UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

Data-driven background estimation study for stealth supersymmetry searches in events with structure-rich jets and low missing energy in pp collisions at 8 TeV

Particle Physics

Proyecto de investigación

Santiago Rafael Paredes Sáenz
Física

Trabajo de titulación presentado como requisito para la obtención del título de

Licenciatura en Física

Quito, 21 de diciembre de 2015

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

HOJA DE CALIFICACIÓN DE TRABAJO DE TITULACIÓN

Data-driven background estimation study for stealth supersymmetry searches in events with structure-rich jets and low missing energy in pp collisions at 8 TeV

Santiago Rafael Paredes Sáenz

Calificación:

Nombre del profesor, Título académico: Edgar Carrera, Ph.D.

Firma del profesor: _____

Quito, 22 de diciembre de 2015

Derechos de Autor

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Firma del estudiante: _____

Nombres y apellidos: Santiago Rafael Paredes Sáenz

Código: 00103811

Cédula de Identidad: 1713475893

Lugar y fecha: Quito, diciembre de 2015

RESUMEN

En vista de que las búsquedas de supersimetría concentradas en encontrar energía faltante no han sido exitosas hasta el momento, modelos alternativos cómo la super-simetría sigilosa, que permiten poca, o ninguna, energía faltante han tomado mucho interés. Una estimación experimental de eventos de fondo fue desarrollada para una señal de teoría de supersimetría sigilosa con cuatro jets en el estado final. Cómo parte de este proceso, eventos de supersimetría sigilosa fueron generados mediante herramientas computacionales y estas simulaciones fueron analizadas. Datos del experimento CMS en CERN de colisiones protón-protón a $\sqrt{s} = 8$ TeV y 21.79 fb^{-1} fueron analiadosLa concordancia entre las distribuciones de la estimación de eventos de fondo y la de los verdaderos eventos de fondo experimentales sugiere que esta técnica, y los límites en las variables que fueron definidos en este estudio, proveerían una aproximación adecuada de los eventos de fondo que simulan el decaimiento estudiado de un squark, y podría ser usado cómo parte de una búsqueda completa de supersimetría sigilosa.

Palabras clave: Modelo estándar, supersimetría, supersimetría sigilosa, sub-estructura de jets, estimación de eventos de fondo.

ABSTRACT

As searches for supersymmetry (SUSY) looking for missing energy have proven unsuccessful so far, SUSY models that allow low, or non, missing energy, like stealth supersymmetry, have acquired increasing interest. A data-driven background estimation was performed for a stealth supersymmetry signal with four jets and low missing energy in the final state. As part of this process, stealth SUSY events were generated via software and later studied. Data from the CMS experiment of pp collisions at $\sqrt{s} = 8$ TeV and 21.79 fb^{-1} were analyzed. The close agreement between the estimated background distributions and true background from data suggests that this technique, and the variable thresholds that were defined in this work, would provide an appropriate representation of the background events which mimic the squark decay under study.

Key words: Standard model, supersymmetry, stealth supersymmetry, jet substructure, background estimation.

AGRADECIMIENTOS

A mis compañeros y amigos, con quienes hemos festejado y sufrido día a día en estos años universitarios. En especial a mi familia, que con apoyo y motivación me ayudaron a seguir la carrera que amo.

A mis profesores y otros mentores, quienes con su conocimiento y enseñanzas convirtieron estos últimos años en retos y aventuras. En especial Yuri Gershtein, quién además de haber supervisado este proyecto, fue un gran guía durante mi estadía en Fermilab.

Un agradecimiento especial para mi director en este, y algunos otros, proyectos Edgar Carrera, quién con su guía y apoyo me ha ayudado a conseguir más metas de las que podría haber imaginado al iniciar mi carrera en Física.

Contents

1 Introduction and Motivation

1.1 The Standard Model of Particle Physics

Since its appearance in 1970's, the Standard Model of Particle Physics has passed every experimental test. This elegant theory describes all the elementary particles of nature, as well as how they interact with each other. Moreover, it predicted the existence of particles that had not been observed, and through exhaustive experimental procedures were later found (such is the case of the top quark and the higgs boson among others).

The Standard Model (SM) states that all matter is composed of elementary particles, which are divided into two groups: quarks and leptons. Additionally, the forces between these particles are mediated by bosons. There are three generations of quarks: up(u)-down(d), charm(c)-strange(s) and top(t)-bottom or -beauty(b); with the corresponding leptons: electron(e)-electron neutrino(ν_e), muon(μ)- muon neutrino(ν_μ) and tau(τ)-tau neutrino(ν_τ). Finally, there are four gauge bosons: the photon γ, which mediates the electromagnetic force; the gluon g, which is responsible for the strong nuclear force, and the W and Z bosons, which mediate the weak nuclear force. Separate from these is the Higgs boson(H), which is responsible for the mechanism that allows other elementary particles to have mass[1][2].

Figure 1: Periodic table of elementary particles. Image credit: Wikipedia Commons.

This model of our universe can be nicely, and compactly, summarized by the la-

grangian from which its fields and interactions are derived:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}D\psi + h.c. + \bar{\psi}_i y_{ij}\psi_j\phi + h.c. + |D_\mu\phi|^2 - V(\phi) \tag{1}$$

As a quick overview, and speaking quite roughly, the terms that contain either an $F_{\mu\nu}$ or a D (with D denoting the covariant derivative) describe gauge fields: photon, W and Z bosons and gluon. The terms containing a ψ describe matter particles, and the ones with ϕ describe the higgs boson. Also note that here $h.c.$ stands for Hermitian Conjugate. The interactions between fields are described by the terms containing various of these symbols[2]. Over and over again have scientists proven that the predictions of the SM are correct, and that it does indeed describe our universe. Nevertheless, despite the elegance and simplicity of the SM, we know that it is an incomplete model, as there are phenomena in nature that it cannot describe. The hierarchy problem and the existence of dark matter are two notable unresolved issues in physics, which the SM, in its current state, cannot explain. Furthermore, the SM is a precise tool to describe the universe and even make predictions about it, but the theory itself raises a few questions: why are there three generations of quarks and leptons even though just the first generation is necessary to create all ordinary stable matter? What sets the scale of the Higgs boson mass? This type of questions is precisely what is most interesting for scientists, since their solutions might shed light over new physics.

1.2 Limitations of the Standard Model

Two of these unsolved problems, dark matter and the hierarchy problem, are specially relevant to this work, hence, a closer look at them is in order.

1.2.1 Dark Matter

In a galaxy, stars orbit around its center due to its gravitational potential, which is given by its distribution of mass. In the same way, galaxies orbit around the center of clusters due to the overall distribution of mass of the galaxy cluster. On the other hand, astronomers can determine the speed at which galaxies orbit by measuring the Doppler shift of the light they emit [3]. Hence, by measuring the orbiting speeds of galaxies in a cluster, it is possible to estimate the mass present in the cluster, as well

as obtain clues about its distribution.

Using this method, in 1933 Fritz Zwicky studied the Coma Cluster and found that the orbiting speeds of galaxies did not follow the pattern expected from the distribution of masses of all its stars. In fact, he determined that the total mass in the cluster was around 400 times more than the mass of all the stars. This large amount of unaccounted-for mass, known only by its gravitational effects[1], was appropriately named Dark Matter [3].

Since Zwicky's observations, astronomers have studied the rotation curves of galaxies (which depict the orbiting speed of a galaxy as a function of the distance r to its center). Judging by the fact that most of the mass of a galaxy is in its center, its orbiting speed is expected to decay as $\frac{1}{\sqrt{r}}$. Surprisingly, in most cases this speed increases with the radius, hence suggesting that the distribution of matter is closer to a spherical halo around the galaxy instead of the disk-shaped structure that the stars in a galaxy show. This evidence suggests that there is dark matter around the galaxy itself extending far outside its nucleus.

The SM cannot account for that many virtually non-interacting particles, consequently scientists now believe that dark matter is built from fairly massive particles known as WIMPs (Weakly Interacting Massive Particles, with estimated masses of around 100 - 200 GeV[2]) [3].

1.2.2 The Hierarchy Problem

In order for our universe to exist as we know it, certain values of constants of nature have to be tuned within a certain range. Some of this constants are the gravitational constant, Plank's constant and, interestingly enough, the mass of the Higgs boson.

In Quantum Field Theory it is quite common to calculate certain physical quantities using a perturbative approach called Feynman Calculus [1][4]. To calculate the mass of a particle we must take all loop order diagrams. The diagram in Figure 2 shows the propagator for the higgs boson together the first order loop correction diagram.

Roughly speaking, an integration must be done over all possible momenta of the

---

[1]It is possible that Dark Matter could interact very weakly with some other force.

[2]In the field of high-energy physics it is customary to express the mass of particles in MeV, GeV, etc. when the actual units of mass would be MeV/c^2. This is justified as it is usual take the value of the speed of light as $c = 1$.
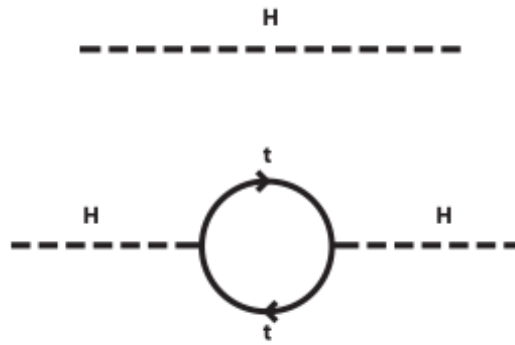
Figure 2: Feynman diagram of the first order loop for Higgs boson. The above diagram shows the propagator at tree level, while below is the one loop order correction with a top quark. Image credit: Wikipedia Commons

virtual particles involved in the loop, for example, the top and anti-top quarks in Figure 2. Note that, here, only the loop with a top quark is shown since it couples most strongly to the higgs field, but such loops can contain other particles as well. These integrals are often divergent. For fermions, these divergences can be cleverly avoided with a process known as renormalization, but when renormalizing the Higgs mass, it ranges out of the acceptable value necessary for our universe to behave as it does. An explanation to why the Higgs mass is right in the acceptable range would then be that the loop order corrections almost magically cancel each other out exactly, a circumstance known as fine tuning. Many scientists consider fine tuning a poor approach to the hierarchy problem, and argue that a much more satisfying answer would involve a mechanism that drives the fine tuning, which allows the cancellation of all the corrections.[1][5]

1.3   Supersymmetry

The symmetry principles from which the SM is derived require that the theory is unchanged for different states of a same system. Supersymmetry differs from classic symmetries because it introduces a more radical symmetry between matter fields, spin 1/2 Dirac fields, and gauge boson fields with spin 1; in other words, it proposes a symmetry between fermions and bosons. In general, superpartners differ by 1/2 in spin, so that 1/2 spin fermions can have 1 spin SUSY partners, and bosons with spin 1 could have superpartners with spin 3/2. Wess' and Zumino's model[6] (proposed in 1974)

is considered the first theory to apply the principle of Supersymmetry (SUSY). This principle poses an elegant solution to the issues discussed above and many others, not to mention that it allows for the development of theories to achieve the highly sought-for unification of gravity with the other three fundamental forces. In supersymmetry, for each fermion, its partner is denoted by adding an "s" prefacing their name. A squark would be the partner of a quark, a slepton of a lepton, a stop of a top, etc. On the other hand, the partner of a boson has the ending "ino" after the boson's name. It follows that the partners for the Higgs, the gluon and the photon would be the Higgsino, gluino and photino, respectively.
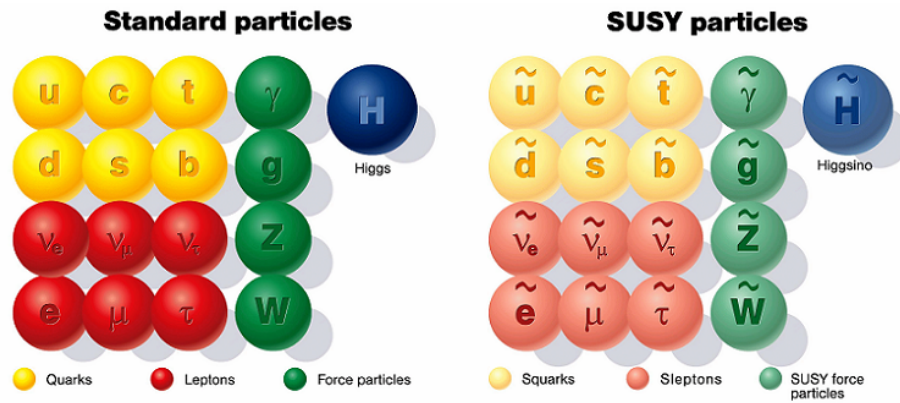

Figure 3: Proposed periodic table of elementary particles including supersymmetric partners. Image credit: DESY Lab. Hamburg, Germany.

If our universe is in fact supersymmetric, we would have to add a few terms to the standard model lagrangian we discussed in equation 1. The simplest supersymmetry model can be represented by:

$$\mathcal{L}_{SUSY} = -\partial^\mu \phi^{*i} \partial_\mu \phi_i + i\psi^{\dagger i} \bar{\sigma}^\mu \partial_\mu \psi_i + F^{*i} F_i \tag{2}$$

Here, the standard model fields are denoted by lower indexes: ϕ_i and ψ_i; while the supersymmetry terms carry raised indices: ϕ^{*i} or $\psi^{\dagger i}$.

We know that if our universe is indeed supersymmetric, SUSY has to be a spontaneously broken symmetry [7]. If it was a perfect symmetry, the masses of each elementary particle would be the same as their partner's. The fact that we have only observed standard particles (as opposed to SUSY particles) indicates that the symmetry is broken and the SUSY partners are a lot more massive than their standard

counterparts.

By adding this new set of particles, SUSY theories can provide elegantly simple solutions to some of the biggest mysteries in physics. These elusive particles could well be what constitute the also elusive dark matter. The lightest supersymmetric particle, or LSP, in most of the proposed theories, is colorless, neutral and stable, which are the exact qualities of the WIMPs which are thought to compose dark matter [3]. Moreover, if we tackle the hierarchy problem by including SUSY particles, for each loop order correction there is another exact contribution with opposite sign (due to the SUSY partner), which cancels out the divergences precisely.
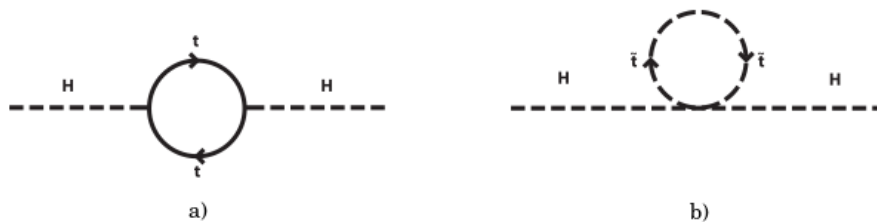

Figure 4: The first order loop correction for the Higgs boson mass on a) is exactly canceled by the contribution of the supersymmetric partner b). Image credit: Wikipedia Commons

The most common method to search for SUSY particles in collider experiments is to analyze events where there is a large amount of missing transverse energy, E_T^{miss} for short (defined as $\vec{E}_T^{miss} = -\sum_i \vec{p}_t(i)$). Since the particles in the beams have little momentum in the direction perpendicular to the beam, after the collision we expect that the vectorial sum of transverse momentum of all particles in the event will add up to zero, unless a particle was missed by the detector (such is the case for neutrinos for example). MET is then defined as. This technique assumes that the SUSY particles are very weakly interacting, so they escape the detector unperturbed. This assumption is true for most supersymmetric theories. Nevertheless, there are other SUSY theories that do not require large amounts of E_T^{miss}, which need to be studied with a different approach.

As elegant as Supersymmetry is, the disappointing reality for physicists is that, to this day, no evidence of this principle has been observed in nature. Colliders like the Tevatron at Fermilab and the Large Hadron Collider (LHC) at CERN have put such theories to the test, without success so far. Experiments have disproved various super-

symmetric theories, but SUSY as a principle still stands, hence the research interest is being drifted towards SUSY models which require low E_T^{miss}, among other conditions.

1.4   Stealth Supersymmetry

Stealth SUSY introduces a new hidden sector of particles. The lightest hidden sector supersymmetric partner (LHSP) decays into its SM partner and the usual SUSY LSP; thus, the LSP doesn't necessarily carry off a large momentum, which could yield events with low E_T^{miss} in the final state. In this context, SM "visible-sector" particles are referred to as lightest visible supersymmetric partner or LVSP.

For this study we consider a model which contains an initial squark $\tilde{q}$, a bino-like particle named neutralino, represented by $\widetilde{\chi}_1^0$; a hidden sector containing a singlet state S, and its superpartner $\widetilde{S}$, the singlino. The model can be characterized by the mass of these components[8][9].

A simple example of a gluino decay in the frame of stealth supersymmetry is shown in Figure 5. Among other alternative low-E_T^{miss} SUSY models, stealth supersymmetry is characterized by providing a natural scenario for SUSY with artificial tunning.
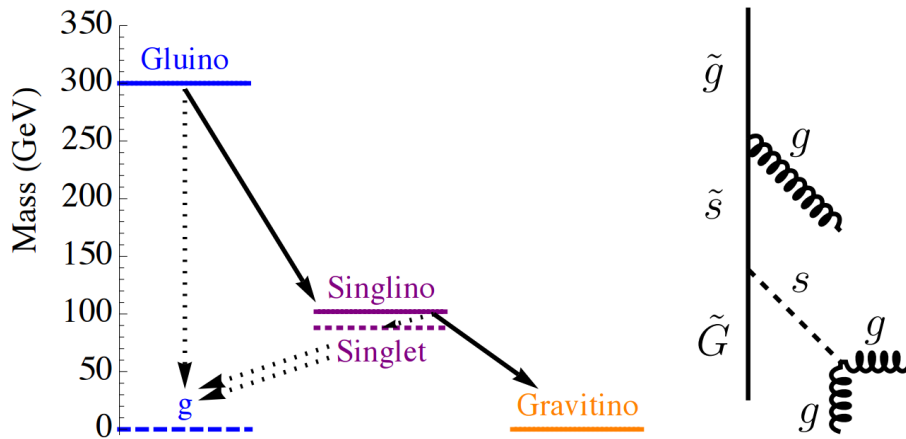

Figure 5: An example spectrum and decay chain for Stealth SUSY with gluino LVSP[8].

2 SUSY Signal and Background Processes

2.1 Signal

The analyzed decay process, depicted in Figure 6, starts with the production of a squark pair after the pp collision. Each quark decays as $\tilde{q} \to q\widetilde{\chi}_1^0$. The produced quark, as dictated by Quantum Chromodynamics (QCD) [1], decays into a shower of particles known as a jet, in a process known as hadronization. The neutralino then decays into the hidden sector particle $\widetilde{S}$ by $\widetilde{\chi}_1^0 \to \gamma\widetilde{S}$; the singlino $\widetilde{S}$ in turn decays into a gravitino $\widetilde{G}$ and the hidden sector singlet particle S, the latter of which decays into two gluons that hadronize into jets.
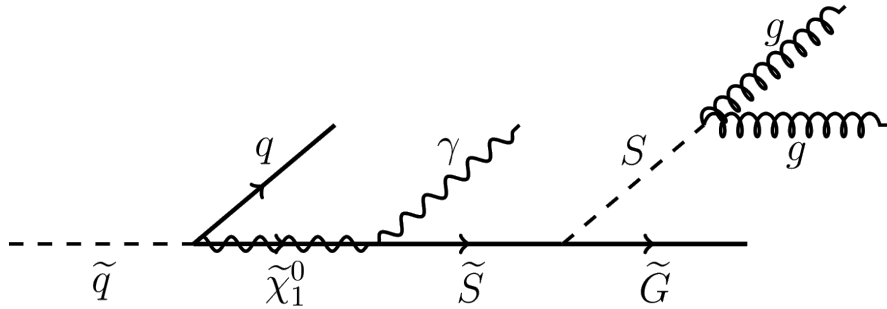

Figure 6: Feynman diagram of the decay process $\tilde{q} \to q\widetilde{\chi}_1^0$, followed by $\widetilde{\chi}_1^0 \to \gamma\widetilde{S}$ and subsequent decay $\widetilde{S} \to S\widetilde{G}$ and final S decay via $S \to gg$.

In this study we consider a squark of mass $M_{\tilde{q}} = 1000$ GeV. Based on the parameters described in [8], we took the masses of the $\widetilde{S}$ and S masses to be 100 GeV and 90 GeV respectively (with the small mass splitting between this two being enforced by a supersymmetry[8]), while the $\widetilde{\chi}_1^0$ mass was taken as 150 GeV, and the gravitino mass was taken to be zero . Due to the mass difference between the $\tilde{q}$ and the $\widetilde{\chi}_1^0$ in this model, the quark (and hence the jet to which it decays) could achieve fairly high momentum up to several hundreds of GeV. On the other hand, the relatively small mass difference between the $\widetilde{S}$ and S dictates that the escaping gravitino will not carry large amounts of energy (compared to the other components in this decay), thus yielding low E_T^{miss} for this process.

The final products of the decay of each squark would be a gravitino (which would show up as E_T^{miss}), a photon and three jets: one product of the first quark emitted, and two from the gluons from the singlet. Nevertheless, the large energy available

for the decay of the $\widetilde{\chi}_1^0$ will be transfered to the photon and gluons, making them boosted, meaning that they will be closely collimated due to their high momenta. The gluons then decay into jets, which have very little spacial separation (due to their high momentum), and hence merge into a single jet of particles. Depending on how the momentum is split in each vertex, even the photon can merge into the jet and go undetected. This means that, effectively, this decay would yield a gravitino, one or no photons and two jets per squark produced. As the model assumes that the *pp* collision produces a squark pair, the final state of this process will include four jets, either zero, one, or two photons, and some E_T^{miss}.

The squark and neutralino masses in this study were specifically chosen to complement the work referenced in [9]. Figure 7 shows the limits on squark and neutralino masses set by the mentioned study. Note that CMS has no constraint in the high-squark low-neutralino mass sector, even though it is accessible at the LHC. In this study we analyze events precisely in this region using new jet substructure techniques which permit to further scrutinize these events. A brief description of jet substructure can be found later on in this paper, but for a detailed description see [10];
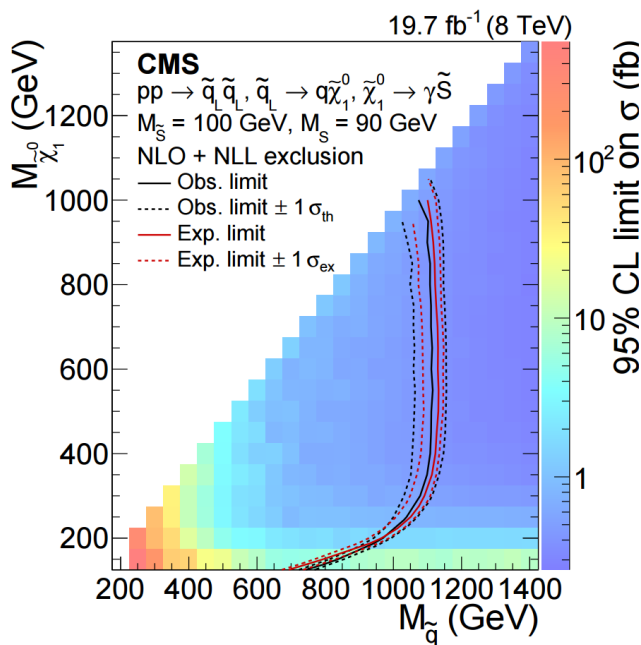

Figure 7: Figure taken from [9]. The 95% confidence level upper limits on the squark pair production cross section as a function of squark and neutralino masses from the γ analysis. The contours show the observed and median expected exclusions assuming the NLO+NLL cross sections, with their one standard deviation uncertainties.

2.2   Background Processes

If a proton-proton collision produces a squark pair, and each decays through the process described above, this event is considered a signal event. However, as we only observe the final decay products and not the intermediate states. There will be other processes that mimic signal events (i.e., which have a final state with four jets, zero, one or two photons and some E_T^{miss}); such processes are tagged as background.

As the final products in our signal consist of mostly jets and some photons, the background is dominated by QCD events with quarks and gluons hadronizing to produce jets. Moreover, it is possible that a jet is misidentified as a photon by the particle detector, which also emulates a signal event.

Being able to discern signal events from background processes is a key component of any particle collider experiment analysis. A technique commonly adopted is to use Monte Carlo models to generate a dataset of how the collisions would look if it only included the background events, and then compare it to the actual experimental data to determine if there is enough discrepancy between the experiment and background data to assure that the signal process does indeed occur.

An alternative approach, and the main objective of this study, is to estimate the background for a process using the experimental data itself. Data-driven background estimation has the advantage that it is not limited by the accuracy of the models used to generate the simulated background events. However, using data to estimate the background also poses some extra challenges, which require a crafty use of the CMS data. Thus, in order to fully explain the method used to perform data-driven background estimation, a brief description of the LHC and the CMS experiment is necessary.

3   The LHC and the CMS Detector

The Large Hadron Collider, measuring 27 kilometers in circumference, is the world's largest particle accelerator. This colossal machine is located at the French-Swiss border, and it is managed by CERN [12]. The LHC accelerates two beams of protons at a speed close to the speed of light in opposite directions around the ring, and collides the protons

at four specific locations where particle detectors are located. Each detector gathers data for a separate experiment. This data is stored in computing facilities around the world, and is analyzed by collaborations of scientists from all over the globe. A more complete description of the LHC machine can be found in [11].

CMS, short for Compact Muon Solenoid, is one of the four experiments that study the LHC's proton-proton (pp) collisions. The data used for this study was collected by the CMS experiments in the first run of the LHC, meaning that the collisions had a center of mass energy of 8 TeV. As most particle detectors, the CMS detector has a cylindrical geometry, and is located so that the collision point is around its center. From the center outwards, cylindrical layers of different particle detectors are disposed so that the particles bursting from the collision are detected in one of the layers. Figures 8 and 9 show the geometry and configuration of the CMS detector.

A key component of the detector, the one that gives the whole experiment its name, is the giant superconducting solenoid that is located inside the detector, with some components placed in the region inside it and some outside. The superconducting electromagnet provides a fairly uniform magnetic field of 3.8 T in its inside region.

From the center, the first layer is a silicone detector that tracks the trajectory of charged particles flying through it. The magnetic field in this region bends the trajectory. Using the curvature of this trajectory, the particle's momentum can be calculated. The next layer, the electromagnetic calorimeter (ECAL for short), is made of lead tungstate scintillating crystals. Here, both photons and electrons are detected via the decay of the particle and subsequent production of electron-positron pairs inside the crystal, which deposit their energy in the scintillator. Next is the hadronic calorimeter (HCAL), made of brass and plastic scintillators, which detect charged and neutral hadrons. The last piece of the detector is the muon chambers, which are the only components outside the solenoid and measure the trajectory and energy of muons traveling through them.

Data from multiple of these layers are often combined to obtain more information about the detected particles. For example, if an energy deposit in the ECAL matches a trajectory in the tracker, called a track, it is likely that the detected particle was an electron or positron (which one can be determined by the curvature in trajectory), but if the energy deposit does not match any track, it is likely that the detected particle is

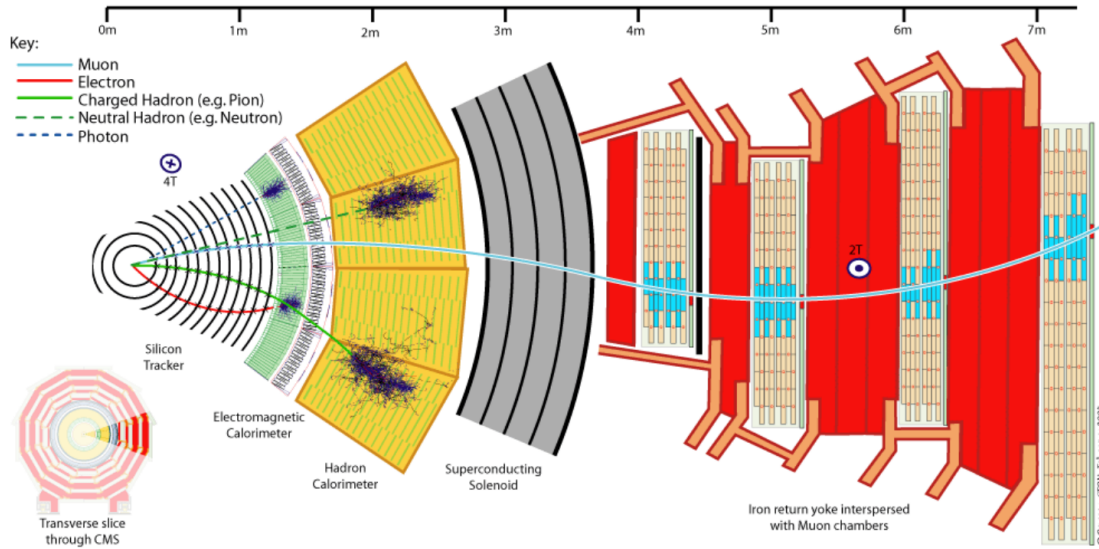a photon. A thorough description of the detector can be found in [14].

Figure 8: Diagram of a radial section of the CMS detector. Image credit: CERN, Geneva, CH.

The coordinate system used in the CMS experiment are spherical coordinates with the beam along the $\theta = 0°$, or z axis, directions. The azimuthal angle is Φ, and the polar angle is usually replaced by the pseudorapidity η[3], an equivalent quantity. A diagram of the coordinates is shown in Figure 10.

4   Analysis

4.1   Simulated Signal Events

To perform a data-driven background estimation for the stealth supersymmetry process in question, the first step was to generate a Monte Carlo sample of the signal, with squark mass equal to 1 TeV and a singlino of 150 GeV of mass. This file contains 10,000 events simulated using MadGraph[17], and the hadronization was done using Pythia 8 [4][18]. By studying this generated sample we were able to obtain the characteristics of the squark decay of interest, and based on this information we found which variables would be the most helpful in separating our signal from the SM background.

---

[3]The pseudorapidity is defined as $\eta = -ln \left[\tan \frac{\theta}{2}\right]$

[4]This Pythia version was chosen specifically for its hadronization algorithm, which is important for this analysis since the Jet Substructure variables will play a key role during background exclusion.
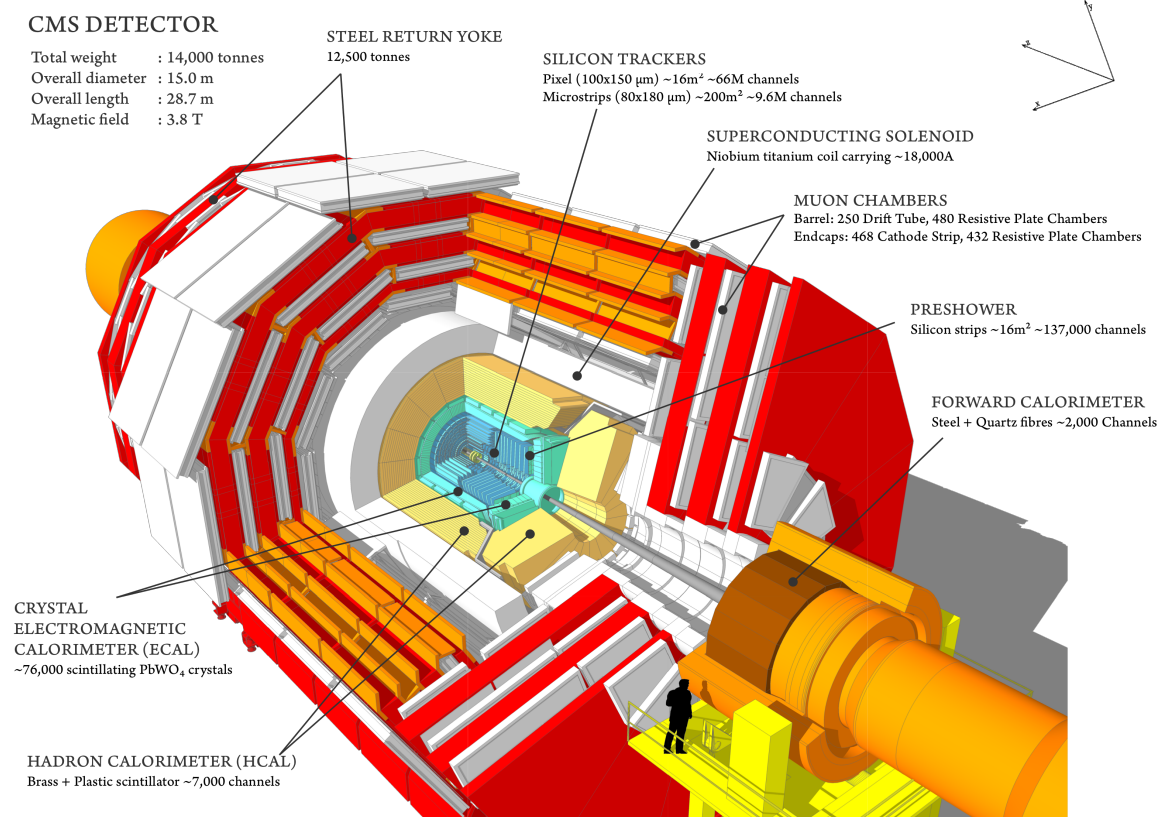
CMS DETECTOR

Total weight   : 14,000 tonnes
Overall diameter : 15.0 m
Overall length   : 28.7 m
Magnetic field   : 3.8 T

STEEL RETURN YOKE
12,500 tonnes

SILICON TRACKERS
Pixel (100x150 μm) ~16m² ~66M channels
Microstrips (80x180 μm) ~200m² ~9.6M channels

SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying ~18,000A

MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 468 Cathode Strip, 432 Resistive Plate Chambers

PRESHOWER
Silicon strips ~16m² ~137,000 channels

FORWARD CALORIMETER
Steel + Quartz fibres ~2,000 Channels

CRYSTAL
ELECTROMAGNETIC
CALORIMETER (ECAL)
~76,000 scintillating PbWO$_4$ crystals

HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator ~7,000 channels

Figure 9: Model of the CMS detector shown at scale. The different layers of detector equipment are shown.Image credit: CERN, Geneva, CH.
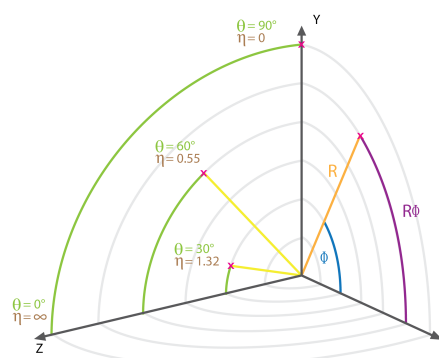
Figure 10: Coordinate system used in the CMS experiment. The x axis points parallel to the ground while the z axis is oriented in the same direction as the proton beam. [13]

This dataset was made using the CMS Remote Analysis Builder CRAB [19], which allows to perform data analysis jobs using CERN's GRID[16] computing network. As we developed this analysis mostly while working at the LHC Physics Center (LPC) at Fermilab, the scheduler CONDOR [19] was used to perform all the jobs locally at the LPC's computing facilities.

4.2   Data Selection Criteria

The first clue comes from analyzing the expected products of the squark decay. After the process depicted in Figure 6 has taken place, the detector would read a signal consisting of one jet from the emitted quark, a photon from the neutralino, and a large-area jet from both gluons. This process is mirrored on the decay of the other squark. We then require signal events to have four jets and two isolated (i.e. separated from the large jet) photons. The parameters to define an isolated photon depend on various factors such as the resolution of the detector, the energy of the collisions, etc. Thus, there are official recommendations for defining isolated photons, which are listed in [15].

We are interested in knowing how the photons that came from a neutralino are reconstructed, and if they are isolated. To find an answer to this question, we compared the data of the generated and reconstructed photons in the dataset. The event generator software creates all the particles of an event, then simulates how they interact with the detector, and finally simulates how the detector reconstructs the identity of the particle. The dataset contains variables for both the initially created particles (we call these generated), and the reconstructed from the detector simulator. We plotted histograms of the variation in η and Φ variables for each generated and reconstructed photons (shown in Figure 11), where a clear peak was observed. The width of these peaks was taken to be the parameter to match a reconstructed photon with a generated one. Knowing how many neutralino photons were properly reconstructed and passed the isolation conditions, we determined a cut of 1 isolated photon per event.

Following the kinematics of the decay for the chosen mass point, we note that the 1TeV squark decays into a quark and a neutralino. While generating these events, the neutralino was made to have mass of 150 GeV, which means that it will have
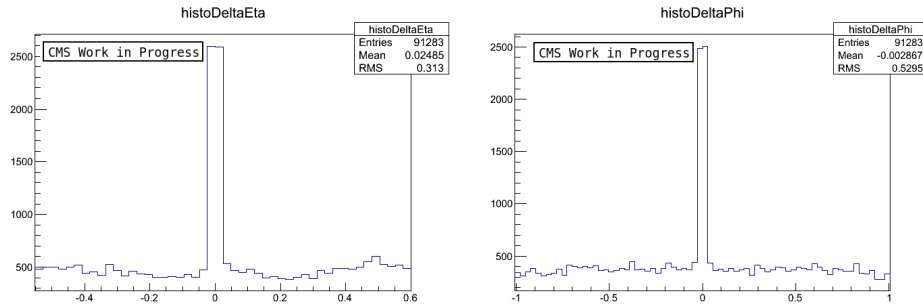
(a) Variation in η between recon-   (b) Variation in Φ between recon-
structed and generated photons.   structed and generated photons.

Figure 11: Determination of the photon matching parameters for generated sample.

a momentum of around 350 GeV, while the quark, and the jet to which it decays, will carry away 500 GeV of energy. The energy available for the decay $\widetilde{\chi}_1^0 \rightarrow \gamma \widetilde{S}$ is split between the photon and the singlino. The $\widetilde{S}$ has a set mass of 100 GeV in this simulation, leaving it with close to 150 GeV of momentum, while the photon is emitted with momentum 250 GeV. Due to the small mass difference between the $\widetilde{S}$ and S of 10 GeV, the final combined jet from the gluons will have a mass of around 140 Gev. Hence we conclude that both jets, products of this squark decay, possess a rather high energy (500 GeV from the quark jet plus 140 GeV from the gluon jet). Note that we have analyzed only one branch of the decay, and that a similar process is mirrored in the decay of the other squark produced in the collision, so that a signal event would have two quark jets and two gluon jets. As we expect high energy jets in a signal event, we set an additional requirement for signal which excludes events where the sum of the momentum of all the jets is less than 750 GeV. This is done using the HT variable, defined as[21]

$$HT = \sum_{all\ jets} |\vec{P}t_{JET}| \qquad (3)$$

After this analysis we also concluded that at least two jets on signal events must have mass of at least 50 GeV, since we expect two jets from the singlet decay with a combined mass of 140 GeV.

The last piece of the puzzle comes from the jets' structure. A shower of particles seen in the detector has to pass trough several filters and processes (e.g. vertex location, jet trimming, jet grooming, etc.) to be tagged as a jet. New algorithms, which are still

in test phase, can help scrutinize the jet even more, shedding light on its substructure. Of a particular interest to this study were the tau variables. The value of τ_N inversely quantifies how consistent a jet's structure is with having N axes [5]. For example, a jet product of the decay of a single quark will have a low τ_1, but a high τ_2, τ_3, etc. It follows that for our signal, the quark jet will have a low value τ_1, while the large jet coming from the two gluons will have a low τ_2 value. A complete description of jet substructure can be found in [10].

To determine the optimal use of such variables, we classified all the jets in the sample as jets coming from a quark, or jets coming from the neutralino decay. To characterize jets as quark jets we computed the spatial separation ΔR between each quark (which came from an squark, since the generated sample saves the particles predecessors) and all the jets in each event, and stored the minimum value. We then plotted these values and found a threshold for this parameter, shown in Figure 12. For the classification as neutralino jet we used a slightly different procedure. First, for each jet we found the closest two gluons from a singlet and plotted this to find a threshold for ΔR. Then, depending on whether one or two gluons met the ΔR requirement, the jets were classified as G jets or GG jets.
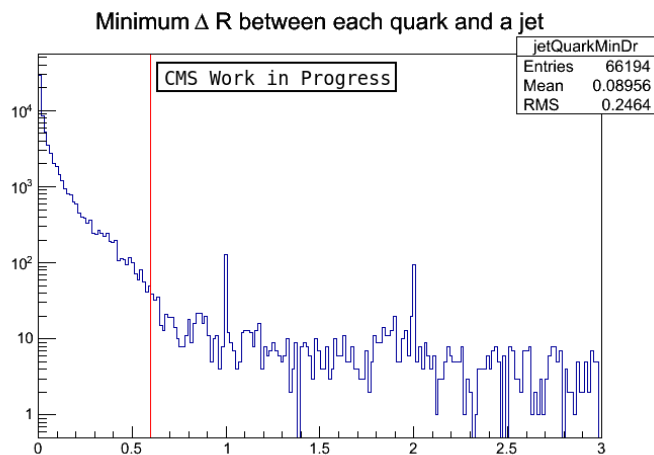
Figure 12: Plot used to determine a match as between a quark and a jet in the generated sample.

For this analysis we found that a cut in the ratio τ_3/τ_1 was more efficient than a cut in just one variable. This decision was reached after plotting of jet Pt vs the

---

[5]The definition of this variable is $\tau_N = \frac{1}{d_0} \sum_k p_{T,k} min\left[\Delta R_{1,k}, \Delta R_{2,k}, \ldots, \Delta R_{N,k}\right]$, where d_0 is a constant and ΔR is the distance from the k-th component of the jet to the N-th axis.
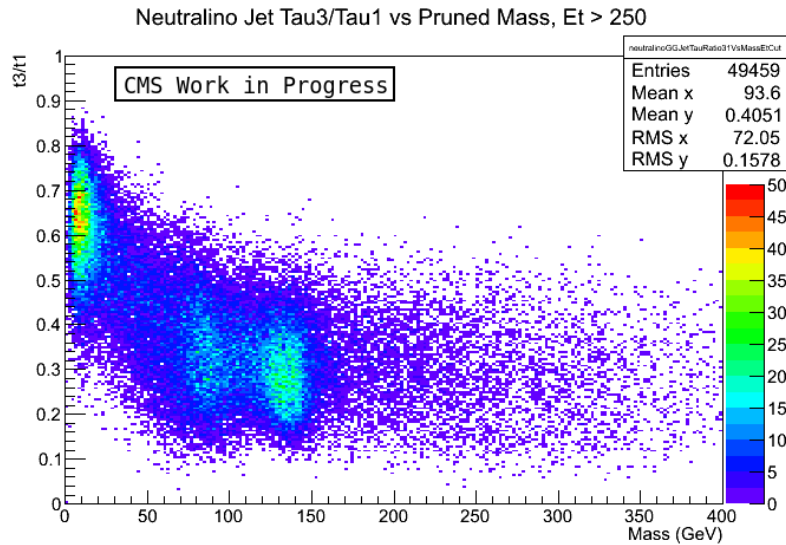
Figure 13: Plot of the gluon-jet mass vs τ_3/τ_1.

tau ratio, shown in Figure 13, where three spots are clearly visible, two with low ratio corresponding to the two sub-jets from gluons, and one with higher value corresponding to background quark jets. In contrast, Figure 14 shows the same distribution for jets coming from the squark. The final tau ratio cut requires that the event has two jets with $\tau_3/\tau_1 < 0.35$ to assure that the event has jets consistent with two separate gluons merged into one. A plot to test the results of this cut is depicted in Figure 15.

4.3 Experimental Data Events

In the same way we used CERN's GRID to obtain the data set of simulated events, the GRID was also used to transform the reconstructed data from the CMS detector to the final files to be analyzed as ROOT [6] files (sample in annexes section). The names of the datasets used for this analysis are listed in Table 1.

CMS data set names
/HT/Run2012A-22Jan2013-v1/AOD
/JetHT/Run2012B-22Jan2013-v1/AOD
/JetHT/Run2012C-22Jan2013-v1/AOD
/JetHT/Run2012D-22Jan2013-v1/AOD

Table 1: Data sets used for analysis.

---

[6]ROOT is a software framework developed by CERN and used to perform data processing and analysis. ROOT itself is composed mostly of C++ code, but it can be integrated with other languages such as PYTHON. [20]
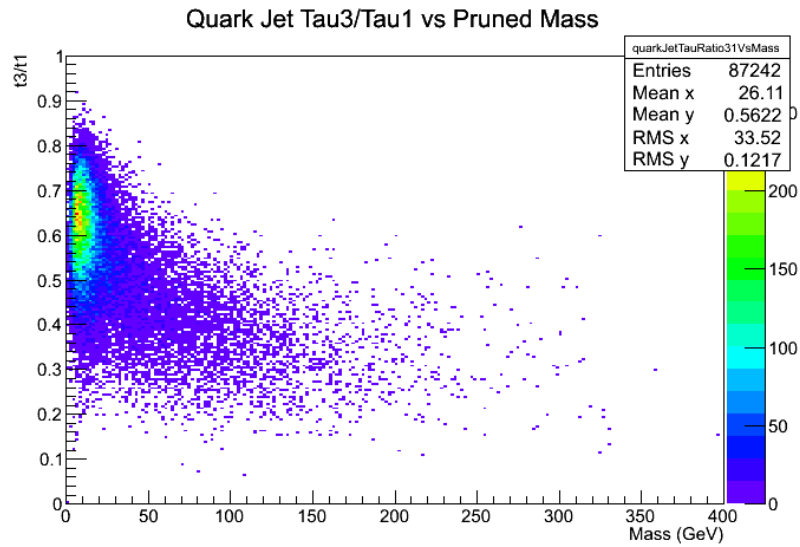
Figure 14: Plot of the quark-jet mass vs τ_3/τ_1.
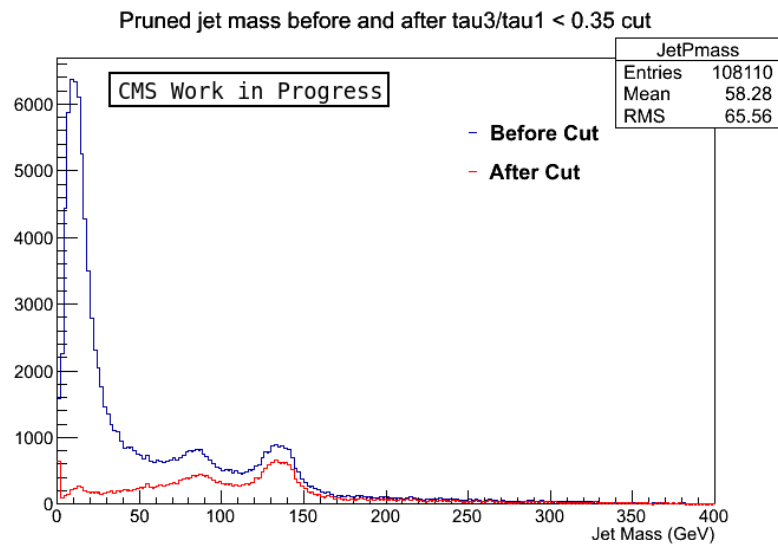
Figure 15: Plot of the mass of all jets in the simulation before and after applying the τ_3/τ_1 cut.

After obtaining the proper files for the analysis, we restricted our search only to events which produced high momentum particles. This was done by requiring events to have HT values above 750 GeV, since the kinematics of the signal will achieve this limit most of the time.

4.4 Background Estimation

The heart of this study is at the estimation of background events by using the experimental data. To achieve this, we first define regions of phase space where either the signal or background is dominant; in our case the variables to define this space were the number of jets and the jet mass. The thresholds for the number of jets was 4, and for the jet mass we tested the limits at 100 GeV, 150 GeV and 200 GeV. Based on this cuts, the signal dominant section is where events have 4 or more jets and jets have at least the required mass. In contrast the background dominated region is defined by low jet count (4 or less) and low jet mass. [22]
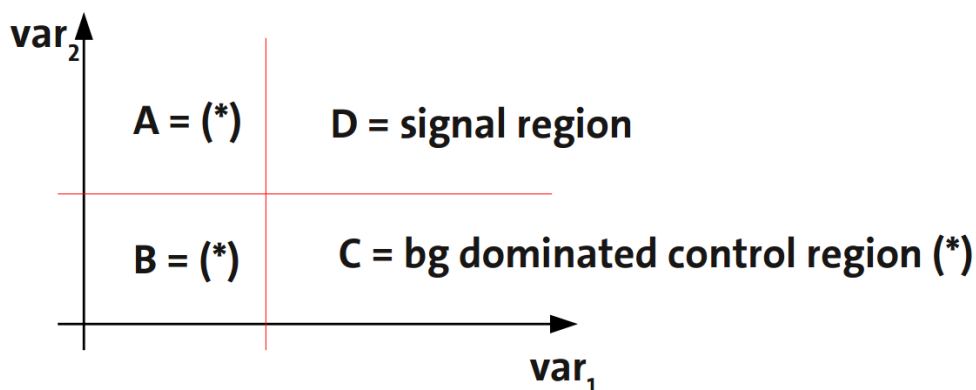
Figure 16: Diagram of the phase space used to define the signal and background dominant regions. Regions marked with (*) all are dominated by background events [22].

We then calculate the probability of background jets being present in the signal region. Specifically, we calculated the probability of one, two, or three (we studied each case separately) jets from the background region having mass > 100 GeV, 150 GeV and 200 GeV (all possible combinations). This probability was calculated for jets ordered in values of Pt and η, Figure 17. This was done for events with 2, 3 and 4 jets to make sure these three follow a similar pattern throughout the distributions, which

indicates that our discriminative variables are independent. These plots are depicted in Figures 17 and Figures 18.
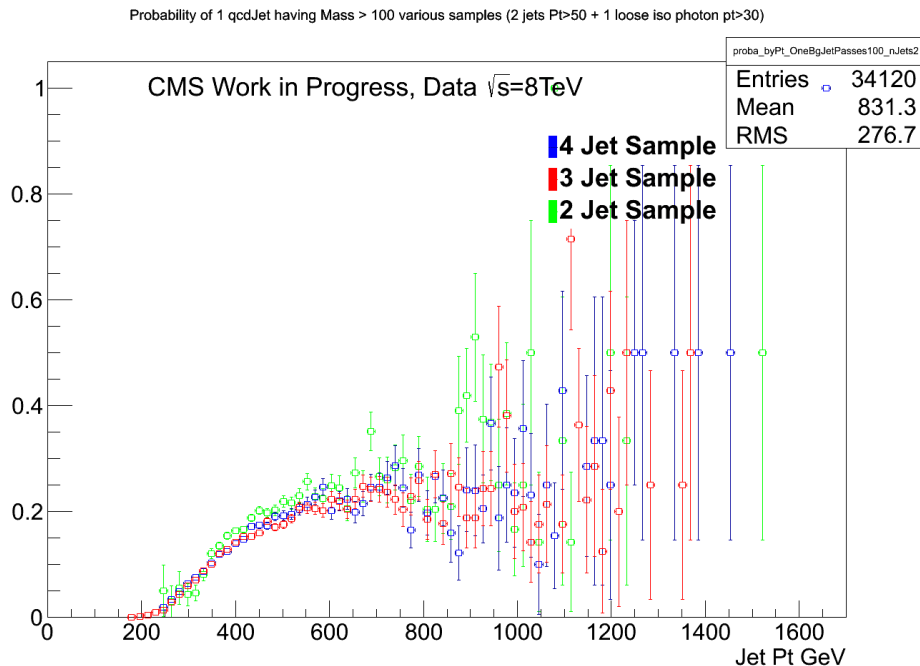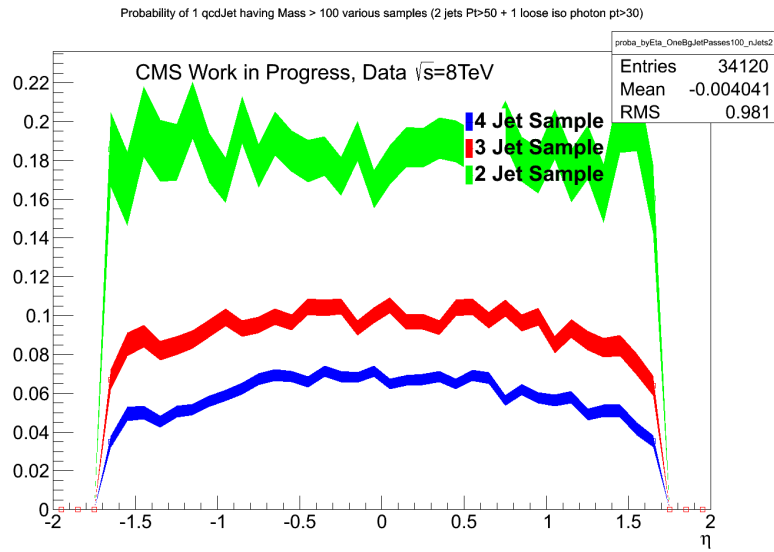

Figure 17: Sample of a plot of the probability of 1 background jet having mass greater than 100 GeV vs Pt, for events with two, three or four jets. Plots with more jets passing the threshold, and other mass limits were plotted and can be found as annexes. Note that the error bars in this plot correspond simply to statistical uncertainties.

Next we used these values of probability to make a prediction of the total HT of events with either one, two or three jets that pass the mass cut. This prediction was done simply by scaling the unaltered HT distribution, i.e. without any cuts applied, with the appropriate value of probability. Finally, this values of scaled HT were compared to the HT distribution of events with either one, two or three jets that pass the mass cut, to test the validity of the prediction of HT. The agreement between the actual data and the prediction indicates that this method of background estimation is accurate, and that, if applied to data in the signal region, it will properly estimate the background in this region as well. A sample of one of such plots is shown in Figure 19, while the full set of plots for all the mass limits and number of jets in the sample can be found in Figures 20 to 38 in the annexes section.

Figure 18: Sample of a plot of the probability of 1 background jet having mass greater than 100 GeV vs Eta, for events with two, three or four jets. Plots with more jets passing the threshold, and other mass limits were plotted and can be found as annexes. Note that the error bands in this plot correspond simply to statistical uncertainties.

5   Conclusions

A data-driven background estimation was performed for a stealth supersymmetry signal of the decay process $\tilde{q} \to q\widetilde{\chi}_1^0$, followed by $\widetilde{\chi}_1^0 \to \gamma\widetilde{S}$ and subsequent decay $\widetilde{S} \to S\widetilde{G}$ and final S decay via $S \to gg$. The close agreement between the estimated HT and jet HT data shown in figures 20 to 29 suggest that this technique, and the variable thresholds that where defined in this work, will provide an appropriate representation of the background events which mimic the squark decay in study, and could be considered as part of a full-blown search for stealth supersymmetry study through the channel depicted in Figure 6.
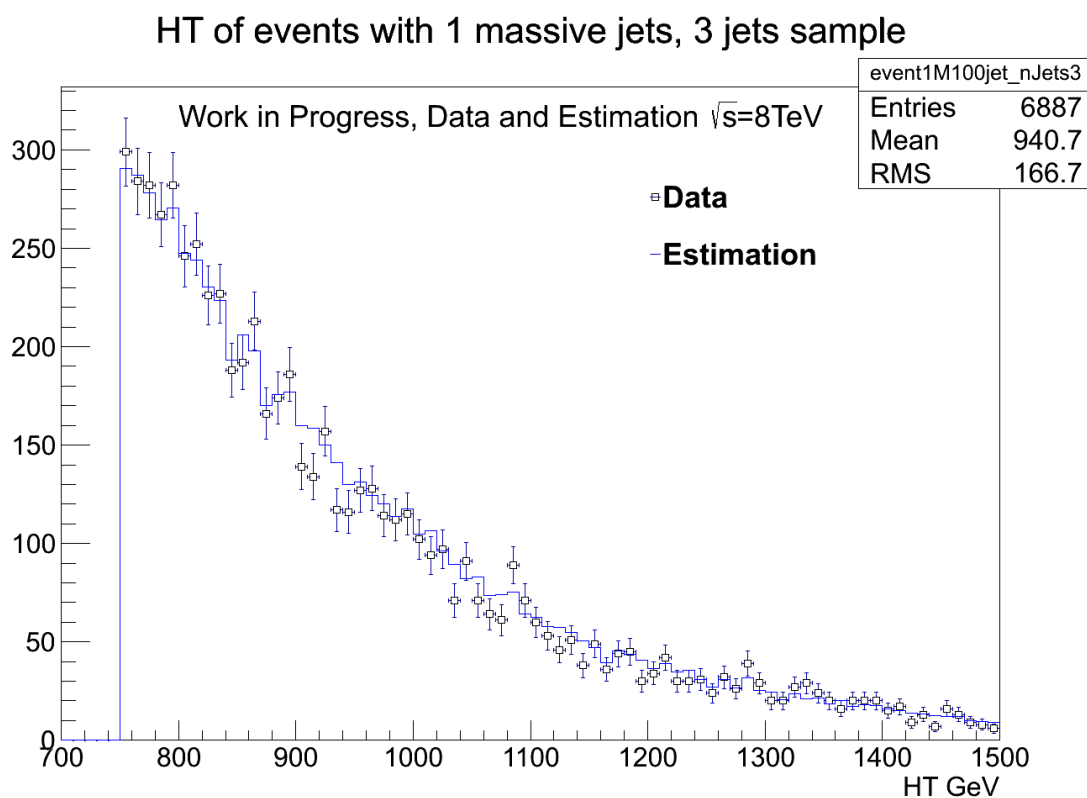
Figure 19: Sample of a plot of the prediction and real data of HT of events with 1 jet with mass > 100 GeV in events with 3 jets. Plots with more jets passing the threshold, and other mass limits, were plotted and can be found as annexes.

References

[1] Griffiths, D, *Introduction Elementary Particles* Second Edition. Physics textbook. Wiley, second edition, 2008.

[2] R. C. Slansky, "Lecture Notes - From Simple Field Theories To The Standard Model," Los Alamos Sci. **11**, 54 (1984).

[3] Katherine Garrett, Gintaras Duda. "Dark Matter: A Primer." *Adv.Astron.2011:968283,2011* arXiv:1006.2483 [hep-ph]

[4] Halzen, F., and Alan D. Martin. "Quarks and Leptons: An Introductory Course in Modern Particle Physics." New York: Wiley, 1984

[5] S. Iso, "What Can We Learn from the 126 GeV Higgs Boson for the Planck Scale Physics? - Hierarchy Problem and the Stability of the Vacuum -," doi:10.1142/9789814566254_0025. arXiv:1304.0293 [hep-ph].

[6] Wess, J.; Zumino, B. (1974). "Supergauge transformations in four dimensions". Nuclear Physics B 70: 39. Bibcode:1974NuPhB..70...39W. doi:10.1016/0550-3213(74)90355-1.

[7] Y.Shirman, "Introduction to Supersymmetry and Supersymmetry Breaking", doi:10.1142/9789812838360_0008. arXiv:0907.0039 [hep-ph].

[8] Joshua T. Ruderman JiJi Fan, Matthew Reece. "Stealth supersymmetry". *JHEP 1111 (2011) 012,* 2011. arXiv:1105.5135v1 [hep-ph]

[9] CMS Collaboration. "Search for stealth supersymmetry in events with jets, either photons or leptons, and low missing transverse momentum in pp collisions at 8 tev." *Phys. Lett. B 743 (2015) 503,* 2015.

[10] "Towards an understanding of jet substructure" arXiv:1307.0007v2 [hep-ph] http://arxiv.org/pdf/1307.0007v2.pdf

[11] Brüning, Oliver Sim and Collier, Paul and Lebrun, et al. "LHC Design Report". CERN, Geneva. 2004.

[12] "About CERN". Online. Retrieved on Nov 20, 2015 from http://home.cern/about

[13] Lenzi, T. "Development and Study of Different Muon Track Reconstruction Algorithms for the Level-1 Trigger for the CMS Muon Upgrade with GEM Detectors". arXiv:1306.0858

[14] CMS Collaboration, "The CMS experiment at the CERN LHC", JINST 3 (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[15] CMS Collaboration. "Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at sqrt(s) = 8 TeV".DOI: 10.1088/1748-0221/10/08/P08010 . arXiv:1502.02702 [physics.ins-det]

[16] The LCG TDR Editorial Board. "LHC Computing Grid Technical Design Report", CERN, Geneva. 2004. Eprint version retrieved from https://espace.cern.ch/WLCG-document-repository/Technical_Documents/TDR/LCG_TDR_v1_04.pdf on December 20, 2015.

[17] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, M. Zaro. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations." 2014. arXiv:1405.0301 [hep-ph]

[18] Torbjorn Sjostrand (Lund University), Stephen Mrenna (Fermilab), Peter Skands (Fermilab). *PYTHIA 6.4 Physics and Manual* JHEP 0605:026,2006. DOI:10.1088/1126-6708/2006/05/026. arXiv:hep-ph/0603175

[19] CMSPublic. "Software Guide on CRAB". Online. Retrieved from https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideCrab on December 20, 2015.

[20] Rene Brun and Fons Rademakers, "ROOT - An Object Oriented Data Analysis Framework," Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) 81-86. See also http://root.cern.ch/.

[21] The CMS Collaboration. "Search for new physics with jets and missing transverse momentum in pp collisions at $\sqrt{s} = 7$ tev". Journal of High Energy Physics. `http://arxiv.org/pdf/1106.4503.pdf`

[22] Sander, C. Lecture on Data Driven Background Estimation. Online. Retrieved on Nov 20, 2015 from `http://www.desy.de/~csander/Talks/120223_SFB_DataDrivenBackgrounds.pdf`

HT of events with 0 massive jets, 2 jets sample
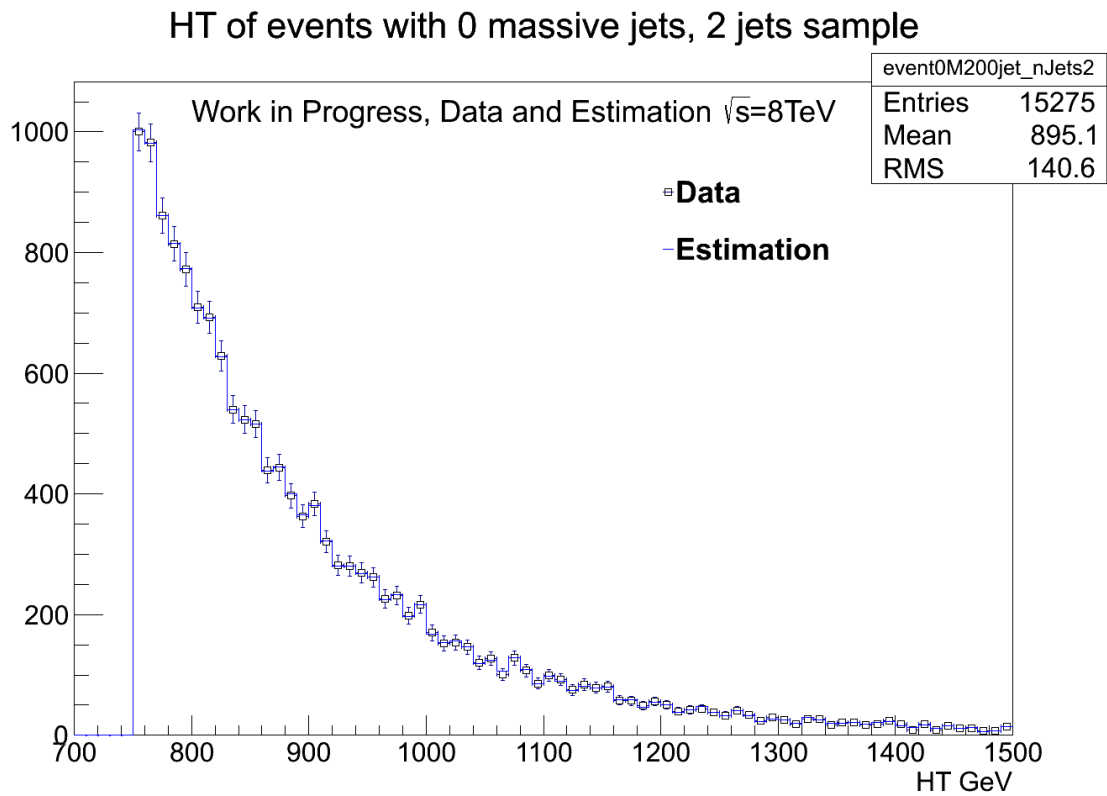
Figure 20: Data and prediction of HT for events with two jets.

6   Annexes

6.1   Plots

HT of events with 0 massive jets, 2 jets sample
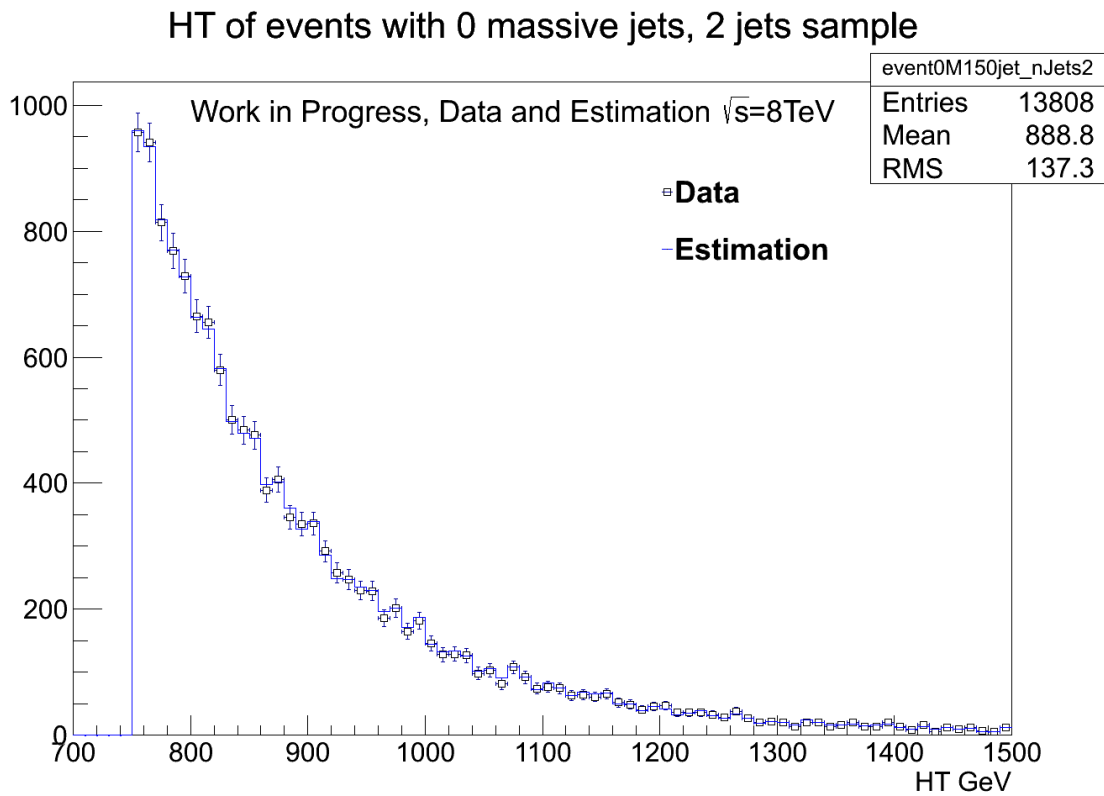
Figure 21: Data and prediction of HT for events with two jets.

HT of events with 0 massive jets, 2 jets sample
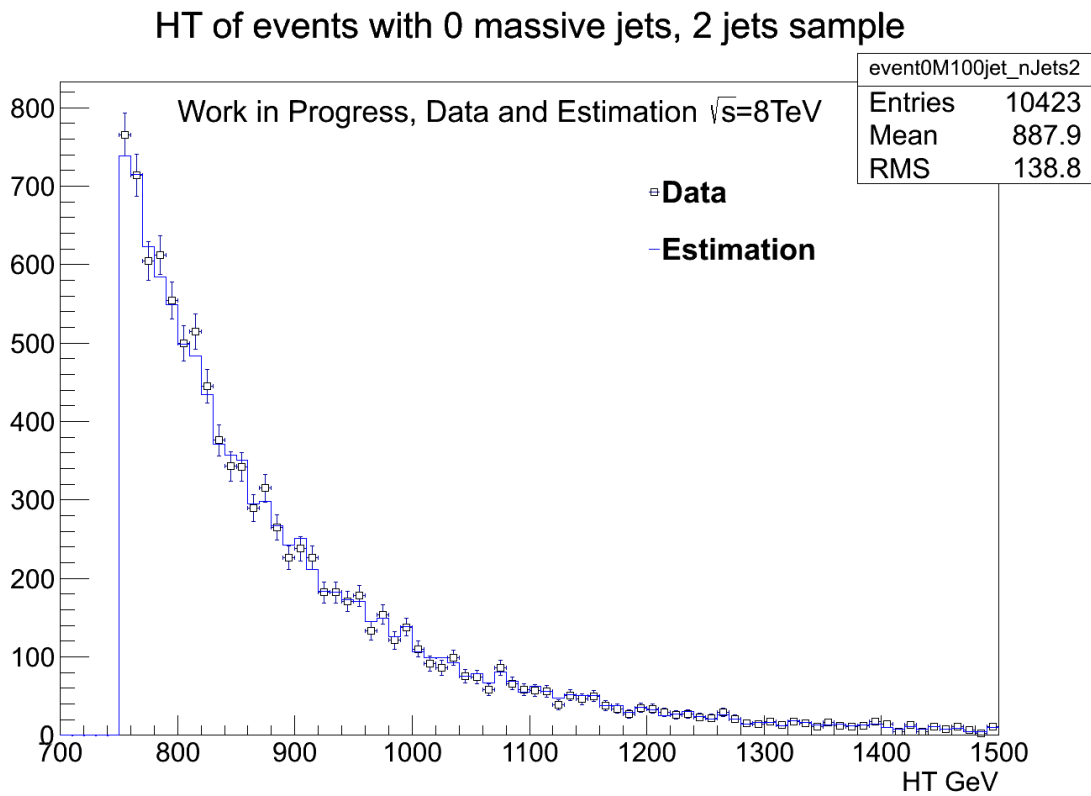
Figure 22: Data and prediction of HT for events with two jets.

HT of events with 1 massive jet
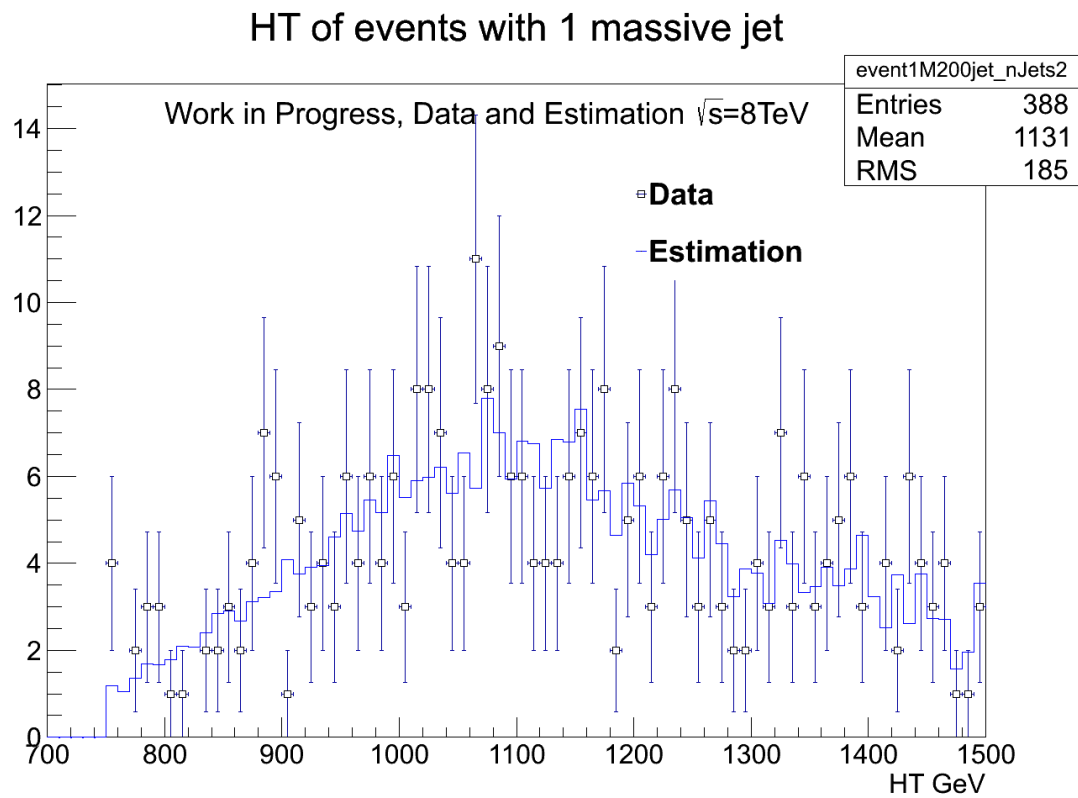
Figure 23: Data and prediction of HT for events with two jets.

HT of events with 1 massive jet
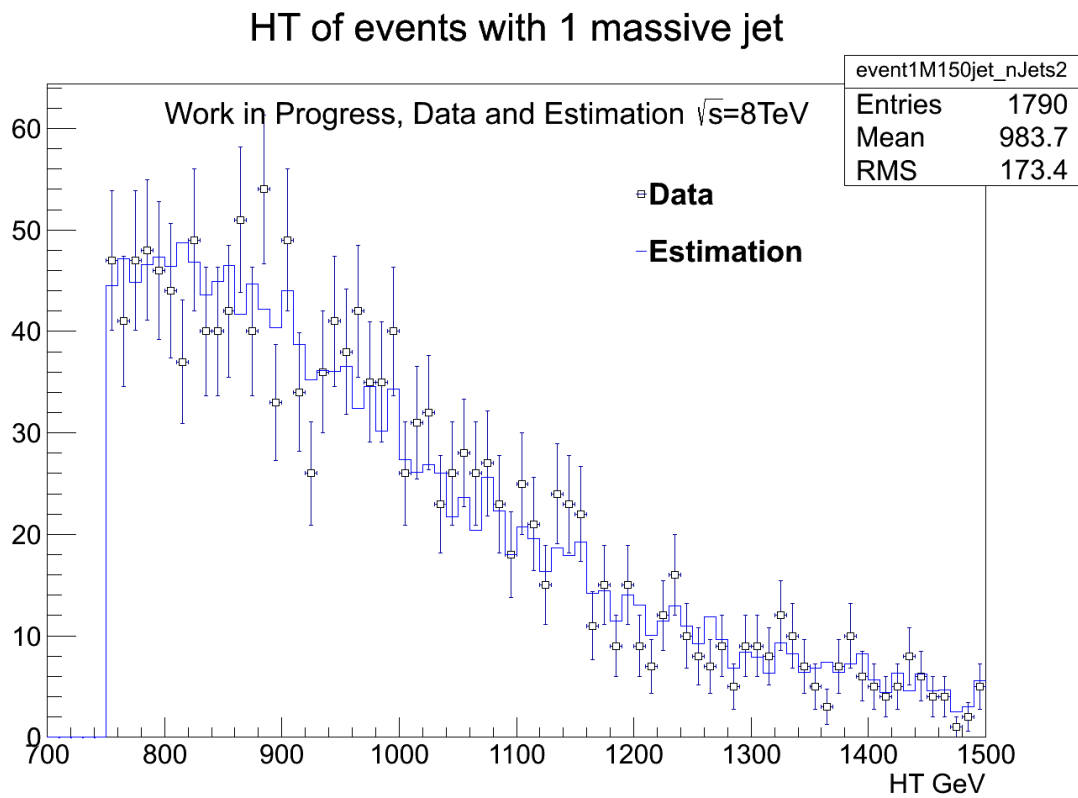
Figure 24: Data and prediction of HT for 'events with two jets.

HT of events with 1 massive jet
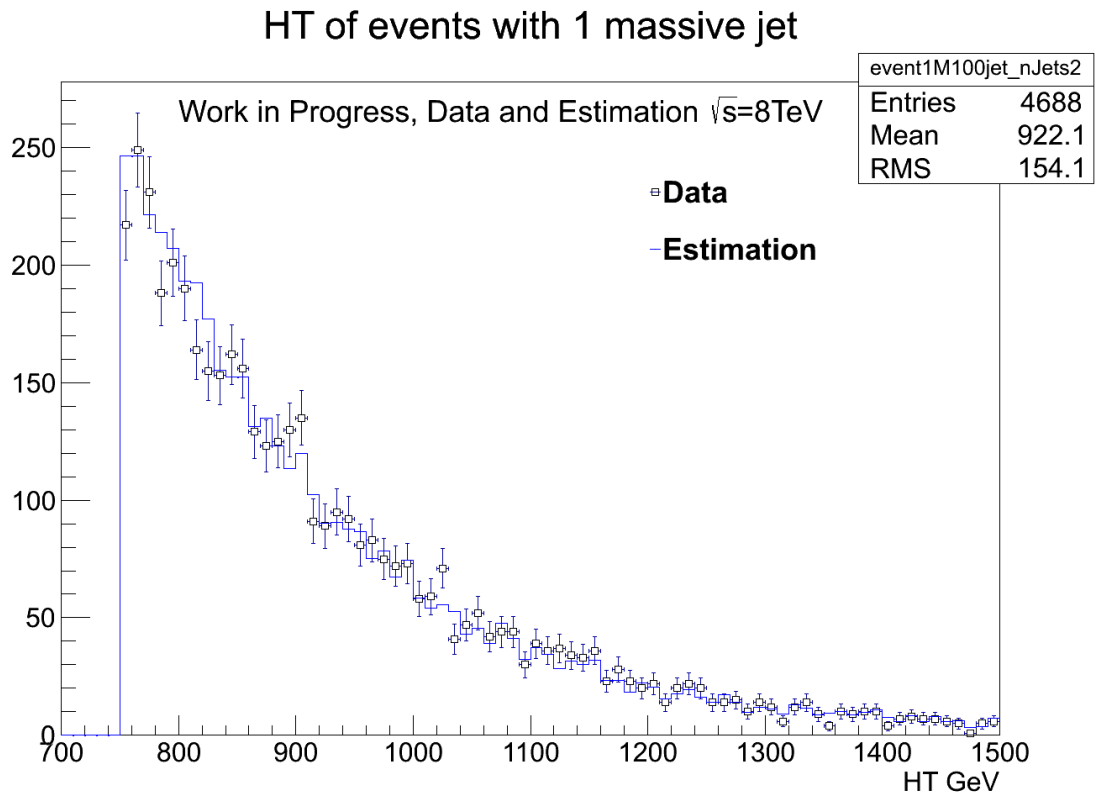
Figure 25: Data and prediction of HT for events with two jets.

HT of events with 2 massive jets, 2 jets sample
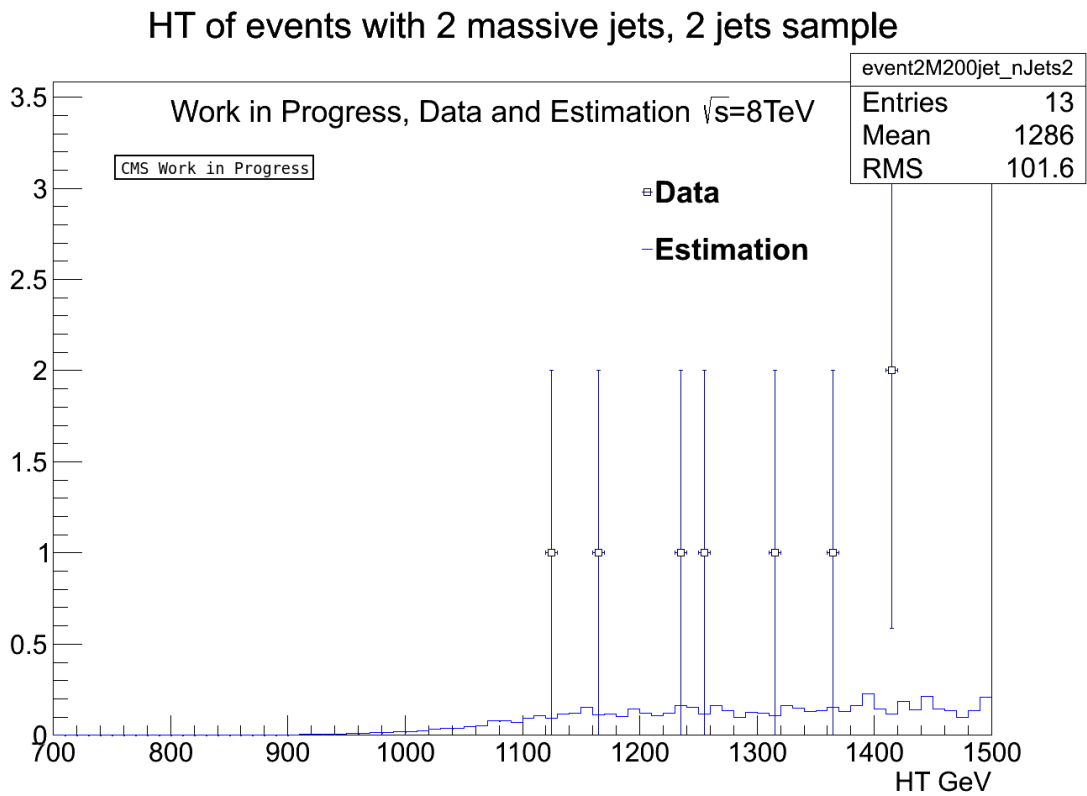
Figure 26: Data and prediction of HT for events with two jets.

HT of events with 2 massive jets, 2 jets sample
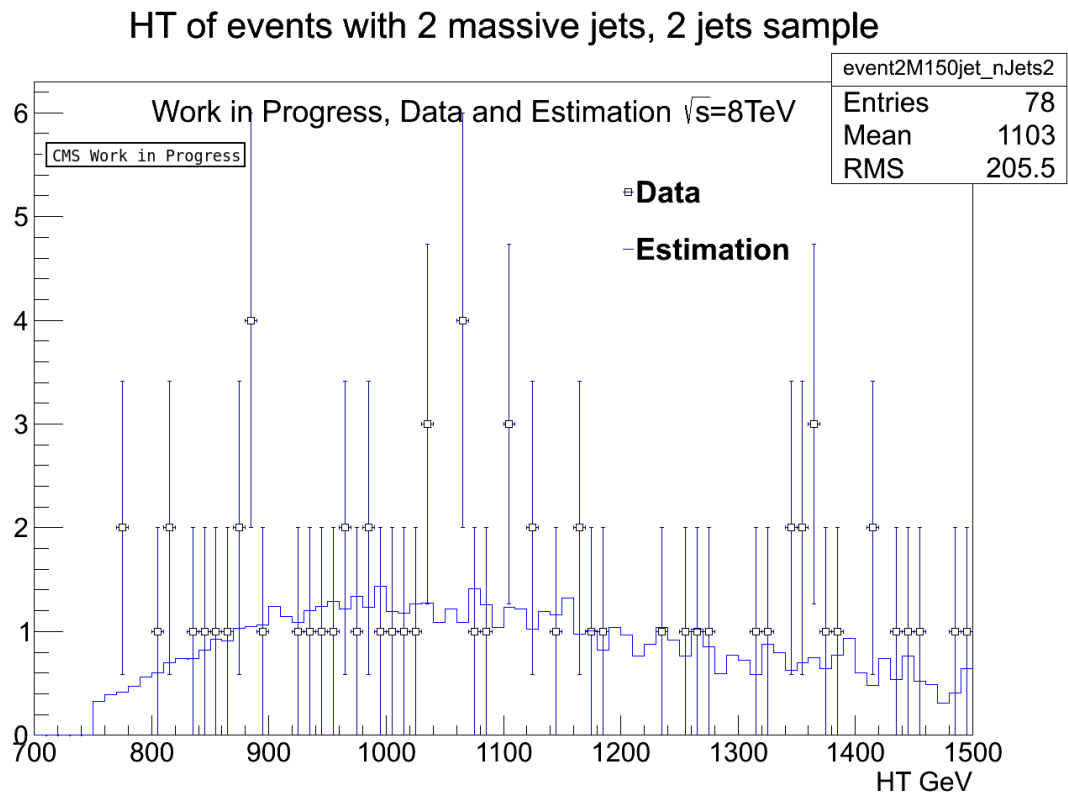
Figure 27: Data and prediction of HT for events with two jets.

HT of events with 2 massive jets, 2 jets sample
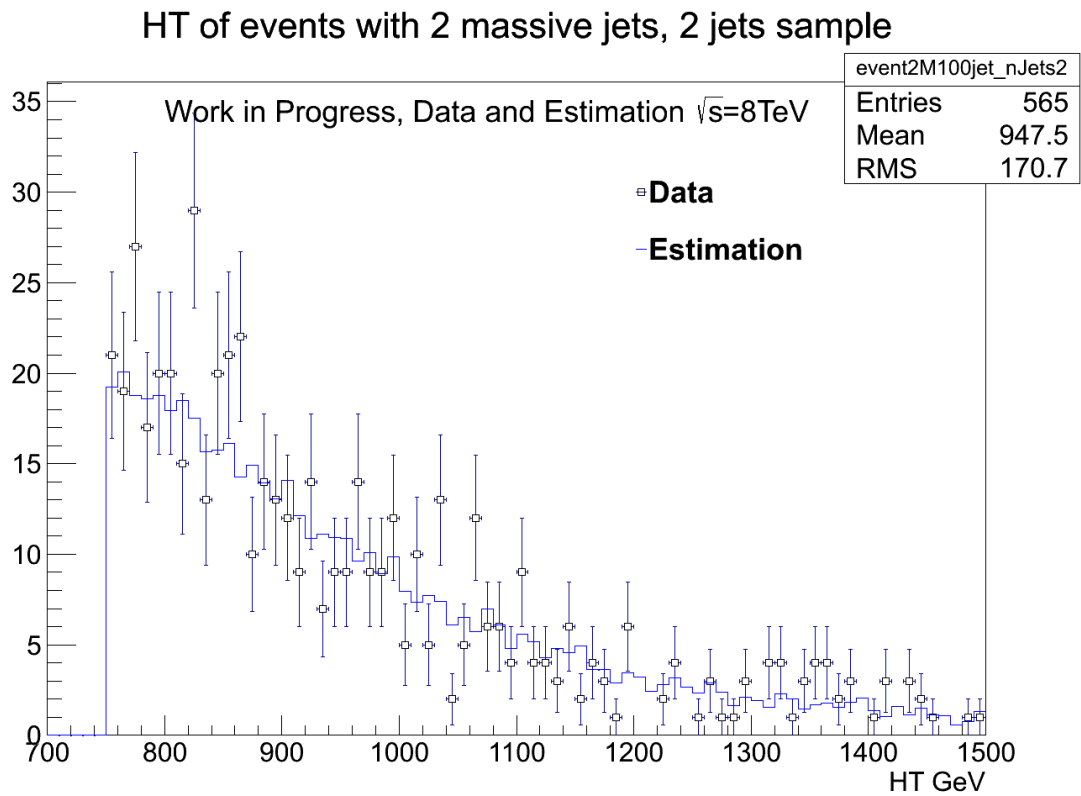
Figure 28: Data and prediction of HT for events with two jets.

HT of events with 0 massive jets, 3 jets sample
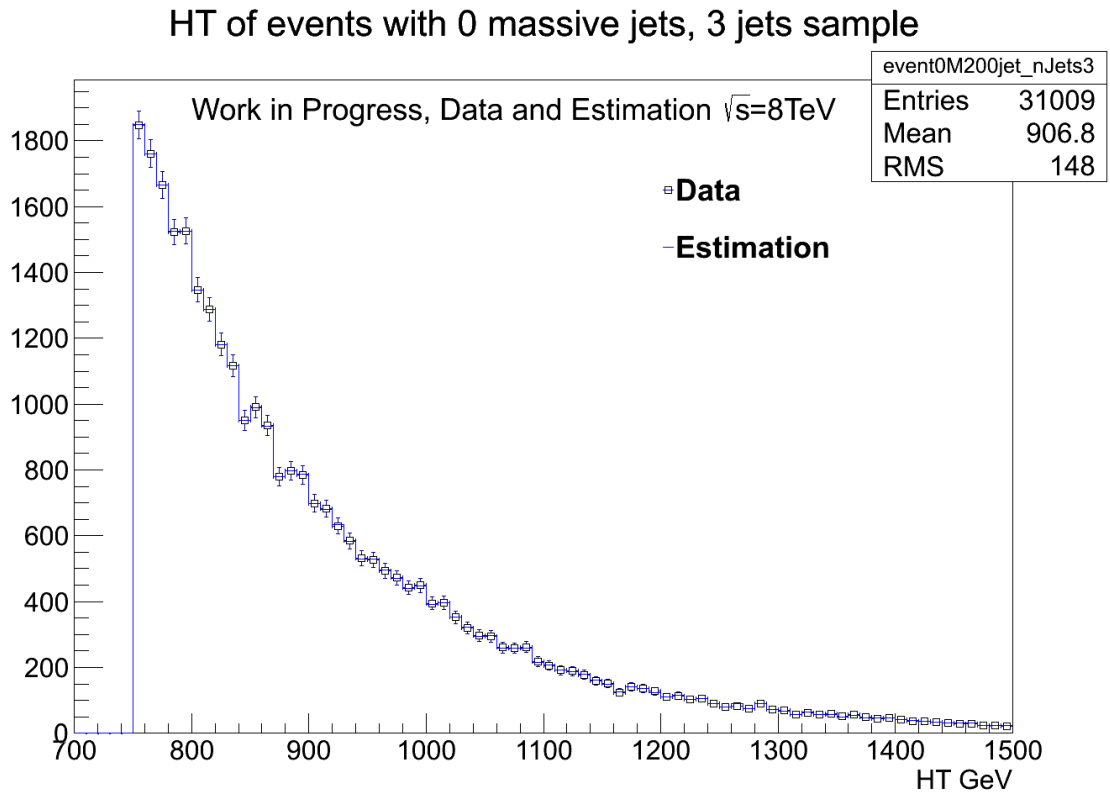
Figure 29: Data and prediction of HT for events with two jets.

7

---

[7]Plots of events with 3 jets passing the 200 GeV and 150 GeV had zero entries and are not shown here.
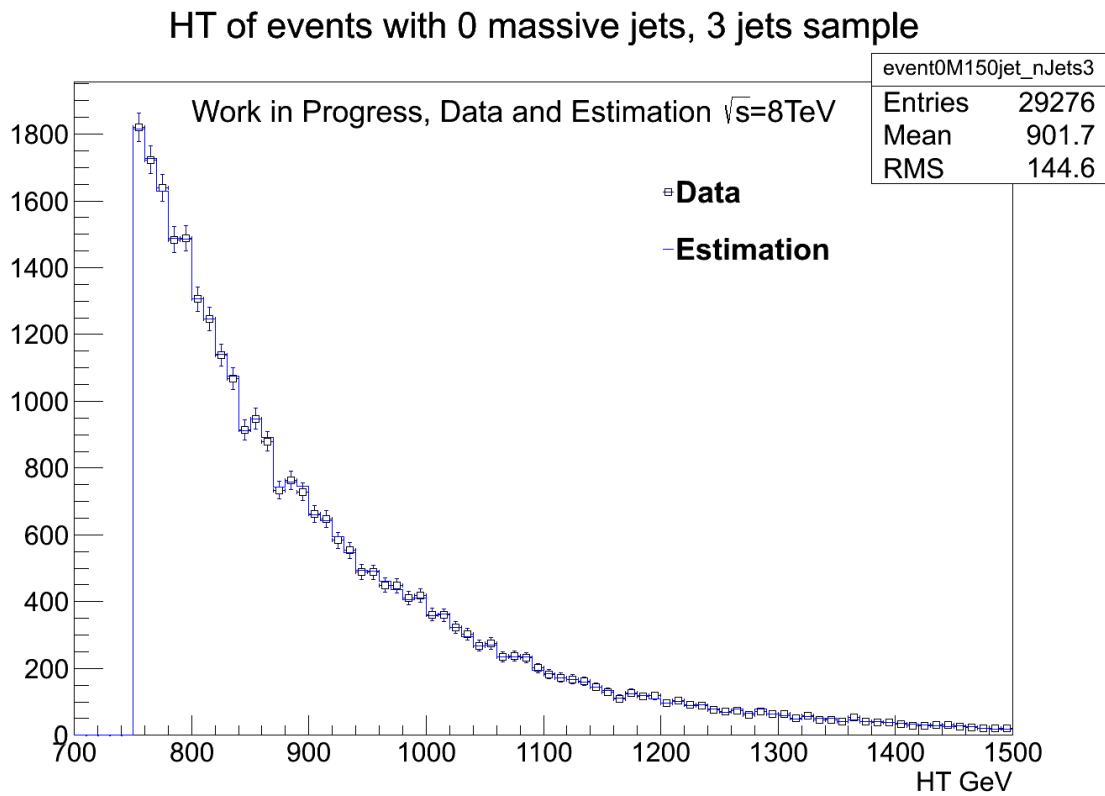
Figure 30: Data and prediction of HT for events with two jets.
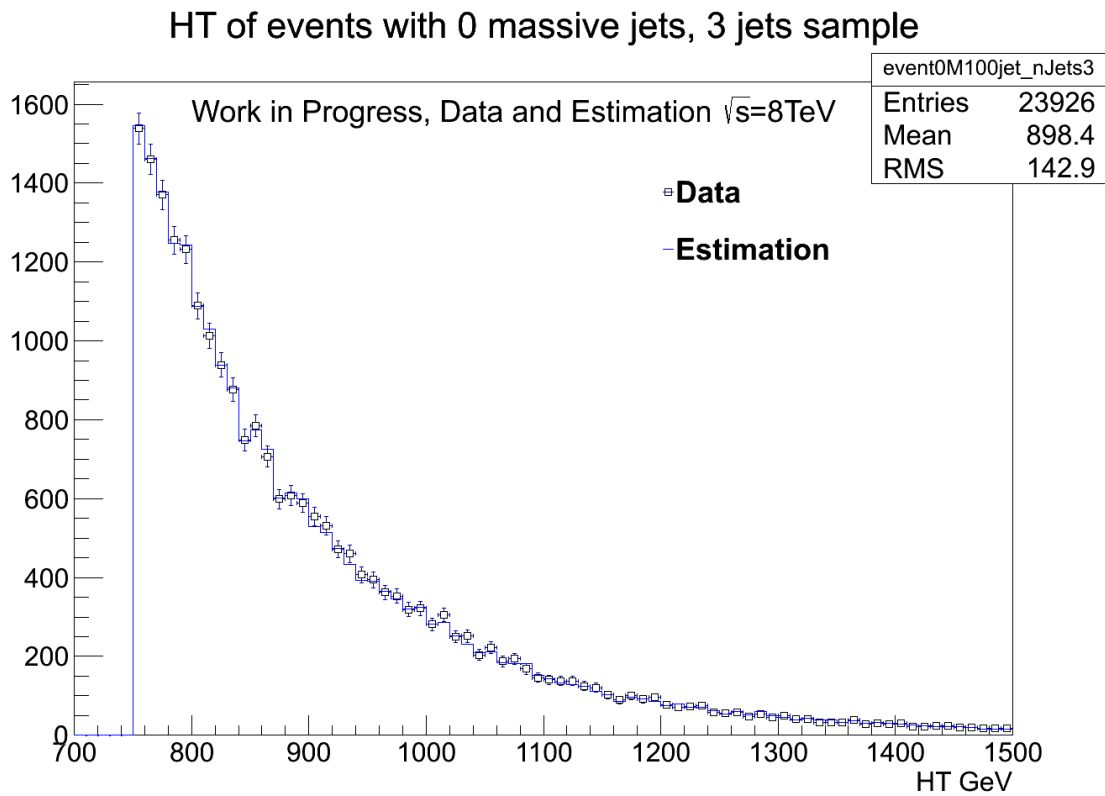
Figure 31: Data and prediction of HT for events with two jets.
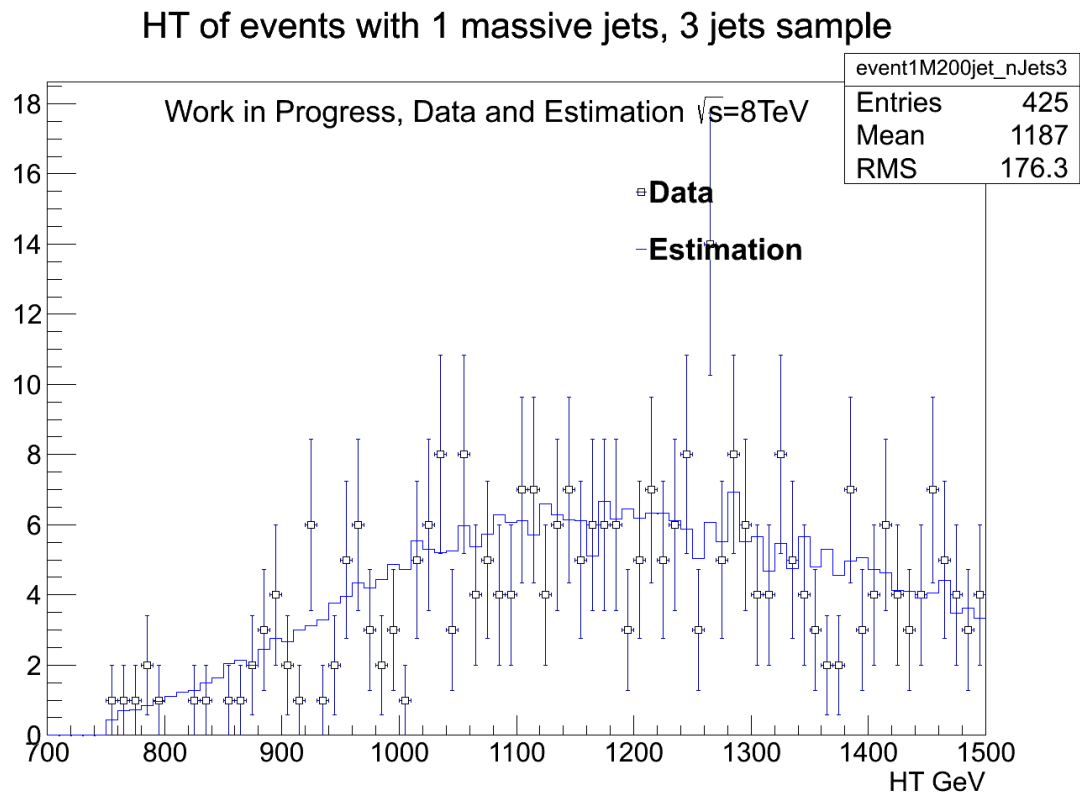
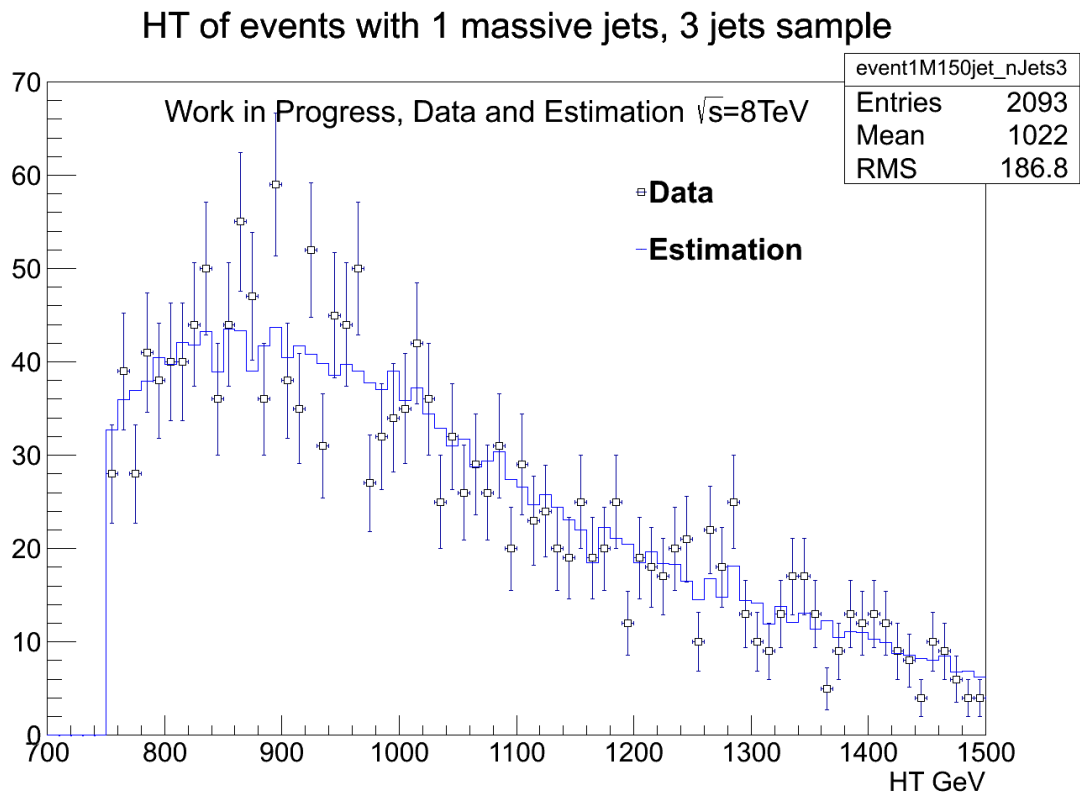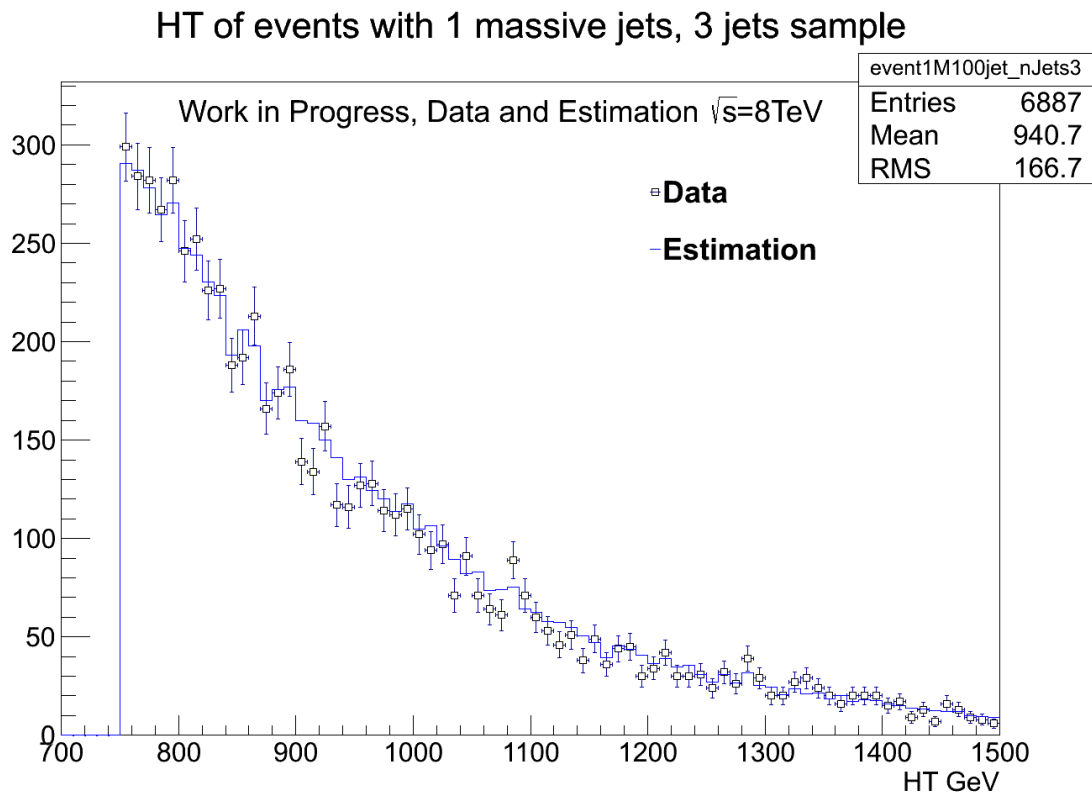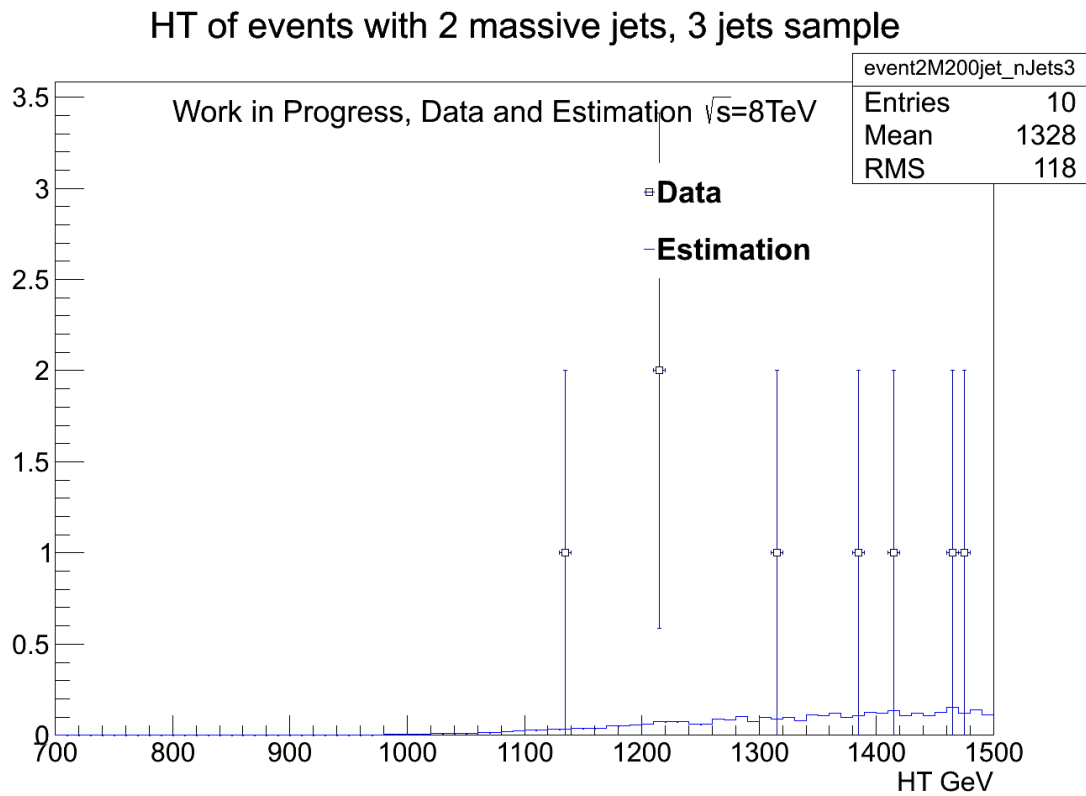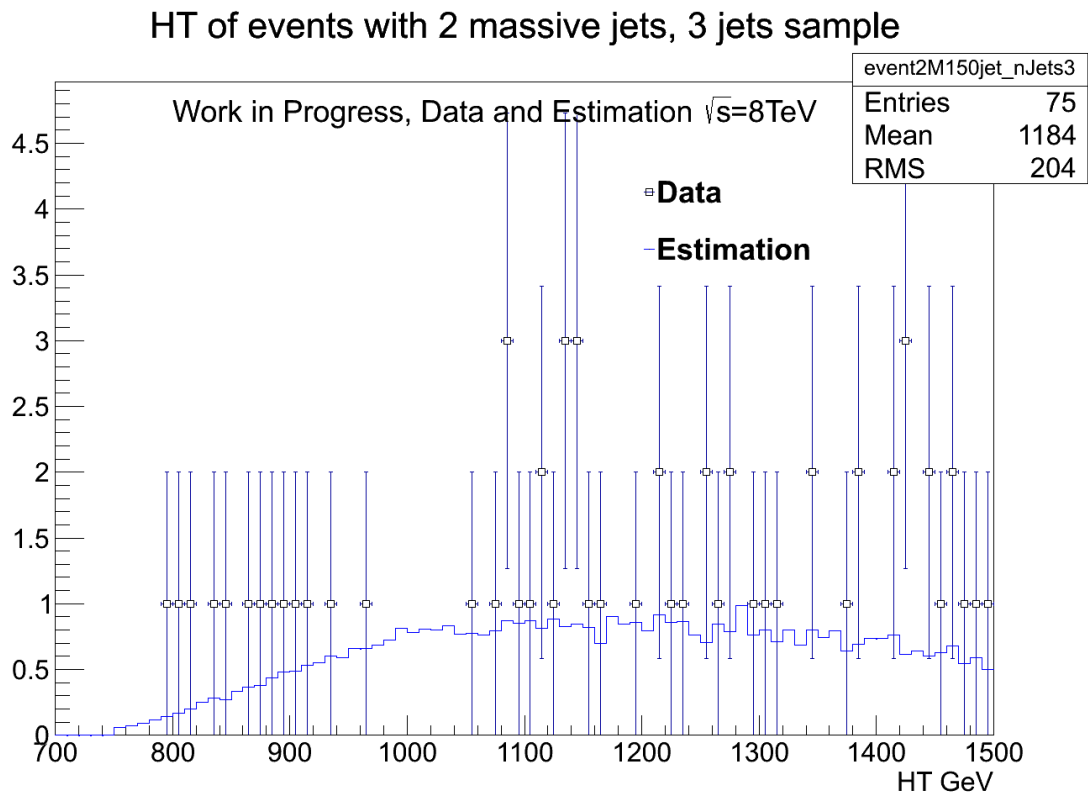Figure 32: Data and prediction of HT for events with two jets.

Figure 33: Data and prediction of HT for events with two jets.

HT of events with 1 massive jets, 3 jets sample

Figure 34: Data and prediction of HT for events with two jets.

HT of events with 2 massive jets, 3 jets sample

Figure 35: Data and prediction of HT for events with two jets.

HT of events with 2 massive jets, 3 jets sample

Figure 36: Data and prediction of HT for events with two jets.
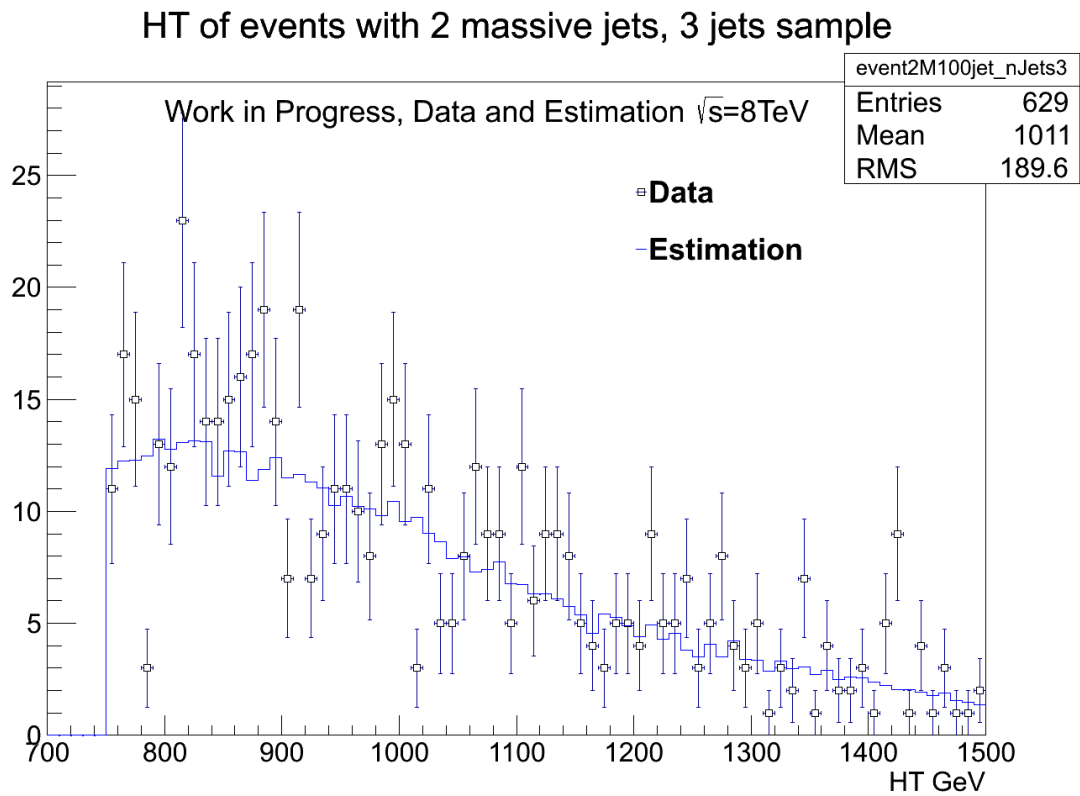
HT of events with 2 massive jets, 3 jets sample

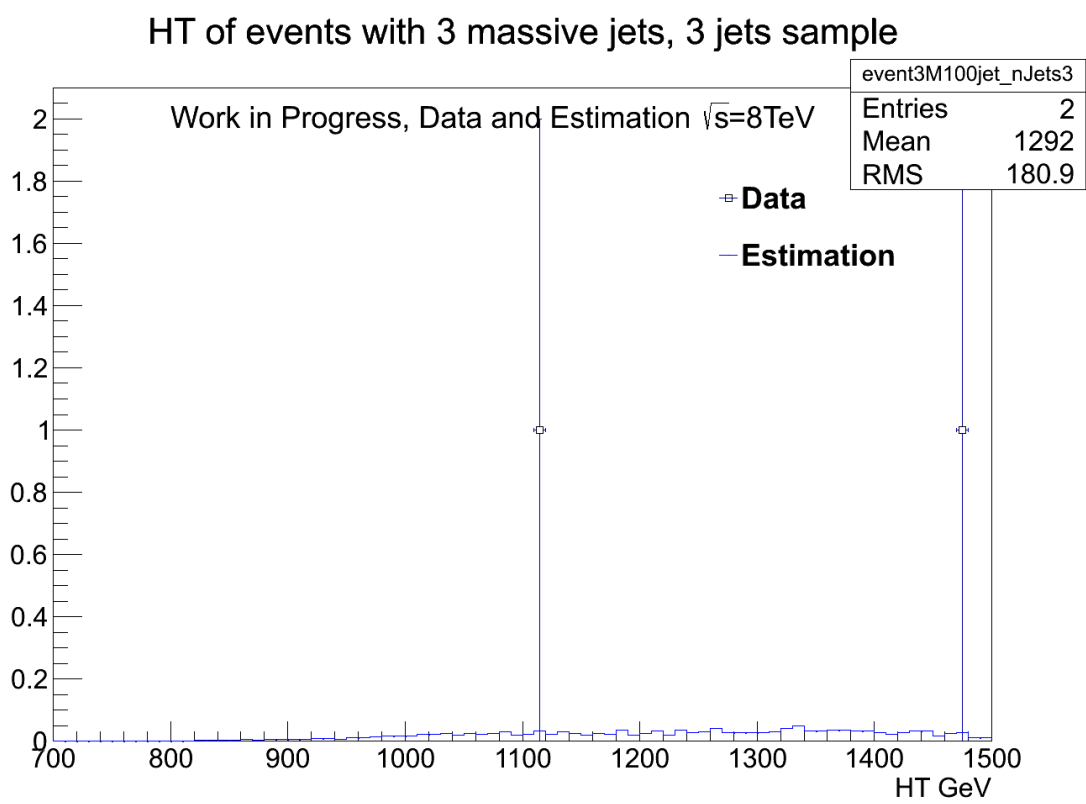
Figure 37: Data and prediction of HT for events with two jets.

Figure 38: Data and prediction of HT for events with two jets.