Universidade de Lisboa
Instituto Superior Técnico

Top quark physics and search for physics beyond the Standard Model at the Large Hadron Collider

Oleksii Toldaiev

Supervisor: Doctor João Manuel Coelho dos Santos Varela
Co-Supervisor: Doctor Michele Gallinaro

Thesis approved in public session to obtain the PhD Degree in
Physics

Jury final classification: Pass with Distinction and Honour

Jury

President: Chairman of the IST Scientific Board

Members of the Committee:
Doutor Jorge Manuel Rodrigues Crispim Romão, Professor Catedrático do Instituto Superior Técnico da Universidade de Lisboa
Doutor António Joaquim Onofre de Abreu Ribeiro Gonçalves, Professor Associado (com Agregação) da Escola de Ciências da Universidade do Minho
Doutor Michele Gallinaro, Professor Auxiliar Convidado do Instituto Superior Técnico da Universidade de Lisboa
Doutor Pedro Vieira de Castro Ferreira da Silva, Research Physicist, European Organization for Nuclear Research (CERN)
Doutor Nuno Teotónio Viegas Guerreiro Leonardo, Investigador Principal do LIP-Laboratório de Instrumentação e Física Experimental de Partículas, Lisboa, individualidade reconhecida na área científica em que se insere a tese.

Funding Institutions
Fundação para a Ciência e a Tecnologia, Fellowship PD/BD/113489/2015 (IDPASC programme)

2020

Abstract

In 2016 the LHC produced a large sample of $\sqrt{s} = 13\,\mathrm{TeV}$ proton-proton collisions for the first time. The energy and luminosity of the data allow exploration of the top quark to a greater detail than ever before, involving standard model (SM) and hypothetical beyond SM (BSM) interactions. This thesis documents a measurement of the top quark pair production cross section in the $\mathrm{t\bar{t}} \to (\ell\nu_\ell)(\tau_\mathrm{h}\nu_\tau)\mathrm{b\bar{b}}$ channel, where ℓ is either an electron or a muon. The analysis is performed using a data sample corresponding to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$ collected in proton-proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ by the CMS detector. Assuming a top quark mass of 172.5 GeV, the measured total $\mathrm{t\bar{t}}$ cross section $\sigma_{\mathrm{t\bar{t}}}(\ell\tau_\mathrm{h}) = 781 \pm 7\,(\mathrm{stat.}) \pm 62\,(\mathrm{syst.}) \pm 20\,(\mathrm{lum.})\,\mathrm{pb}$ is in agreement with the SM expectations. This is the first measurement of the $\mathrm{t\bar{t}}$ production cross section in proton-proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ that explicitly includes τ leptons, and it improves the relative precision with respect to the 7 and 8 TeV results. The ratio of the cross section in the $\ell\tau_\mathrm{h}$ final state to the light dilepton cross section $R_{\ell\tau_\mathrm{h}/\ell\ell} = 0.973 \pm 0.009\,(\mathrm{stat.}) \pm 0.066\,(\mathrm{syst.})$ and the ratio of the partial to the total width of the top quark decay $\Gamma(\mathrm{t} \to \tau\nu_\tau\mathrm{b})/\Gamma_\mathrm{total} = 0.1050 \pm 0.0009\,(\mathrm{stat.}) \pm 0.0071\,(\mathrm{syst.})$ are also measured, improving the precision over the previous results. The higher precision is achieved through a better estimation of the background with misidentified τ leptons. The SM $\mathrm{t\bar{t}}$ process provides a source of W boson decays to leptons, where the ratio of the W boson branching fractions $\mathcal{B}(\mathrm{W} \to \tau\nu_\tau)/\mathcal{B}(\mathrm{W} \to \ell\nu_\ell)$ can be measured to a high precision as a test of the lepton universality in W bosons. Using the methods developed in the cross section analysis, a feasibility study of the test in $\mathrm{t\bar{t}}$ dilepton final states is carried out for the full Run2 data sample. The study demonstrates a possibility to achieve a measurement with the overall relative uncertainty of about 3%. It would be a significant contribution to the current world-best measurements.

Keywords: high energy physics, top quarks, tau leptons, LHC, CMS.

Resumo

No ano 2016, o Large Hadron Collider (LHC) produziu pela primeira vez uma grande amostra de colisões protão-protão $\sqrt{s} = 13\,\mathrm{TeV}$. A energia e a luminosidade dos dados permitem a exploração do quark top com mais detalhes do que nunca, envolvendo o Modelo Padrão (MP) e as hipotéticas interações além do MP (AMP). Esta tese documenta uma medida da secção transversal da produção dos pares de top quarks no canal $\mathrm{t\bar{t}} \to (\ell\nu_\ell)(\tau_\mathrm{h}\nu_\tau)\mathrm{b\bar{b}}$, em que ℓ é um elétrão ou um muão. A análise é realizada usando uma amostra de dados correspondente a uma luminosidade integrada de $35.9\,\mathrm{fb}^{-1}$ coletada de colisões protão-protão em $\sqrt{s} = 13\,\mathrm{TeV}$ pelo detector CMS. Supondo uma massa de top quarks de $172.5\,\mathrm{GeV}$, a secção transversal total $\sigma_{\mathrm{t\bar{t}}}(\ell\tau_\mathrm{h}) = 781 \pm 7\,(\mathrm{stat.}) \pm 62\,(\mathrm{syst.}) \pm 20\,(\mathrm{lum.})\,\mathrm{pb}$ está de acordo com as expetativas do MP. Esta é a primeira medição da secção transversal da produção $\mathrm{t\bar{t}}$ em colisões protão-protão em $\sqrt{s} = 13\,\mathrm{TeV}$ que inclui explicitamente τ leptões e melhora a precisão relativa em relação aos resultados a 7 e 8 TeV. A proporção da secção transversal no estado final $\ell\tau_\mathrm{h}$ para a secção transversal de dileptões leves $R_{\ell\tau_\mathrm{h}/\ell\ell} = 0.973 \pm 0.009\,(\mathrm{stat.}) \pm 0.066\,(\mathrm{syst.})$ e a proporção da largura parcial para a largura total do decaimento do top quark $\Gamma(\mathrm{t}\tau\nu_\tau\mathrm{b})/\Gamma_\mathrm{total} = 0.1050 \pm 0.0009\,(\mathrm{stat.}) \pm 0.0071\,(\mathrm{syst.})$ também são medidos, melhorando a precisão em relação aos resultados anteriores. A maior precisão é alcançada através de uma melhor estimativa do fundo com os τ leptões mal identificados. O processo MP $\mathrm{t\bar{t}}$ fornece uma fonte de decaimento do bosão W para leptões, em que a proporção das taxas de decaimento do bosão W $\mathcal{B}(\mathrm{W} \to \tau\nu_\tau)/\mathcal{B}(\mathrm{W} \to \ell\nu_\ell)$ pode ser medida com alta precisão como teste da universalidade dos leptões nos bosões W. Usando os métodos desenvolvidos na análise de secção transversal, é realizado um estudo de viabilidade do teste nos estados finais de $\mathrm{t\bar{t}}$ com dois leptões para a amostra de dados completa do Run2. O estudo demonstra uma possibilidade de obter uma medição com a incerteza relativa geral de cerca de 3%. O resultado medido dos decaimentos hadrónicos dos taus seria uma contribuição significativa para as medições atuais.

Palavras-chave: física de alta energia, quark top, leptão tau, LHC, CMS.

Resumo Alargado

O progresso da física das altas energias depende em grande parte da capacidade de descoberta da física experimental. A iniciativa mundial de alcançar as novas fronteiras da física de alta energia resultou na construção do Large Hadron Collider (LHC) no complexo de aceleração de partículas na Organização Europeia para a Pesquisa Nuclear (CERN, o acrônimo para Conseil Européen pour la Recherche Nucléaire). O novo acelerador entrou em operação em 2009. E em novembro daquele ano produziu feixes de protões com a energia no referencial do centro de massa de $\sqrt{s} = 2.36\,\mathrm{TeV}$, batendo o recorde anterior de $1.96\,\mathrm{TeV}$ de colisões protão-antiprotão no acelerador Tevatron. Em 2010, o LHC atingiu $7\,\mathrm{TeV}$ de energia nas colisões protão-protão e abriu o novo programa de pesquisa. O estágio Run1 da operação do LHC durou de 2010 a 2012. Durante esse período, o acelerador produziu colisões de 7 e $8\,\mathrm{TeV}$. Forneceu grande número de colisões de protões e de íons para os principais detectores, incluindo a amostra de cerca de $30\,\mathrm{fb}^{-1}$ de colisões protão-protão coletadas pelo Compact Muon Solenoid (CMS). Esses dados serviram para realizar muitas medições, incluindo a descoberta do bosão de Higgs. No ano 2016, o LHC iniciou o estágio Run2 de sua operação e produziu pela primeira vez uma grande amostra de colisões protão-protão $\sqrt{s} = 13\,\mathrm{TeV}$. O Run2 durou até 2018. Nesse período, o LHC forneceu amostras de cerca de $160\,\mathrm{fb}^{-1}$ de colisões protão-protão aos detectores CMS e ATLAS no tunel do acelerador.

A alta energia e luminosidade dos dados permitem a exploração da física do quark top com mais detalhes do que nunca. A partícula elementar com a maior massa, o quark top participa nas interações do regime de alta energia do Modelo Padrão (MP) e nas hipotéticas interações além do MP (AMP). Eles decaem sem hadronização, servindo assim como uma sonda única na física das interações fortes. Portanto, o estudo do top quark contribui para muitas áreas da física de alta energia e estabelece limites rigorosos para a pesquisa teórica.

Esta tese documenta uma medida da seção transversal de produção dos pares de top quarks no canal $\mathrm{t\bar{t}} \to (\ell\nu_\ell)(\tau_\mathrm{h}\nu_\tau)\mathrm{b\bar{b}}$, em que ℓ é um elétron ou um muão, que é realizada usando uma amostra de dados correspondente a uma luminosidade integrada de $35.9\,\mathrm{fb}^{-1}$ coletada a $\sqrt{s} = 13\,\mathrm{TeV}$ pelo detector CMS no ano de 2016. A medição é realizada nos dados de apenas um ano do Run2, pois tal é suficiente para reduzir a incerteza estatística a um nível desprezável. Envolve exclusivamente leptões e quarks de terceira geração que, devido às suas grandes massas, podem ser particularmente sensíveis às contribuições do AMP e fornece acesso a física interessante de alta energia no MP. A análise baseia-se nas ferramentas usadas pelo LIP nas medições em 7 e $8\,\mathrm{TeV}$ de Run1, melhorando a abordagem das principais

incertezas sistemáticas. Supondo uma massa de quark top de 172.5 GeV, a medida seção transversal total $\sigma_{\mathrm{t\bar{t}}}(\ell\tau_{\mathrm{h}}) = 781 \pm 7\,(\mathrm{stat.}) \pm 62\,(\mathrm{syst.}) \pm 20\,(\mathrm{lum.})\,\mathrm{pb}$ está de acordo com as expetativas do MP. Esta é a primeira medição da seção transversal da produção $\mathrm{t\bar{t}}$ em colisões protão-protão em $\sqrt{s} = 13\,\mathrm{TeV}$ que inclui explicitamente τ leptões e melhora a precisão em relação dos resultados em 7 e 8 TeV. A proporção da seção transversal no estado final $\ell\tau_{\mathrm{h}}$ para a seção transversal de leves dileptões $R_{\ell\tau_{\mathrm{h}}/\ell\ell} = 0.973 \pm 0.009\,(\mathrm{stat.}) \pm 0.066\,(\mathrm{syst.})$ e a proporção da largura parcial para a largura total do decaimento do quark top $\Gamma(\mathrm{t} \to \tau\nu_{\tau}\mathrm{b})/\Gamma_{\mathrm{total}} = 0.1050 \pm 0.0009\,(\mathrm{stat.}) \pm 0.0071\,(\mathrm{syst.})$ também são medidos, melhorando a precisão em relação aos resultados anteriores. A maior precisão é alcançada através de uma melhor estimativa do fundo com os τ leptões incorrectamente identificados.

Os métodos usados para estimar esse fundo são cuidadosamente estudados na tese. O método escolhido para a análise restringe tanto o efeito desse fundo na medida de seção transversal que deixa de ser uma contribuição dominante para a incerteza sistemática geral. A maior incerteza restante vem da identificação do decaimento de um τ leptão em hadrões. Uma redução significativa nessa incerteza permitiria perseguir objetivos físicos mais avançados na análise. Uma dessas possibilidades é desenvolvida como um teste de universalidade dos leptões nos $\mathrm{t\bar{t}}$ canais.

O processo MP $\mathrm{t} \to (\tau\nu_{\tau})\mathrm{b}$ envolve diretamente o canal $\mathrm{W} \to \tau\nu_{\tau}$ do decaimento do boson W, que apresenta um interesse em vista é particularmente interessante tendo em conta dos desvios observados na física do quark b e os resultados mais recentes sobre a universalidade dos leptões neste canal. O decaimento do par dos quarks top fornece uma fonte de decaimento do bosão W para leptões, onde a proporção das taxas de decaimento do bosão W $\mathcal{B}(\mathrm{W} \to \tau\nu_{\tau})/\mathcal{B}(\mathrm{W} \to \ell\nu_{\ell})$ pode ser medida com alta precisão como um teste da universalidade dos leptões nos bosões W. Essa medida é feita apenas nos estados finais com decaimento hadrônico do τ, porque não é influenciada pelo modelização do sinal.

Os produtos adicionais dos decaimentos dos top quarks ajudam a selecionar uma amostra pura de eventos de sinal. E a simetria do canal permite a medição em relação ao estado final com dileptões leves para reduzir os efeitos de incertezas sistemáticas. A fim de reduzir as incertezas sistemáticas devido à identificação de leptão τ e outras fontes, a medição é realizada em uma proporção dupla com os canais Drell–Yan (DY) que incluem leptões τ. Desta forma, a relação entre os bosões W é medida em relação à relação em DY. Porém, como a universalidade lepton é confirmada com alta precisão no DY, ela não introduz uma incerteza adicional nessa medida.

Usando os métodos desenvolvidos na análise de seção transversal, é realizado um estudo de viabilidade do teste nos estados finais de $\mathrm{t\bar{t}}$ com dois leptões para a amostra de dados completa do Run2. É realizado em amostras de dados simulados.

Confirma-se que as regiões de controle concordam com os dados no número de eventos esperados e no formato das distribuições dos parâmetros que são usados na medição.

A análise mostra que a precisão da medição é limitada pela incerteza estatística. Portanto, com dados suficientes, a incerteza geral pode ser significativamente reduzida. O estudo mostra uma possibilidade de obter uma medição com a incerteza relativa geral de cerca de 3% com a amostra de dados completa do Run2. Seria o resultado medido em decaimentos hadrónicos dos leptões τ mais preciso num acelerador de hadrões e uma contribuição significativa para as melhores medições atuais.

Fui o autor principal das medidas relacionadas com seção transversal do par de quark top e ao estudo de viabilidade do teste de universalidade de leptões no decaimento do bosão W. Durante a revisão interna da medição da seção transversal na colaboração CMS e a revisão da publicação, fui a pessoa de contato para a análise e o responsável por preparar as respostas às perguntas levantadas e a implementação de sugestões.

Palavras-chave: física de alta energia, quark top, leptão tau, LHC, CMS.

To my parents, brother, and relatives.

Acknowledgements

The research for this doctoral thesis was carried out during the years 2015–2020 in the Compact Muon Solenoid (CMS) experiment at CERN and the CMS group of the Laboratório de Instrumentação e Física Experimental de Partículas (LIP). The research was funded by the Fundação para a Ciência e Tecnologia (FCT) through the PhD scholarship PD/BD/113489/2015 within the IDPASC program. Travels to various schools, workshops and CERN were funded by FCT and LIP.

Arriving to the end of this formative road, I am grateful to my supervisors, Prof. João Varela and Prof. Michele Gallinaro, for the advise, encouragement, and reminders of deadlines. A special thank you for the opportunity to participate in the prototype work for the CT-PPS project, that has been a valuable experience in detector physics and data acquisition. I am also grateful to other senior members of the LIP CMS group, in particular Dr. Pedro Silva, Dr. Nuno Leonardo, and Dr. Pedrame Bargassa, for many consultations, discussions, and suggestions.

This work would not be possible without all the fantastic effort of the CMS collaboration. All analyses in the thesis have been performed on the computing infrastructure provided by LIP, CMS, and CERN. Thank you to the IT teams for their careful and continual contribution. A special thanks to the IT team at LIP.

I would like to thank the members and stuff of the LIP laboratory for the support in professional and other matters. And a special thank you to João Pedro for the help in countless issues.

Without a doubt this work would be much harder if not for my office colleagues and the great time we had together. In particular I would like to thank Pietro Vischia, Cristovão Silva, and Bruno Galinhas for many discussions on physics analysis, and Lara Lloret for the energetic support in many issues, like getting me into CERN. Thank you to Federico, Daniele, Jonathan, Matilde, Giles, Diogo, and our tenacious master students.

Finally, I am grateful to my parents and brother for all the encouragement.

Contents

Abstract . . . i
Resumo . . . ii
Resumo Alargado . . . iii
Dedication . . . vi
Acknowledgements . . . vii
Contents . . . ix
List of Tables . . . xiii
List of Figures . . . xvii

1 Introduction **1**

2 Experimental high energy physics **5**
2.1 The standard model of particle physics . . . 5
2.2 The top quark . . . 6
2.3 The tau lepton . . . 8

3 Experimental apparatus **11**
3.1 The Large Hadron Collider (LHC) . . . 11
3.2 The Compact Muon Solenoid (CMS) detector . . . 12

4 Event reconstruction **17**
4.1 Data sample . . . 18
4.2 Event simulation . . . 20
4.3 Trigger . . . 22
4.4 Primary vertex . . . 25
4.5 Lepton reconstruction . . . 25
4.6 Jets . . . 28
4.7 Missing transverse energy . . . 29
4.8 Hadronic τ lepton reconstruction . . . 29

4.9 Calibration of the simulated events 30

5 The estimation of the background from misidentified tau leptons 35

5.1 Direct measurement of the τ_h misidentification probability 36

5.1.1 Event selection 38

5.1.2 Estimation of the misidentified τ_h background 39

5.1.3 The systematic uncertainty in the cross section measurement . 46

5.1.4 Discussion 48

5.2 The fake factor and shape fit methods 49

6 The measurements in the top quark pair dilepton final states containing a tau lepton 53

6.1 Analysis strategy 54

6.2 Event Selection 55

6.3 Background estimation 57

6.3.1 Drell-Yan 57

6.3.2 W+jets 60

6.3.3 Data-driven QCD multijet 62

6.3.4 Single top quark and dibosons 73

6.3.5 The misidentified τ_h background 73

6.4 Event yields and control distributions 77

6.5 Systematic Uncertainties 77

6.6 Fitting procedure 90

6.7 Results 94

6.7.1 Cross section in the fiducial phase space 94

6.7.2 Cross section in the full phase space 98

6.7.3 Ratio to the dilepton cross section and partial width 105

6.8 Summary 107

7 The lepton universality test in top quark pair events 109

7.1 Strategy 110

7.2 Object Definitions 112

7.2.1 The secondary vertex reconstruction 113

7.3 Event selection 115

7.4 Systematic uncertainties 119

7.5 Fit of the double ratio 121

7.6 Results 125

7.7 Discussion 131

8 Summary **135**

Bibliography **137**

List of Tables

2.1 The three classes of the top quark pair decay final states. 7
2.2 The decays of the tau lepton and their branching fractions. The symbols h^- and h^0 signify charged and neutral hadrons. 9
2.3 The lifetime of the τ lepton compared to D and B mesons. 9

4.1 The datasets from the 2016 sample that are used in this thesis. . . . 19
4.2 The main characteristics of the simulated Monte-Carlo datasets that are used in the analysis: the generator, the simulation accuracy, and the cross section of the hard process. The parton showering is performed with PYTHIA (v8.2) [1] in all processes. 22
4.3 The Monte-Carlo simulated datasets that are used in this thesis. . . 23
4.4 Additional Monte-Carlo datasets that contain some of the parton showering uncertainties in the $\mathrm{t\bar{t}}$ modelling, and the W+jets and the multijet QCD datasets for cross-check studies. The W+jets dataset is generated at NLO accuracy with MC@NLO. It is used to validate the shapes of the distributions obtained in the LO dataset of the W+jets events. The multijet QCD MC datasets are used for validation of the data-driven method to estimate the contribution from this background process. 24
4.5 HLT trigger paths used in the analyses. 25
4.6 Dominant hadronic decays of the τ lepton. 29

5.1 The expected and observed numbers of events for the signal and the backgrounds at different selection stages of the $\mathrm{e}\tau_\mathrm{h}$ final state for an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$. 39
5.2 The expected and observed numbers of events for the signal and the backgrounds at different selection stages of the $\mu\tau_\mathrm{h}$ final state for an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$. 42

5.3 Fractions of jet flavours in the control samples and in the $\mathrm{t\bar{t}}$ pre-selection. Values are taken from the simulation. 46

5.4 The weights of the control regions in the calculation of the weighted misidentification probability for the weighted average, as determined by fixing the jet composition of the average to the pre-selection. . . . 46

5.5 Weights for the measured τ_h misidentification probabilities and the estimation of the misidentified background in the $\mathrm{t\bar{t}}$ selection. The columns "Estimation" list the number of events with misidentified τ_h estimated with the measured misidentification probabilities and their weighted average calculated with the coefficients from the columns "Weight". The coefficients adjust the jet compositions in the averages to the $\mathrm{t\bar{t}}$ pre-selections. The genuine τ contributions are obtained from the simulation. The last row gives the estimation in the final selection after the application of the opposite charge requirement ϵ_OC efficiency. 47

6.1 Composition of the DY dilepton control region. Only statistical uncertainties are included. 58

6.2 Composition of the DY $\rightarrow \tau_\ell \tau_\mathrm{h}$ control region. The event yields of MC QCD are given for reference in parentheses. The MC QCD exhibits a large statistical uncertainty, therefore it is not used in the figures for this control region. The statistical uncertainties are given. 60

6.3 Composition of the events in the W+jets control regions with the electron and the muon final states. Event yields for all processes, including QCD, are estimated from MC. Only statistical uncertainties are given. 62

6.4 Event yields for the measurement of the $f_\mathrm{OC/SC}$ ratio between the OC and SC contributions of the multijet QCD background. 70

6.5 Composition of the $\mathrm{e}^\pm\mu^\mp$ control region. The statistical uncertainties are given. 73

6.6 Expected and observed event yields in the $\mathrm{e}\tau_\mathrm{h}$ final state. The signal and SM background processes are given for an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$. Statistical uncertainties are shown. 80

6.7 Expected and observed event yields in the $\mu\tau_\mathrm{h}$ final state. The signal and SM background processes are given for an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$. Statistical uncertainties are shown. 81

6.8 The nuisance parameters of the PLR fit corresponding to the systematic uncertainties (in %). Column "type" shows the treatment of the parameter: rate, shape, prior probability distribution function ($\ln N$ for log-normal distribution). Column "range" lists the 1-sigma deviation for the "rate" parameters, the "shape" parameters are varied with their recommended 1σ deviation in each event. The τ misidentification probabilities in the QCD and W+jets processes are included in the normalization uncertainties. For the simultaneous fit of two final states all nuisances parameters are correlated, except the parameters which correspond to the lepton object in the events: the efficiencies of lepton identification and trigger ("Lepton ID" and "Trigger"). . . 93

6.9 Systematic and statistical uncertainties determined from the fit to the data in the $\mathrm{e}\tau_\mathrm{h}$ and $\mu\tau_\mathrm{h}$ final states, and their combination. Uncertainties are grouped by their origin: experimental, theoretical, normalization, and extrapolation. The uncertainties in the measurement in the dilepton final state [2] used in the partial width ratio estimate are also quoted (column "Dileptons"), where the asymmetric extrapolation uncertainties are symmetrized by adding them in quadrature. As both measurements use the same data, some uncertainties in the $\ell\tau_\mathrm{h}$ and light dilepton final states are correlated, as shown in the last column. 102

6.10 The p-values of the goodness of fit tests in the separate channels and in the simultaneous fit. Three tests show comparable results: Kolmogorov-Smirnov, Anderson-Darling and the saturated model. . . 105

7.1 Expected event yields in the final states used for the double ratio. The signal and SM background processes are given for an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$. Correctly assigned events are designated as "genuine". Statistical uncertainties are shown. 118

List of Figures

1.1 Feynman diagram of the $t\bar{t}$ production with the SM decay into the dilepton final state that includes a τ lepton. 2

3.1 The map of the CERN LHC site at Geneva (left), and the scheme of the accelerator with the four main detectors and interaction points (right). 12

3.2 A schematic representation of the CMS detector and main subdetector systems. 13

3.3 A logitudinal representation of the CMS detector showing the position of the subdetectors in the η plane with the muon detectors highlighted. 14

5.1 Feynman diagram of the $t\bar{t}$ decay into the lepton+jets (semileptonic) final state. This is the main background process whenever a jet is misidentified as a τ_h. It contributes about 80% of all background events to a typical event selection. 36

5.2 Typical distributions of the τ_h misidentification probability per η of the τ_h candidate, as measured in three different event selections in the following analysis. The $e\mu$+jets selection is enriched in $t\bar{t} \to e\nu_e\mu\nu_\mu b\bar{b}$ events. The misidentification probability distributions measured in this channel were used in the Ref. [3], as one of my service tasks in the CMS collaboration. 37

5.3 Main distributions in the $e\tau_h$ pre-selected events. The multijet QCD contribution is taken from simulation and exhibits large statistical fluctuations. Only statistical uncertainties are shown. 40

5.4 Main distributions in the $\mu\tau_h$ pre-selected events. The multijet QCD contribution is taken from simulation and exhibits significant statistical fluctuations. Only statistical uncertainties are shown. 41

5.5 Misidentification probabilities as functions of the jet variables (p_T, η, r_{jet}) measured in the control samples. 44

5.6 Fake factors measured in the multijet QCD and W+jets selections. . 50

6.1 The τ_{h} p_{T} distributions for events of the $\mathrm{e}\tau_{\mathrm{h}}$ (left) and $\mu\tau_{\mathrm{h}}$ (right) final states observed prior to fitting. Distributions obtained from data (filled circles) are compared with simulation (shaded histograms). The last bin includes overflow events. The simulated contributions are normalized to the cross section values predicted in the SM. The main processes are shown: the signal, the other $\mathrm{t\bar{t}}$ processes grouped together, single top quark production, W+jets, DY processes, diboson, and multijet production. The ratio of the data to the total SM prediction is shown in the lower panel. The hatched bands indicate the systematic uncertainties and the statistical uncertainties of all simulated samples. Statistical uncertainties on the data points are not visible because of the scale of the figure. 56

6.2 Distributions of the leading lepton p_{T} and the transverse mass of the leading lepton and $p_{\mathrm{T}}^{\mathrm{miss}}$ system in the dilepton DY+jets control regions: the $\mathrm{e}^{-}\mathrm{e}^{+}$ on the left, the $\mu^{-}\mu^{+}$ is on the right. The hatched band includes statistical uncertainties. 59

6.3 Distributions of the lepton p_{T} and the tau p_{T} in the DY $\rightarrow \tau_{\ell}\tau_{\mathrm{h}}$ control region pre-selection (top) and full selection (bottom): the $\mu\tau_{\mathrm{h}}$ selection is on the left, the $\mathrm{e}\tau_{\mathrm{h}}$ is on the right. The uncertainty band includes statistical uncertainties. 61

6.4 Distributions of the lepton p_{T} (left) and the transverse mass m_{T} of the lepton and $p_{\mathrm{T}}^{\mathrm{miss}}$ system (right) in the W+jet control region with a muon in the final state. The QCD multijet contribution is taken from the simulation, like other processes. The hatched band includes the statistical uncertainty, lepton identification, b tagging, pileup reweighting and luminosity. 63

6.5 Distributions of the lepton p_{T} (left) and the transverse mass m_{T} of the lepton and $p_{\mathrm{T}}^{\mathrm{miss}}$ system (right) in the W+jet control region with an electron in the final state. The QCD multijet contribution is taken from the simulation, like other processes. The hatched band includes the statistical uncertainty, lepton identification, b tagging, pileup reweighting and luminosity. 64

6.6 Comparison between the distributions of the lepton p_{T} and the transverse mass m_{T} in the W+jets events simulated at LO (top) and NLO (bottom) accuracy. The QCD multijet contribution is taken from the simulation, like other processes. 65

6.7 The diagram of the “ABCD” method to estimate the shape and the normalization of the multijet QCD background from data. 66

6.8 Comparison of the m_T distributions in OC and SC selections in the MC QCD: the pre-selection (top) and the W+jets control region (bottom), the $\mathrm{e}\tau_\mathrm{h}$ (left) and $\mu\tau_\mathrm{h}$ (right) channels. The Y axis shows arbitrary units of event yield normalization. The error bars represent statistical uncertainties in the MC datasets. 67

6.9 Comparison of the m_T distributions in OC and SC selections in the data-driven estimation of the multijet QCD (the difference between the observed data and the non-QCD processes taken from MC): the pre-selection (top) and the W+jet control region (bottom), the $\mathrm{e}\tau_\mathrm{h}$ (left) and $\mu\tau_\mathrm{h}$ (right) channels. The Y axis shows arbitrary units of event yield normalization. The error bars represent statistical uncertainties in the MC datasets. 68

6.10 Distributions of the OC/SC factor found for different relative isolation requirements in the $\mathrm{e}\tau_\mathrm{h}$ (left) and $\mu\tau_\mathrm{h}$ (right) pre-selections. Only statistical uncertainties are shown. 69

6.11 Distributions of the OC/SC factor found in the MC QCD for different relative isolation requirements in the $\mathrm{e}\tau_\mathrm{h}$ (left) and $\mu\tau_\mathrm{h}$ (right) pre-selections. Only statistical uncertainties are shown. 70

6.12 Distributions of the muon p_T (top), $p_\mathrm{T}^\mathrm{miss}$ (middle) and m_T (bottom) in the W+jets control region. In the figures on the left the QCD contribution is taken from the simulation, while on the right it is obtained with the data-driven method. 71

6.13 Distributions of muon p_T (top), $p_\mathrm{T}^\mathrm{miss}$ (middle) and m_T (bottom) in the $\mu\tau_\mathrm{h}$ pre-selection. In the figures on the left the QCD contribution is taken from the simulation, while on the right is obtained with the data-driven method. 72

6.14 Distributions of the leading lepton p_T, $p_\mathrm{T}^\mathrm{miss}$, and m_T in the $\mathrm{e}^\pm\mu^\mp$ control region. Only statistical uncertainties are shown. 74

6.15 The dependence of the m_T on the p_T of τ_h candidates. The distributions show the ratios of the m_T distributions in the events passing specified tau p_T thresholds to the m_T distribution of all events in the muon-tau (top) and electron-tau (bottom) main selections. The ratios for the genuine τ_h are on the left, and the misidentified candidates from $\mathrm{t\bar{t}}$ lepton+jets are on the right. Only statistical uncertainties are included. 75

6.16 Comparison of the signal ($\mathrm{t\bar{t}} \to \ell\nu_\ell\tau_\mathrm{h}\nu_\tau\mathrm{b\bar{b}}$) and the main background of misidentified τ_h ($\mathrm{t\bar{t}} \to \ell\nu_\ell\mathrm{q\bar{q}'b\bar{b}}$) in the shapes of the normalized distributions of the transverse mass m_T between the lepton and $p_\mathrm{T}^\mathrm{miss}$ (left), and the $D_\mathrm{jjb}^\mathrm{min}$ parameter of the event categories (right), that is derived from the constraints of the W boson and the top quark masses in the $\mathrm{t\bar{t}} \to \ell\nu_\ell\mathrm{q\bar{q}'b\bar{b}}$ process. In the m_T distribution, the signal may extend beyond the W boson mass endpoint because of the two-neutrino final state, whereas the background process cannot. The last bin in both distributions includes overflow events. In the $D_\mathrm{jjb}^\mathrm{min}$ distribution, the downward arrow points at the threshold of the cut used ($D_\mathrm{jjb}^\mathrm{min} > 60\,\mathrm{GeV}$), and the panel on the right shows the fraction of events in the "signal-like" category where there is only one untagged jet, which amounts to approximately 5% of all background events and 17% of all signal events. 76

6.17 Distributions of the dijet mass (top), the trijet mass (middle), and the threshold parameter D_jjb (bottom) in the $\mu\tau_\mathrm{h}$ pre-selection. The distributions with MC QCD are shown on the left, while the distributions with data-driven QCD are on the right. 78

6.18 Distributions of the dijet mass (top), the trijet mass (middle), and the threshold parameter D_jjb (bottom) in the $\mathrm{e}\tau_\mathrm{h}$ pre-selection. The distributions with MC QCD are shown on the left, while the distributions with data-driven QCD are on the right. 79

6.19 Distributions in the pre-selection of the $\mathrm{e}\tau_\mathrm{h}$ final state. Transverse momentum (left) and pseudo-rapidity (right) of the electron (top) and the leading b-tagged jet (middle), $p_\mathrm{T}^\mathrm{miss}$ and m_T (bottom). Only statistical uncertainties are shown. 82

6.20 Distributions in the pre-selection of the $\mu\tau_\mathrm{h}$ final state. Transverse momentum (left) and pseudo-rapidity (right) of the muon (top) and the leading b-tagged jet (middle), $p_\mathrm{T}^\mathrm{miss}$ and m_T (bottom). Only statistical uncertainties are shown. 83

6.21 Distributions in the final selection of the $\mathrm{e}\tau_\mathrm{h}$ final state. Transverse momentum (left) and pseudo-rapidity (right) of the electron (top) and the leading b-tagged jet (middle), $p_\mathrm{T}^\mathrm{miss}$ and m_T (bottom). Only statistical uncertainties are shown. 84

6.22 Distributions in the final selection of the $\mu\tau_\mathrm{h}$ final state. Transverse momentum (left) and pseudo-rapidity (right) of the muon (top) and the leading b-tagged jet (middle), $p_\mathrm{T}^\mathrm{miss}$ and m_T (bottom). Only statistical uncertainties are shown. 85

6.23 Distributions of the number of reconstructed vertices at the OC tau selection step in the $e\tau_h$ (top) and $\mu\tau_h$ (bottom) final states. The distributions are produced with different PU reweighting: with the nominal weights for a minimum bias cross section of 69.2 (middle), the variations of 5% down (left) and 5% up (right). Only statistical uncertainties are shown. 86

6.24 The transverse mass distributions between lepton (e or μ) and p_T^{miss}, m_T, in the signal-like (upper) and background-like (lower) event categories for the $e\tau_h$ (left) and $\mu\tau_h$ (right) final states observed prior to fitting. Distributions obtained from data (filled circles) are compared with simulation (shaded histograms). The last bin includes overflow events. The simulated contributions are normalized to the cross section values predicted in the SM. The main processes are shown: the signal, the other $t\bar{t}$ processes grouped together, single top quark production, W+jets, DY processes, diboson, and multijet production. The ratio of the data to the total SM prediction is shown in the lower panel. The vertical bars on the data points indicate the statistical uncertainties, the hatched band indicates the systematic uncertainties and the statistical uncertainties in all simulated samples. 87

6.25 The corrections to the nominal jet energy scale for the parton shower systematics: in the barrel on the left, in endcaps on the right, for b jets (top), c jets (middle) and light flavour jets (bottom). 91

6.26 Generated versus fitted signal strength in toy experiments for the $e\tau_h$ (left) and $\mu\tau_h$ (right) final states. The median of toys is represented by blue dots, the bands represent 68% and 95% quantiles. The black line shows the fully correlated case. 95

6.27 Scan of expected profile likelihood ratio in the $e\tau_h$ (top-left), $\mu\tau_h$ (top-right), and both final states combined (bottom). The simultaneous fit is performed by keeping nuisance parameters and signal strength equal across all event categories and final states. The scan with full uncertainty is shown with bold solid line. The impact of the τ_h identification and misidentification uncertainties is shown by the thin solid line. It is produced by setting the τ_h-related nuisance parameters to their post-fit values. A scan of the statistical uncertainty, shown with a dashed line, is produced by setting all nuisance parameters to their post-fit values. 96

6.28 Expected impacts of nuisance parameters on the signal strength in the eτ_h (top-left), $\mu\tau_\mathrm{h}$ (top-right), and both final states combined (bottom). The simultaneous fit is performed by keeping nuisance parameters and signal strength equal across all event categories and final states. 97

6.29 Scans of the expected and observed profile likelihood ratios in the eτ_h (top-left), $\mu\tau_\mathrm{h}$ (top-right), and both final states combined (bottom). 99

6.30 Post-fit values of main nuisance parameters in the eτ_h (top) and $\mu\tau_\mathrm{h}$ (bottom) final states. 100

6.31 Impacts of the nuisance parameters on the signal strength fitted to the observed data in the eτ_h (top-left), $\mu\tau_\mathrm{h}$ (top-right), and both final states combined (bottom). 101

6.32 Covariances of the post-fit values of main nuisance parameters in the eτ_h (top), $\mu\tau_\mathrm{h}$ (middle) and combined $\ell\tau_\mathrm{h}$ (bottom) final states. The omitted parameters correspond to the uncertainties due to the MC statistics, PDF and scale variations make a small contribution to the overall uncertainty, and do not exhibit significant correlations with other parameters. Expected correlation groups are seen in the parameters: the lepton ID and trigger efficiency; the signal strength, τ_h ID efficiency and τ_h misidentification probability. 103

6.33 Post-fit distributions of the m_T with the nuisance parameters set to the values found in the fit to both channels: the background-rich (left) and signal-rich event categories (right), in the eτ_h (top) and $\mu\tau_\mathrm{h}$ (bottom) final states. 104

6.34 The expected and observed dependence of the likelihood on the total $\mathrm{t\bar{t}}$ cross section $\sigma_{\mathrm{t\bar{t}}}$. It is derived from the fiducial phase space by a simple extrapolation. The arrow points at the cross section measured in the light dilepton final state. The goodness of the fit determined with a Kolmogorov–Smirnov method yields a p value of 0.24. 106

7.1 Feynman diagrams of the $\mathrm{t\bar{t}}$ decays into dilepton final states: including a τ lepton (left), and with only light leptons (right). 110

7.2 Distributions of the length of flight (top left) and the significance of the length of flight (top right) for the genuine τ_h decays taken from the SV data in the $\mathrm{t\bar{t}}$ MC datasets. The correlation plot between the length of flight and the energy for the genuine τ_h candidates (bottom). 114

7.3 Comparison of the distributions for the τ_{h} SV saved in the datasets (left) and reconstructed with the geometric algorithm that is developed and discussed in the text (right) for genuine τ_{h} candidates: the length of flight (top), the length of flight significance (middle), and the correlation between the length of flight and energy (bottom). . . 116

7.4 Comparison of the SV flight length significance available in the simulation (left) and reconstructed with the geometric algorithm that is developed and discussed in the text (right) in the DY final states $\tau_{\ell}\tau_{\mathrm{h}}$ and $\mu\mu$ (top), and the $\mathrm{t\bar{t}}$ final states $\mathrm{t\bar{t}} \rightarrow \mathrm{e}\nu_{\mathrm{e}}\tau_{\mathrm{h}}\nu_{\tau}\mathrm{b\bar{b}}$ and $\mathrm{t\bar{t}} \rightarrow \mathrm{e}\nu_{\mathrm{e}}\mathrm{q\bar{q}'b\bar{b}}$ (bottom). The misidentified τ_{h} candidates in $\mathrm{t\bar{t}}$ are separated according to their physical origin: the hard process b jet, the c flavour jets from the W boson decay, other activity in the event. 117

7.5 Distributions of the leading lepton transverse momentum in the light dilepton $\mathrm{t\bar{t}}$ (left) and DY (right) final states. Only statistical uncertainties are shown. 119

7.6 Comparison of the signal and background distributions in the $\mathrm{t\bar{t}} \rightarrow \mathrm{e}\tau_{\mathrm{h}}$ and DY $\rightarrow \tau_{\mu}\tau_{\mathrm{h}}$ event selections: the m_{T} distributions in the $\mathrm{t\bar{t}} \rightarrow \mathrm{e}\nu_{\mathrm{e}}\tau_{\mathrm{h}}\nu_{\tau}\mathrm{b\bar{b}}$ and $\mathrm{t\bar{t}} \rightarrow \mathrm{e}\nu_{\mathrm{e}}\mathrm{q\bar{q}'b\bar{b}}$ processes on the left, the p_{T} of the μ in the DY $\rightarrow \tau_{\mu}\tau_{\mathrm{h}}$ and W+jets processes on the right. 120

7.7 Scan of the expected profile likelihood ratio in the fit of the ratio between dilepton final states with light leptons and with a τ_{h} in $\mathrm{t\bar{t}}$. The statistical uncertainty is shown by fixing the nuisance parameters that encode systematic uncertainties to their post-fit values. 123

7.8 Expected impacts of the nuisance parameters on the signal strength in the fit of the ratio between dilepton final states with light leptons and with a τ_{h} in $\mathrm{t\bar{t}}$. 124

7.9 Scan of the expected profile likelihood ratio in the double ratio fit. The fit is performed in simulated toy distributions, with the expected event yields normalized to a luminosity of $35.9\,\mathrm{fb}^{-1}$. The statistical uncertainty is estimated by fixing the nuisance parameters that encode systematic uncertainties to their post-fit values. The cancellation of the systematic uncertainties is limited by the statistical fluctuations. 126

7.10 Expected impacts of nuisance parameters on the signal strength in the double ratio fit. The fit is performed in simulated toy distributions, with the expected event yields normalized to a luminosity of $35.9\,\mathrm{fb}^{-1}$. 127

7.11 Scan of the expected profile likelihood ratio in the double ratio fit. The fit is performed in simulated toy distributions, with the expected event yields normalized to a luminosity of 145 fb^{-1}. The statistical uncertainty is shown by fixing the nuisance parameters that encode systematic uncertainties to their post-fit values. The cancellation of the systematic uncertainties is limited by the statistical fluctuations. 128
7.12 Expected impacts of nuisance parameters on the signal strength in the double ratio fit. The fit is performed in simulated toy distributions, with the expected event yields normalized to a luminosity of 145 fb^{-1}. 129
7.13 Scan of the expected profile likelihood ratio in the double ratio fit. The fit is performed in simulated toy distributions, with the expected event yields normalized to a luminosity of 1000 fb^{-1}. The statistical uncertainty is shown by fixing the nuisance parameters that encode systematic uncertainties to their post-fit values. The small statistical fluctuations allow the systematic uncertainties to cancel out in the double ratio. 130
7.14 Summary of the expected relative uncertainties in double ratio measurements at different luminosities: 35.9 fb^{-1}, 145 fb^{-1}, and 1000 fb^{-1}. The overall uncertainty is shown by the thick red line, the systematic component is the thin black line, the statistical fluctuations are shown by the dashed black line. 131

Chapter 1

Introduction

The high energy and luminosity of the proton-proton (pp) collisions at the Large Hadron Collider (LHC) provide for the first time access to a large number of events with top quarks. The top quark mass of approximately 172.5 GeV makes it the heaviest particle in the standard model (SM) and the only quark that decays before it could hadronize. The interactions with the top quarks therefore give access to the "bare-quark" physics, and allow to study the high-energy processes in the SM and the hypothetical beyond SM (BSM) contributions.

The top quark (t) decays into a bottom quark (b) and a W boson. The decay modes of the W boson determine the final state signature of the decay. In pp collisions top quarks are produced mainly in pairs ($\mathrm{t\bar{t}}$): $\mathrm{pp} \to \mathrm{t\bar{t}} \to \mathrm{W^+bW^-\bar{b}}$. The dilepton decay process denotes the case where each W boson decays into a lepton and a corresponding neutrino. The studies presented in this thesis focus on the $\mathrm{t\bar{t}} \to (\ell\nu_\ell)(\tau\nu_\tau)\mathrm{b\bar{b}}$ process, depicted in Figure 1.1, where one W boson decays into $\ell\nu_\ell$, ℓ is either an electron (e) or a muon (μ), and the other into a τ lepton and a neutrino ($\tau\nu_\tau$). The τ leptons are detected by their decay products. The studies in this thesis consider only the τ decays into hadrons, τ_h, because their reconstruction does not rely on the signal model. Furthermore, the leptonic τ decays could be distinguished from the hard process leptons only by the energy and impact parameters, which depend on the τ production model. Wherever necessary, the results are extrapolated to all τ leptons using the branching fractions of the τ decay that have been measured to a high precision [4].

Recent checks of lepton flavour universality violation [5, 6, 7, 8, 9, 10, 11, 12] sparked a renewed interest towards measurements involving τ leptons, owing to a potential disagreement with standard model (SM) predictions. The $\mathrm{t} \to (\tau\nu_\tau)\mathrm{b}$ decay exclusively involves third-generation leptons and quarks which, owing to their

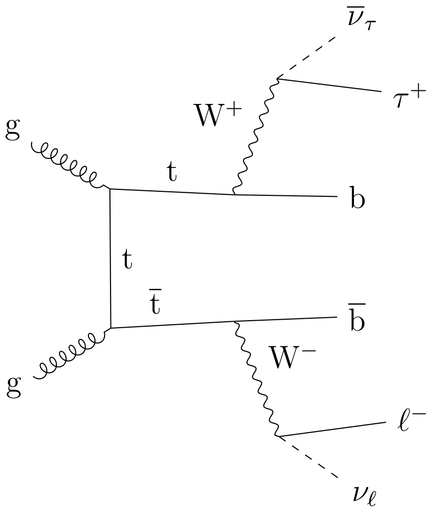

Figure 1.1: Feynman diagram of the $\mathrm{t\bar{t}}$ production with the SM decay into the dilepton final state that includes a τ lepton.

large masses, may be particularly sensitive to BSM contributions. For example, the existence of a charged Higgs boson [13, 14, 15, 16] may give rise to anomalous τ lepton production that could be observed in this decay channel. Also, the SM $\mathrm{t} \to (\tau\nu_\tau)\mathrm{b}$ process directly involves the $\mathrm{W} \to \tau\nu_\tau$ decay, that shows a 2.5σ deviation from the lepton universality principle in the best up to date measurements of the $\mathcal{B}(\mathrm{W} \to \tau\nu_\tau)$ branching fraction [4]. The $\mathrm{t\bar{t}}$ final states provide an opportunity to contribute to these results. Recently a result in the $\mathrm{t\bar{t}}$ channel with leptonic τ decays was presented by ATLAS [17]. It shows a good agreement with the SM to a precision of 1.3%. The measurements with leptonic τ decays rely on assumptions about the signal model to distinguish the prompt leptons from the τ decays. A measurement that is based on hadronic τ_h decays is independent from the kinematic parameters of the signal model and would provide an important contribution to the topic.

This thesis presents the first measurement of the $\mathrm{t\bar{t}}$ production cross section in proton-proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ that explicitly includes τ leptons, $\sigma_{\mathrm{t\bar{t}}}(\ell\tau)$, and also improves the relative precision with respect to the 7 and 8 TeV results [18, 19]. The cross section measurement is carried out in the data sample that was collected in 2016 with the CMS detector at the LHC and corresponds to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$. The cross section is measured by performing a profile likelihood ratio (PLR) fit [20] to the transverse mass of the system containing the light lepton (e or μ) and the missing transverse momentum, in two kinematic categories of the selected events for each of the $\mathrm{e}\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ final states. The cross section is measured in the fiducial phase space of the detector and also extrapolated to the full phase space. The ratio of the cross sections in the $\ell\tau$ and light dilepton [2] final states $\sigma_{\mathrm{t\bar{t}}}(\ell\tau)/\sigma_{\mathrm{t\bar{t}}}(\ell\ell)$, and the ratio of the partial to the total decay width of the top quark $\Gamma(\mathrm{t} \to \tau\nu_\tau \mathrm{b})/\Gamma_{\mathrm{total}}$ are evaluated, improving the precision over previous measurements [21, 22].

Using the data analysis framework developed for the production cross section measurement, a feasibility study of a lepton universality test in $\mathrm{t\bar{t}}$ final states is performed for the full Run2 dataset collected with the CMS detector: a precise measurement of the ratio $\mathcal{B}(\mathrm{W} \to \tau\nu_\tau)/\mathcal{B}(\mathrm{W} \to \ell\nu_\ell)$ that can make a significant contribution to the current world-best results. The methods employed in the lepton universality test in $\mathrm{t\bar{t}}$ events have been developed in the Monte-Carlo (MC) simulation of the events collected in 2016 and 2017 with the CMS detector. The simulation is compared to the collected data in independent control regions. The feasibility of the measurement is studied in the simulated events for the full Run2 dataset that corresponds to a luminosity of $145\,\mathrm{fb}^{-1}$.

I have been the main author of these measurements of the cross section and of the feasibility study in the CMS collaboration. The publication of the cross section measurement passed the internal review in the collaboration, and I have been the contact person at all stages of the review. It included development of all analysis steps, reports, and subsequent presentations of the analysis for the pre-approval and the approval by the collaboration, the preparation of the answers to the questions raised by the collaboration members during the review, and implementation of the suggestions. I have also prepared the answers to the questions from the JHEP referee for the analysis paper, that were reviewed and approved by the analysis review committee within the collaboration. With the help from the CMS collaboration and the LIP group the analysis has been carried to a large extend by my effort.

The thesis is organized as follows: the relevant information about the SM and the physics of the $\mathrm{t\bar{t}}$ channel is presented in Chapter 2; the CMS detector at LHC is briefly described in Chapter 3; details about the event reconstruction and selection, and the simulated event samples used in the data analysis are provided in Chapter 4; the main background estimation methods are discussed in Chapter 5; Chapter 6 presents the measurements of the $\mathrm{t\bar{t}}$ production cross sections in the full phase space and in the fiducial region of the detector, the ratio of the cross sections in the $\ell\tau_{\mathrm{h}}$ and $\ell\ell$ dilepton $\mathrm{t\bar{t}}$ final states, and the ratio of the partial width $\Gamma(\mathrm{t} \to \tau\nu_\tau \mathrm{b})$ to the total decay width of the top quark; the feasibility study of the lepton universality test in $\mathrm{t\bar{t}}$ final states is carried out in Chapter 7; and Chapter 8 summarizes the results.

Chapter 2

Experimental high energy physics

2.1 The standard model of particle physics

The study of the world around us leads to an image of elementary objects that are bound by universal interaction forces. Stronger forces neutralize themselves by bringing the objects of their action into composite stable states. Atoms are such states, that consist of electrons and nuclei bound by the electromagnetic force. The nuclei themselves consist of nucleons, protons and neutrons, that are bound together by an expression of the strong force. The interactions between such composite states produce the phenomena observed in normal conditions. Collisions of composite objects at energies greater than the binding force energy reveal their internal structure. Analysis of the exotic states produced in such collisions leads to the discovery of the elementary content and fundamental structure of nature.

With few exceptions, all current observations fit into the standard model (SM) of elementary particles to a high precision [4]. Under the SM all matter particles have spin 1/2 and fall into two categories according the forces that affect them: leptons interact only electroweakly, quarks interact electroweakly and strongly. They further divide into two types according to the electric charge: leptons with the charges -1e and 0e, and quarks with the charges $+2/3$e and $-1/2$e. Corresponding anti-particles have charges of opposite signs. Finally, there are three flavour families of the matter particles, that differ only by the masses of their members. Other quantum numbers follow this description and can be found in Ref. [4].

The charged leptons in the three flavour families are: electron with the mass of about 0.5 MeV, muon with the mass of 105 MeV, and tau with the mass of 1.7 GeV. The corresponding neutrinos have very small masses that are not measured yet. The quark flavour families consist of: the up ($+2/3$e) and down ($-1/3$e) quarks with masses of about 2.2 MeV and 4.7 MeV respectively, the charm (+2/3e) and strange (−1/3e) quarks with masses of about 1.3 GeV and 95 MeV, and the top (+2/3e) and bottom (−1/3e) quarks with masses of about 173 GeV and 4 GeV. The quarks can form stable states bound by the strong force that contributes a large fraction of the stable state mass. This is the case for protons and other long-lived particles.

The interactions between the matter particles are carried by vector bosons that follow gauge symmetries: the four bosons of the electroweak force follow the $SU(2) \times$ U(1) group, and the eight gluon bosons g of the strong force follow $SU(3)$ group. The SM includes a complex scalar Higgs doublet field with an unstable potential that spontaneously develops a vacuum expectation value of about 246 GeV. Interactions with the Higgs field break a part of the electroweak symmetry at the energies bellow the vacuum expectation value. After the breaking, three out of four electroweak bosons dynamically acquire masses, corresponding to the $\mathrm{W}^{\pm}$ and Z with the masses of about 80 and 91 GeV respectively, and only one neutral scalar particle H remains from the complex Higgs doublet with the mass of 125 GeV. The massive bosons W and Z propagate the short-range weak interactions. The fourth boson of the electroweak field stays massless and corresponds to the photon, γ, that mediates the electromagnetic interactions.

The dynamic description of the interactions between particles is given in the quantum field theory, where excitations of fermionic fields correspond to the particles of matter, and bosonic gauge fields mediate the interactions between them.

The top quarks (t) and tau leptons (τ) from the third flavour family have the largest masses that determine their special role in the SM and in the models of the physics beyond the SM.

2.2 The top quark

In the SM, the left-handed top quark has a charge of $+2/3e$ and is the $T_3 = +1/2$ member of the weak-isospin doublet containing the bottom quark, while the right-handed top is an $SU(2)_L$ singlet. With the mass of about 172.5 GeV, the top quark is the heaviest particle in the SM. It is the only quark that decays before hadronization. Practically all top quarks decay into a bottom quark and a W boson, $\mathrm{t} \to \mathrm{bW}^{+}$, that corresponds to a negligible flavour mixing term in the CKM matrix, $|V_{\mathrm{tb}}| \gg |V_{\mathrm{ts}}|, |V_{\mathrm{td}}|$. The top quark phenomenology provides a unique opportunity to test the theory of the strong interactions, and obtain the information about fundamental interactions at the scale of electroweak symmetry-breaking and beyond.

In pp collisions most top quarks are produced in pairs ($\mathrm{t\bar{t}}$) through the processes of quark annihilation $\mathrm{q\bar{q}} \to \mathrm{t\bar{t}}$ and gluon fusion $\mathrm{gg} \to \mathrm{t\bar{t}}$. About 85% of $\mathrm{t\bar{t}}$ events were produced through the $\mathrm{q\overline{q}}$ process in the 1.96 TeV $\mathrm{p\overline{p}}$ collisions at the Tevatron. At LHC energies the gluon fusion process becomes dominant, contributing about 80% of $\mathrm{t\overline{t}}$ events at 7 TeV and 90% at 13 TeV. In the subsequent decay of the top quarks into bottom quarks and W bosons, $\mathrm{pp} \rightarrow \mathrm{t\overline{t}} \rightarrow \mathrm{W^+bW^-\overline{b}}$, the decays of the W bosons define the signature of the process. The final states produced by the SM $\mathrm{t\overline{t}}$ decays fall into one of three classes listed in Table 2.1.

Table 2.1: The three classes of the top quark pair decay final states.

Name	Process	Branching fraction [%]
Dileptonic	$\mathrm{t\overline{t}} \rightarrow \ell^+\nu_\ell \mathrm{b}\ell'^-\overline{\nu}_{\ell'}\overline{\mathrm{b}}$	10.5
Semileptonic	$\mathrm{t\overline{t}} \rightarrow \mathrm{q\overline{q}'b}\ell^-\overline{\nu}_\ell\overline{\mathrm{b}} + \ell^+\nu_\ell\mathrm{bq''\overline{q}'''\overline{b}}$	43.8
Fully hadronic	$\mathrm{t\overline{t}} \rightarrow \mathrm{q\overline{q}'bq''\overline{q}'''\overline{b}}$	45.7

The large mass of the top quark defines its phenomenology in the SM and the hypothetical BSM interactions. As the only quark that does not hadronize, it gives access to the “bare quark” physics and allows to set stringent limits on the predictions in quantum chromodynamics [4]. The top quark probes the properties of the Higgs boson sector in the SM, the CKM quark mixing, and the electroweak interactions at high energy scale. It plays an important role in searches for BSM processes, as a part of the signal process, or as the main background in the search for such contributions.

Since the discovery of the top quark [23, 24], the top quark pair and the single top quark processes have been studied in $\mathrm{p\overline{p}}$, pp, pPb, and $PbPb$ collisions at the Tevatron and the LHC accelerators. The Tevatron produced $\mathrm{p\overline{p}}$ collisions at the $\sqrt{s} = 1.8$ and 1.96 TeV energies. The LHC performed pp collisions at the $\sqrt{s} = 7$ TeV and 8 TeV during the Run1 stage of its operation, starting in 2010. The Run2 of the LHC operation reached the $\sqrt{s} = 13$ TeV centre-of-mass energy of pp collisions, producing data samples collected by the general purpose detectors at the LHC that correspond to luminosities of about 150 fb^{-1}. Also, there was a dedicated low energy run of pPb collisions at 5.02 TeV and 8.16 TeV. The measurements of the top quark properties and production cross section in these data show a good agreement with the SM expectations.

The most precise measurements of the $\mathrm{t\overline{t}}$ production cross section are performed in the light dilepton final states, in particular the $\mathrm{e}\mu$ final state. The combination of the CMS and ATLAS results in $\mathrm{e}\mu$ channel at 8 TeV produces a measurement with relative uncertainty of 3.5% challenging theoretical uncertainties [25]. Precise measurements in the dilepton final states at 13 TeV yield results with uncertainties of about 3% [26, 2, 27, 28], consistent with the expected $\sigma_{\mathrm{t\overline{t}}} = 832^{+20}_{-29}\,(\mathrm{scale}) \pm 35\,(\mathrm{PDF}+\alpha_S)$ pb at NLO accuracy [29]. The lepton+jets final states have enough kinematic constraints to reconstruct the decay completely and deliver the most precise measurements of the top quark mass. However, advanced methods in the underconstrained dilepton final states deliver precise results too. The results at 13 TeV show a good agreement with the expected value of $m_{\mathrm{t}} = 172.5\,\mathrm{GeV}$ [30, 2, 28]. The large integrated luminosities allowed to carry out a wide range of differential cross section measurements at the Tevatron and LHC data samples. Such measurements test perturbative QCD predictions and reduce the uncertainty in the description of the $\mathrm{t\bar{t}}$ system, improving the background estimation in Higgs physics and searches for rare processes and BSM contributions. The $\sigma_{\mathrm{t\bar{t}}}$ has been measured as a function of the properties of leptons, jets, and $\mathrm{t\bar{t}}$ system at 13 TeV [31, 32, 33, 34, 35, 36]. The CMS group at Laboratório de Instrumentação e Física Experimental de Partículas (LIP) performed measurements of the $\mathrm{t\bar{t}}$ production cross section and the top quark mass in the dilepton final state with a τ lepton in the Run1 data of pp collisions at 7 and 8 TeV [18, 19, 37]. This final state has been used to set a limit on a hypothetical charged Higgs boson contribution $\mathrm{t} \rightarrow \mathrm{H}^{+}\mathrm{b} \rightarrow \tau\nu\,\mathrm{b}$ [38]. The work of the group has been the foundation for my analysis of the 13 TeV data.

The $\mathrm{t} \rightarrow \tau\nu_{\tau}\mathrm{b}$ decay exclusively involves third-generation leptons and quarks. It can be particularly sensitive to certain BSM processes. For example, the models with a charged Higgs boson [13, 14, 15, 16] may give rise to anomalous τ lepton production that could be observed in this decay channel.

In the SM, this final state is produced through the W boson decay into a τ lepton. It gives a direct access to the $\mathcal{B}(\mathrm{W} \rightarrow \tau\nu_{\tau})$ branching fraction. The production of the SM $\mathrm{t\bar{t}}$ pairs serves as a source of a large number of W boson decays with additional features in the final state that can help to select a pure sample of events. Therefore, it can be used to make a precise measurement of the W boson branching fractions and deliver a test of the lepton flavour universality in W boson decays. These features are investigated in this work.

2.3 The tau lepton

The mass of the τ lepton is about 1.78 GeV, which makes it the only lepton that can decay into hadrons and other leptons. The final states with the largest branching fractions of the τ decay are given in Table 2.2. All final states include a tau neutrino ν_{τ} that caries the tauonic number of the original state. The light lepton final states have an additional neutrino, ν_{e} or ν_{μ}, that corresponds to the produced light lepton, e or μ. The hadronic final states with the largest branching fractions consist of one or three charged hadrons, denoted by $h^{\pm}$, and up to two neutral hadrons.

The τ lepton has a short lifetime of about 0.3 ps. With energies of about 30 GeV, typical to the top quark and W boson decays, the boosted τ leptons propagate for about 2 mm from the point of production before decaying. The distance of the τ lepton flight promotes itself in the impact parameters of the final state particles. In the case of the decay into three charged hadrons (a 3-prong τ_h), the vertex of the τ lepton decay can be reconstructed. This final state is produced through a sequential two-body decay $\tau^- \to \nu_\tau a_1^- \to \nu_\tau \pi^- \rho^0 \to \nu_\tau \pi^- \pi^- \pi^+$, that can be distinguished in a Dalitz plot of the kinematic parameters of the final state charged hadrons. The main background for the identification of 3-prong τ_h comes from heavy mesons because their lifetimes are similar to the τ lepton. A comparison of the lifetime values is given in Table 2.3.

Table 2.2: The decays of the tau lepton and their branching fractions. The symbols h^- and h^0 signify charged and neutral hadrons.

Decay mode	Branching fraction (%)
$\tau^- \to \nu_\tau \overline{\nu}_\mathrm{e} \mathrm{e}^-$	17.4
$\tau^- \to \nu_\tau \overline{\nu}_\mu \mu^-$	17.8
Total leptonic modes	35.2
$\tau^- \to \nu_\tau h^-$	11.6
$\tau^- \to \nu_\tau h^- h^0$	26.0
$\tau^- \to \nu_\tau h^- h^0 h^0$	10.8
$\tau^- \to \nu_\tau h^- h^+ h^-$	9.8
$\tau^- \to \nu_\tau h^- h^+ h^- h^0$	4.8
Other hadronic modes	1.7
Total hadronic modes	64.8

Table 2.3: The lifetime of the τ lepton compared to D and B mesons.

	τ	D^0	D^+	D_s^+	B^0	B^+	B_c^+
Lifetime [ps]	0.3	0.4	1.0	0.5	1.5	1.6	0.5

The leptonic final states of the τ decay can be distinguished from the prompt light leptons that originate in the hard process by a smaller transverse momentum of the visible decay products and displaced secondary vertices of their charged tracks. This difference relies on the assumed production models of the τ and the prompt light leptons. On top of that, leptons with low energy and a sizeable impact parameter have larger misidentified backgrounds. Together with small branching fractions, these shortcomings make the analysis with leptonic τ decays impractical, and not suitable when loose assumptions on the signal model are required. However, these features can be exploited in an analysis, but to disentangle the τ from the prompt light leptons in this way becomes more complicated and biased to the signal model.

The jet produced by the hadronic τ decay, τ_{h}, has a distinct signature with a small multiplicity of charged hadrons and neutral pions that decay into photons. This gives the τ_{h} the distinct signature of a jet with a smaller cone when compared to quark- and gluon-originated jets. It makes τ_{h} the main channel to identify τ leptons.

Measurements involving τ_{h} face large backgrounds of jets misidentified as τ that come from hadronic processes with large cross sections, such as multijet QCD, W+jets, and hadronic final states in $\mathrm{t\bar{t}}$. These processes produce jets in a wide spectrum of kinematic ranges and other characteristics that covers the τ_{h} signature. The probability to misidentify a jet as a τ_{h} can be up to 1%, and it varies by as much as an order of magnitude, owing to the diverse structure of different kinds of jets [3]. Even the specific properties of the τ decays with three charged hadrons in the final state can be mimicked by the jets originating from D and B mesons that are produced in the hadronic $\mathrm{t\bar{t}}$ and W+jets processes.

The described characteristics present inherent challenges for the identification of the τ lepton in pp collisions. Due to the complex nature of jets, there is no common estimation of the τ_{h} misidentification probability that would work in all final states [3]. The estimation of the misidentified τ_{h} background must be suited to the particular composition of jets in the events of interest. A detailed description of the studies performed on the τ_{h} identification and misidentification is discussed in Chapter 5.

Chapter 3

Experimental apparatus

3.1 The Large Hadron Collider (LHC)

The LHC at CERN is the largest hadronic accelerator in the world. It is designed to collide beams of protons or heavy ions. Therefore it can produce three types of collisions: proton-proton, proton-ion, and ion-ion. This thesis analyses the data from the proton-proton collisions.

The LHC is located at CERN near Geneva. It is installed in a tunnel of 26.6 km in circumference, approximately 100 m below ground level, that was initially built in the CERN's accelerator complex for the Large Electron-Positron accelerator (LEP) [39]. A detailed description of the LHC is presented in Refs. [40, 41]. The geographic location and the scheme on the accelerator with the locations of main detectors are shown in Figure 3.1. There are four main beam crossing points hosting four main detector experiments, and a number of other access points to the accelerator tunnel. The Compact Muon Solenoid (CMS) detector is installed in an underground cavern at the point 5 on the accelerator ring.

The protons for the collisions are obtained from ionized hydrogen atoms. Then they go through a cascade of accelerators gaining higher kinematic energy on each stage: Linac2 [42] brings the energy up to 50 MeV, Proton Synchrotron (PS) Booster [43] — 1.4 GeV, PS [44] itself — 25 GeV, Super Proton Synchrotron [45] (SPS) — 450 GeV, and then they are injected into the LHC. Once in the LHC, the proton beams are accelerated to the operating energy and are focused to collide in the interaction points of the accelerator ring. The centre-of-mass energy of the proton-proton collisions produced by the LHC reaches 13 TeV, which is the highest energy of particles collisions in a laboratory environment. The peak luminosity is about $2 \times 10^{34}\,\mathrm{cm}^{-1}\,\mathrm{s}^{-1}$, which resulted in about $150\,\mathrm{fb}^{-1}$ datasets of 13 TeV proton-proton collisions collected

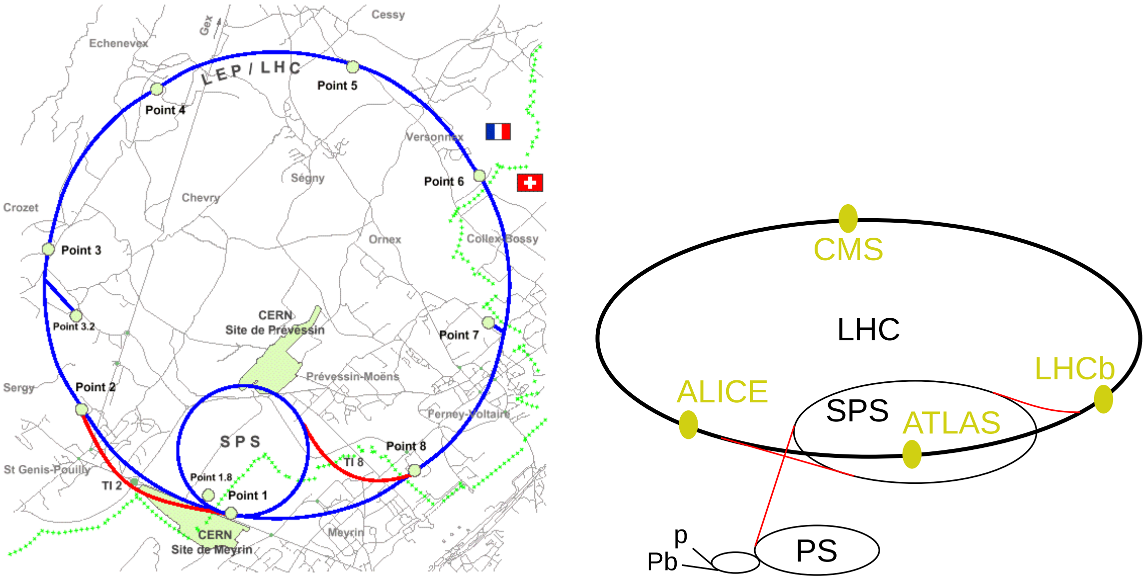

Figure 3.1: The map of the CERN LHC site at Geneva (left), and the scheme of the accelerator with the four main detectors and interaction points (right).

by the detectors at the LHC. The large instantaneous luminosity corresponds to a large background of pileup events, that is about 20–30 collisions per bunch crossing.

The LHC machine has a capacity for further performance upgrade at the High Luminosity stage (HL-LHC) scheduled to start in 2025 [46]. The program of the upgrade includes 14 TeV of centre-of-mass energy of the proton-proton collisions, a factor of five increase in the instantaneous luminosity, and an order of magnitude increase in the integrated luminosity, that corresponds to datasets of about 3000 fb^{-1}. The increased instantaneous luminosity implies a larger background of pileup events: 140–200 events per bunch crossing, which is about an order of magnitude larger than the current level.

3.2 The Compact Muon Solenoid (CMS) detector

The central feature of the CMS apparatus is a superconducting solenoid, 13 m in length and 6 m in diameter, which provides an axial magnetic field of 3.8 T. Inside the solenoid, various particle detection systems are employed. The detection systems are arranged in two main zones: barrel and endcaps, with $|\eta| < 1.4$ and $1.4 < |\eta| < 2.5$, respectively. Where the pseudorapidity η is defined as $\eta = -\ln[\tan(\theta/2)]$,

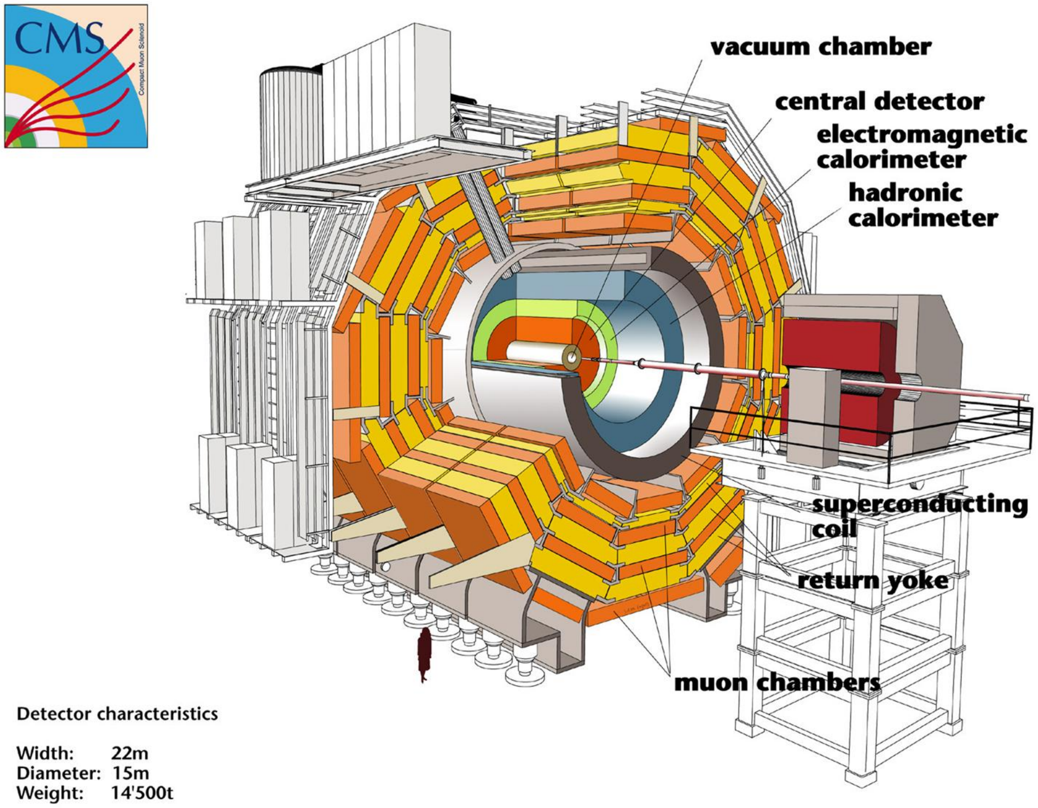

Figure 3.2: A schematic representation of the CMS detector and main subdetector systems.

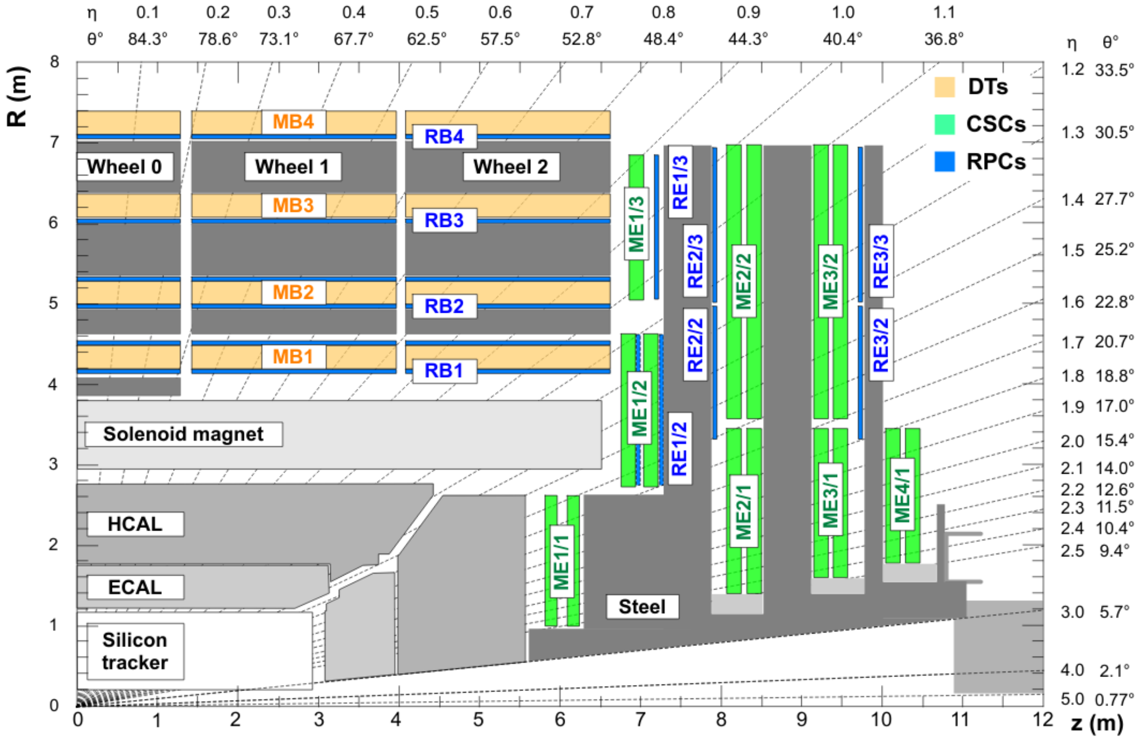

Figure 3.3: A logitudinal representation of the CMS detector showing the position of the subdetectors in the η plane with the muon detectors highlighted.

with θ being the polar angle of the trajectory of the particle with respect to the counterclockwise beam direction. The detector is nearly hermetic, allowing for energy balance measurements in the plane transverse to the beam directions. A schematic drawing of the detector is shown in Figure 3.2. Figure 3.3 shows the layout of the detector components in the η plane. The following paragraphs give a brief overview of the subdetectors with the focus on the features relevant to the reconstruction of tau leptons. A more detailed description of the CMS detector, and the definition of the coordinate system can be found in Ref. [47].

Charged particle trajectories are measured by the silicon pixel and strip tracker, that are the closest to the beamline and cover $0 < \varphi < 2\pi$ in azimuth and $|\eta| < 2.5$. Tracks of charged hadrons are reconstructed with an efficiency typically 80-90%, depending on the transverse momentum p_{T} and η of the hadron. The resolution of the reconstruction of the transverse and longitudinal impact parameters of single isolated charged pions with 1 GeV energy is on the order of 0.1–1 mm. Detailed descriptions of the tracker and the performance studies are given in Ref. [48, 49]

The tracker comprises a significant amount of material in front of the electromagnetic calorimeter (ECAL) and other detectors, mostly the mechanical structure, the services, and the cooling system. The minimum value of the radiation length is $0.4\,X_0$ at $|\eta| \approx 0$, rises to about $2.0\,X_0$ at $|\eta| \approx 1.4$, and decreases to about $1.3\,X_0$ at $|\eta| \approx 2.5$. It leads to a nearly 40% probability for the $\gamma \to \mathrm{e}^{-}\mathrm{e}^{+}$ conversion of the photons that originate from $\pi^0 \to \gamma\gamma$ decays. This feature promotes itself in the τ lepton reconstruction, since a number of τ decay modes with significant branching fractions include the π^0 in the final state.

A lead tungsten crystal electromagnetic calorimeter consists of a barrel and two endcap sections that cover $|\eta| < 1.479$ and $1.479 < |\eta| < 3.0$, respectively. With a small radiation length and a small Moliere radius 2.3 cm, the lead tungsten crystals provide a compact calorimeter with excellent two-shower separation. The ECAL thickness is more than 25 radiation lengths. The studies in the decays of Z boson and 125 GeV Higgs boson show that the raw energy resolution of the calorimeter is 2–5% for electrons, and 1.1–2.6% in the barrel and 2.2–5% in the endcaps for photons [50].

A brass/scintillator sampling hadron calorimeter (HCAL) surround the tracking volume and the ECAL, covering $|\eta| < 3.0$. The scintillation light is converted by the wavelength-shifting fibres and channeled to photodetectors via clear fibres. Taking the average length the particle travels through the matter before disintegration in a nuclear interaction as the interaction length, the thickness of the HCAL varies in the range 7–11 interaction lengths depending on η. In the barrel the relative energy resolution for charged pions is about 20% at 20 GeV energy and 10% at 100 GeV [51].

The calorimetry provides high-resolution energy and direction measurements of electrons and hadronic jets. After jet energy calibration, the final uncertainty in the jet energy scale is below 3% across the phase space considered in the studies of this thesis ($p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 2.5$) [52].

Muon detection systems are located outside of the solenoid and embedded in the steel return yoke. They consist of three types: drift tube stations (DTs) at $|\eta| < 1.2$, cathode strip chambers (CSCs) at $0.9 < |\eta| < 2.4$, resistive plate chambers (RPCs) at $|\eta| < 1.9$. The average hit reconstruction efficiency is about 94–96% for RPCs, and 97% for DTs and CSCs. The information from the muon systems is matched with the tracks reconstructed in the tracker for a robust identification of muon candidates. The magnitudes of the momentum scale corrections are about 0.2% and 0.3% in the barrel and endcap, respectively. The resolution for muons with momenta up to approximately 100 GeV is 1% in the barrel and 3% in the endcap [53].

A two-level trigger system selects the most interesting proton-proton collision events for use in physics analyses [54, 55]. The level 1 (L1) trigger is designed to handle a GHz collision rate and reduce the front end readout bandwidth to 100 kHz of 1 MB events. It is implemented in custom hardware that selects detector signals consistent with streamlined definitions of physical objects: electrons/photons, muons, jets, etc. The 100 GB/s output of the L1 trigger is further reduced to 1000 Hz rate for offline storage in the high level trigger (HLT). The HLT implements more refined reconstruction algorithms in software that runs on a farm of off-the-shelf computers. The HLT decision whether to retain an event is final in the data acquisition process. The gathered data are processed and analysed offline.

In order to cope with the high luminosity in the HL-LHC phase, both with respect to the data acquisition speed and the radiation hardness, the CMS detector will undergo a profound upgrade of all of its component subdetectors. It will include the upgrades of the pixel tracker [56], the inner and outer tracker [57, 58, 59], the ECAL [60], the new high granularity HCAL [61], the upgrade of the outer muon detectors [62], and the installation of the new MIP Timing Detector (MTD) [63] that will provide a precise timing measurement for the charged particles that cross it. The timing will improve the separation between the particles that originate from different collisions, thus it will constrain the significant pileup background. The MTD detector will be installed between the tracker and the ECAL. The CMS group in the LIP laboratory takes a significant part in the development of the electronics readout for this detector.

Chapter 4

Event reconstruction

The $\mathrm{t\bar{t}} \to \ell\nu_\ell\tau_\mathrm{h}\nu_\tau\mathrm{b\bar{b}}$ signal event topology is defined by the presence of two b quark jets from the t quark decays, one W boson decaying leptonically into $\mathrm{e}\nu$ or $\mu\nu$, and a second W boson decaying into $\tau_\mathrm{h}\nu$. Among these final state objects, the light leptons, e and μ, leave the cleanest signatures in the detector, and serve for triggering the collection of relevant events during the data-taking. The collected raw data from the CMS detector are re-processed with the best algorithms offline to reconstruct physical objects suitable for the analysis. The offline reconstruction is stored in the Analysis Object Data format (AOD) and the mini-AOD data format [64] that leaves out non-essential information such as details in the reconstruction of tracks, hence providing smaller and more manageable datasets.

In each event, all objects are reconstructed with a particle-flow (PF) algorithm [65]. PF combines the information from all subdetectors to identify and reconstruct all types of particles in the event, namely charged and neutral hadrons, photons, muons, and electrons, together referred to as PF objects. These objects are used to construct a variety of higher-level objects and observables, including jets and missing transverse momentum ($\vec{p}_\mathrm{T}^\mathrm{miss}$), which is the negative vector sum of transverse momenta of all reconstructed PF objects. Jets are clustered from the PF objects by the anti-k_T algorithm [66] with the distance parameter $R = 0.4$. Parameters of jets and the tracks associated with jets provide input variables for b tagging discriminators. The reconstructed vertex with the largest value of summed physics-object p_T^2 is taken to be the primary pp interaction vertex. The reconstructed energy of the objects is corrected to account for the deviations in the detector performance.

Each type of objects is required to pass a special identification algorithm that verifies the quality of the reconstructed object. In simple cases it can be just a set of heuristic requirements, but usually it is a multivariate algorithm trained to distiguish the signal and the common background in simulation and observed data. The performance of the identification algorithm is quantified by its efficiency and misidentification probability: the fraction of genuine objects that pass the algorithm is the identification efficiency (true positive case), and the fraction of background objects that mistakenly pass the algorithm is the misidentification probability (false positive case). The algorithm parameters can be adjusted resulting in a looser or tighter requirement, i.e. pass more or less events. Increasing the efficiency to pass signal events always leads to an increase in the misidentification probability. A set of algorithm parameters that results in a particular efficiency and misidentification is called a “working point” of the algorithm. In practice a dedicated group in the collaboration develops an identification algorithm and studies its performance for a number of working points that are useful in a wide range of analyses. These algorithms and working points are further cross-checked in numerous analyses that use the corresponding objects, assuring the quality of the reconstructed data.

The complexity of the pp events and the detector response makes a direct theoretical prediction of experimentally measured distributions impractical. Instead, the theory is presented by a simulation of the collision events that includes the detector response. The simulation is further calibrated to account for particularities of the detector performance in any given data-taking period, and remove the residual differences between the reconstruction in the simulation and the observed data.

Recorded and simulated data samples, triggers, reconstruction tools, and the calibration of the simulation are described in the following.

4.1 Data sample

The data sample was collected with the CMS detector in 2016 and corresponds to the luminosity of $35.9\,\mathrm{fb}^{-1}$. The data recorded with the single light lepton triggers, “Single Electron” and “Single Muon”, are used as the starting point for further event selection.

Each dataset is fully identified by a logical name within the CMS computing infrastructure. The logical names for the datasets collected with the detector are composed of three parts: the group of triggers that collected the events (for example, “SingleElectron”), the data-taking epoch (“Run2016D”) and the event reconstruction campaign (“03Feb2017-v1”), and the dataset format (“MINIAOD”). The data-taking conditions can vary between epochs: the pileup, the parameters of triggers, other particular circumstances can be different. Technical issues encountered in the detector operation during the data-taking period are corrected in the event reconstruction campaign or within the physics analysis. The February 2017 campaign includes the best reconstruction calibrations, that were also cross-checked with the July 2018 legacy reconstruction campaign. The exact logical names of the datasets used in the thesis are listed in Table 4.1.

Table 4.1: The datasets from the 2016 sample that are used in this thesis.

Primary set	Logical names
Single Electron	`/SingleElectron/Run2016B-03Feb2017_ver2-v2/MINIAOD`
	`/SingleElectron/Run2016C-03Feb2017-v1/MINIAOD`
	`/SingleElectron/Run2016D-03Feb2017-v1/MINIAOD`
	`/SingleElectron/Run2016E-03Feb2017-v1/MINIAOD`
	`/SingleElectron/Run2016F-03Feb2017-v1/MINIAOD`
	`/SingleElectron/Run2016G-03Feb2017-v1/MINIAOD`
	`/SingleElectron/Run2016H-03Feb2017_ver2-v1/MINIAOD`
	`/SingleElectron/Run2016H-03Feb2017_ver3-v1/MINIAOD`
Single Muon	`/SingleMuon/Run2016B-03Feb2017_ver2-v2/MINIAOD`
	`/SingleMuon/Run2016C-03Feb2017-v1/MINIAOD`
	`/SingleMuon/Run2016D-03Feb2017-v1/MINIAOD`
	`/SingleMuon/Run2016E-03Feb2017-v1/MINIAOD`
	`/SingleMuon/Run2016F-03Feb2017-v1/MINIAOD`
	`/SingleMuon/Run2016G-03Feb2017-v1/MINIAOD`
	`/SingleMuon/Run2016H-03Feb2017_ver2-v1/MINIAOD`
	`/SingleMuon/Run2016H-03Feb2017_ver3-v1/MINIAOD`

The analysis uses only the events that were recorded under good data-taking conditions, when the sub-detectors and other systems performed well. The good conditions are encoded in "lumisection certificates" for a couple levels of quality. The analysis uses the events that pass the "golden json" lumisection certificate of proton-proton collisions. It selects the events collected under the best conditions of all detector sub-systems. The data passing the certificate correspond to a luminosity of about $35.9\,\mathrm{fb}^{-1}$.

Additional requirements are used in order to reduce the contamination from beam-gas interactions which may produce anomalous events containing halo muons, low-quality tracks, etc. They are applied prior to any event selection, and discard less than 0.1% of events. These requirements are commonly referred to as "missing E_{T} filters":

- primary vertex filter sets basic quality requirements on the reconstructed vertices

- beam halo filter removes the background of secondary particles that are generated outside the CMS cavern, in the collisions of the beam with residual gas inside the LHC vacuum chamber, etc
- HCAL noise filters (HBHE and HBHEiso) aim to eliminate anomalous signals due to rare electronics noise activity in HCAL
- ECAL Trigger Primitive (TP) filter discards events with falsely reconstructed high $p_{\mathrm{T}}^{\mathrm{miss}}$ that appears when certain ECAL channels saturate the energy range of their readout
- bad PF muon filter removes events with punch through pions mistakenly identified as muons and included in the calculation of $p_{\mathrm{T}}^{\mathrm{miss}}$, even though they do not pass all PF requirements for muons.

Overlap events between the electron and muon samples are naturally removed by the exclusive event selections employed in the thesis, which require only one well reconstructed light lepton in the event, either an electron or a muon, and no other leptons that could pass the requirements of the triggers.

4.2 Event simulation

The analysis makes use of simulated samples of $\mathrm{t\bar{t}}$ events, as well as other processes that result in reconstructed τ leptons in the final state. These samples are used to design the event selection, to calculate the acceptance for $\mathrm{t\bar{t}}$ events, and to estimate most of the backgrounds in the analysis.

Signal $\mathrm{t\bar{t}}$ events are simulated with the POWHEG event generator (v2) [67, 68, 69, 70, 71] at next-to-leading-order (NLO) accuracy in quantum chromodynamics (QCD). The parton showers are modelled using PYTHIA (v8.2) [1] with the CUETP8M2T4 underlying event (UE) tune [72]. The background samples used in the measurement of the cross section are simulated with POWHEG and MC@NLO (v2.2.2) [73]. The MC@NLO generator with MLM matching [74] is used for the simulation of W boson production in association with jets (W+jets), and Drell–Yan (DY) production in association with jets at leading-order (LO) accuracy. Here, only the leptonic decays of DY events and W bosons are simulated, and up to four additional jets are included. The diboson processes are produced with NLO accuracy: WW with POWHEG, WZ and ZZ with MC@NLO with FxFx matching [75]. The POWHEG generator is used for the simulation of t-channel single top quark production and single top quark production associated with a W boson (tW) [76, 77].

The single top quark s-channel sample is produced with MC@NLO at NLO accuracy with FxFx matching scheme. The simulated events are produced with a top quark mass of $m_{\mathrm{t}} = 172.5\,\mathrm{GeV}$. The generated events are subsequently processed with PYTHIA using the underlying event tune CUETP8M1 to provide the showering of the partons, and to perform the matching of the soft radiation with the contributions from direct emissions included in the matrix-element (ME) calculations. The default parton distribution functions (PDFs) are the NNPDF3.0 [78]. The τ decays are simulated with PYTHIA, which correctly accounts for the τ lepton polarization in describing the kinematic properties of the decay. The CMS detector response is simulated with GEANT4 [79]. Additional pp interactions in the same or nearby bunch crossings (pileup, PU) are superimposed on the hard collision. Simulated events are reweighted to match the distribution of the number of pileup collisions per event in data. This distribution is derived from the instantaneous luminosity and the inelastic cross section [80].

The next-to-next-to-leading-order (NNLO) expected SM $\mathrm{t\bar{t}}$ pair production cross section of $832^{+20}_{-29}\,(\mathrm{scale}) \pm 35\,(\mathrm{PDF}+\alpha_S)$ [29] ($m_{\mathrm{t}} = 172.5\,\mathrm{GeV}$) is used for the normalization of the number of $\mathrm{t\bar{t}}$ events in the simulation. The first uncertainty includes the uncertainties in the factorization and renormalization scales, while the second is associated with possible choices of PDFs and the value of the strong coupling constant (α_S). The proton structure is described by the CT14 (NNLO) PDF set with the corresponding PDF and α_S uncertainties [81]. The W+jets and DY+jets backgrounds are normalized to their NNLO cross sections calculated with FEWZ (v3.1) [82]. The t-channel and the s-channel single top quark production are normalized to the NLO calculations obtained from HATHOR (v2.1) [83, 84]. The production of tW is normalized to the NNLO calculation [85, 86]. Finally, the production of diboson pairs is normalized to the NLO cross section prediction calculated with MCFM [87, 88] (v7.0). The information about the simulated processes is collected in Table 4.2.

Within the computing infrastructure the simulated datasets are identified by logical names similarly to the data recorded by the detector. The three-part names designate: the simulated hard process and the generator with its main parameters, the sample generation campaign, and the data format. The simulated samples used in the thesis are essentially of the same format as the recorded data plus additional information related to the simulation (“MINIAODSIM”). They were generated in the Summer16 MiniAODv2 MC production campaign.

The datasets essential to the analysis are listed in Table 4.3, together with their normalization cross sections. The $\mathrm{t\bar{t}}$ process including all final states is given in one inclusive dataset. The W+jets process is binned by the number of hard process jets in order to produce more events with higher numbers of the jets, because they

Table 4.2: The main characteristics of the simulated Monte-Carlo datasets that are used in the analysis: the generator, the simulation accuracy, and the cross section of the hard process. The parton showering is performed with PYTHIA (v8.2) [1] in all processes.

Process	Generator	Accuracy	Cross section [pb]
$\mathrm{t\bar{t}}$	POWHEG	NLO	831.76
Single top tW	POWHEG	NLO	35.85
Single top *t*-channel	POWHEG	NLO	80.95
Single top *s*-channel	MC@NLO	NLO	136.02
W+jets	MadGraph	LO	52940
DY+jets	MadGraph	LO	6225.42
ZZ	POWHEG, MC@NLO	NLO	3.784
WW	POWHEG	NLO	62.175
WZ	POWHEG, MC@NLO	NLO	23.768

make a larger contribution to the event selection. The sample of W+jets events is obtained by combining the inclusive W+jets dataset with the W+n-jets datasets. In the combination the events of the inclusive dataset are required to have less hard process jets at the generation level then in any of W+n-jets, which is defined by the "number of partons" (NUP) parameter in the simulation.

Some of the $\mathrm{t\bar{t}}$ simulation uncertainties are obtained from separate datasets. Additional datasets are used in cross-check studies. They are listed in Table 4.4.

4.3 Trigger

The CMS trigger system is a two-level online system designed to select interesting physics events from the produced collisions. The Level 1 (L1) trigger makes a rapid decision based on the limited event information from algorithms implemented in hardware, and the High Level Trigger (HLT) fully reconstructs the events already pre-selected by L1. Events passing the HLT paths specified in Table 4.5 are selected and analyzed offline. For the $\mathrm{e}\tau_\mathrm{h}$ ($\mu\tau_\mathrm{h}$) final state, data are collected with a trigger requiring at least one isolated electron (muon) with a threshold of $p_\mathrm{T} > 27(24)$ GeV. To meet the HLT time limit the isolation variables are calculated from the information of the separate subdetectors instead of the full PF reconstruction. The single-lepton triggers are used because most of light leptons in signal events pass their p_T thresh-

Table 4.3: The Monte-Carlo simulated datasets that are used in this thesis.

Dataset	Cro
$t\bar{t}$	
`/TT_TuneCUETP8M2T4_13TeV-powheg-pythia8/...`	
Single top	
`/ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1/...`	
`/ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1/...`	
`/ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1/...`	
`/ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1/...`	
`/ST_s-channel_4f_leptonDecays_13TeV-amcatnlo-pythia8_TuneCUETP8M1/...`	
W+jets	
`/WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/...`	
`/W1JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/...`	
`/W2JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/...`	
`/W3JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/...`	
`/W4JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/...`	
DY+jets	
`/DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/...`	
`/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/...`	
Dibosons	
`/ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/...`	
`/ZZTo2L2Nu_13TeV_powheg_pythia8/...`	
`/WWToLNuQQ_13TeV-powheg/...`	
`/WWTo2L2Nu_13TeV-powheg/...`	
`/WZTo3LNu_TuneCUETP8M1_13TeV-powheg-pythia8/...`	
`/WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/...`	
`/WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8/...`	
`/WZTo1L3Nu_13TeV_amcatnloFXFX_madspin_pythia8/...`	

Table 4.4: Additional Monte-Carlo datasets that contain some of the parton showering uncertainties in the $t\bar{t}$ modelling, and the W+jets and the multijet QCD datasets for cross-check studies. The W+jets dataset is generated at NLO accuracy with MC@NLO. It is used to validate the shapes of the distributions obtained in the LO dataset of the W+jets events. The multijet QCD MC datasets are used for validation of the data-driven method to estimate the contribution from this background process.

Dataset	Cross section [pb]
$t\bar{t}$ modeling systematics	831.76
`/TT_TuneCUETP8M2T4_13TeV-powheg-fsrdown-pythia8...`	
`/TT_TuneCUETP8M2T4_13TeV-powheg-fsrup-pythia8...`	
`/TT_TuneCUETP8M2T4_13TeV-powheg-isrdown-pythia8...`	
`/TT_TuneCUETP8M2T4_13TeV-powheg-isrup-pythia8...`	
`/TT_TuneCUETP8M2T4down_13TeV-powheg-pythia8...`	
`/TT_TuneCUETP8M2T4up_13TeV-powheg-pythia8...`	
`/TT_hdampDOWN_TuneCUETP8M2T4_13TeV-powheg-pythia8...`	
`/TT_hdampUP_TuneCUETP8M2T4_13TeV-powheg-pythia8...`	
$t\bar{t}$ cross-check	831.76
`/TT_TuneCUETP8M2T4_GluonMoveCRTune_13TeV-powheg-pythia8...`	
`/TT_TuneCUETP8M2T4_QCDbasedCRTune_erdON_13TeV-powheg-pythia8...`	
`/TTJets_TuneCUETP8M2T4_13TeV-amcatnloFXFX-pythia8...`	
`/TT_TuneEE5C_13TeV-powheg-herwigpp...`	
W+jets cross-check	
`/WJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/...`	47447.2
QCD multijet	
`/QCD_HT100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/...`	27540000
`/QCD_HT200to300_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/...`	1717000
`/QCD_HT300to500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/...`	351300
`/QCD_HT500to700_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/...`	31630
`/QCD_HT700to1000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/...`	6802
`/QCD_HT1000to1500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/...`	1206
`/QCD_HT1500to2000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/...`	120.4
`/QCD_HT2000toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/...`	25.25

olds, and these triggers have high efficiencies with small and well-studied systematic uncertainties.

Table 4.5: HLT trigger paths used in the analyses.

for data	for MC
`HLT_Iso(Tk)Mu24_v*`	`HLT_Iso(Tk)Mu24_v4`
`HLT_Ele27_WPTight_Gsf_v*`	`HLT_Ele27_WPTight_Gsf_v7`

4.4 Primary vertex

A reconstructed vertex is required to have a position along the z-coordinate within the nominal detector center ($|z| < 24\,\mathrm{cm}$), and a radial position within the beam spot ($|d_0| < 2\,\mathrm{cm}$). There must be more than four degrees of freedom in the vertex fit (ndof > 4). From the set of the vertices passing these criteria, the vertex with the maximum p_T sum of tracks associated with the vertex is taken to be the primary pp interaction vertex (PV).

Other pp interactions in the bunch crossing do not pose any interest and only clutter the detection environment. These pileup (PU) pp collisions are inevitable in the case of large instantaneous luminosity, when many protons are collided simultaneously in bunches to have a good chance of producing an interesting pp collision per bunch crossing. The high frequency of collisions adds out of time pileup, when particles from previous bunch crossing pollute the currently detected event. Owing to the high resolution of the CMS detector, the charged particles from PU vertices are distinguished from the particles originating from the PV. With the information about the charged particles, the neutral PU activity can be estimated and removed from the detected signal.

4.5 Lepton reconstruction

Electron or muon candidates reconstructed with the PF are required to originate from the primary vertex, pass quality selection criteria, and be isolated relative to other activity in the event. The relative isolation is based on PF objects within a cone of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2} = 0.4$ around the electron or muon track, and defined as $I_\mathrm{rel} = (E_\mathrm{ch} + E_\mathrm{nh} + E_\mathrm{ph} - 0.5\,E_\mathrm{ch}^\mathrm{PU})/p_\mathrm{T}$, where E_ch is the transverse energy deposited by charged hadrons from the primary vertex, E_{nh} and E_{ph} are the respective transverse energies of the neutral hadrons and photons, and $0.5\,E_{\mathrm{ch}}^{\mathrm{PU}}$ is the estimation of the contribution of neutral particles from pileup vertices, calculated as half of the energy of the charged particles from pileup; p_{T} is the electron or muon transverse momentum. The lepton isolation requirements are used to suppress backgrounds from multijet production. The charge misidentification probability for electrons and muons is less than 0.5% and 0.1%, respectively, and is measured from Z boson decays and simulation [89, 90, 53].

Muons are reconstructed by performing a simultaneous global track fit that includes the hits in the silicon tracker and in the muon detector system [91, 92, 93]. Requirements that correspond to the so-called "tight" muon selection are furthermore imposed on the quality of the track fit, number of hits in the pixel, tracker, and number of matched muon segments. They include:

- PF identification of the tracker muon and global muon fit that includes the outer muon detectors;
- $p_{\mathrm{T}} > 26\,\mathrm{GeV}$, $|\eta| < 2.4$;
- χ^2 the global-muon track fit: $\chi^2/\mathrm{ndof} < 10$;
- number of muon detector layers ("stations") hit in the global muon fit > 0;
- number of matched segments in the crossed station > 1;
- muon detector track with transverse impact parameter (IP): $d_{\mathrm{xy}} < 2\,\mathrm{mm}$;
- longitudinal distance of the tracker track with respect to the primary vertex (PV): $|d_{\mathrm{z}}| < 5\,\mathrm{mm}$;
- number of valid hits in the pixel detector > 0;
- number of tracker layers with hits > 5;
- the relative isolation: $I_{\mathrm{rel}} < 0.15$ in the $\Delta R < 0.4$ cone.

Electrons are required to pass "tight" electron selection, that makes use of the information from the tracker and the calorimeters. About 94% of the energy from a prompt electron or photon is deposited in 3-by-3 clusters of ECAL crystals, and 97% is deposited in 5-by-5 clusters. But the material in front of the calorimeter results in bremsstrahlung and photon conversions. Because of the strong magnetic field the energy reaching the calorimeter is spread in the azimuthal angle φ. The spread energy is clustered by building a cluster of clusters, a "supercluster", which is extended in φ. The identification requirements include [94, 95, 90]:

- $p_{\mathrm{T}} > 30\,\mathrm{GeV}$, $|\eta| < 2.4$;
- exclude the ECAL barrel-endcaps transition region in the range of $1.4442 < |\eta_{\mathrm{supercluster}}| < 1.566$;
- the impact parameter of the electron track reconstructed with the Gaussian Sum Filter (GSF) [96] $d_{\mathrm{B}} < 0.02\,\mathrm{cm}$;
- the ratio between the energy deposits in the HCAL H and the ECAL E that correspond to the electron must be $H/E < 0.0414$ for in the barrel, and < 0.0641 in the endcaps;
- the spread $|\Delta\eta_{\mathrm{supercluster}}| < 0.00308$ for the barrel, and < 0.00605 for the endcaps;
- the spread $|\Delta\phi_{\mathrm{supercluster}}| < 0.0816$ for the barrel, and < 0.0394 for the endcaps;
- weighted cluster RMS along η and inside the 5-by-5 block of the ECAL crystals: $\sigma_{\eta\eta} < 0.00998$ for the barrel, and < 0.0292 for the endcaps;
- the ratio between the energy E reconstructed with the ECAL and the momentum magnitude p at the PV reconstructed with the tracker $|1/E - 1/p| < 0.0129$, which accounts for the Bremsstrahlung radiation;
- the number of missing hits in the tracker layers closer to the PV that are expected in the fit hypothesis: ≤ 1;
- pass the conversion veto that confirms that there are no photon conversion vertices associated with the cluster;
- the relative PF isolation parameter: $I_{\mathrm{rel}} < 0.0588$ for the barrel, and $I_{\mathrm{rel}} < 0.0571$ for the endcaps.

Events with additional "loose" global muons with a transverse momentum $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and a relative isolation $I_{\mathrm{rel}} < 0.25$ within $|\eta| < 2.5$ are rejected. A similar veto is applied on "loose" electron candidates with $p_{\mathrm{T}} > 15\,\mathrm{GeV}$, $|\eta| < 2.5$ and isolation $I_{\mathrm{rel}} < 0.0994$ ($I_{\mathrm{rel}} < 0.107$) in barrel (endcaps). The "loose" identification requirements use exactly the same parameters as the "tight" ones, but the thresholds are less restrictive. The thresholds of the "loose" identification are tuned to pass about 90% of genuine objects.

4.6 Jets

The signal final state is expected to contain two b quark flavour jets produced directly from each t quark decay. Jets are clustered from PF candidates using the anti-k_T algorithm with the cone parameter value of 0.4 (AK4) [66]. The anti-k_T algorithm is a sequential recombination algorithm that clusters soft particles with the hard ones. Key feature of the anti-k_T algorithm is that the soft particles do not modify the shape of the jet, while the hard particles do, the jet boundary is resilient with respect to soft radiation, but flexible with respect to hard radiation.

The pileup contribution to the jet energy is mitigated with the charged hadron subtraction (CHS) technique. The charged particles from the PU vertices are simply removed. The neutral PU component is estimated under the assumption of diffuse noise: the PU particles are uniformly distributed on the scale of jet radius, the fluctuations in this distribution are small, the jet clustering does not depend significantly on the noise particles, there are many jets formed from the PU particles. The correction to jet p_T is $p_\mathrm{T}^\mathrm{sub} = p_\mathrm{T} - \rho A^j$, where A^j is the jet area, ρ is the noise level in the event. The ρ is taken from the median p_T^j/A^j of all jets in the event.

Jets reconstructed under this procedure are designated as “AK4PFchs jets” in the CMS data, since they are clustered with the AK4 algorithm from the PF objects, and cleaned from PU with the CHS technique.

The reconstructed jet energy is corrected in a factorized approach to estimate the genuine jet energy. Each level of the correction is essentially a scaling of the jet four momentum with a scale factor that depends on various jet parameters (p_T, η, area, flavor, etc) and takes care of a different effect: the pileup contribution, measured in the MC simulation and verified in data; the detector response, $p_\mathrm{T}^\mathrm{reco}/p_\mathrm{T}^\mathrm{gen}(p_\mathrm{T}, \eta)$, measured in the MC; the residual MC-Data correction for the detector response.

The jets are further selected with identification quality requirements that are based on such quantities as the electromagnetic fraction and the fraction of energy deposited in the “hottest” (i.e. with the largest energy deposit) HCAL photodetectors. Jets are required to pass the “loose” identification criteria [52, 97] and be separated from any isolated light lepton ℓ by $\Delta R(\mathrm{jet}, \ell) > 0.4$.

The b tagging algorithm used (“CSVv2” in Ref. [98]) combines the information of displaced tracks and secondary vertices associated with the jet in a multivariate technique. Jets are said to be b-tagged if the b-tag discriminator passes the “medium” working point. This working point provides a b tagging efficiency of about 66% with a corresponding light-flavour misidentification rate of 1%.

4.7 Missing transverse energy

Missing transverse momentum $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is one of the signatures of the $\mathrm{t\bar{t}}$ dilepton final state. The transverse mass of the $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ and light lepton system is used in the following analyses, because it provides kinematic discrimination between the signal and background processes. The transverse mass is defined as

$$m_{\mathrm{T}} = \sqrt{2|\vec{p}_{\mathrm{T}}^{\ell}||\vec{p}_{\mathrm{T}}^{\mathrm{miss}}|(1 - \cos\Delta\varphi)}, \tag{4.1}$$

where $\Delta\varphi$ is the azimuthal angle difference between the lepton transverse momentum vector, $\vec{p}_{\mathrm{T}}^{\ell}$, and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$. The reconstructed $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ includes energy scale corrections of the reconstructed objects. This definition is called "PF Type-1 MET" in the dedicated performance studies [99]. The systematic uncertainties in the energy scales are also propagated to the $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$.

4.8 Hadronic τ lepton reconstruction

About two thirds of the τ lepton decays produce a neutrino and hadrons in the final state, τ_{h}, predominantly one or three charged ($h^{\pm}$) mesons and up to two neutral (h^0) mesons, mostly charged and neutral pions ($\pi^{\pm}$ and π^0). These distinct features of the tau hadronic decays have been exploited for the identification of hadronically decaying tau leptons. The main hadronic tau decay modes are listed in Table 4.6.

Table 4.6: Dominant hadronic decays of the τ lepton.

Decay mode	Branching fraction (%)
$\tau^- \to h^- \nu_\tau$	11.6
$\tau^- \to h^- h^0 \nu_\tau$	26.0
$\tau^- \to h^- h^0 h^0 \nu_\tau$	10.8
$\tau^- \to h^- h^+ h^- \nu_\tau$	9.8
$\tau^- \to h^- h^+ h^- h^0 \nu_\tau$	4.8
Other hadronic modes	1.7
Total	64.8

Hadronic τ lepton decays are reconstructed with the hadron-plus-strips (HPS) algorithm [3], which starts from reconstructed jets. In each jet, a charged hadron is combined with other nearby charged hadrons or photons to identify the decay modes.

The identification of π^0 mesons is enhanced by clustering electrons and photons in "strips" along the track bending direction to take into account possible broadening of calorimeter signatures by early showering photons. The τ_h candidates are selected from the following combinations of charged hadrons and strips that correspond to the τ decay modes: single hadron, hadron plus a strip, hadron plus two strips, and three hadrons.

A multivariate analysis of these HPS τ_h candidates is used to reduce the contamination from quark and gluon jets. A boosted decision tree is trained using a sample of DY events with τ_h decays as signal and a sample of QCD multijet events as background, both from simulation. Input variables include the multiplicity and the transverse momenta of the electron and photon candidates in the vicinity of the τ_h, the kinematic properties of hadrons and strips, and the τ_h lifetime information, such as the impact parameter of the leading track and the length of flight significance in the τ_h candidates with three charged hadrons. Additional requirements are applied to discriminate genuine τ_h leptons from prompt electrons and muons. The τ_h identification efficiency of this algorithm is estimated to be approximately 60% for $p_\mathrm{T} > 20\,\mathrm{GeV}$. It is measured in a sample enriched in $\mathrm{Z} \to \tau\tau \to \mu\tau_\mathrm{h}$ events with a "tag-and-probe" technique [3]. The corresponding probability for a generic hadronic jet to be misidentified as τ_h is less than 1% [3].

The τ_h charge is taken as the sum of the charges of the corresponding charged hadrons. The misidentification probability for the charge is less than 1%. It is estimated from $\mathrm{Z} \to \tau\tau \to \mu\tau_\mathrm{h}$ events with same-charge μ and τ_h candidates.

4.9 Calibration of the simulated events

Event weights are applied to the MC datasets to tune the simulated detector performance to the real data-taking conditions. The weights either only reshape the simulation and keep the overall number of simulated events unchanged, by adjusting some of the parameters such as the simulated number of pileup vertices. Or they correct the trigger or object identification efficiency, which affects the number of selected events. The weights are calculated either for per-event parameters, such as PU vertices, or per a specific object, such as jet energy. Under the assumption of no correlation between corrected parameters, the product of the weights yields the full correction weight. Most of parameters are uncorrelated because of their different physical nature. In other cases, such as the jet energy scale and resolution, or the $\vec{p}_\mathrm{T}^\mathrm{miss}$, the correlations are properly taken into account by considering the relevant parameters together and propagating the corrections between them.

Listed in the order of significance, the corrections to the simulated events are the following:

- **Tau identification efficiency:** the scale factor of 95% with 5% uncertainty is applied to the processes with genuine τ leptons to correct the simulated τ_{h} identification efficiency in the 2016 dataset. It is measured with a tag-and-probe technique in samples enriched in $\mathrm{Z} \to \tau_{\ell}\tau_{\mathrm{h}}$ events [3].

- **Pile-up (PU) distribution:** additional minimum bias interactions are superimposed on simulated events to include the effect of in-time and out-of-time PU in the events. The number of proton-proton interactions per bunch crossing is not well reproduced in the simulation, because the challenging magnitude of the instantaneous luminosity at the LHC leads to rapidly changing data-taking conditions. As a consequence there is a difference in the number of reconstructed primary vertices between the observed data and the simulated samples. To bring the data and the simulation to consistency, the number of primary vertices in the simulation is reweighted to match the distribution observed in data. A minimum bias cross section of 69.2 with a 4.6% uncertainty is used to estimate the PU distribution following the studies from Ref. [100].

- **Top p_{T} reweighting:** in 13 TeV runs the p_{T} spectra of top quarks in data was found to be significantly softer than those predicted by various MC simulations based on either LO or NLO matrix elements interfaced with parton showers. The effect was confirmed (albeit to a lesser degree) by ATLAS. Although predictions at NNLO+NNLL [101] and approx. NNNLO in pQCD [86] provide a much improved description, a residual discrepancy remains [33, 102, 103, 104]. A reweighting procedure is included as one of systematic uncertainties to cover the difference between the predicted and observed spectra.

- **Lepton identification, isolation, and trigger efficiency:** the difference in the performance of the lepton trigger, identification, and isolation efficiencies are measured in data and simulation with a tag-and-probe method in $\mathrm{Z} \to \ell^{+}\ell^{-}$ events [89, 90, 53]. The simulated events are corrected with the corresponding data-to-simulation scale factors.

- **Jet energy scale and resolution:** the corrections to the jet energy scale (JES), jet energy resolution (JER), and $p_{\mathrm{T}}^{\mathrm{miss}}$ scales are estimated according to the prescription described in Ref. [52]. A factorized combination of corrections fixes different effects in the JES: the pileup contribution, the detector response in the recorded data, the residual data-to-simulation disagreement in the detector response. And a (p_{T}, η)-dependent procedure broadens the JER in the simulation adjusting it to the detector performance in the real data-taking: the JER is simply scaled according to the measured data-to-simulation scale factors for the jets that match to the generator level jets, otherwise the JER is broadened stochastically to correspond to the width in the data. All these corrections are propagated to the $p_{\mathrm{T}}^{\mathrm{miss}}$ scale.

Although the jets in signal events are expected to be produced from b-quark fragmentation and subsequent hadronization, no flavor specific jet energy corrections are applied.

- **The b tagging efficiency:** the data-simulation difference in the b tagging (mistagging) is accounted for by applying (p_{T}, η)-dependent scale factors measured in a variety of control samples enriched in b quarks (c and light-flavour quarks) [98]. The data-to-simulation scale factors for b, c, and light-flavour jets are applied to the simulation.

  The event weight w_b corrects the probability to encounter a given configuration of tagged and mistagged jets. It is calculated from the tagging probabilities of separate jets as follows:

$$
\begin{aligned}
P(\mathrm{data}) &= \prod_{i \in \mathrm{tagged}} \epsilon_i \prod_{j \in \mathrm{not\ tagged}} (1 - \epsilon_j) \\
P(\mathrm{MC}) &= \prod_{i \in \mathrm{tagged}} \mathrm{SF}_i \epsilon_i \prod_{j \in \mathrm{not\ tagged}} (1 - \mathrm{SF}_j \epsilon_j) \\
w_b &= P(\mathrm{data}) \Big/ P(\mathrm{MC}).
\end{aligned}
$$

  Where ϵ is the b tagging (mistagging) efficiency for b flavour (light flavour) jets, and SF is the scale factor between an observed and a simulated jet. Both are functions of jet p_{T}, η, and flavour. Therefore, the $P(\mathrm{data})$ is the probability that the b tagging algorithm will return a given combination of tagged jets in real data. And the $P(\mathrm{MC})$ is the probability for the same combination of b tags in the simulation. This formula corrects the tagging probability of the combination of jets in the event, having the correction SF_i for each individual jet.

- **Integrated luminosity:** the number of events simulated in each process is scaled to correspond to the integrated luminosity of the 2016 data sample. The effective generated luminosity is computed as the sum of the per-event weights. The general formula of the weight applied to the ith simulated dataset is:

  $$w_i = \mathcal{L} \cdot \sigma_i \Bigg/ \sum_{e=1}^{N_{\text{gen}}} w_e$$

  where $\mathcal{L}$ is the $35.9\,\text{fb}^{-1}$ of the integrated luminosity in 2016, σ_i is the theory prediction for the cross section of the process, and w_e are the event weights.

Every calibration correction includes a certain systematic uncertainty measured in the corresponding study. These uncertainties are included in the following measurements.

Chapter 5

The estimation of the background from misidentified tau leptons

The τ lepton decay into hadrons (τ_{h}) can be mimicked by a regular hadronic jet originating from a background process. A typical magnitude of the jet-to-tau misidentification probability is less than 1%. But since there are multijet processes with large cross sections, such as hadronic $\mathrm{t\bar{t}}$ decays, multijet QCD, and W+jets, this background is substantial in the studies that have τ_{h} in the final state.

In the $\mathrm{t\bar{t}} \to \ell\nu_{\ell}\tau_{\mathrm{h}}\nu_{\tau}\mathrm{b\bar{b}}$ cross section measurement the largest contribution of misidentified τ_{h} comes from the $\mathrm{t\bar{t}}$ lepton+jets final state $\mathrm{t\bar{t}} \to \ell\nu_{\ell}\mathrm{q\bar{q}'b\bar{b}}$, shown in Figure 5.1, that has a signature very similar to the signal and a much larger production cross section. The jets produced in this process originate from the b quarks, the c and light-flavour quarks from the hadronic W boson decay, and from the initial- and final-state radiation quarks and gluons. Each of these types of jets has a significant τ_{h} misidentification probability. Therefore the analysis adjusts the estimation of this background to the composition of jets in the final state, instead of relying on one simple control region to estimate the contribution from this background, such as the $\mathrm{t\bar{t}} \to \mathrm{e}\nu_{\mathrm{e}}\mu\nu_{\mu}\mathrm{b\bar{b}}$ light dilepton final state.

Depending on the properties of the jet, the misidentification probability can differ by as much as an order of magnitude, as demonstrated in Figure 5.2 and can also be seen in Ref. [3]. The high variability of the τ_{h} misidentification probability poses a challenge to correctly estimate the background of misidentified τ_{h} in a particular final state. The estimation must be suited to the particular composition of jets in the considered events, and it must be performed with a data-driven approach, since the modelling of misidentified objects is not reliable.

This chapter presents an analysis of the jets misidentified as τ_{h} and the methods

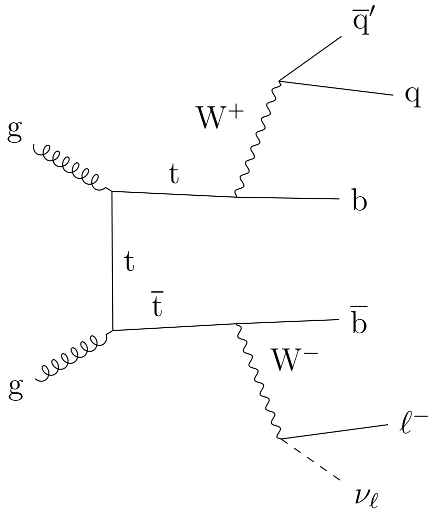

Figure 5.1: Feynman diagram of the $t\overline{t}$ decay into the lepton+jets (semileptonic) final state. This is the main background process whenever a jet is misidentified as a τ_h. It contributes about 80% of all background events to a typical event selection.

to measure their misidentification probabilities with the focus on the direct measurement method that was employed in Run1 analyses [18, 19, 105]. The method estimates the misidentification probabilities of each jet flavour and combines them according to the jet composition in the signal region. This approach touches upon all the details of the determination of this background and serves well to study the factors that affect the τ_h misidentification probability. After that, two ways to improve the background estimation are discussed: the "fake factor" method that builds on top of the direct measurement, and the "shape fit" that overcomes the limits of the others by making the measurement independent from the major factors that affect the τ_h misidentification probability, such as the jet flavour of the misidentified τ_h candidate.

5.1 Direct measurement of the τ_h misidentification probability

The τ_h misidentification probability of a jet depends on its characteristics: the flavour, kinematic parameters, etc. With the jet-to-tau misidentification probability known as a function of these parameters, the number of events with misidentified τ_h

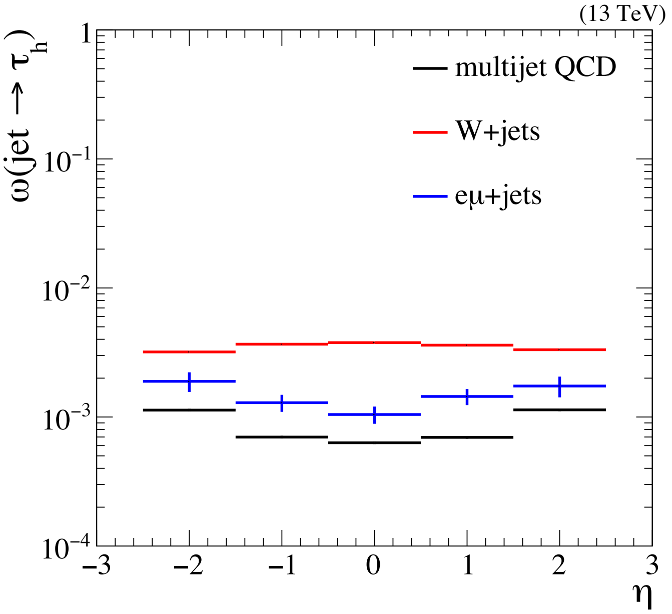

Figure 5.2: Typical distributions of the τ_h misidentification probability per η of the τ_h candidate, as measured in three different event selections in the following analysis. The eμ+jets selection is enriched in $\mathrm{t\bar{t}} \to \mathrm{e}\nu_\mathrm{e}\mu\nu_\mu\mathrm{b\bar{b}}$ events. The misidentification probability distributions measured in this channel were used in the Ref. [3], as one of my service tasks in the CMS collaboration.

that pass into a given selection can be estimated from the jets in the pre-selection, i.e. the same event selection without the τ_h identification requirement:

$$N^{\mathrm{mis}-\tau_\mathrm{h}} = \sum^{\mathrm{events}} \sum_{j}^{\mathrm{jets}} \rho(k_j), \tag{5.1}$$

where $\rho(k_j)$ is the misidentification probability for the jth jet with k_j characteristics that can include jet flavour, p_T, η, and other parameters of the jet structure. Since the probabilities are small, they are simply summed up for all jets in processes without genuine τ_h. The misidentified jets in the processes with genuine τ_h are neglected, because the contribution of real τ_h outweighs the misidentified objects by at least two orders of magnitude, as discussed in Ref. [3] and Chapter 4.

This section presents a direct measurement of the τ_h misidentification probabilities, and studies their application to the $\mathrm{t\bar{t}}$ cross section measurement in the dilepton final state that includes a τ lepton that decays into hadrons. The systematic uncertainty and overall performance of the method are discussed at the end.

5.1.1 Event selection

The events for the measurement in the $\mathrm{t\bar{t}}$ dilepton final state with a τ_h are selected by requiring one isolated electron (muon) with transverse momentum $p_\mathrm{T} > 30$ (26) GeV and $|\eta| < 2.4$, at least two jets with $p_\mathrm{T} > 30\,\mathrm{GeV}$ and $|\eta| < 2.5$, and exactly one τ_h candidate with $p_\mathrm{T} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$. The missing transverse momentum in the event is required to pass $p_\mathrm{T}^\mathrm{miss} \geq 40\,\mathrm{GeV}$, that reduces the background of multijet events. The τ_h candidate and the light lepton are required to have opposite electric charges (OC). At least one jet is required to be identified as originating from b quark hadronization ("b-tagged") by passing the "medium" working point of the "CSVv2" b flavour discriminator [98]. The τ_h candidate is required to pass the "medium" working point of the MVA-based anti-jet discriminator. The b-tagged jets are required to be separated by $\Delta R > 0.4$ from the τ_h jet. Electrons and muons are required to be separated from any jet and from the τ_h candidate in the η-φ plane by $\Delta R > 0.4$. Events with any additional loosely isolated electron (muon) of $p_\mathrm{T} > 15$ (10) GeV are rejected. An electron is considered loosely isolated if $I_\mathrm{rel} < 0.0994$ in the barrel or $I_\mathrm{rel} < 0.107$ in the endcaps. A muon is loosely isolated if $I_\mathrm{rel} < 0.25$ in either the barrel or the endcaps. The rest of the requirements for the reconstructed final state objects are given in Chapter 4.

In order to apply the method of directly measured misidentification probabilities, a pre-selection is considered where the requirement of the τ_h anti-jet discriminator is omitted. The $\mathrm{t\bar{t}} \to \ell\nu_\ell \mathrm{q\bar{q}'b\bar{b}}$ process contributes the dominant fraction of events in the pre-selection.

The expected and observed event yields corresponding to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$ are listed in Tables 5.1 and 5.2 for the $\mathrm{e}\tau_\mathrm{h}$ and $\mu\tau_\mathrm{h}$ final states, respectively. Figures 5.3 and 5.4 show the distributions of the main kinematic parameters at the pre-selection. Under the assumption of the SM signal, both final states exhibit a good agreement between the data and simulation. However, the processes with misidentified τ_h must be estimated independently from the assumptions about the signal. These processes are determined directly from data, as discussed in the following.

Table 5.1: The expected and observed numbers of events for the signal and the backgrounds at different selection stages of the $\mathrm{e}\tau_\mathrm{h}$ final state for an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$.

Process	1 e lepton	≥ 3 jets	$p_\mathrm{T}^\mathrm{miss} \geq 40\,\mathrm{GeV}$	≥ 1 b tag	1 τ_h	OC
$\mathrm{t\bar{t}} \to \mathrm{e}\nu_\mathrm{e}\tau_\mathrm{h}\nu_\tau\mathrm{b\bar{b}}$	176738	82133	67271	53396	6266	6150
other $\mathrm{t\bar{t}}$ backgrounds						
$\mathrm{t\bar{t}} \to \mathrm{e}\nu_\mathrm{e}\mathrm{q\bar{q}'b\bar{b}}$	1246360	958744	668375	513864	3170	2253
$\mathrm{t\bar{t}}$ other dilepton	111931	50701	42255	33129	134	84
$\mathrm{t\bar{t}}$ multijet	80109	58713	46559	35585	560	490
other backgrounds						
DY+jets	9406010	247402	82919	10759	217	180
W+jets	106029000	1116060	698876	75283	212	133
QCD multijet	5132386	421571	149616	41964	0	0
Single top	390131	111388	77801	54877	666	586
MC total	122572665	3046712	1833672	818857	11225	9876
Data	141794839	3468656	1887525	805904	11406	9874

5.1.2 Estimation of the misidentified τ_h background

The selected sample of events is composed of the signal (about 60% of all events), the background processes with genuine τ_h (5%), and the background with misidentified τ_h (35%). In particular, the background processes that contribute events with genuine τ_h are the single top quark production and the Drell–Yan $\tau\tau$ final state. The background of misidentified τ_h comes from events with an electron or a muon, and three or more

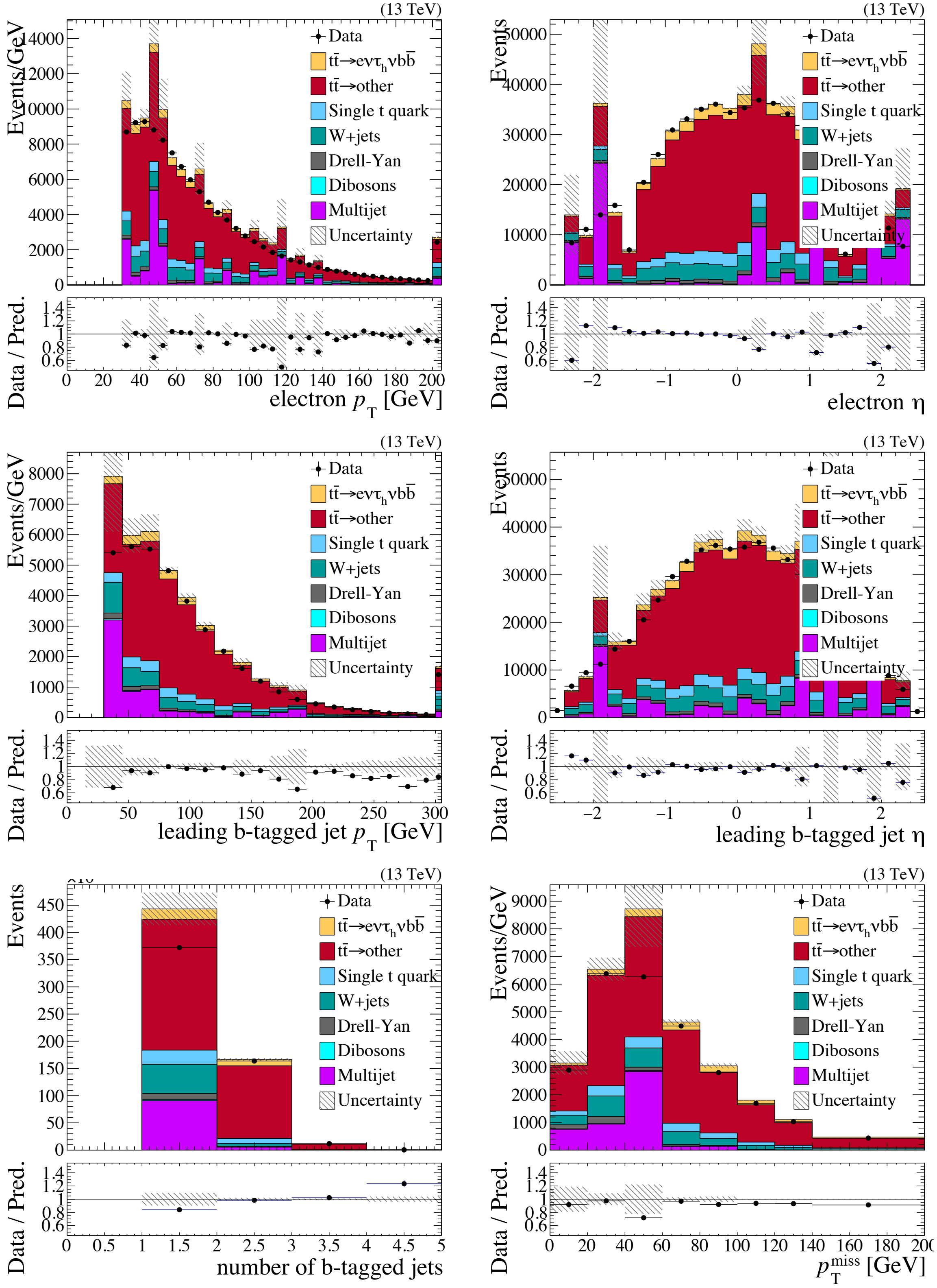

Figure 5.3: Main distributions in the eτ_h pre-selected events. The multijet QCD contribution is taken from simulation and exhibits large statistical fluctuations. Only statistical uncertainties are shown.

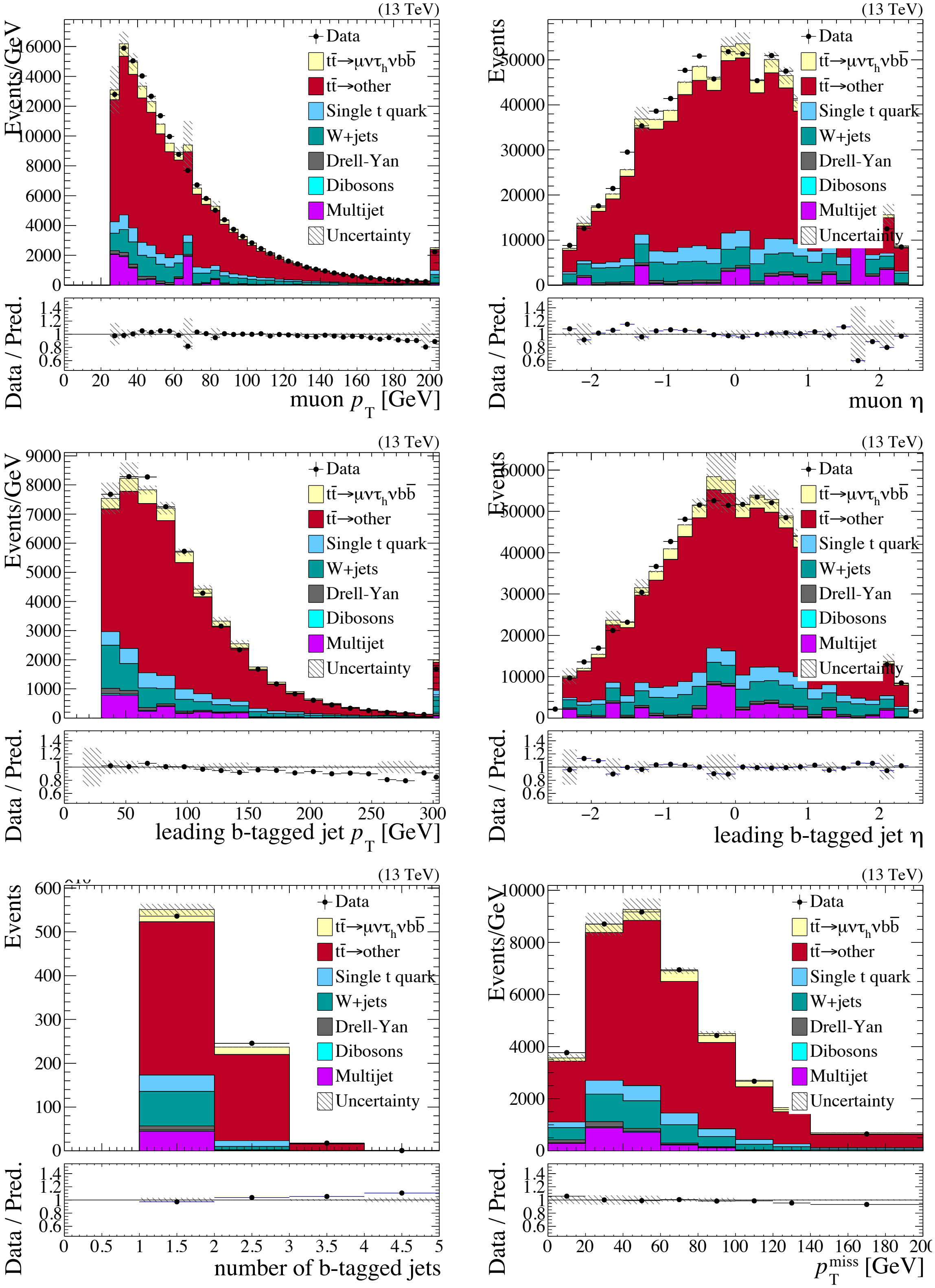

Figure 5.4: Main distributions in the $\mu\tau_h$ pre-selected events. The multijet QCD contribution is taken from simulation and exhibits significant statistical fluctuations. Only statistical uncertainties are shown.

Table 5.2: The expected and observed numbers of events for the signal and the backgrounds at different selection stages of the $\mu\tau_\mathrm{h}$ final state for an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$.

Process	1 μ lepton	≥ 3 jets	$p_\mathrm{T}^\mathrm{miss} \geq 40\,\mathrm{GeV}$	≥ 1 b tag	1 τ_h	OC
$\mathrm{t\bar{t}} \to \mu\nu_\mu\tau_\mathrm{h}\nu_\tau\mathrm{b\bar{b}}$	283160	130788	107188	85388	10118	9963
other $\mathrm{t\bar{t}}$ backgrounds						
$\mathrm{t\bar{t}} \to \mu\nu_\mu\mathrm{q\bar{q}'b\bar{b}}$	2013684	1542979	1082045	833525	5212	3735
$\mathrm{t\bar{t}}$ other dilepton	181540	81429	67872	53506	208	129
$\mathrm{t\bar{t}}_{jj}$	141642	102680	81237	61921	984	859
other backgrounds						
DY+jets	11311029	198157	84125	11117	330	301
QCD multijet	8233294	428361	177013	51535	36	20
W+jets	205885000	1566790	995209	109789	384	271
Single top	648487	177144	124291	87759	1037	902
MC total	228697836	4228328	2718980	1294540	18339	16198
Data	240100613	4732560	2940479	1294791	17709	15355

jets, where at least one jet passes the b tagging discriminant and one is misidentified as the τ_h jet. The dominant contribution of this background is from the $\mathrm{t\bar{t}} \to \ell\nu_\ell\mathrm{q\bar{q}'b\bar{b}}$ process, the multijet QCD processes, and the W boson production in association with jets.

The processes with genuine τ_h can be modelled reliably, therefore they are taken from the simulation with necessary corrections. The contribution of misidentified τ_h backgrounds is estimated with the misidentification probabilities measured in observed data.

The τ_h misidentification probability ω is parametrized as a function of the jet flavour, p_T, η, and jet radius defined as $r_\mathrm{jet} = \sqrt{\sigma_{\eta\eta}^2 + \sigma_{\phi\phi}^2}$, where $\sigma_{\eta\eta}^2$ and $\sigma_{\phi\phi}^2$ are standard deviations in the energy distributions of the jet constituents. Three flavours are considered to account for the significant differences in the corresponding final state signatures: jets originating from b quarks, light quarks, and gluons. Quarks and gluons produce jets with different structures, because of the difference in colour. The b quarks produce long-lived heavy hadrons that contribute particular signatures in the final state jets, distinct from the other quarks.

The misidentification probabilities are measured in three independent samples from data: QCD multijet, W+jets, and eμ+jets. The events selected in these samples are expected to be enriched in different types of jets: originating from gluons, light quarks, and b quarks, respectively. The selected events are also expected to have a negligible contribution of genuine τ_h that can be subtracted using the MC estimation. Specific event selection requirements for each sample are the following.

- W+jets:
  - events are collected with the inclusive single muon HLT described in Chapter 4;
  - one isolated muon with $p_\mathrm{T} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$ that triggers the HLT;
  - at least one jet with $p_\mathrm{T} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$;
  - the leptons are separated in η and ϕ from the jets by requiring $\Delta R > 0.4$ between them.
- QCD multijet:
  - events are collected with the HLT that requires a PF jet with $p_\mathrm{T} > 140\,\mathrm{GeV}$;
  - at least two jets with $p_\mathrm{T} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$, one of which triggers the HLT;
  - to avoid trigger bias, the jet matched to the HLT is not considered in the measurement of the τ_h misidentification probability.
- eμ+jets:
  - collected with the inclusive single muon HLT;
  - one isolated muon and one isolated electron of opposite charges with $p_\mathrm{T} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$;
  - at least one jet with $p_\mathrm{T} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$;
  - the jets are required to be separated from the leptons by $\Delta R > 0.4$.

In each sample, the number of jets passing the τ_h identification requirements is divided by the number of all jets in bins of p_T, η, r_jet to give the misidentification probability ω:

$$\omega(p_\mathrm{T}, \eta, r_\mathrm{jet}) = \left.\frac{N_{\tau_\mathrm{h}}}{N_\mathrm{jets}}\right|_{(p_\mathrm{T}, \eta, r_\mathrm{jet})}. \tag{5.2}$$

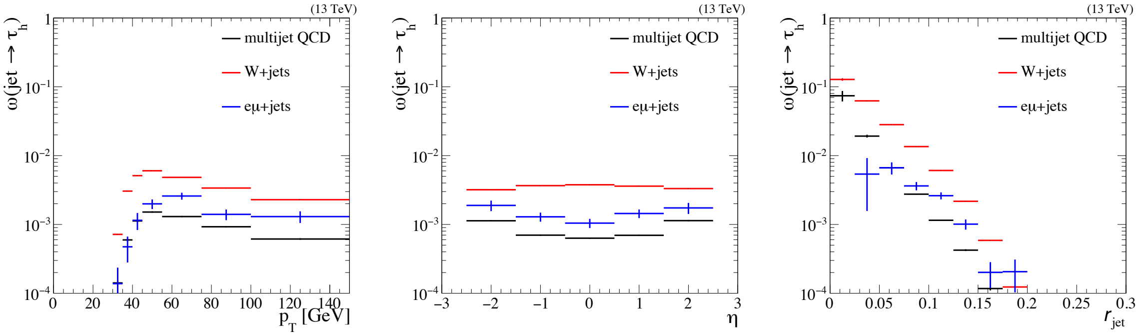

Figure 5.5: Misidentification probabilities as functions of the jet variables (p_{T}, η, r_{jet}) measured in the control samples.

The number of genuine τ_{h} among the selected jets is taken from simulation and is subtracted in the numerator and the denominator before calculating the ratio. It is negligible for all samples except W+jets where a small fraction of events with genuine τ_{h} comes from $\mathrm{t\bar{t}}$ processes. The bins are chosen with statistically significant size: more than 1000 jets in a bin and at least 10 misidentified τ_{h}. Comparisons of the misidentification probabilities as function of the three jet variables (p_{T}, η, r_{jet}) measured in each sample are shown in Figure 5.5.

The misidentification probability $\omega(\mathrm{t\bar{t}})$ for the jets in the $\mathrm{t\bar{t}}$ pre-selection is estimated from the probabilities measured in the control samples. The estimated distribution is applied to every jet in observed data at the pre-selection to obtain the number of jets that pass the τ_{h} selection requirement:

$$N(\mathrm{mis}{-}\tau_{\mathrm{h}}) = \sum^{\mathrm{events}} \sum_{j}^{\mathrm{jets}} \omega(j) - N(\tau_{\mathrm{h}}) \tag{5.3}$$

The estimated number $N(\mathrm{mis}{-}\tau_{\mathrm{h}})$ of misidentified τ_{h} is multiplied by the efficiency $\epsilon_{\mathrm{OC}} \simeq 0.7$ of the opposite charge (OC) requirement that is obtained from the simulation to determine the background with misidentified τ_{h} that passes all requirements of the main event selection.

The $N(\tau_{\mathrm{h}})$ is a small (20%) contamination of genuine τ_{h} contribution inside the estimated misidentified τ_{h} background, which is obtained from MC by applying the same misidentification probability to MC events with genuine τ_{h}.

Due to the large fractions of gluon and b flavour jets, the probabilities $\omega(\mathrm{QCD})$ and $\omega(\mathrm{e}\mu\mathrm{+jets})$ measured in the multijet QCD and eμ+jets samples tend to underestimate the actual $N(\mathrm{mis}{-}\tau_{\mathrm{h}})$ in the $\mathrm{t\bar{t}}$ selection. On the contrary, the $\omega(\mathrm{W+jets})$

probability tends to overestimate the $N(\mathrm{mis}{-}\tau_\mathrm{h})$ value. In order to account for the difference in the jet flavours, the $\omega(\mathrm{t\bar{t}})$ probability is estimated with a weighted average of the misidentification probabilities measured in the control samples. The weights for $\omega(\mathrm{W+jets})$, $\omega(\mathrm{e}\mu\mathrm{+jets})$, and $\omega(\mathrm{QCD})$ are chosen to reproduce the jet flavour composition in the pre-selection.

The correspondence between the measured τ_h misidentification probability and the jet flavour composition in the samples is expressed by:

$$\begin{bmatrix} \omega(\mathrm{W{+}jets}) \\ \omega(\mathrm{e}\mu\mathrm{{+}jets}) \\ \omega(\mathrm{QCD}) \end{bmatrix} = \begin{bmatrix} \%_q(\mathrm{W{+}jets}) & \%_b(\mathrm{W{+}jets}) & \%_g(\mathrm{W{+}jets}) \\ \%_q(\mathrm{e}\mu\mathrm{{+}jets}) & \%_b(\mathrm{e}\mu\mathrm{{+}jets}) & \%_g(\mathrm{e}\mu\mathrm{{+}jets}) \\ \%_q(\mathrm{QCD}) & \%_b(\mathrm{QCD}) & \%_g(\mathrm{QCD}) \end{bmatrix} \times \begin{bmatrix} \rho_q \\ \rho_b \\ \rho_g \end{bmatrix}. \tag{5.4}$$

Where ωs are the overall τ_h misidentification probabilities measured in the control samples, and $\rho_{\{q,b,g\}}$ are the unknown probabilities for different jet flavours, that are multiplied by the % fractions of the light quark, b quark, and gluon jets in the samples.

Finally, the weights W for the average are determined by imposing that the final estimate should have the same result as applying the unknown τ_h misidentification probabilities of the light quark, b quark, and gluon jets to the corresponding quark and gluon jet fractions in the pre-selection:

$$\begin{split} W(\mathrm{W{+}jets}) \times \omega(\mathrm{W{+}jets}) + W(\mathrm{e}\mu\mathrm{{+}jets}) \times \omega(\mathrm{e}\mu\mathrm{{+}jets}) + W(\mathrm{QCD}) \times \omega(\mathrm{QCD}) = \\ \rho_q \times \%_q(\mathrm{t\bar{t}}) + \rho_b \times \%_b(\mathrm{t\bar{t}}) + \rho_g \times \%_g(\mathrm{t\bar{t}}) = \omega(\mathrm{t\bar{t}}). \end{split} \tag{5.5}$$

From Eq. 5.4 these weights can be found in the following system that fixes the jet composition of the average to the $\mathrm{t\bar{t}}$ pre-selection:

$$\begin{bmatrix} \%_q(\mathrm{t\bar{t}}) \\ \%_b(\mathrm{t\bar{t}}) \\ \%_g(\mathrm{t\bar{t}}) \end{bmatrix} = \begin{bmatrix} \%_q(\mathrm{W{+}jets}) & \%_q(\mathrm{e}\mu\mathrm{{+}jets}) & \%_q(\mathrm{QCD}) \\ \%_b(\mathrm{W{+}jets}) & \%_b(\mathrm{e}\mu\mathrm{{+}jets}) & \%_b(\mathrm{QCD}) \\ \%_g(\mathrm{W{+}jets}) & \%_g(\mathrm{e}\mu\mathrm{{+}jets}) & \%_g(\mathrm{QCD}) \end{bmatrix} \times \begin{bmatrix} W(\mathrm{W{+}jets}) \\ W(\mathrm{e}\mu\mathrm{{+}jets}) \\ W(\mathrm{QCD}) \end{bmatrix}. \tag{5.6}$$

The jet flavour compositions in the multijet QCD, W+jets, eμ+jets, and $\mathrm{t\bar{t}}$ pre-selection events are taken from simulation. They are listed in Table 5.3. The weights determined by the solution of Eq. 5.6 are given in Table 5.4.

The uncertainty in the weighed average is determined as the square root of the weighted sum of squared differences between the $N(\mathrm{mis}{-}\tau_\mathrm{h})$ values estimated with the average and separate misidentification probabilities:

$$\delta N(\mathrm{mis}{-}\tau_\mathrm{h}) = \sqrt{\sum W(s)\Delta N(\mathrm{s})^2 \Big/ \sum W(s)}. \tag{5.7}$$

Table 5.3: Fractions of jet flavours in the control samples and in the $t\bar{t}$ pre-selection. Values are taken from the simulation.

Sample	b quarks	light quarks	gluons
QCD multijet	0.044	0.133	0.748
W+jets	0.051	0.673	0.286
$e\mu$+jets	0.513	0.223	0.244
$t\bar{t} \rightarrow e\tau_h$ pre-selection	0.397	0.345	0.234
$t\bar{t} \rightarrow \mu\tau_h$ pre-selection	0.407	0.340	0.236

Table 5.4: The weights of the control regions in the calculation of the weighted misidentification probability for the weighted average, as determined by fixing the jet composition of the average to the pre-selection.

Sample	W(QCD)	W(W+jets)	W($e\mu$+jets)
$e\tau_h$	-0.04	0.23	0.75
$\mu\tau_h$	-0.03	0.22	0.77

Where the index s covers the control regions, and ΔNs are the corresponding deviations from the $N(\text{mis}-\tau_h)$ determined with the weighted average. This approach assumes a good characterization of the jet-to-tau misidentification probability with the chosen jet parameters, and takes into account the asymmetry between the multijet QCD, W+jets, and $e\mu$+jets weight factors.

The estimation of the misidentified τ_h background with the weighted averages is shown in Table 5.5. The result is close to the estimation from the simulation listed in Tables 5.1 and 5.2. The difference between the two estimations is approximately the value of the systematic uncertainty in the weighted average. Therefore, the simulation serves as a cross-check for the data-driven method of the weighted averages.

5.1.3 The systematic uncertainty in the cross section measurement

The observed cross section is determined by counting the number of $e\tau_h$ and $\mu\tau_h$ events consistent with originating from $t\bar{t}$, subtracting the contributions from other

Table 5.5: Weights for the measured τ_{h} misidentification probabilities and the estimation of the misidentified background in the $\mathrm{t\bar{t}}$ selection. The columns "Estimation" list the number of events with misidentified τ_{h} estimated with the measured misidentification probabilities and their weighted average calculated with the coefficients from the columns "Weight". The coefficients adjust the jet compositions in the averages to the $\mathrm{t\bar{t}}$ pre-selections. The genuine τ contributions are obtained from the simulation. The last row gives the estimation in the final selection after the application of the opposite charge requirement ϵ_{OC} efficiency.

	$\mathrm{e}\tau_{\mathrm{h}}$		$\mu\tau_{\mathrm{h}}$	
Sample	Weight	Estimation	Weight	Estimation
QCD multijet	-0.04	3654	-0.03	5507
W+jets	0.23	6207	0.22	9428
eμ+jets	0.75	4886	0.77	7486
Weighted average $N(\mathrm{mis}{-}\tau_{\mathrm{h}})$		4946 ± 566		7672 ± 767
$N(\mathrm{mis}{-}\tau_{\mathrm{h}}) - N_{\mathrm{MC}}(\tau_{\mathrm{h}})$		4008 ± 566		6123 ± 767
Opposite Charge		2845 ± 402		4347 ± 545

processes, and correcting for the efficiency of the event selection:

$$\sigma^{\mathrm{fid}}_{\mathrm{t\bar{t}}}(\ell\tau_{\mathrm{h}}) = \frac{N - B}{\mathcal{L} \cdot \mathcal{B} \cdot \epsilon}, \tag{5.8}$$

where N is the number of observed candidate events, B is the estimate of the background, $\mathcal{L}$ is the integrated luminosity, $\mathcal{B}$ is the product of all branching fractions in the $\mathrm{t\bar{t}} \to \ell\nu_\ell\tau_{\mathrm{h}}\nu_\tau\mathrm{b\bar{b}}$ decay, and ϵ is the overall efficiency of the detector, that includes the efficiencies of all reconstruction algorithms.

The value $\sigma^{\mathrm{fid}}_{\mathrm{t\bar{t}}}(\ell\tau_{\mathrm{h}})$ is the cross section of events that can be observed by the detector, i.e. that pass into the fiducial phase space of the detector defined by the geometrical and kinematic cuts in the event selection. It can be extrapolated to the full phase space with the detector acceptance A, the fraction of the signal events produced in the fiducial phase space. Then the full cross section is given by:

$$\sigma^{\mathrm{fid}}_{\mathrm{t\bar{t}}}(\ell\tau_{\mathrm{h}}) = \frac{N - B}{\mathcal{L} \cdot \mathcal{B} \cdot \epsilon \cdot A} \tag{5.9}$$

The acceptance is measured in the signal simulation.

The misidentification probability uncertainty enters the cross section measurement only in the estimation of the number of background events B. The relative effect of this uncertainty on the signal measurement is $\delta B_{\mathrm{mis}-\tau_{\mathrm{h}}}/(N-B)$, where $\delta B_{\mathrm{mis}-\tau_{\mathrm{h}}}$ is the uncertainty in the number of events with misidentified τ_{h}. From the uncertainties estimated in Table 5.5 and the event yields in Tables 5.1 and 5.2, the estimation of the misidentified τ_{h} background contributes about 5-6% of the uncertainty in the cross section measurement.

The main sources of this uncertainty are the large variability between the misidentified probabilities measured in different control samples, and the statistical uncertainty in the control samples, specifically in the number of misidentified τ_{h} candidates. The statistical uncertainty in the measured probabilities becomes a significant issue because the parameter space of the measurement is large and requires many events to fill: three kinematic parameters (p_{T}, η, and r_{jet} of the jet) and three jet flavours.

5.1.4 Discussion

The direct measurement of the τ_{h} misidentification probabilities is a general method to estimate the background with misidentified τ_{h} in any event selection. It provides an approximation result with a systematic uncertainty that originates from the fundamental difficulty of the issue.

The crucial part of the method is the choice of the jet parameters to characterize the misidentification probabilities. The presented measurements include only basic parameters: the flavour, the kinematic parameters, and a geometric parameter r of the jet. However, jets have a complex structure with more characteristics that can significantly affect the τ_{h} misidentification probability. For example, the presented model of the jet-to-tau misidentification probability could be improved by adding the c quark flavour of jets and the τ_{h} decay modes, at least the number of charged particles in the τ_{h} candidates.

To account for the incomplete set of parameters in the misidentification probabilities, the estimation of the systematic uncertainty must be conservative. The worst case scenario would be to use the diametrical difference in the measured probabilities, i.e. the difference between the largest and the smallest misidentification probability. The systematic uncertainty can be reduced by expanding the τ_{h} misidentification model to cover all jet parameters. But additional jet parameters will increase the measurement phase space for the τ_{h} misidentification probabilities beyond the available number of events in the control samples, or the number of control samples themselves. An addition of another jet flavour, such as the c quark flavour, would require a new control region. These limitations make an expansion of the method to more parameters unfeasible on practice.

The τ_{h} misidentification uncertainty is dominant in the $\mathrm{t\bar{t}}$ cross section measurement. Two approaches to improve the estimation of the misidentified τ_h background are discussed in Section 5.2, including the method of the "shape fit" that is used in the cross section measurement in Chapter 6.

5.2 The fake factor and shape fit methods

Jets are complex objects that are characterized by many parameters. The τ_h itself is a particular kind of jet that includes significantly different decay modes and a number of other parameters. This complexity presents a fundamental difficulty for the modelling and the estimation of the probability to misidentify a jet as a τ_h, as discussed in the direct measurement of the misidentification probabilities in Section 5.1.

Two approaches to improve the estimation of the contribution from the background with misidentified τ_h are considered in the following. The "fake factor" method reduces the dependence on jet parameters by employing a ratio between the probabilities for a jet to pass two different working points of the τ_h identification algorithm [106**?**]. Such "differential" approach accounts for some of the features of the identification and makes the method more robust with respect to the jet characteristics. Another approach is to constrain the contributions from the background processes with a shape fit to τ_h-independent parameters in the event, instead of estimating the misidentification probability for particular jets.

In the fake factor method a given control region with a negligible fraction of genuine τ_h (for example, any of the control regions from Section 5.1) is separated in two categories that do not overlap: the "tight" category T where the τ_h candidates pass the tight working point of the identification requirement, and the "loose" category L where they pass only the loose working point and do not pass the tight. The loose-to-tight ratio between the misidentification probabilities $\epsilon_{T/L}$ in these control regions is calculated:

$$\epsilon_{T/L} = \frac{\omega_T(\mathrm{jet} \to \tau_\mathrm{h})}{\omega_L(\mathrm{jet} \to \tau_\mathrm{h})} = \frac{N_T(\mathrm{jet} \to \tau_\mathrm{h})}{N_L(\mathrm{jet} \to \tau_\mathrm{h})}. \tag{5.10}$$

The same two categories are constructed in the signal region. The signal is measured in the T category, and the L category serves as a statistically independent side region to estimate the misidentified background. The number of events with misidentified τ_h in the T category is estimated as:

$$N_T(\mathrm{mis}{-}\tau_\mathrm{h}) = \epsilon_{T/L} N_L(\mathrm{mis}{-}\tau_\mathrm{h}) \tag{5.11}$$

$$N_L(\mathrm{mis}{-}\tau_\mathrm{h}) = N_L(\mathrm{obs.}) - N_L(\tau_\mathrm{h}). \tag{5.12}$$

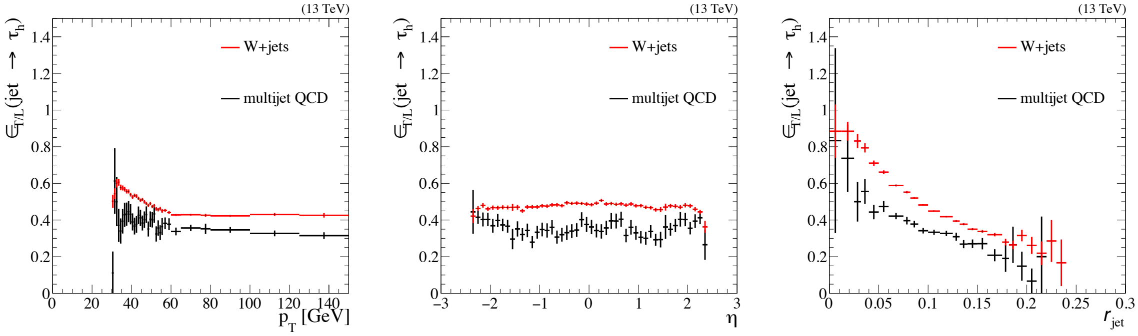

Figure 5.6: Fake factors measured in the multijet QCD and W+jets selections.

Where N_L(obs.) are all observed events in the L category, and $N_L(\tau_\mathrm{h})$ is the number of events with genuine τ_h taken from the simulation. The number of signal events in the L category is included with a large variation that is propagated as a systematic uncertainty into the T category in order to account for the unknown cross section of the signal. The magnitude of the variation must cover the overall uncertainty in the final signal measurement. Since the τ_h candidates in the L category do not pass the tight working point of the identification, the contribution of signal events in this category is small and the corresponding uncertainty is negligible.

The advantage of the method comes from a good separation of the signal and background in the orthogonal τ_h identification requirements: most of genuine τ_h pass the tight identification requirements, the misidentified candidates dominate the loose-only selection. The method also assumes that, whatever the underlying nature of the misidentified objects is, the $\epsilon_{T/L}$ does not change significantly, since it corresponds to the difference between working points of the same identification algorithm. Figure 5.2 demonstrates that the difference between the fake factors measured in the W+jets and multijet QCD control regions is only about 20%.

In order to verify the validity of the method and improve the systematic uncertainty, the same estimation can be performed from additional control regions constructed by inverting some of the selection requirements that are relevant for the signal process. For example a cut on the $p_\mathrm{T}^\mathrm{miss}$ can be included, since the $\mathrm{t\bar{t}} \to \ell\nu_\ell\tau_\mathrm{h}\nu_\tau\mathrm{b\bar{b}}$ has more neutrinos than the main background $\mathrm{t\bar{t}} \to \ell\nu_\ell\mathrm{q\bar{q}'b\bar{b}}$. The main requirement for the side regions is to have the same composition of jets as in the signal region.

In order to account for the difference in the jet composition the $\epsilon_{T/L}$ ratio can be defined for different jet flavours:

$$\omega_T(\text{jet} \to \tau_\mathrm{h}) = \rho_T(\mathrm{q}) \times \%_\mathrm{q} + \rho_T(\mathrm{g}) \times \%_\mathrm{g} \tag{5.13}$$
$$= \epsilon_{T/L}(\mathrm{q})\rho_L(\mathrm{q}) \times \%_\mathrm{q} + \epsilon_{T/L}(\mathrm{g})\rho_L(\mathrm{g}) \times \%_\mathrm{g} \tag{5.14}$$
$$= \bar{\epsilon}_{T/L}\omega_L(\text{jet} \to \tau_\mathrm{h}). \tag{5.15}$$

Where $\epsilon_{T/L}(\{\mathrm{q},\mathrm{g}\})$ are the loose-to-tight ratios for jets of quark and gluon flavours, $\omega_{\{L,T\}}$ are the misidentification probabilities in the selection regions, $\rho_{\{L,T\}}(\{\mathrm{q},\mathrm{g}\})$ are the misidentification probabilities of specific jet flavours, $\%_{\{\mathrm{q},\mathrm{g}\}}$ are the fractions of jets in a given control region, and $\bar{\epsilon}_{T/L}$ is the loose-to-tight ratio averaged over the jet flavours.

The $\bar{\epsilon}_{T/L}$ is a hyperbolic function of jet flavour fractions:

$$\bar{\epsilon}_{T/L}(\%_\mathrm{q}/\%_\mathrm{g}) = \frac{\epsilon_{T/L}(\mathrm{q})\rho_L(\mathrm{q}) \times \%_\mathrm{q} + \epsilon_{T/L}(\mathrm{g})\rho_L(\mathrm{g}) \times \%_\mathrm{g}}{\rho_L(\mathrm{q}) \times \%_\mathrm{q} + \rho_L(\mathrm{g}) \times \%_\mathrm{g}} \tag{5.16}$$
$$= \frac{\epsilon_{T/L}(\mathrm{q})\frac{\rho_L(\mathrm{q})}{\rho_L(\mathrm{g})} \times \frac{\%_\mathrm{q}}{\%_\mathrm{g}} + \epsilon_{T/L}(\mathrm{g})}{\frac{\rho_L(\mathrm{q})}{\rho_L(\mathrm{g})} \times \frac{\%_\mathrm{q}}{\%_\mathrm{g}} + 1}. \tag{5.17}$$

It can be fitted in three contreol regions to determine the parameteres $\epsilon_{T/L}(\mathrm{q})$, $\epsilon_{T/L}(\mathrm{g})$, $\rho_L(\mathrm{q})/\rho_L(\mathrm{g})$. Then the number of events with misidentified τ_h in the signal region is estimated with the averaged $\bar{\epsilon}_{T/L}$:

$$N_T(\text{mis}-\tau_\mathrm{h}) = \bar{\epsilon}_{T/L}(\%_\mathrm{q}/\%_\mathrm{g})N_L(\text{mis}-\tau_\mathrm{h}) \tag{5.18}$$

Where $N_L(\text{mis}-\tau_\mathrm{h})$ is the number of events with misidentified τ_h in the L signal region of the observed data, and $\%_\mathrm{q}/\%_\mathrm{g}$ is the fraction of the quark flavour over the gluon flavour jets in the pre-selection obtained from the simulation.

This procedure can easily include kinematic and other parameters of jets, such as p_T, η, r_jet, etc. The jet flavour is different because it cannot be reconstructed with a certainty, and it must be accounted for indirectly in a fit to data. With more flavours Eq. 5.13 becomes more involved and the determination of the $\bar{\epsilon}_{T/L}$ function requires more control regions. Usually $\bar{\epsilon}_{T/L}$ does not depend significantly on the jet flavour. In such cases, the small flavour dependence is simply taken as a systematic uncertainty.

The “differential” approach of the fake factor method makes it resilient to the dependance of the τ_h misidentification probability on the jet flavour and other parameters. However, it does not eliminate the fundamental dependence on the large space of jet parameters. And the uncertainty in the $\bar{\epsilon}_{T/L}$ due to jet flavours can still be too large. Also the statistical fluctuations that correspond to the number of the τ_{h} candidates in the L category propagate as an uncertainty in the denominator of the $\epsilon_{T/L}$. The equivalent uncertainty in the direct measurement of the misidentification probability was negligible, because the denominator in the probability corresponds to a large number of all jets in the selected events.

The issue of the complex jet structure affecting the τ_{h} misidentification probability can be circumvented by considering whole physical processes that produce misidentified τ_{h}, instead of modelling the misidentification of particular jets in the selected events. This way the flavour composition and other parameters of the jets are embedded in the background estimation automatically. The relevant misidentification probability can be measured directly in a region enriched with events from the same background processes as present in the signal selection. Or the misidentification can be constrained in a shape fit to a distribution that distinguishes the signal and the backgrounds, but that is not affected by the τ_{h} candidates. In the case of the $\mathrm{t\bar{t}} \to \ell\nu_{\ell}\tau_{\mathrm{h}}\nu_{\tau}\mathrm{b\bar{b}}$ final state, the main background $\mathrm{t\bar{t}} \to \ell\nu_{\ell}\mathrm{q\bar{q}'b\bar{b}}$ can be distinguished by the mass constraints of the W boson and the top quark in the fully hadronic branch of the decay. Also the distribution of missing p_{T} is different in the signal and the background because of the additional ν_{τ} neutrino in the dilepton final state. Section 6 develops this approach and successfully employs it to significantly reduce the uncertainty in the measurement of the $\mathrm{t\bar{t}}$ production cross section due to the misidentified τ_{h} background.

Chapter 6

The measurements in the top quark pair dilepton final states containing a tau lepton

This chapter presents three measurements in the $\mathrm{t\bar{t}}$ dilepton final state that includes a τ lepton, $\mathrm{t\bar{t}} \rightarrow \ell\nu_{\ell}\tau\nu_{\tau}\mathrm{b\bar{b}}$ where ℓ represents a light lepton, an electron or a muon (e or μ). The $\mathrm{t\bar{t}}$ production cross section $\sigma_{\mathrm{t\bar{t}}}(\ell\tau)$ is measured in the fiducial phase space of the detector and extrapolated to the full phase space. Also, the ratio of the cross sections in the $\ell\tau$ and light dilepton [2] final states $\sigma_{\mathrm{t\bar{t}}}(\ell\tau)/\sigma_{\mathrm{t\bar{t}}}(\ell\ell)$, and the ratio of the partial to the total decay width of the top quark $\Gamma(\mathrm{t} \rightarrow \tau\nu_{\tau}\mathrm{b})/\Gamma_{\mathrm{total}}$ are evaluated.

The analysis in this chapter relies on the information presented previously. In order to make the text more explicit for the reader, some information is repeated. For example, most of the event selection requirements are the same as in Chapter 5.

The measurements are performed in the data sample of 13 TeV pp collisions that was collected in 2016 with the CMS detector at the LHC and corresponds to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$. The observed data, the simulation, triggers, reconstruction tools, and the final state object definitions are described in Chapter 4.

This is the first measurement of the $\mathrm{t\bar{t}}$ production cross section in 13 TeV collisions at the LHC that explicitly includes τ leptons. It improves over the results obtained at 7 and 8 TeV [18, 19] owing to the significantly reduced uncertainty in the estimation of the misidentified τ_{h} background. The improved precision propagates to the measurements of the ratio to the light dilepton final state cross section and the partial decay width of the top quark.

The measurement strategy is presented in Section 6.1. The event selection for the study is given in Section 6.2. Section 6.3 describes the estimation of backgrounds with the corresponding corrections, whether they are taken from the simulation or estimated from the data. The kinematic distributions of the selected events are presented in Section 6.4. Sources of systematic uncertainties are described in Section 6.5. Section 6.6 covers the profile likelihood ratio (PLR) method which is used for the cross section measurement and the estimation of the uncertainties. The measurements of the cross sections, the ratio to the light dilepton cross section, and the ratio of the partial to the total top quark decay width are discussed in Section 6.7. The results are summarized in Section 6.8.

6.1 Analysis strategy

The τ leptons decay before reaching the detector, and experimentally they are identified by their decay products, light leptons or hadrons. The light leptons from the τ decay can be distinguished from the prompt leptons from the hard process under special assumptions about the kinematic distribution or other parameters of the signal. In order to avoid a bias to the signal model the presented analyses employ only the τ leptons that decay into hadrons (τ_h). The resulting measurements are extrapolated to all τ leptons with the branching fractions of the τ decay that have been measured to a high precision [4]. The hadrons from the τ lepton decay are reconstructed as a narrow jet with a distinct signature in the detector. The dominant background arises from events where a jet is misidentified as a τ_h. For the $\mathrm{t\bar{t}}$ dilepton final state it is mostly the semileptonic $\mathrm{t\bar{t}} \to (\ell\nu_\ell)(\mathrm{q\bar{q}'})\mathrm{b\bar{b}}$ process.

The background with the misidentified τ_h can be constrained with a direct measurement of the background in a control region enriched in background events, or with a shape fit to a distribution that distinguishes signal and background. In both cases the estimation of the misidentified background contribution does not depend on the particular properties of jets and τ_h candidates in the selected events. The background is estimated for the whole physical process that produces misidentified τ_h candidates. Therefore, the problem of the large space of jet parameters is avoided.

The cross section is measured by performing a profile likelihood ratio (PLR) fit [20] to the transverse mass of the system containing the light lepton (e or μ) and the missing transverse momentum, in two kinematic categories of the selected events for each of the $\mathrm{e}\tau_\mathrm{h}$ and $\mu\tau_\mathrm{h}$ final states. The shape of the transverse mass distribution distinguishes the signal and background processes. The kinematic categories are constructed to separate the signal and the main background process, the semileptonic $\mathrm{t\bar{t}}$ final state, by employing the constraints of the top quark and W boson masses in the hadronic branch of the decay. The PLR fit to the transverse mass and the kinematic categories constrain the misidentified τ_{h} background without any special assumptions about the signal model. Section 6.3.5 demonstrates that the transverse mass does not depend significantly on the τ_{h} candidate parameters in the event. And the event categories are not affected by the τ_{h} identification requirements, because they are constructed from the general information about the jets in the events. Therefore, the two approaches cross-check each other, because they are not correlated.

6.2 Event Selection

For the $\mathrm{e}\tau_{\mathrm{h}}$ ($\mu\tau_{\mathrm{h}}$) final state, data are collected with a trigger requiring at least one isolated electron (muon) with a threshold of $p_{\mathrm{T}} > 27\ (24)\,\mathrm{GeV}$. The recorded events are reprocessed offline to reconstruct physical objects with the algorithms and techniques described in Chapter 4.

The events for the measurement are selected by requiring one isolated electron (muon) with transverse momentum $p_{\mathrm{T}} > 30\ (26)\,\mathrm{GeV}$ and $|\eta| < 2.4$, at least two jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 2.5$, and exactly one τ_{h} candidate with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$. The τ_{h} candidate and the light lepton are required to have opposite electric charges (OC). At least one jet must be identified as originating from b quark hadronization ("b-tagged") by passing the "medium" working point of the "CSVv2" b flavour discriminator [98], that provides the efficiency of about 66% with the corresponding light-flavour misidentification rate of 1%. The τ_{h} candidate is required to pass the "medium" working point of the MVA-based anti-jet discriminator, that provides the signal efficiency of 60% with less than 1% misidentification probability for a generic jet [3]. The b-tagged jets are required to be separated by $\Delta R > 0.4$ from the τ_{h} jet. Electrons and muons are required to be separated from any jet and from the τ_{h} candidate in the η-φ plane by a distance of $\Delta R > 0.4$. Events with any additional loosely isolated electron (muon) of $p_{\mathrm{T}} > 15\ (10)\,\mathrm{GeV}$ are rejected. An electron is considered loosely isolated if $I_{\mathrm{rel}} < 0.0994$ in the barrel or $I_{\mathrm{rel}} < 0.107$ in the endcaps. A muon is loosely isolated if $I_{\mathrm{rel}} < 0.25$ in either the barrel or the endcaps.

These selection requirements are motivated by the kinematic parameters of the $\mathrm{t\bar{t}}$ decay, available triggers, and features of the detector. The "medium" working points of the identification algorithm were chosen as a good compromise to study the tighter and looser options. But the looser option shows no improvement. And the event selection with "tight" working points has issues in the modelling of W+jets and $\mathrm{t\bar{t}}$ backgrounds because of insufficient number of simulated events. Since the main advantage from the "tight" working point is a reduction of the misidentified τ_{h} background, and since this background is already well constrained with our method, the analysis employs the "medium" working points to avoid the issues with the simulation statistics.

The selected events exhibit a good agreement between the observed data and the expectation, as shown in Figure 6.1 for the p_{T} distribution of the τ_{h} candidate. The signal contributes about 60% of the selected events in either of the dilepton final states. The dominant background contribution comes from other $\mathrm{t\bar{t}}$ decays, mostly from lepton+jets final states where a jet is misidentified as a τ_{h} candidate.

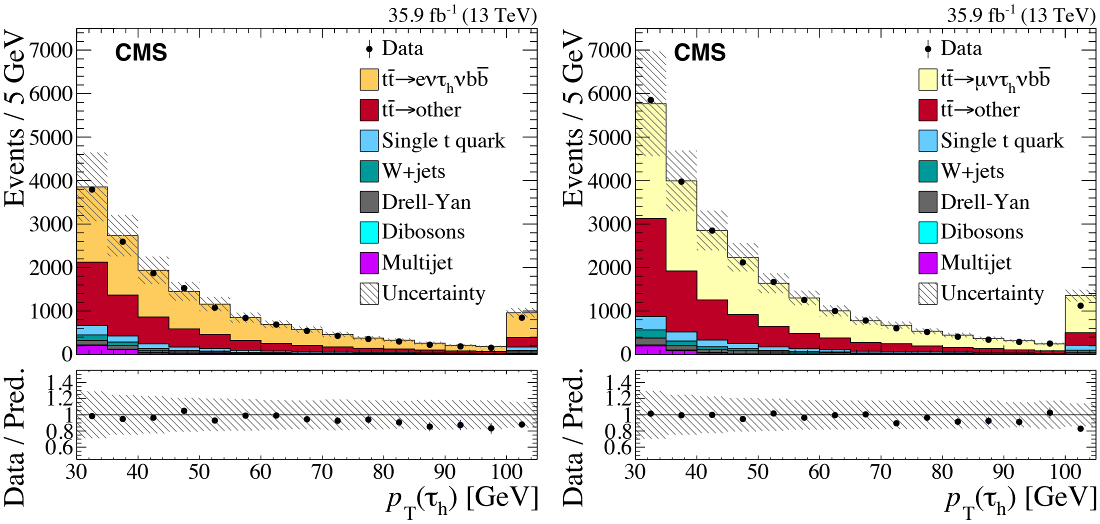

Figure 6.1: The τ_{h} p_{T} distributions for events of the $\mathrm{e}\tau_{\mathrm{h}}$ (left) and $\mu\tau_{\mathrm{h}}$ (right) final states observed prior to fitting. Distributions obtained from data (filled circles) are compared with simulation (shaded histograms). The last bin includes overflow events. The simulated contributions are normalized to the cross section values predicted in the SM. The main processes are shown: the signal, the other $\mathrm{t\bar{t}}$ processes grouped together, single top quark production, W+jets, DY processes, diboson, and multijet production. The ratio of the data to the total SM prediction is shown in the lower panel. The hatched bands indicate the systematic uncertainties and the statistical uncertainties of all simulated samples. Statistical uncertainties on the data points are not visible because of the scale of the figure.

In order to discriminate against the main background of misidentified τ_{h} from the $\mathrm{t\bar{t}}$ lepton+jets process, the constraints from the top quark and W boson masses in the $\mathrm{t} \to \mathrm{bW} \to \mathrm{b(q\bar{q}')}$ decay are used. Two event categories are defined according to the kinematic properties of jets in the event. Jet triplets are constructed for each combination of one b-tagged jet and two untagged jets, chosen from all jets in the event, including the τ_h candidate. The distance parameter for each triplet is calculated as

$$D_\mathrm{jjb} = \sqrt{(m_\mathrm{W} - m_\mathrm{jj})^2 + (m_\mathrm{t} - m_\mathrm{jjb})^2}, \tag{6.1}$$

where $m_\mathrm{t} = 172.5\,\mathrm{GeV}$ and $m_\mathrm{W} = 80.385\,\mathrm{GeV}$ are, respectively, the masses of the top quark and the W boson [4], m_jj is the invariant mass of the two untagged jets, and m_jjb is the invariant mass of the jet triplet. The event is assigned to the "signal-like" category if there is only one untagged jet, or if the minimum parameter value $D_\mathrm{jjb}^\mathrm{min}$ among all jet triplets is larger than 60 GeV. Otherwise, it is assigned to the "background-like" event category. The threshold of 60 GeV provides an optimal choice between the separation of the signal and background event categories, and the maximization of the yields in the categories to reduce the statistical fluctuations.

6.3 Background estimation

The main background contribution comes from events with one lepton and three or more jets, where one of the jets is falsely identified as a τ_h. It is dominated by the lepton+jets $\mathrm{t\bar{t}}$ process. Misidentified τ_h candidates also come from multijet QCD and W+jet background processes. There is a small contribution from processes with genuine hadronic τ_h: tW single top quark production, $\tau_\ell\tau_\mathrm{h}$ from DY decays, $\mathrm{t\bar{t}} \to \tau_\ell\tau_\mathrm{h}\mathrm{b\bar{b}}$, and diboson processes. All processes, except the multijet QCD, are estimated from simulation after applying appropriate corrections. Background processes contribute roughly 40% of the selected events: 25% is due to the semileptonic $\mathrm{t\bar{t}}$, 5% from the multijet QCD and W+jets, 5% from the single top quark processes, and 5% from other processes with genuine taus. Details of the event selection composition are given in Section 6.4.

6.3.1 Drell-Yan

Drell-Yan background constitutes a small fraction of about 4% in the selected events, most of which are the DY $\to \tau_\ell\tau_\mathrm{h}$ events. This contribution with genuine τ_h is well modelled and is taken from the simulation.

Additionally two control regions are consider in order to confirm the normalization of the Drell-Yan MC and to estimate the uncertainty of the normalization: 1) a simple control region of dilepton DY processes and 2) a control region with DY $\to \tau_\ell\tau_\mathrm{h}$ events.

The first control region is selected by requiring events with two well isolated light leptons (electrons or muons) of opposite charge and $p_\mathrm{T} > 30\,\mathrm{GeV}$, of which at least one matches the single-lepton HLT. The selections essentially contain only DY → $\ell^-\ell^+$ events. The event yields are listed in Table 6.1. The uncertainty of the DY normalization is determined by comparing the simulation to the same selection in data. It includes the statistical and the systematical uncertainties in the lepton identification, luminosity, and pileup corrections.

Figure 6.2 shows the distributions of the leading lepton p_T and the transverse mass m_T of the leading lepton and $p_\mathrm{T}^\mathrm{miss}$ system in the selected events with nominal normalization of DY simulation. The nominal normalization is found to underestimate data by factor of 1.05 ± 0.05. This factor and corresponding uncertainty are applied to the simulation throughout the analysis. The difference in the shape distributions of simulated events and data is due to missing additional corrections in the generated mass and p_T of the DY system. Since the difference is small and the overall DY contribution is small in the main selection, the shape difference is neglected.

Table 6.1: Composition of the DY dilepton control region. Only statistical uncertainties are included.

Process	DY → $\mathrm{e}^-\mathrm{e}^+$	DY → $\mu^-\mu^+$
DY → $\ell^-\ell^+$ + jets	7297917 ± 3303	15090721 ± 4702
DY → $\tau_\ell\tau_\mathrm{h}$ + jets	3765 ± 76	6785 ± 101
$\mathrm{t\bar{t}}$	57084 ± 146	97463 ± 188
Dibosons	17963 ± 39	32620 ± 51
Single top	5787 ± 32	9604 ± 41
W → $\ell\nu$ + jets	1039 ± 118	437 ± 48
QCD multijet	102389 ± 55549	15575 ± 8805
Total	7485944 ± 55647	15253206 ± 9984
Data	7095254	16254370

The second control region is selected by inverting the b tagging requirement of the main $\mathrm{t\bar{t}}$ selection and requiring the transverse mass of the lepton and $p_\mathrm{T}^\mathrm{miss}$ system to be less than 40 GeV. Events must have only one well isolated lepton, no b-tagged jets, one τ_h with medium ID and opposite charge to the lepton, and the transverse mass of the lepton and $p_\mathrm{T}^\mathrm{miss}$ system must be less than 40 GeV to reduce the contribution of W+jets events. Also, as a pre-selection stage, a selection with no requirements on the overall number of jets is considered. In both selection stages the DY → $\tau_\ell\tau_\mathrm{h}$ process contributes large fractions of the selected events. Main backgrounds are multijet QCD and W+jets. The event yields in both electron and muon channels

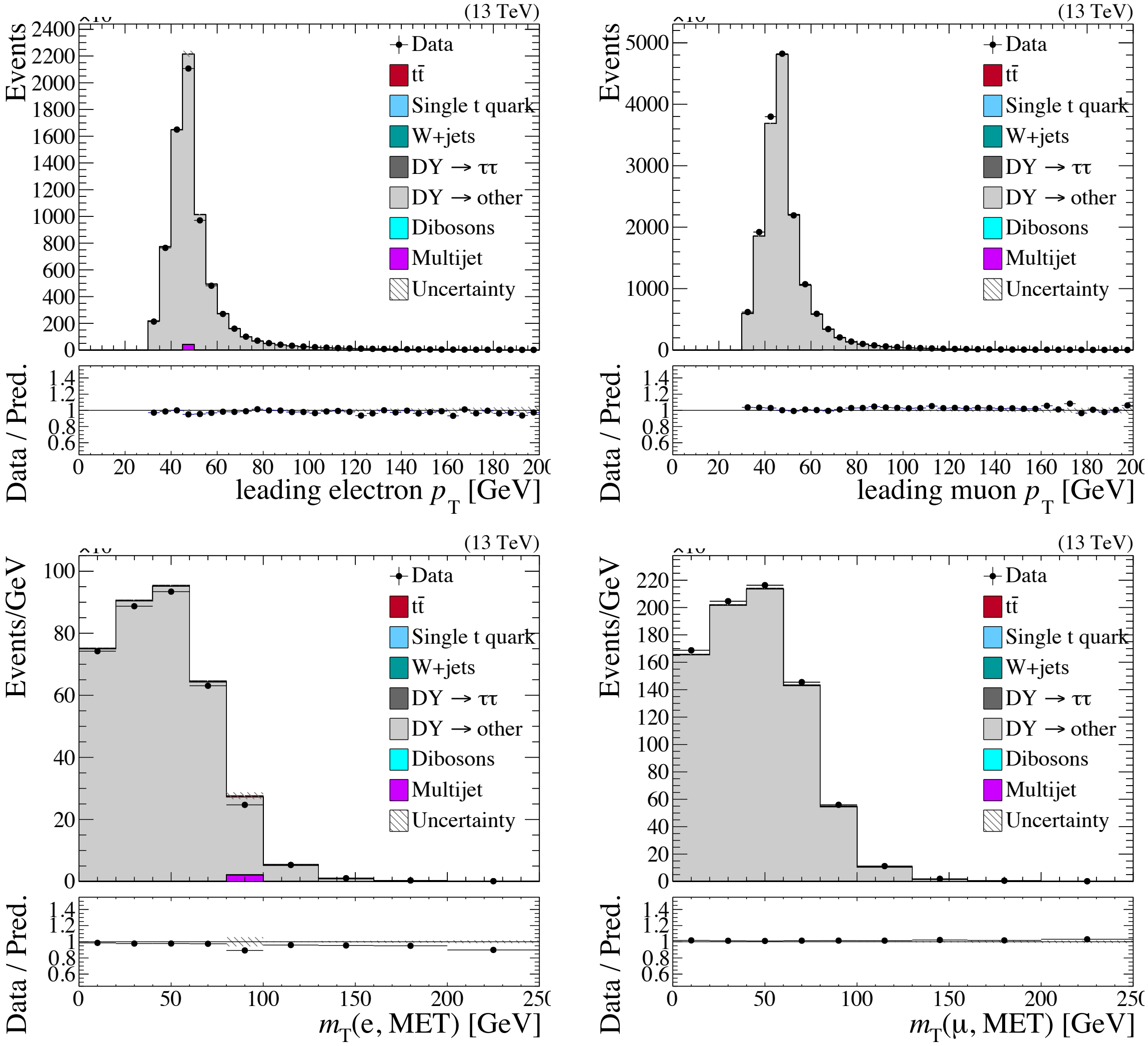

Figure 6.2: Distributions of the leading lepton p_{T} and the transverse mass of the leading lepton and $p_{\mathrm{T}}^{\mathrm{miss}}$ system in the dilepton DY+jets control regions: the e^-e^+ on the left, the $\mu^-\mu^+$ is on the right. The hatched band includes statistical uncertainties.

are given in the Table 6.2. The distributions of the light lepton p_{T} in both selection stages are show in Figure 6.3.

Table 6.2: Composition of the DY → $\tau_\ell\tau_{\mathrm{h}}$ control region. The event yields of MC QCD are given for reference in parentheses. The MC QCD exhibits a large statistical uncertainty, therefore it is not used in the figures for this control region. The statistical uncertainties are given.

Process	pre-DY → $\tau_e\tau_h$	DY → $\tau_e\tau_h$	pre-DY → $\tau_\mu\tau_h$	DY → $\tau_\mu\tau_h$
DY → $\tau_\ell\tau_{\mathrm{h}}$ + jets	31084 ± 220	3136 ± 74	72416 ± 325	4906 ± 88
DY → other	13906 ± 192	1065 ± 45	9485 ± 208	685 ± 38
W → $\ell\nu$ + jets	43304 ± 529	3648 ± 69	72509 ± 705	5344 ± 81
QCD multijet	40777 ± 453	3116 ± 115	32953 ± 537	2592 ± 122
(QCD multijet MC)	(38411 ± 29786)	(185 ± 90)	(16268 ± 6989)	(9188 ± 5333)
$\mathrm{t\bar{t}}$	3286 ± 36	2252 ± 30	4786 ± 43	3261 ± 36
Dibosons	1073 ± 21	301 ± 11	1618 ± 25	421 ± 13
Single top quark	686 ± 11	249 ± 7	1045 ± 13	360 ± 8
Total	134116 ± 756	13767 ± 163	194812 ± 969	17569 ± 179
Data	132629	13456	197473	17721

6.3.2 W+jets

The W+jet events pass the $\mathrm{t\bar{t}}$ selection when one of the jets is misidentified as the τ_{h} candidate. It is a small background process, contributing about 2-3% of the selected events. The process is well modelled, and the contribution is estimated from the simulated events.

As all processes with misidentified τ_{h}, the event yield of the W+jets process is varied by a separate misidentification probability parameter in the PLR fit as described in Section 6.6. Since the fit determines the overall event yield, the main concern in the estimation of the W+jets background is the determination of the m_{T} shape in this process. The modelling of the m_{T} distribution is confirmed in the W+jets control region defined by requiring events with the following requirements: an e or a μ, one τ_{h} candidate of opposite charge to the light lepton and without the anti-jet discriminator requirement of the τ_{h} identification algorithm, no b-tagged jets. The τ_{h} candidate without the anti-jet requirement is included to study the charge distribution of the misidentified τ_{h} with respect to the light lepton in this process.

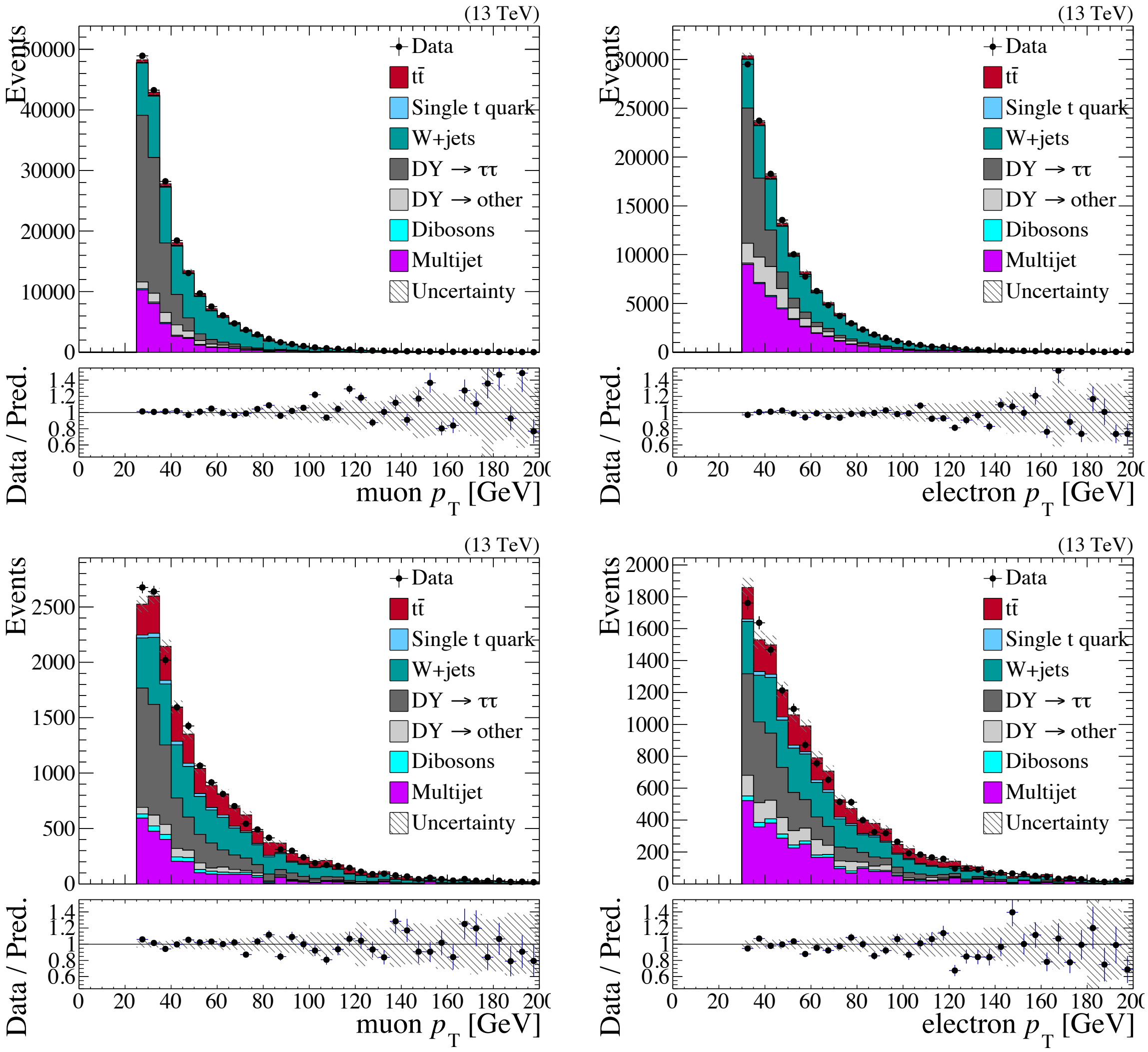

Figure 6.3: Distributions of the lepton p_T and the tau p_T in the DY $\rightarrow \tau_\ell\tau_\mathrm{h}$ control region pre-selection (top) and full selection (bottom): the $\mu\tau_\mathrm{h}$ selection is on the left, the $\mathrm{e}\tau_\mathrm{h}$ is on the right. The uncertainty band includes statistical uncertainties.

The selections mostly contain W+jets and QCD events. The compositions of the event yields are given in Table 6.3. The control distributions in this selection show a good shape agreement between data and MC: Figure 6.4 for the selection with a μ, Figure 6.5 for the selection with an e in the final state. The W+jets contribution is dominant at large values of the lepton p_{T}, $p_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T}.

Table 6.3: Composition of the events in the W+jets control regions with the electron and the muon final states. Event yields for all processes, including QCD, are estimated from MC. Only statistical uncertainties are given.

Process	Electron	Muon
$\mathrm{W} \to \ell\nu + \mathrm{jets}$	4402035 ± 5504	7162457 ± 7376
$\mathrm{W} \to \tau_{\ell}\nu + \mathrm{jets}$	205426 ± 1073	342420 ± 1463
$\mathrm{W} \to \tau_{\mathrm{h}}\nu + \mathrm{jets}$	1756 ± 102	254 ± 44
QCD multijet	1641522 ± 171568	879317 ± 127706
DY + jets	531004 ± 1312	651180 ± 1494
$\mathrm{t\bar{t}}$	165555 ± 260	239504 ± 310
Dibosons	49336 ± 173	71732 ± 210
Single top quark	39701 ± 82	58978 ± 99
Total	7036335 ± 171665	9405842 ± 127936
Data	7048070	9461244

The distribution of W+jets in the LO MC shows a good agreement with the MC@NLO sample generated with NLO accuracy, as shown in Figure 6.6.

6.3.3 Data-driven QCD multijet

The background from the multijet QCD processes is determined from data as it provides a more accurate description with a smaller statistical uncertainty. The method implements the general "ABCD" approach, outlined in Figure 6.7. The shape of the m_{T} distribution is obtained from a sample of events containing lepton and τ_{h} candidates of the same charge (SC). It is estimated by subtracting from the data all other processes taken from the simulation, including the fully hadronic final states in $\mathrm{t\bar{t}}$, single top quark, and dibosons. The normalization of the SC m_{T} distribution obtained in this way is corrected by the OC-to-SC ratio factor, $f_{\mathrm{OC/SC}}$. The ratio is determined in a control region with a relaxed τ_{h} identification and an inverted lepton isolation requirement, where the multijet contribution is dominant. All other event selection requirements remain the same as in the main selection.

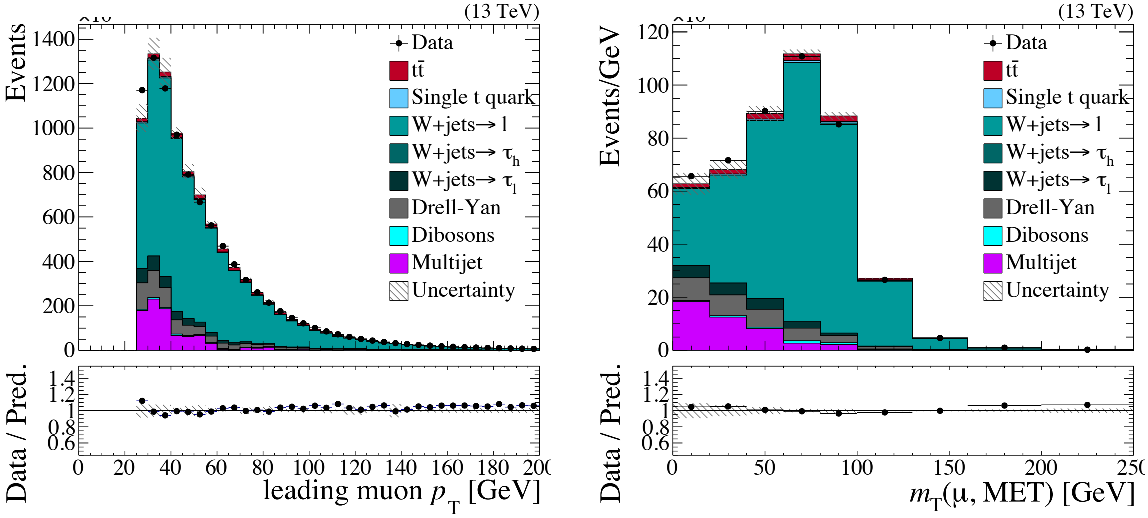

Figure 6.4: Distributions of the lepton p_{T} (left) and the transverse mass m_{T} of the lepton and $p_{\mathrm{T}}^{\mathrm{miss}}$ system (right) in the W+jet control region with a muon in the final state. The QCD multijet contribution is taken from the simulation, like other processes. The hatched band includes the statistical uncertainty, lepton identification, b tagging, pileup reweighting and luminosity.

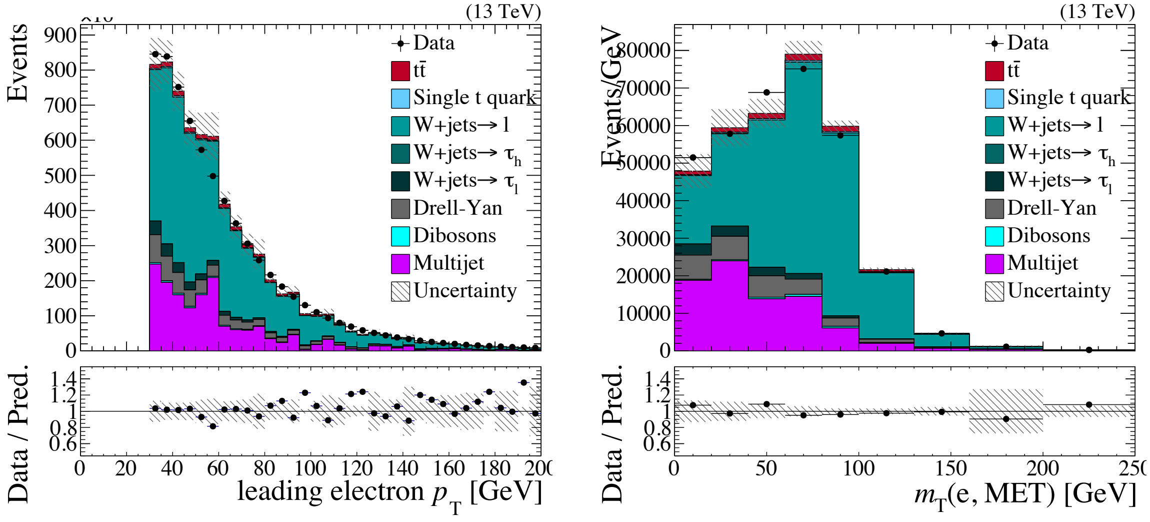

Figure 6.5: Distributions of the lepton p_{T} (left) and the transverse mass m_{T} of the lepton and $p_{\mathrm{T}}^{\mathrm{miss}}$ system (right) in the W+jet control region with an electron in the final state. The QCD multijet contribution is taken from the simulation, like other processes. The hatched band includes the statistical uncertainty, lepton identification, b tagging, pileup reweighting and luminosity.

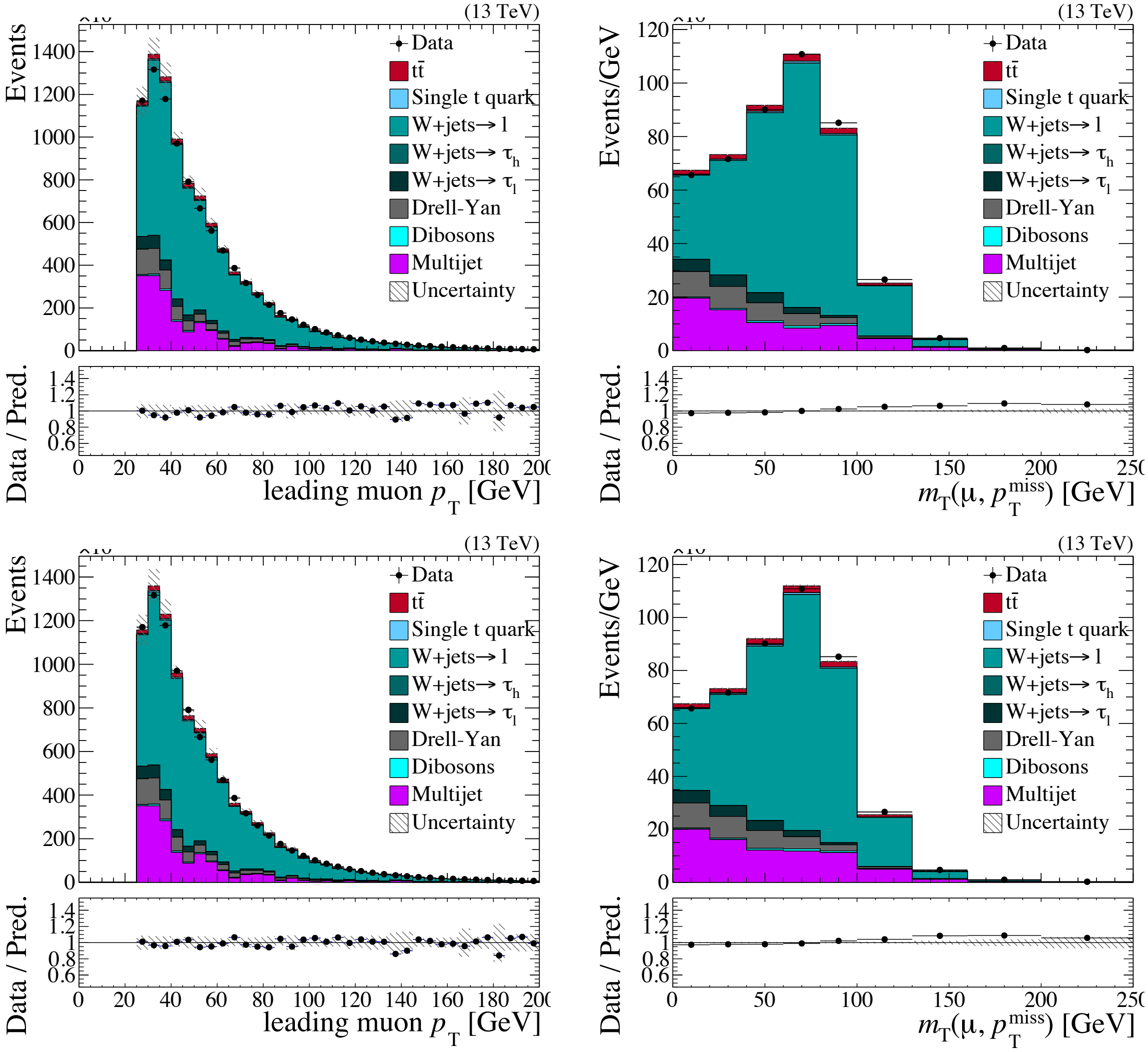

Figure 6.6: Comparison between the distributions of the lepton p_T and the transverse mass m_T in the W+jets events simulated at LO (top) and NLO (bottom) accuracy. The QCD multijet contribution is taken from the simulation, like other processes.

The ratio is measured to be $f_{\mathrm{OC/SC}} = 1.05 \pm 0.05\,(\mathrm{stat} + \mathrm{syst})$, and found to be in agreement with the simulation. Therefore the multijet QCD distribution $m_{\mathrm{T}}^{\mathrm{QCD}}$ is determined as:

$$m_{\mathrm{T}}^{\mathrm{QCD}}\Big|_{\mathrm{OC}} = f_{\mathrm{OC/SC}} \times (m_{\mathrm{T}}^{\mathrm{Data}} - m_{\mathrm{T}}^{\mathrm{otherbkg}})\Big|_{\mathrm{SC}}. \tag{6.2}$$

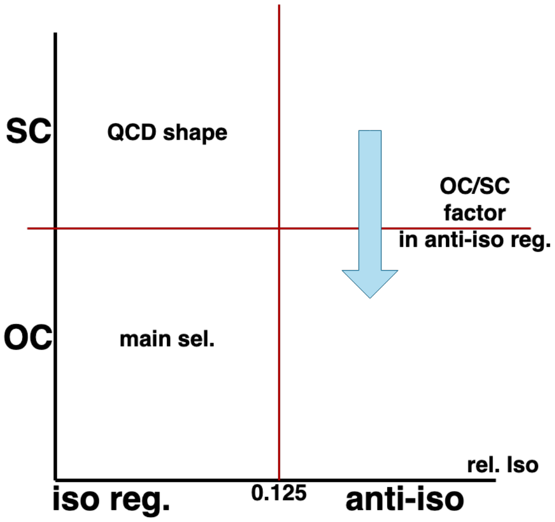

Figure 6.7: The diagram of the “ABCD” method to estimate the shape and the normalization of the multijet QCD background from data.

This method relies on the assumption that the light lepton and the τ_{h} candidate are not correlated in the multijet QCD process. Therefore, this process must produce similar shapes of the m_{T} distribution in the OC and SC selections, and the event yields must be approximately the same.

The assumption about the m_{T} distribution shapes is confirmed in control regions with significant contributions from the multijet QCD process: the $\mathrm{t\bar{t}}$ pre-selection without the requirement of the anti-jet discriminator for the τ_{h} candidate, and the W+jets control region. Figure 6.8 shows a comparison between the m_{T} shapes of the simulated multijet QCD contributions in the OC and SC selections. Figure 6.9 shows the same comparison for the data-driven m_{T} shapes. The figures demonstrate a good agreement between the OC and SC selections.

The ratio between the event yields $f_{\mathrm{OC/SC}}$ is studied in an independent control region that is derived from the pre-selection by inverting the relative isolation requirement for the light lepton: $I_{\mathrm{rel}} > 0.15$ for muons, $I_{\mathrm{rel}}^{\mathrm{barrel}} > 0.0588$ and

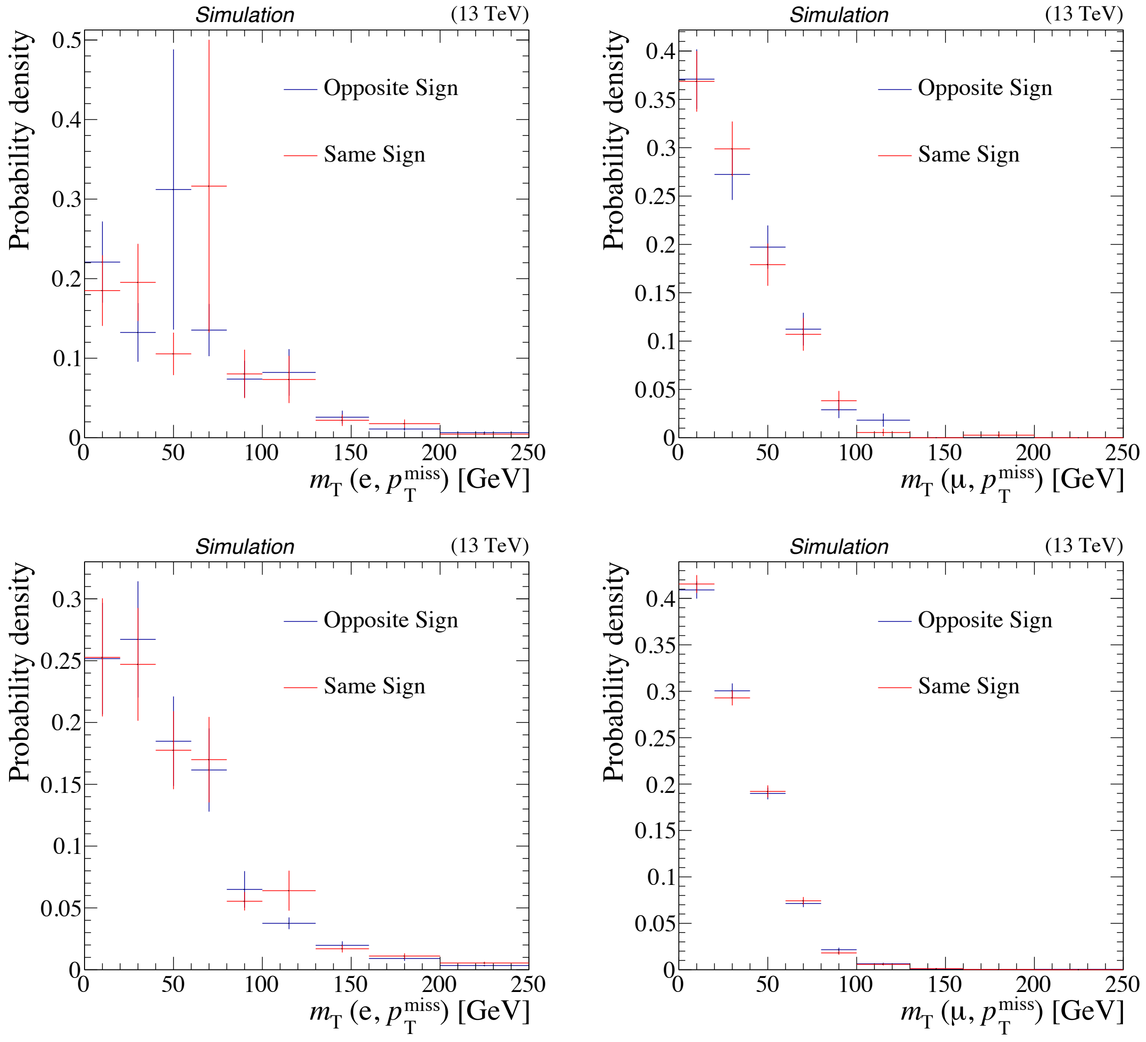

Figure 6.8: Comparison of the m_T distributions in OC and SC selections in the MC QCD: the pre-selection (top) and the W+jets control region (bottom), the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels. The Y axis shows arbitrary units of event yield normalization. The error bars represent statistical uncertainties in the MC datasets.

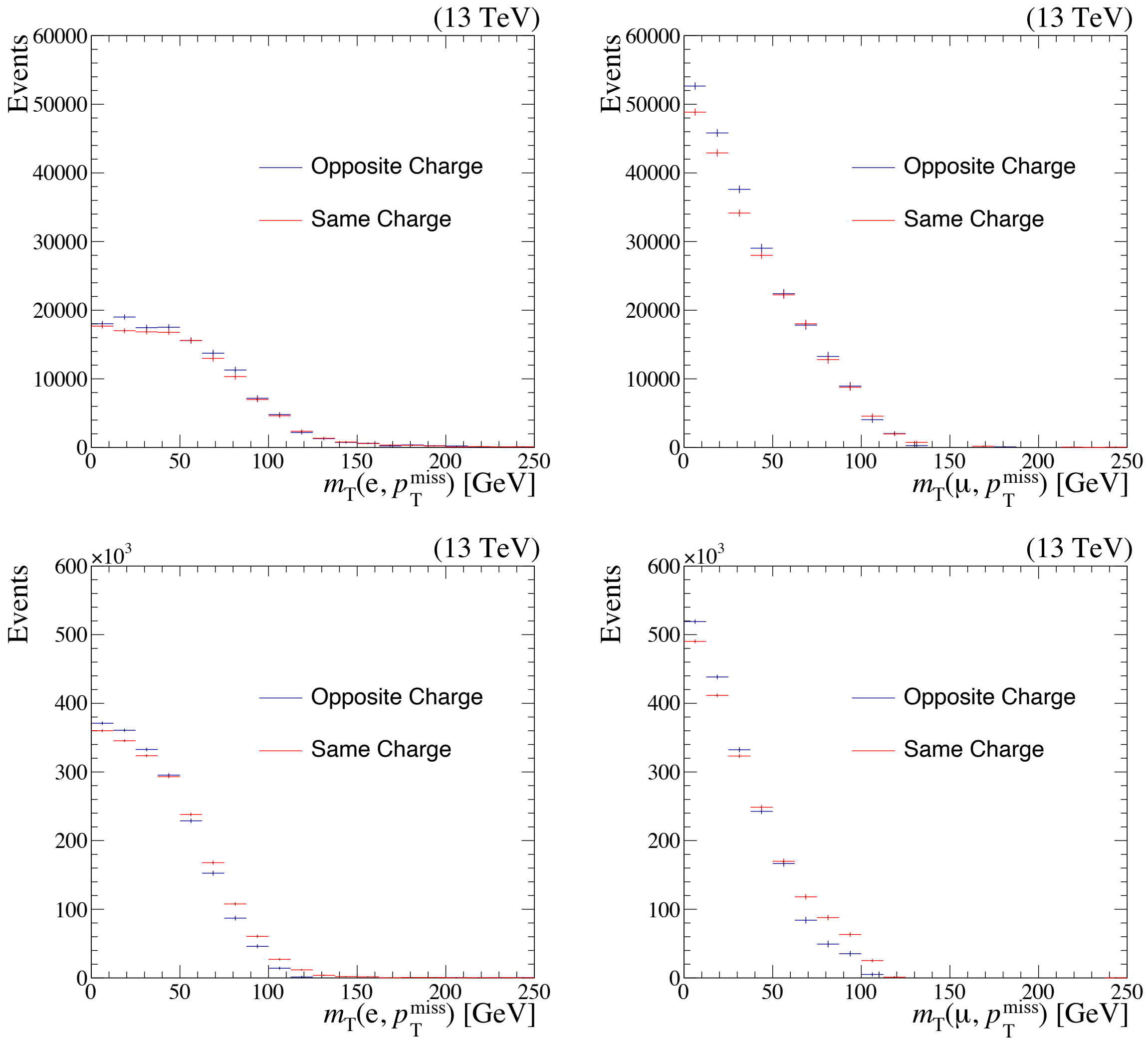

Figure 6.9: Comparison of the m_T distributions in OC and SC selections in the data-driven estimation of the multijet QCD (the difference between the observed data and the non-QCD processes taken from MC): the pre-selection (top) and the W+jet control region (bottom), the eτ_h (left) and $\mu\tau_h$ (right) channels. The Y axis shows arbitrary units of event yield normalization. The error bars represent statistical uncertainties in the MC datasets.

$I_{\mathrm{rel}}^{\mathrm{endcap}} > 0.0571$ for electrons. The relative isolation of the light lepton does not correlate with the charge of the τ_{h} candidate in the multijet QCD process. Therefore, the $f_{\mathrm{OC/SC}}$ ratio is independent from the relative isolation, as verified in Figure 6.11 for the MC QCD and in Figure 6.10 for the data-driven estimation. The $f_{\mathrm{OC/SC}}$ factor from the region with the inverted relative isolation can be extrapolated to the main selection. The corresponding event yields and the calculated resulting factors $f_{\mathrm{OC/SC}} = 1.05 \pm 0.05\,(\mathrm{stat} + \mathrm{syst})$ are given in Table 6.4. The uncertainty in $f_{\mathrm{OC/SC}}$ includes the statistical uncertainty and the systematic uncertainties of lepton identification, luminosity, b tagging, jet energy scale and resolution, and pileup reweighting.

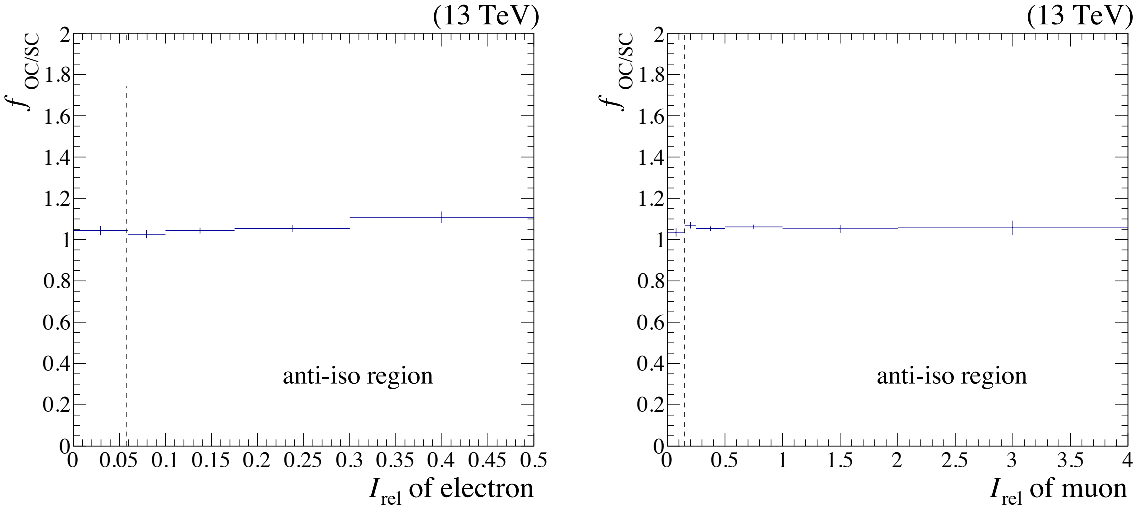

Figure 6.10: Distributions of the OC/SC factor found for different relative isolation requirements in the eτ_{h} (left) and $\mu\tau_{\mathrm{h}}$ (right) pre-selections. Only statistical uncertainties are shown.

Figures 6.13 and 6.12 demonstrate the agreement of the data-driven method and the MC simulated multijet QCD in the pre-selection and W+jets control regions respectively.

As one of the processes with misidentified τ_{h}, the normalization of the multijet contribution is varied in the fit as a separate nuisance parameter, as described in Section 6.5.

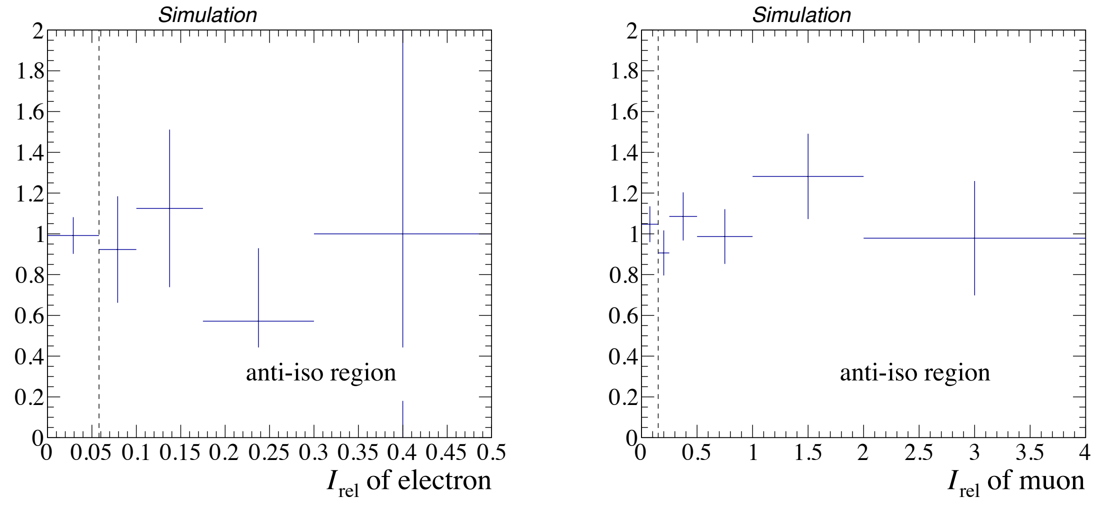

Figure 6.11: Distributions of the OC/SC factor found in the MC QCD for different relative isolation requirements in the $\mathrm{e}\tau_{\mathrm{h}}$ (left) and $\mu\tau_{\mathrm{h}}$ (right) pre-selections. Only statistical uncertainties are shown.

Table 6.4: Event yields for the measurement of the $f_{\mathrm{OC/SC}}$ ratio between the OC and SC contributions of the multijet QCD background.

Process	$\mathrm{e}\tau_{\mathrm{h}}$ anti-iso. presel. OC	 SC	$\mu\tau_{\mathrm{h}}$ anti-iso. presel. OC	 SC
$\mathrm{t\bar{t}} \to \ell\tau\nu\nu\mathrm{b\bar{b}}$	892.0	498.5	2329.5	1350.4
$\mathrm{t\bar{t}} \to \ell\nu\mathrm{q\bar{q}'b\bar{b}}$	11567.3	9300.1	30403.1	24509.9
Other $\mathrm{t\bar{t}}$	1620.7	1491.1	5239.9	4657.1
Single top quark	1037.9	797.1	2658.4	2160.4
DY+jets	452.3	262.5	347.1	330.4
W+jets	2491.2	1476.2	5302.2	4033.4
Dibosons	72.1	67.6	208.0	159.8
Total MC	18133.6	13893.4	46489.2	37202.2
Data	66679.0	60567.0	158153.0	142689.0
Data – MC	48545.4	46673.6	111663.8	105486.8
$f_{\mathrm{OC/SC}}$		1.04		1.06

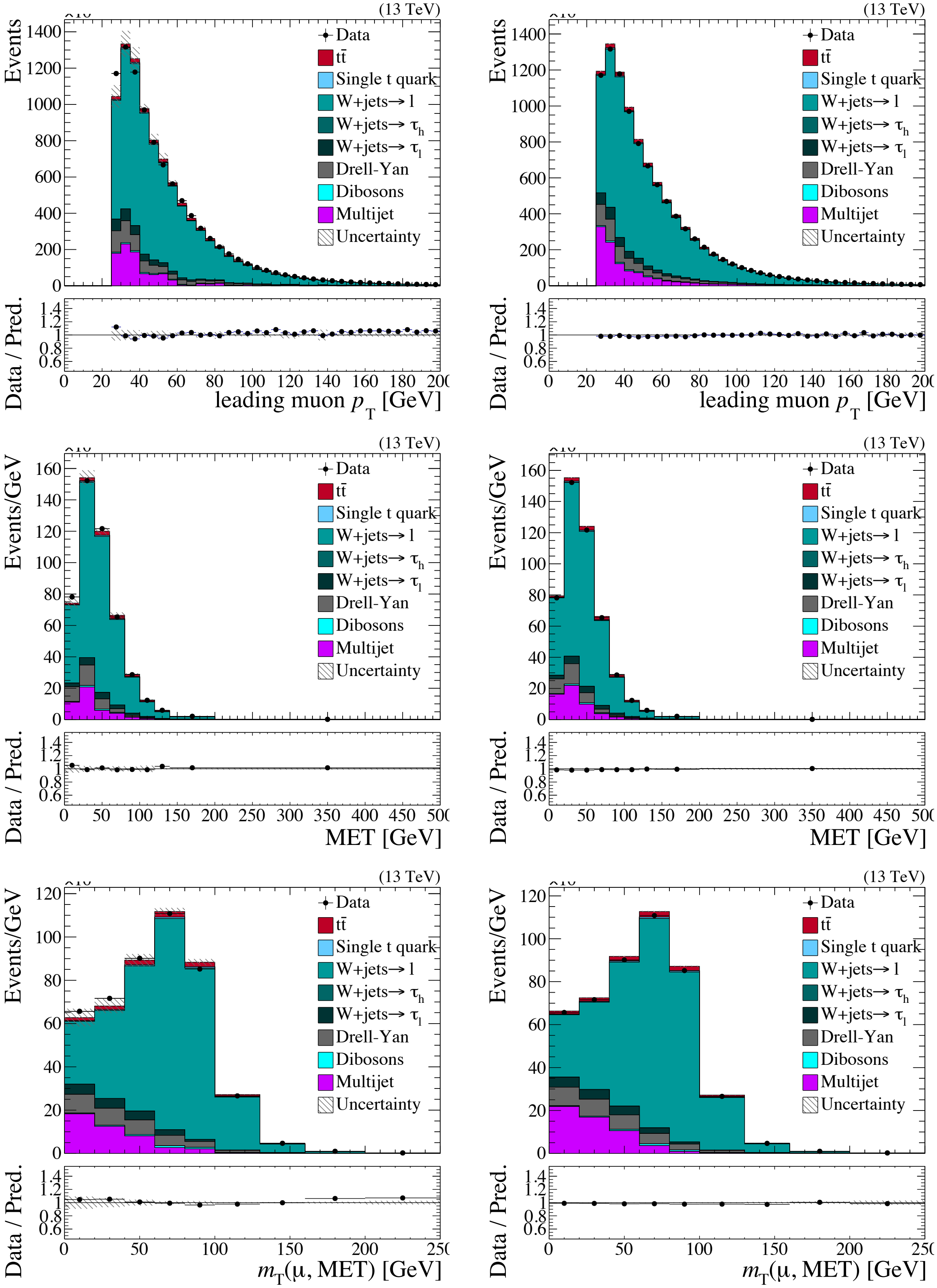

Figure 6.12: Distributions of the muon p_T (top), p_T^{miss} (middle) and m_T (bottom) in the W+jets control region. In the figures on the left the QCD contribution is taken from the simulation, while on the right it is obtained with the data-driven method.

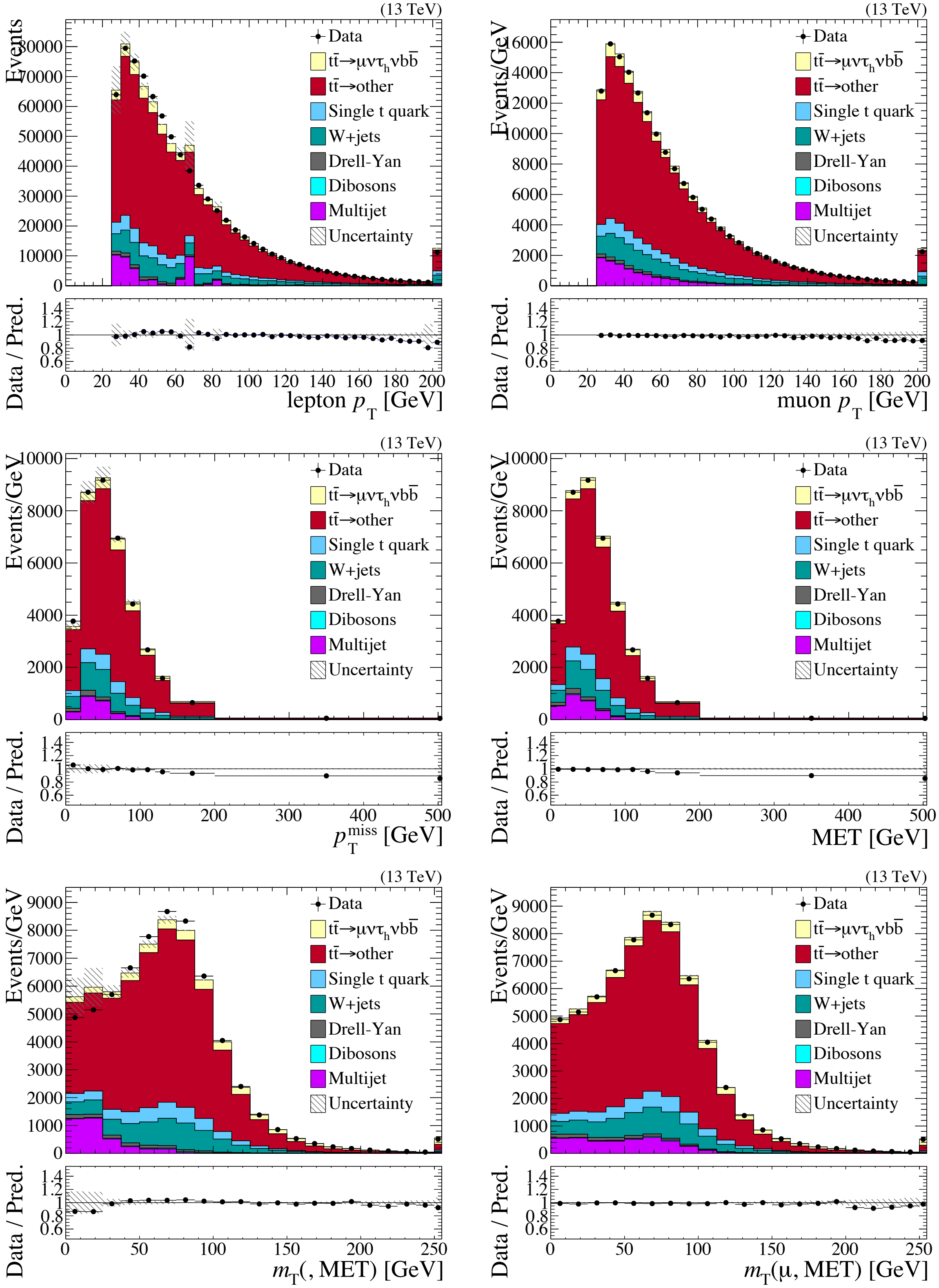

Figure 6.13: Distributions of muon p_{T} (top), $p_{\mathrm{T}}^{\mathrm{miss}}$ (middle) and m_{T} (bottom) in the $\mu\tau_{\mathrm{h}}$ pre-selection. In the figures on the left the QCD contribution is taken from the simulation, while on the right is obtained with the data-driven method.

6.3.4 Single top quark and dibosons

A small background contribution with genuine τ_h comes from the single top quark processes, mostly the tW process, and the diboson processes. These backgrounds are well modelled and estimated from simulation. The shape of the m_T distribution and the normalization uncertainty are verified in the $\mathrm{e}^{\pm}\mu^{\mp}$ control region that is selected by requiring events with one electron and one muon of opposite charges and $p_\mathrm{T} > 30\,\mathrm{GeV}$.

The resulting selection mostly consists of $\mathrm{t\bar{t}}$, Drell-Yan $\tau_\mathrm{e}\tau_\mu$, single-top processes and dibosons. The event yields are listed in Table 6.5. Lepton p_T, $p_\mathrm{T}^\mathrm{miss}$, and m_T distributions are shown in Figure 6.14. Even though the contribution from the single top quark and diboson processes is small, the large number of selected events suppresses statistical fluctuations enough to verify that the overall normalization and the shape of the m_T distribution are well modelled. Taking into account the uncertainties in the lepton identification, luminosity, pileup, and top quark p_T modelling in the control region, the overall uncertainty in the normalization of these backgrounds is estimated to be less than 10%.

Table 6.5: Composition of the $\mathrm{e}^{\pm}\mu^{\mp}$ control region. The statistical uncertainties are given.

Process	Event yield
$\mathrm{t\bar{t}} \to \mathrm{e}^{\pm}\mu^{\mp}\nu_\mathrm{e}\nu_\mu\mathrm{b\bar{b}}$	112889 ± 202
$\mathrm{t\bar{t}}$ other	161 ± 8
Dibosons	12166 ± 49
Single top quark	11410 ± 44
DY+jets	5667 ± 93
W+jets	538 ± 54
Total	142831 ± 239
Data	143096

6.3.5 The misidentified τ_h background

The background processes with misidentified τ_h is the main challenge in the measurement: they contribute a large fraction of the selected events, and they are hard to model or estimate. The jet-to-tau misidentification probability changes significantly with the properties of the selected jets, as seen in Refs. [3, 107] and discussed

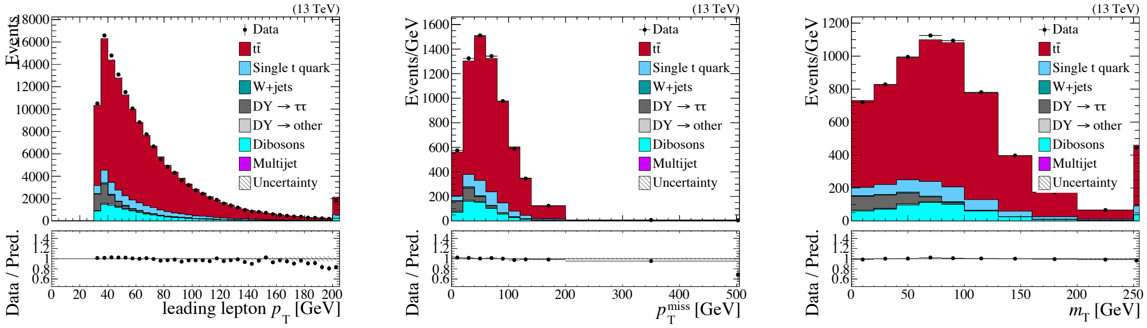

Figure 6.14: Distributions of the leading lepton p_T, $p_\mathrm{T}^\mathrm{miss}$, and m_T in the $\mathrm{e}^{\pm}\mu^{\mp}$ control region. Only statistical uncertainties are shown.

in Chapter 5. Depending on the selection, the probability can vary by up to an order of magnitude. In addition, the simulation can mismodel the dependance of the misidentification probability on the jet p_T and other parameters. Therefore, a correct estimation of this background can be obtained in one of the following ways:

- constrain the background in a shape fit to a distribution that distinguishes signal and background, but does not depend on the p_T or other properties of the τ_h candidate;
- measure the background directly in the selected events by splitting them into background-rich and signal-rich categories.

This analysis implements both approaches, which cross-check each other. The contribution with misidentified τ_h is determined in a fit of the m_T distribution to the observed data. The m_T distribution provides separation between the signal and background processes, as shown in Figure 6.16 on the left. And it does not significantly depend on the p_T and η of the τ_h candidate, or other jet characteristics in the kinematic ranges of this study, as shown in Figure 6.15. Therefore, the fit to this distribution constrains the overall misidentification probability and the event yield from the corresponding background processes. The fit is performed in two event categories, “signal-like” and “background-like”, that provide an additional constraint on the background processes, independent from the details of the m_T distribution.

The event categories are defined in Section 6.2 according to the kinematic properties of jets in the event that expose the constraints of the W boson and the top quark in the $\mathrm{t\bar{t}} \rightarrow \ell\nu_\ell \mathrm{q\bar{q}'b\bar{b}}$ process. For each triplet of one b-tagged and two untagged jets in the event, the distance parameter D_jjb (Eq. 6.1) is calculated that

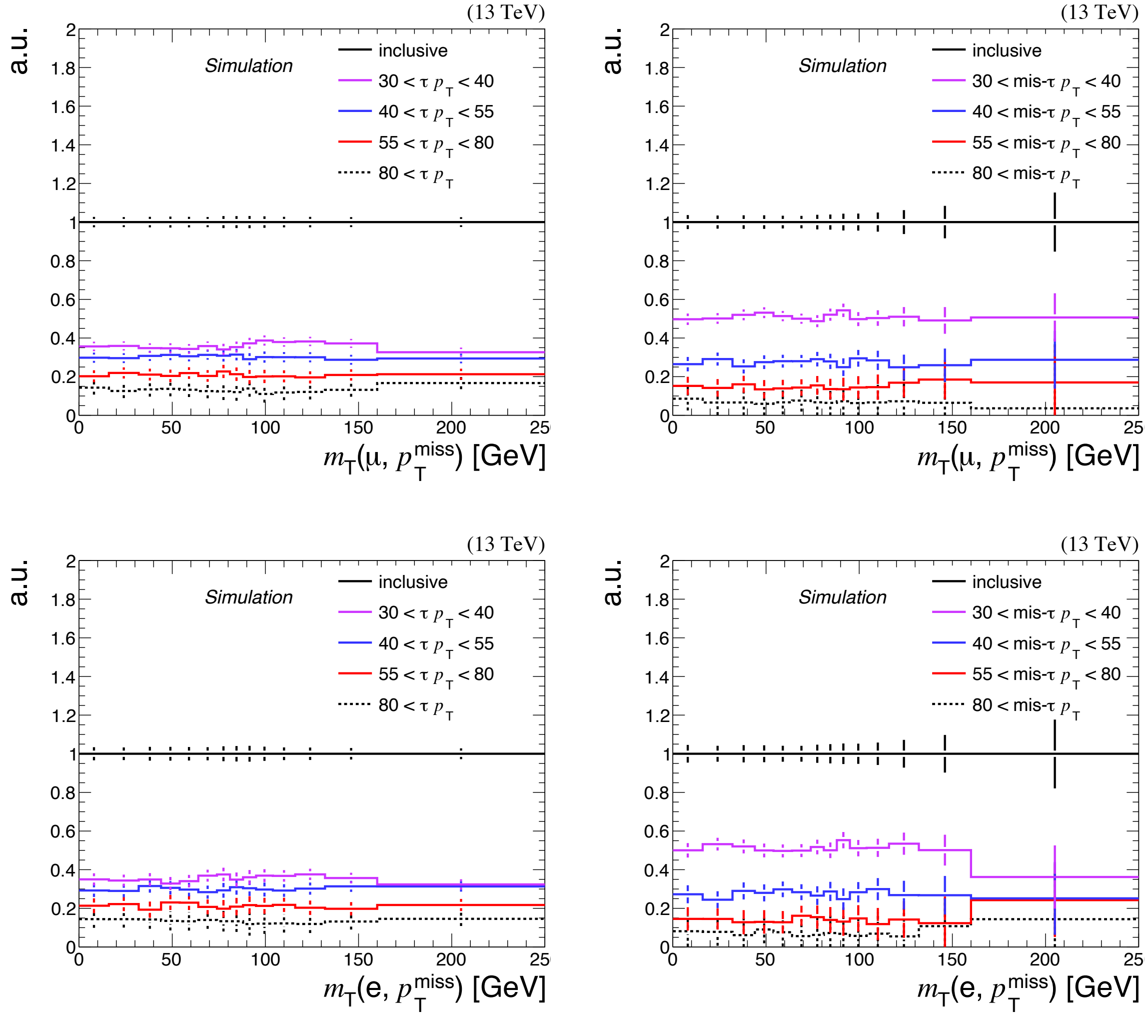

Figure 6.15: The dependence of the m_T on the p_T of τ_h candidates. The distributions show the ratios of the m_T distributions in the events passing specified tau p_T thresholds to the m_T distribution of all events in the muon-tau (top) and electron-tau (bottom) main selections. The ratios for the genuine τ_h are on the left, and the misidentified candidates from $\mathrm{t\bar{t}}$ lepton+jets are on the right. Only statistical uncertainties are included.

separates the signal from the background, as shown in Figure 6.16 on the right. The event is assigned to the “signal-like” category if there is only one untagged jet, or if the minimum parameter value $D_{\mathrm{jjb}}^{\mathrm{min}}$ is larger than 60 GeV. Otherwise, it is assigned to the “background-like” event category.

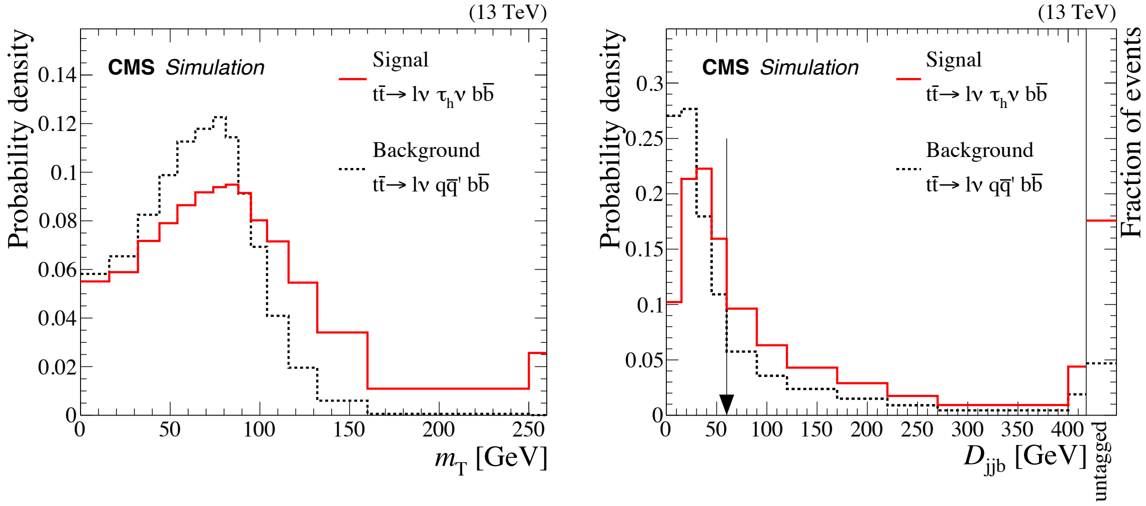

Figure 6.16: Comparison of the signal ($\mathrm{t\bar{t}} \to \ell\nu_\ell\tau_\mathrm{h}\nu_\tau\mathrm{b\bar{b}}$) and the main background of misidentified τ_h ($\mathrm{t\bar{t}} \to \ell\nu_\ell\mathrm{q\bar{q}'b\bar{b}}$) in the shapes of the normalized distributions of the transverse mass m_T between the lepton and $p_\mathrm{T}^\mathrm{miss}$ (left), and the $D_{\mathrm{jjb}}^{\mathrm{min}}$ parameter of the event categories (right), that is derived from the constraints of the W boson and the top quark masses in the $\mathrm{t\bar{t}} \to \ell\nu_\ell\mathrm{q\bar{q}'b\bar{b}}$ process. In the m_T distribution, the signal may extend beyond the W boson mass endpoint because of the two-neutrino final state, whereas the background process cannot. The last bin in both distributions includes overflow events. In the $D_{\mathrm{jjb}}^{\mathrm{min}}$ distribution, the downward arrow points at the threshold of the cut used ($D_{\mathrm{jjb}}^{\mathrm{min}} > 60\,\mathrm{GeV}$), and the panel on the right shows the fraction of events in the “signal-like” category where there is only one untagged jet, which amounts to approximately 5% of all background events and 17% of all signal events.

The simulation models well the kinematic properties of jets that are used in the construction of the event categories. Since the calculation of the di-jet and tri-jet masses uses only kinematic parameters of the jets without any information of τ identification, the procedure can be verified in the pre-selection that is defined exactly like the main selection of events except for the requirement of the anti-jet discriminant for the τ_h candidates. The $\mathrm{t\bar{t}}$ lepton+jets final state contributes most of events in the pre-selection. Figure 6.18 (6.17) shows the distributions with the MC and data-driven QCD backgrounds in the $\mathrm{e}\tau_\mathrm{h}$ ($\mu\tau_\mathrm{h}$) pre-selection.

In the fit, the event yields of the background processes with a misidentified τ_h are determined by adjusting the normalization of the shapes of the m_T distributions in each of these processes independently. The normalization factors are introduced as nuisance parameters with pre-fit constraints of the magnitude determined from studies in other processes [3]. The corresponding uncertainties are discussed in Section 6.5.

6.4 Event yields and control distributions

The event yields in the pre-selection and the main selection including the event categories are listed in Tables 6.6 and 6.7 for the $\mathrm{e}\tau_\mathrm{h}$ and $\mu\tau_\mathrm{h}$ final states, respectively.

Control distributions of the main kinematic parameters are presented in the pre-selection for both final states in Figures 6.19 and 6.20, and the main selection in Figures 6.21 and 6.22. The m_T distribution (Eq. 4.1) that is used to extract the cross-section in the shape fit is presented in the signal-like and background-like event categories, for the $\mathrm{e}\tau_\mathrm{h}$ and $\mu\tau_\mathrm{h}$ final states in Figure 6.24. The hatched band in the figures of the m_T distribution includes all pre-fit systematic uncertainties, that are discussed in Section 6.5.

A good shape agreement is observed between the data and the expected sum of signal and background distributions within the statistical and systematic uncertainties.

6.5 Systematic Uncertainties

The main sources of systematic uncertainty are from τ_h identification and misidentification, b tagging, estimation of pileup in the pp collisions, jet energy scale (JES), and jet energy resolution (JER). Other sources of uncertainty are from the lepton identification, trigger efficiency, and the measurement of the integrated luminosity. Theoretical uncertainties are also included in the event simulation. Uncertainties are applied in a coherent way to signal and background processes. The corresponding corrections and their uncertainties are measured in dedicated studies, which are described below.

The uncertainty in the efficiency of τ_h identification is 5% for all τ_h with $p_\mathrm{T} > 20\,\mathrm{GeV}$ and is applied to all processes with a genuine τ_h. It is measured with a tag-and-probe technique in samples enriched in $\mathrm{Z} \to \tau_\ell\tau_\mathrm{h}$ events [3]. The τ_h charge

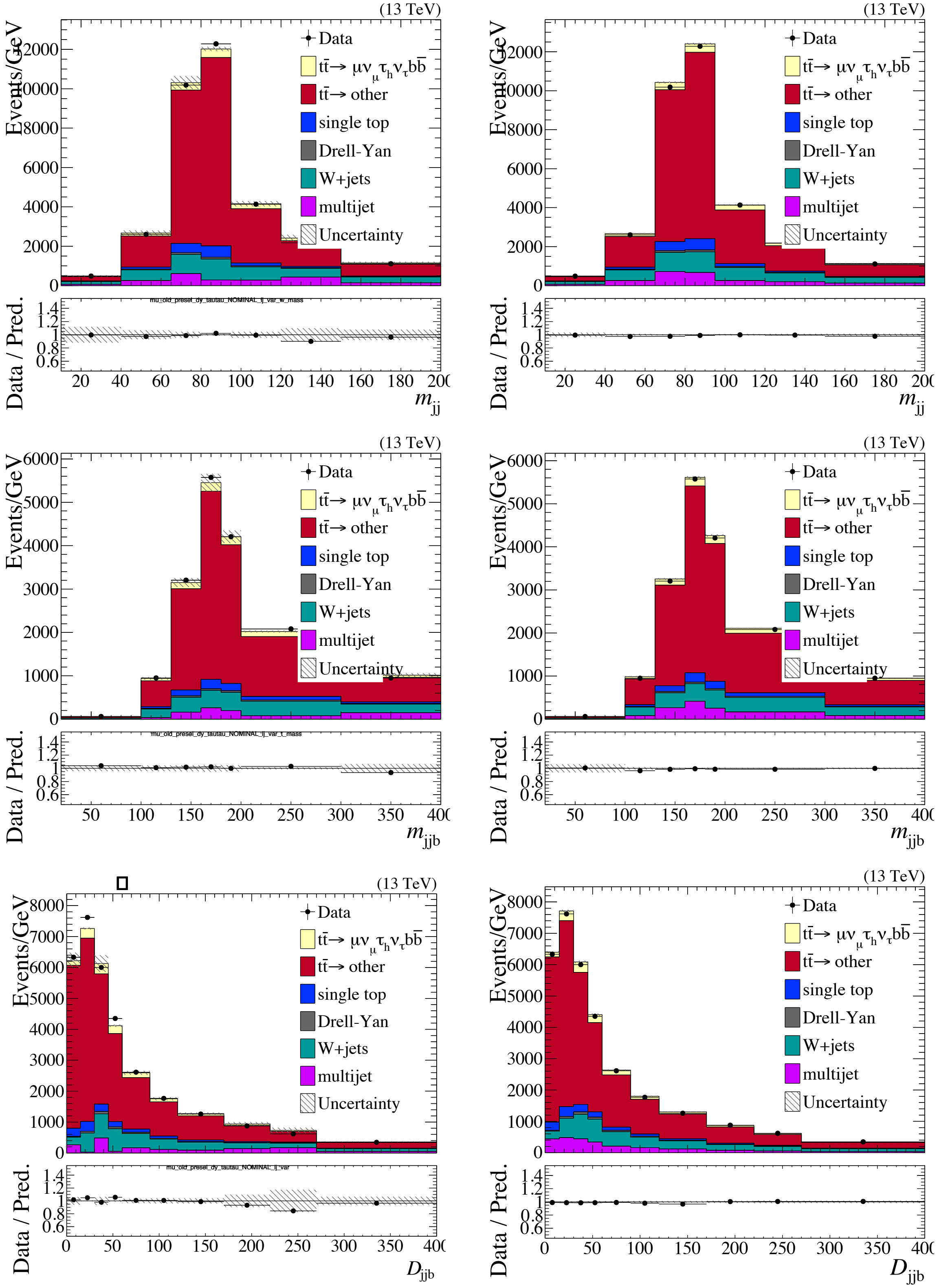

Figure 6.17: Distributions of the dijet mass (top), the trijet mass (middle), and the threshold parameter D_{jjb} (bottom) in the $\mu\tau_{\text{h}}$ pre-selection. The distributions with MC QCD are shown on the left, while the distributions with data-driven QCD are on the right.

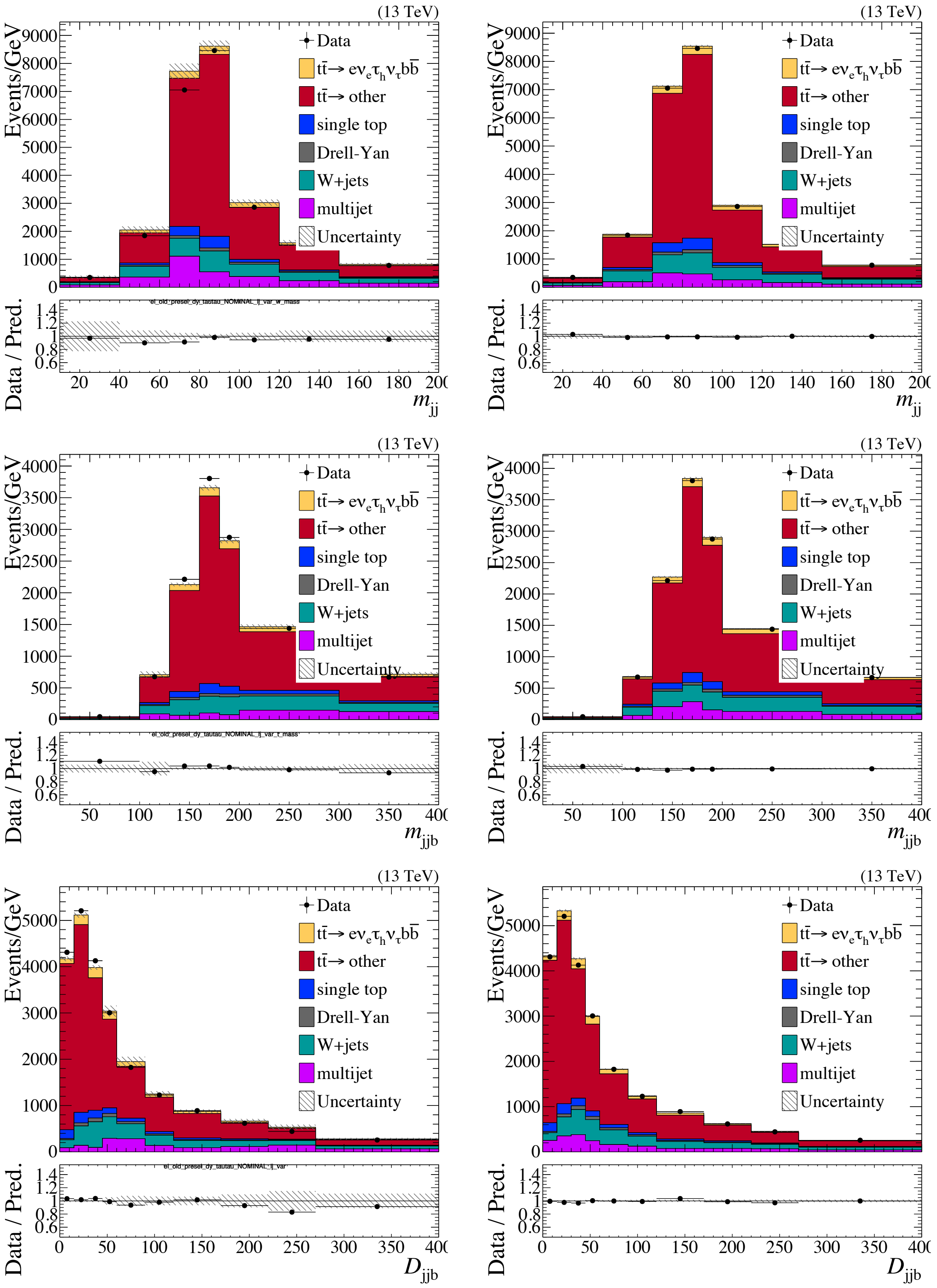

Figure 6.18: Distributions of the dijet mass (top), the trijet mass (middle), and the threshold parameter D_{jjb} (bottom) in the $\mathrm{e}\tau_{\mathrm{h}}$ pre-selection. The distributions with MC QCD are shown on the left, while the distributions with data-driven QCD are on the right.

Table 6.6: Expected and observed event yields in the $e\tau_h$ final state. The signal and SM background processes are given for an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$. Statistical uncertainties are shown.

Process	Pre-selection	Main selection	Background-like	Signal-like
$t\bar{t} \to e\tau_h\nu_e\nu_\tau b\bar{b}$	31065 ± 108	8759 ± 56	3439 ± 35	5319 ± 43
$t\bar{t} \to e\nu_e q\bar{q}' b\bar{b}$	339925 ± 356	4059 ± 39	2451 ± 30	1607 ± 24
$t\bar{t} \to \tau_e\tau_h\nu_\tau\nu_e\nu_\tau b\bar{b}$	1774 ± 26	500 ± 13	192 ± 8	308 ± 10
$t\bar{t} \to \tau_e\nu_\tau\nu_e q\bar{q}' b\bar{b}$	18893 ± 84	216 ± 9	132 ± 7	84 ± 6
$t\bar{t} \to$ other	22620 ± 92	181 ± 8	69 ± 5	113 ± 6
Single top → $e\tau_h$+X	1437 ± 16	498 ± 9	178 ± 5	320 ± 7
Single top → $eq\bar{q}'$+X	31525 ± 76	355 ± 8	170 ± 6	185 ± 5
Single top → other	2784 ± 22	58 ± 3	24 ± 2	34 ± 2
DY → $\tau_e\tau_h$	1834 ± 58	347 ± 25	123 ± 15	224 ± 19
DY → other	10211 ± 184	111 ± 14	30 ± 7	82 ± 12
W → $e\nu_e$	55568 ± 260	400 ± 34	179 ± 31	221 ± 14
W → $\tau_e\nu_\tau$	3360 ± 62	27 ± 6	14 ± 5	13 ± 3
W → $\tau_h\nu_\tau$	41 ± 6	3 ± 2	1 ± 1	2 ± 1
Dibosons	2461 ± 33	53 ± 4	16 ± 2	37 ± 3
QCD multijet	33277 ± 853	447 ± 66	71 ± 45	376 ± 49
Total	556775 ± 999	16015 ± 107	7090 ± 75	8925 ± 77
Data	548107	15420	6787	8633

Table 6.7: Expected and observed event yields in the $\mu\tau_\mathrm{h}$ final state. The signal and SM background processes are given for an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$. Statistical uncertainties are shown.

Process	Pre-selection	Main selection	Background-like	Signal-like
$\mathrm{t\bar{t}} \rightarrow \mu\tau_\mathrm{h}\nu_\mu\nu_\tau\mathrm{b\bar{b}}$	45855 ± 130	13029 ± 67	5137 ± 43	7892 ± 52
$\mathrm{t\bar{t}} \rightarrow \mu\nu_\mu\mathrm{q\bar{q}'b\bar{b}}$	496455 ± 427	6117 ± 47	3674 ± 37	2443 ± 30
$\mathrm{t\bar{t}} \rightarrow \tau_\mu\tau_\mathrm{h}\nu_\tau\nu_\mu\nu_\tau\mathrm{b\bar{b}}$	2628 ± 31	731 ± 16	280 ± 10	451 ± 12
$\mathrm{t\bar{t}} \rightarrow \tau_\mu\nu_\tau\nu_\mu\mathrm{q\bar{q}'b\bar{b}}$	28227 ± 102	347 ± 11	209 ± 9	139 ± 7
$\mathrm{t\bar{t}} \rightarrow$ other	32780 ± 110	259 ± 10	88 ± 6	171 ± 8
Single top $\rightarrow \mu\tau_\mathrm{h}$+X	2001 ± 19	691 ± 11	235 ± 6	456 ± 9
Single top $\rightarrow \mu\mathrm{q\bar{q}'}$+X	45690 ± 90	516 ± 9	230 ± 6	286 ± 7
Single top → other	3968 ± 26	87 ± 4	35 ± 3	52 ± 3
DY $\rightarrow \tau_\mu\tau_\mathrm{h}$	2623 ± 68	581 ± 31	193 ± 18	388 ± 26
DY → other	8236 ± 190	32 ± 14	11 ± 9	21 ± 10
W $\rightarrow \mu\nu_\mu$	81338 ± 318	594 ± 28	238 ± 21	355 ± 18
W $\rightarrow \tau_\mu\nu_\tau$	4771 ± 66	45 ± 7	16 ± 4	29 ± 5
W $\rightarrow \tau_\mathrm{h}\nu_\tau$	3 ± 2	0 ± 0	0 ± 0	0 ± 0
Dibosons	3254 ± 38	80 ± 5	23 ± 2	57 ± 4
QCD multijet	51245 ± 1019	354 ± 79	122 ± 56	231 ± 57
Total	809075 ± 1191	23462 ± 126	10490 ± 86	12972 ± 92
Data	799953	23016	9931	13085

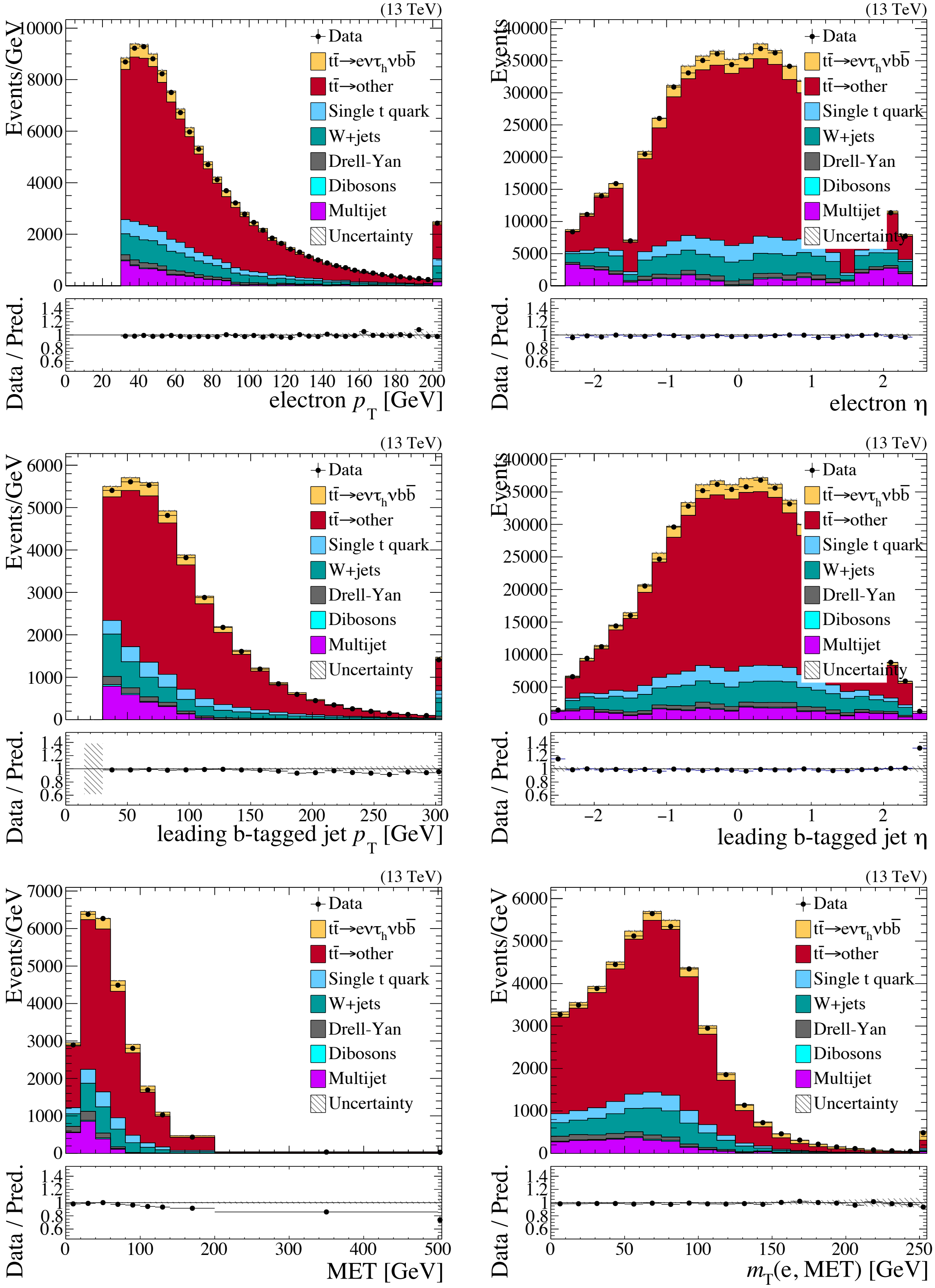

Figure 6.19: Distributions in the pre-selection of the $e\tau_h$ final state. Transverse momentum (left) and pseudo-rapidity (right) of the electron (top) and the leading b-tagged jet (middle), p_T^{miss} and m_T (bottom). Only statistical uncertainties are shown.

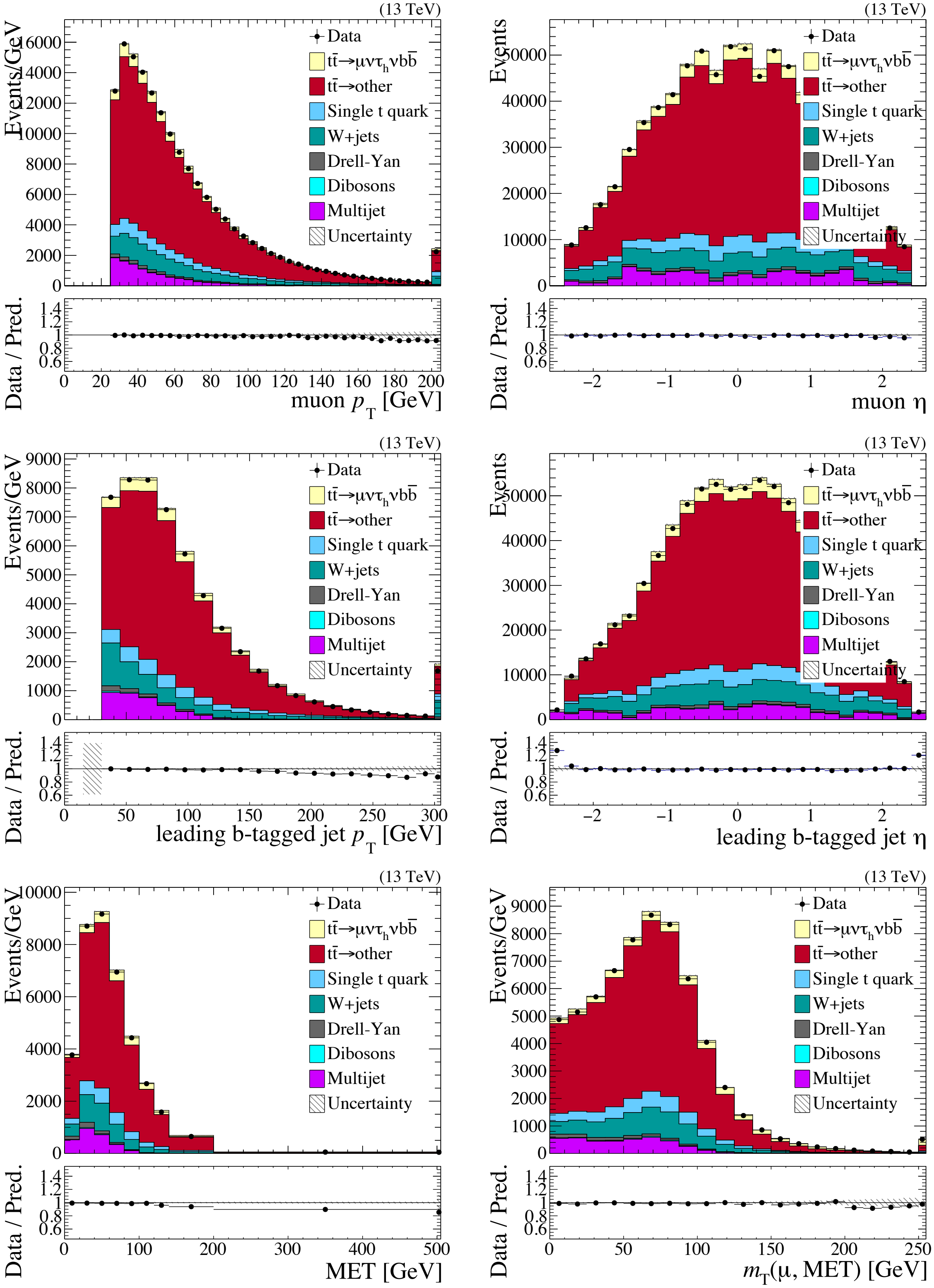

Figure 6.20: Distributions in the pre-selection of the $\mu\tau_\mathrm{h}$ final state. Transverse momentum (left) and pseudo-rapidity (right) of the muon (top) and the leading b-tagged jet (middle), $p_\mathrm{T}^\mathrm{miss}$ and m_T (bottom). Only statistical uncertainties are shown.

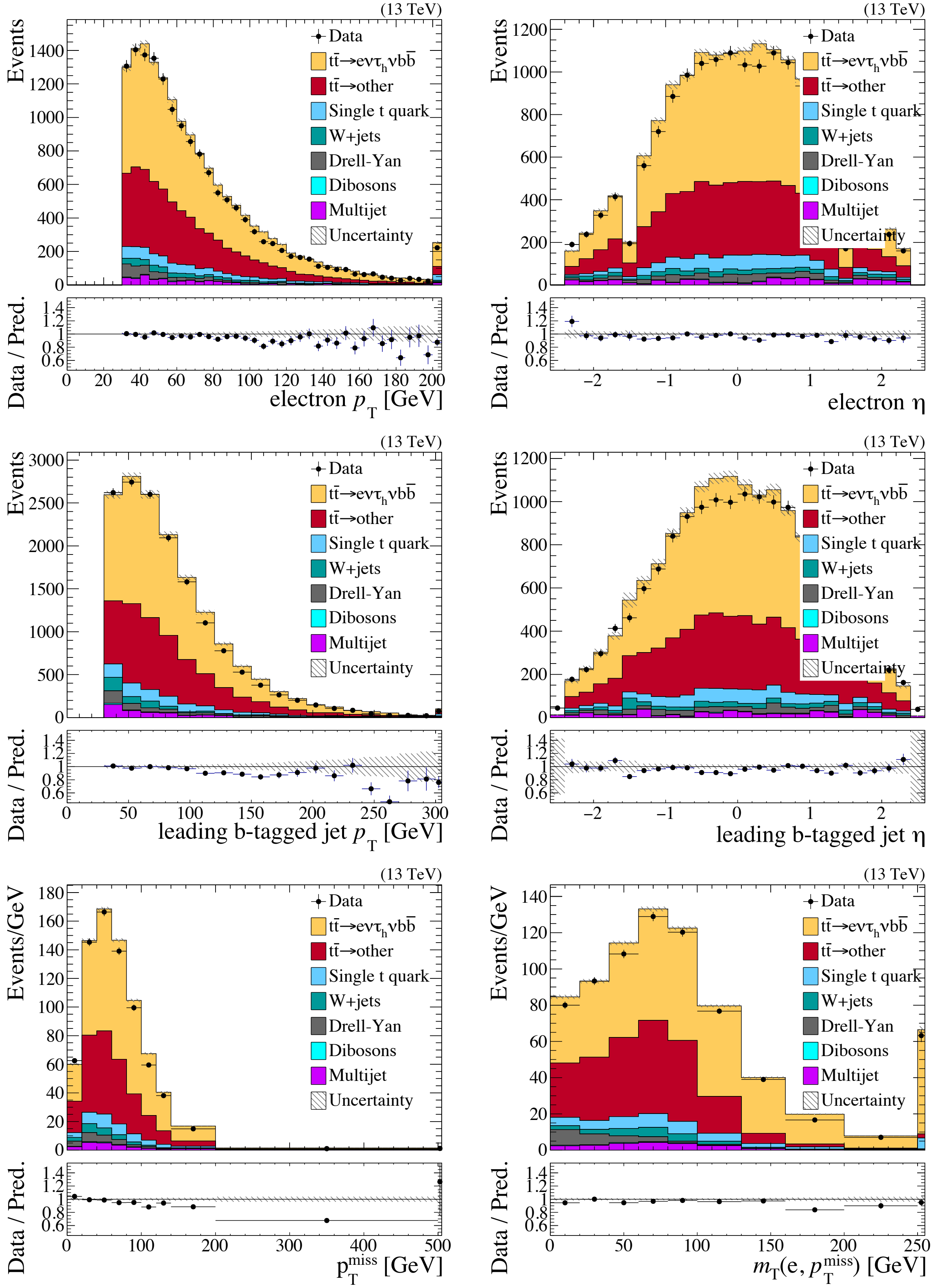

Figure 6.21: Distributions in the final selection of the $e\tau_h$ final state. Transverse momentum (left) and pseudo-rapidity (right) of the electron (top) and the leading b-tagged jet (middle), p_T^{miss} and m_T (bottom). Only statistical uncertainties are shown.

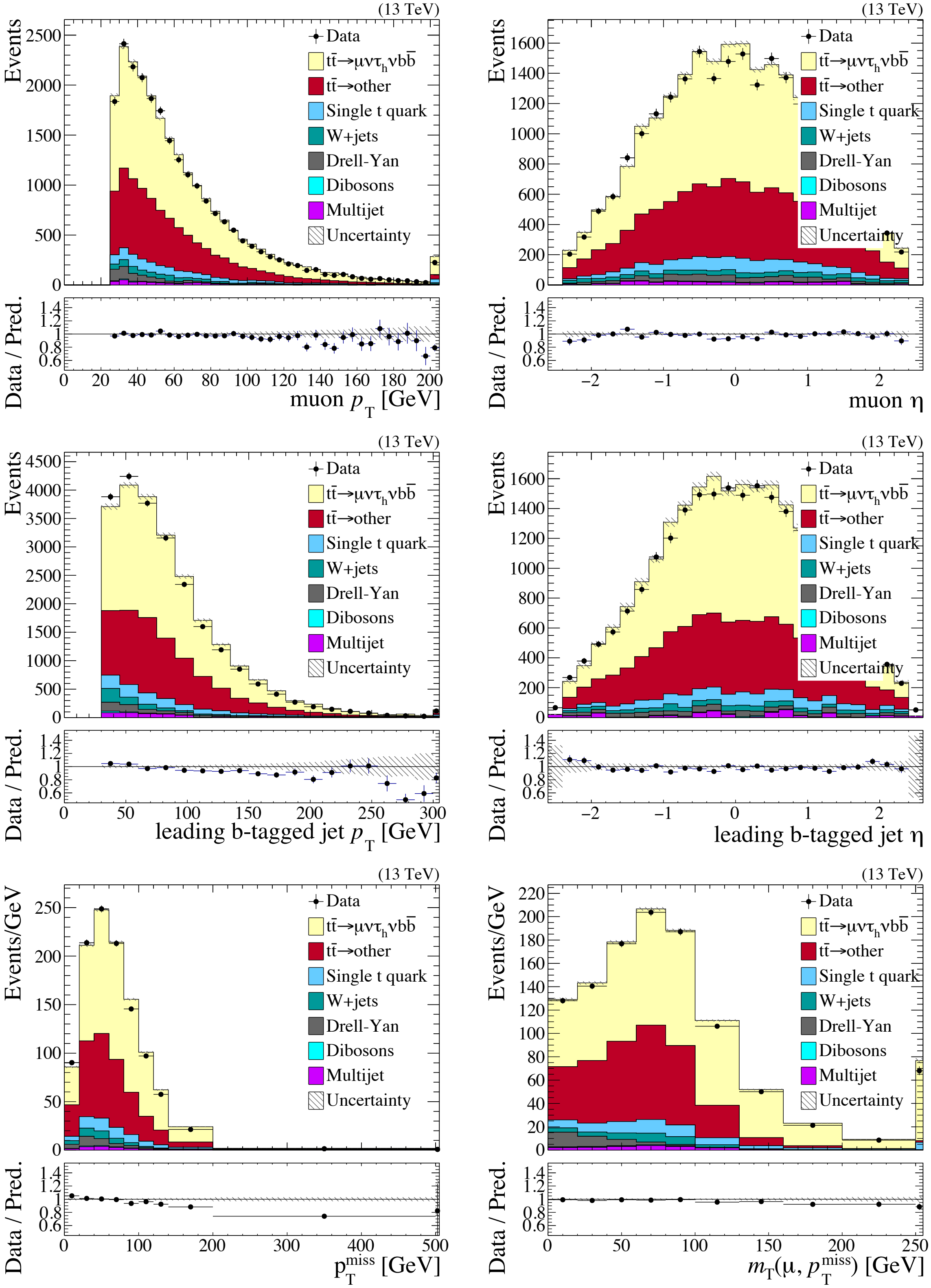

Figure 6.22: Distributions in the final selection of the $\mu\tau_h$ final state. Transverse momentum (left) and pseudo-rapidity (right) of the muon (top) and the leading b-tagged jet (middle), p_T^{miss} and m_T (bottom). Only statistical uncertainties are shown.

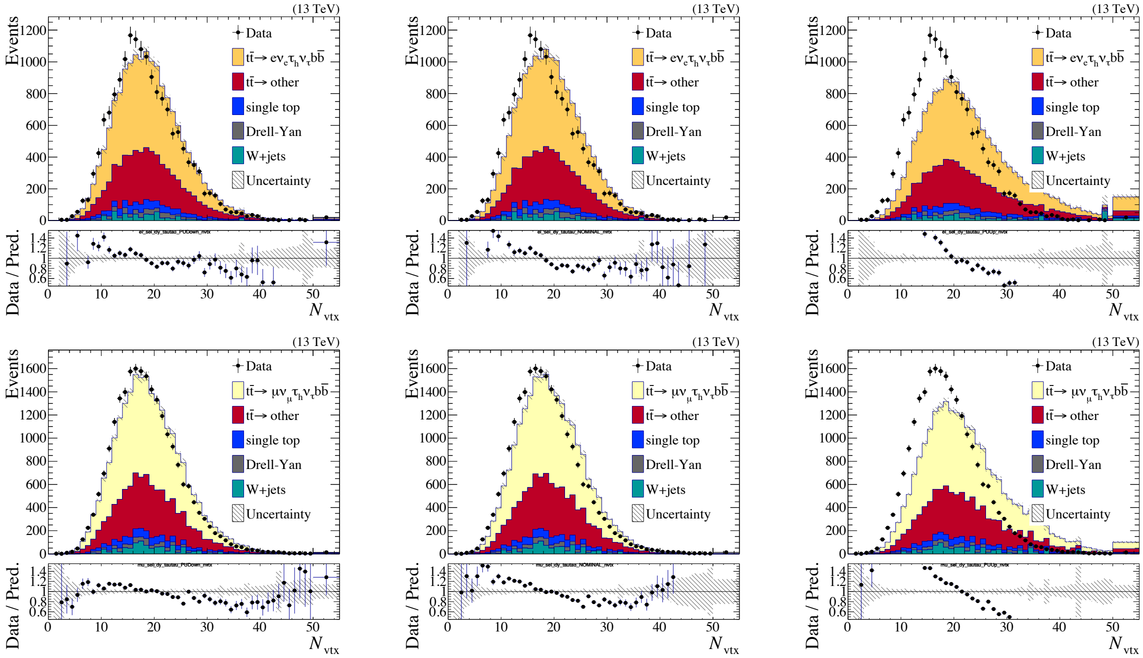

Figure 6.23: Distributions of the number of reconstructed vertices at the OC tau selection step in the $e\tau_h$ (top) and $\mu\tau_h$ (bottom) final states. The distributions are produced with different PU reweighting: with the nominal weights for a minimum bias cross section of 69.2 (middle), the variations of 5% down (left) and 5% up (right). Only statistical uncertainties are shown.

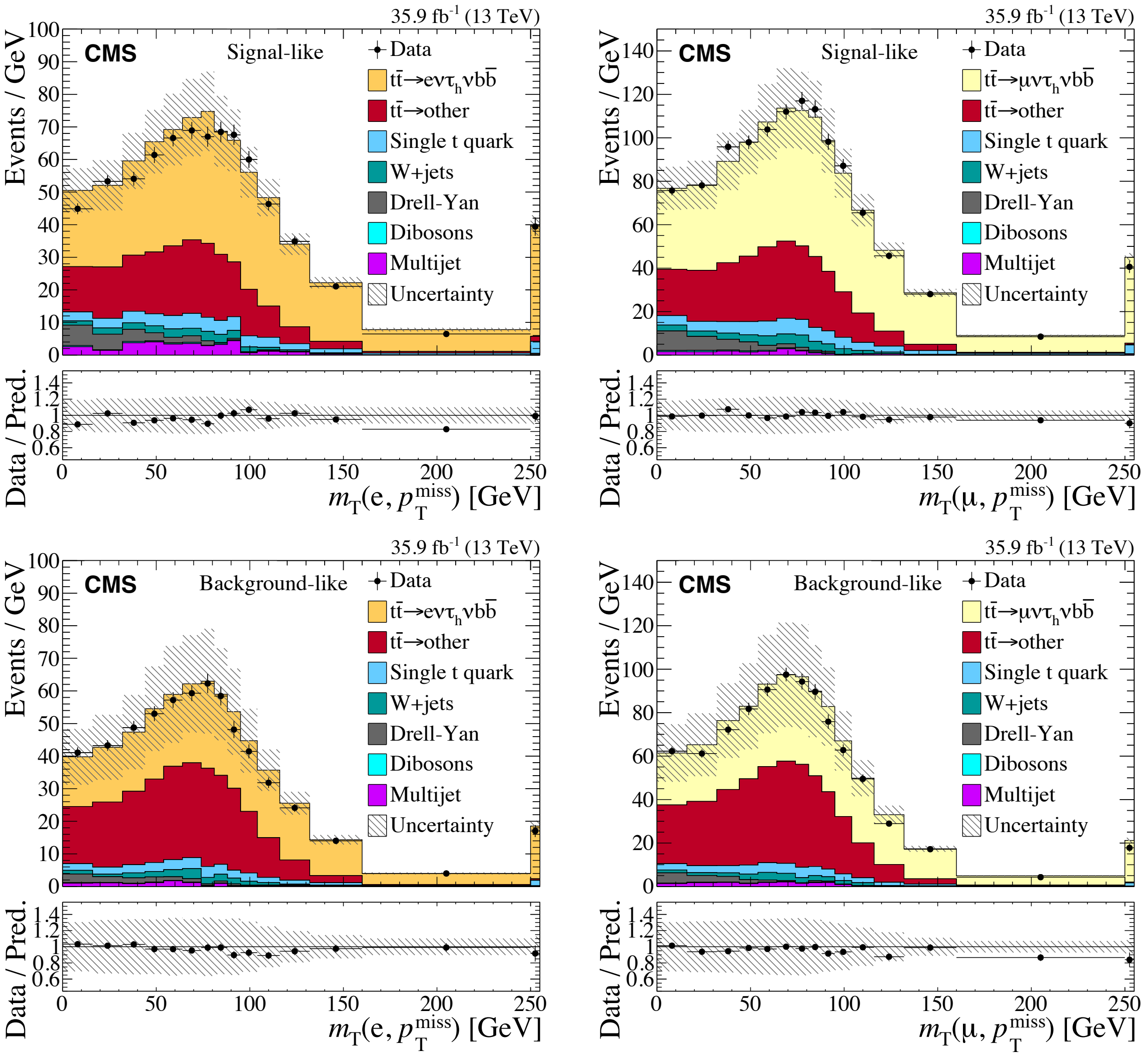

Figure 6.24: The transverse mass distributions between lepton (e or μ) and $p_{\mathrm{T}}^{\mathrm{miss}}$, m_{T}, in the signal-like (upper) and background-like (lower) event categories for the eτ_{h} (left) and $\mu\tau_{\mathrm{h}}$ (right) final states observed prior to fitting. Distributions obtained from data (filled circles) are compared with simulation (shaded histograms). The last bin includes overflow events. The simulated contributions are normalized to the cross section values predicted in the SM. The main processes are shown: the signal, the other $\mathrm{t\bar{t}}$ processes grouped together, single top quark production, W+jets, DY processes, diboson, and multijet production. The ratio of the data to the total SM prediction is shown in the lower panel. The vertical bars on the data points indicate the statistical uncertainties, the hatched band indicates the systematic uncertainties and the statistical uncertainties in all simulated samples.

misidentification probability, estimated to be less than 1%, is considered a part of the τ_{h} identification efficiency uncertainty. The correction to the reconstructed energy of the τ_{h} jet (τ energy scale) and the corresponding uncertainty is estimated in a fit of the data in distributions sensitive to the τ energy, such as the τ_{h} visible mass [3]. The fitted corrections are 0.995 ± 0.012 for the τ decay mode with 1 charged track and no neutral particles, 1.011 ± 0.012 for other decay modes with 1 charged particle, and 1.006 ± 0.012 for all decay modes with 3 charged particles.

The dominant background contribution arises from processes where a jet is misidentified as τ_{h}, mainly lepton+jets $\mathrm{t\bar{t}}$, W+jets, and multijet QCD production. The τ_{h} misidentification probability and its uncertainty in these processes are directly measured in the fit. The misidentification probability is varied within ±50% of the expected values in all processes with a jet falsely identified as the τ_{h} candidate. The variation covers the differences between expected and observed misidentification probabilities and the possible dependence on other kinematic properties of the τ_{h} candidate [3]. The misidentification probability is significantly constrained in the fit and is not the dominant source of the uncertainty in the final result. It is the main improvement in systematic uncertainties with respect to previous results.

The uncertainties related to b tagging (mistagging) efficiencies are estimated from a variety of control samples enriched in b quarks (c and light-flavour quarks) [98]; the data-to-simulation scale factors for b, c, and light-flavour jets are applied to the simulation and the corresponding uncertainties are included in the fit. These uncertainties contribute about 2% to the cross section uncertainty measured by the fit in both $\mathrm{e}\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ final states.

The uncertainties in the JES, JER, and $p_{\mathrm{T}}^{\mathrm{miss}}$ scales are estimated according to the prescription described in Ref. [52]. The uncertainty in the JES is evaluated as a function of jet p_{T} and η. The JES and JER uncertainties are propagated to the $p_{\mathrm{T}}^{\mathrm{miss}}$ scale.

The lepton trigger, identification, and isolation efficiencies are measured in data and simulation with a tag-and-probe method in $\mathrm{Z} \to \ell^+\ell^-$ events [89, 90, 53]. The simulated events are corrected with the corresponding data-to-simulation scale factors. The uncertainties in the scale factors are included as systematic uncertainties in the measurement.

The uncertainty in the integrated luminosity is estimated to be 2.5% [108].

The pileup distribution is estimated from the measured luminosity in each bunch crossing multiplied by the average total inelastic cross section. It is used to model the pileup in simulation with an uncertainty obtained by varying the inelastic pp cross section extracted from a control region by its uncertainty of ±4.6% [100].

The measurement includes the uncertainty in the modelling of the b quark fragmentation, which covers e^+e^- data [109, 110, 111, 112] at the Z pole with the Bowler–Lund [113] and Peterson [114] parametrizations, and the uncertainties in the semileptonic b-flavoured hadron branching fractions according to their measured values [4]. An uncertainty in the modelling of the p_T distribution of the top quark in $t\bar{t}$ processes is included to cover the difference between the predicted and observed spectra [33, 102, 103, 104]. The fit is sensitive to the top quark p_T as it affects the shape of the m_T distribution. The top quark p_T variation also covers the slight trend of the τ_h p_T distribution.

The cross section is measured by the fit in the fiducial phase space of the detector. The fiducial cross section is extrapolated to the full phase space by correcting for the acceptance of the $t\bar{t}$ signal process. The fit and the acceptance include the following modelling uncertainties: the renormalization and factorization scales, and PDFs including α_S. The uncertainty in the PDF is estimated by using the CT14 (NNLO) set as alternative PDFs. The renormalization and factorization scales in the ME calculations are varied independently by factors of 0.5 and 2.0 from their nominal values, and the envelope of the variations is included in the measurement. The scale is varied by factors of 0.5 and 2.0 in the parton shower (PS) simulation of final-state and initial-state radiation, FSR and ISR. The h_{damp} parameter regulating the real emissions in POWHEG (ME-PS matching) is varied from its central value of $1.58\,m_t$ using samples with h_{damp} set to $0.99\,m_t$ and $2.24\,m_t$ ($m_t = 172.5\,\mathrm{GeV}$), as obtained from tuning this parameter to $t\bar{t}$ data at $\sqrt{s} = 8\,\mathrm{TeV}$ [115]. The underlying event tune is varied within its uncertainties [72, 115]. The effect of these uncertainties on the final state objects is included in the fit in the fiducial phase space by adding the corresponding systematic variations normalized to the nominal acceptance. Therefore, the measurement in the fiducial phase space is performed with the nominal acceptance and its uncertainties are only included in the extrapolation to the full phase space. The uncertainties in the fit are not correlated with the acceptance uncertainty in the extrapolation to the full phase space.

The theoretical uncertainties are implemented by reweighting the simulated events with the corresponding scale factors. The differences between weighted and unweighted distributions are taken as the uncertainties in the modelling. Separate datasets with varied parameters are used for determination of the FSR, ISR, ME-PS matching, and underlying event uncertainties.

The uncertainty due to the limited statistics in the simulated datasets is included in the fit with the Barlow-Beeston approach [116, 117].

The changes in the parton shower parameters that implement the systematic uncertainties (underlying event, ME-PS matching, initial and final state radiation) also affect the JES and the probability of b tagging. Therefore, the nominal corrections to JES and b tagging are adjusted to the difference in these systematics. The tagging efficiency that is used for the b tagging corrections is re-measured in the datasets simulated for the parton shower systematics. The nominal JES correction is multiplied with the ratio of JES measured in systematic and nominal $\mathrm{t\bar{t}}$ datasets, shown in Figure 6.25.

6.6 Fitting procedure

The $\mathrm{t\bar{t}}$ production cross section is extracted from a PLR fit to the binned distribution of the transverse mass of the lepton and $p_{\mathrm{T}}^{\mathrm{miss}}$ system in two kinematic event categories, for each of the $\mathrm{e}\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ final states. The transverse mass is defined as $m_{\mathrm{T}} = \sqrt{2|\vec{p}_{\mathrm{T}}^{\ell}||\vec{p}_{\mathrm{T}}^{\mathrm{miss}}|(1-\cos\Delta\varphi)}$, where $\Delta\varphi$ is the azimuthal angle difference between the lepton transverse momentum vector, $\vec{p}_{\mathrm{T}}^{\ell}$, and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$. The m_{T} distribution provides separation between signal and background processes and does not significantly depend on p_{T} and η of the τ_{h} candidate, or other jet characteristics in the kinematic ranges of this study.

The cross section is measured in the fit with the signal strength parameter that encodes the ratio of the signal cross section to the value expected in the SM. The expected number of events in a given bin of the m_{T} distribution is parametrized as a function of signal strength and nuisance parameters. The nuisance parameters encode the effects of systematic uncertainties. The signal strength is a free parameter in the fit. The fitted variables do not significantly depend on the kinematic properties of the τ lepton in the specific $\mathrm{t\bar{t}}$ signal model considered here, i.e. $\mathrm{t\bar{t}} \to (\ell\nu_{\ell})(\tau_{\mathrm{h}}\nu_{\tau})\mathrm{b\bar{b}}$. The likelihood function is defined as a product of Poisson distributions of the expected numbers of events and nuisance constraints in bins of the m_{T} distribution. Based on the likelihood function, the PLR test statistic is defined as the ratio between the maximum of the likelihood for a given value of the signal strength and the global maximum of the likelihood function. The effects of the systematic uncertainties on the signal strength are determined within this approach.

The expected number of events in a given bin k of the m_{T} distribution in both event categories of one of the final states is given as:

$$\hat{N}_k(s,\theta_i) = \hat{N}_k^s(s,\theta_i) + \hat{N}_k^b(\theta_i) = s\hat{S}_k \cdot \prod_i (1+\sigma_i^s\theta_i) + \sum_{b\in\mathrm{bkg}} \hat{B}_k \cdot \prod_i (1+\sigma_i^b\theta_i) \tag{6.3}$$

Where $\hat{S}_k$ and $\hat{B}_k$ are expectations in the given bin k for the signal and background processes respectively, the 1σ effect of ith source of the systematic uncertainty on the signal or background expectations (σ_i^s and σ_i^b, respectively) is parametrized with the

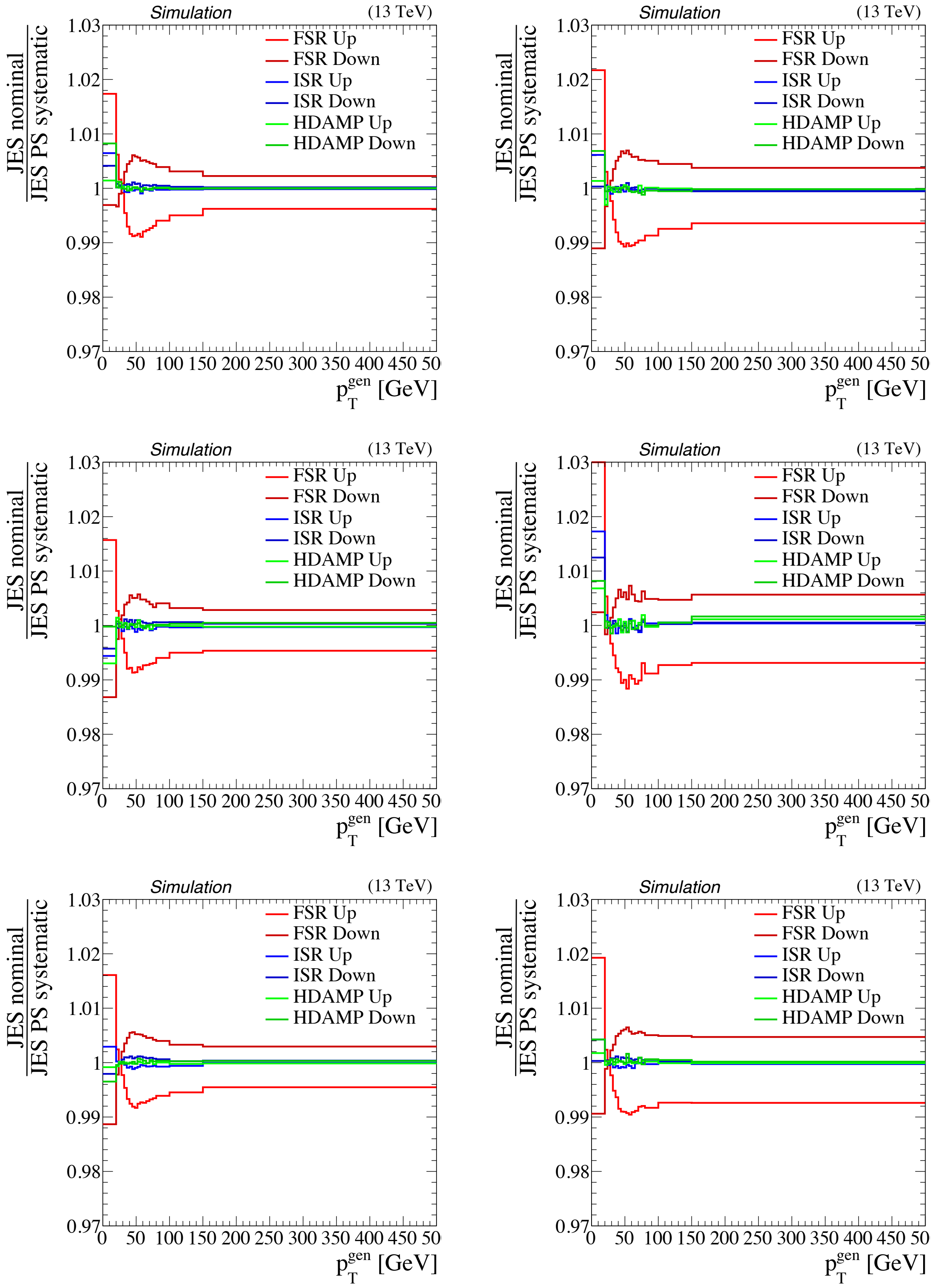

Figure 6.25: The corrections to the nominal jet energy scale for the parton shower systematics: in the barrel on the left, in endcaps on the right, for b jets (top), c jets (middle) and light flavour jets (bottom).

nuisance parameter θ_i, and s is the ratio of the signal cross section to the expected theoretical value, i.e. the signal strength. In the measurement of the ratio of the partial to the total width of the top quark decay the signal strength encodes the $\mathcal{B}(\mathrm{t} \to \tau\nu_\tau \mathrm{b})$. In this case the $\mathrm{t\bar{t}}$ dilepton final states with two τs are included with the s^2 factor, that shifts the result of the fit by about 0.1%, a negligible effect in comparison with uncertainties. Certain systematic uncertainties affect both the normalization and shape of the distributions. In this case a "vertical" interpolation scheme is used for the shapes within the range of the uncertainty. Shape variations are interpolated quadratically when they are below 1σ, and extrapolated linearly above 1σ.

For the model given by Eq. 7.4, the likelihood function is defined as a product of Poisson distributions of the expected numbers of events and nuisance constraints in bins of the m_T distribution:

$$\mathcal{L}(s,\theta_i) = \prod_k \mathcal{P}_{oisson}\left[N_k|\hat{N}_k(s,\theta_i)\right] \cdot \prod_i pdf(\theta_i,0,1) \tag{6.4}$$

where N_k is the number of observed events in the bin k.

The probability distribution functions of the nuisance constraints and their magnitudes are listed in Table 6.8 for experimental and theoretical uncertainties. The theoretical uncertainties in the hard process modeling (listed as "Hard process modeling" in Table 6.8) affect the kinematic distribution of the final states objects and the acceptance of the signal process. Since the observed data cover only the visible part of the phase space and does not constrain the kinematics in the whole phase space, the effect on the acceptance is "externalized" from the fit by normalizing the variations of the corresponding systematic uncertainties to the nominal acceptance. The extrapolation of the cross-section measured by the fit in the fiducial phase space to the full phase space is carried out in Section 6.7 according to the acceptance estimated in the nominal $\mathrm{t\bar{t}}$ MC with the externalized theoretical uncertainties of the modeling included in the uncertainty.

Based on the likelihood function, the PLR test statistic is defined as:

$$\lambda(s) = \frac{\mathcal{L}(s,\tilde{\theta}_i(s))}{\mathcal{L}(\hat{s},\hat{\theta}_i)} \tag{6.5}$$

The quantities $\hat{s}$ and $\hat{\theta}_i$ give the global maximum of the likelihood function, and $\tilde{\theta}_i(s)$ maximize the likelihood for a given value of s. This approach includes the effect of systematic uncertainties on the signal strength: the test statistic $\lambda(s)$ is broader due to the adjustment of the systematic uncertainties in $\tilde{\theta}_i(s)$. It also allows to estimate

Table 6.8: The nuisance parameters of the PLR fit corresponding to the systematic uncertainties (in %). Column “type” shows the treatment of the parameter: rate, shape, prior probability distribution function ($\ln N$ for log-normal distribution). Column “range” lists the 1-sigma deviation for the “rate” parameters, the “shape” parameters are varied with their recommended 1σ deviation in each event. The τ misidentification probabilities in the QCD and W+jets processes are included in the normalization uncertainties. For the simultaneous fit of two final states all nuisances parameters are correlated, except the parameters which correspond to the lepton object in the events: the efficiencies of lepton identification and trigger (“Lepton ID” and “Trigger”).

Uncertainty	Type	Range (%)
τ_h identification	rate	5
τ_h energy scale	shape+rate	
τ_h misidentification	rate, $\ln N$	50
Pileup	shape+rate	
Lepton ID	shape+rate	
Trigger	shape+rate	
JES	shape+rate	
JER	shape+rate	
b tagging	shape+rate	
Top p_T	shape+rate	
DY normalization	rate, $\ln N$	5
Single top normalization	rate, $\ln N$	10
Diboson normalization	rate, $\ln N$	10
W+jet normalization	rate, $\ln N$	50
QCD normalization	rate, $\ln N$	50
Parton Showering		
b fragmentation	shape+rate in $\mathrm{t\bar{t}}$ processes	
Peterson variation	shape+rate in $\mathrm{t\bar{t}}$ processes	
Semileptonic BR of B hadrons	shape+rate in $\mathrm{t\bar{t}}$ processes	
Hard process modeling		
α_S variation	shape in $\mathrm{t\bar{t}}$ processes	
PDF	shape in $\mathrm{t\bar{t}}$ processes	
Final state radiation	shape in $\mathrm{t\bar{t}}$ processes, $\ln N$	
Initial state radiation	shape in $\mathrm{t\bar{t}}$ processes, $\ln N$	
ME-PS matching (h-damp)	shape in $\mathrm{t\bar{t}}$ processes, $\ln N$	
underlying event (Pythia tune)	shape in $\mathrm{t\bar{t}}$ processes, $\ln N$	
Matrix Element (ME) QCD scale		
ME factorization scale	shape in $\mathrm{t\bar{t}}$ processes	
ME renormalization scale	shape in $\mathrm{t\bar{t}}$ processes	
ME correlated variation	shape in $\mathrm{t\bar{t}}$ processes	

the effect of a given systematic separately by fixing the others in the fit. More details on the method can be found in Ref. [20].

The computation of PLR is performed in RooStats [118] through the Higgs Combination tool [116, 117, 119].

The bias and the expected uncertainty of the fit are examined with toy experiments. A number of toy experiments is generated for hypothetical values with signal strengths $s = 0.9, 1.0, 1.1$. The generated and the fitted signal strengths show a linear correlation with no particular bias in the fit, as shown in Figure 6.26 for both final states. The distribution also reflects the expected uncertainty of the fit.

Along with independent fits in the $\mathrm{e}\tau_\mathrm{h}$ and $\mu\tau_\mathrm{h}$ final states, a simultaneous fit is performed in both final states. In this case, both final states enter the fitting procedure effectively making a fit in 4 event categories: 2 final states with 2 categories each. The signal strength and nuisance parameters are kept the same across all categories in the fit, except for the nuisance parameters that encode the uncertainties in the electron and muon reconstruction.

The negative log-likelihood scans in the separate final states and in the simultaneous fit are shown as a function of the cross section in Figure 6.27. The width of the curves corresponds to the expected uncertainty of the measurement. The scan for the statistical uncertainty is obtained by fixing all nuisance parameters to their post-fit values. The effects from the variations of the nuisance parameters on the signal strength are shown in Figure 6.28.

6.7 Results

The number of observed signal events is measured by the PLR fit in the fiducial phase space visible to the detector. The fit is performed in the $\mathrm{e}\tau_\mathrm{h}$ and $\mu\tau_\mathrm{h}$ final states separately, as well as in the $\ell\tau_\mathrm{h}$ combined final state.

From the number of signal events the cross section in the fiducial phase space is estimated. It is extrapolated to the full phase space with the signal acceptance of the detector. The ratio of the cross section in the full phase space and the light dilepton cross section measured in Ref. [2] is calculated, and used in the estimation of the partial width of the top quark decay into τ lepton.

6.7.1 Cross section in the fiducial phase space

The fiducial cross section of the $\mathrm{t\bar{t}}$ production is extracted from the fit to the observed data in the acceptance region of kinematic phase space defined by the event

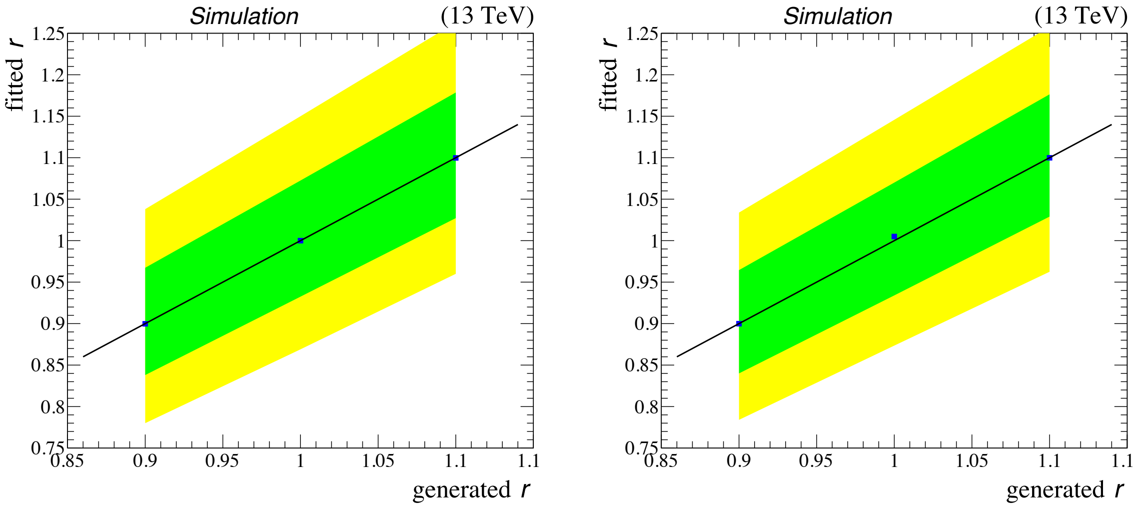

Figure 6.26: Generated versus fitted signal strength in toy experiments for the eτ_{h} (left) and $\mu\tau_{\mathrm{h}}$ (right) final states. The median of toys is represented by blue dots, the bands represent 68% and 95% quantiles. The black line shows the fully correlated case.

selection criteria described earlier. Therefore, the estimate of the fiducial cross section includes the branching fractions of the final states, trigger, lepton identification and isolation, and the overall reconstruction efficiency. It is purely an experimental result equivalent to

$$\sigma_{\mathrm{t\bar{t}}}^{\mathrm{fid}} = (N - B)/(\mathcal{L}\epsilon\mathcal{B}), \tag{6.6}$$

where N is the number of observed candidate events and B is the estimate of the number of background events extracted from the fit, $\mathcal{L}$ is the integrated luminosity, ϵ is the total event reconstruction efficiency, and $\mathcal{B}$ is the branching fraction of the relevant $\mathrm{t\bar{t}} \to \ell\tau_{\mathrm{h}}$ final state. The cross sections in the fiducial phase space for the individual eτ_{h} and $\mu\tau_{\mathrm{h}}$ final states, as well as the $\ell\tau_{\mathrm{h}}$ combined final state, are measured from the PLR fit to be:

$$\sigma_{\mathrm{t\bar{t}}}^{\mathrm{fid}}(\mathrm{e}\tau_{\mathrm{h}}) = 133.2 \pm 1.9\,(\mathrm{stat.}) \pm 10.9\,(\mathrm{syst.}) \pm 3.3\,(\mathrm{lum.})\,\mathrm{pb}, \tag{6.7}$$

$$\sigma_{\mathrm{t\bar{t}}}^{\mathrm{fid}}(\mu\tau_{\mathrm{h}}) = 135.2 \pm 1.5\,(\mathrm{stat.}) \pm 9.9\,(\mathrm{syst.}) \pm 3.4\,(\mathrm{lum.})\,\mathrm{pb}, \tag{6.8}$$

$$\sigma_{\mathrm{t\bar{t}}}^{\mathrm{fid}}(\ell\tau_{\mathrm{h}}) = 134.5 \pm 1.2\,(\mathrm{stat.}) \pm 9.5\,(\mathrm{syst.}) \pm 3.4\,(\mathrm{lum.})\,\mathrm{pb}. \tag{6.9}$$

Table 6.9 lists the systematic uncertainties in the signal strength after the fit. The effect of the uncertainties on the signal strength is estimated by a likelihood

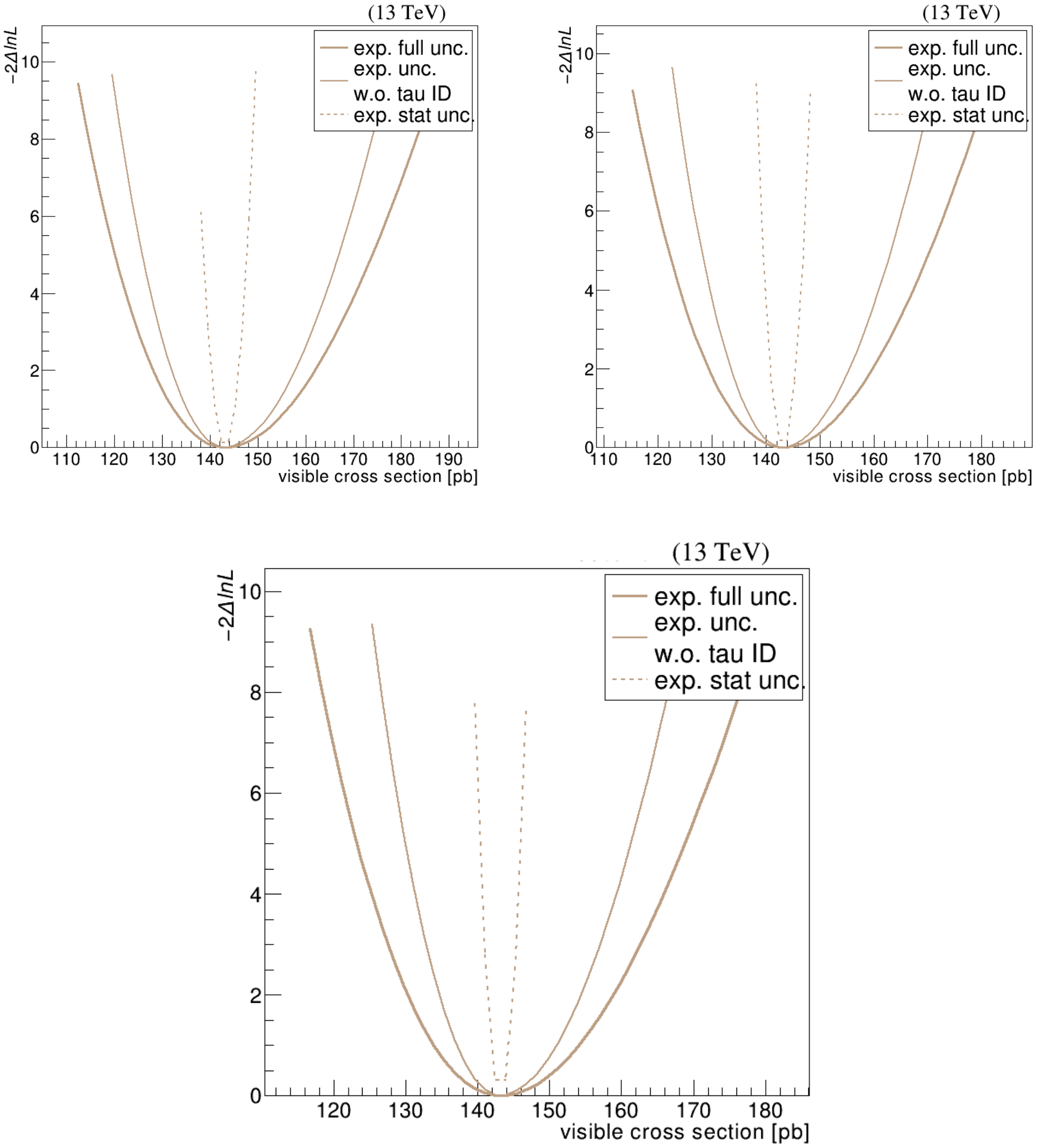

Figure 6.27: Scan of expected profile likelihood ratio in the $e\tau_{\mathrm{h}}$ (top-left), $\mu\tau_{\mathrm{h}}$ (top-right), and both final states combined (bottom). The simultaneous fit is performed by keeping nuisance parameters and signal strength equal across all event categories and final states. The scan with full uncertainty is shown with bold solid line. The impact of the τ_{h} identification and misidentification uncertainties is shown by the thin solid line. It is produced by setting the τ_{h}-related nuisance parameters to their post-fit values. A scan of the statistical uncertainty, shown with a dashed line, is produced by setting all nuisance parameters to their post-fit values.

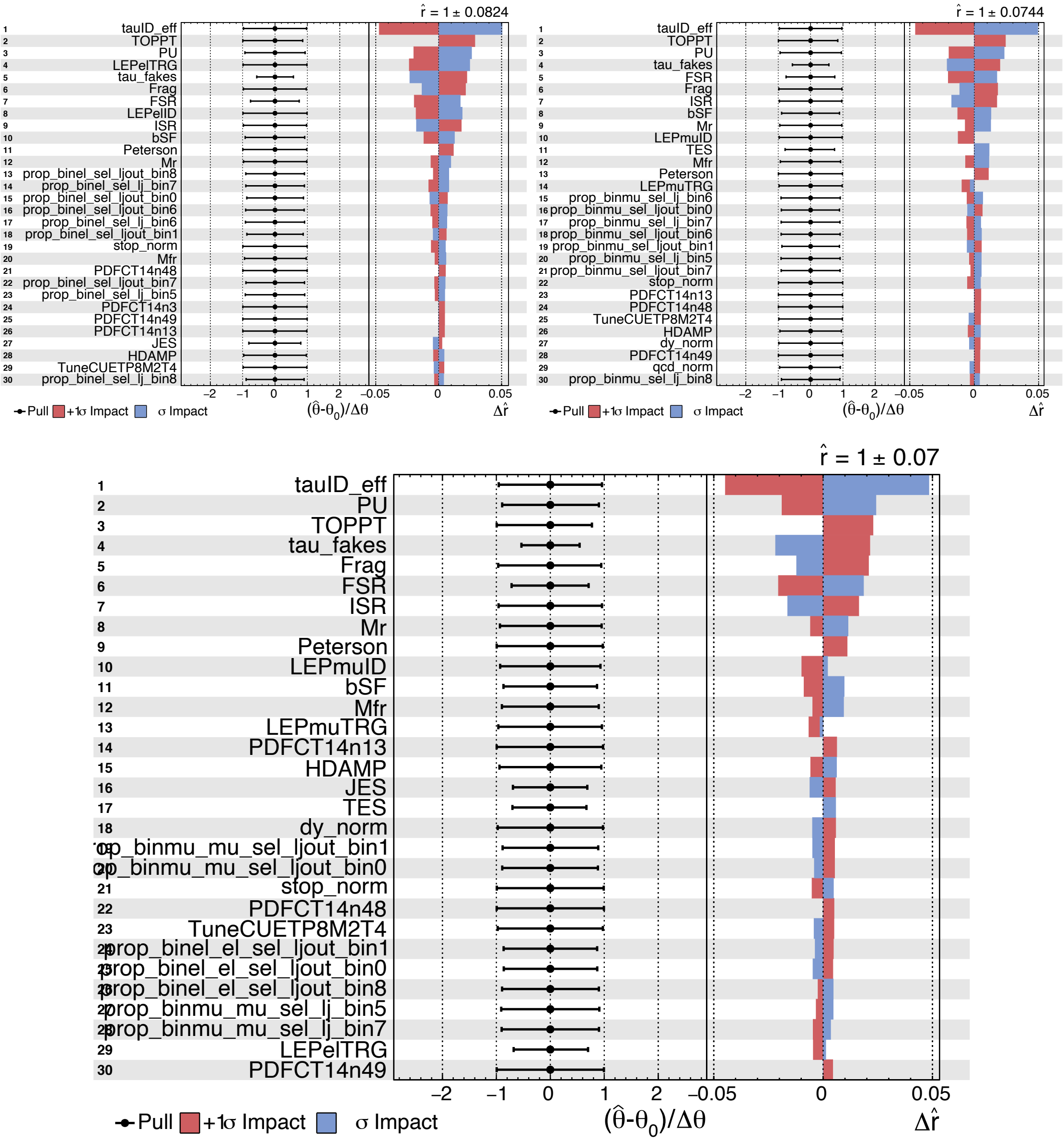

Figure 6.28: Expected impacts of nuisance parameters on the signal strength in the $e\tau_{\mathrm{h}}$ (top-left), $\mu\tau_{\mathrm{h}}$ (top-right), and both final states combined (bottom). The simultaneous fit is performed by keeping nuisance parameters and signal strength equal across all event categories and final states.

scan where only one nuisance parameter (or a group of them) is varied at once while the others are fixed to their nominal postfit values. The largest experimental uncertainties are from τ_h identification and misidentification, and pileup estimation. The largest theoretical uncertainties are due to the modelling of top quark p_T in $\mathrm{t\bar{t}}$ processes, b quark fragmentation, and PS modelling (ISR and FSR).

The likelihood scan of the fit to the observed data in the fiducial phase-space is shown in Figure 6.29 for the $\mathrm{e}\tau_\mathrm{h}$ and $\mu\tau_\mathrm{h}$ final states separately, and for the simultaneous fit in both final states. It is shown in comparison with the expectations. Overall, the fit is in agreement with the SM expectations.

Figure 6.30 reports the values of post-fit nuisances in separate final states. Their impacts on the signal strength are reported in Figure 6.31, for the $\mathrm{e}\tau_\mathrm{h}$, $\mu\tau_\mathrm{h}$, and both final states combined. The breakdown of the systematic uncertainties in the signal strength fitted to the data is given in Table 6.9.

The correlations of the nuisance parameters in the fit are shown in Figure 6.32.

The post-fit distributions with comparison to the data are shown in Figure 6.33. The goodness of the fit tests are reported in Table 6.10.

6.7.2 Cross section in the full phase space

The fiducial $\mathrm{t\bar{t}}$ cross section is extrapolated to the full phase space with the detector acceptance $A_\mathrm{t\bar{t}}$:

$$\sigma_\mathrm{t\bar{t}} = \sigma^\mathrm{vis}_\mathrm{t\bar{t}} / A_\mathrm{t\bar{t}} = (N - B)/(\mathcal{L}\epsilon\mathcal{B}A_\mathrm{t\bar{t}}). \tag{6.10}$$

The acceptance $A_\mathrm{t\bar{t}}$ is the fraction of signal events produced in the fiducial phase space, and it is determined with respect to all signal events in the nominal $\mathrm{t\bar{t}}$ simulation. It includes kinematic selection cuts and is evaluated for the different signal final states as:

$$A_\mathrm{t\bar{t}}(\mathrm{e}\tau_\mathrm{h}) = 0.1687 \pm 0.0004\,(\mathrm{stat.}) \pm 0.0060\,(\mathrm{syst.}), \tag{6.11}$$

$$A_\mathrm{t\bar{t}}(\mu\tau_\mathrm{h}) = 0.1756 \pm 0.0004\,(\mathrm{stat.}) \pm 0.0065\,(\mathrm{syst.}), \tag{6.12}$$

$$A_\mathrm{t\bar{t}}(\ell\tau_\mathrm{h}) = 0.1722 \pm 0.0003\,(\mathrm{stat.}) \pm 0.0062\,(\mathrm{syst.}), \tag{6.13}$$

where the systematic uncertainties include the modelling uncertainties that are described in Section 6.5 and listed as “Extrapolation uncertainties” in Table 6.9.

The cross section values in the full phase space are obtained from the extrapolation of the fiducial cross sections using the acceptances $A_\mathrm{t\bar{t}}$ estimated from the

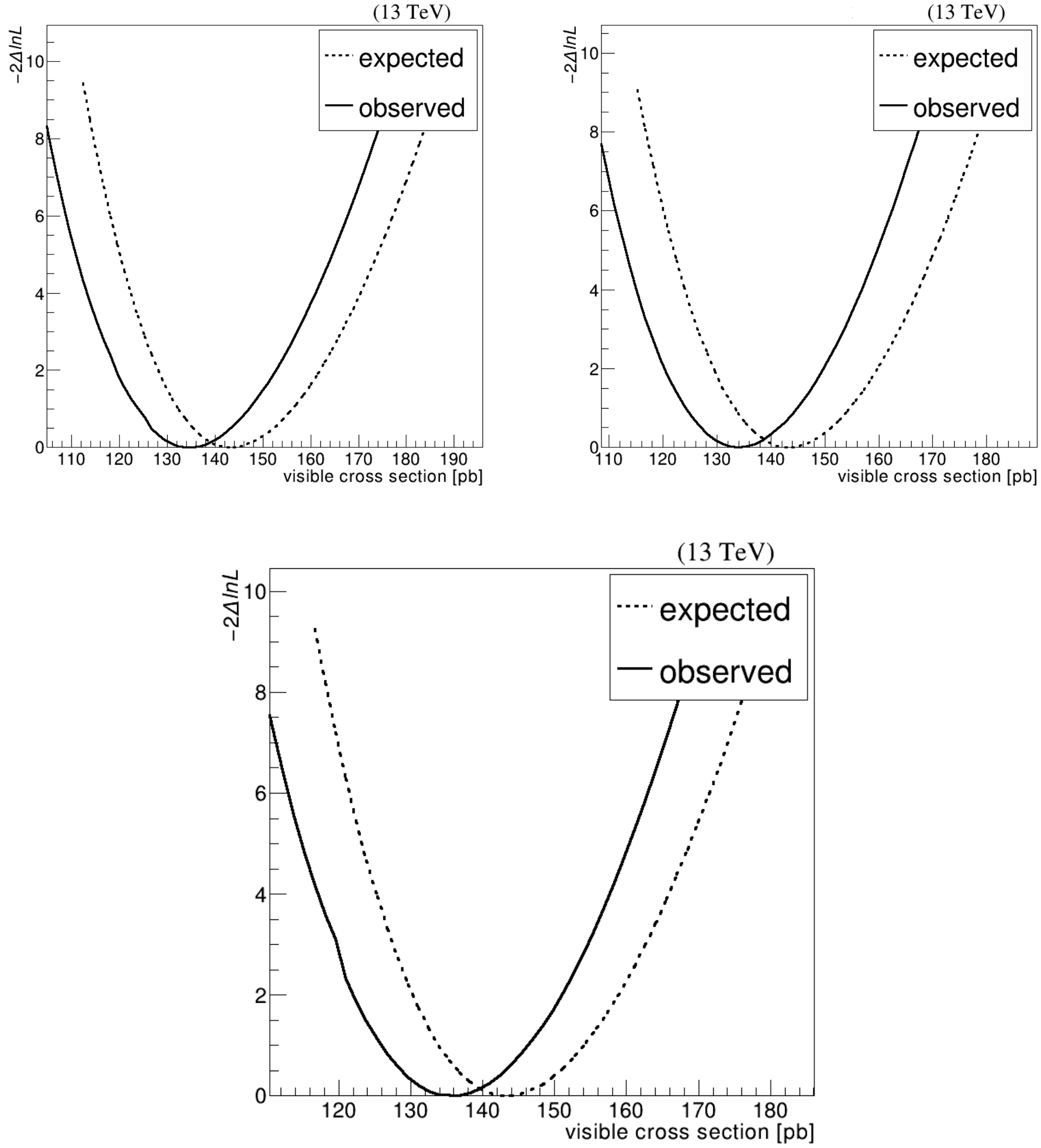

Figure 6.29: Scans of the expected and observed profile likelihood ratios in the $e\tau_\mathrm{h}$ (top-left), $\mu\tau_\mathrm{h}$ (top-right), and both final states combined (bottom).

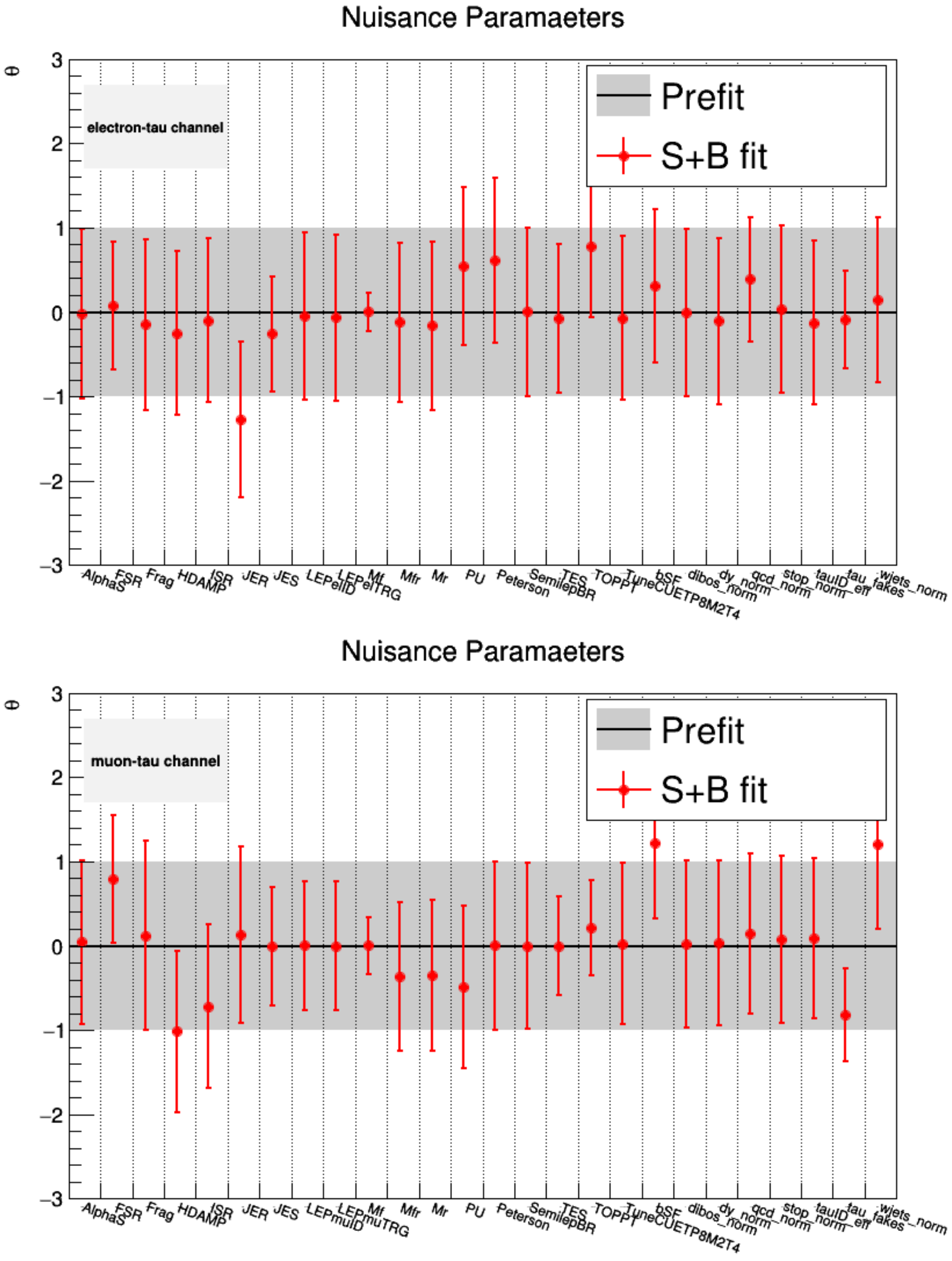

Figure 6.30: Post-fit values of main nuisance parameters in the $e\tau_\mathrm{h}$ (top) and $\mu\tau_\mathrm{h}$ (bottom) final states.

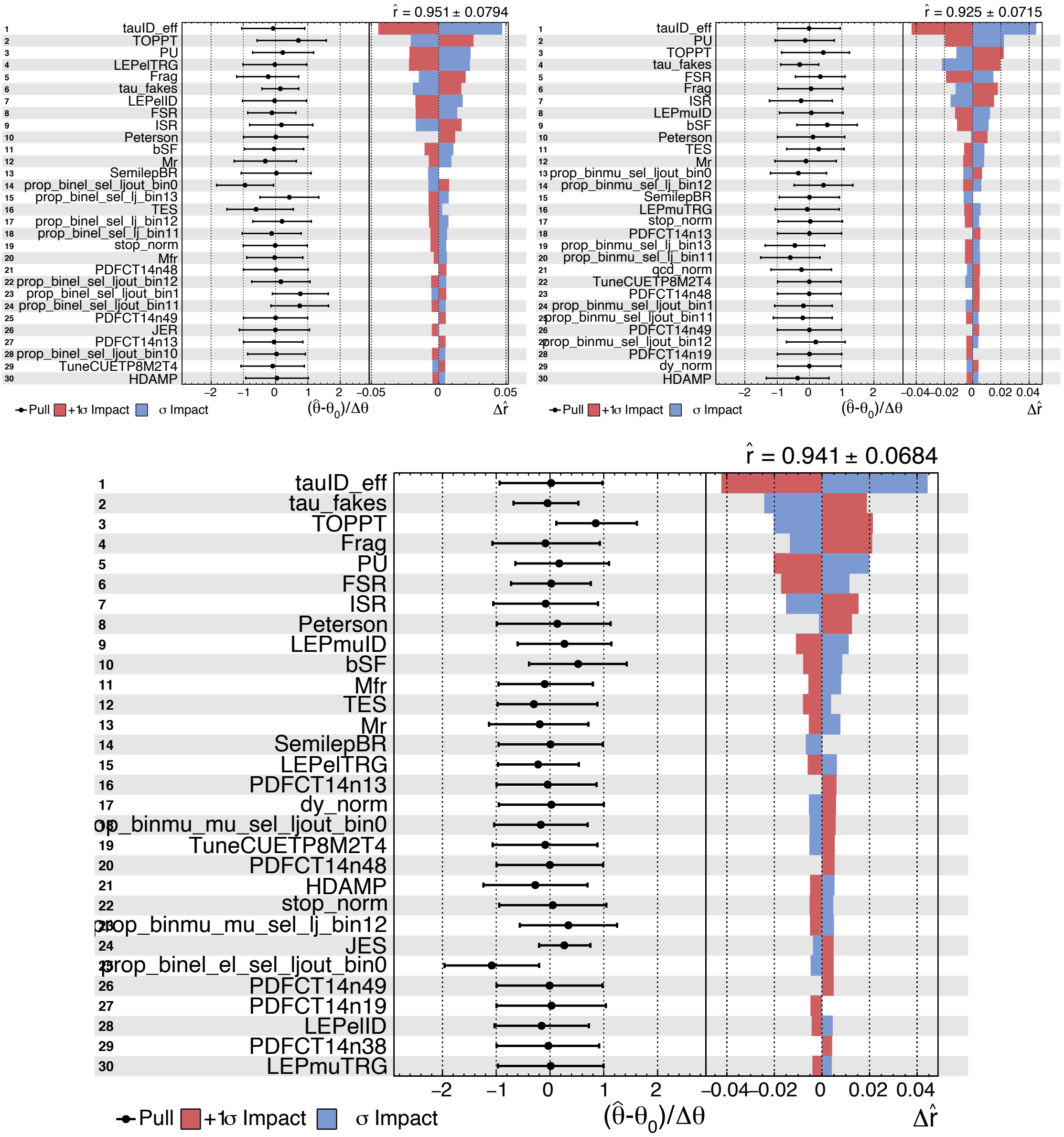

Figure 6.31: Impacts of the nuisance parameters on the signal strength fitted to the observed data in the $e\tau_h$ (top-left), $\mu\tau_h$ (top-right), and both final states combined (bottom).

Table 6.9: Systematic and statistical uncertainties determined from the fit to the data in the $\mathrm{e}\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ final states, and their combination. Uncertainties are grouped by their origin: experimental, theoretical, normalization, and extrapolation. The uncertainties in the measurement in the dilepton final state [2] used in the partial width ratio estimate are also quoted (column "Dileptons"), where the asymmetric extrapolation uncertainties are symmetrized by adding them in quadrature. As both measurements use the same data, some uncertainties in the $\ell\tau_{\mathrm{h}}$ and light dilepton final states are correlated, as shown in the last column.

Source	Uncertainty [%]				
	$\mathrm{e}\tau_{\mathrm{h}}$	$\mu\tau_{\mathrm{h}}$	Combined	Dileptons	Correlation
Experimental uncertainties					
τ_{h} jet identification	4.7	4.5	4.5		0
τ_{h} jet misidentification	2.2	2.3	2.3		0
Pileup	2.5	2.2	2.3	0.1	1
Lepton identification and isolation	1.8	1.1	1.2	2.0	1
b tagging efficiency	1.1	1.2	0.9	0.4	1
τ_{h} energy scale	0.7	0.8	0.8		0
Trigger efficiency	2.3	0.6	0.7	0.3	0
Drell–Yan background	0.4	0.4	0.6	0.9	1
$\mathrm{t\bar{t}}$ background	1.0	0.8	0.6	0.2	0
tW background	0.6	0.5	0.5	1.1	1
W+jets background	0.1	0.4	0.5	0.2	0
Multijet background	0.1	0.5	0.4	<0.1	0
Jet energy scale	0.1	0.2	0.4	0.4	1
Jet energy resolution	0.6	0.3	0.1	0.4	1
Electron momentum scale	0.1	0.1	0.1	0.1	1
Muon momentum scale	0.1	0.1	0.1	0.1	1
Diboson background	<0.1	<0.1	<0.1	0.2	1
Theoretical uncertainties					
b fragmentation	2.3	2.0	2.4	0.7	1
Top quark p_{T} modelling	2.7	2.3	2.2	0.5	1
$\mathrm{t\bar{t}}$ FSR scale	1.7	1.9	1.7	0.8	1
tW FSR scale	<0.1	<0.1	<0.1	0.1	1
$\mathrm{t\bar{t}}$ ISR scale	1.7	1.6	1.5	0.4	1
tW ISR scale	<0.1	<0.1	<0.1	0.1	1
$\mathrm{t\bar{t}}$ ME scale	1.1	1.2	1.1	0.2	1
tW ME scale	<0.1	<0.1	<0.1	0.2	1
Drell–Yan ME scale	<0.1	<0.1	<0.1	0.1	1
Semileptonic b hadron branching fraction	0.8	0.6	0.7	0.1	1
Underlying event	0.5	0.5	0.6	0.3	1
ME-PS matching	0.4	0.4	0.5	0.2	1
Colour reconnection	<0.1	<0.1	<0.1	0.3	1
PDFs	1.5	1.5	1.6	1.1	1
Normalization uncertainties					
Statistical	1.4	1.1	0.9	0.2	0
MC statistical	2.0	1.6	1.6	1.1	0
Integrated luminosity	2.5	2.5	2.5	2.5	1
Extrapolation uncertainties					
$\mathrm{t\bar{t}}$ ME scale	0.3	0.4	0.3	0.3	0
PDFs	1.2	1.4	1.3	1.0	0
Top quark p_{T} modelling	1.0	1.1	1.1	0.5	0
$\mathrm{t\bar{t}}$ ISR scale	0.5	0.3	0.3	0.1	0
$\mathrm{t\bar{t}}$ FSR scale	1.9	2.0	1.9	0.1	0
Underlying event	0.3	0.2	0.2	<0.1	0

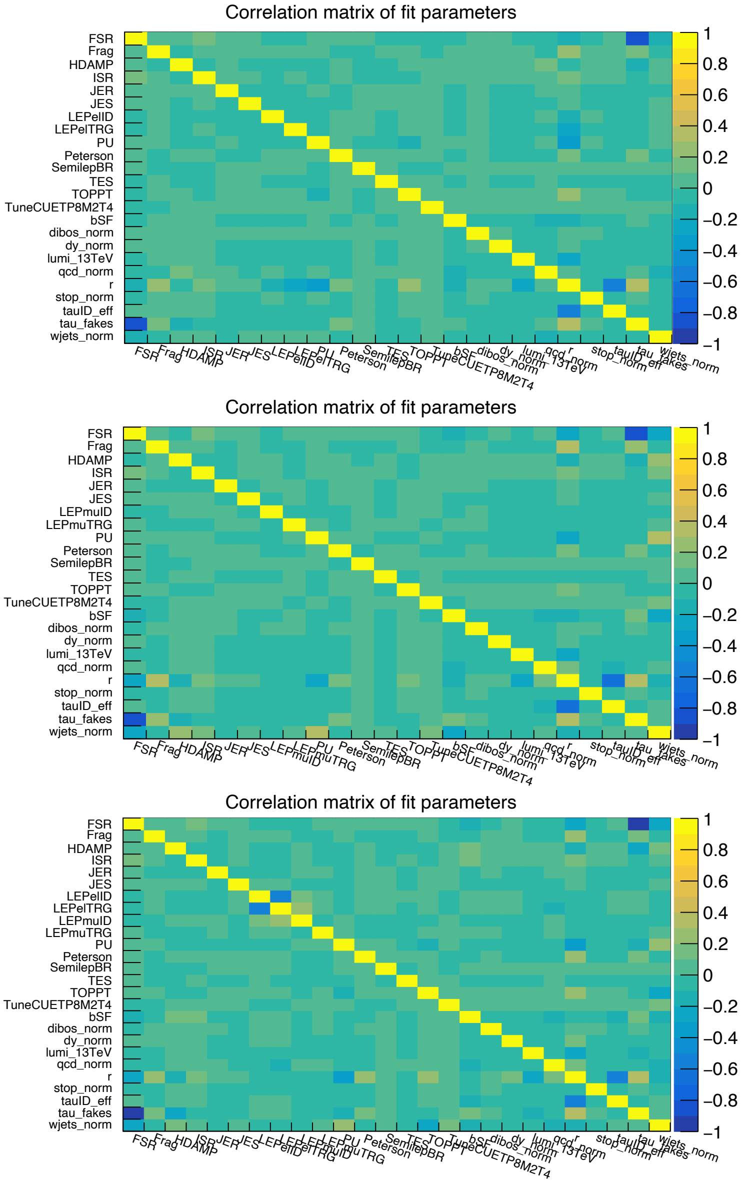

Figure 6.32: Covariances of the post-fit values of main nuisance parameters in the eτ_h (top), $\mu\tau_\mathrm{h}$ (middle) and combined $\ell\tau_\mathrm{h}$ (bottom) final states. The omitted parameters correspond to the uncertainties due to the MC statistics, PDF and scale variations make a small contribution to the overall uncertainty, and do not exhibit significant correlations with other parameters. Expected correlation groups are seen in the parameters: the lepton ID and trigger efficiency; the signal strength, τ_h ID efficiency and τ_h misidentification probability.

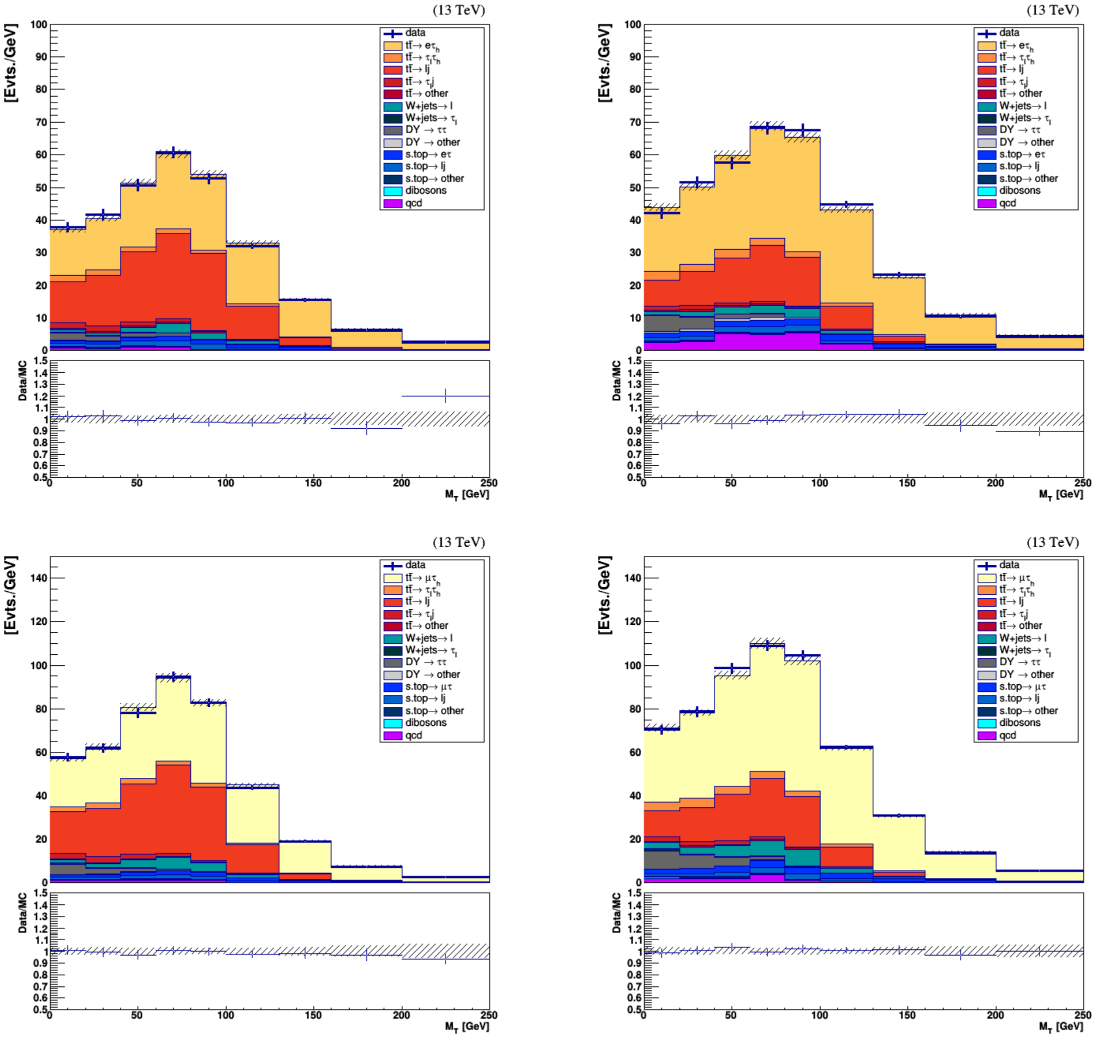

Figure 6.33: Post-fit distributions of the m_T with the nuisance parameters set to the values found in the fit to both channels: the background-rich (left) and signal-rich event categories (right), in the eτ_h (top) and $\mu\tau_\mathrm{h}$ (bottom) final states.

Table 6.10: The p-values of the goodness of fit tests in the separate channels and in the simultaneous fit. Three tests show comparable results: Kolmogorov-Smirnov, Anderson-Darling and the saturated model.

Method	p-value [%]		
	$\mathrm{e}\tau_\mathrm{h}$	$\mu\tau_\mathrm{h}$	Combined
Kolmogorov-Smirnov	11	25	24
Anderson-Darling	5	57	17
saturated model	9	37	16

simulation:

$$\sigma_{\mathrm{t\bar{t}}}(\mathrm{e}\tau_\mathrm{h}) = 789 \pm 11\,(\text{stat.}) \pm 71\,(\text{syst.}) \pm 20\,(\text{lum.})\,\text{pb}, \tag{6.14}$$

$$\sigma_{\mathrm{t\bar{t}}}(\mu\tau_\mathrm{h}) = 770 \pm 8\,(\text{stat.}) \pm 63\,(\text{syst.}) \pm 20\,(\text{lum.})\,\text{pb}, \tag{6.15}$$

$$\sigma_{\mathrm{t\bar{t}}}(\ell\tau_\mathrm{h}) = 781 \pm 7\,(\text{stat.}) \pm 62\,(\text{syst.}) \pm 20\,(\text{lum.})\,\text{pb}. \tag{6.16}$$

The expected and observed dependence of the likelihood on the cross section in the full phase space in the $\ell\tau_\mathrm{h}$ combined final state are shown in Figure 6.34. The result of the fit is consistent with the predicted SM $\mathrm{t\bar{t}}$ production cross section of $832^{+20}_{-29}\,(\text{scale}) \pm 35\,(\text{PDF}+\alpha_S)$ pb [29]. Using simulated $\mathrm{t\bar{t}}$ samples with different m_t values, we find that the cross section changes by 1.5% per $\Delta m_\mathrm{t} = 1\,\text{GeV}$.

6.7.3 Ratio to the dilepton cross section and partial width

The ratio of the cross section in the $\ell\tau_\mathrm{h}$ final state divided by the cross section measured in the light dilepton final state [2] yields a value of $R_{\ell\tau_\mathrm{h}/\ell\ell} = 0.973 \pm 0.009\,(\text{stat.}) \pm 0.066\,(\text{syst.})$, consistent with unity as expected from lepton flavour universality. Since both measurements are performed in the same data-taking period with the same reconstruction algorithms, the uncertainty in the ratio includes the correlations between common sources of uncertainties as indicated in Table 6.9. The relative systematic uncertainty in the ratio is 6.8%. About 5% comes from the uncertainties in the τ_h identification efficiency (4.5%) and misidentification probability (2.3%) in the $\mathrm{t\bar{t}}$ events. The rest comes from the uncorrelated uncertainties in the ratio and the treatment of the correlated uncertainties in the calculation of the ratio. In particular, the triggers are not the same, and a small contribution comes from the uncertainties in the extrapolation to the full phase space that are considered uncorrelated because the two measurements extrapolate from different fiducial phase spaces.

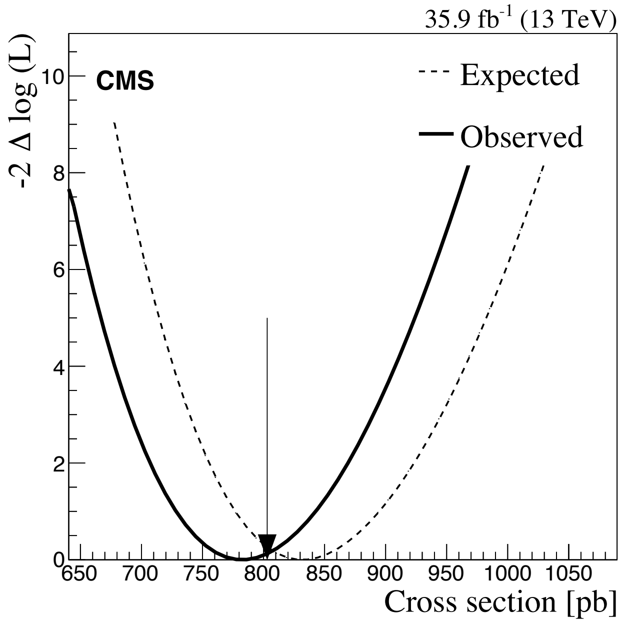

Figure 6.34: The expected and observed dependence of the likelihood on the total $\mathrm{t\bar{t}}$ cross section $\sigma_{\mathrm{t\bar{t}}}$. It is derived from the fiducial phase space by a simple extrapolation. The arrow points at the cross section measured in the light dilepton final state. The goodness of the fit determined with a Kolmogorov–Smirnov method yields a p value of 0.24.

The measurement also provides an estimate of the ratio of the partial to the total width of the top quark decay, $R_\Gamma = \Gamma(\mathrm{t} \to \tau\nu_\tau \mathrm{b})/\Gamma_{\mathrm{total}}$. The ratio is calculated as $R_\Gamma = \sigma_{\mathrm{t\bar{t}}}(\ell\tau_{\mathrm{h}})\mathcal{B}(\mathrm{W} \to \tau\nu_\tau)/\sigma_{\mathrm{t\bar{t}}}(\ell\ell)$, where the cross section measured in the $\ell\tau_{\mathrm{h}}$ final state is multiplied by the branching fraction $\mathcal{B}(\mathrm{W} \to \tau\nu_\tau)$ and divided by the inclusive $\mathrm{t\bar{t}}$ cross section measured in the dilepton final state [2]. The W boson branching fraction $\mathcal{B}(\mathrm{W} \to \tau\nu_\tau)$ that is included in the signal acceptance is cancelled out in the multiplication. The estimate yields the value $R_\Gamma = 0.1050 \pm 0.0009\,(\mathrm{stat.}) \pm 0.0071\,(\mathrm{syst.})$, improving over the previous measurements [22, 21, 4]. The result is dominated by the systematic uncertainty and it is consistent with the SM value of 0.1083 ± 0.0002 [4]. While in Ref. [22] the partial width is evaluated for hadronic decays of τ leptons, here R_Γ is measured for all τ decays by using the $\mathcal{B}(\tau \to \tau_{\mathrm{h}}\nu_\tau) = 64.8 \pm 0.1\%$ branching fraction [4].

6.8 Summary

A measurement of the top quark pair production cross section in the $\mathrm{t\bar{t}} \to (\ell\nu_\ell)(\tau_{\mathrm{h}}\nu_\tau)\mathrm{b\bar{b}}$ channel, where ℓ is either an electron or a muon, is performed in proton-proton collisions at LHC, using a data sample corresponding to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$ obtained at $\sqrt{s} = 13\,\mathrm{TeV}$ [120]. Events are selected by requiring the presence of an electron or a muon, and at least three jets, of which at least one is b tagged and one is identified as a τ lepton decaying to hadrons (τ_{h}). The largest background contribution arises from $\mathrm{t\bar{t}}$ lepton+jets events, $\mathrm{t\bar{t}} \to (\ell\nu_\ell)(\mathrm{q\bar{q}'})\mathrm{b\bar{b}}$, where one jet is misidentified as the τ_{h}. The background contribution is constrained in a fit to the distribution of the transverse mass of the light lepton and missing transverse momentum system in two event categories, constructed according to the kinematic properties of the jets in the $\mathrm{t\bar{t}}$ lepton+jets final state. The signal is included in the fit as a free parameter without constraining the kinematic properties of the τ lepton. Assuming a top quark mass of 172.5 GeV, the measured total $\mathrm{t\bar{t}}$ cross section $\sigma_{\mathrm{t\bar{t}}}(\ell\tau_{\mathrm{h}}) = 781 \pm 7\,(\mathrm{stat.}) \pm 62\,(\mathrm{syst.}) \pm 20\,(\mathrm{lum.})\,\mathrm{pb}$ is in agreement with the standard model expectation. This is the first measurement of the $\mathrm{t\bar{t}}$ production cross section in proton-proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ that explicitly includes hadronically decaying τ leptons, and it improves the relative precision with respect to the 7 and 8 TeV results [18, 19]. The higher precision is achieved through a shape fit to the kinematic distributions of the events, thus better constraining the backgrounds. However, the result is still limited by the systematic uncertainty. Therefore, it cannot be improved by simply adding more data. The measurement of the ratio of the cross section in the $\ell\tau_{\mathrm{h}}$ final state to the light dilepton cross section [2] yields a value of $R_{\ell\tau_{\mathrm{h}}/\ell\ell} = 0.973 \pm 0.009\,(\mathrm{stat.}) \pm 0.066\,(\mathrm{syst.})$, consistent with lepton universality. The ratio of the partial to the total width of the top quark $\Gamma(\mathrm{t} \rightarrow \tau\nu_\tau \mathrm{b})/\Gamma_{\mathrm{total}} = 0.1050 \pm 0.0009\,(\mathrm{stat.}) \pm 0.0071\,(\mathrm{syst.})$ is measured with respect to the $\mathrm{t}\bar{\mathrm{t}}$ inclusive cross section extrapolated from the light dilepton final state, improving the precision over the previous measurements [21, 22].

Chapter 7

The lepton universality test in top quark pair events

The lepton flavour universality principle in the standard model postulates that the interactions with all three families of leptons (electrons, muons, taus) happen at the same rate, the three lepton families differ only by their masses. The current world-best tests confirm this principle to a high precision with few deviations. A notable example is the deviation of about $+2.5\sigma$ from unity in the ratio $\mathcal{B}(\mathrm{W} \to \tau\nu_\tau)/\mathcal{B}(\mathrm{W} \to \ell\nu_\ell)$ of W boson decays, where ℓ is either e or μ [4]. The measurements of the W boson branching fractions were carried out with 1-2% precision at LEP [121] in diboson channels. The same deviation is suggested by recent measurements of $\mathrm{b} \to \mathrm{c}\ell^-\overline{\nu}_\ell$ decays from Babar, Belle, and LHC (LHCb) [9, 122], which observe other anomalies in lepton flavour measurements as well.

The large luminosity and energy of the pp collisions data at LHC open for the first time a possibility for a precise measurement of this ratio in $\mathrm{t\bar{t}}$ decays. The ratio between the W boson branching fractions can be measured from the ratio between the numbers of events observed in the $\mathrm{t\bar{t}}$ dilepton final states with and without the τ lepton, shown in Figure 7.1. In this chapter I present methods to carry out the corresponding analysis in CMS data and employ them in a feasibility study for the full Run2 dataset. The results show that a significant contribution can be made to the current world-best measurements of the lepton flavour universality in W boson decays.

A test of lepton flavour universality at LHC is an excellent topic from both experimental and theoretical points of view: a precise measurement at a hadron accelerator that can improve the current world-best average beyond 1-2% uncertainty, that would contribute to the understanding of the lepton flavour universality hy-

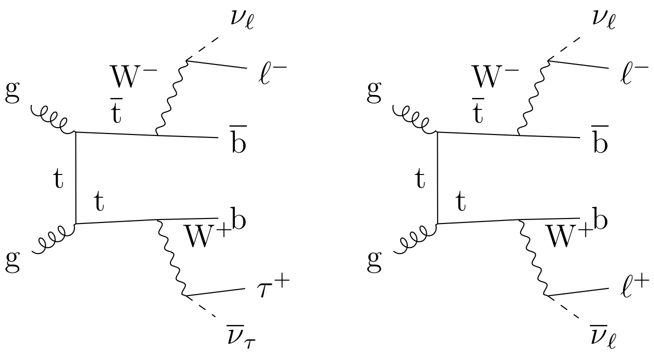

Figure 7.1: Feynman diagrams of the $\mathrm{t\bar{t}}$ decays into dilepton final states: including a τ lepton (left), and with only light leptons (right).

pothesis in the SM. If the central value of a new high-precision measurement were equal to the central value of the LEP2 measurement, the world-average would yield an observation of a BSM contribution to the W boson decays with a high statistical significance.

The feasibility of the same measurement has been studied analytically by the ATLAS collaboration [123]. There is a new measurement of the ratio between leptonic branching fractions of the W boson from ATLAS [**?**] that was performed in leptonic τ decays. A measurement in hadronic τ_h decays is more robust since it is not biased to the signal model. It would make a significant contribution to the current results on the lepton flavour universality.

In the following the general strategy of the measurement is presented in Section 7.1; additional requirements to the reconstructed objects are listed in Section 7.2; the event selection requirements are given in Section 7.3; Section 7.4 presents the treatment of the specific systematic uncertainties; the fitting procedure is discussed in Section 7.5; the results are in Section 7.6 with the discussion in Section 7.7 .

7.1 Strategy

Under the assumption of the SM $\mathrm{t\bar{t}}$ decay, the ratios between W boson branching fractions are extracted from the ratio of the event yields N in $\mathrm{t\bar{t}}$ dilepton final states with and without a τ lepton:

$$\frac{N(\mathrm{t\bar{t}} \to \ell\tau_\mathrm{h}\nu_\ell\nu_\tau \mathrm{b\bar{b}})}{N(\mathrm{t\bar{t}} \to \ell\ell\nu_\ell\nu_\ell \mathrm{b\bar{b}})} \approx \frac{\mathcal{B}(\mathrm{W} \to \tau\nu_\tau)}{\mathcal{B}(\mathrm{W} \to \ell\nu_\ell)}. \tag{7.1}$$

The events in the two $\mathrm{t\bar{t}}$ dilepton final states are selected with the same requirements, except for the second lepton: a τ_h or a light lepton. The systematic uncertainties that correspond to the same requirements cancel out in the ratio. The largest uncertainty of about 5% in the τ_h identification propagates to the final result.

In order to constrain this uncertainty, a symmetric double ratio with Drell–Yan final states is used:

$$\frac{N(\mathrm{t\bar{t}} \to \ell\tau_\mathrm{h}\nu_\ell\nu_\tau\mathrm{b\bar{b}})}{N(\mathrm{t\bar{t}} \to \ell\ell\nu_\ell\nu_\ell\mathrm{b\bar{b}})}\frac{N(\mathrm{DY} \to \ell\ell)}{N(\mathrm{DY} \to \tau_\ell\tau_\mathrm{h})} \approx \frac{\mathcal{B}(\mathrm{W} \to \tau\nu_\tau)}{\mathcal{B}(\mathrm{W} \to \ell\nu_\ell)}\frac{\mathcal{B}(\mathrm{DY} \to \ell\ell)}{\mathcal{B}(\mathrm{DY} \to \tau_\ell\tau_\mathrm{h})}. \tag{7.2}$$

In this construction, the ratio of the W boson branching fractions is measured with respect to the ratio of the Z boson branching fractions, that have been measured to a very high precision [4] and do not introduce a noticeable uncertainty in the result. At the same time, if the statistical fluctuations are negligible and the background contributions to the selected events are small, the experimental and theoretical uncertainties cancel out in the double ratio, and the ratio between W boson branching fractions can be measured to a high accuracy.

Therefore, four sets of requirements are needed to select events in four dilepton final states in $\mathrm{t\bar{t}}$ and DY, designated as: $\mathrm{t\bar{t}} \to \mathrm{e}\tau_\mathrm{h}$, $\mathrm{t\bar{t}} \to \mathrm{e}\mu$, and $\mathrm{DY} \to \tau_\mu\tau_\mathrm{h}$, $\mathrm{DY} \to \mu\mu$. The muon final states are chosen because of the higher reconstruction efficiency and lower p_T threshold in the single muon trigger. Since the leptons from the DY decay have low energies in comparison with the single-lepton HLT thresholds, the final state with μ leptons is selected in the DY events. In order to have a pure selection of light dilepton $\mathrm{t\bar{t}}$ events, the $\mathrm{t\bar{t}}$ final state with an electron and a muon is chosen. In order to cancel out the systematic uncertainties that correspond to light leptons in the double ratio, the $\mathrm{t\bar{t}}$ final state with a τ_h is selected with an electron as the light lepton. In order to have pure selections of events with τ_h, the τ_h candidates are required to have 3 prongs with a well reconstructed secondary vertex that corresponds to the τ decay.

The events are selected with the same requirements as those in the measurement of the $\mathrm{t\bar{t}}$ production cross section. The complete selection requirements are discussed in Section 7.3. The resulting selections with light dilepton events consist almost entirely of signal events. The selection of $\mathrm{DY} \to \tau_\mu\tau_\mathrm{h}$ events contains approximately 90% of DY events. The remaining 10% are from W+jets events with a jet misidentified as a τ_h. The selection of $\mathrm{t\bar{t}} \to \mathrm{e}\tau_\mathrm{h}$ events has a significant background with misidentified τ_h from the $\mathrm{t\bar{t}} \to \mathrm{e}\nu\mathrm{q\bar{q}'b\bar{b}}$ process. To constrain this remaining background in the final states with τ_h, the double ratio is measured using a shape fit to the m_T distribution, as discussed in Section 7.5.

7.2 Object Definitions

The object definitions are the same as in the measurement of the top quark pair production cross section described in Chapter 6, except for a tighter selection of τ_h candidates and reduced τ_h p_T threshold. In order to get a pure sample of $\mathrm{DY} \to \tau_\ell\tau_\mathrm{h}$ events and improve the selection of $\mathrm{t\bar{t}} \to \ell\tau_\mathrm{h}$ events, only the 3-prong τ_h candidates that have a well reconstructed secondary vertex (SV) are considered.

The selection of 3-prong τ_h candidates is the only significant change to the standard object definitions given in Chapter 4. The three tracks of the charged particles reconstructed in the detector allow the determination of the point of the τ lepton decay. With an energy of about 30 GeV, the τ lepton propagates for about 2 mm before the decay that is observed as a SV in the pp collision event. The SV is characterized by the distance from the primary vertex (PV) of the collision, the length of flight. The length of flight has a positive or negative sign, that is determined by the projection of the τ lepton flight direction on the τ_h momentum direction. The statistical significance of the length of flight is given by the ratio between the length of flight and the reconstruction uncertainty of the PV and SV projected onto the flight direction of the τ lepton. The significance represents the length of flight measured in the units of the reconstruction uncertainty. The τ_h candidates are required to have a length of flight significance of 3 or more.

Most of hadronic jets originate from light quarks or gluons that hadronize immediately without forming long-lived hadrons, and do not feature a displaced vertex in the final state tracks. Therefore a requirement of a significant SV discards many misidentified τ_h candidates from jets radiated in the multijet QCD, W+jets, DY+jets processes.

The misidentified τ_h candidates that originate from b flavour jets have a genuine SV and cannot be removed with this requirement. But the background of the b flavour jets in $\mathrm{t\bar{t}}$ events can be constrained by other means: with a high-efficiency requirement of two b-tagged jets, or by exploiting the symmetry between the opposite and same charge selections of the light lepton and the τ_h candidate in the event. Since the $\mathrm{t\bar{t}}$ processes produce the b quarks in pairs, the misidentified τ_h candidates that originate from these quarks have either positive or negative charge with the same probability.

A more cumbersome background of misidentified τ_h originates from the c flavour jets. Background processes with such misidentified τ_h candidates are present in both $\mathrm{t\bar{t}}$ and DY event selections. The main background in the $\mathrm{t\bar{t}} \to \mathrm{e}\tau_\mathrm{h}$ event selection, the $\mathrm{t\bar{t}} \to \mathrm{e}\nu\mathrm{q\bar{q}'b\bar{b}}$ process, has a high probability to produce a c quark in the hadronic W boson decay. The $\mathrm{DY} \to \tau_\ell\tau_\mathrm{h}$ selection includes a contribution from the W+jets process that can produce a c quark in association with the W boson. The jets that originate from the c quarks also have a genuine SV and a significant probability to be misidentified as a τ_{h} candidate. The background processes produce the light leptons and the c quarks with opposite charges, mimicking the signal even more. However, the contributions of the backgrounds with c jets are small, and a precise measurement in the double ratio can be achieved without removing these backgrounds completely.

7.2.1 The secondary vertex reconstruction

The τ_{h} candidates can have poor quality SVs in the reconstructed datasets, because the SV is reconstructed for the τ_{h} seed jet, before the τ identification algorithm is applied and the jet particles are separated into the signal candidates that correspond to the τ_{h} decay and the background candidates. Several physical features of the τ_{h} SV can validate whether the SV is reconstructed correctly: the significance of the reconstructed length of flight is expected to have values between -1 and 20 with a peak around 1 for true τ_{h} and a negligible amount of misidentified τ_{h} above 3, the average length of flight must be about 2 mm, and the length of flight must correlate with the energy of the τ_{h} candidate. Figure 7.2 shows the distributions of these parameters for the genuine τ_{h} candidates in the $\mathrm{t\bar{t}}$ MC datasets. The average length of flight is less than 0.2 mm and the significance values are up to 50. It indicates an incorrect behavior of the reconstruction algorithm and a poor quality of the τ_{h} SV.

In order to validate the correctness of the SV reconstruction and improve the separation between the signal and the background, an additional algorithm was applied offline to reproduce the SV from the tracks already identified as τ_{h} candidates. The algorithm reconstructs the SV following the geometry of the $\tau \to \tau_{\mathrm{h}}\nu_{\tau}$ decay, and it takes into account that the precision of the CMS tracker is better in the transverse plane than along the beam axis.

The three tracks that make up the τ_{h} candidate converge in the SV point. But in order to improve the reconstruction of the SV, two restrictions of the geometry of the decay are considered: the three tracks of the τ_{h} are perpendicular with the vectors of their impact parameters, and each track with the corresponding impact parameter vector lie in one plane with the line of the τ flight between the PV and the SV. With these restrictions in mind, the first step of the algorithm corrects the impact parameters according to the signal geometry. It determines the line of the τ lepton flight by minimizing the sum of the angles between the tracks and their corresponding impact parameters in the transverse plane with respect to the flight line. With a perfect reconstruction this sum must be equal to zero. The minimization starts by assuming the direction of the τ flight to be the same as the reconstructed

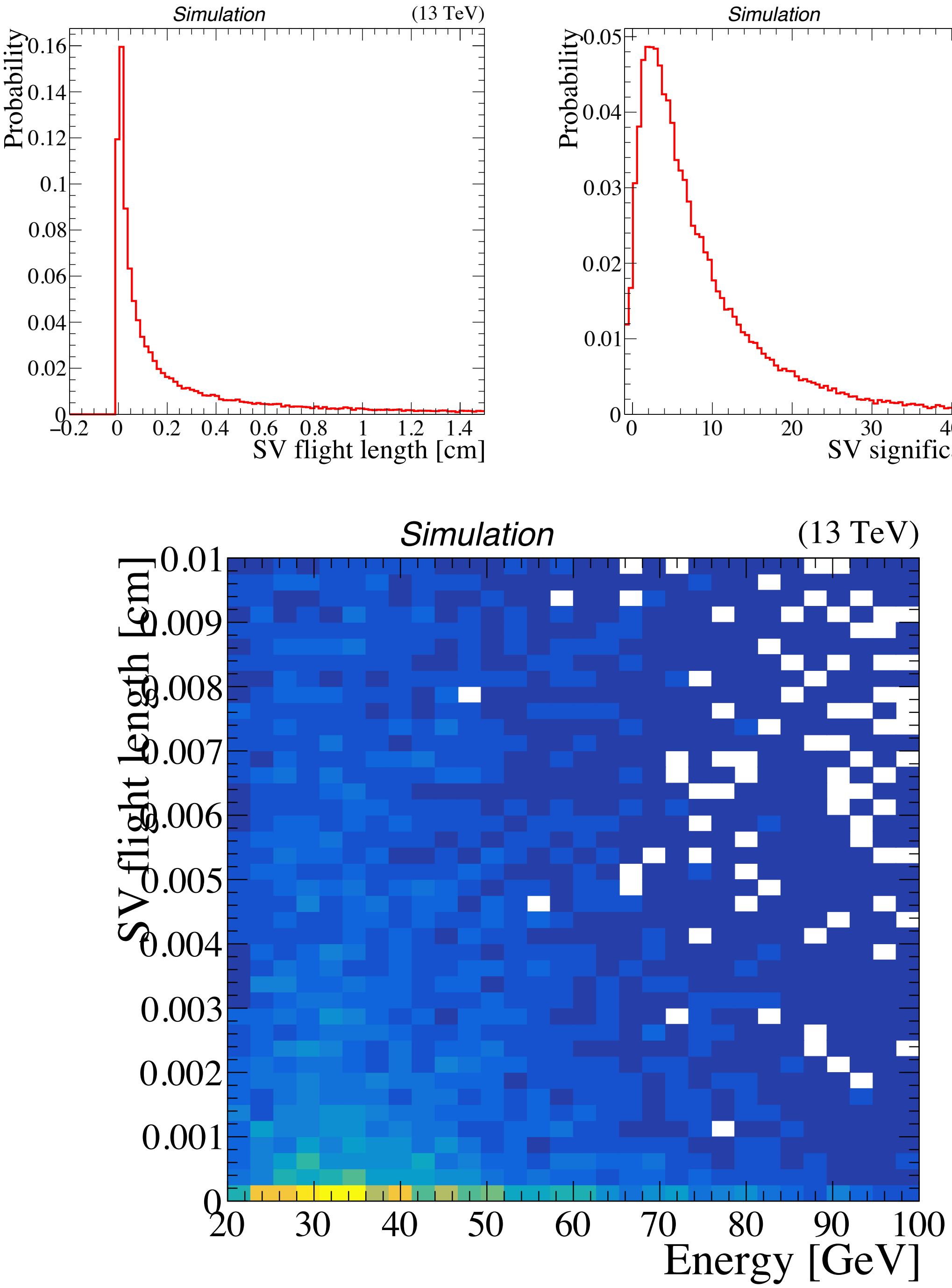

Figure 7.2: Distributions of the length of flight (top left) and the significance of the length of flight (top right) for the genuine τ_h decays taken from the SV data in the $\mathrm{t\bar{t}}$ MC datasets. The correlation plot between the length of flight and the energy for the genuine τ_h candidates (bottom).

τ_h direction. Then a stochastic search in the solid angle of the τ direction finds the minimal angle sum and the best direction of the τ lepton. Each impact parameter is substituted by its projection on the corresponding track component in the plane transverse to the τ flight line. Then, the closest approach point between each pair of tracks is found. The arithmetic average of the three points determines the SV. The uncertainty is estimated from the deviation of the points from the average.

Figure 7.3 shows that the algorithm improves the reconstruction with respect to the default SV saved in the datasets. The average length of flight is 2 mm, the distribution of the SV significance spans through a reasonable range of values, and the length of flight does correlate with the energy of the τ_h candidate.

Figure 7.4 shows the separation between the genuine and misidentified τ_h provided by the SV significance in DY+jets and $\mathrm{t\bar{t}}$ final states. Even though the default SV does distinguish the background in DY, its discriminatory power is not enough in the complex environment of $\mathrm{t\bar{t}}$ processes. The geometric algorithm performs well in both cases.

7.3 Event selection

In order to constrain all systematic uncertainties, the ratio of the W boson branching fractions is measured in a double ratio of four $\mathrm{t\bar{t}}$ and DY dilepton final states, with and without τ leptons:

$$\frac{\mathrm{t\bar{t}} \to \ell\tau_\mathrm{h}}{\mathrm{t\bar{t}} \to \ell\ell} \frac{\mathrm{DY} \to \ell\ell}{\mathrm{DY} \to \tau_\ell\tau_\mathrm{h}} \tag{7.3}$$

Where ℓ denotes a light lepton, and τ_h is a τ lepton that decayed into hadrons.

This analysis is performed only in the DY final states with μ as the light lepton, and the e in the $\mathrm{t\bar{t}}$ final states. Muons provide a greater number of recorded events because of a better reconstruction efficiency and a lower trigger p_T threshold. The lower p_T threshold is particularly important to collect the DY events.

The events in the four final states that make up the double ratio are selected with the following requirements:

- $\mathrm{t\bar{t}} \to \mathrm{e}\tau_\mathrm{h}$: one well isolated e with $p_\mathrm{T} > 30\,\mathrm{GeV}$ that corresponds to the HLT; at least 1 b-tagged jet with $p_\mathrm{T} > 20\,\mathrm{GeV}$; one τ_h candidate with $p_\mathrm{T} > 20\,\mathrm{GeV}$ and a charge opposite to that of the e.

- $\mathrm{t\bar{t}} \to \mathrm{e}\mu$: a well isolated e and a μ of opposite charge with $p_\mathrm{T} > 30\,\mathrm{GeV}$, at least one of them must correspond to the triggered HLT object; at least 1 b-tagged jet with $p_\mathrm{T} > 20\,\mathrm{GeV}$.

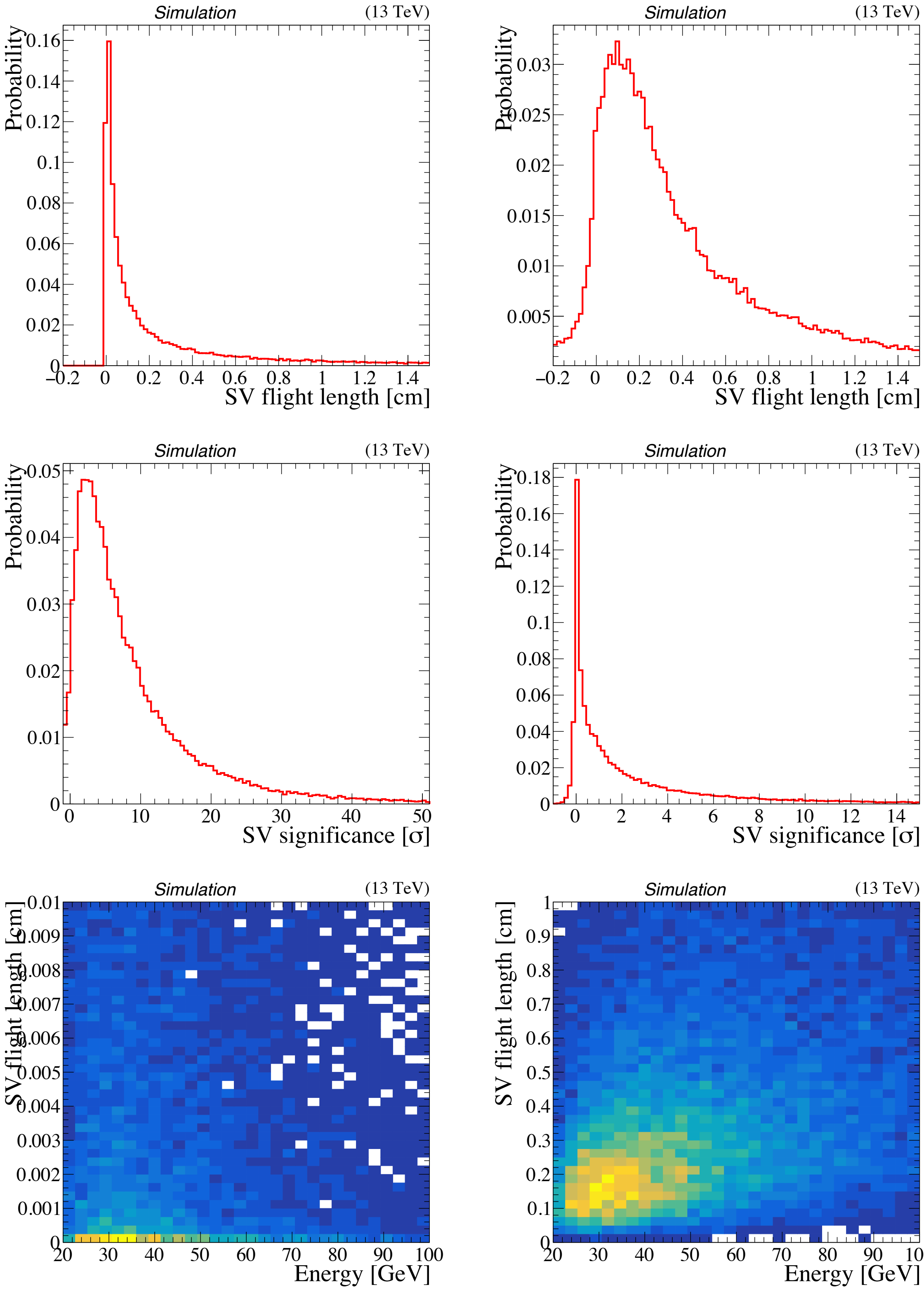

Figure 7.3: Comparison of the distributions for the τ_h SV saved in the datasets (left) and reconstructed with the geometric algorithm that is developed and discussed in the text (right) for genuine τ_h candidates: the length of flight (top), the length of flight significance (middle), and the correlation between the length of flight and energy (bottom).

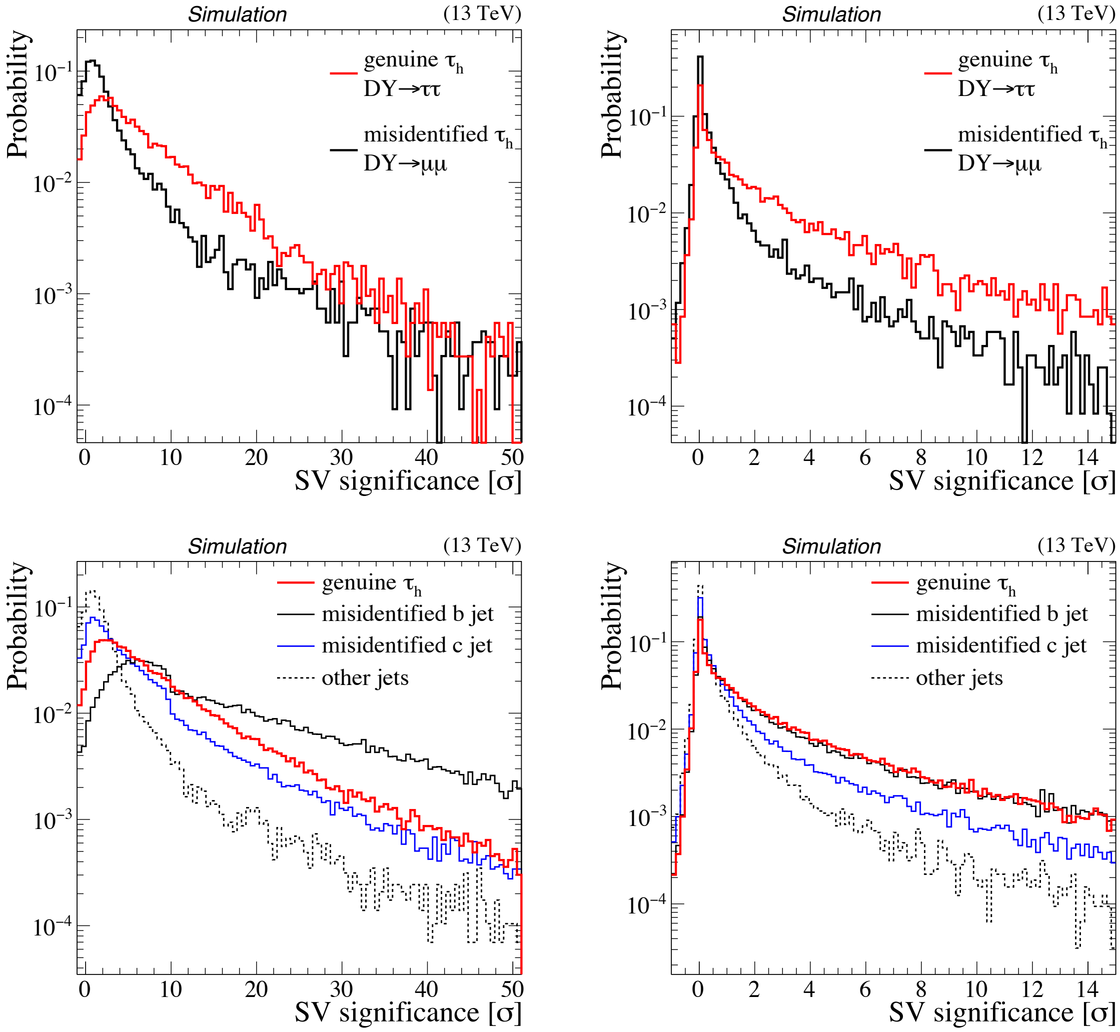

Figure 7.4: Comparison of the SV flight length significance available in the simulation (left) and reconstructed with the geometric algorithm that is developed and discussed in the text (right) in the DY final states $\tau_\ell\tau_\mathrm{h}$ and $\mu\mu$ (top), and the $\mathrm{t\bar{t}}$ final states $\mathrm{t\bar{t}} \to \mathrm{e}\nu_\mathrm{e}\tau_\mathrm{h}\nu_\tau\mathrm{b\bar{b}}$ and $\mathrm{t\bar{t}} \to \mathrm{e}\nu_\mathrm{e}\mathrm{q\bar{q}'b\bar{b}}$ (bottom). The misidentified τ_h candidates in $\mathrm{t\bar{t}}$ are separated according to their physical origin: the hard process b jet, the c flavour jets from the W boson decay, other activity in the event.

- DY → $\tau_\mu \tau_\mathrm{h}$: one well isolated μ with $p_\mathrm{T} > 30\,\mathrm{GeV}$; a τ_h candidate with an opposite charge and $p_\mathrm{T} > 20\,\mathrm{GeV}$; no b-tagged jets; the transverse mass of the μ and $p_\mathrm{T}^\mathrm{miss}$ system must be less than 40 GeV.

- DY → $\mu\mu$: two well isolated μ of opposite charges with $p_\mathrm{T} > 30\,\mathrm{GeV}$; no b-tagged jets; the transverse mass of the leading μ and $p_\mathrm{T}^\mathrm{miss}$ system must be less than 40 GeV.

The transverse mass for a system of two transverse momenta $\vec{p}_\mathrm{T}^\mathrm{\,a}$ and $\vec{p}_\mathrm{T}^\mathrm{\,b}$ is defined as $m_\mathrm{T} = \sqrt{2|\vec{p}_\mathrm{T}^\mathrm{\,a}||\vec{p}_\mathrm{T}^\mathrm{\,b}|(1 - \cos\Delta\varphi)}$, where $\Delta\varphi$ is the angle between the two transverse momenta.

The compositions of the selected events are given in Table 7.1. A comparison to the observed data is presented only for the light dilepton final states, as the final states with τ_h are studied only in the simulation. Figure 7.5 shows the control distributions in the light dilepton final states.

Table 7.1: Expected event yields in the final states used for the double ratio. The signal and SM background processes are given for an integrated luminosity of 35.9 fb^{-1}. Correctly assigned events are designated as “genuine”. Statistical uncertainties are shown.

Process	$\mathrm{t\bar{t}} \to \mathrm{e}\tau_\mathrm{h}$	$\mathrm{t\bar{t}} \to \mathrm{e}\mu$	DY → $\mu\tau_\mathrm{h}$	DY → $\mu\mu$
$\mathrm{t\bar{t}} \to \mathrm{W^+W^-b\bar{b}}$ genuine	1900 ± 40	186000 ± 400	360 ± 20	202000 ± 400
$\mathrm{t\bar{t}} \to \mathrm{W^+W^-b\bar{b}}$ misidentified	800 ± 20	2000 ± 40	280 ± 20	2000 ± 30
tW genuine	190 ± 10	17000 ± 90	80 ± 10	37000 ± 200
tW misidentified	70 ± 5	2000 ± 50	30 ± 5	0 ± 0
Other single top	0	50 ± 10	10 ± 5	0 ± 0
DY → $\tau_\ell \tau_\mathrm{h}$	100 ± 50	400 ± 100	15000 ± 500	0
DY → $\mu\mu$	10 ± 10	0	50 ± 50	41510000 ± 30000
Other DY	0	50 ± 50	0	1000 ± 500
W + jets	0	120 ± 120	3700 ± 650	3200 ± 700
Total	3050 ± 70	209000 ± 500	19500 ± 850	$(41760 \pm 30) \times 10^3$
Data		209387		42831×10^3

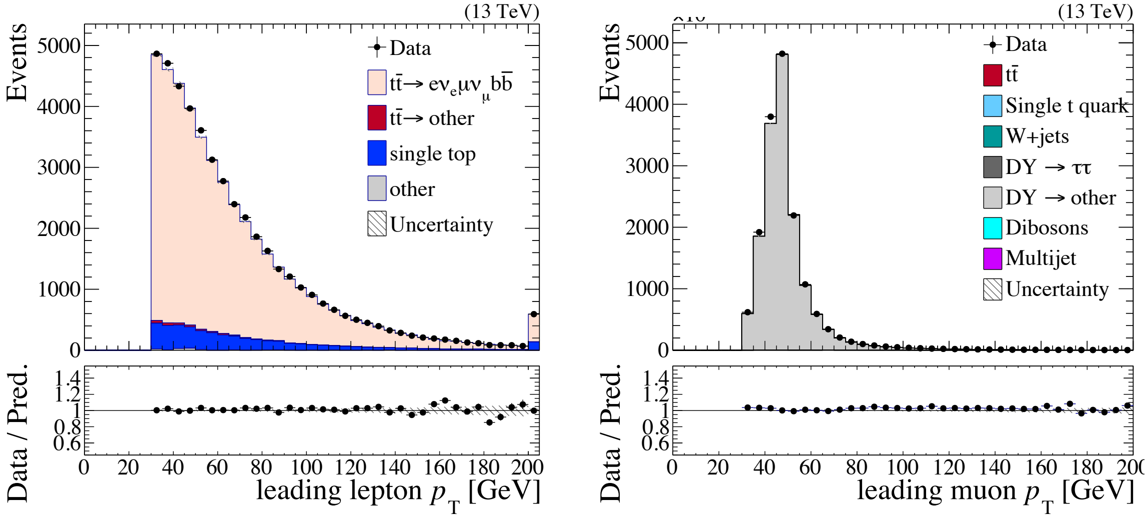

Figure 7.5: Distributions of the leading lepton transverse momentum in the light dilepton $\mathrm{t\bar{t}}$ (left) and DY (right) final states. Only statistical uncertainties are shown.

7.4 Systematic uncertainties

The measurement is affected by the systematic uncertainties in the event reconstruction (experimental uncertainties) and the modelling of the simulated events (theoretical uncertainties). These uncertainties are treated the same way as in the measurement of the $\sigma_{\mathrm{t\bar{t}}}(\ell\tau_{\mathrm{h}})$ production cross section.

In the double ratio almost all experimental and theoretical uncertainties cancel out. The correlations between the leptons in DY can introduce small deviations, but they are found negligible in this study. All systematic uncertainties cancel out in the ratio between the $\mathrm{t\bar{t}} \to \mathrm{e}\tau_{\mathrm{h}}$ and DY $\to \tau_\ell\tau_{\mathrm{h}}$ final states, except the uncertainties in the estimation of background events with misidentified τ_{h} because of the different nature of these backgrounds in the two final states. The events with misidentified τ_{h} in the $\mathrm{t\bar{t}} \to \mathrm{e}\tau_{\mathrm{h}}$ selection come from the semileptonic $\mathrm{t\bar{t}} \to \mathrm{e} + \mathrm{jets}$ process, where most of the misidentified τ_{h} candidates come from the hadronic products of the W boson decay. In the DY $\to \tau_\mu\tau_{\mathrm{h}}$ selection the misidentified τ_{h} are from the W+jets process, mostly from the c quark that is produced in association with the W boson, while the W boson itself decays into $\mu\nu$.

Although tighter requirements for τ_{h} candidates reduce the fraction of the misidentified backgrounds in the selected events, these backgrounds remain not negligible. In order to constrain their effect, the fit is performed to the m_{T} distribution in the

$\mathrm{t\bar{t}} \to \mathrm{e}\tau_\mathrm{h}$ final state. The m_T distribution is also used in the $\mathrm{DY} \to \tau_\mu\tau_\mathrm{h}$ final state. These distributions distinguish the signal and background processes, as shown in Figure 7.6. With a large number of selected events, these backgrounds are well constrained in the fit and they do not introduce a large uncertainty on the measured double ratio. As a cross check, the fit has also been performed to the p_T distribution of the μ in the $\mathrm{DY} \to \tau_\mu\tau_\mathrm{h}$ final state. The background with misidentified τ_h can have a different effect on the p_T and the m_T distributions. However, thanks to the purity of the $\mathrm{DY} \to \tau_\mu\tau_\mathrm{h}$ event selection, the difference between the results is negligible.

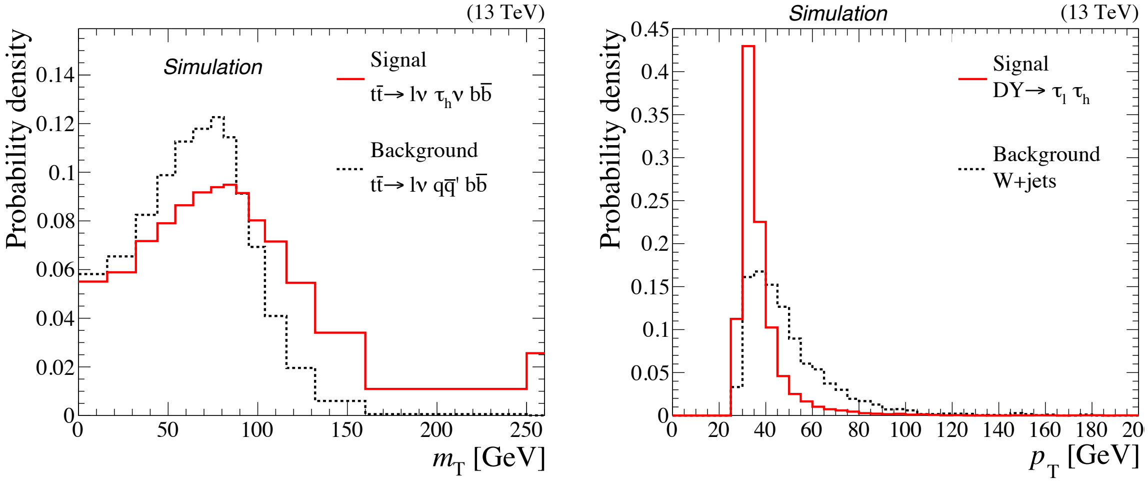

Figure 7.6: Comparison of the signal and background distributions in the $\mathrm{t\bar{t}} \to \mathrm{e}\tau_\mathrm{h}$ and $\mathrm{DY} \to \tau_\mu\tau_\mathrm{h}$ event selections: the m_T distributions in the $\mathrm{t\bar{t}} \to \mathrm{e}\nu_\mathrm{e}\tau_\mathrm{h}\nu_\tau\mathrm{b\bar{b}}$ and $\mathrm{t\bar{t}} \to \mathrm{e}\nu_\mathrm{e}\mathrm{q\bar{q}'b\bar{b}}$ processes on the left, the p_T of the μ in the $\mathrm{DY} \to \tau_\mu\tau_\mathrm{h}$ and W+jets processes on the right.

The efficiency of the τ_h identification and its uncertainty are considered independent from the p_T and other parameters of the τ_h in this analysis. Therefore, the τ_h identification uncertainty cancels out between the events of $\mathrm{t\bar{t}} \to \mathrm{e}\tau_\mathrm{h}$ and $\mathrm{DY} \to \tau_\ell\tau_\mathrm{h}$ in the double ratio. However, the spectra of τ_h momenta are different in these processes and there is a correlation between τ_ℓ and τ_h in the DY process. Therefore it is important to consider a possible p_T dependence in the efficiency and the uncertainty of the τ_h ID requirement. One can expect this dependence to affect the fit in a trivial way, because the fitted m_T distribution practically does not correlate with the τ_h in the event. Also, due to a large number of events, the p_T spectrum of the τ_h in the DY process covers the high p_T tail of the τ_h in the $\mathrm{t\bar{t}}$ process. This dependence is further discussed in the Section 7.7.

7.5 Fit of the double ratio

The double ratio is measured in a profile likelihood ratio (PLR) fit to the binned distributions of the m_{T} in the $\mathrm{t\bar{t}} \to \mathrm{e}\tau_{\mathrm{h}}$ and the DY $\to \tau_\ell\tau_{\mathrm{h}}$ events, and the leading muon p_{T} in the $\mathrm{t\bar{t}} \to \mathrm{e}\mu$ and DY $\to \mu\mu$ events. The PLR is implemented in the same way as in the measurement of the $\sigma_{\mathrm{t\bar{t}}}(\ell\tau_{\mathrm{h}})$ cross section in Section 6.7, except for the definition of the parameter of interest, the "signal strength" parameter in Section 6.6.

The model of expected event yields in each bin k of the distributions is constructed with the templates taken from simulation:

$$\hat{N}_k(n_s,\theta_i) = \hat{N}^s_k(n_s,\theta_i) + \hat{N}^b_k(\theta_i) = \sum_{s\in\mathrm{sig}} n_s \cdot \hat{S}_k \cdot \prod_i (1+\sigma^s_i\theta_i) + \sum_{b\in\mathrm{bkg}} \hat{B}_k \cdot \prod_i (1+\sigma^b_i\theta_i) \tag{7.4}$$

The templates for the signal and background processes are $\hat{S}_k$ and $\hat{B}_k$ respectively. The systematic uncertainties are included in the fit as θ_i nuisance parameters. The $\sigma_i^{\{s,b\}}$ value corresponds to the magnitude of the 1σ effect of the ith systematic uncertainty on the templates. The n_s symbol denotes the two different signal yield normalization parameters that are applied to the signal processes in the four selected final states:

$$n_s = \begin{cases} n_{\mathrm{t\bar{t}}} \cdot r_{\mathrm{DY}} \cdot r & \mathrm{t\bar{t}} \to \mathrm{e}\tau_{\mathrm{h}} \\ n_{\mathrm{t\bar{t}}} & \mathrm{t\bar{t}} \to \mathrm{e}\mu \\ n_{\mathrm{DY}} \cdot r_{\mathrm{DY}} & \mathrm{DY} \to \tau_\mu\tau_{\mathrm{h}} \\ n_{\mathrm{DY}} & \mathrm{DY} \to \mu\mu \end{cases} \tag{7.5}$$

The parameters $n_{\mathrm{t\bar{t}}}$ and n_{DY} correspond to the common normalization in the $\mathrm{t\bar{t}}$ and DY processes. The parameter r_{DY} corresponds to the ratio between the DY processes. Under the assumption of the SM branching fractions in DY, this ratio fits the deviation from the nominal expected τ_{h} ID efficiency in DY events. The parameter r encodes the difference between the W boson branching fractions. It is applied to the $\mathrm{t\bar{t}}$ processes with τ leptons. The dilepton final states with two τ leptons, such as $\mathrm{t\bar{t}} \to \tau_\ell\nu_\tau\nu_\ell\tau_{\mathrm{h}}\nu_\tau\mathrm{b\bar{b}}$, are multiplied by r^2 corresponding to two $\mathrm{W} \to \tau\nu_\tau$ decays. The same r parameter could be applied to the single top processes that produce the signal final state, slightly improving the constraints in the fit. However, I take a conservative approach and do not consider this small contribution as a part of the signal. The normalization parameters $n_{\mathrm{t\bar{t}}}$ and n_{DY}, and the r_{DY} ratio parameter are treated as additional nuisance parameters. The ratio parameter r is included in the fit as the parameter of interest.

With this model, a likelihood function for a given set of observed distributions N_k is defined:

$$\mathcal{L}(r,\theta_i) = \prod_k \mathcal{P}_{\text{oisson}} \left[N_k | \hat{N}_k(r,\theta_i) \right] \cdot \prod_i pdf(\theta_i, 0, 1) \tag{7.6}$$

The PLR test statistic is defined for a given value of r as the ratio between the maximum likelihood value for this r and the global maximum value of the likelihood:

$$\lambda(r) = \frac{\mathcal{L}(r, \tilde{\theta}_i(r))}{\mathcal{L}(\hat{r}, \hat{\theta}_i)} \tag{7.7}$$

The quantities $\hat{r}$ and $\hat{\theta}_i$ give the global maximum of the likelihood function, and $\tilde{\theta}_i(r)$ maximizes the likelihood for a given value of r.

By fixing the nuisance parameters to their post-fit values the effect of separate uncertainties can be studied. This approach includes the effect of systematic uncertainties on the parameter of interest: here, the test statistic $\lambda(r)$ is broader due to the adjustment of the systematic uncertainties in $\tilde{\theta}_i(r)$.

More details on the method can be found in Ref. [20]. The computation of PLR is performed in RooStats [118] with the Higgs Combination tool [116, 117, 119].

As a confirmation of the expected cancellation of systematic uncertainties, a fit to the ratio between the $\mathrm{t\bar{t}}$ final states alone is performed: $\mathrm{t\bar{t}} \to \mathrm{e}\tau_\mathrm{h} / \mathrm{t\bar{t}} \to \mathrm{e}\mu$.

The ratio is fitted as a part of the normalization parameters for the signal processes in the model of the expected numbers of events in the m_T distributions. Exactly like in the fit to the double ratio (7.5), but without the reference to the DY final states:

$$n_s = \begin{cases} n_{\mathrm{t\bar{t}}} \cdot r_{\mathrm{t\bar{t}}} & \mathrm{t\bar{t}} \to \mathrm{e}\tau_\mathrm{h} \\ n_{\mathrm{t\bar{t}}} & \mathrm{t\bar{t}} \to \mathrm{e}\mu \end{cases} \tag{7.8}$$

The normalization $n_{\mathrm{t\bar{t}}}$ is applied to all $\mathrm{t\bar{t}}$ processes and is included in the fit as a nuisance parameter. The ratio $r_{\mathrm{t\bar{t}}}$ is applied to the $\mathrm{t\bar{t}}$ processes with τ leptons, and it is the parameter of interest in the fit.

The fit is performed in the simulation that is normalized to the luminosity of the full Run2 dataset, and it shows that all uncertainties not related to the τ_h candidate cancel out. The NLL scan of the $r_{\mathrm{t\bar{t}}}$ parameter in Figure 7.7 shows the overall uncertainty of about 8%. Figure 7.8 with the breakdown of the systematic uncertainties

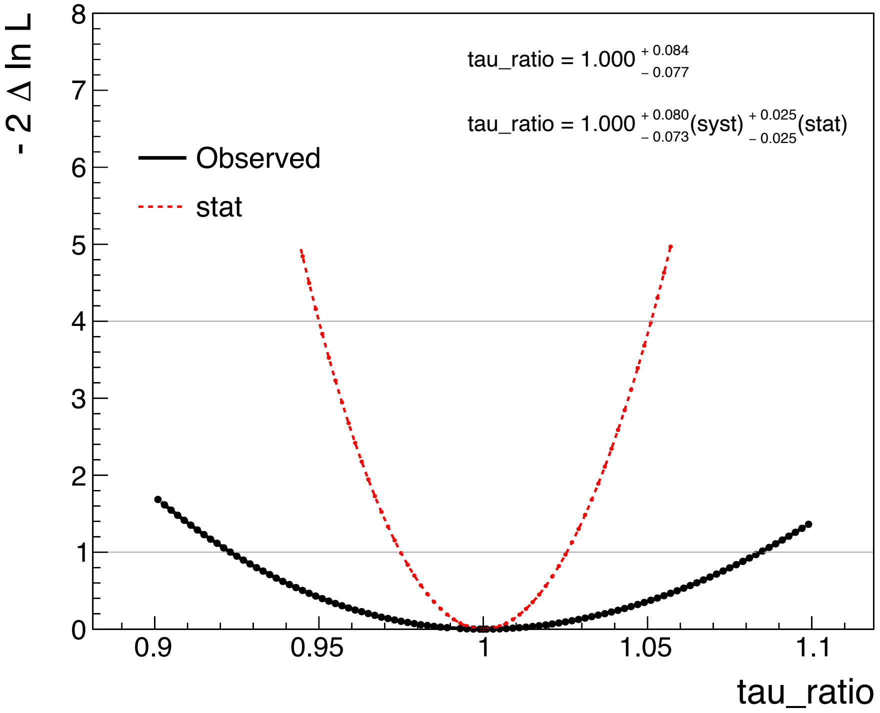

Figure 7.7: Scan of the expected profile likelihood ratio in the fit of the ratio between dilepton final states with light leptons and with a τ_h in $\mathrm{t\bar{t}}$. The statistical uncertainty is shown by fixing the nuisance parameters that encode systematic uncertainties to their post-fit values.

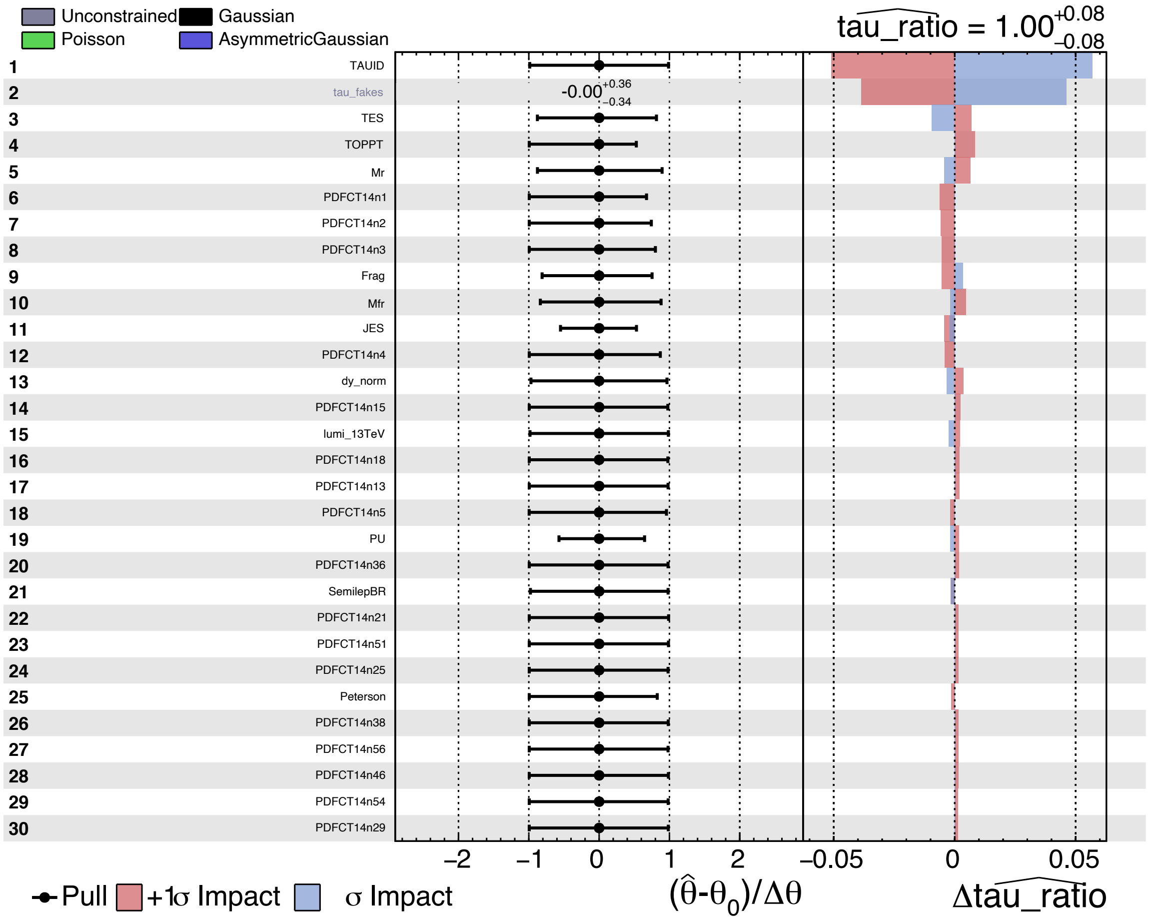

Figure 7.8: Expected impacts of the nuisance parameters on the signal strength in the fit of the ratio between dilepton final states with light leptons and with a τ_h in $t\bar{t}$.

shows that the uncertainty of the τ_h ID and the uncertainty in the τ_h misidentification are by far the dominant ones in the fit, and other uncertainty sources are negligible.

Therefore, we confirm the expected performance of the constructed fitting procedure: the systematic uncertainties cancel out in the ratio between the $\mathrm{t\bar{t}}$ dilepton final states.

7.6 Results

The fit of the double ratio with DY final states leads to the cancellation of the uncertainties related to the τ_h selection. However, since the fitted model is more complex, the impact of the statistical uncertainty on the result can be large. To evaluate the contribution of statistical uncertainty, the fit is performed in the sample of simulated events, that is normalized to the luminosity of $35.9\,\mathrm{fb}^{-1}$ that corresponds to the size of the 2016 year dataset, and then compared to the full Run2 luminosity.

For the fit in $35.9\,\mathrm{fb}^{-1}$, the NLL scan of the double ratio r parameter and the basic breakdown of the uncertainty in the result are shown in Figure 7.9. The statistical uncertainty corresponds to about 30% of all uncertainty. Full listing of the post-fit pulls of the nuisance parameters and their impacts on the parameter of interest is presented in Figure 7.10. Some of the nuisance parameters have not cancelled out because of the large statistical uncertainty: the statistical fluctuations in different final states propagate to the fit of the parameter of interest. The fitting software picks up these fluctuations in the evaluation of the impacts of the nuisance parameters. The statistical fluctuations are independent between different event selections, therefore they are not cancelled in the double ratio. With more events the statistical fluctuations decrease and the double ratio is able to constrain all systematic uncertainties.

In the fit to the luminosity of $145\,\mathrm{fb}^{-1}$ that corresponds to the full Run2 data sample, the overall relative uncertainty is about 3.5%, much smaller than in the case of the luminosity accumulated in 2016. The NLL scan of the double ratio and the uncertainty breakdown are shown in Figure 7.11. The statistical uncertainty is still about 30% of the overall uncertainty. The full breakdown of the impacts of the nuisance parameters on the signal strength is listed in Figure 7.12.

Increasing the luminosity further to a value of $1000\,\mathrm{fb}^{-1}$ reduces the overall uncertainty down to about 2.0%, as shown in Figure 7.13.

The uncertainties evaluated in the feasibility study are summarized in Figure 7.14. The reduction of the statistical fluctuations increases the constrain on the systematic uncertainties in the double ratio. At the luminosity of $145\,\mathrm{fb}^{-1}$ about 10 thousand

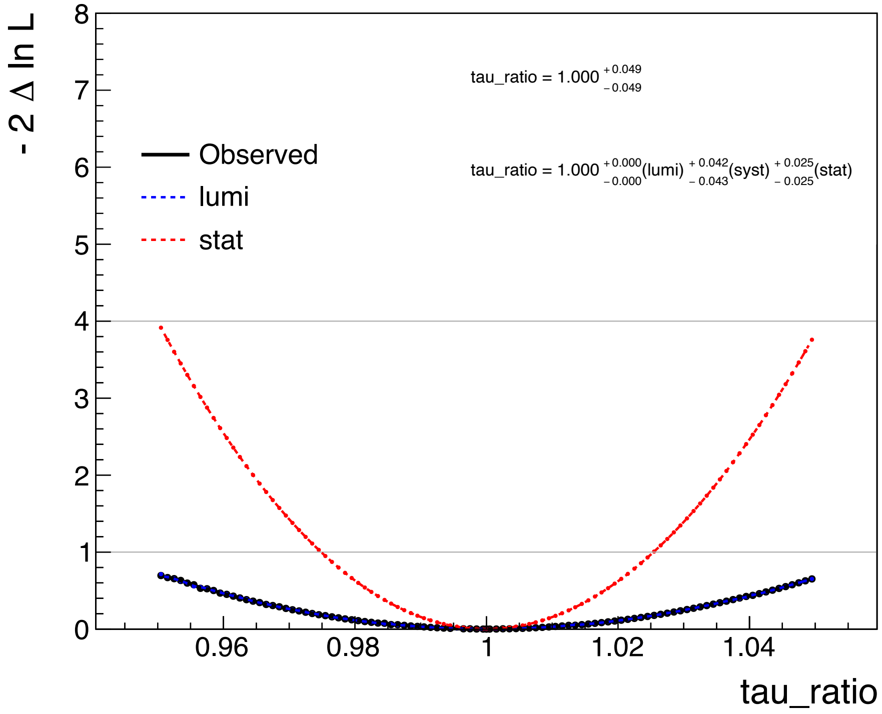

Figure 7.9: Scan of the expected profile likelihood ratio in the double ratio fit. The fit is performed in simulated toy distributions, with the expected event yields normalized to a luminosity of $35.9\,\text{fb}^{-1}$. The statistical uncertainty is estimated by fixing the nuisance parameters that encode systematic uncertainties to their post-fit values. The cancellation of the systematic uncertainties is limited by the statistical fluctuations.

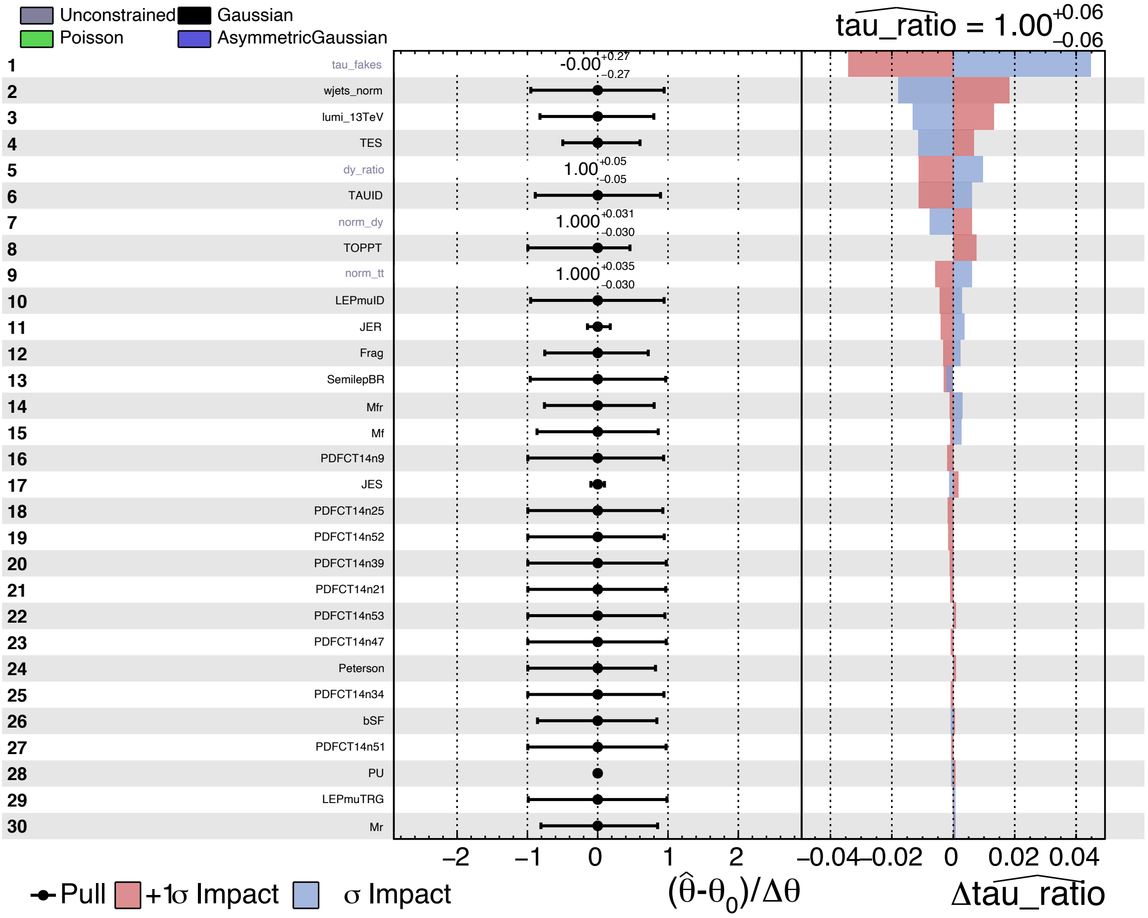

Figure 7.10: Expected impacts of nuisance parameters on the signal strength in the double ratio fit. The fit is performed in simulated toy distributions, with the expected event yields normalized to a luminosity of 35.9 fb^{-1}.

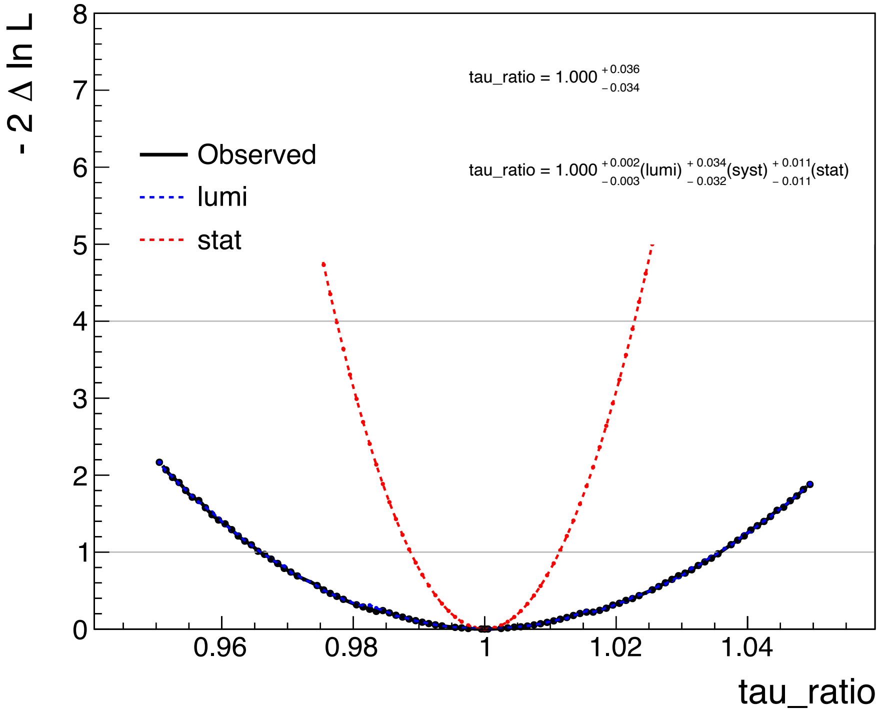

Figure 7.11: Scan of the expected profile likelihood ratio in the double ratio fit. The fit is performed in simulated toy distributions, with the expected event yields normalized to a luminosity of $145\,\mathrm{fb}^{-1}$. The statistical uncertainty is shown by fixing the nuisance parameters that encode systematic uncertainties to their post-fit values. The cancellation of the systematic uncertainties is limited by the statistical fluctuations.

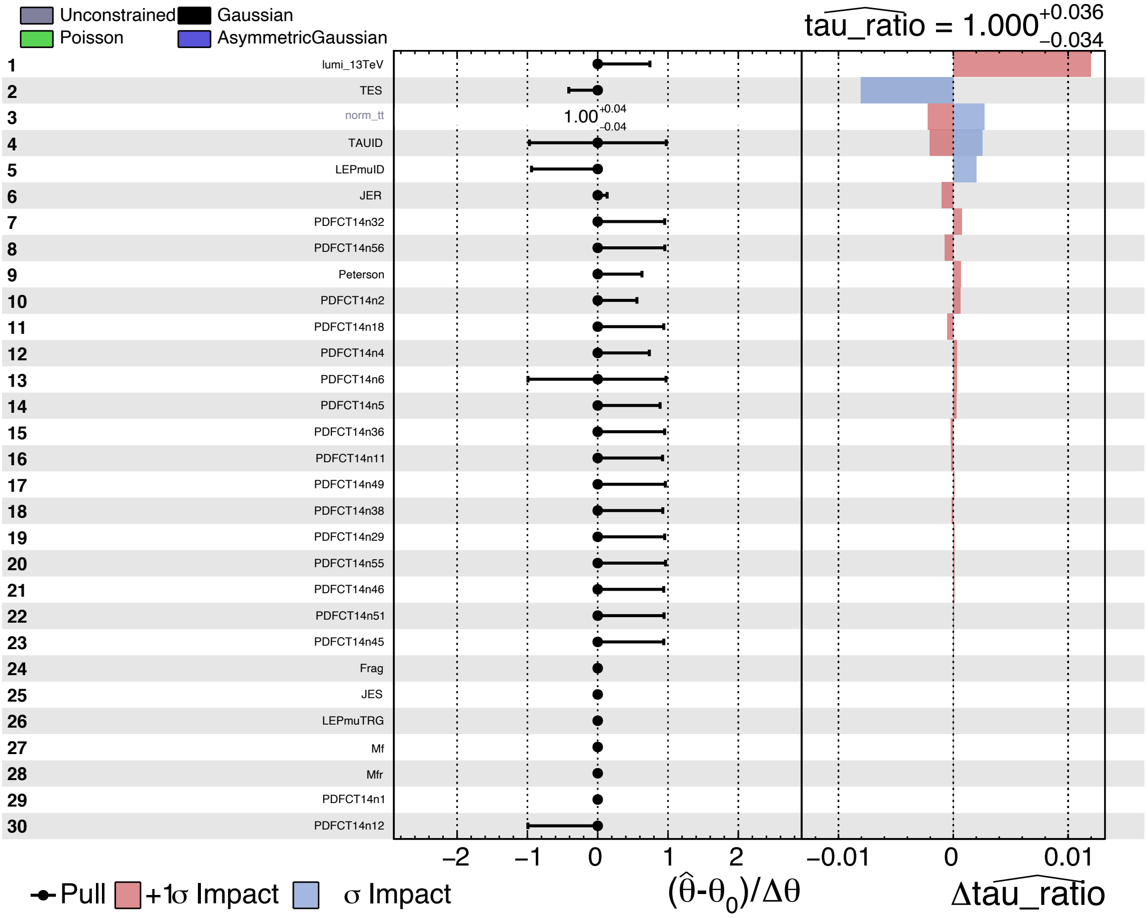

Figure 7.12: Expected impacts of nuisance parameters on the signal strength in the double ratio fit. The fit is performed in simulated toy distributions, with the expected event yields normalized to a luminosity of 145 fb^{-1}.

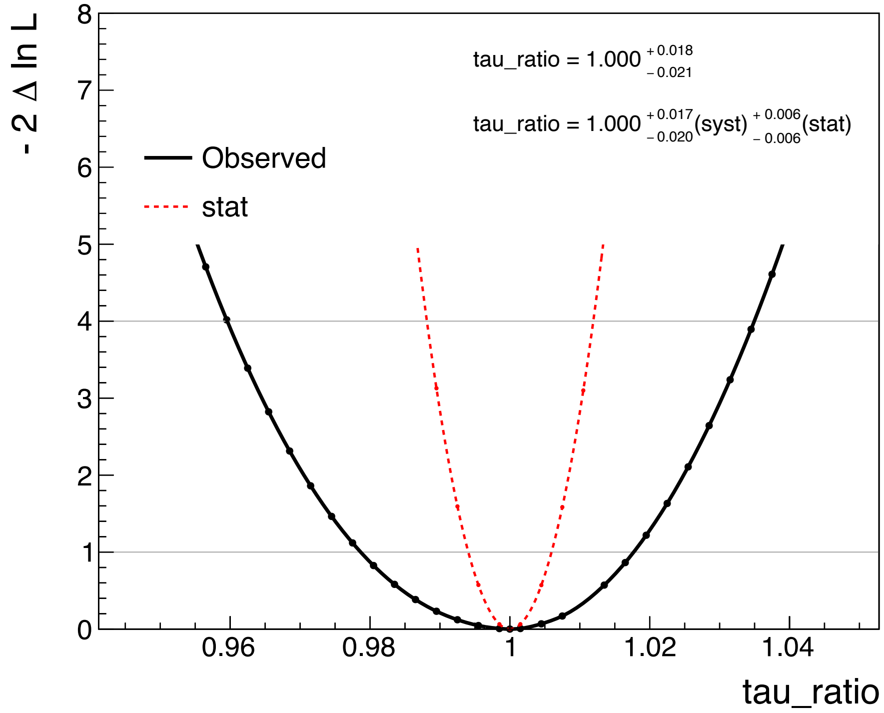

Figure 7.13: Scan of the expected profile likelihood ratio in the double ratio fit. The fit is performed in simulated toy distributions, with the expected event yields normalized to a luminosity of 1000 fb^{-1}. The statistical uncertainty is shown by fixing the nuisance parameters that encode systematic uncertainties to their post-fit values. The small statistical fluctuations allow the systematic uncertainties to cancel out in the double ratio.

events are expected in the $\mathrm{t\bar{t}} \to \ell\tau_\mathrm{h}$ events. It corresponds to a statistical uncertainty of about 1% and a small enough overall uncertainty to make a significant contribution to the current best results.

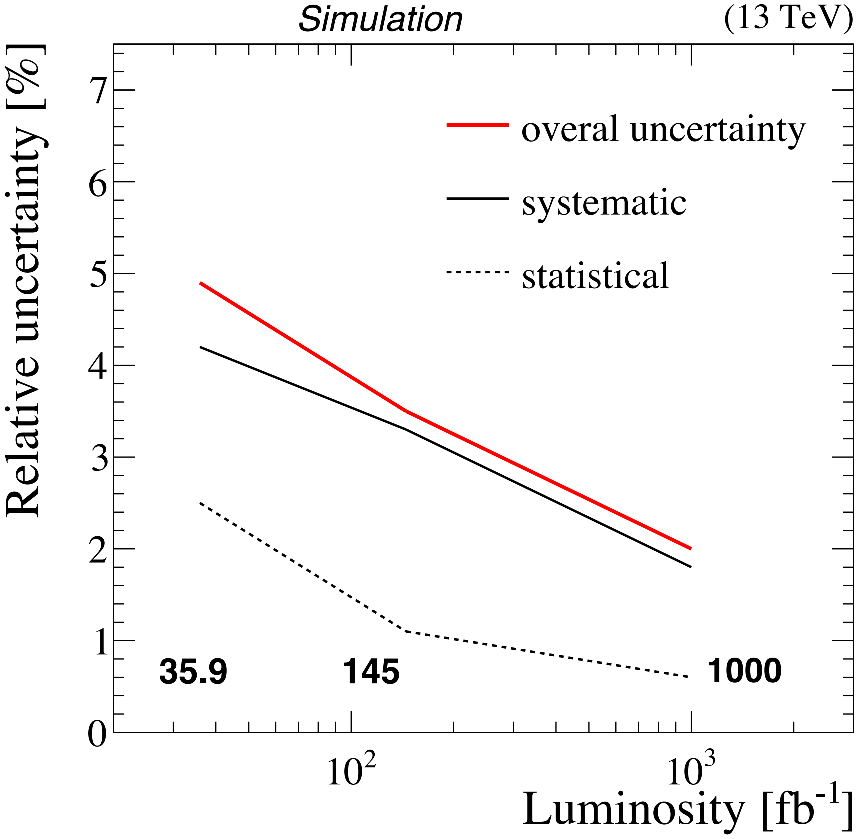

Figure 7.14: Summary of the expected relative uncertainties in double ratio measurements at different luminosities: 35.9 fb^{-1}, 145 fb^{-1}, and 1000 fb^{-1}. The overall uncertainty is shown by the thick red line, the systematic component is the thin black line, the statistical fluctuations are shown by the dashed black line.

7.7 Discussion

The fits performed in the $\mathrm{t\bar{t}}$ final states alone and in all four final states demonstrate that the systematic uncertainties cancel out in the measurement of the ratio parameters, if the statistical fluctuations are negligible. With the current event selection requirements, a measurement in the full Run2 dataset will have a relative uncertainty of about 3.5%. A result with this precision would be the most precise measurement of the ratio between the W boson branching fractions $\mathcal{B}(\mathrm{W} \rightarrow \tau_{\mathrm{h}}\nu_{\tau\mathrm{h}})/\mathcal{B}(\mathrm{W} \rightarrow \ell\nu_{\ell})$ made in the final states with hadronic τ lepton decays at a hadron accelerator up to date.

As demonstrated in Section 7.6, the proposed event selection is pure enough but lacks statistics. Even with the full Run2 dataset the uncertainty is dominated by the statistical fluctuations. There are a number of possibilities to enhance the number of selected events without loosing the purity of the selection:

- A fit using both flavours of light leptons, electrons and muons. The final states with electrons suffer from high p_{T} thresholds in available single-electron triggers. The yield of the $\mathrm{t\bar{t}}$ event selection with muons would be about 50-60% larger than the selection with electrons that is considered in this analysis. An increase of the overall event sample by 50% would significantly reduce the statistical fluctuations.

- An advanced HLT with lower p_{T} thresholds for the leptons. The DY processes, especially the DY $\rightarrow \tau_{\ell}\nu_{\ell}\nu_{\tau}\tau_{\mathrm{h}}\nu_{\tau}$ final state, are sensitive to the lepton p_{T} thresholds. The $\mathrm{t\bar{t}}$ processes also have events for which the lepton p_{T} is below 30 GeV. A $\ell\tau_{\mathrm{h}}$ or a ℓ+jets HLT with a low lepton p_{T} threshold could significantly increase the number of selected events.

- A low p_{T} threshold for τ_{h} candidates. The τ_{h} candidates have a softer p_{T} distribution than the light leptons in both $\mathrm{t\bar{t}}$ and DY events, because of the neutrino in the τ decay. Therefore, the event selection is sensitive to the cuts on the τ_{h} p_{T}. A threshold of 10 GeV would increase the event yield by about 50% in comparison with the usual threshold of 20 GeV. However, the τ_{h} identification algorithms are validated only for the kinematic range starting at 20 GeV [3]. The validity of the algorithms at low p_{T} thresholds must be additionally confirmed. But there are ongoing similar projects to develop low-p_{T} lepton identification at CMS, with the goal to increase the signal acceptance in B physics.

- An improved reconstruction of the τ_{h} secondary vertex with a better treatment of the track reconstruction uncertainties in the detector. Then the cut on the significance of the SV length of flight would not have to be so high to obtain a pure selection of genuine τ_{h}.

- The event selection can be optimized with machine learning algorithms. In particular, the τ_h identification algorithm can be retrained specifically against the misidentified backgrounds from the semileptonic $\mathrm{t\bar{t}}$ final state, such as jets that originate from b and c quarks. Then the algorithm should include the parameters relevant to these jets: the momenta and impact parameters of tracks that form the τ_h candidate, and also the momenta and impact parameters of the tracks in the isolation cone around the signal tracks.

- A great number of events have 1-prong τ_h candidates: the τ lepton decays into the 1-prong hadronic final states about twice more frequently than in the 3-prong final states. But the background of misidentified 1-prong τ_h candidates is larger and harder to distinguish from the genuine τ_h final states. However, a number of kinematic distributions can purify the event selection in $\mathrm{t\bar{t}}$ and DY final states without relying solely on the purity of the τ_h candidates.

- The measurement can be carried out with more data in Run3. In order to perform this measurement in Run3, either the single lepton triggers must keep low p_T thresholds or special complex triggers can be developed to record events from both $\mathrm{t\bar{t}} \to \ell\tau_\mathrm{h}$ and $\mathrm{DY} \to \tau_\ell\tau_\mathrm{h}$ processes.

A careful study of the possible dependence of the τ_h identification efficiency on the p_T and other kinematic parameters of the τ_h candidates is necessary to confirm that the effect is negligible, or to include this effect in the fit.

Since the uncertainty comes from the statistical limitations, the methods can be improved with an optimization of the event selection, and can be employed in precise tests of the lepton universality in the future Run3 data.

The current result serves as a good test of the methods needed for a precise measurement of the lepton universality in W boson decays with hadronic τ lepton decays at LHC. The achieved measurement with 3.5% uncertainty can be reduced further down to the level of 2–3% with addition of other light lepton final states and better event selection strategy. A measurement in hadronic τ decays is not biased to the signal model. It would make a significant contribution to the current world-best measurements.

Chapter 8

Summary

The presented work includes a detailed study of the top quark pair ($\mathrm{t\bar{t}}$) final state with tau leptons (τ) in the final state, $\mathrm{t\bar{t}} \rightarrow \ell\nu_{\ell}\tau_{\mathrm{h}}\nu_{\tau}\mathrm{b\bar{b}}$ channel where ℓ denotes a light lepton, i.e.either an electron (e) or a muon (μ). The top quark pair production cross section [120], $\sigma_{\mathrm{t\bar{t}}}(\ell\tau_{\mathrm{h}})$, the ratio to the cross section in the light dilepton final state [2] $R_{\ell\tau_{\mathrm{h}}/\ell\ell}$, and the ratio of the partial to the total width of the top quark $\Gamma(\mathrm{t} \rightarrow \tau\nu_{\tau}\mathrm{b})/\Gamma_{\mathrm{total}}$ are measured in the data sample collected by the CMS detector at LHC in 2016. In order to establish the methods for the cross section measurement, the identification of the τ leptons and the corresponding backgrounds of misidentified objects are studied extensively. Finally, a feasibility study to carry out a lepton universality test in the $\mathrm{t\bar{t}}$ dilepton final states is presented for the full Run2 data sample.

The measurement of the $\mathrm{t\bar{t}}$ production cross section is the first in $\sqrt{s} = 13\,\mathrm{TeV}$ proton-proton collisions that includes τ leptons in the final state. It improves the relative precision with respect to the Run1 results [18, 19]. The estimated ratio to the light dilepton final state and the ratio of the partial to the total width of the top quark improve over the previous measurements [21, 22].

The methods developed for the cross section measurement can be employed to perform a precise test of the lepton universality principle in the ratio between W boson branching fractions $\mathcal{B}(\mathrm{W} \rightarrow \tau\nu_{\tau})/\mathcal{B}(\mathrm{W} \rightarrow \ell\nu_{\ell})$. The presented analysis shows a possibility to measure this ratio with a relative uncertainty of about 3% in the full Run2 data sample, that would make a significant contribution to current world-best results [4].

The thesis presents detailed information about the τ leptons and the corresponding backgrounds in the environment of the $\mathrm{t\bar{t}}$ final states produced in high energy proton-proton collisions, that can be useful for any analysis involving τ leptons.

Bibliography

[1] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159, 2015.

[2] Albert M Sirunyan et al. Measurement of the $\mathrm{t\bar{t}}$ production cross section, the top quark mass, and the strong coupling constant using dilepton events in pp collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 79:368, 2019.

[3] Albert M Sirunyan et al. Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13:P10005, 2018.

[4] Particle Data Group, M. Tanabashi, et al. Review of particle physics. *Phys. Rev. D*, 98:030001, 2018.

[5] R. Aaij et al. Test of lepton flavor universality by the measurement of the $\mathrm{B}^0 \to \mathrm{D}^{*-}\tau^+\nu_\tau$ branching fraction using three-prong τ decays. *Phys. Rev. D*, 97:072013, 2018.

[6] Roel Aaij et al. Search for lepton-universality violation in $\mathrm{B}^+ \to \mathrm{K}^+\ell^+\ell^-$ decays. *Phys. Rev. Lett.*, 122:191801, 2019.

[7] J. P. Lees et al. Evidence for an excess of $\bar{\mathrm{B}} \to \mathrm{D}^{(*)}\tau^-\bar{\nu}_\tau$ decays. *Phys. Rev. Lett.*, 109:101802, 2012.

[8] J. P. Lees et al. Measurement of an excess of $\bar{\mathrm{B}} \to \mathrm{D}^{(*)}\tau^-\bar{\nu}_\tau$ decays and implications for charged Higgs bosons. *Phys. Rev. D*, 88:072012, 2013.

[9] Simone Bifani, Sébastien Descotes-Genon, Antonio Romero Vidal, and Marie-Hélène Schune. Review of lepton universality tests in B decays. *J. Phys. G*, 46:023001, 2019.

[10] M. Huschle et al. Measurement of the branching ratio of $\bar{\mathrm{B}} \to \mathrm{D}^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{\mathrm{B}} \to \mathrm{D}^{(*)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle. *Phys. Rev. D*, 92:072014, 2015.

[11] Y. Sato et al. Measurement of the branching ratio of $\bar{\mathrm{B}}^0 \to \mathrm{D}^{*+}\tau^-\bar{\nu}_\tau$ relative to $\bar{\mathrm{B}}^0 \to \mathrm{D}^{*+}\ell^-\bar{\nu}_\ell$ decays with a semileptonic tagging method. *Phys. Rev. D*, 94:072007, 2016.

[12] S. Hirose et al. Measurement of the τ lepton polarization and $R(\mathrm{D}^*)$ in the decay $\bar{\mathrm{B}} \to \mathrm{D}^*\tau^-\bar{\nu}_\tau$. *Phys. Rev. Lett.*, 118:211801, 2017.

[13] Abdelhak Djouadi. The anatomy of electroweak symmetry breaking. II. The Higgs bosons in the minimal supersymmetric model. *Phys. Rept.*, 459:1, 2008.

[14] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, Marc Sher, and Joao P. Silva. Theory and phenomenology of two-Higgs-doublet models. *Phys. Rept.*, 516:1, 2012.

[15] Albert M Sirunyan et al. Search for charged Higgs bosons in the $\mathrm{H}^\pm \to \tau^\pm\nu_\tau$ decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 07:142, 2019.

[16] Morad Aaboud et al. Search for charged Higgs bosons decaying via $\mathrm{H}^\pm \to \tau^\pm\nu_\tau$ in the τ+jets and τ+lepton final states with 36 fb^{-1} of pp collision data recorded at $\sqrt{s} = 13$ TeV with the ATLAS experiment. *JHEP*, 09:139, 2018.

[17] Test of the universality of τ and μ lepton couplings in W boson decays from $t\bar{t}$ events at 13 TeV with the ATLAS detector. Technical Report ATLAS-CONF-2020-014, CERN, Geneva, Jun 2020.

[18] Serguei Chatrchyan et al. Measurement of the top quark pair production cross section in pp collisions at $\sqrt{s} = 7$ TeV in dilepton final states containing a τ. *Phys. Rev. D*, 85:112007, 2012.

[19] Vardan Khachatryan et al. Measurement of the $\mathrm{t\bar{t}}$production cross section in pp collisions at $\sqrt{s} = 8$ TeV in dilepton final states containing one τ lepton. *Phys. Lett. B*, 739:23, 2014.

[20] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011. [Erratum: 10.1140/epjc/s10052-013-2501-z].

[21] Timo Antero Aaltonen et al. Study of top-quark production and decays involving a tau lepton at CDF and limits on a charged-Higgs boson contribution. *Phys. Rev. D*, 89:091101, 2014.

[22] Georges Aad et al. Measurements of the top quark branching ratios into channels with leptons and quarks with the ATLAS detector. *Phys. Rev. D*, 92:072005, 2015.

[23] F. Abe et al. Observation of top quark production in $\bar{p}p$ collisions. *Phys. Rev. Lett.*, 74:2626–2631, 1995.

[24] S. Abachi et al. Observation of the top quark. *Phys. Rev. Lett.*, 74:2632–2637, 1995.

[25] CMS Collaboration. Combination of ATLAS and CMS top quark pair cross section measurements in the emu final state using proton-proton collisions at 8 TeV. 9 2014.

[26] Vardan Khachatryan et al. Measurement of the $t\bar{t}$ production cross section using events in the eμ final state in pp collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 77:172, 2017.

[27] Morad Aaboud et al. Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector. *Phys. Lett. B*, 761:136, 2016. [Erratum: Phys.Lett.B 772, 879–879 (2017)].

[28] Georges Aad et al. Measurement of the $t\bar{t}$ production cross-section and lepton differential distributions in $e\mu$ dilepton events from pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. 10 2019.

[29] Michal Czakon and Alexander Mitov. Top++: A program for the calculation of the top-pair cross-section at hadron colliders. *Comput. Phys. Commun.*, 185:2930, 2014.

[30] Albert M. Sirunyan et al. Measurement of the $t\bar{t}$ production cross section using events with one lepton and at least one jet in pp collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 09:051, 2017.

[31] CMS Collaboration. First measurement of the differential cross section for $\mathrm{t\bar{t}}$ production in the dilepton final state at $\sqrt{s} = 13$ TeV. 8 2015.

[32] A.M. Sirunyan et al. Measurements of differential cross sections of top quark pair production as a function of kinematic event variables in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 06:002, 2018.

[33] A. M. Sirunyan et al. Measurement of normalized differential $\mathrm{t\bar{t}}$ cross sections in the dilepton channel from pp collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 04:060, 2018.

[34] Morad Aaboud et al. Measurements of top-quark pair differential cross-sections in the $e\mu$ channel in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector. *Eur. Phys. J. C*, 77(5):292, 2017.

[35] M. Aaboud et al. Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector. *JHEP*, 11:191, 2017.

[36] Georges Aad et al. Measurements of top-quark pair differential and double-differential cross-sections in the ℓ+jets channel with pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector. *Eur. Phys. J. C*, 79(12):1028, 2019.

[37] Serguei Chatrchyan et al. Measurement of the Top-Quark Mass in $t\bar{t}$ Events with Dilepton Final States in pp Collisions at $\sqrt{s} = 7$ TeV. *Eur. Phys. J. C*, 72:2202, 2012.

[38] Vardan Khachatryan et al. Search for a charged Higgs boson in pp collisions at $\sqrt{s} = 8$ TeV. *JHEP*, 11:018, 2015.

[39] *LEP design report*. CERN, Geneva, 1984. Copies shelved as reports in LEP, PS and SPS libraries.

[40] Brüning, Oliver Sim and Collier, Paul and Lebrun, P and Myers, Stephen and Ostojic, Ranko and Poole, John and Proudlock, Paul. *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004.

[41] Lyndon Evans and Philip Bryant. LHC machine. *Journal of Instrumentation*, 3(08):S08001–S08001, aug 2008.

[42] Linear accelerator 2. Sep 2012.

[43] The Proton Synchrotron Booster. Jul 2012.

[44] Proton Synchrotron. 1959.

[45] The Super Proton Synchrotron. Jan 2012.

[46] High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1. *CERN Yellow Rep. Monogr.*, 4:1, 2017.

[47] S. Chatrchyan et al. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[48] Serguei Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014.

[49] Vardan Khachatryan et al. CMS Tracking Performance Results from Early LHC Operation. *Eur. Phys. J. C*, 70:1165, 2010.

[50] Serguei Chatrchyan et al. Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7$ TeV. *JINST*, 8:P09009, 2013. [JINST8,9009(2013)].

[51] S. Abdullin et al. Design, performance, and calibration of CMS hadron-barrel calorimeter wedges. *Eur. Phys. J.*, C55(1):159–171, 2008.

[52] Vardan Khachatryan et al. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *JINST*, 12:P02014, 2017.

[53] A. M. Sirunyan et al. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13:P06015, 2018.

[54] Vardan Khachatryan et al. The CMS trigger system. *JINST*, 12:P01020, 2017.

[55] G Bauer, B Beccati, U Behrens, K Biery, O Bouffet, J Branson, S Bukowiec, E Cano, H Cheung, M Ciganek, S Cittolin, J A Coarasa, C Deldicque, A Dupont, S Erhan, D Gigi, F Glege, R Gomez-Reino, D Hatton, A Holzner, Y L Hwong, L Masetti, F Meijers, E Meschi, R K Mommsen, R Moser, V O'Dell, L Orsini, C Paus, A Petrucci, M Pieri, A Racz, O Raginel, H Sakulin, M Sani, P Schieferdecker, C Schwick, D Shpakov, M Simon, and K Sumorok. The data-acquisition system of the CMS experiment at the LHC. *Journal of Physics: Conference Series*, 331(2):022021, dec 2011.

[56] E. Migliore. CMS pixel detector design for HL-LHC. *Journal of Instrumentation*, 11(12):C12061–C12061, dec 2016.

[57] Ernesto Migliore. The CMS Tracker Upgrade for the High-Luminosity LHC. In *Connecting the Dots and Workshop on Intelligent Trackers*, 11 2019.

[58] Malte Backhaus. The Upgrade of the CMS Inner Tracker for HL-LHC. Technical Report CMS-CR-2019-011, CERN, Geneva, Feb 2019.

[59] Jelena Luetić. The CMS Outer Tracker Upgrade for the High Luminosity LHC. *PoS*, Vertex 2017:003, 2018.

[60] Federico Ferri. The CMS ECAL Phase-2 upgrade for high precision energy and timing measurements. *Nucl. Instrum. Meth. A*, 958:162159, 2020.

[61] Arabella Martelli. The CMS HGCAL detector for HL-LHC upgrade. In *5th Large Hadron Collider Physics Conference*, 8 2017.

[62] CMS Collaboration. The Phase-2 Upgrade of the CMS Muon Detectors. Technical Report CERN-LHCC-2017-012. CMS-TDR-016, CERN, Geneva, Sep 2017. This is the final version, approved by the LHCC.

[63] Cristián H. Peña and. Precision timing with the CMS MIP timing detector. *Journal of Physics: Conference Series*, 1162:012035, jan 2019.

[64] G Petrucciani, A Rizzi, and C Vuosalo. Mini-AOD: a new analysis data format for CMS. *Journal of Physics: Conference Series*, 664(7):072052, 2015.

[65] A. M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12:P10003, 2017.

[66] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-$k_\mathrm{T}$jet clustering algorithm. *JHEP*, 04:063, 2008.

[67] Paolo Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.

[68] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.

[69] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.

[70] John M. Campbell, R. Keith Ellis, Paolo Nason, and Emanuele Re. Top-pair production and decay at NLO matched with parton showers. *JHEP*, 04:114, 2015.

[71] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction. *JHEP*, 09:126, 2007.

[72] Peter Skands, Stefano Carrazza, and Juan Rojo. Tuning PYTHIA 8.1: the Monash 2013 tune. *Eur. Phys. J. C*, 74:3024, 2014.

[73] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[74] Johan Alwall et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur. Phys. J. C*, 53:473, 2008.

[75] Rikkert Frederix and Stefano Frixione. Merging meets matching in MC@NLO. *JHEP*, 12:061, 2012.

[76] Emanuele Re. Single-top Wt-channel production matched with parton showers using the POWHEG method. *Eur. Phys. J. C*, 71:1547, 2011.

[77] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. NLO single-top production matched with shower in POWHEG: s- and t-channel contributions. *JHEP*, 09:111, 2009. [Erratum: 10.1007/JHEP02(2010)011].

[78] Richard D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015.

[79] S. Agostinelli et al. GEANT4—a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250, 2003.

[80] Albert M Sirunyan et al. Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV. *JHEP*, 07:161, 2018.

[81] Sayipjamal Dulat, Tie-Jiun Hou, Jun Gao, Marco Guzzi, Joey Huston, Pavel Nadolsky, Jon Pumplin, Carl Schmidt, Daniel Stump, and C. P. Yuan. New parton distribution functions from a global analysis of quantum chromodynamics. *Phys. Rev. D*, 93:033006, 2016.

[82] Ye Li and Frank Petriello. Combining QCD and electroweak corrections to dilepton production in FEWZ. *Phys. Rev. D*, 86:094034, 2012.

[83] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, and M. Wiedermann. HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR. *Comput. Phys. Commun.*, 182:1034, 2011.

[84] P. Kant, O. M. Kind, T. Kintscher, T. Lohse, T. Martini, S. Mölbitz, P. Rieck, and P. Uwer. HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions. *Comput. Phys. Commun.*, 191:74, 2015.

[85] Nikolaos Kidonakis. Two-loop soft anomalous dimensions for single top quark associated production with a W or H. *Phys. Rev. D*, 82:054018, 2010.

[86] Nikolaos Kidonakis. NNLL threshold resummation for top-pair and single-top production. *Phys. Part. Nucl.*, 45:714, 2014.

[87] John M. Campbell and R. K. Ellis. MCFM for the Tevatron and the LHC. *Nucl. Phys. Proc. Suppl.*, 205-206:10, 2010.

[88] John M. Campbell, R. Keith Ellis, and Ciaran Williams. Vector boson pair production at the LHC. *JHEP*, 07:018, 2011.

[89] V. Khachatryan et al. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8\,\mathrm{TeV}$. *JINST*, 10:P06005, 2015.

[90] Electron and photon performance in CMS with the full 2016 data sample. CMS Detector Performance Report CMS-DP-2017-004, 2017.

[91] Performance of muon identification in pp collisions at $\sqrt{s} = 7$ TeV. Technical Report CMS-PAS-MUO-10-002, CERN, Geneva, 2010.

[92] Muon identification and isolation efficiency on full 2016 dataset. Technical Report CMS-DP-2017-007, CERN, Mar 2017.

[93] Muon POG. Baseline muon selections for run-ii. Twiki, CERN, 2017.

[94] Electron reconstruction and identification at sqrt(s) = 7 TeV. Technical Report CMS-PAS-EGM-10-004, CERN, Geneva, 2010.

[95] EGamma POG. Cut based electron id for run 2. Twiki, CERN, 2017.

[96] Wolfgang Adam, R. Frühwirth, Are Strandlie, and T. Todor. Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC. 1 2005.

[97] Jet algorithms performance in 13 TeV data. Technical Report CMS-PAS-JME-16-003, CERN, Geneva, 2017.

[98] A. M. Sirunyan et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *JINST*, 13:P05011, 2018.

[99] Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector. Technical Report CMS-PAS-JME-16-004, CERN, Geneva, 2016.

[100] M. Aaboud et al. Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV with the ATLAS detector at the LHC. *Phys. Rev. Lett.*, 117:182002, 2016.

[101] Michal Czakon, David Heymes, and Alexander Mitov. High-precision differential predictions for top-quark pairs at the LHC. *Phys. Rev. Lett.*, 116:082003, 2016.

[102] Vardan Khachatryan et al. Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV. *Phys. Rev. D*, 95:092001, 2017.

[103] Vardan Khachatryan et al. Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8\,\mathrm{TeV}$. *Eur. Phys. J. C*, 75:542, 2015.

[104] Vardan Khachatryan et al. Measurement of the $\mathrm{t\bar{t}}$ production cross section in the all-jets final state in pp collisions at $\sqrt{s} = 8$ TeV. *Eur. Phys. J. C*, 76:128, 2016.

[105] Pietro Vischia. Search for a charged Higgs boson in $\tau\nu_{tau}$ and $t\bar{b}$ decays in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV with the CMS detector, Aug 2016. Presented 27 Jul 2016.

[106] Cristovao Beirao Da Cruz E Silva. Search for Direct Stau Pair Production at 8 TeV with the CMS Detector, 2017.

[107] Performance of tau identification with 2016 data at $\sqrt{s} = 13$ TeV. Technical Report CMS-DP-2017-006, CERN, Geneva, Mar 2017.

[108] CMS luminosity measurements for the 2016 data taking period. CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017.

[109] G. Abbiendi et al. Inclusive analysis of the b quark fragmentation function in Z decays at LEP. *Eur. Phys. J. C*, 29:463, 2003.

[110] A. Heister et al. Study of the fragmentation of b quarks into B mesons at the Z peak. *Phys. Lett. B*, 512:30, 2001.

[111] Koya Abe et al. Measurement of the b quark fragmentation function in Z^0 decays. *Phys. Rev. D*, 65:092006, 2002. [Erratum: 10.1103/PhysRevD.66.079905].

[112] J. Abdallah et al. A study of the b-quark fragmentation function with the DELPHI detector at LEP I and an averaged distribution obtained at the Z pole. *Eur. Phys. J. C*, 71:1557, 2011.

[113] M. G. Bowler. e^+e^- production of heavy quarks in the string model. *Z. Phys. C*, 11:169, 1981.

[114] C. Peterson, D. Schlatter, I. Schmitt, and P. M. Zerwas. Scaling violations in inclusive e^+e^- annihilation spectra. *Phys. Rev. D*, 27:105, 1983.

[115] Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV. CMS Physics Analysis Summary CMS-PAS-TOP-16-021, 2016.

[116] The LHC Higgs Combination Group The ATLAS Collaboration, The CMS Collaboration. Procedure for the LHC Higgs boson search combination in Summer 2011. Technical report, CERN, Geneva, Aug 2011.

[117] Georges Aad et al. Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV. *JHEP*, 08:045, 2016.

[118] L. Moneta, K. Belasco, K. S. Cranmer, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke, and M. Wolf. The RooStats Project. In *13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT2010)*. SISSA, 2010. PoS(ACAT2010)057.

[119] CMS Higgs Combine community. Combine. GihubBook, CERN, 2017.

[120] Albert M Sirunyan et al. Measurement of the top quark pair production cross section in dilepton final states containing one τ lepton in pp collisions at $\sqrt{s} =$ 13 TeV. *JHEP*, 02:191, 2020.

[121] S. Schael et al. Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP. *Phys. Rept.*, 532:119–244, 2013.

[122] Y. Amhis et al. HeavyFlavour Averaging Group (HFLAV). Average of $R(D)$ and $R(D^*)$ for spring 2019. Technical report, 2019. Available at `https://hflav-eos.web.cern.ch/hflav-eos/semi/spring19/html/RDsDsstar/RDRDs.html`.

[123] S. Dysch and T.R. Wyatt. A self-calibrating, double-ratio method to test tau lepton universality in W boson decays at the LHC. *Eur. Phys. J. C*, 80(2):155, 2020.