

Observation of single top quark production in association with a Z boson

Willem Verbeke*, on behalf of the CMS Collaboration

Ghent University

E-mail: willem.verbeke@cern.ch

Single top quark production in association with a Z boson (tZq) is observed. Proton-proton collision events containing three leptons (electrons or muons) are analyzed. The data correspond to an integrated luminosity of 77.4 fb^{-1} , collected with the CMS detector throughout 2016 and 2017. The analysis is significantly more sensitive than any previous search for tZq because of the usage of machine learning for lepton identification, and a redesigned search strategy. The observed (expected) signal significance of tZq is 8.2 (7.7) standard deviations above the background only hypothesis. The tZq cross section is measured to be $\sigma(\text{pp} \rightarrow \text{tZq} \rightarrow \text{t}\ell\ell\text{q}) = 111 \pm 13 \text{ (stat)} \text{ }^{+11}_{-9} \text{ (syst) fb}$, with an uncertainty more than twice smaller than that on any previous measurement.

*European Physical Society Conference on High Energy Physics - EPS-HEP2019 -
10-17 July, 2019
Ghent, Belgium*

*Speaker.

This document is based on the results presented in [1]. A more detailed description of the analysis can be found there, in particular information about the simulations that were used, the estimation of systematic uncertainties and the event reconstruction.

1. Introduction

Over the last years the CERN LHC has delivered an unprecedented volume of proton-proton (pp) collision data. This large dataset, collected at a center of mass energy of 13 TeV, makes it possible to probe increasingly rare standard model (SM) processes. An extremely rare process is the production of a single top quark in association with a Z boson (tZq). The tZq production cross section is about 50 times smaller than that of the Higgs boson. In addition tZq is a process uniquely sensitive to the presence of new physics because it involves a myriad of SM interactions, namely the ttZ, WWZ and tbW couplings, and the bW \rightarrow tZ scattering amplitude [2, 3].

We present a search for tZq targeting the leptonic decays of both the top quark and the Z boson, in events with three leptons (electrons or muons). Previously both ATLAS and CMS carried out searches for tZq in this channel using about 36 fb^{-1} of 13 TeV pp collision data [4, 5]. They achieved observed signal significances of respectively 4.2 and 3.7 standard deviations over the background only hypothesis. In both cases the cross section was measured with a relative uncertainty of approximately 35%. We present a new search for tZq based on 77.4 fb^{-1} of pp collision data collected by CMS. By more than doubling the integrated luminosity, and especially through the usage of machine learning for the identification of leptons and a redesigned analysis strategy we significantly increase the reach of this search compared to all previous searches. [1]

2. Data analysis

The full details on the event reconstruction are given in [1]. The analysis uses events with three leptons and at least two jets, one of which is tagged as originating from the hadronization of a b quark. At least one opposite sign and same flavor pair of leptons is required to be present and have a mass within a 30 GeV window around the Z boson mass.

A crucial element of this search is the distinction of leptons originating from the decay of electroweak bosons from genuine leptons coming from hadron decays and jet constituents that are misidentified as leptons. The former class of leptons is called prompt, while the latter are called nonprompt. The limiting factor in the previous CMS search for tZq was a large background from nonprompt leptons, with a large associated uncertainty on its prediction. In order to suppress this background, forests of gradient boosted decision trees (BDT) are employed. The BDTs capitalize on the properties of the jet closest to the lepton in terms of ΔR , the isolation of the lepton from other particles, information on the impact parameter of the lepton with respect to the primary vertex, the p_T and $|\eta|$ of the lepton. Additional inputs to the BDTs are information on the ECAL shower shape of electrons and the quality of the matching between tracks in the muon system and inner tracker for muons.

The efficiency of selecting prompt leptons as a function of the misidentification probability for nonprompt leptons in simulated events is shown in Fig. 1 for varying thresholds on the BDTs, and for the cut-based identification used in the previous tZq measurement by CMS. It can be seen that

the chosen electron (muon) identification threshold corresponds to a 12 (8)% increase in the signal efficiency, while reducing the background by a factor 2 (8).

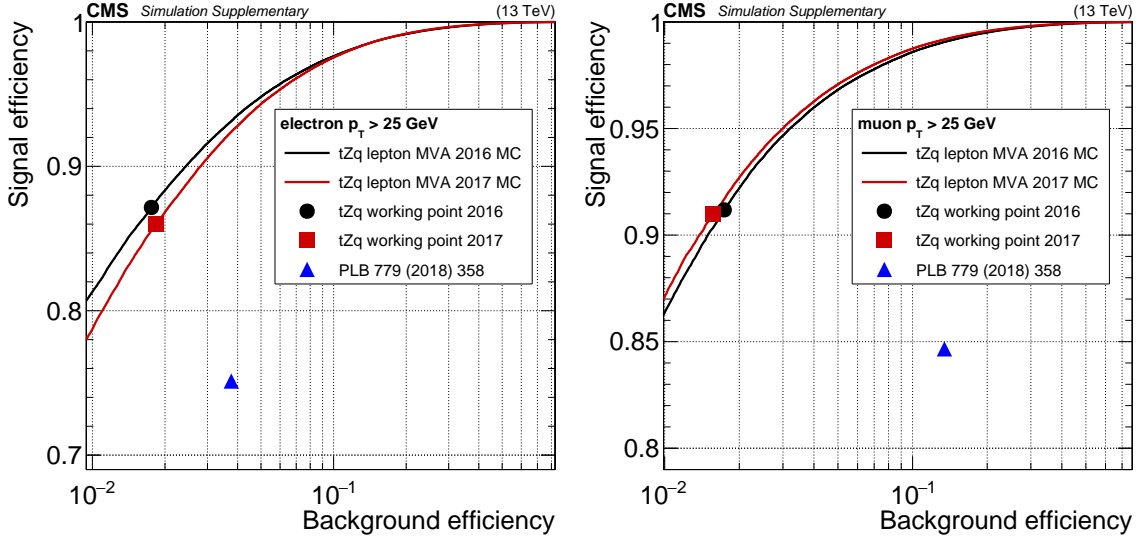


Figure 1: Efficiency of selecting prompt electrons (left), and muons (right) from simulated tZq and $t\bar{Z}$ events as a function of the misidentification probability for nonprompt leptons from simulated $t\bar{t}$ events for the lepton identification BDTs (black and red lines for respectively 2016 and 2017 simulations), and for the cut-based lepton identification used in the previous CMS search for tZq (blue triangle). The black and red points correspond to the cut on the lepton discriminator chosen for this analysis. [6]

Events containing three leptons passing a set of loose criteria, and a tight threshold on the BDT are selected for analysis. These events are subsequently categorized according to the number of jets and b-tagged jets. Events with two or three jets, exactly one of which is b-tagged (SR-2/3j-1b) contain the majority of the tZq signal and form the most important event category. The other two categories consist of events with four or more jets, exactly one of which is b-tagged (SR-4j-1b), and events with two or more b-tagged jets (SR-2b). In each of these event categories a BDT is trained to separate signal from background events. These BDTs exploit a plethora of kinematic differences between tZq and the various backgrounds. One of the most distinct features of tZq is the presence of a forward jet in most of the events, leading to high jet $|\eta|$ values and events with high dijet invariant masses. The complete list of input variables can be found in [1].

Background processes in the three lepton final state can be subdivided into those that contain three prompt leptons and those that contain at least one nonprompt lepton. The contributions from the former category are estimated using simulation. The largest background contribution in SR-2/3j-1b comes from WZ production. Its normalization is measured, and its simulation validated in a dedicated control region containing events with three leptons without b-tagged jets. The smaller background contribution from ZZ production is similarly measured and validated in a control region of four lepton events. In SR-4j-1b and SR-2b the largest background contribution is $t\bar{t}Z$ production. The normalization of this process is implicitly constrained in the final fit since it is the dominant contribution in low BDT bins. Other backgrounds with three prompt leptons include $t\bar{t}W$, tWZ and $t\bar{t}H$ production, asymmetric photon conversions, production of multiple electroweak bosons, and other rare processes. All of these are estimated from simulation.

Background events with nonprompt leptons mainly come from the $t\bar{t}$ and Drell-Yan processes. The contribution of this class of backgrounds is estimated directly from data by measuring the probability that a lepton that passed a loosened identification requirement passes the full lepton selection criteria. The background can then be predicted using a sideband of events that fail the nominal lepton selection while passing the loosened lepton selection.

3. Results

The tZq production cross section is measured with a simultaneous binned maximum likelihood fit to the BDT distributions in each of the three signal regions and the WZ and ZZ control regions. Several sources of systematic uncertainty affect the measurement and are taken into account as nuisance parameters in the fit. A detailed list is given in [1]. The most important uncertainty sources are the estimation of the nonprompt lepton background, the measurement of lepton selection efficiencies, the modeling of final-state radiation and the jet energy scale.

From the fit, the tZq cross section is found to be

$$\sigma(pp \rightarrow tZq \rightarrow t\ell^+\ell^-q) = 111 \pm 13 \text{ (stat)} \pm 9 \text{ (syst)} \text{ fb}$$

where ℓ means a muon electron or τ lepton, for dilepton invariant masses above 30 GeV. This corresponds to a signal strength of :

$$\mu = \frac{\sigma(pp \rightarrow tZq \rightarrow t\ell^+\ell^-q)}{\sigma^{\text{SM}}(pp \rightarrow tZq \rightarrow t\ell^+\ell^-q)} = 1.18 \pm 0.14 \text{ (stat)} \pm 0.11 \text{ (syst)} \pm 0.04 \text{ (theo)},$$

consistent with the SM expectation. The observed (expected) signal significance is 8.2 (7.7) standard deviations over the background only hypothesis, marking the first observation of the tZq process.

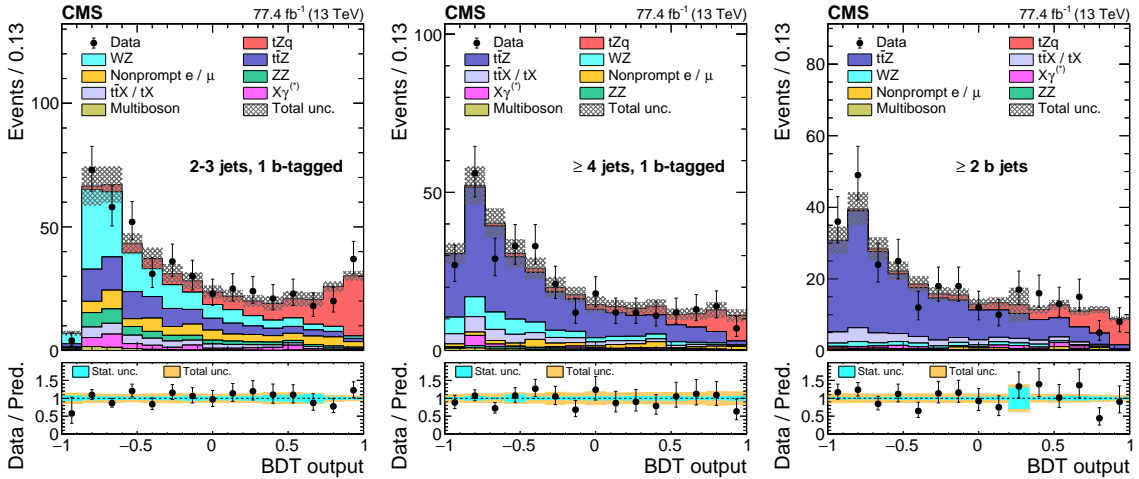


Figure 2: Observed and post-fit expected BDT distributions for events in SR-2/3j-1b (left), SR-4j-1b (middle), and SR-2b (right). [1]

The observed and expected BDT distributions are shown in Fig. 2 for the three signal regions. Two of the most important discriminating variables used in the BDT are the $|\eta|$ of the highest p_T

jet that is not b-tagged (recoiling jet) and the maximum invariant mass made up of any pair of jets in the event. The observed and expected distributions of these variables are shown in Fig. 3. This figure additionally shows the p_T distribution of the Z boson, reconstructed from a pair of leptons, in SR-2/3j-1b events with high BDT scores. The distribution of this variable is expected to be modified for tZq events in the presence of new physics [3].

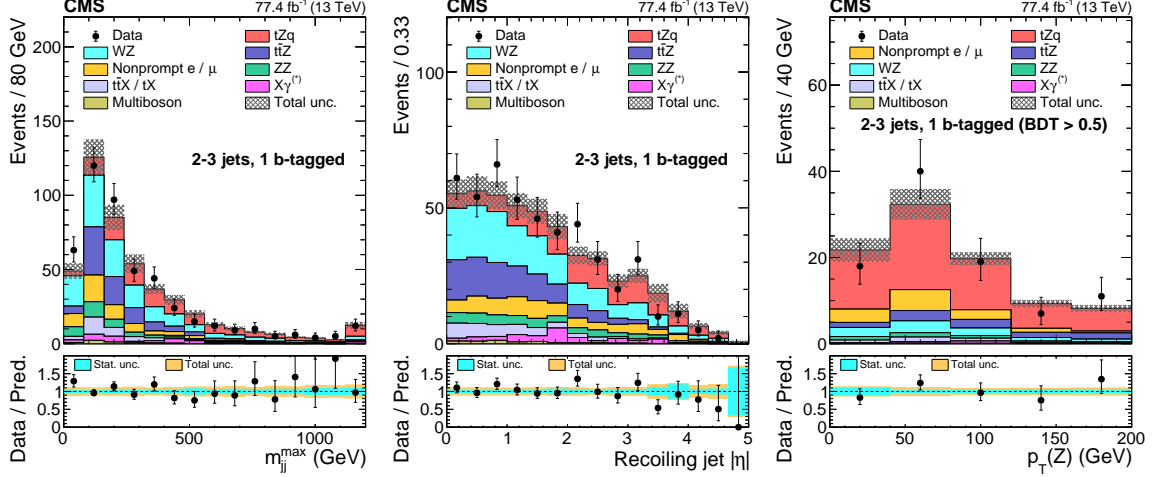


Figure 3: Observed and post-fit expected distributions in SR- 2/3j-1b events for the two most important variables used in the BDT discriminant, the maximum dijet invariant mass among all pairs of jets in the event (left), and the $|\eta|$ of the recoiling jet (middle). The rightmost plot shows the p_T distribution of the Z boson, reconstructed from two leptons for events in SR-2/3j-1b with a BDT score above 0.5. [1]

References

- [1] CMS Collaboration. Observation of Single Top Quark Production in Association with a Z Boson in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 122(13):132003, 2019.
- [2] John Campbell, R. Keith Ellis, and Raoul Rontsch. Single top production in association with a Z boson at the LHC. *Phys. Rev.*, D87:114006, 2013.
- [3] Celine Degrande, Fabio Maltoni, Ken Mimasu, Eleni Vryonidou, and Cen Zhang. Single-top associated production with a Z or H boson at the LHC: the SMEFT interpretation. *JHEP*, 10:005, 2018.
- [4] ATLAS Collaboration. Measurement of the production cross-section of a single top quark in association with a Z boson in proton-proton collisions at 13 TeV with the ATLAS detector. *Phys. Lett.*, B780:557–577, 2018.
- [5] CMS Collaboration. Measurement of the associated production of a single top quark and a Z boson in pp collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett.*, B779:358–384, 2018.
- [6] CMS Collaboration. Supplementary material for TOP-18-008 : Observation of Single Top Quark Production in Association with a Z Boson in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV.