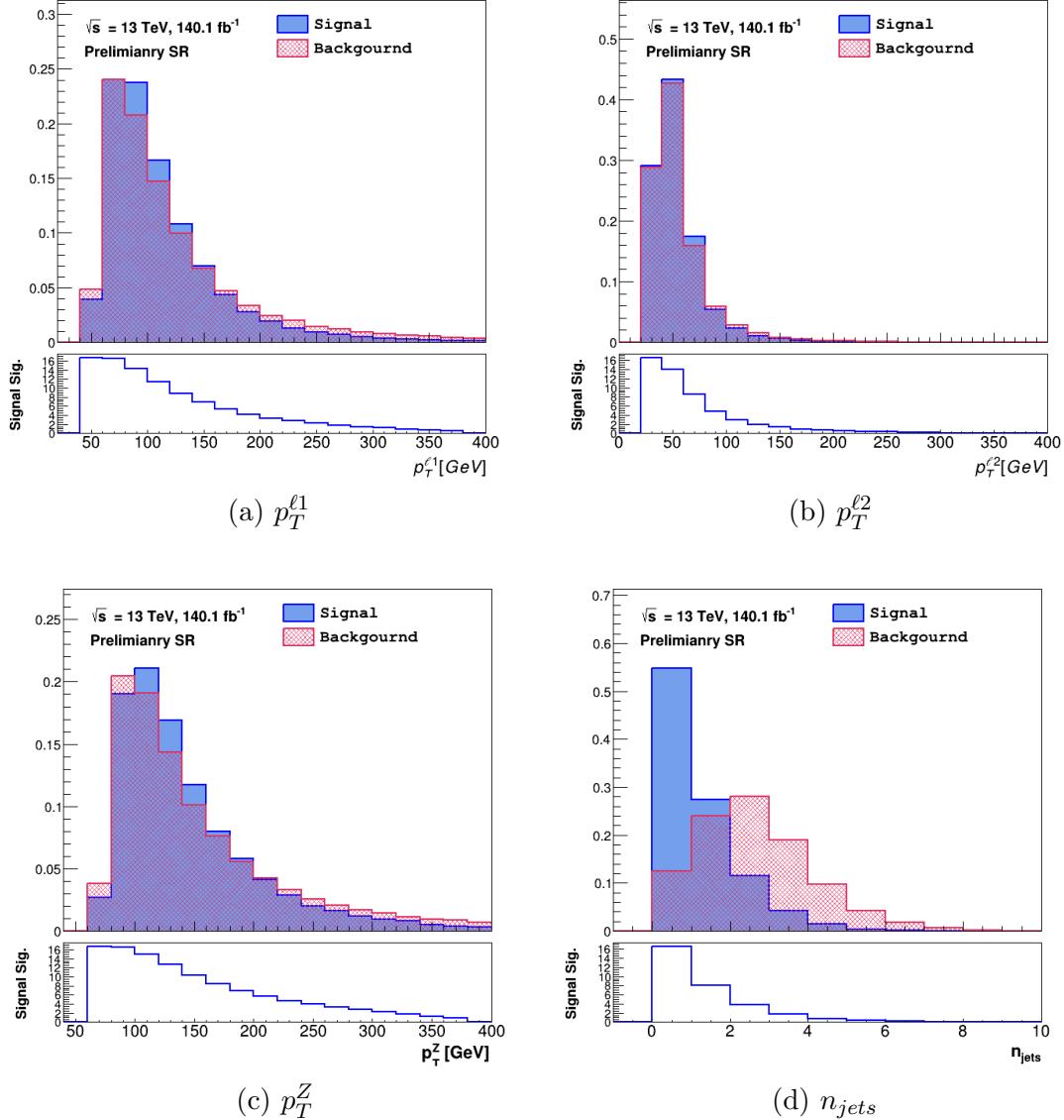


The plots of E_T^{miss} and its related kinematic variables E_T^{miss}/H_T and $\sigma(E_T^{miss})$ indicate high discriminating power in order to impose cuts that provide greater values of signal significance (Fig.3.2). However, the variables E_T^{miss} and $\sigma(E_T^{miss})$ are highly correlated (Fig.C.2). Additionally, the equivalent variable for $\sigma(E_T^{miss})$ has not been utilised in truth level kinematics in order to define the fiducial phase space (Table 2.2). Hence the $\sigma(E_T^{miss})$ is rejected in the optimization process.

On the other hand, the variables $p_T^{\ell 1}$, $p_T^{\ell 2}$ and p_T^Z do not show the same potential to discriminate between background and signal.

Lastly, the $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ variable is of particular interest since it is closely related to the desired ZZ signal process signature. However, the determination of a cut value may be difficult in terms of accuracy since it may need an analysis that

Figure 3.3: Plots for the kinematic variables $p_T^{\ell 1}$, $p_T^{\ell 2}$, p_T^Z and n_{jets} . in the Signal Region. The plots have been created using the preliminary SR defined by the cuts shown in Table 2.1. The bottom plots correspond to the integrated signal significance given by (3.1).



scans the $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ values with sufficient decimal accuracy (Fig.3.2).

In conclusion, the $p_T^{\ell 1}$, $p_T^{\ell 2}$ and p_T^Z variables do not show any potential to have great discriminating power. On the other hand the E_T^{miss} , E_T^{miss}/H_T and $\sigma(E_T^{miss})$ variables indicate a distinctive power in terms of signal to background separation but are highly correlated to each other. The $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ kinematic variable is of particular interest. Further confirmation of the variables studied in the current section, is given from the results in Figure C.1, obtained by utilizing the TMVA Machine Learning library presented in Appendix C.

Taking the aforementioned remarks under consideration, the variables that have been selected during the SR optimization process are the following:

1. E_T^{miss}
2. E_T^{miss}/H_T
3. $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$

Two different optimization methods are presented. The first is a simple optimization using brute force algorithms while the second is based a certain type of heuristic algorithms called *Particle Swarm Optimization* (PSO) algorithms. Both optimization procedures have been performed creating custom algorithms written in C++.

3.2 Brute Force Optimization

A simple way to make an estimation of the optimal set of kinematic variables that maximizes the signal significance in the SR, is by utilizing brute force techniques.

Specifically a brute force optimization (BRO) algorithm has been constructed in C++ in order to find the optimal set of kinematic variables. Following the BRO methodology, a different set of values is assigned to each kinematic variable that has been selected for optimization. The BRO algorithm simply scans over a set of possible combinations of values in the $(\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss}), E_T^{miss}, E_T^{miss}/H_T)$ search space, through iterative processes. Every set of values tried by the algorithm is imposed as a set of cuts on the SR.

Afterwards, the signal and background are calculated leading in a signal significance value Z , calculated according to equation (3.1). The set of kinematic variable values that results in the maximum signal significance, is the optimal set. This set of values is assigned to the kinematic variables and imposed as additional cuts on top of the preliminary cuts shown in Table 2.1.

However, one can easily understand that the BRO analysis can easily get out of hand in terms of time efficiency. For example, assigning 20 different values to every kinematic variable for 3 different variables, results in $20^3 = 8000$ combinations. For every combination, the signal, background and signal significance are estimated by iterating over files with millions of entries. Hence, a BRO analysis is followed by successively "zooming" in the areas that generate higher values of maximum significance starting from a relatively wide region motivated by the results in the no-cut SR (Fig.3.2 & 3.3).

Table 3.2: The search space used for BRO analysis and the corresponding full and N-1 selection results. The uncertainties presented are purely statistical. The notation used to indicate the search space points, refers to the lower and upper value and to the step used respectively.

BRO SCAN I	
Variable	Points in search space
$\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$	(2.2 : 3.1, 0.1)
E_T^{miss}	(95 : 150, 10)
E_T^{miss}/H_T	(0.5 : 0.9, 0.1)

Results				
$\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ [rad]	E_T^{miss} [GeV]	E_T^{miss}/H_T	Signal	Z
2.2	100	0.7	1739.7 ± 14.5	41.47
-	100	0.7	1739.7 ± 14.5	41.47
2.3	-	0.8	1470.5 ± 13.3	35.96
2.5	120	-	1701.9 ± 13.5	31.16

Firstly, a preliminary scan is carried. The points in the $(\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss}), E_T^{miss}, E_T^{miss}/H_T)$ search space, are selected based on the results summarized in Table 3.1. The "N-1 Cuts" method is followed in order to investigate the impact of each individual variable where "N" refers to the total number of cuts. The preliminary scan results are presented in Table 3.2.

As expected from the significance behavior shown in Figure 3.2d, neglecting the $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ cut has little impact on the signal significance. On the other hand, neglecting the other two cuts, clearly has an impact on the signal significance.

However, a more extensive scan is needed especially for the $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ which does not show any particular preference in terms of signal significance according to Figure 3.2d. Consequently a more thorough search has been carried out based on the results obtained from the preliminary scan (Table 3.2). The full and "N-1" selections as well as the corresponding results, are shown in Table 3.3.

Table 3.3: A more extensive the search space used for BRO analysis and the corresponding full and N-1 selection results. This scan has been carried out as an improvement to the preliminary BRO analysis presented in Table 3.2. The uncertainties presented are purely statistical. The notation used to indicate the search space points, refers to the lower and upper value and to the step used respectively.

BRO SCAN II					
Variable	Points in search space				
$\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$	(2.2 : 3.1, 0.05)				
E_T^{miss}	(90 : 110, 1)				
E_T^{miss}/H_T	(0.6 : 0.8, 0.01)				

Results					
$\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ [rad]	E_T^{miss} [GeV]	E_T^{miss}/H_T	Signal	Z	
2.70	101	0.74	1576.9 ± 13.8	41.72	
-	101	0.74	1610.8 ± 13.9	41.70	
2.70	-	0.76	1702.0 ± 14.3	36.92	
2.55	110	-	1970.5 ± 14.7	31.06	

Consequently, utilising a higher resolution search space, a set of variable cuts that attributes a higher significance value has been obtained. Specifically the optimal set of cuts in the search space $(\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss}), E_T^{miss}, E_T^{miss}/H_T)$, as well as the signal significance they attribute, is

$$(\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss}), E_T^{miss}, E_T^{miss}/H_T) = (2.70, 101, 0.74) \longrightarrow 41.72$$

The N-1 method, further confirms that the $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ variable has little impact in the Signal Region definition while neglecting the other variables clearly has impact in the SR definition.

Following the strategy described valuable time and resources can be saved. However, even by following this strategy, this method remains relevant because only three variables are utilised in the optimization process. If more variables are used in the analysis, it is practically impossible to obtain good results by performing a BRO analysis.

Consequently, a method based on a specific type of heuristic algorithm is studied, called *Particle Swarm Optimization* (PSO) algorithm. This method is utilized

with the same three variables as well, in order to further optimize the Signal Region, and at the same time, to serve as a cross-reference with the results obtained by the BRO method.

3.3 Particle Swarm Optimization

The Particle Swarm Optimization (PSO) algorithm, is a type of heuristic stochastic algorithm inspired by the social behaviour of swarms of birds. It has been proposed by Kennedy and Eberhart [10]. In PSO, a population of candidate solutions called *swarm*, represented as *particles*, explores a search space to find the optimal solution to a given problem. Each particle adjusts its position and velocity based on its own experience and the best solution found by the entire swarm. The PSO algorithm is widely used due to its fast convergence and effective performance using a small size of individuals as opposed to genetic algorithms that generally demand much higher values of individuals in order to have a sufficient performance. Still, a major drawback of PSO is getting stuck in local minima. A more elaborate description of the PSO algorithm and its technical features can be found in Appendix D.

A private code in C++ based on the standard PSO algorithm has been constructed in order to further optimize the Signal Region (SR). Every particle in the swarm, is represented by a set of values, one for each kinematic variable used for the SR optimization.

In the present work, the PSO analysis is carried out in the same 3-dimensional search space ($\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$, E_T^{miss} , E_T^{miss}/H_T) used for optimization in the BRO analysis (Fig.3.4). This choice has been made in order to further improve the BRO analysis results and cross-validate with them as well. In other words, every particle represents a set of cut values to be imposed in the Signal Region. Every time a set of cut values is selected by the PSO procedure, it is imposed on top of the preliminary SR criteria defined in Table 2.1.

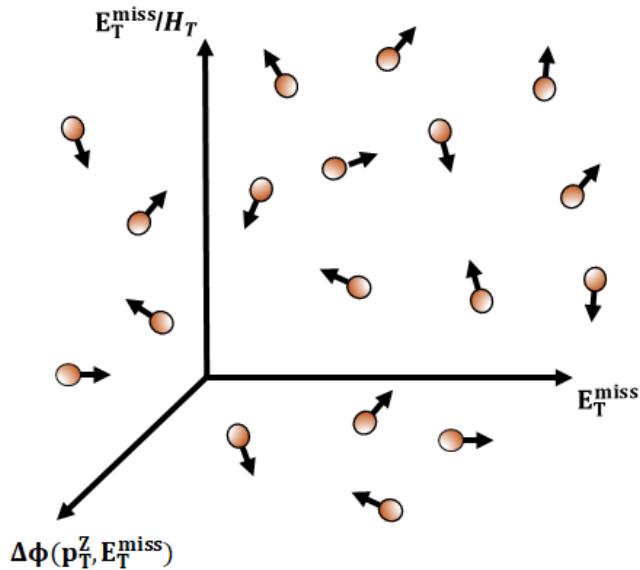


Figure 3.4: Visualization of a swarm of particles in the PSO Search Space.

Firstly, the particles are initialized randomly based on a uniform distribution function. For every particle, its set of cut values is applied as additional cuts in

the SR resulting in a corresponding number of background and signal events. For each particular set of signal and background values, the algorithm fitness value is calculated, i.e. the signal significance in the present work, according to the expression (3.1), same as in the BRO analysis. This process is repeated for every particle and a signal significance is calculated for each one of them. Afterwards, the personal best position of each particle as well as the global (swarm) best position is updated based on the signal significance. This process is repeated by updating the *position* and *velocity* for all the particles comprising the swarm.

For a particle i , the position and velocity are updated based on three fundamental criteria expressed by the following equation (D.2):

$$\begin{cases} \mathbf{v}_i^{t+1} \leftarrow \omega \mathbf{v}_i^t + c_{max} r_1 (\mathbf{p}_{best,i}^t - \mathbf{x}_i^t) + c_{max} r_2 (\mathbf{g}_{best}^t - \mathbf{x}_i^t) \\ \mathbf{x}_i^{t+1} \leftarrow \mathbf{x}_i^t + \mathbf{v}_i^{t+1} \end{cases}$$

where the symbols r_1 and r_2 denote two randomly generated numbers according to a uniform distribution $U(0, 1)$ and t is the iteration number.

The first component encourages the particle to follow its previous velocity \mathbf{v}_i^t . The second component encourages the particle to move towards its personal best position $\mathbf{p}_{best,i}^t$ meaning the set of variable values that attribute the maximum signal significance. The third, component encourages the particle to move towards the best position among all the particles \mathbf{g}_{best}^t .

If a new particle position generates a higher personal or (and) global signal significance, the $\mathbf{p}_{best,i}^t$ or (and) \mathbf{g}_{best}^t position, as well as the corresponding maximum significance, is updated.

The values of the ω and c_{max} parameters must be chosen carefully to balance *exploration* and *exploitation* in the optimization process.

The parameter ω is called inertial weight and it controls the influence of the previous velocity on the current velocity. A higher ω value encourages exploration meaning searching the search space for new maxima.

On the other hand, the parameter c_{max} encourages exploitation meaning exploiting the information already acquired by the particle itself and the other particles in the swarm. A higher value of c_{max} means that the particle focuses more on local maxima already acquired in the search space.

The algorithm terminates after no new global best position has been achieved for a substantial number of iterations resulting in a plateau (Fig.D.2).

The boundaries of the search space, have been restricted enough in order to avoid areas that attribute zero events since no significance can be calculated for zero events. Moreover, the algorithm kept getting stuck in the lower boundary of the $(\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss}))$ variable since there is no strong preference for a particular cut value (Sec.3.1) but this behaviour has been highly suppressed by modifying the velocity update formula (Appendix D.3).

Finally, like in the case of the (ω, c_{max}) parameter pair, there is no golden rule for the swarm size as well as the iteration number which results in a signal significance plateau. Numerous trials have been carried out and for a swarm size of $n = 40$ particles the process resulted in a plateau for a typical value of 40 iterations like in the example shown in Figure D.2.

Following the methodology described, numerous combinations of parameters and search spaces have been tried out in order to grow confidence on the algorithms' performance. In Table 3.4, two representative search spaces are presented.

Figure 3.5: Example of a PSO progress in the 2D Plane (E_T^{miss} , E_T^{miss}/H_T) of the SSI defined in Table 3.4. The swarm consists of 30 particles. As the scan progresses, the swarm converges towards a candidate maximum corresponding to a higher signal significance value. The PSO scan in the SR converged to the set of cut values (2.537, 101.5, 0.737) corresponding to 41.86 signal significance and attributing 1598 signal events.

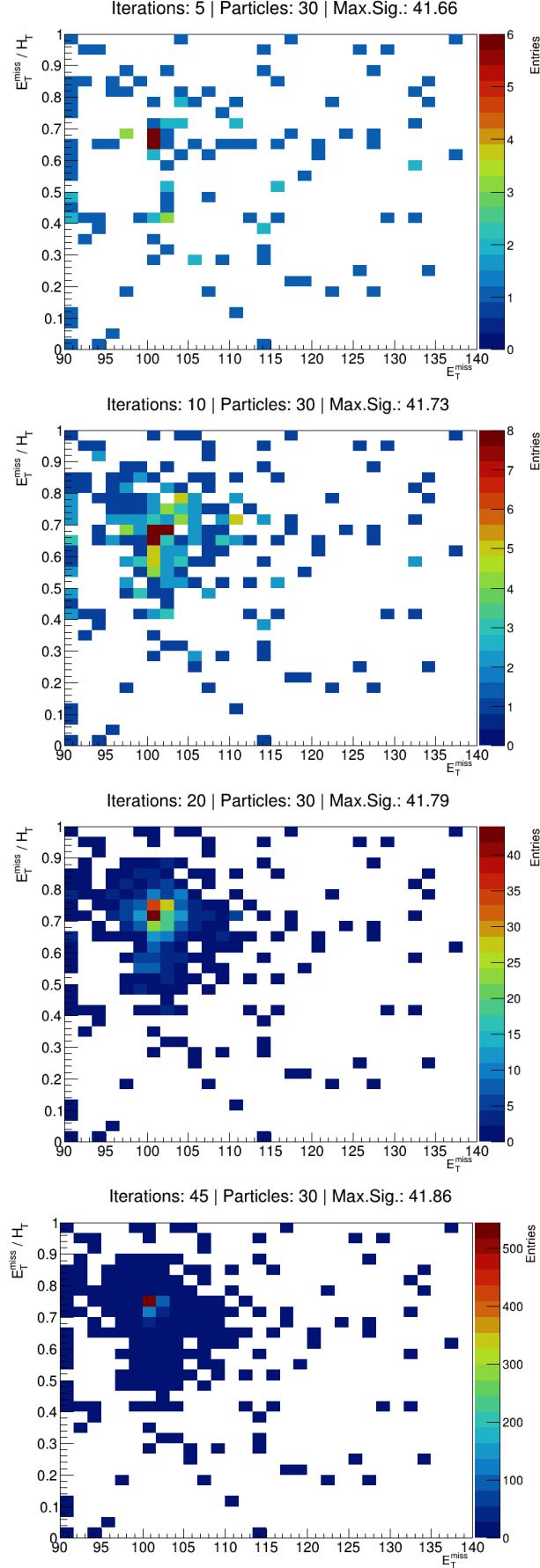


Table 3.4: Two representative search spaces (SS) used for PSO analysis. The SS I is relatively broad while the SS II is more restricted in order to focus in the area that generally generates higher signal significance values. The SS II area is motivated by the results obtained in SS I as well as the results obtained from the BRO analysis shown in Table 3.3. Every set of cuts is imposed on top of the preliminary SR cuts defined in Table 2.1.

PSO SEARCH SPACE I		PSO SEARCH SPACE II	
Boundaries		Boundaries	
$\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ [rad]	[2.2, 3.0]	$\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ [rad]	[2.2, 3.0]
E_T^{miss} [GeV]	[90, 140]	E_T^{miss} [GeV]	[96, 106]
E_T^{miss}/H_T	[0, 1.]	E_T^{miss}/H_T	[0.6, 0.8]

Firstly, a relatively broad search space (SS I) is presented. This SS I is motivated by the kinematic variables selection results shown in Figures 3.2. Results obtained using this search space are shown in Table 3.5. From the results, it can be deduced that there is a strong preference for optimal values $E_T^{miss} = 101.5$ GeV and $E_T^{miss}/H_T = 0.738$ since the algorithm converges to these values almost always in a relatively extensive search space. Any other values appearing occasionally, always result in lower significance values. As a result, through this process some confidence has been grown concerning these two variables. On the other hand, there is no strict preference for the $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ variable as expected from the BRO analysis. However, the values attributing better significance seem to be around the 2.7 rad area. An example of the PSO scan progress in SS I, is presented in Figure 3.5.

Afterwards, a search in a more restricted area is presented (SS II). This relatively restricted area is motivated by the results of the BRO analysis (Table 3.3) as well as the preliminary PSO analysis presented in Table 3.5. Due to the preliminary analysis results, more confidence has been grown in the E_T^{miss} and E_T^{miss}/H_T optimal cut values. For this reason, really strict boundaries have been chosen for these two variables. On the other hand, broader boundaries have been chosen for the $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ variable since various local maximum have been attributed although the maximum significance value appears around the 2.7 value. Results obtained in the more restricted SS II are presented in Table 3.5. Using a search space with more strict boundaries on the E_T^{miss} and E_T^{miss}/H_T variables, the results obtain are mostly the same.

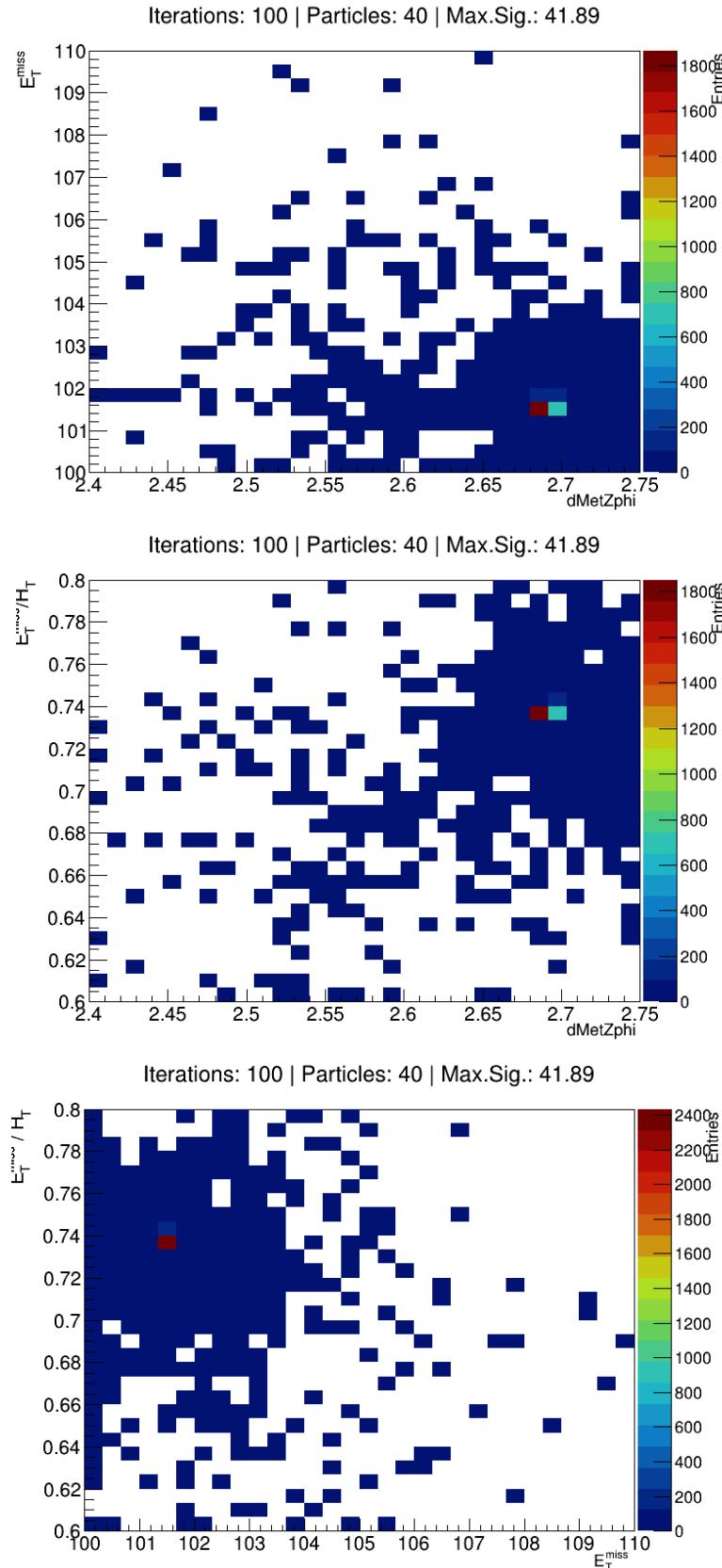
From the numerous results obtained, the best ones are included. According to the results obtained from both BRO and PSO analysis, the set of cut values imposed on top of the preliminary cuts (Table 2.1) to define an optimal Signal Region is

$$(\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss}), E_T^{miss}, E_T^{miss}/H_T) = (2.690, 101.5, 0.738)$$

where the values that are greater than the threshold cut values, are selected as the Signal Region phase space.

In conclusion, the results from both search spaces presented in this work, further confirm the BRO analysis result (Section 3.2), that the variable $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ does not have high impact on the Signal Region definition. The PSO algorithm has been utilized as a complementary method alongside the BRO method serving as means to cross-validate the results. Using a BRO algorithm for the three variables under study in the present work, is a much simpler method to follow. However, if

Figure 3.6: Plots from a PSO scan after 100 iterations with a swarm of 40 particles resulting in a maximum significance value of 41.895. The maximum significance corresponds to the best vector estimation ($\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$, E_T^{miss} , E_T^{miss}/H_T) = (2.690, 101.5, 0.738) which attributes a signal yield of 1578 when imposed on top of the preliminary SR criteria define in Table 2.1.



the study included more than 4 variables it would have been practically impossible in terms of time efficiency. In this case, a heuristic algorithm like the PSO is crucial in order to efficiently perform an analysis that includes many variables.

Table 3.5: Results in the two search spaces shown in Table 3.4 for various combinations of parameters. The configuration and results are presented in descending maximum significance order. Iteration numbers denoted with "*" refer to PSO scans terminating due to reaching a global significance plateau of 15 consecutive iteration points. The ω parameter values termed with "var" refer to a variable ω choice according to the expression D.2.

Results from the PSO Search Space I

iterations	swarm	ω	c_{max}	$\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ [rad]	E_T^{miss} [GeV]	E_T^{miss}/H_T	new $p_{best,i}$	new g_{best}	Signal	Z
80	20	0.7	1.47	2.6898	101.54	0.7376	1120	29	1577.2 ± 13.8	41.894
46*	20	0.7	1.47	2.6844	101.13	0.7376	177	12	1584.0 ± 13.9	41.863
45	30	0.7	1.47	2.5369	101.53	0.7375	349	26	1598.2 ± 13.9	41.859
79*	20	0.6	1.62	2.4724	101.53	0.7375	576	28	1603.3 ± 13.9	41.848
48*	20	var	1.47	2.6152	101.53	0.7376	153	12	1591.4 ± 13.9	41.803
69*	20	0.6	1.62	2.4177	101.51	0.6631	350	30	1784.7 ± 14.6	41.751
80	20	0.9	1.47	2.2009	101.17	0.6667	138	10	1792.5 ± 14.7	41.713

Results from the PSO Search Space II

iterations	swarm	ω	c_{max}	$\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ [rad]	E_T^{miss} [GeV]	E_T^{miss}/H_T	new $p_{best,i}$	new g_{best}	Signal	Z
200	40	0.8	1.47	2.6898	101.54	0.7375	677	34	1578.4 ± 13.8	41.895
150	40	0.8	1.47	2.6898	101.54	0.7370	545	25	1578.6 ± 13.8	41.883
45	40	0.7	1.47	2.6993	101.51	0.7375	395	23	1574.9 ± 13.8	41.882
82	40	0.6	1.62	2.3899	101.53	0.7375	1029	37	1605.9 ± 13.9	41.851
62	20	0.7	1.47	2.4962	101.54	0.7375	254	20	1601.9 ± 13.9	41.847
42	40	0.6	1.62	2.4288	101.54	0.7371	291	16	1609.3 ± 13.9	41.836
40	50	0.7	1.47	2.2206	101.92	0.7374	406	15	1786.2 ± 14.6	41.836
87*	20	0.7	1.47	2.4940	101.51	0.6578	355	22	1789.9 ± 14.6	41.747

3.4 Optimal Signal Region Definiton

In the present section two different analysis methods, Brute Force and Particle Swarm Optimization (BRO & PSO), have been carried out in order to optimize the Signal Region in terms of signal significance. That is, finding the set of optimal cut values in the SR, for which the discrimination between the desired inclusive ZZ signal process becomes maximum.

The BRO analysis is pretty simple and despite being potentially limited in terms of time efficiency, it proved really effective since only three variables have been used during the optimization process. The PSO algorithm, served as method to confirm the results obtained using the BRO method and to potentially improve these results. Effectively, no significant improvement has been achieved by the PSO algorithm but it has indeed confirmed the optimal set of cut values obtained from the BRO analysis. As a result best cuts obtained by the BRO analysis are used after further confirmation by the PSO analysis. As already mentioned, these cuts are imposed on top of the preliminary definition cuts presented in Table 2.1 resulting in the optimal phase space definition for the SR presented in Table 3.6.

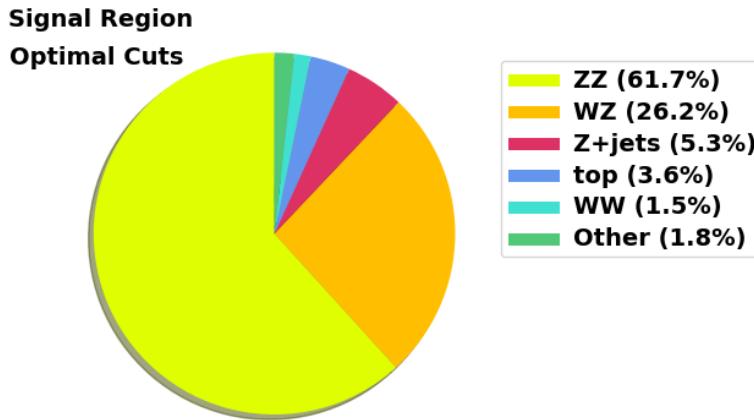
A comparison between the MC event yields of the preliminary and optimized signal regions shown in Figures 2.2 and 3.7, makes the impact of the optimization methodology followed in this section, evident. The event yields visualized in Figure 3.7 correspond to the optimized SR yields presented in Table F.4. The signal yield

Table 3.6: The optimal SR defined after the Signal Region optimization analysis. The results obtained by the BRO and PSO analyses, are added on top of the selection criteria presented in the preliminary SR definition shown in Table 2.1.

Inclusive $ZZ \rightarrow \ell\ell\nu\nu$ Optimal SR definition criteria	
Variable	Criterion
Preselection	
Trigger	At least one lepton trigger
Vertex	At least one vertex with two or more tracks
Pair selection	
Third lepton	veto
b-jets	veto
Dilepton invariant mass	$80 < m_{ll} < 100$ GeV
Leading lepton p_T	$p_T^{\ell 1} > 30$ GeV
Subleading lepton p_T	$p_T^{\ell 2} > 20$ GeV
E_T^{miss}	$E_T^{miss} > 101$ GeV
E_T^{miss}	$E_T^{miss}/H_T > 0.74$ GeV
$\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$	$\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss}) > 2.7$ rad
$\Delta R_{\ell\ell}$	$\Delta R_{\ell\ell} < 1.8$

for the desired process $ZZ \rightarrow \ell\ell\nu\nu$ in the optimized SR is 1577 ± 14 corresponding to a $\sim 61.7\%$ of the total MC event yield in the optimized SR.

Figure 3.7: Pie chart of the MC event yields in the Optimal Signal Region phase space defined by the criteria shown in Table 3.6. Comparing this pie chart to the one corresponding to the preliminary criteria SR phase space (Fig.2.2), makes evident the impact of the SR optimization results.



4 Study of the Background

The $ZZ \rightarrow \ell\ell\nu\nu$ as opposed to the 4ℓ channel, is characterized by large background due to the neutrino system originating from one of the Z resonances. In order to accurately estimate the signal, an accurate discrimination between signal and background must be achieved. Hence, in order to accurately estimate the signal yield, the various background types in the SR must be accurately studied as well. During the background study presented in this section, both data and Monte Carlo samples are utilized.

A region called *Control Region* (CR), suitable to study each major background separately, is created following the same logic used to define a signal region, in order to study it more accurately. Each CR phase space is defined with suitable criteria in order to be maximally enriched with the corresponding background which is to be studied in the CR. Moreover, the CR are required to be *SR-like* since the information provided by each CR is ultimately utilized in the SR to achieve a more accurate signal estimation. At the same time, more "loose" criteria are imposed in order to gain more statistics resulting in a more accurate study of each background.

In each CR both data and MC samples are utilized in order study the data-to-prediction discrepancy for the corresponding dominant background studied. This discrepancy is quantified by defining a corresponding factor called *Scaling Factor* for each major background expressing the discrepancy between the data and the model prediction. That is, the factor by which the model prediction needs to be corrected in order to accurately predict the data.

Ultimately, the information obtained for all the background types is finally utilized in the Signal Region (SR) by correcting each major background type in the SR with the corresponding scaling factors acquired from the background study procedure utilizing the CR.

There are three major background sources:

1. Events containing WZ boson production resulting in the process $WZ \rightarrow \ell\nu\ell\ell$.
In such processes, a lepton may have not been identified resulting in the misidentification of a true lepton pair alongside missing transverse energy mimicking the desired signal signature.
2. Events containing non-resonant lepton pairs (mainly from WW, Wt, t, $t\bar{t}$) originating from different resonances. In such processes, two true leptons may be identified alongside missing transverse energy. Specifically, in the process $WW \rightarrow \ell\nu\ell\nu$ where the signature is identical to the desired signal process studied in the present work. However, the charged leptons originate from different W resonances in contrast with the signal signature where they originate from the same Z resonance. These processes are highly correlated with events containing top quarks since WW events originate from them.
3. Events containing Z and jets which is similar to the desired signal signature since the jets are mismeasured resulting in momentum imbalance which is falsely identified as the presence of neutrinos passing through the detector.

A scaling factor is estimated for each background source described by defining a suitable CR for it. There are numerous other background sources. However, their contribution is really small resulting in insufficient statistics in order to study them separately.

Before utilising the scaling factors in the Signal Region, one can define a *Validation Region* (VR) in order to evaluate their impact. The VR must be a region defined in a completely phase space completely independent from the CRs and the SR. The scaling factors can be tried out in there in order to evaluate the Monte Carlo prediction with respect to the data sample before a scaling is carried out in the Signal Region (Section 4.5).

Finally, it is important to note that in the present work, the data sample in the SR is kept blinded. Data samples in the CR are utilized instead. Consequently, the corresponding MC prediction in the SR is entrusted as the "data" sample. Hence, the "data" in the SR are entirely based on the MC prediction.

4.1 Control Regions

In total, 7 regions have been created consisting of the Signal Region and 6 background (Control) Regions. As previously mentioned, the CRs have been created according to the major background contributions. The CR created to study the non-resonant background type has been splitted into to subregions to study WW and top events separately. Moreover, the CR concerning the Z+jets background, has been further splitted to three subregions according to the number of jets taking part in the event. The samples have been separated in the following background-specific CRs¹:

1. Control Region containing three-lepton events originating from WZ (3ℓ).
2. Control Region enriched in non-resonant lepton pairs originating from WW ($e\mu A$).
3. Control Region enriched in non-resonant lepton pairs originating from events containing top quarks ($e\mu B$).
4. Control Region containing a Z boson and zero jets (Zjets0).
5. Control Region containing a Z boson and one jet (Zjets1).
6. Control Region containing a Z boson and at least two jets (Zjets2).

A more extensive description of each Control Region definition is presented subsequently.

4.1.1 3 lepton Control Region (3ℓ CR)

Firstly the 3ℓ Control Region defined to study the background containing 3-lepton events, is presented. The 3ℓ CR, as any other region, is required to be SR-like since the 3-lepton background is studied in order to accurately estimate it in the SR resulting to a more accurate signal estimation. At the same time, the phase space is more "loose" compared to the SR phase space to attribute sufficient statistics. However, an additional criterion of requiring more than 2 leptons is imposed in order to create a region enriched in 3-lepton events to study them accurately. Moreover an additional cut $m_{TW} > 30$ GeV is imposed to further suppress Zjets

¹Big thanks to Angelos Tsiamis, PhD candidate at AUTh, for providing the Control Region samples ☺

contributions ². Hence, based on the SR phase space criteria (Table 3.6), a phase

Table 4.1: Definition of the 3 lepton Control Region. The corresponding event yields are presented in Figure 4.1 and Table F.1.

3ℓ Control Region
More than 2 leptons
$80 < M_{ll} < 100$ GeV
$E_T^{miss} > 70$ GeV
$E_T^{miss}/H_T > 0.3$
$\Delta R_{\ell\ell} < 2$
$\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss}) > 2.2$
b-jet veto
$m_{TW} > 30$ GeV

space enriched in 3-lepton events originating from WZ boson production, is constructed. The cuts imposed in order to define the 3ℓ CR, are summarized in Table 4.1. The resulting MC yields are shown in the pie chart presented in Figure 4.1. From the pie chart it is evident that the phase space defined for the 3ℓ CR is dominated by WZ events making it suitable to study the 3-lepton background type.

Figure 4.1: Pie chart of the MC event yields corresponding to the 3ℓ Control Region according to the definition criteria shown in Table 4.1.



4.1.2 Non-Resonant Control Region (e μ A & e μ B CR)

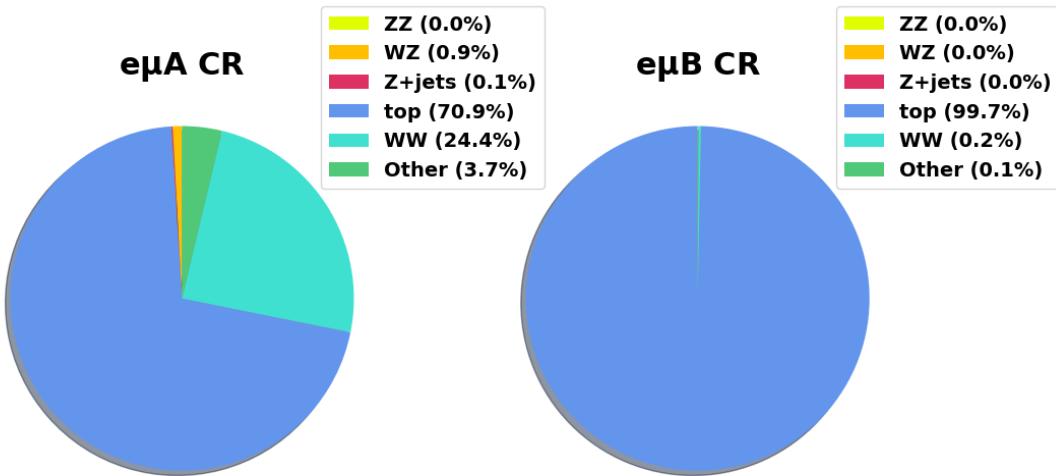
A phase space suitable to study background containing lepton pairs originating from different resonances, must be constructed as well. A similar SR-like phase space like the one used for 3-lepton events, is utilized. However, a veto on same flavor leptons is imposed instead of the more than 2 leptons requirement, in order for the phase space to be maximally enriched in leptons originating from different resonances. The result is a phase space containing WW and mainly top events ($t\bar{t}$, single-top and Wt events). Imposing a veto on events containing b-jets slightly

²The m_{TW} variable is defined as $m_{TW} = \sqrt{2p_T^{\ell 3} E_T^{miss} [1 - \cos(\Delta\phi(\vec{p}_T^{\ell 3}, \vec{E}_T^{miss}))]}$

Table 4.2: Definition of the Non-Resonant Control Regions $e\mu A$ and $e\mu B$ enriched in WW and top events respectively. The corresponding event yields are presented in Figure 4.2 and Table F.2.

NR Control Regions	
$e\mu A$	$e\mu B$
same flavor leptons veto	
$80 < M_{ll} < 100$ GeV	
$E_T^{miss} > 70$ GeV	
$E_T^{miss}/H_T > 0.3$	
$\Delta R_{ll} < 2$	
$\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss}) > 2.2$	
b-jet veto	$n_{bjets} > 0$

Figure 4.2: Pie charts of the MC event yields corresponding to the Non-resonant subregions according to the definition criteria shown in Table 4.2.



suppresses the top events and results in a larger WW portion. However, the top events still dominate the NR CR by a factor of ~ 3 compared to WW events. Consequently, the NR CR is splitted in two subregions according to the b-jet veto requirement and each background type is studied separately. In the first subregion called $e\mu A$, a b-jet veto is imposed resulting in a region containing an adequate amount of WW events to study them. In the second subregion called $e\mu B$, no b-jet requirement is imposed resulting in a region dominated by top events. The criteria to define the non-resonant CR are presented in Table 4.2.

The results are presented in Figure 4.2. It is evident that the $e\mu A$ CR is by no means dominated by WW events but the top events cannot be suppressed further.

4.1.3 Z+jets Control Regions (Zjets0, Zjets1 & Zjets2 CR)

The Z+jets background highly resembles the signal since the jets measurement produces fake missing transverse energy. Hence, the CR defined must be dominated by Z+jets events which are similar to signal but, at the same time to have as little signal events as possible. Obviously, there is a conflict between these two requirements. In the SR high values of both E_T^{miss} and E_T^{miss}/H_T are required. For this reason, to define the Zjets region, a combination of high values of E_T^{miss}

with low values of E_T^{miss}/H_T is selected as well as the opposite. More specifically the following requirement is imposed in order to define the Zjets CR:

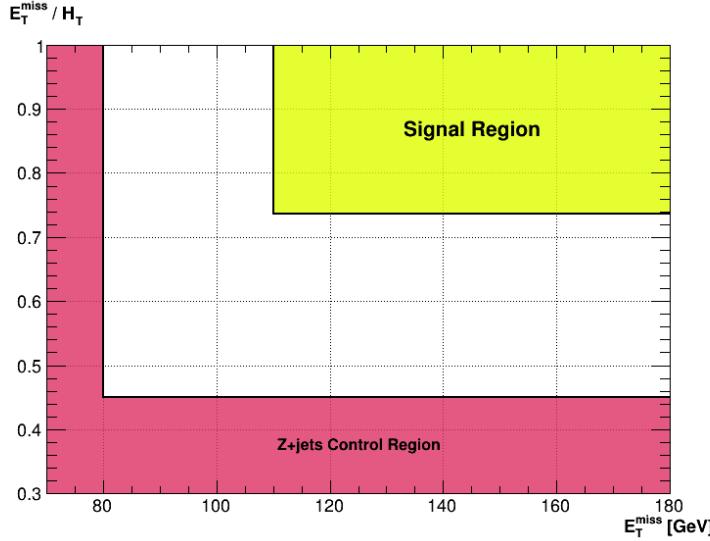
$$70 < E_T^{miss} < 80 \text{ GeV and } 0.3 < E_T^{miss}/H_T$$

or

$$80 < E_T^{miss} \text{ GeV and } 0.3 < E_T^{miss}/H_T < 0.45$$

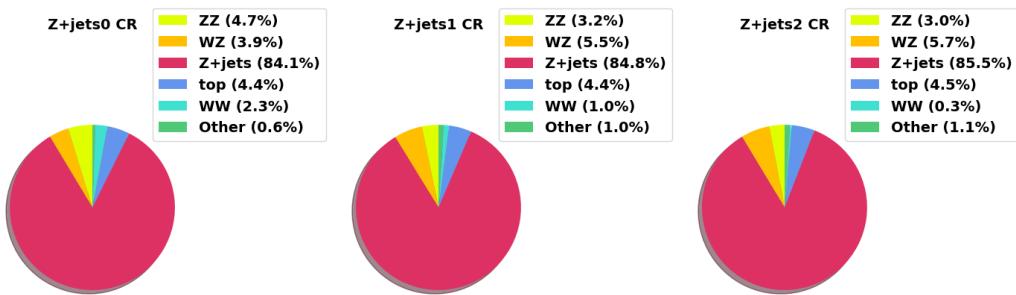
This selection of criteria attributes a characteristic "L-shaped" region in the

Figure 4.3: The Signal and Z+jets regions plotted in the 2-d (E_T^{miss} , E_T^{miss}/H_T). The cuts imposed on the Zjets Region ensure low signal contribution in the Z+jets control region.



E_T^{miss} , E_T^{miss}/H_T phase space as shown in Figure 4.3. Moreover, from further stud-

Figure 4.4: Pie charts of the MC event yields corresponding to the Z+jets subregions according to the definition criteria shown in Table 4.3.



ies of the Z+jets region, it has been deduced that the jet multiplicity has major impact on the MC prediction performance. For this reason, the Z+jets region has been further splitted as well, in three subregions according to the jets multiplicity. Firstly, a subregion called Zjets0 has been created by imposing a jet veto. Secondly, a subregion called Zjets1 has been created by requiring exactly one jet. Finally, a subregion called Zjets2 by requiring at least two jets.

The rest kinematic variables are kept similar to those used for the SR definition. The full selection of criteria in order to define the Z+jets CR, are summarized in Table 4.3.

Table 4.3: Definition of the three Z+jets Control Regions Zjets0, Zjets1 and Zjets2 according to their jet multiplicity. The event yields due to these criteria are presented in Figure 4.4 and Table F.3.

Zjets Control Regions		
Zjets0	Zjets1	Zjets2
$80 < M_{ll} < 100 \text{ GeV}$		
$(70 < E_T^{\text{miss}} < 80 \text{ GeV} \text{ and } 0.3 < E_T^{\text{miss}}/H_T)$		
or		
$(80 \text{ GeV} < E_T^{\text{miss}} \text{ and } 0.3 < E_T^{\text{miss}}/H_T < 0.45)$		
$\Delta R_{\ell\ell} < 1.8$		
$\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{\text{miss}}) > 2.2$		
b-jet veto		
$n_{\text{jets}} = 0$	$n_{\text{jets}} = 1$	$n_{\text{jets}} > 1$

In conclusion, by defining a Control Region enriched in every major background type, a more accurate study of each type is realised. Consequently, a more accurate signal-to-background discrimination in the Signal Region can be achieved leading to a more accurate signal estimation. In the following sections, these Control Regions are utilized following the methodology presented, to acquire the most accurate estimation possible, ultimately leading, to a cross section estimation.

4.2 Direct estimation of Scaling Factors

As already mentioned, the discrepancy between data and MC prediction for every major background is quantified by defining a corresponding scaling factor expressing the factor by which the MC prediction of each background must be corrected, in order to accurately predict the data.

A Control Region (CR) has been created for each major background in order to study it more accurately. In each CR both MC and data samples are utilized. Obviously, the higher the abundance of a CR in its dominant background, the more reliable the estimation of the corresponding scaling factor is.

Hence, the analysis strategy is the following: a successive estimation of the scaling factor of each CR is carried out according to the abundance of each CR in decreasing order. The CRs are accessed successively according to the descending abundance order shown in Table 4.4 and a scaling factor is obtained for each major background contribution. In each CR, the scaling factor is estimated by subtracting the rest (non-dominant) Monte Carlo events from the corresponding CR data sample.

The abundance of each CR can be expressed as the ratio of the dominant background over the total events sum in the CR:

$$\text{abundance} = \frac{\text{dominant background}}{\text{total events}} \quad (4.1)$$

Following the concept of (4.1), the abundance value for each Control Region is shown in Table 4.4. The values presented are calculated before any scaling is carried out.

Table 4.4: Control Regions abundance in their corresponding dominant background. The abundance of each CR is estimated before any scaling factor is calculated. The Non-resonant CR scaling factors are estimated first regardless of the $e\mu A$ CR having the lowest abundance since the regions are highly correlated to each other. The values are the result of the phase spaces defined in section 4.1. The corresponding event yields tables are presented in Appendix F.

Control Region	Abundance (%)	Scaling Factor
Non-Resonant		
$e\mu B$	$N_{top}/N_{non-top} = 99.7$	μ_{top}
$e\mu A$	$N_{WW}/N_{non-WW} = 24.4$	μ_{WW}
3-lepton		
3ℓ	$N_{WZ}/N_{non-WZ} = 89.0$	μ_{WZ}
Z+jets		
Z_{jets2}	$N_{Z_{jets2}}/N_{non-Z_{jets2}} = 85.5$	$\mu_{Z_{jets2}}$
Z_{jets1}	$N_{Z_{jets1}}/N_{non-Z_{jets1}} = 84.8$	$\mu_{Z_{jets1}}$
Z_{jets0}	$N_{Z_{jets0}}/N_{non-Z_{jets0}} = 84.1$	$\mu_{Z_{jets0}}$

Finally, all the scaling factors obtained from the CRs, are utilized in the Signal Region in order to account for all the major background contributions. The corrected, background contributions can now be subtracted from the SR data sample in order to acquire a more accurate estimation of the signal.

According, to Table 4.4, the most reliable scaling factor to estimate is the scaling factor μ_{top} , since the $e\mu B$ is the CR with the highest abundance in its dominant background. The top background consists of single top, $t\bar{t}$ and Wt events. Hence, starting from the $e\mu B$ CR which is enriched in top events, all the remaining events are subtracted from the $e\mu B$ data sample in order to make a data-driven estimation of the total number of top events:

$$N_{top}^{Data, est} = N_{e\mu B}^{Data} - N_{e\mu B, non-top}^{MC, est} \quad (4.2)$$

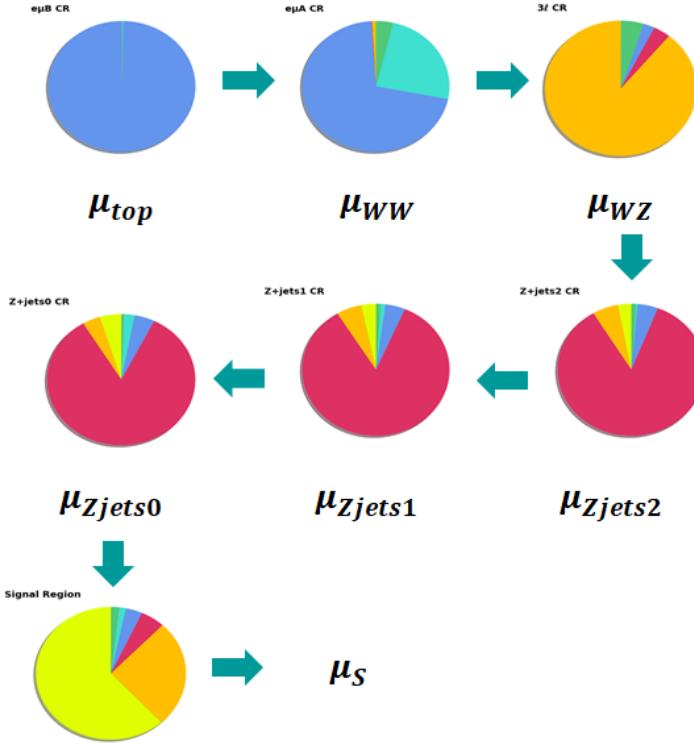
The scaling factor μ_{top} can now be calculated for the $e\mu B$ CR accounting for the discrepancy between the data-driven and the MC estimation of the top events:

$$\begin{aligned} \mu_{top} &= \frac{N_{top}^{Data, est}}{N_{e\mu B, top}^{MC, est}} \\ &= \frac{N_{e\mu B}^{Data} - N_{e\mu B, non-top}^{MC, est}}{N_{e\mu B, top}^{MC, est}} \end{aligned} \quad (4.3)$$

Afterwards, the scaling factor μ_{WW} concerning the $e\mu A$ subregion is calculated which it is highly correlated with the $e\mu B$ CR, since it is a part of the Non-Resonant CR.

In the same manner, a scaling factor μ_{WZ} is estimated in the 3ℓ CR followed by the scaling factors of the rest CRs according to the descending abundance order of Table 4.4 .

Figure 4.5: Visual representation of the procedure followed in order to estimate the scaling factors according to the corresponding CR abundance presented in Table 4.4. The CR regions are utilized in order to make a data-driven estimation of each background contribution in order to ultimately utilize it in the SR to accurately estimate the "signal strength" factor μ_S .



As soon as the value of a scaling factor has been obtained, it can be exploited to account for the corresponding background type in the next CR. For instance, the μ_{top} scaling factor can be exploited in order to correct the top contribution in the $e\mu A$ CR to achieve a more accurate estimation of the μ_{WW} scaling factor. In the same manner, both the μ_{top} and μ_{WW} can be exploited to scale the top and WW contributions in the 3ℓ CR and so on.

Finally, in the Signal Region, all the scaling factors are utilized in order to scale the MC events of all the major background types. Subtracting all the scaled background contributions from the SR data sample attributes a data-driven estimation of the Signal Yield for the desired $ZZ \rightarrow \ell\ell\nu\nu$ process:

$$\begin{aligned}
 N_{Signal}^{Data, est} = & N_{SR}^{Data} - N_{SR, top}^{MC, est} \cdot \mu_{top} \\
 & - N_{SR, Zjets2}^{MC, est} \cdot \mu_{Zjets2} - N_{SR, WZ}^{MC, est} \cdot \mu_{WZ} \\
 & - N_{SR, Zjets1}^{MC, est} \cdot \mu_{Zjets1} - N_{SR, Zjets0}^{MC, est} \cdot \mu_{Zjets0} \\
 & - N_{SR, WW}^{MC, est} \cdot \mu_{WW} - N_{SR, other}^{MC, est}
 \end{aligned} \tag{4.4}$$

The N_{other} events refer to all the sum of all the contributions suffering from insufficient statistics to estimate a scaling factor for.

As soon as the signal yield has been estimated using the expression (4.4), the *Signal Strength* μ_S can be estimated in the same manner as in the relation (4.3):

$$\mu_S = \frac{N_{Signal}^{Data, est}}{N_{Signal}^{MC, est}} \tag{4.5}$$

Figure 4.6: Plots of the Control Regions as a function of $\Delta\phi(p_T^Z, \vec{E}_T^{miss})$ utilized for the direct estimation procedure. In each particular region the background for which, a scaling factor has been obtained, is corrected according to the direct estimation procedure visualized in Figure 4.5. The bottom plots show the data-over-prediction ratio for each bin. A fit function is fitted to the ratio points (red line). A line compatible with zero slope expresses consistent MC prediction performance along the variable range.

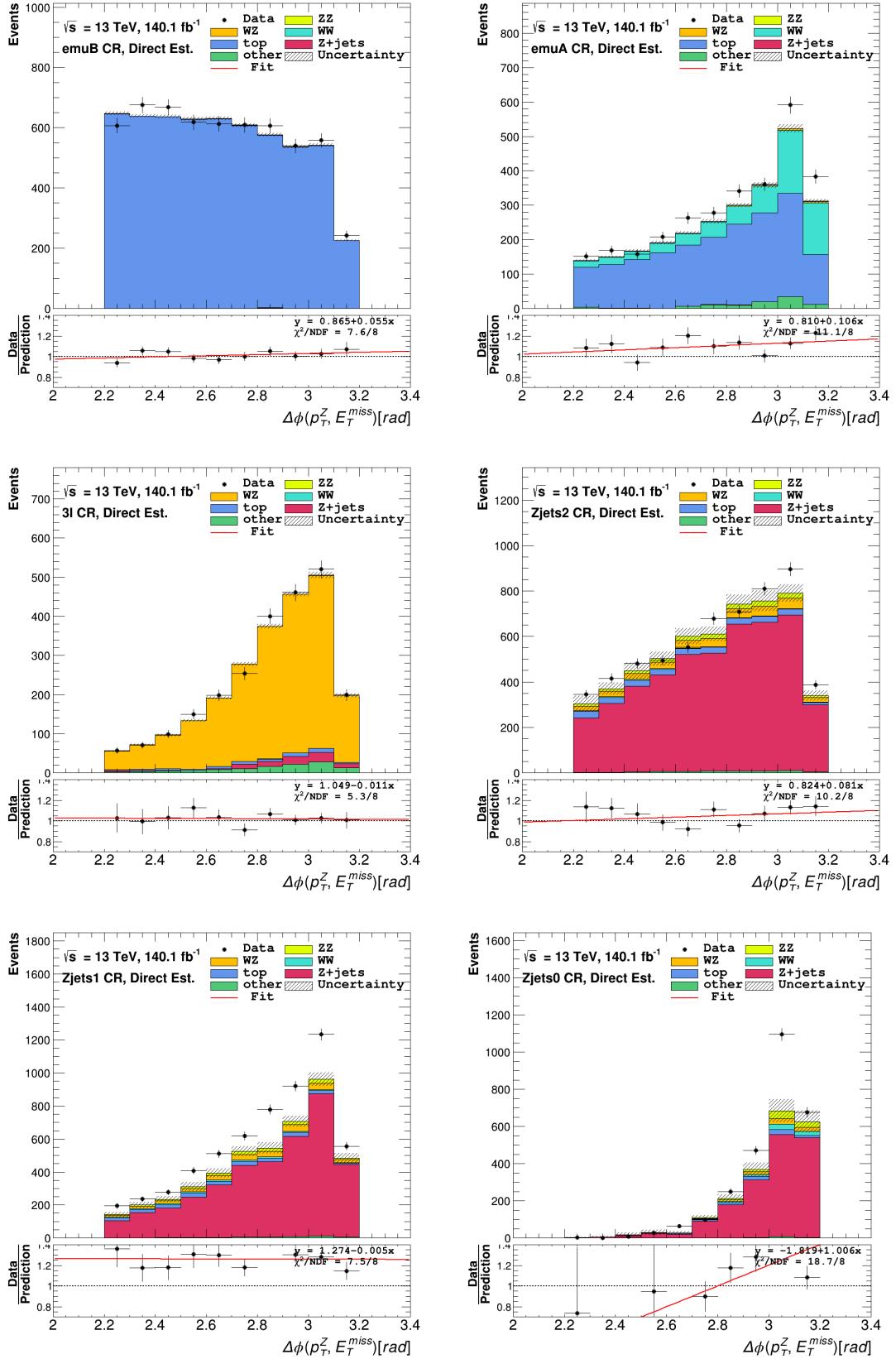
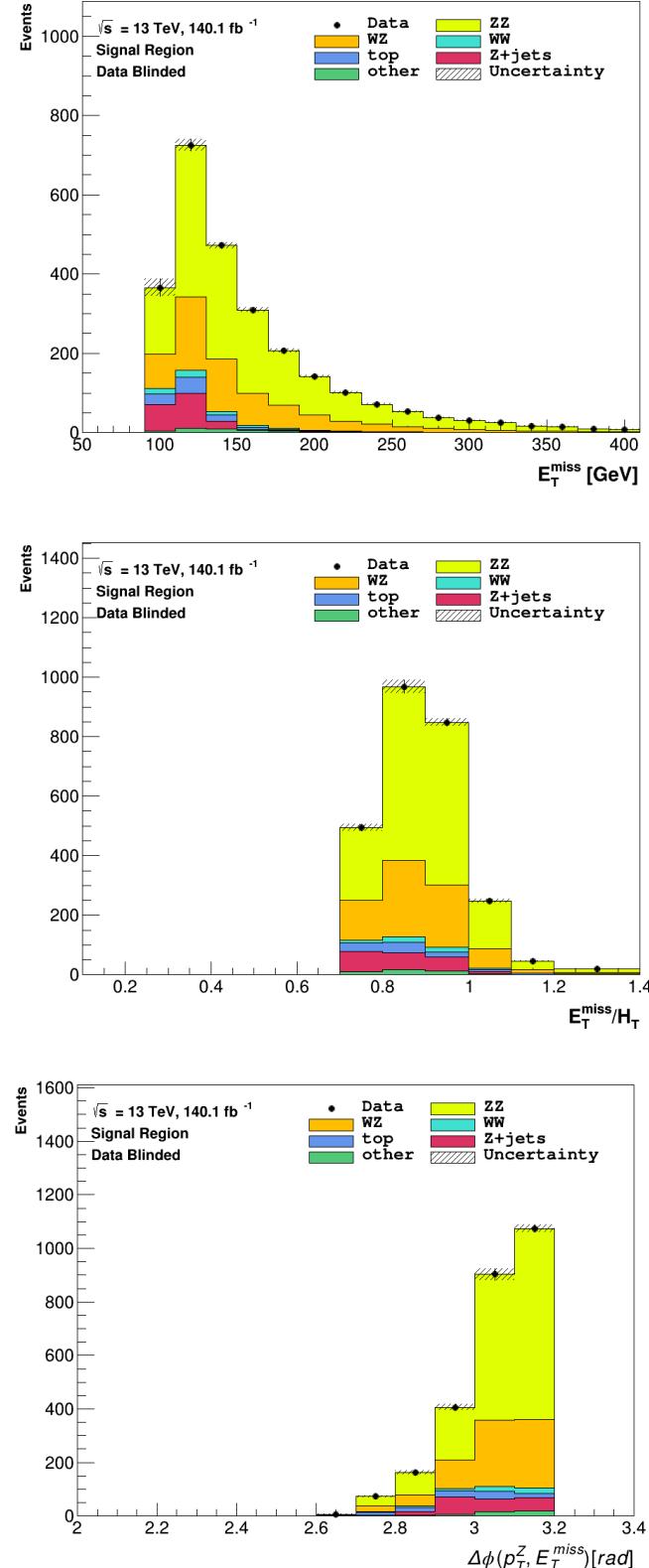


Figure 4.7: Plots of the optimal Signal Region defined in section 3.4, as a function of E_T^{miss} , E_T^{miss}/H_T and $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$. The plots provide a visual representation of the event yields of the desired signal process $ZZ \rightarrow \ell\ell\nu\nu$ along the kinematic variables used for optimization. Since the total scaled MC events are used as pseudo-data due to the real data sample being blinded in the SR, the simulation events predictions are identical to the corresponding data points.



However, it is really important to mention that in the present work, the SR data sample is blinded. Consequently, the total scaled MC yield in the SR is considered as the data sample instead. Hence, a trivial value $\mu_S = 1$ is expected for the signal strength according to equation (4.5).

In Figure 4.6, the Control Regions utilized for the direct estimation procedure, are plotted as a function of the $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ kinematic variable. In each plot the background types for which a scaling factor has been already obtained, are scaled in accordance with the direct estimation procedure described. The bottom plots show the data-over-prediction ratio for each bin. A fit function is fitted to the ratio points (red line). A line compatible with zero slope expresses consistent MC prediction performance along the variable range. By utilizing all the scaling factors afterwards, a slope closer to zero is expected. The scaling factor performance is evaluated by utilizing them in a new, independent and SR-like region as presented in section 4.5. Additional plots of the scaling factors performance in the CR as well, can be found in Appendix G.

The CR plots provide a qualitative visual representation of the extent to which, the dominant background in each particular region needs to be corrected by the corresponding scaling factor in order to accurately predict the data represented by the black dots. Specifically, the WW, Zjets1 and Zjets0 MC background predictions show the largest discrepancy with the data according to the corresponding CR plots. Consequently, they must be corrected by a higher scaling factor value.

Concerning the Signal Region, the plots presented in Figure 4.7, provide a visual representation of the distribution of the signal events $ZZ \rightarrow \ell\ell\nu\nu$ versus the kinematic variables used for optimization in section 3. The cuts imposed after the optimization procedure on these variables (Table 3.6), are obvious in the plots. It has been already mentioned that the total scaled MC events are used as pseudo-data since the real data sample in the SR is kept blinded. Consequently, the simulation events perfectly predict the "data" points comprised by the the total scaled MC events as well.

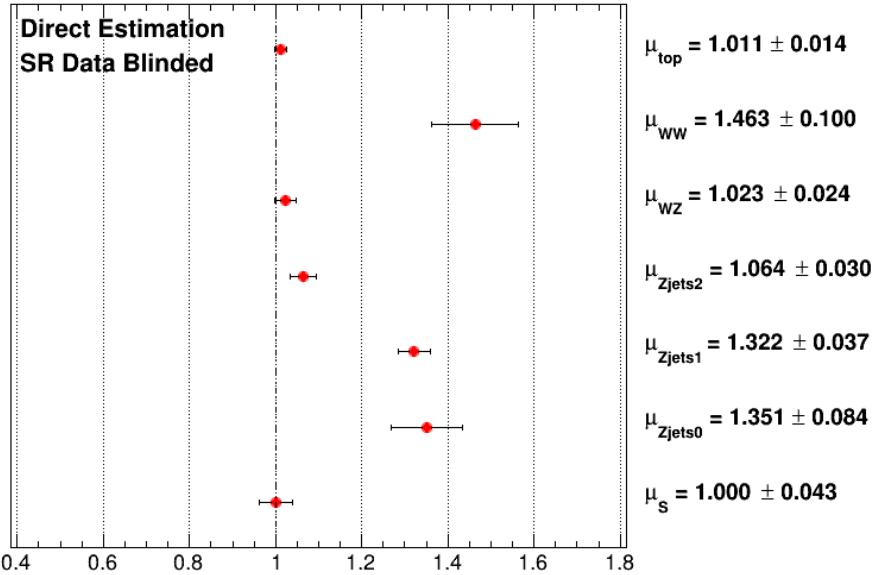
Table 4.5: Scaling factors estimated directly by subtracting the background from each region data sample. The scaling factors are presented according to the successive estimation order shown in Table 4.4. Due to the data being blinded, the MC prediction is considered as the observed signal. Consequently equation (4.5) leads to the trivial value $\mu_S = 1$.

Direct Estimation of μ_S	
Region	Scaling Factor
e μ B	$\mu_{top} = 1.011 \pm 0.014$
e μ A	$\mu_{WW} = 1.463 \pm 0.100$
3 ℓ	$\mu_{WZ} = 1.023 \pm 0.024$
Zjets2	$\mu_{Zjets2} = 1.064 \pm 0.030$
Zjets1	$\mu_{Zjets1} = 1.322 \pm 0.037$
Zjets0	$\mu_{Zjets0} = 1.351 \pm 0.084$
SR	$\mu_S = 1.000 \pm 0.043$

The results obtained by following the analysis methodology described in the current section, are presented in Table 4.5. A visual representation of the results is shown in Figure 4.8. The result on the signal strength is $\mu_S = 1.000 \pm 0.043$. The analysis presented in this section, generally attributes good results. However, it is more than evident that no control region is totally pure in its corresponding

dominant background in order to independently estimate each scaling factor. For this reason a Maximum Likelihood analysis is carried out to perform a simultaneous fit of the scaling factors.

Figure 4.8: Visual representation of the scaling factors by directly extracting the background from each CR data sample. The scaling factors visualized in this figure correspond to the results presented in Table 4.5.



4.3 Maximum Likelihood Analysis

Since every Control Region is not totally pure, a totally independent estimation of each scaling factor is not possible. Hence, a Maximum Likelihood function has been constructed in order to perform a simultaneous fit. Following the Maximum Likelihood method, the scaling factors are varied simultaneously and the set of scaling factors that results in the maximum likelihood function is the *Maximum Likelihood Estimate* (MLE) of the scaling factors. Consequently, one can acquire a more accurate estimation of the scaling factors in comparison with the successive direct estimation procedure described in section 4.2. A more extensive description of the Maximum Likelihood function and its properties can be found in Appendix A.

Since there are 7 scaling factors to fit and 7 available regions in total, no *degrees of freedom* are left. Consequently, the regions (both signal and control regions) are splitted in half in terms of their data event yields. The regions have been splitted with respect to the $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{\text{miss}})$ kinematic variable. This variable has been selected since it does not have major impact on the phase space of the regions (Section 3).

Following the splitting methodology described results in 14 regions, 12 control and 2 signal regions, containing approximately half the data yields with respect to their original one. As a result, there are now 14 available regions, i.e. 14 data points, to fit the 7 scaling factors. The result is a fit with $14 - 7 = 7$ degrees of freedom.

A likelihood function has been constructed for each region with custom private code in C++. For practical purposes, a negative log-likelihood function is constructed. Consequently, its minimization attributes the *best-fit* set of scaling factors instead.

In every region, the likelihood function receives the data event yield N^{obs} and the corresponding model prediction N^{pred} to be fitted to the data points. The observed and predicted events used are the yields obtained according to every region definition (Section 3.4 & 4.1). The corresponding yields are presented in Appendix F. As already mentioned, the data sample in the Signal Region is kept blinded and the total scaled MC prediction is utilized as pseudo-data instead. The model prediction contains the set of scaling factors to be estimated by the fit:

$$\begin{aligned} N_R^{pred} = & N_{R, \text{Signal}}^{MC, \text{est}} \cdot \mu_{\text{Signal}} + N_{R, \text{top}}^{MC, \text{est}} \cdot \mu_{\text{top}} \\ & + N_{R, \text{Zjets2}}^{MC, \text{est}} \cdot \mu_{\text{Zjets2}} + N_{R, \text{WZ}}^{MC, \text{est}} \cdot \mu_{\text{WZ}} \\ & + N_{R, \text{Zjets1}}^{MC, \text{est}} \cdot \mu_{\text{Zjets1}} + N_{R, \text{Zjets0}}^{MC, \text{est}} \cdot \mu_{\text{Zjets0}} \\ & + N_{R, \text{WW}}^{MC, \text{est}} \cdot \mu_{\text{WW}} + N_{R, \text{other}}^{MC, \text{est}} \end{aligned} \quad (4.6)$$

where R denotes the corresponding region.

Hence, the total negative log-likelihood function is given by the following expression:

$$-\ln \mathcal{L}_{total} = - \sum_{R=1}^{14} (N_R^{obs} | N_R^{pred}) \quad (4.7)$$

The number of events is considered to be following a Poisson distribution. Consequently, the total negative log-likelihood is written as follows:

$$\begin{aligned} -\ln \mathcal{L}_{total} = & - \sum_{R=1}^{14} \ln [\text{Poisson}(N_R^{obs} | N_R^{pred})] \\ = & - \sum_{R=1}^{14} \ln \left[\text{Poisson}\left(N_R^{obs} \middle| \sum_i \hat{\mu}_i \cdot N_{R, i}^{MC, \text{est}} + N_{R, \text{other}}^{MC, \text{est}}\right) \right] \end{aligned} \quad (4.8)$$

where the index i refers to each type of event contribution including the signal events.

The Poisson distribution is given by the following expression according to its definition:

$$\begin{aligned} \text{Poisson}(N^{obs} | N^{pred}) &= \frac{e^{-N^{pred}} \cdot (N^{pred})^{N^{obs}}}{N^{obs}!} \Rightarrow \\ -\ln [\text{Poisson}(N^{obs} | N^{pred})] &= -N^{obs} \cdot \ln N^{pred} + N^{pred} + \ln(N^{obs}!) \end{aligned} \quad (4.9)$$

where the term $\ln(N^{obs}!)$ is omitted since $\ln(N^{obs}!) \rightarrow \infty$ for any typical value of N^{obs} .

Hence, by utilising a likelihood function, the data points are compared with the events predicted by the model. The set of estimators $\hat{\mu}_i$ that minimizes the negative log-likelihood $-\ln \mathcal{L}_{total}$, given in equation (4.8) is the *maximum-likelihood estimate* (MLE) of the desired scaling factors μ_i corresponding to the *best-fit* model. As already mentioned, due to the data being blinded in the SR, a signal strength value of $\mu_S = 1$ is expected.

Lastly, the statistical error of each scaling factor has been obtained by scanning the likelihood function with respect to the corresponding factor keeping the others fixed to the best-fit values according to the property expressed in equation (A.9). The likelihood scan for the signal strength μ_s is visualized in Figure 4.9. The $-ln\mathcal{L}$ between the best-fit value and the corresponding 1σ values differs by $1/2$.

Following the methodology presented in this section to perform a simultaneous fit of the scaling factors, the results attributed are presented in Table 4.6.

Figure 4.9: Example of a likelihood scan on μ_s to retrieve its statistical uncertainty. The uncertainty on μ_s can be read off from the two points where $ln\mathcal{L}$ has a higher value by $1/2$ from its minimum value at the MLE value of μ_s .

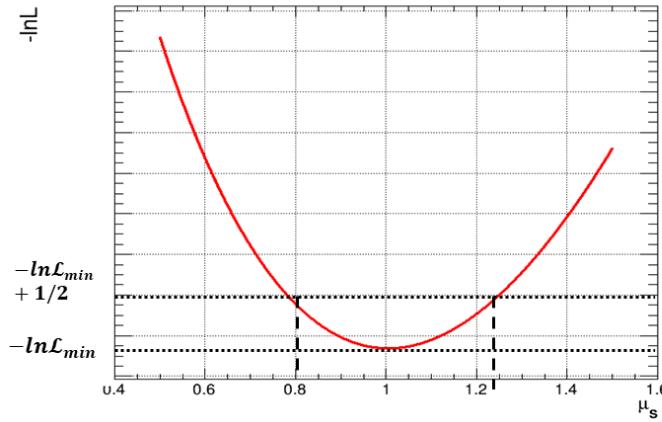


Figure 4.10: Visual representation of the scaling factors obtained by performing a simultaneous MLE fit with all regions including the SR. The scaling factors visualized in this figure correspond to the results presented in Table 4.6. The uncertainties are statistical obtained from scanning the likelihood function with respect to each particular scaling factor as shown in the example for μ_s in Figure 4.9.

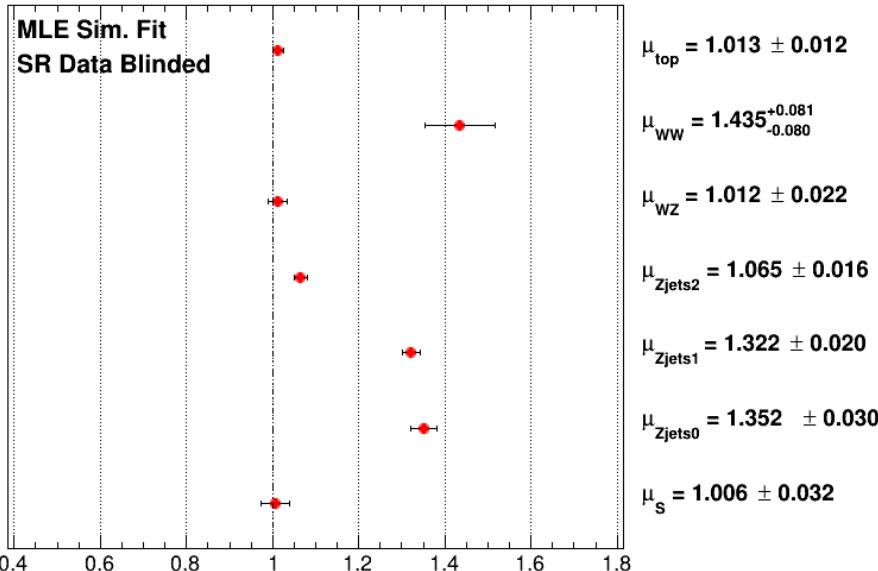


Table 4.6: Scaling factors estimated by performing a simultaneous fit analysis in all the regions including the SR. The total negative log-likelihood function has been estimated using equation (4.8). The uncertainties are statistical obtained from the likelihood scan.

MLE Simultaneous Fit	
Region	Scaling factor
eμB	$\mu_{top} = 1.013 \pm 0.012$
eμA	$\mu_{WW} = 1.435^{+0.081}_{-0.080}$
3ℓ	$\mu_{WZ} = 1.012 \pm 0.022$
Zjets2	$\mu_{Zjets2} = 1.065 \pm 0.016$
Zjets1	$\mu_{Zjets1} = 1.322 \pm 0.020$
Zjets0	$\mu_{Zjets0} = 1.352 \pm 0.030$
SR	$\mu_S = 1.006 \pm 0.032$

4.4 Remarks on the Signal Strength Estimation

Comparing the results obtained by following the direct estimation and MLE methodology presented in tables 4.5 and 4.6 correspondingly, it is evident that most values are almost identical. One can notice a substantial difference only for the μ_{WW} , μ_{WZ} and μ_S scaling factors. The impact on these scaling factors by following the MLE methodology (Section 4.3), as opposed to directly estimating them (Section 4.2), is expected. The maximum likelihood function has been constructed in order to perform a fit of the scaling factors since no independent estimation can be achieved by directly estimating each one of them separately.

Concerning the signal strength scaling factor which is the parameter of interest, the trivial value $\mu_S = 1.000 \pm 0.043$ has been obtained following the direct estimation procedure, while the value $\mu_S = 1.006 \pm 0.032$ has been obtained following the MLE fit methodology. As already mentioned, this value was expected since the data sample in the SR is kept blinded and the total scaled MC prediction is used as pseudo-data in the SR instead.

The μ_{WW} factor is related to the e μ A Control Region which is by no means dominated by WW events in order to independently estimate the μ_{WW} scaling factor. Instead, it is dominated by top events by a factor of ~ 3 compared to the WW events (Fig.4.2).

Concerning the μ_{WZ} scaling factor, one can argue that the impact of the MLE fit is related to the interplay between the 3 ℓ CR (Fig.4.1) constructed to study the WZ background, and the Signal Region which contains a large number of WZ events $\sim 26\%$ (Fig. 3.7). Moreover, the 3 ℓ CR is the region containing the largest percentage of other background ($\sim 5\%$) which is impossible to define a scaling factor for in order to correct the corresponding MC prediction.

Moreover, one can notice that the statistical uncertainties obtained by the likelihood scan following the MLE methodology, are significantly smaller. This behaviour is to be expected since a more sophisticated approach is followed compared to simple error propagation in the direct estimation procedure. By splitting the data samples in half, the information they contain is better utilized.

Finally, concerning the MLE fit methodology, it is important to note that using a custom brute force algorithm scanning over scaling factor values is only a naive approach. It is practically impossible to perform a more sophisticated analysis

including more parameters (e.g. systematic uncertainties) following this methodology. In practice statistical frameworks like TRExFitter are used.

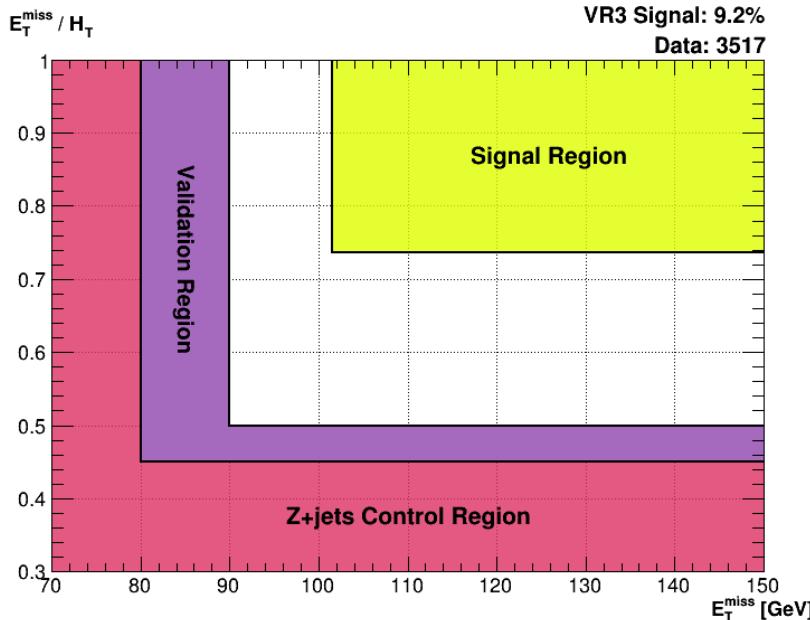
4.5 Validation Region

During the background study, each major background contribution has been studied by taking under consideration the discrepancy between data and MC prediction for each one of them. This discrepancy has been quantified by defining the corresponding scaling factors (Sec. 4.2).

Ultimately, the purpose of the background study is to provide a more accurate estimation of the desired signal process $ZZ \rightarrow \ell\ell\nu\nu$ by correcting the various background types in the Signal Region according to the corresponding scaling factor for each.

In order to evaluate the performance of the scaling factors in terms of correcting the model prediction, a new region has been defined. This region is called *Validation Region* (VR) and it is required to be totally independent from any other region defined in order to provide an independent evaluation of the model correction performance. As every other region defined, it must be *SR-like* since it ultimately serves as a cross-validation for the scaling performance in the SR. However, care must be taken in defining the VR in order to achieve minimal signal presence since the data sample in the VR is accessed.

Figure 4.11: Representation of the Validation Region in the 2-d (E_T^{miss} , E_T^{miss}/H_T) plane. The cuts imposed in the VR ensure that it is totally independent from the Zjets CR and the SR. The VR corresponds to the criteria shown in Table 4.7. The MC signal abundance is 9.2% of the total MC events in the VR. The MC events yields are scaled according to the best estimated scaling factors obtained using the MLE method (Table 4.6).



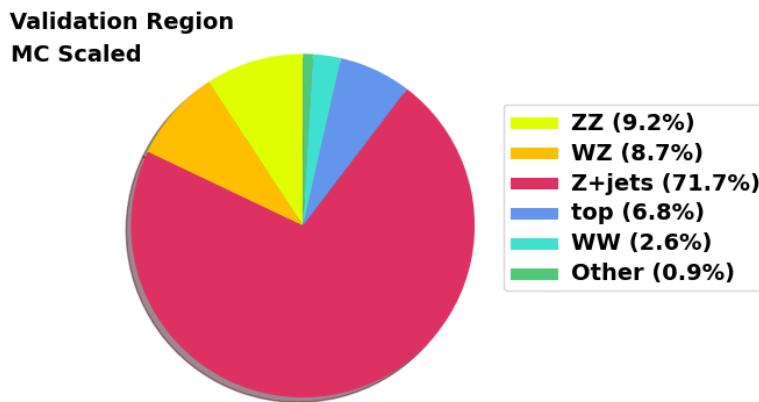
In order to define a suitable phase space the E_T^{miss} and E_T^{miss}/H_T kinematic variables are selected. After a thorough study, a characteristic "L-shaped" phase space similar to the one defined in the Zjets CR is selected (Fig.4.11). The criteria

on other variables are the same criteria used to define the Zjets region. The full selection criteria for the VR, are presented in Table 4.7 attributing a total of 3517 data events and a signal-to-total MC events ratio of 9.2%. Obviously, this percentage has been achieved in exchange for a considerable amount of statistics removed. A more extensive study of the Validation Region is given in Appendix E. The event yields are scaled according to the best-fit scaling factors obtained from the maximum likelihood analysis (Table 4.6).

Table 4.7: Definition criteria for the Validation Region. The criteria are motivated by the requirement for an independent region similar to the SR with minimal signal presence at the same time.

Validation Region
$80 < M_{ll} < 100 \text{ GeV}$
$80 < E_T^{\text{miss}} < 90 \text{ GeV}$ and $0.45 < E_T^{\text{miss}}/H_T$
or
$90 < E_T^{\text{miss}} \text{ GeV}$ and $0.45 < E_T^{\text{miss}}/H_T < 0.5$
$\Delta R_{\ell\ell} < 2$
$\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{\text{miss}}) > 2.2$
b-jet veto

Figure 4.12: Pie chart of the Validation Region event yields based on the definition criteria shown in Table 4.7 but scaled with the best-fit scaling factors shown in Table 4.6.

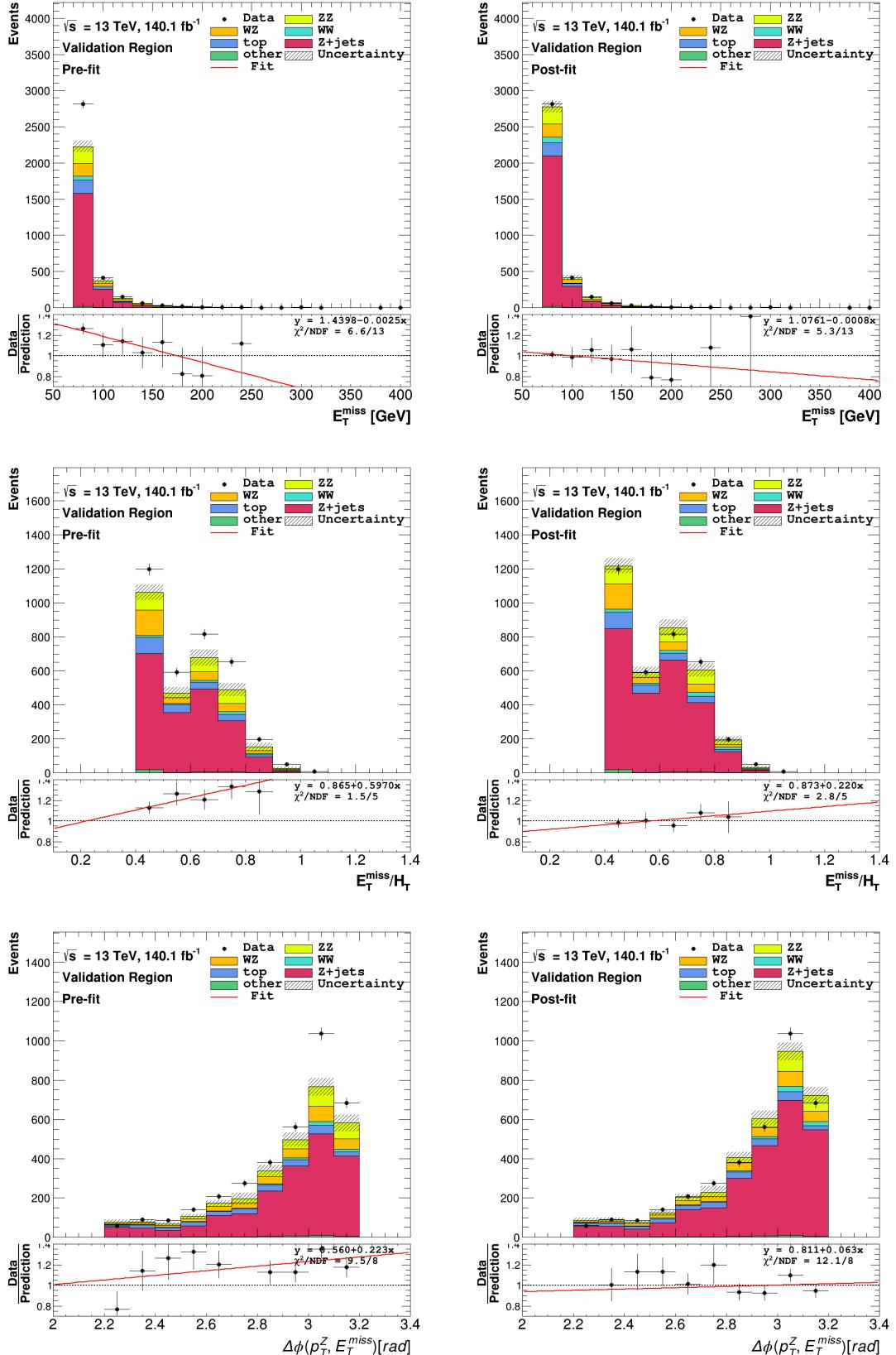


The ratio of the total data-to-MC event yields has shifted from 1.223 before scaling to 1.008 after scaling meaning that the scaling factors perform properly in terms of minimising the discrepancy between data and MC prediction.

Moreover, the scaling performance in the different areas of the phase space must be evaluated. Specifically, the VR is plotted versus different kinematic variables of interest in order to evaluate the scaling performance in the phase space areas of interest before and after scaling.

The results for the kinematic variables selected for Signal optimization (Sec.3.1), are presented in Figure 4.13. Additional plots on VR as well as plots for the scaling performance on the Control Regions, are presented in Appendix G.

Figure 4.13: Pre-fit and Post-fit plots of Validation Region versus E_T^{miss} , E_T^{miss}/H_T and $\Delta\phi(p_T^Z, E_T^{miss})$. The pre-fit plots correspond to the event yields shown in Table F.4. In the post-fit plots, every bin has been scaled using the best-fit values obtained from the maximum likelihood analysis results, shown in Table 4.6. The scaled bins are in good agreement with the data. The uncertainties are statistical. A fit function is used in order to evaluate the scaling performance (red line). Additional plots on VR are presented in the Appendix section G.3.



A fit function fitted to the Data/MC bins, is used as criterion to evaluate the scaling performance. A slope close to zero indicates consistent scaling performance throughout the entire variable. From studying these plots, it is evident that the slope gets closer to zero after scaling. This behaviour indicates consistent scaling performance independently of the kinematic variables.

Finally, it is important to note that the MC prediction for the signal process in the CR is non-negligible and has a significant contribution ($\sim 9.2\%$). As already mentioned, the SR data sample is kept blinded meaning that the SR relies entirely on the total events MC prediction used as pseudo-data. Consequently, only the CR data samples have contributed to the signal strength μ_S estimation. Utilizing the information that the real SR data sample has to offer as well, may result in a better result in terms of scaling performance.

5 Study of Signal Strength Uncertainties

The accurate estimation of the signal strength μ_S is of great importance since it ultimately leads to the cross section measurement of the desired signal process $ZZ \rightarrow \ell\ell\nu\nu$. Of course, the signal strength is subject to various sources of uncertainties. Firstly, the uncertainty due to data statistics has been studied in order to be compared with the statistical uncertainty obtained from the point estimation of μ_S (Tab. 4.6). Afterwards, two additional sources of systematic uncertainty have been studied. The first systematic uncertainty is the result of finite MC statistics while the second refers to the systematic uncertainty originating from the Control Regions phase space definitions. The results are presented in this section. As in the rest of this work, the data sample in the Signal Region is kept blinded.

The uncertainty due to data statistics and due to MC statistics is further studied by performing ensemble tests. An ensemble test is performed by repeating *pseudo-experiments*. In each pseudo-experiment various outcomes called *pseudo-data* are generated by varying the real data. For each outcome (pseudo-experiment), the log-likelihood function given by expression (4.8), is scanned with respect to the signal strength μ_S keeping the other scaling factors fixed to their best-fit values. By scanning the log-likelihood function, the best-fit value for μ_S is obtained for each particular outcome. By repeating an adequate number of pseudo-experiments, one can gain confidence on the uncertainty received by the real outcome. Firstly, the data yields in each region are fluctuated according to a Gaussian distribution while the Monte Carlo samples used for the model prediction are kept fixed to their nominal values. Following this methodology, the uncertainty due to data statistics is further studied. Secondly, the MC sample yields in each region are fluctuated according to a Gaussian distribution while the data samples are kept fixed to their nominal (real) values in order to evaluate the systematic uncertainty due to finite MC statistics.

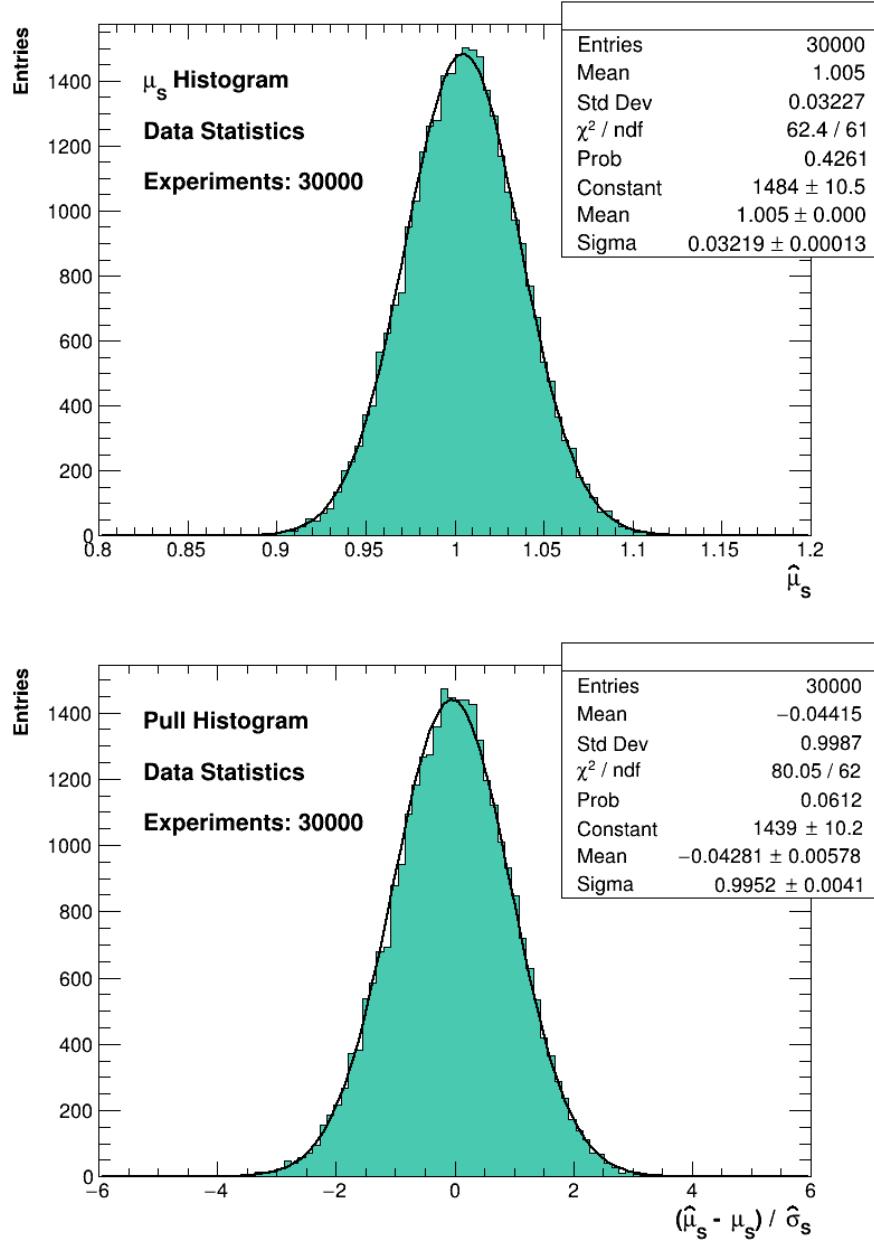
Finally, since there is no way to confirm if the Control Regions defined are the optimal, the impact of the Control Regions is studied in terms of their phase space definitions. Two slightly varied regions are created for the 3ℓ , NR and Zjets Control Regions. All different combinations of nominal and varied CR are tried out attributing different signal strength μ_S values.

As already mentioned, due to data being blinded in the SR, the total scaled MC event yield is used as the "real" nominal data during the study of the uncertainties described in the present section.

5.1 Uncertainty due to Data Statistics

Following, the Maximum Likelihood Analysis described in section 4.3, an estimation of the signal strength μ_S has been achieved resulting in the value $\mu_S = 1.006 \pm 0.032$ (Tab.4.6). In order to confirm this result, an ensemble test is performed by repeating pseudo-experiments where the data yields of each region are varied randomly resulting in a new data outcome. Each outcome is generated according to a Gaussian distribution with the real data central value and uncertainty as mean and standard deviation parameters correspondingly. The μ_S distribution obtained by the varied data yields is compared with the point estimation of the μ_S .

Figure 5.1: Plot of the $\hat{\mu}_S$ and pull term given in equation (5.1) for 30000 pseudo-experiments. The histograms are fitted with a Gaussian function. In each pseudo-experiment the data yields are varied according to a Gaussian distribution. A value of $\hat{\mu}_S = 1.005 \pm 0.032$ is obtained in good agreement with the point estimation result $\mu_S = 1.006 \pm 0.032$ from the maximum likelihood analysis. A pull term distributed according to a standard Gaussian further confirms the validity of the result.



As already mentioned, the maximum likelihood methodology presented in section 4.3, is repeated for each new outcome generated in every pseudo-experiment. A log-likelihood function is constructed from the individual region likelihood functions according to equation (4.8). However, only the parameter of interest μ_S is varied while keeping the rest scaling factors fixed to their best-fit values shown in Table 4.6. This procedure is repeated in every pseudo-experiment and finally a distribution for the signal strength μ_S is obtained. For a substantial amount of pseudo-experiments, the distribution is expected to be Gaussian-shape. A fit is performed with a Gaussian function and its mean and standard deviation param-

eters are compared with the real data results. A standard deviation similar to the error due to data statistics obtained from the μ_S point estimation $\mu_S = 1.006 \pm 0.032$, is expected.

To evaluate the ensemble tests performance, the pull term for each pseudo-experiment is calculated as well:

$$\text{pull} = \frac{\hat{\mu}_S - \mu_S}{\sigma_{\hat{\mu}_S}} \quad (5.1)$$

If the resulting $\hat{\mu}_S$ generated repeatedly from the ensemble test, is distributed according to a Gaussian with mean μ_S and width $\sigma_{\hat{\mu}_S}$, then the pull term given in (5.1) will be distributed as a standard Gaussian with mean equal to zero and width equal to one. Consequently, a pull term distributed according to a standard Gaussian, further confirms the point estimation of μ_S .

In total, 30000 pseudo-experiment have been performed. The corresponding outcomes distribution is presented in Figure 5.1. The Gaussian fit parameters attribute the value $\hat{\mu}_S = 1.005 \pm 0.032$. The result is in really good agreement with the point estimation result $\mu_S = 1.006 \pm 0.032$ obtained from the maximum likelihood analysis (Section 4.3). Moreover, the pull term given by equation (5.1) is distributed according to a Gaussian-shaped distribution with mean equal to -0.043 and a standard deviation equal to 0.995 which can be characterised as approximately standard Gaussian. Hence, the pull distribution further confirms the result of the μ_S estimation which corresponds to a 3.18% uncertainty due to the data statistics.

5.2 Uncertainty due to MC Statistics

During the analysis presented in the present work, various MC samples have been utilised together with the data samples. The MC samples are crucial to better understand the various background contributions as well as the signal in order to achieve an accurate signal estimation which ultimately leads to a cross section measurement. As a result, a further study of the MC samples to access their behaviour is really important. Consequently, an ensemble-test by generating pseudo-samples based on the nominal MC samples, is performed in this case as well.

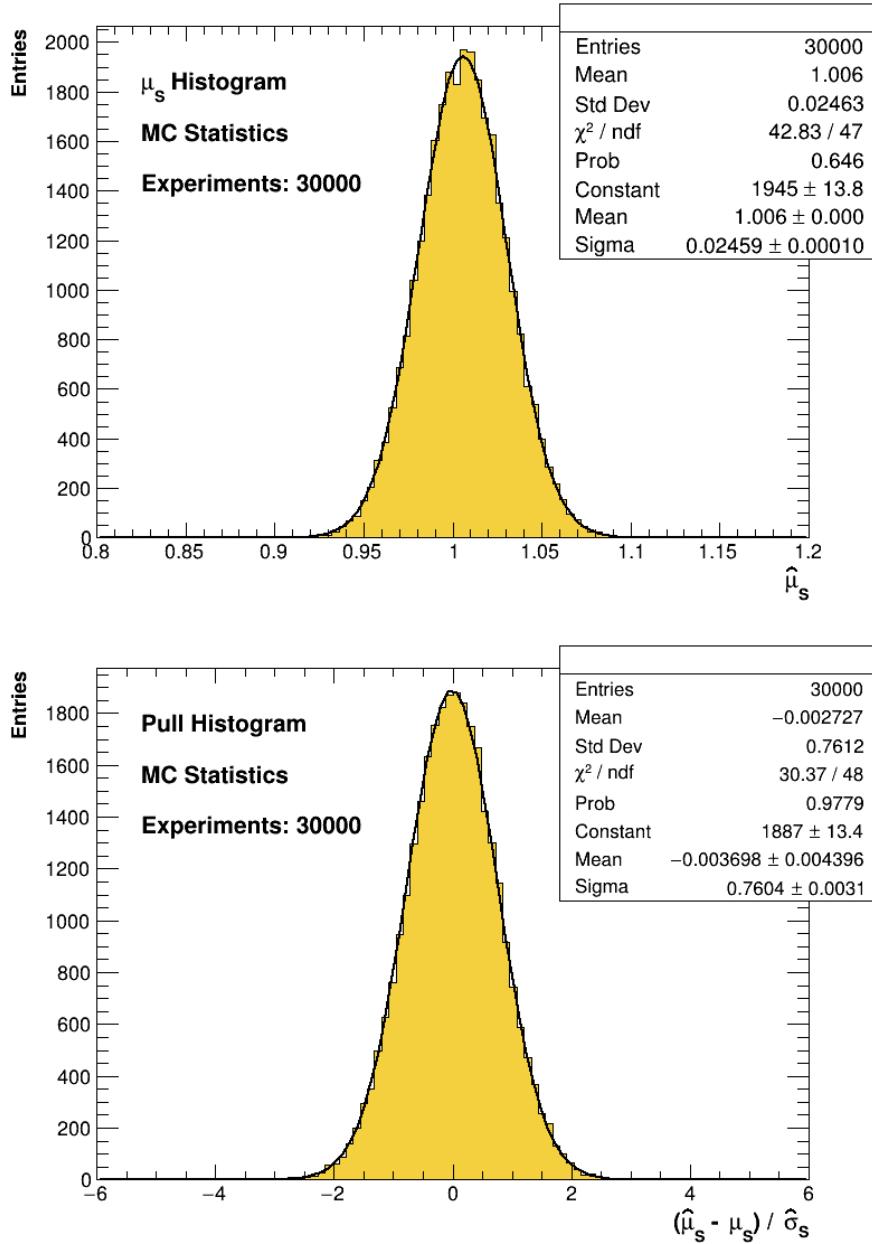
Apart from generating MC pseudo-events instead of data sample pseudo-events, the procedure is identical to the one followed for the data statistics uncertainty. Here, for each pseudo-experiment, the various types of MC sample event yields are varied according to a Gaussian distribution with the corresponding yield mean and standard deviation as its parameters while the nominal data values are used. For each pseudo-experiment the corresponding total log-likelihood is constructed from the individual region likelihood functions according to equation (4.8). The parameter of interest μ_S is scanned while the rest scaling are kept fixed to their best-fit values. The corresponding μ_S and pull term distributions obtained by repeating pseudo-experiments are obtained for this case as well. A fit with a Gaussian function is performed. The fit parameters on the μ_S histogram ultimately attribute the central value and uncertainty of the signal strength.

In total, 30000 pseudo-experiment have been performed. The corresponding outcomes distribution is presented in Figure 5.2. A value $\hat{\mu}_S = 1.006 \pm 0.025$ has resulted from the Gaussian fit parameters. The pull histogram attributes a Gaussian-shaped distribution with a mean value equal to -0.004 with a standard deviation

of 0.760. This distribution deviates from a Gaussian one. That is, no further confidence on the result can be gained based on the pull histogram.

Consequently, an additional uncertainty of 2.50% on the signal strength μ_s must be taken under consideration. This uncertainty is a source of systematic uncertainty due to the finite MC sample statistics.

Figure 5.2: Plot of the $\hat{\mu}_s$ and pull term given in equation (5.1) for 30000 pseudo-experiments. The histograms are fitted with a Gaussian function. In each pseudo-experiment the data yields are varied according to a Gaussian distribution. A value of $\hat{\mu}_s = 1.006 \pm 0.025$. The pull term distribution deviates from a standard Gaussian one.



5.3 Uncertainty due to Control Regions

During the background study, a Control Region has been defined for each major background type. These *nominal* Control Regions, considered the optimal, have

been utilised in order to ultimately estimate the signal strength μ_S . The nominal CR definitions are given in section 4.1.

In order to evaluate the impact of the CR definitions on the μ_S estimation, their phase spaces are slightly varied. Repeating the scaling factors estimation with a combinations of CR variants results in a new μ_S . Its deviation from the nominal one is considered an additional source of uncertainty.

Specifically, the kinematic variables E_T^{miss} and E_T^{miss}/H_T which have major impact on each CR phase space, are varied. The variations are imposed on the 3ℓ , NR and Zjets Control Regions. The NR and Zjets CR are comprised of the $e\mu B$, $e\mu A$ and Zjets0, Zjets1, Zjets2 subregions correspondingly. However, they are treated as whole since the same cuts are imposed on the subregions. The cuts imposed on E_T^{miss} and E_T^{miss}/H_T variables to create variant CR are shown in Table 5.1. The event yields attributed from the new variants are presented in Tables F.5, F.6 and F.7.

Table 5.1: Variants to the nominal CR. The cut variants replace the nominal cuts on the corresponding kinematic variables and are imposed together with the rest cuts imposed for the nominal definitions presented in section 4.1. The corresponding nominal cuts are given as well.

3 ℓ and NR Variants		
Nominal	Var I	Var II
$70 < E_T^{miss}$	$80 < E_T^{miss}$	$85 < E_T^{miss}$
$0.3 < E_T^{miss}/H_T$	$0.5 < E_T^{miss}/H_T$	$0.55 < E_T^{miss}/H_T$

Zjets Variants		
Nominal	Var I	Var II
$(70 < E_T^{miss} < 80 \text{ and } 0.3 < E_T^{miss}/H_T)$ or $(80 < E_T^{miss} \text{ and } 0.3 < E_T^{miss}/H_T < 0.45)$	$(70 < E_T^{miss} < 80 \text{ and } 0.5 < E_T^{miss}/H_T)$ or $(80 < E_T^{miss} \text{ and } 0.3 < E_T^{miss}/H_T < 0.45)$	$(70 < E_T^{miss} < 85 \text{ and } 0.3 < E_T^{miss}/H_T)$ or $(85 < E_T^{miss} \text{ and } 0.3 < E_T^{miss}/H_T < 0.50)$

In total, two different variations are imposed in each of the three regions 3ℓ , NR and Zjets resulting in two more regions apart from the nominal ones. The result is 9 regions in total (the 3 nominal and 6 new variant ones). Every possible combination of the three Control Regions is tried out resulting in $3^3 = 27$ combinations. For each set of CR definitions, a total negative log-likelihood function is constructed by adding each corresponding CR likelihood and finally the SR likelihood function as well, using the expression (4.8). The likelihood function is then scanned around μ_S keeping the other factors fixed to their best-fit values shown in Table 4.6. For each set of varied regions, a new μ_S value is obtained summing up to a total of 27 values.

Since the criteria selected to create CR variants are arbitrary, the maximum deviation from the nominal value $\mu_S = 1.006$ is considered as an additional source of systematic uncertainty due to the CR definitions. Following the presented methodology, the maximum deviation from the μ_S nominal value is 0.00487 resulting in a additional 0.49% systematic uncertainty on μ_S due to the CR definitions.

In conclusion, the uncertainties on μ_S that have been studied in the present work are summarised in Table 5.2. The combined quadrature sum uncertainty is given as well.

Concerning the uncertainties due to data and MC statistics, it is important to mention that in principle, an identical analysis methodology like the one followed

for the real data, should be followed for each pseudo-experiment. That is, performing a simultaneous fit by varying all scaling factors simultaneously. However, it is practically impossible to repeat this procedure for an adequate number of pseudo-experiments in order to gain sufficient statistics.

Table 5.2: Summary of the three uncertainties taken under consideration for the μ_S . The first one corresponds to the statistical uncertainty due to the data Poisson statistics. The second corresponds to the systematic uncertainty due to the finite MC samples statistics. Finally, the third corresponds to the systematic uncertainty produced due to the impact that the control region definitions have on the μ_S .

Summary of μ_S Uncertainties	
Source	Value (%)
Data Statistics	3.18
MC Statistics	2.50
CR Definitions	0.49
Combined	4.07

6 Fiducial Cross Section Measurement

The aim of the present work study is to measure the cross section of the inclusive ZZ process identified by the two channels $ZZ \rightarrow \ell^+ \ell^- \ell'^+ \ell'^-$ and $ZZ \rightarrow \ell^+ \ell^- \nu \bar{\nu}$. In principle, calculating the cross section of the desired process is pretty straightforward. However, the detector and selection efficiencies are required to be taken under consideration in order to measure the cross section at "truth-level". That is, to measure the cross section that corresponds to a phase space totally independent from any detector and selection effects. This phase space is called *fiducial* and in it expresses the prediction for the total signal produced by nature independently of the entire experimental procedure (Section 2.4). The entirety of these effects for the inclusive ZZ process, is represented by the correction factor \mathcal{C}_{ZZ} .

The cross section estimation is obviously directly related to the observed reconstructed signal yield which is by definition related to the MC signal prediction through the signal strength: $N_{ZZ, meas}^{fid} = \mu_S \cdot N_{ZZ, MC}^{fid}$, as shown in equation (4.5). As already mentioned, since the SR data sample is blinded, the total scaled MC event yield in the SR, is used as pseudodata instead. The scaling factors used, have been obtained by the best-fit values presented in Table 4.6.

The integrated fiducial cross section of the inclusive $ZZ \rightarrow \ell \ell \nu \bar{\nu}$ process is then given by the following expression:

$$\begin{aligned}\sigma_{ZZ}^{fid, meas} &= \frac{N_{Signal}^{obs}}{\mathcal{L} \times \mathcal{C}_{ZZ}} \\ &= \frac{\mu_S \cdot N_{Signal}^{MC}}{\mathcal{L} \cdot N_{Signal}^{MC} / N_{Signal}^{fid}} \\ &= \frac{\mu_S \cdot N_{Signal}^{fid}}{\mathcal{L}} \\ &= \mu_S \cdot \sigma_{ZZ}^{fid, MC}\end{aligned}\tag{6.1}$$

where \mathcal{L} is the integrated luminosity corresponding to the data sample used to obtain a measurement. In the present work the full Run II data sample collected by the ATLAS experiment during the period 2015 - 2018 are utilised, corresponding to an integrated luminosity of $140.1 fb^{-1}$.

As already mentioned, the detector and selection efficiencies are required to be taken under consideration in order to estimate the fiducial cross section. The entirety of these effects for the inclusive ZZ process, is represented by the correction factor \mathcal{C}_{ZZ} . The \mathcal{C}_{ZZ} factor is estimated as the ratio of the reconstructed MC signal yield over the MC predicted signal yield for truth-level kinematics:

$$\mathcal{C}_{ZZ} = \frac{N_{Signal}^{MC, SR}}{N_{Signal}^{fid}}\tag{6.2}$$

The truth events have been estimated by the simulation samples described in section 2.2. Taking under consideration both the QCD qq, gg and EWK processes results in a value of $N_{Signal}^{fid} = 2277.28 \pm 17.72$ as presented in Table 6.1. ³

³Big thanks to Angelos Tsiamis, PhD candidate at AUTH, for offering the optimal simulation samples ready for analysis and providing valuable guidance to obtain the correct results ☺

Table 6.1: Inclusive ZZ process signal yields in the Fiducial Volume (FV) for the two QCD processes as well as the EWK process. The signal yields from the three different processes sum up to the total reconstructed signal yields in the fiducial phase space N_{Signal}^{fid} . The uncertainties are purely statistical.

FV Inclusive ZZ Signal Yields	
$qq \rightarrow ZZ \rightarrow \ell\ell\nu\nu$	2050.20 ± 17.72
$gg \rightarrow ZZ \rightarrow \ell\ell\nu\nu$	222.74 ± 0.33
$qq \rightarrow \ell\ell\nu\nu jj$	4.34 ± 0.04
Total	2277.28 ± 17.72

The signal strength value used in the cross section estimation as well as the rest scaling factor values obtained by performing a maximum likelihood simultaneous fit, is presented in Table 4.6. The signal strength value as well as its total combined uncertainty of $\sim 4.07\%$ summarized in Table 5.2, is $\mu_S = 1.006 \pm 0.041$.

Regarding the integrated luminosity, for the full Run2 data sample, a value of $\mathcal{L} = 140.1 \pm 1.2 \text{ fb}^{-1}$ corresponding to an uncertainty of $\sigma_{\mathcal{L}}/\mathcal{L} = 0.83\%$, has been used [11].

Moreover, the signal region has attributed a reconstructed MC signal yield $N_{Signal}^{MC} = 1577.03 \pm 13.84$ for the inclusive ZZ process (Table F.4). This result along with the fiducial phase space signal yield result in a correction factor value of $C_{ZZ} = 0.69 \pm 0.01$. The entirety of quantities associated with the cross section estimation are summarised in Table 6.2.

Table 6.2: Quantities associated with the cross section estimation and their corresponding uncertainties.

Cross section measurement	
μ_S	1.006 ± 0.041
N_{Signal}^{fid}	2277.3 ± 17.7
N_{Signal}^{MC}	1577.0 ± 13.8
C_{ZZ}	0.69 ± 0.01
$\mathcal{L} [\text{fb}^{-1}]$	140.1 ± 1.2

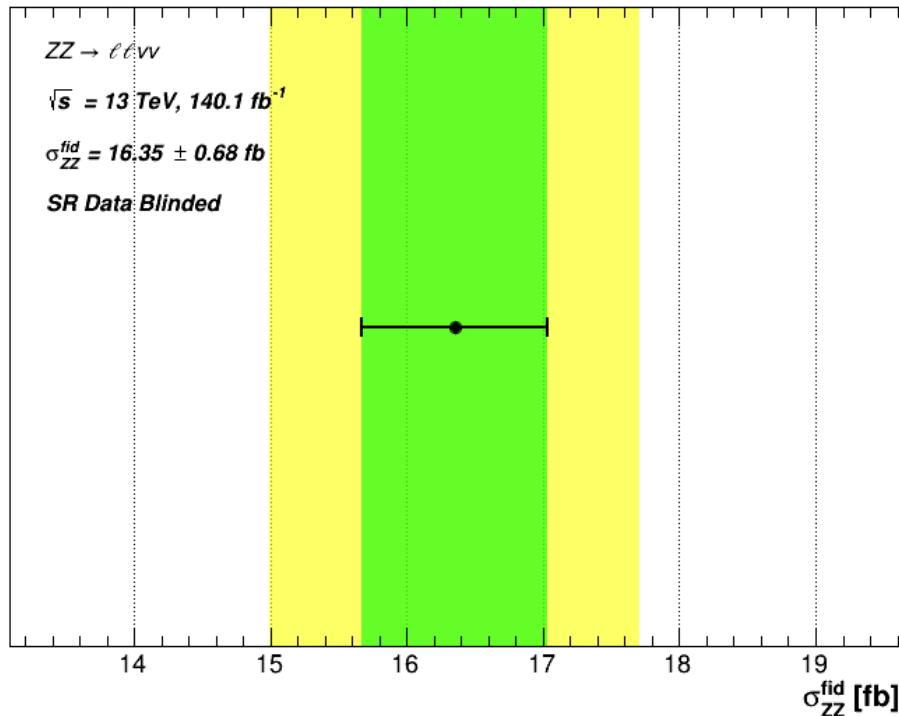
In order to estimate the cross section for the inclusive ZZ process, one only needs the μ_S , N_{Signal}^{fid} and \mathcal{L} quantities. Inserting them in expression (6.1) a measurement of the desired cross section is achieved:

$$\begin{aligned} \sigma_{ZZ}^{fid} &= 16.35 \pm 0.52 \text{ (stat.)} \pm 0.42 \text{ (syst.)} \pm 0.14 \text{ (lumi.)} \\ &= 16.35 \pm 0.68 \text{ (tot.) fb} \end{aligned} \quad (6.3)$$

where only the μ_S and \mathcal{L} uncertainties have been taken under consideration. The uncertainty on N_{Signal}^{fid} due to MC statistics is not taken under consideration since it is already accounted for in the μ_S uncertainty. The resulting relative uncertainty on the cross section measurement is $\sim 4.16\%$. A visual representation of the result is presented in Figure 6.1.

Since the measurement relies entirely on the MC prediction pseudo-data instead of the real data sample in the SR, the cross section result corresponds to the theoretically predicted value. Utilizing the information offered by the SR data sample, one can obtain a value based on the measurement on the SR.

Figure 6.1: Measurement of the inclusive ZZ process fiducial cross section. The 1σ and 2σ regions are plotted as well. The cross section measurement corresponds to the results of (6.3).



7 Conclusion and Remarks

In the present work, a study of the signal and background for the inclusive ZZ process has been conducted. In total 6 different Control Regions (CR) enriched in the corresponding major backgrounds have been studied. From each CR a scaling factor has been estimated accounting for the discrepancy between data and Monte Carlo events for each one of the major background types.

The Signal Region (SR) has been further studied in terms of signal-to-background discrimination using Brute Force methodology as well as methodology based on a type of heuristic algorithms called Particle Swarm Optimization (PSO) algorithms. This study has been carried out for three kinematic variables of interest and resulted in the following optimal set of cuts:

$$(\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss}), E_T^{miss}, E_T^{miss}/H_T) = (2.690, 101.5, 0.738)$$

where the values that are greater than the presented threshold cut values, are selected in order to optimize the Signal Region phase space. The optimal Signal Region used throughout the present work has been defined by imposing these cuts on top of the preliminary cuts shown in Table 2.1. Although, these cut values have been acquired much easier by utilising a custom brute force algorithm, it would have been practically impossible to do so for more variables.

The custom algorithm created based on the PSO algorithm, has been utilized for 3 variables as well to cross validate with the brute force algorithm. However, this algorithm can be further utilized for many more variables to obtain potentially obtain better results concerning the Signal Region study.

Concerning the background study, performing a simultaneous fit by utilizing a maximum likelihood function (Sec.4.3), resulted in significantly lower statistical uncertainty on the desired signal strength parameter in comparison with the direct estimation procedure (Sec.4.2).

The strategy of blinding the data sample in the Signal Region has been followed. The total scaling MC event yields has been selected as pseudo-data in the SR instead. As a result, the value obtained for the signal strength is $\mu_S = 1.006 \pm 0.041$ corresponding to a total relative uncertainty $\sim 4.07\%$ shown in Table 5.2. The rest of the scaling factors as well as μ_S obtained as the MLE best-fit values, are presented in Table 4.6. However, additional sources of uncertainty like the theoretical PDF systematic uncertainty, have been neglected. Considering additional sources of uncertainty would obviously result to measurement suffering by a larger uncertainty value.

During the background study, the performance of the scaling factors has been accessed in a Validation Region which is entirely independent from any region but similar to the SR. Confidence has been gained in their effectiveness by adjusting the MC predictions using the best-fit scaling factor values. These favorable results are detailed in Section 4.5.

After evaluating the background study results, a measurement of the cross section of the inclusive ZZ process has been achieved attributing the following result:

$$\begin{aligned} \sigma_{ZZ}^{fid} &= 16.35 \pm 0.52 \text{ (stat.)} \pm 0.42 \text{ (syst.)} \pm 0.14 \text{ (lumi.)} \\ &= 16.35 \pm 0.68 \text{ (tot.)} fb \end{aligned}$$

suffering by a relative uncertainty $\sim 4.16\%$. The uncertainties taken under consideration are the aforementioned total combined uncertainty of $\sim 4.07\%$ on μ_S

as well as the uncertainty on integrated luminosity of $\sim 0.83\%$ reported on [11]. The cross section measurement as well as the 1σ and 2σ uncertainty region are presented in Figure 6.1.

The cross section result relies entirely on the MC prediction in the SR due to the corresponding data sample being blinded. Hence, this measurement corresponds to the MC predicted fiducial cross section for the $ZZ \rightarrow \ell^+ \ell^- \nu \bar{\nu}$ process. This is evident from expression (6.1) since the measurement has been made by correcting the theoretically predicted cross section with a scaling factor μ_S equal to one due to keeping the SR data sample blinded. However, following this exact methodology presented in this work, one can utilize the SR data sample instead of the MC prediction, to measure the fiducial cross section of the $ZZ \rightarrow \ell^+ \ell^- \nu \bar{\nu}$ process based on the information that the SR data sample has to offer.

Appendices

A Parameter estimation using the Maximum Likelihood Method

Parameter estimation, a process also called *fitting*, is a fundamental process in experimental physics. Its goal is to obtain the best possible representation of the observed data by adjusting the parameters of a model used to describe the data. Put another way, its goal is to minimize the discrepancy between the observed data and the model by adjusting the model parameters accordingly. The two most common *frequentist* methods for parameter estimation are the methods of *maximum likelihood* and *least squares*. In the present work, the maximum likelihood method is utilized in order to obtain the *best-fit* values for a set parameters called *scaling factors*, used during the study of background contamination as an integral part of the study of the desired process $ZZ \rightarrow \ell\ell\nu\nu$.

A.1 The Maximum Likelihood Function

Assuming there is a dataset of N measured quantities $\mathbf{x} = (x_1, x_2, \dots, x_N)$, where the measured quantities x_i are statistically independent and each follow the probability density function $f(x; \boldsymbol{\vartheta})$ where $\boldsymbol{\vartheta} = (\theta_1, \theta_2, \dots, \theta_m)$ is a set of m parameters with unknown values to be estimated. The joint probability density function for the observed values \mathbf{x} is given by the likelihood function

$$\mathcal{L}(\mathbf{x}; \boldsymbol{\vartheta}) = \prod_{i=1}^N f(x_i; \boldsymbol{\vartheta}) \quad (\text{A.1})$$

The set of parameter values $\hat{\boldsymbol{\vartheta}}$, for which the likelihood function $\mathcal{L}(\mathbf{x}; \boldsymbol{\vartheta})$ obtains its global maximum, make up the *maximum-likelihood estimate* (MLE) of the parameters $\boldsymbol{\vartheta}$. Consequently, for a set of estimators $\hat{\boldsymbol{\vartheta}}$ that better describes the observed data, higher values of the likelihood function $\mathcal{L}(\mathbf{x}; \boldsymbol{\vartheta})$, are expected.

Hence, the estimated values $\hat{\boldsymbol{\vartheta}}$ of the parameters are obtained by finding the global maximum of the likelihood function. In practice, it is often more convenient to work with logarithm likelihood function, called the *log-likelihood* and to search for the minimum negative log-likelihood function defined as follows:

$$-\ln \mathcal{L}(\mathbf{x}; \boldsymbol{\vartheta}) = \sum_{i=1}^N \ln f(x_i; \boldsymbol{\vartheta}) \quad (\text{A.2})$$

Unless the minimum occurs at the boundary of the allowed range of values for $\boldsymbol{\vartheta}$, finding the MLE of the parameters $\boldsymbol{\vartheta}$ obviously means that the negative log-likelihood satisfies the following equations:

$$-\frac{\partial \ln \mathcal{L}(\mathbf{x}; \hat{\boldsymbol{\vartheta}})}{\partial \theta_j} = 0 \quad \text{for } j = 1, \dots, m \quad (\text{A.3})$$

Moreover, the likelihood function must be constructed by normalised functions $f(x_i; \boldsymbol{\vartheta})$ that do not depend on the parameters $\boldsymbol{\vartheta}$:

$$\int f(x_i; \boldsymbol{\vartheta}) dx = 1. \quad \text{so that} \quad \int \mathcal{L}(\mathbf{x}; \boldsymbol{\vartheta}) dx_1 dx_2 \dots dx_m = 1 \quad (\text{A.4})$$

A.2 Uncertainty estimation assuming Gaussian Errors

In practice, according to the *Central limit theorem* (CLT), one can assume that the measurements x_i are spread around an unknown true value θ following a Gaussian distribution with known width σ_i regardless of the their individual probability functions. Then, the probability density function is given by

$$f(x_i; \theta, \sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x_i-\theta)^2}{2\sigma_i^2}} \quad (\text{A.5})$$

which leads to the following likelihood function:

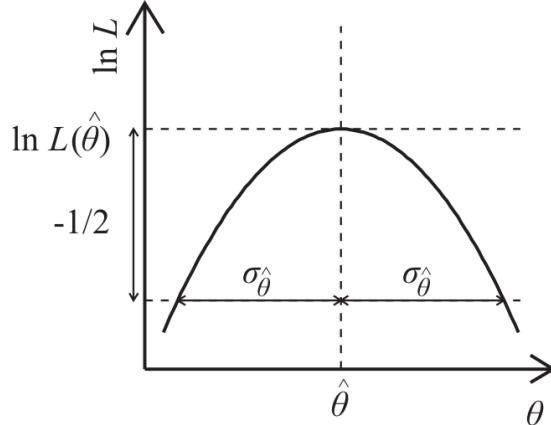
$$\mathcal{L}(\theta) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x_i-\theta)^2}{2\sigma_i^2}} \quad (\text{A.6})$$

and the corresponding log-likelihood function is

$$\ln \mathcal{L}(\theta) = -\frac{1}{2} \sum_{i=1}^N \frac{(x_i-\theta)^2}{\sigma_i^2} + \text{const.} \quad (\text{A.7})$$

which represents a parabola equation. This parabola will have a maximum at some value $\hat{\theta}_i$ which corresponds to the MLE of the parameter θ_i .

Figure A.1: [12] The uncertainty $\sigma_{\hat{\theta}}$ can be read off from the two points where $\ln \mathcal{L}$ drops by $1/2$ from its maximum value at $\hat{\theta}$.



Moreover, the width of the Gaussian defines the standard deviation $\sigma_{\hat{\theta}}$. Using equation (A.7), it is pretty straightforward to show that the following expression holds true:

$$\ln \mathcal{L}(\hat{\theta}_i \pm \sigma_{\hat{\theta}_i}) - \ln \mathcal{L}(\hat{\theta}_i) = -1/2 \quad (\text{A.8})$$

Hence, by adding $-1/2$ to the log-likelihood function at the MLE point, one can retrieve the standard deviation of an estimator $\hat{\theta}_i$ as illustrated in Figure A.1. As already mentioned though, in many practical cases including the present work, the negative log-likelihood is used. In order to estimate the standard deviation of the desired parameter $\sigma_{\hat{\theta}_i}$, the following expression needs to be taken under consideration:

$$\ln \mathcal{L}(\hat{\theta}_i \pm \sigma_{\hat{\theta}_i}) - \ln \mathcal{L}(\hat{\theta}_i) = +1/2 \quad (\text{A.9})$$

Finally, there is a chance that the log-likelihood is not totally symmetric around the minimum in contrast to the log-likelihood plot of Figure A.1. As a result, the lower and upper standard deviation values may differ.

B Claiming discovery in a counting experiment

In order to claim the discovery of a new signal process, one defines a null hypothesis H_0 corresponding to the background-only hypothesis. On the other hand, an alternative hypothesis H_1 is defined including both background and the desired signal processes. To quantify the level of agreement between the measurement and a given hypothesis H , one computes a p-value. Essentially, the p-value expresses the probability, under the assumption of H being true, of encountering data that exhibit equal or greater inconsistency with what H predicts. In other words, a lower p-value means that such an extreme observed outcome would be very unlikely assuming the hypothesis H .

In a counting experiment like the one that the present work refers to, one can assume that the n observed events in a signal sample follow a Poisson distribution with an expectation value of

$$E[n] = \mu s + b \quad (\text{B.1})$$

where s and b correspond to the signal and background model prediction respectively. The parameter μ is called the *strength parameter* of the desired signal process. From equation (B.1), it is evident that a value $\mu = 0$ corresponds to the background-only hypothesis.

In addition to the n measurements in the signal sample, one can perform additional *control* measurements to constrain the background b which can be treated as a nuisance parameter. Assuming m events in this measurement, one can consider them Poisson distributed with a mean value

$$E[m] = \tau b \quad (\text{B.2})$$

where τ is a corresponding scaling factor referring to the background b instead.

Consequently, the data consist of two measurements, n and m , while μ is the parameter of interest and b is treated as a nuisance parameter. Hence, based on the Poisson distribution definition, the total likelihood function is the product of the two corresponding Poisson terms:

$$\mathcal{L}(\mu, b) = \frac{(\mu s + b)^n}{n!} e^{-(\mu s + b)} \frac{(\tau b)^m}{m!} e^{-\tau b} \quad (\text{B.3})$$

Assuming a counting experiment where the mean background b is known with negligible uncertainty and can be treated as a constant, the data consist only of the n measurement. Hence the likelihood function is simply:

$$\mathcal{L}(\mu, b) = \frac{(\mu s + b)^n}{n!} e^{-(\mu s + b)} \quad (\text{B.4})$$

To test a hypothesized value of μ , the profile likelihood ratio is considered:

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\theta})}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \quad (\text{B.5})$$

where $\hat{\theta}$ denotes the value of θ that maximizes \mathcal{L} for the specified μ . The $\hat{\mu}$ and $\hat{\theta}$ are the MLE estimators, i.e. the parameters that maximize the likelihood function.

A widely established methodology to claim a discovery is to perform a frequentist significance test using a likelihood ratio as test statistic. Rejecting the $\mu = 0$

hypothesis leads to signal discovery. Assuming $\mu \geq 0$, one can use the following test statistic for discovery, to test the $\mu = 0$ hypothesis

$$q_0 = \begin{cases} -2\ln\lambda(0), & \hat{\mu} \geq 0 \\ 0, & \hat{\mu} < 0 \end{cases} \quad (\text{B.6})$$

where $\lambda(0)$ is the profile likelihood function ratio given by equation (B.5) corresponding to the background-only hypothesis for $\mu = 0$.

According to equations (B.4) and (B.5), the discovery statistic given in equation (B.6) for $\hat{\mu} > 0$, can be written as follows:

$$\begin{aligned} q_0 &= -2\ln\frac{\mathcal{L}(0)}{\mathcal{L}(\mu)} \\ &= -2\ln\frac{b^n}{(\hat{\mu}s + b)^n e^{-\hat{\mu}s}} \\ &= -2n \ln\frac{b}{\hat{\mu}s + b} - 2\hat{\mu}s \end{aligned} \quad (\text{B.7})$$

where $\hat{\mu}$ is the strength parameter expressing the discrepancy between the measurement and the model prediction for signal. Hence, making a measurement of the signal by subtracting the background model prediction from the signal data sample as $n - b$, the strength parameter is written as follows:

$$\hat{\mu} = \frac{n - b}{s} \quad (\text{B.8})$$

Consequently, the discovery statistic to test the background-only hypothesis expressed as $\mu = 0$, is written as

$$q_0 = 2\left(n \ln\frac{n}{b} + b - n\right) \quad (\text{B.9})$$

To quantify the level of disagreement between the measurement and the background-only hypothesis, one utilizes the observed value of q_0 to compute the p -value as follows

$$p_0 = \int_{q_0,obs}^{\infty} f(q_0|\mu = 0) dq_0 \quad (\text{B.10})$$

where $f(q_0|\mu = 0)$ denotes the pdf of the q_0 statistic under the assumption of the background-only hypothesis.

In particle physics, the p -value is usually converted into an equivalent significance Z , defined such that a Gaussian distributed variable found Z standard deviations above its mean, has an upper-tail probability equal to p . This statement is expressed as

$$Z = \Phi^{-1}(1 - p) \quad (\text{B.11})$$

where $\Phi^{-1}(1 - p)$ is the inverse of the cumulative distribution of the standard Gaussian.

For sufficiently large background b , one can use the following simple asymptotic formula for the Z_0 significance:

$$Z_0 = \sqrt{q_0} \quad (\text{B.12})$$

Utilizing equation (B.9), one obtains the following expression for significance Z_0 , concerning the background-only hypothesis:

$$Z_0 = \sqrt{2 \left(n \ln \frac{n}{b} + b - n \right)} \quad (\text{B.13})$$

In the present work, the sensitivity of the experiment is expressed by reporting the median (expected) significance that one would obtain for a variety of signal hypotheses. The median significance is estimated by using a single representative set of simulated data instead of an ensemble of simulated data. This representative data set is called "*Asimov*" data set. Replacing the n data value by the Asimov value $s + b$ one assumes the nominal signal hypothesis $\mu = 1$ according on Eq.B.8. Hence, one obtains an expression for the median significance based on equation (B.13):

$$\text{med}[Z_0|\mu = 1] = \sqrt{2((s+b)\ln(1+s/b) - s)} \quad (\text{B.14})$$

Expanding the logarithm in s/b one finds that $Z_0 \approx s/\sqrt{b}$ which is widely used when $s \ll b$. The final result (B.14) for the median significance is the ultimate criterion used throughout the entire present work in order to discriminate the desired signal process $ZZ \rightarrow \ell\ell\nu\nu$ from the background.

C Signal Region Optimization with Boosted Decision Trees

A complementary method to the ones followed in section 3 based on kinematic variables, is presented in the current appendix. This method for the signal region optimization utilizes the Machine Learning (ML) library TMVA [13], offered by the CERN ROOT framework. Only MC samples are used in this method. The analysis is performed in the same preliminary SR defined for optimization in Table 2.1. The MC samples for the desired signal process and corresponding background types in the SR are given as input to the algorithm.

Specifically, the analysis utilizes the *Boosted Decision Trees with Gradient Boosting* (BTG) as *classifier*, implemented in the TMVA library, in order to classify events as signal or background correspondingly according to various kinematic variables given as input.

A decision tree, is a hierarchical binary structure classifier used for making decisions. This process involves a series of yes/no decisions on one single variable at a time, that divide the input space into distinct regions. Each region is eventually classified as signal or background based on the majority of training events in the final *leaf node*.

Boosting extends the decision tree concept to several trees making up a *forest*. The trees in the forest are finally combined into a single classifier which is given by the average of the individual decision trees.

Various configuration settings called *hyperparameters*, have been tried out to define an effective classifier. The classifier parameters used in the present work in order to define the classifier, are shown in Table C.1. The NTrees is the number of trees in the forest. The MinNodeSize is the percentage of training events required in a leaf node. The BoostType is the boosting type for the trees in the forest. The gradient boosting has been selected in the present work. Finally, Shrinkage refers to the learning rate for the Gradient Boosting algorithm.

Table C.1: Hyperparameters configuration used to define an effective BDT classifier.

BDT Classifier Hyperparameters	
Parameter	Value
NTrees	400
MinNodeSize	10%
BoostType	Grad
Shrinkage	0.2

The samples are splitted in a *training* sample and a *testing* sample. The training sample is used to create the model that classifies the events as signal or background, called *classifier*, i.e. the BTG classifier. The testing sample is used to evaluate the input variables given. In this approach, several kinematic variables are used as input in order to evaluate their discriminating power in terms of signal-to-background separation.

The MC events in SR are given as a function of some input variables used for classification in Figure C.1. The kinematic variables show consistent behaviour compared to the plots presented in section 3.1. More specifically the algorithm ranks the kinematic variables according to their discriminating power. The variable

with the highest *separation* value has the greatest impact in terms of discriminating signal from background events. The results are shown in Table C.2.

Figure C.1: Plots of the kinematic variables E_T^{miss} , E_T^{miss}/H_T , $\sigma(E_T^{miss})$, $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$, p_T^Z and n_{jets} for the MC signal and background yields in the Signal Region utilizing the TMVA library. The plots have been created using the preliminary SR defined by the cuts shown in Table 2.1.

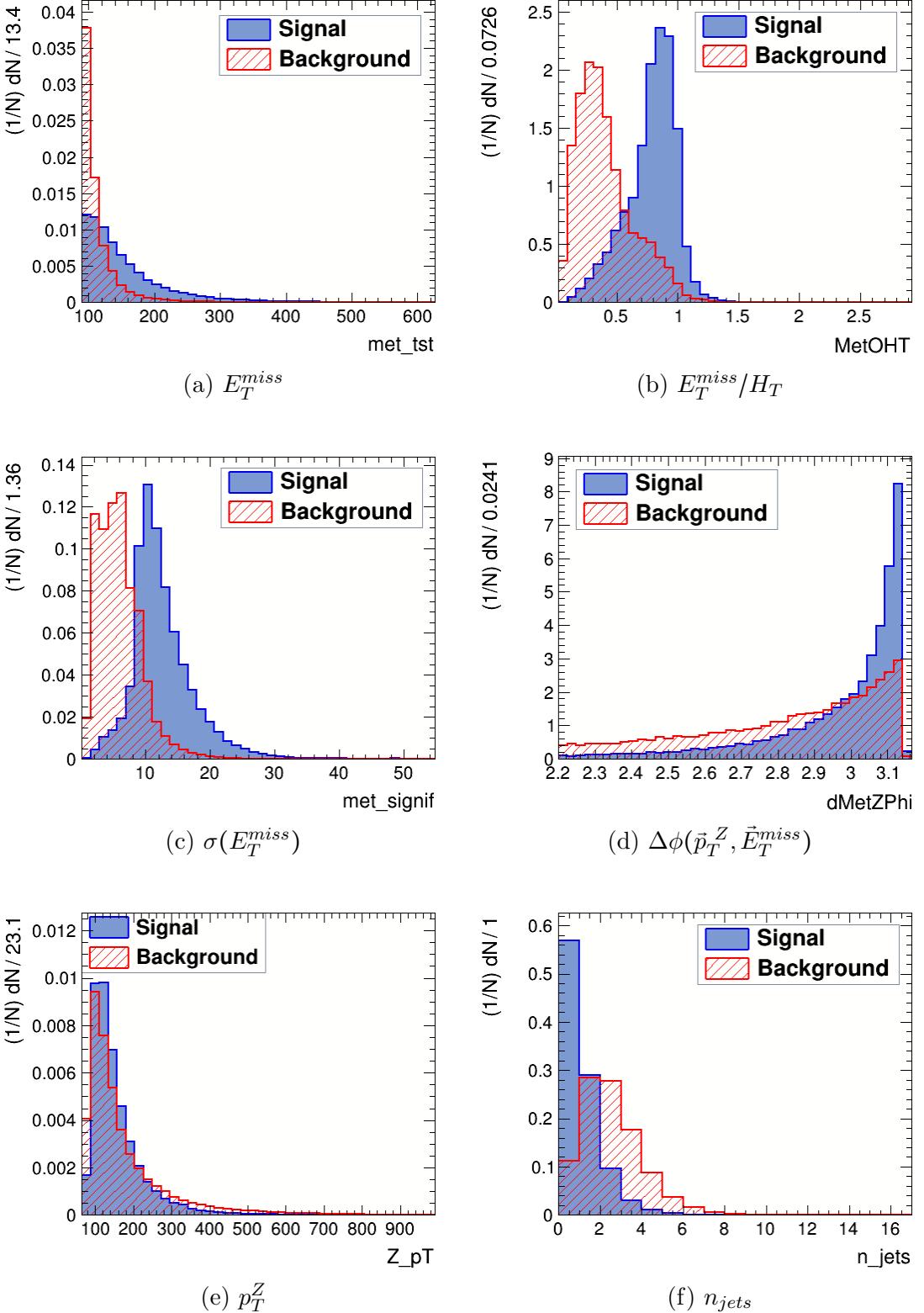


Figure C.2: Correlation coefficient matrices of the kinematic variables in the Signal Region, defined in Table 2.1. The correlation coefficients are calculated between any two variables given as input according to the formula C.1. The matrices have been created by the training process for the MC signal and background events.

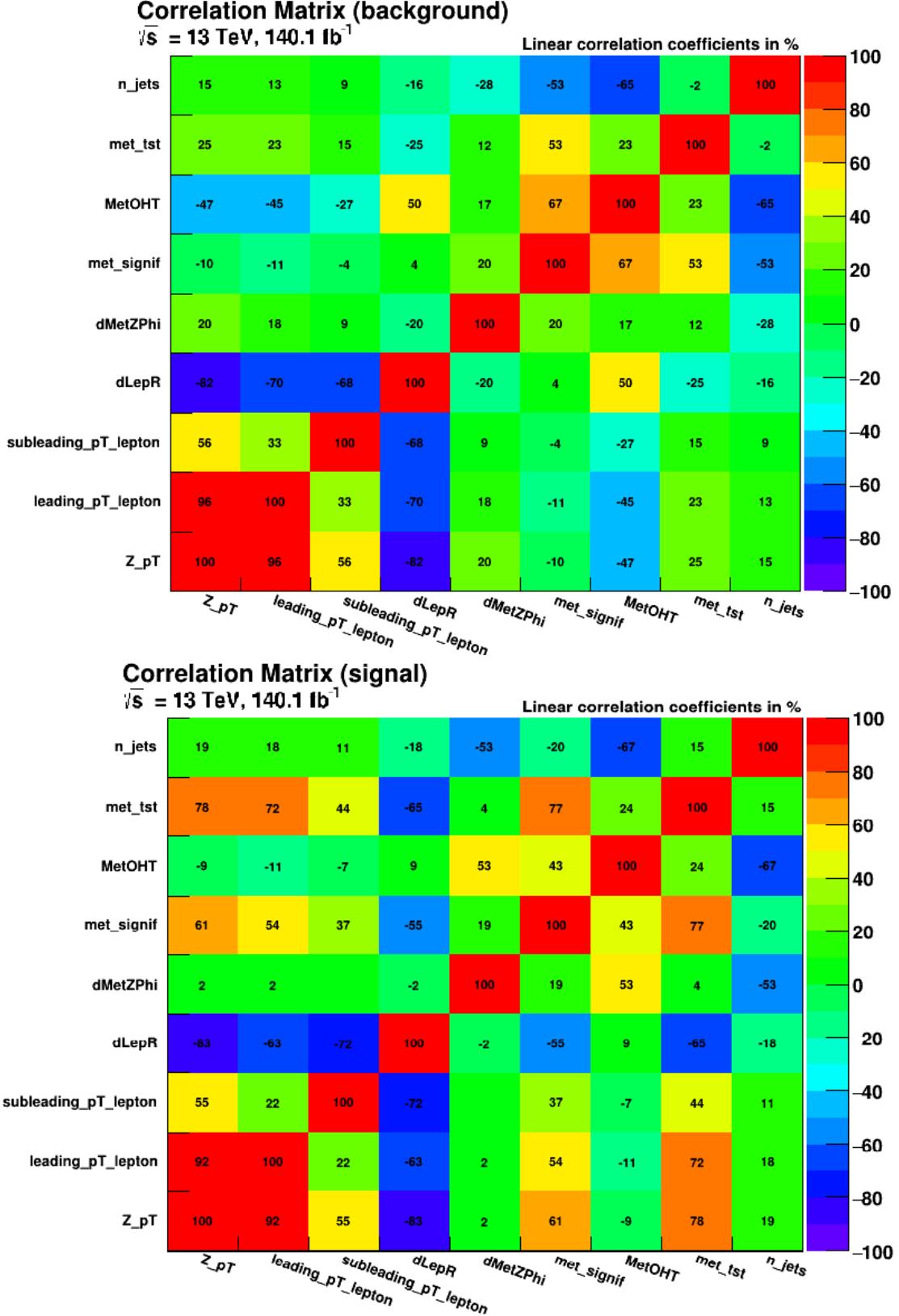


Table C.2: Ranking of the kinematic variables given as input in terms of their signal-to-background discriminating power. Variables with higher separation value are considered to have better discriminating power.

Kinematic Variables Ranking	
Variable	Separation
$\sigma(E_T^{miss})$	0.5041
E_T^{miss}/H_T	0.4513
n_{jets}	0.3050
E_T^{miss}	0.2126
$\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$	0.1119
p_T^Z	0.0354
$p_T^{\ell 1}$	0.0259
$\Delta R_{\ell\ell}$	0.0202
$p_T^{\ell 2}$	0.0061

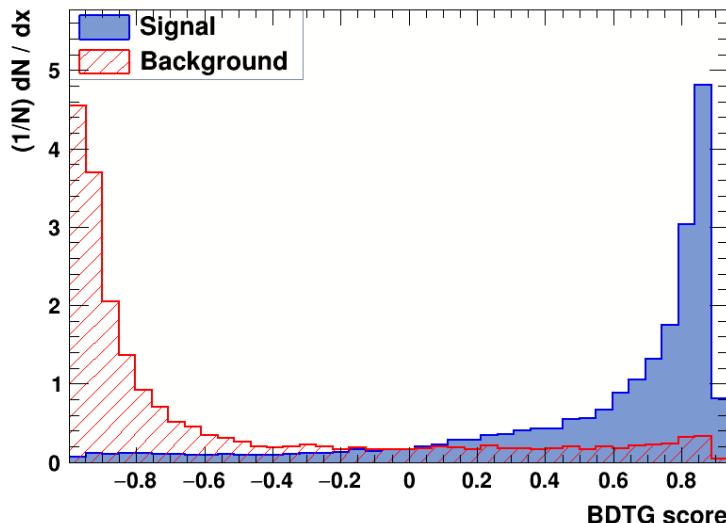
The training provides the correlation matrices for the input variables with respect to the events classified as signal or background. The correlation matrices contain the linear correlation coefficient between any two variables X and Y used as input in the training process. The correlation coefficient is given by the following formula:

$$\rho(X, Y) = \frac{cov(X, Y)}{\sigma_X \sigma_Y} \quad (\text{C.1})$$

The correlation matrices obtained by the training sample for the MC signal and background events, are presented in Figure C.2 accordingly.

The classifier response can be evaluated by the classifier response distribution. The classifier selected in the present work has been defined by the hyperparameters shown in Table C.1. In the classifier response distribution, the MC events are plotted according to the classifiers' confidence on characterizing them as signal or background events accordingly.

Figure C.3: Response distribution of the BTG classifier defined in Table C.1. Events with higher BTG score are considered to have higher probability of being signal events according to the BTG classifier. The classifier shows good performance in terms of signal-to-background discrimination.



The classifiers' performance is quantified by a classification score spanning from -1 to +1. Higher score values imply a higher probability of an event being signal while lower values indicate higher probability for an event to be background. The classifier response distribution is presented in Figure C.3.

According to the response distribution, the BTDG classifier expresses effectiveness in discriminating signal from background events since the majority of the MC signal events are characterized by a score close to +1 indicating higher probability of being a signal event. On the other hand, the majority of MC background events is characterized by score values close to -1 indicating higher probability that these events are actually background according to the classifier. Moreover, there is minimal overlap between signal and background events.

D Particle Swarm Optimization Algorithm

The Particle Swarm Optimization (PSO) algorithm is a type of heuristic algorithm inspired by the social behavior of swarms of birds. It has been proposed by Kennedy and Eberhart [10]. The PSO algorithm is widely used due to its fast convergence and sufficient performance using a small size of individuals as opposed to genetic algorithms that generally demand much higher values of individuals in order to have a sufficient performance. Still, a major drawback of PSO is getting stuck in local minima.

In PSO, a population of candidate solutions called *swarm*, represented as *particles*, explores a search space to find the optimal solution to a given problem. Each particle adjusts its position and velocity based on its own experience and the best solution found by the entire swarm. In the present work, every particle constitutes an expression of a set of cut values on the kinematic variables ($\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$, E_T^{miss} , E_T^{miss}/H_T).

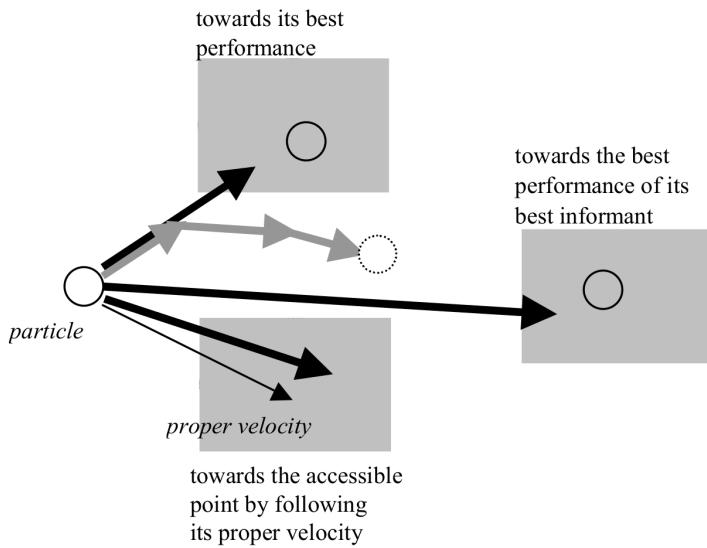


Figure D.1: [14] The three fundamental criteria according to which, the next position of the particle is decided.

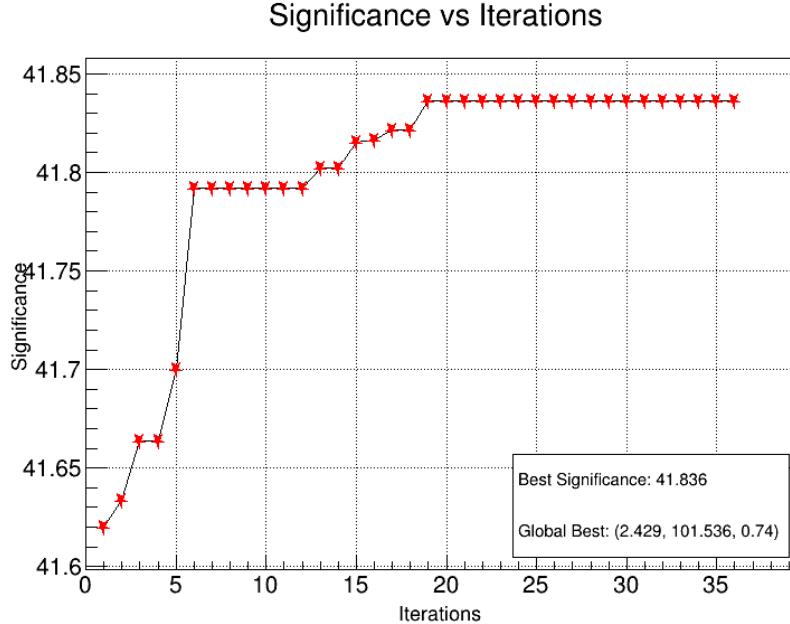
The PSO algorithm employs a cooperative, swarm-based approach, where particles communicate and share information to guide the search towards the most promising regions of the search space.

Through iterations, PSO converges towards an optimal solution by continually updating the particle positions. A basic concept of the procedure followed by a PSO algorithm can be summarized in the following steps:

1. A population of particles with random positions and velocities is initialized in the search space.
2. The fitness of each particle's current position in the search space is evaluated.
3. The personal best solution of each particle is updated according to its fitness value.
4. The global best solution among all the personal best solutions is determined.
5. The positions and velocities of all particles are updated using mathematical equations that take into account the personal and global best solutions.

6. Steps 2-5 are repeated until a convergence criterion is met (e.g. the algorithm reaches a plateau on global significance - Fig. D.2) or for a specified number of iterations.

Figure D.2: Example of a swarm (global) significance plateau. The best signal significance of the swarm is plotted versus the iteration number. The significance value clearly reaches a plateau. The global best vector corresponds to the set of optimal cut values in the search space ($\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$, E_T^{miss} , E_T^{miss}/H_T) presented in section 3.3.



D.1 Initialization

During the initialization process, the particles are simply placed randomly according to a uniform distribution in the search space. Moreover the velocity of each particle is initialized in order to avoid too many particles leaving the search space.

Hence the initialization procedure of each particle i , is described by the following expression:

$$\begin{cases} \mathbf{x}_i^0 \leftarrow U(\mathbf{x}_{min}, \mathbf{x}_{max}) \\ \mathbf{v}_i^0 \leftarrow U((x_{min} - x_{max})/2, (x_{max} - x_{min})/2) \end{cases} \quad (\text{D.1})$$

where U refers to a uniform distribution function.

Concerning the swarm size, the optimal size proposed is between 20 and 30 particles by experimenters which have been proven adequate for most problems. However, in the present work, the complexity of the problem requires 40 to 50 particles in order to scan a sufficiently large portion of the search space.

D.2 Search Space

Assuming, a particle i at iteration t , its position and velocity can be denoted as \mathbf{x}_i^t and \mathbf{v}_i^t respectively. The particle moves to its next position according to its prior personal best position ($\mathbf{p}_{best,i}$), the global best position (\mathbf{g}_{best}) of the swarm as well

as its previous velocity:

At each iteration the position \mathbf{x} and velocity \mathbf{v} is calculated for each particle i:

$$\begin{cases} \mathbf{v}_i^{t+1} \leftarrow \omega \mathbf{v}_i^t + c_{max} r_1 (\mathbf{p}_{best,i}^t - \mathbf{x}_i^t) + c_{max} r_2 (\mathbf{g}_{best}^t - \mathbf{x}_i^t) \\ \mathbf{x}_i^{t+1} \leftarrow \mathbf{x}_i^t + \mathbf{v}_i^{t+1} \end{cases}$$

where the symbols r_1 and r_2 denote two random numbers generated according to a uniform distribution $U(0, 1)$.

The three components of the velocity expression, represent the three fundamental criteria taken into account by the particle in order to update its position. The first component encourages the particle to follow its previous velocity \mathbf{v}_i^t . The second component encourages the particle to move towards its personal best position $\mathbf{p}_{best,i}^t$ meaning the set of variable values that attribute the maximum signal significance. The third, component encourages the particle to move towards the best position among all the particles \mathbf{g}_{best}^t . The three fundamental criteria are shown in Figure D.1.

The values of the ω and c_{max} parameters must be chosen carefully to balance *exploration* and *exploitation* int the optimization process.

The parameter ω is called inertial weight and it controls the influence of the previous velocity on the current velocity. A higher ω value encourages exploration meaning searching the search space for new maxima.

As the PSO algorithm progresses, the swarm gains information about an increasingly larger portion of the search space. For this reason, one may want to progressively decrease the value of the ω parameter in order to encourage exploration in the initial stages of the search and as the procedure continues, to encourage exploitation. Hence, in the earlier stages, the particles prioritize exploration in order to find new local maxima and gain new information. As the algorithm progresses the swarm prioritizes a more extensive scan of the local maxima that are potential global maxima of the search space. A formula for ω that expresses this mechanism is the following [15]:

$$\omega = \omega_{max} - \frac{(\omega_{max} - \omega_{min})}{t_{max}} \cdot t \quad (D.2)$$

where ω_{max} and ω_{min} are the values that the inertial weight starts and ends up with respectively. The symbols t and t_{max} refer to the current and total number of iterations.

On the other hand, the parameter c_{max} encourages exploitation meaning exploiting the information already acquired by the particle itself and the other particles in the swarm. A higher value of c_{max} means that the particle focuses more on local maxima already acquired in the search space. This parameter should be neither too small or too large.

The c_{max} and ω parameters should be neither too small or too large. Moreover, the values of ω and c_{max} shouldn't be chosen independently in order to achieve good convergence [16]. Other studies propose a variable ω parameter decreasing as the process evolves [15].

Two good examples for the pair of values (ω, c_{max}) have proved to be $(0.7, 1.47)$ and $(0.8, 1.62)$ [16]. Numerous trials have been carried out with these two pairs of parameters (ω, c_{max}) as well as other combinations with values close to them in order to gain confidence on the algorithms' performance. Most combinations

including a variable ω parameter, resulted in fast convergence to high signal significance values. However, the combination (0.7, 1.47) generally resulted in higher significance values.

Finally, the number of the particles of the swarm plays a crucial role in the evolution of the system towards the best solution. Again, there is no golden rule for the swarm number and various swarm numbers have been tried out. While, a bigger swarm may result in larger areas of the search space scanned, a smaller swarm results in more faster iterations since less particles need to update their position and velocity for each iteration. Of course the swarm size must not be too small but rather large enough in order to attribute information for an adequate portion of the search space.

Concerning the swarm size, the optimal size proposed is between 20 and 30 particles by experimenters which have been proven adequate for most problems. However, in the present work, the complexity of the problem requires 40 to 50 particles in order to scan a sufficiently large portion of the search space.

D.3 Search Space Confinement

During the PSO procedure, the position update formula of a particle given by the expression (eq.D.2) may make the particle move towards a point that lies outside the search space initially defined. In many cases, this behaviour may have a negative impact on the optimization procedure.

For example, in the present work, leaving the search space may result in a Signal Region that produces zero events. Therefore, it is crucial to add a mechanism that prevents the particles from leaving the search space.

A very simple mechanism is the *interval confinement*. In this case, when a particle position update indicates a point outside the search space, the new position is assigned the nearest border position along the direction proposed by the velocity.

However, this simple method has a disadvantage as well. A very common scenario is that the velocity keeps its previous direction. Hence, the particle is brought back to the border but its velocity proposes a new position beyond the border again.

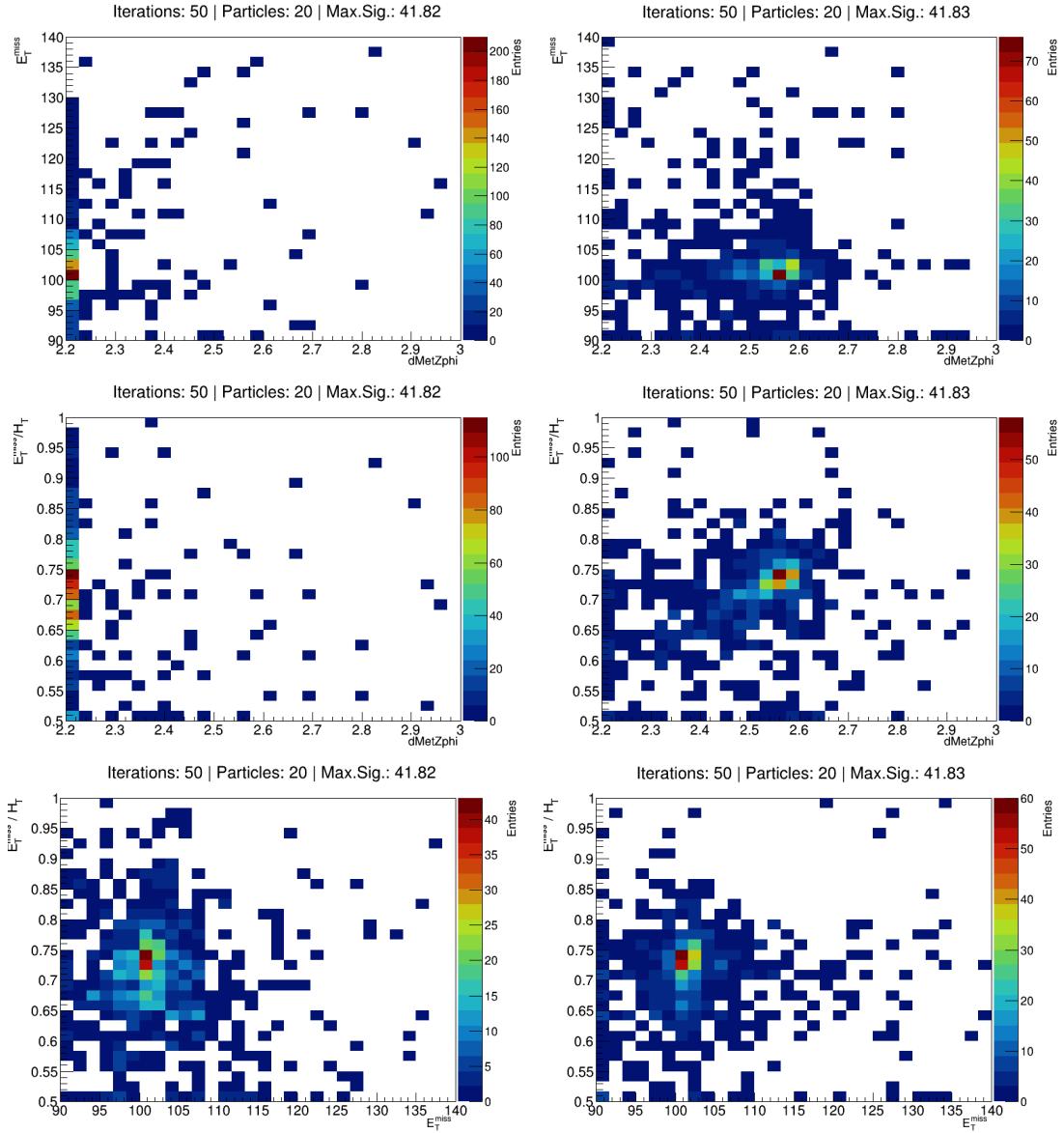
A simple solution is to simply cancel the velocity by setting its value to 0. Numerous trials with this choice did not solve the issue as the particles kept getting stuck on the borders since the velocities may keep their previous direction.

This behavior is mostly observed for the $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ variable which shows no strong preference for a specific value cut (see section 3.1). The particles keep getting driven and stuck on the lower boundary of the $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ dimension of the search space since values on the upper boundary generate lower signal significance. This behaviour is highly suppressed by updating the velocity according to the mechanism introduced by the expression (eq.D.3).

This formula provides an adequate solution by assigning to the velocity vector that leads the particle beyond the search space, its opposite velocity multiplied by a randomly generated number. The complete confinement mechanism is then described by the following particle update expression:

$$x_i^{t+1} \notin [x_{min}, x_{max}] \Rightarrow \begin{cases} x_i^{t+1} < x_{min} \Rightarrow x_i^{t+1} \leftarrow x_{min} \\ x_i^{t+1} > x_{max} \Rightarrow x_i^{t+1} \leftarrow x_{max} \\ v_i^{t+1} \leftarrow -r \cdot v_i^{t+1} \end{cases} \quad (\text{D.3})$$

Figure D.3: The left column of plots presents a typical example of a swarm getting stuck on the search space border due to assigning zero to the updated velocity as a part of the *interval confinement mechanism*. The right column of plots presents an example of a swarm searching by updating its velocities according to equation (D.3). The plots show the 2-dimensional search spaces between all three variables used for optimization after 50 updates of their position and velocity (iterations).



(a) 2-dimensional plots of the swarm scanning the search space by assigning zero to the velocity when it exceeds the boundaries. (b) 2-dimensional plots of the swarm scanning the search space by using the formula (D.3) when it exceeds the boundaries.

where r is a number randomly generated according to a uniform distribution $\sim U(0, 1)$.

In contrast with the $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ variable, the other two variables E_T^{miss} and E_T^{miss}/H_T scan the space normally before converging to a maximum.

An example of the difference between assigning zero to the velocity that exceeds the boundaries and using the update formula (D.3), is presented in Figure D.3.

The zero velocity case presented in the left column (Fig. D.3a) corresponds to a global best set of cuts of (2.20, 101, 0.74) resulting in 41.82 signal significance. The 2.20 value obtained by assigning zero to the velocity corresponds to the lower boundary of the $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ dimension of the search space.

The most vulnerable variable subjected to this behaviour is the $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$ for which, the swarm gets stuck at the lower boundary. However, by updating the velocity according to equation (D.3), the swarm normally searches the search space (top and middle diagrams).

The scan using the confinement mechanism formula (D.3) presented in the right column (Figure D.3b) corresponds to an optimal set of cuts (2.57, 101, 0.74) resulting in 41.83 signal significance.

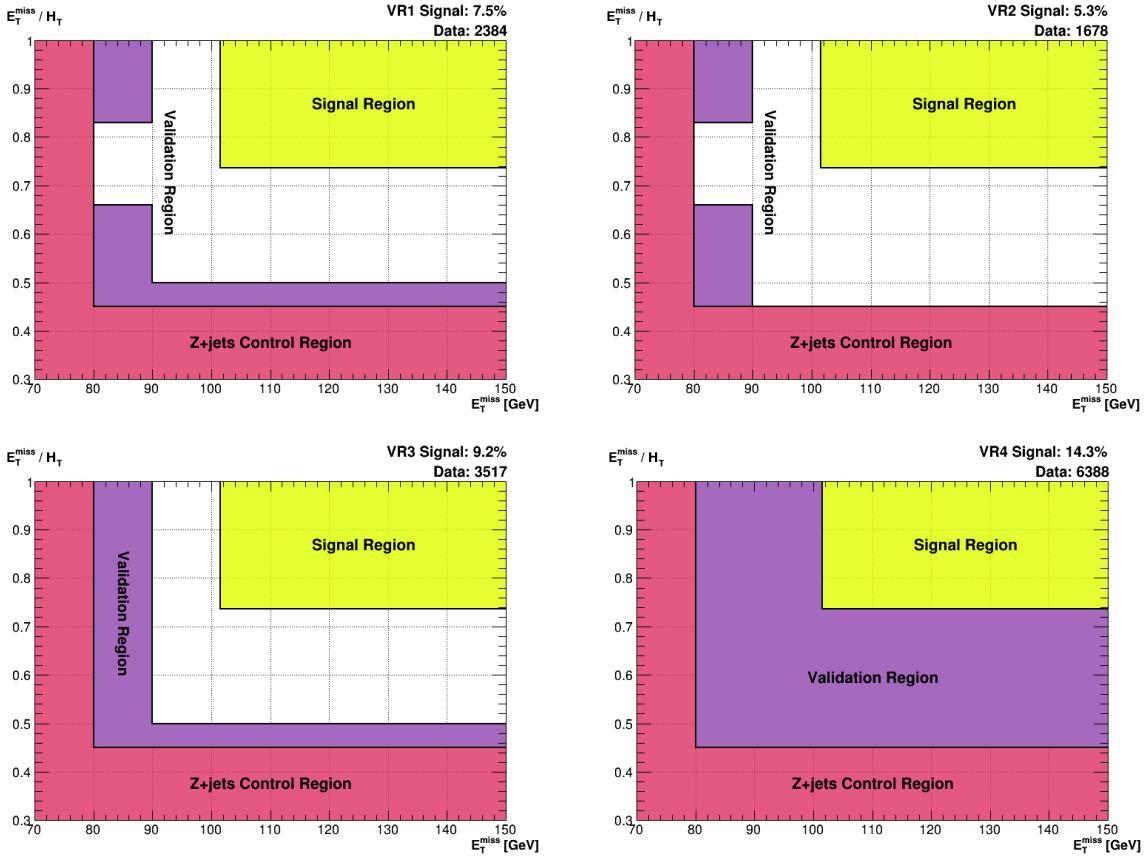
The E_T^{miss} and E_T^{miss}/H_T variables converge to a maximum by normally updating their values in both cases (bottom plots).

E Study of the Validation Region

During the background estimation, various types of background have been studied. A scaling factor has been estimated for each major background contribution as well as for the signal yield. The scaling factors account for the discrepancy between the observed events in the data sample and the predicted MC yields. In order to evaluate their performance in predicting the corresponding data before applying them in the Signal Region, one can define a *Validation Region* (VR).

The VR is required to be similar to the SR and totally independent from any other region defined as well. Moreover, the VR must contain minimal signal events since the data sample is accessed.

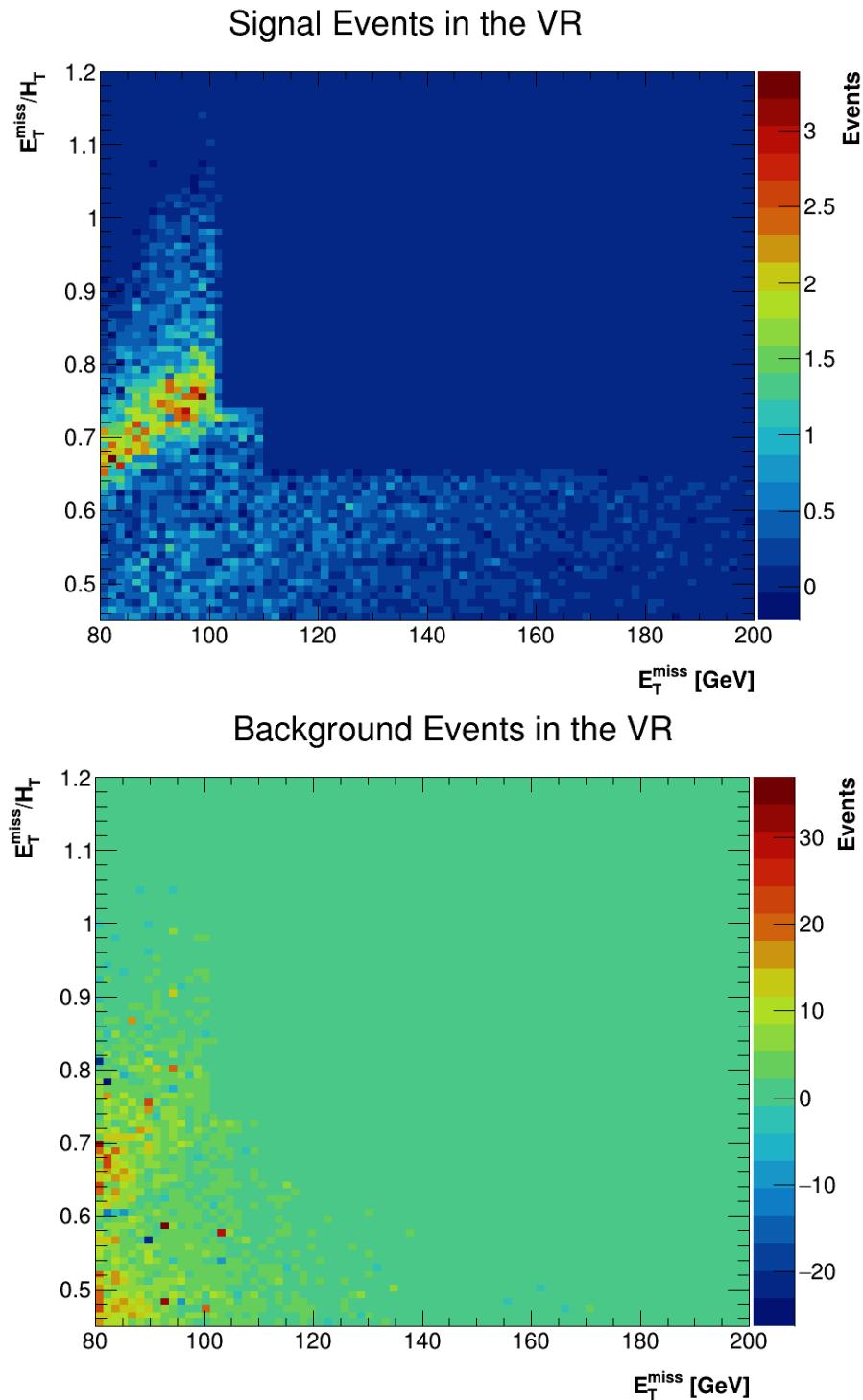
Figure E.1: Validation Regions defined for different cuts on the 2-d (E_T^{miss} , E_T^{miss}/H_T) plane and the corresponding signal percentage and data yield. The regions plotted correspond to the cuts shown in Table E.1. No upper limits are imposed. The plane upper limits on the plots serve visual purposes only. The signal percentage refers to the ratio over the unsealed total events.



An unexploited area has occurred as a result of defining the SR and Zjets region as shown in Figure 4.3. One can try out different cut combinations in this area to define a VR motivated by the aforementioned criteria.

However, since the VR is pretty close to the SR, isolating the signal events while maintaining a substantial amount of statistics is a difficult task. For this reason, various candidate regions have been studied. Combinations of high E_T^{miss} and E_T^{miss}/H_T simultaneously, result in regions enriched in signal. At the same time, a broad range of the (E_T^{miss} , E_T^{miss}/H_T) phase space must be included in order to achieve a region similar to the SR.

Figure E.2: Distribution of the signal and background in the Validation Region. The high concentration of the signal in a relatively restricted area of the VR suggests the isolation of the area defined by the cuts $0.6 < E_T^{miss} < 0.8$ and $E_T^{miss} < 101$.



Four different regions considered representative are presented in Figure E.1. The corresponding cuts imposed on the E_T^{miss} and E_T^{miss}/H_T variables, are presented in Table E.1. The corresponding data yields are presented along with the signal-to-total yield ratio and the data/prediction ratio. The MC yields presented are scaled according to the best-fit values obtained by utilising a likelihood function (see Table 4.6). The rest of the criteria are the same used to define the primary Zjets region as well (see Table 4.3).

Table E.1: Data yield, signal abundance and Data/Prediction for different phase spaces in the VR. Four representative validation regions are presented. The MC yields are scaled by utilising the best-fit factors obtained during the MLE analysis (see Table 4.6). The regions presented in this table correspond to the plots shown in Figure E.1.

Validation Region Study					
Cuts					
Cut1: $80 < E_T^{miss} < 90$ & $0.45 < E_T^{miss}/H_T$			Cut4: $80 < E_T^{miss} < 101$ & $0.45 < E_T^{miss}/H_T$		
Cut2: $80 < E_T^{miss} < 90$ & $0.83 < E_T^{miss}/H_T$			Cut5: $90 < E_T^{miss} & 0.45 < E_T^{miss}/H_T < 0.5$		
Cut3: $80 < E_T^{miss} < 90$ & $0.45 < E_T^{miss}/H_T < 0.66$			Cut6: $101 < E_T^{miss} & 0.45 < E_T^{miss}/H_T < 0.74$		
Region	Cut	Data Yield	Signal / Total (%)	Data / Scaled Prediction (%)	
VR1	Cut2 or Cut3 or Cut5	2384	7.5	99.5	
VR2	Cut2 or Cut3	1678	5.3	99.7	
VR3	Cut1 or Cut5	3517	9.2	100.0	
VR4	Cut4 or Cut6	6388	14.3	98.2	

During, the study of the VR a more thorough study of the signal distribution in the $(E_T^{miss}, E_T^{miss}/H_T)$ plane has been carried out. The signal and background events have been plotted in the 2D plane in order to get a sense of the areas enriched with signal (Figure E.2).

Isolating the region enriched in signal, indeed reduces the signal-to-total events ratio (VR1). Moreover, totally excluding regions with high E_T^{miss} values further reduces the ratio (VR2). However, the requirement of a VR similar to the SR suggests a region extending over a wide continuous range (VR3). The total region occupied between the Zjets and Signal regions (VR4) is presented in Figure E.1 for comparison.

Taking under consideration all the criteria mentioned, the VR3 is considered the most representative region to test the scaling factors with, since it spans in a wide range of E_T^{miss} , E_T^{miss}/H_T meaning it is adequately similar to the SR. At the same time, it balances out a substantial amount of statistics with an acceptable signal presence.

F Event Yields Tables

The event yields obtained from all the regions, Control, Signal and Validation, are summarized. The yields refer to every process events as they are estimated in the phase space defined by each region definition criteria without scaling of any kind. Both Data and Monte Carlo samples are presented for the Control and Validation Regions while the data sample is blinded for the Signal Region.

F.1 Control Region Event Yields

Table F.1: Event yields for Data and Monte Carlo Samples in the 3ℓ Control Region. The yields are presented as estimated according to the region phase space definition (Table 4.1) without scaling of any kind.

3 ℓ CR Event Yields	
Event Type	3ℓ
DATA	2409.00 ± 49.08
Signal ZZ	
llvv	2.19 ± 0.54
llvvjj	0.10 ± 0.01
Signal(ZZ)	2.29 ± 0.54
WZ	
WZ	2098.99 ± 10.23
Z+jets	
Z+jets0	35.74 ± 8.28
Z+jets1	34.48 ± 5.81
Z+jets2	15.70 ± 4.09
Z+jets Total	85.92 ± 10.91
Top	
single top	50.27 ± 1.47
ttbarV-ttbarVV	0.53 ± 0.07
Wt	6.16 ± 1.06
Top Total	56.96 ± 1.82
WW	
WW	0.6 ± 0.14
Other	
llll	77.29 ± 0.90
llqq	0.67 ± 0.14
VVV	13.41 ± 0.11
W+jets	0.00 ± 0.00
Ztt	0.00 ± 0.00
WZjj	22.45 ± 0.16
lllljj	0.28 ± 0.02
llvvjj+WW	0.00 ± 0.00
Other Total	114.10 ± 0.93
WZ / Total (%)	88.98
Signal / Total (%)	0.10
DATA / MC	1.02

Table F.2: Event yields for Data and Monte Carlo Samples in the $e\mu A$ and $e\mu B$ Control Regions. The yields are presented as estimated according to each region phase space definition (Table 4.2) without scaling of any kind.

Non-Resonant CR Event Yields		
Event Type	$e\mu A$	$e\mu B$
DATA	2903.00 ± 53.88	5736.00 ± 75.74
Signal ZZ		
llvv	0.41 ± 0.20	0.00 ± 0.00
llvvjj	0.00 ± 0.00	0.00 ± 0.00
Signal(ZZ)	0.41 ± 0.20	0.00 ± 0.00
WZ		
WZ	22.30 ± 1.02	0.87 ± 0.17
Z+jets		
Z+jets0	0.63 ± 0.43	0.00 ± 0.00
Z+jets1	2.22 ± 1.33	0.12 ± 0.05
Z+jets2	0.40 ± 0.36	0.15 ± 0.06
Z+jets Total	3.25 ± 1.45	0.27 ± 0.08
Top		
single top	1447.51 ± 8.07	5097.42 ± 14.99
ttbarV-ttbarVV	1.91 ± 0.17	8.40 ± 0.36
Wt	385.42 ± 7.67	547.89 ± 9.23
Top Total	1834.84 ± 11.13	5653.71 ± 17.61
WW		
WW	632.07 ± 4.55	13.03 ± 0.71
Other		
llll	0.85 ± 0.09	0.03 ± 0.02
llqq	0.02 ± 0.02	0.00 ± 0.00
VVV	2.83 ± 0.06	0.19 ± 0.01
W+jets	41.26 ± 16.27	2.95 ± 1.96
Ztt	0.91 ± 1.51	0.24 ± 0.12
WZjj	0.14 ± 0.01	0.02 ± 0.00
lllljj	0.01 ± 0.00	0.00 ± 0.00
llvvjj+WW	50.86 ± 1.40	1.13 ± 0.20
Other Total	96.88 ± 16.40	4.56 ± 1.98
NR / Total (%)	95.26	99.90
Signal / Total (%)	0.02	0.00
DATA / MC	1.12	1.01

Table F.3: Event yields for Data and Monte Carlo Samples in the Zjets Control Regions. The yields are presented as estimated according to each region phase space definition (Table 4.3) without scaling of any kind.

Z+jets CR Event Yields			
Event Type	Zjets0	Zjets1	Zjets2
DATA	2691 ± 51.87	5733 ± 75.72	5769 ± 75.95
Signal ZZ			
llvv	97.69 ± 3.59	143.87 ± 4.00	151.62 ± 2.48
llvvjj	0.04 ± 0.01	1.17 ± 0.03	10.31 ± 0.09
Signal(ZZ)	97.73 ± 3.59	145.04 ± 4.00	161.93 ± 2.48
WZ			
WZ	79.37 ± 2.01	248.24 ± 3.47	308.69 ± 3.03
Z+jets			
Z+jets0	1730.10 ± 99.64	0.00 ± 0.00	0.00 ± 0.00
Z+jets1	0.00 ± 0.00	3798.46 ± 89.31	0.00 ± 0.00
Z+jets2	0.00 ± 0.00	0.00 ± 0.00	4659.24 ± 110.95
Z+jets	1730.1 ± 99.64	3798.46 ± 89.31	4659.24 ± 110.95
Top			
single top	62.87 ± 1.66	152.35 ± 2.61	215.33 ± 3.11
ttbarV-ttbarVV	0.02 ± 0.01	0.08 ± 0.04	0.40 ± 0.07
Wt	28.12 ± 2.10	44.09 ± 2.56	28.35 ± 2.13
Top Total	91.01 ± 2.68	196.53 ± 3.66	244.09 ± 3.77
WW			
WW	47.68 ± 1.22	45.98 ± 1.19	18.39 ± 0.79
Other			
llll	6.22 ± 0.26	23.34 ± 0.49	24.17 ± 0.44
llqq	3.30 ± 0.37	16.93 ± 0.95	17.92 ± 1.04
VVV	0.17 ± 0.01	0.66 ± 0.03	1.37 ± 0.04
W+jets	2.46 ± 2.38	2.97 ± 2.27	4.55 ± 2.30
Ztt	0.00 ± 0.00	-0.09 ± 0.16	0.00 ± 0.00
WZjj	0.19 ± 0.02	2.20 ± 0.05	9.61 ± 0.11
lllljj	0.00 ± 0.00	0.04 ± 0.01	0.26 ± 0.02
llvvjj+WW	0.07 ± 0.05	0.04 ± 0.04	0.00 ± 0.00
Other Total	12.41 ± 2.43	46.10 ± 2.52	57.89 ± 2.57
Z+jets / Total (%)	84.1	84.78	85.49
Signal / Total (%)	4.75	3.24	2.97
DATA / MC	1.31	1.28	1.06

F.2 Signal and Validation Region Event Yields

Table F.4: Event yields in the Signal and Validation Regions. The yields are presented as estimated according to the VR phase space definition (Tables 3.6 & 4.7) without scaling of any kind. The data sample is blinded in the SR.

Signal and Validation Region Event Yields		
Event Type	SR	VR
DATA	-	3517 ± 59.30
Signal ZZ		
llvv	1571.71 ± 13.84	317.13 ± 5.98
llvvjj	5.31 ± 0.07	3.36 ± 0.05
Signal(ZZ)	1577.03 ± 13.84	320.49 ± 5.98
WZ		
WZ	669.20 ± 6.14	299.75 ± 3.81
Z+jets		
Z+jets0	101.90 ± 23.97	954.47 ± 65.46
Z+jets1	31.65 ± 8.69	683.71 ± 44.86
Z+jets2	1.16 ± 0.72	288.51 ± 36.94
Z+jets	134.71 ± 25.51	1926.70 ± 87.54
Top		
single top	67.66 ± 1.73	180.42 ± 2.83
ttbarV-ttbarVV	0.24 ± 0.06	0.13 ± 0.04
Wt	23.04 ± 1.80	53.91 ± 2.82
Top Total	90.94 ± 2.50	234.47 ± 3.99
WW		
WW	37.36 ± 1.08	62.42 ± 1.42
Other		
llll	34.21 ± 0.63	20.48 ± 0.46
llqq	0.90 ± 0.22	7.16 ± 0.57
VVV	5.21 ± 0.06	1.13 ± 0.03
W+jets	1.80 ± 1.94	1.09 ± 2.99
Ztt	0.00 ± 0.00	0.00 ± 0.00
WZjj	2.71 ± 0.06	2.89 ± 0.06
lllljj	0.04 ± 0.01	0.07 ± 0.01
llvvjj+WW	0.07 ± 0.05	0.11 ± 0.06
Other Total	44.94 ± 2.06	32.94 ± 3.08
Signal / Total (%)	61.74	11.14
DATA / MC	1.13	1.22

F.3 Control Region Variants Event Yields

Table F.5: Event yields for Variants of the 3ℓ CR. The Variants have been created by altering the E_T^{miss} and E_T^{miss}/H_T requirements in comparison with the nominal ones. The cuts imposed to produce these event yields are shown in Table 5.1.

Process	3ℓ I	3ℓ II
DATA	1221.00 ± 34.94	948.00 ± 30.79
Signal ZZ		
llvv	1.71 ± 0.50	1.54 ± 0.49
llvvjj	0.05 ± 0.01	0.04 ± 0.01
Signal(ZZ)	1.76 ± 0.50	1.58 ± 0.49
WZ		
WZ	1072.48 ± 7.57	833.04 ± 6.70
Z+jets		
Z+jets0	8.66 ± 4.27	3.86 ± 3.44
Z+jets1	4.79 ± 2.12	2.00 ± 1.47
Z+jets2	1.38 ± 0.58	0.67 ± 0.49
Z+jets Total	14.83 ± 4.80	6.53 ± 3.77
Top		
single top	24.82 ± 1.05	18.31 ± 0.90
ttbarV-ttbarVV	0.23 ± 0.04	0.19 ± 0.04
Wt	3.29 ± 0.68	2.43 ± 0.58
Top Total	28.34 ± 1.26	20.93 ± 1.07
WW		
WW	0.33 ± 0.10	0.26 ± 0.09
Other		
llll	36.64 ± 0.63	27.94 ± 0.55
llqq	0.19 ± 0.07	0.11 ± 0.06
VVV	8.19 ± 0.09	6.93 ± 0.08
W+jets	0.00 ± 0.00	0.00 ± 0.00
Ztt	0.00 ± 0.00	0.00 ± 0.00
WZjj	7.45 ± 0.10	5.41 ± 0.08
lllljj	0.06 ± 0.01	0.04 ± 0.01
llvvjj+WW	0.00 ± 0.00	0.00 ± 0.00
Other Total	52.54 ± 0.65	40.43 ± 0.57
WZ / Total (%)	91.64	92.27
Signal / Total (%)	0.15	0.17
DATA / MC	1.04	1.05

Table F.6: Event yields for Variants of the Non-Resonant CR. The Variants have been created by altering the E_T^{miss} and E_T^{miss}/H_T requirements in comparison with the nominal ones. The cuts imposed to produce these event yields are shown in Table 5.1.

Process	NR I	NR II
DATA	2912.00 ± 53.96	2013.00 ± 44.87
Signal ZZ		
llvv	0.40 ± 0.20	0.46 ± 0.19
llvvjj	0.00 ± 0.00	0.00 ± 0.00
Signal(ZZ)	0.40 ± 0.20	0.46 ± 0.19
WZ		
WZ	11.99 ± 0.79	8.71 ± 0.70
Z+jets		
Z+jets0	0.34 ± 0.32	0.34 ± 0.32
Z+jets1	1.25 ± 0.734	1.22 ± 0.73
Z+jets2	0.02 ± 0.02	0.00 ± 0.00
Z+jets Total	1.61 ± 0.80	1.57 ± 0.80
Top		
single top	1924.55 ± 9.21	1295.56 ± 7.55
ttbarV-ttbarVV	3.63 ± 0.25	3.04 ± 0.22
Wt	381.28 ± 7.68	266.03 ± 6.39
Top Total	2309.45 ± 12.00	1564.63 ± 9.90
WW		
WW	322.98 ± 3.25	226.88 ± 2.72
Other		
llll	0.49 ± 0.07	0.39 ± 0.06
llqq	0.00 ± 0.00	0.00 ± 0.00
VVV	2.15 ± 0.06	1.88 ± 0.05
W+jets	17.66 ± 10.45	16.37 ± 9.54
Ztt	0.00 ± 0.00	0.00 ± 0.00
WZjj	0.05 ± 0.01	0.04 ± 0.01
lllljj	0.00 ± 0.00	0.00 ± 0.00
llvvjj+WW	28.90 ± 1.04	20.79 ± 0.88
Other Total	49.27 ± 10.50	39.47 ± 9.58
NR / Total (%)	97.65	97.27
Signal / Total (%)	0.01	0.02
DATA / MC	1.08	1.09

Table F.7: Event yields for Variants of the Zjets CR. The Variants have been created by altering the E_T^{miss} and E_T^{miss}/H_T requirements in comparison with the nominal ones. The cuts imposed to produce these event yields are shown in Table 5.1.

Process	Zjets I	Zjets II
DATA	8421.00 ± 91.77	16656 ± 129.06
Signal ZZ		
llvv	322.88 ± 5.25	593.831 ± 7.51
llvvjj	11.01 ± 0.10	14.81 ± 0.11
Signal(ZZ)	333.89 ± 5.25	608.64 ± 7.51
WZ		
WZ	494.00 ± 4.22	857.07 ± 5.93
Z+jets		
Z+jets0	1709.36 ± 98.73	2292.28 ± 114.17
Z+jets1	1329.63 ± 48.18	4363.31 ± 96.47
Z+jets2	2881.53 ± 73.85	4947.34 ± 116.90
Z+jets Total	5920.52 ± 132.38	11602.90 ± 189.76
Top		
single top	280.93 ± 3.55	559.55 ± 5.00
ttbarV-ttbarVV	0.45 ± 0.08	0.63 ± 0.09
Wt	66.66 ± 3.18	135.77 ± 4.55
Top Total	348.04 ± 4.77	695.95 ± 6.76
WW		
WW	78.91 ± 1.58	150.11 ± 2.18
Other		
llll	40.36 ± 0.61	69.22 ± 0.81
llqq	22.71 ± 1.17	43.55 ± 1.54
VVV	1.93 ± 0.46	3.10 ± 0.05
W+jets	7.08 ± 3.31	13.26 ± 4.48
Ztt	0.00 ± 0.00	-0.09 ± 0.17
WZjj	10.81 ± 0.11	14.69 ± 0.13
lllljj	0.28 ± 0.24	0.38 ± 0.0
llvvjj+WW	0.07 ± 0.05	0.15 ± 0.07
Other Total	83.24 ± 3.57	144.26 ± 4.81
Z+jets / Total (%)	81.56	82.53
Signal / Total (%)	4.60	4.33
DATA / MC	1.16	1.18

G Regions as a function of kinematic variables

G.1 Additional Plots on Control Regions

Figure G.1: Non-Resonant and 3ℓ CR plots before and after scaling as a function of the missing transverse energy E_T^{miss} . The pre-fit plots correspond to the event yields presented in Tables F.1 and F.2. In the post-fit plots, the background contributions in every bin are scaled according to the best-fit scaling factors obtained by the MLE simultaneous fit (Table 4.6). The uncertainties are statistical. A fit function is used in order to evaluate the scaling performance (red line).

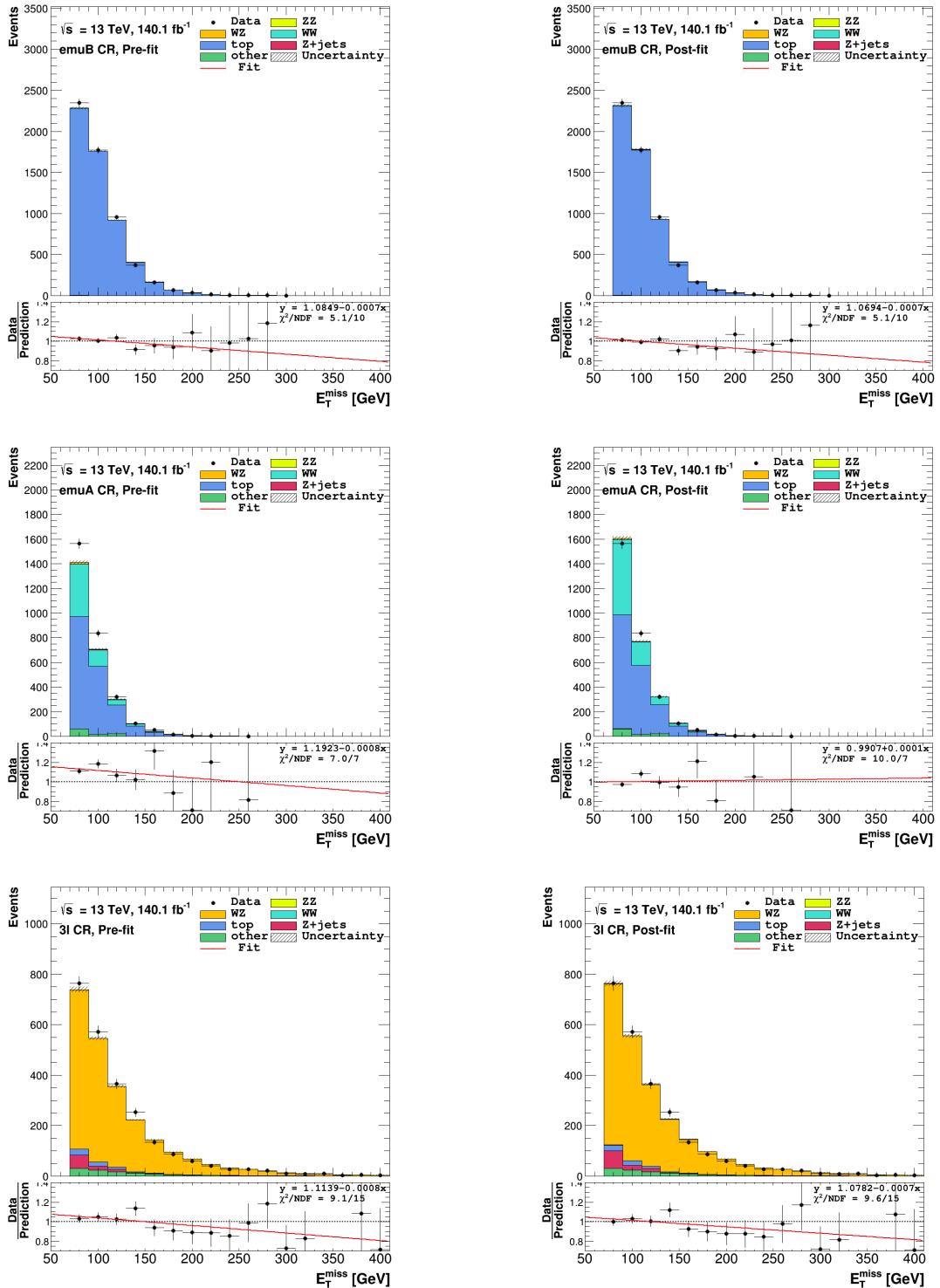


Figure G.2: Z+jets CR plots before and after scaling as a function of the missing transverse energy E_T^{miss} . The pre-fit plots correspond to the event yields presented in Tables F.3. In the post-fit plots, the background contributions in every bin are scaled according to the best-fit scaling factors obtained by the MLE simultaneous fit (Table 4.6). The uncertainties are statistical. A fit function is used in order to evaluate the scaling performance (red line).

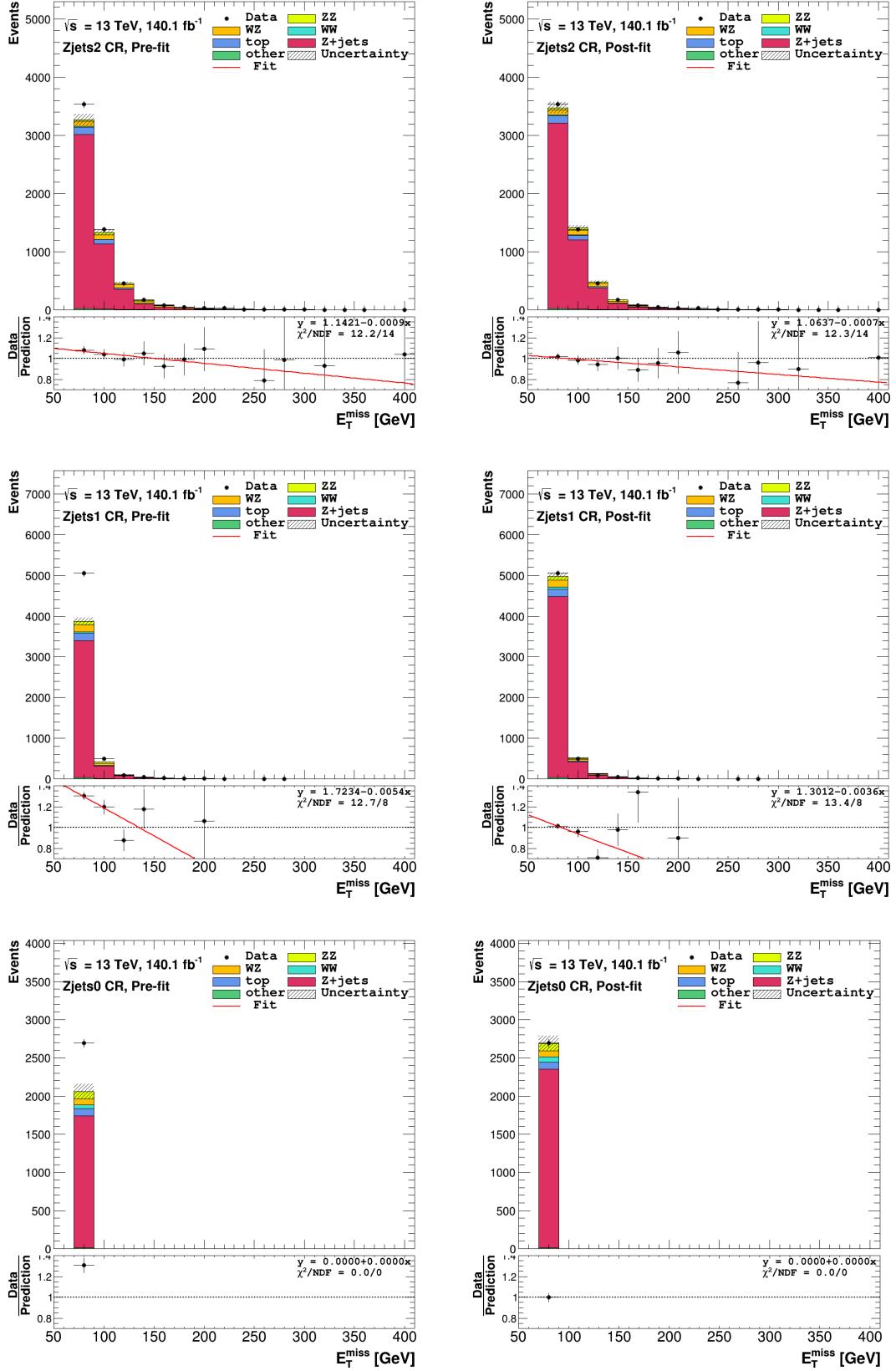


Figure G.3: Non-Resonant and 3ℓ CR plots before and after scaling as a function of E_T^{miss}/H_T . The pre-fit plots correspond to the event yields presented in Tables F.1 and F.2. In the post-fit plots, the background contributions in every bin are scaled according to the best-fit scaling factors obtained by the MLE simultaneous fit (Table 4.6). The uncertainties are statistical. A fit function is used in order to evaluate the scaling performance (red line).

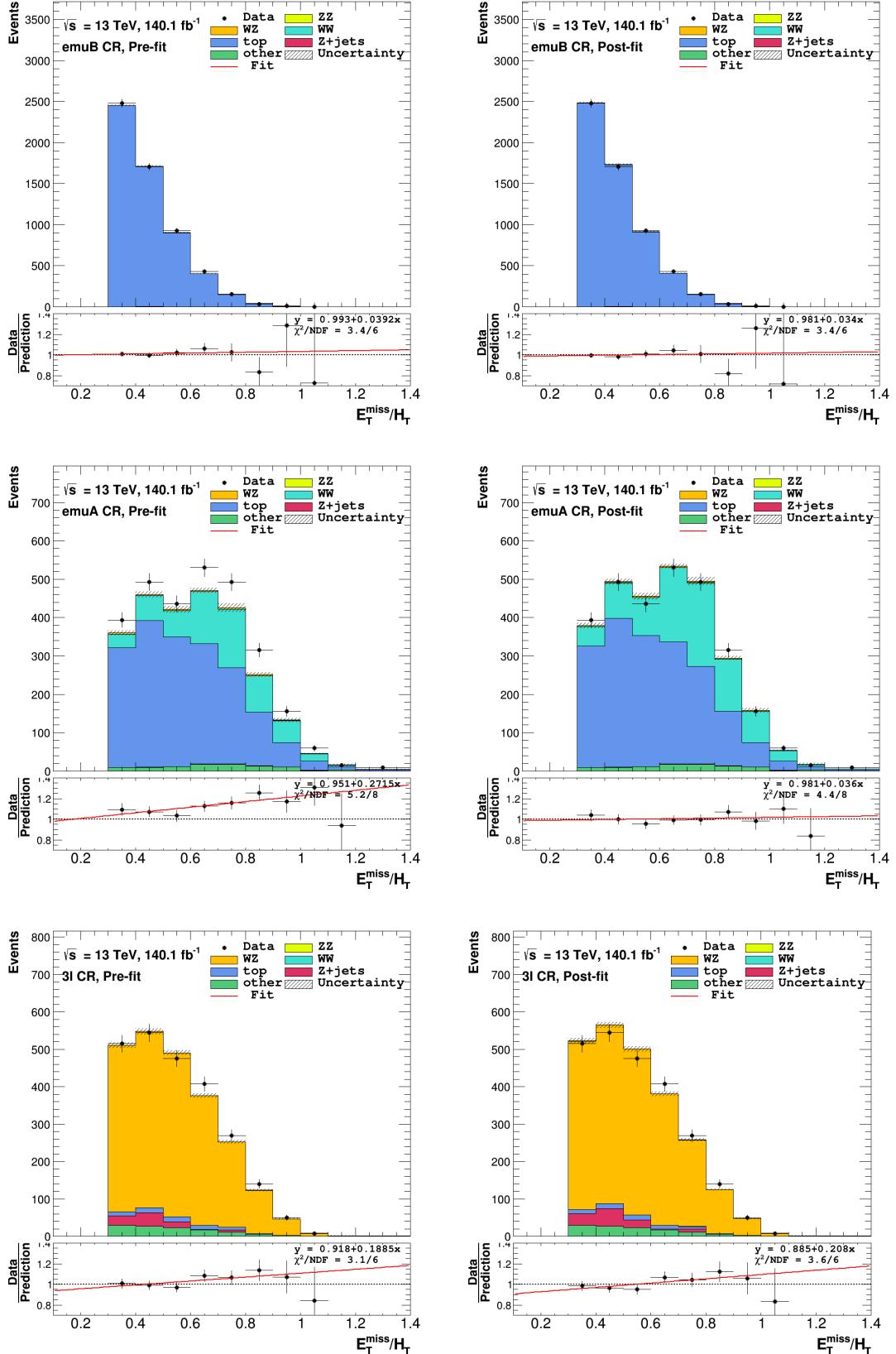


Figure G.4: Z+jets CR plots before and after scaling as a function of E_T^{miss}/H_T . The pre-fit plots correspond to the event yields presented in Tables F.3. In the post-fit plots, the background contributions in every bin are scaled according to the best-fit scaling factors obtained by the MLE simultaneous fit (Table 4.6). The uncertainties are statistical. A fit function is used in order to evaluate the scaling performance (red line).

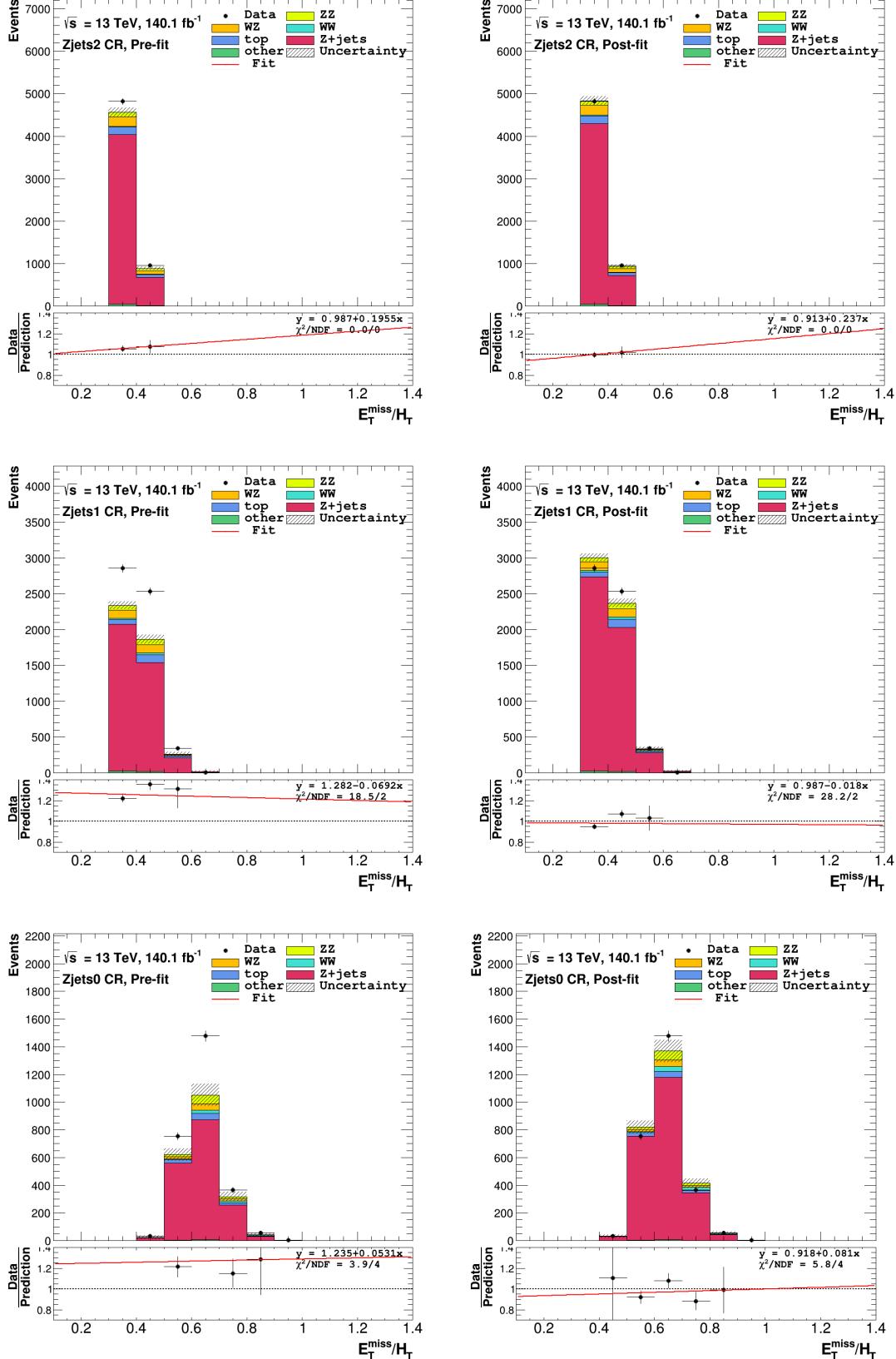


Figure G.5: Non-Resonant and 3ℓ CR plots before and after scaling as a function of $\Delta\phi(\vec{p}_T^Z, \vec{E}_T^{miss})$. The pre-fit plots correspond to the event yields presented in Tables F.1 and F.2. In the post-fit plots, the background contributions in every bin are scaled according to the best-fit scaling factors obtained by the MLE simultaneous fit (Table 4.6). The uncertainties are statistical. A fit function is used in order to evaluate the scaling performance (red line).

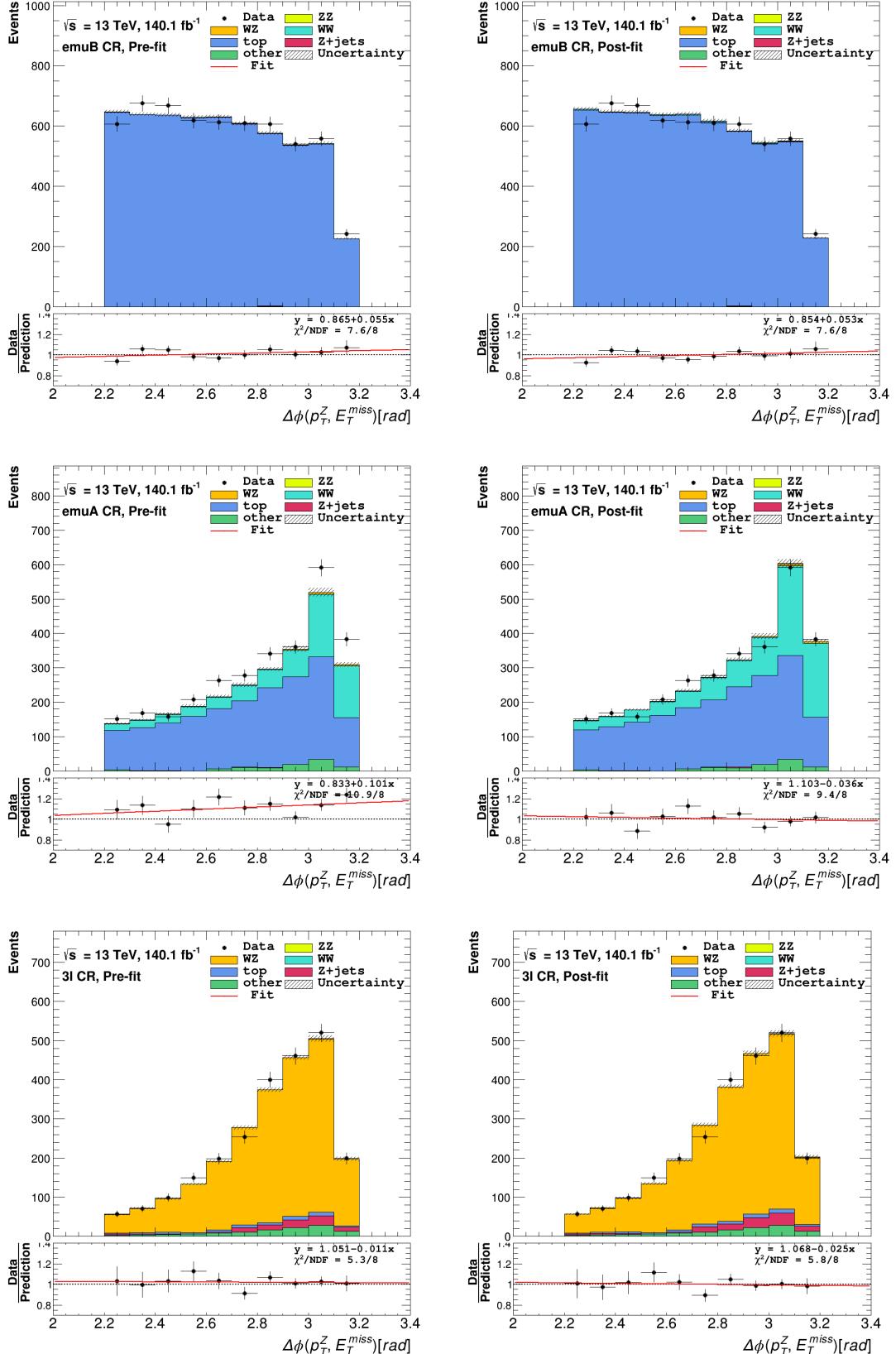


Figure G.6: Z+jets CR plots before and after scaling as a function of $\Delta\phi(p_T^Z, \vec{E}_T^{\text{miss}})$. The pre-fit plots correspond to the event yields presented in Tables F.3. In the post-fit plots, the background contributions in every bin are scaled according to the best-fit scaling factors obtained by the MLE simultaneous fit (Table 4.6). The uncertainties are statistical. A fit function is used in order to evaluate the scaling performance (red line).

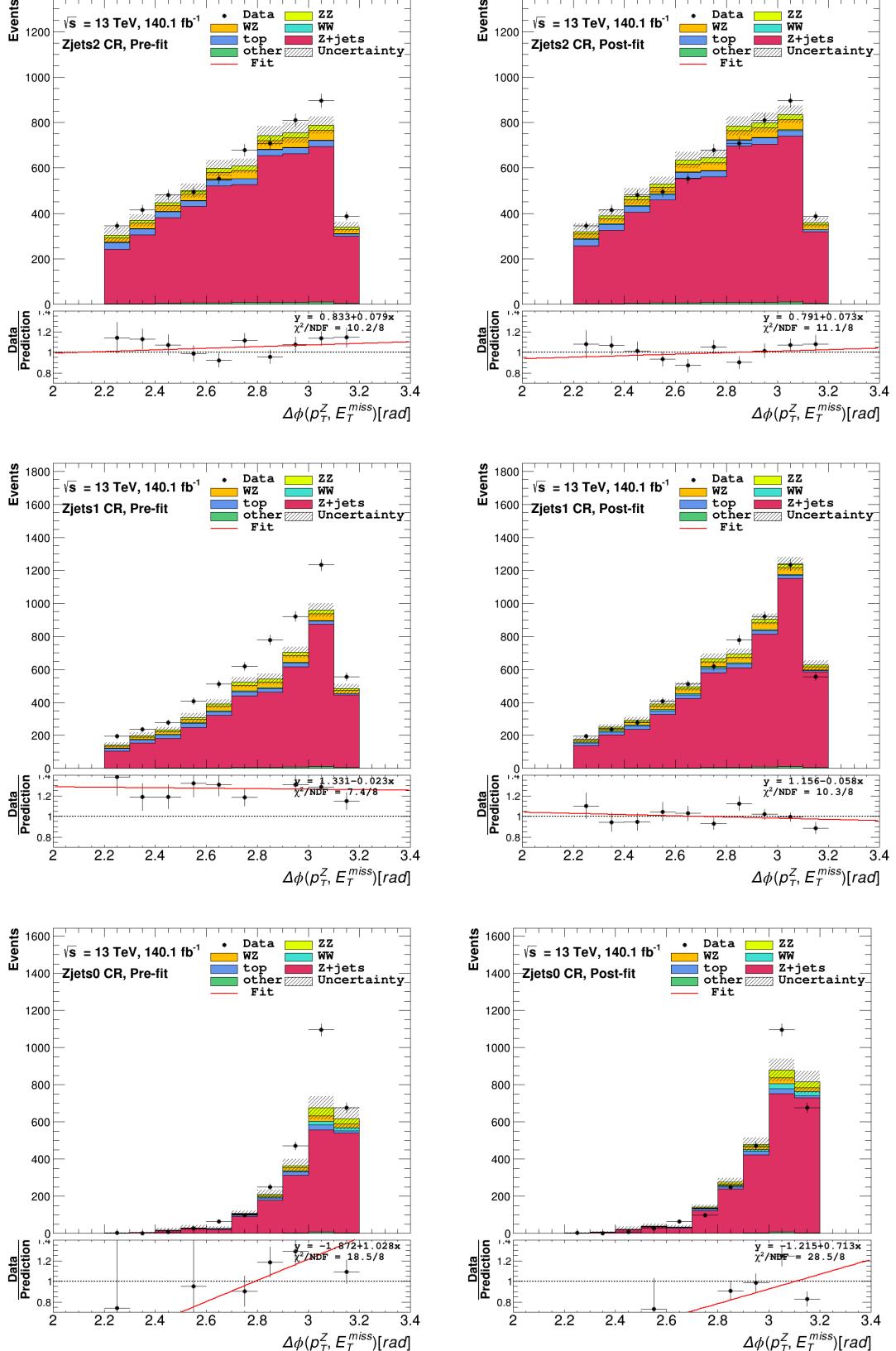


Figure G.7: Non-Resonant and 3ℓ CR plots before and after scaling as a function of $\Delta R_{\ell\ell}$. The pre-fit plots correspond to the event yields presented in Tables F.1 and F.2. In the post-fit plots, the background contributions in every bin are scaled according to the best-fit scaling factors obtained by the MLE simultaneous fit (Table 4.6). The uncertainties are statistical. A fit function is used in order to evaluate the scaling performance (red line).

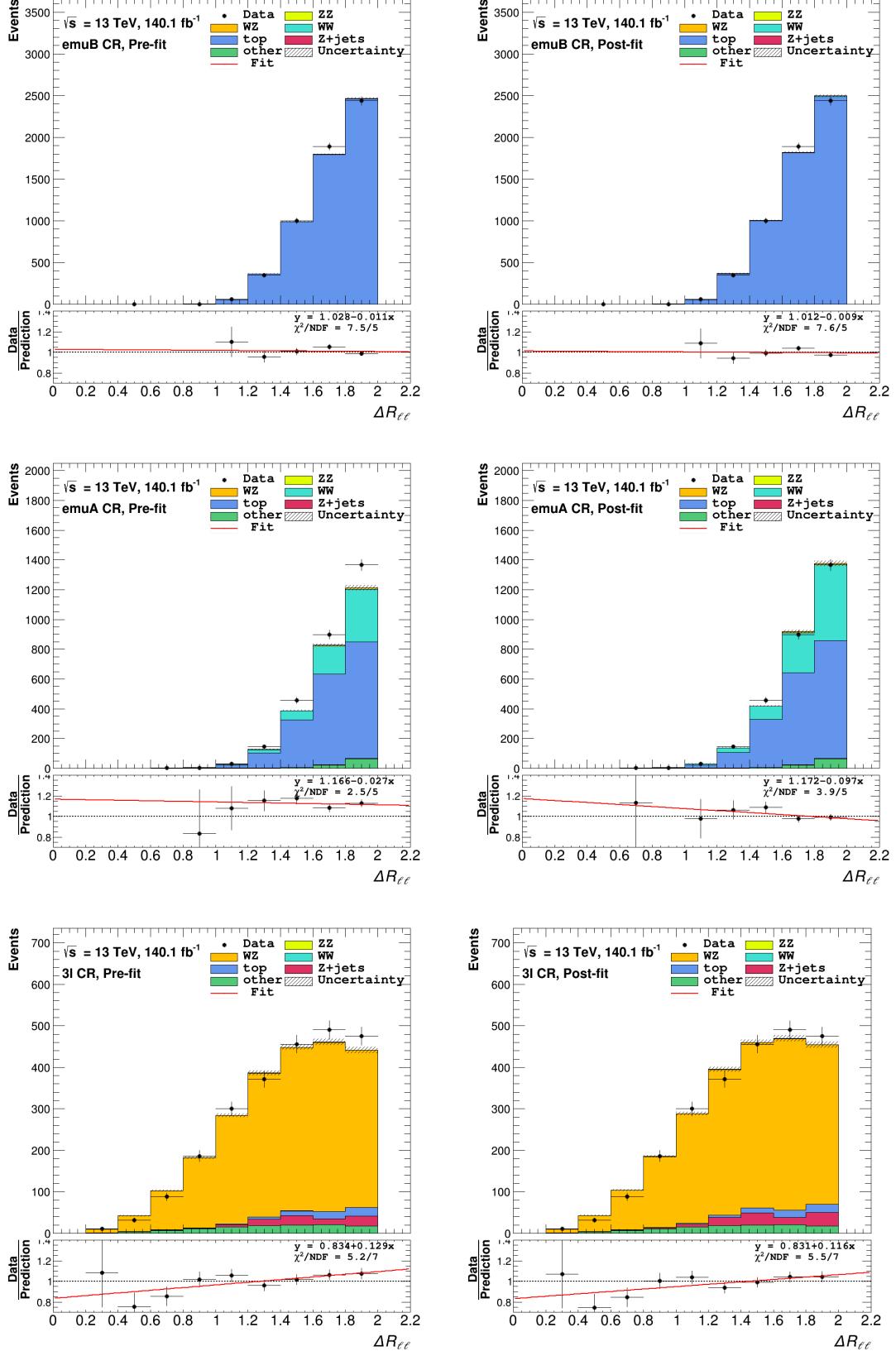


Figure G.8: Z+jets CR plots before and after scaling as a function of $\Delta R_{\ell\ell}$. The pre-fit plots correspond to the event yields presented in Tables F.3. In the post-fit plots, the background contributions in every bin are scaled according to the best-fit scaling factors obtained by the MLE simultaneous fit (Table 4.6). The uncertainties are statistical. A fit function is used in order to evaluate the scaling performance (red line).

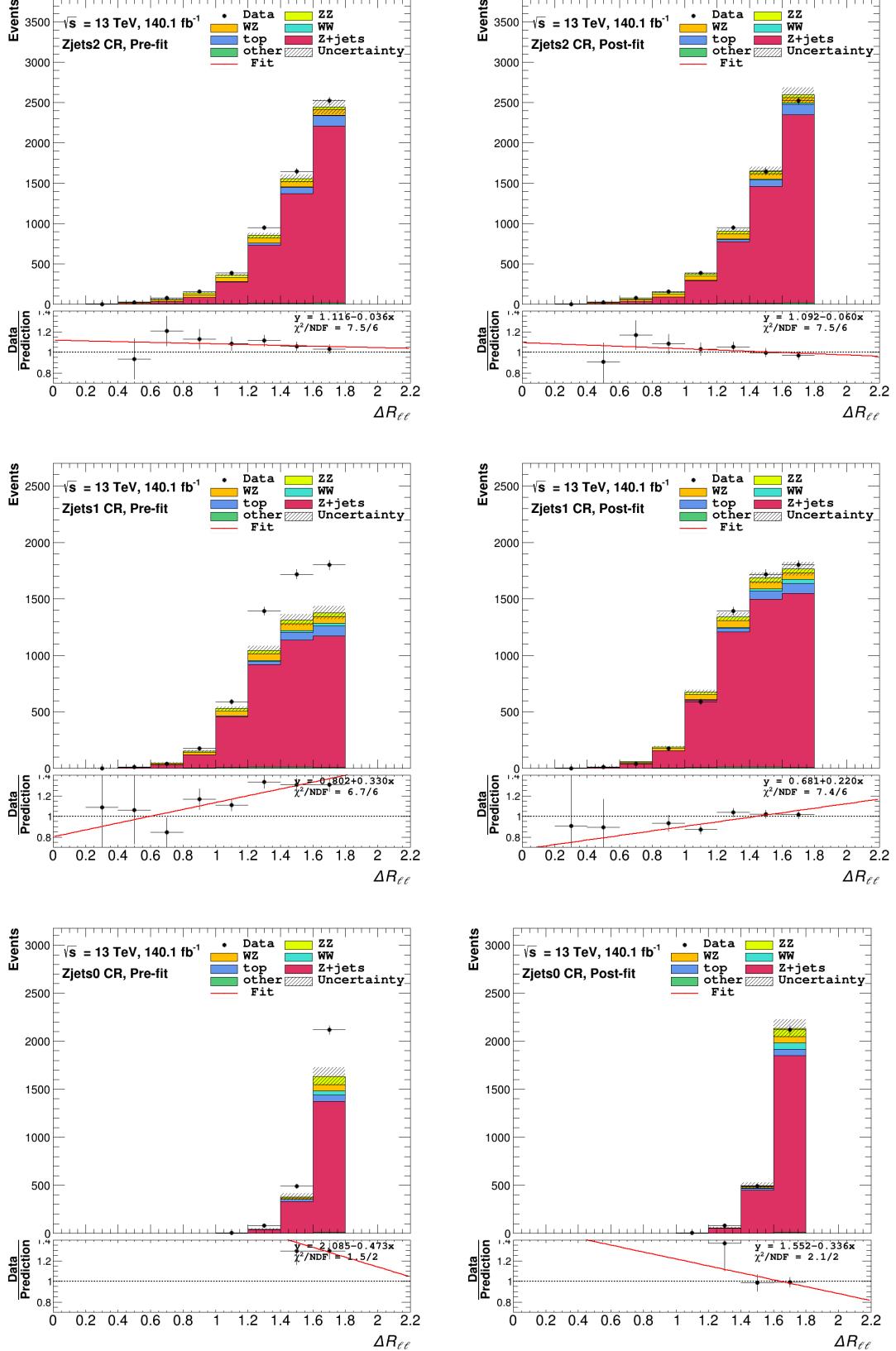


Figure G.9: Non-Resonant and 3ℓ CR plots before and after scaling as a function of p_T^Z . The pre-fit plots correspond to the event yields presented in Tables F.1 and F.2. In the post-fit plots, the background contributions in every bin are scaled according to the best-fit scaling factors obtained by the MLE simultaneous fit (Table 4.6). The uncertainties are statistical. A fit function is used in order to evaluate the scaling performance (red line).

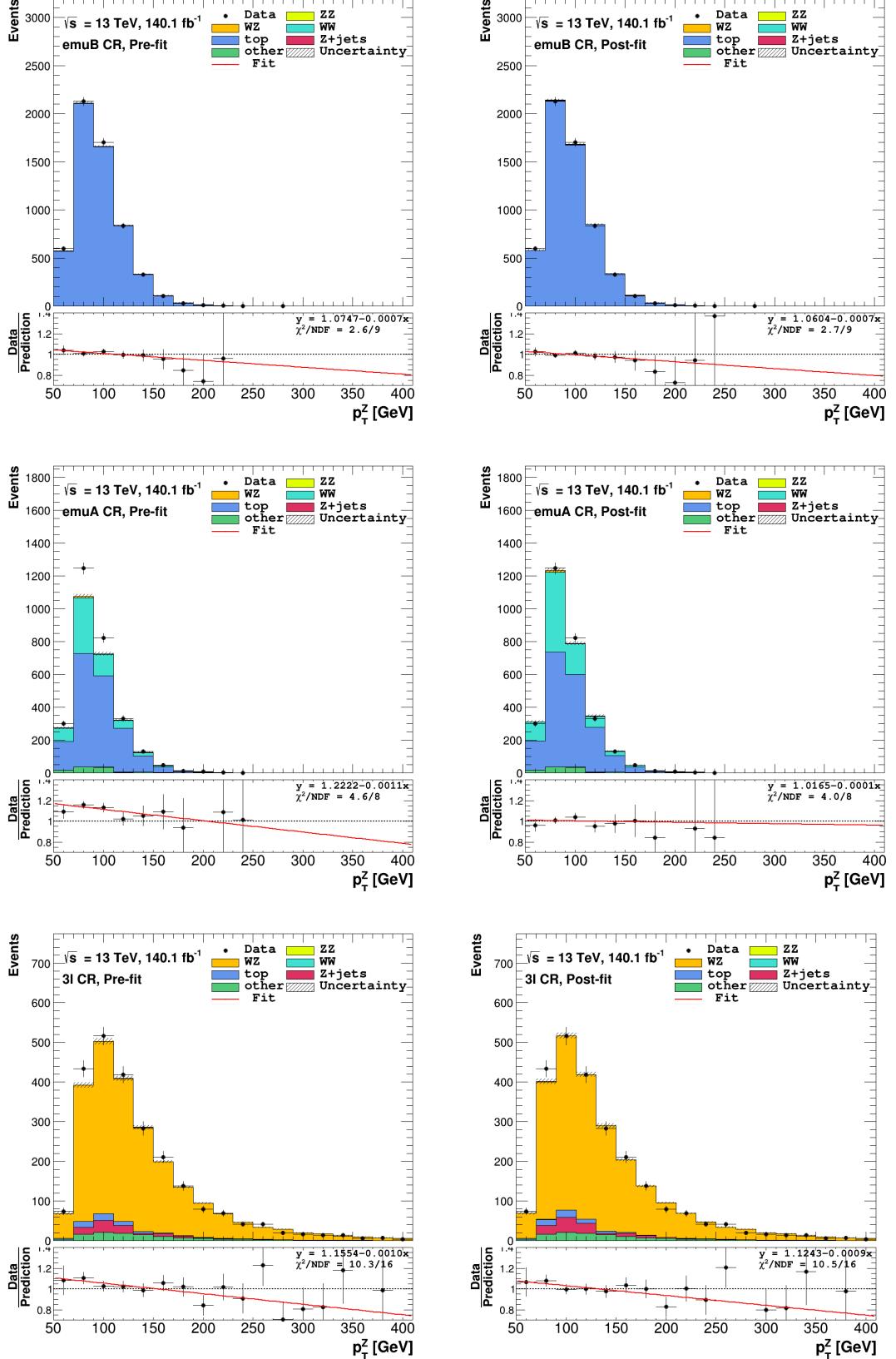


Figure G.10: Z+jets CR plots before and after scaling as a function of p_T^Z . The pre-fit plots correspond to the event yields presented in Tables F.3. In the post-fit plots, the background contributions in every bin are scaled according to the best-fit scaling factors obtained by the MLE simultaneous fit (Table 4.6). The uncertainties are statistical. A fit function is used in order to evaluate the scaling performance (red line).

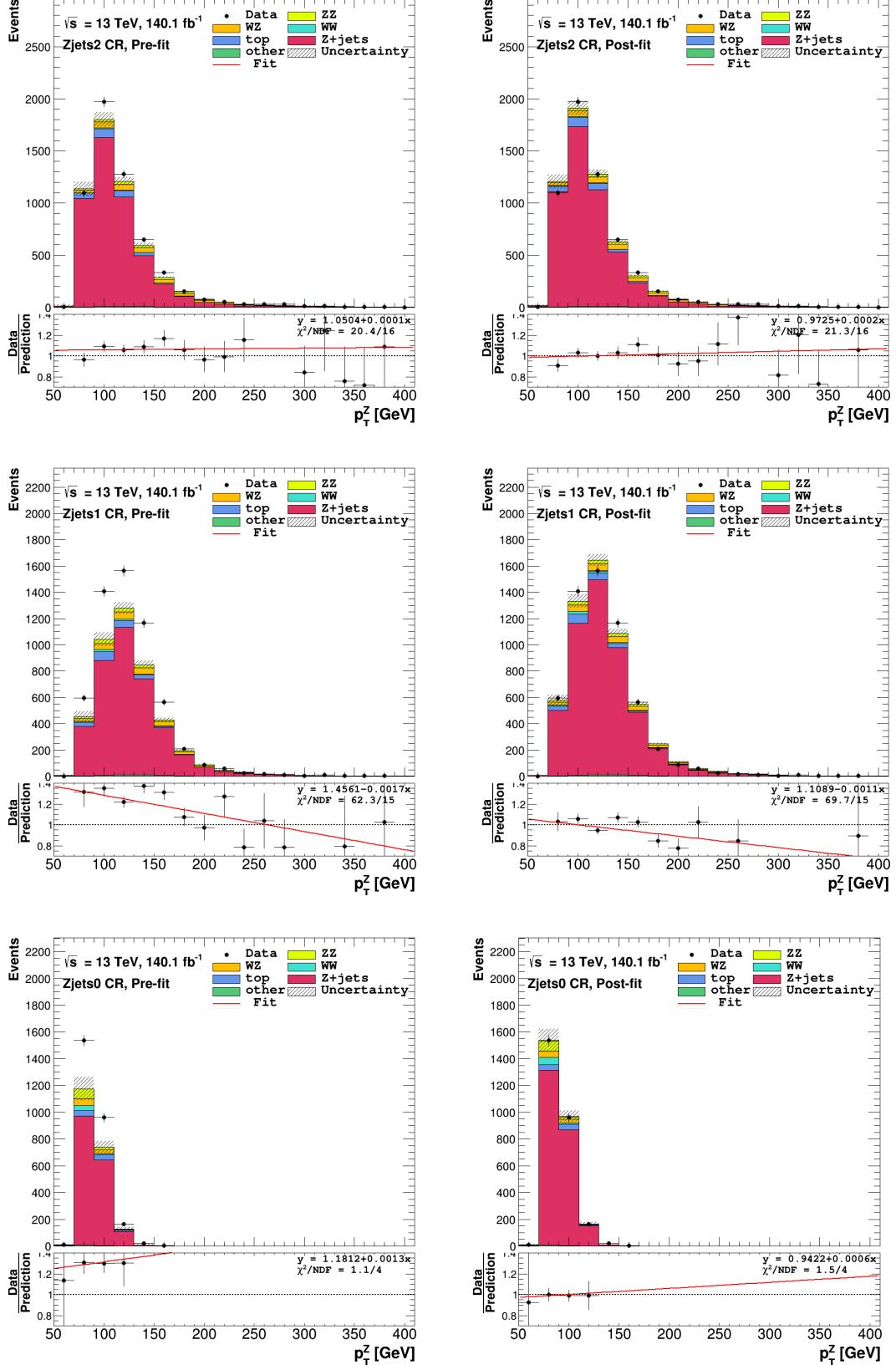


Figure G.11: Non-Resonant and 3ℓ CR plots before and after scaling as a function of n_{jets} . The pre-fit plots correspond to the event yields presented in Tables F.1 and F.2. In the post-fit plots, the background contributions in every bin are scaled according to the best-fit scaling factors obtained by the MLE simultaneous fit (Table 4.6). The uncertainties are statistical. A fit function is used in order to evaluate the scaling performance (red line).

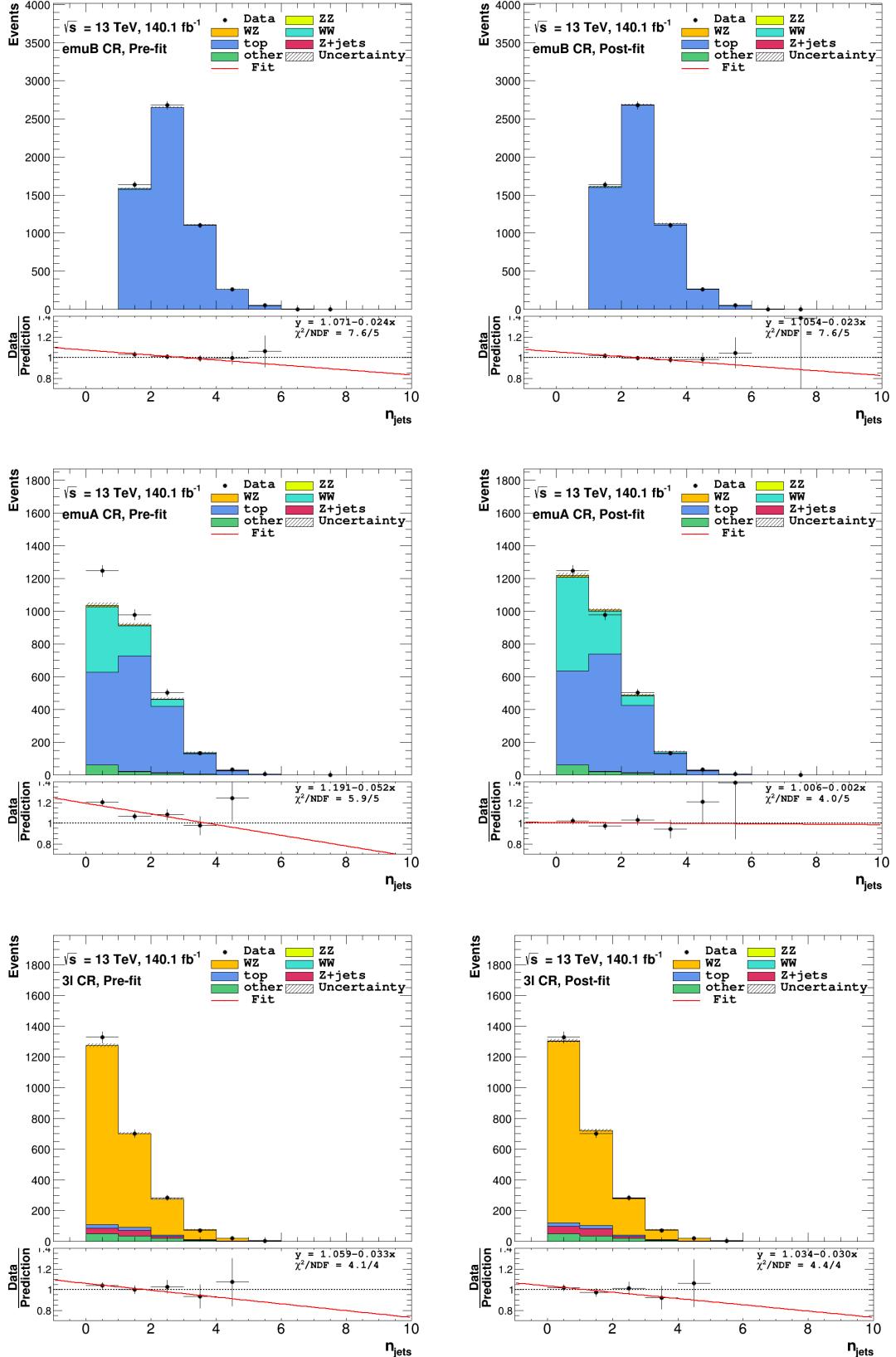


Figure G.12: Z+jets CR plots before and after scaling as a function of n_{jets} . The pre-fit plots correspond to the event yields presented in Tables F.3. In the post-fit plots, the background contributions in every bin are scaled according to the best-fit scaling factors obtained by the MLE simultaneous fit (Table 4.6). The uncertainties are statistical. A fit function is used in order to evaluate the scaling performance (red line).

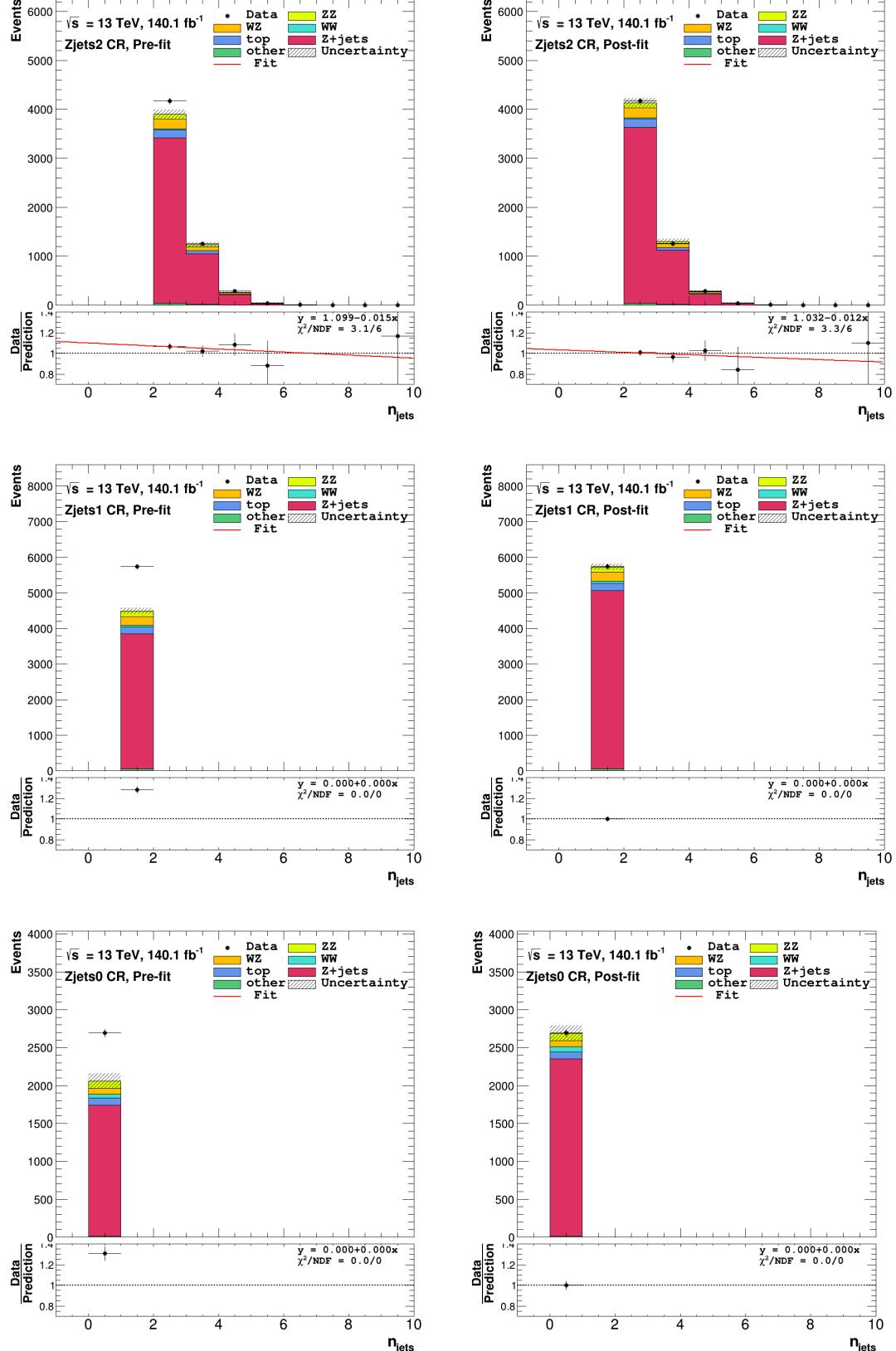


Figure G.13: Non-Resonant and 3ℓ CR plots before and after scaling as a function of missing transverse energy significance $\sigma(E_T^{miss})$. The pre-fit plots correspond to the event yields presented in Tables F.1 and F.2. In the post-fit plots, the background contributions in every bin are scaled according to the best-fit scaling factors obtained by the MLE simultaneous fit (Table 4.6). The uncertainties are statistical. A fit function is used in order to evaluate the scaling performance (red line).

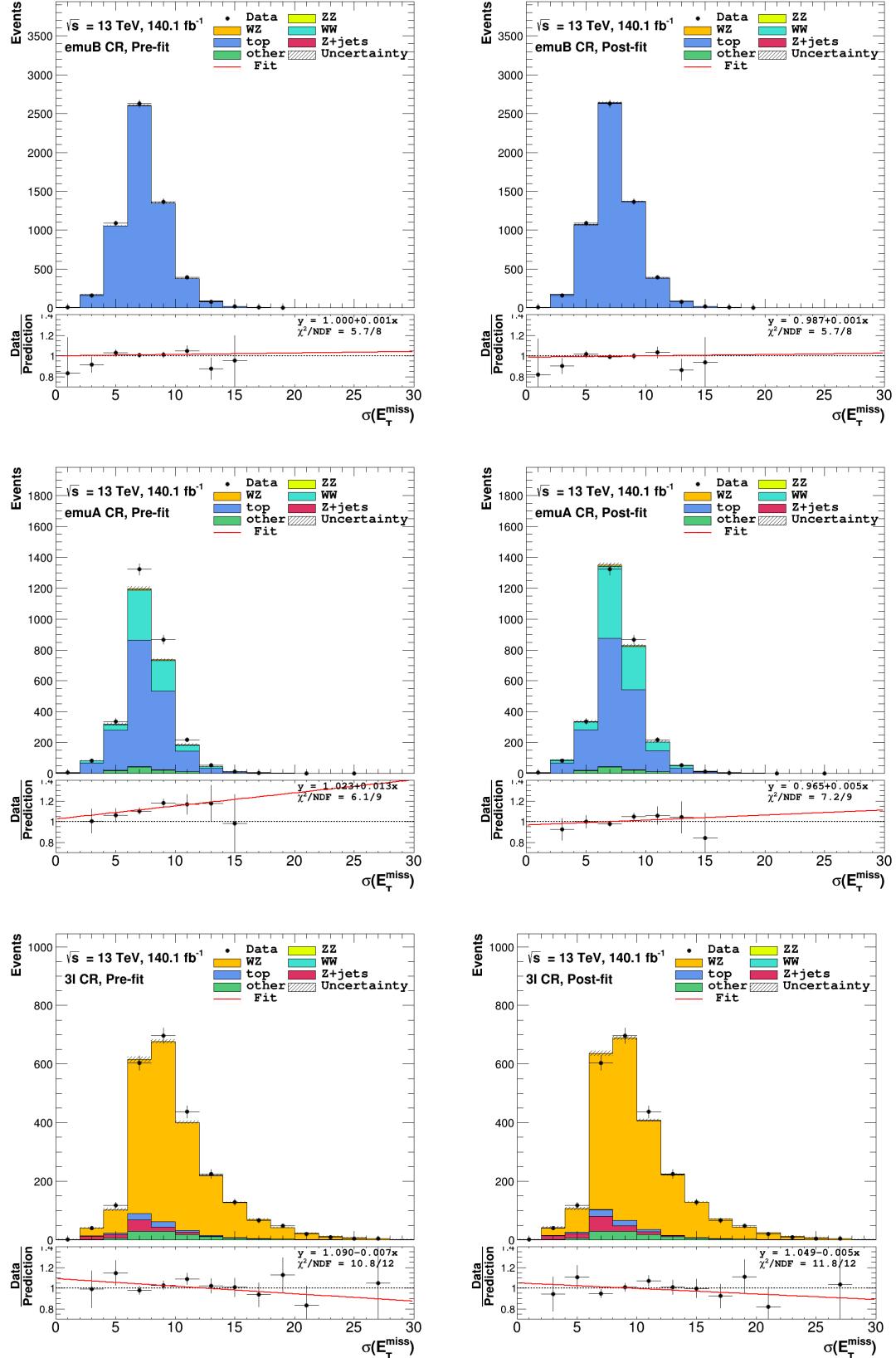
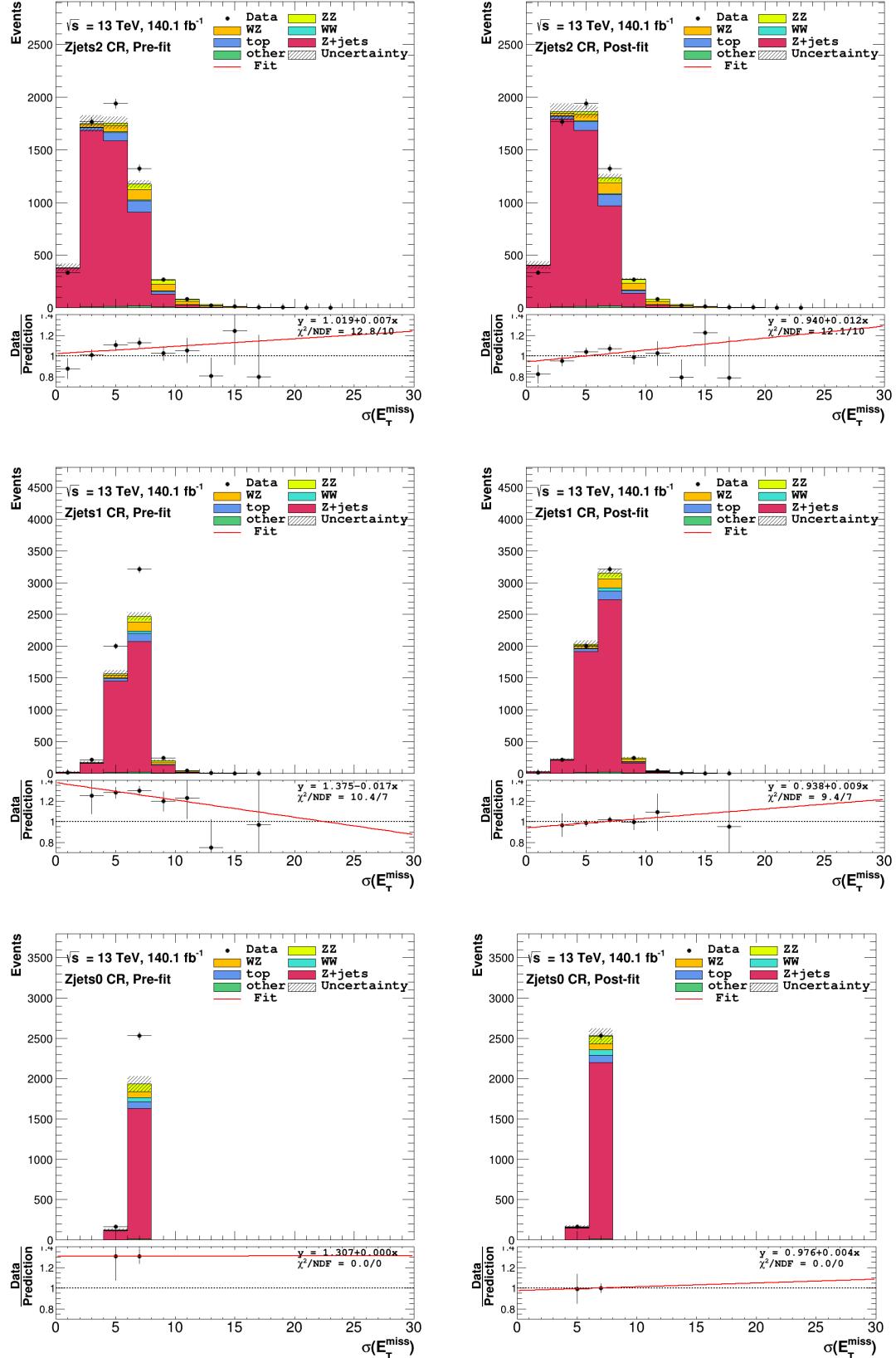
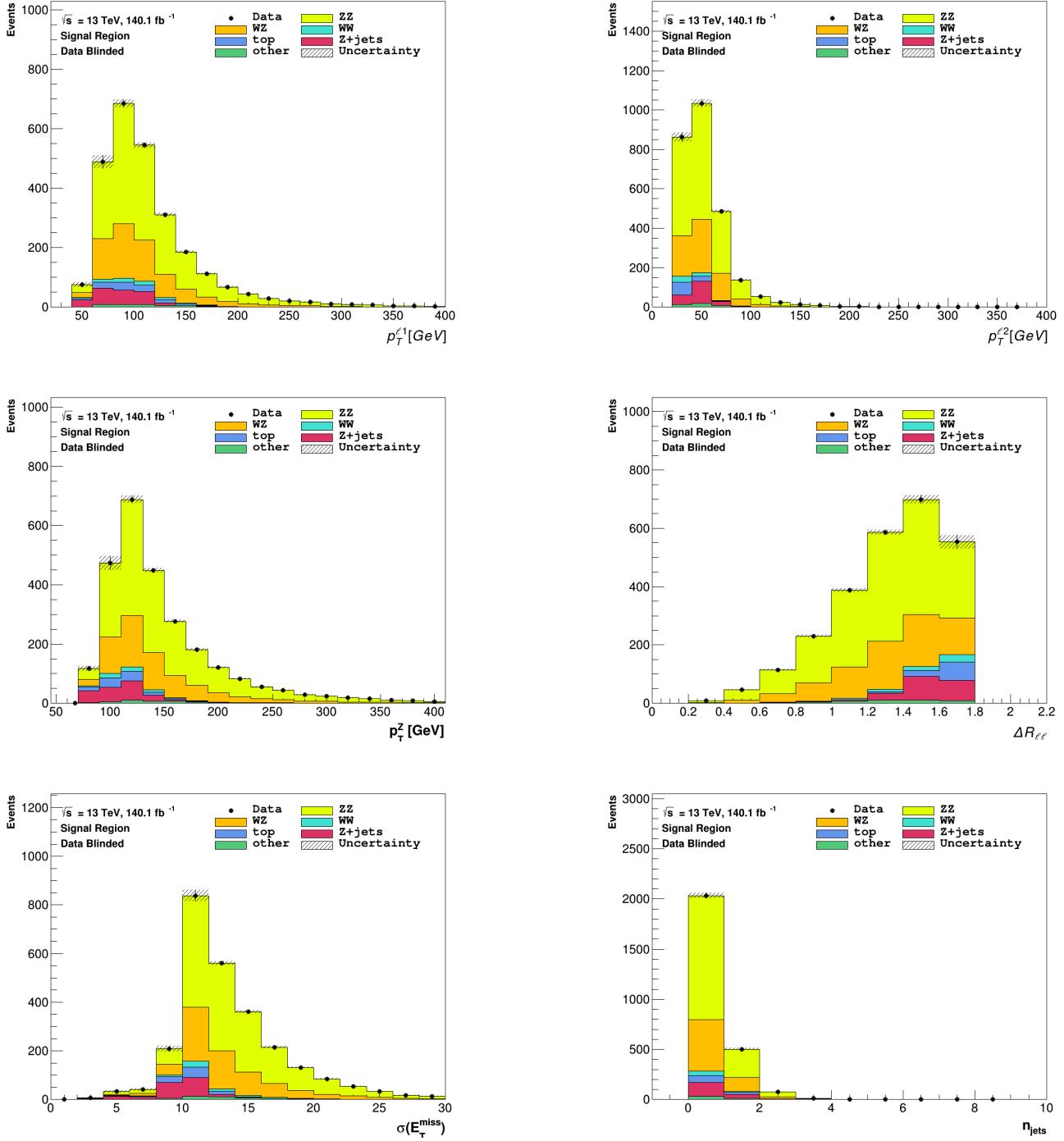


Figure G.14: Z+jets CR plots before and after scaling as a function of missing transverse energy significance $\sigma(E_T^{miss})$. The pre-fit plots correspond to the event yields presented in Tables F.3. In the post-fit plots, the background contributions in every bin are scaled according to the best-fit scaling factors obtained by the MLE simultaneous fit (Table 4.6). The uncertainties are statistical. A fit function is used in order to evaluate the scaling performance (red line).



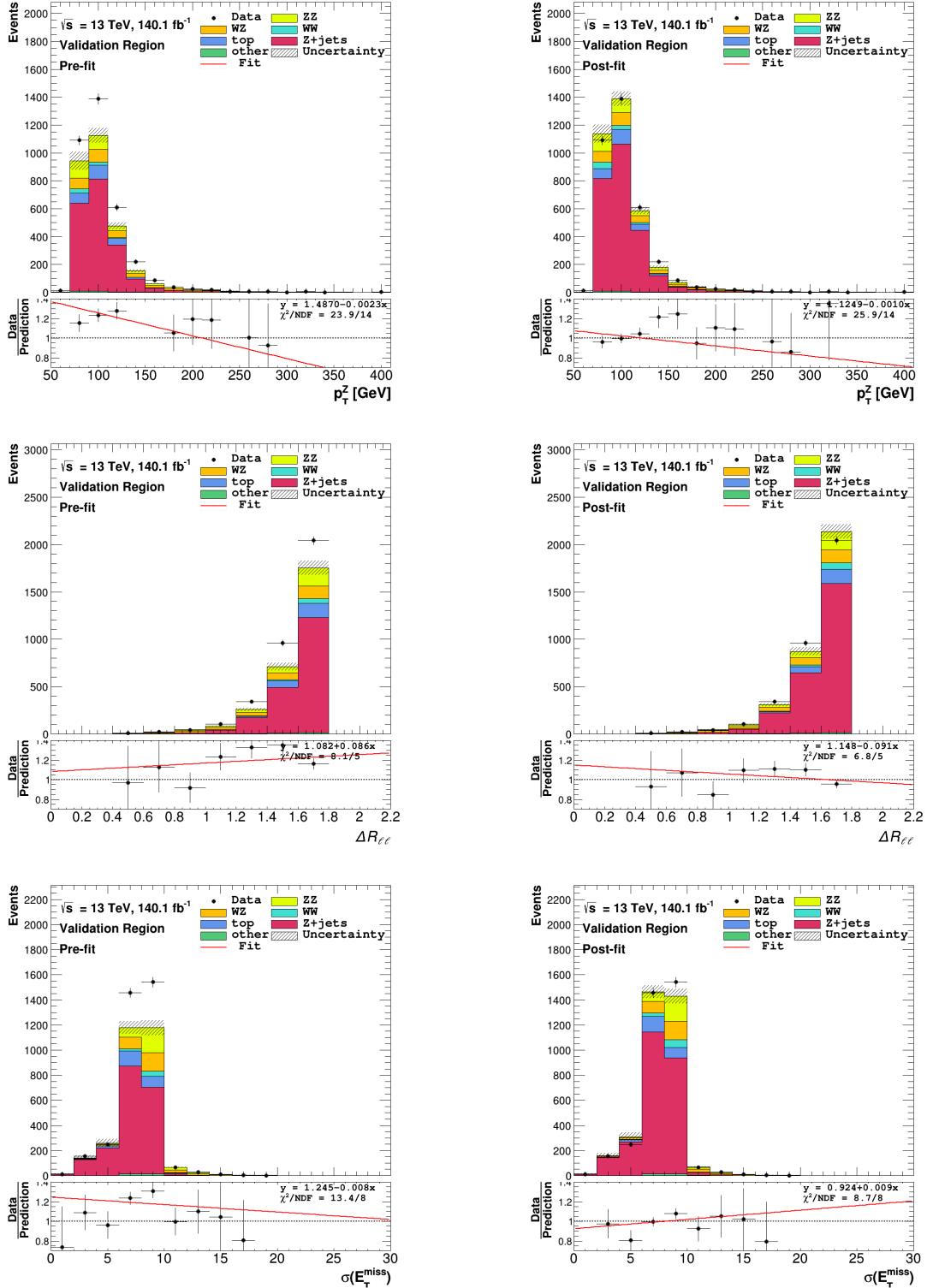
G.2 Additional Plots on Signal Region

Figure G.15: Plots of the optimal Signal Region defined in section 3.4, as a function of $p_T^{\ell 1}$, $p_T^{\ell 2}$, p_T^Z , $\Delta R_{\ell\ell}$, $\sigma(E_T^{miss})$ and n_{jets} . The plots provide a visual representation of the event yields of the desired signal process $ZZ \rightarrow \ell\ell\nu\nu$ for various kinematic variables. Since the total scaled MC events are used as pseudo-data due to the real data sample being blinded in the SR, the simulation events predictions are identical to the corresponding data points. The event yields have been scaled according to the simultaneous fit results shown in Table 4.6.



G.3 Additional Plots on Validation Region

Figure G.16: Pre-fit and Post-fit plots of Validation Region versus p_T^Z , $\Delta R_{\ell\ell}$ and $\sigma(E_T^{miss})$. The pre-fit plots correspond to the event yields shown in Table F.4. In the post-fit plots, every bin has been scaled using the best-fit values obtained from the maximum likelihood analysis results, shown in Table 4.6. The scaled bins are in good agreement with the data. The uncertainties are statistical. A fit function is used in order to evaluate the scaling performance (red line).



References

- [1] ATLAS Collaboration. “Measurement of the ZZ production cross section in proton-proton collisions at $\sqrt{s} = 8$ TeV using the $ZZ \rightarrow \ell^-\ell^+\ell^-\ell^+$ and $ZZ \rightarrow \ell^-\ell^+\nu\bar{\nu}$ decay channels with the ATLAS detector”. In: *Journal of High Energy Physics* 2017.1 (Jan. 2017). URL: <https://doi.org/10.1007%2Fjhep01%282017%29099>.
- [2] ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *Journal of Instrumentation* 3.08 (Aug. 2008). URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08003>.
- [3] Giles Strong. “On the impact of selected modern deep-learning techniques to the performance and celerity of classification models in an experimental high-energy physics use case”. In: *Machine Learning: Science and Technology* 1 (Sept. 2020). DOI: <10.1088/2632-2153/ab983a>.
- [4] ATLAS Collaboration. “Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV”. In: *The European Physical Journal C* 76.5 (May 2016). URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-016-4120-y>.
- [5] ATLAS Collaboration. “Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data”. In: *The European Physical Journal C* 72.3 (Mar. 2012). URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-012-1909-1>.
- [6] ATLAS Collaboration. “Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Physical Review D* 96.7 (Oct. 2017). URL: <https://doi.org/10.1103%2Fphysrevd.96.072002>.
- [7] ATLAS Collaboration. *Selection of jets produced in 13 TeV proton-proton collisions with the ATLAS detector*. Tech. rep. CERN, 2015. URL: <https://cds.cern.ch/record/2037702>.
- [8] ATLAS Collaboration. “Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *The European Physical Journal C* 78.11 (Nov. 2018). URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-018-6288-9>.
- [9] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. In: *The European Physical Journal C* 71.2 (Feb. 2011). URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-011-1554-0>.
- [10] J. Kennedy and R. Eberhart. “Particle swarm optimization”. In: *Proceedings of ICNN’95 - International Conference on Neural Networks*. Vol. 4. 1995, 1942–1948 vol.4. DOI: <10.1109/ICNN.1995.488968>.
- [11] ATLAS Collaboration. *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*. 2022. arXiv: [2212.09379 \[hep-ex\]](2212.09379 [hep-ex]).
- [12] Olaf Behnke and Lorenzo Moneta. “Parameter Estimation”. In: *Data Analysis in High Energy Physics*. John Wiley & Sons, Ltd, 2013. Chap. 2, pp. 27–73. ISBN: 9783527653416. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527653416.ch2>.
- [13] A. Hoecker et al. *TMVA - Toolkit for Multivariate Data Analysis*. 2009. arXiv: [physics/0703039 \[physics.data-an\]](physics/0703039 [physics.data-an]).
- [14] Maurice Clerc. “On a Table Corner”. In: *Particle Swarm Optimization*. John Wiley & Sons, Ltd, 2006. Chap. 2, pp. 29–36. ISBN: 9780470612163. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470612163.ch2>.
- [15] Shigang Wang, Fangfang Zhou, and Fengjuan Wang. “Effect of Inertia Weight w on PSO-SA Algorithm”. In: *International Journal of Online and Biomedical Engineering (iJOE)* 9.S6 (June 2013), pp. 87–91. URL: <https://online-journals.org/index.php/i-joe/article/view/2923>.
- [16] Maurice Clerc. “First Formulations”. In: *Particle Swarm Optimization*. John Wiley & Sons, Ltd, 2006. Chap. 3, pp. 37–50. ISBN: 9780470612163. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470612163.ch3>.
- [17] Fakoudis Dionysios. *Determination of Signal and Background contributions in the process $pp \rightarrow ZZ \rightarrow 2 \text{ leptons} + 2 \text{ neutrinos}$ in the ATLAS experiment*. 2021. URL: <https://ikee.lib.auth.gr/record/335745>.

-
- [18] ATLAS Collaboration. “Measurement of ZZ production in pp collisions at $\sqrt{s} = 7$ TeV and limits on anomalous ZZZ and ZZ γ couplings with the ATLAS detector”. In: *Journal of High Energy Physics* 2013 (Mar. 2013). URL: <https://doi.org/10.1007/2Fjhep03%282013%29128>.
 - [19] ATLAS Collaboration. “Measurement of ZZ production in the $\ell\ell\nu\nu$ final state with the ATLAS detector in pp collisions at $\sqrt{s} = 13$ TeV”. In: *Journal of High Energy Physics* 2019.10 (Oct. 2019). URL: <https://doi.org/10.1007/2Fjhep10%282019%29127>.
 - [20] ATLAS Collaboration. “Search for large missing transverse momentum in association with one top-quark in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Journal of High Energy Physics* 2019.5 (May 2019). URL: <https://doi.org/10.1007/2Fjhep05%282019%29041>.
 - [21] ATLAS Collaboration. *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*. Tech. rep. Geneva: CERN, 2019. URL: <https://cds.cern.ch/record/2677054>.
 - [22] CMS Collaboration. “Measurements of the ZZ production cross sections in the $2\ell 2\nu$ channel in proton-proton collisions at $\sqrt{s} = 7$ TeV and 8 TeV and combined constraints on triple gauge couplings”. In: *The European Physical Journal C* 75.10 (Oct. 2015). URL: <https://doi.org/10.1140/2Fepjc%2Fs10052-015-3706-0>.
 - [23] Particle Data Group. “Review of Particle Physics”. In: *Progress of Theoretical and Experimental Physics* 2022.8 (Aug. 2022), p. 083C01. ISSN: 2050-3911. URL: <https://doi.org/10.1093/ptep/ptac097>.
 - [24] ATLAS Collaboration. *Measurement of the ZZ production cross section in the llvv channel in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*. Tech. rep. Mar. 2012. URL: <https://cds.cern.ch/record/1430736>.
 - [25] M. Petrov. *Search for heavy ZZ Resonances in the $\ell^+\ell^-\nu\bar{\nu}$ channel using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector at the LHC*. University of Oxford, 2018. URL: <https://ora.ox.ac.uk/objects/uuid:1bd2a3be-342e-4712-9159-ccb62b2a6899>.
 - [26] ATLAS Collaboration. *The ATLAS Experiment at the CERN Large Hadron Collider: A Description of the Detector Configuration for Run 3*. 2023. arXiv: [2305.16623 \[physics.ins-det\]](https://arxiv.org/abs/2305.16623).
 - [27] ATLAS Collaboration. “Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *The European Physical Journal C* 78.11 (Nov. 2018). URL: <https://doi.org/10.1140/2Fepjc%2Fs10052-018-6288-9>.
 - [28] Yann Coadou. “Boosted Decision Trees”. In: *Artificial Intelligence for High Energy Physics*. WORLD SCIENTIFIC, Feb. 2022, pp. 9–58. URL: https://doi.org/10.1142/2F9789811234033_0002.