



Facultad de Ciencias

**Medida de la sección eficaz de pares de
quarks top-antitop con métodos
multidimensionales en el experimento CMS
de LHC (CERN)**

**(Top quark pair cross section measurement
with Multi Variable Methods in the CMS
experiment of the LHC (CERN))**

Trabajo de Fin de Máster
para acceder al

**MÁSTER EN FÍSICA, INSTRUMENTACIÓN Y MEDIO
AMBIENTE**

Autor: Ignacio Suárez Andrés

Directores: Javier Fernández Menéndez
Celso Martínez Rivero

Julio - 2015

Contents

1	Resumen	4
2	Introduction	5
3	Physical context	6
3.1	Standard Model and extensions	6
3.2	Top quark processes	6
4	Large Hadron Collider	8
4.1	CMS experiment	9
4.1.1	Tracker	10
4.1.2	Electromagnetic calorimeter	11
4.1.3	Hadronic calorimeter	11
4.1.4	Muon system	12
4.1.5	Trigger	13
4.1.5.1	Level 1 Trigger	13
4.1.5.2	High Level Trigger	13
4.1.6	Identification	13
5	Physical objects and processes	15
5.1	Observables	15
5.1.1	Muons	15
5.1.1.1	Reconstruction	15
5.1.1.2	Identification	15
5.1.1.3	Isolation	16
5.1.2	Electrons	16
5.1.2.1	Reconstruction	16
5.1.2.2	Identification	17
5.1.2.3	Isolation	17
5.1.3	Jets	17
5.1.4	Missing transverse energy	18
5.2	Event selection	18
5.3	Processes	20
6	Boosted Decision Trees	22
6.1	Motivation	22
6.2	Decision trees	22
6.3	Boosting	24
6.4	Parameters	26
7	Analysis and results	29
7.1	$t\bar{t}$	30
7.1.1	Cut on BDT output	30

7.1.2	Fit signal to BDT output	31
7.2	$t\bar{t}W$	35
7.2.1	Cut on BDT output	35
7.2.2	Fit signal to BDT output	36
7.3	$t\bar{t}Z$	39
7.3.1	Cut on BDT output	39
7.3.2	Fit signal to BDT output	40
8	Conclusions	44
9	References	45

1 Resumen

El Gran Colisionador de Hadrones (LHC) ha comenzado recientemente su *Run II*, produciendo colisiones protón-protón con una energía en el centro de masas de 13 TeV, tras haber alcanzado hasta 8 TeV en el *Run I*. Este trabajo debe entenderse en este contexto científico e histórico, en el que las búsquedas de nueva física juegan un papel central en los experimentos de colisionadores.

Para poder descubrir nuevas partículas elementales como las predichas por varias extensiones del Modelo Estándar, se requiere un conocimiento muy preciso de dicho modelo. Los procesos que involucran producción de quarks top, como $t\bar{t}$, $t\bar{t}W$ y $t\bar{t}Z$, suponen un ejemplo interesante ya que son fuentes relevantes de fondo en las búsquedas de partículas supersimétricas (SUSY), siendo además procesos importantes por sí mismos.

Esta constante necesidad de obtener medidas más precisas, especialmente de procesos con baja sección eficaz, sirve de motivación para explorar técnicas avanzadas que puedan mejorar nuestro rechazo de sucesos de fondo al realizar medidas. El método aquí presentado se denomina Boosted Decision Trees (BDTs), un tipo de algoritmo de Análisis Multivariable (MVA) escogido por ser relativamente intuitivo en comparación con otros métodos populares como las redes neuronales artificiales. Se proporciona una descripción exhaustiva de los Boosted Decision Trees, entendidos como la aplicación del algoritmo de *boosting* AdaBoost a simples árboles de decisión para mejorar drásticamente su poder de clasificación.

Se utilizaron Boosted Decision Trees en el análisis de tres diferentes procesos de producción de quarks top: $t\bar{t}$ en el canal con dos leptones de carga opuesta y distinto sabor en el estado final, $t\bar{t}W$ en el canal trileptónico y $t\bar{t}Z$ con cuatro leptones. En cada caso se emplearon dos interpretaciones distintas de la salida de los BDTs: como una única variable discriminante que permite sustituir el análisis habitual con cortes secuenciales y como una función de los datos cuya forma tiene una gran dependencia con los sucesos de señal, lo que se puede aprovechar para reescalar la simulación de la señal de modo que se ajuste correctamente a los datos.

Los resultados para el análisis de $t\bar{t}$ son dos medidas de sección eficaz consistentes con las predicciones teóricas y con medidas experimentales publicadas por la colaboración CMS. Para $t\bar{t}W$ y $t\bar{t}Z$, se ha realizado con éxito una observación de ambos procesos a pesar de su baja sección eficaz, demostrando la potencial utilidad de los métodos de Análisis Multivariable para futuros análisis más avanzados.

Palabras clave: Física de Altas Energías, Física de Partículas, CMS, Top, Multileptones

2 Introduction

The Large Hadron Collider (LHC) has recently started its Run II, producing proton-proton collisions with a centre-of-mass energy of 13 TeV, after achieving up to 8 TeV in Run I. This work is to be understood within this scientific and historical context, in which new physics searches play a central role in collider experiments.

In order to discover new elementary particles predicted by extensions of the Standard Model, very precise knowledge is required about the Standard Model itself. Processes involving top quark production, such as $t\bar{t}$, $t\bar{t}W$ and $t\bar{t}Z$, are an interesting example since they are relevant background sources in searches for Supersymmetric (SUSY) particles, while also being important processes on their own.

The constant need for more precise measurements, specially in low cross section processes, serves as motivation to explore advanced techniques for improving our removal of background events during measurements. The method presented here is known as Boosted Decision Trees (BDTs), a type of Multivariate Analysis (MVA) algorithm chosen for being relatively intuitive as opposed to other popular method such as Artificial Neural Networks. A thorough description of Boosted Decision Trees is provided, understanding them as the application of the boosting algorithm AdaBoost to simple decision trees to drastically improve their separation power.

Boosted Decision Trees were used in the analysis of three top quark production processes: $t\bar{t}$ in the opposite-sign different-flavour dilepton channel, $t\bar{t}W$ in the trilepton channel and $t\bar{t}Z$ in the four-lepton channel. In each case, the output from the BDTs was interpreted in two different ways: as a single discriminating variable that substitutes a sequential cut analysis and as a function of the data whose shape is greatly influenced by signal events, which can be taken advantage of to rescale signal simulations so that they fit the data.

The results for the $t\bar{t}$ analysis are two cross section measurements that are consistent with theoretical predictions and experimental measurements published by the CMS collaboration. For $t\bar{t}W$ and $t\bar{t}Z$, a successful observation was made despite their low cross section, demonstrating the potential usefulness of MVA methods in further analyses.

Keywords: High Energy Physics, Particle Physics, CMS, Top, Multileptons

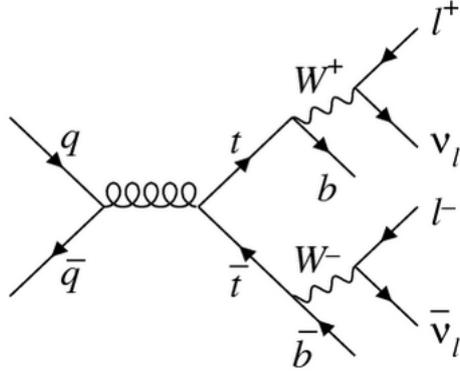


Figure 2: Leading-order Feynman diagram for $t\bar{t}$ production in proton-proton collisions, with W bosons decaying leptonically.

Measurements of related processes involving an additional vector boson, such as $t\bar{t}W$ or $t\bar{t}Z$ production (shown in figure 3), also have an impact in searches for new physics. More precisely, their cross section imposes constraints on the values of dimension-six operators that appear in extensions of the Standard Model lagrangian [3][4]. Within the Standard Model, these processes provide information about the vector and axial components of the coupling of top quarks to vector bosons [5][6] and act as background sources in measurements of other rare processes with multileptonic final states such as $t\bar{t}H$ production.

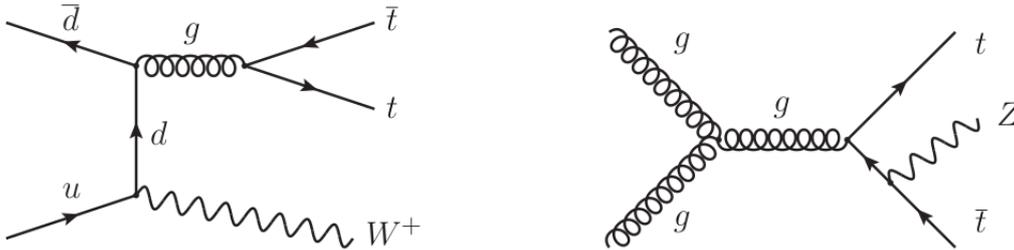


Figure 3: Leading-order Feynman diagrams for $t\bar{t}W$ and $t\bar{t}Z$ production in proton-proton collisions.

4 Large Hadron Collider

The Large Hadron Collider or LHC is a circular particle accelerator built by CERN (*Conseil Européen pour la Recherche Nucléaire*) in Geneva, Switzerland. This accelerator is located in the same tunnel that used to contain its predecessor, LEP, with a length of 26.7 km and being between 45 and 170 m deep under Swiss and French territory. [7]

Four main detectors can be found in LHC, coinciding with the interaction points of the proton beams as shown in figure 4.

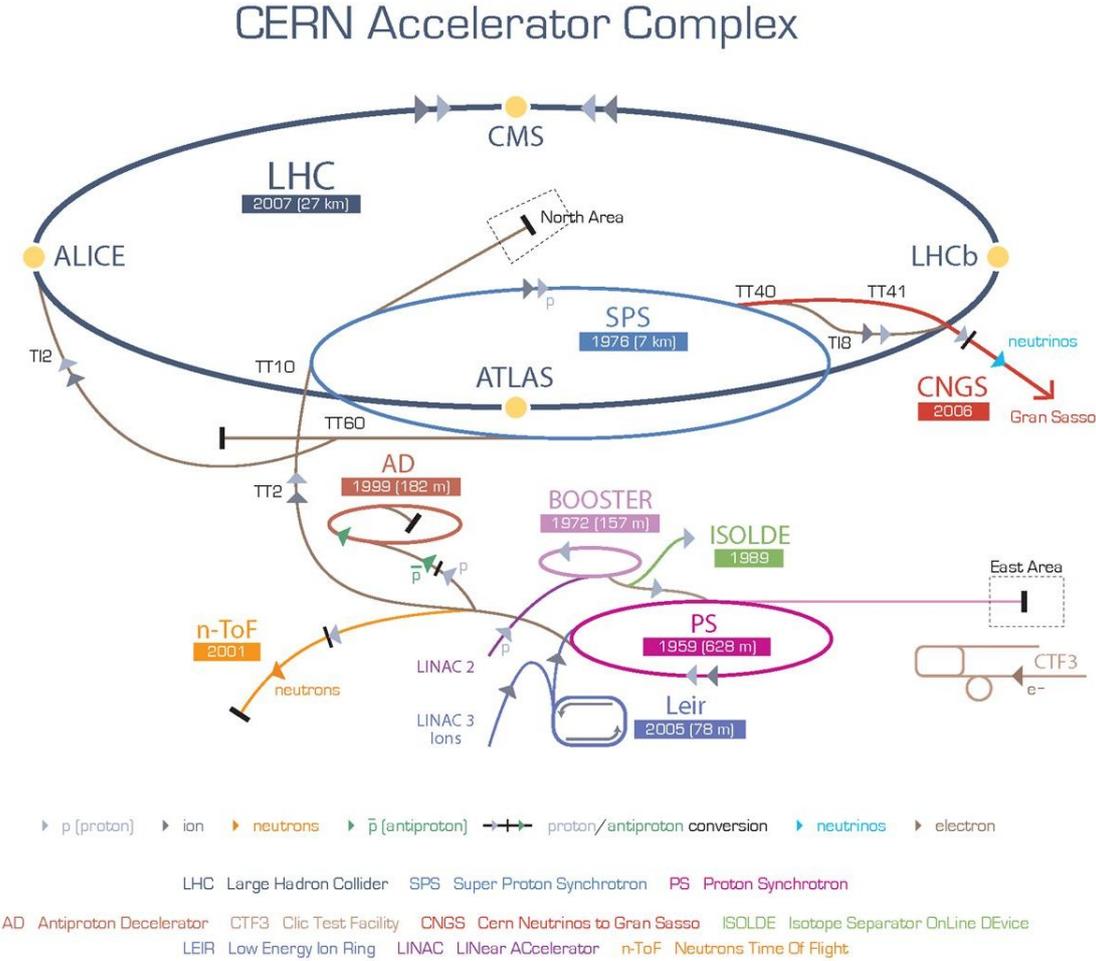


Figure 4: Relative location of detector experiments and complementary accelerators of LHC.

ATLAS (A Toroidal LHC Apparatus) is a general-purpose detector, mainly used in searches for new physics, such as dark matter candidates, and testing of Standard Model predictions.

CMS (Compact Muon Solenoid) is the other general-purpose detector, allowing for double-checking of discoveries. It will be further described below, as this work is based on data collected by it.

ALICE (A Large Ion Collider Experiment) is specialised in heavy-ion physics, exploring properties of the strong interaction such as the predicted quark-gluon plasma state.

LHCb (LHC - beauty) is dedicated to the study of b -quark physics. More precisely, decay of some B mesons allow for measurements of CP violation, potentially shedding some light on the mystery of matter-antimatter asymmetry.

LHC is designed to collide opposing proton beams with a maximum centre-of-mass energy of $\sqrt{s} = 14$ TeV, although it currently reaches 13 TeV. In 2012, when the data used in this work were collected, the value of this energy was 8 TeV, about four times the highest energy previously reached, in Tevatron at Fermilab.

Protons travel around the LHC ring in bunches, contained by multipolar magnets. These bunches cross at the interaction points where the detectors are located every 25 ns (50 ns in 2012), producing an average of 15 primary vertices per crossing. This phenomenon, called *pile-up* and illustrated in figure 5, introduces a need for high resolution and occupancy limit in the detectors. The event rate is about 10^9 Hz.

The amount of data generated is measured in instantaneous luminosity, which describes the number of collisions that take place per unit of area and time. It is dependent on several parameters such as the crossing frequency, the effective size of the proton bunches and the number of protons in each bunch. The number of data taken over a period of time is expressed in terms of (integrated) luminosity, such as the 20 fb^{-1} collected by CMS in 2012. Figure 6 shows how luminosity was accumulated during the years 2010, 2011 and 2012.

4.1 CMS experiment

CMS [8] is a cylindrical detector located 100 m deep under the south-eastern edge of the French village of Cessy, between the Jura Mountains and Lake Geneva. It is 21.6 metres long, has a diameter of 14.6 m and weights around 12500 tonnes.

CMS contains a superconducting solenoid with a length of 12.5 m and a diameter of 6 m, and within which a uniform magnetic field of 3.8 T is created. Its goal is to deviate electrically charged particles in order to measure the sign of their charge and their momentum. The return yoke of the magnet curves muons in the opposite way, improving their measurements, and isolates the muon chambers.[9, 10]

The coordinate system in CMS is taken so that the x axis points to the centre of LHC, the y axis points upwards and the z axis runs counter-clockwise along the beam pipe, if seen from above. The centre is defined as the nominal point of interaction between the proton beams. The θ angle is usually substituted by the pseudorapidity $\eta = -\ln[\tan(\theta/2)]$, which has a value of 0 at the x - y plane and tends to infinity in the z axis. The subdetectors of CMS cover the range up to $|\eta| \leq 5$.

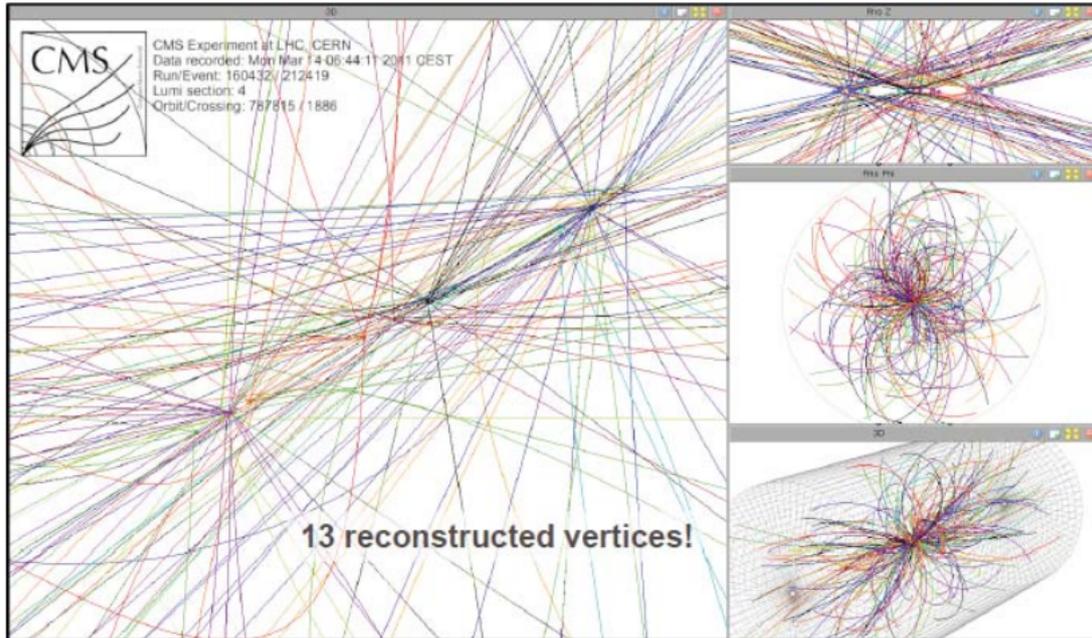


Figure 5: Result of a bunch crossing in 2011 ($\sqrt{s} = 7$ TeV) as measured by the CMS detector, with a pile-up of 13 reconstructed primary vertices.

The x - y plane is called the transverse plane, where some important observables are defined, such as the transverse momentum, measured as $p_T(\text{GeV}/c) = 0,3 \cdot B(\text{T}) \cdot R(\text{m})$ using the radius R of the curved track of a charged particle.

The CMS detector is designed to meet the following goals:

- Good muon identification and momentum resolution for a broad momentum and angle spectrum, good mass reconstruction for muon pairs ($\approx 1\%$ uncertainty at $100 \text{ GeV}/c^2$), correct charge assignment for muons up to $1 \text{ TeV}/c$.
- Good momentum and efficiency of reconstruction of charged particles from their tracks. Efficient b - and τ -tagging, with high detector granularity near the interaction point.
- Good electromagnetic energy resolution, good resolution in diphoton and dielectron mass reconstruction ($\approx 1\%$ uncertainty at $100 \text{ GeV}/c^2$), wide angular coverage, efficient π^0 rejection and good photon and lepton isolation in high luminosity conditions.
- Good resolution in missing transverse energy and dijet mass reconstruction, requiring hadronic calorimeters with the largest coverage possible and fine lateral segmentation.

A three-dimensional representation of CMS can be seen in figure 7, showing the subdetectors that are described in the following paragraphs.

4.1.1 Tracker

The tracker provides a precise measurement of the trajectories of charged particles, which are curved because of the magnet. Impacts in the various layers of the tracker are used to define

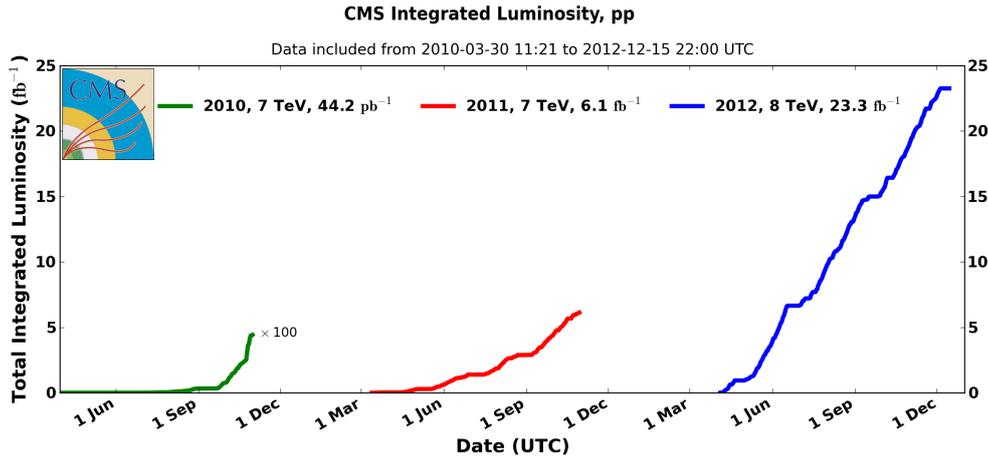


Figure 6: Integrated luminosity accumulated during the years 2010, 2011 and 2012.

tracks, with information about the charge and p_T of the particle, in addition to contributing to its identification.

This subdetector is the closest one to the centre of the detector, with a length of 5.8 m, a maximum radius of 1.13 m and a pseudorapidity range of $|\eta| < 2.5$.

The tracker has two distinct parts: the pixel tracker and the silicon strip tracker. The pixel tracker is in the innermost part of the detector. These pixels are silicon microstrips that provide a very high resolution, even more so in the inner layers, where they are smaller and their shape is closer to a square in order to keep occupancy levels constant across the layers of the detector. The silicon strip tracker is built around the pixel tracker. Although it has lower resolution, it covers a much larger volume, being the largest silicon-based tracker in the world, with a detection area of 200 m². Both parts are divided into a barrel (curved part of the cylinder) and two endcaps (bases) sections, while the barrel of the silicon strip tracker also has an inner and an outer section with different resolutions.

4.1.2 Electromagnetic calorimeter

The electromagnetic calorimeter, or ECAL, is based around the used of scintillator PbWO₄ crystals. This material was chosen due to its fast response, as upon receiving an impact it emits 80% of the scintillation light in 25 ns, which the smallest time between collisions that LHC was designed for. The barrel section covers the range $|\eta| < 1.5$, while the endcaps section extends this up to 3.0.

4.1.3 Hadronic calorimeter

The hadronic calorimeter, or HCAL, is designed to fit between the ECAL and the solenoid. The barrel area reaches up to $|\eta| < 1.3$ while the endcaps cover $1.3 < |\eta| < 3.0$. The detection is performed by scintillator plates made of different types of plastic. The main absorber material for slowing down hadrons is brass, as it is a non-magnetic alloy. In the $|\eta| < 1.26$ region, there is an additional calorimeter outside the solenoid with iron as the absorber, in order to

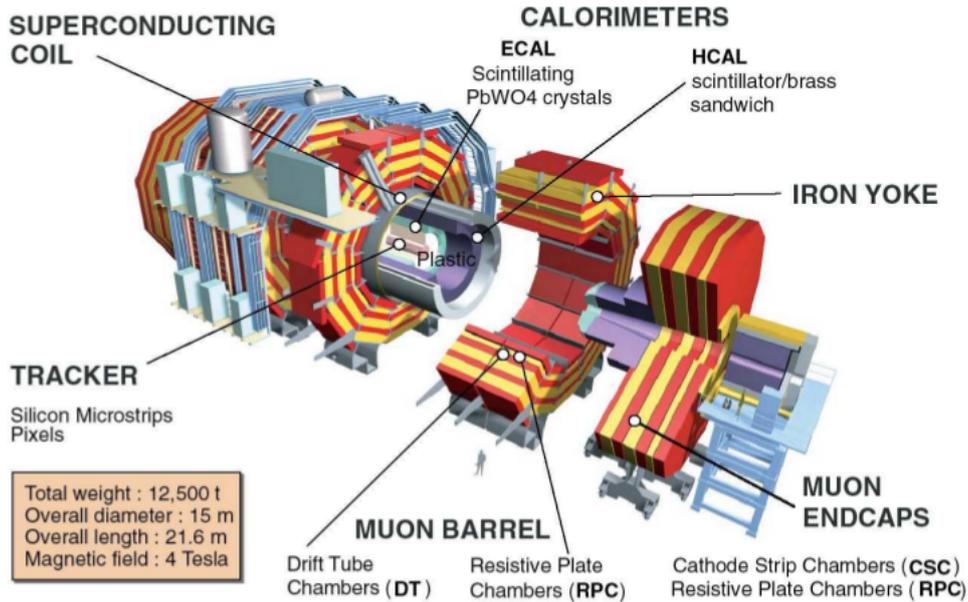


Figure 7: 3D view of CMS and its parts.

stop high- p_T hadrons. Finally, the forward detectors, closest to beam pipe cover the region $3.0 < |\eta| < 5.0$. They are made of quartz fibres that collect Cherenkov radiation and steel was chosen as the absorber material.

4.1.4 Muon system

The muon system is contained within the iron return yoke of the solenoid, where almost exclusively muons can reach. It possesses three different types of subdetectors.

Drift tubes (DTs) are located in the central region $|\eta| < 1.2$, where the muon rate, the neutrons background and the residual magnetic field are low. Muons crossing this detector ionize a gaseous mixture, after which electrons drift towards the anode wire within a maximum of 380 ns. Some tubes provide r and ϕ information and others measure the z coordinate, allowing for 3D reconstruction of trajectories.

The cathode strip chambers (CSCs) cover the endcaps region, $0.9 < |\eta| < 2.4$, due to their flat shape, their larger muon acceptance and their ability to function inside a strong magnetic field. The grids formed by anode wires and cathode strips give a bidimensional measurement in each plane of the subdetector. The resistive plate chambers (RPCs) complement measurements in the $|\eta| < 1.6$ region with excellent temporal resolution, of about 1 ns, which is necessary to match muons to the bunch crossing they came from. In this case, electrons from the gas mixture are collected in aluminum strips, with a response time comparable to that of scintillators.

4.1.5 Trigger

LHC generates events with a frequency of 20 MHz (when the time interval between crossings is 50 ns), with each event occupying 1.5 MB. In order to be able to process such a humongous amount of data, it needs to be reduced as soon as possible, following criteria that ensure that potentially interesting events are not discarded by the filters. This whole process is known as the *trigger*.

4.1.5.1 Level 1 Trigger

The Level 1 Trigger is implemented directly on the hardware. It filters events based on simple criteria applied to measurements performed by the subdetectors, such as a minimum p_T of charged particles. The event output frequency of this level is no larger than 100 kHz.

4.1.5.2 High Level Trigger

The next step, the High Level Trigger or HLT, is software-based. It uses information from fully reconstructed objects, having more complex criteria which might include lepton isolation or origin point of jets. The event frequency of the final output is between 0.1 and 1 MHz.[11]

4.1.6 Identification

Since more refined methods require low-momentum (up to a few GeV/ c) particles, identification in CMS is performed basically by checking which subdetector a particle has interacted with, as shown in figure 8.

- Muons, as electrically charged particles, are detected in both the tracker and the ECAL. Save for those with low p_T , they are expected to reach the muon system, being the only ones to do so due to their minimum interaction with matter.
- Electrons are detected in the tracker and leave an energy cluster in the ECAL, where they are stopped.
- Photons do not leave a track since they are neutral particle, but as they are affected by electromagnetic interactions, they are detected in the ECAL.
- Charged hadrons are detected in the tracker and leave part of their energy in both the ECAL and the HCAL. Neutral hadrons, on the other hand, are only detected in the HCAL.
- Neutrinos and some hypothetical particles not included in the Standard Model are only affected by the weak interaction and cannot be detected by CMS.

However, this is only a simplified version describing individual particles, as hadrons are contained within jets, π^0 might decay into a pair of photons, electrons emit bremsstrahlung radiation, photons are prone to pair production, and so on.

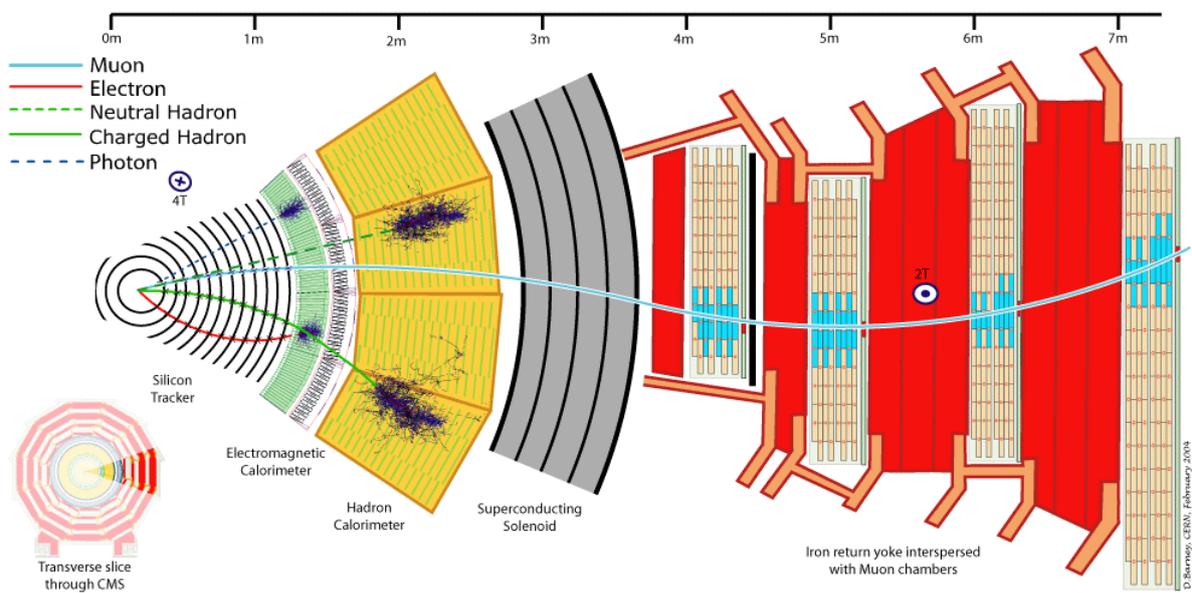


Figure 8: Interaction between different types of particles and the subdetectors of CMS.

5 Physical objects and processes

5.1 Observables

5.1.1 Muons

5.1.1.1 Reconstruction

Muon reconstruction begins using information from the muon chambers, defining a *stand-alone muon*. Impacts in the first station are used as seeds for local fits that are gradually extended outwards, accounting for the effects of the magnetic field and interaction with the iron yoke.

Once the stand-alone muon has been reconstructed, the algorithm searches for compatible extrapolated trajectories from the tracker. This matching depends on the quality of the fit between the extrapolated track and the reconstruction in the muon chambers, assuming the muon is originated at the nominal interaction point. The result of this matching provides what is called a *global muon*. The benefits of measuring muons with two subdetectors are made clear in figure 9.

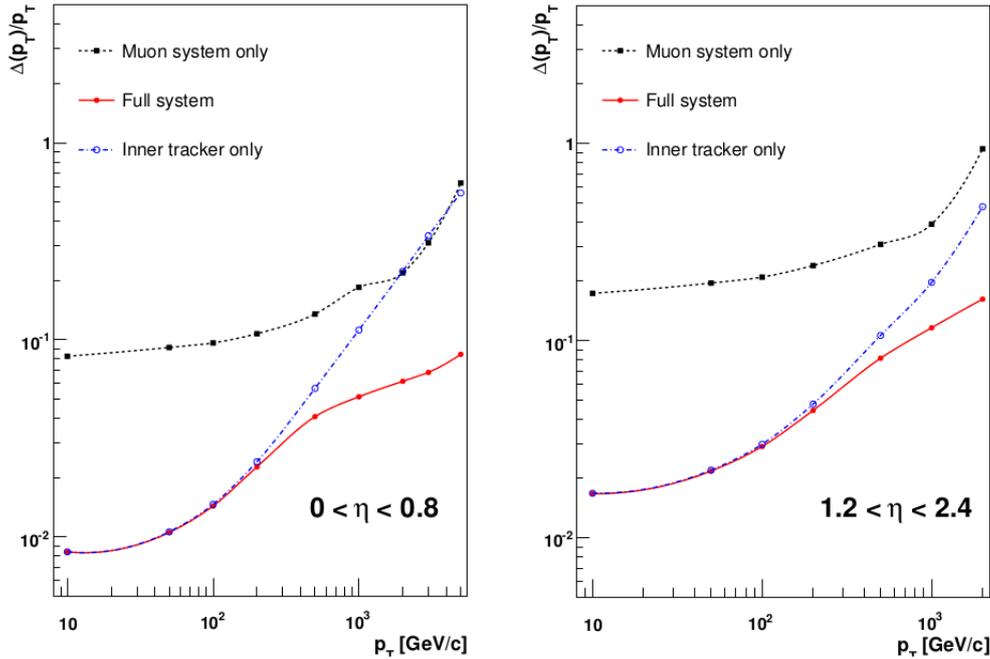


Figure 9: Relative uncertainty in muon p_T measurements, depending on which subdetectors were used for the reconstruction. The improvement in momentum resolution is especially significant for high- p_T muons.

5.1.1.2 Identification

In parallel, muon identification algorithms are run in order to detect muons with loose conditions, such as those not fully reconstructed as stand-alone muons. This is often necessary

when a high number of muons in the final state are required, as some of them are likely to have low momentum and not be able to reach the outermost layers of CMS.

Therefore, these algorithms work from the inside out starting from the tracker, searching for matching energy deposits in the calorimeters, compatible with a minimum ionizing particle. When the muon chambers are reached, this trajectory is extended as much as possible through sequential χ^2 fits, giving a larger weight to impacts in the outermost layers as it is more likely to belong to a muon.

Objects identified as muons may also be rejected depending on their impact parameter, i.e. the distance from their reconstructed track to the primary vertex, as a large value indicates they were originated in heavy quark decays or even in other vertices from the pile-up.

5.1.1.3 Isolation

Muon isolation is especially important in the low- p_T region, below 30 GeV/c, since their primary source becomes decays of heavy quarks instead of weak bosons. Therefore, these muons are recognisable due to being part of jets, whereas they would otherwise be only accompanied by particles created in the underlying event (secondary interaction between other components of the colliding protons), initial or final-state radiation or coming simply from the pile-up.

In order to discard these muons generated as a part of jets, a cone is defined around them, with size $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ and with its axis aligned with the initial direction of the muon, as shown in figure 10. If the sum of energy or p_T contributions coming from particles other than the muon is below a certain threshold, the muon is considered isolated. Originally, these energy or p_T measurements were taken directly from the calorimeters or the tracker; however, the values of the fully reconstructed particles are used since the development of the Particle Flow algorithm [12].

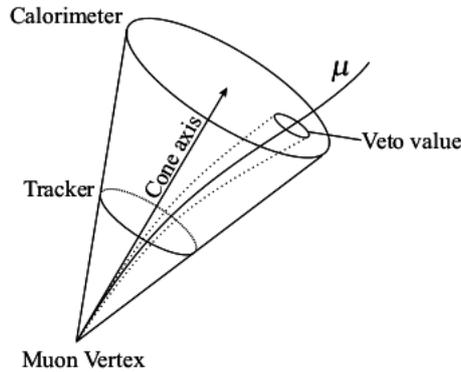


Figure 10: Isolation cone.

5.1.2 Electrons

5.1.2.1 Reconstruction

Electron reconstruction is highly conditioned by bremsstrahlung emission, which causes energy to be scattered in an inverse proportion to the electron's momentum. Therefore, measurement in the ECAL are not in the form of individual hits, but rather of superclusters (clusters of clusters) generated by electromagnetic showers. The mean positions of these superclusters are used in the search for compatible hits in the pixel detector, which are then taken as seeds for fitting tracks in the silicon strip tracker.

5.1.2.2 Identification

An electron candidate must have been fully reconstructed by the previous method. Additionally, it must meet a series of criteria, in order to reject possible *fakes*, mainly jets.

Some of the variables used are the sizes $\sigma_{\phi\phi}$ and $\sigma_{\eta\eta}$, the distance in the $\eta - \phi$ plane between the mean position of the supercluster and the track extrapolation, the ratio between the measured energy and the initial momentum or the fraction of energy that leaks into the hadronic calorimeter.

5.1.2.3 Isolation

In a similar manner to muons, electron isolation is also quantified by defining cones of size $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ around them and summing all p_T contributions inside of it, while removing that of the electron itself. Impact parameters are again taken into account in order to verify whether these additional particles in the cone come from the same primary vertex.

5.1.3 Jets

When the products of an interaction process include high-energy gluons or quarks, they initiate a series of successive decays known as parton showers. When energy has decreased enough, the confinement property of Quantum Chromodynamics causes the gluons and quarks generated to be grouped into hadrons. The final state measured by the detectors is an approximately conical stream of hadrons and other decay subproducts, called a jet, which carries over the momentum of the original quark or gluon. Jet production cross section dominates specially at high p_T , due to the large QCD cross sections.

Jet identification uses information from the tracker and both calorimeters, since low momentum charged particles rarely reach the calorimeters and neutral particles cannot be measured by the tracker. This is done by matching tracks to clusters of energy deposits in the calorimeters, ideally refining the information obtained with the Particle Flow algorithm.

There are several options for clustering all of its components into a single jet object. One of them repeatedly generates cones around particles, finding new candidates for jet ingredients, until energy and orientation of the jets are stabilized and all objects have been successfully clustered. Other possibilities combine distance between particles with their transverse energy in order to leave two objects as separate jets or merge them.

Jets coming from b quarks may be identified, in a process called b -tagging, through properties such as their decay time. This is the main way to detect the generation of top quarks in an event. The variable used here is called *Track Counting High Efficiency* (TCHE) [13], which takes the value of the second highest significance of the 3D impact parameter of all tracks in

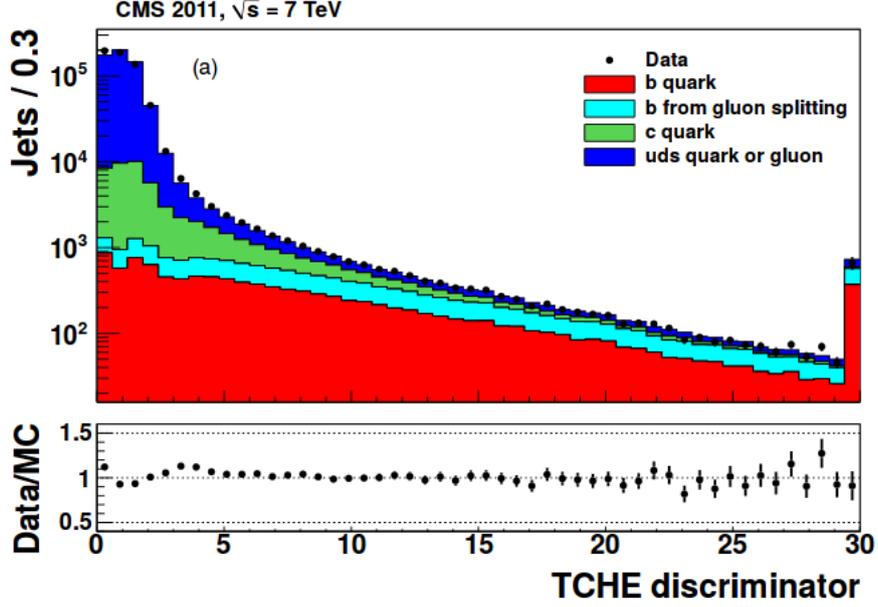


Figure 11: Distribution of TCHE for heavy and light quarks.

the jet. This significance is defined as the ratio between the impact parameter and its estimated uncertainty. The distributions of the TCHE discriminator for different quark flavours are shown in figure 11.

5.1.4 Missing transverse energy

Since neutrinos as well as several Supersymmetric particles cannot be directly detected, their presence is determined through momentum conservation. Although total momentum along the z axis is unknown due to its dependence on parton distribution functions, total transverse momentum in the collision is 0. Assuming ultra-relativistic particles, missing energy in the transverse plane is calculated as

$$\vec{E}_T^{\text{miss}} = - \sum \vec{p}_{Tc} = - \sum_n (E_n \sin \theta_n \cos \phi_n \hat{\mathbf{i}} + E_n \sin \theta_n \sin \phi_n \hat{\mathbf{j}}). \quad (1)$$

This calculation uses either p_T measurements from tracks or energy deposits in the calorimeters, depending on the level of reconstruction of each object.

5.2 Event selection

Data events are required to fire off at least one of the triggers listed in table 1. The general preselection for the three analyses uses events with at least two leptons and also at least two jets, optimizing lepton selection in the following way:

1. Muons

lepton type	trigger paths
SingleElectron	HLT_Ele27_WP80
SingleMu	HLT_IsoMu24_eta2p1
DoubleElectron	HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_ _Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL
DoubleMu	HLT_Mu17_Mu8 HLT_Mu17_TkMu8
MuEG	HLT_Mu17_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoV HLT_Mu8_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoV

Table 1: Trigger paths used to select events.

- $p_T > 20$ GeV/ c for the most energetic lepton, $p_T > 10$ GeV/ c for the second one.
- $|\eta| < 2.4$, due to the geometry of the muon system.
- More than 10 impacts in the inner tracker, with at least one hit in the pixel detector.
- Global fit $\chi^2/\text{d.o.f.} < 10$, rejecting muons from in-flight decays of heavy quarks and hadrons reaching the muon system.
- Impact parameter in the transverse plane $|d_0| < 0.02$ cm. This rejects up to 87% of *fakes*, while keeping about 93% of real muons.
- Longitudinal impact parameter $|d_z| < 0.1$ cm, selecting muons from the main event taken place at the primary vertex.
- Relative p_T uncertainty better than 10%. Besides improving p_T distributions, this also affects the quality of E_T^{miss} measurements.
- $\chi^2/\text{d.o.f.} < 20$ in the *Kink-finder* [14], which searches for impacts around reconstructed trajectories in order to detect in-flight decays.

For further rejection of jets, an MVA-based algorithm was used for describing isolation.

2. Electrons

First of all, electrons have been identified using a MVA method with the following variables:

- Supercluster η width $\sigma_{i\eta i\eta}$.
- Supercluster ϕ width $\sigma_{i\phi i\phi}$.
- $\Delta\eta_{in}$, distance in η between the track and the supercluster.
- $\Delta\phi_{in}$, distance in ϕ between the track and the supercluster.
- Fraction of the total momentum lost due to bremsstrahlung.
- Number of additional clusters from bremsstrahlung.

Afterwards, the selection is made even tighter with a series of cuts:

- $p_T > 20 \text{ GeV}/c$.
- $|\eta| < 2.5$.
- $\sigma_{i\eta i\eta} < 0.15(0.03)$ in the barrel (endcaps) region.
- $|\Delta\phi_{in}| < 0.15(0.010)$, $|\Delta\phi_{in}| < 0.007(0.009)$.
- $E^{\text{HCAL}}/E^{\text{ECAL}} < 0.12(0.10)$.
- $\sum_{\text{tracker}} E_T/p_T < 0.2$, $\sum_{\text{ECAL}} E_T/p_T < 0.2$, $\sum_{\text{HCAL}} E_T/p_T < 0.2$.
- $\text{Iso}_{\text{PF}}/p_T < 0.13(0.09)$.

For the $t\bar{t}$ analysis, the number of leptons required is exactly two, more precisely one electron and one muon with opposite charge.

For $t\bar{t}W$, exactly three leptons are required, with their total charge adding up to 1 or -1. Additionally, the event must not contain a pair of same-flavour opposite-sign leptons whose reconstructed mass falls within $15 \text{ GeV}/c^2$ of the Z boson mass, and at least two jets must be marked as b -tagged, which means having a TCHE value larger than 2.1.

Finally, the $t\bar{t}Z$ selection requires exactly four leptons, with zero total charge. Out of these four leptons, there must be a same-flavour, opposite-sign pair inside the Z mass peak.

5.3 Processes

The list of physical processes included in the analysis, along with their cross section at $\sqrt{s} = 8 \text{ TeV}$ and the Monte Carlo generator used for the simulations, is shown in table 2.

Process	cross section (pb)	Generator
$t\bar{t}$	248	MADGRAPH [15]
$t\bar{t}W$	0.232	MADGRAPH
$t\bar{t}Z$	0.206	MADGRAPH
tW	2.34	POWHEG [16][17]
Drell-Yan	4390	MADGRAPH
$WW \rightarrow 2l2\nu$	5.81	MADGRAPH
WZ	4.85	MADGRAPH
ZZ	2.03	MADGRAPH
$WW\gamma$	0.528	MADGRAPH
WWW	0.0822	MADGRAPH
WWZ	0.0633	MADGRAPH
WZZ	0.0192	MADGRAPH
ZZZ	0.00459	MADGRAPH
$t\bar{t}\gamma$	1.44	MADGRAPH
$t\bar{t}WW$	0.00204	MADGRAPH
$(t\bar{t})H \rightarrow (t\bar{t})WW \rightarrow (t\bar{t})2l2\nu$	0.744	POWHEG
$H \rightarrow ZZ \rightarrow 4l$	0.0524	POWHEG
$W+\text{jets} \rightarrow l\nu+\text{jets}$	37.5	MADGRAPH
$W\gamma \rightarrow l\nu\gamma$	8.12	MADGRAPH
$W\gamma^* \rightarrow l\nu\gamma^*$	554	MADGRAPH
$Z\gamma \rightarrow ll\gamma$	133	MADGRAPH

Table 2: List of processes, 8 TeV cross section and Monte Carlo generator used. The parton showers and hadronization part of the simulation was performed with Pythia [18].

6 Boosted Decision Trees

Multivariate Analysis methods combine several variables, which by themselves might not be enough to efficiently select signal events from data, into a single variable that acts as a powerful discriminator. Since Boosted Decision Trees are the MVA algorithm chosen in this work, they will be thoroughly described in this section.

6.1 Motivation

Decision trees are easy to understand even by non-experts, as they can be considered similar to human procedures for making decisions, such as medical diagnosis. Explanations are made simple by the use of graphical representations, such as those of borders between different classes in low-dimensionality problems or the trees themselves. Although boosting introduces more complicated concepts, the fundamental idea is still intuitive.

Performance of Boosted Decision Trees usually falls behind that of an Artificial Neural Network (ANN), another commonly used Multivariate Analysis method, perfectly optimized for the task at hand. However, BDTs require much less fine-tuning in comparison, being widely regarded as the best “out-of-the box” classifier, since finding a set of parameters that lead to a satisfactory BDT is not as extremely time-consuming as in the case of ANNs.

Training of Boosted Decision Trees is not significantly impeded by the presence of irrelevant variables. Due to the use of maximum separation criteria, these variables are largely ignored. Actually, input variables may be ranked in terms of how often they were used for making decisions during the training, taking into account the separation achieved and the weight of the corresponding trees. This is called variable importance.

Computation time required for growing small trees, like the ones the boosting algorithm AdaBoost is specially designed for, is very short, therefore it may be possible to create a large set of Boosted Decision Trees in a matter of minutes when running in a single machine. This allows the user to test the impact of gradually altering several parameters.

Finally, Boosted Decision Trees are able to adapt to complex boundaries between classes, as shown in figure 12a. However, they may not do so in the most optimal way, requiring more time than classifiers specialized in specific shapes, as exemplified in figure 12b.

6.2 Decision trees

Decision trees are mostly used as a classification algorithm, although there are adaptation for regression problems. The fundamental idea behind them is to perform a series of simple splits in the variable space, establishing a list of binary conditions which are combined into the final classification criterion.

Figure 13a shows a generic example with two classes. Originally, there is a single *root node* which contains all the events to be classified. The variable space is partitioned into a grid, from which the single-variable cut that maximizes separation between classes is chosen, thus splitting the node into two. This process is repeated until some stopping condition is met, such as a high degree of purity or a small number of events in the node. The final nodes, called *leaf*

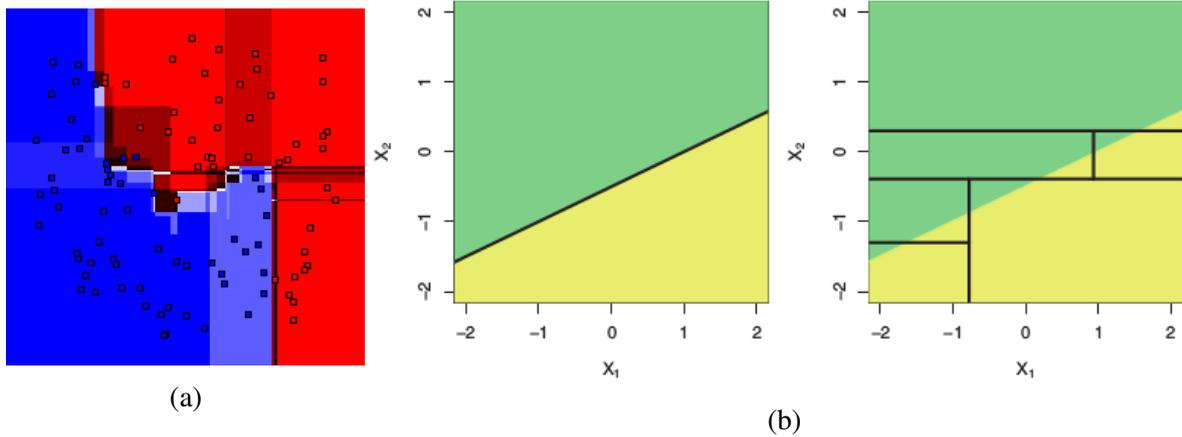


Figure 12: (a) Binary classification with non-linear boundaries using Boosted Decision Trees. (b) Describing a linear boundary is a trivial matter for a linear discriminant, while a decision tree requires several steps for a rough approximation.

nodes, are associated with a single class, therefore representing the output of the decision tree as a function of the split criteria used from the root node up to their own parent node.

A bidimensional graphical example is also presented in figure 13b, illustrating the effects of splitting nodes.

The maximum separation criterion used for selecting the best way to split a node is one of several magnitudes related to misclassification or purity [19]. The simplest one is the misclassification error, expressed as

$$E_m = 1 - \max_k(\hat{p}_{mk}), \quad (2)$$

where \hat{p}_{mk} is the fraction of events in node m that belong to class k . However, it is rarely used due to its low sensitivity for building trees and generally discarded in favour of one of the two following quantities.

The Gini index is defined as

$$G_m = \sum_{k=1}^K \hat{p}_{mk}(1 - \hat{p}_{mk}) \quad (3)$$

while cross-entropy takes the form of

$$D_m = - \sum_{k=1}^K \hat{p}_{mk} \log \hat{p}_{mk}. \quad (4)$$

Both indices tend to 0 when \hat{p}_{mk} is approximately 0 or 1, so they can be said to quantify the purity of node m , i.e. the predominance of one class over others. They are also known to take similar values, as shown in figure 14.

A single decision tree has many limitations, such as a prediction power inferior to that of many other classifiers or a strong tendency to overfitting, even after “pruning” irrelevant nodes.

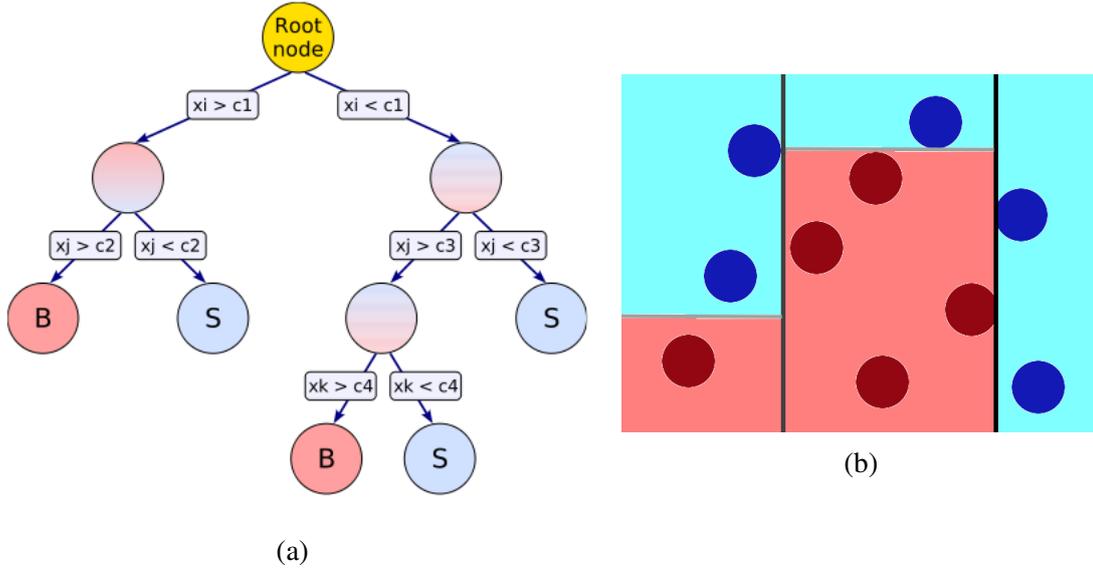


Figure 13: (a) Example of a decision tree with generic variables and binary classification between signal and background.
 (b) Bidimensional graphical example illustrating the partitioning of the variable space. Darker boundaries indicate splits closer to the root node.

They also provide unstable models, with a high sensitivity to small variations in the training dataset.

These reasons have led to the development of several methods that enhance the robustness and overall performance of decision trees by combining the output of different trees through a majority vote. One of these methods is called *bagging*, from *bootstrap aggregating*, and it generates a large number of variations of the original tree by randomly choosing events with the bootstrapping method, possibly also limiting the set of variables used in each tree. The method is used in this work, *boosting*, follows a different approach, instead sequentially growing trees so that each attempts to correct the classification mistakes of the previous one.

6.3 Boosting

The most commonly used method is Adaptive Boosting or AdaBoost [20]. In each iteration, events mislabeled by the previous classifier (decision tree, in this case) are assigned a larger weight, so that they are focused on in the next iteration. Figure 15 is useful for understanding how this process takes place and the way it progresses towards the final output given by the combination of all classifiers.

The AdaBoost algorithm follows these steps[21]:

1. One starts with a dataset $\{\vec{x}_i, y_i\}$, with $\vec{x}_i \in \vec{X}$ as variables and $y_i \in \{-1, +1\}$ as class labels.
2. All weights are initialized with the same value, $w_1(i) = 1/N$, with N being the number

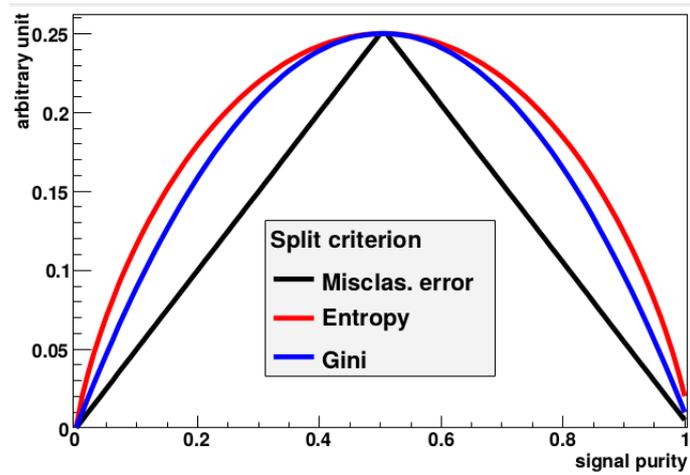


Figure 14: Values, in arbitrary scale, of three indices used as separation criteria as a function of signal purity.

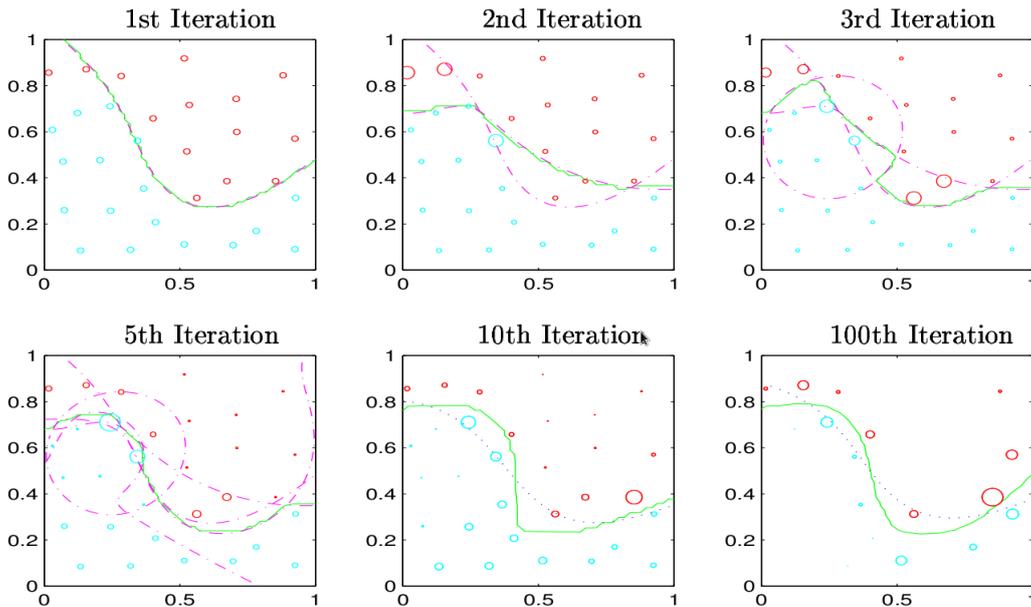


Figure 15: Application of AdaBoost for classifying two types of points, whose size represents their weight relative to the first iteration. Dashed lines mark the borders given by individual classifiers, while the green solid line describes the output of the combined classifier. The dotted lines in the last two graphs show the output given by the bagging method.

of events.

3. For each classifier $t = 1, \dots, T$:

- The classifier is trained according to the weight distribution w_t .
- An output $h_t(\vec{x}_i)$ is obtained.
- The misclassification error is calculated with weights ε_t , where $\mathbf{I}[y_i \neq h_t(\vec{x}_i)]$ is 1 when event \vec{x}_i has been misclassified and 0 otherwise.

$$\varepsilon_t = \sum_{i=1}^N w_t(i) \mathbf{I}[y_i \neq h_t(\vec{x}_i)] \quad (5)$$

- The coefficient α_t is calculated from the misclassification error, with parameter β regulating learning speed, generally set to $1/2$.

$$\alpha_t = \beta \ln \left(\frac{1 - \varepsilon_t}{\varepsilon_t} \right) \quad (6)$$

- Weights are updated according to whether events were correctly classified or not (represented by the product $y_i \cdot h_t(\vec{x}_i)$), conditioned by α_t . Z_t is a normalization coefficient.

$$w_{t+1}(i) = \frac{w_t(i)}{Z_t} \exp[-\alpha_t y_i h_t(\vec{x}_i)] \quad (7)$$

4. The output of the combined classifier takes the form of

$$H(\vec{x}) = \frac{1}{\sum_{t=1}^T \alpha_t} \sum_{t=1}^T \alpha_t h_t(\vec{x}). \quad (8)$$

Since the classifier created in each iteration focuses on events that difficult to label correctly, after a number of steps the new classifiers only work for a limited set of events, misclassifying many others. This is the reason why the contributions of each classifier is weighted by α_t , mitigating this effect and giving more importance to classifiers that work best for the whole dataset in general. Figure 16 shows how this coefficient evolves, along with the misclassification error, as more trees are grown.

AdaBoost performs best with weak classifiers that properly label a moderately large fraction of the events, rather than capturing every detail from the training dataset. For Boosted Decision Trees, this often means limiting the depth of the trees to about three or even less layers of node splitting.

6.4 Parameters

The most important parameters to be tuned in BDTs are the number of trees and their maximum depth.

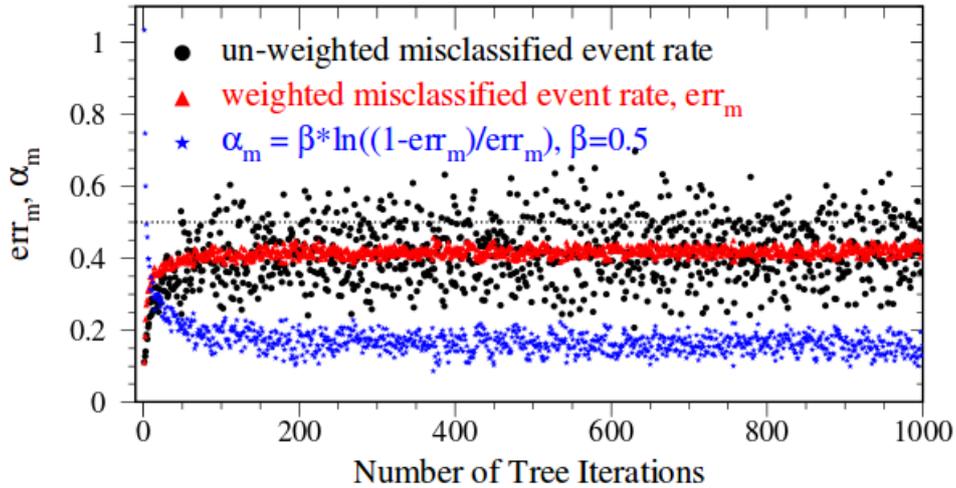


Figure 16: Example of the evolution of misclassification error and weights assigned to each tree during the training process. Simple misclassification error, which may rise above 0.5 for certain overly specialised trees, is shown in black. Weighted misclassification error is shown by the red points, which stabilize at a value below, but not very far from, 0.5. α_t weights, in blue, take larger values for the first trees and gradually decreases as weighted misclassification error increases.

An excess in the number of trees has little effect thanks to their weights limiting the impact of trees with bad generalization properties. However, it can lead to overtraining in extreme cases. The total separation power improves more slowly the larger the number of trees becomes, so there is an ample margin for the user to notice when enough trees have been grown. An example of this is presented in figure 17.

The maximum depth of the trees should be preferably as small as possible, since AdaBoost is oriented at weak classifiers. Trees with a single split (called *stumps*) are often enough, though the final choice is highly dependent on the specifics of the problem.

The learning parameter β is usually set at the standard value of 0.5, sometimes 1. However, it was noticed in this work that when one of the classes, such as signal or background, is significantly scarce, lowering this number (and adding more trees to compensate) prevents the trees from starting to overtrain too fast.

Other technical details will not be elaborated upon, as they were not tuned in this work during the training and optimization of the BDTs. These would include the lower limit on the number of events in a node for it to be divisible, the number of values of each variable to be considered when splitting nodes or the choice of separation index.

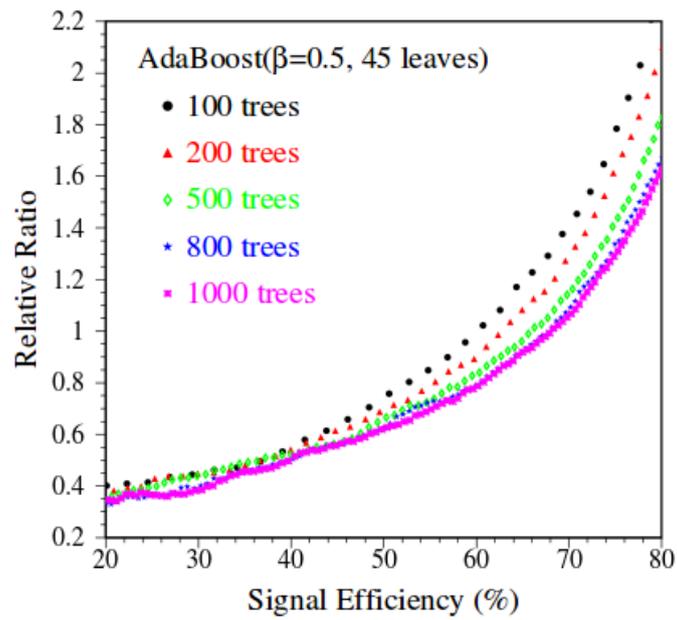


Figure 17: Improvement of the separation power of a BDT as the number of trees is increased. This data is taken from the MiniBooNE [22] experiment, which explains the large number of leaves, since more than 80 input variables are used. The vertical axis represents the ratio between the fraction of signal and background events kept for a given efficiency, multiplied by a constant value.

7 Analysis and results

This analysis was based on the implementation of Boosted Decision Trees in the Toolkit for Multivariate Analysis (TMVA) package [23] for ROOT [24].

Each of the three analyses was performed in two ways for cross-checking:

1. Cut method.

After training the BDT, the output is a single variable with more separation power than the input variables. For this reason, the usual method of applying a series of cuts to the most discriminating variables is substituted by a single cut on the BDT output, selecting only events with positive values of this variable.

2. Fit method.

The shape of the BDT output variable is expected to be greatly influenced by the signal in the positive side. Assuming the simulation of all backgrounds is correct, the signal can be rescaled to best fit the data, thus giving a data-driven estimation of expected signal events.

The following input variables were considered:

- Highest TCHE value in the event.
- Second highest TCHE value in the event.
- p_T of jet with highest TCHE value.
- p_T of jet with second highest TCHE value.
- Highest p_T of all leptons.
- p_T of pair of most energetic leptons.
- m_{ll} , invariant mass of pair of most energetic leptons.
- E_T^{miss} , missing transverse energy.
- M_T , transverse mass of all leptons, jet and E_T^{miss} , reconstructed using only the transverse component of the momenta.

And additionally, in the $t\bar{t}W$ analysis with three leptons:

- Highest p_T of the two same-sign leptons in the event.
- Second highest p_T of the two same-sign leptons in the event.

The cross section of a process is calculated as

$$\sigma = \frac{N_{\text{data}} - N_{\text{background}}}{\varepsilon \mathcal{L}} \quad (9)$$

where $N_{\text{background}}$ is estimated from Monte Carlo simulations, \mathcal{L} is the integrated luminosity (19.5 fb⁻¹) and ε is the selection efficiency, i.e. the fraction of simulated signal events that pass the selection cuts.

For the fit method, the cross section is calculated instead as

$$\sigma = \frac{N_{\text{signal}}^{\text{fitted}}}{\varepsilon \mathcal{L}} \quad (10)$$

since the method is meant to directly provide the number of signal events after rescaling the corresponding simulation.

An estimation of systematic uncertainties affecting the cross section measurement was considered to be beyond the scope of this work and only the statistical and luminosity (and fit, where applicable) components of the total uncertainty have been calculated.

7.1 $t\bar{t}$

For the $t\bar{t}$ analysis, the Boosted Decision Trees were trained using only tW , WW and Drell-Yan processes, being the most prominent background sources within the already exclusive selection. The number of trees was chosen to be 500, their maximum depth was limited to 3 layers and the learning parameter β was set at 0.1 to help deal with overtraining issues.

The distributions of all variables for the $t\bar{t}$ analysis selection, with opposite-sign $e\mu$ plus two or more jets, are shown in figures 18 and 19. The importance of these variables in the training, as provided by TMVA, is listed in table 3.

The reference values to compare the measurements presented here are the theoretical estimation [25]

$$\sigma_{t\bar{t}}^{\text{theo}} = 252.9_{-8.6}^{+6.4} (\text{scale}) \pm 11.7 (\text{PDF} + \alpha_s) \text{ pb} \quad (11)$$

and the experimental measurement [26]

$$\sigma_{t\bar{t}}^{\text{exp}} = 239 \pm 2 (\text{stat.}) \pm 11 (\text{syst.}) \pm 6 (\text{lumi.}) \text{ pb}, \quad (12)$$

both assuming a top quark mass of 172.5 GeV/ c^2 .

7.1.1 Cut on BDT output

The output of the BDT trained for the $t\bar{t}$, having selected events with a BDT value greater than 0, is shown in figure 20a. With this method, the cross section was measured to be

$$\sigma_{t\bar{t}}^{\text{cut}} = 230.9 \pm 1.3 (\text{stat.}) \pm 10.2 (\text{lumi.}) \text{ pb}, \quad (13)$$

slightly lower but still compatible with the theoretical prediction and published measurements.

Variable	Importance
Highest TCHE	0.2091
M_T	0.1538
$p_T^{\ell\ell}$	0.1308
E_T^{miss}	0.1200
$m_{\ell\ell}$	0.1096
p_T of jet with 2nd highest TCHE	0.09551
p_T of jet with highest TCHE	0.07437
Second highest TCHE	0.07311
$p_T^{\ell,\text{max}}$	0.03375

Table 3: Variable importance in the training of the BDT for $t\bar{t}$.

7.1.2 Fit signal to BDT output

Alternatively, the full output of the BDT was used to fit the $t\bar{t}$ simulation to the data by rescaling it while preserving its shape. The distributions of the BDT output after this fitting is shown in figure 20b and the cross section estimated by this method is

$$\sigma_{t\bar{t}}^{\text{fit}} = 238.6 \pm 2.0 (\text{fit}) \pm 1.1 (\text{stat.}) \pm 10.5 (\text{lumi.}) \text{ pb}, \quad (14)$$

which is consistent with previous estimations.

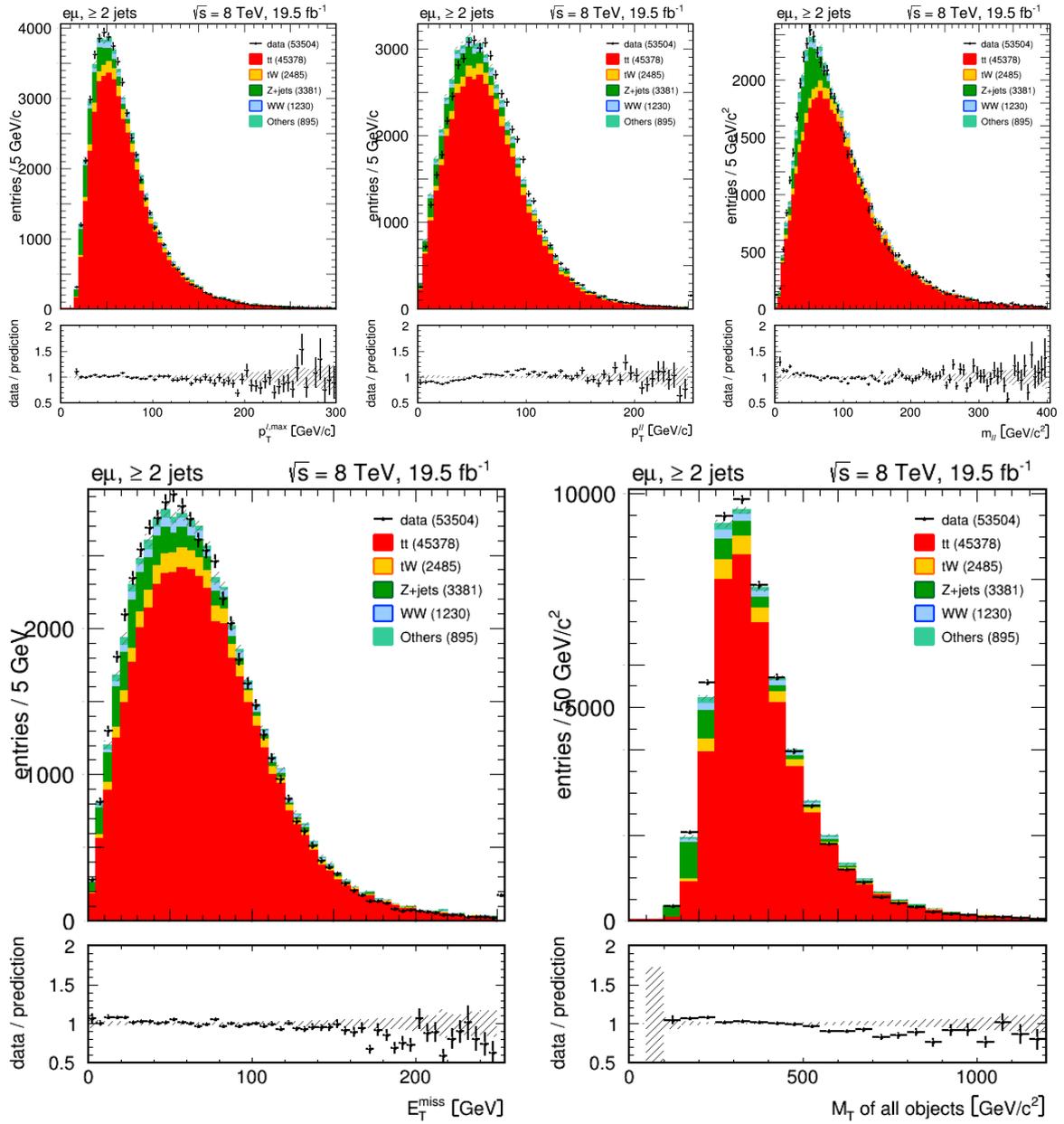


Figure 19: Input variables unrelated to the TCHE value of jets, for the $t\bar{t}$ analysis selection.

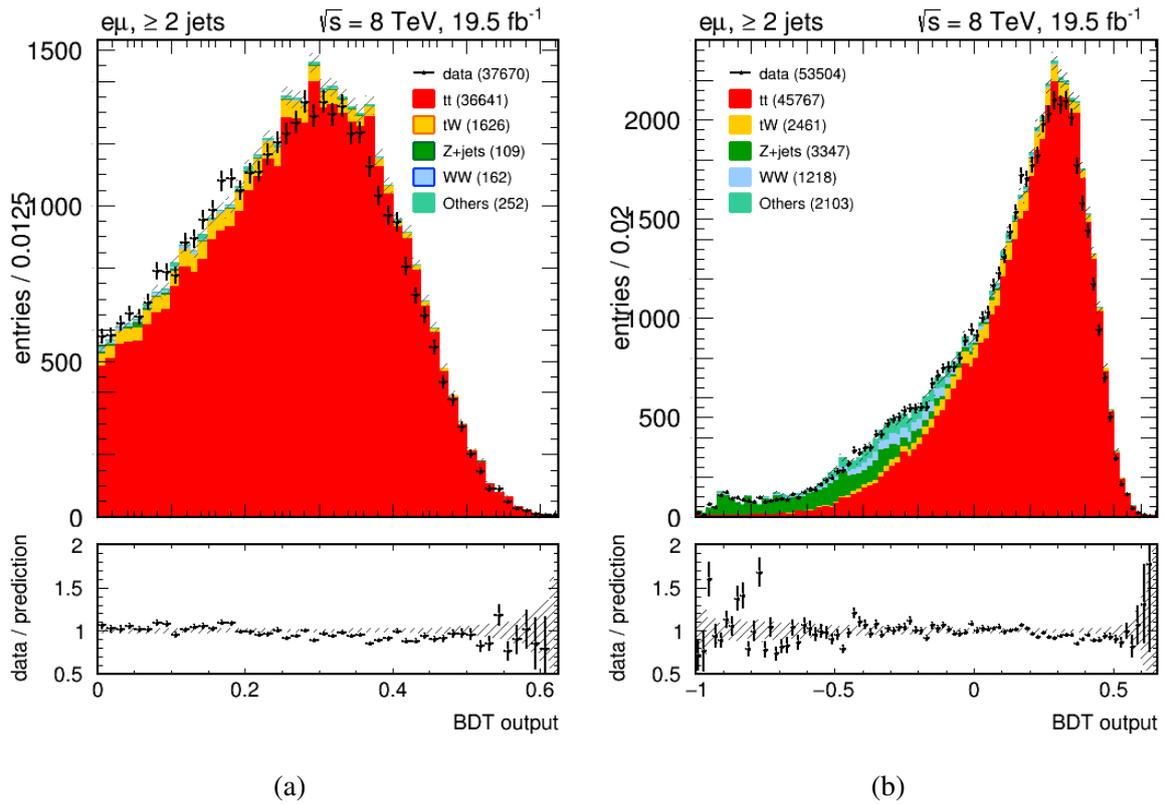


Figure 20: (a) Final result after selecting positive-valued events from the BDT output for $t\bar{t}$.
 (b) Result of fitting the $t\bar{t}$ simulation to match the data.

7.2 $t\bar{t}W$

In the $t\bar{t}W$ analysis, all of the available processes, listed in table 2, were considered for the training. An ensemble of 5000 trees was grown, each with a maximum depth of only 2 layers and a β parameter of 0.01, required by the small number of events left after the selection.

The distributions of the variables in the $t\bar{t}W$ analysis are shown in figures 21 and 22. The importance of these variables is listed in table 4.

The theoretical estimation for the cross section at 8 TeV is [27]

$$\sigma_{t\bar{t}W}^{\text{theo}} = 232 \pm 67 \text{ (scale)} \pm 30 \text{ (PDF)} \text{ fb} \quad (15)$$

while the most recent experimental measurement, combining the same-sign dilepton channel and the trilepton channel, is [28]

$$\sigma_{t\bar{t}}^{\text{exp}} = 382_{-102}^{+117} \text{ fb}, \quad (16)$$

which suggests a possible excess over the theoretical prediction.

Variable	Importance
Lowest p_T of the two same-sign leptons	0.1668
Second highest TCHE	0.1402
$m_{\ell\ell}$	0.1132
p_T of jet with 2nd highest TCHE	0.1099
E_T^{miss}	0.1017
$p_T^{\ell\ell}$	0.07694
Highest TCHE	0.07639
Highest p_T of the two same-sign leptons	0.06874
M_T	0.06342
p_T of jet with highest TCHE	0.04928
$p_T^{\ell,\text{max}}$	0.03337

Table 4: Variable importance in the training of the BDT for $t\bar{t}Z$.

7.2.1 Cut on BDT output

The result of applying the BDT trained for the $t\bar{t}W$ analysis to the corresponding selection of events and rejecting those with a BDT output lower than 0 is presented in figure 23a. The final number of total events is extremely low due to the strict requirements imposed to obtain a significant fraction of signal events.

The cross section measurement is

$$\sigma_{t\bar{t}W}^{\text{cut}} = 370 \pm 200 \text{ (stat.)} \pm 20 \text{ (lumi.)} \text{ fb}, \quad (17)$$

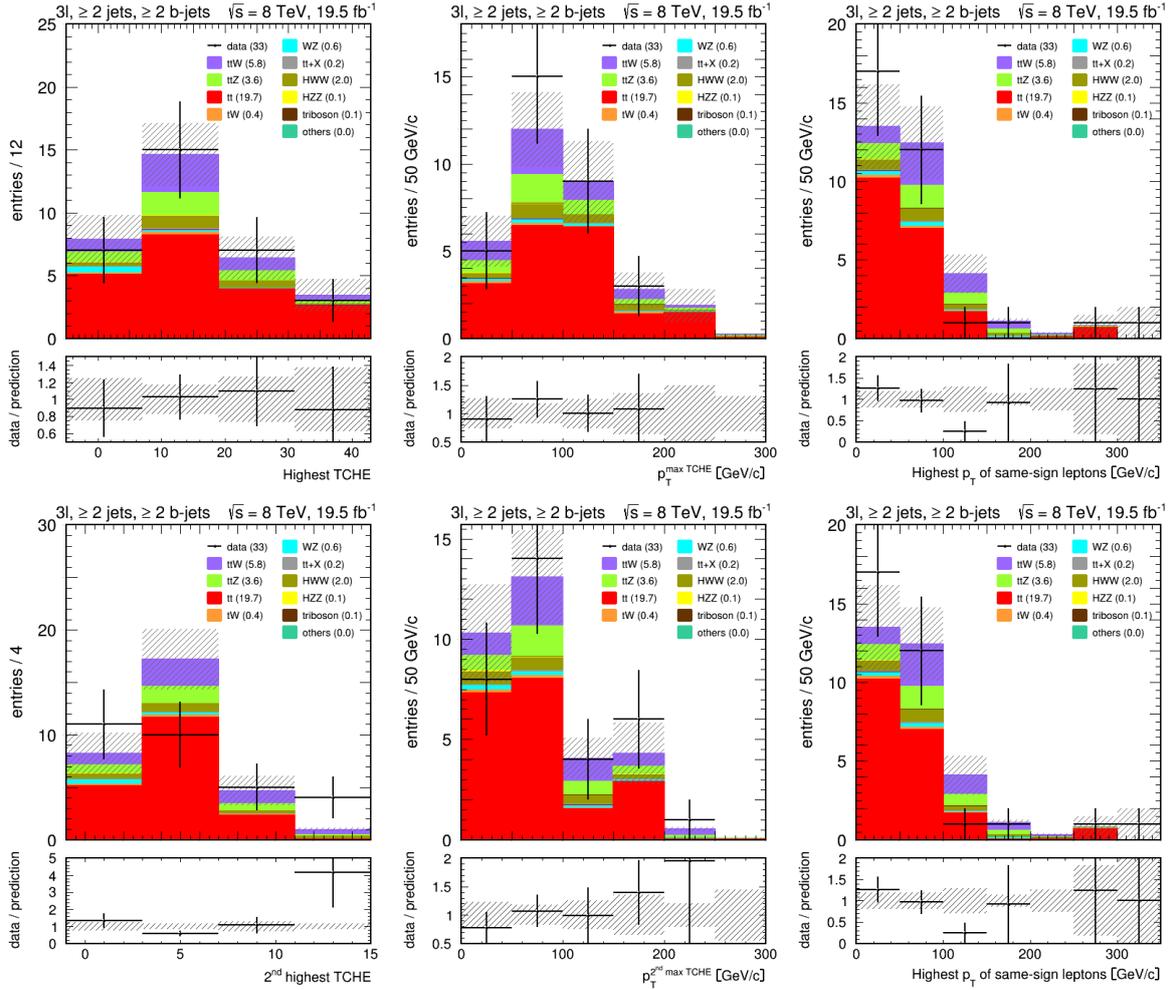


Figure 21: Input variables related to the TCHE value of jets, as well as the transverse momentum of the two same-sign leptons in the event, for the $t\bar{t}W$ analysis selection.

with an uncertainty that is large enough for the measurement to be fully compatible with both the theoretical prediction and the excess hypothesis. However, it can be considered as an observation of a low cross section process and a strong step towards a precise measurement in the future.

7.2.2 Fit signal to BDT output

After applying the fit method, the cross section was measured with even larger uncertainty:

$$\sigma_{t\bar{t}W}^{\text{fit}} = 330 \pm 270 \text{ (fit)} \pm 110 \text{ (stat.)} \pm 10 \text{ (lumi.) fb} \quad (18)$$

Nevertheless, figure 23b illustrates the observation previously noted, as the shape of the data would be difficult to explain without considering $t\bar{t}W$ production.

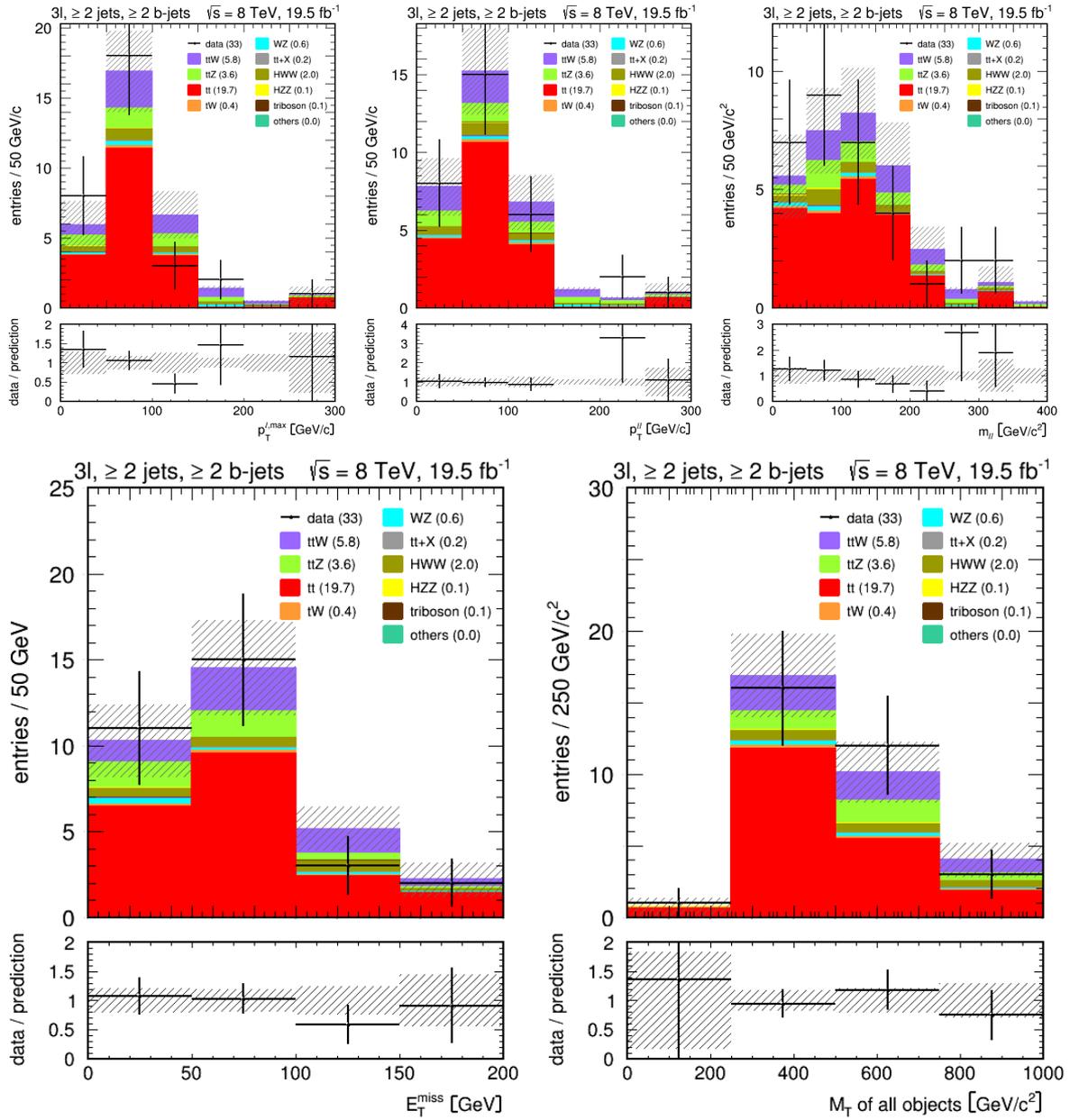


Figure 22: Input variables unrelated to the TCHE value of jets, for the $t\bar{t}W$ analysis selection.

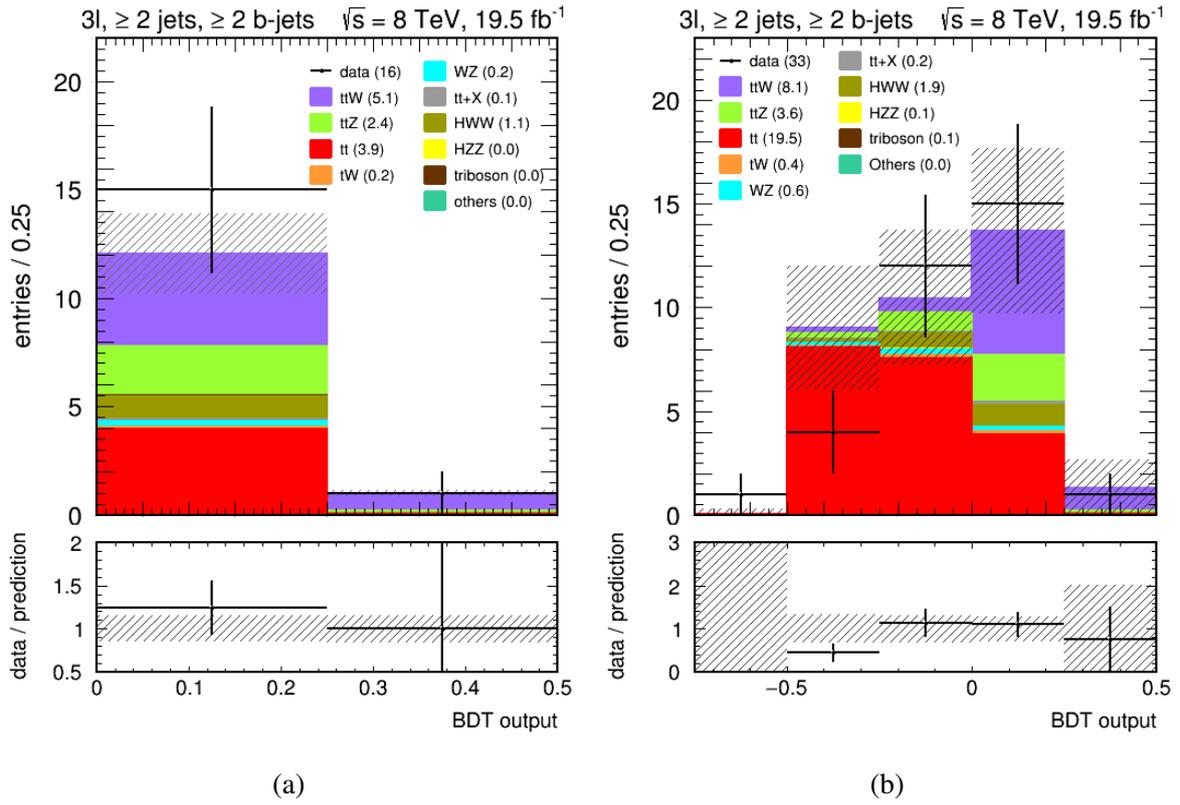


Figure 23: (a) Final result after selecting positive-valued events from the BDT output for $t\bar{t}W$.
 (b) Result of fitting the $t\bar{t}W$ simulation to match the data.

7.3 $t\bar{t}Z$

Like in the $t\bar{t}W$ analysis, all processes with events that passed the selection were used in the training. The number of trees was extended to 10000, with up to 4 layers each, and the β parameter was lowered to 0.001, as larger values led to quick overfitting due to the number of events being even lower, specially for $t\bar{t}Z$ events.

The distributions of the variables in the $t\bar{t}Z$ analysis are presented in figures 24 and 25, whereas their importance in the training is listed in table 5.

The theoretical prediction for the cross section is [29]

$$\sigma_{t\bar{t}Z}^{\text{theo}} = 205.7_{-8.6}^{+6.4} \text{ (scale) fb} \quad (19)$$

while the most recent experimental measurement, which combines the opposite-sign dilepton, trilepton and four-lepton channel, is [28]

$$\sigma_{t\bar{t}}^{\text{exp}} = 242_{-55}^{+65} \text{ fb}, \quad (20)$$

with a value slightly above the theoretical prediction, though compatible within uncertainty limits.

Variable	Importance
E_T^{miss}	0.2038
$p_T^{\ell, \text{max}}$	0.1568
M_T	0.1489
p_T of jet with 2nd highest TCHE	0.1484
p_T of jet with highest TCHE	0.1168
Highest TCHE	0.1009
$p_T^{\ell\ell}$	0.0689
$m_{\ell\ell}$	0.0472
Second highest TCHE	0.0084

Table 5: Variable importance in the training of the BDT for $t\bar{t}Z$.

7.3.1 Cut on BDT output

After the selection using the BDT output values, as presented in figure 26a, the cross section measurement is

$$\sigma_{t\bar{t}Z}^{\text{cut}} = 190 \pm 150 \text{ (stat.)} \pm 10 \text{ (lumi.) fb}, \quad (21)$$

which may again be considered as an observation of a process with low cross section which is consistent with theoretical predictions and published measurements of the CMS collaboration.

7.3.2 Fit signal to BDT output

Finally, the measured cross section after performing the fit shown in figure 26b is

$$\sigma_{iZ}^{\text{fit}} = 200 \pm 190 \text{ (fit)} \pm 110 \text{ (stat.)} \pm 10 \text{ (lumi.) fb}, \quad (22)$$

with an even larger uncertainty, once more derived from the restrictiveness of the selection.

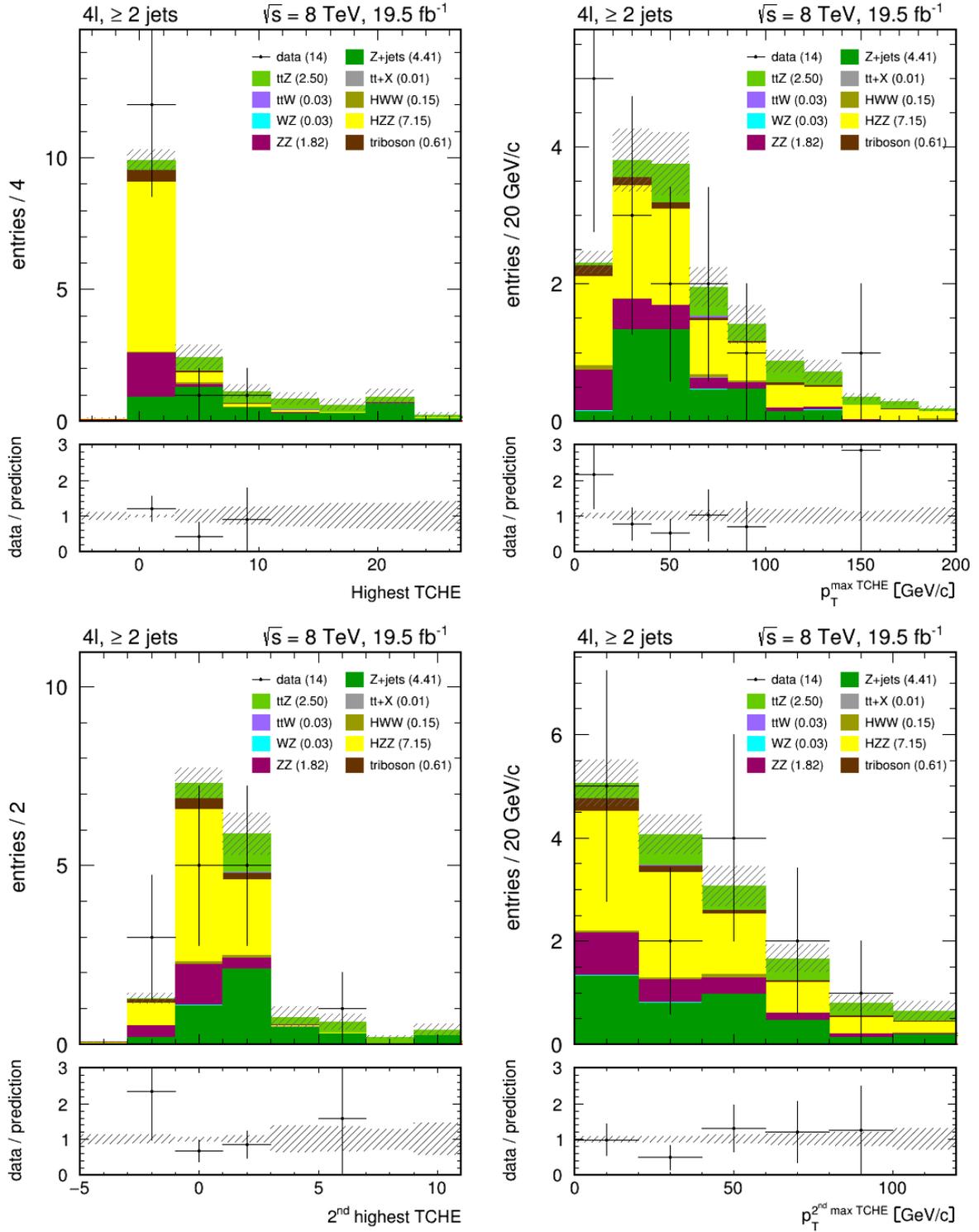


Figure 24: Input variables related to the TCHE value of jets, for the $t\bar{t}Z$ analysis selection.

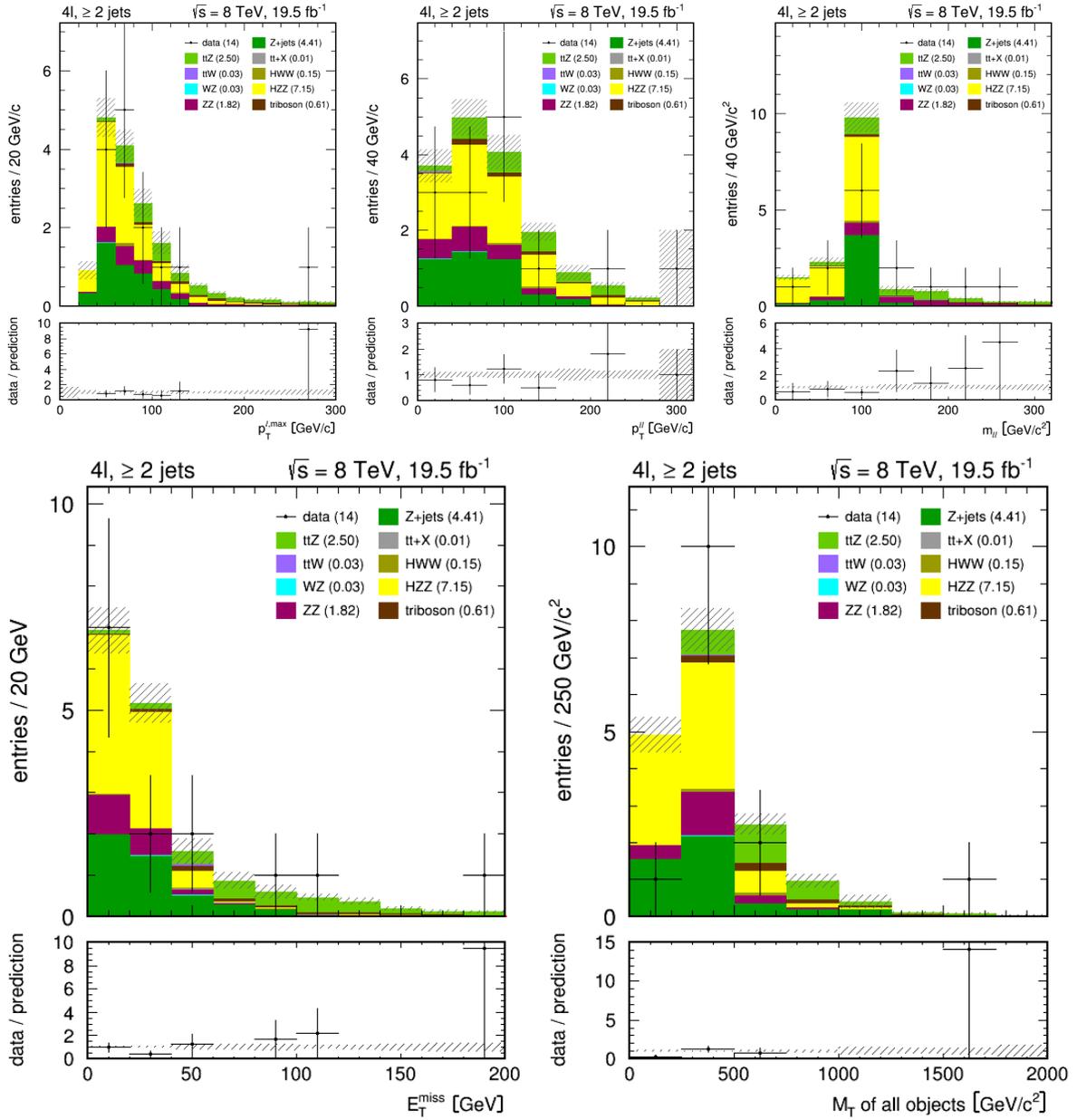


Figure 25: Input variables unrelated to the TCHE value of jets, for the $t\bar{t}Z$ analysis selection.

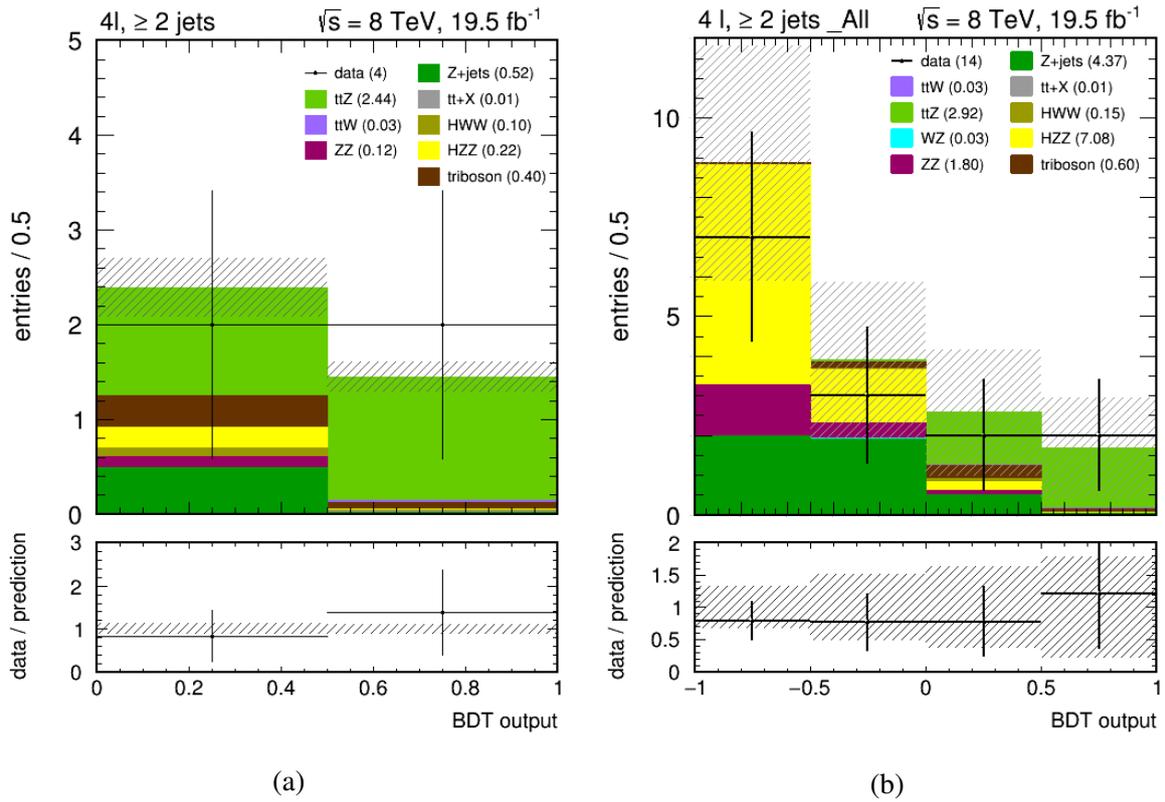


Figure 26: (a) Final result after selecting positive-valued events from the BDT output for $t\bar{t}Z$.
 (b) Result of fitting the $t\bar{t}Z$ simulation to match the data.

8 Conclusions

The Multivariate Analysis method known as Boosted Decision Trees has been thoroughly described from its basic principles: decision trees and the AdaBoost algorithm. This method, chosen for its relative simplicity, has been applied for signal discrimination in the analysis of three processes related to top quark production: $t\bar{t}$ with two different-flavour leptons in the final state, $t\bar{t}W$ with three leptons and $t\bar{t}Z$ with four leptons.

The output values of the Boosted Decision Trees have been interpreted both as a combined discriminating variable and as a function of the data that can be used to rescale signal simulations.

The $t\bar{t}$ cross section has been measured, obtaining values which are consistent with theoretical predictions and officially published measurements of the CMS collaboration:

$$\sigma_{t\bar{t}}^{\text{cut}} = 230.9 \pm 1.3 \text{ (stat.)} \pm 10.2 \text{ (lumi.) pb}, \quad (23)$$

$$\sigma_{t\bar{t}}^{\text{fit}} = 238.6 \pm 2.0 \text{ (fit)} \pm 1.1 \text{ (stat.)} \pm 10.5 \text{ (lumi.) pb}, \quad (24)$$

On the other hand, a clear observation of $t\bar{t}W$ and $t\bar{t}Z$, two processes whose cross section is about 1000 times smaller, was successfully made. The values obtained for the $t\bar{t}W$ cross section are

$$\sigma_{t\bar{t}W}^{\text{cut}} = 370 \pm 200 \text{ (stat.)} \pm 20 \text{ (lumi.) fb} \quad (25)$$

and

$$\sigma_{t\bar{t}W}^{\text{fit}} = 330 \pm 270 \text{ (fit)} \pm 110 \text{ (stat.)} \pm 10 \text{ (lumi.) fb}, \quad (26)$$

while for $t\bar{t}Z$, the results were

$$\sigma_{t\bar{t}Z}^{\text{cut}} = 190 \pm 150 \text{ (stat.)} \pm 10 \text{ (lumi.) fb} \quad (27)$$

and

$$\sigma_{t\bar{t}Z}^{\text{fit}} = 200 \pm 190 \text{ (fit)} \pm 110 \text{ (stat.)} \pm 10 \text{ (lumi.) fb}, \quad (28)$$

all of them compatible with theoretical predictions and recent measurements, although having large uncertainties due to the small number of events in their selection channels, with three and four leptons in the final state respectively.

The fact that this estimation was possible indicates the potential usefulness of MVA methods for more refined analyses aimed at obtaining precise measurements of low cross section processes.

9 References

- [1] W. Beenakker, R. Hopker, T. Plehn, and P. Zerwas, “Stop decays in SUSY QCD,” *Z.Phys.* **C75** (1997) 349–356, arXiv:hep-ph/9610313 [hep-ph].
- [2] A. Pompos, *Search for the supersymmetric partner of the top quark in dilepton events produced in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV*. PhD thesis, Purdue University, 2002. FERMILAB-THESIS-2002-46.
- [3] K. Whisnant, J.-M. Yang, B.-L. Young, and X. Zhang, “Dimension-six CP conserving operators of the third family quarks and their effects on collider observables,” *Phys.Rev.* **D56** (1997) 467–478, arXiv:hep-ph/9702305 [hep-ph].
- [4] E. L. Berger, Q.-H. Cao, and I. Low, “Model Independent Constraints Among the Wtb , Zb anti- b , and Zt anti- t Couplings,” *Phys.Rev.* **D80** (2009) 074020, arXiv:0907.2191 [hep-ph].
- [5] E. Malkawi and C. Yuan, “A Global analysis of the top quark couplings to gauge bosons,” *Phys.Rev.* **D50** (1994) 4462–4477, arXiv:hep-ph/9405322 [hep-ph].
- [6] R. Röntsch and M. Schulze, “Constraining couplings of top quarks to the Z boson in $t\bar{t} + Z$ production at the LHC,” *JHEP* **1407** (2014) 091, arXiv:1404.1005 [hep-ph].
- [7] L. Evans and P. Bryant, “LHC Machine,” *JINST* **3** (2008) S08001.
- [8] CMS Collaboration, S. Chatrchyan *et al.*, “The CMS experiment at the CERN LHC,” *JINST* **3** (2008) S08004.
- [9] CMS Collaboration, G. Bayatian *et al.*, *CMS Physics Technical Design Report, volume I: Detector Performance and Software*. 2006. CERN-LHCC-2006-001, CMS-TDR-008-1.
- [10] CMS Collaboration, G. Bayatian *et al.*, *CMS Physics Technical Design Report, volume II: Physics Performance*. 2007. CERN-LHCC-2006-021, CMS-TDR-008-2.
- [11] CMS Trigger and Data Acquisition Group Collaboration, W. Adam *et al.*, “The CMS high level trigger,” *Eur.Phys.J.* **C46** (2006) 605–667, arXiv:hep-ex/0512077 [hep-ex].
- [12] CMS Collaboration, F. Beaudette, “The CMS Particle Flow Algorithm,” arXiv:1401.8155 [hep-ex].
- [13] CMS Collaboration, S. Chatrchyan *et al.*, “Identification of b-quark jets with the CMS experiment,” *JINST* **8** (2013) P04013, arXiv:1211.4462 [hep-ex].
- [14] B. Lewis, “A new target reconstruction and target scatter algorithm,” 2006. Experiment 949, Technical Note K-063.

- [15] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, “MadGraph 5 : Going Beyond,” *JHEP* **1106** (2011) 128, arXiv:1106.0522 [hep-ph].
- [16] P. Nason, “A New method for combining NLO QCD with shower Monte Carlo algorithms,” *JHEP* **0411** (2004) 040, arXiv:hep-ph/0409146 [hep-ph].
- [17] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method,” *JHEP* **0711** (2007) 070, arXiv:0709.2092 [hep-ph].
- [18] T. Sjostrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual,” *JHEP* **0605** (2006) 026, arXiv:hep-ph/0603175 [hep-ph].
- [19] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*. Springer, 2013.
- [20] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *Journal of Computer and System Sciences* **55** no. 1, (August, 1997) 119–139.
- [21] Y. Freund and R. E. Schapire, “A short introduction to boosting,” *Journal of Japanese Society for Artificial Intelligence* **14** no. 5, (September, 1999) 771–780.
- [22] B. P. Roe, H.-J. Yang, and J. Zhu, “Boosted decision trees, a powerful event classifier,” in *Statistical problems in particle physics, astrophysics and cosmology*, pp. 139–142, Oxford University. Imp. Coll. Pr., 2006.
- [23] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, “TMVA: Toolkit for Multivariate Data Analysis,” *PoS ACAT* (2007) 040, arXiv:physics/0703039.
- [24] R. Brun and F. Rademakers, “ROOT - an object oriented data analysis framework,” in *Proceedings of AIHENP’96 Workshop*, pp. 81–86, Nucl. Inst. & Meth. in Phys. Res. A 389. 1997. See also <http://root.cern.ch/>.
- [25] M. Czakon and A. Mitov, “Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders,” *Comput.Phys.Commun.* **185** (2014) 2930, arXiv:1112.5675 [hep-ph].
- [26] CMS Collaboration, S. Chatrchyan *et al.*, “Measurement of the $t\bar{t}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s} = 8$ TeV,” *JHEP* **1402** (2014) 024, arXiv:1312.7582 [hep-ex].
- [27] J. M. Campbell and R. K. Ellis, “ $t\bar{t}W^{+-}$ production and decay at NLO,” *JHEP* **1207** (2012) 052, arXiv:1204.5678 [hep-ph].
- [28] CMS Collaboration, “Observation of a Z and measurement of a W boson produced in association with a top quark pair in events with multiple leptons at CMS.” CMS-TOP-14-021.

- [29] M. Garzelli, A. Kardos, C. Papadopoulos, and Z. Trocsanyi, “ $t\bar{t}W^\pm$ and $t\bar{t}Z$ Hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects,” *JHEP* **1211** (2012) 056, arXiv:1208.2665 [hep-ph].