



FACULTÉ
DES SCIENCES

UNIVERSITÉ LIBRE DE BRUXELLES



CERN-THESIS-2021-160
24 September 2021

Measurement of the transverse momentum of Drell-Yan lepton pairs over a wide mass range in proton-proton collisions at $\sqrt{s} = 13$ TeV in CMS

Thesis presented by Louis MOUREAUX

in fulfilment of the requirements of the PhD Degree in Science
("Docteur en Sciences")
Academic year 2020-2021

Supervisor: Prof. Laurent FAVART

Inter-university institute for high energies

Thesis jury:

Prof. Pascal VANLAER (Université libre de Bruxelles, Chair)
Prof. Riccardo ARGURIO (Université libre de Bruxelles, Secretary)
Prof. Laurent FAVART (Université libre de Bruxelles)
Dr. Maarten BOONEKAMP (CEA Saclay)
Dr. Philippe GRAS (CEA Saclay)
Prof. Francesco HAUTMANN (University of Oxford)
Prof. Vincent LEMAÎTRE (Université catholique de Louvain)



Acknowledgements

For every Ph. D. student there is someone special, and that person for me was Laurent Favart. The many hours spent using little drawings on the blackboard to understand the physics of some particular kinematic configurations were invaluable. I am immensely grateful to Laurent for proposing very interesting subjects for my master and Ph. D. theses, his careful and smooth supervision, and unfailing support in the last five years.

This thesis would for sure have been very different without my best office, conference, shift, walking, debugging, pain, joy, and chocolate provider mate Buğra Bilin. What could have been a simple professional relationship quickly became a true friendship, and for several parts of the analysis it is impossible to tell whose idea it was in the first place. Buğra, I hope you will never need to open that restaurant, but if you ever do I will be your first client.

I learned a lot by working together with Anastasia Grebenyuk between two P&C meetings, trying to nail down the differences between our two codes. So many bugs were caught this way! Nastja was also instrumental in obtaining funding for this thesis. Thank you for everything! It was a pleasure to watch the two new members of the Brussels team, Max Vanden Bemden and Itana Bubanja, progressively getting their bearings in the collaboration. I'm sure you will rock the Run 2 analysis.

I am also grateful to all other attendants of the Thursday meetings, and especially Philippe Gras for his help with Shears and unfolding and Jelena Mijuskovic for showering the GENEVA samples. The discussions held there with Dan Arcaro, Larry Sulak, Qun Wang, and occasionally others, were very interesting. Many thanks to Hugues Brun for the discussions on muon performance during the early stages of the analysis and to Laurent Thomas for his help with photon-induced lepton production.

I would like to thank the members of my jury: Pascal Vanlaer, Riccardo Argurio, Maarten Boonekamp, Philippe Gras, Francesco Hautmann, and Vincent Lemaître, for the interest that they have shown in my work and for the fruitful discussion that we had at the private defense.

A measurement is nothing without prediction to compare to, and for this

Hannes Jung and Armando Bermudez Martinez were especially helpful. I am also grateful to the authors of GENEVA, Simone Alioli, Christian Bauer, and Frank Tackmann, for generating samples and providing them to us, and also for their investigation and fix of the QED FSR bug. Many thanks also to Ignazio Scimemi and Alexey Vladimirov for providing tables for the ARTEMIDE prediction.

CMS is a large collaboration and some members have the ingrate task of implementing management policies, reading through the analysis notes and coordinating the various teams. I would like to thank the SMP-V conveners Alberto Belloni, Lorenzo Bianchini, and Cecile Caillol; the SMP conveners Paolo Azzurri, Guillermo Ceballos, Josh Bendavid, and Sunil Bansal; and the GEN conveners Alexander Grohsjean, Qiang Li, and Efe Yazgan for their comments and support. I would also like to thank the members of our ARC Slawek Tkaczyk, Marc Dunser, Sunil Bansal, and Arie Bodek for their in-depth review of our analysis.

The time spent in the office with Nicolas Postiau and Hugo Delannoy discussing all kinds of topics between physics (obviously), teaching (a lot), programming (and segfaults), or politics was very enjoyable. But the most dangerous person in the lab is probably Yannick Allard, with which a five minutes discussion quickly becomes hours. Yannick's contagious enthusiasm for electronics has been amazing. I wish we had more time for our Trackella or bubble chamber ideas and I'm looking forward to the new power supplies for the Franck-Hertz lab.

For the good times we had remaking the world (and mainly the GEM DAQ), I would like to thank Laurent Pétré. Some good ideas now used in production at P5 can be traced back to our endless discussions between late and very late in the evening. I should also mention some other GEM DAQ colleagues Misha, Jared, Camilla, Pieter, and Evaldas with whom I wish I could have worked more (one cannot do everything!). I also really appreciated working together with Pranjal, Avijit, and Sree; I wish you all the best for the future.

The life of an institute like IIHE depends on a number of hands and heads for logistics, infrastructure, and support. For the ones I know I would like to thank the ever smiling Audrey as well as Marleen and Sofie; our XROOTD experts and T2B admins Olivier and Romain, and also Denis, Adriano, and Shkelzen; and finally Patrick for the drive to CERN when a strike was threatening my flight.

During my time at ULB, I worked as an assistant for several courses. I deeply enjoyed the sparkles in the eyes of some of the students I had the chance to meet. I wish more of them would have chosen particle physics. I would also like to thank some of the colleagues I enjoyed teaching with: Isabelle, Maxime, Barbara, Juanan, Ioana, Daniel, Stéphane.

These Acknowledgements couldn't be complete without mentioning the constant support from my family through the highs and lows of the past four, nine, and even twenty years. Special thanks go to my parents, brother, real and adoptive

grand-parents, aunts, and close and remote cousins.

Last but not least, I am thankful to FNRS for providing the funding for this thesis through two FRIA grants, as well as support for traveling to CERN and attending schools and conferences.

Contents

1	Introduction	1
2	Theory elements	3
2.1	The Standard Model of particle physics	3
2.1.1	Quantum electrodynamics	5
2.1.2	Quantum chromodynamics	8
2.1.3	Electroweak unification	11
2.2	The Drell-Yan process	12
2.2.1	Collinear factorization	13
2.2.2	The partonic cross section	16
2.3	Beyond fixed order	18
2.3.1	Transverse momentum resummation	20
2.3.2	Parton showers	24
2.3.3	Parton branching solution of DGLAP evolution	27
2.4	Small q_T factorization	27
2.5	Summary	30
3	Experimental setup	33
3.1	The Large Hadron Collider	33
3.1.1	Injection chain	34
3.1.2	Luminosity	35
3.2	The CMS detector	38
3.2.1	Superconducting solenoid magnet	39
3.2.2	Inner tracker	40
3.2.3	Electromagnetic calorimeter	42
3.2.4	Hadron calorimeter	45
3.2.5	Muon detectors	47
3.2.6	Trigger system	50
3.2.7	Luminosity meters	51

4	Drell-Yan measurement	53
4.1	Observables	54
4.2	Existing measurements	57
4.3	Data and signal samples	57
4.4	Event reconstruction and selection	59
4.4.1	Tracks and vertices	60
4.4.2	Electrons	60
4.4.3	Electron triggers	62
4.4.4	Muons	67
4.4.5	Muon triggers	73
4.4.6	Jets	74
4.4.7	Event selection	76
4.4.8	Backgrounds	81
4.5	Kinematic distributions	84
4.5.1	Individual objects	85
4.5.2	Measured observables	89
4.6	Additional background checks	94
4.6.1	“Fake” electrons	94
4.6.2	Top quark pair production	104
4.7	Unfolding	111
4.7.1	Generator level phase space	112
4.7.2	Response matrices	113
4.7.3	Number of iterations	118
4.7.4	Modeling uncertainties	119
4.8	Uncertainties	122
4.9	Combination	126
4.9.1	Comparison at the detector level	126
4.9.2	Combination of the results	130
4.9.3	Uncertainties in the combined results	135
4.10	Cross section ratios	142
4.11	Predictions	149
4.11.1	MadGraph5_aMC@NLO	149
4.11.2	ArTeMiDe	150
4.11.3	CASCADE	155
4.11.4	MiNNLO _{PS}	156
4.11.5	Geneva	156
4.11.6	Cross section ratios	158
4.11.7	Formal accuracy	158
4.12	Results	161
4.12.1	Transverse momentum	161

4.12.2	φ_η^* variable	172
4.12.3	Transverse momentum, one or more jets	182
5	Conclusion	191
A	Statistical framework	197
A.1	Properties of the variance	198
A.2	Weighted events	199
A.3	Systematic uncertainties	199
A.4	Differential cross sections	200
A.5	Particular covariance matrices	201
A.6	Correlation and interpretation	202

Chapter 1

Introduction

Particle physics, or the study of the elementary constituents of matter, has emerged as a field of its own in the second half of the XXst century. The first sources of particles were found in nature, but it was soon realized that particle accelerators could provide much larger and better controlled samples. With the growing energy of collisions, more and more particles were discovered. Efforts to classify them lead to the early formulations of what is now called the Standard Model of particle physics, the evolving theory that describes fundamental particles and their known interactions (except, at the time of this writing, for gravitation).

What is considered a fundamental particle is not fixed and has evolved over time. In particular, protons and neutrons were found to be composite bound states “made” of quarks and gluons. This discovery allowed the description of a large family of hundreds of particles, hadrons, as assemblies of only six more fundamental quarks and eight gluons. This unification, however, comes at the cost of the additional theoretical complexity of relating quarks to hadrons, a problem not yet fully solved.

The discovery of the heaviest fundamental particles known today were made at hadron colliders: the W and Z bosons were first observed at Sp $\bar{p}$ S, the top quark was found at TeVatron and the Higgs boson was discovered at the LHC. The interpretation of these discoveries and searches for new physics rely on a thorough understanding of the behaviour of the collided objects — for the discoveries mentioned above, protons and antiprotons. On the theoretical side, this knowledge is provided by factorization theorems that provide the link between hadrons, quarks and gluons and allow accurate predictions of cross sections. On the experimental side, the theory is constantly being pushed to its limits by comparing its predictions to ever more precise measurements.

Since 2010, the Large Hadron Collider [1] has been in operation at CERN, Geneva, colliding protons at record center-of-mass energies of $\sqrt{s} = 7, 8$, and 13 TeV at four interaction points, around each of which is located an experiment:

ALICE [2], ATLAS [3], CMS [4], and LHCb [5]. One of the main goals of the LHC, the discovery of the Higgs boson, was achieved in 2012 [6, 7] by the ATLAS and CMS collaborations. Extensive searches have so far revealed no signs of physics beyond the Standard Model. In parallel to searches for new particles, measurements of Standard Model processes are carried out, reaching or exceeding the precision of previous collider results thanks to the large center-of-mass energy, luminosity, and constantly improving detectors and experimental methods.

This thesis is part of the experimental programme to test theoretical results at the highest energies currently available at colliders. The kinematic properties, and in particular the transverse momentum, of Z bosons produced at the LHC are used as a proxy to study the production of heavy particles in proton-proton collisions at $\sqrt{s} = 13$ TeV. Using data collected in 2016 by the CMS detector, the analysis presented in this thesis provides a precision measurement of the transverse momentum of pairs of electrons and muons produced in the decay of off-shell Z/γ^* bosons with invariant masses between 50 and 1000 GeV.

This thesis follows the traditional organization of theses in experimental high-energy particle physics: the theoretical context of the work is explained first, in chapter 2, followed by a presentation of the experimental setup in chapter 3 and the data analysis in chapter 4, in which the results are also compared to predictions. Conclusions are given in chapter 5.

Chapter 2

Theory elements

This chapter summarizes the status of the theory related to the production of heavy particles at hadron colliders, with a particular focus on the Drell-Yan process which is at the core of this thesis. Important aspects of the evolving theory of fundamental interactions, the Standard Model of particle physics, are recalled in section 2.1. The relation between the Drell-Yan process and the fields appearing in the Standard Model, obtained through factorization, is explored in section 2.2. The tools that allow calculations beyond an expansion in powers of the coupling constants, required for accurate predictions of transverse momentum distributions, are described in section 2.3. An important class of generalized factorization theorems is covered in section 2.4. The last section of this chapter, 2.5, explains how the different approaches are combined to provide realistic predictions valid over the whole phase space.

The main references used in this chapter are the books by Schwartz [8] for general results in quantum field theory and Campbell, Hutson, and Krauss [9] for topics specific to QCD.

2.1 The Standard Model of particle physics

The Standard Model (SM) encodes the current understanding of the physics of very small objects — particles — in a compact, unified description. The SM is an evolving theory that has been progressively extended to incorporate new effects as they were discovered in experiments. Of the known fundamental interactions, only gravitation could so far not be included in the SM.

The SM is built around a collection of fundamental quantum fields coupled with each other, listed in figure 2.1. The spin- $\frac{1}{2}$ (fermionic) fields describe “matter” and the integer spin (bosonic) fields represent interactions. Particles correspond to excitations of the fields and in practice, the two names are interchangeable. Only

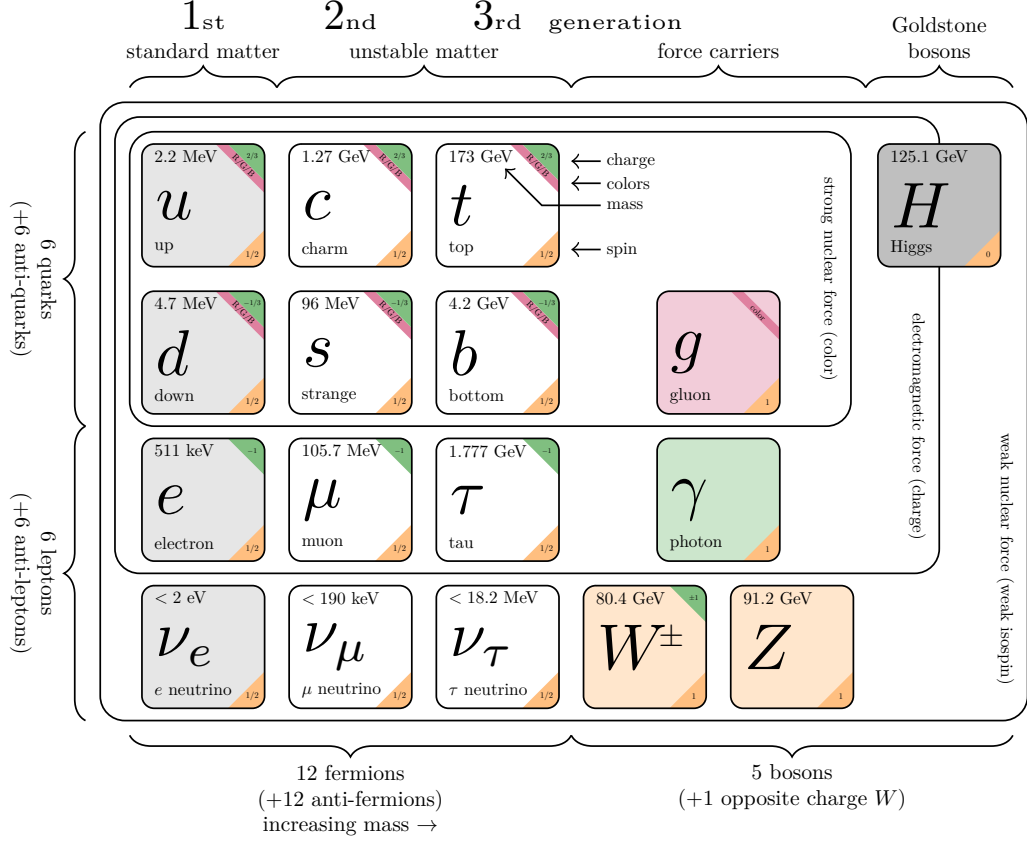


Figure 2.1: Summary of the Standard Model of particle physics. Updated from [10] with the latest world averages [11].

four of the matter particles are abundant in nature: the electron and its neutrino, and the up and down quarks that are constituents of the proton, bound together by a strong gluon field. The two other neutrinos are also relatively common in nature, but their interactions are rare and they are therefore extremely difficult to detect. The last six fermions in the SM are unstable, ephemeral particles that are only produced in high-energy collisions.

Matter particles interact through the exchange of virtual bosons: the electromagnetic force is mediated by massless photons exchanged between charged particles, the strong nuclear force by eight gluons and the weak interaction by three heavy bosons noted Z and $W^\pm$. The explanation of weak interactions required the introduction the Brout-Englert-Higgs [12, 13] scalar field and the associated symmetry breaking mechanism. This prediction was confirmed in 2012 by the ob-

servation of the Higgs boson by the ATLAS [6] and CMS [7] collaborations, forever changing the understanding of the origin of mass.

The next sections describe the interactions between the particles in the Standard Model, starting with quantum electrodynamics in section 2.1.1, where an important aspect of all quantum field theories, the running of the coupling constants, is illustrated. The theory of the strong nuclear force, quantum chromodynamics, is described in section 2.1.2. Finally, section 2.1.3 covers the weak interaction and its unification with electromagnetism.

2.1.1 Quantum electrodynamics

Quantum electrodynamics (QED) describes electromagnetic interactions. It is based on the realization that the normalization of fields are defined up to a phase and the assumption that two points of space-time do not necessarily share the same phase — how would they conspire to do so if they are causally disconnected? Granting this freedom of choosing the phase independently for each point of space-time almost directly leads to the formulation of QED. More specifically, QED is a gauge theory with a local symmetry obeying the $U(1)$ group. The gauge transformations are closely related to the ones appearing in classical electromagnetism. In its simplest formulation, QED contains the massless photon field A_μ and one massive spinor field ψ representing the electron. The Lagrangian density is given by:

$$\mathcal{L}_{\text{QED}} = -\frac{1}{4}F_{\mu\nu}^2 + i\bar{\psi}\not{D}\psi - m_e\bar{\psi}\psi, \quad (2.1)$$

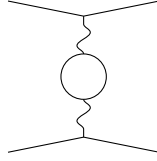
with $D_\mu = \partial_\mu + ieA_\mu$ the covariant derivative and e the absolute charge of the electron. Other notations are to be found in reference [8]. More matter fields can trivially be added by adding the corresponding kinematic and mass terms $i\bar{\psi}\not{D}\psi + m^2\bar{\psi}\psi$ to the Lagrangian density and adjusting the mass and charge accordingly.

All QED interactions between charged particles happen through the photon field, even when no measurable photon is involved. This leads to the idea that electromagnetic interactions are mediated by the exchange of “virtual” photons. The price to pay for this interpretation is that the mass of the virtual photons is not constrained and can be arbitrary large or even not real. At the level of the quantum mechanical amplitude $\mathcal{A}$, the interaction between an electron and a muon can be thought of as a series of photon exchanges between them:

$$\mathcal{A}[e\mu \rightarrow e\mu] \propto 1 + \text{diagram 1} + \text{diagram 2} + \text{diagram 3} + \dots$$

Because interactions between fermions and photons happen through the $e\bar{\psi}A\psi$ terms in the Lagrangian, each $\sim\!<$ vertex contributes a factor e to the amplitude. Since additional internal vertices always come in pairs, adding more vertices corresponds to an expansion of the interaction cross section in powers of $\alpha \equiv e^2/4\pi \approx 1/137$. The small value of the coupling motivates an expansion of the calculations in powers of α , or equivalently in the number of (virtual and real) photons.

A general feature of quantum field theories is that their parameters depend on a dimensionful “renormalization scale” μ_R with the dimensions of an energy. The role of the scale is to set the level of detail at which the theory is probed, which can be thought of in terms of space-time distances through $\lambda_R = \mu_R^{-1}$ — large scales corresponding to small distances. The scale dependence of the parameters can be justified by considering the same interaction between two charged particles as above. When the exchanged energy is small, the wavelength of the virtual photon is large and any quantum fluctuations happening around the electron and the muon do not contribute: their classical charge should be recovered as $\mu_R \rightarrow 0$. At high energy, the photon might polarize vacuum through the creation of a short-lived electron-positron or muon-antimuon pair:



As it turns out, this vacuum polarization diagram contributes constructively to the interaction amplitude: the strength of electromagnetic interactions increases at high energy. More precisely, in the high-energy limit the amplitude is enhanced by a factor:

$$1 + n_f \frac{\alpha}{3\pi} \log \frac{\mu_R^2}{\mu_0^2} + \mathcal{O}(\alpha^2), \quad (2.2)$$

where n_f is the number of fermion species that contribute to the loop diagrams, μ_R can be identified with the energy of the exchanged photon and the signification of μ_0 will be explained later.

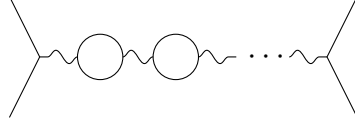
The enhancement of equation (2.2) can be absorbed in the coupling:

$$\alpha \rightarrow \alpha_{\text{eff}}(\mu_R^2) = \alpha_0 \left[1 + n_f \frac{\alpha_0}{3\pi} \log \frac{\mu_R^2}{\mu_0^2} + \mathcal{O}(\alpha_0^2) \right]. \quad (2.3)$$

With this definition, the coupling $\alpha_{\text{eff}}(\mu_0^2) = \alpha_0$ can be measured at a scale of reference μ_0 and “evolved” towards the scale μ_R for which a prediction is to be made. It can be shown that the evolution of α_{eff} is independent of the process

considered, which ultimately has to do with the way counter-terms are introduced in the Lagrangian to regularize certain divergences.

The expression given in equation (2.3) above is problematic for evolving between vastly different scales. Indeed, at every order α_0^n there is a diagram that gives a factor $\log^n \mu_R^2/\mu_0^2$ (the diagram has been rotated by 90° for clarity):



In the limit $\mu_R \rightarrow \infty$, the logarithms become larger than α and ruin the accuracy of the calculation. Fortunately, it is possible to sum the logarithmic terms to all orders in α . One starts by remarking that the above n -loops diagram contributes the following term:

$$\left(n_f \frac{\alpha_0}{3\pi} \log \frac{\mu_R^2}{\mu_0^2} \right)^n \quad (2.4)$$

Summing the series to $n = \infty$ is then trivial and gives the running of α in the leading-logarithmic (LL) approximation:

$$\alpha_{\text{eff}}(\mu_R) = \frac{\alpha_0}{1 - n_f \frac{\alpha_0}{3\pi} \log \frac{\mu_R^2}{\mu_0^2}}. \quad (2.5)$$

Different logarithms appear when accounting for more complicated loops, giving rise to more logarithmic approximations referred to as NLL, NNLL, etc.

Another, more convenient way of deriving equation (2.5) is by using the renormalization group equation, which follows from the requirement that the Lagrangian may not depend on the arbitrary scale μ_R , even when written in terms of α (dropping the “eff” from now on):

$$\mu_R^2 \frac{d\alpha}{d\mu_R^2} = \beta_{\text{QED}}(\alpha), \quad (2.6)$$

where the β_{QED} function encodes the loop calculations. Usually, β is calculated order-by-order in α , $\beta_{\text{QED}} = \sum_{k=0}^{\infty} \beta_k \left(\frac{\alpha}{4\pi}\right)^{k+2}$. The first coefficients in pure QED with any number of fermions are $\beta_0 = \frac{4}{3}n_f$ and $\beta_1 = 4n_f$, and are known up to five loops (β_4) [14].

Some remarks are in order before closing this section:

- Because the β function of QED is positive, the coupling increases with the scale. This running is small: $\alpha \approx \frac{1}{137}$ at $\mu_R = 0$ and $\alpha \approx \frac{1}{128}$ at $\mu_R = m_Z$ (where m_Z is the mass of the Z boson), and α remains small at every conceivable energy.

- The coupling is not the only parameter that evolves with the energy scale. In fact, every free parameter of the Lagrangian is allowed to do so. For QED, the masses do not vary strongly within scales accessible to experiments and their running can usually be neglected. This is not the case for the strong interaction and so the masses of the quarks do depend on the scale.
- Which terms are absorbed in the coupling defines a “renormalization scheme”. There are infinitely many possible definitions and today most predictions use the so-called $\overline{\text{MS}}$ scheme. The choice of scheme is particularly important when dealing with quark masses.
- As encoded in the renormalization group equation, the scale μ_R can in principle be chosen arbitrarily without changing the results. However, calculations performed at a fixed order in α do depend on the choice of scale. When performing predictions, it is useful to remember its physical interpretation as the resolution at which the theory is probed. The scale is thus often chosen to be a mass or energy entering the process of interest.

2.1.2 Quantum chromodynamics

The formulation of quantum chromodynamics (QCD) resembles that of QED. The main difference is the group upon which it is built: the local symmetry is $SU(3)$ instead of $U(1)$. $SU(3)$ is a non-abelian group whose fundamental representation has three elements, called “colors” in the context of QCD. There are three copies of the fermion (quark) fields, one per color, that are intermixed by the symmetry. The equivalents of the photon are eight gluons, corresponding to the dimension of the adjoint representation of $SU(3)$. Noting a the color index of gluons and i, j that of the fermions and summing over repeated indices, the Lagrangian density reads:

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4}(G_{\mu\nu}^a)^2 + i\bar{\psi}_i \not{D}_{ij} \psi_j - m_q^2 \bar{\psi}_i \psi_i, \quad (2.7)$$

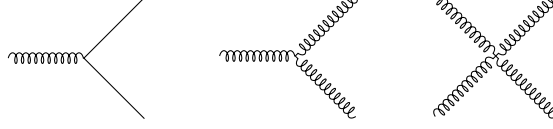
The covariant derivative is given by $D_{\mu,ij} = \partial_\mu \delta_{ij} + ig_s G_\mu^a T_{ij}^a$, where the G^a are the 8 gluon fields, T^a are the generators of $SU(3)$ in the adjoint representation and $\alpha_s \equiv g_s^2/4\pi$ is the strong coupling. The gluon kinematic term is slightly different from the QED case because of the non-abelian nature of the theory:

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s f^{abc} G_\mu^b G_\nu^c, \quad (2.8)$$

with f^{abc} the structure constants of $SU(3)$, with the indices a, b , and c running from 1 to 8. Like for QED, additional quark fields can be added by merely repeating the ψ -terms. In the Standard Model, all quarks share the same coupling g_s to gluons.

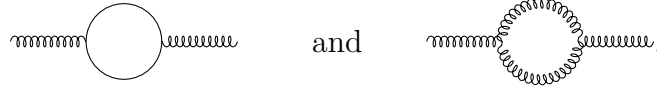
The difference between the covariant derivatives of QED and QCD has profound consequences. The presence of the $g_s f^{abc} G_\mu^b G_\nu^c$ term results in interactions

between gluons in addition to the interactions between gauge bosons and fermions also found in QED. The three vertices allowed in QCD are as follows, with gluons represented as helices and fermions as straight lines:



When calculating matrix elements, the three-branches vertices carry a factor g_s while using the “cross” vertex multiplies the amplitude by g_s^2 .

The major difference between QED and QCD appears in the evolution of α_s . The triple gluon vertex implies that several types of loops enter the LL calculation; ignoring the quartic vertex and ghosts:



The contribution of the additional loops is large and negative. Indeed, the QCD β function reads at leading orders [14]:

$$\beta_{\text{QCD}}(\alpha_s) = - \left(11 - \frac{2}{3}n_f \right) \left(\frac{\alpha_s}{4\pi} \right)^2 - \left(102 - \frac{38}{3}n_f \right) \left(\frac{\alpha_s}{4\pi} \right)^3 - \mathcal{O}(\alpha_s^4). \quad (2.9)$$

With at most 6 fermions contributing in the Standard Model, the QCD β function is negative, which implies that the strong coupling decreases monotonically at high energies. It actually becomes zero in the high energy limit, a feature known as “asymptotic freedom”. In the other direction, the strong coupling diverges and becomes infinite at around $\mu_R = \Lambda_{\text{QCD}} \approx 200 \text{ MeV}$. Below a few times this scale, using expansions in the coupling constant is no longer possible. The strength of the coupling gives rise to “confinement”: at low energy, gluons and quarks always appear as parts of color neutral bound states (hadrons) and are called “partons”. The value of the strong coupling extracted from data is $\alpha_s(m_Z^2) = 0.118 \pm 0.001$. Its running is illustrated in figure 2.2.

The divergence of α_s at small energies, or equivalently for long distances and times, forbids the use of an expansion in powers of α_s at scales under a few Λ_{QCD} . This regime includes most hadron wave functions, masses, and widths, as well as most low-energy interactions. Increasing success in predicting these variables is obtained by solving the Lagrangian in a discretized space-time known as the lattice. For the physics dominated by large scales studied in this thesis, lattice calculations are either in their infancy or simply do not exist. All methods discussed in the next sections separate out perturbative and non-perturbative parts, using expansions in powers of α_s for the former and parameterizations constrained from measurements for the latter.

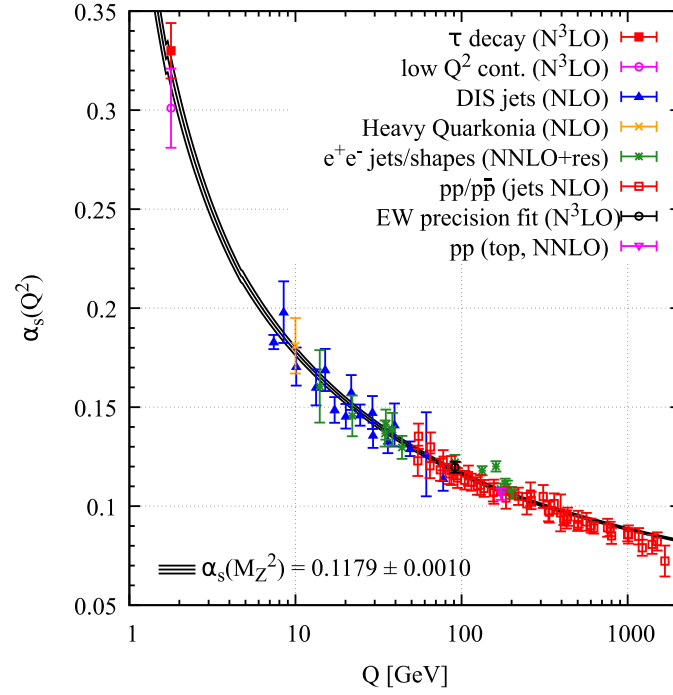


Figure 2.2: Running of the strong coupling constant α_s as a function of the scale $Q = \mu_R$. The data points correspond to experimental input used in the fit. Reproduced from [11].

2.1.3 Electroweak unification

The weak and electromagnetic interactions are unified in a single spontaneously broken gauge theory. At very large energies, a local symmetry exists that relates pairs of particles transforming in the fundamental representation of the group $SU(2) \times U(1)$. Each charged lepton has for partner the corresponding neutrino. Quarks are similarly grouped in doublets: up with down, charm with strange, and top with bottom. Four bosons allow the symmetry to be local, with the same role as photons in QED and gluons in QCD. At the very high energies considered so far, all these fields are massless. Finally, a pair of two complex scalar “Higgs” fields is present that interacts with every other doublet.

The dynamics of the Higgs fields is non-trivial and a potential forces them to acquire an average value different from 0 in vacuum, breaking the symmetry between electroweak partners. Because of the non-zero vacuum expectation value, interaction terms between the various fields and the Higgs turn into standard mass terms proportional to their coupling to the Higgs field. Three of the four electroweak gauge bosons also acquire a mass, each absorbing one degree of freedom from the Higgs fields to form the $W^\pm$ and Z bosons. The last electroweak gauge boson remains massless and becomes the photon of QED. Oscillations of the last remaining degree of freedom of the Higgs fields around its vacuum expectation value create one last (scalar) particle, the Brout-Englert-Higgs boson [12, 13], which interacts with every particle in the Standard Model with a coupling proportional to their mass.

The gauge fields associated with $U(1)$ and $SU(2)$ before electroweak symmetry breaking are written B_μ and W_μ^a with $a = 1, 2, 3$, respectively. The conserved quantity associated with B_μ is called the hypercharge Y . The charge associated with the $SU(2)$ symmetry is called the weak isospin T_3 . The vacuum expectation value of the Higgs fields causes the fields and charges to mix with each other. The gauge bosons recombine to form four mass eigenstates:

$$W^\pm = \frac{1}{\sqrt{2}}(W^1 \mp W^2) \quad (2.10)$$

$$Z = \cos \theta_W W^3 + \sin \theta_W B \quad (2.11)$$

$$A = \sin \theta_W W^3 + \cos \theta_W B \quad (2.12)$$

The first two mass eigenstates correspond to the massive $W^\pm$ bosons with electric charges ± 1 , the Z field corresponds to the massive neutral Z boson, and the A field is the massless QED photon. The parameter θ_W that controls the relative importance of hypercharge and weak isospin is called the weak mixing angle. Its value is measured to be $0.22 < \sin^2 \theta_W < 0.24$ depending on the renormalization scheme and is known to five significant digits in individual schemes [11].

The electric charge of fermions is related to their hypercharge and weak isospin through:

$$Q = T_3 + \frac{Y}{2}. \quad (2.13)$$

Similar relations (also involving θ_W) hold for the couplings of the weak bosons $W^\pm$ and Z to quarks and between themselves. Finally, the Brout-Englert-Higgs boson couples to all massive fields in the Standard Model with a coupling proportional to their mass. The detailed formulation of the Standard Model can be found along with all couplings and Feynman rules in quantum field theory textbooks, e.g. references [8, 15].

At energies much smaller than the mass of the Z and $W^\pm$ bosons, the weak interaction is highly suppressed. It can, however, be observed when neither the electromagnetic nor the strong nuclear forces are allowed, for instance in β radioactive decays and the decay of many particles, including muons and pions. At energies around the mass of the Z and $W^\pm$ bosons, the weak and electromagnetic interactions are comparable in strength. In interactions between charged particles, as is the main topic of the present work, the interference between the Z boson and virtual photons is also sizeable.

2.2 The Drell-Yan process

A process of tremendous importance at the LHC is the production of a pair of leptons of opposite charges, a process named Drell-Yan after the name of the two theorists who gave the first successful interpretation [16] in the parton model of contemporary observations in proton-uranium collisions [17]. In the proton-proton collisions considered in this dissertation, the Drell-Yan process can be written as follows:

$$p + p \rightarrow \ell^+ \ell^- + X, \quad (2.14)$$

where $\ell^+ \ell^-$ represents either a pair of electrons or a pair of muons (excluding taus for simplicity) and X means that other products of the collision are in principle ignored. The importance of the Drell-Yan process stems from its easily measured experimental signature, sizeable cross section, and the existence of the well-developed theoretical framework outlined in the remainder of this chapter.

Before delving into the theory, it is necessary to define the kinematics. It is customary to work in the center-of-mass frame of the incoming protons and take the z axis in the direction of the beams. With this orientation, the x and y axes define the “transverse” plane in which the total momentum is zero. The center-of-mass energy of the colliding protons is noted $\sqrt{s}$ and q is the four-momentum of the lepton pair, with invariant mass $Q^2 \equiv q^2 < s$. There is no consensus on the notation for the projection of q in the transverse plane, with q_T often used

by theorists and p_{T} by experimentalists. As a consequence, q_{T} is preferred in this chapter while p_{T} will be used in the rest of this thesis.

The rapidity y , not to be confused with the spatial y coordinate, is defined as:

$$y = \frac{1}{2} \log \frac{q^0 - q^z}{q^0 + q^z}. \quad (2.15)$$

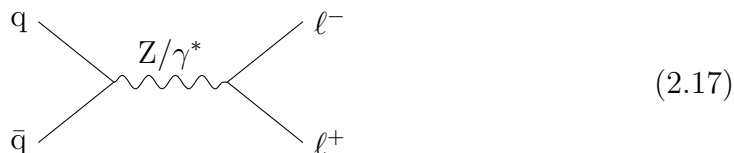
This definition is not restricted to q and can be used for any four-vector. It is also useful to introduce two variables x_1 and x_2 :

$$x_1 = \frac{Q}{\sqrt{s}} e^{+y}, \quad x_2 = \frac{Q}{\sqrt{s}} e^{-y}. \quad (2.16)$$

With this definition, Q and s are related through $Q^2 = x_1 x_2 s$ and one can think of x_1 and x_2 as the fraction of the momentum of each proton “used up” to create the state of mass Q . This idea will be precised in the next section.

2.2.1 Collinear factorization

The interactions of the two protons entering the collision are not straightforward to calculate in the Standard Model because protons do not appear explicitly in the Lagrangian. The lowest Feynman diagram associated with the Drell-Yan process in the Standard Model is the annihilation of a quark-antiquark pair through a Z/γ^* boson, $q\bar{q} \rightarrow Z/\gamma^* \rightarrow \ell^+\ell^-$:



The corresponding cross section is of order α^2 . The mass dependence is given by the photon and Z propagators and their interference, resulting in a steeply falling cross section except for a peak at $Q = m_Z$.

Relating the quark and anti-quark appearing in the initial state of the diagram above to the colliding protons is not trivial because the protons are bound states whose properties cannot be calculated with an expansion in α_s . However, in the limit of short durations or large energies, α_s vanishes and quarks and gluons behave as free particles. If Q is large with respect to Λ_{QCD} , one may expect that this is a good approximation at the time of the creation of the Z/γ^* boson. Before the high-energy part of the interaction, the two protons approaching each other are separated by a space-like interval and so are their remnants after the collision. If the high-energy part of the interaction happens in a time μ_F^{-1} , one can suppose that

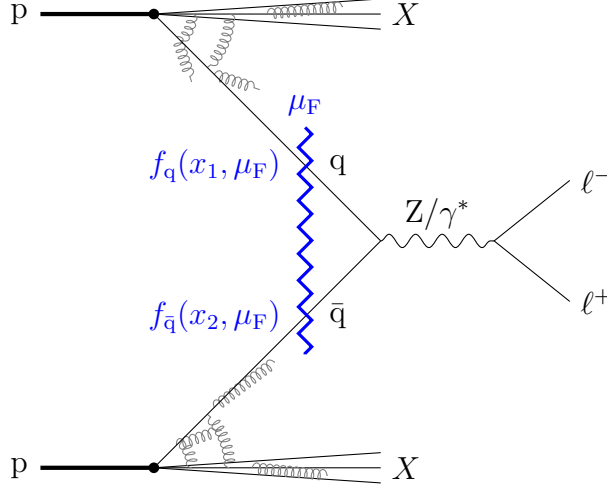


Figure 2.3: Schematic illustration of collinear factorization.

the protons are independent from each other for time scales longer than μ_F^{-1} . This statement can be rephrased to be about the energy scale: interactions happening at an energy scale smaller than μ_F can only affect a single proton. This separation between “soft” (small) and “hard” (large) energy scales is represented as a broken line in figure 2.3.

With the protons being independent from each other at soft scales, one can factorize the cross section into the probabilities $f_q(x_1, \mu_F)$ and $f_{\bar{q}}(x_2, \mu_F)$ of finding a quark and an antiquark in the protons at a scale μ_F and a coefficient $\hat{\sigma}$ describing the hard part of the collision. The parton distribution functions (PDF) f_a depend on the fraction x_a of the momentum of the proton carried by the parton a . In the general case of the production of a hard final state H , one obtains:

$$d\sigma_{pp \rightarrow H+X} = \sum_{a,b} \int_0^1 dx_a dx_b f_a(x_a; \mu_F) f_b(x_b; \mu_F) d\hat{\sigma}_{ab \rightarrow H}(x_a, x_b; \mu_R, \mu_F) \quad (2.18)$$

The sum runs over all parton flavors that can contribute to the production of H . The “partonic cross section” $\hat{\sigma}_{ab \rightarrow H}$ describes the interaction between the two partons, as if they were free. It depends on their energies through x_1 and x_2 , and on the scale μ_F at which factorization is performed. In practice, calculations are limited to a few orders in α_s and thus they also depend on a renormalization scale μ_R . This “collinear factorization” result has been proven to all orders for the Drell-Yan process [18] and appears to work in many other cases.

From the discussion above, it is clear that the factorization scale should be chosen to the value of some high-energy quantity appearing in the process under consideration. For Drell-Yan, $\mu_F = Q$ is a popular choice. For more complicated

processes, there are often several ways of choosing the scale. Since there is often no reason to prefer a choice over another, the choice of scale is a theoretical uncertainty. Conventionally, variations of μ_F by factors two and one half are considered to estimate the scale dependence of any prediction. The same applies to the renormalization scale: while it is usually taken to be identical to μ_F , there is no reason for this to be the best choice and μ_R is thus varied independently. Scale uncertainties will be heavily used in section 4.11.

The dependence of the parton distribution functions f_a in the factorization scale μ_F can be calculated using the renormalization group equation. The diagrams analogous to the loops for the coupling constants are the emission of collinear partons off the lines entering the hard process $\hat{\sigma}$. The following “splittings” appear at leading order:

$$P_{q \rightarrow qg} = P_{\bar{q} \rightarrow \bar{q}g} = \text{[diagram of a quark line emitting a gluon]} \quad (2.19)$$

$$P_{g \rightarrow gg} = \text{[diagram of a gluon splitting into two gluons]} \quad (2.20)$$

$$P_{g \rightarrow q\bar{q}} = P_{g \rightarrow \bar{q}q} = \text{diagram} \quad (2.21)$$

Since splittings can change the nature of a parton, the renormalization group equation introduce interplay between the different PDFs, in the form of the celebrated Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equation [19, 20, 21]:

$$\mu_F^2 \frac{\partial f_b(x, \mu_F)}{\partial \mu_F^2} = \sum_c \frac{\alpha_s(\mu_F^2)}{2\pi} \int_x^1 \frac{d\xi}{\xi} P_{a \rightarrow bc}(x/\xi, \mu_F) f_a(\xi; \mu_F). \quad (2.22)$$

This right hand side of this equation is most easily read backwards. A parton a starting with a momentum fraction ξ may undergo a splitting to flavors bc , losing energy and ending up at momentum fraction $x < \xi$. This happens with a probability parameterized by the “splitting function” $P_{a \rightarrow bc}$. Finally, the second produced parton c is ignored.

While the factorization theorem and the evolution equation of the PDF look nice, they aren't of much use if the PDFs are unknown. Fortunately, PDFs are universal: when factorisation applies, they do not depend on the hard process of interest. The same PDFs also appear in similar factorization theorems, for instance in electron-proton collisions. This remarkable universality allows the extraction of the PDFs using well-known processes to perform predictions with different final states or at different scales. For instance, PDFs extracted from HERA deep inelastic scattering data were widely used in the beginning of the LHC era [22]. Modern PDF sets are extracted in global analyses taking into account data from multiple experiments in several final states, and commonly include evolution up to NNLO. Most sets are made available through the LHAPDF library [23], which provides a unified software interface. Recent PDF sets agree within a few percent

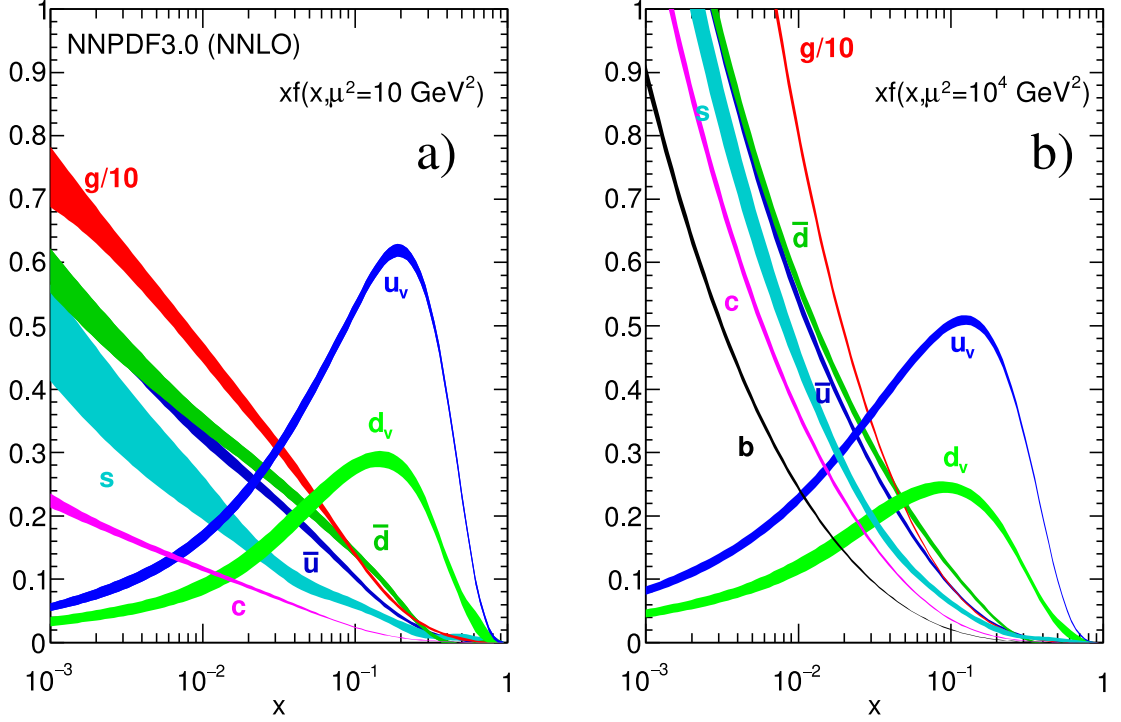


Figure 2.4: Parton density functions as obtained from the NNPDF 3.0 global fit [24], at scales $\mu_F = \sqrt{10}$ GeV (left) and 100 GeV (right). Reproduced from [11].

in x regions well covered by experiments, but predictions are much less accurate at the boundaries of the phase space, $x \rightarrow 0$ and $x \rightarrow 1$. As an example, figure 2.4 shows the results of a global fit performed by the NNPDF Collaboration [24].

2.2.2 The partonic cross section

The Drell-Yan cross section can be calculated by expanding the hard scattering cross section from the factorization formula (2.18) in powers of the coupling constants. At order α_s^0 , the only contribution is from quark-antiquark initial states with the diagram (2.17). The kinematics are simple: q_T is zero and the rapidity is solely given by the PDFs. At order α_s^1 , the picture change dramatically with the appearance of initial-state radiation diagrams, for instance:

$$(2.23)$$

This diagram has two very important consequences:

1. The antiquark in the initial state is replaced by a gluon. The large gluon PDF, about ten times larger than the antiquark PDFs for the kinematic regime considered in this thesis, compensates the additional power of α_s . Because of this, the corrections at order α_s^1 increase the total cross section by nearly a factor two.
2. The additional degrees of freedom added by the quark allow the Z/γ^* boson to be produced with non-zero transverse momentum. The full kinematics of the Z/γ^* boson thus only appear at order α_s^1 .

Calculations of the partonic cross section can be systematically improved by continuing the expansion to higher orders in α_s . The fully differential Drell-Yan cross section has presently been calculated up to order α_s^2 [25, 26]. Even at this order, the q_T distribution is unphysically large in the limit $q_T \rightarrow 0$. This is a sign of a deeper problem whose solution is the topic of section 2.3. Because of this issue, the validity of predictions performed order-by-order in α_s is limited to roughly $q_T \gtrsim Q/2$.

A class of prediction that resembles the fixed-order method mentioned above can be constructed by combining cross sections for different parton multiplicities in the final state, carefully removing divergent regions and avoiding double counting. While the formal accuracy of these “merged” predictions is not better than that of the most inclusive cross section, the hope is to include the most phenomenologically relevant higher order terms of the series. This procedure is most often employed in Monte-Carlo event generators to describe events with many jets: current technology allows the generation of Drell-Yan samples with up to four final-state partons at leading order or two partons at next-to-leading order, reaching contributions of order α_s^4 and α_s^3 , respectively. The limiting factor is the fast growth of the number of Feynman diagrams with the number of involved particles.

In addition to the development in powers of α_s , the effects of the first electroweak corrections at order α^3 have been investigated [27]. Their structure is different from QCD initial-state radiation because both the quarks and the leptons can emit photons. Several types of contributions appear. The emission of a photon off the initial-state quarks resembles the corresponding gluon case. Photons emitted by the leptons have the main consequence of modifying the kinematics: the invariant mass of the pair $m_{\ell\ell}$ can no longer be identified with that of the boson, $m_{\ell\ell} \neq m_{Z/\gamma^*}$. Finally, photon-initiated diagrams with the Z/γ^* in the t channel and the loop diagrams contributing at this order blur the picture of the production of a single electroweak boson, $q\bar{q} \rightarrow Z/\gamma^* \rightarrow \ell\ell$ — though the combined leptonic-photonic final state is still a color singlet. At LHC energies, electroweak

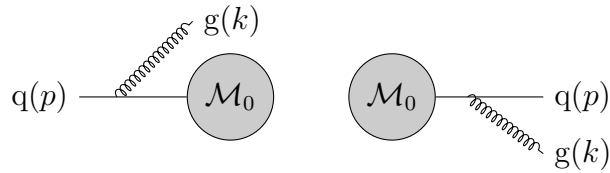
corrections smear the Z mass peak and reduce the total cross section by a few percents, especially at large $m_{\ell\ell}$.

Predictions using the factorization theorem (2.18) together with a partonic cross section calculated at order α_s^1 or α_s^2 and parton density functions extracted from data provide a description of the total cross section as well as $d\sigma/dy$ and $d\sigma/dQ$. The q_T -differential cross section at large q_T can be calculated using the same expansion in powers of α_s . For $q_T \lesssim Q/2$, the behaviour predicted by fixed-order calculations is unphysical. Other techniques are needed in this region, as described in the next section.

2.3 Beyond fixed order

The shortcomings caused by divergences appearing in fixed order calculations limit their domain of applicability when predicting differential cross sections. In the case of the Drell-Yan cross section, problems arise in the region of small to moderate q_T , $q_T \lesssim Q$, where the bulk of the cross section is located. In order to obtain differential predictions in this region, it is necessary to switch to a different expansion: that of large logarithms. Thanks to the remarkable fact that identical logarithmically divergent terms appear at every order in α_s for $q_T \rightarrow 0$, it is possible to sum them to all orders in α_s , a procedure known as q_T resummation and detailed in section 2.3.1. Parton showers, a closely related class of predictions used to implement resummation on top of fixed-order Monte-Carlo generators, are covered in section 2.3.2. In section 2.3.3, an implementation of the parton shower formalism as part of the DGLAP evolution is described. Before turning to these topics however, a more general discussion of the singularities hampering fixed-order calculations is in order.

One of the configurations leading to divergences is when an incoming or outgoing parton radiates another as depicted below for a quark emitting a gluon:



The gray circle stems for the matrix element $\mathcal{M}_0$ without emission, whose structure is technically not relevant but for which the Drell-Yan process will be assumed in the rest of the discussion. Even though a single quark line is represented, $\mathcal{M}_0$ can have any number of incoming and outgoing particles of any kind; emissions off these other particles have been ignored because the corresponding terms are not divergent in the region analyzed below. The four-momenta p and k are assigned to the quark and the gluon, respectively.

The addition of the gluon radiation modifies the matrix element with an additional quark propagator and coupling constant. Noting ϵ the polarization of the gluon, one obtains at leading order:

$$\mathcal{M}_1(p, k) = g_s \frac{p \cdot \epsilon}{p \cdot k} \mathcal{M}_0(p + k). \quad (2.24)$$

The subscript on $\mathcal{M}$ indicates the number of emitted gluons. Taking the square of the matrix element and averaging over helicities, one obtains the following modification to the cross section:

$$d\sigma_1 \approx \frac{\alpha_s}{2\pi} C_F P_{q \rightarrow qg}(z) dz \frac{dk_T^2}{k_T^2} d\sigma_0. \quad (2.25)$$

The two new variables z and k_T describe the kinematics of the emitted gluon in the directions parallel and perpendicular to the momentum of the quark, respectively. In the longitudinal direction, the cross section is driven by the DGLAP splitting function $P_{q \rightarrow qg}(z)$, where z is the momentum fraction conserved by the quark after emitting the gluon. The transverse momentum k_T is simply the projection of the gluon momentum in the plane perpendicular to the quark direction. Momentum conservation implies that after the emission, the quark gains a transverse momentum k_T with respect to its original direction, balancing that of the emitted gluon.

Equation (2.25) contains two divergences: one for $k_T \rightarrow 0$ and another for $z \rightarrow 1$, hidden in the expression for $P_{q \rightarrow qg}$ at leading order:

$$P_{q \rightarrow qg}(z) = \frac{1 + z^2}{1 - z}, \quad (2.26)$$

where terms at $z = 1$ have been ignored. The z divergence is an example of the “soft” divergences that occur when a massless particle is produced at zero energy. The case $k_T \rightarrow 0$ corresponds to a gluon emitted in the same direction as the quark, and the divergence is classified as “collinear”.

Before analyzing how the soft and collinear limits affect the kinematics of the electroweak boson in the Drell-Yan process, a small detour is needed to explain the interpretation of the associated divergences provided by the KLN theorem [28, 29]. Simply put, the theorem states that for well chosen observables, the infinities in real emission diagrams are compensated by the ones appearing in loops diagrams calculated at the same order in the coupling constants. This cancellation can only happen if the observable of interest O ignores particles emitted in the divergent limits. Formally, if Φ is a set of variables describing an event with n particles and $\Psi = (z, k_T)$ describes the kinematics of an additional particle, it is required that in the soft and collinear limits, O is not modified by the additional emission:

$$O_n(\Phi) = O_{n+1}(\Phi, \Psi_{\text{soft}}) = O_{n+1}(\Phi, \Psi_{\text{collinear}}). \quad (2.27)$$

In the definition above, O_n and O_{n+1} are the observable as defined for the kinematics of n and $n + 1$ particles, respectively. In the notations of equation (2.25), Ψ corresponds to z and k_T (possibly supplemented with an azimuthal angle). Ψ_{soft} is the limit $z \rightarrow 0$ and $\Psi_{\text{collinear}}$ the limit $k_T \rightarrow 0$. Variables that satisfy equation (2.27) are called “infrared-safe”. Perturbation theory is only reliable for the calculation of infrared-safe observables.

From the discussion above, it is clear that the number of particles n is not infrared safe, and neither are z and k_T . It is thus not surprising that the cross section for the emission of exactly one parton is badly divergent. In the case of the Drell-Yan process, an infrared-safe observable related to the emission of initial-state gluons is provided by the transverse momentum q_T of the electroweak boson. Momentum conservation implies that q_T is equal to the (vector) sum of the k_T of all emissions. The next sections describe how the emission of many gluons in the soft collinear limit is accounted for in modern predictions.

2.3.1 Transverse momentum resummation

As mentioned above, the cross section of equation (2.25) is divergent in the collinear region of small transverse momentum k_T and the soft region of small longitudinal momentum fraction $1 - z$ of the gluon. This divergence can be cured by considering the effect of many gluons on the q_T distribution. This is achieved by resumming the leading term for $k_T \rightarrow 0$ and $z \rightarrow 1$ to all orders in α_s , accounting for the emission of any number of soft and collinear gluons. The procedure is illustrated below by obtaining a simple resummed cross section in the double leading logarithm approximation (DLLA) [30].

The calculation starts by integrating away the unwanted z dependence. The dominant contribution is the logarithmic term appearing at the upper edge of the integration limits. Writing Q^2 the partonic center-of-mass energy, the available phase space for the emission is limited by the energy taken away by k_T : $z < 1 - k_T^2/Q^2$. Performing the integral yields:

$$\int_0^{1-k_T^2/Q^2} P_{q \rightarrow qg}(z) dz = \int_0^{1-k_T^2/Q^2} \frac{1+z^2}{1-z} dz \quad (2.28)$$

$$= 2 \log \left(\frac{Q^2}{k_T^2} \right) + \text{sub-leading terms.} \quad (2.29)$$

Plugging the result into equation (2.25), one obtains:

$$d\sigma_1 \approx \frac{\alpha_s}{2\pi} C_F \log \left(\frac{Q^2}{k_T^2} \right) \frac{dk_T^2}{k_T^2} d\sigma_0. \quad (2.30)$$

Integrating over k_T produces a squared logarithm, which is the dominant contribution for $k_T \rightarrow 0$. In the DLLA, all terms not containing this double logarithm are neglected.

At this point, two terms contribute to the total cross section: the no-emission cross section $d\sigma_0$ and the one-emission cross section $d\sigma_1$. The KLN theorem guarantees that when integrated over k_T , the divergent part of $d\sigma_1$ cancels with the first loop correction to $d\sigma_0$, $d\sigma_V$. The remainder is a simple correction of order α_s to $d\sigma_0$:

$$\int_0^{Q^2} \frac{d\sigma_V + d\sigma_1}{dk_T^2} dk_T^2 = \text{finite correction of order } \alpha_s. \quad (2.31)$$

In particular, the KLN theorem guarantees that these corrections contain no potentially large logarithm, which allows them to be neglected in the DLLA. Splitting the integral at q_T^2 , one obtains:

$$\int_0^{q_T^2} \frac{d\sigma_V + d\sigma_1}{dk_T^2} dk_T^2 \approx - \int_{q_T^2}^{Q^2} \frac{d\sigma_1}{dk_T^2} dk_T^2 + \mathcal{O}(\alpha_s) \quad (2.32)$$

$$\approx -\frac{\alpha_s}{2\pi} C_F \log^2 \left(\frac{Q^2}{q_T^2} \right) d\sigma_0, \quad (2.33)$$

where the finite contribution of $d\sigma_V$ above q_T has been neglected as well.

This result can be generalized to any number of emitted gluons. For n gluons, one obtains an expression where the prefactor of $d\sigma_0$ is simply raised to the n th power and divided by a factor $n!$. Adding together the contributions of any number of gluons, one obtains for the prefactor:

$$\Sigma(q_T^2) = \exp \left[-\frac{\alpha_s}{2\pi} C_F \log^2 \left(\frac{Q^2}{q_T^2} \right) \right]. \quad (2.34)$$

This function called the Sudakov form factor appears in many problems involving two scales — here q_T and Q .

The final resummed result in the DLLA is obtained by differentiating the Sudakov form factor with respect to q_T :

$$\frac{d\sigma_{\text{resum.}}}{dq_T^2} = \frac{d\Sigma(q_T^2)}{dq_T^2} d\sigma_0 = \frac{\alpha_s}{\pi} C_F \frac{1}{q_T^2} \log \left(\frac{Q^2}{q_T^2} \right) \exp \left[-\frac{\alpha_s}{2\pi} C_F \log^2 \left(\frac{Q^2}{q_T^2} \right) \right] d\sigma_0. \quad (2.35)$$

The spectrum predicted by equation (2.35) contains the main features of the physical distribution for $q_T \lesssim Q/2$. At very small q_T , the divergence of the one-parton cross section is tamed down by the exponential factor. The slope of the cross section is positive up to a maximum called the Sudakov peak located around $q_T \approx 5 \text{ GeV}$ at the LHC. Above this value, the cross section steadily decreases with

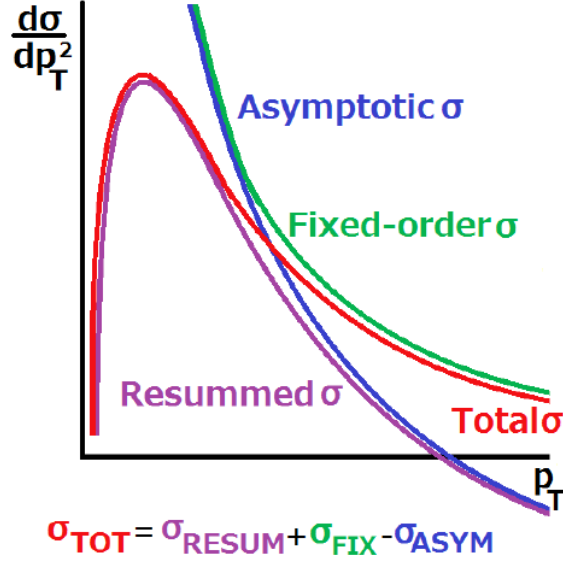


Figure 2.5: Illustration of the general behaviour the q_T spectrum as predicted by resummed calculations matched to fixed order.

increasing q_T and eventually becomes negative at the q_T integration limit used in equation (2.31), $q_T = Q$. In this region of large q_T , it is possible to “match” the resummed result with a fixed order prediction to obtain a well-behaved cross section for $q_T \rightarrow \infty$. This general behaviour is shown in figure 2.5, where the contributions from the resummed cross section σ_{resum} , the fixed order one σ_{fix} , and their matching σ_{asym} are also visible.

The result derived above has a number of shortcomings that limit its applicability. Firstly, all terms smaller than the logarithm have been neglected, while they bring important corrections both to the argument of the exponential and to the cross section, including a second large logarithm (not squared) that can be traced back to the sub-leading terms of equation (2.29). Another significant issue is that the q_T shifts induced by multiple gluons have been treated as uncorrelated, while in reality their vector sum can cancel. Taking this into account produces a non-zero cross section for $q_T \rightarrow 0$. Finally, the running of α_s has been completely ignored. In particular, at very small k_T gluons emissions are non-perturbative and a parameterization is needed. Solving these issues is beyond the scope of this basic introduction and leads to the complete leading order-leading logarithm (LO + LL) approximation. Extending the fixed order calculation beyond leading order introduces two logarithms for each additional order, of the form:

$$\alpha_s^n \log^m \left(\frac{Q^2}{q_T^2} \right) \frac{dq_T^2}{q_T^2}, \quad (2.36)$$

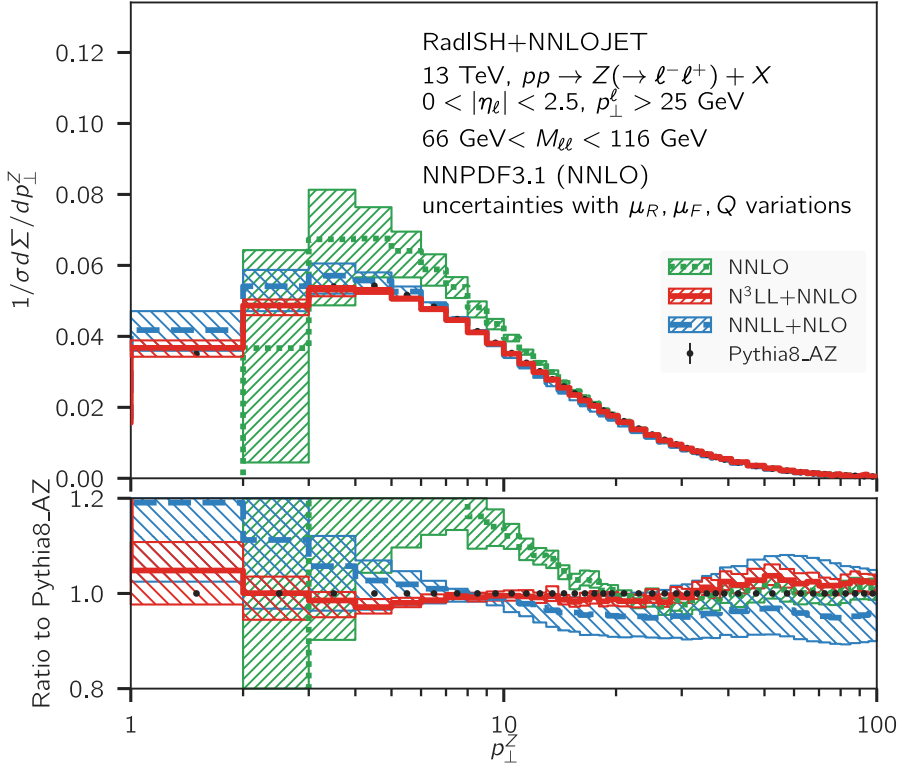


Figure 2.6: RADISH + NNLOJET predictions of the Z/γ^* boson transverse momentum distribution at $\sqrt{s} = 13$ TeV. Reproduced from [31].

with $m \leq 2n - 1$. In general, a resummation of logarithms with the largest m is called leading logarithm (LL) approximation and by including lower powers one obtains the NLL, NNLL, etc., predictions.

Perturbation theory cannot be applied to the region of small k_T — corresponding to large distances in the transverse plane. The breakdown can be overcome using a parameterization of the k_T distribution in the non-perturbative region. The necessary convolution between the distributions of the two incoming partons spreads the non-perturbative contribution in the q_T distribution, and it contributes significantly roughly up to the Sudakov peak. A more systematic treatment of the small- k_T region is obtained in small- q_T factorization, covered in section 2.4.

Figure 2.6 shows predictions of the Z/γ^* boson transverse at $\sqrt{s} = 13$ TeV performed in reference [31] by matching resummation calculated in the RADISH [32, 33] approach to a fixed-order calculation at order α_s^3 , achieving NNLO accuracy for the q_T distribution. The predictions are compared to a prediction that was tuned to reproduce a q_T measurement by the ATLAS Collaboration [34] (PYTHIA8_AZ). All predictions agree in the large- q_T region dominated by the fixed-order calculation, down to $q_T = 20$ GeV. For smaller values, the uncertainties in the NNLO

result blow up, while the matched NNLL and N³LL predictions remain stable and agree with the PYTHIA8 pseudo-data for $q_T \rightarrow 0$.

2.3.2 Parton showers

A very important class of predictions closely related to the resummation of large logarithms is achieved by means of “parton shower” algorithms that are able to add multiple parton emissions on top of events generated by a fixed-order calculation, thereby enhancing its logarithmic accuracy. The construction of the algorithm starts from a simple generalization of equation (2.25), swapping k_T^2 for a generic scale t that tends to zero in the collinear limit (for instance, $t = k_T^2$ or θ^2 , with θ the angle at which the gluon is emitted):

$$d\sigma_1 \approx \frac{\alpha_s}{2\pi} C_F P_{q \rightarrow qg}(z) dz \frac{dt}{t} d\sigma_{q\bar{q}}. \quad (2.37)$$

Starting from a scale t_0 , one can find the probability $\Delta(t_0, t)$ that no emission occurs between scales t_0 and t . It is given by the Sudakov form factor $\Delta(t_0, t)$ already encountered in equation (2.34), here written in terms of the scale t :

$$\Delta(t_0, t) = \exp \left[-\frac{\alpha_s}{2\pi} C_F \int_{t_0}^t \frac{dt'}{t'} \int_{z_-(t')}^{z_+(t')} dz P_{q \rightarrow qg}(z) \right], \quad (2.38)$$

where $z_{\pm}(t)$ encode the phase space boundaries. The Sudakov form factor written in this way can be interpreted as the probability that no gluon is emitted between scales t_0 and t . In this scheme, the probability $\mathcal{P}_{t_0}(t)$ that a gluon is emitted at scale t is given by differentiation with respect to t ,

$$\mathcal{P}_{t_0}(t) = -\frac{\partial}{\partial t} \Delta(t_0, t). \quad (2.39)$$

This expression is readily generalizable to include photon emissions or radiations off gluons, using the corresponding couplings and DGLAP splitting functions (2.19)–(2.21).

Parton showers use the probability distributions $\mathcal{P}_{t_i}(t_{i+1})$ to generate successive branchings at scales $t_0 > t_1 > \dots > t_{\min}$ on top of input events, implementing explicit momentum conservation at each step. The scale $t_{\min}$ at which recursion stops is typically of order 1 GeV and is justified by the breakdown of perturbation theory around this scale. As is made obvious by the parallel between the derivation of $\mathcal{P}$ and analytical resummation, in the limit of infinitely many events the first branching corresponds to leading logarithmic accuracy. Branchings at smaller scales resum additional logarithms that appear at larger orders in α_s . In this way, the logarithmic accuracy of a parton shower is better than a simple LL calculation.

When applied to initial state partons, the parton shower algorithm conflicts with the DGLAP evolution equations (2.22) since it effectively evolves the partons to smaller scales. This can be solved by weighting the argument of the exponential in the expression of the Sudakov form factor by a ratio of the appropriate PDFs for the considered splitting evaluated at t and t_0 . Another aspect of initial-state showers is that the transverse momentum of the emitted partons is to be interpreted as a k_T for the quarks at scale t_0 , i.e. the final state products should be boosted by that amount. At the other end of the evolution, the distribution of the parton k_T at the cutoff scale $t_{\min}$ is unknown and the parameters of an ansatz have to be adjusted to data [35]. The $\langle k_T(t_{\min}) \rangle$ obtained for proton-proton collisions is larger than the mass of the proton [36].

The parton shower algorithm is not limited to emissions around a single quark. It can be applied simultaneously to all initial and final-state particles and recursively to the partons emitted in branchings, exploiting all of the DGLAP splitting functions (2.19) to (2.21). The resulting emission graph is illustrated in figure 2.7 for an event with two quarks at the initial scale. Because the emitted partons are nearly collinear with their parents, they end up forming jets. When they reach the cutoff scale $t_{\min}$, partons can be clustered to form hadrons. When this final step is applied, parton shower codes allow the generation of physical events that can be further processed by simulating their interactions within a detector.

Since the parton shower algorithm is applied on top of another prediction, there is a risk of counting parton emissions twice: once in the underlying prediction and a second time in the shower. This is the case when the matrix element includes terms at several orders in α_s , for instance in NLO event generators or when several parton multiplicities are included. Double counting is avoided by “matching” the fixed order calculation to the parton shower. At NLO, this is achieved by a modification of the matrix element calculation (MC@NLO method [37]) or by vetoing certain radiations in the parton shower (POWHEG method [38]). In both cases, exactly one emission in the shower is effectively upgraded to full LO accuracy. Since other emissions are suppressed by additional α_s factors, they do not affect the formal accuracy of the fixed order prediction.

In order to generate fully exclusive events with particles as they would be seen by a detector, parton showers are typically supplemented with a description of other, perturbative and non-perturbative, effects. This includes among others the k_T distributions of initial-state partons at the cutoff scale, interaction of multiple partons, a description of the remnants of the interacting protons, hadronization, and decay of unstable particles. The effective physics models used for this purpose make extensive use of ad hoc parameters that need to be adjusted so that predictions match observations, a process called “tuning”. For each implementation of the parton shower algorithm, sets of parameter values (“tunes”) have been

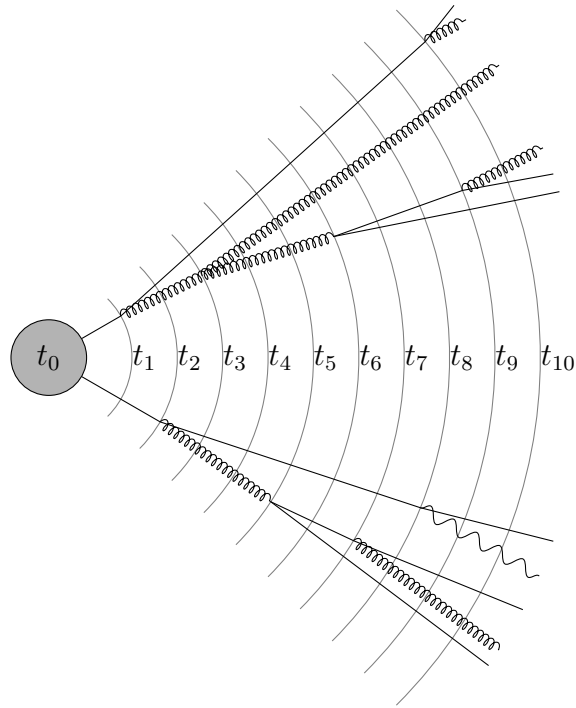


Figure 2.7: Illustration of the possible first ten branchings in the showering of an event with two final-state quarks. The distance from the hardest scattering (gray disc) corresponds to decreasing scales $t_0 > t_1 > \dots > t_{10}$. One branching occurs at each scale chosen by the algorithm.

obtained. Inputs of hadronization models can be extracted from e^+e^- collision data [36], while $p\bar{p}$ and pp data are used to tune multiple parton interaction and the initial state shower, for instance in references [36, 39, 40, 41]. Specialized tunes are also sometimes used to improve the description of specific processes [34, 42].

Commonly used implementations of the parton shower algorithm and associated effects are provided by PYTHIA [35, 43], HERWIG [44, 45], and SHERPA [46]. They all share the same underlying principle of the parton shower, even if they differ in the formalism and approximations used. Despite their intrinsic weaknesses of a formal accuracy so far limited to the leading logarithm and the need for tuning of their non-perturbative parameters, predictions based on parton showers matched to fixed-order calculations have been immensely successful in describing a wide variety of processes in electron-electron, electron-hadron and hadron-hadron collisions. They are routinely used at the LHC to produce simulated samples of processes of interest for searches and measurements.

2.3.3 Parton branching solution of DGLAP evolution

As already mentioned, there is a non-trivial interplay between the initial state parton shower and DGLAP evolution of the PDFs. It was hence proposed [47, 48] to perform the two evolutions simultaneously. Starting with a parameterization of the k_T and x distributions at a scale $t_{\min}$, they are evolved up to the scale of interest Q by solving the DGLAP equations using a parton shower algorithm, updating k_T at each step. A fit to HERA data was performed in reference [49], using a Gaussian initial k_T distribution with standard deviation $\sigma = 0.35 \text{ GeV}$ at $t_{\min} = 1.4 \text{ GeV}^2$. The x evolution was calculated at next-to-leading order in α_s .

The method sketched above is used to obtain a joint distribution in x and k_T at every factorization scale μ_F and for every parton species a , $f_a(x, k_T; \mu_F)$. Because they include a k_T dependence in addition to the traditional x dependence of the PDFs, these distributions are called “transverse momentum-dependent PDFs”, or TMDs for short. For the most part, they replace the initial-state parton shower and can be similarly matched with fixed-order calculations. This allows their combination with existing fixed-order Monte-Carlo generators.

2.4 Small q_T factorization

The two resummation methods described in the previous section require non-perturbative input in the region of small parton transverse momenta, $k_T \lesssim \Lambda_{\text{QCD}}$. It is natural to think that the missing piece includes the non-perturbative “intrinsic” motion of the partons inside the protons. Like the collinear momentum fraction x , this should be an intrinsic property of protons that can only be ex-

tracted from data (barring eventual calculations on the lattice). This motivates the insertion of the transverse momentum as a second parameter to the parton density functions, turning $f_a(x; \mu_F)$ into $f_a(x, k_T; \mu_F, \zeta)$. Like in the parton branching approach, these new functions are called transverse momentum-dependent distributions (TMDs). TMD factorization for Drell-Yan has been considered early on [50, 51]. More recently, several new definitions of the TMDs have been given (e.g. in [52, 53, 54, 55]) that were eventually shown to be equivalent [56]. See e.g. [57] for a recent review. In the following, the “CSS” formalism developed in reference [52] is used in the equations, but the discussion applies to all equivalent definitions of the TMDs. Like PDFs, TMDs defined in this way do not depend on the nature of the hard process.

TMD factorization theorems for the Drell-Yan process have been proven for $q_T^2 \ll Q^2 \ll s$. The results are often written in impact parameter space, using the vector $\mathbf{b}_T = (b_x, b_y)$ conjugate to transverse momentum. The TMD distribution functions $\tilde{f}_q$ depend on two scales μ and ζ used to regularize infrared, ultraviolet, and rapidity divergences. Up to corrections suppressed by powers of q_T/Q , the Drell-Yan cross section can be written as:

$$\frac{d^3\sigma}{dQ^2 dy dq_T^2 d\varphi} = \frac{8\pi^2\alpha^2}{9Q^2 s} \sum_q H_{q\bar{q}}(Q^2; \zeta) \int \frac{d^2\mathbf{b}_T}{(2\pi)^2} e^{i\mathbf{q}_T \cdot \mathbf{b}_T} \tilde{f}_q(x_q, b_T; \mu, \zeta) \tilde{f}_{\bar{q}}(x_{\bar{q}}, b_T; \mu, \zeta), \quad (2.40)$$

where φ is the azimuthal angle associated with $\mathbf{q}_T$ and, like in the collinear factorization case, the sum runs over all quark flavours. The hard factor $H_{q\bar{q}}$ encodes the process-dependent part of the calculation. It is similar to yet different from the hard scattering cross section $\hat{\sigma}_{q\bar{q}}$ appearing in collinear factorization: for instance, real emission graphs like (2.23) are absent. Because factorization is only valid in the Sudakov region $q_T^2 \ll Q^2$, TMD-based predictions must be complemented by resummed and fixed order calculations performed in collinear factorization if a prediction of the whole q_T spectrum is desired.

The TMD distributions are related to the collinear PDFs $f_a(\xi; \mu)$, which is often exploited to reduce the complexity of the extraction from data. The formal relation is written by expanding the TMDs as power series in b_T . The term in the series at order b_T^0 is the following:

$$\tilde{f}_q(x, b_T; \mu, \zeta) = \sum_a \int_x^1 \frac{d\xi}{\xi} \tilde{C}_{q/a}(x/\xi, b_T; \mu, \zeta) f_a(\xi; \mu) + \mathcal{O}(b_T^2 \Lambda_{\text{QCD}}^2) \quad (2.41)$$

The “hard coefficients” $\tilde{C}_{q/a}$ in the above expression are related to the DGLAP splitting functions $P_{a \rightarrow bc}(x/\xi, \mu_F)$ and can be calculated order-by-order in α_s .

Higher order terms in the b_T expansion (“power corrections”) are not known and need to be parameterized and extracted from data.

Since two scales are required when evaluating TMD distributions, their evolution is less straightforward than for collinear PDFs. The evolution between two points in the (ζ, μ) plane can be performed along a variety of paths, some of which enter the non-perturbative region of large b_T [58]. All such paths are equivalent when all orders in α_s are taken into account, but this is not the case when the series are truncated and the choice of evolution path can result in differences between TMD implementations.

The TMD evolution equations are different from the DGLAP equations and do not contain a mixing between flavours or different values of x . In the CSS formalism, they are given in terms of functions $\tilde{K}$, γ_q and γ_K calculable order-by-order in perturbation theory:

$$\frac{\partial \ln \tilde{f}_q(x, b_T; \mu, \zeta)}{\partial \ln \sqrt{\zeta}} = \tilde{K}(b_T; \mu), \quad (2.42)$$

$$\frac{d\tilde{K}(b_T; \mu)}{d \ln \mu} = -\gamma_K(\alpha_s(\mu)), \quad (2.43)$$

$$\frac{d \ln \tilde{f}_q(x, b_T; \mu, \zeta)}{d \ln \mu} = \gamma_q(\alpha_s(\mu)) - \frac{1}{2} \gamma_K(\alpha_s(\mu)) \ln \frac{\zeta}{\mu^2}. \quad (2.44)$$

The specific form taken by the TMD evolution equations differs between different approaches of factorization. Of course, if the relation (2.41) between TMDs and PDFs is used, DGLAP-style convolutions appear.

TMD distributions have been extracted from semi-inclusive deep inelastic scattering and Drell-Yan measurements in electron-proton and proton-proton data, respectively. Existing fits differ in the TMD formalism used, the order at which the perturbative pieces are calculated, and the data used for the extraction. Since the $\mathbf{b}_T$ integration region in equation (2.40) extends to the non perturbative regime $b_T \gtrsim 1 \text{ GeV}^{-1}$, the behaviour of the TMD at large b_T has to be parameterized. Various functional forms, often resembling a Gaussian distribution in the large- b_T limit, are used in the literature. Modern analyses find that TMDs defined in a way equivalent to the CSS formalism can be used without $\mathcal{O}(q_T/Q)$ corrections up to $q_T \lesssim 0.2 Q$ [59, 60].

Figure 2.8 shows the results of global fits of CSS TMDs calculated and evolved at NN(N)LO [60, 61, 62] as well as a comparison to the parton branching (PB) TMD fit at NLO [49]. The distributions are represented in momentum space by Fourier-transforming back the functions of equation (2.40). The momentum conjugate to b_T for a single TMD is written k_T . For the plot, Q is fixed to m_Z and x is chosen to represent the production of a Z boson in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ and $y = 0$. One notices a discrepancy between CSS TMDs in the

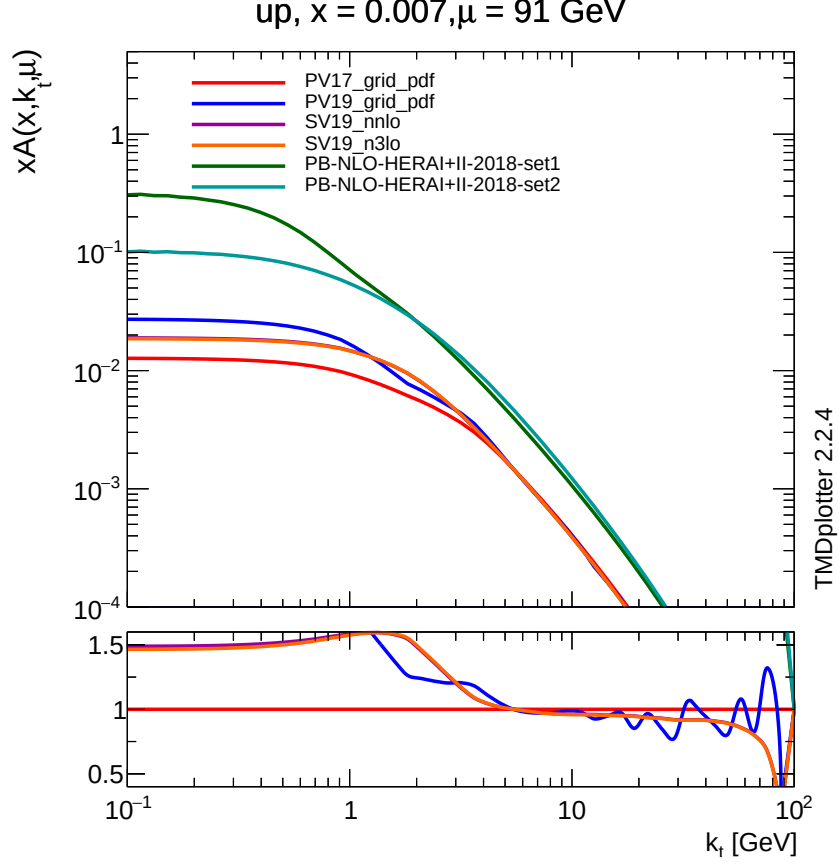


Figure 2.8: Central values of TMD distributions for the up quark evaluated at $x = 0.007$ and $Q = \mu = m_Z$, corresponding to the production of a central Z boson in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$. The different lines correspond to references [61] (PV17), [60] (PV19), [62] at NNLO and N³LO (SV19), and [49] (PB). The ratio to the PV17 prediction is shown in the lower panel. Plot made using TMDPLOTTER/TMDLIB [63, 64].

non-perturbative region and up to $k_T \sim 5 \text{ GeV}$, while they broadly agree at large k_T despite visible oscillations for one of them. In contrast, the PB TMDs have a vastly different normalization, indicating that they are not interchangeable with CSS TMDs as is.

2.5 Summary

The Drell-Yan process provides a probe of the dynamics of partons in the initial state of high-energy hadron-hadron collisions. The collinear factorization theo-

rem (2.18) allow the separation of the low and high energy parts of the interaction, using the non-perturbative parton distribution functions for the former and enabling the use of perturbative methods based on Feynman diagrams for the latter. Calculations at NLO, or NNLO in α_s are successful in predicting the Drell-Yan cross section and the invariant mass and rapidity distributions. It is also possible to merge predictions with different number of partons in the final state, reaching higher order in part of the phase space. Fixed order and merged prediction also provide accurate predictions for the high- q_T tail of the q_T -differential cross section.

The prediction of the q_T spectrum for $q_T \lesssim Q$ requires the resummation of multiple parton emissions to all orders in α_s , using a systematic expansion in powers of $\log Q/q_T$ instead of α_s . Analytic resummation of q_T for the Drell-Yan process is available up to the third logarithmic order, N³LL. It is also possible to resum parton emissions using the parton shower algorithm, formally including the leading logarithm only but in practice providing a much better approximation. Both analytical resummation and the parton shower approximation require the inclusion of non-perturbative elements to describe partons at small scale and/or small transverse momentum, sometimes identified with the “intrinsic” transverse momentum of the partons in the colliding hadrons.

The apparition of non-perturbative elements in q_T resummation has prompted the inclusion of an additional k_T dependence in the PDFs. Two frameworks using such transverse momentum-dependent distributions have been introduced above. In the parton branching approach, a parton shower algorithm is hooked into the DGLAP evolution of the PDFs. In small- q_T (CSS) factorization, a dedicated theorem allows predictions of the q_T spectrum up to $q_T \approx 0.2 Q$. Both parton branching and CSS TMDs are expected to improve the description of the q_T distribution at small q_T with respect to other approaches of q_T resummation.

In order to obtain a valid prediction over the whole q_T spectrum, several techniques have been combined. This can be achieved by matching a q_T -resummed or parton shower calculation to a fixed-order cross section, avoiding double counting and dead regions. Existing techniques to match parton showers to fixed order are also used for parton branching TMDs. For CSS TMDs, a matching against predictions in collinear factorization is expected to be possible, which has to the knowledge of the author not been achieved so far.

Chapter 3

Experimental setup

The work presented in this thesis is performed in the context of the proton-proton collisions performed at the Centre Européen de Recherche sur le Noyau (CERN) using the Large Hadron Collider (LHC). The accelerator is described in the first section of this chapter, where the important concepts of luminosity and pileup are also introduced. An introduction to the Compact Muon Solenoid detector, installed at one of the interaction points of the LHC to study the outcome of the collisions, is provided in the second section.

3.1 The Large Hadron Collider

The Large Hadron Collider [1] is a circular proton-proton, proton-lead and lead-lead collider located in the former tunnel of the LEP accelerator near Geneva, Switzerland. The LHC is located underground at a depth between 45 and 175 m below the surface. It has a circumference of 27 km and is not entirely circular due to the straight tunnel sections around the location of former LEP experiments. Eight major access points are distributed along the tunnel, four of which are used for experiments: ALICE, ATLAS, CMS and LHCb. The LHC is connected to the Super Proton Synchrotron (SPS) and the rest of the CERN accelerator complex through two tunnels, each approximately 2.5 km long. An overview of the LHC complex in the Geneva region is shown in figure 3.1.

The two beams of the LHC circulate in two beam pipes that meet each other at the four collision points. Between 2015 and 2018 (“Run 2”), protons in each beam of the LHC had an energy of 6.5 TeV, corresponding to the largest proton-proton center-of-mass energy ever achieved at a collider: $\sqrt{s} = 13$ TeV. During the 2016 data-taking period [65] considered in this thesis, each beam was made of up to 2220 bunches of around 100 billion protons. The nominal bunch separation was 25 ns. 146 days were dedicated to physics operation with stable beams 49 % of the



Figure 3.1: Aerial picture of the Geneva region, with the LHC indicated in yellow.

time.

3.1.1 Injection chain

The LHC uses several pre-accelerators to prepare and accelerate the beams, part of the CERN accelerator complex sketched in figure 3.2. The source of the protons is a simple hydrogen bottle. The gas is turned into a plasma by heating it and electrons are stripped by applying an electric field. The 100 keV protons are then accelerated by the LINAC2 linear accelerator up to 50 MeV. Several shots from the LINAC2 are accumulated in the four rings of the Proton Synchrotron Booster until around 10^{12} protons per ring is reached. The Booster then accelerates the protons to 1.4 GeV and injects the four bunches into the Proton Synchrotron (PS). The PS accumulates eight bunches and accelerates them further to reach 26 GeV, also splitting them to obtain 48 bunches of 10^{11} protons.

Before they can be injected in the LHC, the protons are accelerated to 450 GeV in the repurposed Super Proton Synchrotron (SPS). The nominal number of

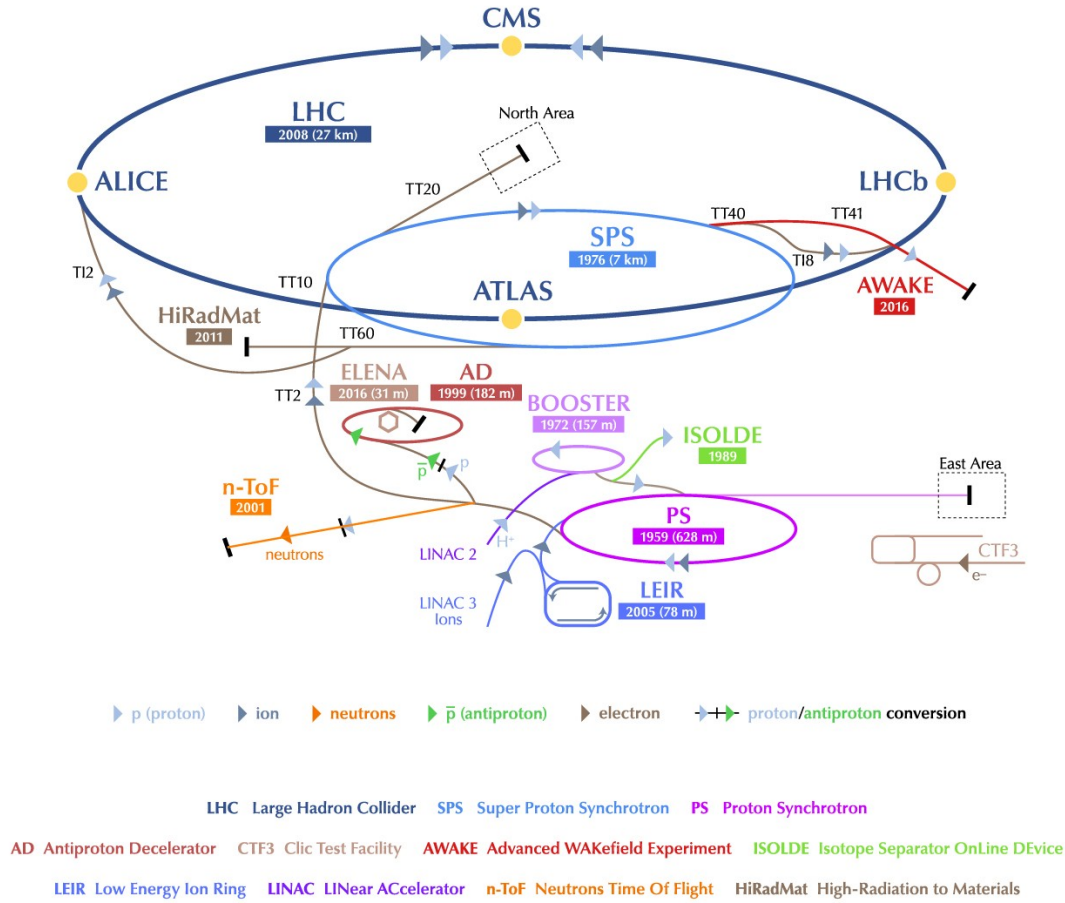


Figure 3.2: The CERN accelerator complex in January 2016 [66].

bunches that can be accelerated by the SPS is 288, but in 2016 a gas leak forced a two-third reduction of the SPS capacity and only 96 bunches at a time were accelerated. Each train of 96 bunches is in turn injected into the LHC. When the desired number of bunches is reached, the LHC accelerates them to 6.5 TeV and starts collisions after a brief stabilization and adjustment period. In 2016, the average time between a beam dump and stable beams for collisions was of 7.1 h, with stable beams kept for 11.2 h on the average.

3.1.2 Luminosity

A key metric for collider experiments is the instantaneous luminosity $d\mathcal{L}/dt$ provided by the accelerator at the interaction point, which controls the number N of

interactions in a time interval t_1 to t_2 through the formula:

$$N = \sigma \int_{t_1}^{t_2} \frac{d\mathcal{L}}{dt} dt, \quad (3.1)$$

where σ is the cross section of the process of interest. Knowing the frequency f of bunch crossings (which depends on the number of bunches and their revolution frequency), the luminosity can be written using the per-bunch-crossing luminosity $\mathcal{L}_{bc}$:

$$\frac{d\mathcal{L}}{dt} = f \mathcal{L}_{bc}. \quad (3.2)$$

The per-bunch-crossing luminosity on the number of protons n_1 and n_2 in each of the two colliding bunches and the effective area of overlap between the bunches A_{eff} , through:

$$\mathcal{L}_{bc} = \frac{n_1 n_2}{A_{\text{eff}}}. \quad (3.3)$$

The effective area can be measured in situ by moving the beams with respect to each other and monitoring the rate of interactions. In the van der Meer method [67, 68] used at the LHC, one assumes that the two-dimensional bunch density profiles $f_k(x, y)$ (with $k = 1, 2$ and x, y perpendicular to the beams) factorize as:

$$f_k(x, y) = f_{k,x}(x) f_{k,y}(y). \quad (3.4)$$

The x and y profiles are obtained by moving the beams in the corresponding direction. The effective area is then calculated as:

$$\frac{1}{A_{\text{eff}}} = \iint dx dy \frac{1}{f_1(x, y) f_2(x, y)} = \int dx \frac{1}{f_{1,x}(x) f_{2,x}(x)} \int dy \frac{1}{f_{1,y}(y) f_{2,y}(y)}. \quad (3.5)$$

Van der Meer scans of the beam profiles provide an absolute reference for the calibration of the luminosity. In 2016, they were performed in three dedicated runs at the beginning of the year. The integrated luminosity is extrapolated from the van der Meer scans by monitoring the rates of certain observables during data taking [69]. Figure 3.3 shows the instantaneous luminosity delivered by the LHC to ATLAS and CMS in 2016 as a function of time.

At the LHC interaction points within ATLAS and CMS, the per-bunch-crossing luminosity $\mathcal{L}_{bc}$ is large enough for several proton-proton interactions to occur in each bunch crossing, a property known as “pileup”. Running with pileup allows the observation of rarer processes at the cost of the added difficulty of separating the collision of interest from spurious ones happening in the same bunch crossing. In 2016, an average of 23 interactions per bunch crossing was observed by CMS, with some events reaching twice this value. The pileup distribution is shown in figure 3.4

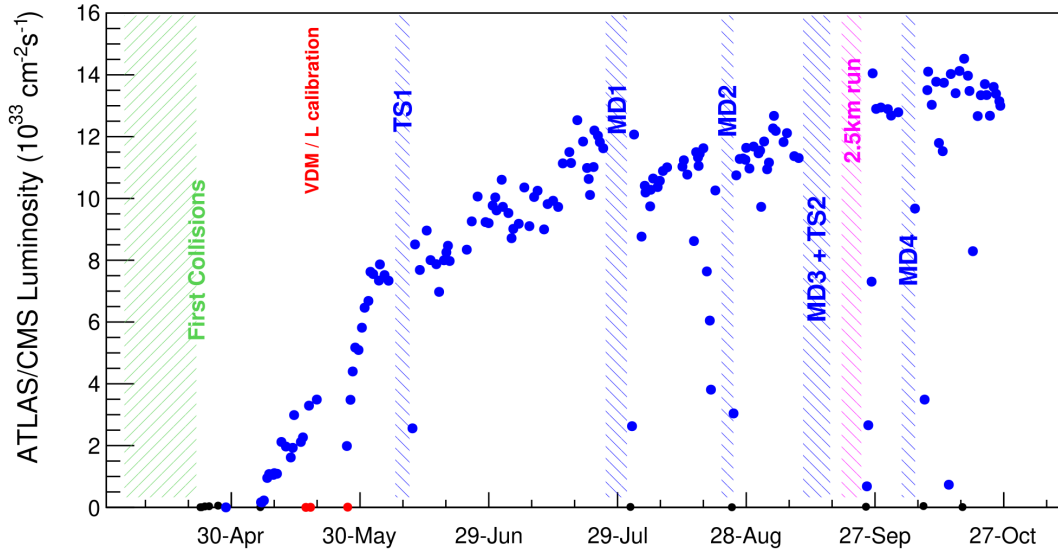


Figure 3.3: Peak instantaneous luminosity delivered to ATLAS and CMS for each run in 2016 [65]. The three special runs for van der Meer scans are shown in red.

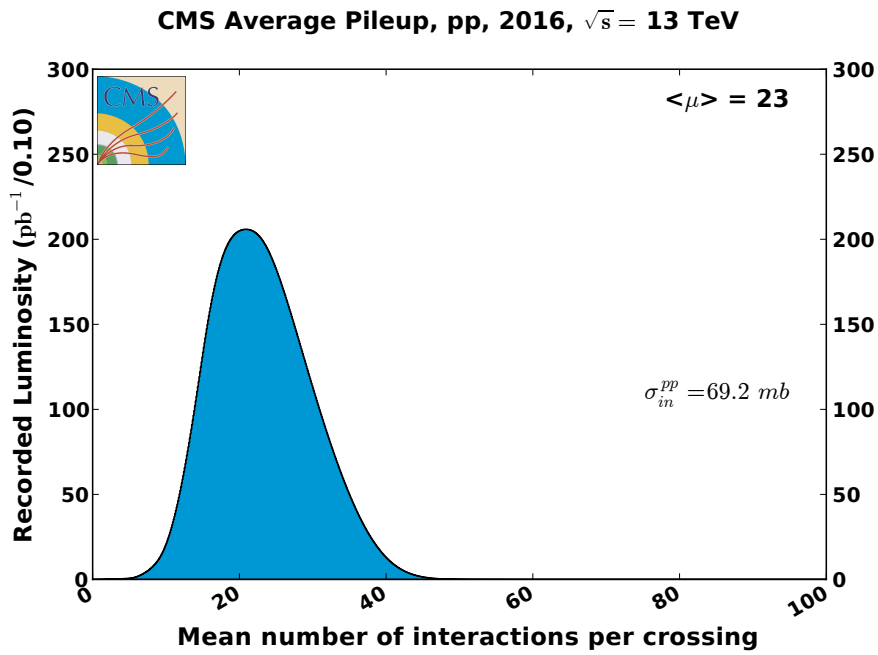


Figure 3.4: Pileup distribution observed by CMS in 2016, assuming a minimum bias cross section of 69.2 mb [70].

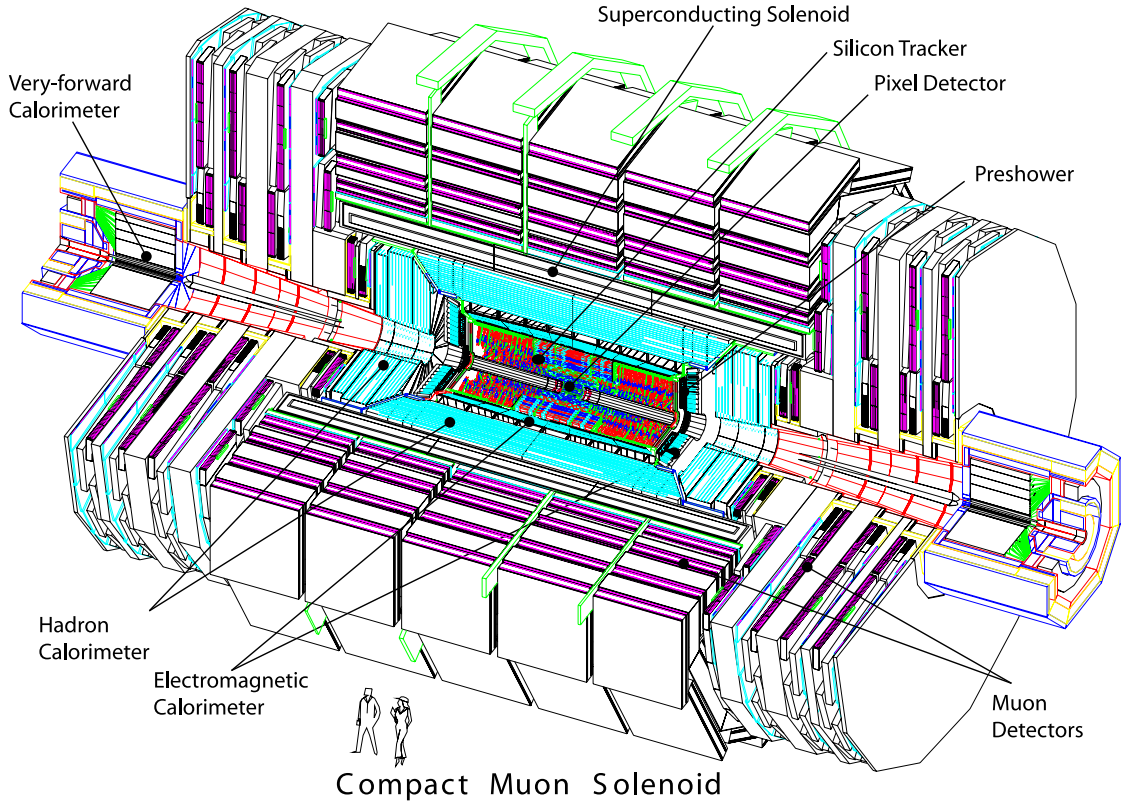


Figure 3.5: General layout of the CMS detector [4]. The beams enter the detector on the left and right of the drawing, and cross in the middle.

3.2 The CMS detector

The Compact Muon Solenoid (CMS) detector [4] is located at interaction point 5 of the LHC, about 100 m underground. CMS was designed to provide robust detection and identification of most particle types produced in proton-proton collisions, including indirect detection of weakly interacting particles (neutrinos or more exotic states) through missing transverse momentum in the transverse plane and tagging of the short-lived b quark and τ lepton. CMS operates in a high pileup environment in order to permit the discovery of new particles. This programme requires the use of a hermetic detector with a wide angular coverage, excellent tracking and vertexing performance, and extensive calorimetry. As indicated in its name, CMS puts a special emphasis on the identification and reconstruction of muons.

A central element in the design of the CMS detector is its large superconducting solenoid magnet that provides a magnetic field of 3.8 T in and around the interaction region, giving sufficient bending power to accurately measure the momentum

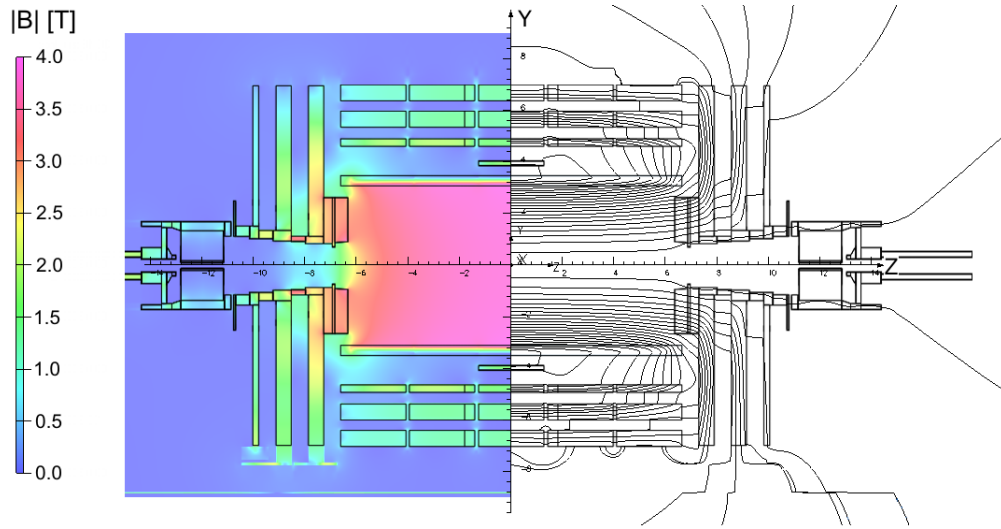


Figure 3.6: Map of the magnetic field induced by the CMS superconducting solenoid [71]. The left part of the picture shows the field strength and the right part the field lines.

and charge of particles up to several TeV. The detector is organized in concentric cylinders around the beam pipe (“barrel”) closed by flat disks (“endcaps”) at each end. The innermost layers, encountered first by particles produced in collisions, consist in a silicon-based tracker, followed by the electromagnetic and hadronic calorimeters, the solenoid magnet, and four layers of muon detectors. This general organization is illustrated in figure 3.5; the following sections describe the key features of each part of the detector.

3.2.1 Superconducting solenoid magnet

The central element of the CMS detector is its superconducting solenoid magnet, located between the hadron calorimeter and the muon systems. The coil is made of four layers of NbTi conductor reinforced with an aluminium core. It is 12.5 m long, with a diameter of 6.3 m and a thickness of 31.2 cm. Outside the magnet, the field is guided by a 10 000 t iron return yoke in which the intensity of the field reaches 1.8 T.

Figure 3.6 shows a map of the magnetic field produced by the solenoid. The interaction region, the tracker and most of the calorimeter system are immersed in a nearly uniform 3.8 T magnetic field. The iron return yoke is visible outside of this region. The quasi-uniformity of the magnetic field greatly simplifies the reconstruction of the trajectory of charged particles, for which an arc of a circle is a good approximation.

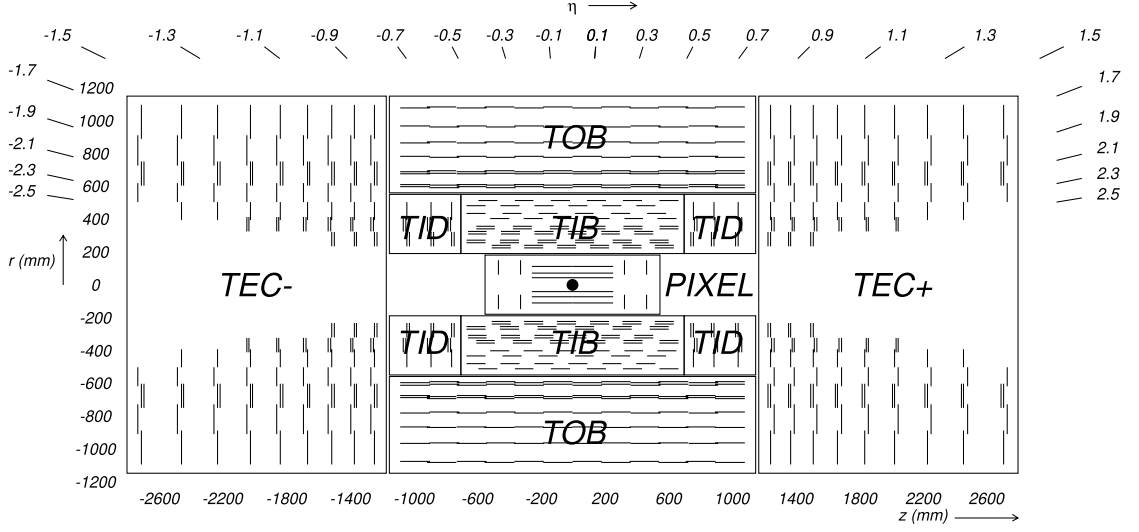


Figure 3.7: General layout of the CMS tracker [4]. The interaction point is indicated by the dot at the center of the detector.

3.2.2 Inner tracker

The CMS tracker was designed to provide an accurate measurement of several hits along the trajectory of charged particles produced in the collisions, with 1000 of them entering the tracker volume for each event at design luminosity. A high granularity and low occupancy is required to allow the separation of particles from pileup collisions. In addition, the tracking system has to be able to measure displaced vertices with micrometer precision to permit the tagging of b jets and hadronic decays of τ leptons. These requirements were met by the use of a full-silicon tracker, for the first time at a collider experiment.

The innermost three layers of the tracker, critical for the reconstruction of primary and secondary vertices as well as for track seeding, consist of pixel detectors with a pixel size of $100 \times 150 \mu\text{m}^2$. In the outer layers, strip sensors with a pitch of 80.5 to $205 \mu\text{m}$ are used. The design meets the resolution, occupancy, and radiation hardness requirements identified before the start of the LHC.

Figure 3.7 shows the layout of the sensors of the CMS tracker in the (z, r) plane. The 1440 pixel modules are grouped in four partitions that can be operated independently: the two halves of the barrel (BPIX $\pm$) on each side of the collision point and the two endcaps (FPPIX $\pm$). The 15148 strip modules are similarly grouped in four partitions covering the inner barrel and dees (TIB/TID), the outer barrel (TOB), and each of the two endcaps (TEC $\pm$), respectively. The spatial layout of the tracker is such that particles produced in the interaction region with a pseudorapidity of up to $|\eta| < 2.5$ cross between 12 and 14 modules.

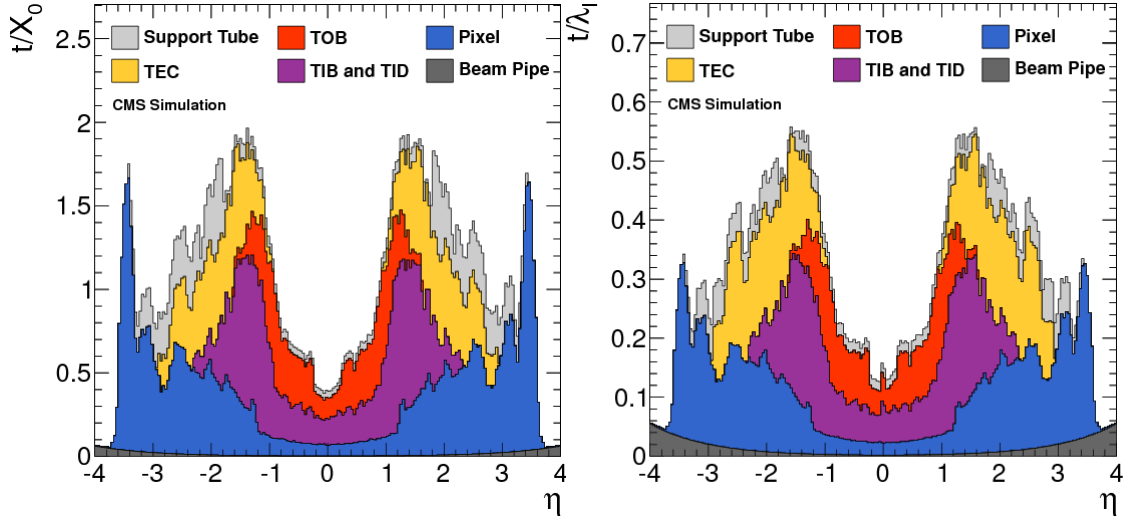


Figure 3.8: Simulated material budget of the CMS tracker, in units of radiation lengths (left) and nuclear interaction lengths (right) [72].

Interactions in the tracker material causes changes of direction through multiple scattering, bremsstrahlung, and photon conversions. Figure 3.8 shows that the tracker is the main contributor to the material budget in front of the calorimeters. The smallest amount of material is found in the central region, $|\eta| < 1$. It is up to four times larger at more forward pseudorapidities.

In 2015 and the first half of the 2016 data taking period, the tracker was affected by an efficiency loss in the strip detectors. The issue was eventually found to be rooted within the APV25 chips used for readout in that part of the detector. In normal operation, the feedback capacitor of the pre-amplifier of the APV25 chips discharges after a signal is recorded, which can take several bunch crossings. At the start of Run 2, improvements in the cooling system allowed to decrease the operating temperature of the tracker from 4°C to -15°C [73] to lower the impact of radiation damage caused by the larger foreseen instantaneous luminosity. This change had the adverse effect of increasing the discharge time of the feedback capacitor, which together with the higher occupancy caused by the increased instantaneous luminosity contributed to maintaining the pre-amplifier close to saturation, causing a luminosity-dependent hit inefficiency. Once the problem was understood, it was solved by lowering the pre-amplifier voltage feedback (VFP) starting with global run 278 769 (at the end of run F). Following this change, the hit efficiency in the inner layers was almost 100 % regardless of the instantaneous luminosity [73]. Figure 3.9 shows the signal-to-noise ratio in the first layer of the TOB before and after the fix.

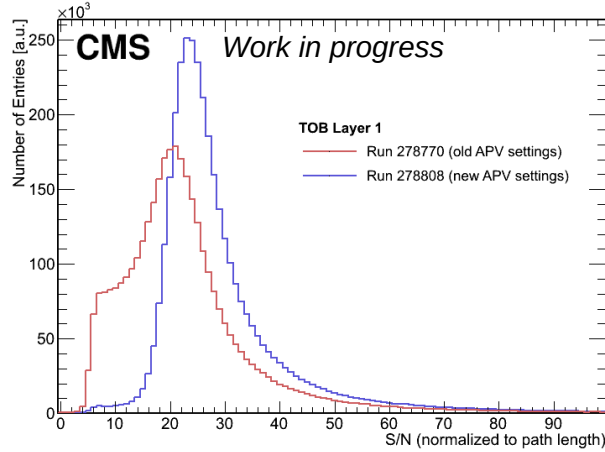


Figure 3.9: Signal-to-noise ratio in the first layer of the TOB before (red) and after (blue) the VFP change [73].

3.2.3 Electromagnetic calorimeter

The task of the electromagnetic calorimeter (ECAL) of CMS is to precisely measure the energy of photons and electrons. This is achieved by absorbing their energy in crystals that return it by scintillating in the visible spectrum. Lead tungstate (PbWO_4) crystals were chosen for the scintillator material owing to their high density, short radiation length and small Molière radius. This allows for a finely segmented, compact calorimeter with good shower containment properties. Lead tungstate has the disadvantage of producing a low light output of about 4.5 electrons per MeV at 18°C , subject to large variations with temperature. Thermal insulation and an extensive water cooling system maintain the crystals at a constant $(18.00 \pm 0.05)^\circ\text{C}$. The crystals and electronics of the ECAL provide a fast response to incoming particles, with 85% of the light collected and digitized in 25 ns, allowing its use in the trigger system.

The ECAL barrel covers central pseudorapidities, up to $|\eta| < 1.479$. It is made of 61 200 crystals with a segmentation in $\phi \times \eta$ of approximately 0.0172×0.0172 , corresponding to one degree in ϕ . The radiation length of the ECAL barrel crystals is $25.8 X_0$ for a crystal length of only 23 cm. The light emitted by each crystal is collected and amplified in two avalanche photodiodes chosen to obtain an average gain of 50, powered by custom power supplies capable of maintaining a stable voltage of up to 500 V with a precision better than 20 mV. The average signal from the photodiodes is pre-amplified, shaped and then amplified and digitized. The amplification and digitization steps are performed in parallel by three amplifiers with gains 1, 6, and 12 connected to three 12-bit analog-to-

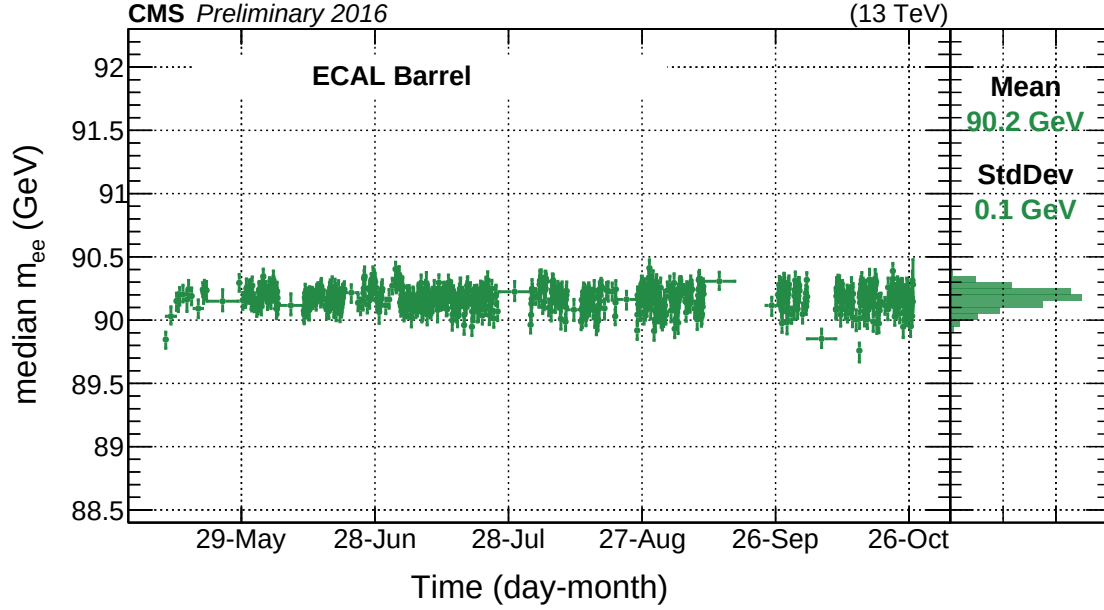


Figure 3.10: Time stability of the mass of Z bosons reconstructed from electron pairs in 2016 [74].

digital converters (ADC). The largest value from a non-saturated ADC is selected, allowing for readout with a high dynamic range.

The endcaps, covering $1.479 < |\eta| < 3.0$, are made of four half-discs (“dees”). Each dee is covered with 3662 crystals arranged in a square lattice. Endcap crystals have a depth of 22 cm, corresponding to a radiation length of $24.7 X_0$. The light from each crystal is collected and amplified in a single vacuum phototriode connected to the same high dynamic range readout electronics as in the barrel.

In front of each endcap of the ECAL is a “preshower” sampling calorimeter that covers the pseudorapidity range $1.653 < |\eta| < 2.6$. The preshowers are made of two layers of lead, 2 and 1 X_0 thick respectively, for a total thickness of 20 cm. Silicon strip sensors with a pitch of 1.9 mm are placed after each layer, providing fine-grained information about the beginning of the electromagnetic shower that helps identify photon pairs from π^0 decays and improves the resolution in the position of electrons and photons hitting the calorimeter.

The exposure of the PbWO_4 crystals of the ECAL to radiation degrades their optical properties. The transparency of the calorimeter is constantly monitored by injecting laser pulses into the crystals, providing a scan of the complete ECAL in about 30 minutes. The observed time-dependent variation of the transparency is accounted for in the calibration in order to achieve a stable response over time, as illustrated in figure 3.10.

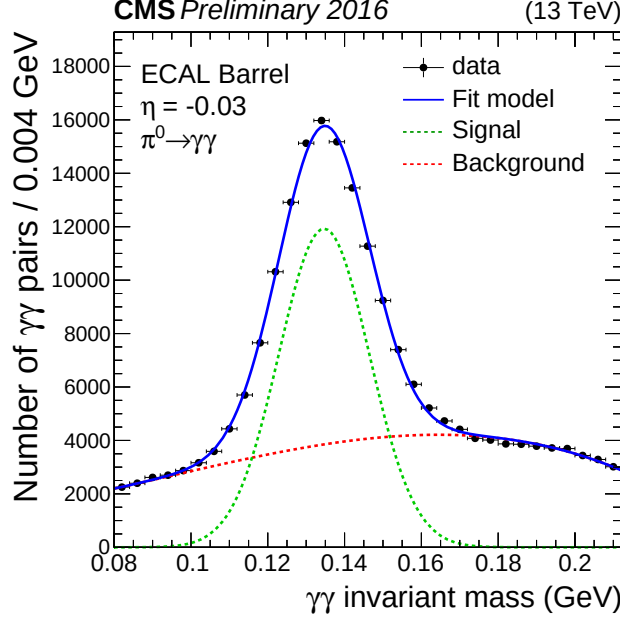


Figure 3.11: Invariant mass of photon pairs detected in the ECAL, with one photon in the barrel at $\eta = -0.03$ and a mass of the pair around the mass of the π^0 [74].

The factors affecting the energy resolution of the ECAL can be grouped in three categories:

- Stochastic effects S , scaling with energy like $\sqrt{E}$, are related to statistical fluctuations of the number of secondary particles in the shower and the number of photoelectrons produced by the photodetectors, which are proportional to the energy of the incoming particle.
- Noise effects N , independent of energy, originate in the electronics noise and contribution of pileup particles.
- “Constant” effects C , scaling linearly with energy and dominant at high energy. In the ECAL, they originate mainly from longitudinal non-uniformity of the crystals, intercalibration errors and the small shower leakage in the back of the crystals.

The relative energy resolution of the ECAL was measured in test beams before the start of the LHC and found to be, for E in GeV:

$$\frac{\Delta E}{E} = \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \oplus C \approx \frac{2.8\%}{\sqrt{E}} \oplus \frac{12\%}{E} \oplus 0.30\%. \quad (3.6)$$

Figure 3.11 shows an example of the typical invariant mass resolution achieved for photon pairs around the mass of the π^0 in 2016 data.

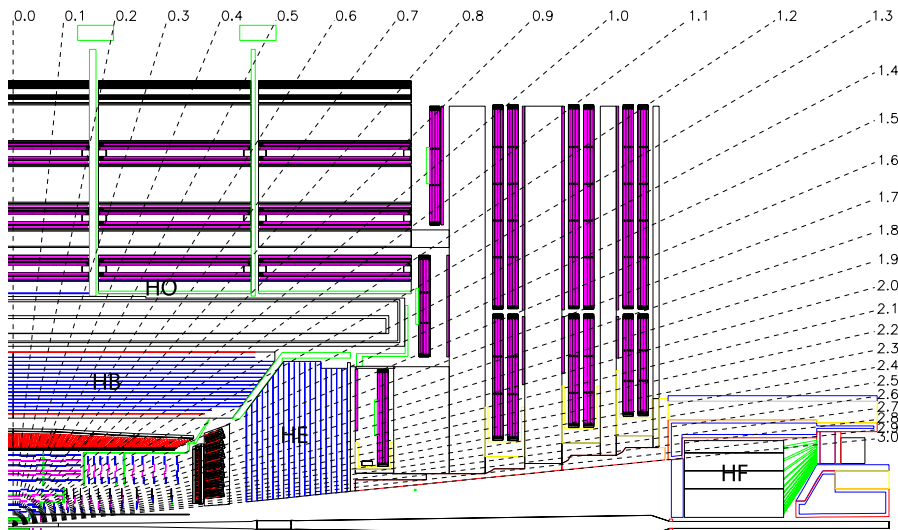


Figure 3.12: Transverse view of the CMS experiment, showing the location of the HCAL [4].

3.2.4 Hadron calorimeter

Located just behind the ECAL, the CMS hadron calorimeter (HCAL) is used to identify signal hadrons and hadronic jets and to measure their energy. It is divided in four regions: the barrel HB and outer barrel HO (covering the region $|\eta| < 1.3$), the endcaps HE ($1.3 < |\eta| < 3$) and two forward calorimeters HF ($3 < |\eta| < 5.2$). The location of the HCAL within the CMS detector is shown in figure 3.12.

The HB and HE are sampling calorimeters with an absorber made of brass plates with plastic scintillator active layers. The brass alloy used has a density of 8.53 g/cm^3 , a radiation length of 1.49 cm and an interaction length of 16.42 cm . Each brass plate has a thickness of about 5 cm . In the barrel, steel plates are used instead in the front and in the back to support the structure mechanically. The total absorber thickness of the HB is 5.82 interaction length at $\theta = 90^\circ$ and increases with θ like $1/\sin \theta$ to reach 10.6 interaction lengths at $|\eta| = 1.3$. The ECAL adds about one interaction length. In the endcaps, the thickness of the plates is about 8 cm for a total of about 10 interaction lengths.

The scintillation light produced in the active layers is collected and extracted using optical fibers. The scintillator is divided in tiles with independent readout, providing a granularity of $\Delta\phi \times \Delta\eta = 0.087 \times 0.087$. The segmentation in HE is between 5 and 10° (0.087 to 0.17 radians) in ϕ depending on $|\eta|$ and between 0.9 and 0.1 in η . This segmentation is illustrated in figure 3.13.

The shower containment of HB is not sufficient for an accurate measurement of hadron energies, especially around $\theta = 0$. The situation is improved by the

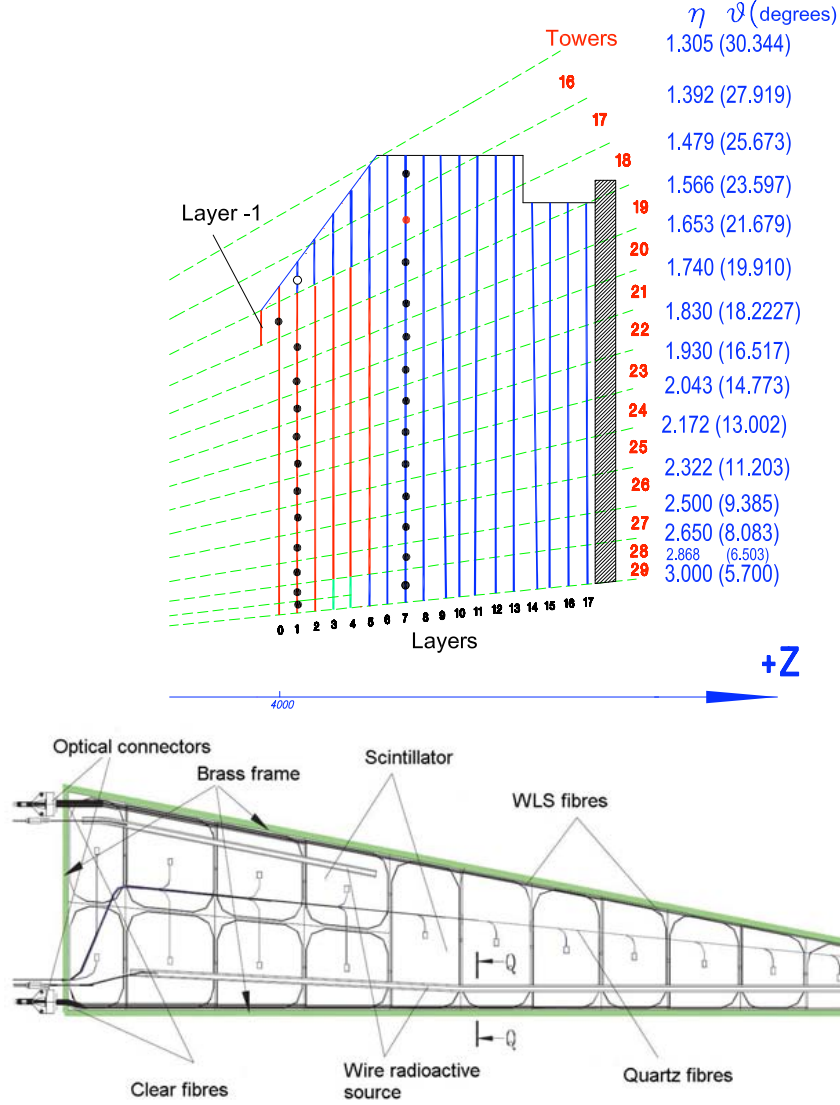


Figure 3.13: HCAL endcap segmentation [4]. Top: transverse view of one half of an endcap, showing the subdivisions in η and z . Bottom: front view of one "tray" of HE scintillator as seen from the interaction point. The ϕ and η segmentation is visible as well as the readout (WLS) fibres.

addition of one or two active layers outside the solenoid, whose material acts as an absorber with a depth of $1.4/\sin\theta$ interaction lengths. The “outer HCAL” HO uses 1 cm-thick plastic scintillator plates based on the same principle as in the barrel and endcap. This structure is doubled for $|\eta| < 0.307$, with 20 cm of iron between the two active layers, ensuring a total depth of over 11 interaction lengths for the complete calorimeter system over all of the barrel and endcap (except in the transition region). The $\phi \times \eta$ segmentation in HO is 0.087×0.087 as in the barrel, but coverage is limited by the mechanical and service structures present around the solenoid.

The last parts of the HCAL are the forward calorimeters HF. The harsh radiation environment and large rate found above $|\eta| = 3$ called for a design different from HB and HE. The HO absorber is made of steel, in which are inserted quartz fibers running parallel to the beam pipe and arranged in a $5\text{ cm} \times 5\text{ cm}$ grid in the transverse plane. Half of the fibers run through the entire depth of the absorber, while the other half starts after 22 cm. This allows a coarse separation between hadronic and electromagnetic showers. For readout, the fibers are grouped in “towers” with a granularity of $\Delta\phi \times \Delta\eta = 0.175 \times 0.175$ except in the innermost and outermost rings.

All sections of the HCAL are capable of providing fast information to the trigger system, which is used for triggers based on the presence of high-energy jets or large missing transverse energy, distinguish between electrons and jets, and for isolation requirements in muon triggers.

3.2.5 Muon detectors

Muon reconstruction was of central importance in the design of CMS. The muon detectors are located between and just outside the magnet return yokes both in the barrel and the endcaps, covering the range $|\eta| < 2.4$. The large covered area of about $25\,000\text{ m}^2$ lead to the choice of three technologies all based on gas detectors: drift tube (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC). Their layout within CMS is shown in figure 3.14.

Drift tubes are used exclusively in the barrel region where the signal and background rates are low and the magnetic field is small. They are organized in four stations, all of which contain chambers with four layers of DT measuring the azimuthal angle of the hits. Each of the first three stations also contain four layers providing measurements of the z coordinate. Each layer is made juxtaposed rectangular tubes mechanically supported by a honeycomb structure (see figure 3.15). The tube cross section is $13 \times 42\text{ mm}^2$, with wire lengths ranging between two and three meters depending on the orientation and location of the chamber. Drift tubes provide a good timing resolution and are capable of associating the measured hits with the correct bunch crossing for trigger.

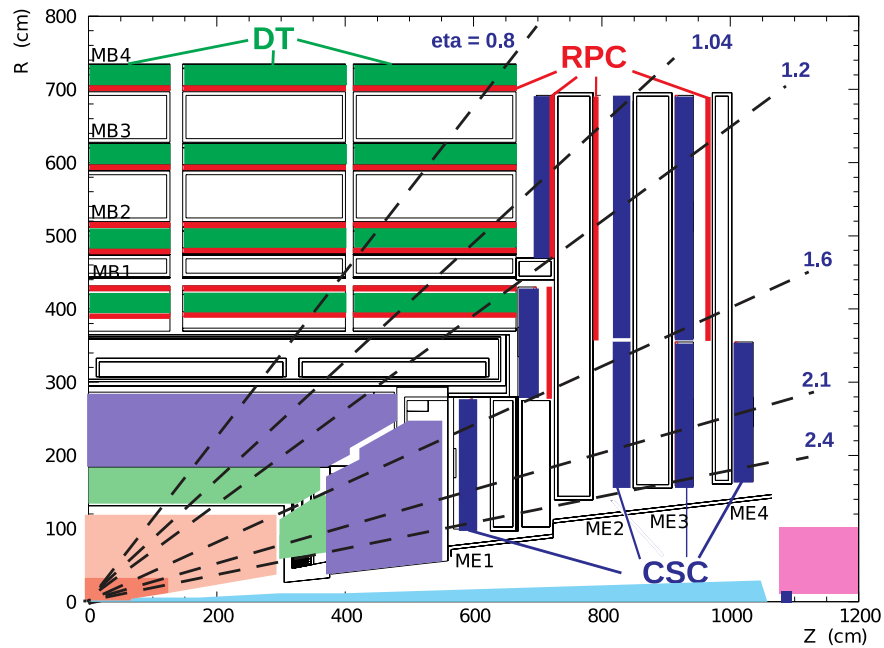


Figure 3.14: Transverse view of the CMS muon system [75].

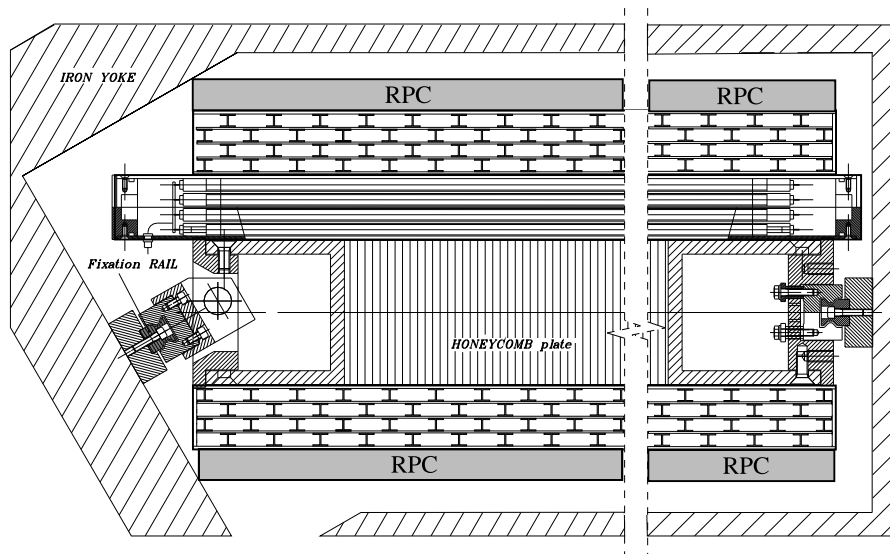


Figure 3.15: Structure of a DT chamber seen in the (r, ϕ) plane [4]. From top to bottom, one see a RPC chamber, four DT layers with the wires along the z axis (providing measurements of ϕ), eight layers perpendicular to the beam with the support structure in the middle, and a second RPC chamber.

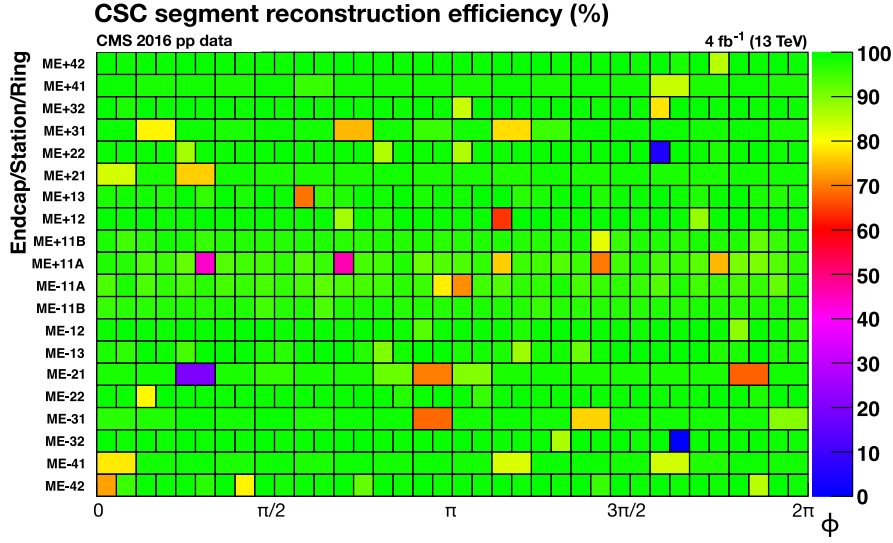


Figure 3.16: Track segment reconstruction efficiency of each CSC chamber in 2016 [76].

The cathode strip chamber technology is used in the endcaps. 468 trapezoidal chambers are used to provide continuous ϕ coverage in the region $1 < |\eta| < 2.4$. Each chamber consists of 7 layers of radial cathode strips separated by 6 layers of anode wires running in the perpendicular direction, providing a three-dimensional determination of the location of hits. There are about 220 000 strips and 180 000 anode wire channels in the whole system. Each CSC records several hits that are used to reconstruct a local muon track segment. Except for a few bad chambers, this process is almost 100 % efficient as seen in figure 3.16. Like the DT, the CSC provide timing information, with a comparable resolution.

The resistive plate chambers supplement the DT and CSC with higher-resolution timing measurements for $|\eta| < 1.6$, of the order of a few nanoseconds. Six layers of RPC are glued to DT chambers in the barrel region and three layers next to the CSC in the endcaps. Each RPC consists of two gas volumes separated by a plane of readout strips providing rather coarse location information. The gas volumes are operated in avalanche mode, with resistive bakelite plates between the electrodes and the gas volume limiting the propagation of the discharges. A schematic view of the main components of an RPC chamber is shown in figure 3.17.

Together, the muon detectors provide a large number of points along the trajectory of muons. Locally, the DT and CSC electronics reconstructs “segments” from hits in a single station. They are then matched together to reconstruct tracks and measure the transverse momentum of muons in near real time, providing information to the trigger system. Offline, combining measurements from all muon

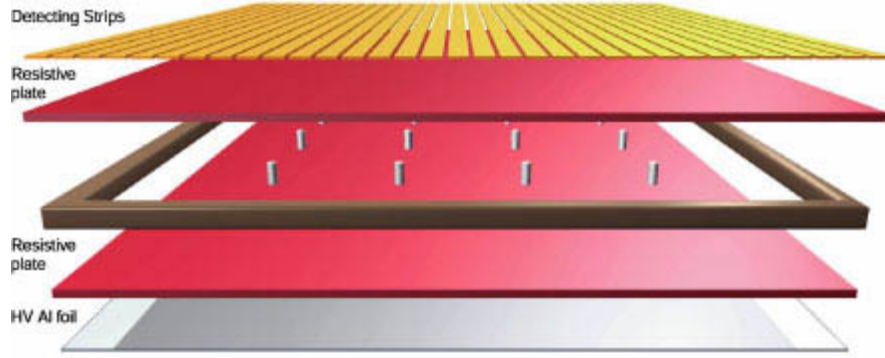


Figure 3.17: Schematic view of the layers surrounding a gas volume in an RPC chamber [77]. The second gas volume is located above the one shown here.

chambers allows for accurate determination of the muon kinematics.

3.2.6 Trigger system

The separation between the bunches of the LHC is of 25 ns and it is impossible to read the complete detector data at this rate, let alone store them for analysis. Fortunately, only a small fraction of the collisions are interesting. An elaborate trigger system has thus been implemented to retain bunch crossings with certain features. The CMS trigger runs in two stages, the first (level one, L1) implemented in dedicated hardware and the second (high level trigger, HLT) based on a computer farm using general-purpose processors.

The L1 trigger works by reconstructing coarse-grained “trigger primitives” in the calorimeters and muon systems and matching them into physics objects such as electrons and muons. Bunch crossings passing configurable cuts (about one in 400, reducing the rate to 100 kHz) are retained by the L1 trigger, and an acceptance signal (L1A) is sent to all parts of the detector. Upon receiving a L1A, the detector electronics start uploading the data for the specified bunch crossing to the central data acquisition system. An important limitation of the L1 trigger system is that two L1A signals must always be separated by at least two bunch crossings.

The data received from the detector after a L1A is regrouped in a single file, which is then staged for processing by the HLT filter farm. The HLT thus has access to the complete event information. However, time constraints only allow for minimal reconstruction at the HLT, in particular when it comes to tracking and the evaluation of expensive multivariate discriminants. Various optimizations are used to speed up the process: reconstruction is first attempted around the L1 primitives, less detailed algorithms are used, and early exit is performed when possible. The HLT reduces the trigger rate to around one kilohertz.

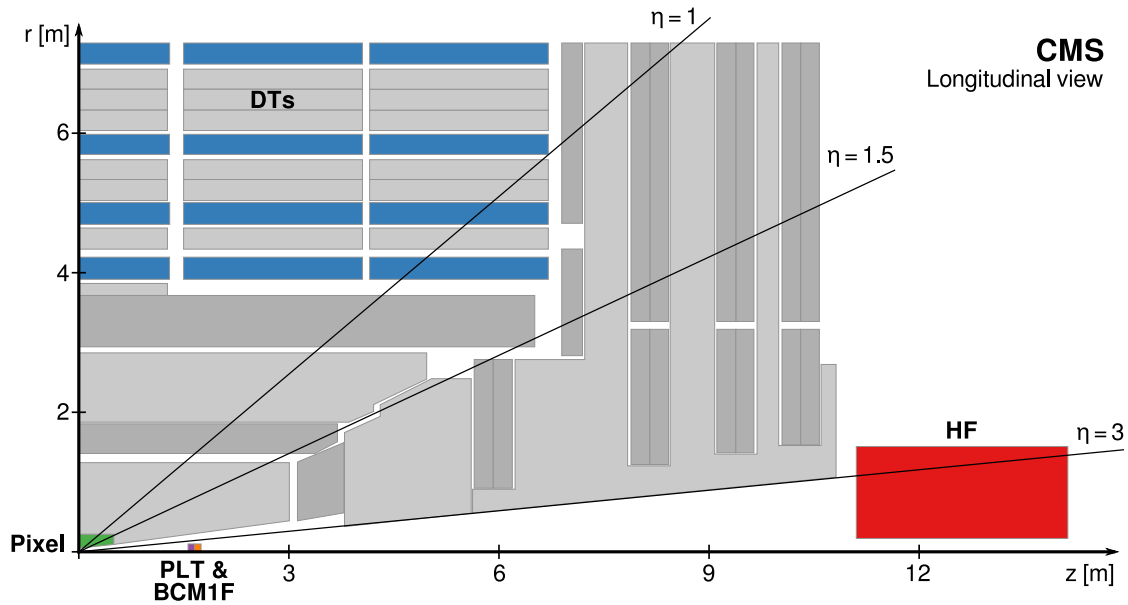


Figure 3.18: CMS subsystems used for luminosity monitoring [69]. The two RAMSES monitors are located behind the HF.

3.2.7 Luminosity meters

Several detectors in CMS are used to monitor the instantaneous luminosity [69]: the pixel detector, the drift tubes, the HCAL, two dedicated devices mounted behind the pixel endcaps (PLT and BCM1F), and two radiation monitors also used for safety (RAMSES). Luminosity monitoring differs from normal data taking in being time-based rather than event-based: rather than the properties of individual events, one is interested in the number of events in a given period of time. Therefore dedicated readout mechanics are used for luminosity in some of the detectors.

Luminosity monitoring in the pixel relies on its low occupancy. On the average, around 100 pixels hits are recorded per proton-proton collision and the risk of overlap between hits is negligibly small. Counting the number of pixel hits in minimum-bias events thus provides a metric that scales linearly with the instantaneous luminosity. A similar approach is provided by counting the number of reconstructed vertices (see section 4.4.1).

The forward hadron calorimeter (HF) is equipped with dedicated readout channels for luminosity monitoring. The HF computes histograms of the number of channels passing a programmable threshold and transmits them at fixed time intervals. This accumulation is performed at 40 MHz, independently of the trigger system. Linearity is ensured by only using those channels that have a small enough occupancy.

The drift tubes (DT) are also used to provide a measurement of the luminosity based on the number of muons crossing them. The number of muon tracks reconstructed for the L1 trigger is recorded and saved at regular intervals. Because the muon rate is low, the DT cannot be calibrated in van der Meer scans, and are used only to monitor the stability of the other systems.

The PLT, or Pixel Luminosity Telescope, uses 16 sets of three 64 mm^2 pixel sensors mounted perpendicular to the beam pipe around $|\eta| = 4.2$, separated by approximately 7.5 cm in the z direction. The sensor readout is performed at 40 MHz with a granularity of 80 rows and 52 columns. Triple coincidence between three sensor planes is used to measure the charge particle rate and thus the instantaneous luminosity. The Fast Beam Conditions Monitoring system (BCM1F) uses 24 diamond sensors mounted next to the PLT to provide fast (6 ns) measurements of incoming particles, allowing separation of collision products from beam-induced backgrounds. Each diamond has two readout channels.

The last sensors used by CMS to monitor the luminosity, part of the RAMSES system, were originally intended for radiation safety monitoring. Ten ionization chambers are installed in the experimental cavern to measure the ambient dose equivalent rate. Two of them, located behind the HF, are close enough to the beams to provide sensitivity to the luminosity. Similarly to the DT, the rate is too low to calibrate the RAMSES detectors in van der Meer scans and they are used to assess the stability of the system with time.

The CMS luminosity measurements in Run 2 [69] use a combination of all the aforementioned methods. In 2016, the relative luminosity collected by CMS in runs approved for physics use was 36.3 fb^{-1} , known with a precision of 1.2% . The absolute calibration using van der Meer scans contributes 1% to the uncertainty, dominated by the transverse factorizability of A_{eff} , equation (3.5). The “integration” over the complete run period contributes 0.7% to the total uncertainty.

Chapter 4

Drell-Yan measurement

Thanks to its excellent muon systems, electromagnetic calorimeters and inner tracker, the CMS detector is well suited for the detection of electron and muon pairs produced in the Drell-Yan process. The large integrated luminosity (36.3 fb^{-1} [69])¹ recorded in 2016 allows for precision measurements of differential distributions of this process. In a previous publication [79], the CMS Collaboration reported a measurement of the total and differential Drell-Yan cross sections as a function of the transverse momentum and rapidity of the produced lepton pairs, for invariant masses of the pair around the Z peak, $76 < m_{\ell\ell} < 106 \text{ GeV}$. In this chapter, the analysis of reference [79] is extended to invariant masses down to 50 GeV and up to 1000 GeV in order to test how predictions handle the change in one of the main scales of the process².

The general procedure used for cross section measurements is to count events reconstructed with kinematics within the target phase space, and use the relation between the number of events N , the cross section σ and the integrated luminosity $\mathcal{L}$:

$$N = \mathcal{L} \sigma \varepsilon \mathcal{A}, \quad (4.1)$$

where ε is the efficiency of the detection and reconstruction, and $\mathcal{A}$ is the acceptance of the selection. The luminosity is measured centrally for all analyses. The acceptance $\mathcal{A}$ is kept close to one by choosing a phase space close to the measurement, which also minimizes the associated uncertainty.

The main source of systematic uncertainty that can be acted upon at the level of the analysis is the efficiency. In order to achieve a high precision, care must be taken that the efficiency is understood correctly. The detailed modeling of efficiency corrections used in reference [79] is adopted in this thesis.

¹An earlier and less precise measurement reported 35.9 fb^{-1} [78]. This value is still used by a few plots in this thesis.

²This analysis was made public as CMS-SMP-20-003 [80] and is documented in the internal Analysis Note AN-2020/026.

This chapter starts with the description, in section 4.1, of the observables considered for the measurement. Existing results are summarized in section 4.2. The next sections describe the methods used to perform the measurement, following the logical flow of the analysis starting with the selection of data and Monte-Carlo samples (section 4.3), the reconstruction and selection of events (section 4.4), the discussion of control distributions (section 4.5), a discussion of two important background contributions (section 4.6), corrections for resolution, efficiency and acceptance effects (section 4.7), and the combination of the electron and muon channels (section 4.9). The treatment of the uncertainties affecting the measurement are described in section 4.8. The calculation of cross section ratios sensitive to the differences between different mass windows is explained in section 4.10. Finally, the predictions to which the results are compared are introduced in section 4.11 and the comparisons are discussed in section 4.12.

4.1 Observables

The main observable considered in this analysis is the transverse momentum of Drell-Yan lepton pair, $p_T(\ell\ell)$. As described in detail in chapter 2, p_T is sensitive to many aspects of QCD: the p_T -differential Drell-Yan cross section at small values of p_T can only be described accurately using soft gluon resummation, either directly or through the evolution of the TMD PDFs. At large p_T , the cross section is dominated by events with multiple jets that balance the lepton pair in the transverse plane; describing this part of the phase space requires merging samples with different parton multiplicities or using predictions calculated at a high order (NNLO or more) in α_s . Predictions that aim to cover the full p_T range accessible to measurements combine several techniques and the p_T range between the two extremes is sensitive to their matching. By measuring p_T -differential cross sections, we aim to provide experimental input about the latest related theoretical development in QCD.

Precise measurements of $p_T(\ell\ell)$ depend on the energy scale of the leptons, whose typical experimental resolution is much worse than that of angular variables. Several alternative variables have been proposed to improve experimental resolution while keeping sensitivity to the QCD effects mentioned above. In a dedicated study [81], the variable φ_η^* was found to be a good compromise between sensitivity and resolution. It is defined in terms of angular variables only, thus avoiding any dependence on the lepton energy scale:

$$\varphi_\eta^* \equiv \tan\left(\frac{\pi - \Delta\phi}{2}\right) \sin(\theta_\eta^*), \quad (4.2)$$

where $\Delta\phi$ is the opening angle between the leptons in the transverse plane and

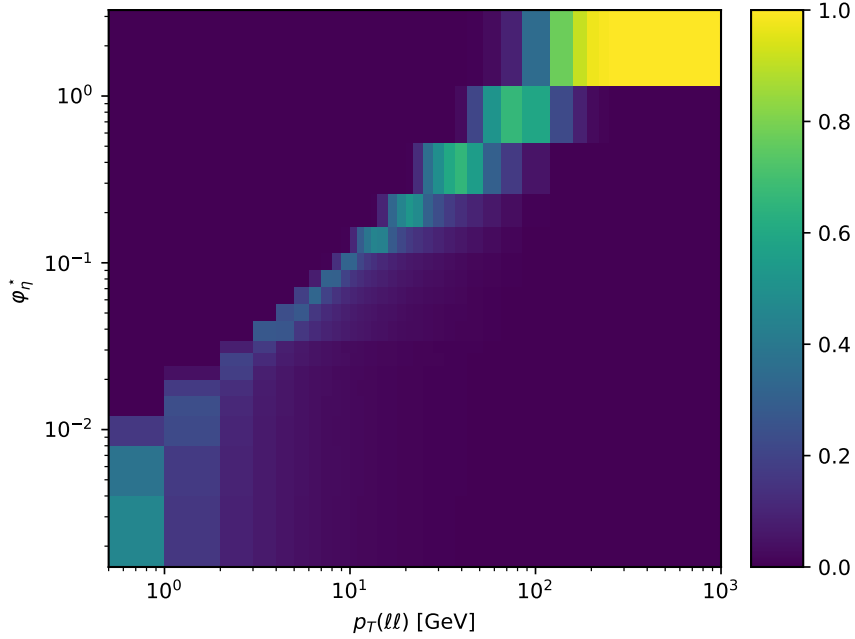


Figure 4.1: Correlation between $p_T(\ell\ell)$ and φ_η^* at the generator level, using the prediction described in section 4.3 with the phase space definition of section 4.7.1 for $76 < m_{\ell\ell} < 106$ GeV. Each bin represents the probability that an event generated with a given value in p_T has a certain value of φ_η^* . The linear scaling between the two variables is visible.

θ^* is defined from $\cos\theta_\eta^* = \tanh[(\eta^+ - \eta^-)/2]$. As is illustrated in figure 4.1, the φ_η^* variable is highly correlated with $p_T(\ell\ell)$. It is useful to keep in mind that φ_η^* can be approximated as $\varphi_\eta^* \sim p_T(\ell\ell)/m_{\ell\ell}$, with equality achieved when the two leptons have equal (scalar) transverse momenta. As has become common in recent measurements, this study includes φ_η^* differential cross sections in addition to $p_T(\ell\ell)$.

The third variable considered in this work is the transverse momentum of the lepton pair, with an additional requirement of at least one jet in the event selection. With this selection, the $p_T(\ell\ell)$ distribution peaks at the jet p_T threshold, where the dominant contribution is from events with exactly one jet in the recoil. Away from the peak, the cross section is dominated by events with two jets or more. At low $p_T(\ell\ell)$, an additional jet is needed to balance the one required in the selection. At large $p_T(\ell\ell)$, the transverse momentum of the pair is often balanced by more than one jet. Feynman diagrams for the production of a Z/γ^* boson in association with one and two jets are shown in figure 4.2. Some of these diagrams have gluons in

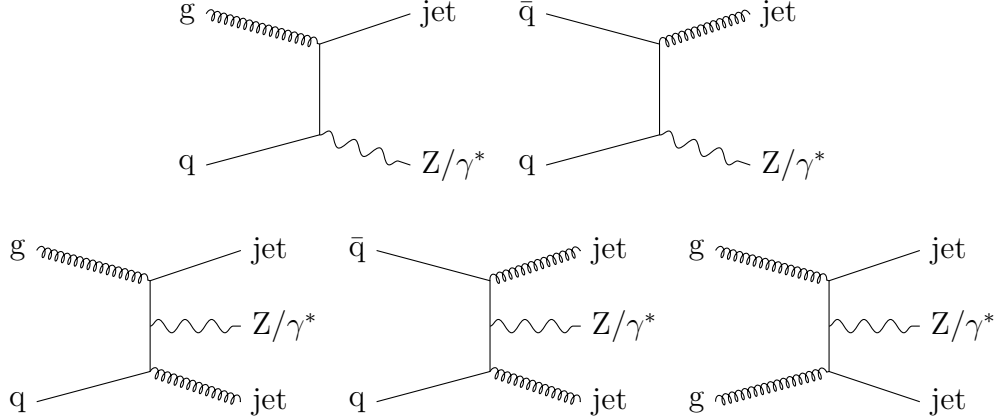


Figure 4.2: Examples of Feynman diagrams for $Z/\gamma^* + 1 \text{ jet}$ (top) and 2 jets (bottom).

the initial state. Therefore, measuring $p_T(\ell\ell)$ distributions when one jet is present probes resummation and TMD distributions for gluons. This is complementary to the measurement of inclusive $p_T(\ell\ell)$ which, at low p_T , is dominated by $q\bar{q}$ -initiated events.

Differential cross sections in the three variables described above are measured for five windows in the invariant mass of the lepton pair $m_{\ell\ell}$: $50 < m_{\ell\ell} < 76$, $76 < m_{\ell\ell} < 106$, $106 < m_{\ell\ell} < 170$, $170 < m_{\ell\ell} < 350$, and $350 < m_{\ell\ell} < 1000 \text{ GeV}$. This allows to compare transverse momentum for different Björken x and different scales of the hard scattering. In particular, the renormalization and factorization scales μ_R and μ_F are often chosen as $\mu_R = \mu_F = m_{\ell\ell}$. The handling of the scale dependence of the transverse momentum distributions depends on the approaches used for resummation and the fixed-order calculation. Performing a measurement at different scales should highlight differences between the approaches.

In order to get more direct information about the scale dependence of the spectra, cross section ratios with respect to the Z peak region are also measured as a function of the variables used for the differential cross sections,

$$\frac{\frac{d\sigma}{dp_T(\ell\ell)}(m_{\ell\ell})}{\frac{d\sigma}{dp_T(\ell\ell)}(m_{\ell\ell} = m_Z)}, \quad (4.3)$$

and likewise for φ_η^* and when requiring at least one jet. The Z mass has been selected in the denominator because it is the region in which most differential measurements have been performed so far. In addition, it is the region with the smallest experimental uncertainties, allowing for the most precise measurements of the ratios.

4.2 Existing measurements

The CMS Collaboration has performed several measurements of kinematic distributions of Drell-Yan lepton pairs, most of them around the Z pole. Measurements have been performed at $\sqrt{s} = 8$ [82] and 13 TeV [79]. Depending on the measurement, different variables were considered for single- and double-differential cross sections: Z rapidity y_Z , transverse momentum $p_T(\ell\ell)$ and φ_η^* . Preliminary results at 7 TeV [83] also include $p_T(\ell\ell)$ -differential cross sections measurements out of the Z peak region, in five $m_{\ell\ell}$ bins ranging from 30 to 1500 GeV. In addition, CMS has performed measurements of the mass shape of Drell-Yan events at the same centre-of-mass energies [84, 85, 86, 87, 88]. Similar measurements have been performed by the ATLAS [89, 90, 91, 34, 92, 93, 94] and LHCb [95, 96, 97, 98, 99] Collaborations.

This analysis is heavily based on the latest CMS Drell-Yan measurement on the Z pole [79]. In particular, it relies on its very detailed and precise studies of lepton reconstruction efficiencies. This choice has far-reaching consequences for the event selection: the same version of the datasets and reconstruction must be used and some of the cuts cannot be changed easily. Extensive validation has been performed to check the synchronization of the reconstruction in data and Monte-Carlo by comparing the results of the two analyses with exactly the same selection. This validation reached perfect agreement in data and reasonable agreement in Monte-Carlo, where the residual differences are explained by slightly different choices affecting only the simulation.

4.3 Data and signal samples

The analysis uses data collected in 2016 and reconstructed in the 03Feb2017 campaign in MINIAOD format, as listed in table 4.1. This version of the reconstruction was selected to enable the use of the very precise lepton reconstruction efficiency measurements performed for the CMS differential Drell-Yan measurement on the Z pole [79]. The `DoubleElectron`, `DoubleMuon` and `SingleElectron` datasets are used as a source for events selected by the corresponding triggers as described below. The data retained for analysis after quality control correspond to an integrated luminosity of 36.3 fb^{-1} [69].

For signal events, several Monte-Carlo samples generated using MADGRAPH5_AMC@NLO [100] interfaced with PYTHIA8 [43] are combined. They contain all three leptonic channels of the Drell-Yan process, $Z/\gamma^* \rightarrow e^+e^-, \mu^+\mu^-, \tau^+\tau^-$. The samples contain events with 0, 1, and 2 final state partons at the matrix element level, generated at NLO in α_s . The τ decays are handled by the TAUOLA [101] package and include all decay modes. Matching to parton

Table 4.1: List of datasets used in the analysis.

DAS path	Run range	$\mathcal{L}$ [fb ⁻¹]
/SingleMuon/Run2016B-03Feb2017_ver2-v2	273150-275376	5.4
/SingleMuon/Run2016C-03Feb2017-v1	275656-276283	2.4
/SingleMuon/Run2016D-03Feb2017-v1	276315-276811	4.3
/SingleMuon/Run2016E-03Feb2017-v1	276831-277420	4.1
/SingleMuon/Run2016F-03Feb2017-v1	277932-278808	3.1
/SingleMuon/Run2016G-03Feb2017-v1	278820-280385	7.5
/SingleMuon/Run2016H-03Feb2017_ver2-v1	281613-284035	8.5
/SingleMuon/Run2016H-03Feb2017_ver3-v1	284036-284044	0.22
/DoubleMuon/Run2016B-03Feb2017_ver2-v2	273150-275376	5.4
/DoubleMuon/Run2016C-03Feb2017-v1	275656-276283	2.4
/DoubleMuon/Run2016D-03Feb2017-v1	276315-276811	4.3
/DoubleMuon/Run2016E-03Feb2017-v1	276831-277420	4.1
/DoubleMuon/Run2016F-03Feb2017-v1	277932-278808	3.1
/DoubleMuon/Run2016G-03Feb2017-v1	278820-280385	7.5
/DoubleMuon/Run2016H-03Feb2017_ver2-v1	281613-284035	8.5
/DoubleMuon/Run2016H-03Feb2017_ver3-v1	284036-284044	0.22
/DoubleEG/Run2016B-03Feb2017_ver2-v2	273150-275376	5.4
/DoubleEG/Run2016C-03Feb2017-v1	275656-276283	2.4
/DoubleEG/Run2016D-03Feb2017-v1	276315-276811	4.3
/DoubleEG/Run2016E-03Feb2017-v1	276831-277420	4.1
/DoubleEG/Run2016F-03Feb2017-v1	277932-278808	3.1
/DoubleEG/Run2016G-03Feb2017-v1	278820-280385	7.5
/DoubleEG/Run2016H-03Feb2017_ver2-v1	281613-284035	8.5
/DoubleEG/Run2016H-03Feb2017_ver3-v1	284036-284044	0.22

shower is performed using the FxFx scheme [102]; the CUETP8M1 [39] tune of PYTHIA8 and the NLO NNPDF 3.0 PDF set [24] are used. Twelve samples generated with the same settings but using different constraints on the generated mass are combined to provide statistical power over the whole phase space, as summarized in table 4.2. Events in regions of overlap are weighted to reproduce the cross section obtained from the sample with the least restrictive cuts.

Monte-Carlo events are fed to a detailed GEANT4 [103] simulation of CMS. The signal created by the particles as they go through the detector is simulated and the same algorithms as in data are applied to reconstruct high level objects. The simulation is performed using an arbitrary number of pileup collisions that does not, in general, match the one observed in data. Since the efficiency of the reconstruction algorithms is highly dependent on the number of pileup collisions, Monte-Carlo events are reweighted according to the pileup profile deduced from the instantaneous luminosity delivered to CMS by the LHC. This procedure, known as “pileup reweighting”, makes use of the total proton-proton cross section, which is also its dominant systematic uncertainty.

4.4 Event reconstruction and selection

The most widely used event reconstruction algorithm in CMS is the so-called Particle Flow algorithm [104], which attempts to link together as much information coming from as many different subsystems as possible. Particle Flow is used both at the HLT and in analysis. The parts of the algorithm relevant to the present analysis are outlined in this section, and in particular the reconstruction and cal-

Table 4.2: Simulated signal samples used in the analysis.

$m_{\ell\ell}$ [GeV]	σ [pb]	N_{events}
10 – 50	18760.0	30 920 596
50 – ∞	5931.9	122 055 388
100 – 200	(not used)	30 910 238
200 – 400	(not used)	6 102 487
400 – 500	(not used)	287 262
500 – 700	(not used)	280 940
700 – 800	(not used)	276 235
800 – 1000	(not used)	271 768
1000 – 1500	(not used)	258 620
1500 – 2000	(not used)	258 625
2000 – 3000	(not used)	255 342

ibration of electrons, muons, and jets. The performance of the algorithms, and in particular their efficiency, and the corrections applied to simulated events are also covered. Finally, the event selection used for the measurement and the various backgrounds that contaminate the selected sample are described.

4.4.1 Tracks and vertices

The reconstruction of the trajectory of charged particles passing through the tracker material, or “tracks” [105], is a central step of the Particle Flow algorithm because tracks enter the definition of most objects used at the analysis level including electrons, muons, and jets. Track reconstruction starts by finding “hits”, the positions at which charged particles crossed a layer of the tracker. Hit triplets compatible with tracks originating from the interaction region are then used to seed a combinatorial Kalman filter and find more hits associated with the track. For seeds that lead to a track, a Kalman filter is used to accurately reconstruct the parameters of the track. The whole procedure is repeated several times with progressively loosened cuts on the seeds and the track candidates. Hits associated with tracks at a given iteration are not considered in the next. This allows to save CPU time by finding “easy” tracks first and reducing the combinatorial complexity of the later steps.

The points at which the proton collisions took place, called “vertices” [105], are characterized by the emission of several tracks originating from hard scatterings, underlying events, and pileup collisions. Reconstructing the vertices is required among others for the rejection of pileup particles and the jet energy scale, see section 4.4.6. This is achieved by clustering tracks based on the z coordinate of their point of closest approach to the beam axis. The vertex associated with the hard scattering, called the “primary” vertex, is chosen to be the one with the largest $\sum p_T^2$ of the associated tracks [104]. The vertex reconstruction algorithm is nearly 100 % efficient as soon as more than two tracks are used.

4.4.2 Electrons

Electrons are reconstructed as part of the Particle Flow algorithm as described in [106]. Most of the reconstruction is shared with photons, because their experimental signatures are similar. The algorithm starts by forming clusters of neighbouring ECAL crystals whose signal exceeds a fixed threshold. The clusters with the most energy are used as seeds to build a “supercluster” (SC) that groups clusters within a fixed window around the seed. The location of the supercluster and tracks with $p_T > 2 \text{ GeV}$ that are compatible with an electron trajectory are used as seeds for a specialized track reconstruction algorithm, the Gaussian Sum Filter (GSF). Electron pairs created by photon conversions in the tracker material

are recovered using another dedicated algorithm. Finally, the superclusters, GSF tracks and conversion tracks are assembled together into electrons (when a GSF track is present) and photons (when there is none).

The charge of the reconstructed electrons is not straightforward to measure from the curvature of the associated GSF track because of bremsstrahlung energy losses. In order to improve the precision of this information, two other methods are used: one based on the curvature of the non-GSF track associated with the electron and the other on the relative position of the supercluster and initial direction of the GSF track. The three methods can be combined either by selecting the charge chosen by two of them or by requiring that all three agree. The first method is used in this analysis because it doesn't incur an efficiency loss, but this comes at the cost of a worse charge reconstruction, especially in the endcaps.

The electron energy is measured by combining information from the curvature of the track with the energy measured in the ECAL [107]. The tracking information improves the resolution at low p_T ; for the energies considered in this analysis, the ECAL measurement dominates the combination. A three-step multivariate regression trained in the simulation is used to correct the supercluster energy for losses in the detector material, achieving a resolution better than 2 % in the barrel and 5 % in the endcaps for $p_T > 15$ GeV. The relative resolution improves as the energy increases, as can be seen in figure 4.3. Additional corrections for residual differences between data and simulation are derived using the mass distribution of Z boson candidates. This calibration proceeds in three steps. First, any time-dependent drift of the energy scale in data is taken into account. Second, the energy resolution in simulation and the energy scale in data are adjusted simultaneously. In the third step, the energy measurement at high p_T is improved by correcting for systematic differences in energy scale depending on the gain used in the ECAL readout electronics. After these corrections, the relative electron energy resolution is between 2 % and 5 % depending on pseudorapidity and energy losses in the tracker material.

Additional criteria are applied to electron candidates to reduce background contributions, based on the shape of the electromagnetic shower, the difference between the inverses of the ECAL-only and tracker-only energies, the difference between the direction of the track and the supercluster position, and isolation of the candidate. Four sets of cuts are defined by the group responsible for electron objects within CMS, with target signal efficiencies of 95 (“**veto**”), 90 (“**loose**”), 80 (“**medium**”), and 70 % (“**tight**”). The misidentification rate is higher for looser criteria. In the present analysis, electron candidates are required to pass the **medium** cuts. The **veto** working point is also used to reject additional electrons (see section 4.4.7).

The electron reconstruction and identification efficiencies measured for the on-

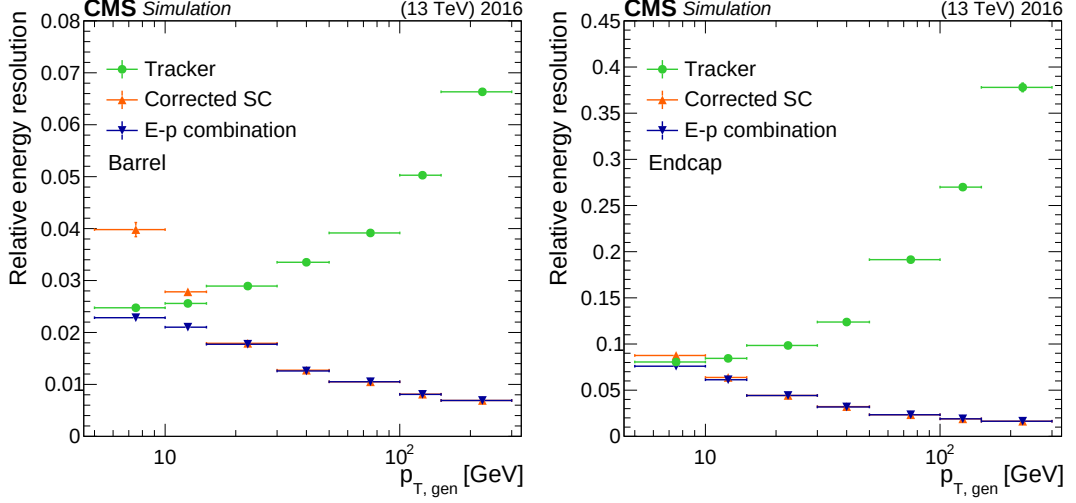


Figure 4.3: Simulated electron energy resolution as a function of the generated p_T , in the barrel (left) and endcaps (right). The contributions from the tracker and ECAL supercluster are shown, as well as their combination. Taken from [106].

shell analysis [79, 108] are used. They are factored out in two components, reconstruction and identification. The reconstruction efficiency is provided as a function of the pseudorapidity of the supercluster η_{SC} and the run period (B–F and G–H). The difference between data and simulation is about 2 %. The identification efficiency is provided in 30×16 bins in η_{SC} and p_T , separately for runs B–F and G–H; it is shown in figure 4.4 and the difference between data and simulation, ranging from 10 % at low p_T to 2 % at high p_T , is shown in figure 4.5. The electron identification efficiency ranges from 60 % at 20 GeV to 88 % at high p_T . It is smaller in the barrel-endcap transition region, visible around $|\eta| = 1.5$, and in the endcaps. The scale factors range from 90 % at low p_T to one at high p_T . In the barrel, they are closer to one for runs G–H than for runs B–F, but the opposite is true in the endcaps, especially at low p_T . For both the reconstruction and identification efficiencies, scale factors are applied to Monte-Carlo events. The scale factors are applied twice, once for each of the two electrons used in the event selection.

4.4.3 Electron triggers

The event selection starts by requiring the events to have fired specific triggers. The choice of triggers has a large impact on several aspects of the analysis: the lepton p_T cuts, for instance, are strongly dependent on this choice. In order to maximize efficiency and acceptance, unprescaled triggers with p_T cuts as small as possible are used. Since the final selection requires either two electrons or two

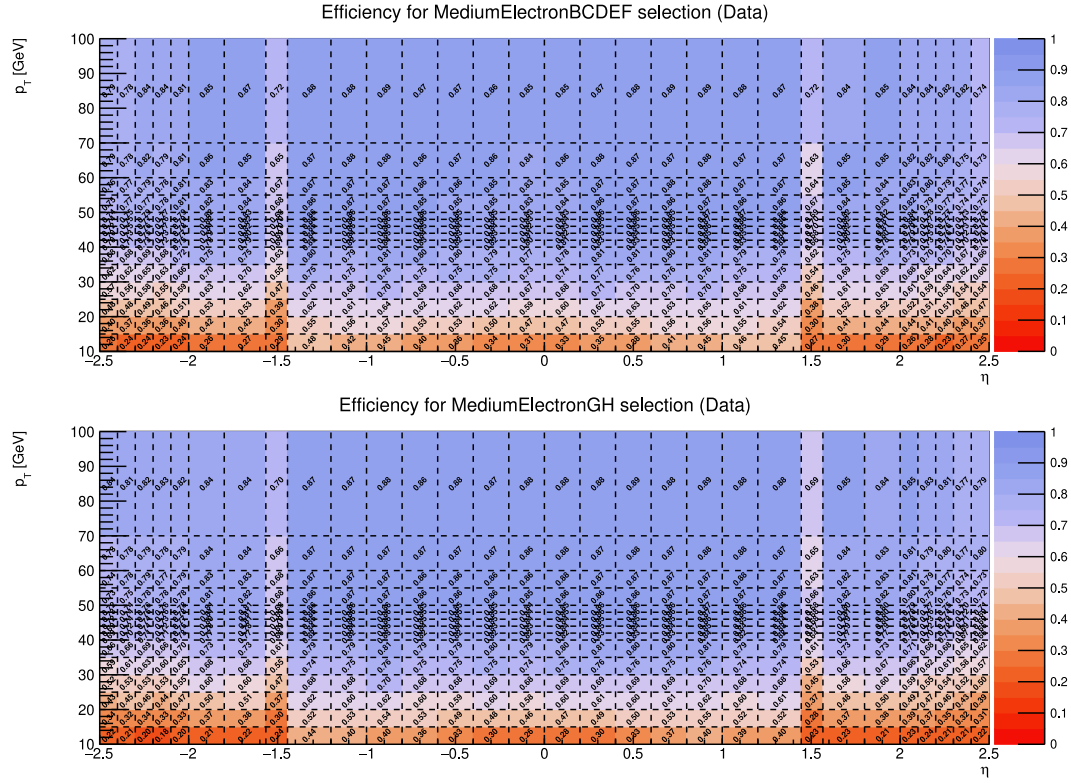


Figure 4.4: Data efficiency of the medium electron identification criteria in bins of η_{SC} and p_T , for runs B–F (top) and G–H (bottom). Plots by D. Hsu [108].

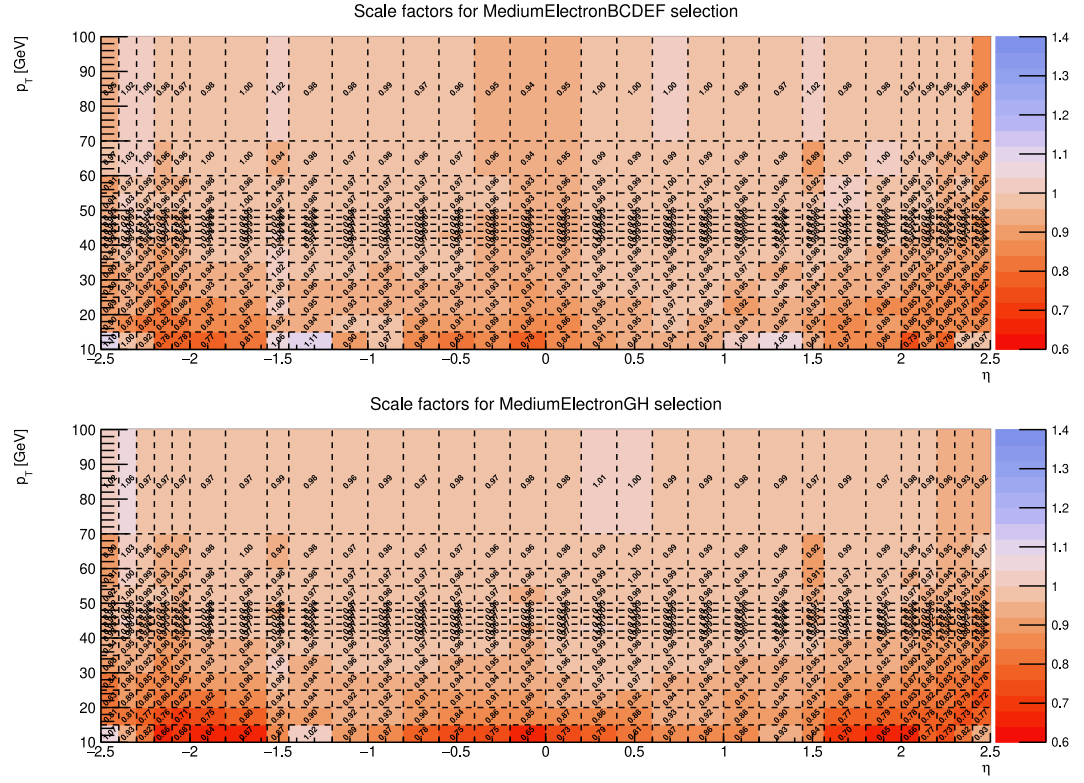


Figure 4.5: Scale factors for electron medium identification in bins of η_{SC} and p_T , for runs B–F (top) and G–H (bottom). Plots by D. Hsu [108].

muons as part of the offline selection, double lepton triggers are used.

In the electron channel, the lowest unprescaled double electron trigger available during the whole year is used: `HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ`. This HLT trigger path requires two electrons with E_T greater than 23 and 12 GeV respectively, loose calorimeter (“`CaloIdL`”) and tracker (“`TrackIdL`”) based identification, and very loose isolation (“`IsoVL`”). In addition, loose compatibility between the vertices associated with the two tracks is required (“`DZ`”). These criteria, implemented at the HLT, are much tighter than the hardware trigger used at L1.

To estimate the efficiency of this trigger, it is factorized into three components: the reconstruction efficiency for each electron (“leg”) as a function of the pseudorapidity of the ECAL supercluster η_{SC} and track p_T , and the efficiency of the DZ filter as a function of the number of vertices n_{vertex} in three categories (endcap-endcap, endcap-barrel and barrel-barrel):

$$\epsilon_{\text{trigger}} = \epsilon_{\text{leg } 1}(\eta_{SC}, p_T) \cdot \epsilon_{\text{leg } 2}(\eta_{SC}, p_T) \cdot \epsilon_{\text{DZ}}(n_{\text{vertex}}, \text{category}). \quad (4.4)$$

The efficiencies with respect to the offline reconstruction are measured in data and Monte-Carlo using the tag-and-probe method [84]. The results are shown in figures 4.6 and 4.7. With respect to p_T , the efficiency is almost 0 below the p_T cut, before raising rapidly (“turn-on”) and reaching a “plateau” at nearly 100 % efficiency. The spread of the turn-on region is important when defining the offline p_T cuts, as it is generally avoided in analyses due to the difficulties involved in modeling it. A decrease of the efficiency at large vertex multiplicity is also visible. It is well described in simulation when both electrons are in the barrel but corrections are needed when at least one of them is in the endcaps.

The efficiencies measured in data and simulation are similar but not identical. To ensure that the efficiency correction derived from the Monte-Carlo will be correct, “scale factors” are assigned to Monte-Carlo events: weights derived from the ratio of the efficiencies in data and Monte-Carlo. This way, if the measured efficiency is 80 % in data and 95 % in the simulation, each event passing the trigger in Monte-Carlo contributes for only $80/95 = 85$ % of one event, and the overall efficiency matches the one measured in data. This procedure introduces uncertainties related to the estimation of the efficiencies, which propagate to the final results. Scale factors are widely used in the present analysis, and many of the considered uncertainty sources arise from their usage.

In 2016, aging crystals in the ECAL endcaps caused a time shift at the level of the L1 trigger (see section 3.2.6), in turn causing electron triggers to sometimes be fired one bunch crossing too early (“pre-firing”) [110]. The interplay of this timing mismatch and the rule that at most one L1A signal may be sent in three bunch crossings causes events to be lost completely. Indeed, let us consider an “interesting” event happening at bunch crossing b_x . If the trigger fires too early,

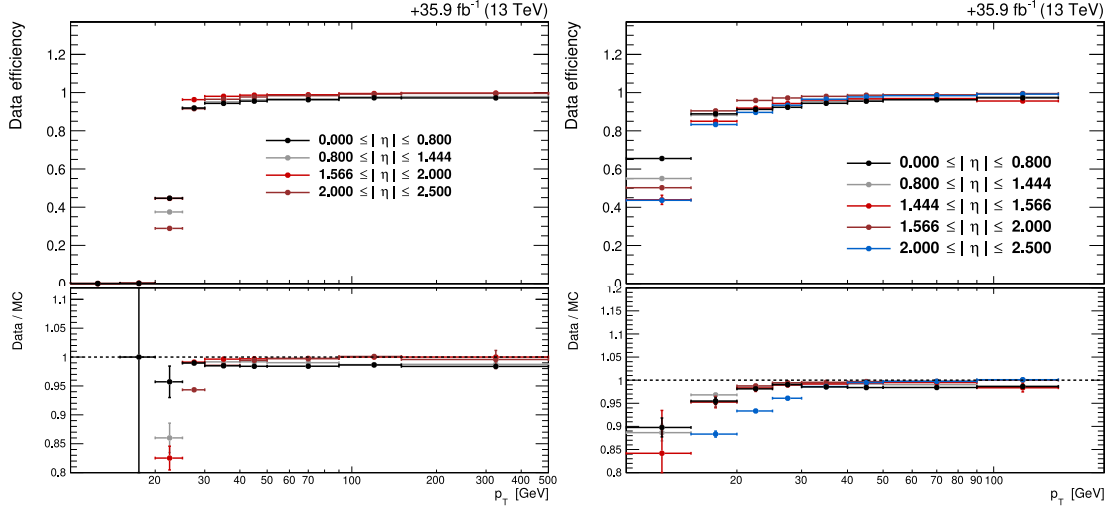


Figure 4.6: Efficiencies of the first (left) and second (right) legs of the electron triggers, as a function of the η_{SC} and p_T of the leg. Plots adapted from reference [109].

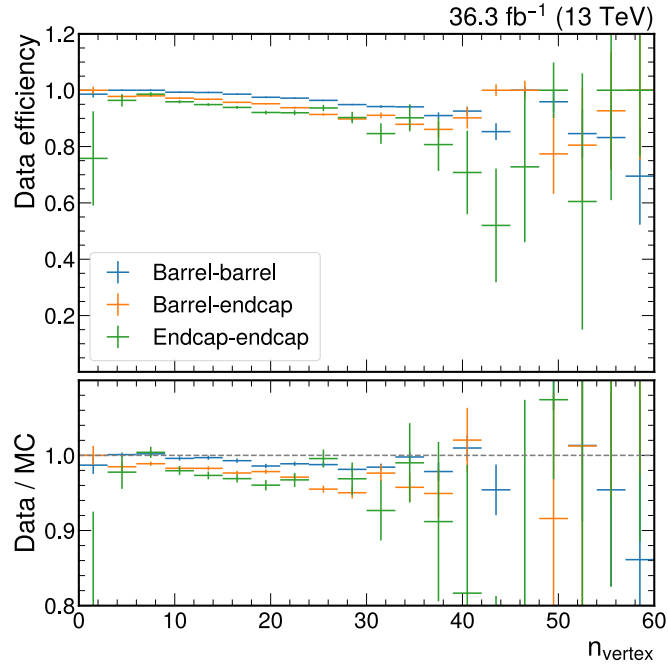


Figure 4.7: Efficiency of the DZ filter in the electron trigger, as a function of the number of reconstructed vertices n_{vertex} . Plot adapted from reference [109].

an L1A signal is sent for bunch crossing $b_x - 1$. As a consequence, anything happening in bunch crossings b_x and $b_x + 1$ is masked, including the interesting event. Other triggers fired by the interesting event are ignored. The event content read from $b_x - 1$ will likely not contain any signal matching the L1 seeds and be silently discarded by the HLT.

The prefiring rate cannot be estimated using the standard tag-and-probe technique because there is no tag: the whole event is lost. Fortunately, there is one situation under the trigger rules where one event cannot be pre-fired: if an L1A signal was sent for bunch crossing $b_x - 3$, pre-firing with respect to bunch crossing b_x will fall within the dead time of $b_x - 3$ and be ignored. Using these “un-pre-fireable” events and the information about L1 seeds for $b_x - 1$, one can measure the pre-firing probability associated with physics objects leaving a signal in the ECAL: electrons and photons, and jets. The rates are shown in figure 4.8.

4.4.4 Muons

Three types of muon candidates are reconstructed in CMS, all of which require at least a signal in the muon systems [111], with or without an associated tracker track. Reconstruction starts either from a track or from a group of DT or CSC segments. In the first case, a track is extrapolated “inside-out” to the muon system where it is matched to DT or CSC segments. Such candidates are called “tracker muons”. Reconstruction can also proceed by fully ignoring the tracker, reconstructing a track in the muon systems alone to form a “standalone muon”. Standalone muons are also matched to tracker tracks in an “outside-in” algorithm to generate “global muons”. A large majority of muons are reconstructed as tracker and/or global muons, and candidates reconstructed using several algorithms are merged.

The charge and momentum of muon candidates are reconstructed simultaneously by fitting the muon track using a Kalman filter. The TuneP algorithm [112] chooses between four different fits to optimize the quality of the determination of q/p . For $p_T < 200$ GeV, the momentum determination is dominated by the reconstruction of the tracker track. At larger p_T , hits in the muon system contribute significant information and performing a global fit becomes important. At extreme energies, muons can lose a significant fraction of their energy to an electromagnetic shower in the calorimeters, magnet, and return yokes [113]. Such cases are addressed by two specialized fits. The momentum resolution in simulation is better than 2% for muons in the central barrel [111] with $p < 100$ GeV and degrades with increasing momenta, saturating at 5–6% as seen in figure 4.9. Performance is better in the barrel than in the endcaps.

In addition to the above, the momentum calibration is refined using a technique known as the Rochester correction [114]. This calibration is performed using

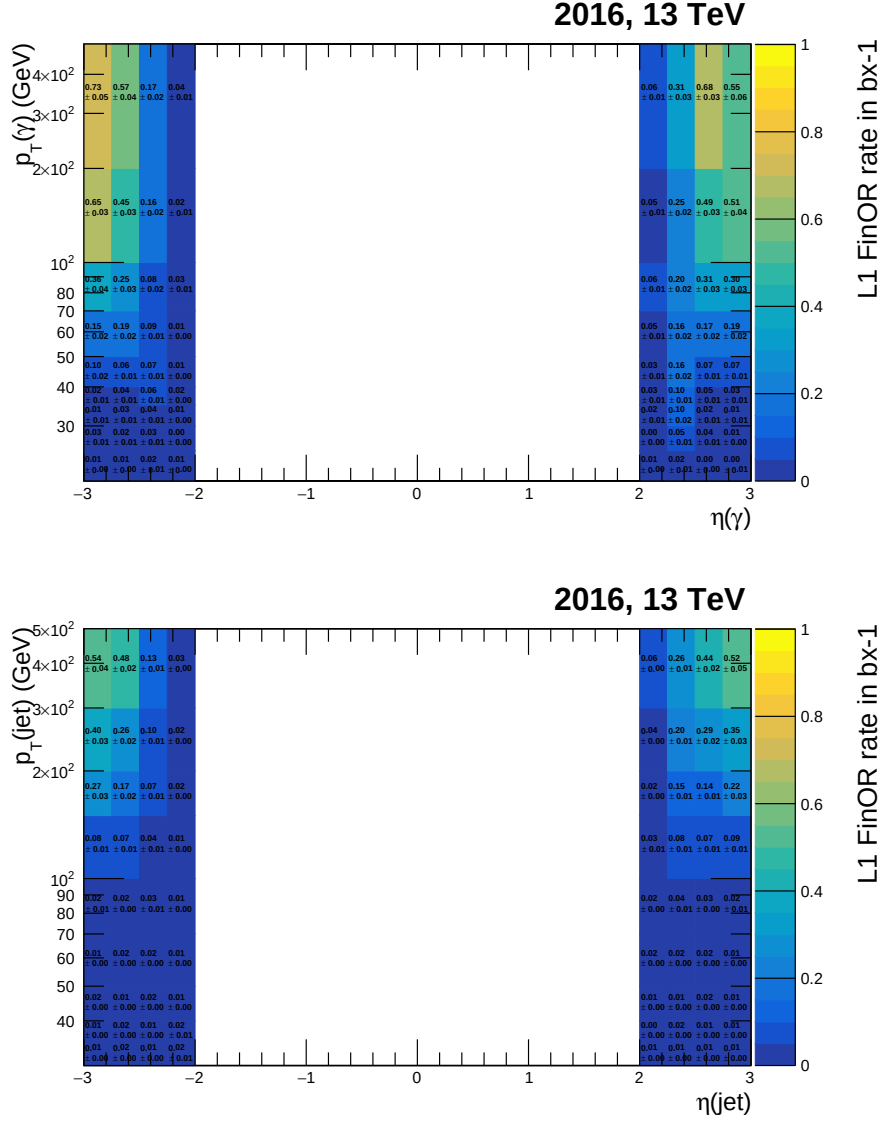


Figure 4.8: Pre-firing probability for electron and photons (left) and jets (right), as a function of η and p_T . Pre-firing in the barrel (middle area) is negligible. Plots from reference [110].

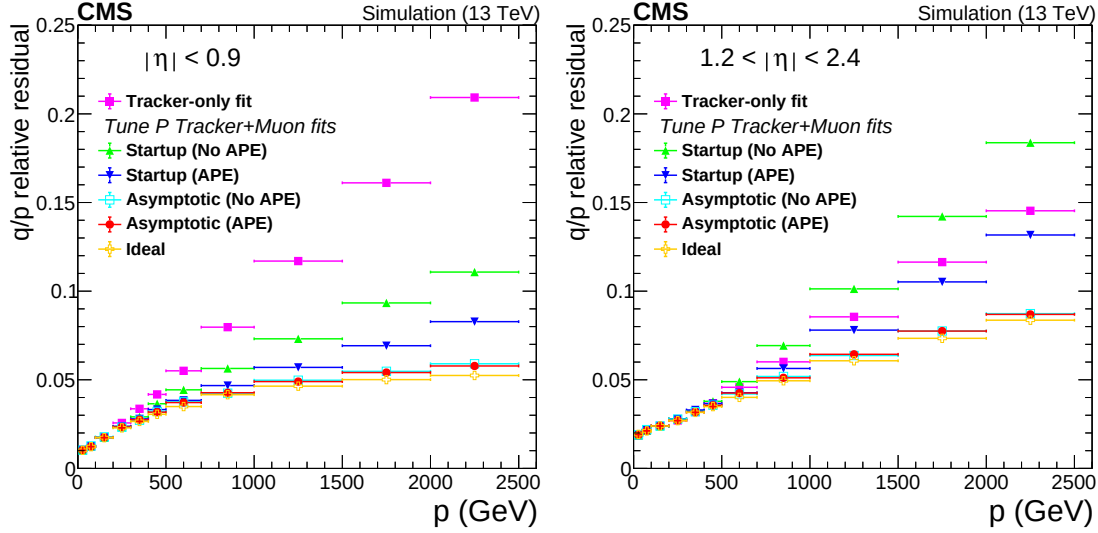


Figure 4.9: Muon momentum resolution in simulation in the central barrel (left) and outer endcap (right). The nominal curve is “Asymptotic (ACE)”; others correspond to various alignment scenarios. Curves using the tracker only are also shown for reference. Taken from [113].

a Monte-Carlo sample smeared using functions representative of the detector resolution, representing “perfect alignment”. In a two-steps procedure, the average inverse muon transverse momentum $\langle 1/p_T \rangle$ of Z events in data and simulation is first required to match the one derived from perfect alignment. This correction is derived in bins of charge, η and ϕ . In addition, the resolution of muon p_T in simulated events is smeared to yield the same mass resolution as in data. After this step, the average Z mass in simulation agrees with the “perfectly aligned” sample in each bin of η and ϕ , but this is not the case in data because of mismodeling of the detector efficiency in simulation. The second step corrects for this residual bias. The Rochester procedure greatly improves the agreement between data and simulation around the Z peak.

The quality of muon identification is assessed based on which algorithm was used for reconstruction as well as variables based on the matching between the tacker and muon tracks and the quality of the fits. Like for electrons, several working points exist with different performance characteristics: “loose”, “medium”, “tight”, and more specialized variations [111]. In the present analysis, muon candidates are selected using a variation of the **medium** identification with a relaxed cut on the fraction of valid tracker hits, 49% instead of the usual 80% (“**medium2016**”) [79, 108], to mitigate the tracking inefficiency caused by saturation effects in the APV25 readout chip [73]. These settings were chosen to be about

99.5 % efficient for muons from Z and W decays. For the rejection of additional leptons discussed in section 4.4.7, the `loose` identification criteria are used.

The `medium2016` identification criteria have no isolation requirement, meaning that they will select a muon within a jet just as well as a prompt muon from the hard interaction. In order to reject muons from hadron decays, the Particle Flow (PF) isolation is used. PF isolation works by summing the momenta of all PF candidates other than muons in a cone of radius $\Delta R = 0.4$ around the muon of interest. The sum is first separated into a charged and a neutral components:

$$I_{\text{PF}}^{\text{abs}} = I_{\text{charged}}^{\text{abs}} + I_{\text{neutral}}^{\text{abs}}. \quad (4.5)$$

The charged component can be reconstructed accurately because pileup particles entering the cone originate from a different vertex. The sum runs over of all charged hadrons whose track originates from the primary vertex: $I_{\text{charged}}^{\text{abs}} = \sum_{h^\pm, \text{primary}} p_{\text{T}}(h^\pm)$. The neutral component, on the other hand, is built mainly from energy deposits in the calorimeters that do not have associated tracks. The pileup contribution is therefore removed stochastically using the measured contribution from charged pileup particles:

$$I_{\text{neutral}}^{\text{abs}} = \max \left(0, \sum_{h^0} p_{\text{T}}(h^0) + \sum_{\gamma} p_{\text{T}}(\gamma) - \Delta\beta \sum_{h^\pm, \text{pileup}} p_{\text{T}}(h^\pm) \right). \quad (4.6)$$

The first two sums cover neutral particles, including contributions from the primary vertex and from pileup. The third term accounts for the neutral pileup, scaling the contribution from charged pileup particles by a transfer factor $\Delta\beta = 0.5$ extracted from simulation. The isolation variable used for analysis is relative to the muon transverse momentum, $I_{\text{PF}}^{\text{rel}} = I_{\text{PF}}^{\text{abs}}/p_{\text{T}}(\mu)$. In the present case, muon candidates are required to satisfy “`tight`” isolation, $I_{\text{PF}}^{\text{rel}} < 0.15$.

The muon reconstruction and identification efficiency measurements performed for the on-shell analysis [79, 108] are used. The reconstruction efficiency is measured as a function of the muon η and the run period. The difference between data and simulation is about 0.5 % for runs B–F and less than 0.1 % for runs G–H. The identification efficiency, including isolation, is measured as a function of η and p_{T} using a coarser binning than for electrons, 13×11 ; it is shown in figure 4.10. It ranges from 75 % at low p_{T} to over 95 % above 40 GeV. At low p_{T} , the efficiency is better in the endcaps than in the barrel. The efficiency in simulation closely follows the data and the two differ by a few percent depending on the run era and the kinematics, as shown in figure 4.11. The largest corrections, of about 5 %, are required for high- p_{T} muons for almost all η bins in runs B–F and a few bins in runs G–H. Like for electrons, the efficiency in simulation is corrected by the application of scale factors for each measured muon.

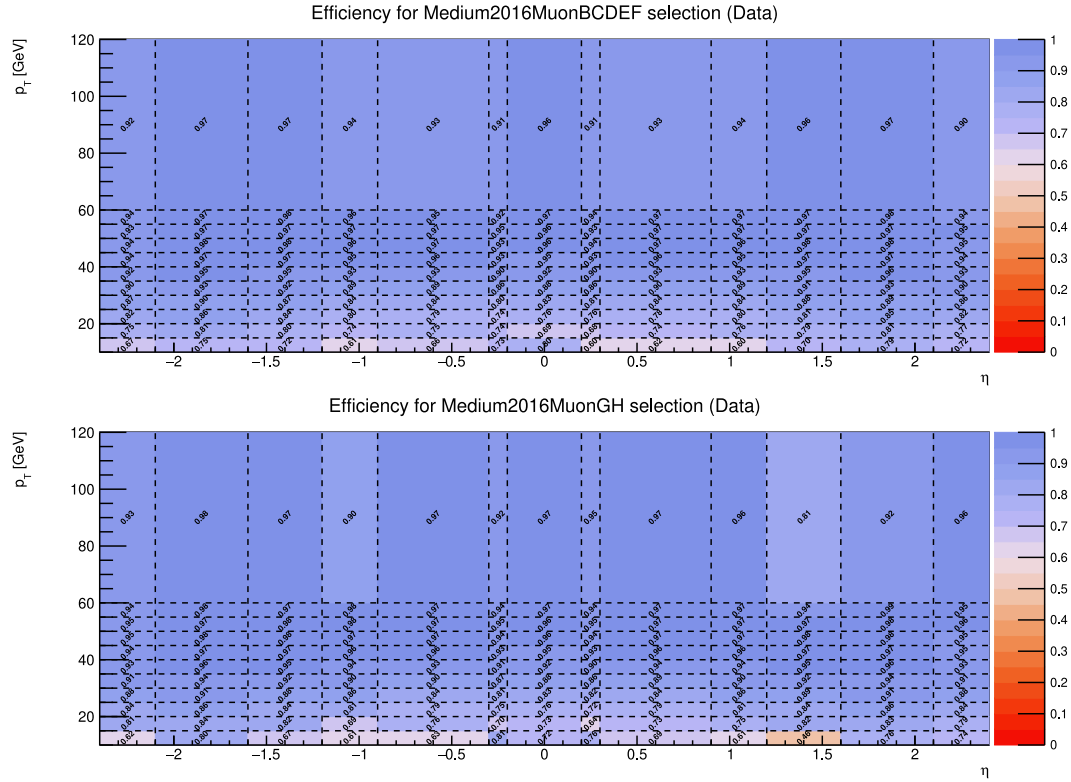


Figure 4.10: Data efficiency of the `medium2016` muon identification criteria and tight PF isolation, in bins of η and p_T for runs B–F (top) and G–H (bottom). Plots by D. Hsu [108].

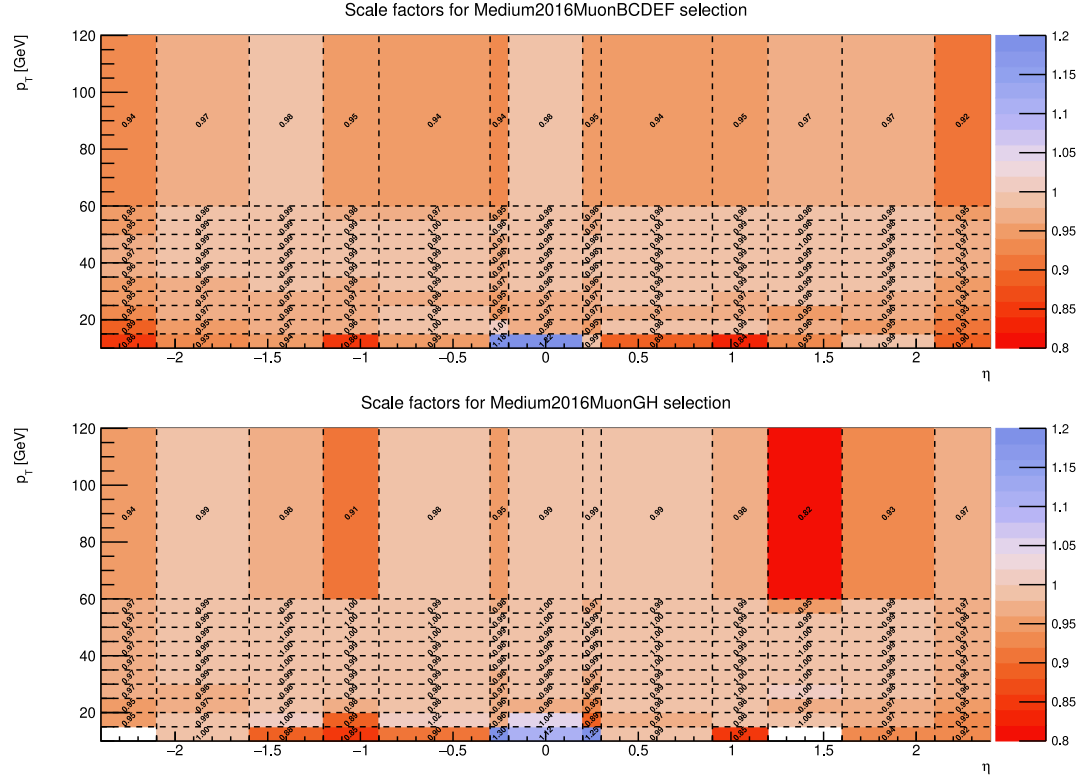


Figure 4.11: Scale factors for muon `medium2016` identification and tight PF isolation, in bins of η_{SC} and p_T for runs B–F (top) and G–H (bottom). Plots by D. Hsu [108].

4.4.5 Muon triggers

In the muon channel, a mix of four HLT triggers is used: two double muon and two single muon triggers. The triggers depend on the run: for runs B to F, a logical OR of the following is used:

```
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL
HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL
HLT_IsoMu24
HLT_IsoTkMu24
```

The first two triggers require two muons with p_T larger than 17 and 8 GeV, respectively, and “very very loose” tracker-based isolation. The other two require a single isolated muon or tracker muon with p_T larger than 24 GeV. These cuts are much tighter than the ones imposed on trigger primitives at the L1. For Monte-Carlo and runs G and H in data, double muon triggers requiring loose compatibility between the vertices associated with the two tracks are used: `HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ` and `HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ`. The single muon triggers are unchanged. The events passing these triggers are accessed by combining the single and double muon datasets. The overlap between the samples is resolved by not considering events from single muon datasets also present in the double muon datasets.

In the beginning of the 2016 data-taking campaign, a misconfiguration of the Endcap Muon Track Finder (EMTF) was preventing it from reconstructing more than one muon falling in the same station of the CSC [115]. This inefficiency is not included in the simulation. In order to mitigate the effect, events with the two leading muons falling in the same endcap are rejected if they have $|\eta| > 1.2$ and an azimuthal separation $\Delta\varphi < 60^\circ$. This results in a small efficiency reduction but it effectively suppresses the pathological case.

Like in the electron channel, the efficiencies of the combinations (“soups”) of triggers are measured in data and Monte-Carlo. The reference trigger method is used. The procedure involves three steps. First, the efficiency $\epsilon_{\text{HLT_Mu17}}$ of the “reference” single muon trigger, `HLT_Mu17`, is measured against offline reconstruction using the tag-and-probe method. The reference trigger must fire whenever any trigger in the soup passes (except for prescale) and thus has looser cuts than any of them. The second step is to calculate the efficiency ϵ_{ref} of `HLT_Mu17` in double muon events, as:

$$1 - \epsilon_{\text{ref}} = (1 - \epsilon_{\text{HLT_Mu17}}(\mu_1))(1 - \epsilon_{\text{HLT_Mu17}}(\mu_2)). \quad (4.7)$$

Finally, the efficiency $\epsilon_{\text{soup|ref}}$ of the soup is measured in events passing the reference trigger `HLT_Mu17`. Mathematically, the absolute efficiency of the soup then given

by:

$$\epsilon_{\text{soup}} = \epsilon_{\text{soup|ref}} \cdot \epsilon_{\text{ref}}. \quad (4.8)$$

The measured efficiencies are shown in figure 4.12 as a function of the absolute pseudorapidity of the two muons. Except for a few bins, they are above 98 %. Like in the electron channel, the Monte-Carlo is corrected to match the data in efficiency; the scale factors are taken with respect to the DZ triggers used in the Monte-Carlo, and deviate from unity by about 2 % for runs B to F and less than one percent for runs G and H.

4.4.6 Jets

Jets are reconstructed by applying the anti- k_T algorithm (as implemented in the FASTJET library [117]) on all particle flow candidates [118]. For this analysis, the cone size parameter is set to $R = 0.4$. Jet momentum is determined as the vector sum of the momentum of all particles clustered in the jet, which are mostly charged hadrons (65 %), photons (25 %), and neutral hadrons (10 %). Charged and neutral hadrons are reconstructed from energy deposits in the HCAL associated or not with a track, respectively. Photons within jets are reconstructed from energy deposits in the ECAL. Rejection of jets arising from instrumental noise is performed based on their energy composition [119], for which the “loose” working point is used. The use of PF candidates greatly improves the jet energy and spatial resolutions over jets solely built using calorimeter information.

Because they group many particles, often of low p_T , jets are very sensitive to contamination from pileup collisions. Indeed, the jet algorithm considers pileup particles as part of the jets, which increases the apparent energy. A first correction is performed by removing particles not associated with the primary vertex (“charged-hadron subtraction”) [120]. This reduces the contamination by about two thirds [118]. A second correction is applied to account for contamination from neutral particles, using the effective area of the jet in the (η, ϕ) plane and the average energy density in the event to subtract the expected pile-up energy. The subtracted energy is measured in minimum bias events in bins of the effective area, the number of pileup vertices, and the jet p_T and η . After these corrections, pileup contributions have been removed but the jet energy is not yet calibrated.

The calibration of the jet energy scale is performed in three steps (only the first is needed in simulation). First, a correction factor for the jet energy is derived in bins of p_T and η by matching reconstructed and generated jets in simulated multijets events. At this point, the energy scale in simulation is fixed and only residual corrections in data are performed. The second step in data consists in a second correction factor, depending only on η , that is extracted from the p_T balance in dijet events to correct for residual effects of the detector response. This allows an

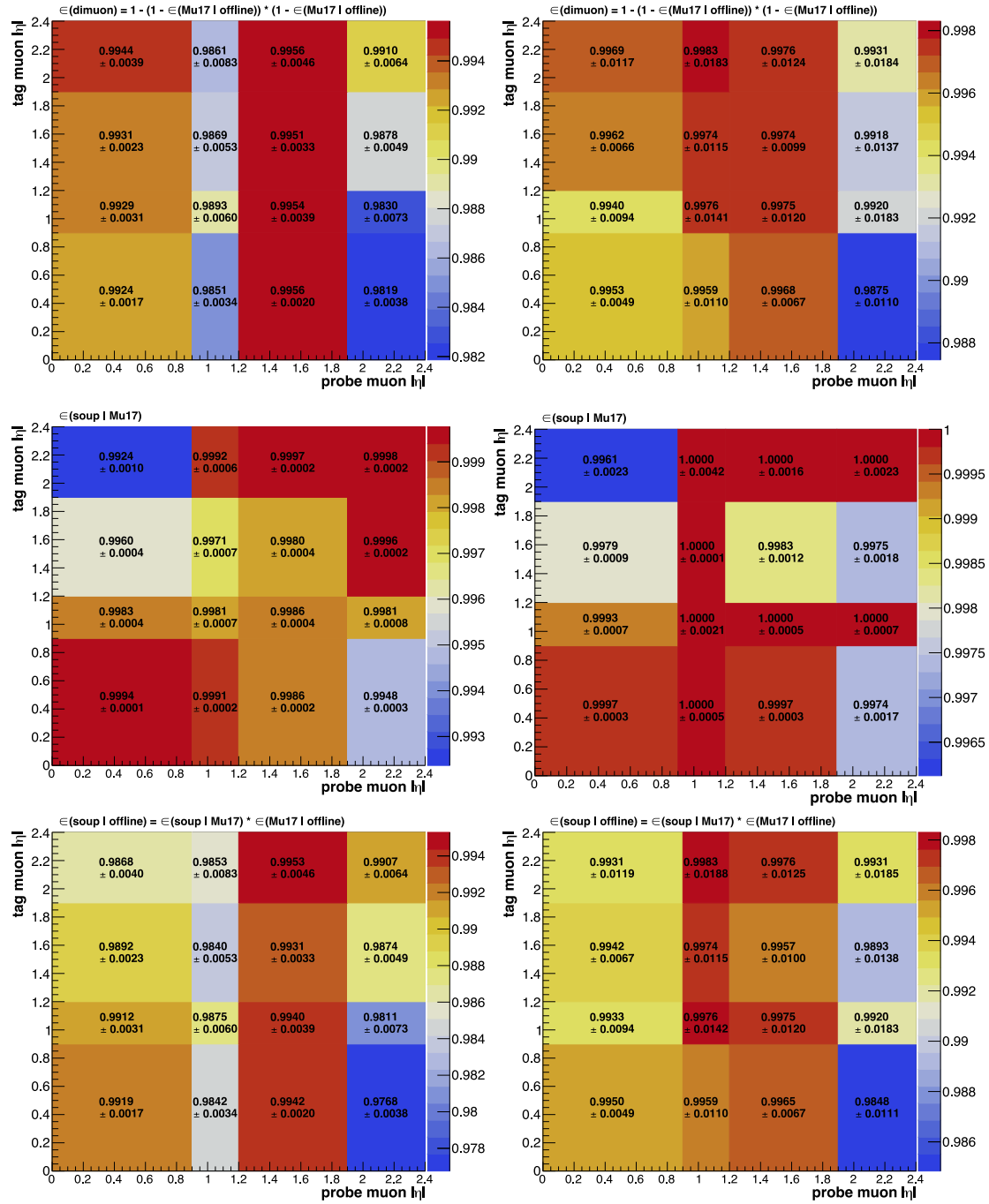


Figure 4.12: Muon trigger efficiencies at different steps of the reference trigger methods, for runs B to F (left) and G–H (right) in data. Top: efficiency of HLT_Mu17 in dimuon events, with respect to offline reconstruction (ϵ_{ref}). Center: efficiency of the soup with respect to HLT_Mu17 ($\epsilon_{\text{soup}|\text{ref}}$). Bottom: efficiency of the soup with respect to offline reconstruction (ϵ_{soup}). Plots from reference [116].

intercalibration of the relative energy of jets at different pseudorapidities, based on a reference region $|\eta| < 1.3$. The last step in data consists in the calibration of the absolute energy scale, performed for the reference region $|\eta| < 1.3$ by exploiting the balance in p_T between jets and well-reconstructed probes: photons and leptonically decaying Z bosons. The resolution of the jet energy assignment is measured in dijet and $\gamma + \text{jets}$ events. It is better in simulation than in data and simulated jets are smeared to better reproduce the actual resolution.

In addition to biasing the apparent energy of jets, pileup particles can be clustered in jets of their own. This can happen either as a stochastic accumulation of pileup particles in the same region or from QCD jets from a pileup interaction. The energy spectrum of pileup jet is rapidly falling and charged hadron subtraction reduces the measured energy, so pileup jets appear mostly at low p_T . In order to further reduce the contamination from pileup jets, a multivariate discriminant called “pileup jet ID” [119] is trained on Z + jets simulated events. It uses as inputs variables related to the composition, geometric shape, energy distribution and number of particles in the jets, as well as the number of vertices. The output of the pileup jet ID is a number between -1 and $+1$, where -1 is associated with pileup jets and $+1$ with jets from the primary vertex. The distribution of the output is shown in figure 4.13. By rejecting jets with small values of the pileup jet ID, it is possible to reduce the pileup jet contamination to the level required by the analysis.

Jets are the generic manifestation of the production of an energetic gluon or quark in the hard process. It is often desirable to know the type of parton from which a jet originated and this is possible to some extent. In CMS, discrimination between different types of jets is based on the presence of mesons with a resolvable mean free path: in particular, B mesons (often produced in b jets) have a mean free path of around 0.5 mm, and D mesons (produced in c jets) of around 0.3 mm [11]. These distances are one order of magnitude longer than the spatial resolution of hits in the barrel of the pixel detector [105]. The discrimination algorithms used in CMS [121] attempt to reconstruct the secondary vertices that correspond to B and D meson decays within jets. Several multivariate discriminants (“jet taggers”) are available for analysis; in this measurement the CSVv2 tagger is used, which is capable of discriminating between light (u, d, s, g), c and b jets. The algorithm is trained in simulated events and validated in data. The efficiency of the algorithm is measured in data using selections designed to produce enriched samples of c and b jets and per-jet scale factors are derived to correct the simulation.

4.4.7 Event selection

The final event selection, summarized in table 4.3 starts by requiring that events pass the triggers discussed in sections 4.4.3 and 4.4.5. For events passing electron

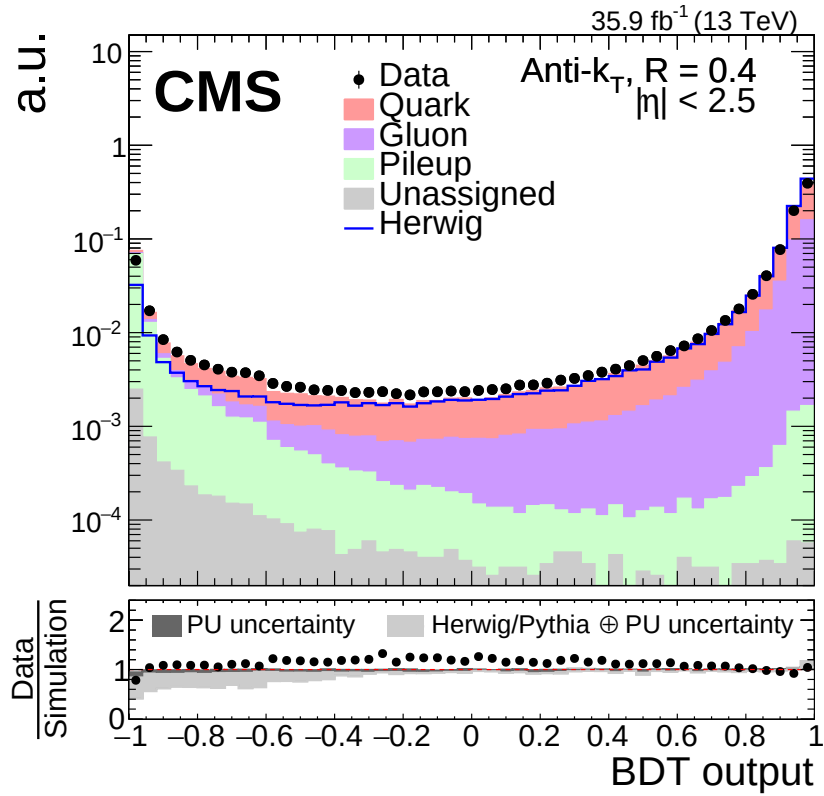


Figure 4.13: Output of the pileup jet ID multivariate discriminant for anti- k_T jets with cone size parameter $R = 0.4$, $30 < p_T < 50$ GeV, and $|\eta| < 2.5$. Taken from [119].

triggers, the selection attempts to find a pair of electrons. Likewise, events passing muon triggers are searched for a pair of muons. Up to small differences dictated by detector effects, the kinematic requirement on the candidates are identical in the two channels. This facilitates their combination later in the analysis. The considered leptons, selected using `medium` identification for electrons and `medium2016` for muons, are first required to be within the acceptance of the tracker, $|\eta| < 2.4$. This guarantees a good quality of the tracks used in the reconstruction and efficient isolation. For electrons, the pseudorapidity of the ECAL supercluster η_{SC} is used. Furthermore, electrons reconstructed within the barrel-endcap transition region $1.4442 < |\eta_{\text{SC}}| < 1.566$ are not considered.

The p_{T} requirements used in the selection should be as low as possible in order to maximize the acceptance, while being restrictive enough to be on the plateau of the trigger efficiencies. The most constraining trigger is the one used in the electron channel, for which the online p_{T} cuts are 23 and 12 GeV. For the leading lepton, the lowest p_{T} cut that can reasonably be used with this trigger is 25 GeV, as has been the case for the on-shell measurement [79]. This guarantees that the first leg of the trigger is fully efficient for central rapidities (see figure 4.6). This p_{T} cut of 25 GeV has implications on the smallest dilepton mass accessible to the measurement because the Jacobian peak in lepton p_{T} distributions is located at $p_{\text{T}}(\ell) = \frac{m_{\ell\ell}}{2}$: the smallest mass for which a meaningful measurement can be performed without large acceptance corrections is 50 GeV.

The cut for the second lepton in the pair is often set to the same value as for the leading one. At low masses of the pair, this choice would however reject most events, as can be seen in figure 4.14. In order to optimize the acceptance in this region, asymmetric cuts are used. The subleading lepton is required to have $p_{\text{T}} > 20$ GeV, which increases the acceptance while still providing good efficiency for trigger, reconstruction and identification.

Events with two leptons passing the aforementioned η_{SC} and p_{T} cuts are considered for analysis. Additional cuts are used to reduce backgrounds. First, the two leptons are required to carry opposite charges, which reduces the contribution from fake leptons as discussed in section 4.6.1. The second cut consists in discarding events with a third isolated lepton. Electrons (muons) used for this veto are required to pass `veto (loose)` identification and have a p_{T} of at least 10 GeV. The same η_{SC} cuts as for signal leptons are used. This selection slightly reduces the contamination from the ZZ and ZW background processes.

In addition to the inclusive measurement of dilepton p_{T} and φ_{η}^* , a measurement of $p_{\text{T}}(\ell\ell)$ when at least one jet is present is performed. The jet used for this selection is required to have a transverse momentum larger than 30 GeV and a rapidity within the tracker acceptance, $|y| < 2.4$. Pileup jets are suppressed by the relatively large p_{T} cut and by requiring a value of the pileup jet ID larger

Table 4.3: Summary of the object and event selection used in the analysis.

Two electrons	Double electron trigger medium identification $ \eta_{\text{SC}} < 2.4$ $ \eta_{\text{SC}} \notin [1.4442, 1.566]$ $p_{\text{T}} > 25, 20 \text{ GeV}$ — or —
Two muons	Single muon or double muon trigger medium2016 identification tight isolation ($I_{\text{PF}}^{\text{rel}} < 0.15$) $ \eta < 2.4$ $p_{\text{T}} > 25, 20 \text{ GeV}$ — with —
Lepton pair	Opposite charges $50 < m_{\ell\ell} < 1000 \text{ GeV}$
Jets	Anti- k_{T} , $R = 0.4$ loose identification $ y < 2.4$ $p_{\text{T}} > 30 \text{ GeV}$ Pileup jet ID > -0.2
No additional electron	veto identification $ \eta_{\text{SC}} < 2.4$ $ \eta_{\text{SC}} \notin [1.4442, 1.566]$ $p_{\text{T}} > 10 \text{ GeV}$
No additional muon	loose identification tight isolation ($I_{\text{PF}}^{\text{rel}} < 0.15$) $ \eta < 2.4$ $p_{\text{T}} > 10 \text{ GeV}$
No b-tagged jet	CSVv2 medium $ y < 2.4$ $p_{\text{T}} > 20 \text{ GeV}$ Pileup jet ID > -0.2

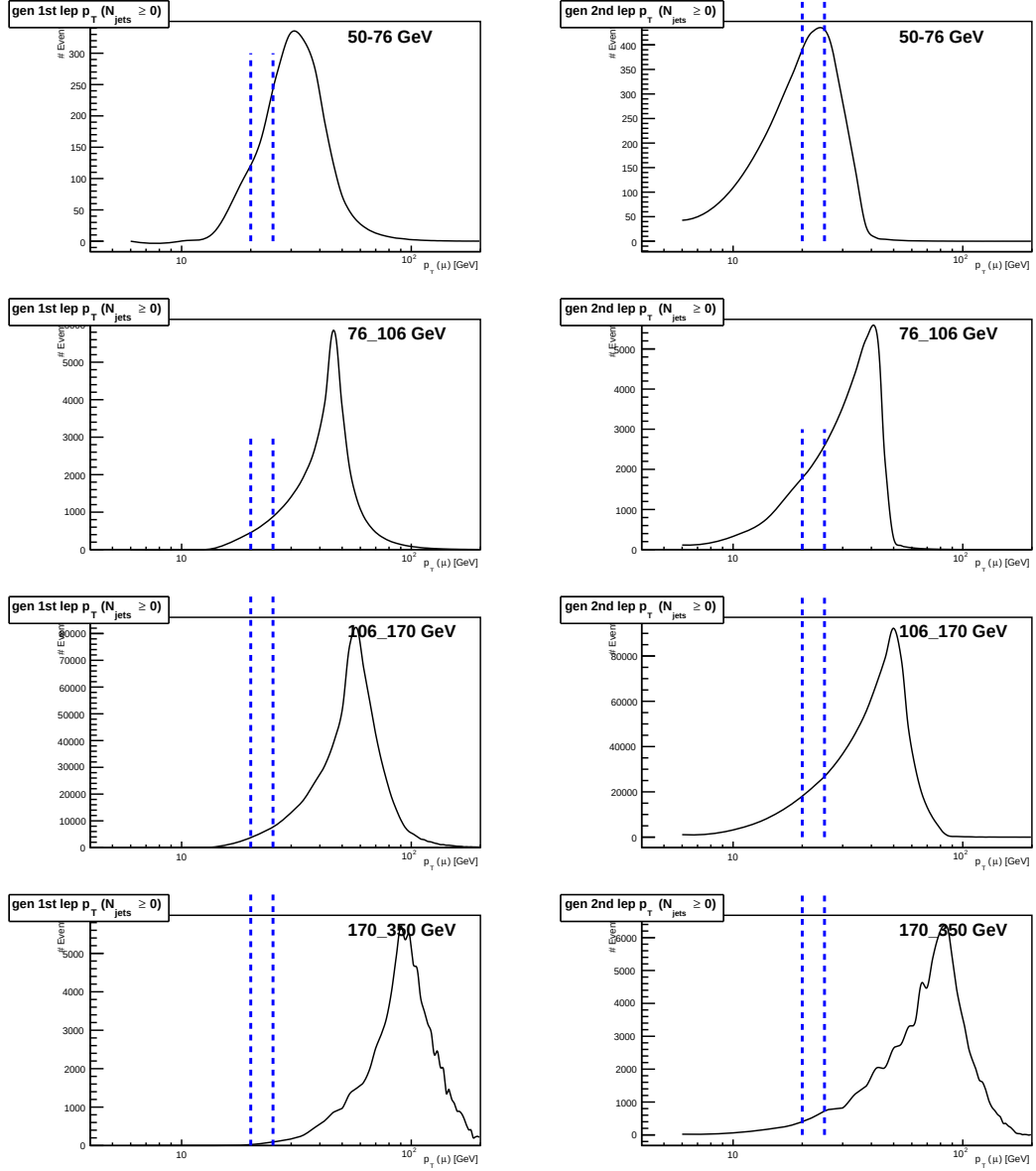


Figure 4.14: Distributions of the generated lepton p_T for the leading (left) and subleading (right) leptons, for several masses as indicated in the legends. The blue lines show possible values for the cuts at 20 and 25 GeV.

than -0.2 . This cut was optimized to produce a stable average number of jet per reconstructed Z boson with respect to the number of reconstructed vertices. It falls between the `loose` and `medium` cuts defined by the JETMET POG, for which no efficiency correction is necessary [122].

In order to reduce contamination from the production of pairs of top quarks, selected events are required to contain no b-tagged jet. Jets used for this veto follow the same requirements as listed above, except that their transverse momentum is allowed to be smaller, down to 20 GeV. The `medium` working point of the CSVv2 tagger is used, as will be further justified in section 4.6.2. The efficiency of this requirement in simulation is corrected for by reweighting selected events by a weight w whose calculation follows method 1a in the BTV POG recommendations [123]; it is summarized below. The POG provides measurements of the ratio of the b tagging efficiency in data and simulation, $SF = \varepsilon_{\text{data}}/\varepsilon_{\text{MC}}$. These numbers are used to scale simulated events on a jet-by-jet basis: for each tagged jet, the event is scaled by the scale factor; for untagged jets, it is scaled by the ratios of the simulated and measured non-tagging efficiencies. The final event weight w is calculated as follows:

$$w = \prod_{i=\text{tagged}} SF_i \prod_{j=\text{not tagged}} \frac{1 - \varepsilon_j}{1 - SF_j \varepsilon_j}. \quad (4.9)$$

The tagging efficiencies ε_j have to be computed separately for the phase space of each analysis because they depend strongly on the average composition of the jets, and in particular on the fraction of gluon jets. They are extracted from simulation in bins of η and p_T .

The overall efficiency of the event reconstruction and selection, as calculated in simulation by dividing the reconstructed-level prediction by the generator-level cross section in the fiducial phase space of section 4.7.1, is shown in figure 4.15. The wiggles around $m_{\ell\ell} = 90$ GeV are caused by the finite resolution of the mass measurement at reconstructed level. In the electron channel, the efficiency is about 35 % at $m_{\ell\ell} = 50$ GeV and reaches a plateau above 150 GeV, with $\varepsilon = 54$ %. The efficiency is larger and varies less in the muon channel, between 80 % at $m_{\ell\ell} = 50$ GeV and 83 % on the plateau reached above the Z peak.

4.4.8 Backgrounds

All but one background contributions in the phase space described above are estimated using Monte-Carlo samples for which the same detector simulation as for signal events was used. The same reconstruction algorithms and corrections as in the signal samples are used. In order to obtain accurate predictions of background contributions, the cross sections used for the normalization of the samples are where applicable calculated at NLO or NNLO in the strong coupling constant.

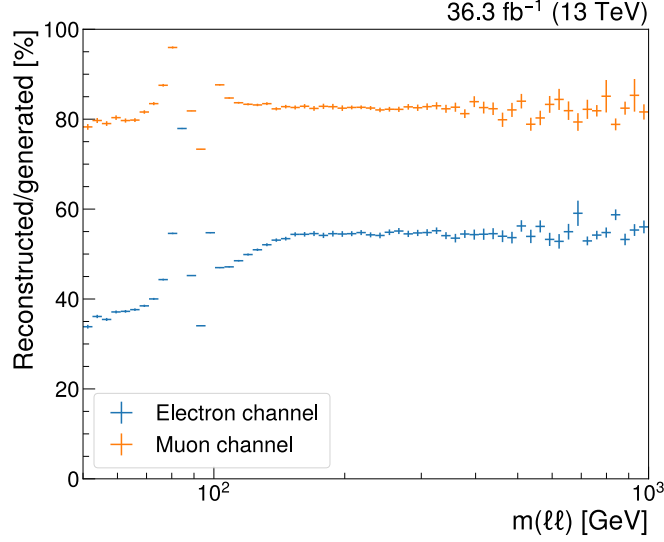


Figure 4.15: Efficiency of event reconstruction and selection as a function of the dilepton invariant mass, calculated in simulation by dividing the number of reconstructed events by the number of generated events in the each bin. The electron and muon channels are shown. The error bars represent the statistical uncertainties.

The dominant background process is the production of a pair of top quarks that subsequently decay leptonically as shown in figure 4.16a. It is simulated at NLO in α_s using version two of the POWHEG BOX [38, 124, 125, 126] event generator and normalized to the cross section calculated at NNLO + NNLL using TOP++ version 2.0 [127]. The parton shower and hadronization are performed using PYTHIA8 [43] using the CUETP8M2T4 tune [42]. Related processes are produced in the same setup: single top production in the t channel [128] and in association with a W boson [129], shown in figure 4.16b. Finally, the production of a single top quark in the s channel, with leptonic decay, is simulated at NLO in α_s using MADGRAPH5_AMC@NLO matched to PYTHIA8 using the CUETP8M1 tune [39]. These processes are normalized to the cross sections listed in table 4.4. The single top cross sections are calculated at NLO in α_s using HATHOR version 2.1 [130, 131] and the cross section of tW associated production is taken from [132].

The production of a Z boson in association with another electroweak boson (Z or W), shown in figure 4.16c, is simulated at leading order with PYTHIA8 using the CUETP8M1 tune and normalized to the NLO cross sections shown in table 4.4 calculated using MCFM 6.6 [133]. The production of a pair of W bosons is simulated at NLO in α_s using the POWHEG BOX [134] interfaced to

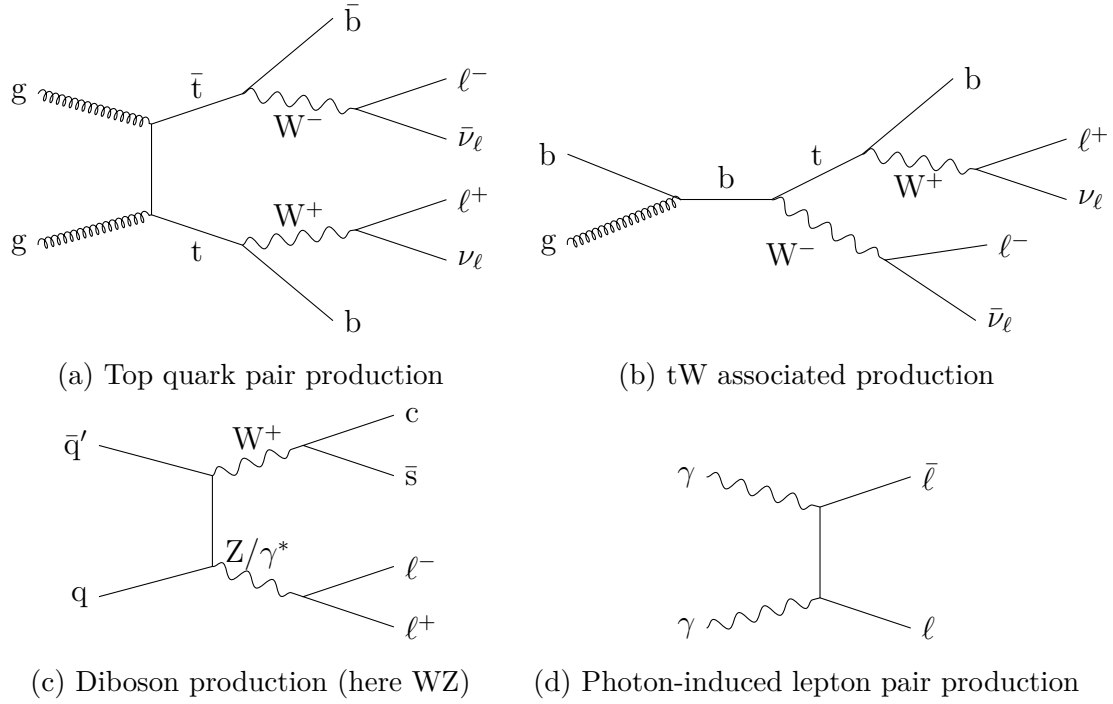


Figure 4.16: Example Feynman diagrams for four background processes considered in the analysis.

Table 4.4: List of background samples used in the analysis, their generator and the cross section used.

Process	Generator	$\sigma \cdot \mathcal{B}$ [pb]
$t\bar{t}$	POWHEG	831.7
tW	POWHEG	35.6
$\bar{t}W$	POWHEG	35.6
$t \rightarrow \ell X$ (s channel)	AMC@NLO	10.32
$\bar{t} \rightarrow X$ (t channel)	POWHEG	80.5
$t \rightarrow X$ (t channel)	POWHEG	136.02
$WW \rightarrow 2\ell 2\nu$	POWHEG	12.21
ZZ	PYTHIA8	15.4
WZ	PYTHIA8	23.5
$\gamma\gamma \rightarrow ee, \mu\mu$	CEPGEN	Table 4.5

PYTHIA8 and normalized to the NNLO cross section calculated in [135].

The photon-induced production of a lepton pair, shown in figure 4.16d, is simulated using the method from LPAIR [136, 137] as reimplemented in the CEP-GEN [138] event generator, interfaced to PYTHIA6 [35]. Three cases are considered depending on whether the protons are broken in the process: fully elastic, elastic-inelastic, and fully inelastic. PYTHIA6 is used only for inelastic cases. The samples are produced for several dilepton invariant mass ranges and normalized to the cross sections predicted by the generator listed in table 4.5.

The Z decay to a pair of τ leptons is considered a background and is handled as part of the unfolding procedure described in section 4.7. As a final background contribution, the misidentification of a jet as an electron is estimated using a data-driven method based on e^+e^+ and e^-e^- events, as described in section 4.6.1. The corresponding misidentification of a jet as a muon is negligible.

4.5 Kinematic distributions

In order to understand the background contamination and the available statistics, it is useful to compare kinematic distributions of measured and simulated events. This is done by superimposing the observed number of events with the prediction from our background estimation and signal model, presented as a stack of histograms. The distributions cover variables sensitive to detector effects to check that they are well understood and modeled in simulation and that no significant background contribution is missing. Obtaining fair agreement between data and simulation is also needed to ensure that efficiency corrections will be correct even after averaging them over variables that are not measured.

After distributions sensitive to reconstruction of the jets and leptons, the histograms used as input for the measurement will be presented. They are used to

Table 4.5: Cross sections used for the three components of the photon-induced background. The relative uncertainty in these numbers is 6 %.

$m_{\ell\ell}$ [GeV]	Per channel σ [pb]		
	Elastic	Elastic-inelastic	Inelastic
50 – 120	0.173	0.454	0.311
120 – 200	1.93×10^{-2}	6.12×10^{-2}	5.02×10^{-2}
200 – 400	2.26×10^{-3}	1.91×10^{-2}	1.75×10^{-2}
400 – 800	3.37×10^{-4}	2.54×10^{-3}	2.53×10^{-3}
800 – 1400	5.48×10^{-5}	2.27×10^{-4}	2.38×10^{-4}

check that the signal dominates the backgrounds and that bins are wide enough to avoid statistical fluctuations. The plots are reconstructed level also provide a qualitative comparison to a prediction — although the level of agreement cannot be inferred from the plots as only statistical uncertainties are shown at this stage.

4.5.1 Individual objects

Control plots for individual electrons and muons are shown in figure 4.17 as a function of the p_T and pseudorapidity. The histograms are filled once per selected lepton, i.e. twice for every event. The data is shown as black dots and the prediction from signal and background Monte-Carlo samples as stacked colored histograms. The background contributions are small because the selection is restricted to the Z peak region. The number of events is shown in a logarithmic scale to make the backgrounds and the tails of the distributions visible. To facilitate comparisons, the ratio of simulation over data is shown in the bottom panel.

Since the lepton energy and efficiency corrections depend on pseudorapidity, the η distributions are sensitive to a misdescription of these in simulation. For instance, figure 4.8 shows that prefiring in the ECAL only happens in the endcaps. If this effect was neglected, a disagreement between data and simulation would be visible at large $|\eta_e|$. The pseudorapidity distributions shown in figure 4.17 are similar between data and simulation and between the channels. The gaps visible in the electron channel around $|\eta_e| = 1.5$ are caused by the removal of the barrel-endcap transition region. As expected due to the larger efficiency in the muon channel, the event yields are different between electrons and muons.

The distributions of the electron and muon p_T exhibit a rising spectrum up to about $p_T = 45$ GeV before rapidly falling towards larger p_T . The slope around 45 GeV is dictated by the shape of the invariant mass distribution around the Z pole. At large p_T , the slope of the distribution becomes smaller. In this region, $p_T(\ell)$ is related to the transverse momentum of the pair. The ratios show that the simulation describes the p_T distribution relatively well. Below 45 GeV, the Monte-Carlo undershoots the data, and above this value the yield is overestimated. It shall be noted that this discrepancy is enhanced by the large slope of the distribution³. The same effect is visible in both channels, indicating that this behavior is not caused by a misdescription of detector effects — this will be further quantified in section 4.9.1.

³Consider two distributions $f(x)$ and $g(x)$ with g slightly shifted with respect to f : $g(x) = f(x - \delta x)$. The ratio of the two is approximately given by:

$$\frac{g(x)}{f(x)} \approx \frac{f(x) - f'(x)\delta x}{f(x)} = 1 - \frac{f'(x)}{f(x)}\delta x,$$

i.e. the small difference δx is enhanced by a factor $f'(x)/f(x)$.

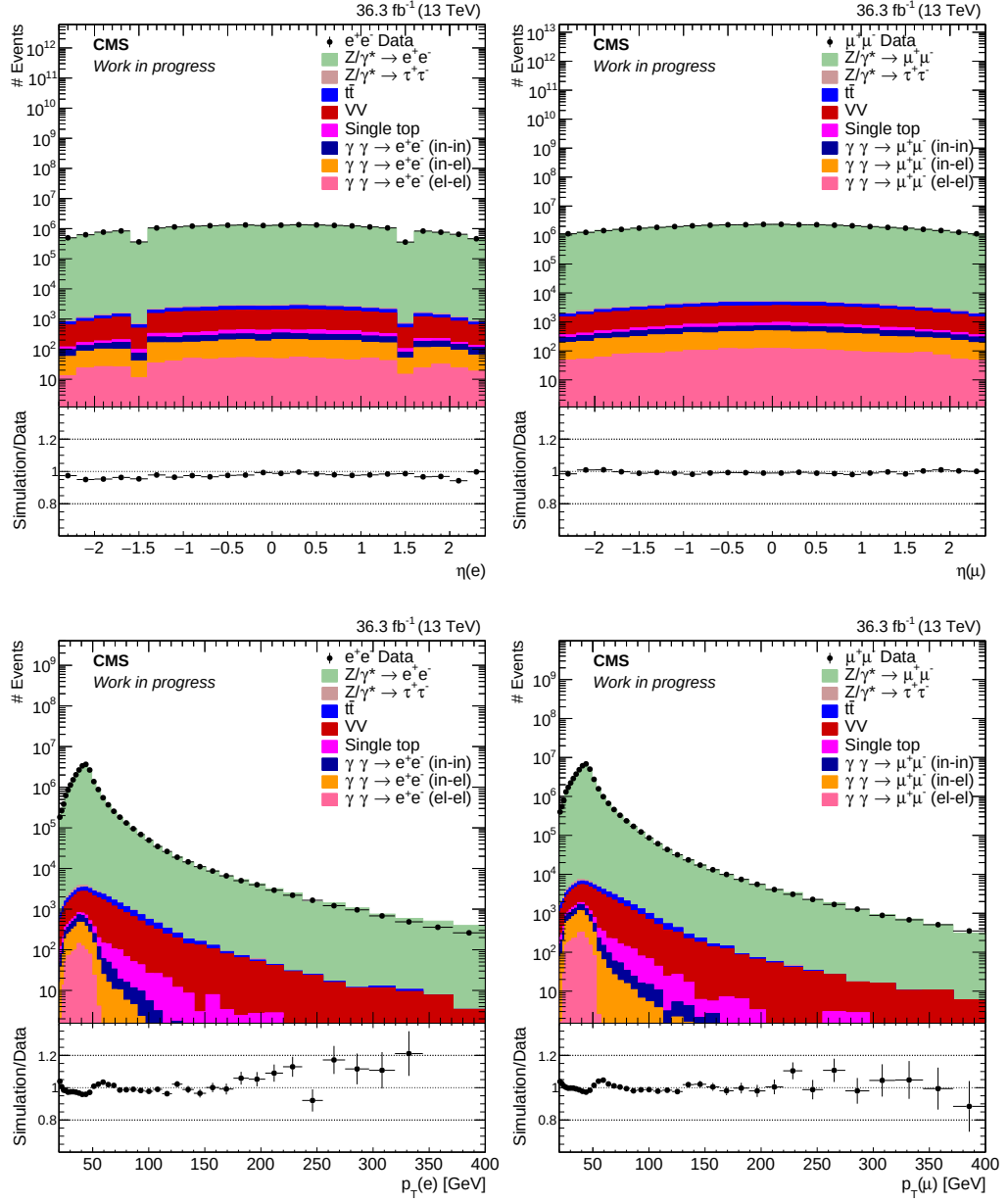


Figure 4.17: Electron (left) and muon (right) pseudorapidity (top) and transverse momentum (bottom) distributions for events with $76 < m_{\ell\ell} < 106$ GeV.

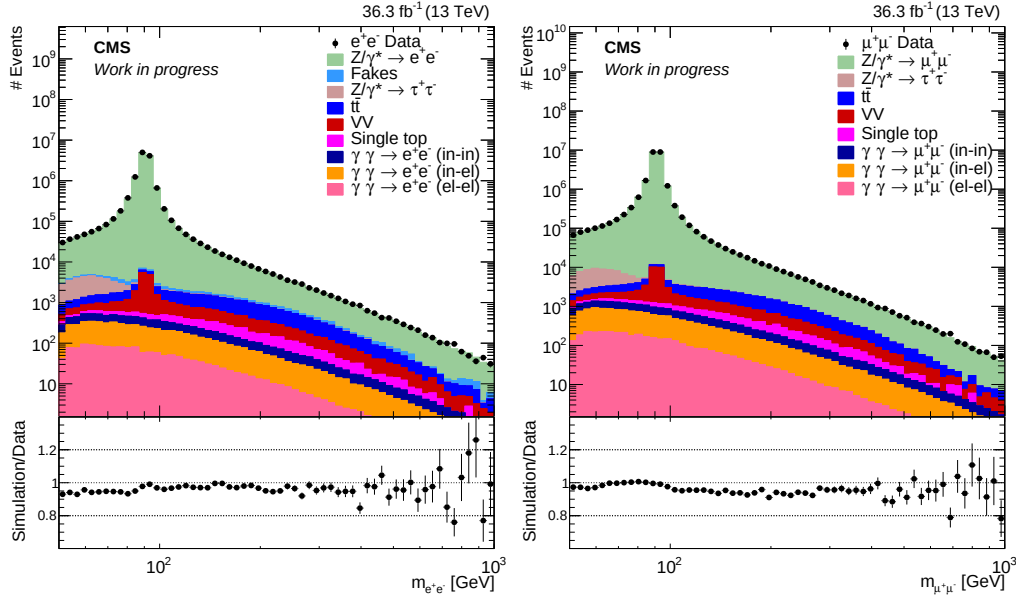


Figure 4.18: Invariant mass of lepton pairs in the electron (left) and muon (right) channels.

The level of description out of the Z peak region is probed by plotting the distribution of the invariant mass of the lepton pair, $m_{\ell\ell}$, shown in figure 4.18. The Z peak is visible around $m_{\ell\ell} = 91$ GeV. The comparison between data and simulation show that the levels of description achieved in and out of the Z peak region are similar, with no significant feature visible in the ratios. One can also see that backgrounds become non-negligible away from the peak: in particular, $Z \rightarrow \tau\tau$ amounts to 10 % of the observed events below the peak and $t\bar{t}$ production becomes important at high mass. Together with the distributions of lepton η and p_T , these distributions show that the simulation describes the main features in the lepton kinematic distributions correctly.

Jets are used in the analysis to define the selection used for the measurement of the $p_T(\ell\ell)$ distribution with one jet and to implement the veto on b jets. Three jet kinematic distributions are shown in figure 4.19: the number of jets present in the events and the rapidity and transverse momentum of the jet with the largest p_T (“leading”). As for the leptons, the selection is restricted to the Z peak region. The distribution of the number of jets decreases steadily when more jets are required, the cross section being divided by a factor ~ 5 for each additional jet. The event yield in the Monte-Carlo provides a good description of the data up to 3 jets, above which it the ratio starts to drop. This behaviour was already noticed in measurements of Drell-Yan in association with jets [139, 140]. One can also see

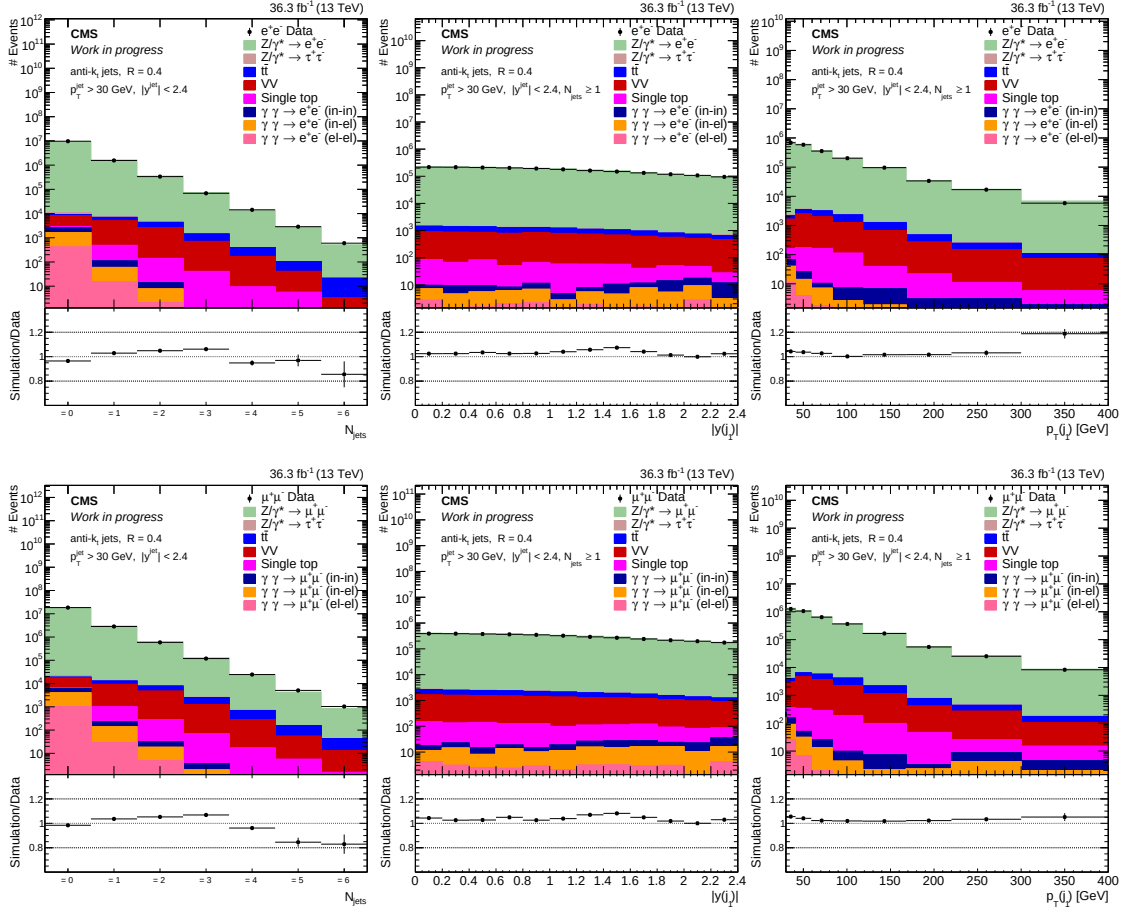


Figure 4.19: Jet kinematic distributions in the electron (top) and muon (bottom) channels for events with $76 < m_{\ell\ell} < 106$ GeV. From left to right: jet multiplicity, absolute rapidity and transverse momentum of the leading jet.

that the relative background contribution increases with the number of jets, with the dominant ones being $t\bar{t}$ and diboson production. The jet rapidity distributions don't present any remarkable feature and show that the yield is consistent between the barrel and endcaps. Finally, the p_T distributions of the leading jet are also well reproduced by simulation, showing in particular that the jet energy scale and resolution are consistent. The simulation thus reproduces the main jet kinematic distributions correctly.

4.5.2 Measured observables

Having gained confidence in the reconstruction procedure used for the leptons from which we build the Z boson and the jets that are part of the phase space definition, we now turn to reviewing the reconstructed level comparisons for the observables that will be used for the measurement: the p_T of the lepton pair, the φ_η^* variable and the p_T of the lepton pair with at least one jet. We will always start the discussion in the Z peak region before moving to the mass bins below and above.

Transverse momentum

The $p_T(\ell\ell)$ distributions are shown in figure 4.20 for the electron channel and figure 4.21 for the muon channel. The variable bin widths tend to flatten the distributions, whose shape on the plots is not representative of the differential cross section. The binning on the Z peak has been determined according to the detector resolution, and adjacent bins have been merged to produce the off-shell binnings.

In the Z peak region, the number of events from Drell-Yan is large: the bin with the smallest number of events contains about 450 events. Accordingly, the relative background contributions are small: the dominant one is from the diboson processes, and amount to a few percent in the high p_T . Below the Z peak, where there is less Drell-Yan data, several background contributions become significant: photon-induced processes in the first bins, the $Z \rightarrow \tau\tau$ process and, above 50 GeV, $t\bar{t}$ production. The total background fraction is approximately constant in p_T , at around 10%. With the exception of $Z \rightarrow \tau\tau$, the same background contributions are visible above the Z peak, totaling up to 35% in the high p_T tail of the last two mass bins.

φ_η^* variable

Figures 4.22 and 4.23 show the reconstructed level distributions of the φ_η^* variable. Since φ_η^* is highly correlated to $p_T(\ell\ell)$, $\varphi_\eta^* \sim p_T(\ell\ell)/m_{\ell\ell}$, the comments that we can

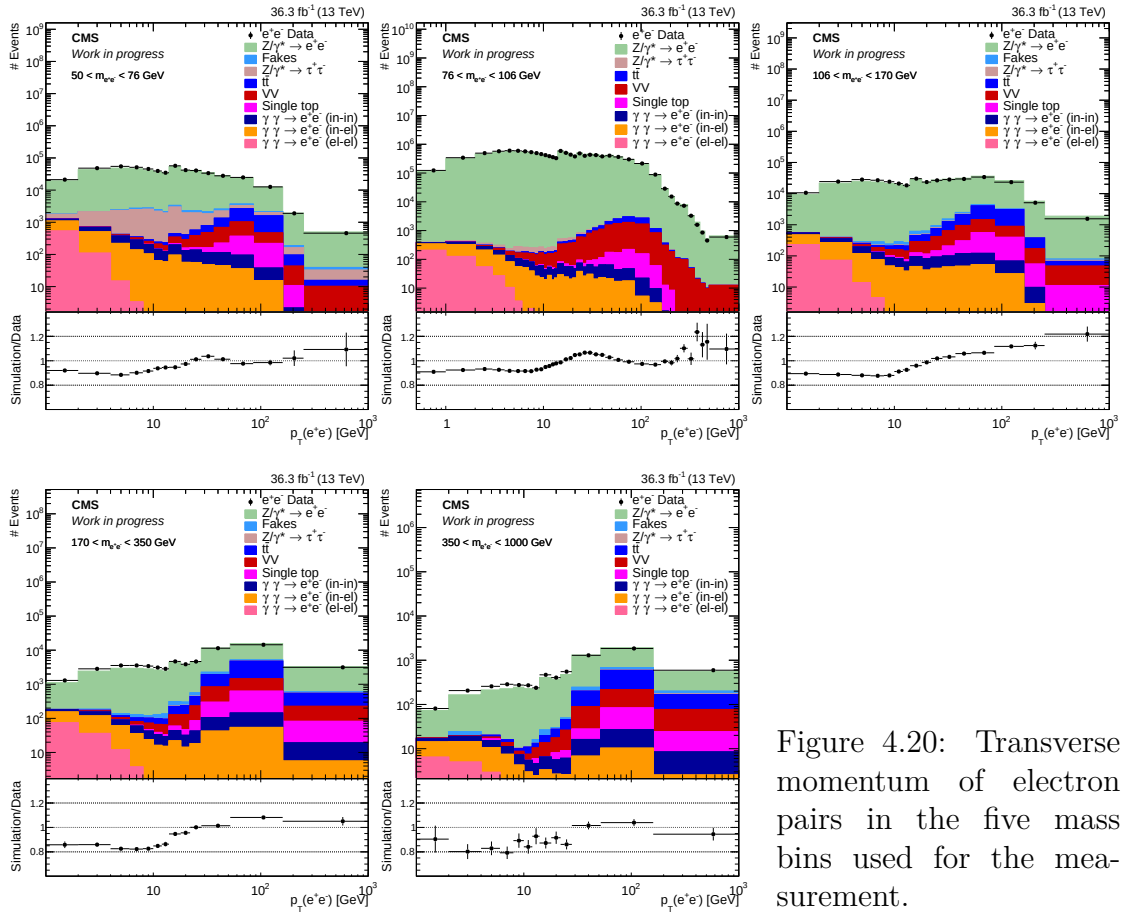


Figure 4.20: Transverse momentum of electron pairs in the five mass bins used for the measurement.

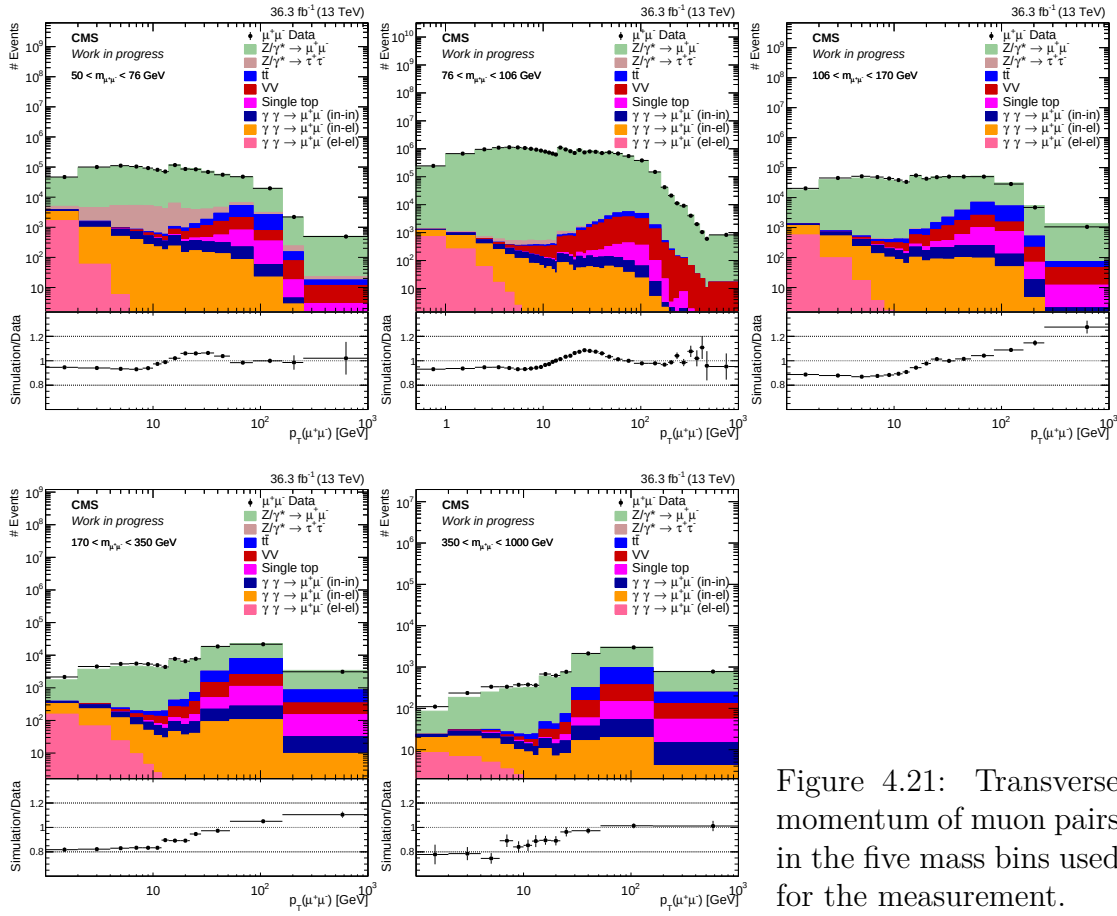


Figure 4.21: Transverse momentum of muon pairs in the five mass bins used for the measurement.

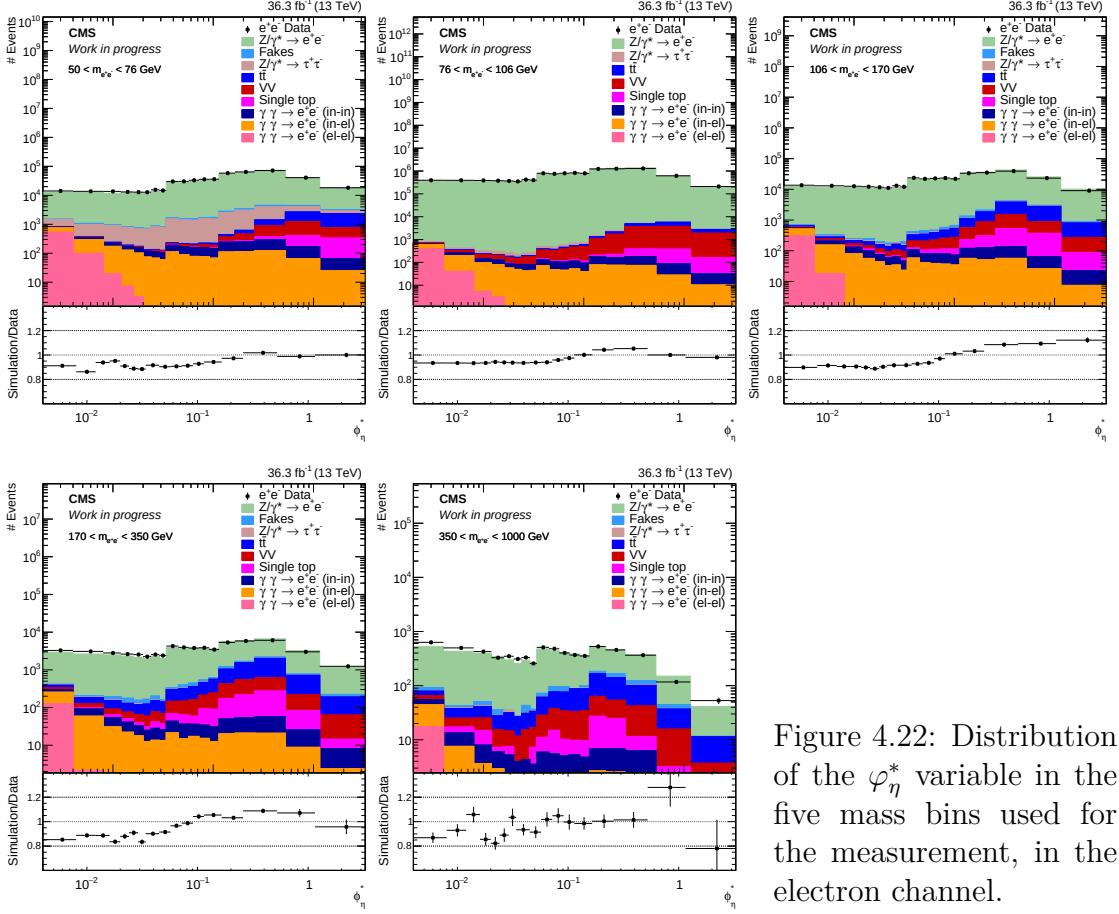


Figure 4.22: Distribution of the φ_η^* variable in the five mass bins used for the measurement, in the electron channel.

make here are very similar to what was said for p_T ; in particular, the background rates are identical. One remarkable observation is related to the behavior of the elastic photon-induced background at high mass: we can see that it is only present in the first p_T bin. This is due to the combination of two effects. First, φ_η^* is measured with greater precision than p_T : events produced with $p_T = \varphi_\eta^* = 0$ are often measured with $p_T \neq 0$, but fall much more often in the first φ_η^* bin; this is taken into account when unfolding. The second effect is more important for the physical interpretation of the results: since $\varphi_\eta^* \sim p_T/m_{\ell\ell}$, the whole p_T spectrum is compressed into the low φ_η^* region when $m_{\ell\ell}$ is large. At high mass, this reduces the discriminating power of φ_η^* with respect to p_T , especially for the low p_T end of the spectrum which is of particular interest in this analysis.

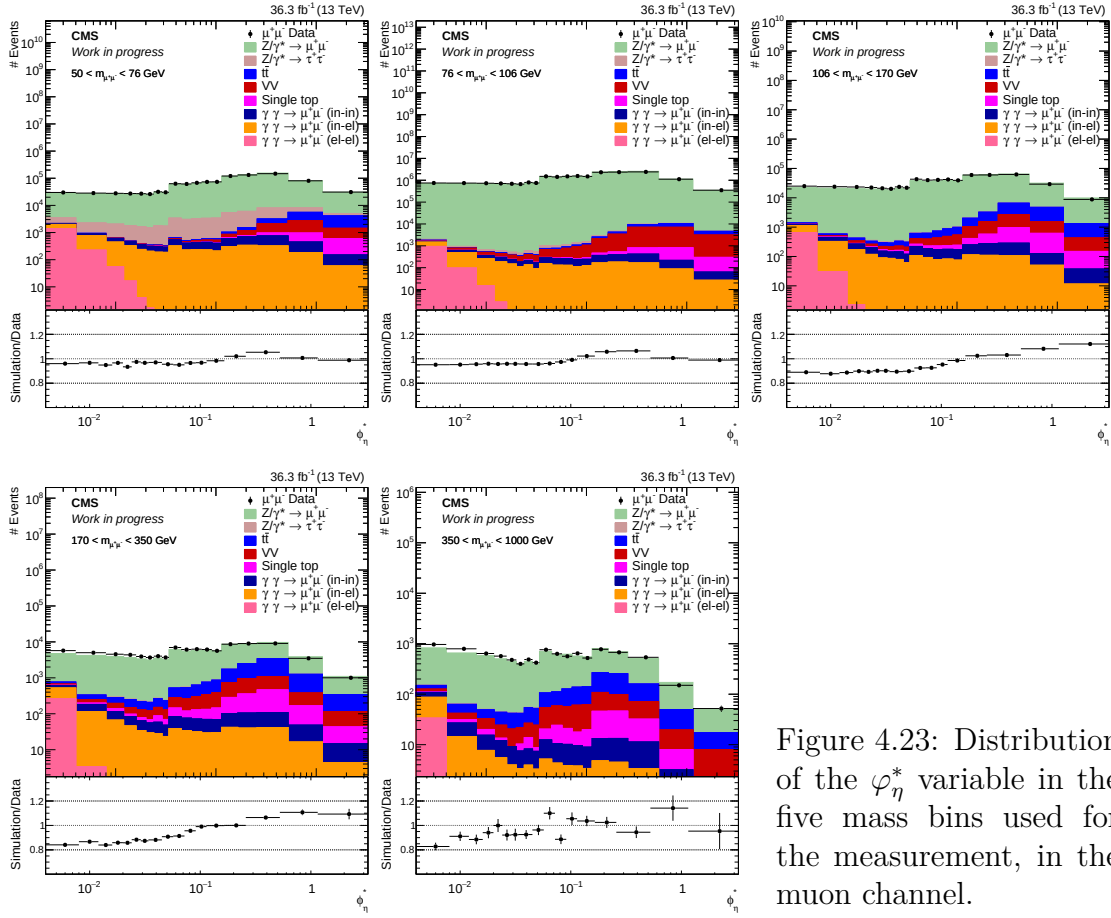


Figure 4.23: Distribution of the φ_η^* variable in the five mass bins used for the measurement, in the muon channel.

Transverse momentum, one jet

The last variable measured in the present analysis is the transverse momentum of the lepton pair, $p_T(\ell\ell)$, with an additional requirement of at least one jet in the event. The corresponding distributions at the reconstructed level are shown in figure 4.24 for the electron channel and figure 4.25 for the muon channel. These distributions seemingly peak at around 70 GeV due to the non-constant bin width; when dividing by the bin width, the peak is as expected around the jet p_T cut of 30 GeV.

The background contributions in this region of the phase space are similar to the inclusive ones described above. The resonant contributions from Drell-Yan and diboson still dominate the Z peak region. In the low mass region, $Z \rightarrow \tau\tau$ and $t\bar{t}$ production are the dominant backgrounds, with a total contribution of about 10 %. The biggest difference lies in the high mass, where $t\bar{t}$ is even more important than in the inclusive case. The presence of this background and the low statistics prevent the inclusion of the fifth mass bin, 350 to 1000 GeV, in the measurement.

4.6 Additional background checks

The background processes considered in the analysis have already been listed in section 4.4.8. This section provides more in-depth explanations for two of them: the details of the data-driven estimation of the contribution from “fake” electrons and cross checks of the description of processes involving a top quark by the simulation.

4.6.1 “Fake” electrons

The signature in the detector associated with an electron can easily be generated by other particles, for instance jets and converting photons. This leads to additional processes contributing to the electron channel: $W(\rightarrow e\bar{\nu}_e) + \text{jet}$ where the jet is misidentified as an electron or QCD events with the misidentification of two jets. Such electrons are said to be “fake” because they have no counterpart in the hard process; this section describes the procedure used to estimate their contribution.

The estimation of the contribution of fake electrons is based on events with two electrons of the same sign, e^+e^+ or e^-e^- . In the Standard Model, processes exist that can generate two electrons of the same sign: mainly diboson production, ZZ and ZW where the two bosons decay into electrons. The inclusive cross section for these processes are small and further suppressed by the requirement of a leptonic decay. Another level of suppression is achieved at the detector level by the requirement that at most two isolated leptons are reconstructed, rendering the contribution negligible as has been checked in the muon channel. As a result,

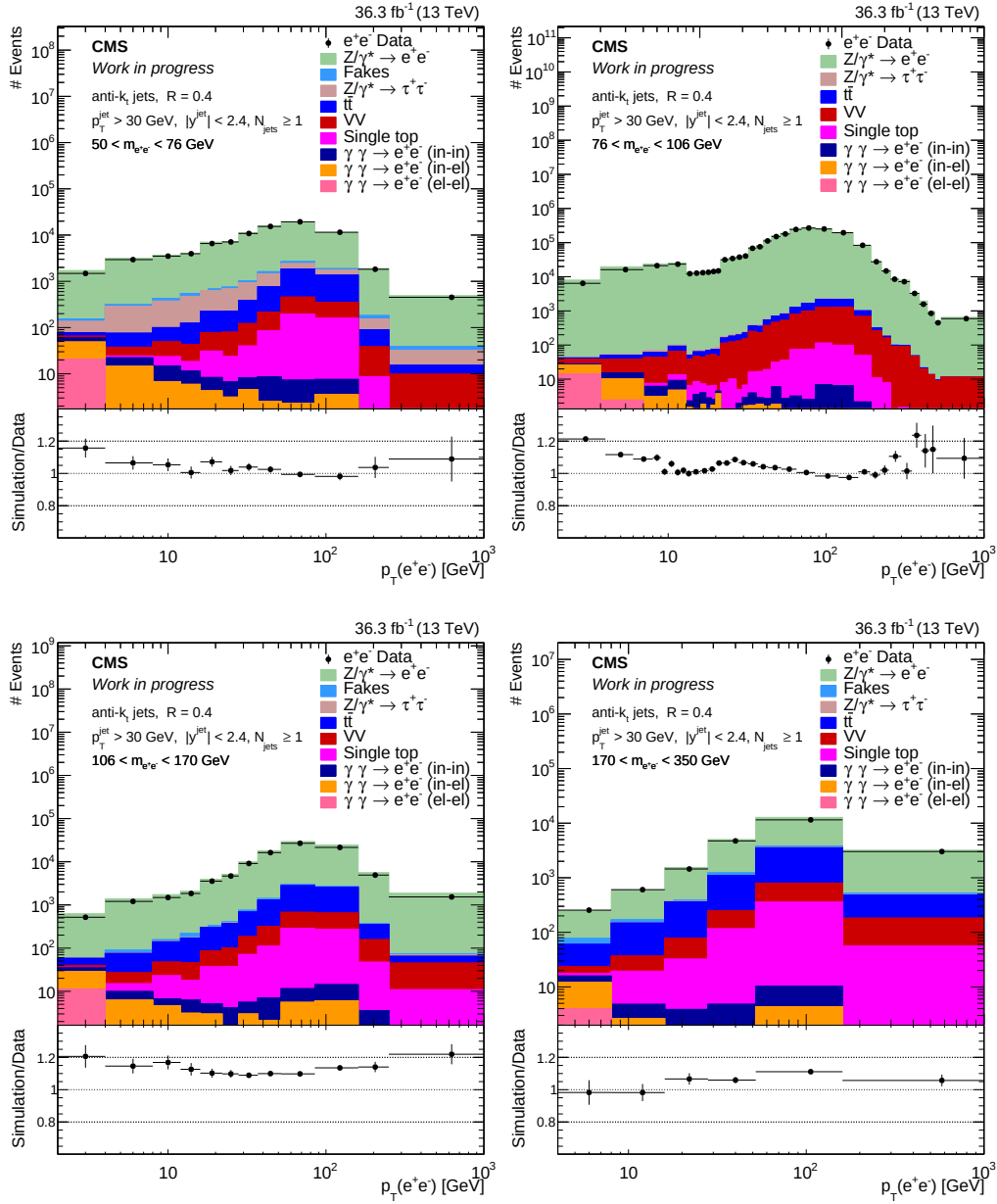


Figure 4.24: Distribution of the transverse momentum of electron pairs, with at least one jet required, in the four mass bins used for the measurement.

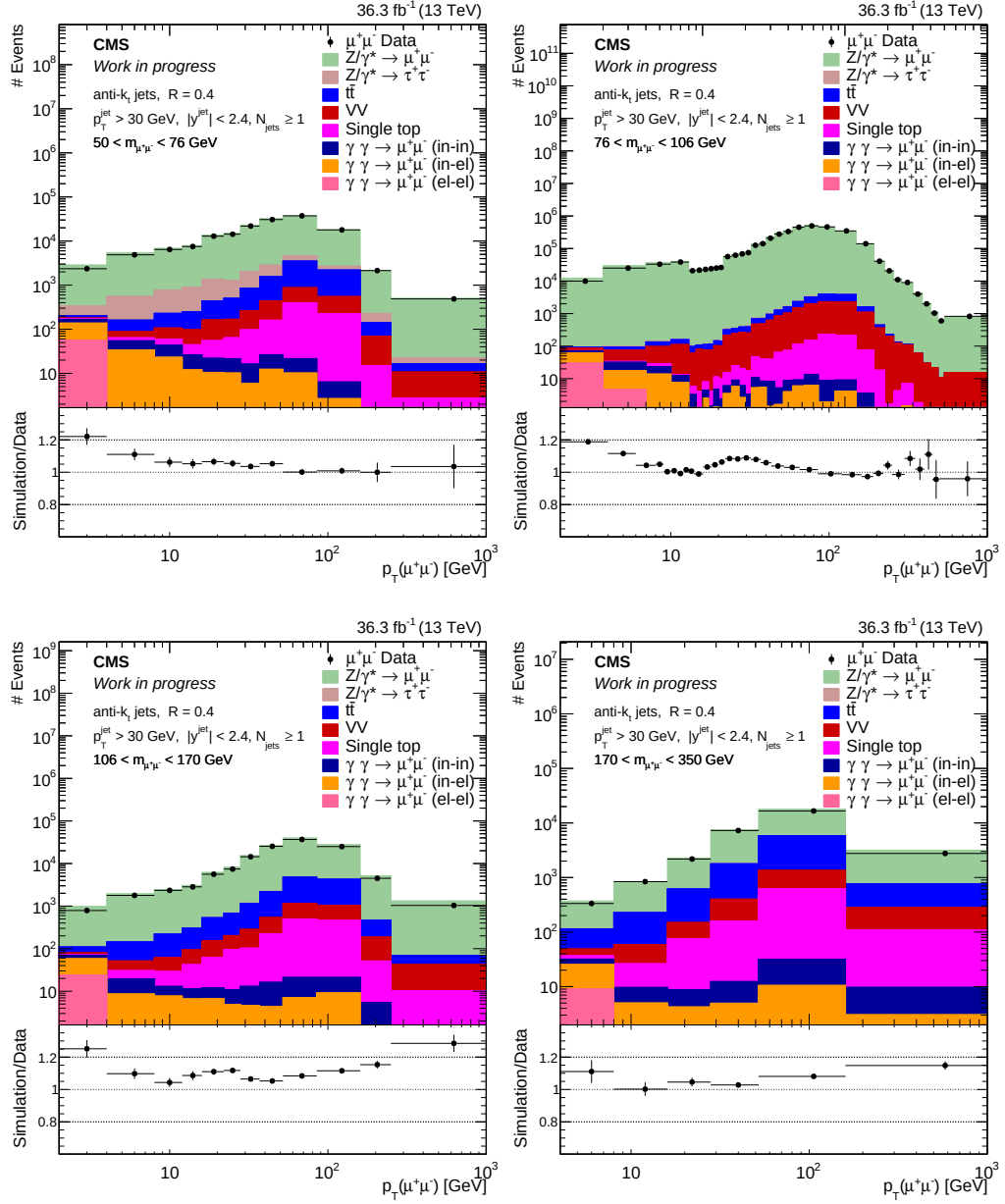


Figure 4.25: Distribution of the transverse momentum of muon pairs, with at least one jet required, in the four mass bins used for the measurement.

the contributions to e^+e^+ and e^-e^- events mostly contain two kinds of processes: e^+e^- events where the charge of one electron is misidentified and events with one or more particles mimicking the signature of an electron. By carefully studying the rate of charge misidentification, one can therefore isolate the contribution from fake electrons. This measurement can then be used as an estimate of the fake electron background in the e^+e^- (signal) region.

The invariant mass distribution of e^+e^+ and e^-e^- events is shown in figure 4.26 along with the prediction from the set of Monte-Carlo samples used for the signal region. The Z peak is visible both in data and in simulation. Since the invariant mass distribution of fake electrons is not expected to contain the Z resonance, this shows that charge misidentification is important. The simulation underestimates the data everywhere including on the Z peak where simulation underestimates the data by about 15 %. This is larger than the expected contribution of the background from fake electrons in the vicinity of the Z peak, which by looking at the sidebands is at most of the order of one percent. In order to isolate the contribution from fake electrons, a better description of charge misidentification needs to be obtained.

Charge misidentification

This section describes the treatment of charge misidentification. After studies in the Monte-Carlo, charge misidentification will be measured in data and simulation by comparing the same sign and opposite sign event yields in the Z peak region. As mentioned above, this region is essentially free of contamination from fake electrons. The results derived in this section will then be extrapolated to estimate the contribution of charge misidentification outside of the Z peak and thereby isolate the background from fake electrons.

The charge misidentification probability $P(e)$ of individual electrons is studied in the Monte-Carlo using the generator truth information. Reconstructed leptons are matched to their closest generated counterpart (in ΔR distance) and the charges are compared. The obtained probability is shown as a function of $|\eta|$ and p_T in figure 4.27 for the leading electron. The rate depends strongly on $|\eta|$ and is larger in the endcaps, where the lever arm usable to determine the track curvature is smaller and the amount of material in front of the ECAL is larger. A dependence on p_T is also visible. When focusing on individual $|\eta|$ bins, as exemplified in figure 4.28, one sees that $P(e)$ can be described by an affine function of p_T . These results are consistent across all studied dielectron invariant mass bins and between the leading and subleading electrons.

Given the probability $P(e)$ that a single electron is reconstructed with the wrong charge, the probability $P[\text{OS} \rightarrow \text{SS}]$ that an e^+e^- pair is reconstructed as

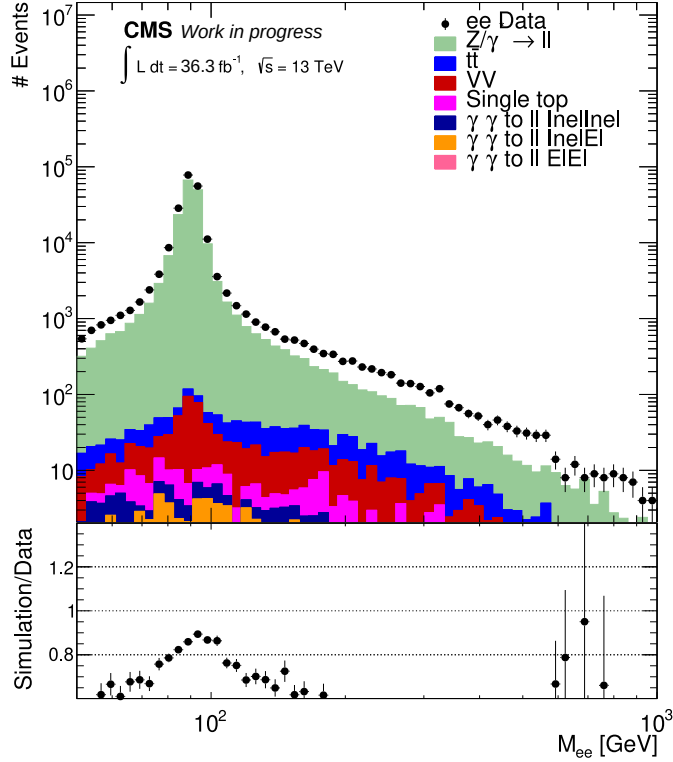


Figure 4.26: Dielectron invariant mass with the same sign selection, without correcting the description of charge misidentification in the Monte-Carlo.

two candidates of the same sign is modelled as follows:

$$P[\text{OS} \rightarrow \text{SS}] = P(e_1)(1 - P(e_2)) + P(e_2)(1 - P(e_1)). \quad (4.10)$$

One critical point to enable the extrapolation of the misidentification probabilities derived on the Z peak to higher mass is that a smooth function of p_T should be used for its modelling. Indeed, the average electron p_T at high mass is larger than in the Z peak region: using binned probabilities, for instance with the binning of figures 4.28, would not describe the evolution of the probability within bins. Even worse, because the lepton p_T distribution is steeply falling, the bin contents would be dominated by the probability in the low p_T part, resulting in a systematic underestimation of the $P(e)$ at high p_T . This problem does not appear at low p_T , where binned probabilities could in principle be used.

The observation of a linear behavior in figure 4.28 motivates the use of an affine function of p_T with piece-wise coefficients depending on $|\eta|$ for the charge

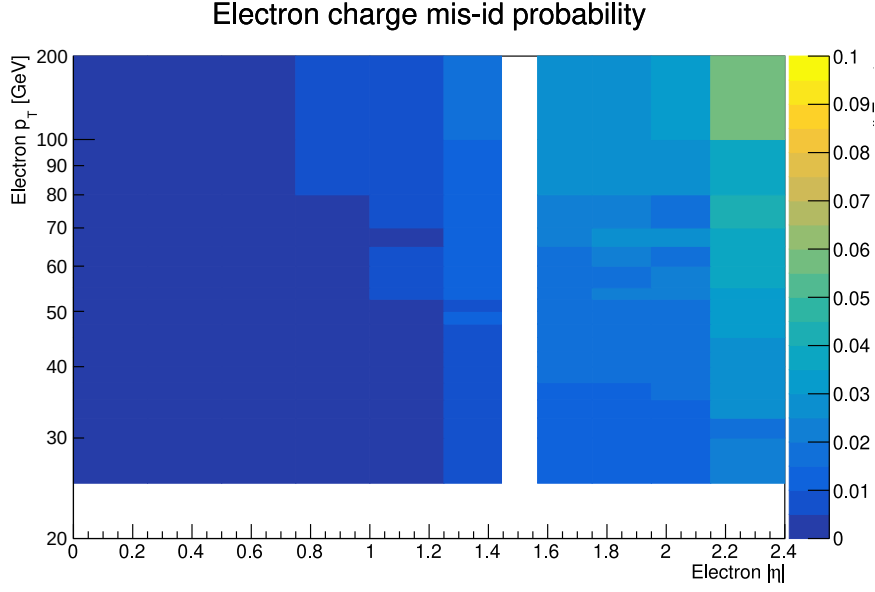


Figure 4.27: Charge misidentification probability for the leading electron in the Monte-Carlo, as a function of its $|\eta|$ and p_T .

misidentification probability:

$$P(e) = a(|\eta|)(p_T - 45 \text{ GeV}) + b(|\eta|). \quad (4.11)$$

The p_T shift by 45 GeV is such that b is representative of the average charge misidentification rate for on-shell events. The 20 parameters $a(|\eta|)$ and $b(|\eta|)$ are measured in data using the same binning as in figure 4.27. Their extraction is described below.

The values of a and b entering equation (4.11) are measured using reconstructed level events in the Z peak region, $m_{ee} \in \{91 \pm 10\}$ GeV. The small width of the window in invariant mass was chosen to minimize contributions from processes other than Drell-Yan in the same-sign selection. The values are extracted separately in data and simulation, by performing a simultaneous fit of all parameters. The target function of the fit is a χ^2 test between histograms in $(|\eta|, p_T)$ of the leading lepton in same sign data and opposite sign data scaled by $P[\text{OS} \rightarrow \text{SS}]$, schematically requiring that:

$$\text{same sign} = P[\text{OS} \rightarrow \text{SS}] \times \text{opposite sign}. \quad (4.12)$$

The rescaling of opposite sign events is performed event by event according to the kinematics of the two electrons. This allows the fit to be sensitive to both electrons even if the χ^2 test is performed using the kinematics of the leading one only. In

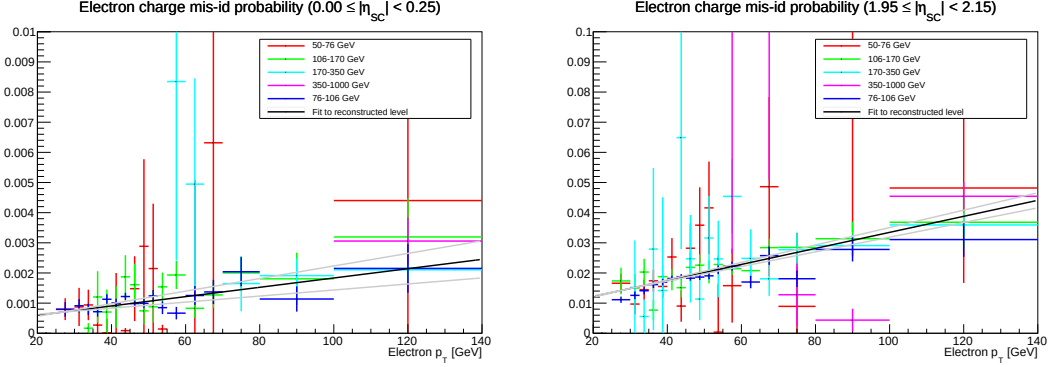


Figure 4.28: Electron charge misidentification probability for the leading electron in the Monte-Carlo, as a function of p_T in the most central (left) and forward (right) pseudorapidity bins. The colors represent different mass ranges and the black line illustrates the results of the fit to reconstructed level information; see the text for details.

addition, the event by event reweighting allows to obtain the best constraint from high- p_T electrons.

The results of each fit — in data and Monte-Carlo — are ten linear relations between $P(e)$ and $p_T(e)$, one for each $|\eta|$ bin. Two of them are shown in figure 4.28, where it can be seen that they reproduce the probabilities extracted using generator level information in the Monte-Carlo. Figure 4.29 summarizes the results for both data and Monte-Carlo. The fitted values of the coefficients are listed in table 4.6. These results show that the Monte-Carlo predicts a smaller charge misidentification probability than observed in most of the phase space, and especially in the forward regions. This is in line with the underestimation of the Z peak in the same-sign region in figure 4.26 that motivated a detailed study of charge misidentification.

Two techniques can be used to estimate charge misidentification using the measured values of a and b . In the first, the Monte-Carlo is corrected by reweighting opposite-sign events by a scale factor corresponding to the ratio of $P[\text{OS} \rightarrow \text{SS}]$ in data and Monte-Carlo. The second method is purely data driven: it uses the observed e^+e^- data, scaling each event by $P[\text{OS} \rightarrow \text{SS}]$ to account for the misidentification probability. A closure test is performed by comparing the two methods with the observed same sign data in the Z peak region from which the corrections were derived, as is done in figure 4.30 for the $p_T(ee)$ and invariant mass distributions.

The $p_T(\ell\ell)$ distribution shown in figure 4.30 shows that the overall normalization is correct for both estimates. The residual disagreement for the Monte-Carlo based estimate is consistent with what is seen in the opposite sign data and is

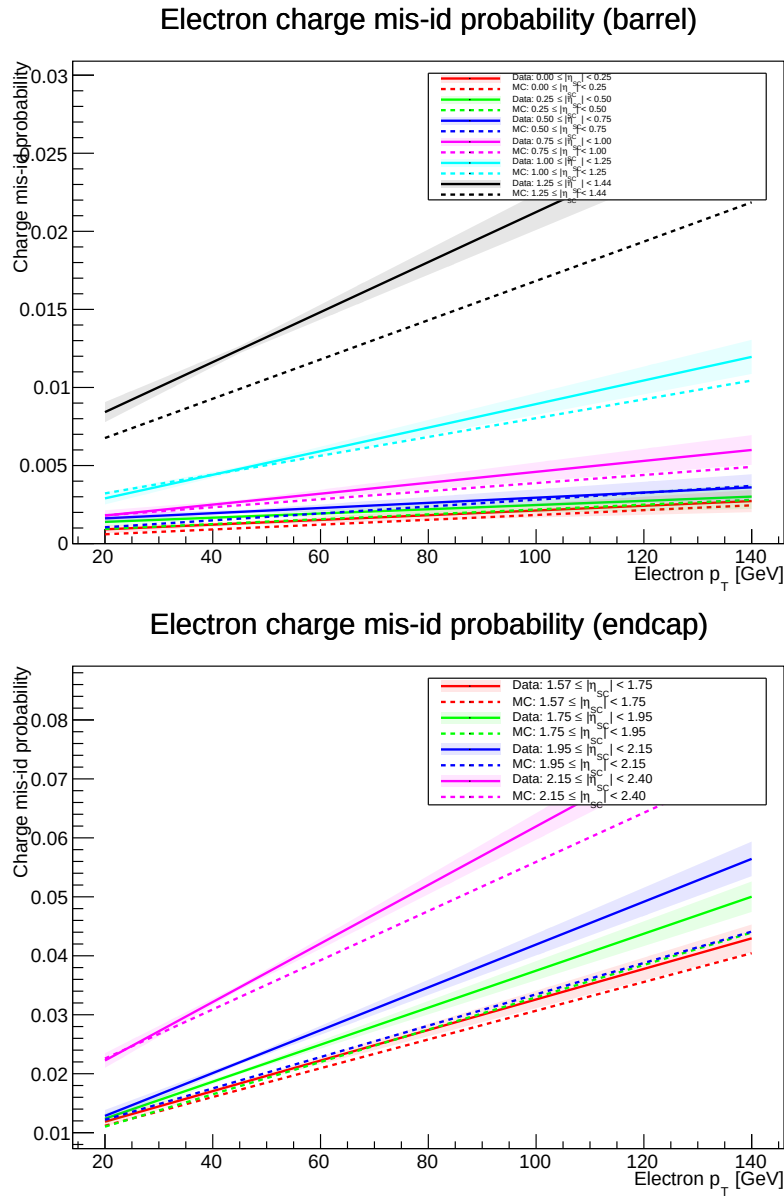


Figure 4.29: Electron charge misidentification probabilities as extracted from reconstructed level information, in the barrel (top) and endcap (bottom). Results in the Monte-Carlo are shown as dashed lines and results in data as solid colored lines. The uncertainty bands correspond to the statistical uncertainty of the fit.

attributed to mismodeling of the distribution by the Monte-Carlo at the generator level. The data-driven estimate provides a better description of the data and has a better statistical power. In the invariant mass distribution, the Monte-Carlo based estimate behaves better than the data-driven one around the Z pole, while they give similar results away from the peak. The difference is caused by a systematic bias in the electron energy scale for charge misidentified electrons. The effect on the electron energy is about 2%, which is in our case negligible. The closure tests presented in figure 4.30 show that both the Monte-Carlo based and the data-driven methods are valid to estimate charge misidentification for the distributions measured in this analysis. Since the data-driven estimate provides a better statistical power and does not depend on the predicted $p_T(\ell\ell)$ distribution, it is preferred over the correction of the Monte-Carlo.

The uncertainty bands shown in figure 4.29 correspond to the statistical uncertainty extracted from the fit. In order to check the stability of the results, the fit is repeated with different cuts on the invariant mass, shifting each end of the considered range by ± 5 GeV. The observed deviations of the fit results were within the statistical uncertainty, indicating that the latter provides a good estimate of the total uncertainty. In order to facilitate the treatment in the later stages of the analysis, a conservative value of 20% was assigned to the uncertainty on the charge misidentification estimate. This was motivated by the statistical uncertainty being of that order at $p_T(e) \sim 300$ GeV — a value above which there are very few events in the main selection.

Table 4.6: Parameters of the electron charge misidentification probability fit as extracted from data and Monte-Carlo.

$ \eta_{SC} $	Data		Simulation	
	a [%/100 GeV]	b [%]	a [%/100 GeV]	b [%]
0.0000 – 0.2500	0.151 ± 0.065	0.129 ± 0.009	0.155 ± 0.054	0.098 ± 0.010
0.2500 – 0.5000	0.135 ± 0.072	0.173 ± 0.010	0.155 ± 0.060	0.133 ± 0.011
0.5000 – 0.7500	0.165 ± 0.077	0.203 ± 0.011	0.221 ± 0.065	0.160 ± 0.012
0.7500 – 1.0000	0.350 ± 0.087	0.267 ± 0.012	0.259 ± 0.074	0.246 ± 0.013
1.0000 – 1.2500	0.755 ± 0.100	0.479 ± 0.014	0.602 ± 0.088	0.472 ± 0.015
1.2500 – 1.4442	1.601 ± 0.167	1.242 ± 0.023	1.258 ± 0.138	0.991 ± 0.023
1.5660 – 1.7500	2.591 ± 0.217	1.834 ± 0.029	2.438 ± 0.197	1.727 ± 0.032
1.7500 – 1.9500	3.137 ± 0.241	2.022 ± 0.030	2.743 ± 0.204	1.788 ± 0.031
1.9500 – 2.1500	3.632 ± 0.274	2.193 ± 0.034	2.659 ± 0.229	1.884 ± 0.034
2.1500 – 2.4000	4.962 ± 0.338	3.464 ± 0.041	4.169 ± 0.289	3.299 ± 0.041

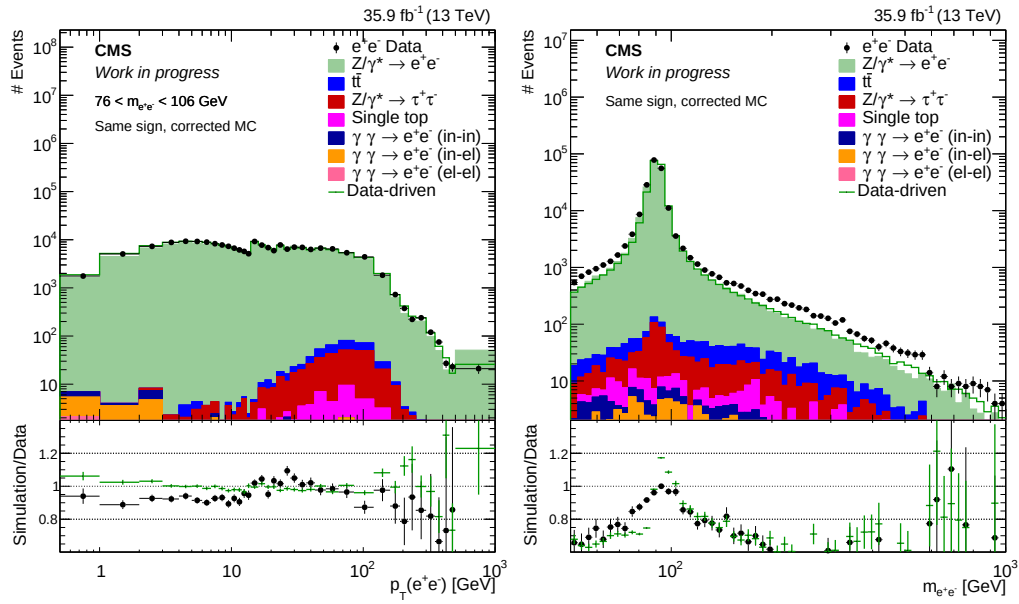


Figure 4.30: Transverse momentum of same sign electron pairs in the Z peak region (left) and the invariant mass of same sign electron pairs (right). Two estimates of the contribution from charge misidentification are shown: the corrected Monte-Carlo as a stack of histograms, and the data-driven estimate as a solid green line.

Results

With the improved knowledge of charge misidentification gained in the previous section, it is now possible to estimate the contribution of fake electrons to the electron channel. This is done by subtracting the charge misidentification contribution from the observed distributions in the same sign region and using the difference as the estimate for fake electrons in the opposite sign region — the difference between the observations and the prediction from charge misidentification visible in the right panel of figure 4.30. The conservative uncertainty used for charge misidentification is sufficient to remove the need from considering additional sources of uncertainty related to the transfer from the same sign to the opposite sign region. The remainder of this section uses control distributions to examine the effects of fake electrons.

Interesting distributions to consider are the subleading electron p_T in mass bins above the Z peak, shown in figure 4.31. The subleading electron is more likely to be fake than the leading one, both because of the steeply falling jet p_T spectrum. At low p_T in the same sign region, the observed event yield is significantly above the charge misidentification estimate. The difference is interpreted as the contribution from fake electrons; it vanishes at high p_T . When used in the opposite sign region (light blue contribution in the bottom plots of figure 4.31), this contribution fills a gap at low p_T and improves the agreement between data and simulation in this region. This gives confidence that estimating the contribution from fake electrons using same sign data is solid and can be relied upon for background estimation.

Control distributions with respect to the transverse momentum of same sign electron pairs are shown in figure 4.32. These plots show that the contribution from fake electrons appears at high $p_{T(ee)}$. This is because fake electrons appear primarily at small p_T and, to produce a pair of high invariant mass, the other electron needs to have $p_T(e) \sim m_{ee}$. The pair is thus very imbalanced, which creates a large $p_{T(ee)}$. Another feature visible in the plots is an apparent disagreement between the data-driven and the Monte-Carlo based estimates for charge misidentification. The difference between the two is similar to what is seen in the opposite sign region and can be attributed to the mismodeling of the dilepton p_T by the Monte-Carlo.

4.6.2 Top quark pair production

The dominant background over much of the phase space, and especially at high mass, originates from the production of a top-antitop ($t\bar{t}$) quark pair. Indeed, the $t\bar{t}$ cross section is largest at $p_T(\ell\ell) \approx 50$ GeV [141], in contrast with the Drell-Yan cross section that peaks around $p_T(\ell\ell) \approx 5$ GeV. In addition, the $m_{\ell\ell}$ distribution in $t\bar{t}$ events falls less quickly than the Drell-Yan one up to $m_{\ell\ell} \approx 300$ GeV. Without

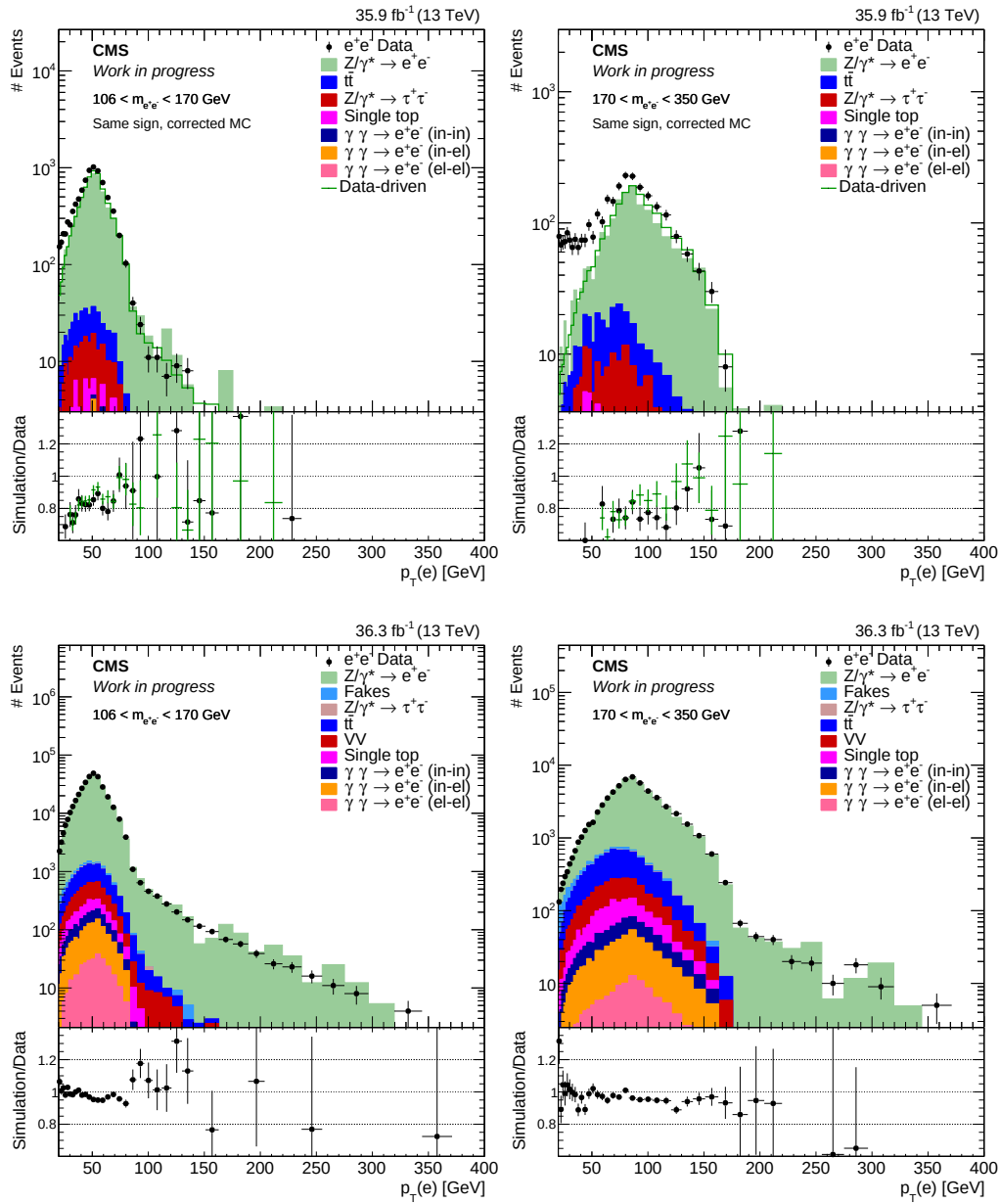


Figure 4.31: Subleading electron p_T in the same sign region (top) and the signal region (bottom), in the mass bins 106 to 170 (left) and 170 to 350 GeV (right). For the same sign region, two estimates of the contribution from charge misidentification are shown: the corrected Monte-Carlo as a stack of histograms, and the data-driven estimate as a solid green line.

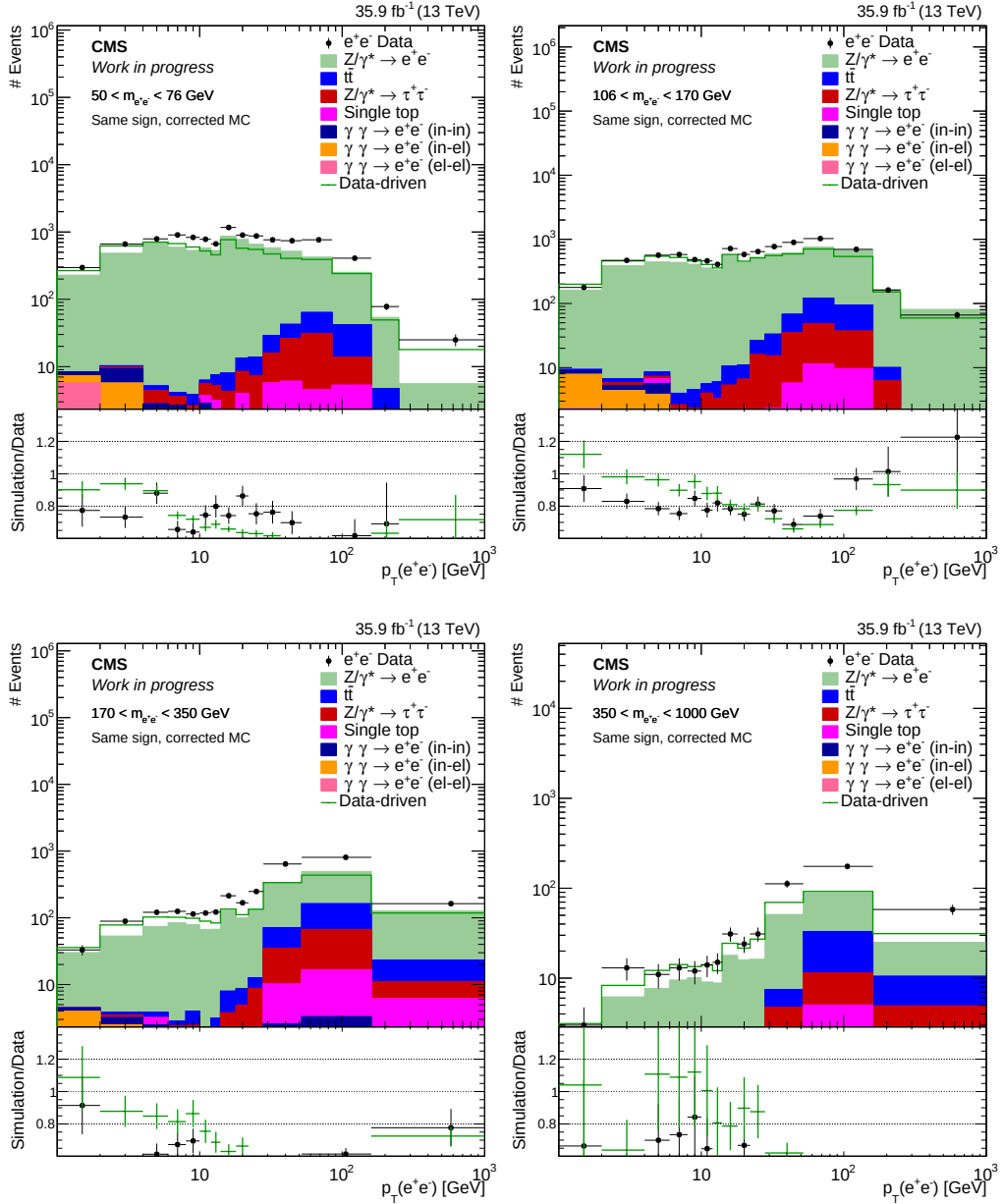


Figure 4.32: Transverse momentum of electron pairs in the same sign region, in the mass bins 50 to 76 (top left), 106 to 170 (top right), 170 to 350 (bottom left), and 350 to 1000 GeV (bottom right). Two estimates of the contribution from charge misidentification are shown: the corrected Monte-Carlo as a stack of histograms, and the data-driven estimate as a solid green line.

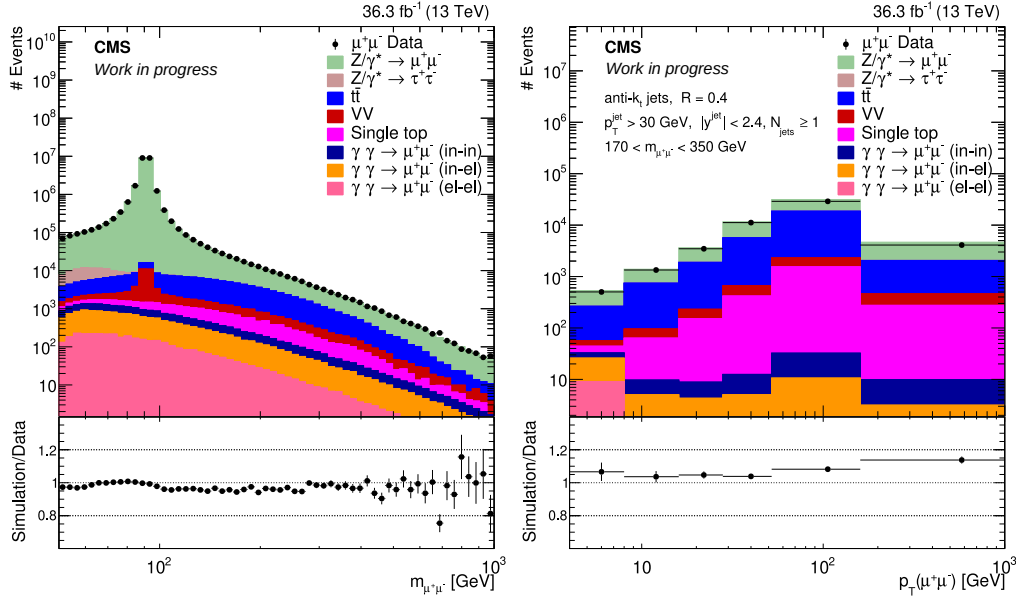


Figure 4.33: Data to simulation comparisons of the invariant mass of muon pairs (left) and their p_T in the mass range from 170 to 350 GeV (right), without rejecting events containing b-tagged jets.

the veto on b jets, $t\bar{t}$ events account for more than half of the observations at high mass and high $p_T(\ell\ell)$. In this section, the handling of this background is discussed: the first part covers the veto on b jets used to reduce the contamination and the second describes the validation of the Monte-Carlo sample in data.

Veto on b jets

Top decays almost always involve the production of a bottom quark [11]. For top quark pairs, events without a b quark represent only 1 % of the total. This makes b tagging extremely relevant to distinguish between Drell-Yan and $t\bar{t}$ production. It was already mentioned in section 4.4.7 that events with a b-tagged jet are rejected as part of the selection. Without this cut, the $t\bar{t}$ contribution would be larger than the signal at high mass and high p_T . This can be seen in figure 4.33, which shows distributions without applying the b veto.

The jet tagger used in this analysis is the Combined Secondary Vertex version 2 (CSVv2) algorithm [121], for which three optimized working points are provided: **loose**, **medium** and **tight**. The **loose** working point is the most efficient (81 % in $t\bar{t}$ simulated events but also comes with the largest mistagging rates (37 % for c jets and 8.9 % for light jets). In contrast, the **tight** working point has a b jet tagging efficiency of only 41 % and the mistagging rate is below 3 %. The choice

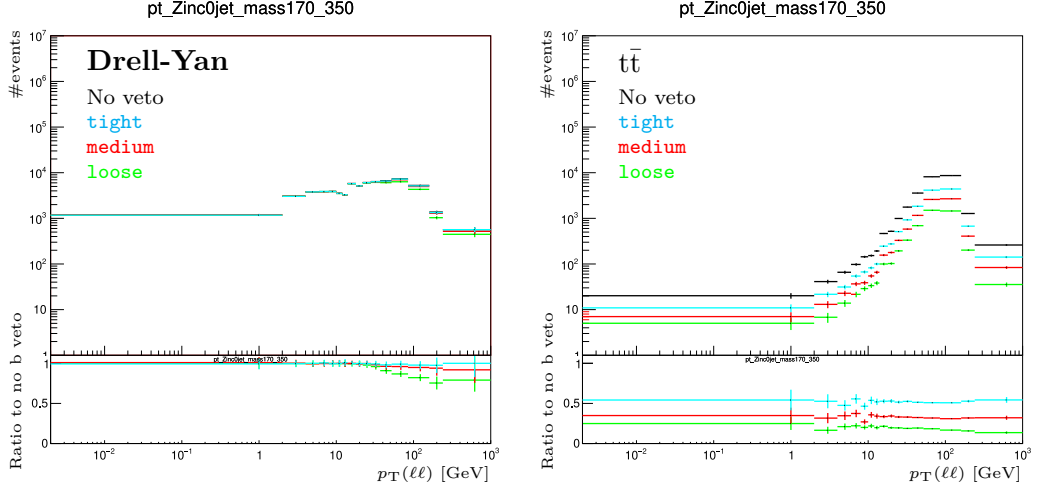


Figure 4.34: Effect of the b veto on the number of events for the three b tagging working points, in Drell-Yan (left) and $t\bar{t}$ (right) simulation for the mass bin between 170 and 350 GeV in the muon channel.

of the working point is based on a balance between reduction of the background and rejection of signal events. Figure 4.34 shows the reduction in the number of events caused by vetoing events with at least one b jet for the three working points, separately for the Drell-Yan and the $t\bar{t}$ Monte-Carlo. A reduction of the $t\bar{t}$ event yield is visible as expected. For signal events, the **loose** working point visibly reduces the selection efficiency at high $p_T(\ell\ell)$, while the effect is much smaller for the **medium** and **tight** working points. The **medium** working is chosen, bringing the background contamination to an acceptable level while having a small effect on Drell-Yan.

Validation in data

In this section, the validation of the Monte-Carlo sample used to estimate the $t\bar{t}$ process is discussed. The validation uses a slightly different phase space than the measurement, based on $e\mu$ events. This selection is optimized to select $t\bar{t}$ events. The single muon triggers **HLT_IsoMu24** and **HLT_IsoTkMu24** are used. One muon with $p_T > 25$ GeV is required, in association with an electron with $p_T > 20$ GeV; the other cuts on leptons are the same as for electrons and muons in the signal region, including the requirement for opposite charges. The choice of this electron-muon final state almost completely eliminates the contribution from the Drell-Yan process, which always yields two leptons of the same flavor; on the contrary, the two leptons produced from leptonic top decays in the $t\bar{t}$ process are independent and can hence be of different flavors. In addition to the changes to the selection of

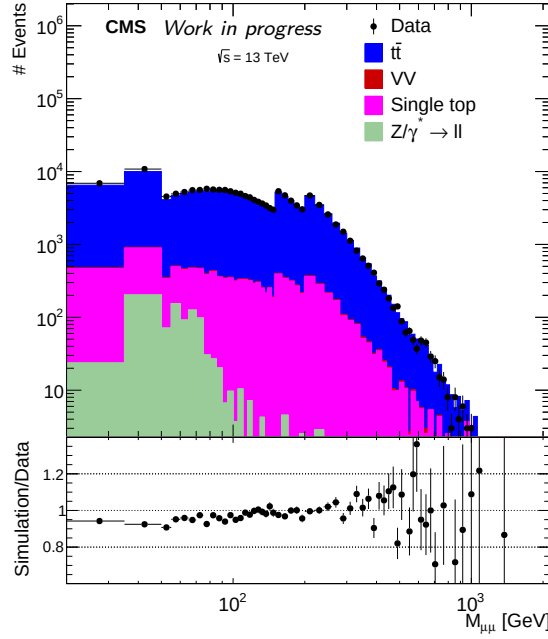


Figure 4.35: Invariant mass of $e\mu$ lepton pairs within the top-enriched phase space. The jumps in the number of events are caused by the irregular binning.

the leptons, the veto on b jets is inverted and is turned into a requirement for at least one b -tagged jet with $p_T > 20$ GeV. The trigger, lepton and b tagging scale factors are adjusted to match this alternative selection. The selected events are highly enriched in the $t\bar{t}$ process, as can be seen in figure 4.35.

The key variable that needs to be checked to validate the description of the $t\bar{t}$ background for the Drell-Yan measurement is the transverse momentum of the lepton pair. Validating $p_T(\ell\ell)$ will also affect φ_η^* because of the correlation between the two. Since the definition of the control region already requires one jet, the cross check can be used for both the inclusive measurement of $p_T(\ell\ell)$ and the measurement when at least one jet is required. The distributions of $p_T(\ell\ell)$ in the control region for four $m_{\ell\ell}$ bins are shown in figure 4.36. Within the limits of the statistical uncertainties, good agreement is observed between data and the POWHEG prediction. This result is consistent with differential $t\bar{t}$ measurements performed by CMS at 13 TeV [141] (esp. figures 41 and 42), which validates the use of POWHEG without further correction to estimate the $t\bar{t}$ background.

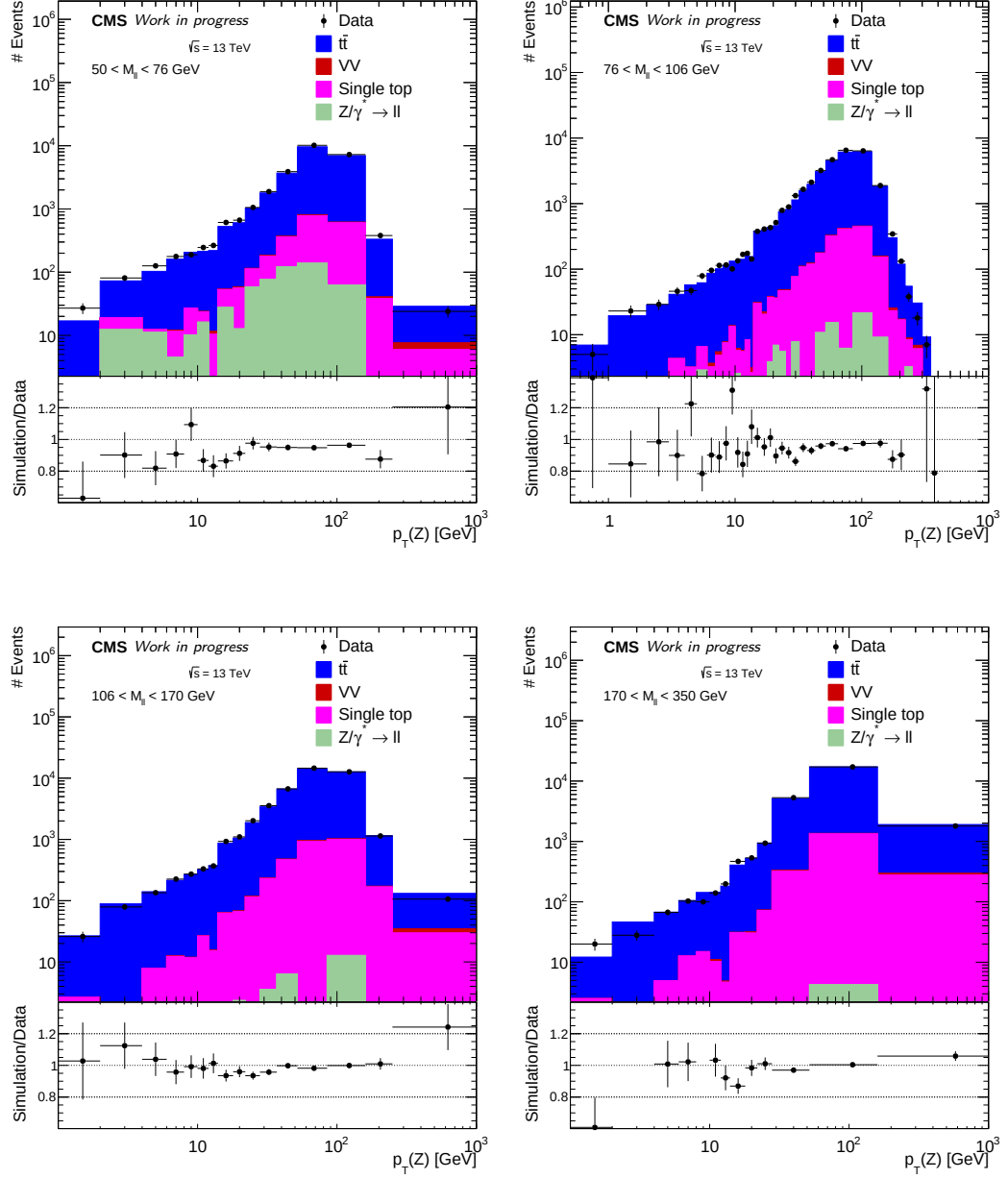


Figure 4.36: Transverse momentum of $e\mu$ lepton pairs in the top-enriched phase space, in the mass bins 50 to 76 (top left), 76 to 106 (top right), 106 to 170 (bottom left), and 170 to 350 GeV (bottom right).

4.7 Unfolding

The distributions measured in data contain contributions from backgrounds, which are subtracted, and signal. Interpretation of the signal at this stage requires a precise knowledge of the detector efficiency and resolution as well as the acceptance of the selection. This dependence on experimental effects is not desirable because it makes comparisons to theory predictions and between experiments extremely difficult. The observed distributions are therefore corrected for all detector effects through a procedure known as “unfolding” [142].

The definition of the phase space in which the unfolded results are provided is given in terms of the particles produced by an event generator (the details for the present analysis will be given in section 4.7.1). Using the simulated response of the detector and reconstruction algorithms to generated events, one can deduce for each measured distribution the conditional probability $P(\text{reco } i | \text{gen } j)$ that an event generated in bin j is reconstructed in bin i . The observed distribution $\mathbf{r}$ is smeared with respect to the “true” distribution $\mathbf{t}$ in an affine way: $\mathbf{r}_i = \sum_j P(\text{reco } i | \text{gen } j) \mathbf{t}_j + \mathbf{f}_i$, or in matrix form $\mathbf{r} = \mathbf{S}\mathbf{t} + \mathbf{f}$ with $S_{ij} = P(\text{reco } i | \text{gen } j)$. The smearing matrix $\mathbf{S}$ includes the per-bin detection efficiency $\varepsilon_j = \sum_i S_{ij}$, smaller than one since not all generated events pass the selection at reconstructed level (“misses”). Its off-diagonal elements represent migrations between bins caused by imperfect detector resolution. The “fakes” vector $\mathbf{f}$ represents events that do not enter the definition of the phase space at generator level but are detected within the acceptance. Fakes are subtracted from the measured data in the same way as background processes.

Because they are extracted from a finite-size Monte-Carlo sample, the elements of $\mathbf{S}$ are not known with absolute statistical precision. This causes its exact inverse to be subject to large fluctuations that need to be regularized. Over the years, several methods have been developed to effectively calculate $\tilde{\mathbf{S}}^{-1}\mathbf{r}$. The simplest is to neglect resolution effects and correct only for the efficiency, $\tilde{\mathbf{S}}^{-1}_{ij} = \delta_{ij}/\varepsilon_j$. This “bin-by-bin” unfolding method is only valid when bins are wide with respect to the resolution. It is also possible to use the exact inverse of $\mathbf{S}$ if the statistical fluctuations are small. When this is not the case, as in the present analysis, more advanced methods are needed, the most notable of which are D’Agostini iteration with early stopping [143, 144, 145] and methods based on Tikhonov regularization [146, 147, 148]. This measurement uses the D’Agostini method as implemented in the `RooUnfold` package [149].

The D’Agostini method inverts the conditional probability $P(\text{reco } i | \text{gen } j)$ by applying the Bayes theorem to find $P(\text{gen } j | \text{reco } i)$. Once the inverse is known,

the unfolded distribution $\mathbf{t}^{\text{data}}$ is calculated as:

$$\mathbf{t}_j^{\text{data}} = \frac{1}{\varepsilon_j} \sum_i \mathbf{r}_i^{\text{data}} \text{P}(\text{gen } j | \text{reco } i). \quad (4.13)$$

The inversion using the Bayes theorem reads:

$$\text{P}(\text{gen } j | \text{reco } i) = \frac{\text{P}(\text{reco } i | \text{gen } j) \text{P}(\text{gen } j)}{\sum_k \text{P}(\text{reco } i | \text{gen } k) \text{P}(\text{gen } k)}. \quad (4.14)$$

The prior used to apply the theorem, i.e. the values of $\text{P}(\text{gen } j)$ in the right-hand side of the above equation, is taken to be the prediction of the simulation. This introduces a bias because the prediction is different from the true distribution. In order to remove the bias, one proceeds iteratively by using the newly calculated $\mathbf{t}^{\text{data}}$ to construct a more realistic prior. Successive applications of the procedure successfully eliminate bias and it has been shown that in the limit of an infinite number of iterations, the results converge to the maximum likelihood estimate [143].

Regularization in the D’Agostini method is achieved by stopping the procedure before convergence in order to limit the appearance of the associated statistical fluctuations. Choosing the number of iterations carefully is important to achieve the desired balance between bias and uncertainty. The method used in this analysis is discussed in section 4.7.3.

Unfolding corrects for the detector efficiency and the acceptance of the selection. The unfolded distributions correspond to the true number of events that happened in the considered phase space. In each bin of the distribution of each variable x , the true number of events N_{unfolded} provided by the corresponding element of $\mathbf{t}^{\text{data}}$ is turned into a differential cross section by dividing it by the integrated luminosity $\mathcal{L}$ and the bin width Δx :

$$\frac{d\sigma}{dx} = \frac{1}{\mathcal{L}} \frac{N_{\text{unfolded}}}{\Delta x}. \quad (4.15)$$

4.7.1 Generator level phase space

The generator level selection in which the measurement is performed are chosen to match the most important cut of the selection in data, keeping the extrapolation of the observation and the associated model dependence to a minimum. The phase space used for the measurement is defined as follows:

1. Photons lying within a $\Delta R(\ell, \gamma) < 0.1$ cone around each prompt lepton in the event record are merged with the closest lepton, adding their four-momenta to that of the original (“bare”) lepton. The resulting objects are

called “dressed” leptons. Prompt particles are the ones not originating from the decay of another particle: this requirement prevents the inclusion of $Z \rightarrow \tau\tau$ events and photons from hadron decays. The dressing is performed separately for electrons and muons. In case there is an ambiguity between two bare leptons of the same flavour, the one closest to the photon is chosen (this can happen when the leptons are close to each other in ΔR distance).

2. In each channel, the two dressed leptons with largest p_T are considered. They are required to have a pseudorapidity $|\eta(\ell)| < 2.4$ and opposite charges; the one with largest p_T is required to have $p_T(\ell) > 25 \text{ GeV}$, and the second one $p_T(\ell) > 20 \text{ GeV}$. In addition, the invariant mass of the pair must be within the measured range, $50 < m_{\ell\ell} < 1000 \text{ GeV}$. If a pair satisfying the above requirements is found, it is used to reconstruct the kinematics of the produced Z/γ^* boson. Cases where a boson is found in each channel are extremely rare in Drell-Yan samples, and in such cases the electron channel is chosen.
3. For the requirement of at least one jet also used when measuring $p_T(\ell\ell)$, stable visible particles (all particles with $c\tau > 1 \text{ cm}$, except neutrinos) are clustered with the anti- k_T algorithm, using a cone size parameter of 0.4. Jets are required to have a rapidity $|y(\text{jet})| < 2.4$, a transverse momentum $p_T(\text{jet}) > 30 \text{ GeV}$ and to be at least 0.4 units away from each of the two selected leptons in ΔR distance.

The phase space definition described above is defined using stable particles after parton showering, hadronization and decay of short lifetime particles, except for the selection of the photons and the requirement that leptons do not originate in τ decays. This makes the results as robust as possible against the development of improved theoretical models, for instance the inclusion of higher electroweak orders in matrix element generators. Still with the goal of facilitating comparisons against future predictions, the selection has been implemented in a RIVET [150] routine.

4.7.2 Response matrices

Before unfolding the distributions, it is important to check that the smearing matrix S is well-behaved. This is usually done by factoring out efficiencies to build the “response matrix” R , $R_{ij} = \frac{1}{\varepsilon_j} S_{ij}$. In response matrices, a good reconstruction is indicated by diagonal elements (“stability”) close to one. The response matrices for the measured distributions are shown in figures 4.37 and 4.38 ($p_T(\ell\ell)$), figures 4.39 and 4.40 (φ_η^*), and figures 4.41 and 4.42 ($p_T(\ell\ell)$ with a jet requirement),

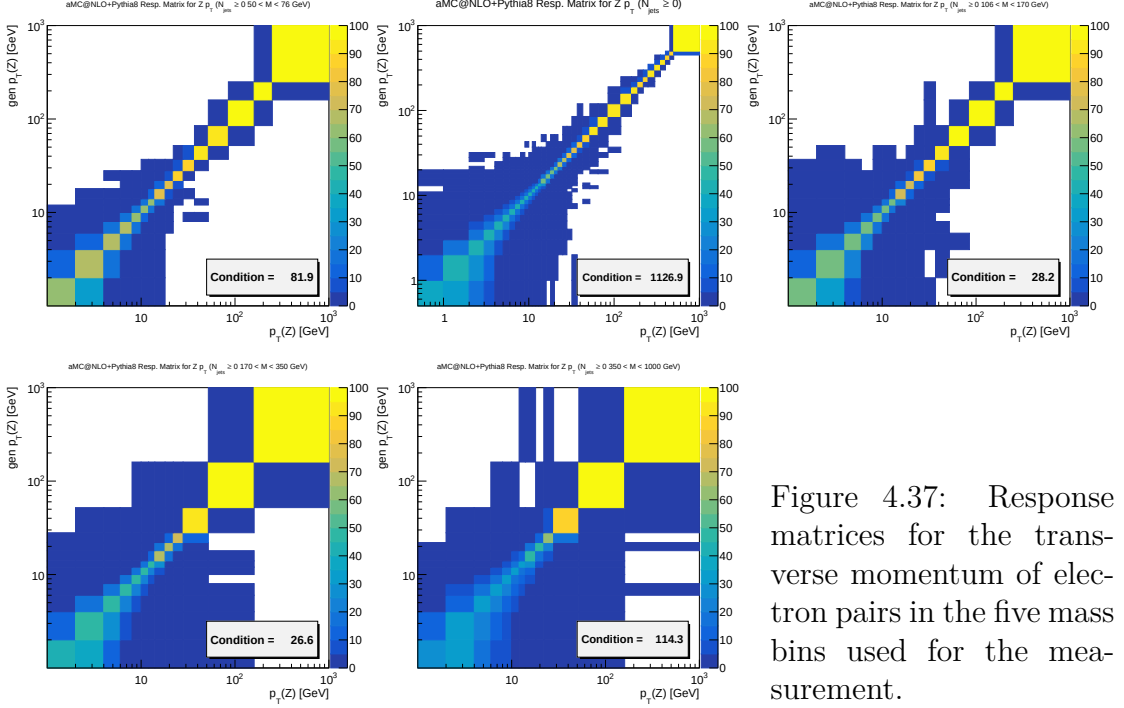


Figure 4.37: Response matrices for the transverse momentum of electron pairs in the five mass bins used for the measurement.

for the electron and muon channels. The “true” value is represented on the vertical axis and the reconstructed one on the horizontal axis. The response matrices are filled with events that pass the cuts both at the generated level and at the reconstructed level. The values shown in the plots are percentages.

The response matrices for $p_T(\ell\ell)$ in the electron channel show that the binning is at the limit of the detector resolution in the low p_T part of the spectrum. This is especially true for the Z peak region, where the bins at low p_T are only 1 GeV wide, and at high mass. At high p_T , the binning is dictated by the number of events rather than the resolution, and the response matrices are almost diagonal. The degradation of the $p_T(\ell\ell)$ resolution at high mass in the electron channel despite the improving electron energy resolution at high p_T (figure 4.3) deserves an explanation. Consider a dielectron pair at small $p_T(ee)$ with respect to the invariant mass m_{ee} . In this case, the most probable situation is that the two electrons are back-to-back in the transverse plane and each electron carries half of the mass, $p_T(e) \approx m_{ee}/2$. In terms of these quantities, the transverse momentum of the pair is given by:

$$p_T(ee) \approx |p_T(e_1) - p_T(e_2)| (\approx 0). \quad (4.16)$$

Writing the relative energy resolution on individual electrons as $\rho \equiv \frac{\Delta p_T(e)}{p_T(e)}$ and

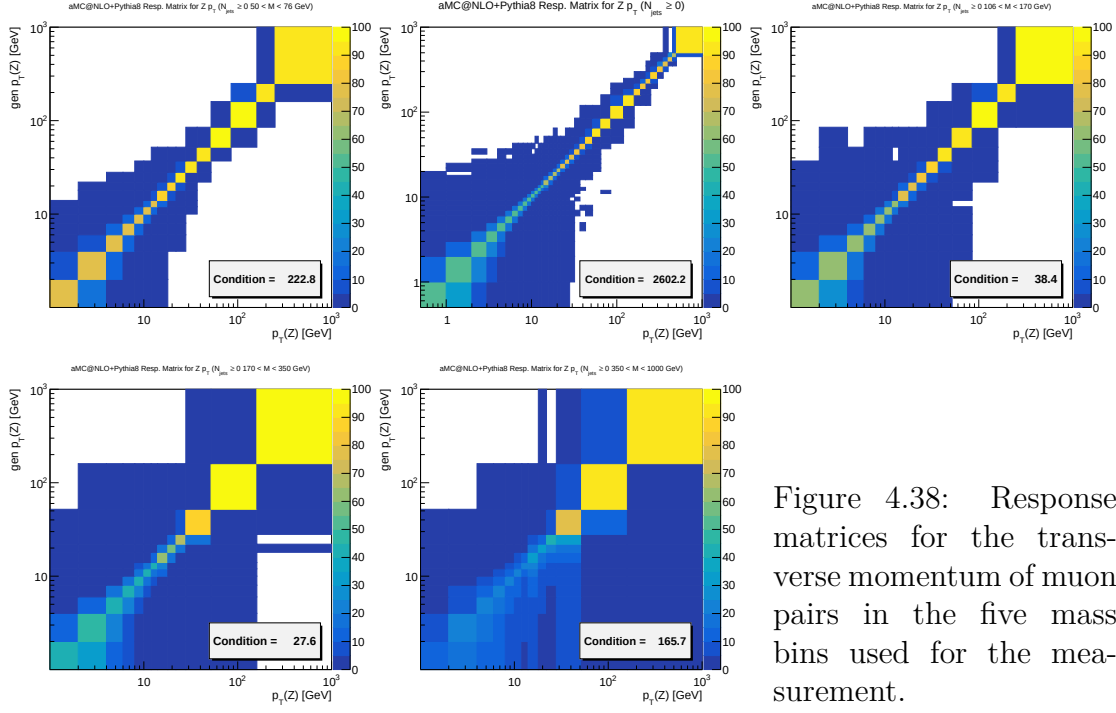


Figure 4.38: Response matrices for the transverse momentum of muon pairs in the five mass bins used for the measurement.

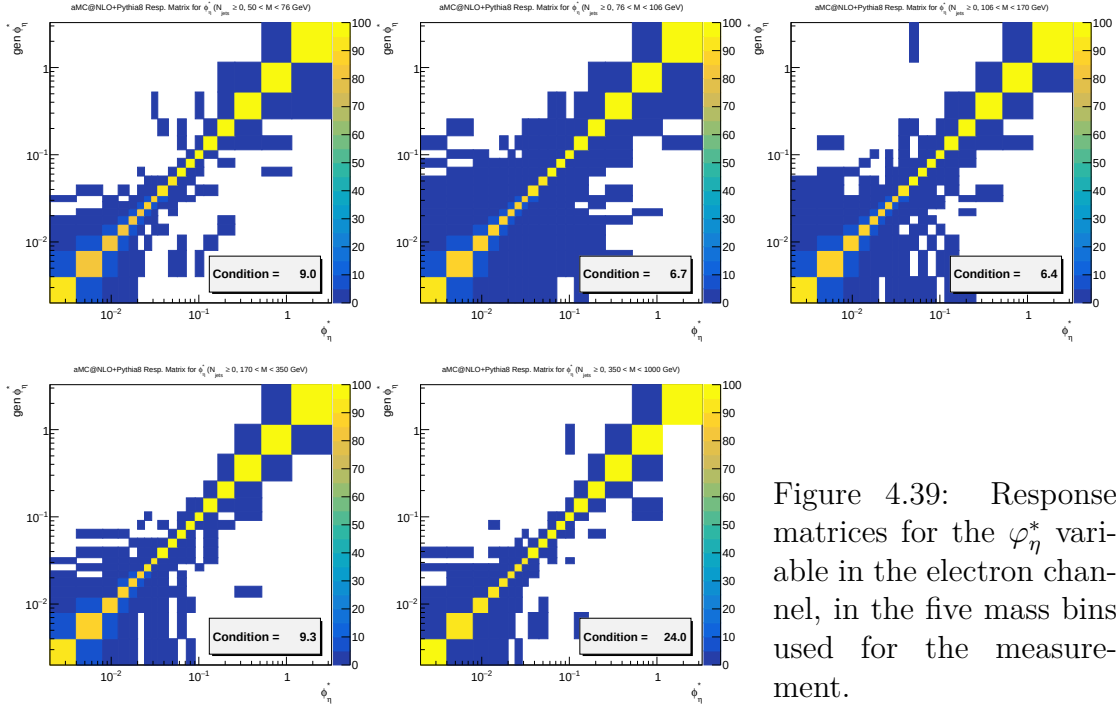


Figure 4.39: Response matrices for the φ_{η}^* variable in the electron channel, in the five mass bins used for the measurement.

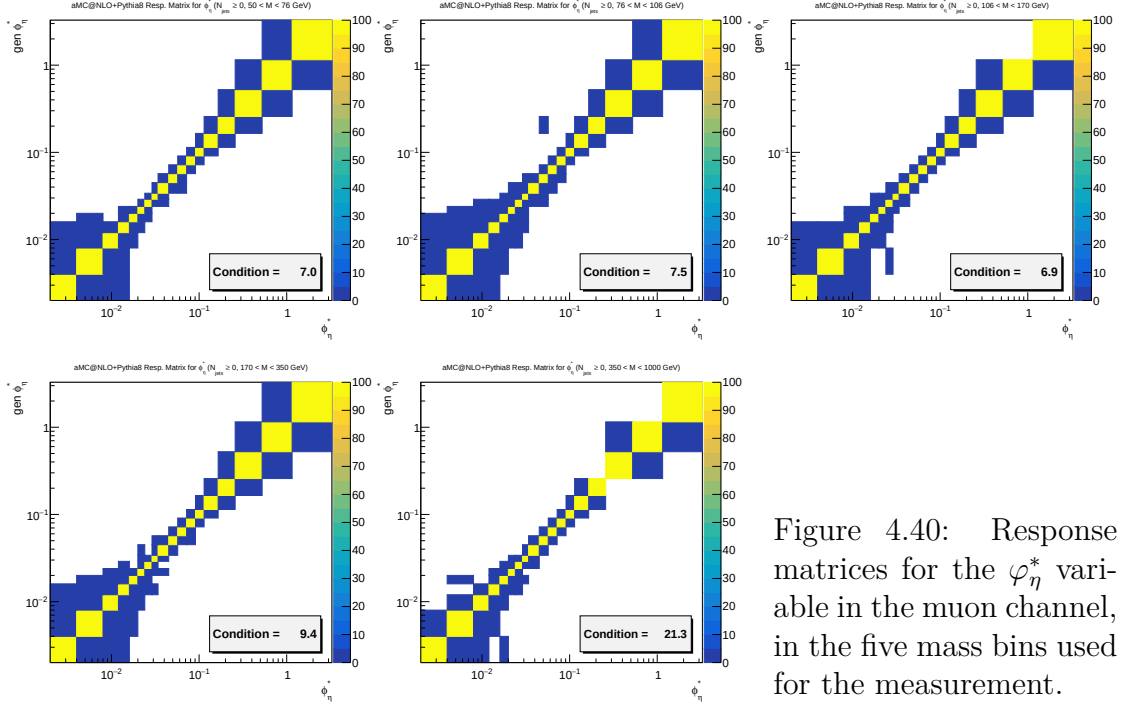


Figure 4.40: Response matrices for the φ_{η}^* variable in the muon channel, in the five mass bins used for the measurement.

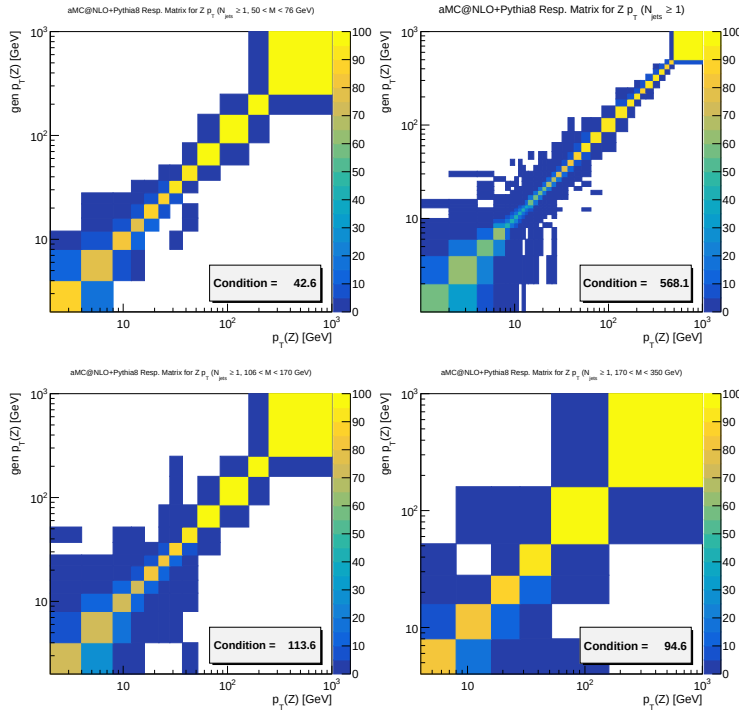


Figure 4.41: Response matrices for the transverse momentum of electron pairs in the four mass bins used for the measurement, with at least one jet required in the events.

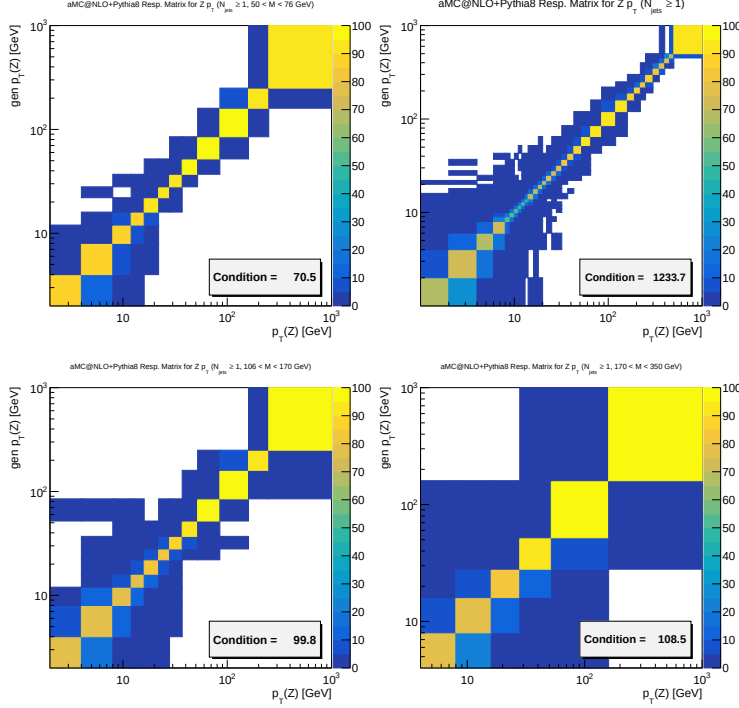


Figure 4.42: Response matrices for the transverse momentum of muon pairs in the four mass bins used for the measurement, with at least one jet required in the events.

disregarding correlations, the uncertainty in this calculation is:

$$\Delta p_T(ee) \approx \sqrt{(\Delta p_T(e_1))^2 + (\Delta p_T(e_2))^2} \approx \frac{\rho m_{ee}}{\sqrt{2}}. \quad (4.17)$$

For a mass of 350 GeV and a typical electron energy resolution of $\rho = 1\%$, the relation above gives an absolute $p_T(ee)$ uncertainty of $\Delta p_T(ee) \approx 2.5$ GeV (for $p_T(ee) \ll m_{ee}$). The energy resolution is thus fully expected to be a limiting factor for the bin width of 2 GeV used in the measurement.

At small p_T , the energy resolution for muons is better than for electrons, which translates into better behaved (more diagonal) response matrices at low mass and up to $m_{\mu\mu} = 170$ GeV. The resolutions in the electron and muon channels are still equivalent for the mass range between 170 and 350 GeV. In the last mass bin, the degrading muon p_T resolution combines with the larger invariant mass to create a response with very large migrations. At low p_T , the diagonal elements are of the order of only 20 %. Trying to use such a response for unfolding would result in a large model dependence. In order to avoid this, the measurement in the highest mass bin is restricted to the electron channel.

In the φ_η^* response matrices, the diagonal elements are always larger than 80 % in the electron channel and larger than 95 % in the muon channel. This experimental resolution is what the definition of φ_η^* was optimized for: being based on angular variables only, φ_η^* doesn't depend on the energy scale. The precision

of angular variables is driven by track angular resolution, which is much better constrained than lepton energy scales. The small resolution difference between electrons and muons is caused by the difficulties in reconstructing electron tracks due to their tendency to emit bremsstrahlung photons. The resolution for $p_T(\ell\ell)$ with at least one jet is the same as in the inclusive, but wider bins have been chosen because of the smaller number of events. As a consequence, the response matrices for this variable are almost exclusively diagonal.

4.7.3 Number of iterations

The number of iterations used for unfolding is chosen by performing a closure test on the unfolded data: the unfolded distribution is folded back by a direct application of the response matrix and compared to the original. If the pseudo-inverse $\tilde{S}^{-1}$ effectively inverses S , one should have up to small differences due to regularization:

$$S\tilde{S}^{-1}\mathbf{r} \approx \mathbf{r}. \quad (4.18)$$

The number of iterations is chosen by comparing the two sides of this equation and using the smallest number of iterations for which the differences are smaller than the statistical uncertainty, with a minimum of four iterations to minimize model dependence. The number of iterations determined using this procedure are listed in table 4.7, where one can see that convergence is fast for distributions with nearly diagonal response matrices and slower when the problem is more difficult, in particular for $p_T(\ell\ell)$ in the Z peak region.

Bottom-line test

The bottom-line test [142] is a cross check of the unfolding procedure. The basic idea is that unfolding must not bring additional information with respect to the detector-level distributions. In order to quantify the amount of information given by the distributions, the bottom-line test uses two χ^2 compatibility tests between the data and a prediction (MC), using covariance matrices at the reconstructed level C_r and generated level C_t (see also appendix A):

$$\begin{aligned} \chi_r^2 &\equiv (\mathbf{r}_{MC} - \mathbf{r}_{data})^T C_r^{-1} (\mathbf{r}_{MC} - \mathbf{r}_{data}) \\ \chi_t^2 &\equiv (\mathbf{t}_{MC} - \mathbf{t}_{data})^T C_t^{-1} (\mathbf{t}_{MC} - \mathbf{t}_{data}) \end{aligned}$$

The first test is performed at the detector level between the predicted MC yields and the observed number of events (minus background). The second test is performed at the generator level between the MC truth and the unfolded data. Any covariance matrix can be used as long as it is propagated consistently with the

data. If unfolding introduces no bias, the two tests should give identical results and the ratio χ_r^2/χ_t^2 should be equal to one up to corrections due to regularization.

The bottom-line test has been performed for every distribution considered in the measurement. The statistical uncertainty in the data has been used for the covariance matrix and the tests were performed against the predictions of the MADGRAPH sample. The ratio χ_r^2/χ_t^2 is in all cases close to unity, with deviations ranging from less than 10^{-6} (always in the muon channel) to 3×10^{-3} . Before reaching this level of agreement, the bottom-line test has given hints to issues in the chosen number of iterations and in the calculation of the covariance matrix.

4.7.4 Modeling uncertainties

Because of the bias inherent to unfolding, the results depend on the prediction from the MADGRAPH sample used to build the prior and the response matrices. This bias is expected to be larger in regions where large migrations occur such as at low $p_T(\ell\ell)$. They can also be significant in regions where MADGRAPH predicts the slope of the measured distributions incorrectly. This can be explained by considering two bins b_1 and b_2 with 10 times more events in b_1 than in b_2 . If 10 % of the events are reconstructed in the wrong bin, the total contribution in b_1 will be 1 % while it will be 50 % in b_2 . In this hypothetical example, a small change in the slope is predicted by the model would have a large effect on the value of b_2 .

Table 4.7: Number of iterations used for unfolding.

Variable	$m_{\ell\ell}$ [GeV]	Electrons	Muons
$p_T(\ell\ell)$	50 – 76	4	4
	76 – 106	25	19
	106 – 170	4	4
	170 – 350	4	9
	350 – 1000	4	n/a
φ_η^*	50 – 76	4	4
	76 – 106	5	4
	106 – 170	5	4
	170 – 350	5	4
	350 – 1000	5	4
$p_T(\ell\ell)$ (one jet)	50 – 76	4	4
	76 – 106	15	8
	106 – 170	4	4
	170 – 350	4	4

after unfolding.

A commonly used method to estimate the bias due to unfolding is to repeat detector simulation and unfolding using an alternative prediction to build the prior and the response matrices. This method works to the extent that the two predictions use different models to describe the underlying physics. Producing the alternative Monte-Carlo is, however, a computationally intensive task and the produced sample uses valuable disk space. Given the large number of simulated events required for unfolding in this analysis, the production of samples using an alternative generator was considered impractical. Instead, a pseudo-sample is constructed by reweighting the MADGRAPH samples to match the data at the reconstructed level. In essence, the weights are calculated by taking the ratio of the data and the original prediction as a function of each considered variable. Alternate response matrices and priors are then built by reweighting the simulated events independently for each variable. The unfolding is repeated with this new model and the difference with respect to the central values is used as the modeling uncertainty.

Calculation of the weights

The weights are calculated independently for each distribution. They are based on histograms filled with a finer binning than for the measurement, splitting each bin in two to gain access to slope information within the bins. The points from the histograms in data and simulation are then interpolated to smooth out statistical fluctuations, by constructing second order polynomials to approximate the neighbourhood of each point. The steeply falling behavior of the $p_T(\ell\ell)$ and φ_η^* spectra is regularized by performing the interpolation in log-log space, where the smoother behavior of the distributions is suited for a description by polynomials. An example of the result is shown in the top panel of figure 4.43.

To obtain the weights used to construct the alternative model for the unfolding uncertainty, the curve obtained by interpolating the data points is divided by its equivalent in simulation. An additional normalization factor is used to ensure that the total cross section does not change when applying the weights. The normalization factor is calculated by requiring that the following equality holds:

$$\sum_{\text{bins}} \text{weight} \times \text{number of events} = \sum_{\text{bins}} \text{number of events}, \quad (4.19)$$

where the sums run over bins in the reconstructed level distribution and the number of events is the prediction from simulation. For this calculation, the weights are evaluated at the center of the bins.

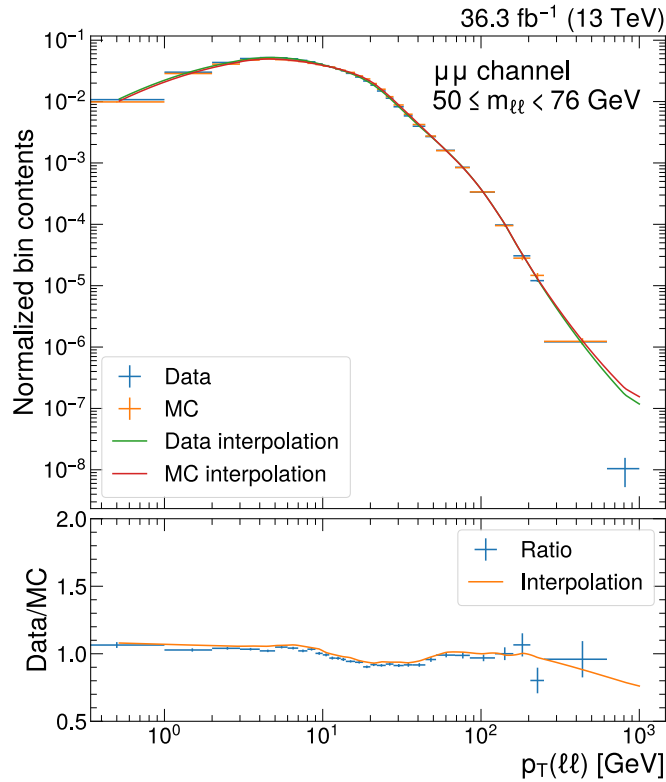


Figure 4.43: Example calculation of the weights used for the unfolding uncertainty. The top panel shows distributions of reconstructed-level $p_T(\ell\ell)$ in data (after background subtraction) and simulation, and their interpolation. The bottom panel shows the ratio of the two histograms and the calculated weights (“interpolation”).

Reweighting

The alternative model is constructed for each measured distribution by reweighting simulated events. When applying the weights, one has to be careful to preserve the efficiencies: changing it would introduce artificial changes not related to the model used for unfolding. The efficiencies are modeled in the smearing matrix as $\varepsilon_j = \sum_i S_{ij}$; in terms of Monte-Carlo events, it is equal to the (effective) fraction of generated events that were reconstructed in any bin. In order to preserve the efficiencies, it is therefore essential to reweight all generated events, not only the ones that pass the selection. Put another way, it is the whole physics model that is reweighted and not just the fraction of it visible in the detector.

As discussed above, the calculation of the weights is based on reconstructed level information. Events are therefore reweighted based on their reconstructed level properties when they are available. Events that were missed by the detector are instead reweighted based on their generator level information. This is in general not valid, as the response matrices show that the reconstructed p_T can be significantly different from the generated one. However, in the case of this analysis, the weights are nearly constant in regions of the response matrices where significant migrations are observed. This makes it possible to neglect the difference between the weights evaluated at the reconstructed and at the generated p_T , and therefore to use weights derived from the reconstructed level to reweight events not reconstructed by the detector. The problem does not appear for φ_η^* because the response matrices are diagonal.

4.8 Uncertainties

The measurement is affected by a number of statistical and systematic uncertainties, the most important of which are the uncertainties in the luminosity and lepton efficiency measurements. The uncertainties are modeled as a number of independent sources whose effects are propagated to the final results and added in quadrature (see appendix A for the theoretical background). Most uncertainties are estimated by varying the corresponding parameter up and down by one standard deviation and repeating the analysis. This results in two new estimates of the cross sections. The differences $\Delta_\alpha \sigma_i$ with respect to the central values σ_i (α being the index of the considered uncertainty) are symmetrized by taking the average and taken as the standard deviation of the measurement. The sign of the deviation that corresponds to varying the parameter upwards is also recorded in each bin in order to handle correlations properly when constructing the covariance matrix of the results. The paragraphs below list the sources considered in the analysis; a summary is also provided in table 4.8.

Luminosity The integrated luminosity $\mathcal{L}$ recorded by CMS in 2016 was of 36.3fb^{-1} with a relative uncertainty of 1.2% [69]. This number is used to convert the number of events to a cross section and to normalize the background Monte-Carlo samples, resulting in a flat 1.2% uncertainty where backgrounds are negligible and a slightly larger uncertainty elsewhere. This rise can be explained as follows.

Neglecting the acceptance and efficiency corrections and data-driven backgrounds, the measurement of the cross section σ is performed as follows:

$$\sigma = \frac{1}{\mathcal{L}}(N_{\text{data}} - \mathcal{L}\sigma_{\text{background}}) = \frac{N_{\text{data}}}{\mathcal{L}} - \sigma_{\text{background}}, \quad (4.20)$$

where N_{data} is the number of observed events and $\sigma_{\text{background}}$ is the total background cross section. The absolute luminosity uncertainty $\Delta_{\mathcal{L}}\sigma$ does not depend on the amount of background:

$$\Delta_{\mathcal{L}}\sigma = \frac{N_{\text{data}}\Delta\mathcal{L}}{\mathcal{L}^2}. \quad (4.21)$$

When calculating the relative uncertainty however, the background fraction appears in the denominator:

$$\frac{\Delta_{\mathcal{L}}\sigma}{\sigma} = \frac{\Delta\mathcal{L}/\mathcal{L}}{1 - \frac{\mathcal{L}\sigma_{\text{background}}}{N_{\text{data}}}}. \quad (4.22)$$

The denominator makes the relative luminosity uncertainty raise when background contamination is larger.

Lepton identification efficiencies These values are used to correct the efficiency in simulation. Their uncertainties are split in a large number of systematic and statistical components that are all propagated separately. The five systematic sources are the choice of the tag lepton in the tag-and-probe method and the modeling of the signal, background, final state radiation, and mass resolution. One statistical source per bin of the scale factors is also considered: there are 143 in the muon channel and 480 in the electron channel. In most of the phase space, this uncertainty is around 1% and is dominated by the systematic components.

Lepton reconstruction and trigger efficiencies The measurement of these efficiencies is used to correct the simulation. The uncertainties in their measurements are propagated by varying the corresponding scale factors up and down by one standard deviation. This uncertainty is around 0.5%.

Lepton energy scale The uncertainty in the energy scale of the leptons is important for several aspects of the analysis. A change in the energy scale can affect events with leptons near the p_T cuts, events at the borders of the mass bins, and the reconstructed value of $p_T(\ell\ell)$. The uncertainty in the lepton energy scale is estimated separately for electrons and muons by repeating the analysis with the leptons scaled up and down by conservative estimates of the uncertainty. The electron energy scale in simulation is varied by 0.18 % for electrons in the barrel and 0.25 % for electrons in the endcaps. The uncertainty is larger for electrons measured using a smaller gain in the ECAL readout electronics, which happens approximately for $p_T > 250$ GeV. For electrons with $p_T > 250$ GeV, the uncertainty is raised to 0.75 %. For muons, the energy scale is varied up and down by the total uncertainty in the Rochester correction, which includes statistical and systematic components.

Lepton energy resolution The uncertainty in the correction of the lepton energy resolution performed in Monte-Carlo is estimated in the signal samples by smearing the leptons with respect to their generated level counterpart. Muons are smeared by 0.6 % and electrons by 6 %.

Jet energy scale The energy scale of the jets enters the selection for the measurement of $p_T(\ell\ell)$ with at least one jet and indirectly affects the inclusive measurements through the application of the veto on b-tagged jets. The uncertainty in this calibration is estimated by scaling the jet momenta in data using centrally provided run-dependent uncertainties. This uncertainty is negligible for the inclusive distributions, but raises to 10 % at low p_T in the selection with at least one jet.

Jet energy resolution The resolution of the jet energy assignment is different in data and simulation, which is corrected for by smearing simulated jets. The factors used for this smearing are varied by their uncertainty.

Pileup The uncertainty in the pileup reweighting procedure used in simulation is estimated by constructing two alternative pileup profiles constructed by varying the minimum bias cross section up and down by its uncertainty of 5 % [151] and repeating the analysis for each of them.

Backgrounds The uncertainty in the background estimation is estimated by scaling the estimated total background cross section up and down by 6 %. This corresponds to the largest uncertainty in the cross sections used for normalization,

which is 6 % for the $t\bar{t}$ process [152, 127] and takes into account the scale and PDF uncertainties.

B jet tagging The up and down uncertainties in the b tagging scale factors are applied and propagated to the measurement. This uncertainty is smaller than 0.1 % for the total cross section.

Pre-firing The uncertainty in the electron pre-firing scale factors is estimated using the number measured centrally and propagated to the analysis as up and down variations. This uncertainty is about 0.3 % for the total cross section, both in the electron and the muon channels.

Unfolding model As explained in section 4.7.4, the modelling uncertainty of the unfolding procedure is estimated for each distribution by reweighing the signal Monte-Carlo in such a way that it matches the data and using it as an alternative model for the unfolding. The difference with respect to the central value is taken as the (symmetric) uncertainty. This uncertainty is smaller than 0.1 % for the total cross section.

Data statistical uncertainty The statistical uncertainty in the data is estimated by taking the square root of the number of events measured in each bin. Through its use of the non-diagonal effective inverse of the smearing matrix $\tilde{S}^{-1}$, unfolding introduces correlations between bins. The covariance matrix at the reconstructed level is diagonal because the events are independent from each other and each event contributes to a single bin: $C_{\mathbf{r}}^{\text{data stat}} = \text{diag}((\Delta\mathbf{r}_{\text{data}})^2)$ (where $\Delta\mathbf{r}_{\text{data}}$ is the vector of per-bin uncertainties and the square is applied element-wise). This uncertainty is propagated to the unfolded level by transforming the covariance matrix using the Jacobian matrix $\partial\mathbf{t}/\partial\mathbf{r}$ (see equation (A.22)):

$$C_{\mathbf{t}}^{\text{data stat}} = \frac{\partial\mathbf{t}}{\partial\mathbf{r}} C_{\mathbf{r}}^{\text{data stat}} \left(\frac{\partial\mathbf{t}}{\partial\mathbf{r}} \right)^T. \quad (4.23)$$

The Jacobian is different from $\tilde{S}^{-1}$ because in the D’Agostini method, $\tilde{S}^{-1}$ depends on the observed distribution [149].

Monte-Carlo statistical uncertainty The statistical uncertainty in the Monte-Carlo enters the measurement through the estimation of the smearing matrix S . It is propagated to the unfolded results using ROOUNFOLD [149].

When displayed, uncertainties are grouped to avoid having too many entries. The lepton energy scale and resolution systematics are shown together as “lepton energy”. Likewise, the jet energy scale and resolutions are grouped under “jet energy”. The uncertainties in lepton identification, reconstruction, trigger, and pre-firing efficiencies are shown as “efficiency”. Finally, the small uncertainties in the b tagging efficiency and number of pileup vertices are categorized as “others”. The different sources within each group are independent and therefore they are added in quadrature. No grouping is used for the sources not listed in this paragraph.

The total uncertainty is obtained by calculating the covariance matrix associated with each source and taking the sum. For statistical uncertainties, the covariance is readily obtained from equation (4.23). For systematic uncertainties, a fully correlated covariance matrix is calculated from the deviations $\Delta_\alpha \sigma_i$ according to equation (A.24):

$$C_{\alpha,ij} = \Delta_\alpha \sigma_i \Delta_\alpha \sigma_j. \quad (4.24)$$

Since the variations $\Delta_\alpha \sigma_i$ are allowed to be positive or negative, the off-diagonal coefficients of the covariance matrices can also be of either sign.

4.9 Combination

In the Standard Model, electrons and muons share the same couplings to electroweak bosons and therefore the Z boson branching fractions to electrons and muons are identical up to mass effects, insignificant at the energy scales and precision considered in this analysis. The results obtained in the two channels should therefore be compatible. In addition, the results obtained in the two channels are to a large extent independent from each other. By combining them into a single measurement, systematic and statistical uncertainties will therefore be reduced. In this section, the two channels are first compared at the detector level. Their combination is described afterwards.

4.9.1 Comparison at the detector level

Checking the compatibility between the channels at the detector level is useful not only to show that combining them makes sense, but also to gain confidence in the treatment of the efficiency corrections that are applied in the simulation. Since the trigger, reconstruction, and isolation efficiencies are extracted and applied independently for electrons and muons, any disagreement at the detector level could point to a methodological problem.

Comparing electrons and muons at the reconstructed level is complicated by the vastly different efficiencies: between 35 and 55 % in the electron channel and

Table 4.8: Summary table of all uncertainty sources considered in the analysis and the number of independent components considered in the propagation. When available, a reference to the section describing the source in detail is also provided. “Covariance” in the “Components” column means that the corresponding sources use a covariance matrix to handle correlations between bins.

Source	Components	Section or reference
Luminosity	1	[69]
Trigger and reconstruction efficiencies		
Electrons	1	4.4.3
Muons	1	4.4.5
ECAL pre-firing	1	4.4.3
Lepton identification efficiency		4.4.2, 4.4.4, [108]
Monte-Carlo generator	1	
Final state radiation modeling	2	
Resolution modeling	2	
Background modeling	2	
Tag selection	2	
Statistical: electrons	480	
Statistical: muons	143	
Lepton energy scale		
Electrons	1	4.4.2
Muons	1	4.4.4
Lepton energy resolution		
Electrons	1	4.4.2
Muons	1	4.4.4
Jet energy scale	1	} 4.4.6
Jet energy resolution	1	
Jet b tagging	1	4.4.7
Pileup reweighting	1	4.3
Background cross section	1	4.4.8
Unfolding model	1	4.7.4
Monte-Carlo statistics		
Electrons	Covariance	4.8
Muons	Covariance	4.8
Data statistics		
Electrons	Covariance	4.8
Muons	Covariance	4.8

about 80 % in the muon channel, as can be seen in figure 4.15. In order to cancel this effect, the following double ratio between the numbers of events N_{data} observed in data after background subtraction and the numbers of events N_{MC} predicted by the Drell-Yan simulation is used:

$$R = \frac{N_{\text{MC}}(\mu\mu)/N_{\text{MC}}(ee)}{N_{\text{data}}(\mu\mu)/N_{\text{data}}(ee)}. \quad (4.25)$$

The ratio in the numerator cancels the modeling of the cross section in the simulation as long as it is identical between the two channels. Likewise, the ratio in the denominator allows to ignore the measured physics. Provided that the cross sections for the two channels are identical, the double ratio can also be written in terms of efficiencies ε in data and simulation:

$$R = \frac{\varepsilon_{\text{MC}}(\mu\mu)/\varepsilon_{\text{data}}(\mu\mu)}{\varepsilon_{\text{MC}}(ee)/\varepsilon_{\text{data}}(ee)}. \quad (4.26)$$

With this reorganization, the double ratio can be interpreted as a comparison of the level of description of the total efficiency in the electron and muon channels. The ratio is equal to one if the simulation reproduces the data efficiency perfectly. Any significant deviation from unity may hint at problems in the selection, efficiency measurements, or in the application of the scale factors.

When considering the double ratio, it is important to keep track of the uncertainties. Some of them cancel out, most notably the ones related to the normalization of the Monte-Carlo: the luminosity and the signal cross section. On the other hand, statistical uncertainties and the ones arising from lepton reconstruction efficiency and resolution add up. These effects have been taken into account when plotting the uncertainties for comparison.

Plots of the double ratio as a function of the invariant mass of the lepton pair, the number of reconstructed jets, and the p_{T} of the leading and subleading leptons are shown in figure 4.44. The systematic uncertainties are also shown. The fluctuations in the uncertainty estimates in the tails of the distributions are caused by statistical fluctuations. In the invariant mass plot, the large lepton energy scale uncertainties in the Z peak region are caused by the relatively fine binning in the rapidly varying spectrum of the edges of the peak. The double ratio is compatible with one within the statistical and systematic uncertainties for the control distributions except $m_{\ell\ell}$, for which R is larger than one below the Z peak and smaller than one between 100 and 200 GeV.

The double ratio plots for the $p_{\text{T}}(\ell\ell)$ distributions are shown in figure 4.45 for the first four invariant mass ranges (statistical uncertainties in the highest mass bin are too large for the double ratio to be a meaningful test). The results obtained for $p_{\text{T}}(\ell\ell)$ are representative of the ones for φ_{η}^* and in the one jet region. As for the

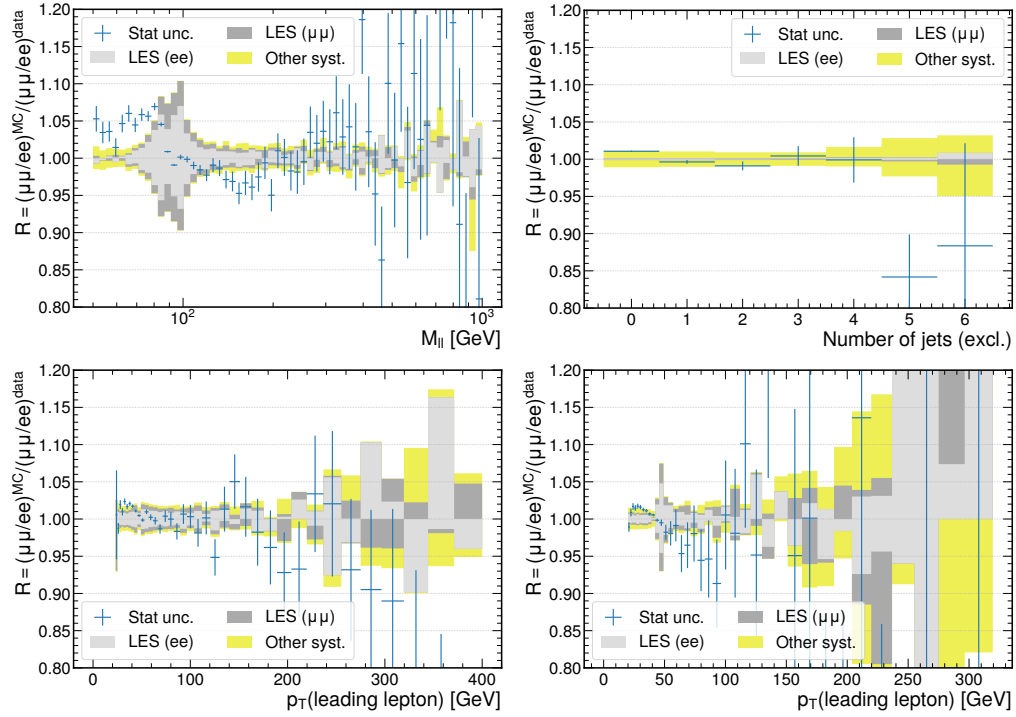


Figure 4.44: Double ratio R as a function of the dilepton invariant mass (top left), the number of jets (top right), and the p_T of the leading and subleading leptons (bottom left and bottom right). The observed values are shown as blue points and the systematic uncertainties as colored boxes around one.

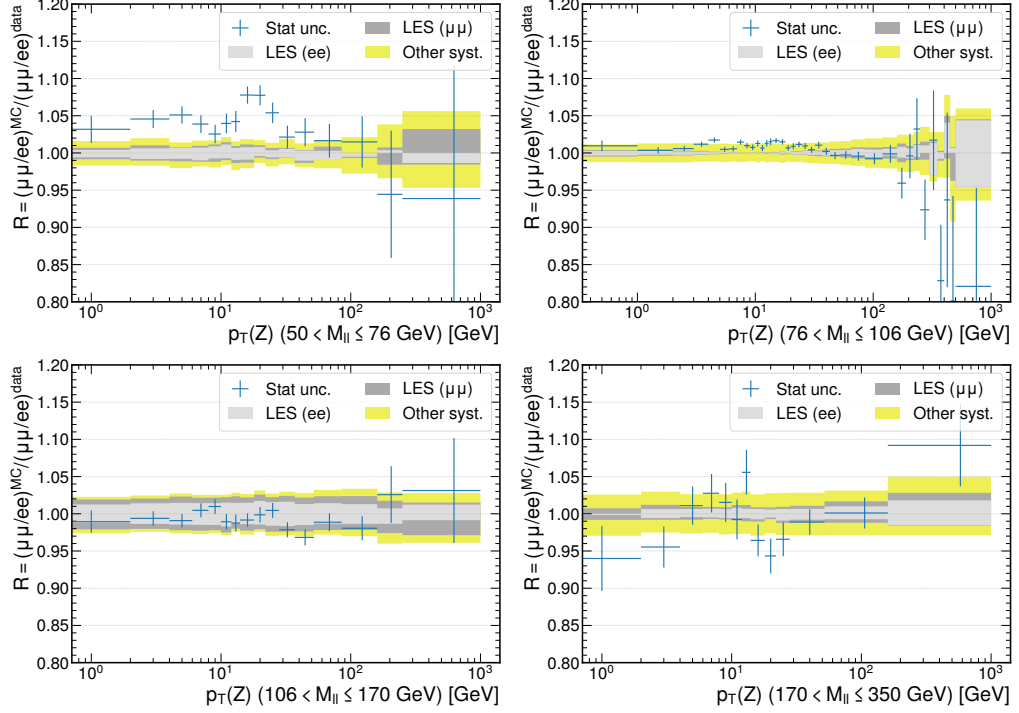


Figure 4.45: Double ratio R as a function of the dilepton p_T in four mass ranges as indicated in the horizontal axis labels. The observed values are shown as blue points and the systematic uncertainties as colored boxes around one.

control distributions of figure 4.44, the values of the double ratio are compatible with one within the statistical and systematic uncertainties except in the lowest mass bin. In that region, particular attention will be given to the compatibility between the two channels after unfolding.

4.9.2 Combination of the results

The combination σ_{comb} of the two unfolded cross sections σ_e and σ_μ in the two channels is performed using a per-bin weighted average. In a bin i , the combination is calculated as:

$$\sigma_{comb,i} = \frac{(\Delta\sigma_{\mu,i})^2 \sigma_{e,i} + (\Delta\sigma_{e,i})^2 \sigma_{\mu,i}}{(\Delta\sigma_{\mu,i})^2 + (\Delta\sigma_{e,i})^2}, \quad (4.27)$$

where $\Delta\sigma_{a,i}$ represents the total uncertainty of the cross section in bin i of channel a . In parallel to the combination, the results in the two channels are compared against each other by plotting their ratio with respect to the combined cross section, as shown in figure 4.46 for $p_T(\ell\ell)$, figure 4.47 for φ_η^* , and figure 4.48 for $p_T(\ell\ell)$ with at least one jet. A quick inspection shows that, depending on the bins of in-

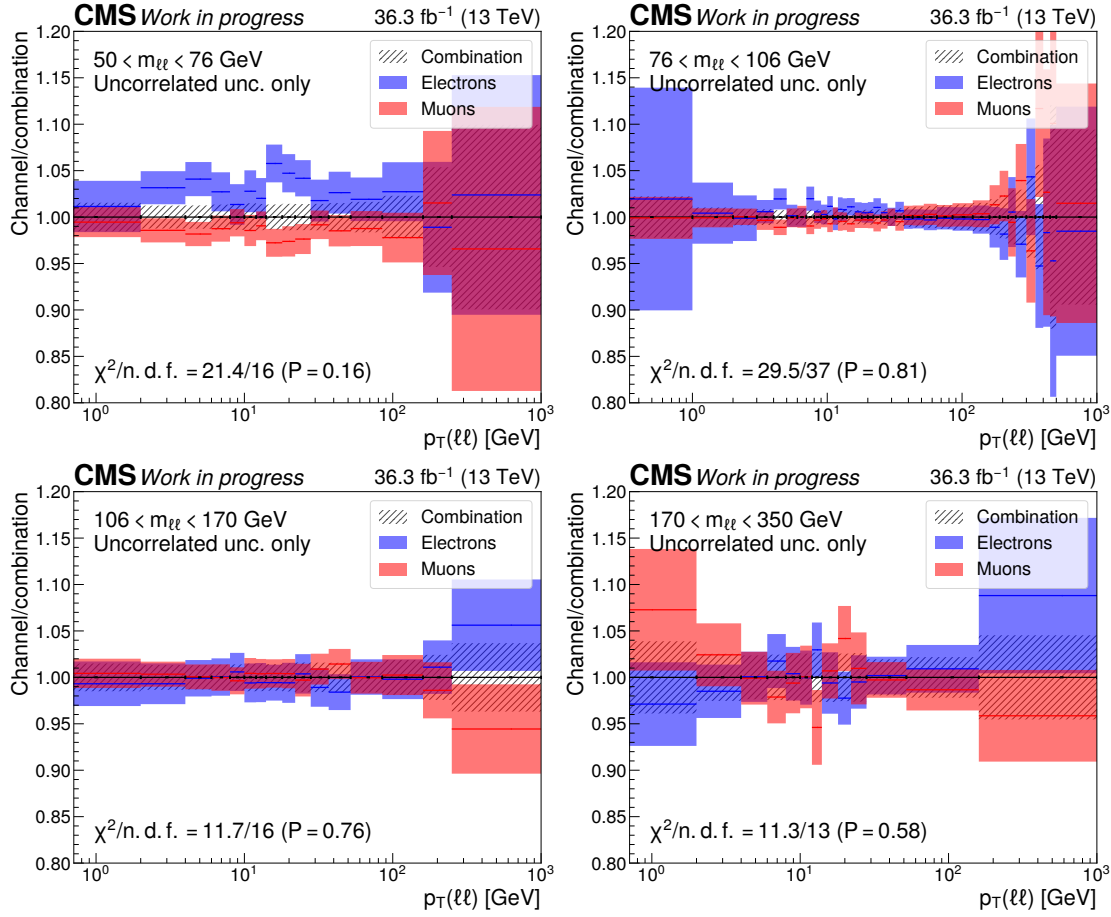


Figure 4.46: Comparison between individual channels and the combined cross sections for the transverse momentum measurement, in the four mass bins where the two channels are used. Only the uncertainties that are not correlated between the channels are displayed.

terest, the combination is dominated by either the electron or the muon channel, indicating that the combination is indeed useful.

The uncertainties shown in figures 4.46 to 4.48 correspond to the sources that apply to a single channel, summed in quadrature. They include the lepton trigger, reconstruction and identification efficiencies, lepton energy scale and resolution, and the statistical uncertainties in the data and simulated samples. In addition, the results of a χ^2 compatibility test between the channels, taking all uncertainties into account, are also displayed. They show that the unfolded cross sections are compatible between the two channels, including at low mass where tension is visible in the double ratio. The worst results are obtained for φ_η^* at high mass, with a difference at the level of 1.3 standard deviations between the channels. This shows

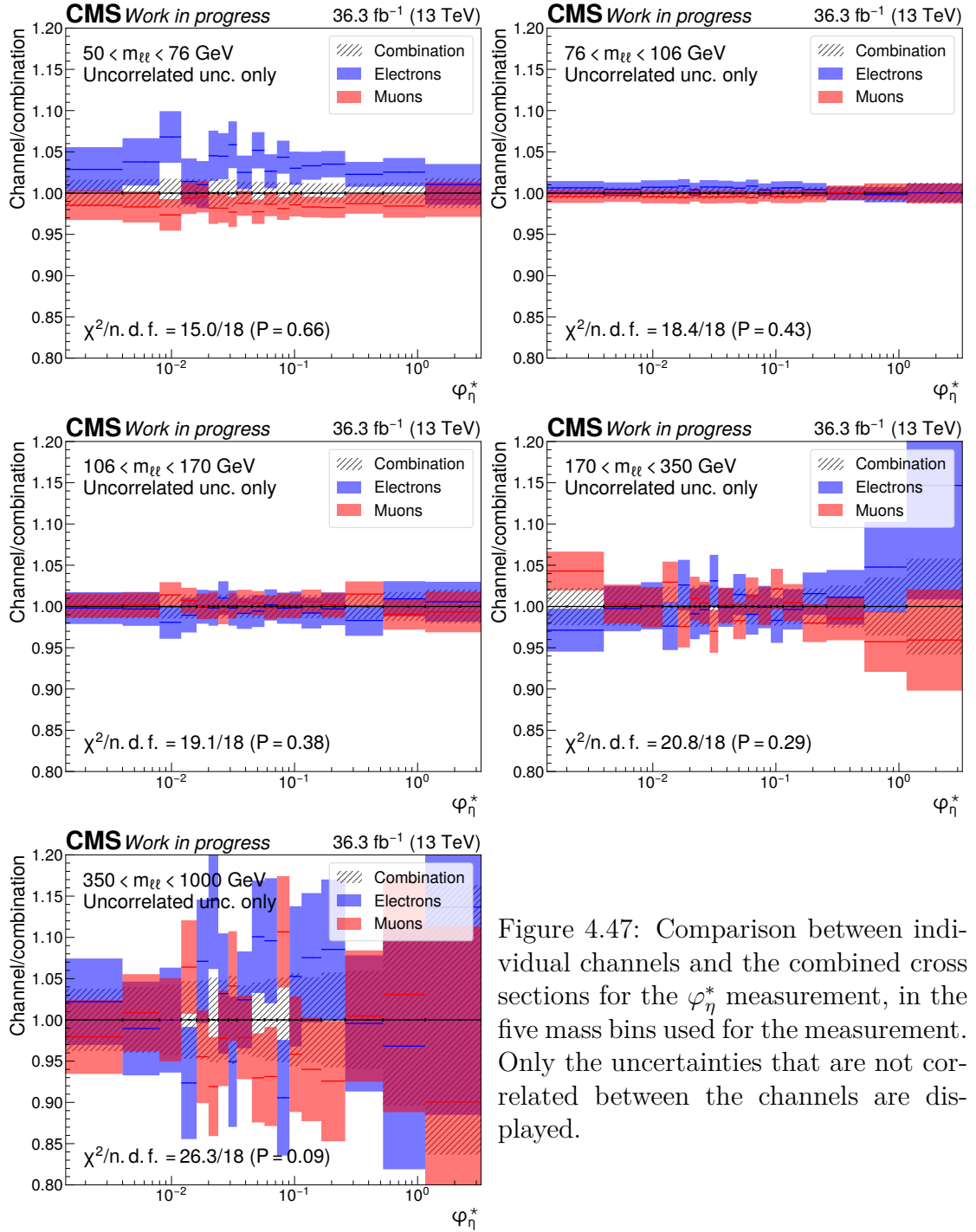


Figure 4.47: Comparison between individual channels and the combined cross sections for the φ_{η}^* measurement, in the five mass bins used for the measurement. Only the uncertainties that are not correlated between the channels are displayed.

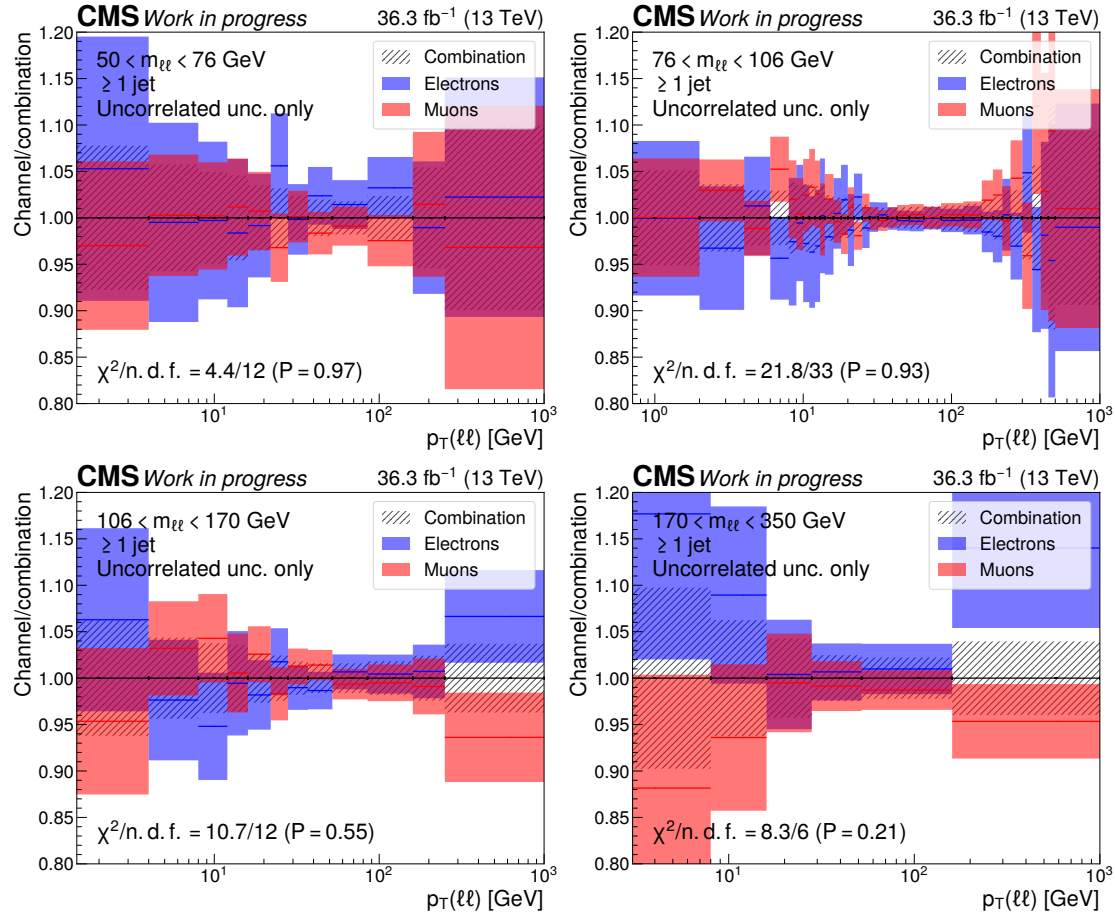


Figure 4.48: Comparison between individual channels and the combined cross sections for the transverse momentum measurement, in the four mass bins used for the measurement. Only the uncertainties that are not correlated between the channels are displayed.

that the channels are compatible and that a combination can be performed.

Uncertainty propagation to the combination is performed in different ways depending on the nature of source. For sources like the luminosity that are correlated between the bins of the distributions, the deviations with respect to the central values are averaged using the same weights as for the cross section:

$$\Delta_\alpha \sigma_{\text{comb},i} = \frac{(\Delta\sigma_{\mu,i})^2 \Delta_\alpha \sigma_{e,i} + (\Delta\sigma_{e,i})^2 \Delta_\alpha \sigma_{\mu,i}}{(\Delta\sigma_{\mu,i})^2 + (\Delta\sigma_{e,i})^2}, \quad (4.28)$$

where α represents the index of the source. The sign of the variations $\Delta_\alpha \sigma_{a,i}$ is preserved to handle possible anti-correlations between the channels. The sign of the combined uncertainty is also tracked, which will be important for the ratio measurements. Sources such as the electron trigger efficiency, that are correlated between the bins but are present in one channel only, are propagated using the same formula with the corresponding variation $\Delta_\alpha \sigma_{a,i}$ in the other channel set to 0.

The only sources that are not fully correlated between the bins are statistical uncertainties from the observed numbers of events and the Monte-Carlo samples used to construct the response matrices. As mentioned in section 4.8, the corresponding covariance matrices (between bins) C_k are provided by the unfolding library. There is no correlation between the individual C_k , and thus they are treated as four different sources. Their propagation is performed by rewriting equation (4.27) in matrix form, with the components of the vectors representing different bins:

$$\sigma_{\text{comb}} = W_e \sigma_e + W_\mu \sigma_\mu. \quad (4.29)$$

With respect to equation (4.27), the bin index i has disappeared and the weights $w_{a,i} = (\Delta\sigma_{b,i})^2 / ((\Delta\sigma_{\mu,i})^2 + (\Delta\sigma_{e,i})^2)$ (with b being the channel other than a) have been grouped in two diagonal matrices $W_a = \text{diag}(w_{a,1}, w_{a,2}, \dots)$. This algebraic manipulation allows to see that a covariance matrix C_k that applies to the channel a can be propagated using equation (A.20):

$$C_{\text{comb},k} = \sum_{a=e,\mu} W_a C_k W_a. \quad (4.30)$$

The result of the propagation is a covariance matrix where the i th diagonal element has been multiplied by $w_{a,i}^2$. Since W_a is diagonal, the correlation matrix after combination is the same as before the combination.

The two seemingly very different approaches described above are particular cases of the more general method used in [153, 154, 155], which is capable of handling arbitrary correlations between the bins and the channels. The methods discussed above are easier to reason about and implement than the general case.

In addition, sign information for correlated sources is preserved, which allows a more precise treatment of correlations between different observables.

The uncertainties in the combination shown in figures 4.46 to 4.48 show significant reductions with respect to the per-channel uncertainties at high $p_T(\ell\ell)$ and φ_η^* in all mass windows. The gains in other regions are small or negligible.

4.9.3 Uncertainties in the combined results

The magnitude of the uncertainties obtained for the combined results is discussed in this section. They are shown in figures 4.49 and 4.50 for the p_T distribution, 4.51 and 4.52 for φ_η^* , and 4.53 and 4.54 for transverse momentum of the lepton pair with at least one jet. Two plots illustrate the precision of each measurement: the breakdown of the total uncertainty, using the grouping described in section 4.8 and the correlation matrix for the total uncertainty, calculated from the covariance matrix C as (no summation on repeated indices):

$$\text{Corr}_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}. \quad (4.31)$$

The interpretation of the correlation matrix is explained in appendix A.6.

Over much of the phase space, the dominant uncertainty for the p_T and φ_η^* distributions is caused by the luminosity measurement, closely followed by the efficiency measurements for lepton triggers, reconstruction, and selection. In the first three mass bins shown in figures 4.49 and 4.51, luminosity and efficiency each contribute about 1% to the total uncertainty, which is about 2% — slightly smaller in the Z peak region where the number of events available for tag-and-probe is large. Other uncertainty sources become significant at high p_T /large φ_η^* : data and Monte-Carlo statistical uncertainties because of the smaller number of events and background normalization. The correlation matrices shown in the right column of the figures translate the fact that the total uncertainty is dominated by systematics with large correlations between the bins: the total uncertainty is almost fully correlated, except for the high p_T regions where statistical uncertainties are significant.

In the two bins of highest mass, figures 4.50 and 4.52, statistical and background uncertainty sources become more important. The unfolding model uncertainty is also significant for transverse momentum in the mass window $350 < m_{\ell\ell} < 1000$ GeV where bin widths are small with respect to the experimental resolution (figure 4.37). The statistical uncertainties cause a reduction of the correlations between bins with respect to lower masses, especially for $350 < m_{\ell\ell} < 1000$ GeV. The correlations between adjacent bins visible in the lower right plot of figure 4.50 are introduced by the unfolding procedure.

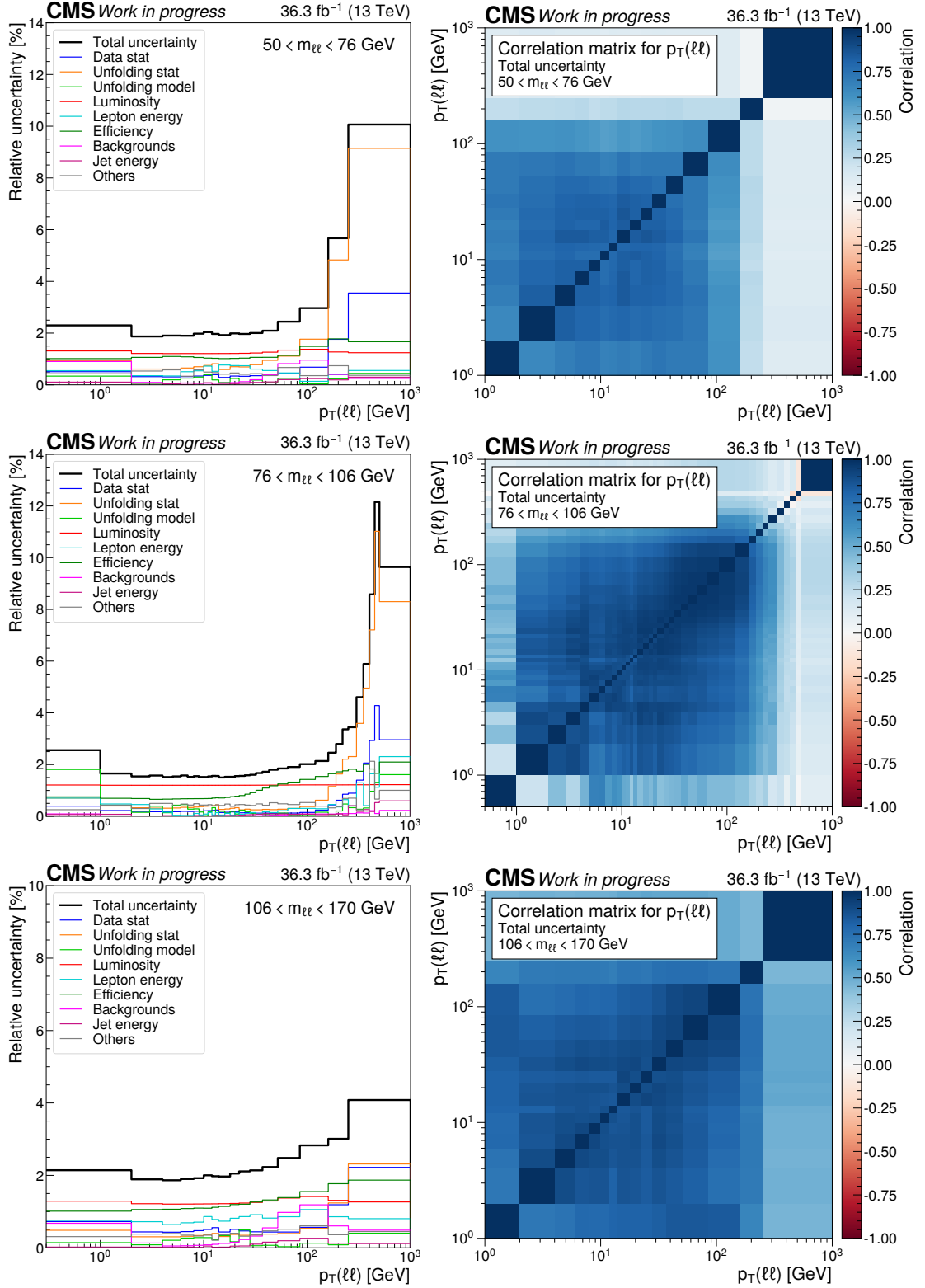


Figure 4.49: Breakdown of uncertainties (left) and correlation matrices (right) for the transverse momentum measurement in the first three mass bins.

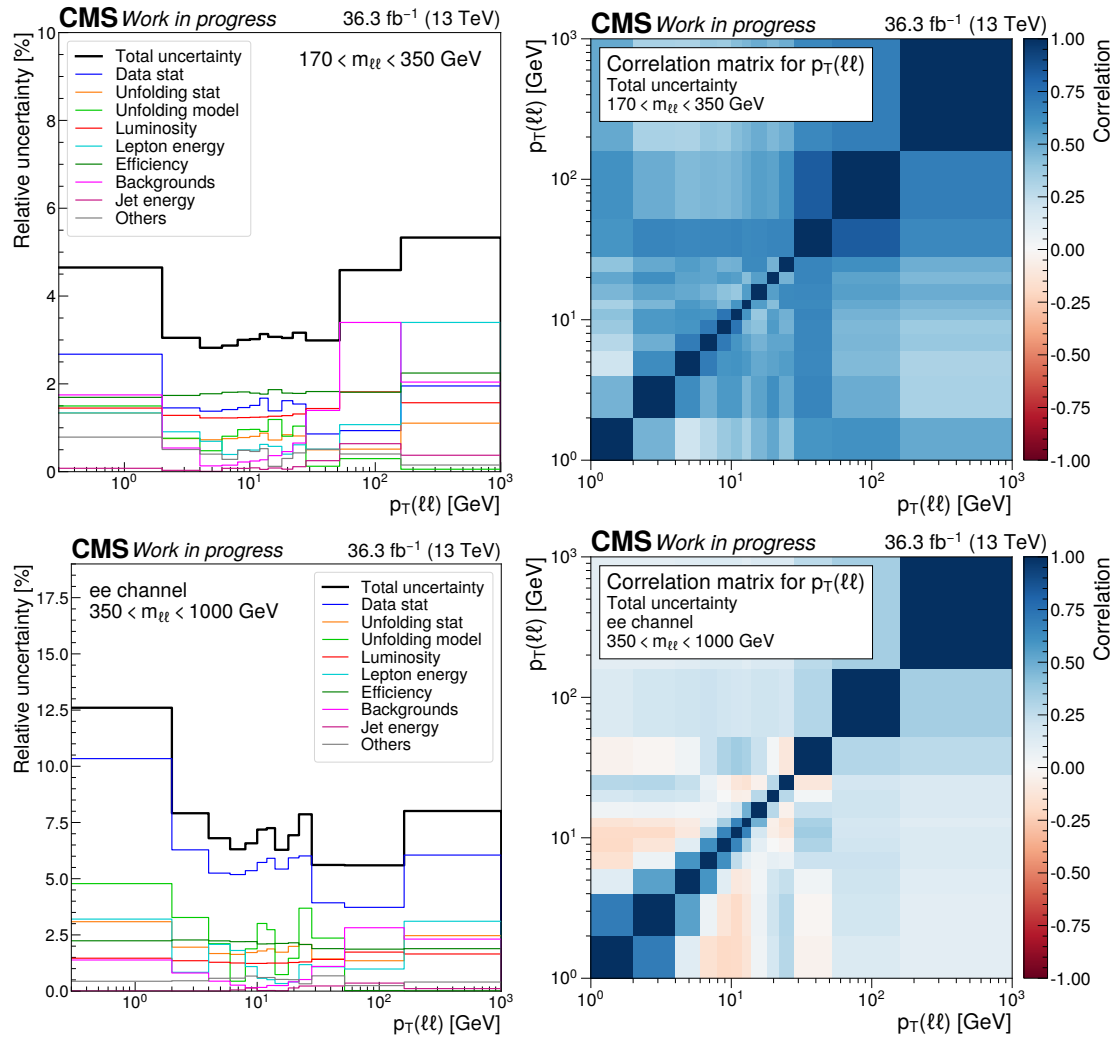


Figure 4.50: Breakdown of uncertainties (left) and correlation matrices (right) for the transverse momentum measurement in the last two mass bins.

Two regions can be distinguished for the uncertainty of the p_T distributions with one or more jets, figures 4.53 and 4.54. For $p_T(\ell\ell)$ smaller than the minimum jet p_T of 30 GeV, the uncertainty is 7–8% and is dominated by the jet energy calibration. This is caused by events with a jet that passes or not the p_T cut depending on which variation of the energy scale is considered. For $p_T(\ell\ell) \gg 30$ GeV, the cross section is dominated by configurations with at least one high- p_T jet that passes the cut regardless of the jet energy scale. This causes a sharp drop in the jet energy scale uncertainty starting at $p_T(\ell\ell) \approx 30$ GeV. The dominant uncertainties at large $p_T(\ell\ell)$ are the same as without requiring a jet: luminosity,

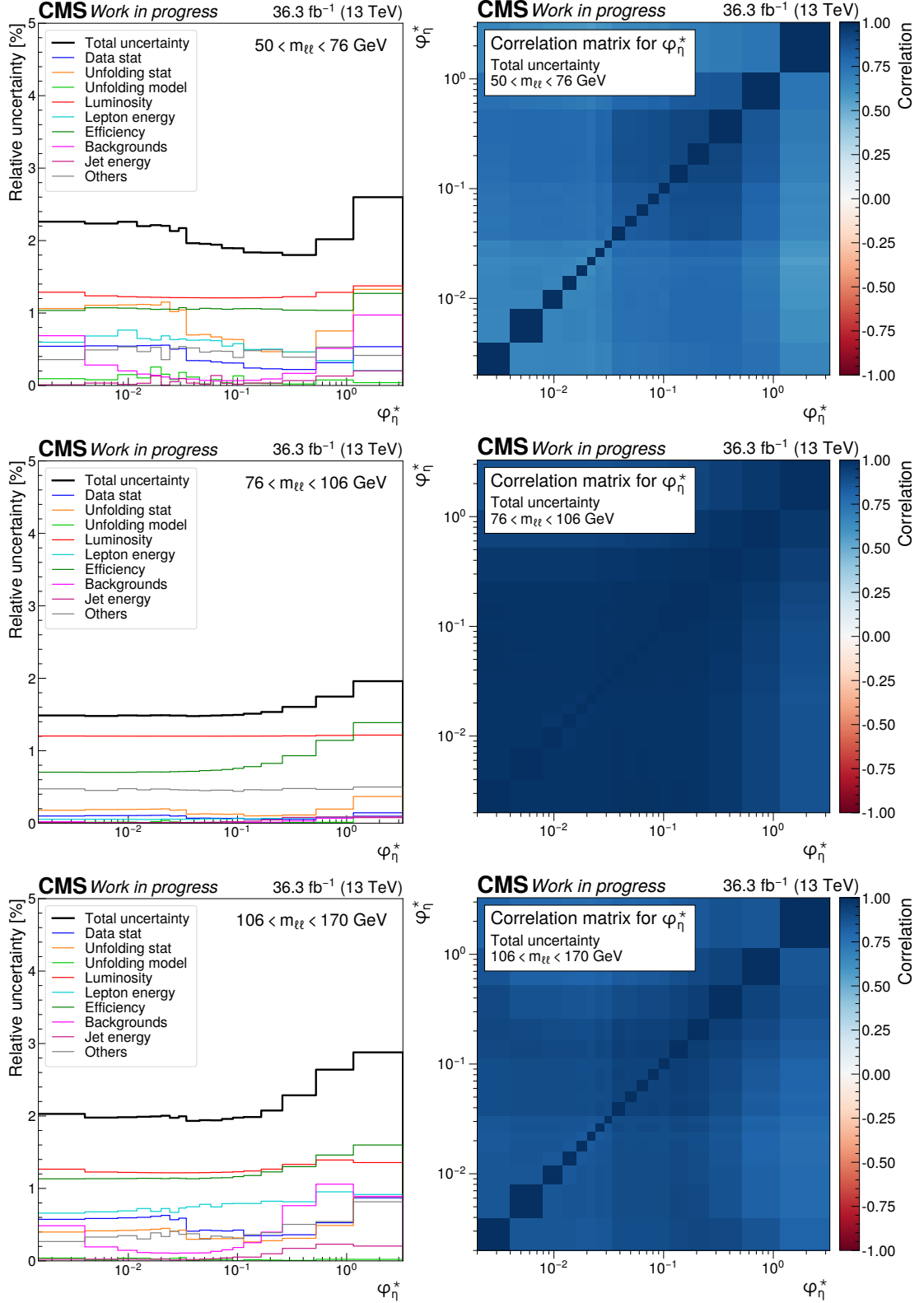


Figure 4.51: Breakdown of uncertainties (left) and correlation matrices (right) for the measured φ_{η}^* distributions in the first three mass bins.

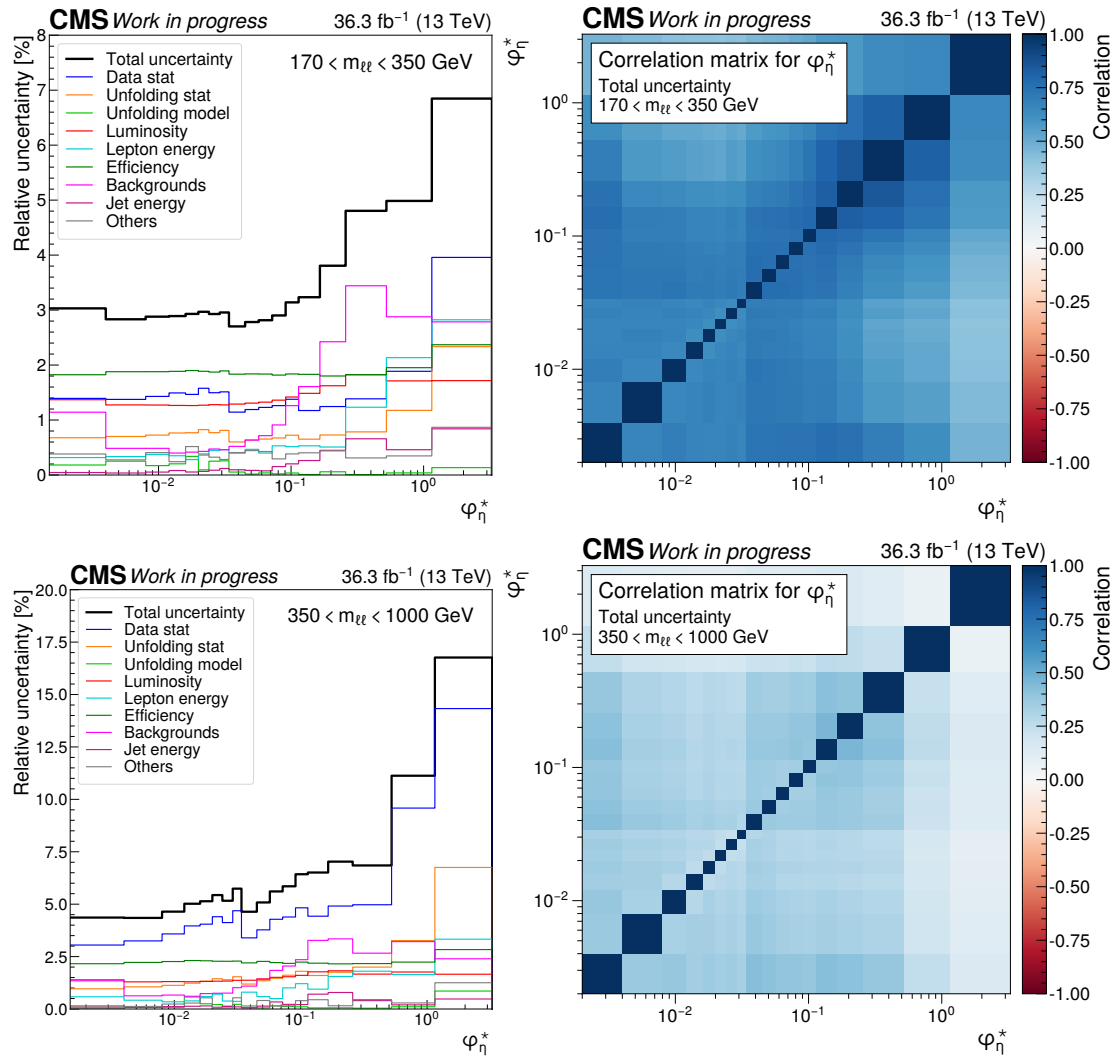


Figure 4.52: Breakdown of uncertainties (left) and correlation matrices (right) for the measured φ_{η}^* distributions in the last two mass bins.

efficiency and statistical. For p_T with at least one jet, the correlations between the p_T bins are significantly smaller than for p_T without a jet requirement. This is caused by the larger statistical uncertainty and the change in the dominant uncertainty around 30 GeV.

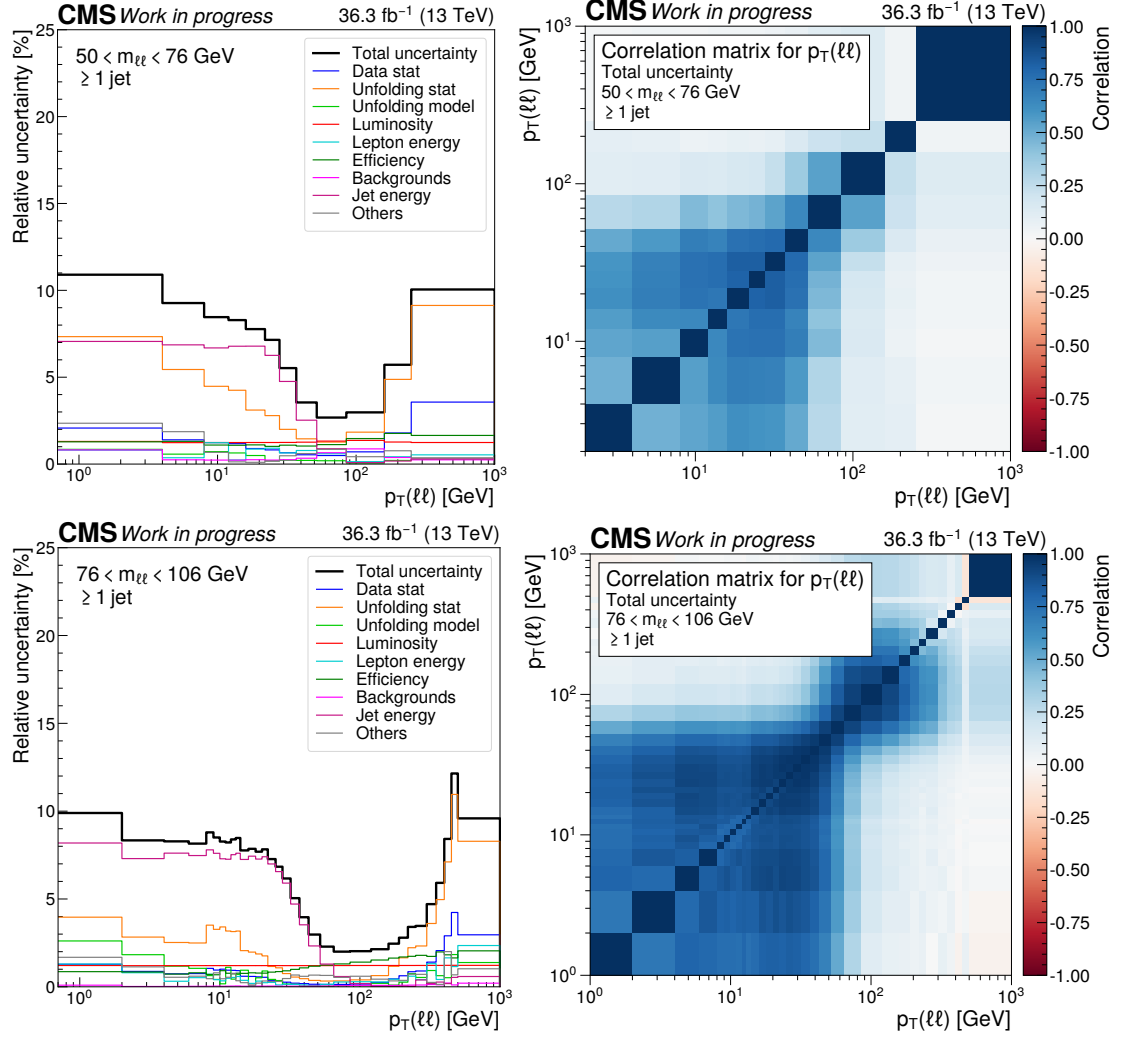


Figure 4.53: Breakdown of uncertainties (left) and correlation matrices (right) for the transverse momentum measurement with one or more jets in the first two mass bins.

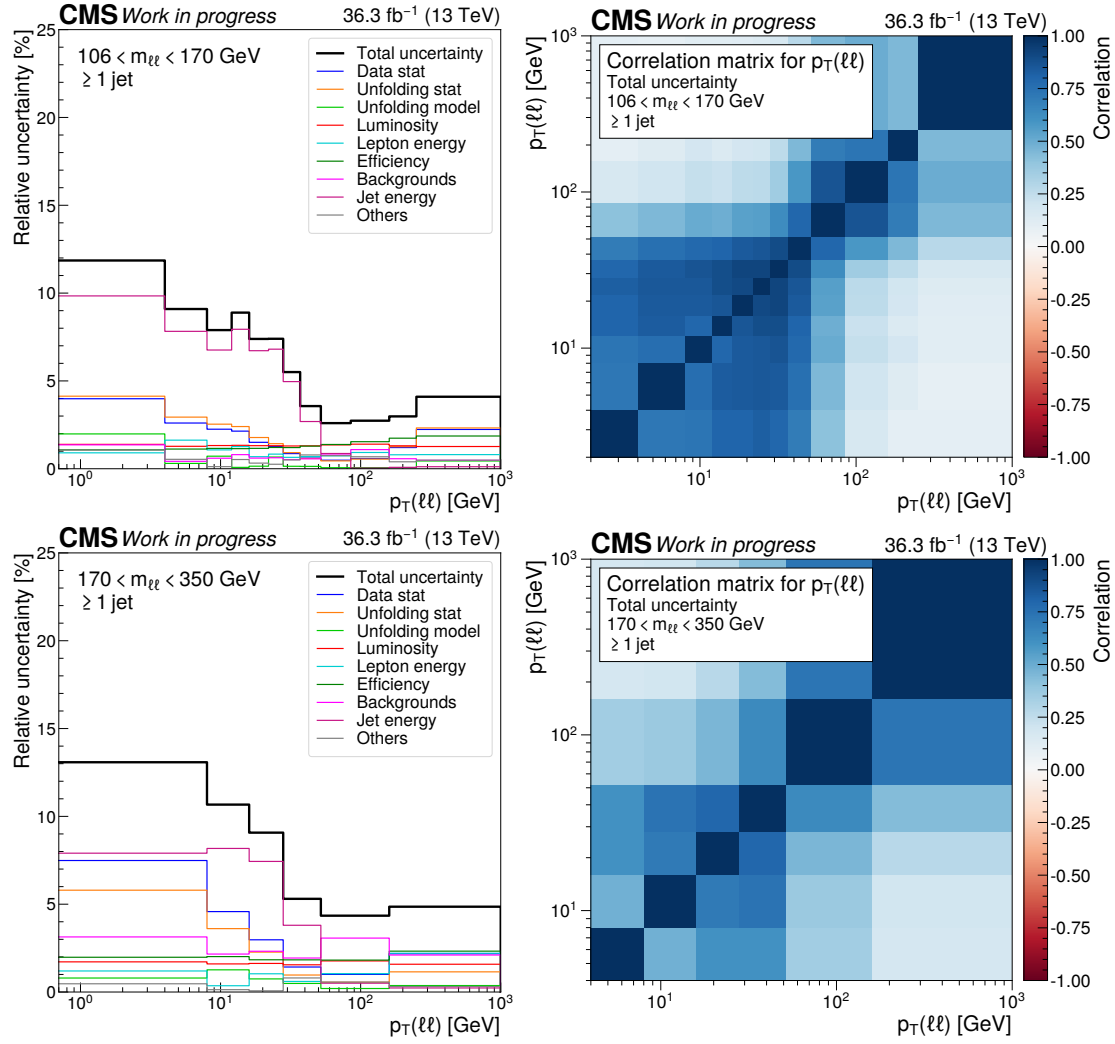


Figure 4.54: Breakdown of uncertainties (left) and correlation matrices (right) for the transverse momentum measurement with one or more jets in the last two mass bins.

4.10 Cross section ratios

In addition to the absolute cross section measurements, the ratios of the differential distributions with respect to the Z peak region are also calculated:

$$\frac{\frac{d\sigma}{dp_T(\ell\ell)}(m_{\ell\ell})}{\frac{d\sigma}{dp_T(\ell\ell)}(m_{\ell\ell} = m_Z)}, \quad (4.32)$$

and likewise for φ_η^* and when requiring at least one jet. The division is performed bin by bin using the binning of the numerator, which is in all cases coarser than in the Z peak region.

The calculation of the uncertainty in the ratios is done in two steps. First, the rebinning of the denominator is taken into account. This is achieved by realizing that any rebinning is a linear operation that can be described with a matrix M, like in the following example where the number of bins is divided by two:

$$\begin{pmatrix} x_1 + x_2 \\ x_3 + x_4 \\ x_5 + x_6 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix}. \quad (4.33)$$

This example can easily be generalized to arbitrary combinations of bins. The rebinning matrix M is then used to propagate the uncertainties, using $\Delta_k \sigma_{\text{rebinned}} = M \Delta_k \sigma$ and $C_{\text{rebinned},k} = M C_k M^T$ for uncertainties that are correlated and uncorrelated, respectively (see equation (A.20)).

To explain the second step, it is first necessary to introduce some notations. The calculation of the ratio $\mathbf{r}$ is recast as a vector function f of two variables $\mathbf{n}$ and $\mathbf{d}$ representing the numerator and denominator, respectively:

$$r_i = f_i(\mathbf{n}, \mathbf{d}) = n_i / d_i. \quad (4.34)$$

Correlated uncertainties are propagated to the ratio by linearizing f , as in equation (A.22):

$$\Delta_\alpha r_i = \frac{\partial f_i(\mathbf{n}, \mathbf{d})}{\partial n_j} \Delta_\alpha n_j + \frac{\partial f_i(\mathbf{n}, \mathbf{d})}{\partial d_j} \Delta_\alpha d_j = \frac{\Delta_\alpha n_i}{d_i} - \frac{n_i}{d_i^2} \Delta_\alpha d_i. \quad (4.35)$$

Since $\Delta_\alpha n_i$ and $\Delta_\alpha d_i$ are allowed to be negative, the uncertainty can increase if the variations of the numerator and denominator have opposite signs. This happens, for instance, for events near the boundaries of the Z peak mass window: if varying the energy scale pushes an event outside the Z peak window, the numerator

increases and the denominator decreases. These cases are, however, not dominant and cancellations occur for most uncertainty sources.

For uncorrelated uncertainties, using equation (A.22) gives the following formula to calculate the covariance matrix (no summation of repeated indices):

$$C_{\mathbf{r},\alpha,ij} = \frac{1}{d_i} C_{\mathbf{n},\alpha,ij} \frac{1}{d_j} + \frac{n_i}{d_i^2} C_{\mathbf{d},\alpha,ij} \frac{n_j}{d_j^2}. \quad (4.36)$$

One can recognize the same factors as in equation (4.35), applied once for each dimension of the covariance matrix. Unlike equation (4.35), no cancellation is possible for uncorrelated uncertainties and the result is always a larger uncertainty.

The total uncertainty in the ratio is calculated by calculating the covariance matrices of correlated uncertainties and adding them to the sum of the covariance matrices of uncorrelated uncertainties, as is done for the cross section measurements. The uncertainties obtained for the cross section ratios are shown in figures 4.55 and 4.56 for transverse momentum, figures 4.57 and 4.58 for φ_η^* , and figure 4.59 for $p_T(\ell\ell)$ with at least one jet. The presentation is identical to the corresponding plots for the cross section measurements.

The cross section ratios reach a precision of about 1 % for the two mass bins around the Z peak region. It is interesting to compare these uncertainties to the ones obtained for the absolute cross section measurements. One notices the nearly complete cancellation of the luminosity uncertainty, except for a small contribution in the regions where backgrounds are large. The efficiency uncertainties are also largely suppressed. For p_T and φ_η^* , the total uncertainty is dominated by the uncertainty in the lepton energy calibration, the background cross sections and statistical uncertainties. Significant cancellations are not expected for any of these sources. As explained above, the uncertainty in the lepton energy scale can increase when taking the ratio. Because background contamination is very small in the Z peak region, the corresponding uncertainty is tiny and doesn't cancel with the larger uncertainties in the other mass ranges. Finally, the statistical uncertainties cannot cancel because the events entering different mass bins are different.

Performing the same analysis for the transverse momentum measurement with at least one jet (figure 4.59), one finds that the jet energy scale and resolution uncertainties almost completely cancel, resulting in a smaller uncertainty than for the cross section measurement for $p_T(\ell\ell) < 30$ GeV. The gain is unfortunately partially cancelled by rather the large statistical uncertainties; cross section ratios for $p_T(\ell\ell)$ with at least one jet could greatly benefit from an extension of the measurement to a larger dataset.

The correlation matrices for the ratios show a significant reduction in the bin-to-bin correlations with respect to the cross sections, caused by the cancellation of the luminosity uncertainty and the larger contribution of statistical uncertainties.

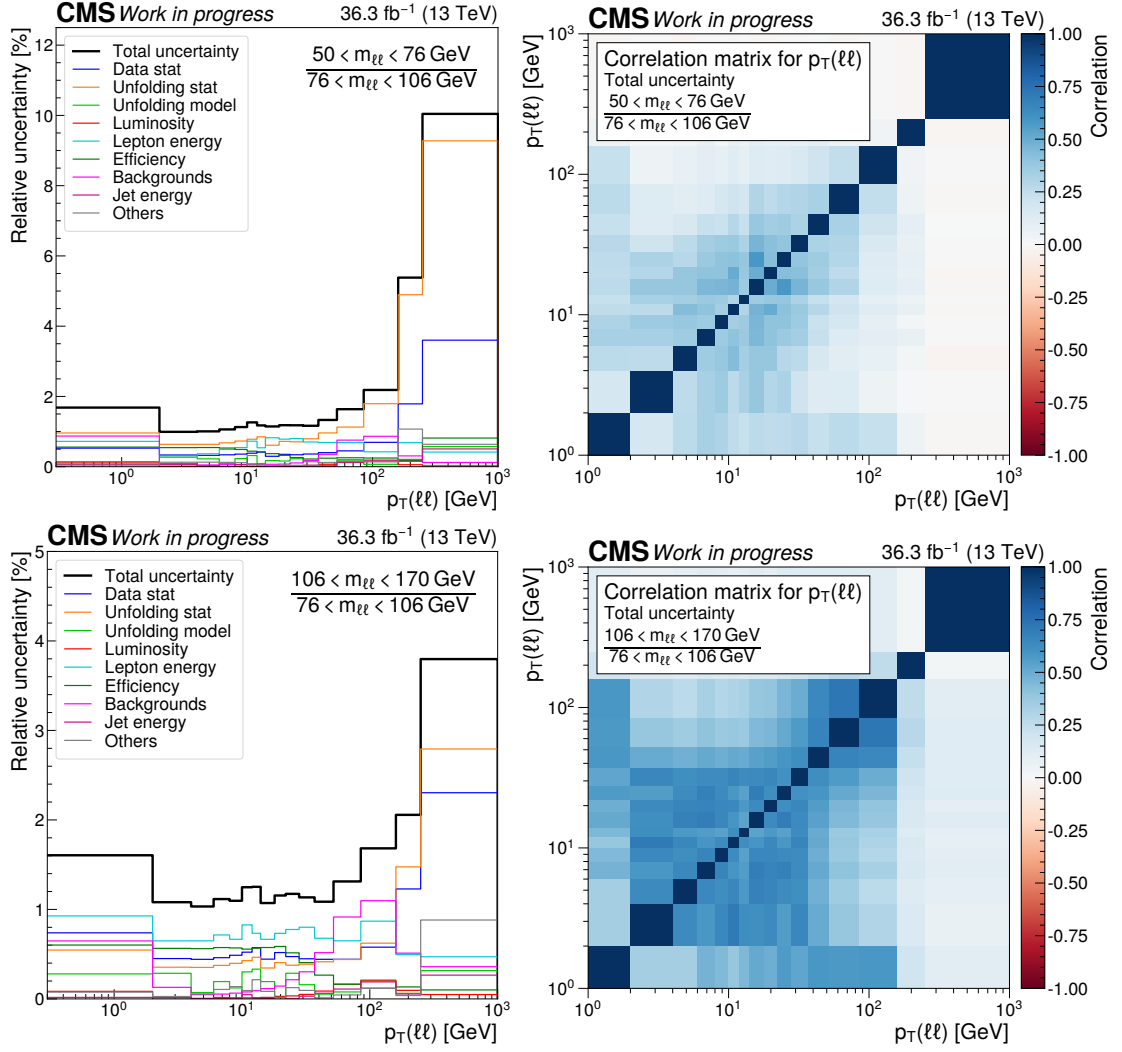


Figure 4.55: Breakdown of uncertainties (left) and correlation matrices (right) for the transverse momentum ratio measurement in the first two mass bins.

One may thus expect fluctuations limited to a single bin, whereas for the absolute cross sections it was not the case.

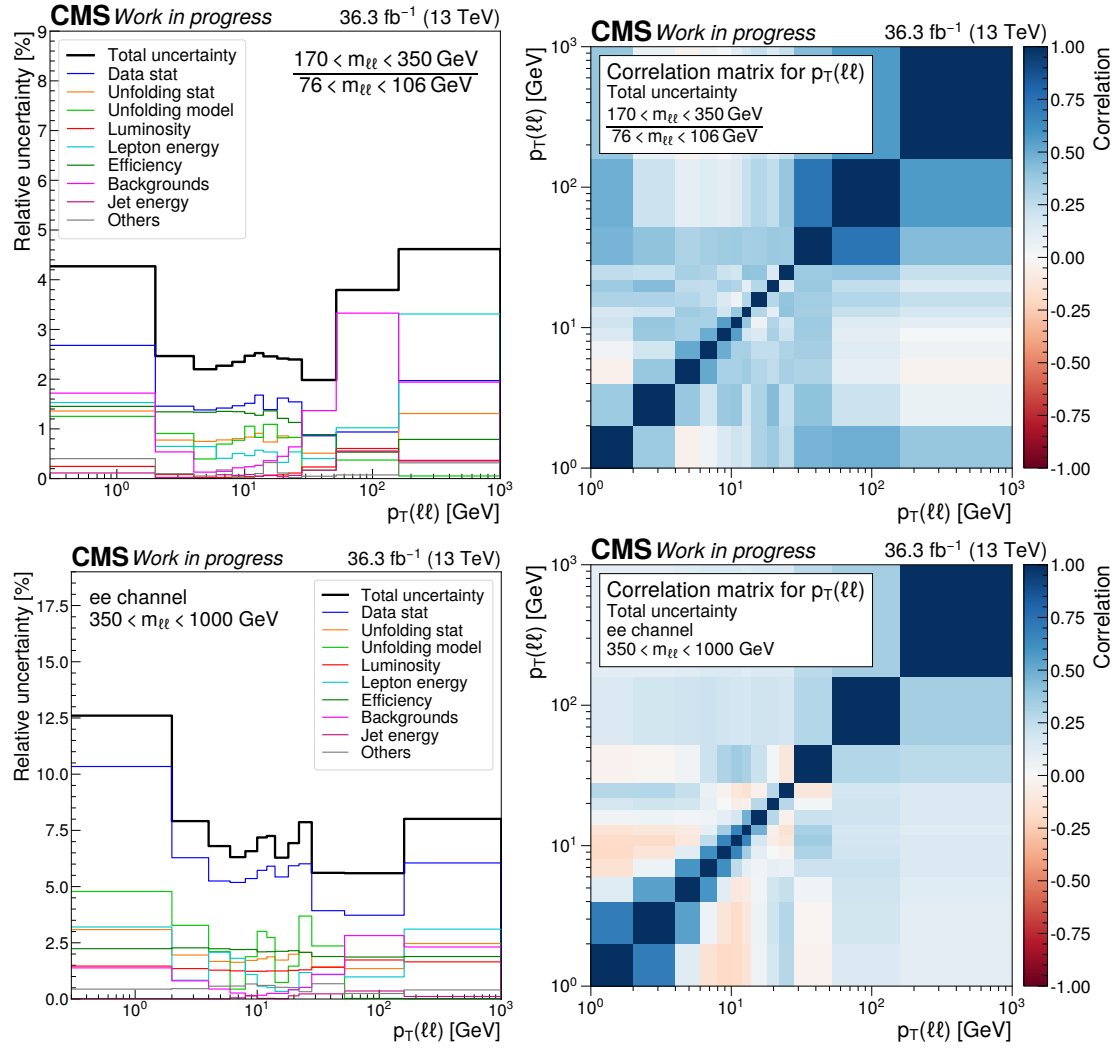


Figure 4.56: Breakdown of uncertainties (left) and correlation matrices (right) for the transverse momentum ratio measurement in the last two mass bins.

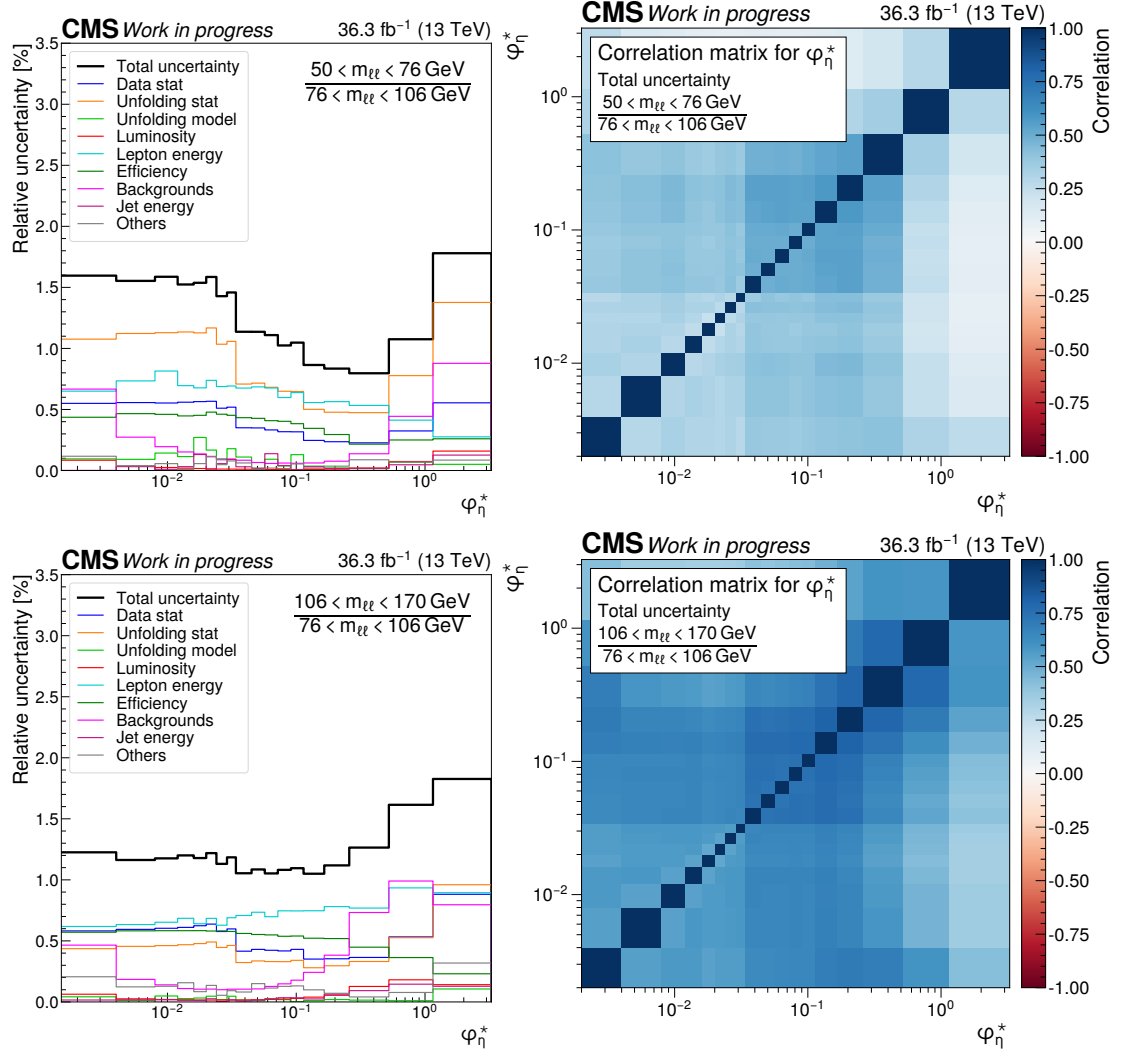


Figure 4.57: Breakdown of uncertainties (left) and correlation matrices (right) for ratios of the measured φ_{η}^* distributions in the first two mass bins.

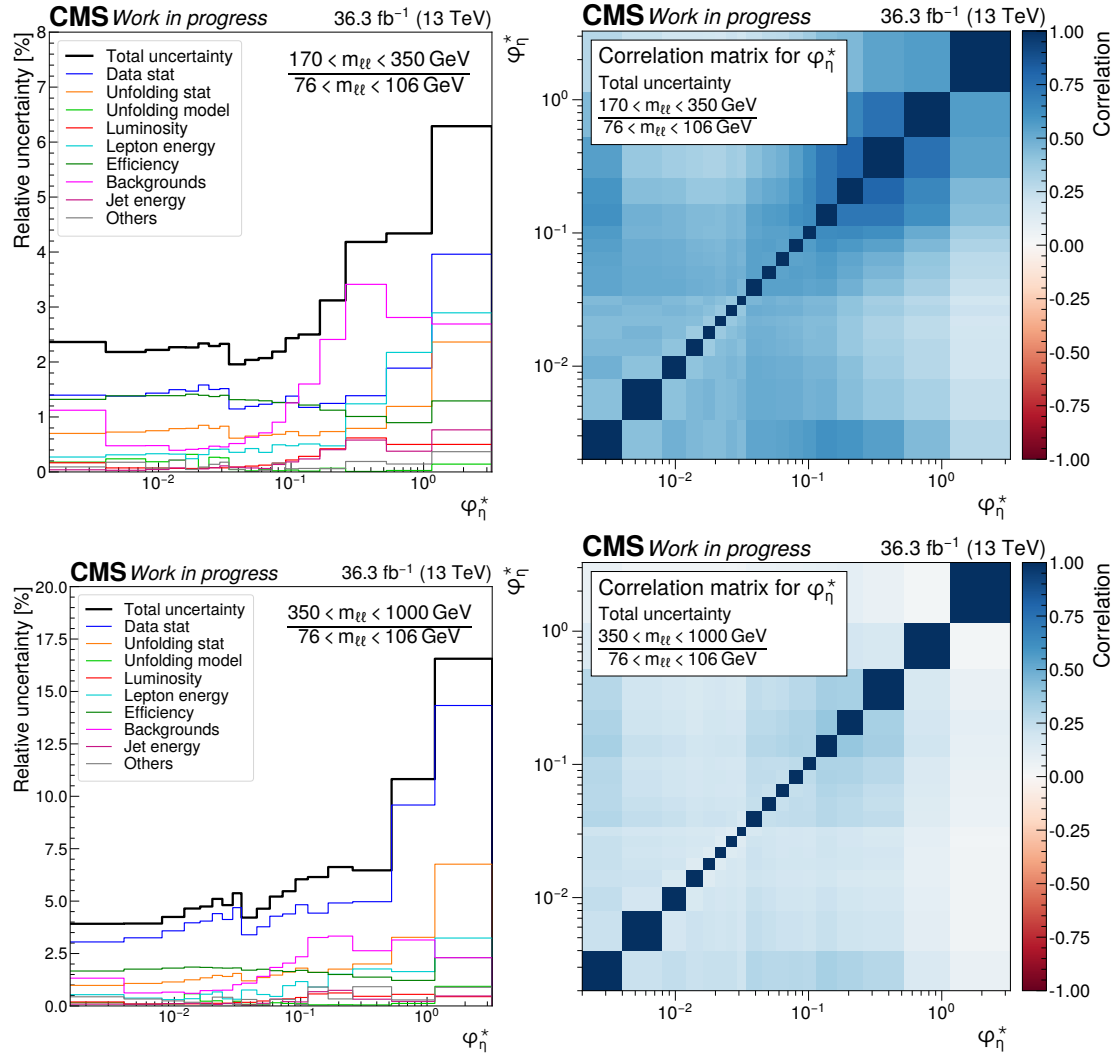


Figure 4.58: Breakdown of uncertainties (left) and correlation matrices (right) for ratios of the measured φ_{η}^* distributions in the last two mass bins.

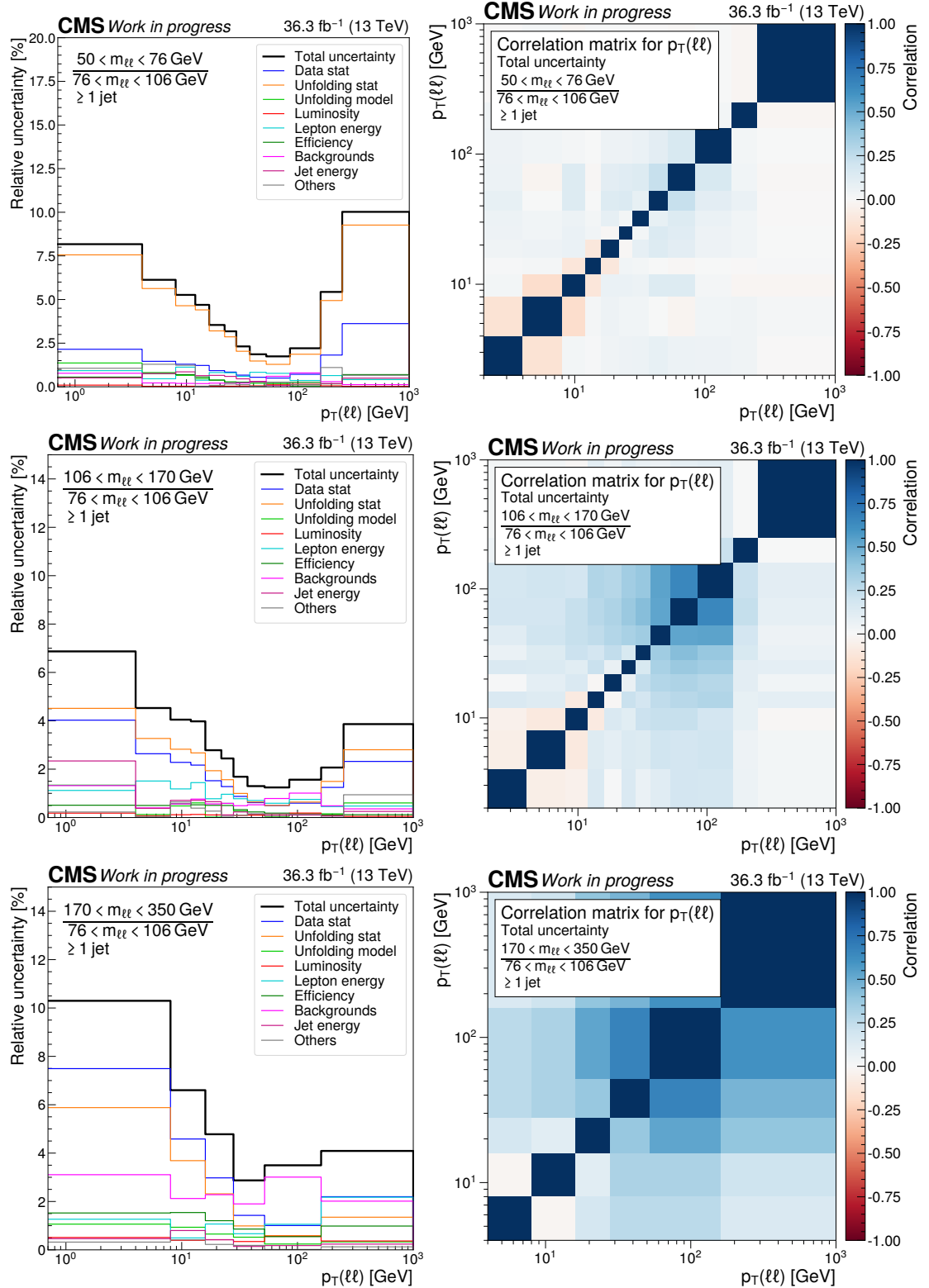


Figure 4.59: Breakdown of uncertainties (left) and correlation matrices (right) for the transverse momentum ratio measurement with one or more jets.

4.11 Predictions

The predictions to which the measurement is compared are described in this section. The first one, described in section 4.11.1, is extracted from the Monte-Carlo samples used to model the signal for unfolding. The ARTEMIDE analytical prediction, based on TMD distributions, is discussed in section 4.11.2, including an extensive discussion of the QED corrections applied on top. This is followed in section 4.11.3 by the description of the CASCADE Monte-Carlo prediction performed in the framework of parton branching TMDs. Sections 4.11.4 and 4.11.5 cover two Monte-Carlo predictions at NNLO using the MINNLO_{PS} and GENEVA methods, respectively. Finally, the calculation of the cross section ratios is covered in section 4.11.6 and the formal accuracy achieved by the calculations for the measured observable is discussed in section 4.11.7.

4.11.1 MadGraph5_aMC@NLO

MADGRAPH5_AMC@NLO [100] (hereunder shortened as MADGRAPH) is a general-purpose Monte-Carlo generator capable of calculating generic cross sections in the Standard Model and simple extensions. It uses the Feynman rules to obtain the list of Feynman diagrams contributing to a process at a given order in α_s , calculate the corresponding amplitudes, and perform the phase space integration. This fully automated method allows the calculation of cross sections with half a dozen particles in the final state; the rise of computational complexity makes more complicated processes prohibitively expensive. MADGRAPH is capable of generating events at leading or next-to-leading order⁴ in α_s , matched to a parton shower. Samples merging several jet multiplicities are also supported through the FxFx method [102].

The technical setup used for the MADGRAPH event samples in this analysis has already been described in section 4.3, and the main points are summarized below. Multiple samples binned in $m_{\ell\ell}$ are used, based on a merged calculation of the Drell-Yan cross section for 0, 1, and 2 accompanying partons at next-to-leading order in α_s . The events are showered using the PYTHIA8 parton shower [43] with the CUETP8M1 tune [39], providing leading logarithmic accuracy at small $p_T(\ell\ell)$, hadronization and other non-perturbative effects including the interaction of multiple parton interactions within a single proton-proton collision.

The uncertainty in the MADGRAPH prediction is estimated using the built-in reweighing capability of the generator. When generating events, MADGRAPH can record additional weights in addition to the central one, corresponding to variations of parameters of the calculation. This feature is used to obtain pseudo-samples:

⁴The recently added support for NLO electroweak corrections [156] (incompatible with parton showers at the time of this writing) is not used in the present analysis.

- Nine pseudo-samples with the factorization and renormalization scales μ_R and μ_F independently varied by factors $\frac{1}{2}$, 1, or 2. The envelope of the variations is quoted as an uncertainty, excluding the two cases where the scales are varied in opposite directions. This method is conventional and is also used for other predictions. The scale uncertainty is the largest theoretical uncertainty, amounting to about 5 % for inclusive distributions.
- 100 pseudo-samples corresponding to the 100 replicas of the NNPDF 3.0 PDF set used for the central value, describing the uncertainties in the extraction of the PDFs. Their standard deviation is quoted as an uncertainty according to the PDF4LHC recommendation [157]. This uncertainty is much smaller than the scale uncertainty.
- Two pseudo-samples corresponding to values of $\alpha_s(m_Z) = 0.118 \pm 0.001$ varied up and down by its uncertainty. The value of α_s affects the PDF determination as well as the MADGRAPH calculation, but the effect on the predicted cross section is negligible. The uncertainty is nevertheless propagated, also according to the PDF4LHC recommendations.

Since the MADGRAPH predictions are obtained using generated event samples, they are also subject to a statistical uncertainty. Despite the large number of simulated events used in the analysis, this uncertainty is the second largest in regions of the phase space where the cross section is small: large p_T , high $m_{\ell\ell}$, and for distributions with at least one jet.

4.11.2 ArTeMiDe

The ARTEMiDE [62] calculation of the Drell-Yan transverse momentum is the only prediction not based on a Monte-Carlo event generator considered in this analysis. ARTEMiDE uses TMD distributions in a framework equivalent to the CSS formalism sketched in section 2.4 to obtain the p_T spectrum of the Z/γ^* boson in the limit $p_T \ll m_{\ell\ell}$. The TMD evolution uses NNLL resummation performed in impact parameter space. A prominent theoretical features of ARTEMiDE is its use of the “ ζ ” scale-setting prescription following a survey of the TMD evolution equations in the (μ_F^2, ζ) plane of the two scales used in TMD factorization [58]. The TMD distributions themselves are derived from the NNPDF 3.1 PDF set [158] using equation (2.41) with an educated guess for the first power correction and a functional parametrization of the remainder. The parameters of the TMD are extracted from an extensive survey of semi-inclusive deep inelastic scattering, deep inelastic scattering, and Drell-Yan data, including LHC Drell-Yan measurements at $\sqrt{s} = 7, 8, \text{ and } 13 \text{ TeV}$ for $m_{\ell\ell}$ between 46 and 150 GeV (the 13 TeV data correspond to the forward region only). Based on comparisons with experimental

data, the results are found to be valid for $p_T \lesssim 0.2 m_{\ell\ell}$. ARTEMiDE doesn't predict φ_η^* nor the p_T distribution with at least one jet.

Two sources of uncertainty are considered by the authors of ARTEMiDE and included for this analysis. The first uncertainty corresponds to variations of the parameters used to describe the non-perturbative region and the second to variations of the scale. No uncertainty in the PDF set on which the extraction is based is considered, even though it would be dominant [62]. Therefore, the ARTEMiDE uncertainties should be considered with caution.

QED corrections

In predictions based on Monte-Carlo generators, like the MADGRAPH sample described in section 4.11.1, corrections for QED final-state radiation are implemented as part of the parton shower. In particular, PYTHIA implements matrix element corrections to describe the first photon emission in the decay of Z/γ^* bosons at leading order in QED. Emissions outside the $\Delta R < 0.1$ cone used for dressing reduce the value of $m_{\ell\ell}$ with respect to m_{Z/γ^*} and may change $p_T(\ell\ell)$ as well. This effect turns out to be important, especially near the lower side of the Z peak. Since QED corrections are not modeled in ARTEMiDE itself, in the present work they are derived from the MADGRAPH sample as described below.

The model used to extract the effect of the photon emissions off the final state leptons and apply it to ARTEMiDE is probabilistic. It is assumed that, given an event with certain m_{Z/γ^*} and $p_T(Z/\gamma^*)$, there exists a probability density for the values of $m_{\ell\ell}$ and $p_T(\ell\ell)$ after photon emissions off leptons in the final state are taken into account, written as:

$$\mathcal{P}(m_{\ell\ell}, p_T(\ell\ell) | m_{Z/\gamma^*}, p_T(Z/\gamma^*)). \quad (4.37)$$

The choice of using a four dimensional function with dependence in both mass and transverse momentum is motivated by the effect of photon emissions on the kinematics of the lepton pair. A photon carrying significant energy away reduces the mass of the lepton pair, but is also likely to change its p_T .

The extraction of $\mathcal{P}$ uses the Les Houches event records [159] produced by MADGRAPH and PYTHIA8. In this format, events are described using a graph structure where each particle has “mothers” and “daughters”. The description of each event is recorded at several steps of its generation: by MADGRAPH with the kinematics of the fixed-order prediction and PYTHIA at various stages of the showering and hadronization. Event records produced by PYTHIA contain the history of the showering encoded using mother-daughter relationships. In order to extract $\mathcal{P}$, QED final-state radiation has to be isolated from the rest of the event record.

In PYTHIA, the QED and QCD showers are interleaved and there is no conceptual separation between them. This allows photon emission from quarks to be treated in the same framework as photon emissions from leptons. In event records, this is reflected by the inclusion of event kinematics before and after showering, with no distinction between QED and QCD effects. This complicates the definition of $p_T(Z/\gamma^*)$, to which both the QCD initial-state shower and QED final-state shower contribute. The following procedure is used to isolate the QED part of the shower:

1. The event record is searched for a Z boson decaying into two leptons (MADGRAPH never encodes γ^* bosons), or alternatively a vertex with two quark mothers and two lepton daughters that MADGRAPH happens to produce. At this vertex, the transverse momentum of the lepton pair is determined only by MADGRAPH: it is zero for events where no additional parton was produced and larger otherwise.
2. All fully showered and hadronized particles (“status 1” in Les Houches terminology) descending from the Z/γ^* are collected. In most cases, this includes two leptons and a small number of photons. Since final-state particles are used, their kinematics take all corrections applied by PYTHIA into account.
3. The kinematics of the Z/γ^* boson, $p_T(Z/\gamma^*)$ and m_{Z/γ^*} , are obtained by summing the four-momenta of the collected particles. With respect to $p_T(\ell\ell)$ obtained from leptons dressed using the cone algorithm, photons generated beyond $\Delta R = 0.1$ are added and photons emitted from quarks that happen to fall within the $\Delta R < 0.1$ cone around a lepton are removed.

The above steps are heavily dependent on the layout of the event graph and the event generators being used. It isn’t surprising since $\mathcal{P}$ has, strictly speaking, no physical meaning. It does, however, fulfill its intended purpose of approximating QED radiation as predicted by the parton shower.

Figure 4.60 shows the values of $\mathcal{P}$ extracted from the MADGRAPH sample in the electron channel (contrary to a common belief, the results are not strongly channel-dependent). Several structures are visible for migrations in p_T , depending on which mass ranges are considered. For over 90 % of the events, photon emissions don’t change anything and they remain in the same (m, p_T) bin after showering. Most events that do migrate end up in a neighbouring p_T bin and stay in the same mass bin. Finally, in about 1 % of the cases, the emission of a photon pushes the event to a lower mass bin. This last possibility is the most interesting for phenomenology.

Photon that cause an event to migrate to another mass bin must carry a significant energy. Since mass bins are significantly wider than p_T bins, the photons

also cause a smearing of the p_T spectrum — typically towards larger p_T , as can be seen in figure 4.60. When the mass difference is large, for instance for migrations between non-adjacent mass windows like in the upper left plot of figure 4.60, so must be the energy of the emitted photon. This creates a distortion of the p_T spectrum for migrating events, pushing the mean p_T towards the value of the mass gap. The same is true for migrations between the Z peak region and the mass window just below it, 50–76 GeV, because the bulk of the cross section in the Z peak region is far from the edge, around 91 GeV.

Migrations between mass bins caused by QED emissions are very important in the region just below the Z peak. Even if only 1 % of the events from the peak migrate, this amounts to a significant contribution between 50 and 76 GeV. Figure 4.61 shows that for p_T around 25 GeV, about 45 % of the events originate from the Z peak region. For predictions to work in this region, it is thus essential that QED corrections of some sort are applied.

In the present case of ARTEMIDE, the QED corrections are implemented by applying the migration probabilities $\mathcal{P}$ described above to the prediction, including migrations between mass bins. Writing $d\sigma_{\text{AR}}/dp_T$ the uncorrected cross section, the following formula is used:

$$\frac{d\sigma_{\text{AR+QED}}}{dp_T(\ell\ell)}(m_{\ell\ell}) = \sum_{m_{Z/\gamma^*}} \sum_{p_T(Z/\gamma^*)} \mathcal{P}(m_{\ell\ell}, p_T(\ell\ell) | m_{Z/\gamma^*}, p_T(Z/\gamma^*)) \frac{d\sigma_{\text{AR}}}{dp_T}(m_{Z/\gamma^*}). \quad (4.38)$$

This corresponds to a simple matrix product if the indices in mass and transverse momentum are thought of as a single dimension. The prediction of ARTEMIDE is flavor-independent, but QED corrections introduce a tiny difference between electrons and muons (less than 0.2 %). When comparing against the combined cross section, the average of the two channels is used.

The uncertainty in the QED correction are estimated from two sources. Firstly, a systematic uncertainty is obtained by extracting the migration probabilities from an event sample generated using the POWHEG generator and showered using the same PYTHIA version and tune as the MADGRAPH sample. The small difference between the two is quoted as an uncertainty. Second, the statistical uncertainty in $\mathcal{P}$ is propagated to the end results.

In principle, the definition of the phase space could be adjusted to reduce migrations by recovering more photons. This can be done, for instance, by increasing the radius in the (ϕ, η) plane of the cone used to dress the leptons. This was studied in reference [153], where it was shown that even a cone with $\Delta R < 0.4$ recovers only about half the events that migrate with $\Delta R < 0.1$. Widening the cone also has the negative effect of capturing more photons unrelated to the leptons, which distorts the invariant mass distribution above the Z peak. Alternative dressing definitions using jet algorithms or following the procedure outlined in reference [160]

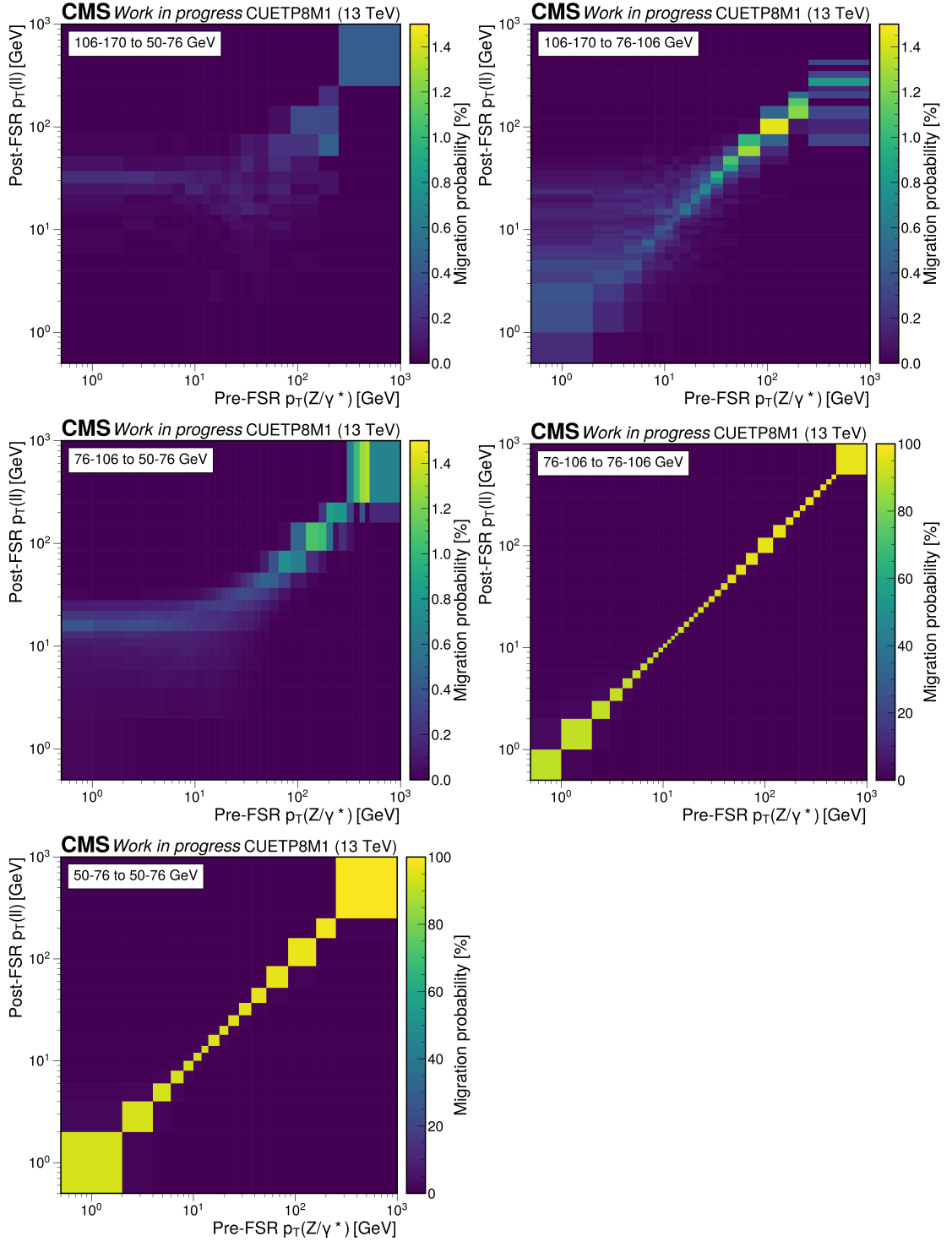


Figure 4.60: Probabilities $\mathcal{P}$ that events migrate in p_T and invariant mass due to photon radiation in the final state. Rows correspond to pre-radiation masses and columns to masses after radiation.

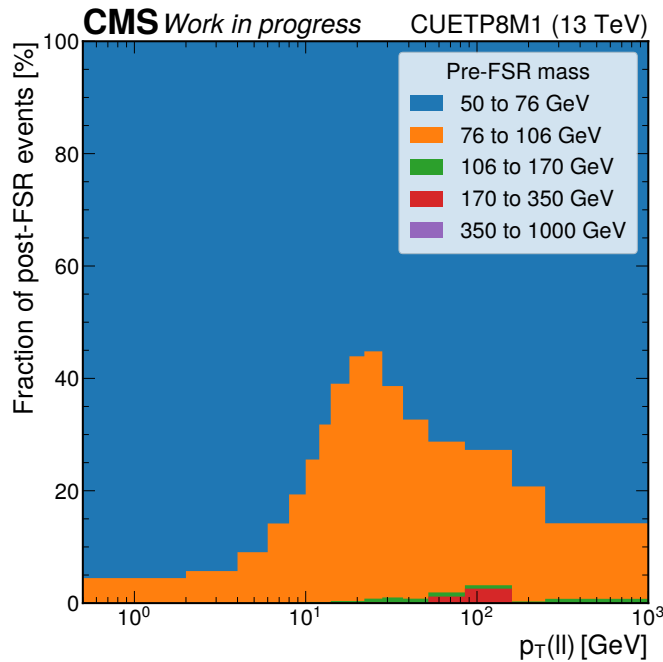


Figure 4.61: Pre-radiation origin of events with a mass between 50 and 76 GeV after QED emissions are taken into account.

could possibly further reduce the migrations; this has yet to be investigated.

4.11.3 CASCADE

CASCADE [161, 162] is a Monte-Carlo generator that matches TMD distributions in the parton branching method (see section 2.3.3) with leading order and next-to-leading order event samples produced externally. Two predictions are used in the present analysis, based on MADGRAPH5_AMC@NLO samples with zero and one parton generated at NLO, respectively. Two samples are used instead of a single merged prediction because support for merging several jet multiplicities was only implemented recently [163]. The PB2018 TMD “set 2” [49] extracted from HERA deep inelastic scattering data is used; CASCADE is the only prediction that doesn’t use LHC data to constrain its description of the initial state dynamics. The final-state showering of the events is performed with PYTHIA version 6.428 [35].

Most of the uncertainties in the CASCADE prediction are inherited from the underlying MADGRAPH calculation and are the same as discussed in section 4.11.1. The only exception is that the PDF uncertainties are replaced by uncertainties in the TMD distributions. The PB2018 TMDs use the Hessian method [157] for uncertainties: 30 replicas of the TMDs are stored, each corresponding to a

1σ variation in parameter space. The total TMD uncertainty is calculated by adding in quadrature the difference between each replica and the central value. Like for MADGRAPH, this uncertainty is small in the phase space covered by the measurement.

4.11.4 MiNNLO_{PS}

The MiNNLO_{PS} [164] method allows the generation of event samples at NNLO matched to a parton shower. It is an extension of the earlier MiNLO [165] and MiNLO' [166] jet merging methods. The MiNLO method is used to calculate cross sections with multiple jets. The coupling constant α_s is evaluated independently at each vertex at a scale that depends on the kinematic configuration, while retaining formal NLO accuracy. Sudakov form factors are used to interpolate between the scales. With MiNLO, an event sample generated with n partons in the final state also predicts finite cross sections for numbers of jets smaller than n . MiNLO' extends MiNLO to obtain formal NLO accuracy for observables involving fewer than n jets. A MiNLO' prediction for $Z + 1$ parton provides NLO accuracy for observables with zero and one jets, and LO accuracy for two jets.

The MiNNLO method stems from the realization that the only missing piece to achieve NNLO accuracy on top of a MiNLO' prediction for $Z + 1$ parton is the inclusion of one additional order for inclusive observables. This was originally achieved by reweighting the 0 jet cross section to NNLO while preserving NLO accuracy for $Z + 1$ jet [167, 168]. In reference [164], the methodology was improved to avoid the computationally intensive reweighting step, leading to MiNNLO_{PS}.

Event generation in the MiNNLO_{PS} method uses the POWHEG method [38, 124] to match the calculation with parton showers and is logically implemented within the POWHEG Box [125] framework. For the sample used in this analysis, Drell-Yan events are generated using the NNLO version of the NNPDF 3.1 PDF set [158] and showered using PYTHIA version 8.240 [43] with the CP5 tune [40]. Final-state radiation in the Z/γ^* decay is handled using the PHOTOS [169] interface to PYTHIA. The uncertainties in the PDF (and the value of $\alpha_s(m_Z)$) and the perturbative scales are estimated using the Hessian method [157] with 100 eigenvectors of NNPDF 3.1 and seven variations of μ_R and μ_F , respectively. The PDF uncertainties are small with respect to the scale uncertainties.

4.11.5 Geneva

GENEVA is a Monte-Carlo generator for the Drell-Yan process matching analytic resummation to a NNLO fixed order prediction. It was originally developed [170] using NNLL' resummation of N -jettiness τ_N and later extended [171] to include p_T resummation at N³LL in the RADISH formalism [32, 33]. The N -jettiness τ_N [172]

is a positive-definite event shape variable designed to be small in events with exactly N jets. Writing p_k the momenta of all particles in an event, N -jettiness is defined as:

$$\tau_N = \frac{2}{(\sum_k p_k)^2} \sum_k \min(q_a \cdot p_k, q_b \cdot p_k, q_1 \cdot p_k, \dots, q_N \cdot p_k). \quad (4.39)$$

In the definition, the momenta q_a and q_b give the beam directions and $q_1, \dots, q_N$ are arbitrary momenta giving the direction of the jets; see reference [172] for details.

In GENEVA, the phase space is split according to resolution variables r_N that distinguish between zero, one and two jets phase spaces. For GENEVA with τ_0 resummation, $r_N = \tau_N$. For p_T resummation, p_T is used for r_0 while τ_1 is still used for r_1 . Two cuts on the resolution variables are used to define three regions:

$$\Phi_0 \text{ events:} \quad r_0 < r_0^{\text{cut}}, \quad (4.40)$$

$$\Phi_1 \text{ events:} \quad r_0 > r_0^{\text{cut}}, r_1 < r_1^{\text{cut}}, \quad (4.41)$$

$$\Phi_2 \text{ events:} \quad r_0 > r_0^{\text{cut}}, r_1 > r_1^{\text{cut}}. \quad (4.42)$$

Configurations with N partons that cause the event to fall in a region Φ_M with $M < N$ are treated as configurations with M partons; in other words, some emissions are considered unresolved. This allows the evaluation of infrared-safe observables as defined in section 2.3 by combining the three regions:

$$\frac{d\sigma}{dO} = \int d\Phi_0 \frac{d\sigma}{dO_0}(\Phi_0) + \int d\Phi_1 \frac{d\sigma}{dO_1}(\Phi_1) + \int d\Phi_2 \frac{d\sigma}{dO_2}(\Phi_2). \quad (4.43)$$

As explained in section 2.3.1, the cuts on r_0 and r_1 introduce large logarithms. In GENEVA, they are resummed analytically to avoid unphysically large differential cross sections.

The results are compared to two predictions using event samples generated using GENEVA with τ_0 and p_T resummation as described in references [170] and [171], respectively. The PDF4LHC15 PDF set [157] and the world average value of $\alpha_s(m_Z) = 0.118$ are used. The events are showered using a specially modified version of PYTHIA version 8.235, which is also used for non-perturbative effects and QED radiation in the initial and final states. It should be noted that the GENEVA authors do not make any claim on the formal logarithmic accuracy of the p_T spectrum obtained after showering. The theory uncertainties are estimated using the conventional 7-fold variations of μ_R and μ_F and variations of scales used for resummation as described in reference [170]. In addition, the statistical uncertainty of the samples is taken into account. The PDF uncertainties are not available but are expected to be similar to the ones found for MINNLO_{PS}. To calculate the total uncertainty, the scale, resummation, and statistical uncertainties are added in quadrature.

Not all GENEVA samples were fully available at the time of submitting this thesis, and as a consequence the GENEVA- τ prediction is based on the electron channel only while the GENEVA- q_T results use the muon channel only. For GENEVA- q_T , a small number of events with very large weights cause significant statistical fluctuations; discussions with the authors to solve this issue are ongoing.

4.11.6 Cross section ratios

The predictions for the cross section ratios are calculated as the ratio of the central values of the cross sections. The statistical uncertainty is propagated using equation (4.36). Where the Hessian method is used, each eigenvector is propagated according to equation (4.35). For PDF uncertainties using the replicas method, the ratio is calculated independently for each replica and the standard deviation of the ratios is used for the uncertainty. Scale variations are handled in a similar way: the ratio is calculated for each variation and the envelope is taken afterwards. These procedures correspond to considering all theoretical uncertainties as fully correlated between the different mass bins.

4.11.7 Formal accuracy

In order to understand the main differences between the predictions described above, it is helpful to consider their formal accuracy in terms of included powers of α_s and resummed logarithms. For this discussion, it is important to note that the order at which a calculation describes an observable depends on the variable considered: it is clear that even an hypothetical N⁵LO calculation could not reliably predict correlations between the seventh and eighth jets. To facilitate reasoning about observables, it is sometimes chosen to refer to the first order in α_s at which an observable is non-trivial as the leading order (LO). The next orders are called NLO, NNLO, etc., as usual. This convention is adopted in the rest of this chapter. When no observable is referred to, the total cross section is assumed.

The first order at which the p_T distribution is non trivial is $\mathcal{O}(\alpha_s^1)$: a quark or gluon needs to be radiated in the initial state for the Z/γ^* boson to acquire non-zero p_T . NLO accuracy for the total cross section thus corresponds to LO for transverse momentum. With this in mind, one finds that the predictions described in the previous sections all reach NLO accuracy for the p_T distribution (and hence φ_η^*) except the CASCADE sample with 0 partons at NLO, which is limited to leading order for the p_T distribution. Since the domain of validity of ARTEMIDE is fully contained in the region dominated by resummation, the question of its perturbative accuracy for transverse momentum doesn't really make sense. The formal accuracy achieved by the various predictions is summarized in table 4.9.

At high p_T , the cross section is dominated by events with several jets, as shown in the left panel of figure 4.62. As a consequence, AMC@NLO, which merges several parton multiplicities at NLO, is expected to provide a better accuracy in this region than other predictions for which events with two jets are predicted at LO (MINNLO_{PS}, GENEVA) or by the parton shower only (CASCADE).

The measurement is also performed for p_T with an additional requirement of at least one jet with $p_T(\text{jet}) > 30 \text{ GeV}$. The part of the $p_T(\ell\ell)$ distribution above the jet threshold, $p_T(\ell\ell) > 30 \text{ GeV}$, is non-trivial at order $\mathcal{O}(\alpha_s^1)$ and thus the predictions have the same formal accuracy as discussed above for the inclusive distribution. This region is referred to as “leading jet” in table 4.9. The part of the distribution below the jet threshold, $p_T(\ell\ell) < 30 \text{ GeV}$, only becomes non trivial when two jets can be emitted, which starts at order $\mathcal{O}(\alpha_s^2)$ — “second jet” in table 4.9. The right panel of figure 4.62 shows that the contribution from events with two jets or more amount for about 25 % of the cross section in this region; the second jet is likely to have a p_T smaller than the 30 GeV required in the selection. The only prediction describing this part of the phase space beyond leading order is MADGRAPH, which obtains NLO accuracy by merging cross sections with 0, 1,

Table 4.9: Perturbative orders included by the predictions considered in the analysis. ARTEMIDE is not listed because its theoretical framework is only valid in a domain dominated by resummation and non-perturbative effects.

Prediction	LO	NLO	NNLO
MADGRAPH			
Total cross section	•	•	
Leading jet/ p_T distribution	•	•	
Second jet	•	•	
CASCADE (0 parton sample)			
Total cross section	•	•	
Leading jet/ p_T distribution	•		
Second jet			
CASCADE (1 parton sample)			
Leading jet/ p_T distribution	•	•	
Second jet	•		
MINNLO _{PS} , GENEVA			
Total cross section	•	•	•
Leading jet/ p_T distribution	•	•	
Second jet	•		

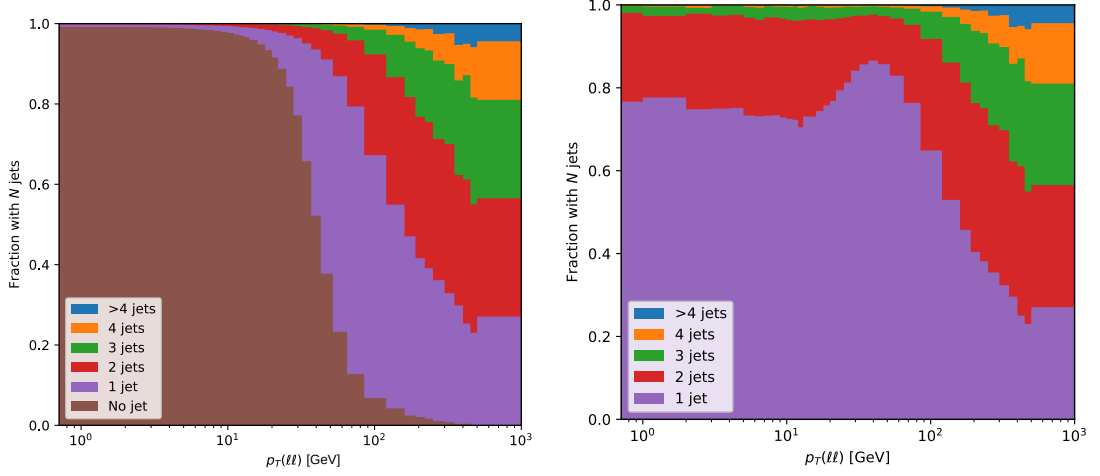


Figure 4.62: Fraction of events with N jets at generator level in the AMC@NLO sample in the Z peak region, as a function of $p_T(\ell\ell)$ for the inclusive cross section (left) and $p_T(\ell\ell)$ for events with at least one jet (right). The jets are required to pass the selection of section 4.7.1, in particular $p_T > 30$ GeV.

and 2 partons in the final state.

The accuracy of the resummation implemented by the predictions is more varied. It is clear that MADGRAPH and CASCADE achieve leading logarithmic accuracy through their reliance upon a parton shower algorithm for resummation⁵. It is generally accepted that the accuracy of parton showers goes beyond the leading logarithm (see e.g. reference [35]) in a different way than analytical resummation. In the following, this accuracy is referred to as “PS” for parton shower.

Two predictions come with explicit claims of a logarithmic accuracy beyond LL for the p_T distribution: ARTEMIDE and GENEVA- q_T , both of which resum the p_T spectrum to N³LL. However, the logarithmic accuracy of GENEVA- q_T is not guaranteed after applying the parton shower [171] and the QED corrections might similarly spoil that of ARTEMIDE. Instead of p_T , GENEVA- τ_0 resums the 0-jettiness τ_0 to NNLL⁷. While in this case resummation remains accurate after showering, resumming τ_0 is not equivalent to a direct resummation of the p_T spectrum. Finally, the inclusion of high-order resummation in the Sudakov form factors used in MINNLO_{PS} may raise hopes of logarithmic accuracy beyond PS, but to the author’s knowledge this has not been studied in detail.

⁵It has recently been suggested that CASCADE achieves NLL accuracy [173].

4.12 Results

4.12.1 Transverse momentum

The measured cross sections in the five considered mass bins are shown in figures 4.63 to 4.67 along with the theory predictions described in the previous section. The ratios of these distributions to the one in the Z peak region, $76 < m_{\ell\ell} < 106$ GeV, are shown in figures 4.68 to 4.71. The data are shown using filled black dots, with error bars representing the statistical uncertainty in the data. The total, systematic plus statistical uncertainty, is displayed as a hatched rectangle around the central value. The horizontal error bars represent the bin widths. More details of the comparison appear in the ratios of prediction/data in the lower two panels of each figure. The uncertainties in the predictions are displayed as colored rectangles showing their various components. In most cases, the scale uncertainties are dominant, followed by statistical uncertainties. Since the measurement covers B rken- x values between 10^{-3} and 0.1, where the PDFs are well constrained by experimental data, the PDF uncertainties are small. For ARTEMIDE, the domain of validity of TMD factorization ($p_T \lesssim 0.2 m_{\ell\ell}$) and the prediction prior to QED corrections are also shown.

Starting from $p_T = 0$, the $p_T(\ell\ell)$ distributions raise until a maximum is reached between 4 and 6 GeV, then fall steeply as p_T becomes larger. Non-perturbative contributions are expected to be significant at small p_T , between $p_T = 0$ and the peak. The largest part of the cross section is located above this region, between the peak and a few tenths of GeV. Events in this region typically do not contain a high- p_T jet and the corresponding configurations are treated using resummation. For larger values of $p_T(\ell\ell)$, the cross section is dominated by events with one or more jets that can reliably be predicted with fixed-order or merged calculations. The three regions of low, moderate, and large p_T are discussed separately below.

Momentum conservation in the transverse plane implies that lepton pairs produced at large p_T are necessarily balanced by one or more high- p_T jets on the opposite side. This makes fixed-order calculations particularly relevant in the region of large $p_T(\ell\ell)$. Referring to table 4.9, all predictions except CASCADE share the same formal accuracy for this part of the phase space, which translates to very similar results: all predictions generally agree with the data within their (rather large) scale uncertainty. Some differences beyond the scale uncertainties remain, for instance for MINNLO_{PS} in the Z peak region and MADGRAPH for $106 < m_{\ell\ell} < 170$ GeV. Finally, one finds that the scale uncertainty is larger for NNLO predictions than for MADGRAPH, suggesting that the inclusion of $Z/\gamma^* + 2$ partons diagrams at NLO and $Z/\gamma^* + 3$ partons at LO capture important effects of N³LO at large p_T .

The ratio measurements show that the part of the cross section at large p_T increases dramatically for large $m_{\ell\ell}$. This growth is well described by all predictions,

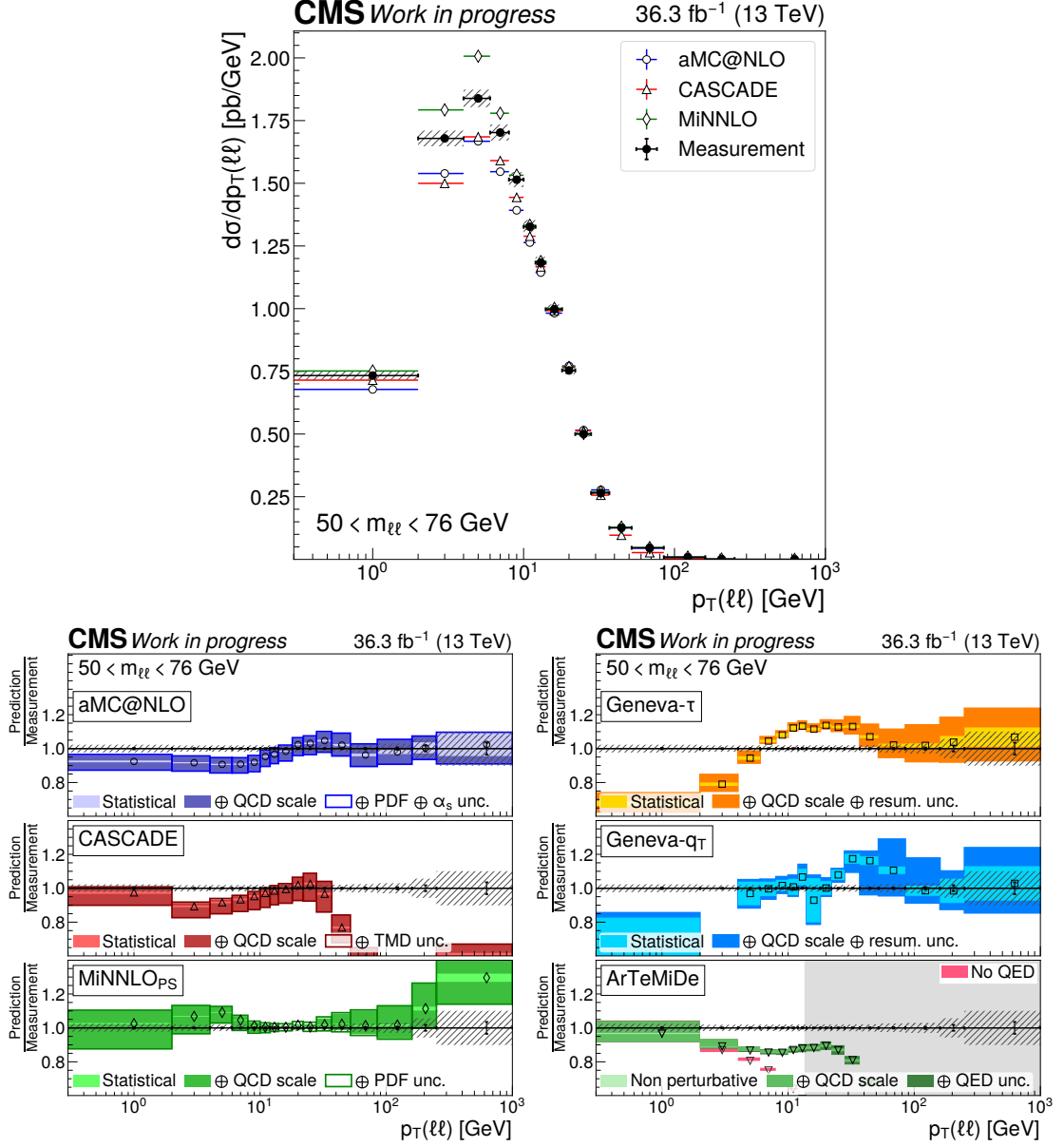


Figure 4.63: Measured cross section as a function of $p_T(\ell\ell)$ for $50 < m_{\ell\ell} < 76$ GeV (top) and comparisons to predictions (bottom).

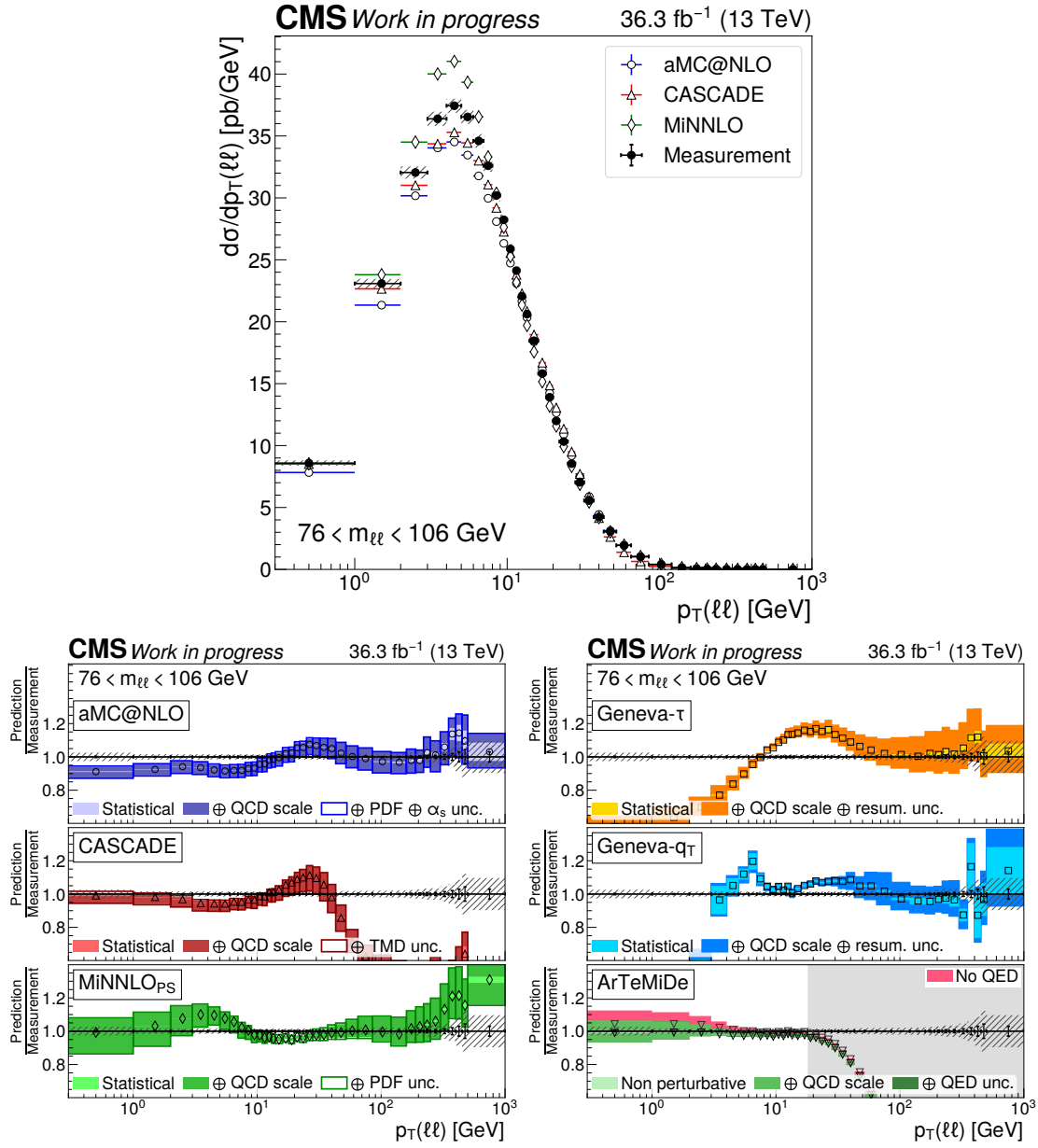


Figure 4.64: Measured cross section as a function of $p_T(\ell\ell)$ for $76 < m_{\ell\ell} < 106$ GeV (top) and comparisons to predictions (bottom).

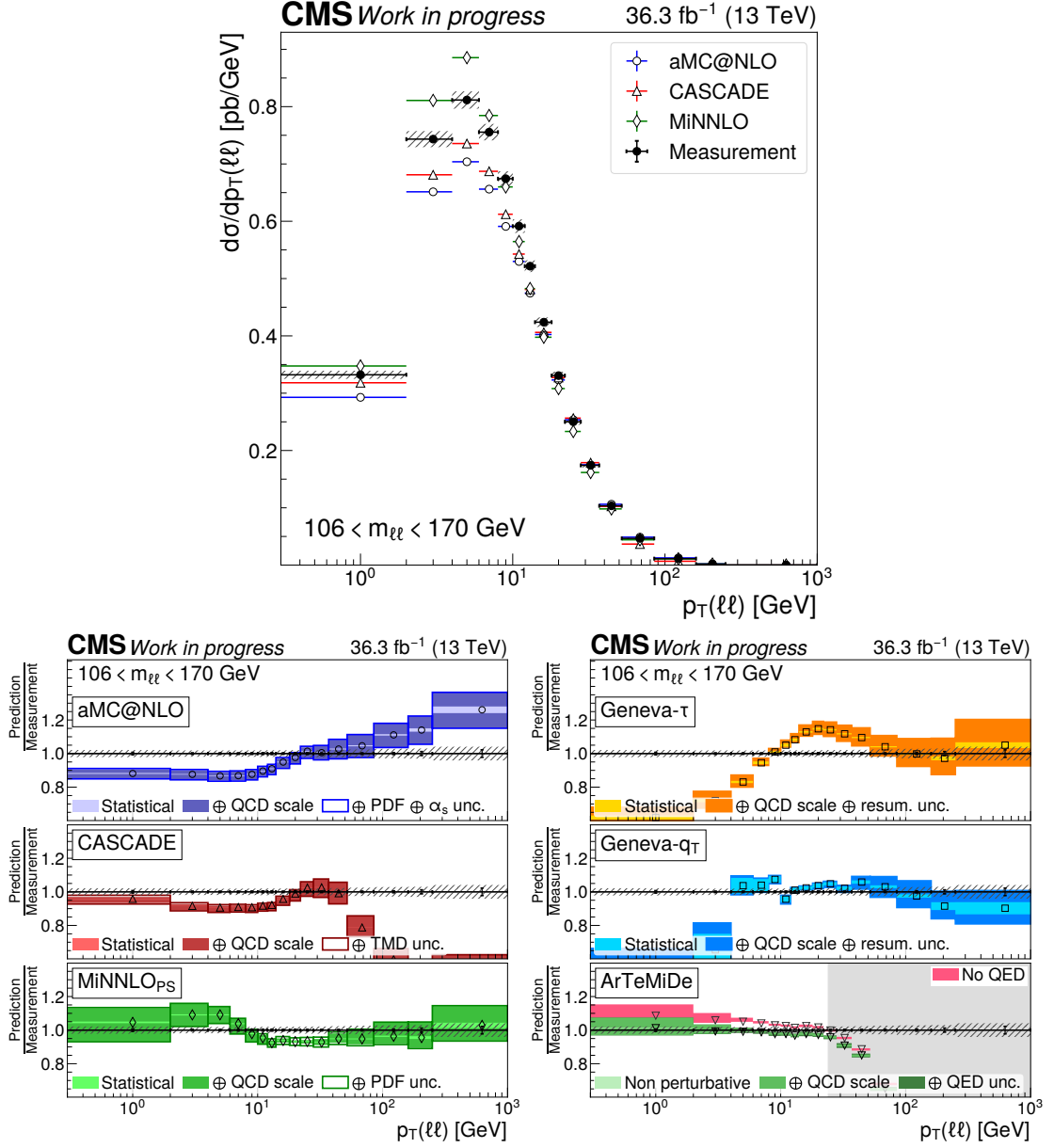


Figure 4.65: Measured cross section as a function of $p_T(\ell\ell)$ for $106 < m_{\ell\ell} < 170$ GeV (top) and comparisons to predictions (bottom).

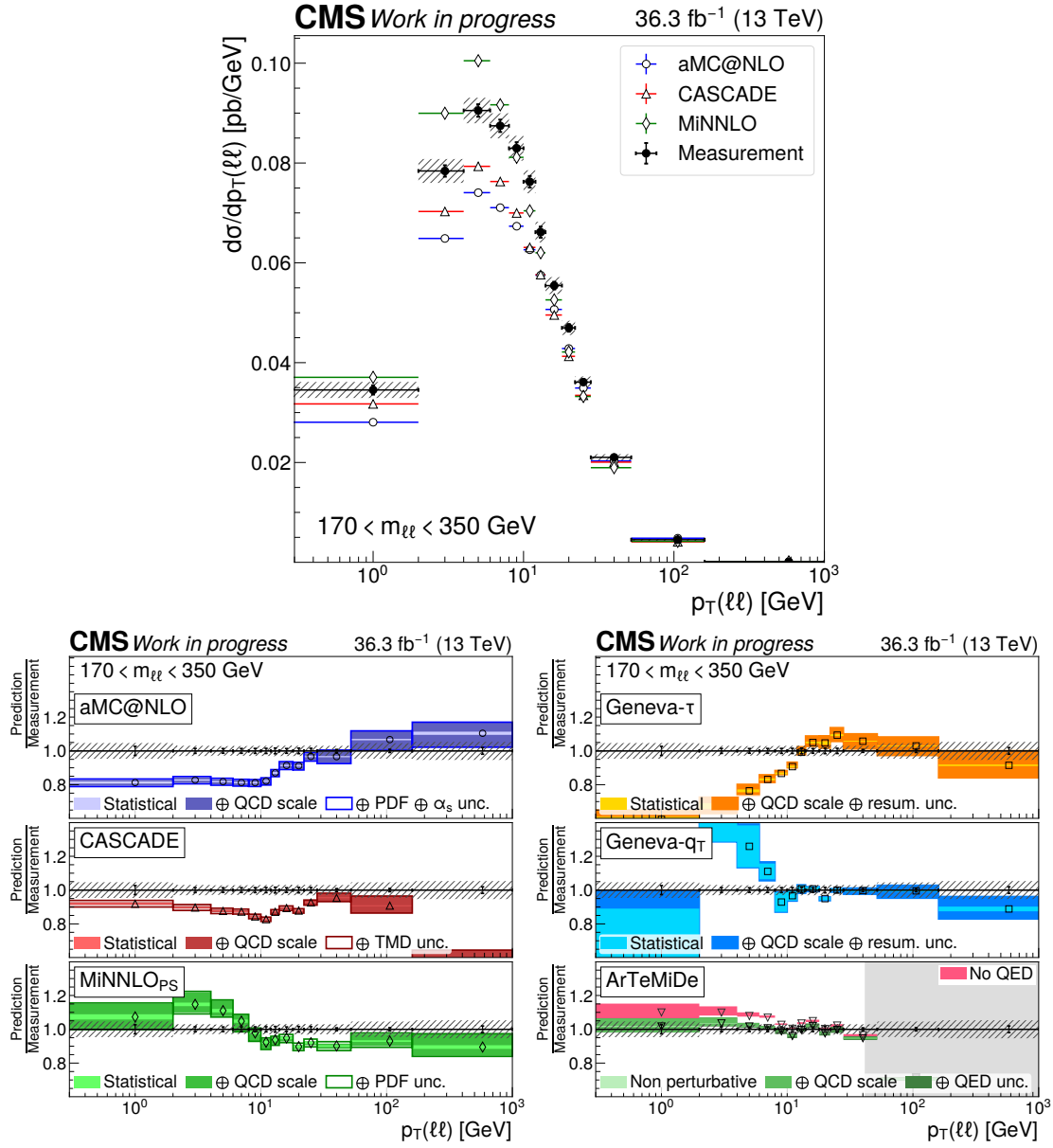


Figure 4.66: Measured cross section as a function of $p_T(\ell\ell)$ for $170 < m_{\ell\ell} < 350$ GeV (top) and comparisons to predictions (bottom).

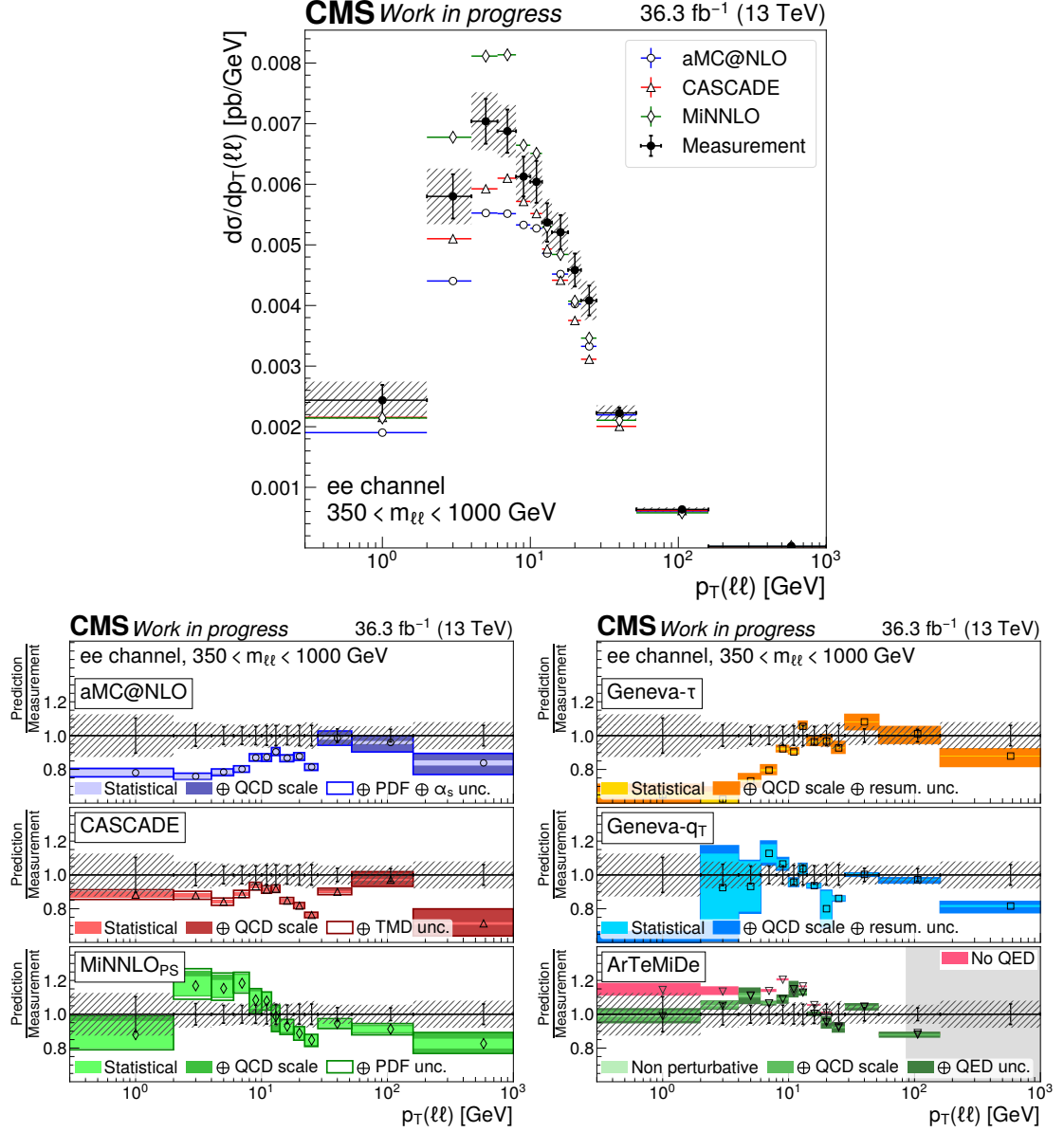


Figure 4.67: Measured cross section as a function of $p_T(\ell\ell)$ for $350 < m_{\ell\ell} < 1000$ GeV (top) and comparisons to predictions (bottom).

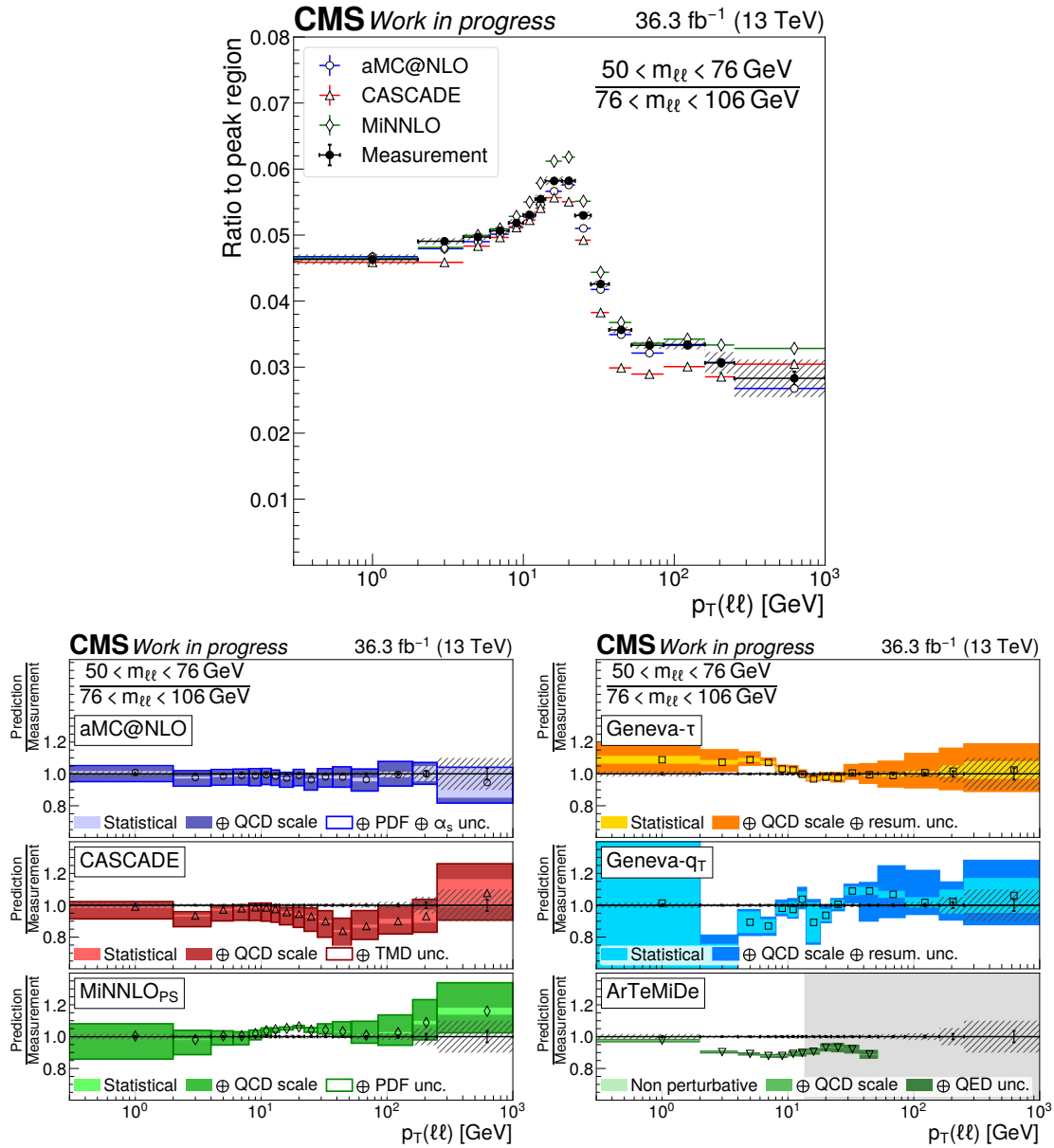


Figure 4.68: Measured cross section ratio as a function of $p_T(\ell\ell)$ for $50 < m_{\ell\ell} < 76$ GeV (top) and comparisons to predictions (bottom).

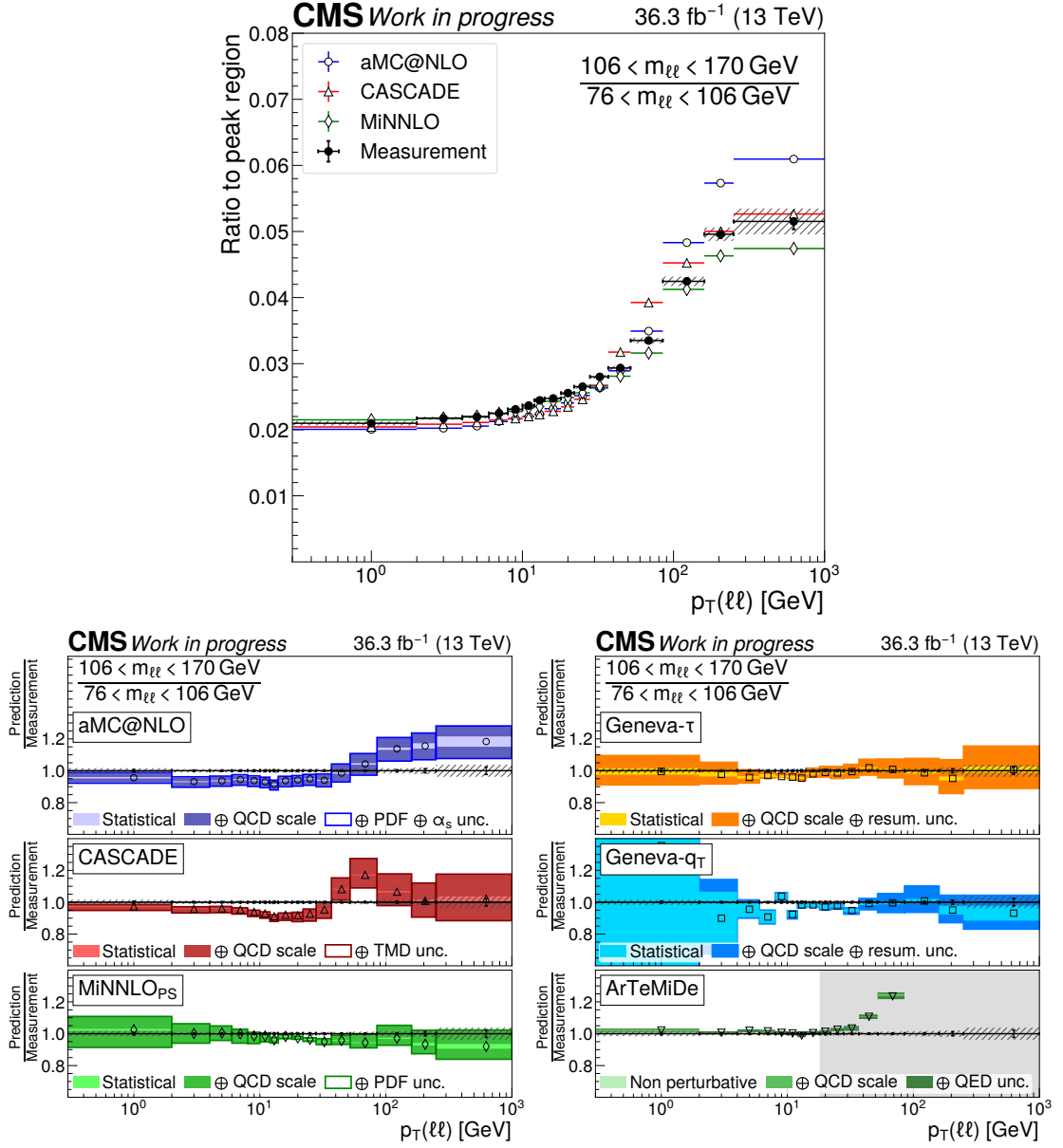


Figure 4.69: Measured cross section ratio as a function of $p_T(\ell\ell)$ for $106 < m_{\ell\ell} < 170$ GeV (top) and comparisons to predictions (bottom).

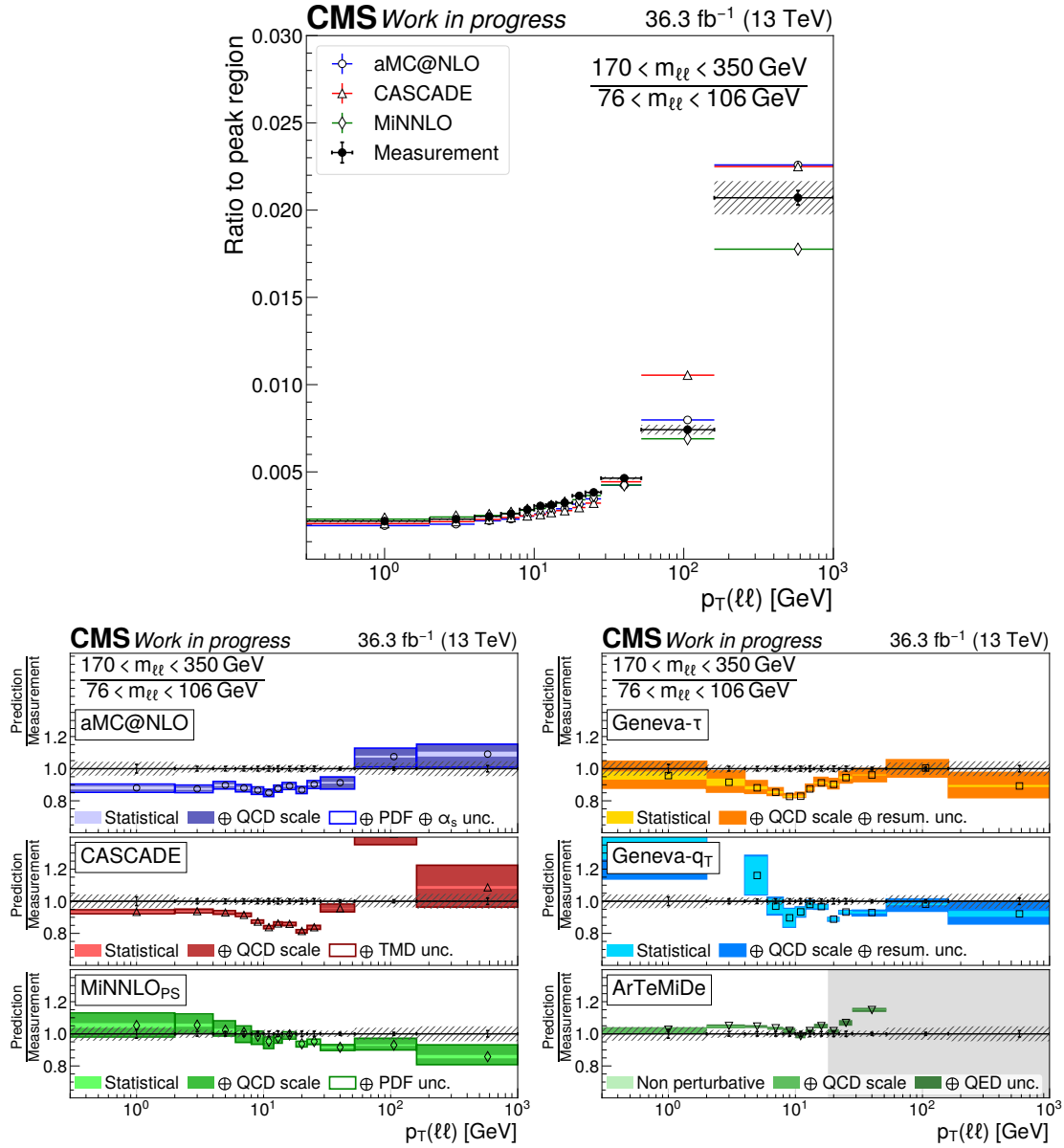


Figure 4.70: Measured cross section ratio as a function of $p_T(\ell\ell)$ for $170 < m_{\ell\ell} < 350$ GeV (top) and comparisons to predictions (bottom).

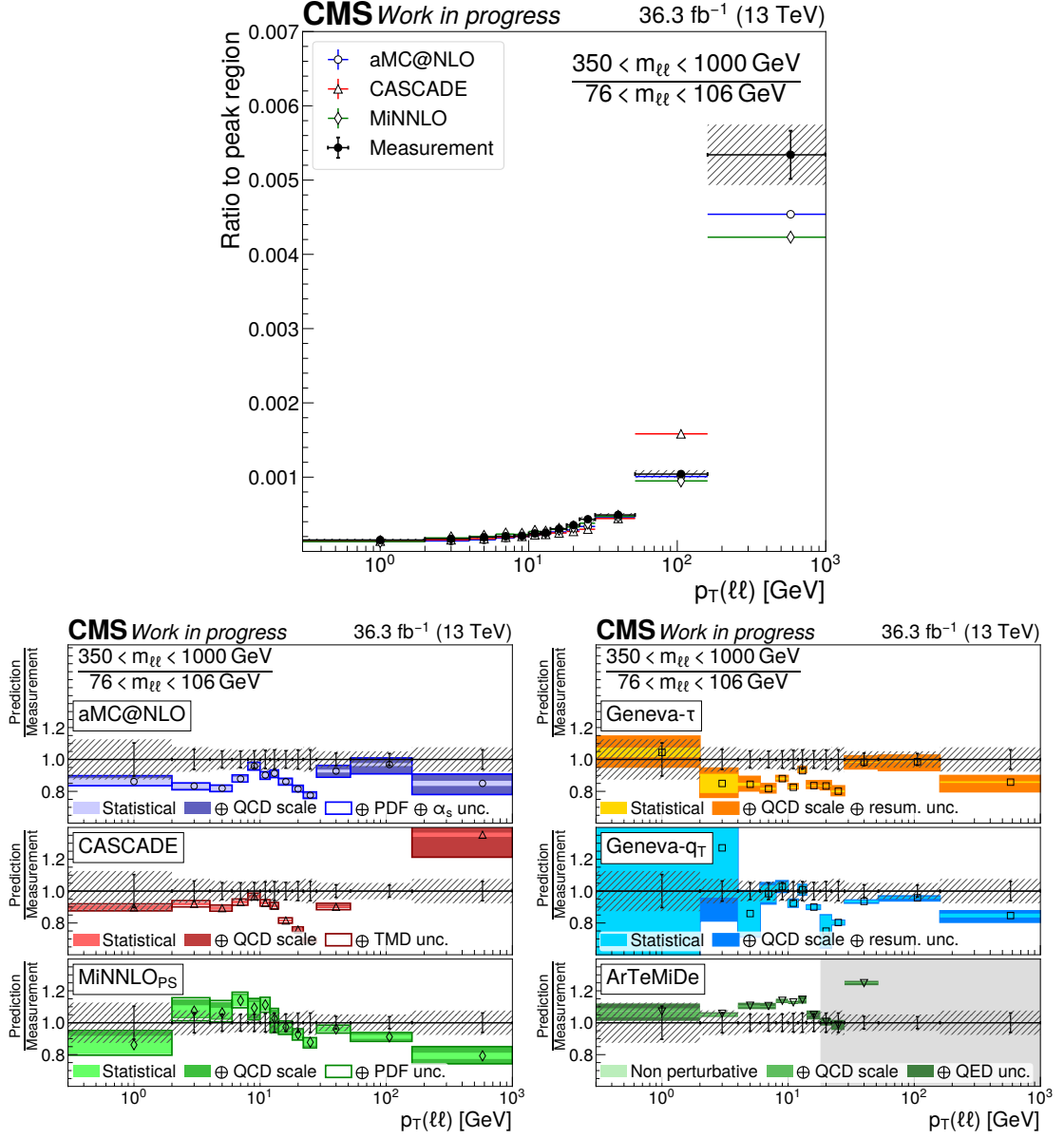


Figure 4.71: Measured cross section ratio as a function of $p_T(\ell\ell)$ for $350 < m_{\ell\ell} < 1000$ GeV (top) and comparisons to predictions (bottom).

including the lower-order CASCADE.

Contrary to the high p_T tail, differences between predictions are clearly visible in the moderate p_T region. The best results are obtained by the predictions at NNLO that explicitly resum p_T : MINNLO_{PS}, GENEVA- q_T , and ARTEMiDE in its domain of validity. They are in agreement in shape and normalization with the data. The reduced scale uncertainty at NNLO is also visible in the intermediate p_T region: it is so small that PDF uncertainties become dominant for MINNLO_{PS} (PDF uncertainties are not available for ARTEMiDE and GENEVA). The second best predictions at moderate p_T are obtained from MADGRAPH and CASCADE, which both rely on a parton shower algorithm for resummation — either standalone or embodied in the PB TMD formalism. They both provide a reasonable prediction in the two lowest mass window, but a significant disagreement appears for masses above the Z peak region. Finally, GENEVA- τ provides by far the worst prediction of the p_T spectrum in the resummation region. Reference [79] shows that a better agreement is obtained in the Z peak region when using a smaller value of $\alpha_s = 0.114$.

As was shown in figure 4.61, migrations in invariant mass caused by QED final-state radiation contribute significantly to the moderate p_T region in the mass window below the Z pole, $50 < m_{\ell\ell} < 76$ GeV. They distort the measured $p_T(\ell\ell)$ spectrum, which causes the unexpected shape observed when comparing it to the peak region in figure 4.68. Except for ARTEMiDE, QED final-state radiation is implemented during the showering step of event generation, using PYTHIA or PHOTOS as described in section 4.11. These corrections appear sufficient to obtain a good agreement with the data. On the other hand, the ARTEMiDE prediction corrected using equation (4.38) underestimates the cross section. It is not clear whether this is caused by the correction procedure or ARTEMiDE itself.

The effect of QED final-state radiation at larger $m_{\ell\ell}$ is smaller, as can be seen by comparing the ARTEMiDE curves with and without QED corrections in figures 4.64 to 4.67. It is, however, significantly larger than the total experimental uncertainty. Predictions not taking QED corrections into account are expected to predict a too large cross section in this region. PYTHIA seems to provide a good enough approximation; this was studied for instance in reference [174].

Values of the transverse momentum under a few GeV are sensitive to non-perturbative contributions including the motion of partons in the colliding protons. In this low p_T region, TMD parton distributions are expected to provide a consistent description of observations. The two predictions that integrate them, CASCADE and ARTEMiDE, indeed predict the correct cross section at small p_T . This is also true of MINNLO_{PS}. The prediction of MADGRAPH is too small in the first few bins, especially at high mass. Finally, GENEVA- τ predicts cross sections more than 30 % too small for $p_T < 3$ GeV. The large statistical uncertainties of

GENEVA- q_T in this region prevent from drawing a conclusion.

4.12.2 φ_η^* variable

The measured φ_η^* -differential cross sections are shown in figures 4.72 to 4.76, and the corresponding cross section ratios in figures 4.77 to 4.80. The plots follow the same layout as for transverse momentum, except that ARTEMiDE is not present. The φ_η^* distributions have the inverted shape of an S, with two plateaus at small and large values and a smooth transition between them. The plateau at small φ_η^* corresponds to the region where non-perturbative effects are important, the transition region to the part of the spectrum requiring resummation, and large φ_η^* to the region where fixed-order predictions are used. This division is very similar to the one used when discussing the p_T distribution, and most of the comments from the previous section also apply to φ_η^* . The main differences are summarized below.

The φ_η^* variable was introduced in reference [81] as an alternative to p_T with better experimental resolution and only slightly worse sensitivity to the modeling of the transverse momentum distribution. However, when comparing the measured distributions of φ_η^* to the distributions of p_T , one finds that the differences with respect to predictions are significantly smaller for φ_η^* . This is not fully compensated by the reduction of the experimental and theoretical uncertainties and the φ_η^* measurements have a smaller discriminant power between predictions than the p_T distributions.

Another effect in the φ_η^* distributions that had not been anticipated when this variable was proposed is the shrinking of the size of the region of non-perturbative effects when large $m_{\ell\ell}$ are considered. This can be seen by comparing the width of the small- φ_η^* plateau between figures 4.72 and 4.76: while it extends up to approximately $\varphi_\eta^* = 2 \times 10^{-2}$ in the lowest mass bin, the plateau ends around $\varphi_\eta^* = 5 \times 10^{-3}$ at high mass. The boundary of the plateau at large φ_η^* , where resummation effects become negligible and fixed-order calculations are sufficient, also moves to smaller φ_η^* values when considering larger $m_{\ell\ell}$ values. This tendency of the φ_η^* distribution to move towards zero when $m_{\ell\ell}$ increases explains the shape of the cross section ratios, opposite to what is seen for p_T — compare e.g. figures 4.69 and 4.78. This dependence of the spectrum in the considered invariant mass window calls for the development of alternative variables explicitly taking the $m_{\ell\ell}$ dependence into account.

One positive point about φ_η^* is that the shape of the φ_η^* distributions is much less affected by QED final state radiation than $p_T(\ell\ell)$, in particular in the mass window just below the Z peak. The main difference between p_T and φ_η^* in this respect can be understood by considering photon radiation off the leptons in the approximation of equation (2.25), here rewritten for a photon and using the angle

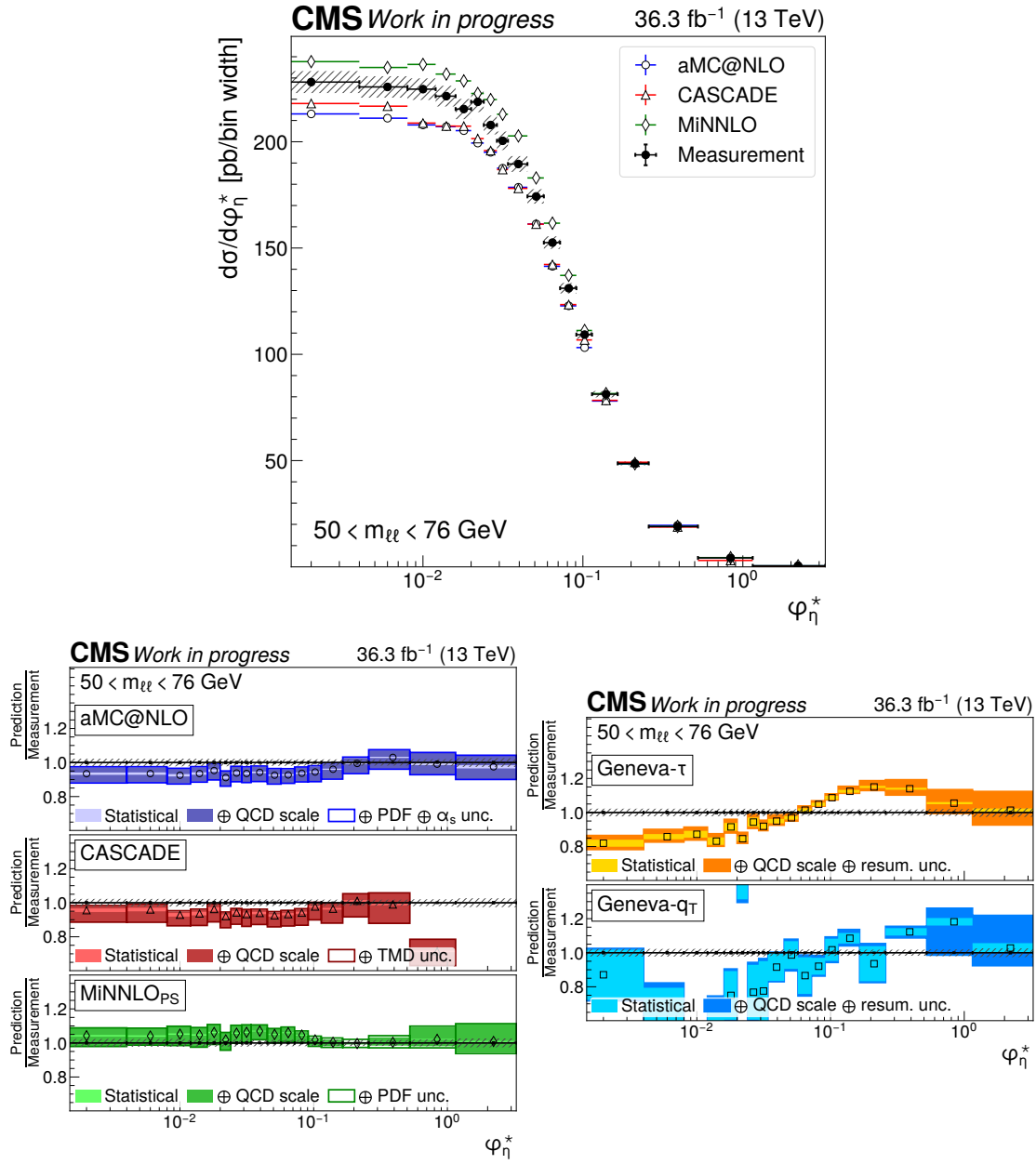


Figure 4.72: Measured cross section as a function of ϕ_η^* for $50 < m_{\ell\ell} < 76$ GeV (top) and comparisons to predictions (bottom).

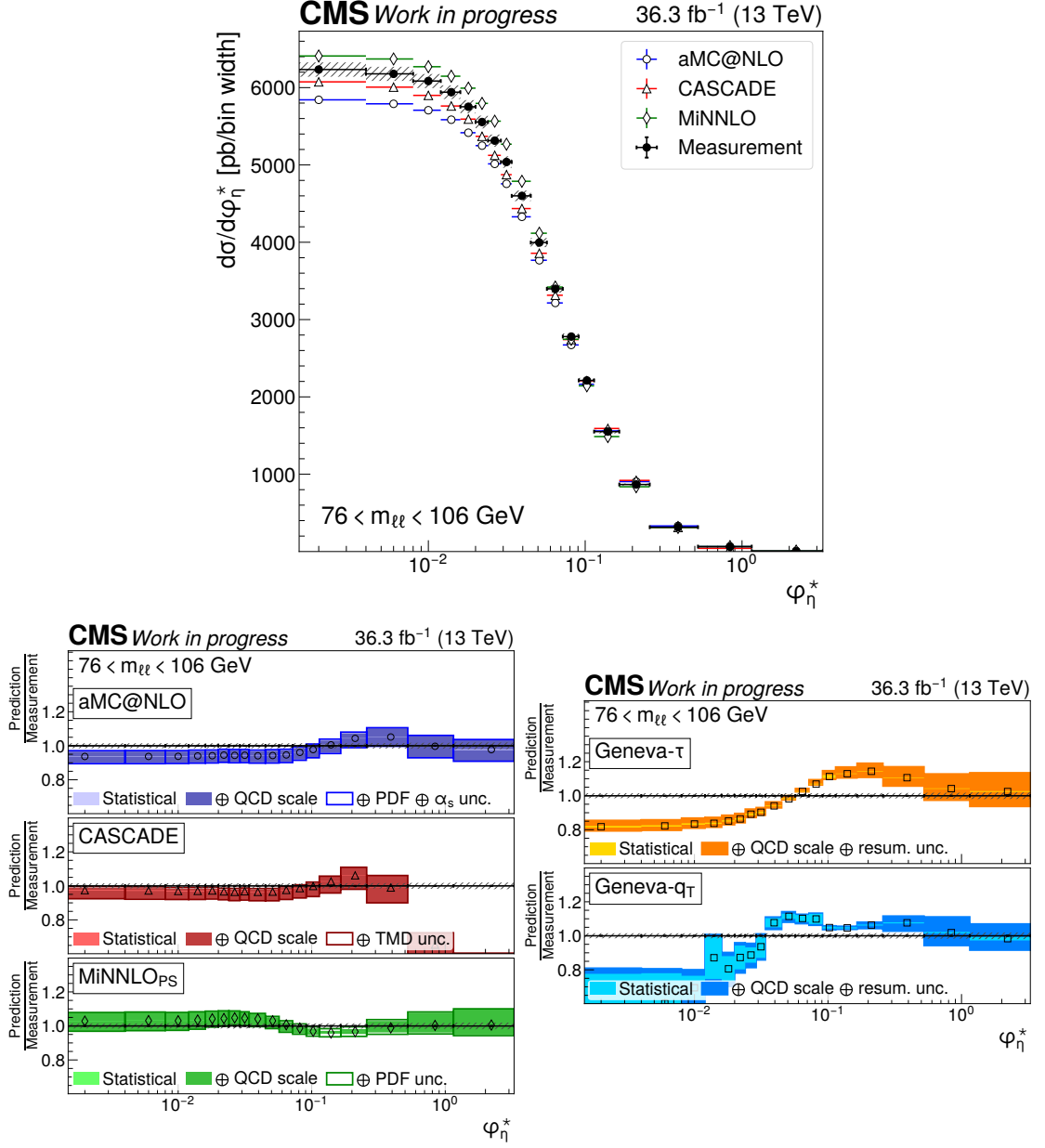


Figure 4.73: Measured cross section as a function of φ_η^* for $76 < m_{\ell\ell} < 106$ GeV (top) and comparisons to predictions (bottom).

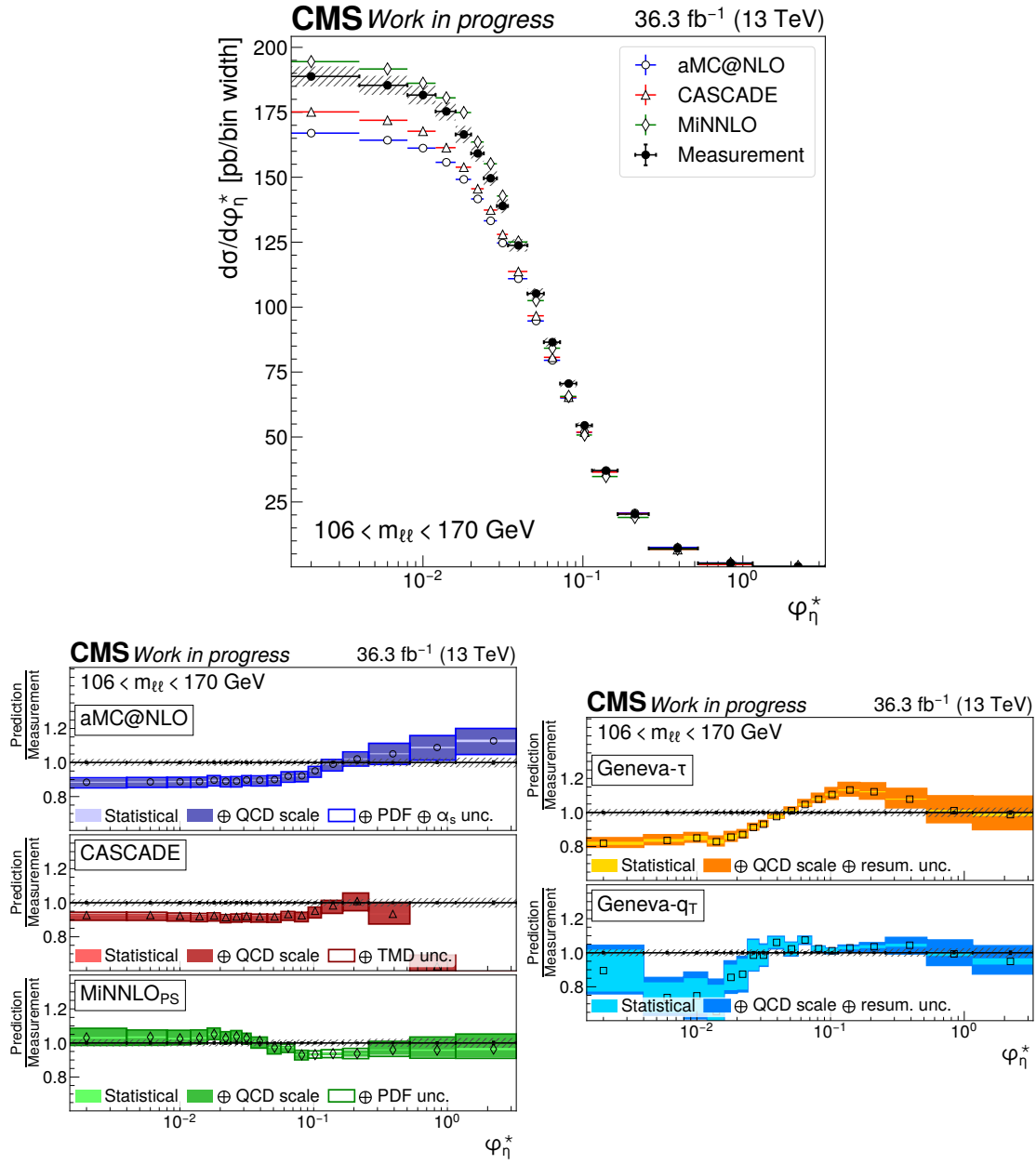


Figure 4.74: Measured cross section as a function of φ_η^* for $106 < m_{\ell\ell} < 170$ GeV (top) and comparisons to predictions (bottom).

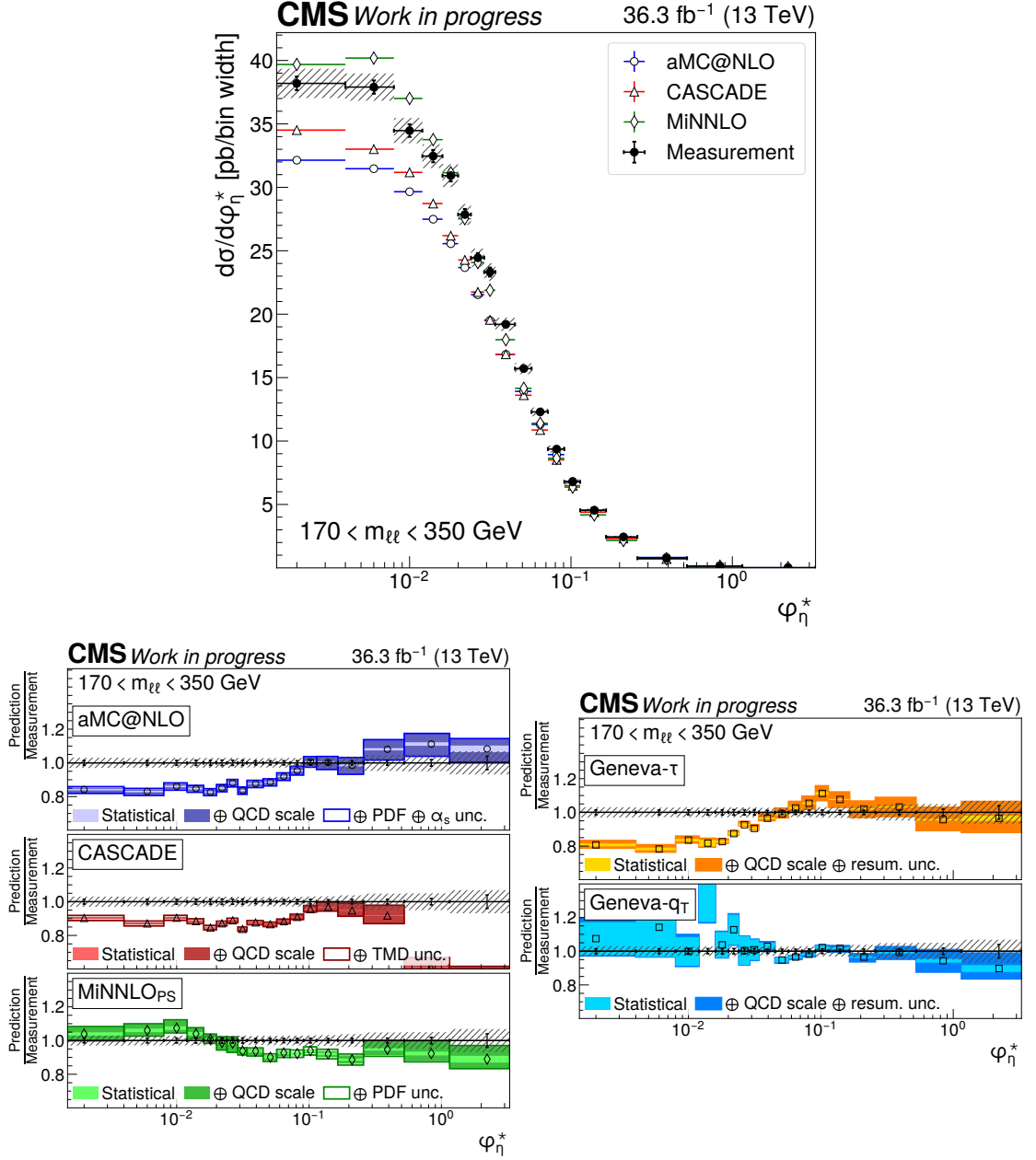


Figure 4.75: Measured cross section as a function of ϕ_η^* for $170 < m_{\ell\ell} < 350$ GeV (top) and comparisons to predictions (bottom).

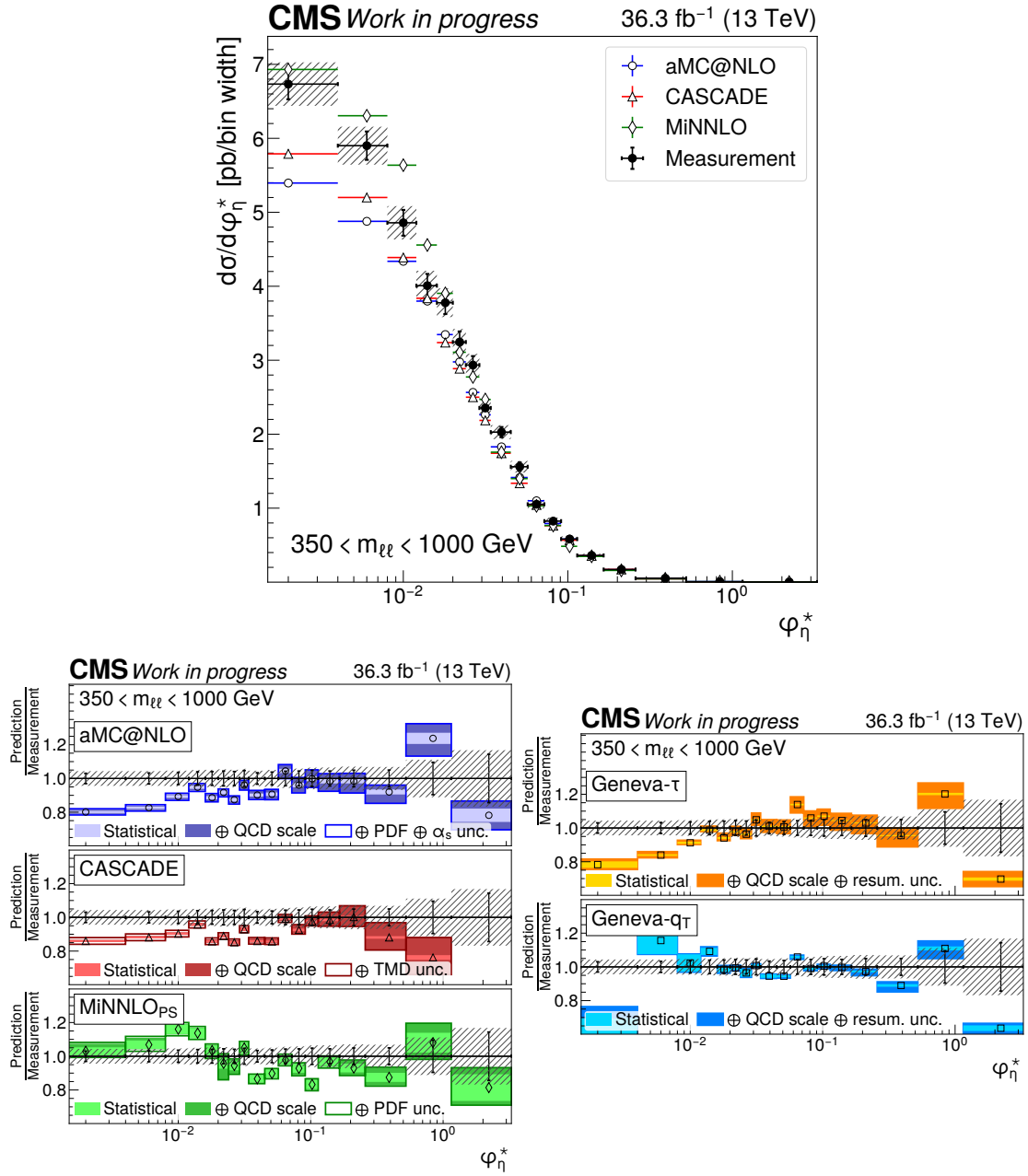


Figure 4.76: Measured cross section as a function of φ_η^* for $350 < m_{\ell\ell} < 1000$ GeV (top) and comparisons to predictions (bottom).

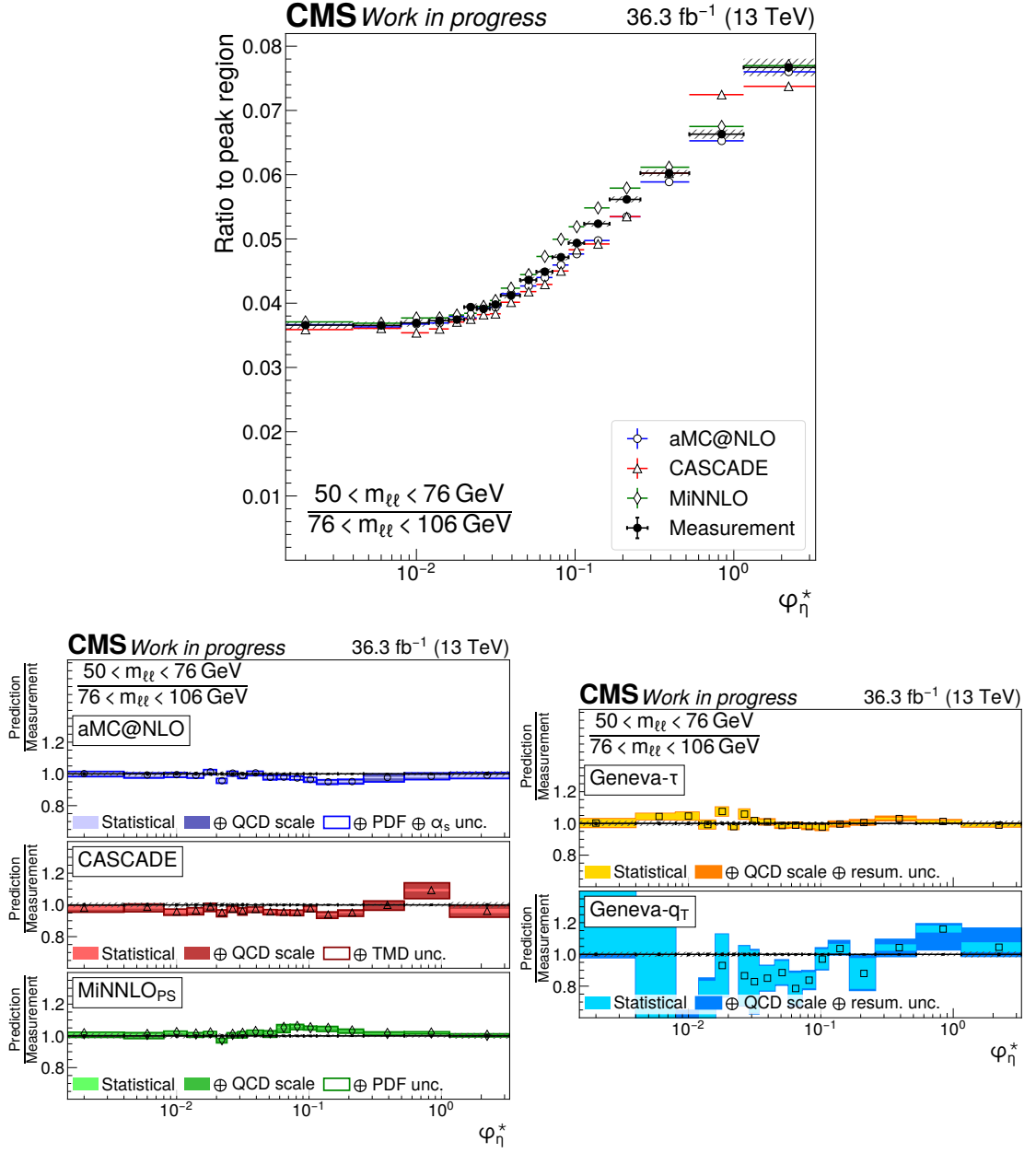


Figure 4.77: Measured cross section ratio as a function of φ_{η}^* for $50 < m_{\ell\ell} < 76 \text{ GeV}$ (top) and comparisons to predictions (bottom).

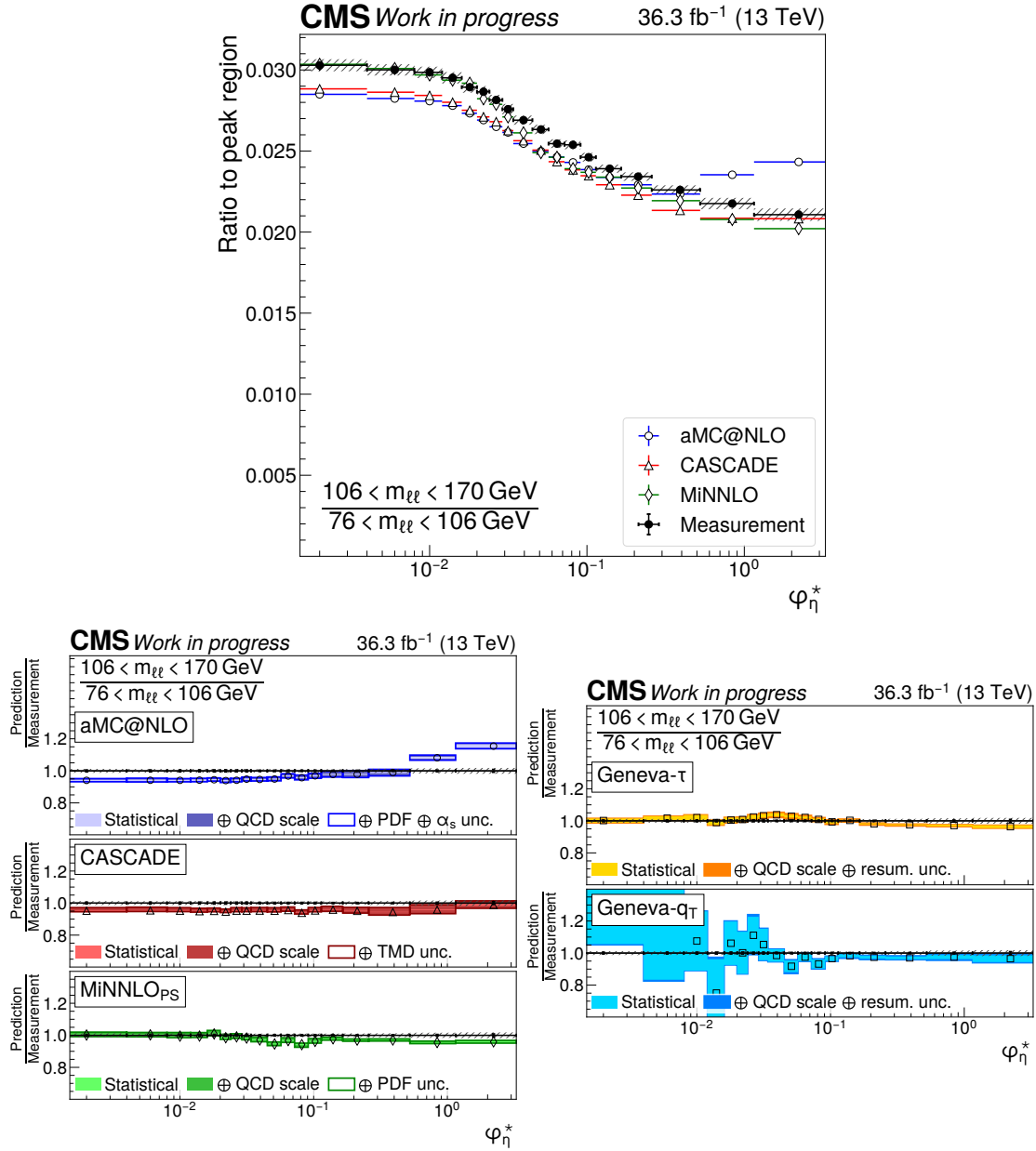


Figure 4.78: Measured cross section ratio as a function of φ_η^* for $106 < m_{\ell\ell} < 170 \text{ GeV}$ (top) and comparisons to predictions (bottom).

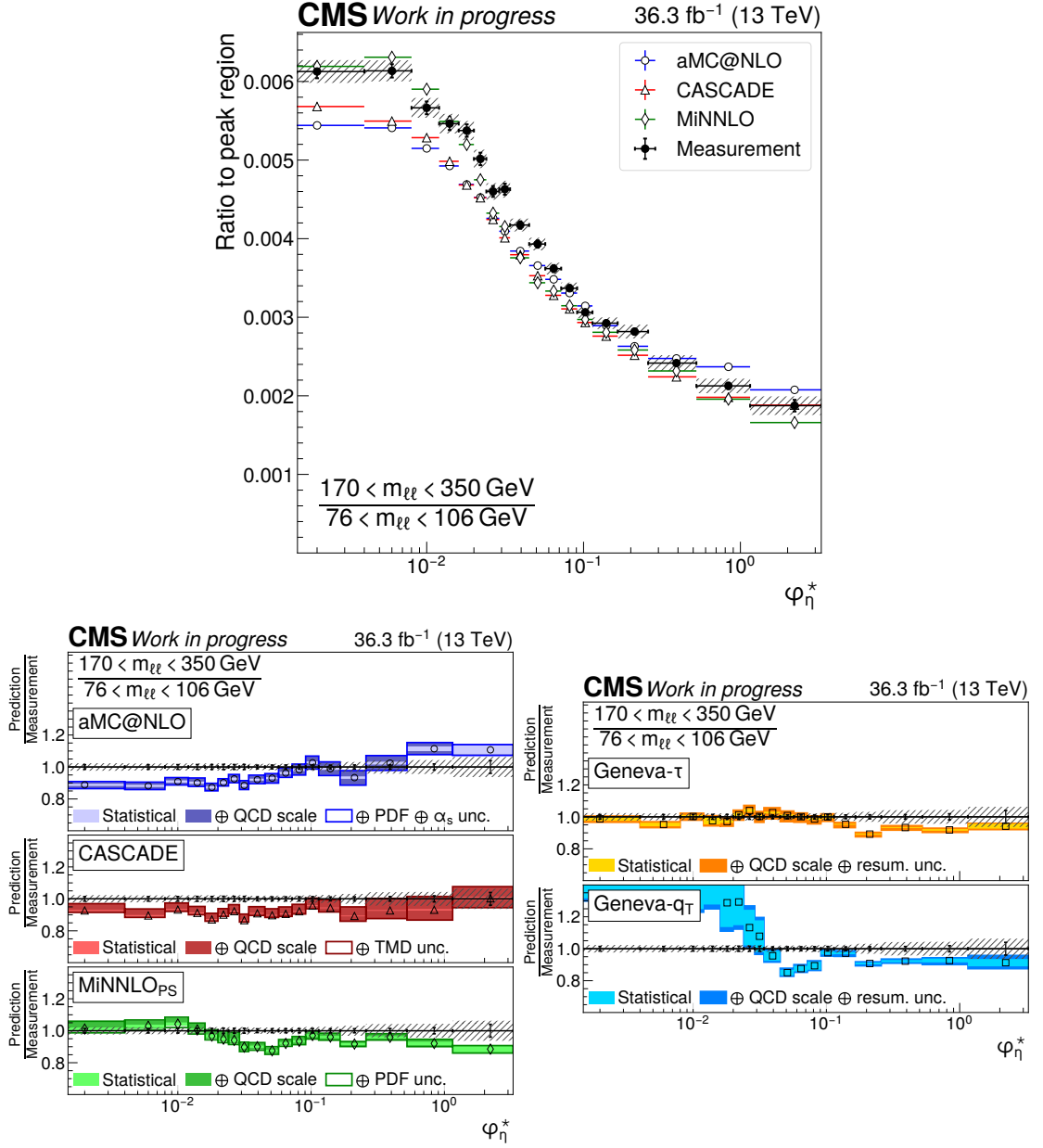


Figure 4.79: Measured cross section ratio as a function of φ_η^* for $170 < m_{\ell\ell} < 350$ GeV (top) and comparisons to predictions (bottom).

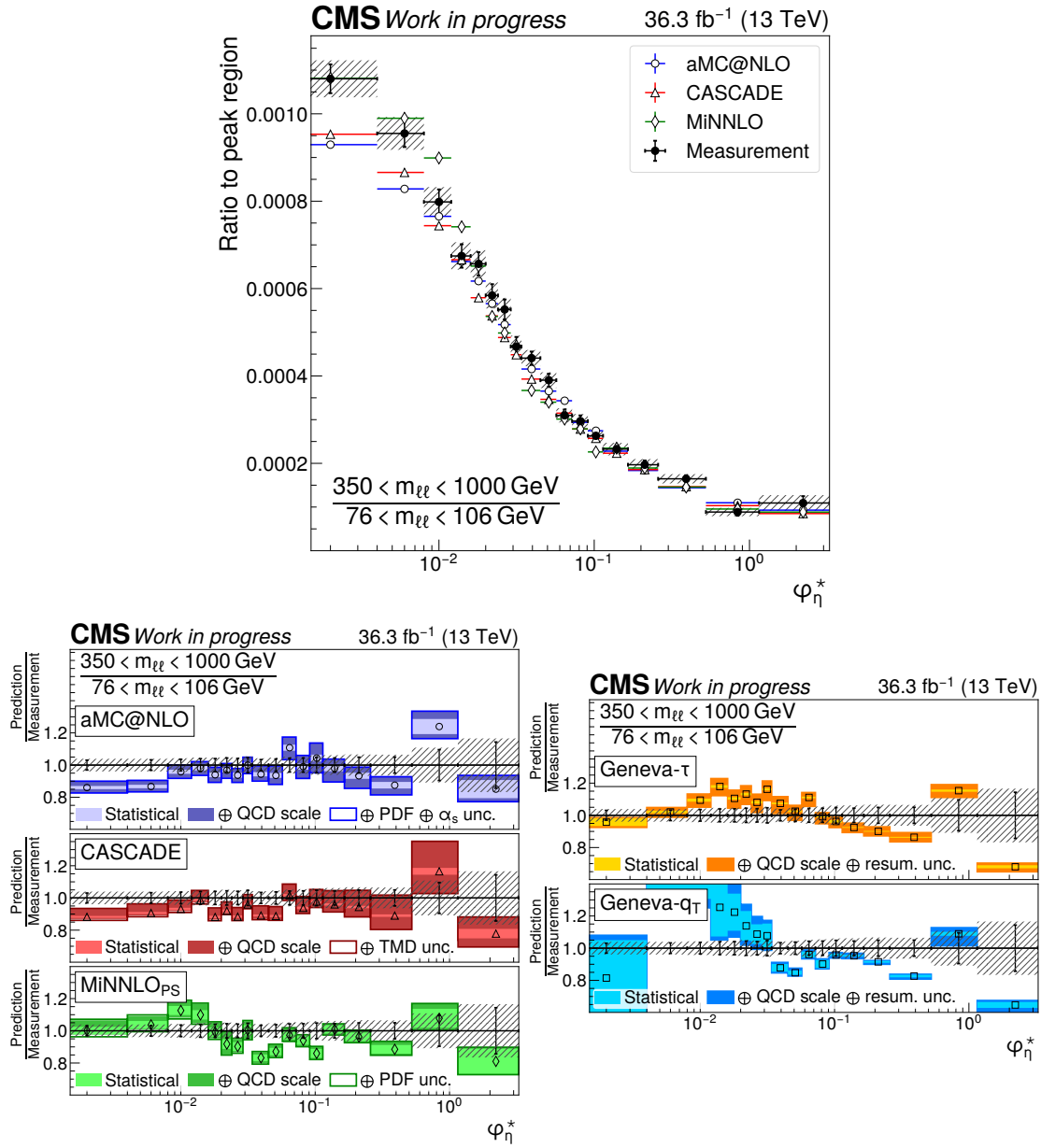


Figure 4.80: Measured cross section ratio as a function of φ_η^* for $350 < m_{\ell\ell} < 1000$ GeV (top) and comparisons to predictions (bottom).

θ between the photon and the lepton:

$$d\sigma_1 \approx \frac{\alpha}{2\pi} \frac{1+z^2}{1-z} dz \frac{d\theta^2}{\theta^2} d\sigma_0. \quad (4.44)$$

Considering that the photon carries away a significant energy, as required for an event to migrate between mass bins, fixes the value of z away from the pole at $z = 1$. The only remaining degree of freedom is the angle θ , which features a collinear pole. This implies that most photons are emitted in a direction nearly collinear to one of the leptons (which is still true after resumming multiple photon emissions). The cones with radius $\Delta R < 0.1$ used in the phase space definition capture some, but not all, of these photons. Those who escape affect the energy of the emitting lepton, but much less its direction because they are still almost collinear. Since φ_η^* only uses angular variables, its value is less sensitive to final-state QED corrections than transverse momentum.

4.12.3 Transverse momentum, one or more jets

The distribution of the lepton pairs in transverse momentum, when at least one jet with $p_T > 30 \text{ GeV}$ is required in the event selection, are shown in figures 4.81 to 4.84. The corresponding ratios are shown in figures 4.85 to 4.87. The cross sections feature a peak at $p_T \simeq 40 \text{ GeV}$, slightly above the jet p_T cut. Three main regions can be distinguished. At small p_T , the cross section is dominated by events with two jets almost balancing each other in the transverse plane. For intermediate p_T , at and around the peak, configurations with a jet recoiling against the lepton pair dominate. In the tail of large p_T , events with multiple jets dominate. Following reference [175], the boundaries between the regions can be set around $p_T(\ell\ell) = 20$ and 70 GeV .

An important observation made in reference [175] is that, at low $p_T(\ell\ell)$, jet distributions receive a large (about 50 %) contribution from double parton scattering (DPS) — two hard parton-parton interactions occurring in the same proton-proton collision. In these events, one of the parton-parton interactions corresponds to the Drell-Yan process and the other to dijet production. The possibility of DPS is not accounted for in the factorization formulas discussed in chapter 2, although dedicated theorems are available [176]. For the predictions used in this analysis except CASCADE, DPS is simulated using the model implemented in PYTHIA (CASCADE does not implement DPS). This dependence on the PYTHIA modeling limits the interpretability of the measurement in this region.

For all predictions, the level of description achieved at large $p_T(\ell\ell)$ is identical to what is observed without requiring a jet, owing to the fact that the cross section at large $p_T(\ell\ell)$ is dominated by events with at least one high- p_T jet. The only exception is CASCADE, for which a sample with one parton generated at NLO

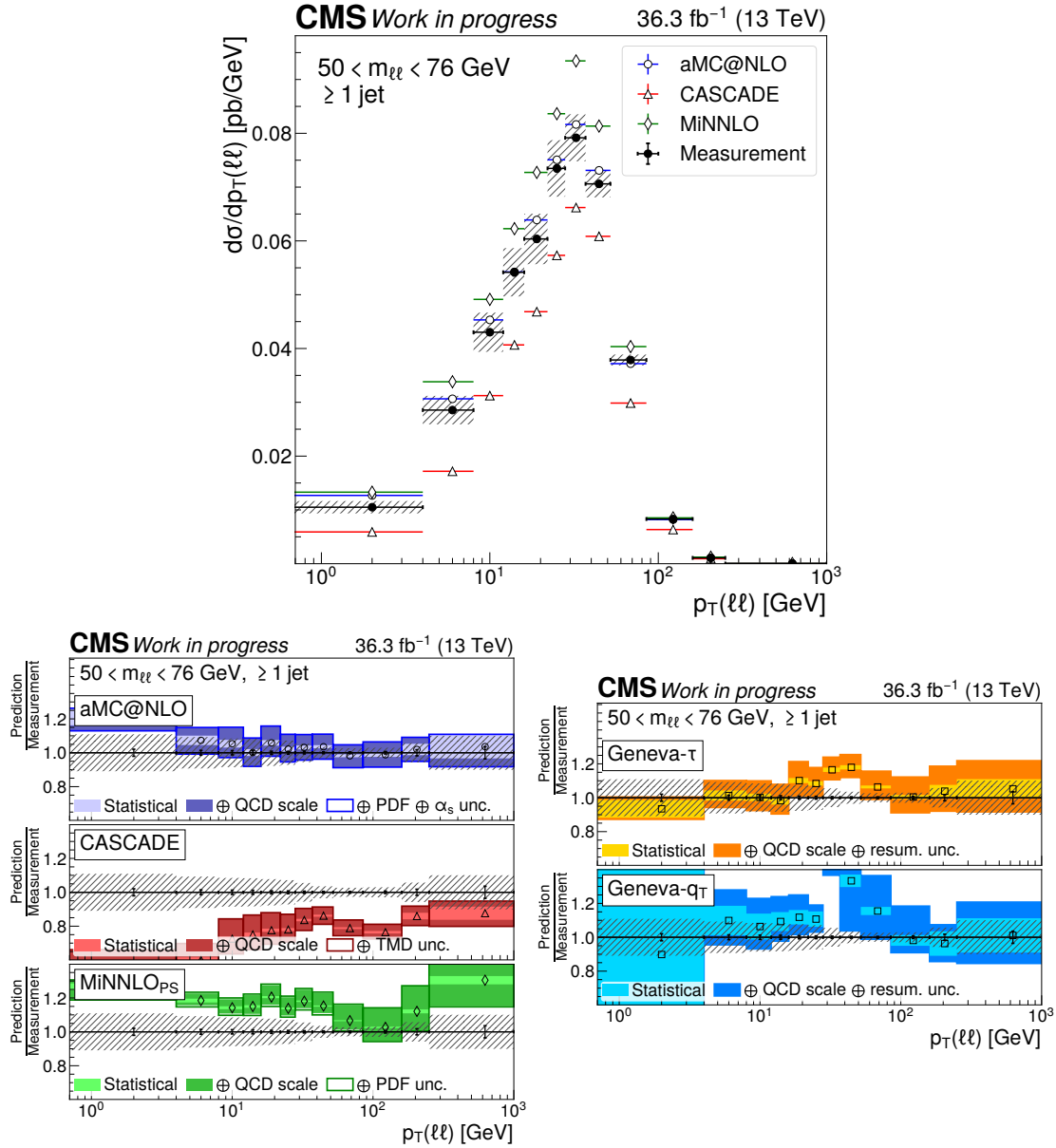


Figure 4.81: Measured cross section as a function of $p_T(\ell\ell)$, with at least one jet, for $50 < m_{\ell\ell} < 76$ GeV (top) and comparisons to predictions (bottom).

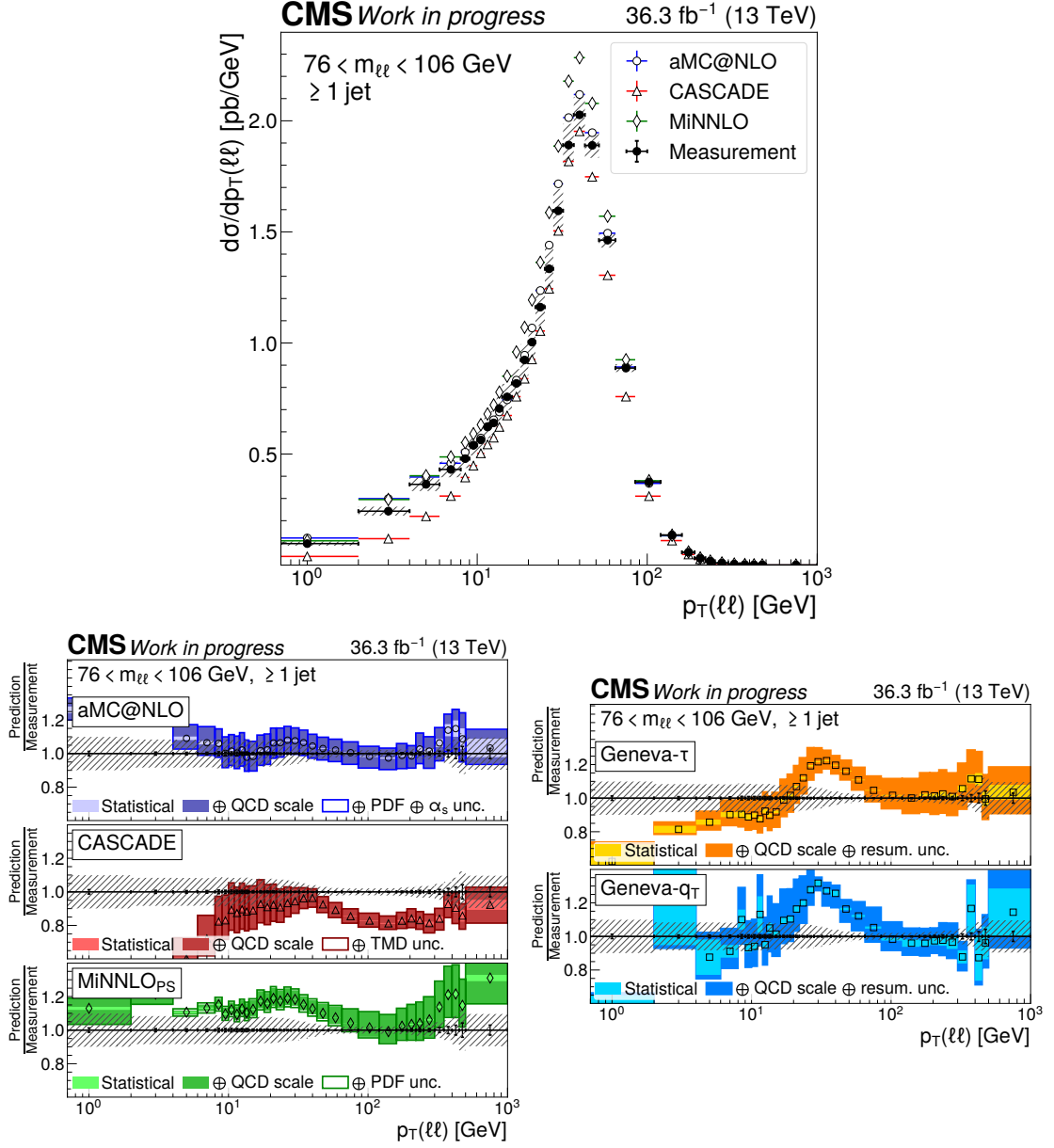


Figure 4.82: Measured cross section as a function of $p_T(\ell\ell)$, with at least one jet, for $76 < m_{\ell\ell} < 106$ GeV (top) and comparisons to predictions (bottom).

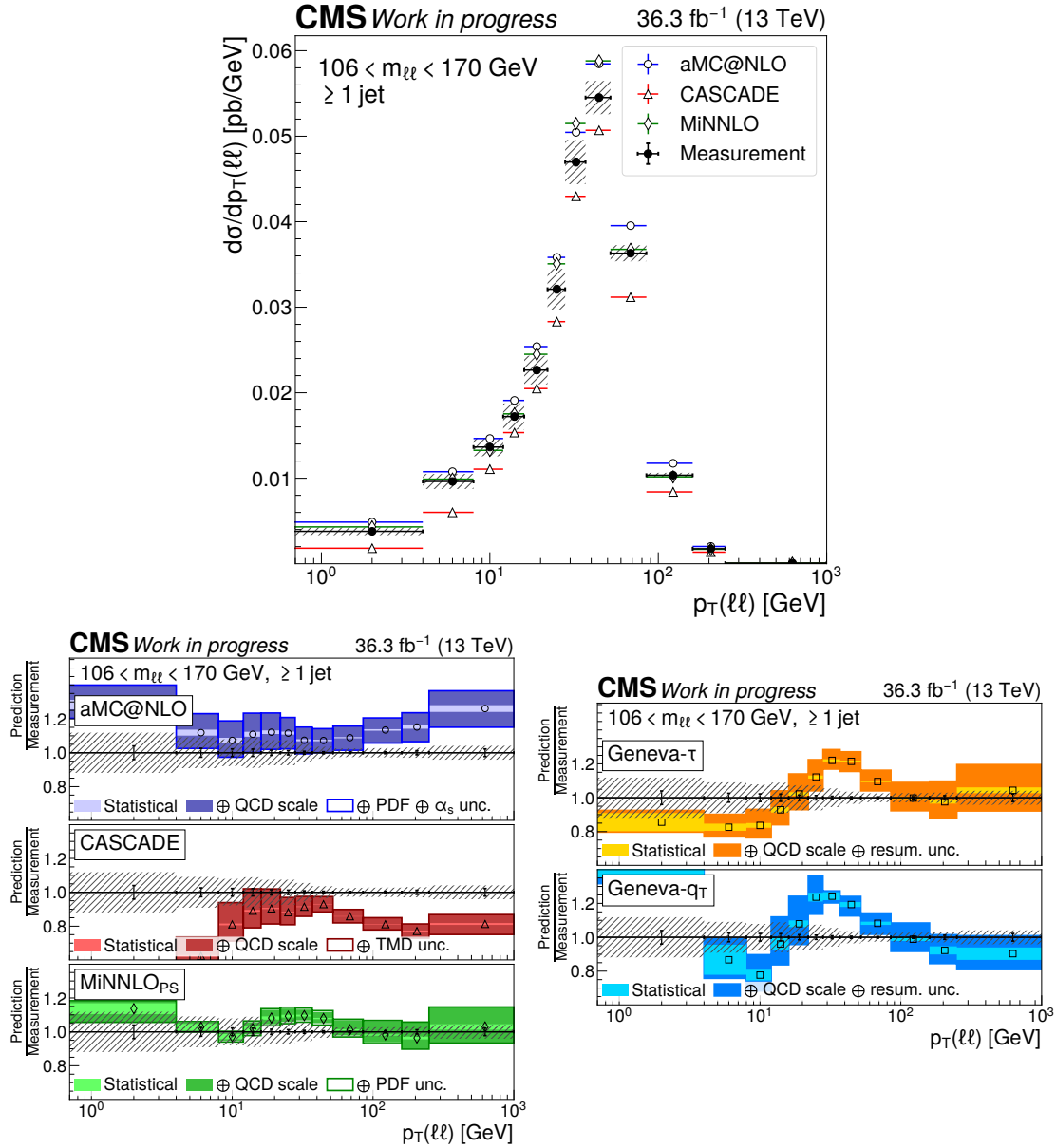


Figure 4.83: Measured cross section as a function of $p_T(\ell\ell)$, with at least one jet, for $106 < m_{\ell\ell} < 170$ GeV (top) and comparisons to predictions (bottom).

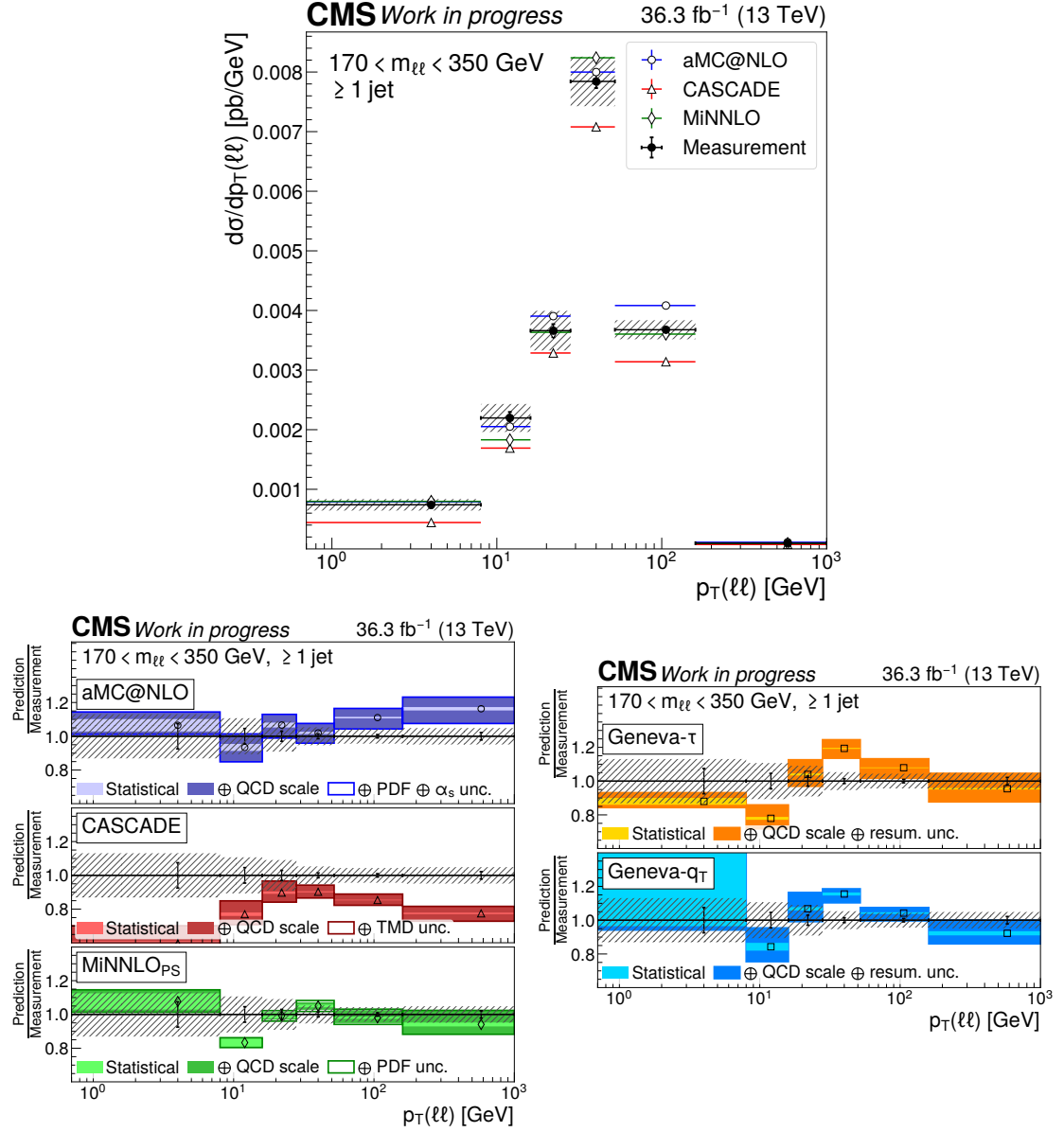


Figure 4.84: Measured cross section as a function of $p_T(\ell\ell)$, with at least one jet, for $170 < m_{\ell\ell} < 350$ GeV (top) and comparisons to predictions (bottom).

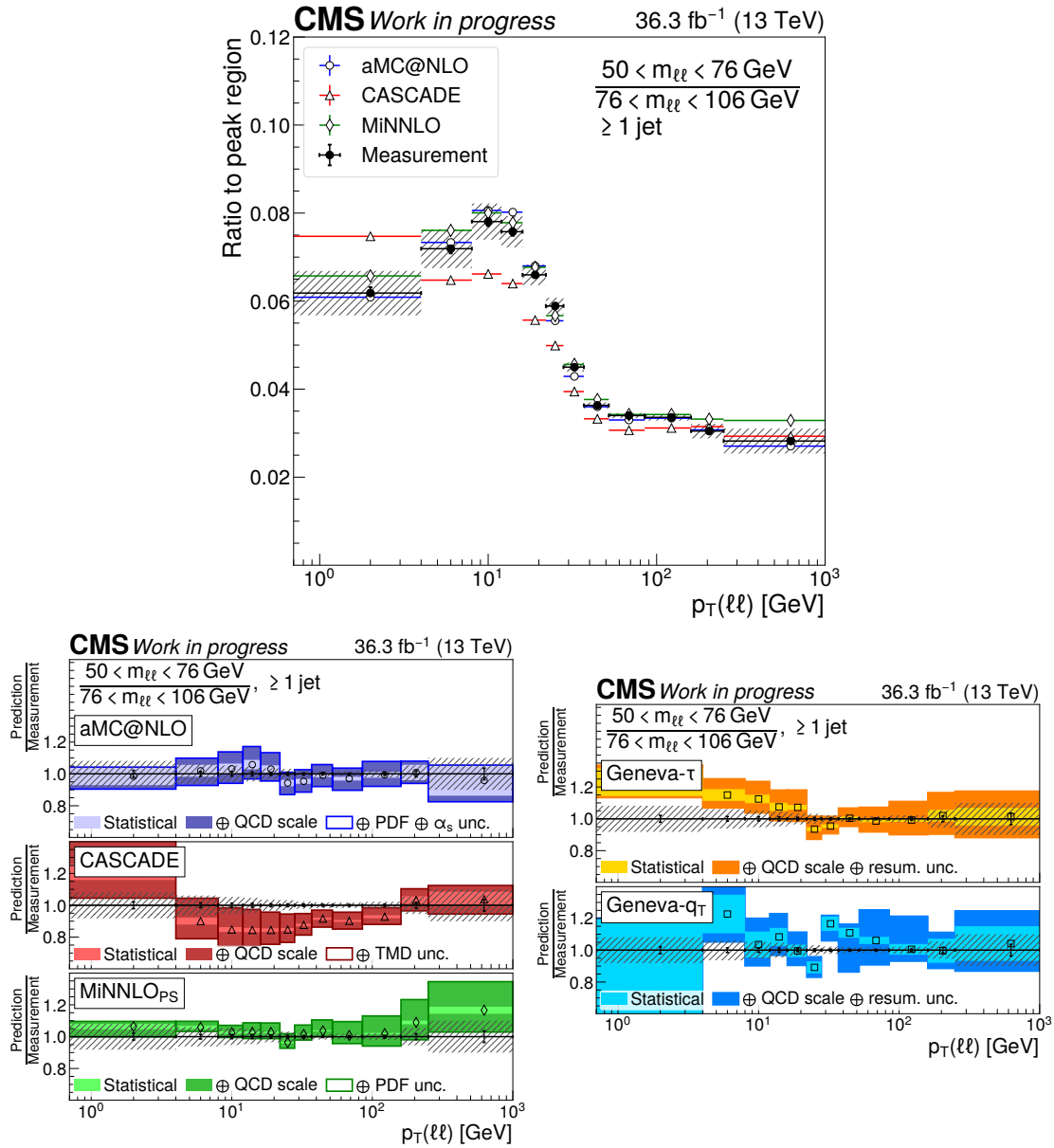


Figure 4.85: Measured cross section ratio as a function of $p_T(\ell\ell)$, with at least one jet, for $50 < m_{\ell\ell} < 76$ GeV (top) and comparisons to predictions (bottom).

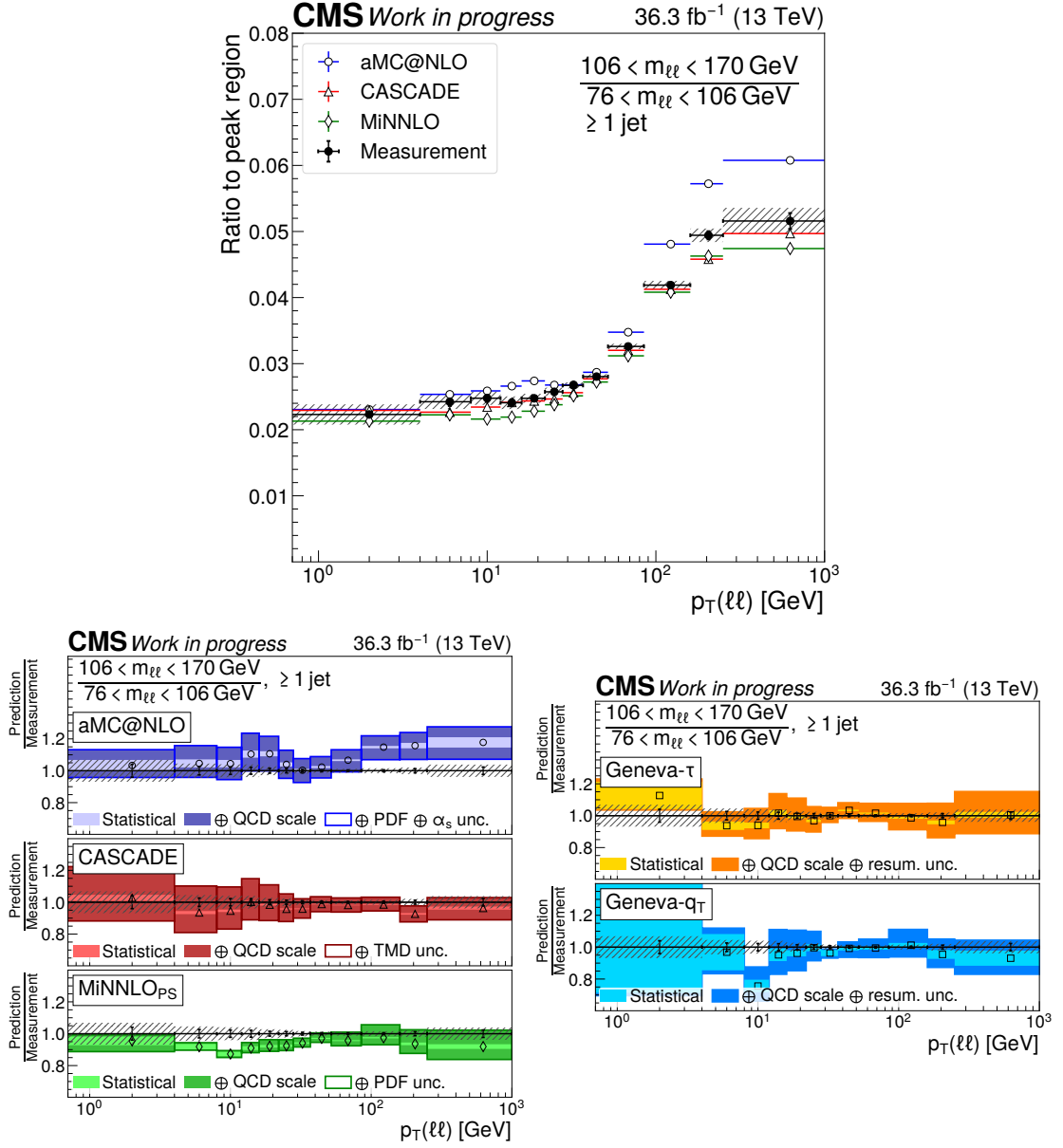


Figure 4.86: Measured cross section ratio as a function of $p_T(\ell\ell)$, with at least one jet, for $106 < m_{\ell\ell} < 170$ GeV (top) and comparisons to predictions (bottom).

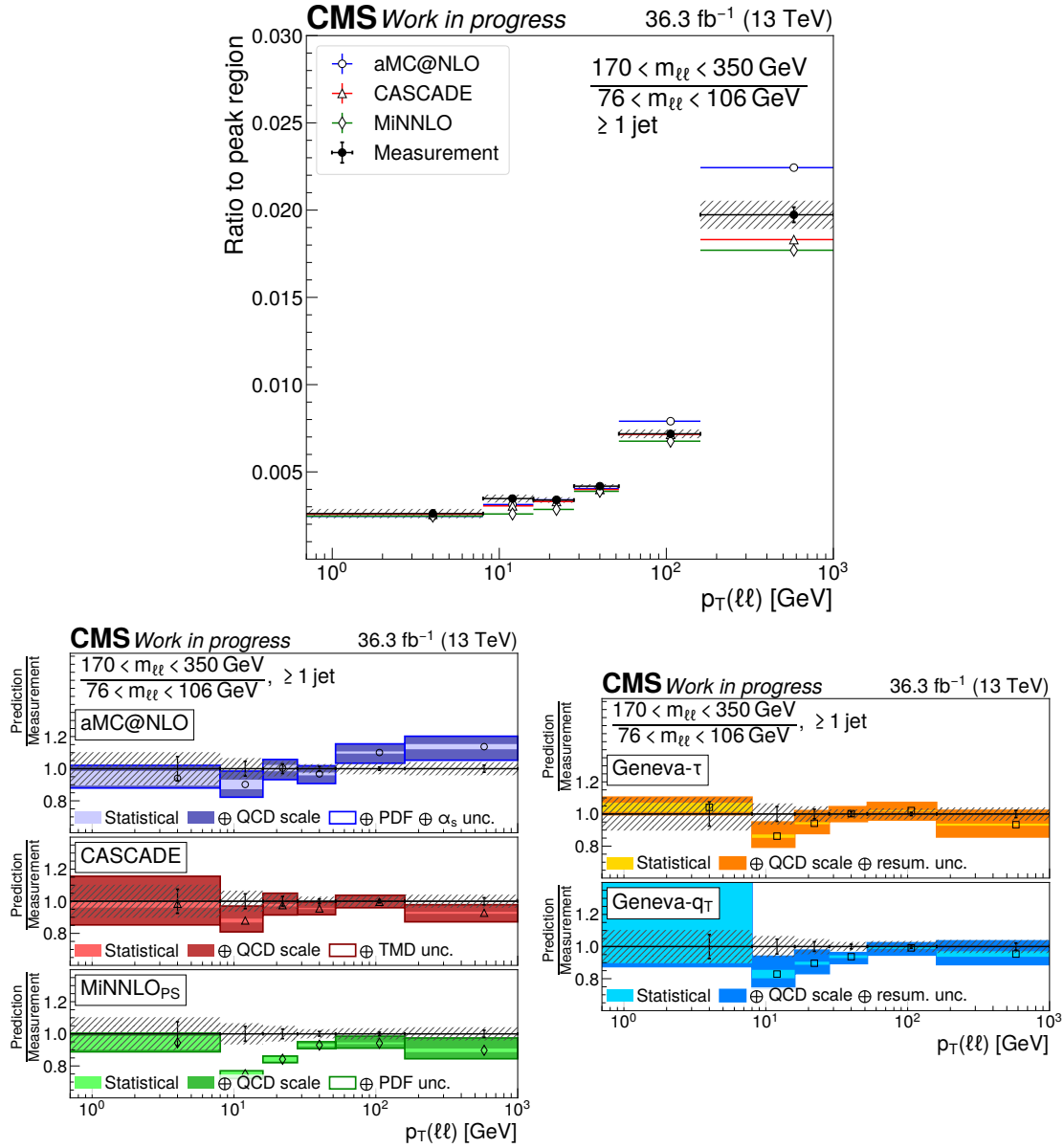


Figure 4.87: Measured cross section ratio as a function of $p_T(\ell\ell)$, with at least one jet, for $170 < m_{\ell\ell} < 350$ GeV (top) and comparisons to predictions (bottom).

is used. Despite this change, the cross section predicted by CASCADE at large $p_T(\ell\ell)$ is too small compared to data, especially for invariant masses above the Z peak region.

More interesting behavior appear in the region of intermediate p_T , where the only prediction to provide a description of the data within the scale uncertainty for all mass windows is AMC@NLO. Except at low mass, CASCADE manages the same achievement. The predictions of MINNLO_{PS} and GENEVA are too large in this region. GENEVA- τ and GENEVA- q_T show the same trend with respect to data, which could be a hint that the cause of this behaviour is identical between the two.

The cross section ratios show that despite their failures at describing the absolute cross sections, all predictions agree with the evolution observed in data. This is the case even in the low- p_T region where DPS is important.

Chapter 5

Conclusion

The Drell-Yan process, or the production of a pair of leptons through the production of a virtual Z/γ^* boson in hadron-hadron collisions, is one of the primary probes of the proton structure. Well-developed theoretical tools are available to predict differential distributions, using first principles where possible and relying on effective models or parameterizations otherwise. The relatively large cross section and easily identified final state contributed to making Drell-Yan one of the primary processes through which QCD effects in the initial state are probed experimentally.

In this thesis, a measurement of differential Drell-Yan cross section is performed, repeated in five regions in invariant mass: $50 < m_{\ell\ell} < 76$, $76 < m_{\ell\ell} < 106$, $106 < m_{\ell\ell} < 170$, $170 < m_{\ell\ell} < 350$, and $350 < m_{\ell\ell} < 1000$ GeV. The dimuon and dielectron channels are measured separately and then combined. In each mass interval, differential cross sections are measured as a function of the transverse momentum $p_T(\ell\ell)$ of the lepton pairs, the φ_η^* variable, and $p_T(\ell\ell)$ with an additional phase space requirement of at least one high- p_T jet. 36.3 fb^{-1} of data corresponding to proton-proton collisions at $\sqrt{s} = 13$ TeV collected in 2016 by the CMS experiment are used. A precision of 2 % for the absolute cross sections at and around the Z peak is achieved, with strong correlations between the bins. Ratios of the distributions with respect to the Z peak region are also measured, for which the total uncertainty reaches 1 % around the Z peak.

The largest uncertainty for the absolute cross sections originates from the luminosity measurement. This precision is achieved thanks to the large statistics available for analysis and a thorough control of the lepton trigger, reconstruction, identification, and isolation efficiencies. For invariant mass regions above and below the Z peak, understanding and reducing the contribution from background processes is also important to reach high precision. The main backgrounds, $t\bar{t}$ and single top production, are suppressed using a veto on b-tagged jets. The photon-induced production of lepton pairs, non-negligible for $p_T(\ell\ell) < 2$ GeV, is

also subtracted based on a Monte-Carlo simulation of this process. In the electron channel, a data-driven method based on events with two electrons reconstructed with the same sign is used to estimate the contribution from misidentified leptons.

At small p_T , the bin width with which $p_T(\ell\ell)$ can be measured is limited by the momentum resolution of the detector and lepton reconstruction algorithms. Unless the resolution is significantly improved, it will remain the limiting factor in the achievable sensitivity to the non-perturbative ingredients of predictions of the p_T spectrum for large scales. The φ_η^* variable, designed to be sensitive to the same effects as p_T , provides less stringent constraints on models than p_T despite smaller experimental uncertainties. Since φ_η^* is not affected by momentum resolution, it may provide more information than p_T at high mass.

The results are compared to six predictions using different techniques to predict the transverse momentum of the lepton pairs. The first prediction uses event samples generated using MADGRAPH5_AMC@NLO with FxFx matching and merging of up to two final state partons to the PYTHIA8 parton shower, in a setup typical of the event samples used by LHC experiments during Run 2. This prediction describes globally well the $p_T(\ell\ell)$ distributions in the low mass and Z peak regions, even if the spectrum is slightly too hard. Above the Z peak, MADGRAPH5_AMC@NLO underestimates the cross section at low p_T while being compatible with the data at high p_T . MADGRAPH5_AMC@NLO also provides a good description of the transverse momentum of the lepton pair when a jet is required, although the underlying event contribution may be overestimated by the CUETP8M1 tune of PYTHIA8.

A second prediction is obtained by matching the parton branching TMD method to a NLO calculation performed by MADGRAPH5_AMC@NLO within the CASCADE framework. The initial-state shower is performed within the TMD and the final-state shower is handled by PYTHIA6. CASCADE provides a better description of p_T spectra at low and moderate p_T than the MADGRAPH5_AMC@NLO samples, especially for masses above the Z peak region. At high p_T , CASCADE underestimates the cross section because of the lack of multiparton merging. CASCADE also lacks a description of multiple parton interactions, which results in a severe underestimation of the cross section at small $p_T(\ell\ell)$ when at least one jet is required.

The measurement is also compared to a Monte-Carlo sample generated in the MINNLO_{PS} approach, providing a prediction of the total cross section at NNLO and matched to the PYTHIA8 parton shower. MINNLO_{PS} provides a good description of the inclusive $p_T(\ell\ell)$ distributions, although localized disagreements are visible at high p_T and at high mass. However, when requiring at least one jet, MINNLO_{PS} overestimates the cross section by nearly 20 % at and under the Z peak.

Two predictions of the GENEVA the Monte-Carlo generator are also considered, using a calculation of the cross section at NNLO with N -jettiness or transverse momentum resummed at NNLL and N^3 LL, respectively. The prediction using N -jettiness resummation is found not to provide a good description of the measured p_T spectrum in any mass window. Switching to using p_T resummation gives a faithful description of the inclusive p_T distribution, but discrepancies persist at low p_T — maybe caused by the matching to the parton shower. This indicates that resummation of a single variable is not sufficient to predict the fully differential Drell-Yan cross section; joint resummation of τ and p_T might improve the situation [177]. When at least one jet is required in the event selection, both Geneva- τ and Geneva- q_T give similar results.

The last prediction uses the ARTEMIDE computer code to calculate the transverse momentum distribution of lepton pairs in the framework of small- q_T factorization. Additional corrections for QED final-state radiation are applied. The ARTEMIDE prediction is based on a fit to semi-inclusive scattering, deep inelastic scattering, and Drell-Yan data. Within the range of validity of small- q_T factorization, $p_T \lesssim 0.2 m_{\ell\ell}$, the extrapolation of this fit to $\sqrt{s} = 13$ TeV is in agreement with the measurement, including for invariant masses not considered in the fit (above 170 GeV). The only exception to this generally good agreement is found in the mass bin between 50 and 76 GeV, which is subject to large QED corrections. In order to test the applicability and universality of the TMD formalism, it will be important to use TMDs fits based on Drell-Yan in prediction of other processes, for instance Higgs or dijet production, and compare them to measurements.

An unanticipated theoretical difficulty in predicting the p_T spectrum of Drell-Yan lepton pairs out of the Z peak region is the importance of QED final-state radiation. The associated correction is especially important for masses smaller than m_Z where migrations from the Z peak become important. Parton shower codes such as PYTHIA8 that implement photon emission in the Z decay at leading order provide a good description of the related effects, but this would need confirmation in a dedicated experimental effort using the large number of Z boson decays recorded at the LHC. Such a study is also important to validate the implementation of NLO electroweak corrections available in MADGRAPH5_AMC@NLO since version 3.0 [156], especially once matching to parton showers becomes available.

The comparisons performed in this thesis show that some classes of predictions give more accurate results in certain regions of the phase space. In particular, predictions that use TMDs (ARTEMIDE, CASCADE) provide a better description of the cross section at small p_T than models based on parton showers. The intermediate region is best predicted using explicit resummation in p_T , as in GENEVA- q_T . At large p_T , the best accuracy is obtained with the MADGRAPH5_AMC@NLO

sample that includes events with two final-state partons at NLO. This prediction also provides the best description of the cross section when one jet is required.

To illustrate the strengths of the different approaches, one can imagine a calculation that would combine them to obtain the best possible prediction of the distributions measured in this thesis. It could look as follows:

- The total cross section would be calculated at NNLO to reduce the scale uncertainties where possible.
- In order to improve the prediction of jet distributions, a merging technique would be used to obtain NLO accuracy for the second parton.
- Explicit resummation of p_T would be performed to NNLL or better.
- TMD parton distribution functions would be used for the non-perturbative part of resummation.
- Final-state radiation, hadronization and multiple parton interactions would be described by a parton shower code. QED FSR in the Z/γ^* decay would be included at least at LO.

While all the ingredients listed above are available independently, it is currently not possible to combine all of them in a single Monte-Carlo prediction. For instance, the parton showers matched to currently available NNLO-accurate Monte-Carlo generators need to be vetoed and truncated; the implementation of this feature in TMD PDFs determined *a priori* seems conceptually difficult.

By measuring transverse momentum spectra at invariant masses significantly larger than the mass of the Z , the analysis presented in this thesis provides constraints on the evolution of the models with respect to the mass. Since none of the considered predictions provides a complete description of the data, additional phenomenological studies will be needed to understand their weaknesses and eventually improve their accuracy.

The analysis presented in this thesis is based on data collected in 2016. Meanwhile, the Run 2 of the LHC was completed, during which CMS recorded physics data corresponding to an integrated luminosity of 137 fb^{-1} [69, 178, 179], four times the amount collected in 2016. These new data will allow the use of smaller mass windows and finer binnings in the high- p_T tails of the distributions. The increased statistics will also permit the measurement of triple-differential cross sections in p_T , rapidity, and $m_{\ell\ell}$. Different choices of binning in multiple dimensions incur different balances between sensitivity in x (mass and rapidity) and k_T (p_T). With its use of a coarse binning in $m_{\ell\ell}$ and a fine binning in p_T , this thesis is primarily

focused on k_T sensitivity; this choice could be revisited, for instance to investigate the correlation between x and k_T in TMDs at constant scale.

A significant reduction of the binning in invariant mass would increase the dependence in the lepton energy scale, especially around the Z peak where the slope in $m_{\ell\ell}$ is large. This would result in larger experimental uncertainties, although with strong anti-correlations between $m_{\ell\ell}$ bins. Migrations between mass windows would also require the use of two-dimensional unfolding, complicating the analysis procedure. A categorization in the Z rapidity would also result in increased uncertainties because of the generally worse status of the detector at large $|\eta_\ell|$ with respect to efficiency, resolution, pre-firing, etc. Since the dependence of detector effects with respect to pseudorapidity is usually better known than with respect to energy, using a fine binning in $|y_Z|$ within rather coarse mass windows may be a better option from the experimental point of view.

In addition to fully reconstructed data, CMS records partially reconstructed events passing looser selections. These data should potentially allow the use of lower p_T thresholds on the leptons and, consequently, a measurement at smaller invariant mass and Björken x . Obtaining a thorough control of backgrounds and experimental uncertainties will, however, not be an easy task.

Initial comparisons between measurements by ATLAS and CMS at the Z pole have proven to be difficult because of differing phase space definitions, inconsistent treatment of the backgrounds, and incompatible binnings. Coordination between the experiments is needed if the possibility to perform a future LHC-wide combination, accurately treating correlated uncertainties, is desired. Such a combination would for instance be beneficial for TMD fits, where it could be used as a single data source for the two measurements.

Appendix A

Statistical framework

The data analysis presented in this thesis is implicitly based on the theory of estimation in statistics. This appendix explains the theory behind the treatment of uncertainties, taking the measurement of a differential cross section in collision data as an example. In this appendix, the main properties of the uncertainties in such a measurement are explained: statistical and systematic uncertainties, the covariance and correlations matrices and their interpretation in terms of the variability of the results. A general discussion of estimation theory can be found in chapters 2 and 7 of reference [180].

For the statistician, a measurement corresponds to the estimation of a parameter θ from the observed data $\{X_i\}$. This is performed using a function $\hat{\theta}$ of the data called an estimator. In the case of a cross section measurement, the estimated quantity is the cross section σ and a possible estimator is given by¹:

$$\hat{\sigma} = \frac{N}{\mathcal{L}} = \frac{1}{\mathcal{L}} \sum_i \Theta(X_i), \quad (\text{A.1})$$

where N is the number of events passing the selection and $\mathcal{L}$ is the luminosity. The corrections for acceptance and efficiency have been omitted for simplicity. In the second form, the calculation of N has been made explicit using an auxiliary function Θ that evaluates to one for events that pass the selection and to zero otherwise. This expression makes it clear that $\hat{\sigma}$ is in fact an average, with the luminosity playing the role of the sample size.

Useful estimators $\hat{\theta}$ are random variables because they use the randomly sampled observations $\{X_i\}$. In the limit of a large number of observed events, $\mathcal{L} \rightarrow \infty$, the expectation of the estimator from equation (A.1) is equal to the true value σ_0 of the cross section: $E[\hat{\sigma}] = \sigma_0$. Asymptotically, the distribution of $\hat{\sigma}$ tends to a

¹In this appendix, $\hat{\sigma}$ stands for the statistical estimator used to measure the cross section and not the partonic cross section as in chapter 2.

normal law with variance:

$$V[\hat{\sigma}] = E[(\hat{\sigma} - \sigma_0)^2] = \frac{N}{\mathcal{L}^2}. \quad (\text{A.2})$$

This expression gives the intrinsic statistical uncertainty of the measurement. It is often given for a counting experiment, where the standard deviation of the observed number of events is $\Delta N \equiv \sqrt{V[N]} = \sqrt{N}$.

A.1 Properties of the variance

The variance has the following properties:

1. The standard deviation ΔX of a random variable X , usually used when quoting uncertainties, is defined as the square root of its variance:

$$\Delta X \equiv \sqrt{V[X]}. \quad (\text{A.3})$$

2. The variance of a sum of two independent random variables X and Y is the sum of the variances:

$$V[X + Y] = V[X] + V[Y]. \quad (\text{A.4})$$

3. When a random variable X is multiplied by a constant a , the variance is scaled by a^2 :

$$V[aX] = a^2 V[X]. \quad (\text{A.5})$$

Moreover, if X is a normally distributed variable with mean μ and $Y = f(X)$ is calculated through the application of a sufficiently smooth function f , the expectation and variance of Y can be approximated as follows:

$$E[Y] \approx f(\mu), \quad (\text{A.6})$$

$$V[Y] \approx f'(\mu)^2 V[X]. \quad (\text{A.7})$$

These expressions generalize as follows when f is a function of several independent variables, say X_1 and X_2 :

$$E[Y] \approx f(\mu_1, \mu_2), \quad (\text{A.8})$$

$$V[Y] \approx \left[\frac{\partial f(\mu_1, \mu_2)}{\partial X_1} \right]^2 V[X_1] + \left[\frac{\partial f(\mu_1, \mu_2)}{\partial X_2} \right]^2 V[X_2]. \quad (\text{A.9})$$

In general, a variable Y calculated in this way does not follow a normal distribution. However, when the standard deviation ΔX of X is small with respect to the non-linearity of f , it is often a good approximation. This property is the basis of “error propagation” techniques as used in this thesis.

A.2 Weighted events

In a more realistic analysis, an event-by-event correction may be applied to account for the dependence of the detection efficiency $\varepsilon(X_i)$ in the event kinematics. The estimator $\hat{\sigma}$ is modified as follows:

$$\hat{\sigma} = \frac{1}{\mathcal{L}} \sum_i \frac{1}{\varepsilon(X_i)} \Theta(X_i). \quad (\text{A.10})$$

This amounts to considering each event with a weight $1/\varepsilon(X_i)$. The variance of this new estimator can be found by applying equations (A.4) and (A.5):

$$V[\hat{\sigma}] = \frac{1}{\mathcal{L}^2} \sum_i \left[\frac{1}{\varepsilon(X_i)} \right]^2 \Theta(X_i). \quad (\text{A.11})$$

The apparition of the square of the weight makes weighted events particularly bad for the statistical uncertainty. This can be seen by considering a sample of 100 observed events, 99 of which have weight 1 and the last one has weight 10. The variance obtained from this sample is 199 and the relative uncertainty is:

$$\frac{\Delta\sigma}{\sigma} = \frac{\sqrt{199}}{109} = 13\%. \quad (\text{A.12})$$

This is the same as the statistical uncertainty obtained from an unweighted sample with only 60 events.

A.3 Systematic uncertainties

In a real measurement, the luminosity and the efficiency ε are obtained from measurements and are thus also random variables that follow some probabilistic distribution. Because they are not of interest for the analysis but nevertheless affect the precision with which it can be measured, they are called *nuisance parameters*.

Assuming that their distribution is sufficiently close to a Gaussian, the central values $\mathcal{L}_c$ and ε_c of the nuisance parameters can be used — following equation (A.8) — to estimate the cross section. With these modifications, the estimator $\hat{\sigma}$ becomes:

$$\hat{\sigma} = \frac{1}{\mathcal{L}_c} \sum_i \frac{1}{\varepsilon_c(X_i)} \Theta(X_i). \quad (\text{A.13})$$

Under the assumption that the nuisance parameters are independent, the variance of the estimate provided by $\hat{\sigma}$ can be calculated using equations (A.9) and (A.11):

$$V[\hat{\sigma}] = \frac{1}{\mathcal{L}_c^2} \sum_i \left[\frac{1}{\varepsilon_c(X_i)} \right]^2 \Theta(X_i) + V[\mathcal{L}] \left[\frac{\partial \hat{\sigma}}{\partial \mathcal{L}_c} \right]^2 + V[\varepsilon] \left[\frac{\partial \hat{\sigma}}{\partial \varepsilon_c} \right]^2, \quad (\text{A.14})$$

where it is implicit that partial derivatives are evaluated at the central value.

The partial derivatives in equation (A.14) are often troublesome to calculate analytically. Instead, they are usually evaluated numerically by varying each nuisance parameter by its standard deviation. This corresponds to using the method of finite differences, but the choice of the standard deviation has an important advantage: writing $\Delta_\varepsilon \hat{\sigma}$ the cross section obtained by varying ε by one standard deviation, the method of finite differences gives the following for the last term of equation (A.14):

$$V[\varepsilon] \left[\frac{\partial \hat{\sigma}}{\partial \varepsilon} \right]^2 \approx (\Delta_\varepsilon)^2 \left[\frac{\Delta_\varepsilon \hat{\sigma}}{\Delta_\varepsilon} \right]^2 = (\Delta_\varepsilon \hat{\sigma})^2. \quad (\text{A.15})$$

This makes the calculation very straightforward and intuitive.

The fact that each nuisance parameter appears as an independent term in the expression of the variance justifies the use of uncertainty breakdown tables and plots in which the contribution of individual terms are provided. The values of the “partial” standard deviations $\Delta_\varepsilon \hat{\sigma}$ are usually used, which are then said to be “added in quadrature” to calculate the total uncertainty. While the standard deviation is useful to get a sense of the magnitude of the uncertainty, the variance is more appropriate mathematically.

A.4 Differential cross sections

This thesis considers the measurement of differential cross sections, or more precisely of cross sections averaged over bins in some kinematic variable. In statistics, such an analysis is modelled as a simultaneous estimation of several quantities σ_k using estimators $\hat{\sigma}_k$ ($k = 1, 2, \dots, d$). The one-dimensional estimator used so far is straightforward to generalize:

$$\hat{\sigma}_k = \frac{1}{\mathcal{L}_c} \sum_i \frac{1}{\varepsilon_c(X_i)} \Theta_k(X_i). \quad (\text{A.16})$$

In additions to implementing cuts, the functions Θ_k now decide in which bin an event appears. Since each event is counted in a single bin, only one of the Θ_k can be non-zero for any event.

The differences with respect to the one-dimensional case appear when considering the uncertainties. The multidimensional equivalent of the variance is the covariance C , which is a $d \times d$ symmetric matrix defined as:

$$C_{kl} = E[(\hat{\sigma}_k - \sigma_{0,k})(\hat{\sigma}_l - \sigma_{0,l})], \quad (\text{A.17})$$

where $\sigma_{0,k}$ is the true cross section in bin k . The apparition of a matrix can be seen in the Gaussian approximation: the generalization of the normal distribution

to multiple dimensions is the multivariate normal distribution, of law:

$$\mathcal{N} \exp \left(-\frac{1}{2} (\mathbf{X} - \mathbf{X}_0)^T \mathbf{C}^{-1} (\mathbf{X} - \mathbf{X}_0) \right), \quad (\text{A.18})$$

with $\mathbf{X}$ the vector of random variables, $\mathbf{X}_0$ their central values and $\mathcal{N}$ a normalization constant.

The variance is a special case of the covariance matrix in dimension one. It can also be seen from the definition (A.17) that the diagonal elements of the covariance matrix correspond to the variance in each bin, translating the fact that the one-dimensional case is identical to the multidimensional one when only one of the bins is considered. In differential measurements, the quoted uncertainties correspond to the square roots of the diagonal elements of the covariance matrix.

Other basic properties of the covariance are similar to that of the variance. The covariance of the sum of two independent random vectors $\mathbf{X}$ and $\mathbf{Y}$ is the sum of their covariances:

$$\mathbf{C}[\mathbf{X} + \mathbf{Y}] = \mathbf{C}[\mathbf{X}] + \mathbf{C}[\mathbf{Y}]. \quad (\text{A.19})$$

Multiplying a random variable by a constant matrix $\mathbf{A}$ gives:

$$\mathbf{C}[\mathbf{A}\mathbf{X}] = \mathbf{A}\mathbf{C}[\mathbf{X}]\mathbf{A}^T. \quad (\text{A.20})$$

This last property is also useful to calculate, for instance, the sum of two correlated random scalars X and Y : the sum $X + Y$ can be rewritten as the matrix product $(1, 1) \cdot (X, Y)^T$ and its variance can be calculated using equation (A.20) if the covariance of (X, Y) is known. Similarly, equation (A.19) is a particular case of equation (A.20).

Under a sufficiently smooth transformation $\mathbf{Y} = f(\mathbf{X})$, the analogous of equations (A.8)–(A.9) in multiple dimensions involve the Jacobian matrix $J_{ij} = \partial f_i / \partial X_j$:

$$E[\mathbf{Y}] \approx f(\mathbf{X}_0), \quad (\text{A.21})$$

$$\mathbf{C}[\mathbf{Y}] \approx \mathbf{J}(\mathbf{X}_0) \mathbf{C}[\mathbf{X}] \mathbf{J}^T(\mathbf{X}_0). \quad (\text{A.22})$$

A.5 Particular covariance matrices

Two special cases of the covariance matrix are of particular interest. The first is the covariance of the random vector $\mathbf{Y}$ obtained by applying a function f to a single random variable X , $\mathbf{Y} = f(X)$. This problem appears for instance when propagating the uncertainty associated with nuisance parameters in the multidimensional generalization of equation (A.14). Using equation (A.22), one obtains for the covariance matrix of $\mathbf{Y}$:

$$\mathbf{C}[\mathbf{Y}] = (\nabla f) V[X] (\nabla f)^T = (\nabla f \Delta X) \otimes (\nabla f \Delta X). \quad (\text{A.23})$$

The covariance is thus a tensor product and looks as follows if $\mathbf{Y}$ has d dimensions:

$$C[\mathbf{Y}] = V[X] \begin{pmatrix} (\partial_1 f)^2 & (\partial_1 f)(\partial_2 f) & (\partial_1 f)(\partial_3 f) & \dots & (\partial_1 f)(\partial_d f) \\ (\partial_2 f)(\partial_1 f) & (\partial_2 f)^2 & (\partial_2 f)(\partial_3 f) & & (\partial_2 f)(\partial_d f) \\ (\partial_3 f)(\partial_1 f) & (\partial_3 f)(\partial_2 f) & (\partial_3 f)^2 & & (\partial_3 f)(\partial_d f) \\ \vdots & & & \ddots & \vdots \\ (\partial_d f)(\partial_1 f) & (\partial_d f)(\partial_2 f) & (\partial_d f)(\partial_3 f) & \dots & (\partial_d f)^2 \end{pmatrix}, \quad (\text{A.24})$$

with all partial derivatives evaluated at the central value of X . This “fully correlated” tensor product form is stable under further applications of equation (A.22), which makes it possible to propagate only one side of the tensor product and calculate the covariance matrix at the end.

Like equation (A.14), fully correlated covariance matrices are well suited for numerical evaluation using the method of finite differences. Similar to (A.15), one evaluates the rate of change of $\Delta \mathbf{Y}$ of f around the central value X_0 as $\Delta \mathbf{Y} = f(X_0 + \Delta X) - f(X_0)$, which gives:

$$C[\mathbf{Y}]_{ij} = V[X] \frac{\Delta Y_i}{\Delta X} \frac{\Delta Y_j}{\Delta X} = \Delta Y_i \Delta Y_j \quad (\text{A.25})$$

This method is used extensively thorough this thesis.

The second remarkable form of covariance matrices is the fully diagonal case:

$$C[\mathbf{X}] = \begin{pmatrix} V[X_1] & 0 & 0 & \dots & 0 \\ 0 & V[X_2] & 0 & & 0 \\ 0 & 0 & V[X_3] & & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & V[X_d] \end{pmatrix}. \quad (\text{A.26})$$

This form arises when statistically independent random variables are used as the components of a random vector. It appears in the statistical uncertainty of binned measurements where each event appears in a single bin. Unlike equation (A.24), this “uncorrelated” form is in general destroyed when equation (A.22) is used.

A.6 Correlation and interpretation

The off-diagonal elements in the covariance matrix are always bounded by the corresponding diagonal elements:

$$C_{ij}^2 \leq C_{ii} C_{jj}. \quad (\text{A.27})$$

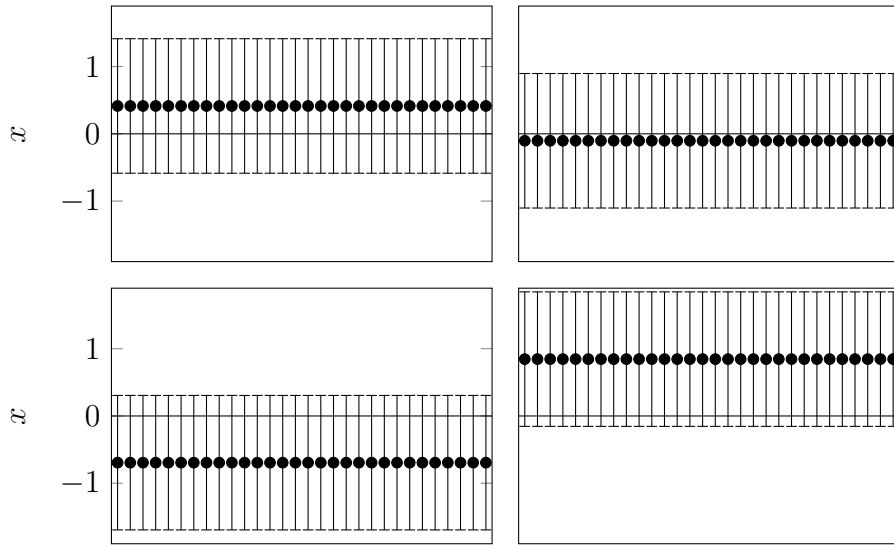


Figure A.1: Four draws from a 30-dimensional normal distribution with mean 0, variance 1 and 100 % correlation (i.e. $C_{ij} = 1$ for all i and j). Each point represents one dimension of the drawn vector.

This motivates the definition of the correlation matrix Corr :

$$\text{Corr}_{ij} \equiv \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}. \quad (\text{A.28})$$

Diagonal elements in the correlation matrix are always equal to one, and off-diagonal elements are in the interval $[-1, 1]$. For a fully correlated covariance matrices, every element of the correlation matrix is equal to ± 1 , while for uncorrelated covariance matrices all off-diagonal elements are zero.

The correlation matrix thus describes whether the multidimensional estimator used in a differential cross section measurement behave more like (A.24) or (A.26). The first case of a full correlation is illustrated on figure A.1 by drawing four samples from a multivariate normal distribution with all correlation coefficients equal to one. All points randomly move up or down in each sample, but without any variability among the bins. If instead a correlation coefficient -1 had been used between two bins, they would always move in opposite directions with respect to the true value. Finally, in the absence of correlation (figure A.2), each bin moves up or down independently of the others. In general, measurements are only partially correlated, corresponding to a mix between the two cases.

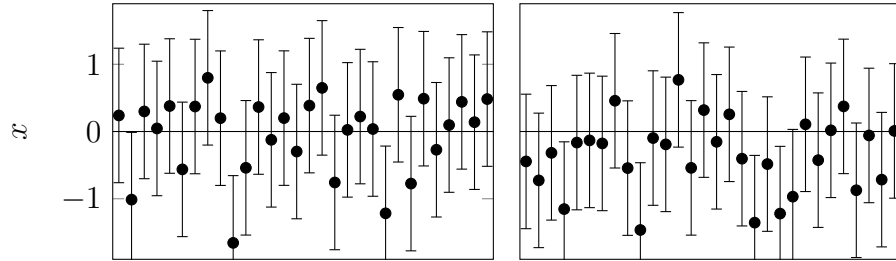


Figure A.2: Two draws from a 30-dimensional normal distribution with mean 0, variance 1 and zero correlation (i.e. $C_{ij} = \delta_{ij}$). Each point represents one dimension of the drawn vector.

Bibliography

- [1] L. Evans, P. Bryant et al., “LHC Machine”, *JINST* **3** (2008) S08001, doi:10.1088/1748-0221/3/08/S08001.
- [2] ALICE Collaboration, “The ALICE experiment at the CERN LHC”, *JINST* **3** (2008) S08002, doi:10.1088/1748-0221/3/08/S08002.
- [3] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider”, *JINST* **3** (2008) S08003, doi:10.1088/1748-0221/3/08/S08003.
- [4] CMS Collaboration, “The CMS Experiment at the CERN LHC”, *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [5] LHCb Collaboration, “The LHCb Detector at the LHC”, *JINST* **3** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.
- [6] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Phys. Lett. B* **716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.
- [7] CMS Collaboration, “Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC”, *Phys. Lett. B* **716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.
- [8] M. D. Schwartz, “Quantum Field Theory and the Standard Model”. Cambridge University Press, 2014. ISBN 978-1-107-03473-0.
- [9] J. Campbell, J. Huston, and F. Krauss, “The Black Book of Quantum Chromodynamics: A Primer for the LHC Era”. Oxford University Press, 2017. ISBN 978-0-19-965274-7.
- [10] C. Burgard, “Standard model of physics”, 2016. Accessed July 2021. <https://texample.net/tikz/examples/model-physics/>.

- [11] Particle Data Group Collaboration, “Review of Particle Physics”, *PTEP* **2020** (2020), no. 8, 083C01, doi:10.1093/ptep/ptaa104.
- [12] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons”, *Phys. Rev. Lett.* **13** (1964) 321–323, doi:10.1103/PhysRevLett.13.321.
- [13] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, *Phys. Rev. Lett.* **13** (1964) 508–509, doi:10.1103/PhysRevLett.13.508.
- [14] F. Herzog et al., “The five-loop beta function of Yang-Mills theory with fermions”, *JHEP* **02** (2017) 090, doi:10.1007/JHEP02(2017)090, arXiv:1701.01404.
- [15] M. E. Peskin and D. V. Schroeder, “An Introduction to quantum field theory”. Addison-Wesley, Reading, USA, 1995. ISBN 978-0-201-50397-5.
- [16] S. D. Drell and T.-M. Yan, “Massive Lepton-Pair Production in Hadron-Hadron Collisions at High Energies”, *Phys. Rev. Lett.* **25** (1970) 316–320, doi:10.1103/PhysRevLett.25.316. [Erratum: *Phys. Rev. Lett.* **25**, 902 (1970)].
- [17] J. H. Christenson et al., “Observation of massive muon pairs in hadron collisions”, *Phys. Rev. Lett.* **25** (1970) 1523–1526, doi:10.1103/PhysRevLett.25.1523.
- [18] J. C. Collins, D. E. Soper, and G. F. Sterman, “Factorization for Short Distance Hadron - Hadron Scattering”, *Nucl. Phys. B* **261** (1985) 104–142, doi:10.1016/0550-3213(85)90565-6.
- [19] V. N. Gribov and L. N. Lipatov, “Deep inelastic e p scattering in perturbation theory”, *Sov. J. Nucl. Phys.* **15** (1972) 438–450.
- [20] Y. L. Dokshitzer, “Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.”, *Sov. Phys. JETP* **46** (1977) 641–653.
- [21] G. Altarelli and G. Parisi, “Asymptotic Freedom in Parton Language”, *Nucl. Phys. B* **126** (1977) 298–318, doi:10.1016/0550-3213(77)90384-4.
- [22] S. Alekhin et al., “The PDF4LHC Working Group Interim Report”, technical report, 2011. arXiv:1101.0536.

- [23] A. Buckley et al., “LHAPDF6: parton density access in the LHC precision era”, *Eur. Phys. J. C* **75** (2015) 132, doi:10.1140/epjc/s10052-015-3318-8, arXiv:1412.7420.
- [24] NNPDF Collaboration, “Parton distributions for the LHC Run II”, *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.
- [25] K. Melnikov and F. Petriello, “Electroweak gauge boson production at hadron colliders through $O(\alpha_s^2)$ ”, *Phys. Rev. D* **74** (2006) 114017, doi:10.1103/PhysRevD.74.114017, arXiv:hep-ph/0609070.
- [26] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, “FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order”, *Comput. Phys. Commun.* **182** (2011) 2388–2403, doi:10.1016/j.cpc.2011.06.008, arXiv:1011.3540.
- [27] Y. Li and F. Petriello, “Combining QCD and electroweak corrections to dilepton production in the framework of the FEWZ simulation code”, *Phys. Rev. D* **86** (2012) 094034, doi:10.1103/PhysRevD.86.094034, arXiv:1208.5967. Also references therein.
- [28] T. Kinoshita, “Mass singularities of Feynman amplitudes”, *J. Math. Phys.* **3** (1962) 650–677, doi:10.1063/1.1724268.
- [29] T. D. Lee and M. Nauenberg, “Degenerate Systems and Mass Singularities”, *Phys. Rev.* **133** (1964) B1549–B1562, doi:10.1103/PhysRev.133.B1549.
- [30] Y. L. Dokshitzer, D. Diakonov, and S. I. Troian, “On the Transverse Momentum Distribution of Massive Lepton Pairs”, *Phys. Lett. B* **79** (1978) 269–272, doi:10.1016/0370-2693(78)90240-X.
- [31] W. Bizon et al., “The transverse momentum spectrum of weak gauge bosons at N³LL + NNLO”, *Eur. Phys. J. C* **79** (2019) 868, doi:10.1140/epjc/s10052-019-7324-0, arXiv:1905.05171.
- [32] P. F. Monni, E. Re, and P. Torrielli, “Higgs Transverse-Momentum Resummation in Direct Space”, *Phys. Rev. Lett.* **116** (2016) 242001, doi:10.1103/PhysRevLett.116.242001, arXiv:1604.02191.
- [33] W. Bizon et al., “Momentum-space resummation for transverse observables and the Higgs $p_\perp$ at N³LL+NNLO”, *JHEP* **02** (2018) 108, doi:10.1007/JHEP02(2018)108, arXiv:1705.09127.

- [34] ATLAS Collaboration, “Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector”, *JHEP* **09** (2014) 145, doi:10.1007/JHEP09(2014)145, arXiv:1406.3660.
- [35] T. Sjostrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual”, *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.
- [36] P. Skands, S. Carrazza, and J. Rojo, “Tuning PYTHIA 8.1: the Monash 2013 Tune”, *Eur. Phys. J. C* **74** (2014) 3024, doi:10.1140/epjc/s10052-014-3024-y, arXiv:1404.5630.
- [37] S. Frixione and B. R. Webber, “Matching NLO QCD computations and parton shower simulations”, *JHEP* **06** (2002) 029, doi:10.1088/1126-6708/2002/06/029, arXiv:hep-ph/0204244.
- [38] P. Nason, “A New method for combining NLO QCD with shower Monte Carlo algorithms”, *JHEP* **11** (2004) 040, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.
- [39] CMS Collaboration, “Event generator tunes obtained from underlying event and multiparton scattering measurements”, *Eur. Phys. J. C* **76** (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.
- [40] CMS Collaboration, “Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements”, *Eur. Phys. J. C* **80** (2020) 4, doi:10.1140/epjc/s10052-019-7499-4, arXiv:1903.12179.
- [41] CMS Collaboration, “Development and validation of HERWIG 7 tunes from CMS underlying-event measurements”, *Eur. Phys. J. C* **81** (2021) 312, doi:10.1140/epjc/s10052-021-08949-5, arXiv:2011.03422.
- [42] CMS Collaboration Collaboration, “Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV”, Technical Report CMS-PAS-TOP-16-021, CERN, Geneva, 2016.
- [43] T. Sjöstrand et al., “An introduction to PYTHIA 8.2”, *Comput. Phys. Commun.* **191** (2015) 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.
- [44] M. Bahr et al., “Herwig++ Physics and Manual”, *Eur. Phys. J. C* **58** (2008) 639–707, doi:10.1140/epjc/s10052-008-0798-9, arXiv:0803.0883.
- [45] J. Bellm et al., “Herwig 7.0/Herwig++ 3.0 release note”, *Eur. Phys. J. C* **76** (2016) 196, doi:10.1140/epjc/s10052-016-4018-8, arXiv:1512.01178.

- [46] Sherpa Collaboration, “Event Generation with Sherpa 2.2”, *SciPost Phys.* **7** (2019) 034, doi:10.21468/SciPostPhys.7.3.034, arXiv:1905.09127.
- [47] F. Hautmann et al., “Soft-gluon resolution scale in QCD evolution equations”, *Phys. Lett. B* **772** (2017) 446–451, doi:10.1016/j.physletb.2017.07.005, arXiv:1704.01757.
- [48] F. Hautmann et al., “Collinear and TMD Quark and Gluon Densities from Parton Branching Solution of QCD Evolution Equations”, *JHEP* **01** (2018) 070, doi:10.1007/JHEP01(2018)070, arXiv:1708.03279.
- [49] A. Bermudez Martinez et al., “Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method”, *Phys. Rev. D* **99** (2019) 074008, doi:10.1103/PhysRevD.99.074008, arXiv:1804.11152.
- [50] J. C. Collins, D. E. Soper, and G. F. Sterman, “Factorization for One-Loop Corrections in the Drell-Yan Process”, *Nucl. Phys. B* **223** (1983) 381–421, doi:10.1016/0550-3213(83)90062-7.
- [51] J. C. Collins, D. E. Soper, and G. F. Sterman, “Transverse Momentum Distribution in Drell-Yan Pair and W and Z Boson Production”, *Nucl. Phys. B* **250** (1985) 199–224, doi:10.1016/0550-3213(85)90479-1.
- [52] J. Collins, “Foundations of perturbative QCD”. Cambridge University Press, 2013. ISBN 978-1-107-64525-7, 978-0-521-85533-4, 978-1-139-09782-6.
- [53] S. M. Aybat and T. C. Rogers, “TMD Parton Distribution and Fragmentation Functions with QCD Evolution”, *Phys. Rev. D* **83** (2011) 114042, doi:10.1103/PhysRevD.83.114042, arXiv:1101.5057.
- [54] M. G. Echevarria, A. Idilbi, and I. Scimemi, “Factorization Theorem For Drell-Yan At Low q_T And Transverse-Momentum Distributions On-The-Light-Cone”, *JHEP* **07** (2012) 002, doi:10.1007/JHEP07(2012)002, arXiv:1111.4996.
- [55] T. Gehrmann, T. Lubbert, and L. L. Yang, “Transverse parton distribution functions at next-to-next-to-leading order: the quark-to-quark case”, *Phys. Rev. Lett.* **109** (2012) 242003, doi:10.1103/PhysRevLett.109.242003, arXiv:1209.0682.
- [56] J. Collins and T. C. Rogers, “Connecting Different TMD Factorization Formalisms in QCD”, *Phys. Rev. D* **96** (2017) 054011, doi:10.1103/PhysRevD.96.054011, arXiv:1705.07167.

- [57] I. Scimemi, “A short review on recent developments in TMD factorization and implementation”, *Adv. High Energy Phys.* (2019) 3142510, doi:10.1155/2019/3142510, arXiv:1901.08398.
- [58] I. Scimemi and A. Vladimirov, “Systematic analysis of double-scale evolution”, *JHEP* **08** (2018) 003, doi:10.1007/JHEP08(2018)003, arXiv:1803.11089.
- [59] I. Scimemi and A. Vladimirov, “Analysis of vector boson production within TMD factorization”, *Eur. Phys. J. C* **78** (2018) 89, doi:10.1140/epjc/s10052-018-5557-y, arXiv:1706.01473.
- [60] A. Bacchetta et al., “Transverse-momentum-dependent parton distributions up to N³LL from Drell-Yan data”, *JHEP* **07** (2020) 117, doi:10.1007/JHEP07(2020)117, arXiv:1912.07550.
- [61] A. Bacchetta et al., “Extraction of partonic transverse momentum distributions from semi-inclusive deep-inelastic scattering, Drell-Yan and Z-boson production”, *JHEP* **06** (2017) 081, doi:10.1007/JHEP06(2017)081, arXiv:1703.10157. [Erratum: JHEP 06, 051 (2019)].
- [62] I. Scimemi and A. Vladimirov, “Non-perturbative structure of semi-inclusive deep-inelastic and Drell-Yan scattering at small transverse momentum”, *JHEP* **06** (2020) 137, doi:10.1007/JHEP06(2020)137, arXiv:1912.06532.
- [63] F. Hautmann et al., “TMDlib and TMDplotter: library and plotting tools for transverse-momentum-dependent parton distributions”, *Eur. Phys. J. C* **74** (2014) 3220, doi:10.1140/epjc/s10052-014-3220-9, arXiv:1408.3015.
- [64] N. A. Abdulov et al., “TMDlib2 and TMDplotter: a platform for 3D hadron structure studies”, 2021, arXiv:2103.09741.
- [65] J. Wenninger, “Operation and Configuration of the LHC in Run 2”, CERN Accelerator Note CERN-ACC-NOTE-2019-0007, CERN, 2019.
- [66] C. De Melis, “The CERN accelerator complex”, CERN Open Photograph OPEN-PHO-ACCEL-2016-001, CERN, 2016.
- [67] S. van der Meer, “Calibration of the Effective Beam Height in the ISR”, ISR Report CERN-ISR-PO-68-31, 1968.
- [68] C. Rubbia, “Measurement of the Luminosity of p $\bar{p}$ Collider with a (Generalized) Van Der Meer Method”, CERN p $\bar{p}$ Note 38, 1977.

- [69] CMS Collaboration, “Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS”, 2021, [arXiv:2104.01927](#). Submitted to *Eur. Phys. J. C*.
- [70] CMS Collaboration, “Public CMS Luminosity Information”, 2021. Accessed August 2021. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [71] CMS Collaboration, “Precise Mapping of the Magnetic Field in the CMS Barrel Yoke using Cosmic Rays”, *JINST* **5** (2010) T03021, doi:10.1088/1748-0221/5/03/T03021, [arXiv:0910.5530](#).
- [72] CMS Collaboration, “Tracker Material Budget plots”, 2020. Accessed July 2021. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/TrackerMaterialBudgetplots>.
- [73] CMS Collaboration, “Operation and Performance of the Strip Tracker with pp collisions in LHC Run1 and Run2”, (2021). TRK-20-002 (version 1).
- [74] CMS Collaboration, “ECAL 2016 refined calibration and Run2 summary plots”, CMS Detector Performance Summary CMS-DP-2020-021.
- [75] CMS Collaboration, “CMS reconstruction improvement for the muon tracking by the RPC chambers”, *JINST* **8** (2013) T03001, doi:10.1088/1748-0221/8/03/T03001, [arXiv:1209.2646](#).
- [76] CMS Collaboration, “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JINST* **13** (2018), no. 06, P06015, doi:10.1088/1748-0221/13/06/P06015, [arXiv:1804.04528](#).
- [77] CMS Collaboration, “Resistive Plate Chambers”. Accessed August 2021. <https://cms.cern/detector/detecting-muons/resistive-plate-chambers>.
- [78] CMS Collaboration, “CMS Luminosity Measurements for the 2016 Data Taking Period”, CMS Physics Analysis Summary CMS-PAS-LUM-17-001, CERN, 2017.
- [79] CMS Collaboration, “Measurements of differential Z boson production cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **12** (2019) 061, doi:10.1007/JHEP12(2019)061, [arXiv:1909.04133](#).

- [80] CMS Collaboration, “Measurement of mass dependence of the transverse momentum of Drell Yan lepton pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis Summary CMS-PAS-SMP-20-003, CERN, 2021.
- [81] A. Banfi et al., “Optimisation of variables for studying dilepton transverse momentum distributions at hadron colliders”, *Eur. Phys. J. C* **71** (2011) 1600, doi:10.1140/epjc/s10052-011-1600-y, arXiv:1009.1580.
- [82] CMS Collaboration, “Measurement of the Z boson differential cross section in transverse momentum and rapidity in proton-proton collisions at 8 TeV”, *Phys. Lett. B* **749** (2015) 187, doi:10.1016/j.physletb.2015.07.065, arXiv:1504.03511.
- [83] CMS Collaboration, “Measurement of the double differential Drell-Yan cross section at low and high masses of the di-muon system at $\sqrt{s} = 7$ TeV”, CMS Physics Analysis Summary CMS-PAS-FSQ-13-003, CERN, 2014.
- [84] CMS Collaboration, “Measurement of the inclusive W and Z production cross sections in pp collisions at $\sqrt{s} = 7$ TeV”, *JHEP* **10** (2011) 132, doi:10.1007/JHEP10(2011)132, arXiv:1107.4789.
- [85] CMS Collaboration, “Measurement of the Drell-Yan cross section in pp collisions at $\sqrt{s} = 7$ TeV”, *JHEP* **10** (2011) 007, doi:10.1007/JHEP10(2011)007, arXiv:1108.0566.
- [86] CMS Collaboration, “Measurement of the differential and double-differential Drell-Yan cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV”, *JHEP* **12** (2013) 030, doi:10.1007/JHEP12(2013)030, arXiv:1310.7291.
- [87] CMS Collaboration, “Measurements of differential and double-differential Drell-Yan cross sections in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *Eur. Phys. J. C* **75** (2015) 147, doi:10.1140/epjc/s10052-015-3364-2, arXiv:1412.1115.
- [88] CMS Collaboration, “Measurement of the differential Drell-Yan cross section in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **12** (2019) 059, doi:10.1007/JHEP12(2019)059, arXiv:1812.10529.
- [89] ATLAS Collaboration, “Measurement of the inclusive $W^\pm$ and Z/γ^* cross sections in the electron and muon decay channels in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector”, *Phys. Rev. D* **85** (2012) 072004, doi:10.1103/PhysRevD.85.072004, arXiv:1109.5141.

- [90] ATLAS Collaboration, “Measurement of the high-mass Drell–Yan differential cross-section in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector”, *Phys. Lett. B* **725** (2013) 223, doi:10.1016/j.physletb.2013.07.049, arXiv:1305.4192.
- [91] ATLAS Collaboration, “Measurement of the low-mass Drell–Yan differential cross section at $\sqrt{s} = 7$ TeV using the ATLAS detector”, *JHEP* **06** (2014) 112, doi:10.1007/JHEP06(2014)112, arXiv:1404.1212.
- [92] ATLAS Collaboration, “Measurement of the transverse momentum and ϕ_η^* distributions of Drell–Yan lepton pairs in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”, *Eur. Phys. J. C* **76** (2016) 291, doi:10.1140/epjc/s10052-016-4070-4, arXiv:1512.02192.
- [93] ATLAS Collaboration, “Precision measurement and interpretation of inclusive W^+ , W^- and Z/γ^* production cross sections with the ATLAS detector”, *Eur. Phys. J. C* **77** (2017) 367, doi:10.1140/epjc/s10052-017-4911-9, arXiv:1612.03016.
- [94] ATLAS Collaboration, “Measurement of the transverse momentum distribution of Drell–Yan lepton pairs in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Eur. Phys. J. C* **80** (2020) 616, doi:10.1140/epjc/s10052-020-8001-z, arXiv:1912.02844.
- [95] LHCb Collaboration, “Inclusive W and Z production in the forward region at $\sqrt{s} = 7$ TeV”, *JHEP* **06** (2012) 058, doi:10.1007/JHEP06(2012)058, arXiv:1204.1620.
- [96] LHCb Collaboration, “Measurement of the cross-section for $Z \rightarrow e^+e^-$ production in pp collisions at $\sqrt{s} = 7$ TeV”, *JHEP* **02** (2013) 106, doi:10.1007/JHEP02(2013)106, arXiv:1212.4620.
- [97] LHCb Collaboration, “Measurement of the forward Z boson production cross-section in pp collisions at $\sqrt{s} = 7$ TeV”, *JHEP* **08** (2015) 039, doi:10.1007/JHEP08(2015)039, arXiv:1505.07024.
- [98] LHCb Collaboration, “Measurement of forward W and Z boson production in pp collisions at $\sqrt{s} = 8$ TeV”, *JHEP* **01** (2016) 155, doi:10.1007/JHEP01(2016)155, arXiv:1511.08039.
- [99] LHCb Collaboration, “Measurement of the forward Z boson production cross-section in pp collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **09** (2016) 136, doi:10.1007/JHEP09(2016)136, arXiv:1607.06495.

- [100] J. Alwall et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.
- [101] N. Davidson et al., “Universal Interface of TAUOLA Technical and Physics Documentation”, *Comput. Phys. Commun.* **183** (2012) 821–843, doi:10.1016/j.cpc.2011.12.009, arXiv:1002.0543.
- [102] R. Frederix and S. Frixione, “Merging meets matching in MC@NLO”, *JHEP* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.
- [103] GEANT4 Collaboration, “GEANT4—a simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003) 250–303, doi:10.1016/S0168-9002(03)01368-8.
- [104] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, *JINST* **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.
- [105] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* **9** (2014) P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.
- [106] CMS Collaboration, “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC”, *JINST* **16** (2021) P05014, doi:10.1088/1748-0221/16/05/P05014, arXiv:2012.06888.
- [107] CMS Collaboration, “Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV”, *JINST* **10** (2015) P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.
- [108] D. G. Hsu, “Precision measurement of and search for dark matter in the transverse momentum spectra of Z bosons”. PhD thesis CERN-THESIS-2019-378, MIT, 2019.
- [109] Junho Kim and Jihun Kim, “2016 Dielectron Trigger SF and POG Medium WP ID SF”. <https://indico.cern.ch/event/697095/#2-2016-double-electron-trigger>.
- [110] CMS Collaboration, “Reweighting recipe to emulate Level 1 ECAL prefiring”, 2021. Accessed August 2021. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/L1ECALPrefiringWeightRecipe>.

- [111] CMS Collaboration, “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JINST* **13** (2018) P06015, doi:10.1088/1748-0221/13/06/P06015, arXiv:1804.04528.
- [112] CMS Collaboration, “Performance of CMS Muon Reconstruction in pp Collision Events at $\sqrt{s} = 7$ TeV”, *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.
- [113] CMS Collaboration, “Performance of the reconstruction and identification of high-momentum muons in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JINST* **15** (2020) P02027, doi:10.1088/1748-0221/15/02/P02027, arXiv:1912.03516.
- [114] A. Bodek et al., “Extracting Muon Momentum Scale Corrections for Hadron Collider Experiments”, *Eur. Phys. J. C* **72** (2012) 2194, doi:10.1140/epjc/s10052-012-2194-8, arXiv:1208.3710.
- [115] CMS Collaboration, “The EMTF cancel-out bug, wrongly referred to as the “high-Pt” problem”, 2017. Accessed May 2021. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/EndcapHighPtMuonEfficiencyProblem>.
- [116] S. Chhibra Singh, A. Vartak, and S. Lowette, “Search for a dark photon resonance in the dimuon final state in proton-proton collisions at $\sqrt{s} = 13$ TeV”, CMS Analysis Note AN-2018/338, 2019.
- [117] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet user manual”, *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.
- [118] CMS Collaboration, “Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s} = 13$ TeV”, *JINST* **15** (2020) P05002, doi:10.1088/1748-0221/15/05/P05002, arXiv:1910.00079.
- [119] CMS Collaboration, “Pileup mitigation at CMS in 13 TeV data”, *JINST* **15** (2020) P09018, doi:10.1088/1748-0221/15/09/P09018, arXiv:2003.00503.
- [120] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”, *JINST* **12** (2017) P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.
- [121] CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, *JINST* **13** (2018) P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.

- [122] CMS Collaboration, “Jet identification in high pile-up environment (Pile-upJetID)”, 2020. Accessed August 2021. <https://twiki.cern.ch/twiki/bin/view/CMS/PileupJetID>.
- [123] CMS Collaboration, “Methods to apply b-tagging efficiency scale factors”, 2020. Accessed May 2021. <https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods>.
- [124] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”, *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.
- [125] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”, *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.
- [126] S. Frixione, P. Nason, and G. Ridolfi, “A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction”, *JHEP* **09** (2007) 126, doi:10.1088/1126-6708/2007/09/126, arXiv:0707.3088.
- [127] M. Czakon and A. Mitov, “Top++: A program for the calculation of the top-pair cross-section at hadron colliders”, *Comput. Phys. Commun.* **185** (2014) 2930, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.
- [128] S. Alioli, P. Nason, C. Oleari, and E. Re, “NLO single-top production matched with shower in POWHEG: s- and t-channel contributions”, *JHEP* **09** (2009) 111, doi:10.1088/1126-6708/2009/09/111, arXiv:0907.4076. [Erratum: JHEP 02, 011 (2010)].
- [129] E. Re, “Single-top Wt-channel production matched with parton showers using the POWHEG method”, *Eur. Phys. J. C* **71** (2011) 1547, doi:10.1140/epjc/s10052-011-1547-z, arXiv:1009.2450.
- [130] M. Aliev et al., “HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR”, *Comput. Phys. Commun.* **182** (2011) 1034–1046, doi:10.1016/j.cpc.2010.12.040, arXiv:1007.1327.
- [131] P. Kant et al., “HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions”, *Comput. Phys. Commun.* **191** (2015) 74–89, doi:10.1016/j.cpc.2015.02.001, arXiv:1406.4403.

- [132] N. Kidonakis, “Top Quark Production”, in *Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons*. 2013. [arXiv:1311.0283](#). [doi:10.3204/DESY-PROC-2013-03/Kidonakis](#).
- [133] J. M. Campbell, R. K. Ellis, and W. T. Giele, “A multi-threaded version of MCFM”, *Eur. Phys. J. C* **75** (2015) 246, [doi:10.1140/epjc/s10052-015-3461-2](#), [arXiv:1503.06182](#).
- [134] T. Melia, P. Nason, R. Rontsch, and G. Zanderighi, “W+W-, WZ and ZZ production in the POWHEG BOX”, *JHEP* **11** (2011) 078, [doi:10.1007/JHEP11\(2011\)078](#), [arXiv:1107.5051](#).
- [135] T. Gehrmann et al., “W⁺W⁻ Production at Hadron Colliders in Next to Next to Leading Order QCD”, *Phys. Rev. Lett.* **113** (2014) 212001, [doi:10.1103/PhysRevLett.113.212001](#), [arXiv:1408.5243](#).
- [136] J. A. M. Vermaseren, “Two photon processes at very high-energies”, *Nucl. Phys. B* **229** (1983) 347, [doi:10.1016/0550-3213\(83\)90336-X](#).
- [137] S. Baranov, O. Duenger, H. Shooshtari, and J. A. M. Vermaseren, “LPAIR: A generator for lepton pair production”, in *Workshop on Physics at HERA*, p. 1478. 1991.
- [138] L. Forthomme, “CepGen - A generic central exclusive processes event generator for hadron-hadron collisions”, 2018, [arXiv:1808.06059](#).
- [139] CMS Collaboration, “Measurement of differential cross sections for Z boson production in association with jets in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Eur. Phys. J. C* **78** (2018) 965, [doi:10.1140/epjc/s10052-018-6373-0](#), [arXiv:1804.05252](#).
- [140] CMS Collaboration, “Measurement of multi-differential cross sections for the production of a Z boson in association with jets in proton-proton collisions at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis Summary CMS-PAS-SMP-19-009, CERN, 2021.
- [141] CMS Collaboration, “Measurements of $t\bar{t}$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons”, *JHEP* **02** (2019) 149, [doi:10.1007/JHEP02\(2019\)149](#), [arXiv:1811.06625](#).
- [142] CMS Collaboration, “Recommendations on Unfolding”, 2021. Accessed May 2021. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/ScrecUnfolding>.

- [143] G. D’Agostini, “A multidimensional unfolding method based on bayes’ theorem”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **362** (1995) 487–498, doi:10.1016/0168-9002(95)00274-X.
- [144] W. H. Richardson, “Bayesian-based iterative method of image restoration”, *J. Opt. Soc. Am.* **62** (1972) 55–59, doi:10.1364/JOSA.62.000055.
- [145] L. B. Lucy, “An iterative technique for the rectification of observed distributions”, *Astronomical Journal* **79** (1974) 745, doi:10.1086/111605.
- [146] V. A. Il’in, “Tikhonov’s work on methods of solving ill-posed problems”, *Russian Mathematical Surveys* **22** (1967) 142–149, doi:10.1070/rm1967v022n02abeh001216.
- [147] S. Schmitt, “TUnfold: an algorithm for correcting migration effects in high energy physics”, *JINST* **7** (2012) T10003, doi:10.1088/1748-0221/7/10/T10003, arXiv:1205.6201.
- [148] A. Hocker and V. Kartvelishvili, “SVD approach to data unfolding”, *Nucl. Instrum. Meth. A* **372** (1996) 469–481, doi:10.1016/0168-9002(95)01478-0, arXiv:hep-ph/9509307.
- [149] T. Adye, “Unfolding algorithms and tests using RooUnfold”, in *PHYSTAT 2011*, pp. 313–318. CERN, Geneva, 2011. arXiv:1105.1160. doi:10.5170/CERN-2011-006.313.
- [150] C. Bierlich et al., “Robust Independent Validation of Experiment and Theory: Rivet version 3”, *SciPost Phys.* **8** (2020) 026, doi:10.21468/SciPostPhys.8.2.026, arXiv:1912.05451.
- [151] CMS Collaboration, “Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV”, *JHEP* **07** (2018) 161, doi:10.1007/JHEP07(2018)161, arXiv:1802.02613.
- [152] CMS Collaboration, “NNLO+NNLL top-quark-pair cross sections”, 2021. Accessed July 2021. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TtbarNNLO>.
- [153] A. Léonard, “Measurement of Z boson production in association with jets at the LHC and study of a DAQ system for the Triple-GEM detector in view of the CMS upgrade”. PhD thesis CERN-THESIS-2015-200, Brussels U., 2015.

- [154] F. Zhang, “Measurement of jet production in association with a Z boson at the LHC and jet energy correction calibration at high level trigger in CMS”. PhD thesis CERN-THESIS-2017-086, Peking U., Beijing, Brussels U., 2017.
- [155] Q. Wang, “Measurement of the differential crosssection of Z boson production in association with jets at the LHC”. PhD thesis, Peking U., Beijing, Brussels U., 2018.
- [156] R. Frederix et al., “The automation of next-to-leading order electroweak calculations”, *JHEP* **07** (2018) 185, doi:10.1007/JHEP07(2018)185, arXiv:1804.10017.
- [157] J. Butterworth et al., “PDF4LHC recommendations for LHC Run II”, *J. Phys. G* **43** (2016) 023001, doi:10.1088/0954-3899/43/2/023001, arXiv:1510.03865.
- [158] NNPDF Collaboration, “Parton distributions from high-precision collider data”, *Eur. Phys. J. C* **77** (2017) 663, doi:10.1140/epjc/s10052-017-5199-5, arXiv:1706.00428.
- [159] J. Alwall et al., “A Standard format for Les Houches event files”, *Comput. Phys. Commun.* **176** (2007) 300–304, doi:10.1016/j.cpc.2006.11.010, arXiv:hep-ph/0609017.
- [160] CMS Collaboration, “Measurements of production cross sections of the Higgs boson in the four-lepton final state in proton–proton collisions at $\sqrt{s} = 13$ TeV”, *Eur. Phys. J. C* **81** (2021) 488, doi:10.1140/epjc/s10052-021-09200-x, arXiv:2103.04956.
- [161] H. Jung et al., “The CCFM Monte Carlo generator CASCADE version 2.2.03”, *Eur. Phys. J. C* **70** (2010) 1237–1249, doi:10.1140/epjc/s10052-010-1507-z, arXiv:1008.0152.
- [162] S. Baranov et al., “CASCADE3 A Monte Carlo event generator based on TMDs”, *Eur. Phys. J. C* **81** (2021) 425, doi:10.1140/epjc/s10052-021-09203-8, arXiv:2101.10221.
- [163] A. B. Martinez, F. Hautmann, and M. L. Mangano, “TMD Evolution and Multi-Jet Merging”, 2021, arXiv:2107.01224.
- [164] P. F. Monni et al., “MiNNLO_{PS}: a new method to match NNLO QCD to parton showers”, *JHEP* **05** (2020) 143, doi:10.1007/JHEP05(2020)143, arXiv:1908.06987.

- [165] K. Hamilton, P. Nason, and G. Zanderighi, “MINLO: Multi-Scale Improved NLO”, *JHEP* **10** (2012) 155, doi:10.1007/JHEP10(2012)155, arXiv:1206.3572.
- [166] K. Hamilton, P. Nason, C. Oleari, and G. Zanderighi, “Merging H/W/Z + 0 and 1 jet at NLO with no merging scale: a path to parton shower + NNLO matching”, *JHEP* **05** (2013) 082, doi:10.1007/JHEP05(2013)082, arXiv:1212.4504.
- [167] K. Hamilton, P. Nason, E. Re, and G. Zanderighi, “NNLOPS simulation of Higgs boson production”, *JHEP* **10** (2013) 222, doi:10.1007/JHEP10(2013)222, arXiv:1309.0017.
- [168] A. Karlberg, E. Re, and G. Zanderighi, “NNLOPS accurate Drell-Yan production”, *JHEP* **09** (2014) 134, doi:10.1007/JHEP09(2014)134, arXiv:1407.2940.
- [169] E. Barberio and Z. Was, “PHOTOS - a universal Monte Carlo for QED radiative corrections: version 2.0”, *Comput. Phys. Commun.* **79** (1994) 291, doi:10.1016/0010-4655(94)90074-4.
- [170] S. Alioli et al., “Drell-Yan production at NNLL'+NNLO matched to parton showers”, *Phys. Rev. D* **92** (2015) 094020, doi:10.1103/PhysRevD.92.094020, arXiv:1508.01475.
- [171] S. Alioli et al., “Matching NNLO to parton shower using N³LL colour-singlet transverse momentum resummation in GENEVA”, ZU-TH 8/21, 2021. arXiv:2102.08390.
- [172] I. W. Stewart, F. J. Tackmann, and W. J. Waalewijn, “N-Jettiness: An Inclusive Event Shape to Veto Jets”, *Phys. Rev. Lett.* **105** (2010) 092002, doi:10.1103/PhysRevLett.105.092002, arXiv:1004.2489.
- [173] A. M. van Kampen, “Drell-Yan transverse spectra at the LHC: a comparison of parton branching and analytical resummation approaches”, in *28th International Workshop on Deep Inelastic Scattering and Related Subjects*. 8, 2021. arXiv:2108.04099.
- [174] CMS Collaboration, “Study of Final-State Radiation in Decays of Z Bosons Produced in pp Collisions at 7 TeV”, *Phys. Rev. D* **91** (2015) 092012, doi:10.1103/PhysRevD.91.092012, arXiv:1502.07940.
- [175] CMS Collaboration, “Azimuthal correlations in Z+jets events at 13 TeV”, CMS Physics Analysis Summary CMS-PAS-SMP-21-003, CERN, 2021.

- [176] P. Bartalini and J. R. Gaunt, eds., “Multiple Parton Interactions at the LHC”, volume 29. WSP, 2018. doi:10.1142/10646, ISBN 978-981-322-775-0, 978-981-322-777-4.
- [177] G. Lustermans, J. K. L. Michel, F. J. Tackmann, and W. J. Waalewijn, “Joint two-dimensional resummation in q_T and 0-jettiness at NNLL”, *JHEP* **03** (2019) 124, doi:10.1007/JHEP03(2019)124, arXiv:1901.03331.
- [178] CMS Collaboration, “CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis Summary CMS-PAS-LUM-17-004, CERN, 2018.
- [179] CMS Collaboration, “CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis Summary CMS-PAS-LUM-18-002, CERN, 2019.
- [180] F. James, “Statistical methods in experimental physics”. 2006. ISBN 978-981-256-795-6, 978-981-270-527-3.