Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

Michal Dubovský

Autoreferát dizertačnej práce

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

na získanie akademického titulu philosophiae doctor

v odbore doktorandského štúdia:

Jadrová a subjadrová fyzika

Miesto a dátum:

Bratislava, 15.5.2019

(2. strana autoreferátu)

Dizertačná práca bola vypracovaná
v dennej forme doktorandského štúdia

na Katedre jadrovej fyziky a biofyziky.

Predkladateľ: Mgr. Michal Dubovský
Katedra jadrovej fyziky a biofyziky
Fakulta matematiky, fyziky a informatiky
Univerzita Komenského
Mlynská dolina F1
842 48 Bratislava

Školiteľ: prof. RNDr. Stanislav Tokár DrSc.
Katedra jadrovej fyziky a biofyziky
Fakulta matematiky, fyziky a informatiky
Univerzita Komenského
Mlynská dolina F1
842 48 Bratislava

Študijný odbor: jadrová a subjadrová fyzika

Študijný program: jadrová a subjadrová fyzika (Jednoodborové štúdium, doktorandské III. st., denná forma)

Predseda odborovej komisie

1.1 Introduction

The top quark is the heaviest known elementary particle. It was discovered in 1995 by the experiments CDF[1] and D0[2] at Tevatron collider. Increasing the center of mass energy at hadron collider experiments, together with larger amount of collected data, brought new opportunities in the top-quark physics during the last years. While some processes with the top quark in the final state were not experimentally achievable with the data from Tevatron, they can be studied with a high precision at Large Hadron Collider (LHC).

The top quark can be produced alone, in so called single-top production, via weak interaction. It can be also produced in pairs via strong interaction, which is the process with the largest cross section with the top quark in the final state (in pp collisions at energies recently reached by LHC). Another production mechanism is associated production of the top-quark pair and other particles, such as photon, Z and W bosons, or Higgs boson. These processes will be labelled $t\bar{t}X$, where $X = W$, Z, H or γ. The cross sections of the $t\bar{t}X$ processes are significantly lower than the cross section of the top-quark pair production and their observation was not possible with the data collected by Tevatron (1983-2011) or LHC in Run I (2011-2012).

The LHC Run II. (2015-2018), delivered ≈ 140 fb^{-1} of pp collisions data at $\sqrt{s} = 13$ TeV to both ATLAS and CMS experiments. Considering the significant increase in $t\bar{t}X$ cross sections with respect to 8 TeV pp collisions and significantly larger integrated luminosity collected at 13 TeV, the observations of these processes became possible now, for the first time ever. This brought a lot of attention of ATLAS and CMS collaborations to these processes, aiming for measurement of their total and differential cross sections. These measurements provide an unique opportunity for precision tests of the Standard Model. The cross section of these processes depends on the top-quark coupling constant to the particle X. Therefore it can be sensitive to a new physics beyond the Standard Model which could modify the coupling constant. Any deviation from the value predicted by the Standard Model would be a sign of a new physics.

The associated production of the top-quark pair and Z boson is a rare process predicted by the Standard Model with the cross section

$$\sigma_{t\bar{t}Z} = 0.863\ ^{+8.5\%}_{-9.9\%}\ (\text{scale})\ ^{+3.2\%}_{-3.2\%}\ (\text{PDF} + \alpha_S)\ \text{pb}\ [3].$$

The gg fusion contributes by 70 %, $q\bar{q}$ annihilation contributes by 30 %. The precise knowledge of $t\bar{t}Z$ cross section is important for a few reasons. If the Z boson originates from the final state radiation, the cross section of the process is sensitive to the top-quark coupling constant to the Z boson. According to the Standard Model the coupling constant depends on the top-quark weak isospin. Measuring the $t\bar{t}Z$ cross section, the information about the coupling constant can be extracted. Any deviation from the Standard

Model prediction could be a sign of a new physics. Another reason to study the $t\bar{t}Z$ process is its role in other measurements. The $t\bar{t}Z$ produces up to four promt leptons in the final state. A similar signature can be also expected from other Standard Model rare processes such as $t\bar{t}H$ and also by processes predicted by super symmetry and other theories beyond the Standard Model. Being an important background in these measurements and searches, as well as providing a good ground for the precision Standard Model testing, makes the $t\bar{t}Z$ process important in the high energy physics and motivates its detailed study.

1.2 $t\bar{t}Z$ decay channels

Both top quark and Z boson are unstable particles with a very short lifetime in the order of 10^{-25} s. The top quark and Z boson decay before reaching a sensitive part of a detector. Therefore, only their decay products can be observed. With the probability of 99.8 %, the top quark decays into b quark and W boson. The W boson subsequently decays either into a quark anti-quark pair ($\mathcal{BR}(t \to bq\bar{q'}) = 67.41 \pm 0.27$ % [4]) or into lepton and anti neutrino ($\mathcal{BR}(t \to be^-\nu_e) = 10.71 \pm 0.16$ %, $\mathcal{BR}(t \to b\mu^-\nu_\mu) = 10.63 \pm 0.15$ %, $\mathcal{BR}(t \to b\tau^-\nu_\tau) = 11.38 \pm 0.21$ % [4]). The Z boson decays into $q\bar{q}$ pair (69.91 ± 0.06 % [4]), $\nu\bar{\nu}$ pair (20.00 ± 0.06 % [4]) or $\ell^+\ell^-$ pair (6.729 ± 0.008 % for e^+e^- and $\mu^+\mu^-$ together, 3.370 ± 0.008 % for $\tau\tau$ [4]).

Given a short mean life time of the τ leptons, $\tau_\tau \approx 2.9{\times}10^{-13}$ s, and their difficult experimental identification and reconstruction, this analysis does not target the decay channels with the τ leptons in the final state. The $t\bar{t}Z$ decay channels with the Z boson decaying into quarks or neutrinos suffer from a high background rate and their reconstruction and identification is very challenging. Therefore, only the decay channels with the Z boson decaying into e^+e^- and $\mu^+\mu^-$ are considered in this analysis.

Based on the decay channel of the top-quark pair, three $t\bar{t}Z$ decay channels are allowed for leptonically decaying Z boson.

The dilepton (2ℓ) channel, with the top-quark pair decaying hadronically. In the final state there are two b-jets and another four light jets from the top-quark pair decay, together with the two leptons from the Z boson decay. The branching ratio of the 2ℓ channel is 3.1 %. The 2ℓ decay channel is characteristic by a high background rate. The main background processes are Z+jets and dileptonically decaying top-quark pair with additional jets.

The trilepton channel is formed by leptonically decaying Z boson and lepton+jets decay of the top-quark pair. In the final state there are three charged leptons, one from the top quark and two from the Z boson, one neutrino, two b-jets and two light jets. The background rate is significantly lower compared to the 2ℓ channel and the branching ratio is 2.3 %. The main background processes are WZ, tWZ, tZ and fake leptons.

The tetralepton channel is formed by leptonically decaying Z boson and

dileptonically decaying top-quark pair, producing four leptons (two from the Z boson, two from the top-quark pair), two neutrinos and two b-jets. The 4ℓ channel has the smallest background rate from all the channels. The main sources of background are ZZ, tWZ and fake leptons. The branching ratio of the channel is 0.43 %.

1.3 Recent $t\bar{t}Z$ cross section measurements

Because of a low cross section the $t\bar{t}Z$ process was not experimentally available at Tevatron. During the LHC Run I, the ATLAS and CMS collaborations published results of $t\bar{t}Z$ searches (cross section measurements), reaching less than 5 σ signal significance (exclusion of background only hypothesis). At LHC Run II the ATLAS and CMS both published cross section measurements with 2015 datasets, still not reaching 5 σ significance. The first observation of the $t\bar{t}Z$ was made with 2015+2016 dataset by CMS collaboration [5], followed by ATLAS collaboration [6]. The results presented in the first part of this thesis were published in Reference [6]. The results presented in the second part of the thesis will be published in an currently ongoing ATLAS measurement of differential and total $t\bar{t}Z$ cross section.

All the $t\bar{t}Z$ cross section measurements published until now are summarized in Table 1. All of them are in a good agreement with the Standard Model predictions. The Standard Model predicts cross sections 0.137 pb at 7 TeV [7], 0.206 pb at 8 TeV [7] and 0.863 pb at 13 TeV [3] for pp collisions.

$\sqrt{s}$	$\mathcal{L}$ [fb^{-1}]	Collab.	Channels	Signif.	Cross section [pb]	Ref.
7 TeV	4.7	ATLAS	3ℓ	-	< 0.74 at 95 % CL	[8]
7 TeV	5	CMS	3ℓ	3.3 σ	$0.28^{+0.14}_{-0.11}(\text{stat.})^{+0.06}_{-0.03}(\text{syst.})$	[9]
8 TeV	20.3	ATLAS	2ℓ, 3ℓ, 4ℓ	4.2 σ	$0.176^{+0.058}_{+0.052}$	[10]
8 TeV	19.5	CMS	3ℓ, 4ℓ	3.1 σ	$0.20^{+0.08}_{-0.07}(\text{stat.})^{+0.04}_{-0.03}(\text{syst.})$	[11]
13 TeV	3.2	ATLAS	3ℓ, 4ℓ	3.9 σ	$0.92\pm0.29(\text{stat.})\pm0.10(\text{syst.})$	[12]
13 TeV	2.7	CMS	3ℓ, 4ℓ	3.6 σ	$1.07^{+0.35}_{-0.31}(\text{stat.})^{+0.17}_{-0.14}(\text{syst.})$	[13]
13 TeV	12.9	CMS	3ℓ, 4ℓ	3.9 σ	$0.70^{+0.16}_{-0.15}(\text{stat.})^{+0.14}_{-0.12}(\text{syst.})$	[14]
13 TeV	35.9	CMS	3ℓ, 4ℓ	> 5 σ	$0.99^{+0.09}_{-0.08}(\text{stat.})^{+0.12}_{-0.10}(\text{syst.})$	[5]
13 TeV	36.1	ATLAS	2ℓ, 3ℓ, 4ℓ	8.9 σ	$0.95\pm0.08(\text{stat.})\pm0.10(\text{syst.})$	[6]
13 TeV	77.5	CMS	3ℓ, 4ℓ	> 5 σ	$1.00^{+0.06}_{-0.05}(\text{stat.})^{+0.07}_{-0.06}(\text{syst.})$	[15]

Table 1: Measurements of the $t\bar{t}Z$ cross section performed by ATLAS and CMS collaboration until now. The last CMS publication at 77.5 fb^{-1}also provided the first differential cross section measurement in the variables related to the leptons originating from the Z boson decay.

2 2ℓ channel analysis at 2015+2016 dataset

The $t\bar{t}Z$ cross section measurement performed by ATLAS collaboration at 36.1 fb^{-1}of data corresponding to 2015+2016 dataset included three $t\bar{t}Z$

decays channels (2ℓ, 3ℓ and 4ℓ) and two decay channels of $t\bar{t}W$ (2ℓ SS and 3ℓ). The $t\bar{t}W$ cross section measurement was included in the analysis since many aspects of the analyses were the same and the fit could profit from including other regions in order to constrain effects of some systematic uncertainties. The knowledge of $t\bar{t}Z$ and $t\bar{t}W$ cross sections results is also important for effective field theory interpretation.

The thesis is focused mostly on the analysis in the 2ℓ channel, which is the channel with the highest background rate. There are 2 leptons from the Z boson decay, 2 b quarks and 4 light quarks from the top-quark pair decay. The leptons from Z boson can be measured directly in the detector, while the quarks hadronize and produce streams of particles, jets. The event selection is based on this final state signature.

2.1 Event Selection

Single lepton triggers are used to choose the initial dataset. Exactly two leptons in the event are required. The leptons must be of the same flavour and opposite sign of the electric charge. The difference between the lepton pair invariant mass and the Z boson mass (91.2 GeV) is required to be less than 10 GeV. Transverse momentum of the leading (subleading) lepton is required to be higher than 30 (15) GeV. In addition to these cuts three signal regions are defined based on the number of jets and b-jets in the event. At most one b-jet is allowed not to be b-tagged, or one light jet is allowed not to be reconstructed. The 2ℓ-Z-6j1b signal region contains events with exactly one b-jet and at least 6 jets, targeting the case with one b-jet not being b-tagged. The 2ℓ-Z-5j2b signal region contains events with at least two b-jets and exactly 5 jets, targeting the case with one light jet not being reconstructed. The 2ℓ-Z-6j2b signal region contains events with at least two b-jets and at least 6 jets, targeting the ideal case, when both b-jets are tagged and all 4 light jets are reconstructed.

2.2 Event Yields

Applying the already mentioned selection to data and Monte Carlo samples, the yields summarized in Table 2 can be obtained. The contribution of $t\bar{t}$ background is estimated using a data driven technique, which will be described later. The 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 the modelling of heavy-flavour jets is problematic and can be mismodelled by Monte Carlo. The last two rows show the signal purity, S/(S+B), and statistical only significance of signal, $\mathrm{S}/\sqrt{S+B}$, in other words, what is the ratio of the signal to the expected statistical uncertainty of the total number of events in the region.

The fraction of signal events is too low to observe the signal from the

	2ℓ-Z-6j1b	2ℓ-Z-5j2b	2ℓ-Z-6j2b
$t\bar{t}Z$	80.7 ± 11.2	71.4 ± 4.48	126 ± 16.7
$t\bar{t}$	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 2: 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.

event yields only. In the most sensitive region, 2ℓ-Z-6j2b, the statistical only significance of the signal is 2.9 σ. Once the systematic uncertainties are included, the total significance would be strongly reduced if this approach was used. Applying further cuts in order to enhance the signal fraction in signal regions, the number of available events would be decreased significantly and thus it would not be beneficial. A more sophisticated approach is necessary to increase the signal sensitivity. A multivariate analysis is employed for this task.

2.3 Neural Network

The first multivariate analysis that was tested and optimized to separate signal and background in the 2ℓ channel is a Neural Network, implemented in NeuroBayes package [16]. The Neural Network is an algorithm that combines a set of input variables, $\vec{x}$, to calculate the output variable y related to the probability for an event to be the signal. It is a way how to combine a set of input variables in order to obtain a single variable with better signal vs. background separation than any input variable alone.

The set of input variables listed in Table 3 was considered. The motivation of variables is summarized in the Thesis. Some of them target reconstruction of the Z boson, top quark or W bosons from the top-quark decay. Other variables are sensitive to a boost in the direction of z-axis, which is higher for the Z+jets background compared to the $t\bar{t}$ or $t\bar{t}Z$. Some

variables aim for reconstruction of various object pairs which are expected to have different properties between the signal and background events. A subset of variables uses the fact that the total energy of the collisions is higher in the signal events. The variables have been ranked by the total correlation to the desired NN output (1 for signal, -1 for background). The ranking of the variables in 3 analysis regions is also summarized in Table 3.

The multivariate analysis is used in two phases: training and testing (application) phase. In the training phase, a known sample is provided to the MVA (usually Monte Carlo simulation), so the MVA can learn signal and background patterns. In the testing (application) phase the MVA response is checked on a statistically independent sample in order to validate the MVA ability to separate signal from background. If the number of events in the training dataset is too low or if too many input variables are used in the training, the MVA can learn statistical fluctuations of the training sample and its performance on an independent sample is worse. The effect of MVA being trained on the statistical fluctuations of the training sample is called an overtraining. In order to reduce the overtraining, it is desired to keep the number of input variables as low as possible. Another reason to keep the number of input variables low is to avoid a possible missmodelling of the probability density function in the space of input variables. The higher is the number of input variables, the higher is the risk of missmodelling.

In order to estimate the optimal number of NN input variables, the expected statistical only significance obtained from the given region was used. The NN was trained firstly using only the best variable from the ranking. The distribution of the NN output was scanned and it was searched for the cut reaching the highest statistical significance of the signal ($S/\sqrt{S+B}$). Once the cut was found, process was repeated for the first two variables from the ranking. The variables were being added one by one until the full set of the variables was included and $max(S/\sqrt{S+B})$ was estimated for all numbers of input variables.

This dependency of $max(S/\sqrt{S+B})$ on the number of NN input variables is shown in Figure 1.

Figure 1 shows that for more than 12 variables there is only a very small improvement in $max(S/\sqrt{S+B})$. In other words, adding more than 12 variables is not very beneficial and could lead to problems related to the overtraining and missmodelling of the input variables. The set of 12 variables from the top of the ranking was used as the optimal set of the NN input variables.

The distributions of the NN output for 3 signal regions obtained from the simulation and data, together with the overtraining checks are shown in Figure 2. No significant overtraining or missmodelling of the NN output was found.

Variable	Definition	6j1b	5j2b	6j2b
p_{T}^{ll}	p_{T} of the lepton pair	9	3	1
p_{T}^{3jet}	p_{T} of the third jet	17	9	17
p_{T}^{4jet}	p_{T} of the fourth jet	2	1	4
p_{T}^{6jet}	p_{T} of the sixth jet	8	-	10
ΔR_{ll}	ΔR between the two leptons	15	17	18
$N_{jetpairs}^{Vmass}$	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)	1	2	2
$N_{bjj}^{top-mass}$	number of 3 jets combinations (with exactly 1 b-jet) close to the top-quark mass ($\lvert M_{bjj} - M_{top}\rvert < 15$ GeV) and ($\lvert M_{jj} - M_W\rvert < 15$ GeV)	4	5	5
M_{jj}^{MindR}	mass of the combination between any two jets with the smallest ΔR	6	7	8
M_{uu}^{Ptord}	mass of the two untagged jets with the highest p_T	10	12	16
M_{bb}	mass of the two jets with the highest b-tagging weight (output from MV2C10 tagger)	12	4	3
ΔR_{bb}	cone between two jets with the highest b-tagging weight in the event	16	14	15
$\mathrm{Cent}_{\mathrm{jet}}$	scalar sum of p_{T} divided by sum of E for all jets	7	6	6
ΔR_{ave}^{jj}	average ΔR for all jet pairs	21	15	20
$\max M_{lepb}^{MindR}$	maximum mass between a lepton and the b-tagged jet with the smallest ΔR	13	13	13
$H1$	Second Fox-Wolfram moment	5	8	11
$H1_{jet}$	Second Fox-Wolfram moment built from jets only	23	19	22
H_{T}^{6jets}	sum of jet p_{T}, up to 6 jets	20	-	23
$p_{\mathrm{T}}^{jjj_1}$	p_{T} of 3 jet system formed adding the nearest 2 jets in ΔR to the jet with highest p_{T}	22	16	12
M_{jjj_1}	invariant mass of the 3 jet system used for $p_{\mathrm{T}}^{jjj_1}$	11	11	7
$\eta_{\ell\ell}$	η of dilepton system	3	10	9
$p_{\mathrm{T}}^{\mathrm{jjj}_2}$	p_{T} of 3 jet system built in the same way as jjj_1, not considering the jets included in jjj_1	19	-	21
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	18	-	19
M_W^{avg}	sum of the two closest 2 jet invariant masses from from jjj_1 and jjj_2 divided by 2	14	-	14
p_{T}^{b1}	p_{T} of the first b-jet. Jets are ordered according to the p_T	not considered		
p_{T}^{b2}	p_{T} of the second b-jet. Jets are ordered according to the p_T	-	18	-

Table 3: The definitions of input variables used in the Neural Network training and their ranking in the three analysis regions. Jets and leptons are ordered by their p_T from the highest one. To supress 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.

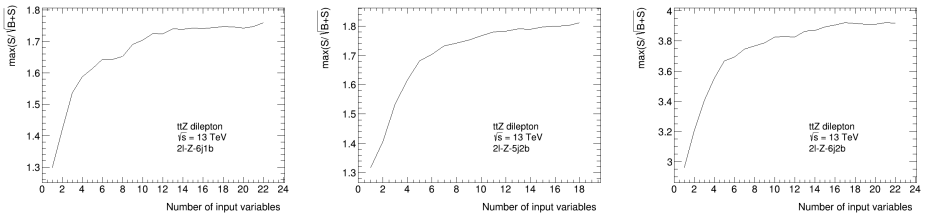

Figure 1: Dependency of estimated signal statistical significance on the number of input variables. For a given number of input variables the distribution of NN output is scanned with very narrow steps (0.001). At each step the $S/\sqrt{S+B}$ for events with NN output higher than the current point is calculated. The maximal value obtained for a given number of input variables is shown in the graphs above.

2.4 Boosted Decision Tree

As an alternative MVA method, the Boosted Decision Tree (BDT) was tested and optimized by another member of the analysis team. The set of input variables used in the BDT training and their ranking is shown in Table 4. The overtraining checks and distributions of the BDT output in simulation and data are shown in Figure 3.

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	$N_{jetpairs}^{Vmass}$	dR_{jj}^{ave}	$N_{jetpairs}^{Vmass}$
5	dR_{jj}^{ave}	H_{T}^{6jets}	dR_{jj}^{ave}
6	p_{T}^{4jet}	$N_{jetpairs}^{Vmass}$	p_{T}^{4jet}
7	dR_{ll}	M_{jj}^{MindR}	ΔR_{bb}
8	p_{T}^{ll}	dR_{ll}	p_{T}^{ll}
9	p_{T}^{6jet}	ΔR_{bb}	M_{bb}
10	M_{W}^{avg}	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	$M_{\mathrm{jj}}^{\mathrm{MindR}}$	$M_{\mathrm{uu}}^{\mathrm{Ptord}}$	$\mathrm{max}M_{\mathrm{lepb}}^{\mathrm{MindR}}$
14	$\mathrm{max}M_{\mathrm{lepb}}^{\mathrm{MindR}}$	p_{T}^{5jet}	$M_{\mathrm{jj}}^{\mathrm{MindR}}$
15	$M_{\mathrm{uu}}^{\mathrm{Ptord}}$		M_{W}^{avg}
16			p_{T}^{1b-jet}
17			$N_{bjj}^{top-mass}$

Table 4: Ranking of the variables used for BDT training.

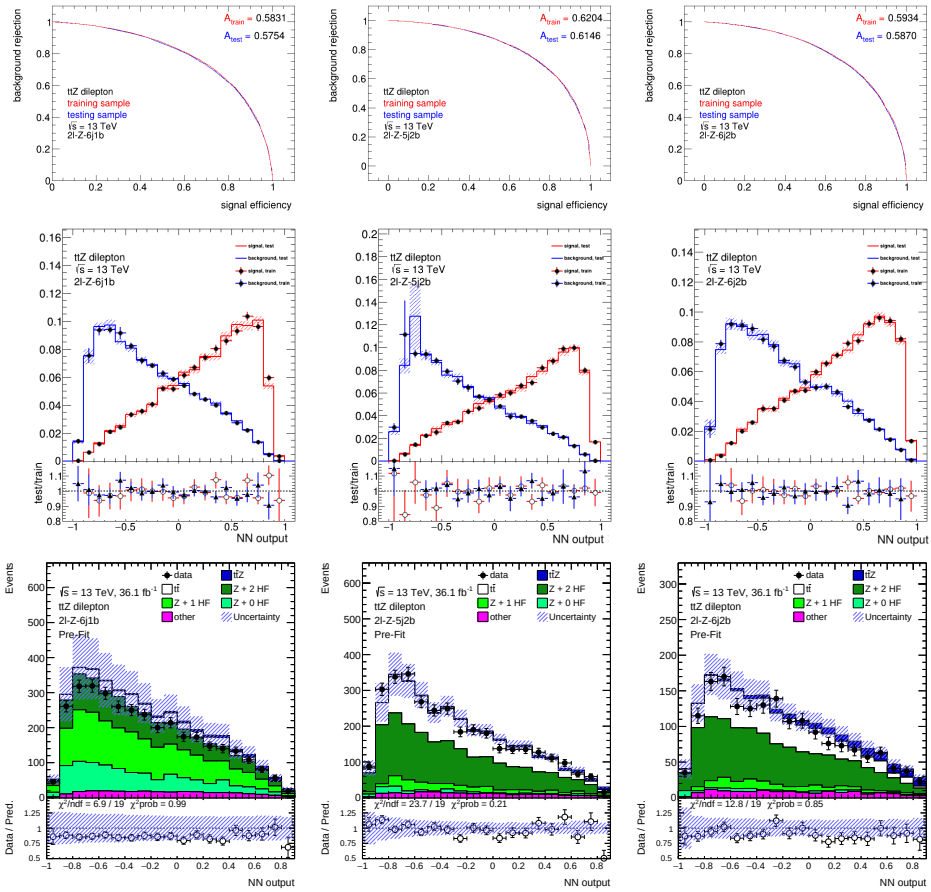

Figure 2: The ROC curves (first row), normalized distributions of the NN output (second row) for signal and background for both testing (solid line) and training samples (points) and distribution of NN output in simulation and data (third row). The error bars show only statistical uncertainty in the middle row and full statistical + systematic uncertainty in the last row. The 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 3.

2.5 Data driven $t\bar{t}$ background estimate

The dileptonically decaying top-quark pair is the second largest background in the 2ℓ channel. In order to reduce the systematic uncertainties related to the $t\bar{t}$ background and avoid a possible missmodelling of this background in Monte Carlo simulation, a data driven technique was used.

In the 2ℓ $t\bar{t}Z$ signal regions the leptons are required to be of the same flavour. Because of the lepton universality, the $t\bar{t}$ process contributes to

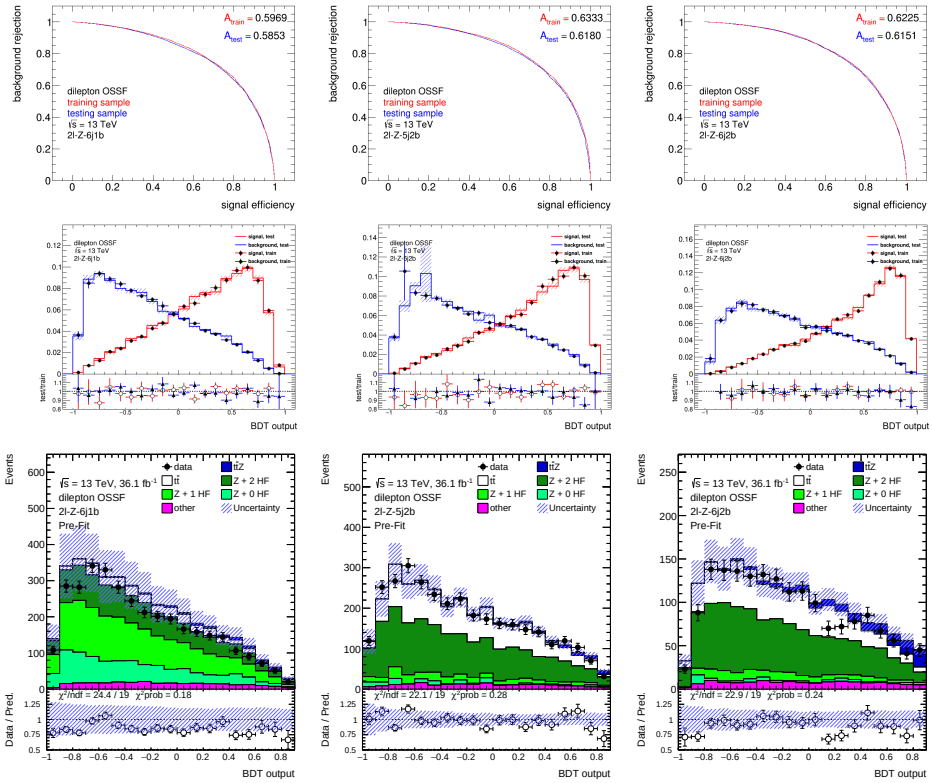

Figure 3: The ROC curves (first row) and normalized distributions of BDT output (second row) for signal and background for both testing (solid line) and training samples (points) and distribution of BDT output in simulation and data (third row). The error bars show only statistical uncertainty in the middle row and full statistical + systematic uncertainty in the last row. The 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 4.

same flavour (ee or $\mu\mu$) and different flavour channels ($e\mu$) equally. The $e\mu$ events from data are used in the analysis as the data driven estimate of the $t\bar{t}$ background in the regions with the same flavour of the leptons.

The non-$t\bar{t}$ background is subtracted from the $e\mu$ data. The rest is considered to be $t\bar{t}$ decaying into $e\mu$ (and two neutrinos and b-jets). Possible different detector acceptance for $e\mu$ events compared to the $\ell\ell$ events is taken into account, scaling the $e\mu$ data by the factor of

$$C_{t\bar{t}} = \frac{N_{t\bar{t}}^{\ell\ell}}{N_{t\bar{t}}^{e\mu}} = 0.981 \pm 0.030, \quad (1)$$

where $N_{t\bar{t}}^{\ell\ell}$ is number of $t\bar{t}$ events passing the $\ell\ell$ selection obtained from the Monte Carlo and $N_{t\bar{t}}^{e\mu}$ is the number of $t\bar{t}$ events passing the $e\mu$ selection obtained from the Monte Carlo. The uncertainty comes from a limited number of events in Monte Carlo and from a systematic uncertainty. The ratio was calculated inclusively from all regions. The ratios obtained from individual signal regions are consistent with this value, however, they suffer from a higher statistical uncertainty due to a limited number of events in individual signal regions. Shapes of the MVA input variables as well as distributions of MVA outputs were checked in $t\bar{t}$ Monte Carlo simulation in order to search for possible differences between $e\mu$ and $\ell\ell$ events in the shape of the variables. No statistically significant difference was found.

2.6 Fitting procedure and systematic uncertainties

A nuisance parameter fit is employed to extract the signal cross section. The fit has three free parameters: the signal cross section and normalizations of Z + 1HF and Z + 2HF background (associated production of Z boson and heavy flavour jets). The systematic uncertainties are taken into account via set of nuisance parameters, with prior Gaussian distributions. The value of a nuisance parameter equal to +1 means $+1\sigma$ systematic variation from the nominal distribution and the Gaussian term in the likelihood decreases the total likelihood by the factor of exp(-1/2) in this case. The fit searches for the maximum of the likelihood in order to extract the free parameters of the fit, as well as the post-fit values of the nuisance parameters.

All experimental uncertainties related to reconstruction and identification of the objects (jets, leptons, missing transverse energy) are considered in the analysis. The uncertainty related to the luminosity $\pm 2.1\%$ is considered in the fit as well as the uncertainty related to a different pile-up profiles in data and simulation.

The set of theoretical uncertainties related to signal modelling, uncertainties of scale choice, parton distribution functions uncertainties and cross section uncertainties are taken into account as the theoretical systematic uncertainties.

2.7 Asimov fit results

In order to make a choice between the NN and BDT, to optimize the binning choice and test the fit machinery, so-called Asimov fit was used. The Asimov fit is fit of Monte Carlo to the same Monte Carlo. The obtained mean values of the fit parameters are the same as these predicted by the Monte Carlo, but the fitted uncertainty provides an estimate of the expected uncertainty on the signal cross section, or any other parameter which is aimed to be extracted from the fit. The Asimov fit is used in order to find an analysis setup providing the best result, which means to minimize the expected signal

cross section uncertainty in the case of this analysis.

The optimal choice of the binning was found to be 19 bins of the equal width, covering the interval from -1.0 to 0.9 with the first bin including underflow and the last bin including overflow. The following values of the signal cross section were obtained from the Asimov fit when using NN and BDT.

$$\mu_{t\bar{t}Z}^{\mathrm{NN}} = 1.000^{+0.218}_{-0.212}\ (\mathrm{stat.})\ ^{+0.265}_{-0.233}\ (\mathrm{syst.}) \ = \ 1.000^{+0.343}_{-0.315} \tag{2}$$

$$\mu_{t\bar{t}Z}^{\mathrm{BDT}} = 1.000^{+0.207}_{-0.201}\ (\mathrm{stat.})\ ^{+0.233}_{-0.207}\ (\mathrm{syst.}) \ = \ 1.000^{+0.312}_{-0.288} \tag{3}$$

The Monte Carlo prediction of the $t\bar{t}Z$ cross section is taken as the unit of $\mu_{t\bar{t}Z}$. Since the fit of the BDT output distribution results in the lower signal cross section uncertainty, the BDT was chosen as the MVA used for the analysis.

2.8 Fit to data

Fitting the BDT output distributions (Figure 3, the bottom row) to data, the signal and Z+1HF and Z+2HF normalizations, summarized in Table 5, were obtained. The fitted value of the $\mu_{t\bar{t}Z}$ corresponds to the $t\bar{t}Z$ cross section

$$\sigma_{t\bar{t}Z}^{2l,measured} = \sigma_{t\bar{t}Z}^{theory} \times \mu_{t\bar{t}Z} = 0.636^{+0.152}_{-0.148}(\mathrm{stat.})^{+0.203}_{-0.190}(\mathrm{syst.})\ \mathrm{pb} = 0.636^{+0.254}_{-0.241}\ \mathrm{pb}. \tag{4}$$

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

Table 5: The fitted signal and Z+HF normalizations obtained from fit to data of the BDT output in the 2ℓ channel. The Monte Carlo prediction of the cross section is taken as the unit of μ.

The fitted signal cross section value is in agreement with the Standard Model prediction. The impact of the leading 10 systematic uncertainties is shown in Figure 4. The signal significance (exclusion of background only hypothesis) is 3.0 σ.

2.9 Results obtained from other channels

The $t\bar{t}Z$ cross section was measured in three decay channels: 2ℓ, 3ℓ and 4ℓ. The measurement was combined with the $t\bar{t}W$ cross section measurement,

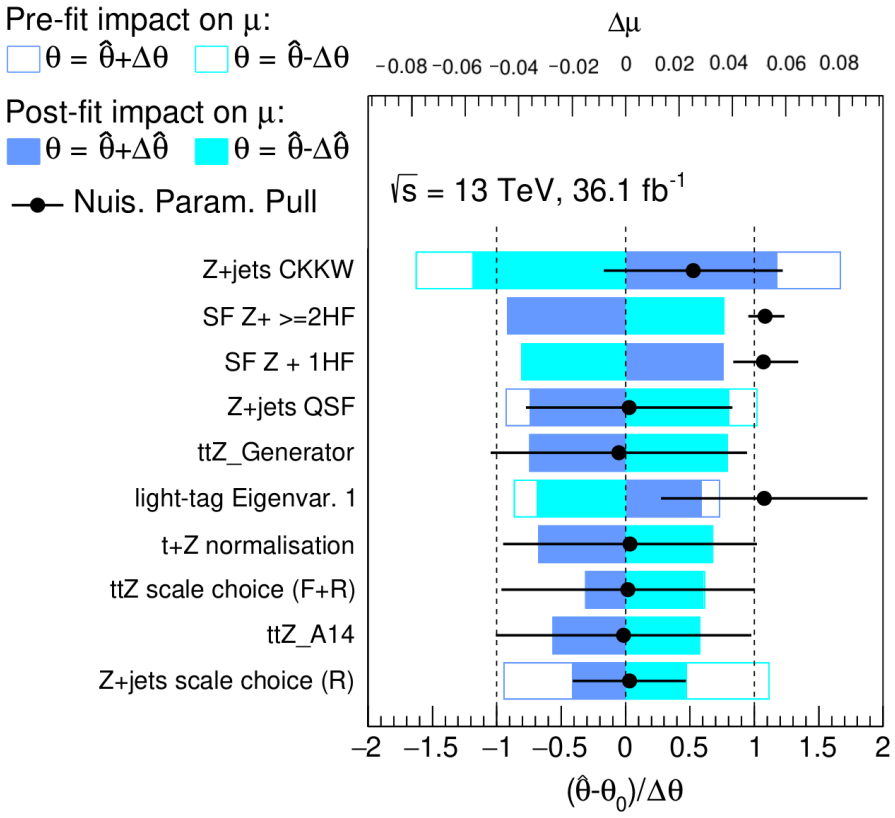

Figure 4: Effects of 10 leading systematic uncertainties in the fit of the BDT output distribution. The leading systematic uncertainty is shower matching scale for Z+jets background, followed by normalizations of Z+1 HF and Z+2HF backgrounds.

including 2ℓ same-sign and 3ℓ channels of the $t\bar{t}W$ into the fit. The results obtained from the fit in individual channels of the analysis are summarized in Table 6.

The author of the thesis worked only on the analysis in the 2ℓ channel. Other results are obtained from the other members of the analysis team and they are summarized in References [6, 17].

Channel	Cross section [pb]
$t\bar{t}Z$ 2ℓ	$0.64^{+0.15}_{-0.15}(\text{stat.})^{+0.20}_{-0.19}(\text{syst.}) = 0.64^{+0.25}_{-0.24}$
$t\bar{t}Z$ 3ℓ	$0.95 \pm 0.11(\text{stat.}) \pm 0.11(\text{syst.}) = 0.95^{+0.16}_{-0.15}$
$t\bar{t}Z$ 4ℓ	$1.07^{+0.25}_{-0.22}(\text{stat.})^{+0.10}_{-0.11}(\text{syst.}) = 1.07^{+0.26}_{-0.25}$
$t\bar{t}W$	$0.85^{+0.20}_{-0.19}$

Table 6: The fitted values of $t\bar{t}Z$ and $t\bar{t}W$ cross sections obtained from individual fits in three $t\bar{t}Z$ channels and in two decay channels of $t\bar{t}W$.

Performing the combined fit with all the $t\bar{t}Z$ and $t\bar{t}W$ signal and control regions included in the fit, the following $t\bar{t}Z$ and $t\bar{t}W$ cross section values were obtained

$$\sigma_{t\bar{t}W}^{measured} = 0.87 \ \pm 0.13(\text{stat.}) \ \pm 0.14(\text{syst.}) \ \text{pb} \ = 0.87 \ \pm 0.19 \ \text{pb} \quad (5)$$

$$\sigma_{t\bar{t}Z}^{measured} = 0.95 \ \pm 0.08(\text{stat.}) \ \pm 0.10(\text{syst.}) \ \text{pb} \ = 0.95 \ \pm 0.13 \ \text{pb}. \quad (6)$$

The obtained $t\bar{t}Z$ cross section is in a good agreement with the Standard Model prediction, while the $t\bar{t}W$ cross section deviates by more than 1σ from the prediction [18, 3].

$$\sigma_{t\bar{t}W}^{\text{theory}} = 0.601^{+0.078}_{-0.069}(\text{scale}) \pm 0.017(\text{PDF}) \pm 0.017(\alpha_S)\text{pb} \quad (7)$$

$$\sigma_{t\bar{t}Z}^{\text{theory}} = 0.863^{+0.073}_{-0.085}(\text{scale})^{+0.028}_{-0.028}(\text{PDF} + \alpha_S)\text{pb} \quad (8)$$

Obtained signal significances are 8.9 σ for $t\bar{t}Z$ and 4.3 σ for $t\bar{t}W$.

2.10 Post-fit yields and BDT output distributions in the 2ℓ channel

Applying the signal and Z+jets background normalizations obtained from the combined fit, as well as the values of nuisance parameters with their uncertainties, the Monte Carlo prediction has been corrected, resulting in the event yields shown in Table 7 and BDT output distribution showed in Figure 5. A good agreement between data and simulation is observed in all regions.

	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 7: 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.

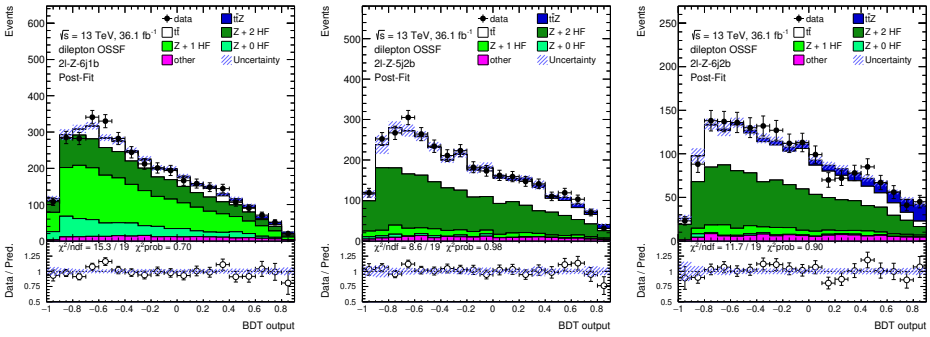

Figure 5: The Post-Fit distributions of BDT output in three dilepton signal regions.

3 4ℓ channel analysis in full Run II dataset

After finishing the analysis of the 2015 and 2016 dataset, the new analysis began, processing the full Run II dataset (2015-2018) corresponding to 139 fb^{-1}. The aim of the analysis is to improve the precision of the total cross section measurement and to measure a differential cross section of the $t\bar{t}Z$. The dilepton channel was excluded from the analysis because of its high background rate. Taking into account the results obtained from the previous analysis, summarized in Table 6, the 4ℓ channel can profit the most from the increasing number of events in data, since it was limited by the statistical uncertainty.

3.1 Event Selection

The following event selection is applied in order to target the events from the 4ℓ channel and to suppress a background, mostly from the ZZ background.

- exactly 4 leptons
- $p_{\mathrm{T}}^{1lep} > 27$ GeV, $p_{\mathrm{T}}^{2lep} > 20$ GeV, $p_{\mathrm{T}}^{3lep} > 10$ GeV
- at least one opposite sign same flavour (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 8. The OSSF lepton pair with the invariant mass closest to the Z boson mass is considered to originate from 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. If the Z_2 pair invariant mass is consistent with the Z-boson mass, the high cut on missing transverse energy is applied in order to suppress the ZZ background. If the mass is not consistent with the Z-boson mass, the cut is relaxed.

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, 40 GeV]	-	-

Table 8: The definitions of four signal regions and ZZ control region of tetralepton channel.

3.2 Event Yields

Applying the 4ℓ selection to the data and simulation, the event yields shown in Table 9 were obtained.

	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: 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 writing the thesis. Good data/MC agreement is observed in all regions.

3.3 Asimov fit results

Since the full Run II data analysis is still in progress and some samples for theoretical systematic uncertainties are not yet available, only Asimov fit results will be presented. Two free parameters are considered in the 4ℓ channel fit: signal and ZZ background normalizations. Contrary to the 2ℓ channel fit in the previous analysis, only one bin per region is used. There is no need to use an MVA in the 4ℓ channel, because of a low background rate in the signal regions.

Fitting the expected event yields (data-driven estimate of the fake leptons and Monte Carlo estimate of other processes) shown in Table 9, the following expected signal and ZZ normalizations were extracted from the Asimov fit.

$$\mu_{t\bar{t}Z}^{4\ell,\text{Asimov}} = 1.000^{+0.142}_{-0.132}\ (\text{stat.})\ ^{+0.058}_{-0.050}\ (\text{syst.})\ = 1.000^{+0.153}_{-0.142} \tag{9}$$

$$\mu_{ZZ} = 1.000^{+0.073}_{-0.068} \tag{10}$$

The expected signal significance is 9.6 σ.

Since theoretical uncertainties related to signal and background modelling are missing, the Result 9 is not the final result and the uncertainty is underestimated. In order to estimate the final uncertainty, considering also the systematic uncertainties which are currently not available, the impact of these uncertainties on the cross section uncertainty was assumed to be the same as in the previous analysis. The effects of the currently missing systematic uncertainties were checked in the previous analysis and they were added in square to the systematic uncertainty obtained from the Asimov fit

shown in Equation 9. The following estimate of the final $t\bar{t}Z$ cross section precision was obtained

$$\mu_{t\bar{t}Z}^{4\ell,\text{expected}} = 1.000^{+0.142}_{-0.132}\ (\text{stat.})\ ^{+0.074}_{-0.068}\ (\text{syst.})\ = 1.000^{+0.160}_{-0.149}. \quad (11)$$

This is just an approximate estimate of the final uncertainty based on the assumption that the effect of the missing systematic uncertainties was not changed by changes in the event selection and new calibrations. The uncertainty needs to be evaluated again once the samples for the currently missing systematic uncertainties become available.

4 Conclusion

The $t\bar{t}Z$ cross section measurement performed at 36.1 fb^{-1} of pp data collected by the ATLAS experiment in 2015 and 2016 is the first observation of the $t\bar{t}Z$ process by the ATLAS collaboration and the second observation, after the CMS paper [5], in general. The cross section was measured in combination of 3 $t\bar{t}Z$ decay channels (2ℓ, 3ℓ and 4ℓ) and two $t\bar{t}W$ decay channels (2ℓ SS and 3ℓ). The $t\bar{t}Z$ signal significance obtained from the measurement is 8.9 σ, together with the cross section value

$$\sigma_{t\bar{t}Z}^{measured} = 0.95\ \pm 0.08(\text{stat.})\ \pm 0.10(\text{syst.})\ \text{pb}\ = 0.95\ \pm 0.13\ \text{pb}. \quad (12)$$

The result is in a good agreement with the Standard Model prediction.

The thesis is focused on the analysis in 2ℓ channel. Because of a high background rate, a multivariate analysis needs to be used in this channel. Neural Network and Boosted Decision Tree algorithms were studied and optimized. Based on the results of the Asimov fit, the BDT was chosen as the final MVA choice. Fitting the BDT output distributions in 3 signal regions of the 2ℓ channel, the following value of the $t\bar{t}Z$ cross section with 3.0 σ significance was obtained

$$\sigma_{t\bar{t}Z}^{2l,measured} = 0.636^{+0.152}_{-0.148}(\text{stat.})^{+0.203}_{-0.190}(\text{syst.})\ \text{pb} = 0.636^{+0.254}_{-0.241}\ \text{pb}. \quad (13)$$

The fitted value is in a good agreement with the Standard Model prediction.

The other part of the thesis describes the measurement at the full Run II LHC dataset corresponding to 139 fb^{-1} in 4ℓ channel. Since some of the samples for theoretical systematic uncertainties estimate are still missing, only Asimov fit results are presented. Considering the same effects of the missing systematic uncertainties as in the previous analysis, the expected value of the fitted signal cross section is

$$\mu_{t\bar{t}Z}^{4\ell,\text{expected}} = 1.000^{+0.142}_{-0.132}\ (\text{stat.})\ ^{+0.074}_{-0.068}\ (\text{syst.})\ = 1.000^{+0.160}_{-0.149}. \qquad (14)$$

where the theoretical prediction of the $t\bar{t}Z$ cross section is taken as the unit of $\mu_{t\bar{t}Z}^{4\ell,\text{expected}}$.

References

[1] CDF Collaboration. Observation of top quark production in pbar-p collisions. *Phys. Rev. Lett. 74, 2626*, 1995. arXiv:hep-ex/9503002.

[2] D0 Collaboration. Observation of the top quark. *Phys. Rev. Lett. 74, 2632*, 1995. arXiv:hep-ex/9503003.

[3] 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.

[4] C. Patrignani et al. (Particle Data Group). *Review of particle physics.* CERN, Geneva, 2016.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] NeuroBayes twiki page. `https://twiki.cern.ch/twiki/bin/view/Main/NeuroBayes`, accessed: 2019-01-10.

[17] Olga et al. Bessidskaia Bylund. Measurement of the $tt\bar{Z}$ and $tt\bar{W}$ production cross sections in multilepton final states using 36.1 fb^{-1} of pp collisions at 13 TeV at the LHC. Technical Report ATL-COM-PHYS-2016-1730, CERN, Geneva, Nov 2016. ATL-COM-PHYS-2016-1730.

[18] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2016. arXiv:1610.07922.