NATIONAL TECHNICAL UNIVERSITY OF ATHENS
SCHOOL OF APPLIED MATHEMATICAL AND PHYSICAL SCIENCES
DEPARTMENT OF PHYSICS

SEARCH FOR SECOND GENERATION LEPTOQUARK PARTICLES IN ATLAS/LHC EXPERIMENT

A dissertation by

Georgios D. Zacharis, MSc

to obtain the degree of
Doctor of Philosophy
in the subject of Physics
under the authority of
Professor Th. Papadopoulou

National Technical University of Athens

December 2018

SEARCH FOR SECOND GENERATION LEPTOQUARK PARTICLES IN ATLAS/LHC EXPERIMENT

Ph.D Thesis

Georgios D. Zacharis

Advisor: Theodora Papadopoulou

Professor, NTU-Athens

Exam committee:

.....	.....	.....
Nikos Tracas Professor, NTUA	Dimitrios Fassouliotis Assoc. Professor, NKUA	Evangelos Gazis Professor, NTUA

.....	.....	.....	.....
Theodoros Alexopoulos Professor, NTUA	Ioannis Gkialas Professor, Un. of Aegean	Stavros Maltezos Professor, NTUA	Aristotelis Kyriakis Researcher, NCSR "Demokritos"

.....

Georgios D. Zacharis

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ
ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΑΝΑΖΗΤΗΣΗ ΤΩΝ ΛΕΠΤΟΚΟΥΑΡΚ ΣΩΜΑΤΙΔΙΩΝ ΤΗΣ ΔΕΥΤΕΡΗΣ ΓΕΝΙΑΣ ΣΤΟ ΠΕΙΡΑΜΑ ATLAS/LHC

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Γεώργιος Δ. Ζαχαρής

Φυσικός Πανεπιστημίου Πατρών

ΕΠΙΒΛΕΠΟΥΣΑ:
Θ. ΠΑΠΑΔΟΠΟΥΛΟΥ
ΚΑΘΗΓΗΤΡΙΑ Ε.Μ.Π.

Αθήνα, 25 Δεκεμβρίου 2018

.....

Γεώργιος Δ. Ζαχαρής

(2018) ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ.

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ
ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΑΝΑΖΗΤΗΣΗ ΤΩΝ ΛΕΠΤΟΚΟΥΑΡΚ ΣΩΜΑΤΙΔΙΩΝ ΤΗΣ ΔΕΥΤΕΡΗΣ ΓΕΝΙΑΣ ΣΤΟ ΠΕΙΡΑΜΑ ATLAS/LHC

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Γεώργιος Δ. Ζαχαρής

Φυσικός Πανεπιστημίου Πατρών

ΤΡΙΜΕΛΗΣ ΣΥΜΒΟΥΛΕΥΤΙΚΗ ΕΠΙΤΡΟΠΗ:

1. Θ. Παπαδοπούλου, Καθ. Ε.Μ.Π. (Επιβλέπουσα)
2. Ν. Τράκας, Καθ. Ε.Μ.Π.
3. Δ. Φασουλιώτης, Αν. Καθ. Ε.Κ.Π.Α.

ΕΠΤΑΜΕΛΗΣ ΣΥΜΒΟΥΛΕΥΤΙΚΗ ΕΠΙΤΡΟΠΗ:

1. Ν. Τράκας, Καθ. Ε.Μ.Π.
2. Δ. Φασουλιώτης, Αν. Καθ. Ε.Κ.Π.Α.
3. Ε. Γαζής, Καθ. Ε.Μ.Π.
4. Θ. Αλεξόπουλος, Καθ. Ε.Μ.Π.
5. Σ. Μαλτέζος, Καθ. Ε.Μ.Π.
6. Ι. Γκιάλας, Καθ. Παν. Αιγαίου
7. Α. Κυριάκης, Ερευνητής Β΄ ΕΚΕΦΕ "Δημόκριτος"

Αθήνα, 25 Δεκεμβρίου 2018

Ευχαριστίες

Η παρούσα εργασία αφιερώνεται στους γονείς μου, Μαρία και Δημήτρη, καθώς και στην αδερφή μου Άννα για την άνευ όρων αγάπη και υποστήριξή τους, καθώς και τη χρηματοδότηση της έρευνάς μου κατά τις μακρές περιόδους που το εκπαιδευτικό σύστημα της πατρίδας μου αδυνατούσε να το πράξει. Τους χρωστάω τα πάντα.

Αφιερώνεται ακόμα στη μνήμη του Πέτρου Σκαμαγκούλη, ενός πολλά υποσχόμενου νέου θεωρητικού φυσικού και καλού φίλου που έφυγε από κοντά μας απρόσμενα νωρίς βυθίζοντας στη θλίψη όλους όσοι τον γνώριζαν. Θα μας λείπει για πάντα.

Θέλω να ευχαριστήσω την καθηγήτρια Θεοδώρα Παπαδοπούλου που επέβλεψε την παρούσα διδακτορική διατριβή δίνοντάς μου την ευκαιρία να συμμετέχω σε αυτήν την περιπέτεια. Ήταν δύσκολα, αλλά έμαθα πολλά. Η καθοδήγηση και η υποστήριξή της ήταν μείζονος σημασίας για την επιτυχία αυτής της προσπάθειας.

Επίσης, θα ήθελα να εκφράσω τις εγκάρδιες ευχαριστίες μου στον δρ. Νεκτάριο Μπενέκο και τον δρ. Ηλία Παναγούλια για την πολύτιμη και πολύπλευρη βοήθειά τους, τα σχόλια και την ιώβεια υπομονή τους. Εκτός από δύο αφοσιωμένους συνεργάτες, γνώρισα και δύο καλούς φίλους.

Τέλος, ευχαριστώ τον φίλο και συνάδελφο θεωρητικό φυσικό Γιάννη Μητσούλα για τα σχόλια, τις παρατηρήσεις και την απαραίτητη βοήθειά του στο θεωρητικό μέρος αυτής της εργασίας.

Χωρίς αυτούς τους ανθρώπους, τίποτα δε θα μπορούσε να έχει γίνει πραγματικότητα.

Acknowledgements

This project is dedicated to my parents, Dimitris and Maria, and my sister Anna for their unconditional love and support, as well as the financial support of my research for long periods of time during which my country 's educational system was not able to do so. I owe them everything.

It is also dedicated to the memory of Petros Skamagoulis, a prominent young theoretical physicist and a good friend, who left us unexpectedly soon adding the burden of grief on our shoulders. It was a loss for Physics, but a greater loss for us. He is always missed.

I would like to thank my supervisor professor Theodora Papadopoulou for giving me the opportunity to get into this adventure. Her guidance and support was of critical importance for the success of this attempt.

I would also wish to cordially thank dr. Nektarios Benekos and dr. Ilias Panagoulias for their help in so many levels, dedication, comments and patience of Job. They honoured me with their sincere friendship.

Finally, I thank my friend and colleague theoretical physicist Giannis Mitsoulas for his comments, observations and the necessary help in the theoretical section of this work.

They all made everything happen.

"...that the powerful game is on and you may contribute a verse..."

♠ Robin Williams as John Keating in "*Dead Poets Society*"

"We 're looking for a needle in a haystack and we do not know what that needle looks like."

♣ Hugh Laurie as Gregory House in "*House, M.D.*"

Contents

Abstract

Imaging an archaeologist (like dr. Jones) searching for a rumored rare manuscript in a buried ancient library. Now imaging there are tens of millions of ancient scrolls, all of them looking similar to each other. Add to the scenario the fact they are all torn apart in extra small pieces. Indy 's job is, for proving its existence, to find and gather all the pieces that specific scroll consists of without knowing how many they are or how exactly the scroll looks like.

Trying to discover a new particle looks something like that. It 's not an easy task at all. Leptoquarks (the scroll) are not stable. This means that one must be in a position to tell with great confidence that a new unstable particle has been discovered at the very same time it has already decayed to other particles in an infinetesimal fraction of a moment.

The main idea behind this is to record and identify stable known particles like electrons, muons etc which have been theoretically calculated and expected to be the debris of the decay of the new particle (scroll pieces) as its unique "signature". A rather non trivial work considering that already known bosons (all the other scrolls) can decay to the same particles with the same characteristics (look-alike pieces). This exactly is the most strenuous issue about detector 's recorded data analysis since one must identify something about which can never be entirely sure what exactly looks like.

Preface

The present dissertation consists of the following Chapters:

- **Κεφάλαιο 1: Ανάλυση Δεδομένων του Ανιχνευτή ATLAS στα 8 TeV**

  Στο πρώτο κεφάλαιο γίνεται μία σύντομη παρουσίαση στην ελληνική γλώσσα της ανάλυσης σε ενέργεια κέντρου μάζας $\sqrt{s} = 8$ TeV και συνολική φωτεινότητα 20.3 fb^{-1} που πραγματοποιήθηκε για την ανίχνευση του ζεύγους μποζονίων λεπτοκουάρκ δεύτερης γενιάς στην τοπολογία με ελλείπουσα εγκάρσια ενέργεια.

- **Chapter 2: Introduction**

  A brief introduction about the roots of scientific thinking in the ancient world, as well as the beginning of elementary particles Physics starting at the end of 19th century it takes place. It is also presented list of the different kinds of particle accelerating machines from the very beginning until nowadays and (distanced) future plans.

- **Chapter 3: Standard Model of the Elementary Particles**

  The main features of elementary particles Standard Model (SM) and its greatest successes are presented. The most significant still unanswered by SM questions are also presented, as well as some of the most important beyond SM theories and their implications integrating leptoquark boson.

- **Chapter 4: LHC Accelerator and the ATLAS Detector**

  Describes the Large Hadron Collider (LHC) and its components. ATLAS detector 's components are also extensively introduced since their function and efficiency is crucial for data acquisition. All the data used for the both analyses have been collected by ATLAS.

- **Chapter 5: ATLAS Muon Spectrometer CSC Sub-detectors Monitoring**

  Focuses onto CSC sub-detectors. CSCs are a part of ATLAS Muon Spectrometer (all ATLAS sub-detectors dedicated for muon identification and recording). Their performance, along with the other MS components, is crucial for muon objects acquisition and quality check. CSC code mantainance and offline data quality was named as my ATLAS collaboration Qualification Task.

- **Chapter 6: LHC Data LQ Analysis at 8 TeV, Run I**

  In the sixth chapter, it is presented the analysis for the pair scalar leptoquark bosons (LQ) of second generation done with the data collected by ATLAS during Run I period. The center-of-mass energy of both proton beams at $\sqrt{s} = 8$ TeV and a total luminosity of 20.3 fb^{-1}. The results and conclusions ot this analysis are discussed at the end of the section.

- **Chapter 7: LHC Data LQ Analysis at 13 TeV, Run II**

  Like in Ch. 6, here it is presented the analysis again for second generation pair scalar LQs based on the data collected by ATLAS during Run II period, for both beams center-of-mass energy at $\sqrt{s} = 13$ TeV and a total luminosity of 36.1 fb^{-1}. This chapter also concludes with the results of this analysis.

- **Appendix A: Supplementary Material of 8 TeV Analysis**

  Contains more plots and information about LQ analysis of $\sqrt{s} = 8$ TeV.

- **Appendix B: Supplementary Material of 13 TeV Analysis**

  This is the final chapter containing supplementary information about LQ analysis of $\sqrt{s} = 13$ TeV.

Κεφάλαιο 1

Ανάλυση Δεδομένων του ATLAS στα 8 TeV

1.1 Εισαγωγή

Στα πλαίσια του παρόντος διδακτορικού πραγματοποιήθηκαν δύο αναλύσεις των δεδομένων του ανιχνευτή ATLAS του επιταχυντή LHC για την ανίχνευση του υποθετικού σωματιδίου λεπτοκουάρκ.

Η πρώτη ανάλυση βασίστηκε στα δεδομένα που κατέγραψε ο ανιχνευτής κατά την πρώτη περίοδο λήψης και καταγραφής δεδομένων (Run 1) το 2012 με ενέργεια κέντρου μάζας και των δύο δεσμών πρωτονίων ίση με 8 TeV και συνολική φωτεινότητα ίση με 20.3 fb^{-1}.

1.2 Ανάλυση Δεδομένων του ανιχνευτή ATLAS στα 8 TeV (Περίοδος Ι)

Η παρούσα ανάλυση επικεντρώνεται στην έρευνα για την ανίχνευση ζεύγους βαθμωτών[1] λεπτοκουάρκ σωματιδίων δεύτερης γενιάς. Η θεωρία προβλέπει ότι τα λεπτοκουάρκ είναι ασταθή σωματίδια και, εφόσον παραχθούν, διασπώνται ακαριαία σε άλλα σταθερά παράγωγα σωματίδια. Προβλέπονται διαφορετικές πιθανές τοπολογίες διάσπασης για κάθε γενιά λεπτονίων και κουάρκ.

Στη συγκεκριμένη ανάλυση λαμβάνεται ως δεδομένο ότι η μίξη λεπτονίων και κουάρκ προερχόμενων από διαφορετικές γενιές είναι απαγορευτική. Εδώ η τοπολογία που ερευνάται είναι το ζεύγος βαθμωτού λεπτοκουάρκ δεύτερης γενιάς με ελλείπουσα εγκάρσια ενέργεια (E_T^{miss}). Οι κυριότεροι τρόποι παραγωγής ενός τέτοιου ζεύγους που προέρχονται από συγκρούσεις δεσμών πρωτονίων είναι μέσω της αλληλεπίδρασης δύο γκλουονίων (gg) ή ενός κουάρκ κι ενός αντι-κουάρκ ($q\overline{q}$), ένα από κάθε δέσμη. Το πρώτο λεπτοκουάρκ διασπάται σε ένα μιόνιο κι ένα κουάρκ δεύτερης γενιάς - charm ή strange (ή τα αντι-σωματίδιά τους) - και το δεύτερο σε ένα νετρίνο μιονίου (ν_μ) κι επίσης ένα κουάρκ δεύτερης γενιάς (ή τα αντι-σωματίδιά τους). Η θεωρία προβλέπει επίσης ότι το ζεύγος αποτελείται από ένα λεπτοκουάρκ (LQ) κι ένα αντι-λεπτοκουάρκ ($\overline{LQ}$).

Δεδομένου ότι ο ανιχνευτής δεν είναι κατασκευασμένος για την ανίχνευση νετρίνων, αυτά γίνονται αντιληπτά ως ελλείπουσα εγκάρσια ενέργεια υπολογιζόμενη μέσω των αντίστοιχων Αρχών Διατήρησης της Ενέργειας και της Ορμής. Επίσης, σύμφωνα

[1] τα βαθμωτά λεπτοκουάρκ έχουν σπιν 0

με το φαινόμενο της ασυμπτωτικής ελευθερίας, ένα κουάρκ δεν είναι δυνατό να παρατηρηθεί ποτέ ελεύθερο. Αυτός που είναι ανιχνεύσιμος είναι ένας πίδακας (τζετ) που προέρχεται από ένα κουάρκ μέσω της διαδικασίας της αδρονοποίησης. Γι' αυτό, τα τελικά παράγωγα σωματίδια της τοπολογίας που μελετάται εδώ και μπορούν να ανιχνευθούν ευθέως είναι μιόνια, πίδακες και η ελλείπουσα εγκάρσια μάζα (μέσω της απουσίας της).

Το κλάσμα διάσπασης (β) εκφράζει το ποσοστό που το αρχικό ζεύγος λεπτοκουάρκ διασπάται σε μιόνιο-πίδακα και σε νετρίνο μιονίου-πίδακα. Προφανώς οι τιμές που μπορεί να πάρει είναι $0 \leq \beta \leq 1$. Δεδομένου ότι η πραγματική τιμή αυτής της παραμέτρου είναι άγνωστη, κατά σύμβαση αποδίδεται η τιμή $\beta = 0.5$ που είναι και η τιμή με τη μεγαλύτερη ευαισθησία ανίχνευσης. Ο λόγος είναι ότι θεωρητικά η διάσπαση των δύο λεπτοκουάρκ είναι ισοπίθανη.

Εκτός από τα πραγματικά καταγεγραμμένα δεδομένα του ανιχνευτή, στην παρούσα ανάλυση χρησιμοποιήθηκαν αρχεία Monte Carlo (MC) για την προσομοίωση σήματος λεπτοκουάρκ, καθώς και αρχεία προσομοίωσης γνωστών διαδικασιών του Καθιερωμένου Προτύπου (ΚΠ) οι οποίες καταλήγουν στην ίδια τελική κατάσταση με αυτή που μελετάται και συμβάλλουν ως υπόβαθρο. Στη συνέχεια ακολουθούν οι ενότητες που περιγράφονται τα κριτήρια επιλογής των αντίστοιχων φυσικών αντικειμένων (μιονίων, πιδάκων κλπ), τα κριτήρια επιλογής τως γεγονότων[2], οι ορισμοί των περιοχών ελέγχου και περιοχών σήματος, η βελτιστοποίησή τους, η στατιστική ανάλυση, καθώς και τα τελικά αποτελέσματα και συμπεράσματα.

1.2.1 Αρχεία δεδομένων και αρχεία Monte Carlo

Τα δεδομένα του ανιχνευτή που χρησιμοποιήθηκαν για αυτή τη μελέτη έχουν περάσει από τα επίσημα ποιοτικά κριτήρια. Τα κυριότερα υπόβαθρα που συμμετέχουν στη συγκεκριμένη τοπολογία είναι τα W+πίδακες, Z/γ^*+πίδακες, $t\bar{t}$ (όπου και τα δύο ταυ λεπτόνια διασπώνται λεπτονικά), μονό τοπ (γεγονότα με ένα τοπ κουάρκ), καθώς και διμποζονικά γεγονότα (που περιέχουν δύο μποζόνια της ασθενούς αλληλεπίδρασης, δηλ. WW, ZZ ή WZ). Επιπλέον, υπάρχει και το υπόβαθρο που προέρχεται από την Κβαντική ΧρωμοΔυναμική (ΚΧΔ), το οποίο όμως δεν είναι δυνατό να αναπαραχθεί με κάποια γεννήτρια τυχαίων γεγονότων και, ως εκ τούτου, γίνεται μία έμμεση εκτίμηση της συνολικής συνεισφοράς του μέσω των πραγματικών δεδομένων (μέθοδος ΑΒΓΔ η οποία αναφέρεται αναλυτικότερα παρακάτω).

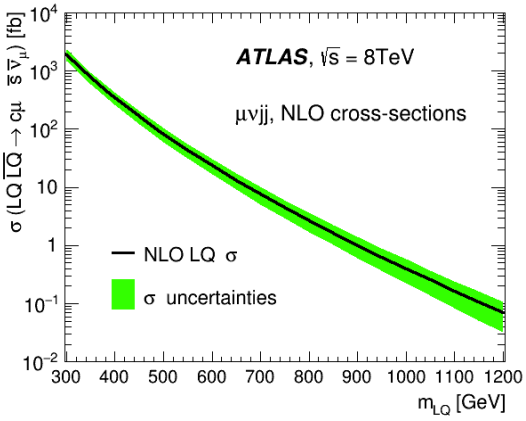
Η γεννήτρια Pythia 8.165 χρησιμοποιήθηκε για την παραγωγή των αρχείων του σήματος. Συνολικά παρήχθησαν 19 τέτοια αρχεία ξεκινώντας από ονομαστική μάζα λεπτοκουάρκ ίση με 300 GeV και καταλήγοντας σε μάζα ίση με 1200 GeV με βήμα 50 GeV. Η παραγωγή τους έγινε κεντρικά από το πείραμα ATLAS για λογαριασμό της ομάδας ανάλυσης των λεπτοκουάρκ. Στο Σχ. 1.1 παρουσιάζονται οι ενεργές διατομές παραγωγής του ζεύγους υπολογισμένες σε πρώτο όρο διαταραχής από τα 300 GeV μέχρι τα 1200 GeV.

1.2.2 Προσομοίωση και επικύρωση σήματος

Προκειμένου να παραχθεί το σήμα που πρόκειται να χρησιμοποιηθεί στην ανάλυση, δοκιμάστηκαν διαφορετικές εκδόσεις της γεννήτριας Pythia. Συγκεκριμένα δοκιμάστηκε η έκδοση Pythia 6 - PDF[3] CTEQ6L1, η έκδοση Pythia 6 - PDF D6 και η

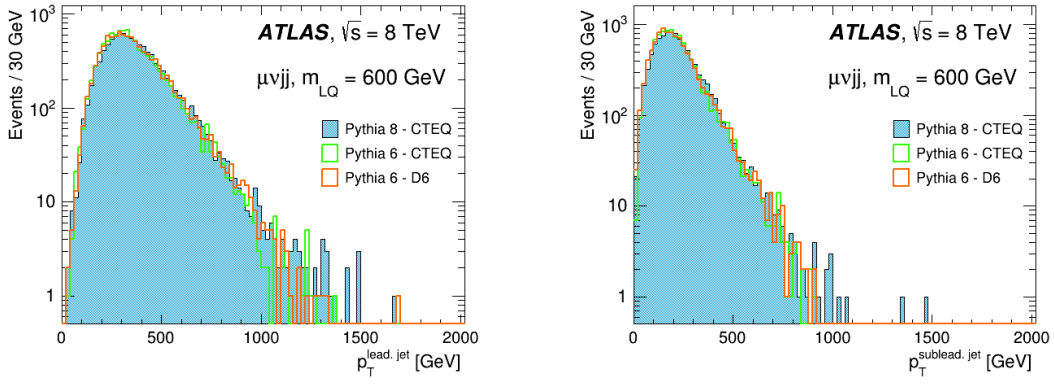
[2]ως γεγονός ορίζεται το σύνολο των πρωτογενών και δευτερογενών παραγόμενων σωματιδίων ως αποτέλεσμα της σύγκρουσης δύο πρωτονίων

[3]PDF (Parton Distribution Function): η συνάρτηση που καθορίζει την κατανομή της ορμής στα παρτόνια μέσα στα πρωτόνια



Σχήμα 1.1: Ενεργές διατομές παραγωγής του ζεύγους λεπτοκουάρκ υπολογισμένες σε πρώτο όρο διαταραχής και η στατιστική αβεβαιότητά τους.

έκδοση Pythia 8 - PDF CTEQ6L1. Ως παράδειγμα, στο Σχ. 1.2 παρουσιάζονται οι συγκρίσεις κατανομών των εγκάρσιων ορμών των δύο ανακατασκευασμένων πιδάκων με τις μεγαλύτερες εγκάρισες ορμές (p_T) κάθε γεγονότος για μάζα λεπτοκουάρκ ίση με 600 GeV.



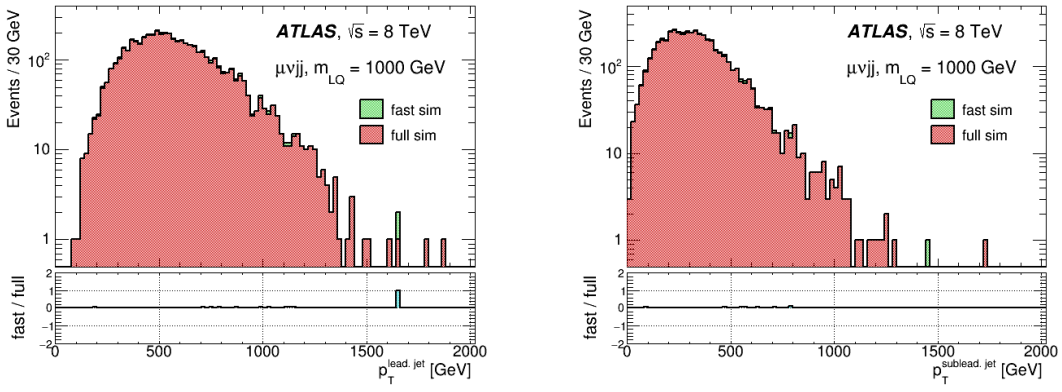
(a) (b)

Σχήμα 1.2: Εγκάρσιες ορμές (p_T) του πίδακα (α) με τη μεγαλύτερη ορμή και (β) με τη δεύτερη μεγαλύτερη ορμή του γεγονότος για μάζα λεπτοκουάρκ ίση με 600 GeV.

Μια άλλη μελέτη που έγινε, αφορούσε τη σύγκριση μεταξύ γρήγορης και πλήρους προσομοίωσης σήματος. Στο Σχ. 1.3 εμφανίζονται οι ίδιες μεταβλητές με πριν για μάζα λεπτοκουάρκ ίση με 1000 GeV, με τη διαφορά τώρα ότι η έκδοση της γεννήτριας είναι η ίδια και η σύγκριση έγκειται ανάμεσα στη γρήγορη και την πλήρη προσομοίωση.

Γίνεται προφανές ότι οι διαφορές είναι αμελητέες, κυρίως στην περιοχή ενδιαφέροντος σήματος για κάθε μάζα. Γι' αυτό, η παραγωγή των σημάτων έγινε με τη γρήγορη προσομοίωση για οικονομία χρόνου και υπολογιστικών πόρων.

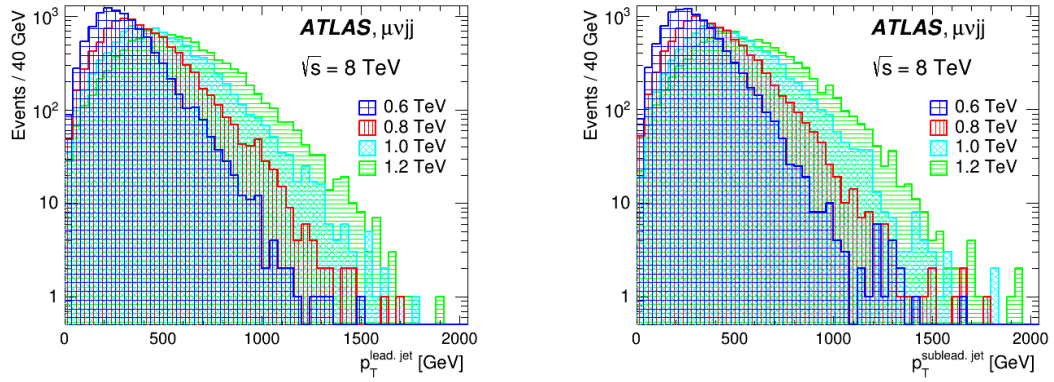
Το τελευταίο βήμα στη μελέτη του παραγόμενου σήματος είναι η επικύρωσή του. Στο Σχ. 1.4 εμφανίζονται ξανά οι δύο πίδακες του γεγονότος με τις μεγαλύτερες



(a) (b)

Σχήμα 1.3: Εγκάρσιες ορμές (p_T) του πίδακα (α) με τη μεγαλύτερη ορμή και (β) με τη δεύτερη μεγαλύτερη ορμή του γεγονότος για μάζα λεπτοκουάρκ ίση με 1000 GeV.

εγκάρσιες ορμές για τέσσερις διαφορετικές μάζες (600 GeV, 800 GeV, 1000 GeV και 1200 GeV αντίστοιχα).



(a) (b)

Σχήμα 1.4: Εγκάρσιες ορμές (p_T) του πίδακα (α) με τη μεγαλύτερη ορμή και (β) με τη δεύτερη μεγαλύτερη ορμή του γεγονότος για τέσσερις μάζες λεπτοκουάρκ (600 GeV, 800 GeV, 1000 GeV, 1200 GeV) που έχουν παραχθεί με γρήγορη προσομοίωση μέσω της γεννήτριας Pythia 8.165 - PDF CTEQ6L1.

Γίνεται φανερό ότι η συμπεριφορά των κατανομών είναι η αναμενόμενη με το μέγιστό τους να μειώνεται όσο μεγαλώνει η μάζα και, ταυτόχρονα, να αυξάνεται η εμβέλειά τους σε όλο και μεγαλύτερες ενέργειες. Γενικότερα, και οι υπόλοιπες κινηματικές μεταβλητές εμφανίζουν την αναμενόμενη συμπεριφορά. Τελικά το σήμα λεπτοκουάρκ που χρησιμοποιήθηκε σε αυτή την ανάλυση παρήχθη με τη μέθοδο της γρήγορης προσομοίωσης μέσω της γεννήτριας Pythia 8.165 και κάνοντας χρήση των CTEQ6L1 PDF.

1.2.3 Επιλογή φυσικών αντικειμένων και γεγονότων

Παρακάτω παρατίθενται τα κριτήρια επιλογής μιονίων, πιδάκων (φυσικά αντικείμενα) και γεγονότων συνολικά.

1.2.3.1 Μιόνια

- εγκάρσια ορμή (p_T) > 30 GeV
- ψευδοωκύτητα (η) < 2.5
- πρέπει να έχει καταγραφεί από τους ανιχνευτές B-layer, Pixel, SCT, Si, TRT
- εγκάρσια παράμετρος πρόσκρουσης (z_0) < 1.0 mm
- διαμήκης παράμετρος πρόσκρουσης (d_0) < 0.2 mm
- απομόνωση[4] (p_T^{cone20}/p_T^{μ}) < 0.2
- μιόνια τριών σταθμών (εσωτερικό, μέσο και εξωτερικό μιονικό φασματόμετρο)
- ταυτοποίηση του μιονικού σκανδαλιστή (εξακρίβωση ότι το συγκεκριμένο μιόνιο είναι αυτό που ενεργοποίησε τον μιονικό σκανδαλιστή)

1.2.3.2 Πίδακες

- ενέργεια του πίδακα > 0
- εγκάρσια ορμή (p_T) > 20 GeV
- ψευδοωκύτητα (η) < 2.8
- πηλίκο κορυφής (JVF) > 0.5
- εγκάρσια ορμή (p_T) > 30 GeV
- αποκλείονται γεγονότα με μιόνιο μέσα στη στερεά γωνία $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ μικρότερη από 0.4 με τον πίδακα με τη μεγαλύτερη εγκάρσια ορμή στο κέντρο της
- αποκλείονται γεγονότα με ηλεκτρόνιο μέσα στη στερεά γωνία ΔR μικρότερη από 0.4 με τον πίδακα με τη μεγαλύτερη εγκάρσια ορμή στο κέντρο της

1.2.3.3 Γεγονότα

- το γεγονός πρέπει να είναι στη λίστα με τις καλές περιόδους καταγραφής (GRL)
- η πρώτη κορυφή του γεγονότος πρέπει να έχει τουλάχιστον τρεις σχετιζόμενες τροχιές
- γεγονότα με κακούς πίδακες αποκλείονται
- γεγονότα με σφάλμα καταγραφής στο tile καλορίμετρο και στο καλορίμετρο Αργού αποκλείονται
- ανολοκλήρωτα γεγονότα αποκλείονται
- γεγονότα με πίδακες καταγεγραμμένους από προβληματικούς υπο-ανιχνευτές αποκλείονται
- εφαρμόζεται η λίστα με τους σκανδαλιστές

[4]ορίζεται ως το πηλίκο του αθροίσματος των εγκάρσιων ορμών όλων των τροχιών του γεγονότος που βρίσκονται μέσα στον αντίστοιχο κώνο της στερεάς γωνίας ΔR με κέντρο το μιόνιο προς την εγκάρσα ορμή του μιονίου

- το γεγονός πρέπει να περιέχει τουλάχιστον ένα "χαλαρό" μιόνιο
- το γεγονός πρέπει να περιέχει τουλάχιστον ένα μιόνιο σήματος ("μοναχικό") ή προερχόμενο από το καλορίμετρο
- ο μιονικός σκανδαλιστής πρέπει να είναι ο EF_mu18_MG_medium
- το καταγεγραμμένο μιόνιο πρέπει να είναι αυτό που προκάλεσε τον προαναφερόμενο μιονικό σκανδαλισμό
- το γεγονός πρέπει να περιέχει ακριβώς ένα μιόνιο και κανένα ηλεκτρόνιο
- το γεγονός πρέπει να περιέχει τουλάχιστον δύο πίδακες σήματος
- η εγκάρσια ελλείπουσα ενέργεια πρέπει να είναι μεγαλύτερη από 40 GeV
- η γωνία $\Delta\phi$ μεταξύ του πίδακα με τη μεγαλύτερη εγκάρσια ορμή και της εγκάρσιας ελλείπουσας ενέργειας στο επίπεδο πρέπει να είναι μεγαλύτερη από 0.5
- η γωνία $\Delta\phi$ μεταξύ του μιονίου με τη μεγαλύτερη εγκάρσια ορμή και της εγκάρσιας ελλείπουσας ενέργειας στο επίπεδο πρέπει να είναι μεγαλύτερη από 0.8
- η εγκάρσια μάζα μιονίου-ελλείπουσας εγκάρσιας ενέργειας (m_T) πρέπει να είναι μεγαλύτερη από 50 GeV

1.2.4 Υπόβαθρο ΚΧΔ

Το υπόβαθρο ΚΧΔ δεν είναι δυνατόν να παραχθεί με προσομοίωση προερχόμενη από κάποια γεννήτρια τυχαίων γεγονότων. Γι' αυτό υπολογίζεται έμμεσα από τα πραγματικά δεδομένα με τη μέθοδο ΑΒΓΔ.

Σύμφωνα με αυτή τη μέθοδο, επιλέγονται δύο μεταβλητές που θεωρούνται ανεξάρτητες, έτσι ώστε να δημιουργείται ένας φασικός χώρος. Οι δύο αυτές μεταβλητές είναι η ελλείπουσα εγκάρσια ενέργεια (E_T^{miss}) και η απομόνωση του μιονίου p_T^{cone20}/p_T^{μ}. Απαιτώντας να ισχύει ταυτόχρονα $\mathrm{E}_T^{miss} > 50$ GeV και $p_T^{cone20}/p_T^{\mu} < 0.2$ δημιουργείται μια περιοχή σήματος στον αντίστοιχο φασικό χώρο. Όλα τα γεγονότα που περνάνε αυτά τα κριτήρια θεωρούνται ως γεγονότα σήματος. Αντιστρέφοντας κάθε φορά μία από τις ανισώσεις (ή και τις δύο) δημιουργούνται οι υπόλοιπες τρεις περιοχές ελέγχου. Τα γεγονότα που ικανοποιούν αυτά τα κριτήρια θεωρούνται ως γεγονότα υποβάθρου. Έτσι δημιουργούνται οι παρακάτω περιοχές:

- $\mathrm{E}_T^{miss} > 50$ GeV, $p_T^{cone20}/p_T^{\mu} < 0.2$ (περιοχή σήματος Α)
- $\mathrm{E}_T^{miss} > 50$ GeV, $p_T^{cone20}/p_T^{\mu} > 0.2$ (περιοχή ελέγχου Β)
- $\mathrm{E}_T^{miss} < 50$ GeV, $p_T^{cone20}/p_T^{\mu} > 0.2$ (περιοχή ελέγχου Γ)
- $\mathrm{E}_T^{miss} < 50$ GeV, $p_T^{cone20}/p_T^{\mu} < 0.2$ (περιοχή ελέγχου Δ)

Οι αριθμοί γεγονότων από αυτές τις τέσσερις περιοχές συνδέονται με τη σχέση

$$N_A \cdot N_\Gamma = N_B \cdot N_\Delta \tag{1.1}$$

κι έτσι το πλήθος γεγονότων του υποβάθρου ΚΧΔ ισούται με

$$N_A = \frac{N_\Delta}{N_\Gamma} \cdot N_B \simeq 0.2411 \cdot N_B \tag{1.2}$$

Ο λόγος που ο υπολογισμός των γεγονότων της περιοχής σήματος γίνεται σε σχέση με την περιοχή ελέγχου Β είναι ότι αυτή είναι η πιο πολυπληθής σε σύγκριση με τις υπόλοιπες δύο περιοχές ελέγχου Γ και Δ, όπως και ότι περιέχει περισσότερα γεγονότα σήματος. Τα παραπάνω κριτήρια εφαρμόζονται στο τέλος της αλληλουχίας προεπιλογής. Το ποσοστό του υποβάθρου ΚΧΔ σε σχέση με το συνολικό τελικό υπόβαθρο είναι $\sim 17.4\%$. Τα αποτελέσματα παρουσιάζονται στον Πίν. 1.1

Περιοχή	Δεδομένα	Προσομοιωμένα υπόβαθρα	Υπόβαθρο ΚΧΔ
Περιοχή Σήματος Α ($E_T^{miss} > 50$ GeV, $p_T^{cone20}/p_T^{\mu} < 0.2$)	$1,865,890 \pm 12,376$	$1,396,160 \pm 11,055$	$294,969 \pm 4,490$
Περιοχή Ελέγχου Β ($E_T^{miss} > 50$ GeV, $p_T^{cone20}/p_T^{\mu} > 0.2$)	$1,236,849 \pm 10,242$	$13,419 \pm 1,233$	$1,223,430 \pm 10,162$
Περιοχή Ελέγχου Γ ($E_T^{miss} < 50$ GeV, $p_T^{cone20}/p_T^{\mu} > 0.2$)	$493,601 \pm 6,296$	$1,157 \pm 320$	$492,444 \pm 6,288$
Περιοχή Ελέγχου Δ ($E_T^{miss} < 50$ GeV, $p_T^{cone20}/p_T^{\mu} < 0.2$)	$314,172 \pm 5,093$	$195,443 \pm 4,234$	$118,729 \pm 2,923$

Πίνακας 1.1: Πλήθος γεγονότων ανά περιοχή πραγματικών δεδομένων, προσομοιωμένων υποβάθρων και του υποβάθρου ΚΧΔ στο τέλος της αλληλουχίας προ-επιλογής. Τα εμφανιζόμενα σφάλματα είναι μόνο στατιστικά.

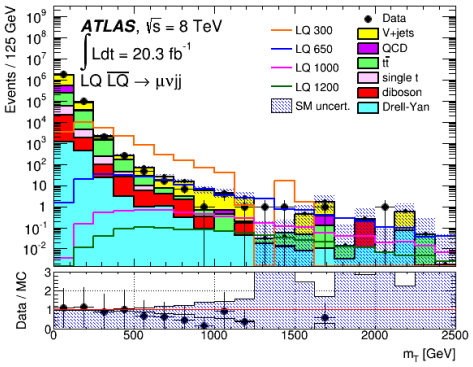
1.2.5 Μεταβλητές και εναπομείναν σήμα

Εκτός από τις βασικές κινηματικές μεταβλητές, όπως η εγκάρσια ορμή (p_T) και η αζιμουθιακή γωνία (ϕ), χρησιμοποιούνται κι άλλες. Κάποιες από αυτές αναφέρονται παρακάτω:

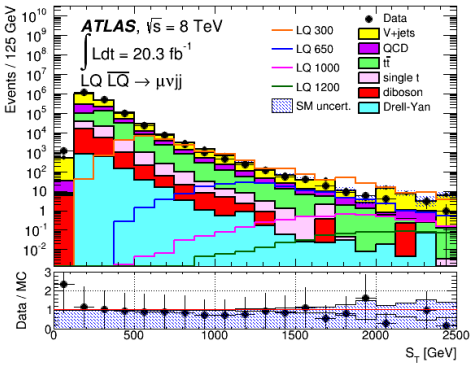
- $L_T = p_T^{\ell} + E_T^{miss}$, το βαθμωτό άθροισμα της εγκάρσιας ορμής του φορτισμένου λεπτονίου (μιονίου) και της ελλείπουσας εγκάρσιας ενέργειας
- $H_T = p_T^{jet_1} + p_T^{jet_2}$, το βαθμωτό άθροισμα της εγκάρσιας ορμής των δύο πιδάκων του γεγονότος με τη μεγαλύτερη εγκάρσια ορμή
- $S_T = L_T + H_T$, η συνολική βαθμωτή εγκάρσια ορμή
- N_{jets} (N_{muons}) = το πλήθος των πιδάκων (μιονίων) ανά γεγονός
- $m_T = \sqrt{2p_T^{\ell}E_T^{miss}[1 - cos\Delta\phi(\ell, E_T^{miss})]}$, η εγκάρσια μάζα μιονίου-ελλείπουσας εγκάρσιας ενέργειας. Αυτή η μεταβλητή χρησιμοποιείται για να διαχωρίσει τα γεγονότα στα οποία το μποζόνιο W διασπάται λεπτονικά με αυτά στα οποία διασπάται αδρονικά,
- οι δύο μάζες λεπτοκουάρκ m_{LQ_1} και m_{LQ_2}

Σε κάθε γεγονός, κάθε μία από τις μάζες λεπτοκουάρκ είναι ο συνδυασμός ενός από του δύο πίδακες με ένα από τα λεπτόνια της τελικής κατάστασης (μιόνιο και νετρίνο μιονίου). Σύμφωνα με τη θεωρία, τα δύο παραγόμενα λεπτοκουάρκ πρέπει να έχουν την ίδια μάζα. Ως εκ τούτου, σε κάθε γεγόνος γίνονται όλοι οι δυνατοί συνδυασμοί. Ο συνδυασμός που περιέχει το μιόνιο είναι η αναλλοίωτη μάζα λεπτοκουάρκ (m_{LQ}), ενώ αυτός με την ελλείπουσα εγκάρσια ενέργεια είναι η εγκάρσια μάζα λεπτοκουάρκ ($m^T_{LQ} = \sqrt{2p^j_T E^{miss}_T [1 - cos\Delta\phi(j, E^{miss}_T)]}$)[5]. Στο τέλος επιλέγεται αυτός ο συνδυασμός που έχει τη μικρότερη διαφορά μάζας, γεγονός το οποίο με τη σειρά του συνεπάγεται ότι, σε κάθε γεγονός, σε κάθε μάζα μπορεί να αντιστοιχεί η αναλλοίωτη ή η εγκάρσια μάζα λεπτοκουάρκ.

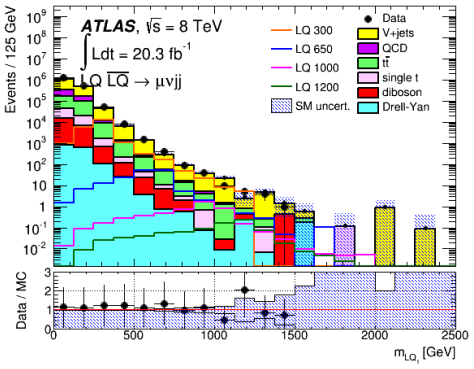
Στην παρούσα ανάλυση οι μεταβλητές που χρησιμοποιούνται για τον διαχωρισμό πιθανού σήματος και υποβάθρου είναι οι εξής: m_T, S_T και m_{LQ_1}. Στο Σχ. 1.5 φαίνονται οι κατανομές αυτών των τριών μεταβλητών στο τέλος της αλληλλουχίας προ-επιλογής σε συνδυασμό με τέσσερα προσομοιωμένα σήματα για τις μάζες 300 GeV, 650 GeV, 1000 GeV και 1200 GeV αντίστοιχα.



(a)



(b)



(c)

Σχήμα 1.5: (α) Εγκάρσια μάζα μιονίου-ελλείπουσας εγκάρσιας ενέργειας (m_T), (β) συνολική βαθμωτή εγκάρσια ορμή (S_T) και (γ) η μάζα λεπτοκουάρκ m_{LQ_1}. Οι κατανομές είναι μετά την εφαρμογή όλων των κριτηρίων της αλληλουχίας προ-επιλογής για συνολική φωτεινότητα 20.3 fb^{-1}.

[5] η μάζα m^T_{LQ} δεν πρέπει να συγχέεται με την εγκάρσια μάζα μιονίου-ελλείπουσας εγκάρσιας ενέργειας m_T

Από την αρχή μέχρι το τέλος της εφαρμογής των κριτηρίων προ-επιλογής, το σήμα που απομένει είναι περίπου σταθερό για όλες τις μάζες λεπτοκουάρκ και ίσο με $\sim$ 59%. Στον Πίν. 1.2 φαίνονται αναλυτικά τα αποτελέσματα για κάθε μάζα.

Μάζα λεπτοκουάρκ [GeV]	Αριθμός αναμενόμενων γεγονότων **πριν** την προ-επιλογή	Αριθμός αναμενόμενων γεγονότων **μετά** την προ-επιλογή	Εναπομείναν σήμα [%]
300	$37,367.5 \pm 378.5$	$21,655.1 \pm 288.1$	57.95
350	$14,852.4 \pm 152.3$	$8,641.5 \pm 116.2$	58.18
400	$6,850.4 \pm 68.9$	$4,031.2 \pm 52.9$	58.85
450	$3,216.5 \pm 32.6$	$1,914.9 \pm 25.2$	59.53
500	$1,611.3 \pm 16.4$	952.6 ± 12.6	59.12
550	833.7 ± 8.5	505.1 ± 6.6	60.59
600	466.3 ± 4.7	279.5 ± 3.7	59.94
650	262.0 ± 2.7	156.8 ± 2.1	59.85
700	150.0 ± 1.5	89.5 ± 1.2	59.67
750	88.7 ± 0.9	53.0 ± 0.7	59.75
800	52.0 ± 0.5	30.9 ± 0.4	59.42
850	32.3 ± 0.3	19.1 ± 0.3	59.13
900	19.7 ± 0.2	11.9 ± 0.2	60.41
950	12.4 ± 0.1	7.4 ± 0.1	59.68
1000	7.9 ± 0.1	4.8 ± 0.1	60.76
1050	5.1 ± 0.1	3.0 ± 0.0	58.82
1100	3.2 ± 0.0	1.9 ± 0.0	59.38
1150	2.1 ± 0.0	1.2 ± 0.0	57.14
1200	1.4 ± 0.0	0.8 ± 0.0	57.14

Πίνακας 1.2: Αριθμός των αναμενόμενων γεγονότων πριν και μετά την εφαρμογή της αλληλουχίας προ-επιλογής.

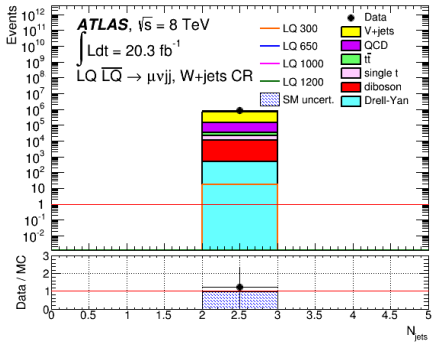
1.2.6 Περιοχές ελέγχου

Η κεντρική ιδέα πίσω από τις περιοχές ελέγχου (ΠΕ)[6] είναι η εφαρμογή συγκεκριμένων κριτηρίων έτσι ώστε να εξαληφθεί σχεδόν το σύνολο του σήματος και να παραμείνουν μόνο γεγονότα υποβάθρου με ένα κυρίαρχο σε κάθε περίπτωση. Με αυτόν τον τρόπο γίνεται εφικτό να ελεγχθεί ο βαθμός συμφωνίας των ήδη γνωστών διαδικασιών του ΚΠ με την ίδια τελική κατάσταση σε σχέση με τα πραγματικά δεδομένα εφόσον δεν αναμένεται η ύπαρξη κάποιου σήματος.

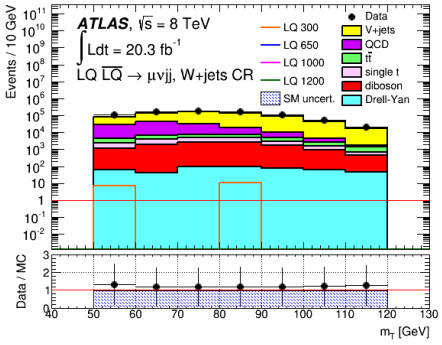
Σε αυτή την ανάλυση ορίζονται δύο περιοχές ελέγχου με αμελητέα επιμόλυνση σήματος για να ελεγχθεί η ακρίβεια των γεγονότων υποβάθρου. Στην πρώτη, το κύριο υπόβαθρο είναι το W+πίδακες, ενώ στη δεύτερη είναι το τοπ - αντι-τοπ κουάρκ. Παρακάτω ορίζονται αυτές οι δύο περιοχές:

- $N_{jets} = 2$, 40 GeV $\leq m_T \leq$ 120 GeV και $S_T <$ 225 GeV (Σχ. 1.6)

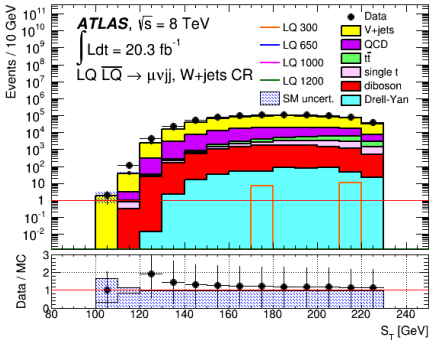
[6]δεν πρέπει να συγχέονται με τις ομώνυμες περιοχές στον υπολογισμό του υποβάθρου ΚΧΔ



(a)



(b)



(c)

Σχήμα 1.6: Περιοχή ελέγχου W+πίδακες.

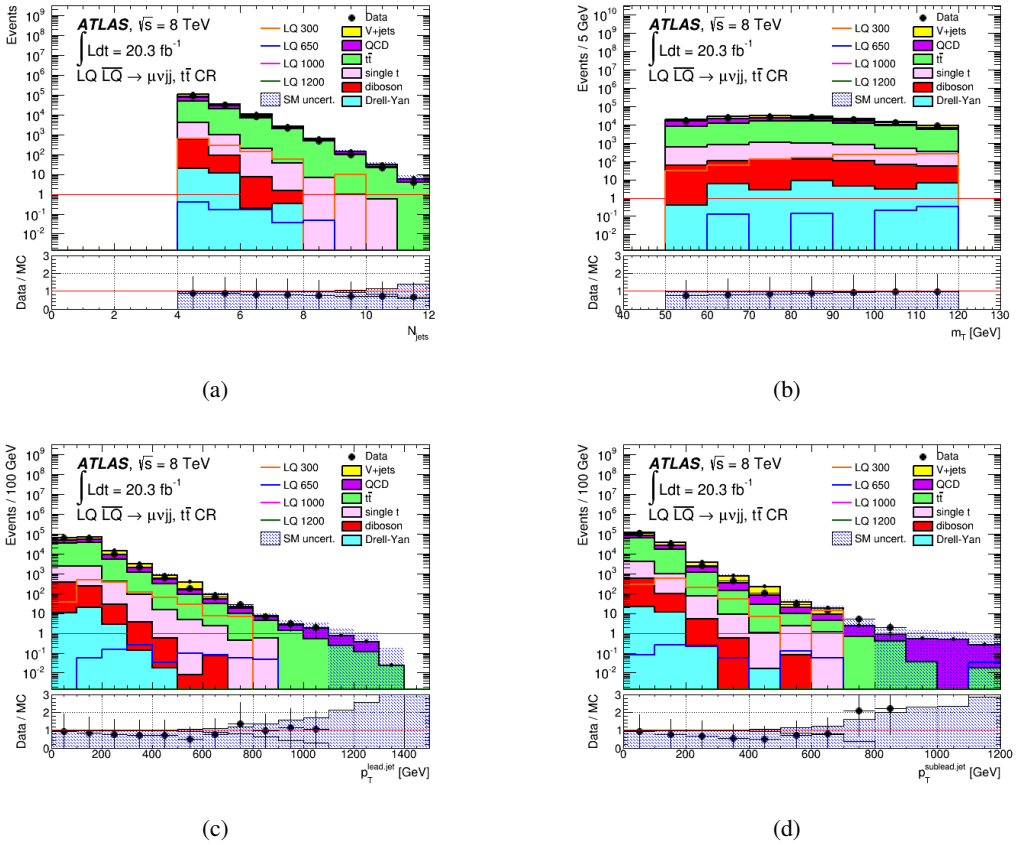
- $N_{jets} \geq 4$, 40 GeV $\leq m_T \leq$ 120 GeV, η εγκάρσια ορμή του πρώτου πίδακα $>$ 50 GeV και η εγκάρσια ορμή του δεύτερου πίδακα $>$ 40 GeV (Σχ. 1.7)

Γίνεται φανερό ότι στις ΠΕ, και κυρίως στην W+πίδακες, στην κατανομή της εγκάρσιας ορμής p_T, όπως και της συνολικής εγκάρσιας ορμής S_T του γεγονότος που περιέχει την πρώτη, τα πραγματικά δεδομένα δε συμφωνούν απολύτως με το συνολικό αναμενόμενο υπόβαθρο. Αυτό πιθανότατα συμβαίνει εξ' αιτίας κακής περιγραφής των πιδάκων από τις γεννήτριες που τα παρήγαγαν.

1.2.7 Περιοχές σήματος

1.2.7.1 Ορισμός

Σε αντιδιαστολή με τις ΠΕ, ορίζονται 19 περιοχές σήματος (ΠΣ), μία για κάθε μάζα από τα 300 GeV μέχρι τα 1200 GeV με βήμα 50 GeV. Αυτές είναι οι περιοχές στις οποίες το σήμα λεπτοκουάρκ έχει μεγάλη στατιστική εφαρμόζοντας συγκεκριμένα κριτήρια στις μεταβλητές m_T, S_T και m_{LQ_1}.



Σχήμα 1.7: Περιοχή ελέγχου τοπ - αντι-τοπ κουάρκ

1.2.7.2 Βελτιστοποίηση

Προκειμένου να μεγιστοποιηθεί η διάκριση μεταξύ σήματος και υποβάθρου γίνεται χρήση της εξίσωσης Z

$$Z = \frac{S}{\sqrt{B}} \tag{1.3}$$

Η βελτιστοποίηση των τιμών αυτών των τριών μεταβλητών γίνεται για δύο διαφορετικές υποθέσεις. Η πρώτη είναι ότι αυτές είναι μεταξύ τους ανεξάρτητες και η δεύτερη ότι είναι εξαρτημένες ελέγχοντας σε ποιον βαθμό ισχύει αυτό.

1.2.7.3 Υπόθεση ανεξάρτητων μεταβλητών

Σε αυτή την περίπτωση υπολογίζεται το πηλίκο του σήματος προς την τετραγωνική ρίζα του συνολικού υποβάθρου (Εξ. (1.3)) για κάθε μία μεταβλητή ανεξάρτητα θεωρώντας ότι μία διαφορετική επιλογή τιμής για κάποια από αυτές δεν επηρεάζει τις υπόλοιπες.

Στον Πίν. 1.3 εμφανίζονται οι βέλτιστες τιμές των τριών ανεξάρτητων μεταβλητών για τις οποίες η διάκριση σήματος υποβάθρου μεγιστοποιείται, καθώς και το ποσοστό της αντίστοιχης αποτελεσματικότητας (το ποσοστό του σήματος που επιβιώνει μετά την εφαρμογή των παραπάνω κριτηρίων).

1.2.7.4 Υπόθεση εξαρτημένων μεταβλητών

Για να μελετήσουμε τη συγκεκριμένη υπόθεση, κατασκευάζουμε ένα τρισδιάστατο ιστόγραμμα του οποίου κάθε άξονας αντιστοιχεί σε κάθε μία από τις τρεις μεταβλητές που χρησιμοποιούμε για τον ορισμό των Περιοχών Σήματος. Με αυτόν τον τρόπο μπορούμε εύκολα να διαπιστώσουμε το σημείο στο οποίο συμβαίνει η ταυτόχρονη μεγιστοποίηση του πηλίκου της Εξ. (1.3) και από τις συντεταγμένες του προκύπτουν οι βέλτιστες τιμές για κάθε μία από αυτές. Στο Σχ. 6.34 φαίνονται δύο διαφορετικές όψεις του ίδου τρισδιάστατου ιστογράμματος για την ΠΣ8 στα 650 GeV. Εύκολα διακρίνεται το κόκκινο σημείο στο οποίο έχουμε τη μεγιστοποίηση του σήματος έναντι της τετραγωνικής ρίζας του υποβάθρου ταυτόχρονα και για τις τρεις μεταβλητές.

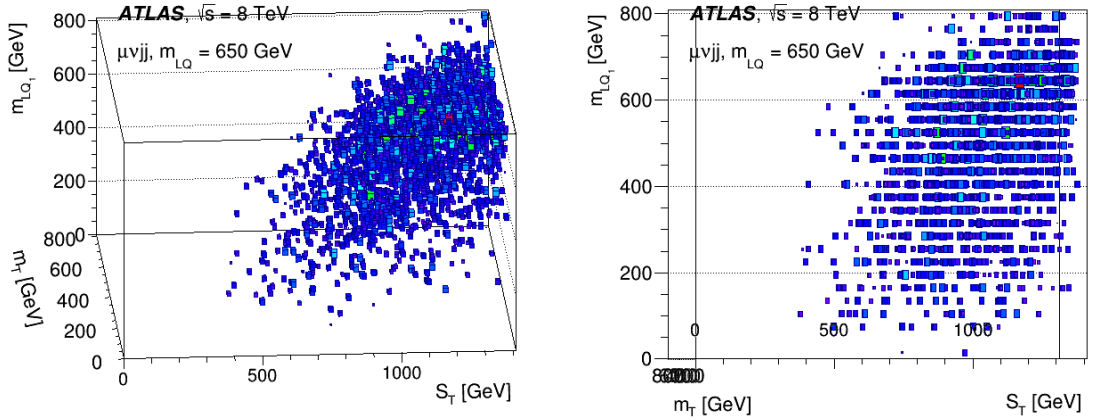
Στον Πίν. 1.4 εμφανίζονται οι βέλτιστες τιμές των τριών εξαρτημένων μεταβλητών για τις οποίες η διάκριση σήματος υποβάθρου μεγιστοποιείται, καθώς και το ποσοστό της αντίστοιχης αποτελεσματικότητας.

1.2.7.5 Εφαρμοζόμενη μέθοδος

Συγκρίνοντας τα αποτελέσματα των δύο προηγούμενων πινάκων γίνεται φανερό ότι οι τιμές των μεταβλητών S_T και m_{LQ_1} δεν αλλάζουν σημαντικά. Αντίθετα, οι τιμές της m_T παρουσίαζουν μεγαλύτερη απόκλιση ανάμεσα στις δύο υποθέσεις. Γενικά,

Μάζα [GeV] [ΠΣ]	$m_T >$ [GeV]	Z	Αποτ. (%)	$S_T >$ [GeV]	Z	Αποτ. [%]	$m_{LQ_1} >$ [GeV]	Z	Αποτ. [%]
300 [ΠΣ1]	340	118.50	21.3	565	50.65	60.4	260	40.17	63.3
350 [ΠΣ2]	355	70.21	29.6	655	30.27	59.1	310	21.97	60.0
400 [ΠΣ3]	365	40.46	35.1	735	18.68	62.5	355	13.65	59.1
450 [ΠΣ4]	385	22.89	38.0	795	11.41	67.6	410	8.31	52.7
500 [ΠΣ5]	465	14.19	33.2	895	7.19	64.7	440	5.41	58.5
550 [ΠΣ6]	465	8.64	38.7	1135	4.93	41.2	495	3.53	54.6
600 [ΠΣ7]	565	5.71	31.0	1135	3.40	52.3	535	2.51	56.2
650 [ΠΣ8]	565	3.78	37.2	1220	2.38	53.4	595	1.76	51.1
700 [ΠΣ9]	565	2.29	40.3	1310	1.66	53.2	645	1.29	49.5
750 [ΠΣ10]	565	1.50	45.6	1310	1.17	64.8	665	0.93	55.3
800 [ΠΣ11]	565	0.98	52.0	1390	0.77	64.5	665	0.62	64.9
850 [ΠΣ12]	565	0.61	53.9	1540	0.54	57.6	745	0.44	58.7
900 [ΠΣ13]	575	0.41	57.9	1540	0.38	66.4	790	0.30	56.9
950 [ΠΣ14]	770	0.27	40.8	1745	0.26	52.9	850	0.22	53.8
1000 [ΠΣ15]	880	0.19	33.9	1790	0.19	55.9	850	0.16	60.3
1050 [ΠΣ16]	880	0.13	36.9	1790	0.14	64.3	945	0.11	50.0
1100 [ΠΣ17]	880	0.08	39.1	1790	0.09	70.3	1050	0.08	38.0
1150 [ΠΣ18]	880	0.06	43.4	1790	0.06	75.0	1050	0.06	48.0
1200 [ΠΣ19]	880	0.04	46.6	2130	0.04	54.9	1115	0.05	43.3

Πίνακας 1.3: Οι περιοχές σήματος με τις βελτιστοποιημένες τιμές για τις μη σχετιζόμενες μεταβλητές.



(a) (b)

Σχήμα 1.8: Ταυτόχρονη βελτιστοποίηση των m_T, S_T και m_{LQ_1} από το τρισδιάστατο ιστόγραμμα της ΠΣ8 στα 650 GeV. Το κόκκινο σημείο εμφανίζεται στην πάνω δεξιά γωνία του ιστογράμματος.

ωστόσο, συνάγεται ότι οι και οι τρεις μεταβλητές δεν είναι εντελώς ανεξάρτητες, αν και η μεταξύ τους συσχέτιση δεν παρουσιάζεται ως ιδιαίτερα ισχυρή.

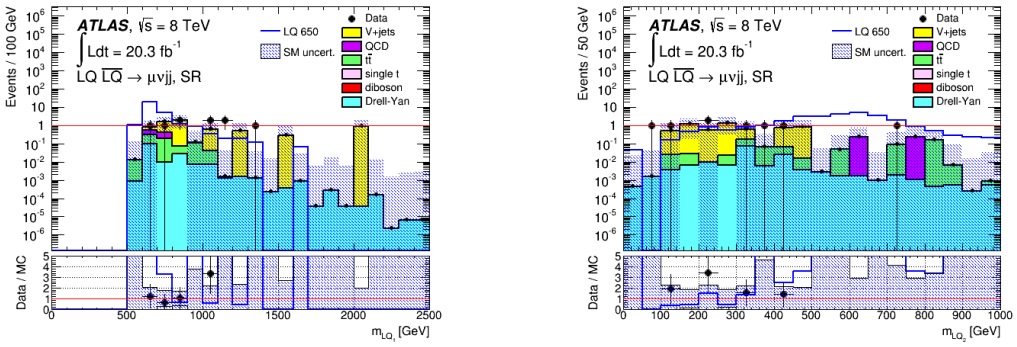
Η διάκριση σήματος-υποβάθρου που προκύπτει από τη δεύτερη υπόθεση είναι ανάλογη αυτής της πρώτης (S_T και m_{LQ_1}) ή μικρότερη (m_T). Αντίθετα, η αποτελεσματικότητα που προκύπτει από τις δύο υποθέσεις είναι σχεδόν η ίδια (S_T και m_{LQ_1}), αλλά αισθητά μεγαλύτερη για την m_T. Τελικά, θεωρήθηκε προτιμότερη η μεγιστι-

Μάζα [GeV] [ΠΣ]	$m_T >$ [GeV]	Z	Αποτ. (%)	$S_T >$ [GeV]	Z	Αποτ. [%]	$m_{LQ_1} >$ [GeV]	Z	Αποτ. [%]
300 [ΠΣ1]	180	78.00	65.5	570	53.06	59.0	270	40.01	58.0
350 [ΠΣ2]	150	25.21	84.9	690	29.73	53.2	330	20.04	47.6
400 [ΠΣ3]	240	27.58	64.7	780	18.15	52.4	360	13.53	57.0
450 [ΠΣ4]	270	16.18	62.9	840	11.37	58.8	420	8.12	48.5
500 [ΠΣ5]	210	5.99	83.1	1020	6.76	42.3	480	4.73	41.2
550 [ΠΣ6]	240	4.08	80.5	1170	4.74	36.9	540	2.67	31.8
600 [ΠΣ7]	300	3.47	74.1	1170	3.31	47.5	570	2.17	42.1
650 [ΠΣ8]	390	2.95	63.0	1200	2.29	55.8	630	1.50	36.3
700 [ΠΣ9]	300	1.18	81.4	1110	1.39	76.8	630	1.24	53.9
750 [ΠΣ10]	330	0.86	80.7	1440	1.09	48.2	690	0.86	48.3
800 [ΠΣ11]	420	0.70	70.7	1470	0.75	55.1	750	0.55	44.0
850 [ΠΣ12]	450	0.46	68.5	1500	0.53	61.9	840	0.28	28.6
900 [ΠΣ13]	570	0.41	58.2	1770	0.35	42.2	870	0.24	34.5
950 [ΠΣ14]	480	0.21	72.6	1890	0.22	38.4	870	0.20	48.1
1000 [ΠΣ15]	480	0.14	75.4	2010	0.13	34.2	930	0.14	44.7
1050 [ΠΣ16]	540	0.10	70.2	2040	0.10	40.8	990	0.09	39.1
1100 [ΠΣ17]	870	0.07	40.2	2070	0.06	45.6	1050	0.08	38.0
1150 [ΠΣ18]	1020	0.05	32.1	2040	0.05	55.5	1140	0.04	25.4
1200 [ΠΣ19]	900	0.04	44.7	2520	0.03	23.7	1360	0.01	11.6

Πίνακας 1.4: Οι περιοχές σήματος με τις βελτιστοποιημένες τιμές για τις σχετιζόμενες μεταβλητές.

ποίηση της διάκρισης σήματος-υποβάθρου σε βάρος μιας μερικής μεγαλύτερης απώλειας αποτελεσματικότητας σήματος. Για αυτόν τον λόγο οι τιμές που χρησιμοποιήθηκαν τελικά για τον ορισμό όλων των ΠΣ είναι αυτές που προκύπτουν από την υπόθεση των μη σχετιζόμενων μεταβλητών.

Στο Σχ. 1.9 φαίνονται οι κατανομές των μαζών του ζεύγους λεπτοκουάρκ στην ΠΣ8 που αντιστοιχεί σε μάζα ίση με 650 GeV. Δεν παρατηρείται κάποιο σημαντικό πλεόνασμα σήματος για κάποια από τις δύο κατανομές που να αποτελεί ένδειξη για κάτι περισσότερο από μία στατιστική διακύμανση. Στην περιοχή ευαισθησίας του συγκεκριμένου σήματος γύρω από τα 650 GeV δεν υπάρχει πλεόνασμα και όπου υπάρχει, αυτό συμβαίνει μακριά από τα 650 GeV και με πολύ μικρή στατιστική.



(a) (b)

Σχήμα 1.9: Οι κατανομές μαζών του ζεύγους λεπτοκουάρκ στην ΠΣ8 ονομαστικής μάζας ίση με 650 GeV ($m_T > 565$ GeV, $S_T > 1220$ GeV και $m_{LQ_1} > 595$ GeV) πριν τη στατιστική ανάλυση.

1.2.8 Στατιστική ανάλυση

Η καλύτερη μέθοδος διαχωρισμού σήματος λεπτοκουάρκ σε σχέση με το υπόβαθρο του ΚΠ είναι τα στατιστικά τεστ του πηλίκου πιθανότητας, όπως αυτή δίνεται στην Εξ. (1.4). Σε αυτή την εξίσωση το x συμβολίζει τα πραγματικά δεδομένα, το b το εκτιμώμενο υπόβαθρο και το s το προβλεπόμενο σήμα. Το Λ(x) είναι η συνάρτηση πιθανότητας (εδώ η συνάρτηση κατανομής πιθανότητας Poisson).

$$\Lambda(\mathbf{x}) = \frac{\frac{(s+b)^{\mathbf{x}} e^{-(s+b)}}{\mathbf{x}!}}{\frac{b^{\mathbf{x}} e^{-b}}{\mathbf{x}!}} \tag{1.4}$$

Στα πειράματα του LHC είναι σύνηθες να χρησιμοποιείται ο λογάριθμος του πηλίκου πιθανότητας (LLR - Logarithmic Likelihood Ratio) που φαίνεται στην Εξ. (1.5)

$$LLR(\mathbf{x}) = -2\ell og(\Lambda(\mathbf{x})) \tag{1.5}$$

Η συχνότητα της τιμής της ποσότητας LLR προσδιορίζεται μέσω ψευδοπειραμάτων τύχης. Τα επίπεδα εμπιστοσύνης (CL) ορίζονται ως το ολοκλήρωμα της κανονικοποιημένης κατανομής πιθανότητας των τιμών LLR με κάτω όριο την παρατηρούμενη τιμή LLR και άνω όριο το άπειρο, όπως φαίνεται και στις Εξ. (1.6) και Εξ. (1.7) για

την υπόθεση σήματος + υποβάθρου και μόνο υποβάθρου αντίστοιχα

$$CL_{s+b} = \int_{LLR(s+b|\mathbf{x})}^{\infty} \mathbf{P}(s+b|x')d(LLR(s+b|x') \tag{1.6}$$

$$CL_{b} = \int_{LLR(b|\mathbf{x})}^{\infty} \mathbf{P}(b|x')d(LLR(b|x') \tag{1.7}$$

Το τελικό επίπεδο εμπιστοσύνης ορίζεται ως το πηλίκο του επιπέδου εμπιστοσύνης σήματος + υποβάθρου προς το αντίστοιχο του υποβάθρου μόνο (Εξ. 1.8)

$$CL_{s} = \frac{CL_{s+b}}{CL_{b}} \tag{1.8}$$

Τα τελικά αποτελέσματα προκύπτουν μέσω του πακέτου HistFitter το οποίο ενσωματώνει τα πακέτα λογισμικού HistFactory και RooStats, τα οποία με τη σειρά τους βασίζονται στα RooFit και ROOT αντίστοιχα. Στον Πίν. 1.5 παρουσιάζονται τα τελικά αποτελέσματα μετά τη στατιστική ανάλυση για μία ΠΣ, συγκεκριμένα την ΠΣ8.

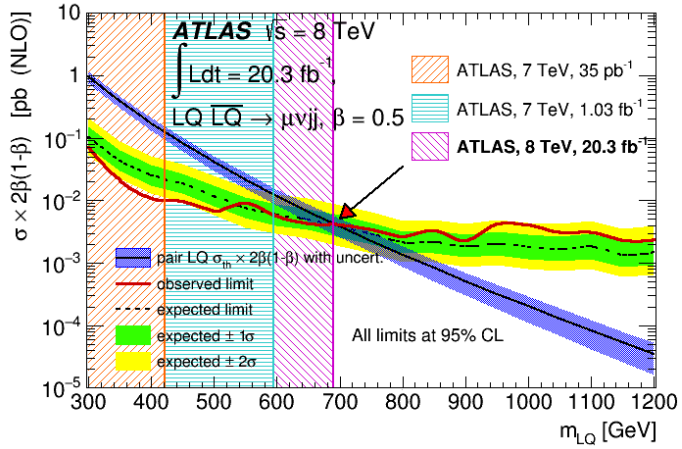
Αριθμός γεγονότων	ΠΣ8 [650 GeV]	W+πίδακες [ΠΕ]	Τοπ - αντι-τοπ κουάρκ [ΠΕ]
Παρατηρούμενα γεγονότα	22	1,205,164	221,094
Προσαρμοσμένα γεγονότα	$22.9^{+77.2}_{-22.9}$	1,205,174 ± 1,096	221,096.30 ± 472.10
Προσ. γεγ. W+πίδακες	15.0 ± 1.1	746,868 ± 9,662	53,799 ± 696
Προσ. γεγ. Z+πίδακες	2.5 ± 0.2	157,049 ± 3,500	13,052 ± 291
Προσ. διμποζονικά γεγ.	–	17,741 ± 395	1,068 ± 24
Προσ. μονό τοπ κουάρκ γεγ.	–	14,808 ± 330	7,043 ± 157
Προσ. γεγ. Drell-Yan	0.9 ± 0.1	2,011 ± 45	239 ± 5
Προσ. γεγ. ΚΧΔ	2.8 ± 0.2	245,453 ± 5,470	35,056 ± 781
Προσ. τοπ - αντι-τοπ κουάρκ γεγ.	1.6 ± 0.1	21,243 ± 139	110,839 ± 725
Προσ. γεγ. σήματος 650 GeV	0.1	1	3
Προσομοιωμένα αναμενόμενα γεγονότα	54.5	1,205,165	221,103
Αναμ. γεγ. W+πίδακες	15.0	746,808	53,794
Αναμ. γεγ. Z+πίδακες	2.5	157,068	13,054
Αναμ. διμποζονικά γεγ.	–	17,743	1,068
Αναμ. μονό τοπ κουάρκ γεγ.	–	14,810	7,044
Αναμ. γεγ. Drell-Yan	0.9	2,012	239
Αναμ. γεγ. ΚΧΔ	2.8	245,483	35,060
Αναμ. τοπ - αντι-τοπ κουάρκ γεγ.	1.6	21,243	110,835
Αναμ. γεγ. σήματος 650 GeV	31.7	–	8

Πίνακας 1.5: Περιοχή σήματος: αποτελέσματα προσαρμογής συνολικής φωτεινότητας 20.3 fb^{-1}. Τα αποτελέσματα προκύπτουν από τις περιοχές ελεγχου και την περιοχή σήματος στα 650 GeV χρησιμοποιώντας την προσαρμογή αποκλεισμού. Οι ονομαστικές προσομοιωμένες αναμενόμενες τιμές (κανονικοποιημένες στις αντίστοιχες ενεργές διατομές) δίνονται προς σύγκριση. Τα αναφερόμενα σφάλματα είναι τα στατιστικά και τα συστηματικά. Οι αβεβαιότητες των αποτελεσμάτων είναι εξ' ορισμού συμμετρικές. όπου τα αρνητικά σφάλματα κόβονται όταν φτάνουν στο μηδέν.

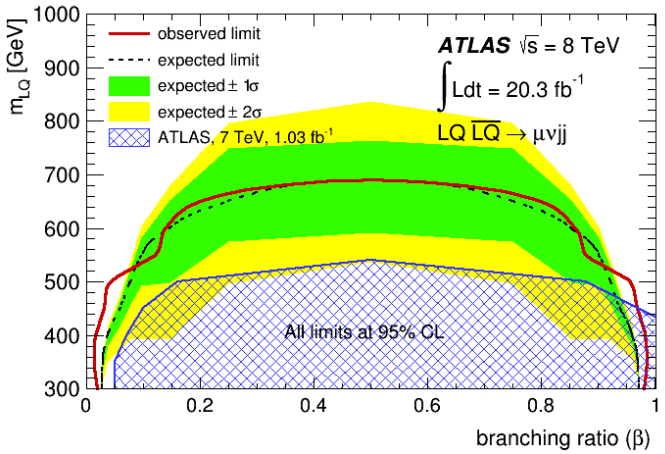
1.3 Αποτελέσματα και συμπεράσματα

Κατόπιν της στατιστικής αξιολόγησης, παρουσιάζονται τα αποτελέσματα των νέων εξαγόμενων ορίων για $\beta = 0.5$ (Σχ. 1.10). Όπως φαίνεται σε αυτό το ιστόγραμμα, δεν υπάρχει κάποια ένδειξη πιθανού σήματος λεπτοκουάρκ. Τα νέα όρια φτάνουν μέχρι την ενέργεια των ~ 690 GeV επεκτείνοντας το προηγούμενο όριο από παλιότερη ανάλυση κατά ~ 100 GeV. Το νέο αυτό όριο αντιστοιχεί σε ενεργό διατομή λεπτοκουάρκ ~ 0.008 pb. Αυτό συνεπάγεται ότι, αν τελικά υπάρχει, θα πρέπει να έχει μάζα μεγαλύτερη από 690 GeV και ενεργό διατομή παραγωγής μικρότερη από 8 fb.

Στο Σχ. 1.11 παρουσιάζεται η αποκλεισμένη μάζα λεπτοκουάρκ σε συνάρτηση με τα νέα όρια της τιμής του κλάσματος διάσπασης β.



Σχήμα 1.10: Όρια ενεργών διατομών υπολογισμένων σε δεύτερο όρο διαταραχής συναρτήσει της μάζας λεπτοκουάρκ (συγκριτικά αποτελέσματα της παρούσας με προηγούμενες αναλύσεις). Η πράσινη (κίτρινη) ζώνη αντιστοιχεί στο σύνολο της στατιστικής και συστηματικής αβεβαιότητας για απόκλιση $\pm 1(2)\ \sigma$. Οι αναμενόμενες ενεργές διατομές παραγωγής βαθμωτού ζεύγους λεπτοκουάρκ και η αντίστοιχη αβεβαιότητα εμφανίζονται με τη μπλε διαγράμμιση.



Σχήμα 1.11: Αποκλεισμένη μάζα ως συνάρτηση του κλάσματος διάσπασης β.

Chapter 2

Introduction

2.1 Ancient, helinistic and medieval periods

Scientific thinking has its roots in the greek world with prosocratic philosophers (apr. 800 BC - 400 BC). After all, the origin of the word ”physics” is derived from the greek word F'ush (fisi)[1]. Philosophers like Thales, Heraclitus, Zinon, Anaxagoras and Democritus were the first to set questions about the begining of Cosmos as an early Cosmology and formulate theories trying to explain why Nature appears to be the way it is. It is remarkable that Democritus gave shape to the first atomic[2] theory before 400 B.C., the main concept of which is very close to the modern one. Anaximenes of Miletus coined the first model for the origin of Cosmos theorizing that everything in the world is composed of air and considering celestial bodies as balls of fire in the sky (fire - earth - air). This was a radical cultural turn from theology and mythology to philosophy as the precursor of what was about to evolve into Science. They first conceived the idea that behind a natural process there is some explicable cause and not a divine act. These group of people are also called ”pro-scientists”, because they lacked the scientific method, but their way of thinking was very close to the modern basic characteristics of Science.

The death of Alexander the Great in 323 BC was succeeded by the rise of the new helinistic world. Mathematicians like Eratosthenes and Hypatia, engineers and inventors like Archimedes and Heron and astronomers like Hipparchus and Ptolemy pushed their fields to the next level of rigid scientific method through the application of mathematical proof and the observation of Nature. Them along with the Arabs of Middle Ages who followed like al-Khwarizmi[3] and al-Khayyami were going to be the paramount stone over which the european Renaissance of 14th century would flourish for the next three centuries.

2.2 The modern era of science

The modern scientific era begins in 1609, the year that Galileo invented the first telescope for the observation of the solar system. Introducing the use of scientific tools for observation and measurment beyond mathematical proof and prilosophic

[1] Nature in greek
[2] atom (άτομο in greek): something that cannot be divided
[3] the word "algorithm" is an altered form of his name

arguments would be a cornerstone in Natural Sciences. Galileo was one of the first in a line of giants of modern Science who followed his steps, like Newton and Huygens up to Maxwell and all the protagonists of the 20th century who, intentionally or not, revolutionized Physics with the theory of General Relativity and Quantum theory.

2.2.1 Brief history of elementary particle physics

In particle physics, an "elementary particle" is a particle that does not have any internal compositeness or, at least, it is considered not to have. In history of particle physics there have been many examples that particles which had been considered to be fundamental later were proved to be consisted of other "more fundamental" particles. A clasic example is that of the nucleons of atomic nuclei, proton and neutron, until the discovery that the parton hypothesis was correct.

2.2.1.1 Discovering the first particles

The start of particle physics could be tracked around 1895 with the discovery of X-rays by Wilhelm Röntgen which later would be identified as high energy photons, the particles of light. Two years after that, in 1897, a major breakthrough took place with the discovery of the electron by J.J. Thomson at the famous Cavendish laboratory. Another critical experiment for the first steps of elementary particle physics was held at the same place. That was the discovery of the alpha particles (helium nuclei) by Ernest Rutherford in 1899. Rutherford was also responsible for the next steps in 1911, the discovery of the atomic nucleus through scattering and the discovery of proton in 1919. It was proved that more of the 99% of the atomic mass was concentrated in a very confined space in the center of the atom and, therefore, the first atomic model close to experimental data was constructed.

2.2.1.2 First period of Quantum Mechanics

At the same time, the beginning of the 20th century, it had become obvious that the world below atomic level appeared to have some very bizarre and counter-intuitive properties. Subatomic particles had a totally unintuitive behavior. All known laws of Physics until the 19th century could not be applied in the case of elementary particles. For that, a radically new framework was needed in order new phenomena to be explained adequately. The first hint was the phenomenon of ultraviolet catastrophe. The Rayleigh - Jeans law, an approximation to the spectral radiance of electromagnetic radiation as a function of wavelength from a black body at a given temperature through classical arguments, was failing at high (ultraviolet) frequencies. The solution came from Max Planck who in 1900 proposed that the problem could be solved by the assumption that the thermal electromagnetic radiation spectrum of a black body is not continuous, but comes to very small discrete energy packages, the quanta[4]. Each of these quanta comes with specific energy which is given by the formula

$$\varepsilon = hf \tag{2.1}$$

where h is Planck 's constant (h = $6.626 \cdot 10^{-34}$ J·s) and f the frequency of the specific package. From that point onwards, the concept and the principles of Quantum Mechanics (QM) had prevailed into microcosm. Until the mid '20s, the first

[4]quantum (plural "quanta" in latin): amount or package

period of QM had been concluded passing on the torch to its possible (many) interpretations with the Copenhagen interpretation being the most dominant up to this day, although there is an ongoing debate for the final conclusions. Many great physicists made significant contributions, like Werner Heisenberg (*matrix mechanics, uncertainty principle*), Erwin Schrödinger (*wave mechanics*), Wolfgang Pauli (*Pauli exclusion principle*), Paul Dirac (*theoretical prediction of anti-matter*), Niels Bohr (*Bohr atomic model*), Max Born (*probability dension function of Schrödinger 's equation*), Luis de Broglie (*material waves*) etc. Even Albert Einstein contributed with his photoelectric effect, despite his later strong opposition to QM because of its probabilistic interpetation and quantum leaps.

The mid of 19th century was the threshold for another significant revolution in the core of Physics in general and James Clerk Maxwell with his electromagnetic theory was the pioneer. Until then, the scientific method was first the experimental observation and then the explanation of new physics through new possible theories in a phenomenological approach. We should not forget after all that the end of the 19th century was naively believed by many physicists to be the end of new physics. It was a strong belief among physicists back then that all the physical laws had already been discovered and the only remaining task was increasing the decimal numbers of the results and, thus, the accuracy. Following Maxwell, mainly Max Planck with his quanta in 1900 and more revolutionary Einstein initially with his Special Theory of Relativity [1] in 1905 (his *annus mirabilis*[5]) managed to reverse the procedure. Since then until today theoretical physicists coin new theories based on pure mathematics and experimentalists follow to confirm or cancel them through experiments. The breakthroughs of elementary particles physics during the entire 20th and 21st century up to our days have been based on this approach using QM and Special Relativity as main tools. The latter is a fundamental theory for elementary particles given the high (relativistic) velocity developing in accelerators and, thus, the energy they acquire through γ Lorentz factor

$$E = mc^2 = \gamma m_0 c^2 = \frac{m_0 c^2}{\sqrt{1-\beta^2}} = \frac{m_0 c^2}{\sqrt{1-(\frac{v}{c})^2}} \tag{2.2}$$

where m_0 is the particle rest mass, v its velocity and c the absolute barrier of speed of light. In case of a proton in the LHC 6.5 TeV beam acquiring a final speed of 0.999999989c, a γ Lorentz factor of $\sim$ 6742 is derived from Eq. (2.2). This implies that a proton with a rest mass of 938 MeV being accelerated up to that velocity reaches a final mass of $\sim$ 6324 GeV, the total rest mass of $\sim$ 33 lead (Pb) atoms.

2.2.1.3 Second period of Quantum Mechanics

During the second period of QM after 1925, the main interest was shifted from the entity of particle to that of quantum field. The result of theoretical phycisists like Dirac, Hans Bethe, Richard Feynman, Victor Weisskopf, Julian Schwinger, Freeman Dyson, Sin-Itiro Tomonaga and Ernst Jordan was the Quantum Field Theory (QFT) approach. Quantum ElectroDynamics (QED) was the "jewel of Physics" according to Feynman for which he along with Schwinger and Tomonaga were awarded the 1965 Nobel prize in Physics. QED is still the physical theory with the most accurate

[5] miraculous year in latin

predictions of quantities like the electron anomalous magnetic moment and hydrogen 's energy levels Lamb shift. Another QFT success was Quantum ChromoDynamics (QCD) development for the description of strong interaction among quarks and gluons. The theory as we know it today was formulated by David Politzer, David Gross and Frank Wilczek in 1975 (2004 Nobel prize in Physics). The next significant step was the elementary particle physics Standard Model development which is being analysed thoroughly in Chapter 3. The final success of this long sequence was the discovery of the so-called Higgs boson [2] which emerged as the last missing piece of Standard Model rooting back to 1965 and the work of Peter Higgs, Francois Englert, Robert Brout, Gerald Guralnik, Carl Hagen and Tom Kibble. Higgs and Englert have been awarded the 2013 Nobel prize in Physics.

2.2.1.4 Particle accelerating machines

Finally, there is a last topic it should be discussed. During the first era of elementary particle physics (until $\sim$ the 1920s) the research had been conducted only through the study of cosmic rays. These consist of high velocity and energy charged particles coming from space. Physicists of the time were using the products of the stars as their laboratory. The research was held through prompt balloons sent up in the atmosphere in order to measure cosmic rays at different altitudes and latitudes (e.g. Arthur Compton). The emerging problem was that the energy of cosmic rays could reach up to a certain point, so it became obvious the necessity for a new way of research. The solution was the accelerator machines which could accelerate and guide charged particles. Below are listed the different kinds of these constructions.

Early accelerators

- **Cyclotrons**: The particles are held to a spiral trajectory by a static magnetic field and accelerated by a rapidly varying (radio frequency) electric field. Lawrence was awarded the 1939 Nobel prize in Physics for this invention.
- **Other early accelerator types**: Linear accelerators (Linac) is a type of particle accelerator that greatly increases the kinetic energy of charged subatomic particles or ions by subjecting the charged particles to a series of oscillating electric potentials along a linear beamline. On the other hand, Betatron is a type of cyclic particle accelerator. It is essentially a transformer with a torus-shaped vacuum tube as its secondary coil. An alternating current in the primary coils accelerates electrons in the vacuum around a circular path. The betatron was the first machine capable of producing electron beams at energies higher than could be achieved with a simple electron gun.
- **Synchrotrons**: It is a particular type of cyclic particle accelerator, descended from the cyclotron, in which the accelerating particle beam travels around a fixed closed-loop path. The magnetic field which bends the particle beam into its closed path increases with time during the accelerating process, being synchronized to the increasing kinetic energy of the particles. The synchrotron is one of the first accelerator concepts to enable the construction of large-scale facilities, since bending, beam focusing and acceleration can be separated into different components. The most powerful modern particle accelerators use versions of the synchrotron design. The largest synchrotron-type accelerator, also the largest particle accelerator in the world, is the LHC.

Modern accelerators

- **High intensity hadron accelerators (Meson and neutron sources)**: Two examples are the J-PARC (Japan Proton Accelerator Research Complex) and ISIS neutron source. The first is a high intensity proton accelerator facility and aims for the frontier in materials and life sciences, and nuclear and particle physics. It uses high intensity proton beams to create high intensity secondary beams of neutrons, hadrons, and neutrinos. The second is a pulsed neutron and muon source and uses the techniques of muon spectroscopy and neutron scattering to probe the structure and dynamics of condensed matter on a microscopic scale ranging from the subatomic to the macromolecular. At ISIS the neutrons are created by accelerating bunches of protons in a synchrotron, then colliding these with a heavy tungsten metal target, under a constant cooling load to dissipate the heat from the 160 kW proton beam. The impacts cause neutrons to spall off the tungsten atoms and the neutrons are channelled through guides, or beamlines, to around 20 instruments, each individually optimised for the study of different types of interactions between the neutron beam and matter. The target station and most of the instruments are set in a large hall. Neutrons are a dangerous form of radiation, so the target and beamlines are heavily shielded with concrete.

- **Electron and low intensity hadron accelerators**: SLAC (Stanford Linear Accelerator Center) Linac was a radiofrequency linear accelerator that accelerated electrons and positrons up to 50 GeV. Tevatron was a circular particle accelerator (now inactive since 2011) and holds the title of the second highest energy particle collider in the world, after the LHC. It was a synchrotron that accelerated protons and antiprotons in a 6.86 km ring to energies of up to 1 TeV, hence its name.

  ILC (International Linear Collider) is a possible future linear collider foe electrons and positrons with a length between 30 and 50 km and an initial collision energy of 500 GeV up to 1000 GeV after the upgrade.

  Another accelerator of this category is CERN Proton Synchrotron (PS). It usually accelerates either protons delivered by the Proton Synchrotron Booster (PSB) or heavy ions from the Low Energy Ion Ring (LEIR). In the course of its history it has juggled many different kinds of particles, feeding them directly to experiments or to more powerful accelerators. With a circumference of 628 metres, PS has 277 conventional (room-temperature) electromagnets, including 100 dipoles to bend the beams round the ring. The accelerator operates at up to 25 GeV. In addition to protons, it has accelerated electrons, positrons, antiprotons, alpha particles (helium nuclei), oxygen and sulphur nuclei.

Colliders

- **Electron - positron colliders**: The most known electron - positron collider was LEP (Large Electron Positron). LEP collided electrons with positrons at energies that reached 209 GeV. It was a circular collider with a circumference of 27 kilometres built in a tunnel roughly 100m underground and passing through Switzerland and France. LEP was used from 1989 until 2000. Around 2001 it was dismantled to make way for the LHC which re-used the LEP tunnel. To date, LEP is the most powerful accelerator of leptons ever built.

- **Electron - proton colliders**: HERA (Hadron - Electron Ring Accelerator) was a particle accelerator at DESY in Hamburg. It began operating in 1992. At HERA, electrons or positrons were collided with protons at a center of mass energy of 318 GeV. It was the only lepton - proton collider in the world while operating. Also, it was on the energy frontier in certain regions of the kinematic range. HERA was closed down on 30 June 2007.

  LHeC (Large Hadron electron Collider) is another project in development. This collider woulb possibly be the cleanest high resolution microscope of the world, a device which would be dedicated to Higgs physics, searches for new phenomena and high precision electroweak and QCD physics, such as the high density states of matter at low relative parton momenta x.

- **Hadron colliders**: Besides LHC, another important hadron collider is the SPS (Super Proton Synchrotron) at CERN. It has been used to accelerate protons and antiprotons, electrons and positrons (for use as the injector for LEP) and heavy ions. From 1981 to 1991, the SPS operated as a hadron (more precisely, proton - antiproton) collider (as such it was called $Sp\overline{p}S$), when its beams provided the data for the UA1 and UA2 experiments, which resulted in the discovery of the W and Z bosons. These discoveries and a new technique for cooling particles led to the 1984 Nobel prize in Physics for Carlo Rubbia and Simon van der Meer.

And, finally, here is a brief reference of some machines for the very far and distanced future.

Hypothetical accelerators

- **Eloisatron**: Eurasiatic Long Intersecting Storage Accelerator was a project of INFN headed by Antonio Zichichi at the Ettore Majorana Foundation and Centre for Scientific Culture in Erice, Sicily. The center-of-mass energy was planned to be 200 TeV and the size was planned to span parts of Europe and Asia.

- **Fermitron**: This was an accelerator sketched by Enrico Fermi on a notepad in the 1940s proposing an accelerator in stable orbit around the Earth.

- **Undulator radiation collider**: It is a design for an accelerator with a center-of-mass energy around the GUT scale ($\sim 10^{13}$ TeV, almost a trillion times more powerful than LHC of 2017 at 13 TeV). It would be light-weeks across and require the construction of a Dyson swarm around the Sun.

- **Planckatron**: An accelerator with a center-of-mass energy of the order of the Planck scale ($\sim 10^{16}$ TeV, almost a thousand trillion times more powerful than LHC at 13 TeV). This is the energy scale for the straight forward discovery of the hypothetical strings of string theories. It is estimated that the radius of the Planckatron would have to be roughly the radius of the Milky Way. It would require so much energy to run that it could only be built by at least a Kardashev[6] Type II civilization [3].

[6]Nikolai Semenovich Kardashev (born April 25, 1932 in Moscow, Soviet Union) is a Russian astrophysicist

Chapter 3

Standard Model of the Elementary Particles

The Standard Model (SM) [4,8–13] ("Standard Theory" would be more accurate according to Iliopoulos, given its up to now establishment and strong experimental verification) of elementary particle physics is the theoretical framework which describes in a unique mathematical formulation the building blocks or matter (first generation fermions, Table 3.1) and their interaction via the exchange of mediators (gauge bosons, Table 3.2) that define the dynamics of the system. Each of these particles has its relevant anti-particle with all quantum numbers identical and opposite electric charge subjected to the same interaction. SM is a gauge quantum field, where gauge has the meaning that the Lagrangian is invariant under certain Lie groups of local transformations. The first step towards the SM was Sheldon Glashow 's discovery in 1961 of a way to combine the electromagnetic and weak interactions. Steven Weinberg and Abdus Salam incorporated in 1967 the Higgs mechanism into Glashow's electroweak interaction, giving it its modern form. Glashow, Salam and Weinberg have been awarded the 1979 Nobel prize in Physics for their work.

3.1 Physics behind Standard Model

In terms of Quantum Field Theory the particles are described as quanta of the field, while the interactions emerging through the principle of the Lagrangian invariance under local gauge transformations of a symmetry group. An easy way to understand the concept of a field in QFT is the following. You can imagine the field as a pot full of water. As long as the water is cold, nothing happens. As its temperature starts increasing, bubbles keep emerging and immediately dissapearing randomly. We could say that quantum field is like the boiling water and the particles (its excitations) are the random rising bubbles.

In Table 3.2 appear the mediators of three forces. The hypothetical mediator boson of gravity, graviton, is not included since it still remains theoretical. After all, gravity is the weakest force of all known four and, because of that, its contribution in quantum level is totally negligible[1]. On the other hand, the newly discovered Higgs boson (2012) has its place in that boson table, even if it is not a mediator of any kind of force. This particle completes the SM puzzle as its last missing piece. Higgs boson is the excitation of the Higgs field and it has been identified as the

[1] gravity 's still theoretical mediator boson is the massless graviton (G) and it is the only with spin equal to 2; instead, γ, $W^{\pm}$, Z^0 and gluons have spin 1

massive remaining term after spontaneous breaking of Higgs Lagrangian symmetry takes place. Through this mechanism all massive particles acquire their mass. The difference between scalar and gauge (or vector) bosons is that the first have spin equal to 0 and the latter have spin equal to 1. Bosons have integer spin and obey to Bose - Einstein statistics. Instead, fermions' spin is half-integer and they follow Fermi - Dirac statistics.

	Leptons			**Quarks**		
Fermion generation	Flavor	charge [e]	Mass [GeV]	Flavor	charge [e]	Mass [GeV]
1st	$\begin{pmatrix}\nu_e\\ e\end{pmatrix}_L, e_R$	$\begin{pmatrix}0\\ -1\end{pmatrix}, -1$	$\begin{pmatrix}< 3\cdot 10^{-9}\\ 511\cdot 10^{-6}\end{pmatrix}$	$\begin{pmatrix}u(up)\\ d(down)\end{pmatrix}_{L,R}$	$\begin{pmatrix}2/3\\ -1/3\end{pmatrix}$	$\begin{pmatrix}2.3\cdot 10^{-3}\\ 4.8\cdot 10^{-3}\end{pmatrix}$
2nd	$\begin{pmatrix}\nu_\mu\\ \mu\end{pmatrix}_L, \mu_R$	$\begin{pmatrix}0\\ -1\end{pmatrix}, -1$	$\begin{pmatrix}< 0.19\cdot 10^{-3}\\ 113.4\cdot 10^{-3}\end{pmatrix}$	$\begin{pmatrix}c(charm)\\ s(strange)\end{pmatrix}_{L,R}$	$\begin{pmatrix}2/3\\ -1/3\end{pmatrix}$	$\begin{pmatrix}1.275\\ 0.095\end{pmatrix}$
3rd	$\begin{pmatrix}\nu_\tau\\ \tau\end{pmatrix}_L, \tau_R$	$\begin{pmatrix}0\\ -1\end{pmatrix}, -1$	$\begin{pmatrix}< 18\cdot 10^{-3}\\ 1.777\end{pmatrix}$	$\begin{pmatrix}t(top)\\ b(bottom)\end{pmatrix}_{L,R}$	$\begin{pmatrix}2/3\\ -1/3\end{pmatrix}$	$\begin{pmatrix}173.21\\ 4.6\end{pmatrix}$

Table 3.1: Summary of basic properties of elementary fermions in Standard Model.The first generation quarks and leptons appearing in red are the building blocks of matter.

Interaction	Gauge bosons	Scalar bosons	Charge [e]	Mass [GeV]	Range [m]
Electromagnetic	γ (photon)	—	0	0	∞
Weak	$W^{\pm}$	—	± 1	80.385 ± 0.015	10^{-18}
	Z^0	—	0	91.1876 ± 0.0021	
Strong	g (gluon)	—	0	0	$\leq 10^{-15}$
—	—	H^0 (Higgs)	0	124.98 ± 0.28	—

Table 3.2: Summary of basic properties of elementary bosons - interactions mediators in Standard Model.

The indexes L and R denote the left and right fermions chirality respectively. Chirality should not be confused with helicity. The helicity of a particle is right-handed (positive) if the direction of its spin is the same as the direction of its motion and left-handed (negative) if the direction of its spin is the opposite to that of the direction of its motion. The chirality of a particle is a more abstract concept. It is determined by whether the particle transforms in a right- or left-handed representation of the Poincaré group. The massive electron, muon and tau, as well as all the quarks, can have both chiralities. On the other hand, the massless neutrinos in SM appear only with left chirality[2] since they move with the speed of

[2]in the case of massless fermions, chirality and helicity are identical

light and only one reference frame can exist. Of course, this is the case if their mass is indeed absolutely zero. A left-handed fermion and its right-handed partner have exactly the same mass. Particles and anti-particles (e.g. electrons and positrons) have the opposite chirality.

A synoptic form of SM Lagrangian making use of the Dirac notation is the following

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\overline{\psi}\not{D}\psi + \psi_i y_{ij}\psi_j\phi + h.c. + |D_\mu\phi|^2 - V(\Phi) \tag{3.1}$$

The first term expresses the forces involved in elementary particles world: the electromagnetic, weak and strong interactions. The second term stands for the way these three forces act on matter particles, quarks and leptons. The third describes how quarks and leptons obtain their mass through Higgs boson and the last two describe the Higgs mechanism.

In the elementary particle theory, and consequently in SM, Group Theory is of great importance. U(1), U(2), SU(2) and SU(3) are the main groups which describe elementary particles and their properties.

3.2 Little about Group Theory

U(1) group consists of unitary matrices set of order 1. The matrices of U(1) group are hermitian ($\theta = \theta^+$) because they are unitary, so they can be written in the form of $U = e^{i\theta}$. In general, the number of the generators of a U(n) group is is n^2, so in this case there is only one hermitian unitary matrix of order 1, for instance a real constant θ. In other words, the U(1) group intoduces a phase (a "gauge") $e^{i\theta}$.

In correspondence with U(1), the U(2) group consists of the set of hermitian unitary matrices of order 2. Given that $U(2) = SU(2)\times U(1)$, the four generators are the three generators of SU(2) group which are presented below and $e^{i\theta}$ of U(1).

SU(2) group is a subgroup of U(2) given that it consists of hermitian unitary matrices 2×2 and additionally their determinant must be $\det(\sigma_i) = +1$. This is the reason why SU(2) generators are $n^2\text{-}1 = 2^2\text{-}1 = 3$. Like all SU(n) groups, it is traceless ($\mathrm{Tr} = \sum \alpha_{km} = 0$, where $0 \leq k = m \leq n$). The most common base consists of the three Pauli matrices σ_i

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Finally, SU(3) group consists of all hermitian unitary 3×3 matrices with a determinant $\det(\lambda^i) = +1$. Additionally, now the generators are $n^2\text{-}1 = 3^2\text{-}1 = 8$. Its most common base are the eight Gell-Mann matrices λ^i

$$\lambda^1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda^2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda^4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

$$\lambda^5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \lambda^6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \lambda^7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix},$$
$$\lambda^8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

SM is described by $SU(3)_C \times SU(2)_L \times U(1)_Y$ symmetry. The $SU(3)_C$ group, where the index C stands for the QCD number of color, has the Gell-Mann generator matrices which correspond to the 8 massles $G_\mu^{1,2...8}$ gluons of strong interaction. This is responsible for binding together protons and neutrons in an atom nucleus and confining quarks into protons and neutrons. Likewise, the $SU(2)_L$ group consists of the three Pauli matrices which are been identified as the $W^\pm$ and Z^0 massive gauge bosons - carriers of the weak interaction - only involved with and explaining the phenomenon of radioactive nuclei decay. The L index denotes the left chirality. Finally the single generator of $U(1)_Y$ group corresponds to the massless photon of electromagnetism, the interaction essentially explaining every physical phenomenon outside the range of a nucleus (~ 1 fm $= 10^{-15}$ m) and gravity. The index Y stands for the electroweak hypercharge Y_W[3]

$$Y_W = 2(Q - T_3) \tag{3.2}$$

where Q is the electric charge and T_3 stands for the weak isospin. The reason a lizzard remains stuck to a vertical wall against gravity is due to the fact that protons and electrons in the atoms between them are been attracted through constant exchange of photons - they are known as Van der Walls forces, very weak electromagnetic interaction, but still stronger than gravity. The same thing happens when a magnet stays on the fridge instead of dropping down under its own weight neutrilizing the attraction of the entire Earth, only in that case we observe the other side of the coin - the magnetic force.

The combination of $SU(2)_L \times U(1)_Y$ stands for the unification of weak interaction and electromagnetism, the so called electroweak interaction. Weak hypercharge Y_W of electroweak interaction is the equivalent of Q for electromagnetic, T_3 for weak and color for strong interaction respectively.

3.3 Is Standard Model the final answer?

It is a fact that during the last decades SM has been a very succesful and self-consistent mathematical framework. The list of its successful predictions is rather long. Nevertheless, it cannot be ignored the fact that this framework has its limitations. Below there is a brief reference to the answers SM managed to provide so far, along with the most significant and still open issues, as well as their possible solutions.

[3] not to be confused with Y of strong interaction; Y = S + C + B' + T + B where S stands for strangeness, C for charm, B' for bottomness, T for topness and B for the baryon number

3.3.1 Successes

The first experimentally verified prediction back in 1973, in a neutrino experiment in the Gargamelle bubble chamber at CERN, was the neutral weak currents caused by Z boson, for instance in elastic scattering

$$\nu_e \, e^- \rightarrow \nu_e \, e^- \tag{3.3}$$

The Z^0 boson itself as well as the $W^{\pm}$ bosons, responsible for the charged weak currents, were discovered in 1983 at CERN.

SM has embraced the idea of hadrons compositeness. The first step was to theorize the existence of point-like particles (Feynman called them "partons" at the time) as the building blocks of baryons and mesons. Following that, the first evidence of such particle was the discovery of up and down quarks, the first generation, as well strange quark, the first quark of second generation, in 1968 at SLAC experiment. After that, charm quark followed in 1974 at SLAC and BNL and bottom quark of third generation in 1977. The last but not least to be found was top quark in 1995 at CDS and D0 given it is the most massive and, thus, the most difficult to be produced. With the discovery of top quark, the third generation was completed as theoretically expected for symmetry reasons. The term "partons" is still in use today meaning the hadrons-consisting quark-gluon "soup". At the same time, the correspondence between the lepton-quark generations, at least about the number of generations, has been proven.

Finally, the last missing chapter that had been theorized by many theoretical physicists since 1965 was the Higgs boson as an excitation of the so-called Higgs field. The existence of at least one Higgs boson has been undoubtly verified in 2012 at ATLAS and CMS experiments of LHC/CERN. There are many theories that predict a whole family of Higgs bosons, but the already discovered boson is called the SM Higgs boson.

3.3.2 Pending questions

At this point, it is widely admitted that SM is not a universal theory despite its success. The reason is that even for masses well bellow the Planck scale where gravity is no longer negligible ($m_{Planck} = \sqrt{\frac{hc}{2\pi G}} \sim 10^{19}$ GeV), only electromagnetic and weak interactions are of a common origin. Additionally, the coupling constants associated to each of the three SM interactions, when extrapolated at higher energy with the renormalization group equations, are not converging perfectly to the same one point as they ideally would be expected to do.

Another hint attesting against SM as the final answer are various experimental observations that SM has nothing to say about. For instance, the neutrino oscillation observed in Super-Kamiokande Observatory is a fact (2015) and arises questions about extremely low neutrinos masses - but nevertheless with non zero mass - which SM takes for granted to be equal to zero.

SM includes many free parameters like Cabibbo - Kobayashi - Maskawa (CKM) mixing angles for flavor-changing weak decays or the masses and mixing angles of the neutrinos. They still wait to be experimentally determined.

Another case is the abundance of repulsive dark energy (68%) and attractive dark matter (27%) in the observable universe which have been verified to be a

fact through gravity lenses phenomenon that distort very far star light trajectories passing by a very large mass (e.g our Sun) very close to us. They both make up to $\sim 95\%$ of universal density, leaving the rest $\sim 5\%$ for the known baryonic mass. We know they are there, we just don 't know what they are made of.

The baryon asymmetry problem refers to the imbalance in baryonic and antibaryonic matter in the known universe. SM does not provide an obvious explanation for why this should be so and it is a natural assumption that the universe be neutral with all conserved charges. The Big Bang should have produced equal amounts of matter and antimatter. Since this does not seem to have been the case, it is likely some physical laws must have acted differently or did not exist for matter and antimatter. This issue is also known as the CP-violation which is about the CP-symmetry violation (or charge conjugation parity symmetry). CP-symmetry states that the laws of physics should be the same if a particle is interchanged with its antiparticle (C-symmetry), while its spatial coordinates are inverted ("mirror" or P-symmetry). The discovery of CP-violation in 1964 in the decays of neutral kaons resulted in the 1980 Nobel prize in Physics for its discoverers James Cronin and Val Fitch.

As of all the above, the final conclusion cannot be other than SM has been a very succesful tool up to now, but it is definitely not the end of the road for the construction of a final Grand Unified Theory (GUT)[4].

3.4 A possible solution

It is a fact that there is a striking coincidence between the existence of three lepton and quark families (or generations). Of course, it is not by any means final that there will not arise a fourth lepton or quark generation or why they should be exactly three. To put it in other words, there would be no need for the second and third generation, since all known matter consists only of the first. Nevertheless, a symmetry between them is however not included in SM, since there is no direct coupling between quark and lepton generations. There are some Beyond Standard Model (BSM) theories, which try to deal with these limitations by predicting new interactions arising from a more general symmetry group. The easiest way is by gathering leptons and quarks in the same multiplets, so such theories naturally allow lepton-quark transitions, which are mediated by new bosons usually called leptoquarks [5, 6].

3.5 Leptoquark bosons phenomenology and Beyond Standard Model (BSM) theories

The quarks and leptons despite being a representation of fields that are independent of each other, unexplained symmetries can be interpreted between their generation structure. An important indication is the cancellation of quarks' and leptons' contribution to the triangle anomalies of gauged currents, which are key contributions to the renormalization and consistency of SM. This remarkable balance between the contributions from leptons and quarks implies that these two particle types are part of a symmetry in a more fundamental theory.

[4] a GUT should unify electromagnetism, weak and strong interactions; it is supposed to be the intermediate step between SM and a Theory of Everything (ToE) which would incorporate gravity as well

Leptoquarks (LQ) are proposed particles that are responsible for the related interactions between leptons and quarks. They carry fractional electric charge, as well as both lepton and baryon numbers, and they are colour-triplet bosons. They can be either scalar or vector bosons and they are expected to couple directly to lepton-quark pairs, as shown in Fig. 3.1. The coupling strength between scalar LQs and the lepton-quark pair depend on a single Yukawa coupling $\lambda_{LQ-\lambda-q}$, while the additional magnetic moment and electric quadrupole moment interactions of vector LQs lead to their coupling strength with lepton-quark pairs being governed by two couplings, κ_G and λ_G. These respective coupling constants for scalar and vector LQs, as well as the branching ratio β into different leptons and quarks are model dependent. Such particles are predicted as a possible solution in numerous SM extensions.

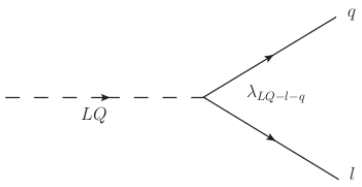


Figure 3.1: Feynman diagrams showing the Yukawa coupling $\lambda_{LQ-\lambda-q}$ between a leptoquark, a lepton and a quark.

The term "BSM Physics" is used to cover experimental observations that cannot be explained by SM, as well as a wide range of theories. BSM theories can represent any extension to SM, regardless of whether they include any proposed solutions to the unresolved questions by SM or not. It can also refer to any deeper explanation on observations that are already included in SM, but only with ad hoc parametrizations such as the Yukawa couplings. BSM theories have created a rather extended theoretical space for new particles and new interactions, nevertheless some signatures of new physics are at the TeV energy scale range and thus within the energy range of current generation of particle colliders.

In the following, some common BSM theories in which new interactions between leptons and quarks naturally arise are briefly discussed.

3.5.1 Grand Unified Theories (GUT)

Part of the limitations of SM can be overcome by considering new theories based on a larger gauge symmetry group G. Such models are referred to as Grand Unified Theories (GUT). GUTs denote a class of BSM theories that attempt to unify the strong and electroweak interactions at high energies. This unification of gauge couplings implies that these interactions are subsets of the same gauge symmetry at a larger scale, where there exist only one interaction with one universal gauge coupling. The breaking of this larger gauge symmetry is predicted to be at the energy scale of 10^{16} GeV (also commonly known as the "GUT scale" or Λ_{GUT}). The group G is chosen such that the additional symmetries could explain the arbitrary features of SM.

In order to be consistent with experimental observations, G must include SM symmetry group (which is of rank 4), to which it can be reduced after spontaneous

symmetry breaking. By choosing a gauge group based on a single coupling strength, the unification of all SM interactions is made possible. Thus, SM would be a low energy approximation of GUTs at the electroweak scale. The idea of a larger underlying symmetry that is broken at lower energies is also present in SUSY, hence it is worth noting that GUT and SUSY are not mutually exclusive from each other.

The gauge group G is required to include complex irreducible representations which can accommodate the particles of SM. By gathering a fermion family within the same multiplet, a connection between leptons and quarks can be introduced, as well as mediators (bosons) for this connection. Among the groups of rank 4 involving only one coupling strength, SU(5) is the only one meeting the requirements of a GUT.

The simplest model based on SU(5) was proposed by Georgi and Glashow in 1973 [7]. This model includes $SU(3)_C \times SU(2)_L \times U(1)_Y$ as a subgroup and can accommodate a complete fermion family with the association of the complex representations $\overline{5}$ and 10[5]. Translated in the representation $SU(3)_C \times SU(2)_L$, $\overline{5}$ and 10 become:

$$\overline{5} = (\overline{3}, 1) + (1, 2) = \overline{d} + \begin{pmatrix} \nu_\ell \\ \ell^- \end{pmatrix} \tag{3.4}$$

$$10 = (\overline{3}, 1) + (3, 2) + (1, 1) = \overline{u} + \begin{pmatrix} u \\ d \end{pmatrix} + \ell^+ \tag{3.5}$$

where u and d are the up and down quarks and $\begin{pmatrix} \nu_\ell \\ \ell^- \end{pmatrix}$ the weak isospin doublet respectively.

The SU(5) group contains 24 generators, including those of SM, namely the eight Gell-Mann matrices of $SU(3)_C$, the three Pauli matrices of $SU(2)_L$ and the phase of $U(1)_Y$. Since these generators must be represented by traceless matrices (as they are hermitian operators) and since the electric charge operator Q is one of them, the sum of the charges of the particles included in the $\overline{5}$ representation must vanish. This leads to the following equation:

$$Tr(Q) = 3Q_{\overline{d}} + Q_{\ell^-} = 0 \;\; \rightarrow \;\; Q_d = \frac{Q_{\ell^-}}{3} = -\frac{1}{3}e \tag{3.6}$$

where $Q_{\overline{d}}$ and Q_{ℓ^-} are the electric charges of a anti-down quark ($\overline{u}$) and a negatively charged lepton respectively. Thus, the fractional charge of quarks arises as a consequence from the preceding relation. Performing a similar calculation for the 10 representation, we conclude to the following result

$$Q_u = -2Q_d = +\frac{2}{3}e \tag{3.7}$$

explaining the electric charge difference between the two quarks. Finally, combining Eq. (3.6) and Eq. 3.7 we take back

$$Q_p = 2Q_u + Q_d = \frac{4}{3}e + (-\frac{1}{3}e) \;\; \rightarrow \;\; Q_p = +e \tag{3.8}$$

[5] $\overline{5}$ is the fundamental SU(5) representation consisting of 5×5 matrices and 10 is the adjoint representation consisting of 10×10 matrices; in adjoint representation 10 and $\overline{10}$ coincide

and, thus, unravelling the mystery of $Q_p = -Q_e$.

The 24 generators correspond to as many bosons which are included in the 24 adjoint representation. Decomposed in the representation $SU(3)_C \times SU(2)_L$, 24 becomes:

$$24 = (8,1) + (1,3) + (1,1) + (3,2) + (\overline{3},2) \tag{3.9}$$

where (8,1) includes the eight gluons, (1,3) + (1,1) corresponds to the electroweak bosons (γ, Z, $W^{\pm}$). (3,2) and ($\overline{3}$,2) contains twelve new bosons, 6 X and 6 Y of electric charge $\pm\frac{4}{3}$ and $\pm\frac{1}{3}$ respectively, and which are sensible to both $SU(3)_C$ and $SU(2)_L$ gauge transformations.

The new X and Y bosons can mediate lepton-quark or quark-quark transitions, which violate the conservation of the lepton (L) and baryon (B) numbers while conserving B-L. The most popular example of the consequences of such quark-quark transitions is decay of the proton into mesons and leptons, as represented in Fig. 3.2. Since its experimental lower limit is $\tau_{proton} > 10^{31}$ years, high lower bound on the mass of the X and Y bosons are set, namely $m_{X,Y} > 10^{15}$ GeV. At that point, all the known interactions besides gravity are being characterized by the same Yukawa coupling (they are of the same strength). Of course, such a scale is yet, or for the foreseeable future, not accessible by experiments. Propagated down to the electroweak scale, the predictions of the Georgi and Glashow SU(5) model are not in agreement with experimental data.

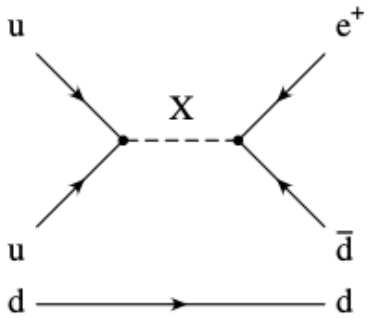


Figure 3.2: Decay of the proton (uud) into a positron and neutral pion ($\pi^0 \to d\bar{d}$), through the mediation of the X boson. This process conserves the electrical charge, but violates both lepton and baryon numbers. In general direct observations of some massive particles that are predicted in GUTs are beyond the reach of foreseeable particle colliding experiments. Observations of proton decay, neutrinos properties as well as electric dipole moments of SM particles may be used as indirect probes to the presence of such particles.

Other GUT models including one coupling constant and based on gauge groups of rank greater than 4 have been postulated. For instance, the SO(10) group of rank 5 allows to build a theory where all the fermions are included within the same irreducible representation, namely 16. Several ways to break SO(10) down to the electroweak scale are possible, one of them leading to the Pati - Salam GUT model, of group $SU(3)_C \times SU(2)_L \times SU(2)_R$, and which interprets the lepton sector as the fourth color. As it is expected, models based on even larger symmetry group, like for instance SU(15) or E(6), which can also gather all the fermions within their fundamental representation can be constructed. However, their main drawback is the introduction of numerous new parameters.

3.5.2 Supersymmetry (SUSY)

Supersymmetry (SUSY) is a very popular concept that many proposed theories are built on[6] as an extension of SM. It allows the unification of matter and interactions through the existence of individual super-partners (sparticles) for all particles in SM, where they differ by a half-integer in spin from the respective original SM particle. A supersymmetric transformation acts on the spin of a particle and associates a fermion to a boson and a boson to a fermion. These superpartners are predicted to have the same mass as their SM partners in theories with unbroken supersymmetry, however none of these have been seen in experiments with their SM partners, as they should if they exist. This implies that if all SM particles have a respective supersymmetric super-partner, they must be much heavier compared to the SM particle as predicted by theories with spontaneously broken supersymmetry, potentially at the TeV scale.

SUSY models in general contain explanations for matter/antimatter asymmetries with extra CP-violation processes. They also predict the unification of the electroweak and strong gauge couplings at high energies. It is worth mentioning that the introduction of superpartners cancels the divergences of the Higgs mass corrections, as their contribution counteract that of SM particles. This is due to the fact that loop corrections have amplitudes of opposite signs depending on whether the contributing particle is a boson or a fermion. Since no supersymmetric particle has yet been observed, the supersymmetry must be broken. Various symmetry breaking scenarios are possible leading to numerous supersymmetric models. Furthermore, the precise convergence of the three coupling constants contributes to make SUSY models serious candidates for a universal theory of all interactions.

While there are many currently existing supersymmetric models, the "Minimal Supersymmetric Standard Model" (MSSM) is the simplest and one of the best studied candidates. It includes only a minimal number of extra particles to extend SM with supersymmetry, while also proposing solutions to the SM open questions in similar ways as all other non-minimal SUSY theories. In MSSM, the Higgs potential is constrained, and thus no more arbitrary. In order to explain the stability of the proton in MSSM, a new quantum number is introduced, the R-parity, as defined by the following relation:

$$R = (-1)^{3B+2S+L} \tag{3.10}$$

where S is the spin of the particle, and B and L the baryon and lepton numbers, respectively. The R-parity is discrete and multiplicative. It is equal to 1 for SM particles and -1 for supersymmetric particles. Although R-parity is supposed to be conserved, its violation is theoretically allowed, but yet leads to contrasting phenomenologies.

For a given process, the conservation of R-parity is achieved in the same way as in the case of spatial parity. The initial and final parities are equal to the product of the parities of the initial and final states respectively. As a result, when R-parity is conserved, SUSY particles can only be produced in pairs and can decay only to another SUSY particle in association with a particle from SM. The lightest SUSY particle (LSP), arising from successive SUSY particle decays, must thus be stable. In order to fit observations, LSP must be neutral and only interacting through weak

[6] on the contrary, super-asymmetry has been proposed only by dr. Sheldon Cooper so far

interactions and gravitational interactions. That is the reason it is often also called a weakly interacting massive particle (WIMP).

In the case of violated R-parity, SUSY particles are allowed to decay exclusively into SM particles and, by considering the opposite process, the single production of SUSY particles is thus also allowed. As a consequence, the lightest SUSY particle is no more stable and cannot describe dark matter.

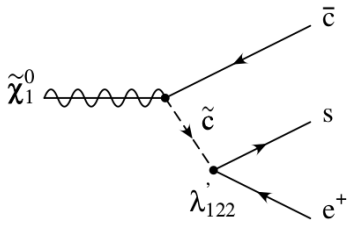


Figure 3.3: Decay of the neutralino $\tilde{\chi}_1^0$ (the supersymmetric partner of SM neutral bosons) into a positron and two quarks. R-parity is assumed to be violated. In the process, the supersymmetric partner of the quark $\tilde{c}$ (the squark) is decaying to a positron and a quark through a leptoquark-like coupling.

Assuming the violation of R-parity, supersymmetric bosons are thus allowed to decay to SM fermions by violating the lepton and baryon numbers and involving a leptoquark-like coupling. Fig. 3.3 represents an example of a R-parity violated process, namely the decay of a neutralino (the superpartner of SM neutral bosons) to a squark and a quark (which conserves R-parity), followed by the decay of the squark (the superpartner of a quark) to an electron and a quark, which violates R-parity through a leptoquark-like coupling (λ_{122}). Experimental and theoretical constraints on the leptoquark-lepton-uark coupling can thus be translated into constraints on the squarks.

3.6 Leptoquark effective models

The search for leptoquarks can be carried out by relying on effective models and, thus, independently of theories beyond SM, like those presented in Sec. 3.5. However, constraints on the nature of the leptoquarks and on their interactions with SM particles have to be imposed so as to be consistent with experimental observations. The following description is based on [14].

A general effective leptoquark model, as proposed by Buchmüller, Rückl and Wyler (BRW) [15], requires leptoquarks to have normalizable interactions, to obey SM gauge groups $SU(3)_C \times SU(2)_L \times U(1)_Y$ symmetries and to couple only to SM fermions and gauge bosons.

In order to ensure proton stability, leptoquarks are further constrained to conserve lepton and baryon numbers separately. This leads to seven scalar and seven vector leptoquarks carrying the fermionic number $F = 3B + L$ of either $|F| = 0$ or $|F| = 2$.

Table 3.3 lists the resulting seven scalar (S) and seven vector (V) leptoquarks, along with their electric and weak charges, their fermionic numbers and their decay products. Within each isospin family, mass degeneracy is assumed for experimental searches, which is theoretically motivated when omitting loop corrections. The symbols referring to the leptoquarks in Table 3.3 therefore designates any of its electric

charge state. In addition, the L and R subscripts on the leptoquark symbol denotes to the chirality of the coupled lepton. For instance, the symbol $S_{\frac{1}{2},L}$ both refers to a scalar leptoquark of electric charge -$\frac{5}{3}$ and -$\frac{2}{3}$, which couples to a left-handed lepton. Weak isospin is appended as a subscript and is equal to 0, $\frac{1}{2}$ or 1.

LQ type	$\|F\|$	Q	T_3	Decay products
$S_{0,L}$	2	-1/3	0	$\ell_L^- u_L, \nu_L d_L$
$S_{0,R}$	2	-1/3	0	$\ell_R^- d_R$
$\widetilde{S}_{0,R}$	2	-4/3	0	$\ell_R^- d_R$
$S_{1,L}$	2	-4/3	-1	$\ell_L^- d_L$
	2	-1/3	0	$\ell_L^- u_L, \nu_L d_L$
	2	+2/3	+1	$\nu_L u_L$
$S_{\frac{1}{2},L}$	0	-5/3	-1/2	$\ell_L^- \overline{u}_L$
$S_{\frac{1}{2},R}$	0	-5/3		$\ell_R^- \overline{u}_R$
	0	-2/3	+1/2	$\ell_L^- \overline{d}_L$
$\widetilde{S}_{\frac{1}{2},L}$	0	-2/3	-1/2	$\ell_L^- \overline{d}_L$
	0	+1/3	+1/2	$\nu_L \overline{d}_L$
$V_{\frac{1}{2},L}$	2	-4/3	-1/2	$\ell_L^- d_R$
$V_{\frac{1}{2},R}$	2	-4/3		$\ell_R^- d_L$
	2	-1/3	+1/2	$\ell_R^- u_L$
$\widetilde{V}_{\frac{1}{2},L}$	2	-1/3	-1/2	$\ell_L^- u_R$
	2	+2/3	+1/2	$\nu_L u_R$
$V_{0,L}$	0	-2/3	0	$\ell_L^- d_R, \nu_L \overline{u}_R$
$V_{0,R}$	0	-2/3	0	$\ell_R^- d_L$
$\widetilde{V}_{0,R}$	0	-5/3	0	$\ell_R^- \overline{u}_L$
$V_{1,L}$	0	-5/3	-1	$\ell_L^- \overline{u}_R$
	0	-2/3	0	$\ell_L^- \overline{d}_R, \nu_L \overline{u}_R$
	0	+1/3	+1	$\nu_L \overline{d}_R$

Table 3.3: Classification of the leptoquarks in the Buchmüller - Rückl - Wyler effective model. The content of seven scalar (S) and seven vector (V) leptoquarks is listed along with their fermionic numbers (F = 3B + L), their electric and weak charges (Q and the third component T_3 of the weak isospin) and their possible decay products. The symbols referring to the leptoquarks denote any of their electric charge state, the appended L and R subscripts designate to the chirality of the coupled lepton and the appended number is the weak isospin quantum number.

Although all 14 leptoquarks appear in a GUT based on SU(15), only a subset of the BRW leptoquarks are generally included in a specific fundamental model (for example, V_0 is part of the Pati - Salam GUT model).

The branching fractions of the leptoquark decays into a charged lepton and quark or neutrino and quark ($\beta = \mathrm{Br}(\mathrm{LQ} \to \ell \mathrm{q})$) are determined by the respective leptoquark-lepton-quark coupling. By construction, it can thus only take the discrete values 0, $\frac{1}{2}$ or 1.

Under the preceding assumptions, leptoquarks could in principle decay into any combination of a quark and a lepton, but LQs with masses as low as $\mathcal{O}(100\text{GeV})$ are only allowed to couple to one generation of quarks and leptons, since they otherwise would generate lepton number violation and sizable flavor-changing neutral currents. An example of flavor-changing neutral current mediated by a leptoquark is provided in Fig. 3.4, where the decay of the B^0 meson into a tau and a anti-muon is represented. Under this effective model, leptoquarks can thus be indexed by the generation of the fermion to which it couples. There are therefore three generations of leptoquarks.

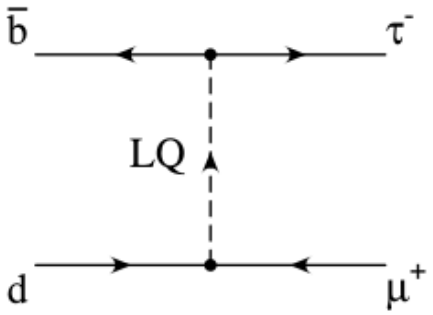


Figure 3.4: Example of flavor-changing neutral current. A B^0 meson ($d\bar{b}$) decaying to a τ-lepton and an anti-muon.

Additional low energy constraints, like for instance chirally suppressed meson decays (e.g. $\pi \to$ e ν) or virtual-loop contributions to the g-2 of the muon, entail the necessity of assuming that leptoquarks have pure chiral couplings to SM fermions. Augmented by the two previous requirements, the BRW model is called the minimal Buchmüller - Rückl - Wyler effective model (mBRW).

The interaction of the leptoquarks, as defined under the mBRW model, with SM bosons is completely determined in the case of scalar leptoquarks, since it only relies on SM interactions. The coupling between vector leptoquarks and SM bosons is however more complicated, as it depends on trilinear and quartic couplings, which might require the introduction of anomalous couplings, which are free parameters. Four anomalous independent coupling are associated to the electroweak sector and two additional are introduced for the strong sector. More details on the interaction of vector leptoquarks with SM bosons are provided in [16, 17].

The assumptions of the mBRW effective model allows relatively small LQ masses in reach of hadron colliders like LHC. Enriched phenomenology can be predicted when reasonably relaxing some of its constraints. For instance, assuming that leptoquarks not only couple to SM bosons but also to other unknown new fields enables the search for generic models, in which the branching fraction $\beta = \text{Br}(\text{LQ} \to \ell\text{q})$ can take arbitrary values. It is worth mentioning that under this assumption, the leptoquarks $\tilde{S}_{\frac{1}{2}}$ and S_0 could be associated to the $\tilde{u}$ and $\tilde{d}$ squarks (superpartners of quarks) respectively.

The search for second generation scalar leptoquarks performed in this thesis relies on the mBRW with the exception that β is assumed to be a free parameter, which allows a more general investigation.

3.7 Physics in hadron colliders

The primary feature for any physical process to take place in any type of collider is the production rate $\mathscr{R}$. This is given by

$$\mathscr{R} = \frac{dN}{dt} = \mathscr{L} \cdot \sigma \tag{3.11}$$

This means that the rate with which a certain process is produced depends on the product of the luminosity, $\mathscr{L}$, with the cross-section of the specific process, σ. The former is an accelerator-dependent parameter, while the latter is physics related. The physical meaning of luminosity is the number of collisions taking place in a specific area per second. The more dense the beam (more protons focused in a small area) the more collisions will occur. On the other hand, the cross-section is a parameter[7] that expresses numerically the possibility for a specific process to take place. The bigger the cross-section (the area around the interaction) the higher the probability these two protons will give the expected calculated result. The cross-section is a center-of-mass energy depending parameter which means that for a given interaction it gets bigger as the energy increases. In conclusion, the intended result is to acquire as much as possible higher center-of-mass beam energy (more energetic protons) and, at the same time, more dense beams (more protons in a confined space) to increase the possibility for the protons to collide each other.

The luminosity that the accelerator provides to the experiments is determined by the fine-tuning of the parameters of the machine. In detail, at a hadron collider where two beams, structured as bunches of protons, are made to collide under an angle, the luminosity is defined as

$$\mathscr{L} = \frac{N^2 \cdot n_b \cdot f \cdot \gamma}{4\pi \cdot \epsilon_n \cdot \beta^*} \cdot F \tag{3.12}$$

where N is the number of particles in a bunch, n_b the number of bunches per beam, f is the revolution frequency, γ is the relativistic factor, ϵ_n is a factor called normalised emittance that gives the spatial spread of the beam, β^* is the beta function at the interaction point which reflects the transverse size of the beam (the collisions are not frontal, but off-center so as to provide a larger effective bunch area[8]), while F is a reduction factor due to the angle at which the two beams are colliding.

All these parameters are fine-tuned in order to provide maximum luminosity to an experiment. However, the maximum luminosity that is provided by the accelerator is not fully recorded by the experiments, due to dead-time issues that arise in each experiment. For example, Fig. 3.5 shows the total integrated luminosity as a function of time, for the full Run 2 of the ATLAS detector at the LHC machine. Different coloured contours show the luminosity that LHC provided and what was recorded by ATLAS. Furthermore, from the luminosity recorded by ATLAS, a small fraction of it is not good for physics analysis purposes, since for example, some data are of low quality due to temporary malfunctions of some ATLAS sub-system.

[7] measured in barn = 10^{-28}m^2

[8] the geometrical cross-section of each collision is not a circle, but an ellipse

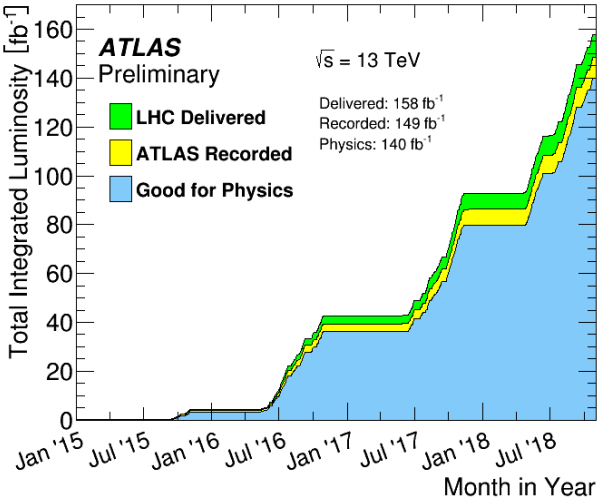


Figure 3.5: The total integrated luminosity as a function the month for Run 2 until October 2018 data taking periods. The different contours show the luminosity that the LHC delivered (green), the ATLAS detector recorded (yellow) and which fraction of the recorded luminosity is good for physics (blue). The All Good Data Quality criteria require all reconstructed physics objects to be of good data quality.

Contrary to lepton colliders, in a hadron collider the beam particles are composite. For example, the protons that are colliding in the LHC machine can be imagined as a composite object of three bound "valence" quarks and a sea of virtual quark - anti-quark pairs, as well as gluons that are continuously created and annihilated due to quantum fluctuations. These quantum fluctuations are non-perturbative given the low energies involved. This internal structure of the collided particle has as a consequence the fact that the energy given to the particle is split amongst its constituents. Thus, in a hadron collider the actual collisions happen between two constituents, one from each hadron, that carry a fraction of the incoming hadron momentum χ_1 and χ_2. In such a case, the cross-section of a process Q can be defined as:

$$\sigma_Q(s, \mu_R, \mu_f) = \sum_\alpha \int_0^1 d\chi_1 d\chi_2 \cdot f_\alpha(\chi_1, \mu_f) f_\alpha(\chi_2, \mu_f) \cdot \hat{\sigma}_Q(\hat{s}, \mu_R, \mu_f) \tag{3.13}$$

where α represents the different types of partons (quarks or gluons) and s is the squared center-of-mass energy provided by the collider. The function f_α is referred to as the parton density function (PDF), which in defined as the density of the partons in the hadron that carry χ fraction of the hadron total momentum. The use of the PDFs is advantageous since it decouples the non-perturbative effects from the hard-scattering process described by the term $\hat{\sigma}_Q$ in Eq. 3.13. The cross-section of hard scattering, where "hard" is being interpated as large momentum transfer between the constituents of the two hadrons, in such a scenario is calculated in a reduced energy $\hat{s} = \chi_1 \chi_2 s$. Finally, the terms μ_f and μ_R are referred to as factorisation and renormalisation scale. The factorisation scale is introduced in order to renormalise effects as the collinear gluon splitting and the emission of soft gluons, when imposing the assumption that these quantities are finite at a given scale. In addition, the renormalisation scale is introduced to deal with divergences in the computation of the scattering amplitude of the hard process. This is due to the fact that the

hard scattering cross-section $\hat{\sigma}_Q$ can be computed using perturbation theory up to a certain order.

Especially for the protons, the PDFs have been studied by deep-inelastic scattering experiments, as well as from measurements at hadron colliders [18]. Several collections of PDFs, referred to as PDF sets, exist and slightly vary due to the different approaches that have been used to extract this information from the data. For the theoretical predictions with the use of Monte Carlo simu1ation a specific set is used per process. The differences between the various PDF sets for the same process are usually regarded as theoretical uncertainty when the analysis is performed.

3.7.1 Leptoquarks in hadron colliders

Both single and pair production of leptoquarks can occur at a hadron collider in pp collisions like LHC.

The single leptoquark production depends on the unknown leptoquark-lepton-quark coupling ($\lambda_{LQ-\ell-q}$) and therefore relies on the theoretical model considered. Fig. 3.6 and Fig. 3.7 show an example of single LQ production, in s- and t-channels respectively.

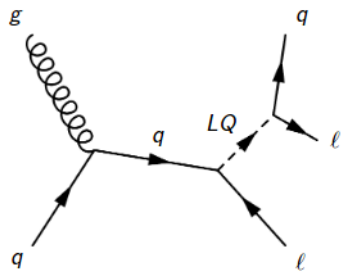


Figure 3.6: s-channel diagram for single LQ production

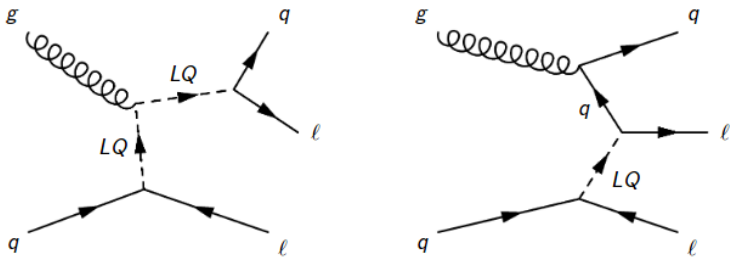


Figure 3.7: t-channel diagram for single LQ production

By contrast, the pair LQ production is predominantly realized in the s-channel, via the strong coupling, independently of $\lambda_{LQ-\ell-q}$, and thus only depends on the

assumed leptoquark mass. The additional contribution to the LQ pair production due to t-channel lepton exchange with a cross-section proportional to $\lambda^2_{LQ-\ell-q}$ could actually also be considered, but turns out to be negligible in the case of the search for second or third generation leptoquarks. When adopting the general assumption that the leptoquark couples to the lepton and the quark of the same generation, this process is further suppressed by vanishing quark parton distribution functions (PDF) at high Bjorken x and, thus, negligible in comparison with uncertainties on the predicted cross-section.

Searches for leptoquarks at electron - proton colliders (such as HERA) or e^+e^- colliders (such as LEP), where leptoquarks can only be singly produced, have maximized sensitivities for leptoquarks of the first generation. On the contrary, pair leptoquark production allows searches for the three generations of leptoquarks.

In LHC, pair LQ production (Fig. 3.8) can lead to three characteristic final states: $\ell^+\ell^-$qq, $\ell^\pm\nu_\ell$qq and $\nu_\ell\nu_\ell$qq. The analysis presented in this thesis describes the search for second generation scalar pair LQ production in the decay mode (channel) $LQ\overline{LQ} \rightarrow \mu\nu_\mu$qq. In the following, the branching fraction β is restrained to the case of second generation fermions and is thus defined by $\beta = \mathrm{Br}(LQ \rightarrow \mu q)$. The branching ratios for the three possible pair LQ decay modes are expressed by the formula below as function of β:

$$Br(LQ\overline{LQ} \rightarrow \mu\mu qq) = \beta^2 \tag{3.14}$$

$$Br(LQ\overline{LQ} \rightarrow \mu\nu_\mu qq) = 2\beta(1-\beta) \tag{3.15}$$

$$Br(LQ\overline{LQ} \rightarrow \nu_\mu\nu_\mu qq) = (1-\beta)(1-\beta) \tag{3.16}$$

The branching fraction of the $LQ\overline{LQ} \rightarrow \mu\nu_\mu$qq mode is thus maximized for $\beta = 0.5$, whereas in the $LQ\overline{LQ} \rightarrow \mu\mu$qq mode, the best search sensitivity is obtained for $\beta = 1$. At $\beta = 0$, only the mode $LQ\overline{LQ} \rightarrow \nu_\mu\nu_\mu$qq has a non-vanishing branching ratio. In this thesis we investigate only the $\mu\nu_\mu$qq topology.

The main background processes from SM to the pair production of leptoquarks followed by their decay into the $\mu\nu_\mu$qq final state is the W. At higher energies, the production of top quark pairs also contributes significantly to this final state.

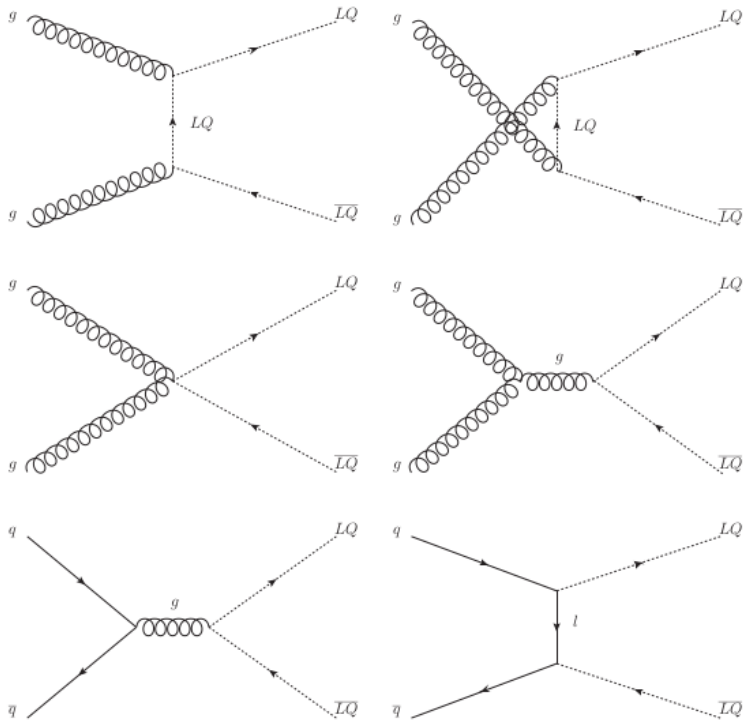


Figure 3.8: Dominant leading order Feynman diagrams for the LQ pair-production of scalar LQ from gluon - gluon (gg) fusion and quark - anti-quark ($q\overline{q}$) annihilation.

3.8 Recent experimental results at LHC

At the time (2018) Large Hadron Collider is one of only two hadron colliders[9], but the biggest in the world. Its two biggest experiments which still search for leptoquarks are ATLAS (A Toroidal Lhc ApparatuS) and CMS (Compact Muon Solenoid). All searches have been conducted with the LHC proton - proton (pp) beam. The accelerator have started with a nominal beam energy at 3.5 TeV (7 TeV totally). The first main upgrade have raised the total beam energy at 8 TeV (4 TeV per beam) and the last upgrade until now at total energy of 13 TeV (6.5 TeV per beam).

Some of the more recent results about leptoquark searches are cited below. CMS searches at center-of-mass total beam energy at 8 TeV and integrated luminosity of $\mathscr{L} = 19.7$ fb^{-1} have been concluded with an exclusion limit of 1010 GeV for the first generation pair leptoquark, dielectron channel (eejj). The same limit for first generation pair leptoquark, the e + E_T^{miss} channel (e$\nu_e$jj) was at 850 GeV. In the same way, the second generation dimuon channel ($\mu\mu$jj) exclusion limit was at 1080GeV and the μ + E_T^{miss} channel ($\mu\nu_\mu$jj) was at 760 GeV. In a more recent CMS paper for LQ searches at 13 TeV and $\mathscr{L} = 2.6$ fb^{-1} (Fig. 3.10) there have been only first generation results. For the eejj channel the exclusion limit has been raised up to 1130 GeV and regarding the e$\nu_e$jj channel is at 920 GeV.

[9]the other is Relativistic Heavy Ion Collider (RHIC) in BNL, but opareting only with heavy ion beams

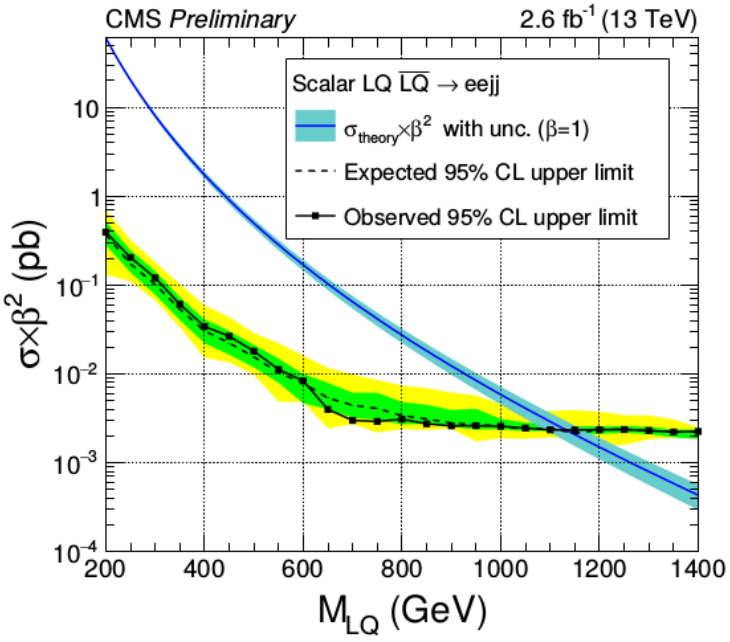


Figure 3.9: CMS results at 13 TeV, integrated luminosity at 2.6 fb^{-1}.

On the other hand, the ATLAS paper at 8 TeV and $\mathscr{L} = 20.0$ fb^{-1} (Fig. **??**) embodies the results of three channels, one for each generation. The first dielectron channel has an exclusion limit at 1050 GeV and the second dimuon channel is at 1000 GeV. The last is the dineutrino-dijet third generation channel ($b\nu_\tau\bar{b}\bar{\nu}_\tau$) which contains the LQ mass between 200 GeV and 640 GeV (200 GeV $< m_{LQ_3} <$ 640 GeV). Finally, the ATLAS 13 TeV and 3.2 fb^{-1} search contains only the dielectron and dimuon channels with exclusion limits at 1100 GeV and 1050 GeV respectively.

As discussed earlier, the coupling strength between scalar LQs and lepton-quark pairs depend only on a single Yukawa coupling $\lambda_{LQ-\ell-q}$, while the equivalent coupling strength for vector LQs are governed by two couplings, κ_G and λ_G, due to the extra magnetic moment and electric dipole moment interactions of vector LQs. These couplings for vector LQs are model dependent and their variation between different models can yield cross-section values that differ by one to two orders of magnitude.

Calculations for the LO production cross-sections for pair-produced scalar LQs at the LHC has been performed by Krämer et al [19]. The two primary production processes, gluon fusion and quark - anti-quark annihilation, are respectively predicted by

$$\hat{\sigma}_{gg}^{LO} = \frac{\alpha_s^2\pi}{96\hat{\sigma}}\left[v(41-31v^2)+(18v^2-v^4-17)\log\frac{1+v}{1-v}\right] \tag{3.17}$$

$$\hat{\sigma}_{qq}^{LO} = \frac{\alpha_s^2\pi}{\hat{\sigma}}\frac{2}{27}v^3 \tag{3.18}$$

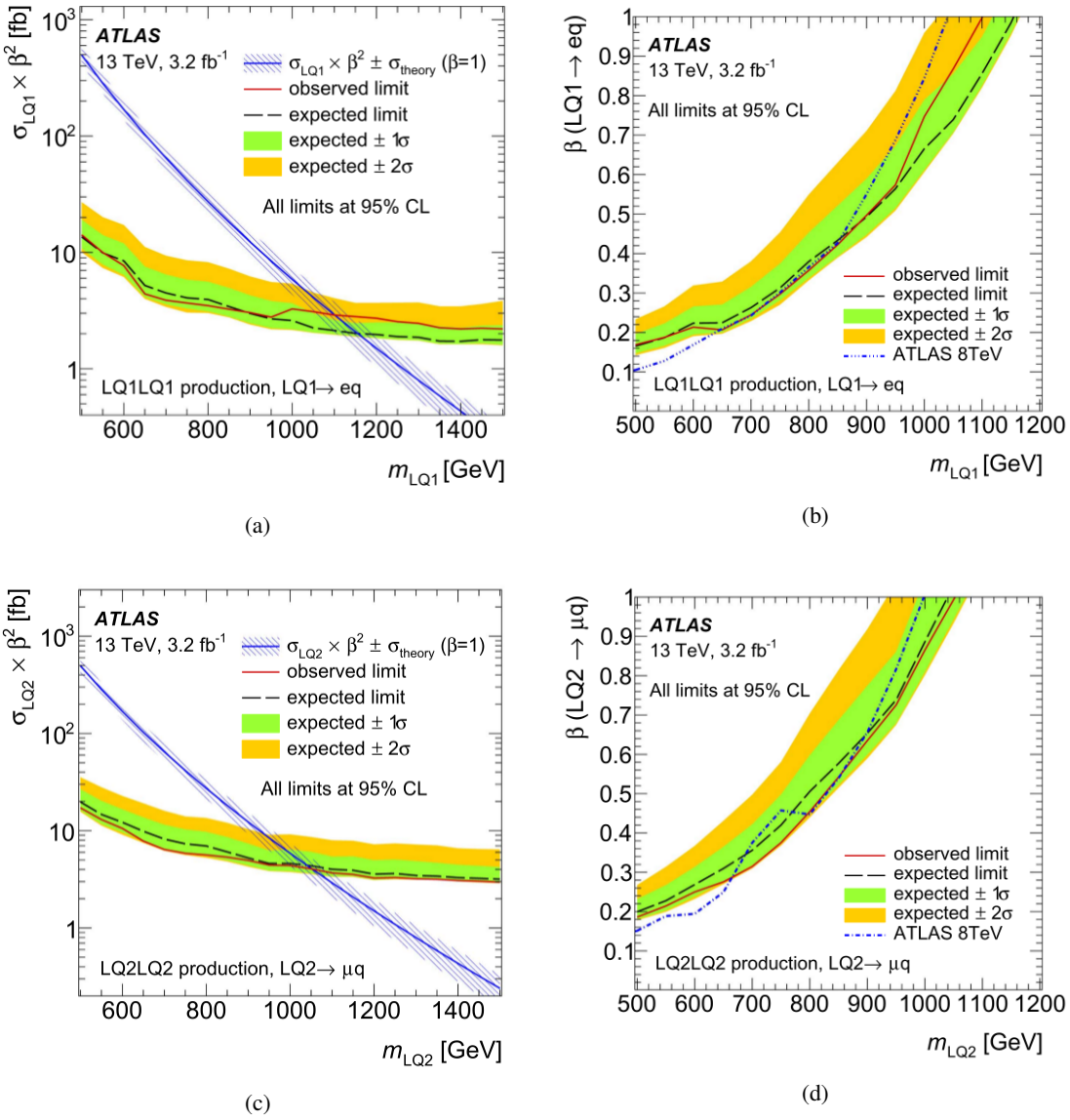


Figure 3.10: CMS results at 13 TeV, integrated luminosity at 2.6 fb^{-1}.

where $v = \sqrt{1 - \frac{4m_{LQ}^2}{\hat{s}}}$ denotes the velocity of the LQ, with $\hat{s}$ being the collision energy squared. Production at low LQ mass values are dominated by the gluon fusion mechanism, while the contribution from quark - anti-quark annihilation becomes significant at higher LQ mass values, up to 30% at $m_{LQ} = 1.5$ TeV. QCD radiative corrections can be expressed as

$$\sigma(pp \rightarrow LQ\overline{LQ}) = \sigma_{gg} + \sigma_{gq} + \sigma_{q\bar{q}} \tag{3.19}$$

which include virtual corrections, gluon bremsstrahlung as well as contributions from quark-gluon collisions. The perturbative expansion of the total partonic production

cross-section can be written as

$$\hat{\sigma}_{ij} = \frac{\alpha_s^2(m_{LQ}^2)}{m_{LQ}^2}\left\{f_{ij}^B(v) + 4\pi\alpha_s(m_{LQ}^2)\left[f_{ij}^{v+s}(v) + f_{ij}^h(v)\right]\right\} \tag{3.20}$$

where i,j denotes g,q, while f^B, f^{v+s} and f^h respectively represents the Born cross-section term for parton subprocesses and the virtual + soft and hard gluon corrections. These scaling functions have dependence on $\sqrt{\hat{s}}$ through v and can be expressed as follows near the productions threshold with $v \ll 1$

$$f_{gg}^B = \frac{7\pi v}{384} \qquad\qquad f_{q\bar{q}}^B = \frac{\pi v^3}{54}$$

$$f_{gg}^{v+s} = f_{gg}^B\left[\frac{11}{336v}\right] \qquad f_{q\bar{q}}^{v+s} = f_{q\bar{q}}^B\left[\frac{1}{48v}\right] \tag{3.21}$$

$$f_{gg}^h = f_{gg}^B\left[\frac{3}{2\pi^2}\log^2(8v^2) - \frac{183}{28\pi^2}\log^2(8v^2)\right]$$

$$f_{q\bar{q}}^h = f_{q\bar{q}}^B\left[\frac{2}{3\pi^2}\log^2(8v^2) - \frac{107}{36\pi^2}\log^2(8v^2)\right] \tag{3.22}$$

by choosing m_{LQ} as the renormalization and factorization scale. The dependence of these functions on the scale reduces greatly when these next-to-leading order (NLO) terms are included, which in turn enhances the production of leptoquarks.

Chapter 4

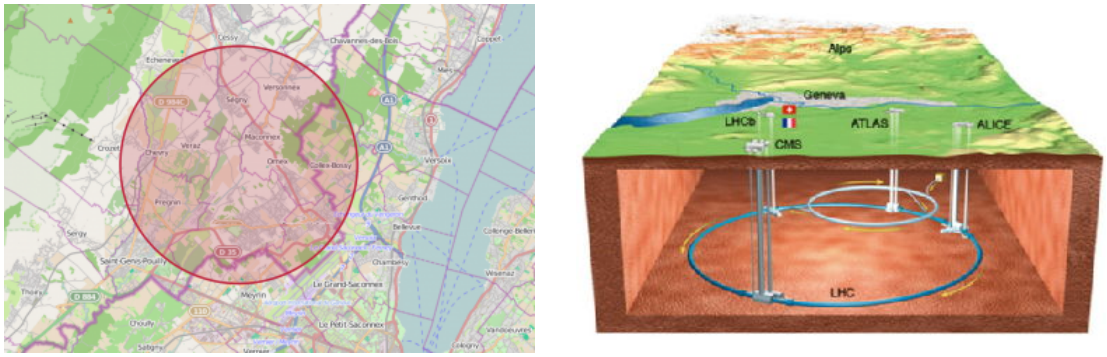
LHC Accelerator and the ATLAS Detector

In this chapter a detailed analysis of the characteristic properties of LHC hadron collider and ATLAS detector is performed. The searches for the hypothetical leptoquark boson in this thesis made use of the data collected by the ATLAS detector during the second phase of Run 1 at center-of-mass energy of 8 TeV and, also, during the first phase of Run 2 at 13 TeV. We follow the protons from their production through ionization of hydrogen atoms and acceleration until their collision with the protons of the opposite beam. Then, a full presentation and explanation of the individual parts (sub-detectors) which ATLAS detector is constructed from follows. Their task is the collisions monitoring through recording and storing all their remnants. This is the most crucial level, since all physics analyses which take place at the end of this procedure depend absolutely on the accuracy of the monitoring of all the remaining particles and their physical properties. In case of an assuming problematic collision recording, the specific collision results (this is called an event) is rejected as non trustworthy and, thus, not suitable for physics analysis.

4.1 LHC complex

The Large Hadron Collider (LHC) [20–23] is a circular proton - proton (pp) collider located at CERN in Geneva (Switzerland). The ring of 27 km of circumference is built in the already existing LEP (Large Electron Positron) experiment tunnel 1 as deep as 175 m beneath the Franco-Swiss border, next to the southmost point of lake Léman (lake of Geneva), Fig. 4.1a. Most of the tunnel in under France (Fig. 4.1b).

The project was approved in December 1994 and aimed to extend the particle physics searches to higher energy scale and to facilitate the studies on rare, undiscovered phenomena, as well as precision measurements for the establishment of the previously observed ones. The main difference between LEP and LHC, a hadron - lepton and a purely hadron - hadron collider, is that hadrons have internal structure and, thus, there are more ways of interaction and possible different outcomess. On the other hand, a lepton is a point-like entity with defined initial state. In general, a hadron collider is more adequate for new discoveries. On the contrary, a collider with one or both lepton beams is better for accurate measurements of already known



(a) (b)

Figure 4.1: CERN map. (a) The pink circle denotes the undergound main LHC tunnel. On the right is lake Léman. (b) The four biggest LHC experiments with their subterrain shafts and collider 's tunnel.

particles and phenomena. As discussed in the previous chapter, LHC is contributing to the searches of any physics process by providing the luminosity at a certain center-of-mass energy, designed to reach a nominal value of $\mathscr{L} = 10^{34}$ cm^{-2}s^{-1} = 10^{-5} fb^{-1}s^{-1} and $\sqrt{s} = 14$ TeV, which is the machine related parameters in the rate of a process.

At the time of writing (December 2018), LHC is about to complete almost nine very successful years of operation (2010-2018) completing the first period (Run 1) of data taking, colliding protons at $\sqrt{s} = 7$ TeV and 8 TeV, as well as providing ^{82}Pb^{+} - ^{82}Pb^{+} heavy ion collisions at $\sqrt{s} = 2.76$ TeV and p - ^{82}Pb^{+} collisions at $\sqrt{s}$ = 5.06 TeV. During the period 2013-2014 the first full Long Shutdown (LS1) was completed when many sub-systems were upgraded. In 2015 the second data taking period (Run 2) was started with the center-of-mass energy of both beams at 13 TeV. Since then there have been two intervals of lead ion beams between pp beams and two short-term technical stops increasing gradually the integrated luminosity of the detector. At this point (mid-December 2018) LHC has already stopped for the second full Long Shutdown (LS2) and it will remain without beams until the spring of 2020. After that, Run 3 will start with a total center-of-mass energy at $\sqrt{s} = 14$ TeV. So far the LHC total delivered integrated luminosity has reached 158 fb^{-1}[1]. From that, the ATLAS recorded luminosity is at 149 fb^{-1} and 140 fb^{-1} have been qualified as good data for physics analyses. Theoretically, to accumulate integrated luminosity of $\sim$ 160 fb^{-1} with the maximum future steady rate of $5 \cdot 10^{-5}$ fb^{-1}s^{-1}, the detector should operate continuously for approximately 37 days.

The full accelerator complex, Fig. 4.2, starts from the production of protons. They are the nuclei of ionized hydrogen atoms (H^{+}). These protons are accelerated to 50 MeV via the LINAC-2 before being injected into a series of accelerators since they gain their final energy gradually. The first step is the Proton Synchrotron Booster (BOOSTER, up to 1.4 GeV). It follows the Proton Synchrotron (PS, up to 25 GeV) and finally the Super Proton Synchrotron (SPS, up to 450 GeV). Finally the beam prepared by the SPS is been injected into LHC where the proton beam is accelerated in both opposite directions forming bunches of protons with a time spacing from each other of 25 ns. The two beams travel into ultrahigh vacuum guided by the magnetic

[1] $\mathscr{L}_{int} = \int \mathscr{L} dt$; its unit is cm^{-2} or b^{-1} = 10^{28}m^{-2}

field of superconducting magnets freezed down to -271.3°. They are parallel for most of the circumferance except from four interaction points (IPs) within the accelerator ring corresponding to the four major experiments hosted at the LHC: ATLAS , CMS, LHCb and ALICE. The beams cross each other in these points and the detectors record the collisions.

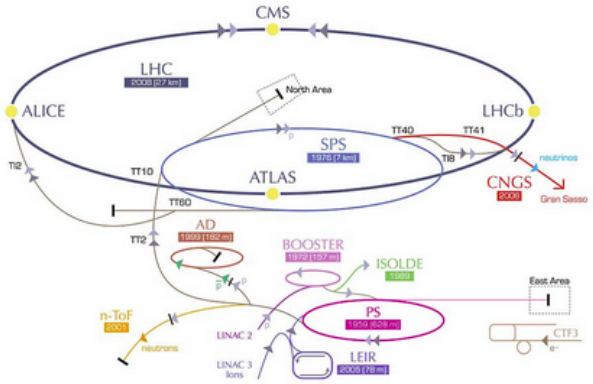


Figure 4.2: Schematic view of the accelerator chain starting from the proton production up to the collisions at the interaction points corresponding to the major LHC experiments.

Table 4.1 summarises the LHC peak performance during the Run 1 and the most recent phace of Run 2 (2017) in terms of the machine parameters. The data-taking period of 2010 was mainly used for the machine commissioning and testing with a center-of-mass energy of 7 TeV, delivering 48.1 pb^{-1} . On the other hand, the 2011 and 2012 datasets are mainly used for physics analysis purposes with a center-of-mass energy of 7 TeV and 8 TeV and delivering 5.6 fb^{-1} and 23.3 fb^{-1}, respectively. During 2017 the detector has managed to succeed in or even surpass most of the planned characteristic parameters. The last column of the table refers to the values of settings at which the LHC was designed to operate for Run 2 period.

	2010	2011	2012	2017	Nominal
Bunch spacing [ns]	150	50	50	25	25
Number of bunches	368	1380	1380	2736	2808
β^* [for ATLAS and CMS]	3.5	1.0	0.6	0.4	0.55
Max bunch intensity [protons/bunch]	$1.2 \cdot 10^{11}$	$1.45 \cdot 10^{11}$	$1.7 \cdot 10^{11}$	$1.25 \cdot 10^{11}$	$1.14 \cdot 10^{11}$
ε_n [mm.mrad]	≈ 2.0	≈ 2.4	≈ 2.5	≈ 3.2	3.75
Peak luminosity [$cm^{-2}s^{-1}$]	$2.1 \cdot 10^{32}$	$3.7 \cdot 10^{33}$	$7.7 \cdot 10^{33}$	$1.4 \cdot 10^{34}$	$1.0 \cdot 10^{34}$

Table 4.1: The values of the LHC machine parameters during the 2010, 2011, 2012 and 2017 data taking period. The last colunmn represents the values for the same parameters for which the machine was designed to operate.

4.2 ATLAS detector

The ATLAS (A Toroidal Lhc ApparatuS) [24] is a multi-purpose detector which is designed to be sensitive in the widest possible range of potential new physics, as well as a platform to extend the understanding of our current knowledge. It covers a broad spectrum of physics studies that can be done through the LHC collisions, from precision measurements of the Standard Model and the Higgs boson discovery up to New Physics studies and Beyond Standard Model theories testing.

The detector 's length is 44 m and has a diameter of 25 m. It has a total weight of 7000 tons. The beam pipe is placed in the center and all the sub-detectors have been built in a cylundrical symmetry in concentric layers (like an onion) around it covering a full ϕ of 360°. Beginning from the center and directing towards the external layers, the first part is the Inner Detector (ID) [25, 26] that engulfs the beam pipe. Its task is the reconstruction of the trajectories of the charged particles produced by the collisions. ID can measure precisely the charged particles' momenta due to the fact it is immersed in a magnetic field generated from the magnet system [27]. The next layer consists of the calorimetry system. This is an electromagnetic and a hadronic calorimeter for measuring the energy of the particles which interact either electromagnetically or hadronically. They interact with the calorimeters material and they stop inside them while depositing their energy. Through the energy calculation an estimation of the total energy of the particles can be extracted. Finally, the last cylindrical layer of the ATLAS detector is the muon spectrometer [28]. As it becomes obvious from the name, it is absolutely specialized into muon leptons. It captures the muons coming from the verteces of interest.

Besides the sub-detectors from which ATLAS consists of, another very important and crucial sub-system is that of Trigger and Data Acquisition system (TDAQ) [29]. More than a billion collisions are produced in the heart of ATLAS every second. Each of these collisions produce many particles each one of which has a specific trajectory, energy and many other details necessary for the reconstruction of each event. In order for a collision to be recorded for further analysis, it must fullfill some very specific and elementary criteria. If not, it is just an extra burden and it should be rejected immediately. In a different case, the amount of collected information would be unimaginably enormous making every attempt of later data processing prohibitive. For this reason, the vast majority of the produced collisions is been indeed rejected and never recorded in ATLAS storage systems. TDAQ is responsible for this selection (triggering) and record of all interesting events (data aqcuisition) by using signals from certain detector sub-systems and redirecting the data flow towards the permanent storage facilities.

Fig. 4.3 shows a cut-away view of the ATLAS detector with its geometrical dimensions of the full system, as well as its described subsystems.

In the coordinate system which is used throughout ATLAS the interaction point (IP), the epicenter of the event, is defined as its origin. The beam pipe is defined as the z-axis and the transverse x-y plane is perpendicular to it. One of the most common characteristic property of the particles is the momenta measured in the x-y plane and for this reason is called as transverse momenta. It is denoted as p_T. The positive x-axis points towards the center of the LHC. In an analogous way, the positive of the y-axis points towards the surface of the Earth. ATLAS detector

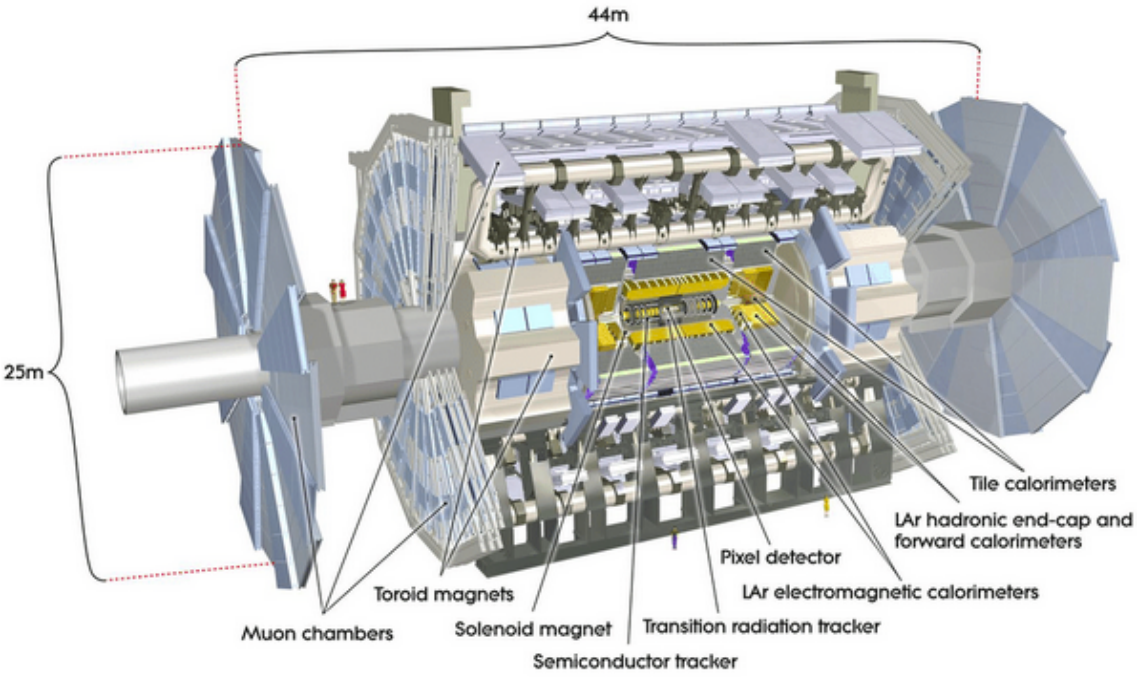


Figure 4.3: A cut-away view of the full ATLAS detector.

can be devided to three parts. The Side A and C which are the two end-caps and the cylindrical Side B. It can be pictured like a barrel layed on its cylindrical side. The convention is that positive z-axis points towards the A end-cap and, of course, negative z-axis towards C end-cap. Given the cylindrical symmetry of the whole construction it is much more convenient that the transverse plane to be described in r-ϕ coordinates where ϕ is the azimuthal angle measured from the x-axis around the beam pipe and r is the radial dimension measuring the distance from it. In this scheme the polar angle θ is defined as the angle starting from the positive z-axis. However, the usual thing is that the polar angle is reported in terms of pseudorapidity $\eta = -\ell n \tan(\frac{\theta}{2})$. Therefore, the solid angle ΔR in the η-ϕ space is defined as $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. Practically, η and ϕ are the two most common coordinates used in both ATLAS and CMS in order to express the way the particles-debris of a collision are been distributed in the cylindrical coordinates system. Fig. 4.4 presents a vertical cross-section of ATLAS detector where all the different sub-detectors and the dedicated particles they detect appear.

4.2.1 Magnet system

Four large superconducting magnets constitute the ATLAS magnetic system. They are designed that way in order to provide a magnetic field mostly orthogonal to the particle trajectory. In Fig. 4.5 the magnets can be seen. The ATLAS magnet system consists of three parts: the barrel toroid (red lines), the end-caps toroid (green lines) and the central solenoid (CS) magnet (blue cylinder). All the toroids are air-core and they develope a magnetic field of 4.1 T. On the other hand, the CS developes a magnetic field of 2 T.

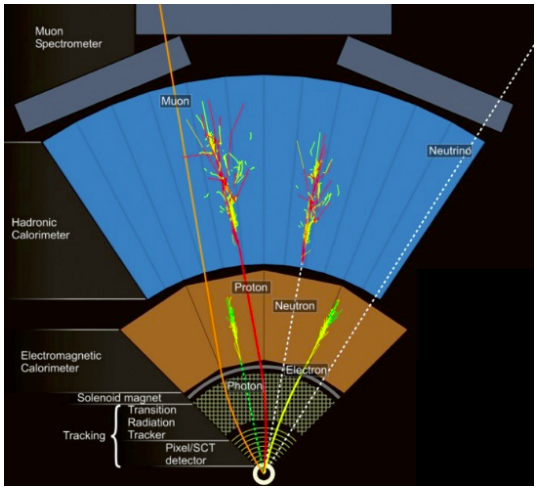


Figure 4.4: A vertical cross-section of ATLAS detector.

The end-cap toroids are rotated by 22.5° with respect to the central barrel in order to improve the overlap of the respective magnetic fields and achieve higher uniformity. The CS is 5.3 m long with a diameter of 2.4 m, while the end-caps are 5 m long and have an external diameter of 10.7 m and an internal bore of 1.65 m.

This structure has many advantages in extending the coverage of the Muon Spectrometer at larger pseudorapidity values (particles very close to z-axis), as well as having no magnetic field inside the calorimetry system and, thus, not degrading its performance.

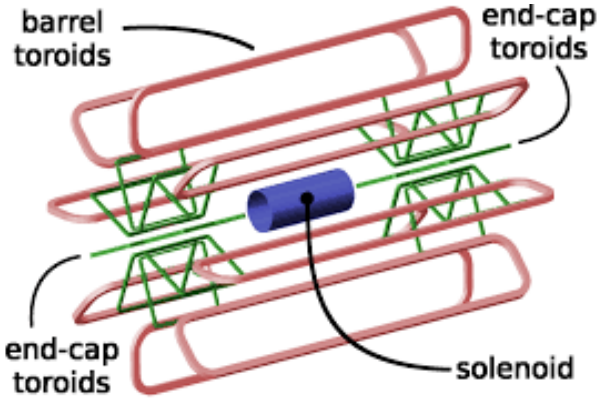


Figure 4.5: Sketch of the ATLAS magnet system (solenoid and torroids).

4.2.2 Inner detector

Starting from the inside of ATLAS, the first component is the Inner Detector. The ID is designed to efficiently reconstruct the trajectories of charged particles in $|\eta| < 2.5$ and for p_T from 500 MeV up to few TeV. In addition, it has the capability of acquiring the track impact parameters from primary and secondary vertices allows for the identification of short-lived particles. By primary vertex it is ment the pp

IP and by secondary vertices the points where produced unstable particles decay to other stable particles.

The ID has a cylindrical shape 7 m long with its center at the IP, its radius is 1.15 m and is fully contained within the magnetic field provided by the solenoid described previously. It consists of three detector technologies: a Pixel detector, which is the innermost detector, a SemiConductor Tracker (SCT) and a Transition Radiation Tracker (TRT). The first two technologies of the ID consist of radiation-hard semiconductor based detectors that provide high precision measurements. Radiation-hard materials are the materials that appear small radiation fatigue. The TRT technology is based on drift tubes. These are tubes containing a stretched wire within a gas volume. When a charged particle passes through the volume it knocks electrons off the atoms of the gas. These follow the electric field ending up at the positively-charged wire. TRT allows for a larger number of measurements to be acquired compared to the few layers that the first two have, but with lower intrinsic resolution.

All three technologies have a barrel part, organized in concentric layers, and two end-cap counterparts made of disks that are arranged orthogonal to the beam direction on each side of the barrel. Fig. 4.6 shows a cut-open view of the ID in the barrel and end-cap regions. The total achieved resolution from the information from all ID sub-detectors combined on the track momentum is $\frac{\sigma_{p_T}}{p_T} = 0.05\% \mathrm{p}_T$ (GeV)$\oplus$0.1%. The Trigger sub-system is been fed with p_T from ID sub-systems signals and used as input for selecting only interesting events.

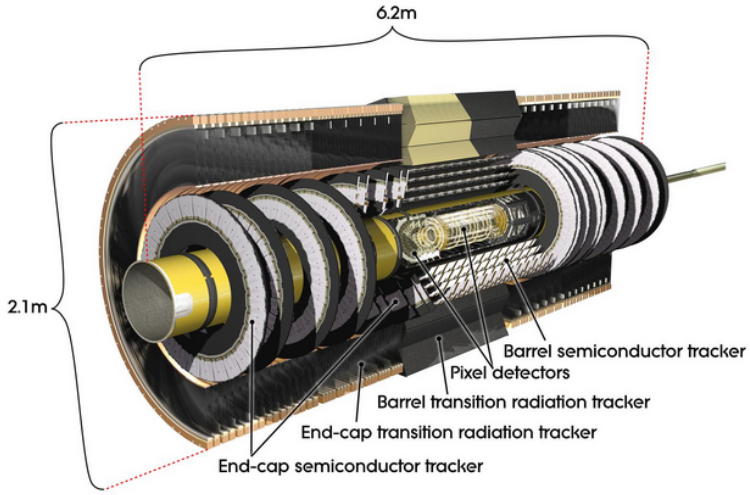


Figure 4.6: Cut-open view of the Inner Detector displaying the barrel and end-cap components of different sub-detectors and tracking elements.

- **Pixel detector**: Pixel detector offers two dimensional spatial information based on silicon chip technology segmented in a matrix of pixels. It is located at the core of the inner detector, closest to the nominal interaction point.The size of each silicon pixel is of 50×400 $\mu\mathrm{m}^2$ with a thickness of 250 μm, with its shortest direction aligned in the direction of the bending plane of the particle,

which ensures the best performance. The basic module of a Pixel sensor is a plane consisting of 144×328 pixels. In total about 1800 modules are assembled in the barrel and end-cap regions summing up to about 80 millions channels. The readout chips typically have large areas and each element are equipped with individual circuits. They are also radiation hardened to withstand over 300 kGy[2] of ionizing radiation, as well as over $5 \cdot 10^{14}$ neutrons per cm^2 over 10 years of operation.

As mentioned, the detector has both a barrel and an end-cap part. In the central barrel region, the modules are organized in three cylinders of length of 80 cm and radii of 5.05 cm (called B-layer), 8.85 cm (called layer 1) and 12.25 cm (called layer 2). In the end-caps, the modules are organized in disks of 34 cm radius and placed at $|z| = 49.5$, 58.0 and 65.0 cm from the center of ATLAS.

The high granularity of the Pixel detector contributes in the reconstruction of tracks, since it provides at least three (one per layer) precise hits close to the interaction point. These hits are of utmost importance for the reconstruction of the primary vertices by measuring the impact parameter of the tracks which are defined as the minimum distance of the track to the primary vertex. It also determines the ID 's ability to detect short lived particles such as B hadrons and τ leptons. Overall, the pixel detector offers an intrinsic accuracy of 10 μm in r-ϕ and 115 μm in z.

- **SemiConductor Tracker**: following Pixel detector, this is the second layer of ID. It is based on silicon technology and it consists of four concentric layers in the barrel region covering a range of $|\eta| < 1.4$ and of nine disks in each end-cap region spanning $1.1 < |\eta| < 2.5$. Each one of these layers consists of two detector modules of micro-strip detectors mounted back-to-back and placed with the micro-strips tilted by a small angle ($\sim$ 40 mrad) to the magnetic field (and to the beam line) in order to provide three dimensional information. The double barrel layers are mounted at radial distances of 30.0, 37.3, 44.7 and 52.0 cm. The end-cap SCT disks, consisting of modules organized radially, are mounted at z distances between $\pm$85 cm and 272 cm.

  The SCT has a total number of 15392 sensors organized in 4088 modules. Each module measures 6.36×6.40 cm^2 with a thickness of 350 μm and holds 780 strips with 80 μm strip pitch. This results in about 6.3 million readout channels.

  The SCT is highly granular and can achieve a spatial resolution of 17 μm in the strip pitch direction and of 580 μm in the z direction. The three dimensional spatial information for (at least) four points per track over the full ID coverage helps with the track reconstruction in term of increased impact parameter resolution, as well as the measurement of the momentum. Both Pixel detector and SCT are designed using materials with the lowest possible coefficient of thermal expansion, to account for issues such as operation in cryogenic temperatures, removal of heat generated by electronics, as well as detector leakage current. Prior to the LHC operation, prototypes of these modules had been tested in magnetic fields with test beam conditions and have successfully demonstrated the required performance in resolution, signal to noise, as well as speed. Modules containing detectors and their corresponding electronics have also been

[2]Gray (Gy) is an SI unit of ionizing radiation dose; it is defined as the absorption of 1 Joule of radiation energy per kg of matter

tested with the expected level of radiation during LHC operation and have shown to perform successfully within specifications.

- **Transition Radiation Tracker**: TRT is the outermost sub-detector of ID complex. It is based on the multi-wire proportional chamber technology utilizing drift tubes. This provides a relatively cheap, radiation-hard solution to extend the continuous tracking, both in barrel and end-cap regions. The barrel section is constructed out of 5×10^4 straws of 114 cm length and 4 mm diameter, parallel to the beam axis, from a radius 56 cm, up to 107 cm, thus covering a region of $|\eta| < 0.7$. In addition, each end-cap section consists of 32×10^4 straws of length of 37 cm oriented in the radial direction, at z distances between $\pm$84 cm and 271 cm, covering $0.7 < |\eta| < 2.5$.

  Each of the straws is filled with a gas mixture of 70% Xe, 20% $C0_2$, 10% CF_4. The space between them is filled with polyethylene that helps to produce the transition radiation; the photon emission depends on the Lorentz boost of the charged particles, which at the energies available, is only present for electrons. Measuring the drift time between straws, the TRT adds on average 36 hits in the collection of points from Pixel and SCT, with a spatial resolution of $\sim$ 130 μm.

  The sub-detector, during signal processing, can make use of two different thresholds. The high threshold signals are used to identify photons that arise from transition radiation. Therefore, apart from aiding the tracking of charged particles, the TRT can act as a direct electron identification tool, to separate electrons from the pool of the other charged particles. The continuous tracking component of the TRT detector assures a good performance on pattern recognition. Spacing of straws are optimized for tracking within the available radial space and the straws are distributed over maximum possible path lengths. This reduces the effect of loopers and interactions that can saturate small regions of the detector, as well as enhancing the performance of pattern recognition.

4.2.3 Calorimeter system

The second layer of ATLAS "onion" is ATLAS calorimeter system. It consists of an inner electromagnetic calorimeter which is surrounded by a hadronic calorimeter. Both are symmetric in ϕ, fully encircling the region around the beam line. The two systems are divided in a central part and two symmetric end-caps allowing for a pseudorapidity coverage of $|\eta| < 5$. The central electromagnetic calorimeter has a finer segmentation compared to the hadronic and the forward (end-cap) calorimeters, especially in the 17 region covered by the ID. This provides better a precision measurement of the photons and electrons, while still allowing for a precise measurement of the kinematic properties of jets and the calculation of the missing energy. Fig. 4.7 illustrates the cut-away view of the calorimetry system, on which the different sub-systems are explicitly shown. Signals from the calorimetry system are also redirected during data taking to the Trigger system providing input to the algorithms that identify interesting events on-the-fly.

- **Electromagnetic Calorimeter (ECAL)**: a liquid Argon (LAr) sampling detector with accordion shaped electrodes and 1.8 mm thick lead plates operates as

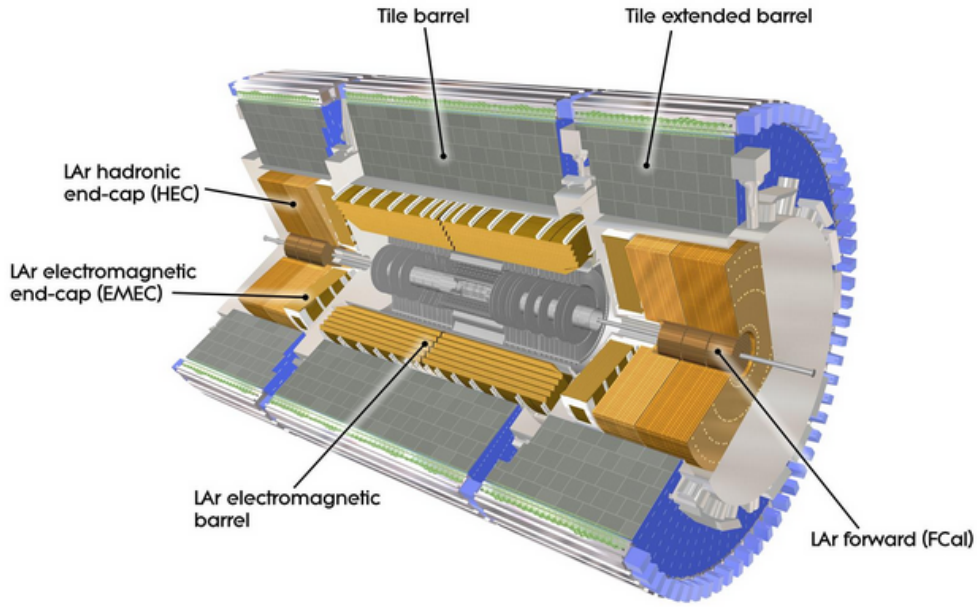


Figure 4.7: A cut-away view of the ATLAS barrel calorimeter system.

absorber. The LAr has been chosen due to its linear behaviour in high ionization yield as well as its stable and radiation resistant characteristics, The shape of the accordion electrodes provides full ϕ symmetry without any azimuthal cracks. The material and thickness of the absorber has been optimized to allow for low cost while being fully efficient and providing high energy resolution. Therefore, the total thickness of the ECAL depends on η and increases from 22 to 33 radiation lengths in the barrel and from 24 to 38 radiation lengths in the end-cap, which allows for the full containment of electron and photons with p_T up to a few TeV. The energy resolution provided by the ECAL is $\sigma_E/E = 9\%/\sqrt{E} \oplus 0.3\%$, with the energy in GeV.

The barrel part of the ECAL ($|\eta| < 1.5$) consists of two identical "half"-barrels with a small gap of few mm at z = 0, while each end-cap is divided into two co-axial cylinders covering the region of $1.375 < |\eta| < 3.2$. In the central region ($|\eta| < 1.8$), a pre-sample layer is present, the signals from which are used to calibrate the energy losses from electrons and photons due to their interactions with the material between the ECAL and the IP. This layer is in principle an active liquid argon layer of 1.1 cm thickness in the barrel region and 0.5 cm thickness in the end-cap.

The special characteristic of the ECAL is the segmentation (Fig. 4.8) of its different longitudinal layers. This segmentation provides an excellent granularity allowing for precise energy and spatial resolution of different energies. The first layer is about four radiation lengths thick and the finely segmented readout of the absorber is in the η direction with strips of thickness $\Delta\eta = 0.003$, the aim of which is the precise measurement of the shower properties and is important for photon and electron identification. The middle layer of the ECAL contains most of the thickness of the calorimeter and is divided in squared cells in the η-ϕ plane with size $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$. The last layer is coarsely seg-

mented and the information that provides is used to estimate the energy loss beyond the ECAL.

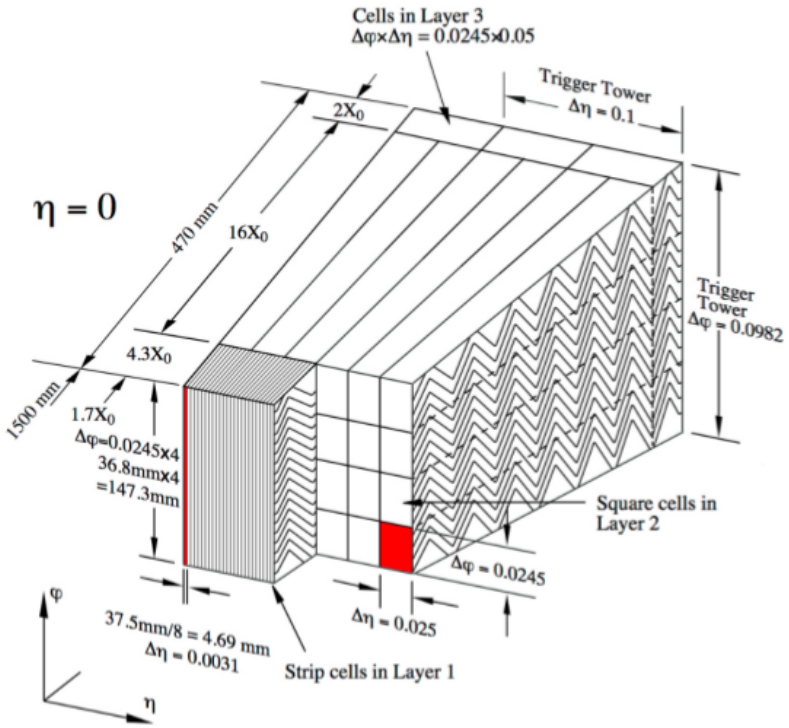


Figure 4.8: Schematic of the segmentation of the EM calorimeter.

- **Hadronic Calorimeter**: in contrast to the homogeneous ECAL, the hadronic calorimeter consists of independent sampling calorimeters following different technologies. This comes as a result of the fact that at different η regions the energy flux and the performance requirements vary. HCAL 's main role is the identification of hadrons (p, n, π) and jets.

  1. A Tile Calorimeter (TILE) is used in the central region covering the region $0 < |\eta| < 1.7$. It is structured as a cylinder of 11.4 m length and inner and outer radii of 2.28 m and 4.25 m respectively. It follows the sampling calorimetry technology with iron tiles as absorber and plastic scintillators as sensitive material. The Tile Calorimeter is organized in periodic towers in z and ϕ direction.
  2. A Hadronic End-cap Calorimeter (HEC) covers $1.5 < |\eta| < 3.2$ using copper plates as absorber and LAr as active material, due to the high radiation resistance of LAr.
  3. A Forward Calorimeter (FCAL) is used in the most forward region covering $3.1 < |\eta| < 4.9$. The passive material of this calorimeter is tungsten rods embedded in a matrix of copper, while again LAr is used as the active material.

The depth of the hadronic calorimeter is 7.4 interaction lengths λ. This allows a precise measurement of the missing energy arising from undetected particles (neutrinos), as well a shielding for the Muon Spectrometer from the rest of the detector, allowing only for muons with at least p_T of a few GeV to reach it.

The resolution achieved is $\sigma_E/E = {}^{50\%}/\sqrt{E} \oplus 3\%$ for the TILE and HEC and $\sigma_E/E = {}^{100\%}/\sqrt{E} \oplus 10\%$ for the FCAL.

4.2.4 Muon spectrometer

The last layer of ATLAS detector is Muon Spectrometer (MS). This is the outermost layer of ATLAS detector. It covers the region of $|\eta| < 2.7$ and its goal is to provide precise measurements of the muon momenta up to the high energy of few TeV with the help of the magnetic field generated by the toroids (as described in Sec. 4.2.1) to curve the muon trajectories. Fig. 4.9 shows the cut-away view of the ATLAS detector with all the sub-systems that constitute MS. The MS can operate as a stand-alone trigger sub-system providing in this way crucial input to the Trigger sub-system during data taking. The MS sub-systems consist of four technologies providing precision measurements, as well as trigger capabilities. The MS is structured in three regions, covering different η ranges: a barrel ($|\eta| < 1.05$) and two end-caps ($1.05 < |\eta| < 2.7$).

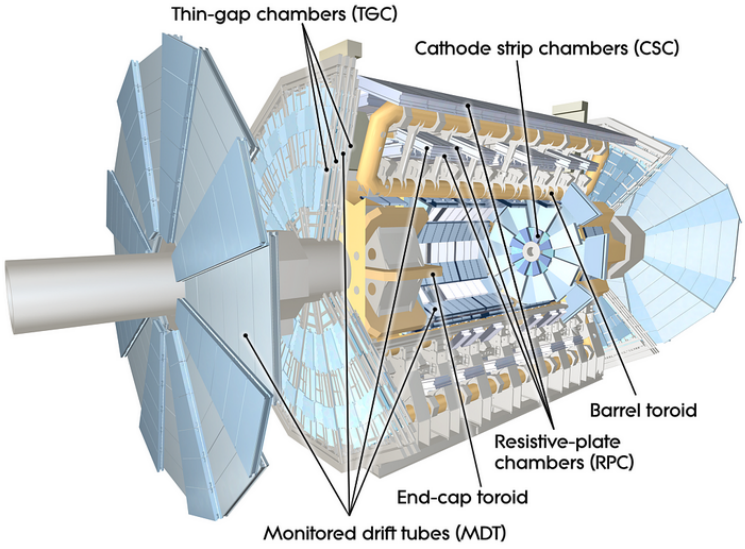


Figure 4.9: A cut-away view of the ATLAS Muon Spectrometer.

Triggering Chambers

- **Resistive Plate Chambers (RPC)**: RPC are located in the main body (barrel region) of the MS covering up to $|\eta| = 1.05$. The chambers consist of two parallel high resistive electrode plates that are separated by insulating spacers, creating a gap that is filled with a gas mixture of 94.7% $C_2H_2F_4$, 5% iC_4H_{10}, 0.3% SF_5. The chambers operate with an electric field of ~ 4.9 kV/mm and the readout of the plates is done through strips mounted at the back of the plates, orhogonal to each other, thus providing η-ϕ measurement.

RPC perform excellently as a fast trigger system in the barrel and end-cap regions providing a fast momentum estimation to the hardware-based trigger within 2 ns. In addition, they provide measurements in the azimuthal coordinate ϕ with a spatial resolution of $\sim$ 10 mm, bunch crossing identification and well defined p_T thresholds for triggering.

- **Thin Gap Chambers (TGC)**: TGC are mounted on both end-caps of MS covering the region $1.05 < |\eta| < 2.7$ (2.4 for triggering). There are totally four layers. They are multi-wire proportional chambers with the characteristic property that the distance between the wire and the cathode is smaller than the distance between wires. By proportional it is meant that the signal gathered is analogous to the gas ionization from the detected particles. The reason for this structure is to achieve a fast charge collection time. A wire of 50 μm diameter is placed between the two graphite cathode planes. The gas gap of 2.8 mm is filled with a mixture of 55% $C0_2$, 45% nC_5H_{12}. An operational voltage of 2.9 kV is applied on the anode wires which give the chambers signals.

  TGC provide a timing resolution comparable to the RPC and a spatial one varying in the range of 2-7 mm for both the radial coordinate and the one orthogonal to it.

Tracking Chambers

- **Cathode Strip Chambers (CSC)**: CSC are on the first layer of the end-caps of MS. There are located only in the end-caps and not in the barrel. They cover a region of $2.0 < |\eta| < 2.7$, the region very close to the beam axis. They are installed in the first layer of MS, because in this region the counting rate that can be handled by the counter-part sub-system, the Monitored Drift Tubes (MDT), is exceeded. Thus, they provide a safe operation with counting rates of about 1 kHz/cm^2.

  CSC are multi-wire proportional chambers with wires oriented in the radial direction. The chambers have two cathodes segmented in strips with orthogonal directions. CSC are filled with a gas mixture of 80% Ar, 20% $C0_2$ and they have an anode pitch of 2.5 mm and a cathode strip readout strip pitch of $\sim$ 5.308-5.567 mm. By interpolating the charge between the different strips, a precise measurement of the ϕ coordinate can be achieved.

  For the collection of the signal an upper limit of 40 ns is set for the drift time. This allows for very high acquisition rates in the η range they cover. The spatial resolution achieved is 40 μm in the bending plane as well as $\sim$ 5 mm in the non-bending one. A more thorough analysis of CSC sub-detectors follows in Ch. 5.

- **Monitored Drift Tube Chambers (MDT)**: MDT sub-system is mounted on both barrel and end-cap regions and positioned in the ϕ direction. They are the majority of MS sub-detectors. Tracking in the MS is based on the magnetic detection of muon tracks in large superconducting air-core toroid magnets. Over the range $|\eta| \leq 1.0$, magnetic bending is provided by the large barrel toroid which provides a field integral between 2 and 6 Tm. For $1.4 \leq |\eta| \leq 2.7$ muon tracks are bent by two smaller end-cap toroid magnets contributing between 4

and 8 Tm. The bending in the region $1.0 \leq |\eta| \leq 1.4$, usually referred to as the transition region, is provided by a combination of barrel and end-cap fields.

Each chamber contains layers of aluminium drift tubes with a diameter of 29.970 mm. They are filled with a gas mixture of 93% Ar, 7% CO_2 at a pressure of 3 bar. The high-energy muons ionize the gas mixture in the tube and the electrons from the ionisation are collected at the central wire made from tungsten-rhenium (W-Re). The wire has a diameter of 50 μm. It is at a high voltage of 3080 V as delivered by the Power Supply (PS) System of the MDT. The cathode wire is held in place by a cylindrical end-plug, as shown in Fig. 4.10. The transfer of power from the PS to the chamber and from the signal of the chamber to the readout chain are placed at opposite ends of the tube.

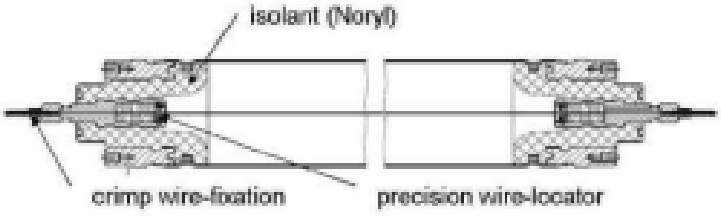


Figure 4.10: Cut-away view of the inside of a Monitored Drift Tube.

Most of the MDT chambers consist of two Multi-Layers (ML), each having four rows of tubes for the innermost chambers and three rows of tubes for the rest. Depending on the relative position of the chamber in MS, the length of the tubes may vary between 0.7 and 6.3 m, while to maintain a constant resolution, a mechanical spacer is used for fixing the position of the layers. Fig. 4.11 shows a schematic of a MDT multi-layer consisting of rows of tubes glued on the supporting frame.

Due to their robustness and simple operation, the MDT chambers cover the larger area of the MS and they are used to measure the coordinate in the bending plane with a precision of 35 μm.

Overall, in the barrel region, the muon tracks are measured by the MDT and the RPC sub-systems with the chambers assembled in three concentric layers. The MDT provides the η coordinate and the RPC provides the ϕ coordinate together with timing information. In the end-cap region, MDT and CSC sub-systems provide the η coordinate, while the ϕ coordinate is taken from the TGC sub-system.

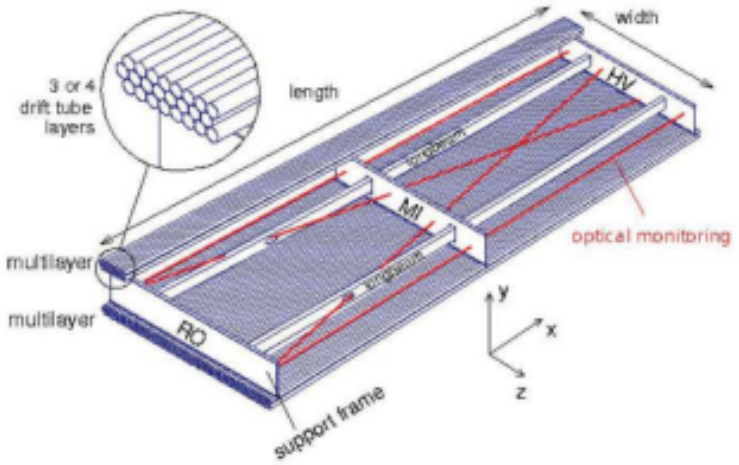


Figure 4.11: An MDT Multi-Layer consisting of three or four rows of tubes, glued on a support frame.

Apart from the combined operation mode with the rest of ATLAS, MS is designed with the capability to measure in stand-alone mode muons with p_T from 3 GeV up to a few TeV with a momentum resolution $\sim 10\%$ at $p_T = 1$ TeV. At lower energies, the resolution improves if MS is operated in combined mode with the ATLAS ID, reconstructing the full information from tracks in MS and ID.

Most of the sub-systems of the ATLAS detector have undergone major upgrades during the LS1, between the first and the second period of data taking. Furthermore, plans have been made for the possible upgrades in the future. The MS is not an exception to this.

4.2.5 Data acquisition and trigger system

Trigger and Data Acquisition (TDAQ) system of ATLAS is of great importance, since this is the system in charge of the real time selection (trigger) and permanently storage for later analysis only of the events that contain interesting characteristics. This procedure is crucial because of the vast amount of information. For a given luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$, the IP bunch crossing rate is 40 MHz with an average of 25 collisions per bunch crossing. On the other hand, ATLAS storage capacity for offline data processing is only up to 200-300 Hz or $\approx$ 300 MB/s.

The procedure of real time data rejection of uniteresting collisions is achieved through a three-level decision chain. In each level, the rate is reduced either by fine-tuning the decisions made or applying selections on the output of the previous level. The first level (L1) is hardware-based. Its input is received signals from $\approx 10^6$ RPC and TGC channels and $\approx$ 7200 calorimeter trigger towers. Each of these L1 sub-systems have the task to identify certain detector signals that could be associated to physics objects, like photons and high p_T electrons, muons, τ leptons decaying to hadrons and undetected ("missing") transverse energy. The last one is been associated with neutrinos which they almost do not interact at all with any of the sub-detectors. The combination of these signals, mainly by demanding to coincide of by veto, results to L1 triggering. This way Regions of Interest (RoI) are

created. The result is that the initial 40 MHz is limited down to 75 kHz. All the events that have survived L1 are been through the read out of all sub-systems and the information is stored temporarily into Readout Buffers (ROB). Intermediate buffers ("derandomisers") are also used in order data rate is such that the readout drivers (ROD) can handle it. The necessity for the derandomisers is due to the fact that some electronics return multiplex signals.

The second step of TDAQ is the Level 2 (L2). All the remaining data at ROB are been rejected or passed to the next level. L2 is a software-based system which uses specific algorithms in order to perform reconstruction within RoIs existing from L1. L2 is been fed with the ROB information like transverse momentum p_T, energy sums and posistion of the candidate particles and narrows down the rate to $\approx$ 3 kHz. In case that an event is rejected by L2, the data stored are discarded as well. If not, then the data stored in ROBs are transferred by data acquisition (DAQ) system to sub-farm input (SFI) storage. This storage is accessible only by the third level of triggering, the Event Filter (EF) processing units.

EF is also a software-based selection level. Offline algorithms seeded by the information passed from the L2 run on computing farms to reconstruct the selected data and also performing refined alignment and calibration. The final output are transferred for permanent storage. The output data rate at this level is reduced to about 200 Hz with an average recording rate of $\approx$ 300 MB/s.

The three levels of the Trigger and Data Acquisition system along with the designed input and recording rates at each level are shown in Fig. 4.12.

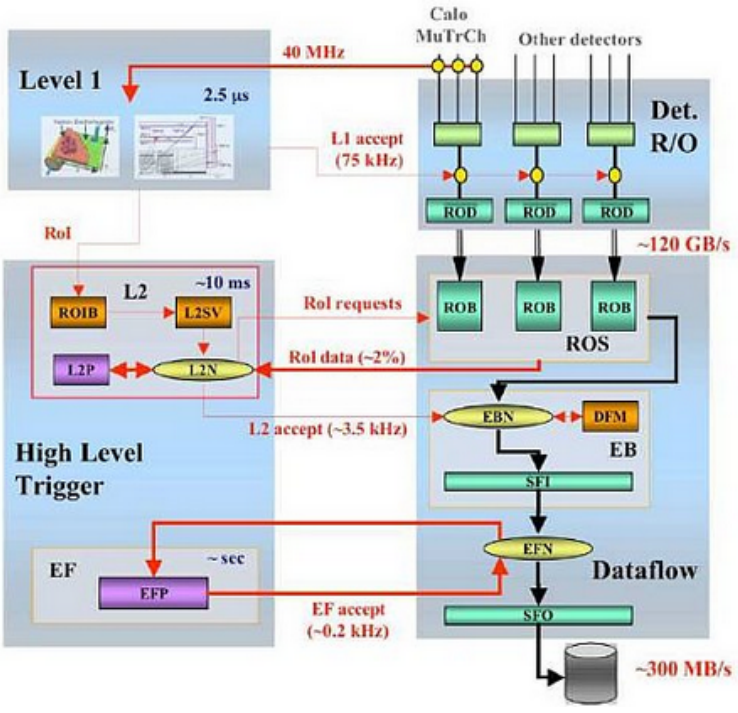


Figure 4.12: A schematic overview of the ATLAS Trigger and Data Acquisition System.

Chapter 5

ATLAS Muon Spectrometer CSC Subdetectors Monitoring

In the present chapter offline handling of data recorded from the CSC (Cathode Strip Chambers) is been described [30]. During 2015 and 2016, my main task in CSC sub-detectors along with Ilias Panagoulias was the mantainance of their monitoring plots and algoritms and, thus, securing their proper function. This work was recognised on behalf of the experiment as our ATLAS Qualification Task resulting to our official ATLAS collaboration membership since 2016.

The chambers cover the forward and backward η regions from 2.0 to 2.7 and serve as the first muon station over that η range (Fig. 5.1).

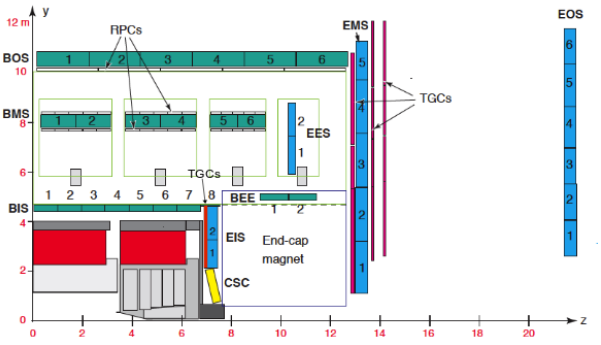


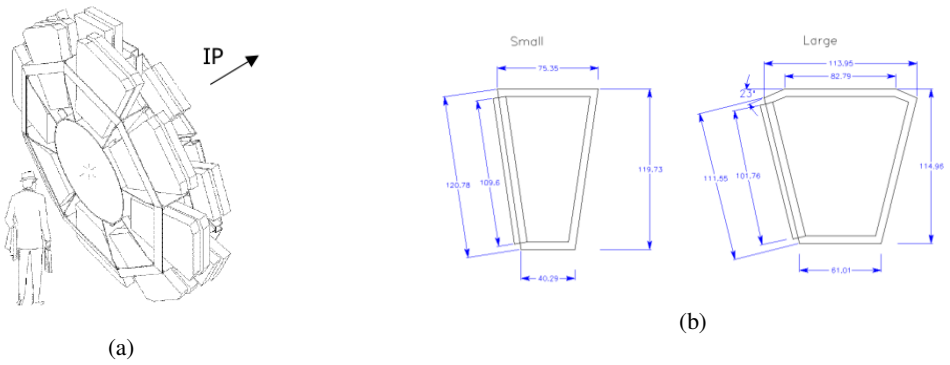
Figure 5.1: Side view of the present ATLAS muon spectrometer layout. CSC sub-detectors appear with yellow.

5.1 Detector characteristics

5.1.1 Geometry and readout

The CSC detector consists of 32 chambers, 16 on each end-cap (Fig. 5.2a). The 16 are arranged in eight ϕ-sectors on each of two wheels with an azimuthal overlap to provide hermetic coverage in ϕ. The two wheels are named CSS (*Small Chambers) and CSL (*Large chambers) (Fig. 5.2b) following the overall ATLAS geometry of and a chamber is identified by specifying the wheel, η-side (- or +) and ϕ-sector

(1-8)[1]. There are four planes of wires within each chamber with orthogonal strips on the two cathode planes providing measurements in the η and ϕ directions. Thus a readout plane may be identified by the wire plane number (1-4) and orientation (η, ϕ). Strips are numbered consecutively starting from one in each plane. There are 192 strips in the η planes and 48 in the ϕ planes.



(a) (b)

Figure 5.2: (a) CSC sub-detectors array of each end-cap, (b) small and large CSC chambers dimensions.

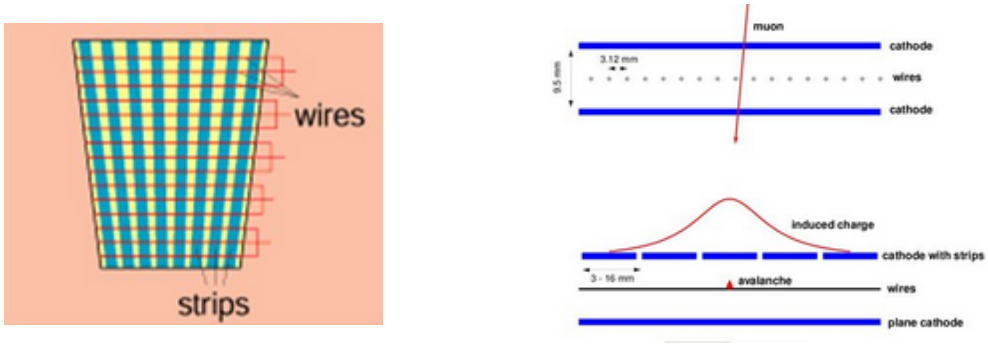
The η-strips are relatively narrow ($\sim$ 5.6 mm for the Small Chambers and $\sim$ 5.3 mm for the Large) so that the signal induced by a single charged particle typical registers on three or more strips making it possible to distinguish isolated tracks from those overlapped by other tracks. The latter are called spoiled. In the former, unspoiled, cases (80-90% of the tracks) it is possible to use distribution of charges on adjacent strips to determine the track position to high precision, better than 100 μm. The ϕ-measuring strips are much wider ($\sim$ 12.9 mm for the Small chambers, and $\sim$ 21.0 mm for the Large) and the signal from a single track is generally confined to one or two strips, making this precision measurement less significant. The spoiled η and most or all of the ϕ measurements have a resolution of a few mm. For this reason, the η coordinate is called the precision coordinate.

In Fig. 5.3b appears a cross section of a CSC chamber depicting the strips and wires array, as well as an example of a random muon trajectory. The muon is been accelerated by the existing electric field created by cathode strips and anode wires and deposits its charge on an η-strip element. Typically, the charge distribution is expected to be Gaussian and, thus, there is one strip element with maximum charge deposition and a few adjacent to it with less charge (Fig. 5.4).

The readout for each strip includes ADC (Analog-to-Digital-Converter) sampling at regular intervals. Typically four samples are read out. The time distribution is used to determine the time at which the particle interacted with the detector. The shape of the signal may also be used to determine whether there are contributions from multiple particles at different times.

The default reconstruction includes steps to construct each of the following in turn: digits, 1D clusters, 2D segments (i.e. one position and one direction), and 4D segments. Different track finding and fitting algorithms may use any of these elements; most typical at present is to use the clusters.

[1]the chambers from the upper side of the imaginary horizontal line which devides the wheel in the middle define the positive η side and vice versa



(a) (b)

Figure 5.3: (a) CSC strips and wires creating a grid for muon position and direction measurement, (b) a cross section of the same grid.

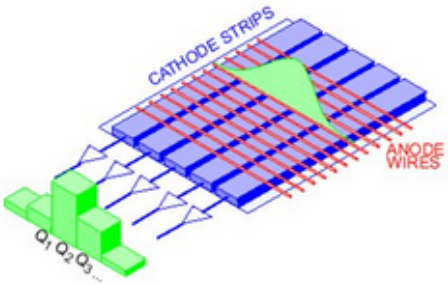


Figure 5.4: Gaussian distribution of charge deposition on strip.

5.1.2 Momentum resolution

The CSC radial (η) position measurement is essential to the muon spectrometer momentum measurement in the region $|\eta| > 2.0$. MDT's following the forward toroid can be used to predict the CSC position for a straight track and the deviation x from that prediction provides a measurement of the momentum. The toroid provides a field integral of approximately 6 Tm over a path of 4 m. The relation between the position deviation and momentum is approximately

$$x = H \cdot \frac{1}{p} \tag{5.1}$$

where H = 3.6 TeV · mm. For instance, a muon having a total momentum of p = 3.6 TeV (corresponding to p_T = 1.0 - 0.5 TeV for η = 2.0 - 2.7) will deviate by 1.0 mm from a straight trajectory. Thus, if we want 10% resolution for such muons (0.027 $\frac{1}{TeV}$ in $\frac{1}{p}$) for such muons we must keep the combined position resolution and misalignment under 100 μm.

5.1.3 Timing resolution

The CSC timing measurement is another important one especially for Event Filter (EF) trigger[2] and beam halo events veto. These events are fake (they are also called machine-induced background or MIB) and must be vetoed. MIB is produced through several interactions, but mainly through protons scattering off LHC collimators. MIB particles often arrive at a high radius and can increase the fake trigger rate of muon chambers. The time measurement is relative to the start of the readout window. This time is determined by the latency setting which is adjusted so that times peak in the middle of the window. We have 4 samples at 50 ns, covering 150 ns, so the center is at 75 ns. This is our T0 which will be subtracted.

Still, if the first sample is the largest and the interpolation indicates peaking before the first sample, then the time is set to 0 as a flag and as an aproximation. Same for the peaking times after the last sample.

If the trigger comes between two 50 ns samples, then we subtract 25 ns from the measured time because our window is shifted relative to the trigger by that amount. This is called trigger phase = 1. Unfortunately, the flag values of 0 and 150 are also shifted. As a result, we have valid times starting from -25 ns for phase 1. On the contrary, the flag of phase 0 is still at 0 ns.

Parabola interpolation of ADC samples followed by bipolar correction provides a time measurement of each channel (strip). For the cluster, we consider the measured time of peak charge strip as a time of cluster. Each 2D segment (η/ϕ) has 3 or 4 hits (clusters) associated. The unweighted average of the 3 or 4 cluster times can be used as a time of segment.

To use the time measurement correctly, T0 values should be subtracted. T0 is the average peaking time for segments from the IP. It is the sum of the T0 base value of the channel and 12.5 ns times the T0 phase of the readout fiber of the channel. The T0 base value depends on the shaping time of the analog electronics and is expected to be constant over time. The T0 phase, however, can change from one run to the next. It takes on the possible values of 0 or 1 and shifts the times by 0 or 12.5 ns. In one chamber, there are 5 T0 phases to be calibrated: 4 precision layers (η) and 1 non-precision layer (ϕ). Note that the 4 precision layers are however split in the middle and combined. This gives us channel 1-92 for layers 1(2) and 3(4) in one constant, channel 93-192 of layers 1(2) and 3(4) in the next, etc.

5.2 Data structures

5.2.1 Hits

Hits, class name CSCSimHit, record the simulated (GEANT) interaction of a particle with the CSC detector. They hold the identity of the particle, the starting and end positions of the GEANT step, the mean time and the energy deposited in that step.

5.2.2 Digits

Three separate classes are used to describe the response of a single strip: the original digit class CscDigit, the prep data digit CsctripPrepData and the RDO

[2]the ATLAS trigger system has 3 levels; the first (L1) is hardware-based using information from the calorimeter and MS, the second (L2) and third (EF) are software-based using information from all sub-detectors; L2 and EF together are the High Level Trigger (HLT)

description CscRawData. The original digit class holds the total charge and time and is presently used only to hold the output from the first stage of digitization. Apparently the time is always left zero. The RDO description is obtained either by running the second step of digitization or by conversion from bytestream, the format of the data read from the detector. The prep data digit is an unpacked and calibrated form of this data. Both RDO and prep data descriptions include the charge for each time sample. Although the latter are prep data objects and hence can be used as input for track reconstruction, the nature of the detector suggests it is preferable to first cluster the strips.

All digit containers hold a separate collection of digits for each chamber. The time samples for each strip are combined to a single measurement of charge and time, the "strip fit".

5.2.3 Clusters

A cluster is intended to provide a measurement of the coordinate where a charged particle crossed a wire plane. Existing algorithms construct 1D clusters which use information from one cathode plane to provide a single coordinate (i.e. that associated with the plane). In principle it is also possible to construct 2D clusters with include information from both planes and thus provide measurements of both coordinates but this is not currently done. We often drop the qualification when referencing 1D clusters.

Clusters are stored as prep data objects of type "CscPrepData". Each cluster holds identifiers for the digits from which it was constructed and a fit of the measured coordinate. The measurement is expressed as a 1D vector with a 1D error matrix with a surface specifying the orientation.

5.2.4 Segments

A CSC segment provides a measurement of position and direction within a chamber and is formed by combining one cluster from each of the planes in that chamber. Both 2D and 4D segments (Fig. 5.5) are supported: the former measure the position and direction for one orientation (η or ϕ) and the latter provide a complete measurement of both coordinates and directions at a specified plane, typically at the center of the chamber.

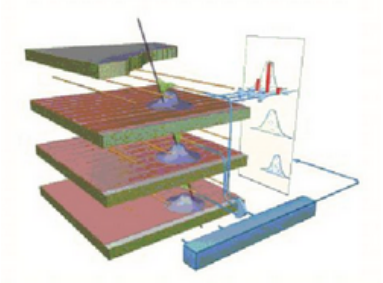


Figure 5.5: Graphic representation af a 4D segment consisting of four layers of CSC for the determination of muon 's direction.

The class describing CSC segments is the common muon segment class: MuonSegment. It holds the list of contributing prep data objects, typically clusters and

fit, i.e. the above measurements and their error matrix. The ordering is the same as that for the clusters.

The muon segments for a chamber are stored in a collection of type MuonSegmentCombination and these are held in the collection of type MuonSegmentCombinationCollection. The muon segment combination is a collection of vectors of muon segments. For 4D segments, a single vector is used and, for 2D segments, the ϕ and η segments are written to separate vectors if the algorithm looks for both. All is by convention or choice of the algorithm author. There is nothing in the classes to require or even encourage this organization. Unlike the digits and clusters, the current algorithms do not write empty containers (muon segment combinations) when no segments are found in a chamber. Instead one MuonSegmentCombination is written for each chamber that has one or more segments. If no segments are found in any chamber, then an empty MuonSegmentCombinationCollection should be written to affirm that the algorithm has been run.

5.3 Data quality monitoring plots

In the following section some plots for CSC data recorded monitoring are presented.

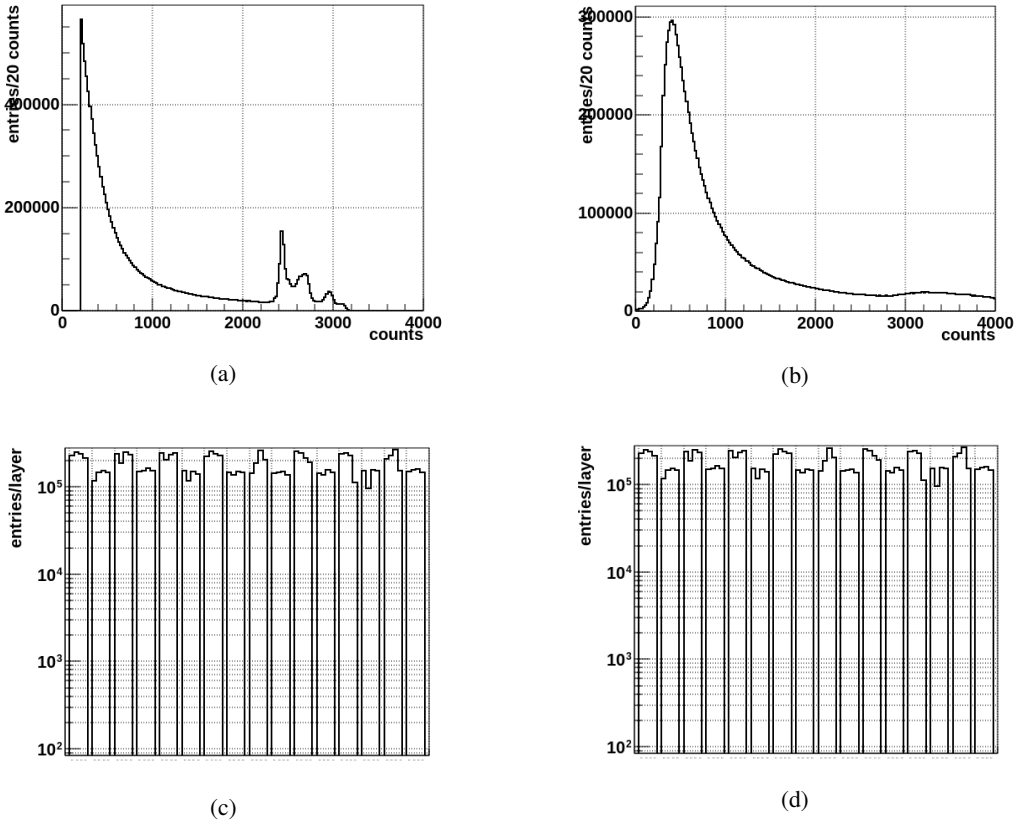
5.3.1 Cluster

- Fig. 5.6a: 1D maximum strip charge; secondary peaks at the high end tail (charge saturation at the upper limit of the ADC range)
- Fig. 5.6b: 1D total strip charge (Landau distribution)
- Fig. 5.6c: 1D maximum strip charge occupancy varying from large in turn to small chamber, each chamber consisting of four layers
- Fig. 5.6d: 1D total strip charge occupancy varying from large in turn to small chamber, each chamber consisting of four layers
- Fig. 5.7a: 2D maximum strip charge per layer and chamber
- Fig. 5.7b: 2D total strip charge per layer and chamber

5.3.2 Plot addition

As mentioned previously, our main CSC sub-detectors work was the mantainance of their monitoring plots and algoritms. More rarely, the addition of a new monitoring plot was considered as useful. A characteristic example is the following:

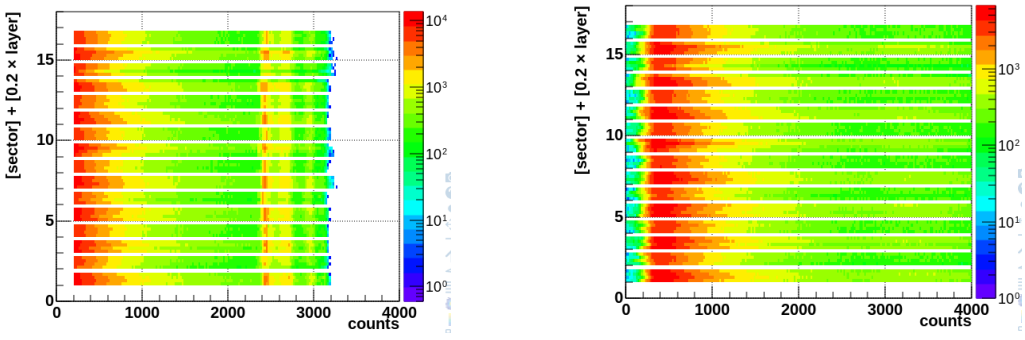
Fig. 5.8 depicts an 1D monitoring plot which concetrates all η- and ϕ-strips from all layers and chambers ([192 η-strips + 48 ϕ-strips] $\times$ 4 layers $\times$ 16 chambers per end-cap) in totally 15,360 x-axis channels and the corresponding y-axis cluster hits. The main idea behind these two plots (one for each ATLAS end-cap) is to make possible an easy and quick detection of a potential burst appearing in the form of non expected spike coming from any possible CSC malfunction. If this is the case, then a crosscheck with the rest of the plots can reveal the problematic η- or ϕ-strip, as well as the layer and the chamber it is located.



(a) (b)

(c) (d)

Figure 5.6



(a) (b)

Figure 5.7

5.3.3 Raw data

- Fig. 5.9a: 1D total signal occupancy, on the left of 0 of x-axis corresponds to 48 ϕ-strips per layer and on the right to 192 η-strips per layer

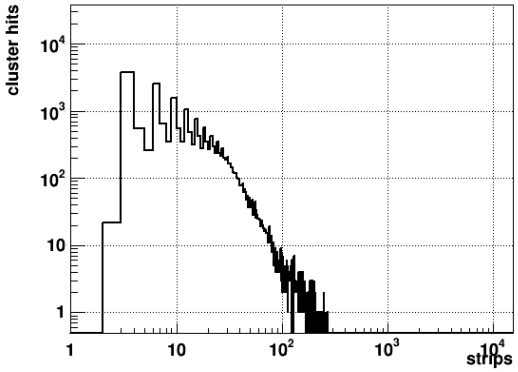
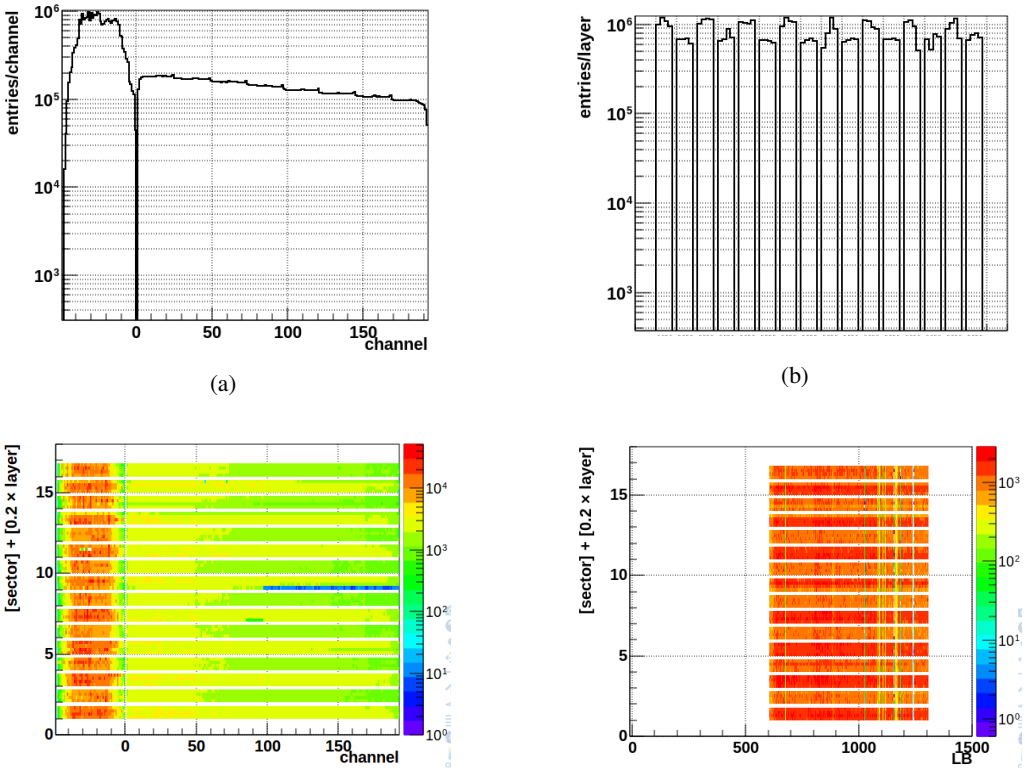


Figure 5.8



(a) (b)

(c) (d)

Figure 5.9

- Fig. 5.9b: 1D signal occupancy per layer and chamber
- Fig. 5.9c: 2D plot analogous to Fig. 5.9a
- Fig. 5.9d: 2D layer signal occupancy per lumi block (LB)

Chapter 6

LHC Data LQ Analysis at 8 TeV, Run I

This analysis [31] focuses on the search for scalar pair-produced second generation leptoquark (LQ) with missing transverse energy (E_T^{miss}). The missing energy denotes the existence of a neutrino which cannot be detected by any ATLAS sub-detector system. As a consequence, this is a search for final states with two leptons of the second generation and two jets. In particular, the considered topology consists of a pair of an LQ and an anti-LQ ($\overline{LQ}$). The first decays to a muon or a neutrino and a jet which comes from a c- or an s-quark through fragmentation and hadronization[1] (at high-energy colliders, jets of hadrons are the observable counterparts of the perturbative concepts of quarks and gluons). There is no specific preference in which LQ decays to particles or anti-particles. That means we can have as final state products either muons (μ^-) or antimuons (μ^+), neutrinos (ν_μ) or antineutrinos ($\overline{\nu}_\mu$) etc. Theoretically, both LQs of the pair have the same probability of decaying. This is expressed through the branching ratio (or branching fraction) β, thus $\beta = 0.5$ for the LQ pair (in principal β is an unknown quantity where $0 \leq \beta \leq 1$). By convention, $\beta = 1$ denotes the LQ decaying to a charged lepton and $\beta = 0$ that decaying to a neutral lepton. Of course, at the end of this analysis a full scan of β takes place constructing the β vs LQ mass phase space.

Previous LQ searches have been performed by ATLAS with the data collected at $\sqrt{s} = 7$ TeV and integrated luminosity $\mathscr{L} = 35$ pb^{-1}, excluding with 95% CL the existance of second generation LQs for masses below 422 GeV ($\beta = 0.5$) [32]. Later results from ATLAS for the same topology at $\sqrt{s} = 7$ TeV and $\mathscr{L} = 1.03$ fb^{-1} excluded LQ masses below 594 GeV [33].

This analysis uses proton - proton collision data at $\sqrt{s} = 8$ TeV corresponding to an integrated luminosity of 20.3 fb^{-1}, recorded by the ATLAS detector in 2012. Monte Carlo (MC) simulated samples of the LQ signals at various masses, as well as SM processes that contribute to the same final state as backgrounds, were generated, simulated and reconstructed, as descibed in Sec. 6.1. Selection criteria as well as quality criteria of the relevant physics objects and events are described in Sec. 6.3. In Sec. 6.7 are shown the main sources of systematic uncertainties. In Sec. 6.5 the signal regions, defined by optimizing the statistical significance, are described. In Sec. 6.6 data and MC expectations are compared and validated using standard candles. In Sec. 6.10 the final results are presented.

[1] hadronization is the process of the formation of hadrons out of quarks and gluons; the tight cone of particles created during the hadronization of a single quark is called a jet

6.1 Monte Carlo and data samples

This analysis is based on all of the data that the ATLAS detector has recorded with $\sqrt{s} = 8$ TeV during 2012. The run ranges with their respective integrated luminosities are displayed in Table 6.1. The official data quality selection is applied to the data.

Data period	Run range	Integrated luminosity [pb^{-1}]
A	200804 – 201556	794.02
B	202660 – 205113	5094.68
C	206248 – 207397	1406.02
D	207447 – 209025	3288.39
E	209074 – 210308	2526.28
G	211522 – 212272	1274.81
H	212619 – 213359	1444.93
I	213431 – 213819	1016.26
J	213900 – 215091	2596.34
L	215414 – 215643	839.76
Total		**20281.49**

Table 6.1: The data samples used in this analysis in their respective data periods, run ranges when recorded and their integrated luminosities.

The main backgrounds to LQ signatures are W+jets, Z/γ*+jets[2] events, $t\bar{t}$ events, where both top quarks decay leptonically, single top and diboson events. In addition there are also events coming from QCD processes and they add to the total background. This latter source of background will be evaluated through data-driven techniques. Background and signal processes in this analysis are estimated by using Monte Carlo simulated samples produced with the ATLAS simulation infrastructure within the ATLAS mc12a production campaign.

The LQ signals events at each LQ mass point (300-1200 GeV, at 50 GeV steps) were generated using Pythia 8.165 [34] together with the ATLAS Underlying Event Tune AU2 [35] and the CTEQ6L1 [36] PDF[3] set. The W+jets samples were produced with ALPGEN 2.14 interfaced with JIMMY 4.31, also with the AUET2 MC tune applied (Fig. 6.1). The MC samples used to represent the Z+jets, as well as the $t\bar{t}$ background, were produced with PowHEG [37, 38] interfaced with Pythia 8.165. A second set of $t\bar{t}$ samples was produced with SHERPA 1.4.1 and was used to crosscheck the Monte Carlo modeling. MC Samples representing the WW, WZ and ZZ diboson decays were generated with HERWIG 6.52 [39] and using the AUET2 [40] MC tune. On the other hand, the Drell-Yan backgrounds were all generated with SHERPA 1.4.1 [41]. In particular, the samples used to describe the Drell-Yan processes were generated with the massive c, b treatment instead of the conventional massless treatment. Samples of single top-quark events in the s-channel, t-channel and Wt were generated with MC@NLO 4.01 [42, 43] using the AUET2 MC tune.

[2] or vector+jets (*V+jets*) since W and Z are vector bosons

[3] PDF (Parton Distribution Function): the function defining momentum distribution among partons consisting a proton

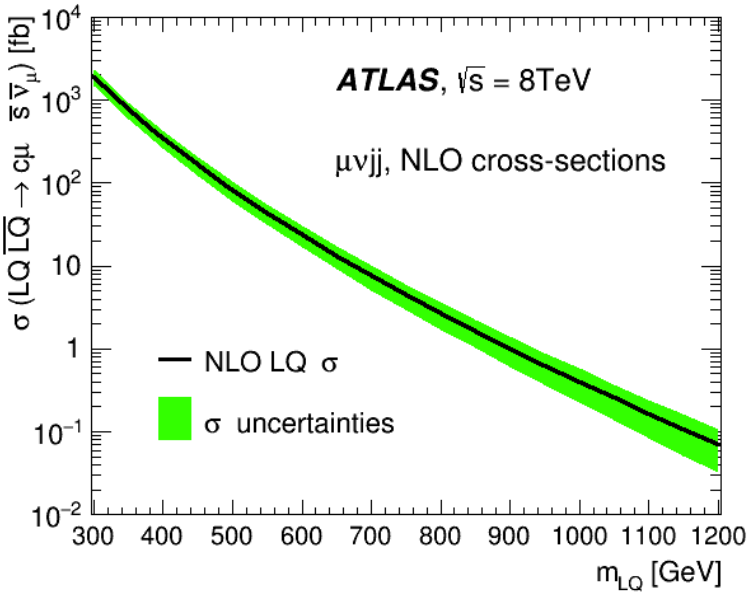


Figure 6.1: LQ NLO production cross-sections and their uncertainties.

The hadronization and parton showering of the samples produced with MC@NLO are produced using HERWIG 6.52 coupled to JIMMY 4.31. The choice of Parton Distribution Functions (PDF) used to produce the MC simulated samples is generator dependent: AcerMC, Pythia, HERWIG and ALPGEN use CTEQ6L1, while MC@NLO uses CT10 [44]. The QCD background has been estimated using data-driven techniques which are presented in detail in Sec. 6.4.2.

The predicted cross-sections at $\sqrt{s} = 8$ TeV of the LQ signal and background processes used in the analysis, as long as the request IDs and scale factors are shown in Tables 6.2 and 6.3 respectively. Each event of a produced MC sample can "weight" more or less than 1 and, thus, the total event weight of a sample is not equal to the absolute number of the events containing.

Process	Generator	request ID	σ [pb]	k-factor$\times$Filtering efficiency
W$\mu\nu$+0 jets	Alpgen/Jimmy	107690	8016.81	1.19$\times$1.00
W$\mu\nu$+1 jets	Alpgen/Jimmy	107691	1571.43	1.19$\times$1.00
W$\mu\nu$+2 jets	Alpgen/Jimmy	107692	475.90	1.19$\times$1.00
W$\mu\nu$+3 jets	Alpgen/Jimmy	107693	133.83	1.19$\times$1.00
W$\mu\nu$+4 jets	Alpgen/Jimmy	107694	35.51	1.19$\times$1.00
W$\mu\nu$+5 jets	Alpgen/Jimmy	107695	10.54	1.19$\times$1.00
W$\tau\nu$+0 jets	Alpgen/Jimmy	117700	8035.80	1.19$\times$1.00
W$\tau\nu$+1 jets	Alpgen/Jimmy	117701	1579.80	1.19$\times$1.00
W$\tau\nu$+2 jets	Alpgen/Jimmy	117702	477.55	1.19$\times$1.00
W$\tau\nu$+3 jets	Alpgen/Jimmy	117703	133.79	1.19$\times$1.00
W$\tau\nu$+4 jets	Alpgen/Jimmy	117704	35.58	1.19$\times$1.00
W$\tau\nu$+5 jets	Alpgen/Jimmy	117705	10.54	1.19$\times$1.00
Z$\mu\mu$	PowHEG/Pythia	147771	1240.00	1.00$\times$1.00
Z$\tau\tau$	PowHEG/Pythia	147772	1241.20	1.00$\times$1.00
WW	Herwig	105985	67.92	1.00$\times$1.00
ZZ	Herwig	105986	7.27	1.00$\times$1.00
WZ	Herwig	105987	12.01	1.00$\times$1.00
$t\bar{t}$	PowHEG/Pythia	117050	137.30	1.00$\times$1.00
single top [s-chan/W$\mu\nu$]	MC@NLO/Jimmy	108345	0.61	1.00$\times$1.00
single top [s-chan/W$\tau\nu$]	MC@NLO/Jimmy	108346	0.56	1.00$\times$1.00
single top [Wt]	MC@NLO/Jimmy	108347	22.37	1.00$\times$1.00
single top [t-chan/μ]	MC@NLO/Jimmy	17361	9.46	1.00$\times$1.00
single top [t-chan/τ]	MC@NLO/Jimmy	117362	9.48	1.00$\times$1.00
DY0120	Sherpa	180767	9.9821	1.00$\times$1.00
DY0180	Sherpa	180768	1.602	1.00$\times$1.00
DY0250	Sherpa	180769	0.567	1.00$\times$1.00
DY0400	Sherpa	180770	$9.49\cdot10^{-2}$	1.00$\times$1.00
DY0600	Sherpa	180771	$2.0\cdot10^{-2}$	1.00$\times$1.00
DY0800	Sherpa	180772	$4.28\cdot10^{-3}$	1.00$\times$1.00
DY1000	Sherpa	180773	$1.54\cdot10^{-3}$	1.00$\times$1.00
DY1250	Sherpa	180774	$4.5\cdot10^{-4}$	1.00$\times$1.00
DY1500	Sherpa	180775	$1.49\cdot10^{-4}$	1.00$\times$1.00
DY1750	Sherpa	180776	$5.31\cdot10^{-5}$	1.00$\times$1.00
DY2000	Sherpa	180777	$2.05\cdot10^{-5}$	1.00$\times$1.00
DY2250	Sherpa	180778	$8.18\cdot10^{-6}$	1.00$\times$1.00
DY2500	Sherpa	180779	$3.35\cdot10^{-6}$	1.00$\times$1.00
DY2750	Sherpa	180780	$1.4\cdot10^{-6}$	1.00$\times$1.00
DY3000	Sherpa	180781	$1.01\cdot10^{-6}$	1.00$\times$1.00

Table 6.2: Predicted cross-sections at $\sqrt{s}$ = 8 TeV, IDs, k-factors, filtering and efficiencies for the background processes used in the analysis. The cross-section for W^+W^- process is scaled by a factor 1.2 in order to take into account the ATLAS [45] and CMS [46] measurements.

6.2 Simulation and validation of the signal Monte Carlo samples

6.2.1 Pythia versions comparison

In the present analysis, the theoretical model of leptoquarks has been validated with Pythia generator. Various Pythia versions have been tested in order to find which is the one which describes better the characteristic kinematic properties of LQs and their final state products. In 2011 the Pythia LQ model was affected by a bug that was creating wrongly high levels of hadronic activity. The Pythia D6 tune was put in place by the Pythia authors in order to cure this problem.

Fig. 6.2 shows the transverse momentum (p_T) of the truth leading and subleading jet, as well as truth muon 's p_T and η produced with Pythia 6 (using CTEQ6L1 and D6 PDF set) and Pythia 8 (CTEQ6L1 PDF set) superimposed for comparison reasons. The signal samples have been produced for a nominal mass of 600 GeV for both LQs. In truth level, the jets are essentially the correspondent quarks in which both LQs decay to.

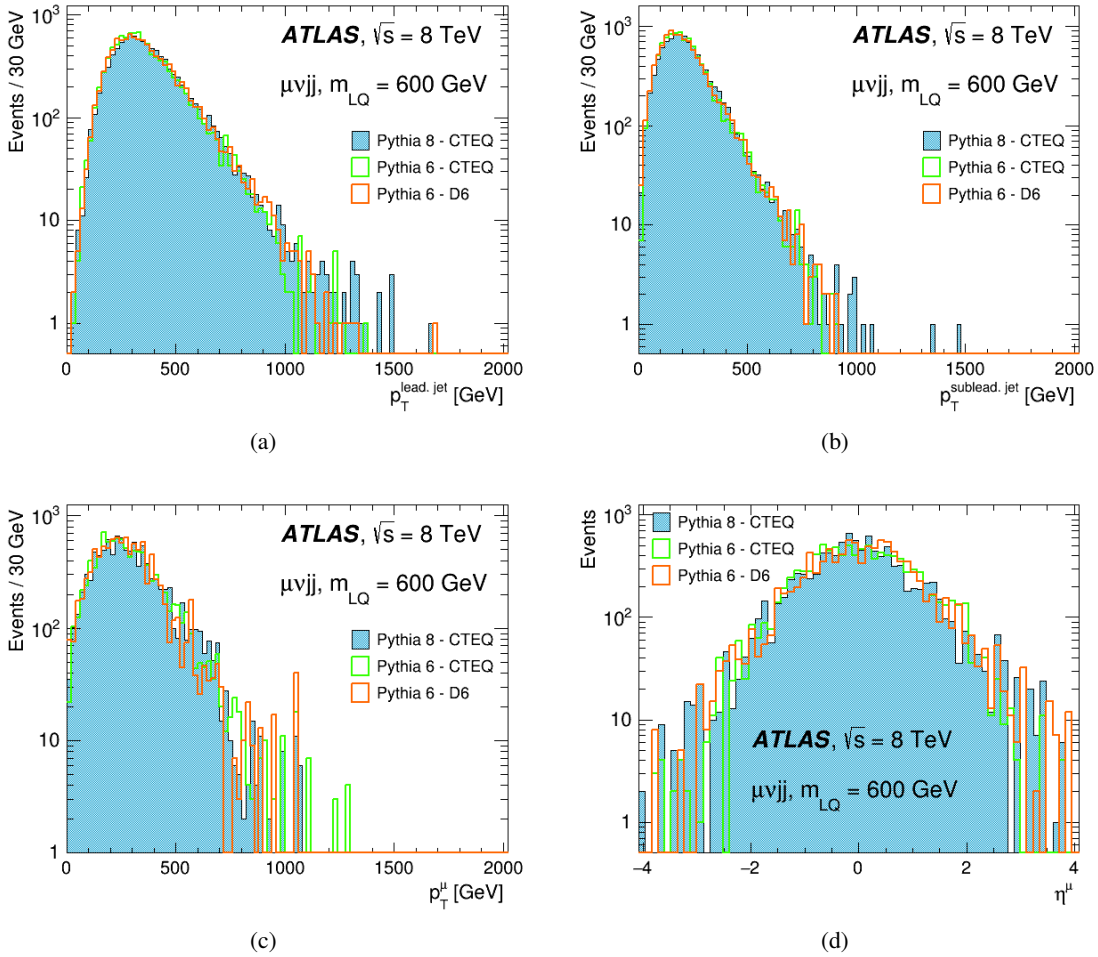


Figure 6.2: Transverse momentum (p_T) of (a) leading jet, (b) subleading jet and (c) muon in truth level in LQ events. Plot (d) shows the pseudorapidity (η) distribution of muon.

In Fig. 6.3 we see the same kinematics for the same objects, as well as neutrino, but in parton level. The parton level jets are been constructed by identifying the two highest-p_T truth quarks and making sure that these jets are the closest to the previous quarks (they are inside the solid angle with the quark in its center and the smallest radius). The neutrino corresponds to the E_T^{miss}. In the parton jets the high hadronic activity in low energy bins with Pythia 6 is obvious. A remedy to that problem of Pythia 6 is to exclude the final state radiation (FSR). FSR is gluon emission by a quark deriving directly from an LQ boson. This phenomenon can be seen in analogy to bremsstrahlung (where gluon corresponds to photon). FSR does not affect the flavor of the quark, but only its energy. By excluding FSR we make sure that the quark coming from LQ is indeed the final particle and not the quark minus a gluon. In Fig. 6.4 it is making obvious that jets without final state radiation fixes the problem for all Pythia versions.

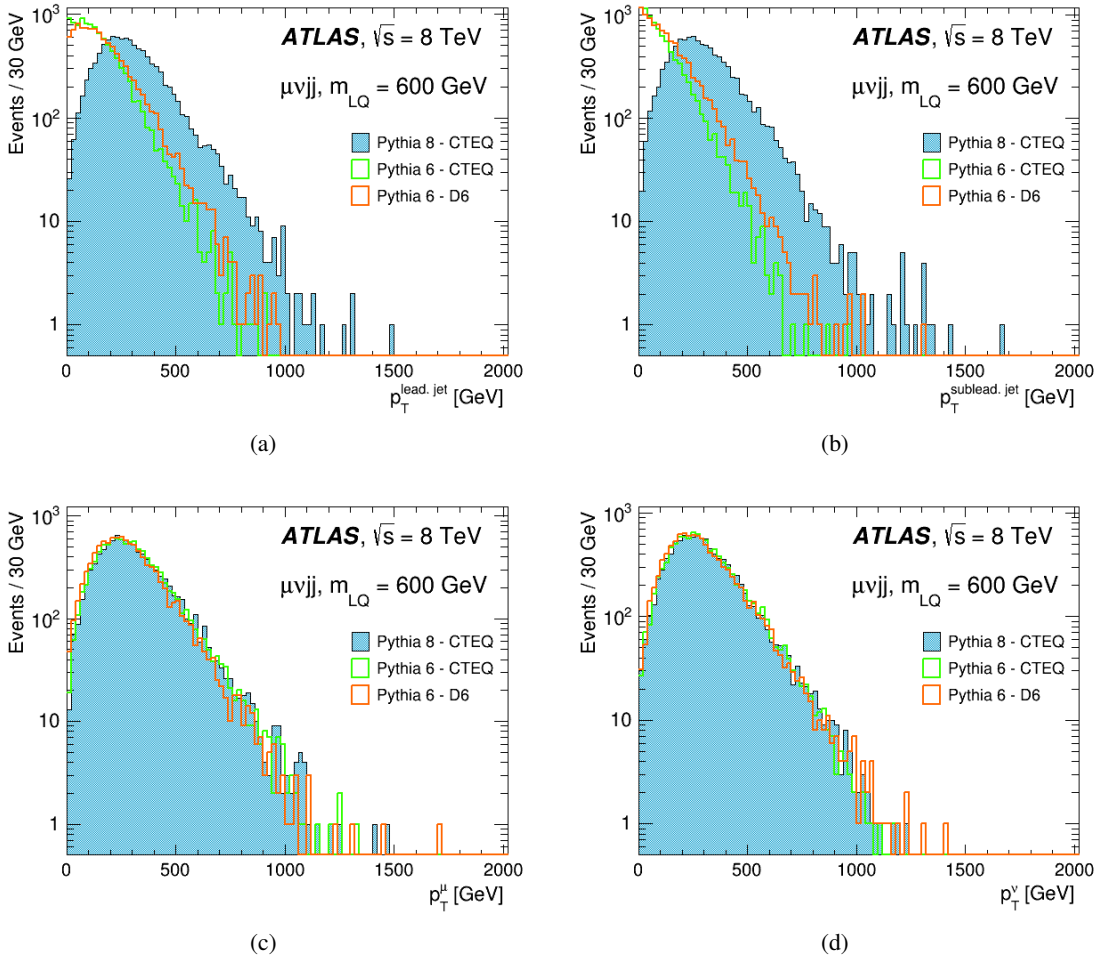
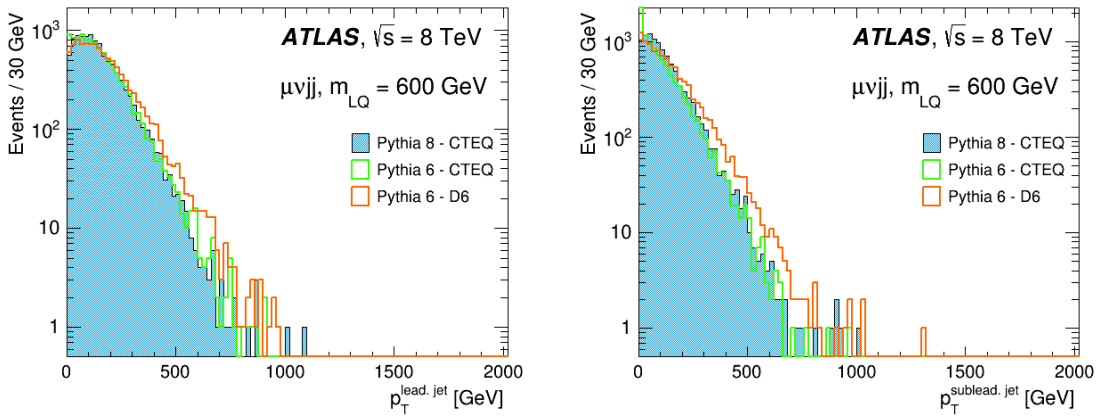


Figure 6.3: Transverse momentum (p_T) of (a) leading jet, (b) subleading jet, (c) muon and (d) neutrino in parton level in LQ events.



(a) (b)

Figure 6.4: Transverse momentum (p_T) of (a) leading jet and (b) subleading jet in parton level without FSR.

6.2.2 Fast and full simulation comparison

The official final signal used in the analysis can be produced centrally by ATLAS using fast or full simulation procedure. Fast simulation is, obviously, faster than full, so the question is about the deegree of disagreement between these two methods. In order to check the outcome, few samples with a small number - 5000 - of events have been requested for production with Pythia 8.165 generator in both modes. Fig. 6.5 shows the two leading jets and muon p_T, as well as E_T^{miss} corresponding to neutrino for a nominal LQ mass of 1 TeV.

In Fig. 6.6 appear the two parton-level reconstructed LQ masses in fast and full simulation for 0.3, 0.6 and 1 TeV respectively. The mass reconstruction was performed with the use of mass window combination. Further information about this method can be found below. Fig. 6.7 and 6.8 are 2-dimensional plots for 1 TeV LQs depicting the mass in contour and lego mode respectively, as well as η and ϕ distributions of the first over the second hypothetical boson. In general, there are rather small discrepancies between the two modes. Therefore, the final decision for the final LQ signal samples was to be produced in fast simulation. More plots about different Pythia versions and fast/full simulation are in App. A.1 and App. A.2 respectively.

6.2.3 Signal validation

The LQ signal used in the analysis has been validated through the production of some essential variables and kinematical properties of LQ bosons for 10,000 events with Pythia 8.165 at fast simulation for four LQ masses. Fig. 6.9 - Fig. 6.12 follow where both leading quarks, muon, neutrino and reconstructed LQs variables at truth level are been shown.

All the p_T distributions are normal as the nominal mass increases. The ϕ plots are isotropic since no specific polar angle should be prefered. The η distributions are mirror-symmetrical to a vertical axis identified with $\eta = 0$. Their maximum is

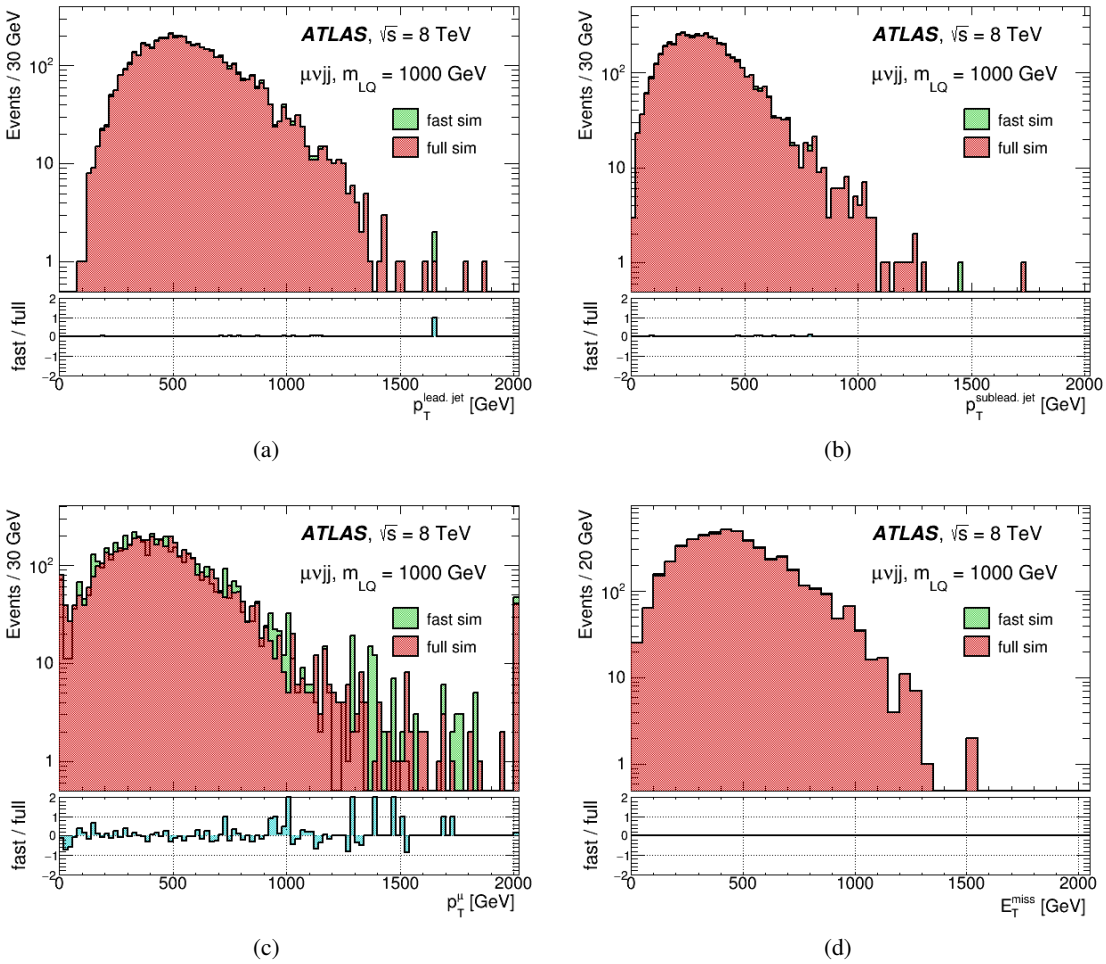


Figure 6.5: Transverse momentum (p_T) of (a) leading jet, (b) subleading jet and (c) muon and (d) E_T^{miss} produced in fast and full simulation mode.

obtained for small pseudorapidity values and the bins occupancy gets smaller for higher η values. This is something we expect since these are signal events in which the majority of hypothetical LQs should be produced almost vertical towards the ATLAS beam axis and in a back-to-back direction to each other. The truth LQ masses are confined in a small number of bins around the nominal value of each sample with minimum dispersion. This is expected since a truth reconstructed mass is the straight combination of a truth quark and a truth lepton at generator level. One final note is that Pythia allows only the combination of a c-quark to a μ (coming from an LQ decay) and an $\bar{s}$-quark to an $\overline{\nu}_\mu$ (products of an $\overline{LQ}$ decay) or a $\bar{c}$-quark to an $\overline{\mu}$ and an s-quark to a ν_μ. Of course, this is something that should not affect the result over kinematical variables given that it depends only on the λ-coupling of pair LQ production and not on the possible combinations of the produced stable particles of the final state. Additionally, it is noted that the LQ decay cross-section is the same for particles and antiparticles. In any case, and when it comes to the real data analysis, all the combinations are done freely according to other physical criteria applied, which will be mentioned later.

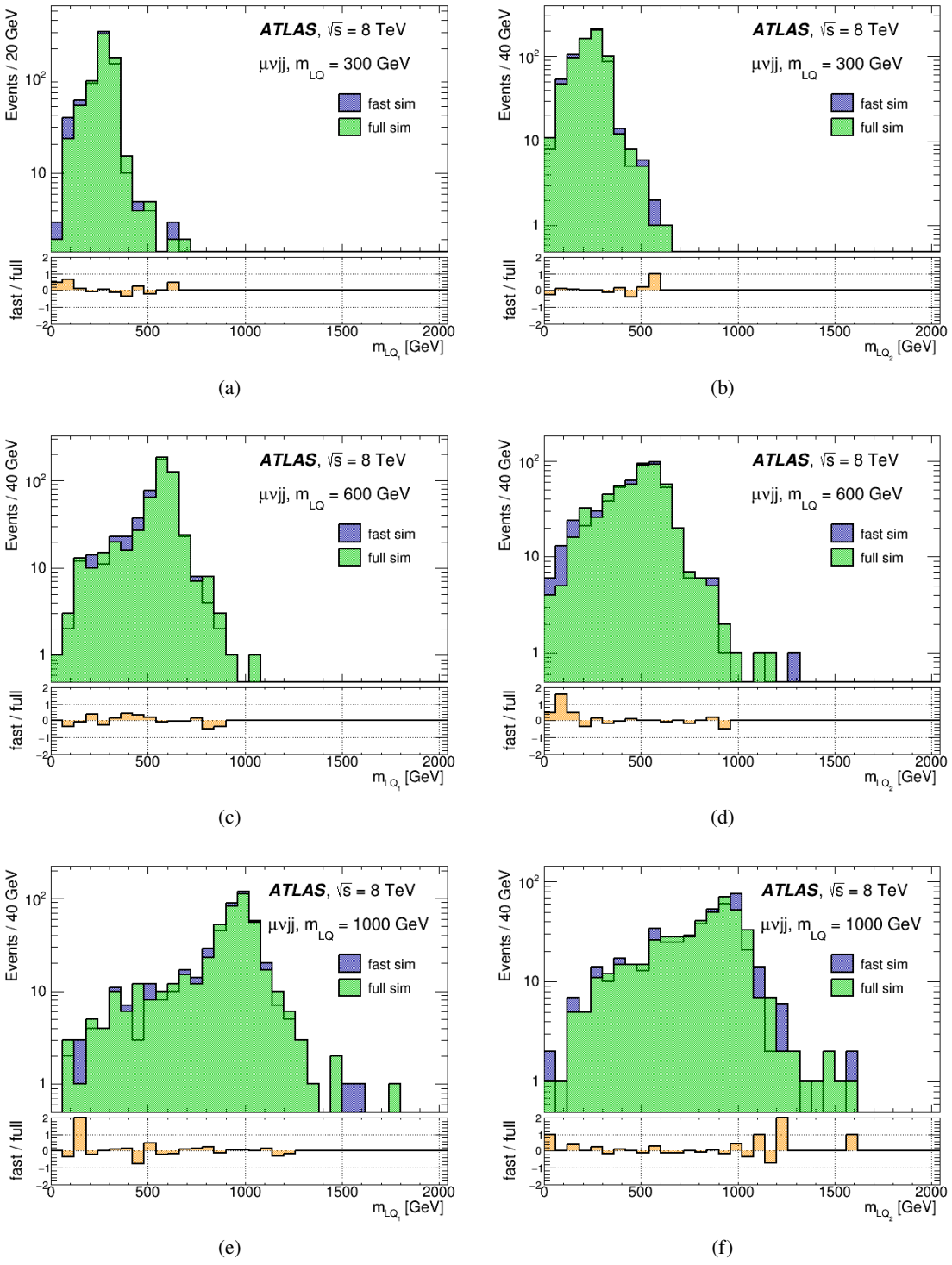


Figure 6.6: Reconstructed LQ masses for (a) - (b) 300 GeV, (c) - (d) 600 GeV and (e) - (f) 1000 GeV respectively.

LQ mass [GeV]	Request ID	σ [pb]	σ uncert. [%]	k-factor$\times$Filtering efficiency
300	188633	1.890	21.91	1.00×1.00
350	188634	0.770	22.89	1.00×1.00
400	188635	0.342	24.04	1.00×1.00
450	188636	0.163	25.78	1.00×1.00
500	188637	$8.20 \cdot 10^{-2}$	26.63	1.00×1.00
550	188638	$4.31 \cdot 10^{-2}$	28.05	1.00×1.00
600	188639	$2.35 \cdot 10^{-2}$	29.50	1.00×1.00
650	188640	$1.32 \cdot 10^{-2}$	31.12	1.00×1.00
700	188641	$7.61 \cdot 10^{-3}$	32.71	1.00×1.00
750	188642	$4.48 \cdot 10^{-3}$	34.38	1.00×1.00
800	188643	$2.69 \cdot 10^{-3}$	36.17	1.00×1.00
850	188644	$1.64 \cdot 10^{-3}$	38.01	1.00×1.00
900	188645	$1.01 \cdot 10^{-3}$	40.06	1.00×1.00
950	188646	$6.34 \cdot 10^{-4}$	42.05	1.00×1.00
1000	188647	$4.01 \cdot 10^{-4}$	44.17	1.00×1.00
1050	188648	$2.56 \cdot 10^{-4}$	46.34	1.00×1.00
1100	188649	$1.65 \cdot 10^{-4}$	48.71	1.00×1.00
1150	188650	$1.07 \cdot 10^{-4}$	51.23	1.00×1.00
1200	188651	$6.96 \cdot 10^{-5}$	53.90	1.00×1.00

Table 6.3: Predicted LQ signal cross-sections at $\sqrt{s}$ = 8 TeV and other numbers for all the masses used in the analysis acquired with Pythia 8.165 generator package. The cross-sections and their uncertainties have been calculated up to NNLO with Krämer tool.

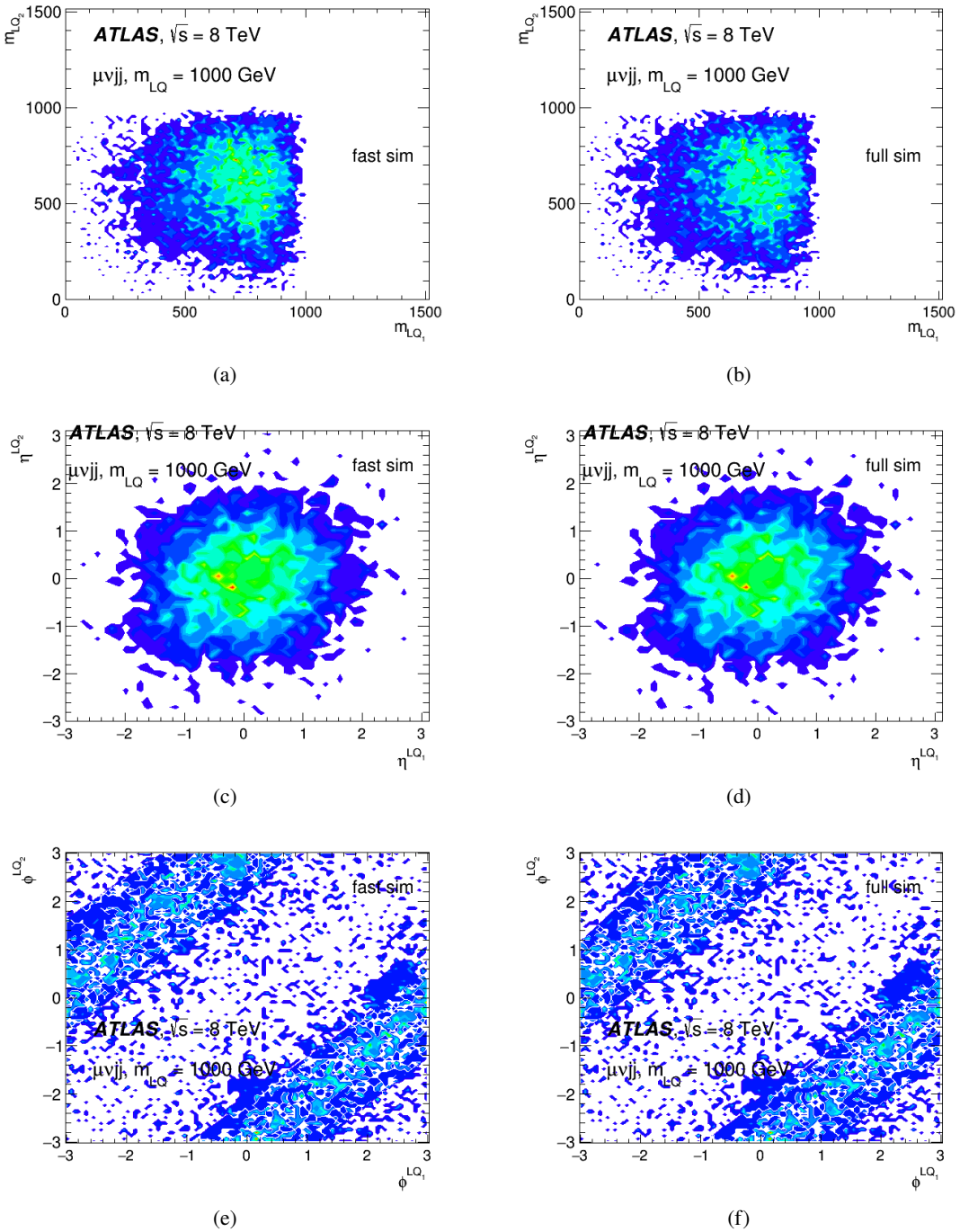
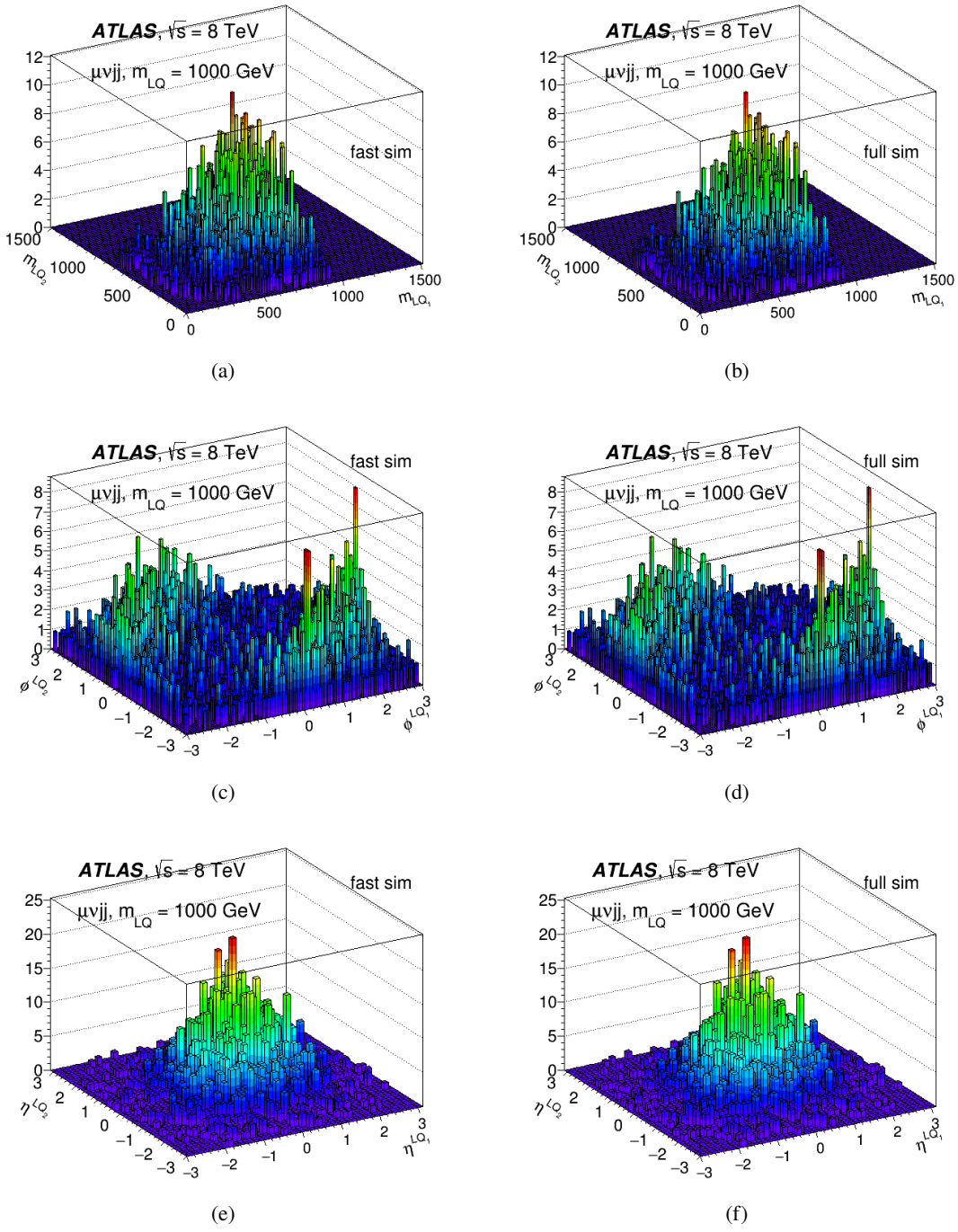


Figure 6.7: 2-D LQ_1 vs LQ_2 plots of (a) - (b) mass, (c) - (d) ϕ and (e) - (f) η produced in fast and full simulation mode.

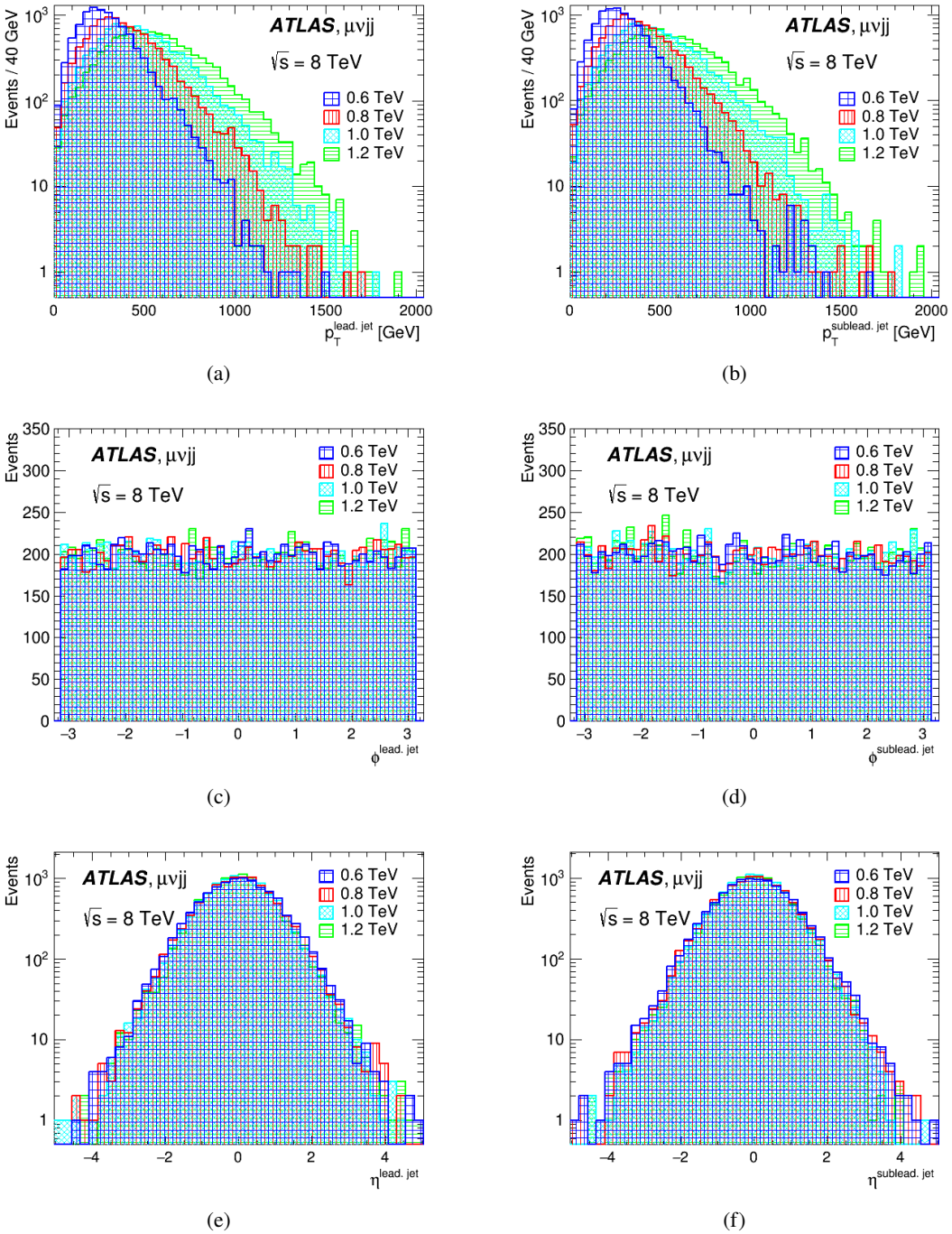


(a) (b)

(c) (d)

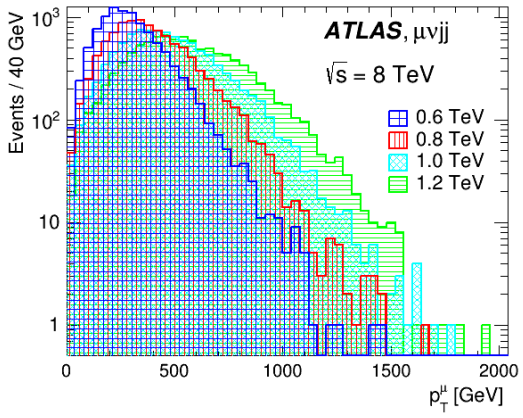
(e) (f)

Figure 6.8: 2-D LQ_1 vs LQ_2 plots of (a) - (b) mass, (c) - (d) ϕ and (e) - (f) η produced in fast and full simulation mode.

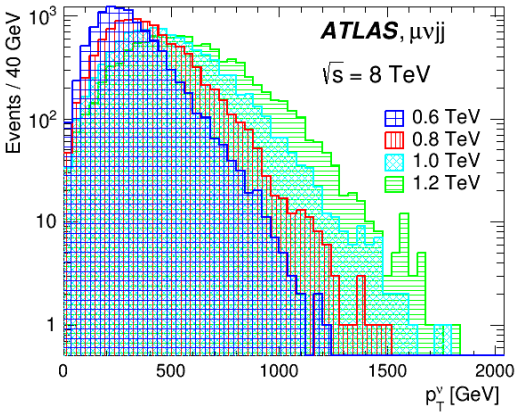


(a) (b) (c) (d) (e) (f)

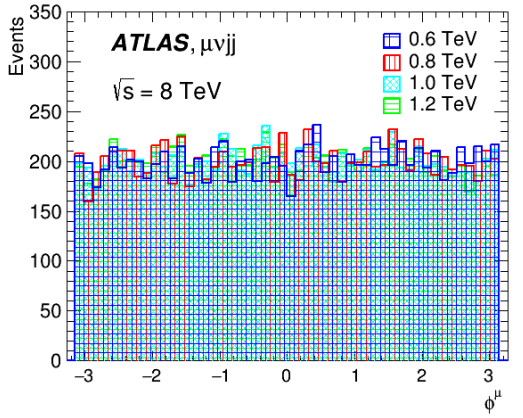
Figure 6.9: Both leading quarks (a) - (b) p_T, (c) - (d) ϕ and (e) - (f) η.



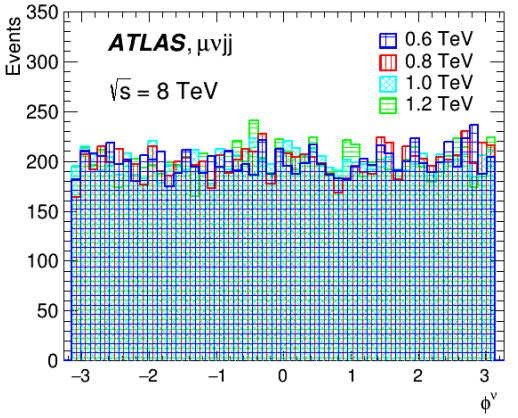
(a)



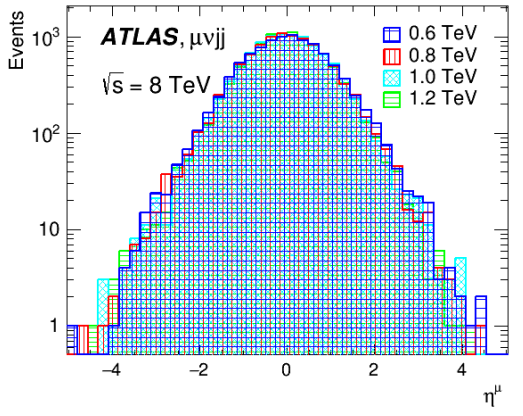
(b)



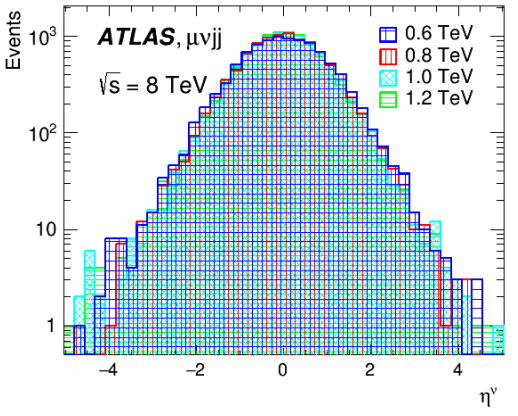
(c)



(d)

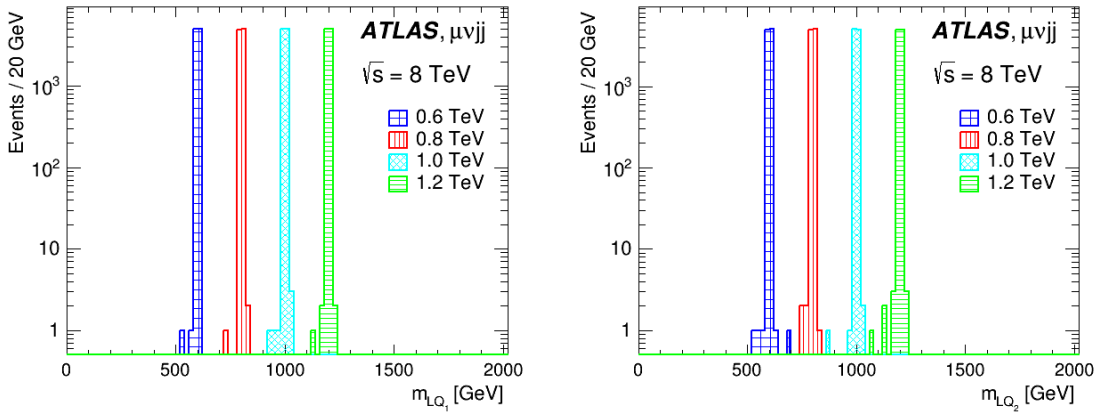


(e)



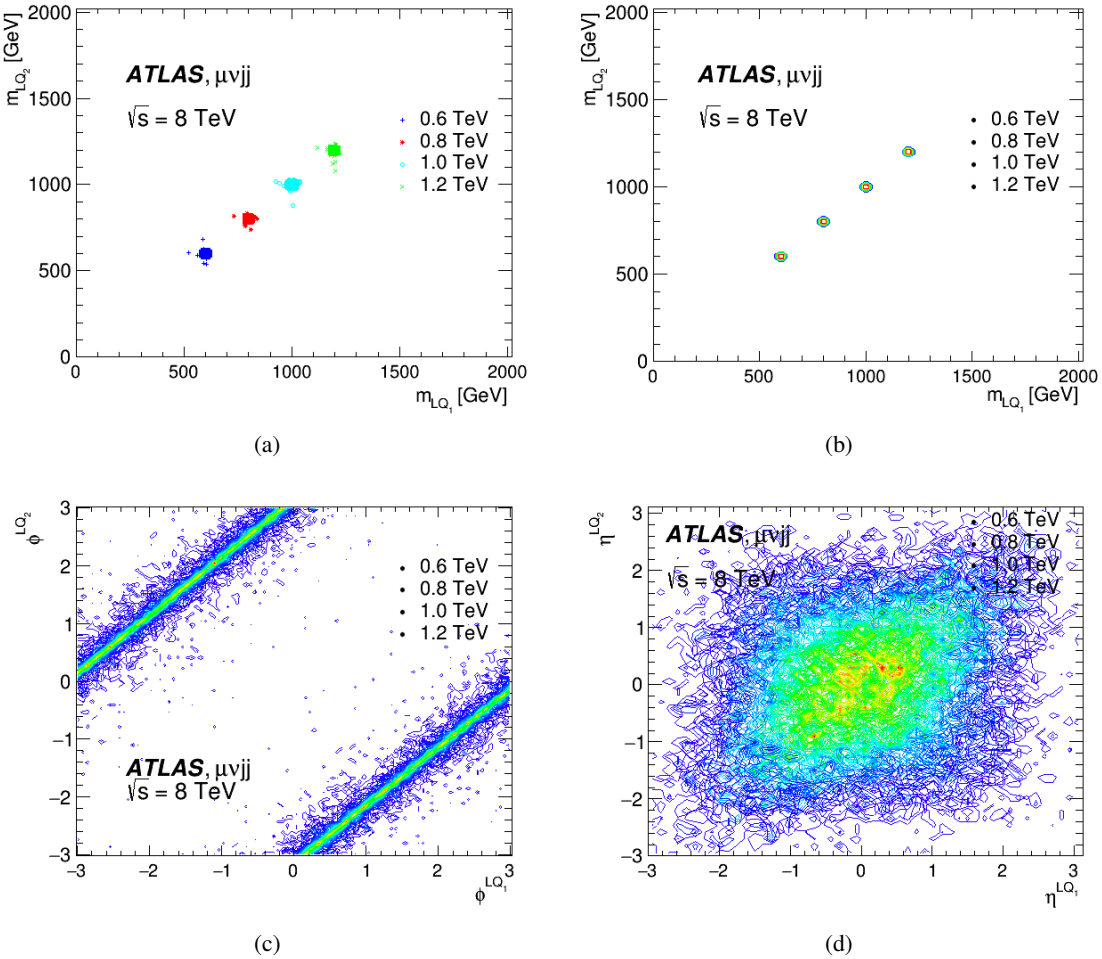
(f)

Figure 6.10: muon and neutrino (a) - (b) p_T, (c) - (d) ϕ and (e) - (f) η.



(a) (b)

Figure 6.11: Reconstructed truth LQ masses.



(a) (b)

(c) (d)

Figure 6.12: 2-D truth reconstructed LQ_1 vs LQ_2 plots of (a) - (b) mass, (c) ϕ and (d) η.

6.3 Event and object selection

6.3.1 Trigger and data quality

A data event, in order to be recorded as good, is required to have all relevant components of the ATLAS detector in good working condition. For the present analysis an online trigger requires at least one muon, which is fully efficient for muons with transverse momentum above 40 GeV [47]. The trigger used is EF_mu18_MG_medium[4]. Every time the proton bunches cross each other multiple collisions can occur, which create multiple primary vertices. The primary vertex of the event, from which the leptons are required to originate, is defined as the reconstructed vertex with the largest sum of squared transverse momenta of its associated tracks. Given that many particles produced by a primary vertex of a proton - proton collision are not stable and can decay in an infinitessimal amount of time, many secondary vertices can be created as well in close proximity to the primary one. For this reason, we must be in a position to tell the difference among them and use only the primary vertex products in the analysis. Events are selected if they contain a primary vertex with at least three associated tracks satisfying $p_{T,track} > 0.4$ GeV, at least one muon and at least two jets. Monte Carlo simulated events were corrected to better describe the data. More specifically, a weighting factor[5] was applied event by event in order to match the number of primary vertices in data. A weighting factor was also applied in order to improve the modeling of the vertex position. Scale factors were applied to account for differences in trigger, reconstruction, isolation and identification efficiencies. The energy and momentum of the selected physics objects were corrected to match the resolution and scales measured in data.

6.3.2 Muon selection

Muon tracks are reconstructed independently in both the Inner Detector and the Muon Spectrometer (Combined Muons - CB [48]). After that, a combined fit to the measurements from both sub-systems is applied. Minimum requirements are imposed on the number of hits in each of the ID sub-systems to ensure that tracks are well measured. Most of cosmic muons are being absorbed by the overlying earth. For those which manage to get through this and happen to coincide with the proton - proton collisions, the additional requirements on the muon longitudinal and transverse impact parameters $|d_0| < 0.2$ mm and $|z_0| < 1.0$ mm are applied to reject them. Individual muons must also pass the relative-isolation requirement of $p_T^{cone20}/p_T^{\mu} < 0.2$, where p_T^{cone20}/p_T^{μ} is the ratio of the sum of transverse momenta of all the tracks within a solid angle (cone) of $\Delta R < 0.2$ around the muon track over muon 's transverse momentum. In order to ensure optimal momentum resolution, muons are required to register at least three hits in each of the inner, middle and outer stations of the MS and at least one hit in two of the layers of the muon trigger chambers. The hit requirement at the inner station in the forward region corresponds to at least two hits in the Cathode Strip Chanbers (CSC). The muons selected according to those requirements are expected to have p_T resolution at ~ 1 TeV ranging from 10% to 25%. They are referred to as three-station muons. For the

[4] it requires a track recorded from ID matching to a track from inner muon triggering chambers; it depends on 3-station muons; its efficiency based on stand-alone muons from data is $\sim$ 70% in barrel and $\sim$ 90% in forward calorimeter of a threshold of $p_T > 20$ GeV

[5] the so-called *"event weight"* containing MC weight, pile up weight, trigger weight and vertex weight

central region where $|\eta| < 1.05$, muons with at least five precision hits in each of the inner and outer stations are also selected in addition to the three-station muons and are referred as two-station muons. These two-station muons also require at least one hit in one layer of the trigger chambers and have slightly worse resolution compared to the three-station muons. The standalone momentum measurements from the ID and MS of each muon must not differ by more than five (three) times the sum in quadrature of the standalone uncertainties for the three-station (two-station) muons. The intrinsic position resolution and effects of known misaligned chambers are included in the simulation and muons passed through them are rejected. Muon selection cut-flow is in Table 6.4. A detailed description of the muon performance in ATLAS can be found in [49].

Muon selection	Event number	Jet selection	Event number
all muons	10,120	all jets	32,253
only central muons	9,892	E > 0	32,253
$p_T > 30$ GeV	9,697	good jet	32,253
$\eta < 2.5$	9,697	$\eta < 2.8$	32,253
B-layer	9,697	$\|\text{JVF}\| > 0.5$	31,712
Pixel	9,697	$p_T > 30$ GeV	31,712
SCT	9,697	$\Delta R_{muon,lead.jet} < 0.4$	31,712
Si	9,697	$\Delta R_{elec,lead.jet} < 0.4$	31,712
TRT	9,697		
$\|z_0\| < 1$mm	9,695		
$\|d_0\| < 0.2$mm	9,592		
$p_T^{cone20}/p_T^{\mu} < 0.2$	9,432		
3-station muons	8,091		
trigger matched	7,972		

Table 6.4: The results are from Pythia 8 LQ signal sample of m_{LQ} = 300 GeV. **muon selection**: B-layer, pixel, SCT, Si and TRT are MCP quality cuts referring to sub-detectors, sensors etc. The 3-station muon refers to at least three (inner, middle and outer) MS hits. The trigger matched cut validates that the selected muon is the one that fired the muon trigger. **jet selection**: the energy > 0 confirms it is not a noise channel.

6.3.3 Jet selection

Jets are reconstructed from energy clusters detected in the calorimeter using the AntikT4 algorithm [50]. Jets used in the analysis must satisfy $p_T > 30$ GeV and $|\eta| < 2.8$. Jets reconstructed in a cone $\Delta R < 0.4$ around one selected lepton (muon) are removed at the analysis level as not to mistake it for a lepton. Jets selection cut-flow is also in Table 6.4 Additional jet quality criteria are also applied to remove fake jet signals caused by detector effects [51].

6.3.4 Event selection

In order an event to be selected must fulfill some specific criteria. The following list contains these fulfilled criterias:

- event must be in Good Runs List (GRL)

- event must have removed jets that are double counted in events simulated with the generator AlpGen
- the first primary vertex (PV) of the event must contain at least 3 associated tracks
- events with "bad" jets must be removed (E_T^{miss} cleaning)
- events with liquid argon (LAr) and tile calorimeters errors are removed
- incomplete or tile corrupted events are removed
- events with jets falling in instrumented areas with problems are removed
- trigger menus are applied
- event must contain at least one loose muon
- event must contain at least one LQ signal (stand alone or calorimeter) muon
- muon trigger must be EF_mu18_MG_medium
- passed muon triggered must be matched to the previous one
- event must contain only one muon and no electron
- event must contain at least two signal jets
- E_T^{miss} must be larger than 40 GeV
- the angle in the transverse plane between leading p_T jet and E_T^{miss} must be greater than 0.5
- the angle in the transverse plane between leading p_T muon and E_T^{miss} must be greater than 0.8
- lepton-E_T^{miss} mass (m_T) (definition in Sec.6.5.1) must be greater than 50 GeV

Event selection	Event number
events after muon trigger	16,777,216
bad event and tile error	16,777,216
$N_{muons} > 0$	4,915,295
exactly 1 muon	4,747,909
pass EF_mu18_MG_medium trigger	4,747,434
trigger matched	4,563,483
$N_{muons} = 1$ and $N_{electrons} = 0$	4,563,443
$N_{jets} \geq 2$	4,563,443
$E_T^{miss} > 40$ GeV	3,016,673
$\Delta\phi_{lead.jet,E_T^{miss}} > 0.5$	2,725,828
$\Delta\phi_{lead.muon,E_T^{miss}} > 0.8$	2,110,178
$m_T > 50$ GeV	1,865,890

Table 6.5: The results are from real data.

At this point the event pre-selection ends and follow the definition of control and signal regions.

6.3.5 Pile-up reweighting

The MC simulation does not describe the pile-up conditions exactly. The reason is that it is often produced before a data taking is finished. Since pile-up can affect variables such as electron or muon isolation, jet p_T or jet multiplicity, it is important to bring the pile-up conditions in MC to agreement with data. In order to do it, MC events are weighted. The weights are chosen in such a way that the distribution of a pile-up related variable in MC matches data. This procedure of MC weighting is usually called pile-up reweighting (PRW).

6.4 Determination of background yields

Yields from all background processes except QCD (multijet) background are determined using the MC simulated samples described previously. The final yields are determined after the scale factors described in Sec. 6.3.5 are applied.

6.4.1 Simulated backgrounds

The simulated backgrounds used in this analysis are listed in Tab. 6.2 and Tab. 6.3. Their yields are determined with all of the analysis requirements listed in Sec. 6.3 applied to the simulated events, then multiplied by the luminosity and predicted cross-sections. The control regions defined for the W+jets and $t\bar{t}$ processes that are described in Sec. 6.6 are used to validate the modelling of these two most dominant backgrounds. Normalisation scale factors are also obtained for these background samples in their respectively defined control regions.

6.4.2 QCD background

QCD (multijet) background cannot be acquired through simulation. Therefore, this is done with ABCD data driven method. Two mutually independent variables are been selected. In this case, the variables defining a high signal purity region are E_T^{miss} and p_T^{cone20}/p_T^{μ} muon isolation. Of course, it is rather obvious that, in this section, signal refers to the events originating from QCD background and has nothing to do with leptoquarks. The phace space of the orthogonal variables consists of three control regions and one signal region (Fig. 6.13). The signal region (region A) recquires $E_T^{miss} > 50$ GeV and valid $p_T^{cone20}/p_T^{\mu} < 0.2$ isolation. On the other hand, control region B consists of $p_T^{cone20}/p_T^{\mu} > 0.2$ and $E_T^{miss} > 50$ GeV. Control region C is the opposite of signal region A ($p_T^{cone20}/p_T^{\mu} > 0.2$ and $E_T^{miss} < 50$ GeV). Finally, the last control region D consists of $p_T^{cone20}/p_T^{\mu} < 0.2$ and $E_T^{miss} < 50$ GeV. The events contained into these regions theoretically do not contaminate each other. The number of events of each region satisfy the following equation

$$N_A \cdot N_C = N_B \cdot N_D \tag{6.1}$$

and, therefore, the number of events of QCD background is given by

$$N_A = \frac{N_D}{N_C} \cdot N_B = QCD\_scaleFactor \cdot N_B \tag{6.2}$$

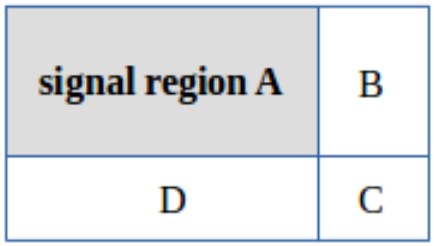


Figure 6.13: Schematic illustration of ABCD data-driven method phace space.

N_A is been calculated in respect to control region B since it has the richest statistics and it contains more signal events. The shape of the actual QCD distribution is acquired by applying

$$QCD\_scaleFactor = \frac{N_D}{N_C} \simeq 0.2411 \tag{6.3}$$

factor to the control region B QCD distribution. This last one is been constructed throught the subtraction of all the simulated background samples from real data for $p_T^{cone20}/p_T^{\mu} > 0.2$ and $E_T^{miss} > 50$ GeV. All the calculations appearing in Tab. 6.6 have been applied to the final pre-selection cut. QCD background is estimated to be equal to $\sim 17.4\%$ of the total (simulated and data-driven) background.

Region	Data	simulated background	QCD background
Signal Region A ($E_T^{miss} > 50$ GeV, $p_T^{cone20}/p_T^{\mu} < 0.2$)	$1,865,890 \pm 12,376$	$1,396,160 \pm 11,055$	$294,969 \pm 4,490$
Control Region B ($E_T^{miss} > 50$ GeV, $p_T^{cone20}/p_T^{\mu} > 0.2$)	$1,236,849 \pm 10,242$	$13,419 \pm 1,233$	$1,223,430 \pm 10,162$
Control Region C ($E_T^{miss} < 50$ GeV, $p_T^{cone20}/p_T^{\mu} > 0.2$)	$493,601 \pm 6,296$	$1,157 \pm 320$	$492,444 \pm 6,288$
Control Region D ($E_T^{miss} < 50$ GeV, $p_T^{cone20}/p_T^{\mu} < 0.2$)	$314,172 \pm 5,093$	$195,443 \pm 4,234$	$118,729 \pm 2,923$

Table 6.6: Number of events per region of real data, inclusive simulated backgrounds and data-driven QCD background calculated at the final pre-selection cut. The errors appeared are only statistical.

6.5 Searching for leptoquarks

6.5.1 Definitions of additional observables

In addition to the common kinematic variables, p_T, E_T^{miss}, η and ϕ, several additional variables are used at various stages in the analysis. The first set of these variables are event quantities which are large in magnitude for events with massive particles. These are:

- $L_T = p_T^{\ell} + E_T^{miss}$, the scalar sum of lepton (muon) transverse momentum and missing transverse energy,
- $H_T = p_T^{jet_1} + p_T^{jet_2}$, the scalar sum of the two leading jets transverse momenta,
- $S_T = L_T + H_T$ and
- N_{jets} (N_{muons}) = total number of jets (muons) per event.

A second category of additional variables is related to mass reconstruction. There are three main variables in this category:

- $m_T = \sqrt{2p_T^{\ell}E_T^{miss}(1 - cos\Delta\phi(\ell, E_T^{miss}))}$, the standard lepton-missing transverse energy mass combination. This is used to distinguish between events with and without leptonic decay of a real W boson. It must not be confused with any LQ mass combination,
- m_{LQ_1} and m_{LQ_2}

These two mass variables combine the leading two jets, the lepton and the E_T^{miss} of the event into the lepton-jet and E_T^{miss}-jet pairs which best reconstruct the two hypothesized LQs of the specific event. This is done by constructing the invariant mass and the transverse mass of all the possible lepton and jet combinations in every event. That is, the muon is associated to the leading (subleading) jet and the invariant LQ mass (m_{LQ}) calculated, while the E_T^{miss} is associated to the subleading (leading) jet to calculate the transverse LQ mass ($m_{LQ}^T = \sqrt{2p_T^j E_T^{miss}(1 - cos\Delta\phi(j, E_T^{miss}))}$[6]). According to theoretical prediction the two pair-produced LQs must have the same mass, thus we calculate the absolute difference between the invariant and the transverse mass in each of these two combinations. The combination giving the smallest difference is taken as the good lepton-jet assignment (mass window method). As a result, in each single event if m_{LQ_1} corresponds to m_{LQ}, then m_{LQ_2} corresponds to m_{LQ}^T and vice versa. This algorithm gives the correct pairing in $\sim$ 90% of the cases as measured on the signal MC samples.

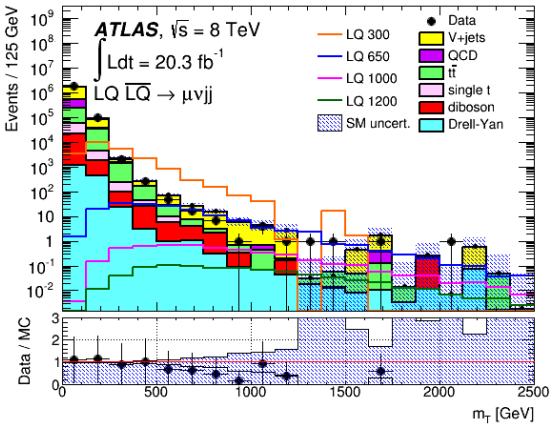
There are also some secondary mass combinations:

- $m_{LQ}^{avg} = \frac{m_{LQ_1} + m_{LQ_2}}{2}$, the average LQ mass
- m_{LQ}^{max} and m_{LQ}^{min}, maximum and minimum LQ masses; all possible mass combinations are been calculated for each event and the maximum and minimum of them are been assigned respectively,

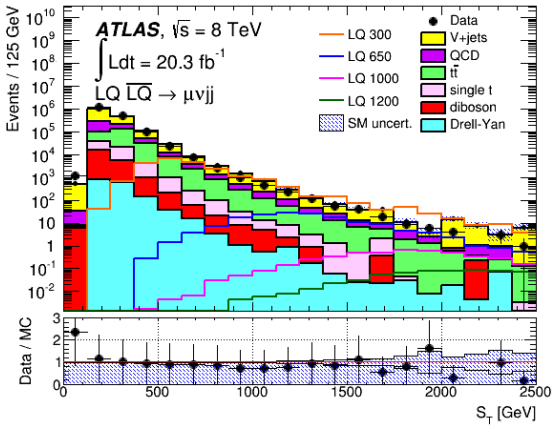
[6] m_{LQ}^T should not be confused with the non-LQ m_T mass

- $m_{LQ}^{\ell-jet_1}$, $m_{LQ}^{\ell-jet_2}$, $m_{LQ}^{\nu-jet_1}$, $m_{LQ}^{\nu-jet_2}$, arbitrary combinations using as criteria the specific LQ components independently from mass calculations.

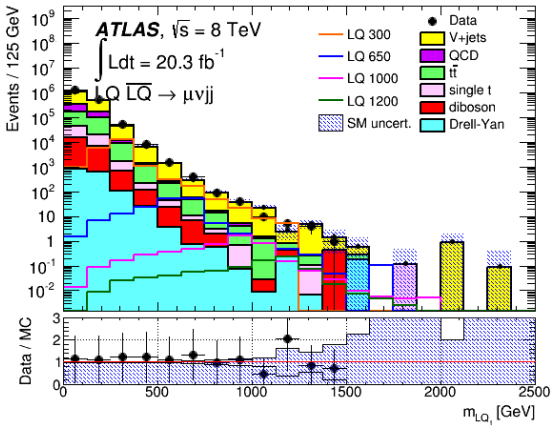
In the present analysis the variables used for their high degree of potential LQ signal separetion from background are m_T, S_T and m_{LQ_1} (Fig. 6.14). In these histograms appear four different LQ signal masses at 300, 650, 1000 and 1200 GeV respectively. There have been applied specific cuts to these three variables, different for each mass point in order to remove most of the background events and preserve most of the signal in each case. More variables distributions at the end of pre-selection can be found in App. A.3. Sec. 6.8 is dedicated to signal regions cuts optimization technics.



(a)



(b)



(c)

Figure 6.14: (a) Lepton-E_T^{miss} mass (m_T), (b) total scalar momentum (S_T) and (c) m_{LQ_1} after applying the pre-selection cuts with signal masses at 300, 650, 1000 and 1200 GeV and an integrated luminosity of 20.3 fb^{-1}.

6.5.2 Signal acceptances

The signal acceptance of the selection cuts is almost fixed to a mean value of $\sim$ 59% for all mass and, certainly, it does not get smaller than $\sim$ 57% (for $m_{LQ} \geq$ 1150 GeV) (Table 6.7).

LQ mass [GeV]	Number of expected events ***before*** pre-selection cut-flow	Number of expected events ***after*** pre-selection cut-flow	LQ signal acceptance [%]
300	37,367.5 $\pm$ 2,644.0	21,655.1 $\pm$ 2,012.5	57.95
350	14,852.4 $\pm$ 1,755.3	8,641.5 $\pm$ 1,339.2	58.18
400	6,850.4 $\pm$ 1,243.3	4,031.2 $\pm$ 954.6	58.85
450	3,216.5 $\pm$ 879.7	1,914.9 $\pm$ 680.0	59.53
500	1,611.3 $\pm$ 640.0	952.6 $\pm$ 491.7	59.12
550	833.7 $\pm$ 481.9	505.1 $\pm$ 374.2	60.59
600	466.3$\pm$ 358.0	279.5 $\pm$ 281.8	59.94
650	262.0 $\pm$ 279.6	156.8 $\pm$ 217.5	59.85
700	150.0 $\pm$ 211.9	89.5 $\pm$ 169.5	59.67
750	88.7 $\pm$ 168.7	53.0 $\pm$ 131.2	59.75
800	52.0 $\pm$ 128.3	30.9 $\pm$ 102.6	59.42
850	32.3 $\pm$ 100.7	19.1 $\pm$ 81.9	59.13
900	19.7 $\pm$ 80.8	11.9 $\pm$ 66.2	60.41
950	12.4 $\pm$ 64.7	7.4 $\pm$ 53.0	59.68
1000	7.9 $\pm$ 52.7	4.8 $\pm$ 43.2	60.76
1050	5.1 $\pm$ 42.5	3.0 $\pm$ 34.8	58.82
1100	3.2 $\pm$ 33.8	1.9 $\pm$ 27.7	59.38
1150	2.1 $\pm$ 27.8	1.2 $\pm$ 22.8	57.14
1200	1.4 $\pm$ 22.3	0.8 $\pm$ 18.6	57.14

Table 6.7: Number of expected events determined from MC simulated samples before and after pre-selection. The errors appeared are only statistical.

6.6 Control regions

6.6.1 Definition

The main idea behind control regions is to apply specific cuts in order to form regions where almost all possible signal has been excluded and they are populated only by background events with one dominating background in each case. Thus, it is possible to check the degree of agreement of already known SM background in respect to real data given that no signal excess is expected. Of course, in case this agreement is not adequate it is feasibile to apply some extra coefficients to improve fitting.

For present analysis, two main and one secondary control regions with negligible signal contamination are defined to validate modeling accuracy of MC simulated background events, as well as to derive correction scale factors. There are two W+jets control regions, one corresponding to W boson 's leptonic and one to its hadronic decay. W bosons can decay to a lepton and a neutrino (μ^- and ν_μ in our case) or to two quarks (which are been identified as jets through hadronization procedure). The two W+jets control regions are as follow

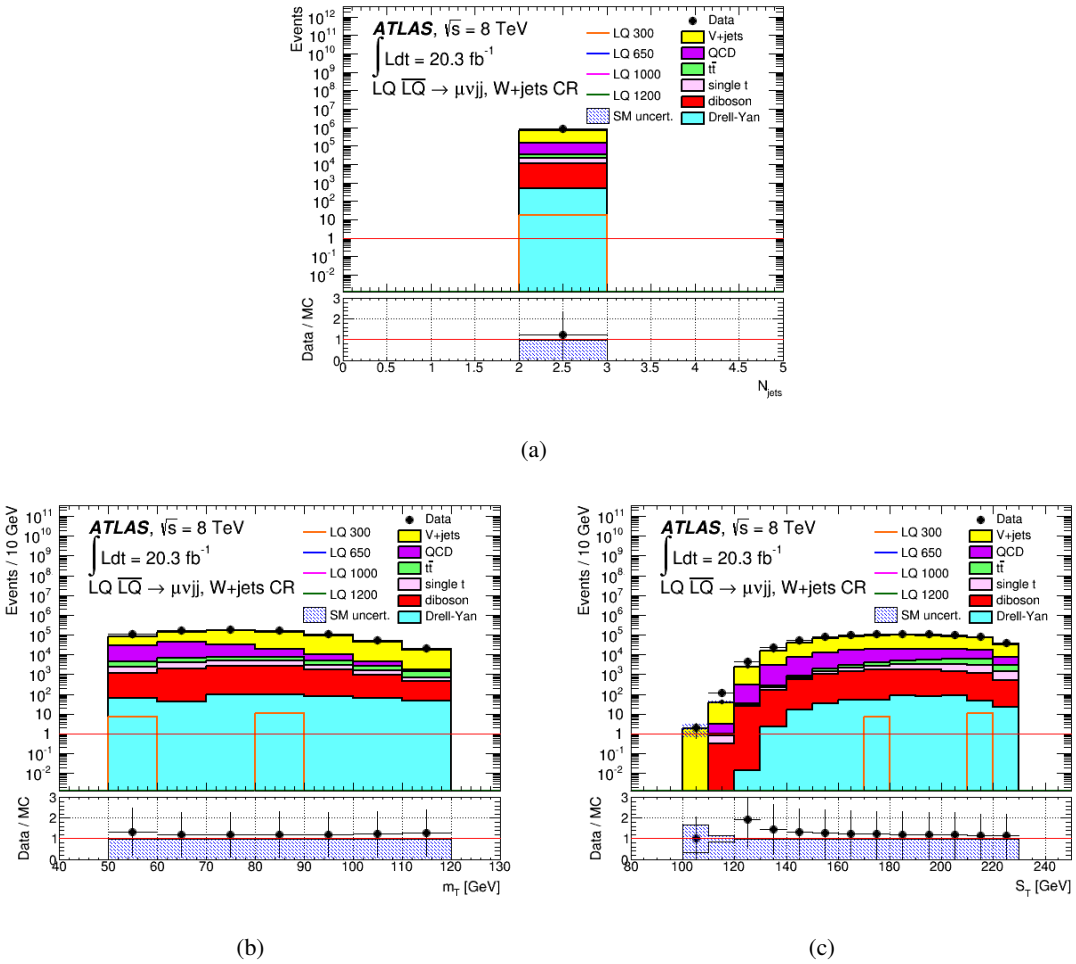
- W+jets (CR A): number of required jets $N_{jets} = 2$, 40 GeV $< m_T <$ 120 GeV and $S_T <$ 225 GeV (W leptonic decay, Fig. 6.15)

- W+jets (CR B): number of required jets $N_{jets} \geq 3$, 40 GeV $< m_T <$ 120 GeV and $S_T <$ 225 GeV (W hadronic decay, App. A.4)

m_T variable request has an 80 GeV width symmetrically distributed around a mean value of 80 GeV. Due to W 's rest mass of 80.385 GeV, such a request results to as less of "W events" exclusion as possible permitting a sufficient statistic population. Besides the number of requested jets, the rest of both W+jets control regions are identical as expected.

The last is $t\bar{t}$ control region

- $t\bar{t}$ (CR C): number of required jets $N_{jets} \geq 4$, 40 GeV $< m_T <$ 120 GeV, leading jet $p_T >$ 50 GeV and subleading jet $p_T >$ 40 GeV (Fig. 6.19)



(a)

(b)

(c)

Figure 6.15: W+jets (CR A) control region for W boson 's leptonic decay.

where the dominant background is $t\bar{t}$.

At the end, only control region A (leptonic W+jets, Fig. 6.16 - Fig. 6.18) and control region C ($t\bar{t}$, Fig. 6.19 - Fig. 6.21) have been used for the final statistical analysis and exclusion limits extraction. Table 6.8 contains all the background, data and signal yields aqcuired at the end of pre-selection cutflow, control and signal

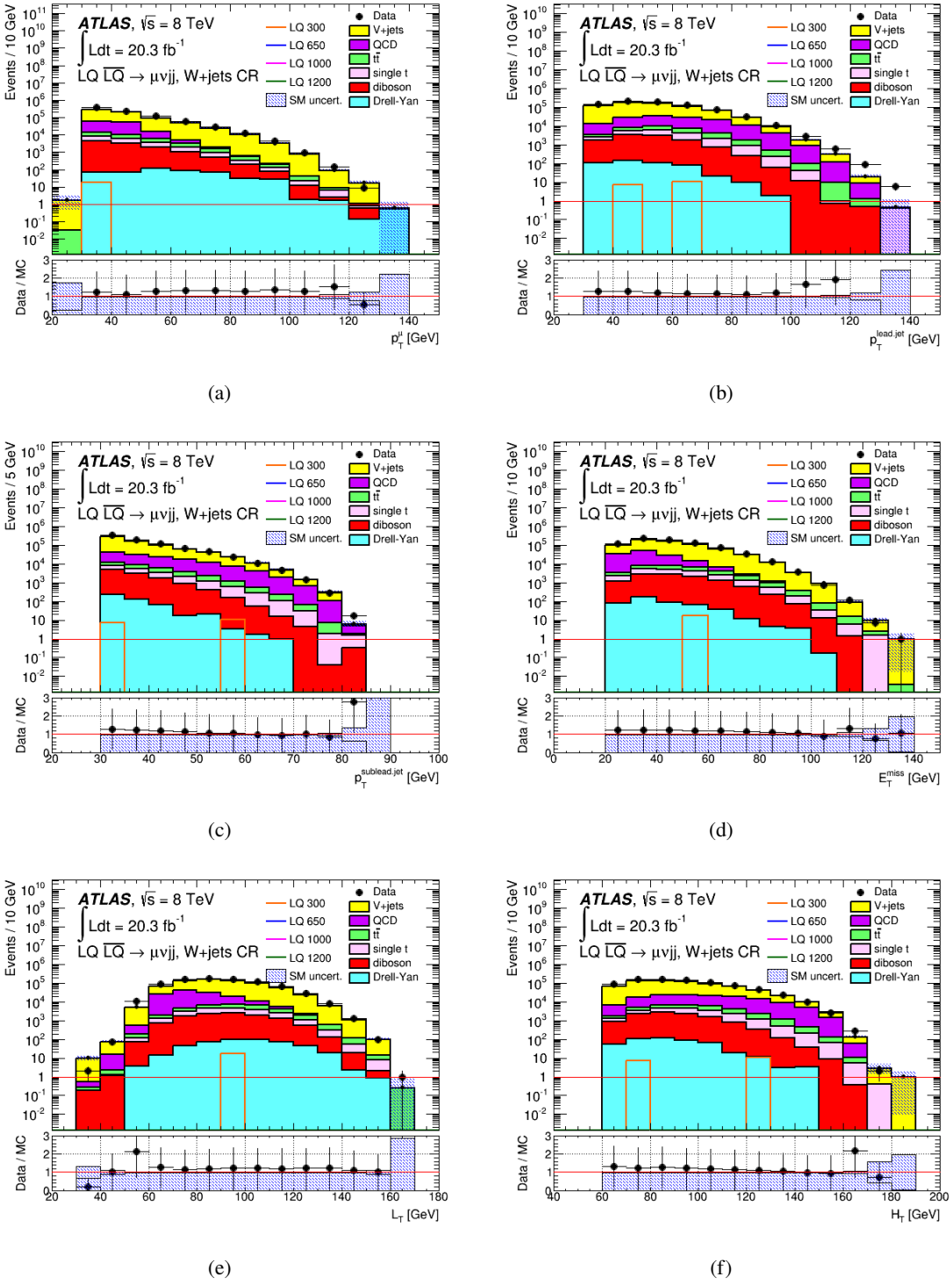
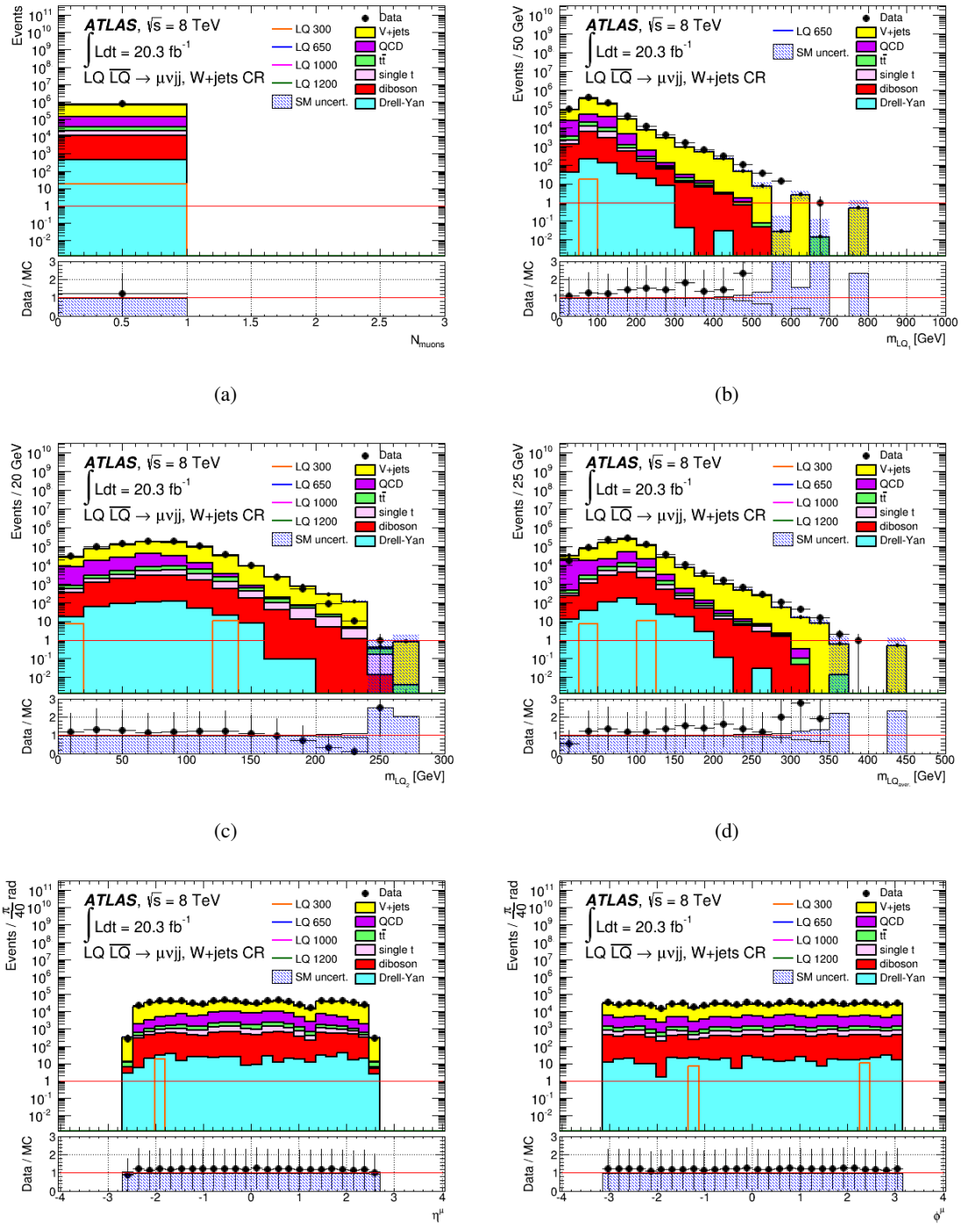


Figure 6.16: W+jets (CR A) control region for W boson 's leptonic decay.

regions. It becomes obvious that in both jet p_Ts and the variables containing one or both jet p_Ts (e.g. L_T, H_T, S_T, LQ masses etc) data and MC do not agree completely, especially in W+jets CR. This is probably due to poor generator description of



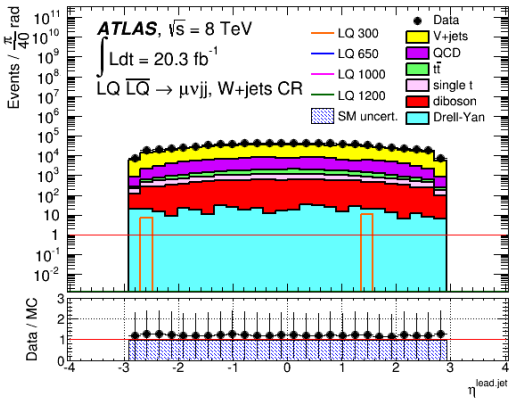
(a) (b)

(c) (d)

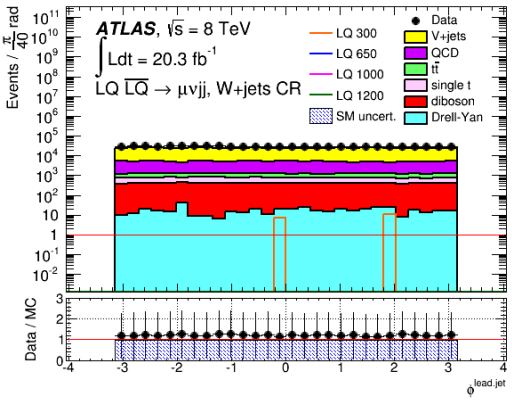
(e) (f)

Figure 6.17: W+jets (CR A) control region for W boson 's leptonic decay.

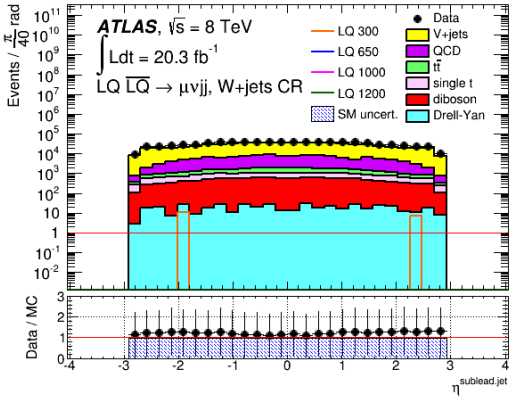
jets. These results have only undergone through event and objects selections and the application of various weights. The correspondent yields after final statistical analysis follow in Sec. 6.9.



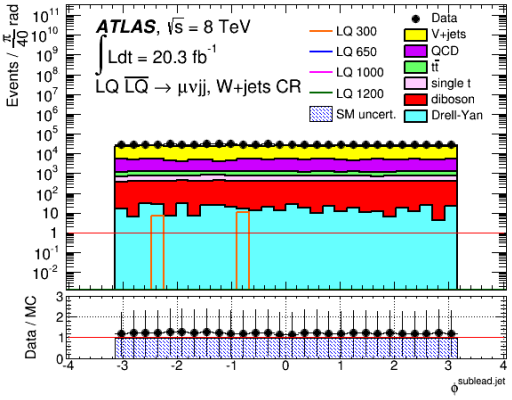
(a)



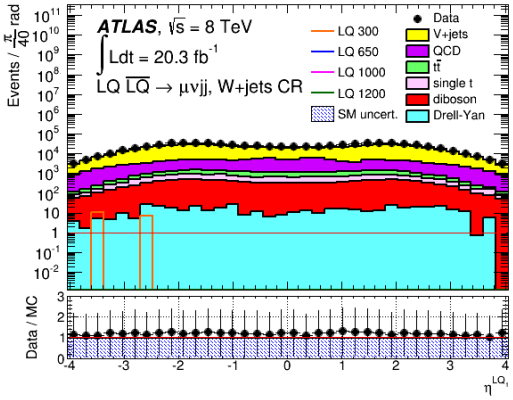
(b)



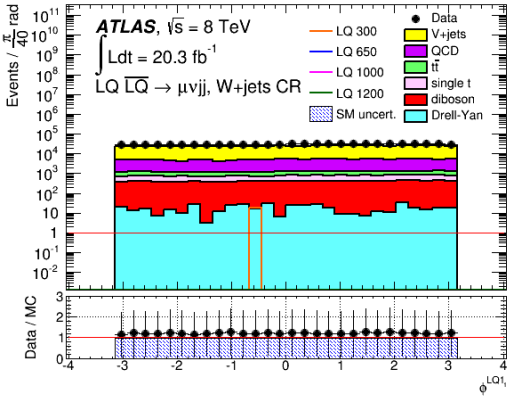
(c)



(d)



(e)

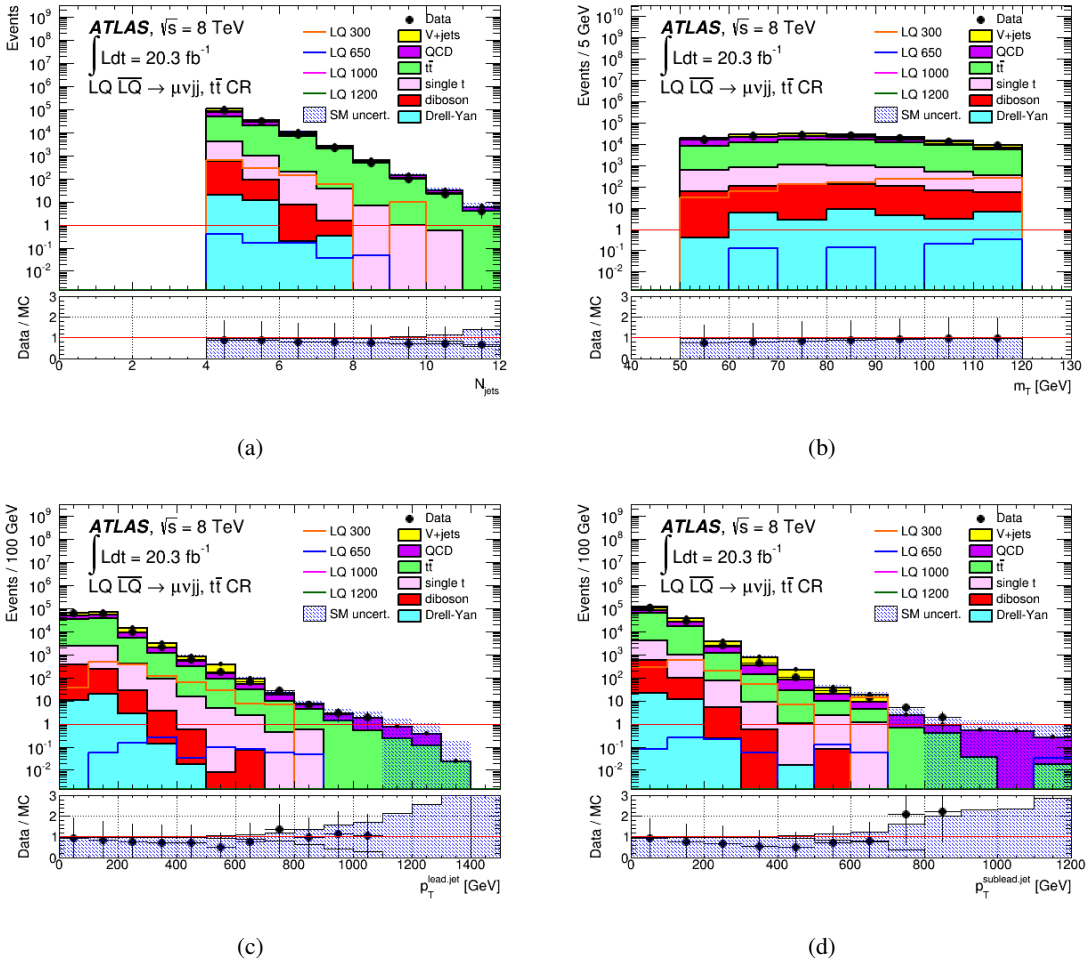


(f)

Figure 6.18: W+jets (CR A) control region for W boson 's leptonic decay.

Event yields	End of pre-selection	W+jets [CR A]	$t\bar{t}$ [CR C]	SR
V+jets	$1,117,590 \pm 9,567$	$542,828 \pm 3,200$	$37,377 \pm 2,064$	5.5 ± 7.5
$t\bar{t}$	$214,424 \pm 4,570$	$13,484 \pm 459$	$73,383 \pm 2,703$	0.6 ± 3.3
Diboson	$21,977 \pm 1,225$	$10,937 \pm 447$	613 ± 215	–
Single top	$40,544 \pm 1,839$	$10,098 \pm 406$	$4,559 \pm 658$	–
Drell-Yan	$1,633 \pm 383$	473 ± 88	32 ± 35	0.2 ± 4.4
QCD	$294,969 \pm 4,490$	$163,778 \pm 1,752$	$12,895 \pm 1,173$	0.5 ± 2.7
TOTAL BACKGROUND	$1,691,137 \pm 12,147$	$741,598 \pm 3,735$	$128,859 \pm 3,736$	6.8 ± 18.0
Data	$1,865,887 \pm 12,376$	$834,746 \pm 3,997$	$133,579 \pm 3,708$	9 ± 9
LQ 300	$21,656 \pm 2,013$	18 ± 6	$1,110 \pm 308$	44 ± 20
LQ 350	$8,641 \pm 1,339$	–	256 ± 130	61 ± 36
LQ 400	$4,031 \pm 955$	–	88 ± 73	36 ± 32
LQ 450	$1,915 \pm 680$	–	32 ± 39	29 ± 38
LQ 500	953 ± 492	–	8 ± 15	19 ± 33
LQ 550	505 ± 374	–	4 ± 10	17 ± 39
LQ 600	280 ± 282	0.1 ± 0.3	1 ± 5	28 ± 65
LQ 650	157 ± 218	–	0.8 ± 3.6	30 ± 78
LQ 700	90 ± 170	–	0.3 ± 1.7	23 ± 72
LQ 750	53 ± 131	–	0.1 ± 0.6	17 ± 64
LQ 800	31 ± 103	–	0.1 ± 0.6	12 ± 56
LQ 850	19 ± 82	–	–	8 ± 48
LQ 900	12 ± 66	–	–	6 ± 41
LQ 950	7 ± 53	–	–	4 ± 35
LQ 1000	5 ± 43	–	–	3 ± 29
LQ 1050	3 ± 35	–	–	2 ± 25
LQ 1100	2 ± 28	–	–	1 ± 21
LQ 1150	1 ± 23	–	–	1 ± 17
LQ 1200	1 ± 19	–	–	1 ± 14

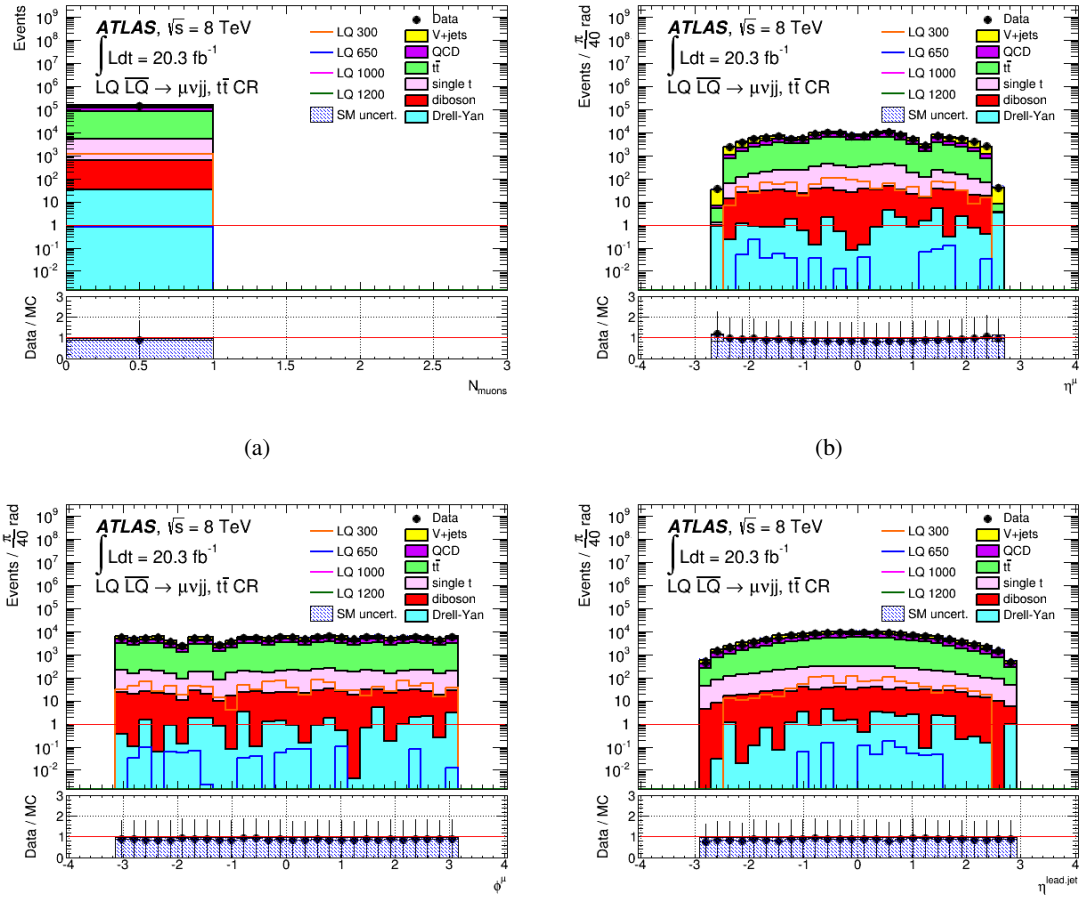
Table 6.8: Summary of yields before signal regions (at the end of pre-selection cut flow), from control and signal regions. The errors appeared are only statistical.



(a) (b)

(c) (d)

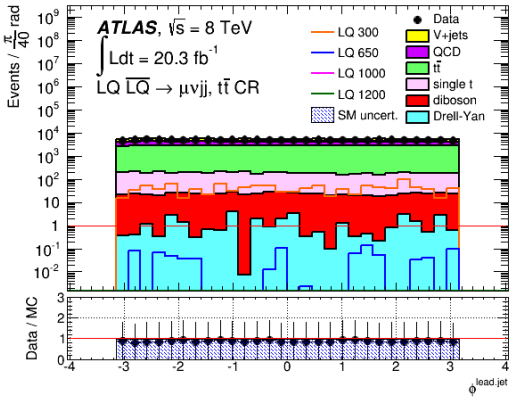
Figure 6.19: $t\bar{t}$ control region



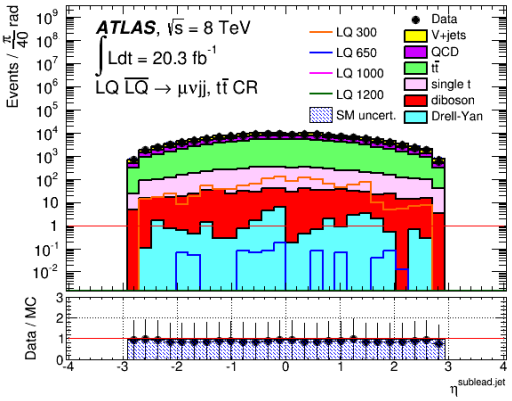
(a) (b)

(c) (d)

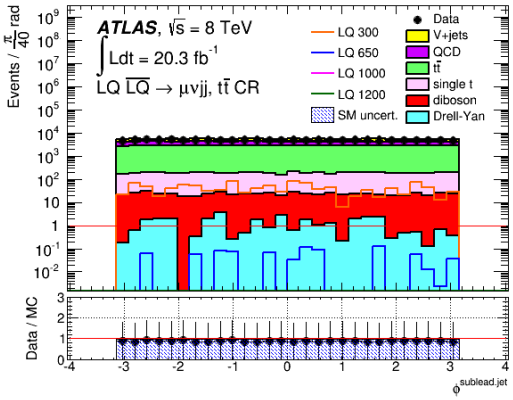
Figure 6.20: $t\bar{t}$ control region



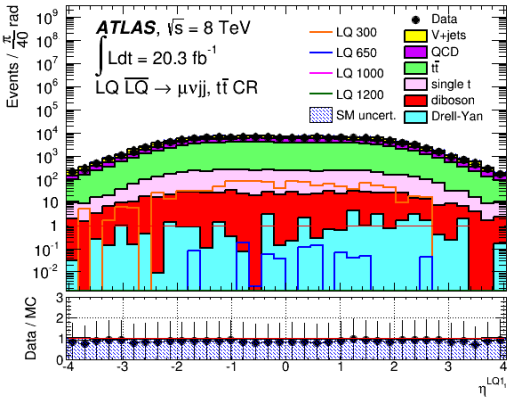
(a)



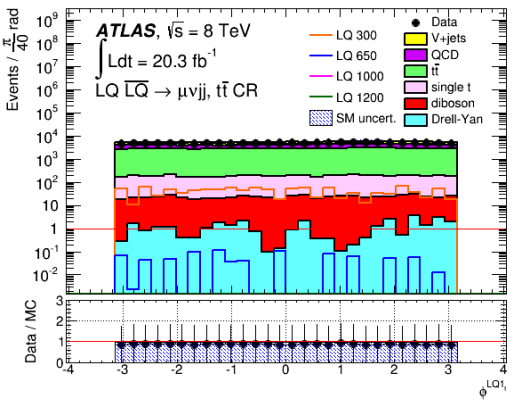
(b)



(c)



(d)



(e)

Figure 6.21: $t\bar{t}$ control region

6.7 Systematics

6.7.1 Jet uncertainties

There are two main sources of jets uncertainties considered: Jet Energy Scale (JES) and Jet Energy Resolution (JER) uncertainties. The JES and its uncertainty are determined by using various different techniques. Absolute in-situ calibrations are performed using standard candles or physics objects that have better experimental resolution. They are based on the momentum conservation in the event (balance in Z+jets or γ+jets events). Inter-calibration of different detector parts are performed using the same principle and take advantage of multi-jet final states. Additional factors are used to correct for pile-up effects and are derived as a function of the number of primary vertices in the event (in-time pile-up) or as a function of the expected interactions per bunch crossing (which help to constrain the out-of-time pile-up). MC simulations are also used to calibrate the jet response and are derived as a function of the jet flavor. Since the energy resolutions of reconstructed jets in data and MC agree within their uncertainties, the jet energy resolution in MC does not need to be corrected. The uncertainties on JER are used as an extra source of systematic errors.

6.7.2 Muon uncertainties

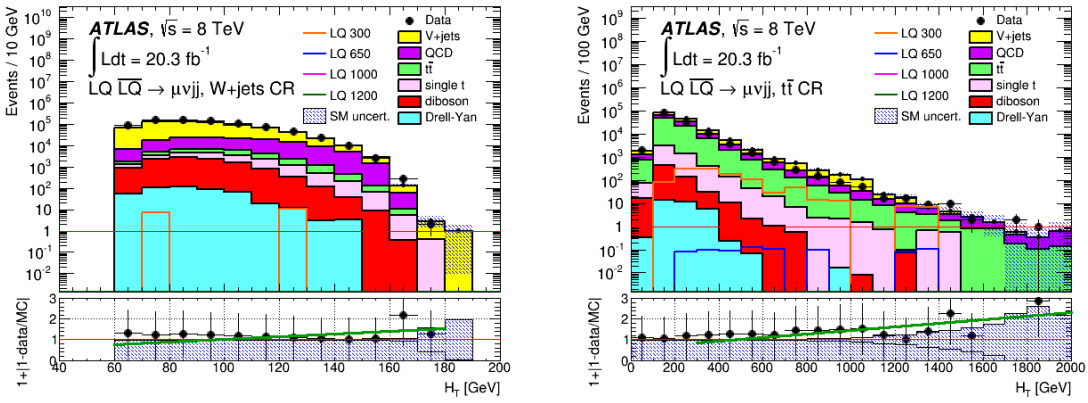
Muons have scaling and smearing corrections applied on their p_T in order to minimize the differences in resolution between data and MC events. The uncertainty on those corrections is propagated as systematic uncertainty. Differences in the identification efficiency, as well as on trigger selection efficiency are taken into account by using scaling factors. The uncertainties on those scaling factors are also propagated as systematic.

In the present analysis there is not a full systematic uncertainty calculation. There has only been a general estimation of 10% overall systematic uncertainty for the final limits extraction.

6.7.3 Residual data-MC differences

An estimation of remaining differences between data and MC can be done by using some relevant kinematic distributions in the control regions area. The distribution used for this purpose is that of the H_T variable for W+jets and $t\bar{t}$ backgrounds, where H_T is the sum of transverse momenta for the first two leading jets. This variable has been chosen for its well measured properties since it does not contain the elusive neutrino as E_T^{miss}. The fitted quantity used as a linear function in both control regions is $1+|1-\frac{data}{MC}|$. The fit is performed for 60 GeV $\leq H_T \leq$ 190 GeV in W+jets and for $H_T \geq$ 300 GeV in $t\bar{t}$. Finally, the bin-per-bin coefficients have been extracted from the H_T distribution in $t\bar{t}$ region given its bigger statistics over W+jets background and they have been used for the rest of the variables fitting in control regions, at the end of pre-selection as well as the signal regions. In Fig. 6.22 appear the H_T distributions for both contol regions after fitting with the use of least squares method.

The best fit values for W+jets control region are



(a) (b)

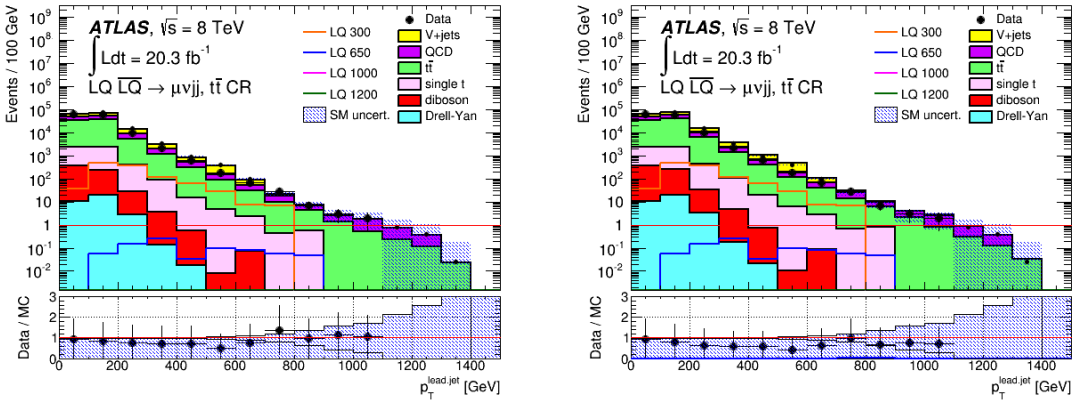
Figure 6.22: Fit of $1+|1-data/MC|$ as a function of H_T in the (a) W+jets and (b) $t\bar{t}$ control regions.

- $p_0 = (37.1 \pm 27.1) \times 10^{-2}$
- $p_1 = (65.5 \pm 21.2) \times 10^{-4}$

and for $t\bar{t}$ control region are

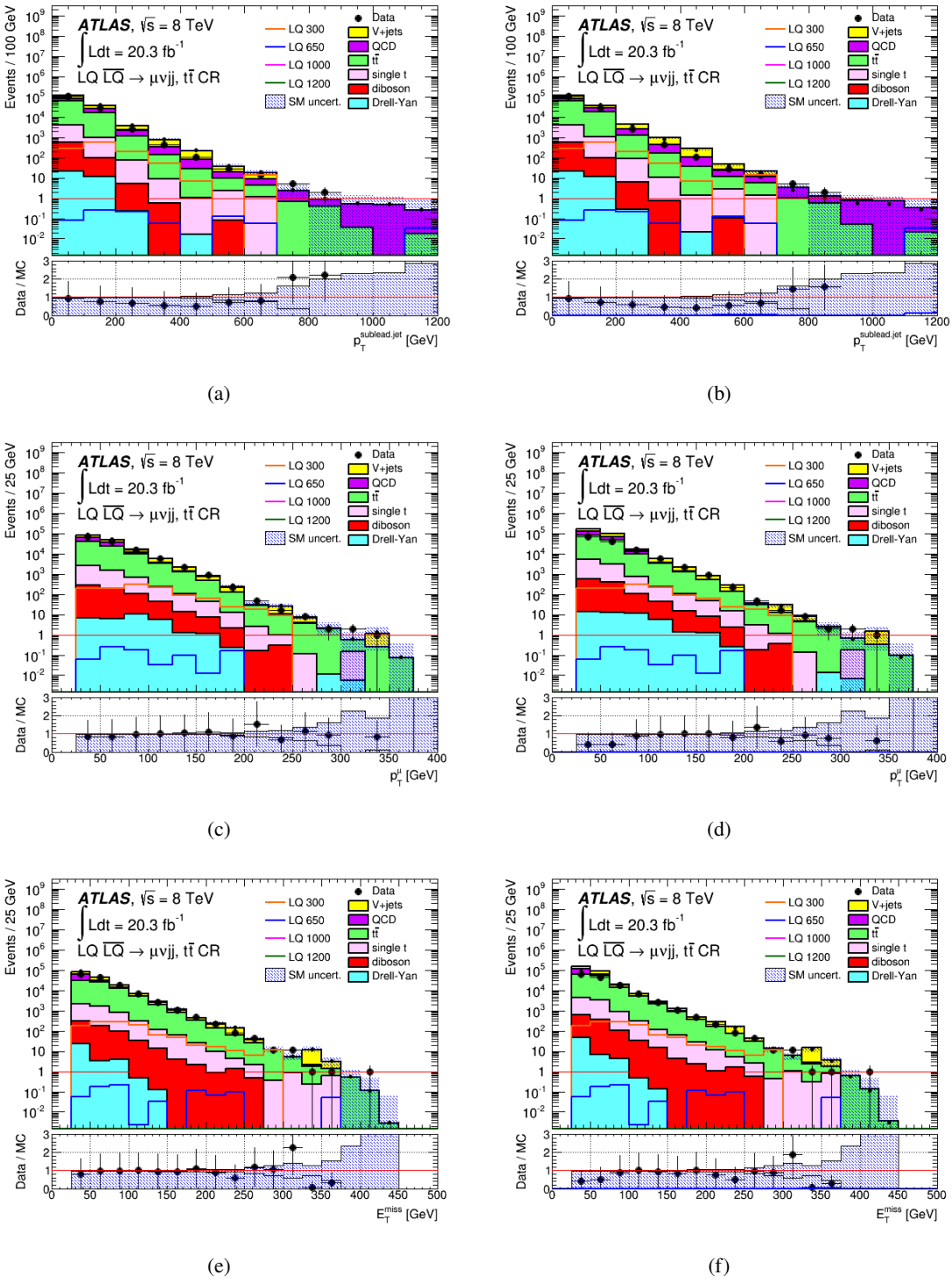
- $p_0 = (61.3 \pm 29.0) \times 10^{-2}$
- $p_1 = (85.6 \pm 23.0) \times 10^{-5}$

where p_0 and p_1 are the coefficients of linear equation $y = p_0 \cdot x + p_1$ ($x \in$ [60,190] for W+jets CR and $x \in$ [300,2000] for $t\bar{t}$ CR). Some kinematic distributions pre-and post-fit comparisons follow (Fig. 6.23 - Fig. 6.31). The improvement from the fitting is negligible (e.g in W+jets CR and SR) or non-existent at all ($t\bar{t}$ CR). For that, the distibutions used for the final statistical evaluation was decided not to include this data-MC fiting.



(a) (b)

Figure 6.23: $t\bar{t}$ control region data-MC residuals bin-per-bin fitting based on H_T variable. On the left column appear the pre-fitted distributions and on the right the post-fitted respectively.

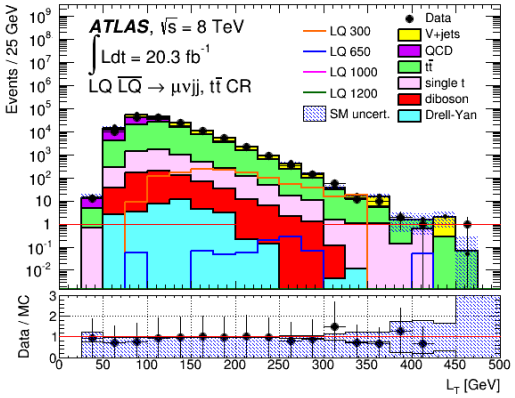


(a) (b)

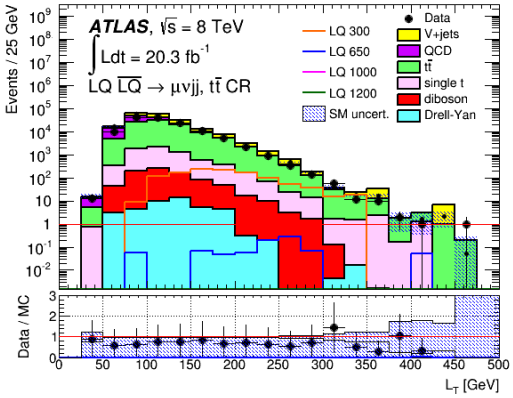
(c) (d)

(e) (f)

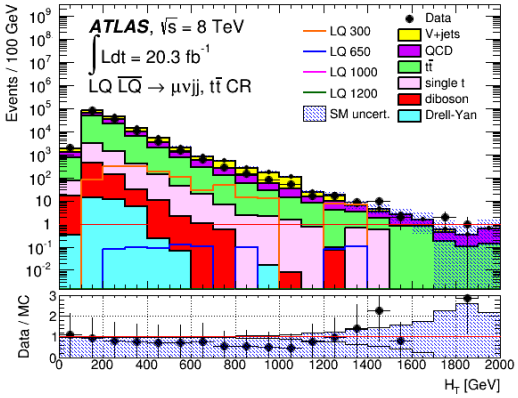
Figure 6.24: $t\bar{t}$ control region data-MC residuals bin-per-bin fitting based on H_T variable. On the left column appear the pre-fitted distributions and on the right the post-fitted respectively.



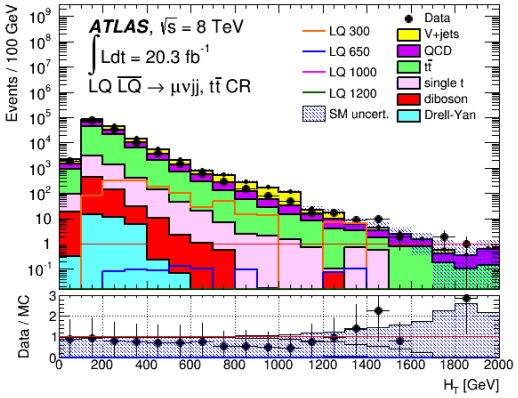
(a)



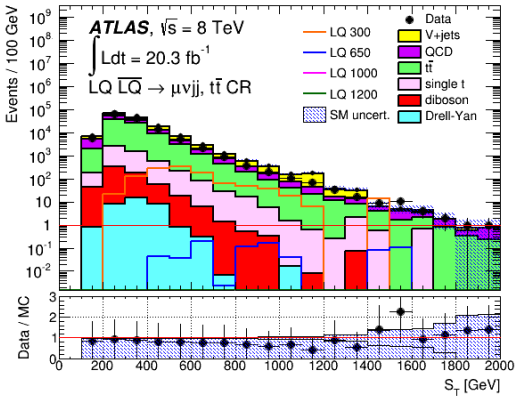
(b)



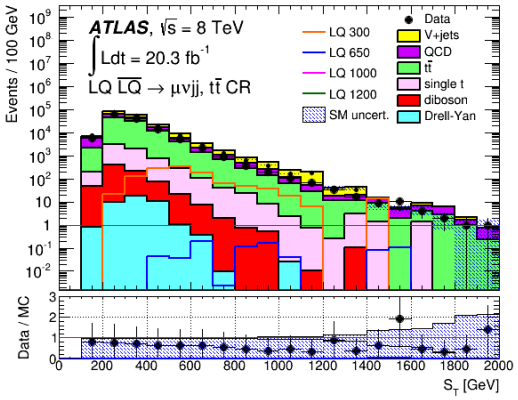
(c)



(d)

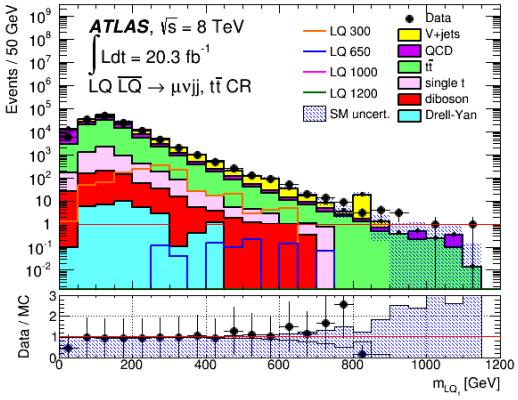


(e)

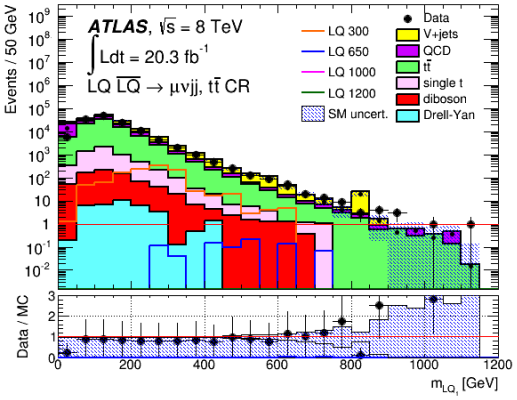


(f)

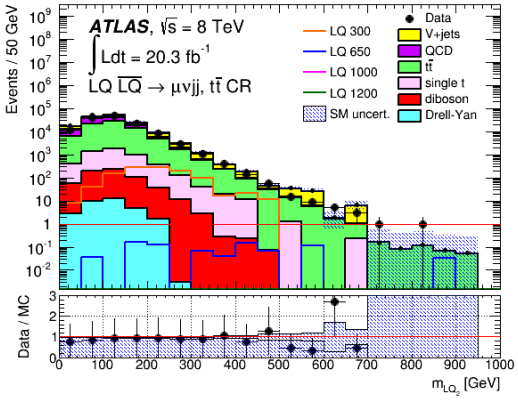
Figure 6.25: $t\bar{t}$ control region data-MC residuals bin-per-bin fitting based on H_T variable. On the left column appear the pre-fitted distributions and on the right the post-fitted respectively.



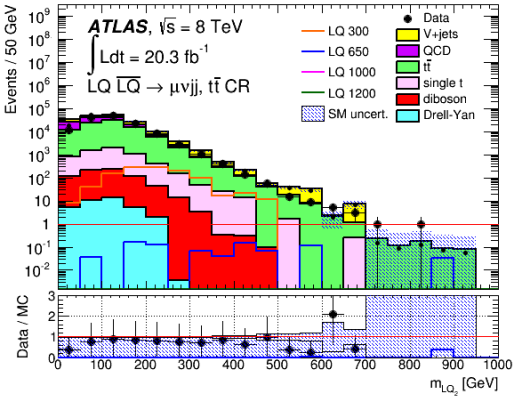
(a)



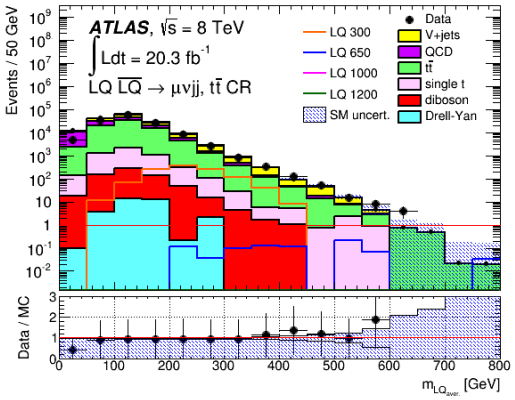
(b)



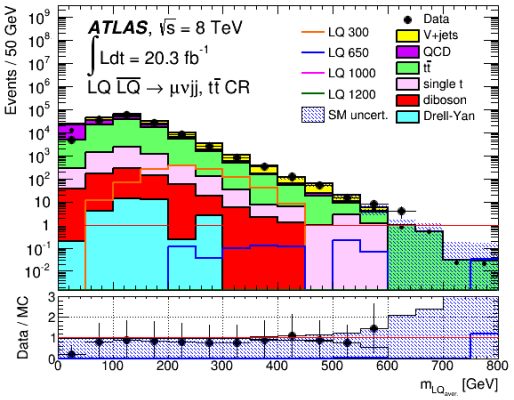
(c)



(d)



(e)



(f)

Figure 6.26: $t\bar{t}$ control region data-MC residuals bin-per-bin fitting based on H_T variable. On the left column appear the pre-fitted distributions and on the right the post-fitted respectively.

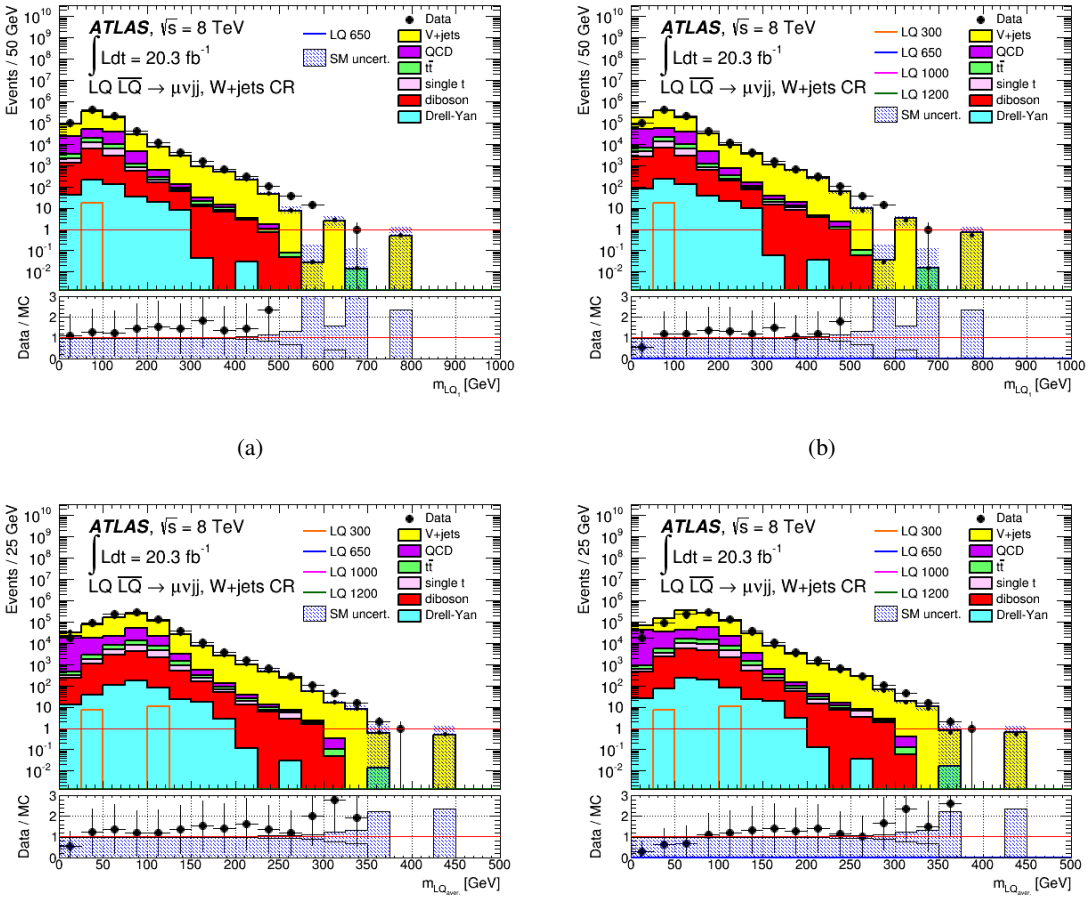
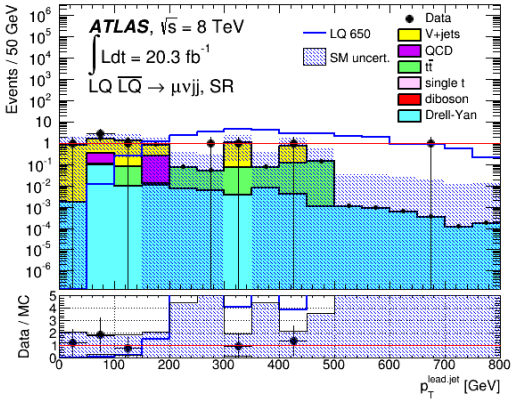
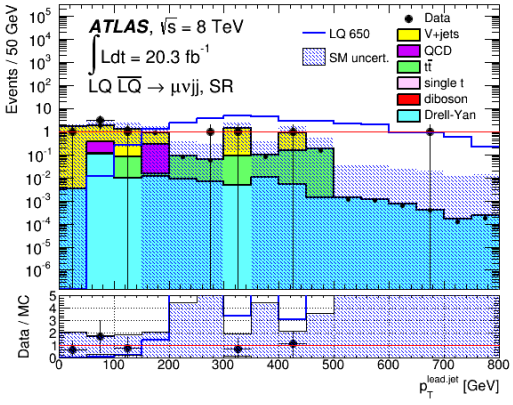


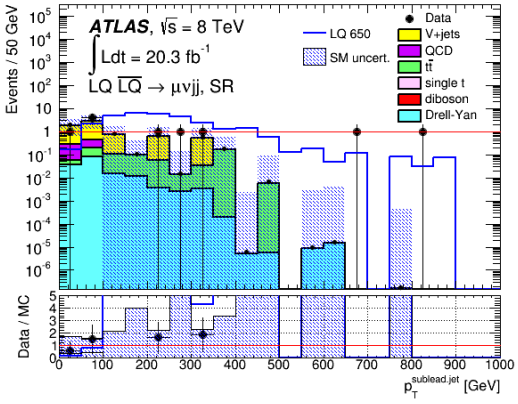
Figure 6.27: W+jets control region data-MC residuals bin-per-bin fitting based on H_T variable. On the left column appear the pre-fitted distributions and on the right the post-fitted respectively.



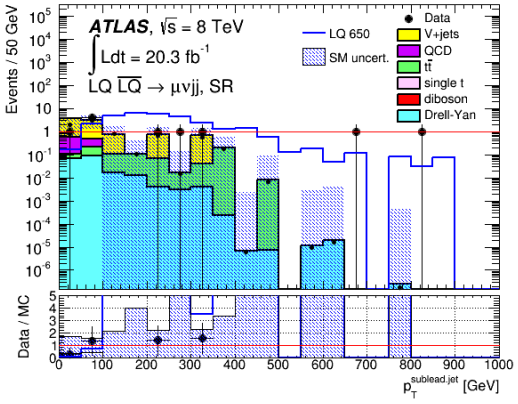
(a)



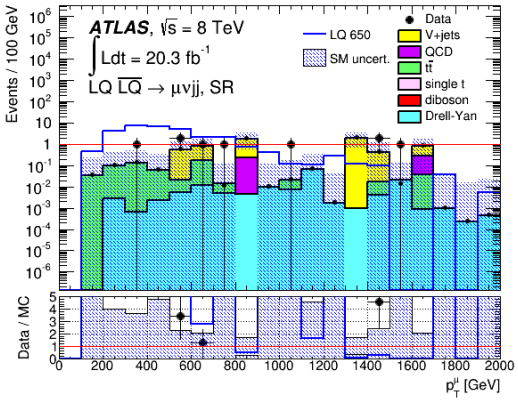
(b)



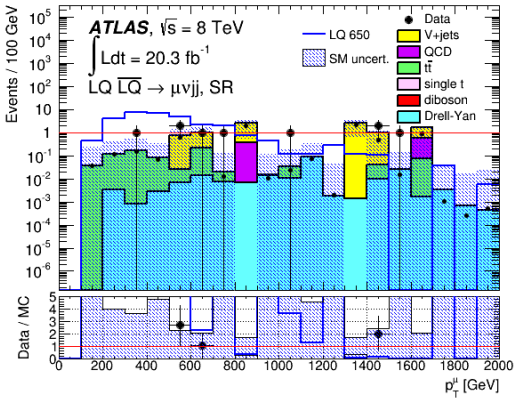
(c)



(d)

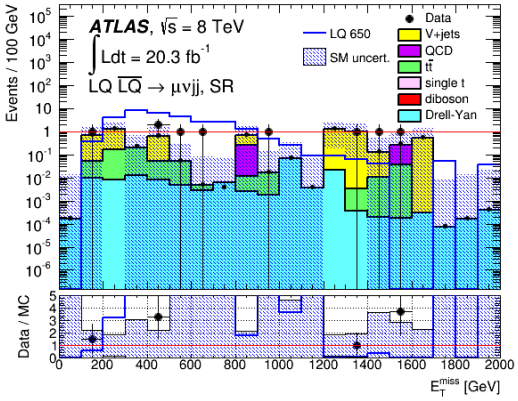


(e)

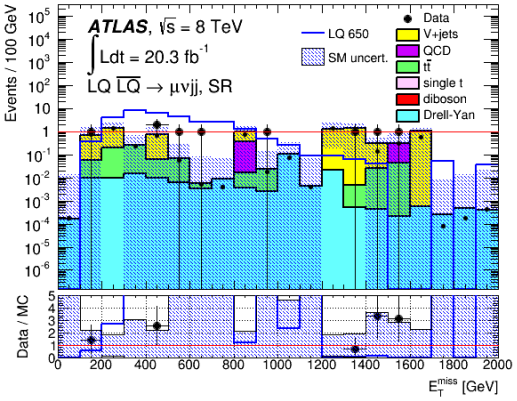


(f)

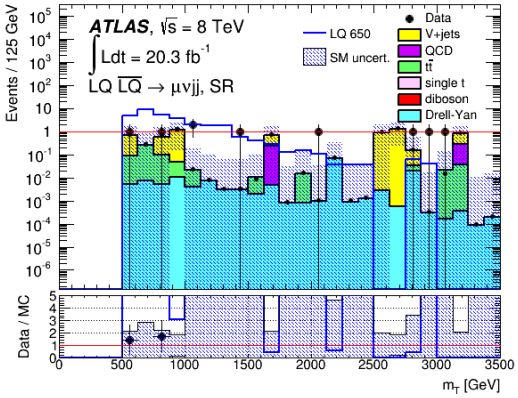
Figure 6.28: SR8 (650 GeV) data-MC residuals bin-per-bin fitting based on H_T variable. On the left column appear the pre-fitted distributions and on the right the post-fitted respectively.



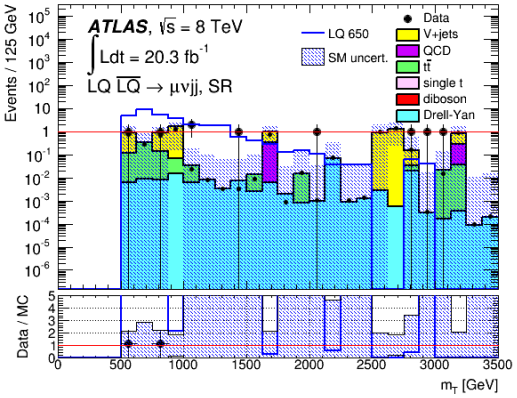
(a)



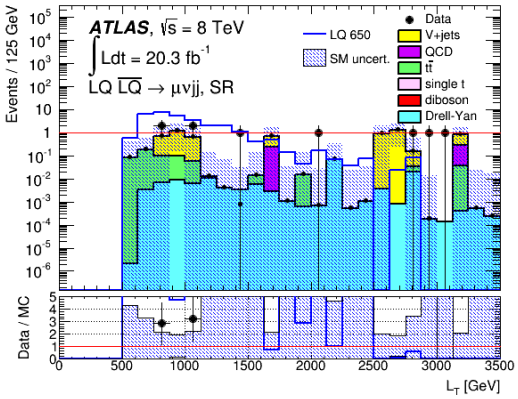
(b)



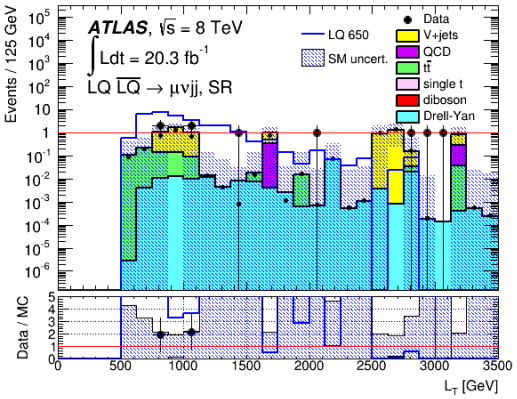
(c)



(d)

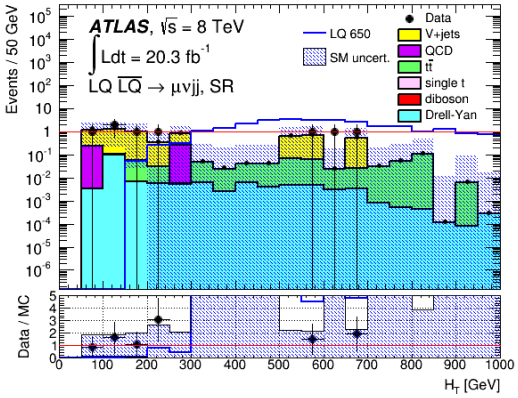


(e)

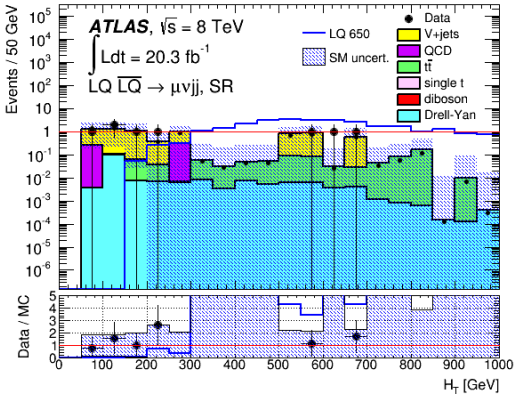


(f)

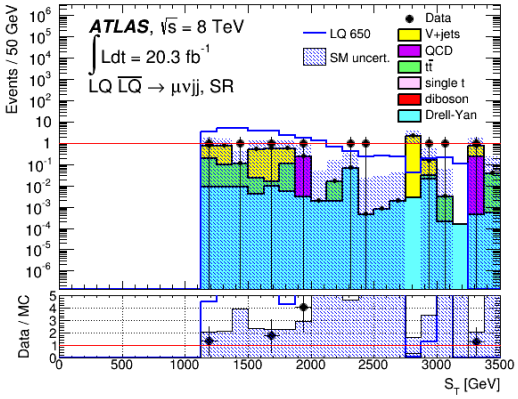
Figure 6.29: SR8 (650 GeV) data-MC residuals bin-per-bin fitting based on H_T variable. On the left column appear the pre-fitted distributions and on the right the post-fitted respectively.



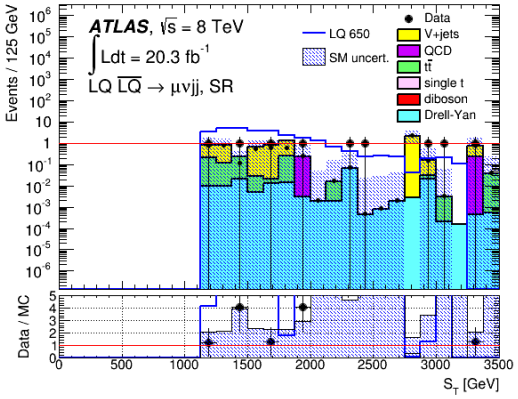
(a)



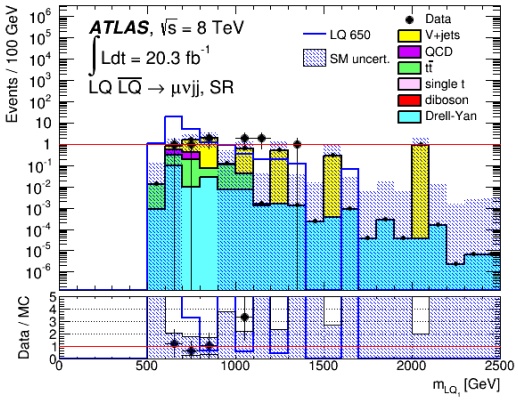
(b)



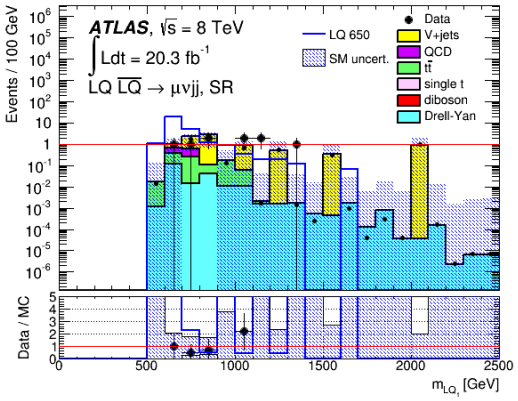
(c)



(d)



(e)



(f)

Figure 6.30: SR8 (650 GeV) data-MC residuals bin-per-bin fitting based on H_T variable. On the left column appear the pre-fitted distributions and on the right the post-fitted respectively.

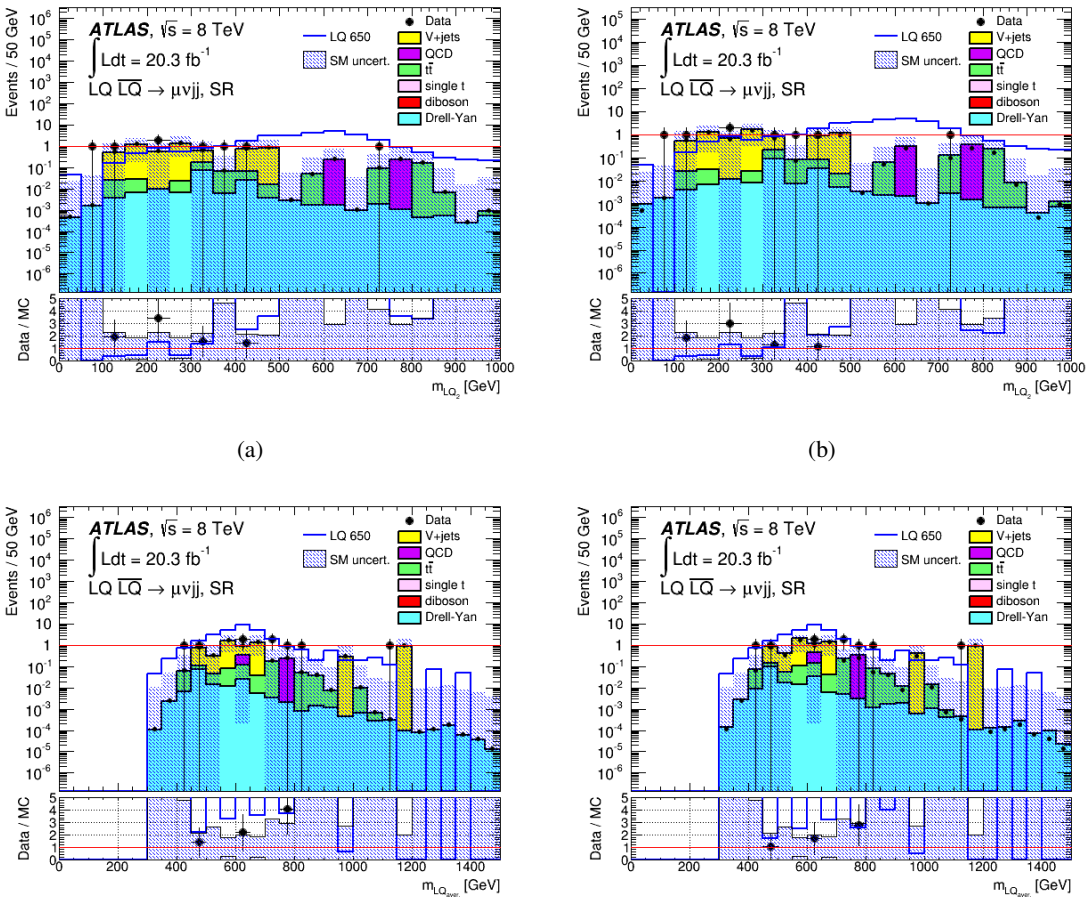


Figure 6.31: SR8 (650 GeV) data-MC residuals bin-per-bin fitting based on H_T variable. On the left column appear the pre-fitted distributions and on the right the post-fitted respectively.

6.8 Signal regions

6.8.1 Definition

In the present analysis there have been defined a set of 19 different signal regions for a more detailed scanning starting from 300 GeV until 1200 GeV with a step of 50 GeV. For their definition there have been used the three variables mentioned in Sec. 6.5.1, m_T, S_T and m_{LQ_1}. In the next section follows the discussion for the selected cut values of these variables for signal regions optimization.

6.8.2 Optimization

The main idea behind signal regions optimization is to select the variables cut values used for their definition in such a way so as to maximize signal-to-SM-background discrimination (significance) maximizing at the same time the ratio of the signal left over total signal (efficiency).

There are two main significance formulas used, Eq. (6.4)

$$Z_1 = \frac{S}{\sqrt{B}} \tag{6.4}$$

and Eq. (6.5) [52]

$$Z_2 = \sqrt{2[(S+B)\cdot \ell n(1+\frac{S}{B}) - S]} \tag{6.5}$$

where S and B are, respectively, the event yields after selection for signal and backgrounds. Efficiency 's definition is the following

$$eff = \frac{S_{left}}{S_{total}} \times 100\% \tag{6.6}$$

where S_{left} is the signal left after each variable cut and S_{total} is the total signal at the end of pre-selection before any signal region cut. Both Z_1 and Z_2 equations produce similar results. For that, the optimization procedure was performed for both formulas in the three dimensional phase space constructed by m_T, S_T and m_{LQ_1}. Two different studies have been held: the first assumes the three variables to be mutually uncorrelated and, thus, the signal significance and efficiency calculations can be done independently; the second makes the assumption that they are not independent and, thus, the cut value assigned to one of them affects the cut values applied to the other two. During the second study, all the variables significance and efficiency calculations are done simultaneously.

In both cases, all the results have been derived from calculations which have made use of the 19 official produced MC LQ signal samples.

6.8.2.1 Uncorrelated variables hypothesis

In the first case, the significance calculation was performed for both formulas (Eq. (6.4) and Eq. (6.5)) for the three signal region definition variables independently for each mass point. Since they are not correlated, the optimization performed over one of the variables does not affect the others and, thus, the calculations are done sequentially. The results are presented and compared in Table 6.9 and Table 6.10.

LQ mass [GeV] (signal region)	$m_T >$ [GeV]	Z_1	Eff. (%)	$S_T >$ [GeV]	Z_1	Eff. [%]	$m_{LQ_1} >$ [GeV]	Z_1	Eff. [%]
300 [SR1]	340	118.50	21.3	565	50.65	60.4	260	40.17	63.3
350 [SR2]	355	70.21	29.6	655	30.27	59.1	310	21.97	60.0
400 [SR3]	365	40.46	35.1	735	18.68	62.5	355	13.65	59.1
450 [SR4]	385	22.89	38.0	795	11.41	67.6	410	8.31	52.7
500 [SR5]	465	14.19	33.2	895	7.19	64.7	440	5.41	58.5
550 [SR6]	465	8.64	38.7	1135	4.93	41.2	495	3.53	54.6
600 [SR7]	565	5.71	31.0	1135	3.40	52.3	535	2.51	56.2
650 [SR8]	565	3.78	37.2	1220	2.38	53.4	595	1.76	51.1
700 [SR9]	565	2.29	40.3	1310	1.66	53.2	645	1.29	49.5
750 [SR10]	565	1.50	45.6	1310	1.17	64.8	665	0.93	55.3
800 [SR11]	565	0.98	52.0	1390	0.77	64.5	665	0.62	64.9
850 [SR12]	565	0.61	53.9	1540	0.54	57.6	745	0.44	58.7
900 [SR13]	575	0.41	57.9	1540	0.38	66.4	790	0.30	56.9
950 [SR14]	770	0.27	40.8	1745	0.26	52.9	850	0.22	53.8
1000 [SR15]	880	0.19	33.9	1790	0.19	55.9	850	0.16	60.3
1050 [SR16]	880	0.13	36.9	1790	0.14	64.3	945	0.11	50.0
1100 [SR17]	880	0.08	39.1	1790	0.09	70.3	1050	0.08	38.0
1150 [SR18]	880	0.06	43.4	1790	0.06	75.0	1050	0.06	48.0
1200 [SR19]	880	0.04	46.6	2130	0.04	54.9	1115	0.05	43.3

Table 6.9: Signal regions optimized with m_T, S_T and m_{LQ_1} (derived from Eq. (6.4) significance formula).

As one can see, the results are almost identical. In Fig. 6.33 significance graphics with their correspondent efficiencies for the three used observables appear for an LQ nominal mass at 650 GeV. Fig. 6.32 concentrates all the cut values for each signal region which have been derived sequentially given the hypothesis that the variables do not affect each other.

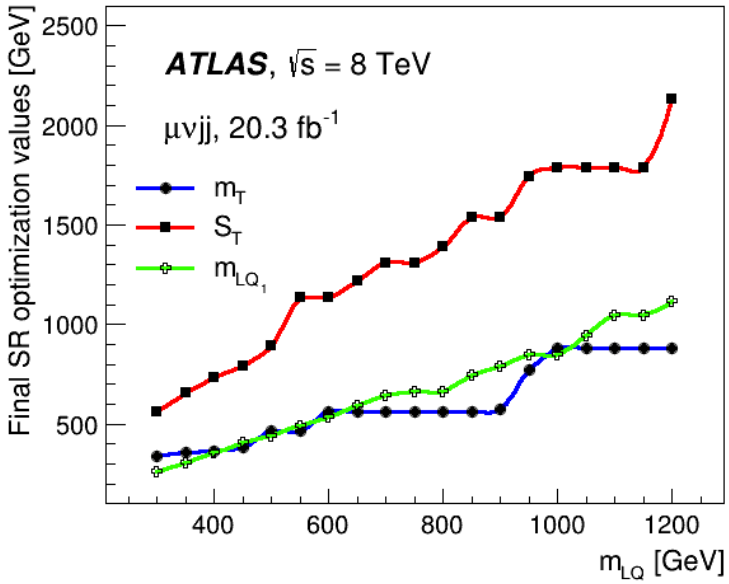


Figure 6.32: Signal regions definition variables cuts after optimization. The values have been acquired under the uncorrelated variables hypothesis.

LQ mass [GeV] (signal region)	$m_T >$ [GeV]	Z_2	Eff. (%)	$S_T >$ [GeV]	Z_2	Eff. [%]	$m_{LQ_1} >$ [GeV]	Z_2	Eff. [%]
300 [SR1]	240	85.76	44.2	550	50.64	64.4	255	39	65.8
350 [SR2]	315	52.63	37.4	665	28.93	59.1	310	21.43	60.0
400 [SR3]	345	32.97	39.1	715	18.05	67.0	355	13.37	59.1
450 [SR4]	375	19.74	40.1	795	11.10	67.5	410	8.15	52.7
500 [SR5]	395	12.46	44.8	895	7.02	64.5	440	5.33	58.5
550 [SR6]	465	7.87	38.7	1135	4.78	41.2	480	3.48	58.9
600 [SR7]	565	5.20	31.0	1135	3.33	52.3	535	2.48	56.2
650 [SR8]	565	3.54	37.2	1220	2.32	53.4	595	1.74	51.1
700 [SR9]	565	2.20	40.3	1265	1.63	58.7	645	1.28	49.5
750 [SR10]	565	1.46	45.6	1310	1.15	64.8	665	0.92	55.3
800 [SR11]	565	0.96	52.0	1390	0.77	64.5	665	0.62	64.8
850 [SR12]	565	0.61	53.8	1540	0.53	57.6	745	0.44	58.7
900 [SR13]	575	0.41	57.9	1540	0.38	66.4	790	0.30	56.9
950 [SR14]	770	0.27	40.8	1745	0.26	52.9	850	0.22	53.8
1000 [SR15]	880	0.18	33.9	1790	0.19	55.9	850	0.16	60.3
1050 [SR16]	880	0.12	36.9	1790	0.14	64.3	945	0.11	50.0
1100 [SR17]	880	0.08	39.1	1790	0.09	70.3	1050	0.08	38.0
1150 [SR18]	880	0.06	43.4	1790	0.06	75.0	1050	0.06	48.0
1200 [SR19]	880	0.04	46.6	2130	0.04	54.9	1115	0.04	43.3

Table 6.10: Signal regions optimized with m_T, S_T and m_{LQ_1} (derived from Eq. (6.5) significance formula).

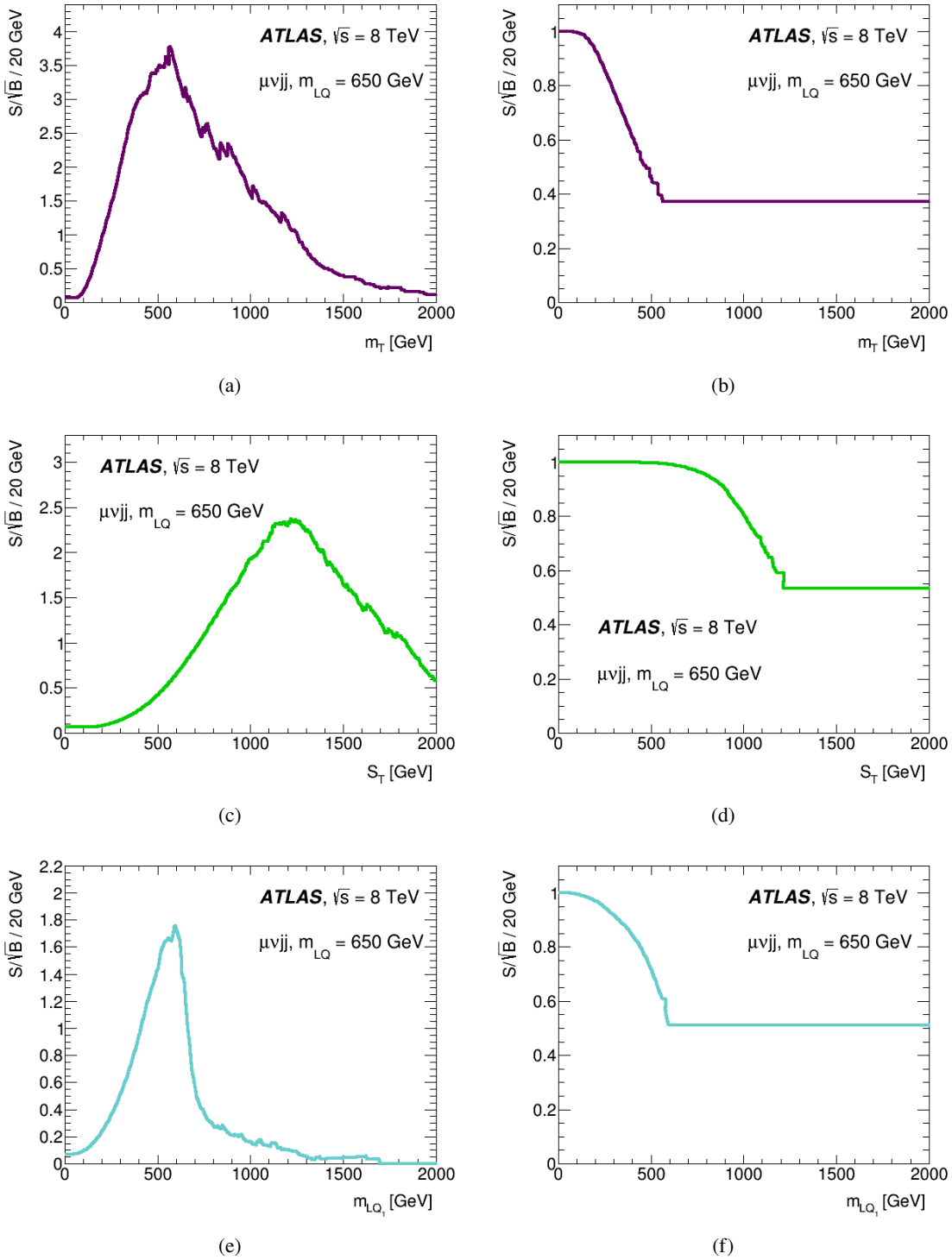


Figure 6.33: (a) m_T, (c) S_T and (e) m_{LQ_1} significance ($Z_1 = \frac{S}{\sqrt{B}}$) and correspondent (b) m_T, (d) S_T and (f) m_{LQ_1} efficiencies for signal mass point at 0.65 TeV.

6.8.2.2 Correlated variables hypothesis

The second study has been based on the assumption the m_T, S_T and m_{LQ_1} variables are correlated. In that case, the optimization cannot be performed sequentially,

but simultaneously over all three of them while trying to determine the best cut values for each mass point at the same time. Here the calculations were based on Z_1 significance formula, since the differences among Z_1 and Z_2 have been proved negligible during the first study.

In order to meet the requirements of the hypothesis, the three definition variables construct a 3-D phase space of cut values. That way it is made feasible the detection of the three-coordinates point where signal versus background discrimination is maximized. Afterwards, these cut values are feeded to Z_1 formula for the significance calculation and the correspondent efficiency for each variable separately. Fig. 6.34 presents two aspects of the same 3-D histogram containing m_T, S_T and m_{LQ_1}. This histogram is used for the simultaneous optimization. In Fig. 6.34a the angle is suitable for the better axis reading and in Fig. 6.34b is easier pointing the red hot spot since it is turned that way so as to watch it from the side (notice that the m_T axis cannot be seen since it is lost in histogram 's perspective).

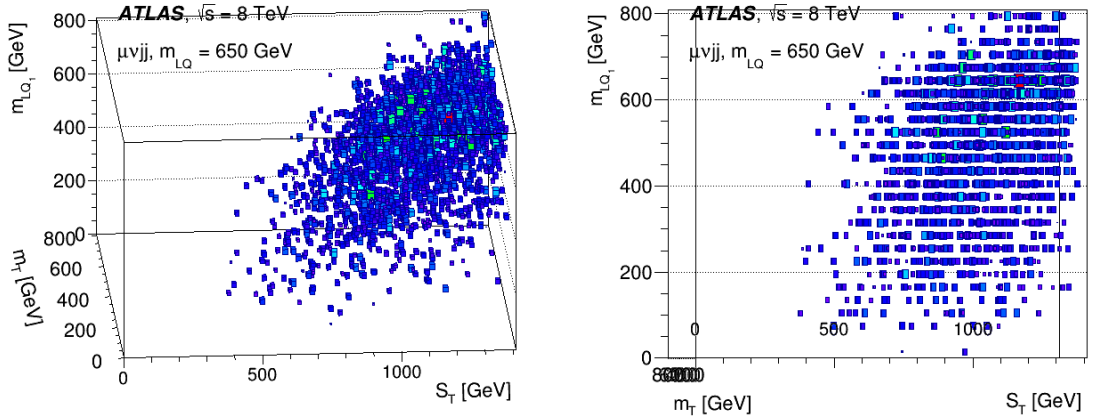


Figure 6.34: m_T, S_T and m_{LQ_1} simultaneous optimization 3-D histogram for signal mass point at 650 GeV. The red hot spot appears on the upper right corner of the histogram.

It follows Table 6.11 with the significance and efficiency results. In Fig. 6.35 the cut values for all signal regions are presented which have been acquired from the simultaneous signal optimization. In App.A.7 the rest 3D plots for all the SRs are presented.

6.8.2.3 Followed method

Comparing the acquired numbers from Table 6.9 and Table 6.11 it becomes obvious that the S_T and m_{LQ_1} cut values are not significantly different. The m_T values have a larger deviation between the two methods. In general, the above results imply that the variables are not totally independant, although that dependancy does not appear to be very strong.

The significance from the second method is analogous to that of the first one (S_T, m_{LQ_1}) or smaller (m_T). On the other hand, the efficiency appears almost analogous to the first method for S_T and m_{LQ_1}), but significantly larger for the m_T variable.

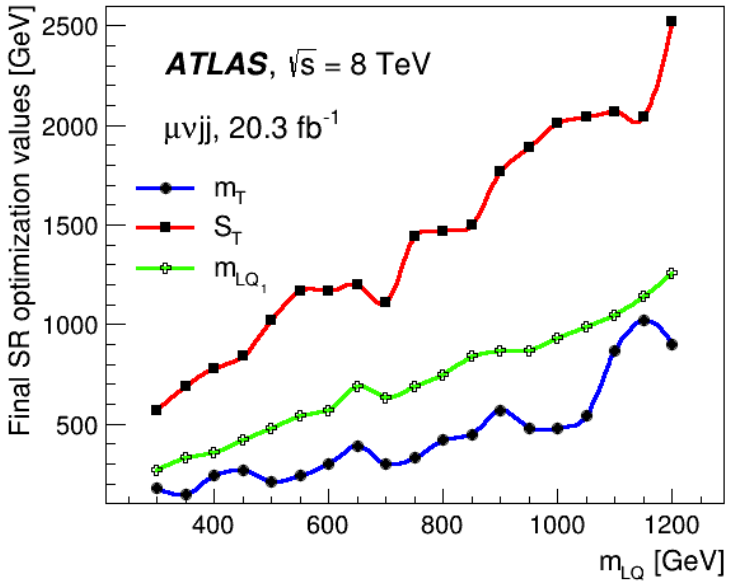


Figure 6.35: Signal regions definition variables cuts after optimization. The values have been acquired under the correlated variables hypothesis.

At the end, it was considered more preferable to maximize the signal-background discrimination in expence to a somewhat larger signal efficiency loss. Thus, the final

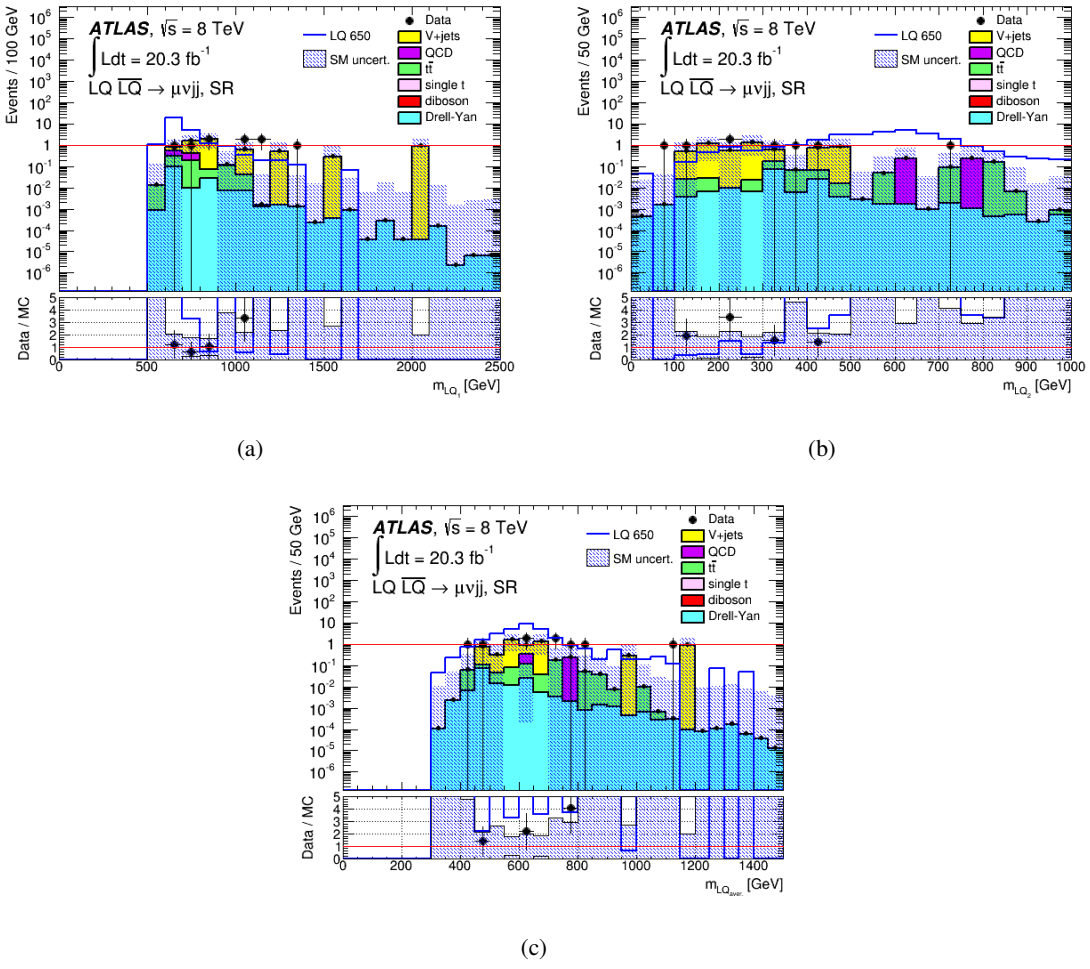
LQ mass [GeV] (signal region)	$m_T >$ [GeV]	Z_1	Eff. (%)	$S_T >$ [GeV]	Z_1	Eff. [%]	$m_{LQ_1} >$ [GeV]	Z_1	Eff. [%]
300 [SR1]	180	78.00	65.5	570	53.06	59.0	270	40.01	58.0
350 [SR2]	150	25.21	84.9	690	29.73	53.2	330	20.04	47.6
400 [SR3]	240	27.58	64.7	780	18.15	52.4	360	13.53	57.0
450 [SR4]	270	16.18	62.9	840	11.37	58.8	420	8.12	48.5
500 [SR5]	210	5.99	83.1	1020	6.76	42.3	480	4.73	41.2
550 [SR6]	240	4.08	80.5	1170	4.74	36.9	540	2.67	31.8
600 [SR7]	300	3.47	74.1	1170	3.31	47.5	570	2.17	42.1
650 [SR8]	390	2.95	63.0	1200	2.29	55.8	630	1.50	36.3
700 [SR9]	300	1.18	81.4	1110	1.39	76.8	630	1.24	53.9
750 [SR10]	330	0.86	80.7	1440	1.09	48.2	690	0.86	48.3
800 [SR11]	420	0.70	70.7	1470	0.75	55.1	750	0.55	44.0
850 [SR12]	450	0.46	68.5	1500	0.53	61.9	840	0.28	28.6
900 [SR13]	570	0.41	58.2	1770	0.35	42.2	870	0.24	34.5
950 [SR14]	480	0.21	72.6	1890	0.22	38.4	870	0.20	48.1
1000 [SR15]	480	0.14	75.4	2010	0.13	34.2	930	0.14	44.7
1050 [SR16]	540	0.10	70.2	2040	0.10	40.8	990	0.09	39.1
1100 [SR17]	870	0.07	40.2	2070	0.06	45.6	1050	0.08	38.0
1150 [SR18]	1020	0.05	32.1	2040	0.05	55.5	1140	0.04	25.4
1200 [SR19]	900	0.04	44.7	2520	0.03	23.7	1360	0.01	11.6

Table 6.11: Signal regions optimized with m_T, S_T and m_{LQ_1} (derived from Eq. (6.4) significance formula).

cut values applied for the definition of signal regions are that of Table 6.9.

6.8.3 Signal Region histograms

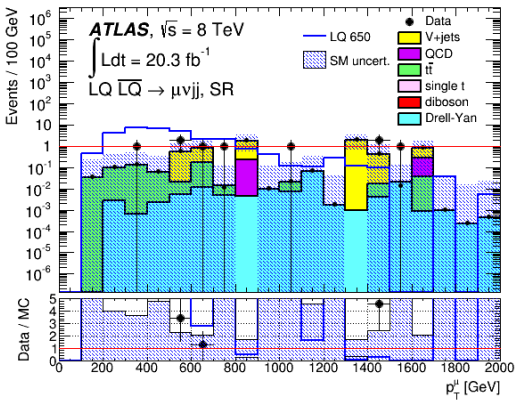
Fig. 6.36 of the two LQ mass distributions in SR8 (650 GeV) before the statistical analysis follows. SR8, like all the rest SRs, has been defined with the use of Table 6.10. **No significant excess in bins in the vicinity of the area of given signal sensitivity has been observed for both SR LQ masses that could suggest something more than a statistical fluctuation**. In case of some signal excess, this is observed to regions far from 650 GeV and with a rather small statistic. The red line denotes "1" (only one event) to make easier histogram reading given the SRs small bin event population. In Fig. 6.37 and Fig. 6.38 there are more distributions in SRs along with the already presented in Sec. 6.7.3. In Sec. 6.9 takes place an extensive statistical analysis with the use of HistFitter package for further investigation and new exclusion limits extraction. More plots on significance and efficiency calculations can be found in App. A.5 and App. A.6 respectively.



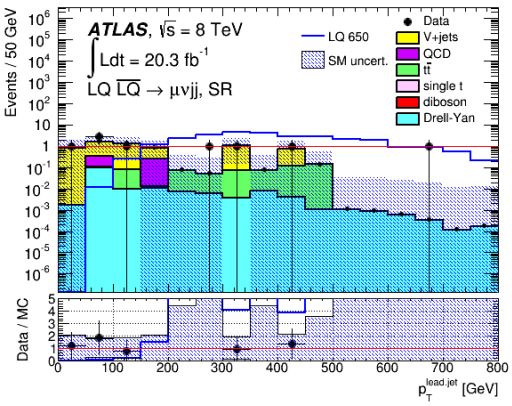
(a) (b)

(c)

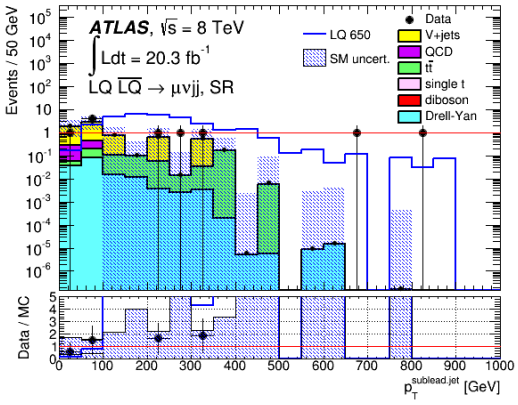
Figure 6.36: Both LQ mass distributions, (a) - (b), obtained through mass window method along with average LQ mass (c) in SR8 for a nominal LQ signal mass of 650 GeV ($m_T > 565$ GeV, $S_T > 1220$ GeV and $m_{LQ_1} > 595$ GeV) before statistical analysis.



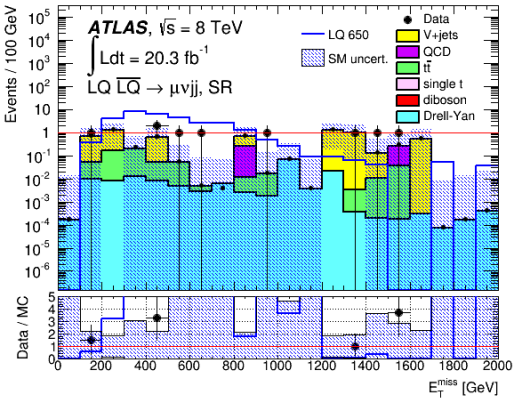
(a)



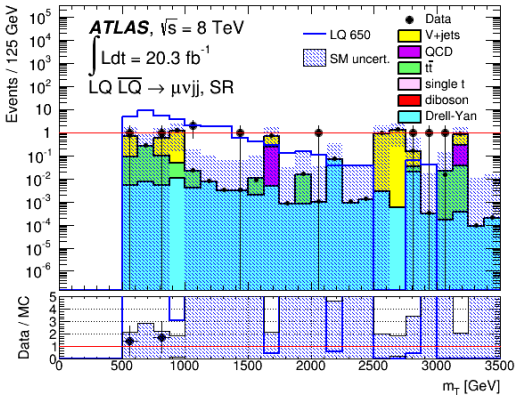
(b)



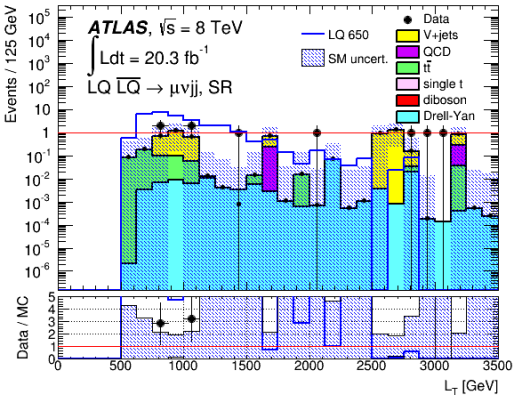
(c)



(d)



(e)



(f)

Figure 6.37: Distributions obtained in SR8 for a nominal LQ signal mass of 650 GeV before statistical analysis.

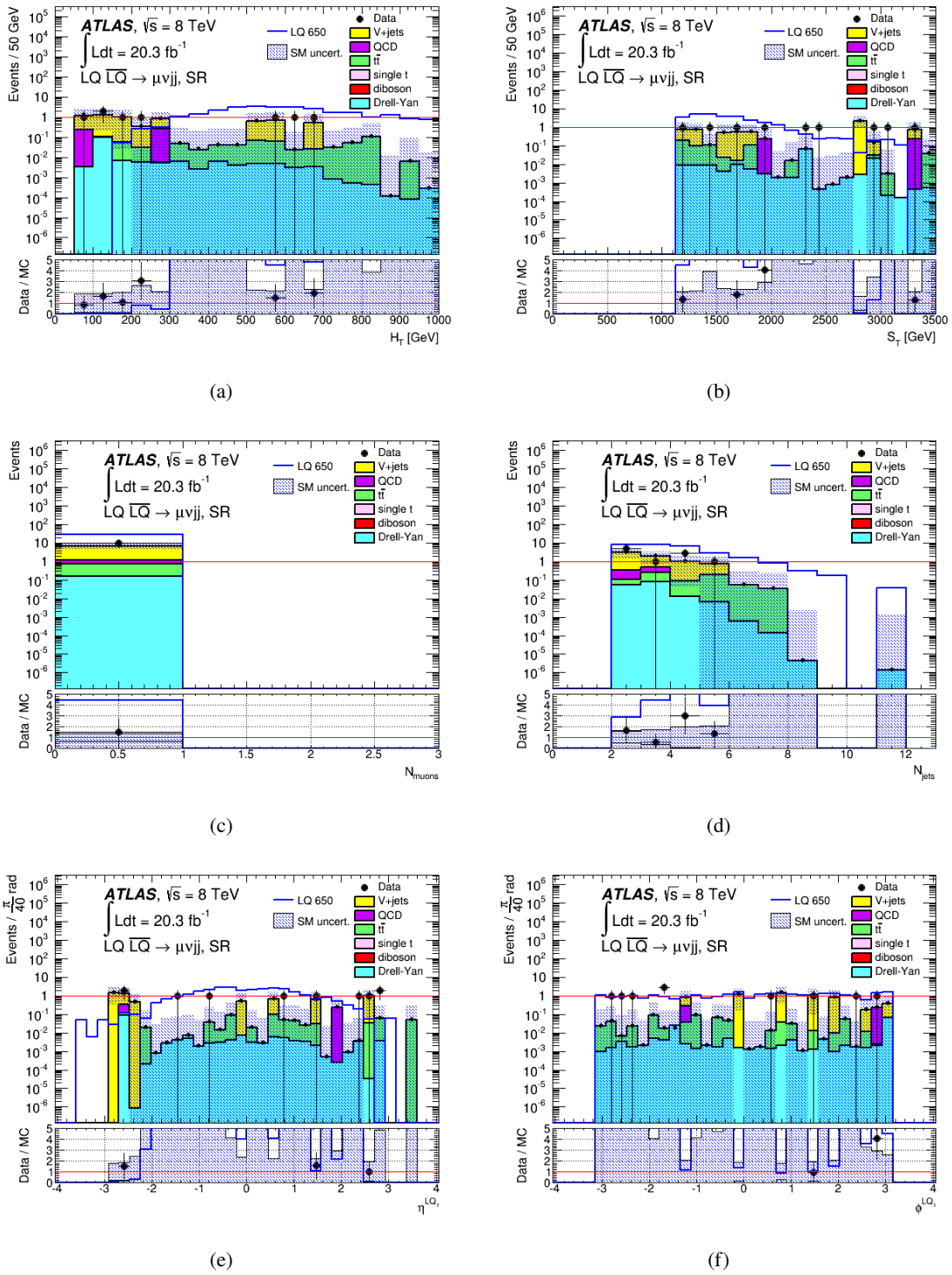


Figure 6.38: Distributions obtained in SR8 for a nominal LQ signal mass of 650 GeV before statistical analysis.

6.9 Statistical analysis

6.9.1 Statistical signal significance calculation

The optimal method to distinguish a leptoquark signal above the standard model background is given by the likelihood ratio test statistic as shown in Eq. (6.7). In this equation, $\mathbf{x}$ represents the data, b is the estimated background yield, s is the predicted signal yield and $\mathcal{L}$ is the likelihood function, which for this application is a Poisson probability distribution function. The fully expanded likelihood ratio test statistics is shown in Eq. (6.8)

$$\Lambda(\mathbf{x}) = \frac{\mathcal{L}(s+b|\mathbf{x})}{\mathcal{L}(b|\mathbf{x})} \tag{6.7}$$

$$\Lambda(\mathbf{x}) = \frac{\frac{(s+b)^{\mathbf{x}}e^{-(s+b)}}{\mathbf{x}!}}{\frac{b^{\mathbf{x}}e^{-b}}{\mathbf{x}!}} \tag{6.8}$$

It is more common in LHC collaborations to use the logarithm of the likelihood ratio (LLR) or more specifically negative two multiplied by the logarithm of Λ as shown in Eq. (6.9)

$$LLR(\mathbf{x}) = -2\ell og(\Lambda(\mathbf{x})) \tag{6.9}$$

Pseudo-experiments are performed to determine the frequency of the LLR value for the measured data. In each experiment the data is sampled from a Poisson distribution whose argument is the sum of the signal and background for signal+background hypotheses. Systematic uncertainties are introduced into the calculation by sampling the fractional uncertainty on each background or signal yield for each source of systematic uncertainty. A symmetric Gaussian distribution of the uncertainty is assumed unless asymmetric uncertainties are provided (e.g. JES). Correlated systematics between backgrounds and the signal are commonly sampled (e.g. luminosity) whereas uncorrelated systematics (e.g. $\sigma_{t\bar{t}}$ or QCD normalization) are sampled independently. Confidence levels are defined by integrating the normalized probability distribution of LLR values from the observed LLR value to infinity as shown in Eq. (6.10) and (6.11) for the signal+background and background-only hypothesis, respectively

$$CL_{s+b} = \int_{LLR(s+b|\mathbf{x})}^{\infty} \mathbf{P}(s+b|x')d(LLR(s+b|x') \tag{6.10}$$

$$CL_{b} = \int_{LLR(b|\mathbf{x})}^{\infty} \mathbf{P}(b|x')d(LLR(b|x') \tag{6.11}$$

The purely frequentist confidence level (CL_{s+b}) is known to be unstable if the background model dramatically disagrees with the data. The modified-frequentist approach offers a solution to this feature by defining a new confidence level, CL_s, as the ratio of the confidence levels for the signal+background and background-only hypotheses as shown in Eq. (6.12)

$$CL_s = \frac{CL_{s+b}}{CL_b} \tag{6.12}$$

In the event that no data excess is observed the production cross-section is excluded at 95% confidence level when 1-CL_s = 0.95. Conversely, if an excess is observed in the data, the background-only hypothesis can be excluded at the 3σ and 5σ confidence level when 1-CL_b equals 1-2.7 $\cdot$ 10^{-3} (0.9973) and 1-4.3 $\cdot$ 10^{-7} (0,99999957) respectively. It follows that 3σ implies that the possibility of a false excess identification is one out of $\sim$ 370 and for 5σ is one out of $\sim$ 2,325,581. Thus, 3σ CL is usually taken as a possible indication, but for claiming a discovery, 5σ CL is the standard demand.

6.9.2 HistFitter model-independent signal fit

An analysis searching for new physics phenomena typically sets model-independent upper limits on the number of events beyond the expected number of events in each SR. In this way, for any signal model of interest, anyone can estimate the number of signal events predicted in a particular signal region and check if the model has been excluded by current measurements or not.

Setting the upper limit is accomplished by performing a model-independent signal fit. For this fit strategy, both the CRs and SRs are used, in the same manner as for the model-dependent signal fit. Signal contamination is not allowed in the CRs, but no other assumptions are made for the signal model, also called a "dummy signal" prediction. The SR in this fit configuration is constructed as a single-bin region, since having more bins requires assumptions on the signal spread over these bins. The number of signal events in the signal region is added as a parameter to the fit. Otherwise, the fit proceeds in the same way as the model-dependent signal fit.

6.9.3 Model-independent upper limit

To obtain the 95% CL upper limit on the number of events in a "beyond the Standard Model prediction" for each SR, the fit in the SR proceeds in the same way as the background-only fit, except that the number of events observed in the signal region (evaluated in one bin) is added as an input to the fit. The signal strength parameter is constrained to be non-negative. The statistical test uses the model-independent signal fit configuration.

By normalizing the signal-strength from the fit to the integrated luminosity of the data sample and accounting for the uncertainty on the recorded luminosity, this can be interpreted as the upper limit on the visible cross-section of new physics, σ_{vis}. Here σ_{vis} is defined as the product of acceptance, reconstruction efficiency and production cross-section.

HistFitter includes a script to calculate and present the upper limits on the number of signal events and on the visible cross-section.

As discussed previously, the profile-likelihood based hypothesis tests use the background-level estimates obtained from a background-only fit to both the CRs and SRs (the best estimates available). For consistency, both the observed and expected upper limit (or p-value) determination use the same background-level estimates, such that the expected limit is the most compatible and predictive assessment for the observed limit. As a consequence, the expected upper limit depends indirectly on the observed data.

6.9.4 Fit to data in the control regions

Both control regions (W+jets and $t\bar{t}$) can be used for constraining the two main background sources, as well as the magnitude and correlations of the systematic uncertainties. Data and Monte Carlo predictions are fitted simultaneously in control and signal region under analysis using the model in Eq. (6.13). The control regions are used in order to normalize the main sources of backgrounds to data. The MC predictions for each source of background are used together with their systematic uncertainties. The fit is performed using only the expected data and MC yields (i.e. each control region contributes as one single bin to the fit). The fit procedure is performed with HistFitter package which uses the software packages HistFactory [53] and RooStats [54], which are based on RooFit [55] and ROOT [56,57]. The calculated scale factors with their uncertainties are presented in Table 6.12

Control Region	Scale Factor [SF]
μ_{W+jets}	$1.22^{+0.04}_{-0.04}$
$\mu_{t\bar{t}}$	$0.88^{+0.06}_{-0.06}$

Table 6.12: Scale factors (SF) for each control region with their statistical uncertainties calculated and used by HistFitter package.

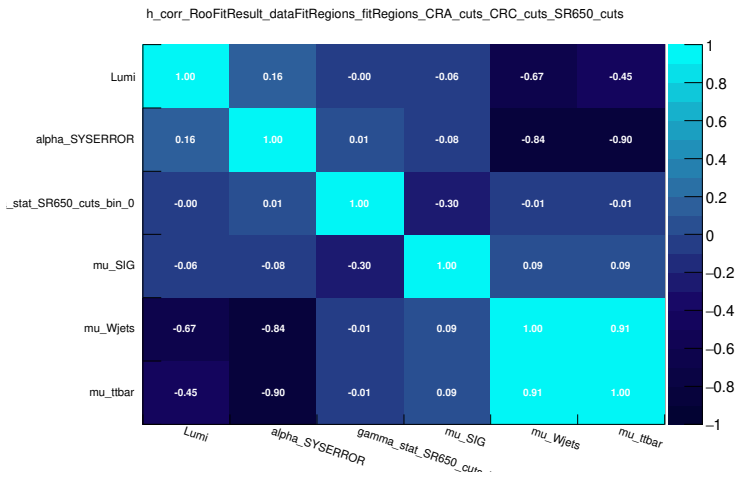


Figure 6.39: Correlation matrix for SR at 650 GeV.

These scale factors are been derived from Eq. (6.13)

$$\mu_p = \frac{N_p(CR, obs.)}{MC_p(CR, raw)} \tag{6.13}$$

where p stands for the process (W+jets or $t\bar{t}$ in our case). N_p(CR,obs.) is the number of observed data in the correspondent CR and MC_p(CR,raw) is the raw and

unnormalized estimated contribution from the process to CR as obtained from MC simulation.

In Fig. 6.39 appear the correlations between the various systematic sources in SR at 650 GeV. The systematic uncertainties are calculated by using both the Negative Log Likelihood (NLL) and a Profile Log Likelihood (PLL). The PLL estimates were used in the analysis.

The yields in the control regions before and after the fit are shown separately for each background component in Table 6.13. The observed events are the real events before any fitting followed by fitted, the real events after the fitting processing. This result is presented separately for each different background process and LQ signal. At the bottom of the table appear the simulated expected events for the same categories as above for comparison. As one can easily ascertain, the main and almost the only difference for all SRs has to do with the LQ signal (the theoretically expected is far larger than the finally observed one as expected).

Event yields	SR8 [650 GeV]	W+jets [CR A]	$t\bar{t}$ [CR C]
Observed events	22	1,205,164	221,094
Fitted events	$22.9^{+77.2}_{-22.9}$	$1,205,174 \pm 1,096$	$221,096 \pm 472$
Fitted W+jets events	15.0 ± 1.1	$746,868 \pm 9,662$	$53,799 \pm 696$
Fitted Z+jets events	2.5 ± 0.2	$157,049 \pm 3,500$	$13,052 \pm 291$
Fitted Diboson events	–	$17,741 \pm 396$	$1,068 \pm 24$
Fitted Single Top events	–	$14,808 \pm 330$	$7,043 \pm 157$
Fitted DY events	0.9 ± 0.1	$2,011 \pm 45$	239 ± 5
Fitted QCD events	2.8 ± 0.2	$245,453 \pm 5,470$	$35,056 \pm 781$
Fitted $t\bar{t}$ events	1.6 ± 0.1	$21,243 \pm 139$	$110,839 \pm 725$
Fitted LQ650 events	0.1	1	3
MC exp. events	54.5	1,205,165	221,103
MC exp. W+jets events	15.0	746,807	53,794
MC exp. Z+jets events	2.5	157,068	13,054
MC exp. Diboson events	–	17,743	1,068
MC exp. Single Top events	–	14,810	7,044
MC exp. DY events	0.9	2,011	240
data-driven exp. QCD events	2.8	245,483	35,060
MC exp. $t\bar{t}$ events	1.6	21,243	110,835
MC exp. LQ650 events	31.7	–	8

Table 6.13: Signal region: Fit results for an integrated luminosity of 20.3 fb^{-1}. The results are obtained from the control regions and signal region at 650 GeV using the exclusion fit. Nominal MC expectations (normalised to MC cross-sections) are given for comparison. The errors shown are the statistical plus systematic uncertainties. Uncertainties on the fitted yields are symmetric by construction, where the negative error is truncated when reaching to zero event yield.

The rest of the signal regions yields tables are presented in App. A.8. Fig. 6.40 concentrates the final event yield for each signal region separately. Each bin represents a signal region and has a width of 50 GeV given the used step from the produced MC LQ signal samples. This histogram makes also obvious the good data

agreement to Standard Model MC backgrounds, as well as that the expected LQ signal is well above both of data and MC. The exclusion limit is approximately equal to 700 GeV. Beyond that, the expected signal becomes more and more unsignificant denoting the sensitivity of this analysis.

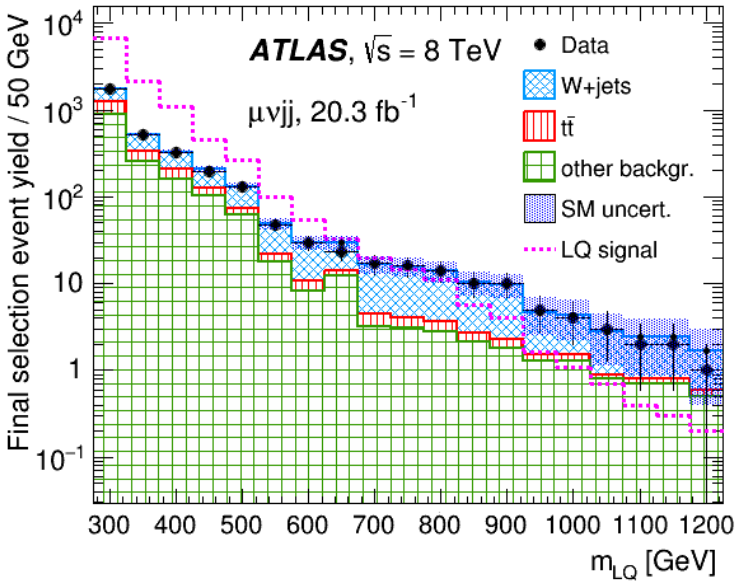


Figure 6.40: Final event yield per signal region with statistical SM uncertainty.

6.10 Final results and conclusions

Following the statistical evaluation, the cross-section limits are presented for the $\beta = 0.5$ scenario (Fig. 6.41 and Fig. 6.42a). There is not any indication of a potential signal excess larger than 2σ. **There has been an exclusion limit improvement up to $\sim$ 690 GeV (corresponding to an NLO $\sigma_{LQ} \simeq$ 8 fb) in comparison to previous analyses in lower energy and luminosities**. There has also been a branching ratio (β) limit improvement in respect to LQ mass appearing in Fig. 6.42b.

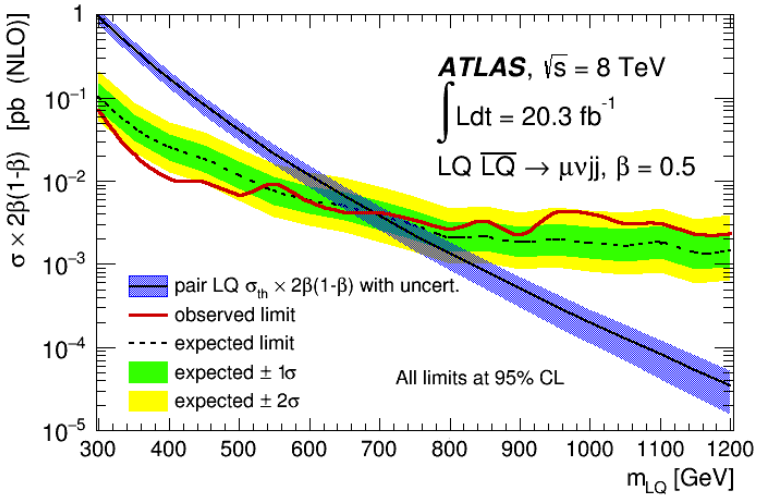
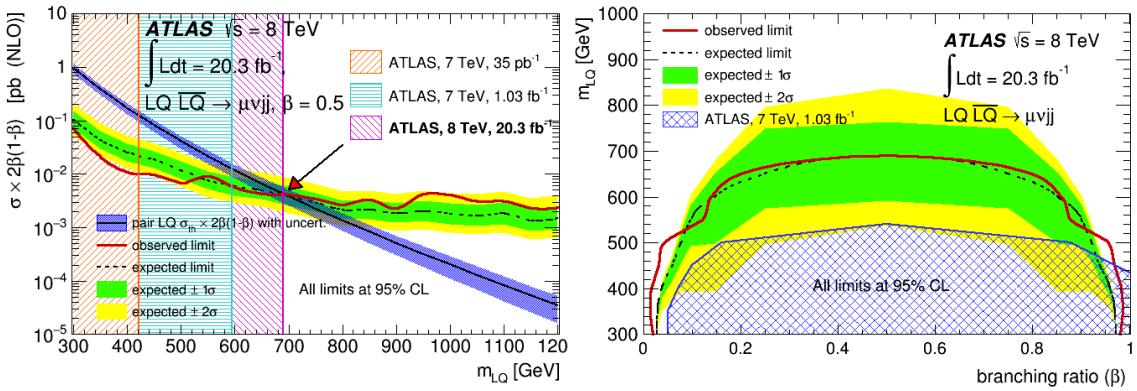


Figure 6.41: Cross-section limits as a function of mass. The $\pm$ 1(2) σ green (yellow) uncertainty bands on the expected limit represents all sources of systematic and statistical uncertainty. The expected NLO production cross-section for scalar leptoquark pair-production and its corresponding theoretical uncertainty due to the choice of PDF set and renormalisation/factorisation scale (blue dashed lines) are also included.



(a) (b)

Figure 6.42: (a) Cross-section limits as a function of mass, results of current and previous analyses; (b) excluded LQ mass as a function of branching ratio β.

Chapter 7

LHC Data LQ Analysis at 13 TeV, Run II

This is the pair LQ analysis of the recorded LHC data at 13TeV (Run 2) with total integrated luminosity of 36.1 fb^{-1} [58].

7.1 Analysis overview

7.1.1 Final state

The studied case is again the second generation scalar LQ decaying like the following

$$LQ\overline{LQ} \rightarrow \ell^- j \; \bar{\nu}_\ell \bar{j}' \quad or \quad \nu_\ell j' \; \ell^+ \bar{j} \tag{7.1}$$

where ℓ is a muon (μ), ν_ℓ a muon neutrino (ν_μ) and j, j' are two second-generation jets originating from a strange or a charm quark. The final state must contain exactly one high-p_T muon, E_T^{miss} corresponding to neutrino and at least two jets ($\mu\nu$jj channel). Scalar LQ cross-sections are in general expected to be significantly smaller than vector LQs (for more information, see App. B.1).

7.1.2 Background processes

The main backgrounds are W+jets and $t\bar{t}$. Considerable contributions come from Z+jets, single top and diboson production. Further small contributions come from $W \rightarrow \tau\nu$ and $Z \rightarrow \tau\tau$ production with additional jets (Sec. 7.4 for details).

QCD background is considered as a part of "fake" background. The reason for using this more general term is just that QCD events are a large part of this background, but not all of it.

7.1.3 Discrimination between signal and background

A major difference between LQ and background events is the presence of jet-lepton pairs coming from the decay of the parent LQ, giving a peak in the reconstructed jet-lepton mass spectrum for the signal. The reconstruction of these masses provides a very important variable used to distinguish between signal and background events.

In this analysis the same reconstructed LQ masses, S_T and m_T variables described in Sec. 6.5.1 are used. However, a cut-based analysis using these variables does not

provide sufficient discrimination power. This observation was made in the 2011 7 TeV LQ search [59,60] and is confirmed in Sec. 7.7. Thus one needs to simultaneously exploit several variables, i.e. to use a multivariate (MVA) technique for the signal discrimination. For this analysis, a Boosted Decision Tree (BDT) as implemented in TMVA [61] is used.

A further means for signal vs background discrimination (in some cases) is the use of (anti-) b-tagging. This has not been used in the past in order to keep the search as general as possible and to not reduce the sensitivity to a signal with b-quarks in the LQ decay. While the benchmark model used for the signal generation assumes coupling to leptons and quarks of the same generation this is not necessarily the case and in fact experimentalists have been encouraged by the theory community to also consider final states with leptons and quarks of different generations. However, a b-jet veto (requirement) is useful in the definition of control regions for the W+jets ($t\bar{t}$) background in the lepton-neutrino channels.

7.1.4 Background estimation and statistical analysis

This analysis makes use of control regions for the main backgrounds, there definitions are given in Sec. 7.1.5. The CRs are defined such that the background process under consideration is enhanced while the signal contamination is kept negligible.

A frequentist statistical analysis of the results will be performed using the HistFitter package [62], a tool based on the RooStats framework. Two fits are performed: first a fit including only the control regions to obtain normalisation factors of the corresponding backgrounds and then an exclusion fit including only the signal regions to extract the signal strength. A profile likelihood is constructed simultaneously from all regions considered in a given fit taking the statistical and systematic uncertainties into account as nuisance parameters. The remaining background contributions are small and are estimated purely based on simulation.

In case an excess is observed, the statistical tools can be used to quantify that excess. In the absence of a significant observed excess, exclusion limits are set on the new physics model parameter of interest. Details on the statistical analysis procedure will be given in Sec. 7.11.

7.1.5 Regions of phasespace

The exact definitions of muons, jets and E_T^{miss} together with the corresponding kinematic cuts are described in Sec. 7.3. However, at this stage already it is useful to present regions of phase space the analysis makes use of. They are in part inherited from previous analyses (LQ searches at 13 and 7 TeV), but some alterations have been made in view of the usage of a BDT.

The W+jets CR (W CR) is defined by restricting m_T to a window around the W-peak and requiring the absence of b-jets in the event. To suppress the fake background, the E_T^{miss}-significance, calculated as $E_T^{miss}/\sqrt{p_T^{j_1}+p_T^{j_2}+p_T^{\ell_1}}$ is required to be greater than 4.

The $t\bar{t}$ CR is defined by the same cut on m_T as for the W CR, but requiring at least two b-tagged jets.

The signal region is given by $m_T > 130$ GeV, $E_T^{miss} > 150$ GeV and E_T^{miss} significance greater than 3, in order to reduce the contribution from fake backgrounds.

A single BDT-bin, optimised per mass point, defines the signal region. Table 7.1 concentrates control and signal regions.

W CR	$t\bar{t}$ CR	SR
40 GeV $<$ m_T $<$ 130 GeV no b-jets E_T^{miss} significance $>$ 4	40 GeV $<$ m_T $<$ 130 GeV $\geq$ 2 b-jets	$m_T > 130$ GeV $E_T^{miss} > 150$ GeV E_T^{miss} significance > 3 1 BDT bin

Table 7.1: Regions of phasespace used in the LQ search.

7.1.6 Variables, notation

This section explains the meaning of some important variables used in the present LQ search. The lepton-jet and E_T^{miss}-jet pairing is such that the absolute value of the m_{LQ} and m_{LQ}^T is minimised. Single objects are denoted by letters (j_1, j_2, ℓ_1, ℓ_2, μ for leading jet, subleading jet, leading lepton, subleading lepton and muon respectively).

- E_T^{miss} significance $= \frac{E_T^{miss}}{\sqrt{p_T^{j_1}+p_T^{j_2}+p_T^{\ell_1}}}$,
- $m_{LQ}^{diff} = |\ m_{LQ}\text{-}m_{LQ}^T\ |$,
- $m_{LQ}^{relDiff} = \frac{|m_{LQ}-m_{LQ}^T|}{|m_{LQ}+m_{LQ}^T|}$,
- $\Delta\phi(j,j) = \Delta\phi$ between leading and subleading jet,
- $\Delta\eta(j,j) = \Delta\eta$ between leading and subleading jet,
- $\Delta\phi(\ell,\ell) = \Delta\phi$ between leading and subleading lepton,
- $\Delta\eta(\ell,\ell) = \Delta\eta$ between leading and subleading lepton,
- m_{jj} = two leading jets invariant mass,
- $\min(|\Delta\phi(j_1,\ell_1)|, |\Delta\phi(j_1,\ell_2)|)$ = minimum $\Delta\phi$ between leading jet and a lepton,
- $\min(\Delta R(j_1,\ell_1), \Delta R(j_1,\ell_2))$ = minimum ΔR between leading jet and a lepton,
- $\min(|\Delta\phi(j_2,\ell_1)|, |\Delta\phi(j_2,\ell_2)|)$ = minimum $\Delta\phi$ between subleading jet and a lepton,
- $min(\Delta R(j_2,\ell_1), \Delta R(j_2,\ell_2))$ = minimum ΔR between subleading jet and a lepton,
- $\Delta\phi_{LQ_1,LQ_2}, \Delta\eta_{LQ_1,LQ_2}, \Delta R_{LQ_1,LQ_2} = \Delta\phi, \Delta\eta, \Delta R$ between the two LQ candidates.

The rest of the variables are identical to that of the previous LQ analysis at 8 TeV in Sec. 6.5.1.

7.2 Data and MC samples

7.2.1 Data

The present analysis uses pp collision data collected during the period of 2015 and 2016 at $\sqrt{s} = 13$ TeV with 25 ns bunch spacing. In 2015, data were recorded in periods D3-J6, runs 276262-284484. In 2016, data were recorded in periods A3-L11, runs 297730-311481. The integrated luminosity of the 2015 and 2016 data is 3.2 fb^{-1} and 32.9 fb^{-1} respectively.

Data is reconstructed by the ATLAS Athena framework [63] release 20.7. The analysis uses EXOT9.

7.2.2 MC samples

Signal and most of the background event yields are estimated with the use of MC [64]. The ATLAS detector response was simulated with Geant4 [65] and all MC events were reconstructed in the same way as data.

7.2.2.1 Signal MC samples

The NLO MadGraph model used is described in [66]. The version used in the ATLAS software is LQnomix_NLO, to be found in MadGraphModels. The simulation is done with MadGraph5_aMC@NLO version 2.4.3 in MCProd 19.2.5.14.3, using MadSpin for the decay of LQs, top-quarks and W-bosons and showering the events with Pythia8.212 with the A14 [68] tune for NNPDF23LO. The NNPDF 3.0 NNLO [67] Parton Distribution Function (PDF) set is used.

Since this is the first time this MG model is used in ATLAS, as a cross check two LQ mass points (1 and 1.5 TeV) were generated with Pythia8 [69] at LO (Sec. 7.2.2.4).

7.2.2.2 MadGraph and Pythia signal samples validation

MadGraph generator has the ability to produce particles and anti-particles at the same time in each different sample. More precisely, the final products of the LQ decays are μ^-, ν_μ, c-quark, s-quark and μ^+, $\overline{\nu}_\mu$, $\overline{c}$-quark, $\overline{s}$-quark as well. Of course, the pairing is done according to theory which predicts an LQ decaying to particles (μ^- or ν_μ and a quark) and an $\overline{LQ}$ decaying to anti-particles (μ^+ or $\overline{\nu}_\mu$ and an $\overline{q}$) respectively. This is a major difference in comparison to Pythia generator which allows only one combination in all the events. That means that in Pythia signal samples, an LQ decays strictly to a μ^- and a quark and an $\overline{LQ}$ to an μ^+ and an $\overline{q}$. Table 7.2 summarizes the possible compination for each generator.

	Pythia	MadGraph	
LQ	μ^-, c	μ^-, c	ν_μ, s
$\overline{LQ}$	$\overline{\nu}_\mu$, $\overline{s}$	$\overline{\nu}_\mu$, $\overline{s}$	μ^+, $\overline{c}$

Table 7.2: LQ possible combinations in the same MC sample per signal generator.

Fig. 7.1 depicts the reconstructed truth m_{LQ_1} and m_{LQ_1} and truth m_{LQ_2} and m_{LQ_2} respectively for both MadGraph combinations for three mass points, 0.5, 1.0 and 1.5 TeV. Each sample consists of $\sim$ 1,000 events, $\sim$ 500 events per combination. Truth LQ masses consist of a lepton and a quark. The distributions for both combinations are almost identical.

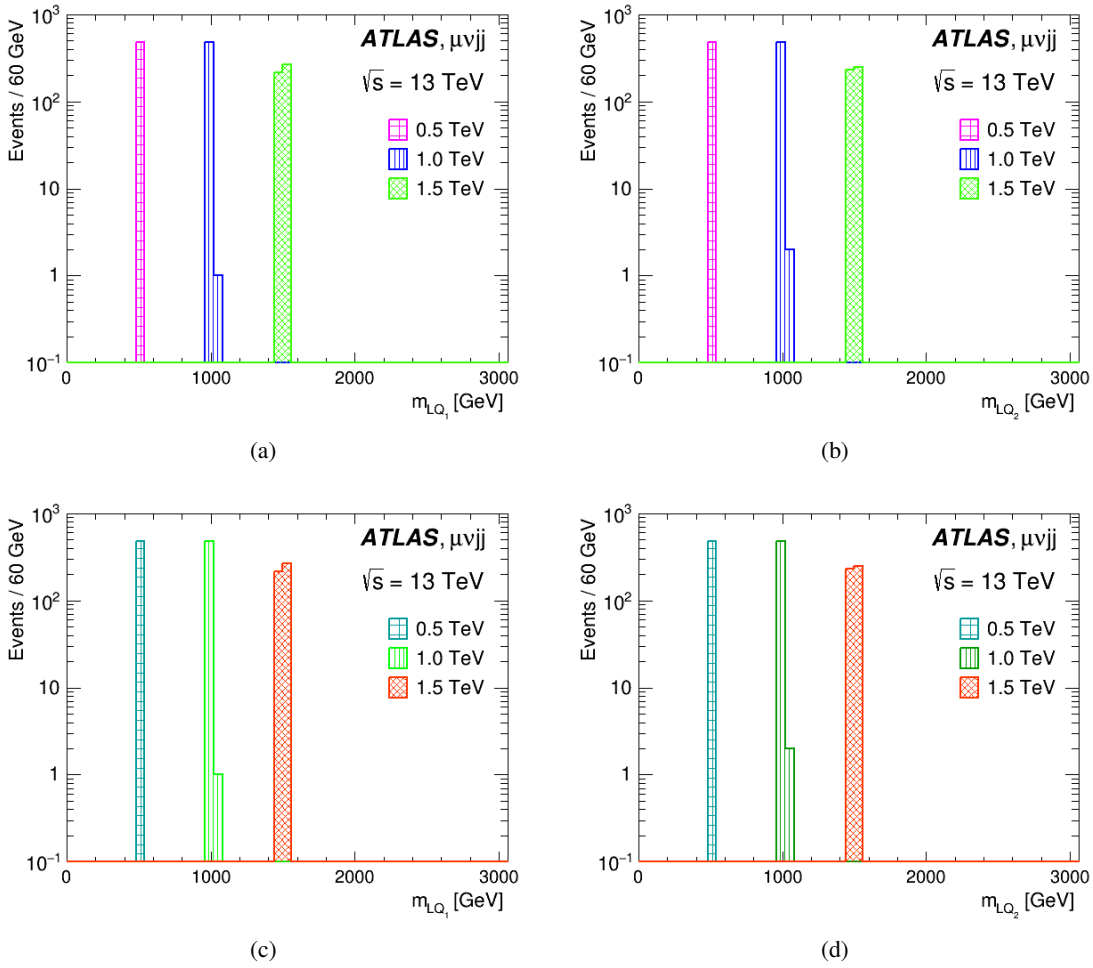


Figure 7.1: MadGraph samples: LQ decaying to μ^- and c (a), $\overline{LQ}$ decaying to $\overline{\nu}_\mu$ and $\overline{s}$ (b), LQ decaying to ν_μ and s (c), $\overline{LQ}$ decaying to μ^+ and $\overline{c}$ (d).

The width of the LQ is calculated using three parameters: the coupling λ, the branching ratio β and m_{LQ}. For the decay involving a charged lepton ($\beta = 1$) the partial width is given by

$$\Gamma_\ell = \frac{m_{LQ}^4 \lambda^2 \beta}{16\pi m_{LQ}^3} = \lambda^2 \beta \frac{m_{LQ}}{16\pi} \tag{7.2}$$

For a decay involving a neutrino ($\beta = 0.5$) the formula is the following

$$\Gamma_\nu = \frac{m_{LQ}^4 \lambda^2 (1-\beta)}{16\pi m_{LQ}^3} = \lambda^2 (1-\beta) \frac{m_{LQ}}{16\pi} \tag{7.3}$$

and the total width for the pair LQ decay is given by

$$\Gamma_\ell + \Gamma_\nu = \lambda^2 \frac{m_{LQ}}{16\pi} \sim 0.0018 \cdot m_{LQ} \tag{7.4}$$

and thus, for a nominal m_{LQ} of 1000 GeV, the LQ decay width would be equal to $\sim$ 1.8 GeV.

In Fig. 7.2 appear the reconstructed m_{LQ_1} and m_{LQ_2} consisting of a lepton and a jet, again for the same two combinations and for the same three mass points. In this case, the selected jet is the one closest to the truth quark coming directly from the LQ. This way the jet is been identified as the one coming from the right quark.

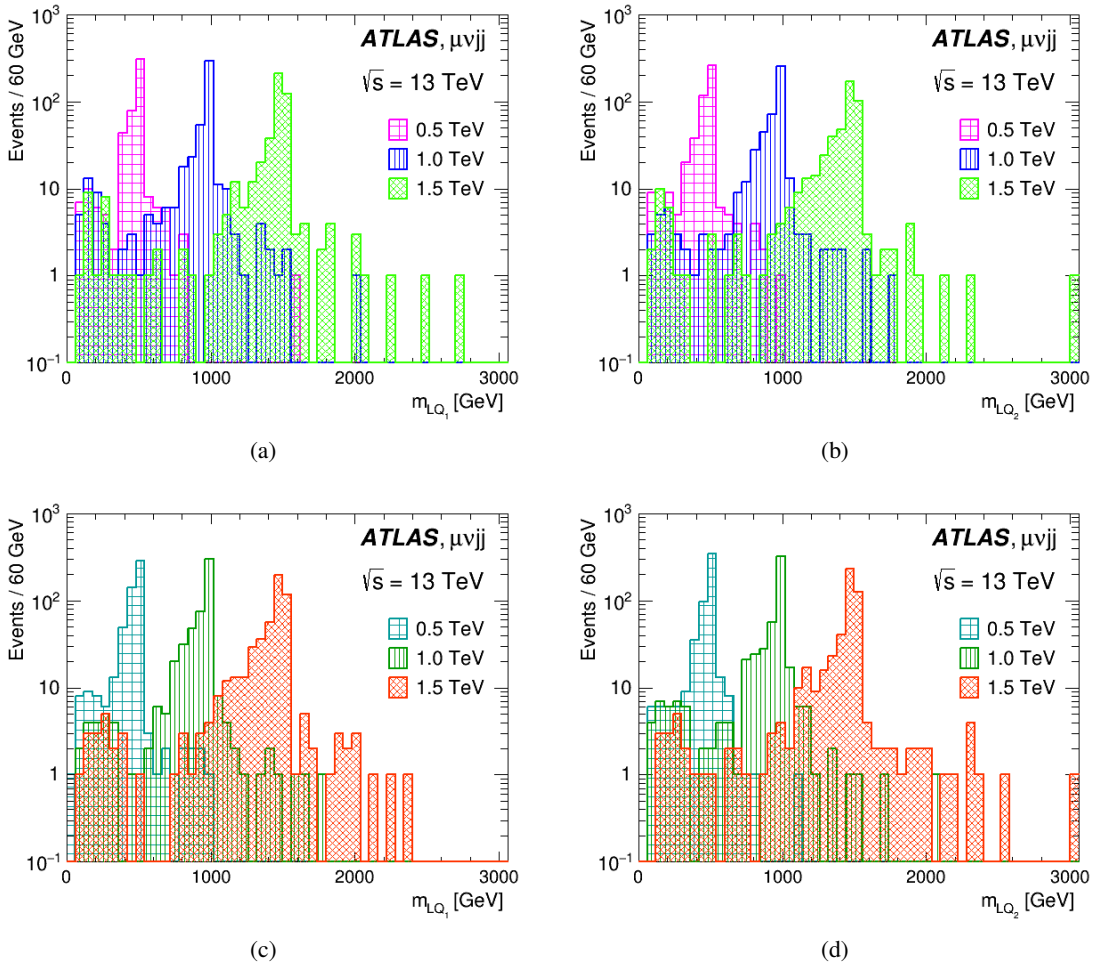


Figure 7.2: MadGraph samples: LQ decaying to μ^- and c-related jet (a), $\overline{LQ}$ decaying to $\overline{\nu}_\mu$ and $\overline{s}$-related jet (b), LQ decaying to ν_μ and s-related jet (c), $\overline{LQ}$ decaying to μ^+ and $\overline{c}$-related jet (d).

In Fig. 7.3 appear the reconstructed m_{LQ_1} and m_{LQ_2} consisting of a lepton and a truth quark or a lepton and a jet produced with Pythia8 generator. In this case, there are only two mass points, 1.0 TeV and 1.5 TeV. The rest of the validation plots are presented in App. B.2, App. B.3 and App. B.3.1.

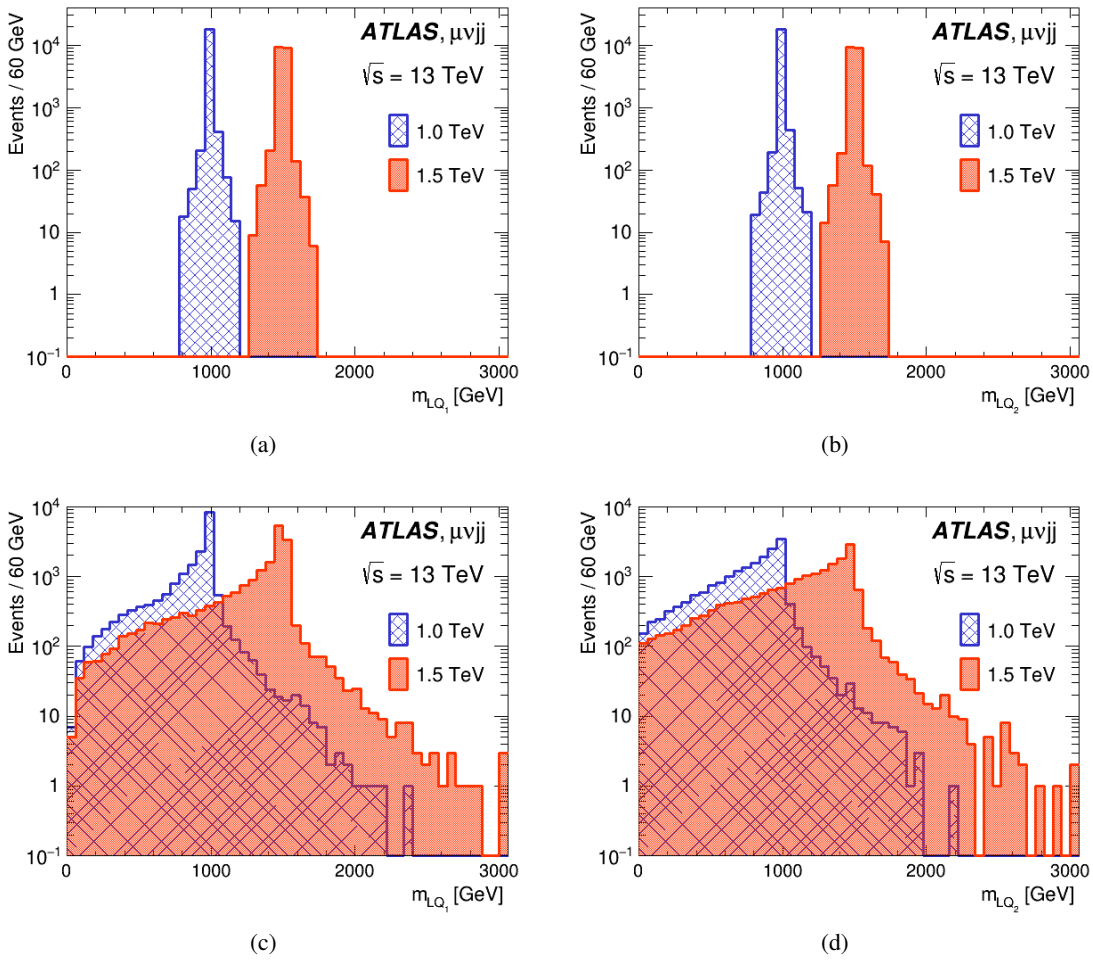


Figure 7.3: Pythia8 samples: (a) LQ decaying to μ^- and c, (b) $\overline{LQ}$ decaying to $\overline{\nu}_\mu$ and $\overline{s}$, (c) LQ decaying to μ^- and c-related jet, (d) $\overline{LQ}$ decaying to $\overline{\nu}_\mu$ and $\overline{s}$-related jet.

7.2.2.3 Signal samples information

The final MadGraph signal samples used in the analysis are produced with a mass step of 50 GeV ranging from m_{LQ} = 200 GeV up to m_{LQ} = 1500 GeV and a mass step of 100 GeV ranging up to m_{LQ} = 2200 GeV. The reasoning is that at low masses the cross-section falls more steeply and that the region just above the current limit around m_{LQ} = 1000 GeV is highly interesting. The mass region beyond m_{LQ} = 1500 GeV will likely be difficult to probe with this data set and therefore the step size can be larger. The request therefore consists of 34 mass points per generation, with 150,000 events each.

7.2.2.4 Comparison to Pythia8.165 samples

As mentioned previously, Pythia8 samples have been generated as well for both generations, for LQ masses of 1 TeV and 1.5 TeV. The cross-sections given in this table are the NLO cross-sections calculated with a dedicated tool [70] that uses the CTEQ6L1 PDF. These cross-sections are higher than those calculated by MadGraph

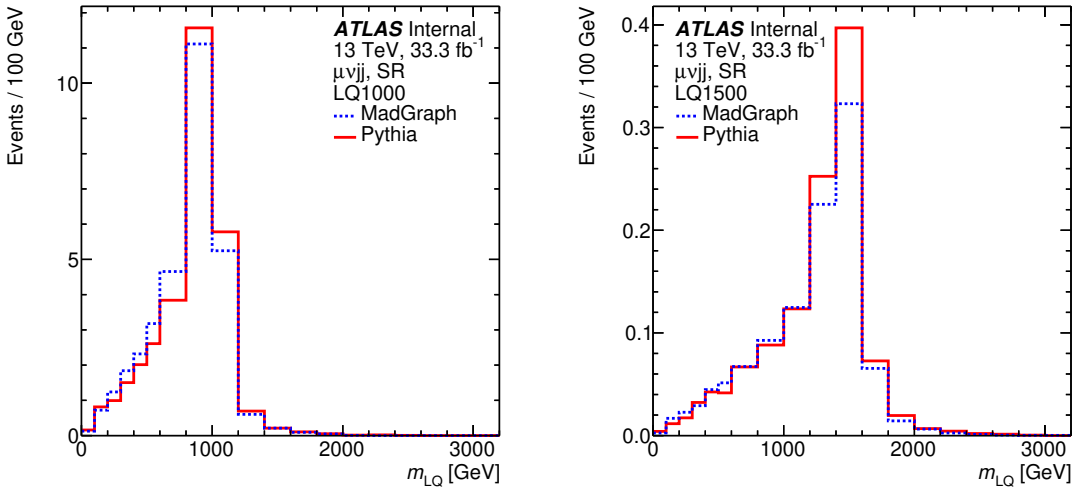
	MadGraph		Pythia8	
mass [GeV]	DSID	σ [pb]	DSID	σ [pb]
200	306742	58.321		
250	306743	19.869		
300	306744	7.8754		
350	306745	3.5078		
400	306746	1.7087		
450	306747	0.88173		
500	306748	0.48123		
550	306749	0.27533		
600	306750	0.163		
650	306751	0.099098		
700	306752	0.061852		
750	306753	0.039378		
800	306754	0.025775		
850	306755	0.016902		
900	306756	0.011413		
950	306757	0.0077933		
1000	306758	0.0053518	305820	0.00586
1050	306759	0.0037112		
1100	306760	0.0026188		
1150	306761	0.0018387		
1200	306762	0.0013137		
1250	306763	0.00093751		
1300	306764	0.00067825		
1350	306765	0.00049102		
1400	306766	0.00035888		
1450	306767	0.00026404		
1500	306768	0.00019359	305821	0.000240
1600	306769	0.00010677		
1700	306770	$5.9234 \cdot 10^{-5}$		
1800	306771	$3.3154 \cdot 10^{-5}$		
1900	306772	$1.8745 \cdot 10^{-5}$		
2000	306773	$1.0687 \cdot 10^{-5}$		
2100	306774	$6.1567 \cdot 10^{-6}$		
2200	306775	$3.5434 \cdot 10^{-6}$		

Table 7.3: Signal MadGraph and Pythia samples and their cross-sections.

(using NNPDF3.0).

As a sanity check, relevant distributions for the MadGraph and Pythia samples are compared (at reconstruction level) for these two mass points. Some differences are expected, among other things because of the LO vs NLO generation (Table 7.3). However, there should not be any dramatic differences.

Fig. 7.4 presents the m_{LQ} normalized distribution comparison between Pythia8 and MadGraph generators for 1.0 TeV and 1.5 TeV. The rest of comparison plots are presented in App. B.4.



(a) (b)

Figure 7.4: Comparison of m_{LQ} distribution for the MadGraph (blue) and Pythia8 (red) signal simulation of an LQ of (a) 1 TeV and (b) 1.5 TeV mass.

7.2.2.5 Background processes

This section lists MC samples for the following processes: Z/γ^*+jets, W+jets, $t\bar{t}$, single top, diboson.

Events containing W or Z bosons with associated jets are simulated [71] using the Sherpa 2.2.1 generator. Matrix elements are calculated for up to 2 partons at NLO and 4 partons at LO using the Comix [72] and OpenLoops [73] matrix element generators and merged with the Sherpa parton shower [74] using the ME+PS@NLO prescription [75]. The CT10nlo PDF set is used in conjunction with dedicated parton shower tuning developed by the Sherpa authors. The NNPDF 3.0 NNLO PDF set is used for the hard scattering process. To ensure sufficient statistics in all of the phase space to be explored in the analysis, the generation is split into bins of (truth-level) $\max(H_T, p_T^V)$. It is also filtered for quark flavours with the following three filters: c-veto + b-veto, c-filter + b-veto, b-filter.

For the generation of $t\bar{t}$ and single top-quarks in the Wt and s-channel the Powheg-Box v2 [76–78] generator is used. The single-top samples use the CT10 [44] PDF set in the matrix element calculations, the $t\bar{t}$ samples use NNPDF3.0. Electroweak t-channel single top-quark events are generated using the Powheg-Box v1

generator. This generator uses the 4-flavour scheme for the NLO matrix elements calculations together with the fixed 4-flavour PDF set CT10f4. For all top processes, top-quark spin correlations are preserved (for t-channel, top quarks are decayed using MadSpin [79]).

For the single-top samples, the parton shower, fragmentation, and the underlying event are simulated using Pythia6.428 [80] with the CTEQ6L1 [81] PDF sets and the corresponding Perugia 2012 tune (P2012) [82]. The top mass is set to 172.5 GeV. The EvtGen v1.2.0 program [83] is used for properties of the bottom and charm hadron decays. The hdamp parameter, which controls the p_T of the first additional emission beyond the Born configuration, is set to the mass of the top quark. The main effect of this is to regulate the high-p_T emission against which the $t\bar{t}$ system recoils.

For the $t\bar{t}$ samples, Pythia8 is used instead of Pythia6, with the A14 tune and the NNPDF2.3 LO PDF set.

Diboson processes with 4 charged leptons, 3 charged leptons+1 neutrino or 2 charged leptons and 2 neutrinos are simulated using the Sherpa 2.1.1 generator. Matrix elements contain all diagrams with four electroweak vertices. They are calculated for up to 1 (4ℓ, 2ℓ+2ν) or 0 partons (3ℓ+1ν) at NLO and up to 3 partons at LO using the Comix and OpenLoops matrix element generators and merged with the Sherpa parton shower using the ME+PS@NLO prescription. The CT10 PDF set is used in conjunction with dedicated parton shower tuning developed by the Sherpa authors. The generator cross-sections are used in this case (already at NLO). The Sherpa diboson sample cross-section has been scaled down to account for its use of alphaQED=1/129 rather than 1/132 corresponding to the use of current PDG parameters as input to the G_μ scheme.

7.2.2.6 Pile-up reweighting

In the presented search, the chosen variable is the average expected number of pp collisions in a bunch-crossing, $\langle\mu\rangle$. The weights are retrieved with the use of an official tool, the PileupReweighting tool [84].

7.3 Event selections and object definitions

This section provides an overview of muon and jet selection criteria together with the whole event selection chain. Each event must contain exactly one muon and no electrons. Below all the steps are presented in detail.

7.3.1 Basic event selection

The following is a very basic event selection focused mainly on the event quality and the triggers used for data during 2015 and 2016 running periods.

- event must be in a run of GRL
- veto LAr noise burst events, incomplete events and data corruption at the event level are vetoed
- HLT_mu26_imedium || HLT_mu50 for 2015 data
- HLT_mu26_ivarmedium || HLT_mu50 for 2016 data

- remove duplicated events

7.3.1.1 Data quality

The collected data are divided into Luminosity Blocks (LBs), which consist of one minute of approximately constant instantaneous luminosity and data taking conditions, such as detector status and the trigger menu. Only LBs where the LHC has declared "stable beams" and having components of the detector declared to be operating as expected have been used. Events are required to belong to an LB with good data quality for tracking, calorimetry and muon spectrometer, as well as to the relevant GRLs of 2015 and 2016 respectively.

7.3.1.2 Data corruption

Events with corrupted data due to LAr, Tile and SCT are rejected. Events with LAr noise bursts are rejected as well. Incomplete events are also rejected. These can appear in the LB taking place just after a TTC restart.

7.3.1.3 Vertices

In an event, there are tens of vertices originating in pp collisions. The primary vertex of interest satisfies the following criteria:

- has the largest $\sum p_T^2$ of associated tracks among all vertices,
- has at least two associated tracks with $p_T > 0.4$ GeV.

7.3.2 Muon selection

The muon definition consists of three steps: muon reconstruction, identification and isolation [85,86]. Only combined muons from the region $|\eta| < 2.5$ are used. The p_T cut ($p_T > 65$ GeV) is driven by the trigger used.

The presented analysis uses the Medium identification WP which is the default selection for muons in ATLAS. The efficiency of this WP is more than 98% for muons with $p_T > 65$ GeV.

Additional background suppression can be achieved by requiring muons to be isolated. The presented analysis uses the LooseTrackOnly isolation WP.

To make sure that the muon cadidates originate in the hardest primary vertex, cuts are applied on the transverse (d_0) and longitudinal (z_0) impact parameters. Namely, the significance of d_0, d_0^{sig}, is required to be smaller than 3 mm and $|\Delta z_0 sin\theta|$ to be smaller than 0.5 mm. Table 7.4 summarizes muon selection criteria.

7.3.3 Jet selection

Jets used in the presented analysis are reconstructed from topological clusters (topoclusters) [87–89] with the anti-k_T algorithm [90] with a distance parameter of $R = 0.4$. The jet calibration procedure substantially relies on the 2012 approach [91]. (But here EM jets are used rather than LC jets.)

For baseline jets the only requirement that is applied is a cut of $p_T > 20$ GeV. The final jet object definition requires $p_T > 60$ GeV to reduce the probability of

selecting pile-up jets and $|\eta| < 2.5$ as b-tagging information will be used and is only available in the region covered by the tracker.

7.3.3.1 Jet cleaning

The presented analysis uses the Loose jet cleaning criterium. The jet cleaning is performed after the overlap removal (Sec. 7.3.5) on jets that have a $p_T > 20$ GeV. In general, an event is rejected if it contains any such jet that does not pass the loose cleaning criteria. An exception to that are jets with 20 GeV $< p_T <$ 60 GeV and $|\eta| < 2.48$: if such a jet is marked as bad, the event is only rejected if this jet is not marked as a pile-up jet by the JetVertexTagger9, i.e. it passes the JVT cut. The recommended cut value is 0.59 and a jet passes the cut (i.e. is not considered pile-up) if its jvt attribute is greater than this threshold.

7.3.3.2 B-tagging

The analysis makes use of b-tagging in the definition of control regions. B-tagging can be a means of increasing the signal vs background discrimination for the benchmark signal model. The effect of b-tagging in different phase space regions has been studied and the results are summarised here. It is performed using the BTagging-Tool. Scale factors to adjust the efficiency in simulation to that observed in data are obtained form the BTaggingEfficiencyTool. The tagger used is MV2c10, corresponding to a c-quark fraction in the training sample of 7% [92]. The 77% efficiency, fixed cut working point is used. The b-quark origin is checked for all jets passing the final selection criteria described previously.

B-tagging effect can be considerable in the lepton-neutrino channel. It is especially helpful to increase the purity of the W+jets and $t\bar{t}$ CRs, as will be illustrated later.

Fig. 7.5 shows the ϕ-distribution of the leading muon in the W CR (i.e. for events with 40 GeV $< m_T <$ 130 GeV) for different requirements on the number of b-tagged jets. Fig. 7.5a shows the distribution without any requirement on b-tagging. A control region defined in this way would suffer from a large contamination of $t\bar{t}$ events. This contamination can be reduced very effectively by requiring that none of the jets is b-tagged, as can be seen from Fig. 7.5b. On the other hand, requiring explicitly that b-tagged jets be present in the event selects $t\bar{t}$ events with high purity and can be used to define a CR for $t\bar{t}$. Fig. 7.5c shows the distribution for events with at least one b-tagged jet and it is seen that the $t\bar{t}$ fraction is significantly increased. When requiring at least two b-jets the event sample consists almost entirely of $t\bar{t}$

Criterion	Cut
Type	combined
p_T	> 65 GeV
$\|\eta\|$	< 2.5
Impact parameter	$d_0^{sig} < 3$ mm, $\|\Delta z_0 sin\theta\| < 0.5$ mm
Quality	Medium
Isolation	LooseTrackOnly

Table 7.4: Muon object selections used in the LQ search.

events, cf. Fig. 7.5d.

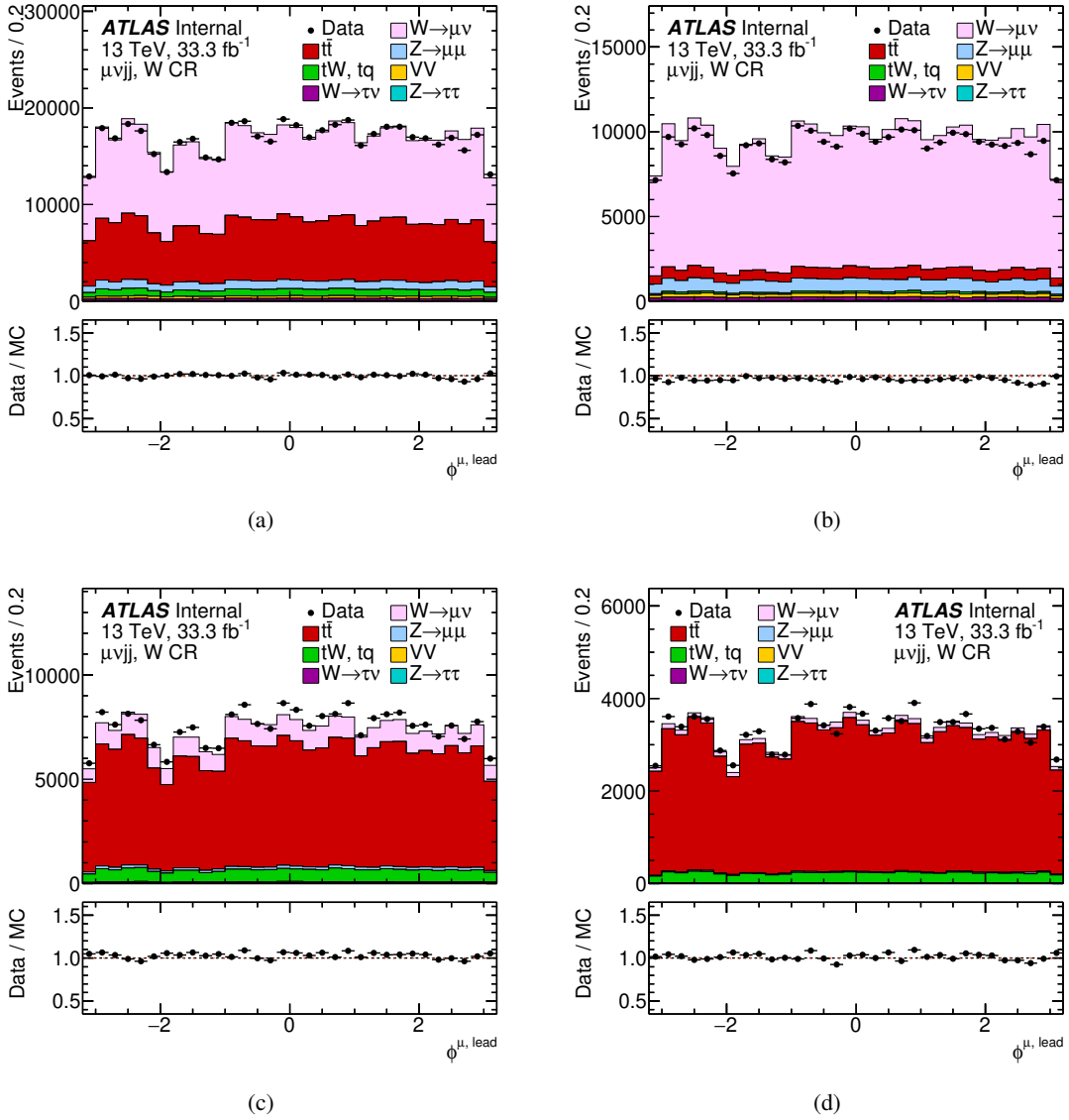


Figure 7.5: Distribution of the azimuthal angle ϕ of the leading muon in the W CR for different selections on the number of b-tagged jets: (a) no requirement on b-tagging, (b) no b-tagged jet, (c) at least one b-tagged jet, (d) at least two b-tagged jets.

7.3.4 Missing transverse energy

The x and y components of MET [93, 94] are calculated according to:

$$E^{miss}_{x(y)} = E^{miss,e}_{x(y)} + E^{miss,\gamma}_{x(y)} + E^{miss,\tau}_{x(y)} + E^{miss,jets}_{x(y)} + E^{miss,\mu}_{x(y)} + E^{miss,soft}_{x(y)} \tag{7.5}$$

where each object term is given by the negative vectorial sum of the momenta of the respective calibrated objects. In the present analysis, the terms corresponding to electrons, muons and jets (after the baseline selection) are considered whereas

the ones that correspond to photons and taus are not built explicitly (they enter via $E_{x(y)}^{miss,soft}$). The soft term is calculated with the use of tracks reconstructed by the Inner Detector. For its calculation, the following tracks are rejected:

- Tracks within ΔR(track, electron/photoncluster) < 0.05
- Tracks within ΔR(track, τ) < 0.2
- Tracks associated with jets using the ghost-association technique [95, 96] are removed.

In addition, ID tracks associated to combined or segment-tagged muons are replaced with the combined ID+MS fit.

No overlap removal is applied before passing the containers to the tool, as it has its own overlap removal implemented. The tool is configured to use the default JVT cut (0.59) for the jet term.

7.3.5 Overlap removal

Overlaps between objects after the baseline selection (and after the E_T^{miss} calculation) are treated in the following way, according to the outcome of the harmonisation effort as documented in [97] and implemented in a common tool [98].

The tool is configured in a standard way in which information about the jet flavour is not taken into account (bJetLabel=" "). For the jet-lepton overlap removal, the "boostedLeptons" option is applied, i.e. a sliding ΔR cone is used, given by ΔR slide = min(0.4, 0.04 + $^{10}/_{p_T[GeV]}$), where the p_T is that of the lepton.

First, electrons are removed if they share a track with a muon. Then, overlaps between electrons and jets are removed in a two-step procedure: jets which are within $\Delta R < 0.2$ of an electron are removed, then electrons are removed if they are within ΔR slide of one of the remaining jets. Finally, muon-jet overlaps are resolved as follows: if a muon and a jet are closer than ΔR slide, the jet is rejected if it has less than 3 tracks, otherwise the muon is rejected.

For the definition of signal and control regions only objects that survive the overlap removal are considered.

7.4 Background estimation

7.4.1 V+jets background

The estimation of the contributions from V+jets processes in the signal regions are estimated in a semi-data-driven way. The simulations are constrained with data in control regions that are orthogonal to the signal regions. These control regions are defined such that the purity in the background process of interest is enhanced while a potential signal contamination is reduced to a negligible level.

They are used to extract the normalisation for the respective background, as well as to correct shape mis-modelling of the data by the simulation. The Sherpa 2.2.1 V+jets samples do not describe variables related to the jet activity in the event very well. This is a known and understood issue related to the choice of the renormalisation scale. It has been found that a reweighting in the invariant di-jet mass, m_{jj}, mitigates the mismodelling to a very high degree. The shape corrections

are derived from the data-MC-ratio in the control region prior to the fit. Afterwards, fits to single-bin control regions and signal regions are performed, details of which are described later.

The W CR uses events with 40 GeV $< m_T <$ 130 GeV in which there are no b-tagged jets. In addition, the E_T^{miss} is required to exceed 40 GeV and its significance has to be greater than 4.

7.4.2 $t\bar{t}$ background

The background contribution from $t\bar{t}$ production is estimated in a similar way as the V+jets backgrounds. Control regions orthogonal to the signal region are defined and a normalisation factor is extracted by performing a fit to all control regions.

The $t\bar{t}$ CR IS defined by the same cut on m_T as the W CR, except that instead of vetoing events with b-jets, at least two b-jets are required.

7.4.3 Fake background

Background contributions resulting from the mis-identification of other objects as leptons or from non-prompt leptons are difficult to model in simulations and are therefore estimated in a data driven way. Examples of fake signatures are (light-flavour) jets with a large fraction of energy in the electromagnetic calorimeter such that they are reconstructed as leptons, or actual leptons from the decay of hadrons inside (heavy-flavour) jets. The fake probability is much higher for electrons than for muons due to the similarities in the signatures of electrons and jets or photons. As in previous version of this analysis (and in similar analyses) the fake background contribution is assumed to be negligible in the muon channels.

7.5 Systematic uncertainties

This section presents sources of systematic uncertainty considered in the LQ search. The uncertainties are divided into three categories: experimental uncertainties and theoretical uncertainties on the background and the signal respectively.

7.5.1 Experimental uncertainties

The experimental systematic uncertainties are summarised below. The object related uncertainties are evaluated with the use of official ATLAS performance groups packages as collected in the AnalysisBase release 2.4.29. These per-object uncertainties are all propagated from the object level to the final spectra.

7.5.1.1 Luminosity uncertainty

This uncertainty is applied in the fit as an uncertainty in the MC normalisation to account for uncertainties in the ATLAS luminosity measurement. The (preliminary) uncertainty in the combined 2015+2016 integrated luminosity is 2.1%.

7.5.1.2 Pile-up reweighting uncertainty

An uncertainty due to the MC reweighting of the h i distribution is included. It is estimated by varying the DataScaleFactor from 1.0/1.09, which is the default, to 1.0 and 1.0/1.18, respectively.

7.5.1.3 Muon-related uncertainties

Seven sources of experimental uncertainties related to muon reconstruction are considered in this analysis. All uncertainties are provided by the Muon Combined Performance (MCP) group. They are uncertainties in the determination of the MS momentum scale, MS momentum resolution, ID momentum resolution and uncertainties in the determination of the four efficiency scale factors-trigger, identification, isolation and track-to-vertex association (TTVA).

7.5.1.4 Jet-related uncertainties

There are two main sources of uncertainty related to the jet reconstruction on ATLAS: jet energy scale and resolution. Both are provided by the JetEtMiss group as recommendations for the full dataset.

Another source of uncertainty related to jets is the one in the scale factor of the b-tagging efficiency, provided by the b-tagging calibration group. It is taken into account in regions where events with b-tagged jets are explicitly selected. Similarly, an uncertainty in the SF of the b-tagging inefficiency is taken into account where a veto on b-tagged jets is applied.

An additional small source of uncertainty is the efficiency determination of the jet cleaning algorithm. This uncertainty is not considered at the moment.

JES Uncertainty The uncertainty due to the JES is the largest experimental uncertainty in the search. It is mostly derived from various in-situ techniques. By default, it is split into ~ 70 components in order to keep track of correlations among various effects affecting the JES. In principle, one should propagate all the components through the analysis separately. This would yield the most correct treatment of their correlations that is available. However, it might become too complicated to handle all these components. Therefore, reduced sets of the JES uncertainty components are derived by the JetEtMiss. Besides others, JetEtMiss provides a set with four components only. From the technical point of view, it is desirable to use it. However, reduced sets do not contain complete information on the correlations. In order to estimate the sensitivity of the analysis to the jet-jet correlations, three additional sets are provided by the JetEtMiss group. Each of those sets contains four components only. The sets are built with different jet-jet correlation assumptions in such a way that they probe the sensitivity of an analysis to jet correlations. The loss of correlations is indeed found to have no impact on the uncertainty.

JER Uncertainty JER uncertainty is split into about ten components. In parallel to this full set, JetEtMiss provides a set containing just one component of the JER uncertainty. It is just a quadrature sum of the uncertainty components contained in the full set. It should be applied to MC only and in just one direction (up). The presented analysis uses this one-component set.

7.5.1.5 E_T^{miss} uncertainty

The uncertainties due to electron, muon and jet energy scale and resolution are propagated to E_T^{miss} by recomputing E_T^{miss} for each component of those uncertainties. Thus, the only part that remains for discussion is an uncertainty on the soft term. There are three components of the uncertainty: one due to E_T^{miss} scale and two

components of the E_T^{miss} resolution uncertainty. The resolution uncertainty is split into the longitudinal and perpendicular components with respect to an axis in the transverse plane. This axis is defined along the direction of a vectorial sum of all the hard objects in the event (electrons, muons and jets).

7.5.2 Theoretical uncertainties on the background

7.5.2.1 Modelling of V+jets production

The Sherpa2.2.1 V+jets samples include event weights reflecting variations of the PDF set, the value of α_s and the renormalisation and factorisation scales. These weights are used to estimate the uncertainty in the number of selected events in each bin of a given distribution in a given phase space region.

PDF Uncertainties PDF variations include 100 replicas of the nominal NNPDF3.0 PDF as well as central values for two different PDF set, MMHT2014nnlo68cl and CT14nnlo. The NNPDF intra-PDF uncertainty is estimated as the standard deviation of the set of 101 NNPDF3.0 sets. The envelope of the differences between the nominal NNPDF set and the other two PDF sets is used as an additional uncertainty.

α_s Uncertainty The effect of varying α_s from the nominal value of 0.118 up and down to 0.117 and 0.119 respectively is obtained from weights calculated using the same PDF set as the nominal but with these different values of α_s.

Scale Uncertainties The samples include weights for a "7-point" variation of the renormalisation (μ_R) and factorisation (μ_F) scale, i.e. varying the scales either together or independently by a factor of 2 up and down. The envelope of all these variations is taken as an estimate of the scale uncertainty.

Weights for the matching and the resummation scale are not yet implemented in the Sherpa 2.2.1 samples. The recommendation is to use Sherpa 2.1 samples that were generated with variations of these scales. However, it was found already in the 2015 analysis that these samples do not have sufficient statistics to yield a meaningful estimate of the uncertainty in the signal region. Since these uncertainties are expected to be much smaller than in particular the renormalisation scale, they are not taken into account at the moment.

Reweighting Uncertainty In addition to these uncertainties related to choices that have to be made in the simulation, a further uncertainty from the reweighting that will be applied to the simulation is considered. This is taken as the full size of the correction, i.e. the difference between the reweighted and the unweighted distributions.

7.5.2.2 Modelling of $t\bar{t}$ production

The uncertainties in the $t\bar{t}$ modelling are assessed from a number of simulation samples with different settings that are provided centrally. These samples are listed in Table 7.5. There are two samples that correspond to the up- and down-variation of different, but correlated sources of uncertainty-radiation parameters, scales, hdamp parameter. In addition, there is a sample produced with a different generator (MCatNLO) to assess differences in the modelling of the hard-scattering

and a sample that was done with a different showering programme (HERWIG7 instead of Pythia).

Moreover, the nominal sample contains weights for the PDF sets MMHT, CT14 and PDF4LHC as well as replicas and error sets for NNPDF3.0 and PDF4LHC, respectively. The default PDF is NNPDF3.0, the uncertainty is estimated from the PDF4LHC15 error sets, following the top group recommendations.

DSID	Comment	Generator
410511	variations (hdamp, scale, ISR/FSR)	PowhegPythia8
410512	variations (hdamp, scale, ISR/FSR)	PowhegPythia8
410225	Hard Scatter Generator	aMC@NLO+Pythia8
410525	Fragmentation/Hadronisation Model	Powheg+Herwig7
410527	Fragmentation/Hadronisation Model, dilepton filtered	Powheg+Herwig7
410248	colour reconnection	PowhegPythia8

Table 7.5: Alternative $t\bar{t}$ MC samples used to assess the $t\bar{t}$ modelling uncertainties. The first column gives the internal ATLAS dataset number (DSID), the second a brief description of the difference with respect to the nominal sample, the third gives the generator.

7.5.3 Theoretical signal uncertainties

The signal samples contain the same set of weights as the V+jets samples and the estimation of the uncertainty is done analogously. The scale uncertainty is again the largest contribution. Another sizeable contribution is the PDF uncertainty.

7.6 Control region plots

In this section, the MC modelling of various distributions in comparison to data in the control regions is discussed. In the interest of reading flow, only a few selected distributions are shown here. A more complete set is collected in App. B.5.

All the distributions shown in this section are based on the full (2015+2016) 13 TeV dataset corresponding to 36.1 fb^{-1}. All plots in this section show the distributions before applying a reweighting to the V+jets or $t\bar{t}$ background.

Object calibration, smearing and selection scale factors are applied according to the CP group twikies and their Moriond recommendations and with the use of AnalysisBase release 2.4.29.

The comparisons are done for the W (Fig. 7.6) and $t\bar{t}$ (Fig. 7.7 and Fig. 7.8) CRs. In each figure the grey band displays the total (experimental and theoretical) systematic uncertainty, i.e. quadrature sum of all its components. The hatched orange band shows the statistical uncertainty of MC. The black error bars display the statistical uncertainty of data.

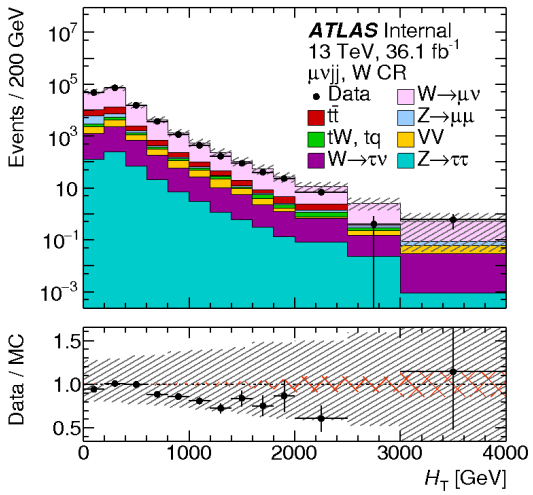
The systematic uncertainties are large depending on the distribution and region of the spectrum. This is driven by the large renormalisation scale uncertainty on

the W+jets background. Most of the deviations of the simulation from the data are covered by these large uncertainties. However, there are some clear trends in the data-to-MC ratios in some variables and in some cases the discrepancies are on the edge of the uncertainty band or even beyond it, e.g. for the distribution of H_T and m_{jj}. More precisely, the discrepancies occur for variables related to the jet activity, whereas lepton variables are generally well described.

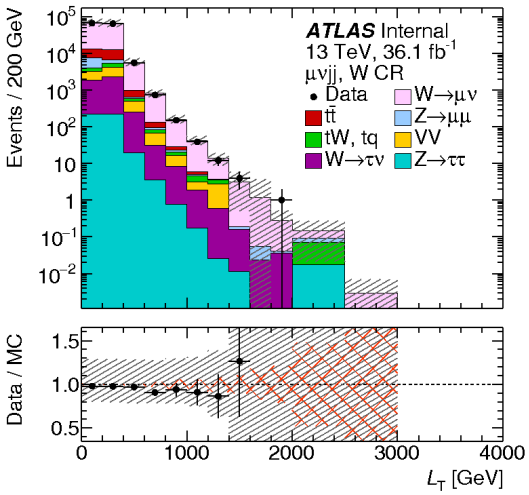
In the $t\bar{t}$ CR most variables, in particular angular or distance variables, are well described. However, there is a slight mismodelling in the (sub)leading jet p_T distribution and - possibly related - in variables such as S_T, H_T and the various LQ mass variables.

Comparisons have been done between $t\bar{t}$ samples generated with Powheg+Pythia6 and Powheg+Pythia8 respectively and the agreement is better with the latter.

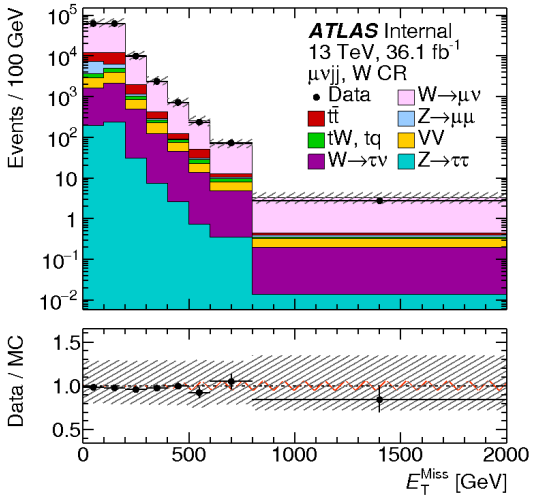
Since the two samples have different settings for the hdamp parameter, this could indicate some relation of the mismodelling and this parameter.



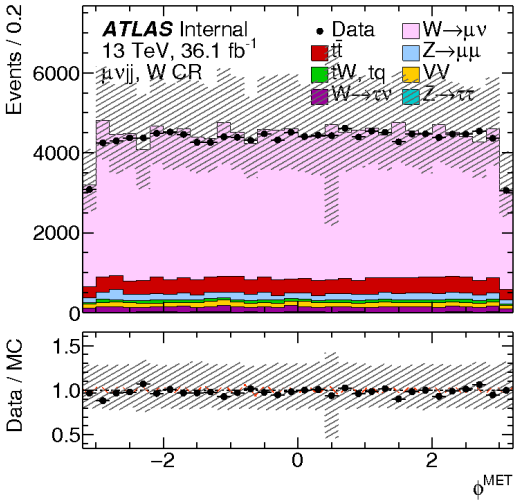
(a)



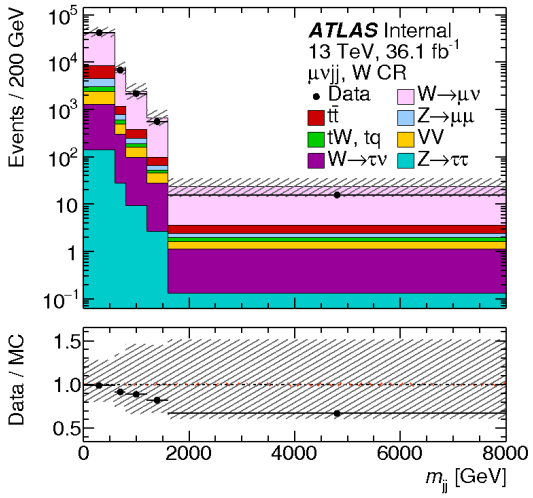
(b)



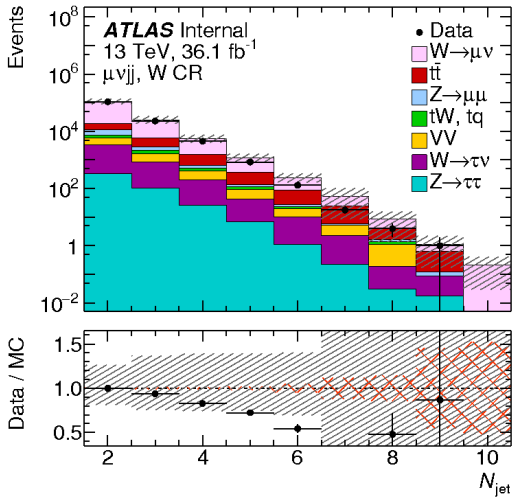
(c)



(d)



(e)



(f)

Figure 7.6: Kinematic distributions in the W CR. The grey band displays the total systematic uncertainty and the hatched orange band shows the statistical uncertainty in the simulation.

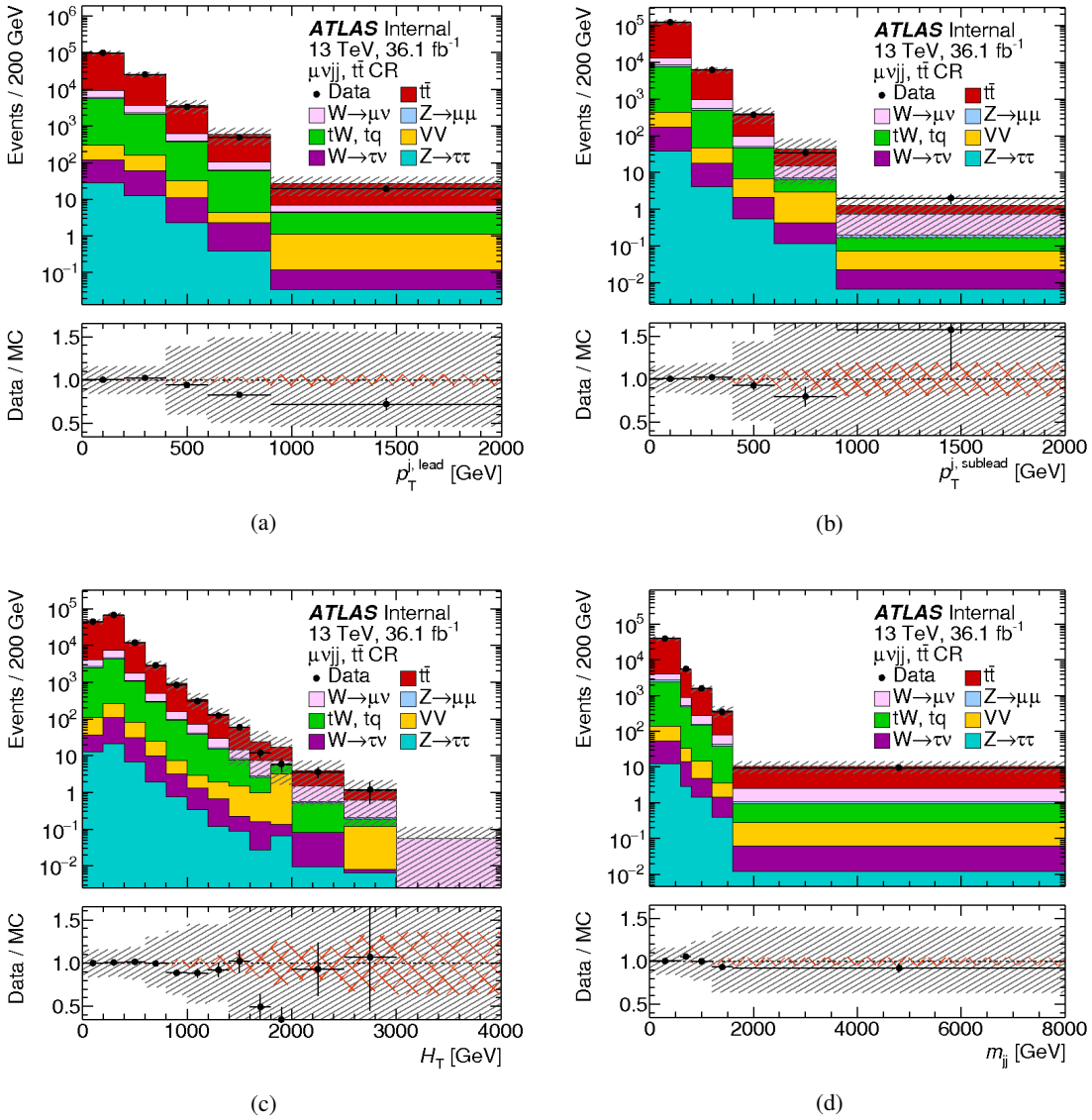


Figure 7.7: Kinematic distributions in the $t\bar{t}$ CR. The grey band displays the total systematic uncertainty and the hatched orange band shows the statistical uncertainty in the simulation.

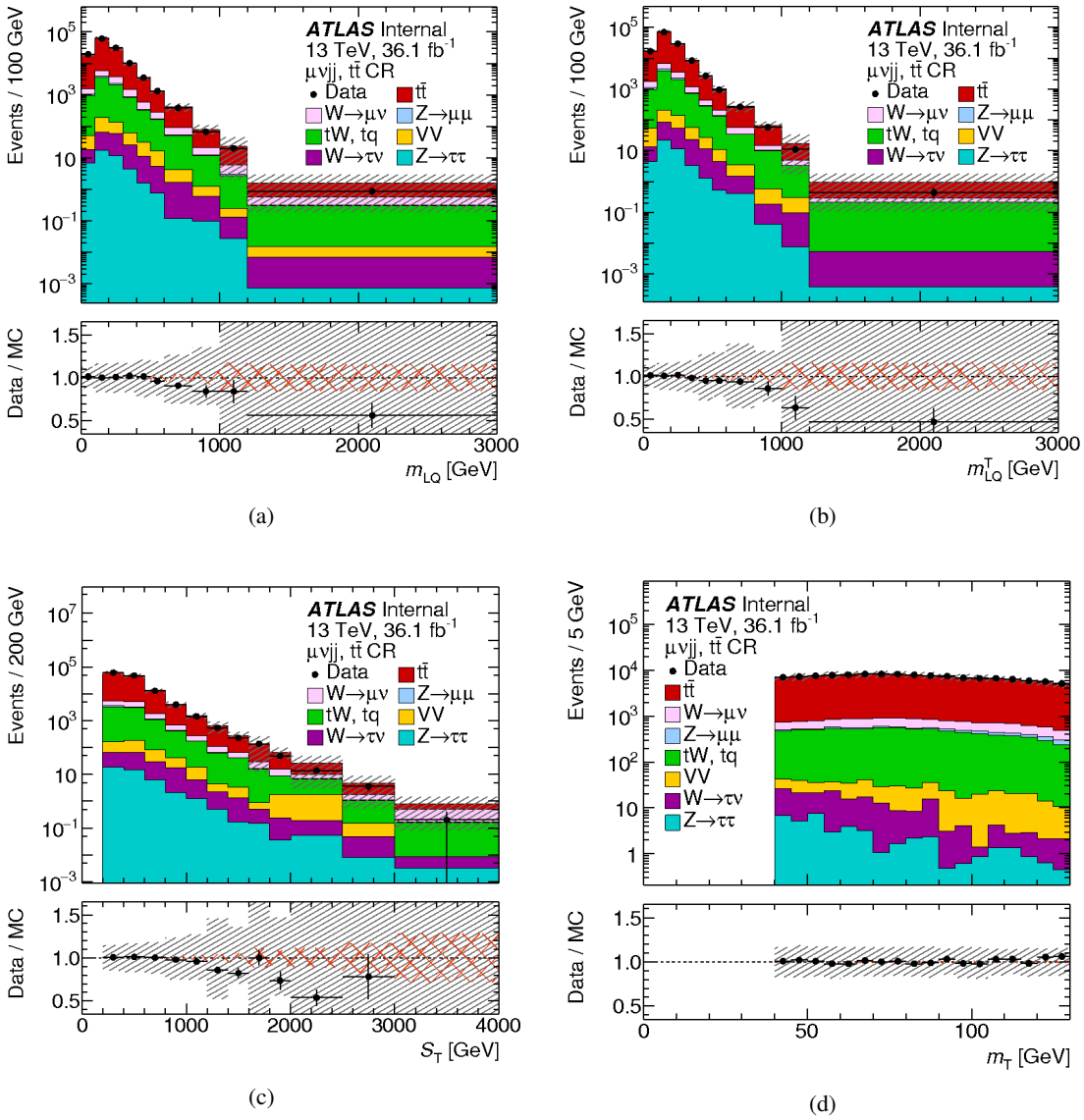


Figure 7.8: Kinematic distributions in the $t\bar{t}$ CR. The grey band displays the total systematic uncertainty and the hatched orange band shows the statistical uncertainty in the simulation.

7.7 Signal and background comparison

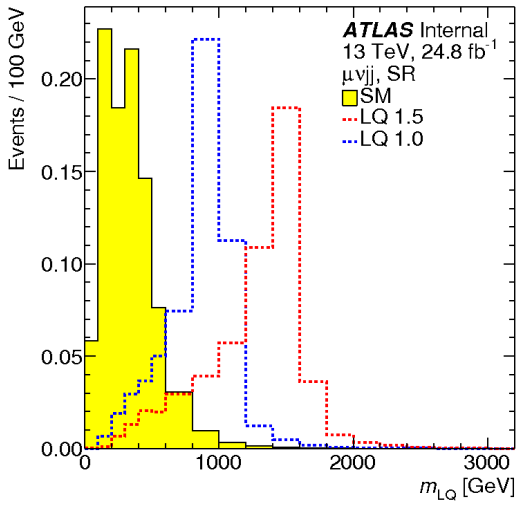
Comparisons of signal and background have been studied with the goal to provide insight into which variables provide a good separation between the two and are, thus, good candidates to be used as input to the BDT. This section shows a few selected distributions. The full set can be found in App. B.6 and App. B.7. The comparisons are done both for distributions normalised to unit area for a pure shape comparison and for distributions normalised to luminosity. The luminosity-normalised comparisons are done for events with $m_T > 130$ GeV. The shape comparisons are done in the SRs as defined in the cut-based analysis of 2015 data which have an additional cut of $S_T > 600$ GeV, but the general conclusions will not be greatly altered by this.

LQ mass variables and S_T provide good discrimination to lesser extent H_T, L_T, the single object p_T spectra and E_T^{miss}. Angular variables reveal some differences in the signal and background topologies: in the background, the two jets and the two leptons occur mostly back-to-back (in ϕ), whereas in the signal their azimuthal distance is more evenly distributed. On the other hand, the closed azimuthal distance between a jet and any lepton is typically smaller in background events than in signal events. Possible candidates are $min(|\Delta\phi(j_1,\ell_1)|,|\Delta\phi(j_1,\ell_2)|)$ or $min(|\Delta\phi(j_2,\ell_1)|,|\Delta\phi(j_2,\ell_2)|)$ (Fig. 7.9). The same distributions normalised to luminosity are shown in Fig. 7.10. The choice of input variables to the BDT will be discussed in more detail below.

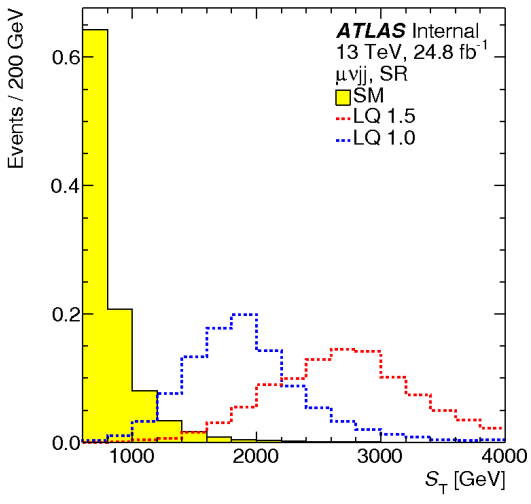
7.7.1 Candidate variables

Based on the comparisons between signal and background, a set of 14 variables is identified that might be considered as inputs to the BDT. These variables ordered by expected importance are listed below:

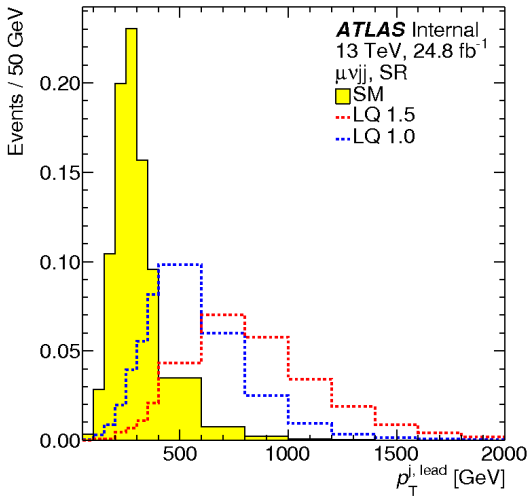
$$m_{LQ},\ m_{LQ}^{T},\ m_T,\ E_T^{miss},\ p_T^{j_2},\ p_T^{\ell_1},\ S_T,\ min(|\Delta\phi(j_2,\ell_1)|,|\Delta\phi(j_2,\ell_2)|),$$
$$min(|\Delta\phi(j_1,\ell_1)|,|\Delta\phi(j_1,\ell_2)|),\ H_T,\ L_T,\ \Delta\phi_{\ell\ell},\ \Delta\phi_{jj},\ \Delta\eta_{jj}.$$



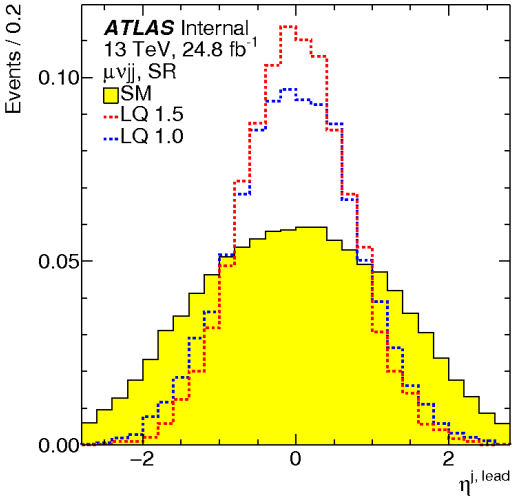
(a)



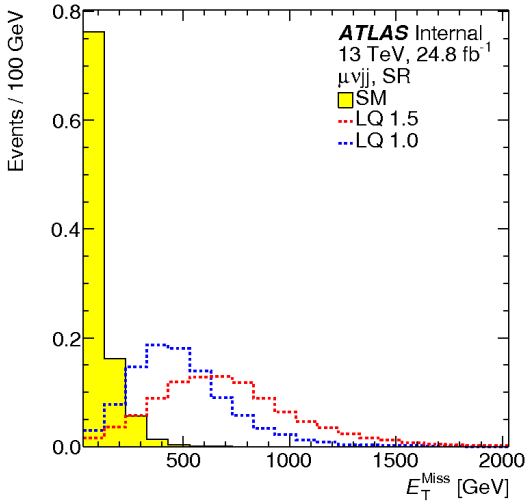
(b)



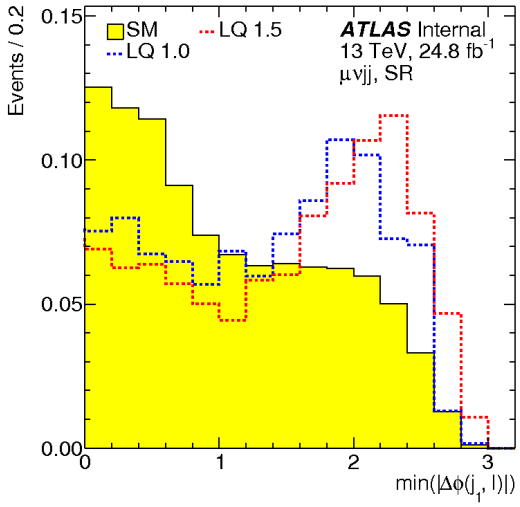
(c)



(d)



(e)



(f)

Figure 7.9: Kinematic distributions in the 2015 SR normalised to unit area.

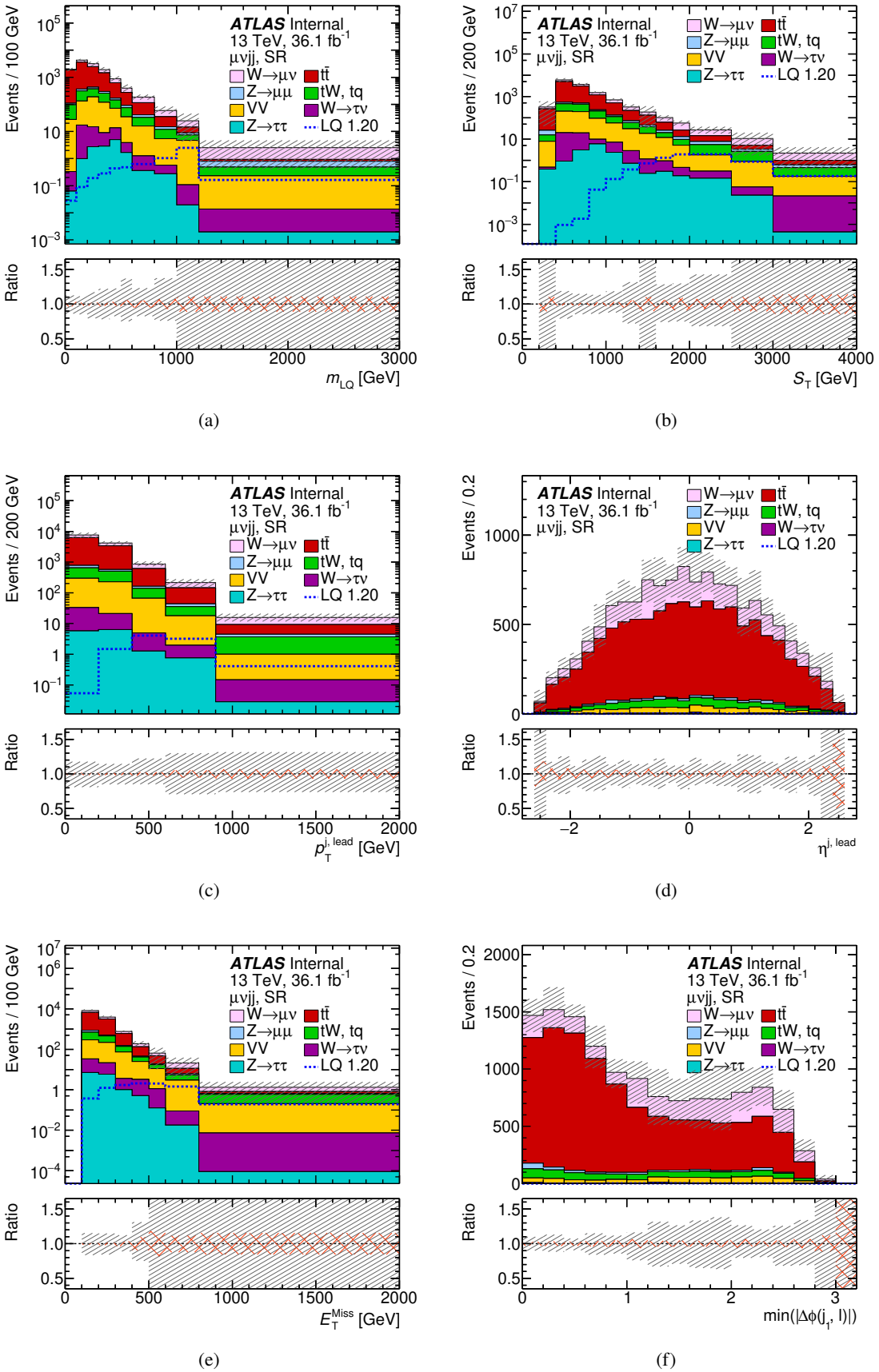


Figure 7.10: Kinematic distributions in the 2015 SR normalised to luminosity. The grey band displays the total systematic uncertainty. The hatched orange band shows the statistical uncertainty in the simulation.

7.8 V+jets reweighting

As was observed previously, the Sherpa V+jets samples model jet related variables poorly, which is a consequence of the choice that was made for the renormalisation scale. The V+jets CRs are used to derive a reweighting of the simulation to improve the description of data. This is important for the training of the BDT and it will also be used to correct the simulation before performing the fit described later. This procedure has been agreed upon within the L+X subgroup where several analyses are faced with this mismodelling issue.

The studies are in App. B.8 and they were done at an earlier stage of the analysis with a slightly different version of the ntuples and only with the 2016 data, but this should have no effect on the conclusions. The derivation of weights from a fit to the data/MC ratio corresponds to a newer version of the ntuples.

7.9 Boosted Decision Trees (BDT)

The present analysis uses Boosted Decision Trees (BDT) to increase the sensitivity to the LQ signal. This section describes the setup that is being used, how the input variables are chosen and what the output looks like.

7.9.1 General setup

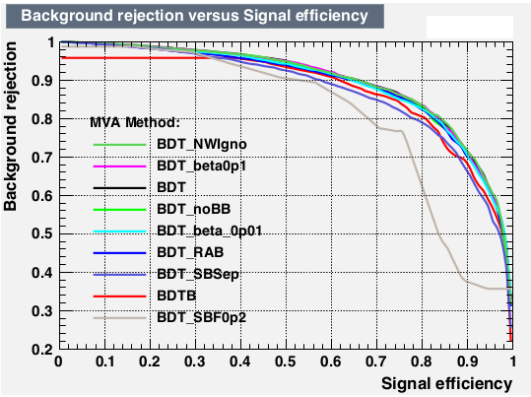
The TMVA implementation of a BDT is used with most settings chosen to be the TMVA default configuration. In particular, 850 decision trees are used, boosted adaptively with bagging, i.e. the classifier is retrained on a randomly selected subsample of the training sample. Here, a bagging sample size of 50% of the full training sample is used. The separation type used is the Gini index [99].

Some care has to be taken in principle when using MC samples with negative event weights (which is the case in this analysis both for signal and background), since this is a concept that is not foreseen in machine learning (ML) techniques. There is no canonical recommendation as to how to treat such cases. The default in TMVA is the "InverseBoostNegWeights" option, i.e. boosting the events with the inverse of the usual weight. Another option in TMVA is to ignore events with negative weights in the training (but applied in the evaluation). The resulting ROC curve is the green one labelled "BDT_NWIgno" in Fig. 7.11a to be compared to the black one which corresponds to the default setting. Both give very similar results. In general, little difference is observed between various different configurations and the default one is always among the best performing ones.

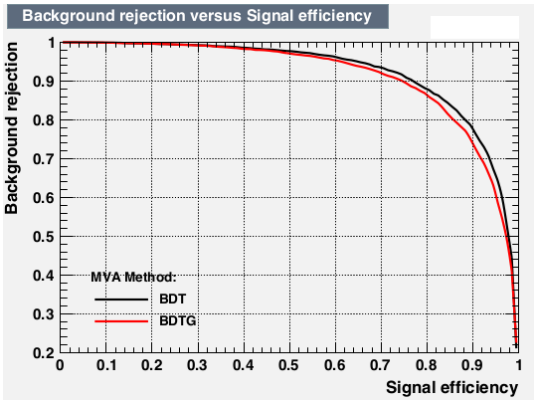
It has also been suggested to use "gradient boost" with the "pray" option, for which the negative weights are used as they are. This has been tested and results in a worse ROC curve than the default setting, as can be seen in Fig. 7.11b: the red ROC curve uses the gradient boost and praying, the black one is the nominal, i.e. adaptive boost and inverse weights.

The default values for the maximum tree depth and the minimum leaf size are 3 and 2.5% respectively. Following advice from the ML forum, in this analysis values of 2 and 5% are used in order to increase the robustness against overtraining in the presence of negative weights. Fig. 7.11c illustrates that the ROC has little dependence on the node size (blue and green vs black respectively). Moreover, increasing

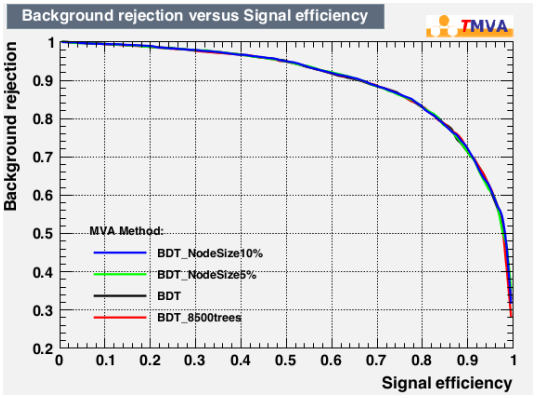
the number of decision trees by a factor of 10 (red line) does not change the ROC curve either.



(a)



(b)



(c)

Figure 7.11: Comparisons of ROC curves for different TMVA configuartions. The nominal one is always the black line.

In the exchange with the ML forum, it was also pointed out that if the inputs are similar for both negative and positive weights, the negative weights could be ignored in the training. The shapes of relevant variables have been compared for events with negative and positive weights, respectively, and overall little significant differences are found. Some examples are shown in Fig 7.12. The black points are the events with positive and the red ones the events with negative weights. On the left are the distributions normalised to luminosity, on the right normalised to unit area. The top plots are for a signal with an LQ mass of 1.3 TeV and the bottom plots are for $t\bar{t}$. Given the overall agreement between events with negative and positive weights, the negative event weights are ignored in the training and only taken into account in the testing.

To make full use of the MC statistics, two BDTs are used: one is trained on events with even event numbers and evaluated on events with odd event numbers and for the other BDT it is the other way round.

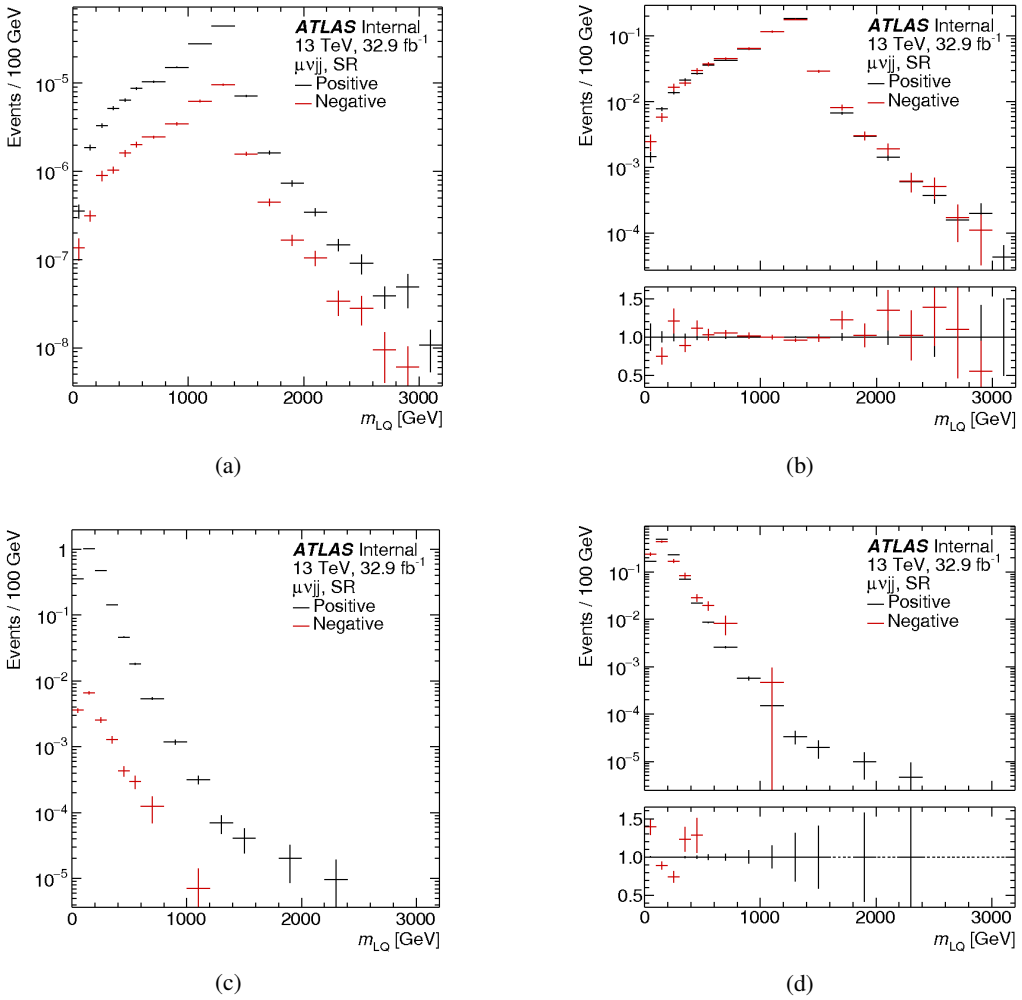


Figure 7.12: Comparisons of distributions with positive (black) and negative (red) weights normalised to luminosity (left) or to unit area (right) for a signal sample for an LQ mass of 1.3 TeV (top) and $t\bar{t}$ SR (bottom).

7.9.2 Correlations between variables

Apart from a good modelling of the individual distributions, it is also important for the BDT training that the correlations between input variables are modelled well. This has been studied by taking all pairs of input variables and looking at the profile of one as a function of the other. A few examples are shown in Fig. 7.13. The complete set of the variables can be found in Sec. 7.9.3. For these plots, the V+jets simulation has been reweighted in m_{jj} using the weights derived from a fit of the data-to-MC ratio. In general, a very good description of the correlations is observed for all regions.

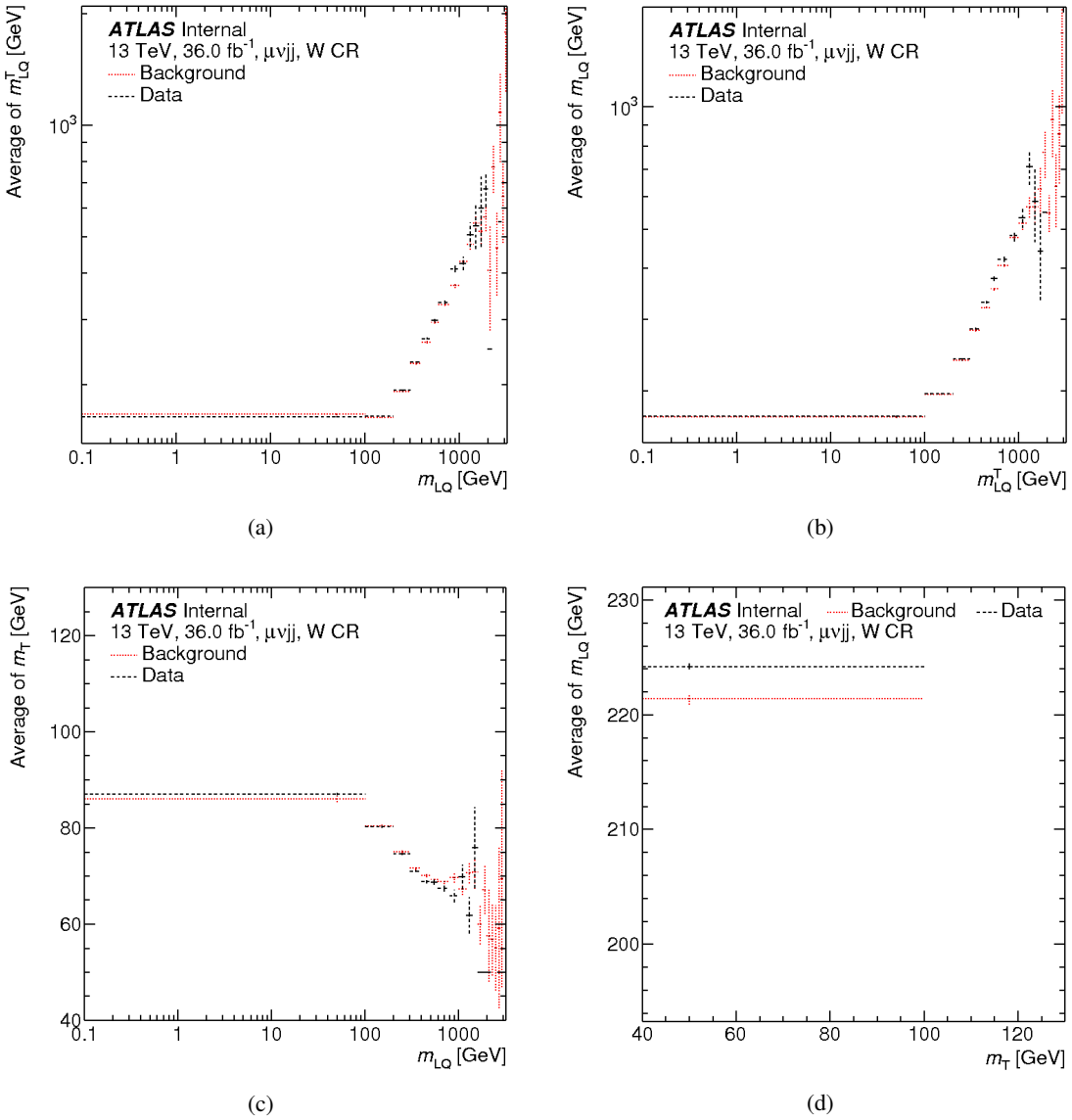


Figure 7.13: Correlation of BDT input variables in the W CR.

7.9.3 Input variables selection

After the V+jets modelling has been improved by the m_{jj} reweighting and the description of the correlations of the input variable candidates has been checked, further studies are done to reduce the set of input variables to the smallest set that provides good discrimination between signal and background for all mass points.

The decision which variables to use is based on comparing ROC curves of background rejection vs signal efficiency for different sets of variables in order to single out variables that to not add to the discrimination. The detailed procedure is as follows.

- The input variables are ordered according to some prior knowledge of their importance. The order is given with the first variable listed being the most important one.
- For N variables in the original set, BDT is trained N-1 times, each time removing one variable from the list, per default the least important one. However, at each step (when L variables are left), it is checked whether the prior ordering is reasonable. To do so, BDT is trained L times with L-1 variables and a different variable is removed each time. The resulting ROC curves are compared to the one obtained when training with L variables. The ROC curve obtained when the variable that was assigned the lowest importance a priori is removed should be closest to the ROC curve for L variables.
- In the end, that set of K variables is retained for which removing another variable changes the ROC curve significantly.
- The steps above are repeated for several signal points.

In the following, the outcome will be presented for three LQ mass points: 400 GeV, 800 GeV and 1.3 TeV.

Fig. 7.14 presents the ROC curve scan. From the plots in the left column, it is concluded that only the first six variables out of the original set are needed and these are m_{LQ}, m^{T}_{LQ}, m_T, E^{miss}_T, $p^{j_2}_T$ and $p^{\ell_1}_T$. The transverse mass significantly improves the ROC curve for all mass points. The impact of including the subleading jet p_T is largest for the medium and high mass range, whereas the leading lepton p_T contributes most at low mass.

The ROC curve using these six variables (black) is shown in the plots in the right column together with the ROC curves that are obtained when one of the six variables at a time is removed. All of the variables add to the discrimination, especially at low mass, where m_{LQ} is the most important variable.

In general, very good discrimination can be achieved especially for the higher mass points. As a result, the above set of six variables is used as inputs to the BDT training (also see App. B.9).

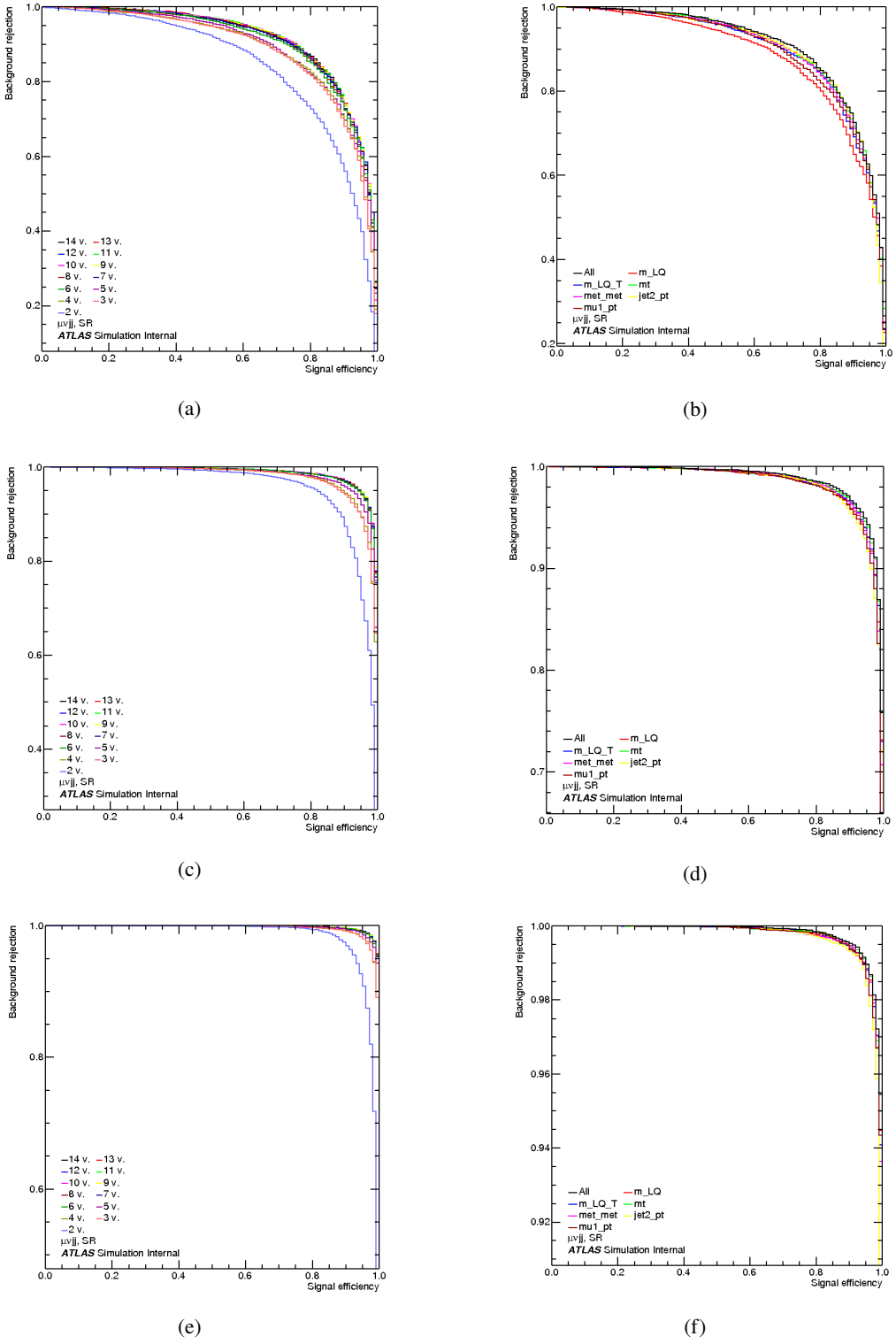


Figure 7.14: ROC curves for different sets of input variables and three mass points (400 GeV (top), 800 GeV (middle) and 1.3 TeV (bottom)). Left: Scan over original set of input candidates removing one by one successively. Right: Comparison of ROC curve for the final set of variables (black) with those where one variable at a time is removed.

7.9.4 Overtraining checks

To check that no overtraining occurs, BDT is evaluated on the training as well as the test sample and the obtained distributions are compared. Examples for the case of training on events with even event numbers are shown in Fig. 7.15 for three mass points (400 GeV, 800 GeV and 1200 GeV). Good agreement is observed, indicating that there is no overtraining. This also holds true for the other mass points and also the BDT trained on events with odd event numbers.

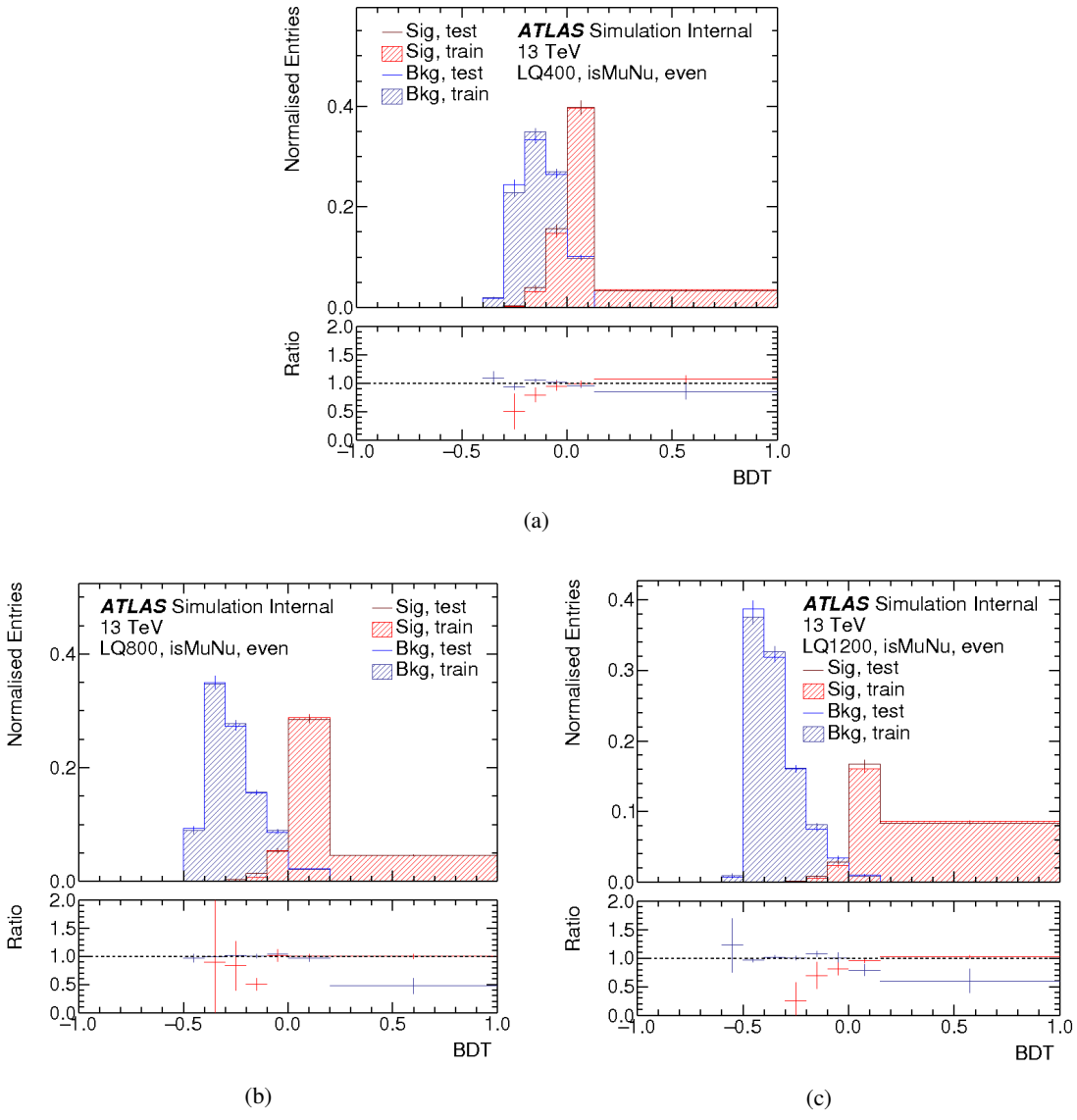


Figure 7.15: Distributions of the BDT output for signal (red) and background (blue) as evaluated on the training and the test sample for a nominal LQ mass of 400 GeV, 800 GeV and 1.2 TeV respectively.

7.9.5 Comparison of BDTs trained on even and odd event numbers

In this section, the output spectra of the two BDTs used are compared to each other. In principle, there should be no significant difference between the two if the

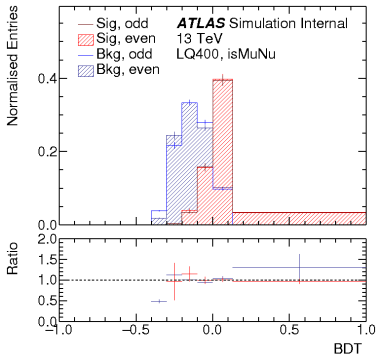
splitting of the MC samples based on the event numbers is random. Fig. 7.16 show examples for the mass points of 400 GeV, 600 GeV, 800 GeV, 1.0 TeV, 1.2 TeV, 1.4 TeV and 1.6 TeV.

The two BDTs always give very similar results for the signal samples. For the background, there are cases with statistically significant differences between the spectra, for example at 1.2 TeV, 1.4 TeV or 1.6 TeV.

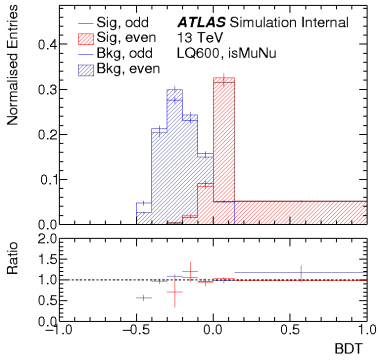
We have not found an explanation why the two spectra might be different. The impact on the result, however, is expected to be small, since the differences occur mostly in a range of the BDT spectrum that is not sensitive to the signal.

7.9.6 Output

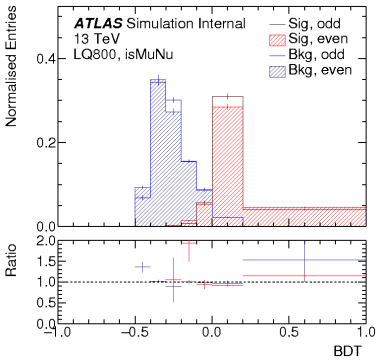
Fig. 7.17 - Fig. 7.21 show some of the distributions of the BDT output in the signal regions for LQ mass hypothesis from 200 GeV up to 1500 GeV with a pace of 50 GeV and from 1500 GeV up to 1800 GeV with a pace of 100 GeV. In all cases the backgrounds are accumulated at negative values, while the signal accumulates at positive values.



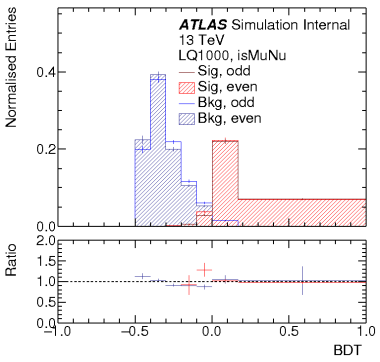
(a)



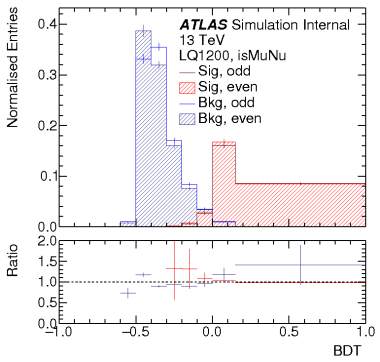
(b)



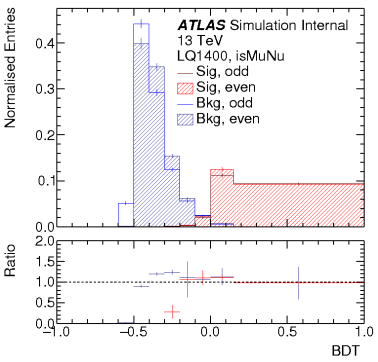
(c)



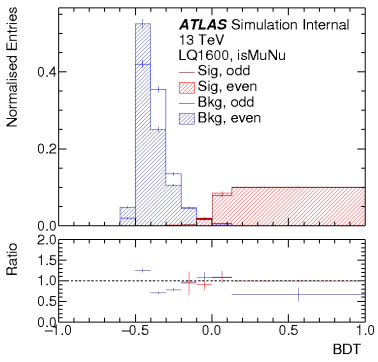
(d)



(e)



(f)



(g)

Figure 7.16: Distributions of the BDT output for signal (red) and background (blue) when trained on events with even or odd events numbers for an LQ mass of 0.4 TeV, 0.6 TeV, 0.8 TeV, 1.0 TeV, 1.2 TeV, 1.4 TeV and 1.6 TeV respectively.

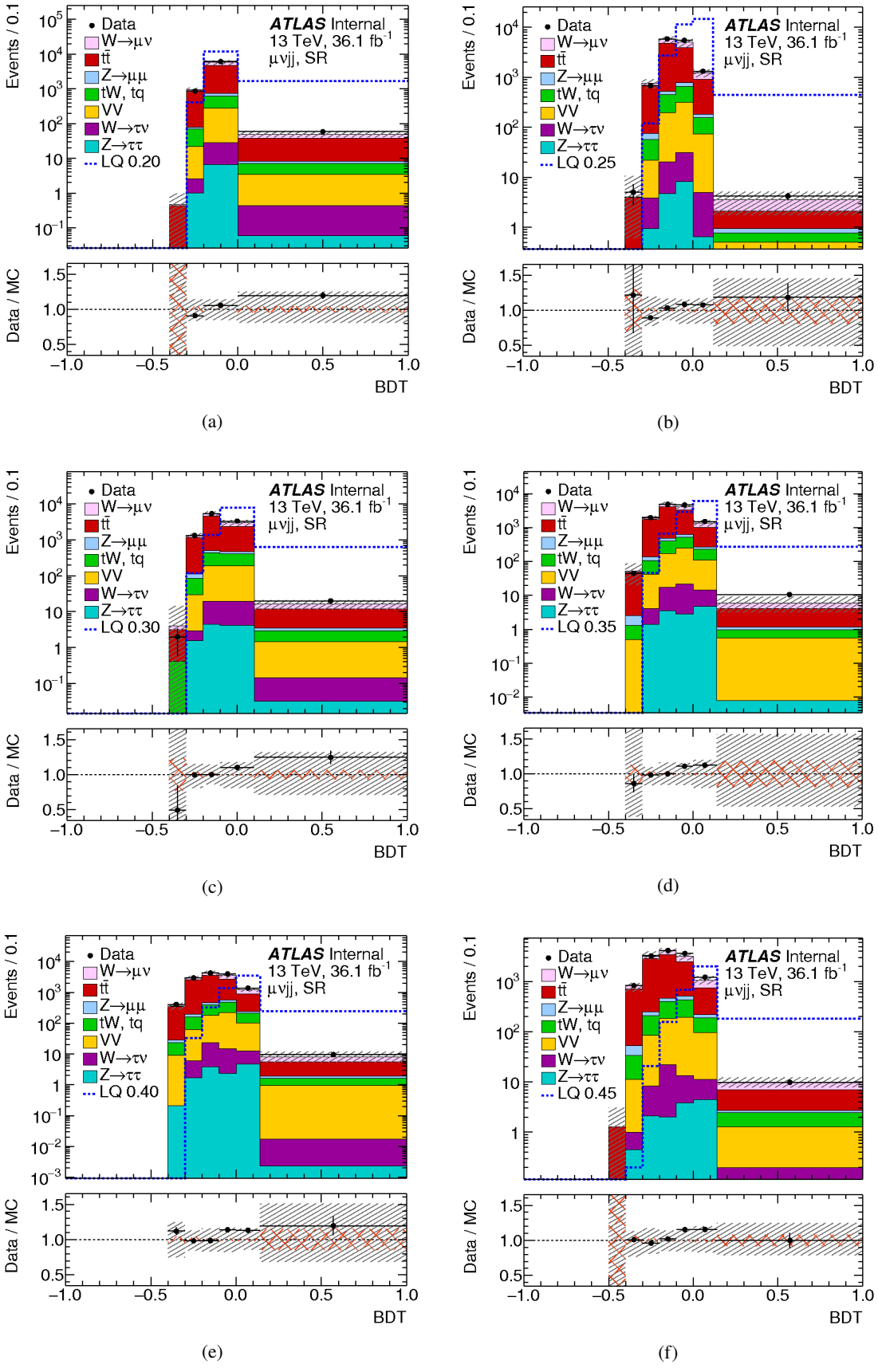
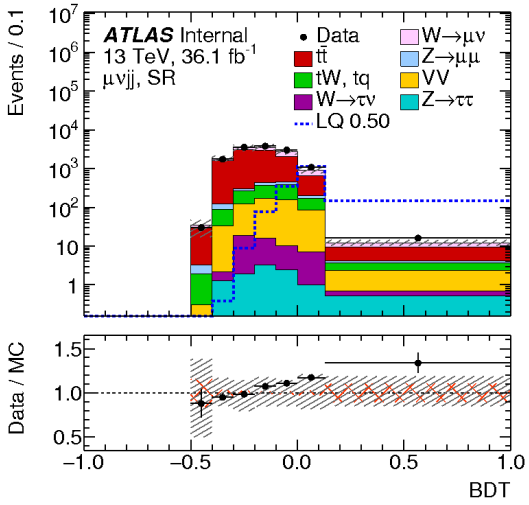
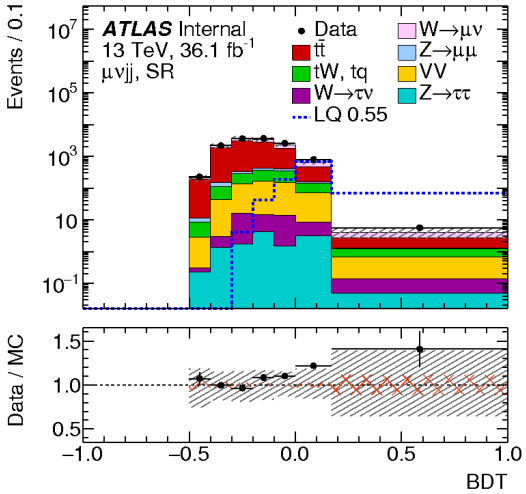


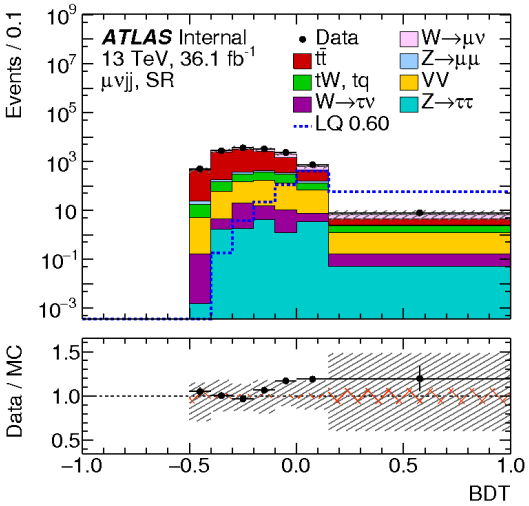
Figure 7.17: Distributions of the BDT output for different LQ masses.



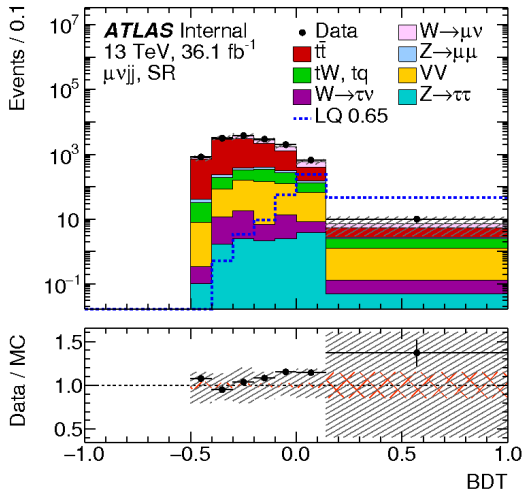
(a)



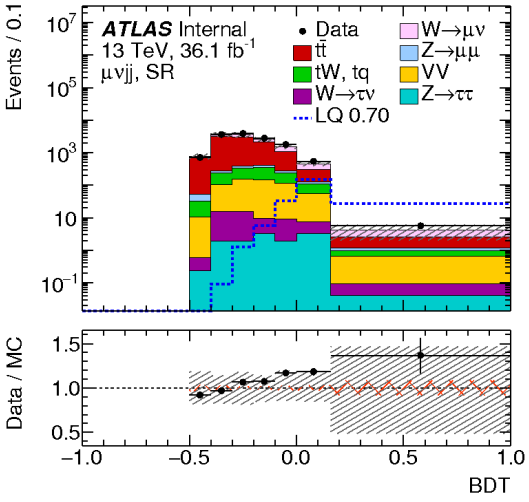
(b)



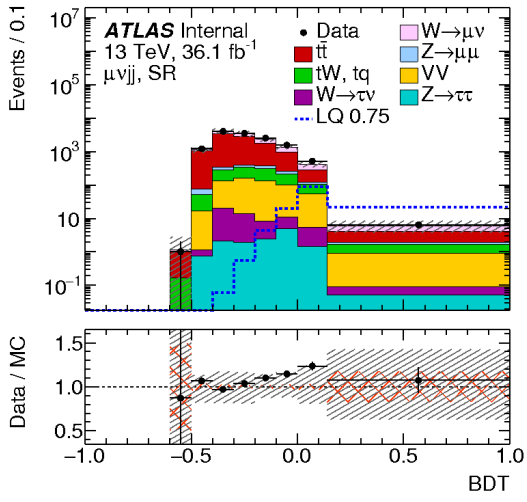
(c)



(d)



(e)



(f)

Figure 7.18: Distributions of the BDT output for different LQ masses.

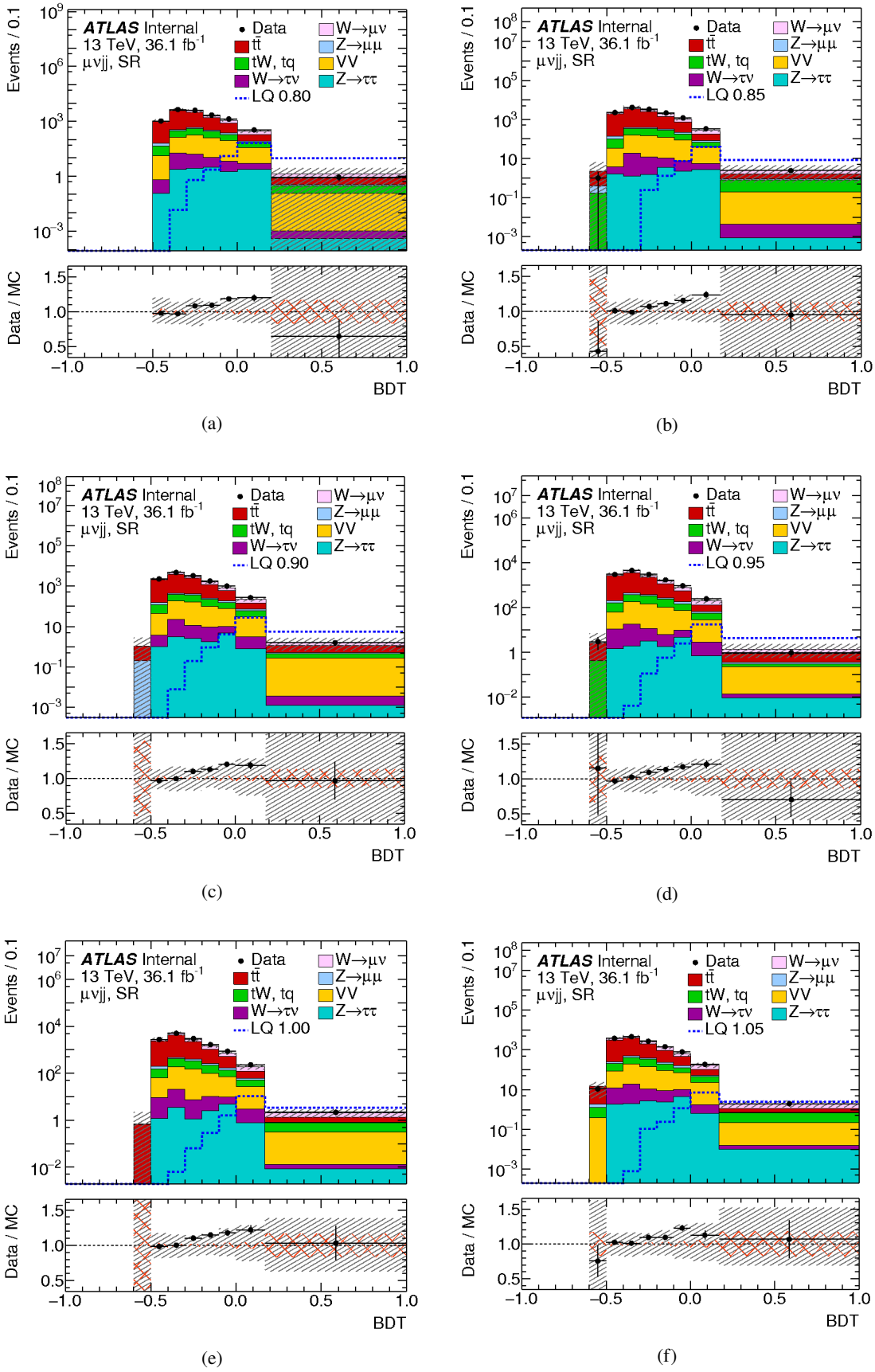
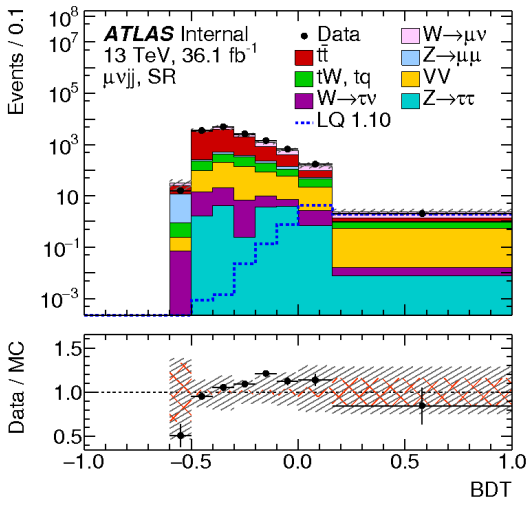
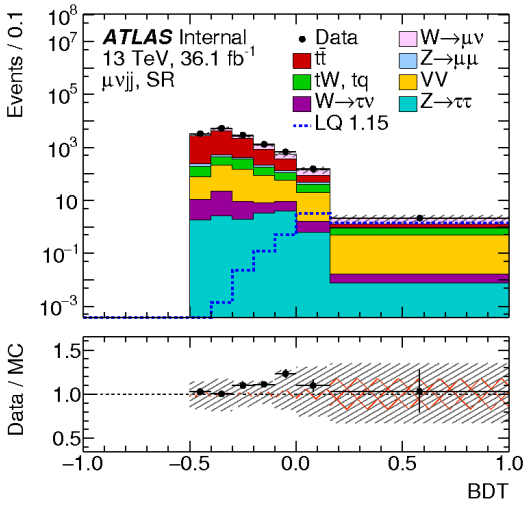


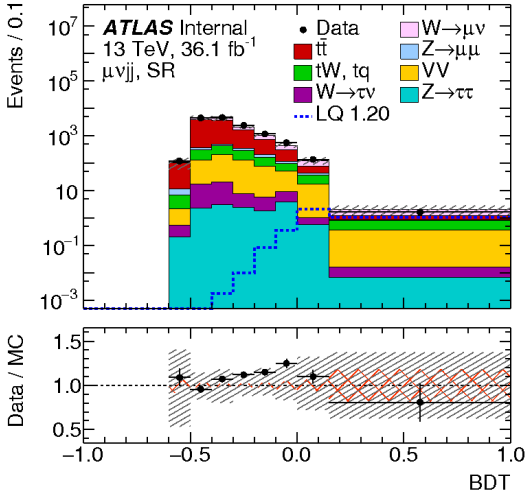
Figure 7.19: Distributions of the BDT output for different LQ masses.



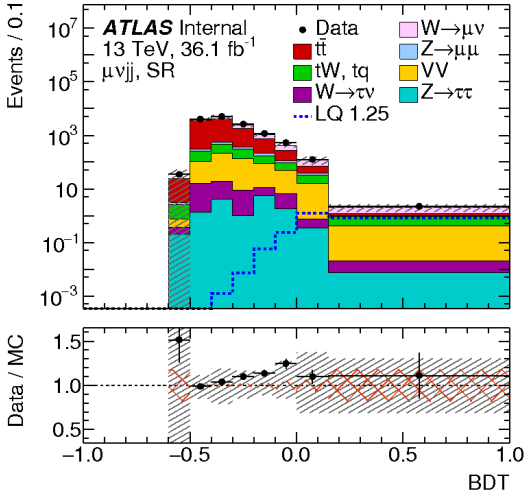
(a)



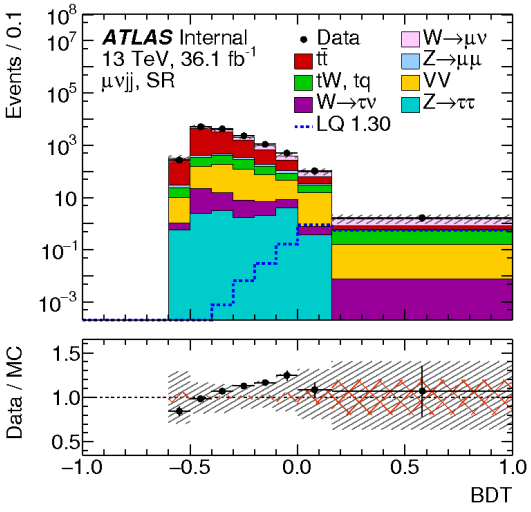
(b)



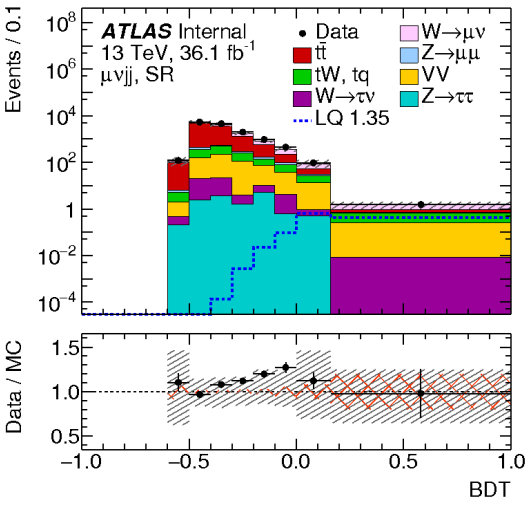
(c)



(d)



(e)



(f)

Figure 7.20: Distributions of the BDT output for different LQ masses.

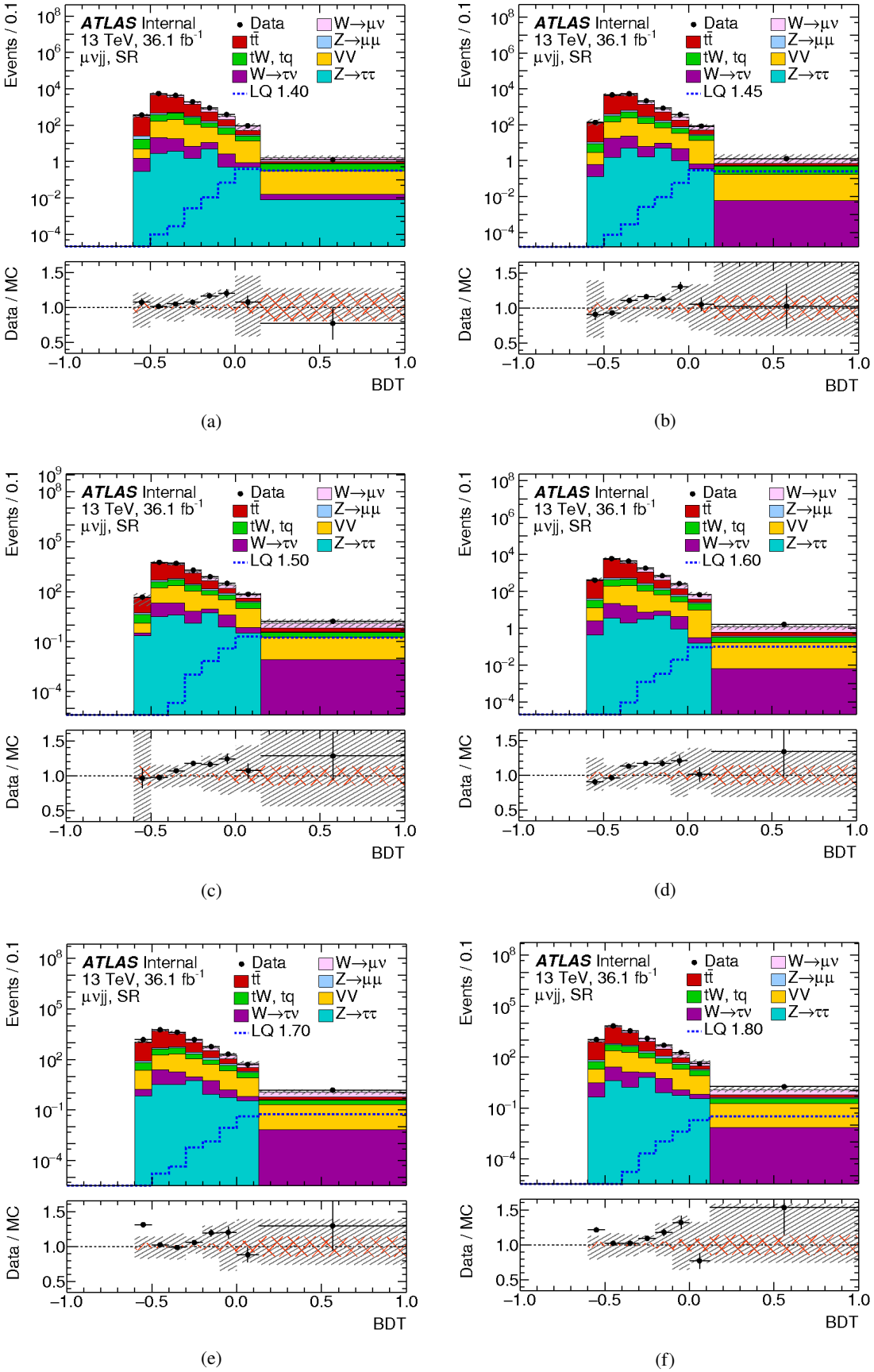


Figure 7.21: Distributions of the BDT output for different LQ masses.

7.10 Validation distributions

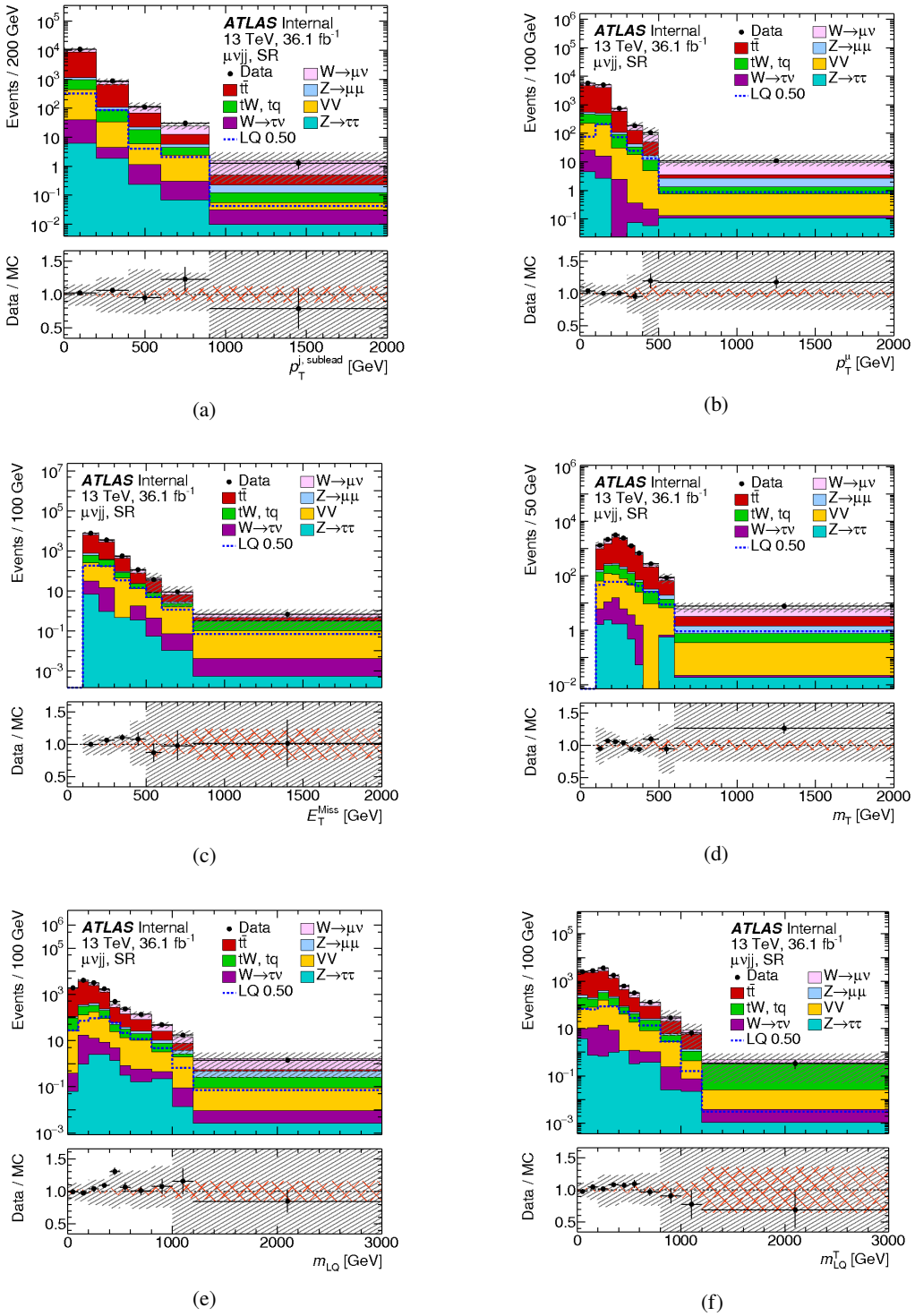


Figure 7.22: Distribution of most relevant BDT input variables in the validation regions for an LQ mass hypothesis of 500 GeV.

Distributions of BDT input variables for events with a BDT below 0 are presented.

As seen in previous section, this part of the spectrum is not sensitive to the signal and can thus be used as a validation region.

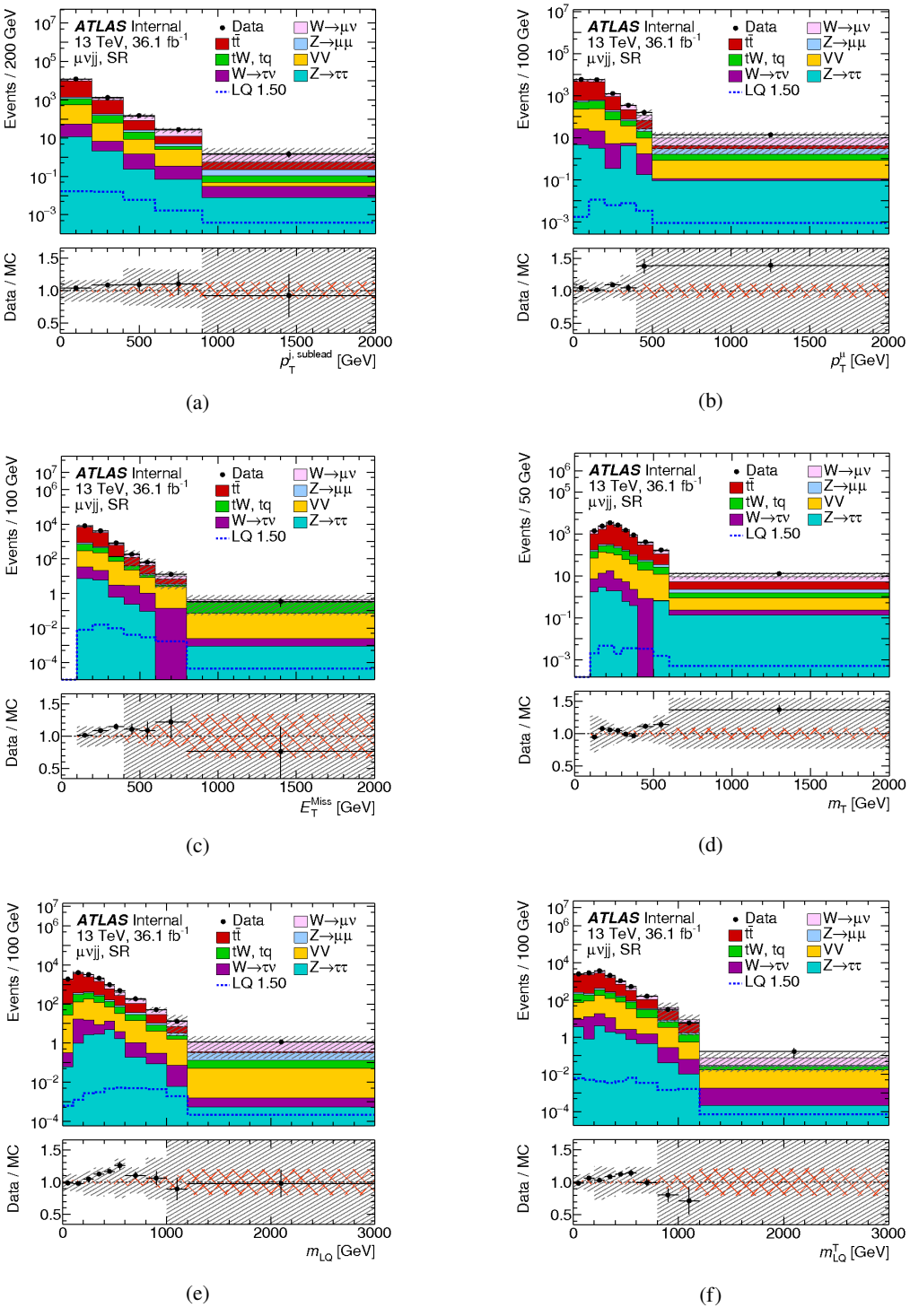


Figure 7.23: Distribution of most relevant BDT input variables in the validation regions for an LQ mass hypothesis of 1500 GeV.

Fig. 7.22 - 7.23 present example plots from this region are presented for LQ masses of 500 GeV and 1500 GeV. Additional mass points can be found in App.B.10.

7.11 Statistical evaluation

The results of the analysis are interpreted using the profile likelihood method. This section provides more details on the procedures applied in the statistical interpretation.

7.11.1 Post-unblinding observations

After the region of positive BDT score values in the search region was unblinded, no significant excess was observed (Sec. 7.12).

7.11.2 Methodology

The signal and backgrounds are described by a binned Probability Density Function (PDF) built using either the CRs (W+jets and $t\bar{t}$) or one SR. Uncertainties are incorporated in the PDFs as nuisance parameters. The PDF for the fit to the signal regions includes as parameter of interest the signal strength (μ_{sig}).

The SR drives the signal extraction, but adds negligible constraints on the background parameters. On the other hand, the CRs are enriched in background and have a negligible signal contamination and are, thus, used to normalise the predicted backgrounds to data. CRs consist of only one bin each.

The fit is performed in two stages: first, only the CRs are included in the fit to extract the normalisation factors for the two main backgrounds. These normalisation factors are then applied in the signal regions in the form of gaussian nuisance parameters that can change the respective background normalisation. Constraints on other nuisance parameters are not transferred from the CRs. (As will be seen later, there are no pulls or constraints in the CR-only fit, however.) While this in principle means a certain loss of information (and hence a conservative approach), this is outweighed by the simplification of the final fit it provides. This simplification results in a significant reduction of the CPU time needed to perform the fit and produce diagnostic plots like PLL curves which is an important advantage. Furthermore, it removes the difficulties arising from the large control region statistics from the signal extraction fit and thereby cures several of the issues described previously.

In the signal region, a single bin of the (reweighted) distribution of the BDT is used for the hypothesis testing. The bin range has been optimised for each mass point in each channel using the sensitivity measure

$$Z = \sqrt{2[(S+B)\cdot \ell n(1+\frac{S}{B}) - S]} \tag{7.6}$$

where S and B are the signal and background expectations respectively. This quantity was maximised for a set of cuts that in addition fulfil the following requirements:

- there are at least 2 background events expected in the chosen region,
- the MC statistical uncertainty on the estimated number of background events is less than 20% in the chosen region,
- if there are less than 10 background events expected in the chosen region, the MC statistical uncertainty is required to be less than 10%.

All the statistical and systematic uncertainties are included as nuisance parameters which are typically constrained by a Gaussian. The width of each constraint is specified according to the prescription of each uncertainty.

The optimal value and error of the signal strength and nuisance parameters as well as their correlations are determined simultaneously when the PDF is fitted to data. If the fit to data is minimally constrained, the nuisance parameters are simply propagating the effect of the uncertainties and only the signal strength adjusted. However, if the fit is over-constrained, the nuisance parameters (their mean value and error) can also be adjusted in order to optimally describe the data.

The Minuit minimisation strategy is set to 2, which is more robust against difficulties in finding a minimum than the default strategy 1.

The results in this analysis were produced with a modified version of ROOT 6.12.04 in which a cutoff on the tolerance value in Minuit2 has been removed. This in particular helped with cases of asymmetric post-fit errors and post-fit errors larger than 1.

7.11.3 Systematic uncertainties treatment

In this analysis, the nuisance parameters corresponding to systematic uncertainties are not allowed to change the overall normalisation of the backgrounds they affect, in order to remove degeneracies with the background normalisation factors that are extracted from the control regions. The uncertainties may change the relative size of a given background in different regions, but not the total amount. In that sense, they are shape uncertainties, even though only single-bin regions are used in the statistical evaluation. To achieve this, the histograms corresponding to the systematically varied distributions are normalised such that the total event number summed over all regions is conserved. The normalisation uncertainty is introduced via the normalisation scale factors and (their uncertainty). Experimental uncertainties are considered fully correlated (i.e., they are given the same name in HistFitter) between all processes and all regions that are included in a given fit.

Theory and modelling uncertainties are considered fully correlated between different regions in a fit, but uncorrelated between processes. For example, the V+jets scale uncertainties are treated as uncorrelated between W+jets and Z+jets.

In the signal regions, uncertainties that have an effect smaller than 2% on a given process are removed from the uncertainty list for this process. This is done before the normalisation mentioned above.

To improve the fit stability, some symmetrisation is applied to the uncertainties as described in the following. In cases in which both the up and the down variation corresponding to a given uncertainty result in a shift in the same direction, the larger of the two is used as a symmetric uncertainty. If the absolute values of the effects of the up and down variations differ by more than a factor of 5, again the larger of the two is applied as a symmetric uncertainty. The symmetrisation is done both in the CRs and the SRs.

Finally, a cutoff is introduced for the $t\bar{t}$ modelling uncertainties in the SR. This choice was made since the original uncertainties can be as large as hundreds of percent because of poor statistics in the signal region, in particular for the alternative samples used to estimate the variations. These uncertainties therefore have a large statistical component and the use of the full size of the uncertainty would mean a

double counting of the statistical uncertainties, but does not really reflect the uncertainty in the modelling. Different values for the cutoff have been tested (50%, 90%, 100%) and it was found that the resulting limits remain unchanged, but the fit stability increases when a 50% cutoff is chosen. Therefore, the $t\bar{t}$ modelling uncertainties are capped at 50% in the signal regions.

The list of systematics is presented below, all are double-sided:

- muon momentum scale
- muon momentum scale (charge dependent, based on combination of correction on combined (Z scale) and recombination of the corrections)
- muon momentum scale (charge dependent, based on the residual charge-dependent bias after correction)
- muon momentum resolution (Inner Detector measurement)
- muon momentum resolution (Muon Spectrometer measurement)
- muon trigger scaling factor (statistic uncertainty of the extraction method)
- muon trigger scaling factor (systematic uncertainty of the extraction method)
- muon efficiency scale factor (statistic uncertainty of the extraction method)
- muon efficiency scale factor (systematic uncertainty of the extraction method)
- muon isolation scale factor (statistic uncertainty of the extraction method)
- muon isolation scale factor (systematic uncertainty of the extraction method)
- muon track-to-vertex association scale factor (statistic uncertainty of the extraction method)
- muon track-to-vertex association scale factor (systematic uncertainty of the extraction method)
- jet energy scale (η intercalibration)
- jet energy scale (Nuisance Parameter 1)
- jet energy scale (Nuisance Parameter 2)
- jet energy scale (Nuisance Parameter 3)
- jet energy resolution
- b-tagging (in)efficiency scale factor (b jets)
- b-tagging (in)efficiency scale factor (c jets)
- b-tagging (in)efficiency scale factor (light jets)
- b-tagging (in)efficiency scale factor (extrapolation)
- b-tagging (in)efficiency scale factor (extrapolation from charm)
- $t\bar{t}$ modelling: generator comparison

- $t\bar{t}$ modelling: shower description
- $t\bar{t}$ modelling: radiation parameter variations
- V+jets: PDF uncertainty
- V+jets: α_s uncertainty
- V+jets: scale uncertainty
- V+jets: reweighting uncertainty

The $+1\sigma$ boundary of the jet energy resolution uncertainty band is determined by increasing the energy resolution by 1σ of its uncertainty. Since it is impossible to decrease the jet energy resolution, the uncertainty is symmetrized in order to get the -1σ boundary of the uncertainty band.

7.11.4 Likelihood

A maximum likelihood fit is used to constrain the MC expectation and the uncertainties with data. The likelihood is also used to search for the presence of a signal using normalisation factors extracted from the CRs and the SRs.

For the fit to the CRs, the likelihood can be expressed as

$$L(N,\theta^0 \mid \beta,\theta,\gamma) = \prod_j^{CRs} P\left(N_j \mid \left[\sum_\ell^{processes} \beta_\ell B_{\ell j} \times \prod_k^{systs} \nu_{j\ell k}(\theta_k)\right] \times \gamma_j\right) \times \prod_k^{systs} G(\theta_k^0 \mid \theta_k, 1) \tag{7.7}$$

where the number N refers to the observed number of events (in a given CR j), the B is the expected number of background events (for a given process ℓ in CR j) that is normalised by scaling factors β. The β parameters for the two main backgrounds (W+jets and $t\bar{t}$) are floating parameters in the fit. For the rest of the backgrounds β is effectively constant at 1.

The functions ν are the response functions for the nuisance parameters θ, i.e. they model the effect of a given uncertainty on the number of events. These response functions are determined in HistFitter/HistFactory via (vertical) morphing of the template histograms for the nominal distribution and the $\pm$ 1σ variations. The strategy used is the default in HistFactory and corresponds to a 6^{th} order polynomial interpolation with an exponential extrapolation beyond $\pm$ 1σ.

The γ parameters incorporate the effects of finite MC sample statistics. Finally, Gaussian constraint terms are introduced for the nuisance parameters of systematic uncertainties.

In an analogous way, the likelihood for the fit to the SR can be written as:

$$L(N,\theta^0 \mid \mu,\beta,\theta,\gamma) = \prod_j^{CRs} P\left(N_j \mid \left[\mu S_j \times \prod_k^{systs} \nu_{jk,sig}(\theta_k) + \sum_\ell^{processes} \beta_\ell B_{\ell j} \times \prod_k^{systs} \nu_{j\ell k}(\theta_k)\right] \times \gamma_j\right) \times \prod_k^{systs} G(\theta_k^0 \mid \theta_k, 1) \times \prod_m^{M} G(\beta_m^0 \mid \beta_m, 1) \tag{7.8}$$

where the additional parameter μ, the signal strength, is introduced as the parameter of interest. The expectation value for the number of events contains an additional

term μS, where S is the predicted number of signal events modified according to the response functions ν for the signal.

The normalisation factors β of the main backgrounds are included with Gaussian constraint terms. M = 2 refers to the number of background processes for which the normalisation is determined from the two control regions. For the other nuisance parameters no constraints are transferred from the CRs, i.e. the corresponding Gaussians are based on the pre-fit distributions of the nuisance parameters.

7.11.5 Test statistics

The test statistics used for the exclusion fit is shown in Eq. (7.9) and Eq. (7.10). Here $\widehat{\mu}$ and $\widehat{\theta}$ represent, respectively, the values of the signal strength and of the nuisance parameters which maximise the likelihood and $\widehat{\widehat{\theta}}$ are the value of the nuisance parameters which maximise the likelihood for a given signal strength μ.

$$q_\mu = -2\ell n \frac{L(data \mid \mu, \widehat{\widehat{\theta}}_\mu)}{L(data \mid \widehat{\mu}, \widehat{\theta})}, \qquad \widehat{\mu} < \mu \tag{7.9}$$

$$q_\mu = 0, \qquad \widehat{\mu} > \mu \tag{7.10}$$

7.12 Results and interpretation

In this section, we first discuss the fit to the control regions, then the results of the exclusion fit and finally the obtained limits are presented.

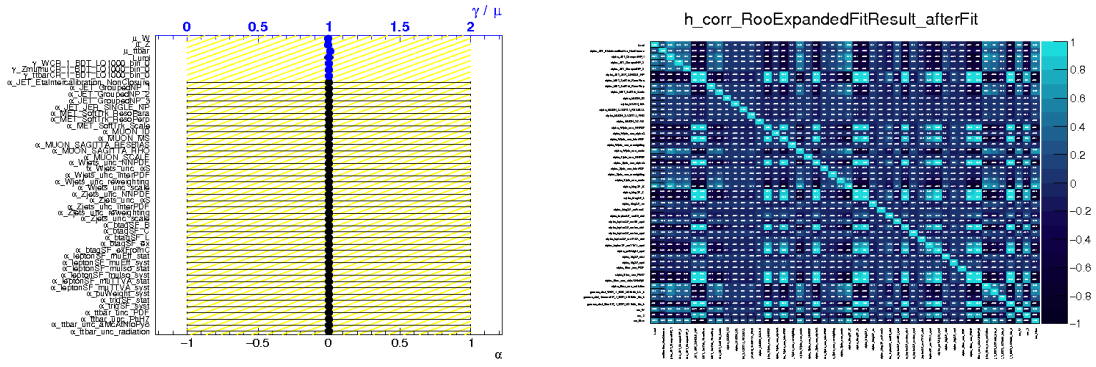
7.12.1 Control region fit

As described previously, a simultaneous fit to both control regions is performed first to determine the background normalisation factors. Fig. 7.24a shows the pulls and constraints on the fit parameters from this fit. No pulls or constraints of any of the systematics are observed. The normalisation factors are found to be close to unity, as can also be seen in Tab. 7.6.

Control Region	Scale Factor (SF)
μ_{Wjets}	$0.995^{+0.018}_{-0.019}$
$\mu_{t\bar{t}}$	$1.009^{+0.019}_{-0.018}$

Table 7.6: Scale factors for both main backgrounds as obtained from the CRs fit.

The correlation matrix is shown in Fig. 7.24b. It shows a number of very high correlations and anti-correlations. However, this is an artefact of the issues with large observed numbers of events (and correspondingly large correlations in the likelihood) and does not reflect the actual correlations between nuisance parameters.



(a) (b)

Figure 7.24: Fit results from CR fit. Left: Pulls and constaints. Right: Correlation matrix.

The Hessian matrix in Minuit2 fails one of the criteria for positive-definiteness leading to a fixed term being added to the diagonal in order to make the matrix inversion possible. This renders the displayed correlation matrix meaningless, however. It is shown just for the shake of completeness.

A more important check of the quality of the fit are the PLL curves for the nuisance parameter scans. Some representative examples are shown in Fig. 7.25. The blue (NLL) curve is obtained from a scan of the respective nuisance parameter while all other parameters are fixed at their values in the global minimum. The red (PLL) curve corresponds to scanning the respective nuisance parameter and re-minimising with respect to the other parameters for each scan point.

All curves show a smooth, parabolic behaviour and the same holds true for the other nuisance parameters not shown here. It is therefore concluded that there are no actual problems in the likelihoods and the normalisation factors obtained from the control region fit can be applied in the signal region fit.

Tab. 7.7 shows the pre- and post-fit event yields in the control regions together with the total uncertainties. There are more than 100,000 events in each region.

A breakdown of the uncertainties for the main backgrounds is shown in Tab. 7.8 for W^-CR and in Tab. 7.9 for the $t\bar{t}$ CR. Uncertainties are displayed when they are larger than 0.5% for at least one of the considered processes.

The uncertainty on the $t\bar{t}$ yield is dominated by the modelling uncertainties, especially the generator comparison and shower modelling. In this CR there is a larger number of uncertainties displayed, mostly due to the presence of uncertainties related to E_T^{miss} and b-tagging. The E_T^{miss} related uncertainties are at the sub-percent level, while the b-tagging uncertainties can be of the order of several % units. The $t\bar{t}$ modelling uncertainties are very small in this region ($< 1\%$), while the V+jets scale uncertainties amount to several % units.

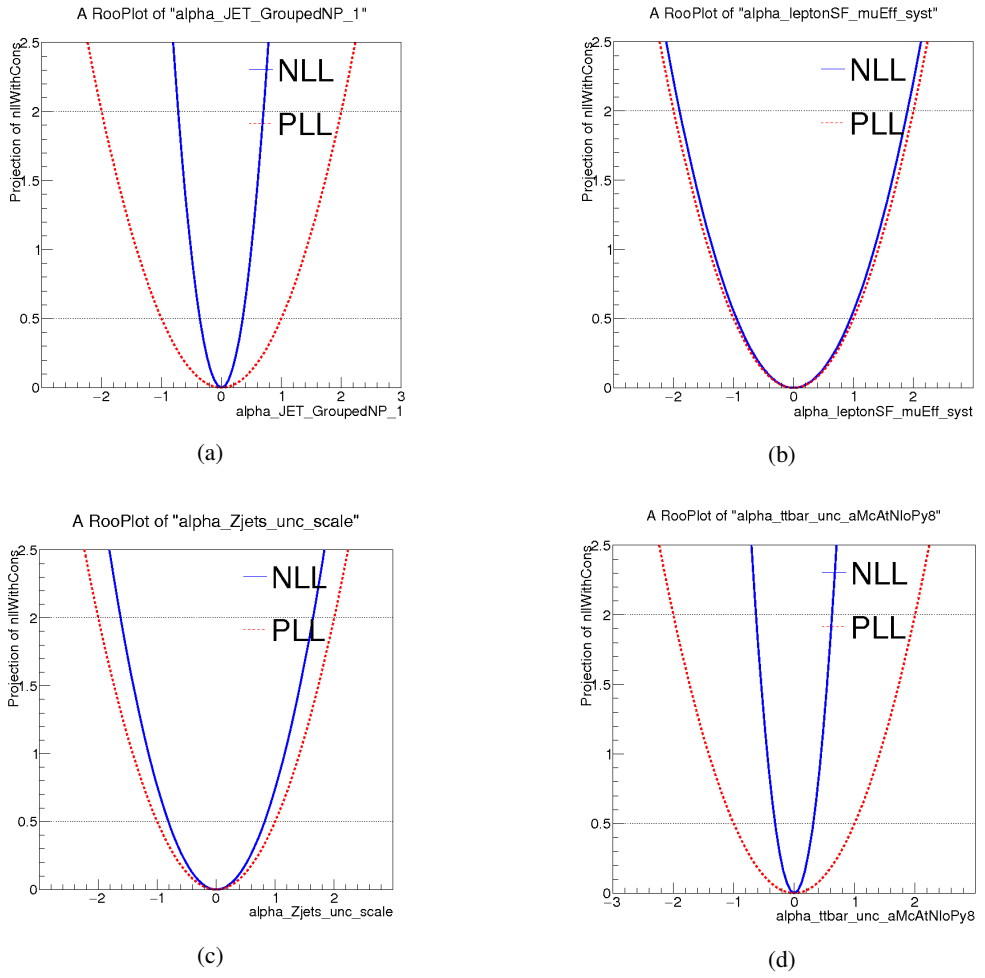


Figure 7.25: Examples of PLL curves. The curves are shown for one of the jet energy scale uncertainty nuisance parameters (top left), muon efficiency scale factor systematic (top right), Z+jets scale uncertainty parameter (bottom left) and the $t\bar{t}$ generator uncertainty (bottom right).

In the W CR, the largest experimental uncertainties are jet and b-tagging related and are at the order of several % units. Other experimental uncertainties are $< 1\%$. Theoretical and modelling uncertainties are a few % units.

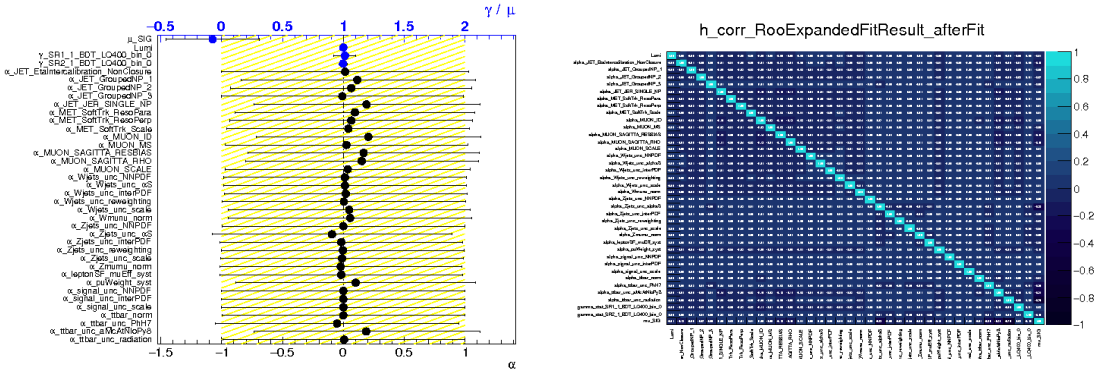
7.12.2 Exclusion fit results

Using the scale factors obtained from the CR fit, a combined fit to both signal regions is performed for each LQ mass point.

Fig. 7.26 shows as an example the post-fit parameters (left) and the correlation matrix (right) for the exclusion fit for a mass point of 400 GeV. There are some pulls of mostly experimental systematic uncertainties, but none that go beyond 0.3σ. None of the parameters is significantly constrained.

The correlations are generally very small, but there are some rather large (anti -)correlations with the signal strength for large the $t\bar{t}$ modelling uncertainties.

For all fits, the PLL curves for all nuisance parameters have been inspected carefully. Some representative examples are shown in Fig. 7.27 for the 400 GeV mass



(a) (b)

Figure 7.26: Post-fit parameters (left) and their correlations matrix (right) for exclusion fit for a leptoquark mass of 400 GeV.

point. They all show smooth parabolic PLL curves giving confidence that there are no critical issues with the fit.

A very similar behaviour for the correlations is observed for other mass points, examples are shown for LQ masses of 800 GeV and 1200 GeV in Fig. 7.28 and Fig. 7.29 respectively. The pulls and constraints typically become less pronounced for higher mass points, as demonstrated for example in Fig. 7.29.

Event Yields	W+jets CR	$t\bar{t}$ CR
Observed events	138,599	128,723
Fitted Bkg events	$138,599.00 \pm 585.76$	$128,723.00 \pm 456.27$
Fitted Z $\rightarrow \mu\mu$ events	$5,329.11 \pm 242.86$	828.85 ± 51.80
Fitted $t\bar{t}$ events	$12,103.04 \pm 717.93$	$114,983.48 \pm 423.63$
Fitted W $\rightarrow \mu\nu$ events	$111,568.69 \pm 1,472.99$	$4,745.30 \pm 219.83$
Fitted VV events	$3,555.74 \pm 93.22$	305.37 ± 12.68
Fitted Wt events	$2,043.13 \pm 91.99$	$7,709.30 \pm 184.38$
Fitted W $\rightarrow \tau\nu$ events	$3,999.29 \pm 91.37$	150.69 ± 7.51
MC exp. events	$138,969.15 \pm 2,294.01$	$127,616.02 \pm 2,069.24$
MC exp. Z $\rightarrow \mu\mu$ events	$5,353.97 \pm 327.12$	832.74 ± 88.81
MC exp. $t\bar{t}$ events	$11,984.27 \pm 1,352.18$	$113,852.72 \pm 1,418.36$
MC exp. W $\rightarrow \mu\nu$ events	$112,032.69 \pm 824.38$	$4,765.14 \pm 546.71$
MC exp. VV events	$3,555.75 \pm 131.44$	305.38 ± 28.02
MC exp. Wt events	$2,043.17 \pm 174.11$	$7,709.36 \pm 282.89$
MC exp. W $\rightarrow \tau\nu$ events	$3,999.30 \pm 122.85$	150.69 ± 15.28

Table 7.7: Event yields in the control regions with total uncertainties.

Uncertainty in Estimate	$W \to \mu\nu$	$t\bar{t}$	$Z \to \mu\mu$	$W \to \tau\nu$
Total background expectation	111,568.69	12,103.04	5,329.11	3,999.29
Total statistical ($\sqrt{Nexp}$)	±334.02	±110.01	±73.00	±63.24
Total background systematic	±1,472.99 [1.32%]	±717.93 [5.93%]	±242.86 [4.56%]	±91.37 [2.28%]
mu_W	±1,798.88 [1.6%]	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
btagSF_B	±96.30 [0.09%]	±232.37 [1.9%]	±15.17 [0.28%]	±3.07 [0.08%]
btagSF_C	±89.40 [0.08%]	±215.42 [1.8%]	±14.09 [0.26%]	±2.85 [0.07%]
btagSF_L	±82.25 [0.07%]	±197.78 [1.6%]	±12.96 [0.24%]	±2.62 [0.07%]
MET_SoftTrk_ResoPerp	±81.99 [0.07%]	±195.64 [1.6%]	±6.45 [0.12%]	±2.62 [0.07%]
JET_JER_SINGLE_NP	±74.98 [0.07%]	±178.52 [1.5%]	±55.77 [1.0%]	±2.39 [0.06%]
JET_GroupedNP_1	±72.68 [0.07%]	±177.82 [1.5%]	±93.58 [1.8%]	±2.32 [0.06%]
puWeight_syst	±53.86 [0.05%]	±127.82 [1.1%]	±55.12 [1.0%]	±1.72 [0.04%]
MUON_SAGITTA_RHO	±40.69 [0.04%]	±95.33 [0.79%]	±12.11 [0.23%]	±1.30 [0.03%]
MET_SoftTrk_ResoPara	±37.84 [0.03%]	±90.23 [0.75%]	±2.98 [0.06%]	±1.21 [0.03%]
MET_SoftTrk_Scale	±31.67 [0.03%]	±76.09 [0.63%]	±4.99 [0.09%]	±1.01 [0.03%]
JET_GroupedNP_3	±26.23 [0.02%]	±63.59 [0.53%]	±31.08 [0.58%]	±0.84 [0.02%]
JET_GroupedNP_2	±21.28 [0.02%]	±52.19 [0.43%]	±33.11 [0.62%]	±0.68 [0.02%]
leptonSF_muEff_syst	±2.40 [0.00%]	±2.74 [0.02%]	±33.03 [0.62%]	±0.08 [0.00%]
trigSF_syst	±1.02 [0.00%]	±0.06 [0.00%]	±27.55 [0.52%]	±0.03 [0.00%]
Zjets_unc_scale	–	–	±155.00 [2.9%]	–
Lumi	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]	±88.05 [2.2%]
Zjets_unc_reweighting	–	–	±51.52 [0.97%]	–
Zjets_unc_alphaS	–	–	±69.67 [1.3%]	–
ttbar_unc_radiation	–	±162.01 [1.3%]	–	–
ttbar_unc_PhH7	–	±323.50 [2.7%]	–	–
ttbar_unc_aMcAtNloPy8	–	±285.34 [2.4%]	–	–

Table 7.8: Breakdown of the dominant systematic uncertainties in the W CR. Note that the individual uncertainties can be correlated and do not necessarily add up quadratically to the total uncertainty. The percentages show the size of the uncertainty relative to the total expected yield.

Uncertainty in Estimate	$t\bar{t}$	$W \to \tau\nu$	$W \to \mu\nu$	$Z \to \mu\mu$
Total background expectation	114,983.48	150.69	4,745.30	828.85
Total statistical ($\sqrt{Nexp}$)	±339.09	±12.28	±68.89	±28.79
Total background systematic	±423.63 [0.37%]	±7.51 [4.99%]	±219.83 [4.63%]	±51.80 [6.25%]
btagSF_B	±232.39 [0.20%]	±3.07 [2.0%]	±96.30 [2.0%]	±15.17 [1.8%]
btagSF_C	±215.42 [0.19%]	±2.85 [1.9%]	±89.40 [1.9%]	±14.09 [1.7%]
btagSF_L	±197.78 [0.17%]	±2.62 [1.7%]	±82.25 [1.7%]	±12.96 [1.6%]
JET_JER_SINGLE_NP	±196.41 [0.17%]	±2.39 [1.6%]	±74.98 [1.6%]	±4.97 [0.60%]
MET_SoftTrk_ResoPerp	±195.64 [0.17%]	±2.62 [1.7%]	±81.99 [1.7%]	±12.91 [1.6%]
JET_GroupedNP_1	±160.02 [0.14%]	±2.32 [1.5%]	±72.68 [1.5%]	±2.81 [0.34%]
puWeight_syst	±126.52 [0.11%]	±1.72 [1.1%]	±53.86 [1.1%]	±2.42 [0.29%]
MUON_SAGITTA_RHO	±123.56 [0.11%]	±1.30 [0.86%]	±40.68 [0.86%]	±5.90 [0.71%]
MET_SoftTrk_ResoPara	±90.23 [0.08%]	±1.21 [0.80%]	±37.84 [0.80%]	±5.96 [0.72%]
MET_SoftTrk_Scale	±76.09 [0.07%]	±1.01 [0.67%]	±31.67 [0.67%]	±4.99 [0.60%]
JET_GroupedNP_3	±59.52 [0.05%]	±0.84 [0.56%]	±26.23 [0.55%]	±0.31 [0.04%]
leptonSF_muEff_syst	±34.54 [0.03%]	±0.08 [0.05%]	±2.40 [0.05%]	±5.57 [0.67%]
trigSF_syst	±25.65 [0.02%]	±0.03 [0.02%]	±1.02 [0.02%]	±4.47 [0.54%]
Zjets_unc_scale	–	–	–	±44.95 [5.4%]
Lumi	±0.00 [0.00%]	±3.32 [2.2%]	±0.00 [0.00%]	±0.00 [0.00%]
Zjets_unc_alphaS	–	–	–	±4.81 [0.58%]
mu_W	0.00 [0.00%]	±0.00 [0.00%]	±76.51 [1.6%]	±0.00 [0.00%]
Wjets_unc_scale	–	±0.00 [0.00%]	±205.11 [4.3%]	–
Zjets_unc_interPDF	–	–	–	±9.48 [1.1%]
Wjets_unc_alphaS	–	±0.00 [0.00%]	±42.24 [0.89%]	–

Table 7.9: Breakdown of the dominant systematic uncertainties in the $t\bar{t}$ CR in the muon channel. Note that the individual uncertainties can be correlated, and do not necessarily add up quadratically to the total uncertainty. The percentages show the size of the uncertainty relative to the total expected yield.

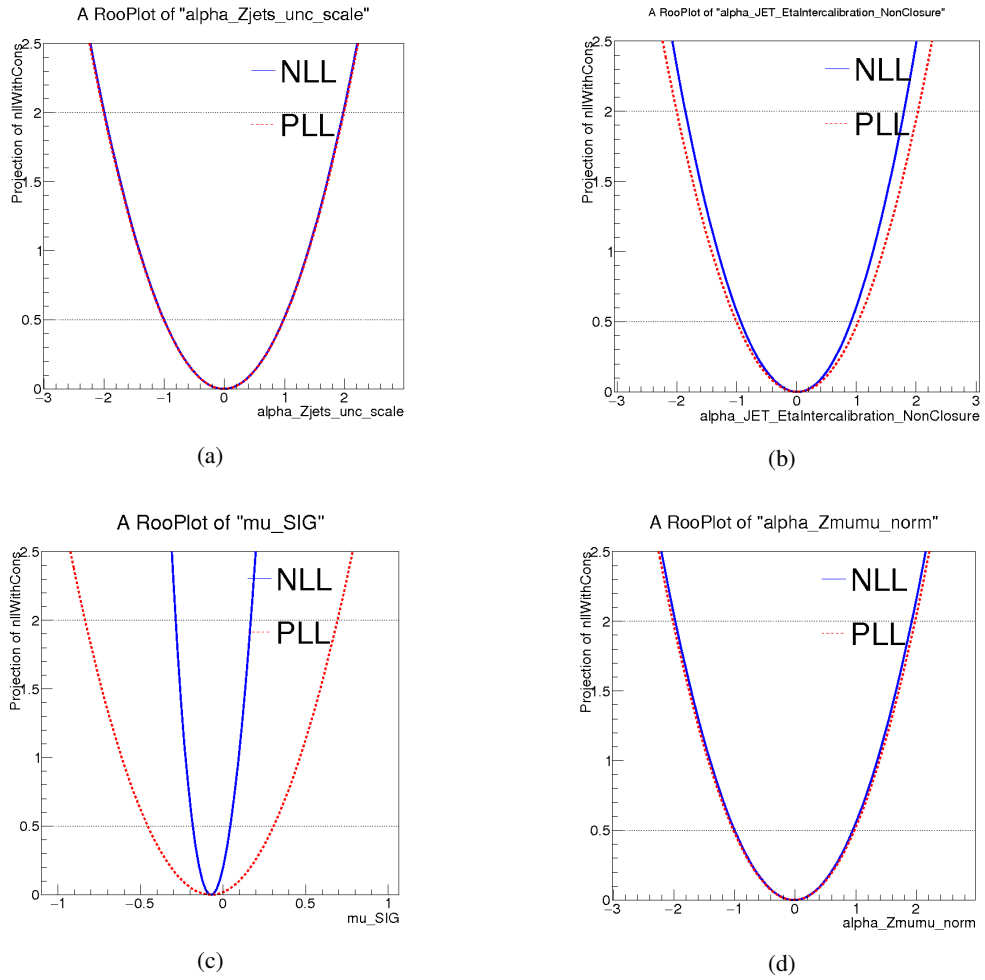


Figure 7.27: Examples of PLL curves for a leptoquark mass of 400 GeV. The curves are shown for Z+jets scale uncertainty parameter (top left), JES η-intercalibration non-closure parameter (top right), singal strength parameter (bottom left) and Z+jets normalisation factor (bottom right).

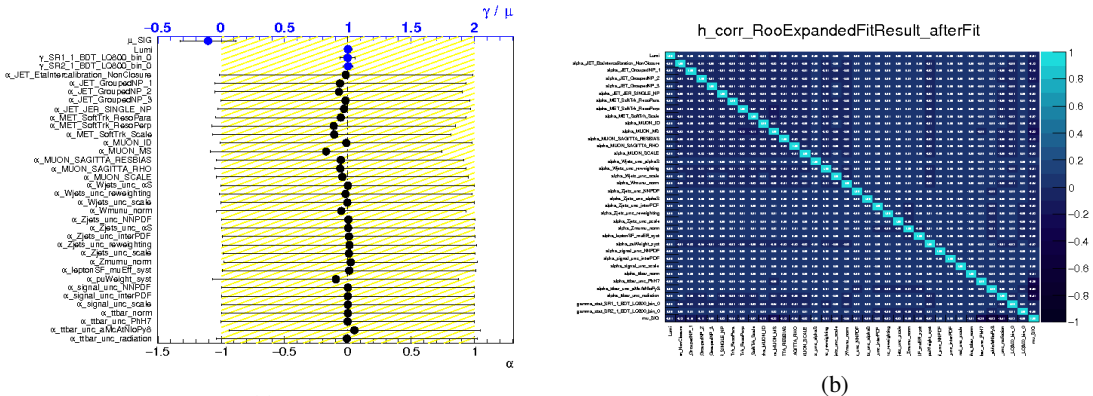
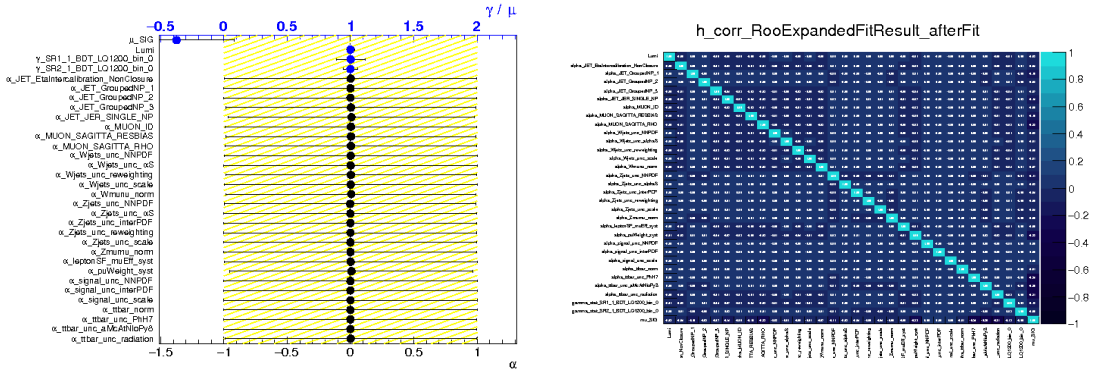


Figure 7.28: Post-fit parameters (left) and their correlations matrix (right) for exclusion fit for a leptoquark mass of 800 GeV.



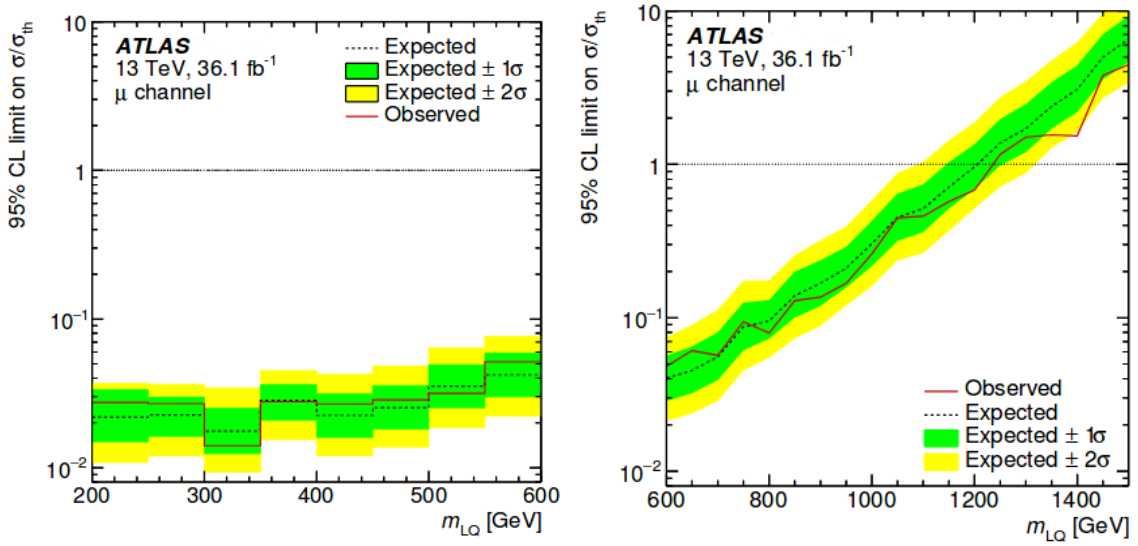
(a) (b)

Figure 7.29: Post-fit parameters (left) and their correlations matrix (right) for exclusion fit for a leptoquark mass of 1200 GeV.

Tables with the pre- and post-fit event yields in all signal regions are collected in App. B.11.

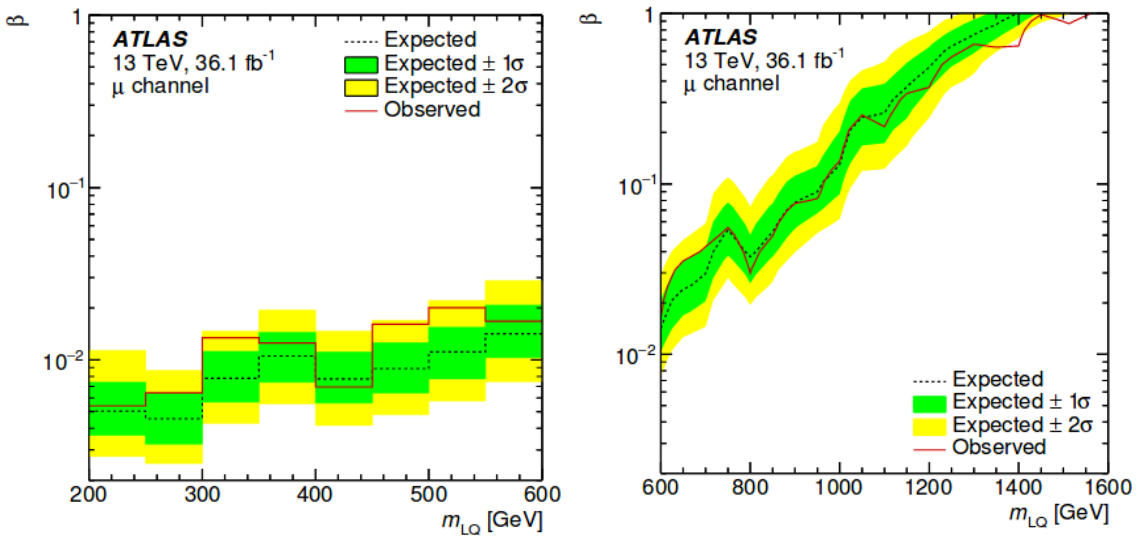
7.12.3 Results and conclusions

The results are translated into exclusion bounds on the signal strength as a function of the LQ mass. The 95% CL scalar LQ production cross-section limit of combined dimuon channel for 200 GeV $\leq m_{LQ} \leq$ 600 GeV is shown in Fig. 7.30a and for 600 GeV $\leq m_{LQ} \leq$ 1500 GeV in Fig. 7.30b. There the excluded LQ mass range extends to ∼1235 GeV. Fig. 7.31a presents β-m_{LQ} exclusion plane for 200 GeV $\leq m_{LQ} \leq$ 600 GeV and Fig. 7.31b for 600 GeV $\leq m_{LQ} \leq$ 1500 GeV. Again there is not any potential signal excess indication.



(a) (b)

Figure 7.30: (a) Cross-section limits as a function of LQ mass for 200 GeV $\leq m_{LQ} \leq$ 600 GeV. (b) Cross-section limits as a function of LQ mass for 600 GeV $\leq m_{LQ} \leq$ 1500 GeV.



(a) (b)

Figure 7.31: (a) Excluded LQ mass as a function of branching ratio β for 200 GeV $\leq$ m$_{LQ}$ $\leq$ 600 GeV. (b) Excluded LQ mass as a function of branching ratio β for 600 GeV $\leq$ m$_{LQ}$ $\leq$ 1500 GeV.

Appendices

Appendix A

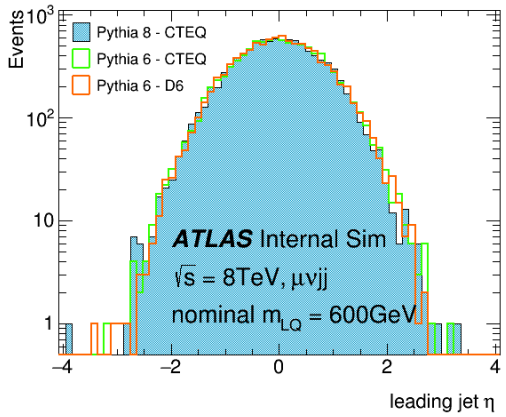
Supplementary Material of 8 TeV Analysis

In this appendix more supplementary material of the 8 TeV, Run I LQ analysis is presented in addition to these of Ch. 6.

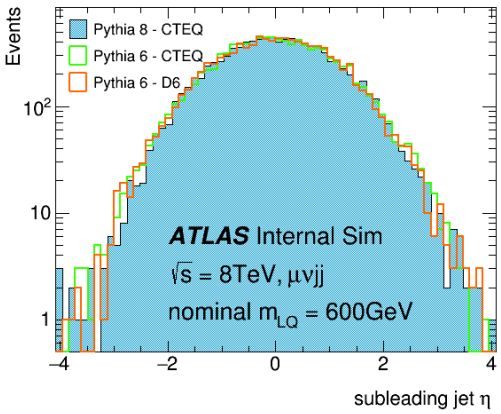
A.1 Different Pythia generator versions comparisons

Below appear η and ϕ distributions produced with Pythia 6 (CTEQ and D6 PDF sets respectively) and Pythia 8 CTEQ in truth and parton level. Especially in parton level, both FSR and no FSR cases have been studied.

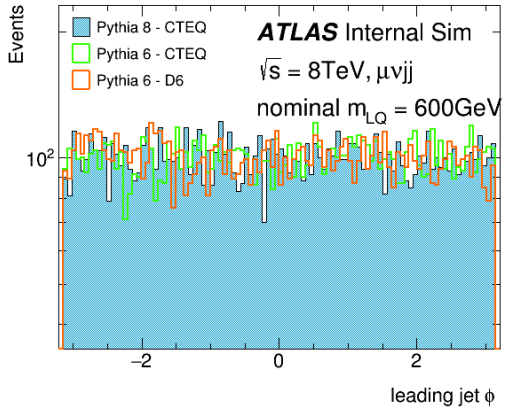
A.1.1 Truth level distributions



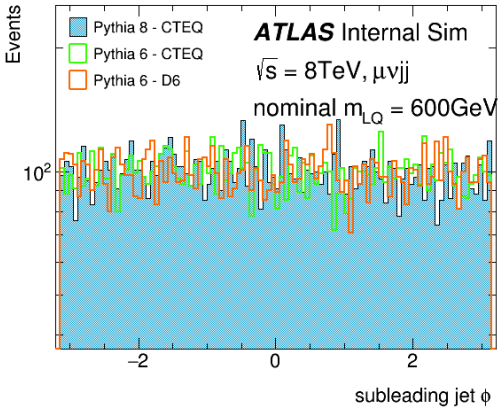
(a)



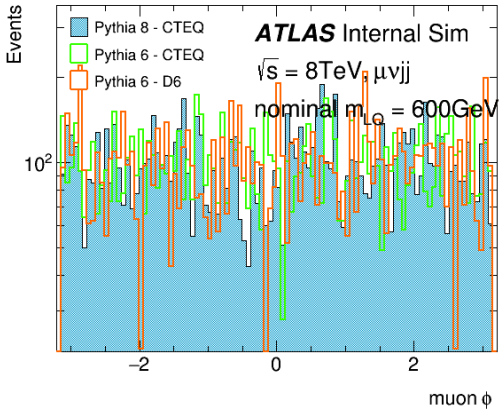
(b)



(c)



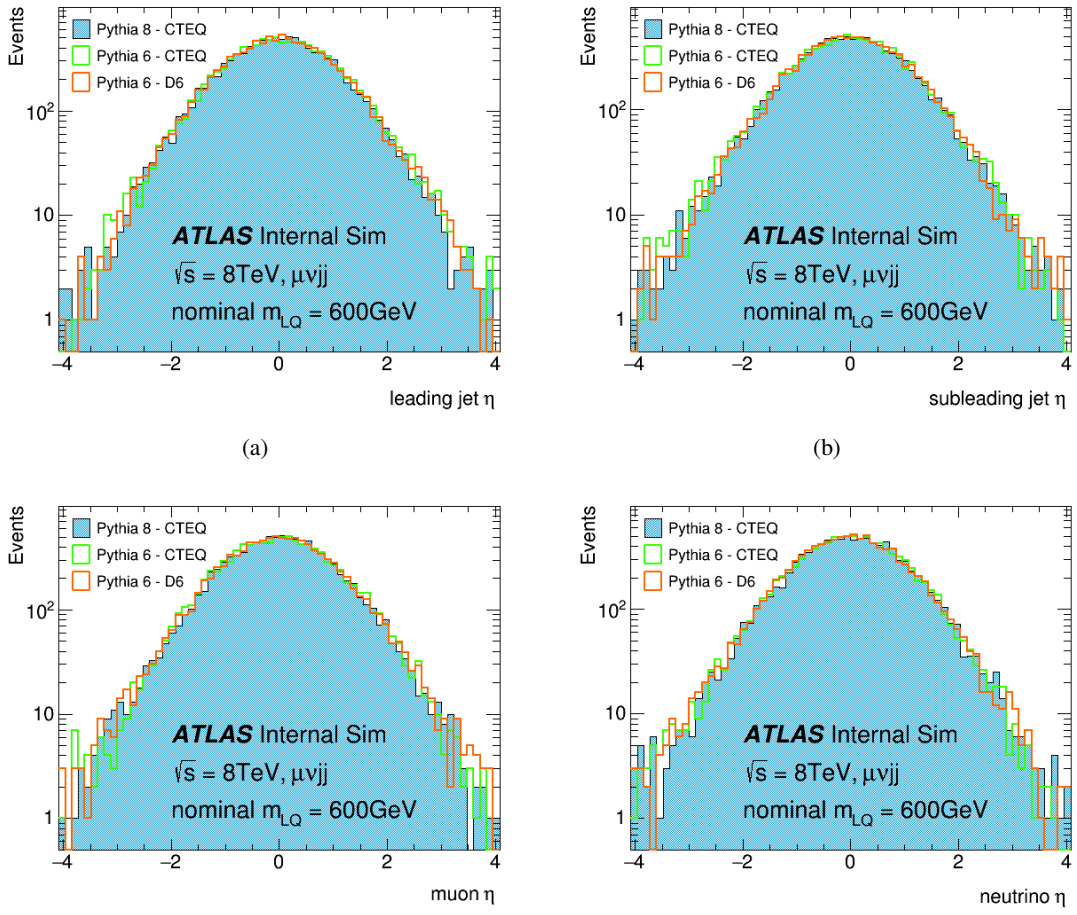
(d)



(e)

Figure A.1: The two truth level (a) leading and (b) subleading quarks η, (c) leading and (d) subleading quarks ϕ and (e) muon ϕ distributions produced with Pythia 6 - D6, Pythia 6 - CTEQ and Pythia 8 versions.

A.1.2 Parton level distributions - no final state radiation



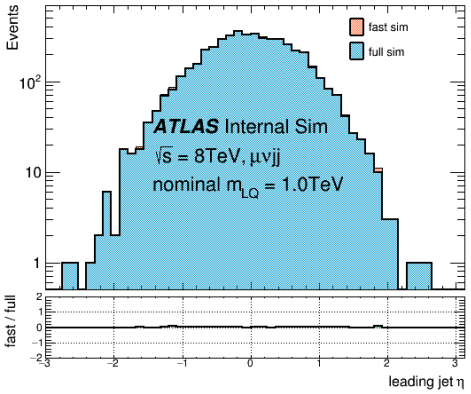
(a) (b)

(c) (d)

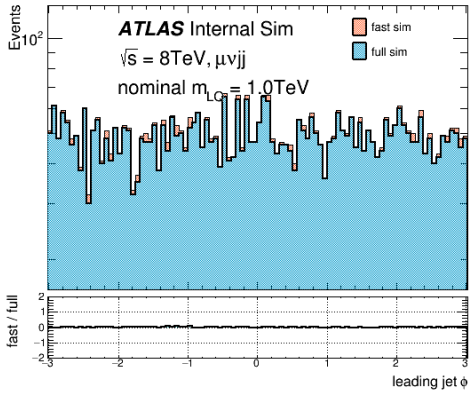
Figure A.2: The two parton level (a) leading and (b) subleading jets, (c) muon and (d) neutrino η distributions produced with Pythia 6 - D6, Pythia 6 - CTEQ and Pythia 8 versions.

A.2 Fast - full simulation comparisons

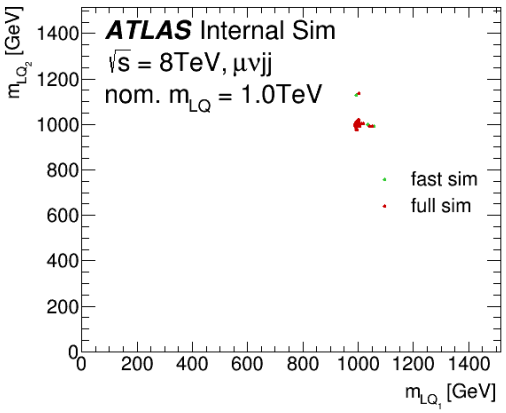
A.2.1 Truth level distributions



(a)



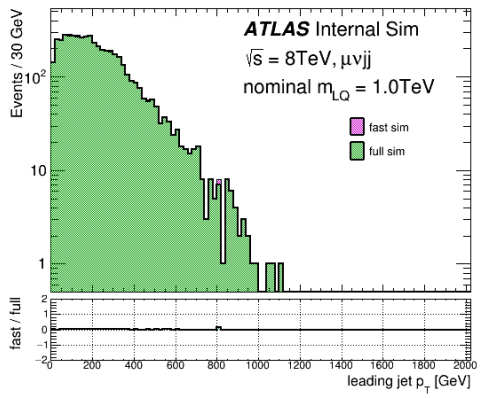
(b)



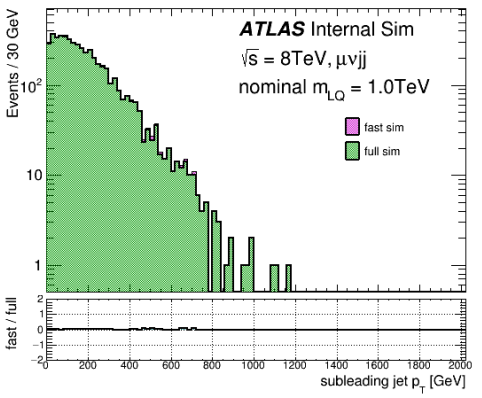
(c)

Figure A.3: Truth level (a) leading quark η and (b) ϕ distributions and (c) m_{LQ_1} vs m_{LQ_2} fast and full simulation.

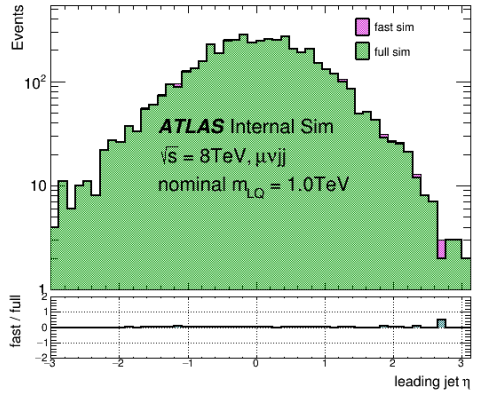
A.2.2 Parton level distributions - final state radiation



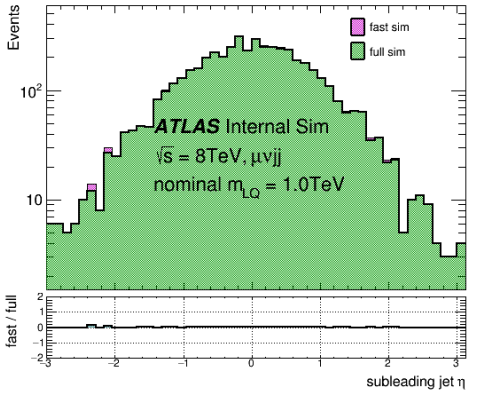
(a)



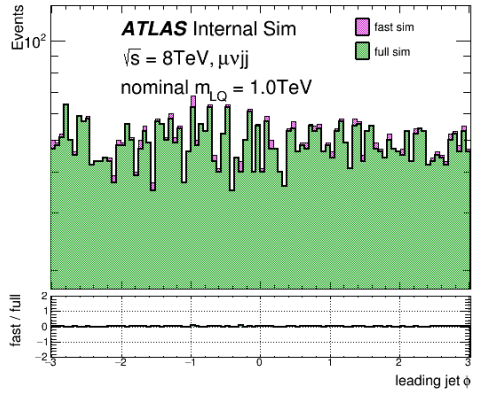
(b)



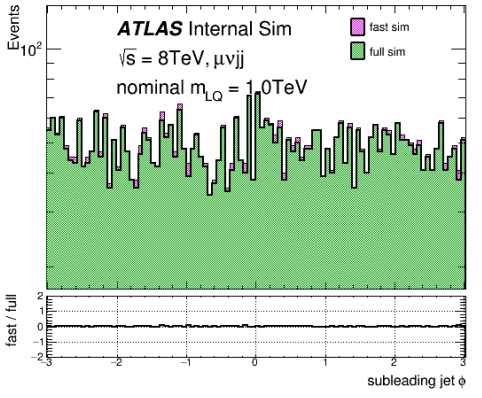
(c)



(d)



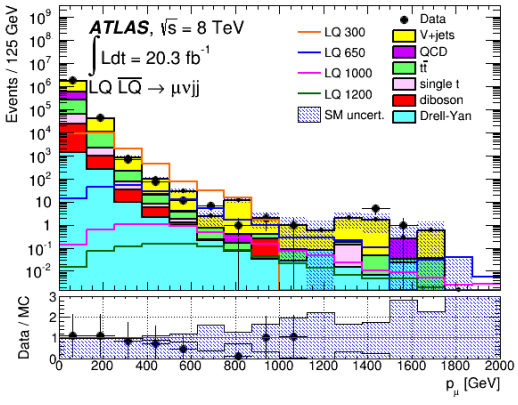
(e)



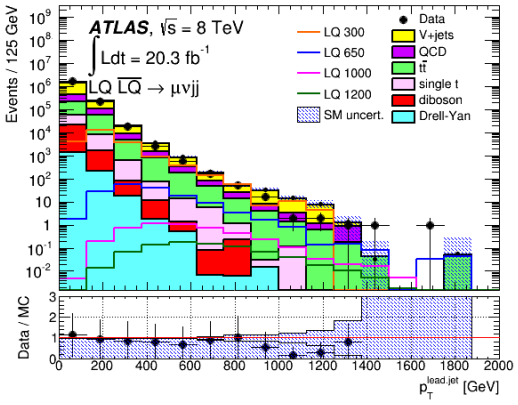
(f)

Figure A.4: Parton level leading and subleading jet (a) - (b) p_T, (c) - (d) η and (e) - (f) ϕ distributions in fast and full simulation mode. Gluon FSR corrections are included.

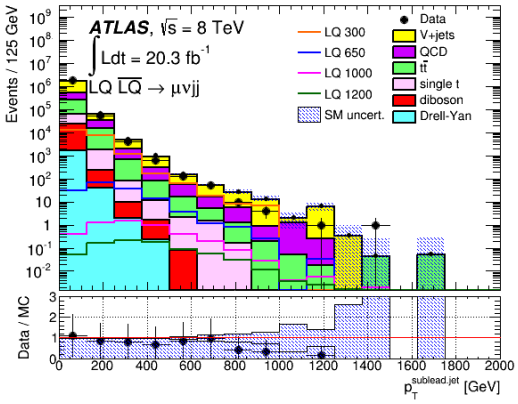
A.3 Pre - selection histograms



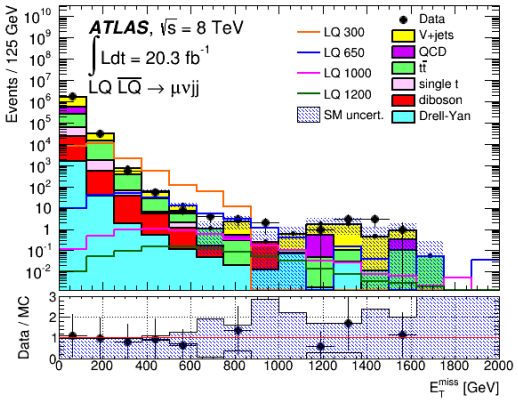
(a)



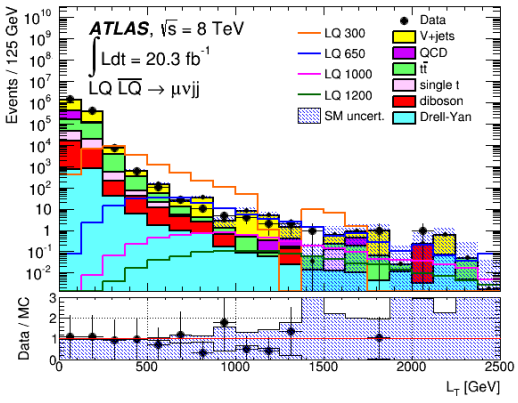
(b)



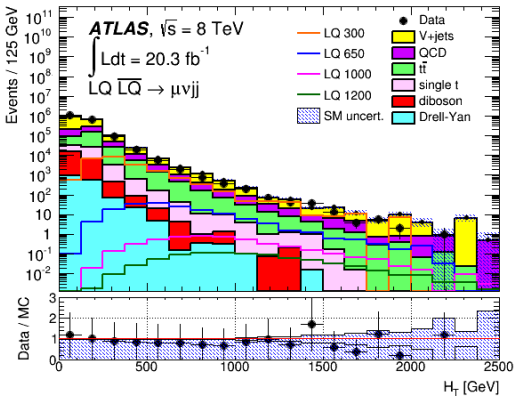
(c)



(d)

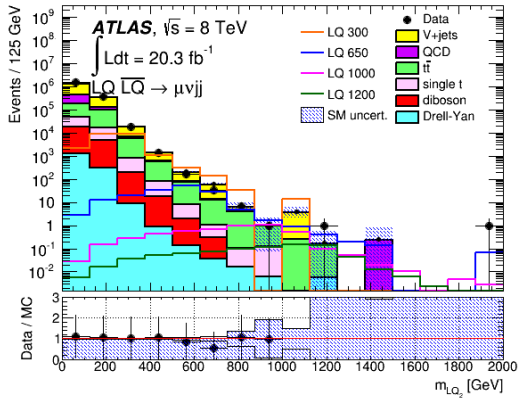


(e)

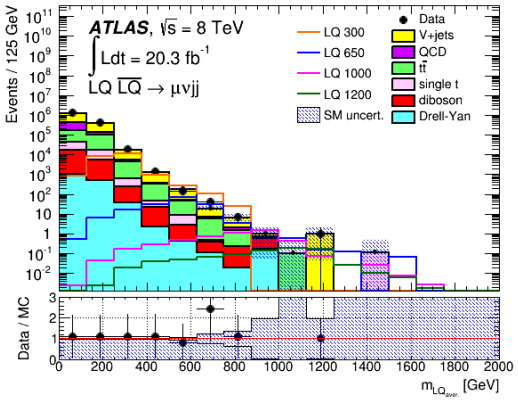


(f)

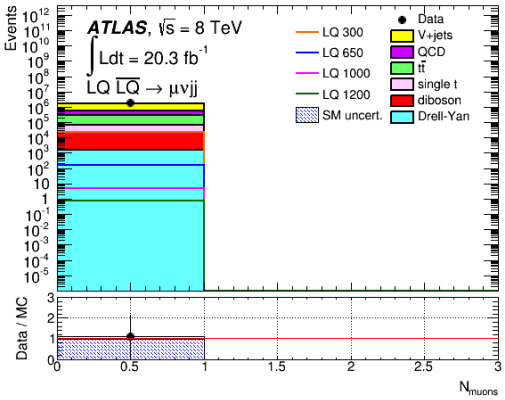
Figure A.5: Distributions at the end of pre - selection.



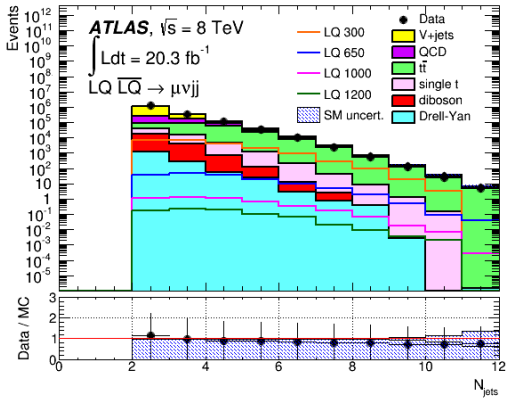
(a)



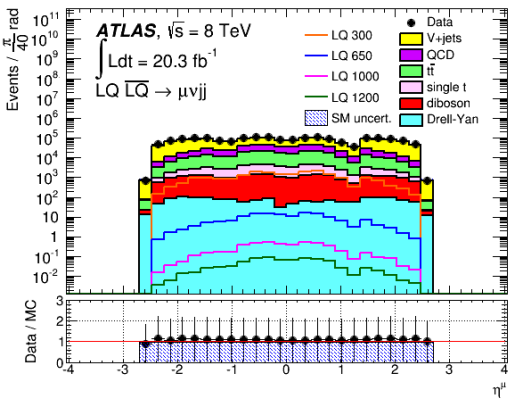
(b)



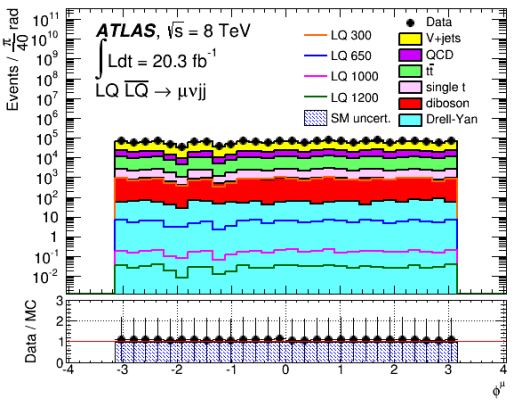
(c)



(d)

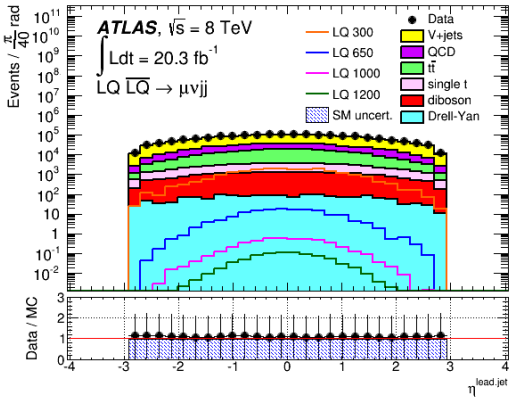


(e)

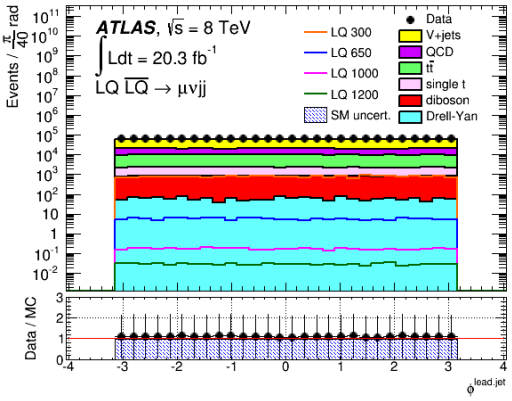


(f)

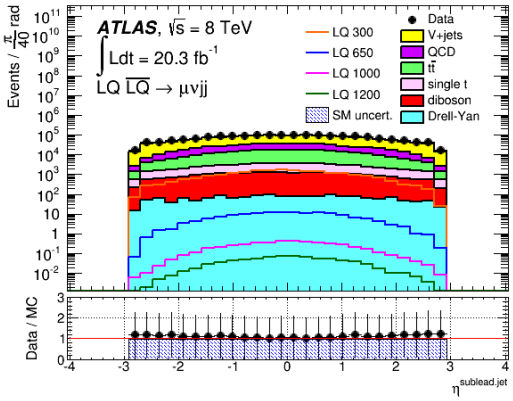
Figure A.6: Distributions at the end of pre - selection.



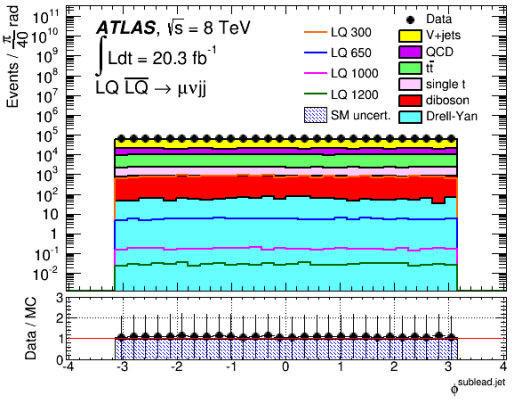
(a)



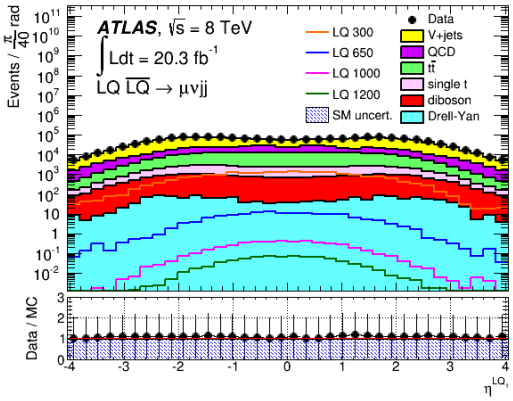
(b)



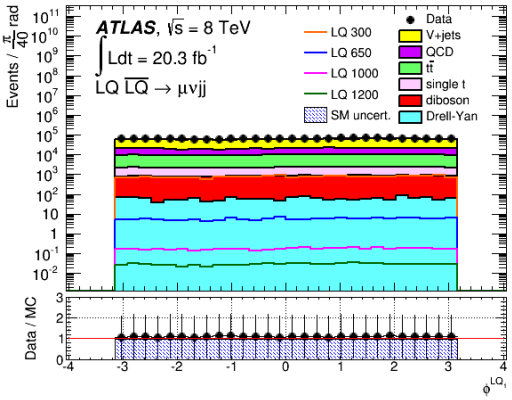
(c)



(d)



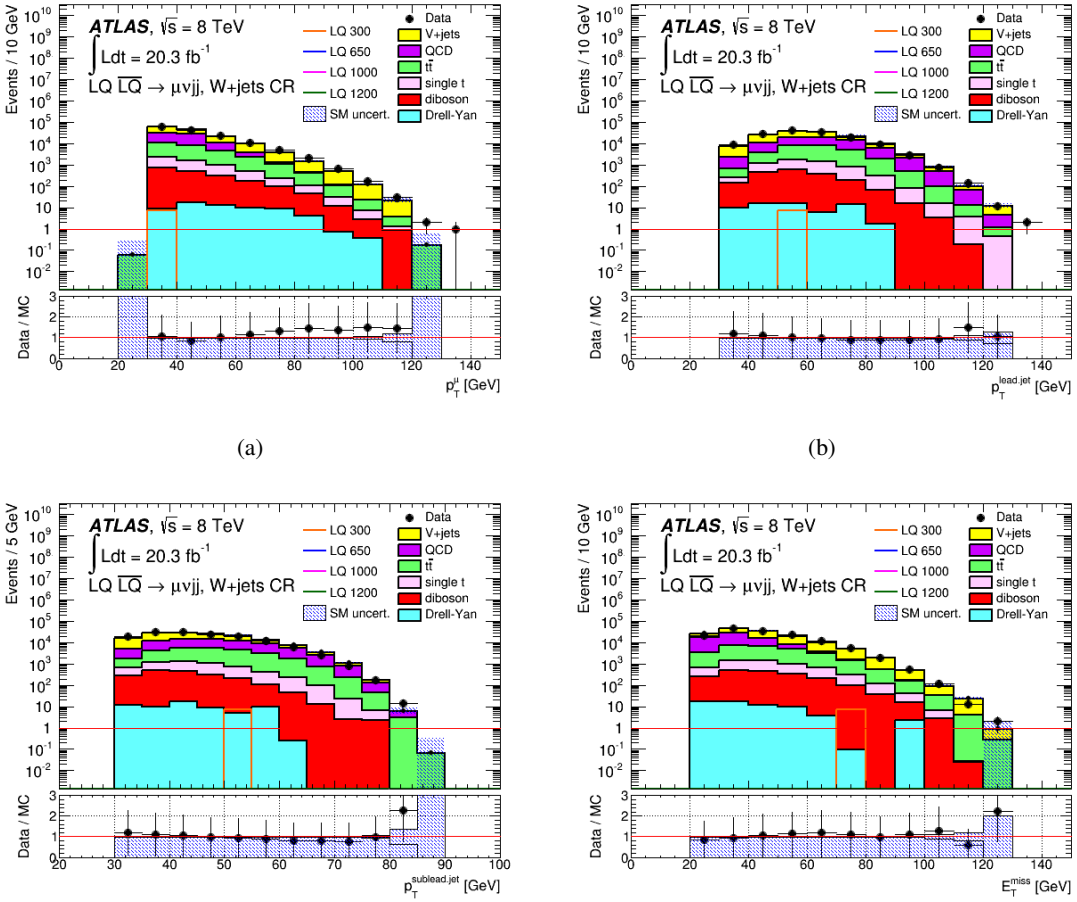
(e)



(f)

Figure A.7: Distributions at the end of pre - selection.

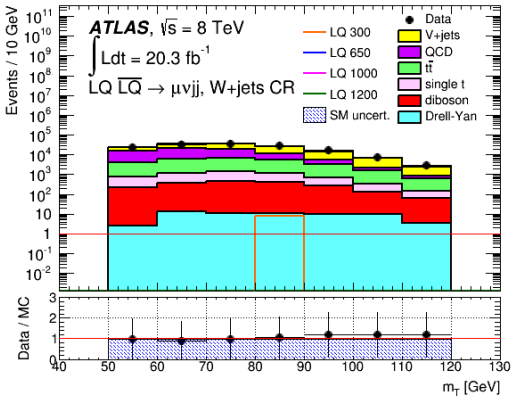
A.4 W+jets (W decaying hadronically) control region histograms



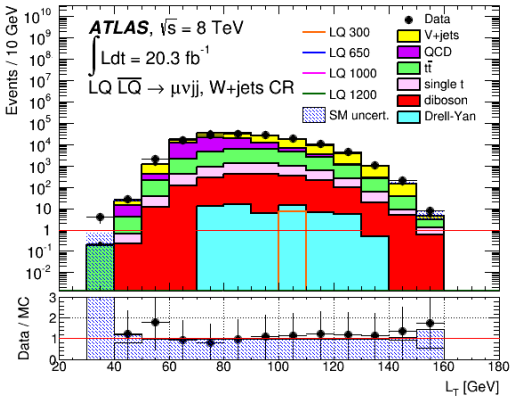
(a) (b)

(c) (d)

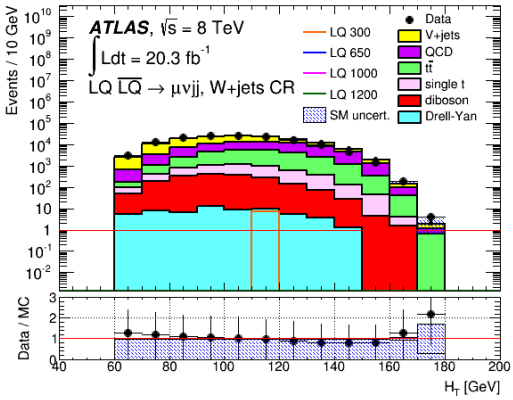
Figure A.8: Distributions at W+jets CR (where W boson decays hadronically).



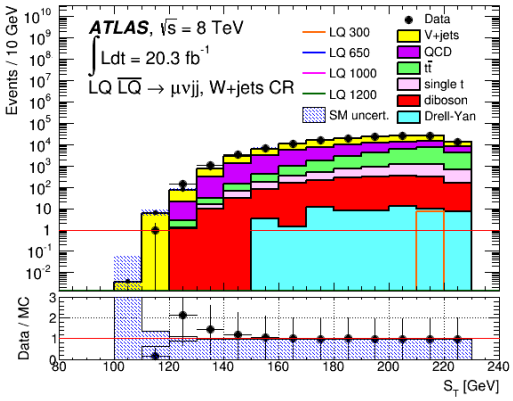
(a)



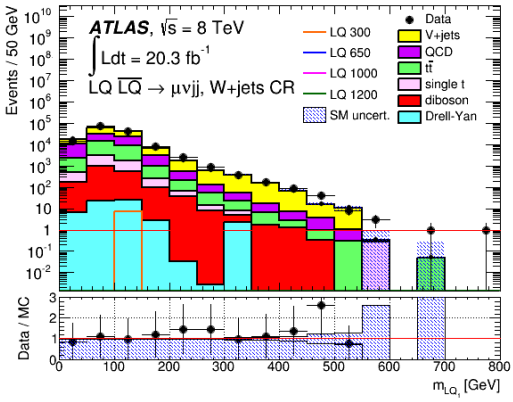
(b)



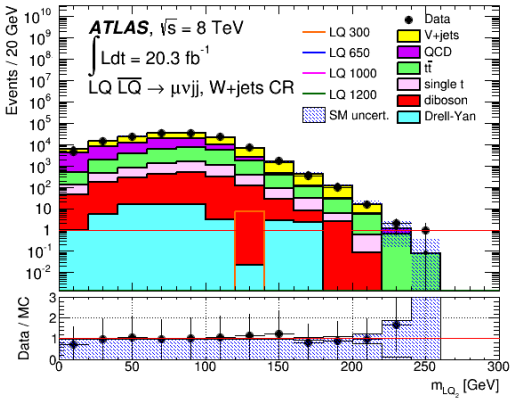
(c)



(d)

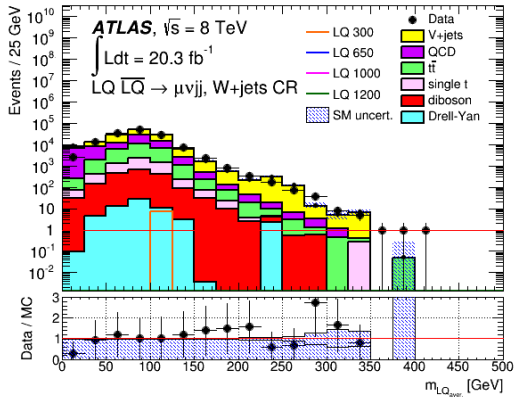


(e)

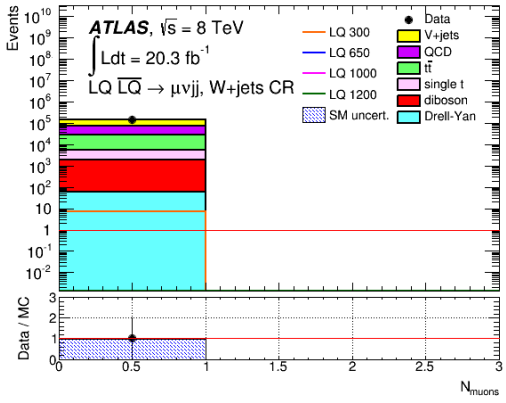


(f)

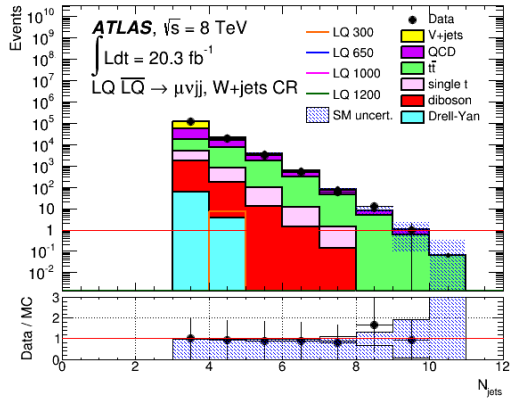
Figure A.9: Distributions at W+jets CR (where W boson decays hadronically).



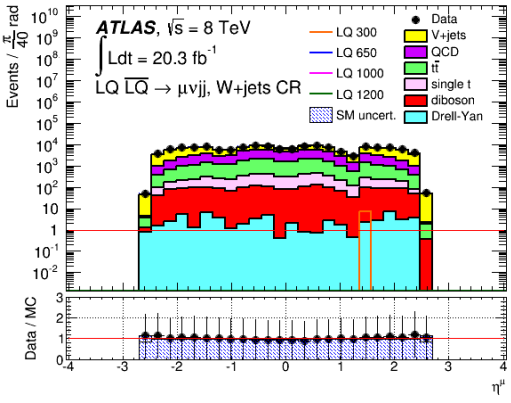
(a)



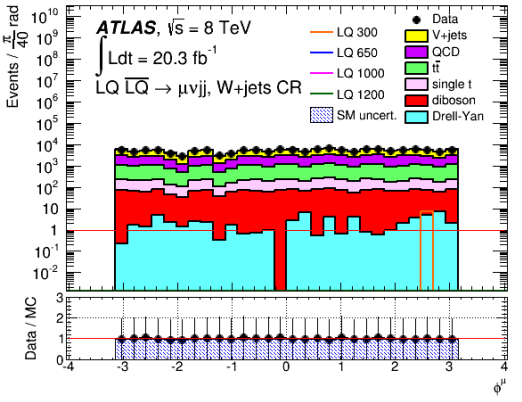
(b)



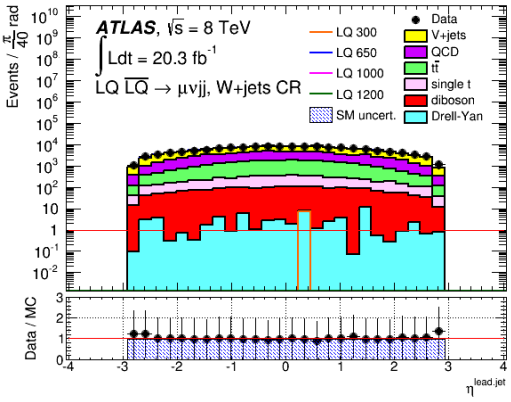
(c)



(d)

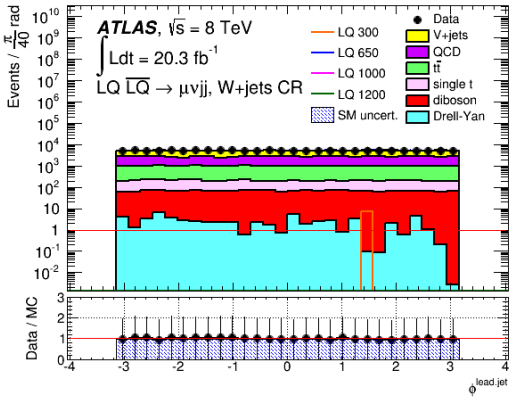


(e)

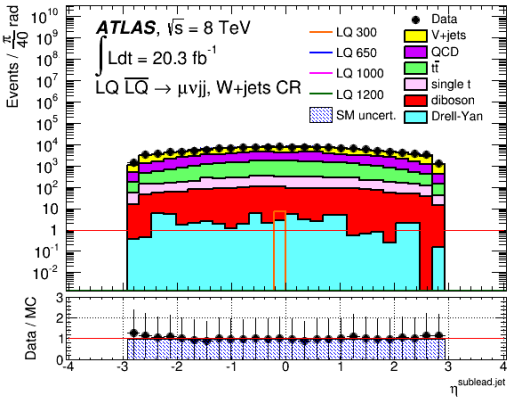


(f)

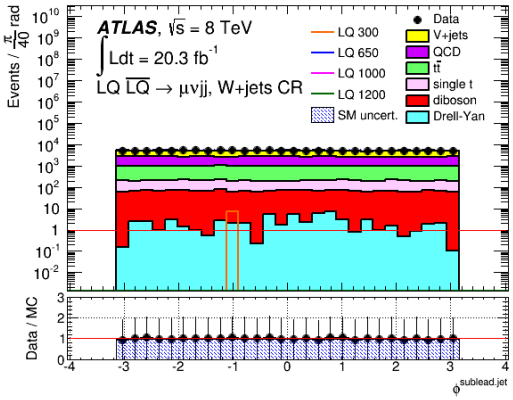
Figure A.10: Distributions at W+jets CR (where W boson decays hadronically).



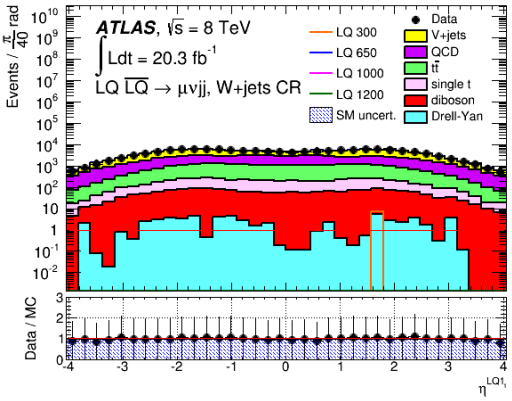
(a)



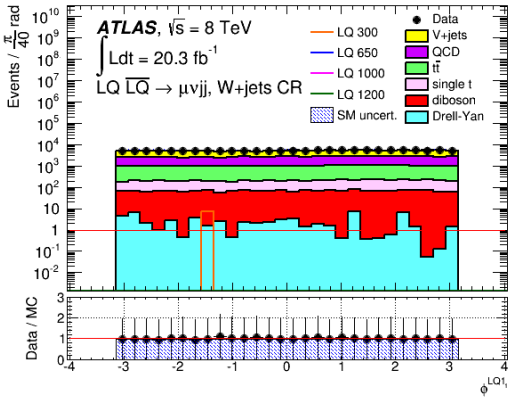
(b)



(c)



(d)



(e)

Figure A.11: Distributions at W+jets CR (where W boson decays hadronically).

A.5 Uncorrelated variables hypothesis - Significance signal plots

At this point there are the plots depicting the values of m_T, S_T and m_{LQ_1} variables under the uncorrelated variables hypothesis which maximize the LQ signal significance for different SRs.

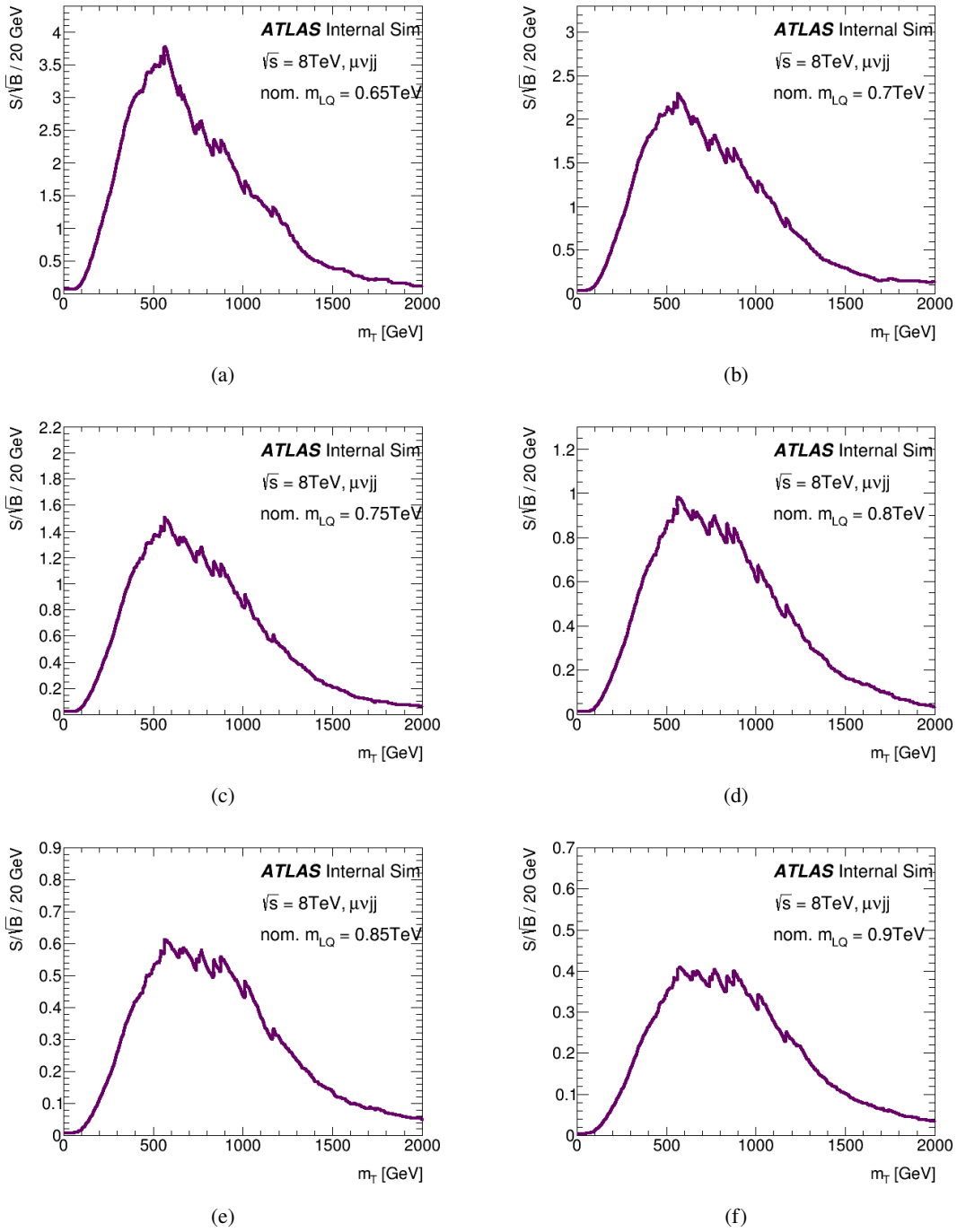
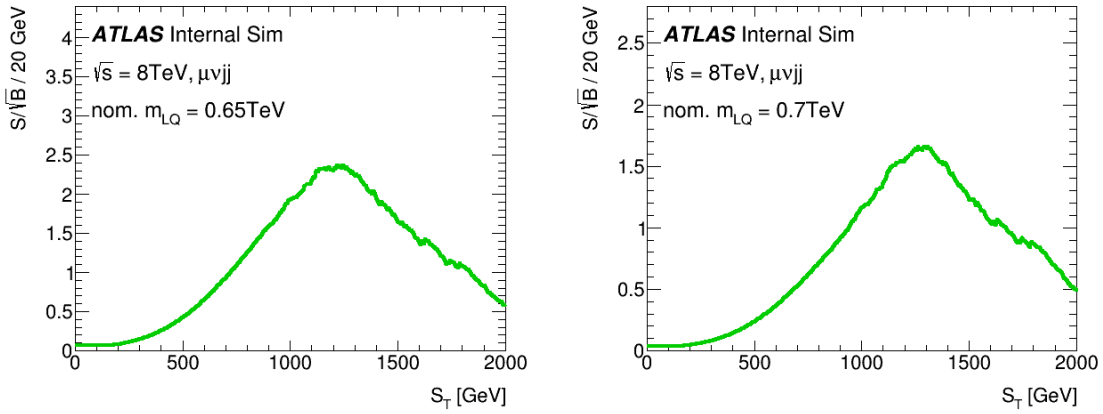
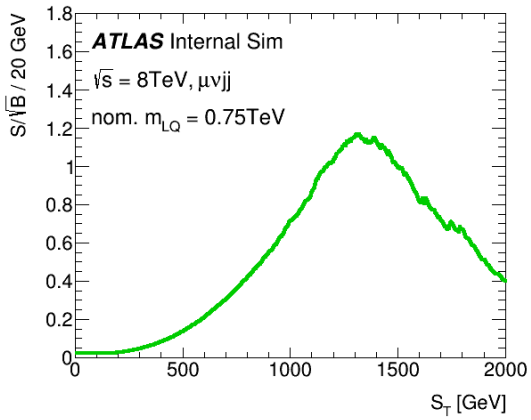


Figure A.12: Significance signal plots on m_T variable at (a) 650 GeV, (b) 700 GeV, (c) 750 GeV, (d) 800 GeV, (e) 850 GeV and (f) 900 GeV.

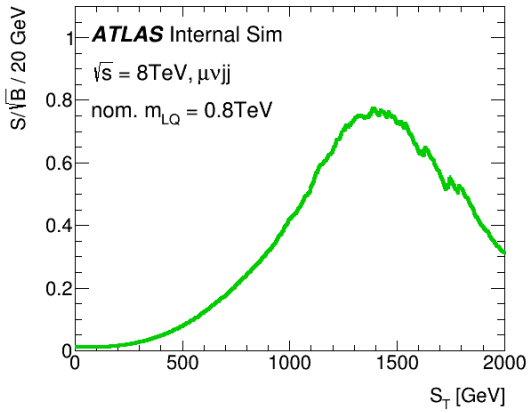


(a)

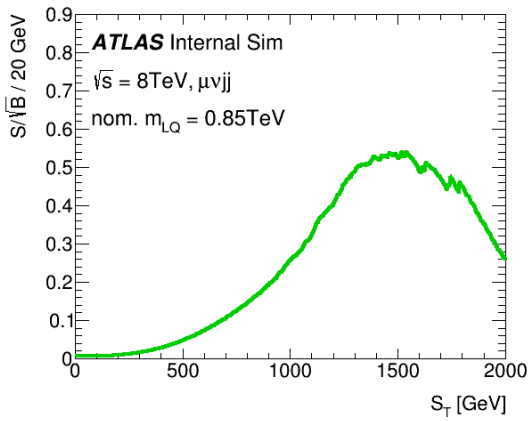
(b)



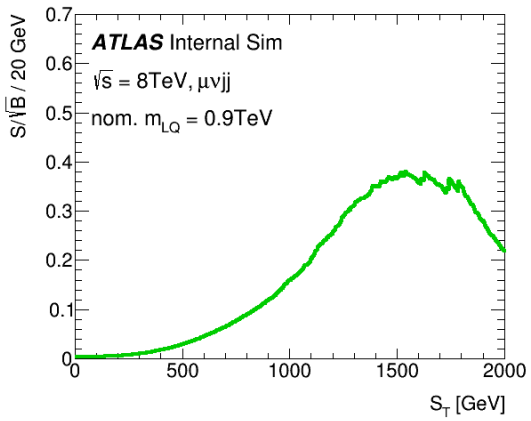
(c)



(d)



(e)



(f)

Figure A.13: Significance signal plots on S_T variable at (a) 650 GeV, (b) 700 GeV, (c) 750 GeV, (d) 800 GeV, (e) 850 GeV and (f) 900 GeV.

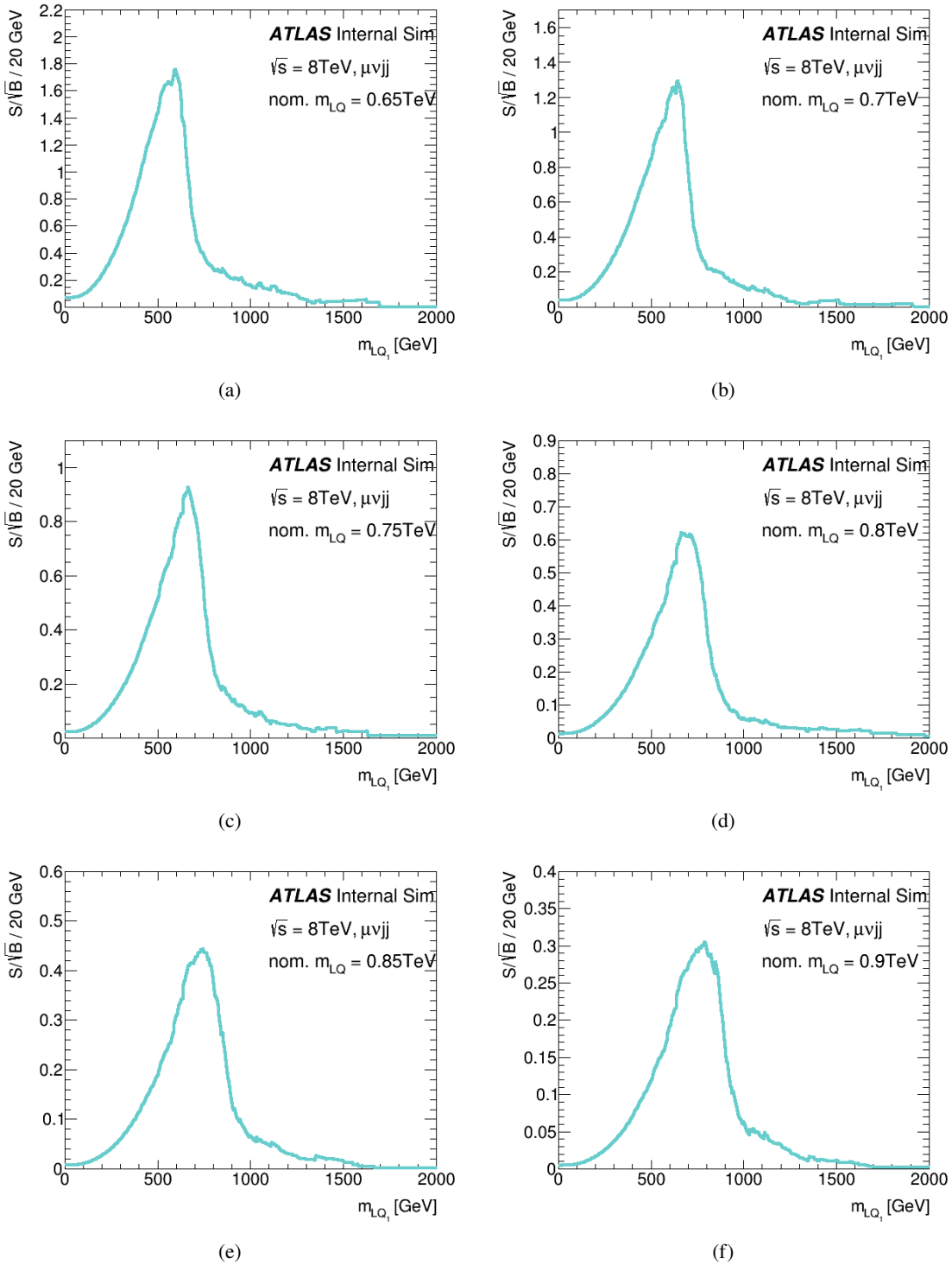


Figure A.14: Significance signal plots on m_{LQ_1} variable at (a) 650 GeV, (b) 700 GeV, (c) 750 GeV, (d) 800 GeV, (e) 850 GeV and (f) 900 GeV.

A.6 Uncorrelated variables hypothesis - Efficiency signal plots

Here appear the plots with the efficiency values of m_T, S_T and m_{LQ_1} variables corresponding to the previously acquired significance values of each SR from Sec. A.5.

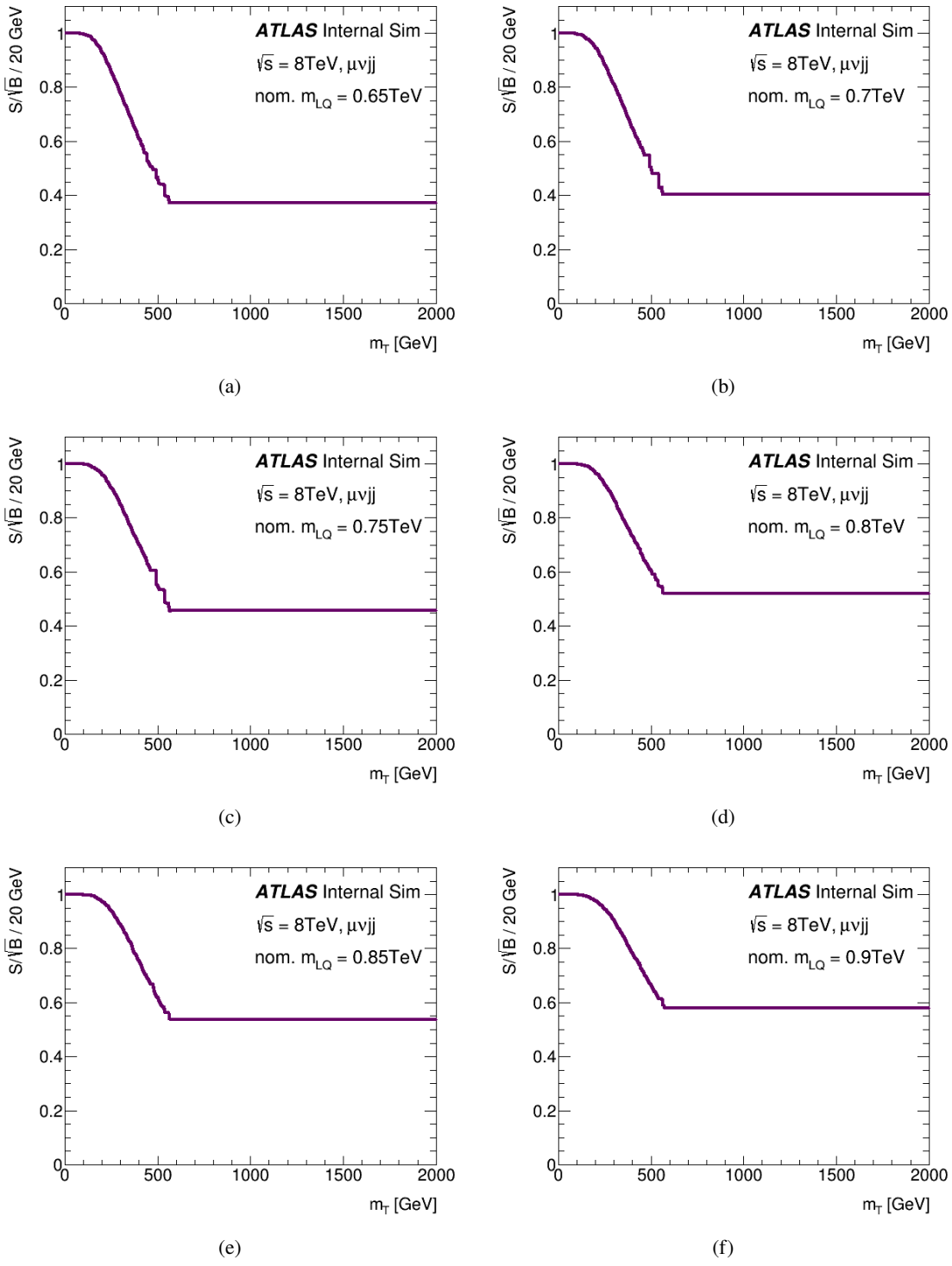
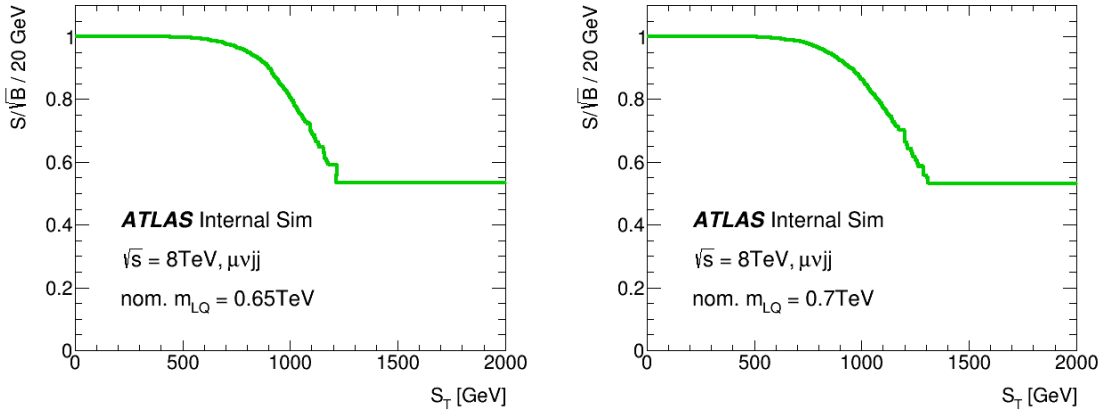
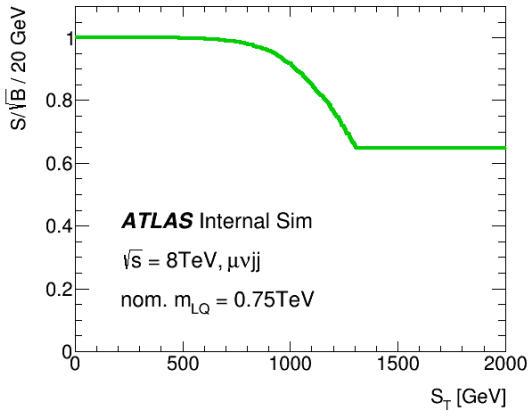


Figure A.15: Efficiency signal plots of m_T variable at (a) 650 GeV, (b) 700 GeV, (c) 750 GeV, (d) 800 GeV, (e) 850 GeV and (f) 900 GeV.

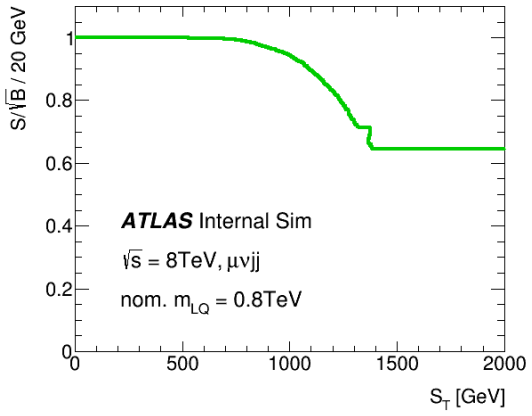


(a)

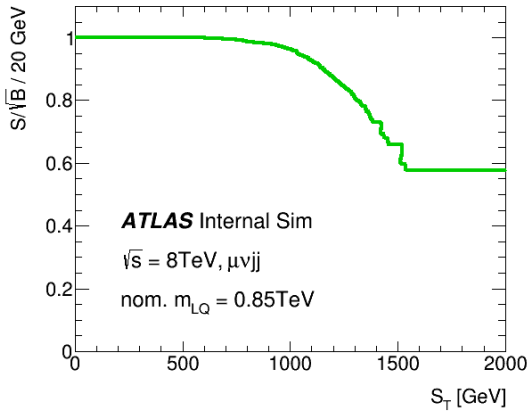
(b)



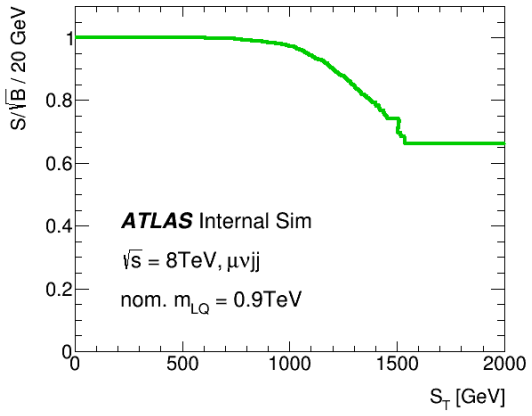
(c)



(d)



(e)



(f)

Figure A.16: Efficiency signal plots of S_T variable at (a) 650 GeV, (b) 700 GeV, (c) 750 GeV, (d) 800 GeV, (e) 850 GeV and (f) 900 GeV.

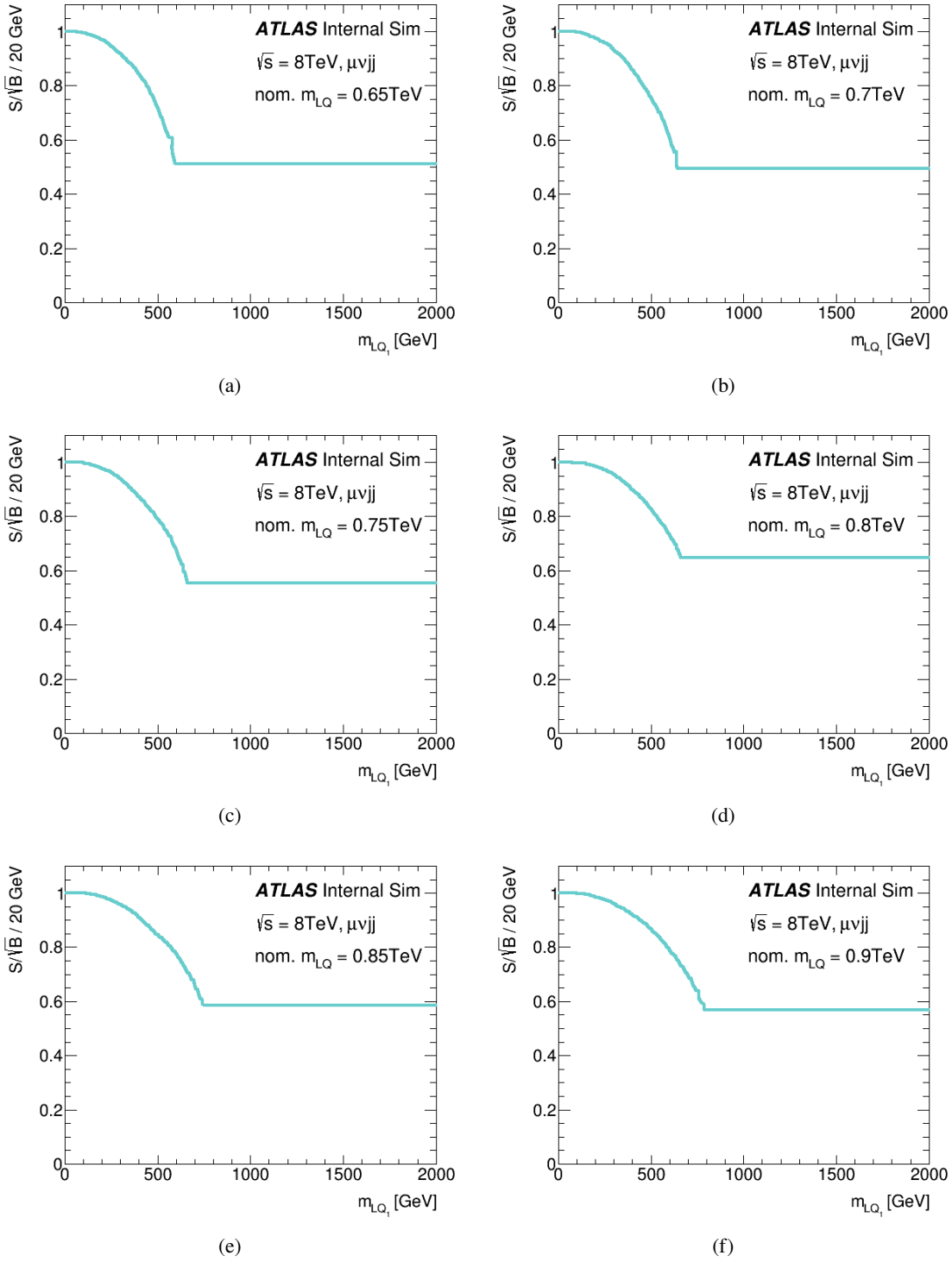
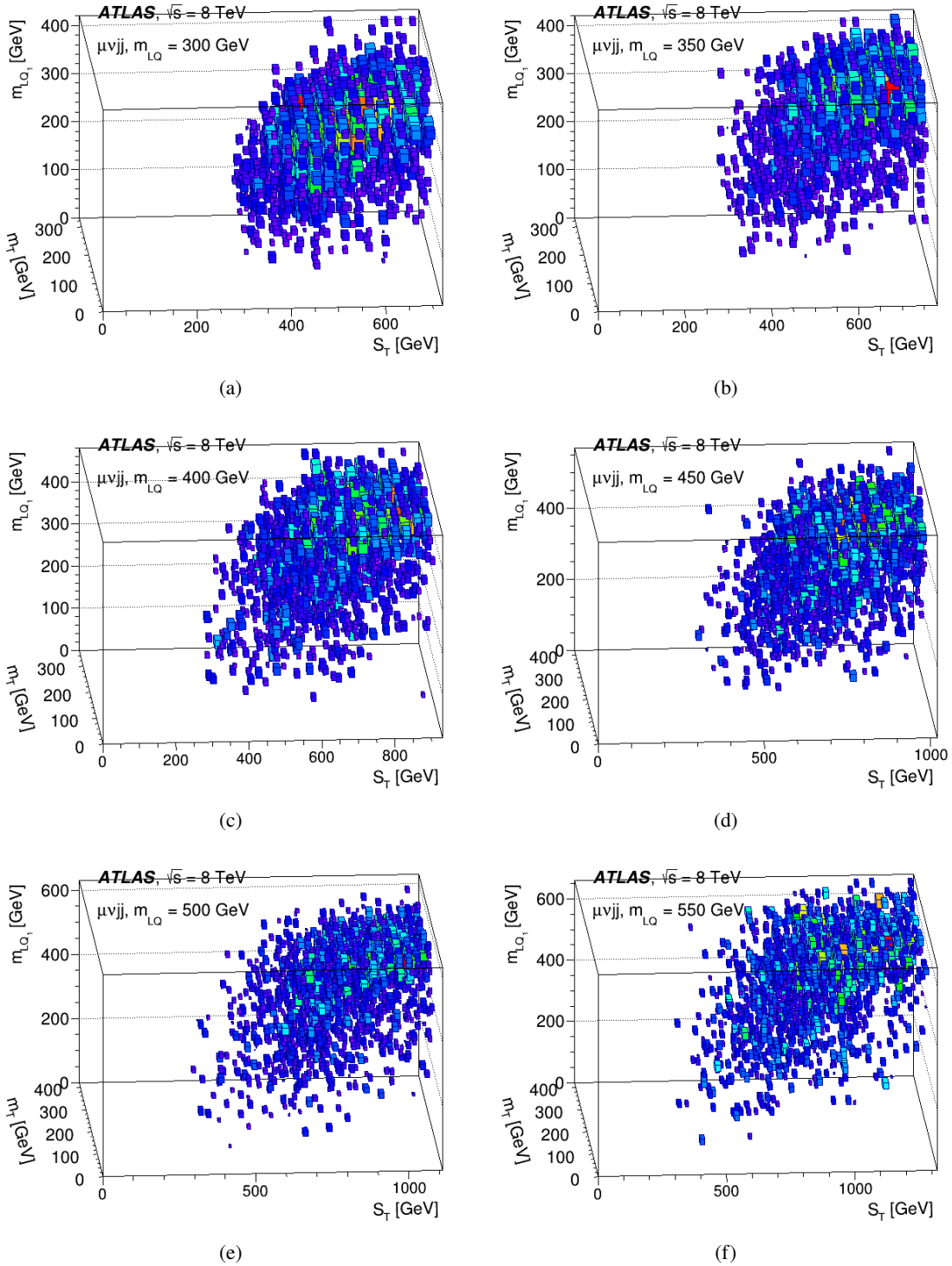


Figure A.17: Efficiency signal plots of m_{LQ_1} variable at (a) 650 GeV, (b) 700 GeV, (c) 750 GeV, (d) 800 GeV, (e) 850 GeV and (f) 900 GeV.

A.7 Correlated variables hypothesis

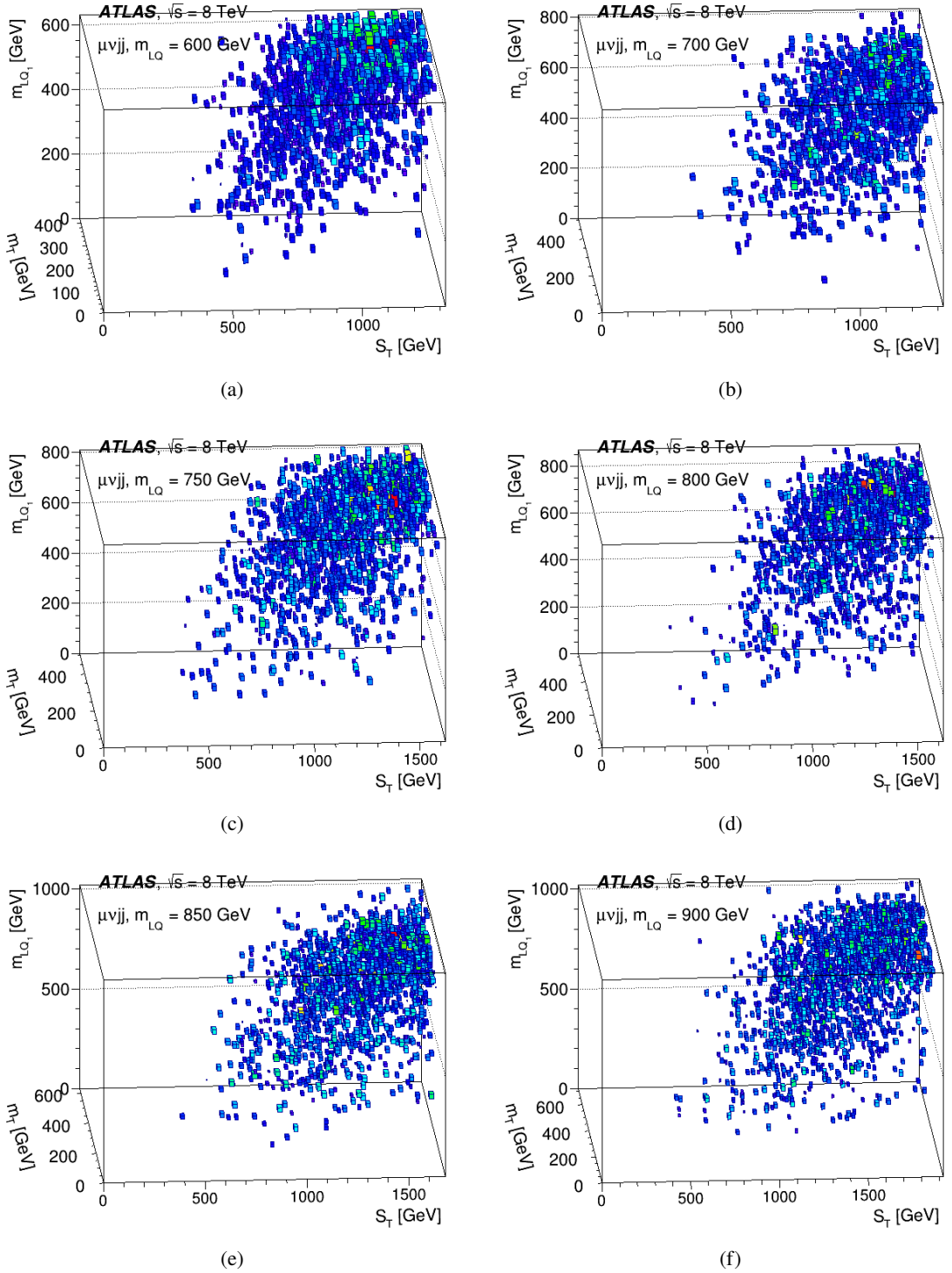
The following 3D plots depict the simultaneous m_T, S_T and m_{LQ_1} variables LQ signal significance maximization under the correlated variables hypothesis for all

SRs.



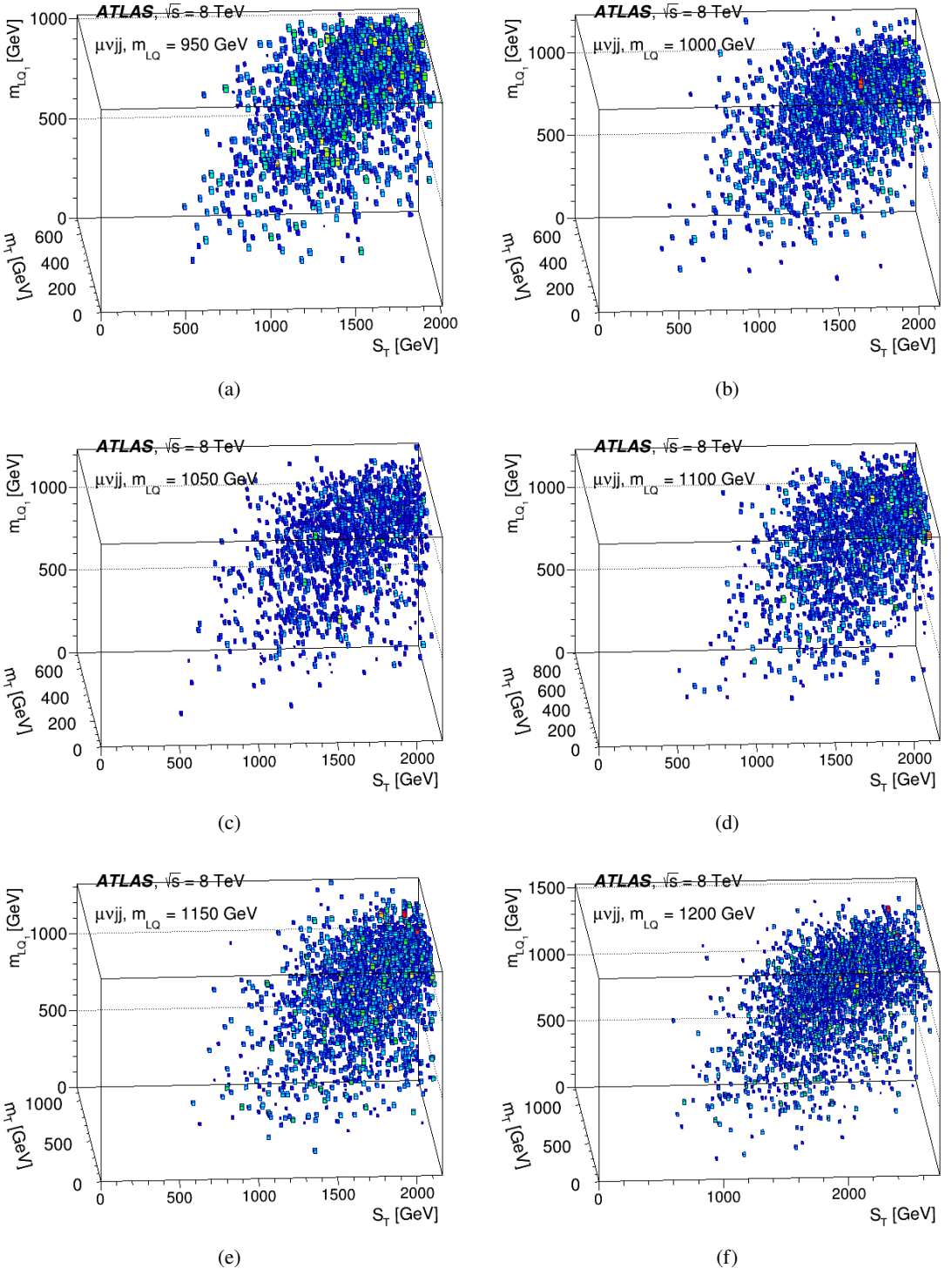
(a) (b) (c) (d) (e) (f)

Figure A.18: S_T, m_T and m_{LQ_1} simultaneous optimization 3-D histogram for different signal mass points. The red hot spot denotes the best values maximizing significant.



(a) (b) (c) (d) (e) (f)

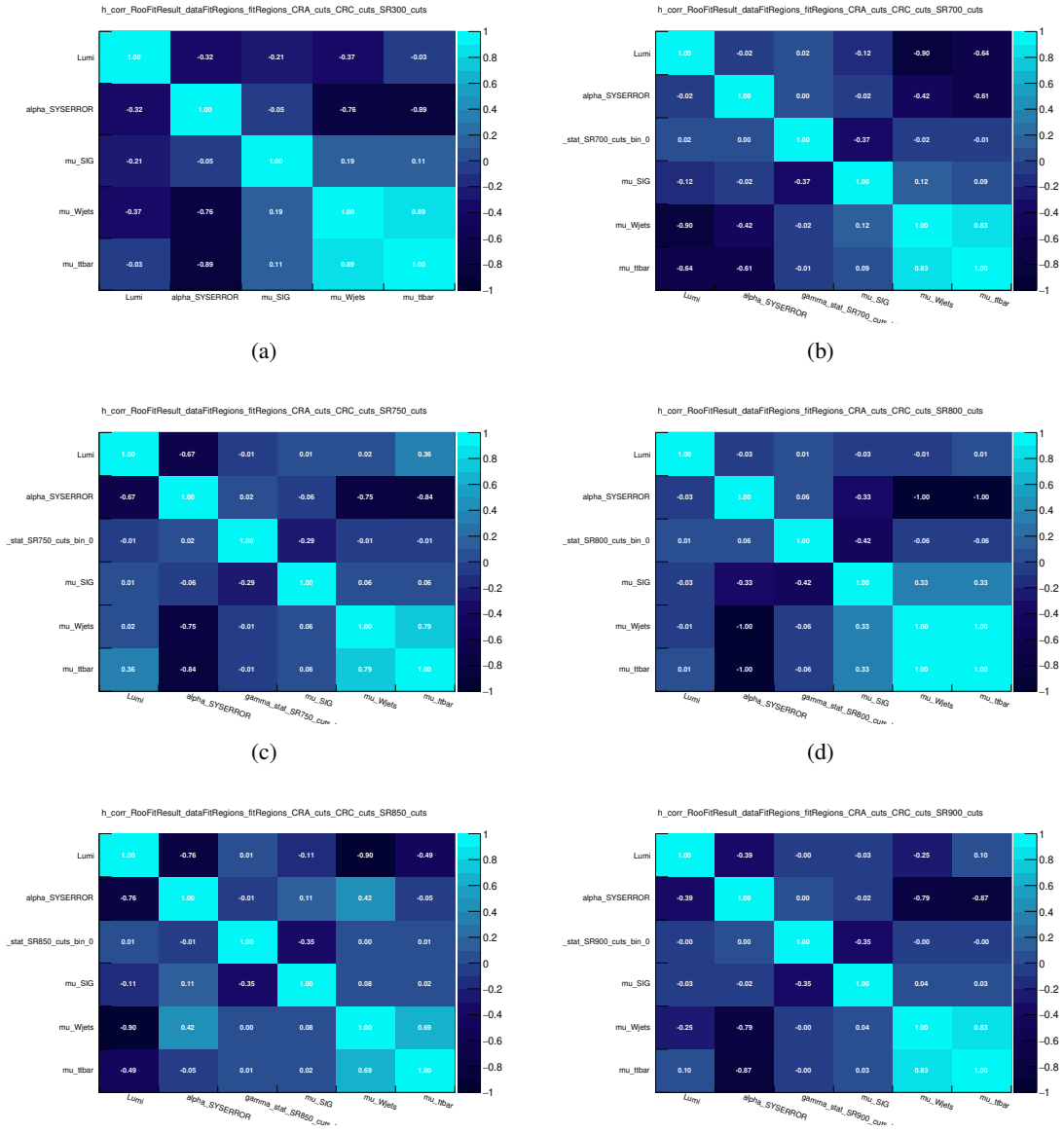
Figure A.19: S_T, m_T and m_{LQ_1} simultaneous optimization 3-D histogram for different signal mass points. The red hot spot denotes the best values maximizing significant.



(a) (b) (c) (d) (e) (f)

Figure A.20: S_T, m_T and m_{LQ_1} simultaneous optimization 3-D histogram for different signal mass points. The red hot spot denotes the best values maximizing significant.

A.8 Event yields after statistical fitting



(a) (b) (c) (d) (e) (f)

Figure A.21: Correlation matrices for (a) SR1 at 300 GeV, (b) SR9 at 700 GeV, (c) SR10 at 750 GeV, (d) SR11 at 800 GeV, (e) SR12 at 850 GeV and (f) SR13 at 900 GeV respectively.

Event yields	SR1 [300 GeV]	W+jets [CR A]	$t\bar{t}$ [CR C]
Observed events	1,732	1,205,164	221,094
Fitted events	$1,733.31 \pm 70.30$	$1,205,156.62 \pm 1,096.51$	$221,093.76 \pm 469.21$
Fitted W+jets events	542.50 ± 11.41	$747,836.05 \pm 9,126.85$	$53,868.26 \pm 657.43$
Fitted Z+jets events	478.95 ± 5.85	$156,691.16 \pm 3,295.44$	$13,022.49 \pm 273.88$
Fitted Diboson events	45.63 ± 0.96	$17,700.26 \pm 372.26$	$1,065.55 \pm 22.41$
Fitted Single Top events	72.29 ± 1.52	$14,774.37 \pm 310.73$	$7,027.37 \pm 147.80$
Fitted DY events	37.87 ± 0.80	$2,006.73 \pm 42.20$	238.71 ± 5.02
Fitted QCD events	213.30 ± 4.49	$244,893.97 \pm 5,150.48$	$34,975.86 \pm 735.59$
Fitted $t\bar{t}$ events	340.16 ± 2.20	$21,254.08 \pm 137.63$	$110,894.58 \pm 718.09$
Fitted LQ300 events	2.61	–	0.94
MC exp. events	8,386.41	1,205,193.28	223,489.99
MC exp. W+jets events	543.80	746,807.50	53,794.17
MC exp. Z+jets events	478.29	157,067.78	13,053.79
MC exp. Diboson events	45.74	17,742.80	1,068.12
MC exp. Single Top events	72.46	14,809.88	7,044.27
MC exp. DY events	37.96	2,011.56	239.28
data - driven exp. QCD events	213.81	245,482.59	35,059.93
MC exp. $t\bar{t}$ events	339.98	21,242.70	110,835.22
MC exp. LQ300 events	6,654.37	28.47	2,395.21

Table A.1: Signal region: Fit results for an integrated luminosity of 20.3 fb^{-1}. The results are obtained from the control regions and signal region at 300 GeV using the exclusion fit. Nominal MC expectations (normalised to MC cross - sections) are given for comparison. The errors shown are the statistical plus systematic uncertainties. Uncertainties on the fitted yields are symmetric by construction, where the negative error is truncated when reaching to zero event yield.

Event yields	SR2 [350 GeV]	W+jets [CR A]	$t\bar{t}$ [CR C]
Observed events	519	1,205,164	221,094
Fitted events	519.52 ± 45.69	$1,205,162.15 \pm 1,097.74$	$221,094.30 \pm 470.05$
Fitted W+jets events	180.65 ± 7.02	$746,556.77 \pm 16,865.38$	$53,776.11 \pm 1,214.85$
Fitted Z+jets events	145.07 ± 3.28	$157,157.94 \pm 6,106.64$	$13,061.28 \pm 507.52$
Fitted Diboson events	18.99 ± 0.74	$17,752.99 \pm 689.82$	$1,068.73 \pm 41.53$
Fitted Single Top events	22.21 ± 0.86	$14,818.38 \pm 575.79$	$7,048.31 \pm 273.87$
Fitted DY events	14.35 ± 0.56	$2,012.71 \pm 78.21$	239.42 ± 9.30
Fitted QCD events	56.56 ± 2.20	$245,623.50 \pm 9,544.12$	$35,080.05 \pm 1,363.09$
Fitted $t\bar{t}$ events	81.67 ± 0.81	$21,239.86 \pm 210.50$	$110,820.41 \pm 1,098.329$
Fitted LQ350 events	0.02	–	–
MC exp. events	2,653.95	1,205,164.81	221,882.12
MC exp. W+jets events	180.55	746,807.50	53,794.17
MC exp. Z+jets events	145.12	157,067.78	13,053.79
MC exp. Diboson events	18.98	17,742.80	1,068.12
MC exp. Single Top events	22.20	14,809.88	7,044.27
MC exp. DY events	14.35	2,011.56	239.28
data - driven exp. QCD events	56.53	245,482.59	35,059.93
MC exp. $t\bar{t}$ events	81.68	21,242.70	110,835.22
MC exp. LQ350 events	2,134.54	28.47	2,395.21

Table A.2: Signal region: Fit results for an integrated luminosity of 20.3 fb^{-1}. The results are obtained from the control regions and signal region at 350 GeV using the exclusion fit. Nominal MC expectations (normalised to MC cross - sections) are given for comparison. The errors shown are the statistical plus systematic uncertainties. Uncertainties on the fitted yields are symmetric by construction, where the negative error is truncated when reaching to zero event yield.

Event yields	SR3 [400 GeV]	W+jets [CR A]	$t\bar{t}$ [CR C]
Observed events	328	1,205,164	221,094
Fitted events	$328.31^{+356.23}_{-328.31}$	$1,205,165.08 \pm 1,122.95$	$221,094.89 \pm 491.90$
Fitted W+jets events	121.00 ± 10.10	$746,737.09 \pm 36,356.65$	$53,789.10 \pm 2,618.85$
Fitted Z+jets events	93.45 ± 4.55	$157,093.46 \pm 13,118.95$	$13,055.92 \pm 1,090.31$
Fitted Diboson events	10.90 ± 0.91	$17,745.70 \pm 1,481.95$	$1,068.29 \pm 89.21$
Fitted Single Top events	14.30 ± 1.19	$14,812.30 \pm 575.79$	$7,045.42 \pm 588.37$
Fitted DY events	10.14 ± 0.85	$2,011.89 \pm 168.01$	239.32 ± 19.99
Fitted QCD events	33.17 ± 2.77	$245,522.73 \pm 20,503.71$	$35,065.66 \pm 2,928.35$
Fitted $t\bar{t}$ events	45.27 ± 0.90	$21,241.92 \pm 422.59$	$110,831.16 \pm 2,204.88$
Fitted LQ400 events	0.08	–	0.02
MC exp. events	1,424.09	1,205,164.81	221,882.12
MC exp. W+jets events	120.98	746,807.50	53,794.17
MC exp. Z+jets events	93.46	157,067.78	13,053.79
MC exp. Diboson events	10.90	17,742.80	1,068.12
MC exp. Single Top events	14.30	14,809.88	7,044.27
MC exp. DY events	10.14	2,011.56	239.28
data - driven exp. QCD events	33.16	245,482.59	35,059.93
MC exp. $t\bar{t}$ events	45.27	21,242.70	110,835.22
MC exp. LQ400 events	1,095.88	28.47	2,395.21

Table A.3: Signal region: Fit results for an integrated luminosity of 20.3 fb^{-1}. The results are obtained from the control regions and signal region at 400 GeV using the exclusion fit. Nominal MC expectations (normalised to MC cross - sections) are given for comparison. The errors shown are the statistical plus systematic uncertainties. Uncertainties on the fitted yields are symmetric by construction, where the negative error is truncated when reaching to zero event yield.

Event yields	SR4 [450 GeV]	W+jets [CR A]	$t\bar{t}$ [CR C]
Observed events	195	1,205,164	221,094
Fitted events	195.95 ± 108.00	$1,205,159.60 \pm 1,095.99$	$221,091.43 \pm 469.70$
Fitted W+jets events	79.96 ± 1.50	$746,896.41 \pm 8,145.06$	$53,800.57 \pm 586.71$
Fitted Z+jets events	53.05 ± 0.58	$157,033.79 \pm 2,943.22$	$13,050.96 \pm 244.61$
Fitted Diboson events	17.19 ± 0.13	$17,738.96 \pm 332.47$	$1,067.88 \pm 20.01$
Fitted Single Top events	5.66 ± 0.11	$14,806.67 \pm 277.52$	$7,042.74 \pm 132.00$
Fitted DY events	6.41 ± 0.12	$2,011.12 \pm 37.69$	239.23 ± 4.48
Fitted QCD events	21.31 ± 0.40	$245,429.47 \pm 4,599.98$	$35,052.34 \pm 656.97$
Fitted $t\bar{t}$ events	22.22 ± 0.14	$21,243.17 \pm 129.98$	$110,837.65 \pm 678.18$
Fitted LQ450 events	–	13,320.01	0.06
MC exp. events	655.03	1,205,164.81	221,882.12
MC exp. W+jets events	79.98	746,807.50	53,794.17
MC exp. Z+jets events	53.05	157,067.78	13,053.79
MC exp. Diboson events	7.19	17,742.80	1,068.12
MC exp. Single Top events	5.66	14,809.88	7,044.27
MC exp. DY events	6.41	2,011.56	239.28
data - driven exp. QCD events	21.32	245,482.59	35,059.93
MC exp. $t\bar{t}$ events	22.22	21,242.70	110,835.22
MC exp. LQ450 events	459.20	28.47	2,395.21

Table A.4: Signal region: Fit results for an integrated luminosity of 20.3 fb^{-1}. The results are obtained from the control regions and signal region at 450 GeV using the exclusion fit. Nominal MC expectations (normalised to MC cross - sections) are given for comparison. The errors shown are the statistical plus systematic uncertainties. Uncertainties on the fitted yields are symmetric by construction, where the negative error is truncated when reaching to zero event yield.

Event yields	SR5 [500 GeV]	W+jets [CR A]	$t\bar{t}$ [CR C]
Observed events	131	1,205,164	221,094
Fitted events	132.20 ± 28.79	$1,205,172.70 \pm 1,096.32$	$221,096.18 \pm 469.08$
Fitted W+jets events	57.27 ± 1.33	$747,177.03 \pm 10,092.47$	$53,820.79 \pm 726.98$
Fitted Z+jets events	32.63 ± 0.44	$156,936.32 \pm 3,646.10$	$13,042.86 \pm 303.02$
Fitted Diboson events	4.99 ± 0.12	$17,727.95 \pm 411.87$	$1,067.22 \pm 24.79$
Fitted Single Top events	4.64 ± 0.11	$14,797.48 \pm 343.79$	$7,038.37 \pm 163.52$
Fitted DY events	4.24 ± 0.10	$2,009.87 \pm 46.70$	239.08 ± 5.55
Fitted QCD events	14.67 ± 0.34	$245,277.13 \pm 5,698.52$	$35,030.58 \pm 813.86$
Fitted $t\bar{t}$ events	13.35 ± 0.09	$21,246.91 \pm 144.39$	$110,857.17 \pm 753.36$
Fitted LQ500 events	0.41	0.01	0.11
MC exp. events	397.63	1,205,164.81	221,882.12
MC exp. W+jets events	57.32	746,807.50	53,794.17
MC exp. Z+jets events	32.61	157,067.78	13,053.79
MC exp. Diboson events	5.00	17,742.80	1,068.12
MC exp. Single Top events	4.65	14,809.88	7,044.27
MC exp. DY events	4.24	2,011.56	239.28
data - driven exp. QCD events	14.69	245,482.59	35,059.93
MC exp. $t\bar{t}$ events	13.35	21,242.70	110,835.22
MC exp. LQ500 events	265.77	28.47	2,395.21

Table A.5: Signal region: Fit results for an integrated luminosity of 20.3 fb^{-1}. The results are obtained from the control regions and signal region at 500 GeV using the exclusion fit. Nominal MC expectations (normalised to MC cross - sections) are given for comparison. The errors shown are the statistical plus systematic uncertainties. Uncertainties on the fitted yields are symmetric by construction, where the negative error is truncated when reaching to zero event yield.

Event yields	SR6 [550 GeV]	W+jets [CR A]	$t\bar{t}$ [CR C]
Observed events	48	1,205,164	221,094
Fitted events	$48.26^{+228.50}_{-48.26}$	$1,205,165.13 \pm 1,098.07$	$221,094.73 \pm 475.96$
Fitted W+jets events	26.36 ± 0.48	$746,961.73 \pm 7,897.16$	$53,805.28 \pm 568.85$
Fitted Z+jets events	6.85 ± 0.07	$157,011.86 \pm 2,844.57$	$13,049.14 \pm 236.41$
Fitted Diboson events	1.05 ± 0.0	$17,736.49 \pm 321.33$	$1,067.74 \pm 19.34$
Fitted Single Top events	0.62 ± 0.01	$14,804.60 \pm 268.21$	$7,041.76 \pm 127.57$
Fitted DY events	2.16 ± 0.04	$2,010.84 \pm 36.43$	239.20 ± 4.33
Fitted QCD events	7.10 ± 0.13	$245,395.19 \pm 4,445.80$	$35,047.44 \pm 634.95$
Fitted $t\bar{t}$ events	4.12 ± 0.02	$21,244.42 \pm 127.05$	$110,844.18 \pm 662.89$
Fitted LQ550 events	0.02	–	–
MC exp. events	148.12	1,205,164.81	221,882.12
MC exp. W+jets events	26.37	746,807.50	53,794.17
MC exp. Z+jets events	6.85	157,067.78	13,053.79
MC exp. Diboson events	1.05	17,742.80	1,068.12
MC exp. Single Top events	0.62	14,809.88	7,044.27
MC exp. DY events	2.16	2,011.56	239.28
data - driven exp. QCD events	7.11	245,482.59	35,059.93
MC exp. $t\bar{t}$ events	4.12	21,242.70	110,835.22
MC exp. LQ550 events	99.84	28.47	2,395.21

Table A.6: Signal region: Fit results for an integrated luminosity of 20.3 fb^{-1}. The results are obtained from the control regions and signal region at 550 GeV using the exclusion fit. Nominal MC expectations (normalised to MC cross - sections) are given for comparison. The errors shown are the statistical plus systematic uncertainties. Uncertainties on the fitted yields are symmetric by construction, where the negative error is truncated when reaching to zero event yield.

Event yields	SR7 [600 GeV]	W+jets [CR A]	$t\bar{t}$ [CR C]
Observed events	29	1,205,164	221,094
Fitted events	29.72 ± 22.61	$1,205,170.98 \pm 1,098.31$	$221,095.94 \pm 470.68$
Fitted W+jets events	19.06 ± 1.13	$746,915.60 \pm 9,275.98$	$53,801.95 \pm 668.17$
Fitted Z+jets events	2.50 ± 0.14	$157,030.69 \pm 3,359.24$	$13,050.70 \pm 279.18$
Fitted Diboson events	0.26 ± 0.02	$17,738.61 \pm 379.47$	$1,067.86 \pm 22.84$
Fitted Single Top events	–	$14,806.38 \pm 316.74$	$7,042.60 \pm 150.6$
Fitted DY events	1.44 ± 0.08	$2,011.08 \pm 43.02$	239.23 ± 5.12
Fitted QCD events	3.95 ± 0.23	$245,424.63 \pm 5,250.19$	$35,051.65 \pm 749.83$
Fitted $t\bar{t}$ events	2.41 ± 0.14	$21,243.98 \pm 139.03$	$110,841.91 \pm 725.39$
Fitted LQ600 events	0.10	0.01	0.04
MC exp. events	83.75	1,205,164.91	221,111.47
MC exp. W+jets events	19.05	746,807.50	53,794.17
MC exp. Z+jets events	2.50	157,067.78	13,053.79
MC exp. Diboson events	0.26	17,742.80	1,068.12
MC exp. Single Top events	–	14,809.88	7,044.27
MC exp. DY events	1.44	2,011.56	239.28
data - driven exp. QCD events	3.95	245,482.59	35,059.93
MC exp. $t\bar{t}$ events	2.41	21,242.70	110,835.22
MC exp. LQ600 events	54.14	1,800.1	16.69

Table A.7: Signal region: Fit results for an integrated luminosity of 20.3 fb^{-1}. The results are obtained from the control regions and signal region at 600 GeV using the exclusion fit. Nominal MC expectations (normalised to MC cross - sections) are given for comparison. The errors shown are the statistical plus systematic uncertainties. Uncertainties on the fitted yields are symmetric by construction, where the negative error is truncated when reaching to zero event yield.

Event yields	SR9 [700 GeV]	W+jets [CR A]	$t\bar{t}$ [CR C]
Observed events	16	1,205,164	221,094
Fitted events	16.59 ± 11.39	$1,205,164.06 \pm 1,098.91$	$221,094.50 \pm 470.25$
Fitted W+jets events	12.11 ± 0.99	$746,849.10 \pm 12,076.22$	$53,797.16 \pm 869.88$
Fitted Z+jets events	–	$157,052.41 \pm 4,351.27$	$13,052.51 \pm 361.63$
Fitted Diboson events	–	$17,741.07 \pm 491.53$	$1,068.01 \pm 29.59$
Fitted Single Top events	–	$14,808.43 \pm 410.28$	$7,043.58 \pm 195.15$
Fitted DY events	0.78 ± 0.06	$2,011.36 \pm 55.73$	239.26 ± 6.63
Fitted QCD events	2.37 ± 0.19	$245,458.57 \pm 6,800.64$	$35,056.49 \pm 971.27$
Fitted $t\bar{t}$ events	1.31 ± 0.10	$21,243.13 \pm 163.79$	$110,837.48 \pm 854.57$
Fitted LQ700 events	0.02	–	0.01
MC exp. events	36.01	1,205,164.81	221,100.07
MC exp. W+jets events	12.11	746,807.50	53,794.17
MC exp. Z+jets events	–	157,067.78	13,053.79
MC exp. Diboson events	–	17,742.80	1,068.12
MC exp. Single Top events	–	14,809.88	7,044.27
MC exp. DY events	0.78	2,011.56	239.28
data - driven exp. QCD events	2.37	245,482.59	35,059.93
MC exp. $t\bar{t}$ events	1.31	21,242.70	110,835.22
MC exp. LQ700 events	19.44	–	5.29

Table A.8: Signal region: Fit results for an integrated luminosity of 20.3 fb^{-1}. The results are obtained from the control regions and signal region at 700 GeV using the exclusion fit. Nominal MC expectations (normalised to MC cross - sections) are given for comparison. The errors shown are the statistical plus systematic uncertainties. Uncertainties on the fitted yields are symmetric by construction, where the negative error is truncated when reaching to zero event yield.

Event yields	SR10 [750 GeV]	W+jets [CR A]	$t\bar{t}$ [CR C]
Observed events	15	1,205,164	221,094
Fitted events	$15.66^{+48.13}_{-15.66}$	$1,205,160.40 \pm 1,095.22$	$221,093.69 \pm 469.37$
Fitted W+jets events	11.59 ± 0.99	$746,859.91 \pm 4,799.58$	$53,797.94 \pm 345.72$
Fitted Z+jets events	–	$157,047.18 \pm 1,732.52$	$13,052.07 \pm 143.99$
Fitted Diboson events	–	$17,740.48 \pm 195.71$	$1,067.98 \pm 11.78$
Fitted Single Top events	–	$14,807.94 \pm 163.36$	$7,043.34 \pm 77.70$
Fitted DY events	0.68 ± 0.06	$2,011.29 \pm 22.1$	239.25 ± 2.64
Fitted QCD events	2.37 ± 0.20	$245,450.40 \pm 2,707.77$	$35,055.33 \pm 386.7$
Fitted $t\bar{t}$ events	1.01 ± 0.09	$21,243.19 \pm 107.87$	$110,837.77 \pm 562.84$
Fitted LQ750 events	0.01	0.01	0.01
MC exp. events	30.17	1,205,164.81	221,097.94
MC exp. W+jets events	11.60	746,807.50	53,794.17
MC exp. Z+jets events	–	157,067.78	13,053.79
MC exp. Diboson events	–	17,742.80	1,068.12
MC exp. Single Top events	–	14,809.88	7,044.27
MC exp. DY events	0.68	2,011.56	239.28
data - driven exp. QCD events	2.37	245,482.59	35,059.93
MC exp. $t\bar{t}$ events	1.01	21,242.70	110,835.22
MC exp. LQ750 events	14.51	–	3.16

Table A.9: Signal region: Fit results for an integrated luminosity of 20.3 fb^{-1}. The results are obtained from the control regions and signal region at 750 GeV using the exclusion fit. Nominal MC expectations (normalised to MC cross - sections) are given for comparison. The errors shown are the statistical plus systematic uncertainties. Uncertainties on the fitted yields are symmetric by construction, where the negative error is truncated when reaching to zero event yield.

Event yields	SR11 [800 GeV]	W+jets [CR A]	$t\bar{t}$ [CR C]
Observed events	13	1,205,164	221,094
Fitted events	$13.89^{+25.07}_{-13.89}$	$1,205,167.88 \pm 1,128.52$	$221,095.36 \pm 473.86$
Fitted W+jets events	10.21 ± 1.34	$746,815.65 \pm 35,612.74$	$53,794.76 \pm 2,565.26$
Fitted Z+jets events	–	$157,065.91 \pm 12,838.69$	$13,053.63 \pm 1,067.01$
Fitted Diboson events	–	$17,742.59 \pm 1,450.29$	$1,068.10 \pm 87.31$
Fitted Single Top events	–	$14,809.70 \pm 1,210.56$	$7,044.18 \pm 575.80$
Fitted DY events	0.60 ± 0.08	$2,011.53 \pm 164.42$	239.28 ± 19.56
Fitted QCD events	2.21 ± 0.29	$245,479.66 \pm 20,065.70$	$35,059.51 \pm 2,865.79$
Fitted $t\bar{t}$ events	0.86 ± 0.08	$21,242.83 \pm 415.65$	$110,835.89 \pm 2,168.69$
Fitted LQ800 events	0.01	0.01	0.01
MC exp. events	25.15	1,205,164.81	221,096.37
MC exp. W+jets events	10.21	746,807.50	53,794.17
MC exp. Z+jets events	–	157,067.78	13,053.79
MC exp. Diboson events	–	17,742.80	1,068.12
MC exp. Single Top events	–	14,809.88	7,044.27
MC exp. DY events	0.60	2,011.56	239.28
data - driven exp. QCD events	2.21	245,482.59	35,059.93
MC exp. $t\bar{t}$ events	0.86	21,242.70	110,835.22
MC exp. LQ800 events	11.27	–	1.59

Table A.10: Signal region: Fit results for an integrated luminosity of 20.3 fb^{-1}. The results are obtained from the control regions and signal region at 800 GeV using the exclusion fit. Nominal MC expectations (normalised to MC cross - sections) are given for comparison. The errors shown are the statistical plus systematic uncertainties. Uncertainties on the fitted yields are symmetric by construction, where the negative error is truncated when reaching to zero event yield.

Event yields	SR12 [850 GeV]	W+jets [CR A]	$t\bar{t}$ [CR C]
Observed events	10	1,205,164	221,094
Fitted events	$10.26^{+19.40}_{-10.26}$	$1,205,170.35 \pm 1,094.03$	$221,095.75 \pm 478.71$
Fitted W+jets events	7.58 ± 1.00	$746,879.46 \pm 13,153.05$	$53,799.35 \pm 947.44$
Fitted Z+jets events	–	$157,043.60 \pm 4,749.86$	$13,051.78 \pm 394.76$
Fitted Diboson events	–	$17,740.07 \pm 536.56$	$1,067.95 \pm 32.30$
Fitted Single Top events	–	$14,807.60 \pm 447.86$	$7,043.18 \pm 213.02$
Fitted DY events	0.20 ± 0.03	$2,011.25 \pm 60.83$	239.25 ± 7.24
Fitted QCD events	1.89 ± 0.25	$245,444.81 \pm 7,423.60$	$35,054.53 \pm 1,060.2$
Fitted $t\bar{t}$ events	0.56 ± 0.07	$21,243.56 \pm 191.99$	$110,839.71 \pm 1,001.71$
Fitted LQ850 events	0.03	–	–
MC exp. events	15.92	1,205,164.81	221,095.61
MC exp. W+jets events	7.59	746,807.50	53,794.17
MC exp. Z+jets events	–	157,067.78	13,053.79
MC exp. Diboson events	–	17,742.80	1,068.12
MC exp. Single Top events	–	14,809.88	7,044.27
MC exp. DY events	0.20	2,011.56	239.28
data - driven exp. QCD events	1.89	245,482.59	35,059.93
MC exp. $t\bar{t}$ events	0.56	21,242.70	110,835.22
MC exp. LQ850 events	5.68	–	0.83

Table A.11: Signal region: Fit results for an integrated luminosity of 20.3 fb^{-1}. The results are obtained from the control regions and signal region at 850 GeV using the exclusion fit. Nominal MC expectations (normalised to MC cross - sections) are given for comparison. The errors shown are the statistical plus systematic uncertainties. Uncertainties on the fitted yields are symmetric by construction, where the negative error is truncated when reaching to zero event yield.

Event yields	SR13 [900 GeV]	W+jets [CR A]	$t\bar{t}$ [CR C]
Observed events	9	1,205,164	221,094
Fitted events	$9.92^{+10.00}_{-9.92}$	$1,205,161.54 \pm 1,094.33$	$221,094.09 \pm 469.80$
Fitted W+jets events	7.54 ± 1.08	$746,816.85 \pm 6,101.64$	$53,794.84 \pm 439.5$
Fitted Z+jets events	–	$157,063.22 \pm 2,176.166$	$13,053.41 \pm 180.86$
Fitted Diboson events	–	$17,742.29 \pm 245.82$	$1,068.08 \pm 14.80$
Fitted Single Top events	–	$14,809.45 \pm 205.19$	$7,044.06 \pm 97.60$
Fitted DY events	0.19 ± 0.03	$2,011.50 \pm 27.87$	239.28 ± 3.32
Fitted QCD events	1.58 ± 0.23	$245,475.47 \pm 3,401.14$	$35,058.91 \pm 485.75$
Fitted $t\bar{t}$ events	0.51 ± 0.07	$21,242.75 \pm 107.65$	$110,835.50 \pm 561.65$
Fitted LQ900 events	0.10	0.01	0.01
MC exp. events	13.95	1,205,164.81	221,095.25
MC exp. W+jets events	7.54	746,807.50	53,794.17
MC exp. Z+jets events	–	157,067.78	13,053.79
MC exp. Diboson events	–	17,742.80	1,068.12
MC exp. Single Top events	–	14,809.88	7,044.27
MC exp. DY events	0.19	2,011.56	239.28
data - driven exp. QCD events	1.58	245,482.59	35,059.93
MC exp. $t\bar{t}$ events	0.51	21,242.70	110,835.22
MC exp. LQ900 events	4.13	–	0.47

Table A.12: Signal region: Fit results for an integrated luminosity of 20.3 fb^{-1}. The results are obtained from the control regions and signal region at 900 GeV using the exclusion fit. Nominal MC expectations (normalised to MC cross - sections) are given for comparison. The errors shown are the statistical plus systematic uncertainties. Uncertainties on the fitted yields are symmetric by construction, where the negative error is truncated when reaching to zero event yield.

Event yields	SR14 [950 GeV]	W+jets [CR A]	$t\bar{t}$ [CR C]
Observed events	4	1,205,164	221,094
Fitted events	4.71 ± 3.39	$1,205,164.70 \pm 1,124.84$	$221,094.72 \pm 473.21$
Fitted W+jets events	3.19 ± 0.72	$746,811.58 \pm 35,062.41$	$53,794.46 \pm 2,525.62$
Fitted Z+jets events	–	$157,066.26 \pm 12,646.27$	$13,053.66 \pm 1,051.02$
Fitted Diboson events	–	$17,742.63 \pm 1,428.56$	$1,068.11 \pm 86.00$
Fitted Single Top events	–	$14,809.73 \pm 1,192.41$	$7,044.20 \pm 567.17$
Fitted DY events	0.07 ± 0.02	$2,011.54 \pm 161.96$	239.28 ± 19.27
Fitted QCD events	1.26 ± 0.29	$245,480.22 \pm 19,764.97$	$35,059.59 \pm 2,822.84$
Fitted $t\bar{t}$ events	0.17 ± 0.04	$21,242.74 \pm 409.04$	$110,835.43 \pm 2,134.19$
Fitted LQ950 events	0.02	–	–
MC exp. events	6.32	1,205,164.81	221,095.10
MC exp. W+jets events	3.19	746,807.50	53,794.17
MC exp. Z+jets events	–	157,067.78	13,053.79
MC exp. Diboson events	–	17,742.80	1,068.12
MC exp. Single Top events	–	14,809.88	7,044.27
MC exp. DY events	0.07	2,011.56	239.28
data - driven exp. QCD events	1.26	245,482.59	35,059.93
MC exp. $t\bar{t}$ events	0.17	21,242.70	110,835.22
MC exp. LQ950 events	1.63	–	0.32

Table A.13: Signal region: Fit results for an integrated luminosity of 20.3 fb^{-1}. The results are obtained from the control regions and signal region at 950 GeV using the exclusion fit. Nominal MC expectations (normalised to MC cross - sections) are given for comparison. The errors shown are the statistical plus systematic uncertainties. Uncertainties on the fitted yields are symmetric by construction, where the negative error is truncated when reaching to zero event yield.

Event yields	SR15 [1000 GeV]	W+jets [CR A]	$t\bar{t}$ [CR C]
Observed events	4	1,205,164	221,094
Fitted events	4.41 ± 3.12	$1,205,163.88 \pm 1,095.86$	$221,094.40 \pm 467.79$
Fitted W+jets events	2.82 ± 0.65	$746,848.58 \pm 13,446.49$	$53,797.13 \pm 968.58$
Fitted Z+jets events	–	$157,052.54 \pm 4,861.65$	$13,052.52 \pm 404.05$
Fitted Diboson events	–	$17,741.08 \pm 549.19$	$1,068.01 \pm 33.06$
Fitted Single Top events	–	$14,808.44 \pm 458.40$	$7,043.58 \pm 218.04$
Fitted DY events	0.06 ± 0.01	$2,011.36 \pm 62.26$	239.26 ± 7.41
Fitted QCD events	1.26 ± 0.29	$245,458.77 \pm 7,598.32$	$35,056.52 \pm 1,085.20$
Fitted $t\bar{t}$ events	0.17 ± 0.04	$21,243.11 \pm 177.45$	$110,837.36 \pm 925.87$
Fitted LQ1000 events	0.10	–	0.02
MC exp. events	5.43	1,205,164.81	221,094.96
MC exp. W+jets events	2.84	746,807.50	53,794.17
MC exp. Z+jets events	–	157,067.78	13,053.79
MC exp. Diboson events	–	17,742.80	1,068.12
MC exp. Single Top events	–	14,809.88	7,044.27
MC exp. DY events	0.06	2,011.56	239.28
data - driven exp. QCD events	1.26	245,482.59	35,059.93
MC exp. $t\bar{t}$ events	0.17	21,242.70	110,835.22
MC exp. LQ1000 events	1.10	–	0.18

Table A.14: Signal region: Fit results for an integrated luminosity of 20.3 fb^{-1}. The results are obtained from the control regions and signal region at 1000 GeV using the exclusion fit. Nominal MC expectations (normalised to MC cross - sections) are given for comparison. The errors shown are the statistical plus systematic uncertainties. Uncertainties on the fitted yields are symmetric by construction, where the negative error is truncated when reaching to zero event yield.

Event yields	SR16 [1050 GeV]	W+jets [CR A]	$t\bar{t}$ [CR C]
Observed events	2	1,205,164	221,094
Fitted events	2.92 ± 2.05	$1,205,151.25 \pm 1,102.20$	$221,092.14 \pm 471.22$
Fitted W+jets events	1.89 ± 0.57	$746,881.40 \pm 13,407.86$	$53,799.49 \pm 965.80$
Fitted Z+jets events	–	$157,036.12 \pm 4,822.28$	$13,051.15 \pm 400.78$
Fitted Diboson events	–	$17,739.23 \pm 544.79$	$1,067.90 \pm 32.79$
Fitted Single Top events	–	$14,806.89 \pm 454.69$	$7,042.85 \pm 216.27$
Fitted DY events	0.04 ± 0.01	$2,011.15 \pm 61.76$	239.24 ± 7.35
Fitted QCD events	0.79 ± 0.24	$245,433.11 \pm 7,536.78$	$35,052.86 \pm 1,076.41$
Fitted $t\bar{t}$ events	0.14 ± 0.04	$21,243.36 \pm 178.27$	$110,838.64 \pm 930.14$
Fitted LQ1050 events	0.06	0.01	0.01
MC exp. events	3.53	1,205,164.81	221,094.89
MC exp. W+jets events	1.89	746,807.50	53,794.17
MC exp. Z+jets events	–	157,067.78	13,053.79
MC exp. Diboson events	–	17,742.80	1,068.12
MC exp. Single Top events	–	14,809.88	7,044.27
MC exp. DY events	0.04	2,011.56	239.28
data - driven exp. QCD events	0.79	245,482.59	35,059.93
MC exp. $t\bar{t}$ events	0.14	21,242.70	110,835.22
MC exp. LQ1050 events	0.67	–	0.11

Table A.15: Signal region: Fit results for an integrated luminosity of 20.3 fb^{-1}. The results are obtained from the control regions and signal region at 1050 GeV using the exclusion fit. Nominal MC expectations (normalised to MC cross - sections) are given for comparison. The errors shown are the statistical plus systematic uncertainties. Uncertainties on the fitted yields are symmetric by construction, where the negative error is truncated when reaching to zero event yield.

Event yields	SR17 [1100 GeV]	W+jets [CR A]	$t\bar{t}$ [CR C]
Observed events	2	1,205,164	221,094
Fitted events	2.40 ± 0.81	$1,205,164.80 \pm 1,099.38$	$221,094.75 \pm 477.82$
Fitted W+jets events	1.62 ± 0.79	$746,829.18 \pm 8,535.25$	$53,795.73 \pm 614.81$
Fitted Z+jets events	–	$157,059.90 \pm 3,081.15$	$13,053.13 \pm 256.07$
Fitted Diboson events	–	$17,741.91 \pm 348.06$	$1,068.06 \pm 20.95$
Fitted Single Top events	–	$14,809.13 \pm 290.52$	$7,043.91 \pm 138.19$
Fitted DY events	0.04 ± 0.02	$2,011.46 \pm 39.46$	239.27 ± 4.69
Fitted QCD events	0.63 ± 0.31	$245,470.27 \pm 4,815.56$	$35,058.17 \pm 687.76$
Fitted $t\bar{t}$ events	0.09 ± 0.04	$21,242.94 \pm 132.80$	$110,836.47 \pm 692.89$
Fitted LQ1100 events	0.02	0.01	0.01
MC exp. events	2.76	1,205,164.81	221,094.83
MC exp. W+jets events	1.63	746,807.50	53,794.17
MC exp. Z+jets events	–	157,067.78	13,053.79
MC exp. Diboson events	–	17,742.80	1,068.12
MC exp. Single Top events	–	14,809.88	7,044.27
MC exp. DY events	0.04	2,011.56	239.28
data - driven exp. QCD events	0.63	245,482.59	35,059.93
MC exp. $t\bar{t}$ events	0.09	21,242.70	110,835.22
MC exp. LQ1100 events	0.37	–	0.05

Table A.16: Signal region: Fit results for an integrated luminosity of 20.3 fb^{-1}. The results are obtained from the control regions and signal region at 1100 GeV using the exclusion fit. Nominal MC expectations (normalised to MC cross - sections) are given for comparison. The errors shown are the statistical plus systematic uncertainties. Uncertainties on the fitted yields are symmetric by construction, where the negative error is truncated when reaching to zero event yield.

Event yields	SR18 [1150 GeV]	W+jets [CR A]	$t\bar{t}$ [CR C]
Observed events	2	1,205,164	221,094
Fitted events	2.39 ± 0.99	$1,205,165.25 \pm 1,124.50$	$221,094.84 \pm 473.17$
Fitted W+jets events	1.62 ± 0.67	$746,808.34 \pm 35023.37$	$53,794.23 \pm 2,522.81$
Fitted Z+jets events	–	$157,067.63 \pm 12,632.58$	$13,053.77 \pm 1,049.88$
Fitted Diboson events	–	$17,742.79 \pm 1,427.01$	$1,068.11 \pm 85.91$
Fitted Single Top events	–	$14,809.86 \pm 1,191.12$	$7,044.26 \pm 566.55$
Fitted DY events	0.04 ± 0.02	$2,011.56 \pm 161.78$	239.28 ± 19.25
Fitted QCD events	0.63 ± 0.26	$245,482.36 \pm 19,743.58$	$35,059.89 \pm 2,819.79$
Fitted $t\bar{t}$ events	0.09 ± 0.04	$21,242.71 \pm 408.60$	$110,835.28 \pm 2,131.89$
Fitted LQ1150 events	0.01	–	0.02
MC exp. events	2.71	1,205,164.81	221,094.81
MC exp. W+jets events	1.63	746,807.50	53,794.17
MC exp. Z+jets events	–	157,067.78	13,053.79
MC exp. Diboson events	–	17,742.80	1,068.12
MC exp. Single Top events	–	14,809.88	7,044.27
MC exp. DY events	0.04	2,011.56	239.28
data - driven exp. QCD events	0.63	245,482.59	35,059.93
MC exp. $t\bar{t}$ events	0.09	21,242.70	110,835.22
MC exp. LQ1150 events	0.32	–	0.03

Table A.17: Signal region: Fit results for an integrated luminosity of 20.3 fb^{-1}. The results are obtained from the control regions and signal region at 1150 GeV using the exclusion fit. Nominal MC expectations (normalised to MC cross - sections) are given for comparison. The errors shown are the statistical plus systematic uncertainties. Uncertainties on the fitted yields are symmetric by construction, where the negative error is truncated when reaching to zero event yield.

Event yields	SR19 [1200 GeV]	W+jets [CR A]	$t\bar{t}$ [CR C]
Observed events	1	1,205,164	221,094
Fitted events	1.83 ± 0.81	$1,205,164.38 \pm 1,094.18$	$221,094.69 \pm 464.76$
Fitted W+jets events	1.14 ± 0.62	$746,807.47 \pm 4,302.95$	$53,794.17 \pm 309.95$
Fitted Z+jets events	–	$157,067.64 \pm 1,541.21$	$13,053.77 \pm 128.09$
Fitted Diboson events	–	$17,742.79 \pm 174.10$	$1,068.11 \pm 10.48$
Fitted Single Top events	–	$14,809.86 \pm 145.32$	$7,044.26 \pm 69.12$
Fitted DY events	0.03 ± 0.01	$2,011.56 \pm 19.74$	239.28 ± 2.35
Fitted QCD events	0.45 ± 0.25	$245,482.37 \pm 2,408.77$	$35,059.89 \pm 344.02$
Fitted $t\bar{t}$ events	0.08 ± 0.04	$21,242.69 \pm 100.56$	$110,835.17 \pm 524.66$
Fitted LQ1200 events	0.13	–	0.04
MC exp. events	1.96	1,205,164.81	221,094.80
MC exp. W+jets events	1.20	746,807.50	53,794.17
MC exp. Z+jets events	–	157,067.78	13,053.79
MC exp. Diboson events	–	17,742.80	1,068.12
MC exp. Single Top events	–	14,809.88	7,044.27
MC exp. DY events	0.03	2,011.56	239.28
data - driven exp. QCD events	0.47	245,482.59	35,059.93
MC exp. $t\bar{t}$ events	0.09	21,242.70	110,835.22
MC exp. LQ1200 events	0.17	–	0.02

Table A.18: Signal region: Fit results for an integrated luminosity of 20.3 fb^{-1}. The results are obtained from the control regions and signal region at 1200 GeV using the exclusion fit. Nominal MC expectations (normalised to MC cross - sections) are given for comparison. The errors shown are the statistical plus systematic uncertainties. Uncertainties on the fitted yields are symmetric by construction, where the negative error is truncated when reaching to zero event yield.

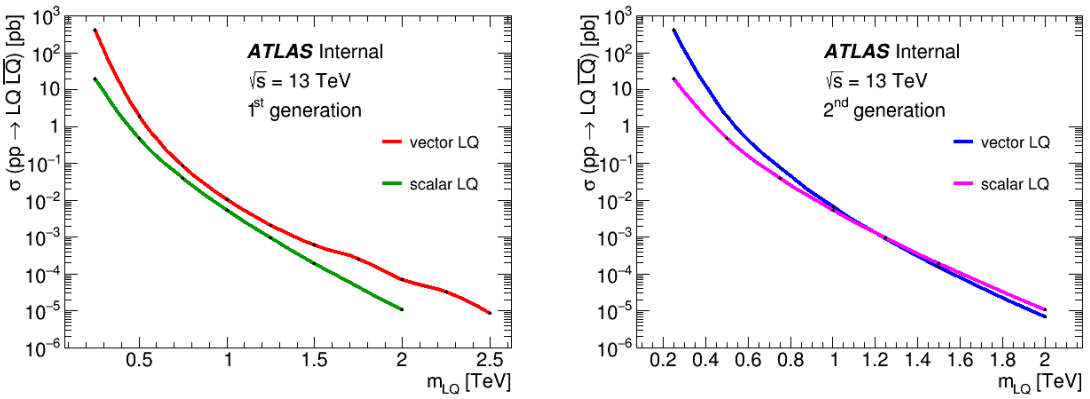
Appendix B

Supplementary Material of 13 TeV Analysis

In the present section more material of the 13 TeV, Run II LQ analysis in addition to these of Ch. 7 can be found.

B.1 vector LQ

In the present search analysis only scalar LQs are studied, since vector LQs production cross sections are in general substantially larger and, thus, more difficult to be discovered. This is due the rates for both gg $\rightarrow$ VV and $q\bar{q} \rightarrow$ VV are larger than their scalar counterparts as they appear in Fig. B.1.

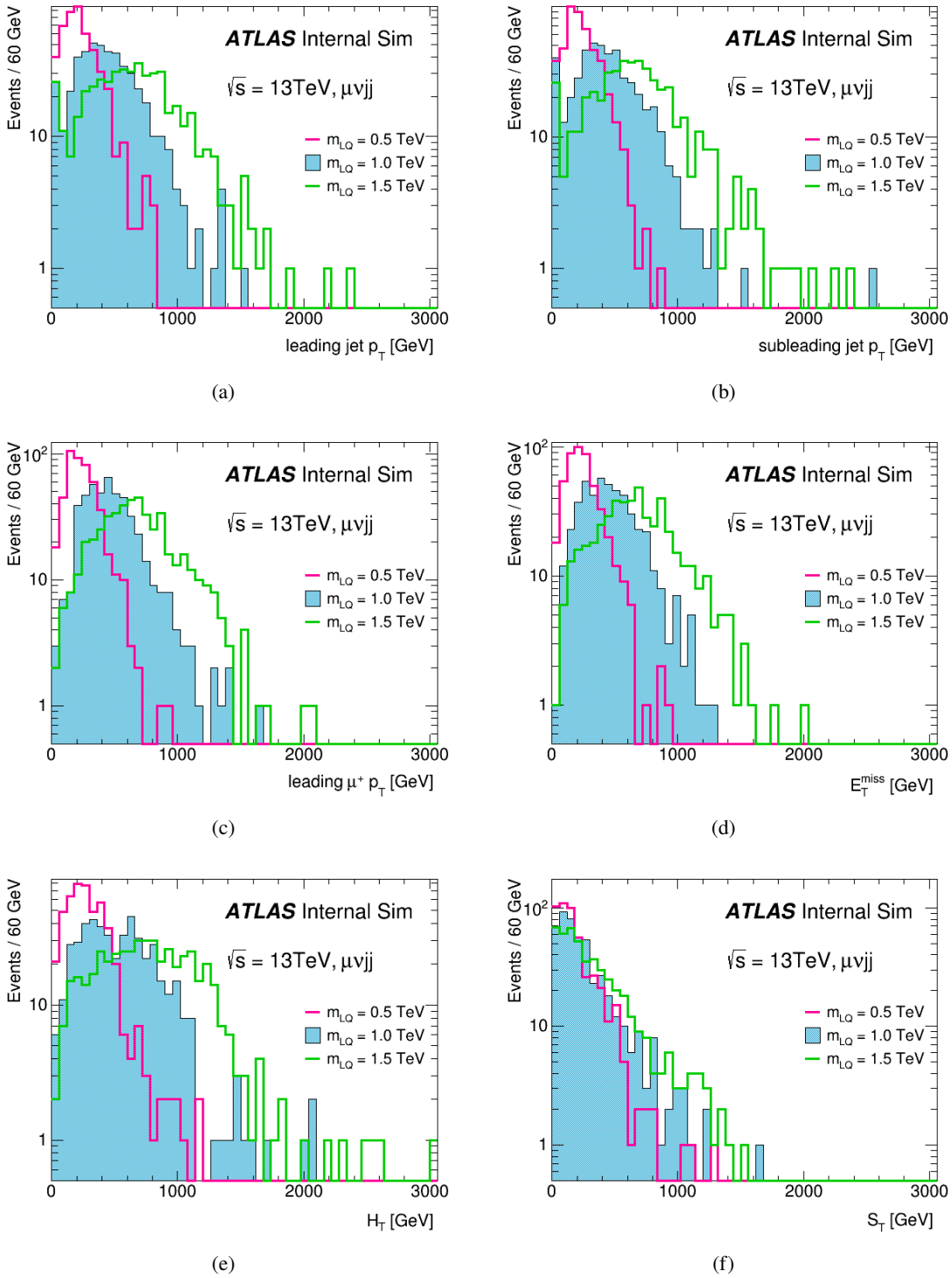


(a) (b)

Figure B.1: Scalar and vector first (a) and second (b) generation pair LQ production cross sections at 13 TeV. The results have been acquired through CalcHEP 3.6.29 software package.

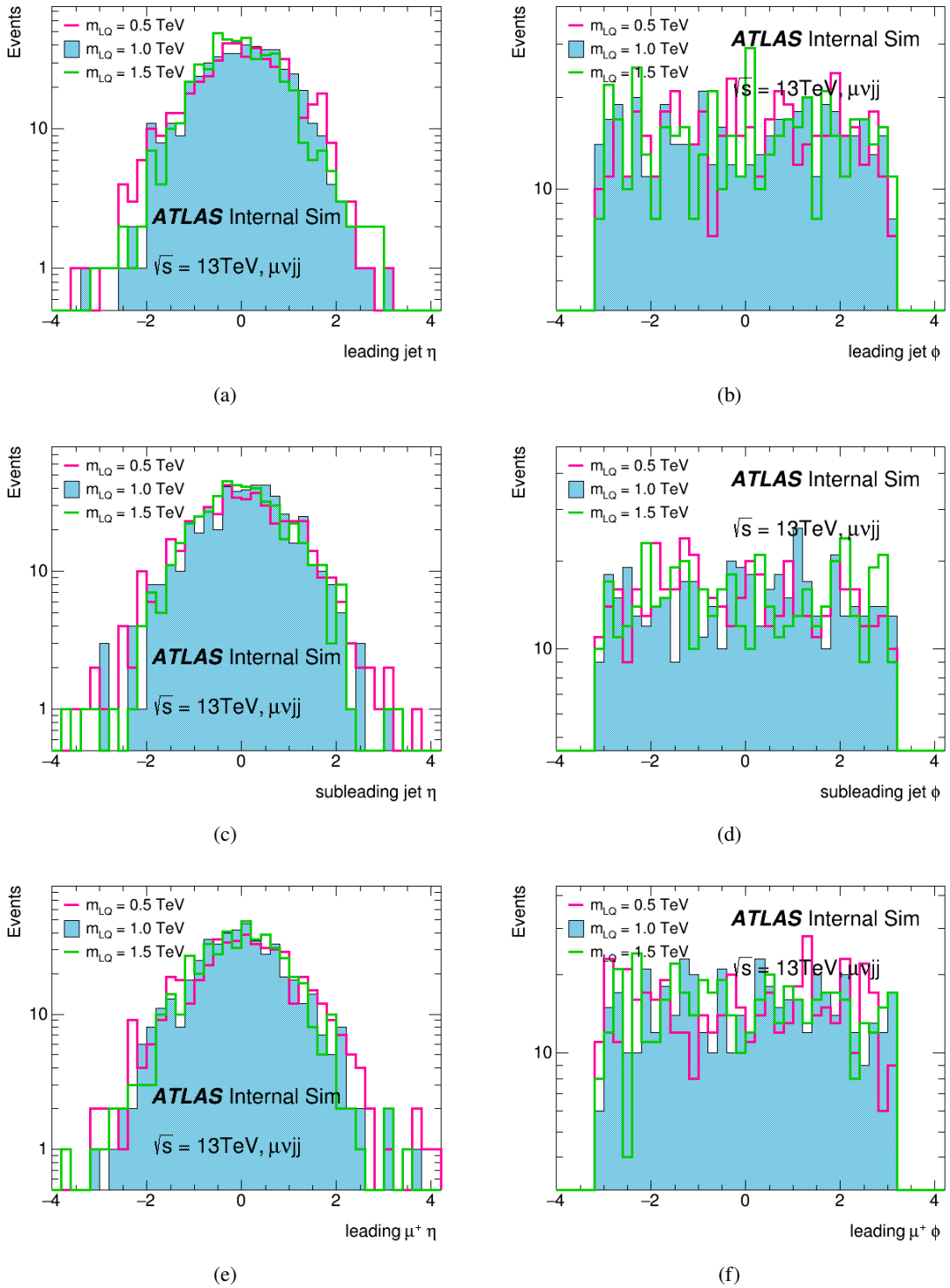
The present cross sections calculations for vector LQs have been done with the use of CalcHEP 3.6.29 software package for effective evaluation and simulation of high energy physics collider processes at parton level.

B.2 MadGraph signal validation histograms



(a) (b) (c) (d) (e) (f)

Figure B.2: MadGraph signal distributions where LQ decays to μ^- and c and $\overline{LQ}$ decays to $\overline{\nu}_\mu$ and $\overline{s}$.

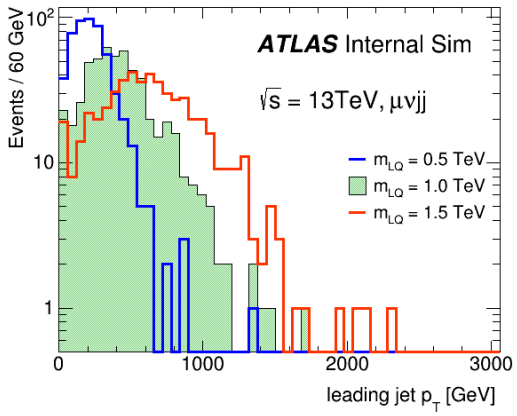


(a) (b)

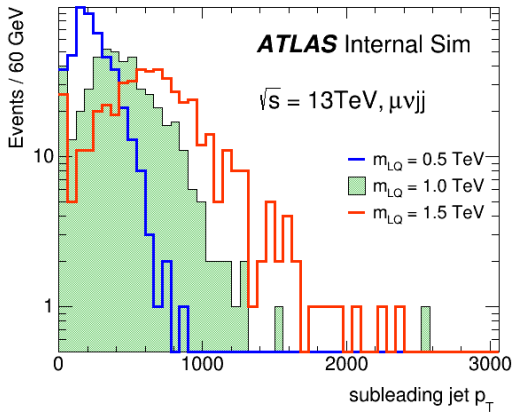
(c) (d)

(e) (f)

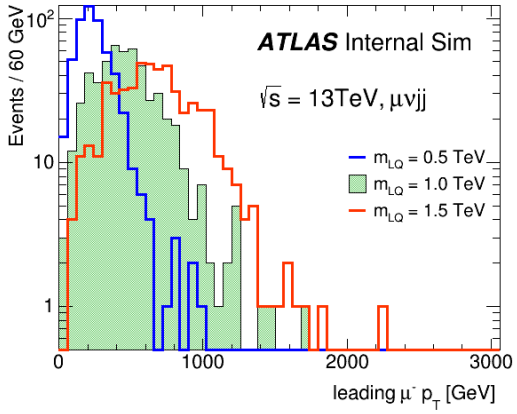
Figure B.3: MadGraph signal distributions where LQ decays to μ^- and c and $\overline{LQ}$ decays to $\overline{\nu}_\mu$ and $\overline{s}$.



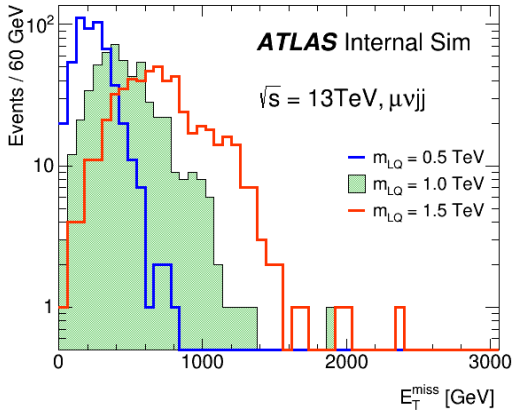
(a)



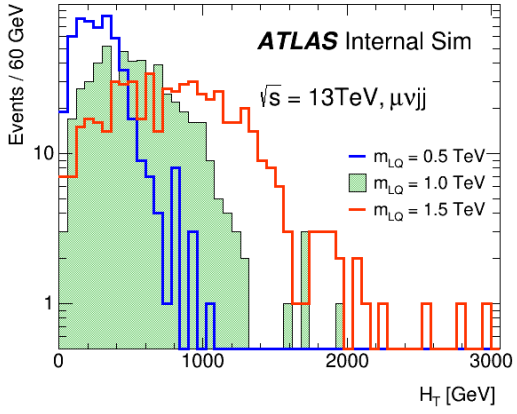
(b)



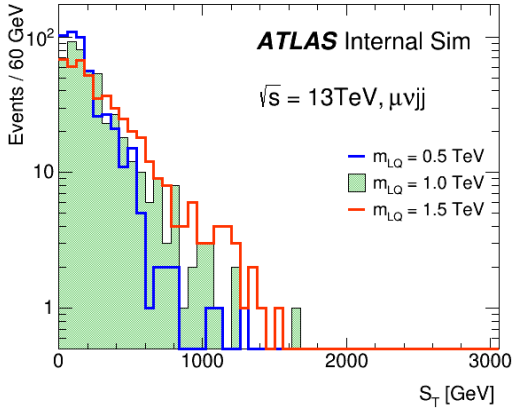
(c)



(d)

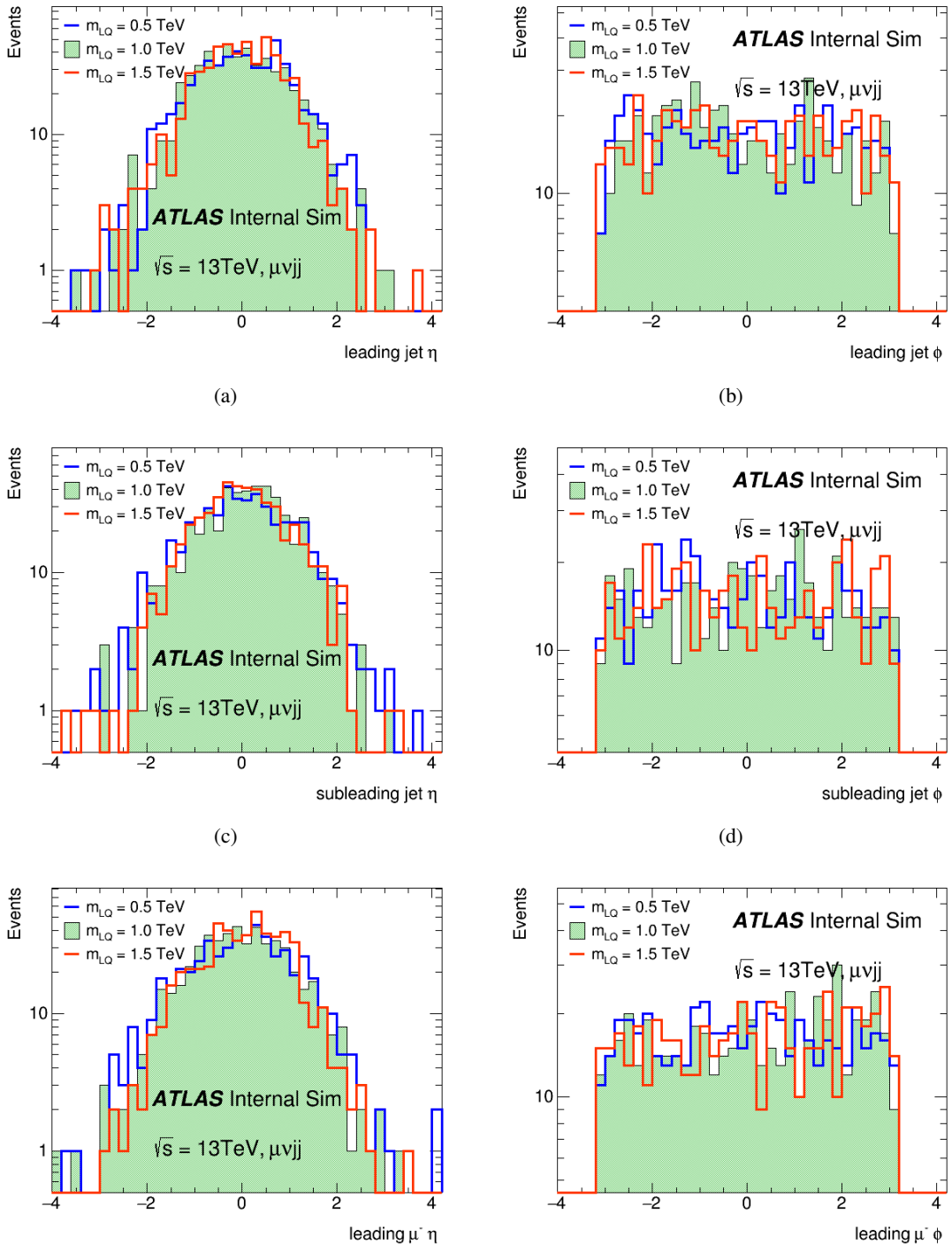


(e)



(f)

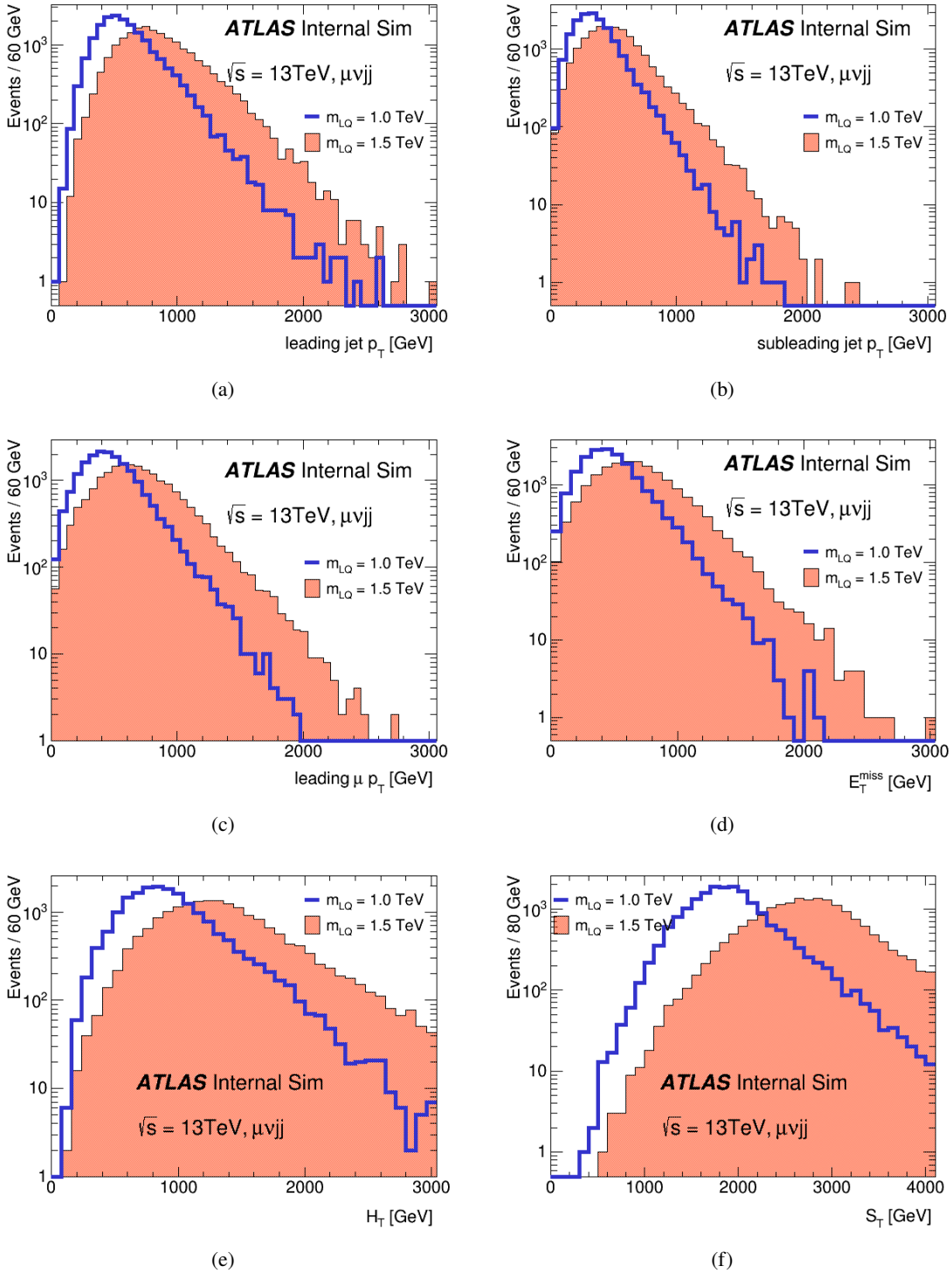
Figure B.4: MadGraph signal distributions where LQ decays to ν_{mu} and s and $\overline{LQ}$ decays to μ^+ and $\bar{c}$.



(a) (b) (c) (d) (e) (f)

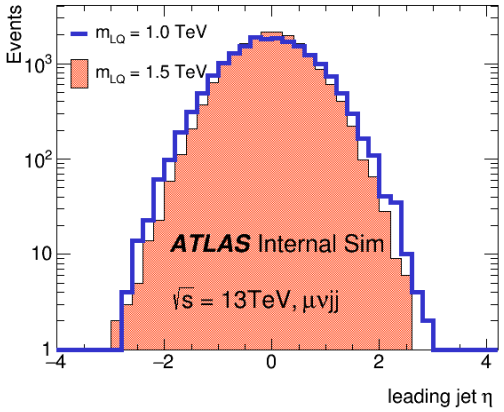
Figure B.5: MadGraph signal distributions where LQ decays to ν_{mu} and s and $\overline{LQ}$ decays to μ^+ and $\overline{c}$.

B.3 Pythia8 signal validation histograms

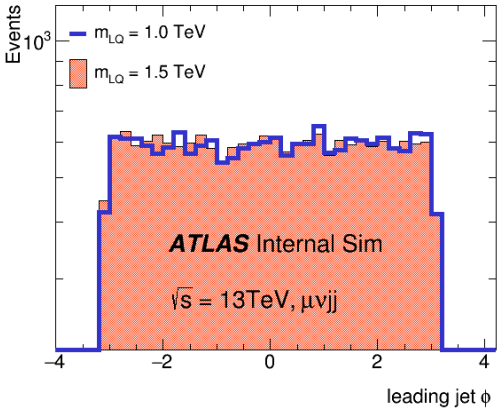


(a) (b) (c) (d) (e) (f)

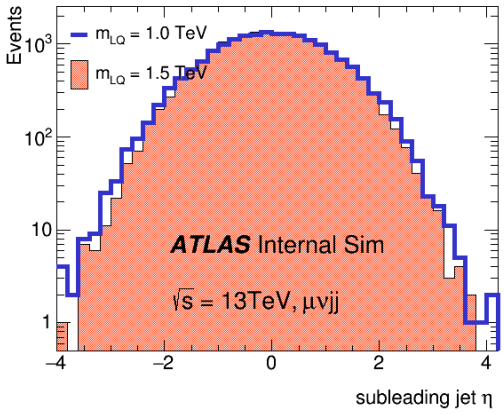
Figure B.6: Pythia signal distributions where LQ decays to μ^- and c and $\overline{LQ}$ decays to $\overline{\nu}_\mu$ and $\overline{s}$.



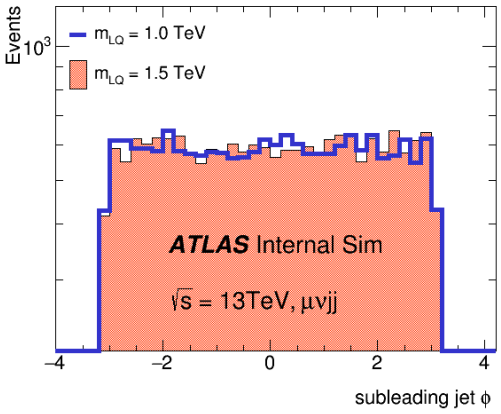
(a)



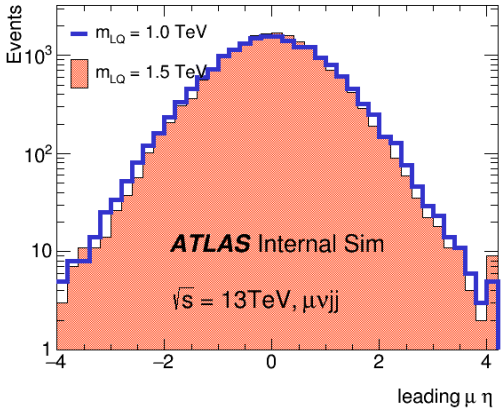
(b)



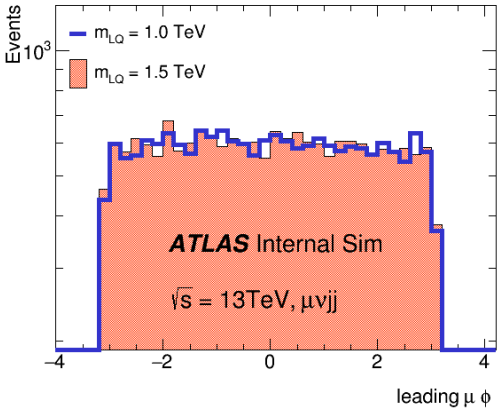
(c)



(d)



(e)



(f)

Figure B.7: Pythia signal distributions where LQ decays to μ^- and c and $\overline{LQ}$ decays to $\overline{\nu}_\mu$ and $\overline{s}$.

B.3.1 Pythia8 NNPDF23LO uncertainties

The present study is about the NNPDF23LO A14 eigentuning for Pythia8 signal samples and only for the dimuon channel. The muon - E_T^{miss} channel is expected to have an analogous behavior.

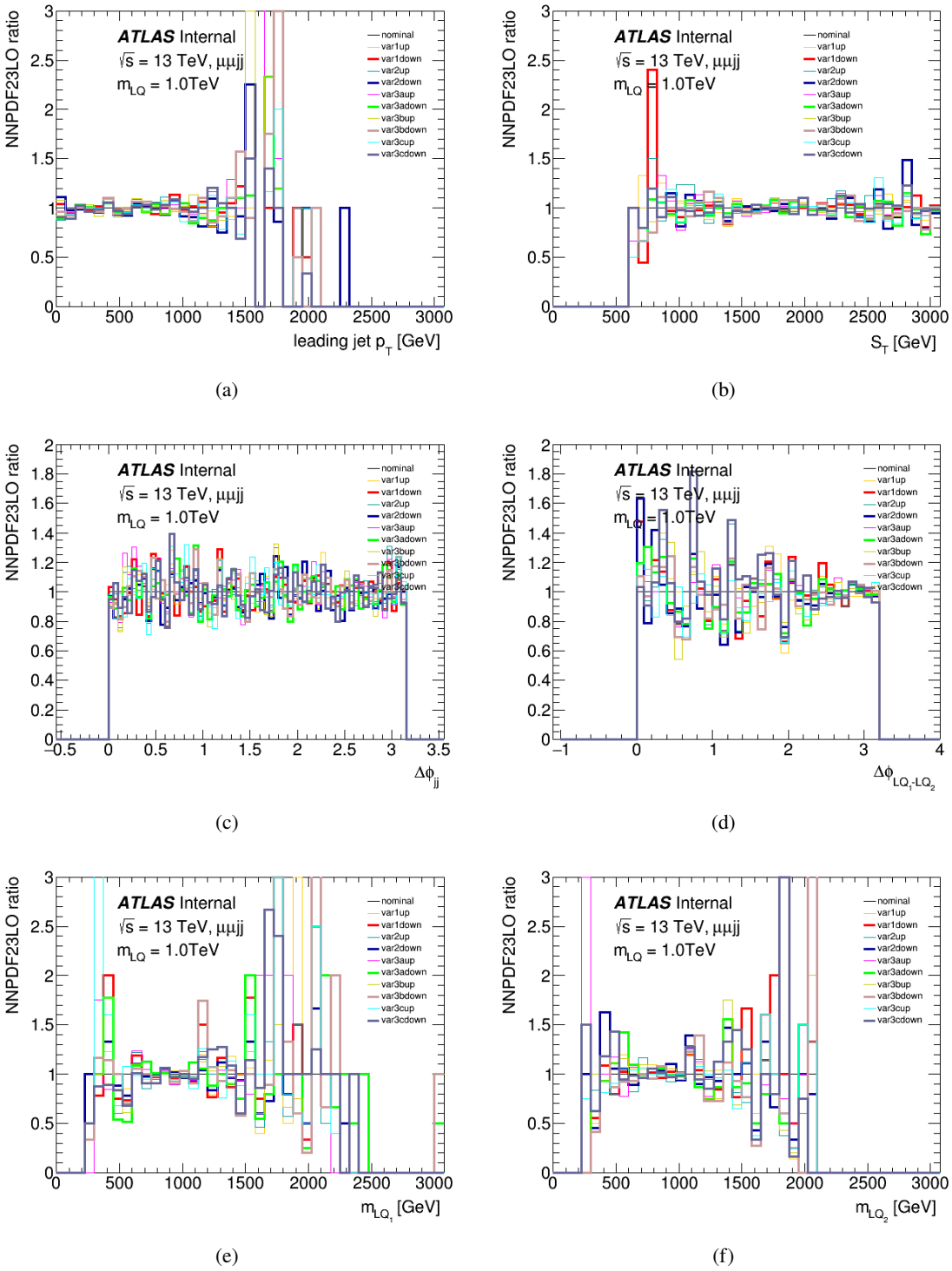
The method for these calculations is the following: a small cutflow has been applied on selected leading leptons and jets. Both samples (1.0 and 1.5 TeV) have 10,000 events at the beginning. Using the cut - and - count method at the end of the cutflow, the percentage of systematic deviations from the nominal NNPDF for all variations are been calculated (Eq. B.1). There are totally 10 different variations compared to the nominal one.

$$syst. = \frac{\sqrt{(alter - nominal)^2 + alter}}{nominal} \cdot 100\% \tag{B.1}$$

This is the applied cutflow:

- leading muon 's $p_T > 40$ GeV
- leading muon 's $\eta < 2.5$
- subleading muon 's $p_T > 40$ GeV
- subleading muon 's $\eta < 2.5$
- leading jet 's $p_T > 50$ GeV
- leading jet 's $\eta < 2.8$
- subleading jet 's $p_T > 50$ GeV
- subleading jet 's $\eta < 2.8$

The maximum systematic deviation for 1.0 TeV sample is 1.95% coming from VAR2: ISF/FSR (jet shapes and substructure) and for 1.5 TeV sample is 1.14% from the same variation.



(a) (b)

(c) (d)

(e) (f)

Figure B.8: Pythia8 signal ratio comparison for different NNPDF23LO variations at 1.0 TeV.

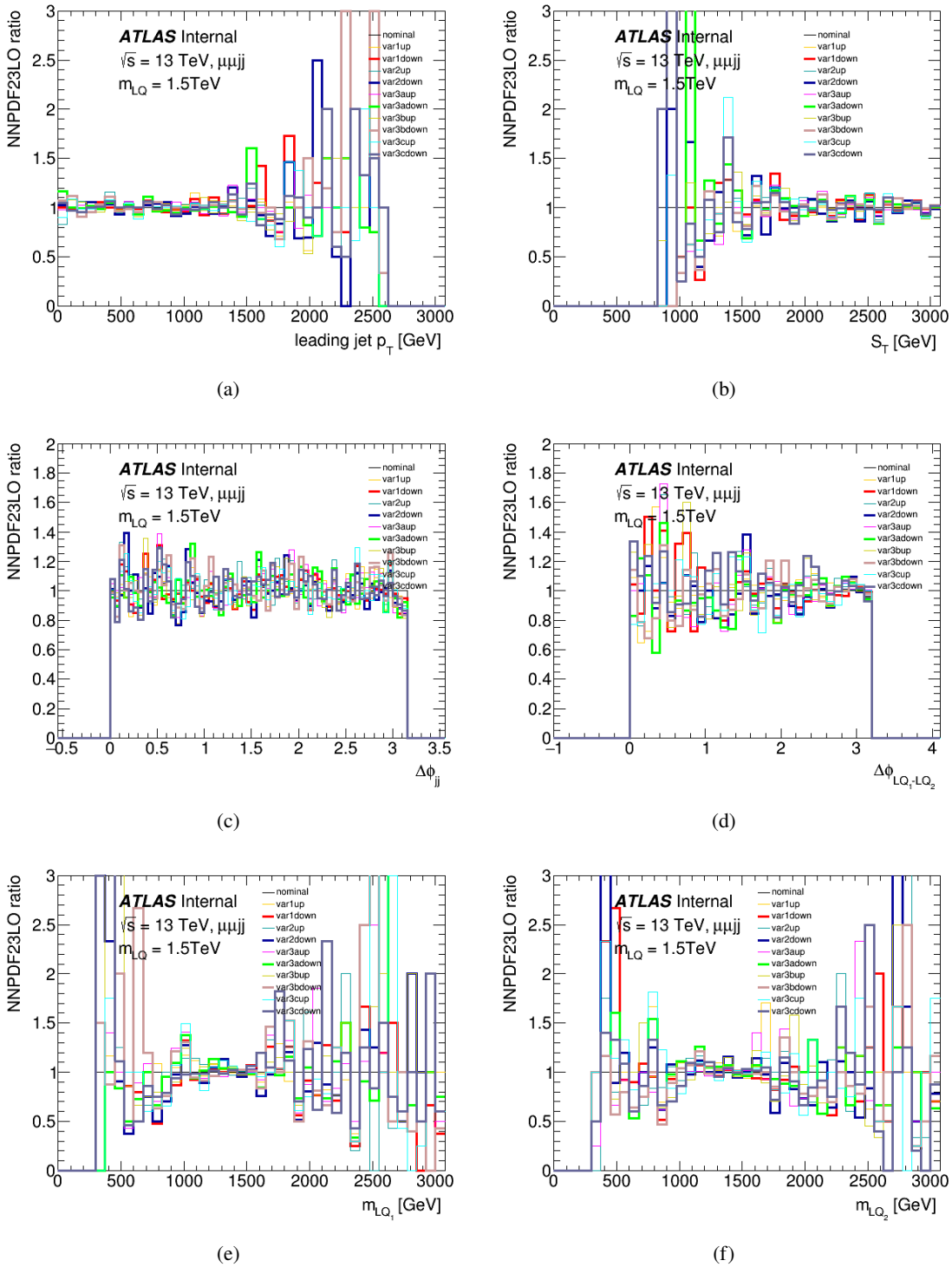


Figure B.9: Pythia8 signal ratio comparison for different NNPDF23LO variations at 1.5 TeV.

B.4 MadGraph - Pythia8 signal plots comparison

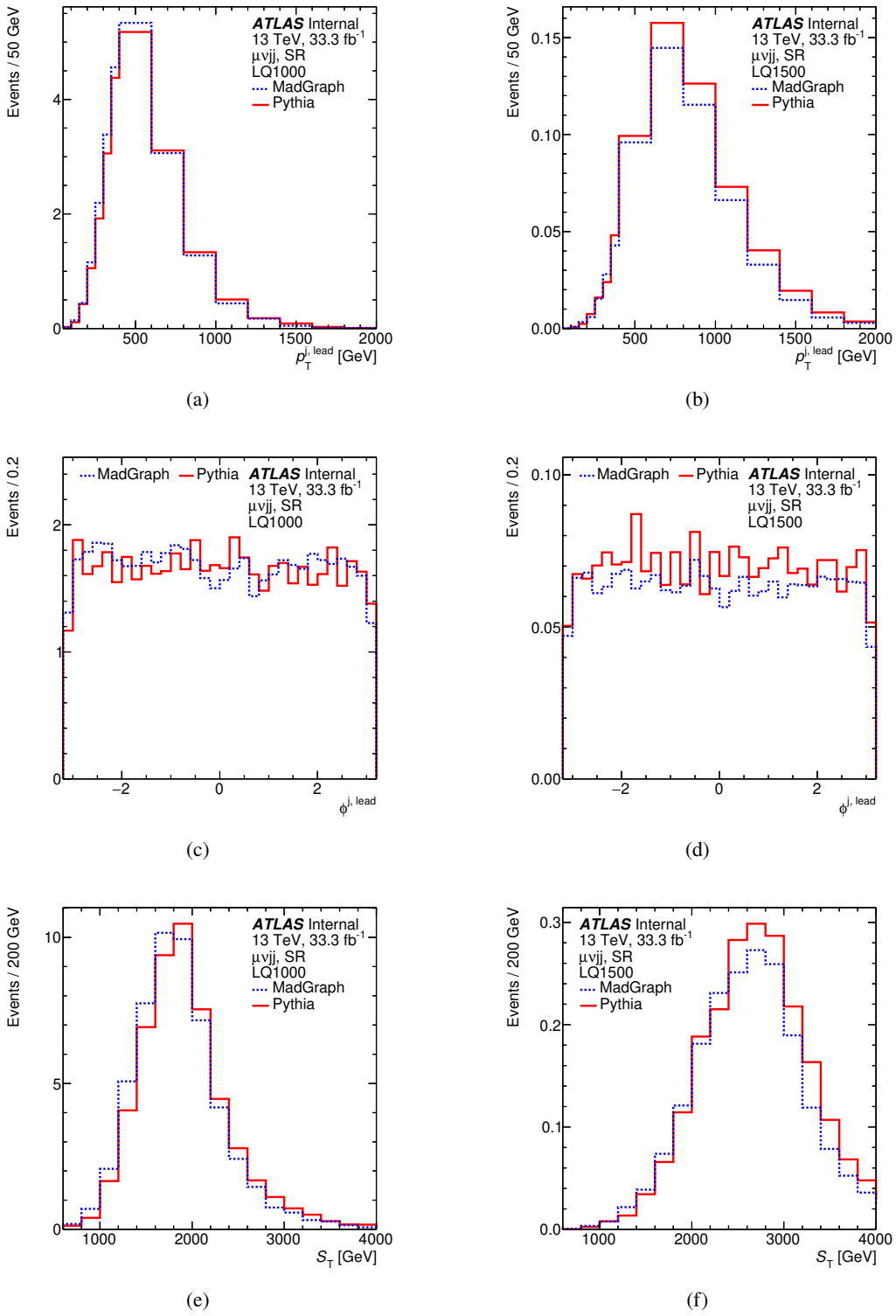


Figure B.10: Comparison of distributions for the MadGraph (blue) and Pythia8 (red) signal simulation of an LQ of 1 TeV (left) and 1.5 TeV (right) mass.

B.5 $t\bar{t}$ control region plots

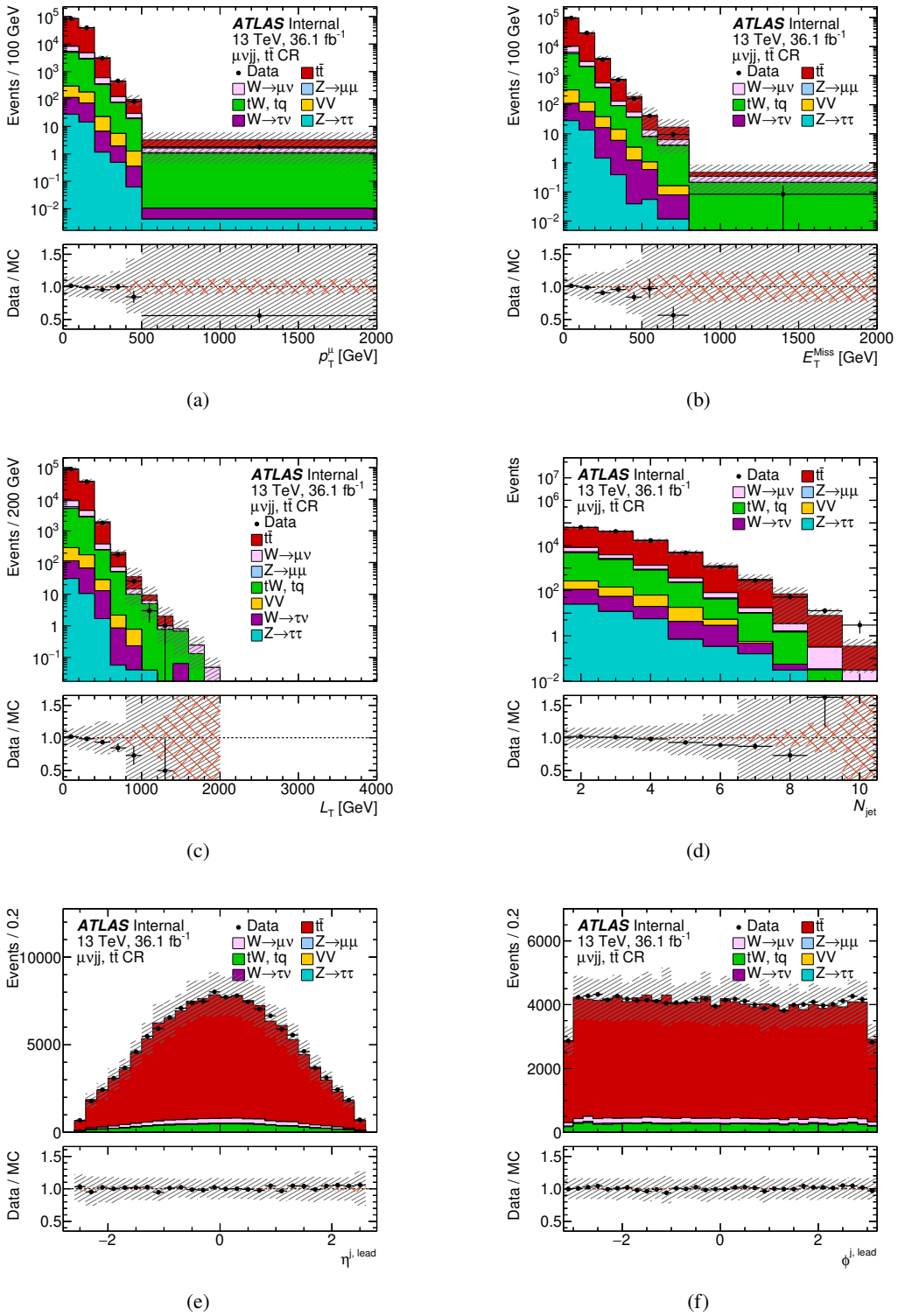
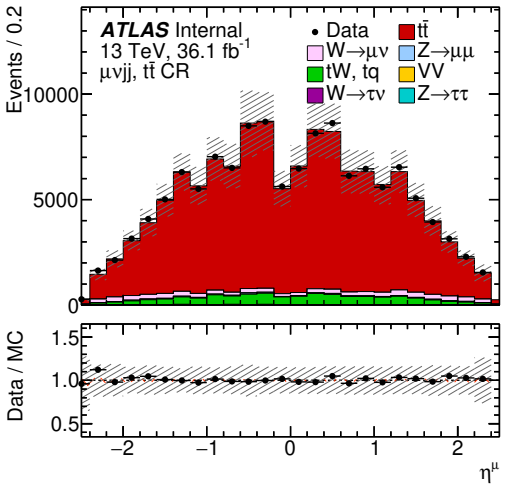
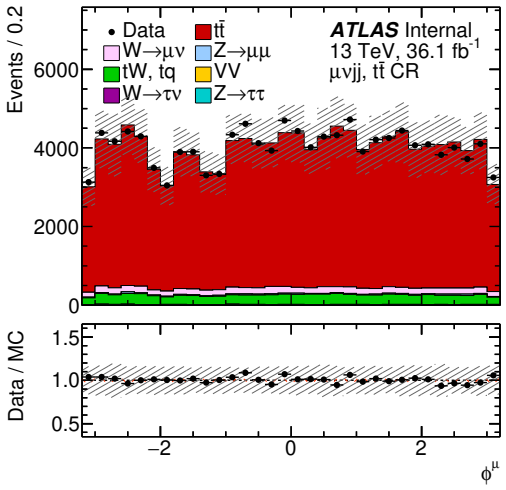


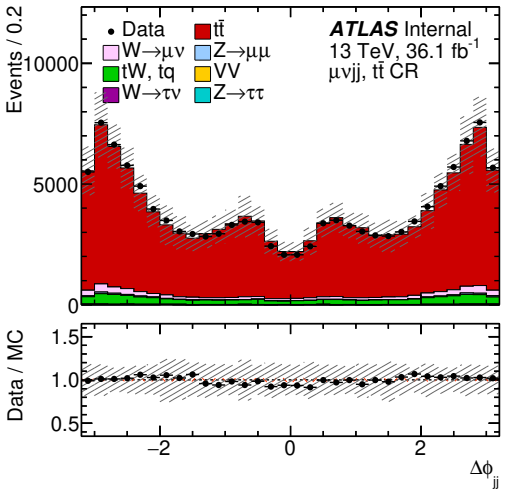
Figure B.11: Kinematic distributions in the $t\bar{t}$ CR. The grey band displays the total systematic uncertainty and the hatched orange band shows the statistical uncertainty in the simulation.



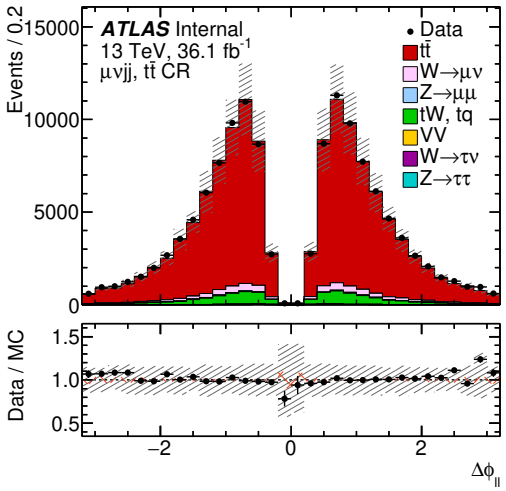
(a)



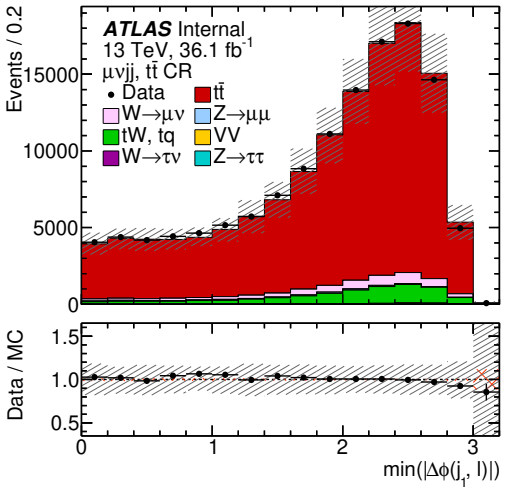
(b)



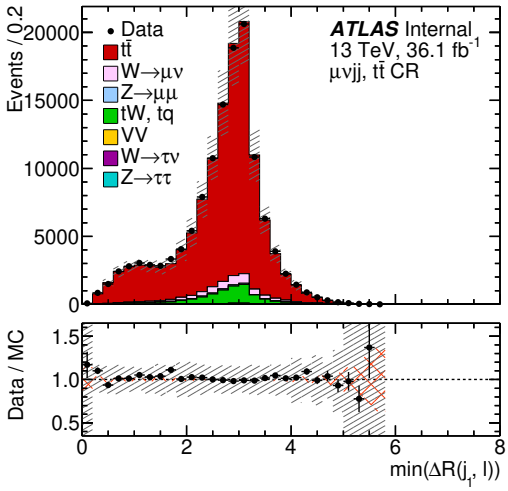
(c)



(d)



(e)



(f)

Figure B.12: Kinematic distributions in the $t\bar{t}$ CR. The grey band displays the total systematic uncertainty and the hatched orange band shows the statistical uncertainty in the simulation.

B.6 Signal - background: shape comparison

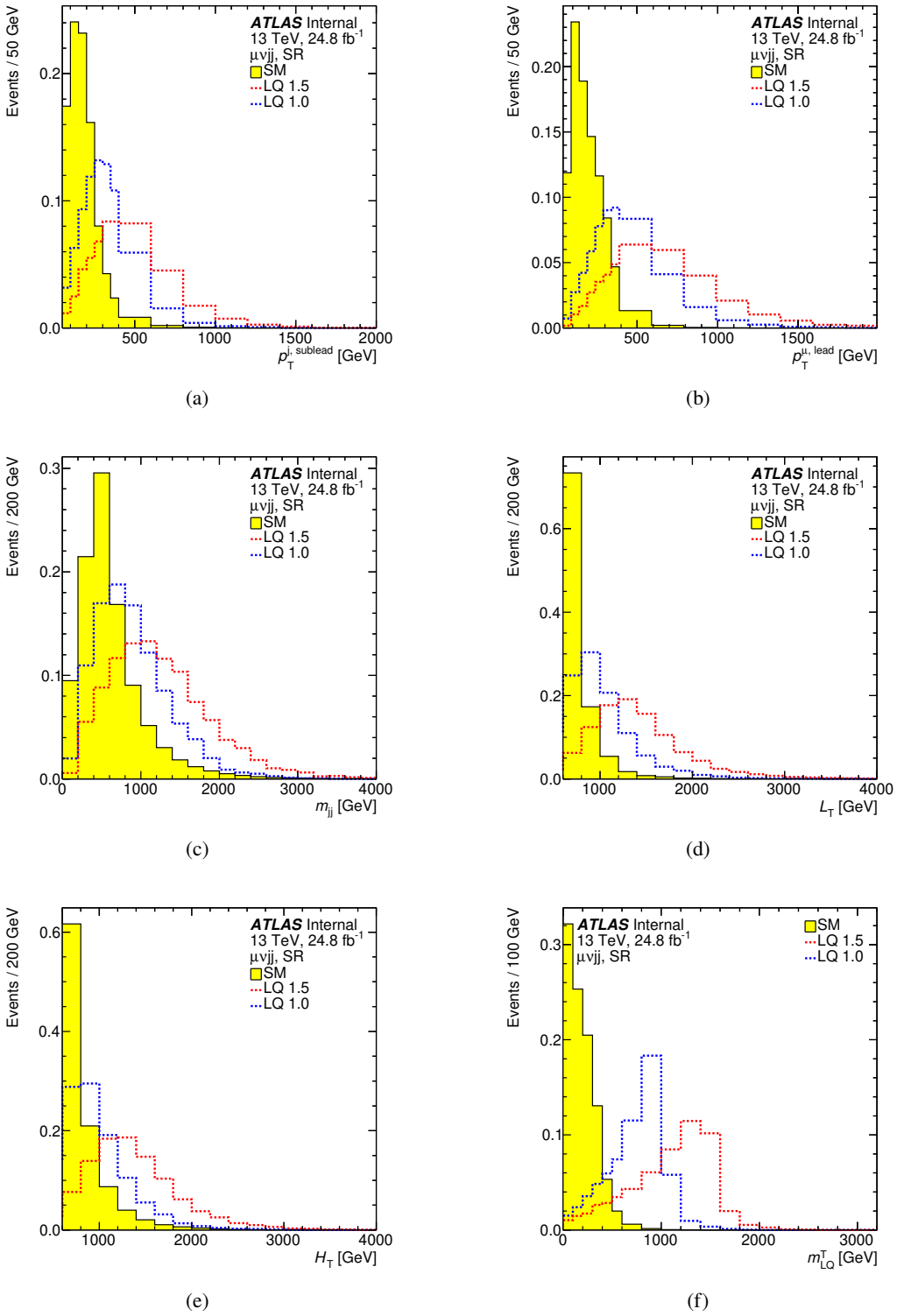


Figure B.13: Kinematic distributions in the 2015 SR.

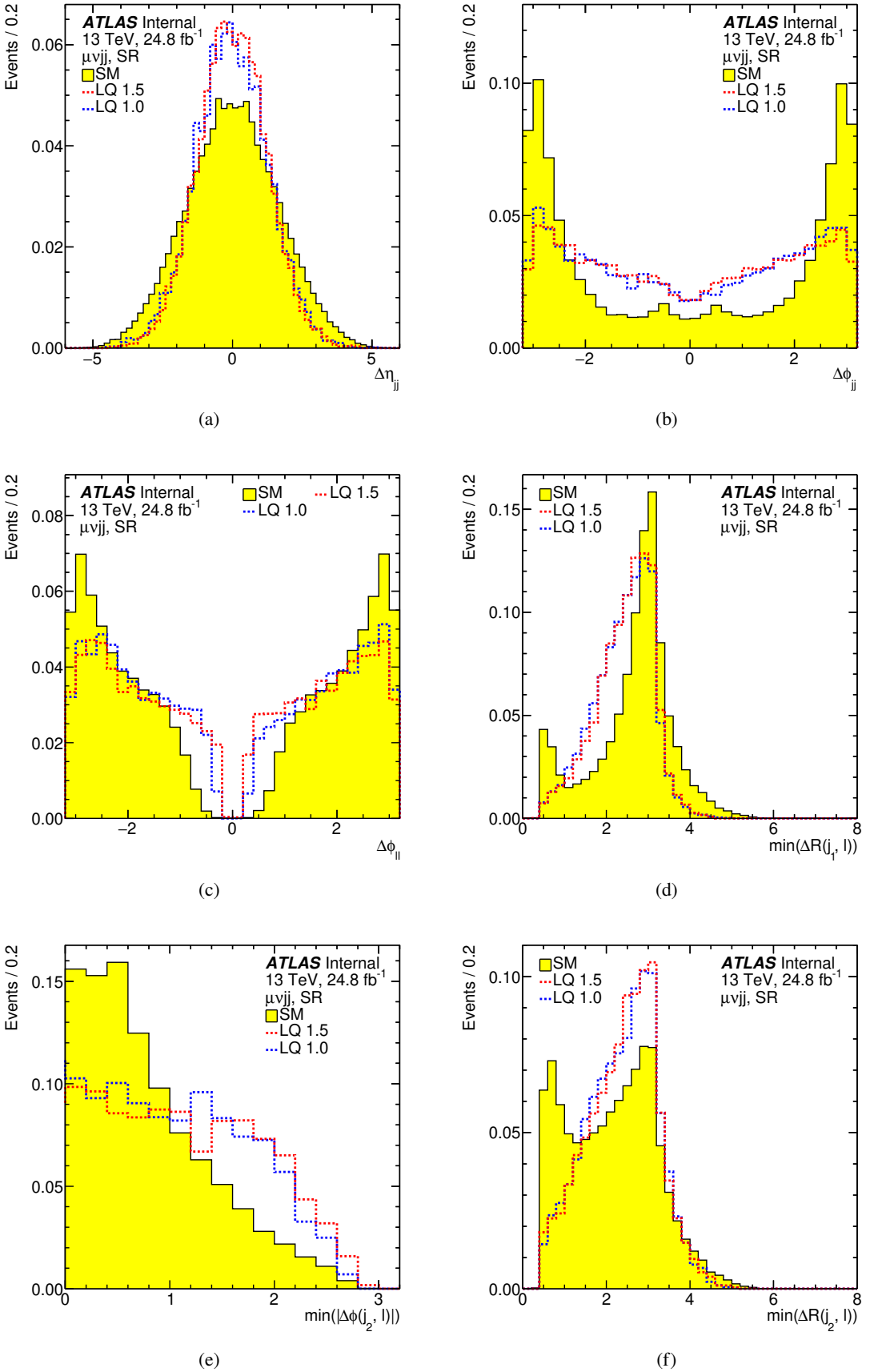


Figure B.14: Kinematic distributions in the 2015 SR.

B.7 Signal - background: absolute distributions comparison

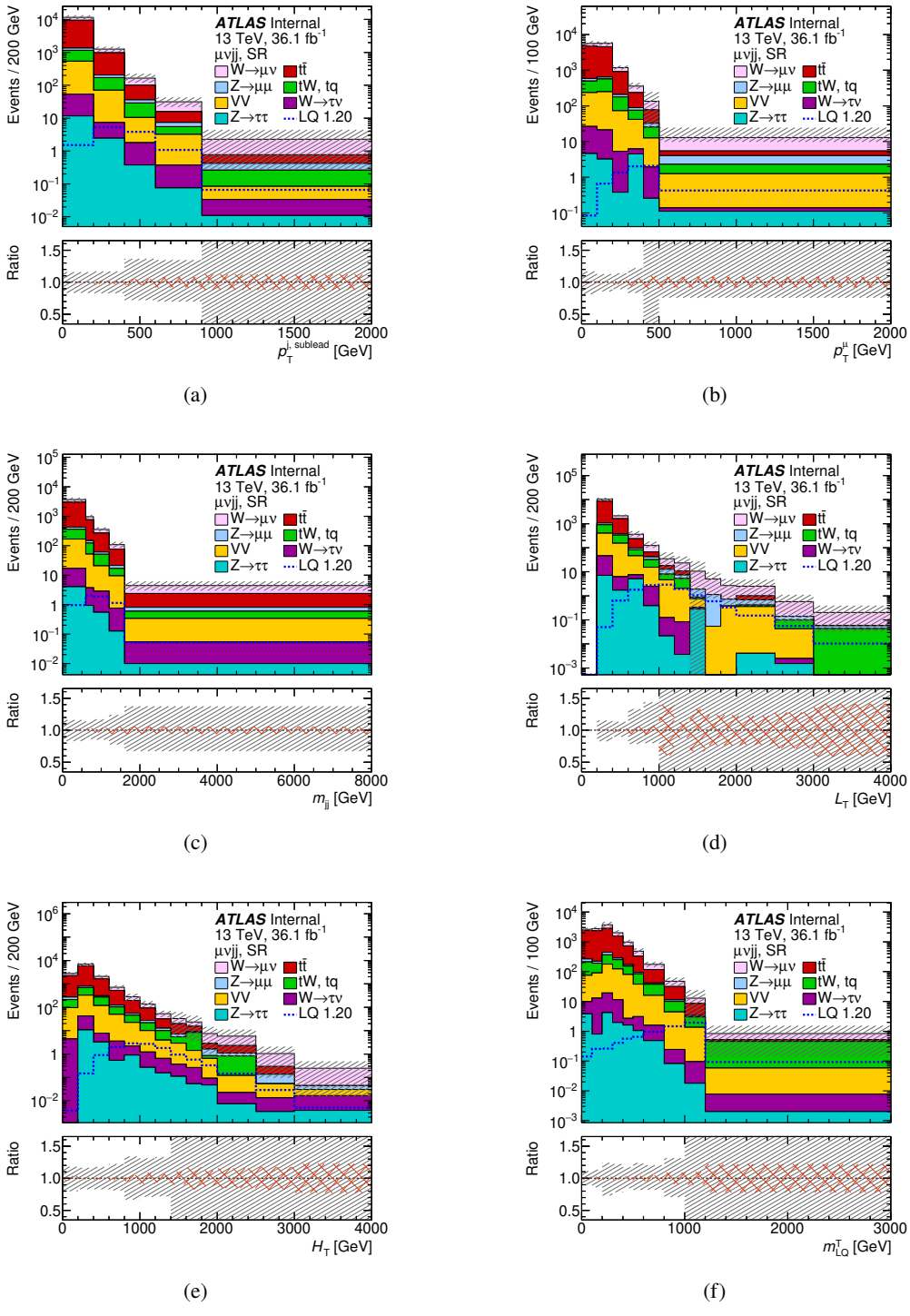
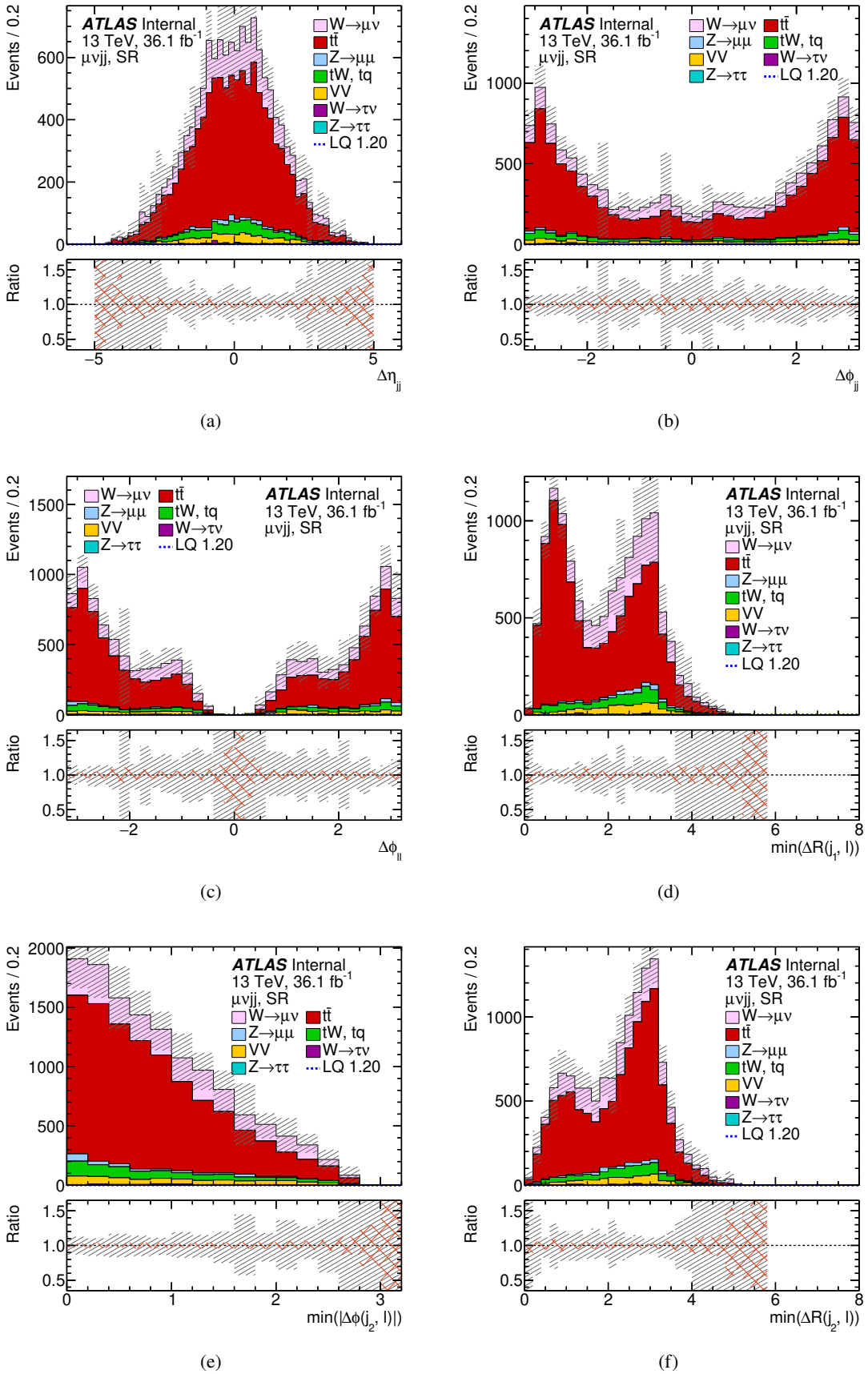


Figure B.15: Kinematic distributions in the 2015 SR. The grey band displays the total systematic uncertainty. The hatched orange band shows the statistical uncertainty in the simulation.



(a) (b) (c) (d) (e) (f)

Figure B.16: Kinematic distributions in the 2015 SR. The grey band displays the total systematic uncertainty. The hatched orange band shows the statistical uncertainty in the simulation.

B.8 V+jets reweighting

B.8.1 Reweighting Effect

Three variables have been investigated in the W+jets CR for the reweighting: leading jet p_T, H_T and m_{jj}. It is found that reweighting in either of the three variables improves the modelling of jet - related quantities, while leaving the di - lepton variables unchanged as desired. However, reweighting in leading jet p_T or H_T introduces a slight slope in the lepton p_T distributions. Moreover, these two variables do not improve the m_{jj} modelling as much as reweighting in m_{jj} improves the leading jet p_T or H_T distribution. Therefore, m_{jj} was chosen as the variable to use for deriving the weights. Overall, the reweighting results in a good description of the relevant variables in the control regions as they appear in Fig. B.17.

B.8.2 Validation

The signal regions differ from the V+jets control regions only in the cut on m_T. To validate the reweighting it is thus tested how weights derived in a certain region of m_T affect the modelling in a different region of m_T. For this test, the weights are derived in "tight" control regions that are more restrictive in mass than the nominal control regions: for W+jets m_T is required to lie between 60 GeV and 100 GeV. The validation regions are then the remaining phase space between these tight CRs and the signal regions, i.e. $100 < m_T\ /\text{GeV} < 130$.

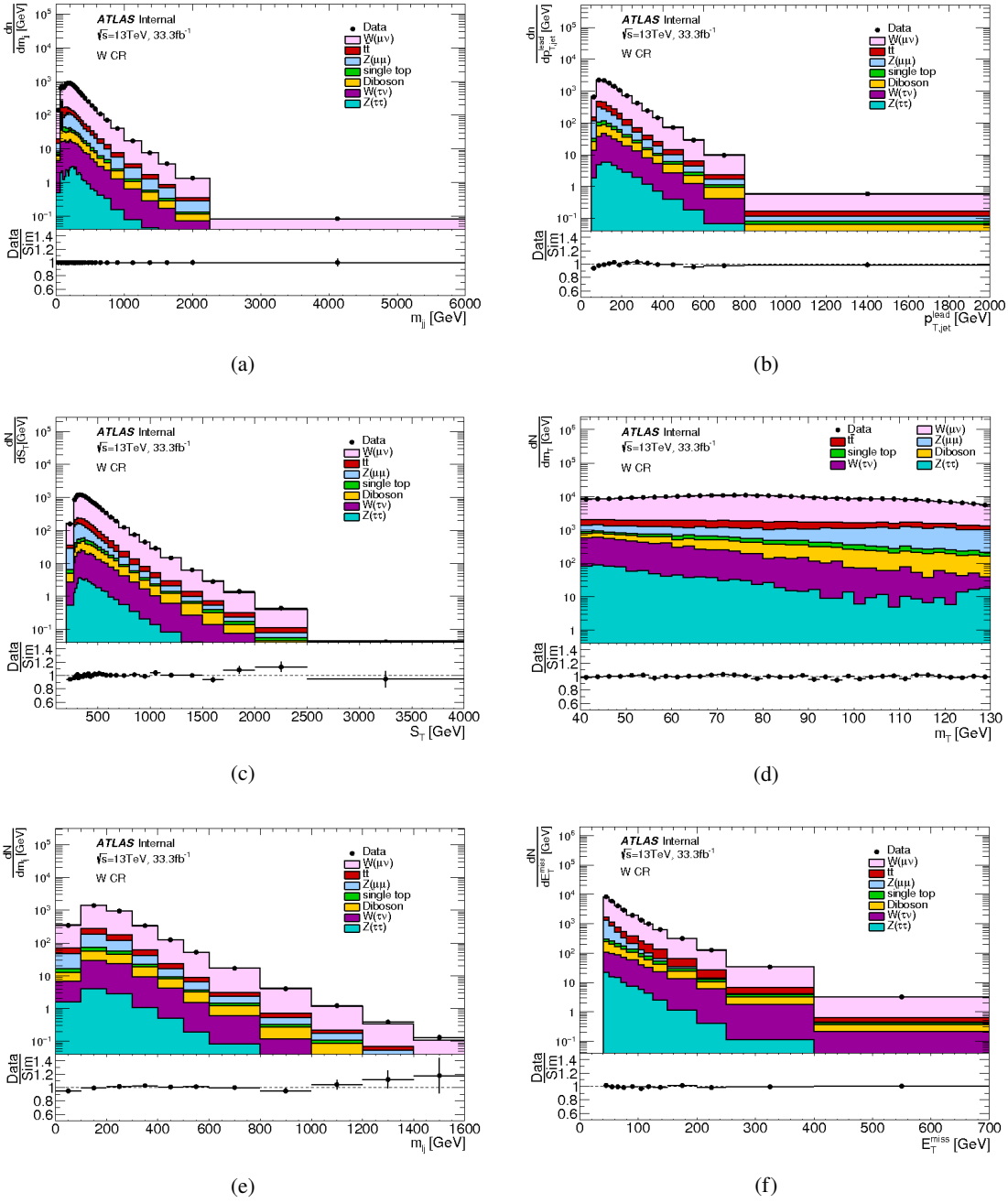
Fig. B.18, B.19 and B.20 show the outcome of this test for the W+jets background. The top row in these figures is always showing the tight CR and the bottom row the validation region. The distributions of m_{jj}, S_T and leading jet p_T are compared before (left) and after (right) reweighting in m_{jj}. The error bars shown include only statistical uncertainties. Again, by construction, the data - to - simulation ratio in m_{jj} is flat at 1 after applying the weights in the tight CR, as this is where the weights are derived for this test. The description of the data is improved for all variables in a similar way in all regions.

Based on these studies, it is concluded that the behaviour is similar in the different mass regions, i.e. it is a viable approach to derive the weights in the CR and apply them in the SR.

B.8.3 Final Weights

For the final weights, the ratio of data to simulation is fitted (using ROOT TH1::Fit() method) to be more independent of the binning. The turn - on at low values of m_{jj} and the tail are fitted separately. Different combinations of functional forms have been tested as shown in Fig. B.21. For the low - mass turn - on, the second order polynomial and a Gaussian provide almost identical results and the polynomial is chosen to obtain the weights. For the fit at higher masses, the second order polynomial is chosen as well, since it provides the smoothest transition to the low - mass region. The stitching point is at 400 GeV.

The top row of Fig. B.22 shows the m_{jj} distribution in the W CR after applying reweighting based on the fits. The description of data is much improved compared to the original situation shown for reference in the bottom row.



(a) (b) (c) (d) (e) (f)

Figure B.17: Distributions of kinematic variables in the W+jets CR after bin - by - bin reweighting in m_{jj}. The error bars reflect only the statistical uncertainty.

B.8.4 Control Plots with Reweighting Applied

In this section side - by - side comparisons of some control distributions before and after the reweighting are shown (Fig. B.23 - B.31).

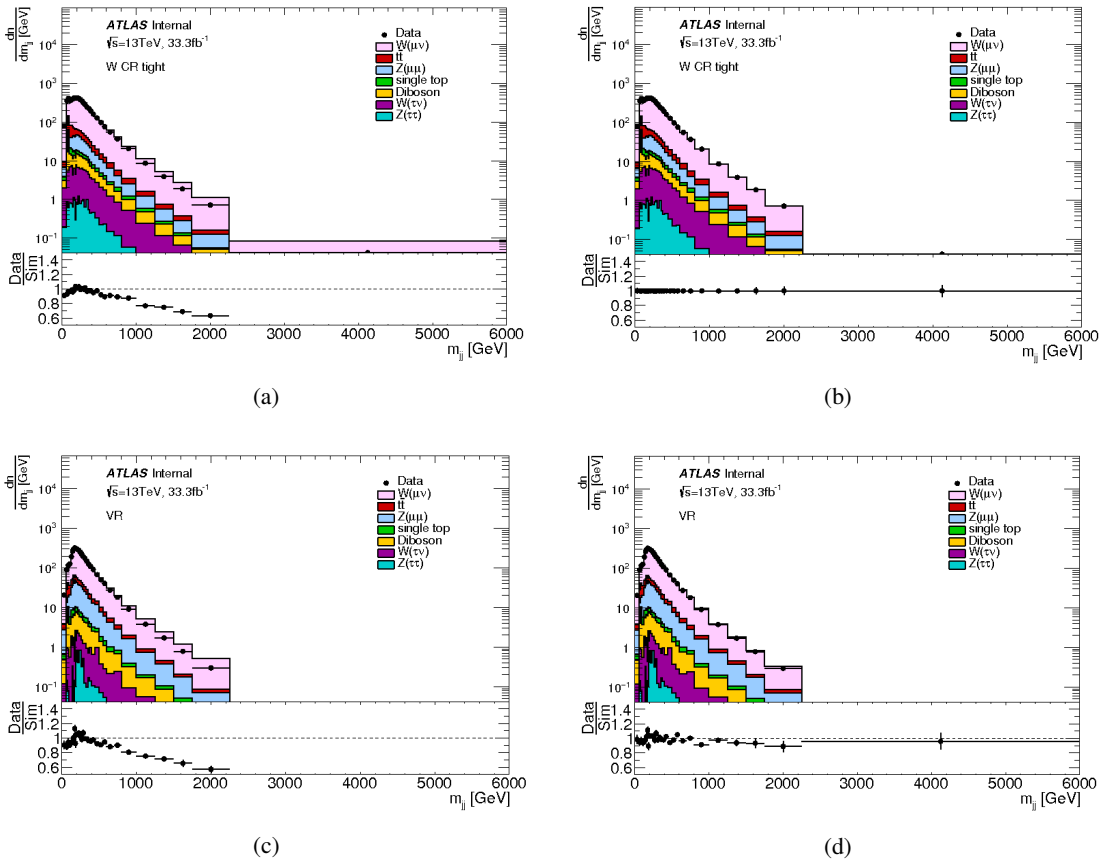
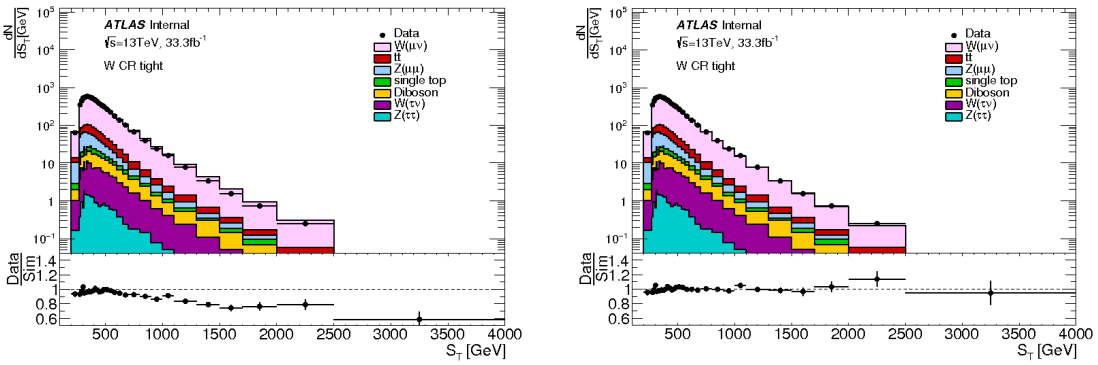
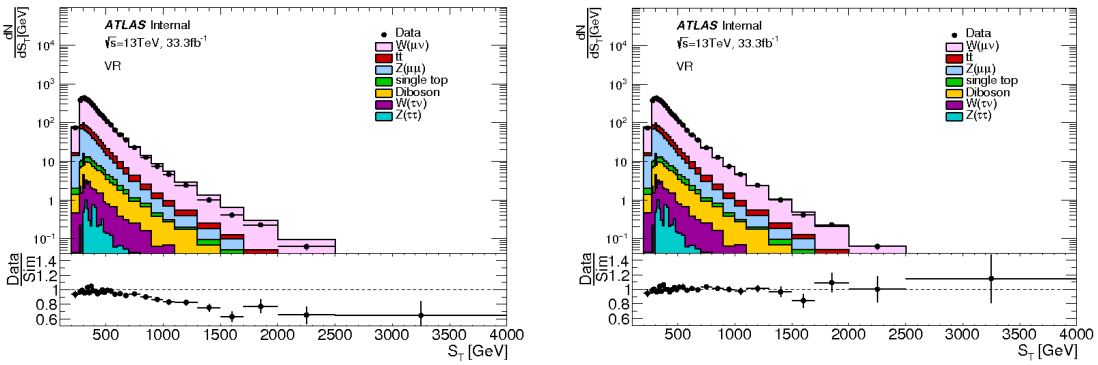


Figure B.18: Distributions of m_{jj} in the tight W CR (top) and the validation region (bottom) before (left) and after (right) reweighting in m_{jj}. The error bars reflect only the statistical uncertainty.



(a) (b)



(c) (d)

Figure B.19: Distributions of S_T in the tight W CR (top) and the validation region (bottom) before (left) and after (right) reweighting in m_{jj}. The error bars reflect only the statistical uncertainty.

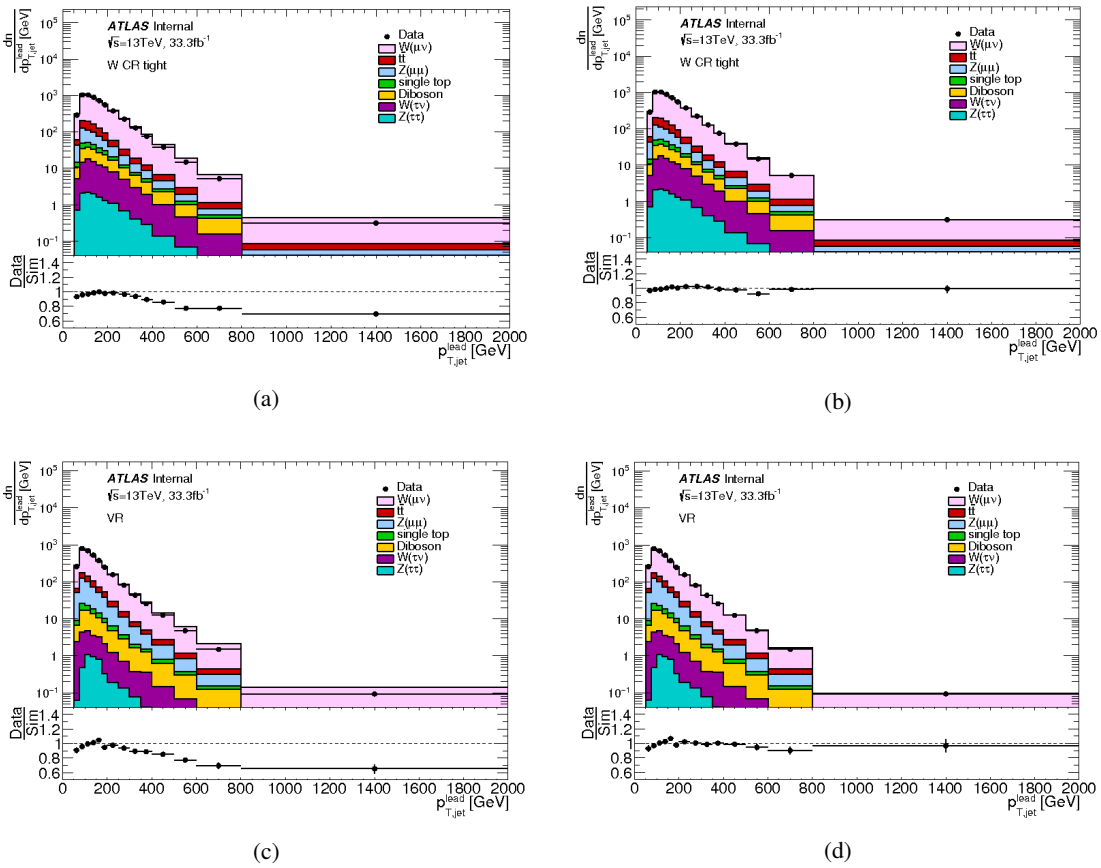


Figure B.20: Distributions of leading jet p_T in the tight W CR (top) and the validation region (bottom) before (left) and after (right) reweighting in m_{jj}. The error bars reflect only the statistical uncertainty.

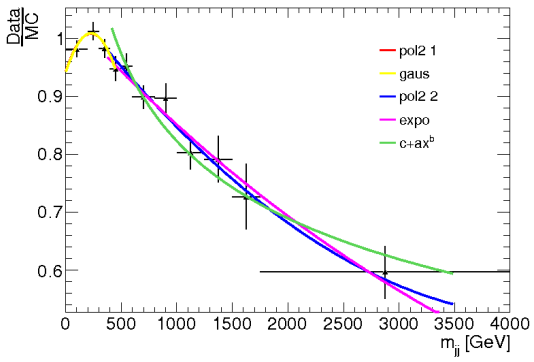
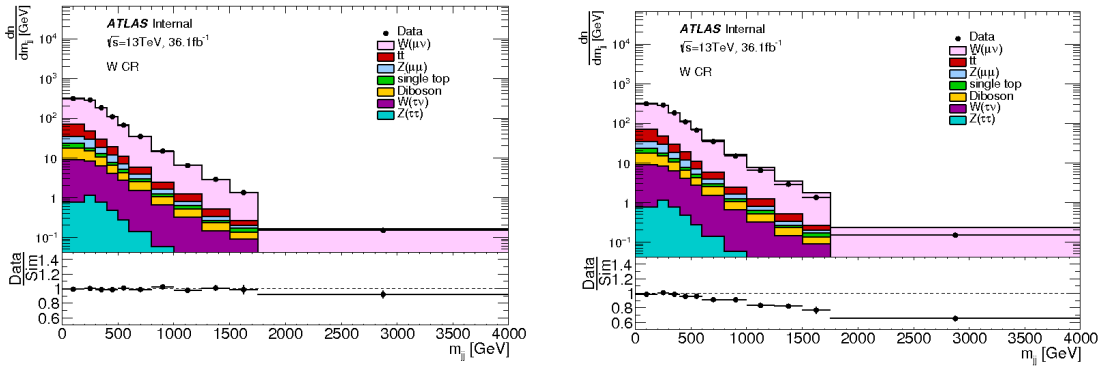


Figure B.21: Fits to the ratio of data and simulation in the W CR using different functional forms.



(a) (b)

Figure B.22: Distribution of m_{jj} after the reweighting based on the fit (left) for the W CR. On the right appears the distribution before any reweighting for reference.

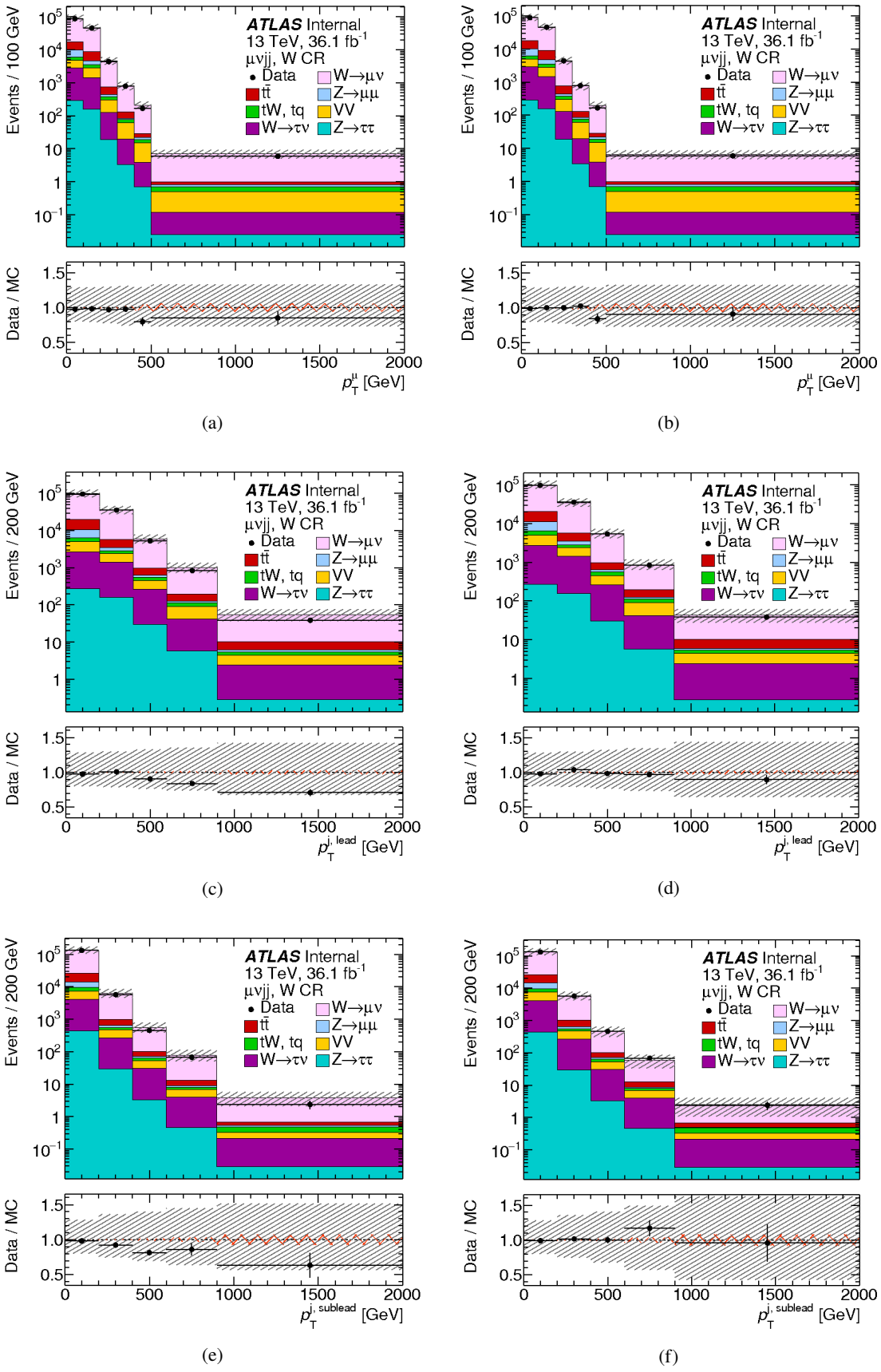


Figure B.23: Kinematic distributions in the W CR before (left) and after (right) reweighting in m_{jj}.

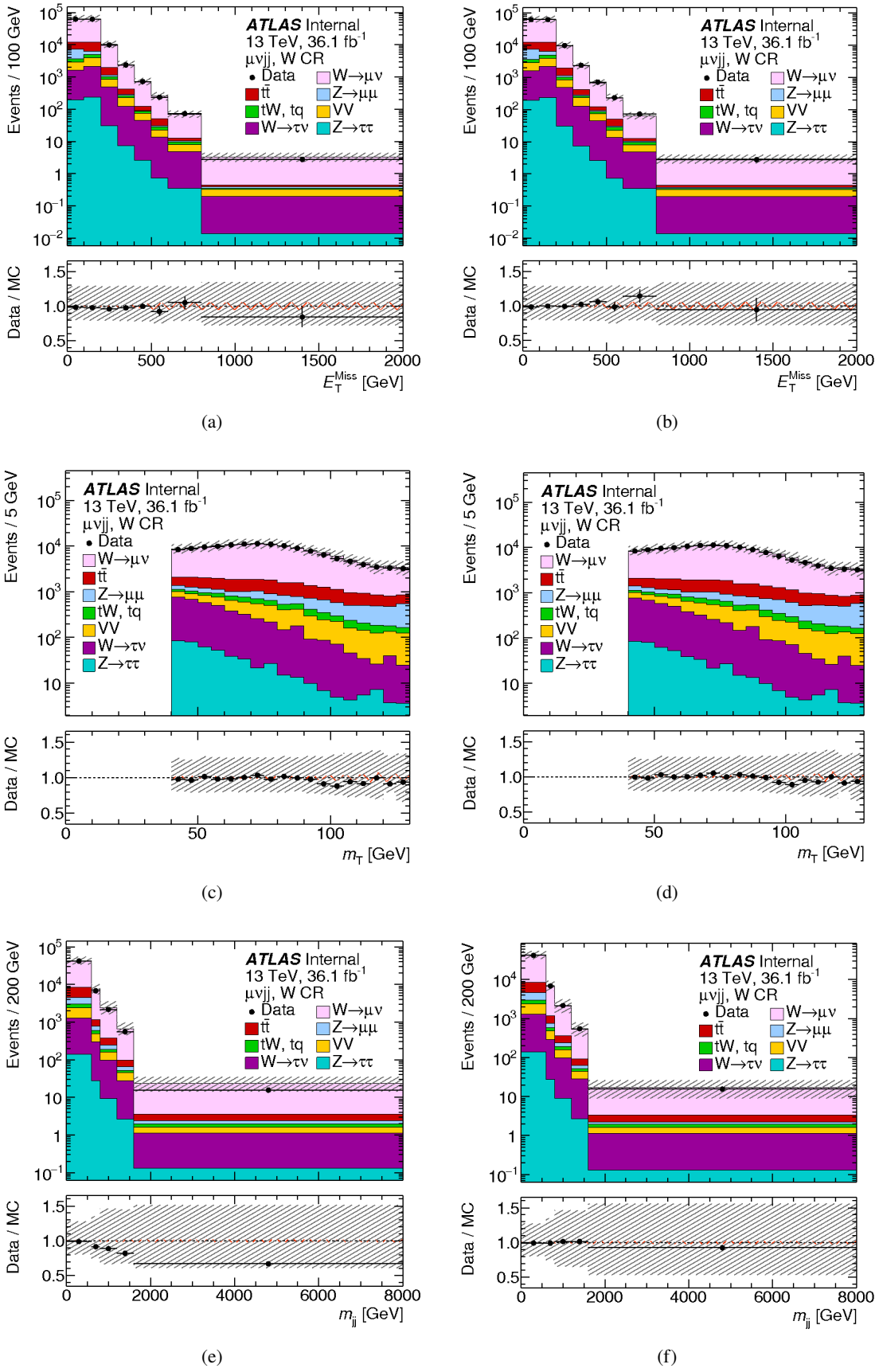


Figure B.24: Kinematic distributions in the W CR before (left) and after (right) reweighting in m_{jj}.

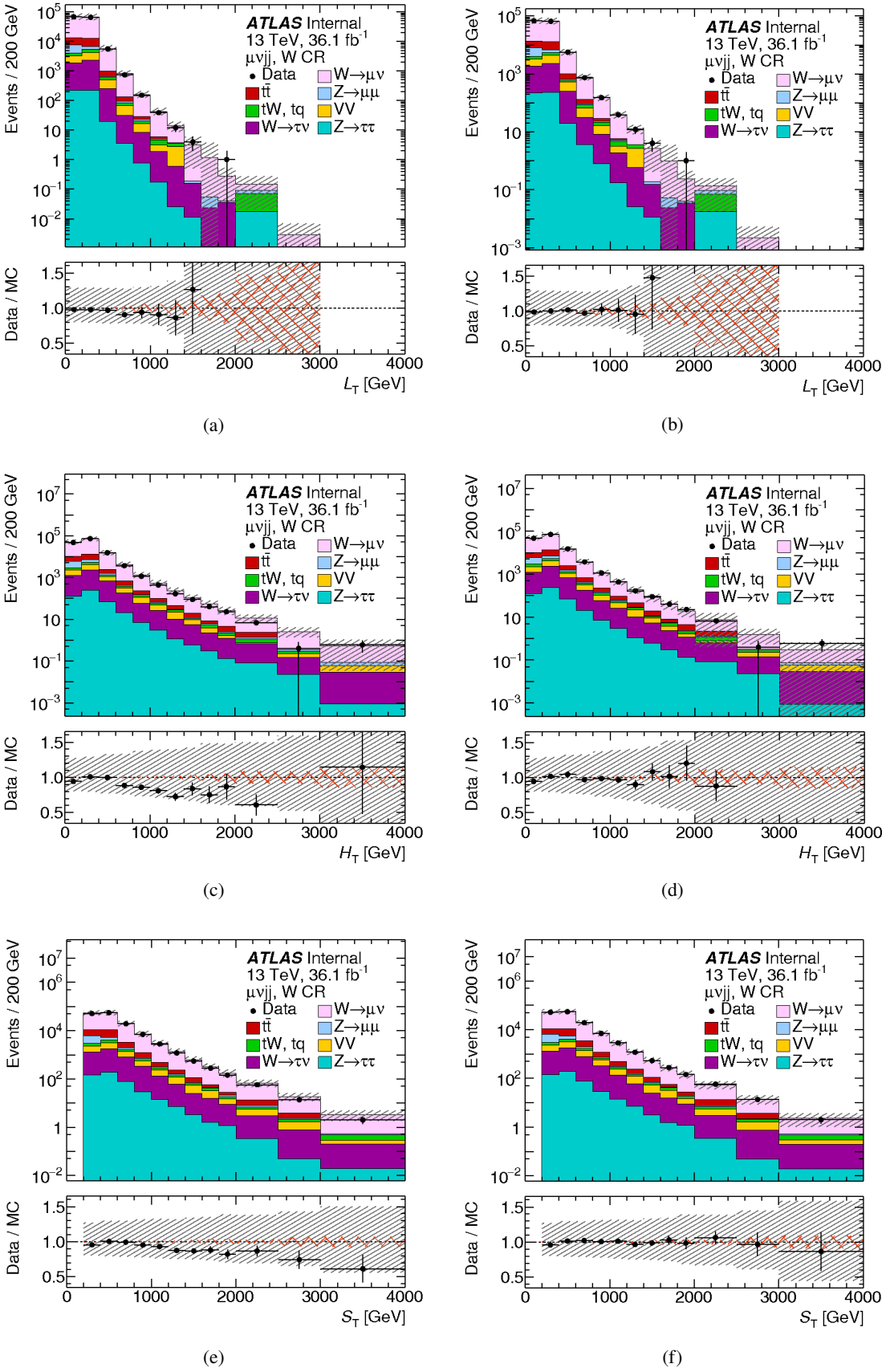


Figure B.25: Kinematic distributions in the W CR before (left) and after (right) reweighting in m_{jj}.

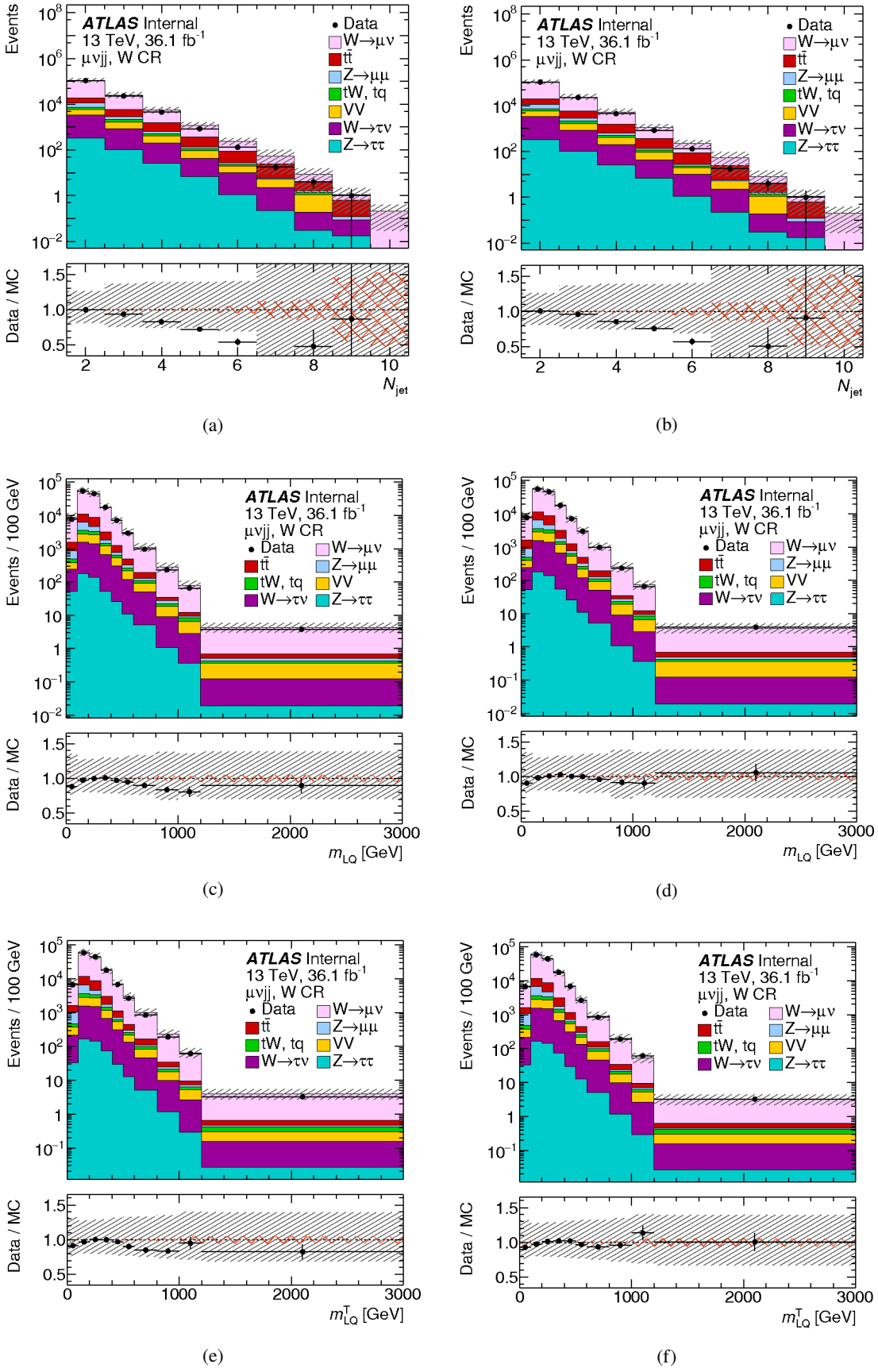


Figure B.26: Kinematic distributions in the W CR before (left) and after (right) reweighting in m_{jj}.

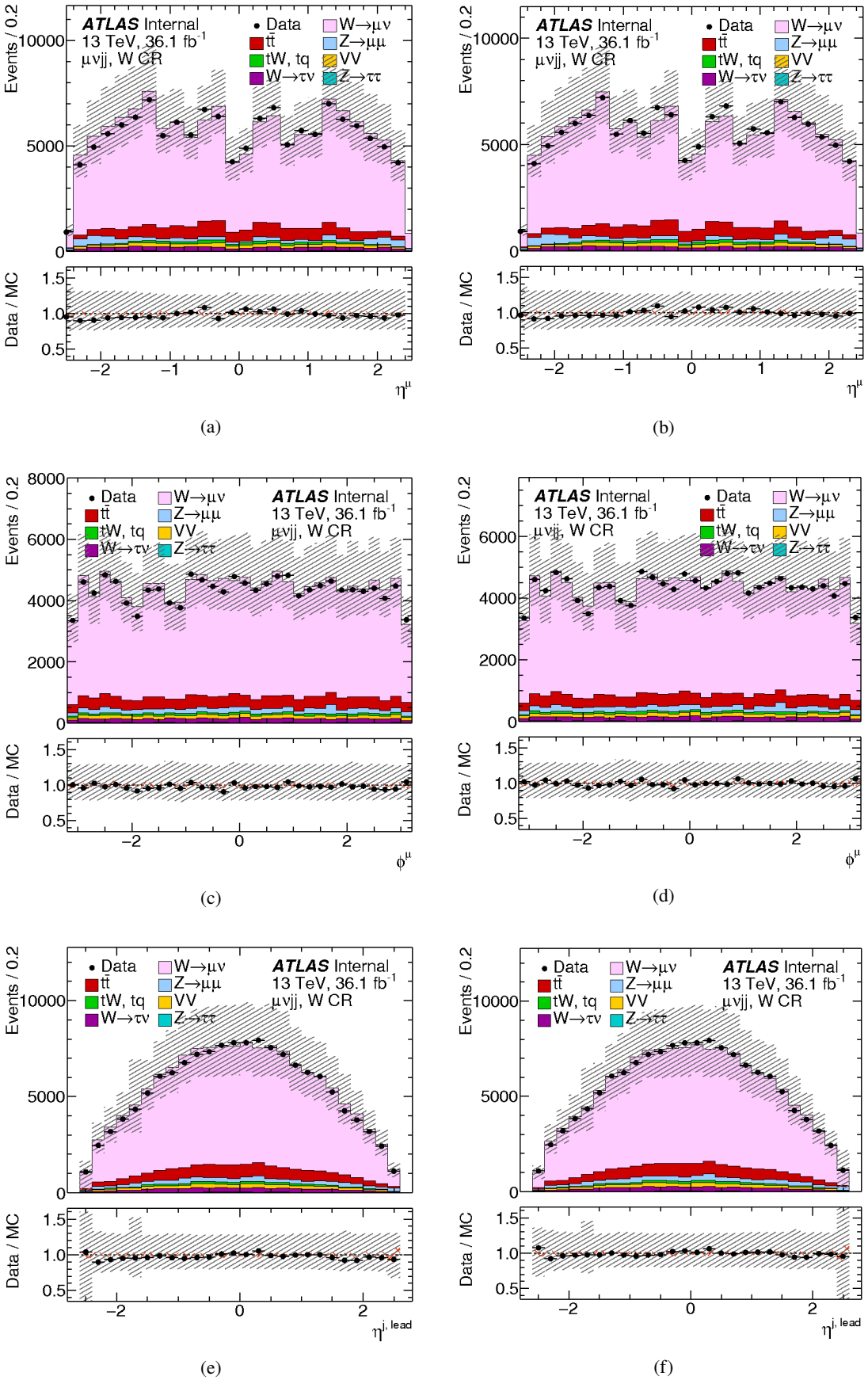


Figure B.27: Kinematic distributions in the W CR before (left) and after (right) reweighting in m_{jj}.

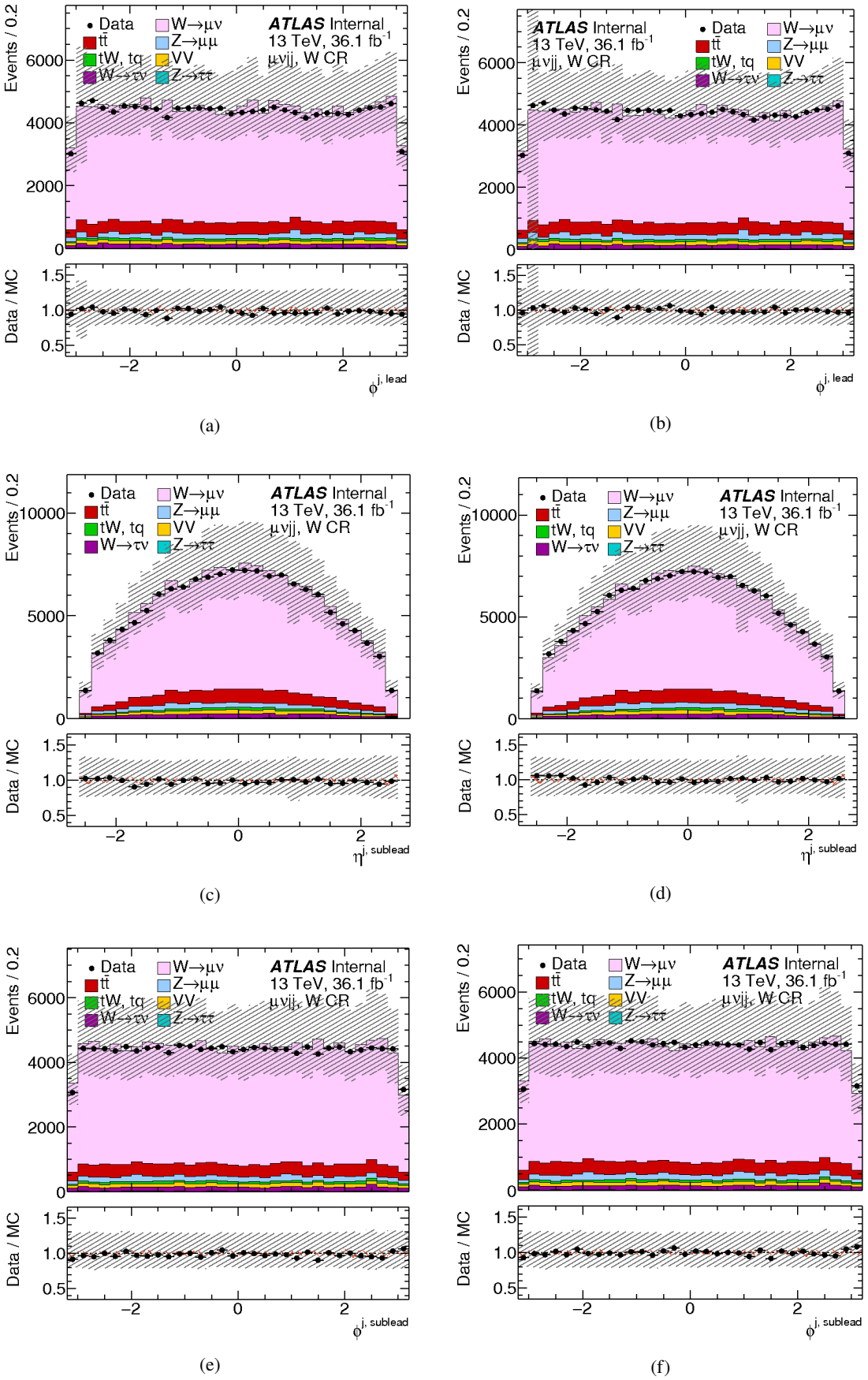
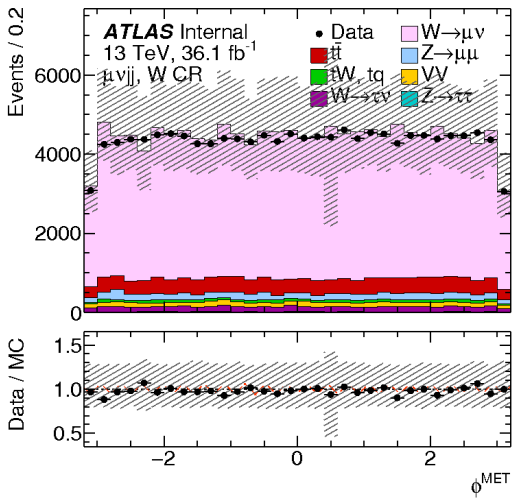
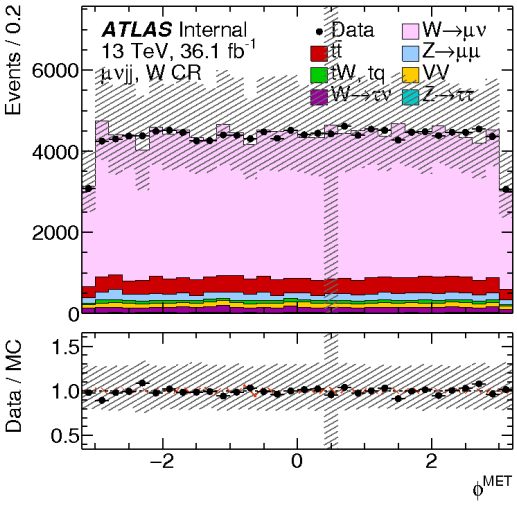


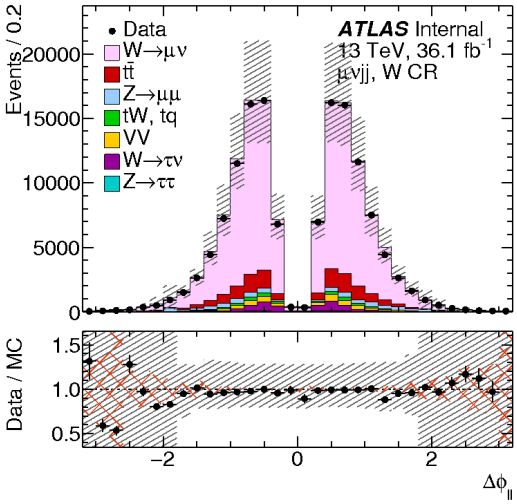
Figure B.28: Kinematic distributions in the W CR before (left) and after (right) reweighting in m_{jj}.



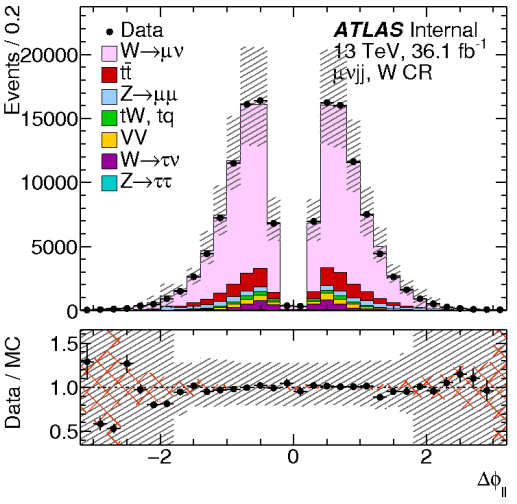
(a)



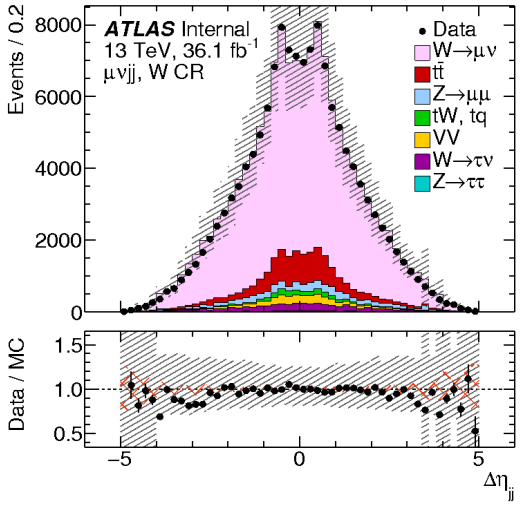
(b)



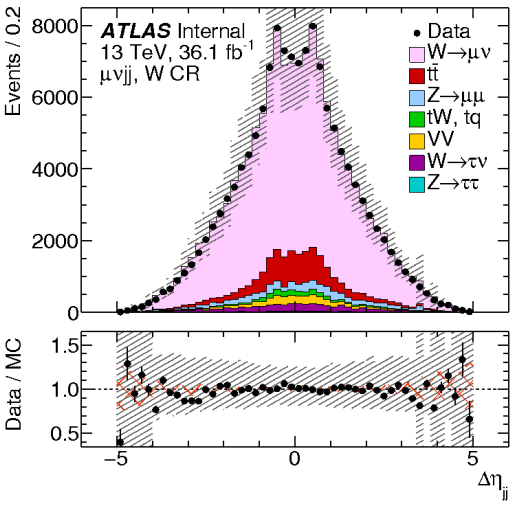
(c)



(d)

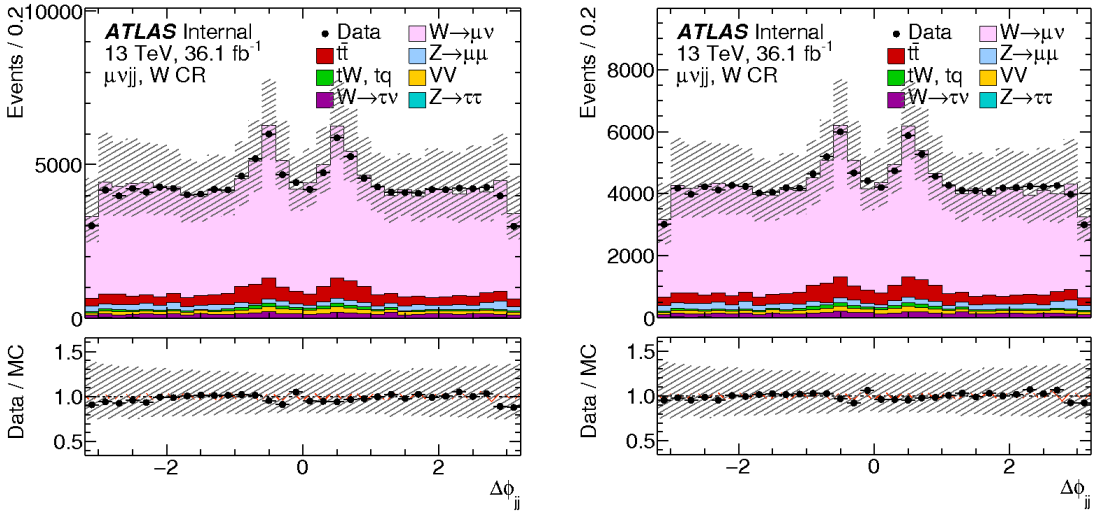


(e)

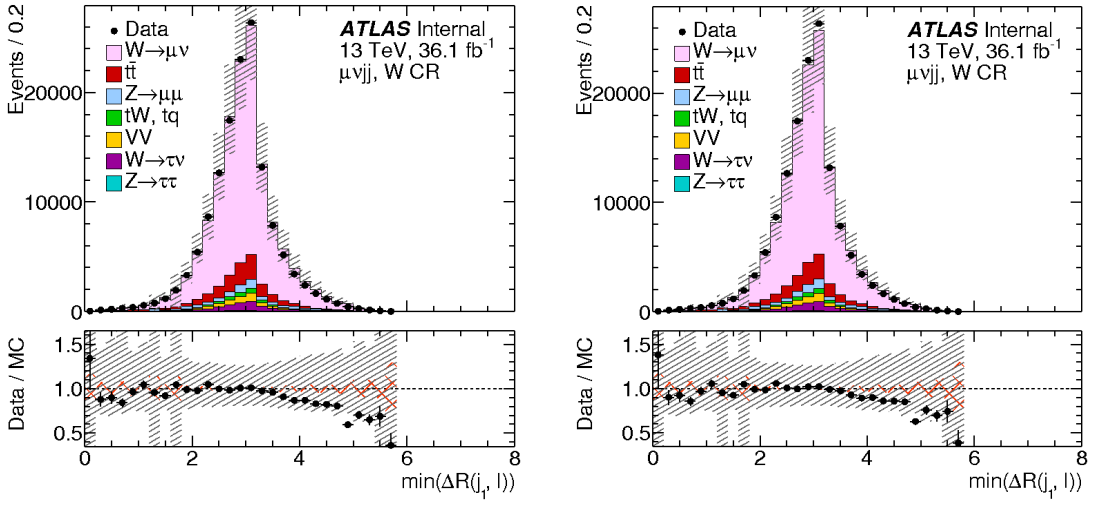


(f)

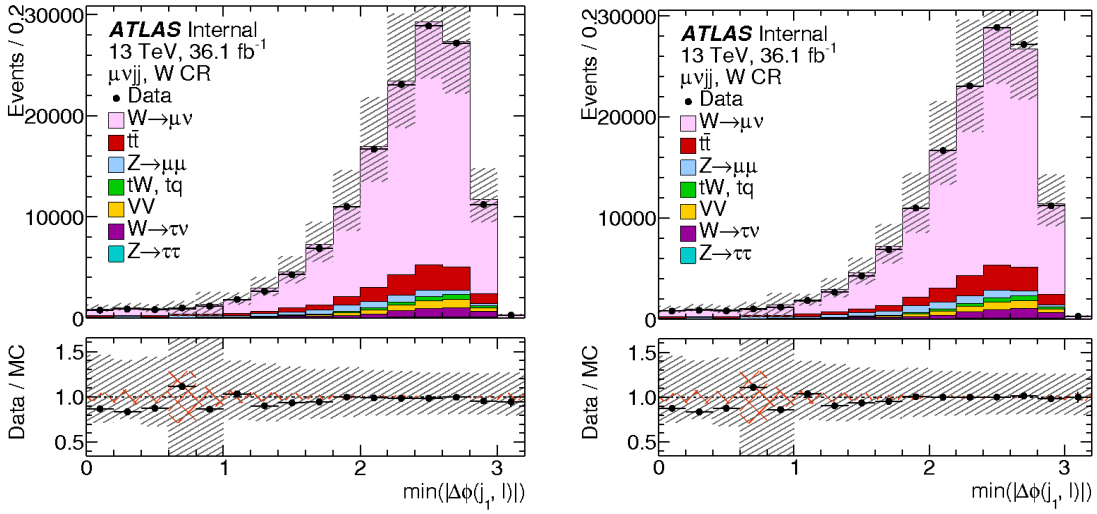
Figure B.29: Kinematic distributions in the W CR before (left) and after (right) reweighting in m_{jj}.



(a) (b)



(c) (d)



(e) (f)

Figure B.30: Kinematic distributions in the W CR before (left) and after (right) reweighting in m_{jj}.

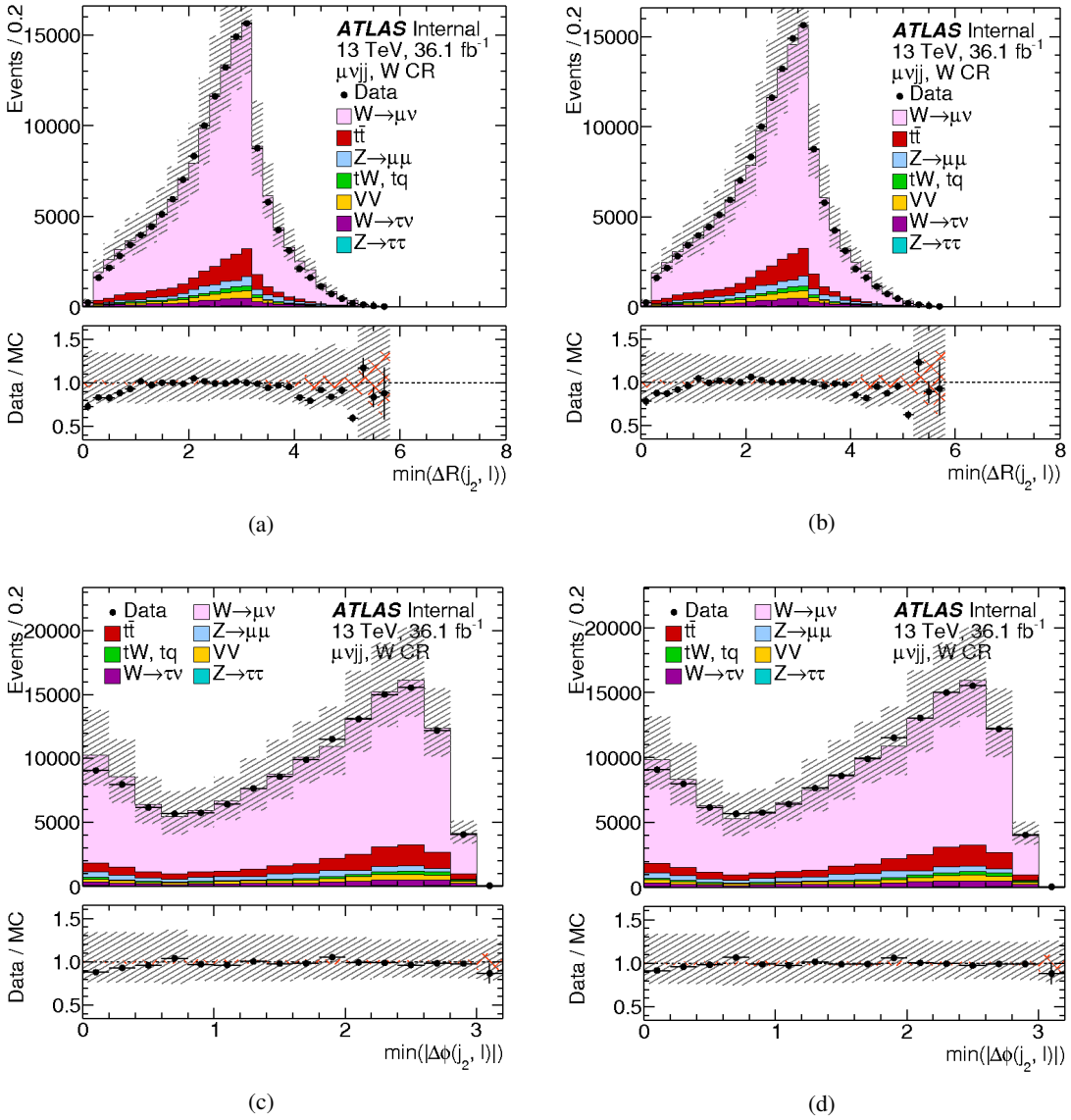


Figure B.31: Kinematic distributions in the W CR before (left) and after (right) reweighting in m_{jj}.

B.9 BDT input variables correlation

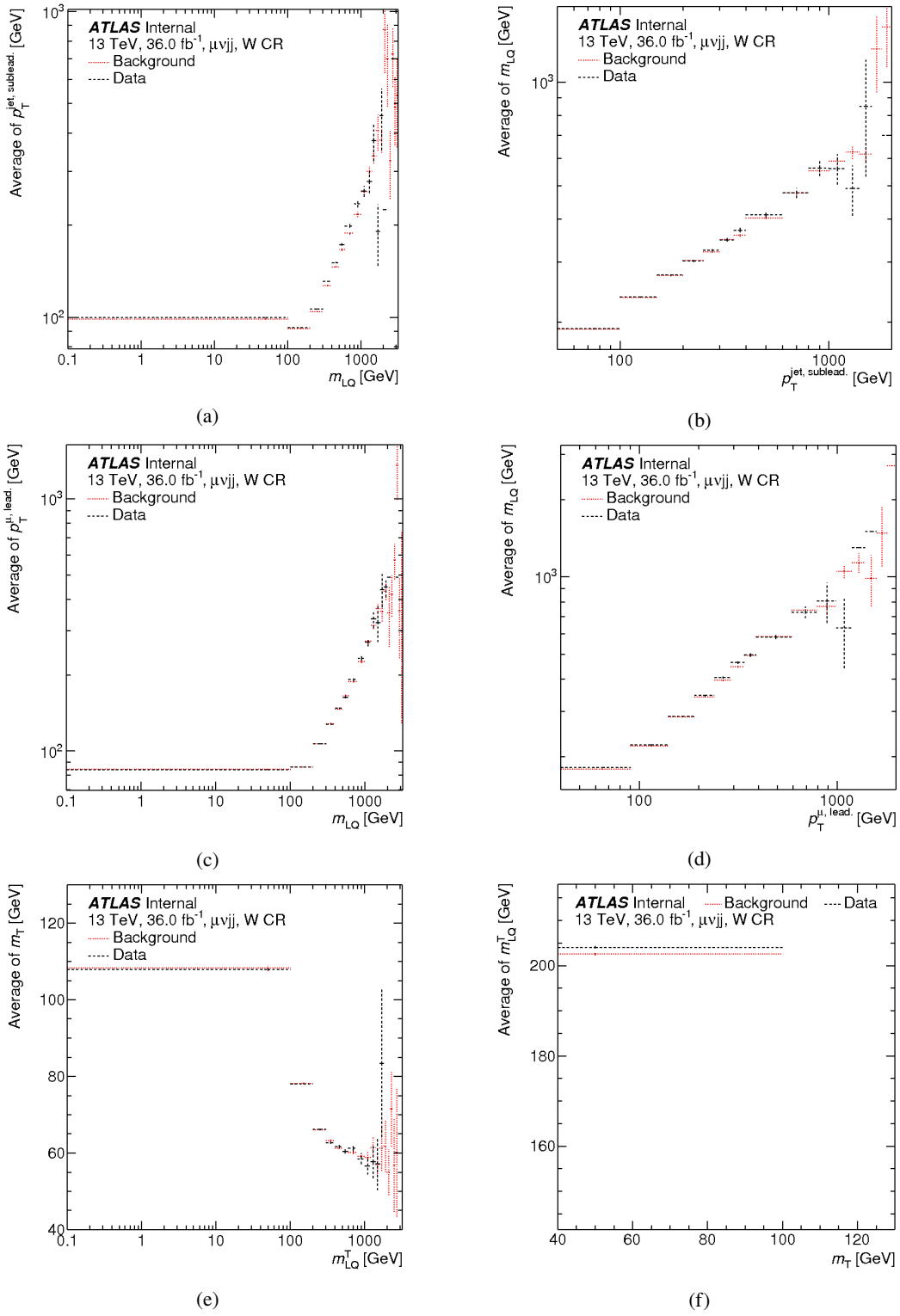
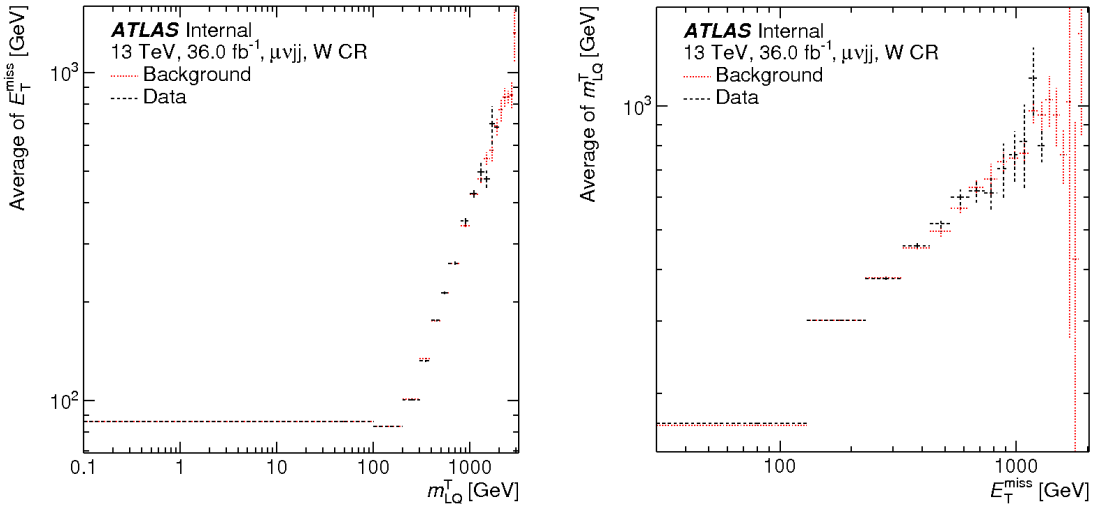
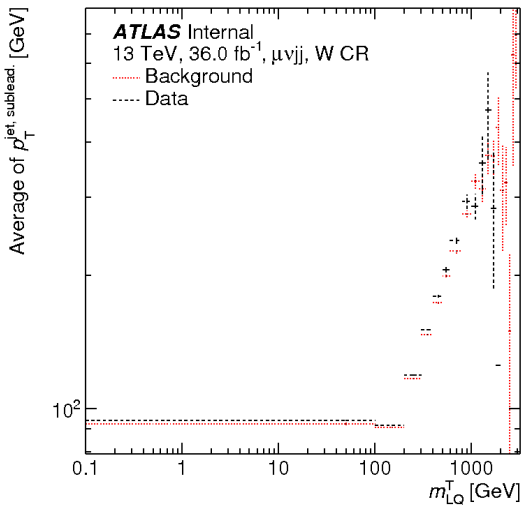


Figure B.32: Correlation of BDT input variables in the W CR.

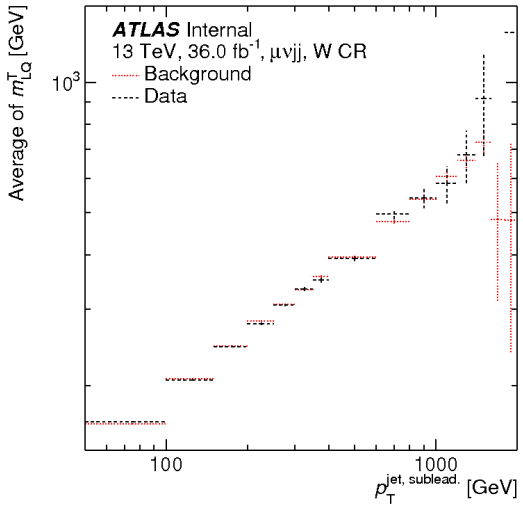


(a)

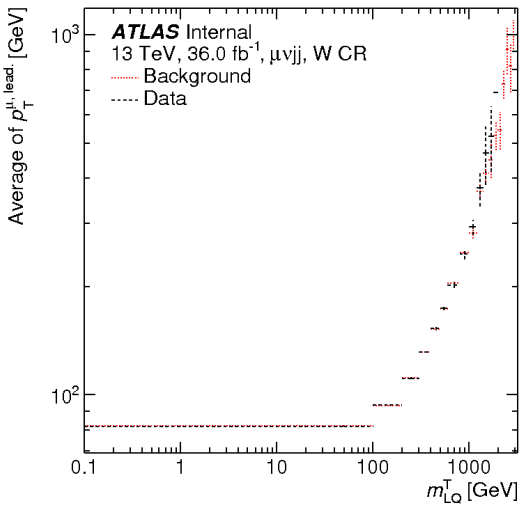
(b)



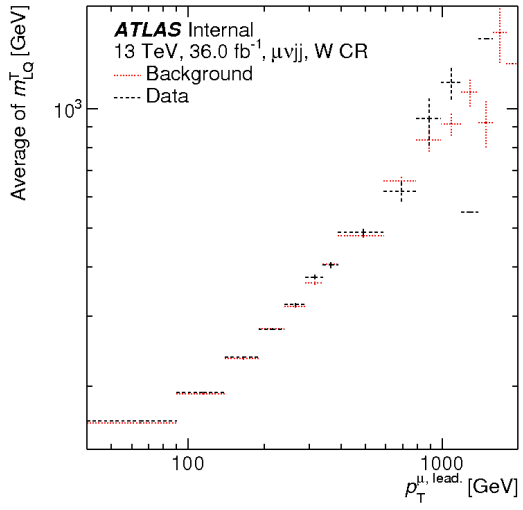
(c)



(d)



(e)



(f)

Figure B.33: Correlation of BDT input variables in the W CR.

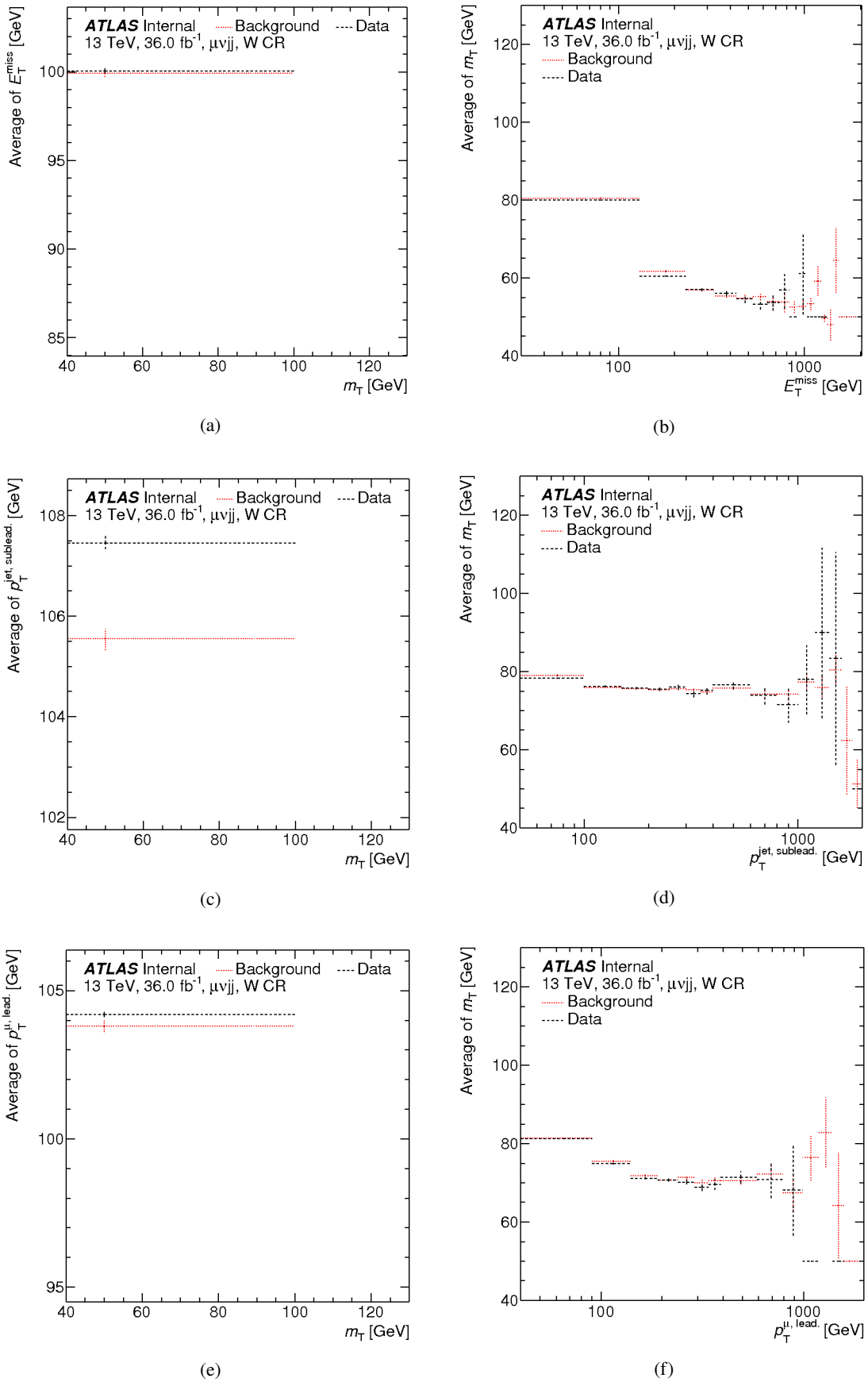
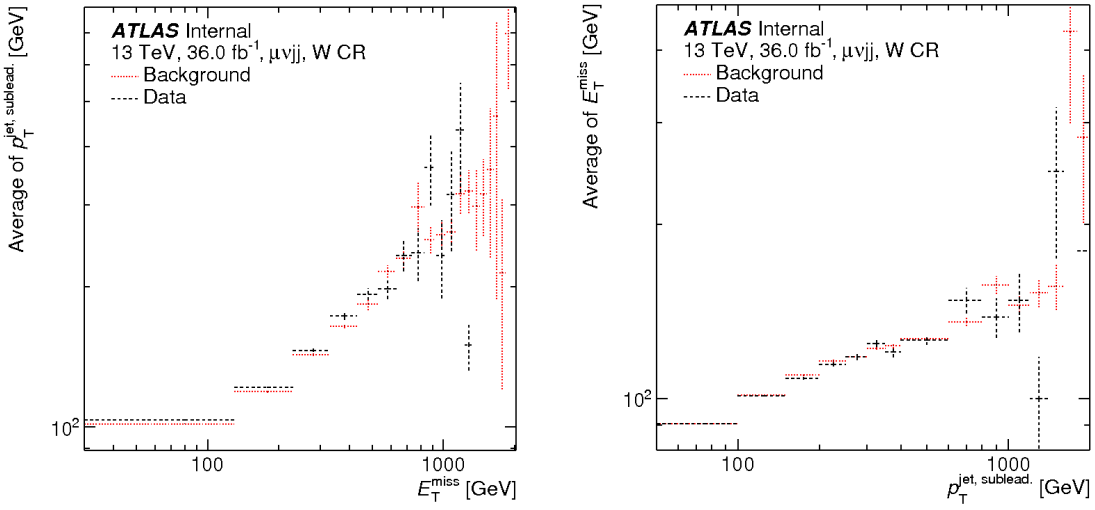
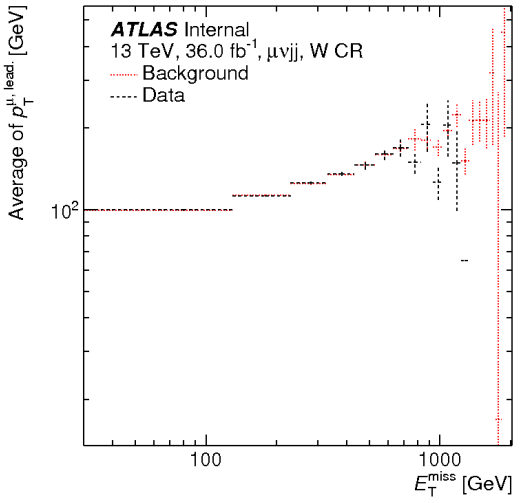


Figure B.34: Correlation of BDT input variables in the W CR.

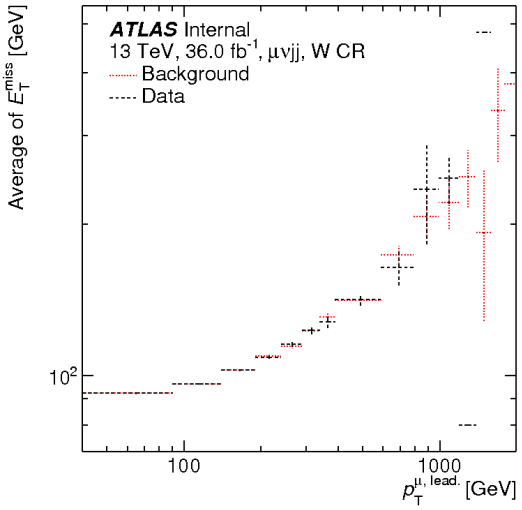


(a)

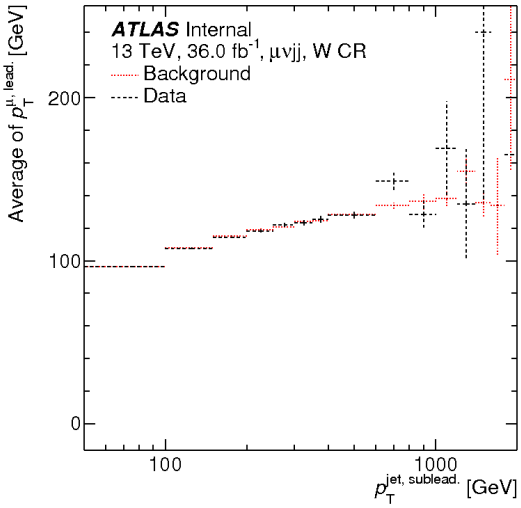
(b)



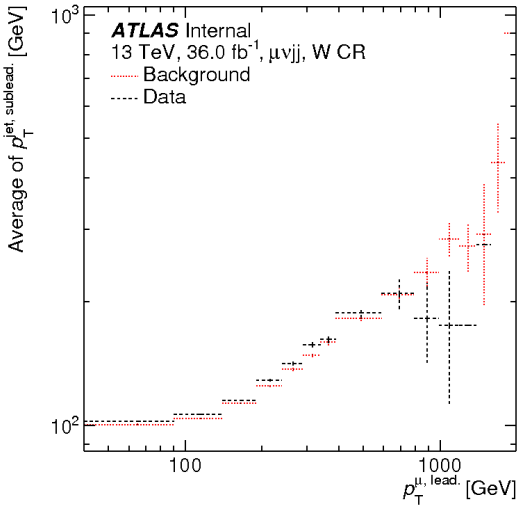
(c)



(d)



(e)



(f)

Figure B.35: Correlation of BDT input variables in the W CR.

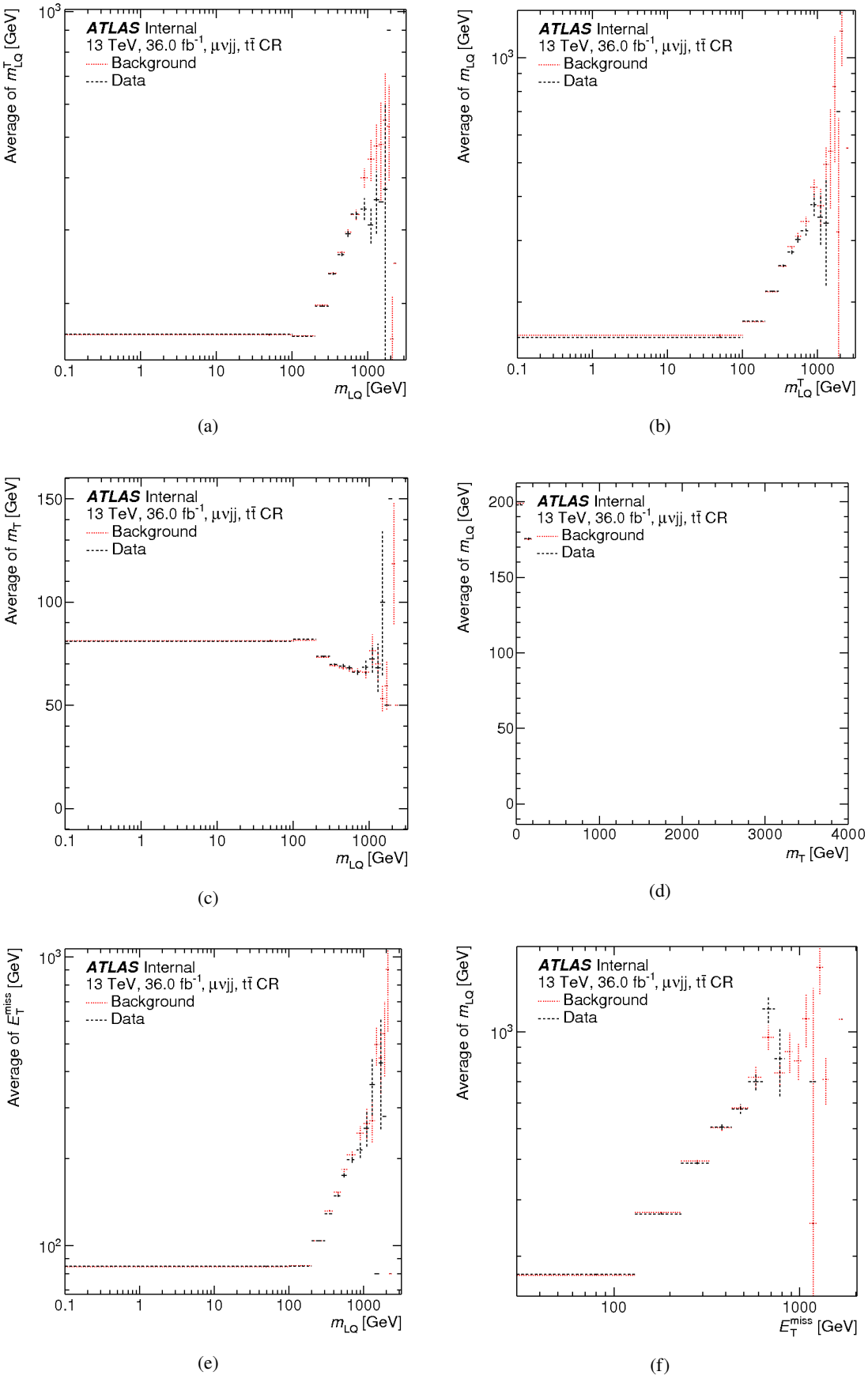
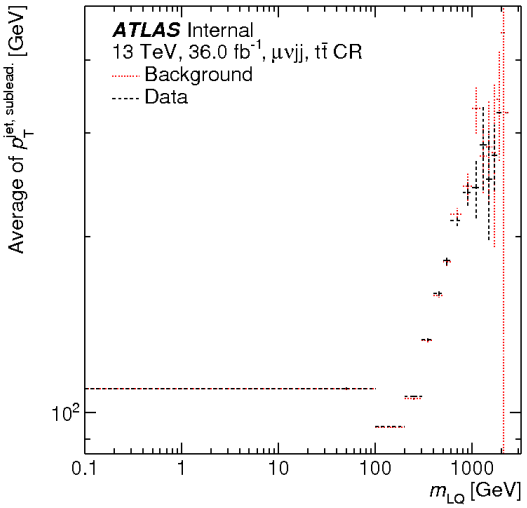
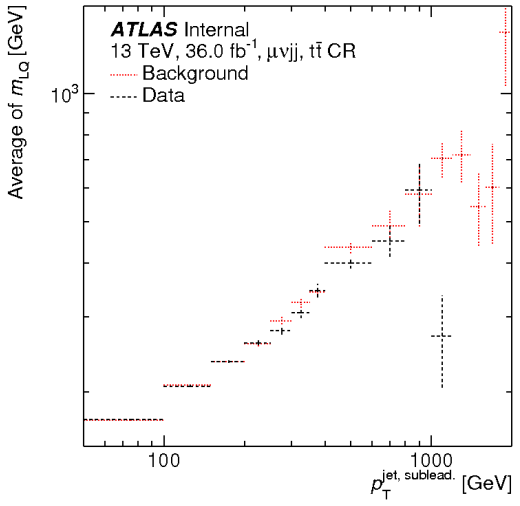


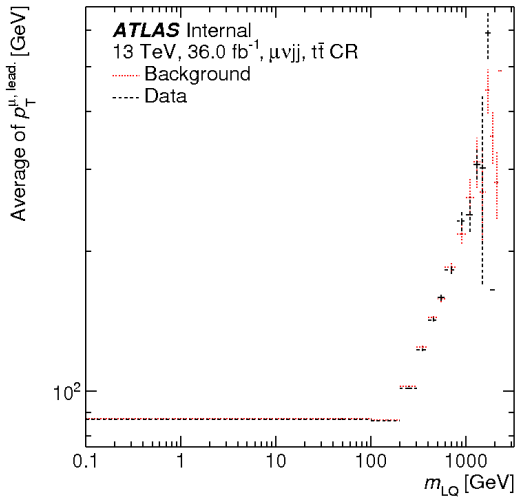
Figure B.36: Correlation of BDT input variables in the $t\bar{t}$ CR.



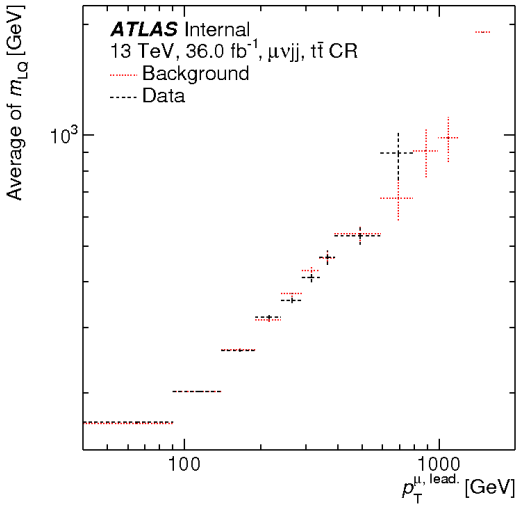
(a)



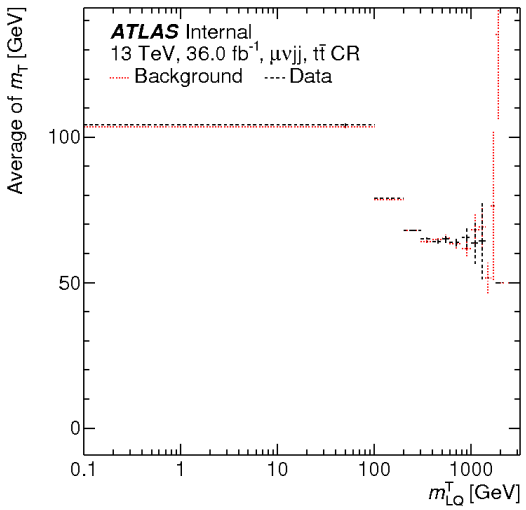
(b)



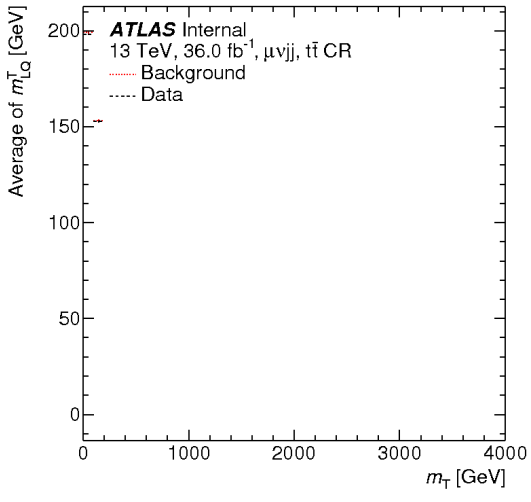
(c)



(d)



(e)



(f)

Figure B.37: Correlation of BDT input variables in the $t\bar{t}$ CR.

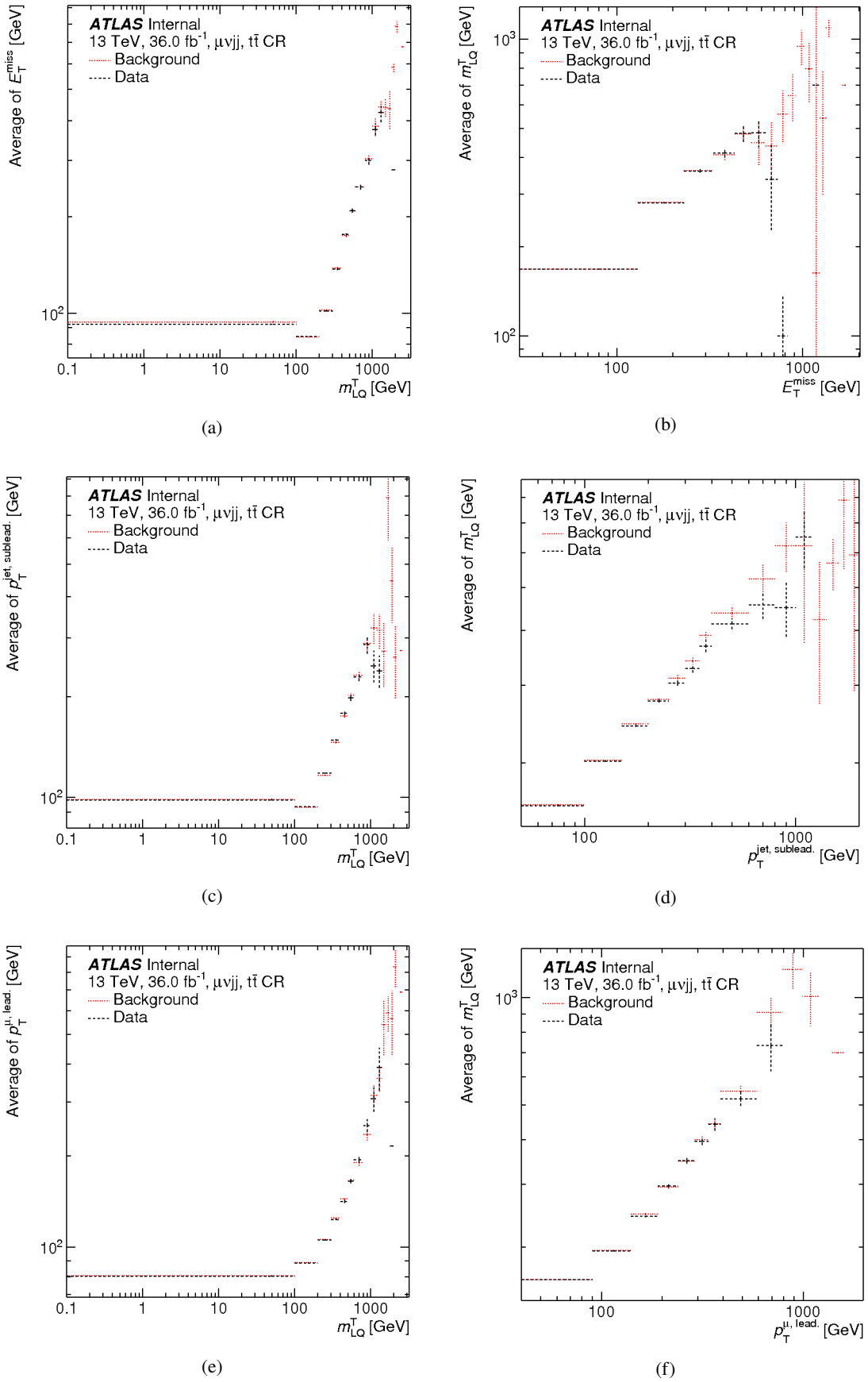
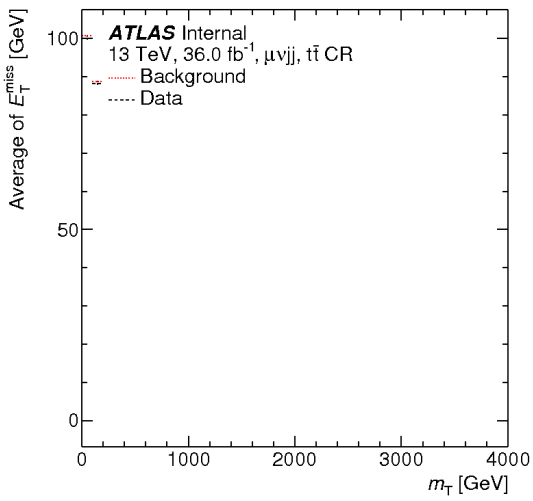
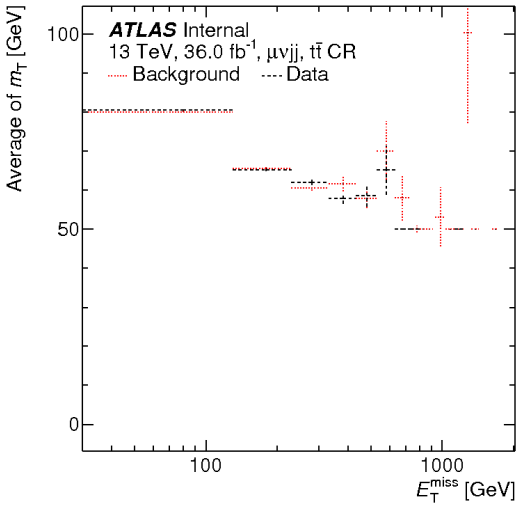


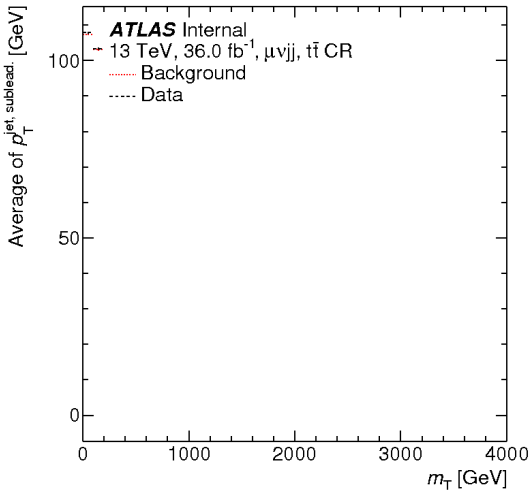
Figure B.38: Correlation of BDT input variables in the $t\bar{t}$ CR.



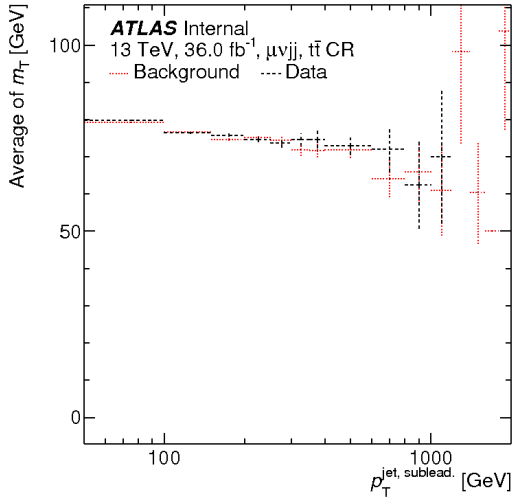
(a)



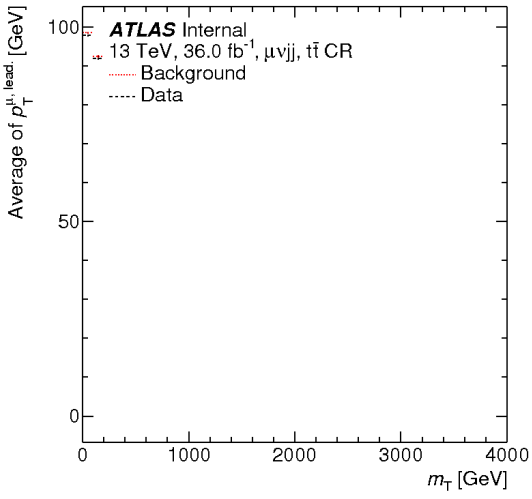
(b)



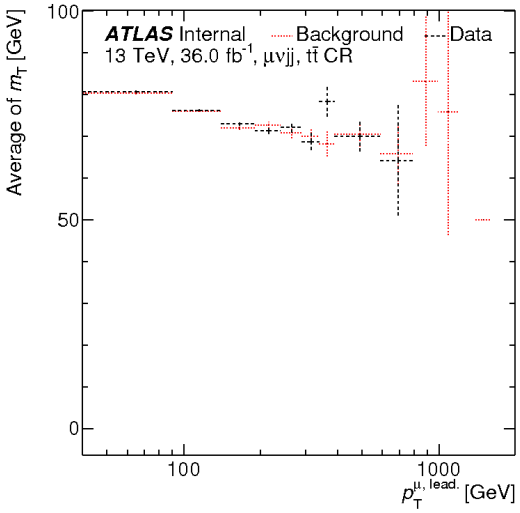
(c)



(d)

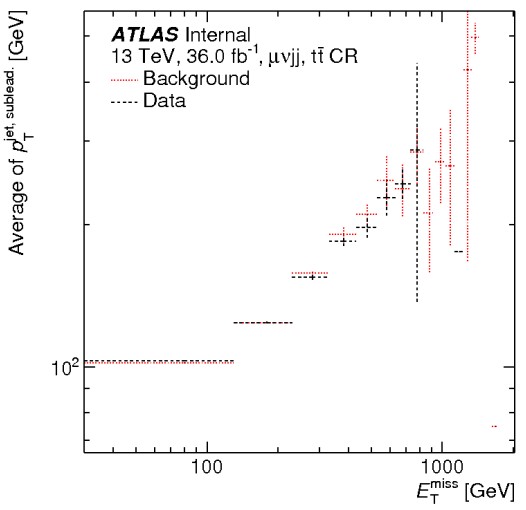


(e)

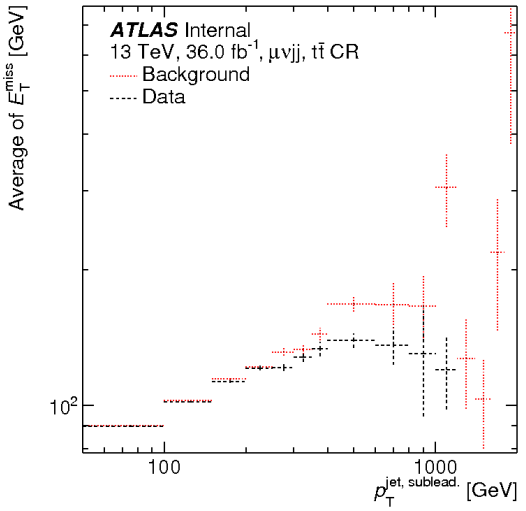


(f)

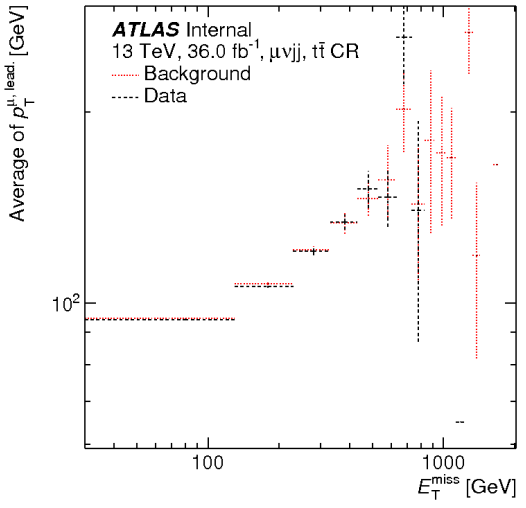
Figure B.39: Correlation of BDT input variables in the $t\bar{t}$ CR.



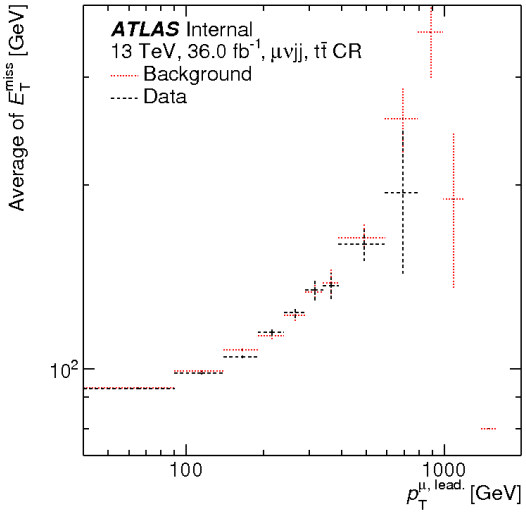
(a)



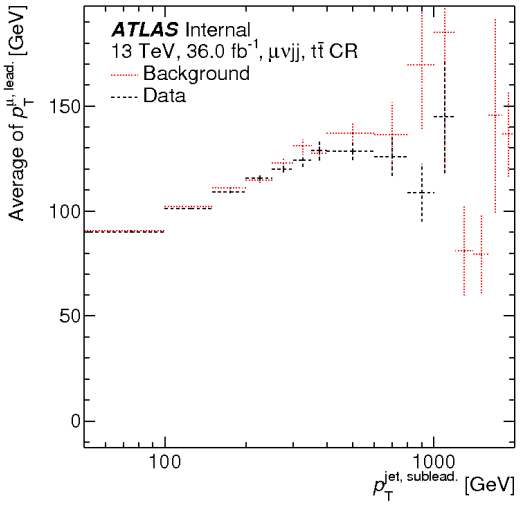
(b)



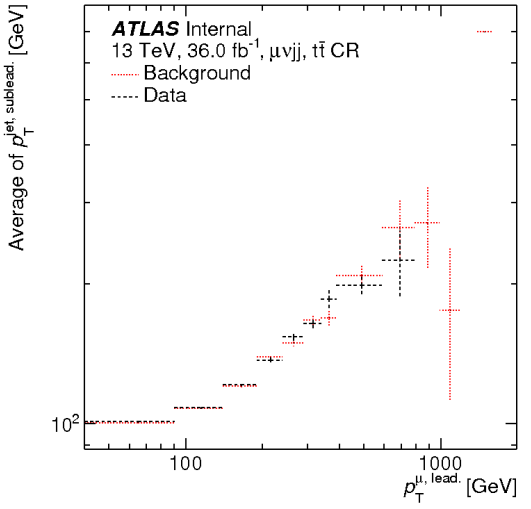
(c)



(d)



(e)



(f)

Figure B.40: Correlation of BDT input variables in the $t\bar{t}$ CR.

B.10 BDT validation distributions

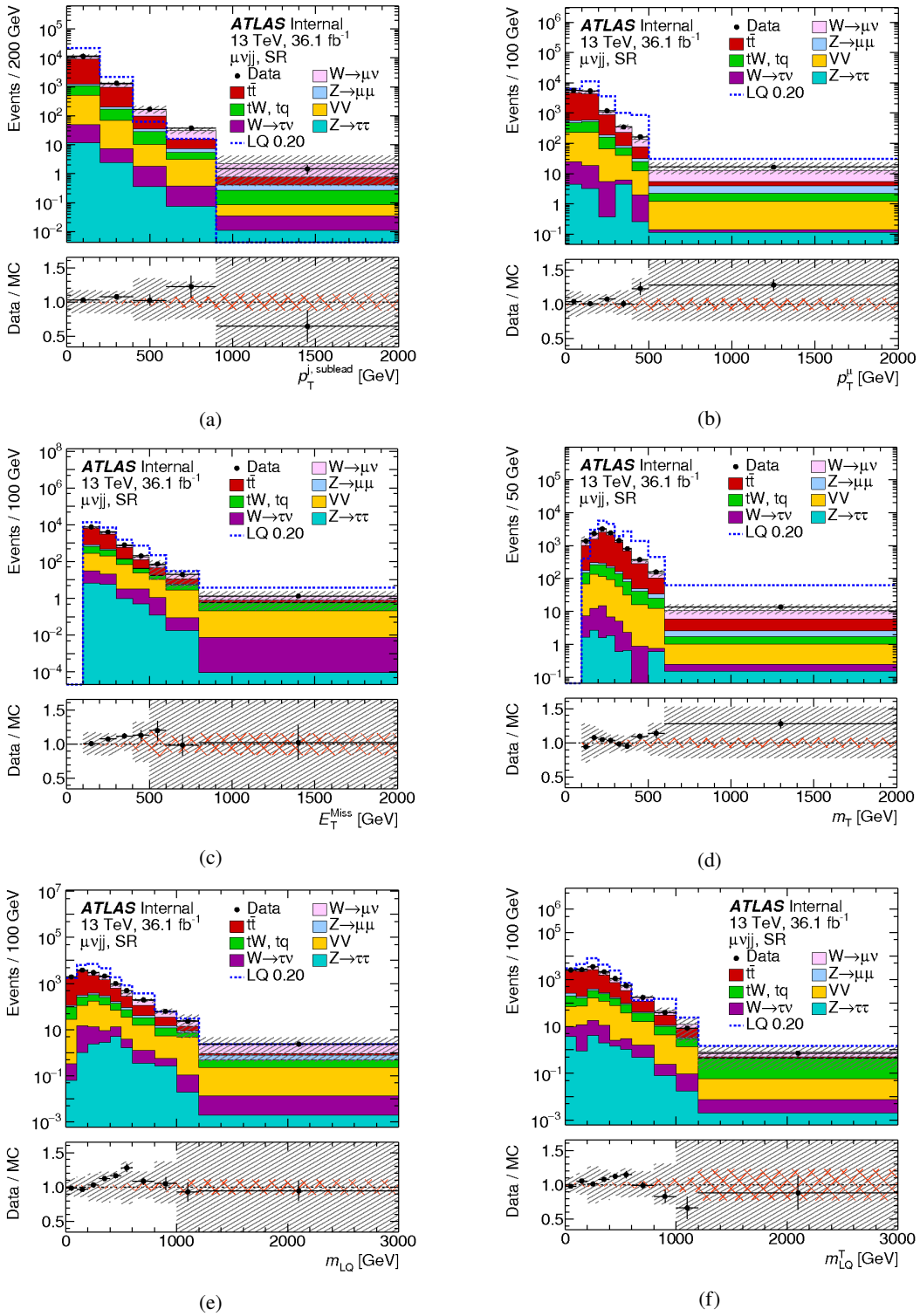


Figure B.41: Distribution of most relevant BDT input variables in the validation regions for an LQ mass hypothesis of 200 GeV.

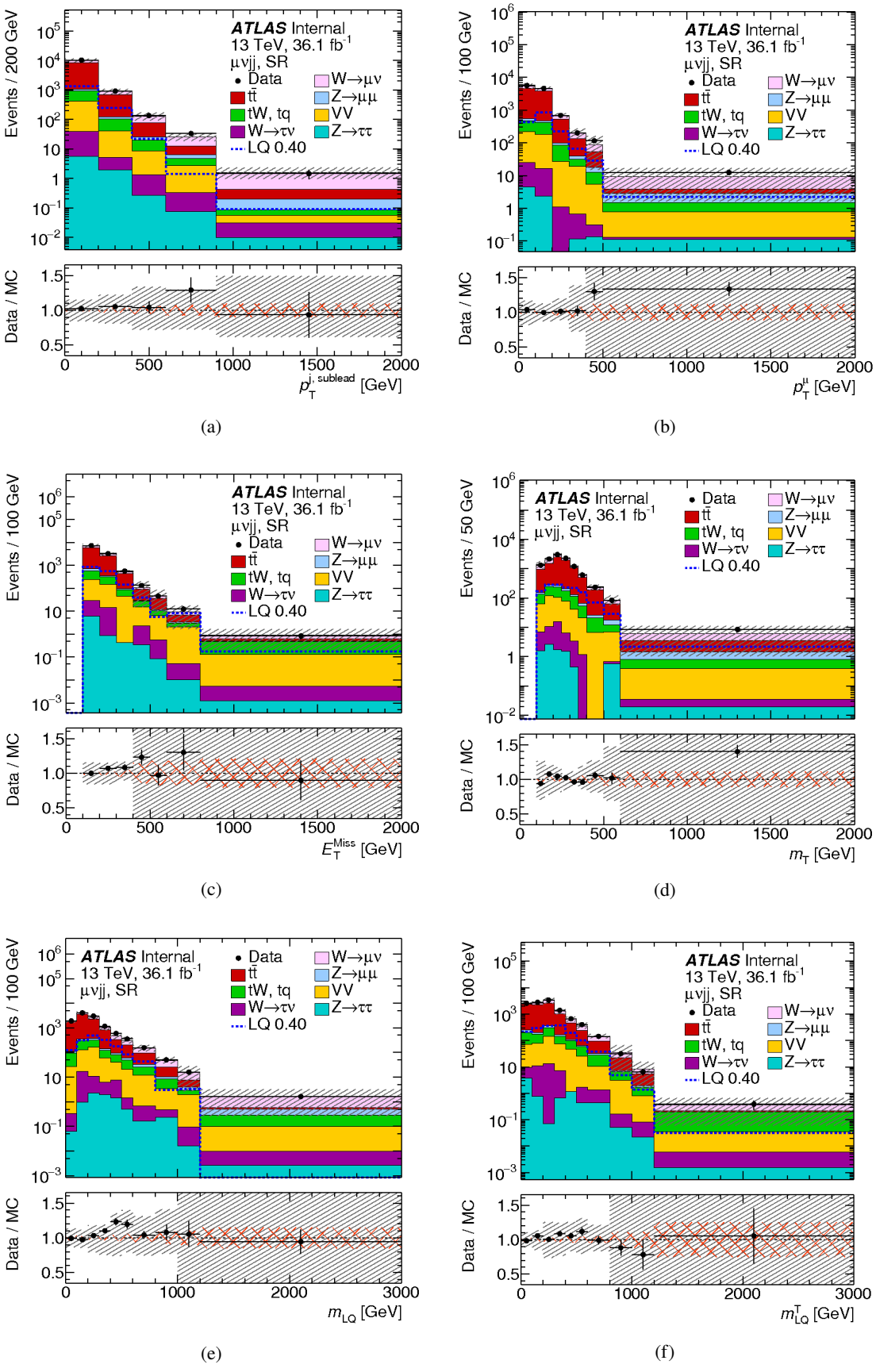
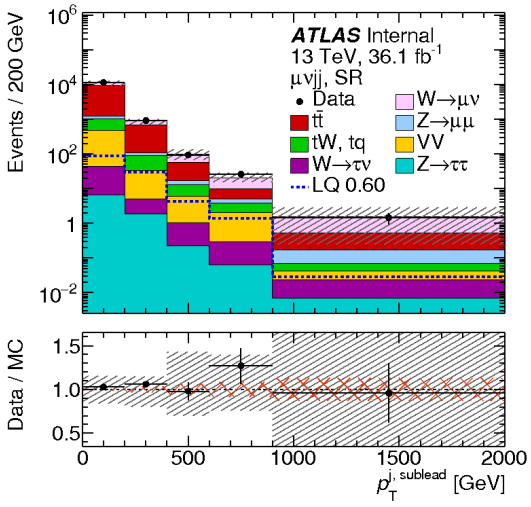
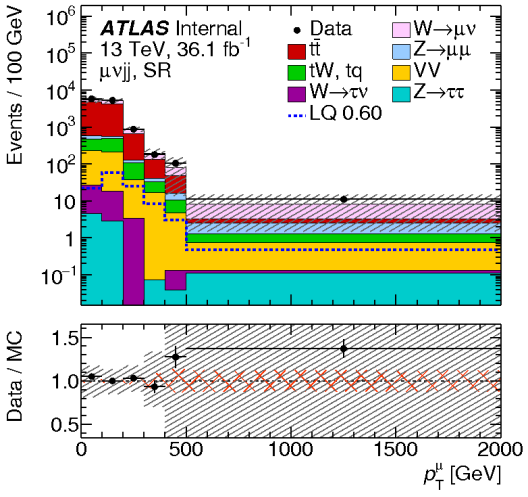


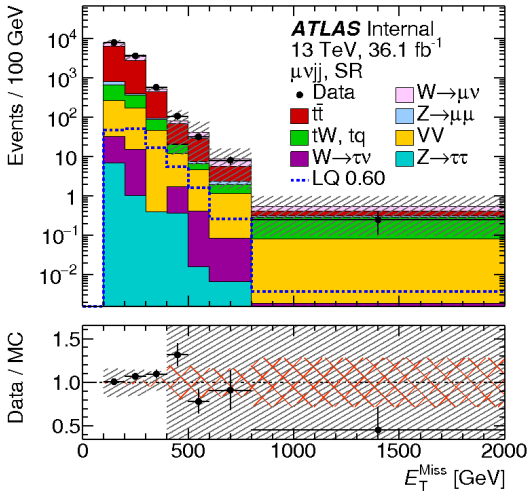
Figure B.42: Distribution of most relevant BDT input variables in the validation regions for an LQ mass hypothesis of 400 GeV.



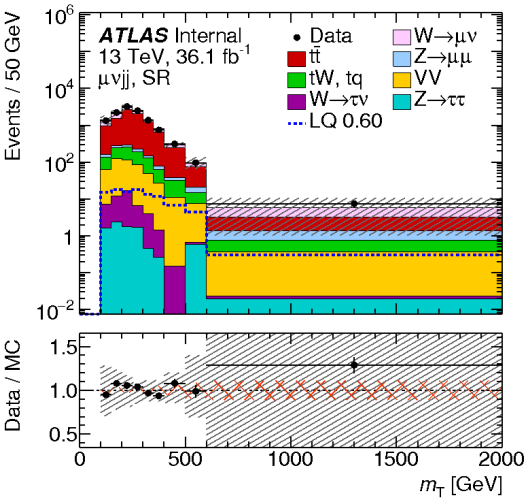
(a)



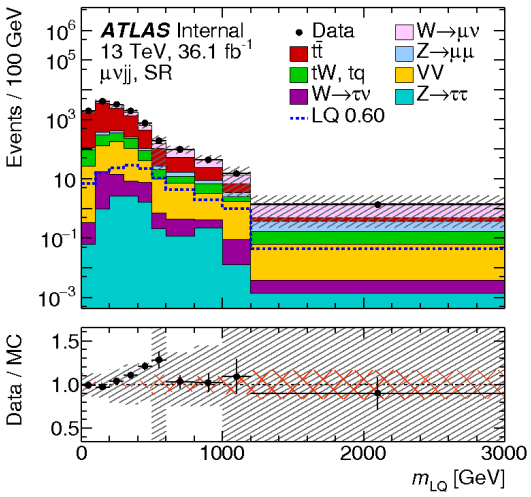
(b)



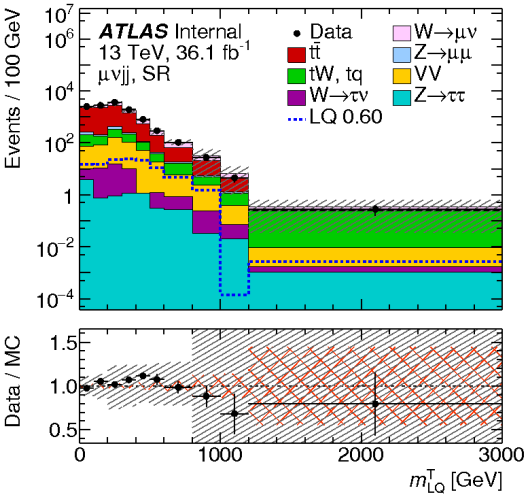
(c)



(d)



(e)



(f)

Figure B.43: Distribution of most relevant BDT input variables in the validation regions for an LQ mass hypothesis of 600 GeV.

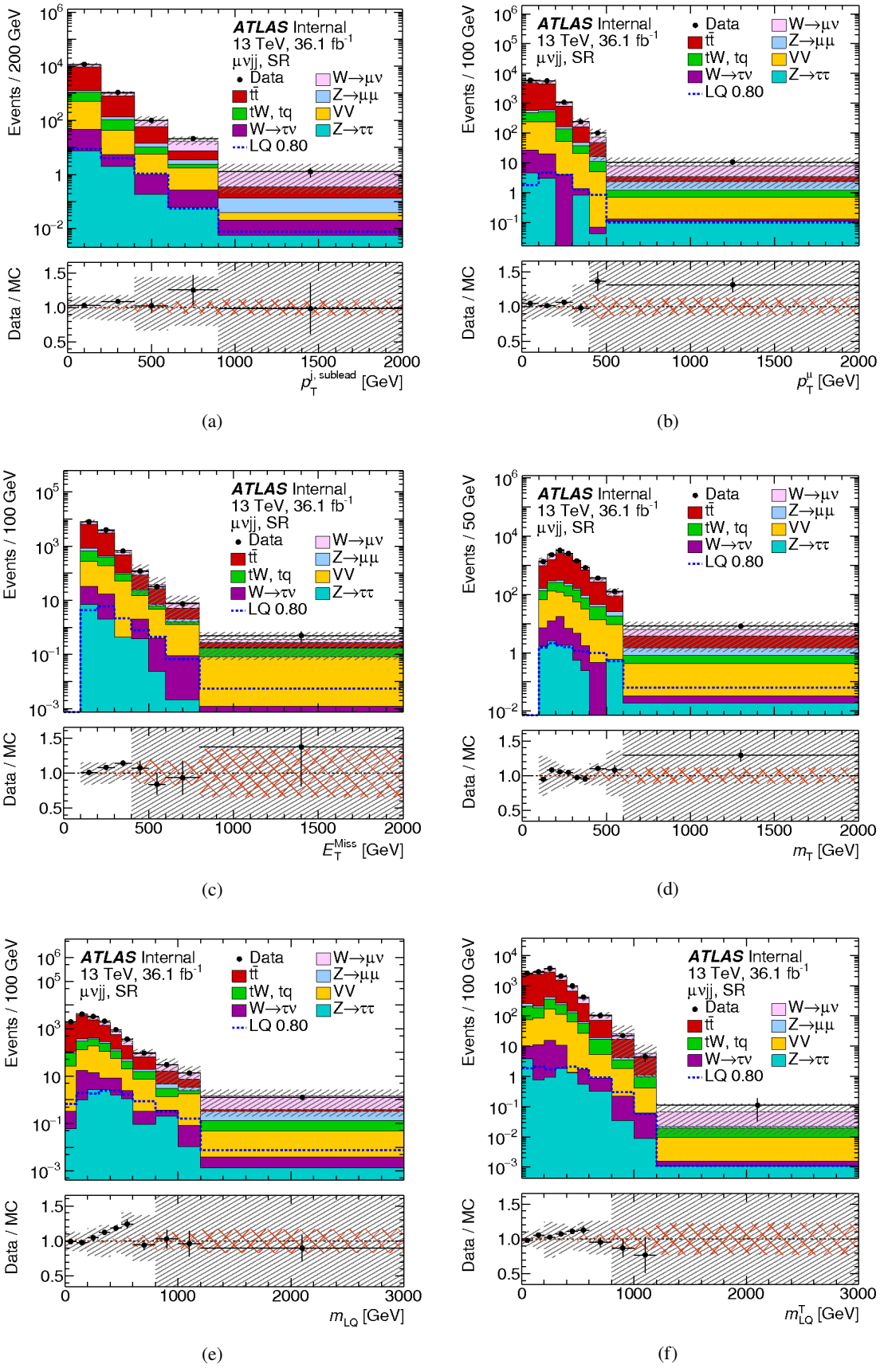
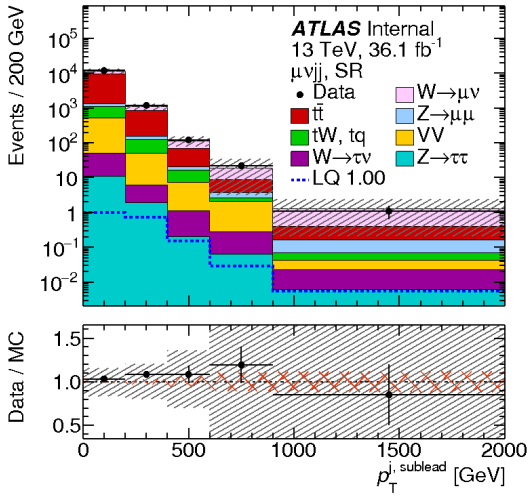
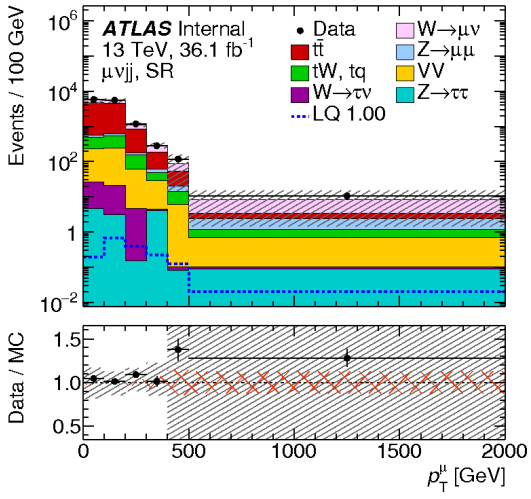


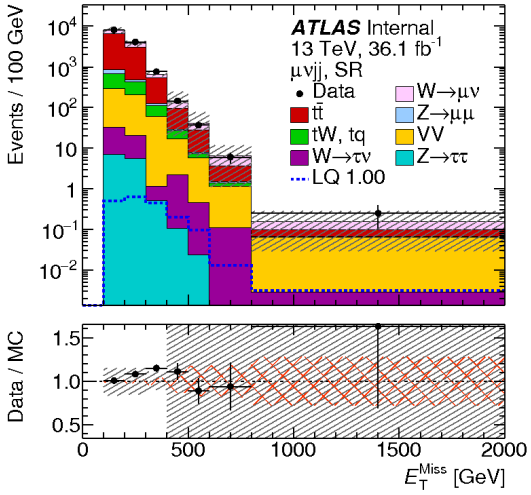
Figure B.44: Distribution of most relevant BDT input variables in the validation regions for an LQ mass hypothesis of 800 GeV.



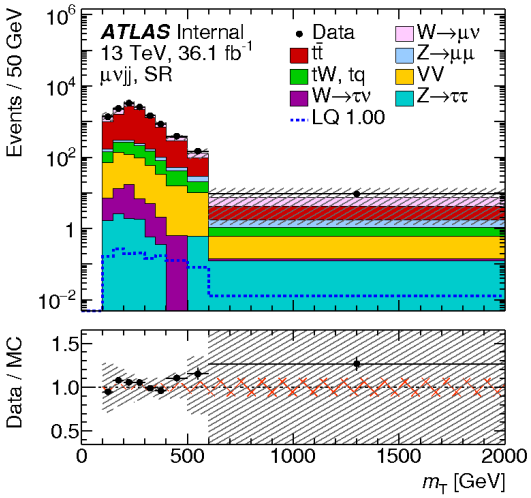
(a)



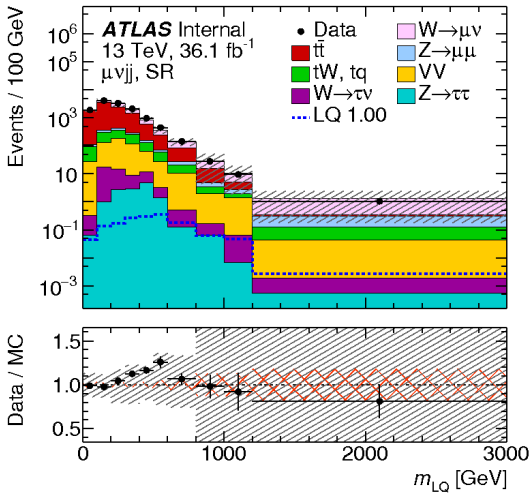
(b)



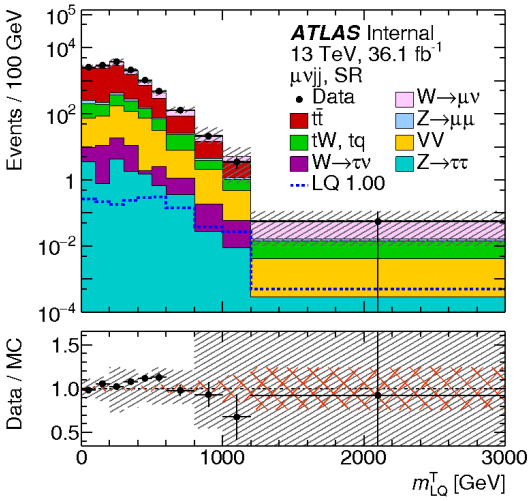
(c)



(d)



(e)



(f)

Figure B.45: Distribution of most relevant BDT input variables in the validation regions for an LQ mass hypothesis of 1000 GeV.

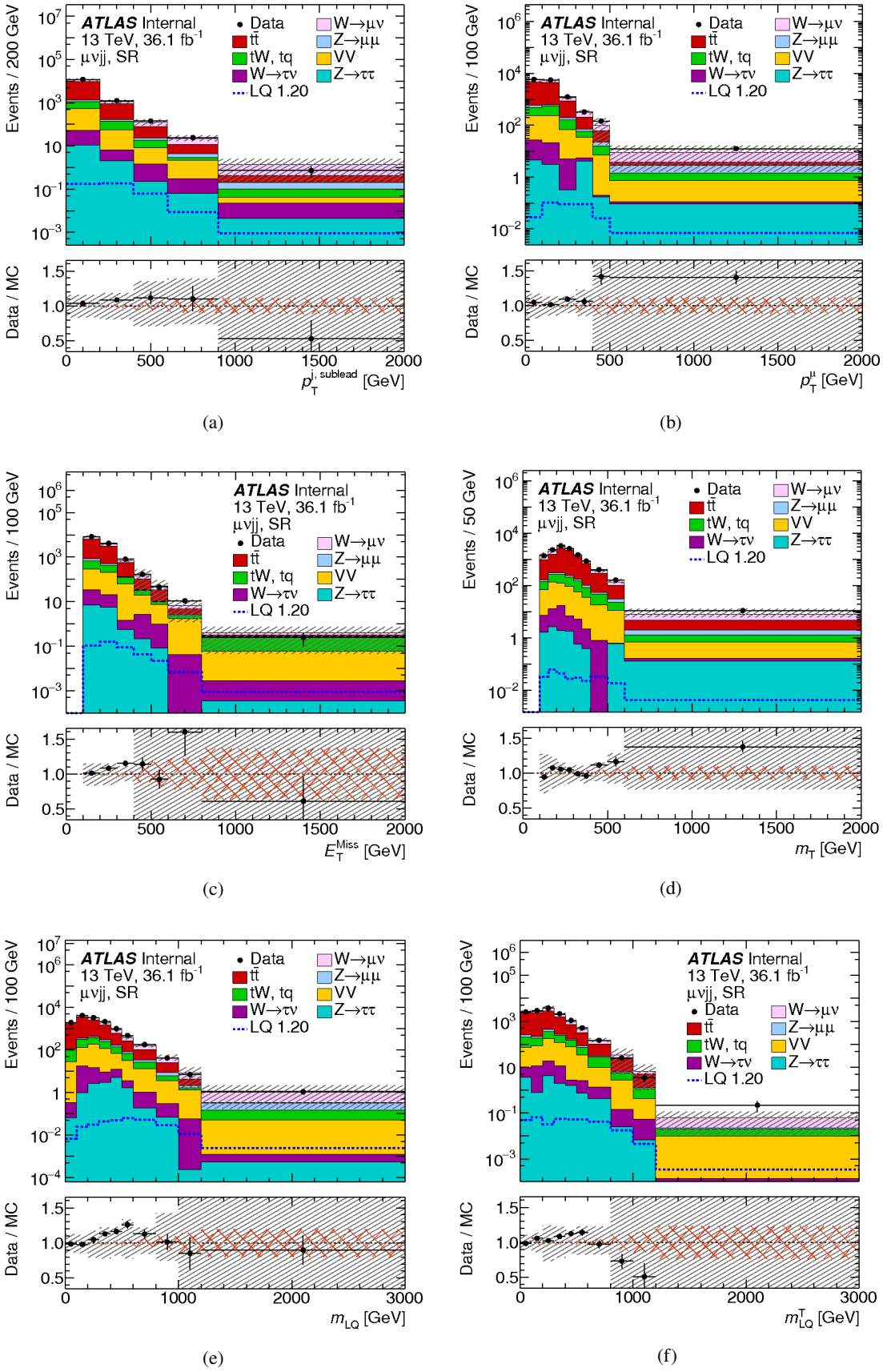
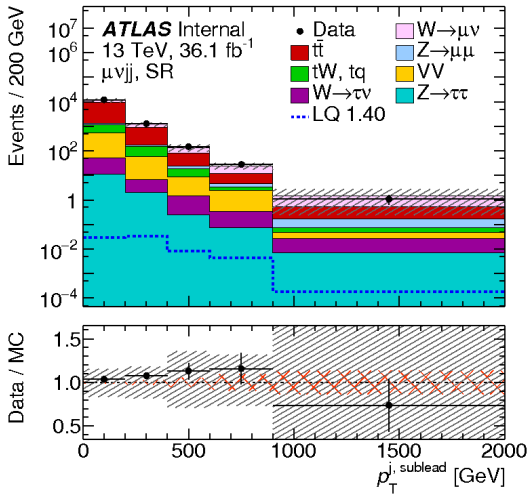
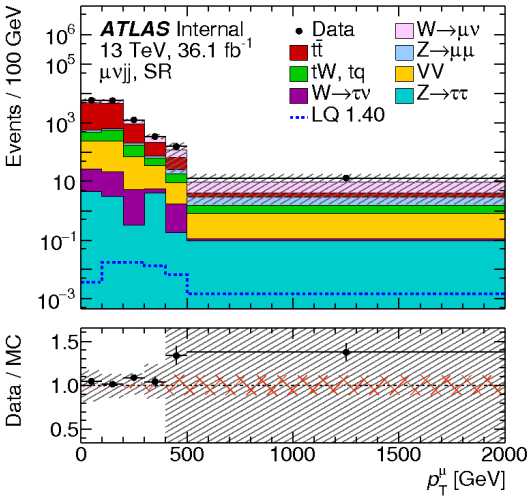


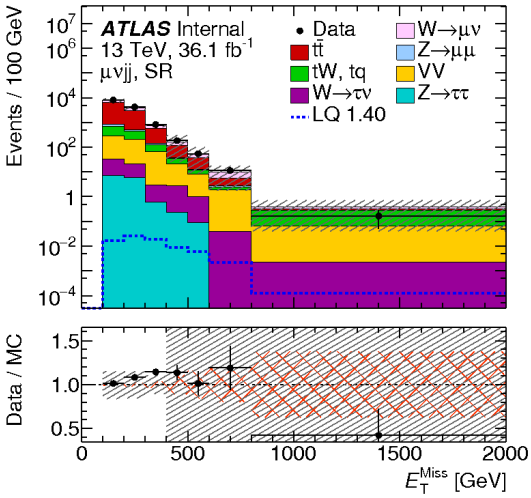
Figure B.46: Distribution of most relevant BDT input variables in the validation regions for an LQ mass hypothesis of 1200 GeV.



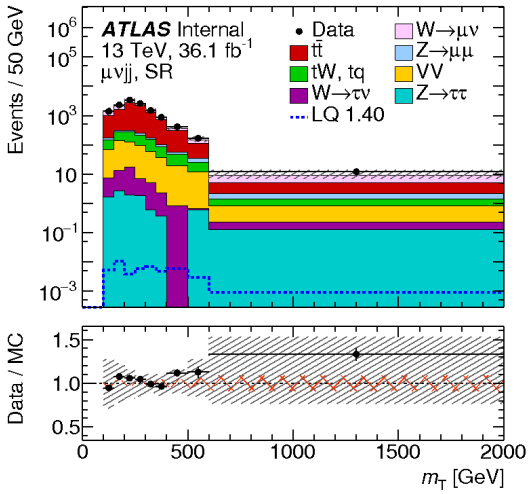
(a)



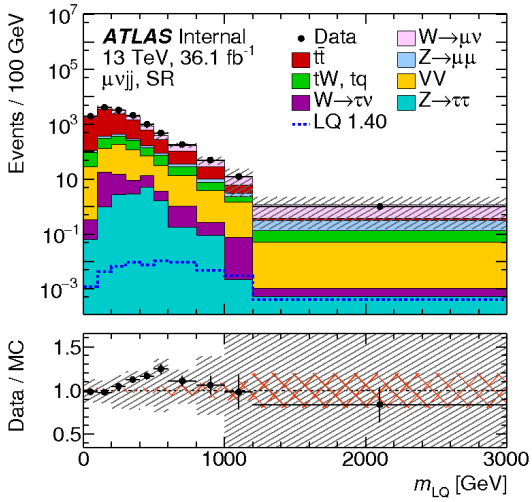
(b)



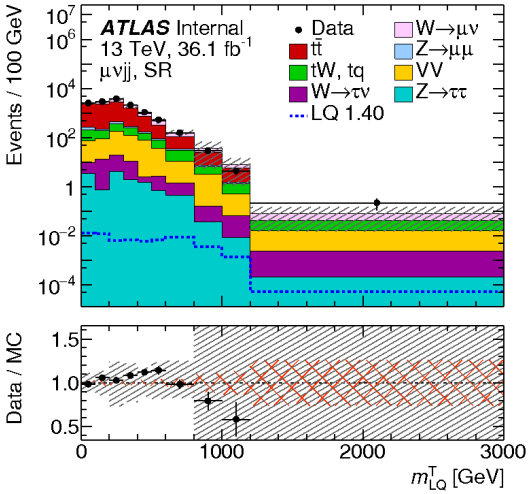
(c)



(d)



(e)



(f)

Figure B.47: Distribution of most relevant BDT input variables in the validation regions for an LQ mass hypothesis of 1400 GeV.

B.11 Final event yields

Event yields	SR 200 GeV	SR 250 GeV	SR 300 GeV
Observed events	590	37	180
Fitted Bkgr events	587.57 ± 31.8	31.98 ± 8.76	176.52 ± 26.01
Fitted W $\rightarrow \mu\nu$ events	129.89 ± 20.17	13.03 ± 2.96	40.64 ± 6.35
Fitted Z $\rightarrow \mu\mu$ events	11.01 ± 1.73	1.52 ± 0.35	5.32 ± 0.79
Fitted $t\bar{t}$ events	370.50 ± 36.09	10.51 ± 7.21	102.17 ± 26.34
Fitted VV events	31.83 ± 3.27	3.39 ± 0.64	12.71 ± 1.36
Fitted W t events	40.22 ± 4.44	2.40 ± 0.47	14.57 ± 1.65
Fitted W $\rightarrow \tau\nu$ events	4.12 ± 0.42	1.13 ± 0.21	1.11 ± 0.12
Fitted LQ events	$3.00^{+18.53}_{-3.00}$	$4.70^{+7.59}_{-4.70}$	$2.31^{+22.20}_{-2.31}$
MC exp. events	494.67 ± 107.70	31.29 ± 11.28	144.71 ± 40.12
MC exp. W $\rightarrow \mu\nu$ events	118.41 ± 23.28	12.76 ± 4.10	36.81 ± 7.06
MC exp. Z $\rightarrow \mu\mu$ events	10.22 ± 1.93	1.49 ± 0.45	4.94 ± 0.86
MC exp. $t\bar{t}$ events	295.14 ± 91.51	10.25 ± 7.72	76.68 ± 36.21
MC exp. VV events	29.69 ± 3.99	3.33 ± 0.89	11.78 ± 1.58
MC exp. W t events	37.35 ± 5.39	2.35 ± 0.65	13.46 ± 1.93
MC exp. W $\rightarrow \tau\nu$ events	3.85 ± 0.51	1.11 ± 0.29	1.03 ± 0.13
MC exp. LQ events	3.85 ± 0.51	3939.58 ± 451.35	5812.36 ± 290.32

Table B.1: Event yields in signal regions before and after the fit for an LQ mass of 200, 250 and 300 GeV.

Event yields	SR 350 GeV	SR 400 GeV	SR 450 GeV
Observed events	90	82	85
Fitted Bkgr events	93.34 ± 12.18	82.63 ± 18.39	79.96 ± 12.21
Fitted W $\rightarrow \mu\nu$ events	25.75 ± 4.65	24.18 ± 5.59	23.90 ± 4.61
Fitted Z $\rightarrow \mu\mu$ events	2.45 ± 0.44	2.77 ± 0.61	1.68 ± 0.28
Fitted $t\bar{t}$ events	48.64 ± 11.70	39.67 ± 15.91	33.60 ± 10.39
Fitted VV events	7.98 ± 1.29	9.49 ± 1.87	9.08 ± 1.21
Fitted W t events	5.86 ± 0.96	6.37 ± 1.28	9.98 ± 1.37
Fitted W $\rightarrow \tau\nu$ events	2.67 ± 0.48	0.15 ± 0.03	1.73 ± 0.23
Fitted LQ events	$7.90^{+7.73}_{-7.90}$	$2.17^{+16.27}_{-2.17}$	$6.65^{+9.70}_{-6.65}$
MC exp. events	53.42 ± 27.37	68.70 ± 31.31	85.22 ± 18.84
MC exp. W $\rightarrow \mu\nu$ events	15.70 ± 7.57	20.59 ± 8.57	24.85 ± 5.37
MC exp. Z $\rightarrow \mu\mu$ events	1.52 ± 0.71	2.38 ± 0.93	1.73 ± 0.31
MC exp. $t\bar{t}$ events	25.09 ± 18.66	31.96 ± 21.25	37.26 ± 14.39
MC exp. VV events	4.95 ± 2.21	8.17 ± 3.03	9.34 ± 1.41
MC exp. W t events	3.63 ± 1.63	5.46 ± 2.07	10.27 ± 1.59
MC exp. W $\rightarrow \tau\nu$ events	2.53 ± 0.30	0.13 ± 0.05	1.78 ± 0.27
MC exp. LQ events	2368.97 ± 278.34	2148.64 ± 205.67	1579.00 ± 89.95

Table B.2: Event yields in signal regions before and after the fit for an LQ mass of 350, 400 and 450 GeV.

Event yields	SR 500 GeV	SR 550 GeV	SR 600 GeV
Observed events	140	47	68
Fitted Bkgr events	128.69 ± 16.24	43.68 ± 9.05	60.61 ± 11.80
Fitted W $\rightarrow \mu\nu$ events	29.77 ± 5.03	13.14 ± 2.36	19.46 ± 3.46
Fitted Z $\rightarrow \mu\mu$ events	4.86 ± 0.75	0.75 ± 0.16	1.61 ± 0.43
Fitted $t\bar{t}$ events	58.29 ± 8.66	18.72 ± 8.02	19.55 ± 10.72
Fitted VV events	18.81 ± 2.37	5.01 ± 0.73	9.07 ± 1.14
Fitted W t events	15.09 ± 1.92	5.19 ± 0.79	9.94 ± 1.27
Fitted W $\rightarrow \tau\nu$ events	1.87 ± 0.24	0.86 ± 0.12	0.97 ± 0.12
Fitted LQ (200 GeV) events	$4.15^{+8.03}_{-4.15}$	$4.70^{+7.59}_{-4.70}$	$4.96^{+9.89}_{-4.96}$
MC exp. events	100.41 ± 21.01	33.05 ± 10.37	56.43 ± 14.60
MC exp. W $\rightarrow \mu\nu$ events	22.34 ± 5.97	10.98 ± 2.47	18.53 ± 3.93
MC exp. Z $\rightarrow \mu\mu$ events	3.85 ± 0.87	0.66 ± 0.16	1.56 ± 0.45
MC exp. $t\bar{t}$ events	45.97 ± 10.71	11.72 ± 8.70	17.01 ± 11.91
MC exp. VV events	14.82 ± 2.97	4.39 ± 0.76	8.78 ± 1.31
MC exp. W t events	11.95 ± 2.38	4.54 ± 0.83	9.61 ± 1.46
MC exp. W $\rightarrow \tau\nu$ events	1.48 ± 0.30	0.76 ± 0.13	0.94 ± 0.14
MC exp. LQ events	1317.91 ± 95.19	567.53 ± 36.91	487.85 ± 27.77

Table B.3: Event yields in signal regions before and after the fit for an LQ mass of 500, 550 and 600 GeV.

Event yields	SR 650 GeV	SR 700 GeV	SR 750 GeV
Observed events	86	47	55
Fitted Bkgr events	74.54 ± 11.83	42.97 ± 8.23	52.89 ± 10.69
Fitted W $\rightarrow \mu\nu$ events	17.83 ± 4.21	15.24 ± 3.62	16.94 ± 4.80
Fitted Z $\rightarrow \mu\mu$ events	2.30 ± 0.39	1.46 ± 0.33	1.66 ± 0.39
Fitted $t\bar{t}$ events	31.76 ± 11.04	17.47 ± 7.24	20.03 ± 9.69
Fitted VV events	10.56 ± 1.37	5.29 ± 0.91	7.17 ± 1.53
Fitted W t events	11.34 ± 1.48	3.03 ± 0.54	6.73 ± 1.44
Fitted W $\rightarrow \tau\nu$ events	0.75 ± 0.10	0.48 ± 0.08	0.35 ± 0.07
Fitted LQ events	$7.04^{+9.35}_{-7.04}$	$0.28^{+6.22}_{-0.28}$	$1.74^{+7.71}_{-1.74}$
MC exp. events	62.40 ± 16.68	34.04 ± 12.08	50.85 ± 19.83
MC exp. W $\rightarrow \mu\nu$ events	15.65 ± 4.60	12.91 ± 4.31	16.41 ± 6.44
MC exp. Z $\rightarrow \mu\mu$ events	2.13 ± 0.41	1.31 ± 0.37	1.63 ± 0.51
MC exp. $t\bar{t}$ events	23.69 ± 13.43	11.93 ± 9.04	18.90 ± 13.44
MC exp. VV events	9.75 ± 1.56	4.75 ± 1.10	7.00 ± 2.08
MC exp. W t events	10.48 ± 1.67	2.71 ± 0.65	6.57 ± 1.95
MC exp. W $\rightarrow \tau\nu$ events	0.69 ± 0.11	0.43 ± 0.09	0.34 ± 0.10
MC exp. LQ events	388.10 ± 36.63	222.45 ± 10.62	186.75 ± 19.66

Table B.4: Event yields in signal regions before and after the fit for an LQ mass of 650, 700 and 750 GeV.

Event yields	SR 800 GeV	SR 850 GeV	SR 900 GeV
Observed events	7	20	13
Fitted Bkgr events	9.57 ± 4.27	21.03 ± 4.99	14.45 ± 4.02
Fitted W $\rightarrow \mu\nu$ events	4.09 ± 1.44	7.62 ± 2.71	4.28 ± 1.13
Fitted Z $\rightarrow \mu\mu$ events	0.28 ± 0.10	0.56 ± 0.09	0.43 ± 0.11
Fitted $t\bar{t}$ events	3.08 ± 3.05	5.97 ± 3.90	5.81 ± 3.38
Fitted VV events	0.80 ± 0.26	1.59 ± 0.20	2.32 ± 0.52
Fitted W t events	1.32 ± 0.44	5.26 ± 0.67	1.59 ± 0.35
Fitted W $\rightarrow \tau\nu$ events	0.00 ± 0.00	0.03 ± 0.00	0.02 ± 0.00
Fitted LQ events	$1.93^{+3.97}_{-1.93}$	$1.02^{+4.87}_{-1.02}$	$1.98^{+3.51}_{-1.98}$
MC exp. events	10.78 ± 5.09	21.04 ± 5.65	13.44 ± 5.36
MC exp. W $\rightarrow \mu\nu$ events	4.63 ± 1.91	7.62 ± 3.12	4.08 ± 1.32
MC exp. Z $\rightarrow \mu\mu$ events	0.31 ± 0.13	0.56 ± 0.09	0.41 ± 0.12
MC exp. $t\bar{t}$ events	3.45 ± 3.11	5.98 ± 3.99	5.19 ± 4.06
MC exp. VV events	0.89 ± 0.35	1.59 ± 0.22	2.22 ± 0.62
MC exp. W t events	1.48 ± 0.58	5.26 ± 0.71	1.52 ± 0.42
MC exp. W $\rightarrow \tau\nu$ events	0.01 ± 0.00	0.03 ± 0.00	0.02 ± 0.01
MC exp. LQ events	76.76 ± 5.17	72.08 ± 4.56	46.12 ± 2.44

Table B.5: Event yields in signal regions before and after the fit for an LQ mass of 800, 850 and 900 GeV.

Event yields	SR 950 GeV	SR 1000 GeV	SR 1050 GeV
Observed events	8	18	16
Fitted Bkgr events	10.56 ± 4.17	18.25 ± 3.81	15.81 ± 4.98
Fitted W $\rightarrow \mu\nu$ events	3.96 ± 1.43	6.49 ± 1.32	5.83 ± 1.89
Fitted Z $\rightarrow \mu\mu$ events	0.36 ± 0.14	0.42 ± 0.09	0.44 ± 0.16
Fitted $t\bar{t}$ events	3.90 ± 3.16	4.99 ± 3.08	3.58 ± 2.47
Fitted VV events	1.67 ± 0.56	2.57 ± 0.34	1.76 ± 0.52
Fitted W t events	0.63 ± 0.21	3.74 ± 0.49	4.15 ± 1.24
Fitted W $\rightarrow \tau\nu$ events	0.04 ± 0.01	0.04 ± 0.01	0.05 ± 0.02
Fitted LQ events	$2.36^{+3.60}_{-2.36}$	$1.81^{+4.11}_{-1.81}$	$0.17^{+4.38}_{-0.17}$
MC exp. events	11.25 ± 6.08	17.41 ± 4.07	14.91 ± 6.88
MC exp. W $\rightarrow \mu\nu$ events	4.19 ± 2.02	6.24 ± 1.38	5.50 ± 2.56
MC exp. Z $\rightarrow \mu\mu$ events	0.38 ± 0.18	0.41 ± 0.09	0.41 ± 0.20
MC exp. $t\bar{t}$ events	4.22 ± 3.79	4.55 ± 3.20	3.34 ± 2.78
MC exp. VV events	1.76 ± 0.80	2.51 ± 0.35	1.66 ± 0.71
MC exp. W t events	0.66 ± 0.30	3.66 ± 0.50	3.94 ± 1.69
MC exp. W $\rightarrow \tau\nu$ events	0.04 ± 0.02	0.04 ± 0.01	0.05 ± 0.02
MC exp. LQ events	35.42 ± 1.68	$28.381.98\pm$	20.56 ± 1.58

Table B.6: Event yields in signal regions before and after the fit for an LQ mass of 950, 1000 and 1050 GeV.

Event yields	SR 1100 GeV	SR 1150 GeV	SR 1200 GeV
Observed events	17	18	14
Fitted Bkgr events	19.12 ± 3.74	18.78 ± 3.82	17.35 ± 4.21
Fitted W $\rightarrow \mu\nu$ events	7.88 ± 1.79	7.39 ± 1.70	6.99 ± 1.79
Fitted Z $\rightarrow \mu\mu$ events	0.41 ± 0.10	0.51 ± 0.14	0.49 ± 0.15
Fitted $t\bar{t}$ events	3.24 ± 2.31	2.80 ± 1.68	3.09 ± 2.27
Fitted VV events	4.18 ± 0.65	4.27 ± 0.77	2.98 ± 0.60
Fitted W t events	3.34 ± 0.52	3.73 ± 0.68	3.72 ± 0.77
Fitted W $\rightarrow \tau\nu$ events	0.07 ± 0.01	0.08 ± 0.02	0.08 ± 0.02
Fitted LQ events	$0.99^{+3.73}_{-0.99}$	$2.41^{+4.01}_{-2.41}$	$3.41^{+4.05}_{-3.41}$
MC exp. events	19.98 ± 4.35	17.32 ± 4.24	17.28 ± 4.85
MC exp. W $\rightarrow \mu\nu$ events	8.25 ± 2.04	6.72 ± 1.91	6.96 ± 2.12
MC exp. Z $\rightarrow \mu\mu$ events	0.41 ± 0.11	0.47 ± 0.14	0.49 ± 0.16
MC exp. $t\bar{t}$ events	3.50 ± 2.42	2.63 ± 1.66	3.08 ± 2.31
MC exp. VV events	4.31 ± 0.73	3.96 ± 0.86	2.97 ± 0.71
MC exp. W t events	3.44 ± 0.59	3.46 ± 0.76	3.71 ± 0.91
MC exp. W $\rightarrow \tau\nu$ events	0.07 ± 0.01	0.08 ± 0.02	0.08 ± 0.02
MC exp. LQ events	16.31 ± 1.42	11.99 ± 1.25	9.12 ± 1.09

Table B.7: Event yields in signal regions before and after the fit for an LQ mass of 1100, 1150 and 1200 GeV.

Event yields	SR 1250 GeV	SR 1300 GeV	SR 1350 GeV
Observed events	19	14	13
Fitted Bkgr events	19.26 ± 4.52	14.19 ± 3.57	16.16 ± 4.51
Fitted W $\rightarrow \mu\nu$ events	7.57 ± 1.93	6.88 ± 1.87	6.59 ± 1.85
Fitted Z $\rightarrow \mu\mu$ events	0.48 ± 0.15	0.37 ± 0.13	0.46 ± 0.16
Fitted $t\bar{t}$ events	2.05 ± 1.47	1.65 ± 1.18	2.32 ± 1.42
Fitted VV events	3.87 ± 0.83	1.35 ± 0.30	2.56 ± 0.66
Fitted W t events	5.18 ± 1.11	3.85 ± 0.86	4.12 ± 1.06
Fitted W $\rightarrow \tau\nu$ events	0.12 ± 0.03	0.09 ± 0.02	0.10 ± 0.03
Fitted LQ events	$2.26^{+4.63}_{-2.26}$	$1.37^{+3.97}_{-1.37}$	$4.45^{+4.27}_{-4.45}$
MC exp. events	17.00 ± 4.94	13.07 ± 3.86	13.27 ± 5.48
MC exp. W $\rightarrow \mu\nu$ events	6.50 ± 2.16	6.21 ± 2.09	5.27 ± 2.34
MC exp. Z $\rightarrow \mu\mu$ events	0.43 ± 0.16	0.34 ± 0.14	0.38 ± 0.18
MC exp. $t\bar{t}$ events	1.98 ± 1.44	1.64 ± 1.15	2.07 ± 1.41
MC exp. VV events	3.41 ± 0.93	1.25 ± 0.34	2.09 ± 0.83
MC exp. W t events	4.57 ± 1.24	3.55 ± 0.95	3.37 ± 1.34
MC exp. W $\rightarrow \tau\nu$ events	0.11 ± 0.03	0.08 ± 0.02	0.09 ± 0.03
MC exp. LQ events	6.82 ± 0.89	4.75 ± 0.68	3.59 ± 0.57

Table B.8: Event yields in signal regions before and after the fit for an LQ mass of 1250, 1300 and 1350 GeV.

Event yields	SR 1400 GeV	SR 1450 GeV	SR 1500 GeV
Observed events	11	11	14
Fitted Bkgr events	15.41 ± 3.91	14.20 ± 5.41	16.63 ± 5.01
Fitted W $\rightarrow \mu\nu$ events	6.25 ± 1.76	6.42 ± 2.33	8.50 ± 2.56
Fitted Z $\rightarrow \mu\mu$ events	0.49 ± 0.18	0.47 ± 0.19	0.53 ± 0.19
Fitted $t\bar{t}$ events	1.80 ± 1.14	1.73 ± 1.46	3.10 ± 1.67
Fitted VV events	2.69 ± 0.62	1.84 ± 0.65	2.17 ± 0.65
Fitted W t events	4.10 ± 0.94	3.65 ± 1.29	2.21 ± 0.66
Fitted W $\rightarrow \tau\nu$ events	0.08 ± 0.02	0.09 ± 0.03	0.11 ± 0.03
Fitted LQ events	$5.43^{+4.03}_{-5.43}$	$3.88^{+4.70}_{-3.88}$	$4.76^{+5.05}_{-4.76}$
MC exp. events	14.12 ± 4.23	10.72 ± 7.68	10.89 ± 6.15
MC exp. W $\rightarrow \mu\nu$ events	5.56 ± 1.97	4.77 ± 3.49	5.51 ± 3.16
MC exp. Z $\rightarrow \mu\mu$ events	0.46 ± 0.18	0.35 ± 0.26	0.36 ± 0.22
MC exp. $t\bar{t}$ events	1.82 ± 1.11	$1.42^{+1.53}_{-1.42}$	2.03 ± 1.79
MC exp. VV events	2.46 ± 0.69	1.38 ± 0.97	1.44 ± 0.78
MC exp. W t events	3.75 ± 1.05	2.74 ± 1.93	1.47 ± 0.80
MC exp. W $\rightarrow \tau\nu$ events	0.07 ± 0.02	0.07 ± 0.05	0.08 ± 0.04
MC exp. LQ events	2.77 ± 0.46	2.06 ± 0.30	1.53 ± 0.19

Table B.9: Event yields in signal regions before and after the fit for an LQ mass of 1400, 1450 and 1500 GeV.

Event yields	SR 1600 GeV	SR 1700 GeV
Observed events	14	13
Fitted Bkgr events	12.23 ± 3.31	12.12 ± 3.64
Fitted W $\to \mu\nu$ events	6.18 ± 1.79	6.32 ± 1.89
Fitted Z $\to \mu\mu$ events	0.32 ± 0.11	0.45 ± 0.18
Fitted $t\bar{t}$ events	2.51 ± 1.61	1.34 ± 1.30
Fitted VV events	1.49 ± 0.35	2.10 ± 0.55
Fitted W t events	1.65 ± 0.39	1.81 ± 0.47
Fitted W $\to \tau\nu$ events	0.08 ± 0.02	0.08 ± 0.02
Fitted LQ events	$1.03^{+3.08}_{-1.03}$	$1.37\_1.37^{+2.16}$
MC exp. events	10.47 ± 3.25	10.03 ± 3.65
MC exp. W $\to \mu\nu$ events	5.19 ± 1.76	4.91 ± 1.96
MC exp. Z $\to \mu\mu$ events	0.30 ± 0.11	0.38 ± 0.17
MC exp. $t\bar{t}$ events	2.11 ± 1.54	1.47 ± 1.25
MC exp. VV events	1.33 ± 0.34	1.71 ± 0.56
MC exp. W t events	1.47 ± 0.38	1.48 ± 0.48
MC exp. W $\to \tau\nu$ events	0.07 ± 0.02	0.07 ± 0.02
MC exp. LQ events	0.89 ± 0.11	0.51 ± 0.07

Table B.10: Event yields in signal regions before and after the fit for an LQ mass of 1600 and 1700 GeV.

Bibliography

[1] "Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?", Annalen Der Physik (1905).

[2] ATLAS Collaboration, G. Aad et al., 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, arXiv:1207.7214. url: https://indico.cern.ch/event/197461/

[3] "On the inevitability and the possible structures of supercivilizations", The search for extraterrestrial life: Recent developments; Proceedings of the Symposium, Boston, MA, June 18-21, (1984) (A86-38126 17-88). Dordrecht, D. Reidel Publishing Co., 1985, p. 497-504.

[4] S. Weinberg, Eur. Phys. J. C34, 5 (2004), [hep-ph/0401010], and references therein.

[5] "Search for New Quarks Suggested by the Superstring", V.D. Angelopoulos (CERN & Oxford U.), John R. Ellis, H. Kowalski, Dimitri V. Nanopoulos, N.D. Tracas (CERN), F. Zwirner (CERN & SISSA, Trieste & INFN, Padua). Oct 1986. 34 pp. Published in Nucl.Phys. B292 (1987) 59-92, CERN-TH-4578/86, MAD/TH/86-26, DOI: 10.1016/0550-3213(87)90637-7

[6] "Indirect Searches for Leptoquarks at Present and Future Colliders", Herbert K. Dreiner, John R. Ellis, Dimitri V. Nanopoulos, N.D. Tracas, N.D. Vlachos (CERN). Sep 1987. 9 pp. Published in Mod.Phys.Lett. 3A (1988) 443, CERN-TH-4837/87, ITP-UH-13/87, MAD/TH/87-25, DOI: 10.1142/S0217732388000520

[7] "Quarks & Leptons: An Introductory Course in Modern Particle Physics", Francis Halzen, Alan D. Martin (1983), John Wiley & sons

[8] S. Weinberg, A Model of Leptons, Phys.Rev.Lett. 19 (1967) 1264-1266.

[9] S. Glashow, Partial Symmetries of Weak Interactions, Nucl.Phys. 22 (1961) 579-588.

[10] G. Zweig, An SU(3) model for strong interaction symmetry and its breaking, Version 1.

[11] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, Advantages of the Color Octet Gluon Picture, Phys.Lett. B47 (1973) 365-368.

[12] D. J. Gross and F. Wilczek, Ultraviolet Behavior of Nonabelian Gauge Theories, Phys.Rev.Lett. 30 (1973) 1343-1346

[13] H. D. Politzer, Reliable Perturbative Results for Strong Interactions, Phys.Rev.Lett. 30 (1973) 1346-1349.

[14] M. Kuze and Y. Sirois, Prog. Part. Nucl. Phys. 50, 1 (2003), [hep-ex/0211048].

[15] W. Buchmüller, R. Rückl and D. Wyler, Phys. Lett. B191, 442 (1987).

[16] J. Blümlein and R. Rückl, Phys. Lett. B304, 337 (1993).

[17] J. Blümlein, E. Boos and A. Kryukov, Phys. Lett. B392, 150 (1997), [hep-ph/9610506].

[18] R. Placakyte, Parton Distribution Functions, arXiv: 1111.5452 [hep-ph].

[19] M.Krämer et al. Pair production of scalar leptoquarks at the CERN LHC. Phys.Rev.D, 71:057503 (2005).

[20] M. Benedikt, P. Collier, V. Mertens, J. Poole, and K. Schindl, LIIC Design Report. 3. The LHC injector chain.

[21] 0. Bulling, P. Collier, P. Lebrun, S. Myers, R. Ostojic, et al., LHC Design Report. 2. The LHC infrastructure and general services.

[22] 0. S. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, et al., LHC Design Report Vol.1: The LHC Main Ring.

[23] L. Evans and P. Bryant, LHC Machine, JINST 3 (2008) 808001.

[24] ATLAS Collaboration, G. Aad et al., The ATLAS Experiment at the CERN Large Hadron Collider, JINST 3 (2008) S08003.

[25] ATLAS Collaboration, G. Aad et al., ATLAS inner detector: Technical design report. Vol. 1.

[26] ATLAS Collaboration, G. Aad et al., ATLAS inner detector: Technical design report. Vol. 2.

[27] ATLAS Collaboration, G. Aad et al., Magnet System. Technical Design Report ATLAS. CERN, Geneva (1997).

[28] ATLAS Collaboration, G. Aad et al., ATLAS muon spectrometer: Technical design report.

[29] ATLAS Collaboration, G. Aad et al., Performance of the ATLAS Trigger System in 2010, Eur.Phys.J. C72 (2012) 1849, arXiv:1110.1530 [hep-ex].

[30] ATLAS Collaboration, G. Aad et al., Technical Design Report for the Phase-II Upgrade of the ATLAS Muon Spectrometer (2017), CERN-LHCC-2017-017 ; ATLAS-TDR-026.

[31] "Search for pair - produced second generation scalar leptoquarks decaying to two jets, one muon and E_T^{miss} in proton - proton collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector.", ATL-COM-PHYS-2018-1004 (2018), url: https://cds.cern.ch/record/2629484

[32] ATLAS Collaboration (2011), CERN-PH-EP-2011-050, arXiv:1104.4481 [hep-ex].

[33] ATLAS Collaboration (2012), CERN-PH-EP-2012-056, arXiv:1203.3172 [hep-ex].

[34] T. Sjostrand, S. Mrenna, and P. Z. Skands, Comput.Phys.Commun. 178 (2008) 852–867, arXiv:0710.3820 [hep-ph].

[35] Tech. Rep. ATL-PHYS-PUB-2012-003, CERN, Geneva (2012).

[36] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, et al., Phys.Rev. D78 (2008) 013004, arXiv:0802.0007 [hep-ph].

[37] P. Nason, A New method for combining NLO QCD with shower Monte Carlo algorithms, JHEP 0411 (2004) 040, arXiv: hep-ph/0409146 [hep-ph].

[38] S. Frixione, P. Nason and C. Oleari, Matching NLO QCD computations with Parton Shower simulations: the POWHEG method, JHEP 0711 (2007) 070, arXiv: 0709.2092 [hep-ph].

[39] G. Corcella, I. Knowles, G. Marchesini, S. Moretti, K. Odagiri, et al., JHEP 0101 (2001) 010, arXiv:hep-ph/0011363 [hep-ph].

[40] Tech. Rep. ATL-PHYS-PUB-2011-008, CERN, Geneva, (2011).

[41] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, et al., JHEP 0902 (2009) 007, arXiv:0811.4622 [hep-ph].

[42] S. Frixione, E. Laenen, P. Motylinski, and B. R. Webber, JHEP 0603 (2006) 092, arXiv:hep-ph/0512250 [hep-ph].

[43] S. Frixione, E. Laenen, P. Motylinski, B. R. Webber, and C. D. White, JHEP 0807 (2008) 029, arXiv:0805.3067 [hep-ph].

[44] H.-L. Lai et al., New parton distributions for collider physics, Phys. Rev. D82 (2010) 074024, arXiv: 1007.2241 [hep-ph].

[45] ATLAS Collaboration, G. Aad et al., Phys.Rev. D87 no. 11, (2013) 112001, arXiv:1210.2979 [hep-ex].

[46] CMS Collaboration, S. Chatrchyan et al., Eur.Phys.J. C73 no. 2, (2013) 2283, arXiv:1210.7544 [hep-ex].

[47] ATLAS Collaboration, G. Aad et al., arXiv:1408.3179 [hep-ex]. Submitted to EPJC.

[48] Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton - proton collision data, ATLAS Collaboration (2014), arXiv:1407.3935 [hep-ph].

[49] ATLAS Collaboration, G. Aad et al., Eur.Phys.J. C74 no. 11, (2014) 3130, arXiv:1407.3935 [hep-ex].

[50] M. Cacciari, G. P. Salam and G. Soyez, The anti-k_t jet clustering algorithm, JHEP 04 (2008) 063, arXiv:0802.1189 [hep-ph].

[51] ATLAS Collaboration, Tech. Rep. ATLAS-CONF-2012-020, CERN, Geneva (2012).

[52] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, Eur.Phys.J. C71 (2011) 1554, arXiv:1007.1727 [physics.data-an].

[53] ROOT Collaboration, K. Cranmer, G. Lewis, L. Moneta, A. Shibata and W. Verkerke, CERN-OPEN-2012-016 (2012).

[54] L. Moneta et al., The RooStats Project, PoS ACAT2010 (2010) 057, arXiv: 1009.1003 [physics.data-an].

[55] W. Verkerke and D. P. Kirkby, eConf C0303241, MOLT007 (2003), [physics/0306116].

[56] R. Brun and F. Rademakers, Nucl.Instrum.Meth. A389, 81 (1997).

[57] I. Antcheva et al., Comput.Phys.Commun. 182, 1384 (2011).

[58] "Searches for scalar leptoquarks and differential cross section measurements in dilepton - dijet events in proton - proton collisions at a centre - of - mass energy of $\sqrt{s}$ = 13 TeV with the ATLAS experiment", ATL-COM-PHYS-2018-171 (2018), url: https://cds.cern.ch/record/2306032/files/ATL-COM-PHYS-2018-171.pdf

[59] G. Aad et al., Search for first generation scalar leptoquarks in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector, Phys. Lett. B709 (2012) 158, [Erratum: Phys. Lett.B711,442(2012)], arXiv: 1112.4828 [hep-ex].

[60] G. Aad et al., Search for second generation scalar leptoquarks in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector, Eur. Phys. J. C72 (2012) 2151, arXiv: 1203.3172 [hep-ex].

[61] A. Hoecker et al., TMVA: Toolkit for Multivariate Data Analysis, PoS ACAT (2007) 040, arXiv: physics/0703039.

[62] M. Baak et al., HistFitter software framework for statistical data analysis, Eur.Phys.J. C75 (2015) 153, arXiv: 1410.1280 [hep-ex].

[63] ATLAS Collaboration, ATLAS computing: Technical design report (2005).

[64] ATLAS Collaboration, The ATLAS Simulation Infrastructure, Eur.Phys.J. C70 (2010) 823, arXiv: 1005.4568 [physics.ins-det].

[65] S. Agostinelli et al., GEANT4: A Simulation Toolkit, Nucl. Instrum. Meth. A506 (2003) 250.

[66] T. Mandal, S. Mitra and S. Seth, Pair Production of Scalar Leptoquarks at the LHC to NLO Parton Shower Accuracy, Phys. Rev. D93 (2016) 035018, arXiv: 1506.07369 [hep-ph].

[67] T. Sjöstrand, S. Mrenna and P. Z. Skands, A Brief Introduction to Pythia 8.1, Comput.Phys.Commun. 178 (2008) 852, arXiv: 0710.3820 [hep-ph].

[68] ATLAS Collaboration, ATLAS Pythia 8 tunes to 7 TeV data, ATL-PHYS-PUB-2014-021 (2014) url: http://cdsweb.cern.ch/record/1966419.

[69] R. D. Ball et al., Parton distributions for the LHC Run II, JHEP 04 (2015) 040, arXiv: 1410.8849 [hep-ph].

[70] M. Krämer, Leptoquark hadroproduction Twiki, [Online; accessed 12-December-2015] (2015), url: http://web.physik.rwth-aachen.de/service/wiki/bin/view/Main/LeptoquarkProduction.

[71] ATLAS Collaboration, "Monte Carlo Generators for the Production of a W or Z/γ^* Boson in Association with Jets at ATLAS in Run 2", ATL-PHYS-PUB-2016-003 (2016), url: https://cds.cern.ch/record/2120133.

[72] T. Gleisberg and S. Höche, Comix, a new matrix element generator, JHEP 0812 (2008) 039, arXiv: 0808.3674 [hep-ph].

[73] F. Cascioli, P. Maierhofer and S. Pozzorini, Scattering Amplitudes with Open Loops, Phys. Rev. Lett. 108 (2012) 111601, arXiv: 1111.5206 [hep-ph].

[74] S. Schumann and F. Krauss, A Parton shower algorithm based on Catani-Seymour dipole factorisation, JHEP 03 (2008) 038, arXiv: 0709.1027 [hep-ph].

[75] S. Höche et al., QCD matrix elements + parton showers: The NLO case, JHEP 04 (2013) 027, arXiv: 1207.5030 [hep-ph].

[76] S. Alioli et al., A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX, J. High Energy Phys. 1006 (2010) 043, arXiv: 1002.2581 [hep-ph].

[77] S. Alioli, S.-O. Moch and P. Uwer, Hadronic top-quark pair-production with one jet and parton showering, JHEP 01 (2012) 137, arXiv: 1110.5251 [hep-ph].

[78] E. Re, Single-top Wt-channel production matched with parton showers using the POWHEG method, Eur. Phys. J. C71 (2011) 1547, arXiv: 1009.2450 [hep-ph].

[79] P. Artoisenet et al., Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations, JHEP 1303 (2013) 015, arXiv: 1212.3460 [hep-ph].

[80] T. Sjostrand, S. Mrenna and P. Z. Skands, Pythia 6.4 Physics and Manual, JHEP 05 (2006) 026, arXiv: hep-ph/0603175.

[81] J. Pumplin et al., New generation of parton distributions with uncertainties from global QCD analysis, JHEP 07 (2002) 012, arXiv: hep-ph/0201195 [hep-ph].

[82] P. Z. Skands, Tuning Monte Carlo Generators: The Perugia Tunes, Phys. Rev. D82 (2010) 074018, arXiv: 1005.3457 [hep-ph].

[83] D. J. Lange, The EvtGen particle decay simulation package, Nucl. Instrum. Meth. A462 (2001) 152.

[84] W. Buttinger, "Using Event Weights to account for differences in Instantaneous Luminosity and Trigger Prescale in Monte Carlo and Data", tech. rep. ATL-COM-SOFT-2015-119, CERN (2015), url: https://cds.cern.ch/record/2014726.

[85] G. Aad et al., Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s} = 13$ TeV, Eur. Phys. J. C76 (2016) 292, arXiv: 1603.05598 [hep-ex].

[86] ATLAS Collaboration, "ATLAS Muon Combined Performance in 2016", MUON-2016-004 (2016), url: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/MUON-2016-004/.

[87] W Lampl et al., "Calorimeter Clustering Algorithms: Description and Performance", tech. rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN (2008), url: https://cds.cern.ch/record/1099735.

[88] ATLAS Collaboration, Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV, Eur. Phys. J. C73 (2013) 2304, arXiv: 1112.6426 [hep-ex].

[89] C. Cojocaru et al., Hadronic Calibration of the ATLAS Liquid Argon End-Cap Calorimeter in the Pseudorapidity Region $1.6 < |\eta| < 1.8$ in Beam Tests, Nucl. Instrum. Meth. A 531 (2004) 400 - 432, arXiv: physics/0407009 [hep-ex].

[90] M. Cacciari, G. P. Salam and G. Soyez, The anti-k_t jet clustering algorithm, JHEP 04 (2008) 063, arXiv: 0802.1189 [hep-ph].

[91] "Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS", tech. rep. ATLAS-CONF-2015-037, CERN (2015), url: http://cds.cern.ch/record/2044941.

[92] ATLAS Collaboration, Optimisation of the ATLAS b - tagging performance for the 2016 LHC Run, ATL-PHYS-PUB-2016-012 (2016), url: http://cds.cern.ch/record/2160731.

[93] ATLAS Collaboration, "Public Plots: Missing Transverse Momentum Distribution and Performance in 2016 data", JETM-2016-008 (2016) url: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/JETM-2016-008/.

[94] "Performance of missing transverse momentum reconstruction for the ATLAS detector in the first proton-proton collisions at at $\sqrt{s}$ = 13 TeV", tech. rep. ATL-PHYS-PUB-2015-027, CERN (2015) url: http://cds.cern.ch/record/2037904.

[95] M. Cacciari and G. P. Salam, Pileup subtraction using jet areas, Phys. Lett. B659 (2008) 119, arXiv: 0707.1378 [hep-ph].

[96] M. Cacciari, G. P. Salam and G. Soyez, The Catchment Area of Jets, JHEP 04 (2008) 005, arXiv: 0802.1188 [hep-ph].

[97] D. Adams et al., "Recommendations of the Physics Objects and Analysis Harmonisation Study Groups 2014", tech. rep. ATL-COM-PHYS-2014-451, CERN (2014), url: https://cds.cern.ch/record/1700874.

[98] S. Farell, "OverlapRemovalTool",
url: https://svnweb.cern.ch/trac/atlasoff/browser/PhysicsAnalysis/AnalysisCommon/AssociationUtils/trunk/doc/README.rst
[Online; accessed 7-April-2017].

[99] C. Gini, Variabilitá e mutabilitá (1912).

Ο Wheeler κι εγώ σκεφτήκαμε ότι το επόμενο θέμα έπρεπε να αφορά την κβαντική θεωρία της ηλεκτροδυναμικής, η οποία παρουσίαζε κάποιες δυσκολίες με την αυτοεπίδραση του ηλεκτρονίου. Θεωρήσαμε καλό να προσπαθήσουμε να ξεπεράσουμε πρώτα τη δυσκολία στην κλασική φυσική και μετά να το δούμε σε σχέση με την κβαντική θεωρία.

Όταν τα καταφέραμε με την κλασική θεωρία, ο Wheeler μου είπε: "Feynman, είσαι νέος και σχετικά άπειρος· πρέπει να δώσεις μια διάλεξη για να συνηθίσεις τις ομιλίες. Θα χρειαστεί να μιλήσεις πολλές φορές για την αντιμετώπιση του προβλήματος! Εν τω μεταξύ, εγώ θα εξετάσω το μέρος που αφορά την κβαντική θεωρία και θα δώσω μια διάλεξη αργότερα".

Έτσι έφτασε η στιγμή και για την πρώτη μου παρουσίαση. Ο Wheeler κανόνισε με τον Eugene Wigner να με βάλει στο πρόγραμμα των διαλέξεων. Μία ή δύο μέρες πριν, είχα συναντήσει τον Wigner στον διάδρομο: "Feynman" μου είπε "νομίζω ότι η δουλειά που κάνεις με τον Wheeler είναι πολύ ενδιαφέρουσα και γι' αυτό έχω καλέσει τον Russell στη διάλεξη". Ο διάσημος αστρονόμος Henry Norris Russell θα ερχόταν να με ακούσει!

"Έδειξε επίσης ενδιαφέρον και ο καθηγητής von Neumann" συνέχισε ο Wigner, ενώ εγώ δεν πίστευα στ' αυτιά μου. Ο von Neumann ήταν ο μεγαλύτερος μαθηματικός του Princeton. "Ίσως έρθει και ο καθηγητής Pauli που τυχαίνει να βρίσκεται εδώ· του έστειλα πρόσκληση". Ο Pauli ήταν διάσημος φυσικός - κι εγώ είχα ήδη γίνει κατάχλωμος. Οπότε ο Wigner πρόσθεσε: "Ο καθηγητής Einstein έρχεται σπάνια στις εβδομαδιαίες διαλέξεις μας, αλλά η εργασία σου είναι τόσο ενδιαφέρουσα, ώστε τον ενημέρωσα σχετικά και θα έρθει".

Πρέπει να είχα γίνει πράσινος, διότι ο Wigner προσπάθησε να με καθησυχάσει λέγοντας: "Όχι, όχι! Μην ανησυχείς! Πρέπει όμως να σε προειδοποιήσω: εάν ο καθηγητής Russell αποκοιμηθεί - που αναμφίβολα θα το πάθει - δε σημαίνει ότι η διάλεξη είναι ανούσια· σε όλες τις διαλέξεις τον παίρνει ο ύπνος. Αντίθετα, αν δεις τον καθηγητή Pauli να κουνάει συνεχώς το κεφάλι του δείχνοντας ότι συμφωνεί με όσα λες, μη δώσεις σημασία· πάσχει από τρομώδη παράλυση...".

Richard P. Feynman, "Σίγουρα θα αστειεύεστε, κύριε Φάινμαν"[1]
εκδόσεις Κάτοπτρο

[1]τίτλος πρωτότυπου: "Surely you 're joking, mr Feynman"· προτείνεται να διαβαστεί σε συνδυασμό με το δεύτερο αυτοβιογραφικό βιβλίο του R. Feynman "Τι σε νοιάζει εσένα τι σκέφτονται οι άλλοι;", εκδόσεις Πολιτεία (τίτλος πρωτότυπου: "What do you care what other people think?")

Δρόμοι παλιοί που αγάπησα και μίσησα ατέλειωτα
κάτω από τους ίσκιους των σπιτιών να περπατώ
νύχτες των γυρισμών αναπότρεπτες κι η πόλη νεκρή
Την ασήμαντη παρουσία μου βρίσκω σε κάθε γωνιά
κάμε να σ' ανταμώσω κάποτε φάσμα χαμένο του πόθου μου κι εγώ
Ξεχασμένος κι ατίθασος να περπατώ
κρατώντας μια σπίθα τρεμόσβηστη στις υγρές μου παλάμες
Και προχωρούσα μέσα στη νύχτα χωρίς να γνωρίζω κανένα
κι ούτε κανένας με γνώριζε.

Μανόλης Αναγνωστάκης, "Δρόμοι παλιοί"