Event selection and categorisation 5

A basic event selection is made for selecting signal like events and is discussed in Section 5.1. The necessary corrections in order to make simulation and data coherent, introduced in Chapter 4, are summarised in Section 4.5 and the resulting data/MC agreement is shown. One of the main background processes entering the analysis are background processes that have prompt leptons contaminated by real leptons either from decays of tau leptons or from hadronized mesons or baryons (so-called “non-prompt leptons") as well as by hadrons or jets misidentified as leptons. These two classes of contamination will be referred to as the not prompt-lepton (NPL) background process. The NPL background process is evaluated with a data-driven method discussed in Section 5.4. The analysis strategy is presented in Section 5.5, defining selection criteria to create signal and background regions to constrain the huge SM background compared to the expected signal.

5.1 Baseline event selection and filters

In this analysis a search is performed in a final state made up of a Z boson and a top quark, associated or not with a jet. The leptonic decay of the Z boson and the top quark is considerred for which the leading order Feynman diagrams can be seen in Figure 5.1 and Figure 5.2. The signal consists of the single top quark production through a FCNC tZq interaction (tZ in the final state) and the top quark pair production where on of the top quarks decays through the FCNC tZq vertex (tZq with $q = c, u$ in the final state). Their final state signatures consist of three leptons, only considering electrons or muons, and a jet originating from a b quark. For FCNC tZq, there is an additional up or charm jet. Leptons from tau decays are not vetoed and are entering the analysis via their leptonic decays. Four different lepton channels based on lepton flavour are considered: 3e, 2e1μ, 1e2μ, and 3μ.

The CMS collaboration recorded in the course of 2016, proton collisions data at a centre-of-mass of 13 TeV with a total recorded integrated luminosity of 35.9 fb^{-1}. The baseline event selection has as goal to substantially reject SM background events, whilst maintaining a high signal efficiency. The CMS trigger system, described in Section 2.2.3, filters out the main fraction of the collision events from uninteresting processes, and dedicated trigger paths are defined to single out the events with our required detector signature.

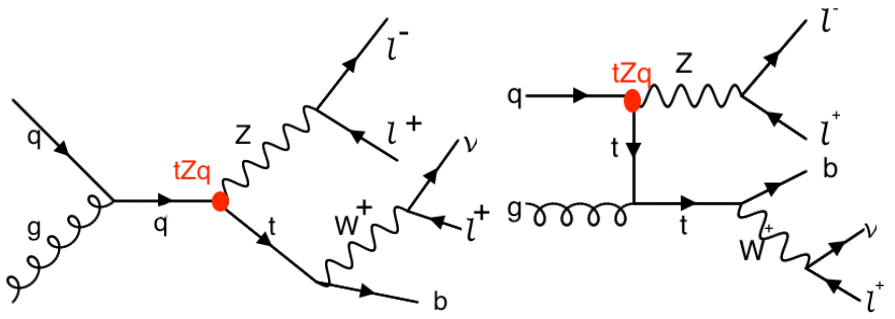

Figure 5.1: Single top quark Feynman diagrams at leading order. The vertex labelled tZq is the sought-for FCNC interaction.

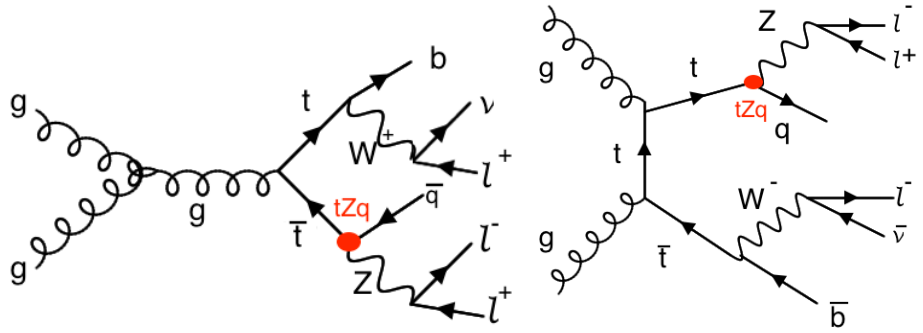

Figure 5.2: Top quark pair Feynman diagram at leading order. The vertex labelled tZq is the sought-for FCNC interaction.

The trigger paths are chosen based on online triggering objects with at least one muon (M), at least one electron (E), at least two muons (MM), at least two electrons (EE), at least one muon and an electron (ME), at least three muons (MMM), at least three electrons (EEE), at least two muons and one electron (MME), or at least two electrons and one muon (EEM). For the MC simulation a simple *or* of all triggers is taken, hence the event is considered when it passes one of the trigger paths. For data however, double counting of the same event has to be taken into account and a procedure to avoid double counting has been put into place. It consists of vetoing in a given dataset the events that are already selected in another, as given in Table 5.1.

For the single lepton triggers, at least one electron (muon) with a transverse momentum p_T higher than 32 (24) GeV is required. The dilepton triggers require the combination of an electron (muon) with $p_T > 23$ GeV and a muon (electron) with $p_T > 8$ GeV, or the combination of an electron (muon) with $p_T > 23$ (17) GeV and an electron (muon) with $p_T > 12$ (8) GeV. Events collected by the trilepton triggers require a combination of an electron (muon) with $p_T > 16$ (12) GeV, a second electron (muon) of $p_T > 12$ (10) GeV, and a third electron (muon) with $p_T > 8$ (5) GeV. The mixed trilepton trigger events require a combination of two electrons (muons) with $p_T > 12$ (9) GeV and a third muon (electron) with $p_T > 8$ (9) GeV. The HLT trigger paths used in data and simulation are summarised in Table 5.2.

Table 5.1: Trigger logic used to select data events in order to avoid double counting.

Dataset	Trigger Logic
1e1μ	EM || EEM || MME
2μ	(MM || MMM) && !(EM || EEM || MME)
2e	(EE || EEE) && !(MM || MMM) && !(EM || EEM || MME)
single μ	M && !(EE || EEE) && !(MM || MMM) && !(EM || EEM || MME)
single e	E && !M && !(EE || EEE) && !(MM || MMM) && !(EM || EEM || MME)

Table 5.2: HLT trigger paths used to select data and simulation events.

Trigger path name	Trigger type
HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_v	ME
HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_DZ_v	ME
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v	ME
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v	ME
HLT_DiMu9_Ele9_CaloIdL_TrackIdL_v	MME
HLT_Mu8_DiEle12_CaloIdL_TrackIdL_v	EEM
HLT_IsoMu24_v	M
HLT_IsoTkMu24_v	M
HLT_Ele32_eta2p1_WPTight_Gsf_v	E
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_v	MM
HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_v	MM
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v	MM
HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v	MM
HLT_TripleMu_12_10_5_v	MMM
HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v	EE
HLT_Ele16_Ele12_Ele8_CaloIdL_TrackIdL_v	EEE

In order to ensure a full trigger efficiency, the offline p_T thresholds are set higher than the online trigger thresholds. Selected offline electrons (muons) are required to have a $p_T >$ 35 (30) GeV and $|\eta| < 2.1$ (2.4). The electrons and muons corresponding to a tight working point, as discussed in Section 4.4.1 (Table 4.1) and Section 4.4.2 (Table 4.3), are used for analysis. Only events with exactly three leptons are being considered for the analysis. Events with extra leptons according to looser working points, as discussed in Section 4.4.1 (Table 4.1) and Section 4.4.2 (Table 4.3), are vetoed. The trigger efficiency estimation is described in Section 5.1.2 and is approximately 100%. To ensure that all reconstructed particles considered for the analysis are corresponding to a proton interaction and to remove signals from beam halo particles as well as detector noise, several filters are used. These are described in Section 5.1.1. In addition to three leptons, the selected events should at least contain one offline jet with a $p_T > 30$ GeV and $|\eta| < 2.4$, for which at least one jet is tagged as coming from a b quark.

5.1.1 Event cleaning

Some events arising from instrumental noise and beam backgrounds might end up in the data [191, 192]. In the ECAL, spurious deposits can appear from non-collision origins such as beam halo particles, or from particles hitting the sensors in the ECAL photo-detectors. Conjointly, dead ECAL cells can cause artificial missing transverse energy. The HCAL can also show spurious energy from particle interactions with the light guides and the photomultiplier tubes of the HF, as well as noisy hybrid photodiodes. In CMS, different algorithms, so-called filters, are developed to identify and suppress these events.

The ECAL electronics noise and spurious signals from particle interactions with photo detectors are mostly removed via topological and timing-based selections using only the ECAL information. The remaining effects such as anomalously high energy crystals and the lack of information for channels due to inefficiencies in the read-out are removed through dedicated events filters. Five ECAL endcap supercrystals have been identified for giving anomalously high energies due to high amplitude pulses in several channels at once, and are masked. Furthermore, the crystal read-out from a small amount of ECAL towers is not available. Nonetheless, their trigger primitive information is still available making it possible to estimate the magnitude of unmeasured energy and when the value is too large, the event is filtered out.

The machine induced particles via for example beam-gas, or beam-pipe interactions, that are flying with the beam, affect the physics analysis. They leave a calorimeter deposit along a line at constant ϕ in the calorimeter, and interactions in the CSCs will often line up with this deposit. This can be seen in Figure 5.3. Therefore, events containing such beam halo particles are removed from the selection with the CSC Beam Halo Filter. This algorithm uses information related to the geometric quantities, energy deposits, and timing signatures. For 2016 proton collision data, the filter rejects 85% in a halo-enriched sample, whereas the mistag probability determined from simulation is found to be less than 0.01%.

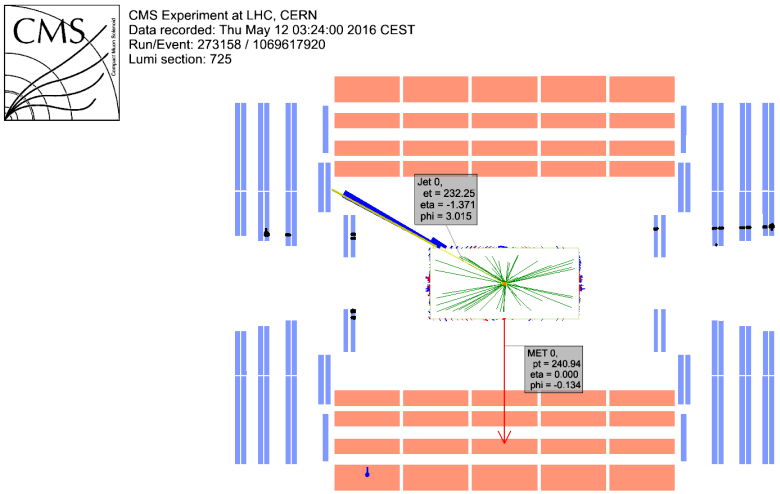

Figure 5.3: Event display of a beam halo event with collinear hits in the CSC (black), missing transverse energy of 250 GeV and a jet of 232 GeV. The hadronic deposit is spread in η, but narrow in ϕ. Figure taken from [191].

Furthermore, there is anomalous high missing transverse energy coming from low quality muons that lead to high-p_{T} tracks, but are considered not good by the particle flow algorithm. These low quality tracks will be mislabelled as charged hadrons and will therefore be used in the calculation of the missing transverse energy. By investigating the purity of the reconstructed tracks and the relative transverse momentum error of the muons, these events can be filtered out.

5.1.2 Estimation of the trigger efficiency

The trigger efficiency in data is estimated using a data sample collected using unprescaled $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers. These allow events with a combination of the missing transverse energy being higher than 110 GeV (120 GeV) and the scalar sum of the transverse momenta of the reconstructed PF jets $H_{\mathrm{T}}^{\mathrm{trig.}}$ being at least 300 GeV (120 GeV), or events for which the calorimeter (PF) $E_{\mathrm{T}}^{\mathrm{miss}}$ is higher than 200 GeV (300 GeV). For an HB-HE cleaned event, the PF missing transverse energy threshold is lowered to 170 GeV. These trigger paths are summarised in Table 5.3 and chosen to be completely uncorrelated with the lepton triggers given in Table 5.2.

Table 5.3: Unprescaled $E_{\mathrm{T}}^{\mathrm{miss}}$ HLT trigger paths for estimating the trigger efficiency. An *or* of the triggers is used to select events.

Trigger path	Requirement
HLT_PFHT300_PFMET110_v*	PF $E_{\mathrm{T}}^{\mathrm{miss}} > 110$ GeV, PF $H_{\mathrm{T}}^{\mathrm{trig.}} > 300$ GeV
HLT_MET200_v*	calorimeter $E_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV
HLT_PFMET300_v*	PF $E_{\mathrm{T}}^{\mathrm{miss}} > 300$ GeV
HLT_PFMET120_PFHT120_IDTight_v*	PF $E_{\mathrm{T}}^{\mathrm{miss}} > 120$ GeV, PF $H_{\mathrm{T}}^{\mathrm{trig.,tightWP}} > 120$ GeV
HLT_PFMET170_HBHECleaned_v*	PF $E_{\mathrm{T}}^{\mathrm{miss}} > 170$ GeV, cleaned for HB/HE anomalous signals

The trigger efficiency is studied on the main background, namely WZ+jets, with all corrections applied. For this study, the events passing a three-lepton cut and at least one jet are being used. The corresponding efficiencies are then calculated as

$$\epsilon_{data} = \frac{\text{Nb. of events passing lepton and MET triggers}}{\text{Nb. of events passing MET triggers}} \tag{5.1}$$

$$\epsilon_{MC} = \frac{\text{Nb. of events passing lepton triggers}}{\text{Nb. of total events}} \tag{5.2}$$

The resulting efficiencies for all lepton channels combined are shown in Table 5.4 and scale factors can be found in Table 5.5, where the scale factors are defined as

$$SF = \frac{\epsilon_{data}}{\epsilon_{MC}}. \tag{5.3}$$

Table 5.4: Trigger efficiencies on data events selected with $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers and WZ simulation for all lepton channels together. The unweighed number of events is quoted. When there are no events passing the cuts, it is indicated with N/A. The uncertainties are statistical uncertainties. Region A contains events after requiring three leptons and at least one jet. Region B has the same requirements as region A, but only events with exactly one jet that is b-tagged are considered. Region C also contains the requirements in region A, where at least one jet should be b-tagged. In region C, no events with a b-tagged jet are allowed.

Region	data		WZ simulations	
	Efficiency	unc.	Efficiency	unc.
A	99%	13%	100%	1%
B	100%	58%	100%	4%
C	96%	26%	100%	3%
D	100%	17%	100%	1%

Table 5.5: Trigger scale factors for each three-lepton lepton channel, after requiring three leptons and jets selection criteria, in the Z mass window.

all	3μ	3e	2e1μ	1e2μ
1.0000	1.0000	0.9541	1.0006	1.0004

More detailed scale factors and efficiencies can be found in Appendix A.

The trigger efficiencies are also measured as a function of the p_{T} of the leptons and the distributions of the scale factors can be found in Figure 5.4. The scale factors are in general close to unity, with the exception of the 3e lepton channel. This is due to a lack of events with a $p_{\mathrm{T}} > 250$ GeV. The trigger efficiencies are measured to be nearly 100% for both simulation and data. The results are dominated by statistics and assigning a large uncertainty to the trigger efficiency based on the dataset collected by $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers, would be over conservative. A one percent uncertainty on the trigger selection for the 2e1μ and 3μ final states, and 5% for the 3e and 1e2μ final states is assigned instead, in accordance the SM tZq search [196]. No scale factors will be applied on simulation as they are close to unity. Control plots are made in the dilepton region to validate all corrections applied to simulation.

5.2 Corrections

Mismatches between data and simulation are corrected via the use of scale factors. These are elaborately discussed in Section 4.4. In this section a short overview of the applied corrections on a dilepton dataset is given. Requiring three leptons enhances the fraction of NPL backgrounds. These backgrounds are not well simulated and are determined in a data-driven way. For this reason, the study of the agreement between data and simulation is performed with events selected with the trigger logic and trigger paths described in Section 5.1, that contain at least one opposite sign same flavour lepton pair that has an invariant mass m_{ll} inside a Z boson mass window of $|m_{\mathrm{ll}} - m_{\mathrm{Z}}| < 7.5$ GeV, and have jets present. The main contributing process in this selection if from the $\mathrm{Z}/\gamma^{*}+$jets process. In the following, the distributions relevant for

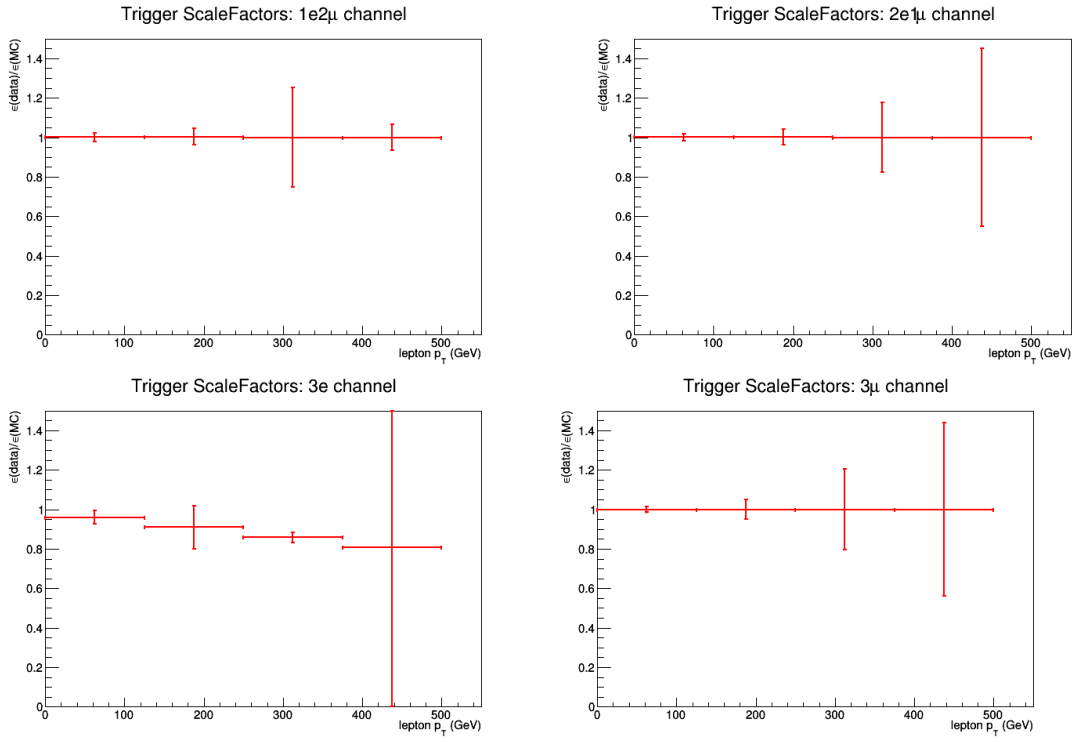

Figure 5.4: The trigger scale factors measured as a function of lepton p_{T} , using the dataset collected by $E_{\mathrm{T}}^{\mathrm{miss}}$ triggers and WZ simulation, after requiring three leptons and jets selection criteria, in the Z mass window. All corrections to simulation are applied. Left, upper: 1e2μ channel. Right, upper: 2e1μ channel. Left, lower: 3e channel. Right, lower: 3μ channel.

each correction are shown for 2e, 1e1μ, and 2μ lepton channels. Events with three leptons are not vetoed from the selection. The collection of these lepton channels are referred to as ”all channels“. In Figure 5.5, the number of events per leptonic decay are shown. The agreement for the three leptonic decays is shown for simulated samples. These samples will however not be used as explained in Section 5.4.

Pileup reweighting

In data, the number of interactions per bunch crossing (pileup) is calculated with a minimum bias cross section of 69.2 mb. The distribution of the number of simulated pileup events is then reweighed to match the expected number of pileup events in data. Pileup reweighting manifests itself as an altered shape of the number of reconstructed primary vertices as can be seen in Figure 5.6.

Note that Figure 5.6 indicates that even after pileup reweighting, the primary vertex multiplicity is not well described by simulation. This is a known effect, and using a minimum bias cross section with a slightly lower value is found to better describe the data. However, the b-tagging scale factors are only provided for the nominal inelastic cross section, and thus this value is used.

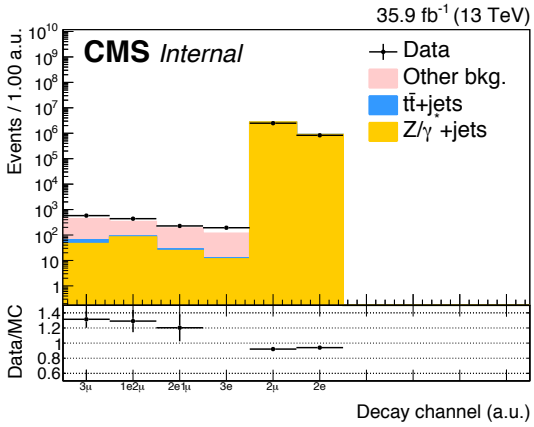

Figure 5.5: Number of events per leptonic-decay after requiring at least two leptons and jets, in the Z boson mass window. The different decays are not exclusive.

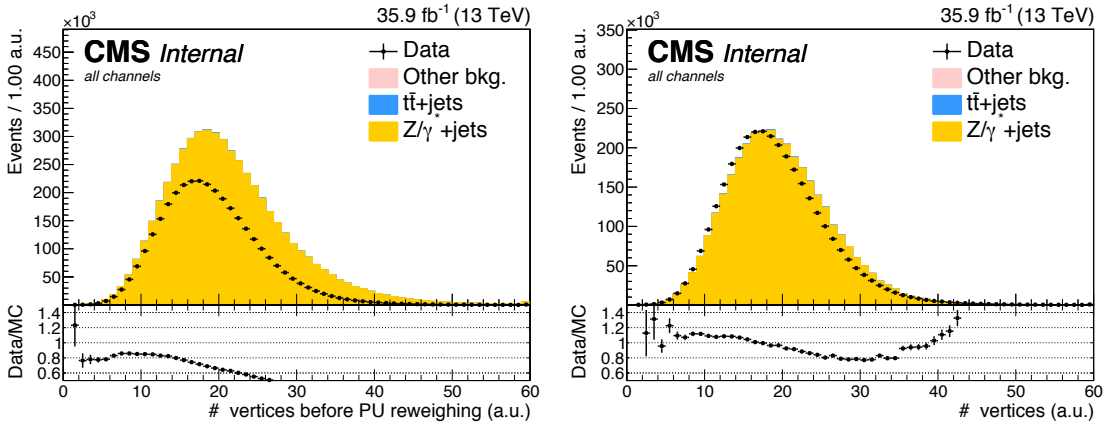

Figure 5.6: Distribution of the number of primary vertices before (left) and after (right) pileup reweighting. Events selected by requiring two leptons and jets, in the Z boson mass window.

Lepton scale factors

The efficiency to select leptons is different in simulation (ϵ_{MC}) compared to the data (ϵ_{data}). This is corrected for by applying lepton scale factors (SF) to the simulation that are defined as

$$SF = \frac{\epsilon_{\mathrm{data}}}{\epsilon_{\mathrm{MC}}}. \tag{5.4}$$

These scale factors are measured for the identification, isolation, tracking and trigger efficiencies of the objects as a function of p_{T} and η (see Section 4.4.1 and Section 4.4.2). Multiplying these scale factors for each lepton provides an overall efficiency per lepton i:

$$SF^{\mu}_{\mathrm{global}} = \prod_{\mathrm{i}}^{\#\mu} SF^{\mu}_{\mathrm{ID}}(p_{\mathrm{T}},\eta)\, SF^{\mu}_{\mathrm{Iso.}}(p_{\mathrm{T}},\eta)\, SF^{\mu}_{\mathrm{Trig.}}(p_{\mathrm{T}},\eta)\, SF^{\mu}_{\mathrm{Track}}(p_{\mathrm{T}},\eta), \tag{5.5}$$

$$SF^{e}_{\text{global}} = \prod_{i}^{\#e} SF^{e}_{\text{ID}}(p_T, \eta)\, SF^{e}_{\text{Iso.}}(p_T, \eta)\, SF^{e}_{\text{Trig.}}(p_T, \eta)\, SF^{e}_{\text{Track}}(p_T, \eta). \quad (5.6)$$

The effect of the scale factors can be found in Figure 5.8 for the electron scaling and in Figure 5.7 for the muons. The trigger efficiencies are estimated in Section 5.1.2.

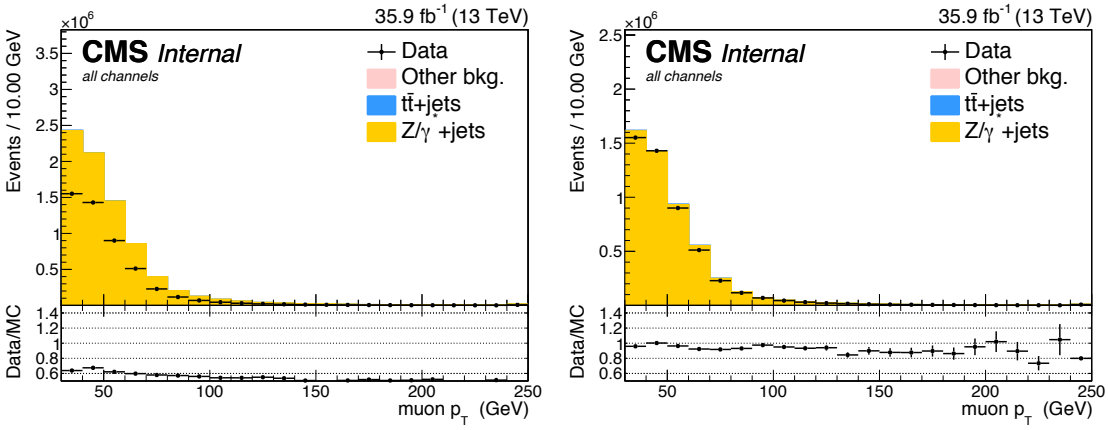

Figure 5.7: Distribution of the p_T of the muons before (left) and after (right) muon scale factors. After requiring dilepton plus jets selection criteria, in the Z boson mass window. Both after the Rochester correction.

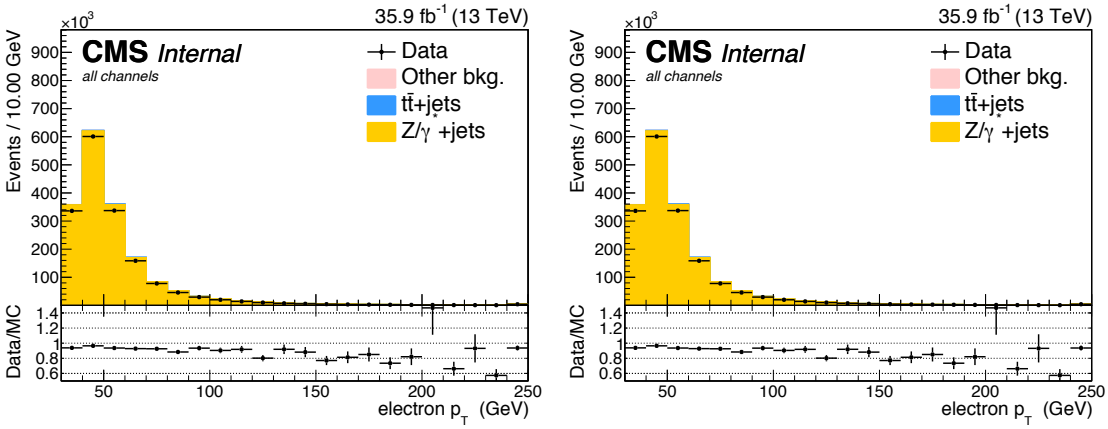

Figure 5.8: Distribution of the p_T of the electrons before (left) and after (right) electron scale factors. After requiring dilepton plus jets selection criteria, in the Z boson mass window. Both after energy scale corrections and smearing.

Additionally, corrections are determined from $Z \to ee$ events for the energy resolution of the leptons. For the electrons, energy smearing and regression is applied [193]. The energy regression uses the detector information to correct the electron energy in order to have the best energy resolution and corrects for material effects in the ECAL, improving the performance. The energy scale and smearing corrects the simulation energies to have identical energy resolution

in simulation and data. For the muons, the p_T is corrected using the Rochester method [194, 195]. This correction is determined from $Z \rightarrow \mu\mu$ events and removes the bias of the muon p_T from any detector misalignment or any possible error of the magnetic field. The effect of the Rochester correction can be found in Figure 5.9.

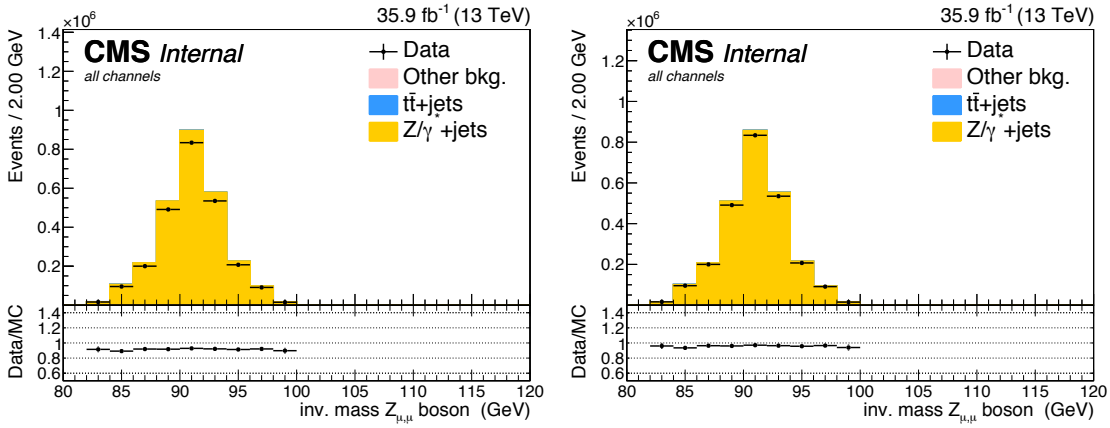

Figure 5.9: Distribution of the mass of the Z boson from muons before (left) and after (right) the Rochester correction. After requiring dilepton plus jets selection criteria, in the Z boson mass window.

CSVv2 shape correction

In order to make the distribution of the CSVv2 b-tagging discriminant in simulation agree with data, jet-by-jet based scale factors are applied. These scale factors are a function of the p_T , η and CSVv2 value of the jet as discussed in Section 4.4.4. The effect of these scale factors on the distribution of the CSVv2 discriminant of all jets can be found in Figure 5.10.

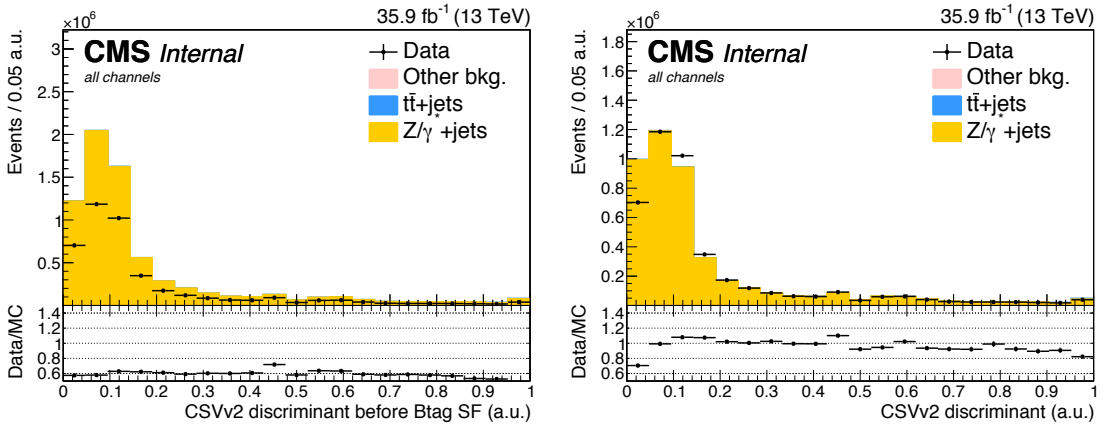

Figure 5.10: Distribution of the CSVv2 discriminant of the jets before (left) and after (right) b-tag scale factors. After requiring dilepton plus jets selection criteria, in the Z boson mass window.

Jet energy

The jet energy in data and simulation is corrected by the measured energy response of the detector. This provides p_T- and η dependent scale factors and are directly taken from the frontier condition database as discussed in Section 4.4.3. The effect of the jet energy corrections on the distribution of the p_{T} of all jets can be found in Figure 5.11 and Figure 5.12.

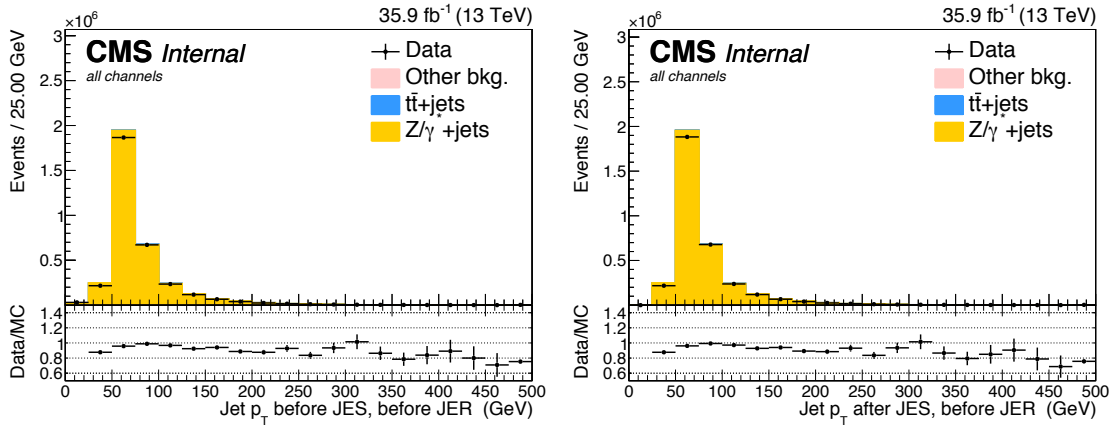

Figure 5.11: Distribution of the p_{T} of the jets before (left) and after (right) jet energy scale corrections.After requiring dilepton plus jets selection criteria, in the Z boson mass window.

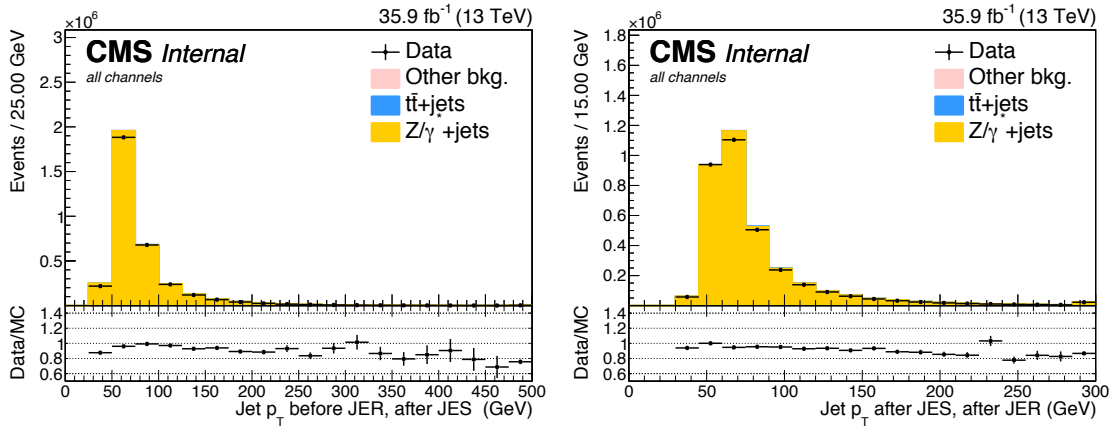

Figure 5.12: Distribution of the p_{T} of the jets before (left) and after (right) jet energy resolution smearing.After requiring dilepton plus jets selection criteria, in the Z boson mass window.

5.3 Event reconstruction

After selecting the events, the objects are assigned to physical particles. The Z boson is reconstructed as the sum of the four vectors of the two same flavour leptons of opposite sign giving the closest value to the Z boson mass. The third remaining lepton is assigned as the lepton coming from the W boson decay. This lepton assignment is validated using simulation by matching the reconstructed objects to their generated counterpart by minimizing ΔR between

the true particle and reconstructed one. The efficiencies derived from the simulated signal samples and the SM tZq background process can be found in Table 5.6 for events selected after the three-leptons and jets requirements. The probability that a lepton is assigned to the wrong boson is of the order of a % for the top quark pair FCNC signal process, and of the order of 2% for the single top quark FCNC signal process. For the SM tZq process, this probability is 3%.

Table 5.6: Efficiencies of assigning the correct leptons in the analysis after requiring three leptons and jets.

Origin	FCNC tZq	FCNC tZ	SM tZq
W boson	99%	98%	97%
Z boson	99%	98%	97%
all leptons in the decay	99%	98%	97%

The SM b jet is assigned to the jet with the highest CSVv2 discriminant. This jet is then removed from the collection of jets. A loop over the jets is performed and the jet that in combination with the reconstructed Z boson, gives the mass closest to the top mass is assigned as the light flavour jet coming from the FCNC decay of the top quark. The SM top quark candidate is reconstructed by summing the third lepton, the SM b jet and the neutrino ($E_{\mathrm{T}}^{\mathrm{miss}}$). The longitudinal momentum of the neutrino is calculated from the missing transverse momentum and the lepton momentum. From momentum conservation follows that

$$\vec{p}_{\mathrm{W}} = \vec{p}_{\ell_{\mathrm{W}}} + \vec{p}_{\nu}, \tag{5.7}$$

and energy conservation requires the W boson mass squared,

$$m_{\mathrm{W}}^2 = \begin{pmatrix} E_{\mathrm{W}} \\ \vec{p}_{\mathrm{W}} \end{pmatrix} = (80.4\ \mathrm{GeV})^2, \tag{5.8}$$

is equal to the sum of the neutrino ν and the lepton ℓ_{W} squared

$$\begin{aligned} m_{\mathrm{W}}^2 &\equiv \left(\begin{pmatrix} E_{\ell_{\mathrm{W}}} \\ \vec{p}_{\ell_{\mathrm{W}}} \end{pmatrix}^2 + \begin{pmatrix} E_{\nu} \\ \vec{p}_{\nu} \end{pmatrix} \right)^2, \\ &= m_{\ell_{\mathrm{W}}}^2 + m_{\nu}^2 + 2.E_{\ell_{\mathrm{W}}} E_{\nu} - 2 \begin{pmatrix} p_{\ell_{\mathrm{W}}}^x \\ p_{\ell_{\mathrm{W}}}^y \\ p_{\ell_{\mathrm{W}}}^z \end{pmatrix} \begin{pmatrix} p_{\mathrm{T},\nu}\ \cos\phi_{\nu} \\ p_{\mathrm{T},\nu}\ \sin\phi_{\nu} \\ p_{\nu}^z \end{pmatrix}. \end{aligned} \tag{5.9}$$

Assuming that the lepton and neutrino are approximately massless, this equation can be solved for the sought-for p_{ν}^z by setting $p_{\mathrm{T},\nu}$ equal to the missing transverse energy:

$$0 = p_{\nu}^z - \frac{2\,\beta\, p_{\ell_{\mathrm{W}}}^z}{E_{\ell_{\mathrm{W}}}^2 - \left(p_{\ell_{\mathrm{W}}}^z\right)^2} p_{\nu}^z - \frac{\beta^2 - E_{\ell_{\mathrm{W}}}^2 p_{\mathrm{T},\nu}^2}{E_{\ell_{\mathrm{W}}}^2 - \left(p_{\ell_{\mathrm{W}}}^z\right)^2}, \tag{5.10}$$

with

$$\beta = \frac{1}{2}m_{\mathrm{W}}^2 + p_{\mathrm{T},\nu}p_{\mathrm{T},\ell_{\mathrm{W}}}\cos\left(\phi_{\ell_{\mathrm{W}}} - \phi_{\nu}\right). \tag{5.11}$$

When the solution of this quadratic solution is complex, only the real part is considered. When there are two real solutions, the solution for p_{ν}^{z} that gives the invariant mass $m_{\mathrm{b}\nu\ell_{\mathrm{W}}}$ closest to the top quark mass (172.9 GeV) is kept. The normalised distributions of the reconstructed top quark mass are shown in Figure 5.13. For the FCNC signal, one can clearly see a peak around the known mass of the top quark. For the $\mathrm{Z}/\gamma^{*}+$jets+jets and $\mathrm{t\bar{t}}$+jets backgrounds, there is not enough events surviving the three lepton requirement to make sensible distributions.

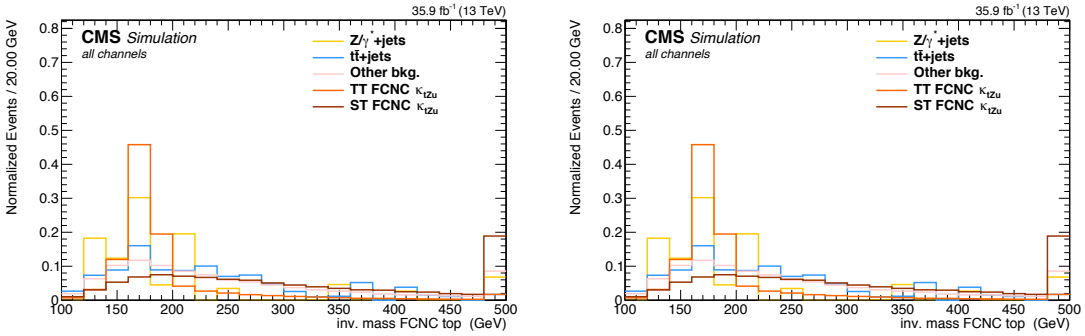

Figure 5.13: Normalised distribution of the top quark mass decaying via the SM Wtb vertex (left) and the FCNC tZq vertex (right). After requiring three leptons and jets, in the Z boson mass window.

5.4 Data driven NPL background process

One of the most important backgrounds consists of events with not prompt-leptons (NPL background process). Their origin lies mostly in instrumental backgrounds and are therefore very difficult to model. The NPL background process is estimated from data for both its shape and its normalisation.

The NPL background process originates from hadronic objects wrongly reconstructed as leptons (so-called fake leptons), or from real leptons coming from the semi-leptonic decay of a b or c hadron and from the conversion of photons that pass the identification and isolation requirements (so-called non-prompt leptons). The dominant source of events contained in this NPL background process depends on the flavour of the leptons and therefore the events with a not prompt-muon (NPμ) are treated differently than those with a not prompt-electron (NPe). For muons, the dominant source is the semi-leptonic decay of heavy flavour hadrons, while for electrons, the dominant sources are hadrons and photon conversions.

The backgrounds causing events that are contained within the NPL background process are mostly $\mathrm{Z}/\gamma^{*}+$jets (Drell–Yan) and $\mathrm{t\bar{t}}$+jets dilepton processes, and in a smaller amount WW processes. All of these processes contain two real leptons and one NPL. Due to the fact that

the probability for a lepton to be a NPL is small, backgrounds containing two or more not prompt-leptons are neglected in this search. The assumption is made that for $Z/\gamma^* +$jets the two leptons compatible with a Z boson decay are the real leptons, and the additional lepton is coming from a NPL source, while for $t\bar{t}+$jets the NPL is associated to the Z boson. This assumption has been validated using Monte Carlo simulations of the $Z/\gamma^* +$jets process and $t\bar{t}$ process by matching the reconstructed leptons to their true initial simulated particles, after requiring exactly three leptons and jets, in the Z boson mass window. For the $Z/\gamma^* +$jets process this assumption is true in 80% of the events, increasing to 100% of the events after requiring one b-tagged jet. For the $t\bar{t}$ process this is true for 60% of the selected events, and this increases to 90% after requiring one b-tagged jet.

Since simulation is not able to reproduce the events causing not prompt-leptons, the shape of the distributions for the NPL background sample is constructed from data. Events are selected from data by requiring exactly two leptons that are identified as real isolated leptons according to the tight working point given in Table 4.1 and Table 4.3. A third lepton, the so-called not prompt-lepton, is added by taking a lepton for which the identification criteria are loosened and the isolation citeria are inverted. The full requirements on the not prompt-leptons are given in Table 5.7 and Table 5.8. For not prompt-electrons, a large fraction is coming from misidentified photons. These are removed by applying a tighter cut on the $1/E - 1/p$ variable, and by limiting the isolation values to be smaller than one, in coherence with the SM tZq search from CMS [196].

The normalisation of the distributions from the NPL background sample is estimated from data through the use of control regions that are fitted simultaneously with the signal regions. These regions are defined in Section 5.5.

Table 5.7: Not prompt-electron requirements used in this analysis. The requirements for electrons are set in the barrel ($|\eta_{supercluster}| \leq 1.479$) and the end caps ($|\eta_{supercluster}| > 1.479$).

Properties	$\lvert\eta_{supercluster}\rvert \leq 1.479$	$\lvert\eta_{supercluster}\rvert > 1.479$
$\sigma_{\eta\eta}$	< 0.011	< 0.0314
$\lvert\Delta\eta_{in}\rvert$	< 0.00477	< 0.00868
$\lvert\Delta\phi_{in}\rvert$	< 0.222	< 0.212
H/E	< 0.298	< 0.101
relative isolation	$[0.0588, 1]$	$[0.0571, 1]$
$\lvert 1/E - 1/p\rvert$ (GeV^{-1})	< 0.0129	< 0.0129
expected missing inner hits	≤ 1	≤ 1
conversion veto	Y	Y
p_T (GeV)	> 35	> 35

Table 5.8: Not prompt-muon requirements used in the analysis.

Properties	modified Loose Muon WP
Global muon or Tracker Muon	Both
Particle Flow muon	Y
χ^2/ndof of global muon track fit	N/A
Nb. of hit muon chambers	N/A
Nb. of muon stations contained in the segment	N/A
Size of the transverse impact parameter of the track wrt. PV	N/A
Longitudinal distance wrt. PV	N/A
Nb. of pixel hits	N/A
Nb. of tracker layers with hits	N/A
Relative Isolation	$[0.15, 1]$
p_{T} (GeV)	> 30

5.5 Analysis Strategy

The baseline selection of this analysis selects events where jets and three leptons are present. Additional leptons with a looser identification are vetoed in order to reduce the contamination of backgrounds with four or more leptons in the final state, e.g. ZZ, $\mathrm{t\bar{t}Z}$, and $\mathrm{t\bar{t}H}$. This makes that most important backgrounds in this search consist of backgrounds that contain three prompt leptons in the final state. These are mainly WZ+jets, $\mathrm{t\bar{t}Z}$ and SM tZq. For these backgrounds, the three lepton topology is identical to the FCNC signal: two opposite sign leptons of the same flavour decaying from the Z boson, and a third additional, high p_{T} lepton coming from the W boson decay.

For the single top quark FCNC final state, one b jet coming from the SM top quark decay is expected. For the top quark pair FCNC signal, an additional light jet is expected. In the $\mathrm{t\bar{t}Z}$ final state, two b jets are present in the final state. However, due to inefficiencies of the b-tagging algorithm, one of the two b jets may be identified as a light quark jet, giving the same final state as the top quark pair FCNC final state. For the WZ+jets final states, one of the b jets produced by gluon splitting, can be b-tagged or light flavour jets coming from the WZ+jets production can be mis-tagged as b jets. The SM tZq final state expects the same signal as the top quark pair FCNC process. Furthermore, the NPL background is responsible for a significant amount of background events.

The analysis strategy is shown in Figure 5.14. Based on jet multiplicity, b-tag information and the Z boson mass, five statistically independent regions are defined after a common selection of exactly three leptons containing one opposite sign same flavour pair that is assigned to the Z boson, at least one jet and at the most three jets, and the transverse mass of the W boson to be maximal 300 GeV. The requirements for each region is shown in Table 5.9. Two signal regions are considered, targeting the final state of each of the signals. The STSR is constructed to target the final state of the single top quark FCNC signal, while the TTSR is constructed to target the top quark pair FCNC signal. In each signal region, a multivariate discriminant based on Boosted Decision Trees (BDT) (see Section 3.3) is used to respectively discriminate single top quark FCNC and top quark pair FCNC signal from backgrounds. The rate of WZ+jet events as well

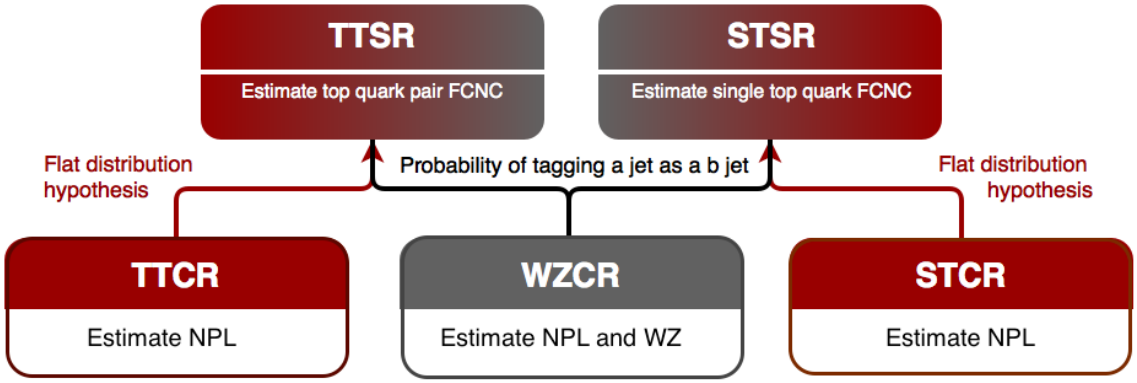

Figure 5.14: The strategy used for the search presented in this thesis. The WZCR region is used to estimate the WZ+jets background process as well as the NPL background coming from the Z/γ^*+jets process. The TTCR and STCR regions are used to estimate the contributions of the NPL background coming from the $t\bar{t}$+jets process.

as that of NPL background process, mainly originating mainly from the Z/γ^*+jets process, is estimated in the WZCR. Here, the transverse mass of the W boson, defined as

$$m_T(W) = \sqrt{\left(p_T(l_W) + p_T(\nu_W)\right)^2 - \left(p_x(l_W) + p_x(\nu_W)\right)^2 - \left(p_y(l_W) + p_y(\nu_W)\right)^2}, \quad (5.12)$$

is used as discriminating variable between the two backgrounds, WZ+jets and NPLs. The normalisation of the backgrounds are then used in the signal regions via the b-tagging information (Section 5.5.1). The NPL background coming from a $t\bar{t}$ process are constrained by two control regions, TTCR and STCR, one for each signal region (TTSR and STSR). The normalisation of the $t\bar{t}$ process, is estimated by subtracting all other background predictions from data (Section 5.5.3). A simultaneous global fit using the `Higgs Combined Tool` (Section 3.4) is performed taking into account each region (STSR, TTSR, WZCR, TTCR and STCR) for the four different lepton channels. The BDTs in the signal regions, as well as the transverse mass of the W boson are discussed in Chapter 6. The number of events for each region are shown in Section 5.5.4.

Table 5.9: The statistically independent regions used in the analysis.

	WZCR	STSR	TTSR	STCR	TTCR
Number of jets	$\geqslant 1$	1	$\geqslant 2$	1	$\geqslant 2$
Number of b jets	0	1	$\geqslant 1$	1	$\geqslant 1$
$\lvert m_Z^{reco} - m_Z \rvert < 7.5$ GeV	Yes	Yes	Yes	No	No
$\lvert m_Z^{reco} - m_Z \rvert < 30$ GeV	Yes	Yes	Yes	Yes	Yes
Number of leptons	3	3	3	3	3

5.5.1 WZCR

The WZCR is constructed by vetoing events with jets tagged as being a b jet, making it statistically independent from the signal regions where at least one b-tagged jet is required. In this region,

a fit is performed on the transverse mass of the W boson, in order to estimate the NPL yield coming from $Z/\gamma^* +$ jets and the WZ+jets backgrounds.

A transfer factor is used to account for going from a region without b-tagged jets to a region with exactly, or at least, one b-tagged jet. For this, the probability of tagging at least one jet with the CSVv2 algorithm at the loose working point is used to calculate the expected number of events, N_b, after b-tagging:

$$N_{\mathrm{b}} = \frac{\sum_{\text{events}} \mathscr{P}_b}{\text{total nb of events}}, \tag{5.13}$$

where $\mathscr{P}_{\mathrm{b}}$ is the probability that an event survives the b-tagging requirement,

$$\begin{aligned} \mathscr{P}_{\mathrm{b}} &= 1 - \mathrm{P}(\text{event doesn't survive b tag}), \\ &= 1 - \left(\prod_{\mathrm{b}} \mathrm{P}(\text{b not b-tagged}) \prod_{\mathrm{c}} \mathrm{P}(\text{c not b-tagged}) \prod_{\mathrm{udsg}} \mathrm{P}(\text{light not b-tagged}) \right), \end{aligned} \tag{5.14}$$

with the products going over all b-, c-, and light jets. The jet flavour is determined by means of matching the reconstructed jet to the generated quarks, based on the distance in the $\eta\phi$ plane. In order to estimate the probability for exactly one b-tagged jet, the expected number of events is corrected by the fraction of events with exactly one jet in the WZCR. The resulting transfer factors are given in Appendix B. The yield of WZ+jets events in the signal region estimated using the above described transfer factor, and the yield calculated with simulated events, are in agreement.

5.5.2 TTSR and STSR

The TTSR is defined to target top quark pair FCNC (tZq), while the STSR focusses on single top quark FCNC (tZ). They have NPL contributions coming from $Z/\gamma^* +$ jets and $\mathrm{t\bar{t}}$+jets events. In this region, the data driven NPL template is split into two templates, based on the presence of the NPL in the Z boson. The NPL associated with W boson is assigned to $Z/\gamma^* +$ jets and estimated in the WZCR, while the NPL associated with Z boson is assigned to $\mathrm{t\bar{t}}$+jets and estimated in the TTCR and STCR.

5.5.3 TTCR and STCR

The TTCR and STCR are constructed with the same selection criteria as TTSR and STSR, but are outside the Z boson mass window (sidebands):

$$7.5\,\mathrm{GeV} < |m_{\mathrm{Z}}^{\mathrm{reco}} - m_{\mathrm{Z}}| < 30\,\mathrm{GeV}, \tag{5.15}$$

where $m_{\mathrm{Z}}^{\mathrm{reco}}$ is the reconstructed mass of the Z boson in the event, and m_{Z} the known mass of the Z boson. These regions are dominated by $\mathrm{t\bar{t}}$+jets (see Appendix B) and are used to estimate the NPL background event yield coming from $\mathrm{t\bar{t}}$+jets in the STSR and TTSR. Since there are few events entering the STCR and TTCR, no shapes are used in the fit. The distribution of the mass of the Z boson is flat for $\mathrm{t\bar{t}}$+jets events, as shown in Figure 5.15, and thus the number of

expected events, N_s, in the signal regions estimated from the number of expected events, N_c, in the control region is obtained as

$$N_s = \frac{15}{60-15} N_c. \tag{5.16}$$

The resulting transfer factors are given in Appendix B. The expected yield in the signal region estimated from the TTCR (STCR) is in agreement with the yield calculated from simulated events.

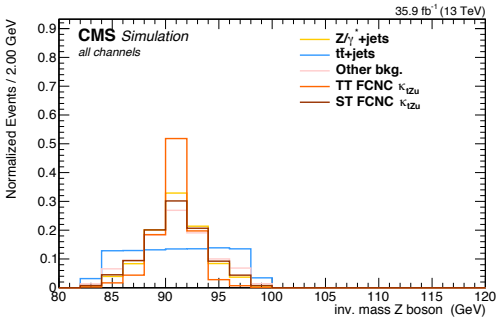

Figure 5.15: The normalized distribution for $Z/\gamma^* +$ jets and $t\bar{t}+$jets events before dividing the events into regions, after $|m_Z^{reco} - m_Z| < 7.5$ GeV. All lepton channels combined.

5.5.4 Event yields

(to be filled in once combine is hacked for this)

Table 5.10: Event yields in the STCR.

Process	all channels	3μ channel	1e2μ channel	2e1μ channel	3e channel
NPL $t\bar{t}$-like	33.00 ± 8.48	3μ channel	1e2μ channel	2e1μ channel	3e channel
$t\bar{t}Z$	0.18 ± 0.21	3μ channel	1e2μ channel	2e1μ channel	3e channel
WZ	3.52 ± 0.25	3μ channel	1e2μ channel	2e1μ channel	3e channel
ZZ	0.31 ± 0.10	3μ channel	1e2μ channel	2e1μ channel	3e channel
Other bkg.	1.66 ± 0.91	3μ channel	1e2μ channel	2e1μ channel	3e channel
tZq	0.31 ± 0.06	3μ channel	1e2μ channel	2e1μ channel	3e channel
κ_{tZu}/Λ	0.42 ± 0.03	3μ channel	1e2μ channel	2e1μ channel	3e channel
κ_{tZc}/Λ	0.32 ± 0.03	3μ channel	1e2μ channel	2e1μ channel	3e channel
Data	32.00 ± 3.28	3μ channel	1e2μ channel	2e1μ channel	3e channel
Total bkg.	38.98 ± 8.67	3μ channel	1e2μ channel	2e1μ channel	3e channel

Table 5.11: Event yields in the STSR

Process	all channels	3μ channel	1e2μ channel	2e1μ channel	3e channel
NPL $Z/\gamma^* + jets$-like	6.12 ± 1.41	3μ channel	1e2μ channel	2e1μ channel	3e channel
$t\bar{t}Z$	3.51 ± 0.34	3μ channel	1e2μ channel	2e1μ channel	3e channel
WZ	6.10 ± 0.66	3μ channel	1e2μ channel	2e1μ channel	3e channel
ZZ	4.60 ± 0.53	3μ channel	1e2μ channel	2e1μ channel	3e channel
Other bkg.	1.25 ± 0.25	3μ channel	1e2μ channel	2e1μ channel	3e channel
tZq	8.03 ± 0.47	3μ channel	1e2μ channel	2e1μ channel	3e channel
NPL $t\bar{t}$-like	9.33 ± 1.55	3μ channel	1e2μ channel	2e1μ channel	3e channel
κ_{tZu}/Λ	11.25 ± 0.17	3μ channel	1e2μ channel	2e1μ channel	3e channel
κ_{tZc}/Λ	18.52 ± 0.30	3μ channel	1e2μ channel	2e1μ channel	3e channel
Data	138.00 ± 14.97	3μ channel	1e2μ channel	2e1μ channel	3e channel
Total bkg.	38.94 ± 2.70	3μ channel	1e2μ channel	2e1μ channel	3e channel

Table 5.12: Event yields in the TTCR.

Process	all channels	3μ channel	1e2μ channel	2e1μ channel	3e channel
NPL $t\bar{t}$-like	29.00 ± 9.62	3μ channel	1e2μ channel	2e1μ channel	3e channel
$t\bar{t}Z$	2.85 ± 0.44	3μ channel	1e2μ channel	2e1μ channel	3e channel
WZ	3.98 ± 0.63	3μ channel	1e2μ channel	2e1μ channel	3e channel
ZZ	0.32 ± 0.08	3μ channel	1e2μ channel	2e1μ channel	3e channel
Other bkg.	3.88 ± 0.66	3μ channel	1e2μ channel	2e1μ channel	3e channel
tZq	0.79 ± 0.13	3μ channel	1e2μ channel	2e1μ channel	3e channel
κ_{tZu}/Λ	0.61 ± 0.05	3μ channel	1e2μ channel	2e1μ channel	3e channel
κ_{tZc}/Λ	0.59 ± 0.05	3μ channel	1e2μ channel	2e1μ channel	3e channel
Data	43.50 ± 3.28	3μ channel	1e2μ channel	2e1μ channel	3e channel
Total bkg..	40.81 ± 9.87	3μ channel	1e2μ channel	2e1μ channel	3e channel

Table 5.13: Event yields in the TTSR.

Process	all channels	3μ channel	1e2μ channel	2e1μ channel	3e channel
NPL $Z/\gamma^* +$ jets-like	9.36 ± 1.62	3μ channel	1e2μ channel	2e1μ channel	3e channel
$t\bar{t}Z$	42.55 ± 2.28	3μ channel	1e2μ channel	2e1μ channel	3e channel
WZ	12.46 ± 1.12	3μ channel	1e2μ channel	2e1μ channel	3e channel
ZZ	4.84 ± 0.35	3μ channel	1e2μ channel	2e1μ channel	3e channel
Other bkg.	5.62 ± 0.82	3μ channel	1e2μ channel	2e1μ channel	3e channel
tZq	16.93 ± 0.89	3μ channel	1e2μ channel	2e1μ channel	3e channel
NPL $t\bar{t}$-like	8.00 ± 1.07	3μ channel	1e2μ channel	2e1μ channel	3e channel
κ_{tZu}/Λ	21.76 ± 0.25	3μ channel	1e2μ channel	2e1μ channel	3e channel
κ_{tZc}/Λ	56.85 ± 0.57	3μ channel	1e2μ channel	2e1μ channel	3e channel
Data	243.00 ± 19.21	3μ channel	1e2μ channel	2e1μ channel	3e channel
Total bkg..	99.76 ± 4.05	3μ channel	1e2μ channel	2e1μ channel	3e channel

Table 5.14: Event yields in the WZCR.

Process	all channels	3μ channel	1e2μ channel	2e1μ channel	3e channel
NPL $Z/\gamma^* +$ jets-like	204.00 ± 22.12	3μ channel	1e2μ channel	2e1μ channel	3e channel
$t\bar{t}Z$	9.57 ± 0.69	3μ channel	1e2μ channel	2e1μ channel	3e channel
WZ	551.63 ± 29.26	3μ channel	1e2μ channel	2e1μ channel	3e channel
ZZ	46.19 ± 2.40	3μ channel	1e2μ channel	2e1μ channel	3e channel
Other bkg.	6.75 ± 0.99	3μ channel	1e2μ channel	2e1μ channel	3e channel
tZq	7.30 ± 0.45	3μ channel	1e2μ channel	2e1μ channel	3e channel
κ_{tZu}/Λ	14.12 ± 0.23	3μ channel	1e2μ channel	2e1μ channel	3e channel
κ_{tZc}/Λ	26.34 ± 0.51	3μ channel	1e2μ channel	2e1μ channel	3e channel
Data	1053.00 ± 34.45	3μ channel	1e2μ channel	2e1μ channel	3e channel
Total bkg.	825.44 ± 37.24	3μ channel	1e2μ channel	2e1μ channel	3e channel