Karlsruher Institut für Technologie

Cross-check of the CMS search for additional MSSM Higgs bosons in the di-τ final state using $\mu \to \tau$ embedded events

Zur Erlangung des akademischen Grades

MASTER OF SCIENCE

von der Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)
genehmigte

MASTERARBEIT

von

B.Sc. Janek Bechtel

geb. 20.02.1991
in Karlsruhe

Abgabedatum: 24. Oktober 2017

Referent: Prof. Dr. Günter Quast
Korreferent: Priv.-Doz. Dr. Roger Wolf

Institut für Experimentelle Teilchenphysik

Abstract

The *Standard Model* of particle physics (SM) is the theory used to describe both the fundamental constituents of matter and the fundamental interactions. While being very successful in describing the observed phenomena, it leaves some questions unanswered. Extensions to the SM such as the *Minimal Supersymmetric extension to the Standard Model* (MSSM) qualify for solving these open questions. A promising signal channel to test the MSSM is the decay of an additional neutral Higgs boson into two τ-leptons. For signal extraction, important background processes such as the decay of the Z boson into two τ-leptons need to be understood. In this thesis, the development and the application of the *embedding* method are described. This data-driven method is developed to improve the modeling of $Z \to \tau\tau$ events at the CMS detector. Embedded events are used to search for a heavy neutral Higgs boson in the context of the MSSM. The results presented in this thesis serve as a cross-check of the background description by simulated $Z \to \tau\tau$ events that is used for the results published by the CMS collaboration.

Contents

CHAPTER 1

Introduction

The goal of physics is to describe all matter and its fundamental interactions to predict the evolution of objects through time and space. For centuries, this was limited to the study of objects that are observable by eye. However, the advent of new discoveries in the early 20th century lead to the development of radical new theories such as quantum mechanics, which forms the basis of modern particle physics. It was found that the principles and models describing macroscopic objects fail to explain the phenomena observed when examining the smallest constituents of matter. New mathematical models were postulated and experimentally validated in a reproducible way to expand the human understanding of the universe to these fundamental objects. Theoretical effort and technological progress in the last century resulted in the discovery of a variety of new particles and in the understanding of the fundamental forces describing their interactions, with the latest discovery being the Higgs boson at the *Large Hadron Collider* (LHC) at the *European Organization for Nuclear Research* (CERN) in 2012 [1, 2].
The Higgs boson, already predicted in 1964 [3–5], was a longed-for missing piece in the description of mass-generation for the constituents of matter and their force carriers. Our current understanding of particle physics, expressed in the *Standard Model* (SM), was repeatedly validated with this discovery. Nonetheless, the SM is still known to be deficient, for example in its excessive need for fine-tuning or its inability to provide a dark-matter candidate. Supersymmetric theories such as the *Minimal Supersymmetric extension to the Standard Model* (MSSM) qualify for solving such problems. The MSSM predicts two additional neutral Higgs bosons and is currently being tested at high accuracy using an increasing amount of high-energy collision data, e.g. taken by the *Compact Muon Solenoid* (CMS) detector at LHC. A promising signal channel to test the MSSM is the decay of a neutral Higgs boson into two τ-leptons, a higher-mass sibling of the electron with an enhanced coupling to a Higgs boson. The signal is predicted to be small compared to the background of well-known non-Higgs SM processes, most notably the decay of the neutral force carrier of the weak force, the Z boson, which also decays into two τ-leptons. An accurate prediction of this background is crucial for the discovery of new particles.
After a brief introduction into the Standard Model of particle physics in the following, the *embedding* method will be introduced. This data-driven method is developed to improve

the prediction of $Z \to \tau\tau$ decays. The focus will then be laid upon the application of the embedding method, including dedicated model-to-data corrections for the use of this method, a validation using simulated events and, finally, the estimation of $Z \to \tau\tau$ events using embedded events as background description for a search for heavy neutral Higgs bosons decaying into τ-leptons.

1.1 The Standard Model of particle physics

The Standard Model of particle physics is the theory used to describe the fundamental constituents of matter and the fundamental forces specifying their interactions, namely the electromagnetic force, the weak force and the strong force. Gravitation, the easiest force to observe in everyday life, is the only interaction not yet described in a consistent framework in combination with the other three. In the realms of particle physics, this force is also by far the weakest and irrelevant for the description of processes concerning the interaction of fundamental particles. The particles in the SM can be categorized into two fundamental groups: Particles with half-integer spin are *fermions*, particles with integer spin are *bosons*. Nature has chosen fermions as constituents of matter, with an example being the electron, and bosons, such as the photon, as force carriers. The SM imposes twelve fermions as constituents of matter, which can be further categorized by the means of their participation in the strong interaction.
Leptons like electron do not participate in the strong interaction and only interact via the electromagnetic and weak force. For this thesis, the two heavy generations of leptons will be of particular importance: The *muon* and *τ-lepton*. These leptons are similar to the electron in their interactions, yet their masses are higher than the electron mass by a factor of around 200 for the muon, and 3500 for the τ-lepton. For this reason, muons and τ-leptons are unstable for everyday purposes, with an average lifetime of $2,2 \cdot 10^{-6}$ s for the muon, and $2,9 \cdot 10^{-13}$ s for the τ-lepton, after which both decay into lighter particles. Whereas the muon lifetime is long enough for it to be directly observable in a particle detector, the τ-lepton usually decays after a distance in the order of some millimeters, making it only observable via its decay products. Additional leptons are three neutrinos that interact only via the weak interaction and carry neither electromagnetic nor strong charge, and only appear at experiments via energy that escapes detection.
Quarks are fermions that additionally participate in the strong interaction and can therefore not be observed in isolation, for reasons that will be evident in the discussion of the strong force below. The six quarks form composite particles with other quarks which are called hadrons, with prominent examples being the pion, proton or neutron.

The three fundamental forces described by the SM are mediated by force carriers, all of which are particles with integer spin and therefore categorized as bosons.

- **The electromagnetic force** is mediated by the photon. Of the forces described by the Standard Model, it is most intuitive and heavily used for technological applications such as the generation of electromagnetic waves for transmission of information. As photons are massless, the electromagnetic force has infinite range.
- **The weak force** is mediated by three massive force carriers: Two charged W bosons and the neutral Z boson. Even though it is comparable in strength to the electromagnetic force, it only has subatomic range due to the high mass of these force carriers. Upon its discovery, the weak force was seen as an independent force due to the drastically different behavior from all other known forces. At high energies, however, the weak and electromagnetic force unify in what is called electroweak unification [6–8], the postulation of which predicted the three force carrying gauge bosons. The coupling of leptons to gauge bosons such as the Z is independent of the lepton flavor in a property called *lepton universality*. The Z boson, which was discovered in 1983 at the SPS collider at CERN [9, 10], is of particular importance for this thesis due to its resonant behavior at collider experiments, a detailed description of which will be given below.
- **The strong force** is mediated by eight gluons. The corresponding charge is the color charge. Two properties distinguish the strong force from all other forces: The coupling strength of the strong force decreases for short distances, resulting in particles that can be described as free when their relative distance approaches the size of a proton. For larger distances, the potential of the strong interaction increases linearly. For this reason, isolated color-charged particles such as quarks cannot be observed, as the energy stored in the field results in the creation of new particles, forming color-neutral composite particles called hadrons.

1.1.1 The Z boson at hadron colliders

The Z boson is one of the heaviest particles in the Standard Model with a mass of $91.2\,\text{GeV}$[1] [**pdg1**], which makes it almost as massive as 100 hydrogen nuclei. As a force carrier of the weak force, the Z boson couples to all fermions. At a particle collider such as the LHC, Z bosons can be created by the annihilation of a quark-antiquark pair. Here, the two quarks annihilate to form leptons, using a photon or Z boson as force carrier. A resonance curve in the invariant mass spectrum of di-lepton events can be observed, as shown in figure 1.1.
The well-measured resonant cross section of the Z boson is used as a tool for analysis at the LHC as a source of high-energy di-lepton events. In this thesis, leptons from the Z peak will be used both for the selection of di-muon events for the embedding method as well as for the derivation of *tag and probe* scale factors as described in [11]. Furthermore, events with a decay of the Z boson into leptons form an important background when

[1] In this thesis, natural units will be used when referring to masses and momenta, i.e. $c = 1$.

searching for signals in di-lepton final states: With a cross section of the $q\bar{q} \to Z \to \ell^-\ell^+$ process in the order of $\mathcal{O}$(nb) at its peak [12], it exceeds possible $\mathcal{O}$(pb) signals of new physics in di-lepton final states by three orders of magnitude.

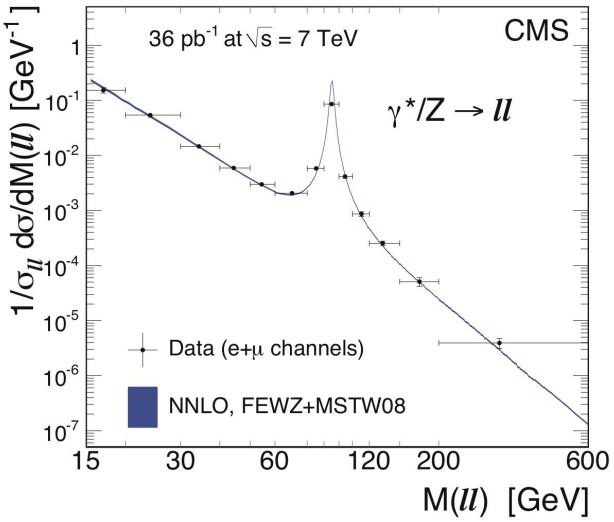


Figure 1.1: Invariant mass spectrum of di-lepton events from $q\bar{q} \to Z \to \ell^-\ell^+$ processes, normalized to the cross section at the Z resonance, as measured in pp collisions with a center of mass energy of $\sqrt{s} = 7\,\text{TeV}$ at CMS. Observed data are shown as black dots, the blue line represents the theory prediction. For invariant masses approaching $m_Z = 91.2\,\text{GeV}$, the cross section is enhanced by three orders of magnitude [13].

1.1.2 The Higgs mechanism in the Standard Model

The structure of the interactions in the SM is described by gauge symmetries, from which the SM derives its accurate predicting power using the classical Lagrange formalism [14]. Here, the three fundamental forces can be attributed to internal local gauge symmetries: A $SU(3)_C$ symmetry describes the structure of strong interactions, a $SU(2)_L$ and $U(1)_Y$ symmetry related to the weak isospin L and hypercharge Y respectively describe all electroweak interactions, leading to the SM as a gauge quantum field theory with the internal symmetries of $SU(3)_C \times SU(2)_L \times U(1)_Y$. Each symmetry requires the presence of massless gauge bosons, which is observed in reality for the electromagnetic and strong interactions and their massless gauge bosons, which are the photon and gluons. However, for the weak interaction, a problem arises due to the non-zero masses of the weak gauge bosons. The theoretical challenge is to keep the theory, which provides accurate results and is exact for the electromagnetic and strong interaction, invariant under gauge transformations while still explaining the masses of the W and Z bosons. This was accomplished by Englert, Brout and Higgs [3–5] via the process of *spontaneous symmetry breaking*. A self-interacting complex scalar field is introduced as a $SU(2)$ doublet $\phi = \begin{pmatrix} \phi_+ \\ \phi_0 \end{pmatrix}$ following

a Lagrangian of the form

$$\mathcal{L}^{\text{Higgs}} = \partial_\mu \phi^\dagger \partial^\mu \phi - V(\phi) \tag{1.1}$$

with a potential of the form

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \tag{1.2}$$

In case that $\lambda > 0$, the lower component of the field ϕ acquires a non-zero vacuum expectation value $v = \sqrt{\frac{\mu^2}{2\lambda}}$, leading to a spontaneous symmetry breaking of $SU(3)_C \times SU(2)_L \times U(1)_Y$ into $SU(3)_C \times U(1)_Q$, where the gauge symmetries relating to gluons and the photon remain unbroken and the electroweak gauge symmetry $SU(2)_L \times U(1)_Y$ does not. An expansion of ϕ around the minimum with a real field H in the lower component

$$\phi = \begin{pmatrix} 0 \\ v + \frac{H}{\sqrt{2}} \end{pmatrix} \tag{1.3}$$

then introduces couplings of the gauge fields of the $W^\pm$ and Z bosons to the quantum vacuum, from which they acquire masses. From the four degrees of freedom of the complex doublet field ϕ, three get consumed by the $W^\pm$ and Z bosons, where they are manifested as additional longitudinal degrees of freedom. The remaining degree of freedom represents the new Higgs field H. The particle predicted by the theory is the scalar Higgs boson. After decades of preparation, the challenging search was successfully concluded in 2012 when a particle with the properties as expected for a SM Higgs boson was discovered at the Large Hadron Collider at CERN [1, 2].

1.2 The Higgs sector in the Minimal Supersymmetric Extension to the Standard Model

Supersymmetric models introduce a new type of symmetry transforming fermions into bosons, and vice versa. In supersymmetry, each particle receives a superpartner with equal quantum numbers and mass and a spin which differs by 1/2. As no light superpartners have been observed yet supersymmetry, if present, must be spontaneously broken [15]. Supersymmetry would solve open questions such as the fact that the mass of the Higgs boson is unstable against divergent radiative corrections, and requires unnatural fine-tuning in the Standard Model. Furthermore, supersymmetry allows for the possibility of a unification of the three gauge interactions, the couplings of which require supersymmetric modifications to unify at high energies.
The MSSM considers only the minimal amount of new particles necessary to realize supersymmetry. These are the known particles and their respective superpartners. The MSSM requires a second doublet in the weak isospin space in addition to the doublet of the SM.

$$\phi^{\text{SM}} = \begin{pmatrix} \phi_+^{\text{SM}} \\ \phi_0^{\text{SM}} \end{pmatrix} \qquad \phi^{\text{MSSM}} = \begin{pmatrix} \phi_0^{\text{MSSM}} \\ \phi_-^{\text{MSSM}} \end{pmatrix} \tag{1.4}$$

The presence of the two doublet fields leads to a total of eight degrees of freedom in the Higgs sector, three of which again get consumed by the masses of the weak gauge bosons, while the remaining five appear as five spin-0 Higgs bosons h, H, A, H^+, H^-. The lightest neutral scalar (CP-even) Higgs boson h is usually interpreted as the already-discovered 125 GeV Higgs boson of the SM. Of the two heavier neutral Higgs bosons H and A, H is scalar (CP-even), and A is pseudoscalar (CP-odd), and $H^\pm$ refers to two charged Higgs bosons. The second Higgs doublet introduces two different vacuum expectation values for up- and down-type fermions. While their quadratic sum is fixed by the known mass of the W boson, their ratio $\tan\beta = \frac{v_u}{v_d}$ is a free parameter of the MSSM. Large values for $\tan\beta$ would enhance the coupling of heavy Higgs bosons to down-type fermions such as the b-quark and τ-lepton, and is therefore one of the most relevant parameters when searching for signatures of H/A bosons decaying into τ-leptons.

1.3 The $H/A \to \tau\tau$ decay channel

This thesis will focus on the search for the two heavy neutral Higgs bosons H and A. As the analysis will not be sensitive to the CP eigenvalue of a possible Higgs candidate, they are referred to as H/A to represent any of the two heavy Higgs bosons, which can be degenerate in mass. For the reasons given below, the most promising decay channel of these particles is the decay into two oppositely charged τ-leptons

$$H/A \to \tau^-\tau^+ \ . \tag{1.5}$$

The first reason is that, with a mass of 1.78 GeV, the τ-lepton is the heaviest known lepton. As the coupling of a Higgs boson to fermions is proportional to their mass, the branching ratio of decays of a Higgs boson into two τ-leptons is significantly higher than into two muons or electrons. Second, while the branching ratio into heavy quarks such as b-quarks is even higher, the τ-lepton provides a better experimental accessibility compared to hadronically interacting quarks. Due to their color charge, quarks will form color-neutral jets which are difficult to distinguish from the quantum chromodynamic (QCD) multijet production in the same proton-proton crossing. A τ-lepton by contrast will decay into a lighter lepton - the muon or electron - or into light isolated hadrons, and is easier to identify using a particle detector such as the CMS detector. Details to the reconstruction of τ-leptons at CMS will be discussed in section 2.2.3. Electrons or muons are comparably easy to identify, as will be explained in detail in the next chapter. Hadronically decaying τ-leptons, which will be referred to as τ_h, decay mainly into one or three charged pions (or, less frequently, kaons) and up to two neutral pions. In di-τ events, the two τ-leptons decay independently, resulting in six exclusive di-τ final states

$$\tau\tau \to \tau_h\tau_h, e\tau_h, \mu\tau_h, e\mu, \mu\mu, ee \ .$$

The most promising final states of di-τ events are decays were one τ-lepton decays into a muon or electron and the second τ-lepton into an isolated jet, representing the two semi-leptonic final states $\mu\tau_h$ and $e\tau_h$. These combine an adequate branching ratio with a good separation from QCD background events and are therefore favorable from an

experimental point of view. The full-hadronic final state $\tau_h\tau_h$ is harder to separate from QCD background but makes up the largest fraction of di-τ decays, which is illustrated in figure 1.2. The search for a $H/A \to \tau\tau$ will therefore focus on these three final states.

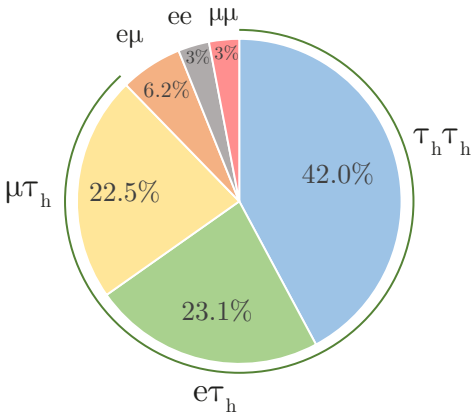


Figure 1.2: Branching ratios of individual di-τ final states [15]. The channels used for the MSSM $H/A \to \tau\tau$ analysis in chapter 6 are emphasized by the green circle.

As a heavy Higgs boson would decay instantly for all experimental purposes, they would manifest themselves as a resonance in the mass spectrum of di-τ events over a background of known processes depending on the di-τ final state.
For heavy neutral Higgs bosons to be produced, their mass has to be supplied as center of mass energy of a hard particle scattering. At the LHC, which is a proton-proton collider, pp-collisions are performed at a center of mass energy of 13 TeV. As the proton itself is composed of quarks and gluons, the proton energy is folded with the probability to find a parton (quark or gluon) with a momentum fraction x. The probability follows an energy-dependent parton density function, which has been determined e.g. in deep inelastic electron-proton scattering [16]. The search will be extended to two production modes of a heavy Higgs boson: The gluon-fusion, were two gluons interact via a virtual fermion loop to merge into a Higgs boson, and the b-associated production. Both production modes are shown in figure 1.3.

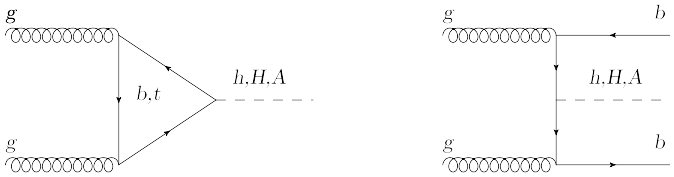


Figure 1.3: The two dominant production modes for neutral heavy Higgs bosons: Production via gluon fusion (left) and b-associated production (right).

The expected cross sections of the two production processes as well as the decay into τ-leptons depend on the choice of parameters for the model, mainly on the value of $\tan\beta$ and the mass of the pseudoscalar Higgs boson A.
As the analysis is not sensitive to the spin of a heavy Higgs boson, observed di-τ events can only be differentiated from $Z \to \tau\tau$ decays using the invariant mass. Especially for low- to mid-mass hypotheses of the new particle, the accurate description of $Z \to \tau\tau$ events belongs to one of the main challenges.

CHAPTER 2

The CMS Experiment at the LHC

The data analyzed in the context of this thesis were taken by the Compact Muon Solenoid (CMS) detector [17], a general-purpose detector located at the Large Hadron Collider (LHC) [18] at CERN in Geneva. The LHC is a 27 km-long ring consisting of superconducting magnets as well as accelerating structures to boost particles to the high energies needed for the gain of new knowledge about the fundamental constituents of the universe. Two beams of protons are accelerated in opposite directions along the ring, guided by strong magnetic fields, before being brought to collision. The center-of-mass energy achieved for proton-proton (pp) collisions in 2016 was 13 TeV. In addition to achieving these high energies, one of the goals of the LHC is to have a high rate of pp-collisions referred to as *instantaneous luminosity*, which is defined as the number of proton-proton collisions per cm^2 and second. In 2016, this luminosity reached record numbers of up to $\mathcal{L} = 1.5 \cdot 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, leading to an average number of 27 pp-collisions per bunch crossing, which occurs at a frequency of 40 MHz [19]. Of those 27, usually one corresponds to a hard scattering process that has the potential of revealing unknown phenomena, while most of the remaining collisions happening correspond to low energy scatterings. These are referred to as *pileup* and complicate the identification and reconstruction of the hard scattering process of interest, making the accurate description of these additional collisions important for the analysis of hard scatterings.

2.1 CMS detector design

One of the collision points of the LHC lies at the center of the CMS detector. Its main purpose is to study the SM and discover physics beyond it. The CMS experiment has been part of the discovery of the Higgs boson in 2012, and is now recording data at higher energies to test the predictions of the SM at even higher accuracy. The detector is build cylinder-symmetrically around the beam pipe in layers of specialized detector components. A view of the detector is shown in figure 2.1. The symmetric design aims to achieve a hermetic measurement of all particles emerging from the pp-collisions collisions in the plane transverse to the beam axis. Quantities used for analysis include the *transverse momentum* p_T, which is the projection of the total momentum $\vec{p}$ onto the transverse plane, the *azimuthal angle* ϕ and the *pseudorapidity* $\eta = -\log(\tan\frac{\theta}{2})$, where θ is the polar angle.

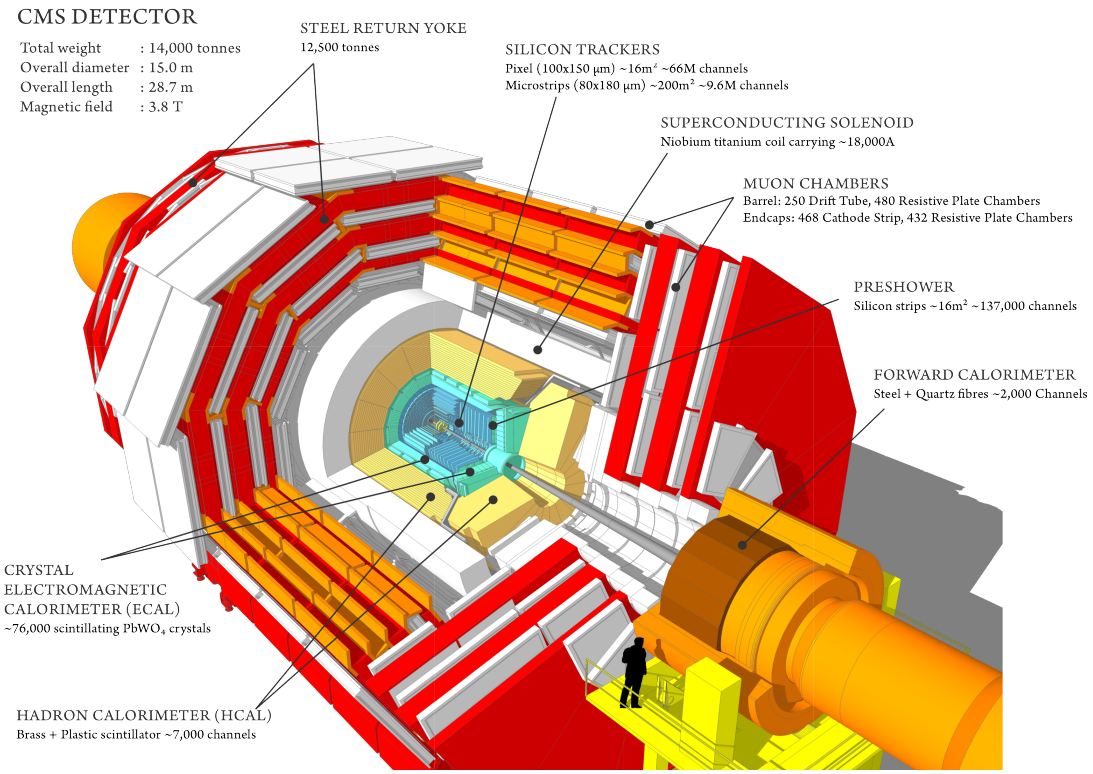


Figure 2.1: Sectional view of the CMS detector. The detector is designed as cylinder-symmetric layers of specialized detector components around the beam pipe of the Large Hadron Collider in which the proton-proton collisions take place [20].

The layers in the CMS detector can be summarized to five main components [21]:

1. The **silicon trackers** allow the reconstruction of accurate tracks of charged particles. For a measurement of the transverse momentum of particles, the CMS detector detects curved trajectories of charged particles in the magnetic field generated by the solenoid magnet described below. For this, CMS utilizes a system of 65 million silicon pixels in the innermost tracking system, followed by ten layers of silicon strips, extending the tracking system to a radius of 1.3 m from the beam pipe. The semi-conducting pixels and strips produce tiny electric signals when transversed by a charged particle, which will then be amplified and recorded to gain a position measurement of the hits with a resolution of 10 µm.

2. The **crystal electromagnetic calorimeter** (ECAL) contains almost 80,000 lead tungstate crystals and allows the measurement of the energy of electromagnetically interacting particles such as photons or electrons. The crystals serve as a calorimeter due to its scintillating behavior, which is the production of light when being transversed by particles. The material has been chosen due to the high density and therefore high stopping power of particles. The crystals of the ECAL are 23 cm

long and contain 28 radiation lengths for electromagnetic showers. The energy of electrons and photons is then measured as the number of scintillated photons.

3. The **hadronic calorimeter** (HCAL) measures hadronically interacting particles, i.e. particles that are composites of quarks and gluons. In contrast to the ECAL, the HCAL uses sampling layers, consisting of alternating layers of high-density brass absorber and plastic scintillator material. Due to the larger interaction length of hadrons in materials, the construction of an HCAL inside the solenoid was a challenge, and with a radius of one meter, the minimum thickness was achieved while still containing ten absorption lengths in the calorimeter.

4. The **superconducting solenoid magnet** lies around the inner tracking and calorimetry systems and provides a strong magnetic field of 3.8 T which is parallel to the beam axis and therefore causes particles to bend along their movement in the transverse plane, allowing the measurement of their transverse momentum. The strong magnetic field is only possible due to its superconducting properties, for which it is cooled to a temperature of 4.6 K.

5. The **muon chambers** qualify the CMS detector for one of its premier tasks, which is the accurate reconstruction of muons. These charged leptons have a high penetrating power in materials and are stopped by neither the electromagnetic nor the hadronic calorimeter. The muon chambers are integrated between the iron return yoke around the solenoid magnet and register hits of muons due to the ionization of gas in a drift tube. The outer magnetic field of the solenoid magnet curves the muon tracks to an S-shaped curve in the transverse plane when they transverse the detector, the curvature of which enables the measurement of muon momenta at enhanced accuracy for high-energy muons.

In the energy regimes relevant for the analysis in this thesis, the CMS detector delivers a momentum resolution of a single charged particle track of $\frac{\sigma(p_T)}{p_T} \approx 0.5\%$ at 10 GeV [14]. As the momentum resolution is proportional to the curvature of the track $1/R$, the relative uncertainty decreases linearly with the strength of the magnetic field used, yet increases with the momentum $\frac{\sigma(p_T)}{p_T} \propto \frac{p_T}{B}$. The energy resolution of the ECAL is $\frac{\sigma(E)}{E} \approx 1\%$ for electrons and photons with an energy of 30 GeV. In contrast to the measurement of momentum, which is inferred from the curvature of a track, the energy measurement is made by counting the resulting products of the particle shower occurring when the electron or photon enters the ECAL. These increase linearly with the energy, resulting in an improving energy resolution for higher-energy particles following the uncertainty of a Poisson distribution $\frac{\sigma(E)}{E} \propto \frac{1}{\sqrt{E}}$. The energy resolution of the HCAL is considerably larger with $\frac{\sigma(E)}{E} \approx 10\%$ for a charged pion at an energy of 100 GeV due to the longer interaction length of hadrons in material compared to electrons or photons, as well as the occurrence of neutral long-living hadrons such as the neutron or the K_L^0 carrying away energy of the initial particle undetected.
In figure 2.2, a longitudinal slice in the z-R-plane of the CMS detector is shown, highlighting the position of the detector components with respect to the radius R and the

pseudorapidity segment η. The sensitive components cover the range of $|\eta| < 2.5$ for silicon tracker and muon chambers, and $|\eta| < 3.0$ for the ECAL and HCAL. An additional Cherenkov forward calorimeter extends the coverage to $|\eta| < 5$. A coverage to higher η values is impossible due to the beam pipe guiding the proton beams. At around $\eta = 1.0$, the detector is divided into the *barrel region* covering the region $|\eta| < 1.0$ and the *endcap* region covering higher $|\eta|$-values.

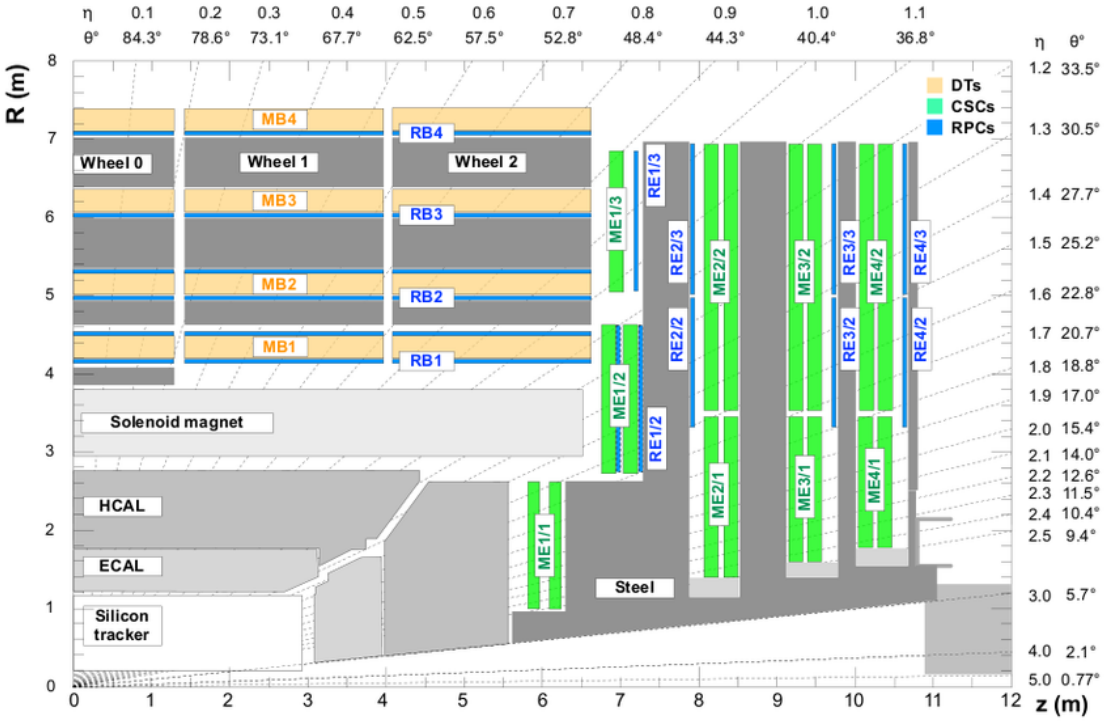


Figure 2.2: Longitudinal slice of one quadrant of the CMS detector. From [22].

2.2 Event reconstruction at CMS

2.2.1 Trigger systems

The rate at which interactions occur at the LHC is 40 MHz, meaning data needs to be read out 40 millions times per second. As reading out the raw event data of each component would result in an unmanageable data flow, *triggers* reduce the amount of data that is written to tape. Only the rare events that have the potential to contain new physics are stored. On a first level, the data flow is reduced by a hardware-based *L1 trigger* to the most interesting 100,000 events per second by utilizing the information from the calorimeter and muon chambers only. This is then further reduced by a *high level trigger*, which has access to the full event data and can afford a longer time to process the event due to the initial reduction of event rate. In the high level step, hits in the

muon chambers can already be matched to tracks in the silicon tracker and a dedicated software runs a check for specific signatures with the potential of containing new physics. Finally, around 300 events per second are stored on tape and used by analysts.

2.2.2 The particle flow algorithm

A variety of particles can emerge as measurable objects for the CMS detector: photons, electrons, muons as well as a variety of charged and neutral hadrons. The challenge is to match all independently measured energy deposits in the calorimeters to the tracks in the silicon detector and muon chambers. As soon as a complete list of all reconstructed particles is available, further quantities such as the identification of jets corresponding to the initial seeding particle can be derived. At CMS, this is done using the *particle flow* algorithm [23]. The algorithm starts by using the tracks in the inner tracker, which are then extrapolated to the ECAL and HCAL and into the muon chambers. This enables the association of energy deposits in the calorimeters to the respective tracks. The matching of a track to energy deposits in both ECAL and HCAL results in the reconstruction as a charged hadron, while energy clusters that are not linked to a track determines the reconstruction as photon (if linked to clusters in the ECAL) or neutral hadrons (if linked to cluster in both ECAL and HCAL). Tracks that can be associated with a track in the muon system determine a muon candidate, the known properties of which can then be used to derive a best estimate of the momentum of the full track corresponding to the muon candidate. Electrons will be identified as tracks in the silicon tracker that can be linked to energy clusters in the ECAL, but not in the HCAL. A slice of the CMS detector with the tracks and respective particle showers in the calorimeters is shown in figure 2.3.

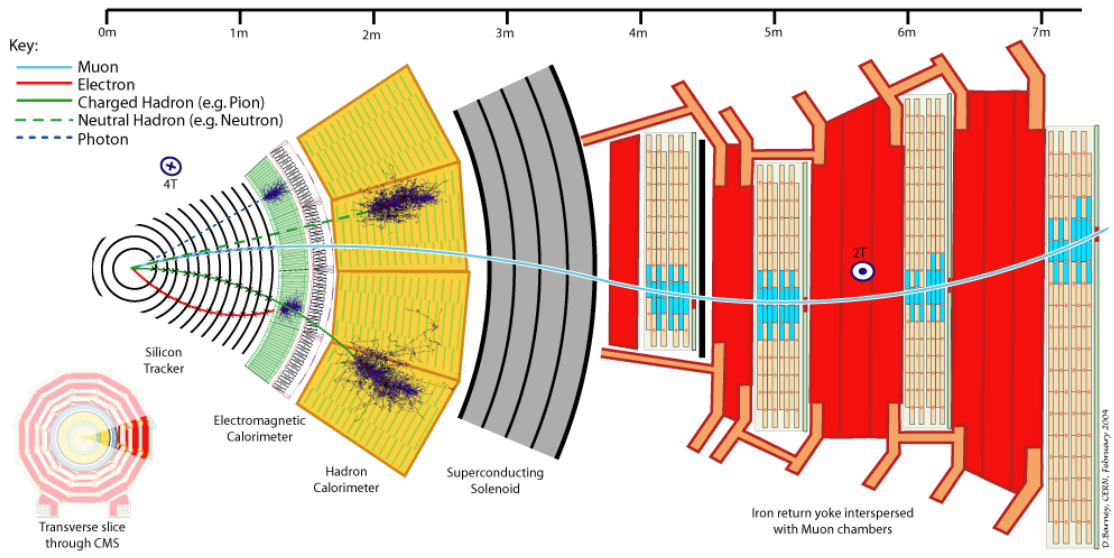


Figure 2.3: Transverse slice of the CMS detector. The tracks of various particle flow candidates and sketches of their showering behavior in the ECAL and HCAL are shown. Figure taken from [24].

As the reconstruction and identification of leptons will be of particular importance for this thesis, a detailed description will be given in the next section. All particles reconstructed by the particle flow algorithm are unambiguously classified and can be linked to a collision vertex. The vertex where the sum of p_T of all associated particle flow candidates is highest is referred to as the primary vertex of the hard interaction. Additional vertices are referred to as pileup, describing additional pp-collisions with lower exchanged transverse momentum.
The particle flow candidates can then be clustered to *jets*, which is for this analysis done with the use of the anti-k_t clustering algorithm with an opening angle of 0.4 [25]. The reconstructed jets then enable the insight of the respective seed of the jets: As quarks or gluons will hadronize before being reconstructed, their existence and momentum can only be inferred to via the reconstruction and momentum of the product of their hadronization, the jet.

2.2.3 Reconstruction of leptons

Muons

The reconstruction of muons is robust against misidentification of other particles as muons, as the requirement of a hit in the outer muon chambers strongly suppresses possible candidates for misidentification. Muons are at first reconstructed independently in two approaches [26]: In the outside-in approach, a standalone muon track is found in the muon system alone, which is then matched to a track in the inner tracker. These muons are then reconstructed as *global muons*. The global fit can improve the momentum resolution of the muon mainly for high-energy muons with $p_T > 200\,\text{GeV}$. In the inside-out approach, all tracks that fulfill a minimum p_T-requirement of $p_T > 0.5\,\text{GeV}$ are considered as muon candidates and are extrapolated to the muon systems. If at least one matching short track stub can be found, the muon qualifies as a *tracker muon*. The reconstruction as tracker muon is more efficient for low-energy muons, as only a single segment in the muon system is required. By combination of the two approaches, the reconstruction of muons with the CMS detector has a very high efficiency of over 99% for muons with $p_T > 5\,\text{GeV}$, giving motivation for the choice of naming of the Compact *Muon* Solenoid. Most muons are in fact reconstructed as both global and tracker muons and therefore merged into a single muon candidate. Muons that are only reconstructed as tracks in the muon system but not in the silicon tracker are discarded as muon candidates due to a possible contamination by cosmic ray muons. Further reduction of both misidentification of particles other than muons, and of muons resulting from decays of other particles in flight are made by setting requirements on the quality of the global muon fit using its χ^2 variable, as well as the restriction of the transverse distance of the reconstructed muon vertex from the primary vertex to $|d_{xy}| < 0.2\,\text{cm}$, both of which only marginally decrease the efficiency of the muon reconstruction. For the muon selection shown in this thesis, working points of the muon identification with an efficiency of around 99% are used.

An additional important way to identify muons from decays such as the $Z \to \tau\tau$ or $H/A \to \tau\tau$ decay as opposed to muons produced by the decay of heavy quarks or hadrons is the requirement of an *isolated muon*. In this thesis, the isolation of muons will be calculated using the particle flow relative isolation I^{μ}_{rel}. It is defined as the sum of all transverse momenta (for charged particles) and transverse energies (for neutral particles), divided by the transverse momentum of the muon. Only particles in a predefined cone in the η-ϕ-plane with radius $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ around the direction of the muon are considered. The optimization of the cone size ΔR results in a value of $\Delta R = 0.4$. The transverse momentum of the muon itself is excluded from the sum. Charged particles are only taken into account if their tracks are associated to the primary vertex of the collision to exclude distortions from particles associated to pileup. As neutral particles do not leave tracks in the silicon tracker, their attributed pileup contribution is estimated using the sum of transverse momenta from all pileup-attributed charged particles. This is then corrected with a factor of 0.5 accounting for the expected fraction of neutral to charged hadrons and then subtracted from $\sum E_T^{\text{neutral}}$, with a maximum of subtracted energy being $\sum E_T^{\text{neutral}}$ itself. The definition of I^{μ}_{rel} is therefore given by

$$I^{\mu}_{\text{rel}} = \frac{1}{p^{\mu}_T} \cdot \left(\sum p_T^{\text{charged,PV}} + \max\left(0, \sum E_T^{\text{neutral}} - 0.5 \cdot \sum p_T^{\text{charged,PU}}\right)\right) \tag{2.1}$$

The abbreviation PV refers to particle flow candidates whose tracks are associated with the primary collision vertex, PU refers to pileup-associated candidates. A usual requirement for a selection of isolated muons is $I^{\mu}_{\text{rel}} < 0.15$.

Electrons

Electrons are reconstructed by the particle flow algorithm by an association of a track in the silicon tracker to a cluster of energy in the ECAL. Due to the low mass of electrons compared to the heavier leptons, the radiation of bremsstrahlung is greatly enhanced for electrons, resulting in $33 - 86\%$ [27] of the electron energy being radiated as photons before the electron reaches the ECAL, depending on the amount of material between the primary vertex and the ECAL and therefore the pseudorapidity of the electron track. It is therefore essential to collect the energy of the radiated photons when calculating the initial energy of the electron. This is done by collecting the energy deposits in the ECAL in a window in the η-ϕ-plane, which is larger in the ϕ-direction, where most of the radiation occurs due to the bending of the electron in the magnetic field in the transverse plane. The ECAL crystal with the highest E_T is then used as a seed for the generation of a cluster in the ECAL.
A final identification of electrons is done in an multivariate (MVA) approach using a boosted decision tree (BDT) to identify possible signal electrons, e.g. from a τ decay, from background such as photon conversions, jets misidentified as electrons or electrons produced in quark decays. This method results in a final electron discriminator with an identification efficiency ranging from 80% (*tight* working point) to 90% (*loose* working point), with a misidentification rate of up to 10%. Furthermore, isolated electrons can be identified using the particle flow relative isolation. The same equation as 2.1 applies

with the cone size chosen for electrons and a usual requirement being $\Delta R = 0.3$ and $I^e_{\text{rel}} < 0.10$ respectively.

Tau-leptons

As stated in section 1.3, τ-leptons decay into leptons with a probability of around 1/3 and into hadrons with a probability of around 2/3, where usually considered final states are one or three charged pions or, seldomly, kaons in addition to up to two neutral pions, clustered in an isolated jet. For τ-decays into leptons, reconstruction of the muon or electron candidates is done as explained in the previous paragraphs. Hadronically decaying τ-leptons (τ_h) pose an additional challenge of differentiating them from quark and gluon jets, which are abundant at a hadron collider. To avoid misidentification of these jets as τ_h, properties of the hadronic decay of the τ-lepton are exploited [28]: The number of hadrons in a τ_h-jet is small compared to usual quark and gluon jets, the lifetime of the τ-lepton is longer then the hadronization time of quarks and gluons, and τ_h-jets are usually more isolated and collimated. The reconstruction again starts from the complete list of particle flow candidates, where charged particles need to pass a minimum threshold of $p_T > 0.5\,\text{GeV}$. Neutral pions decay instantaneously in experimental terms into a pair of photons $\pi^0 \rightarrow \gamma\gamma$, and are reconstructed as strips from photons and electrons (from $\gamma \rightarrow e^- e^+$ pair production) within the seeding jet. Finally, three decay modes are considered that are labeled 'one-prong' , 'one-prong + π^0' and 'three-prong', *prong* referring to a charged hadron ($\pi^\pm$ or $K^\pm$). The final identification of hadronic τ-decays is then done via an MVA-based discriminator using a BDT, with input variables being the charged and neutral p_T-sums, the reconstructed decay mode of the τ_h, the impact parameter in the transverse plane (distance of the reconstructed jet seed to the vertex of the primary collision) and its significance, as well as the flight distance of the τ_h. The BDT is trained on various simulated τ-events such as $Z \rightarrow \tau\tau$ or $H \rightarrow \tau\tau$. The MVA approach then defines several working points of which a *tight* working point will be used for the hadronic τ-leptons in the $\mu\tau_h$ and $e\tau_h$ final states, and a *medium* working point for the $\tau_h\tau_h$ final state. The tight working point results in a very pure τ_h reconstruction, with a rate of quark or gluon jet misidentification of $4 \cdot 10^{-4}$ and a τ_h identification efficiency of 27%, whereas the medium working point provides an efficiency of 51% with a jet$\rightarrow \tau_h$ misidentification rate of $3 \cdot 10^{-3}$ [29]. The slightly looser selection in the full-hadronic decay channel is motivated by the requirement of two reconstructed τ_h-jets, as opposed to only one for the semileptonic final states. For the suppression of electrons and muons misidentified as τ_h, a MVA-based discriminant is used for electrons, and a cut-based discriminant is used for muons. Input variables of the anti-electron discriminant include the fraction of energy deposits in the ECAL over the total energy in ECAL+HCAL, the energy in the ECAL over the momentum of the leading charged particle of the τ_h candidate, as well as the energy fraction carried away by photons or the mass of the τ_h candidate. Discriminating variables for the cut-based separation of muons and τ_h are a veto of events with at least two hits in the muon system in a $\Delta R < 0.3$ cone around the τ_h track or with an energy fraction of the τ_h-track in the ECAL or HCAL of less then 0.2. A tighter separation is chosen by the requirement of no hits in the muons

systems in the $\Delta R < 0.3$ cone. The working points used for the anti-electron discriminant range from *very loose* (efficiency 94%, $e \to \tau_h$ misidentification rate $2 \cdot 10^{-2}$) to *medium* (efficiency 85%, $e \to \tau_h$ misidentification rate $1 \cdot 10^{-3}$) and *tight* (efficiency 78%, $e \to \tau_h$ misidentification rate $6 \cdot 10^{-4}$).
The efficiencies and misidentification rates for the suppression of muons misidentified as τ_h range from *loose* (efficiency 99.5%, $\mu \to \tau_h$ misidentification rate $5 \cdot 10^{-4}$) to *tight* (efficiency 98%, $\mu \to \tau_h$ misidentification rate $3 \cdot 10^{-4}$) [29].

CHAPTER 3

Description of the embedding method

The embedding method is a data-driven procedure to derive an estimate of $Z \to \tau\tau$ events. As has been explained in section 1.3, $Z \to \tau\tau$ events are the most prominent source of background for the $\mu\tau_h$ and $e\tau_h$ final states when analyzing the decay of a Higgs boson in the di-τ channel. The goal of the embedding method is to describe this background as accurately as possible, while using the main parts of the event from data and only while relying on simulation for the decay of the Z boson into τ-leptons and the subsequent τ-decays. This method utilizes on *lepton universality* explained in section 1.1, which refers to the equal couplings of leptons to the Z boson. The subtraction of the two muons makes a $Z \to \mu\mu$ event indistinguishable from a $Z \to \tau\tau$ event with removed τ-leptons. Hence, a $Z \to \mu\mu$ event where the muons are replaced for τ-leptons provides a description of the $Z \to \tau\tau$ event.
For the production of embedded events, events with two muons resembling those of a Z decay are selected from data, and the two muons and calorimetry entries corresponding to them are cleaned from the event. The reconstructed kinematics of the muons are then used to simulate a decay of a Z boson into two τ-leptons with the same kinematics. The cleaned and simulated events are, in a next step, merged to create the hybrid $\mu \to \tau$ embedded event which will be used as a background estimate for analyses in the di-τ channel. Figure 3.1 illustrates the steps that will be explained in detail in the following. In the first run period of the LHC, the embedding method was successfully used for the discovery of the Higgs boson in data taken at CMS at a center of mass energy of 7 and 8 GeV [30]. Since then, the method has been newly developed for the changed run conditions for the Run II-period scheduled from 2015 to 2018. The changed conditions include not only the higher center of mass energy of 13 TeV and increased pileup from an average number of pp-collisions of 14 in Run I to 27 in Run II [19], but also essential developments in the reconstruction software of CMS. The latter were e.g. needed for processing of the increased flux of data given the shorter spacing of the proton bunches in the LHC. Important progress on the redevelopment has been made in [31]. Open questions include the need for a closure study using simulated $Z \to \mu\mu$ events for an increased understanding of the effects of the muon selection, event cleaning and particle generation, as well as the development of a dedicated uncertainty model.

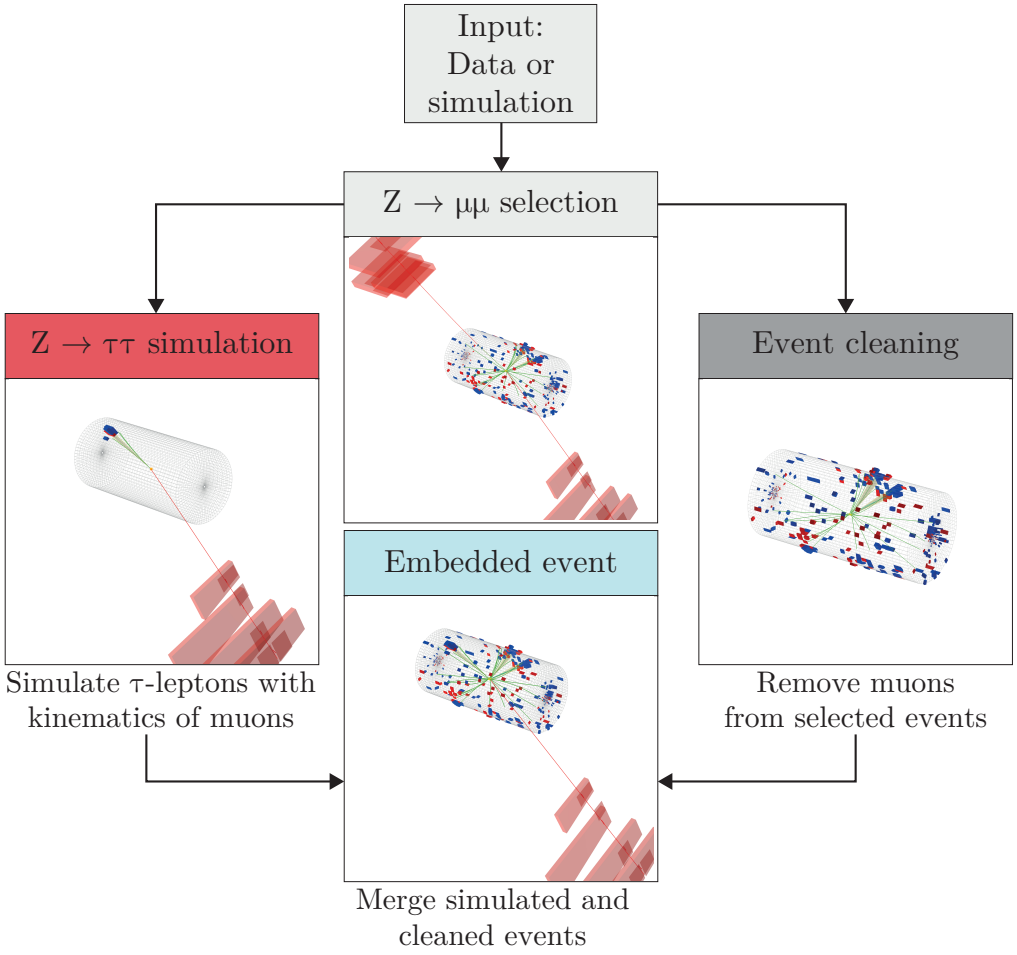


Figure 3.1: Visualization of the four steps applied for the creation of $\mu \to \tau$ embedded events. $Z \to \mu\mu$ events are selected from data. The muon tracks and their footprints are cleaned from the event, and a $Z \to \tau\tau$ decay is simulated using the kinematics of the initially selected muons. The cleaned and simulated events are merged to form a hybrid embedded event which can be used to describe $Z \to \tau\tau$ events at CMS. The example shows the steps to create an embedded event in the $\mu\tau_h$ final state.

Moreover, in the scope of this thesis, the need for corrections has been identified for

1. the change in efficiencies to pass identification, isolation and trigger requirements of the simulated leptons, which can differ from the efficiencies for leptons in data or simulation,
2. reconstruction effects during the initial muon reconstruction,
3. the contamination by di-muon events that are selected from decays of top-quark pairs into muons instead of $Z \to \mu\mu$ events.

One of the main advantages of embedded over simulated $Z \to \tau\tau$ events is the representation of the main parts of the event. Jets resulting from the underlying event of the hard scattering as well as additional collisions are reconstructed from data. As the $Z \to \mu\mu$ selection and an analysis in the di-τ channel make use of the same reconstruction methods w.r.t. jets, no misdescriptions are expected from this. The selection of $Z \to \mu\mu$ events represents, on the one hand, an exclusive sample independent from the selection of $Z \to \tau\tau$ events used for the analysis of di-τ final states and, on the other hand, a sideband region for the search for Higgs bosons. The latter is given by the fact that the decay of the Higgs boson into muons is suppressed due to the coupling of the Higgs boson to the mass of the final state lepton. Experimentally, muons are the particle of choice for the embedding method as they are easy to reconstruct using the CMS detector, have a high energy resolution and interact minimally with the detector material, which minimizes the amount of calorimetry remnants that need to be cleaned.

3.1 $Z \to \mu\mu$ selection

The data used for the selection of di-muon events for the production of embedded events in the scope of this thesis are the full $35.9\,\mathrm{fb}^{-1}$ of data at a center of mass energy of 13 TeV taken by the CMS detector in 2016. To be selected, an event recorded by CMS has to fulfill the following requirements in a specific order:

1. An event has to pass the CMS high level trigger constructed for di-muon events[1]. It is designed to trigger on all events with two loosely isolated muons with a p_T of 17 GeV (8 GeV) for the leading (trailing)[2] muon. The distance of the event vertex to primary vertex along the beam axis must be $|d_z| < 0.2\,\mathrm{cm}$.
2. In the event, two muons must be reconstructed with a respective transverse momentum $p_T > 8\,\mathrm{GeV}$ and a pseudorapidity $|\eta| < 2.5$.
3. A minimum of one Z boson candidate for the two muons with a minimum mass of the di-muon system $m_{\mu\mu} > 20\,\mathrm{GeV}$, a zero charge $q_{\mu_1} + q_{\mu_2} = 0$ and a reconstructed transverse momentum of $p_T > 17\,\mathrm{GeV}$ of the leading muon of the $Z \to \mu\mu$ pair need to be found. If additional $Z \to \mu\mu$ candidates are found, the di-muon pair with the closest invariant mass to the Z mass is used.

[1] `HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ` or `HLT_Mu17_TrkIsoVVL_TrkMu8_TrkIsoVVL_DZ`
[2] Leading (trailing) muon refers to the muon with higher (lower) p_T of the di-muon event.

4. The two muons that are matched to the Z boson candidate must be reconstructed as global as defined in section 2.2.3, referring to a reconstruction as a standalone track in both the muon systems and the inner tracker.

If an event passes these requirements, it is used as a $Z \to \mu\mu$ candidate for the production of embedded events. Events that do not meet these requirements are discarded. The selection is then applied to the seven individual run periods at CMS labeled Run2016B-Run2016H, which make up the full integrated luminosity of the 2016 runtime of the LHC collected at CMS. An overview of the composition of individual run periods and the corresponding number of di-muon events selected for the creation of embedded events can be seen in table 3.1.

Table 3.1: Selected $Z \to \mu\mu$ candidate events used for production of embedded event samples.

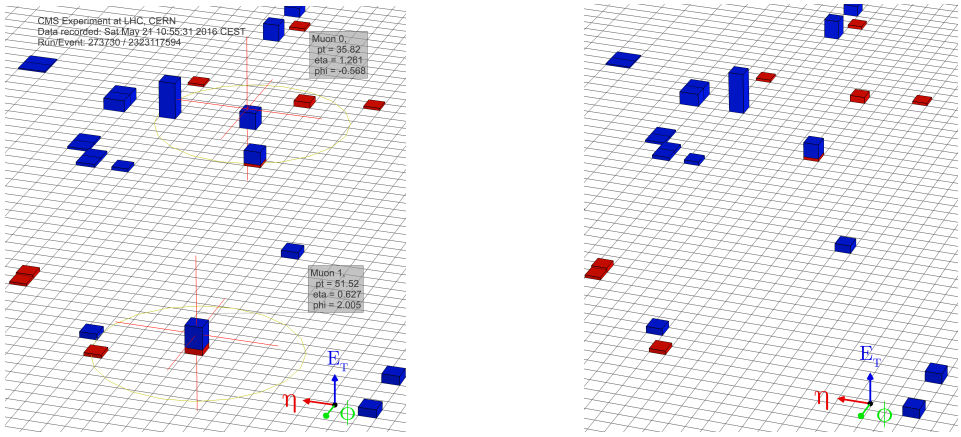
Run label	Luminosity in fb^{-1}	Number of events	Events / fb^{-1} in thousands
Run2016 B	5.71	6,086,165	1,065
Run2016 C	2.57	2,629,705	1,023
Run2016 D	4.24	4,295,319	1,013
Run2016 E	4.03	3,954,476	981
Run2016 F	3.11	2,969,454	955
Run2016 G	7.58	8,046,574	1,061
Run2016 H	8.65	9,712,482	1,123
Total	35.87	37,694,175	1,051

A misidentification rate of di-muon events from processes different from $Z \to \mu\mu$ decays, most notable top-quark pairs decaying into muons, is possible and expected at $\mathcal{O}(1\%)$ of the selected events. A minimization of the contamination by other processes could be achieved by requiring a tight region of the invariant di-muon mass around the mass of the Z boson. In the context of this thesis, a different approach is chosen that considers the embedded events from processes other than $Z \to \mu\mu$ decays as valid backgrounds for di-τ events, and corrects the resulting contamination by a veto of di-τ events in other background descriptions. The correction for contamination by top-quark pairs in embedded samples is explained in section 4.3.

3.2 Event cleaning

Muons in the CMS detector are reconstructed starting from a track in the inner silicon tracker matched to the outer muon tracker. Calorimetry entries are then assigned to the muon by extrapolating the track through the electromagnetic and hadronic calorimeters. As the clusters in the inner and outer tracking system used for the reconstruction of the muon track are known at this stage, the cleaning of the muon track on this level can be achieved by deleting all tracker hits that were used to reconstruct the muon candidate. In a second step, the corresponding entries in the electromagnetic and hadron calorimeter are removed. If the muon is well isolated and all entries in the vicinity of the muon track are only crossed by the track, this calorimetry cleaning can be done unambiguously,

leaving no remnants in neither tracker nor calorimeters. The cleaning of a well-isolated di-muon event is shown on the level of reconstructed energy towers in the η-ϕ-plane in figure 3.2.

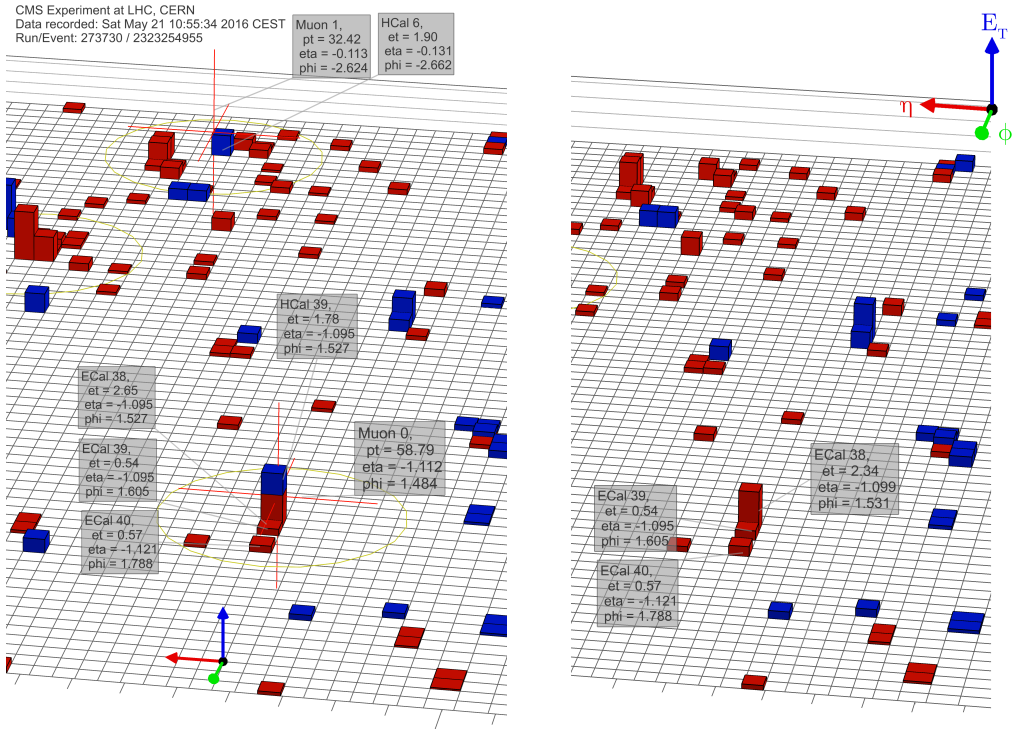


(a) Selected Event. **(b)** Cleaned Event.

Figure 3.2: Example of the cleaning of a well-isolated di-muon event in the η-ϕ-plane. Entries in the electromagnetic calorimeter are shown as red towers, entries in the hadronic calorimeter as blue towers. The two muons are shown as red crosses. The entries of the muons in the tracker as well as the energy deposits of the muons in the hadron calorimeter are removed completely from the event.

Ambiguities in the affiliation of the reconstructed hits in the calorimeter to the muon candidates are introduced when energy deposits in the vicinity of the muon are caused by jets, photons or electrons. In these cases the energy associated with the muon is not definitely known. This ambiguity can cause inefficiencies in the cleaning, which lead to remnants of the muon footprint remaining in the event as shown in figure 3.3. In this example, the muon track is surrounded by energy deposits in the electromagnetic calorimeter. These energy deposits can be caused by the muon, or a photon or hadron in vicinity of the muon track. The cleaning removes energy deposits crossed by the muon track; ambiguous deposits in an area around it, however, remain. Inefficiencies in the cleaning step can lead to the reconstruction of additional particle flow candidates, as will be discussed in a closure study using simulated $Z \to \mu\mu$ events in section 5.1.1. Two possible approaches to improve the performance of the event cleaning can be pursued: Firstly, energy remnants in the ECAL surrounding the muon but not crossed by the muon track are often attributed to final state radiation of the muon in the inner tracker. This is not covered by the muon selection and event cleaning, and leads to a misdescription of the muon energy as well as the number of reconstructed photons in the event. A straight-forward way to avoid this misdescription is to apply a selection requirement vetoing events in which a photon in the vicinity of the muon can be found.
Second the amount of energy cleaned can be extended from entries crossed directly by the muon path to an area surrounding the muon in a predefined cone in the η-ϕ-plane to find an optimal value of energy deposits to be cleaned. The figure of merit of this

optimization is the transverse energy flow around the reconstructed muon tracks. The impact of the embedding method on the energy flow around the embedded objects is studied in section 3.3.



(a) Selected Event. **(b)** Cleaned Event.

Figure 3.3: Example of the cleaning of a di-muon event in the η-ϕ-plane for an event with ambiguous energy deposits around the muon tracks. Electromagnetic calorimetry entries are shown as red, hadronic calorimetry entries as blue towers. The selected di-muon event is shown on the left; the cleaned event on the right. The two muons are marked as red crosses. In the cleaned event, the tracks of the muons are completely cleaned by removing the corresponding hits in the silicon tracker and muon system. The recorded energy in the calorimeters crossed by the muon tracks is removed. Entries in vicinity of the muon, which are not unambiguously attributed to the muon track, remain in the event.

3.3 $Z \to \tau\tau$ simulation

The Z boson candidate reconstructed from the selected muons is used to simulate a decay into two τ-leptons, which further decay into one of the six di-τ final states: $\tau\tau \to \tau_h\tau_h, \mu\tau_h, e\tau_h, e\mu, \mu\mu, ee$. The branching ratios of these final states are given in figure 1.2. To avoid producing a large number of embedded $Z \to \tau\tau$ events that will not pass the criteria imposed in later stages of the analysis, a filter is implemented during the generator step of the embedded event production. The purpose of the filter is to produce exclusive samples for each desired di-τ final state, and to have a high number of embedded events whose kinematics are in regions of interest for the analysis in the di-τ channel. Furthermore, migration effects of the initial muon selection at the boundaries of the imposed muon p_T requirements of 17 GeV and 8 GeV for the leading and trailing muon respectively need to be avoided. The production of embedded events for this thesis was done for the four most sensitive di-τ final states: $\tau_h\tau_h, \mu\tau_h, e\tau_h$, and $e\mu$; the respective p_T- and η-conditions on the generated τ-decay products are summarized in table 3.2. As an additional advantage, enforcing an exclusive decay channel enables the use of all selected $Z \to \mu\mu$ events repeatedly for all four channels. The repeated use of the selected $Z \to \mu\mu$ events increases the number of available events in the embedded event samples.

Table 3.2: Conditions imposed in the generator step. The conditions are chosen as slightly looser than the baseline event selection of the respective final state which will be given in chapter 6 to avoid migration effects at the boundaries. By construction, all embedded events produced fulfill these criteria.

Channel	Transverse momentum		Pseudorapidity	
$\mu\tau_h$	$p_T^{\mu} > 18\,\text{GeV}$	$p_T^{\tau_h} > 18\,\text{GeV}$	$\lvert\eta^{\mu}\rvert < 2.2$	$\lvert\eta^{\tau_h}\rvert < 2.4$
$e\tau_h$	$p_T^{e} > 23\,\text{GeV}$	$p_T^{\tau_h} > 18\,\text{GeV}$	$\lvert\eta^{e}\rvert < 2.2$	$\lvert\eta^{\tau_h}\rvert < 2.4$
$\tau_h\tau_h$	$p_T^{\tau_h} > 38\,\text{GeV}$		$\lvert\eta^{\tau_h}\rvert < 2.2$	
$e\mu$	$(p_T^{e} > 16\,\text{GeV}$	$p_T^{\mu} > 8\,\text{GeV})$	**or** $(p_T^{e} > 11\,\text{GeV}$	$p_T^{\mu} > 16\,\text{GeV})$

The filter is implemented in a way that firstly sets prior conditions on the final state as well as kinematic thresholds depending on the final state, and finally repeats the decay of the two simulated τ-leptons N times, at which $N = 1000$ has been chosen. The last event in the desired final state fulfilling the kinematic conditions is kept, which is equivalent to storing a randomly passing event. Imposing these criteria naturally introduces a bias as the prediction of $Z \to \tau\tau$ decays will be shifted towards τ decay products that are more central and have higher p_T. To correct for this bias, *generator weights* are stored in embedded samples. They contain the probability of this specific event passing the requirements, calculated by counting the number of passing events P during the N trials, and assigning the weight $g = \frac{P}{N}$ to the last passing event which is then stored. Since the choice of the final state is explicitly contained in the filter, the generator weights implicitly contain the branching ratio of the respective final state. Without additional kinematic criteria, the fraction of decays of the simulated τ-leptons is the branching ratio of the same final state, e.g. 22.5% for the $\mu\tau_h$ final state. By imposing further conditions however, less events will be eligible, shifting the weights to lower values. The generator

weights for all final states are shown in figure 3.4. They depend on the transverse momenta and pseudorapidity of the leptons, as an event selected from high-p_T di-muon event reconstructed in the barrel region of the CMS detector will have a higher probability of passing the kinematic thresholds. To illustrate this dependence, number of embedded events and the respective generator weights as a function of the transverse momenta of the two τ decay products is shown in figure 3.5 as an example for the $\mu\tau_h$ final state. The distributions for the additional final states can be found in A.1. The majority of events are selected from di-muons events in which both muons have a p_T of less than 50 GeV. It is possible that the found Z boson candidate during selection will not decay in an eligible event in the final state even after repeating the simulation. In this case, the selected event will be discarded, reducing the number of events in the final state from the total number of selected events. The discarding of a selected $Z \to \mu\mu$ event reflects that the reconstructed Z boson of the di-muon event does not constitute a background candidate for the description of a di-τ event, e.g. due to its high boost along the beam axis or low transverse momentum of the Z decay products. The efficiencies of the procedure applied to the respective final states as well as the number of events per final state are given in table 3.3. For all final states, the number of embedded events available per final state and for the kinematic regime required is higher then the number of expected $Z \to \tau\tau$ events in data after the same selection, which provides excellent statistics for background estimation.

Table 3.3: Number of embedded events and efficiencies of the procedure for the four final states. The right-most column gives the remaining number of events per final state. Due to the tight requirements when selecting τ-leptons in the full-hadronic decay channel $\tau_h\tau_h$, there is a significant loss of events when creating embedded events in this final state. The loss reflects the fact that less $Z \to \tau\tau$ events are expected after typical event selections compared to other final states.

Channel	Efficiency	Number of embedded events
Selected		37,694,175
$\mu\tau_h$	62.86%	23,694,406
$e\tau_h$	59.37%	22,378,770
$\tau_h\tau_h$	22.18%	8,362,033
$e\mu$	76.10%	28,685,023

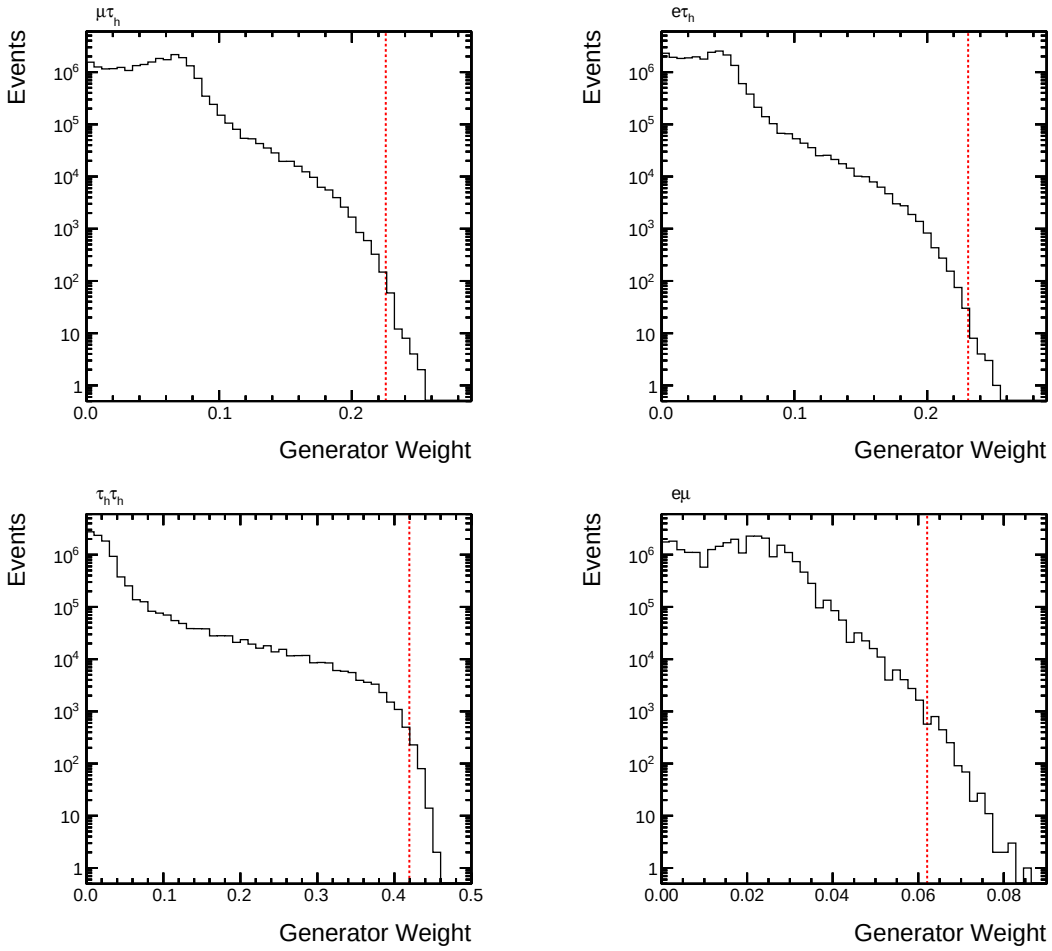
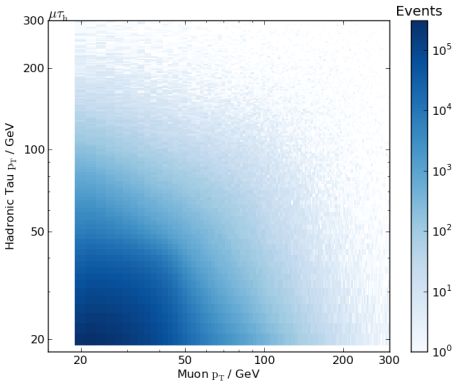
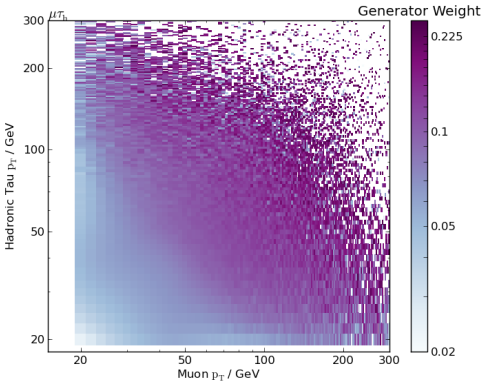


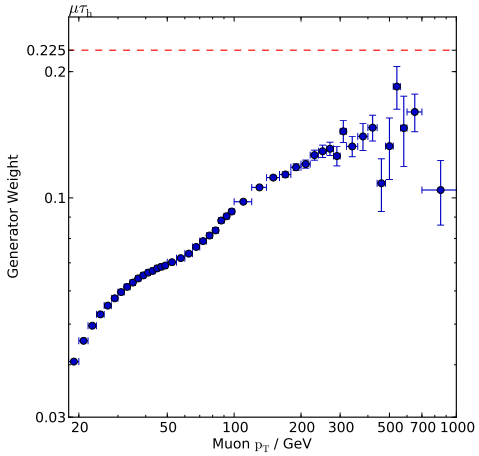
Figure 3.4: Generator weights in the di-τ decay channels. They give the probability of this specific event occurring in the kinematic environment of the selected $Z \to \mu\mu$ event. The generator weights are applied to avoid biases due to the imposed conditions on the embedded $Z \to \tau\tau$ decays in the simulation step. The red dotted lines correspond to the branching ratio of the final state as given in figure 1.2. Events having larger generator weights than the respective branching ratio are expected due to statistical fluctuations in the repetition of the simulation.



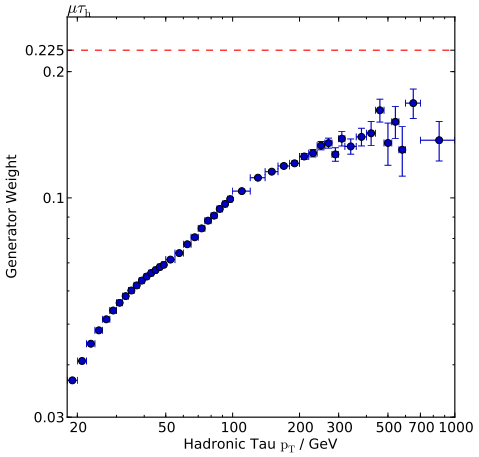
(a) Number of embedded events.



(b) Average generator weight as a function of the p_T of both τ-decay products.



(c) Average generator weight as a function of the p_T of the muon from τ-decay.



(d) Average generator weights as a function of the p_T of the hadronically decaying τ-lepton.

Figure 3.5: Upper row: Number of embedded events in the $\mu\tau_h$ final state as a function of the two transverse momenta of the τ decay products (left) and the respective average of the generator weights (right), reflecting the efficiency during the simulation step of the embedding method. Lower row: Average generator weight as a function of the individual transverse momenta. The distributions are shown inclusively over all η values.

3.4 Mirroring

Mirroring is a tool used to study effects introduced by the CMS object reconstruction as well as by the embedding method. It allows for the simulation of the τ-leptons not in the same direction as the removed muons, but in a direction that corresponds to a completely unbiased environment, while at the same time keeping the properties of the Z boson. Firstly, a plane spanned by the 3-momentum of the reconstructed Z boson with the beam direction is defined. The momentum of the selected muon $\vec{p}(\mu)$ is then decomposed in components parallel and perpendicular to this plane. The perpendicular component $p_T^{\perp}(\mu)$ is mirrored on the Z-beam plane, while the parallel component is not changed. This defines a new direction in which the simulation of the embedded lepton is performed. The pseudorapidity and transverse momentum are not changed by the procedure, whereas the azimuthal angle is changed. Yet, the distribution of events over the azimuthal angle follows a flat distribution and is not subject to selection requirements. A visualization of the mirroring procedure is shown in figure 3.6.

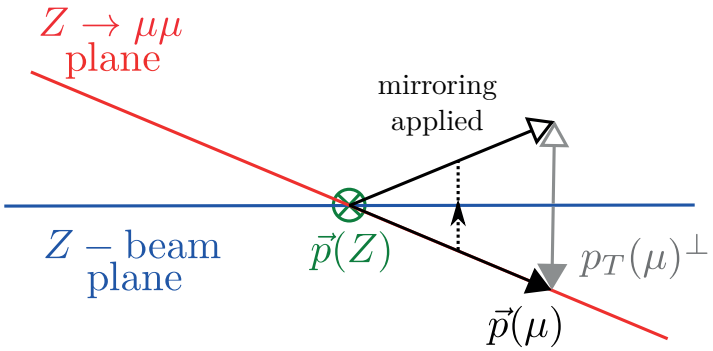


Figure 3.6: Illustration of the mirroring method in a view along the momentum of the reconstructed Z boson with the Z-beam plane appearing as a line. The momentum of the initially selected muon is shown as a black filled arrow. The component of the muon momentum perpendicular to the Z-beam plane is mirrored at the plane. The transformed direction is shown by the unfilled arrow. The p_T and η of the embedded object are not changed with respect to the selected muon; the azimuthal angle ϕ is. Figure taken from [31].

The mirrored objects are simulated in an environment that has neither been influenced by the initial selection nor by the initial reconstruction of the muons. Thus, both physical as well as reconstruction effects are removed. Effects that can be studied using mirroring include the selection bias introduced by imposing isolation criteria during the initial muon selection, effect of the objects reconstruction at CMS, and the impact of inefficiencies in the cleaning step that might change the reconstruction of embedded leptons. Mirroring will be used to study these effects using simulated $Z \to \mu\mu$ events in section 5.1.1.
The mirroring procedure is not applied for the τ-leptons simulated for $\mu \to \tau$ embedded events that are used for the analysis. Potential problems of using the mirroring procedure for the background description can arise due to the contamination of embedded events by

top-quark pair decays. Effects corresponding to a selection bias on the final reconstruction efficiency due to the isolation requirements imposed during the initial muon selection are small for muons. They are determined in the form of tag and probe scale factors specifically derived for the conditions and events used in this thesis, which will be discussed in section 4.1.

CHAPTER 4

Corrections for embedded events as $Z \to \tau\tau$ estimation

4.1 Lepton identification, isolation and trigger scale factors

The simulation of particle decays and their reconstruction at CMS is not a perfect description of data taken by the experiment. For simulated $Z \to \tau\tau$ events, usually scale factors are derived to make up for differences in event description between simulation and data. This concerns mostly the identification, isolation and trigger requirements imposed on particles, where the efficiency can be different depending on the kinematics and respective detector region where the particle is reconstructed. The embedding method aims to improve the understanding of data by avoiding the simulation of a range of complex variables, such as additional collisions, the underlying event of the hard collision, and jet kinematics. The decay of the Z boson itself into two leptons still has to be simulated.
Therefore it is expected that the acceptances of leptons of embedded events is equal to leptons in simulated events. Differences can occur due to a selection bias of embedded events, where the initially selected muons already fulfill p_T, η and isolation requirements. This selection bias can shift the efficiencies for embedded leptons to higher values compared to leptons in simulated $Z \to \tau\tau$ events. Custom scale factors for the muon and electron acceptance were derived using the *tag and probe* method as described in [32, 33]. For this purpose, the Z boson resonance into two leptons is used to select di-lepton events. One lepton is required to pass very tight selection criteria to minimize the chance of misidentification. This lepton is then referred to as *tag* lepton. A second lepton is selected requiring selection criteria that can be very loose, and is referred to as *probe* lepton. The *tag* and *probe* leptons are paired if a matching to the Z resonance is found, i.e. if the leptons have opposite charge, a minimum distance of $\Delta R > 0.5$ in the η-ϕ-plane and an invariant mass compatible with m_Z. The *tag* and *probe* pair then enables a measurement of the efficiency of the *probe* lepton passing the identification, isolation or trigger requirement of interest by counting the number of *probe* leptons that pass the desired criteria $\epsilon = \frac{N_{\text{pass}}}{N_{\text{all}}}$.
For this $\mu \to \mu$ and $\mu \to e$ embedded samples are produced on the 2016 dataset, resulting in events that have a simulated $Z \to \mu\mu$ and $Z \to ee$ decay embedded in a cleaned $Z \to \mu\mu$ event from data. As *probe* leptons, all particle flow electrons and all global

muons are considered that pass minimal p_T and η requirements. *Tag* leptons additionally need to fulfill identification and isolation, as well as tighter p_T and η requirements. All requirements are given in table 4.1.

Table 4.1: Conditions imposed for muons and electrons for the selection of tag and probe leptons. Muon and electron identification is explained in section 2.2.3. The working points have an efficiency of 99% (muon medium ID) and 80% (electron MVA ID).

	Muon				Electron			
	p_T^μ	$\lvert\eta^\mu\rvert$	ID	I_{rel}^μ	p_T^e	$\lvert\eta^e\rvert$	ID	I_{rel}^e
Tag	$> 23\,\text{GeV}$	< 2.2	medium	<0.15	$> 26\,\text{GeV}$	< 2.1	MVA	<0.1
Probe	$> 10\,\text{GeV}$	< 2.4	-	-	$> 10\,\text{GeV}$	< 2.1	-	-

Efficiencies are determined for identification (ID), isolation (Iso) and high-level trigger (HLT) requirements in sequence, i.e. only leptons that pass the ID requirement enter as *probes* for Iso, and only leptons passing both ID and Iso requirements enter enter as *probes* for the derivation of the HLT efficiency. As a consequence, the total efficiency is calculated in a specific order, which is given by

$$\epsilon(\text{HLT,Iso,ID}) = \epsilon(\text{HLT}|\text{Iso,ID}) \cdot \epsilon(\text{Iso}|\text{ID}) \cdot \epsilon(\text{ID}) \ . \tag{4.1}$$

The determination of efficiencies using tag and probe is done in bins of $|\eta|$ and p_T. The former is done to reflect the detector geometry, as efficiencies are expected to change between barrel (lowest $|\eta|$ bin), transition region and endcaps (high $|\eta|$ bins) of the CMS detector, while the latter reflects expected changes in acceptance with increasing p_T.
The resulting efficiencies $\epsilon(\text{ID})$, $\epsilon(\text{Iso}|\text{ID})$ and $\epsilon(\text{HLT}|\text{Iso,ID})$ are shown figure 4.1 for the lowest $|\eta|$ bin, corresponding to the barrel region of the CMS detector. Muon efficiencies are shown on the left, electron efficiencies are shown on the right.
For muons, the efficiencies are calculated for a *medium* identification as explained above, a relative isolation requirement of $I_{\text{rel}}^\mu < 0.15$ and a trigger requirement using the HLT IsoMu22, which is designed to trigger on isolated muons with $p_T > 22\,\text{GeV}$. The trigger turn-on curve for embedded events shows a steeper slope, which is expected due to the selection criteria imposed on the muon during the selection step, in which they already have to fulfill an HLT requirement of $p_T > 17\,\text{GeV}$ for the leading muon, and $p_T > 8\,\text{GeV}$ for the trailing one. Apart from this effect, no differences are observed between embedded and simulated leptons, and the respective scale factors are close to one.
For probed electrons, differences in the efficiency for identification and subsequently isolation between $\mu \to e$ embedded electrons, electrons in data and $Z \to ee$ simulation can be seen. Electrons in $\mu \to e$ embedded events show a higher efficiency and are more isolated than electrons in simulation or data. These effects were not studied further in the scope of this thesis however can serve as a starting point for further studies in the future. The input variables of the MVA-based electron identification and the objects inside the isolation cone of electrons are quantities of interest for this study. The effects of the correction for differences in identification, isolation and trigger efficiency for the transverse momentum of the muons and electrons in the $\mu\tau_h$ and $e\tau_h$ final states of $\mu \to \tau$ embedded events are shown in figure 4.2.

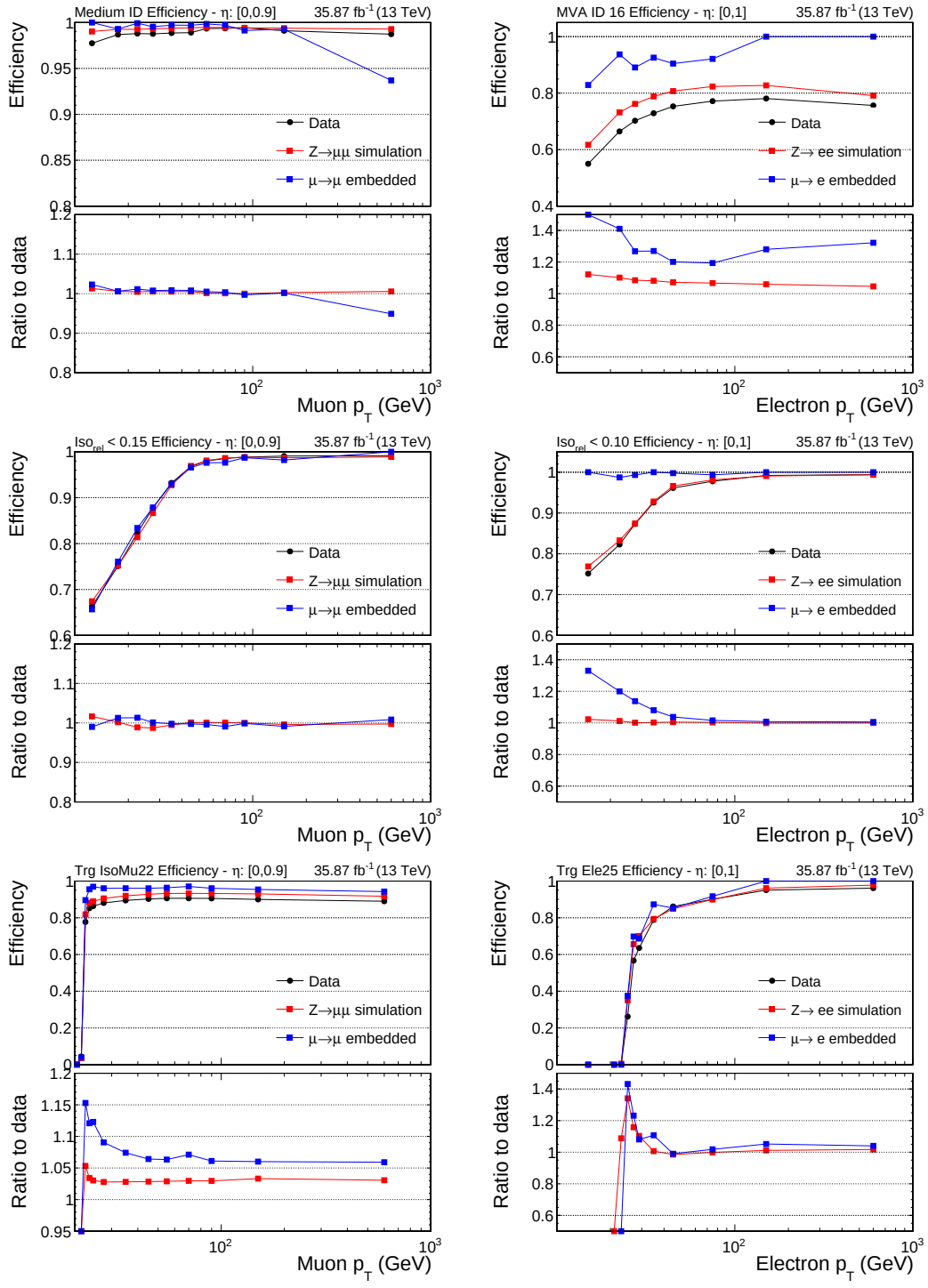


Figure 4.1: Identification, isolation and trigger efficiencies for muons (left) and electrons (right).

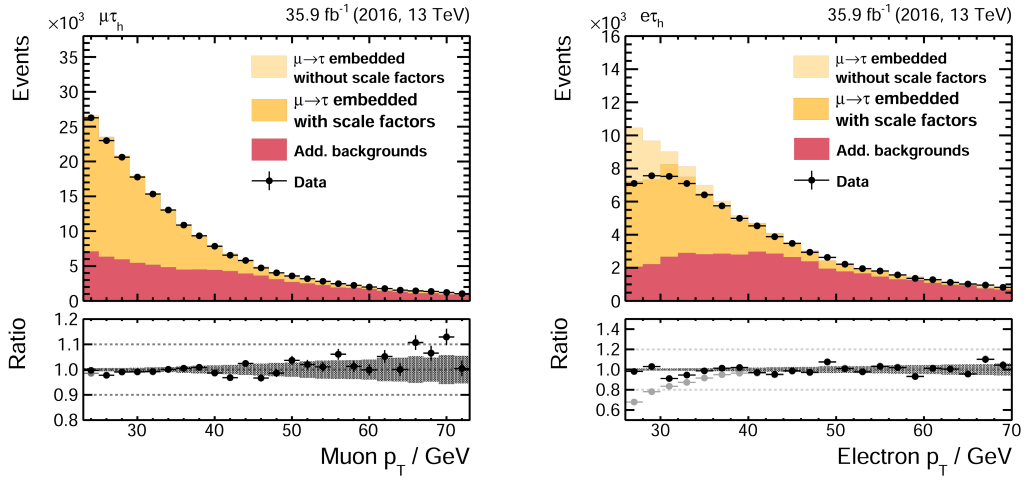


Figure 4.2: Impact of identification, isolation and trigger scale factors derived by tag and probe that were applied on muons (left) and electrons (right). The shape of the $\mu \to \tau$ embedded events is shown as (faded) yellow histogram (before) after the application of identification, isolation and trigger scale factors. The agreement to observed data is shown in the ratio plot as gray (without scale factors) and black dots (with scale factors). The scale factors have a minimal effect for the description of the transverse momentum of the muon from τ-decay. For electrons, a difference between data and embedded events is visible and is corrected by the application of the scale factors derived by tag and probe.

4.2 Initial muon reconstruction

The initial selection step is expected to broaden the energy resolution of the embedded objects. When the initially selected muons are reconstructed, their energy will be folded with a second Gaussian function related to the finite muon energy resolution of the CMS detector as described in chapter 2. This causes an additional smearing effect on the reconstructed di-muon mass. The effect on the derived Z boson kinematics can be modeled by a product of two Gaussians related to the independent reconstruction of the two muons, which is therefore again a Gaussian smearing of the Z peak. This smearing leads to a softening of the Z peak in the mass spectrum of the di-muon system which is shown in figure 4.3. The shown di-muon mass spectrum was derived by applying $\mu \rightarrow \mu$ *embedding* to simulated $Z \rightarrow \mu\mu$ events, i.e. muons will be removed and reinjected into the event in the context of the embedding methods. Details about the sample of simulated events that has been used for this study will be given in chapter 5.

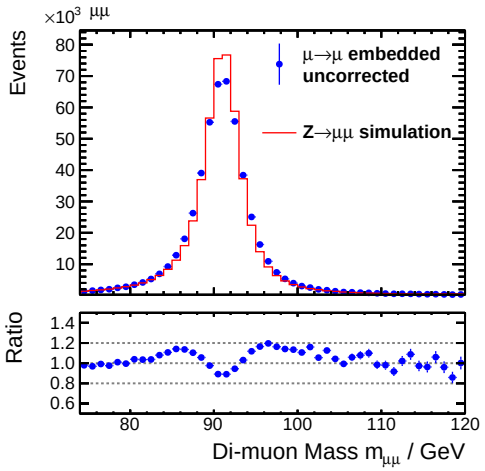


Figure 4.3: Invariant di-muon mass of simulated $Z \rightarrow \mu\mu$ events before and after the removal and reinjection of muons. The softening of the Z peak caused by the additional folding of the resolution of the Z peak with the detector resolution in embedded events is visible.

Usually, folding effects such as the one presented here require unfolding to revert the distortions introduced by the detector. Here, a way is chosen in the form of correction factors which assume both distributions to be Gaussian distributed. Correction factors are developed which shift the smeared 4-vectors of the muons back in direction of the Z mass.
The correction factors are constrained to a region around the Z peak, since selected muons from data do not necessarily come from a Z boson, especially if the di-muon mass is significantly above m_Z. This is explained in section 4.3.

Keeping this in mind, a correction factor has been derived that will have the desired effects on the 4-vectors of the muons.

$$c_1(m_{\mu\mu}) = \frac{A}{A - (A-1) \cdot \exp\left(\frac{-(m_{\mu\mu}-m_Z)^2}{2\cdot\sigma_A^2}\right)} \tag{4.2}$$

$$c_{tot}(m_{\mu\mu}) = \frac{m_{\mu\mu} + (c_1 - 1) \cdot m_Z}{m_{\mu\mu} \cdot c_1} \tag{4.3}$$

The function shown in equation 4.2 will restrict the applications of the corrections around the Z-peak by using a correction factor c_1 that will only be different from 1 around m_Z. Equation 4.3 is using this form factor to derive the final correction c_{tot} depending on the position of the di-muon mass relative to $m_Z = 91.2\,\text{GeV}$ [34]. The correction only depends on the constant mass of the Z boson m_Z and the di-muon mass of the event $m_{\mu\mu}$. The parameters A and σ_A that describe the magnitude of the effect and the deterioration behavior respectively have been determined by χ^2-curve-fitting to the values $A = 1.137$ and $\sigma_A = 5.01\,\text{GeV}$. This fit is shown in figure 4.4, and the resulting distribution of the final correction factors is shown in figure 4.5.
The correction factor c_{tot} is multiplied to the momenta of the two muons $\vec{p}'_\mu = c_{tot} \cdot \vec{p}_\mu$. The muon energy is then recalculated as $E' = \sqrt{m_\mu^2 + (\vec{p}'_\mu)^2}$ with the muon mass $m_\mu = 0.1057\,\text{GeV}$ [35]. These new kinematics are then used for the reconstruction of the Z boson and subsequent simulation of the embedded objects. Figure 4.6 shows the distribution of the visible mass of the di-muon system after the application of the energy-dependent correction factors. The correction has the desired effect of reverting the smearing effect introduced by the initial muon reconstruction and the mass distribution of $\mu \to \mu$ embedded events agrees with the di-muon events selected from simulation. Uncertainties for this correction are derived using the uncorrected distribution as well as the distribution where the correction is applied twice.

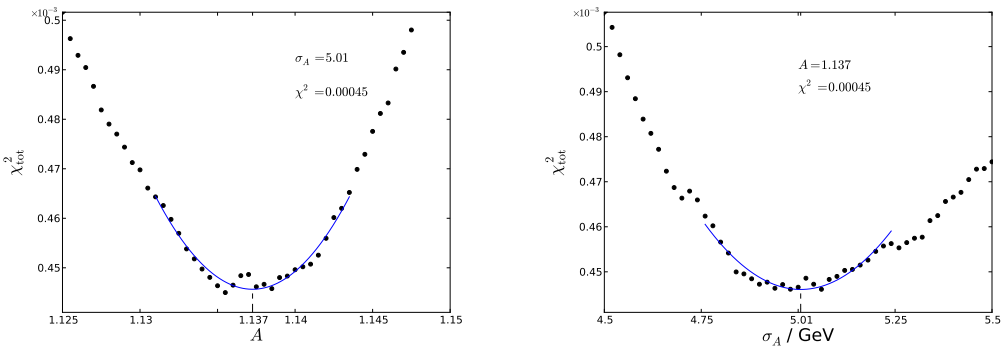


Figure 4.4: One-dimensional projections of the fit of parameters A (left) and σ_A (right), as derived by a fit of the invariant di-muon mass of $\mu \to \mu$ embedded events to the initially selected simulated $Z \to \mu\mu$ events.

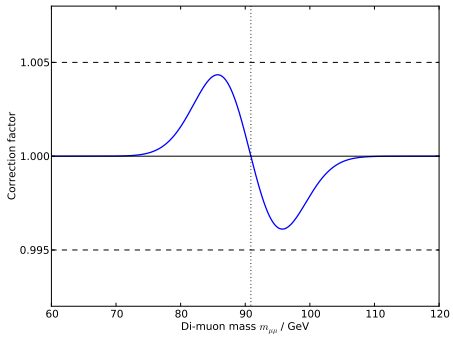


Figure 4.5: Correction factor $c_{tot}(m_{\mu\mu})$ that is multiplied to the momentum of the selected muons to receive the corrected kinematics of the Z decay as calculated from equation 4.3. The corrected kinematics are then used for the simulation of the Z boson decay into leptons. The correction is limited to an area of $\approx 20\,\text{GeV}$ around the Z boson mass and reaches a maximum of 0.4% around 5 GeV above and below m_Z.

The corrections and therefore uncertainties only affect selected events where the di-muon mass is within $\pm 20\,\text{GeV}$ around the Z mass and is limited to a maximum correction factor of 0.4% for events where the reconstructed di-muon mass is $\pm 5\,\text{GeV}$ around the Z peak.
The smearing caused by the initial muon reconstruction as well as the effect of the correction have been studied for $\mu \to \tau$ embedded events, where the resolution of the visible di-τ mass is dominated by the neutrinos escaping detection. This will be discussed in section 5.2. For future productions of embedded events, e.g. on the 2017 dataset, the corrections have been implemented at an early stage in the analysis for both the $\mu \to \mu$ and $\mu \to \tau$ embedding method. The energy of the di-muon system is corrected right after the muons are selected and before the simulation is performed, to ensure the correct description of all subsequent event variables.
The embedded events produced on the 2016 data taken by CMS that will be used for the analysis presented in chapter 6 have not been reproduced with the corrections applied Instead the effect for $\mu \to \tau$ embedded events has been checked to be small. Uncertainties on the energies of the final state leptons are applied which will be discussed in section 6.4.

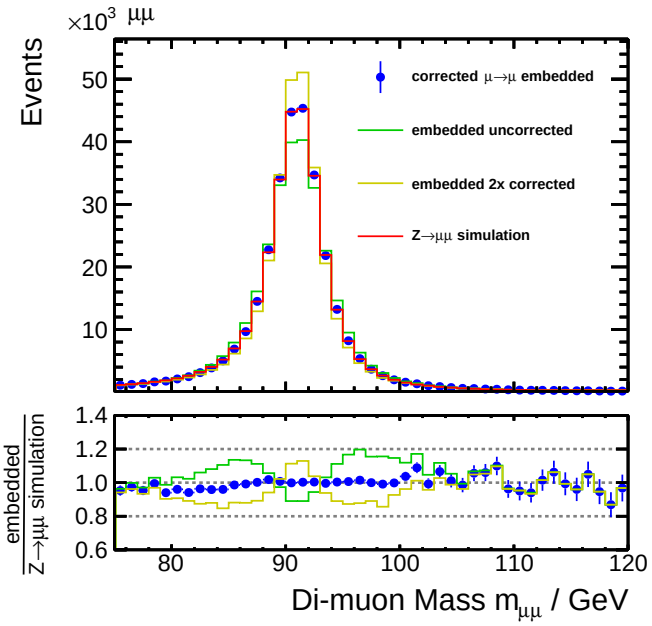


Figure 4.6: Visible mass of the di-muon system with the corrections shown in figure 4.5 applied. The improved agreement with the originally selected di-muon events is visible. The shapes of uncorrected embedded events and the shape with the correction applied twice are used as a systematic uncertainty.

4.3 Correction for contamination by top-quark pair decays

When selecting $Z \to \mu\mu$ events from data, only lower p_T and upper η conditions are imposed on the two selected muons (see section 3.1). Therefore, any central di-muon event with sufficient p_T will fulfill the selection criteria. Next to Z decays, the most prominent source of high-energy di-muon events are decays of top-quark pairs in the decay mode

$$t\bar{t} \to W^+W^-b\bar{b} \to \mu^+\nu_\mu\mu^-\bar{\nu}_\mu + \text{jets} \ .$$

This decay mode has the same branching ratio as

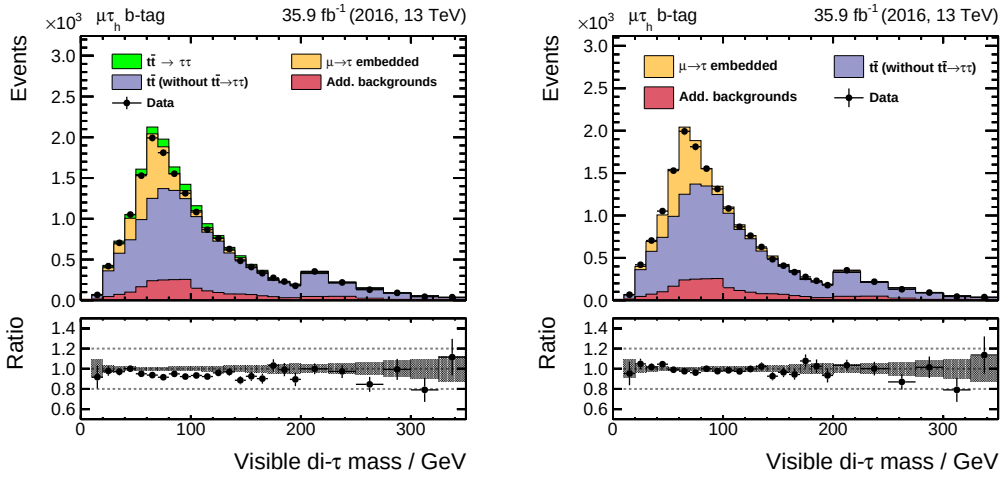
$$t\bar{t} \to W^+W^-b\bar{b} \to \tau^+\nu_\tau\tau^-\bar{\nu}_\tau + \text{jets} \ .$$

These events will be selected and processed by the $\mu \to \tau$ embedding procedure, which leads to the embedded event samples being contaminated by events that do not describe $Z \to \tau\tau$, but rather $t\bar{t} \to \tau\tau$ events.
Events in which a top-quark pair decays into two τ-leptons (muons) will from now on be shortened as $t\bar{t} \to \tau\tau(\mu\mu)$ events. In the use of embedded events during Run I [30] at center of mass energies of 7 and 8 GeV this effect was studied and covered by dedicated uncertainties. For the production of embedded events using Run II-data that is taken at a center of mass energy of 13 TeV, the $t\bar{t}$ contamination of embedded events becomes slightly more pronounced due to the enhancement of the cross-section of the production of top-quark pairs, and dedicated corrections are applied. The contamination due to the selection of $t\bar{t} \to \mu\mu$ events applies to a fraction of around 1% of all embedded events inclusively. As a possible $H/A \to \tau\tau$ signal is expected to be associated with b-quarks, a categorization in chapter 6 is performed depending on the presence of b-quark jets, where the contamination can be much larger, e.g. up to 39% for the `b-tag loose` $\mathtt{m_T}$ category in the $\mu\tau_h$ channel. The categorization is described in chapter 6. When using embedded events as a background estimation for $Z \to \tau\tau$ events, such events will be included in the additional $t\bar{t}$ estimation and therefore double-counted. This leads to an overestimate of events depending on the muon energy and on the presence of b-quark jets in the event. The effect of the $t\bar{t}$ contamination when using embedded events without any further treatment is shown in figure 4.7 (a).

4.3.1 $t\bar{t} \to \tau\tau$ veto approach

To take the contamination into account, a correction is applied not on embedded events, but on the simulated events used for the estimation of $t\bar{t}$ background. As embedded events have advantages over simulated events when describing jets, these advantages can be used to describe a small part of the $t\bar{t}$ background as well by keeping the $\mu \to \tau$ embedded events selected from $t\bar{t} \to \mu\mu$, and instead removing them from the simulation. As the $t\bar{t}$ background estimation is derived from simulation, the removal can be done without further assumptions and uncertainties concerning the reconstruction of the two τ-leptons by using the generator information present in simulated $t\bar{t}$ events to veto all

genuine $t\bar{t} \to \tau\tau$ events. Additional background attributed to $t\bar{t}$ decays in which the W boson decays directly into a muon or electron, such as $t\bar{t} \to \mu(e)\tau_h$ or $t\bar{t} \to e\mu$ are not impacted by this approach and are still obtained from simulation.
This veto thus avoids the double-counting of $t\bar{t} \to \tau\tau$ events and uses embedded events as a background description for $Z/t\bar{t} \to \tau\tau$. The effect of the veto is shown in figure 4.7, where the visible di-τ mass of events in the $\mu\tau_h$ final state is shown. Embedded events are shown by the yellow histogram. The simulated $t\bar{t}$ events are split depending on whether they are genuine di-τ events ($t\bar{t} \to \tau\tau$, green) or whether the background can be attributed to $t\bar{t}$ decays into either a prompt muon and τ_h or a prompt muon and a jet misidentified as τ_h ($t\bar{t}$, purple). Additional background processes, such as W+jets, di-boson or QCD multijet production are shown in red. As embedded events are contaminated by $t\bar{t} \to \tau\tau$, such events are double-counted in the left hand plot. The removal of the $t\bar{t} \to \tau\tau$ events from simulation corrects for the contamination. A selection was applied that requires the presence of at least one b-tagged jet in addition to the baseline event selection of the $\mu\tau_h$ final state, which will be described in chapter 6. The additional requirement of at least one b-tagged jet was imposed to accentuate the effects of the $t\bar{t}$ contamination. In events where a b-tagged jet is present, the contamination of embedded events by $t\bar{t} \to \mu\mu$ events is enhanced by a factor of 25, and 1 in 4 embedded events are selected from a $t\bar{t}$ decay as opposed to a Z decay.



(a) Before $t\bar{t} \to \tau\tau$ veto

(b) After $t\bar{t} \to \tau\tau$ veto

Figure 4.7: Visible di-τ mass of events in the $\mu\tau_h$ final state. Only events with a b-tagged jet are shown to accentuate the contamination by $t\bar{t}$ events. The removal of the simulated $t\bar{t} \to \tau\tau$ events, shown on the right, brings the desired effect of correcting for this contamination.

4.3.2 Validation study in high-purity $t\bar{t} \to \tau\tau$ environment

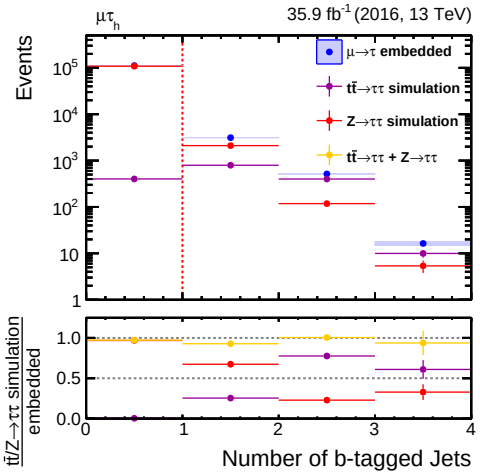
The correction is studied in a high-purity $t\bar{t}$ environment. The goal is to check that the additional embedded events can in fact be attributed to $t\bar{t} \to \tau\tau$ events, which is done by comparing the numbers and characteristics of embedded and $t\bar{t}$ events in a selection where the contribution of Z boson decays is small. Requirements were therefore chosen to ensure a low contribution from selected $Z \to \mu\mu$ events to the embedded events, which was done by using the following selection criteria in addition to the baseline selection in the respective final states. The results presented here will focus on the $\mu\tau_h$ and $e\tau_h$ final states. The selection requirements are:

- Number of b-tagged jets $n_{\text{btag}} \geq 1$
- Visible di-τ mass $m_{\text{vis}} > 90\,\text{GeV}$
- Transverse mass $m_T(\ell, \not{E}_T) > 70\,\text{GeV}$

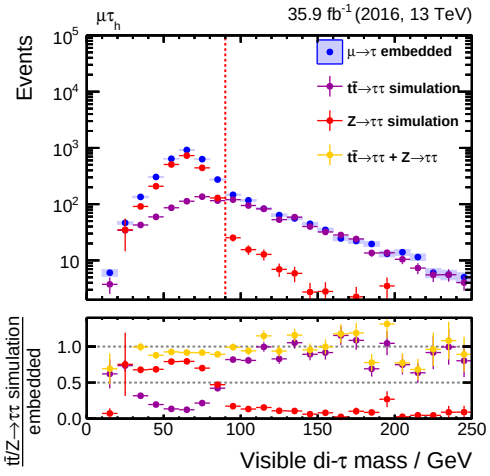
Where the transverse mass of the lepton $m_T(\ell, \not{E}_T)$ is defined as

$$m_T^{\mu(e)} = \sqrt{2 p_T^{\mu(e)} \not{E}_T (1 - \cos(\Delta\phi))} \tag{4.4}$$

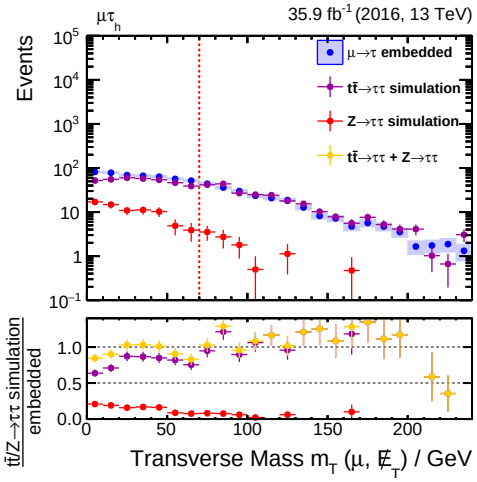
and $p_T^{\mu(e)}$ refers to the magnitude of transverse momentum of the muon (electron), $\not{E}_T$ to the magnitude of missing transverse momentum $\vec{p}_T^{\,\text{miss}}$ and $\Delta\phi$ to the azimuthal angle between the directions of the two momenta.
Only top-quark pair decays into two τ-leptons are considered. This ensures a comparison of $\mu \to \tau$ embedded events to exactly those simulated $t\bar{t}$ events subject to the $t\bar{t} \to \tau\tau$ veto. The selection is visualized in figure 4.8.
The difference between embedded and simulated $Z \to \tau\tau$ events for values of $m_T^{\mu(e)} > 70\,\text{GeV}$, $m_{\text{vis}} > 90\,\text{GeV}$ and $n_{\text{btag}} \geq 1$ is caused by the increasing fraction of $t\bar{t} \to \tau\tau$ events. In addition, the sum of simulated $t\bar{t} \to \tau\tau$ and simulated $Z \to \tau\tau$ events is shown. The distribution of embedded events is in agreement with the distribution of simulated $t\bar{t} \to \tau\tau + Z \to \tau\tau$ events, as both decays of top-quark pairs and decays of the Z boson into muons are selected by the $\mu \to \tau$ embedding method. The events that are not selected from Z boson decays in the embedded event sample can therefore be used to replace the simulated $t\bar{t} \to \tau\tau$ events.
The application of the three selection criteria results in a pure $t\bar{t}$ environment with an expected fraction of $t\bar{t}$ events of 96%. This enables a direct comparison between the remaining simulated $t\bar{t} \to \tau\tau$ and the embedded events, which are now mostly $t\bar{t} \to \tau\tau$ events.
Figure 4.9 shows the distribution of the transverse momentum, transverse mass and visible di-τ mass for both embedded and simulated $t\bar{t}$ events. Additionally, simulated $Z \to \tau\tau$ events are shown to control for effects caused by remaining $Z \to \tau\tau$ events present in the embedded events. Statistics are limited in this study due to the selection criteria that have been applied. The respective statistical uncertainties of embedded and $t\bar{t}$ events are shown by a blue band and purple error bars respectively.



(a) Baseline selection.



(b) Baseline selection + $n_{\text{btag}} \geq 1$.



(c) Baseline selection + $n_{\text{btag}} \geq 1$ + $m_{\text{vis}} > 90\,\text{GeV}$.

Figure 4.8: Variables used to suppress the $Z \to \tau\tau$ background for the study in a pure $t\bar{t}$ environment in the $\mu\tau_h$ channel. Shown are embedded events in blue, simulated $t\bar{t}$ events in purple and simulated $Z \to \tau\tau$ events in red. The sum of simulated events in shown in yellow. The three selection criteria are indicated by the red dotted lines, the distributions are shown consecutively: The distribution of the visible di-τ mass is shown after imposing a $n_{\text{btag}} \geq 1$ requirement, and the distribution of transverse mass is shown after imposing both a $n_{\text{btag}} \geq 1$ and a $m_{\text{vis}} > 90\,\text{GeV}$ requirements. The combination of the three selection criteria leads to a very pure $t\bar{t}$ environment, and enables the comparison of events in embedded samples that come from muons selected in $t\bar{t}$ decays instead of decays of the Z boson.

No significant trend or deviation between the two distributions is seen and the distributions are compatible within the statistical uncertainties. Removing the depicted $t\bar{t}$ simulated events from the respective background estimation properly accounts for the $t\bar{t}$ contamination in the embedded sample. At the same time, it will show lower statistical fluctuations, since the prediction derived by embedded events is described by twice as many events as the prediction by simulated $t\bar{t}$ events. The shown background predictions consist of 1330 unweighted embedded events and 705 unweighted simulated $t\bar{t}$ events, which lowers the statistical uncertainty by a factor of $\sqrt{2}$ when using embedded events for the description of $t\bar{t} \to \tau\tau$ events. Furthermore, systematic uncertainties on the jet energy and b-tag efficiency can be avoided by the use of embedded events. A systematic uncertainty of 10% of the vetoed $t\bar{t}$ events is introduced to account for insufficient knowledge of the $t\bar{t}$ contamination, which will be discussed in detail in chapter 6.

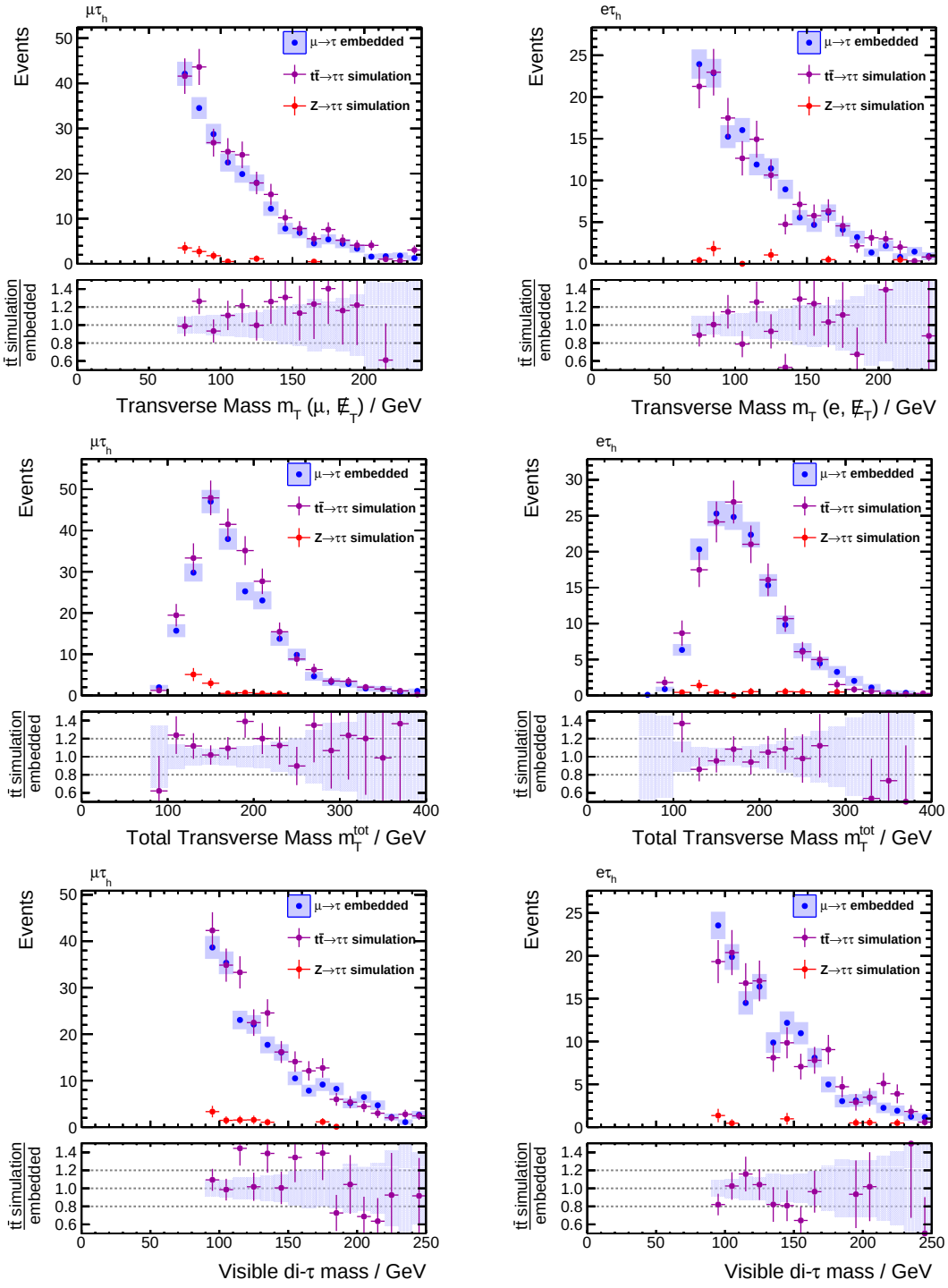


Figure 4.9: Transverse masses (upper row), total transverse mass (middle row) and visible mass (lower row) of the two τ-decay products in the $\mu\tau_h$ (left) and $e\tau_h$ (right) final states after applying a selection to increase the expected fraction of $t\bar{t}$ events to 96%.

CHAPTER 5

Closure study using simulated $Z \to \ell\ell$ events

For a systematic understanding of all differences that the application of the $\mu \to \tau$ embedding method introduces, a validation study has been conducted. A dataset containing around 20 million simulated $Z \to \ell\ell$ has been used for the study, with ℓ referring to a muon, electron or τ-lepton. For details on the dataset used, see A.1. The chosen dataset contains a high-pileup scenario, for which additional pp collisions with small exchanged transverse momentum are added to the hard $q\bar{q} \to Z \to \ell\ell$ interaction to simulate the conditions at the CMS detector, where multiple interactions occur per crossing. During 2016 data-taking, the number of interactions per crossing at the CMS detector followed a Poisson distribution with a mean of 27. [19]. The number of interactions in the dataset follows a roughly flat distribution over a fixed range, which is 28 to 62 for the dataset chosen. The pileup distribution is shown in figure 5.1.
After having selected the simulated $Z \to \mu\mu$ events, the validation study was performed in two ways:

1. As a first study, presented in section 5.1, the reconstructed kinematics of the dimuon system were used to produce $\mu \to \mu$ embedded events. This is the most straight-forward validation method, as both the selected and the embedded event are expected to show equal characteristics.

2. Second, the same selected $Z \to \mu\mu$ events were used to produce $\mu \to \tau$ embedded events just as performed with data. The embedded events were then compared to simulated $Z \to \tau\tau$ events simulated under the same conditions as the selected $Z \to \mu\mu$ events. This will be shown in section 5.2.

As the conditions used for the simulated events are known and no other backgrounds interfere in this controlled environment, this study will serve as a test of the reliability of the embedding method to describe $Z \to \tau\tau$ events.

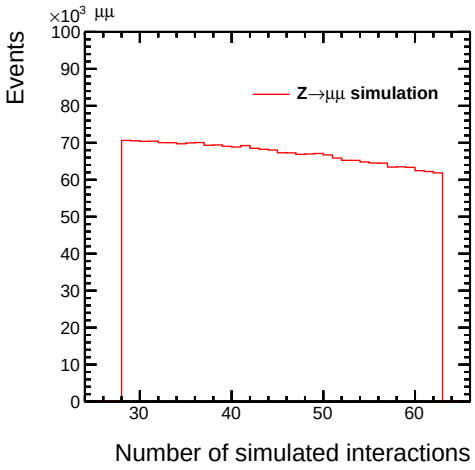


Figure 5.1: Number of simulated interactions for the high-pileup dataset used. It follows a roughly flat distribution over the range of 28 to 62.

5.1 Study of $\mu \to \mu$ embedded events

The first study is performed by removing muons from the simulated $Z \to \ell\ell$ events and simulating the decay of the reconstructed Z boson into two muons again. Ideally, this should not introduce changes to the events other then the expected effects due to the initial reconstruction of the muon, which has been described in section 4.2, and effects due to the cleaning of muon tracks. For the study, an event selection has been applied to account for the p_T, isolation and η requirements during the initial muon selection as applied to data. A slightly tighter event selection with respect to the one shown in section 3.1 is chosen to avoid migration effects at the boundaries of the respective requirements. The event selection is summarized in table 5.1. The selection of well-isolated muons is done using the relative isolation of the muon as defined in equation 2.1. Additionally, just as during the initial selection, only events in which the charges of the two muons are opposite are considered for this study.

Table 5.1: Event selection of the muons used for the comparison of simulated $Z \to \mu\mu$ with $\mu \to \mu$ embedded events.

	Transverse momentum	Pseudorapidity	Relative Isolation
Leading Muon	$p_T^\mu > 23\,\text{GeV}$	$\lvert\eta^\mu\rvert < 2.4$	$I^\mu_{\text{rel}} < 0.15$
Trailing Muon	$p_T^\mu > 10\,\text{GeV}$	$\lvert\eta^\mu\rvert < 2.4$	$I^\mu_{\text{rel}} < 0.15$

5.1.1 Validation of event cleaning

In this section, the effectiveness of the event cleaning is validated using simulated $Z \to \mu\mu$ events. As mirroring (see section 3.4) is a tool to study effects related to the selection requirements imposed during the muon selection and to remnants in the area around

the cleaned muon tracks, $\mu \to \mu$ embedded events were produced with and without the application of mirroring. The purpose of the mirroring method is to check if possible effects caused can be explained by selection or cleaning effects. The relevant variable for this study is the isolation of the muon before and after the embedding method is applied, as well as the hadronic and electromagnetic activity around the embedded muons.
In figure 5.2 the relative isolation of the leading muon as defined in equation 2.1 of the embedded $Z \to \mu\mu$ events is compared to the initially selected simulated $Z \to \mu\mu$ events. The default selection requirement on the muon isolation for the $H/A \to \tau\tau$ analysis of $\mathrm{I}^{\mu}_{\mathrm{rel}} < 0.15$ is illustrated by a red dotted line.

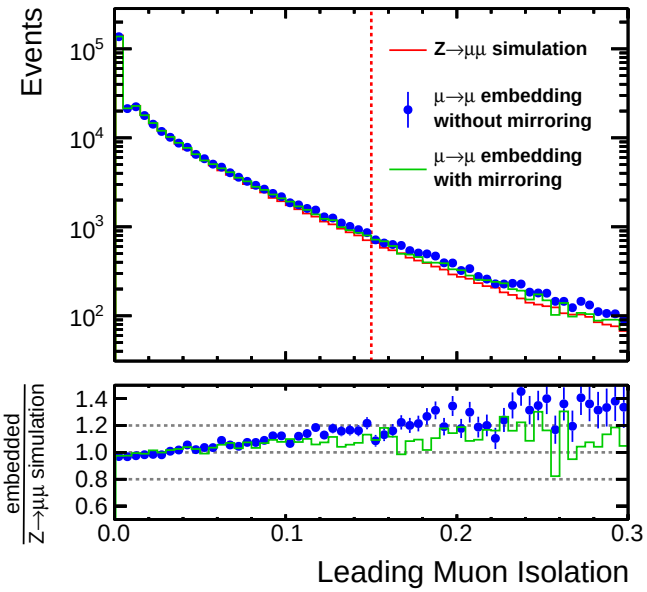


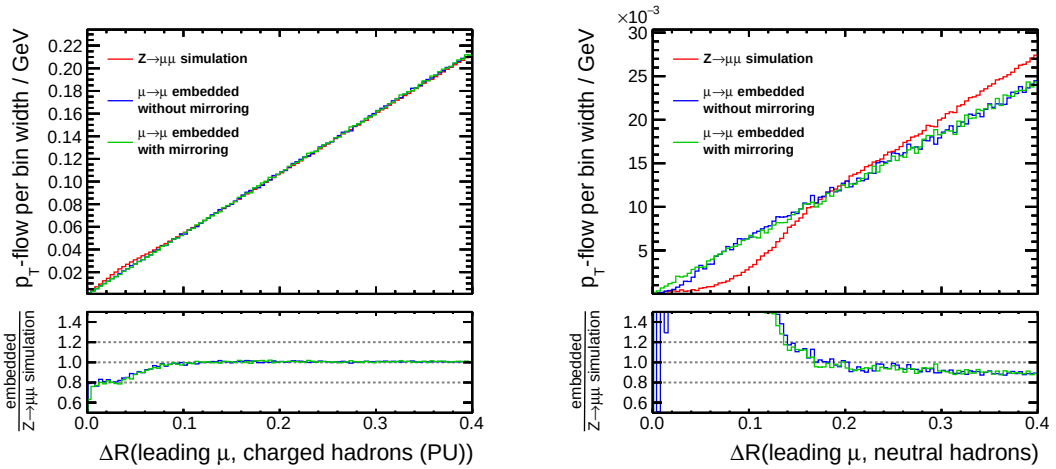
Figure 5.2: Relative isolation I^{μ}_{rel} of the leading muon in embedded events with and without mirroring, compared to the events selected from $Z \to \mu\mu$ simulation. The red dotted line implies the requirement on relative muon isolation of 0.15 used for the event selection. The number of events with an isolation greater than 0.15 increases with the application of the embedding method.

Muons in both unmirrored as well as mirrored embedded events are less isolated than the initially selected simulated $Z \to \mu\mu$ events, however, the effect is larger for unmirrored events. This can be explained by remnants around the initially selected muon that did not get removed in the cleaning step (see section 3.2).
Figure 5.3 shows the p_T-flow distribution in a ΔR cone of 0.4 in the η-ϕ-plane, which is the cone radius in which particles enter the sum used for calculation of the relative isolation. The p_T-flow is defined as the sum of all particle flow candidate p_T-values for each ΔR bin per muon. For a constant track density, one expects a linear rise in p_T-flow with increasing distance as each ΔR bin refers to a ring around the muon direction with

linearly increasing area at higher radius. The p_T-flow distributions shown in figure 5.3 are split into four particle flow candidates:

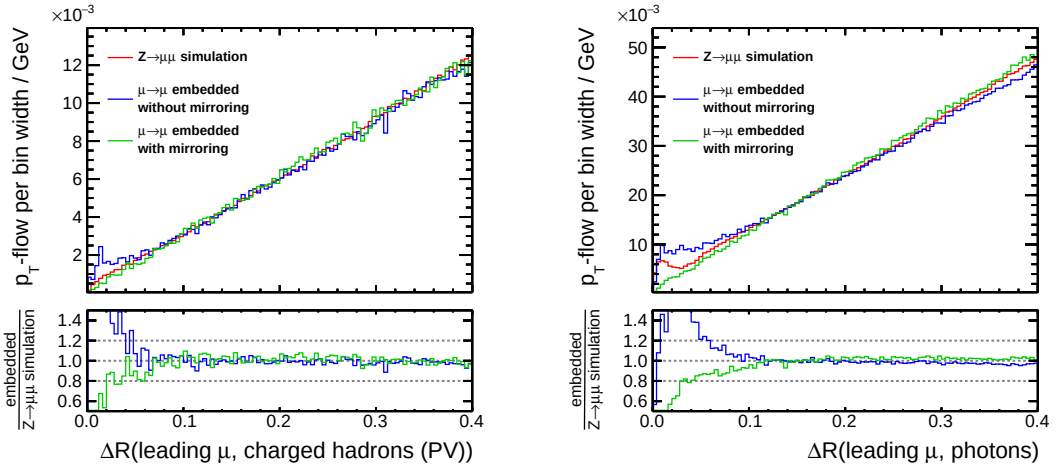
1. Neutral hadrons.
2. Photons.
3. Charged hadrons, whose tracks were matched to leading primary vertex of the $Z \to \mu\mu$ event.
4. Charged hadrons, whose tracks were not matched to the leading primary vertex and are therefore associated with pileup.

The p_T-flow distribution of mirrored embedded events follow a linear rise with increasing ΔR, referring to a constant track density around mirrored muons. The mirroring method therefore enables the study of effects due to the reconstruction of particle flow candidates around the muons. The distribution of p_T-flow of simulated $Z \to \mu\mu$ events shows such reconstruction effects: For charged hadrons from pileup and from the primary vertex as well as photons an excess of p_T-flow of up to 40% over a linear distribution can be seen for $\Delta R < 0.05$. For neutral hadrons on the other hand, the p_T flow is lower in an are of $\Delta R < 0.2$ around simulated muons, which points to a physical or reconstruction effects of less neutral hadron activity around reconstructed muons. These effects were not studied further, but can serve as a starting point for further studies on the reconstruction of muons at CMS. Embedded events do not reproduce these reconstruction effects. As pileup is independent of the direction of the two muons from the leading primary vertex, the p_T-flow distribution of hadrons related to pileup is independent of the application of mirroring (Figure 5.3 (a)), and embedded events show a linear p_T-flow distribution irrespective of the application of mirroring. The same is true for neutral hadrons (Figure 5.3 (b)) where the pileup contribution is dominant. As neutral hadrons do no leave tracks in the pixel detector and can not be matched to a collision vertex, the pileup contribution to the p_T-flow is estimated from the rate of charged hadrons from pileup over charged hadrons from primary vertex, which is 17/1. For these hadrons, both unmirrored and mirrored embedded events show a linearly increasing unbiased distribution.
An effect of the application of the embedding method is visible for low values of ΔR for charged hadrons whose tracks were matched to the primary vertex of the $Z \to \mu\mu$ event (Figure 5.3 (c)), as well as for photons (d). The increase in p_T-flow is limited to an area of $\Delta R < 0.05$ and can be explained by remnants of the cleaning step, as the increase can not be seen in mirrored embedded events. By applying mirroring, the embedded muons are simulated into an environment where the respective remnants do not impact the relative isolation of the muon.
The remnants however do not disappear by applying mirroring. They can lead to additional particle flow candidates being reconstructed for the event irrespective of the application of mirroring, which leads to the number of particle flow candidates being on average higher for embedded events in comparison to the originally selected simulated $Z \to \mu\mu$ events. This can be seen in figure 5.4.



(a) Charged hadrons from PU

(b) Neutral hadrons



(c) Charged hadrons from PV

(d) Photons

Figure 5.3: p_T-flow distributions around the leading muon. Shown are $\mu \to \mu$ embedded events where mirroring has not been applied (blue), and where mirroring has been applied (green). The embedded events are compared to the initially selected muons from $Z \to \mu\mu$ simulation. Mirroring corrects effects due to remaining photon and charged hadron remnants around the cleaned muon track, however does not reproduce the reconstruction features seen in $Z \to \mu\mu$ simulation.

The effects due to inefficiencies in the cleaning step are not solved by applying mirroring. The application of mirroring reduced the bias on muon isolation introduced by the embedding method, however, for the description of well-isolated muons within the selection used for the analysis in chapter 6, it is unlikely to make differences for the description of $Z \to \tau\tau$ background by $\mu \to \tau$ embedded events. As the p_T-flow distribution for mirrored events tends to over-correct the p_T-flow distribution and is not featureless by itself, it is possible that it introduces new effects that require further studies if mirroring were

applied. Thus, mirroring was not used for the production of embedded events using 2016 data. The mirroring method will remain a useful tool for validating the effects of the embedding method.

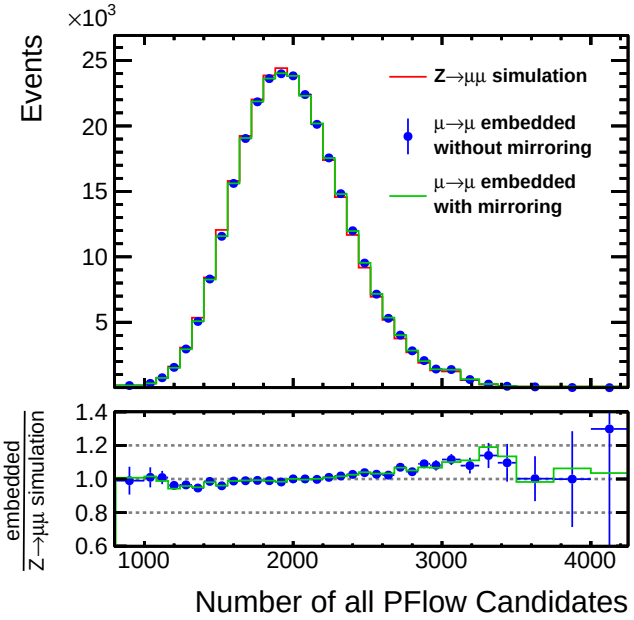


Figure 5.4: Number of particle flow candidates in $Z \to \mu\mu$ events and missing transverse momentum as calculated from all particle flow candidates before and after applying the embedding method. The comparison was made both with and without the application of mirroring. A shift of embedded events to higher number of particle flow candidates is visible. The reconstruction of additional particle flow candidates is independent of the application of mirroring.

5.1.2 Comparison of event kinematics

As muons are reconstructed twice in $\mu \to \mu$ embedded events, an effect is expected which causes a smearing of the Z peak in the invariant di-muon mass spectrum. This has been described in section 4.2 and dedicated corrections are applied for all distributions shown in the following. In addition, the double reconstruction is expected to impact the description of the missing transverse energy $\not{E}_T$ in embedded events. The $\not{E}_T$ as calculated from the sum of transverse energy from all particle flow candidates is shown in figure 5.5. For simulated $Z \to \mu\mu$ events that have a small missing transverse energy $\not{E}_T < 5\,\text{GeV}$, the application of the $\mu \to \mu$ embedding method reduces the number of events in this regime by 10%.

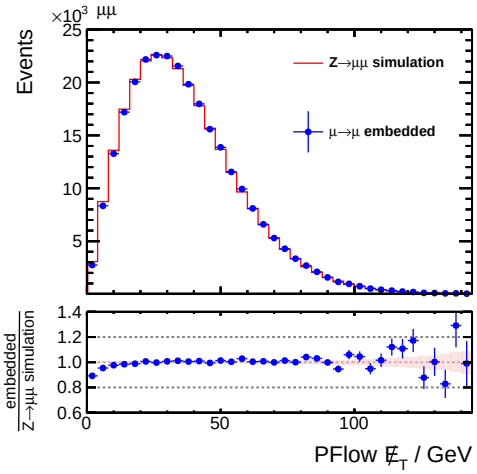
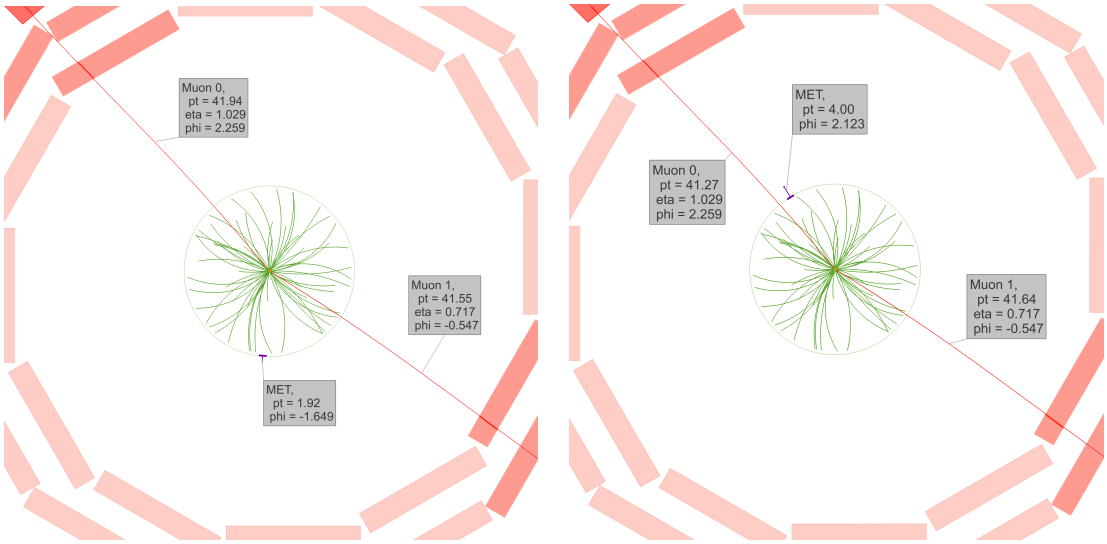


Figure 5.5: Missing transverse momentum $\not{E}_T$ as calculated from all particle flow candidates before and after applying the embedding method. A misdescription of events with $\not{E}_T < 5\,\text{GeV}$ can be seen.

This effect is caused by the additional reconstruction of embedded muons and is illustrated in figure 5.6. Here, a simulated $Z \to \mu\mu$ event with a low missing transverse energy is shown. For the creation of the embedded event, the two muons are cleaned and a decay into two muons with the same kinematics is simulated. The reconstructed transverse momenta of the two embedded muons is changed with respect to the selected muons due to the independent reconstruction of the embedded event. This also randomly shifts the direction and magnitude of the reconstructed missing transverse energy. For events with $\not{E}_T > 5\,\text{GeV}$, these random shifts compensate each other and do not impact the overall description of the $\not{E}_T$ distribution. Yet, if a simulated event has a very low reconstructed $\not{E}_T$ like the event shown in figure 5.6, the shift is more likely to be towards higher values of $\not{E}_T$ and is not fully compensated by other events migrating to lower $\not{E}_T$ values. The effect is not affected by the corrections on the initial muons reconstruction shown in 4.2, as these corrections are applied on the two muons collectively before reconstruction of the Z boson mass. In figure 5.7, the total transverse energy of the event is shown as calculated with and without the inclusion of the two muons from the Z decay.



(a) $Z \to \mu\mu$ simulation **(b)** $\mu \to \mu$ embedded

Figure 5.6: Event display of a simulated $Z \to \mu\mu$ event (left) and the corresponding $\mu \to \mu$ embedded event (right) in the ρ-ϕ-plane of the CMS detector. A simulated event with a low missing transverse energy of $\not{E}_T = 1.9\,\text{GeV}$ was chosen. As the embedded muons are reconstructed independently of the initially selected muons, resolution effects are visible and the reconstructed transverse momenta of the muons are shifted by 1.6% and 0.2% for Muon 0 and Muon 1 respectively. This shifts the missing transverse energy to a higher value of $\not{E}_T = 4\,\text{GeV}$.

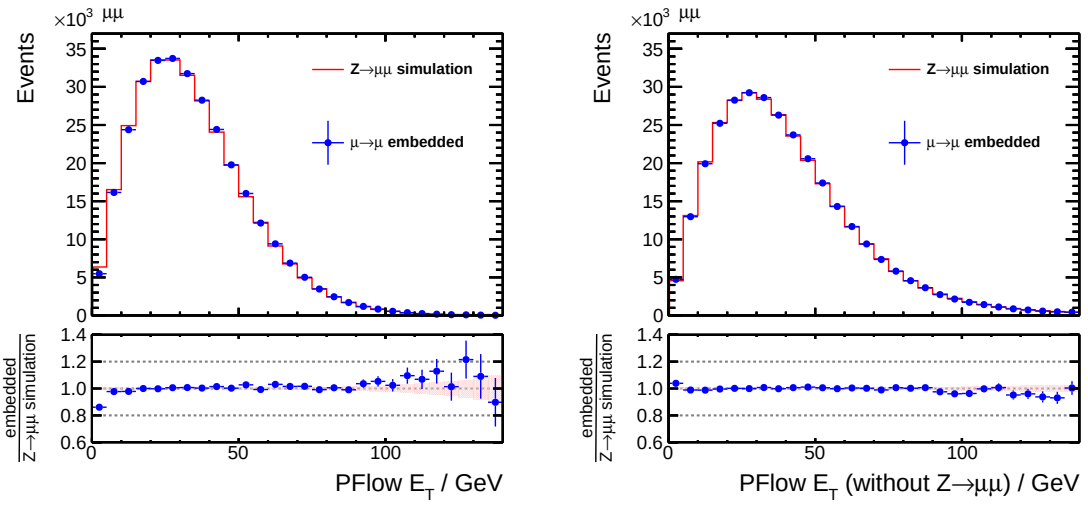


Figure 5.7: Total transverse momentum E_T as calculated from all particle flow candidates before and after applying the embedding method. The shift of events with low E_T is not visible when excluding the muons from the $Z \to \mu\mu$ event.

As the transverse energy is highly correlated with the missing transverse energy, the same effect for events with low E_T can be seen. The effect disappears when the $Z \to \mu\mu$ event is excluded from the calculation. The effect can therefore be fully attributed to the additional muon reconstruction in embedded events. As the effect in the order of $\mathcal{O}(10\%)$ only occurs for the small number of events with $\not{E}_T < 5\,\text{GeV}$ (1.7% in this study), this effect is not expected to have an impact on the estimation of $Z \to \tau\tau$ events by embedded events.
Expected effects have been found in the description of the invariant di-muon mass and events with low missing transverse energy $\not{E}_T$. Corrections for the effect on the invariant di-muon mass have been applied. Further variables that require an accurate description include the kinematics of the two embedded leptons. Figure 5.8 compares the transverse momenta of the two muons before and after the application of the embedding method, as well as the description of high-p_T jets. Moreover, the transverse mass of the leading muon as defined in equation 4.4 is shown. This variable is used for event classification and discrimination for the analysis in the di-τ channel presented in chapter 6, an unbiased description by embedded events is therefore of particular importance. The description of the quantities shown is reproduced by embedded events within the statistical accuracy of the selected sample. Additional event quantities can be found in A.3.

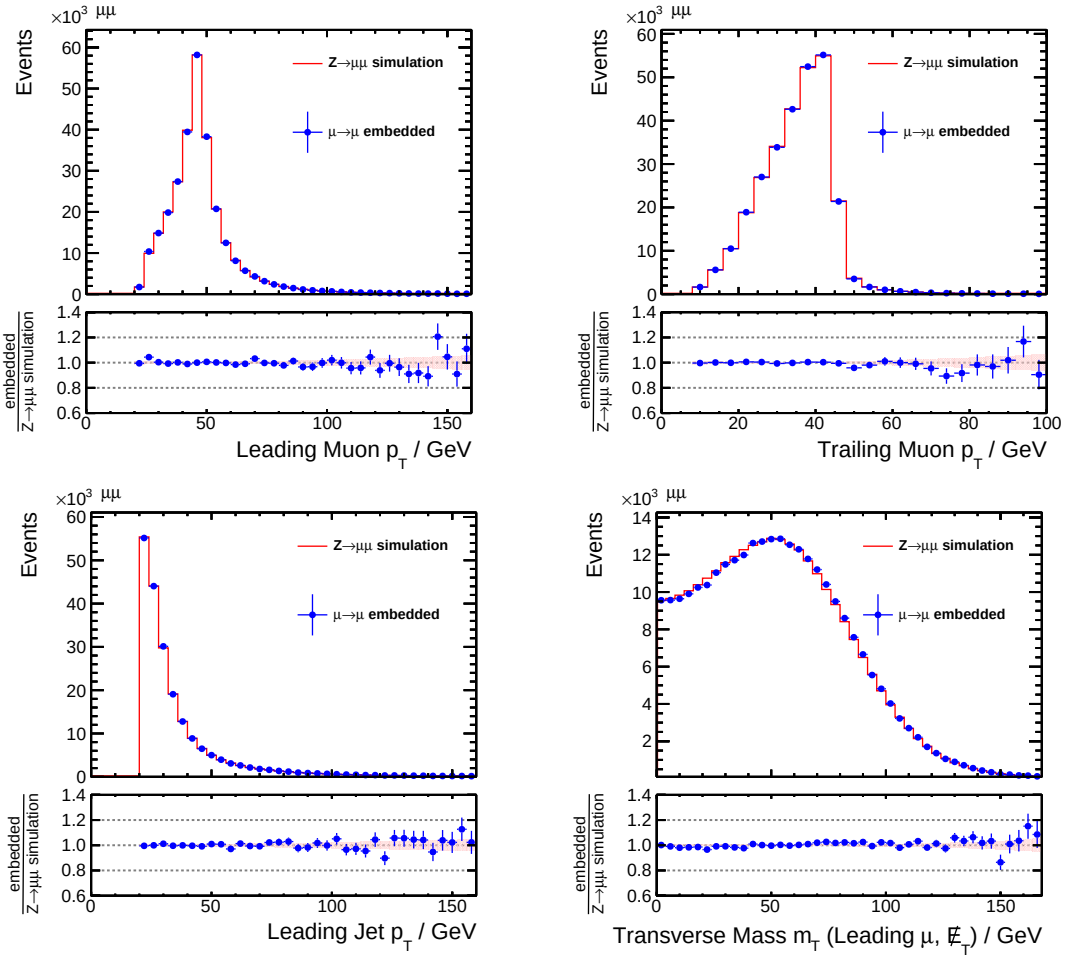


Figure 5.8: Comparison of p_T variables of the simulated and embedded di-muon events, p_T of the leading jet in the respective events as well as missing transverse energy.

5.2 Study of $\mu \to \tau$ embedded events

When applying the $\mu \to \tau$ embedding method, the selected muons are replaced by τ-leptons, which are then used to describe $Z \to \tau\tau$ events. The same $Z \to \mu\mu$ events as in 5.1 are used, this time replacing the muons by τ-leptons and enforcing a desired final state as described in section 3.3. The embedded events are compared to the simulated $Z \to \tau\tau$ events that are produced under the same conditions as the initially selected $Z \to \mu\mu$ events. A graphical description of this study is shown in figure 5.9. In this section, comparisons of $\mu \to \tau$ embedded events in the $\mu\tau_h$ and $e\tau_h$ final states are shown.

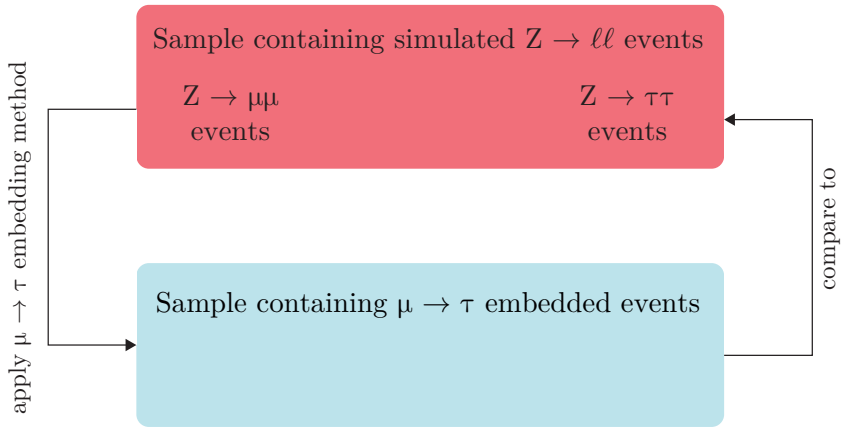


Figure 5.9: Procedure of the study using $\mu \to \tau$ embedded events selected from simulation.

The selection criteria are shown in table 5.2. They are aligned with the criteria used for the analysis in chapter 6, however with a looser p_T selection to increase statistics. Again, an opposite charge of the two τ-decay products is required.

Table 5.2: Event selection of the τ decay products in the $\mu\tau_h$ and $e\tau_h$ final states.

Channel	Transverse momentum		Pseudorapidity		τ_h ID	$I_{\text{rel}}^{\mu/e}$
$\mu\tau_h$	$p_T^{\mu} > 10\,\text{GeV}$	$p_T^{\tau_h} > 20\,\text{GeV}$	$\lvert\eta^{\mu}\rvert < 2.1$	$\lvert\eta^{\tau_h}\rvert < 2.3$	Tight	< 0.15
$e\tau_h$	$p_T^{e} > 22\,\text{GeV}$	$p_T^{\tau_h} > 22\,\text{GeV}$	$\lvert\eta^{e}\rvert < 2.1$	$\lvert\eta^{\tau_h}\rvert < 2.3$	Tight	< 0.10

A smearing effect due to the initial muon reconstruction was visible in the $\mu \to \mu$ embedded events. The same effect was examined for $\mu \to \tau$ embedded events. The effect is expected to be smaller as energy carried away by neutrinos in the decay of the τ-lepton will shift the visible di-τ mass distribution to a lower mean and larger variance, which will dominate the smearing of the Z peak. In figure 5.10, a zoom on the peak region of the visible di-τ mass in shown. Both corrected and uncorrected according to section 4.2 show no deviation from the visible di-τ mass distribution of simulated $Z \to \tau\tau$ events.

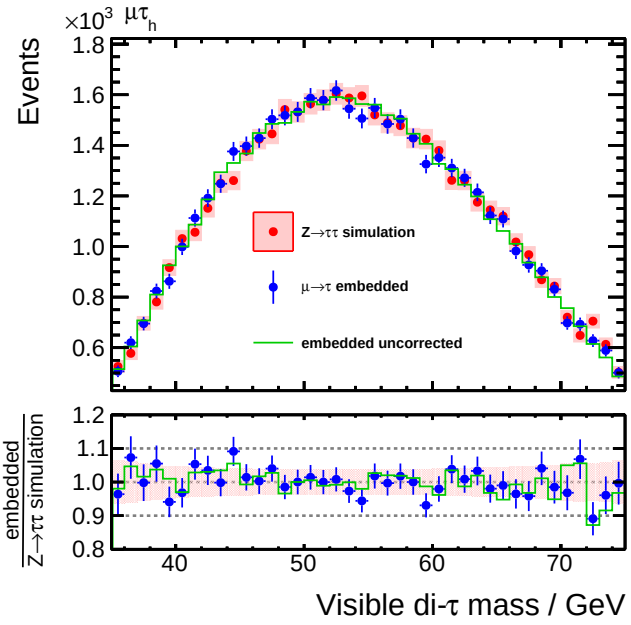


Figure 5.10: Zoom on the peak area of the visible mass of the di-tau system with the applied corrections of figure 4.5. Additionally, shown is the spectrum for uncorrected energies of the selected muons. In contrast to the di-muon mass of $\mu \to \mu$ embedded events in figure 4.6, the smearing caused by the initial muon reconstruction does not impact the visible di-tau mass within the statistical accuracy of this study.

Additional variables of the embedded $Z \to \tau\tau$ events in comparison with simulated $Z \to \tau\tau$ events are shown in figures 5.11 and 5.12 for the $\mu\tau_h$ and $e\tau_h$ decay channels respectively. The distributions of transverse momenta for all τ decay products show no significant deviations within the statistical accuracy. The number of embedded events are larger than the number of simulated $Z \to \tau\tau$ events, as the number of simulated $Z \to \tau\tau$ events is decreased by the respective branching ratio of the di-τ final state. For the production of embedded events however, this decrease is avoided by the enforcing of a final state, such that a selection $Z \to \mu\mu$ event can be used for each final state. The number of simulated events is enough for an investigation of possible effects in energy regimes around the Z mass, where agreement is observed. The effect on missing transverse momentum for low-$\not{E}_T$ events in figures 5.11 (d) and 5.12 (d) is seen just as it was for $\mu \to \mu$ embedded events. Overall, kinematic variables are reproduced well by the embedded events and no visible effects are introduced, which gives confidence in the robustness of the method for as a $Z \to \tau\tau$ background estimate for data.

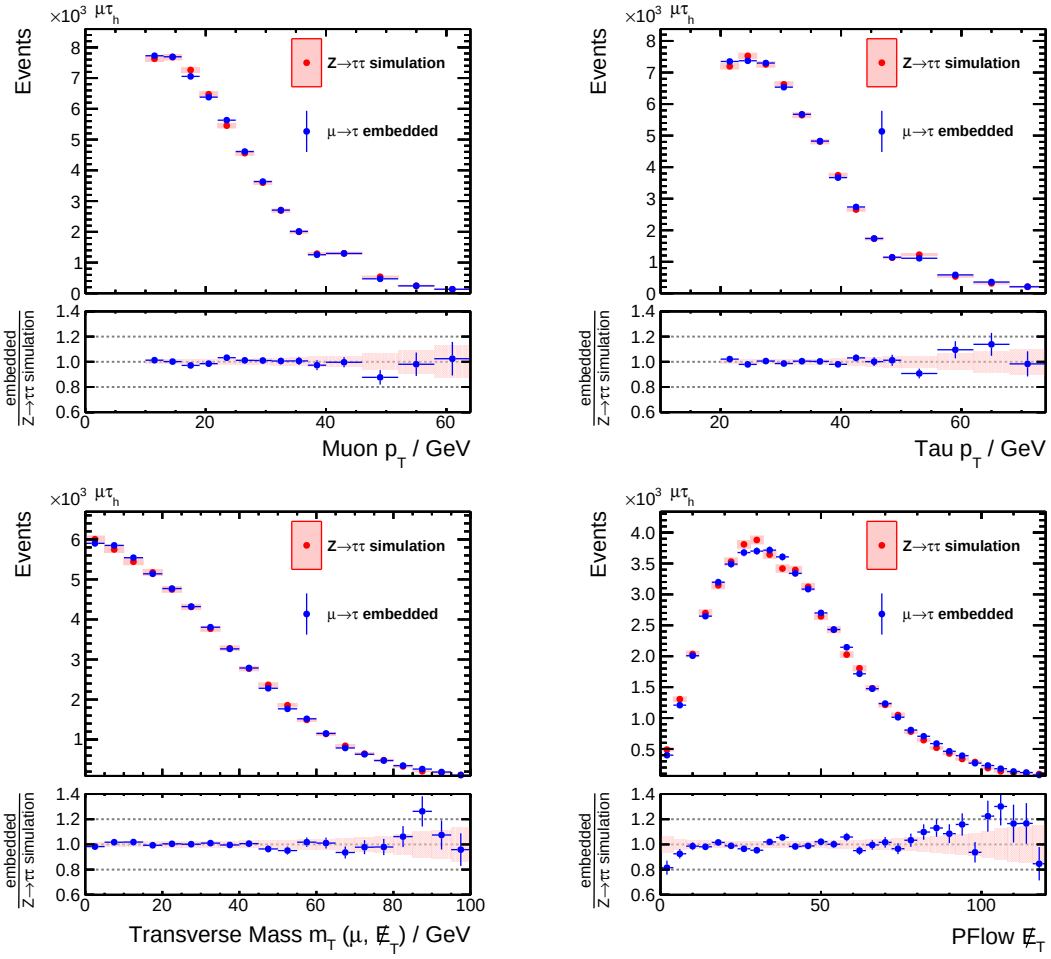
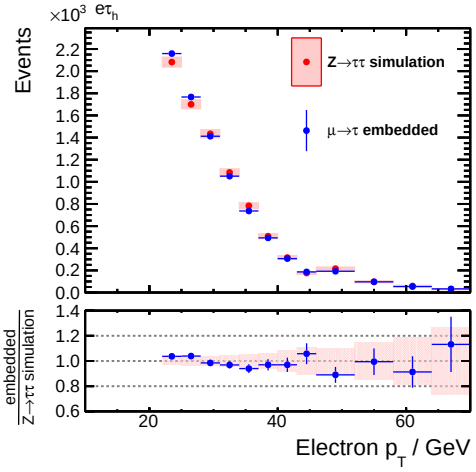
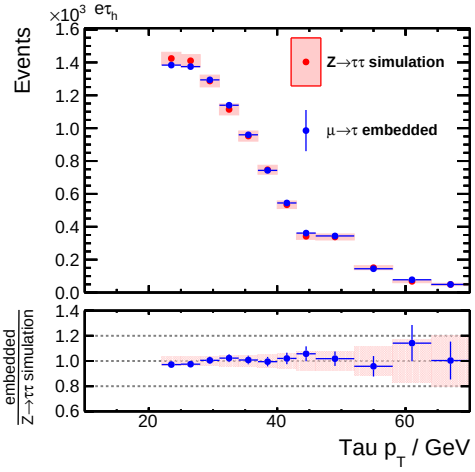


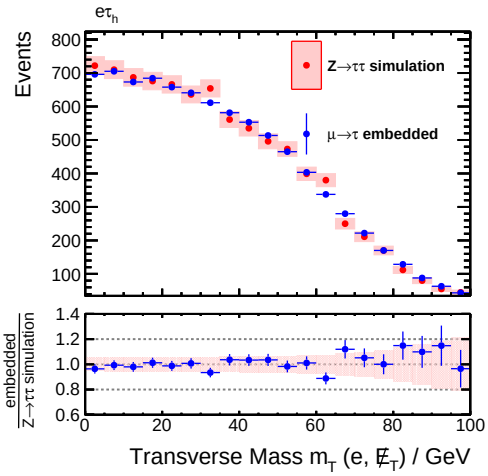
Figure 5.11: Selection of kinematic variables from the study of $\mu \to \tau$ embedded events produced on a sample of simulated $Z \to \mu\mu$ events. The embedded events are compared to simulated $Z \to \tau\tau$ events produced under the same conditions. Shown are di-τ decays in the $\mu\tau_h$ final state. No effects can be seen in the description of the leptons by applying the embedding method. The effect for low values of $\not{E}_T$ is discussed in section 5.1.1.



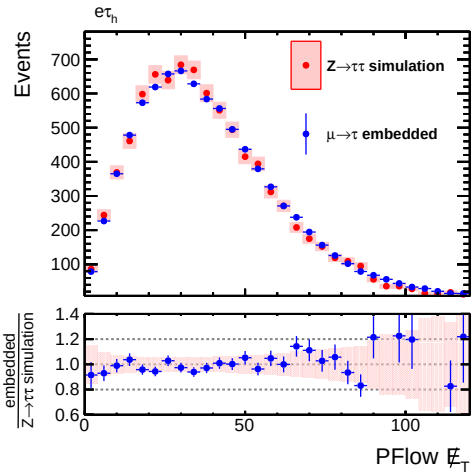
(a) Transverse momentum of the electron.



(b) Transverse momentum of the hadronically decaying τ-lepton.



(c) Transverse mass $m_T(e, \not{E}_T)$.



(d) Missing transverse energy.

Figure 5.12: Selection of kinematic variables from the study of $\mu \to \tau$ embedded events produced on a sample of simulated $Z \to \mu\mu$ events. The embedded events are compared to simulated $Z \to \tau\tau$ events produced under the same conditions. Shown are di-τ decays in the $e\tau_h$ final state. No effects can be seen in the description of the leptons by applying the embedding method. The effect for low values of $\not{E}_T$ is discussed in section 5.1.1.

CHAPTER 6

Validation of the MSSM $H/A \to \tau\tau$ analysis

In this chapter, the analysis and the results of a search for additional neutral Higgs bosons as predicted by the MSSM are presented. The search is conducted in the di-τ final state and features a background description of $Z \to \tau\tau$ events using $\mu \to \tau$ embedded events as a data-driven alternative to simulated $Z \to \tau\tau$ events. The model-independent search for a new particle can in its simplest form be thought of as a counting experiment: All events showing characteristics of an $H/A \to \tau\tau$ event are selected, counted, and then compared with the theoretical predictions of all Standard Model processes which are expected to show the same characteristics. The comparison provides a test of a *background-only* versus *signal+background* hypothesis. For this, the full dataset collected by the CMS experiment in 2016 is used, which amounts to an integrated luminosity of $\mathcal{L} = 35.9\,\text{fb}^{-1}$ at a center of mass energy of 13 TeV. The $\mu \to \tau$ embedded events used for background description have been produced from $Z \to \mu\mu$ events from the same dataset. Those parts of the analysis which are not specific to the description of $Z \to \tau\tau$ events, such as event selection, categorization and description of additional backgrounds, have been investigated in the context of the analysis that is to be published by the CMS collaboration [36–38].
In this chapter, the analysis strategy, background modeling, systematic uncertainties and, finally, the results derived by use of embedded events will be discussed. The results will be compared to the published results. The embedded event sample as described in this thesis has been used in the analysis published by CMS as a cross-check of the $Z \to \tau\tau$ background modeling. As $Z \to \tau\tau$ events form a dominant source of background in $H/A \to \tau\tau$ searches, the validation by an alternative method is of particular importance.

6.1 Analysis strategy

This section serves as a short outline of the analysis strategy used for the search for MSSM $H/A \to \tau\tau$ events using embedded events. Here, the main focus is laid upon the embedded $Z \to \tau\tau$ estimate replacing the simulation-based estimation the published analysis. For a more detailed discussion of the analysis strategy and backgrounds not specific to $Z \to \tau\tau$ events, refer to [36–38].

6.1.1 Event selection

The analysis presented in the following is performed in the three most sensitive di-τ decay channels, which are the $\mu\tau_h$, $e\tau_h$ and $\tau_h\tau_h$ final states. These final states cover 88% of all di-τ events. The decay channels used are illustrated in figure 6.1.

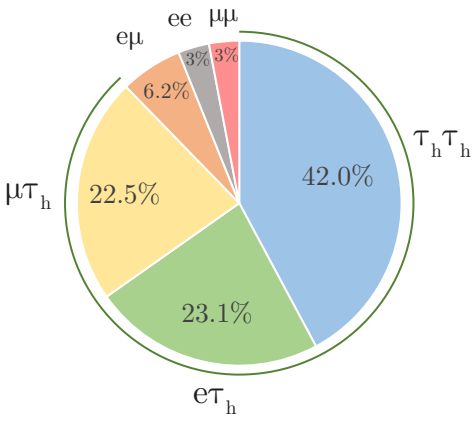


Figure 6.1: Branching ratios of individual di-τ final states. The channels used for the MSSM $H/A \to \tau\tau$ analysis presented in this chapter are emphasized by the green circle.

For each final state, a dedicated event selection is applied. The online selection on trigger level requires at least one muon with $p_T > 22\,\text{GeV}$ and $|\eta| < 2.1$ for the $\mu\tau_h$ final state or electron with $p_T > 25\,\text{GeV}$ and $|\eta| < 2.1$ for the $e\tau_h$ final state. For the $\tau_h\tau_h$ final state, the trigger decision is based on two hadronic τ-leptons $p_T > 35\,\text{GeV}$ and $|\eta| < 2.1$. The offline selection imposes further kinematic requirements which are given in table 6.1.

Table 6.1: Kinematic event selection of the τ decay products in the $\mu\tau_h$, $e\tau_h$ and $\tau_h\tau_h$ final states.

Final state	Transverse momentum		Pseudorapidity	
$\mu\tau_h$	$p_T^{\mu} > 23\,\text{GeV}$	$p_T^{\tau_h} > 30\,\text{GeV}$	$\lvert\eta^{\mu}\rvert < 2.1$	$\lvert\eta^{\tau_h}\rvert < 2.3$
$e\tau_h$	$p_T^{e} > 26\,\text{GeV}$	$p_T^{\tau_h} > 30\,\text{GeV}$	$\lvert\eta^{e}\rvert < 2.1$	$\lvert\eta^{\tau_h}\rvert < 2.3$
$\tau_h\tau_h$	$p_T^{\tau_h} > 40\,\text{GeV}$		$\lvert\eta^{\tau_h}\rvert < 2.1$	

Furthermore, muons and electrons are required to pass the identification criteria with efficiencies of 99% and 80% respectively and must be isolated using the relative isolation variable $I_{\text{rel}}^{\mu/e}$ defined in equation 2.1. The relative isolation requirements used in this analysis are $I_{\text{rel}}^{\mu} < 0.15$ and $I_{\text{rel}}^{e} < 0.1$.
Hadronic τ-decays are identified using the MVA-based discriminant described in section 2.2.3. For the semi-leptonic final states, the *tight* working point is required, while in the $\tau_h\tau_h$ final state, both hadronically decaying τ-leptons are required to pass the *medium* working point. All requirements chosen are summarized in table 6.2. Events not matching the requirements listed in tables 6.1 and 6.2, as well as events where the charges of the two decay products are of equal sign are rejected from the analysis. Events where additional muons and electrons are reconstructed which fulfill looser requirements of $p_T^{\mu/e} > 10\,\text{GeV}$, $|\eta^{\mu}| < 2.4$, $|\eta^{e}| < 2.5$ and $I_{\text{rel}}^{\mu/e} < 0.3$ are rejected in order to suppress background and to avoid an attribution of a single event to more than one final state. The number of embedded events remaining after event selection is shown in table 6.3 in comparison to the number of simulated $Z \to \tau\tau$ events that are used for the published analysis. The number of embedded events is increased by a factor of 4.9 compared to the simulated events.

Table 6.2: Lepton isolation requirements with the respective efficiencies and misidentification rates for the e, μ and τ_h identification as well as anti-muon and anti-electron discriminators for the $\mu\tau_h$, $e\tau_h$ and $\tau_h\tau_h$ final states. Average values are given, as the respective efficiencies and misidentification rates depend on the transverse momentum of the lepton and the detector region where the lepton is reconstructed. The reconstruction of leptons at CMS is described in section 2.2.3.

Final state		μ/e ident.	τ_h ident.	τ_h anti-muon	τ_h anti-electron
$\mu\tau_h$	Efficiency	99%	27%	98%	94%
	Misid. rate	$1 \cdot 10^{-3}$	$4 \cdot 10^{-4}$	$3 \cdot 10^{-4}$	$2 \cdot 10^{-2}$
$e\tau_h$	Efficiency	80%	27%	99%	78%
	Misid. rate	$5 \cdot 10^{-2}$	$4 \cdot 10^{-4}$	$5 \cdot 10^{-4}$	$6 \cdot 10^{-4}$
$\tau_h\tau_h$	Efficiency	-	51%	99%	91%
	Misid. rate	-	$3 \cdot 10^{-3}$	$5 \cdot 10^{-4}$	$4 \cdot 10^{-3}$

Table 6.3: Remaining number of events used for the $Z \to \tau\tau$ estimation by embedded events in this analysis after the selection shown in tables 6.1 and 6.2. In comparison, the number of simulated $Z \to \tau\tau$ events that is used for the analysis published by CMS is given.

Final state	$\mu \to \tau$ embedded events	Simulated $Z \to \tau\tau$ events	$\frac{\#\text{ embedded}}{\#\text{ simulated}}$
$\mu\tau_h$	2,204,022	519,490	4.2
$e\tau_h$	1,739,554	291,302	6.0
$\tau_h\tau_h$	742,532	155,231	4.8
Total	4,686,108	966,023	4.9

6.1.2 Categorization

After having been selected, events are categorized in their respective final states. The purpose of this categorization is to increase sensitivity for the two H/A production modes introduced in section 1.3, namely the gluon-fusion ($gg\phi$) and the b-associated production ($bb\phi$). The symbol ϕ refers to a production of any of the two neutral heavy Higgs bosons $\phi = H/A$. The categorization is made in two steps:

1. Initially, the presence of a jet with $p_T > 20\,\text{GeV}$ and $|\eta| < 2.4$ that is tagged as a b-quark jet results in this event being assigned to the `b-tag` category. Otherwise the event is assigned to the `no b-tag` category. The tagging of b-jets is done using the combined secondary vertex algorithm at a high-purity working point resulting in a b-tagging efficiency of 70% and a light-parton misidentification probability of 1.5% [39]. The `b-tag` category is especially sensitive to the b-associated production, the rate of expected ggϕ to bbϕ is 1:3 in this category. The `no b-tag` category is sensitive to both production modes with an expected rate of 1:1, assuming an equal cross section of the two production modes.

2. Furthermore, the $\mu\tau_h$ and $e\tau_h$ final states are split in two additional categories depending on the transverse mass in the final state as defined in equation 4.4. Events with $m_{\text{T}}^{\mu/e} < 40\,\text{GeV}$ are assigned to the `tight` $\mathtt{m_T}$ category, events with $40 < m_{\text{T}}^{\mu/e} < 70\,\text{GeV}$ to the `loose` $\mathtt{m_T}$ category. The respective categories increase the sensitivity for different mass hypotheses of the H/A. For low mass hypotheses, most signal events are expected in the `tight` $\mathtt{m_T}$ category, whereas the `loose` $\mathtt{m_T}$ category is sensitive to high-mass $H/A \to \tau\tau$ signals. Events with $m_{\text{T}}^{\mu/e} > 70\,\text{GeV}$ are rejected to improve the ratio of potential signal over expected background events, which in this regime are dominated by W boson production in association with jets.

The two semi-leptonic final states are thus split into four categories each and the full-hadronic final state in two categories, resulting in a total number of ten exclusive categories. The categorization of events for the three decay channels is illustrated in figure 6.2.

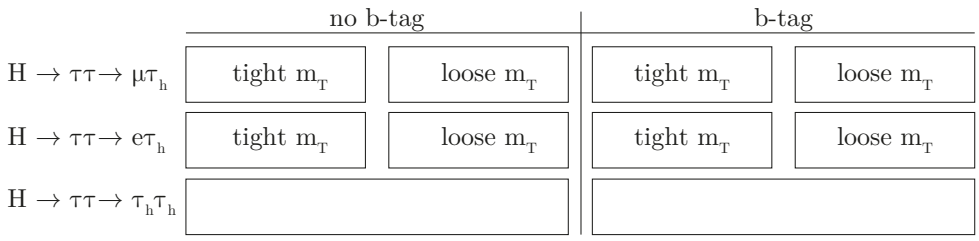


Figure 6.2: Categorization of the analysis in ten sub-categories depending on final state, presence of b-tagged jets, and transverse mass.

6.2 Additional background processes

Below a list of all Standard Model background processes that are considered for the analysis is given. They are sorted according to the presence of a genuine di-τ pair, the presence of a prompt muon or electron, as well as events that pass the selection criteria due to a misidentification of either lepton or jet as hadronic τ-lepton.

- Genuine di-τ events
  - $Z \to \tau\tau$
  - $t\bar{t} \to \tau\tau$
- Events with genuine τ_h and prompt ℓ ($\ell = \mu, e$) or jet misidentified as ℓ
  - $t\bar{t} \to \ell\tau_h$
  - $t\bar{t} \to \tau_h$+jets
- Events with prompt ℓ, where one is misidentified as τ_h
  - $Z \to \ell\ell$
  - $t\bar{t} \to \ell\ell$
- Events with jets misidentified as τ_h (jet$\to \tau_h$ fakes)
  - $Z \to \ell\ell$+jets
  - $t\bar{t} \to \ell$+jets
  - Di-boson
  - Single top
  - W production in association with jets
  - QCD multijet production

For presentation purposes, similar background processes that are estimated by the same method are collected, which results in five background contributions that will be used for referencing and labeling. In this collection, all events in which jets are misidentified as τ_h will be referred to as *jet*$\to \tau_h$ *fakes* and estimated by a data driven *fake-factor* method, which is explained in detail in [36–38]. Background events derived using embedded events incorporate both $Z \to \tau\tau$ and $t\bar{t} \to \tau\tau$ events and will be labeled $\mu \to \tau$ *embedded*. Events in which the Z boson decays into two muons or electrons and the selection therefore is attributed to a $\ell \to \tau_h$ misidentification will be estimated using simulated $Z \to \ell\ell$ events. The remaining background events that do not result from a jet$\to \tau_h$ misidentification are labeled *Electroweak* (W boson, di-boson and single top) and $t\bar{t}$. Background events that are labeled $t\bar{t}$ exclude the decay of a top-quark pair into a $\tau\tau$ final state but incorporate $t\bar{t}$ decays into an $\ell(\ell/\tau_h)$ final state. The five background contributions and their estimation methods for this analysis are summarized in table 6.4.

Table 6.4: Estimation methods for the background processes. The backgrounds are either estimated by data driven methods or simulation.

Background label	Estimated by...
$\mu \to \tau$ embedded	Embedded events (includes $Z \to \tau\tau$ and $t\bar{t} \to \tau\tau$)
jet$\to \tau_h$ fakes	Fake-factor method
$Z \to \ell\ell$	Simulation
Electroweak	Simulation
$t\bar{t}$	Simulation (excludes $t\bar{t} \to \tau\tau$)

As the two most prominent backgrounds ($Z \to \tau\tau$ and jet$\to \tau_h$ fakes) are estimated from data, this analysis demonstrates the up to now largest use of data-driven background estimation methods for a MSSM $H/A \to \tau\tau$ search on the RunII-data taken by the CMS experiment.

6.3 Final discriminator

As a final discriminating variable, the total transverse mass $m_{\mathrm{T}}^{\mathrm{tot}}$ is used, which is defined as

$$m_{\mathrm{T}}^{\mathrm{tot}} = \sqrt{m_{\mathrm{T}}^2(\not{E}_T, \tau_1^{\mathrm{vis}}) + m_{\mathrm{T}}^2(\not{E}_T, \tau_h) + m_{\mathrm{T}}^2(\tau_1^{\mathrm{vis}}, \tau_h)} \; . \tag{6.1}$$

$\not{E}_T$ refers to the magnitude of the missing transverse momentum, τ_1 to the first visible τ decay product (μ, e or τ_h for the $\mu\tau_h$, $e\tau_h$ and $\tau_h\tau_h$ final states), and m_{T} to the transverse mass of two objects (1,2) defined as

$$m_{\mathrm{T}}(1,2) = \sqrt{2p_T(1)p_T(2)(1 - \cos(\Delta\phi(1,2)))} \; . \tag{6.2}$$

All following figures will be shown as function of this discriminating variable.

6.4 Uncertainty model for embedded events

Due to the hybrid nature of embedded events, a dedicated uncertainty model taking the nature of embedded events into account has to be applied. The adaptions from the uncertainty used for simulated events are two-fold: First, uncertainties introduced for corrections on the $Z \to \tau\tau$ estimation that became obsolete with the switch to embedded events are omitted. Second, new uncertainties need to be introduced to account for corrections applied to the embedded events. In this section, only uncertainties that are omitted or introduced for the use of $\mu \to \tau$ embedded events are discussed. For a full discussion of the uncertainty model of $Z \to \tau\tau$ events derived by simulation, as well the additional background estimation methods not concerning $\mu \to \tau$ embedded events, refer to [36–38].

The following uncertainties are obsolete for the description of $Z \to \tau\tau$ background when utilizing embedded events:

- **Jet energy scale**
  Jets are kept from selected the $Z \to \mu\mu$ events in data. They will be reconstructed as data, so the jet energy scale of embedded events is identical to data. A jet energy scale uncertainty does not apply.

- **b-tag scale factors**
  As b-quark jets will not be removed by the event cleaning, b-quark jets will be tagged from data and the b-tag efficiency and misidentification rate will be the same. There are no corresponding scale factors applied for embedded events.

- **Z boson recoil uncertainties**
  Embedded events provide the correct description of the missing transverse energy $\not{E}_T$ without the need for recoil corrections as in the case for simulated events. The corresponding uncertainties do not apply.

- **Drell-Yan reweighting**
  In simulated $Z \to \tau\tau$ events, the complete event characteristics of the collisions of proton bunches at the LHC, each containing 10^{11} protons, are simulated. The simulation of the hadronic processes is a difficult task and causes discrepancies between the simulated events and the observed data. The remedy is a reweighting procedure, where the pileup distribution of simulated events is fitted to the observed distribution. Embedded events do not require this reweighting, and the distribution of pileup and related event characteristics of the proton-proton collisions at the LHC are described without further need for corrections.

The following uncertainties enter for the $Z \to \tau\tau$ background description when using the embedded event sample.

- **Initial muon reconstruction**
  The $Z \to \tau\tau$ decay in embedded samples is simulated using a Z mass reconstructed from the initially selected di-muon system. Any resolution effects on the energy reconstruction of this first step is therefore propagated to the embedded events. This introduces an additional smearing on the Z peak which has to be accounted for. A full description of the effect and the corresponding correction is given in section 4.2. This effect is small and corrected in future embedded samples with correction factors of 0.4% and below on the reconstructed di-muon mass. The effect is therefore well covered for reconstruction of electrons and τ_h, at which uncertainties on the energy reconstruction ranging from 0.7-2.3% are assigned for simulated $Z \to \tau\tau$ events in [36]. The uncertainties due to the initial muon reconstruction are subdominant to these energy scale uncertainties for electrons and hadronic τ-leptons and no new uncertainties are applied. An additional energy scale uncertainty of 1% is assigned for reconstructed muons. The uncertainty affects all categories likewise and is most relevant for embedded events around the Z peak. Figure 6.3 shows the uncertainty bands for the `no b-tag tight` $\mathtt{m_T}$ categories.

- **$t\bar{t}$ contamination**
  When using embedded events, a veto is applied on simulated $t\bar{t}$ events, which is shown in detail in section 4.3. To cover this, a 10% uncertainty is assigned on the shape of the removed events in the $t\bar{t}$ simulation. As these are expected to be the same number and shape as the additional events selected from $t\bar{t}$ in the embedded event sample, the uncertainty is linearly correlated with the $t\bar{t}$ contamination present in a certain category and $m_{\mathrm{T}}^{\mathrm{tot}}$-bin. The uncertainty bands are displayed in figure 6.4. The uncertainty is most prominent in b-tagged categories, which is expected since the number of selected $t\bar{t}$ events in the embedded event sample is highest there. It is on the sub-percent level for the `no b-tag tight` $\mathtt{m_T}$ and $\mathcal{O}(1)\%$ for the `no b-tag loose` $\mathtt{m_T}$ categories. Nevertheless, the uncertainty becomes relevant in the mid-mass ranges of $m_{\mathrm{T}}^{\mathrm{tot}}$ in `b-tag` categories exceeding 5%.

A comparison of distributions as a function of $m_{\mathrm{T}}^{\mathrm{tot}}$ as well as respective statistical and systematic uncertainties between embedded and simulated events is shown in figure 6.5. Given the $t\bar{t}$-contamination present in embedded events, embedded events should not be compared directly to the simulated $Z \to \tau\tau$ events. Instead, a valid comparison is chosen, which is the comparison of the distribution of the sum of expected $Z \to \tau\tau + t\bar{t}$ events, in which the $t\bar{t} \to \tau\tau$ are vetoed from simulated $t\bar{t}$ events when being added to the embedded $Z \to \tau\tau$ distribution.
For b-tag categories, the additional $t\bar{t}$ contamination uncertainty increases the overall uncertainty for $m_{\mathrm{T}}^{\mathrm{tot}}$ values larger than 100 GeV. For `no b-tag` categories, the resulting uncertainties when using embedded events are comparable to the uncertainties applied for the use of simulated $Z \to \tau\tau$ events for $m_{\mathrm{T}}^{\mathrm{tot}} \lesssim 200\,\mathrm{GeV}$ and smaller than the ones used for simulated events for larger $m_{\mathrm{T}}^{\mathrm{tot}}$ values. In this regime, the statistical uncertainties become dominant which are smaller for embedded events.

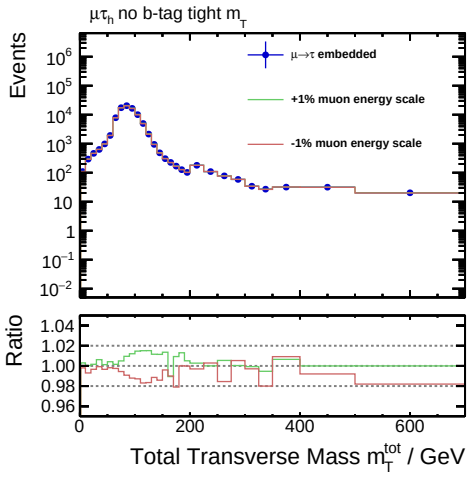


Figure 6.3: $m_{\mathrm{T}}^{\mathrm{tot}}$ distribution of embedded events in the `no b-tag tight` $\mathtt{m_T}$ category of the $\mu\tau_h$ final state. The variation of the muon energy scale by $\pm 1\%$ is shown as green and red lines, illustrating the uncertainty bands of the introduced muon energy scale uncertainty. Events with $m_{\mathrm{T}}^{\mathrm{tot}}$ values ranging between 50 and 200 GeV are most affected by this uncertainty.

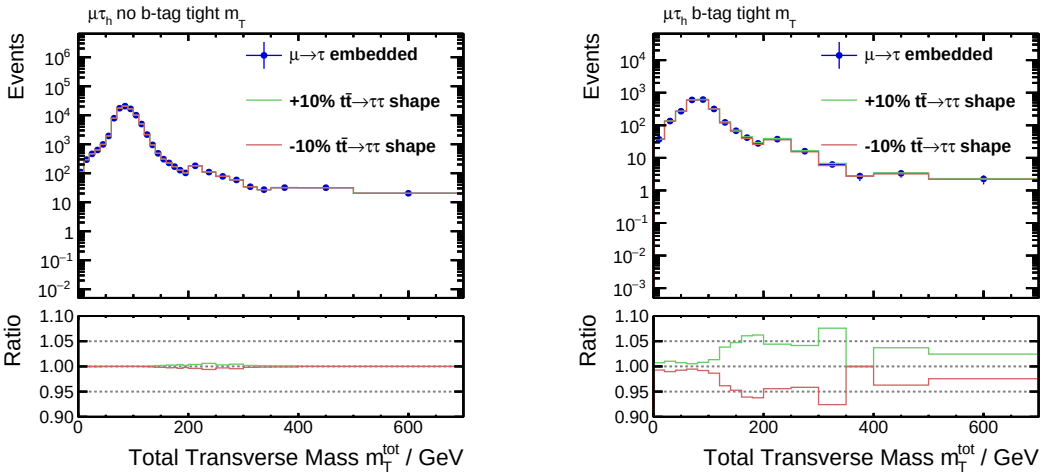


Figure 6.4: $m_{\mathrm{T}}^{\mathrm{tot}}$ distributions of embedded events in the `no b-tag tight` $\mathtt{m_T}$ (left) and `b-tag tight` $\mathtt{m_T}$ (right) categories of the $\mu\tau_h$ final state. In green and red, the variations of $\pm 10\%$ of the shape of the events attributed to $t\bar{t}$ contamination in embedded events are shown. The uncertainty depends on the number of expected $t\bar{t} \to \tau\tau$ decays in the respective category, which is highest in the categories that include b-tagged jets, as can be seen in the right $m_{\mathrm{T}}^{\mathrm{tot}}$ distribution.

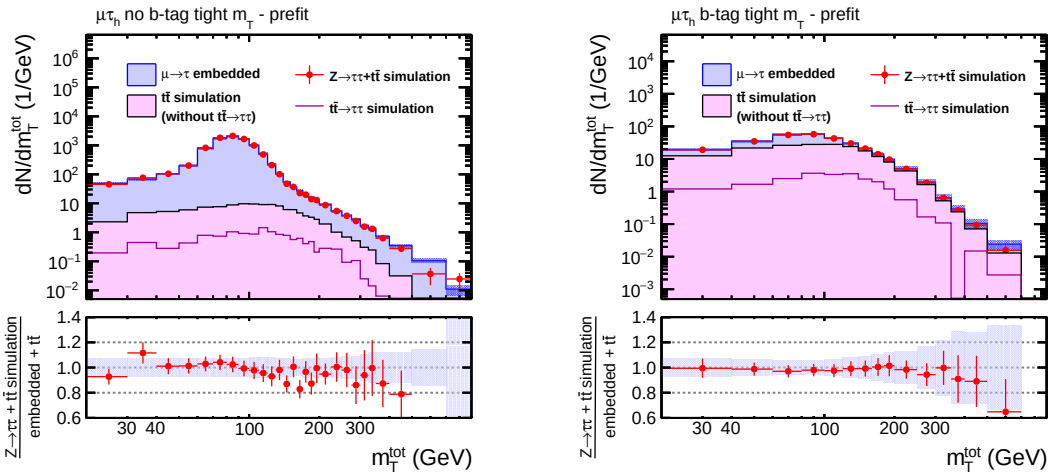


Figure 6.5: Comparison of $m_{\mathrm{T}}^{\mathrm{tot}}$ distributions for the `no b-tag tight` $\mathtt{m_T}$ and `b-tag tight` $\mathtt{m_T}$ categories in the $\mu\tau_h$ final state. For the use of embedded events, $t\bar{t} \to \tau\tau$ events are vetoed from the simulated $t\bar{t}$ events, as they are contained in the embedded event sample. Simulated $t\bar{t}$ events in which this $t\bar{t} \to \tau\tau$ contribution is removed are shown as a bright purple histogram. The $\mu \to \tau$ embedded events are shown as a blue histogram which is stacked on the histogram containing the $t\bar{t}$ events. This stacked histogram represents the background estimate for $Z \to \tau\tau$ and $t\bar{t}$ events of this analysis made with the use of embedded events. The dark purple line indicates the $t\bar{t} \to \tau\tau$ events that were removed for the use of embedded events. The sum of all simulated $t\bar{t}$ events, including $t\bar{t} \to \tau\tau$, with the simulated $Z \to \tau\tau$ events used in the published analysis is shown by the red dots. These red dots therefore represent the background estimate for $Z \to \tau\tau$ and $t\bar{t}$ events of the published analysis. The blue shaded band and red errorbars show the combined statistical and systematic (prefit) uncertainty of $\mu \to \tau$ embedded events and simulated $Z \to \tau\tau$ events respectively.

6.5 Results

6.5.1 Distributions as a function of total transverse mass

After the adaptions described above, the distributions as a function of the discriminating variable $m_{\mathrm{T}}^{\mathrm{tot}}$ are subject to a binned maximum likelihood fit. The fit is applied to all final states and categories. Figure 6.6 shows the distributions for $m_{\mathrm{T}}^{\mathrm{tot}}$ as histograms for the $\mu\tau_h$ and $e\tau_h$ final states in the `no b-tag tight` $\mathtt{m_T}$ category, and for $\tau_h\tau_h$ in the `b-tag` category after this fit has been performed for a background-only hypothesis. The histograms are divided by the respective bin-width for presentation purposes. The stacked multi-colored histograms represent the five background contributions as shown in table 6.4 that add up to the total expected background. Their cumulative uncertainty is indicated by the shaded band. The y-axis on the plot is split between logarithmic and linear scale as is indicated by the black line. This allows a visibility of $m_{\mathrm{T}}^{\mathrm{tot}}$ over the full mass range as well as shapes of selected signal events, while also showing the dominant background processes in the peak region. The signals shown for the $\mu\tau_h$ and $e\tau_h$ final state indicate a decay of a neutral Higgs boson at a mass of 130 GeV and a cross-section of 10 pb for a production via gluon-fusion and a b-associated production respectively. In the $\tau_h\tau_h$ final state, the signal shape of a heavy neutral Higgs boson at a mass of 700 GeV and a cross-section of 0.01 pb is indicated. In the categories shown, $Z \to \tau\tau$ events form a dominant source of background. The categories are thus most affected by the change of the estimation method from simulated to embedded events. The $m_{\mathrm{T}}^{\mathrm{tot}}$ distributions of the ten categories are given in A.18-A.21. All distributions show a good agreement between the background estimation and the observation.

6.5.2 Upper Limits on H/A cross-section

The search for a heavy Higgs boson is performed for the two most relevant Higgs production modes as explained in section 1.3, which are production via gluon-fusion ($gg\phi$) and b-associated production ($bb\phi$). For the two production modes, the search is made model-independently using templates derived by simulating 28 Higgs boson mass peaks each with masses ranging from 90-3200 GeV. The signal events are calculated at leading-order for the gluon-fusion production and at next-to-leading-order for the b-associated production.
No signal is observed. Thus, the results of the search are presented as 95% confidence level (CL) upper limits on the product of the cross-section for the production of a neutral Higgs boson with the branching ratio into a di-τ final state. Figure 6.7 shows the expected and observed limits of the individual final states. In the gluon-fusion production, the expected limits are driven mainly by the $\mu\tau_h$ and $\tau_h\tau_h$ final states for the low and high mass points respectively, whereas the $\tau_h\tau_h$ final state dominates the sensitivity in the b-associated production mode. The $e\tau_h$ final state improves the overall sensitivity of the analysis mainly for the low- to mid-mass hypotheses.

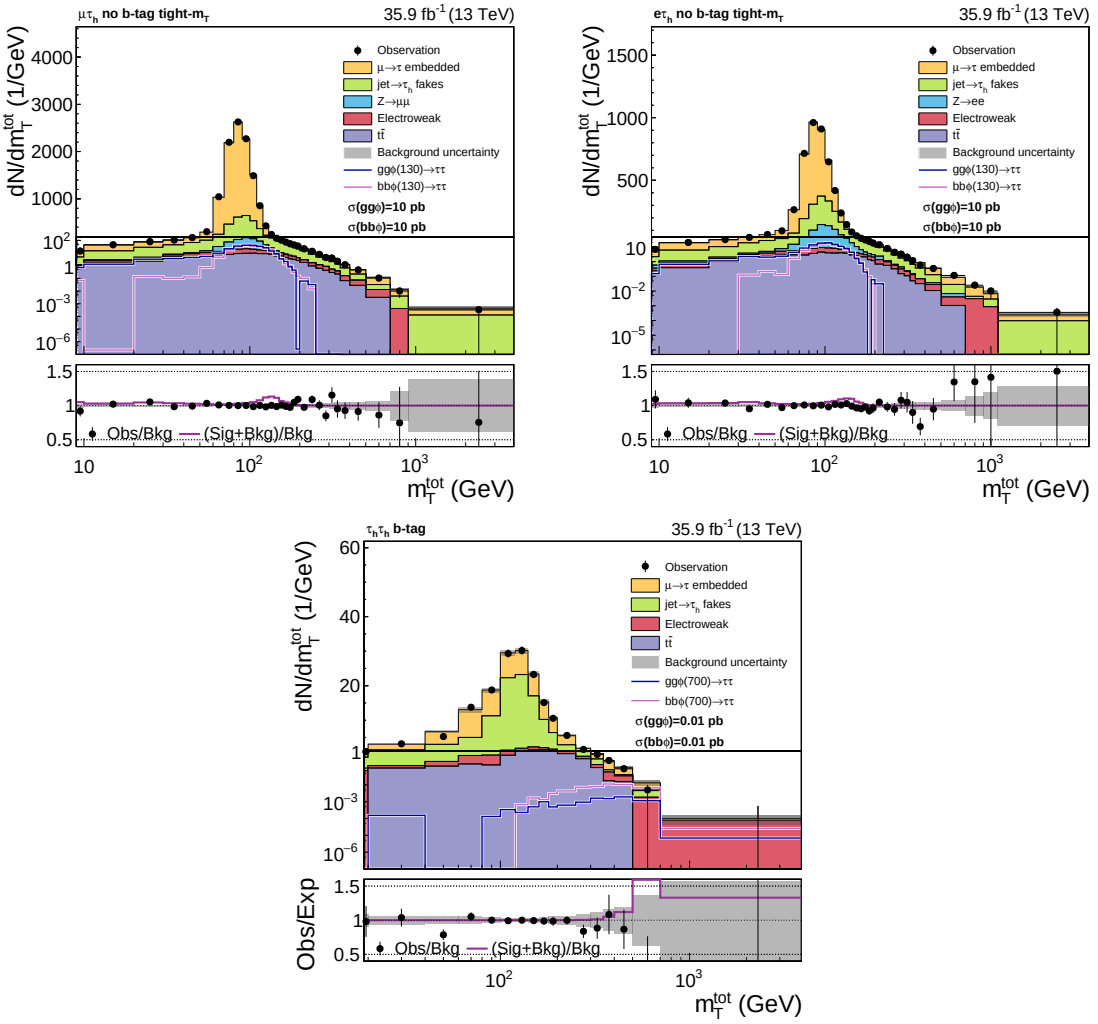


Figure 6.6: $m_{\mathrm{T}}^{\mathrm{tot}}$ distributions in the $\mu\tau_h$ and $e\tau_h$ `no b-tag tight` $\mathtt{m_T}$ and the $\tau_h\tau_h$ `b-tag` categories. The stacked multi-colored histograms represent the background estimation. Observed data are represented by black points, with their statistical uncertainty indicated by error bars.

The observed limits for the combination of all final states are shown in figure 6.8 in comparison to the expected ones. The confidence intervals on the expected limits are shown as a green and yellow band for the $\pm 1\sigma$ and $\pm 2\sigma$ confidence intervals, representing a 68% and 95% confidence interval respectively. No significant deviation between expected and observed limits is observed. The largest deviation is found at a mass of 160 GeV for the gluon-fusion production. This deviation is still contained in the 95% confidence interval.
The relative differences between observed and expected limit is shown in figure 6.9 for both the combination of decay channels, as well as for the channels individually. The largest deviation between expected and observed limit for the gluon-fusion production mode at a mass of 160 GeV is driven by the $\tau_h\tau_h$ final state.

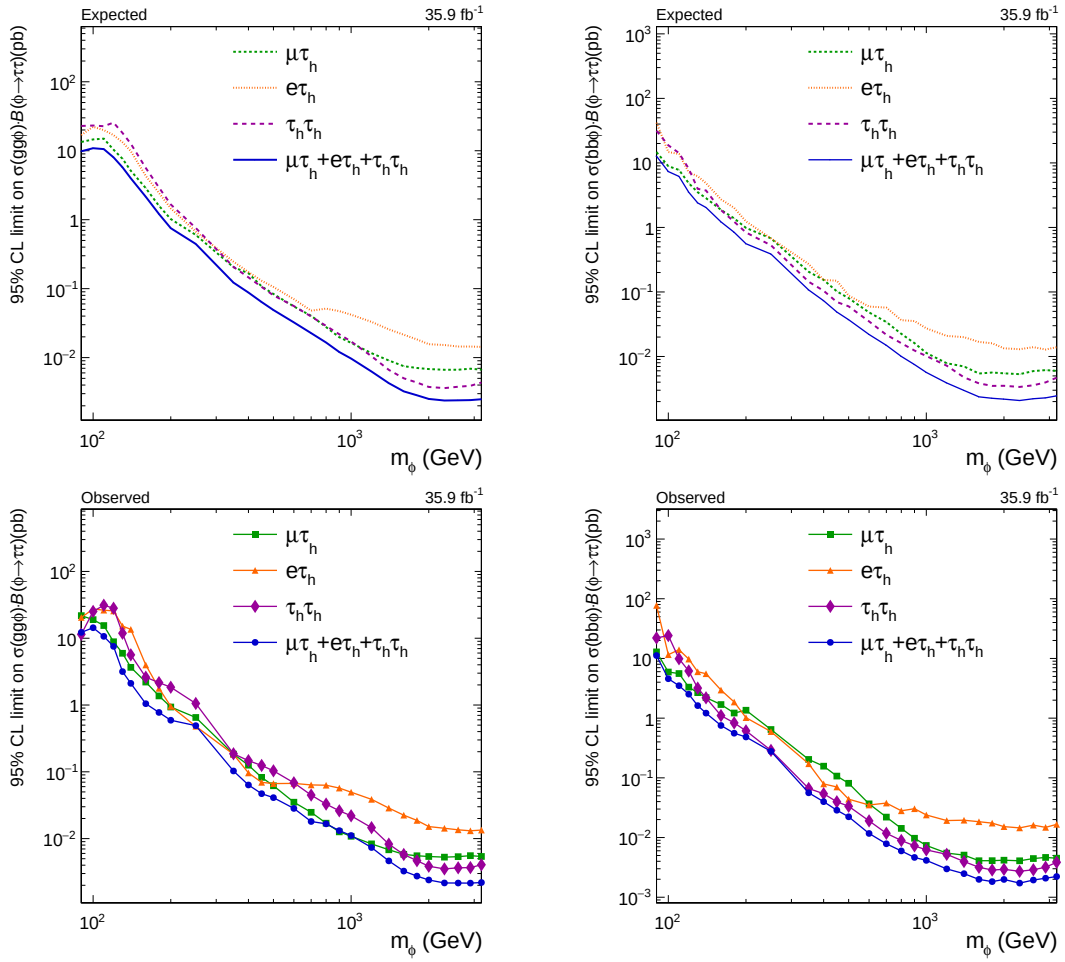


Figure 6.7: Composition of combined expected (upper row) and observed (lower row) limits from the individual di-τ decay channels for the production via gluon-fusion (left) and the b-associated production (right). The sensitivity of the analysis is driven by the $\mu\tau_h$ final state for low mass hypotheses of $\phi = H/A$, and by the $\tau_h\tau_h$ final state for mid- to high-mass hypotheses.

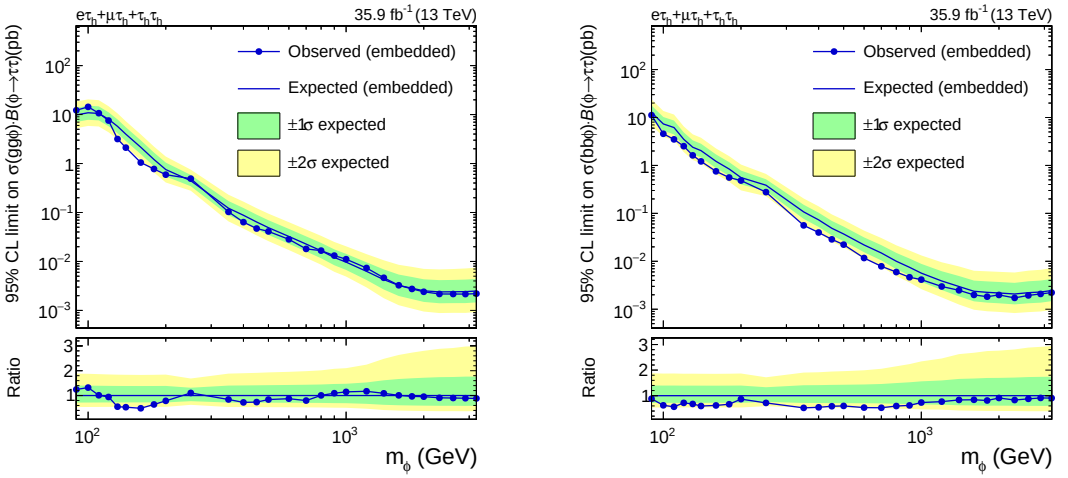


Figure 6.8: Observed 95% CL upper limits derived in this thesis by the use of embedded events as background estimation. Shown are upper limits on the cross-section of a heavy neutral Higgs boson produced via gluon-fusion (left) and b-associated production (right).

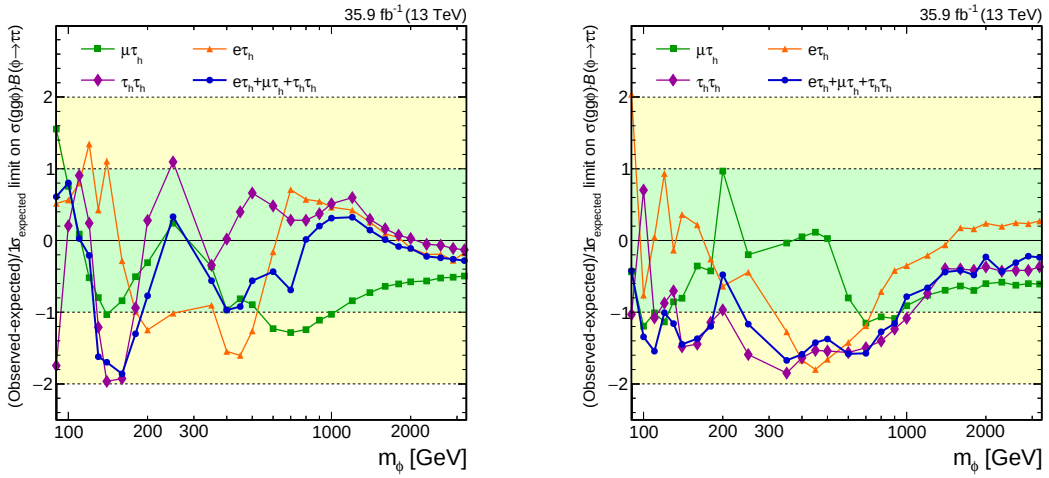


Figure 6.9: Relative differences between observed and expected limits for the individual decay channels as well as the combination of all three decay channels used for the analysis.

6.6 Comparison to the published results

Embedded events served as a cross-check for the estimated $Z \to \tau\tau$ background in the analysis published by CMS [36]. In this section, the results derived by the use of embedded events will be compared to the published analysis, in which the results are derived by the use of simulated $Z \to \tau\tau$ events. The description of jet$\to \tau_h$ fakes, Electroweak and $Z \to \ell\ell$ background is equivalent. All differences in expected and observed limits can be attributed to the different description of $Z \to \tau\tau$ and $t\bar{t} \to \tau\tau$ events and the appropriate uncertainty model.
Figure 6.10 shows the comparison of the expected limits between the published analysis and the limits derived in this thesis for the combination of all final states, whereas the observed limits are shown in figure 6.11. The exclusion sensitivity of the published analysis lies within the 68% confidence interval of this analysis for all mass points and both methods show similar trends between expectation and observation.
The expected limits derived in this thesis show a gain in exclusion sensitivity for all mass points in the gluon-fusion production mode, with a strongest gain for mass hypotheses exceeding $\approx 1000\,\text{GeV}$. In the b-associated production, the sensitivity is comparable for mass points up to $\approx 1000\,\text{GeV}$, and again higher for the heavier masses. For high-masses, the statistical uncertainty and therefore the number of events available for background modeling becomes increasingly important. The improvement in exclusion sensitivity is expected as an effect of the re-binning algorithm changing the number of bins as well as the bin edges of the histograms as a function of $m_{\text{T}}^{\text{tot}}$.

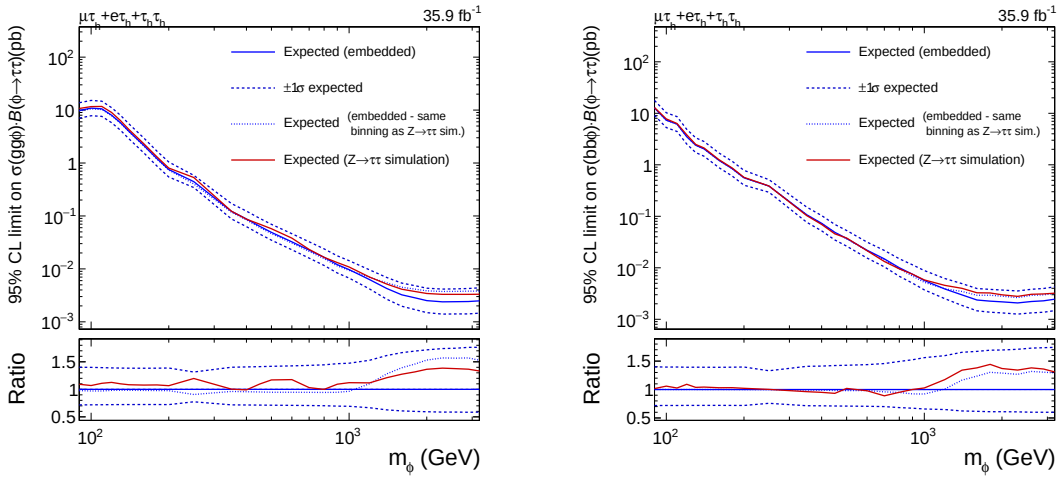


Figure 6.10: Comparison of expected limits between this analysis (embedded) and the published analysis ($Z \to \tau\tau$ simulation) of the combination of all three final states. The exclusion sensitivity of the published analysis lies within the $\pm 1\sigma$-confidence interval of this analysis. The blue dotted line shows the exclusion sensitivity when the same binning of the $m_{\text{T}}^{\text{tot}}$ distributions as used in the published analysis is enforced.

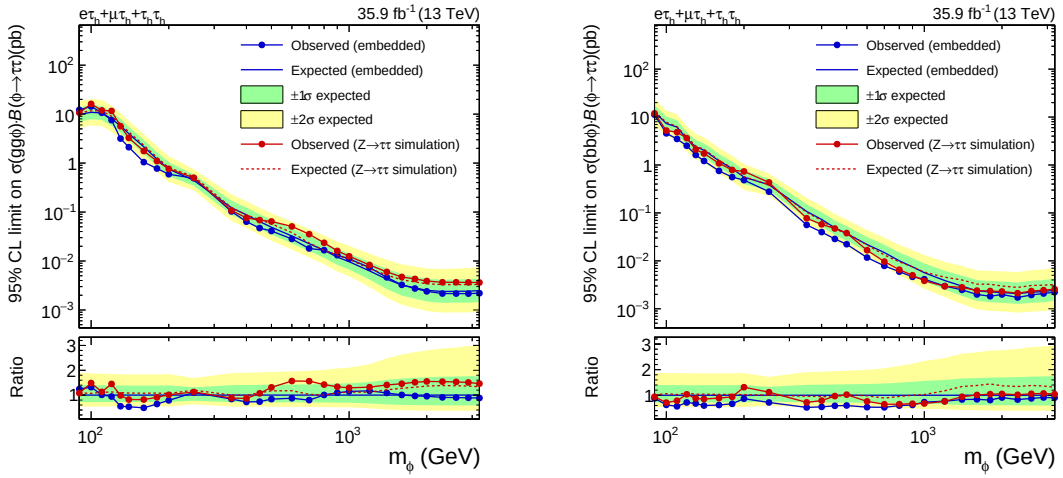
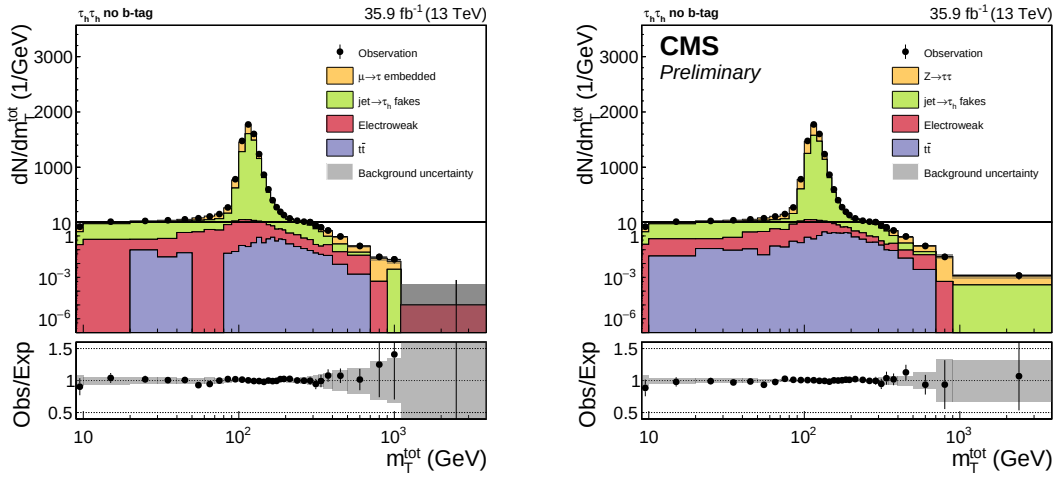


Figure 6.11: Comparison of expected and observed limits between this analysis (embedded) and the published results ($Z \to \tau\tau$ simulation). Agreement between the results derived by using embedded and simulated events is observed for all mass points within the confidence interval of this analysis.

The algorithm starts at the first empty bin and then moves to neighboring bins left and right, which are merged until the bin has a content greater than 0 with a relative bin uncertainty of less than 0.9. The merging is then continued at the next empty bin. The choice to merge with left and/or right bins is done to minimize the number of bins lost by the algorithm. Due to the higher number of events in the embedded event sample, this algorithm can result in different number of bins for the two background estimation methods. The binning has a large impact on the exclusion sensitivity for the high-mass region. The expected exclusion limits of the analysis using embedded events for which the same binning as used in the published analysis is enforced is shown as a blue dotted line. An example for the difference in binning caused by the re-binning algorithm is shown for the `no-btag` category of the $\tau_h\tau_h$ final state in figure 6.12. The algorithm resulted a finer binning for $m_{\mathrm{T}}^{\mathrm{tot}} > 700\,\mathrm{GeV}$. The comparison of the sensitivity of the individual final states is shown in figure 6.13. Again, the sensitivity lies within the confidence interval of this analysis.
Figure 6.14 highlights the differences in trends between observed and expected limits, normalized by the respective confidence interval. The observed 95% CL upper limits follow similar trends around the expectation for most mass points across the three final states.



(a) $\mu \to \tau$ embedded (this analysis) **(b)** $Z \to \tau\tau$ simulation (published analysis)

Figure 6.12: $m_{\mathrm{T}}^{\mathrm{tot}}$ distributions in the $\tau_h\tau_h$ channel for the no b-tag category for a background estimation using embedded (left) and simulated (right) $Z \to \tau\tau$ events as background estimation method. For $m_{\mathrm{T}}^{\mathrm{tot}} > 700\,\mathrm{GeV}$, the re-binning algorithm leads to an additional bin for the histogram using embedded events.

Agreement between the $Z \to \tau\tau$ estimation method used in the published analysis and this analysis is observed. The main goal of this analysis is to provide a cross-check of the published results as well as a proof of the performance of embedded events using Run II-data. Both the uncertainty on the initial muon reconstruction and $t\bar{t}$ contamination are derived as conservative estimates that can be decreased for future analyses, further improving the performance of embedded events. Up to now, the sensitivity derived by the use of embedded events match the one derived by the use of simulated events. Improvements can be seen for the sensitivity in the search for Higgs bosons mainly in the gluon-fusion production mode and in the search for heavy Higgs bosons with a mass exceeding 1000 GeV. This makes embedded events a valid choice for $Z \to \tau\tau$ background description for future analyses.

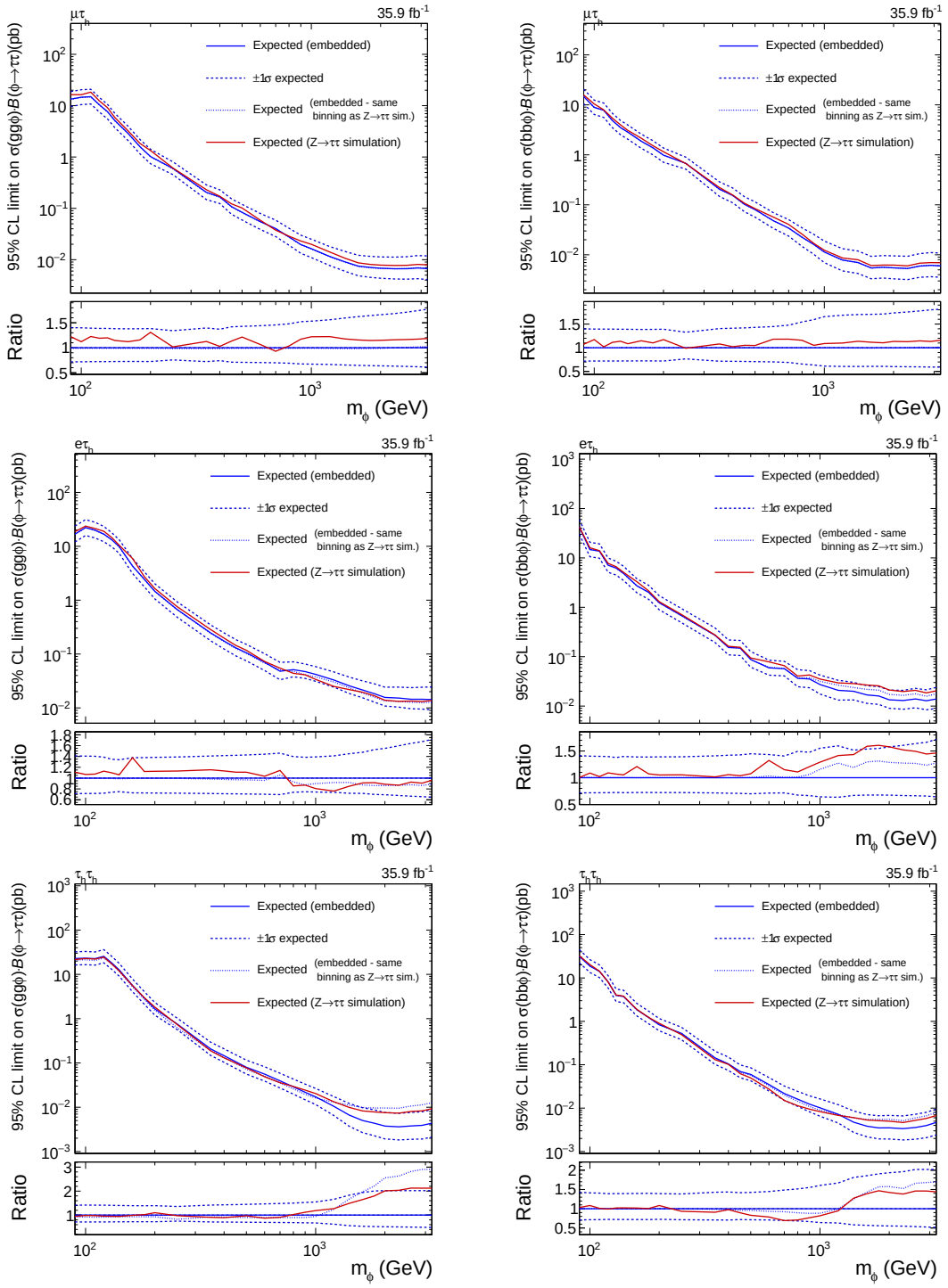


Figure 6.13: Comparison of expected limits between this analysis (embedded) and the published analysis ($Z \to \tau\tau$ simulation) for the $\mu\tau_h$, $e\tau_h$ and $\tau_h\tau_h$ final states for a production via gluon fusion (left) and b-associated production (right).

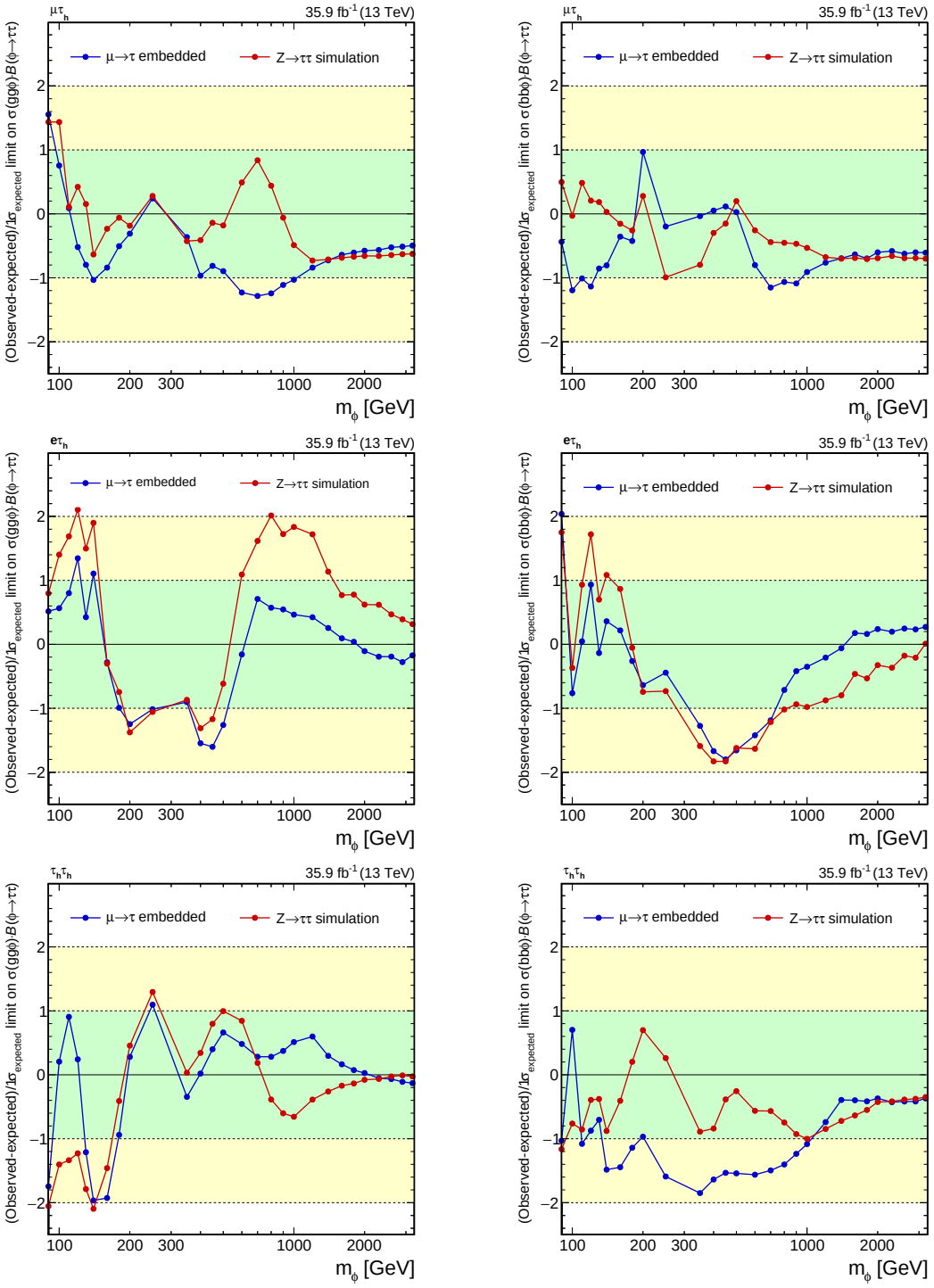


Figure 6.14: Comparison of deviations between expected and observed limits for this analysis and the analysis published by CMS for the $\mu\tau_h$, $e\tau_h$ and $\tau_h\tau_h$ final states for a production via gluon fusion (left) and b-associated production (right).

CHAPTER 7

Conclusion and outlook

Embedded events provide a valuable alternative to simulation for description of the expected background from $Z \to \tau\tau$ decays. The advantages of the method in the description of complicated event characteristics of proton-proton collisions at the LHC, such as of the underlying event and pileup, became evident in the good agreement of embedded events as $Z \to \tau\tau$ estimate of data in the analysis presented. Many corrections that are needed for simulated events become obsolete with the use of embedded events.
One additional advantage of having an alternative method for future analyses is the chance to use the inherent differences of simulated and embedded events when examining possible mis-descriptions in background methods. As a hybrid of both data and simulation, embedded events can be used to disentangle several physics effects occurring in the simulation of collisions at the LHC. In this regard, they have been used for a study on mis-descriptions of τ_h identification in simulated events.
The embedding method is being further developed. In this thesis starting points for improvements have been identified as

1. the re-evaluation of event cleaning of the muon track and its calorimetry entries. As shown in the p_T-flow distributions in section 5.1.1, the cleaning can leave remnants in the vicinity of the muon track that impact the description of muon isolation. The cleaning can be tuned on a simulated $Z \to \mu\mu$ sample to find an optimization for the amount of entries to be cleaned to mitigate such effects. The effect on the muon isolation can then be quantified to check for the necessity of a dedicated uncertainty.

2. a further study on the identification and isolation of electrons in embedded events. In section 4.1, embedded electrons have been shown to be more isolated than electrons in simulation or data, and scale factors have been derived which correct this effect. A starting point for this study is a study of the separate ID input variables to the MVA discriminator used for electron identification.

3. a normalization of embedded events from first principles, starting from the number of selected $Z \to \mu\mu$ events. The number of expected $Z \to \tau\tau$ events can than be obtained by extrapolating the number of selected events onto the full phase space, taking into account the number of events lost due to the selection criteria.

4. a re-evaluation of the uncertainties applied on both $t\bar{t}$ contamination and the initial muon reconstruction. Both uncertainties are used as conservative estimates and can easily be decreased. The correction of $t\bar{t}$ contamination especially is well-founded in the Standard Model and no problems were apparent due to the applied correction.

In the resumption of Run II data-taking by the CMS detector in June 2017, a pragmatic advantage of the embedding method over simulation became evident, as embedded events are produced with comparably small computational effort and are robust against the changing requirements in detector conditions between individual data-taking periods of the CMS detector. This simplifies the analysis of events and decreases both time and analytical effort to obtain first background estimates for new data. Figure 7.1 shows the visible di-τ mass in the $\mu\tau_h$ final state of events selected from the first stable beam data taken in 2017. The remaining backgrounds, such as $Z \to \ell\ell$, $t\bar{t}$ or electroweak decays, are estimated using simulated events that were produced for the the detector conditions at CMS during 2016 data-taking. Due to the use of embedded events, a good agreement between the background estimate and the current data can already be achieved.

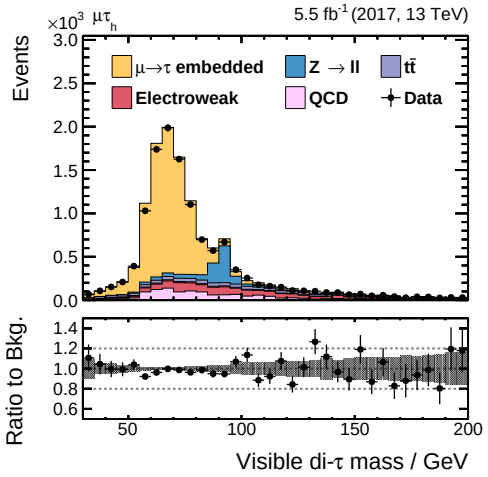


Figure 7.1: The embedding method applied on the first $5.5\,\mathrm{fb}^{-1}$ of the growing Run II-dataset of data taken with the CMS experiment in 2017. The agreement of the background estimate and the data is provided without the use for additional corrections taking into account the changed detector conditions.

With an increasing instantaneous luminosity in the coming years at LHC, the value of the data driven embedding method will rise due to its inherent advantages when describing these complicated hadronic effects. The embedding method has the potential to improve our understanding of $Z \to \tau\tau$ decays and represents a contending candidate for the main

method to be used for the estimation of this background for future analysis in the di-τ channel. The advantages include

1. a lower computational effort compared to simulated event generation.

2. an increase in statistics over simulated events. In the analysis presented in this thesis, the embedded event sample included around five times more $Z \to \tau\tau$ events than the simulated $Z \to \tau\tau$ event sample.

3. a small need for corrections. For simulated events, corrections on the jet energy scale, b-tagging efficiency and recoil of the simulated Z boson need to be applied. Furthermore, a reweighting of the simulated events is necessary to correctly describe the conditions during the proton-proton collisions at the LHC. These corrections become obsolete when using embedded events.

4. an earlier availability of embedded events given the lower computational effort and small need for corrections.

New effects introduced by the updated method are now taken into account and dedicated corrections are supplied, resulting in an overall lower need for corrections when compared to simulated $Z \to \tau\tau$ events.
The successful validation of the $H/A \to \tau\tau$ search using CMS data shows the reliability of embedded events for publication of substantiated physics results. Improvements were already seen in exclusion limits for additional neutral Higgs bosons, which makes the embedding method, at the very least, a mandatory cross-check for estimating the background caused by decays of the Z boson for future analyses of the promising di-τ final state.

CHAPTER 8

Bibliography

[1] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". *Physics Letters B* 716.1 (2012), pp. 1–29. ISSN: 0370-2693.
DOI: https://doi.org/10.1016/j.physletb.2012.08.020.

[2] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". *Physics Letters B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693.
DOI: https://doi.org/10.1016/j.physletb.2012.08.021.

[3] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". *Physical Review Letters* 13 (Aug. 1964), pp. 321–323.
DOI: 10.1103/PhysRevLett.13.321.

[4] P. W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". *Physical Review Letters* 13 (Oct. 1964), pp. 508–509.
DOI: 10.1103/PhysRevLett.13.508.

[5] G. S. Guralnik, C. R. Hagen, and T. W. Kibble. "Global Conservation Laws and Massless Particles". *Physical Review Letters* 13 (Nov. 1964), pp. 585–587.
DOI: 10.1103/PhysRevLett.13.585.

[6] S.L. Glashow. "The renormalizability of vector meson interactions". *Nuclear Physics* 10.Supplement C (1959), pp. 107–117. ISSN: 0029-5582.
DOI: https://doi.org/10.1016/0029-5582(59)90196-8.

[7] S. Weinberg. "A Model of Leptons". *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266.
DOI: 10.1103/PhysRevLett.19.1264.

[8] Abdus Salam and J. C. Ward. "Weak and electromagnetic interactions". *Il Nuovo Cimento (1955-1965)* 11.4 (Feb. 1959), pp. 568–577. ISSN: 1827-6121.
DOI: 10.1007/BF02726525.

[9] UA1 Collaboration. "Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $s^{1/2} = 540$ GeV". *Phys. Lett.* 122B (1983). [,611(1983)], pp. 103–116.
DOI: 10.1016/0370-2693(83)91177-2.

[10] UA2 Collaboration. "Evidence for $Z0 \to e^+e^-$ at the CERN anti-p p Collider". *Phys. Lett.* 129B (1983), pp. 130–140.
DOI: 10.1016/0370-2693(83)90744-X.

[11] CMS Collaboration. "Measurements of inclusive W and Z cross sections in pp collisions at $\sqrt{s} = 7$ TeV". *Journal of High Energy Physics* 2011.1 (Jan. 2011), p. 80. ISSN: 1029-8479.
DOI: 10.1007/JHEP01(2011)080.

[12] Jonathan M. Butterworth, Guenther Dissertori, and Gavin P. Salam. "Hard Processes in Proton-Proton Collisions at the Large Hadron Collider". *Ann. Rev. Nucl. Part. Sci.* 62 (2012), pp. 387–405.
DOI: 10.1146/annurev-nucl-102711-094913. arXiv: 1202.0583 [hep-ex].

[13] CMS Collaboration. "Measurement of the Drell-Yan Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV". *JHEP* 10.arXiv:1108.0566. CMS-EWK-10-007. CERN-PH-EP-2011-093 (Aug. 2011), 007. 41 p.

[14] Roger Wolf. "The Higgs Boson Discovery at the Large Hadron Collider". Springer Tracts of Modern Physics Volume 264. 2015.
DOI: 10.1007/978-3-319-18512-5.

[15] C. Patrignani et al. "2016 Review of Particle Physics". *Chin. Phys. C, 40, 100001* (2016). Particle Data Group.

[16] H1 and ZEUS collaborations. "Combined measurement and QCD analysis of the inclusive e ± p scattering cross sections at HERA". *Journal of High Energy Physics* 2010.1 (Jan. 2010), p. 109. ISSN: 1029-8479.
DOI: 10.1007/JHEP01(2010)109.

[17] The CMS Collaboration. "The CMS experiment at the CERN LHC". *Journal of Instrumentation* 3.08 (2008), S08004.

[18] Lyndon Evans and Philip Bryant. "LHC Machine". *Journal of Instrumentation* 3.08 (2008), S08001.

[19] CMS Collaboration. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. 2017.

[20] CMS Collaboration. https://cms.cern/detector. 2017.

[21] Lucas Taylor. http://cms.web.cern.ch/news/detector-overview. 2011.

[22] Moon Meok Kim et al. "Web-based monitoring tools for Resistive Plate Chambers in the CMS experiment at CERN". 9 (Oct. 2014).

[23] CMS Collaboration. "Particle-flow reconstruction and global event description with the CMS detector". Tech. rep. CMS-PRF-14-001. CMS-PRF-14-001-004. Submitted to JINST. All figures and tables can be found at http://cms-results.web.cern.ch/cms-results/public-results/publications/PRF-14-001. Geneva: CERN, June 2017.

[24] Elizabeth Quigg et al. David Barney. *Interactive Slice of the CMS detector*. https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ShowDocument?docid=4172 accessed 2017/09/26.

[25] Gavin P. Salam. "Towards Jetography". *Eur. Phys. J.* C67 (2010), pp. 637–686. DOI: 10.1140/epjc/s10052-010-1314-6. arXiv: 0906.1833 [hep-ph].

[26] CMS Collaboration. "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV". *Journal of Instrumentation* 7 (2012). DOI: https://doi.org/10.1088/1748-0221/7/10/P10002.

[27] The CMS collaboration. "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV". *Journal of Instrumentation* 10.06 (2015), P06005.

[28] Romeo, Francesco. "Reconstruction and identification of tau leptons in CMS". *EPJ Web Conf.* 126 (2016), p. 04041. DOI: 10.1051/epjconf/201612604041.

[29] The CMS collaboration. "Reconstruction and identification of τ lepton decays to hadrons and $\nu\tau$ at CMS". *Journal of Instrumentation* 11.01 (2016), P01019.

[30] S. Chatrchyan et al. "Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons". *Journal of High Energy Physics* 2014.5 (May 2014), p. 104. ISSN: 1029-8479. DOI: 10.1007/JHEP05(2014)104.

[31] A. Akhmetshin. "Embedding - a Data Driven Method to Estimate the $Z \to \tau\tau$ Background in the $H \to \tau\tau$ Searches". https://ekp-invenio.physik.uni-karlsruhe.de/record/48866?ln=en. 2016.

[32] https://twiki.cern.ch/twiki/bin/view/CMSPublic/TagAndProbe.

[33] N. Adam et al. *Generic Tag and Probe Tool for Measuring Efficiency at CMS with Early Data*. CMS AN-2009/111. 2009.

[34] J. Beringer et al. PR D86, 010001, http://pdg.lbl.gov. 2012.

[35] C. Patrignani et al. (Particle Data Group). "2016 Review of Particle Physics". *Chin. Phys. C* 40, 100001 (June 2016).

[36] CMS Collaboration. "Search for additional neutral Higgs bosons in the di-τ final state in *pp* collisions" (2017). CMS-HIG-17-020, to be published.

[37] M. Spanring J. Brandstetter M. Flechl. "Data-driven background estimation of fake-tau backgrounds in MSSM Higgs boson decays to tau leptons with 2016 data" (2017). CMS AN-2017/104 (private communiation).

[38] A. Akhmetshin et al. “Search for MSSM H/A to tautau using the full 2016 dataset” (2017). CMS AN-2017/071 (private communiation).

[39] CMS Collaboration. “Identification of b-quark jets with the CMS experiment”. *JINST* 8 (2013), P04013.
DOI: 10.1088/1748-0221/8/04/P04013. arXiv: 1211.4462 [hep-ex].

APPENDIX A

Appendix

A.1 Generator weights of embedded events for all final states

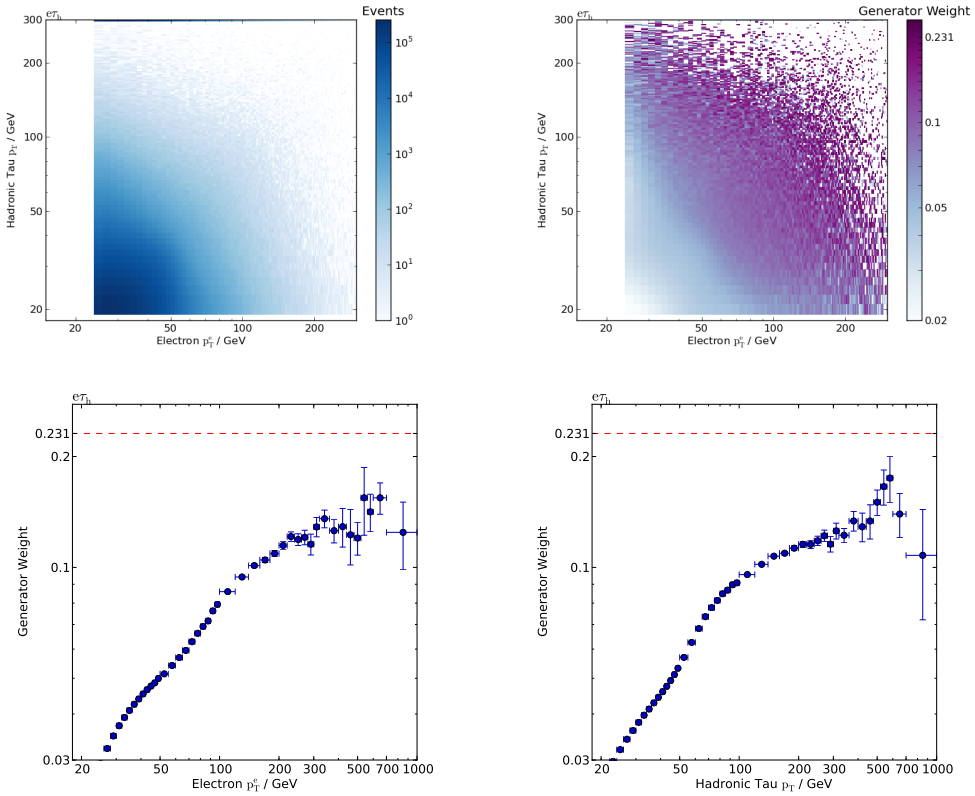


Figure A.1: Number of embedded events in the $e\tau_h$ final state as a function of the two transverse momenta of the τ decay products and the respective averages of the generator weights reflecting the efficiency during the simulation step of the embedding method. The average generator weight as a function of the individual transverse momenta is shown in the lower row.

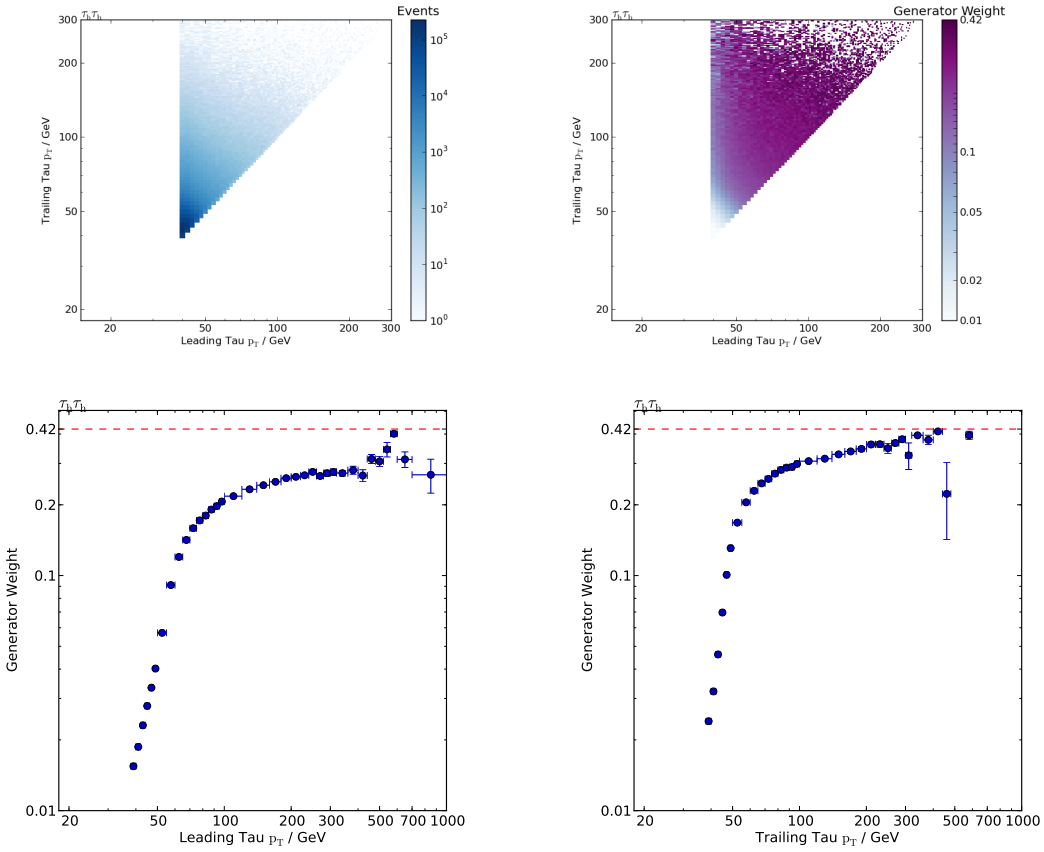


Figure A.2: Number of embedded events in the $\tau_h\tau_h$ final state as a function of the two transverse momenta of the τ decay products and the respective averages of the generator weights reflecting the efficiency during the simulation step of the embedding method. The average generator weight as a function of the individual transverse momenta is shown in the lower row.

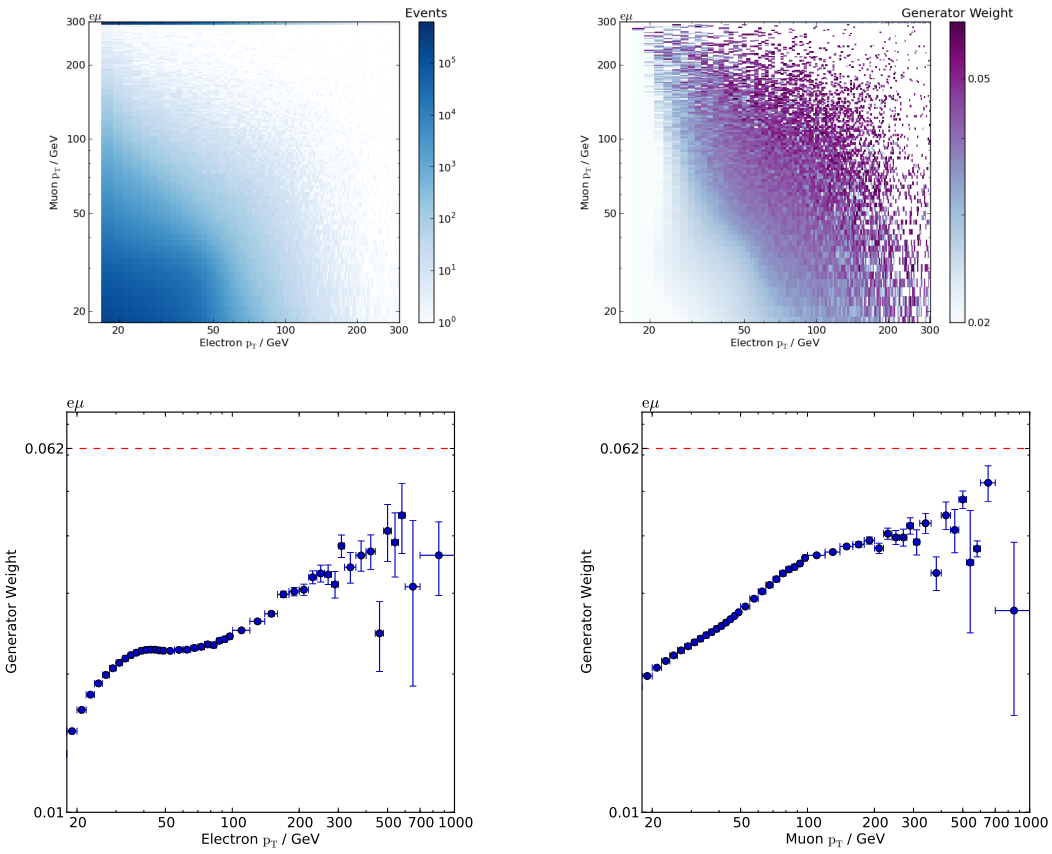


Figure A.3: Number of embedded events in the $e\mu$ final state as a function of the two transverse momenta of the τ decay products and the respective averages of the generator weights reflecting the efficiency during the simulation step of the embedding method. The average generator weight as a function of the individual transverse momenta is shown in the lower row.

A.2 Tag and probe efficiencies

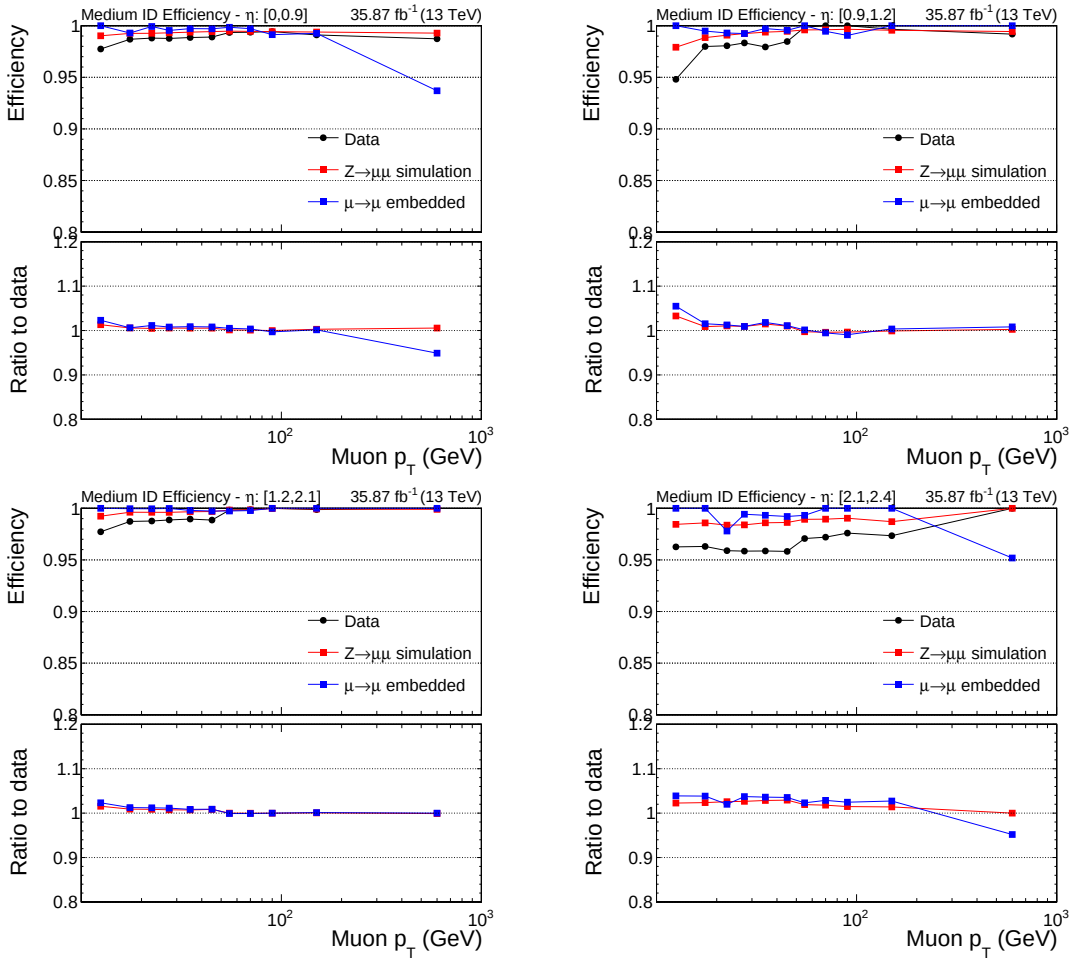


Figure A.4: Muon identification efficiencies for all η-bins.

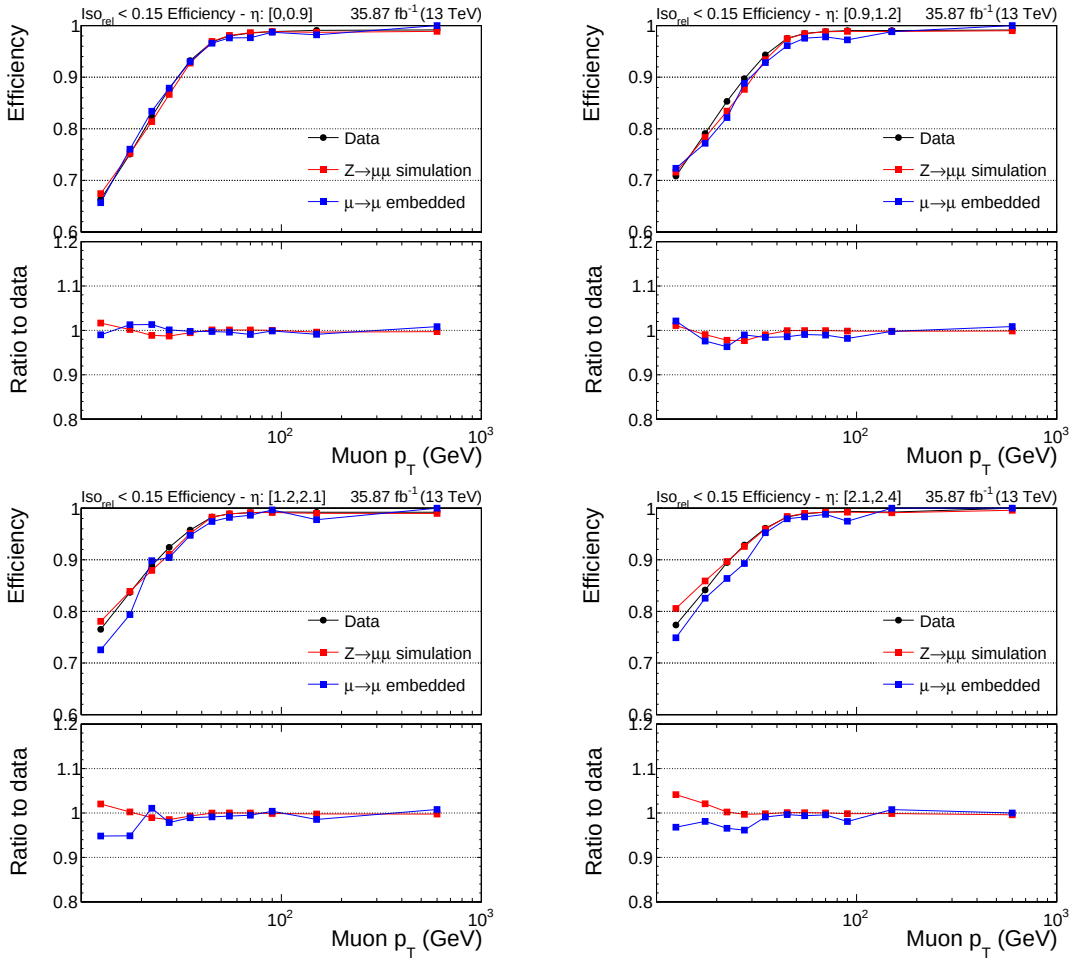


Figure A.5: Muon isolation efficiencies for all η-bins.

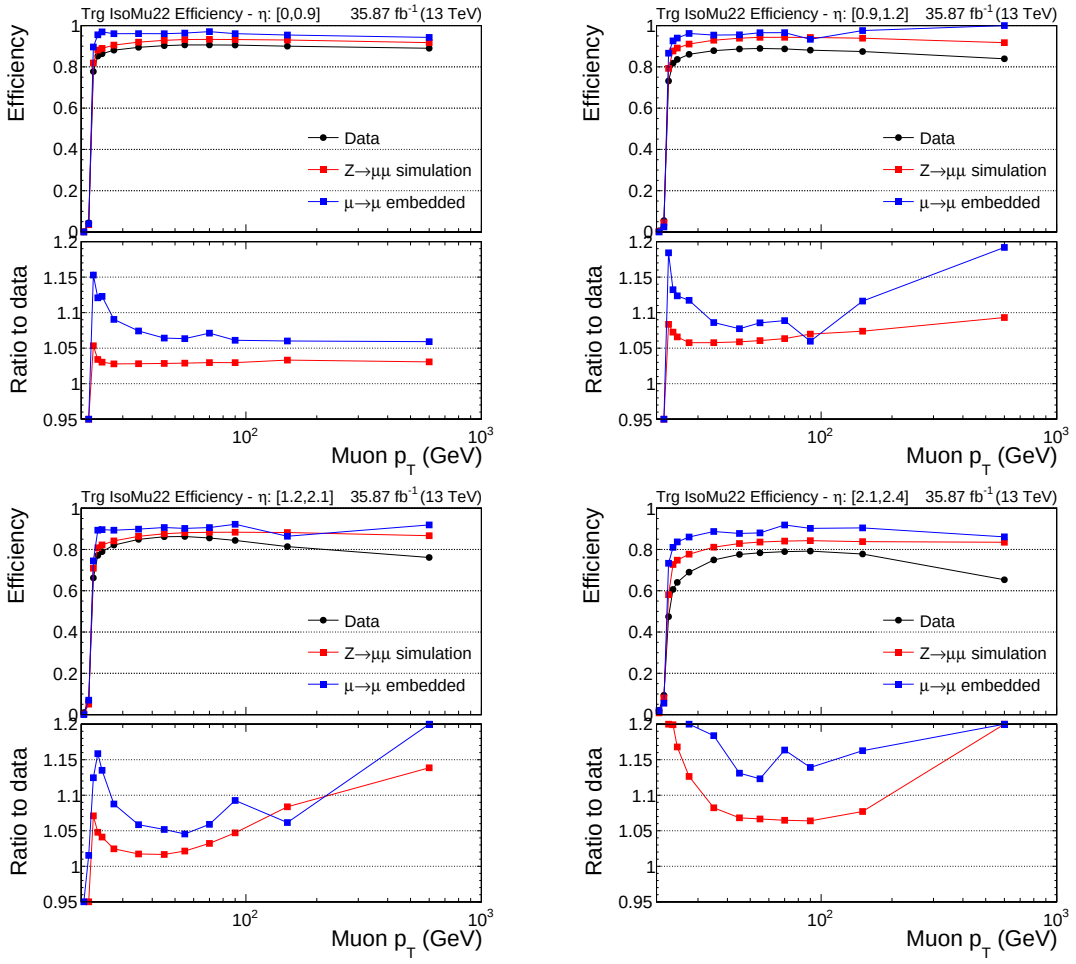


Figure A.6: Muon trigger efficiencies for all η-bins.

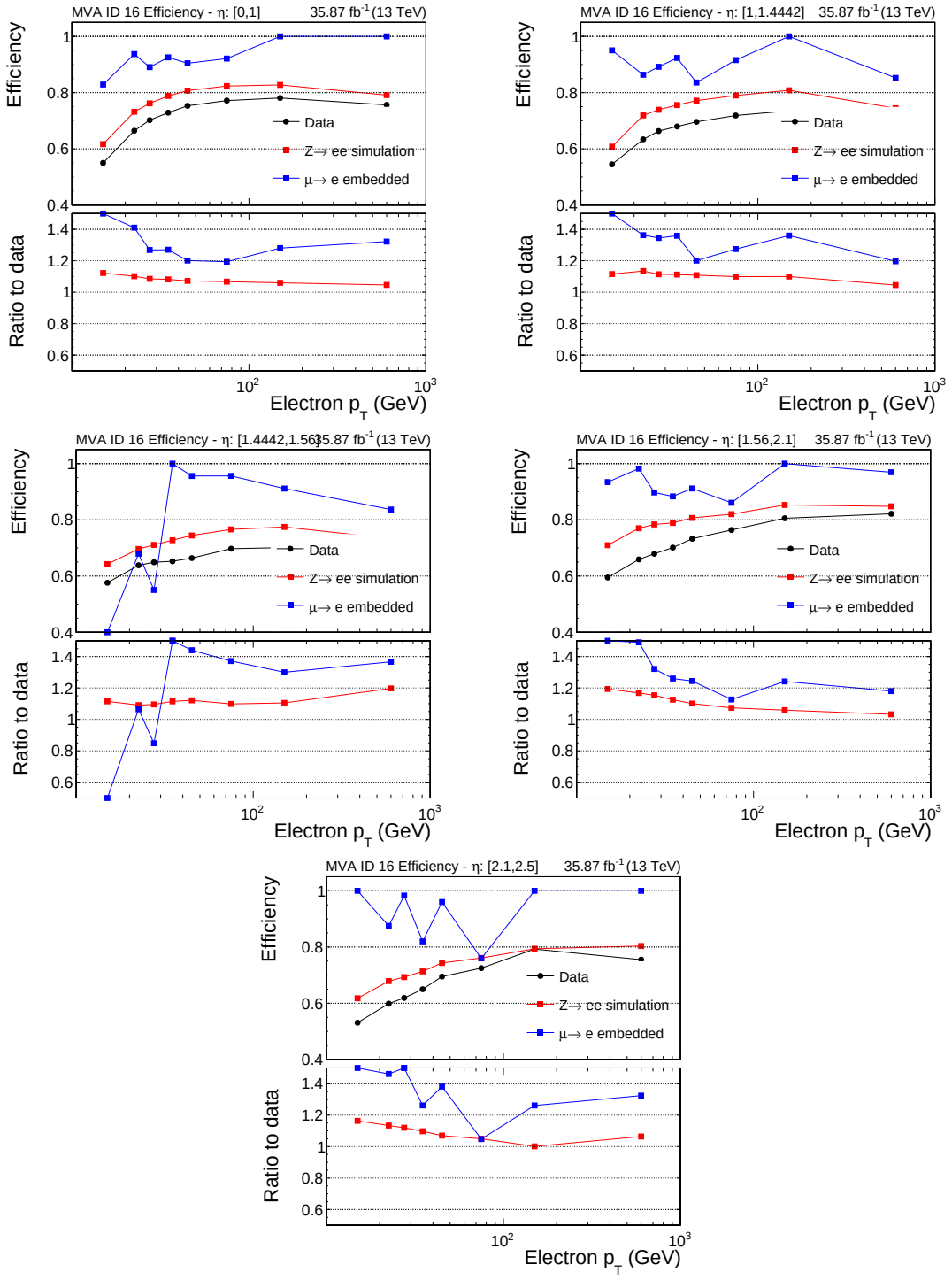


Figure A.7: Electron identification efficiencies for all η-bins.

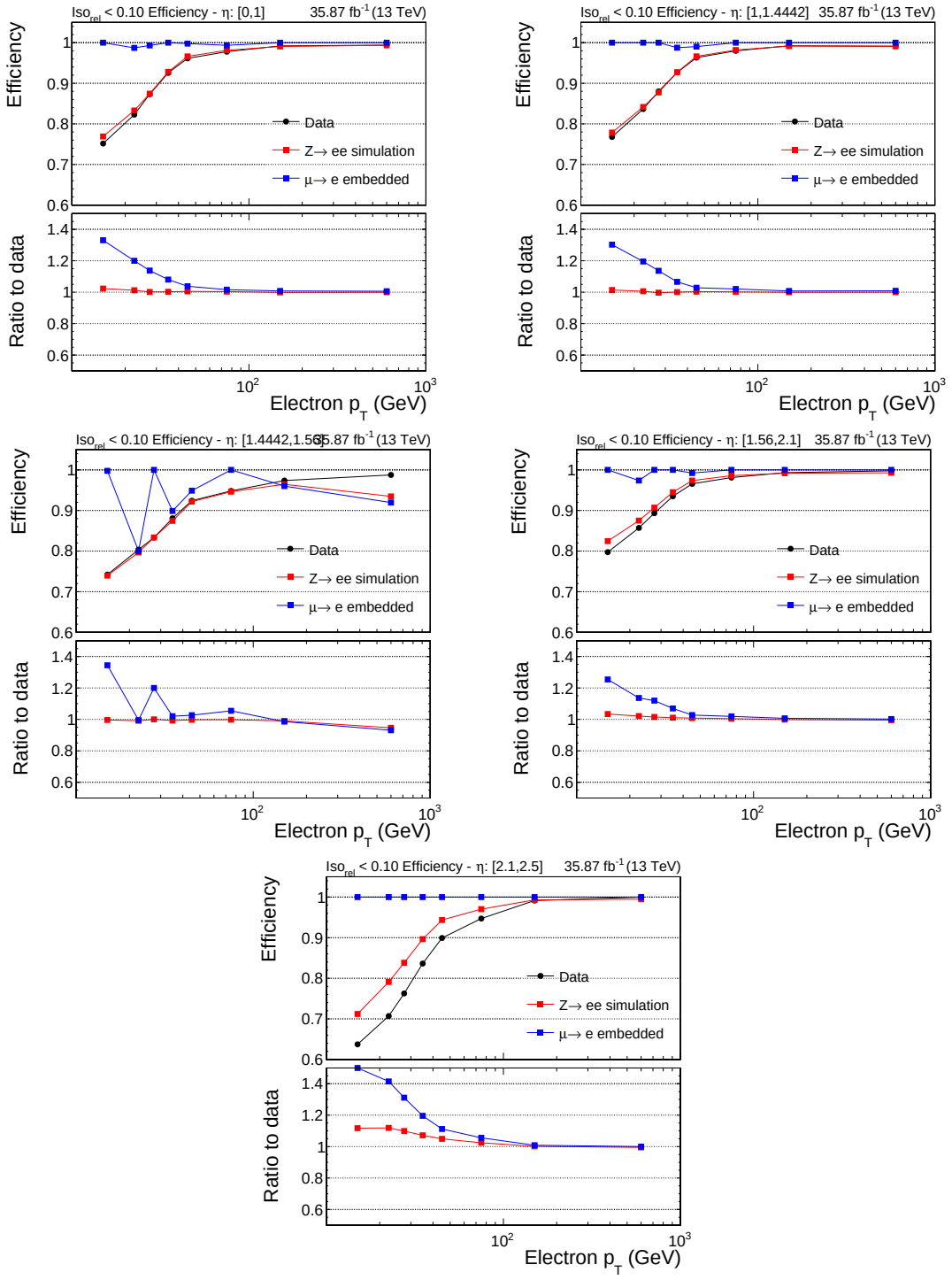


Figure A.8: Electron isolation efficiencies for all η-bins.

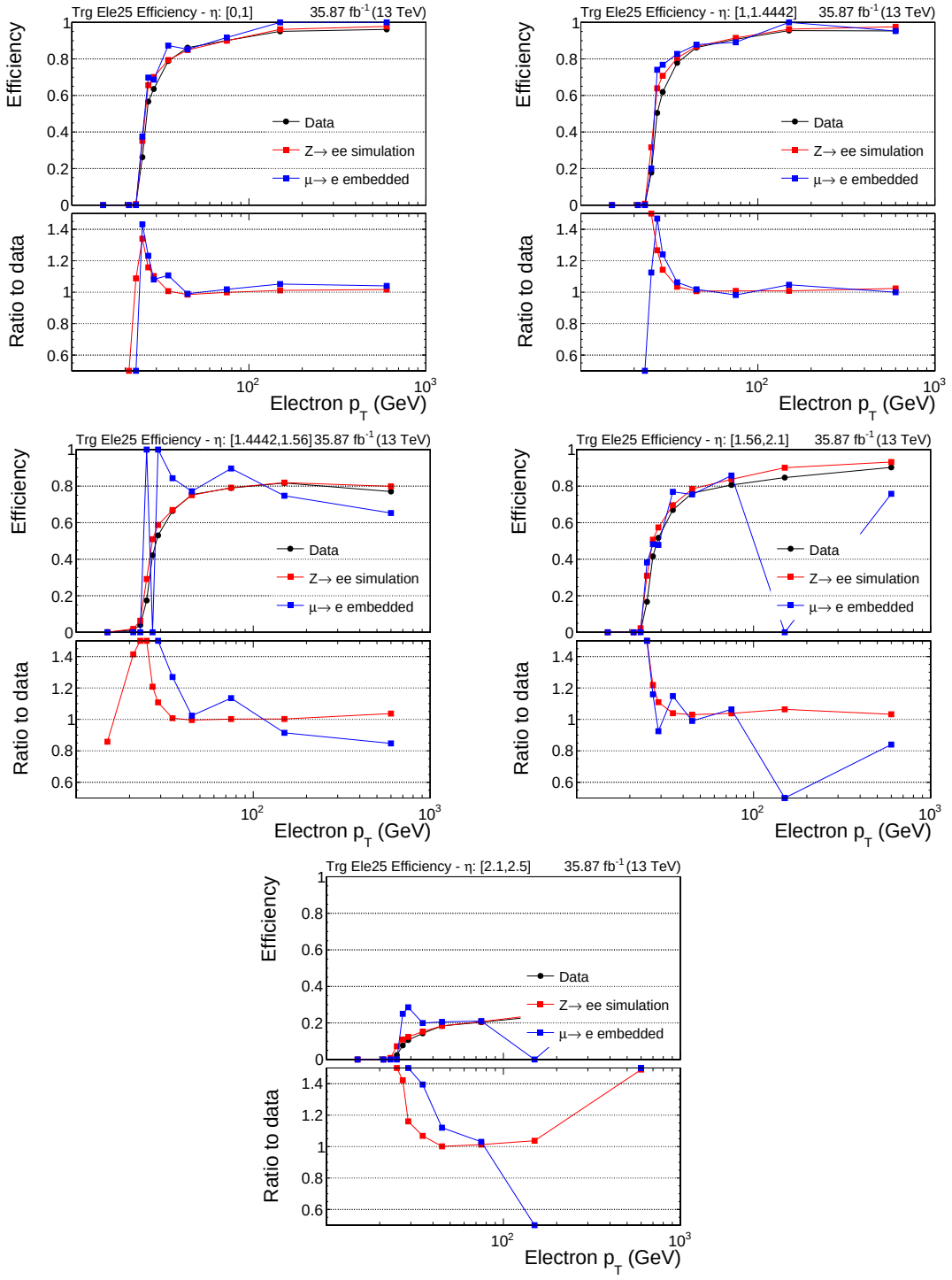
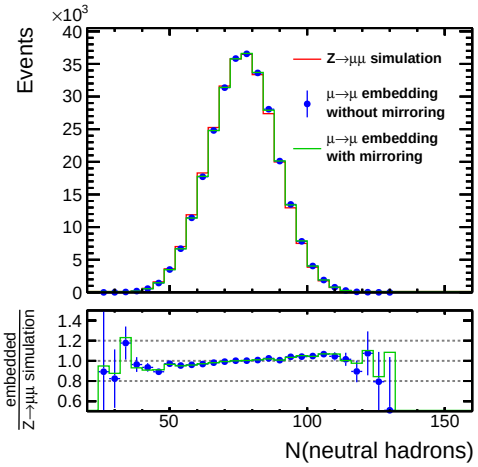


Figure A.9: Electron trigger efficiencies for all η-bins.

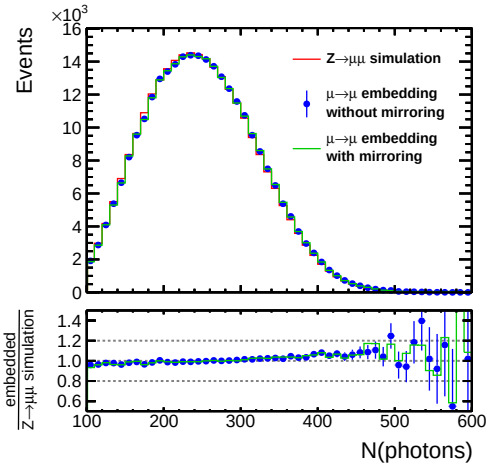
A.3 Additional dataset information and figures for the Monte Carlo validation study

Table A.1: $Z \to \ell\ell$ Monte Carlo sample used for the validation study.

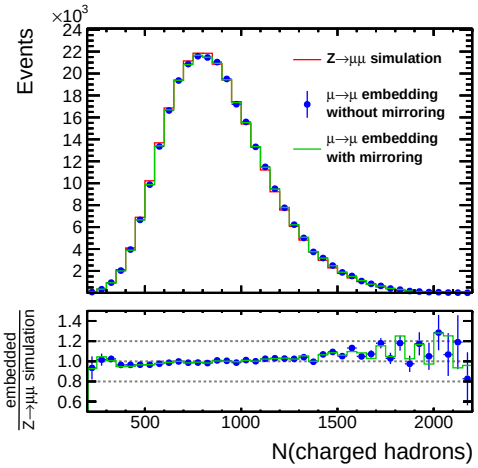
Dataset name	`/DYToLLMCRunIISummer16DR80/AllFinalState-imputFlatPU28to62HcalNZSRAWAODSIM_madgraph_miniAOD-v5/USER`
Number of events	19,963,203
Energy	13 TeV
Global Tag	`80X_mcRun2_asymptotic_2016_TrancheIV_v6`



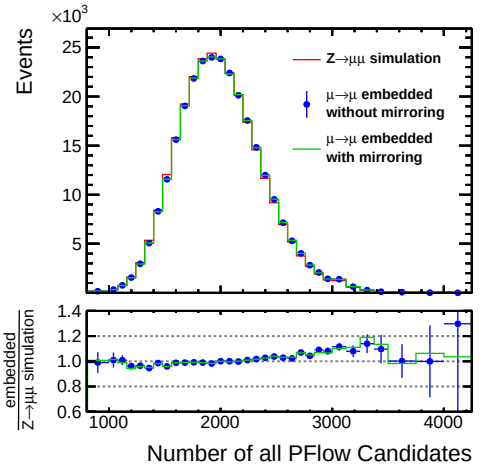
(a) Neutral hadrons



(b) Photons



(c) Charged hadrons



(d) All PF candidates

Figure A.10: Number of particle flow candidates in simulated $Z \to \mu\mu$ events before and after applying embedding. The comparison was made both with and without the application of mirroring.

A.4 m_T^{tot} distributions

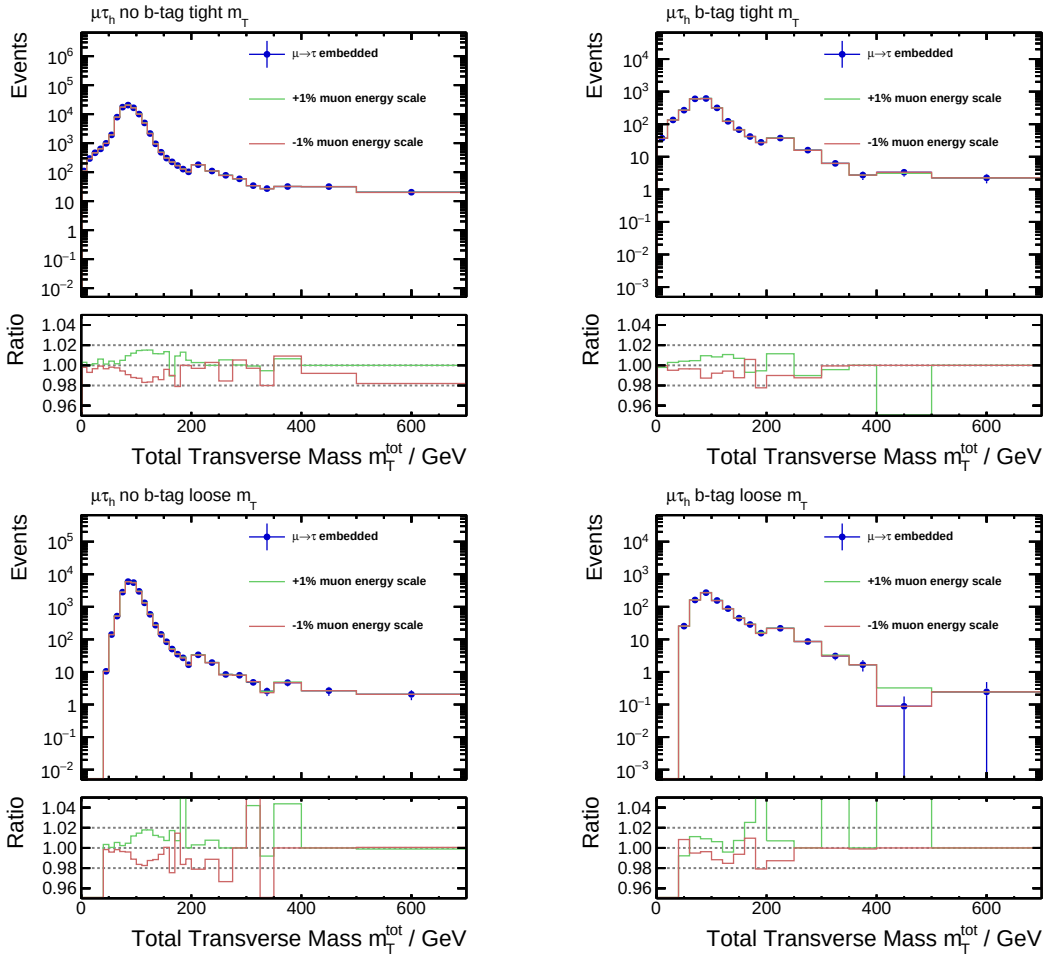


Figure A.11: Uncertainty bands of the introduced muon energy scale uncertainty.

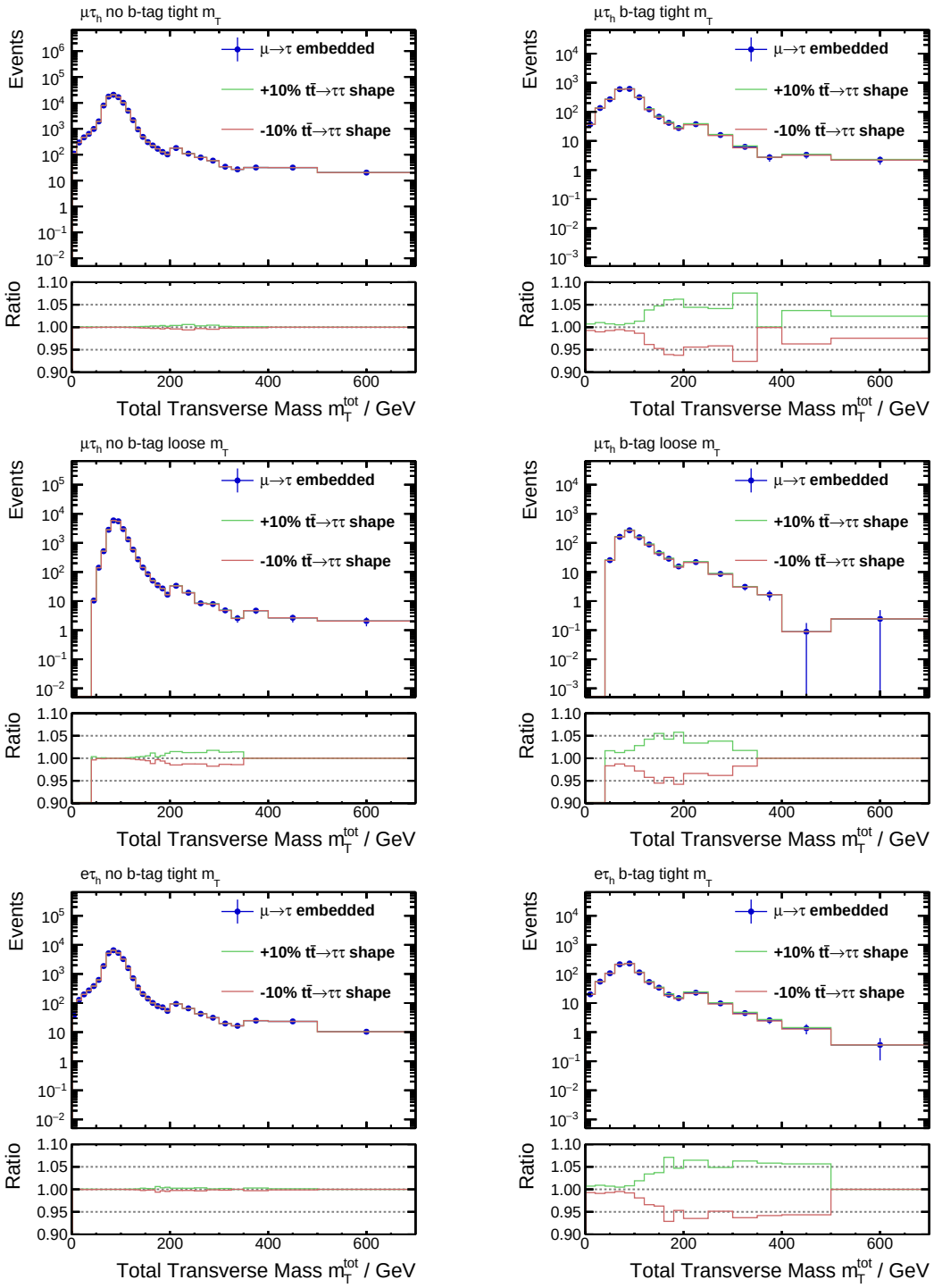


Figure A.12: Uncertainty bands of the introduced $t\bar{t}$ contamination uncertainty.

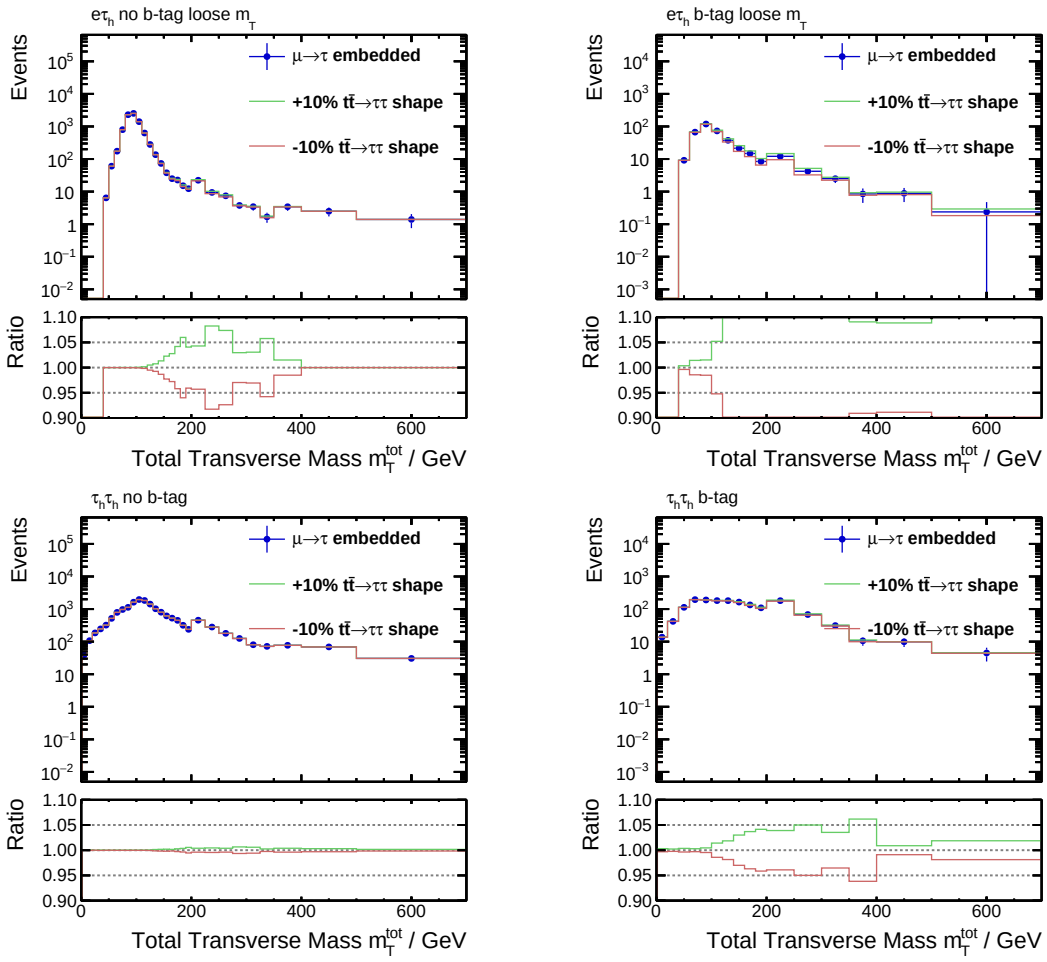


Figure A.13: Uncertainty bands of the introduced $t\bar{t}$ contamination uncertainty.

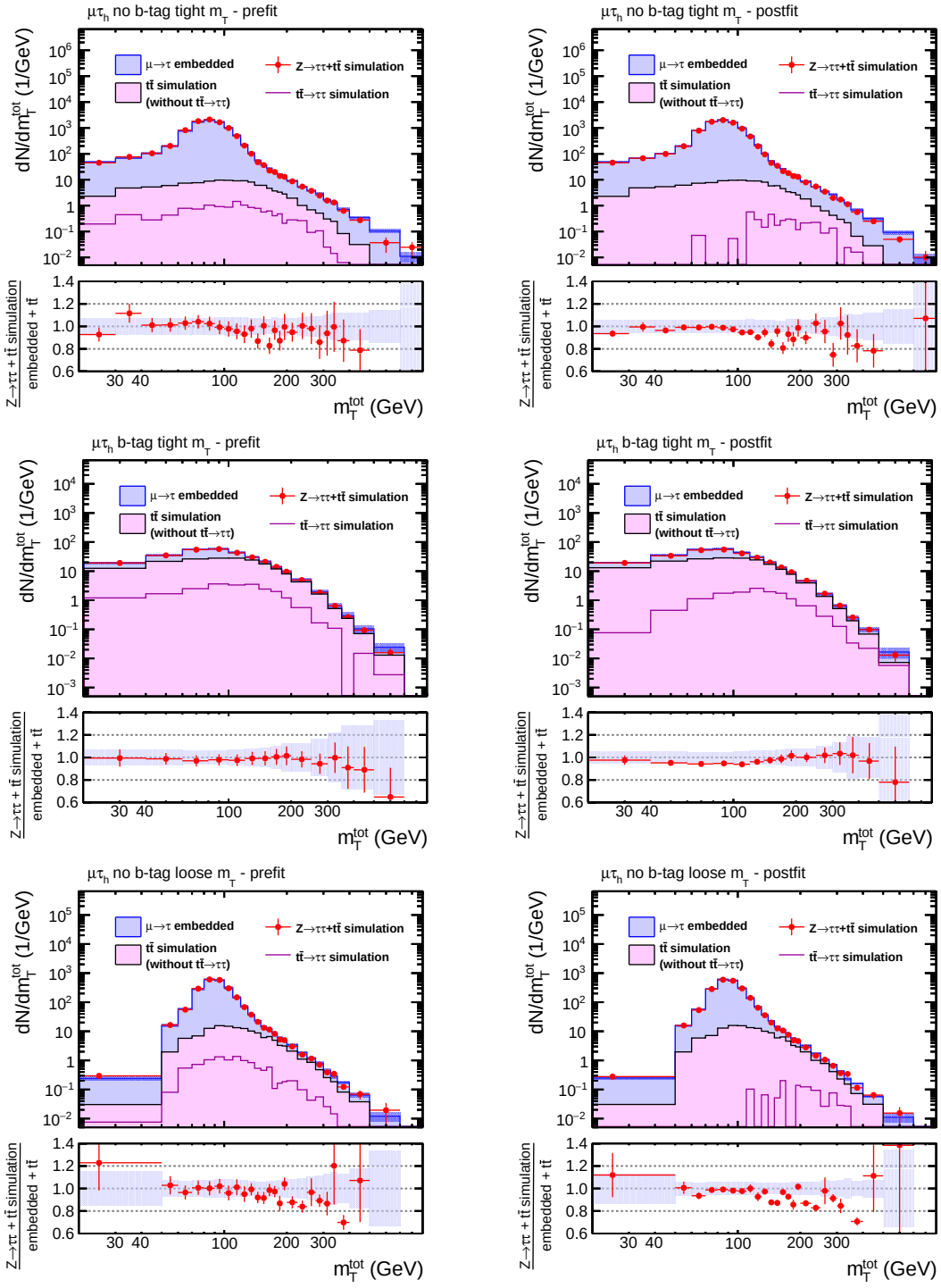


Figure A.14: Comparison of $m_{\mathrm{T}}^{\mathrm{tot}}$ distributions between the respective background estimates with the combined statistical and systematic uncertainties of $\mu \to \tau$ embedded events and simulated $Z \to \tau\tau$ events.

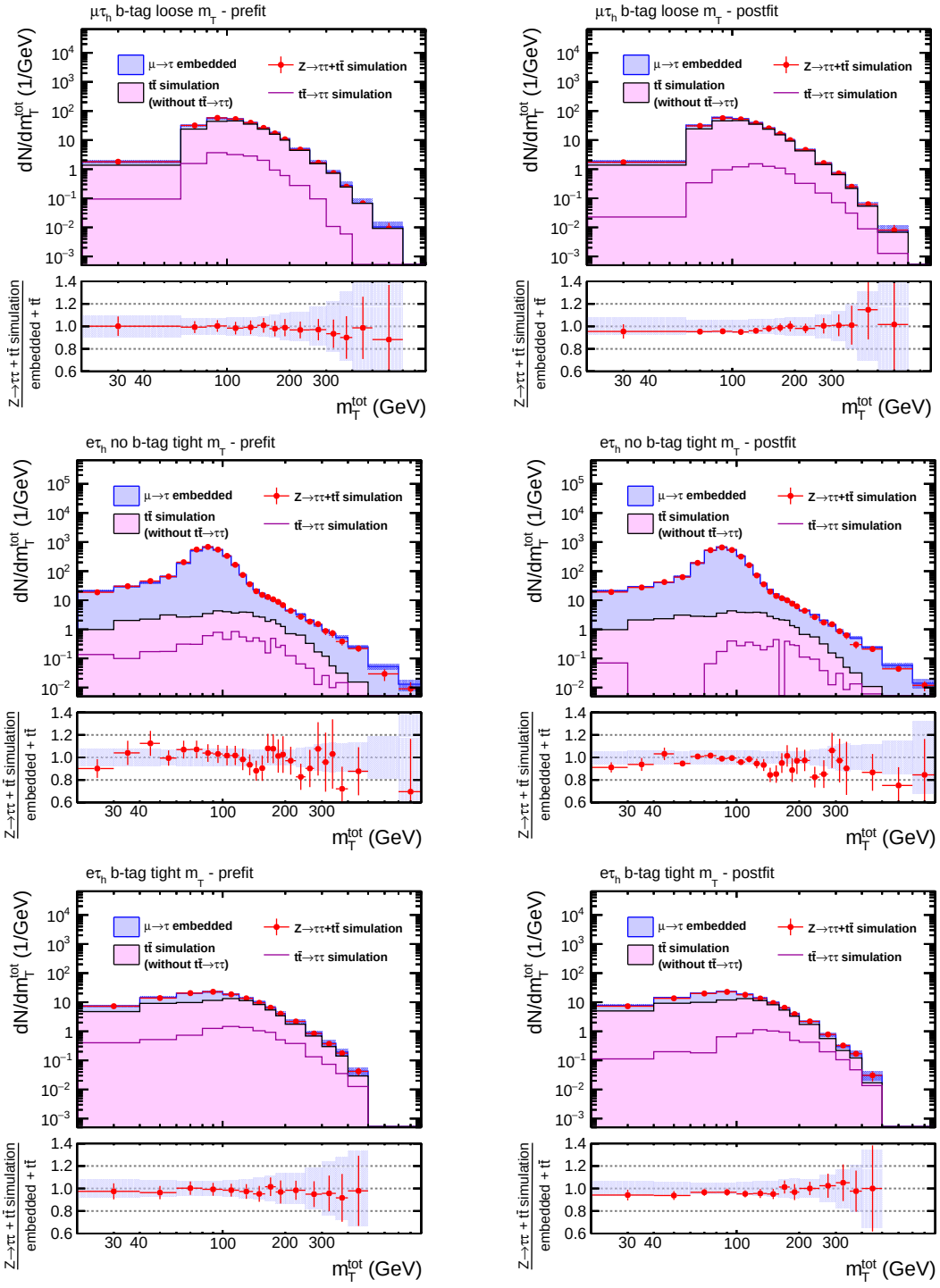


Figure A.15: Comparison of $m_{\mathrm{T}}^{\mathrm{tot}}$ distributions between the respective background estimates with the combined statistical and systematic uncertainties of $\mu \to \tau$ embedded events and simulated $Z \to \tau\tau$ events.

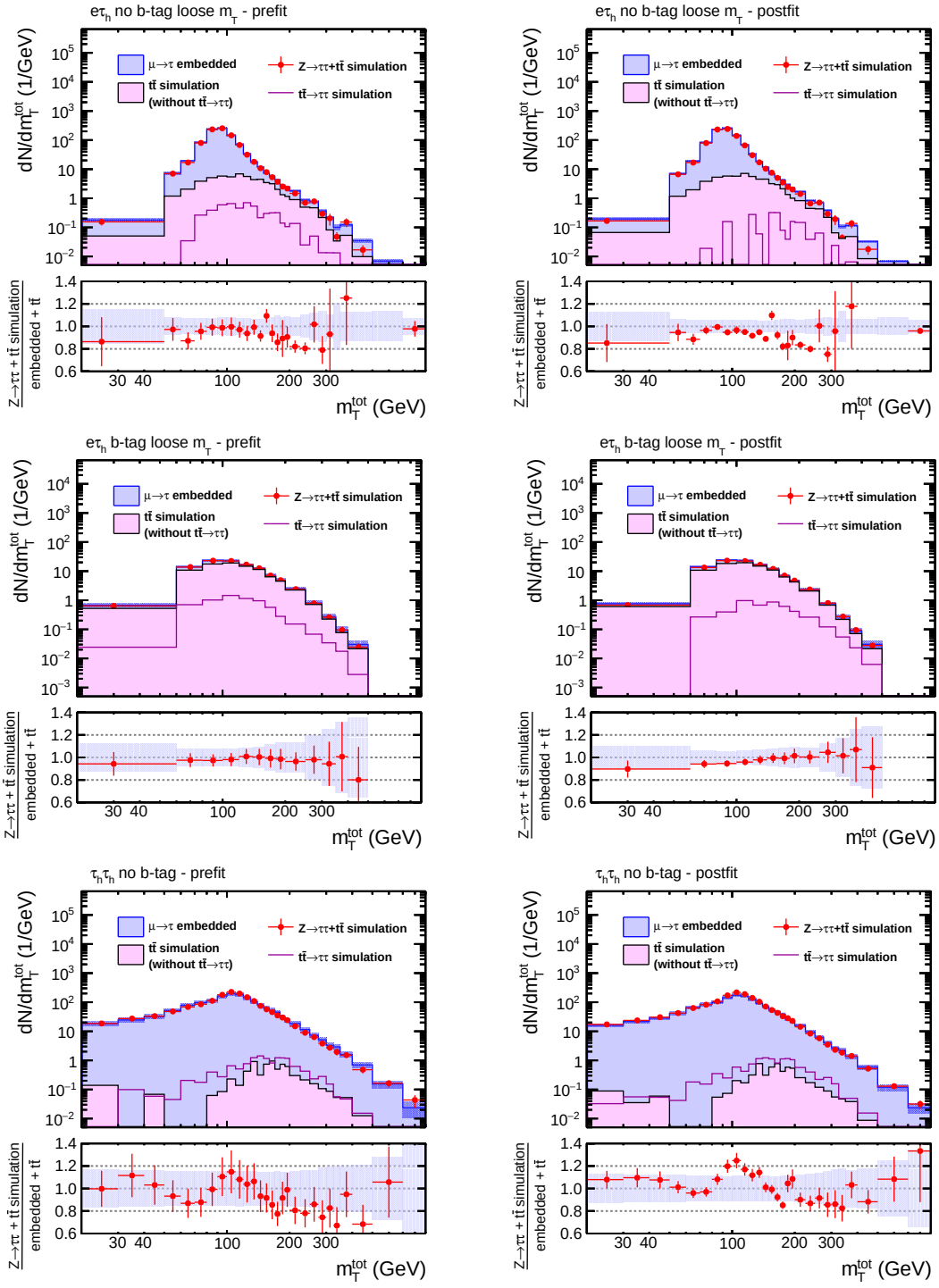


Figure A.16: Comparison of $m_{\mathrm{T}}^{\mathrm{tot}}$ distributions between the respective background estimates with the combined statistical and systematic uncertainties of $\mu \to \tau$ embedded events and simulated $Z \to \tau\tau$ events.

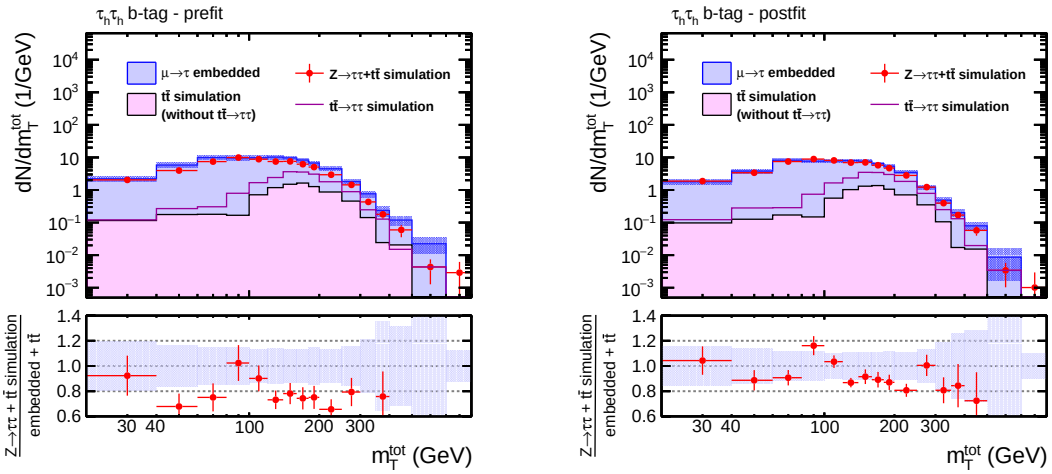


Figure A.17: Comparison of $m_{\mathrm{T}}^{\mathrm{tot}}$ distributions between the respective background estimates with the combined statistical and systematic uncertainties of $\mu \to \tau$ embedded events and simulated $Z \to \tau\tau$ events.

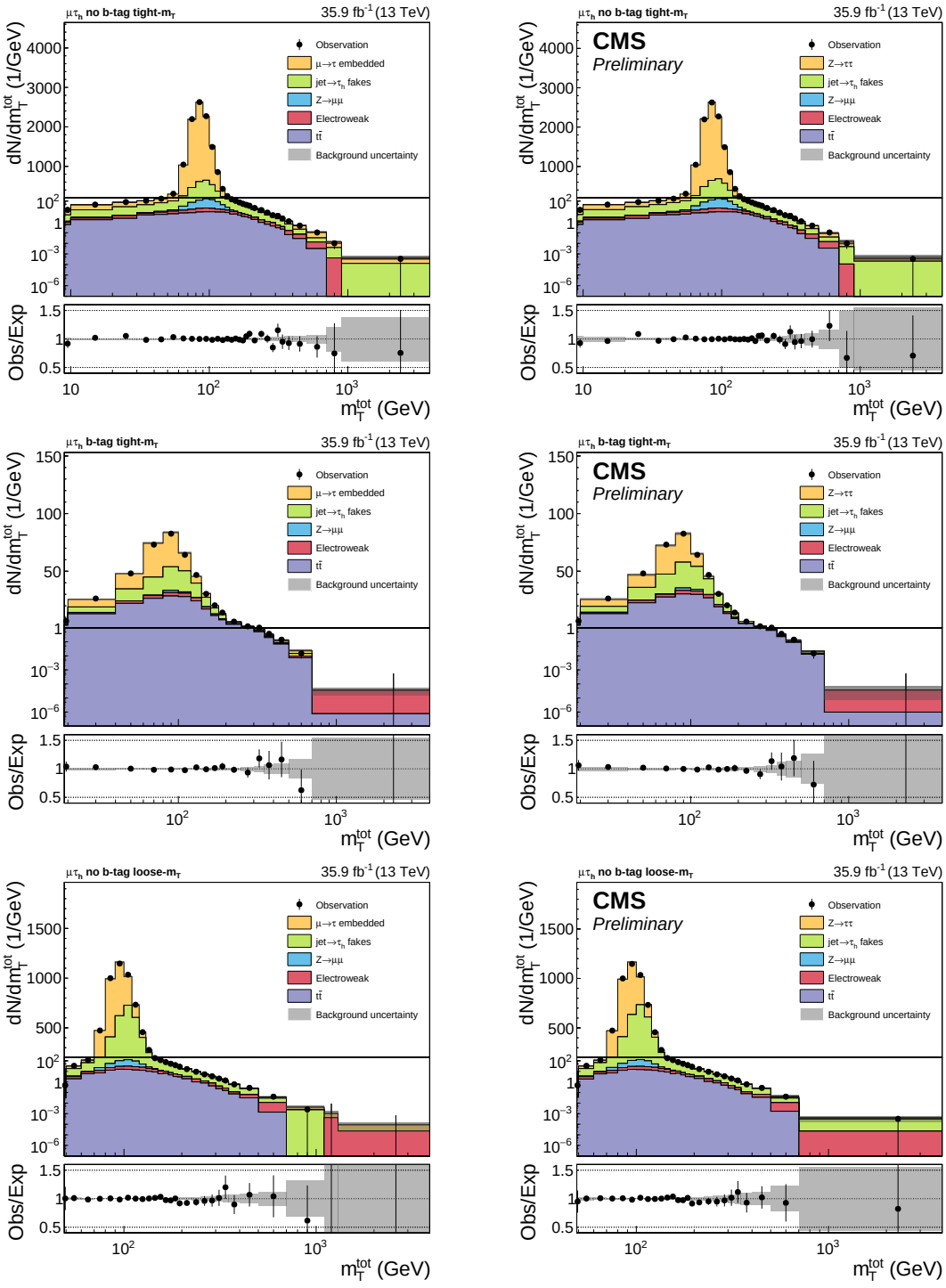


Figure A.18: m_T^{tot} distributions derived by using embedded events (left) and simulated $Z \to \tau\tau$ events (right).

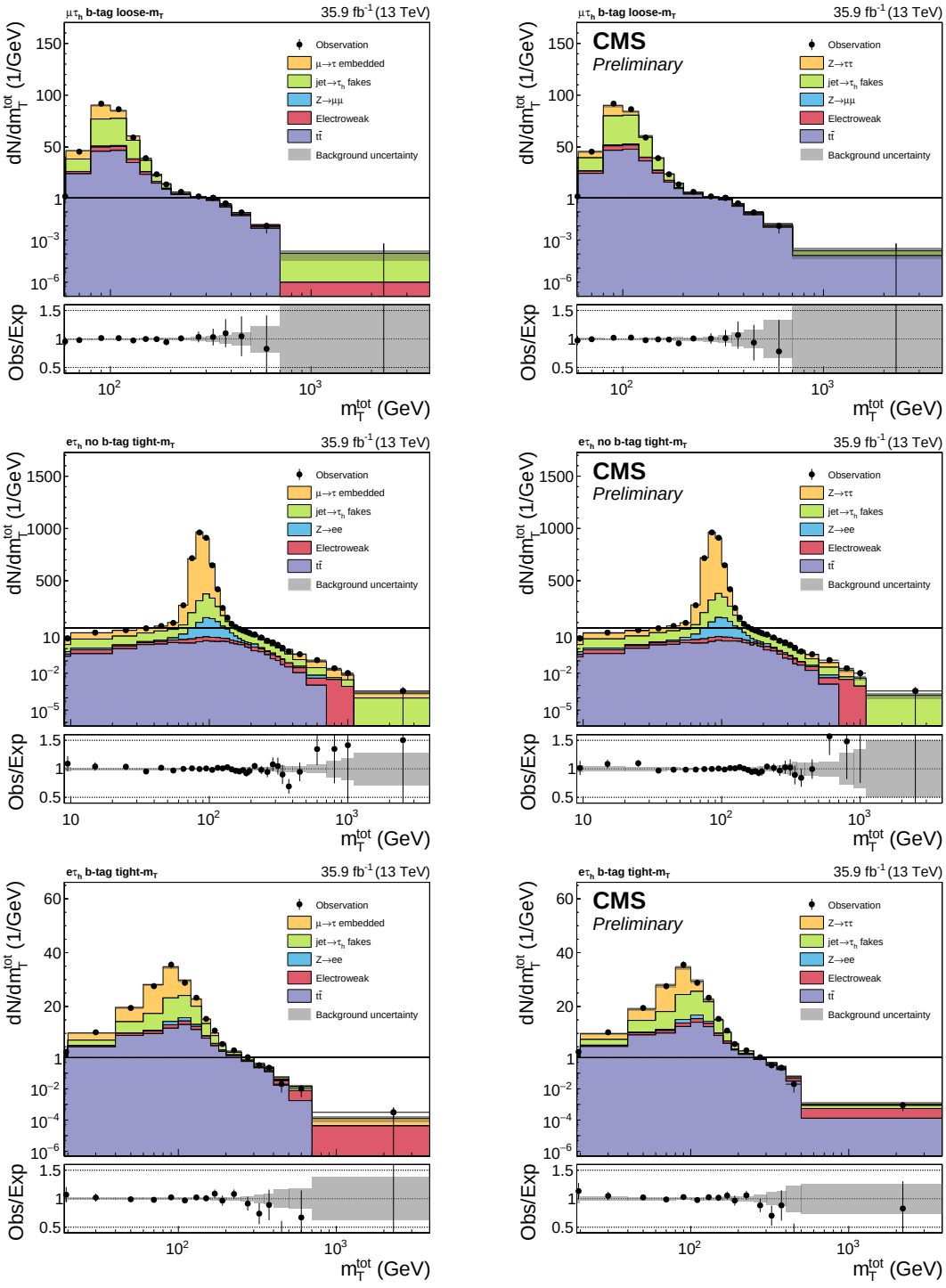


Figure A.19: m_T^{tot} distributions derived by using embedded events (left) and simulated $Z \to \tau\tau$ events (right).

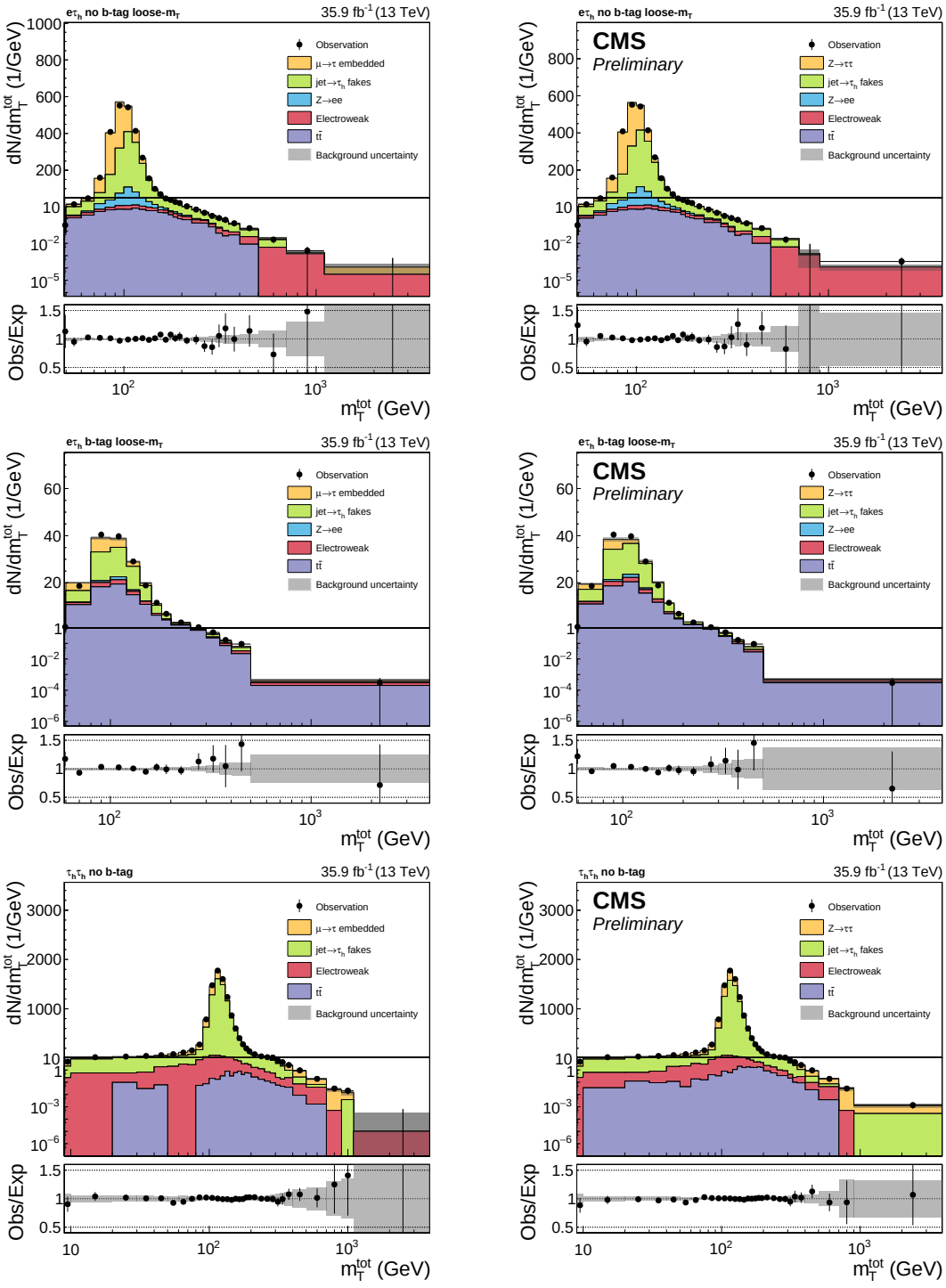


Figure A.20: m_T^{tot} distributions derived by using embedded events (left) and simulated $Z \to \tau\tau$ events (right).

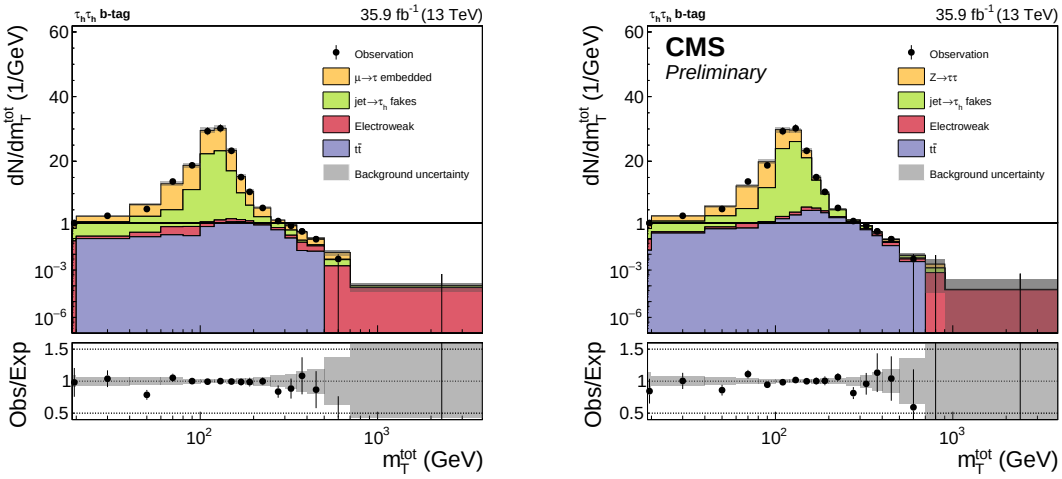


Figure A.21: m_T^{tot} distributions derived by using embedded events (left) and simulated $Z \to \tau\tau$ events (right).

A.5 Unblinded cross-check for all final states

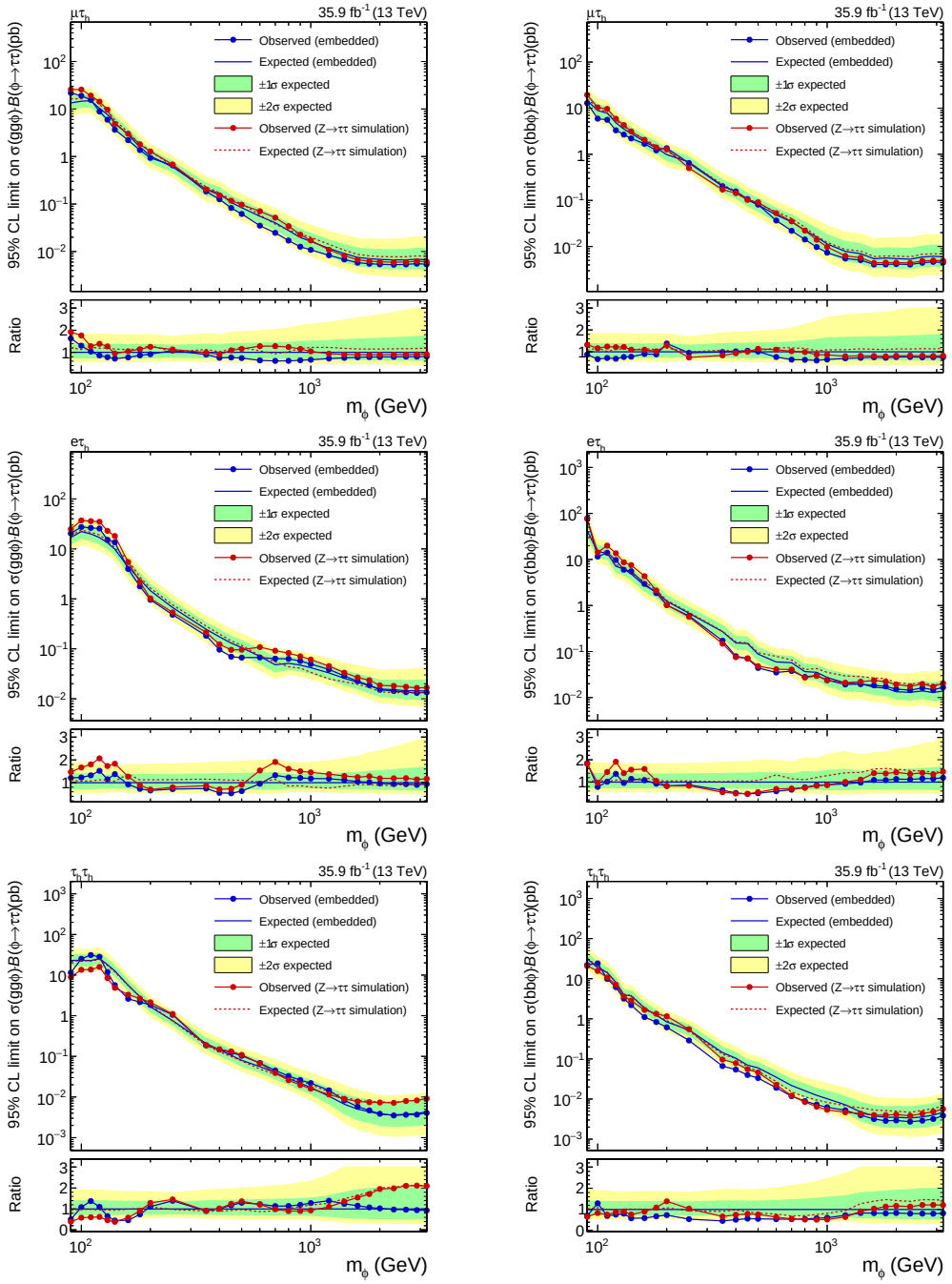


Figure A.22: Comparison of expected and observed limits between this analysis (embedded) and the published results ($Z \to \tau\tau$ simulation).

APPENDIX B

List of Figures

APPENDIX C

List of Tables

Acknowledgements

The results presented in this thesis could only be achieved thanks to the support I received during my year at the Institute for Experimental Particle Physics at the KIT. I would like to express my gratitude to Prof. Dr. Günter Quast for the fast integration is his working group and the excellent support I received. Furthermore, I want to thank Dr. Roger Wolf for his high level of commitment in the supervision of my thesis, including his constructive remarks during presentations and during the proof-reading of this thesis. My thanks also goes to Dr. Stefan Wayand for providing helpful guidance during the first eight months of my thesis. I am grateful to the whole team at the Higgs group, including René Caspart, Dr. Raphael Friese, Artur Gottmann and Sebastian Wozniewski for the great work atmosphere and the unhesitant support in both technical and scientific matters.
The datasets produced in the course of this thesis would not have been possible without the high-performance computing infrastructure as well as the know-how that was made available to me. In this regard, I want to thank our computing team at the institute including Christoph Heidecker and Matthias Schnepf for providing technical support, and the computing resources BwForCluster NEMO in Freiburg as well as the National Analysis Facility in Hamburg.

Erklärung der selbständigen Anfertigung der Dissertationsschrift

Hiermit erkläre ich, dass ich die Dissertationsschrift mit dem Titel

»*Cross-check of the CMS search for additional MSSM Higgs bosons in the di-τ final state using $\mu \to \tau$ embedded events*«

selbständig und unter ausschließlicher Verwendung der angegebenen Hilfsmittel angefertigt habe.

Janek Bechtel
Karlsruhe, den 23. Oktober 2017