Comenius University in Bratislava

Faculty of Mathematics, Physics and Informatics

Study of associated production of the top-quark pair and Z boson

PhD Thesis

Bratislava, 2019

Mgr. Michal Dubovský

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics
Department of Nuclear and Subnuclear Physics

PhD Thesis

Study of associated production of the top-quark pair and Z boson

Mgr. Michal Dubovský

Study program: Nuclear and Subnuclear Physics
Supervisor: prof. RNDr. Stanislav Tokár, DrSc.
Consultant: Mgr. Pavol Bartoš, PhD.
Bratislava, 2019

55610095

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Mgr. Michal Dubovský
Študijný program: jadrová a subjadrová fyzika (Jednoodborové štúdium, doktorandské III. st., denná forma)
Študijný odbor: jadrová a subjadrová fyzika
Typ záverečnej práce: dizertačná
Jazyk záverečnej práce: anglický
Sekundárny jazyk: slovenský

Názov: Study of associated production of the top-quark pair and Z boson in proton-proton collisions at the ATLAS experiment.
Štúdium asociovanej produkcie top kvarkového páru a Z bozónu v protón-protónových zrážkach experimentu ATLAS.

Anotácia: Navrhovaná práca sa zaoberá fyzikou top kvarku v protón-protónových zrážkach na experimente ATLAS. Základná myšlienka tejto tematiky je tom, že top kvark sa produkuje na veľmi malých vzdialenostiach s charakteristickou silnou väzbovou konštantou na úrovni 0,1 (top kvark je perturbatívny objekt). Okrem toho sa top kvark rozpadá pred hadronizáciou, preto jeho vlastnosti nie sú „rozmazané“ hadronizáciou a jednoznačne sa prejavujú v jeho rozpadových produktoch.
Navrhovaná práca má byť zameraná na štúdium asociovanej produkcie top-kvarkového páru a Z bozónu. Účinný prierez tohto procesu je citlivý na väzbovú konštantu top kvarku a Z bozónu, takže meranie tohto účinného prierezu poskytuje možnosť precízneho testu Štandardného modelu (SM) resp. hľadania fyziky za SM.
Navyše, uvedený proces predstavuje významné pozadie pre iné kritické merania experimentu ATLAS, ako je asociovaná produkcia top-kvarkového páru a Higgsovho bozónu a taktiež pre rôzne hľadania fyziky za SM.
Konkrétny obsah práce bude závisieť od konkrétnych potrieb experimentu ATLAS a bude preto špecifikované neskôr

Cieľ: Štúdium asociovanej produkcie top-kvarkového páru a Z bozónu s dôrazom na anomálne väzbové konštanty top kvarku. Ciele budú upresnené v súlade s potrebami experimentu ATLAS.

Literatúra: 1. M. Beneke et al., Top Quark Physics, Proc. of the Workshop on Standard Model Physics at LHC, CERN 2000-004
2.The ATLAS collaboration, Expected Performance of the ATLAS Experiment, CERN-OPEN-2008-020, Geneva 2009
3. J.F. Donoghue, E.Golowich and B.R. Holstein, Dynamics of the Standard Model, Cambridge University Press, New York 1992

Poznámka: Výpočtové zručnosti a dobrá znalosť teórie sú vítané

Kľúčové slová: top kvark, Z bozón, b-jet, anomálne väzbové konštanty

Školiteľ: prof. RNDr. Stanislav Tokár, DrSc. (od 15.09.2015)

55610095

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname: Mgr. Michal Dubovský
Study programme: Nuclear and Subnuclear Physics (Single degree study, Ph.D. III. deg., full time form)
Field of Study: Nuclear And Subnuclear Physics
Type of Thesis: Dissertation thesis
Language of Thesis: English
Secondary language: Slovak

Title: Study of associated production of the top-quark pair and Z boson in proton-proton collisions at the ATLAS experiment.

Annotation: The proposed thesis deals with the top-quark physics in proton-proton collisions at the ATLAS experiment. The basic idea of this subject is in the fact that top quark is produced at very small distances with characteristic strong coupling constant at a level of 0.1 (the top quark is perturbative object).
Moreover, the top quark decays before hadronization therefore its properties are not diluted by the hadronization and should manifest themselves in the top quark decay products.
The thesis is aimed at study of associated top-quark pair production with Z boson, The cross section of this process is sensitive to coupling constant of top quark to Z boson, therefore the cross section measurement provides an opportunity for precision test of the Standard Model (SM) and search for the beyond the SM physics, respectively.
Additionally, the mentioned process is an important background for other crucial ATLAS measurements such as is the associated production of top-quark pair and Higgs boson, as well as are various searches for physics beyond the Standard Model.
The concrete contents of work will depend on specific needs of the ATLAS experiment and will therefore be specified later.

Aim: Study of the associated production of top-quark pair and Z boson with emphasis on the
top quark anomalous coupling constants. The goals will be specified in comply with needs of the experiment ATLAS.

Literature: 1. M. Beneke et al., Top Quark Physics, Proc. of the Workshop on Standard Model Physics at LHC, CERN 2000-004
2.The ATLAS collaboration, Expected Performance of the ATLAS Experiment, CERN-OPEN-2008-020, Geneva 2009
3. J.F. Donoghue, E.Golowich and B.R. Holstein, Dynamics of the Standard Model, Cambridge University Press, New York 1992

Comment: Computational skill and good knowledge of theory is welcome.

Keywords: top quark, Z boson, b-jet, anomalous couplings

Tutor: prof. RNDr. Stanislav Tokár, DrSc. (from 2015-09-15)
Consultant: Mgr. Pavol Bartoš, PhD. (from 2015-09-15)

Acknowledgement

I would like to thank my supervisor Stanislav Tokár for his help and support during my PhD study and working on this thesis. Even though being constantly busy with his duties as a head of the department and many other responsibilities, he always had a time to discuss all the analysis aspects when I needed it.

I would like to thank also all the members of the $t\bar{t}Z$ analysis group in the ATLAS collaboration in both 36 fb^{-1} and full Run II dataset analyses. The measurements would not be possible without cooperation of all of those people and the great effort made by them in all the aspects of the analyses. The main thanks goes to Elizaveta Shabalina and María Moreno Llácer for a lot of their help in the analysis. I cannot imagine my starts of working on this analysis without their help. I would like to thank also Serban Protopopescu who worked on the Boosted Decision Tree optimization in the dilepton channel.

In the second round of the analysis, at the full Run II dataset, I would like to thank Baptiste Ravina for a lot of his work, mainly on the development and maintaining the software for the ntuple production and unfolding part of the analysis. An important part of the work has been done also by Fabio Cardillo, especially in the fake lepton background estimate and the ntuple production. Many thanks goes also to Tom McCarthy and again to Fabio for many useful discussions about the analysis and all their work on the analysis as contact persons. I would like to thank also Dominik Babál, who did a lot of work on the differential cross section measurement in the 4ℓ channel.

Abstrakt

Táto práca prezentuje meranie totálneho účinného prierezu asociovanej produkcie top-kvarkového páru a Z bozónu (tak zvaný $t\bar{t}Z$ proces) v dátach zozbieraných detektorom ATLAS v protón-protónových zrážkach pri ťažiskovej energií 13 TeV.

Prvá časť práce opisuje prvé pozorovanie $t\bar{t}Z$ procesu ATLAS detektorom, v ktorom boli použité dáta zozbierané v rokoch 2015 a 2016, čo zodpovedá integrovanej luminozite 36.1 fb^{-1}. Táto čast je zameraná hlavne na analýzu v dileptonónovom kanáli, kedy sa Z bozón rozpadá na dva nabité leptóny a top-kvarkový pár sa rozpadá hadrónovo. Kvôli veľkému pozadiu v dileptónovom kanáli je treba použiť multivarietnu techniku (MVA). Optimalizácia MVA ako aj setu jej vstupných premenných je opísaná v práci. Fitovaním výstupu z MVA v troch signálnych regiónoch dileptónového kanála bola získaná nasledovná hodnota totálneho účinného prierezu $t\bar{t}Z$ procesu:

$$\sigma_{t\bar{t}Z}^{2l,measured} = 0.64^{+0.15}_{-0.15}(\text{stat.})^{+0.20}_{-0.19}(\text{syst.})\text{ pb} = 0.64^{+0.25}_{-0.24}\text{ pb}. \tag{1}$$

Pozaďová hypotéza bola vylúčaná s vierohnodnosťou 3.0 σ. Kombináciou 2ℓ kanálu s 3ℓ a 4ℓ $t\bar{t}Z$ rozpadovými kanálmi a po pridaní dvoch rozpadových kanálov $t\bar{t}W$, pozaďová hypotéza bola vylúčená na 8.9 σ a získaná hodnota totálneho účinného prierezu $t\bar{t}Z$ bola

$$\sigma_{t\bar{t}Z}^{measured} = 0.95\ \pm 0.08(\text{stat.})\ \pm 0.10(\text{syst.})\text{ pb}\ = 0.95\ \pm 0.13\text{ pb}. \tag{2}$$

Oba výsledky sú v súlade s predpoveďou Štantandardného Modelu.

Druhá časť práce sa zaoberá analýzou celého Run II datasetu, zodpovedajúcemu dátam z rokov 2015-2018 a integrovanej luminozite 139 fb^{-1}. Táto časť práce je zameraná na 4ℓ kanál, kedy sa Z bozón aj top-kvarkový pár rozpadajú dileptónovo. Táto časť práce prezentuje iba výsledky získané z Monte Carlo simulácií. Očakávaná hodnota účinného prierezu $t\bar{t}Z$ procesu získaná zo simulácie zodpovedajúcej plnému Run II datasetu je

$$\mu_{t\bar{t}Z}^{4\ell,\text{expected}} = 1.000^{+0.142}_{-0.132}\ (\text{stat.})\ {}^{+0.074}_{-0.068}\ (\text{syst.})\ = 1.000^{+0.160}_{-0.149}. \tag{3}$$

Klúčové slová: top kvark, Z bozón, $t\bar{t}Z$ asociovaná produkcia

Abstract

The thesis presents the measurement of the total cross section of the top-quark pair and Z boson associated production ($t\bar{t}Z$) at pp collisions data collected by the ATLAS detector at $\sqrt{s} = 13$ TeV.

The first part presents the first observation of the $t\bar{t}Z$ process by the ATLAS detector at 2015 and 2016 data corresponding to the luminosity of 36.1 fb^{-1}. This part is focused on the dilepton channel, with the leptonically decaying Z boson and hadronically decaying top-quark pair. Because of a high background rate in the 2ℓ channel, a multivariate analysis needs to be employed. Optimization of the MVA and the set of its input variables are described in details in the thesis. Fitting the MVA output in 3 signal regions of the 2ℓ channel, the following value of the $t\bar{t}Z$ total cross section has been obtained:

$$\sigma_{t\bar{t}Z}^{2l,measured} = 0.64^{+0.15}_{-0.15}(\text{stat.})^{+0.20}_{-0.19}(\text{syst.})\ \text{pb} = 0.64^{+0.25}_{-0.24}\ \text{pb}. \tag{4}$$

with 3.0 σ significance (exclusion of background only hypothesis).

Combining the 2ℓ channel with 3ℓ and 4ℓ $t\bar{t}Z$ channels and adding two $t\bar{t}W$ channels to the fit, 8.9 σ signal significance and the following value of cross section have been obtained:

$$\sigma_{t\bar{t}Z}^{measured} = 0.95\ \pm 0.08(\text{stat.})\ \pm 0.10(\text{syst.})\ \text{pb}\ = 0.95\ \pm 0.13\ \text{pb}. \tag{5}$$

The both results are compatible with the Standard Model prediction.

The second part of the thesis deals with analysis of the full Run II dataset collected by the ATLAS detector at pp collisions from 2015 to 2018. This part is focused on the 4ℓ channel, with both Z boson and top-quark pair decaying dileptonically. Only expected results obtained from a Monte Carlo simulation are presented. The expected cross section obtained from the 4ℓ channel fit at dataset corresponding to the full Run II dataset is

$$\mu_{t\bar{t}Z}^{4\ell,\text{expected}} = 1.000^{+0.142}_{-0.132}\ (\text{stat.})\ {}^{+0.074}_{-0.068}\ (\text{syst.}) \ = 1.000^{+0.160}_{-0.149}, \tag{6}$$

where the theory prediction has been taken as the unit of $\mu_{t\bar{t}Z}^{4\ell,\text{expected}}$.

Key words: top quark, Z boson, $t\bar{t}Z$ associated production,

Contents

Chapter 1

Introduction

The associated production of the top-quark pair and Z boson, the so-called $t\bar{t}Z$, is a rare process predicted by the Standard Model. The theory prediction of the $t\bar{t}Z$ cross section at 13 TeV pp collisions is 0.863 pb [1], which is approximately 1,000 times less than the cross section of the top-quark pair production. Given very short lifetimes of the Z boson and top quark, various decay channels are allowed. With the current experimental setup and available number of events, only the decay channels with the Z boson decaying into electron-positron or muon-antimuon are possible to identify. It restricts available number of $t\bar{t}Z$ events to 6.7 %, corresponding to the $Z \to e^-e^+/\mu^+\mu^-$ branching ratio [2].

The $t\bar{t}Z$ production provides a unique opportunity to probe the top-quark to Z-boson coupling constant. According to the Standard Model, this coupling constant is related to the top-quark weak isospin. Any deviation from the Standard Model prediction could be a sign of a new physics beyond the Standard Model. The $t\bar{t}Z$ production is also an important background for $t\bar{t}H$ measurements, which probe a coupling constant of the Higgs boson to the top quark [3, 4]. This coupling constant is related to one of the fundamental parameters of the Standard Model, top-quark Yukawa's coupling constant. The $t\bar{t}Z$ is also an important background for some searches for a new physics targeting multilepton final states [5]. A good understanding of the $t\bar{t}Z$ is therefore important for these analyses.

The first attempts to measure the $t\bar{t}Z$ cross section were made by ATLAS [6, 7] and CMS [8, 9] already during the Run I period of data taking at the LHC, at the center of mass energy 7 and 8 TeV.

The measurements at $\sqrt{s} = 7$ TeV were limited to a 3ℓ channel only, aiming at the leptonic decays of the Z boson and lepton+jets decays of the top-quark pair. The theory prediction of the $t\bar{t}Z$ cross section at 7 TeV is 0.137 pb [10], corresponding to $\approx$ 14 expected $t\bar{t}Z$ events in the 3ℓ channel in 5 fb^{-1} of available data[1]. The ATLAS collaboration observed only one event in a signal region, while 1.13 events were expected. The measurement set the upper limit of the $t\bar{t}Z$ cross section at 95 % confidence level to 0.74 pb [6]. The CMS collaboration measured the $t\bar{t}Z$ cross section obtaining the value $\sigma_{t\bar{t}Z} = 0.28^{+0.14}_{-0.11}(\text{stat.})^{+0.06}_{-0.03}(\text{syst.})$ pb, excluding the background only hypothesis at 3.3 σ [8].

In 2012, the Large Hadron Collider (LHC) delivered ≈ 20 fb^{-1} of data at $\sqrt{s} = 8$ TeV, resulting in significantly increased number of available data events. The theory prediction for the $t\bar{t}Z$ cross section at $\sqrt{s} = 8$ TeV energy is 0.206 pb [10]. The ATLAS collaboration measured the $t\bar{t}Z$ cross section in the combination of 2ℓ, 3ℓ and 4ℓ channels, corresponding to the leptonic decays of the Z boson and all-hadron, lepton+jets and dilepton decay channels of the top-quark pair. The ATLAS collaboration achieved the result of $0.176^{+0.058}_{-0.052}$ pb with 4.2 σ significance [7]. The CMS collaboration measured the $t\bar{t}Z$ cross section in 3ℓ and 4ℓ channels, achieving the result of $0.20^{+0.08}_{-0.07}(\text{stat.})^{+0.04}_{-0.03}(\text{syst.})$ pb and 3.1 σ significance [9].

In 2015, the LHC was operational again, reaching the center of mass energy of 13 TeV in pp collisions. The theory prediction of the $t\bar{t}Z$ cross section at 13 TeV is $0.863^{+8.5\%}_{-9.9\%}$ (scale) $^{+3.2\%}_{-3.2\%}$ (PDF + α_S) pb [1]. The ATLAS and CMS collaborations both used 2015 dataset (3.2 fb^{-1} collected by ATLAS, 2.7 fb^{-1} collected by CMS) to measure the $t\bar{t}Z$ cross section at 13 TeV, using the 3ℓ and 4ℓ decay channels. The ATLAS measured the cross section of $0.92 \pm 0.29(\text{stat.}) \pm 0.10(\text{syst.})$ pb, reaching 3.9 σ significance [11]. The CMS measured $1.07^{+0.35}_{-0.31}(\text{stat.})^{+0.17}_{-0.14}(\text{syst.})$ pb with 3.6 σ significance [12].

In 2016, the CMS collaboration published a result from 12.9 fb^{-1} dataset, obtaining the cross section of $0.70^{+0.16}_{-0.15}(\text{stat.})^{+0.14}_{-0.12}(\text{syst.})$ pb with 3.9 σ significance, taking into account the combination of the 3ℓ and 4ℓ channels [13].

Approximately at this point the analysis presented in this thesis began. Although there had already been many measurements of the $t\bar{t}Z$ cross section, none of them had reached 5.0 σ significance, which is the threshold needed to announce observation of a process in the high energy physics. Both CMS and ATLAS already approached this value with the data collected in 2015. Taking into account an increasing number of available data events, being collected by ATLAS detector, the ATLAS collaboration

---

[1] No detector acceptance and efficiency effects are considered. Only the luminosity, $t\bar{t}Z$ cross section and 3ℓ channel branching ratio have been considered.

decided to include the 2ℓ channel into the analysis, aiming for more than 5 σ significance with 2015+2016 dataset. The results of this ATLAS measurement[14] are described in details in this thesis, as well as a current ATLAS measurement at full Run II dataset, corresponding to 139 fb^{-1} of data.

In the meantime, the CMS collaboration published $t\bar{t}Z$ total cross section measurement at 35.9 fb^{-1}, obtaining the result of $0.99^{+0.09}_{-0.08}(\text{stat.})^{+0.12}_{-0.10}(\text{syst.})$ pb and reaching for the first time ever significance more than 5 σ. The combination of the 3ℓ and 4ℓ channels was used in this measurement [15].

The next paper was published in 2019 by the CMS collaboration, using the dataset corresponding to 2015-2017 period of data taking, measuring the total cross section of $1.00^{+0.06}_{-0.05}(\text{stat.})^{+0.07}_{-0.06}(\text{syst.})$ pb. This publication also includes the first $t\bar{t}Z$ differential cross section measurement, using two variables, the transverse momentum of the Z boson and an angular distribution of the negatively charged lepton from the Z boson. The combination of the 3ℓ and 4ℓ channels was used in this measurement [16].

All $t\bar{t}Z$ cross section results already published at the time of writing this thesis (spring 2019) are compatible with the Standard Model predictions [6, 7, 8, 9, 11, 12, 13, 14, 15, 16].

Chapter 2

Standard Model, Top quark, Z boson and $t\bar{t}Z$ production

2.1 Introduction

Human ideas of the structure of matter have been changing significantly through centuries. The first idea of atoms was postulated by Democritus in ancient Greece around 400 BC [17]. On the level of science and technology at those times it was impossible to provide an experimental evidence of atoms. Without a proof, the atomic theory was just one of many theories and it was forgotten for a long time.

At the end of 18th century, John Dalton, an English physicist and chemist, postulated the atomic theory again. He studied chemical reactions of gasses and ratios of gas volumes before and after the reactions. Dalton realised that the ratio of reacting gasses volumes is always a ratio of two small integer numbers. This fact suggested that during chemical reactions, atoms are regrouped to form a new compound but no atom can be created nor destroyed. Although Dalton thought that the atoms could not be split further, his experiments and postulations still meant a significant step in human knowledge on the matter structure.

The structure of atoms was unknown until the end of the 19th century. In 1897, Joseph John Thompson discovered an electron, while studying cathode emission [18]. It was the first elementary particle to be discovered. After his discovery, Thompson proposed a pudding model of the atom. The idea of the pudding model was that the atom was composed of positively charged liquid (pudding) and electrons floating inside the liquid like plums in the pudding. Light emission spectra of atoms were explained as electron oscillations in the liquid.

In 1905, Albert Einstein explained photoelectric effect by a quantum of light, a photon [19]. The photon, an elementary particle intermediating the electromagnetic force in nature, was the second one among the discovered elementary particles.

In 1909, E. Marsden and H.Geiger, both students of E. Rutherford, studied scattering of the alpha particles on a thin golden foil. Aim of the experiment was to measure distribution of the scattering angle and estimate the size of the golden atom from this distribution, still considering Thompson's pudding model of the atom. Surprisingly, back-scattering was observed significantly more often than the theory had predicted [20]. The experiment clearly showed that the positive charge of the atom is centred in a very small volume, now known as the atomic nucleus. When the alpha particle got too close to the nucleus, it was scattered backwards by the electromagnetic field of the nucleus.

In 1917, Rutherford performed the first known nuclear reaction. He bombarded the nitrogen gas by alpha particles. As the results of the reaction, the hydrogen nucleus, the proton, was formed [21].

$$^{14}_{7}N +^{4}_{2} \alpha \rightarrow^{1}_{1} p +^{17}_{8} O \tag{2.1}$$

Rutherford realised that masses of all the other known elements were almost integer multiples of the hydrogen mass. He assumed that all the atomic nuclei were formed by protons and electrons. The neutron still had not been discovered by that time.

In 1920s, various β-decays were already known. The energy distribution of the β particles is continuous for a given atom, while α and γ radiation energy spectra are discrete. The energy carried out by the α and γ radiation summed together with the kinetic energy of a daughter nucleus is equal to difference in the energies of the nuclei before and after the decay. According to the energy conservation principle, it was expected that the β radiation should also have the discrete spectra. In general, two possibilities were considered: either the energy was not conserved in β decay, or there was at least one more particle escaping undetected. Violation of the energy conservation in β decays seemed to be unlikely, since the energy was conserved in all other known processes. In 1930 Wolfgang Pauli postulated existence of a neutrino, a particle without an electric charge escaping undetected from the β decay, carrying a part of the energy and momentum away.

In 1932, the neutron was discovered by James Chadwick [22]. Bombarding a beryllium surface by alpha particles he measured a radiation caused by neutral particles which could not have been photons. Its absorption length in lead was longer than it could be for the photons with the same energy. Chadwick estimated the neutron mass to be approximately the same as the proton mass.

Chadwick's discovery of the neutron was an important milestone for nuclear physics. According to the previous experiments, atoms consist of a nucleus and an electron shell surrounding the nucleus. Electrons are bound in atoms by the electromagnetic force, intermediated by photons. An atomic nucleus is formed by neutrons and protons. Beta decays could have been explained by the weak nuclear force, which was assumed to be a contact force, so a neutron was assumed to decay directly into 3 daughter particles: a proton, positron and neutrino. The protons in atomic nuclei were assumed to be bound by the strong force, intermediated by spinless pions. Although the pions had not been experimentally observed yet, from a known range of the strong nuclear force their mass had been estimated to be $\approx$ 100 MeV. All the particles were assumed to be discovered already, except of the pions. The pions should have been the last piece of the puzzle to be found to have the complete model of nuclear/particle physics. As we know today, it was a completely wrong assumption.

After the neutron discovery a lot of effort was made in order to find the pions. In 1936-1937, a new particle with the mass of 106 MeV was found in cosmic rays [23]. For a moment it seemed to be the last missing piece, the pion, since the mass was in agreement with the expected pion mass and no other particle was expected to be found. Further studies showed the new particle did not interact by the strong interaction. It could not have been the pion. The new particle had similar properties to the electron, differing only by the mass. It is more than 200 times heavier that the electron. The particle was called muon. The muon confused physicists for a long time. It was not predicted by theory and it was difficult to find a place for it in current models.

In 1947 ($\pi^{\pm}$) and in 1950 (π^{0}) the pions were finally discovered. Their discovery was followed by discoveries of kaons and many other new particles.

From mesons to quarks

In 1960s many new mesons and baryons were known. Gell-Man [24] and George Zweig [25] realised there are two groups of nine mesons, each with the same spin and approximately the same mass. The list of these mesons can be found in Table 2.1. Their idea was as follows. If there were 3 fundamental particles (u, d and s quarks) and their antiparticles, all having exactly the same properties, there would be 9 composite particles, eight from octet and one from singlet, with the same mass (energy spectrum). If the symmetry was broken, for example if their electric charges or masses were not exactly the same, differences in mass spectra would be observed, but the particles would still form the octet and singlet. Since the mesons differ by their charge, it could have been expected that at least the electromagnetic interaction breaks the symmetry. The $SU(3)$ symmetry of the three quarks was called flavour, or $SU(3)_{\mathrm{f}}$ symmetry. It is not an exact symmetry, because of the different masses and charges of the u, d and s quarks.

s = 0			s = 1		
meson	quarks	mass	meson	quarks	mass
π^+	$u\bar{d}$	140 MeV	ρ^+	$u\bar{d}$	775 MeV
π^0	$\frac{1}{\sqrt{2}}(u\bar{u} - d\bar{d})$	135 MeV	ρ^0	$\frac{1}{\sqrt{2}}(u\bar{u} - d\bar{d})$	775 MeV
π^-	$d\bar{u}$	140 MeV	ρ^-	$d\bar{u}$	775 MeV
K^+	$u\bar{s}$	494 MeV	K^{*+}	$u\bar{s}$	892 MeV
K^0	$d\bar{s}$	498 MeV	K^{*0}	$d\bar{s}$	896 MeV
$\bar{K}^0$	$s\bar{d}$	498 MeV	$\bar{K}^{*0}$	$s\bar{d}$	896 MeV
K^-	$s\bar{u}$	494 MeV	K^{*-}	$s\bar{u}$	892 MeV
η	$\eta_8.cos(\theta_P) - \eta_1.sin(\theta_P)$	548 MeV	ω	$\frac{1}{\sqrt{2}}(u\bar{u} + d\bar{d})$	783 MeV
η'	$\eta_8.cos(\theta_P) + \eta_1.sin(\theta_P)$	958 MeV	φ	$s\bar{s}$	1019 MeV

$$\eta_8 = \frac{1}{\sqrt{6}}(u\bar{u} + d\bar{d} - 2s\bar{s}) \qquad \eta_1 = \frac{1}{\sqrt{3}}(u\bar{u} + d\bar{d} + s\bar{s})$$

Table 2.1: Mesons known in 1960s grouped together depending on their spin and mass. Pions, kaons and η_8 form the octet and η_1 is the singlet of $SU(3)_f$. Physical particles, mass eigenstates η and η', are linear combinations of η_8 and η_1 [26, 2].

The mass spectra of mesons was not the only thing suggesting the hadrons are composite particles. Collider experiments showed interesting results as well. For electron-proton collisions, the dependency of the cross section on the square of transferred momentum (q^2) was measured. At low values of q^2, the cross section was similar to scattering of two point-like objects, like $e\mu$ scattering for example. For higher q^2, where the wavelength of the electron approaches the proton size, the cross section starts to be different from the point-like scattering. It was expected, since the proton size must have been taken into account, using so-called form-factors. The interesting and surprising fact was, that for very high values of q^2 (above 1 GeV), the cross-section dependency was again similar to scattering of two point-like objects. This suggested that the proton is a composite particle, formed by point-like constituents (quarks) and for high energy of the electrons, the proton-electron scattering can be simplified to the electron scattering on a quark. In some sense it was a similar observation to the back-scattering of alpha particle in Rutheford's experiment. When the alpha particle in Rutheford's experiments has energy high enough, it interacts only with a part of the atom, the atomic nucleus. For a high energy of the electron it interacted only with one constituent of the proton, the quark.

2.2 Standard Model

The Standard Model is the theory of elementary particles, describing their fundamental properties and interactions. It was formulated in 1960s and 1970s, after formulation of the quark hypothesis. The Standard Model combines the electromagnetic and weak interaction and adds Higgs mechanism as a way how particles gain their masses. There are four fundamental forces in nature: gravitational, electromagnetic, weak nuclear force and strong nuclear force. The Standard Model describes all of them except of the gravitational and thus the gravity will not be further discussed in this thesis. The forces differ by their range, strength (coupling constant) and types of particles that are affected by them.

Fundamental particles predicted by the Standard Model can be categorized into three groups based on their spin: scalar boson, fermions and vector bosons.

2.2.1 Fermions

The largest group is formed by particles with spin 1/2, fundamental fermions. The fermions are further divided into two groups: quarks and leptons. Each quark and lepton has its anti-particle.[1] There are three generations of the fermions, each having 2 quarks and 2 leptons. The fermions from higher generations have higher masses.[2] The fermions of the same generation differ by their electric charges.

Quarks

There are six types of quarks: d, u, s, c, b and t. The quarks interact strongly, electromagnetically and weakly. All the quarks carry an electric charge. The up-type quarks (u, c and t) carry electric charge of +2/3, the down-type quarks (d, s and b)

---

[1] This is not fully understood for neutrinos yet.

[2] This applies for the quarks and charged leptons. Ordering of the neutrino masses is not fully understood yet.

carry charge of -1/3.[3]

The quarks form bound states, hadrons. In 1960s, hadrons with three quarks of the same flavour were known. Since the quarks are fermions, they follow Fermi-Dirac statistics, so two quarks cannot be in the same state. The two quarks of the same flavour in a hadron could have been explained by two different orientations of the spin, but the third quark would not be allowed, without an additional quantum number, which would make it different from the others. This quantum number has been called colour and it represents a charge of the strong interaction.

There are three colour charges in the strong interaction: red, green and blue. The anti-quarks carry anti-colours: anti-red, anti-green and anti-blue. Measurable quantities does not depend on the colour and all measurable quantities must be invariant with respect to rotations in the colour space. This is the exact and fundamental $SU(3)$ symmetry of the Standard Model and it is the main principle of the quantum chromodynamics (QCD). It will be further discussed in dedicated QCD chapter 2.2.4.

In the nature, only white (colourless) states are allowed. Since the quarks are colour objects, they cannot exist as free particles. They form colourless objects, hadrons. In principle there are two ways, and their combinations, how the colourless object can be obtained from coloured objects.

The first one is to combine a colour with its anti-colour. This covers mesons, bound states of a quark and anti-quark. The other way how the colourless object can be built is the combination of all three colours, or anti-colours. This covers baryons, bound states of three quarks or three anti-quarks. The protons and neutrons, particles forming atomic nuclei, are baryons formed by uud quarks (proton) and udd quarks (neutron). Except of the valence quarks, the hadrons are formed also by the gluons and sea quark-antiquark pairs, it will be further described in Chapter 2.2.8. There has been already an observation of penta-quark, formed by four quarks and one anti-quark, or four anti-quarks and one quark [27].

A typical time needed for the quarks to form a hadron is $\approx 10^{-24}$ s. The only quark with the lifetime shorter than the mean hadronization time is the top quark. It is the only quark that can be studied as a free quark. All the other quarks can be studied only in hadrons.

---

[3] Of the proton charge (elementary charge). The charge of proton is commonly used in the high energy physics as the unit of the charge.

Leptons

The leptons are grouped into three pairs (generations), each containing a charged lepton and electrically neutral neutrino: e, ν_e, μ, ν_μ, τ and ν_τ. The muons and taus are unstable and decay either into leptons of previous generations, or into hadrons, together with their neutrinos. The mean lifetime of the muon, $\tau_\mu = 2.2$ μs [2], is long enough to enable a direct detection of the muons. In the high energy physics, the muon can be considered to be a stable particle, since the muons with an energy of few GeVs are stable enough to fly tens of kilometers before their decay. However, the tau-lepton mean lifetime is very short, $\tau_\tau = 2.9 \times 10^{-13}$ s [2], which is not enough to enable a direct observation of the tau leptons. The tau-lepton decays either leptonically (35 %), or hadronically (65 %) [2].

Leptons do not interact strongly. Charged leptons interact weakly and electromagnetically. The neutrinos do not have an electric charge so they interact only weakly. The absence of the electric charge makes them difficult to detect. Neutrinos were predicted in 1930, but the first experimental observation of the neutrino was performed almost 30 years later, in 1956 [28]. The most common lepton, observed in the world around us, is the electron. It is charged lepton belonging to the first generation of the leptons.

2.2.2 Vector bosons and particle interactions

The second important group of the elementary particles predicted by the Standard Model are fundamental vector bosons, that are intermediators of the interactions in the Standard Model. There are four fundamental vector bosons in the Standard Model: photons intermediating the electromagnetic interaction, gluons intermediating the strong interaction, Z boson and W bosons intermediating the weak interaction.

2.2.3 Quantum electrodynamics (QED)

Lagrangian of a free fermion field Ψ with mass and kinetics terms is given as

$$\mathcal{L}_{free}(\Psi) = \bar{\Psi}(i\gamma^\mu \partial_\mu - m)\Psi, \tag{2.2}$$

where Ψ is a bispinor of the fermion field and γ^μ are Dirac matrices (4×4 matrices) defined as

$$\gamma^0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \gamma^{1,2,3} = \begin{pmatrix} 0 & \vec{\sigma}_k \\ -\vec{\sigma}_k & 0 \end{pmatrix}, \gamma^5 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}, \tag{2.3}$$

where $\vec{\sigma}_k$ are Pauli's matrices $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

If physics given by measurable quantities does not depend on the phase of Ψ function, the Lagrangian must be invariant under U(1) local transformations

$$\Psi \rightarrow \Psi' = \Psi e^{i\alpha(\mathbf{x})} \tag{2.4}$$

Transforming the free Lagrangian according to this transformation, the following equation can be obtained.

$$\mathcal{L}_{free}(\Psi') = \bar{\Psi}(i\gamma^\mu\partial_\mu - m)\Psi - \bar{\Psi}\gamma^\mu(\partial_\mu e^{i\alpha(\mathbf{x})})\Psi \tag{2.5}$$

In order to obtain a Lagrangian invariant under $U(1)$ transformation of Ψ, a new term, corresponding to interaction of fermion field with a field A_μ, needs to be added. The field A_μ follows the transformation:

$$A_\mu \rightarrow A_\mu + \frac{i}{e}(\partial_\mu e^{i\alpha(\mathbf{x})})e^{-i\alpha(\mathbf{x})} = A_\mu + \frac{i}{e}\partial_\mu\alpha(\mathbf{x}) \tag{2.6}$$

Adding the interaction term of the fermion field with the field A_μ, the QED Lagrangian can be obtained.

$$\mathcal{L}_{QED} = \bar{\Psi}(i\gamma^\mu\partial_\mu - m)\Psi - ie\bar{\Psi}\gamma_\mu\Psi A^\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{2.7}$$

where $F_{\mu\nu} = \partial_\mu A_\nu$ - $\partial_\nu A_\mu$ and e is an electric charge of the fermion.

The second term corresponds to the interaction of the photons with the fermions and it makes the QED Lagrangian invariant with respect to the local $U(1)$ transformations. The last term is not necessary in order to keep the invariance. It is the term corresponding to a free photon field.

The Equation 2.7 can be rewritten into the following form:

$$\mathcal{L}_{QED} = \bar{\Psi}(i\gamma^\mu D_\mu - m)\Psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{2.8}$$

where $D_\mu = \partial_\mu + ieA_\mu$ is the covariant derivative.

The photons are massless particles (no mass term in the Lagrangian), without an electric charge (A_μ is a real function) and with the spin equal to one (A_μ is four-component function). They intermediate the electromagnetic interaction in the Standard Model.

2.2.4 Quantum chromodynamics (QCD)

While there is only one electric charge, there are three colour charges in the strong interaction: green, blue and red. Only the quarks and gluons carry the colour charge. As experiments show, measured physical quantities cannot depend on a colour charge and only white states are allowed to exist as free particles. The $SU(3)$ colour symmetry is a fundamental symmetry of the Standard Model. It means that the Standard Model Lagrangian must be invariant under the $SU(3)_c$ transformations. The three component Ψ for the quarks must be invariant under the following local transformation:

$$\Psi \to \Psi e^{i\frac{1}{2}\lambda_a \alpha^a(\mathbf{x})}, \tag{2.9}$$

where $a = [1,\ 2,\ \ldots\ 8]$ and λ_a are the following matrices [29]. The invariance of the lepton fields with respect to the transformation is guaranteed by the fact that the leptons do not carry the colour charge and therefore they are singlets with respect to the $SU(3)_c$ transformation.

$$\lambda_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \lambda_2 = \begin{bmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \lambda_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \lambda_4 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$\lambda_5 = \begin{bmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{bmatrix} \qquad \lambda_6 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \qquad \lambda_7 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{bmatrix} \qquad \lambda_8 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

Following the procedure of building an invariant Lagrangian, similar to the one used for QED, a new interaction term can be added. The important difference with respect to QED is that the $U(1)$ transformation (Equation 2.4) has only one generator (equal to one), while the QCD has 8 generators of the group transformations, $\frac{1}{2}\lambda_a$ for $a =$ [1,2, ... 8].

When the transformation given by Equation 2.9 is applied to the free fermion field Lagrangian and only linear terms are considered, eight new terms, violating the $SU(3)$ invariance will appear. In order to make the Lagrangian invariant, eight independent fields, corresponding to the gluon octet, have to be added in order to compensate the terms violating the SU(3) invariance. The process of adding the new gauge field is similar to the approach used in the QED.

$$\mathcal{L}_{QCD} = \bar{\Psi}(i\gamma^\mu\partial_\mu - m)\Psi - g\bar{\Psi}\gamma_\mu\Psi A_a^\mu \frac{1}{2}\lambda_a - \frac{1}{4}G^a_{\mu\nu}G_a^{\mu\nu}, \tag{2.10}$$

where $G^a_{\mu\nu} = \partial_\mu A^a_\nu$ - $\partial_\nu A^a_\mu$ - $gf^{abc}A^b_\mu A^c_\mu$ and g is the strong interaction coupling constant and A^μ_a is the gluon field.

The f^{abc} is defined as follows [30]:

$$\begin{aligned} f^{123} &= 1 \\ f^{147} &= \text{-}f^{156} = f^{246} = f^{257} = f^{345} = \text{-}f^{367} = \tfrac{1}{2} \\ f^{458} &= f^{678} = \tfrac{\sqrt{3}}{2} \\ f^{abc} &= 0 \text{ otherwise} \end{aligned}$$

The QCD Lagrangian can be rewritten using the covariant derivative into the more convenient form:

$$\mathcal{L}_{QCD} = \bar{\Psi}(i\gamma^\mu D_\mu - m)\Psi - \frac{1}{4}G^a_{\mu\nu}G_a^{\mu\nu}, \tag{2.11}$$

where $D_\mu = \partial_\mu + igA^a_\mu \frac{1}{2}\lambda_a$ is the covariant derivative.

The gluons are massless particles, with unity spin, intermediating the strong interaction in the Standard Model. There are 8 types of the gluons, differing by the colour charge they carry [30]:

$|R\bar{G}>$, $|R\bar{B}>$, $|G\bar{R}>$, $|G\bar{B}>$, $|B\bar{G}>$, $|B\bar{R}>$,
$\frac{1}{\sqrt{2}}\left(|R\bar{R}> - |G\bar{G}>\right)$, $\frac{1}{\sqrt{6}}\left(|R\bar{R}> + |G\bar{G}> - 2|B\bar{B}>\right)$.

2.2.5 Electroweak interaction

In the QED and QCD the way how to define gauge field was straightforward. There is a known symmetry of the Standard Model and requiring the Lagrangian to be invariant under the transformations corresponding to the symmetry, the gauge fields could be obtained.

In the electroweak interaction the way, how the theory was built, was inverted. Weak interaction was known a long time before the Standard Model. Various β decays, muon decay and weak decays of hadrons had been known. In 1956, the parity violation in weak interaction was discovered, measuring the angle between the intensity of an external magnetic field and electrons emitted from the β decay of ^{60}Co [31]. Because of the parity violation, only left-handed (L) particles and right-handed (R)

antiparticles interact by weak interaction. The left-handed and right-handed stand for chirality eigenstates. For massless particles and also in ultrarelativistic limit for massive particles, the chirality and helicity eigenstates match each others. The left chirality eigenstate u_L is defined by $\frac{1}{2}(1\text{-}\gamma^5)$ operator as follows

$$u_L = \frac{1}{2}(1-\gamma^5)u. \tag{2.12}$$

Considering existence of a charged heavy vector boson W, intermediating the weak interaction, such as β decays, the interaction term given by Equation 2.13 has to be added into the Lagrangian, so far considering only interactions of electrons and electron neutrinos. Similar terms have to be added for interactions of μ and τ leptons. The weak interaction of quarks is different and will be described later.

$$\mathcal{L}_{int} = \frac{g_W}{\sqrt{2}}(\bar{\nu}_e, \bar{e})_L \left(\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} W_\mu^+ + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} W_\mu^- \right) \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L , \tag{2.13}$$

where g_W is a coupling constant, W_μ^+ and W_μ^- are fields corresponding to W bosons.

The interaction Lagrangian rises the question, what symmetry does generate these fields? It is similar to $SU(2)$ symmetry, since these matrices are linear combinations of the first two Pauli's matrices, σ_1 and σ_2. But there is not the third Pauli's matrix σ_3. Adding a new interaction term with σ_3, the Lagrangian would become invariant with respect to the $SU(2)$ transformations. Artificial addition of the term with σ_3 matrix leads to the interaction Lagrangian:

$$\mathcal{L}_{int} = \frac{g_W}{\sqrt{2}}(\bar{\nu}_e, \bar{e})_L \left(\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} W_\mu^+ + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} W_\mu^- + \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} W_\mu^3 \right) \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L . \tag{2.14}$$

The term with σ_3 corresponds to an electrically new neutral field, since there are vertices $W^3 e\bar{e}$ and $W^3 \nu_e \bar{\nu}_e$. Since the field W^3 couples to the neutrino, so it cannot be the photon. The $U(1)$ local symmetry with respect to a hypercharge Y

$$Y = 2(Q - T_3), \tag{2.15}$$

is required where Q is the electric charge of the particle and T_3 is third component of weak isospin (+1/2 for left-handed up-type quarks, -1/2 for left-handed down-type quarks, +1/2 for left-handed neutrinos, -1/2 for left-handed charged leptons and 0 for right-handed fermions).

Adding the term corresponding to the hypercharge $U(1)$ symmetry, the interaction part of the Lagrangian can be obtained:

$$\mathcal{L}_{int} = (\bar{\nu_e}, \bar{e})_L \left(\frac{g_W}{\sqrt{2}} \sigma^+ W_\mu^+ + \frac{g_W}{\sqrt{2}} \sigma^- W_\mu^- + \frac{g_W}{2} \sigma_3 W_\mu^3 + \frac{g'}{2} \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} B_\mu \right) \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L + (\bar{\nu_e}, \bar{e})_L \frac{g'}{2} \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix} B_\mu \begin{pmatrix} \nu_e \\ e \end{pmatrix}_R , \tag{2.16}$$

where σ^+, σ^- and σ_3 are corresponding matrices from Equation 2.14, $\frac{g'}{2}$ is a coupling constant of the new field (factor 1/2 is there for some practical reasons, that will be clear later). The new term, coming from the $U(1)$ symmetry cannot be the photon field, because it interacts with neutrinos. Combining the last two terms, a field interacting only with charged particles and thus the interaction part of the Lagrangian can be obtained, corresponding to the fields Z_μ and A_μ reads

$$\mathcal{L}_{int} = (\bar{\nu_e}, \bar{e}) \left(\frac{-g_W}{2.\cos(\vartheta_W)} \begin{pmatrix} \frac{1}{2}(1-\gamma^5) & 0 \\ 0 & \frac{1}{2}(1-\gamma^5) - 2\sin^2\vartheta_W \end{pmatrix} Z_\mu + q \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} A_\mu \right) \begin{pmatrix} \nu_e \\ e \end{pmatrix} , \tag{2.17}$$

where fields Z_μ and A_μ correspond to physical particles, Z boson and photon and $q = e = g.sin(\vartheta_W)$ is the charge of the electron. They are linear combinations of fields W_μ^3 and B_μ:

$$A_\mu = B_\mu \cos\vartheta_W + W_\mu^3 \sin\vartheta_W \tag{2.18}$$

$$Z_\mu = B_\mu \cos\vartheta_W - W_\mu^3 \sin\vartheta_W \tag{2.19}$$

The $\vartheta_W = \arctan\,(g'/g_W)$ is the weak mixing angle, which can be related to the Z and W boson masses [2]:

$$\sin^2\vartheta_W = 1 - \frac{M_W^2}{M_Z^2} \approx 0.2299, \tag{2.20}$$

where $M_Z = 91.1876 \pm 0.0021$ GeV and $M_W = 80.385 \pm 0.015$ GeV [2].

It should be emphasized that there is an important difference between interactions intermediated by W bosons and those intermediated by Z and γ. The W bosons interact only with left-handed particles and right-handed anti-particles. Z boson interacts with left-handed neutrinos and with both electrons (although the coupling is not the same for the left-handed and right-handed). The photon does not make a difference between right handed and left handed particles.

The electroweak interaction has $SU(2)_L \otimes U(1)_Y$ symmetry, where $SU(2)_L$ part is related to the weak isospin of left-handed particles and $U(1)_Y$ is related to the hypercharge.

Weak interaction of quarks and CKM matrix

The weak interaction is the only interaction allowing the quark flavour to change and thus it is the only way how heavy charged mesons can decay. Replacing $\begin{pmatrix} \nu_e \\ e \end{pmatrix}$ by $\begin{pmatrix} u \\ d \end{pmatrix}$ and adding independent terms for another two quark generations would not work, although it works for leptons. It would not allow decays of s and b quarks, since the only allowed decay vertices would be into the other quarks from the same generation, but it is not kinematically allowed because they are lighter from their isospin partners. In this case charged kaons and b-mesons would be stable, which does not reflect experimental observations. Vertices mixing quarks from different generations must be allowed. The $\begin{pmatrix} \nu_e \\ e \end{pmatrix}$ are replaced by $\begin{pmatrix} u \\ d' \end{pmatrix}$ and similarly for s', c, b' and t for quarks of the higher generations. The important change with respect to the weak interaction of leptons is that these states do not match the eigenstates of mass and the flavours which are conserved in strong and electromagnetic interactions. These states d', s' and b' are linear combinations of d, s and b states. The relation between these two flavour bases is given by Cabbibo-Cobayashi-Maskawa (CKM) matrix as follows.

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{2.21}$$

where q are quark eigenstates of QED, QCD and mass, q' are quark eigenstates of the weak interactions and V_{ij} are elements of CKM matrix. Their absolute values are [2]:

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97343^{+0.00011}_{+0.00012} & 0.22506 \pm 0.00050 & 0.00357 \pm 0.00015 \\ 0.22492 \pm 0.00050 & 0.97351 \pm 0.00013 & 0.0411 \pm 0.0013 \\ 0.00875^{+0.00032}_{-0.00033} & 0.0403 \pm 0.0013 & 0.99915 \pm 0.00005 \end{pmatrix}. \tag{2.22}$$

2.2.6 Coupling constants in the Standard Model

When calculating a cross section or a decay width in the Standard Model, a factor corresponding to a vector boson propagator has to be taken into account in the calculations. Loops in the propagator, together with integral over fourmomentum of the propagator, have to be taken into account.

The Standard Model is based on the $SU(3)_c \times SU(2)_L \times U(1)_Y$ symmetry. Each of these symmetries is related to its coupling constant of the compensating boson field. The coupling constants depend on transferred momentum.

There are important differences between the photon, gluon and Z/W propagators, having important physical consequences, which should be noted.

QED

Considering electron-positron scattering in QED, the Feynman diagrams showed in Figure 2.1 have to be taken into account.

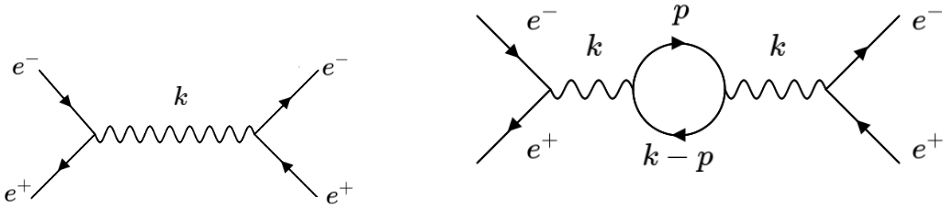


Figure 2.1: Leading order (left) and next to leading order (right) Feynman diagrams for the electron-positron scattering. These are not the only diagrams for this process, but the aim of this chapter is not to calculate the cross section, it is meant to illustrate common patterns in QED diagrams, leading to a dependency of the coupling constant on the transferred momentum.

The matrix element corresponding to the leading order diagram is as follows,

$$M_{fi} = -i[u(e^-_{in})\,(-ie\gamma^\mu)\,\bar{v}(e^+_{in})]\frac{-ig_{\mu\nu}}{k^2}[v(e^+_{out})\,(-ie\gamma^\nu)\,\bar{u}(e^-_{out})], \tag{2.23}$$

where u, v, $\bar{u}$ and $\bar{v}$ are bispinors corresponding to the incoming and outgoing particles. The k is a transferred momentum, the fourmomentum of the photon propagator, which is sum of e^+e^- fourmomenta in this case.

Calculating a contribution from one loop diagram (also called next to leading order, or NLO), only the factor for propagator, $\frac{-ig_{\mu\nu}}{k^2}$ will change, according to the transformation

$$\frac{-ig_{\mu\nu}}{k^2} \to \frac{-ig_{\mu\nu}}{k^2} + \frac{-i}{k^2} I_{\mu\nu}(k^2) \frac{-i}{k^2}, \tag{2.24}$$

where $I_{\mu\nu}(k^2)$ is the factor corresponding to the loop correction.

Calculating the integral and splitting the propagator correction factor between the both vertices, so-called running coupling constant can be obtained. The running coupling constant accounts for higher order corrections of the propagator and it depends on the transferred momentum. For high transferred momentum ($\mu^2 >> M_e^2$) it follows

$$\alpha_{EM}(Q^2) = \frac{\alpha_{EM}(\mu^2)}{1 - \frac{\alpha_{EM}(\mu^2)}{3\pi} ln \frac{Q^2}{\mu^2}}, \tag{2.25}$$

The Equation 2.25 shows that the coupling constant of the electromagnetic interaction, increases for high value of the transferred momentum.

The value of the electromagnetic coupling constant at low transferred momentum, $k^2 = 0$, is $\alpha_{\text{EM}} = (e/4\pi)^2 \approx 1/137$.

QCD

The important difference between QCD and QED is caused by the colour charge of glouns. Gluons themselves are colour objects, so they couple to other gluons, while photons do not carry the electric charge, so they do not couple to other photons. The gluon self interaction adds a new term to loop corrections, corresponding to the gluon loop in the propagator. The NLO Feynman diagrams for $q\bar{q}$ scattering are shown in Figure 2.2

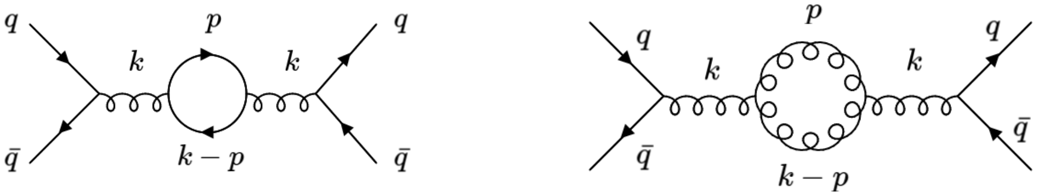


Figure 2.2: Next to leading order Feynman diagrams for $q\bar{q}$ scattering in the QCD with quark (left) and gluon (right) loop. Contrary to the QED, the QCD has the additional term with gluon loop, coming from the gluon self-interaction.

Calculating the correction factor, corresponding to the propagator loops, similar procedure as for the QED has to be applied. Correction to the propagator can be again transformed to the vertex correction, corresponding to the change of the QCD coupling constant

$$\alpha_S(Q^2) = \frac{\alpha_S(\mu^2)}{1 + \frac{\alpha_S(\mu^2)}{4\pi}(11 - \frac{2}{3}n_f) ln\frac{Q^2}{\mu^2}}, \tag{2.26}$$

where $\alpha_S(Q^2)$ and $\alpha_S(\mu^2)$ are strong interaction coupling constants corresponding to Q^2 and μ^2 scales and n_f stands for number of quark flavours. The reference scale is usually taken as $\mu^2 = M_Z^2$. The value of the strong coupling constant at this scale is $\alpha_S(M_Z^2) = 0.1181 \pm 0.0011$ [2]. The strong coupling constant diverges when the denominator approaches zero. The Q^2 value for which the denominator is equal to zero is called Λ^2_{QCD}. Eq. 2.26 can be rewritten as:

$$\alpha_S(Q^2) = \frac{4\pi}{\left(11 - \frac{2}{3}n_f\right) ln\frac{Q^2}{\Lambda^2_{QCD}}}, \tag{2.27}$$

where $\Lambda_{QCD} \approx 300$ MeV [32]. Equation 2.26 has important physical consequences.

1. The QCD coupling constant decreases with rising Q^2. In the limit case $Q^2 \rightarrow \infty$ it approaches zero. This causes asymptotic freedom of quarks. At very high energies, the strong interaction coupling constant is almost zero and quarks behave as free particles.

2. The QCD coupling constant diverges for $Q^2 \rightarrow \Lambda^2_{QCD}$. It is closely related to the quark confinement. The quarks at low energies are strongly bound in hadrons. If the distance between two quarks is too high, a new quark-antiquark pair is created from their potential energy and two new hadrons are formed. This causes also the limited range ($\approx 10^{-15}$m) of the strong interactions.

Weak interactions

The weak interaction was known long before the Standard Model. It was observed in many decays. It was assumed to be a contact interaction, without Z or W bosons, with a direct decay of one fermion into three others. In other words, there was a four fermion vertex, and corresponding four fermion interaction in Lagrangian. This theory provided sufficiently good results for various decays. The factor for W boson propagator in the Standard Model for low transferred momentum $|q^2| << m_W^2$ can be replaced as

$$\frac{-i(g_{\mu\nu} - q_\mu q_\nu/m_W^2)}{q^2 - m_W^2} \rightarrow \frac{ig_{\mu\nu}}{m_W^2} \tag{2.28}$$

The Feynman diagrams for the muon decay in the Standard Model and in Fermi's theory of the weak interaction can be seen in Figure 2.3

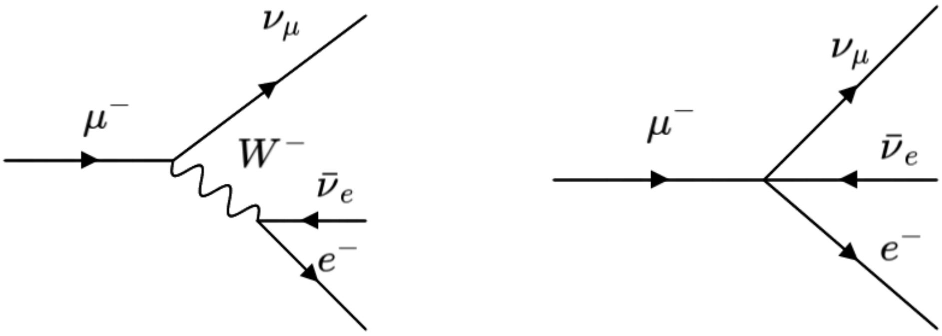


Figure 2.3: Muon decay in the Standard Model (left) and in Fermi's theory with fourfermion vertex (right). For low transferred momentum both theories provide the same result at the tree level.

Calculating the matrix element corresponding to the Standard Model diagram, the following equation can be obtained.

$$M_{fi} = \left[\frac{g_W}{\sqrt{2}}\bar{u}_{\nu_\mu}\gamma^\alpha\frac{1}{2}(1-\gamma^5)u_\mu\right]\frac{g_{\alpha\beta}-q_\alpha q_\beta/m_W^2}{q^2-m_W^2}\left[\frac{g_W}{\sqrt{2}}\bar{u}_e\gamma^\beta\frac{1}{2}(1-\gamma^5)v_{\nu_e}\right] \tag{2.29}$$

Replacing the W-boson propagator factor according to Equation 2.28, the following equation can be obtained.

$$M_{fi} = \frac{g_W^2}{8m_W^2}\left[\bar{u}_{\nu_\mu}\gamma^\alpha(1-\gamma^5)u_\mu\right]\left[\bar{u}_e\gamma_\alpha(1-\gamma^5)v_{\nu_e}\right] \tag{2.30}$$

Looking at the Equation 2.30, three terms can be recognized. The first one corresponds to the coupling constant of weak interaction in Fermi's theory. The second therm corresponds to the muon flux and the last term corresponds to the electron flux. The Fermi's constant for the weak interaction fourfermion vertex is defined as

$$G_F = \frac{g_W^2}{4\sqrt{2}m_W^2} \approx 1.166 \times 10^{-5}\mathrm{GeV}^{-2}\ [2]. \tag{2.31}$$

The value of the weak coupling constant for a low transferred momentum in the Standard model is $g_W \approx 1/40$.

2.2.7 Higgs boson and Higgs mechanism

The Higgs boson was the last discovered Standard Model particle. It was predicted in 1960s [33, 34], however the first experimental evidence was delivered by ATLAS and

CMS experiments in 2012 [35, 36]. The mass of the Higgs boson has been found to be 125.09 ± 0.24 GeV [2]. The Higgs boson was predicted as an important part of the Standard Model, since it enables non zero masses of W and Z bosons through Brout-Englert-Higgs mechanism. The idea of the mechanism is that if there is a scalar field Φ with potential

$$V(\Phi) = \mu^2\Phi^*\Phi + \lambda(\Phi^*\Phi)^2, \tag{2.32}$$

where $\mu^2 < 0$ and λ are both real numbers, the minimum of potential lies in a point with a non-zero value of Φ. The Φ value in the minimum of potential, so-called vacuum expectation value, is $v = e^{i\Theta}\sqrt{-\mu^2/\lambda}$, where the choice of the phase Θ is arbitrary and $\Theta = 0$ is often used.

Rewriting the Lagrangian replacing $\Phi \to v + \varphi$, a new physical Higgs field φ, with zero vacuum expectation value is obtained. The constant v is the vacuum expectation value of the Higgs field. Interaction terms of other particles with the original field Φ will split into two parts. The terms containing v, which are similar to mass terms in a Lagrangian, these terms are responsible for particle masses. The other term corresponds to the interaction with physical Higgs field φ. The "mass" term in Lagrangian and interaction term with Higgs field are described by the same constant for a fermion, coupling constant to the Higgs field, so-called Youkawa coupling constant. The equality of these coupling constants causes the Higgs boson to couple strongly to heavy particles. The heaviest particle with mass lower than $m_H/2$ is b-quark, what makes $H \to b\bar{b}$ decay channel dominant, with $\mathcal{BR}(H \to b\bar{b}) = 60.8$ %. It is followed by $\mathcal{BR}(H \to W^*W) = 23$ % and $\mathcal{BR}(H \to gg) = 7$ % [37]. Although the already mentioned decay branches have the highest branching ratio, they suffer from a large background rate, which makes them difficult to measure. The discovery of the Higgs boson has been performed in $H \to \gamma\gamma$ (via fermion, mostly top-quark, loop) and $H \to Z^*Z \to 4\ell$ channels with significantly lower branching ratios. The theory prediction for the Higgs boson decay width is $\Gamma_H = 42$ MeV [37].

2.2.8 Structure of proton

Protons and antiprotons are particularly important particles in high energy physics, since they are often used in particle colliders, such as the Large Hadron Collider. The proton is composite particle formed by quarks and gluons. The proton has three valence quarks: uud, however, especially at high energies also so-called sea quark-antiquark pairs can be found inside the proton. When calculating a cross section of a process in pp collisions, the contribution from all partons have to be taken into account, integrating over the whole range of x for a given type of the initial particle. The composition of the proton is described by parton distribution functions pdf(x). The pdf is a probability density to find a particle of the given type carrying x part of the proton's momentum. The proton pdf-s depend on the transferred momentum. The parton distribution functions for the proton at $\mu^2 = 10$ GeV^2 and (100 GeV)^2 scales can be found in Figure 2.4

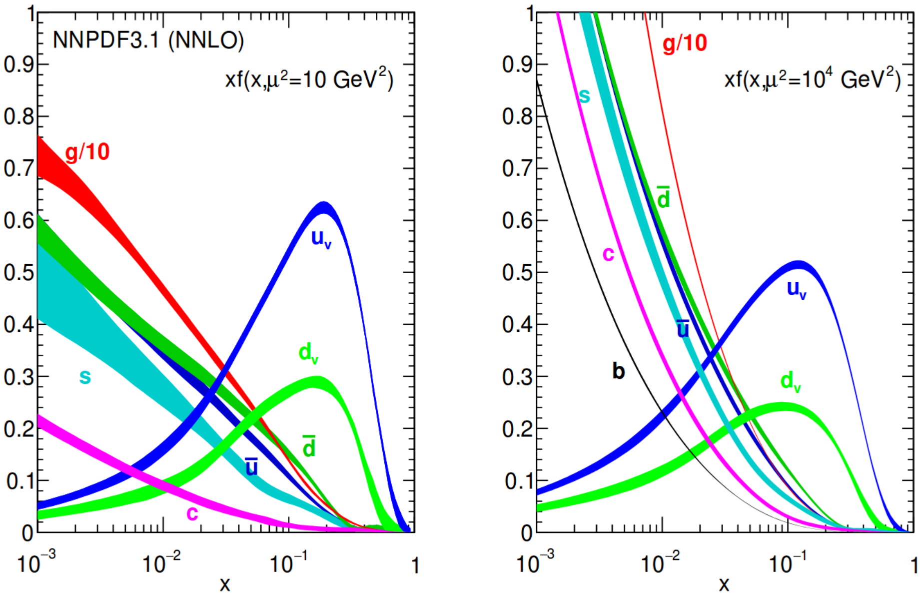


Figure 2.4: Parton distribution functions for the proton at 10 GeV (left) and 100 GeV (right) [38]. Except of valence quarks (u and d), there are gluons together with sea quarks and antiquarks. As the energy of the proton rises, the sea quarks and gluons carry higher part of the energy.

Factorization theorem

In the cross sections calculated from Feynman diagrams, the quarks and gluons (or possibly other elementary particles) play a role of the initial state particles. In order to calculate the cross section of a process in pp interaction, the factorization theorem is employed:

$$\sigma_{pp\to X} = \sum_{i,j} \int \int f_i(x_i, \mu_F) f_j(x_j, \mu_F) \sigma_{ij\to X}(x_i, x_j, \mu_R, \mu_F) dx_i dx_j, \tag{2.33}$$

where $\sigma_{pp\to X}$ is the cross section of the $pp \to X$ process (i.e. measured quantity), the sum runs over all types of partons in the proton, $f_i(x_i, \mu_F)$ is the parton distribution function of i-th type of parton at scale μ_F (so-called factorization scale) and $\sigma_{ij\to X}$ is the cross section of $ij \to X$ interaction, calculated from Feynman diagrams (where ij are partons) and μ_R is the renormalization scale. If all the terms of the perturbation series were summed, the dependency of the cross section on μ_R and μ_F scales would disappear. If only tree level diagrams are considered, the parton-parton cross section does not depend on μ_F.

2.3 Top quark

The top quark belongs to the 3$^{\text{rd}}$ generation of quarks. It was discovered at Tevatron (Fermilab) in 1995 by the experiments CDF[39] and D0[40].

The top quark is the heaviest known elementary particle with the mass of 173.21 ± 0.51 ± 0.71 GeV [2]. The electric charge of the top quark is +2/3 of the elementary charge. Spin of the top quark is 1/2. According to the Standard model, the top quark is weak isospin partner of the b-quark.

Top quarks can be produced as $t\bar{t}$ pair via the strong interaction, alone via the weak interaction (single-top production), or in associated production with other particles, such as Z boson.

The top quark is very unstable particle with the decay width $\Gamma = 1.322$ GeV [41] predicted by the Standard Model. The theoretical calculation of the decay width is performed at NNLO + NNLL resummation precision in QCD and NLO in EW, the top-quark mass of $m^{top} = 172.5$ GeV is considered. The mean lifetime of the top quark, which can be calculated from the decay width, is in the order of $\approx 10^{-25}$ s. Since the mean time of hadronization is $\approx 10^{-24}$ s, the top quark decays as a bare quark, before it could create a hadron. This makes the top quark a unique object, which allows us to study properties of a bare quark.

2.3.1 Top-quark pair production

Top-quark pair production is the process with the highest cross-section with the top quark in the final state. The top-quark pairs are produced via strong interaction, either in quark-antiquark annihilation or gluon-gluon fusion. The fraction of top-quark pairs produced in gluon-gluon fusion rises with energy. In pp collisions at $\sqrt{s} = 13$ TeV, the contribution from gluon-gluon fusion is ≈ 90 % and 10 % comes from quark-antiquark annihilation.[4] The cross-section of $pp \to t\bar{t}$ at $\sqrt{s} = 13$ TeV is 832^{+40}_{-46} pb according to the NNLO calculation with NNLL gluon resumations [42]. The dependency of the top-quark pair production cross section in pp and $p\bar{p}$ collisions on the center of mass energy, together with ATLAS, CMS and Tevatron measurements can be found in Figure 2.6. The observed cross sections are in a good agreement with the theory prediction.

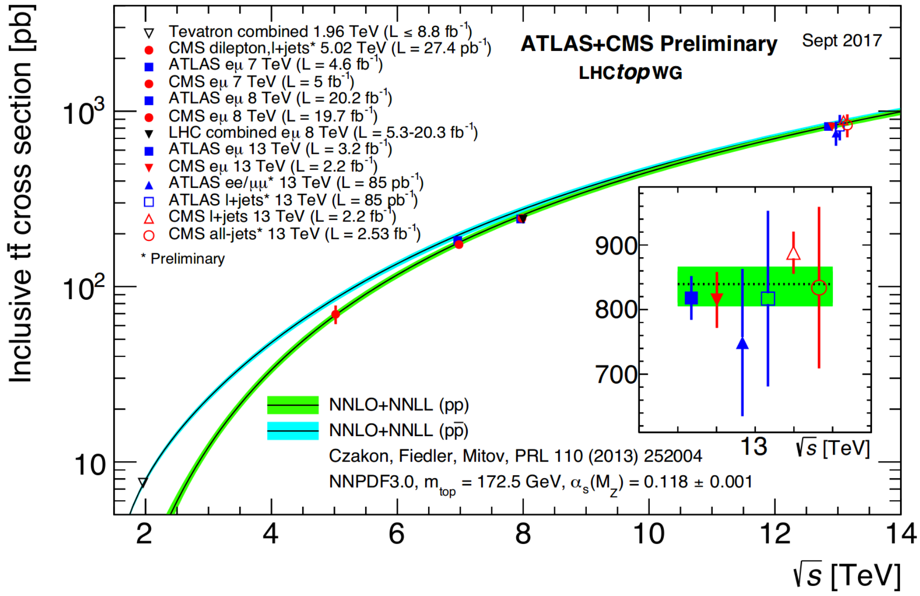


Figure 2.5: Theory prediction of $pp \to t\bar{t}$ cross section depending on the center of mass energy in pp and $p\bar{p}$ collisions, supported by ATLAS and CMS measurements from 7, 8 and 13 TeV and Tevatron combination at 1.96 TeV [43].

2.3.2 $t\bar{t}$+X production

According to the Standard Model, the cross section of $t\bar{t}$+X (X = Z, W, H, γ) depends on the top-quark coupling constant to the X particle. Thus the measurements of the $t\bar{t}$+X cross sections provide an opportunity to verify or measure the values of these coupling constants. The measurement of the $t\bar{t}H$ cross section is particularly important for a precision Standard Model testing, since the coupling constant of the top quark to the Higgs boson, top-quark Yukawa's coupling constant, is one of the Standard Model free parameters. The cross section of $t\bar{t}\gamma$ depends on the photon coupling to the top quark, which is related to its electric charge. The cross sections obtained from theory calculations for proton-proton collisions at $\sqrt{s} = 8$ TeV and $\sqrt{s} = 13$ TeV are summarized in Table 2.2. In the Standard Model, the Z boson is coupled to the top

---

[4] Calculated from Powheg Monte Carlo samples.

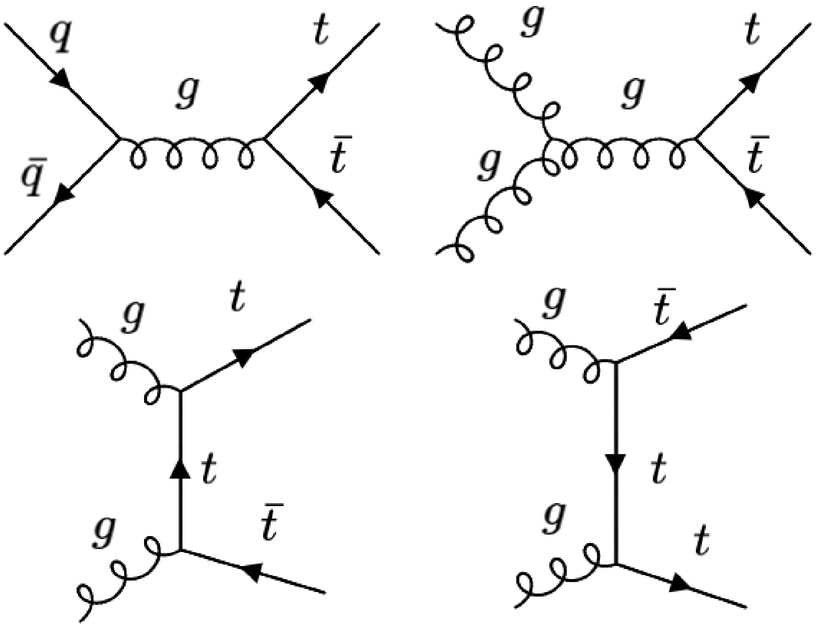


Figure 2.6: Feynman diagrams for top-quark pair production. The quark-antiquark annihilation (upper left) and gluon-gluon fusion (other).

quark via electroweak interactions and it is further discussed in Section 2.5

	8 TeV	13 TeV
$t\bar{t}$	252^{+13}_{-13} pb (NNLO+NNLL)	832^{+40}_{-46} pb (NNLO+NNLL)
$t\bar{t}Z$	215 ± 30 fb (NLO)	863^{+78}_{-90} fb (NLO+NNLL)
$t\bar{t}W$	232 ± 32 fb (NLO)	601^{+80}_{-72} fb (NLO)
$t\bar{t}H$	129^{+11}_{-16} fb (NLO)	499^{+29}_{-46} fb (NLO)
$t\bar{t}\gamma$	151 ± 24 fb (NLO)	521 ± 42 fb (NLO)

Table 2.2: Cross sections for top-quark production in pp collisions at next-to-leading order (NLO) or next-to-next-to leading order (NNLO), or next-to-text leading logarithm precision [44, 42, 45, 1, 46]. Since the photon is massless and thus the total cross section calculation would suffer from collinear (infra-red) radiation terms, only a fiducial cross-section has been calculated for this process. The fiducial volume is defined as top-quark pair decaying into ℓ+jets and $p_{\mathrm{T}}^{\gamma} > 15$ GeV [47, 48]

.

2.3.3 Top-quark decay

The top quark decays with probability of 99.8% into the b-quark and W boson. In the other 0.2% of cases, the top quark decays into d or s quark and W boson. Those top-quark decays are not considered in the analysis. They are difficult to identify in the measured data because of a large background.

W boson from the top-quark decays hadronically ($\mathcal{BR}(t \to bq\bar{q}') = 67.41 \pm 0.27$ % [2]), into two quarks, or leptonically into charged lepton and its anti-neutrino ($\mathcal{BR}(t \to be^-\nu_e) = 10.71 \pm 0.16$ %, $\mathcal{BR}(t \to b\mu^-\nu_\mu) = 10.63 \pm 0.15$ %, $\mathcal{BR}(t \to b\tau^-\nu_\tau) = 11.38 \pm 0.21$ % [2]). In the final state from the top-quark decay, there is one b-jet and two light jets coming from decay of the W boson (hadronic decay), or one b-jet, a charged lepton and its neutrino coming from the W-boson decay (leptonic decay).

The τ leptons are unstable particles with very short mean lifetime $\tau_\tau = 2.9 \times 10^{-13}$ s [2]. Because of the short lifetime, they cannot be measured directly. Various decay modes of τ lepton are allowed, having leptons or hadrons in the final state. The reconstruction of hadronic decay is a challenging problem and these decays are not studied in this analysis. However, the leptonic decays of τ leptons are important, since there is exactly one charged lepton in the final state, which is similar to leptonic decay of the top quark. Although distributions of the lepton fourmomentum and missing transverse energy are different from the direct leptonic decay, it is difficult to recognize the direct top-quark decay into electron or muon from the cascade decay through the τ lepton. The branching ratios of $\tau \to \mu\bar{\nu}_\mu\nu_\tau$ and $\tau \to e\bar{\nu}_e\nu_\tau$ are 17.39 % and 17.82 % [2].

2.3.4 Top-quark pair decay

Combining decay modes of two top quarks, three final states are allowed:

All-hadronic decay with both W bosons decaying into jets. There are two b-jets and four light jets[5] in the final state. The branching ratio of this decay channel is 45.4% [2].

Semi-leptonic decay, also called lepton+jets, one W boson decays into lepton+neutrino and the other one into two quarks. In this case, there are two b-jets, two light jets, charged lepton and missing transverse momentum (neutrino) in the final state. If τ leptons which are difficult to identify are not considered, the branching ratio of the channel is 28.8%. If the leptonic decays of τ leptons are considered, which is the

---

[5] The light jet is the jet initiated by light quark, it means u, d, s or c quark

case of our analysis, the branching ratio is 34.2 %. If all decay channels of τ lepton are considered, the branching is 44.1% [2].

Dileptonic decay, both W bosons decay into lepton and neutrino, so in the final state there are two b-jets, two charged leptons of the opposite sign and missing transverse momentum (two neutrinos). If the τ leptons are not considered, the branching ratio of the channel is 4.6%. If also leptonic decays of τ leptons are considered, the branching ratio of the channels is 6.4 %. If all decay channels of τ lepton are considered, the branching is 10.7% [2].

In general, higher number of leptons in the final state allows better background suppression at the cost of smaller dataset due to decreasing branching ratio. In addition to the small statistics, the dilepton decay of the top-quark pair is also difficult to reconstruct, since there are two neutrinos in the final state, or even more, if at least one lepton originates from a τ-lepton decay. There is no exact solution for the two neutrino problem, however there are some probabilistic-approach algorithms how to reconstruct momenta of the neutrinos, such as Neutrino Weighting Method [49].

2.4 Z boson

Z boson is heavy vector boson and together with the W boson they are the mediators of the weak interactions. Mass of the Z boson is 91.1876 ± 0.0021 GeV [2] and it carries no electric charge. The Z boson is unstable particle with the decay width $\Gamma = 2.4952 \pm 0.0023$ GeV, which corresponds to the mean lifetime of $\approx 2\times10^{-25}$ s [2]. The Z boson couples to all Standard Model particles except of the photon and gluon. There are many decay channels of the Z-boson, the most important (with the highest probability) are the following:

Hadronic decay: Z boson decays into quark-antiquark pair, so in the final state there are two jets from these quarks. The probability of this decay is 69.91 ± 0.06 % [2].

Invisible decay: Z boson decays into neutrino and antineutrino. This decay is very difficult to identify, because usually the only sign is a missing transverse momentum, that appears also in other events and in addition, the detector resolution in missing transverse momentum is worse compared to the momentum resolution for jets and leptons. The probability of the invisible Z-boson decay is 20.00 ± 0.06 % [2].

Leptonic decay: Z boson decays into two charged leptons. In the final state there are two charged leptons of the same flavour, opposite sign and their invariant mass is close to the mass of the Z-boson. In this case, it is important to distinguish between e^+e^-, $\mu^+\mu^-$ and $\tau^+\tau^-$ case. τ leptons are very unstable and decay before reaching the detector, so their reconstruction is challenging problem. Electrons and muons are well reconstructed by the ATLAS detector. Probability of the Z-boson decay into two charged leptons is 10.099 ± 0.011 % [2]. Subtracting the contribution of $Z \to \tau\tau$, 6.729 $\pm$ 0.008 % probability of having e^+e^- or $\mu^+\mu^-$ pair in the final state can be obtained. The $\tau \to \ell + \bar{\nu}_\ell + \nu_\tau$ decays are not considered here, since the neutrinos carry away a part of the τ-lepton fourmomentum and thus the invariant mass of the lepton pair is lower than the Z-boson mass.

Other decay modes are allowed only at next-to-leading order and their probability is far below 1 % [2].

2.4.1 Z+jets production

The most frequent process with the Z boson in the final state, and the most significant background in dilepton channel of $t\bar{t}Z$ production, is the associated production of the Z boson and jets. Leading order Feynman diagrams are shown in Figure 2.7. The total cross section of the Z+jets in pp collisions at 13 TeV is 67^{+4}_{-8} nb at NLO precision, when up to 3 additional partons with $p_\mathrm{T} > 30$ GeV in pseudorapidity range $|\eta| < 4$ are considered [50]. The contribution from events with at least one b-jet (jet initiated by a b-hadron) is ≈ 7 %. The Z+jets production is dominated by quark anti-quark annihilation, which contributes by 76 %, it is followed by quark-gluon interaction (23 %) and gluon-gluon fusion (1 %). [6]

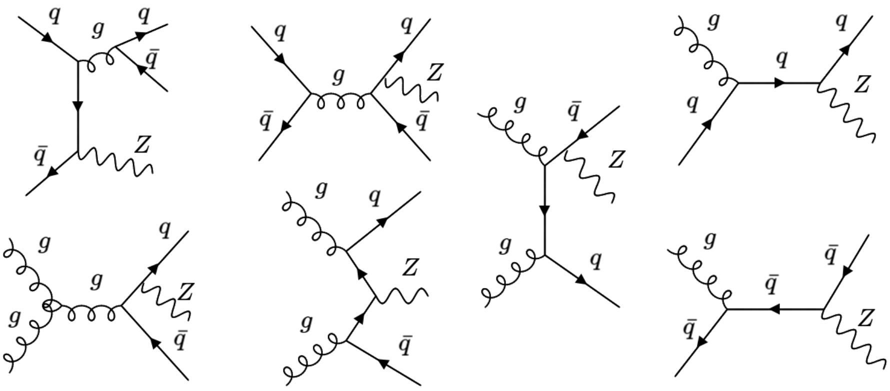


Figure 2.7: Leading order Feynman diagrams for the associated production of the Z boson with jets.

---

[6] Calculated from Z+jets Sherpa 2.2.1 samples.

2.5 $t\bar{t}Z$ associated production

The associated production of the top-quark pair and Z boson is a rare process predicted by the Standard model. The theory prediction for the $t\bar{t}Z$ cross section for proton-proton collisions at $\sqrt{s} = 13$ TeV at NLO+NNLL in QCD and NLO in EW precision is

$$\sigma_{t\bar{t}Z} = 0.863\ ^{+8.5\%}_{-9.9\%}\ (\text{scale})\ ^{+3.2\%}_{-3.2\%}\ (\text{PDF} + \alpha_S)\ \text{pb}\ [1],$$

which is approximately 1,000 times less than the cross section of the top-quark pair production.

The leading order Feynman diagrams of the $t\bar{t}Z$ production are shown in Figure 2.8. The contribution from quark-antiquark annihilation is 30 % and gluon-gluon fusion contributes by 70 % at $\sqrt{s} = 13$ TeV.[7]

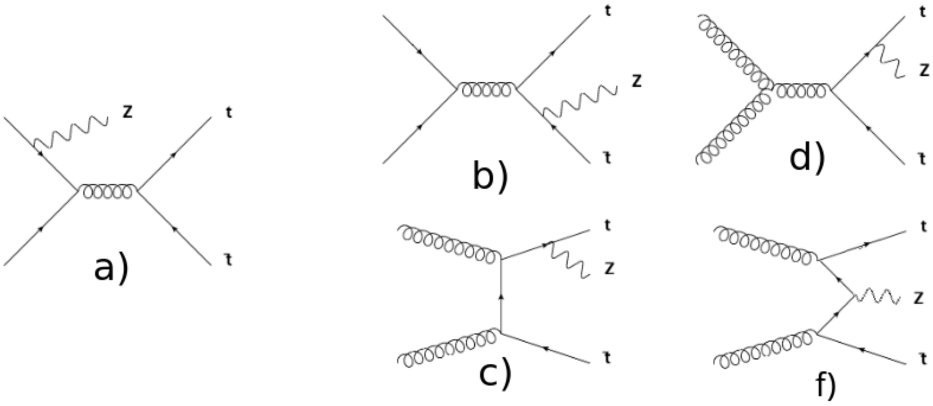


Figure 2.8: The Feynman diagrams for the $t\bar{t}Z$ production. The $t\bar{t}Z$ can be produced either by quark-antiquark annihilation (a-b), or by gluon-gluon fusion (c-f) . Diagram a) shows initial state radiation, the other diagrams final state radiation, which is sensitive to the coupling constant of Z boson and top quark.

According to the Standard Model, the Z boson couples to quarks and leptons. The vertex of fermion and Z boson is shown in Figure 2.9 and according to the Standard Model, the coupling of Z boson with fermion pair is:

$$\frac{-ig_Z}{2}\gamma^{\mu}(c_v^f - c_A^f\gamma^5), \tag{2.34}$$

where $g_Z = \frac{e^2}{4\pi sin(\vartheta_W)cos(\vartheta_W)}$, values c_v^f and c_A^f depend on the particle weak isospin and the weak interaction mixing angle and are summarized in Table 2.3

---

[7] Calculated from MC@NLO $t\bar{t}Z$ samples.

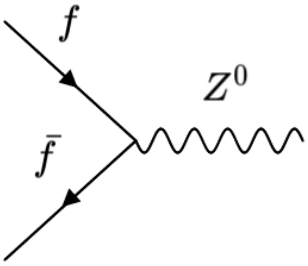


Figure 2.9: Vertex for the interaction between fermion and Z boson. The fermion can be any quark or lepton.

particle	c_v^f	c_A^f
ν_e, ν_μ, ν_τ	$\frac{1}{2}$	$\frac{1}{2}$
e, μ, τ	$-\frac{1}{2} + 2\sin(\vartheta_W)$	$-\frac{1}{2}$
u, c, t	$\frac{1}{2} - \frac{4}{3}\sin^2(\vartheta_W)$	$\frac{1}{2}$
d, s, b	$-\frac{1}{2} - \frac{2}{3}\sin^2(\vartheta_W)$	$-\frac{1}{2}$

Table 2.3: Values of c_v^f and c_A^f for different types of fermions, ϑ_W is the weak mixing angle, the value measured in the experiment is 28.7° [30]

If the Z boson in the $t\bar{t}Z$ production originates from the final state radiation (Feynman diagrams b-f in Figure 2.8), the cross section depends quadratically on the Term 2.34. This allows us to probe the coupling constant of Z boson to top quark, if the contribution of initial state radiation in $t\bar{t}Z$ production is known.

2.6 Decay channels of $t\bar{t}Z$

As mentioned before, the top quark and Z boson are very unstable particles, so they cannot be measured directly. Depending on the decay modes of individual particles various final states are allowed. Invisible decay of the Z boson is very difficult to identify and also the identification of hadronicaly decaying Z boson is challenging problem, because the momentum and energy resolution for jets is worse than that of leptons. The studies focused on the channels with Z decaying into neutrinos showed very low signal sensitivity. Currently only channels with Z boson decaying into e^+e^- or $\mu^+\mu^-$ lepton pairs are considered in the analysis. The $\tau^+\tau^-$ decays are difficult to reconstruct and this analysis does not target these decay channels.

2.6.1 Dilepton opposite-sign same-flavour channel

In this case the top-quark pair decays hadronically and Z boson decays into e^+e^- or $\mu^+\mu^-$ pair. This decay channel has the highest branching ratio, but the background rate is relatively high. Dominant background processes are Z+jets and dilepton decay of $t\bar{t}$ with additional jets. Considering the branching ratios of the $t\bar{t}$-pair and Z boson, the branching ratio of the decay channel is 4.9% considering the τ leptons. If the leptonic decays of τ leptons from $t\bar{t}$ are considered, but $Z \to \tau\tau$ are not, the branching ratio of the channel is 3.1%.

In the final state there are:

- 2 leptons of opposite sign and same flavour with invariant mass close to the mass of Z boson (so-called Z-like pair)
- 2 b-jets
- 4 additional light jets
- no neutrinos

2.6.2 Trilepton channel

The top-quark pair decays semileptonically and the Z boson decays into e^+e^- or $\mu^+\mu^-$ pair. This channel is characterized by reasonably high statistics while the background is relatively low. The dominant background are fake leptons, dibosons (mostly $WZ \to \ell\ell\ell\nu$), tZ and tWZ [51]. If the leptonic decays of τ leptons from $t\bar{t}$ are considered, but $Z \to \tau\tau$ are not, the branching ratio of the channel is 2.3 %.

In the final state there are:

- 3 leptons, (one Z-like pair, one lepton from $t\bar{t}$ decay)
- 2 b-jets
- 2 additional light jets
- 1 neutrino

2.6.3 Tetralepton channel

Top-quark pair decays dileptonically and Z boson decays into e^+e^- or $\mu^+\mu^-$ pair. This channel has a very low background rate, but the branching ratio of this channel is low, so the statistics is very limited. The dominant backgrounds are fake leptons, tWZ and $ZZ \to \ell\ell\ell\ell$. If the leptonic decays of τ leptons from $t\bar{t}$ are considered, but $Z \to \tau\tau$ are not, the branching ratio of the channel is 0.43 %.

In the final state there are:

- 4 leptons (one Z-like pair, one pair of opposite-sign leptons from $t\bar{t}$)
- 2 b-jets
- 2 neutrinos

Chapter 3

The LHC complex and ATLAS experiment

3.1 Large Hadron Collider

The LHC is currently the largest and the most important accelerator for the particle physics in the world. Its circumference is about 27 km and it is 50-175 meters under the Earth surface. It is located on the border of Switzerland and France near Geneva. It was built by the European Organisation for Nuclear Research (CERN). It has replaced the LEP (Large electron-positron collider), which was located in the tunnel before. LEP worked with electron and positron beams and was able to reach the center of mass energy of about 100 GeV in the first phase and 200 GeV in the second phase. It was functional from 1989 to 2000.

The first proton beam at the LHC was accelerated in September 2008. Few days later there was an accident, when large amount of liquid helium escaped and damaged superconducting magnets, their mountings and contaminated the vacuum tube. Due to this accident, the LHC was not able to work during the next year. The first collisions of the two beams were reached in 2010. On 13th March 2011 the LHC started to accelerate the proton beams with the energy of 3.5 TeV per a particle. After a short shutdown, in April 2012 the proton beams were accelerated again, but to 4 TeV per particle. In 2011 and 2012 the LHC was also used to collide the lead ions accelerated to 287 TeV. In 2013 and 2014 there was a next shutdown due to upgrades of the accelerator, to enable to accelerate the protons up to 6.5 TeV. In 2015-2018 the LHC was operational again, accelerating protons to energy 6.5 TeV per particle [52].

There are few pre-accelerators at the LHC. Before the protons enter the LHC, they must be accelerated to 450 GeV by the pre-accelerators. The accelerator complex is showed in Figure 3.1.

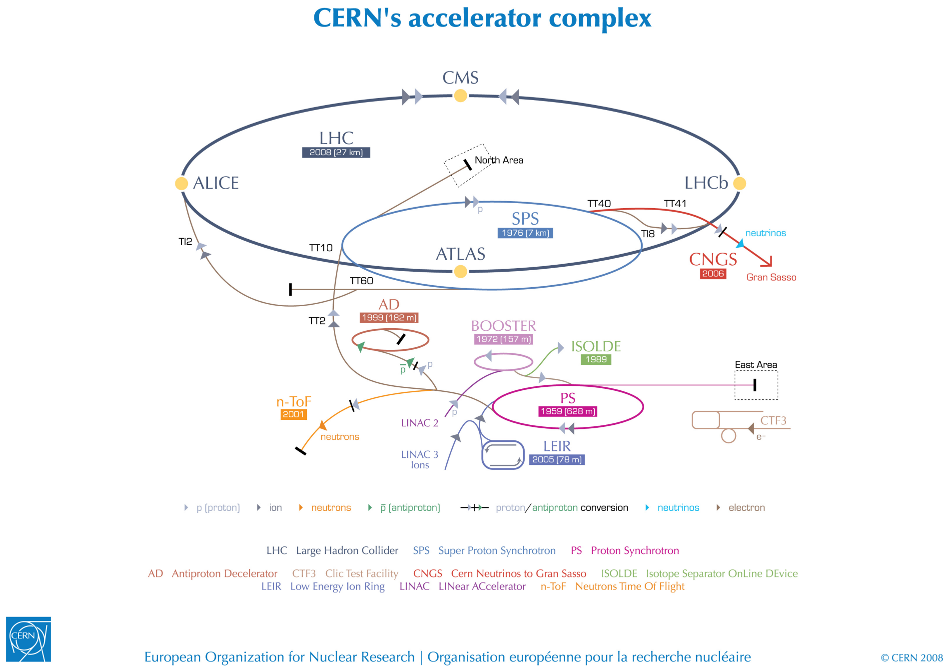


Figure 3.1: LHC and the preaccelerators [53].

The first part of the accelerating complex is a proton source, where the protons are produced by ionizing of hydrogen. The protons from the source continue to the linear accelerator Linac 2 (marked as ”p” in the picture). In the case of heavy ions (Pb) Linac 3 is used (marked as ”Pb” in the picture), which is specially constructed to accelerate heavy ions. The collisions of heavy ions are not important for this analysis, so only the accelerating of the protons is described here. After leaving the Linac, the protons have energy of 50 MeV. This process takes few microseconds [54].

From Linac 2 the proton beam continues to PSB (Proton Synchrotron Booster). It is the synchrotron with the radius of 25 meters, where the protons are accelerated to 1.4 GeV. This takes 530 ms [55].

The protons continue to Proton Synchrotron (marked as ”PS” in the picture). Its radius is 100 meters. It was built in 1959 and was used as a standalone accelerator, a long time before building of LHC. Now it accelerates the protons to the energy of 25 GeV. This takes around 1 second [56].

The next preaccelerator is SPS (Super Proton Synchrotron). The protons stay there for 10.8 s, 7.2 s or 3.6 s for the next accelerating, or they are accelerated immediately. The SPS is the cyclic accelerator with a circumference of 6.9 km. It is the last preaccelerator before the protons enter the LHC. The protons are accelerated to 450 GeV in SPS. The accelerating process in the SPS takes 4.3 seconds. SPS was used in 1981-1984 as a standalone proton-antiproton collider during experiments which led to the discovery of the W and Z bosons [57].

After accelerating the protons in SPS, the proton beam is divided into two parts, which are injected to LHC in opposite directions. There are two parallel tubes in LHC, each accelerating one proton beam. These two tubes collide only in 4 places in LHC, where detectors are placed. 1232 dipole electromagnets create a strong magnetic field which makes the protons to stay in the cyclic tube of LHC. 392 quadrupole magnets focus the proton beams to prevent the collisions with a surface of the tube. To reach the magnetic field strong enough, superconducting magnets are used that need to be cooled down by a liquid helium to 1.8 K [52].

There are 4 detectors at LHC: ATLAS, CMS, LHCb and ALICE. In this thesis, only data and Monte Carlo samples from the ATLAS detector are used, therefore it will described in more details.

Luminosity

The luminosity is the important parameter of particle collider. The luminosity expresses the number of events of the selected type that will be observed. We are able to express the number of the events of the selected type (for example a top-quark pair production) via the formula:

$$\frac{dN}{dt} = L \cdot \ \sigma, \tag{3.1}$$

where N is the number of events, σ is cross-section and L is a parameter of the collider called luminosity. Luminosity depends on the number of particles in a bunch, size of the bunch and a frequency of bunch crossing. It is obvious that the number of events in a selected time interval will be

$$N = \int L \cdot \sigma dt = \sigma \cdot \int L dt = \sigma \cdot \mathcal{L}, \tag{3.2}$$

$\mathcal{L}$ is called integrated luminosity. Its unit is an inverse barn (b^{-1}).

The total integrated luminosity collected at $\sqrt{\mathrm{s}} = 7$ TeV was $\mathcal{L} = 5.5\ \mathrm{fb}^{-1}$ and at $\sqrt{\mathrm{s}} = 8$ TeV it is $\mathcal{L} = 23\ \mathrm{fb}^{-1}$. During years 2015 - 2018 the ATLAS recorded 139 fb^{-1} of data from proton-proton (pp) collisions at $\sqrt{s} = 13$ TeV. Only the data from 13 TeV are used in this thesis.

3.2 A Toroidal LHC Apparatus (ATLAS)

The weight of the detector is about 7000 tons. It has cylindrical shape with the length of 44 meters and 25 meters in the diameter. It was assembled and used first time in 2008. One of its most important results, together wih CMS, was the discovery of Higgs boson in July 2012 [35, 36]. Construction of the ATLAS detector is shown in the Figure 3.2.

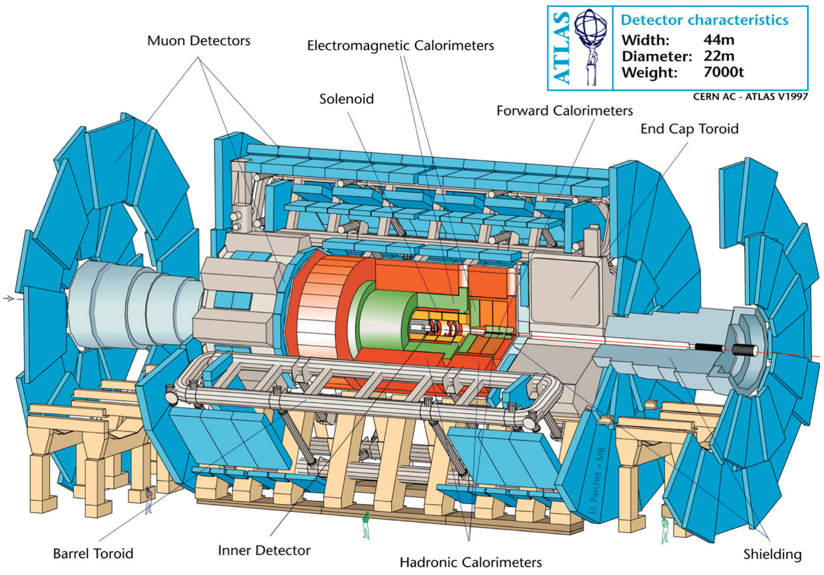


Figure 3.2: The ATLAS detector [58]

The ATLAS detector consists of 4 main parts: an inner detector, calorimeters, muon spectrometer and magnetic systems. Each of these parts consists of several layers. ATLAS is designed to detect all particles created in an event excluding neutrinos.

3.2.1 ATLAS coordinate system

The centre of the ATLAS coordinate system is located in the centre of ATLAS, in the interaction point, where the two proton beams collide. The coordinate system is right-handed, with the x-axis pointing to the middle of LHC, the y-axis pointing upwards and the z-axis pointing in the direction of the beam. The spherical coordinates are usually used in the analyses performed by ATLAS. The angle ϕ is angle from x-axis in xy plane. The ϑ is the angle from z-axis. Since the difference in ϑ of two particles is not Lorentz invariant with respect to boosts in direction of z-axis (direction of the colliding protons), the pseudorapidity η is used. The η is defined as follows

$$\eta = -\ln\left(\tan\frac{\vartheta}{2}\right) \tag{3.3}$$

Since the p_Z of particles depends on the p_Z of initial partons, and usually it is not zero, so-called transverse momentum, p_{T} is often used. The transverse momentum is defined as:

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2} = p.\sin\vartheta. \tag{3.4}$$

3.2.2 The Inner Detector

Inner detector is closest to the beampipe. It is cylindrically shaped with the radius of 2.1 m and length of 6.2 meters. The Inner Detector is able to detect particles in a pseudorapidy range $|\eta| < 2.5$. There is a magnetic field of about 2 Tesla in this area, which causes the trajectory of the charged particles to curve. The main purpose of the inner detector is a precise reconstruction of particle trajectories. From radius and direction of the curvature of the trajectory, the momentum and charge of the particle can be determined [59].

The inner detector consists of the three parts: Pixel Detector, Semiconductor tracker and Transition radiation tracker.

Pixel Detector

It is the first part of the inner detector and it is closest to the beampipe. It has high granularity and precise spatial resolution. The size of one pixel is 50 × 400 μm. It is as close to the interaction point as possible and in Run-I (7 and 8 TeV) it consisted of three barrels with the radii of 5 cm, 9 cm and 12 cm (1456 modules) and three discs on each side between the radii of 9 cm and 15 cm (288 modules). The total number of read out channels is 80 million. In Run-II, one more pixel layer was added. The name of the layer is IBL and its main purpose is to improve b-tagging of the jets, which is also one of the main purposes of the pixel detector. The IBL has additional 12 million readout channels and its distance from the interaction point is 3.3 cm. The size of the IBL pixels is 50 × 250 μm. Precision spatial measurement close to the primary vertex enable us to reconstruct secondary vertices from decay of shortly living b-hadrons. The thickness of each pixel layer is about 2.5 % of radiation length [59, 60].

Semiconductor Tracker (SCT)

SCT consists of eight layers of silicon microstrip detectors, so in the ideal case it provides eight precise measurements per a track, which enables the measurement of the momentum, impact parameter and vertex position with a high precision. The barrel modules are mounted at the radii of 29.9, 37.1, 44.3 and 51.6 cm. SCT is divided into two parts: barrel, which covers the area $|\eta| < 1.4$ and end-cap, which covers $1.4 < |\eta| < 2.5$ [61].

Transition Radiation Tracker (TRT)

It consists of tubes with 4 mm in the diameter filled by a mixture of Xe, CO_2 and CF_4 (70 % : 27 % : 3 %). A gold-plated W-Re wire of 30 μm diameter is placed in the centre of each tube. The TRT is designed to reconstruct charged particles with $|\eta| < 2.0$ and $p_{\mathrm{T}} > 500$ MeV. The barrel part of TRT covers pseudorapidity range of $|\eta| < 1.0$ and consists of 52,544 straws oriented parallel to the beam axis. The end-cap parts consist of 122,880 straws each and cover range $1.0 < |\eta| < 2.0$ [62]. The TRT provides good discrimination between electrons and charged hadrons in wide p_{T} range from 0.5 to 100 GeV, measuring X-rays produced by transition radiation of the particles [63].

3.2.3 Calorimeter

ATLAS has two calorimeters: an electromagnetic (inner) and hadronic (outer) calorimeter. The main purpose of the calorimeters is the measurement of the energy of the particles by absorbing them. Both calorimeters are sampling ones, that means, there is a passive and active medium in both calorimeters. High-density metal plates are used as the passive medium. Liquid argon and scintillating plastic are used as the active medium.

Electromagnetic (EM) calorimeter

The EM calorimeter is designed to detect the particles interacting via an electromagnetic interaction - photons and charged particles, predominantly electrons and positrons. Lead plates are used as the absorber and as the active medium, liquid argon is used. The EM calorimeter consists of two parts. The first one is barrel which covers $|\eta| < 1.475$ and the second one is the end-cap part covering $1.375 < |\eta| < 3.2$. The total thickness of the EM calorimeter is 24 radiation lengths in the barrel part and 26 radiation lengths in the end-cap parts. An angular segmentation of the EM calorimeter is $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$. The energy resolution of the EM calorimeter is given by the formula [64, 65, 66]

$$\frac{\Delta \mathrm{E}}{E} = \frac{11.5\%}{\sqrt{E[\mathrm{GeV}]}} \oplus 0.5\%. \tag{3.5}$$

Hadronic calorimeter

The hadronic calorimeter is used to detect particles which pass through EM calorimeter and interact via a strong interaction. It is crucial to measure the energy of jets[1]. The barrel part covers region $|\eta| < 1.6$, iron is used as an absorber and scintillating tiles are used as an active part. The end-cap parts use copper as the absorber and liquid argon as the active medium due to a higher radiation dose. The thickness of the hadronic calorimeter is 10 hadron interaction lengths.

The resolution in the energy and the angle is lower than in the EM calorimeter. The segmentation is $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ and the average resolution in the jet energy is given by Equations 3.6 and 3.7 [64]:

$$\frac{\Delta \mathrm{E}}{E} = \frac{50\%}{\sqrt{E[\mathrm{GeV}]}} \oplus 3\% \qquad \text{for} \qquad |\eta| < 3 \tag{3.6}$$

$$\frac{\Delta \mathrm{E_T}}{E_\mathrm{T}} = \frac{100\%}{\sqrt{E[\mathrm{GeV}]}} \oplus 10\% \qquad \text{for} \qquad 3 < |\eta| < 5 \tag{3.7}$$

3.2.4 Muon Spectrometer

The muon is the only particle (apart from a neutrino) which is not absorbed in the calorimeters. The muon spectrometer covers the pseudorapidity range $|\eta| < 2.7$ and it is designed to detect the muons and reconstruct them independently from the inner detector. The muon spectrometer is crucial for reconstruction of the muons with high p_T, which are usually sign of an interesting physics. There is a magnetic field of 0.5 T in the Muon Spectrometer, which enables measurement of the moun momentum. Figure 3.3 shows the momentum resolution of the Muon Spectrometer, Inner Detector and their combination.

The Muon Spectrometer consists of four parts: Monitored Drift Tubes (MDT), Cathode Strip Champbers (CSC), Thin Gap Chambers (TGC) and Resistive Plate Chambers (RPC) [67].

---

[1] The jets are showers of the particles, which comes from a quark, hard gloun or tau lepton.

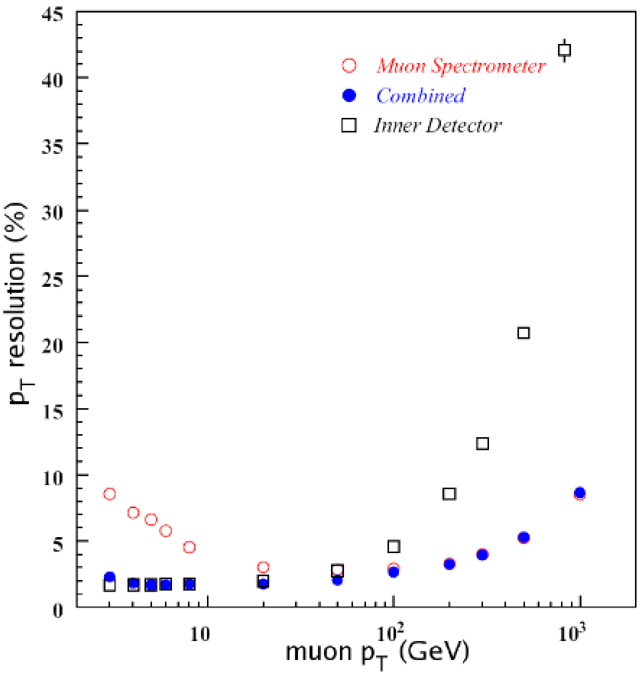


Figure 3.3: Momentum resolution of Muon Spectrometer and Inner Detector [66].

The Monitored Drift Tubes

The Monitored Drift Tubes (MDTs) consist of aluminium tubes with 30 mm in the diameter and with a wire in the centre. The MDTs cover the barrel part and end-cap parts. In the barrel part ($|\eta| < 1.05$), 3 layers of the tubes 5, 7.5 and 10 m from the beam axis are used. In the end-cap parts ($1.05 < |\eta| < 2.7$), there are 3 wheels in the z-axis distance 7.5, 13 and 20 m. The tubes are filled with a mixture of Ar, CO_2 (93 %, 7 %) and they work as drift chambers with maximal drift time of 500 ns. The muon crossing the tube causes a ionization process in the gas and measuring the electron cluster drift time enables to reach the spatial resolution of around 50 μm [66, 68, 67].

The Cathode Strip Chambers

CSCs are used in the end-cap region $2 < |\eta| < 2.7$, they have high spatial resolution and low drift time of electrons (30 ns), that is why they can be used in environment with a high muon flux. The CSC are multi-wire proportional chambers with a segmented cathode. The spatial resolution is around 60 μm [66, 67].

The Resistive Plate Chambers

RPCs are gaseous detectors with worse spatial resolution (about 10 mm) compared to the other parts of Muon Spectrometer, but they are able to work very fast (1.5 ns), so they are used as the first level trigger. RPCs cover the barrel part of the detector, $|\eta| < 1.05$, and they are divided into 3 stations, each with 2 layers located on each side of MDT layers [66, 68, 67, 69].

The Thin Gap Chambers

TGCs are multi-wire proportional chambers providing a triggering and second coordinate measurement in the end-cap regions. The typical spatial resolution is about 2-7 mm and the drift time less than 25 ns. The TGCs cover the range $1.05 < |\eta| < 2.7$ and they are used as the first level trigger in the range of $1.05 < |\eta| < 2.4$ [66, 67, 69].

3.2.5 Magnetic System

The magnetic system at ATLAS provides almost homogeneous magnetic fields which enables the measurement of the momentum and charge of the particles. The magnetic system is divided into 3 parts: the central solenoid, the barrel toroid and the end-cap toroids.

The Central Solenoid

It is located around the inner detector and provides a homogeneous magnetic field of around 2 T, which enables to measure the momentum of the particles in the inner detector. The track of charged particles with the low transverse momentum cannot be measured, because their trajectories are curved too strongly. With rising transverse momentum of the particles detection efficiency is rising and reaches almost 100% at 400 MeV in p_{T}. Due to the reduction of energy losses of the particles that leave the inner detector, thickness of solenoid and cooling system is 0.83 of radiation length at normal incidence. The Central Solenoid shares its cooling system with the electromagnetic calorimeter. The Central solenoid has the radius of 1.22 m and the length of 5.3 m [70, 71].

The Barrel Toroid

It consists of 8 coils with air cores, each is 25 m long and 5 m wide, covering range $|\eta| < 1.3$. They create the magnetic field varying from 2 T to 8 T in the area 26 meters long and 20 meters in the diameter. An electromagnet able to create the homogeneous magnetic field strong enough in such a huge area would be unaffordably expensive. The dependency of magnetic field intensity inside the Barrel Toroid on η is shown in Figure 3.4 [71].

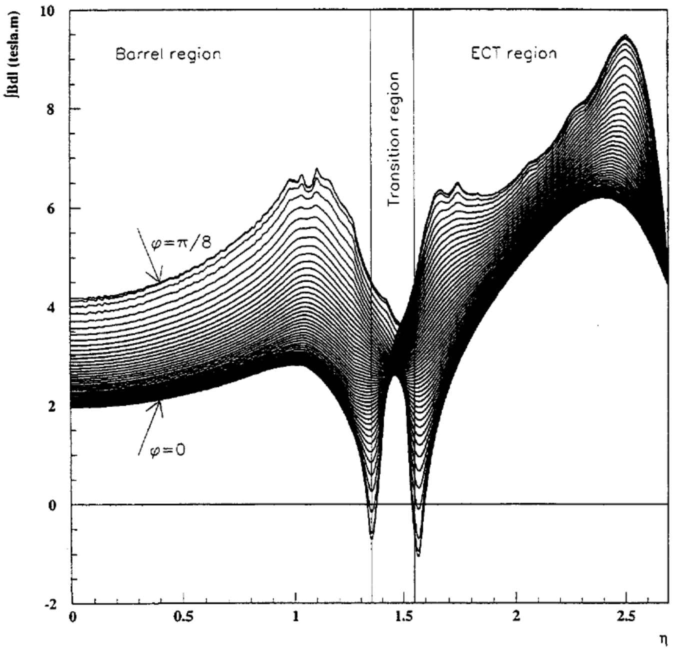


Figure 3.4: Dependency of $\int \vec{B}d\vec{l}$ on η inside the Barrel Toroid and End-Cap Toroids [71].

The End-Cap Toroids

The End-Cap Toroids cover the area $1.6 < |\eta| < 2.7$ and consist of 8 coils. Their length is 5 m and extend radially from 1.65 to 10.7 m [71].

3.3 ATLAS trigger system

Bunch crossing rate at ATLAS is about 40 MHz and if every event was recorded it would be around 1 PB/s of data. The main part of these events are not physically interesting and also it is not possible to store so much events, therefore a technique to decide which events should be recorded is needed. ATLAS trigger system fulfils this function and it has two parts.

3.3.1 Level 1 trigger (L1)

It is designed to reduce the event rate from 40 MHz to $\approx$ 100 kHz. It is a hardware trigger and it is based on a signal from the calorimeters and muon spectrometer. The signals are processed in the Central Trigger Processor with the latency of 2.5 μs. If there is high missing transverse energy (E_{T}), muons, electrons, photons, or jets with high p_{T}, the event is tagged as potentially interesting and continues to High Level Trigger [72, 73].

3.3.2 High Level Trigger (HLT)

It is an off-line software trigger. In its output there is around 1 kHz of events. Compared to the Level 1 trigger, it uses the complete information about the event including the information from the inner detector. The HLT runs at the computer farm with $\approx$40,000 processors nearby the detector. Its latency is $\approx$ 300 ms [74, 73].

Single lepton triggers are used in this analysis. If a muon or electron with momentum above a certain p_{T} threshold is reconstructed in an event, the event is stored. The threshold depends on the type of the lepton and on the year of data taking. It varies from 20 to 26 GeV. The trigger efficiency rises with p_{T}, reaching $\approx$ 100 % for p_{T} higher than the trigger threshold plus 1 GeV [73].

Chapter 4

Object reconstruction

Although there are many particles that can be produced during the collisions, only few of them are stable enough to be measured directly in the detector: p, n, e, μ, γ, $\pi^{\pm}$, $K^{\pm}$ and K^{0}. All the other particles decay before they could reach the detector, except of weakly interacting neutrinos which cannot be measured by the ATLAS detector.

Heavy unstable particles decays into other, more stable particles which can be measured in a detector. In order to reconstruct these particles, multiple objects are defined in ATLAS analyses:

Tracks: When a charged particle flies through the inner detector, it causes a ionization which can be measured. The tracks of these particles in the inner tracker are reconstructed in order to measure momentum and charge of these particles. Only charged particles leave tracks in the inner detector, while neutral particles, such as photons and neutral hadrons fly through the inner detector without leaving any signal.

Electrons: The electron is a stable particle. Since it carries an electric charge, it leaves a signal in the inner detector (track) and in the electromagnetic calorimeter.

Muons: The mean lifetime of the muon, 2.2×10^{-6} s[2], makes it stable enough to be directly measured by ATLAS detector. Because of relatively high mass, ≈ 200 times higher than the electron mass, the muon can penetrate through the all subdetectors of ATLAS, leaving a signal in all sensitive parts of the detector. The muon is the only particle able to reach the outermost layer of the ATLAS detector, the muon spectrometer.

Jets: However the quarks and gluons cannot be observed directly, streams of particles, so called jets is their experimental evidence. The jets consist of hadrons and possibly leptons from decays of the hadrons. The jets are measured from a signal in the inner detector and calorimeters. The jets can be initiated also by hadronically decaying τ-leptons, although these jets are not used in this analysis.

Photons: The photons, particles without an electric charge leave a signal only in the electromagnetic calorimeter. Since the photons are not used in our analysis, they will not be described in more details.

4.1 Track reconstruction

Tracks are reconstructed from the signal in the inner detector. At the beginning of the track reconstruction, hits in the pixel detector and SCT are reconstructed. Each hit is characterized by 3 spatial coordinates (so-called silicon space-point) and a cluster of cells assigned to it. There are two possible types of clusters. Single-particle clusters are clusters created by a charge deposit from a single charged particle. Merged clusters are clusters containing a charge deposit from more than one charged particle. A Neural Network is employed to discriminate between the merged clusters and single-particle clusters. In the next step, combinations of exactly three silicon space-points, so-called track seeds, are created. The track seeds are used as the inputs for Kalman filter [75], which is employed to build track candidates, adding more silicon space-points into the track seeds. The tracks are characterized by five parameters: q/p, ϕ, η, d_0 and z_0, where p is momentum of the track, q is the track charge, ϕ and θ are angular coordinates of the track, d_0 and z_0 are the transverse and longitudinal distances from the primary vertex of the event[1], as shown in Figure 4.1.

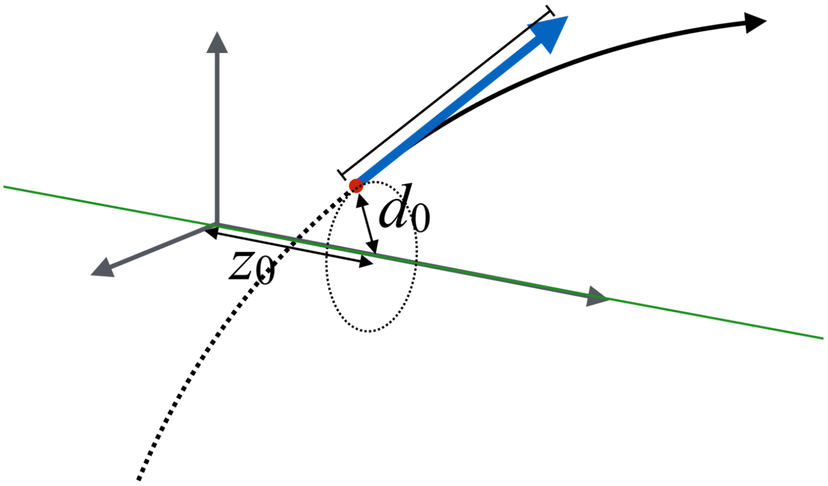


Figure 4.1: z_0 and d_0 of a track [76].

---

[1] The primary vertex is the vertex with the highest p_T sum of the tracks compatible with the vertex. It is assumed to be the vertex from the hard-scattering event.

In the next step, a score for each track candidate is calculated. A higher number of hits increase the score, while crossing a sensitive part of the inner detector without leaving a signal (hole) decreases the score. The χ^2 of the fit is also taken into account in order to penalize the track candidates with a poor fit. The logarithm of the track momentum is also considered to promote energetic tracks [77].

After calculating the score for all the track candidates and ordering them decreasingly by the score, the ambiguities are being solved. If two track candidates share the same track seed, the candidate with the lower score is removed. If two track candidates share the same cluster, which has been identified as a single-particle cluster, the track candidate with the lower score is removed. If the track has more than two shared clusters, it is removed. In addition, all the track candidates are required to meet the following criteria [77]:

- $p_{\mathrm{T}} > 400$ MeV
- $|\eta| < 2.5$
- At least 7 hits in pixel detector and SCT (12 are expected in the ideal case)
- Maximally either one shared pixel cluster or two shared SCT clusters on the same layer
- At most two holes (crossing a sensitive part without leaving a signal) in the SCT and pixel detector

4.2 Electrons

Electrons are reconstructed combining an energy deposit in the electromagnetic calorimeter with a matched track reconstructed in the inner detector. In order to estimate energy deposit of electron candidate in EM calorimeter, the η-ϕ space is divided into grid of 200×256 towers, each of them having a size of 0.025×0.025. A sliding window algorithm [78] is employed to search for the energy deposit matched to the track of the electron candidate, searching for a window position with the maximal energy deposit in the EM calorimeter in each layer. The size of the window is 7×3 in the barrel part and 5×5 in the end-cap parts of the EM calorimeter.

In order to discriminate the real electrons from other particles among the electron candidates, a likelihood based discriminant d, is employed:

$$d = \frac{L_S}{L_S + L_B}, \qquad L_{S(B)}(\vec{x}) = \Pi_{i=1}^{n} P_{S(B),i}(x_i), \tag{4.1}$$

where $\vec{x}$ is a set of input variables, discriminating the electron candidates originating from real electrons against those originating from other particles (fake electrons). The list of these variables can be found in Reference [79]. The $P_{S(B),i}(x)$ are probability density functions of these variables for real (fake) electrons, derived from a simulation. The discriminant d is usually transformed according to the Equation 4.2 in order to obtain a discriminant with a sharp peak around one.

$$d' = -\frac{1}{15}\ln(d^{-1} - 1) \tag{4.2}$$

The distribution of the d' for real and fake electrons is shown in Figure 4.2

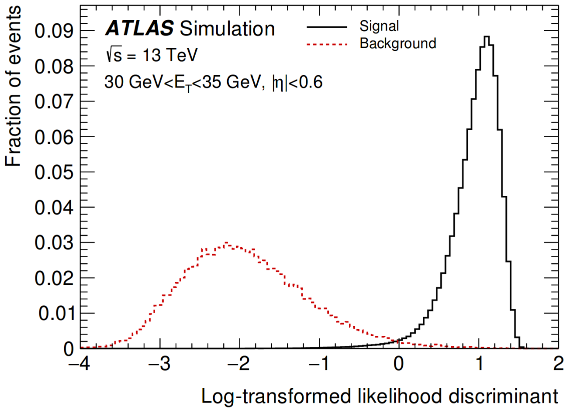


Figure 4.2: Distribution of the electron likelihood discriminant d' for electrons in 30 GeV $< E_{\mathrm{T}} <$ 35 GeV and $|\eta| < 0.6$ for real electrons (black) and fake electrons (red). Both histograms are normalized to unit area [79].

The ATLAS collaboration uses three cut thresholds for d', when defining the electrons: Loose, Medium and Tight. The real and fake electron efficiencies for these lepton definitions as functions of E_{T} of the electron candidate are shown in Figure 4.3

In addition to the likelihood cut, the electron candidates are required to meet isolation criteria. The aim of the isolation criteria is to suppress a fraction of non-promt electrons, originating from photon conversions, heavy hadron decays and light hadrons mis-identified as electrons, which are usually accompanied by other particles.

In this analysis so-called *FixedCutTight* isolation is applied. The *FixedCutTight* includes requirements on both, the calorimeter and inner detector. The sum of p_{T} of all other tracks in cone $\Delta R = 0.2$ around the electron candidate must be below 6 % of

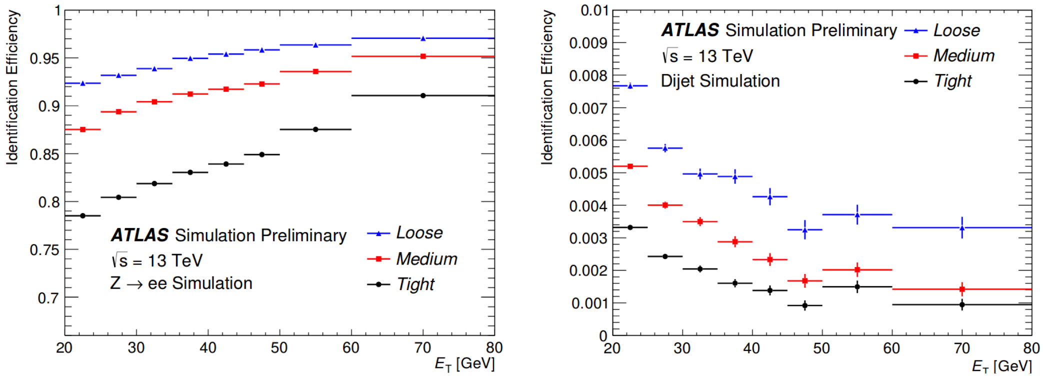


Figure 4.3: Real (left) and fake (right) electron efficiencies for Loose, Medium and Tight electron candidates as a function of E_T [80].

the electron candidate p_T. The similar requirement is applied also for the signal from the EM calorimeter, where sum of the transverse energies of all clusters not assigned to the electron candidate, in the cone $\Delta R = 0.2$ around the electron candidate, must be below 6 % of the candidate E_T [81].

In addition to the likelihood and isolation criteria, the electrons are required to meet the following criteria:

- $p_T > 7$ GeV
- $|\eta| < 2.47$
- electrons with $1.37 < |\eta| < 1.52$ are rejected because of the transition between end-cap and barrel parts of the EM calorimeter
- $|d_0|/\sigma(d_0) < 5$
- $|z_0.\sin(\theta)| < 0.5$ mm

The electron reconstruction efficiencies are measured in regions dominated by $Z \to ee$ and $J/\Psi \to ee$ events. In order to measure the efficiencies, a tag and probe method is used, when one electron from the pair is tight and the other one is loose. Measuring the probability of the loose electron to pass the tight requirement, the reconstruction efficiency can be extracted. Comparing the electron reconstruction efficiencies in data and Monte Carlo, multiplicative scale factor, close to one can be obtained. It is further applied in the Monte Carlo in order to improve the agreement between data and simulation [80].

4.3 Muons

The muon candidates are reconstructed from two independent tracks, in the inner detector and in the muon spectrometer (MS). A global fit is applied to both tracks in order to extract the moun momentum. To suppress muons originating from hadronic decays, a quality of the combined track is taken into account. If the track in the inner detector was caused be a meson which decayed later into the muon detected by MS, the fit of the track is expected to be poor because a part of the hadron momentum has been carried away by a neutrino. The set of cuts on discriminating variables is used in order to suppress the muons from hadronic decays. The complete list of the variables can be found in Reference [82].

The *Medium* muons are required to fulfil criteria on q/p significance (the difference between the q/p ratios calculated from the ID and MS divided by corresponding uncertainty), ρ which is related to a difference in p_{T} measured in the ID and MS and χ^2 of the combined track fit. In addition, the muon candidates used in this analysis are required meet the following [82]:

- $p_{\mathrm{T}} > 7$ GeV
- $|\eta| < 2.5$
- at least one pixel hit, at least five SCT hits
- less than three SCT or pixel holes
- at least 10% of the TRT hits originally assigned to the track included in the final fit (only for $0.1 < |\eta| < 0.9$)
- at least three hits in at least two MDT layers
- q/p significance less than 7

Similarly to electrons, isolation criteria are applied in order to suppress non promt muons. The *FixedCutTightTrackOnly* isolation is used, requiring sum of p_{T} of all other tracks in cone of $\Delta R = \min(10\ \mathrm{GeV}/p_{\mathrm{T}}, 0.3)$ to be less than 6% of the muon p_{T} [80].

The muon reconstruction efficiencies, similarly to the electron efficiencies, are measured in $Z \to \mu\mu$ and $J/\Psi \to \mu\mu$ events. The corresponding multiplicative scale factors derived in these regions are used to correct a Monte Carlo.

4.4 Jets

As a consequence of the quark confinement, the quarks cannot exist as free particles and thus cannot be directly measured in a detector. Quarks create colourless bound states, hadrons. The experimental evidence of quarks and gluons, in high energy physics, are streams (jets) of particles, usually light hadrons. Assuming that the hadronization process has a small impact on the initial parton fourmomentum, the fourmomentum of the original quark or gluon can be determined measuring the jet fourmomentum.

The jet reconstruction algorithm must not be sensitive to a collinear gluon emission (so-called infrared divergence), gluon splitting and pile-up.

Topological cell clustering

In order to reconstruct the jet fourmomentum, calorimeter cells with a signal above a certain threshold are merged into topo-clusters. For each calorimeter cell a signal significance $\sigma^{\mathrm{EM}}_{\mathrm{cell}}$ is calculated [83].

$$\sigma^{\mathrm{EM}}_{\mathrm{cell}} = \frac{E^{\mathrm{EM}}_{\mathrm{cell}}}{\sigma^{\mathrm{EM}}_{\mathrm{noise,cell}}}, \tag{4.3}$$

where $E^{\mathrm{EM}}_{\mathrm{cell}}$ is the reconstructed energy in the cell and $\sigma^{\mathrm{EM}}_{\mathrm{noise,cell}}$ is an expected noise. The cells with $\varsigma^{\mathrm{EM}}_{\mathrm{cell}} > 4$ are used as primary seeds. The neighbouring cells with $\varsigma^{\mathrm{EM}}_{\mathrm{cell}} > 2$ are merged together with the seed. The merging algorithm is stopped once there are no more cells with $\varsigma^{\mathrm{EM}}_{\mathrm{cell}} > 2$ to be merged to a neighbouring cluster. The topo-clusters from the algorithm are then used in an anti-k$_{\mathrm{T}}$ algorithm.

anti-k$_{\mathrm{T}}$ algorithm

In high energy physics, various jet clustering algorithms are employed to build jets from the topo-clusters. A jet clustering algorithm defines a measure of distance between two objects and merges them together if their distance is bellow a certain threshold. The following equation is usually used as the measure of the distance d_{ij} between i and j objects.

$$d_{ij} = \min(k^{2p}_{\mathrm{T}i}, k^{2p}_{\mathrm{T}j}) \frac{\Delta^2_{ij}}{R^2}, \tag{4.4}$$

where $k_{\mathrm{T}i}$ is the transverse momentum of object i, $\Delta^2_{ij} = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ is a cone between the objects and R is a desired radius of the jet [84].

In this analysis the anti-k$_T$ algorithm is used (p = -1) and radius of the jet is $R = 0.4$.

The algorithm loops over all pairs of the objects, looking for the pair with minimal d_{ij}. If $d_{ij} < d_{i,\mathrm{B}}$ the particles are merged together into a pseudo-jet ($d_{i,\mathrm{B}}$ is the distance from the beam). The algorithm stops the merging process when there is no pair of objects with $d_{ij} < d_{i,\mathrm{B}}$.

Jet Vertex Tagger

In order to suppress jets not originating from the hard scattering, the Jet Vertex Tagger (JVT) [85] is employed. There are two sources of the pile-up contributing into the measured signal in calorimeter: in-time and out-of-time pile up. In-time pile up refers to additional pp interactions in the same bunch crossing that triggered the event. Out-of-time pile up refers to energy deposits in calorimeters from previous or next bunch crossings.

The JVT is a multivariate tagger using k-Nearest Neighbour algorithm [86], combining two variables, $corrJVF$ and R_{pT} into a single output value. The k-Nearest Neighbour algorithm looks for 100 closest events in two dimensional $corrJVF$-R_{pT} plane, returning the ratio of signal events with respect to the all events.

$corrJVF$ is an improved version of Jet Vertex Fraction, which is a scalar sum of p_{T} of the tracks associated to the jet coming from the primary vertex, divided by the p_{T} sum of all the tracks associated to the jet. The JVT itself is sensitive to the pile up and its distribution for hard scattering jets depends on the number of interactions per bunch crossing. As a consequence, the efficiency for hard scattering jets passing a cut for $JVT > JVT_{min}$ depends on the pile up. The $corrJVF$, defined by Equation 4.5 fixes this problem, introducing a pile-up dependent term in its definition.

$$corrJVT = \frac{\sum_k p_{\mathrm{T}}^{\mathrm{trk},k}(PV_0)}{\sum_k p_{\mathrm{T}}^{\mathrm{trk},k}(PV_0) + \frac{\sum_{n\geq 1}\sum_l p_{\mathrm{T}}^{\mathrm{trk},l}(PV_n)}{k.n_{\mathrm{trk}}^{\mathrm{PU}}}}, \tag{4.5}$$

where $p_{\mathrm{T}}^{\mathrm{trk},k}(PV_n)$ is p_{T} of the k-th track associated to the n-th vertex ($n = 0$ for the primary vertex), $n_{\mathrm{trk}}^{\mathrm{PU}}$ is the number of tracks associated to other vertices than the hard-scattering vertex (number of pile-up tracks) and $k = 0.01$.

The variable $R_{p\mathrm{T}}$ is defined as the scalar sum of p_{T} of all the tracks associated to the jet that originate from the primary vertex, divided by the p_{T} of the fully calibrated jet.

The fraction of hard-scattering jets accepted by cuts on JVF, corrJVF, $R_{p\mathrm{T}}$ and JVT as a function of pile-up is shown in Figure 4.4. The plot c) in Figure 4.4 shows efficiency for jets originating from other pp interaction than the primary vertex of the event (pile-up jets).

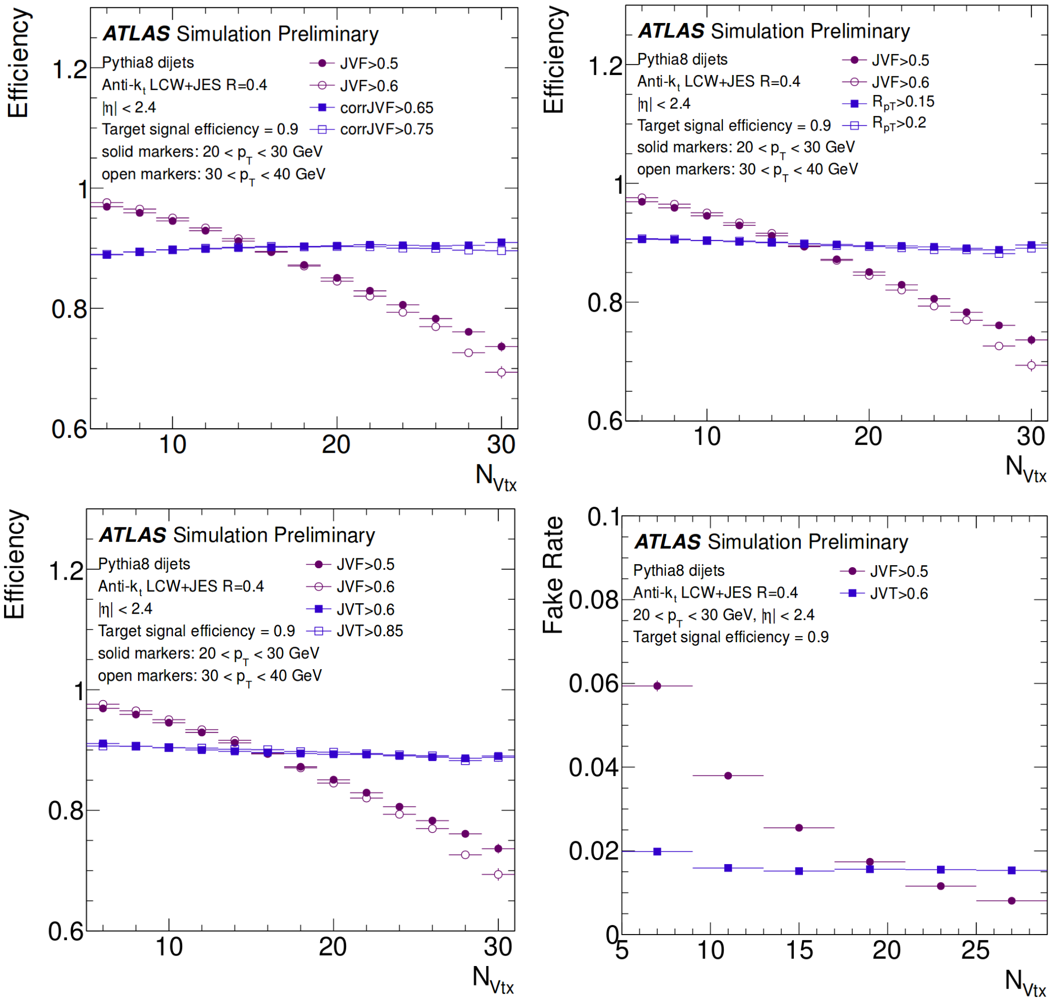


Figure 4.4: Clockwise a-d from the top-left corner: a) Hard-scattering jets efficiency at given JVF and corrJVF cut reaching in average 90 % efficiency. b) Hard-scattering jets efficiency at given JVF and R_{pT} cut reaching in average 90 % efficiency. c) Hard-scattering jets efficiency at given JVF and JVT cut reaching in average 90 % efficiency. d) Pile-up jets efficiency at given JVF and JVT cut reaching in average 90 % efficiency for hard-scattering jets [85].

The JVT response is calibrated using $Z \to \mu\mu$+jets events. Requiring the angle between the lepton pair and the jet to be higher than 2.8 and being balanced in p_{T}, a very pure sample of hard-scattering jets can be obtained. Comparing data with Monte

Carlo prediction in this region, a multiplicative scale factor can be obtained [85].

The JVT > 0.59 cut, reaching 92 % efficiency for the jets from hard scattering, is used in the analysis for the jets with $p_{\mathrm{T}} < 60$ GeV. The JVT cut is dropped for the jets with $p_{\mathrm{T}} > 60$ GeV, since the pile-up contribution in this area is negligible.

The jets in this analysis are required to satisfy $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$ requirements and they are built using the anti-$\mathrm{k_T}$ algorithm with $R = 0.4$.

4.5 *b*-tagging

Presence of a *b*-quark in an event is often sign of an interesting physics, for example in decays of the top quark. Although the mass of the *b*-quark is more than 4 GeV, it is quite stable because of low values of off-diagonal elements of the CKM matrix $|V_{cb}| \approx 0.04$ and $|V_{ub}| \approx 0.003$ [2]. This makes *b*-hadrons being relatively stable with mean lifetime of $\approx 10^{-12}$ s and enable them to travel ≈ 0.5 mm from the point where they have been produced (primary vertex) to a point where they decay, forming a secondary vertex.

Given a spatial resolution of Pixel Detector, the ATLAS is able to reconstruct the secondary vertices from decays of *b*-hadrons. Combining the information about the secondary vertex, possible tertiary vertex from a subsequent *c*-hadron decay and information about transverse and longitudinal impact parameters of tracks assigned to the jet (reconstructed by IP2D and IP3D algorithms), a multivariate discriminant can be obtained [87].

The IP2D and IP3D algorithms [88] reconstruct a track trajectory in two dimensions xy (IP2D) or in three dimensions (IP3D). Reconstructing the point where the track trajectory is closest to the primary vertex, so-called impact parameter, an information about the secondary vertex displacement can be obtained. If the point of the closest approach of the track is in front of the primary vertex in direction of the jet, the impact parameter is positive. In the other case the impact parameter is negative. The positive value of the impact parameter may be a sign of a secondary vertex, while the negative value can be a sign of a pile-up track, track from interaction with detector material or a poor fit of the track. The IP2D algorithm is more robust against the pile-up, since it does not take into account longitudinal position of the closest approach point, which is typically large for pile-up tracks.

The SV (secondary vertex) algorithm [88] provides an additional set of input variables for the *b*-tagging. The algorithm reconstructs all vertices formed by at least two tracks, rejecting the vertices which are likely to originate from light hadrons, such as kaons. The secondary vertex reconstruction efficiency depends on the jet p_{T}, η and type of the jet. The average efficiency reaches 70-80 % for b-jets, ≈30 % for c-jets and ≈3 % for light jets [87].

JetFitter [89] algorithm is used to reconstruct all vertices inside a jet, trying to reconstruct also a tertiary vertex from c-hadron decay.

The b-tagger MV2c10 combines information from IP2D, IP3D, SV and JetFitter algorithms, using a set of 21 variables defined in Reference [87]. In addition to these variables, p_{T} and η of the jet are included, so the b-tagger can learn correlations of input variables with p_{T} and η of the jet. In order to avoid the MV2C10 to get trained on the p_{T} and η distributions which provide certain level of b-jet vs. other jets separation themselves, the background (non-b jets) sample is reweighted based on jet p_{T} and η to follow these distributions from the b-jet sample. The MV2c10 is Boosted Decision Tree based b-tagger, trained on a background sample with 7% of c-jets and 93 % of light jets [90].

The distributions of MV2c10 output for b- c- and ligh jets are shown in Figure 4.5. Efficiencies for b- c- and light jets at various cut thresholds are shown in Table 4.1. Cutting on the MV2c10 output at a certain value, a sample enriched in b-jets can be obtained.

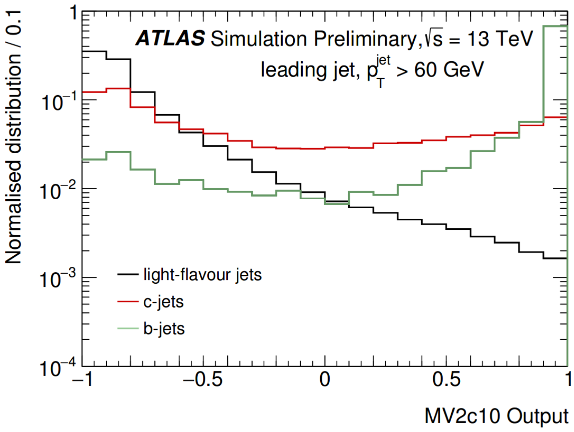


Figure 4.5: Distribution of MV2c10 output for b-jets (green), c-jets (red) and light jets (black) [90]. All the distributions are normalized to unity integral.

WP	Cut value X	b-jet efficiency	c-jet efficiency	light jet efficiency
85 %	0.1758	85 %	32 %	2.9 %
77 %	0.6459	77 %	16 %	0.77 %
70 %	0.8244	70 %	8.3 %	0.26 %
60 %	0.9349	60 %	2.9 %	0.065 %
50 %	0.9769	50 %	0.94 %	0.017 %

Table 4.1: Efficiencies for b-, c- and light jets at various MV2c10 working points. The value of the cut on MV2c10 output reaching the desired efficiencies is shown in the second row [90].

4.6 Missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$)

The ATLAS detector is designed to reconstruct all the particles in an event, except of weakly interacting neutrinos. The neutrinos escape the detector without being detected, carrying away a part of the energy-momentum from the event. Although the neutrinos cannot be detected directly at ATLAS, indirect reconstruction based on the momentum conservation is employed.

The colliding protons have only z-component of the momentum, while the transverse momentum is zero. Because of the momentum conservation, the final state particles must be balanced in p_{T}. If the measured objects are not balanced in p_{T}, it might be a sign of a particle escaping undetected, such as a neutrino.

The missing transverse energy (momentum) is calculated as follows:

$$\vec{p_{\mathrm{T}}^{\mathrm{miss}}} = -\left(\sum_{i}^{\mathrm{hard}} \vec{p_{\mathrm{T},i}} + \sum_{i}^{\mathrm{soft}} \vec{p_{\mathrm{T},i}}\right), \tag{4.6}$$

where the first, hard, term corresponds to fully calibrated objects reconstructed by ATLAS, such as jets, leptons and photons. The other, soft, term corresponds to tracks reconstructed from inner detector, not assigned to any of the reconstructed objects.

In order to take into account a contribution from neutral particles, leaving no signal in the inner detector, the p_{T} of all hard objects, except of the muons, is reconstructed also from the calorimeter signal. The soft term, related to the tracks, is calculated only from the tracks, and thus the soft neutral particles are not taken into account.

The missing transverse energy is reconstructed only in the xy plane. The reconstructed missing transverse energy is characterized by its size (usually refered to as $E_{\mathrm{T}}^{\mathrm{miss}}$) and angle ϕ_{MET} [91].

4.7 Overlap removal

In order to avoid a double counting of the same particle level object in more detector level objects, the overlap removal procedure is applied. The overlap removal algorithm removes step by step objects which are close to each other in the cone ΔR and thus can originate from the same particle, such as electron candidate and jet. The overlap removal is applied in the following steps:

- Electron candidates sharing the same track with a muon candidate are removed.
- All jets with at least one electron candidate in $\Delta R < 0.2$ cone around them are removed.
- If the distance between a jet and electron is $\Delta R < 0.4$, the electron is removed.
- If a muon is in $\Delta R < 0.4$ cone of a jet and the jet has more than 2 associated tracks, the moun is removed.
- If the distance between a jet and muon is $\Delta R < 0.4$, the jet is removed.

Chapter 5

Data and simulated Monte Carlo samples

5.1 Data

During the Run II, in 2015-2018, the LHC delivered 156 fb^{-1} of *pp* collision data to the ATLAS detector. Considering only periods when ATLAS detector was operational and able to record the data, the ATLAS has collected 147 fb^{-1} of data. The collected data from the ATLAS detector are divided into lumi blocks, based on the time of its collection and keeping information about the operational circumstances of each subdetector.

In some of the lumi blocks, the ATLAS magnetic system was switched off, not allowing to measure a charge and momentum of reconstructed particles. These data still can be used by some analyses, not requiring charge and momentum identification of reconstruction objects. Similarly, some analyses might not need another ATLAS subdetectors, which allows them to use the lumi blocks with these parts not being operational. However, given the complexity of this measurement, requiring reconstruction of various objects, all the subsystems of the ATLAS detector are important for this analysis and thus only lumi blocks with the fully operational ATLAS detector can be used. The integrated luminosity formed only by these blocks is 139 fb^{-1} with $\pm$ 1.7 % uncertainty from the full Run II dataset and 36.1 fb^{-1} with $\pm$ 2.1 % uncertainty considering only 2015+2016 period of data taking.

The luminosity is measured by LUCID detector, placed near the beam pipe in a forward region of the ATLAS detector. The LUCID detects particles flying close to the beam, estimating the luminosity from amount of these particles [92].

During the Run II period, the performance of the LHC in terms of the reached luminosity had been improved significantly. The LHC started to collect the data in 2015, with 50 ns frequency of collisions. After a short period of data taking, the bunch spacing was shortened to 25 ns, keeping this frequency for the rest of the Run II data taking. Only data with 25 ns bunch spacing are used in this analysis.

With the increasing of the luminosity, the average number of pp interaction per bunch crossing, pile-up, was increased as well. The higher number of collisions makes an event reconstruction more challenging, worsening also the energy resolution of calorimeter and lepton reconstruction. The pile-up profile for individual years of the Run II, together with the time dependency of the total integrated luminosity collected by the ATLAS detector is shown in Figure 5.1.

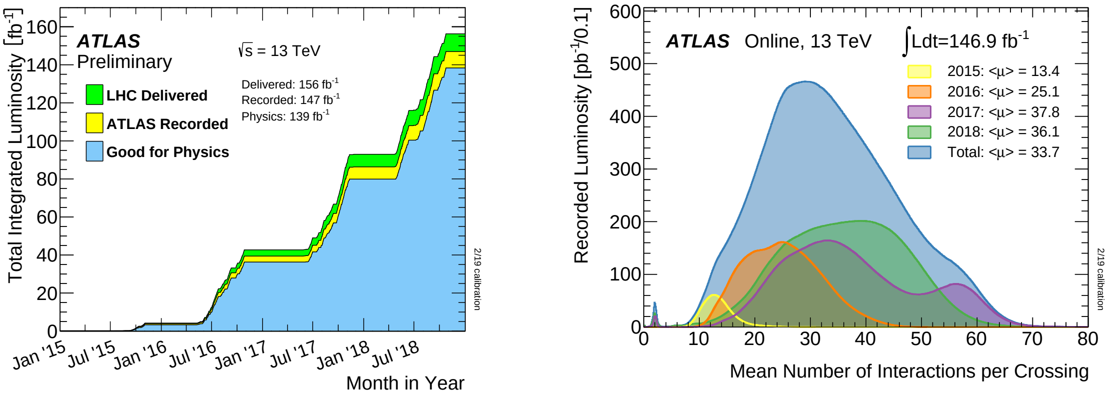


Figure 5.1: The total integrated luminosity delivered to and collected by ATLAS as a function of time (left) and pile-up profile (right) for the Run II dataset collected by ATLAS [93].

5.2 Monte Carlo

In this analysis, the collected data are compared to a simulated Monte Carlo prediction. The Monte Carlo prediction uses random numbers in order to simulate the final state of pp collisions, which are themselves random because of their quantum mechanic behaviour. The Monte Carlo events are used in optimization studies, MVA trainings and in order to get templates for the fit used to extract the $t\bar{t}Z$ cross section.

The generating of Monte Carlo can be split into few parts, each requiring different approaches and algorithms:

Parton Distribution Functions

Since protons are composite particles, the pp collisions must be modelled as collisions of individual partons from both protons. In order to estimate initial momenta of the partons entering the interaction, the parton distribution functions (PDFs), described in Section 2.2.8, are used. This analysis uses Monte Carlo samples generated with the following PDF sets: NNPDF2.3 [94], NNPDF3.0 [94], CTEQ6L1 [95] and CT10 [96].

Hard Scattering (Matrix Element) simulation

The interaction of two partons from the two colliding protons leads to a process described by Feynman diagrams with two initial particles (partons) and set of final state particles such as quarks, leptons, gluons and photons. The input particle momenta are predicted from the PDFs.

A Monte Carlo generator takes into account a calculation of the matrix elements coming from the Feynman diagrams up to a certain level of the perturbation theory, to generate the final state particles, following the distributions predicted by the theory. Some Monte Carlo generators uses a reweighting of the generated events in order to model all the distributions correctly. Negative weights can also appear, typically as a consequence of higher order corrections.

Hadronization and parton showering

The final state quarks and gluons often produce additional final state radiation (FSR) gluons. Photons can be radiated as well, especially by charged leptons. The process of the gluon and photon radiation is simulated at the level of matrix element, if transverse momentum of the radiated particle is higher than a certain threshold. Inclusive calculation of the final state radiation on the matrix element level would lead to infrared divergences. If the momentum of the radiated gluon/photon is bellow the threshold simulated at the ME level, the FSR is modelled by a parton showering algorithm.

Since quarks and gluons are colour objects and thus cannot exist as stable particles, they form hadrons. Although no exact theory of the hadronization process is known, there are two commonly used models: string model used by PYTHIA[97] and cluster model used by HERWIG [98].

Detector response

The particles simulated in the hadronization and parton showering processes interact with material of a detector, leaving energy deposits and signals in active parts of the detector. This process is simulated using GEANT4 [99]. This is the most CPU intensive part of the event generating. In order to save computational sources approximate algorithm of detector response, ATLFAST-II [100], is often used for some samples which do not require a proper detector simulation.

5.2.1 $t\bar{t}$ + V samples

The associated production of the top-quark pair with at least one vector boson (Z, W or photon) are simulated using NNPDF2.3 ($t\bar{t}\gamma$) and NNPDF3.0 (others) sets of the PDFs. The matrix element part is simulated by aMC@NLO [101] generator (NLO in QCD and LO in EW). The top-quark mass is set to 172.5 GeV for all these samples. The parton showering and hadronization is modelled using PYTHIA8 [97] and decays of heavy flavour hadrons are modelled using EvtGen [102]. The A14 [103] tune of Monte Carlo generators is used in the simulation. The off-shell contribution of Z/γ, as well as their interference, is considered in $t\bar{t}Z$ samples for $m_{\ell\ell} > 5$ GeV. The contribution of samples with dilepton invariant mass bellow 5 GeV has been checked and it has been found to be negligible. The samples are normalized according to NLO QCD + EW cross sections.

5.2.2 Z+jets samples

The Z+jets process is simulated using SHERPA 2.2.1 [104] generator at NLO precision up to 2 additional partons and LO up to 4 additional partons. The NNPDF3.0 [94] set of PDFs is used. SHERPA is used also to simulate the parton showering and hadronization. Only the samples with Z boson decaying into charged leptons (e, μ or τ) are considered. In order to provide a sufficient number of events in all parts of the phase space, the samples are sliced based on the value of $max(H_{\mathrm{T}}, p_{\mathrm{T}}^{Z})$, where H_{T} is a scalar sum of p_{T} of all partons in the event and p_{T}^{Z} is Z-boson transverse momentum. The samples are normalized to the NNLO cross section.

5.2.3 $t\bar{t}$ samples

The $t\bar{t}$ sample is generated using POWHEGBOX [105] at the matrix element level, taking into account NLO in QCD precision. The parton showering and hadronization are modelled using Pythia8 [97]. The A14 [103] tune of Monte Carlo generators is used in the simulation. The top-quark mass is set to 172.5 GeV and h_{damp} parameter, handling gluon emissions, is set to 1.5 of the top-quark mass. In order to estimate a contribution from real leptons, the sample containing only dilepton decay of the top-quark pair is considered. The fake lepton contribution is estimated from non-all hadron inclusive samples, containing ℓ+jets and dilepton decays of the top-quark pair. The $t\bar{t}$ samples are normalized to NNLO + NNLL cross section prediction.

The Monte Carlo $t\bar{t}$ samples are used only to train a multivariate technique. In the further steps of the analysis, a data driven (DD) technique of $t\bar{t}$ background estimate is used. The DD $t\bar{t}$ estimate will be described later.

5.2.4 tWZ sample

The associated production of a single top quark with W and Z bosons is generated using NNPDF3.0 [94] set of PDFs, matrix element is calculated by aMC@NLO [101] at NLO in QCD precision and parton showering and hadronization is modelled by Pythia8 [97]. The A14 [103] tune of Monte Carlo generators is used in the simulation and top-quark mass is set to 172.5 GeV. The sample is normalized to the NLO cross section, obtained from the aMC@NLO generator and shown in Table 5.1. A diagram removal is employed in order to remove overlap and interference with $t\bar{t}Z$ and $t\bar{t}$ samples. The Z boson is required to decay leptonicaly in the tWZ sample.

5.2.5 $ZZ \to \ell\ell\ell\ell$ samples

The ZZ decaying into four leptons is a main background in the tetralepton channel. The sample has been generatered using Sherpa 2.2.2 [104] for both matrix element and parton showering and hadronization. The CT10 PDF set [96] has been used. The sample is calculated at LO precision up to two additional partons.

In the 2ℓ OS channel analysis, using only 2015+2016 data, the ZZ background was modelled using Sherpa 2.1 [104] samples, at NLO precision up to one additional parton and LO up to three additional partons. The CT10 PDF set [96] has been used.

5.2.6 Other samples

A contribution from other processes is very small in both 2ℓ and 4ℓ channels. The summary of all the samples with corresponding cross sections and used set of PDFs and generators can be found in Table 5.1

Sample	ME Generator	PDF	Shower	Normalisation	Cross section [pb]
$t\bar{t}Z(Z \to \ell\ell)$	aMC@NLO	NNPDF3.0	Pythia8	NLO	0.1237
$t\bar{t}Z(\to qq)$	aMC@NLO	NNPDF3.0	Pythia8	NLO	0.172039
$t\bar{t}Z(\to \nu\nu)$	aMC@NLO	NNPDF3.0	Pythia8	NLO	0.585758
$t\bar{t}W$	aMC@NLO	NNPDF3.0	Pythia8	NLO	0.6008
tZ non-allhad	MadGraph	CTEQ6L1	Pythia6	LO	0.240
tWZ	aMC@NLO	NNPDF3.0	Pythia8	NLO	0.0156
$\ell\ell\ell\nu$	Sherpa 2.1	CT10	Sherpa	NLO	11.9
$\ell\ell\ell\ell$	Sherpa 2.1	CT10	Sherpa	NLO	11.5
$\ell\ell\nu\nu$	Sherpa 2.1	CT10	Sherpa	NLO	12.8
$gg \to \ell\ell\nu\nu$	Sherpa	CT10	Sherpa	NLO	0.78
$t\bar{t}$	Powheg	NNPDF3.0	Pythia8	NLO	831.76
$4t$	MadGraph	NNPDF2.3	Pythia8	LO	0.0092
$3t$	MadGraph	NNPDF2.3	Pythia8	LO	0.00164
$t\bar{t}WW$	MadGraph	NNPDF3.0	Pythia8	NLO	0.0099
$t\bar{t}H$	aMC@NLO	NNPDF3.0	Pythia8	NLO	0.5065
$t\bar{t}\gamma$	aMC@NLO	NNPDF2.3	Pythia8	NLO	2.982
VH	EvtGen	NNPDF2.3	Pythia8	LO	2.2496
$gg \to H(\to 4\ell)$	Powheg	CTEQ6L1	Pythia 8	NNLO+NNLL	0.0081
Z + jets	Sherpa 2.2.1	NNPDF3.0		NNLO	2107.0
W + jets	Sherpa 2.2.1	NNPDF3.0		NNLO	20080.0
Single top t-channel	Powheg	CT10	Pythia6	LO	70.428
Single top Wt	Powheg + EvtGen	CT10	Pythia6	LO	71.7
Single top s-channel	Powheg + EvtGen	CT10	Pythia6	LO	3.35
Triboson	Sherpa 2.1	CT10	Sherpa	LO	0.0166

Table 5.1: Summary of the Monte Carlo samples used in the 2ℓ OS channel analysis. In the 4ℓ channel analysis, Sherpa 2.2.2 is used to model the diboson samples (ZZ). Other samples are produced using the same generators in both 2ℓ and 4ℓ channels.

5.3 Estimation of a fake lepton background

The fake lepton background plays an important role in the 3ℓ and 4ℓ channel analysis, while its fraction is very small in the 2ℓ channel. The fake lepton background in the 2ℓ channel analysis is estimated from Monte Carlo and ± 50 % normalization uncertainty is assigned to it.

In the 4ℓ channel, the fake lepton background is more significant and more sophisticated approach is necessary. Two methods of the fake lepton background estimate were tested: fully data driven Matrix Method (MM) and semi data driven Fake Factor method (FF). The methods provide consistent results and show a reasonable data/MC agreement in control regions. The Matrix Method, used also in the 3ℓ channel, has been chosen as the nominal method of fake lepton estimate in 4ℓ analysis with full Run II dataset and FF method is used to estimate a systematic uncertainty of the fake lepton contribution.

5.3.1 Matrix Method

The Matrix method is a fully data driven method for the fake lepton estimate. It uses events with one loose lepton to estimate the fake lepton contribution to regions with tight leptons only. The difference between the loose and tight leptons is usually only in an isolation requirement. While the tight lepton definition is consistent with the leptons used in the analysis, no isolation requirement is applied on the loose leptons.

In order to define a probability for a loose lepton to pass a tight selection, the efficiencies r and f are introduced. The real lepton efficiency, r, is the probability for a real loose lepton to pass the tight selection. The fake lepton efficiency, f, is the probability for a fake loose lepton to pass the tight selection. The number of loose and tight leptons as functions of total number of fake and real loose leptons can be written as:

$$N_{\mathrm{loose}} = N_{\mathrm{loose}}^{real} + N_{\mathrm{loose}}^{fake} \tag{5.1}$$

$$N_{\mathrm{tight}} = r.N_{\mathrm{loose}}^{real} + f.N_{\mathrm{loose}}^{fake}, \tag{5.2}$$

where real and fake lepton efficiencies are parametrized in a set of variables characterizing either the loose lepton or the whole event. In the case of this analysis, the efficiencies are parametrized in η and p_{T} of the loose lepton. The real and fake lepton efficiencies for electrons and muons are taken into account as independent functions.

The real lepton efficiencies are determined from $Z \to \ell\ell$ Monte Carlo simulation, measuring the ratio of event yields for TT (tight-tight) and TL (tight-loose) events from the events with exactly two leptons, both truth matched to a real promt lepton at particle level.

The fake lepton efficiencies are measured in dilepton same-sign control regions in data. The promt lepton background estimated from Monte Carlo is subtracted from the data. The following fake lepton control regions have been used:

- Electron control region: exactly one muon and exactly one electron, same-sign of the lepton charges, exactly one b-jet and no other jet
- Exactly two muons, no electron, same-sign of the lepton charges, exactly one b-jet and no other jet

The efficiencies are measured as functions of η and p_{T}, using the binning summarized in Table 5.2

Bin ranges	p_{T} [GeV]	$\lvert\eta\rvert$
Muons	[7, 12, 20, 35, 50, ∞]	[0, 0.5, 1, 1.5, 2, 2.5]
Electrons	[7, 12, 20, 35, 50, ∞]	[0, 0.7, 1.37, 1.52, 2, 2.47]

Table 5.2: p_{T} and $|\eta|$ bin ranges chosen for the fake and real lepton efficiency measurements.

Considering Eq. 5.1 and Eq. 5.2, the fake lepton contribution in tight regions can be estimated as follows:

$$N_{\mathrm{tight}}^{fake} = \frac{f}{r-f}(rN_{\mathrm{loose}} - N_{\mathrm{tight}}) \tag{5.3}$$

5.3.2 Fake Factor method

The Fake Factor method is a semi data driven method of the fake lepton estimate. It uses scale factors derived from data to correct a prediction of Monte Carlo fakes. The basic idea of the FF method is that the fake lepton contribution is split into four components: fake electrons and fake muons and each of them is further split into leptons from decay of heavy hadrons and leptons from other sources. The FF method is based on an assumption that the shapes of the contributions are well modelled in Monte Carlo and only their normalizations have to be corrected. A separate scale factor is assigned to each of the four types of the fake leptons and the Monte Carlo prediction is reweighted according to the corresponding scale factor.

In order to extract the fake factors from data, four control regions are defined, each dominated by different type of the fake leptons. The control regions are designed to target dileptonically decaying Z+jets or $t\bar{t}$ events with one additional fake lepton. The control region definitions are shown in Table 5.3.

	Z+jets CR	$t\bar{t} \to \ell\ell$ CR
Leptons	$= 3$	$= 3$
Leptons	one Z-like pair	no OSSF pair, sum of charges ± 1
W transverse mass	< 50 GeV	-
$E_{\mathrm{T}}^{\mathrm{miss}}$	< 50 GeV	-
$p_{\mathrm{T}}^{\mathrm{1jet}}$	-	> 30 GeV
N_{jets}	≥ 2	≥ 2

Table 5.3: Definitions of fake control regions for deriving the fake factors. "OSSF" stands for opposite sign, same flavour. Both Z+jets and $t\bar{t}$ are further split into electron and muon control regions based on the type of the fake lepton candidate. In Z+jets CR, the lepton not belonging to the Z-like pair (OSSF, $|M_{\ell\ell} - M_Z| < 10$ GeV) is considered to be the fake lepton. For $t\bar{t} \to \ell\ell$ CR, the lepton with the lowest p_{T} from the same-sign lepton pair is considered to be the fake lepton. The W transverse mass stands for the invariant mass built from missing transverse energy and transverse component of the fake lepton energy-momentum. Cuts on the transverse mass and missing transverse energy are applied in order to reduce the WZ background.

Since the number of fake muons originating from other sources was not high enough to enable an estimate of the corresponding scale factor, its value was set to 1.00 and 50% uncertainty was assigned to it. Fitting the event yields in the four control regions, the following values of the scale factors, for fake electrons and fake muons originating from heavy flavour decays and from the other sources, have been obtained. The quoted uncertainties are statistical only.

$$\mu^{e}_{heavy} = 0.62 \pm 0.15 \tag{5.4}$$

$$\mu^{e}_{other} = 1.84 \pm 0.29 \tag{5.5}$$

$$\mu^{\mu}_{heavy} = 1.20 \pm 0.09 \tag{5.6}$$

$$\mu^{\mu}_{other} = 1.00 \pm 0.50 \tag{5.7}$$

Chapter 6

Systematic uncertainties

6.1 Experimental uncertainties

6.1.1 Luminosity

The uncertainty on the luminosity is estimated from a measurement performed by LUCID detector [92].

The systematic uncertainty on luminosity for years 2015 and 2016 of data taking was found to be 2.1 %. The uncertainty on luminosity is in the 2ℓ channel applied to all processes except for the $t\bar{t}$, Z+1HF and Z+2HF backgrounds. The $t\bar{t}$ background is data-driven and thus does not suffer from the uncertainty on luminosity. Normalizations of Z+1 HF and Z+2 HF backgrounds are described by free parameters of the fit. In order to improve the stability of the fit, it is not applied to these backgrounds, since the luminosity uncertainty has only normalization (without any shape) effect. Application of a normalization-only uncertainty on the Z+HF background would have had no impact on the fitted mean value of the $t\bar{t}Z$ cross section and its uncertainty.

The full Run II dataset is used in the 4ℓ channel measurement. The systematic uncertainty on the luminosity from the full Run II was found to be 1.7 %. It was applied to all Monte Carlo samples.

6.1.2 Pile-up

Monte Carlo is often generated before the pile-up profile of data is precisely known. In order to make the pile-up profiles in data and Monte Carlo consistent, the Monte Carlo is reweighted based on the measured pile-up profile in the data. The number of pp interactions during a bunch crossing has an impact on p_{T} resolution of the jets. The uncertainty on jet p_{T} measurement originating from the different pile-up profile in data and Monte Carlo is taken into account using an alternative pile-up reweighting with a different pile-up profile, given by our limited knowledge of the pile-up [106].

The pile-up reweighting systematics is applied to all Monte Carlo samples in this analysis.

6.1.3 Systematic uncertainties related to leptons

When reconstructing the leptons from data and Monte Carlo, various selections have to be made. The leptons are required to fulfil identification and isolation criteria, and in addition, one of the leptons in an event is required to fire a trigger. The leptons in this analysis are also required to pass p_{T} and η cuts.

Efficiencies of all of these selections are known only within an uncertainty which has to be taken into account. In addition, uncertainties on lepton energy resolution and scale have to be taken into account.

Lepton efficiencies

The lepton identification, isolation and trigger efficiencies are measured in $Z \to \ell\ell$ and $J/\Psi \to \ell\ell$ dominated regions. The tag and probe method is employed to measure the efficiencies and corresponding uncertainties. Events with exactly two leptons with an invariant mass close to the mass of Z or J/Ψ are selected. These events are well dominated by real leptons originating from Z and J/Ψ. In order to estimate a background, which is at a level of 0.1 %, a data driven approach is used [82, 80].

One of the leptons, tag, is required to meet a tight selection. The selection criteria on the other, probe, lepton are relaxed. Measuring efficiencies for the probe lepton to pass the tight selection, corresponding scale factors and their uncertainties are obtained [80, 82].

The systematic uncertainties on lepton efficiencies are applied to all Monte Carlo samples in this analysis.

Lepton energy scale and resolution

A different detector response between data and Monte Carlo on lepton energy reconstruction have to be taken into account. Lepton energy scale and resolution are important in order to estimate the fraction of events passing cuts on p_{T} of individual leptons and invariant mass of lepton pairs used in the analysis. The lepton energy scale and resolution have also significant impact on shapes of MVA input variables.

In order to obtain correction factors improving the data vs. Monte Carlo agreement, in the lepton energy scale and resolution, and their uncertainties, $Z \to \ell\ell$ and $J/\Psi \to \ell\ell$ dominated regions are used. Fitting the invariant mass distributions of e^+e^- and $\mu^+\mu^-$ lepton pairs, the correction factors and their uncertainties are obtained. The correction factor corresponding to lepton energy scale is applied to data to improve the agreement between real and measured lepton energy. The correction corresponding to the lepton energy resolution is applied to Monte Carlo in order to match the resolution observed in data events [82, 107]. The lepton uncertainties are treated separately for each lepton in an event.

6.1.4 Systematic uncertainties related to jets

Jet Vertex Tagger

Systematic uncertainties are applied to the Jet Vertex Tagger efficiency, described in Section 4.4. The three sources of systematic uncertainties have been considered: a residual contamination of pile-up jets after the pile-up suppression, differences in Monte Carlo predictions obtained from different Monte Carlo generators and statistical uncertainty of the JVT efficiency measurement [85]. The three sources of the systematic uncertainties are merged into one systematic variation described by one nuisance parameter in a fit.

b-tagging

In order to improve data vs. Monte Carlo agreement in efficiency for jets to pass a cut on MV2c10 b-tagger output at a given efficiency working point, multiplicative scale factors are used. The scale factors for jets originating from b-hadron decays are derived from regions dominated by dilepton $t\bar{t}$. The scale factors for jets originating from c-hadron decays are derived from regions dominated by lepton+jets decays of the top-quark pairs. The hadronically decaying W bosons produce c quarks in one half of the cases. Reconstructing an event kinematics, pairing jets to the b quarks and to the quarks from the W-boson decay, a sample enriched in c-jets can be obtained. The light jet efficiencies are extracted from multijet events. The effieciencies are parametrized in p_{T} and η of the jets [90].

The systematic uncertainties on the b-tagging efficiencies are taken into account as a set of nuisance parameters, corresponding to different sources of systematics, such as limited number of events in Monte Carlo, limited spatial resolution in track reconstruction e.t.c. The systematic uncertainty on the b-jet efficiency is described by 7 parameters, uncertainty related to the c-jet efficiency is described by 4 parameters and a set of 12 parameters are assigned to the light jet efficiency.

Jet energy scale

In order to ensure the same response on the jet p_{T} (or energy) between data and simulation, a careful calibration is needed. Several multiplicative scale factors for jet p_{T} are derived from a simulation. In the first step, a pile-up contribution is subtracted, taking into account the average p_{T} density of the pile-up particles and area of the jet. In the next step, a jet p_{T} response is calibrated as a function of p_{T} and η, using simulated events.

After applying correction scale factors for η and p_{T} of the jet, the dependency on few more variables is taken into account, multiplying the jet p_{T} by additional scale factors depending on fraction of energy measured in the first layer of the hadronic calorimeter and in the last layer of the electromagnetic calorimeter, number of tracks with $p_{\mathrm{T}} > 1$ GeV, weighted transverse distance in η-φ plane between the tracks and the jet axis and finally a number of muon track segments associated with the jet.

After applying all already mentioned corrections derived from the simulation, in situ calibration methods are used to mitigate differences between p_{T} response in Monte Carlo and data. Events balanced in p_{T}, either dijet events with high p_{T} jets or Z+jets and γ+jets events are used to calibrate the energy scale of the jets.

The uncertainty on the jet energy scale has been found to be less than 1 % for 100 GeV $< p_{\mathrm{T}} <$ 500 GeV in $|\eta| < 1.2$ region and about 4.5 % for the jets in this η range and $p_{\mathrm{T}} = 20$ GeV.

The uncertainties on the calibration coming from Monte Carlo simulation, sample statistics, energy scale of other objects (photons and leptons) and jet flavour composition are taken into account varying energy of individual jets in an event. In order to deal with the jet energy scale uncertainties, a set of 20 nuisance parameters is used in the fit, each related to a specific variation of JES uncertainty originating from the already mentioned sources [108].

Jet energy resolution

The jet energy resolution can be parametrized as a function of the jet p_{T} as follows:

$$\frac{\sigma(p_{\mathrm{T}})}{p_{\mathrm{T}}} = \frac{N}{p_{\mathrm{T}}} \oplus \frac{S}{\sqrt{p_{\mathrm{T}}}} \oplus C, \tag{6.1}$$

where N is a constant related to a noise effect (either electronics or pile-up), S is related to statistical fluctuations of energy losses in the active part of the calorimeter and C is a constant term [109].

Worsening the jet energy resolution in Monte Carlo, by additional jet energy smearing, an alternative sample for JER systematics is obtained, having a worse resolution compared to the nominal. A sample with better (more precise) jet energy resolution cannot be obtained from the Monte Carlo. Symmetrizing the systematics variation of JER in the fit, the two sided systematic uncertainty is obtained.

6.1.5 Systematic uncertainties related to missing transverse energy

The missing transverse energy is calculated from so-called hard term corresponding to reconstructed objects in an event and soft term, corresponding to contribution of tracks not assigned to any objects, as shown in Eq. 4.6. The $E_{\mathrm{T}}^{\mathrm{miss}}$ is therefore affected by the uncertainty in measurement of these tracks/objects. While the uncertainties related to the objects are already included in the systematic uncertainties already mentioned in this section, the soft term uncertainty has to be taken into account separately. The uncertainties on soft-term component are derived in $Z \to ee$ dominated regions, where is no physical missing transverse energy.

6.2 Theoretical uncertainties

Except of the already mentioned experimental uncertainties, theoretical uncertainties related to a cross-section uncertainty, uncertainties related to choice of parton distribution functions and uncertainties related to a modelling of matrix element and parton showering have to be taken into account. Various scale uncertainties (factorization, renormalization, resummation and parton matching scales) have to be taken into account as well.

6.2.1 Scale choice

Factorization and renormalization scales

When calculating a cross section of a given process, two types of divergences appear in the perturbation theory. The ultraviolet divergences (UV), related to a large transferred momentum in loops of Feynman diagrams and infrared divergences (IR) related to a radiation of a massless particle with low momentum or collinear massless particle (photon, gluon). In order to avoid the divergences, artificial scales μ_R (renormalization scale) and μ_F (factorization scale) are introduced. The renormalization scale defines size of the strong coupling constant and should correspond to characteristic transferred momentum of the process in question. The factorization scale is a typical scale for a given process, at which the proton structure is probed, limiting soft gluon emissions by the initial state partons (including these into the parton distribution functions). These scales are arbitrary free parameters and if all the terms of the perturbation series were summed, the obtained result would not depend on their choice. However, if the calculation is performed only up to a certain level of precision (LO, NLO, NNLO ...), there is a dependency of the theory prediction on the scale choice.

The nominal choice of the μ_R and μ_F in the simulation is set to one half of the scalar sum of transverse momentum of all final state partons.

Parton matching scale (CKKW)

In the process of parton showering and hadronization, a radiation of gluons and photons by the final state partons have to be taken into account. In order to avoid a double counting between the matrix element and parton showering, the CKKW scale is introduced. The gluons from the final state radiation above the CKKW scale are simulated at the matrix element calculation and the softer emissions are handled during the parton showering and hadronization.

The nominal choice of the CKKW scale for Sherpa 2.2.1 samples is 20 GeV.

Gluon resummation scale (QSF)

In order to sum over soft and collinear gluon radiation corrections from final state partons, at higher orders of perturbation theory, the QSF scale is introduced. The related systematic uncertainty is evaluated varying the scale, usually by a factor of 2 and 0.5.

6.2.2 Uncertainties related to $t\bar{t}$Z

Various theoretical uncertainties related to the $t\bar{t}Z$ modelling have been taken into account. Each of them is described by one nuisance parameter.

- Renormalization scale variations by a factor of 2 and 0.5.

- Factorization scale variations by a factor of 2 and 0.5.
- Variations of both factorization and renormalization scales consistently by a factor of 2 and 0.5.
- Variation of weights related to parton distribution functions. The NNPDF3.0 [94] pdf set has been used as a nominal. NNPDF3.0 provides a set of 100 weights, each of them obtained from a fit to data. The nominal weight is obtained as the arithmetic average of the 100 weights and its uncertainty is considered as the pdf systematics uncertainty.
- Variation of Var3 parameter of A14 tune, handling the gluon radiation during the parton showering and hadronization [103].
- Choice of the matrix element event generator by replacing the nominal aMC@NLO [101] sample by an alternative Sherpa 2.2.1 sample [104].

Only shape and acceptance effects are taken into account in the theoretical uncertainties related to the $t\bar{t}Z$. The effect of systematic uncertainties on the total cross section has been dropped, since the cross section and its uncertainty are extracted from the fit.

6.2.3 Uncertainties related to the Z+jets background

The following scale variations have been taken into account for Z+jets samples, each of them is described by one nuisance parameter:

- Renormalization scale variations by a factor of 2 and 0.5.
- Factorization scale variations by a factor of 2 and 0.5.
- Variations of both factorization and renormalization scales consistently by a factor of 2 and 0.5.
- Variation of CKKW scale from nominal 20 GeV, down to 15 GeV and up to 30 GeV.
- Variation of QSF scale by a factor of 2 and 0.5.

The 10 % cross section uncertainty on the Z+light component has been considered. The normalizations of Z+HF components are set as free parameters of the fit (will be discussed later). The normalization effects of scale uncertainties have been dropped for Z+jets in order to avoid correlations with Z+HF normalizations.

The systematic uncertainties for the Z+jets matrix element and parton showering have not been considered, since the difference between the alternative and nominal samples were not statistically significant. A good agreement between data and simulation have been observed in Z+jets control regions (regions with low MVA output, described later in Section 8), suggesting reasonable Z+jets modelling by the nominal Sherpa samples.

6.2.4 Uncertainties related to tWZ

The tWZ is an important background in the 4ℓ channel, while its contribution to the 2ℓ channel is very small. In the 2ℓ channel it was merged with other small backgrounds and uncertainty ± 50 % was assigned to its normalization. In the 4ℓ channel, the following theoretical uncertainties related to the tWZ are considered, each of them is described by one nuisance parameter:

- Renormalization scale variations by a factor of 2 and 0.5.
- Factorization scale variations by a factor of 2 and 0.5.
- Variations of both factorization and renormalization scales consistently by a factor of 2 and 0.5.
- Variation of weights related to parton distribution functions. Their estimation is similar to the $t\bar{t}Z$ case.

6.2.5 Uncertainties related to ZZ background

The ZZ is an important background in the 4ℓ channel. In the 2ℓ channel it was merged with other diboson samples and uncertainty ± 50 % was assigned to its normalization. In the 4ℓ channel, the following theoretical uncertainties related to the tWZ are considered, each of them is described by one nuisance parameter:

- ± 50 % normalization uncertainty has been assigned to the cross section of ZZ in association with at least one b-jet.
- Variation of weights related to parton distribution functions. Their estimation is similar to the $t\bar{t}Z$ case.

6.2.6 Uncertainties related to the fake lepton background in the 4ℓ channel

The Matrix Method described in Sec. 5.3.1 is used as the nominal method for the fake lepton estimate in the 4ℓ channel. The Fake Factor method described in Sec. 5.3.2 is used as an alternative approach and the difference between the two predictions is taken into account as the systematic uncertainty of the fake lepton background estimate.

6.2.7 Uncertainties related to less significant backgrounds

The contribution of other processes are significantly smaller than the contributions from Z+jets, $t\bar{t}$ and $t\bar{t}Z$. The uncertainties related to the cross section of these processes have been taken into account.

The uncertainty on the $t\bar{t}H$ cross section has been set to $^{+5.8\ \%}_{-9.2\ \%}$ (scale) and ± 3.6 % (pdf) [110].

The uncertainty on the tZ cross section has been set to ± 30 %.

The ± 50 % uncertainty has been assigned to cross section of the diboson background in the 2ℓ channel. In the 4ℓ channel the corresponding diboson uncertainties, described in Sec. 6.2.6, are applied.

All the other processes have been merged together and ± 50 % uncertainty has been assigned to their normalization. This include also fake lepton contribution in the 2ℓ channel.

Chapter 7

Analysis Methods

7.1 Multivariate analysis

7.1.1 Motivation

In high energy physics, it is often necessary to design a selection which suppresses a background while the number of signal events is kept reasonably high. There are two, different, requirements on the selection. The first one is to make the selection criteria tight in order to reduce the background rate and the other one is to make them loose in order to keep a reasonable number of signal events passing the selection. The problem rises an obvious question: How tight selection should be used? The answer strongly depends on specifications of the analysis and whether it is dominated by a systematic uncertainty or not.

A good choice of selection criteria can significantly improve a statistical sensitivity of an analysis. The analyser has to understand the composition of the background in his analysis and differences with respect to the desired signal, so he can design selection criteria providing a good signal to background separation. In some cases simple cuts can be applied in order to get a signal region with a reasonably high number of events and signal purity. When the simple cuts cannot help that much, it is necessary to use more sophisticated approach, a multivariate analysis.

For a simplicity let's assume an analysis measuring the cross section of $Z{+}b\bar{b}$ using events with the leptonically decaying Z boson. The typical signature of the final state are two leptons of the opposite sign and same flavour and two b-jets. The obvious Standard Model background with this signature is the dileptonically decaying top-quark pair. Both processes provide the same particles in the final state, except of neutrinos which cannot be measured directly. Both have two b-jets and the opposite sign lepton pair in the final state. In one half of the cases, the flavour of the leptons from $t\bar{t}$ is the same (τ-s are not assumed). However, there are some differences in the kinematics. The distribution of the dilepton invariant mass has a peak around $M_Z =$ 91.2 GeV for Z+$b\bar{b}$ events, while in the case of top-quark background the distribution is significantly wider and without any clear peak. This fact can be used in order to define the signal region with a low $t\bar{t}$ background contamination.

Requirement on the dilepton invariant mass to be inside the Z-mass window, which means $|M_{\ell\ell} - M_Z| < x$, can significantly suppress the $t\bar{t}$ background rate. The choice of x depends on a desired signal purity, let's use $x = 10$ GeV in this case. The distributions of dilepton invariant mass for signal (red) and background (blue) together with cut values (black dashed lines) are shown in Figure 7.1

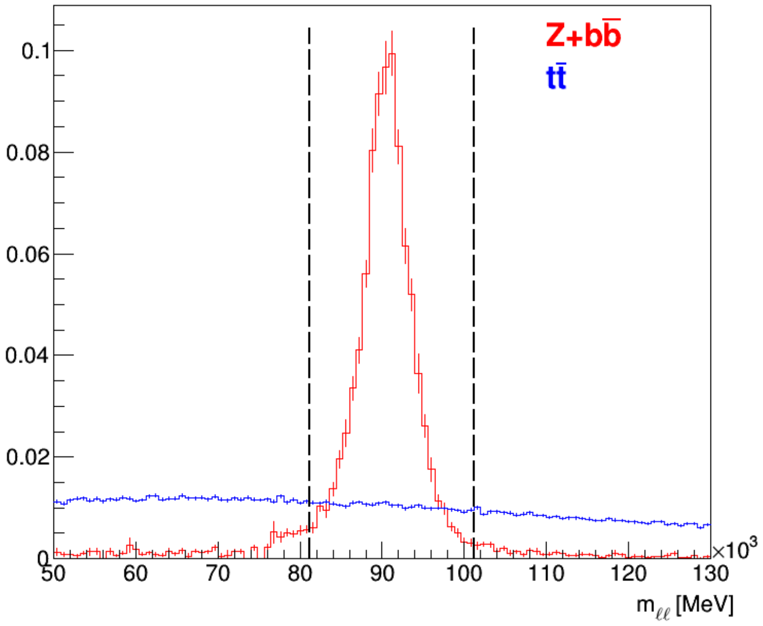


Figure 7.1: Distribution of dilepton invariant mass for Z+$b\bar{b}$ events (red) and $t\bar{t}$ events (blue). The black dashed lines show cuts defining the Z-mass window. Only the events with $M_{\ell\ell}$ in the Z-mass window are accepted.

Although the $t\bar{t}$ rate after the cut is significantly lower, still it is not negligible. The question is: What else can be done in order to suppress the $t\bar{t}$ even more? There are more options, one of them is to cut on the missing transverse energy (MET). In the case of the Z+$b\bar{b}$ there is no physical MET (no neutrinos), while in the case of $t\bar{t}$ events there are two neutrinos in the final state which can be measured as the MET in the detector. Since the MET resolution of the detector is not ideal, a non-zero value of MET can be expected also for $Z{+}b\bar{b}$. The maximal requirement on the MET can be used in order to suppress the $t\bar{t}$ background and get more pure $Z{+}b\bar{b}$ signal region. Looking at the MET distributions in Figure 7.2, the MET < 50 GeV cut can be used in order to cut off a significant part of the background while only a small part of signal events is rejected.

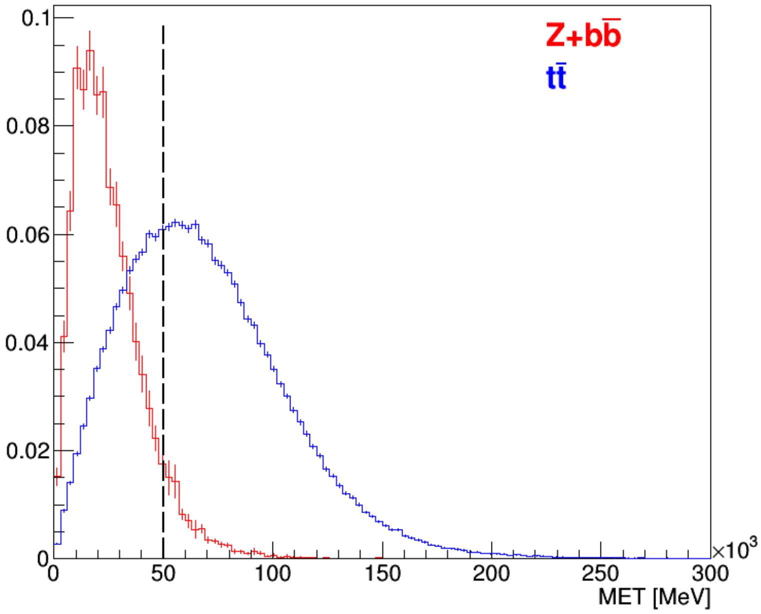


Figure 7.2: Distribution of missing transverse energy for $Z{+}b\bar{b}$ events (red) and $t\bar{t}$ events (blue). The black dashed lines show cuts defining the MET < 50 GeV region.

Let's look at these cuts in two dimensions. The requirements $|M_{\ell\ell} - M_Z| < 10$ GeV and MET < 50 GeV are shown on the top of Figure 7.3. The $Z{+}b\bar{b}$ events are shown in red, the $t\bar{t}$ are shown in blue. It can be seen that the cuts target the region with high signal purity. However, the rectangular shape of the region is not the best choice. In the top right corners, around the points $M_{\ell\ell} = 81.2$ (or 101.2) GeV and MET $=$ 50 GeV, the signal purity is lower than in the middle of the upper edge (around the point MET $= 50$ GeV and $M_{\ell\ell} = 91.2$ GeV). Using some kind of a curved border of the region, the signal purity could be increased while the expected statistics would be kept the same. A suggestion of a such region can be seen on the bottom of Figure 7.3. The obvious questions arises: Is the proposed border the best one, if we want to keep the number of events the same, or could it be improved even more? Can we use some algorithm to build such borders? Would it be possible to build a single variable from the set of input variables (in this case MET and $M_{\ell\ell}$) and to define the signal region cutting on this variable? The answer is: "Use Multivariate Analysis!".

7.1.2 Introduction

There are many kinds of Multivariate Analyses (MVA). The MVA in general is an algorithm that is able to combine a set of input variables into one output variable which is related to the probability for the given event to be a signal.[1] The higher the output is, the higher is the probability that the event belongs to the signal. In other words, the output from the MVA is a measure of the event to look like a signal event.

The obvious and the best choice of such algorithm is to use directly the probability of the event to be the signal. If differential cross-sections of signal and background processes as functions of input variables $\vec{x}$ are know, the following equation can be used to get the probability.

$$P(\vec{x}) = \frac{\sigma^{\mathrm{signal}}(\vec{x})}{\sigma^{\mathrm{signal}}(\vec{x}) + \sigma^{\mathrm{background}}(\vec{x})}, \tag{7.1}$$

The problem is, that the differential cross section is often unknown and only a Monte Carlo simulation with a limited statistics is available. If the number of available signal and background events is high and number of input variables is low, one can approximate the differential cross-section by discrete n-dimensional histogram with a finite binning, where n is the number of the input variables. The choice of the binning depends on the available statistics and it is limited by statistical fluctuations in the individual bins. The average number of events in a bin decreases exponentially with the number of input variables. As a consequence, this approach can be hardly used for more than two or three input variables.

If the low number of input variables is not enough, it is necessary to use a more sophisticated algorithm, which is able to estimate the event probability to be a signal with the limited statistics while a higher number of the input variables is considered. Two kinds of such algorithms, Artificial Neural Network (NN) and Boosted Decision

---

[1] In general, the output does not have to be only one number. In the problem of multiclassification, the MVA can have more outputs. Each of them corresponds to a different output class. For example if we had one signal and two background processes, we could use an MVA with 3 outputs. The first one would be related to the probability to be the signal, the second one would be related to the probability to be the first background and the third output would be related to the other background process. In this analysis we do not use this kind of MVAs.

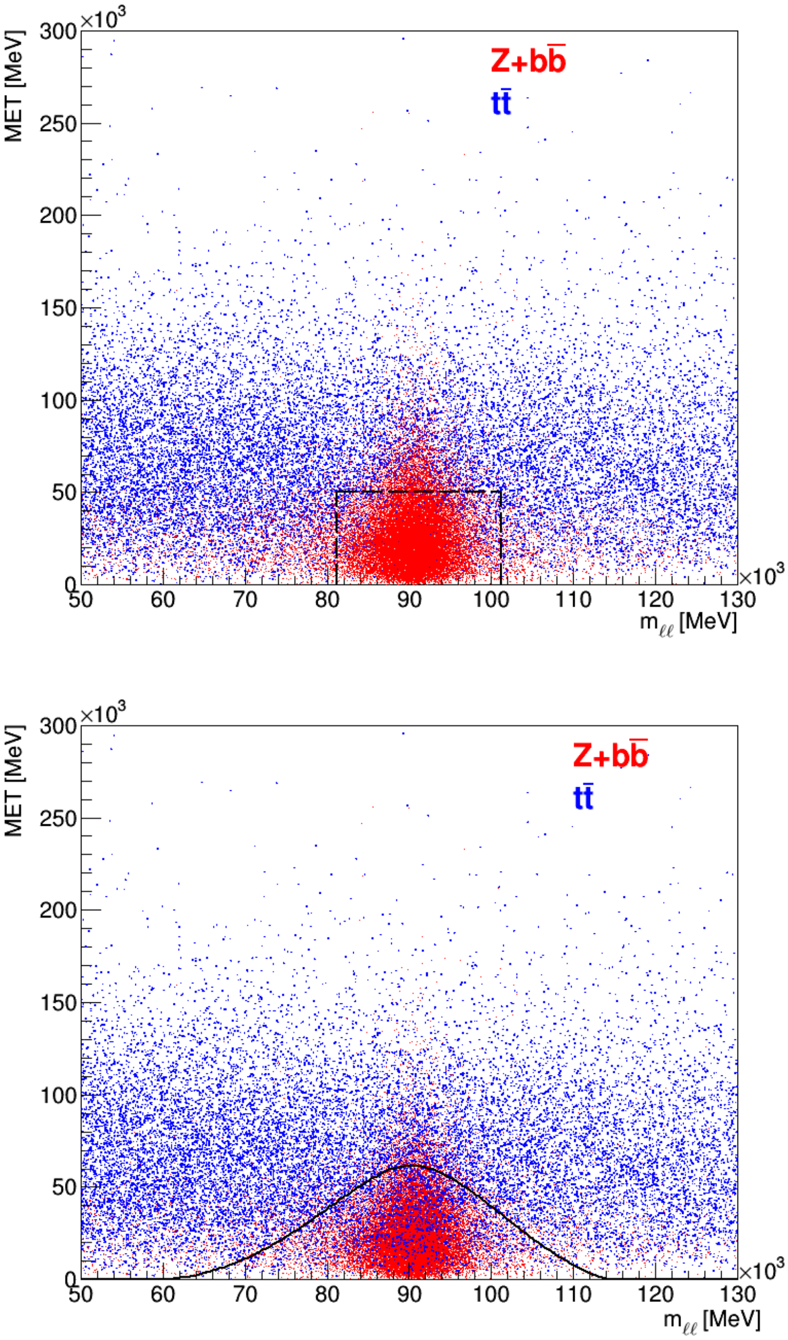


Figure 7.3: Z+$b\bar{b}$ (red) and $t\bar{t}$ events (blue) showed in 2D scatter plot in $m_{\ell\ell}$ vs. MET plane. Two approaches to signal regions definitions are shown: simple cuts (top) and cutting on output from Neural Network (bottom).

Tree (BDT), will be discussed later in this chapter. Even if the MVA output is closely related to the probability to be a signal, usually it is not directly the probability.

When using the MVA, usually two phases are needed: training and application phase. During the training phase, a known set of events (usually a Monte Carlo simulation of the signal and background) is provided to the MVA. Information about the real type (class) of the event is provided, so the MVA can learn on the sample how the signal and background events look like. The technical details of the training strongly depend on a type of the MVA and will be discussed later for NN and BDT. Once the MVA is trained, it is ready to classify an unknown set of events, which is called the application phase.

Overtraining and choice of the input variables

In the ideal case, if the statistics of the training set of events is sufficiently high, the MVA learns only real, physical, differences between signal and background samples and it is not affected by statistical fluctuations. In practice we always have a limited statistics and therefore there are not only real, physical, differences between the signal and background sets, but also random differences caused by the statistical fluctuations, which are characteristic for the given set of events. The effect of MVA being trained to the statistical fluctuations of the training sample is called the overtraining.

Because of the overtraining, the MVA performance on the training set is usually better compared to a statistically independent sample, resulting in a better signal vs. background separation in training dataset. If the difference in the MVA response between the training and the statistically independent sample is large, the MVA is considered to be overtrained. In order to avoid a possible bias caused by the overtraining, the training set of the events cannot be used in the analysis itself to estimate the expected MVA output distribution. For this purpose, a statistically independent sample has to be used.

There are a few ways how to avoid, or at least to reduce, the overtraining. One of them is to increase the statistics in the training set. This is often impossible because of limited computational resources or/and a limited storage space. Another way of the overtraining reduction is to decrease the number of input variables. The lower the number of input variables is, the smaller is the space that has to be exploited by the MVA. A higher number of input variables usually leads to a higher signal to background separation of the MVA output, but on the other hand, the overtraining is also higher. Choice of the number of MVA input variables is usually a trade off between a reasonable signal to background separation and low overtraining.

There is also another, important, reason to keep the number of MVA input variables low, even if a large statistics is available and thus the overtraining is small. When training on the Monte Carlo events, one has to be sure that the simulation describes the data well. If there is a mismodelling, it will be propagated into the expected shape of the MVA output distribution estimated from the Monte Carlo. This would bring a bias into the analysis. The first thing to check is a modelling of individual input variables. In other words, the distribution functions of the input variables $p_i(x_i)$ have to be modelled properly. This can be easily checked from control plots. However, it is necessary to keep in mind that this is not the end of the story. Even if the modelling of the individual variables is proper, they are not independent, it means $p(\vec{x}) \neq \Pi_{i=1}^{n} p_i(x_i)$, where x_i are components of the vector $\vec{x}$ and $\Pi_{i=1}^{n}$ stands for product over all $p_i(x_i)$ terms. Also the n-dimensional distribution function has to be well modelled in the MC. Since the statistics is limited, there is not any way how to verify the $p(\vec{x})$ modelling. In general, adding more input variables increases a risk of $p(\vec{x})$ being mismodelled and therefore it should be avoided if possible.

When the number of input variables is going to be reduced, it is necessary to decide which variables should be kept and which should be removed. There are few algorithms for this task. The simplest of them is to look at the distributions of individual variables and keep those with the most different shapes between the signal and background. As a measure of the differentness, the separation power is usually used, which is defined by:

$$S = \frac{1}{2} \cdot \int \frac{(y_i(x_i)^{sig} - y_i(x_i)^{bkg})^2}{y_i(x_i)^{sig} + y_i(x_i)^{bkg}} dx, \tag{7.2}$$

where $y_i(x_i)^{sig}$ and $y_i(x_i)^{bkg}$ are probability density functions of the variable x_i for signal and background, normalized to unity. The separation power is equal to zero for identical distributions and equal to one for different non-overlaping distributions. If the analytical functions $y_i(x_i)^{sig}$ and $y_i(x_i)^{bkg}$ are not known, and only a MC simulation is available, a discrete form of Eq. 7.2 has to be used, using histograms instead of analytical functions $y_i(x_i)$. The discrete form of the Eq. 7.2 is:

$$S = \frac{1}{2} \cdot \sum_{i}^{bins} \frac{(N_i^{sig} - N_i^{bkg})^2}{N_i^{sig} + N_i^{bkg}}, \tag{7.3}$$

where the sum runs over all bins and N stands for the number of events in the bin.

When choosing the set of input variables, they can be ordered by the separation power and their subset from the top of the ranking can be used. This approach has an important disadvantage, it does not take into account correlations between the variables, which can be often important.

Let's consider set of n input variables $\vec{x} = (x_1, ..., x_n)$, where their ordering is given by their separation decreasingly. Let's assume that $x_3 = (x_1 + x_2)$. If, for example, only first 3 variables are used, one of them is redundant and does not bring any new information. In this case the correlation between the variables would have an negative effect on the MVA performance. It would be better to remove the third variable and to take the fourth instead.

To give an example of positive effect of correlation, let's assume again the set of input variables $\vec{x} = (x_1, ..., x_n)$ ordered by their separation. Let's assume that the variable x_n is uniformly distributed in the interval [0,1] in both signal and background. Since the distributions in background and signal are the same, the separation is equal to zero. Such a variable seems to be useless at the first glance, but it does not have to be. Let's assume variable x_{n-1}, which is defined in the following way: $x_{n-1} = x_n$ (for the signal), $x_{n-1} = 1$ - x_n (for the background). Since x_n has uniform distribution in [0,1], x_{n-1} is also uniformly distributed in this interval for both signal and background. Its separation power is therefore zero. Although the individual separation power of both variables is zero, their combination provides the ideal separation, since there is a clear difference in the relation between them in the signal and background. In this case the relation can be simplified to the correlation, which is +100% in signal and -100% in background. In general, an MVA is able to search for even more complicated relations of input variables.

The ideal way of variable choice would be to train with all possible combinations of input variables, check the separation power of the MVA output and then take the set of input variables providing the highest separation power, or another measurable quantity, for example expected sensitivity of the fit, area under the ROC curve[2]. Although this approach is the best option, it would be very CPU intensive, since it would require to repeat the training and application phase at least $n!$ times. Therefore some approximate variable ranking algorithms have been developed. They usually depend on the type of the MVA, so they will be discussed later in the NN and BDT chapters.

---

[2] ROC (receiver operating characteristic) curve is a dependency of the signal efficiency on the background rejection (fraction of rejected background events). For a variable without any separation, the ROC curve is the straight line between points [0,1] and [1,0], let's call this line the diagonal. The ROC curve of any MVA with a discrimination power will be above the diagonal. The larger is the area bellow the ROC cure, the better separation is provided by the MVA output. The difference between areas under ROC curves built from training and testing sets can be used as a measure of the overtraining.

Input variables transformation

Various training algorithms expect values of the input variables to be in a specific range. If the input variables are far from this range, the training algorithm may not work, or needs significantly more time to converge. In order to overcome such problems, input variables transformations are used. The simplest of them is a variable normalization. The variable is transformed according to a linear function in order to get a variable from the specific range, for example the mean value is subtracted and the result is divided by RMS of the initial distribution. At the output of this procedure, the variable with mean value 0 and RMS 1 is obtained. If the training algorithm works better with certain distribution of input variable (for example Gaussian, or uniform), another transformation follows in order to get variable with the desired distribution.

Another, more sophisticated, input variable transformation is a decorrelation. The input variables covariance matrix C is calculated. Diagonalizing the covariance matrix, the matrix $D = S^T C S$ is calculated. The matrix $(S\sqrt{D}S^T)^{-1}$ is then used for the variable decorrelation and normalization to the unity variance [86]:

$$\vec{x} \rightarrow (S\sqrt{D}S^T)^{-1}\vec{x}, \tag{7.4}$$

where elements of the matrix $\sqrt{D}$ are square roots of elements of matrix D.

Cross training

As discussed before, the training set of events cannot be used in the rest of the analysis, because the overtraining could bring a bias to the analysis. One possibility of avoiding the bias is to use one half of MC events only for the MVA training and use the other half in the rest of the analysis. This would mean to loose one half of the Monte Carlo events, which would lead to a higher MC statistical uncertainty.

In order to be able to use the whole MC statistics, a cross-training can be used. The Monte Carlo is split, usually into two subsets of events. The difference with respect the former approach is that two MVAs are going to be used. The first of them is trained on the first subset and it is used to classify the other subset. The other MVA is trained on the second subset and then used to classify the first one. In this way the whole MC statistics can be used. If running on an independent sample (for example on data), the sample has to be split again into two parts, so both MVAs will be used.

The splitting mode has to be independent on the physics in order to guarantee the same MVA response on the training and testing sets. Usually an event number (a unique integer number identifying the event) is used and events are split into two halves based on the even or odd value of this number. A random number can be used as well, but in this case it is difficult to reproduce the result. In our analysis the event number is used.

7.1.3 Artificial Neural Network

The idea of an Artificial Neural Network was motivated by an attempt to model information processing in a nervous system. Human brain is composed of "computational units", neurons, and connections between them, synapses. In biological neural networks, information is stored in the synapses connecting the neurons. The artificial NNs work in a similar way.

Each node has input and output connections, called synapses. The node (neuron) takes values from the input synapses, performs a mathematical computation and sends an output value to the output synapses. Each neuron is characterized by its integration function and an activation function. The integration function reduces number of input arguments to a single numerical value. This value is used as an input for the activation function. Output value calculated by the activation function is propagated to the output synapses of the neuron. Weighted sum of input values, $f(x_1, ..., x_n) = w_1 x_1 + ... + w_n x_n$, is usually used as the integration function, where x_i are outputs of previous neurons propagated through the synapses and w_i are weights of the synapses. These weights are optimized during the training phase. Sigmoid, $y(t) = 1/(1 + e^{-t})$, or hyperbolic tangent, $y(t) = (e^t - e^{-t})/(e^t + e^{-t})$ functions, are usually used as the activation function, but in principle any continuous and differentiable function can be used.

An example of Neural Network structure is shown in Figure 7.4. The NN consists of nodes and synapses. Nodes are organized into layers. The Neural Network on Figure 7.4 has one input layer, one middle (so-called hidden) layer and one output layer. The Neural Network can have various number of nodes in a layer and also various number of layers. A Neural Network with many hidden layers is called Deep Neural Network (DNN). It is able to learn complicated dependences between input variables. DNNs are usually used with a large set of input variables when individual separation power of input variables is low. These variables are called low-level variables. A good example of low level variables are fourmomenta of particles/objects in an event. A DNN itself is able to build more powerful (high-level) variables, such as $M_{\ell\ell}$ and MET in our model case. Typical disadvantage of DNNs is a need for a large size of the training dataset. The other type of Neural Network is Shallow Neural Network, which has a low number of hiden layers (one or two). Neural Networks are also frequently used in pattern recognition to identify a text or objects from an image. Shallow Neural Network does not require as high number of training events as DNN does, but on the other hand its ability to work with low-level variables is significantly worse. Because of a small number of hidden layers and neurons, it is usually not capable of finding complicated dependencies between a set of low-level variables.

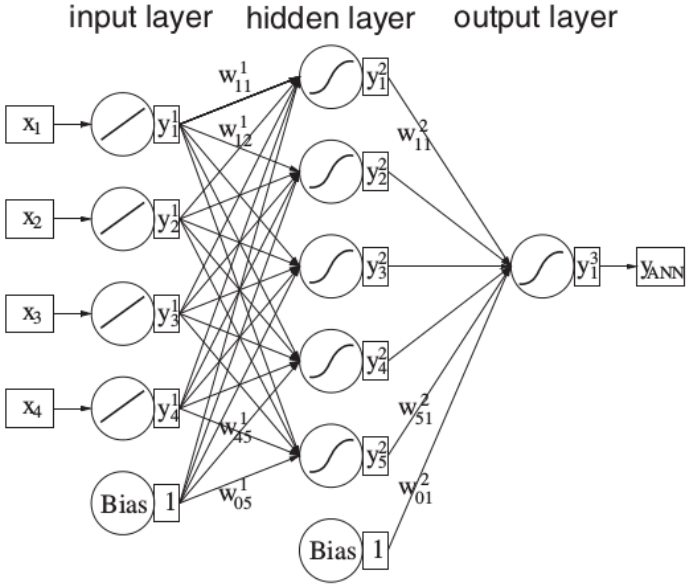


Figure 7.4: Architecture of a Neural Network with a single hidden layer [111].

Loss function and training

The set of synapses weights $\vec{w}$ is being optimized during the training phase, in order to minimize a loss function. The loss function is a measure of MVA ability to correctly classify events. Various kinds of loss function might be used. The simplest case is the sum of squared errors defined as:

$$L(\vec{w}) = \sum_{i}^{\text{events}} (y_i(\vec{w}) - y_i^{\text{truth}})^2, \tag{7.5}$$

where the sum runs over all events, $\vec{w}$ is the vector of synapses weights, y_i^{truth} is a true value target of the event i (usually 1 for signal and 0 (or -1) for a background), y_i is the output from the MVA obtained for the event i. The square loss function is a popular choice for some MVAs, but it is not an ideal choice for Neural Network, because it makes training algorithms to converge slowly. Cross-entropy is a more appropriate choice for Neural Networks and it is defined as follows [111]:

$$L(\vec{w}) = \sum_{i=1}^{\text{events}} [y_i^{\text{truth}} ln(y_i(\vec{w}) + (1 - y_i^{\text{truth}}) ln(1 - y_i(\vec{w}))], \tag{7.6}$$

In order to motivate the choice of loss function, let's assume that we want the MVA output to be the probability of the event to be signal, which means:

$$P_s(\vec{x}, \vec{w}) = y(\vec{x}, \vec{w}) \tag{7.7}$$

$$P_b(\vec{x}, \vec{w}) = 1 - P_s(\vec{x}, \vec{w}) = 1 - y(\vec{x}, \vec{w}), \tag{7.8}$$

where $P_s(\vec{x}, \vec{w})$ and $P_b(\vec{x}, \vec{w})$ stand for the probability of the event being the signal/background, $\vec{x}$ is a vector of input variables and $\vec{w}$ is the vector of synapses weights. The Equations 7.7 and 7.8 can be rewritten as:

$$P(y^{\text{truth}}, \vec{x}, \vec{w}) = y(\vec{x}, \vec{w})^{y^{\text{truth}}} [1 - y(\vec{x}, \vec{w})]^{(1 - y^{\text{truth}})} \tag{7.9}$$

Therefore the probability for the whole set of events is:

$$L(\vec{w}) = \Pi_i^{\text{events}} \left(y_i(\vec{w})^{y_i^{\text{truth}}} [1 - y_i(\vec{w})]^{(1 - y_i^{\text{truth}})} \right) \tag{7.10}$$

Calculating $ln(L(\vec{w}))$, the Equation 7.6 can be obtained.

Training

Minimization of a loss function is a complicated problem of finding the minimum of the many dimensional loss function. Various numerical methods are used in order to find the minimum, where gradient of the loss function often plays the leading role.

The simplest way of finding the minimum is iterative gradient descent method. The minimum can be found starting from a random point and moving against the gradient of the loss function. In this case, the synapses weights are updated as follows

$$\vec{w}_{n+1} = \vec{w}_n - \gamma \nabla L(\vec{w}_n), \tag{7.11}$$

where γ is a learning rate, which is a free parameter of Neural Network and its appropriate choice is analysis dependent.

In order to increase the convergence speed of a learning algorithm and to avoid ending up in a local minimum, a so-called momentum term can be used [112]. When using the momentum, the weights of the network are updated according to Equation 7.12. The direction of the weight update is a linear combination of the gradient from the current step and from the previous step.

$$\vec{w}_{n+1} = \vec{w}_n - \gamma \nabla L(\vec{w}_n) - \alpha \nabla L(\vec{w}_{n-1}), \tag{7.12}$$

where $0 < \alpha < 1$ is the momentum rate. The momentum can significantly decrease a number of steps necessary to reach the minimum, as shown in Figure 7.5

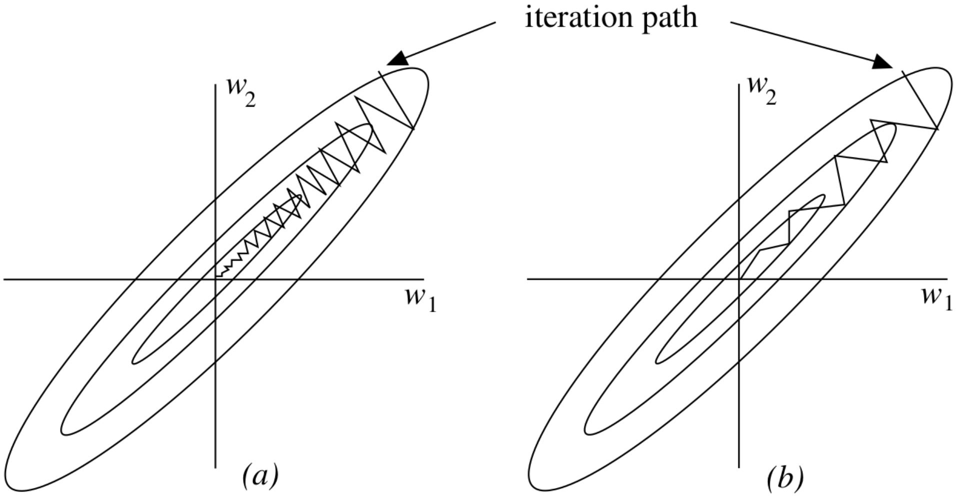


Figure 7.5: Finding a loss function minimum without the momentum term (a) and with the momentum term (b) [112].

The gradient of the loss function is usually not calculated from the whole training sample, but only from a subset of events. In a new iteration a new subset of events is used. The advantage of this approach is that since the calculation of the gradient is not exact, the algorithm is more robust against converging to a local minimum of the loss function caused by a statistical fluctuation of the training dataset. The number of events in the subsample is usually referred to as batch size.

Backpropagation

The so-called backpropagation is used in order to find the gradient of the loss function. When the Neural Network output is evaluated, the input variables are provided to the input nodes and the values are propagated forward through the network (from the input nodes to the output node). When calculating the loss function gradient, the values are propagated backwards (from the output node to the input nodes).

Since the loss function is calculated as a sum of contributions from individual events, the gradient of the loss function can be also calculated as the sum. Let's consider ij component of the gradient, corresponding to the weight of the synapse with input node i and output node j (see Fig. 7.6). An individual contribution of an event to the loss function will be referred to as "error function", E, in the rest of this chapter.

$$\frac{\partial L(\vec{w})}{\partial w_{ij}} = \sum_{k}^{\text{events}} \frac{\partial E(\vec{w})_k}{\partial w_{ij}} \tag{7.13}$$

Let's consider Neural Network with one output node and all nodes using the same activation function $net(t)$ with continues first derivative. A generalization to different types of activation function is trivial and will be discussed later. The partial derivative of the error function E with respect to the weight w_{ij} can be written as

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial (net(x_j))} \frac{\partial (net(x_j))}{\partial x_j} \frac{\partial x_j}{\partial w_{ij}}, \tag{7.14}$$

where:

$\frac{\partial E}{\partial (net(x_j))}$ is the partial derivative of the error function with respect to the output of the j-th node.

$\frac{\partial (net(x_j))}{\partial x_j}$ is the derivative of the j-th node activation function (with respect to its integration function output). Since the activation function, $net(x_j)$, is usually analytical and simple (for example sigmoid, or hyperbolic tangent), the calculation is trivial. The derivative will be marked as $net'(x_j)$.

$\frac{\partial x_j}{\partial w_{ij}}$ is the partial derivative of the j-th node integration function with respect to the weight w_{ij}. For the most common choice of the integration function, $f_j(\vec{x}) = \sum_i w_{ij} net(x_i)$, we get $\frac{\partial x_j}{\partial w_{ij}} = \frac{\partial w_{ij} net(x_i)}{\partial w_{ij}} = net(x_i)$, which is the output of the i-th node.

The synapse corresponding to the w_{ij} weight and nodes around are shown in Figure 7.6.

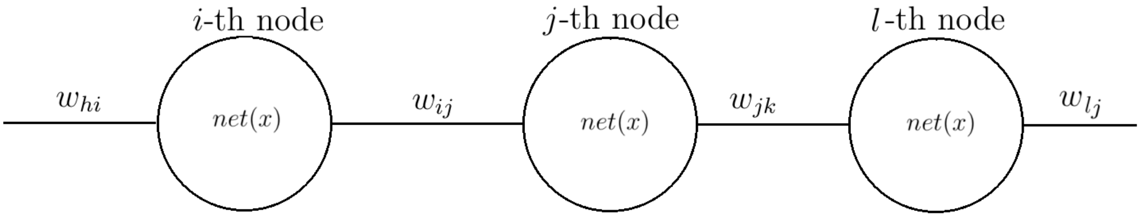


Figure 7.6: The part of neural network that we will use to derive a backpropagation algorithm.

Rewriting the Equation 7.14, the following Equation can be obtained:

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial (net(x_j))} \times net'(x_j) \times net(x_i), \tag{7.15}$$

The calculation of second and third terms in Equation 7.15 is trivial and can be performed during the forward propagation. The first term, corresponding to the partial derivative of the error function E with respect to the output from j-th node, is more complicated:

$$\frac{\partial E}{\partial net(x_j)} = \frac{\partial E}{\partial (net(x_l))}\frac{\partial (net(x_l))}{\partial (net(x_j))} = \frac{\partial E}{\partial (net(x_l))}\frac{\partial (net(x_l))}{\partial x_l}\frac{\partial x_l}{\partial (net(x_j))}, \tag{7.16}$$

where:

$\frac{\partial E}{\partial (net(x_l))}$ is the partial derivative of the error function with respect to the output of the l-th node

$\frac{\partial (net(x_l))}{\partial x_l}$ is the derivative of the l-th node activation function (with respect to its input), similar to the second term in Equation 7.14

$\frac{\partial x_l}{\partial (net(x_j))}$ is a partial derivative of the l-th node integration function with respect to the j-th note output. Since the integration function sums over all the inputs with corresponding weights, the derivative is equal to the weight of the connecting synapse, w_{jl}.

Rewriting the Equation 7.16, we get the recursive algorithm to calculate the derivative of the error function with respect to the output of a node, if the derivative of its neighbour node from the next layer is known:

$$\frac{\partial E}{\partial net(x_j)} = \frac{\partial E}{\partial (net(x_l))} \times net'(x_l) \times w_{jl}, \tag{7.17}$$

When calculating the error function derivative, the following steps are necessary for each event:

1. The values of input variables are propagated forward through the network. Output of each node is stored and also derivative of its activation function in the point of its current input is calculated and stored. As the result of feed-forward propagation, the NN output, y_{NN} is obtained.

2. The derivative of the error function with respect to the NN output is calculated for the current value of output and this derivative is set to the output of the output node.

3. The value of the derivative is propagated backwards (considering Eq. 7.17). The backpropagation is similar to forward propagation with an important difference.

When propagating forward, the output of integration function is provided to the activation function and output of the activation function is propagated through the output synapses. In the backpropagation, the output of the integration function is multiplied by the value of the activation function derivative (which was stored), instead of using the activation function itself. The backpropagation algorithm is illustrated in Figure 7.7.

4. The backpropagation from the previous step allows us to access the error function derivative with respect to the outputs of all nodes. From the forward propagation, we still have the information about output and value of activation function derivative for each node. Using Equation 7.15, we can calculate the error function derivative with respect to weight of any synapse.

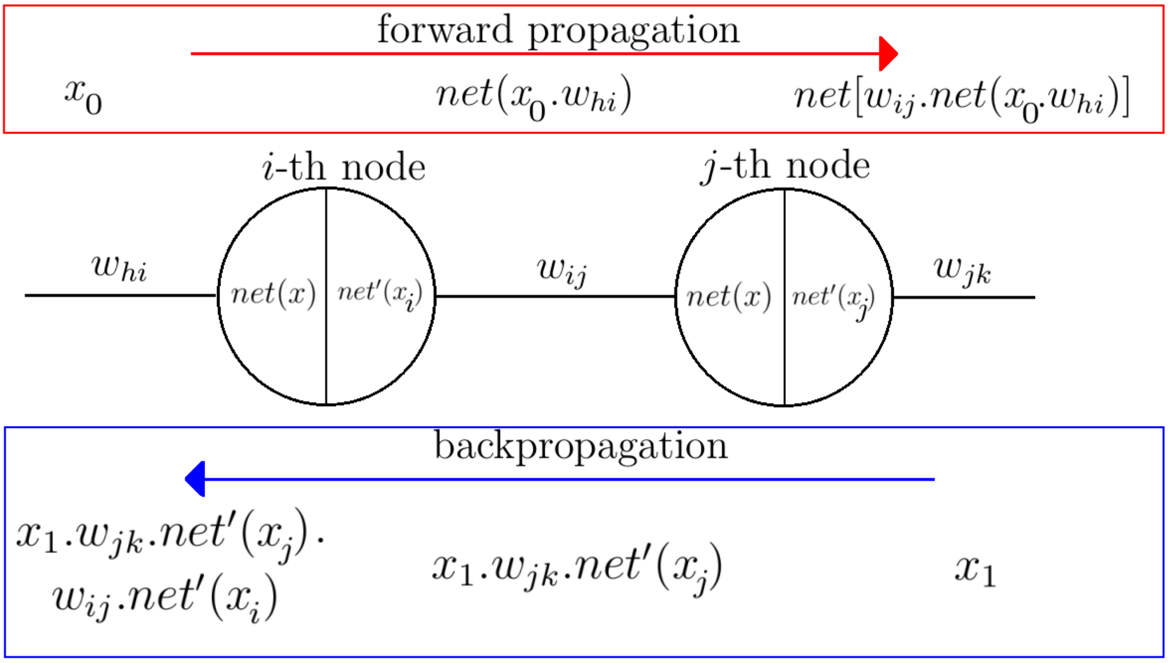


Figure 7.7: Illustration of forward-propagation (top) and backpropagation (bottom). When propagating forward through the network (left → right), each node calculates output its integration function from its input values, uses it as the input for its activation function, and the activation function output is propagated to the next node. In the backpropagation, instead of the activation function, the output of the integration function is multiplied by the value of activation function derivative at the point of the last input.
In our case, $net'(x_i) = net'(x)|_{x=x_0.w_{hi}}$ and $net'(x_j) = net'(x)|_{x=w_{ij}.net(x_0.w_{hi})}$

.

Once the gradient of the loss function is calculated, the network weights are updated according to Equation 7.12. Some neural networks use additional regularization or noise terms in the weight update formula in order to reduce the overtraining or speed up the convergence in a case of low values of the gradient.

Variable ranking

There are multiple approaches to the variable ranking in Neural Networks.

The multilayer perceptron (MLP), the Neural Network implemented in TMVA package [113] in ROOT [114], uses an approach based on the weights of synapses connecting the input variable node from the input layer to the first hidden layer. The higher the sum of squares of these weights, the higher the variable importance. Another approach is to use derivative of the NN output with respect to the input variable, calculate its mean over the training dataset and order the variables by the mean value of the derivative.

The NeuroBayes [115] Neural Network, which is used in this analysis, uses method independent ranking (can be used for any MVA) of the input variables based on the total correlation of the set of variables to the target. The total correlation of set of variables $\vec{x}$ to the target variable y is a measure of shared information between $\vec{x}$ and y:

$$C(\vec{x} \leftrightarrow y) = \int\int p(\vec{x}, y) ln \left(\frac{p(\vec{x}, y)}{p(\vec{x})p(y)} \right) dy d\vec{x}, \tag{7.18}$$

where $p(\vec{x}, y)$, $p(\vec{x})$ and $p(y)$ are distribution functions. If there is no dependency of the target on the input variables, $p(\vec{x}, y) = p(\vec{x})p(y)$ and the total correlation is zero. However, a dependency of the target on the input variables will cause non zero value of the total correlation.

NeuroBayes ranking algorithm calculates the total correlation of the whole set of input variables to the target. Then the algorithm loops over all variables and calculates the total correlation of the rest of the input variables to the target. The variable which removal causes the smallest loss in the total correlation is considered to be the least important variable. It is removed and the algorithm loops again over the remaining variables, looking for the second least significant variable. The algorithm is repeated until only one input variables remains, that is considered to be the most important one. The variable removed right before it is considered to be the second and so on.

7.1.4 Boosted Decision Tree

The Boosted Decision Tree (BDT) is based on a decision tree algorithm, which is illustrated in Figure 7.8.

The decision tree can be used as a binary classifier, which means its output is either 0 (possibly -1) or 1 if an event is background-like or signal-like, as well as regressor, which means that its output is used to predict value of a continues target variable. In our analysis, the decision tree is used for the classification. The differences between classification and regression trees will be discussed later in this chapter, since the regression tree is used also in a training algorithm of classification trees.

Classification Tree

The decision tree is series of simple cuts, each of them splitting a sample into two subsamples, based on true/false value of the cut condition. In Figure 7.8, the whole sample is split into two parts at the "Root node". If $x_i > c_1$, the event belongs to the subsample on the left, otherwise it belongs to the right subsample. The aim of the cut is to enhance the signal fraction in one subsample and to suppress it in the other. For each subsample a further cut is applied in order to reach even better signal/background purity of the subset. After an event has gone through all the cuts (splitting nodes) it ends up in a signal-like or background-like node on the bottom of the tree. These bottom nodes of the tree are called leafs. The output value -1 (1) is returned if the event has ended up in a background-like (signal-like) node.

Depth of the decision tree (number of cuts) is limited by a number of events in a training dataset. If there were too many cuts, the number of events in the leafs would be small and thus the decision would suffer from large statistical fluctuations. When building the decision tree, usually maximal depth of the tree and also minimal number of events per node are defined. When the node reaches the maximal depth, or minimal number of events, the branching is stopped and no other cuts follow.

When building the decision tree, the algorithm starts from the top of the tree. It loops over all input variables and for each of them it searches for the best cut, providing the maximal signal to background separation. The variable and cut providing the best separation are used for the first splitting node. The same algorithm is further applied to both subsamples and it is repeated until the maximum depth or minimal number of events is reached by all nodes on the bottom of the tree. As a measure of signal to background separation, various functions can be used. Those used by the BDT implemented in the TMVA package are listed in Table 7.1.

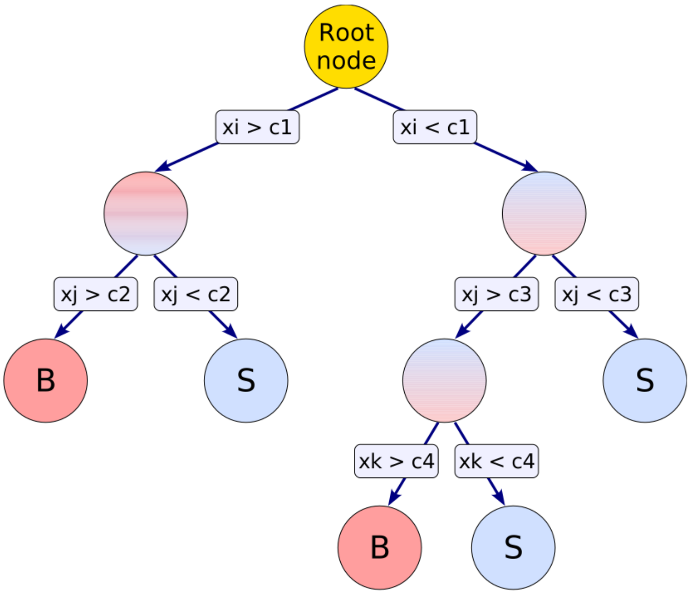


Figure 7.8: Decision tree visualization [86]. The decision tree consists of nodes (cuts) and output leafs (circles on the bottom of the tree). The decision tree returns the same output value for all events in the same leaf.

name	definition
Gini Index (default)	$p.(1\text{-}p)$
Cross entropy	$\text{-}p.\ln(p)\ \text{-}(1\text{-}p).\ln(1\text{-}p)$
Misclassification error	$1\text{-}\max(p,1\text{-}p)$
Statistical significance	$S/\sqrt{S+B}$
Average squared error	$1/N_{events}\cdot\sum_{events}(y-\hat{y})^2$

Table 7.1: Separation criteria used by the BDT training algorithm implemented in the TMVA package [86]. p is the purity, fraction of signal events.

The decision tree splits the n-dimensional space of input variables into hypercubes, each of them classified either as signal or background like.

Regression Tree

The idea of a regression tree is similar to the idea of the classification tree. The regression tree splits the n-dimensional space of input variables into hyper cubes, returning the same output value for all events in the same cube. In order to reach the best performance, it is necessary to build cubes in a way which minimizes target variable variance inside of each cube.

Starting from the full training set of events, the algorithm loops over all input variables and for each of them a set of cut thresholds is tested. For both output branches, the output value minimizing a loss function inside the leaf is returned (arithmetic average for square loss function). The loss function is estimated again from the whole dataset and the cut (value and variable) minimizing the loss function is chosen. The full dataset is now split into two parts, based on the decision applied at the first node. The algorithm is applied again, until a stop criterion is reached by all the leafs.

Forest of decision trees and boosting

The boosted decision tree is a kind of improvement of one decision tree. The BDT algorithm consists of a set of individual decision trees, so-called forest. When classifying an event by the BDT, the total output is calculated as a weighted arithmetic average of response of all the trees in the forest. In order to assign the weight to a tree, the error rate, e_{err} is calculated:

$$e_{err} = \frac{\sum_i^{events} w_i I(s_i, y_i)}{\sum_i^{events} w_i}, \tag{7.19}$$

where i is the event index, w_i is the event weight and function $I(s_i, y_i) = 0$ for correctly classified events ($s_i = y_i$) and 1 for incorrectly classified events. Once the error rate is known, the decision tree weight is calculated as follows:

$$\alpha = \beta.ln(\frac{1 - e_{err}}{e_{err}}), \tag{7.20}$$

where β is a constant.

The idea of the BDT is that a large set of classifiers, which are independent of each other, can classify an event better than a single classifier could. This idea is frequently used for decision trees, but it is a general principle, and it can be used for any MVA classifier. In order to obtain the quasi-independent set of decision trees, so-called boosting is used. The boosting algorithm makes the $(n + 1)$-th tree to focus mostly on the events misclassified by first n trees. Two most common choices of boosting, adaptive boost and gradient boost, are going to be discussed.

Adaptive boost

The adaptive boost algorithm consists of an artificial change of event weights. Once the first decision tree is built, all the events from training sample are classified by it and the sample is split into two parts: events classified correctly and events classified incorrectly. Weights of the misclassified events are artificially increased, so that the next decision tree will focus more on these events and its training will not be affected too much by the events correctly classified by the previous tree. The forest of trees built by the adaptive boost algorithm consists of binary classification trees, so there are only two possible values (usually +1 for signal and -1 for background) at the output of each individual tree.

The boosting guarantees a certain level of independence of the individual trees, since the next tree will more likely classify correctly the events misclassified by the previous tree and those correctly classified previously are more likely to be misclassified.

In the AdaBoost algorithm the event weights are updated as

$$w_{new}(k) = w_{old}(k) \left(\frac{< L >}{1- < L >} \right)^{1-L(k)}, \tag{7.21}$$

where $< L >$ is a mean value of a loss function averaged over whole dataset and $L(k)$ is k-th event contribution to the loss function, divided by maximal contribution to the loss function from all the events. The examples of $L(k)$ can be found in Table 7.2

Linear	$L(k) = \frac{\lvert y(k)-y(k)^{truth}\rvert}{\max\limits_{\text{j} \in \text{events}} \lvert y(j)-y(j)^{truth}\rvert}$
Square	$L(k) = \left[\frac{\lvert y(k)-y(k)^{truth}\rvert}{\max\limits_{\text{j} \in \text{events}} \lvert y(j)-y(j)^{truth}\rvert}\right]^2$
Exponential	$L(k) = 1 - \exp\left[\frac{\lvert y(k)-y(k)^{truth}\rvert}{\max\limits_{\text{j} \in \text{events}} \lvert y(j)-y(j)^{truth}\rvert}\right]$

Table 7.2: Possible loss functions used in AdaBoost algorithm [86].

After building a tree, the boosting algorithm is applied until the desired number of trees is reached.

Gradient boost

The Gradient boost algorithm is a different approach how to obtain a quasi-independent set of decision trees. The idea of Gradient boost is that output of boosted decision tree classifier during the training is given by the equation

$$F_m(\vec{x}) = F_{(m-1)}(\vec{x}) + h(\vec{x}), \tag{7.22}$$

where $F_{m-1}(\vec{x})$ is the output calculated from the trees that already have been built, $h(\vec{x})$ is the output of a new tree which is currently being built and $F_m(\vec{x})$ is output calculated from the already built trees together with the new tree. In the ideal case, $F_m(\vec{x}) = y^{\text{truth}}$, in other words $h(x) = (y^{\text{truth}}$ - $F_{(m-1)}(\vec{x}))$. The $h(\vec{x})$ is not designed to fit the value of the truth target, it is designed to fit so-called residuals, so the value of the loss function for the new $F_m(\vec{x})$ estimator will be minimal. The gradient descent method is used to built the new tree. The new tree is built in the following steps [116]:

1. The gradient of the loss function with respect to the output of all events is calculated,

$$r_{im} = -\left[\frac{\partial E(y^{truth}, F_{(m-1)}(\vec{x_i}))}{\partial F_{(m-1)}(\vec{x_i})}\right], \tag{7.23}$$

where i is index of an event, m is index of a tree.

2. The new decision tree $h(x)$ is built to fit the training set $(\vec{x_i}, r_{im})$, where $\vec{x_i}$ is input variables vector for i-th event and r_{im} is the desired target.

3. The weight γ_m of the new tree is obtained minimizing the loss function:

$$\gamma_m = \min_{\gamma} \sum_{i}^{events} (E(y_i, F_{(m-1)}(\vec{x_i}) + \gamma h(\vec{x_i}))), \tag{7.24}$$

4. Update the model according to the equation

$$F_m(\vec{x}) = F_{(m-1)}(\vec{x}) + \nu\gamma_m h(\vec{x}), \tag{7.25}$$

where $0 < \nu \leq 1$ is a learning rate. The boosted decision tree built using a small learning rate is more robust against the overtraining compared to the BDT built with $\nu = 1$, however, too small learning rate leads to higher CPU time of the training. Introducing the learning rate is called shrinkage.

5. Repeat the process again until a desired number of trees in the forest is reached.

The TMVA package which is used in the analysis uses binomial log-likelihood loss function for the BDT training:

$$L = \sum_{i}^{event} ln\left(1 + e^{-2F(\vec{x_i})y_i^{truth}}\right) \tag{7.26}$$

Variable ranking

The BDT variable ranking is based on counting of the variable occurrence in the cuts, weighted by the separation gain-squared it has achieved and by number of events in each node [86].

7.2 Nuisance parameter fit

In order to estimate the signal fraction in observed events in this analysis, a nuisance parameter fit is used. The nuisance parameter fit allows us to reduce effects of a systematic uncertainties in the case when the prior estimate of the systematics is overestimated and can be measured from data with a higher precision. The nuisance parameter fit is a binned likelihood fit.

Let's assume the simplest case, when only statistical uncertainty of measured data is considered. Giving the expected mean values $\vec{x}$ in bins of a histogram, the probability to observe the data values $\vec{n}$ is given by Poisson distribution as follows.

$$p(\vec{n},\vec{x}) = \Pi_i^{bins} \frac{x_i^{n_i}}{n_i!} e^{-x_i} \tag{7.27}$$

When measuring, for example, a signal cross section of a certain process, the values $\vec{x}$ depend on the cross section σ. Let's assume for now that there is no prior knowledge of the measured cross section and all other parameters of the model, which have an influence on $\vec{x}$, are known precisely. The probability to observe the data points $\vec{n}$ if the cross section is σ, is following:

$$p(\vec{n},\sigma) = \Pi_i^{bins} \frac{x_i(\sigma)^{n_i}}{n_i!} e^{-x_i(\sigma)}, \tag{7.28}$$

where dependency of $\vec{x}$ on σ is usually known from a simulation. Looking for a maximum of probability, the mean value of signal cross section can be measured. Its uncertainty is usually estimated from the points where $p(\vec{n},\sigma)$ reaches exp(-1/2) of the value in maximum, or eventually from interval in which $\int p(\vec{n},\sigma)d\sigma = 0.68$.

The signal cross section in Equations 7.27 and 7.28 is a free parameter of the fit. It means that there is no prior constrain on its value, in other words, there is no direct dependency of the probability on the value of σ, only trough the changes in Poisson distribution terms.

Let's assume a more complicated case. In the high energy physics, there are often lots of uncertainties coming either from a theory or experiment. For example, the cross section of backgrounds is not known with infinite precision, there is an uncertainty on the cross section that needs to be considered in the analysis.

In order to take into account such kind of uncertainty, let's assume the former example, but add a next parameter of the fit, b, which stands for the background cross section. The background cross section is known at a certain level of precision, $b_{prior} \pm \Delta b_{prior}$, where b_{prior} is expected cross section of the background and Δb_{prior} is its uncertainty.

The background cross section is not a free parameter of fit, there is a prior knowledge about it. In order to take into account the systematic variation Δb_{prior} a Gaussian term is added into the Equation 7.28.

$$p(\vec{n}, \sigma, b) = \frac{1}{\sqrt{2\pi}\Delta b} e^{-\frac{(b-b_{prior})^2}{2(\Delta b_{prior})^2}} \Pi_i^{bins} \frac{x_i(\sigma, b)^{n_i}}{n_i!} e^{-x_i(\sigma,b)}, \tag{7.29}$$

Looking for a maximum of $p(\vec{n}, \sigma, b)$ in Equation 7.29, the signal cross section σ and background cross section b can be extracted together with their uncertainties. Let's assume two limit cases:

1. The prior uncertainty on the background cross section, Δb_{prior}, is small and by far under the statistical sensitivity of the analysis. In this case, variations in b will have almost no effect on Poisson term and only the Gaussian term will be influenced. The points where $p(\vec{n}, \sigma, b) = \frac{1}{\sqrt{e}}\max(p(\vec{n}, \sigma, b))$ will be the same as the points where

$$e^{-\frac{(b-b_{prior})^2}{2(\Delta b_{prior})^2}} = \frac{1}{\sqrt{e}}\max\left(e^{-\frac{(b-b_{prior})^2}{2(\Delta b_{prior})^2}}\right) = \frac{1}{\sqrt{e}} \tag{7.30}$$

In this case it is not possible to set a better constrain on the background cross section and the posterior uncertainty will be the same as the prior one, $\Delta b_{prior} \approx \Delta b_{posterior}$.

This is very often the case, since systematic uncertainties are usually predicted from dedicated measurements, which are more sensitive to them than other measurements.

2. The prior uncertainty Δb_{prior} is large and by far above the statistical sensitivity of the measurement. In this case, variations in b will have almost no effect on Gaussian term and only the Poisson term will be influenced. In this case the nuisance parameter behaves almost like a free parameter of the fit and the posterior uncertainty $\Delta b_{posterior}$ will be given by Poisson terms. The posterior uncertainty will be lower then the prior uncertainty, $\Delta b_{posterior} < \Delta b_{prior}$.

In the nuisance parameter fit, there is a set of free parameters, $\vec{\mu}$, typically signal cross section and normalization of major backgrounds, and set of so-called nuisance parameters, $\vec{\Theta}$, for systematic uncertainties. The nuisance parameters are usually defined in the way that $\Theta_j = 0$ means the prior mean value and $\Theta_j = 1$ is $+1\sigma$ variation. In the previous case $\Theta_b = \frac{b - b_{prior}}{\Delta b_{prior}}$.

The likelihood function used for a nuisance parameter fit with nuisance parameters for systematic uncertainties is the following:

$$p(\vec{n}, \vec{\mu}, \vec{\Theta}) = \left(\Pi_j^{syst} \frac{1}{\sqrt{2\pi}} e^{-(\frac{\Theta_j^2}{2})} \right) \Pi_i^{bins} \frac{x_i(\vec{\mu}, \vec{\Theta})^{n_i}}{n_i!} e^{-x_i(\vec{\mu}, \vec{\Theta})} \tag{7.31}$$

Template morphing

In order to be able to use the nuisance parameter fit with Equation 7.31, one has to know $\vec{x}(\vec{\mu}, \vec{\Theta})$ function. For the beginning let's consider just the easiest case, when the fit has only one free parameter, signal cross section and one nuisance parameter, background normalization. For the expected mean value vector $\vec{x}$ we get the following equation,

$$\vec{x}(\vec{\mu}, \vec{\Theta}) = \vec{x}(\sigma_s, \Theta_b) = \sigma_s c_s \vec{x_s} + \sigma_b c_b \vec{x_b} = \sigma_s c_s \vec{x_s} + (1 + \Theta_b \frac{\Delta b_{prior}}{b_{prior}}) b_{prior} c_b \vec{x_b}, \tag{7.32}$$

where σ_s is the signal cross section, σ_b is the background cross section, $\vec{x_s}$ is the signal distribution, $\vec{x_b}$ is the background distribution, Θ_b is the nuisance parameter related to the background cross section uncertainty, b_{prior} and Δb_{prior} are prior background cross section and its uncertainty and c_s and c_b are constants related to the normalization of signal and background templates in Monte Carlo.

For the systematic uncertainties affecting only a template normalization, the relation between the expected mean value vector $\vec{x}$ and values of nuisance parameters is given by Equation 7.32. If a negative normalization of the sample should be avoided, the linear extrapolation is replaced by an exponential:

$$\vec{x}(\vec{\mu}, \vec{\Theta}) = \vec{x}(\sigma_s, \Theta_b) = \sigma_s c_s \vec{x_s} + \sigma_b c_b \vec{x_b} = \sigma_s c_s \vec{x_s} + b_{prior} c_b \vec{x_b} \left(1 + \frac{\Delta b_{prior}}{b_{prior}} \right)^{\Theta_b}. \tag{7.33}$$

For systematics uncertainties affecting also shape of the template, it is necessary to estimate $\vec{x}(\Theta)$ dependency. In this analysis a piecewise linear interpolation and extrapolation is used.

$$\vec{x}(\Theta_i) = \vec{x_s}(\Theta_i) + \vec{x_b} = \vec{x_s}^{\,\text{nominal}} + I_{lin}(\Theta_i, \vec{x_s}^{\,\text{nominal}}, \vec{x_s}^{\,\text{syst-up}_i}, \vec{x_s}^{\,\text{syst-down}_i}) + \vec{x_b}, \tag{7.34}$$

where the systematics described by nuisance parameter Θ_i affects only signal, $\vec{x_s}^{\,\text{nominal}}$ is the signal nominal distribution and $\vec{x_s}^{\,\text{syst-up}_i}$ and $\vec{x_s}^{\,\text{syst-down}_i}$ are $+\sigma$ and -σ systematics variations. The function $I_{lin}(\Theta_i, \vec{x_s}^{\,\text{nominal}}, \vec{x_s}^{\,\text{syst-up}_i}, \vec{x_s}^{\,\text{syst-down}_i})$ is defined in the following way:

$$I_{lin}(\Theta, \vec{x_s}, \vec{x_s}^{\,\text{up}}, \vec{x_s}^{\,\text{down}}) = \begin{cases} \Theta(\vec{x_s}^{\,\text{up}} - \vec{x_s}), & \text{if } \Theta \geq 0 \\ -\Theta(\vec{x_s} - \vec{x_s}^{\,\text{down}}), & \text{if } \Theta < 0 \end{cases} \tag{7.35}$$

The problem with Equation 7.34 is that the systematics is allowed to cause a negative normalization of the sample. In order to avoid the negative normalization, the shape and normalization effects are separated. The shape effect is estimated via the linear extrapolation and interpolation (Equation 7.35) and output histogram is normalized to the nominal distribution integral. The histogram is then multiplied by the normalization term given by exponential extrapolation (Equation 7.36), naturally preventing the integral from reaching negative values.

$$I_{exp}(\Theta, x_s, x_s^{\text{up}}, x_s^{\text{down}}) = \begin{cases} (\frac{x_s^{\text{up}}}{x_s})^{\Theta}, & \text{if } \Theta \geq 0 \\ (\frac{x_s}{x_s^{\text{down}}})^{\Theta}, & \text{if } \Theta < 0 \end{cases} \tag{7.36}$$

Statistical uncertainty of Monte Carlo templates

The statistical uncertainties on Monte Carlo templates are taken into account as separate systematic uncertainty. The uncertainty is decorrelated between the histogram bins, having a separate nuisance parameter, usually labelled γ_i for each bin. The prior distribution of the γ nuisance parameter is described by Poisson and usually $\gamma_i = 1$ corresponds to the nominal value of Monte Carlo prediction in i-th bin.

Pruning and smoothing

The estimate of systematics effects often suffers from a limited number of events in Monte Carlo. If the statistical fluctuations in systematic template are large, they can lead to unphysical pulls and constraints and to double counting of statistical uncertainties of MC. These effects are not desired and there are algorithms for suppressing them. One of the possibilities is to use pruning and smoothing.

The pruning algorithm loops over the systematic uncertainties and compare their effects to a reference threshold. If the effect is lower than the chosen threshold, the systematics is dropped. The pruning allows us to reduce a number of nuisance parameters and thus to decrease a size of the space that has to be exploited by the fitting algorithm. The smaller space usually leads to a higher stability of the fit and lower CPU time.

If the templates for systematics suffer from a large statistical uncertainty, artificial pulls and constraints can appear. If the shape of statistical fluctuations in the templates is similar to the shape of statistical fluctuations in data, the fit will pull the systematics in order to correct the fluctuations of data. These pulls are unphysical and should be avoided. Large statistical fluctuations in the bins for systematics can also cause artificial constrains in the fit and thus also underestimation of the total uncertainty. In order to avoid such effects, several smoothing algorithms have been developed. They smooth statistically insignificant differences between nominal and systematics templates, merging bins together, or replacing bin content by weighted sum of the bin content and contents of neighbouring bins. In this analysis smoothing is used only for systematics which are evaluated from statistically independent samples. It is not used for systematics evaluated by reweighting the same sample, since the difference between nominal and systematics templates are not affected by statistical fluctuations in this case.

Chapter 8

2ℓ channel analysis at 36.1 fb^{-1}

The aim of the analysis in dilepton (2ℓ) channel is to measure the total cross section of $t\bar{t}Z$, targeting the leptonic decay of the Z boson and hadronic decay of the top-quark pair. The dataset corresponding to 2015-2016 period of data taking and luminosity of 36.1 fb^{-1} is used in the 2ℓ channel analysis.

The main sources of backgrounds are Z+jets and dileptonically decaying $t\bar{t}$ with additional jets. In order to suppress these backgrounds, the selection summarized in Section 8.1 is used to define three signal regions.

8.1 Event Selection

If all the decay products of the $t\bar{t}Z$ are reconstructed correctly, in the final state there is a lepton pair from the Z-boson decay, two b-jets from top-quark pair decay and another four light jets from the top-quark pair. No physical missing transverse energy is expected. The lepton pair originating from the Z-boson decay is expected to have an invariant mass close to the Z-boson mass, opposite sign of the electric charge and the same flavour. The event selection is based on the expected final state signature, allowing at most one light jet to be lost because of a limited detector acceptance, or at most one of the b-jets not being b-tagged.

The definitions of three signal regions are summarized in Table 8.2. In order to define b-jets, MV2c10 b-tagger (described in Section 4.5) with the cut reaching 77 % efficiency for b-jets is employed.

Single lepton triggers have been used in the analysis. Since the luminosity in 2016 was significantly higher compared to 2015, the p_{T} threshold of the triggers had to be adapted to the higher event rates. That is the reason why the 2015 data have been collected using different triggers compared to the data from 2016. The triggers used for the both periods of data taking are summarized in Table 8.1. Each data and Monte Carlo event is required to pass at least one trigger.

Year	Electron trigger	Muon trigger

2015	`HLT_e24_lhmedium_L1EM20VH`
	`HLT_e60_lhmedium`
	`HLT_e120_lhloose`
		`HLT_mu20_iloose_L1MU15`
		`HLT_mu50`

2016	`HLT_e26_lhtight_nod0_ivarloose`
	`HLT_e60_lhmedium_nod0`
	`HLT_e140_lhloose_nod0`
		`HLT_mu26_ivarmedium`
		`HLT_mu50`

Table 8.1: Triggers used in the 2ℓ channel analysis with 36.1 fb^{-1} of data. The number right after `mu` or `e` is the minimal p_{T} cut of the given type of lepton that can fire the trigger. `lhloose`, `medium` or `tight` in the name refers to the lepton identification described in Sections 4.2 and 4.3.

Variable	2ℓ-Z-6j1b	2ℓ-Z-5j2b	2ℓ-Z-6j2b
Triggers	Defined in Table 8.1		
Leptons	$= 2$		
Leptons	same-flavour		
Leptons	opposite-sign		
$M_{\ell\ell}$	$\lvert M_{ll} - M_Z\rvert < 10$ GeV		
p_{T} (1st lepton)	> 30 GeV		
p_{T} (2nd lepton)	> 15 GeV		
$n_{b-\mathrm{jets}}$	=1	≥ 2	≥ 2
n_{jets}	≥ 6	=5	≥ 6

Table 8.2: Summary of the event selection in the dilepton channel signal regions.

8.2 Event Yields

The observed and expected number of events in three signal regions of the dilepton channel are shown in Table 8.3.

As Table 8.3 shows, the signal purity in the signal regions is very small, as well as

	2ℓ-Z-6j1b	2ℓ-Z-5j2b	2ℓ-Z-6j2b
$t\bar{t}Z$	80.7 ± 11.2	71.4 ± 4.48	126 ± 16.7
DD $t\bar{t}$	329 ± 20.4	1110 ± 46.8	574 ± 29.3
Z + 2 HF	1060 ± 244	1610 ± 309	913 ± 211
Z + 1 HF	1460 ± 364	245 ± 70.5	126 ± 46.0
Z + 0 HF	794 ± 277	103 ± 55.4	45.2 ± 31.2
other	263 ± 93.8	189 ± 52.0	135 ± 38.0
Total	3980 ± 848	3330 ± 396	1920 ± 285
data	3433	3272	1749
Data/MC	0.86 ± 0.18	0.98 ± 0.13	0.91 ± 0.14
S/(S+B)	2.0 %	2.1 %	6.6 %
S/$\sqrt{S+B}$	1.3	1.2	2.9

Table 8.3: Expected (Monte Carlo) and observed (data) yields in 3 signal regions of the dilepton channel at 36.1 fb^{-1}. Quoted uncertainties include both statistical and systematic uncertainties. The category ”other” includes all other SM background producing at least two promt leptons, as well as background from non-promt and fake leptons. Z + jets contribution is split into 3 parts, based on the number of truth heavy-flavour jets (jets initiated by a c- or b- hadron), since modelling of heavy-flavour jets is problematic and can be mismodelled by Monte Carlo. The $t\bar{t}$ background is estimated using data-driven technique described in Sec. 8.5. The last two rows show signal purity, S/(S+B), and statistical only significance of signal, S/$\sqrt{S+B}$, in other words, what is the ratio of the signal to expected statistical uncertainty of the total number of events in the channel.

the statistical significance of the signal. Including effects of systematics, the expected signal significance would be significantly decreased.

In order to be able to measure the signal cross section of such a rare process at such high background rate, it is necessary to find an algorithm able to separate the signal from background.

8.3 Neural Network

NeuroBayes [115] Neural Network interfaced with TMVA [86] package has been used in order to obtain the variable providing signal vs. background separation.

8.3.1 Considered set of input variables

A set of input variables providing signal versus background discrimination, summarized in Table 8.4, has been tested in the training. There are various motivations for using these input variables.

The important physical difference between $t\bar{t}Z$ and backgrounds, mostly Z+jets and $t\bar{t}$, is in the origin of additional jets. While the signal has six jets initiated by the quarks from the hard scattering (two b-jets and four light jets from the top-quark pair), the backgrounds have only two jets expected from the hard scattering and other jets come mostly from initial or final state radiation, or gluon splitting. This has an important consequences on the jet kinematics. The jets from gluon splitting, or initial/final state radiation are softer compared to the jets initiated by the quarks from a hard scattering. The difference in their energy is the motivation of using p_{T} of the jets, number of jets with $p_{\mathrm{T}} > 40$ GeV and H_{T}^{6jets} which is the scalar sum of p_{T} of leading 6 jets in p_{T}. A significant difference in distribution of these variables between the signal and background can be observed.

The different origin of the jets is also used in the variables targeting the top-quark or W-boson properties, such as $N_{jetpairs}^{Vmass}$, $N_{bjj}^{top-mass}$, M_{jj}^{MindR}, M_{uu}^{Ptord}, M_{bb}, ΔR_{ave}^{jj}, and ΔR_{bb}. In the $t\bar{t}Z$ events, there are two quark pairs, originating from hadronic decay of W bosons. The invariant mass of jet pairs initiated by these quark pairs is expected to be close to the W-mass, $m_W = 80.3$ GeV. In addition, invariant mass of the three jets originating from the same top quark (one b-jet and two light jets), is expected to be close to the top-quark mass $m_{top} = 172.5$ GeV. Because of the finite energy and angular resolution of the calorimeters, the distribution of the jets from W/top invariant masses are expected to be relatively broad, but having a peak at the mass of the top quark or

Variable	Definition
p_{T}^{ll}	p_{T} of the lepton pair
p_{T}^{3jet}	p_{T} of the third jet
p_{T}^{4jet}	p_{T} of the fourth jet
p_{T}^{6jet}	p_{T} of the sixth jet
ΔR_{ll}	ΔR (cone) between the two leptons
$N_{jetpairs}^{Vmass}$	number of jet pairs with mass within a window of 30 GeV around 85 GeV (targeting jets from W boson, and possibly also jets from Z boson)
$N_{bjj}^{top-mass}$	number of 3 jets combinations (with exactly 1 b-jet) close to the top-quark mass ($\|M_{bjj} - M_{top}\| < 15$ GeV) and ($\|M_{jj} - M_{W}\| < 15$ GeV)
$N_{pT>40GeV}^{jet}$	number of jets with $p_T > 40$ GeV
M_{jj}^{MindR}	mass of the combination between the two jets with the smallest ΔR
M_{uu}^{Ptord}	mass of the two untagged jets with the highest p_T
M_{bb}	mass of the two jets with the highest b-tagging weight (output from MV2C10 tagger)
ΔR_{bb}	cone between two jets with the highest b-tagging weight in the event
$\mathrm{Cent}_{\mathrm{jet}}$	scalar sum of p_{T} divided by sum of E for all jets
ΔR_{ave}^{jj}	average ΔR for all jet pairs
$\max M_{lepb}^{MindR}$	maximum mass between a lepton and the b-tagged jet with the smallest ΔR
$H1$	Second Fox-Wolfram moment, given by Equation 8.1
$H1_{jet}$	Second Fox-Wolfram moment built from jets only
H_{T}^{6jets}	scalar p_{T} sum of the first 6 jets
$p_{\mathrm{T}}^{jjj_1}$	p_{T} of 3 jet system formed adding the nearest 2 jets in ΔR to the jet with highest p_{T}
M_{jjj_1}	invariant mass of the 3 jet system used for $p_{\mathrm{T}}^{jjj_1}$
$\eta_{\ell\ell}$	η of the dilepton system
$p_{\mathrm{T}}^{\mathrm{jjj}_2}$	p_{T} of 3 jet system built in the same way as jjj_1, not considering the jets included in jjj_1
M_{jjj_2}	invariant mass of 3 jet system built in the same way as jjj_1, not considering the jets included in jjj_1
M_{W}^{avg}	It is looped over all jet pairs from jjj_1 and jet pairs from jjj_2. The two pairs (one from jjj_1 and one from jjj_2) with minimal difference in the dijet mass is searched for. The variable is defined as the sum of these two dijet invariant masses divided by 2. In the ideal case it matches the W-boson mass.
p_{T}^{b1}	p_{T} of the first b-jet. Jets are ordered by p_T
p_{T}^{b2}	p_{T} of the second b-jet. Jets are ordered by p_T

Table 8.4: The definitions of the input variables used in the Neural Network training. Jets and leptons are ordered by their p_T from the highest one. To supress an effect of the mismodelling in the events with high jet multiplicity, only first 8 jets ordered by p_T are considered when evaluating the variables.

W boson.

The number of jet pairs close to the invariant mass of the W and Z boson, is expected to be higher for the $t\bar{t}Z$ events, since in the ideal case, there are two jet pairs originating from the W-boson decay, while there are no jets from W-boson decay in the Z+jets and dilepton $t\bar{t}$ background. In the phase of the analysis when the Neural Network was being developed, an option to target also decay channels with hadronically decaying Z boson was considered. This is the reason why the W-mass window is centered around 85 GeV instead of 80.3 GeV.

Number of hadronicaly decaying top-quark candidates, $N_{bjj}^{top-mass}$ is defined as the number of three jet combinations, with exactly one b-tagged jet, invariant mass of the three jet system close to the top-quark mass and invariant mass of the two non-b-tagged jets close to the W mass. It is expected to be at least two in the signal events where all the jets from the top-quark pair decay are reconstructed correctly. A higher number of such combinations can arise from another random combination of jets, having by chance invariant masses compatible with W boson and top quark.

Variables M_{jj}^{MindR} and M_{uu}^{Ptord} attempt to reconstruct invariant mass of the W boson. If there is a W boson in an event, the invariant mass of the system of two non-b-tagged jets closest to each other in the cone, or pair of such jets with the highest p_{T} are more likely to have invariant mass close to the W-boson mass.

The variables M_{bb} and ΔR_{bb} are motivated by different origin of the b-jets between the events with top-quark pair ($t\bar{t}Z$ and $t\bar{t}$ in this case) and the Z+jets background, which does not have the top-quark pair. In the case of the Z+jets background, the b-jets are likely to originate from a gluon splitting and consequently having a lower invariant mass and being closer to each other, compared to the $t\bar{t}Z$ and $t\bar{t}$, where the b-jets come from two top quarks, so a higher invariant mass and larger cone between the b-jets can be expected.

The average cone between two jets, ΔR_{ave}^{jj}, provides a discriminating information based on the angular distribution of jets. The variable targets mostly $t\bar{t}Z$ vs. $t\bar{t}$ discrimination. In case of the $t\bar{t}Z$ production, the non zero momentum of Z boson causes the top quarks to be closer to each other, while in the top-quark pair production, the top quarks are very likely to be back-to-back. If the top quarks are closer to each other, smaller average cone between their decay products can be expected, while in the back-to-back case the average cone between the jets is expected to be larger.

Except of the differences in jets, also the origin of the leptons is different between $t\bar{t}Z$ and Z+jets compared to $t\bar{t}$. While in case of the $t\bar{t}Z$ and Z+jets the leptons originate from Z-boson decay, in the case of $t\bar{t}$ they originate from the top-quark pair decay. If the leptons originate from the Z boson, they are expected to be closer to each other in their cone compared to the $t\bar{t}$ case. The variable ΔR_{ll} is designed based on this difference. If a lepton originates from a top quark, the invariant mass of the lepton with the b-jet from the same top-quark is expected to be bellow the top-quark mass. While in the case of a lepton from the Z boson, the invariant mass is expected to be larger. In the analysis, the lepton is paired with the closest b-jet in cone, building the variable $\max M_{lepb}^{MindR}$.

As stated in Section 2.4.1, the Z+jets production is dominated by the quark anti-quark annihilation, where the quark typically carries higher momentum than the anti-quark, since the quark is very likely a valence quark, while the anti-quark is from the sea. As a consequence, the Z+jets system is boosted in the direction of the z-axis. The $t\bar{t}Z$ production is dominated by gluon gluon fusion and thus the boost is typically lower. The lower boost with respect to Z+jets makes the $t\bar{t}Z$ events to be more central. The jets and leptons are more likely to fall into low $|\eta|$ values, while higher values of $|\eta|$ can be expected for Z+jets. The variables $\eta_{\ell\ell}$ and centrality of the event are a measure of the boost.

The second Fox-Wolfram momentum is defined as:

$$H1 = \sum_{i,j} \frac{\vec{p_i} \cdot \vec{p_j}}{E_{viss}^2}, \tag{8.1}$$

where $\vec{p_i}$ and $\vec{p_j}$ are 3-momenta of i-th and j-th object (jet or lepton) and E_{viss} is all visible energy in the event. The Fox-Wolfram moments are often used in order to take into account angular correlations between jets, which are expected to differ between the signal and background [117].

The variables M_{jjj_1} and $p_{\mathrm{T}}^{jjj_1}$ target a top-quark reconstruction. The jet with the highest p_{T} and two closest jets to it in the cone are considered to be from the same top quark. These distributions are therefore expected to be different between the signal and background.

The variables related to jjj_2 are built using a similar logic. From the other jets (not forming jjj_1), the highest p_{T} jet and two closest to it are considered to originate from the other top quark. These three jets are labelled jjj_2. This 3 jet system is used to build M_{jjj_2} and $p_{\mathrm{T}}^{jjj_2}$ variables.

In the next step, it is searched for jet pairs potentially originating from the W-boson decay. It is searched for two jet pairs, one built from jjj_1 and the other from jjj_2, with the minimal difference in their invariant masses. In the ideal case, mass of both jet pairs would be equal to W-boson mass. The arithmetic average of the two masses is used as the input variable M_W^{avg}.

8.3.2 The full set of variables, the variable ranking and sensitivity of the analysis as a function of the number of input variables

The variables are ranked based on their total correlation to the target, using an implementation from NeuroBayes [115] package, described in Section 7.1.3. The ranking of the variables in three signal regions of the analysis is shown in Table 8.5.

The NeuroBayes has been trained using the full set of input variables in order to check distribution of its output and overtraining. The NN output distributions for background and signal events, for both training and testing samples are shown on the bottom of Figure 8.1. The dependency of background rejection on the signal efficiency, so-called ROC curves, are shown on the top of Figure 8.1. As a measure of the NN performance, the area between the diagonal ([1,0] - [0,1] line) and the ROC curve, multiplied by two is shown in the top-right corner. The number is expected to be zero for a random distribution without a signal vs. background separation and one for the ideal separation. The difference between these two numbers for training and testing samples is a measure of the NN overtaining. No significant overtraining has been found.

Although the Neural Network is not overtrained even with the full set of the input variables, it is desired to keep the number of input variables as low as possible in order to avoid a possible bias coming from a mismodelling in the simulation, as described in Section 7.1.2. In order to find an optimal number of the input variables, an estimate of a statistical significance of the measurement has been employed. For a given number of variables, the NN output distribution has been scanned, looking for the optimal cut reaching maximal $S/\sqrt{S+B}$ for events with the NN output higher than the cut. Once the cut is found, the value of the $\max(S/\sqrt{S+B})$ at this point is saved. The $\max(S/\sqrt{S+B})$ is firstly evaluated using only the first variable from the ranking. Once it is found, the process is repeated using the first two variables from top of the ranking, then three e.t.c., until the full set of variables is reached. The dependency of $\max(S/\sqrt{S+B})$ on the number of input variables is shown in Figure 8.2.

Set of the leading 12 input variables (from the Table 8.5) in each region has been chosen as the reasonable trade off between reaching a higher significance and avoiding a possible bias. As it can be seen in Figure 8.2, adding more than 12 variables does not bring any significant improvement in the statistical sensitivity.

rank	2ℓ-Z-6j1b	2ℓ-Z-5j2b	2ℓ-Z-6j2b
1.	$N^{Vmass}_{jetpairs}$	p_{T}^{4jet}	p_{T}^{ll}
2.	p_{T}^{4jet}	$N^{Vmass}_{jetpairs}$	$N^{Vmass}_{jetpairs}$
3.	η^{ll}	p_{T}^{ll}	M_{bb}
4.	$N^{top-mass}_{bjj}$	M_{bb}	p_{T}^{4jet}
5.	$H1$	$N^{top-mass}_{bjj}$	$N^{top-mass}_{bjj}$
6.	M^{mindR}_{jj}	$Centr_{jet}$	$Centr_{jet}$
7.	$Centr_{jet}$	M^{mindR}_{jj}	M^{jjj_1}
8.	p_{T}^{6jet}	$H1$	M^{mindR}_{jj}
9.	p_{T}^{ll}	p_{T}^{3jet}	η^{ll}
10.	M^{pTord}_{uu}	η^{ll}	p_{T}^{6jet}
11.	M^{jjj_1}	M^{jjj_1}	$H1$
12.	M_{bb}	M^{pTord}_{uu}	$p_{\mathrm{T}}^{\mathrm{jjj}_1}$
13.	$MaxM^{mindR}_{lep,b}$	$MaxM^{mindR}_{lep,b}$	$MaxM^{mindR}_{lep,b}$
14.	M^{avg}_{W}	dR_{bb}	M^{avg}_{W}
15.	dR_{ll}	dR^{ave}_{jj}	dR_{bb}
16.	dR_{bb}	$p_{\mathrm{T}}^{\mathrm{jjj}_1}$	M^{pTord}_{uu}
17.	p_{T}^{3jet}	dR_{ll}	p_{T}^{3jet}
18.	M^{jjj_2}	p_{T}^{2bjet}	dR_{ll}
19.	$p_{\mathrm{T}}^{\mathrm{jjj}_2}$	$H1_{jet}$	M^{jjj_2}
20.	H_{T}^{6jets}		dR^{ave}_{jj}
21.	dR^{ave}_{jj}		$p_{\mathrm{T}}^{\mathrm{jjj}_2}$
22.	$p_{\mathrm{T}}^{\mathrm{jjj}_1}$		$H1_{jet}$
23.	$H1_{jet}$		H_{T}^{6jets}

Table 8.5: The ranking of the MVA input variables using the method independent ranking, based on the total correlation of input variables to the target value. The ranking has been performed independently for each analysis region. Some of the variables, related to all six jets are not defined in 2ℓ-Z-5j2b region.

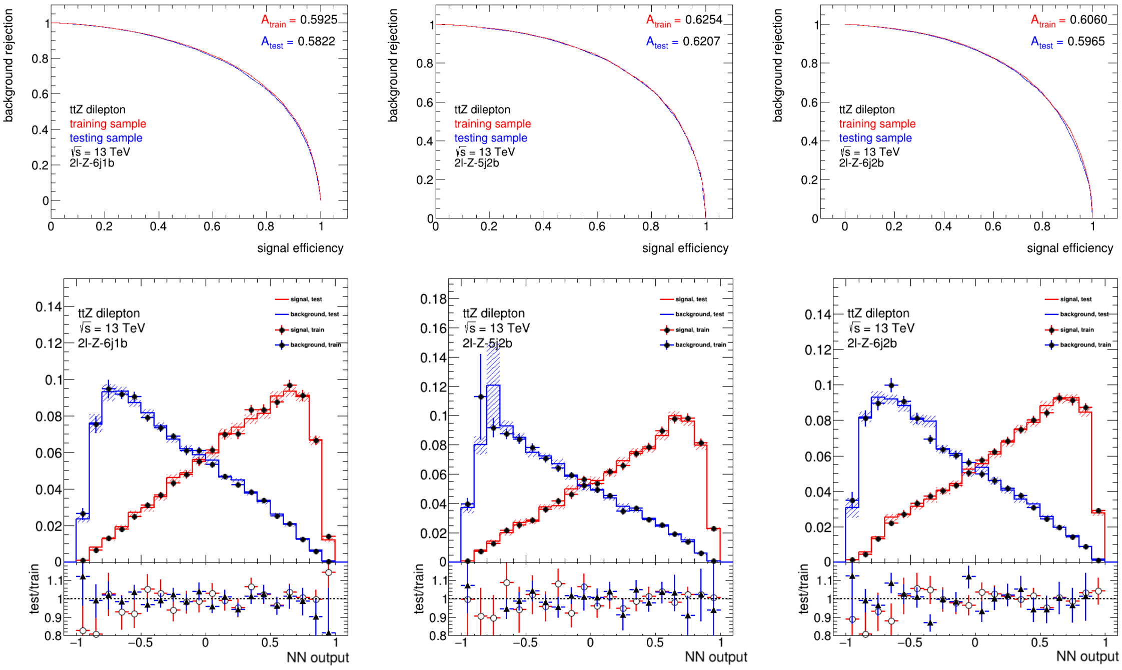


Figure 8.1: The first row shows ROC curves for three signal regions. The second row shows normalized distributions of NeuroBayes output for signal and background for both testing (solid line) and training samples (points). The error bars show only statistical uncertainty. The plots in the first, second and the last column correspond to 2ℓ-Z-6j1b, 2ℓ-Z-5j2b and 2ℓ-Z-6j2b regions. The Neural Network has been trained using the full set of input variables shown in Table 8.5.

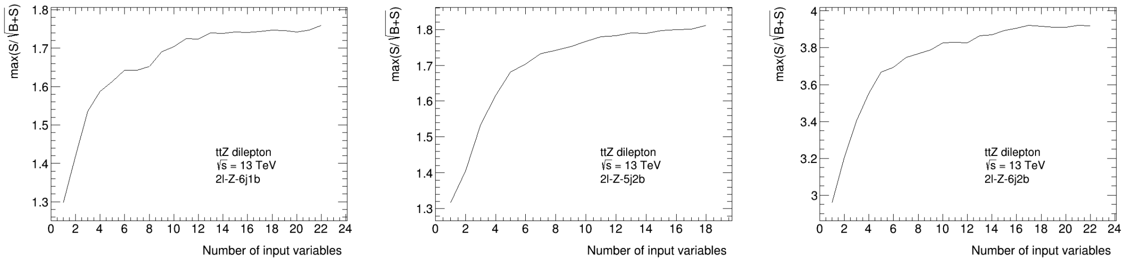


Figure 8.2: Dependency of the estimated signal statistical significance on the number of input variables. For a given number of the input variables the NN output distribution is scanned with very narrow steps (0.001). At each step the $S/\sqrt{S+B}$ for events with the NN output higher than the current point is calculated. The maximal value obtained for a given number of input variables is shown in the graphs above. The final choice is to use 12 variables.

The ideal way how to find the suitable number of the input variables would be to perform a fit with the full systematics for each number of the input variables, instead of calculating only maximal $S/\sqrt{S+B}$, and choose a number of input variables reaching a reasonable signal sensitivity. However, this would be unaffordably CPU intensive and in the time of the NN optimization, a full set of samples for systematic uncertainty estimate was not available.

8.3.3 The final choice of the input variables

The ROC curves and distributions of the NN output for signal and background for both training and testing samples for the final choice of the 12 input variables are shown in Figure 8.3. The correlation matrices of the input variables for signal and background are shown in Figure 8.4

8.3.4 Parameters of the Neural Network

The set of hyperparameters has been taken from the previous ATLAS analysis at 8 TeV [7] and they are summarized in Table 8.6. Further optimizations showed that this set of parameters is suitable also for the 13 TeV measurement. All the input variables are transformed to follow Gaussian distribution with mean value zero and $\sigma = 1$. The signal training samples are reweighted to the integral of background events in order to obtain an NN output with the ideal signal vs. background separation.

In order to avoid a bias caused by an overtraining, the cross training has been used. Available numbers of raw (without reweighting) Monte Carlo events in background and signal samples are summarized in Table 8.7. The Monte Carlo estimate of the $t\bar{t}$ background is used in the training, since the data driven method described in Sec. 8.5 does not provide a sufficient number of events for the NN training.

The total number of raw Monte Carlo signal events before applying any selection is 8,244,600. Based on the Z-decay branch, they are split into the following categories: $Z \to e^+e^-$ (1,408,800), $Z \to \mu^+\mu^-$ (1,409,600), $Z \to \tau^+\tau^-$ (934,300), $Z \to \nu\bar{\nu}$ (1,495,400) and $Z \to q\bar{q}$ (2,996,400).

Parameter	Value
Number of hidden layers	1
Number of nodes in hidden layer	14
Batch size	50 events
Number of iterations	100
Learning speed	0.01
Momentum	0

Table 8.6: Hyperparameters of the Neural Network used in the analysis. The meaning of the parameters is further described in Section 7.1.3.

Region	Signal events	Background events
2ℓ-Z-6j1b	110,894	130,551
2ℓ-Z-5j2b	59,378	83,868
2ℓ-Z-6j2b	144,063	63,701

Table 8.7: Number of Monte Carlo events passing the event selection in the individual analysis regions. The number of events is high enough to train an MVA without any significant overtraining.

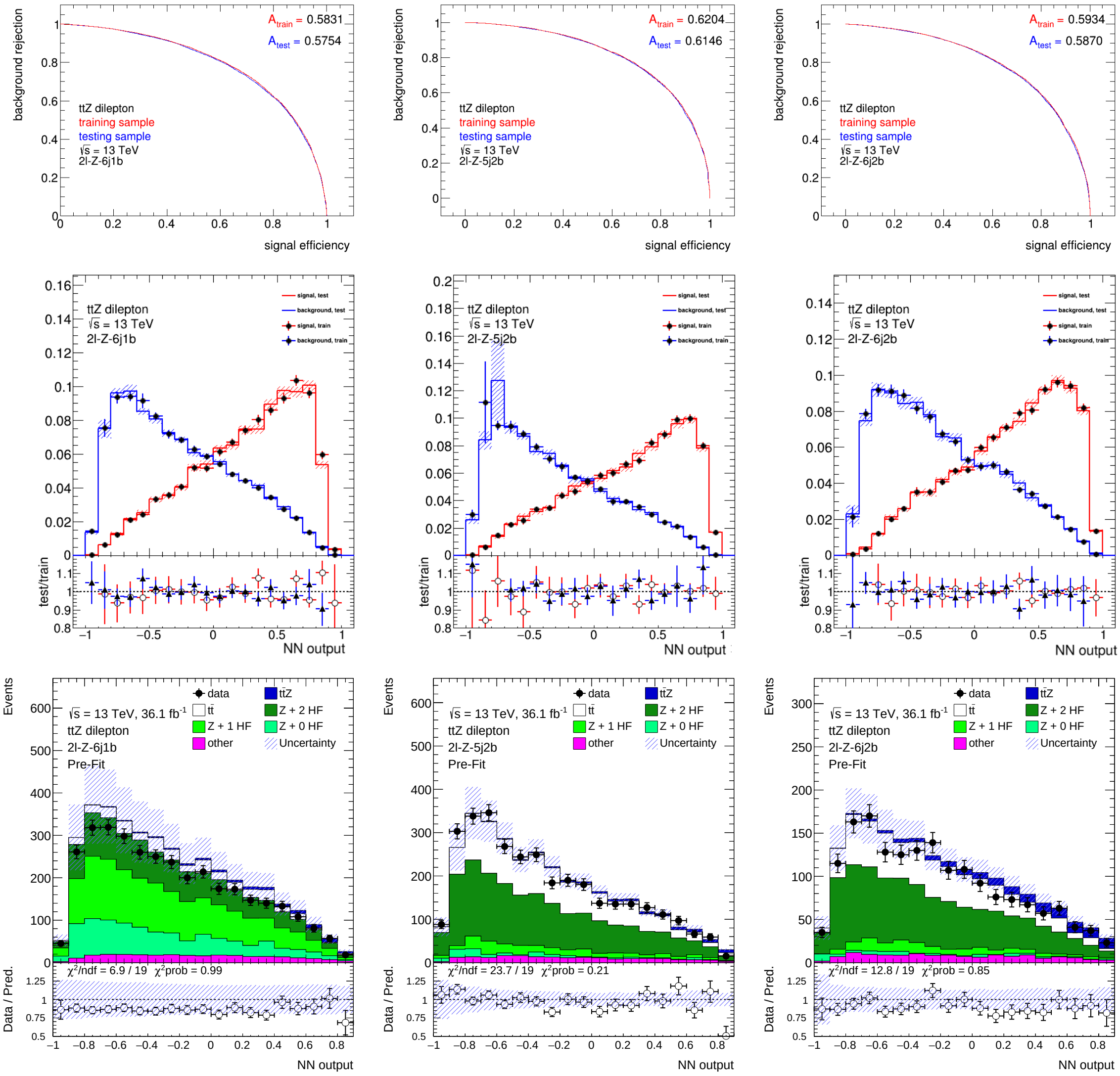


Figure 8.3: The ROC curves (first row), normalized distributions of the NN output (second row) for signal and background for both testing (solid line) and training samples (points) and the distribution of the NN output in simulation and data (third row). The error bars show only statistical uncertainty in the middle row and full statistical + systematic uncertainty in the last row (the purpose of the middle row plots is to check the performance and overtraining, so only the statistical uncertainty is important here, while the purpose of the last row is to check the NN output modelling). The plots in the first, second and the last column correspond to 2ℓ-Z-6j1b, 2ℓ-Z-5j2b and 2ℓ-Z-6j2b regions. The Neural Network has been trained using the set of 12 leading input variables shown in Table 8.5.

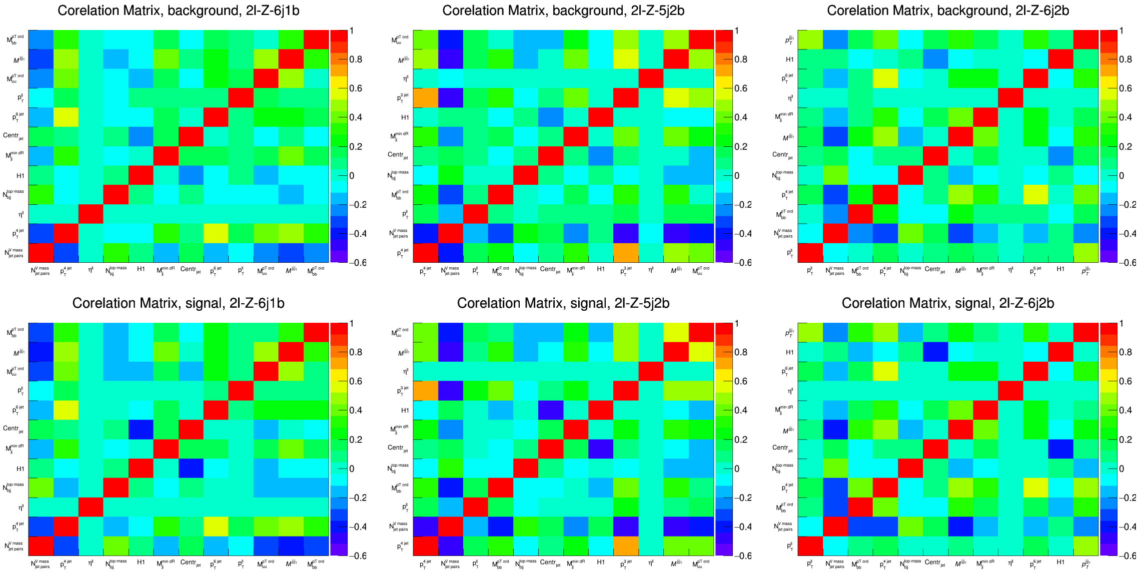


Figure 8.4: The correlation matrices of the NN input variables for background (top) and signal (bottom). The plots in the first, second and the last column correspond to 2ℓ-Z-6j1b, 2ℓ-Z-5j2b and 2ℓ-Z-6j2b regions.

8.4 Boosted Decision Tree

As an alternative MVA method, the Boosted Decision Tree (BDT) has been tested and optimized[1]. With the given set of training events, the BDT has slightly better performance in a sense of the fit sensitivity and areas under the ROC curves, however the Neural Network is more robust against the overtraining. Based on an agreement in the analysis group, the BDT has been chosen as the final MVA method used for the analysis.

The choice of BDT input variables and their ranking are shown in Table 8.8.

rank	2l-Z-6j1b	2l-Z-5j2b	2l-Z-6j2b
1	$\eta^{\ell\ell}$	Cent_{jet}	$H1$
2	Cent_{jet}	$H1$	Cent_{jet}
3	$H1$	$\eta^{\ell\ell}$	$\eta^{\ell\ell}$
4	$N^{Vmass}_{jetpairs}$	dR^{ave}_{jj}	$N^{Vmass}_{jetpairs}$
5	dR^{ave}_{jj}	H_{T}^{6jets}	dR^{ave}_{jj}
6	p_{T}^{4jet}	$N^{Vmass}_{jetpairs}$	p_{T}^{4jet}
7	dR_{ll}	M^{MindR}_{jj}	ΔR_{bb}
8	p_{T}^{ll}	dR_{ll}	p_{T}^{ll}
9	p_{T}^{6jet}	ΔR_{bb}	M_{bb}
10	M^{avg}_{W}	M_{bb}	H_{T}^{6jets}
11	p_{T}^{1b-jet}	p_{T}^{ll}	p_{T}^{6jet}
12	H_{T}^{6jets}	p_{T}^{4jet}	dR_{ll}
13	$M^{\mathrm{MindR}}_{\mathrm{jj}}$	$M^{\mathrm{Ptord}}_{\mathrm{uu}}$	$\max M^{\mathrm{MindR}}_{\mathrm{lepb}}$
14	$\max M^{\mathrm{MindR}}_{\mathrm{lepb}}$	p_{T}^{5jet}	$M^{\mathrm{MindR}}_{\mathrm{jj}}$
15	$M^{\mathrm{Ptord}}_{\mathrm{uu}}$		M^{avg}_{W}
16			p_{T}^{1b-jet}
17			$N^{top-mass}_{bjj}$

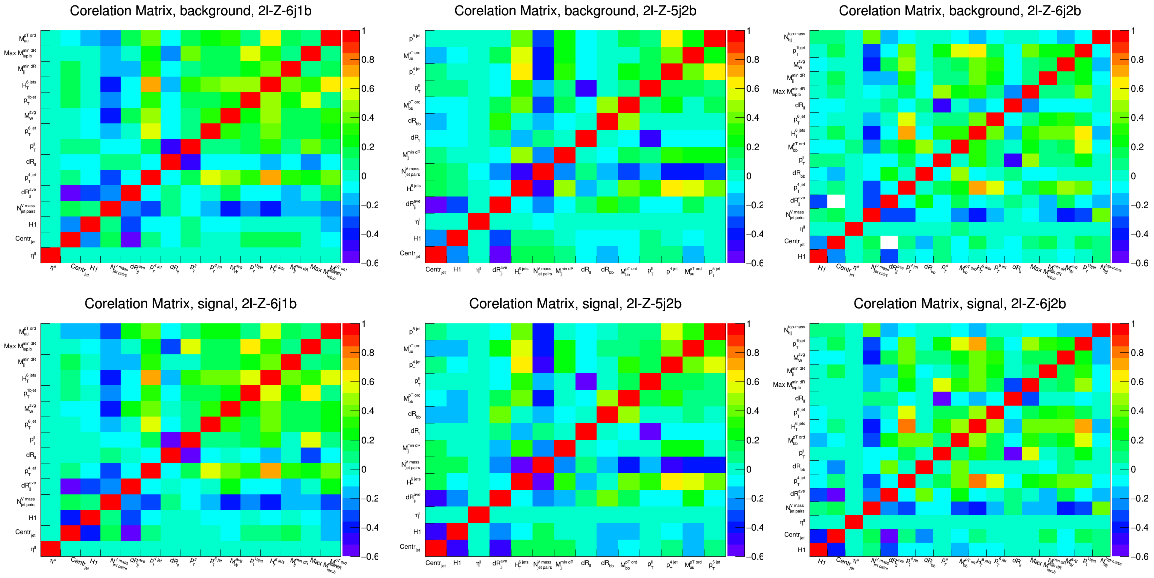
Table 8.8: Ranking of the variables used for the BDT training.

The correlation matrices of the input variables are shown in Figure 8.5. The ROC curves and distributions of the BDT output in signal and background for both training and testing samples, together with control plots of the BDT output in data and the simulation are shown in Figure 8.6. The parameters of the BDT are summarized in

---

[1] The BDT optimization was not done by the author of the thesis.

Table 8.9.


Corelation Matrix, background, 2l-Z-6j1b

Corelation Matrix, background, 2l-Z-5j2b

Corelation Matrix, background, 2l-Z-6j2b

Corelation Matrix, signal, 2l-Z-6j1b

Corelation Matrix, signal, 2l-Z-5j2b

Corelation Matrix, signal, 2l-Z-6j2b

Figure 8.5: Correlation matrices of the BDT input variables for background (top) and signal (bottom). The plots in the first, second and the last column correspond to 2ℓ-Z-6j1b, 2ℓ-Z-5j2b and 2ℓ-Z-6j2b regions.

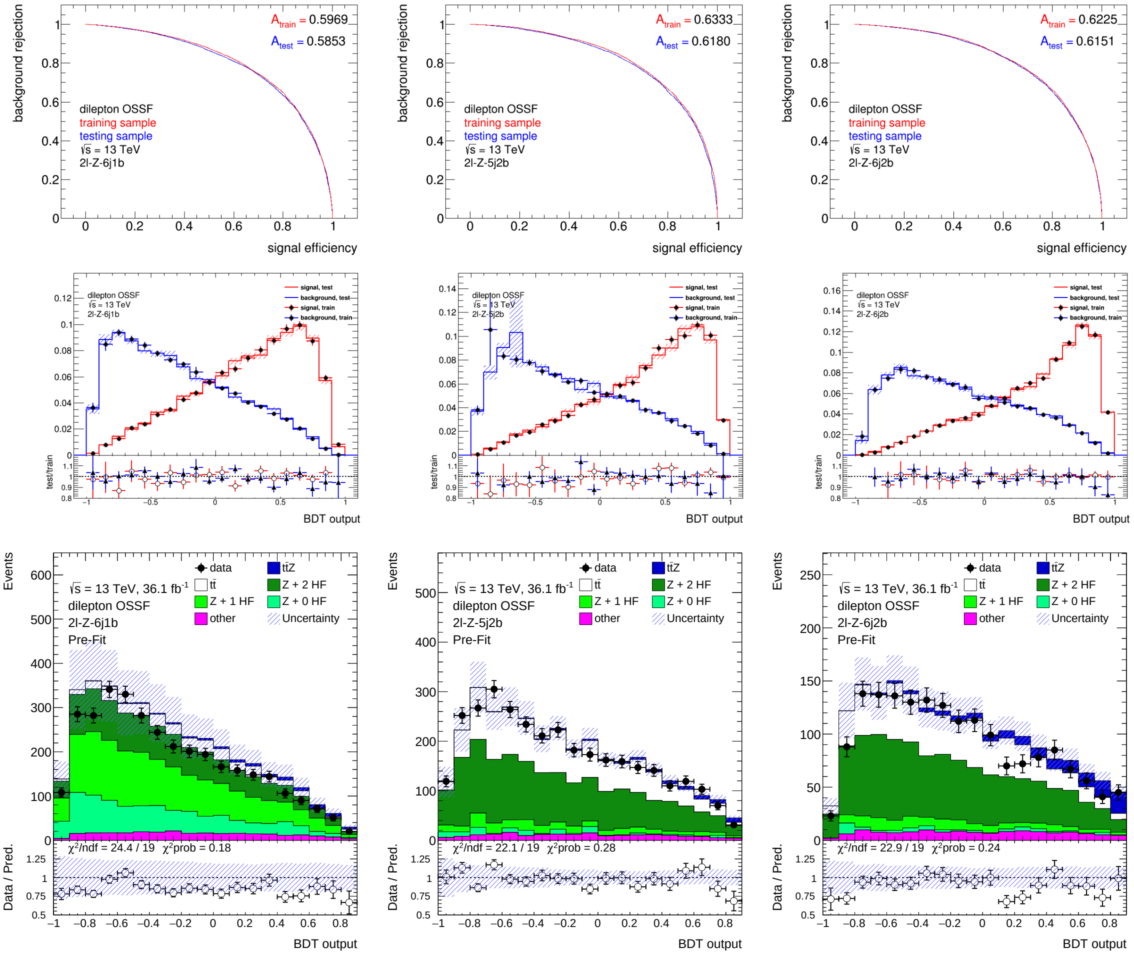


Figure 8.6: The ROC curves (first row) and normalized distributions of the BDT output (second row) for signal and background for both testing (solid line) and training samples (points) and distributions of the BDT output in simulation and data (third row). The error bars show only statistical uncertainty in the middle row and full statistical + systematic uncertainty in the last row (the purpose of the middle row plots is to check the performance and overtraining, so only the statistical uncertainty is important here, while the purpose of the last row is to check the BDT output modelling). The plots in the first, second and the last column correspond to 2ℓ-Z-6j1b, 2ℓ-Z-5j2b and 2ℓ-Z-6j2b regions. The BDT has been trained using the set of input variables shown in Table 8.8.

Option	Values	Description
NTrees	500	Number of trees in the forest
Maxdepth	3	Max depth of decision tree
Minnodesize	5%	Minimimum % of training events required in a leaf node
nCuts	20	Number of grid points in variable range used in finding optimal cut in node splitting
BoostType	Grad	Boosting type for the trees in forest
Shrinkage	0.3	Learning rate for GradBoost Algorithm
SeparationType	GiniIndex	Separation criteria for node splitting
NodePurityLimit	0.5	nodes with purity > NodePurityLimit are signal
SigToBkgFraction	1	Sig to Bkg ratio used in Training

Table 8.9: Settings used in the BDT training.

8.5 Data-driven $t\bar{t}$ background estimate

The top-quark pair production is the second largest background in the dilepton channel. In order to check the top-quark pair production modelling, dedicated $t\bar{t}$ validation regions have been designed.

Because of the lepton universality, dilepton decays of the top-quark pair into e^+e^-, $\mu^+\mu^-$, $e^+\mu^-$ and $e^-\mu^+$ have the same branching ratio and the same event kinematics. While the first two decay modes (let's call them $\ell\ell$) have both leptons of the same type and thus contribute to the background in the dilepton channel of the $t\bar{t}Z$, the decay mode with different flavour of the leptons ($e\mu$) can be easily identified and does not contribute to the $t\bar{t}Z$ signal regions.

The $t\bar{t}$ validation regions are defined by the selection summarized in Table 8.10. The selection is the same as the definition of the $t\bar{t}Z$ signal regions, the only difference is in the lepton flavour requirement. The lepton pair in the $t\bar{t}$ validation regions is required to be of a different flavour. The other cuts are kept the same in order to use the events from the same kinematic region.

Variable	$e\mu$-Z-6j1b-VR	$e\mu$-Z-5j2b-VR	$e\mu$-Z-6j2b-VR
Triggers	Defined in Table 8.1		
Leptons	$= 2$		
Leptons	$e\mu$		
Leptons	opposite-sign		
$M_{\ell\ell}$	$\lvert M_{ll} - M_Z\rvert < 10$ GeV		
p_{T} (1st lepton)	> 30 GeV		
p_{T} (2nd lepton)	> 15 GeV		
$n_{b-\mathrm{jets}}$	=1	≥ 2	≥ 2
n_{jets}	≥ 6	=5	≥ 6

Table 8.10: Summary of the event selection in the $t\bar{t}$ validation regions. The only difference with respect to the definition of signal regions (summarized in Table 8.2) is requirement on the different flavour of the leptons.

The event yields obtained from the simulation and observed data yields are shown in Table 8.11. The validation regions are clearly dominated by the top-quark pair production.

Monte Carlo generators have a problem to describe properly the top-quark pair production in association with additional b-jets. The events with one or two b-jets are

	$e\mu$-Z-6j1b-VR	$e\mu$-Z-5j2b-VR	$e\mu$-Z-6j2b-VR
$t\bar{t}Z$	1.6 ± 0.3	5.6 ± 0.7	4.2 ± 0.7
$t\bar{t}$	304 ± 18.0	1100 ± 65.0	527 ± 31.1
Z + 2 HF	0.08 ± 0.01	0.3 ± 0.04	0.05 ± 0.02
Z + 1 HF	0 ± 0	0 ± 0	0 ± 0
Z + 0 HF	0 ± 0	0 ± 0	0 ± 0
other	33.4 ± 11.9	64.9 ± 20.4	45.1 ± 16.5
Total	339 ± 21.4	1170 ± 67.9	577 ± 35.1
observed	341	1151	596
Data/MC	1.01 ± 0.08	0.98 ± 0.06	1.03 ± 0.08

Table 8.11: Event yields of three $t\bar{t}$ validation regions. The signal contamination is very small and the regions are dominated by $t\bar{t}$.

well modelled, but a disagreement between data and simulation can be observed for three and more b-jets. The number of b-jets observed in the data and simulation is shown in Figure 8.7

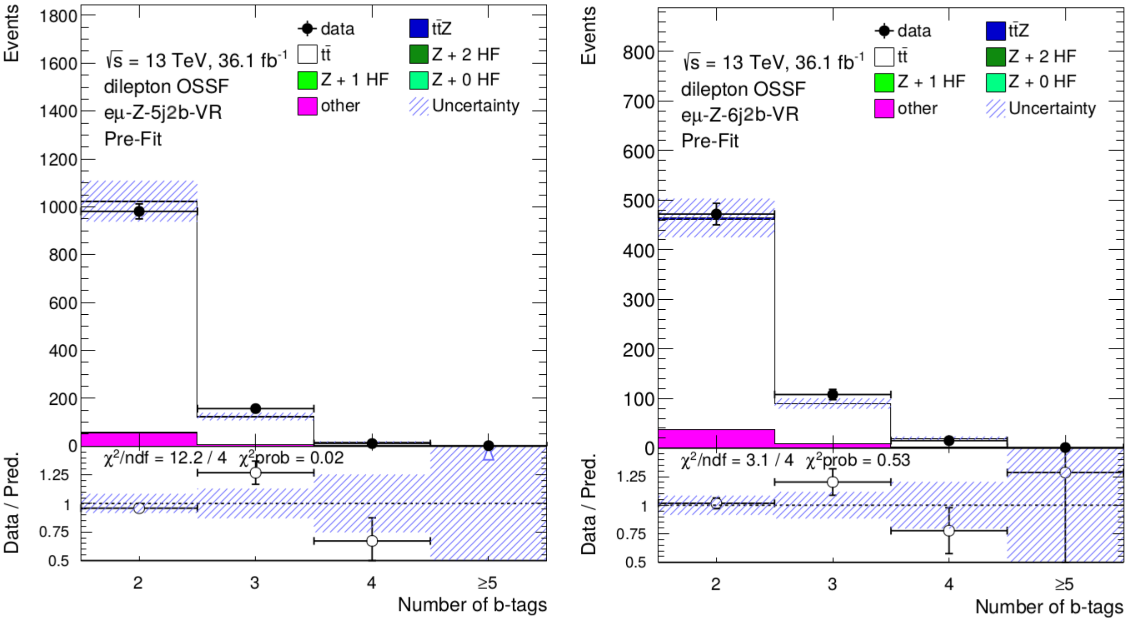


Figure 8.7: Control plots for number of b-jets in $e\mu$-Z-5j2b-VR and $e\mu$-Z-6j2b-VR regions. The mismodelling in $\geq$3b region might seem to be statistically insignificant, but it is observed by many other analyses with significantly larger number of events.

In order to deal with the modelling of the additional b-jets in the top-quark pair production and to reduce effects of modelling and experimental systematic uncertainties on this background, a data-driven estimate of the $t\bar{t}$ background has been employed.

The data events from $e\mu$ validation regions are used in the signal regions as the data-driven $t\bar{t}$ background estimate. The non-$t\bar{t}$-background and $t\bar{t}Z$ contributions in $e\mu$ regions are estimated from Monte Carlo and subtracted from data, so the resulting distributions, (data$_{e\mu}$ - non-$t\bar{t}$-MC$_{e\mu}$), correspond to $t\bar{t}$ only contribution.

The different detector efficiency and slightly different $\tau \to e$ and $\tau \to \mu$ branching ratios are taken into account using a multiplicative scale factor estimated from $t\bar{t}$ Monte Carlo. The $e\mu$ data events are reweighted using the ratio of the $t\bar{t}$ events with $\ell\ell$ lepton pair to the events with $e\mu$ pair:

$$C_{t\bar{t}} = \frac{N_{t\bar{t}}^{\ell\ell}}{N_{t\bar{t}}^{e\mu}} = 0.981 \pm 0.030, \tag{8.2}$$

where $N_{t\bar{t}}^{\ell\ell}$ is Monte Carlo prediction for $t\bar{t}$ events after $\ell\ell$ selection and $N_{t\bar{t}}^{e\mu}$ is Monte Carlo prediction for $t\bar{t}$ after $e\mu$ selection. The uncertainty ± 0.030 comes from statistical and $t\bar{t}$ modelling (Powheg+Pythia8 vs. aMC@NLO+Pythia8) uncertainty of the ratio. It is applied as a systematics uncertainty on the data-driven $t\bar{t}$ normalization in the fit. The ratio is calculated from the sum of events from all three (6j1b, 5j2b and 6j2b) regions together. Calculating the ratio in individual regions is consistent with the value 0.981 ± 0.030, however its uncertainty is larger because of a lower number of available events. Since there is no expected difference between $t\bar{t}$ $\ell\ell$ and $e\mu$ events among the regions, the global value of the ratio is further used in the analysis.

The shapes of Neural Network and Boosted Decision Tree outputs, as well as distributions of all their input variables have been checked in Monte Carlo, comparing their shapes and normalizations in $e\mu$ and $\ell\ell$ events. No statistically significant difference has been found. The plots comparing shape of the variables between $\ell\ell$ and $e\mu$ Monte Carlo for top-quark pair production can be found in Appendix .1.

Because of a low number of events in $e\mu$ data, the data-driven $t\bar{t}$ estimate cannot be used for the NN and BDT training. The low number of available events would lead to a large overtraining. In order to estimate an effect of the overtraining, the Monte Carlo based $t\bar{t}$ estimate is also used in overtraining checks, such as ROC curve plots and comparison of the MVA response on training and testing samples. However, the data-driven $t\bar{t}$ background estimate is used in the rest of the analysis such as in the control plots, yield tables, fit, studies of the fit sensitivity, modelling studies, etc.

8.5.1 Neural Network and Boosted Decision Tree validation

In order to validate the response of the BDT and NN in an independent region and check the agreement between simulation and data, the $e\mu$ validation regions have been used. The distributions of the NN and BDT outputs are shown in Figure 8.8. No statistically significant discrepancies have been found.

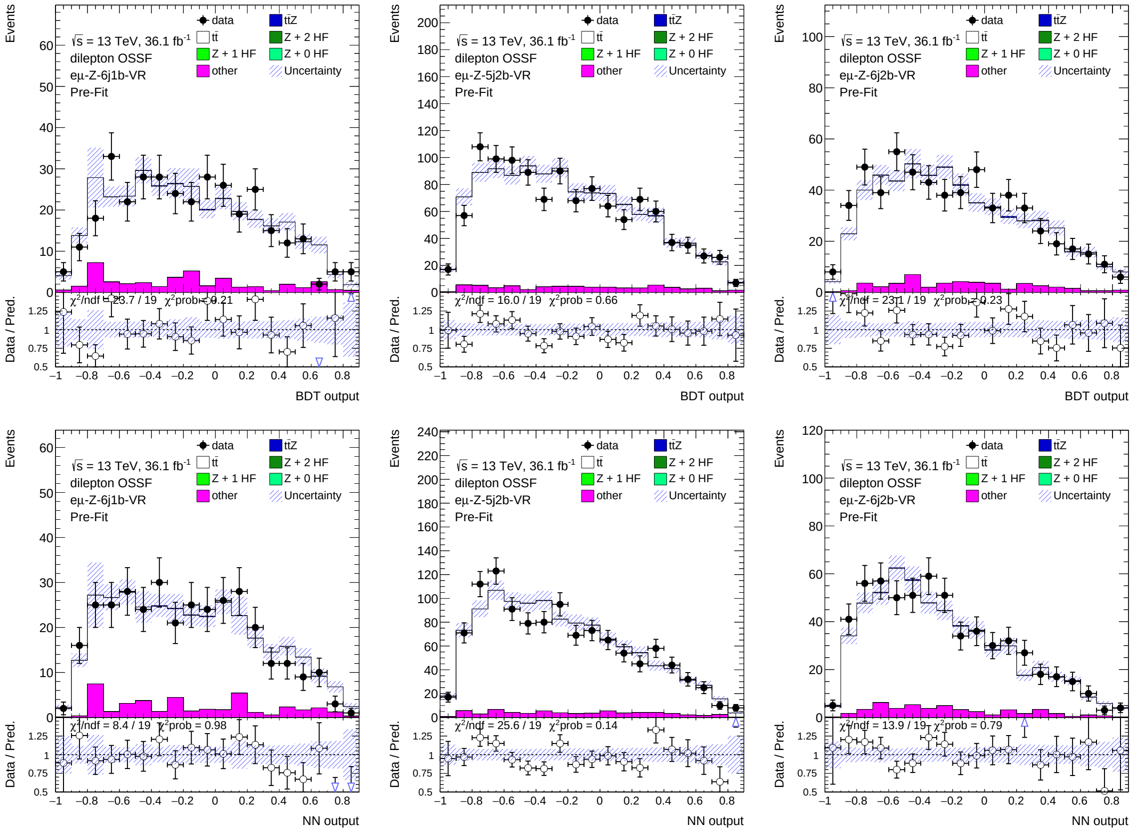


Figure 8.8: The Boosted Decision Tree (top) and Neural Network (bottom) output distributions in three $t\bar{t}$ validation regions: $e\mu$-Z-6j1b-VR region (first column), $e\mu$-Z-5j2b-VR region (second column) and $e\mu$-Z-6j2b-VR region (the last column). The Neural Network has been trained using the 12 leading variables shown in Table 8.5, the Boosted Decision Tree has been trained using the variables from Table 8.8. The shaded bands show both statistical and systematic uncertainty of Monte Carlo. No significant disagreement between data and Monte Carlo has been found.

8.6 Fit strategy

In order to extract the cross section of the $t\bar{t}Z$, the nuisance parameter fit described in Section 7.2 has been used. The distributions of the MVA output in the three analysis regions are fitted. The choice of the MVA type and binning choice will be described in Section 8.7.

In order to reduce systematic uncertainties on the $t\bar{t}$ background and deal with a possible missmodelling, the fully data-driven technique of the $t\bar{t}$ background described in Section 8.5 has been employed to obtain the $t\bar{t}$ templates used in the fit. Limited number of events in the data driven $t\bar{t}$ templates is taken into account as additional systematic uncertainty.

The Z+jets background has been split into three parts, based on a number of truth heavy flavour jets[2] in an event: Z+0 HF, Z+1 HF and Z+2 HF (the last one includes events with two and more truth heavy flavour jets). The normalization of Z+0 HF component has been estimated from Monte Carlo and 10% uncertainty has been assigned to it. The normalizations of Z+1 HF and Z+2 HF are free parameters of the fit, so the fit can extract them from low MVA output regions, which are dominated by Z+jets with a low signal contamination.

The fit has three free parameters: $t\bar{t}Z$ normalization, Z + 1HF normalization and Z + 2HF normalization. Each of the normalizations is described by a μ parameter, which is the ratio between the fitted cross section and the Monte Carlo prediction of the cross section ($\mu = 1$ means the perfect agreement between data and prediction).

The set of systematic uncertainties described in Section 6 has been considered. Each of the systematic variations is described by a nuisance parameter in the fit.

In optimization studies, so-called Asimov fit is used. The Asimov fit is fit of Monte Carlo templates to the same Monte Carlo. The mean values of the fitted parameters agree with their Monte Carlo predictions, but their uncertainties provide an estimate of the final uncertainty of the measurement. The Asimov fit is used also to choose the final MVA choice from the NN and BDT and optimize the binning choice for the MVA output histograms.

---

[2] The jet initiated by a c or b hadron is considered to be a heavy flavour jet. The Monte Carlo information from the parton level event has been employed to assign the hadron to the jets.

8.7 Asimov fit results

8.7.1 Asimov fit results using Neural Network

Binning optimization

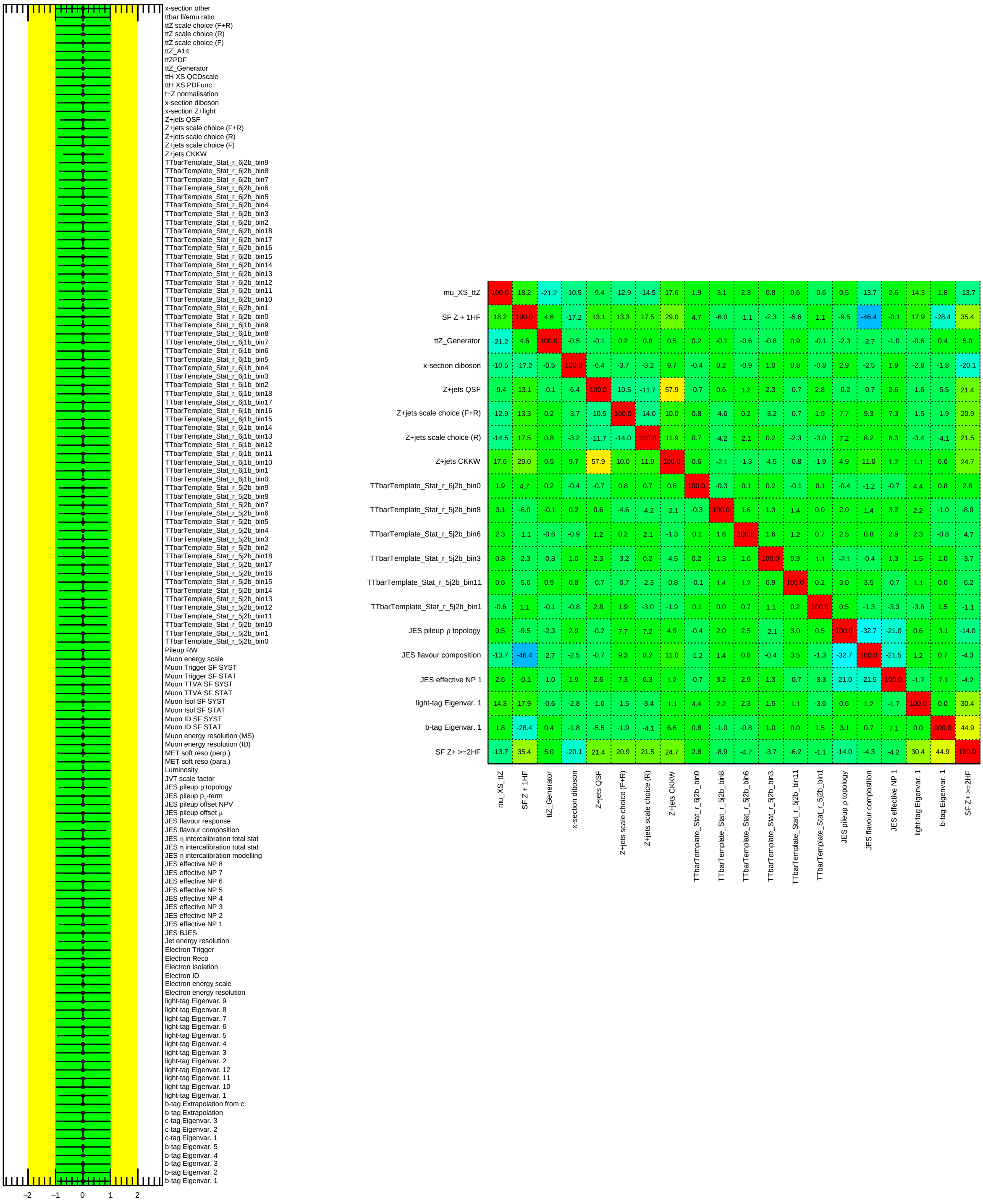
In order to reach the best signal sensitivity, it is necessary to study effects of the binning choice. A wider binning leads to lower statistical fluctuations in individual bins, making the fit more stable and it reduces technical problems with low number of events in the bins and possible negative bin contents caused by the statistical fluctuations of the Monte Carlo events. On the other hand, wide bins are less sensitive to the shape differences between signal and background distributions, which increases the uncertainty of the measurement. The reasonable choice of the binning is a trade of between the stability against the statistical fluctuations and a reasonable sensitivity to the shape difference between signal and background templates.

The following numbers of bins have been tested in the fit of the Neural Network output: 1, 2, 5, 10 and 19. The NN output is from the interval [-1, 1], thus the histograms with 1, 2, 5 and 10 bins have the equally wide bins, covering this interval. However, for 20 bins, the last bin, covering the interval [0.9, 1], contains a very few events for $t\bar{t}$ and other minor background sources, which prevents the fit from converging for this binning choice. If the last two bins ([0.8, 0.9] and [0.9, 1.0]) are merged together, ending up with 19 bins, the fit converges reaching the best signal sensitivity from the tested options. The higher number of bins than 20 were tested as well, but the problems with negative bin contents started to be very significant for such high number of bins and thus the fit is not able to converge. Table 8.12 summarizes expected fitted signal cross section and its uncertainty for the tested binning choices.

The 19 bins from interval [-1, 0.9], with the last bin including also overflow, have been chosen as the final binning choice.

Fit results for the final binning choice

Fitting the Neural Network output distributions showed on the bottom of Figure 8.3, the results summarized in Table 8.13 have been obtained. The plots showing nuisance parameters and uncertainties on their post-fit values, together with the correlation matrix of the fit parameters, are shown in Figure 8.9.



	mu_XS_ttZ	SF Z + 1HF	ttZ_Generator	x-section diboson	Z+jets QSF	Z+jets scale choice (F+R)	Z+jets scale choice (R)	Z+jets CKKW	TTbarTemplate_Stat_r_6j2b_bin0	TTbarTemplate_Stat_r_5j2b_bin8	TTbarTemplate_Stat_r_5j2b_bin6	TTbarTemplate_Stat_r_5j2b_bin3	TTbarTemplate_Stat_r_5j2b_bin11	TTbarTemplate_Stat_r_5j2b_bin1	JES pileup ρ topology	JES flavour composition	JES effective NP 1	light-tag Eigenvar. 1	b-tag Eigenvar. 1	SF Z+ >=2HF
mu_XS_ttZ	100.0	18.2	-21.2	-10.5	-9.4	-12.9	-14.5	17.6	1.9	3.1	2.3	0.8	0.6	-0.6	0.5	-13.7	2.6	14.3	1.8	-13.7
SF Z + 1HF	18.2	100.0	4.6	-17.2	13.1	13.3	17.5	29.0	4.7	-6.0	-1.1	-2.3	-5.6	1.1	-9.5	-46.4	-0.1	17.9	-28.4	35.4
ttZ_Generator	-21.2	4.6	100.0	-0.5	-0.1	0.2	0.8	0.5	0.2	-0.1	-0.6	-0.8	0.9	-0.1	-2.3	-2.7	-1.0	-0.6	0.4	5.0
x-section diboson	-10.5	-17.2	-0.5	100.0	-6.4	-3.7	-3.2	9.7	-0.4	0.2	-0.9	1.0	0.8	-0.8	2.9	-2.5	1.9	-2.8	-1.8	-20.1
Z+jets QSF	-9.4	13.1	-0.1	-6.4	100.0	-10.5	-11.7	57.9	-0.7	0.6	1.2	2.3	-0.7	2.8	-0.2	-0.7	2.6	-1.6	-5.5	21.4
Z+jets scale choice (F+R)	-12.9	13.3	0.2	-3.7	-10.5	100.0	-14.0	10.0	0.8	-4.6	0.2	-3.2	-0.7	1.9	7.7	9.3	7.3	-1.5	-1.9	20.9
Z+jets scale choice (R)	-14.5	17.5	0.8	-3.2	-11.7	-14.0	100.0	11.9	0.7	-4.2	2.1	0.2	-2.3	-3.0	7.2	8.2	6.3	-3.4	-4.1	21.5
Z+jets CKKW	17.6	29.0	0.5	9.7	57.9	10.0	11.9	100.0	0.6	-2.1	-1.3	-4.5	-0.8	-1.9	4.9	11.0	1.2	1.1	6.6	24.7
TTbarTemplate_Stat_r_6j2b_bin0	1.9	4.7	0.2	-0.4	-0.7	0.8	0.7	0.6	100.0	-0.3	0.1	0.2	-0.1	0.1	-0.4	-1.2	-0.7	4.4	0.8	2.8
TTbarTemplate_Stat_r_5j2b_bin8	3.1	-6.0	-0.1	0.2	0.6	-4.6	-4.2	-2.1	-0.3	100.0	1.6	1.3	1.4	0.0	2.0	1.4	3.2	2.2	-1.0	-8.9
TTbarTemplate_Stat_r_5j2b_bin6	2.3	-1.1	-0.6	-0.9	1.2	0.2	2.1	-1.3	0.1	1.6	100.0	1.6	1.2	0.7	2.5	0.8	2.9	2.3	-0.8	-4.7
TTbarTemplate_Stat_r_5j2b_bin3	0.8	-2.3	-0.8	1.0	2.3	-3.2	0.2	-4.5	0.2	1.3	1.6	100.0	0.9	1.1	-2.1	-0.4	1.3	1.5	1.0	-3.7
TTbarTemplate_Stat_r_5j2b_bin11	0.6	-5.6	0.9	0.8	-0.7	-0.7	-2.3	-0.8	-0.1	1.4	1.2	0.9	100.0	0.2	3.0	3.5	-0.7	1.1	0.0	-6.2
TTbarTemplate_Stat_r_5j2b_bin1	-0.6	1.1	-0.1	-0.8	2.8	1.9	-3.0	-1.9	0.1	0.0	0.7	1.1	0.2	100.0	0.5	-1.3	-3.3	-3.6	1.5	-1.1
JES pileup ρ topology	0.5	-9.5	-2.3	2.9	-0.2	7.7	7.2	4.9	-0.4	2.0	2.5	-2.1	3.0	0.5	100.0	-32.7	-21.0	0.6	3.1	-14.0
JES flavour composition	-13.7	-46.4	-2.7	-2.5	-0.7	9.3	8.2	11.0	-1.2	1.4	0.8	-0.4	3.5	-1.3	-32.7	100.0	-21.5	1.2	0.7	-4.3
JES effective NP 1	2.6	-0.1	-1.0	1.9	2.6	7.3	6.3	1.2	-0.7	3.2	2.9	1.3	-0.7	-3.3	-21.0	-21.5	100.0	-1.7	7.1	-4.2
light-tag Eigenvar. 1	14.3	17.9	-0.6	-2.8	-1.6	-1.5	-3.4	1.1	4.4	2.2	2.3	1.5	1.1	-3.6	0.6	1.2	-1.7	100.0	0.0	30.4
b-tag Eigenvar. 1	1.8	-28.4	0.4	-1.8	-5.5	-1.9	-4.1	6.6	0.8	-1.0	-0.8	1.0	0.0	1.5	3.1	0.7	7.1	0.0	100.0	44.9
SF Z+ >=2HF	-13.7	35.4	5.0	-20.1	21.4	20.9	21.5	24.7	2.8	-8.9	-4.7	-3.7	-6.2	-1.1	-14.0	-4.3	-4.2	30.4	44.9	100.0

Figure 8.9: Nuisance parameters plot (left) and correlation matrix (right) for the Asimov fit of the Neural Network output distribution. The black dots (lines) in the nuisance parameter plot show mean value (uncertainty) of the post-fit distribution of the nuisance parameter. The green band shows pre-fit uncertainty on the nuisance parameters.

Number of bins	Fitted signal cross section
1	$1.000\ ^{+1.997}_{-2.008}$
2	$1.000\ ^{+0.535}_{-0.537}$
5	$1.000\ ^{+0.371}_{-0.344}$
10	$1.000\ ^{+0.345}_{-0.319}$
19	$1.000\ ^{+0.343}_{-0.315}$

Table 8.12: The fitted signal cross section and its uncertainty, as the ratio between fitted value and Monte Carlo prediction, extracted from the Asimov fit of the NN output distribution for various binning choices. The uncertainties include both statistical and systematic uncertainty.

Parameter	Value
$\mu_{t\bar{t}Z}$	$1.000\ ^{+0.218}_{-0.212}\ \text{(stat.)}\ ^{+0.265}_{-0.233}\ \text{(syst.)} = 1.000\ ^{+0.343}_{-0.315}$
μ_{Z+1HF}	$1.000\ ^{+0.053}_{-0.053}\ \text{(stat.)}\ ^{+0.276}_{-0.234}\ \text{(syst.)} = 1.000\ ^{+0.281}_{-0.240}$
μ_{Z+2HF}	$1.000\ ^{+0.036}_{-0.036}\ \text{(stat.)}\ ^{+0.151}_{-0.128}\ \text{(syst.)} = 1.000\ ^{+0.155}_{-0.133}$

Table 8.13: The fitted signal cross and Z+HF normalizations obtained from Asimov fit of Neural Network output for the best binning choice (19 bins from -1.0 to 0.9).

8.7.2 Asimov fit results using Boosted Decision Tree

Binning optimization

The binning optimization for the fit of the BDT output was similar to the optimization of the NN output binning. The same binning choices have been tested, leading to the same conclusion: The best binning choice is 19 bins from -1.0 to 0.9. Table 8.14 summarizes the fitted signal cross section and its uncertainty obtained from the Asimov fit of the BDT output distributions for the various binning choices.

Fit results for the final binning choice

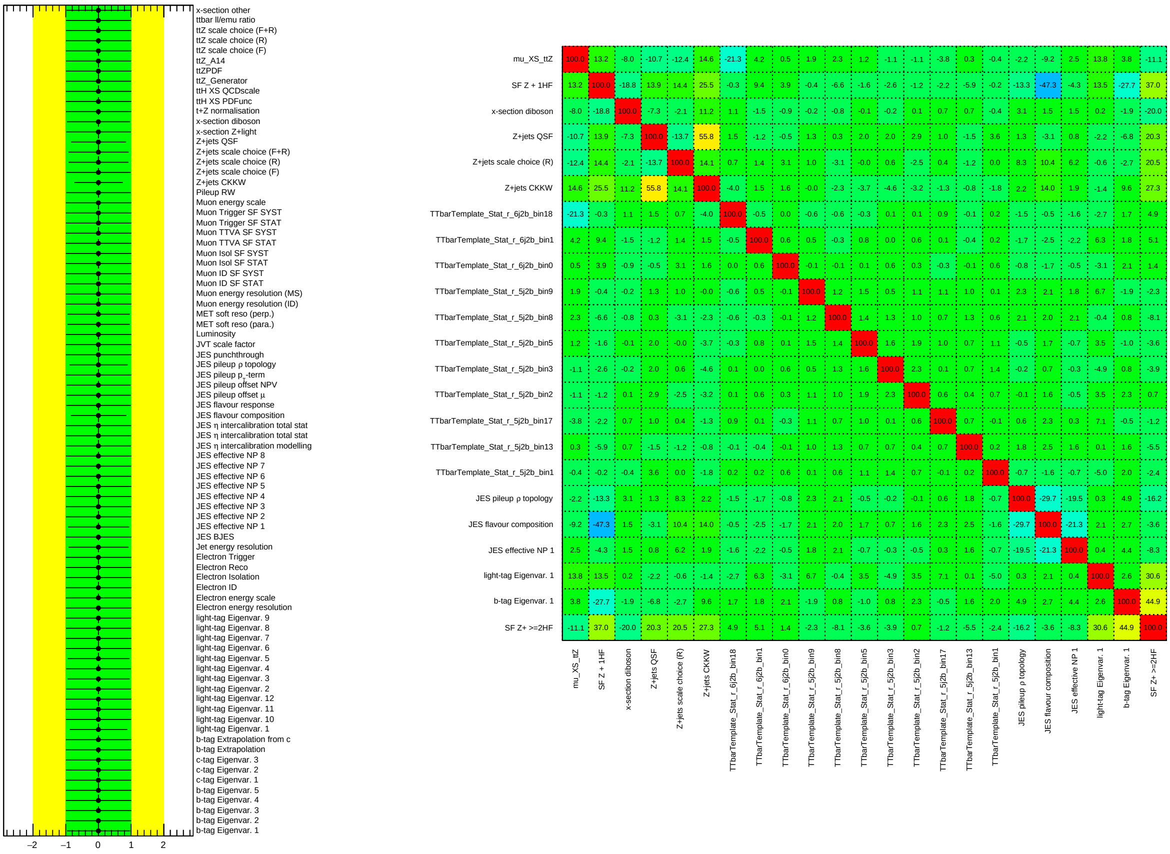
Fitting the Boosted Decision Tree output distributions showed on the bottom of Figure 8.6, the results summarized in Table 8.15 have been obtained. The plots showing nuisance parameters and uncertainties on their post-fit values, together with the correlation matrix of the fit parameters, are shown in Figure 8.10. The effects of 20 leading systematics uncertainties are shown in Figure 8.11. The expected signal significance (exclusion of the background only hypothesis) is 3.8 σ.

Number of bins	Fitted signal cross section
1	$1.000\ ^{+1.997}_{-2.008}$
2	$1.000\ ^{+0.533}_{-0.535}$
5	$1.000\ ^{+0.348}_{-0.325}$
10	$1.000\ ^{+0.320}_{-0.294}$
19	$1.000\ ^{+0.312}_{-0.288}$

Table 8.14: The fitted signal cross section and its uncertainty, as the ratio between fitted value and Monte Carlo prediction, extracted from the Asimov fit of the BDT output distributions for various binning choices. The uncertainties include both statistical and systematic uncertainty.

Parameter	Value
$\mu_{t\bar{t}Z}$	$1.000\ ^{+0.207}_{-0.201}$ (stat.) $^{+0.233}_{-0.207}$ (syst.) $= 1.000\ ^{+0.312}_{-0.288}$
μ_{Z+1HF}	$1.000\ ^{+0.053}_{-0.053}$ (stat.) $^{+0.280}_{-0.238}$ (syst.) $= 1.000\ ^{+0.285}_{-0.244}$
μ_{Z+2HF}	$1.000\ ^{+0.036}_{-0.035}$ (stat.) $^{+0.149}_{-0.127}$ (syst.) $= 1.000\ ^{+0.153}_{-0.133}$

Table 8.15: The fitted signal cross and Z+HF normalizations obtained from Asimov fit of the BDT output for the best binning choice (19 bins from -1.0 to 0.9).



	mu_XS_ttZ	SF Z + 1HF	x-section diboson	Z+jets QSF	Z+jets scale choice (R)	Z+jets CKKW	TTbarTemplate_Stat_r_6j2b_bin18	TTbarTemplate_Stat_r_6j2b_bin1	TTbarTemplate_Stat_r_6j2b_bin0	TTbarTemplate_Stat_r_5j2b_bin9	TTbarTemplate_Stat_r_5j2b_bin8	TTbarTemplate_Stat_r_5j2b_bin5	TTbarTemplate_Stat_r_5j2b_bin3	TTbarTemplate_Stat_r_5j2b_bin2	TTbarTemplate_Stat_r_5j2b_bin17	TTbarTemplate_Stat_r_5j2b_bin13	TTbarTemplate_Stat_r_5j2b_bin1	JES pileup ρ topology	JES flavour composition	JES effective NP 1	light-tag Eigenvar. 1	b-tag Eigenvar. 1	SF Z+ >=2HF
mu_XS_ttZ	100.0	13.2	-8.0	-10.7	-12.4	14.6	-21.3	4.2	0.5	1.9	2.3	1.2	-1.1	-1.1	-3.8	0.3	-0.4	-2.2	-9.2	2.5	13.8	3.8	-11.1
SF Z + 1HF	13.2	100.0	-18.8	13.9	14.4	25.5	-0.3	9.4	3.9	-0.4	-6.6	-1.6	-2.6	-1.2	-2.2	-5.9	-0.2	-13.3	-47.3	-4.3	13.5	-27.7	37.0
x-section diboson	-8.0	-18.8	100.0	-7.3	-2.1	11.2	1.1	-1.5	-0.9	-0.2	-0.8	-0.1	-0.2	0.1	0.7	0.7	-0.4	3.1	1.5	1.5	0.2	-1.9	-20.0
Z+jets QSF	-10.7	13.9	-7.3	100.0	-13.7	55.8	1.5	-1.2	-0.5	1.3	0.3	2.0	2.0	2.9	1.0	-1.5	3.6	1.3	-3.1	0.8	-2.2	-6.8	20.3
Z+jets scale choice (R)	-12.4	14.4	-2.1	-13.7	100.0	14.1	0.7	1.4	3.1	1.0	-3.1	-0.0	0.6	-2.5	0.4	-1.2	0.0	8.3	10.4	6.2	-0.6	-2.7	20.5
Z+jets CKKW	14.6	25.5	11.2	55.8	14.1	100.0	-4.0	1.5	1.6	-0.0	-2.3	-3.7	-4.6	-3.2	-1.3	-0.8	-1.8	2.2	14.0	1.9	-1.4	9.6	27.3
TTbarTemplate_Stat_r_6j2b_bin18	-21.3	-0.3	1.1	1.5	0.7	-4.0	100.0	-0.5	0.0	-0.6	-0.6	-0.3	0.1	0.1	0.9	-0.1	0.2	-1.5	-0.5	-1.6	-2.7	1.7	4.9
TTbarTemplate_Stat_r_6j2b_bin1	4.2	9.4	-1.5	-1.2	1.4	1.5	-0.5	100.0	0.6	0.5	-0.3	0.8	0.0	0.6	0.1	-0.4	0.2	-1.7	-2.5	-2.2	6.3	1.8	5.1
TTbarTemplate_Stat_r_6j2b_bin0	0.5	3.9	-0.9	-0.5	3.1	1.6	0.0	0.6	100.0	-0.1	-0.1	0.1	0.6	0.3	-0.3	-0.1	0.6	-0.8	-1.7	-0.5	-3.1	2.1	1.4
TTbarTemplate_Stat_r_5j2b_bin9	1.9	-0.4	-0.2	1.3	1.0	-0.0	-0.6	0.5	-0.1	100.0	1.2	1.5	0.5	1.1	1.1	1.0	0.1	2.3	2.1	1.8	6.7	-1.9	-2.3
TTbarTemplate_Stat_r_5j2b_bin8	2.3	-6.6	-0.8	0.3	-3.1	-2.3	-0.6	-0.3	-0.1	1.2	100.0	1.4	1.3	1.0	0.7	1.3	0.6	2.1	2.0	2.1	-0.4	0.8	-8.1
TTbarTemplate_Stat_r_5j2b_bin5	1.2	-1.6	-0.1	2.0	-0.0	-3.7	-0.3	0.8	0.1	1.5	1.4	100.0	1.6	1.9	1.0	0.7	1.1	-0.5	1.7	-0.7	3.5	-1.0	-3.6
TTbarTemplate_Stat_r_5j2b_bin3	-1.1	-2.6	-0.2	2.0	0.6	-4.6	0.1	0.0	0.6	0.5	1.3	1.6	100.0	2.3	0.1	0.7	1.4	-0.2	0.7	-0.3	-4.9	0.8	-3.9
TTbarTemplate_Stat_r_5j2b_bin2	-1.1	-1.2	0.1	2.9	-2.5	-3.2	0.1	0.6	0.3	1.1	1.0	1.9	2.3	100.0	0.6	0.4	0.7	-0.1	1.6	-0.5	3.5	2.3	0.7
TTbarTemplate_Stat_r_5j2b_bin17	-3.8	-2.2	0.7	1.0	0.4	-1.3	0.9	0.1	-0.3	1.1	0.7	1.0	0.1	0.6	100.0	0.7	-0.1	0.6	2.3	0.3	7.1	-0.5	-1.2
TTbarTemplate_Stat_r_5j2b_bin13	0.3	-5.9	0.7	-1.5	-1.2	-0.8	-0.1	-0.4	-0.1	1.0	1.3	0.7	0.7	0.4	0.7	100.0	0.2	1.8	2.5	1.6	0.1	1.6	-5.5
TTbarTemplate_Stat_r_5j2b_bin1	-0.4	-0.2	-0.4	3.6	0.0	-1.8	0.2	0.2	0.6	0.1	0.6	1.1	1.4	0.7	-0.1	0.2	100.0	-0.7	-1.6	-0.7	-5.0	2.0	-2.4
JES pileup ρ topology	-2.2	-13.3	3.1	1.3	8.3	2.2	-1.5	-1.7	-0.8	2.3	2.1	-0.5	-0.2	-0.1	0.6	1.8	-0.7	100.0	-29.7	-19.5	0.3	4.9	-16.2
JES flavour composition	-9.2	-47.3	1.5	-3.1	10.4	14.0	-0.5	-2.5	-1.7	2.1	2.0	1.7	0.7	1.6	2.3	2.5	-1.6	-29.7	100.0	-21.3	2.1	2.7	-3.6
JES effective NP 1	2.5	-4.3	1.5	0.8	6.2	1.9	-1.6	-2.2	-0.5	1.8	2.1	-0.7	-0.3	-0.5	0.3	1.6	-0.7	-19.5	-21.3	100.0	0.4	4.4	-8.3
light-tag Eigenvar. 1	13.8	13.5	0.2	-2.2	-0.6	-1.4	-2.7	6.3	-3.1	6.7	-0.4	3.5	-4.9	3.5	7.1	0.1	-5.0	0.3	2.1	0.4	100.0	2.6	30.6
b-tag Eigenvar. 1	3.8	-27.7	-1.9	-6.8	-2.7	9.6	1.7	1.8	2.1	-1.9	0.8	-1.0	0.8	2.3	-0.5	1.6	2.0	4.9	2.7	4.4	2.6	100.0	44.9
SF Z+ >=2HF	-11.1	37.0	-20.0	20.3	20.5	27.3	4.9	5.1	1.4	-2.3	-8.1	-3.6	-3.9	0.7	-1.2	-5.5	-2.4	-16.2	-3.6	-8.3	30.6	44.9	100.0

Figure 8.10: Nuisance parameters plot (left) and correlation matrix (right) for the Asimov fit of Boosted Decision Tree output distribution.

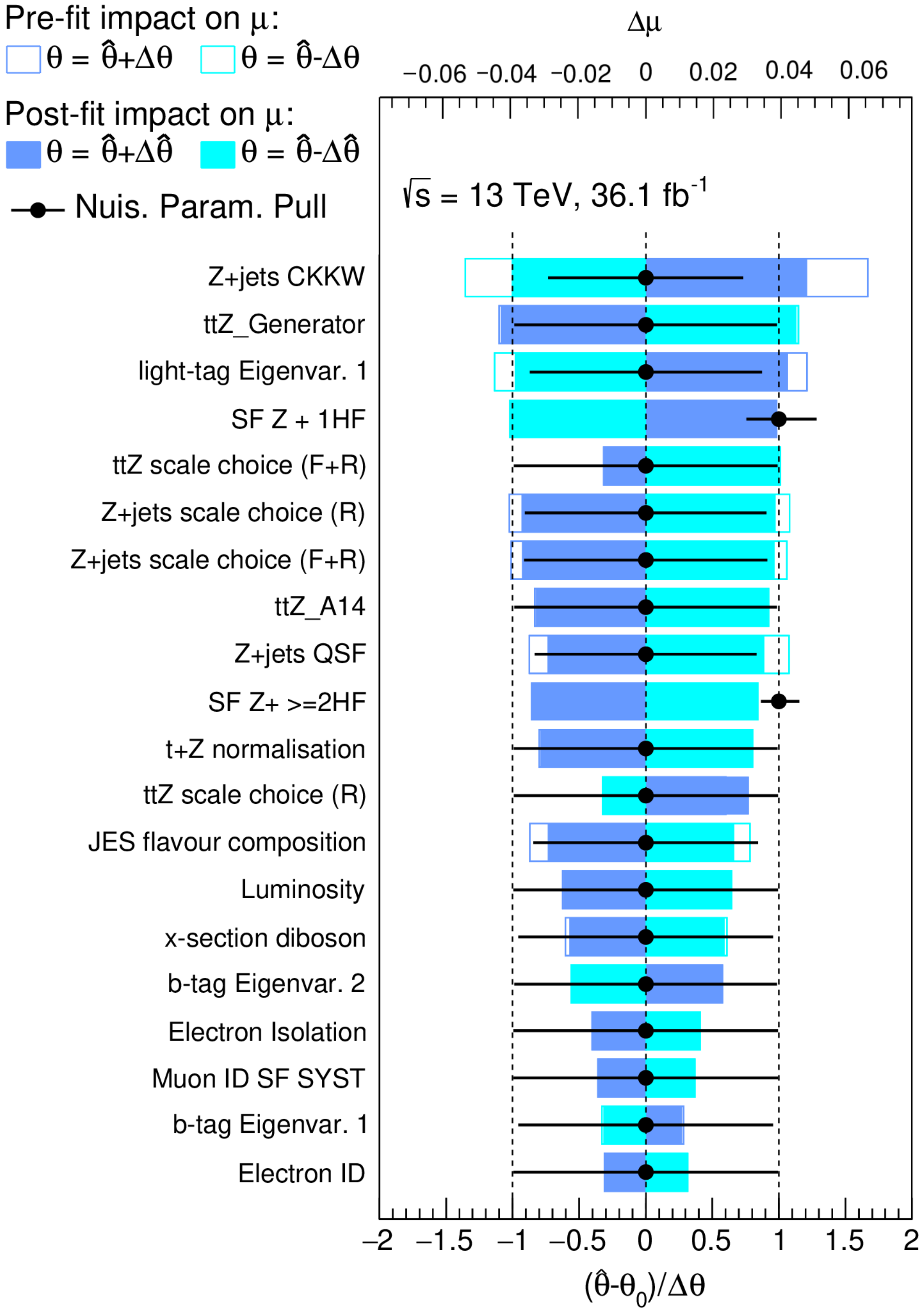


Figure 8.11: Effects of 20 leading systematic uncertainties in Asimov fit of BDT output distribution. The leading systematics uncertainty is shower-matching scale for Z+jets samples, followed by uncertainty related to matrix element modelling in $t\bar{t}Z$ and b-tagging systematics related to misstag rate for light jets. The expected mean value of nuisance parameters for systematics is zero, while the expected mean value of Z+1HF and Z+2HF scale factors is one.

8.7.3 Asimov fit conclusion

Comparing the expected relative uncertainty on the signal cross section obtained from NN fit (Table 8.13) and BDT fit (Table 8.15), the BDT has been chosen as the final choice of the MVA used for the analysis. The expected signal cross section uncertainty is lower in the BDT case. Correlation matrices of the systematic uncertainties are similar for the both MVA methods.

The 19 bins of the same width, covering the BDT output interval from -1 to 0.9, with the last bin including overflow, have been chosen as the final binning choice.

The expected signal significance is 3.8 σ. The expected uncertainty of the measurement is summarized in Table 8.15.

8.8 Fit to data

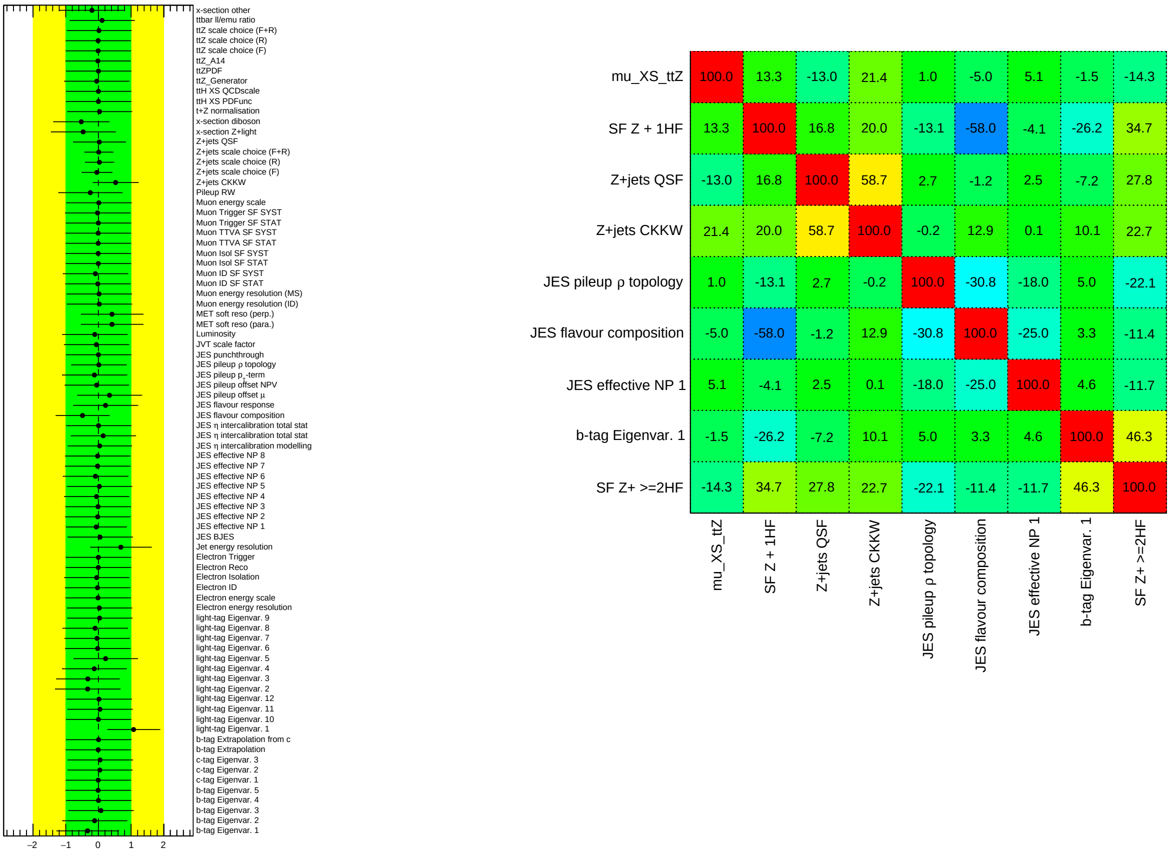
Performing the fit to data on the BDT output distributions, showed in Figure 8.6 - bottom line of plots, in the three signal regions of the dilepton channel, the signal and background normalizations summarized in Table 8.16 have been obtained. The observed signal significance is 3.0 σ (3.8 σ is expected). The fitted $\mu_{t\bar{t}Z}$ value corresponds to the cross section:

$$\sigma_{t\bar{t}Z}^{2l,measured} = \sigma_{t\bar{t}Z}^{theory} \times \mu_{t\bar{t}Z} = 0.636^{+0.152}_{-0.148}(\text{stat.})^{+0.203}_{-0.190}(\text{syst.})\ \text{pb} = 0.636^{+0.254}_{-0.241}\ \text{pb}, \quad (8.3)$$

Parameter	Value
$\mu_{t\bar{t}Z}$	$0.721\ ^{+0.173}_{-0.168}\ (\text{stat.})\ ^{+0.230}_{-0.216}\ (\text{syst.}) = 0.721\ ^{+0.288}_{-0.273}$
μ_{Z+1HF}	$1.072\ ^{+0.270}_{-0.234}$
μ_{Z+2HF}	$1.084\ ^{+0.148}_{-0.132}$

Table 8.16: The fitted signal and Z+HF normalizations obtained from fit to data of the BDT output. The Monte Carlo prediction of the cross section is taken as the unit of μ.

The plot with the post-fit values of nuisance parameters and their uncertainties, together with the correlation matrix of the fit parameters are shown in Figure 8.12. The ranking plot for the systematic uncertainties, showing the effects of 20 leading systematics is shown in Figure 8.13.



	mu_XS_ttZ	SF Z + 1HF	Z+jets QSF	Z+jets CKKW	JES pileup ρ topology	JES flavour composition	JES effective NP 1	b-tag Eigenvar. 1	SF Z+ >=2HF
mu_XS_ttZ	100.0	13.3	-13.0	21.4	1.0	-5.0	5.1	-1.5	-14.3
SF Z + 1HF	13.3	100.0	16.8	20.0	-13.1	-58.0	-4.1	-26.2	34.7
Z+jets QSF	-13.0	16.8	100.0	58.7	2.7	-1.2	2.5	-7.2	27.8
Z+jets CKKW	21.4	20.0	58.7	100.0	-0.2	12.9	0.1	10.1	22.7
JES pileup ρ topology	1.0	-13.1	2.7	-0.2	100.0	-30.8	-18.0	5.0	-22.1
JES flavour composition	-5.0	-58.0	-1.2	12.9	-30.8	100.0	-25.0	3.3	-11.4
JES effective NP 1	5.1	-4.1	2.5	0.1	-18.0	-25.0	100.0	4.6	-11.7
b-tag Eigenvar. 1	-1.5	-26.2	-7.2	10.1	5.0	3.3	4.6	100.0	46.3
SF Z+ >=2HF	-14.3	34.7	27.8	22.7	-22.1	-11.4	-11.7	46.3	100.0

Figure 8.12: Nuisance parameters plot (left) and correlation matrix (right) for the Asimov fit of the Boosted Decision Tree output distribution.

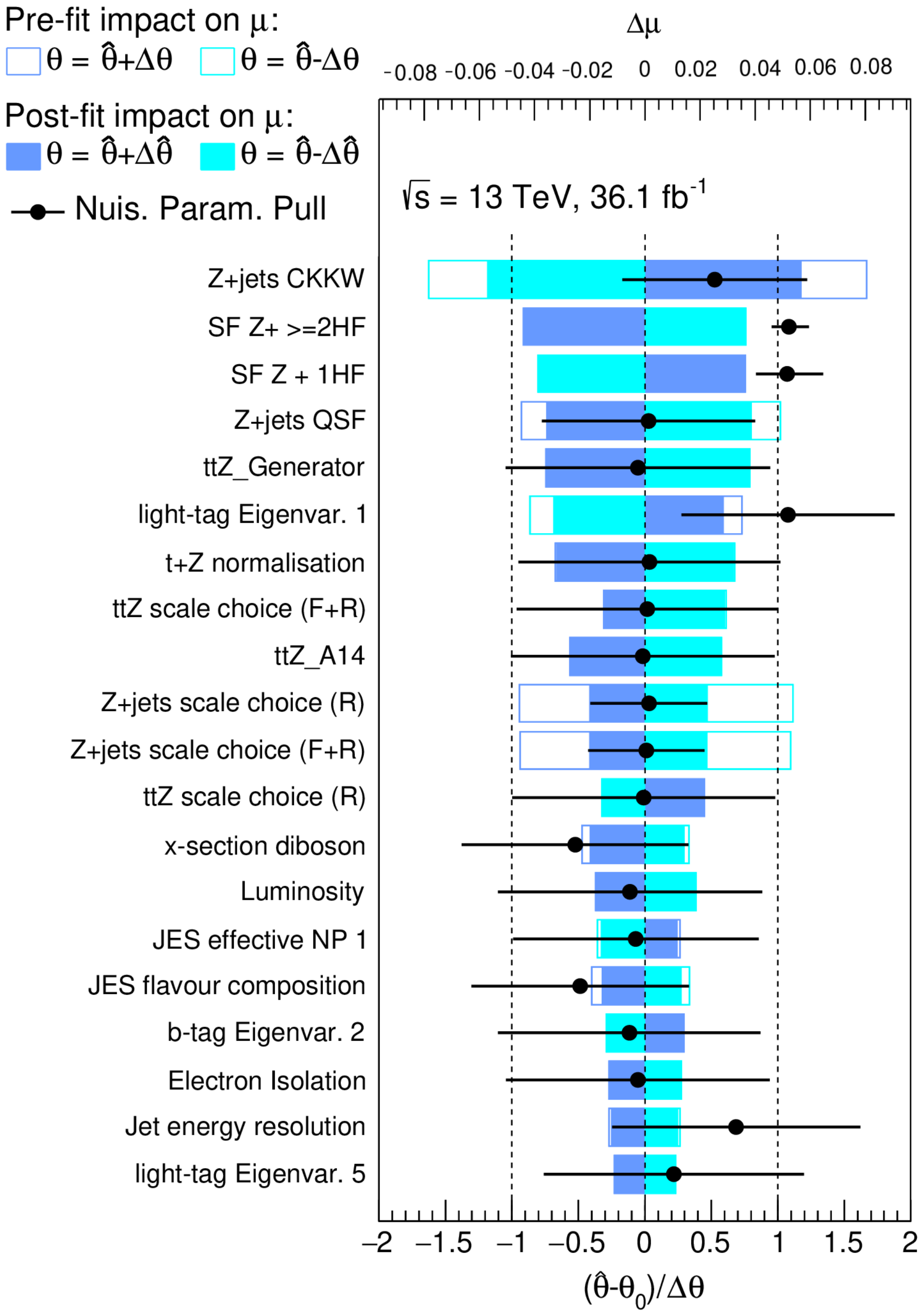


Figure 8.13: Effects of 20 leading systematic uncertainties in the fit of BDT output distribution. The leading systematic uncertainty is the shower matching scale for Z+jets background, followed by the normalizations of Z+1 HF and Z+2HF backgrounds. The expected mean value Z+HF scale factors is one, while expected mean value of the other parameters is zero. The pull in light-tag Eigenvar.1 nuisance parameter is caused by Z+0HF background in 2ℓ-Z-6j1b region, where the Monte Carlo overestimates the data. The pull corrects this discrepancy.

8.9 Results from the $t\bar{t}Z$ trilepton and tetralepton channels, and from $t\bar{t}W$ analysis

The aim of the total cross section measurement was to measure the $t\bar{t}Z$ and $t\bar{t}W$ cross sections in the combined fit.

Two $t\bar{t}W$ channels are used in the analysis. The first $t\bar{t}W$ decay channel is the same-sign dilepton, with the W boson decaying leptonically and the lepton+jets decay of the top-quark pair, where the lepton from top-quark pair and the lepton from the associated W boson have the same sign of charge. The other channel is the trilepton channel, with the W boson decaying leptonically and the top-quark pair decaying dileptonically.

Three channels sensitive to the $t\bar{t}Z$ contribution are used in the fit: 2ℓ, 3ℓ and 4ℓ channels.

The results obtained from the fit in the individual 3ℓ and 4ℓ $t\bar{t}Z$ channels will be summarized in this section, as well as the results of the $t\bar{t}W$ fit.[3]

The results of the combined fit, for all the $t\bar{t}Z$ and $t\bar{t}W$ channels, fitting all the channels together are summarized in Section 8.9.4 and they represent the final result of the cross section measurement, published also in Reference [14].

8.9.1 Trilepton channel analysis

The trilepton channel targets the $t\bar{t}Z$ decay branch with the Z boson decaying leptonically and top-quark pair decaying into lepton + jets. Three signal regions target on-shell Z boson contribution and one signal region (labelled as ”noZ”) targets off-shell Z boson contribution. A WZ production is the leading background in the trilepton channel. In order to check its modelling and obtain its cross section, a dedicated WZ control region (3ℓ-Z-0b3j-CR) is used. Definitions of the four signal regions of the trilepton channel, together with the WZ control region, are summarized in Table 8.17.

The event yields obtained from the simulation and observed numbers of events in the individual regions of the trilepton channel are summarized in Table 8.18.

Contrary to the dilepton channel, the fit in the trilepton channel uses one bin per region, fitting the event yields summarized in Table 8.18. The background rate in the trilepton channel is significantly lower compared to the dilepton channel, which

---

[3] The author of the thesis worked only on the dilepton channel in the 2015 + 2016 data analysis. However, for an overall understanding and overview of the analysis the author considers it to be important to bring also a summary of results obtained from the other channels.

Variable	3ℓ-Z-1b4j	3ℓ-Z-2b3j	3ℓ-Z-2b4j	3ℓ-no$Z$-2b4j	3ℓ-Z-0b3j-CR
Triggers	Defined in Table 8.1				
Leptons	$= 3$				
p_{T} (1st lepton)	> 27 GeV				
p_{T} (2nd and 3rd lepton)	> 20 GeV				
One OSSF lepton pair		required		required	required
$\lvert M_{\ell\ell}^{OSSF}$ - $M_Z\rvert < 10$ GeV		required		vetoed	required
n_{jets}	≥ 4	=3	≥ 4	≥ 4	=3
$n_{b-\mathrm{jets}}$	=1	≥ 2	≥ 2	≥ 2	=0

Table 8.17: Summary of the event selection in four signal regions and one control region of the trilepton channel. OSSF stands for opposite-sign, same-flavour.

	3ℓ-Z-1b4j	3ℓ-Z-2b3j	3ℓ-Z-2b4j	3ℓ-no$Z$-2b4j	3ℓ-Z-0b3j-CR
$t\bar{t}Z$	44.5 ± 5.32	57.0 ± 10.0	16.6 ± 3.30	12.7 ± 2.30	5.10 ± 1.20
$t\bar{t}W$	0.491 ± 0.293	0.519 ± 0.336	0.826 ± 0.424	3.67 ± 1.91	0.184 ± 0.102
$lll\nu$	36.7 ± 12.0	7.05 ± 3.82	3.32 ± 1.62	1.05 ± 0.585	211 ± 22.4
$llll$	3.00 ± 0.616	0.534 ± 0.102	0.679 ± 0.264	0.328 ± 0.189	11.5 ± 1.99
tZ	2.89 ± 0.958	3.41 ± 1.13	3.66 ± 1.19	0.315 ± 0.138	1.42 ± 0.519
tWZ	6.74 ± 1.90	5.77 ± 1.99	2.07 ± 0.445	0.667 ± 0.301	2.17 ± 0.655
$t\bar{t}H$	1.25 ± 0.193	1.42 ± 0.216	0.515 ± 0.0873	4.87 ± 0.656	0.110 ± 0.0344
Other	0.305 ± 0.161	0.375 ± 0.368	0.868 ± 1.04	2.13 ± 1.09	1.53 ± 1.17
Fakes	6.53 ± 2.85	4.01 ± 2.18	1.17 ± 1.15	3.16 ± 1.99	0 ± 0
$\gamma + X$	1.35 ± 1.15	0.492 ± 0.590	0.623 ± 0.649	4.88 ± 2.49	0 ± 0
Total	104 ± 14.2	80.6 ± 10.2	30.4 ± 4.47	33.8 ± 4.70	238 ± 23.0
Observed	86	78	45	37	211
Data/MC	0.83 ± 0.14	0.97 ± 0.16	1.48 ± 0.31	1.09 ± 0.24	0.89 ± 0.11

Table 8.18: Events yields in four signal regions and one control region of the trilepton channel.

makes the fit sensitive enough even with one bin per region. There is no need to use a multivariate analysis. The $t\bar{t}Z$ and WZ cross sections obtained from the fit are summarized in Table 8.19. The signal significance was found to be 6.8 σ. The $\mu_{t\bar{t}Z}$ fitted in the trilepton channel corresponds to the following value of the $t\bar{t}Z$ cross section:

$$\sigma_{t\bar{t}Z}^{3l,measured} = 0.95 \pm 0.11(\text{stat.}) \pm 0.11(\text{syst.})\ \text{pb} = 0.95^{+0.16}_{-0.15}\ \text{pb}, \tag{8.4}$$

Parameter	Value
$\mu_{t\bar{t}Z}$	$1.08\ ^{+0.12}_{-0.12}\ (\text{stat.})\ ^{+0.13}_{-0.12}\ (\text{syst.}) = 1.08\ ^{+0.18}_{-0.17}$
μ_{WZ}	$0.93 \pm 0.07\ (\text{stat.}) \pm 0.10\ (\text{syst.}) = 0.93 \pm 0.12$

Table 8.19: The fitted signal and WZ normalizations obtained from fit to data in the trilepton channel. The Monte Carlo prediction of the cross section is taken as the unit of μ.

The ranking plot for the systematic uncertainties, showing the effects of 15 leading systematic uncertainties, is shown in Figure 8.14.

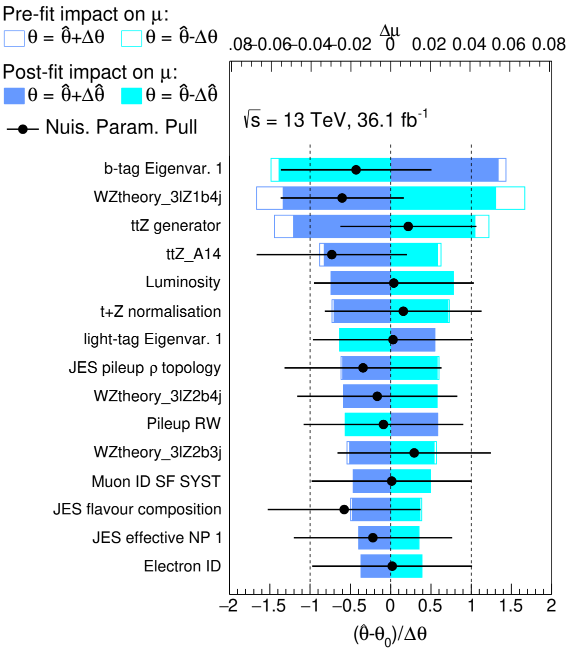


Figure 8.14: Effects of 15 leading systematic uncertainties in the trilepton channel fit. The leading systematic uncertainty is b-tagging efficiency for b-jets, followed by uncertainty of the WZ scale factor extrapolation from WZ control region to 3ℓ-Z-1b4j signal region. The third systematic uncertainty is the matrix element modelling of $t\bar{t}Z$.

8.9.2 Tetralepton channel analysis

The tetralepton channel targets the case with the Z boson decaying leptonically and both W bosons from the top-quark pair also decaying leptonically.

In the final state, there are 2 b quarks, 4 leptons (OSSF lepton pair from Z boson and the other opposite-sign pair from the top-quark pair) and two neutrinos from the top-quark pair. If all the objects are reconstructed correctly, there are two b-jets, one Z-like lepton pair (same flavour, opposite sign, $M_{\ell\ell}$ close to the Z-mass), another OS lepton pair and missing transverse energy E_{T}^{miss}.

In the event reconstruction, the OSSF lepton pair with the invariant mass closest to the Z-mass is considered to be the lepton pair from the Z-boson decay and it is labelled as Z1. The other lepton pair is considered to originate from the top-quark pair and it is labelled Z2. The signal regions are split according to the flavour of the Z2 leptons into different-flavour (DF) and same-flavour (SF) signal regions.

The main background in the tetralepton channel is $ZZ \to \ell\ell\ell\ell$. This background affects mostly the same-flavour signal regions, however it also contributes to different flavour signal regions through $Z \to \tau^+\tau^- \to e^{\mp}\nu_{e^{\pm}}\nu_{\tau^{\mp}}\mu^{\pm}\nu_{\mu^{\mp}}\nu_{\tau^{\pm}}$ decays. The ZZ background is reduced requiring a minimal number of jets and b-jets and applying M_{Z2} and E_{T}^{miss} cuts.

Another important background in the tetralepton channel is tWZ, which is irreducible, since the particles in the final state are similar to the tetralepton $t\bar{t}Z$. The only difference is one b quark, however, in the tetralepton channel, also regions with exactly one b-jet are defined.

Associated production of the Higgs boson and the top-quark pair ($t\bar{t}H$) and fake leptons also contribute to the background.

In order to target the events from the $t\bar{t}Z$ tetralepton channel, the following selection common to all 4ℓ signal and control regions is applied:

- exactly 4 leptons, sum of their charges equal to zero
- at least one OSSF lepton pair
- all OS lepton pairs must satisfy the condition $M_{\ell\ell} > 15$ GeV in order to suppress the background from leptonically decaying resonances and photon conversions
- $p_{\mathrm{T}}^{1\ \mathrm{lep}} > 27$ GeV

In addition to the already mentioned common selection to all signal and control regions, further cuts are applied in order to define the four signal regions and one control region, defined in Table 8.20. The composition of the background is different between same-flavour and different-flavour regions. In order to suppress the ZZ background, contributing mainly to the same-flavour signal regions, combined cuts on the missing transverse energy E_{T}^{miss} and invariant mass of the Z2 lepton pair are applied. If the invariant mass of the Z2 pair is compatible with the Z-boson mass, higher cut on the E_{T}^{miss} is applied in order to cut off the ZZ background. If the Z2 pair is not compatible with the Z-boson mass, the cut is relaxed.

Region	Z_2 leptons	$p_{\mathrm{T}4}$	$p_{\mathrm{T}34}$	$\lvert m_{\ell\ell} - m_{Z_2}\rvert$	E_{T}^{miss}	$N_{b\text{-jets}}$
4ℓ-DF-1b	$e^{\pm}\mu^{\mp}$	> 7 GeV	> 35 GeV	-	-	1
4ℓ-DF-2b	$e^{\pm}\mu^{\mp}$	> 10 GeV	-	-	-	≥ 2
4ℓ-SF-1b	$e^{\pm}e^{\mp}, \mu^{\pm}\mu^{\mp}$	> 7 GeV	> 25 GeV	{ > 10 GeV	> 40 GeV }	1
				< 10 GeV	> 80 GeV	
4ℓ-SF-2b	$e^{\pm}e^{\mp}, \mu^{\pm}\mu^{\mp}$	> 10 GeV	-	{ > 10 GeV	- }	≥ 2
				< 10 GeV	> 40 GeV	
4ℓ-ZZ-CR	$e^{\pm}e^{\mp}, \mu^{\pm}\mu^{\mp}$	> 7 GeV	-	< 10 GeV	[20 GeV; 40 GeV]	-

Table 8.20: Definitions of the four signal regions and ZZ control region of the tetralepton channel. $p_{\mathrm{T}34}$ stands for scalar sum of the third and fourth lepton p_{T}. $p_{\mathrm{T}4}$ stands for transverse momentum of the fourth lepton. The leptons are ordered by their p_{T} decreasingly.

The event yields obtained from the simulation and observed numbers of events in the individual regions of the tetralepton channel are summarized in Table 8.21.

Similarly to the fit in the trilepton channel, one bin per region is used in the tetralepton channel, fitting the event yields summarized in Table 8.21. The signal and ZZ normalizations obtained from the fit are summarized in Table 8.22. The signal significance is 5.7 σ. The $\mu_{t\bar{t}Z}$ fitted in the tetralepton channel corresponds to the following value of the $t\bar{t}Z$ cross section:

$$\sigma_{t\bar{t}Z}^{4l,measured} = 1.07^{+0.25}_{-0.22}(\text{stat.})^{+0.10}_{-0.11}(\text{syst.})\text{ pb} = 1.07^{+0.26}_{-0.25}\text{ pb}, \tag{8.5}$$

The ranking plot for the systematic uncertainties, showing effects of 15 leading systematics, is shown in Figure 8.15.

	4ℓ-SF-1b	4ℓ-SF-2b	4ℓ-DF-1b	4ℓ-DF-2b	4ℓ-ZZ-CR
$t\bar{t}Z$	6.56 ± 0.39	6.14 ± 0.56	7.38 ± 0.42	5.99 ± 0.74	0.13 ± 0.04
ZZ	2.27 ± 0.75	1.06 ± 0.52	0.19 ± 0.06	0 ± 0	145.93 ± 20.84
tWZ	1.60 ± 0.46	0.55 ± 0.26	1.57 ± 0.30	0.51 ± 0.26	0 ± 0
$t\bar{t}H$	0.58 ± 0.07	0.62 ± 0.09	0.68 ± 0.08	0.57 ± 0.08	0 ± 0
Other	0.12 ± 0.03	0.09 ± 0.03	0.20 ± 0.05	0.10 ± 0.02	0.51 ± 0.36
Fakes	1.84 ± 0.82	1.23 ± 0.62	0.93 ± 0.16	0.39 ± 0.11	7.35 ± 9.14
Total	12.97 ± 1.33	9.69 ± 1.17	10.95 ± 0.65	7.57 ± 0.88	153.92 ± 24.71
Observed	18	14	11	5	144
Data/MC	1.39 ± 0.36	1.44 ± 0.42	1.00 ± 0.31	0.66 ± 0.31	0.94 ± 0.17

Table 8.21: Events yields in the four signal regions and one control region of the tetralepton channel.

Parameter	Value
$\mu_{t\bar{t}Z}$	$1.21\ ^{+0.28}_{-0.25}\ \text{(stat.)}\ ^{+0.11}_{-0.12}\ \text{(syst.)} = 1.21\ ^{+0.30}_{-0.28}$
μ_{ZZ}	0.94 ± 0.18

Table 8.22: The fitted signal and ZZ normalizations obtained from fit to data in the tetralepton channel. The Monte Carlo prediction of the cross section is taken as the unit of μ.

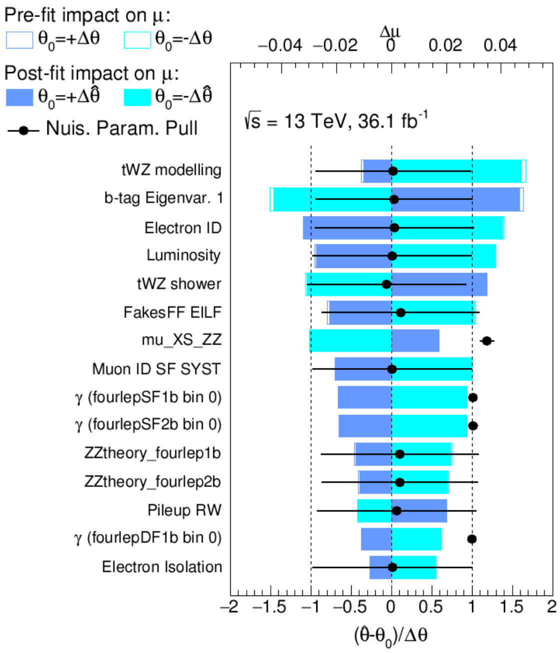


Figure 8.15: Effects of 15 leading systematic uncertainties in the tetralepton channel fit. The leading systematic uncertainty is the matrix element modelling of the tWZ background, followed by the b-tagging efficiency for b-jets and electron identification efficiency. The gammas stand for statistical uncertainty of Monte Carlo templates, it is a ratio between the fitted value (providing the best likelihood) and the value predicted by the Monte Carlo (which suffers from statistical fluctuations). The gammas are expected to have mean values equal to one. The mu_XS_ZZ stands for ZZ normalization (expected mean value is one).

8.9.3 $t\bar{t}W$ signal regions

The $t\bar{t}W$ cross section is measured in two decay channels: same-sign dilepton and trilepton. The same-sign dilepton channel is further split into 12 signal regions, based on the type of the leptons (ee, $e\mu$, $\mu\mu$), b-jet multiplicity (=1b, $\geq 2b$) and sign of the lepton charge (pp for plus-plus and mm for minus-minus). The trilepton $t\bar{t}W$ channel is further split into 4 signal regions, based on the b-jet multiplicity and sign of the lepton charges sum.

The splitting based on the sign of the lepton charge is motivated by an asymmetry in $t\bar{t}W$ production. Since there are pp collisions at the LHC, the environment is not charge symmetric. Contrary to the $t\bar{t}Z$ which is produced either by gg fusion or $q\bar{q}$ annihilation, where the initial state partons are charge symmetric, the $t\bar{t}W$ is produced only in $q'\bar{q}$ interaction, where both q' and $\bar{q}$ have the same sign of the electric charge. Since there are two u quarks, but only one d valence quark in the proton, the production of $t\bar{t}W^+$ has a higher cross section compared to the $t\bar{t}W^-$ production. While the signal is charge asymmetric, the backgrounds are almost charge symmetric. The splitting based on the lepton charge helps to reach a higher signal sensitivity.

The event selection for the dilepton same-sign signal regions is summarized in Table 8.23, the selection applied in trilepton $t\bar{t}W$ signal regions is summarized in Table 8.24.

Requirement	Region
$n_{b-\mathrm{jets}}$	==1 (for the 6 "_1b" regions), ≥ 2 (for the 6 "_2b" regions)
$E_{\mathrm{T}}^{\mathrm{miss}}$	> 20 GeV for the 2 "$\mu\mu$_2b" regions and > 40 GeV for the others
H_T	> 240 GeV for all regions
p_{T} (1st lepton)	> 27 GeV for all regions
p_{T} (2nd lepton)	> 27 GeV for all regions
nJets	≥ 2 for the 2 "mumu_2b" regions and ≥ 4 for the others
Z-veto	for the 8 regions with same flavour leptons ("mumu" and "ee")

Table 8.23: Summary of the event selection in the same-sign dilepton $t\bar{t}W$ signal regions. The Z-veto stands for vetoing a lepton pair compatible with Z-boson mass, $|M_{\ell\ell}^{OSSF}$ - $M_Z|$ $<$ 10 GeV, and it is applied in order to suppress the Z+jets contribution with misidentified sign of a lepton charge in the same-flavour regions. H_{T} stands for the scalar sum of leptons and jets p_{T}.

The event yields in $t\bar{t}W$ signal regions are summarized in Table 8.25 (3ℓ regions),

Variable	3Lp-noZ2b2j	3Lm-noZ2b2j	3Lp-noZ1b2j	3Lm-noZ1b2j
Leptons	==3			
Lepton p_{T}	> 27 GeV			
Z-like OSSF pair	vetoed			
Sum of lepton charges	$+1$	-1	$+1$	-1
H_{T}	> 0 GeV		> 240 GeV	
n_{jets}	2 or 3			
$n_{b-\mathrm{jets}}$	≥ 2	≥ 2	1	1

Table 8.24: Definition of the four trilepton $t\bar{t}W$ signal regions. ”noZ” in the region name stands for Z-like pair veto. It is applied in order to remove an overlap with trilepton $t\bar{t}Z$ signal regions and to suppress WZ and $t\bar{t}Z$ contribution.

Table 8.26 (2ℓ SS $1b$ regions) and Table 8.27 (2ℓ SS $2b$ regions).

	3Lp-noZ2b2j	3Lm-noZ2b2j	3Lp-noZ1b2j	3Lm-noZ1b2j
$t\bar{t}Z$	2.19 ± 0.49	2.05 ± 0.56	3.24 ± 0.55	3.07 ± 0.47
$t\bar{t}W$	5.97 ± 1.10	3.47 ± 1.81	6.32 ± 0.55	3.20 ± 0.37
WZ	0.58 ± 0.30	0.20 ± 0.12	6.22 ± 2.14	5.45 ± 2.01
tZ	0.19 ± 0.09	0.17 ± 0.08	1.00 ± 0.35	0.48 ± 0.17
tWZ	0 ± 0	0.10 ± 0.11	0.46 ± 0.17	0.62 ± 0.21
$t\bar{t}H$	1.02 ± 1.45	1.08 ± 1.44	1.35 ± 1.80	1.42 ± 1.83
Other	0.11 ± 0.08	0.13 ± 0.10	0.23 ± 0.14	0.18 ± 0.11
DDCF	0 ± 0	0 ± 0	0 ± 0	0 ± 0
$\gamma + X$	0.46 ± 0.80	1.96 ± 1.76	1.02 ± 0.83	1.48 ± 1.06
Fakes	2.92 ± 0.51	4.78 ± 0.83	6.41 ± 1.87	7.88 ± 2.77
Total	13.44 ± 3.01	13.93 ± 4.44	26.25 ± 5.23	23.77 ± 6.10
Observed	11	14	30	15
Data/MC	0.82± 0.31	1.00± 0.42	1.14± 0.31	0.63± 0.23

Table 8.25: The event yields in four trilepton $t\bar{t}W$ signal regions.

Performing the fit in 16 $t\bar{t}W$ signal regions, using one bin per region, the $t\bar{t}W$ normalization $\mu_{t\bar{t}W} = 1.41^{+0.34}_{-0.32}$ has been obtained, corresponding to the following value of $t\bar{t}W$ cross section:

$$\sigma_{t\bar{t}W}^{measured} = 0.85\ ^{+0.20}_{-0.19}\ \mathrm{pb} \tag{8.6}$$

The signal significance is 4.3 σ. The ranking plot of systematic uncertainties is shown in Figure 8.16

	$\mu\mu$-SSpp_1b	$\mu\mu$-SSmm_1b	eμSSpp_1b	eμ-SSmm_1b	ee-SSpp_1b	ee-SSmm_1b
$t\bar{t}Z$	1.56 ± 0.31	1.79 ± 0.31	3.81 ± 0.59	3.82 ± 0.60	1.39 ± 0.29	1.46 ± 0.27
$t\bar{t}W$	7.35 ± 0.78	3.73 ± 0.96	13.86 ± 1.28	7.55 ± 0.73	5.29 ± 0.53	2.90 ± 0.32
WZ	2.47 ± 1.90	1.18 ± 0.56	4.46 ± 2.22	3.03 ± 1.29	1.42 ± 0.64	1.03 ± 0.44
tZ	0.11 ± 0.07	0.07 ± 0.04	0.27 ± 0.14	0.22 ± 0.09	0.10 ± 0.06	0.10 ± 0.08
tWZ	0.18 ± 0.12	0.26 ± 0.17	0.52 ± 0.17	0.54 ± 0.20	0.21 ± 0.15	0.19 ± 0.11
$t\bar{t}H$	1.72 ± 0.27	1.81 ± 0.27	3.50 ± 2.18	3.86 ± 0.48	1.55 ± 0.25	1.50 ± 0.25
Other	0.32 ± 0.17	0.35 ± 0.19	0.65 ± 0.34	0.56 ± 0.29	0.36 ± 0.22	0.23 ± 0.13
DDCF	0.04 ± 0.00	0.04 ± 0.00	2.32 ± 0.25	2.29 ± 0.24	1.76 ± 0.20	1.77 ± 0.19
$\gamma + X$	0 ± 0	0 ± 0	4.40 ± 2.02	4.55 ± 2.34	5.18 ± 3.06	3.12 ± 1.89
Fakes	7.13 ± 2.19	5.18 ± 1.39	14.09 ± 3.89	8.18 ± 1.92	1.53 ± 1.10	1.37 ± 0.99
Total	20.89 ± 5.51	14.41 ± 3.93	47.88 ± 8.84	34.59 ± 5.17	18.81 ± 3.68	13.67 ± 2.63
Observed	15	19	46	43	27	14
Data/MC	0.72± 0.26	1.32± 0.47	0.96± 0.23	1.24± 0.27	1.44± 0.39	1.02± 0.34

Table 8.26: The event yields in six $t\bar{t}W$ dilepton same-sign $1b$ signal regions.

	$\mu\mu$-SSpp_2b	$\mu\mu$-SSmm_2b	eμ-SSpp_2b	eμ-SSmm_2b	ee-SSpp_2b	ee-SSmm_2b
$t\bar{t}Z$	3.19 ± 0.50	3.23 ± 0.49	4.56 ± 0.73	4.68 ± 0.76	1.71 ± 0.40	1.78 ± 0.35
$t\bar{t}W$	15.32 ± 1.84	8.09 ± 1.53	18.68 ± 1.88	9.60 ± 1.14	6.79 ± 0.97	3.43 ± 0.50
WZ	0.58 ± 0.28	0.31 ± 0.16	0.98 ± 0.56	0.33 ± 0.21	0.57 ± 0.47	0.16 ± 0.11
tZ	0.39 ± 0.25	0.14 ± 0.09	0.50 ± 0.22	0.15 ± 0.08	0.18 ± 0.08	0.05 ± 0.05
tWZ	0.35 ± 0.27	0.34 ± 0.15	0.40 ± 0.21	0.28 ± 0.13	0 ± 0	0 ± 0
$t\bar{t}H$	2.66 ± 0.37	2.94 ± 0.40	4.72 ± 0.60	4.97 ± 0.60	1.86 ± 0.29	1.72 ± 0.27
Other	1.01 ± 0.53	0.85 ± 0.44	1.62 ± 0.83	1.74 ± 0.89	0.64 ± 0.33	0.73 ± 0.38
DDCF	0.14 ± 0.01	0.14 ± 0.01	2.92 ± 0.31	2.94 ± 0.31	2.15 ± 0.23	2.20 ± 0.24
$\gamma + X$	0 ± 0	0 ± 0	6.34 ± 2.74	2.23 ± 1.86	2.02 ± 1.74	2.79 ± 1.88
Fakes	11.21 ± 2.52	12.95 ± 2.76	15.35 ± 2.09	12.78 ± 1.80	4.23 ± 1.05	3.58 ± 0.90
Total	34.83 ± 6.03	28.99 ± 6.22	56.07 ± 6.03	39.69 ± 5.29	20.15 ± 2.95	16.44 ± 2.71
Observed	43	23	56	39	18	19
Data/MC	1.23± 0.28	0.79± 0.24	1.00± 0.17	0.98± 0.20	0.89± 0.25	1.16± 0.33

Table 8.27: The event yields in six $t\bar{t}W$ dilepton same-sign $2b$ signal regions.

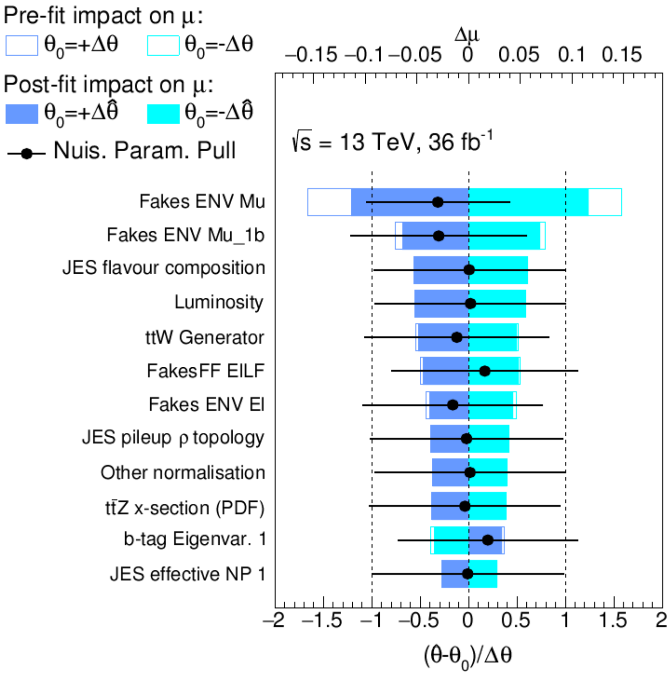


Figure 8.16: The ranking plot showing effects of individual systematic uncertainties of the $t\bar{t}W$ total cross section fit in 16 signal regions.

8.9.4 Combined fit

After optimizing the signal sensitivity of the individual channels, the combined fit has been performed. The combined fit included the three $t\bar{t}Z$ channels: 2ℓ, 3ℓ and 4ℓ, together with 2ℓ SS and 3ℓ channels of the $t\bar{t}W$. All signal and control regions already mentioned in this chapter were included in the fit.

The systematic uncertainties are correlated among the channels and regions, thus the fit is able to constrain effects of some systematic uncertainties and reduce the total systematic uncertainty.

The results of the combined fit are summarized in Table 8.28. The $t\bar{t}Z$ signal significance is 8.9 σ. The $\mu_{t\bar{t}Z}$ and $\mu_{t\bar{t}W}$ values from the table correspond to the following cross section values:

$$\sigma_{t\bar{t}W}^{measured} = 0.87\ \pm 0.13(\text{stat.})\ \pm 0.14(\text{syst.})\ \text{pb}\ = 0.87\ \pm 0.19\ \text{pb} \tag{8.7}$$

$$\sigma_{t\bar{t}Z}^{measured} = 0.95\ \pm 0.08(\text{stat.})\ \pm 0.10(\text{syst.})\ \text{pb}\ = 0.95\ \pm 0.13\ \text{pb} \tag{8.8}$$

The effects of statistical and systematic uncertainties are summarized in Table 8.29.

Parameter	Value
$\mu_{t\bar{t}Z}$	1.08 ± 0.14
$\mu_{t\bar{t}W}$	1.44 ± 0.32
$\mu_{Z+1\text{HF}}$	1.19 ± 0.25
$\mu_{Z+2\text{HF}}$	1.09 ± 0.13
μ_{WZ}	0.91 ± 0.10
μ_{ZZ}	1.11 ± 0.17

Table 8.28: The fitted signal and background normalizations obtained from the combined fit. The Monte Carlo prediction of the cross section is taken as the unit of μ.

Category	$t\bar{t}Z$ POI	$t\bar{t}W$ POI
Luminosity	2.9%	4.5%
Simulated sample statistics	2.0%	5.3%
Data-driven background statistics	2.5%	6.3%
JES/JER	1.9%	4.1%
Flavour tagging	4.2%	3.7%
Other object-related	3.7%	2.5%
Data-driven background normalisation	3.2%	3.9%
Modelling of backgrounds from simulation	5.3%	2.6%
Background cross section	2.3%	4.9%
Fake leptons and charge misID	1.8%	5.7%
$t\bar{t}Z$ modelling	4.9%	0.7%
$t\bar{t}W$ modelling	0.3%	8.5%
Total systematic uncertainty	10.2%	16.0%
Statistical uncertainty	8.4%	15.2%
Total	13.0%	22.2%

Table 8.29: The effects of systematic uncertainties on $t\bar{t}Z$ and $t\bar{t}W$ cross sections.

8.10 Post-Fit results in the dilepton channel

Replacing Pre-Fit (Monte Carlo prediction) values of Z+HF and $t\bar{t}Z$ normalizations, and values of nuisance parameters related to the systematic uncertainties, by the values and corresponding uncertainties obtained from the combined fit, the yields summarized in Table 8.30 have been obtained.

	2ℓ-Z-6j1b	2ℓ-Z-5j2b	2ℓ-Z-6j2b
$t\bar{t}Z$	78.0 ± 11.3	78.6 ± 9.69	122 ± 14.1
$t\bar{t}$	330 ± 9.76	1120 ± 33.0	577 ± 17.1
Z + 2 HF	908 ± 112	1590 ± 104	790 ± 62.5
Z + 1 HF	1470 ± 173	212 ± 52.1	106 ± 26.3
Z + 0 HF	468 ± 96.8	48.6 ± 20.2	18.4 ± 8.97
other	189 ± 65.5	157 ± 42.0	108 ± 28.5
Total	3450 ± 68.2	3210 ± 71.2	1720 ± 44.2
Data	3433	3272	1749
Data/MC	1.00 ± 0.03	1.02 ± 0.03	1.02 ± 0.04

Table 8.30: The Post-Fit event yields of the three signal regions of the dilepton channel. The fitted values of signal, Z + 1 HF and Z + 2 HF normalizations, as well as the fitted values of nuisance parameters related to the systematic uncertainties, have been applied. The uncertainties include both statistical and systematic uncertainties. The fitted event yields obtained from the Monte Carlo are in a very good agreement with the observed number of events.

The Post-Fit plots for the BDT output distributions are shown in Figure 8.17.
The Post-Fit distributions of the BDT input variables are shown in Appendix .1

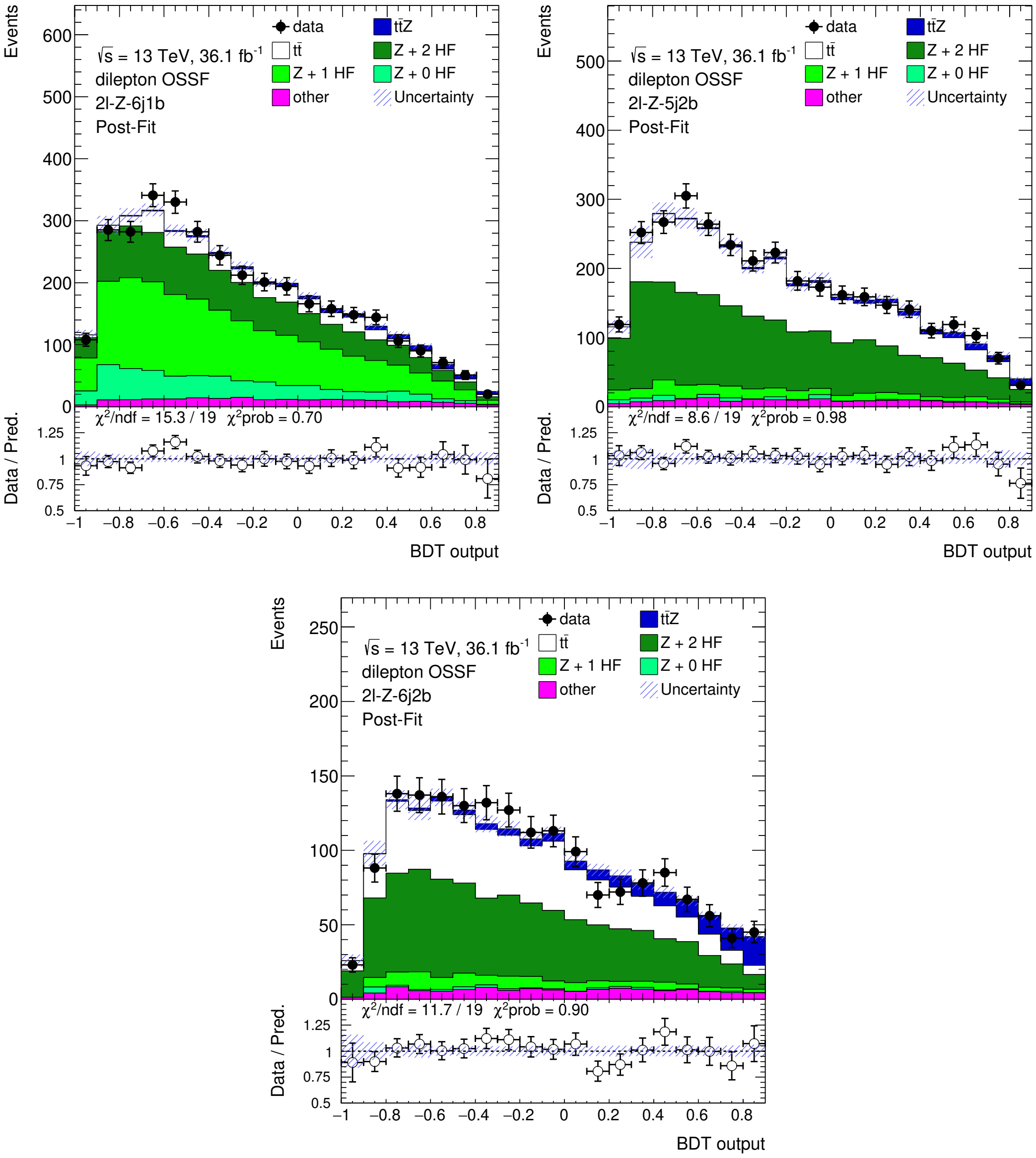


Figure 8.17: The Post-Fit distributions of BDT output in three dilepton signal regions. A good agreement between the data and Monte Carlo prediction can be observed.

Chapter 9

4ℓ channel analysis at full Run II dataset (139 fb^{-1})

After finishing the analysis at 2015 and 2016 data, the increasing luminosity, corresponding to 139 fb^{-1} collected during the Run II (2015-2018) brought new possibilities in the $t\bar{t}Z$ process analysis. The number of events obtained from the full Run II dataset is high enough to enable a differential cross section measurement, as well as to improve a precision of the total cross section measurement, reaching an uncertainty similar, or maybe, even lower than the uncertainty of the theory prediction.

The main focus of the first $t\bar{t}Z$ full Run II dataset analysis is given to the differential cross section measurement which, however, will not be described in this thesis. Since the dilepton channel, used already in the previous analysis, is very challenging because of its high background rate, only the 3ℓ and 4ℓ channels are used in the current analysis. This is the first $t\bar{t}Z$ differential cross-section measurement performed by ATLAS. The CMS collaboration published recently a paper presenting an analysis performed at 77.5 fb^{-1} of data (corresponding to 2015-2017 period of data taking), measuring the total cross section as well as the differential cross section in two variables: p_{T} of the Z-boson and cosine of angle between the negatively charged lepton from the Z boson and the Z boson itself [16].

This thesis will focus mostly on the total cross section measurement in the tetralepton channel. While the total cross section measurement at 36.1 fb^{-1} in the 3ℓ channel was almost limited by the systematic uncertainty, reaching the same statistical and systematic uncertainty as shown in Eq. 8.4, the 4ℓ channel was still dominated by the statistical uncertainty, as shown in Eq. 8.5. Due to the high statistical uncertainty of the previous measurement, a significant improvement can be expected in the 4ℓ channel with the full Run II dataset.

The analysis of the full Run II dataset is still in progress, with some Monte Carlo samples not yet available and several optimization studies in progress. This chapter of the thesis is meant to document the current analysis status from the total cross section point of view, and give an expected uncertainty on the $t\bar{t}Z$ cross section obtained from the tetralepton channel with the full Run II dataset. The thesis does not aim to present a result of a fit to data at the full Run II dataset, since the missing systematic uncertainties might have an effect on both mean value and uncertainty of the fitted cross section. Only the Asimov fit results are going to be presented.

9.1 Selection

The event selection in the tetralepton channel was motivated by the selection previously used at 36.1 fb^{-1} analysis, however, it was improved for the purpose of the differential measurement, taking into account also higher number of available data events.

Since a combined fit of the 3ℓ and 4ℓ channels is planned, the optimization of object definitions, such as b-tagging working point, lepton isolation and identification etc., is based on a compromise between 3ℓ and 4ℓ channel needs. The main difference with respect to the previous analysis was in relaxing the b-tagging requirement, switching to the threshold corresponding to 85 % efficiency for b-jets. Since the main goal of the analysis is the differential cross-section measurement, targeting also $t\bar{t}$ system reconstruction, the events with exactly one jet cannot be used, while they were used in the previous analysis.

The different requirements on the number of jets and b-tagging changed background fractions and compositions. The cuts used previously and summarized in Sec 8.9.2 had to be reoptimized.

As the result of the optimization, the following cuts have been obtained:

- exactly 4 leptons
- $p_{\text{T}}^{1lep} > 27$ GeV, $p_{\text{T}}^{2lep} > 20$ GeV, $p_{\text{T}}^{3lep} > 10$ GeV
- at least one OSSF lepton pair with $|M_{\ell\ell} - M_Z| < 10$ GeV
- sum of lepton charges $== 0$
- $M_{\ell\ell} > 10$ GeV for all OSSF lepton pairs (in order to suppress a background from photon conversions and decay of resonances)

In addition to the already mentioned selection, four signal regions and one control region for the ZZ background are defined by the cuts in Table 9.1. Similarly to the previous analysis, the OSSF lepton pair with the invariant mass closest to the Z boson mass is considered to originate from the Z-boson decay and labelled Z_1. The other lepton pair is considered to originate from the top-quark pair and it is labelled Z_2.

Region	Z_2 leptons	$\lvert m_Z - m_{Z_2} \rvert$	E_{T}^{miss}	$N_{b\text{-jets}}$	N_{jets}
4ℓ-DF-1b	$e^{\pm}\mu^{\mp}$	-	-	$== 1$	≥ 2
4ℓ-DF-2b	$e^{\pm}\mu^{\mp}$	-	-	≥ 2	≥ 2
4ℓ-SF-1b	$e^{\pm}e^{\mp}, \mu^{\pm}\mu^{\mp}$	{ > 10 GeV < 10 GeV	> 50 GeV > 100 GeV }	$== 1$	≥ 2
4ℓ-SF-2b	$e^{\pm}e^{\mp}, \mu^{\pm}\mu^{\mp}$	{ > 10 GeV < 10 GeV	- > 50 GeV }	≥ 2	≥ 2
4ℓ-CR-ZZ	$e^{\pm}e^{\mp}, \mu^{\pm}\mu^{\mp}$	< 10 GeV	20 GeV $< E_{\mathrm{T}}^{miss} <$ 40 GeV	-	-

Table 9.1: The definitions of four signal regions and ZZ control region of tetralepton channel.

9.2 Event yields

Applying the selection summarized in the previous section, the event yields shown in Table 9.2 have been obtained.

9.3 Asimov Fit

In order to extract an expected uncertainty on the signal cross section, the nuisance parameter fit with similar approach to the previous analysis is used. The four signal regions and one control region are included in the fit, each region having one bin, reflecting expected event yields. The fit has two free parameters: the $t\bar{t}Z$ normalization and ZZ normalization. A set of nuisance parameters is used for the systematics.

In the time of the thesis writing, some of the Monte Carlo samples corresponding to a matrix element modelling and parton showering and hadronization were not available. Due to this, the following systematic uncertainties are missing in the fit: $t\bar{t}Z$ matrix element modelling, tWZ matrix element modelling and tWZ parton showering. In order to compensate an effect of missing samples for theory uncertainties on the ZZ background, additional uncertainty ± 50 % have been assigned to ZZ+HF normaliza-

	4ℓ-SF-1b	4ℓ-SF-2b	4ℓ-DF-1b	4ℓ-DF-2b	4ℓ-CR-ZZ
$t\bar{t}Z$	12.37 ± 0.59	22.05 ± 1.03	16.36 ± 0.90	21.47 ± 1.01	0.62 ± 0.10
ZZ	3.42 ± 0.92	4.41 ± 1.41	0.79 ± 0.15	0.11 ± 0.04	480.1 ± 23.0
tWZ	2.71 ± 0.30	1.86 ± 0.25	3.18 ± 0.32	1.74 ± 0.34	0.12 ± 0.09
$t\bar{t}H$	0.45 ± 0.06	0.73 ± 0.09	0.54 ± 0.07	0.71 ± 0.09	0.02 ± 0.01
Fakes	2.03 ± 0.55	1.99 ± 0.79	1.90 ± 0.65	1.03 ± 0.45	22.49 ± 9.93
other	0.03 ± 0.02	0.11 ± 0.06	0.02 ± 0.02	0.09 ± 0.05	0.76 ± 0.68
Total	21.02 ± 1.35	31.15 ± 2.10	22.80 ± 1.27	25.15 ± 1.26	504.1 ± 25.0
data	18	30	31	31	519
Data/MC	0.86 ± 0.21	0.96 ± 0.19	1.36 ± 0.26	1.23 ± 0.23	1.03 ± 0.07

Table 9.2: The yields in 4 signal regions and ZZ control region of the tetralepton channel with the full Run II dataset, corresponding to 139 fb^{-1} of data. The uncertainties include both statistical and systematic uncertainties, except of tWZ, ZZ and tWZ modelling systematics, since the related samples were not yet available in the time of the thesis writing. Good data/MC agreement is observed in all regions.

tion (associated production of the ZZ with b-jets). This uncertainty is considered to be conservative enough.

Performing the Asimov fit, the following signal and ZZ normalizations have been obtained:

$$\mu_{t\bar{t}Z}^{4\ell,\mathrm{Asimov}} = 1.000^{+0.142}_{-0.132}\ (\mathrm{stat.})\ ^{+0.058}_{-0.050}\ (\mathrm{syst.})\ = 1.000^{+0.153}_{-0.142} \tag{9.1}$$

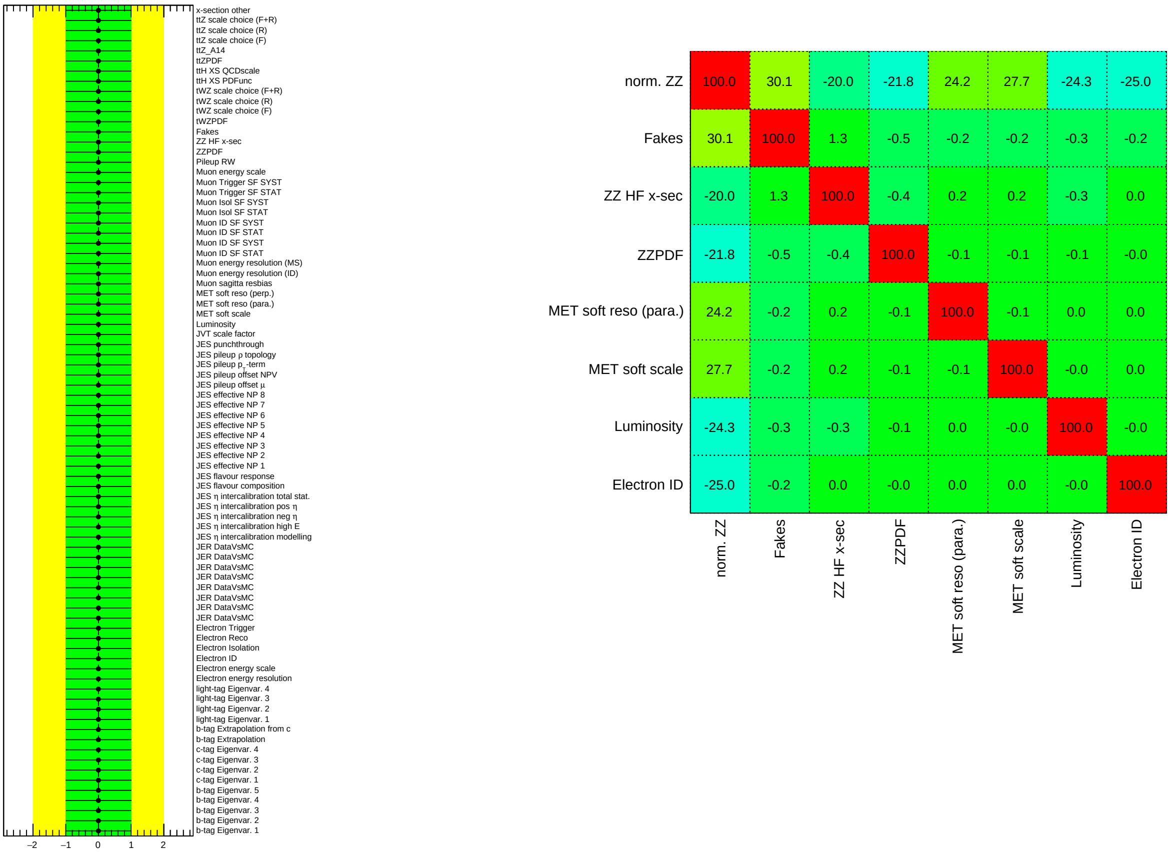
$$\mu_{ZZ} = 1.000^{+0.073}_{-0.068} \tag{9.2}$$

The post-fit values and uncertainties of nuisance parameters, together with correlation matrix of fit parameters are shown in Figure 9.1. Effects of 20 leading systematic uncertainties are shown in the ranking plot in Figure 9.2.

The expected signal significance is 9.6 σ.

9.4 Estimate of the final uncertainty in the 4ℓ channel fit

The uncertainty quoted in Eq. 9.1 is not the final expected uncertainty, since the $t\bar{t}Z$ and tWZ modelling uncertainties are still missing. Although the Monte Carlo



	norm. ZZ	Fakes	ZZ HF x-sec	ZZPDF	MET soft reso (para.)	MET soft scale	Luminosity	Electron ID
norm. ZZ	100.0	30.1	-20.0	-21.8	24.2	27.7	-24.3	-25.0
Fakes	30.1	100.0	1.3	-0.5	-0.2	-0.2	-0.3	-0.2
ZZ HF x-sec	-20.0	1.3	100.0	-0.4	0.2	0.2	-0.3	0.0
ZZPDF	-21.8	-0.5	-0.4	100.0	-0.1	-0.1	-0.1	-0.0
MET soft reso (para.)	24.2	-0.2	0.2	-0.1	100.0	-0.1	0.0	0.0
MET soft scale	27.7	-0.2	0.2	-0.1	-0.1	100.0	-0.0	0.0
Luminosity	-24.3	-0.3	-0.3	-0.1	0.0	-0.0	100.0	-0.0
Electron ID	-25.0	-0.2	0.0	-0.0	0.0	0.0	-0.0	100.0

Figure 9.1: Nuisance parameters plot (left) and correlation matrix (right) for the Asimov fit in tetralepton channel, using event yields corresponding to the full Run II dataset. No significant constrains of the systematic uncertainties have been observed.

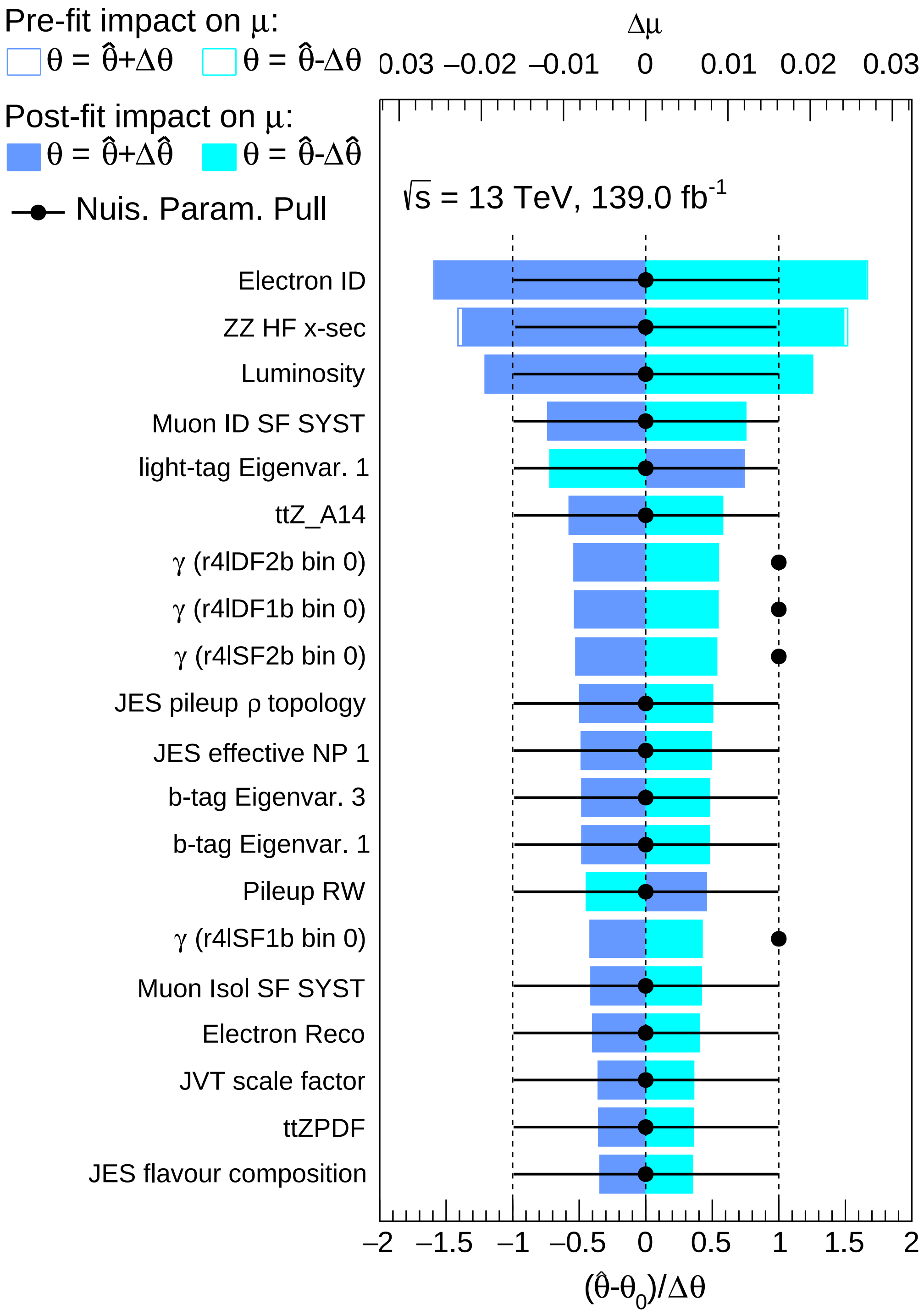


Figure 9.2: Effects of 20 leading systematic uncertainties in the tetralepton channel Asimov fit, using event yields corresponding to the full Run II dataset. The uncertainty on electron identification efficiency is the leading systematic uncertainty, followed by cross section of ZZ+HF background and luminosity uncertainty. The γ parameters correspond to statistical uncertainties of fitted templates, originating from limited number of events in data-driven fake lepton estimate. It is the ratio between the fitted (providing the highest likelihood) and predicted (suffering from statistical fluctuations) Monte Carlo contribution in the bin.

samples for their proper estimate are not available at the time being, their effect can be approximately estimated from the previous analysis.

The currently missing uncertainties were not correlated with any other uncertainties. Based on this fact, their effect can be summed in quadrature with the current value of the expected uncertainty, in order to obtain an estimate of the final expected uncertainty on the $t\bar{t}Z$ cross section. This approach is based on an assumption, that the systematic uncertainties were not changed by the different selections applied in previous analysis and the current one. It is not an exact prediction of this effect and the uncertainty have to be calculated again once the samples for these systematic uncertainties become available.

The effects of the missing uncertainties in the previous analysis on the fitted value of the $t\bar{t}Z$ cross section were following:

- $t\bar{t}Z$ matrix element modelling: ± 3.4 %
- tWZ matrix element modelling: ± 1.4 %
- tWZ parton showering and hadronization: ± 2.8 %

Adding the uncertainties in quadrature to the systematic uncertainty of the current measurement, the following uncertainty on the signal cross section can be expected:

$$\mu_{t\bar{t}Z}^{4\ell,\text{expected}} = 1.000^{+0.142}_{-0.132}\,(\text{stat.})\,{}^{+0.074}_{-0.068}\,(\text{syst.}) \;= 1.000^{+0.160}_{-0.149} \tag{9.3}$$

The expected uncertainty on the $t\bar{t}Z$ cross section obtained from the 4ℓ channel shows significant improvement in the sensitivity with respect to the result from 36.1 fb^{-1}, shown in Eq. 8.8.

The expected uncertainty obtained from the 4ℓ channel (+16%, -15%) is significantly lower compared to the relative uncertainty obtained from the 4ℓ channel fit at 36.1 fb^{-1}, which was +25%, -23%. It approaches the uncertainty obtained from the combined fit at 36.1 fb^{-1}, which was ± 14 %.

Chapter 10

Conclusion

This thesis presents measurements of the total cross section of the top-quark pair and Z boson associated production in two analyses using the data from pp collisions at $\sqrt{s}$ = 13 TeV collected by the ATLAS experiment.

The first analysis was performed at 36.1 fb^1, corresponding to 2015-2016 period of data taking. The analysis uses combination of 2, 3 and 4 lepton final states of $t\bar{t}Z$ and the fit is performed simultaneously with $t\bar{t}W$ cross section fit in dilepton same-sign and trilepton final states. This measurement was the first observation of the $t\bar{t}Z$ process by the ATLAS experiment, reaching the significance (exclusion of background only hypothesis) 8.9 σ and measuring the cross section value

$$\sigma_{t\bar{t}Z}^{measured} = 0.95\ \pm 0.08(\text{stat.})\ \pm 0.10(\text{syst.})\ \text{pb}\ = 0.95\ \pm 0.13\ \text{pb}. \tag{10.1}$$

The author of the thesis was the main analyser in the 2ℓ channel. The 2ℓ channel is characteristic by the high background rate, which makes it necessary to use a multivariate analysis in order to improve a signal sensitivity. The Neural Network has been studied in details in the thesis and the set of its input variables has been studied and optimized. As an alternative to the Neural Network, the Boosted Decision Tree was tested by another member of the analysis team. Based on the expected signal sensitivity using the NN and BDT, showing slightly better performance in the BDT case, the BDT has been chosen as the MVA used for this measurement. The expected signal cross section uncertainty, obtained from the fit in the 2ℓ channel, was $^{+21\%}_{-20\%}$(stat.) $^{+23\%}_{-20\%}$(syst.) with expected significance 3.8 σ. The fit to data in the 2ℓ channel results in 3.0 σ observed signal significance and the total cross section value of

$$\sigma_{t\bar{t}Z}^{2l,measured} = 0.64^{+0.15}_{-0.15}(\text{stat.})^{+0.20}_{-0.19}(\text{syst.})\ \text{pb} = 0.64^{+0.25}_{-0.24}\ \text{pb}. \tag{10.2}$$

Both measured cross sections, the one obtained only from the fit in the 2ℓ channel, as well as the value obtained from the combination of all channels, are in a good agreement with the theory prediction at NLO+NNLL in QCD and NLO in EW precision [1]:

$$\sigma_{t\bar{t}Z}^{\text{theory}} = 0.863^{+0.073}_{-0.085}(\text{scale})^{+0.028}_{-0.028}(\text{PDF} + \alpha_{\text{S}})\text{pb} \tag{10.3}$$

The other part of the thesis presents the current measurement of the $t\bar{t}Z$ total and differential cross section in 4ℓ channel. The analysis is still in progress. The combination of 3ℓ and 4ℓ channels is planned, but only individual results from the 3ℓ and 4ℓ channels are currently available. The expected signal sensitivity of the 4ℓ channel to the total cross section, with the currently implemented systematic uncertainties, is

$$\mu_{t\bar{t}Z}^{4\ell,\text{Asimov}} = 1.000^{+0.142}_{-0.132}\ (\text{stat.})\ {}^{+0.058}_{-0.050}\ (\text{syst.})\ = 1.000^{+0.153}_{-0.142} \tag{10.4}$$

with 9.6 σ significance. Assuming the same effects of the missing systematic uncertainties as in the previous analysis, the expected sensitivity to the total cross section was found to be

$$\mu_{t\bar{t}Z}^{4\ell,\text{expected}} = 1.000^{+0.142}_{-0.132}\ (\text{stat.})\ {}^{+0.074}_{-0.068}\ (\text{syst.})\ = 1.000^{+0.160}_{-0.149}. \tag{10.5}$$

Bibliography

[1] Anna Kulesza, Leszek Motyka, Daniel Schwartländer, Tomasz Stebel, and Vincent Theeuwes. Associated production of a top quark pair with a heavy electroweak gauge boson at NLO+NNLL accuracy. *Eur. Phys. J.*, C79(3):249, 2019. arXiv:1812.08622.

[2] C. Patrignani et al. (Particle Data Group). *Review of particle physics*. CERN, Geneva, 2016.

[3] ATLAS Collaboration. Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector. *Phys. Rev. D*, 97:072003, Apr 2018. arXiv:1712.08891.

[4] CMS Collaboration. Measurement of the associated production of a Higgs boson with a top quark pair in final states with electrons, muons and hadronically decaying τ leptons in data recorded in 2017 at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-HIG-18-019, CERN, Geneva, 2018. CMS-PAS-HIG-18-019.

[5] ATLAS collaboration. Search for direct top squark pair production in events with a Higgs or Z boson, and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector. *Journal of High Energy Physics*, 2017(8):6, Aug 2017. arXiv:1706.03986.

[6] ATLAS collaboration. Search for $t\bar{t}Z$ production in the three lepton final state with 4.7 fb^{-1} of $\sqrt{s} = 7$ TeV pp collision data collected by the ATLAS detector. Technical Report ATLAS-CONF-2012-126, CERN, Geneva, Aug 2012. ATLAS-CONF-2012-126.

[7] ATLAS collaboration. Measurement of the ttW and ttZ production cross sections in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Journal of High Energy Physics*, 2015(11):172, Nov 2015. arXiv:1509.05276.

[8] CMS Collaboration. Measurement of associated production of vector bosons and top quark-antiquark pairs in *pp* collisions at $\sqrt{s}$=7 TeV. *Phys. Rev. Lett.*, 110:172002, Apr 2013. arXiv:1303.3239.

[9] CMS Collaboration. Measurement of top quark-antiquark pair production in association with a W or Z boson in pp collisions at $\sqrt{s} = 8$ TeV. *Eur. Phys. J.*, C74(9):3060, 2014. arXiv:1406.7830.

[10] M. V. Garzelli, A. Kardos, C. G. Papadopoulos, and Z. Trócsányi. $t\bar{t}W$ and $t\bar{t}Z$ hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects. *Journal of High Energy Physics*, 2012(11):56, Nov 2012. arXiv:208.2665.

[11] ATLAS Collaboration. Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ production cross sections in multilepton final states using 3.2 fb^{-1} of *pp* collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector. *Eur. Phys. J.*, C77(1):40, 2017. arXiv:1609.01599.

[12] CMS Collaboration. Measurement of the cross section of top quark pair production in association with a Z boson in pp collisions at 13 TeV. Technical Report CMS-PAS-TOP-16-009, CERN, Geneva, 2016. CMS-PAS-TOP-16-009.

[13] Measurement of the top pair-production in association with a W or Z boson in pp collisions at 13 TeV. Technical Report CMS-PAS-TOP-16-017, CERN, Geneva, 2016. CMS-PAS-TOP-16-017.

[14] ATLAS Collaboration. Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ cross sections in proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector. *Phys. Rev. D*, 99:072009, Apr 2019. arXiv:1901.03584.

[15] CMS collaboration. Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at $\sqrt{s}$ =13 TeV. *Journal of High Energy Physics*, 2018(8):11, Aug 2018. arXiv:1711.02547.

[16] CMS Collaboration. Measurement of top quark pair production in association with a Z boson in proton-proton collisions at $\sqrt{s}$ = 13 TeV. Technical Report CMS-PAS-TOP-18-009, CERN, Geneva, 2019. CMS-PAS-TOP-18-009.

[17] Bernard Pullman. *The Atom in the History of Human Thought.* Oxford University Press, 1998.

[18] J.J. Thompson. Cathode Rays. *Philosophical Magazine*, 1897.

[19] A. Einstein. On a heuristic point of view concerning the production and transformation of light. *Annalen der Physik*, (17):132–148, 1905.

[20] H. Geiger. The Scattering of the α-Particles by Matter. *Proceedings of the Royal Society*, (A84):492–504, 1910.

[21] Ernest Rutheford. Bakerian Lecture: Nuclear constitution of atoms. *Proceedings of the Royal Society of London.*, (Volume 97):97, 1920.

[22] James Chadwick. Possible Existence of Neutron. *Nature*, (Feb. 27):312, 1932.

[23] J. C. Street and E. C. Stevenson. New Evidence for the Existence of a Particle of Mass Intermediate Between the Proton and Electron. *Physical Review*, 52:1003–1004, November 1937.

[24] M. Gell-Mann. A schematic model of baryons and mesons. *Physics Letters*, 8(3):214 – 215, 1964.

[25] G Zweig. An SU_3 model for strong interaction symmetry and its breaking; Version 1. Technical Report CERN-TH-401, CERN, Geneva, Jan 1964. CERN-TH-401.

[26] Bogdan Povh, Klaus Rith, Christoph Scholz, Frank Zetsche, Werner Rodejohann. *Particles and Nuclei.* Number 978-3-662-49583-4 in 1868-4513. Springer-Verlag Berlin Heidelberg, 7 edition, 2015.

[27] Anton Poluektov. First Look at 13 TeV Data and Highlights from the Most Recent Analyses. pages 13–17. 5 p, 2016. https://cds.cern.ch/record/2630651.

[28] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire. Detection of the Free Neutrino: a Confirmation. *Science*, 124(3212):103–104, 1956.

[29] Stanislav Tokar. *Lectures of nuclear and subnuclear physics 3.* 2014.

[30] David Griffiths. *Introduction to elementary particles.* Wiley, 1987.

[31] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental Test of Parity Conservation in Beta Decay. *Physical Review*, 105:1413–1415, February 1957.

[32] Alexandre Deur, Stanley J. Brodsky, and Guy F. de Teramond. The QCD Running Coupling. *Prog. Part. Nucl. Phys.*, 90:1–74, 2016. arXiv:1604.08082.

[33] G. S. Guralnik, C. R. Hagen, and T. W. Kibble. Global Conservation Laws and Massless Particles. *Physical Review Letters*, 13:585–587, November 1964.

[34] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

[35] ATLAS Collaboration. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1):1 – 29, 2012. arXiv:1207.7214.

[36] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012. arXiv:1207.7235.

[37] Abdelhak Djouadi, Jan Kalinowski, Margarete Muehlleitner, and Michael Spira. HDECAY: Twenty$_{++}$ Years After. 2018. arXiv:1801.09506.

[38] Richard D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J.*, C77(10):663, 2017. arXiv:1706.00428.

[39] CDF Collaboration. Observation of top quark production in pbar-p collisions. *Phys. Rev. Lett. 74, 2626*, 1995. arXiv:hep-ex/9503002.

[40] D0 Collaboration. Observation of the top quark. *Phys. Rev. Lett. 74, 2632*, 1995. arXiv:hep-ex/9503003.

[41] Jun Gao, Chong Sheng Li, and Hua Xing Zhu. Top-quark decay at next-to-next-to-leading order in qcd. *Phys. Rev. Lett.*, 110:042001, Jan 2013. arXiv:1210.2808.

[42] ATLAS Collaboration. 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. B761 (2016) 136*, 2017. arXiv:1606.02699.

[43] Richard Hawkings. Inclusive $t\bar{t}$ cross-section measurements at LHC. In *10th International Workshop on Top Quark Physics (TOP2017) Braga, Portugal, September 17-22, 2017*, 2017. arXiv:1711.07400.

[44] ATLAS Collaboration. Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS detector. *Eur. Phys. J. C74 (2014) 3109; Eur. Phys. J. C76 (2016) 642*, 2014. arXiv:1406.5375.

[45] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2016. arXiv:1610.07922.

[46] J R Andersen et al. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. 2013. arXiv:1307.1347.

[47] The ATLAS collaboration. Measurement of the $t\bar{t}\gamma$ production cross section in proton-proton collisions at $\sqrt{s}$=8 TeV with the ATLAS detector. *Journal of High Energy Physics*, 2017(11):86, Nov 2017. arXiv:1706.03046.

[48] Binish Batool, Julien Caudron, Sara Ghasemi, Markus Cristinziani, Mazuza Ghneimat, Ivor Fleck, John Kamal Rizk Meshreki, Amartya Rej, Yichen Li, Maria Moreno Llacer, Elizaveta Shabalina, Arnulf Quadt, Benedikt Volkel, Joshua Wyatt Smith, and Knut Zoch. Measurement of the $t\bar{t}\gamma$ production cross-section in leptonic final states using 36.1 fb^{-1} pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. Technical Report ATL-COM-PHYS-2018-743, CERN, Geneva, Jun 2018. ATL-COM-PHYS-2018-743.

[49] D0 collaboration. *Phys. Rev. Lett.*, 80:2063–2068, Mar 1998. arXiv:hep-ex/9706014.

[50] 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. arXiv:1405.0301.

[51] Olga et al. Bessidskaia Bylund. Measurement of ttZ and ttW production at ATLAS in 13 TeV data, using trilepton and same charge dimuon final states. Technical Report ATL-COM-PHYS-2016-694, CERN, Geneva, Jun 2016. ATL-COM-PHYS-2016-694.

[52] Large hadron collider. `https://home.cern/science/accelerators/large-hadron-collider`, accessed: 2019-02-21.

[53] Christiane Lefèvre. The CERN accelerator complex. Complexe des accélérateurs du CERN. Dec 2008.

[54] Linear accelerator 2. `http://home.cern/about/accelerators/linear-accelerator-2`, accessed: 2019-02-21.

[55] Proton synchrotron booster. `https://home.cern/science/accelerators/proton-synchrotron-booster`, accessed: 2019-02-21.

[56] Proton synchrotron. `https://home.cern/science/accelerators/proton-synchrotron`, accessed: 2019-02-21.

[57] Super proton synchrotron. `https://home.cern/science/accelerators/super-proton-synchrotron`, accessed: 2019-02-21.

[58] The atlas experiment. `http://scipp.ucsc.edu/personnel/atlas.html`, accessed: 2019-02-21.

[59] The inner detector. `http://atlas.cern/discover/detector/inner-detector`, accessed: 2019-02-21.

[60] Aidan Grummer. Operational Experience and Performance with the ATLAS Pixel Detector at the Large Hadron Collider. Technical Report ATL-COM-INDET-2018-032, CERN, Geneva, May 2018. 27 May, 2018 (PM2018 - 14th Pisa Meeting on Advanced Detectors - La Biodola, Isola d'Elba (Italy)).

[61] A Abdesselam et al. The data acquisition and calibration system for the ATLAS semiconductor tracker. *Journal of Instrumentation*, 3(01):P01003–P01003, jan 2008. CERN-ATL-COM-INDET-2007-015.

[62] The ATLAS TRT collaboration. The ATLAS TRT Barrel Detector. *Journal of Instrumentation*, 3(02):P02014–P02014, feb 2008. The ATLAS TRT Barrel Detector.

[63] ATLAS collaboration. Electron efficiency measurements with the ATLAS detector using 2012 LHC proton–proton collision data. *Eur. Phys. J.*, C77(3):195, 2017. arXiv:1612.01456.

[64] *ATLAS calorimeter performance: Technical Design Report.* Technical Design Report ATLAS. CERN, Geneva, 1996. CERN-LHCC-96-040.

[65] ATLAS Collaboration. ATLAS liquid-argon calorimeter: Technical Design Report. Technical report, Geneva, 1996. CERN-LHCC-96-041.

[66] Maaike Limper. *Track and vertex reconstruction in the ATLAS inner detector.* PhD thesis, University of Amsterdam, okt 2009.

[67] The ATLAS Collaboration. The ATLAS experiment at the CERN large hadron collider. *Journal of Instrumentation*, 3(08):S08003–S08003, aug 2008. 10.1088/1748-0221/3/08/s08003.

[68] Sandro Palestini. The muon spectrometer of the ATLAS experiment. *Nuclear Physics B - Proceedings Supplements*, 125:337 – 345, 2003. https://doi.org/10.1016/S0920-5632(03)91013-9.

[69] *ATLAS muon spectrometer: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1997. CERN-LHCC-97-022.

[70] Magnet system. `http://atlas.web.cern.ch/Atlas/TP/NEW/HTML/tp9new/node8.html#SECTION00431000000000000000`, accessed: 2019-02-21.

[71] *ATLAS magnet system: Technical Design Report, 1*. Technical Design Report ATLAS. CERN, Geneva, 1997. CERN-LHCC-97-018.

[72] Will Buttinger. The ATLAS level-1 trigger system. *Journal of Physics: Conference Series*, 396(1):012010, dec 2012. ATL-DAQ-PROC-2012-024.

[73] Martin zur Nedden. The Run-2 ATLAS Trigger System: Design, Performance and Plan. Technical Report ATL-DAQ-PROC-2016-039, CERN, Geneva, Dec 2016. ATL-DAQ-PROC-2016-039.

[74] Thomas Schoerner-Sadenius. The Trigger of the ATLAS experiment. *Mod. Phys. Lett.*, A18:2149–2168, 2003.

[75] R. Frühwirth. Application of Kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 262(2):444 – 450, 1987.

[76] Heather Gray. Track reconstruction with the atlas experiment, 2016. `https://indico.cern.ch/event/504284/contributions/2023875/attachments/1240146/1823137/HGray_Zurich_Tracking.pdf`, accessed: 2019-03-12.

[77] ATLAS collaboration. Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2. *Eur. Phys. J.*, C77(10):673, 2017. arXiv:1704.07983.

[78] W Lampl, S Laplace, D Lelas, P Loch, H Ma, S Menke, S Rajagopalan, D Rousseau, S Snyder, and G Unal. Calorimeter Clustering Algorithms: Description and Performance. Technical Report ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva, Apr 2008. ATL-LARG-PUB-2008-002.

[79] ATLAS collaboration. Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} =$ 13 TeV. *Submitted to: Eur. Phys. J.*, 2019. arXiv:1902.04655.

[80] Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data. Technical Report ATLAS-CONF-2016-024, CERN, Geneva, Jun 2016. ATLAS-CONF-2016-024.

[81] Isolationselectiontool twiki. `https://twiki.cern.ch/twiki/bin/view/AtlasProtected/IsolationSelectionTool`, accessed: 2019-03-12.

[82] ATLAS Collaboration. Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} =$ 13 TeV. *The European Physical Journal C*, 76(5):292, May 2016. arXiv:1603.05598.

[83] ATLAS Collaboration. Topological cell clustering in the atlas calorimeters and its performance in lhc run 1. *The European Physical Journal C*, 77(7):490, Jul 2017. arXiv:1603.02934.

[84] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008. arXiv:0802.1189.

[85] Tagging and suppression of pileup jets with the ATLAS detector. Technical Report ATLAS-CONF-2014-018, CERN, Geneva, May 2014. ATLAS-CONF-2014-018.

[86] J. Stelzer J. Therhaag E. von Toerne H. Voss A. Hoecker, P. Speckmayer. *TMVA 4 Users Guide.* 2017.

[87] Expected performance of the ATLAS b-tagging algorithms in Run-2. Technical Report ATL-PHYS-PUB-2015-022, CERN, Geneva, Jul 2015. ATL-PHYS-PUB-2015-022.

[88] Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data. Technical Report ATLAS-CONF-2011-102, CERN, Geneva, Jul 2011. ATLAS-CONF-2011-102.

[89] G. Aad et al. Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics. 2009. arXiv:0901.0512.

[90] Calibration of light-flavour jet b-tagging rates on ATLAS proton-proton collision data at $\sqrt{s} = 13$ TeV. Technical Report ATLAS-CONF-2018-006, CERN, Geneva, Apr 2018. ATLAS-CONF-2018-006.

[91] ATLAS Collaboration. Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 78(arXiv:1802.08168. 11):903. 66 p, Feb 2018. arXiv:1802.08168.

[92] Oleksandr Viazlo and ATLAS LUCID Collaboration. ATLAS LUCID detector upgrade for LHC Run 2. Technical Report ATL-FWD-PROC-2015-004, CERN, Geneva, Oct 2015. ATL-FWD-PROC-2015-004.

[93] LuminosityPublicResultsRun2 . `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2`, accessed: 2019-03-29.

[94] Richard D. Ball et. al. Parton distributions with lhc data. *Nuclear Physics B*, 867(2):244 – 289, 2013. arXiv:1207.1303.

[95] Jonathan Pumplin et. al. New generation of parton distributions with uncertainties from global QCD analysis. *Journal of High Energy Physics*, 2002(07):012–012, jul 2002. arXiv:hep-ph/0201195.

[96] Hung-Liang Lai, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, Jon Pumplin, and C. P. Yuan. New parton distributions for collider physics. *Phys. Rev.*, D82:074024, 2010. arXiv:1007.2241.

[97] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. A brief introduction to pythia 8.1. *Computer Physics Communications*, 178(11):852 – 867, 2008. arXiv:0710.3820.

[98] Johannes Bellm et al. Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J.*, C76(4):196, 2016. arXiv:1512.0117.

[99] S. Agostinelli at al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303, 2003.

[100] W Lukas. Fast Simulation for ATLAS: Atlfast-II and ISF. Technical Report ATL-SOFT-PROC-2012-065, CERN, Geneva, Jun 2012. ATL-SOFT-PROC-2012-065.

[101] Alwall J et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7):79, Jul 2014. arXiv:1405.0301.

[102] David J. Lange. The EvtGen particle decay simulation package. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 462(1):152 – 155, 2001. BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines.

[103] ATLAS Run 1 Pythia8 tunes. Technical Report ATL-PHYS-PUB-2014-021, CERN, Geneva, Nov 2014. ATL-PHYS-PUB-2014-021.

[104] T Gleisberg, S Höche, F Krauss, M Schönherr, S Schumann, F Siegert, and J Winter. Event generation with SHERPA 1.1. *Journal of High Energy Physics*, 2009(02):007–007, feb 2009. arXiv:0811.4622.

[105] C. Oleari. The POWHEG BOX. *Nuclear Physics B - Proceedings Supplements*, 205-206:36 – 41, 2010. Loops and Legs in Quantum Field Theory, arXiv:1007.3893.

[106] Pile-up subtraction and suppression for jets in ATLAS. Technical Report ATLAS-CONF-2013-083, CERN, Geneva, Aug 2013. ATLAS-CONF-2013-083.

[107] Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2016-015, CERN, Geneva, Aug 2016. ATL-PHYS-PUB-2016-015.

[108] ATLAS Collaboration. Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the atlas detector. *Phys. Rev. D*, 96:072002, Oct 2017. arXiv:1703.09665.

[109] Craig Sawyer, Christopher John Young, Cigdem Issever, and David Lopez Mateos. Monte Carlo Calibration and Combination of In-situ Measurements of Jets in ATLAS. Technical Report ATL-COM-PHYS-2015-071, CERN, Geneva, Jan 2015. ATL-COM-PHYS-2015-071.

[110] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2016. arXiv:1610.07922.

[111] Olaf Behnke, Kevin Kroninger, Gregory Schott, and Thomas Schorner-Sadenius. *Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods.* Wiley-VCH, 1st edition, 2013.

[112] Raul Rojas. *Neural Networks: A Systematic Introduction.* Springer-Verlag, Berlin, 1996.

[113] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, and Helge Voss. TMVA: Toolkit for Multivariate Data Analysis. *PoS*, ACAT:040, 2007.

[114] ROOT. `https://root.cern.ch/`, accessed: 2019-01-10.

[115] NeuroBayes twiki page. `https://twiki.cern.ch/twiki/bin/view/Main/NeuroBayes`, accessed: 2019-01-10.

[116] Jerome Friedman. Stochastic gradient boosting. *Computational Statistics and Data Analysis*, 38:367–378, 02 2002. SLAC-REPRINT-2002-017.

[117] Catherine Bernaciak, Malte Seán Andreas Buschmann, Anja Butter, and Tilman Plehn. Fox-wolfram moments in higgs physics. *Phys. Rev. D*, 87:073014, Apr 2013. arXiv:1212.4436.

.1 Input variables for Neural Network and Boosted Decision Tree

In order to verify the MVA technique response, modelling of input variables needs to be checked. Distributions of MVA input variables have been checked in data and Monte Carlo prediction in order to validate their modelling. No significant mismodelling in MVA input variables has been found. One can see that the uncertainty band is reduced in the Post-Fit plots compared to the Pre-Fit ones and the Data/MC agreement improves after the fit is performed. The Post-Fit plots use values of parameters from the combined fit.

Pre-Fit, Post-Fit (after combined fit) and separation plots for all BDT and NN input variables defined in the Table 8.4 are shown in the Figures 1-24. In addition, shapes of the MVA input variables in $t\bar{t}$ background after $\ell\ell$ and $e\mu$ selections have been checked in order to justify the data driven technique used for the $t\bar{t}$ background estimation. No statistically significant difference between the event kinematics after these selection has been found.

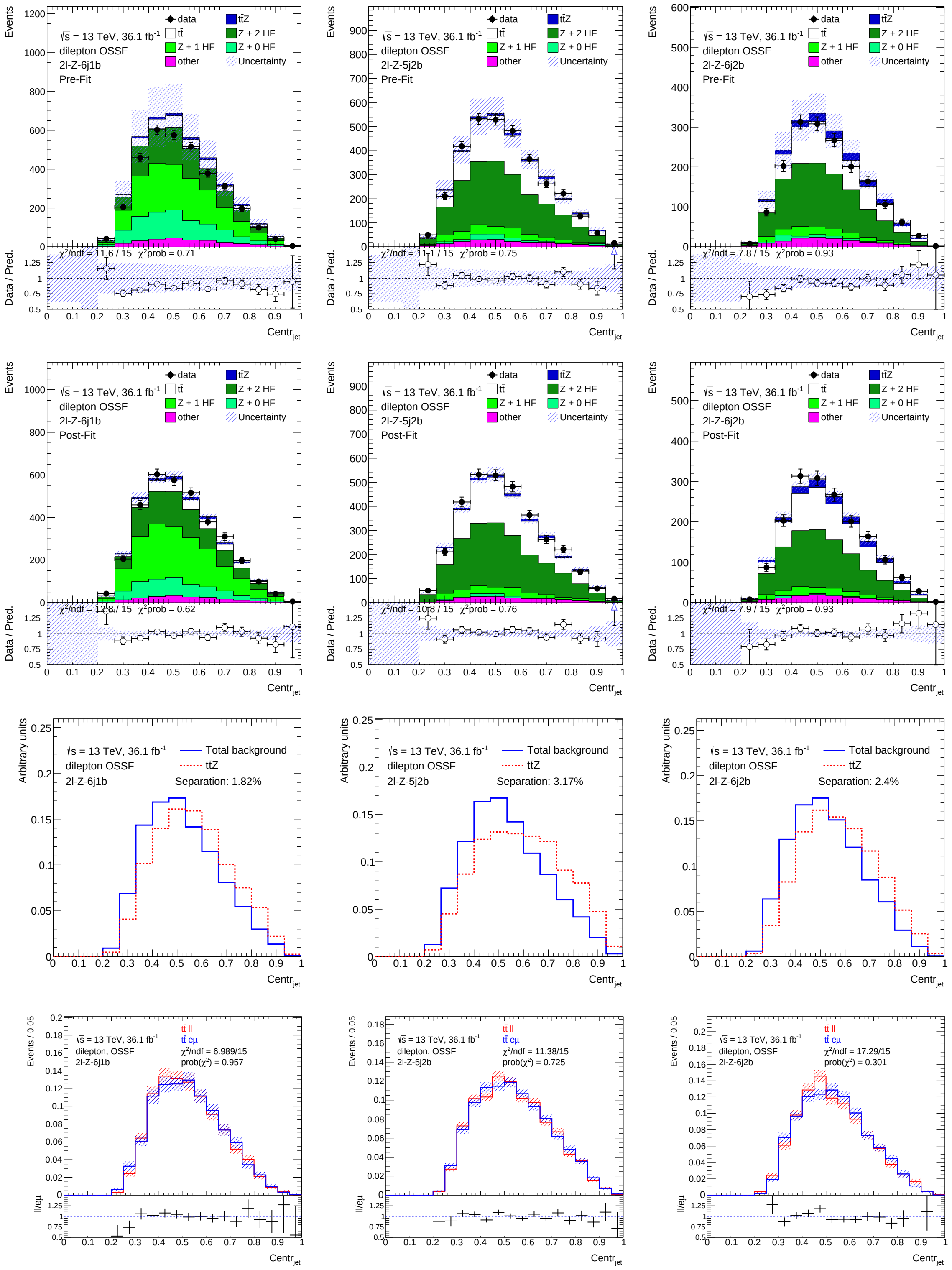


Figure 1: Centrality of the event built from the jets Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in three signal regions of the dilepton channel. The error bars include the systematic uncertainties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

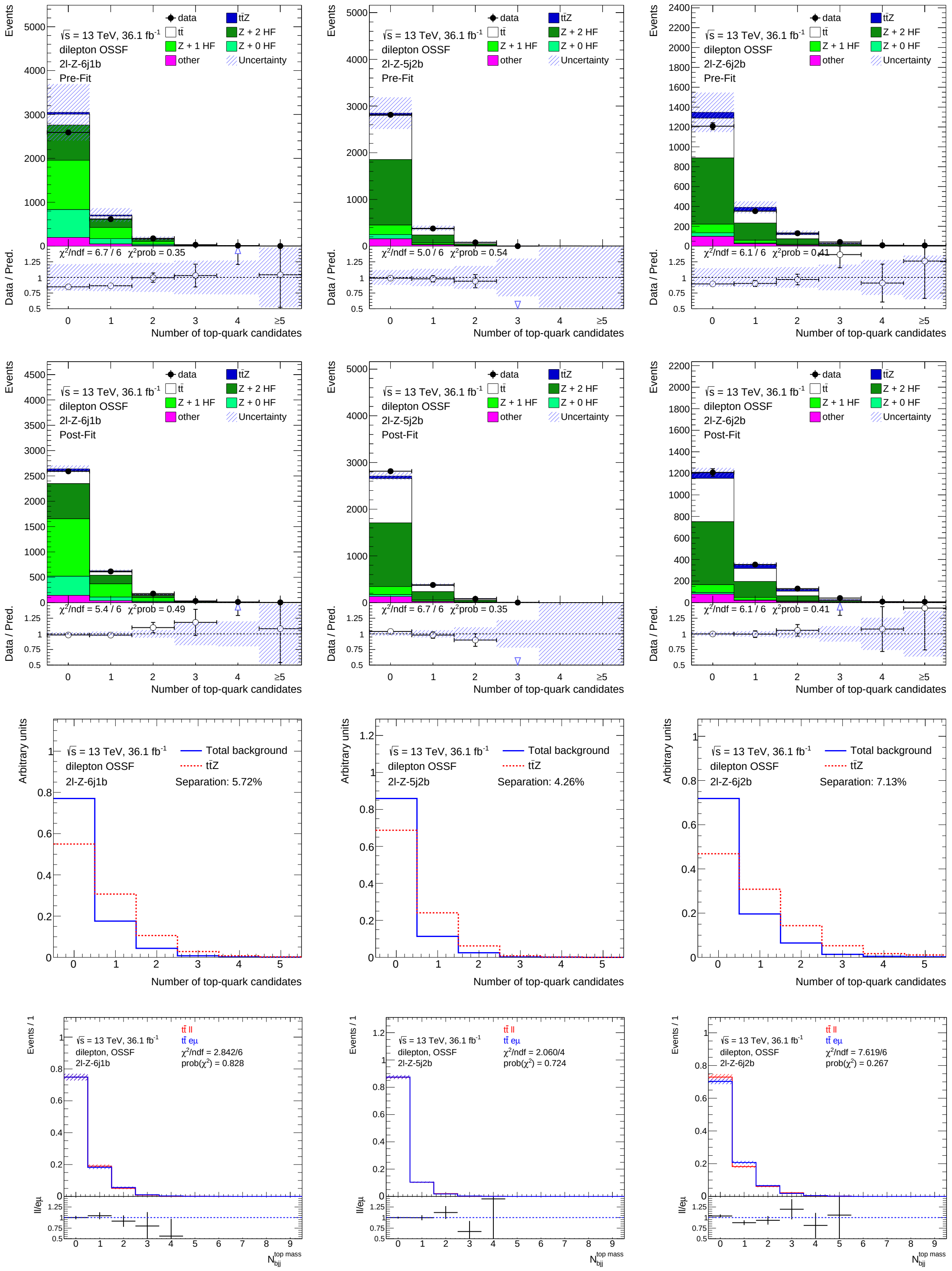


Figure 2: Number of hadronically decaying top-quark candidates, Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in three signal regions of the dilepton channel. The error bars include the systematic uncertanties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

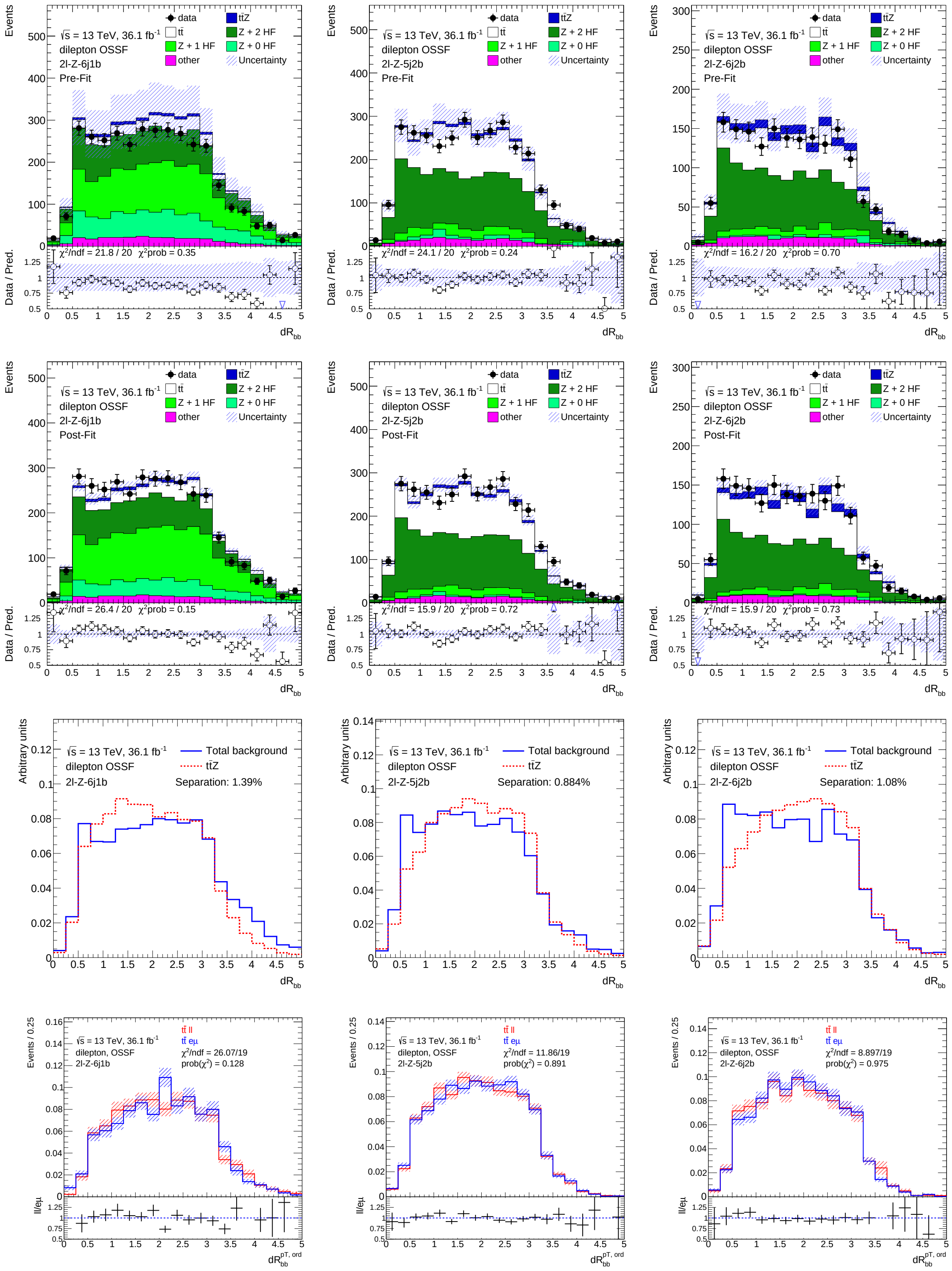


Figure 3: The cone between two jets with the highest b-tagging weights PPre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in three signal regions of the dilepton channel. The error bars include the systematic uncertainties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

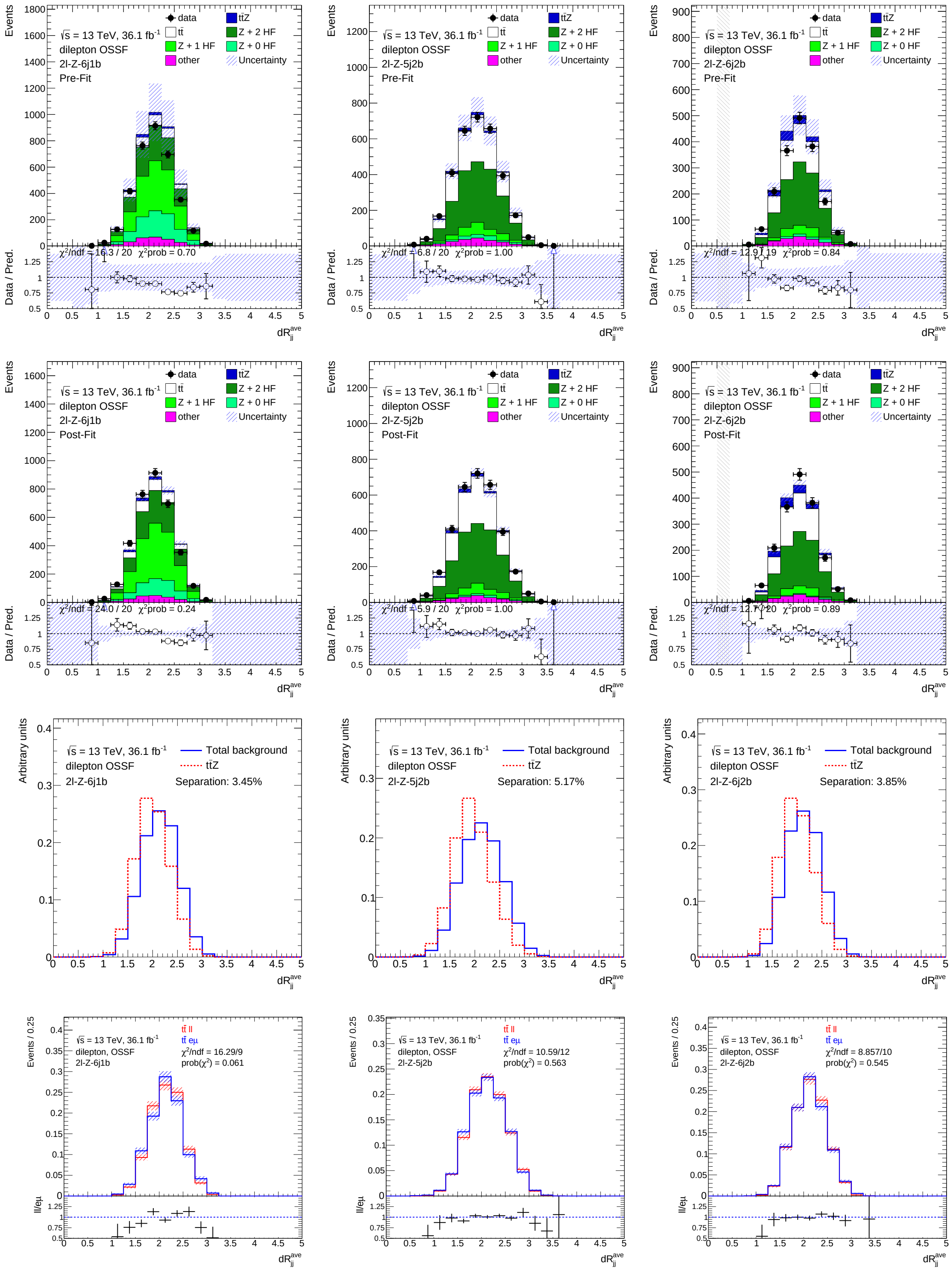


Figure 4: Average cone between any two jets in the event Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in three signal regions of the dilepton channel. The error bars include the systematic uncertainties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

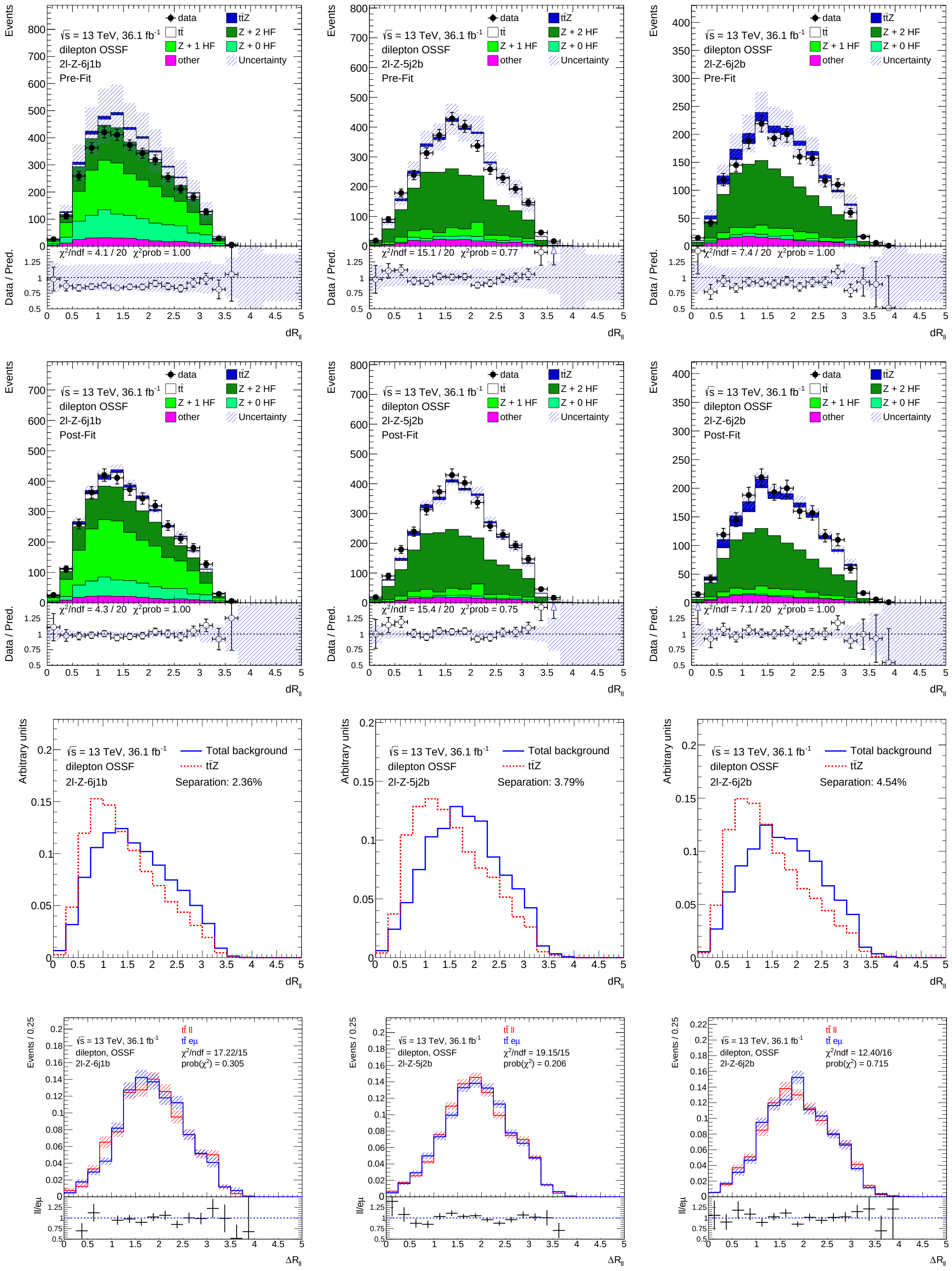


Figure 5: Cone between the leptons Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in three signal regions of the dilepton channel. The error bars include the systematic uncertainties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

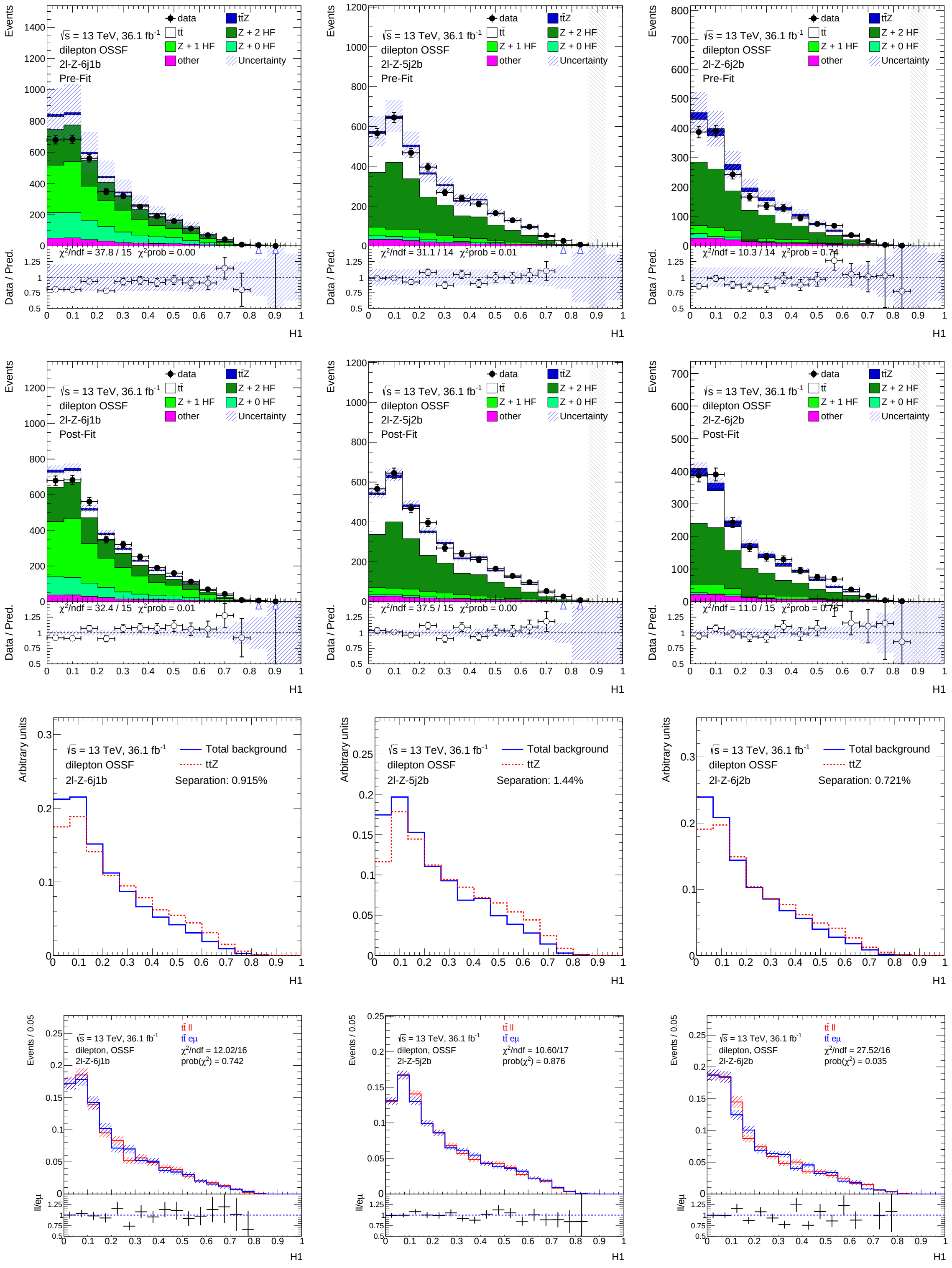


Figure 6: First Fox-Wolfram momentum built from all jets and leptons in the event, Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in three signal regions of the dilepton channel. The error bars include the systematic uncertainties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

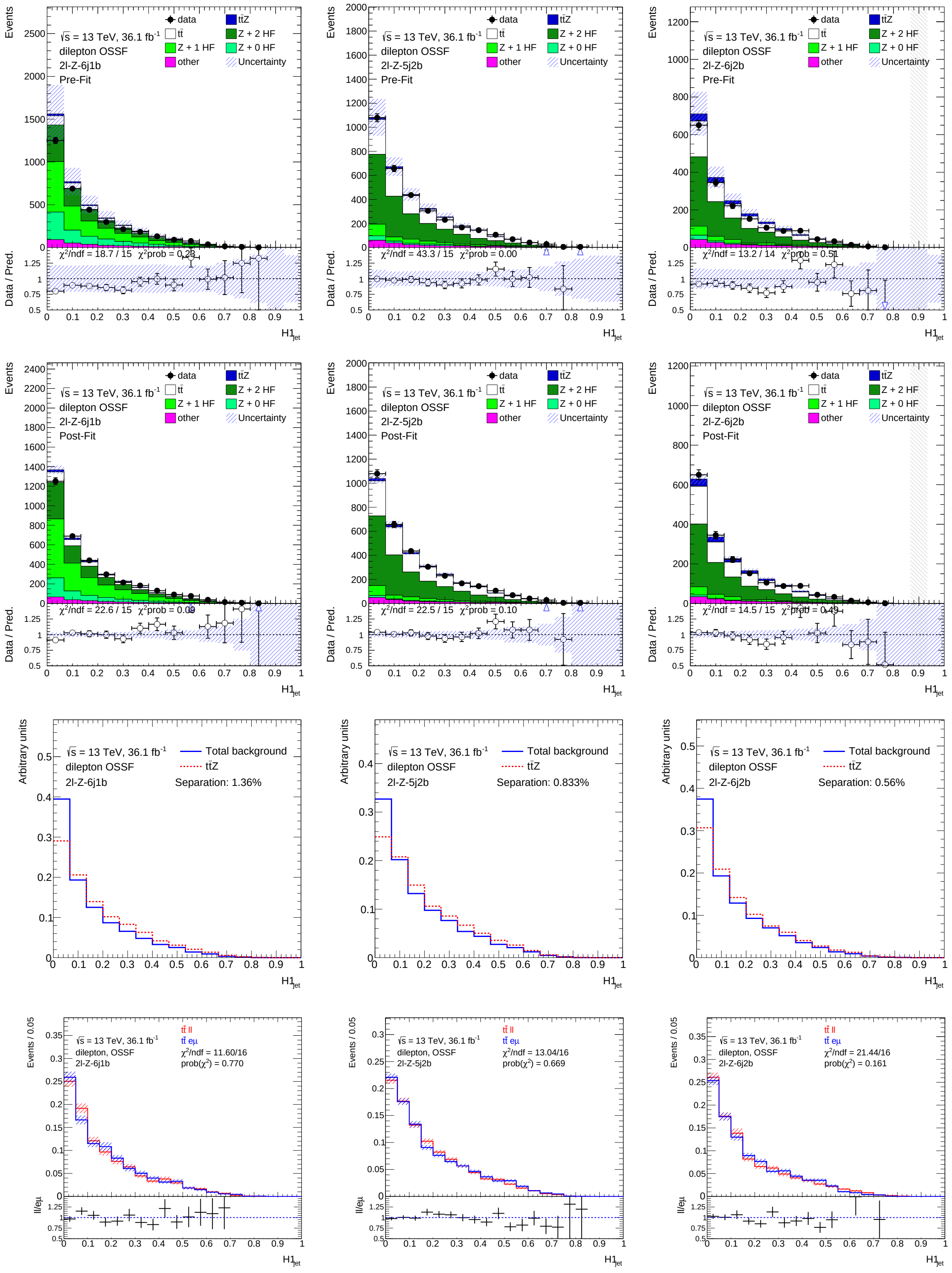


Figure 7: First Fox-Wolfram momentum built from jets, Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in three signal regions of the dilepton channel. The error bars include the systematic uncertanties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

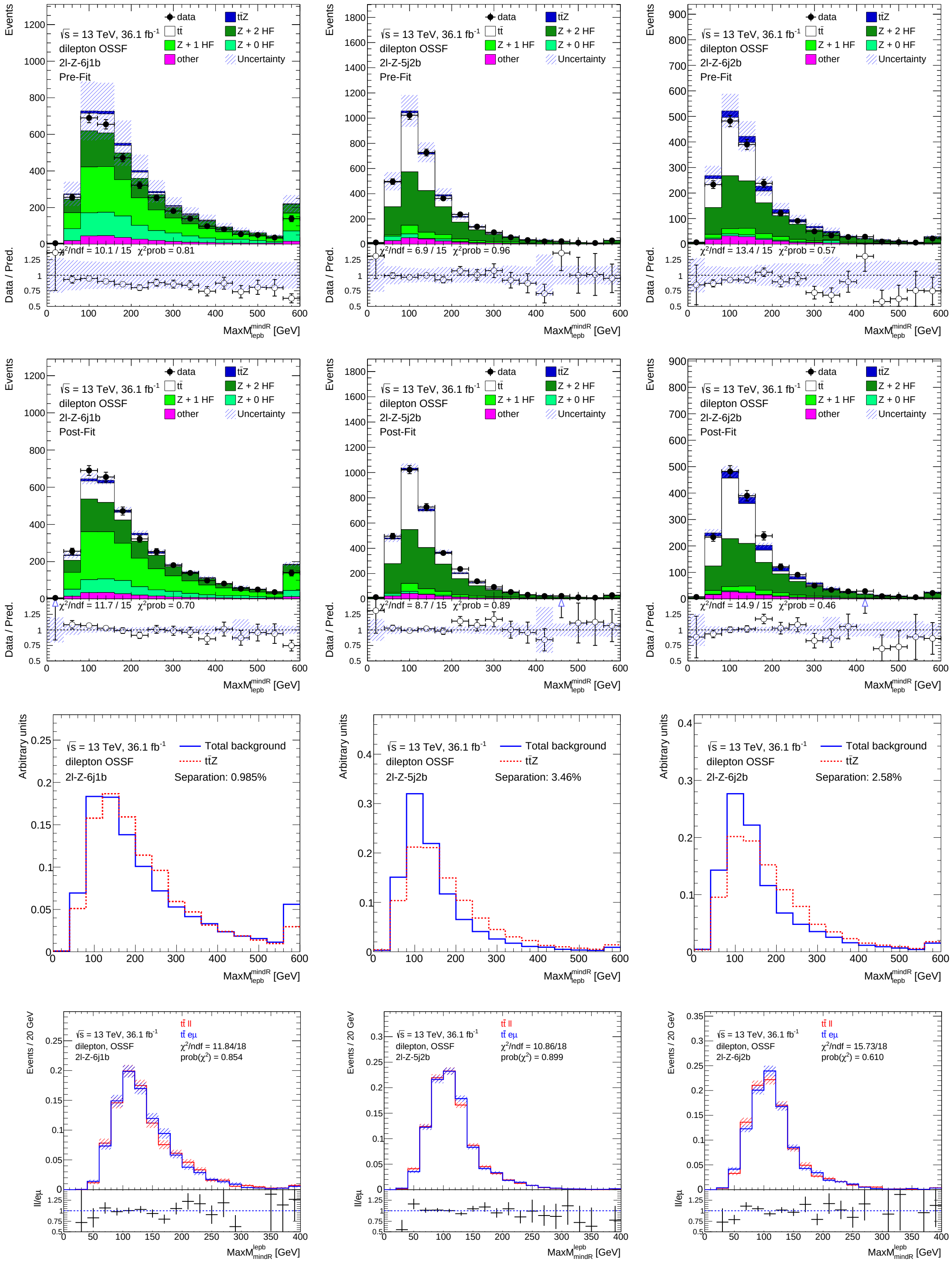


Figure 8: Maximal invariant mass of the lepton and the closest b-jet, Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in three signal regions of the dilepton channel. The error bars include the systematic uncertainties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

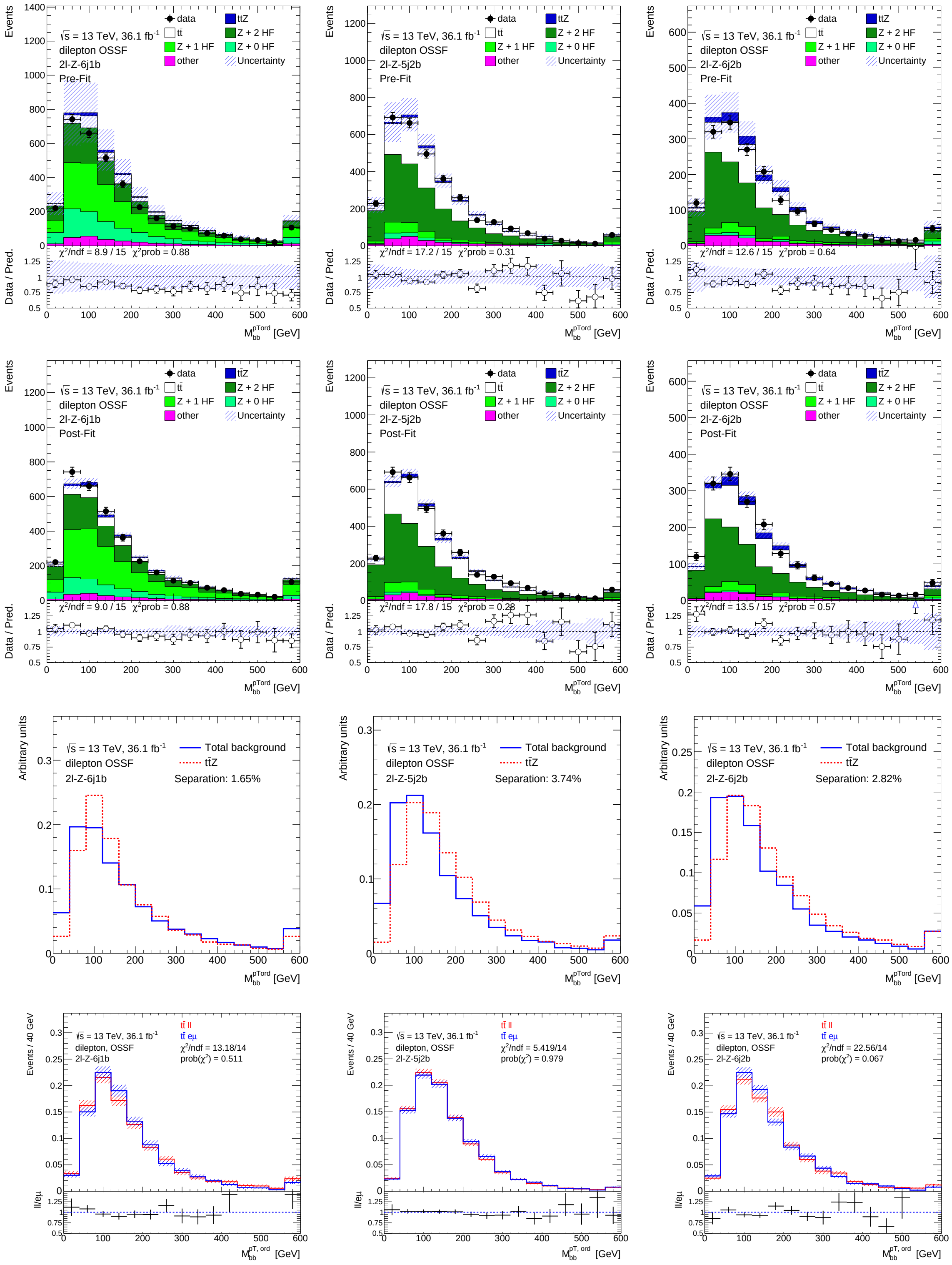


Figure 9: Invariant mass of the two b-jets with the highest b-tagging weight Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in three signal regions of the dilepton channel. The error bars include the systematic uncertainties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

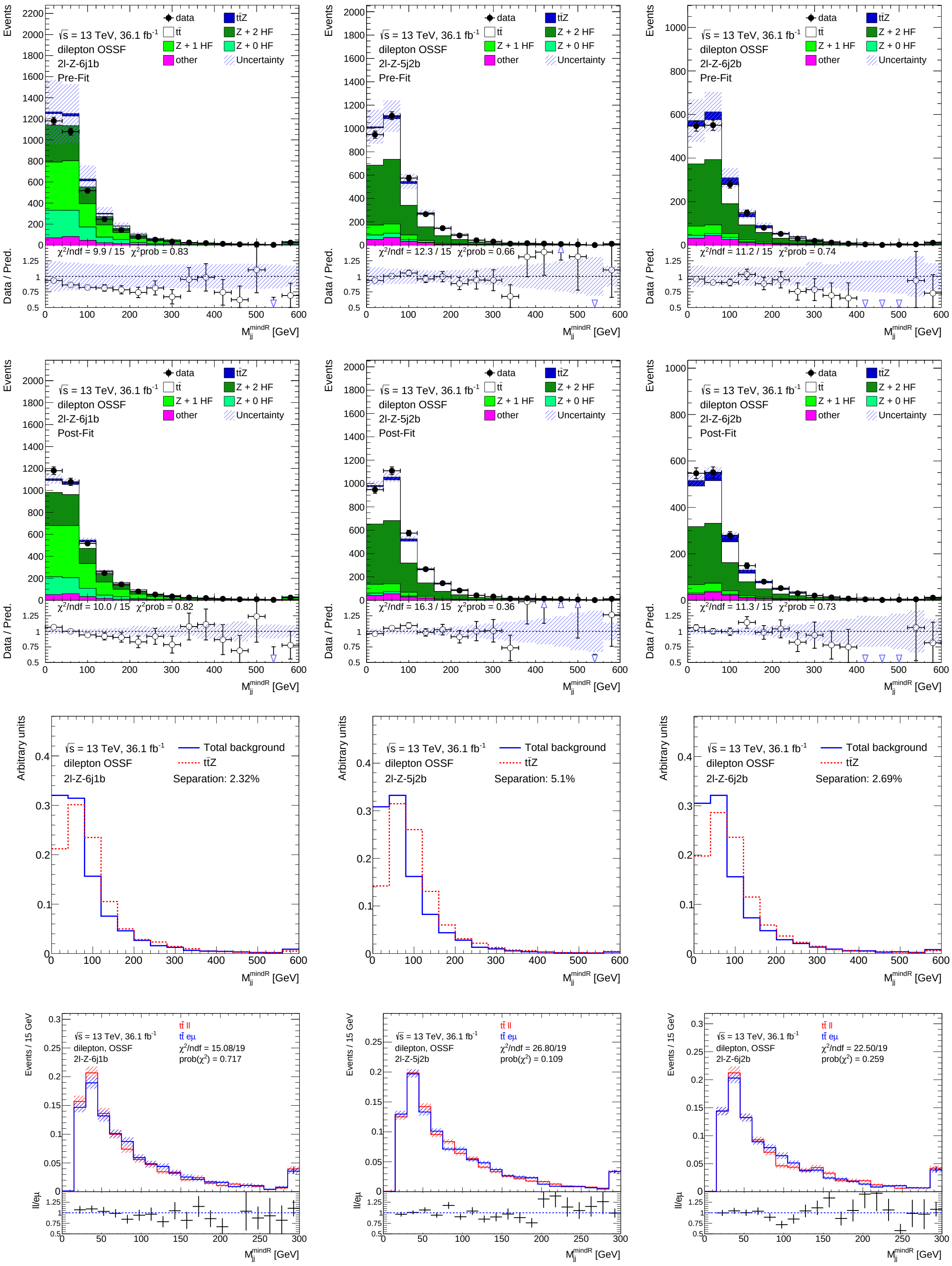


Figure 10: Invariant mass of the two closest non-b-tagged jets, Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in three signal regions of the dilepton channel. The error bars include the systematic uncertainties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

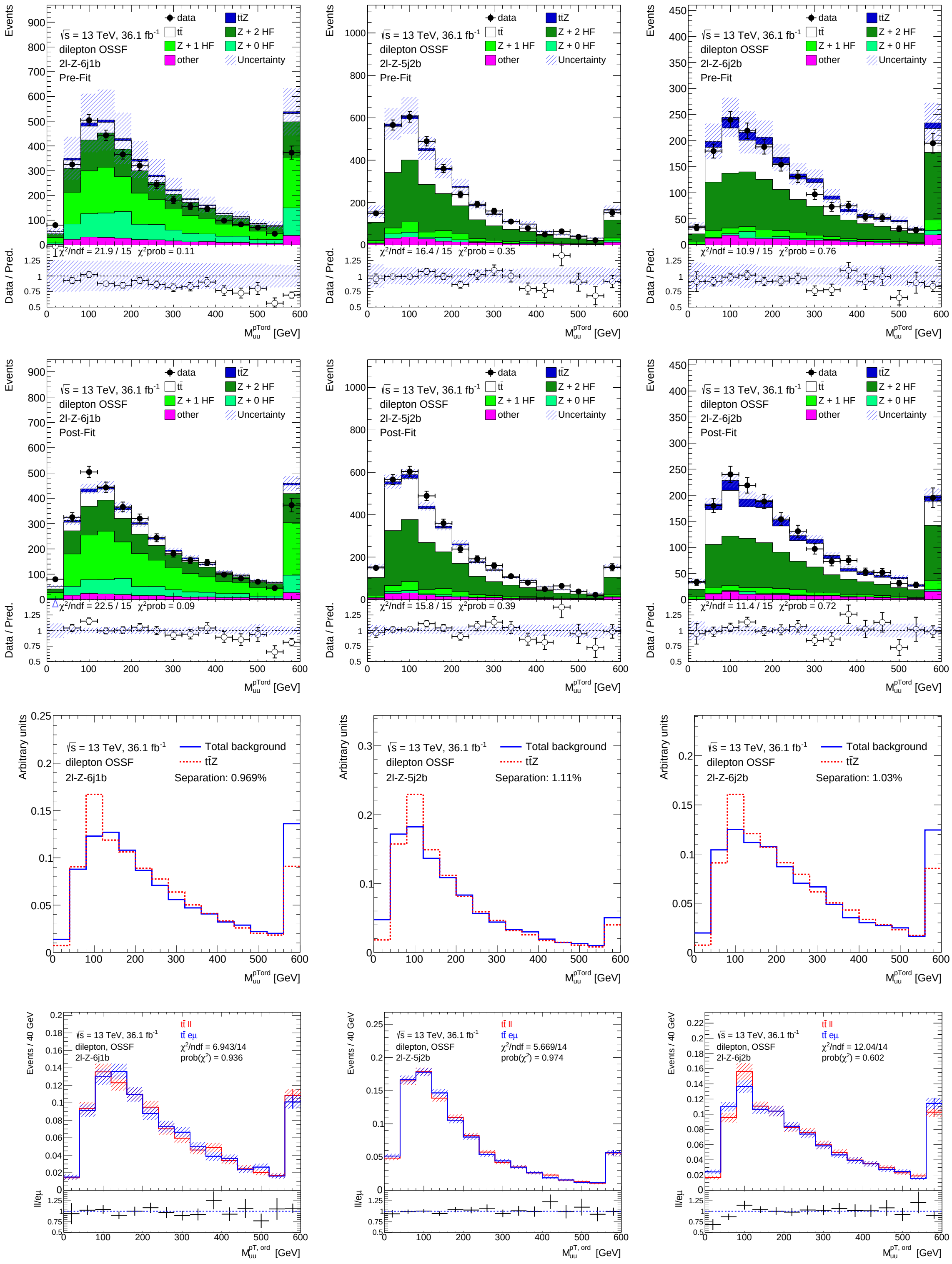


Figure 11: Invariant mass of the two non-b-tagged jets with the highest p_T in the event, Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in three signal regions of the dilepton channel. The error bars include the systematic uncertainties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

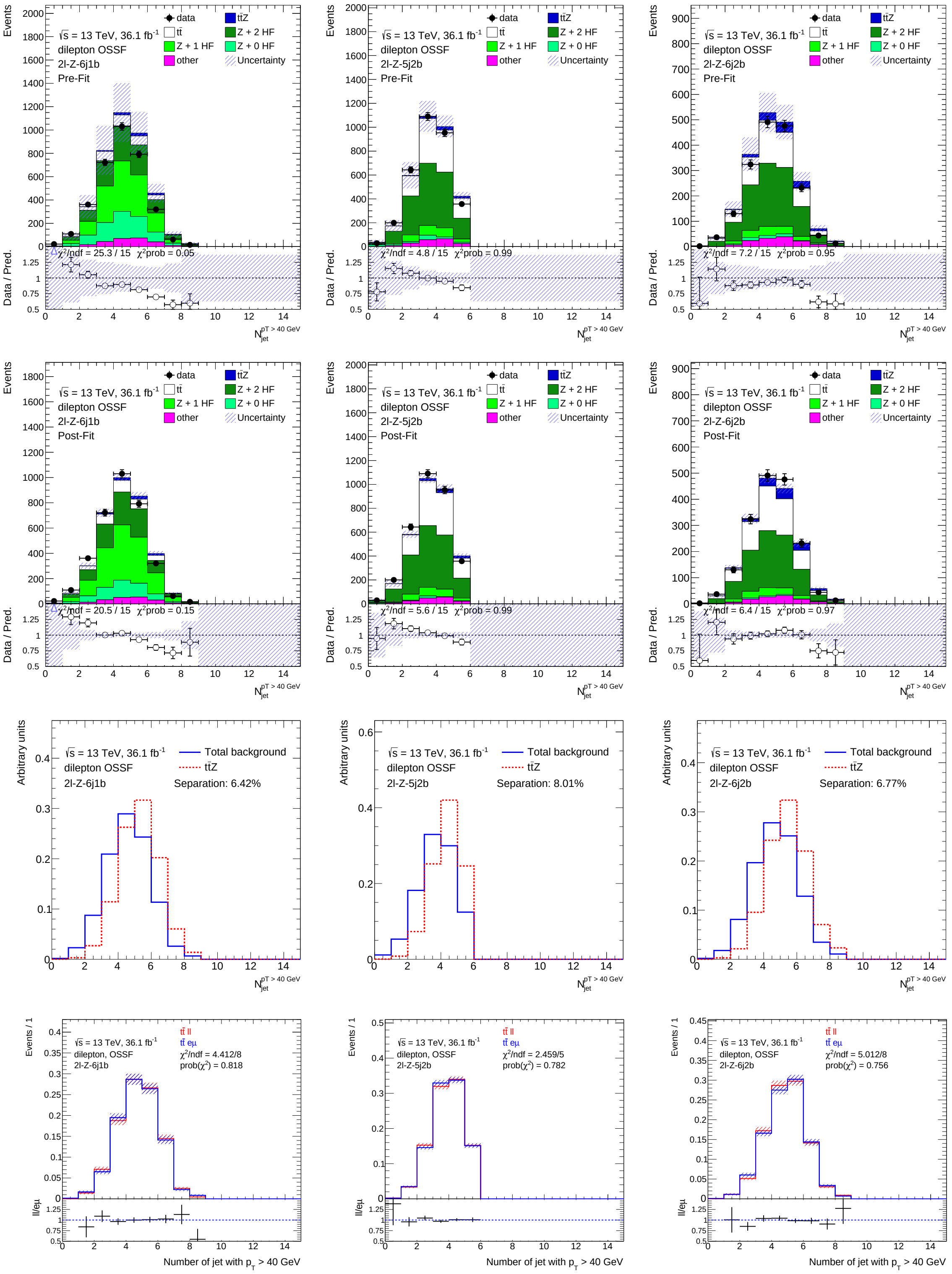


Figure 12: Number of jets with $p_T > 40$ GeV Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in three signal regions of the dilepton channel. The error bars include the systematic uncertainties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

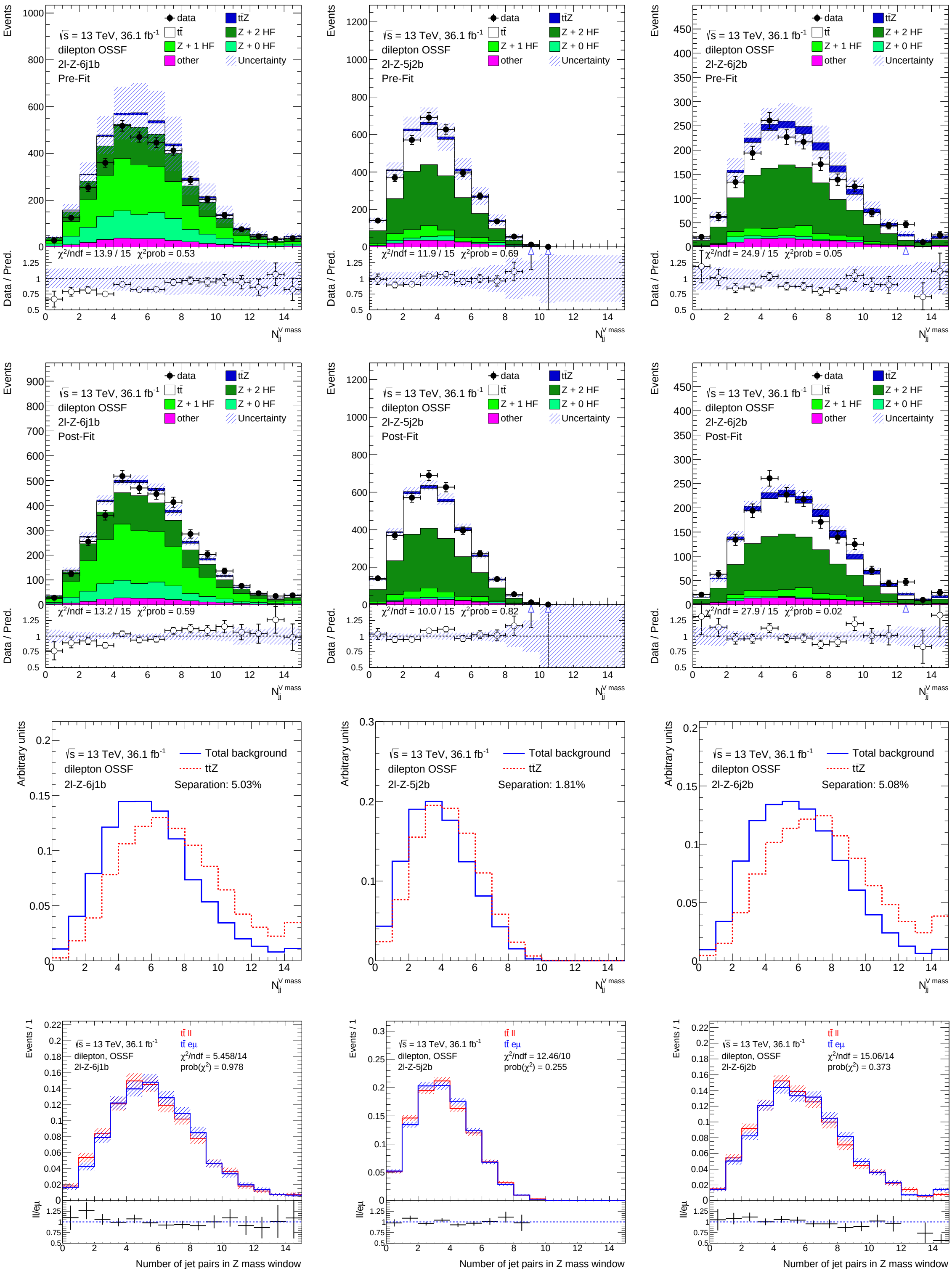


Figure 13: Number of jet pairs with invariant mass close to the mass of Z and W boson, Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in three signal regions of the dilepton channel. The error bars include the systematic uncertainties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

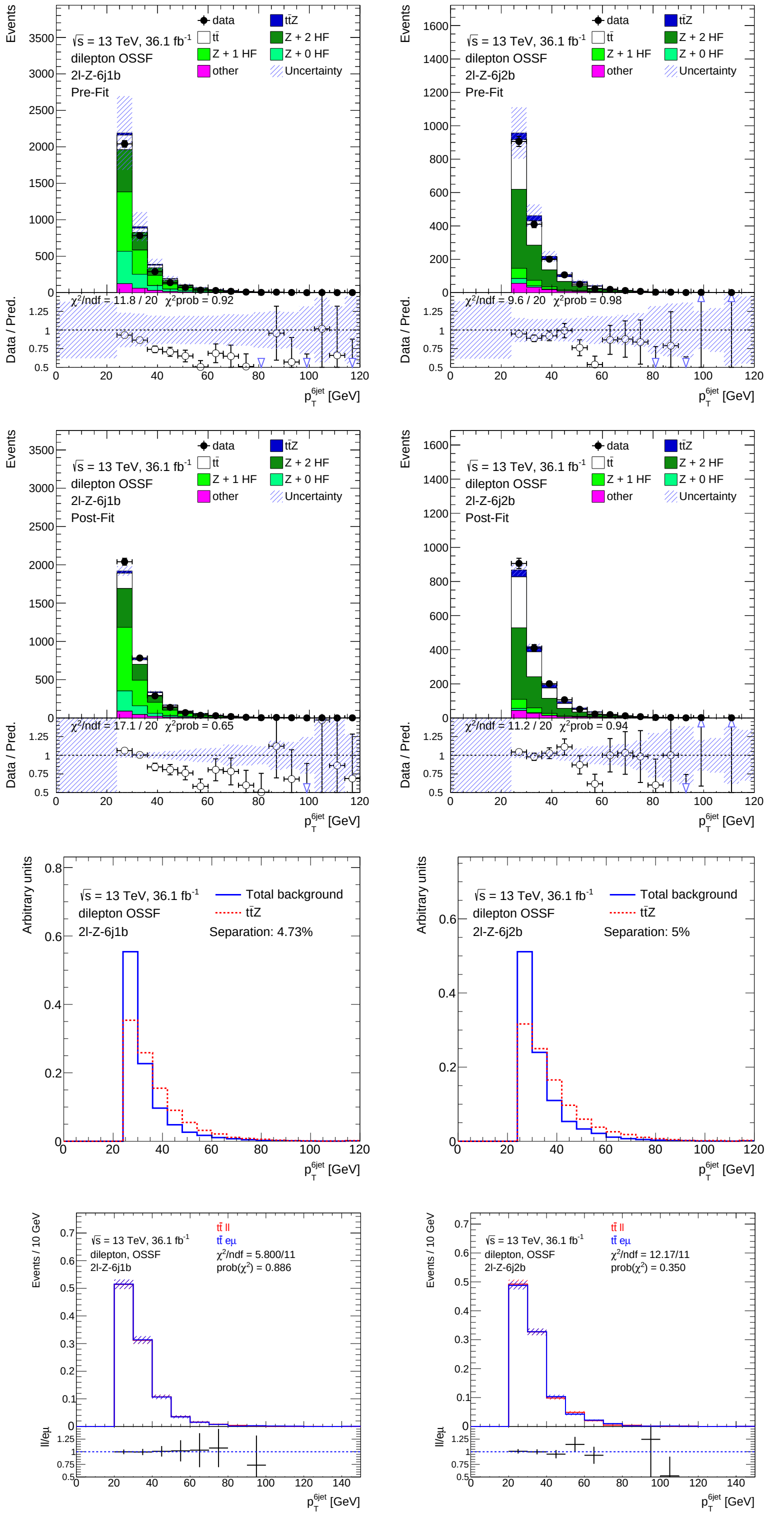


Figure 14: Transverse momentum of the sixth jet Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in two signal regions of the dilepton channel. The error bars include the systematic uncertainties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

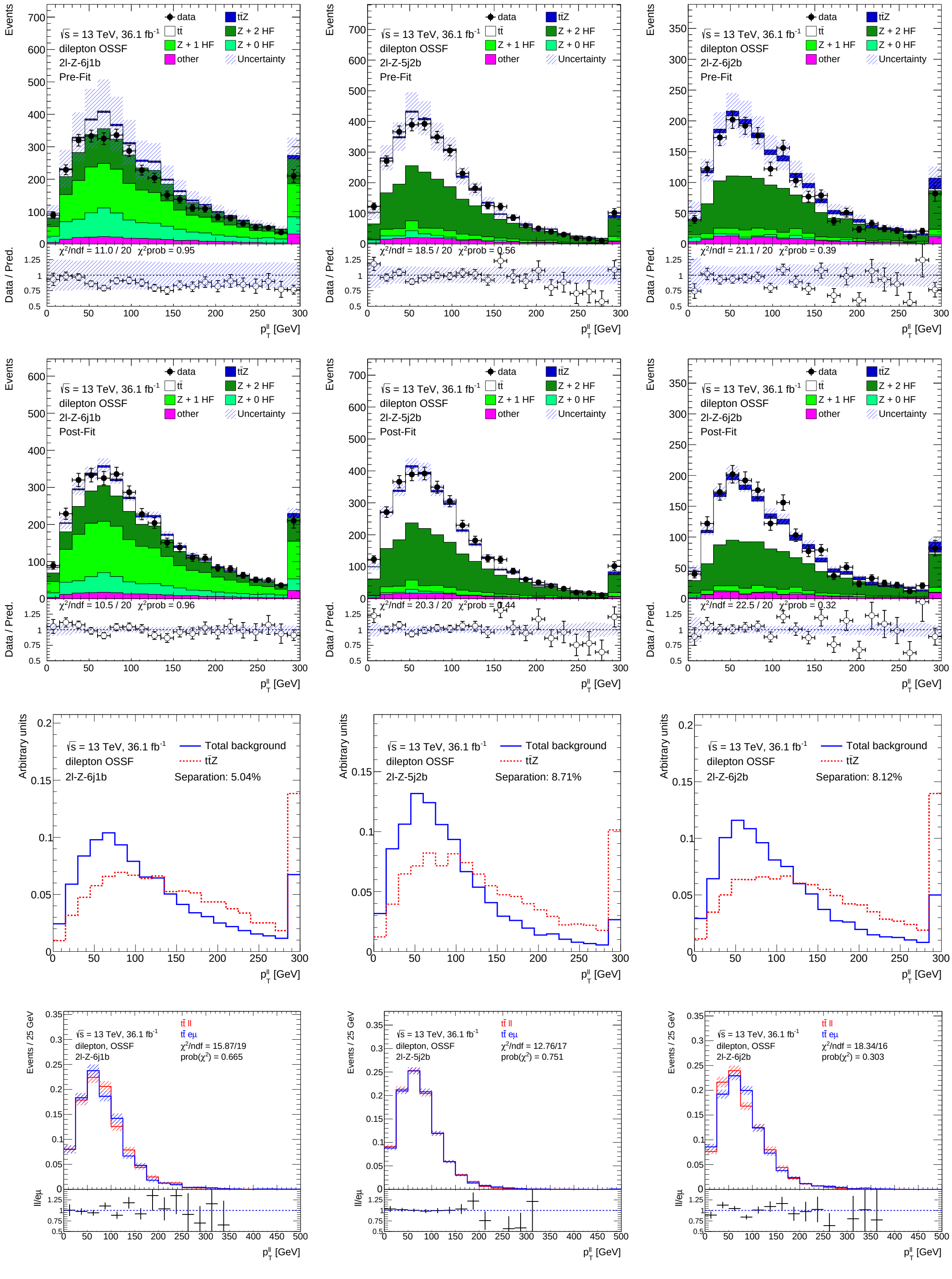


Figure 15: Dilepton transverse momentum, Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in three signal regions of the dilepton channel. The error bars include the systematic uncertainties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

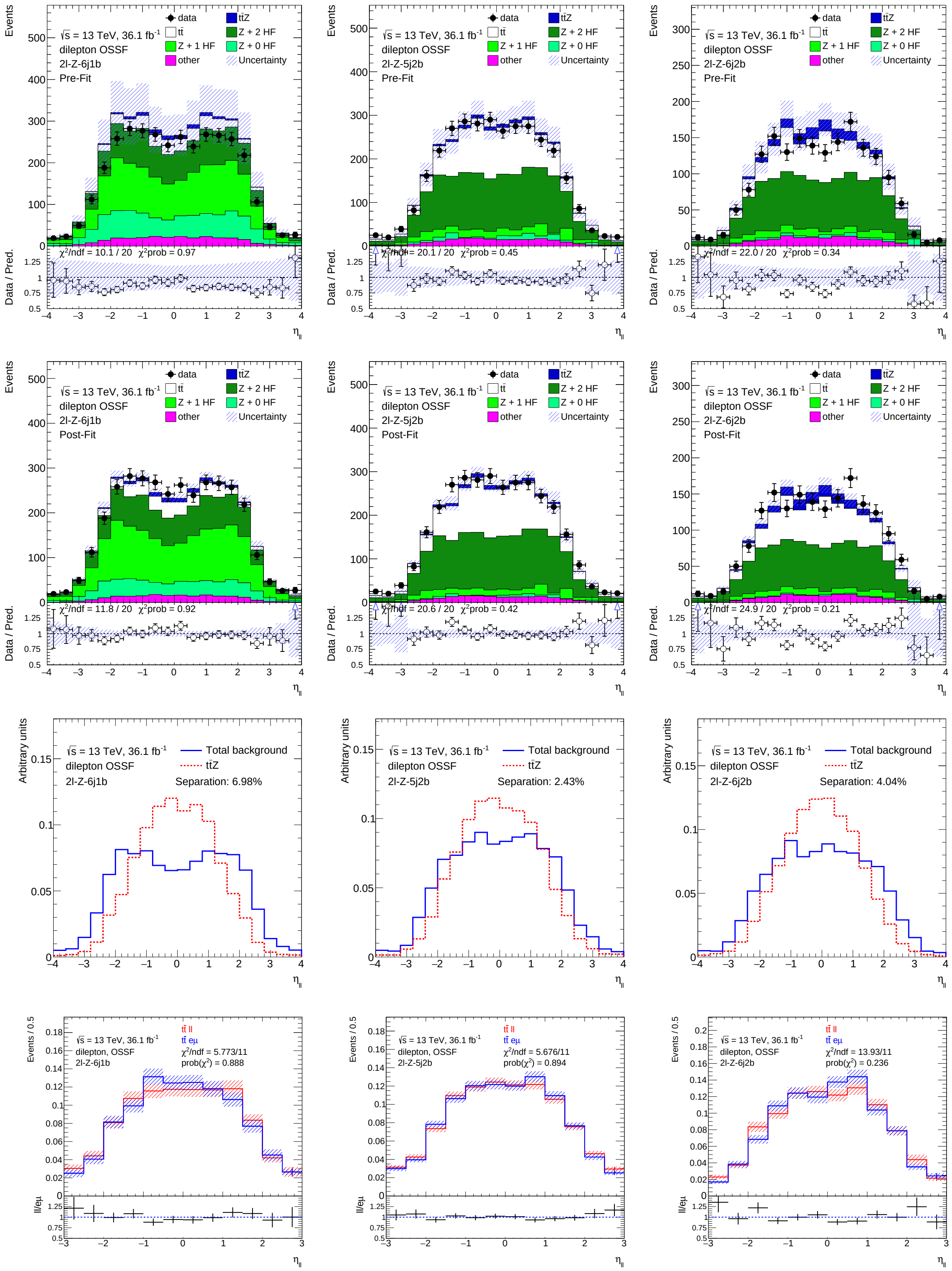


Figure 16: η of the dilepton system, Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in three signal regions of the dilepton channel. The error bars include the systematic uncertanties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

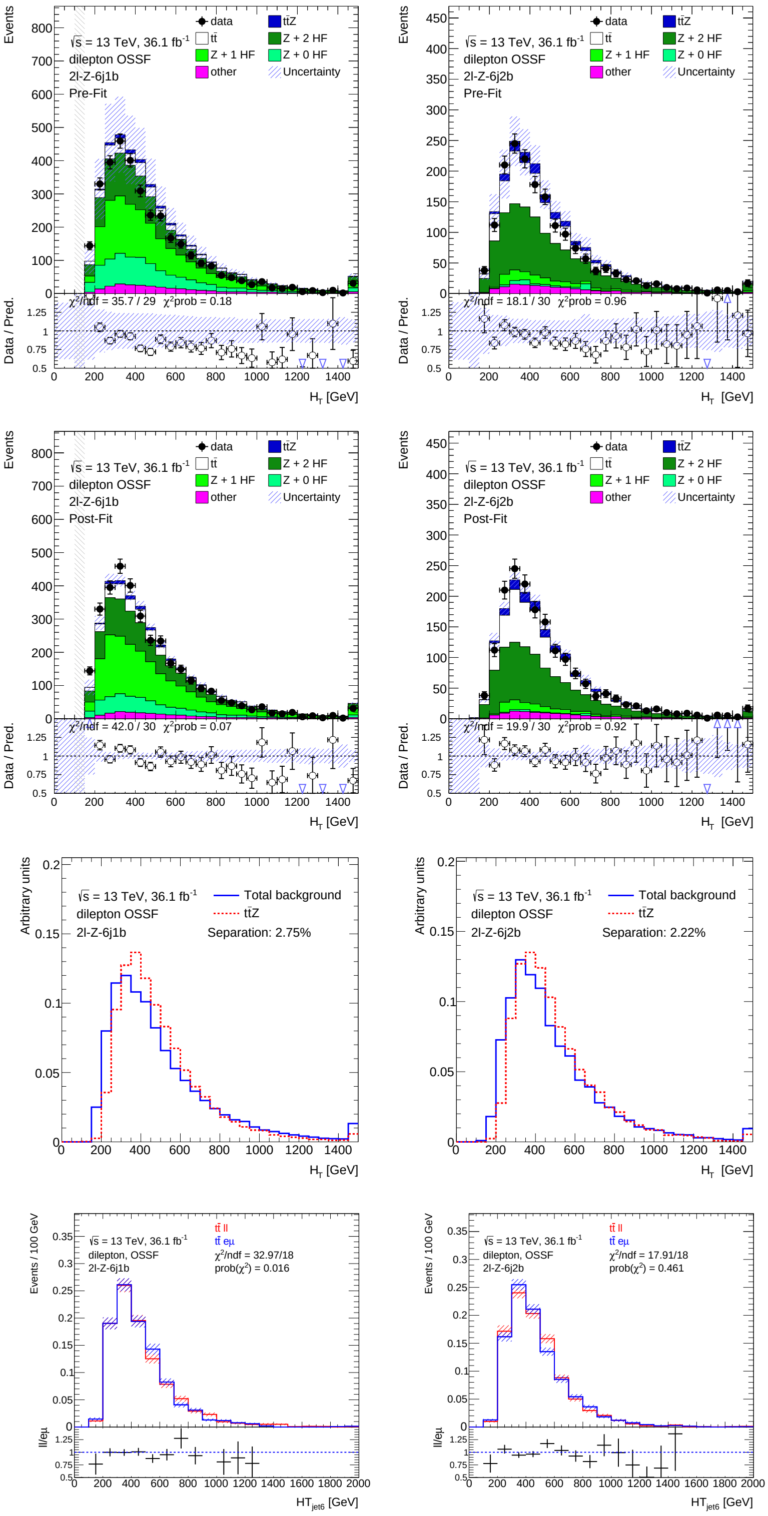


Figure 17: Scalar sum of transverse momenta of 6 leading jets in p_T, Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in two signal regions of the dilepton channel. The error bars include the systematic uncertanties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

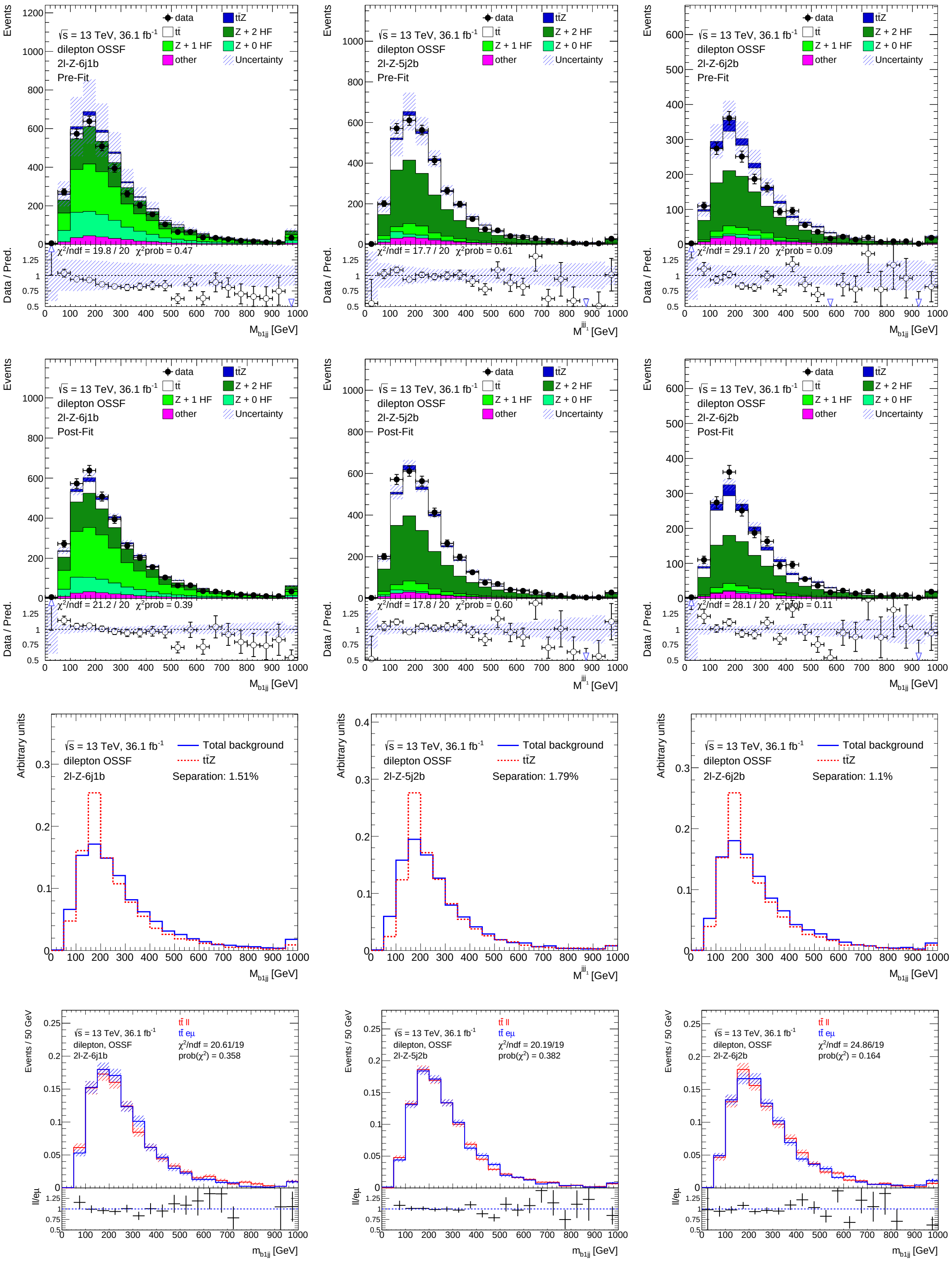


Figure 18: M_{jjj_1}, Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in three signal regions of the dilepton channel. The error bars include the systematic uncertanties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

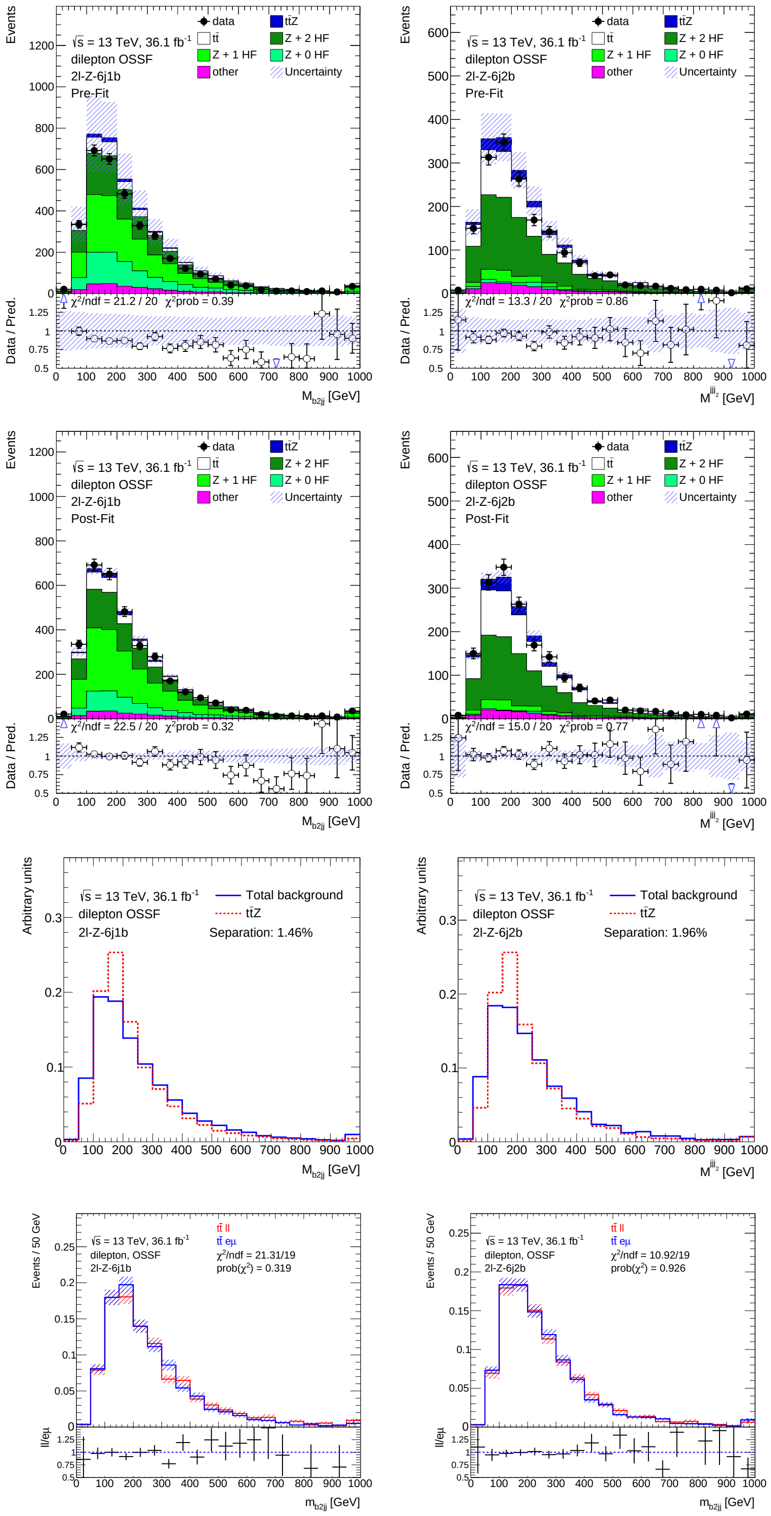


Figure 19: M_{jjj_2}, Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in three signal regions of the dilepton channel. The error bars include the systematic uncertanties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

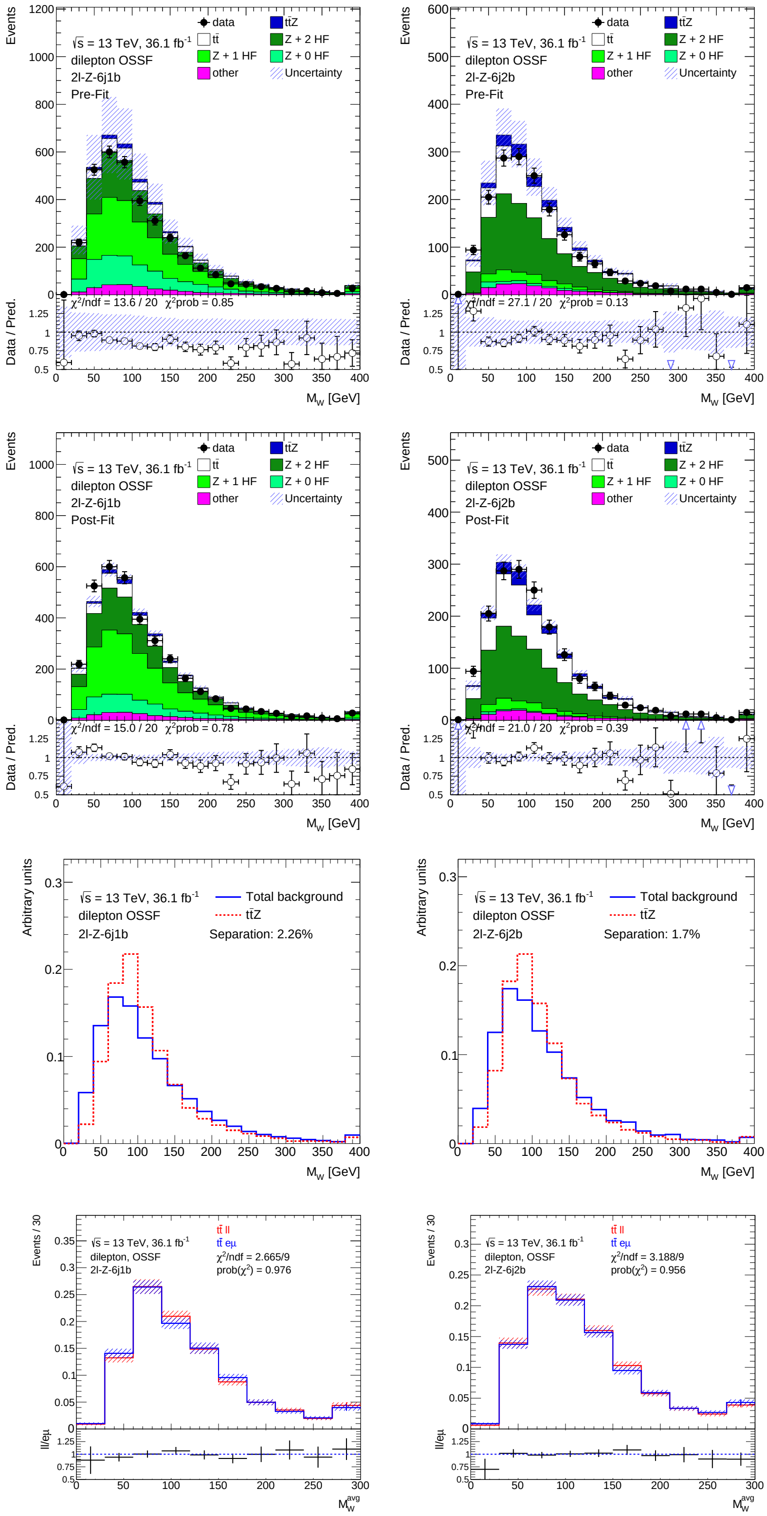


Figure 20: M_W^{avg}, Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in three signal regions of the dilepton channel. The error bars include the systematic uncertainties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

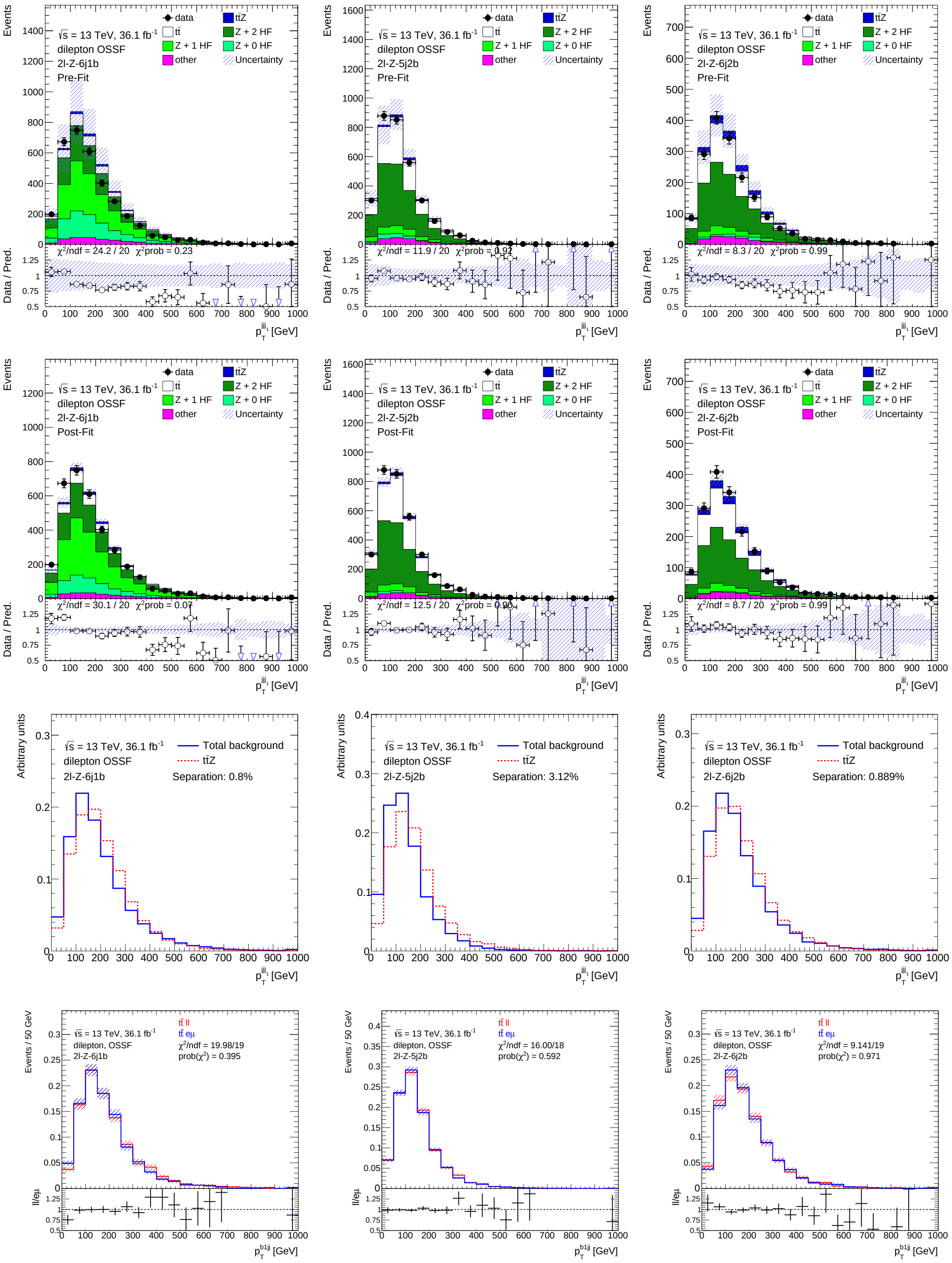


Figure 21: $p_T^{jjj_1}$, Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in three signal regions of the dilepton channel. The error bars include the systematic uncertainties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

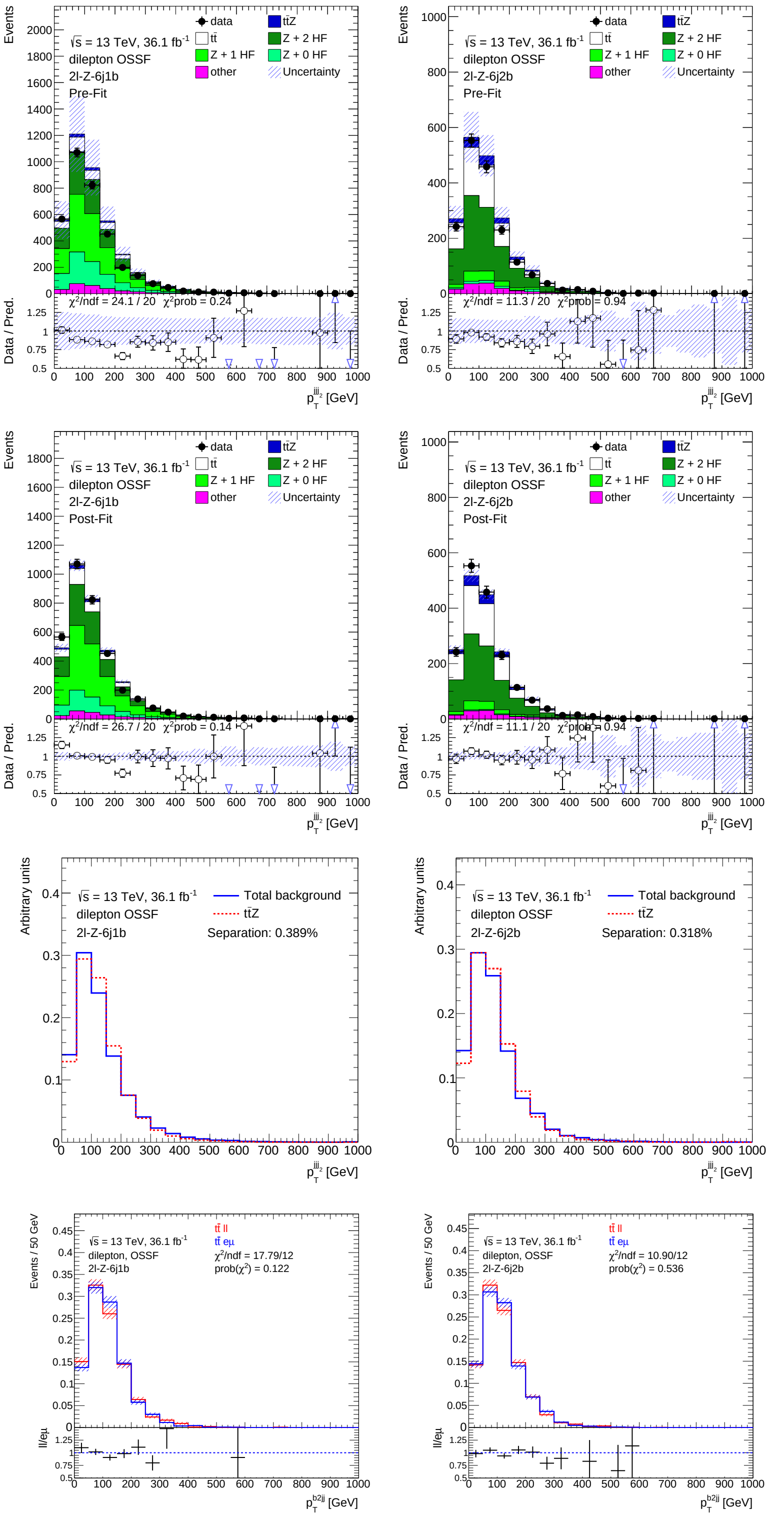


Figure 22: $p_T^{jjj_2}$, Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in two signal regions of the dilepton channel. The error bars include the systematic uncertanties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

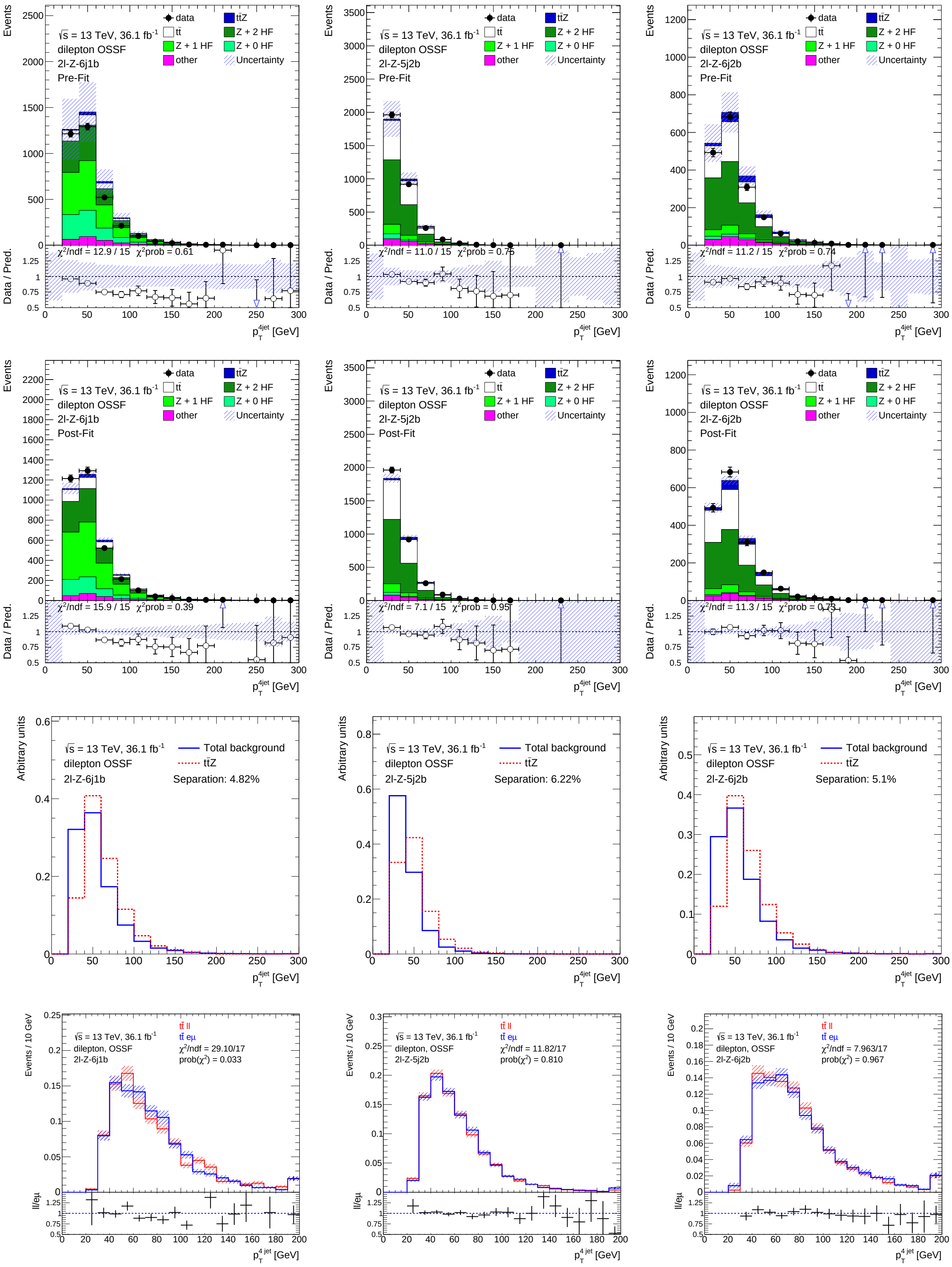


Figure 23: p_T^{4jet}, Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in three signal regions of the dilepton channel. The error bars include the systematic uncertainties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.

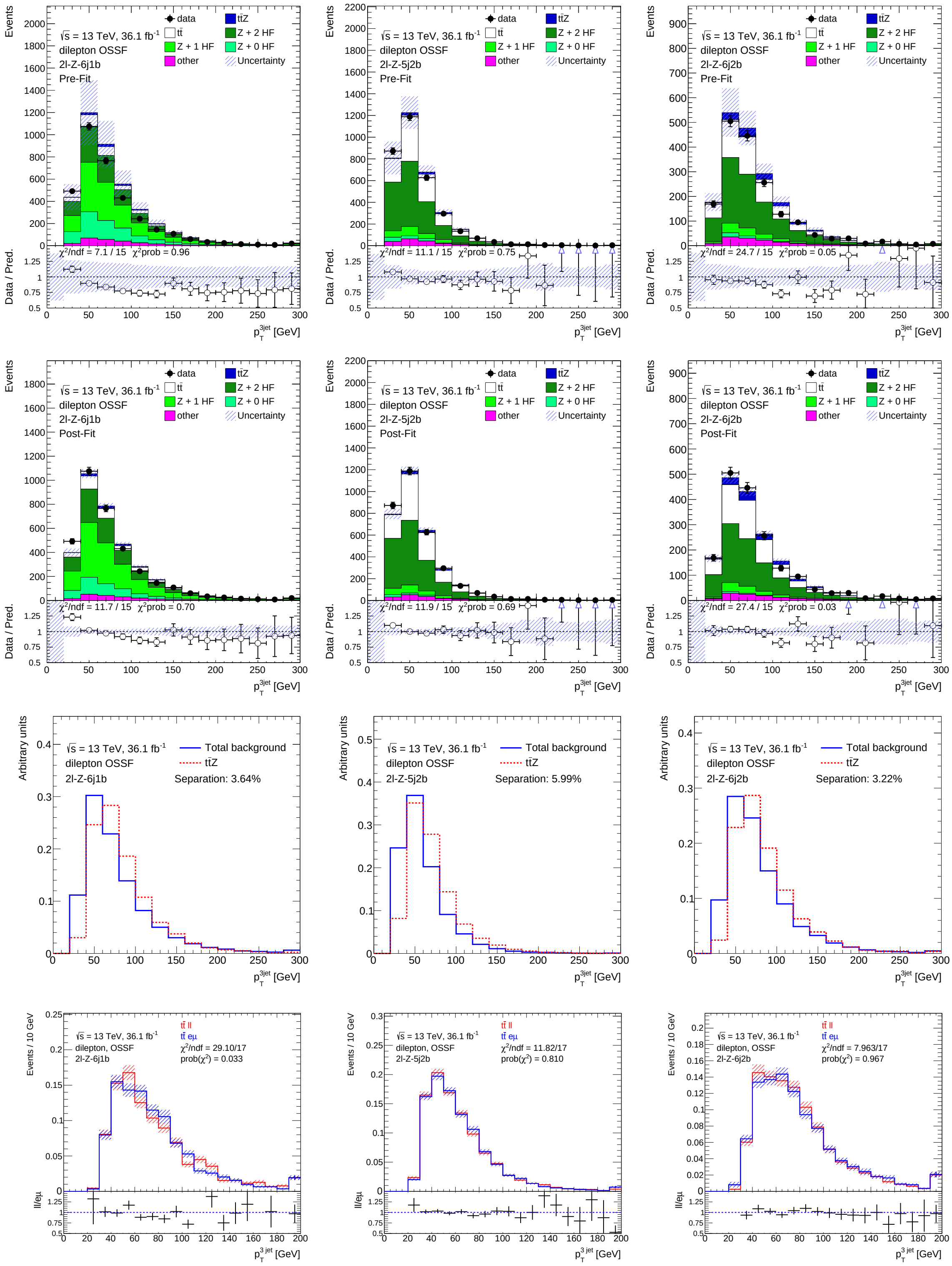


Figure 24: p_T^{3jet}, Pre-Fit (first line) and Post-Fit (second line) plots, separation plots (third line) and comparison of $t\bar{t}$ Monte Carlo predictions for $e\mu$ and $\ell\ell$ selections (fourth line) in three signal regions of the dilepton channel. The error bars include the systematic uncertainties defined in Sec. 6. The data-driven technique described in Section 8.5 is used to estimate the $t\bar{t}$ contribution.